Reconstructive approaches to one- and two-electron density matrix theory

by

John M. Herbert

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

(Chemistry)

at the

University of Wisconsin–Madison

2003
Reconstructive approaches to one- and two-electron density matrix theory

John M. Herbert

Under the supervision of Professor John E. Harriman

University of Wisconsin—Madison

Novel computational methods for electronic structure theory are explored, in which the fundamental variable is either the one- or the two-electron reduced density matrix (1- or 2-RDM), rather than the electronic wavefunction. A unifying theme among these methods is density matrix reconstruction, that is, decoupling approximations that express higher-order density matrices as functionals of lower-order ones.

On the 2-RDM side, a connected (extensive) version of the Contracted Schrödinger Equation (CSE) is developed, in which the basic unknowns are the RDM cumulants through order four. Reconstruction functionals that neglect the 3- and 4-RDM cumulants are examined and revealed to be significantly less accurate than suggested by previous minimal-basis results. Exact 3-RDM cumulants for some four-electron systems are calculated and found to be comparable in importance to unconnected products of lower-order cumulants. Decoupling approximations for the 3- and 4-RDM cumulants are developed based upon a renormalized, diagrammatic perturbation theory for the three- and four-particle Green’s functions, in which the effective, pairwise interaction is extracted from the two-particle cumulant. Diagram rules suitable for both the time-dependent and time-independent versions of this perturbation theory
are derived.

Reconstructive approaches to natural orbital (1-RDM) functional theory are also examined, wherein the 2-RDM is parametrized in terms of the natural orbitals and their (generally fractional) occupancies. It is demonstrated, at the theorem level, that proposed “corrected Hartree” and “corrected Hartree-Fock” natural orbital functionals necessarily violate positivity of the 2-RDM, which is closely related to their failure to respect antisymmetry. Calculations demonstrate that negative eigenvalues of the 2-RDM are associated with a large, stabilizing (but ultimately spurious) contribution to the energy. Nevertheless, a partially self-interaction-corrected version of the corrected Hartree functional is shown to yield high-quality potential energy curves—comparable to multireference configuration interaction—for the strongly-correlated diatomic molecules HF and C₂. A new class of functionals that preserve both antisymmetry and positivity are proposed and subjected to some preliminary tests.
To my parents, Anne and Mike, and my sister Erin, for making this possible.
Acknowledgements

The work described in this thesis has been carried out, of necessity, in the absence of other group members with whom I might have collaborated. However, let me acknowledge in this respect the patient ears of Tolga Gulmen, who has endured enough of my lectures (none-too-cleverly disguised as conversations) on density matrix theory that, for whatever it is worth, he probably deserves honorary status as a Harriman group member. As for my advisor, in retrospect I can fully appreciate the complete independence and intellectual free reign that I was granted as a graduate student, which I expect will contribute substantially to my future success.
<table>
<thead>
<tr>
<th>2.3.1</th>
<th>First-quantized formalism</th>
<th>35</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3.2</td>
<td>Second-quantized formalism</td>
<td>41</td>
</tr>
<tr>
<td>2.3.3</td>
<td>Hole matrices</td>
<td>45</td>
</tr>
<tr>
<td>2.3.4</td>
<td>Spin components</td>
<td>49</td>
</tr>
<tr>
<td>2.3.5</td>
<td>Some $N$-representability constraints</td>
<td>56</td>
</tr>
<tr>
<td>2.4</td>
<td>Intracules and extracules</td>
<td>62</td>
</tr>
<tr>
<td>2.4.1</td>
<td>Definitions</td>
<td>62</td>
</tr>
<tr>
<td>2.4.2</td>
<td>Computational formulas in a Slater-type basis</td>
<td>65</td>
</tr>
<tr>
<td>2.4.3</td>
<td>Approximate intracule-extracule isomorphism</td>
<td>72</td>
</tr>
<tr>
<td>2.5</td>
<td>Reduced density matrix cumulants</td>
<td>75</td>
</tr>
<tr>
<td>2.5.1</td>
<td>Additive versus multiplicative separability</td>
<td>75</td>
</tr>
<tr>
<td>2.5.2</td>
<td>Cumulant formalism</td>
<td>80</td>
</tr>
<tr>
<td>2.5.3</td>
<td>Extensivity</td>
<td>84</td>
</tr>
<tr>
<td>2.5.4</td>
<td>Independence of the cumulants</td>
<td>87</td>
</tr>
<tr>
<td>2.6</td>
<td>Contraction of Grassmann products</td>
<td>89</td>
</tr>
<tr>
<td>2.6.1</td>
<td>Contraction theorems</td>
<td>90</td>
</tr>
<tr>
<td>2.6.2</td>
<td>Generalized contraction theorems</td>
<td>102</td>
</tr>
<tr>
<td>2.6.3</td>
<td>Grassmann product ansatz</td>
<td>107</td>
</tr>
<tr>
<td>2.7</td>
<td>References</td>
<td>110</td>
</tr>
</tbody>
</table>

3 Equations to determine the two-electron density matrix

3.1 Introduction

3.2 The contracted Schrödinger equation

3.2.1 Reduced eigenvalue equations
3.2.2 Density matrix reconstruction functionals ........................................ 134

3.3 An explicitly extensive CSE .............................................................. 137
  3.3.1 Connectivity, diagrams, and the CSE .......................................... 139
  3.3.2 Cancellation of unconnected terms ............................................ 147
  3.3.3 Discussion of the connected equations ....................................... 150

3.4 Summary ............................................................................................. 153

3.5 References ............................................................................................ 154

4 Reconstruction functionals for the three- and four-electron density matrices ......................................................... 160
  4.1 Introduction .......................................................................................... 160
  4.2 Review of proposed reconstruction functionals ................................... 168
    4.2.1 Valdemoro’s unconnected approximations .................................... 168
    4.2.2 Mazziotti’s cumulant-based reconstructions .................................. 174
  4.3 Assessment of the neglected terms ..................................................... 176
    4.3.1 Tests of the Valdemoro 4-RDM functional .................................... 177
    4.3.2 Iterative improvement via CSE(2) ............................................... 189
    4.3.3 Can the cumulant hierarchy be truncated? .................................... 195
  4.4 Renormalized perturbation theory for many-particle propagators ......... 207
    4.4.1 Overview ...................................................................................... 208
    4.4.2 Many-body perturbation theory for the propagators .................... 213
    4.4.3 Connectivity and topology of diagrams ....................................... 226
    4.4.4 Ladder-type expansions for the three-particle propagator ............ 232
    4.4.5 Ladder-type expansions for the four-particle propagator ............ 242
<table>
<thead>
<tr>
<th>4.5</th>
<th>Reconstruction functionals from renormalized perturbation theory</th>
<th>250</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.5.1</td>
<td>The quasi-pair ladder approximation</td>
<td>250</td>
</tr>
<tr>
<td>4.5.2</td>
<td>Factorization approximations for QPL diagrams</td>
<td>256</td>
</tr>
<tr>
<td>4.5.3</td>
<td>Higher-order QPL approximations</td>
<td>261</td>
</tr>
<tr>
<td>4.5.4</td>
<td>Iteration and higher-order correlations</td>
<td>265</td>
</tr>
<tr>
<td>4.6</td>
<td>Summary, outlook, and future directions</td>
<td>267</td>
</tr>
<tr>
<td>4.7</td>
<td>References</td>
<td>273</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5</th>
<th>(One-electron) density matrix functional theory</th>
<th>282</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>282</td>
</tr>
<tr>
<td>5.1.1</td>
<td>Why DMFT?</td>
<td>283</td>
</tr>
<tr>
<td>5.1.2</td>
<td>Reconstructive approach</td>
<td>297</td>
</tr>
<tr>
<td>5.2</td>
<td>Reconstruction functionals for the 2-RDM</td>
<td>302</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Specification of the functionals</td>
<td>302</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Origins of the reconstruction functionals</td>
<td>305</td>
</tr>
<tr>
<td>5.2.3</td>
<td>N-representability</td>
<td>315</td>
</tr>
<tr>
<td>5.3</td>
<td>Qualitative analysis of reconstructed pair densities</td>
<td>322</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Two-electron conditional densities</td>
<td>325</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Intracule densities</td>
<td>331</td>
</tr>
<tr>
<td>5.3.3</td>
<td>Extracule densities</td>
<td>334</td>
</tr>
<tr>
<td>5.3.4</td>
<td>Pair densities from optimized density matrices</td>
<td>336</td>
</tr>
<tr>
<td>5.4</td>
<td>Variational procedure</td>
<td>341</td>
</tr>
<tr>
<td>5.5</td>
<td>N-representability, variational stability, and basis-set dependence</td>
<td>352</td>
</tr>
<tr>
<td>5.5.1</td>
<td>Variational stability</td>
<td>355</td>
</tr>
</tbody>
</table>
5.5.2 \( N \)-representability .................................................. 365

5.6 Application to strongly correlated systems .......................... 376

5.7 A novel, generalized Hartree-Fock functional ....................... 383
  5.7.1 Derivation ............................................................. 384
  5.7.2 \( Q \)- and \( G \)-conditions for GHF(\(x, y\)) ......................... 393

5.8 Summary and conclusion .................................................. 402

5.9 References ................................................................. 408

APPENDICES 424

A Storage and indexing of many-particle tensors 425
  A.1 Non-redundant indexing schemes ................................. 426
  A.2 References ............................................................. 435

B Additional formulas for Grassmann products 436

C Algebraic form of ICSE(2) 442
  C.1 Algebraic translation of the diagrammatic equation .......... 443
  C.2 References ............................................................. 445

D Reduced density matrices from multideterminant wavefunctions 446
  D.1 Density matrix formulas ............................................ 447
  D.2 References ............................................................. 451

E Atomic integrals over Slater-type functions 452
  E.1 One-electron integrals .............................................. 452
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>E.2 Two-electron integrals</td>
<td>453</td>
</tr>
<tr>
<td>E.3 References</td>
<td>454</td>
</tr>
</tbody>
</table>
# List of Tables

## Chapter 1

1.1 Approximate number of independent parameters to determine the wavefunction or 2-RDM, for an $N$-electron singlet state in a basis of $\nu$ orbitals. 5

1.2 Approximate number of independent parameters to determine the wavefunction or 2-RDM, for the water and ethylene molecules. 5

## Chapter 4

4.1 Parameters employed in reconstruction of $D_4^g$. 181

4.2 Largest elements of the CI and reconstructed 2-RDMs, for the Be ground state. 183

4.3 Comparison of CI and reconstructed 2-RDMs for Be, following renormalization. 186

4.4 Energy errors induced by errors in the largest elements of the Be 2-RDM. 187

4.5 Spectrum of $D_2^{g\alpha}$ and reconstructed, renormalized approximations thereto, for the Be ground state. 188

4.6 Spectrum of $D_2^{g\beta}$ and reconstructed, renormalized approximations thereto, for the Be ground state. 189
4.7 Natural occupation numbers of the CI and reconstructed, renormalized 1-RDMs, for the Be ground state. ............................................. 190

4.8 First iterative correction $D_2^{(1)}$ to the CI 2-RDM $D_2^{(0)}$ for the ground state of Be, using Valdemoro’s reconstruction functionals. ................. 193

4.9 Selected matrix elements of $D_3$, $\Delta_3$, $\Sigma_3$ and the reduced Hamiltonian $W_{123}$ for Be/6-31G*. .................................................. 201

4.10 Decomposition of the Be and LiH electronic energies into connected and unconnected contributions, at the three-particle level. ............... 204

Chapter 5 .............................................................................. 282

5.1 Specification of reconstruction functionals, according to Eq. (5.18). ... 304

5.2 Eigenvalues of $\hat{D}_2^{\alpha\alpha}$ and $\hat{Q}_2^{\alpha\alpha}$ for 2-RDM reconstruction functionals. ... 319

5.3 Eigenvalues of $\hat{Q}_2^{\alpha\alpha}$ and $\hat{G}_2^{\alpha\alpha}$ for 2-RDM reconstruction functionals. ... 319

5.4 Total spin expectation values from reconstructed 2-RDMs, within a spin-restricted formalism........................................... 322

5.5 Unoptimized energies for the Be ground state, from 2-RDMs reconstructed CI occupation numbers. .......................... 337

5.6 FCI energies for basis sets used in this section. .......................... 354

5.7 Correlation energies for Be, in selected basis sets. ..................... 355

5.8 LiH($R$ = 1.5953 Å) correlation energies in selected basis sets. ........ 356

5.9 Largest $\alpha$-spin occupations numbers for Be, excluding a unit-occupied core orbital. ....................................................... 361

5.10 Spectrum of $\hat{D}_2^{\alpha\alpha}$ and $\hat{G}_2^{\alpha\alpha}$, for Be/6-311G(2df). .................. 366

5.11 Spectrum of $\hat{D}_2^{\alpha\alpha}$ for LiH/6-311G($d,2p$), at several values of $R$. .... 367
5.12 Spectrum of $\hat{G}_2^{\alpha\alpha}$ for LiH/6-311G(d, 2p), at several values of $R$. 368
5.13 Theoretical bounds on the reconstructed eigenvalues of $\hat{D}_2^{\alpha\alpha}$. 369
5.14 Decomposition of DMFT electronic energies for Be/6-311G(2df) and LiH($R = 1.5953$ Å)/6-311+G(d, 3p). 373
5.15 Smallest eigenvalues of $\hat{Q}_2^{\alpha\alpha}$ for several values of the reconstruction parameter $x$. 398
5.16 Smallest eigenvalues of $\hat{Q}_2^{\alpha\beta}$ for several values of the reconstruction parameter $y$. 399
5.17 Smallest eigenvalues of $\hat{G}_2^{\alpha\alpha}$ for several values of the reconstruction parameter $x$. 400
5.18 Smallest eigenvalues of $\hat{G}_2^{\alpha\beta}$ for several values of the reconstruction parameter $y$. 401
5.19 Summary of the spectra of $\hat{D}_2^{\alpha\alpha}$, $\hat{Q}_2^{\alpha\alpha}$, and $\hat{G}_2^{\alpha\alpha}$ for the functionals introduced in Section 5.2. 403
5.20 Number of negative eigenvalues of $\hat{Q}_2$ and $\hat{G}_2$, for the GHF($x, y$) reconstruction. 404

Appendix A 425

A.1 Storage requirements for spin components of the 2-, 3-, and 4-RDMs, as functions of the number of orbitals $\nu$. 435
List of Figures

Chapter 1

1.1 The hierarchy of quantum-mechanical distribution functions. 11

Chapter 2

2.1 An eigenvalue $n_k$ of $\Delta_1$ as a (double-valued) function of the corresponding eigenvalue $e_k$ of $\text{tr}_2 \Delta_2$. 89

Chapter 3

3.1 Glossary of diagram elements for representing the CSE. 141

Chapter 4

4.1 Ratios of connected to unconnected matrix elements of the $\alpha\alpha\beta$ spin component of $D_3$, for Be. 200
4.2 Magnitudes of the energy contributions from the connected (cumulant) and unconnected parts of the 3-RDM spin components, for Be/6-31G*. 206
4.3 Framework for constructing second-order two-particle diagrams. 218
4.4 Four diagrams in the $\overline{D}_1$ perturbation series. 221
4.5 Two-particle diagrams at second order in $\hat{V}$.

4.6 Framework for constructing $n$th-(renormalized)-order terms in the expansion of $\mathcal{D}_{j_1, \ldots, j_p, k_1, \ldots, k_p}$.

4.7 Two topologically inequivalent embeddings of the same three-particle connectivity.

4.8 An embedding of the connectivity $\{S'_1, S'_2, S'_3\}$ introduced in the text.

4.9 Diagrams in which $v_1^L$ and $v_1^R$ are connected by the same quasiparticle line.

4.10 A three-particle ladder diagram in disguise.

4.11 Diagrammatic representations of various steps in the algorithmic construction of $n$th-order, three-particle ladders.

4.12 Two possibilities for connecting the second interaction in Fig. 4.11(a).

4.13 All independent third-order, three-particle ladders.

4.14 All independent fourth-order, three-particle ladders.

4.15 Steps in the four-particle counting algorithm (first case).

4.16 Possible nascent connectivities associated with connecting the first $m+1$ interactions in Fig. 4.15(b).

4.17 Illustration of the four-particle counting algorithm (second case).

4.18 All independent four-particle ladders at third order.

4.19 Fourth-order, four-particle ladder diagrams arising from Case 1 of the generating algorithm.

4.20 Fourth-order, four-particle ladder diagrams arising from Case 2 of the generating algorithm.
4.21 A few diagrams in the expansion of $\mathbf{X}_2$, at second order in the dressed interaction. .................................................. 251

4.22 Selected diagrams in the expansion of $\mathbf{X}_3$. .................................................. 253

4.23 Three-particle QPL diagrams, through fourth order. ......................... 254

4.24 Third- and fourth-order, four-particle QPL diagrams. ......................... 262

4.25 Alternative ways to apply the NY factorization to the first diagram in Fig. 4.24. .................................................. 264

4.26 Diagrams illustrating the connection between reconstruction and solution of the CSEs or ICSEs. .................................................. 266

Chapter 5

5.1 Spin-restricted STO-6G potential curves and CI natural occupation numbers for $\text{H}_2$. .................................................. 289

5.2 Radial distribution function for the Be ground state. ......................... 326

5.3 Two-electron charge density for the SIC-CH(4/5) functional with $r_1 = 0.1 \text{ } \text{a}_0$ and $\theta_1 = 0 = \varphi_1$. .................................................. 327

5.4 Pair densities for same-spin electrons, integrated over $\varphi_2$, with $\theta_1 = 0 = \varphi_1$ and $r_1 = 0.2655 \text{ } \text{a}_0$. .................................................. 328

5.5 Pair densities for same-spin electrons, with $\theta_1 = 0 = \varphi_1$ and $r_1 = 0.2655 \text{ } \text{a}_0$. .................................................. 328

5.6 Pair densities for same-spin electrons, with $\theta_1 = 0 = \varphi_1$ and $r_1 = 0.2655 \text{ } \text{a}_0$, the most probable radial electron distance. ......................... 329

5.7 Two-electron correlation densities for Be, integrated over $\theta_2$, with $r_1 = 0.2655 \text{ } \text{a}_0$ and $\theta_1 = 0 = \varphi_1$. .................................................. 331
5.8 Selected region of the radial intracule density for Be. 332
5.9 Radial correlation intracule for Be. 333
5.10 Radial correlation intracule for Be. 333
5.11 Radial correlation intracule for Be. 334
5.12 Radial intracule densities for parallel-spin electrons. 335
5.13 Radial intracule densities for two electrons of opposite spin. 335
5.14 Radial (spherically averaged) extracule densities for two electrons of the same spin. 336
5.15 Be parallel-spin intracule densities calculated using CI and DMFT occupation numbers. 338
5.16 Be opposite-spin intracule densities calculated using CI and DMFT occupation numbers. 339
5.17 Be opposite-spin extracule densities calculated using optimized DMFT occupation numbers. 339
5.18 Be parallel-spin extracule densities calculated using CI and DMFT occupation numbers. 340
5.19 LiH/6-311G energy minimization with uniform scaling of the \( n_i \) following each converged occupancy optimization. 349
5.20 DMFT versus FCI energies for Be. 357
5.21 DMFT versus FCI energies for LiH (at \( R = 1.5953 \) Å). 358
5.22 Be correlation energies, as fractions of the FCI correlation energy. 358
5.23 LiH correlation energies (at \( R = 1.5953 \) Å), as fractions of the FCI correlation energy. 359
5.24 Largest LiH/6-311G(d, 2p) occupation numbers, excluding a unit-occupied
core orbital. ........................................ 362
5.25 LiH potential curves. ......................................................... 363
5.26 LiH potential curves. ......................................................... 364
5.27 LiH/6-311G(·, 3p) potential curves. ......................................................... 365
5.28 FH/6-31+G** potentials and orbital energies. ......................................................... 379
5.29 6-31+G** potential curves for the FH molecule. ......................................................... 380
5.30 6-31G* potential curves for the C2 molecule. ......................................................... 381
5.31 Normalizing factors for the GHF(x, y) functional. ......................................................... 391
5.32 Same-spin intracule densities for Be, reconstructed using natural occu-
pation numbers from Bunge’s CI wavefunction. ......................................................... 392
5.33 Same-spin extracule densities for Be, reconstructed using natural occu-
pation numbers from Bunge’s CI wavefunction. ......................................................... 393
5.34 Smallest eigenvalues of $\hat{Q}_2^{\alpha}$ and $\hat{Q}_2^{\beta}$, as functions of the reconstruction
parameters $x$ and $y$. ......................................................... 397
5.35 Smallest eigenvalue of $\hat{G}_2^{\alpha\beta}$, as a function of the reconstruction param-
eter $y$. ......................................................... 397

Appendix A

A.1 Indexing scheme for antisymmetric ordered triples $(i, j, k)$. ......................................................... 430
## List of Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Page No.</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>( i ), ( i )</td>
<td>52</td>
<td>( \alpha ) - and ( \beta )-spin orbital indices</td>
</tr>
<tr>
<td>( a_0 )</td>
<td>33</td>
<td>Atomic unit of distance (bohr)</td>
</tr>
<tr>
<td>( E_h )</td>
<td>33</td>
<td>Atomic unit of energy (hartree)</td>
</tr>
<tr>
<td>( \text{tr}_{n,n+1,...} )</td>
<td>36</td>
<td>Partial trace operator</td>
</tr>
<tr>
<td>( p_{q^+} )</td>
<td>36</td>
<td>Contraction operator</td>
</tr>
<tr>
<td>( \wedge )</td>
<td>47</td>
<td>Antisymmetrized (Grassmann) tensor product</td>
</tr>
<tr>
<td>( ^n )</td>
<td>103</td>
<td>Partial tensor product on ( n ) indices</td>
</tr>
<tr>
<td>( \ominus )</td>
<td>91</td>
<td>Deletion operator</td>
</tr>
<tr>
<td>( \mathcal{P} )</td>
<td>105</td>
<td>Projection operator for pure contractions</td>
</tr>
<tr>
<td>( \mathcal{T} )</td>
<td>105</td>
<td>Projection operator for transvections</td>
</tr>
<tr>
<td>( \mathcal{T} )</td>
<td>44</td>
<td>Time-ordering operator</td>
</tr>
<tr>
<td>( \mathcal{N} )</td>
<td>215</td>
<td>Normal-ordering operator</td>
</tr>
<tr>
<td>( \epsilon(\pi) )</td>
<td>47</td>
<td>Parity of the permutation ( \pi )</td>
</tr>
<tr>
<td>( \mathbb{S}_n )</td>
<td>47</td>
<td>Symmetric group of order ( n! )</td>
</tr>
<tr>
<td>( \hat{P}_n, \overline{P}_n, \hat{P}_n )</td>
<td>47</td>
<td>Sum of signed permutations in ( \mathbb{S}_n )</td>
</tr>
<tr>
<td>( \hat{A}_n, \overline{A}_n, \hat{A}_n )</td>
<td>47</td>
<td>Antisymmetrizing projection operators on ( n ) indices</td>
</tr>
<tr>
<td>( x = (r, \xi) )</td>
<td>32</td>
<td>Composite space-spin coordinate for an electron</td>
</tr>
<tr>
<td>( I(u), I(u) )</td>
<td>63</td>
<td>Intracule (relative coordinate) density</td>
</tr>
<tr>
<td>( X(R), X(R) )</td>
<td>63</td>
<td>Extracule (center-of-mass) density</td>
</tr>
<tr>
<td>( D_p, \hat{D}_p, \mathbf{D}_p )</td>
<td>35,42</td>
<td>( p )-particle reduced density matrix</td>
</tr>
<tr>
<td>( Q_p, \hat{Q}_p, Q_p )</td>
<td>46,47</td>
<td>( p )-hole reduced density matrix</td>
</tr>
<tr>
<td>( \rho_p, \rho_{p}^{\alpha\beta...\gamma} \rho_p )</td>
<td>39</td>
<td>( p )-particle densities</td>
</tr>
<tr>
<td>( D_p^\circ )</td>
<td>63</td>
<td>Spin-traced ( p )-particle charge density matrix</td>
</tr>
<tr>
<td>( \Delta_p, \hat{\Delta}_p, \Delta_p )</td>
<td>81</td>
<td>( p )-particle cumulant (connected part of ( D_p ))</td>
</tr>
<tr>
<td>( \Sigma_p, \Sigma_{p}, \Sigma_p )</td>
<td>172</td>
<td>Unconnected part of ( D_p )</td>
</tr>
<tr>
<td>( \overline{D}_p, \mathbf{D}_p )</td>
<td>44</td>
<td>( p )-particle propagator (Green’s function)</td>
</tr>
<tr>
<td>( W_{12...p} )</td>
<td>39</td>
<td>( p )-particle reduced operator or reduced Hamiltonian</td>
</tr>
<tr>
<td>( \Omega_p, \hat{\Omega}_p )</td>
<td>133</td>
<td>( p )-particle energy density matrix</td>
</tr>
<tr>
<td>( \Omega_p, \Omega_p )</td>
<td>147</td>
<td>Connected and unconnected parts of ( \Omega_p )</td>
</tr>
<tr>
<td>Symbol</td>
<td>Page No.</td>
<td>Meaning</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>$G_2, G_2$</td>
<td>58</td>
<td>Particle-hole density matrix</td>
</tr>
<tr>
<td>$D_4^c$</td>
<td>90</td>
<td>Four-particle Grassmann product ansatz</td>
</tr>
<tr>
<td>$\mathcal{Y}_{l,m;\ell',m'}^{(k)}$</td>
<td>68</td>
<td>Gaunt coefficient</td>
</tr>
<tr>
<td>$j_\ell(x)$</td>
<td>68</td>
<td>Spherical Bessel function of order $\ell$</td>
</tr>
<tr>
<td>$\mathcal{H}_\ell$</td>
<td>69</td>
<td>Hankel transform of order $\ell$</td>
</tr>
<tr>
<td>$\tilde{J}_i, \tilde{K}_i$</td>
<td>287</td>
<td>Hartree-Fock Coulomb and exchange operators</td>
</tr>
</tbody>
</table>
## List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Page no.</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGP</td>
<td>60</td>
<td>Antisymmetrized geminal power</td>
</tr>
<tr>
<td>BFGS</td>
<td>346</td>
<td>Broyden-Fletcher-Goldfarb-Shanno</td>
</tr>
<tr>
<td>CC</td>
<td>137</td>
<td>Coupled-cluster</td>
</tr>
<tr>
<td>CCSD</td>
<td>353</td>
<td>Coupled-cluster singles-doubles</td>
</tr>
<tr>
<td>CCSD(T)</td>
<td>353</td>
<td>Coupled-cluster singles, doubles, and non-iterative triples</td>
</tr>
<tr>
<td>CD</td>
<td>51</td>
<td>Charge density</td>
</tr>
<tr>
<td>CH</td>
<td>303</td>
<td>Corrected Hartree</td>
</tr>
<tr>
<td>CHF</td>
<td>303</td>
<td>Corrected Hartree-Fock</td>
</tr>
<tr>
<td>CI</td>
<td>18</td>
<td>Configuration interaction</td>
</tr>
<tr>
<td>CISD</td>
<td>353</td>
<td>Configuration interaction singles-doubles</td>
</tr>
<tr>
<td>CSE</td>
<td>125</td>
<td>Contracted Schrödinger equation</td>
</tr>
<tr>
<td>DFT</td>
<td>4</td>
<td>Density functional theory</td>
</tr>
<tr>
<td>DMFT</td>
<td>282</td>
<td>Density matrix functional theory</td>
</tr>
<tr>
<td>FCI</td>
<td>3</td>
<td>Full configuration interaction</td>
</tr>
<tr>
<td>GF</td>
<td>43</td>
<td>Green’s function</td>
</tr>
<tr>
<td>GGA</td>
<td>376</td>
<td>Generalized gradient approximation</td>
</tr>
<tr>
<td>HF</td>
<td>3</td>
<td>Hartree-Fock</td>
</tr>
<tr>
<td>HFB</td>
<td>309</td>
<td>Hartree-Fock-Bogoliubov</td>
</tr>
<tr>
<td>HRDM</td>
<td>46</td>
<td>Hole reduced density matrix</td>
</tr>
<tr>
<td>ICSE</td>
<td>148</td>
<td>Irreducible contracted Schrödinger equation</td>
</tr>
<tr>
<td>KS</td>
<td>283</td>
<td>Kohn-Sham</td>
</tr>
<tr>
<td>LDA</td>
<td>284</td>
<td>Local density approximation</td>
</tr>
<tr>
<td>MCHF</td>
<td>303</td>
<td>Modified corrected Hartree-Fock</td>
</tr>
<tr>
<td>MCSCF</td>
<td>350</td>
<td>Multiconfigurational self-consistent field</td>
</tr>
<tr>
<td>MP2</td>
<td>353</td>
<td>Second-order Møller-Plesset perturbation theory</td>
</tr>
<tr>
<td>MR-CISD</td>
<td>379</td>
<td>Multireference configuration interaction singles-doubles</td>
</tr>
<tr>
<td>NO</td>
<td>40</td>
<td>Natural orbital</td>
</tr>
<tr>
<td>NSO</td>
<td>40</td>
<td>Natural spin orbital</td>
</tr>
<tr>
<td>NY</td>
<td>163</td>
<td>Nakatsuji-Yasuda</td>
</tr>
<tr>
<td>QPL</td>
<td>252</td>
<td>Quasi-pair ladder</td>
</tr>
<tr>
<td>RDM</td>
<td>34</td>
<td>Reduced density matrix</td>
</tr>
<tr>
<td>Acronym</td>
<td>Page no.</td>
<td>Meaning</td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>RDMC</td>
<td>81</td>
<td>Reduced density matrix cumulant</td>
</tr>
<tr>
<td>RHF</td>
<td>288</td>
<td>Restricted Hartree-Fock</td>
</tr>
<tr>
<td>RKS</td>
<td>288</td>
<td>Restricted Kohn-Sham</td>
</tr>
<tr>
<td>SD</td>
<td>86</td>
<td>Single determinant</td>
</tr>
<tr>
<td>SIC</td>
<td>304</td>
<td>Self-interaction correction</td>
</tr>
<tr>
<td>STO</td>
<td>65</td>
<td>Slater-type orbital</td>
</tr>
</tbody>
</table>
Papers Published


Chapter 1

Density matrices and electronic structure theory

1.1 Reduced descriptions of electronic structure

Electrons are responsible for chemistry. This ingenuous statement, true basically as a matter of definition, explains at once why electronic structure theory is relevant and also why it is difficult. By their nature, electrons are quantum-mechanical objects, and chemistry—the marriage of atoms, via the quantum superposition phenomenon, to form molecules—is an inherently quantum-mechanical science.\(^1\) Classical physics cannot explain even the simplest chemical reaction, \(\text{H} + \text{H} \rightarrow \text{H}_2\). In the semiclassical pictures of molecular dynamics\(^2\,^3\) and molecular spectroscopy,\(^4\) the nuclei propagate according to essentially classical equations of motion, but they do so on a potential energy surface of quantum-mechanical origin, which nature obtains—and which the theoretical chemist \textit{ought} to obtain—by solving the electronic Schrödinger equation. That such semiclassical descriptions often compare well with experiment is a
testament to the comparatively classical nature of the nuclear problem relative to the electronic one. Even in cases where the nuclear dynamics is not semiclassical, due for example to nearly degenerate electronic potential surfaces, all of the quantities necessary to specify and propagate the vibronic (nuclear vibrations plus electronic degrees of freedom) Hamiltonian are still derivable from clamped-nuclei, Born-Oppenheimer electronic structure calculations only, e.g., as derivatives of the Born-Oppenheimer electronic wavefunctions with respect to the nuclear coordinates.

Thus electronic structure theory informs the whole of chemistry. Insofar as precise, quantitative explanations for chemical phenomena depend upon computation of electronic wavefunctions, however, the deck is stacked against the theorist from the start, because electronic structure theory is inherently more demanding than other problems in theoretical chemistry whose solution also requires elements of quantum mechanics. Certainly, computational quantum mechanics in multiple dimensions is challenging no matter what the degrees of freedom represent, but where nuclear degrees of freedom are concerned, there exists a wealth of of interesting chemical and physical problems whose essence can be reduced to, say, 5–10 quantum-mechanical coordinates, possibly coupled to additional, classical degrees of freedom. Indeed, the theoretical chemist interested in reaction dynamics or vibrational wavefunctions can compile an impressive resumé without ever going beyond a few degrees of freedom. In contrast, the helium atom already has three electronic degrees of freedom (spin not included) and the hydrogen molecule has four, but neither is especially engaging in its own right. We can therefore summarize the difficulty with electronic structure in two observations: first, the problems are intrinsically quantal; and second, there are no interesting, few-dimensional applications.
Although eighty years of theoretical work (aided by several decades of extraordinary growth in microprocessor speed) have produced a variety of sophisticated procedures for calculating electronic wavefunctions, these methods are severely limited by a steep rise in computational expense with respect to system size. For the \textit{ab initio} (that is, non-empirical) methods in common use, this cost increases as $\nu^4 - \nu^7$, where $\nu$ indicates the size of the underlying orbital basis, roughly proportional to the number of electrons $N$. Traditional implementations of correlated electronic structure methods are therefore limited at present to a few tens of electrons, up to perhaps a few hundred electrons at the mean-field Hartree-Fock (HF) level.

Ultimately these limitations stem from the fact that the electronic wavefunction depends upon $N$ electronic coordinates, and therefore the number of unknowns that characterize this wavefunction scales exponentially with the size of the system. In this work, we deal only with the model problem defined by the introduction of a finite spin-orbital basis. Within this model, the exact solutions to Schrödinger’s equation are determinantal expansions, namely, the full configuration interaction (FCI) wavefunctions. For $\nu$ orbitals ($2\nu$ spin orbitals) and $N$ electrons, a crude estimate of the number of free parameters in the FCI wavefunction is $\binom{2\nu}{N}$, though the actual number is somewhat less than this because the $N$ electrons must be coupled to an $\hat{S}^2$ eigenstate. Additional spatial or angular momentum symmetries may reduce the number of parameters further, but taking into account only the $\hat{S}^2$ symmetry, the number of unknowns in the FCI wavefunction is

\begin{equation}
N_{\text{config}} = \frac{2S + 1}{\nu + 1} \left( \frac{\nu + 1}{N/2 - S} \right) \left( \frac{\nu + 1}{N/2 + S + 1} \right)
\end{equation}

for a state with total spin quantum number $S$. Assuming a rough proportionality
between \( \nu \) and \( N \), this number grows like \( N^N \) as \( N \) becomes large.

To emphasize the severity of this configurational proliferation, we have used Eq. (1.1) to calculate the number of independent parameters (configuration coefficients) that determine a singlet (\( S = 0 \)) FCI wavefunction, for a few values of \( N \) and \( \nu \), and listed them in Table 1.1. To connect with a hypothetical application, we list in Table 1.2 the number of parameters that determine the FCI wavefunctions for the ground states of \( \text{H}_2\text{O} \) and \( \text{C}_2\text{H}_4 \), in several standard Gaussian basis sets. These calculations really are hypothetical; the largest two water calculations in Table 1.2, and all but the minimal-basis ethylene calculation, are larger than any FCI calculation that has yet been performed. The current record in this respect\(^7\) is single point on the \( \text{N}_2 \) potential surface, calculated using a 4\( s \)3\( p \)1\( d \) basis, with 10 correlated electrons distributed among 36 orbitals.\(^*\) The corresponding FCI wavefunction is a linear combination of approximately \( 9.7 \times 10^9 \) Slater determinants.\(^1\)

Approximate electronic structure methods reduce the exponential scaling of FCI to a manageable, algebraic dependence, but it is of interest to consider whether electronic structure theory need start from the electronic wavefunction—and its \( N \)-fold coordinate dependence—at all. At a practical level, the success of density functional theory (DFT)\(^{10,11}\) clearly indicates that it does not, for DFT replaces the \( N \)-coordinate wavefunction \( \Psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) \) as the essential player in electronic structure theory with a one-coordinate function \( \rho_1(\mathbf{x}) \), the one-electron density. (\( \mathbf{x} \) represents space- and spin-)

\(^*\)Using the density matrix renormalization group, Chan and Head-Gordon\(^8\) have recently reported a numerically exact solution of the electronic Schrödinger equation for water, using 10 correlated electrons and 41 orbitals. This calculation required 150 hours on 64 parallel processors, so clearly there remains an incentive to explore alternative ansätze.

\(^1\)Extrapolations to the FCI limit, based on limited CI plus perturbation theory, have been carried out using as many as \( 4 \times 10^{15} \) determinants.\(^9\)
Table 1.1: Approximate number of independent parameters to determine the wave-
function or 2-RDM, for an $N$-electron singlet state in a basis of $\nu$ orbitals.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$\nu$</th>
<th>No. parameters</th>
<th></th>
<th>No. parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\Psi$ (FCI)</td>
<td>2-RDM</td>
<td>$\Psi$ (FCI)</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>$2.0 \times 10^1$</td>
<td>$1.6 \times 10^2$</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>$3.4 \times 10^2$</td>
<td>$2.5 \times 10^3$</td>
<td>26</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>$5.4 \times 10^3$</td>
<td>$4.0 \times 10^4$</td>
<td>36</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>$1.7 \times 10^3$</td>
<td>$2.5 \times 10^3$</td>
<td>60</td>
</tr>
<tr>
<td>8</td>
<td>16</td>
<td>$8.7 \times 10^5$</td>
<td>$4.0 \times 10^4$</td>
<td>138</td>
</tr>
<tr>
<td>8</td>
<td>32</td>
<td>$2.9 \times 10^8$</td>
<td>$6.5 \times 10^5$</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>16</td>
<td>$3.5 \times 10^7$</td>
<td>$4.0 \times 10^3$</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>32</td>
<td>$1.6 \times 10^{13}$</td>
<td>$6.5 \times 10^5$</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>64</td>
<td>$2.5 \times 10^{18}$</td>
<td>$1.0 \times 10^7$</td>
<td></td>
</tr>
</tbody>
</table>

Table 1.2: Approximate number of independent parameters to determine the wave-
function or 2-RDM, for the water and ethylene molecules.

<table>
<thead>
<tr>
<th>Basis set</th>
<th>$\nu$</th>
<th>No. parameters</th>
<th>$\nu$</th>
<th>No. parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\Psi$ (FCI)</td>
<td>2-RDM</td>
<td>$\Psi$ (FCI)</td>
</tr>
<tr>
<td>STO-3G</td>
<td>7</td>
<td>$2.0 \times 10^2$</td>
<td>$1.5 \times 10^3$</td>
<td>14</td>
</tr>
<tr>
<td>6-31G</td>
<td>13</td>
<td>$4.3 \times 10^5$</td>
<td>$1.7 \times 10^4$</td>
<td>26</td>
</tr>
<tr>
<td>6-31G*</td>
<td>18</td>
<td>$1.2 \times 10^7$</td>
<td>$6.4 \times 10^4$</td>
<td>36</td>
</tr>
<tr>
<td>6-31++G**</td>
<td>30</td>
<td>$4.0 \times 10^9$</td>
<td>$5.0 \times 10^5$</td>
<td>60</td>
</tr>
<tr>
<td>6-311++G(3df; 2pd)</td>
<td>69</td>
<td>$2.3 \times 10^{13}$</td>
<td>$1.4 \times 10^7$</td>
<td>138</td>
</tr>
</tbody>
</table>
coordinates.) This immediately eliminates the exponential scaling dilemma;* whereas wavefunction quantum mechanics requires the solution of an \( N \)-electron (Schrödinger) equation, DFT requires only the solution of a one-electron (Kohn-Sham\(^{16} \)) equation. There is a high price to be paid for this reduction, however, as one is forced to give up the simple energy functional \( E[\Psi] = \langle \Psi | \hat{H} | \Psi \rangle \) in favor of an unknown functional \( E[\rho_1] \.

One may judge that the latter functional must be substantially more complicated in its form, given the ubiquitousness of empirical parameters in DFT.

The work in this thesis represents a quest for the middle ground. We inquire whether there is some intermediate reduction, which does not give up quite so much detailed information as does DFT, yet to some extent reduces the complexity of the \( N \)-electron quantum-mechanical problem. The reduced quantities that we seek as basic variables for our theory are the reduced density matrices (RDMs),\(^{17-21} \) specifically the one-electron RDM (1-RDM)

\[
D_1(x_1; x'_1) \overset{\text{def}}{=} N \int dx_2 \cdots dx_N \Psi(x_1, \ldots, x_N) \Psi^*(x'_1, x_2, \ldots, x_N) \quad (1.2a)
\]

and the two-electron RDM (2-RDM)

\[
D_2(x_1, x_2; x'_1, x'_2) \overset{\text{def}}{=} \binom{N}{2} \int dx_3 \cdots dx_N \Psi(x_1, \ldots, x_N) \Psi^*(x'_1, x'_2, x_3, \ldots, x_N),
\]

(1.2b)

which are reduced probability distributions stemming from the \( N \)-electron probability distribution (\( N \)-electron density matrix) \( \Psi(x_1, \ldots, x_N) \Psi^*(x'_1, \ldots, x'_N) \).

---

*This is true, provided that the electronic energy functional \( E[\rho_1] \) itself does not build this scaling back into the methodology. Various attempts\(^{12-15} \) to incorporate aspects of many-body theory within DFT, in the interest of rendering the latter more systematic, appear to do just that. As will be seen in this work, there is a similar tendency in density matrix theory to unwittingly build back in the complexity of the many-body problem in the course of direct determination of the two-electron reduced density matrix.
The fundamental theorem\(^{18}\) that makes these quantities interesting is the observation that, owing to the antisymmetry of \(\Psi\), \(D_p\) can be used to compute the exact expectation value of any \(p\)-electron operator that treats all \(p\) electrons equivalently. Since electrons are indistinguishable according to the postulates of quantum mechanics, any valid observable \textit{must} correspond to such an operator, and furthermore it is an empirical fact that all observables in molecular quantum mechanics are expectation values of one- or two-electron operators only. In particular, the kinetic energy is a known functional of either the wavefunction or the 1-RDM:

\[
T[\Psi] \equiv \left\langle \Psi \left| \sum_{i=1}^{N} \left(-\frac{1}{2} \nabla_i^2 \right) \right| \Psi \right\rangle = \int dx \, dx' \left(-\frac{1}{2} \nabla_1^2 \right) D_1(x_1; x'_1) = T[D_1]. \tag{1.3}
\]

Although \(T[D_1]\) affords the same kinetic energy as \(T[\Psi]\), the former is a one-electron expression whereas \(T[\Psi]\) is defined on an \(N\)-electron Hilbert space. Similarly, the electron repulsion energy can be expressed either as an \(N\)-electron functional \(V_{ee}[\Psi]\) or a two-electron functional \(V_{ee}[D_2]\), both of which are known explicitly:

\[
V_{ee}[\Psi] \equiv \left\langle \Psi \left| \sum_{i<j}^{N} r_{ij}^{-1} \right| \Psi \right\rangle = \int dx \, dx' \, r_{12}^{-1} \, D_2(x_1, x_2; x'_1, x'_2) = V_{ee}[D_2]. \tag{1.4}
\]

Since lower-order probability distributions can be obtained from higher-order ones, for example

\[
D_1(x_1; x'_1) = \frac{2}{N-1} \int dx_2 \, D_2(x_1, x_2; x'_1, x_2), \tag{1.5}
\]
it follows that the total electronic energy can be expressed as a known functional of either $\Psi$ or $D_2$. In fact, $D_2$ contains all of the information necessary to compute any observable, insofar as observables involve only one- and two-electron operators.

The foregoing discussion is predicated upon knowledge of the electronic wavefunction, from which follow the RDMs. A long-standing goal of quantum chemistry—and the objective towards which this thesis work is directed—has been the direct computation of RDMs (especially $D_2$) without the necessity of first calculating $\Psi$. We develop this problem further in the next section, but for now let us take stock of the compactness of the 2-RDM relative to the electronic wavefunction. In Tables 1.1 and 1.2 above, we have also listed the number of independent parameters (matrix elements) necessary to fully specify the 2-RDM. Whereas several of the hypothetical FCI calculations in these tables far exceed the present feasible limit of about $10^{10}$ determinants, in each case $D_2$ has fewer than $10^9$ matrix elements.

Novel computational paradigms aside, the compact codification of expectation values by the 1- and 2-RDMs is tremendously appealing from a conceptual perspective, and even within traditional wavefunction quantum mechanics, the RDMs are useful tools for understanding electron correlation. Relative to the electronic wavefunction, the 2-RDM is more closely connected to measurable, readily-interpretable quantities.

---

$D_2$ suffices to calculated observables for any single electronic state, but is insufficient to calculate transition probabilities between electronic states. The latter calculation requires a transition RDM,\(^\text{22,23}\) which is an independent quantity.

\(^1\)The estimates given in Tables 1.1 and 1.2 [computed using Eq. (2.65)] take into account all Hermitian and permutational symmetries, as well as certain spin symmetries, but do not account for other $N$-representability constraints (Section 1.2) that further reduce the number of independent parameters. Strictly speaking, each tabulated value is therefore an upper bound to the true number of independent parameters that determines the 2-RDM. However, insofar as complete $N$-representability constraints are unknown, and known constraints are not easily cast in a form that reduces the number of free parameters, as a practical matter only the aforementioned symmetries are available.
such as the two-electron density \( \rho_2(x_1, x_2) = D_2(x_1, x_2; x_1', x_2') \). As discussed in Subsection 5.2.2, \( \rho_2 \) can be used in conjunction with the 1-RDM to put a solid definition behind the intuitive concepts of a *Fermi hole*\(^{25,26} \)—the depletion of conditional density, relative to the independent-particle result, that arises from the Pauli exclusion principle—and the *Coulomb hole*,\(^{26,27} \) which describes the depletion in conditional density owing to electron correlations beyond the mean-field model. Also available from \( \rho_2 \) are the electron-pair distribution function (intracule density) \( I(u) \) and electron-pair center-of-mass distribution function (extracule density) \( X(R) \), each of which provides an additional tool for visualizing the effects of electron correlation.\(^{28-30} \)

Yet another appealing feature of \( D_2 \) is that it affords an exact, pairwise-additive decomposition of the energy (or any other expectation value) into *natural spin-geminal* contributions. The natural spin geminals \( g_K(x_1, x_2) \) are the eigenfunctions of \( D_2 \),

\[
D_2(x_1, x_2; x_1', x_2') = \sum_K \lambda_K g_K(x_1, x_2) g_K^*(x_1', x_2'),
\]

with occupancies \( \lambda_K \) satisfying\(^31 \)

\[
0 \leq \lambda_K \leq \lfloor N/2 \rfloor.
\]

Defining

\[
\mathcal{E}_K = \int dx_1 dx_2 \frac{|g_K(x_1, x_2)|^2}{r_{12}}
\]

as the electron-repulsion energy associated with \( g_K \), one obtains, for example, a pairwise decomposition of \( V_{ee} \):

\[
V_{ee} = \sum_K \lambda_K \mathcal{E}_K.
\]

The interpretive usefulness of this decomposition has hardly been explored.
1.2 The representability problem

The electronic wavefunction, the 2-RDM, the 1-RDM, and the one-electron density $\rho_1(x) \overset{\text{def}}{=} D_1(x; x)$ establish a hierarchy of electron distribution functions, as depicted schematically in Fig. 1.1. The distributions become more tractable as one moves down this hierarchy, by integrating out certain coordinate dependence [as in Eqs. (1.2) and (1.5)], and finally by dropping the “off-diagonal” (in a continuous representation) part of $D_1(x_1; x'_1)$, to obtain the one-electron density. To gain this tractability, however, one is gradually forced to cede complete knowledge of more and more of the energy functional. At the level of the 1-RDM, for example, the kinetic energy ($T$) and nuclear-electron potential energy ($V_{ne}$) functionals are known, but the electron-repulsion energy functional ($V_{ee}$) is unknown, though $V_{ee}[D_1]$ can be shown to exist, in the abstract sense of a one-to-one correspondence between non-degenerate ground-state energies and 1-RDMs. If we follow this trail of reductionism still further, eliminating the off-diagonal part of $D_1(x_1; x'_1)$, we must sacrifice our knowledge of the exact kinetic energy functional. In DFT, the functionals $T[\rho_1]$ and $V_{ee}[\rho_1]$ are both unknown, though the abstract proof of their existence has been widely publicized.

From the look of Fig. 1.1, it may seem that we have lost nothing by reducing the $N$-electron density matrix $\Psi\Psi^*$ to a 2-RDM, though intuition ought to suggest otherwise. In fact we have given up quite a lot, namely, the simple Hilbert-space boundary conditions that inform us whether $\Psi$ is an acceptable $N$-electron state. The corresponding conditions for acceptable two-electron reduced distributions—the constraints that $D_2$ must satisfy, in order that it derive from some acceptable $N$-electron state—are termed $N$-representability conditions. Although many necessary
Figure 1.1: The hierarchy of quantum-mechanical distribution functions. $D_{1\{i;i'\}}$ means $D_{1}(x_{i};x_{i}')$, etc.
\[ N\text{-representability conditions on the } 2\text{-RDM are known,}^{20,21,34-57} \text{ conditions that are both necessary and sufficient are not known.}^* \text{ This is the fundamental impediment to direct computation of RDMs.} \]

The \( N\)-representability problem is worth restating, and we do this by emphasizing the difference between wavefunction and \( 2\)-RDM variational calculations. Given a trial wavefunction \( \Psi \), the energy associated to this function is computed from the functional

\[
E[\Psi] = \frac{\int dx_1 \cdots dx_N \Psi^*(x_1, \ldots, x_N) \hat{H} \Psi(x_1, \ldots, x_N)}{\int dx_1 \cdots dx_N \Psi^*(x_1, \ldots, x_N) \Psi(x_1, \ldots, x_N)}. \tag{1.10}
\]

Likewise, given an unrelated trial \( 2\)-RDM, the associated energy is obtained from the functional

\[
E[D_2] = \frac{\int dx'_1 dx'_2 \delta(x_1 - x'_1) \delta(x_2 - x'_2) \int dx_1 dx_2 \hat{H}_{12} D_2(x_1, x_2; x'_1, x'_2)}{\int dx_1 dx_2 D_2(x_1, x_2; x_1, x_2)}, \tag{1.11}
\]

in which \( \hat{H}_{12} \) is a certain two-particle Hamiltonian, called a reduced Hamiltonian. The reduced Hamiltonian is introduced as a matter of convenience, so that the known functionals \( T[D_1] \), \( V_ne[D_1] \), and \( V_{ee}[D_2] \) can be expressed as a single, known functional of the \( 2\)-RDM. See Eq. (2.4) for the explicit form of \( \hat{H}_{12} \).

A variational calculation of either the wavefunction or the \( 2\)-RDM is accomplished by minimizing \( E \) with respect to variations in \( \Psi \) or \( D_2 \), but in neither case may these variations be carried out arbitrarily. For \( \Psi \) there are, on the one hand, certain minimal requirements such as continuity, differentiability, and square-integrability, which give \( \Psi \Psi^* \) the character of a probability distribution. Likewise, \( D_2 \) must be positive semidefinite, in the matrix or kernel sense, which insures that its diagonal

---

*In contrast to the situation for the \( 2\)-RDM, any \( 1\)-RDM whose eigenvalues all lie within the interval \([0, 1]\) is necessarily \( N\)-representable.\(^{34}\)*
elements (spin-geminal populations) are positive in any basis, and furthermore guarantees that $\rho_2(x_1, x_2) \geq 0$. On the other hand, not every function $\Psi$ that satisfies the aforementioned requirements is an acceptable $N$-electron state, for Pauli’s postulate informs us that $\Psi$ must also be antisymmetric with respect to coordinate permutations,

$$\Psi(x_1, x_2, x_3, \ldots) = -\Psi(x_2, x_1, x_3, \ldots) = \Psi(x_2, x_3, x_1, \ldots) = \cdots .$$  

(1.12)

From this it follows that the reduced distributions are antisymmetric as well. For example,

$$D_2(x_1, x_2; x'_1, x'_2) = -D_2(x_2, x_1; x'_1, x'_2) = -D_2(x_1, x_2; x'_2, x'_1) = D_2(x_2, x_1; x'_2, x'_1).$$

(1.13)

A positive, antisymmetric, self-adjoint kernel $D_2(x_1, x_2; x'_1, x'_2)$ is an acceptable two-electron density matrix, but these facile conditions are not sufficient to guarantee that $D_2$ has an acceptable $N$-electron pre-image, that is, they do not insure that $D_2$ is a reduced density matrix for any $N$-electron system (for arbitrary $N > 2$). This is the infamous “$N$-representability problem.” It implies that variations of the functional $E[D_2]$ must be confined to some (less than fully-characterized) proper subset of two-electron density matrices satisfying the minimal constraints listed above; otherwise, the variational energy $E[D_2^{\text{sat}}]$ is not an upper bound to the exact energy, and spurious over-binding results in practice. This is analogous to the fact that the ground state for a system of fermions is not that eigenfunction of $\hat{H}$ having the most negative eigenvalue. Instead, the ground state is the lowest-energy antisymmetric eigenfunction; the energy of this true fermionic ground state can always be (anomalously) lowered by admitting an admixture with some lower-energy bosonic wavefunction.
Excepting for Coleman’s indefatigable enthusiasm, there is almost no optimism that a complete and tractable solution to the $N$-representability problem will be found. Nevertheless, quite a lengthy list of necessary $N$-representability conditions can be compiled, whose stringency may be evaluated via partially-constrained 2-RDM variational calculations. (Given their exceptionally high cost, this is how the author envisions the role of such calculations.) As a practical matter, it is not necessary that a given method yield a variational upper bound—most electronic structure methods do not—but lacking such a bound, one must verify empirically that the energy does not diverge with respect to expansion of the variational space. This has been verified in the case of variational 2-RDM calculations that are constrained by the three matrix positivity conditions known as $D$, $Q$, and $G$ (see Subsection 2.3.5 of this work). For light atoms and small molecules, these partially-constrained variational calculations afford 100–125% of the FCI correlation energy, whereas if just one of these three constraints is lifted, variational minimization can yield as much as 800% of the correlation energy. It seems safe to conclude that the $D$-, $Q$-, and $G$-conditions are necessary and almost sufficient to insure $N$-representability of the 2-RDM. As such, the main challenge in density matrix theory—as in wavefunction theory—is to find computationally tractable alternatives to variational calculations, though in density matrix theory the $N$-representability problem introduces an additional layer of complexity; at least some of the known $N$-representability conditions must be enforced when attempting to compute RDMs directly.
1.3 Description and outline of the present work

We emphasized variational 2-RDM calculations in the previous section merely as a means to explain the concept of $N$-representability; no such calculations are undertaken here. As indicated previously, our research seeks a middle ground between wavefunction theory, which is characterized by facile and explicit boundary conditions; a simple, exact electronic energy functional; and disastrous coordinate scaling, and DFT, which also has simple boundary conditions (every reasonable density is $N$-representable\textsuperscript{67}), but trades an explicitly known energy functional for a more tractable coordinate dependence. Referring back to Fig. 1.1, we have indicated two levels of intermediacy, distinguished by whether the 1-RDM or the 2-RDM is the basic variable, and both methodologies are explored in this work. Following a discussion of the formal properties of RDMs in Chapter 2, the next two chapters are dedicated to a discussion of a non-variational methodology for direct determination of $D_2$, based on the Contracted Schrödinger Equation (CSE). Then in Chapter 5 we discuss determination of the electronic energy as a functional of the 1-RDM, in what has come to be called density matrix functional theory (DMFT) or sometimes natural orbital functional theory. Both the theory and practice of DMFT differ markedly from the 2-RDM-based methodology discussed in Chapters 3 and 4, but the unifying theme—encapsulated by Fig. 1.1—is the same: the direct determination of RDMs, and from them, electronic energies, without the necessity of first calculating the electronic wavefunction.

As the title of this work suggests, another unifying theme that will become evident is the important role of what we term density matrix reconstruction, by which we mean the development of approximate formulas for higher-order RDMs $D_p$ in
terms of the lower-order ones, $D_1, \ldots, D_{p-1}$. We designate any approximate relationship $D_p = D_p[D_1, \ldots, D_{p-1}]$ as a reconstruction functional for the $p$-RDM. (Such a functional might alternatively be called a decoupling approximation for $D_p$.) It is a reconstruction because each lower-order RDM is obtained from the $p$-RDM via a partial trace relationship analogous to Eq. (1.5), but in order to determine the RDMs from principles we need to be able to move up the hierarchy in Fig. 1.1. Reconstruction functionals for the 3- and 4-RDMs are the key elements to determination of $D_2$ via the CSE, while our approach to DMFT relies upon a reconstruction of the 2-RDM as a functional of the 1-RDM.

Due to the length of this thesis, for the reader’s convenience we conclude this introductory chapter with a brief travelogue highlighting the journey through density matrix theory that will follow in Chapters 2-5. Special emphasis is given here to those results that originate with the author’s work.

Chapter 2 is dedicated mostly to a careful introduction of RDMs. Particular attention is given, in Section 2.3, to the various representations (Fock space, Hilbert space, position space, etc.) in which RDMs can be formulated. Section 2.4 introduces the electron-pair distribution function $I(u)$ and electron-pair center-of-mass distribution $X(R)$. As mentioned above, these are useful tools in the interpretation of the 2-RDM or two-electron density, as each reduces this density to a one-dimensional function that still encapsulates some aspect of electron correlation. It is precisely for this interpretive value that we introduce $I(u)$ and $X(R)$, which are used in Chapter 5 to test some proposed reconstruction functionals for the 2-RDM. The quantities $I(u)$ and $X(R)$ are under-utilized in quantum chemistry, in the author’s opinion, and are not yet available from standard electronic structure codes. Thus in Chapter 2 we develop the
necessary formulas to evaluate these densities in a basis of Slater-type orbitals, such as
are typically employed in accurate atomic calculations. (The analogous formulas for a
Gaussian basis set have been worked out by others.\textsuperscript{68–72}) Quite unexpectedly, out of
this analysis comes an explanation for an approximate proportionality, $X(R) \propto I(2R)$,
reported previously\textsuperscript{73–75} on the basis of numerous atomic HF calculations, but which
had gone unexplained prior to the author’s work.\textsuperscript{76}

In Section 2.5 we introduce cumulant partitionings for the RDMs. Up to an over-
all antisymmetrization, the cumulant decomposition of the $p$-RDM is essentially the
same as that of the $p$-particle distribution function in classical statistical mechanics\textsuperscript{77}
and is also closely related to the cluster decomposition of the wavefunction\textsuperscript{78} that is
familiar from many-body wavefunction theory. The cumulants are interesting because
they are \textit{additively separable} (that is, extensive) quantities, whereas the RDMs are not.
Hence, a properly size-consistent methodology for electronic structure theory ought to
be formulated in terms of the RDM cumulants rather than the RDMs themselves.\textsuperscript{79,80}

More phenomenologically, one might expect this fermionic generalization of the classi-
cal cumulant formalism to be a useful tool for understanding electron correlation. As
discussed in Subsection 4.3.3, this \textit{may} be the case for dynamical correlation, but the
present version of the cumulant formalism appears to be ill-suited to the discussion of
non-dynamical correlation.

In our discussion of the RDM cumulants we will encounter for the first time
the antisymmetrized (Grassmann) tensor product, which plays a major role in the
formalism of fermion quantum mechanics. In Section 2.6, which is based mainly on
the results in Ref. 81 and some unpublished generalizations thereof, we derive a number
of partial trace relations for Grassmann products of RDMs and other many-particle
tensors. These are perhaps not too exciting in their own right, but some of them have proved to be useful tools.

In Chapter 3 we develop the CSE. This equation, which was introduced nearly simultaneously by Cohen and Frishberg\textsuperscript{82} and by Nakatsuji,\textsuperscript{83} provides an equivalent,\textsuperscript{83,84} density-matrix formulation of the electronic Schrödinger equation in terms of the 2-, 3-, and 4-RDMs.* Given reconstruction functionals for the 3- and 4-RDMs in terms of $D_2$, the CSE provides a means for direct determination of the 2-RDM. Like the configuration interaction (CI) formulation of the Schrödinger equation, however, the CSE contains a host of “unconnected” terms that do not preserve extensivity. While these terms cancel out of the exact equation, they cannot be expected to do so in an approximate treatment. Thus the CSE as originally formulated is not size-consistent,\textsuperscript{87} and it is of interest to obtain connected working equations (in terms of the RDM cumulants) that eliminate all non-extensive terms \textit{a priori}. This exercise, which represents the major original contribution of Chapter 3, was first published in Ref. 88. In Subsection 3.3.3 we describe the fundamental differences between the original CSE and its connected analogue.

Chapter 4 presents our work on the reconstruction problem for the 3- and 4-RDMs. Reconstruction functionals derived previously by Valdemoro and coworkers\textsuperscript{89,90} are reviewed and re-derived in Subsection 4.2.1, which is notable in that our derivation brings to light certain inconsistencies—not readily apparent in the original papers cited above—in the sequence of approximations leading to the final functional

---

*Equivalence of the CSE and the Schrödinger equation is predicated on the assumption that each is solved subject to appropriate boundary conditions. For the CSE, this means the unknown $N$-representability constraints. As pointed out by Harriman\textsuperscript{85} (see also Yasuda\textsuperscript{86}), absent a complete set of $N$-representability constraints, there exists a continuous manifold of spurious, non-$N$-representable solutions to this equation.
expressions. The Valdemoro scheme is tested by reconstructing the 4-RDM, starting from the 2-RDM for an accurate, large-basis CI calculation. Comparing the reconstructed RDMs to the CI ones, we discover that the Valdemoro functionals are not nearly so accurate as suggested by previous minimal-basis results. In light of this, alternative reconstruction procedures based on more readily understood approximations are recommended.

One such reconstruction scheme has been proposed by Mazziotti and relies upon a truncation of the RDM cumulant hierarchy, ultimately neglecting the four-electron cumulant. Given the well-known dominance of pair correlations in electronic structure, this might at first glance be deemed reasonable, but our analysis in Subsection 4.3.3 casts doubt upon the validity of this approximation. Interestingly, the cumulant truncation approximation itself (separate from its use in the CSE) has not been tested directly, so all arguments with regard to the negligibility of the four-electron cumulant remain speculative. In Subsection 4.3.3 we do, however, provide the first analysis of the three-electron cumulant in an extended basis set. Relative to the unconnected part of the 3-RDM, this cumulant is found to be small but definitely not negligible, contributing 10–20 millihartrees to the electronic energy of Be and LiH. This result (as yet unpublished) provides the first quantitative answer to the question “how small are the RDM cumulants?”, which has been raised in some recent papers and also circulated at the past two Sanibel Symposia.

This numerical evidence, combined with the observation from Chapter 3 that the CSE ought to be recast in terms of cumulants only, provides an incentive to develop 3- and 4-RDM reconstruction functionals that account, at least approximately, for both the three- and four-electron cumulants. The latter has been completely neglected in all
CSE calculations to date. In Sections 4.4 and 4.5 we develop a scheme for estimating the three- and four-particle cumulants using perturbation theory. Section 4.4 (originally published in Ref. 98) is concerned mainly with the formal underpinnings of this perturbation theory, which for generality we cast in a form that can be applied either to the RDMs or their time-dependent generalizations, the many-particle Green’s functions. In either case, the three- and four-particle perturbation series are substantially more complicated than their one- or two-particle analogues, and a diagram technique is indispensable.

In Section 4.5 we employ this method to provide a systematic approach to density matrix reconstruction that generalizes the 3-RDM reconstruction functional of Nakatsuji and Yasuda, both to higher orders and—more importantly—to the case of the 4-RDM. Thus we obtain reconstruction functionals that account, at least approximately, for both the three- and the four-particle cumulant. On the one hand, the generality of this approach is quite appealing. However, the author has serious reservations about the utility of this approach, and especially its advantage (or lack thereof) compared to more mainstream many-body methods such as coupled-cluster theory. These concerns, along with a discussion of the future direction of this work, are aired in Section 4.6.

Chapter 5 represents a substantial change of pace, and of paradigm, for in this chapter we discuss energy functionals of the 1-RDM. There are various ways to approach the derivation of an approximate functional $E[D_1]$, but the one that fits with the overarching theme of this work is to imagine a reconstruction function $D_2[D_1]$, or in other words a parametrization of the 2-RDM in terms of the 1-RDM, which automatically affords an energy functional $E[D_1]$. In Section 5.2 we introduce several
reconstruction functionals for DMFT that have been proposed recently, of the “corrected Hartree” and “corrected Hartree-Fock” variety. It is shown immediately (Subsection 5.2.3) that none of these proposals corresponds to an \( N \)-representable 2-RDM. A variety of \( N \)-representability conditions are violated by these functionals, but perhaps the most outstanding transgressions are the failures to preserve antisymmetry and positivity of the reconstructed 2-RDM. We will demonstrate a close relationship between these two egregious \( N \)-representability violations, which indicates one avenue for further functional development.

Elsewhere in the present chapter we mentioned that whenever a particular ansatz violates \( N \)-representability, there is immediately a question as to whether the electronic energy is plausible—or even finite—in the complete-basis limit. In the case of the aforementioned 1-RDM functionals, results of some small active-space calculations initially led to speculation that the electronic energy might diverge as the variational space is enlarged. We have recently addressed this concern by means of a careful analysis of the basis-set dependence of these functionals, using all-electron (rather than active-space) calculations. Not only is the energy found to converge, but we identify one functional (of the corrected Hartree type, with a self-interaction correction) that produces astoundingly accurate diatomic potential curves, both for weakly-correlated molecules like LiH (Subsection 5.5.1) as well as strongly-correlated ones like FH and \( \mathrm{C}_2 \) (Section 5.6). Not only are the resulting potential energy curves qualitatively correct at all values of the internuclear distance, they are quantitatively comparable to multireference CI singles-doubles results. This indicates that the 1-RDM functional accounts for both dynamical and non-dynamical correlation, which is something of a holy grail in this field.
Unfortunately, analysis of the $N$-representability violations manifested by this first generation of 1-RDM functionals reveals that their success is due entirely to a cancellation of errors (Subsection 5.5.2), albeit an extraordinarily accurate and robust one. Therefore there is much room for development of better functionals. We feel that analysis of $N$-representability constraints for the reconstructed 2-RDMs offers a means to gather detailed information about the physical nature of a particular 1-RDM functional, and that this information can be used to guide future development.

Along these lines, an important theme from Section 5.5 is the close connection between the antisymmetry and positivity requirements for the 2-RDM. In light of the large, anomalous energy lowering that is brought about by spurious negative eigenvalues of $D_2$, we undertake in Section 5.7 to derive new density matrix functionals that are manifestly positive and antisymmetric. Preliminary tests indicate that these new functionals yield qualitatively reasonable electron-pair distributions, while $N$-representability violations are smaller than those afforded by the corrected Hartree and corrected Hartree-Fock functionals. The present state of DMFT, and the plan for our own future work in this field, is summarized in Section 5.8.

1.4 References


Chapter 2

Properties of reduced density matrices

2.1 Introduction

In the previous chapter, we observed that the RDMs—to the extent that they can be calculated—supply all of the information necessary to calculate observables for any electronic state. In Sections 2.2 and 2.3 we make this idea precise by formally introducing the RDMs and demonstrating their connection to computation of physical observables. One may envisage a density matrix in a number of ways: as an operator on the abstract Hilbert space or the abstract Fock space, as an integral kernel in either a position or a momentum representation, or as a tensor whose elements represent the populations and coherences for \( p \)-electron states within an \( N \)-electron system (\( N \geq p \)). Often in this work we pass between these notations without comment, so in order to make the connections clear, Sections 2.2 and 2.3 are presented in pedagogical fashion.

Following this formal introduction of the RDMs, the remainder of this chapter is devoted to developing some of their formal properties. In Section 2.4 we introduce the intracule density (pair distribution function) and extracule density (center-of-mass
distribution function) for an electron pair in a many-electron system.\textsuperscript{1–3} Both of these reduced densities stem from the 2-RDM, and each preserves some aspect of electron correlation, even though each depends upon only a single electronic coordinate. The intracule and extracule densities are useful for interpretative purposes, and we will use these quantities in Chapter 5 to analyze two-electron densities, which are otherwise difficult to visualize. In anticipation of this analysis, Subsections 2.4.2 and 2.4.3 are devoted to working out analytic formulas for evaluating the intracule and extracule densities in a basis of Slater-type orbitals. Interestingly, out of these manipulations comes a mathematical justification for an approximate relationship between the intracule and extracule densities that was noticed by Koga\textsuperscript{4–6} in the intracule and extracule densities from atomic HF calculations\textsuperscript{7–10} but was unexplained prior to the author’s work.\textsuperscript{11}

An important property for approximate electronic structure methods is *extensivity,*\textsuperscript{*} which insures that the method reduces to the correct limit for non-interacting sub-systems. Just as in wavefunction-based approaches, an arbitrary, approximate method for calculating RDMs need not satisfy this property. To remedy this, in Section 2.5 we introduce a cumulant\textsuperscript{12} decomposition for the RDMs, and discuss the properties (including manifest extensivity) of the RDM cumulants. The density-matrix cumulant expansion was fleshed out in its modern form by Mazziotti,\textsuperscript{13,14} though its essential elements are implicit in much earlier work.\textsuperscript{15–17} The discussion in Section 2.5 of this work bridges the gap between the older and the newer formalisms.

Section 2.6 introduces the antisymmetric (Grassmann) tensor product, which

\textsuperscript{*}Extensivity is related to the notion of *size-consistency,* though the former concept is somewhat more general; see Section 2.5.
figures prominently in the algebra of fermion operators and matrices. Because of the partial trace relation that connects the antisymmetric RDMs of different orders, it is frequently necessary to consider contractions (partial traces) of antisymmetric kernels or tensors. In Subsections 2.6.1 and 2.6.2, some general formulas for carrying out such contractions are derived, which greatly facilitate later derivations. (The results in Subsections 2.6.1 were mainly published in Ref. 18.) The proofs are somewhat tedious, but having mastered the techniques involved, the reader will appreciate the simplified counting procedure that is introduced in Subsection 2.6.2, by means of which the previous results (and many new ones) can be derived with relatively little effort.

### 2.2 Observables for indistinguishable particles

Consider an observable $\hat{W}$ for a system of $N$ indistinguishable electrons. It is an empirical fact that all such observables in molecular quantum mechanics correspond to some combination of one- and two-electron interactions. Let $\hat{h}$ be a one-electron interaction and $\hat{g}$ a two-electron interaction. Within a first-quantized, Hilbert-space formulation of quantum mechanics, $\hat{g}(x_i, x_j)$ is the pairwise interaction between the $i$th particle and the $j$th particle, where $x_i = (r_i, \xi_i)$ denotes a composite space-spin coordinate for the $i$th particle. We will frequently abbreviate $x_i$ as simply “$i$”. Because the individual electrons are indistinguishable, $\hat{W}$ must treat each of the $N$ electrons, and each of the $N(N-1)/2$ pairs of electrons, equivalently. This requires that $\hat{g}(i, j) = \hat{g}(j, i)$ and also that $\hat{W}$ must consist of the symmetric sum

$$\hat{W}(i, \ldots, N) = \sum_{i=1}^{N} \hat{h}(i) + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{g}(i, j). \quad (2.1)$$
This is the most general form of an observable for electrons. In particular, if one takes
\( \hat{h}(i) = -\frac{1}{2}\nabla_i^2 + \sum_\alpha Z_\alpha r_i^{-1} \) and \( \hat{g}(i,j) = r_{ij}^{-1} \), then \( \hat{W} \) equals the usual spin-independent Born-Oppenheimer electronic Hamiltonian, in atomic units.* Spin-orbit and spin-spin interaction terms are also consistent with Eq. (2.1).

As indicated above, \( \hat{W} = \hat{W}(1,\ldots,N) \) acts on an \( N \)-particle Hilbert space: \( \langle \hat{W} \rangle = \langle \Psi|\hat{W}\Psi \rangle \). Equivalently, this expectation value can be expressed as \( \langle \hat{W} \rangle = \text{tr}(\hat{W} \hat{D}_N) \), the average of \( \hat{W} \) over the pure-state density matrix

\[
D_N(1,\ldots,N;1',\ldots,N') = \Psi(1,\ldots,N) \Psi^*(1',\ldots,N') \ .
\] (2.2)

The essential result that motivates reduced density matrix theory is the observation that, owing to the permutational antisymmetry of \( \Psi \) and the indistinguishability of electrons, \( \langle \hat{W} \rangle \) can always be calculated using a two-electron distribution

\[
D_{2}(1,2;1',2') \overset{\text{def}}{=} \left(\frac{N}{2}\right) \int d\mathbf{x}_3 \cdots d\mathbf{x}_N \Psi(1,\ldots,N) \Psi^*(1',2',3,\ldots,N) \ .
\] (2.3)

\( D_{2}(1,2;1',2') \) is the position-space representation of the two-electron reduced density matrix.

Given \( \hat{h} \) and \( \hat{g} \) (and hence \( \hat{W} \)), define a corresponding reduced operator

\[
\hat{W}_{12}(1,2) \overset{\text{def}}{=} \hat{g}(1,2) + \frac{\hat{h}(1) + \hat{h}(2)}{N - 1}
\] (2.4)

that acts on a two-electron Hilbert space. The expectation value of such an operator

---

*Mostly atomic units are utilized in this work, and the usual symbols “\( E_h \)” for the atomic unit of energy (hartree) and \( a_0 \) for the atomic unit of distance (bohr) are also employed. The author enthusiastically advocates the millihartree \( (1.00 \ mE_h = 0.027 \ eV = 0.63 \ \text{kcal/mol} = 220 \ \text{cm}^{-1}) \) as a convenient energy unit for much of everyday physical chemistry and molecular quantum mechanics.
is defined as its average over the two-electron distribution $D_2$,
\[
\langle \hat{W}_{12} \rangle \overset{\text{def}}{=} \text{tr} (\hat{W}_{12} \hat{D}_2) = \int dx'_1 dx'_2 \, \delta(x_1 - x'_1) \, \delta(x_2 - x'_2) \int dx_1 dx_2 \, \hat{W}_{12}(1,2) \, D_{2(1,2;1',2')} (2.5)
\]
\[
= \int dx_1 dx_2 \left\{ \hat{W}_{12}(1,2) \, D_{2(1,2;1',2')} \right\}_{1'=1,2}
\]
By definition, $\hat{W}_{12} = \hat{W}_{12}(1,2)$ acts on the unprimed coordinates of $D_{2(1,2;1',2')}$, and following integration over these coordinates, one sets $x_i = x'_i$. This careful segregation of the primed and the unprimed coordinates is necessary in case $\hat{W}_{12}$ contains derivative operators, and the final line of Eq. (2.5) introduces a common notation for the aforementioned sequence of operations. The fundamental theorem in reduced density matrix theory is the equality of the two- and the $N$-particle expectation values,\(^1^9\)
\[
\langle \hat{W}_{12} \rangle = \langle \hat{W} \rangle . \tag{2.6}
\]

### 2.3 Reduced density matrices for particles and for holes

The definition of the two-electron density matrix in Eq. (2.3) readily generalizes to a $p$-electron reduced density matrix ($p$-RDM or $p$-matrix), for each $p < N$. The RDMs can be introduced in numerous ways: as integral operators on Hilbert space (in either a position or a momentum representation), as operators on the abstract Fock (“occupation-vector”) space, or as tensors in a finite spin-orbital basis. Several of these formulations are elaborated in this section. Later on we will switch between these representations, as convenient.
2.3.1 First-quantized formalism

In the position representation, the \( p \)-RDM for the pure state \( \Psi \) is defined\(^{19-21}\) as the integral kernel

\[
D_p(1,\ldots,p;1',\ldots,p') \overset{\text{def}}{=} \binom{N}{p} \int \! dx_{p+1} \cdots dx_N \, \Psi(1,\ldots,N) \, \Psi^*(1',\ldots,p',p+1,\ldots,N) \\
= \sum_{i_1,\ldots,i_p,j_1,\ldots,j_p} D_{i_1,\ldots,i_p;j_1,\ldots,j_p} \phi_{i_1}(1) \cdots \phi_{i_p}(p) \, \phi^*_{j_1}(1') \cdots \phi^*_{j_p}(p').
\]

(2.7)

This equation also introduces an expansion of the kernel \( D_p \) in terms of a basis \( \{ \phi_k \} \) of orthonormal spin orbitals. As indicated previously, \( D_2 \) is the quantity of paramount interest, since it provides the most compact codification of expectation values, but in order to determine the 2-RDM it will be necessary to deal with \( D_3 \) and \( D_4 \) as well.

Let \( D_p \) denote the tensor consisting of the \( p \)-RDM expansion coefficients introduced in Eq. (2.7). The elements of this tensor we write in either of two ways,

\[
(D_p)_{i_1,\ldots,i_p}^{j_1,\ldots,j_p} \equiv D_{i_1,\ldots,i_p;j_1,\ldots,j_p},
\]

(2.8)

the latter form being somewhat more convenient for \( p \leq 2 \). In writing \( D_{i_1,\ldots,i_p;j_1,\ldots,j_p} \), the order of the RDM (that is, \( p \)) is implied by the number of indices. The tensor \( D_p \) is self-adjoint and also antisymmetric in both its upper and lower indices.

Note that \( \Psi(1,\ldots,N) \, \Psi^*(1',\ldots,N') \) is just the position-space representation of the \( N \)-electron density matrix for the pure state \( \Psi(1,\ldots,N) \). In Dirac notation, this density matrix is written as an operator \( |\Psi\rangle\langle\Psi| \) on the \( N \)-electron Hilbert space. In this formulation, the \( p \)-RDM is represented by an operator

\[
\hat{D}_p = \binom{N}{p} \, \text{tr}_{p+1,\ldots,N}(|\Psi\rangle\langle\Psi|)
\]

(2.9)

that acts on a \( p \)-electron Hilbert space.
In Eq. (2.9) we have introduced a linear *partial trace* operator $\text{tr}_{p+1, \ldots, N}$ that maps a density matrix for $N$ particles onto a *reduced* density matrix for $p < N$ particles. For definiteness, let us explicitly define the so-called $n$-trace

$$\text{tr}_{q-n+1, \ldots, q} = \int dx_{q-n+1} \cdots dx_q \int dx'_{q-n+1} \cdots dx'_q \, \delta(x_{q-n+1} - x'_{q-n+1}) \cdots \delta(x_q - x'_q),$$

(2.10)

which maps a $q$-electron density matrix onto a $(q-n)$-RDM. For tensors, the $n$-trace is simply the sum over the final (rightmost) $n$ indices. To preserve the binomial coefficient normalization for the RDMs, we define a *contraction operator* $\text{p}_{p-n}^p$, which is merely a scaled $n$-trace:

$$\text{p}_{p-n}^p \overset{\text{def}}{=} \left( \frac{N}{p} \right) \text{tr}_{p-n+1, \ldots, p}.$$  

(2.11)

The scale factor is such that

$$\text{p}_{q}^p \hat{D}_p = \hat{D}_q$$

(2.12)

for each $q < p \leq N$. In other words

$$D_{j_1, \ldots, j_{p-n}; k_1, \ldots, k_{p-n}} = \left( \frac{N}{p} \right) \sum_{i_1, \ldots, i_n} D_{j_1, \ldots, j_{p-n}; i_1, \ldots, i_n; k_1, \ldots, k_{p-n}; i_1, \ldots, i_n}.$$  

(2.13)

The normalization implied by Eq. (2.7) is such that

$$\text{tr}(D_p) = \sum_{k_1, \ldots, k_p} D_{k_1, \ldots, k_p; k_1, \ldots, k_p} = \left( \frac{N}{p} \right)$$

(2.14)

equals the number of distinct $p$-tuples of electrons present in the $N$-electron system. This so-called Löwdin\textsuperscript{19–21} normalization convention is the most convenient for computing expectation values, though unit-normalized RDMs (Coleman’s convention\textsuperscript{22}) and also $[N!(N - p)!]$-normalized RDMs (McWeeny’s convention\textsuperscript{23}) are sometimes encountered in the literature.
The operator $\hat{D}_p$ in Eq. (2.9) may be expanded in terms of direct products

$$|\phi_{j_1} \otimes \phi_{j_2} \otimes \cdots \otimes \phi_{j_p}\rangle = |\phi_{j_1}\rangle \otimes |\phi_{j_2}\rangle \otimes \cdots \otimes |\phi_{j_p}\rangle$$

(2.15)
of one-particle kets,

$$\hat{D}_p = \sum_{i_1, \ldots, i_p; j_1, \ldots, j_p} D_{i_1, \ldots, i_p; j_1, \ldots, j_p} |\phi_{i_1} \otimes \phi_{i_2} \otimes \cdots \otimes \phi_{i_p}\rangle |\phi_{j_1} \otimes \phi_{j_2} \otimes \cdots \otimes \phi_{j_p}\rangle.$$ 

(2.16)

An important point is that $|\phi_{j_1} \otimes \cdots \otimes \phi_{j_p}\rangle$ is a simple direct product, whereas

$$|\phi_{j_1} \cdots \phi_{j_p}\rangle = \frac{1}{\sqrt{p!}} \hat{P}_p |\phi_{j_1} \otimes \cdots \otimes \phi_{j_p}\rangle$$

(2.17)
is standard notation for a normalized, antisymmetric $p$-electron basis function, which represents a proper $p$-electron state. [Here $\hat{P}_p$ denotes a sum of signed permutations of the indices $j_1, \ldots, j_p$; see Eq. (2.48).] The order of the indices in $|\phi_{j_1} \cdots \phi_{j_p}\rangle$ is irrelevant except for a phase factor, though it matters in $|\phi_{j_1} \otimes \cdots \otimes \phi_{j_p}\rangle$. The use of non-antisymmetric basis functions simplifies certain results in Chapter 5, where we discuss approximate parametrizations of $\hat{D}_2$ that violate the antisymmetry requirement. Note that in the $|\phi_{j_1} \otimes \cdots \otimes \phi_{j_p}\rangle$ basis, the expansion coefficients form a tensor $D_p$ that is antisymmetric tensor in both its upper and lower indices, whereas if antisymmetrized basis functions were employed, the summations in Eq. (2.7) would be restricted to $i_1 < i_2 < \cdots < i_p$ and $j_1 < j_2 < \cdots < j_p$. Our reliance on simple direct product states poses no conceptual problem, since matrix elements in the $|\phi_{j_1} \cdots \phi_{j_p}\rangle$ basis differ from those in the $|\phi_{j_1} \otimes \cdots \otimes \phi_{j_p}\rangle$ basis only by a factor of $p!:

$$\hat{D}_p = p! \sum_{i_1 < i_2 < \cdots < i_p; j_1 < j_2 < \cdots < j_p} D_{i_1, \ldots, i_p; j_1, \ldots, j_p} |\phi_{i_1} \otimes \cdots \otimes \phi_{i_p}\rangle |\phi_{j_1} \otimes \cdots \otimes \phi_{j_p}\rangle.$$ 

(2.18)

The utility of the 2-RDM is that it can be used to compute the expectation value of any one- or two-electron observable for an $N$-electron system, such as the
operator $\hat{W}$ defined in Eq. (2.1). The corresponding two-electron reduced operator $\hat{W}_{12}$ was cleverly constructed in order to reproduce the exact expectation value of $\hat{W}$, $w \equiv \langle \Psi | \hat{W} | \Psi \rangle = \text{tr}(\hat{W}_{12} D_2)$. Alternatively, since the 1-RDM can be used to compute the expectation value of any one-particle operator, a more straightforward way to compute $\langle \hat{W} \rangle$ is to partition this operator into a one-electron part

$$\hat{W}_1 \stackrel{\text{def}}{=} \sum_i \hat{h}(i)$$

(2.19)

and a two-electron part

$$\hat{W}_2 \stackrel{\text{def}}{=} \sum_{i<j} \hat{g}(i,j) .$$

(2.20)

Then $w = w_1 + w_2$ consists of a one-electron part

$$w_1 = \langle \Psi | \hat{W}_1 | \Psi \rangle = \text{tr}(\hat{h} \hat{D}_1)$$

(2.21)

and a two-electron part

$$w_2 = \langle \Psi | \hat{W}_2 | \Psi \rangle = \text{tr}(\hat{g} \hat{D}_2) .$$

(2.22)

The succinctness of these expressions is the reason for including the factor of $\binom{N}{p}$ in the definition of the $p$-RDM [Eq. (2.7)].

The matrix elements of $\hat{W}_{12}$ quantify the contribution of each 2-RDM element to the expectation value $w$. An essential feature of certain direct 2-RDM computational methodologies discussed in Chapters 3 and 4 is the need for approximate 3- and 4-RDMs as well. These higher-order RDMs can be used to compute $w$ in their own right, so to assess the quality of approximate 3- and 4-RDMs it is useful to possess three- and four-particle analogues of $\hat{W}_{12}$. This is accomplished easily enough. The appropriate reduced operators are

$$\hat{W}_{123} \stackrel{\text{def}}{=} \frac{2[\hat{h}(1) + \hat{h}(2) + \hat{h}(3)]}{(N-1)(N-2)} + \frac{\hat{g}(1,2) + \hat{g}(1,3) + \hat{g}(2,3)}{N-2} .$$

(2.23)
for the three-electron case and
\[
\hat{W}_{1234} \overset{\text{def}}{=} \frac{6[\hat{h}(1) + \hat{h}(2) + \hat{h}(3) + \hat{h}(4)]}{(N - 1)(N - 2)(N - 3)} + \frac{2[\hat{g}(1,2) + \hat{g}(1,3) + \hat{g}(1,4) + \hat{g}(2,3) + \hat{g}(2,4) + \hat{g}(3,4)]}{(N - 2)(N - 3)}
\]
(2.24)

for four electrons. Notice that these operators are invariant to permutations of the electronic coordinates, hence they do not distinguish between electrons. Generally, the numerical coefficient of the \( \hat{h} \) term in \( \hat{W}_{12\ldots p} \) equals \( \binom{N}{p} / N \) and the coefficient of the \( \hat{g} \) term is \( \binom{N}{p} / \binom{N}{2} \). The reader may verify that
\[
\langle \Psi | \hat{W} | \Psi \rangle = \text{tr}(\hat{W}_{123} \hat{D}_3) = \text{tr}(\hat{W}_{1234} \hat{D}_4)
\]
(2.25)

The \( p \)-RDM is the quantum-mechanical probability distribution for composite \( p \)-particles in an \( N \)-particle system,\(^*\) containing both population and coherence information. The term coherence, an artifact of spectroscopic applications of density matrix theory,\(^{29-31}\) is used generically to mean an off-diagonal density matrix element. For an arbitrary choice of \( p \)-electron states (not necessarily eigenfunctions of \( \hat{D}_p \)), the composite \( p \)-particles exist in a superposition of these states, and coherences may be conceptualized as transition probabilities among these \( p \)-electron states. The values of both populations and coherences depend upon the choice of basis states, hence the \( p \)-electron states themselves are at best only context-dependent models to help us understand the microscopic universe. On the other hand, the density
\[
\rho_p(\mathbf{x}_1, \ldots, \mathbf{x}_p) \overset{\text{def}}{=} D_p(\mathbf{x}_1, \ldots, \mathbf{x}_p; \mathbf{x}_1, \ldots, \mathbf{x}_p)
\]
(2.26)

\(^*\)For example, the kernel \( D_2(1; 2, 1^\prime; 2^\prime) \) represents the distribution of composite 2-particles, that is, electron pairs. Quantum mechanically, a composite \( p \)-particle is fundamentally different than \( p \) individual particles, and obeys different statistics.\(^{24,25}\) For example, a pair of electrons—sometimes termed a pairon\(^{24-26}\)—is neither a fermion nor a boson (despite certain statements in the literature to the contrary\(^{27,28}\)).
for any \( p \) out of the \( N \) electrons is invariant to our choice of model \( p \)-electron states. Note that the density as defined here retains a dependence on the spin coordinates \( \xi_i \). The spin-traced analogue of \( \rho_p \) is a \( p \)-electron charge density, which we denote as \( \rho_p^\circ \). These, along with spin-resolved (spin-indexed) densities, are discussed more fully in Subsection 2.3.4.

To insure that \( \rho_p \geq 0 \) everywhere, we require that \( \hat{D}_p \geq 0 \) as a kernel, operator, or matrix (that is, \( \hat{D}_p \) must be positive semidefinite). This condition also implies that each eigenvalue of \( \hat{D}_p \) is non-negative, which guarantees that the diagonal matrix elements of \( \hat{D}_p \) (populations) are non-negative in any basis.

Owing to the presence of off-diagonal density matrix elements, it is generally incorrect to visualize the composite \( p \)-particles in an \( N \)-electron system as a classical or statistical ensemble composed of well-defined \( p \)-electron states. In particular, the one-electron distribution \( D_1(x_1; x_1') \) is not a statistical ensemble of spin orbitals. \( \hat{D}_p \) can always be diagonalized, however, and in its eigenvector basis all coherences among \( p \)-electron states are zero. It is for this reason that the eigenvectors of \( \hat{D}_p \) are termed the “natural” \( p \)-states.\(^{19,32}\) In particular, the natural expansion of \( D_1 \) is

\[
D_1(x; x') = \sum_k \sum_{\sigma \in \{\alpha, \beta\}} n_{i,\sigma} \chi_{k,\sigma}^\ast(r') \chi_{k,\sigma}(r) \sigma^\ast(\xi') \sigma(\xi). \tag{2.27}
\]

The functions \( \chi_{k,\sigma}(r) \) are called natural orbitals (NOs) and the functions \( \chi_{k,\sigma}(r) \sigma(\xi) \) are called natural spin orbitals (NSOs). We have already introduced, in Eq. (1.6), the natural expansion of \( D_2 \) in terms of natural spin geminals, and noted that this expansion affords a simple sum-over-states formula [Eq. (1.9)] for the energy and other observables. The natural expansion of \( D_1 \) in Eq. (2.27) yields an analogous sum-over-states formula for the expectation value of any one-electron operator.
In this work we consider only $\hat{S}_z$ eigenstates, for which there is no loss of generality in separating the space- and spin-coordinates as indicated in Eq. (2.27). Even so, the spatial parts $\chi_{k,\sigma}$ can have different forms for different spins, hence both the orbitals and their occupation numbers generally retain a spin index. Singlet states are an important exception, for which $n_{i,\alpha} = n_{i,\beta}$ and $\chi_{k,\alpha} = \chi_{k,\beta}$.

It should be obvious to any chemist that the NSO populations must satisfy

$$0 \leq n_{i,\sigma} \leq 1. \quad (2.28)$$

In fact, $\hat{D}_1$ is ensemble $N$-representable if and only if $0 \leq n_{i,\sigma} \leq 1$ for each $i$ and $\sigma$.\textsuperscript{22}

It follows from Eq. (2.28) that the diagonal elements of $\hat{D}_1$, in any spin-orbital basis, are contained in the interval $[0,1]$. The eigenvalues of $\hat{D}_2$, on the other hand, are bounded between zero and $\left\lfloor N/2 \right\rfloor$,\textsuperscript{33} the integer part of $N/2$, but this is not sufficient to insure that $\hat{D}_2$ is $N$-representable. In contrast to the situation with $D_1$, it is likely impossible to formulate necessary and sufficient conditions for the $N$-representability of $D_2$ in terms of its eigenvalues alone; explicit eigenvector information is almost certainly required.\textsuperscript{*}

### 2.3.2 Second-quantized formalism

The position-space, integral-kernel formalism introduced above follows the formulation in Löwdin’s seminal work on RDM theory.\textsuperscript{19–21} Most of the older RDM literature uses this representation. However, the theory can be developed more compactly in second quantization, and the second-quantized formulation is especially useful for introducing perturbative expansions of the RDMs, which is a major topic in Chapter 4. Within

\textsuperscript{*}For an argument to this effect, consult page 80 of Ref. 34.
second quantization, the elements of the tensor $D_p$ introduced above are given by
\[
D_{i_1,\ldots,i_p;j_1,\ldots,j_p} = \frac{1}{p!} \langle \Psi | \hat{a}_{i_1}^\dagger \cdots \hat{a}_{i_p}^\dagger \hat{a}_{j_p} \cdots \hat{a}_{j_1} | \Psi \rangle ,
\] (2.29)
where $\{\hat{a}_k^\dagger\}$ and $\{\hat{a}_k\}$ are, respectively, the creation and annihilation operators for the basis $\{\phi_k\}$.

Risking a slight abuse of notation (forgivable so long as the context is specified), we will recycle the symbols $\hat{W}_1$, $\hat{W}_2$, and $\hat{W}_{12}$ to indicate the second-quantized formulations of the eponymous Hilbert-space operators. Thus we have
\[
\hat{W}_1 = \sum_{ij} h_{ji} \hat{a}_i^\dagger \hat{a}_j ,
\] (2.30)
\[
\hat{W}_2 = \frac{1}{2} \sum_{ijkl} g_{kl;ij} \hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_l \hat{a}_k ,
\] (2.31)
and
\[
\hat{W}_{12} = \frac{1}{2} \sum_{ijkl} W_{kl;ij} \hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_l \hat{a}_k .
\] (2.32)

The coefficients in the first two expansions are simply the spin-orbital matrix elements of $\hat{h}$ and $\hat{g}$:
\[
h_{ji} = \langle \phi_j(1) | \hat{h}(1) | \phi_i(1) \rangle
\] (2.33)
and
\[
g_{kl;ij} = \langle \phi_k(1) \otimes \phi_l(2) | \hat{g}(1,2) | \phi_i(1) \otimes \phi_j(2) \rangle .
\] (2.34)

The expansion coefficients for $\hat{W}_{12}$ are [compare Eq. (2.4)]
\[
W_{kl;ij} = g_{kl;ij} + \frac{h_{ki} \delta_{ji} + h_{lj} \delta_{ik}}{N - 1} .
\] (2.35)

Similarly, the three-electron reduced Hamiltonian [Eq. (2.23)] has matrix elements
\[
W_{ijk;rst} = \frac{2}{(N - 1)(N - 2)} \left[ h_{ir} \delta_{js} \delta_{kt} + h_{js} \delta_{ir} \delta_{kt} + h_{kt} \delta_{ir} \delta_{js} \right] + \frac{g_{ij;rs} \delta_{kt} + g_{jk;rt} \delta_{is} + g_{jk;st} \delta_{ir}}{N - 2} .
\] (2.36)
The coefficient tensors $g$ and $W$ introduced here are self-adjoint but not antisymmetric, although $g_{kl;ij} = g_{lk;ji}$ and therefore $W_{kl;ij} = W_{lk;ji}$. One could antisymmetrize $W$, but in so doing would destroy an important block structure, for if $\hat{W}$ is spin-independent then $W$ is diagonal in the spin variables of each particle. It follows immediately that

$$\langle \hat{W} \rangle = \text{tr}(\hat{W}_{12} \hat{D}_2) = \text{tr}(WD_2) = \sum_{ijkl} W_{ijkl} D_{kl;ij}.$$ (2.37)

Yet another formulation of the RDMs will prove useful in Chapter 4, where we develop a renormalized perturbation theory to compute the 3- and 4-RDMs. In that context, it is convenient to envisage the $p$-RDM as the time-independent limit of the so-called $p$-electron Green’s function ($p$-GF), also called the $p$-electron propagator. Modulo a time ordering and some factors of $i = \sqrt{-1}$ (see below), this generalization is obtained by using time-dependent creation and annihilation operators in Eq. (2.29).

Actually, our interest in the propagators themselves is somewhat incidental. In developing a perturbation theory for the RDMs (Chapter 4), we discovered that when formulated appropriately, the structure of the perturbation series is basically the same for both the time-dependent and time-independent versions of the theory, essentially as a consequence of Wick’s theorem. For the sake of formal manipulations, which are performed diagrammatically, it is therefore equally easy (or equally difficult) to develop both versions simultaneously. The resulting diagrammatic equations can then be translated into time-dependent or time-independent algebraic equations, according to taste, and it is only at this point that the two theories diverge from one another. In the hope that the time-dependent formalism might be useful elsewhere, the entire derivation was therefore formulated in terms of the more general $p$-GF.

The $p$-GF is a $p$-particle integral kernel $\bar{D}_p(x_1, t_1, \ldots, x_p, t_p; x'_1, t'_1, \ldots, x'_p, t'_p)$, in
which each of the coordinates carries a time argument in addition to space- and spin-
coordinates. The similarity between this notation ($\overline{D}_p$) and that for the $p$-RDM ($D_p$) is intentional, and generally we shall use an overline to indicate the time-dependent generalization of a density matrix quantity. For instance, in Section 2.5 we introduce the RDM cumulants $\Delta_{p_i}$; the corresponding Green’s function quantities are denoted $\overline{\Delta}_p$.

Upon introducing a finite spin-orbital basis set, $\overline{D}_p$ is equivalent to a tensor $\overline{D}_p$ whose spin-orbital matrix elements $\overline{D}_{j_1, \ldots, j_p; k_1, \ldots, k_p}$ are time dependent. These elements are [compare Eq. (2.29)]

$$
\overline{D}_{j_1, \ldots, j_p; k_1, \ldots, k_p}(t_1, \ldots, t_p; t'_1, \ldots, t'_p) = \frac{i^p}{p!} \langle \Psi | \hat{T} \left( \hat{a}_{j_1}(t_1) \cdots \hat{a}_{j_p}(t_p) \hat{a}_{k_p}(t'_p) \cdots \hat{a}_{k_1}(t'_1) \right) | \Psi \rangle .
$$

(2.38)

Here $\hat{a}_k(t) \equiv e^{it\hat{H}} \hat{a}_k e^{-it\hat{H}}$ is the Heisenberg representation of $\hat{a}_k$. The operator $\hat{T}$ is a signed permutation that brings the time-dependent operator product into descending time order from left to right, with creation operators $\hat{a}_{k}^\dagger$ on the left in the case of equal times. For example,

$$
\hat{T} \left( \hat{a}_{j}(t) \hat{a}_{k}^\dagger(t') \right) = \begin{cases} 
\hat{a}_{j}(t) \hat{a}_{k}^\dagger(t') & \text{if } t > t' \\
-\hat{a}_{k}^\dagger(t') \hat{a}_{j}(t) & \text{if } t \leq t'
\end{cases} .
$$

(2.39)

From the $p$-GF defined above, one obtains the $p$-RDM as the zero-time limit, with the caveat that this limit must produce the operator ordering in Eq. (2.29). The appropriate choice is

$$
D_{i_1, \ldots, i_p; j_1, \ldots, j_p} = i^{-p} \lim_{t_1, \ldots, t_p, t'_1, \ldots, t'_p \to 0} \overline{D}_{j_1, \ldots, j_p; k_1, \ldots, k_p}(t_1, \ldots, t_p; t'_1, \ldots, t'_p) .
$$

(2.40)
The factors of $i = \sqrt{-1}$, introduced in Eq. (2.38) and implied whenever time arguments are present, are explicitly eliminated in Eq. (2.40). Imaginary factors were introduced in Eq. (2.38) in deference to the usual convention for Green’s functions (see, for example, Ref. 37); these factors cause no trouble in the present work, as we never explicitly evaluate any formulas involving Green’s functions. The reader is cautioned that the factor of $1/p!$ included in Eq. (2.38) is a departure from the usual convention for the $p$-GF, but is perfectly natural here, given our prior definition of the $p$-RDM.

Thus far we have discussed only pure-state RDMs, derivable from a wavefunction $\Psi$. The definitions in Eqs. (2.7) and (2.29), however, are easily extended to the case of an ensemble state $\sum_i c_i |\Psi_i\rangle\langle\Psi_i|$ that cannot be described by a single wavefunction, but instead consists of a convex linear combination of pure-state density matrices. In particular, note that the expectation values used above in conjunction with $\Psi$ are trivially reformulated as expectation values over the $N$-electron density matrix $D_N$; these more general definitions are valid for ensembles as well as for pure states. Although we are most interested in the electronic properties of pure states, in Chapter 3 we will have reason to discuss ensemble solutions to the contracted Schrödinger equation.

### 2.3.3 Hole matrices

It is well known that within the finite-basis formulation of indistinguishable-particle quantum mechanics, a system consisting of $p$ particles (and therefore $R - p$ holes, where $R$ is the number of spin-orbital basis functions) is equivalent to a system of $R - p$ particles and $p$ holes. In terms of the RDMs, which we have tacitly assumed to be particle density matrices, this means that $D_p = Q_{R-p}$, where $Q_{R-p}$ denotes a RDM
for holes. We shall use the term “p-RDM” exclusively to mean a p-electron RDM, so let us introduce the abbreviation p-HRDM for the p-hole reduced density matrix, denoted symbolically as $Q_p$ and with tensor elements

$$Q_{i_1,\ldots,i_p;j_1,\ldots,j_p} \overset{\text{def}}{=} \frac{1}{p!} \langle \Psi | \hat{a}_{i_1} \cdots \hat{a}_{i_p} \hat{a}_{j_p} \cdots \hat{a}_{j_1} | \Psi \rangle . \quad (2.41)$$

Comparison of the Eq. (2.41) with Eq. (2.29), which defines the p-RDM elements, suggests a close relationship between $D_p$ and $Q_p$. Indeed, the expectation value of the fermion anticommutation rule

$$\hat{a}_i \hat{a}_j + \hat{a}_j \hat{a}_i = \delta_{i,j} \quad (2.42)$$

with respect to $|\Psi\rangle$ yields the relationship

$$D_1 + Q_1^\dagger = I , \quad (2.43)$$

where $I$ is an identity matrix. Using this relationship, one can easily show that $D_1 Q_1^\dagger = Q_1^\dagger D_1$, hence $D_1$ and $Q_1^\dagger$ share a common set of eigenvectors. In their mutual NSO representation, Eq. (2.43) is clearly a restatement of the Pauli-Coleman condition, $0 \leq n_{i,\sigma} \leq 1$.

A composite anticommutation rule for two indistinguishable fermions is obtained by repeated application of the one-particle anticommutation relation, to transform a two-hole excitation operator into a sum of excitation operators for particles. The result is

$$\hat{a}_i \hat{a}_j \hat{a}_k^\dagger \hat{a}_l^\dagger = \delta_{ik} \delta_{jl} - \delta_{il} \delta_{jk} + \hat{a}_i^\dagger \hat{a}_k^\dagger \hat{a}_i \hat{a}_j$$

$$+ \hat{a}_i^\dagger \hat{a}_i \delta_{jk} + \hat{a}_k^\dagger \hat{a}_j \delta_{il} - \hat{a}_k^\dagger \hat{a}_i \delta_{jl} - \hat{a}_i^\dagger \hat{a}_j \delta_{ik} . \quad (2.44)$$
The expectation value of this equation yields an element of $Q_2$ on the left; on the right we obtain elements of $I$, $D_1$, and $D_2$:

$$Q_{ij;kl} = \delta_{ik} \delta_{jl} - \delta_{il} \delta_{jk} + D_{l;i} \delta_{jk} + D_{k;j} \delta_{il} - D_{k;i} \delta_{jl} - D_{l;j} \delta_{ik} + D_{kl;ij} .$$

Equation (2.45) can be written in a more compact form if we introduce an antisymmetrized product (Grassmann product) “$\wedge$” for tensors. If $T = (T_{i_1,\ldots,i_n}^{j_1,\ldots,j_m})$ and $S = (S_{j_1,\ldots,j_m}^{i_1,\ldots,i_n})$ are tensors, then their Grassmann product $T \wedge S$ is the tensor with elements\textsuperscript{18,38,39}

$$\left( T \wedge S \right)_{j_1,\ldots,j_{m+n}}^{i_1,\ldots,i_{m+n}} = \left( \hat{A}_{m+n} \otimes \hat{A}_{m+n} \right) T_{i_1,\ldots,i_n}^{j_1,\ldots,j_m} S_{j_{n+1},\ldots,j_{m+n}}^{i_{n+1},\ldots,i_{m+n}}$$

$$= [(m + n)!]^{-2} \sum_{\sigma, \pi \in \mathfrak{S}_{m+n}} \epsilon(\sigma) \epsilon(\pi) T_{\pi(j_1),\ldots,\pi(j_n)}^{\sigma(i_1),\ldots,\sigma(i_n)} S_{\pi(j_{n+1}),\ldots,\pi(j_{m+n})}^{\sigma(i_{n+1}),\ldots,\sigma(i_{m+n})} .$$  

(2.46)

We also define “Grassmann powers” (sometimes called \textit{exterior products})

$$T^{\wedge n} = T \wedge T \wedge \cdots \wedge T \quad \text{n factors}$$

(2.47)

of a tensor $T$. The tensor $T^{\wedge n}$ is to be distinguished from $T^n$, the latter denoting the usual matrix product.

Some comments on this new notation are warranted. Previously, we made reference to the operator $\hat{P}_n$, the sum of signed permutations on $n$ coordinates or $n$ indices. Operating on a function $f = f(1,\ldots,n)$, this operator yields

$$\hat{P}_n f(1,\ldots,n) = \sum_{\pi \in \mathfrak{S}_n} \epsilon(\pi) f(\pi(1),\ldots,\pi(n)) ,$$

(2.48)

where $\mathfrak{S}_n$ is the symmetric group of order $n!$ and $\epsilon(\pi) = \pm 1$ is the parity of the permutation $\pi$. Instead of $\hat{P}_n$, it is sometimes convenient to utilize an antisymmetrizing projection operator

$$\hat{A}_n \overset{\text{def}}{=} \frac{1}{n!} \hat{P}_n ,$$

(2.49)
which has the property $\hat{A}_n^2 = \hat{A}_n$. Thus $\hat{A}_n$ acts as the identity operator on any antisymmetric function, in contrast to $\hat{P}_n$.

We will also utilize $\hat{P}_n$ and $\hat{A}_n$ to antisymmetrize tensor indices. In this context, we may wish to antisymmetrize the upper indices, the lower indices, or both. Thus in Eq. (2.46) we have introduced $\overline{A}_{m+n}$, an antisymmetrizing projection operator for the upper $m + n$ indices, along with the corresponding operator $\overline{A}_{m+n}$ for the lower indices. $\overline{P}_{m+n}$ and $\overline{P}_{m+n}$ are similarly defined. According to Eq. (2.46), when acting on a direct product of tensors whose overall particle number is $m + n$, the symbol "\wedge" replaces the operator $\overline{A}_{m+n} \otimes \overline{A}_{m+n} = (m + n)!^{-2} \overline{P}_{m+n} \otimes \overline{P}_{m+n}$. Note also that the Grassmann product \wedge is commutative, distributive, and associative, and $T \wedge S$ is a linear operator on the $(m + n)$-particle space. If $T$ and $S$ are self-adjoint then so is $T \wedge S$.

Using this new notation, Eq. (2.45) is rewritten as the tensor identity

$$Q_2^\dagger = I \wedge I - 2 I \wedge D_1 + D_2.$$  \hfill (2.50)

Thus the 2-HRDM is a functional of the 1- and 2-RDMs. The appearance of a matrix adjoint in this equation is a somewhat annoying consequence of our definition of $Q_p$. This choice was made for later consistency [specifically, with Eq. (2.77)] and is the standard convention. In the absence of a magnetic field, one may assume that the RDMs are real and symmetric, in which case the adjoint symbols can be omitted.

Using Wick’s Theorem\cite{35,36} for manipulating strings of creation and annihilation operators, it is possible to generalize the above procedure and thereby show that $Q_p = Q_p[D_1, \ldots, D_p]$. (A graphical technique developed by Valdemoro and co-workers\cite{40-43} is helpful for larger $p$.) The resulting HRDM functional can be written in a form
that is highly reminiscent of the binomial theorem, wherein the ordinary product is replaced by an antisymmetrized product. With the convention that $D_0$ and $I^0$ each denote the scalar 1, the result is \cite{13,39,42}

$$Q_p^\dagger = \sum_{n=0}^{p} (-1)^n \left( \begin{array}{c} p \\ n \end{array} \right) [D_n \wedge I^{(p-n)}]. \quad (2.51)$$

Similar arguments lead to a functional

$$D_p^\dagger = \sum_{n=0}^{p} (-1)^n \left( \begin{array}{c} p \\ n \end{array} \right) [Q_n \wedge I^{(p-n)}] \quad (2.52)$$

that expresses particle RDMs in terms of HRDMs of equal and lower order. These relationships, which are equivalent to expectation values of the fermion anticommutation relations for composite $p$-particles, form the basis of a density matrix reconstruction scheme that is introduced in Subsection 4.2.1.

### 2.3.4 Spin components

For computational purposes, RDMs and related quantities are expanded in products of one-particle basis functions. In particular, quantities such as $D_p(x_1, \ldots, x_p; x'_1, \ldots, x'_p)$ that depend on space/spin coordinates $x_i$ must be expanded in a spin-orbital basis. Owing to certain permutational symmetries, these expansions are considerably more compact if one separates the space- and spin-variables, and expands spin-indexed quantities in an orbital basis set. The spin-indexed blocks of the RDMs are known as its spin components. The decomposition of $D_p$ into spin components depends on the assumption that $D_p$ represents an $\hat{S}_z$ eigenstate, for which there is no loss of generality in using the same set of orbitals to expand both $\alpha$- and $\beta$-spin quantities, provided that the orbitals are allowed to have different forms for different spins. Decomposition of $D_p$ into spin blocks is a useful way to weed out matrix elements that are redundant,
by virtue of self-adjointness, antisymmetry, and also certain spin symmetries discussed below. More general treatments of spin bases have been given,\textsuperscript{44,45} but here we are interested only in simple direct-product basis functions [compare Eq. (2.15)]

\[ |\sigma_1 \otimes \sigma_2 \otimes \cdots \otimes \sigma_p\rangle = |\sigma_1\rangle \otimes |\sigma_2\rangle \otimes \cdots \otimes |\sigma_p\rangle , \]

(2.53)

where each \( \sigma_i \in \{\alpha, \beta\} \).

Henceforth we consider only \( \hat{S}_z \) eigenstates. Then we may suppose that the spin-orbital basis \( \{|\phi_1\rangle, \ldots, |\phi_R\rangle\} \) introduced previously is constructed from a basis of orthonormal orbitals \( \{|\chi_1\rangle, \ldots, |\chi_\nu\rangle\} \), where \( R = 2\nu \). Order the spin-orbitals such that

\[ |\phi_\kappa\rangle = \begin{cases} 
|\chi_\kappa\rangle \otimes |\alpha\rangle & \text{if } \kappa \leq \nu \\
|\chi_{\kappa-\nu}\rangle \otimes |\beta\rangle & \text{if } \kappa > \nu. 
\end{cases} \]

(2.54)

Spin-orbitals constructed in this way—as direct products of an orbital \( \chi_k \) and a spin function \( \sigma \in \{\alpha, \beta\} \)—are said to possess \textit{pure spin-type}.

Let \( \hat{\mathcal{K}} \) be an arbitrary two-electron operator [equivalently, \( \mathcal{K}(x_1, x_2; x'_1, x'_2) \) is an arbitrary two-electron kernel]. The \textit{spin components} of \( \hat{\mathcal{K}} \) are the expansion coefficients in the spin basis \( \{|\sigma_1 \otimes \sigma_2\rangle|\sigma'_1 \otimes \sigma'_2\rangle\} \). This expansion is

\[ \hat{\mathcal{K}} = \hat{\mathcal{K}}_{\alpha\alpha\alpha\alpha}|\alpha \otimes \alpha\rangle\langle\alpha \otimes \alpha| + \hat{\mathcal{K}}_{\alpha\beta\alpha\beta}|\alpha \otimes \beta\rangle\langle\alpha \otimes \beta| + \hat{\mathcal{K}}_{\beta\alpha\beta\alpha}|\beta \otimes \alpha\rangle\langle\beta \otimes \alpha| + \hat{\mathcal{K}}_{\beta\beta\beta\beta}|\beta \otimes \beta\rangle\langle\beta \otimes \beta| \]

(2.55)

\[ + \hat{\mathcal{K}}_{\alpha\beta\beta\alpha}|\alpha \otimes \beta\rangle\langle\beta \otimes \alpha| + \hat{\mathcal{K}}_{\beta\alpha\alpha\beta}|\beta \otimes \alpha\rangle\langle\alpha \otimes \beta| + \hat{\mathcal{K}}_{\beta\beta\alpha\beta}|\beta \otimes \beta\rangle\langle\alpha \otimes \beta| + \hat{\mathcal{K}}_{\beta\alpha\beta\beta}|\beta \otimes \beta\rangle\langle\beta \otimes \beta| \].

The coefficients \( \hat{\mathcal{K}}_{\sigma_1\sigma_2\sigma'_1\sigma'_2} \) are yet operators on the orbital space, that is, they possess an expansion in dyads \(|\chi_i \otimes \chi_j\rangle\langle\chi_k \otimes \chi_l|\). [Alternatively, \( \mathcal{K}_{\sigma_1\sigma_2\sigma'_1\sigma'_2}(r_1, r_2; r'_1, r'_2) \) is a spin-indexed spatial kernel.] Note that the two-electron spin basis \( \{|\sigma_1 \otimes \sigma_2\rangle|\sigma'_1 \otimes \sigma'_2\rangle\} \) contains 16 functions, but owing to the requirement that \( m_{s(1)} + m_{s(2)} = m_{s(1')} + m_{s(2')} \) for \( \hat{S}_z \) eigenstates, only the six terms given in Eq. (2.55) may have nonzero coefficients.
Now suppose that $\hat{K}$ is antisymmetric in both its primed and its unprimed co-
ordinates; then only three of the six spin components in Eq. (2.55) are independent.\textsuperscript{44}
We will take as independent spin components $\hat{K}_{aa} \equiv \hat{K}_{aaaa}$, $\hat{K}_{a\beta} \equiv \hat{K}_{a\alpha\beta}$, and $\hat{K}_{\beta\beta} \equiv \hat{K}_{\beta\beta\beta\beta}$. The other nonzero spin component kernels can be obtained as coordinate permutations of these three. For example,

$$K_{\alpha\beta\alpha\beta}(r_1, r_2; r'_1, r'_2) = -K_{\alpha\beta\alpha\beta}(r_1, r_2; r'_1, r'_2).$$

(2.56)

In the case of the 2-RDM ($K = D_2$), the quantity $D_2^{\sigma\sigma'}(r_1, r_2; r'_1, r'_2)$ represents a spin-indexed two-electron probability distribution, where electron 1 has spin $\sigma$ and electron 2 has spin $\sigma'$. This is fine for the parallel-spin case, and we define a two-
electron charge density

$$\rho_2^{\sigma\sigma}(r_1, r_2) \overset{\text{def}}{=} D_2^{\sigma\sigma}(r_1, r_2; r_1, r_2)$$

(2.57)

for parallel-spin electrons. In the opposite-spin case, a trace over spin variables annihilates the spin components $D_2^{\alpha\beta\alpha\alpha}$ and $D_2^{\beta\alpha\alpha\beta}$, hence only $D_2^{\beta\alpha\beta\alpha}$ and $D_2^{\beta\alpha\beta\alpha}$ contribute to the two-electron charge density for opposite-spin electrons. To avoid making an artificial distinction between the two indistinguishable electrons, we define this opposite-
spin density to be

$$\rho_2^{\alpha\beta}(r_1, r_2) \overset{\text{def}}{=} D_2^{\alpha\beta}(r_1, r_2; r_1, r_2) + D_2^{\beta\alpha}(r_1, r_2; r_1, r_2)$$

$$= D_2^{\alpha\beta}(r_1, r_2; r_1, r_2) + D_2^{\alpha\beta}(r_2, r_1; r_1, r_2).$$

(2.58)

(Note the coordinate permutations in the second line.) The total two-electron charge density (2-CD) $\rho_2^\circ = \rho_2^\circ(r_1, r_2)$ is then

$$\rho_2^\circ = \rho_2^{\alpha\alpha} + \rho_2^{\alpha\beta} + \rho_2^{\beta\beta},$$

(2.59)
which should be compared to the spin-dependent two-electron density $\rho_2(x_1, x_2)$ that was defined in Eq. (2.26). Similarly, the 1-CD $\rho_1^0$ arises from the spinless 1-RDM and has spin-indexed components $\rho_1^0$ and $\rho_1^\beta$.

Taking matrix elements of $\hat{K}_{\sigma\sigma'}$ in the $\{|\chi_i \otimes \chi_j\}$ basis, these operators are converted to tensors $K_{\sigma\sigma'}$ having elements $(K_{\sigma\sigma'})_{ij}^{kl}$ with $i, j, k, l \leq \nu$. In contrast, matrix elements of the original operator $\hat{K}$ must be taken in the $\{|\phi_i \otimes \phi_j\}$ basis, so that when we write $(K)^{ij}_{kl}$, the indices refer to spin orbitals, $i, j, k, l \leq 2\nu$. For notational brevity we will sometimes employ the standard notation $\bar{i}$ to index the spin orbital $|\chi_i\rangle \otimes |\beta\rangle$; according to our indexing scheme

$$\bar{i} = i + \nu .$$

(2.60)

Usually, when $\bar{i}$ is used to indicate $|\chi_i\rangle \otimes |\beta\rangle$, the denuded symbol $i$ is taken to mean $|\chi_i\rangle \otimes |\alpha\rangle$, but this is potentially confusing since we use $i$ for an orbital index (no spin implied). Hence, whenever we employ the “$\bar{i}$” notation, we will also use $\check{i}$ to mean $|\chi_i\rangle \otimes |\alpha\rangle$. In terms of a counting index, $\check{i} = i$, according to the ordering introduced in Eq. (2.54).

This notation lets us write down simple correspondences between the matrix elements of $K$ and those of its spin components. For example,

$$(K_{\alpha\beta})^{ij}_{kl} = K^{i\bar{j}}_{k\check{l}} .$$

(2.61)

From these expressions one can easily derive the permutational symmetries that connect the various spin components. The tensor formulation of Eq. (2.56), for example, is

$$(K_{\alpha\beta})^{ij}_{kl} = K^{i\bar{j}}_{k\check{l}} = -K^{\bar{i}j}_{\check{k}l} = -(K_{\alpha\beta\alpha\beta})^{ij}_{kl} .$$

(2.62)
The second equality follows from the antisymmetry of $\mathcal{K}$, while the other two equalities are merely definitions. Similar relations exist among other spin components, ultimately leaving $\mathcal{K}_{\alpha\alpha}$, $\mathcal{K}_{\gamma\beta}$, and $\mathcal{K}_{\beta\beta}$ as the only independent ones. $\mathcal{K}_{\alpha\alpha}$ and $\mathcal{K}_{\beta\beta}$ are antisymmetric in both their primed and their unprimed coordinates, while $\mathcal{K}_{\alpha\beta}$ exhibits no special symmetry. The lack of antisymmetry in $\mathcal{K}_{\alpha\beta}$ is sometimes confusing to those who have not worked through Eq. (2.62). To rationalize this property in one’s mind, realize that an electron with $\alpha$ spin is distinguishable from one with $\beta$ spin, hence there is no requirement that the the wavefunction or RDMs be antisymmetric with respect to permutations of the coordinates of these two electrons.

We can take advantage of spin-component symmetries to store the 2-RDM using the fewest possible array elements. We next determine the required number of elements, as a function of the number of orbitals $\nu$. Consider a typical element $D_{ijkl}^{\alpha\alpha}$, with $i, j, k, l \leq \nu$. Since $D_2^{\alpha\alpha}$ is self-adjoint and antisymmetric, we require only those elements with $i < j$, $k < l$, and $i \leq k$. Furthermore, when $i = k$ we need only the elements with $j \leq l$. The number of elements satisfying these requirements is

$$N_{\alpha\alpha} = \sum_{i=1}^{\nu-1} \sum_{j=i+1}^{\nu} \left[ \sum_{k=i}^{\nu-1} \sum_{l=k+1}^{\nu} \frac{1}{2} - \sum_{t=i+1}^{\nu} \frac{1}{2} \right]$$

$$= \frac{1}{8} \nu (\nu^3 - 2\nu^2 + 3\nu - 2).$$

The second of the terms in brackets subtracts the number of elements with $i = k$ and $j > l$. Clearly $N_{\beta\beta} = N_{\alpha\alpha}$, while a similar exercise demonstrates that

$$N_{\alpha\beta} = \frac{1}{2} \nu^2 (\nu^2 + 1).$$

Thus the total number of independent 2-RDM elements is

$$2N_{\alpha\alpha} + N_{\alpha\beta} = \frac{1}{8} \nu (5\nu^3 - 2\nu^2 + 7\nu - 2).$$
This is comparable to the number of independent two-electron integrals,\(^{46}\)

\[
N_{\text{integrals}} = \frac{1}{8} \nu (\nu + 1) (\nu^2 + \nu + 2),
\]

and represents the number of matrix elements that must be stored in an iterative solution of the contracted Schrödinger equation. The tensor \(D_2\) contains a total of \(16\nu^4\) elements, so as \(\nu \to \infty\) the number of independent elements relative to the number of total elements approaches a ratio of \(5/8 = 0.625\). For singlet states \(D_2^{\alpha \alpha} = D_2^{\beta \beta}\), which further decreases the number of independent elements. A thorough discussion of the indexing schemes required in order to fully exploit permutational and also spin symmetries of \(D_2\) and the higher RDMs can be found in Appendix A.

Finally in this subsection, let us find out how the three independent spin components of the 2-RDM are normalized, in the case that \(\text{tr} \, D_2 = \binom{N}{2}\). This derivation is simplest in second quantization, but in contrast to the discussion in Subsection 2.3.2, here we require creation and annihilation operators for orbitals rather than spin orbitals. The difference is basically one of notation. In terms of the spin-orbital annihilation operators \(\hat{a}_i\) introduced in Subsection 2.3.2, define

\[
\hat{a}_{i,\alpha} \overset{\text{def}}{=} \hat{a}_i, \quad \hat{a}_{i,\beta} \overset{\text{def}}{=} \hat{a}_\gamma.
\]

We will always include an explicit (Greek) spin index to indicate these new operators, whereas the notation \(\hat{a}_i\) will be reserved for a spin-orbital annihilation operator. The anticommutation relations for the orbital creation and annihilation operators are

\[
\hat{a}_{i,\sigma}^\dagger \hat{a}_{j,\sigma'} + \hat{a}_{j,\sigma'}^\dagger \hat{a}_{i,\sigma} = \delta_{ij} \delta_{\sigma\sigma'}
\]
and
\[ \hat{a}_{i,\sigma} \hat{a}_{j,\sigma'} = \hat{a}_{j,\sigma'} \hat{a}_{i,\sigma} (\delta_{ij} \delta_{\sigma\sigma'} - 1) . \] (2.69)

These operators generate elements of the spin components of \( D_2 \) according to
\[ D_{ij;kl}^{\sigma\sigma'} = \frac{1}{2} \langle \Psi | \hat{a}_{i,\sigma} \hat{a}_{j,\sigma'}^\dagger, \hat{a}_{l,\sigma'} \hat{a}_{k,\sigma} | \Psi \rangle . \] (2.70)

It follows from the anticommutation relations given above that
\[ \sum_{jk} \langle \Psi | \hat{a}_{j,\sigma}^\dagger \hat{a}_{k,\sigma}^\dagger \hat{a}_{k,\sigma} \hat{a}_{j,\sigma} | \Psi \rangle = \sum_{k} \sum_{j \neq k} \langle \Psi | \hat{a}_{j,\sigma} \hat{a}_{j,\sigma}^\dagger \hat{a}_{k,\sigma}^\dagger \hat{a}_{k,\sigma} | \Psi \rangle . \] (2.71)

Note that \( \hat{a}_{k,\sigma}^\dagger \hat{a}_{k,\sigma} \) is the number operator for the spin orbital \(|\chi_k\rangle \otimes |\sigma\rangle\), whose occupancy is \( D_{k,k}^\sigma \). Thus
\[ \hat{a}_{k,\sigma}^\dagger \hat{a}_{k,\sigma} | \Psi \rangle = D_{k,k}^\sigma | \Psi \rangle . \] (2.72)

Introducing \( N_\sigma = \text{tr} D_1^\sigma \), the number of electrons with spin \( \sigma \), Eq. (2.71) reduces to
\[ \sum_{jk} \langle \Psi | \hat{a}_{j,\sigma}^\dagger \hat{a}_{k,\sigma}^\dagger \hat{a}_{k,\sigma} \hat{a}_{j,\sigma} | \Psi \rangle = \sum_{k} \sum_{j \neq k} D_{j,j}^\sigma D_{k;k}^\sigma = N_\sigma (N_\sigma - 1) , \] (2.73)
whence
\[ \text{tr} D_2^{\sigma\sigma} = \left( \frac{N_\sigma}{2} \right) . \] (2.74)

Thus \( \text{tr} D_2^{\sigma\sigma} \) equals the number of \( \sigma \)-spin electron pairs. This result is independent of the manner in which the spin angular momenta are coupled, therefore it is valid for any multiplet, irrespective of the value of \( \langle S^2 \rangle \), and holds also for ensembles that are not necessarily \( \hat{S}^2 \) eigenstates. (This follows immediately if the pure-state expectation values in the derivation above are replaced by ensemble expectation values.)

One could obtain \( \text{tr} D_2^{\alpha\beta} \) in a similar fashion. Alternatively, observe that
\[ \text{tr} D_2 = \text{tr} D_2^{\alpha\alpha} + \text{tr} D_2^{\beta\beta} + 2 \text{tr} D_2^{\alpha\beta} = \left( \frac{N_\alpha + N_\beta}{2} \right) . \] (2.75)
Using our result for $\text{tr} \, D_2^{\sigma \sigma}$, one then obtains

$$\text{tr} \, D_2^{\alpha \beta} = \frac{N_{\alpha}N_{\beta}}{2},$$

which equals the number of opposite-spin electron pairs. These results are probably obvious, but the fact that they do not depend upon $\langle \hat{S}^2 \rangle$ is worth noting, since certain other decompositions\textsuperscript{32,47} of the spin-traced $D_2$ do depend upon $\langle \hat{S}^2 \rangle$.

### 2.3.5 Some $N$-representability constraints

Each of the definitions for a RDM that were introduced in Subsections 2.3.1 and 2.3.2 assumes that the electronic wavefunction is known, whereas the goal of this work is to obtain $D_2$ without first calculating $\Psi$. As summarized in Chapter 1, the essence of the so-called $N$-representability problem—the primary nemesis in density matrix theory—is an incomplete knowledge of the boundary conditions that insure that a reduced, two-electron distribution function $D_2(x_1, x_2; x'_1, x'_2)$ corresponds to an actual $N$-electron state. One might, without reference to an underlying wavefunction, define a $p$-electron density matrix to be any positive (semidefinite), antisymmetric, normalized, and self-adjoint kernel $D_p(x_1, \ldots, x_p; x'_1, \ldots, x'_p)$. This is fine for $p$-particle systems, but the upshot of the $N$-representability problem is that these conditions are not sufficient to guarantee that $D_p$ corresponds to a reduced density matrix for any system of $N > p$ electrons.

The antisymmetry, positivity, and self-adjointness requirements, the contraction relationship $\frac{1}{2} \downarrow \hat{D}_2 = \hat{D}_1$, and the trace condition $\text{tr} \hat{D}_2 = \binom{N}{2}$ each constitute $N$-representability conditions on $\hat{D}_2$. A large number of additional, necessary conditions for 2-RDM $N$-representability are known.\textsuperscript{22,34,48–71} Of these, we will focus on three matrix positivity conditions that are relatively simple yet quite restrictive in practice.
These conditions—positivity of the two-particle density matrix ($\hat{D}_2$), positivity of the two-hole density matrix ($\hat{Q}_2$), and positivity of the particle-hole density matrix ($\hat{G}_2$, defined below)—are known in the common parlance as the $D$-, $Q$-, and $G$-conditions, respectively.\textsuperscript{51}

We have already introduced the tensors $\mathbf{D}_2$ and $\mathbf{Q}_2$, but in order to provide a unified presentation we next introduce a tensor $\mathbf{P}_2$ that can be used to define $\mathbf{D}_2$, $\mathbf{Q}_2$, and $\mathbf{G}_2$ all at once. Let $\hat{A}_{i,j}$ denote a product of two creation and/or annihilation operators for some orthonormal spin-orbital basis, and define

$$ P_{ij;kl} \overset{\text{def}}{=} \frac{1}{2} (\Psi | \hat{A}_{i,j} \hat{A}_{k,l}^\dagger | \Psi ) \quad (2.77) $$

to be an element of $\mathbf{P}_2$. Up to an overall constant, $\mathbf{P}_2$ is the metric (overlap) matrix for the states $\hat{A}_{k,l}^\dagger | \Psi \rangle$, which indicates that $\mathbf{P}_2$ must be positive (semidefinite). As we have seen already, the choice $\hat{A}_{i,j} = \hat{a}_i^\dagger \hat{a}_j^\dagger$ defines matrix elements of $\mathbf{D}_2$ (i.e., $\mathbf{P}_2 = \mathbf{D}_2$ in this case), while $\hat{A}_{i,j} = \hat{a}_i \hat{a}_j$ defines matrix elements of $\mathbf{Q}_2$. The choice $\hat{A}_{i,j} = \hat{a}_j^\dagger \hat{a}_i$ generates elements of a tensor that we call $\mathbf{G}_2$. The selection $\hat{A}_{i,j} = \hat{a}_i \hat{a}_j^\dagger$ does not generate an independent positivity condition, as the corresponding metric matrix is positive if and only if $\mathbf{G}_2$ is positive.\textsuperscript{71} We note in passing that Eq. (2.77) can clearly be generalized to a generator for $p$-particle positivity conditions; the three- and four-particle cases have been discussed by several authors.\textsuperscript{69–72}

The condition $\mathbf{D}_2 \geq 0$ is obviously a constraint on $\mathbf{D}_2$, but so are the conditions $\mathbf{Q}_2 \geq 0$ and $\mathbf{G}_2 \geq 0$. This can be seen by rearranging the creation and annihilation operators in Eq. (2.77) to obtain $\mathbf{Q}_2$ and $\mathbf{G}_2$ as functionals of the 2-RDM. The functional $\mathbf{Q}_2(\mathbf{D}_1, \mathbf{D}_2)$ has been given in Eq. (2.45), while the functional $\mathbf{G}_2(\mathbf{D}_1, \mathbf{D}_2)$,
expressed in terms of its matrix elements, is

\[ G_{ijkl} = \frac{1}{2} D_{ik} \delta_{jl} - D_{il} \delta_{kj} \, . \]  

(2.78)

It bears mention that our \( G \)-matrix \( G_2 \) differs from the \( G \)-matrix that was originally introduced into the quantum chemistry literature by Garrod and Percus.\(^{51}\) Let \( G^{GP} \) denote the latter. Modifying the definition in Ref. 51 to account for our normalization convention, the Garrod-Percus \( G \)-matrix has elements

\[ G_{ijkl}^{GP} = \frac{1}{2} D_{ji} \delta_{ik} - \frac{1}{2} D_{kj} \delta_{il} - D_{kj} \delta_{il} \, . \]  

(2.79)

(The factors of \( \frac{1}{2} \) do not appear in Ref. 51 due to the different normalization convention.) That \( G^{GP} \) must be positive follows by recognizing that \( G^{GP} \) can be defined as a special case of the metric tensor \( P_2 \) via the choice \( \hat{A}_{i,j} = \hat{a}_i \dagger \hat{a}_j - \langle \Psi | \hat{a}_i \dagger \hat{a}_j | \Psi \rangle \).

One can partition the tensors \( Q_2 \) and \( G_2 \) into spin components, each of which must be separately positive. The spin components of \( Q_2 \) are

\[ Q_{ij:kl}^{\alpha\alpha} \overset{\text{def}}{=} \langle \chi_i \otimes \chi_j | \hat{Q}_2^{\alpha\alpha} | \chi_k \otimes \chi_l \rangle \]

\[ = D_{kl:ij}^{\alpha\alpha} + \frac{1}{2} \left\{ \delta_{ij} \left[ \delta_{ki} - D_{k:i}^{\alpha} \right] - \delta_{il} \left[ \delta_{kj} - D_{k:j}^{\alpha} \right] + \delta_{kj} D_{l:i}^{\alpha} - \delta_{li} D_{l:j}^{\alpha} \right\} \]  

(2.80)

and

\[ Q_{ij:kl}^{\alpha\beta} \overset{\text{def}}{=} \langle \chi_i \otimes \chi_j | \hat{Q}_2^{\alpha\beta} | \chi_k \otimes \chi_l \rangle \]

\[ = D_{kl:ij}^{\alpha\beta} + \frac{1}{2} \left\{ \delta_{ik} \delta_{jl} - D_{k:i}^{\alpha} \delta_{jl} - D_{l:i}^{\alpha} \delta_{ik} \right\} \, , \]  

(2.81)

while for \( G_2 \) one obtains

\[ G_{ij:kl}^{\alpha\alpha} = D_{il:jk}^{\alpha\alpha} + \frac{1}{2} D_{i:k}^{\beta} \delta_{jl} \]  

(2.82)

and

\[ G_{ij:kl}^{\alpha\beta} = -D_{il:jk}^{\alpha\beta} + \frac{1}{2} D_{i:k}^{\alpha} \delta_{jl} \, . \]  

(2.83)
Note that these are *orbital* matrix elements, whereas Eqs. (2.77)–(2.79) define *spin-orbital* matrix elements.

The $D$, $Q$, and $G$-conditions have a long history as constraints for variational 2-RDM calculations. These conditions are not sufficient to guarantee that the 2-RDM is $N$-representable, thus variation of $D_2$ subject to these constraints does not yield an upper bound to the FCI electronic energy, but one may consider the extent of overcorrelation to be an indicator of the stringency of these or any other partial set of $N$-representability constraints. For light atoms like $^4$Be and its isoelectronic homologues, a variational 2-RDM calculation with the constraints $D_2 \geq 0$, $Q_2 \geq 0$, and $G_2 \geq 0$ yields 100.0\% of the total electronic energy, while for heavier atoms (e.g., $^{24}$Mg and $^{28}$Si) the same technique affords energies that are overcorrelated by only a few percent. Nakatsuji and co-workers have examined a large number of molecules containing one or two first-row atoms plus hydrogens, and have found that variational 2-RDM calculations subject to the aforementioned constraints recover no more than 120\% of the correlation energy $E_c = E_{\text{FCI}} - E_{\text{HF}}$. This indicates that the $D$, $Q$, and $G$-conditions, taken together, are quite restrictive. In contrast, if the $G$-constraint alone is lifted, the resulting variational calculations yield 200–800\% of the correlation energy. This observation complements the mathematical result that the $G$-condition encompasses several other $N$-representability conditions, including $N$-representability of the 1-RDM.

It is obvious from mathematical considerations that $D_2$, $Q_2$ and $G_2$ must be positive, but one may inquire whether there is any discernable physics in these con-
ditions. In fact there is, and a physical interpretation of these constraints may be put forth on the basis of the fact that operator positivity implies that the diagonal elements of the corresponding matrix are positive in any basis. Thus $\hat{D}_2 \geq 0$, for example, insures that the geminal populations are positive. According to Eq. (2.45), the condition $(Q_2)^{ij}_{ij} \geq 0$, for $i \neq j$, implies that\(^{54}\)
\[ 2D_{ij;ij} \geq D_{i;i} + D_{j;j} - 1. \tag{2.84} \]
The left side of this inequality, including the factor of two, is the population of a normalized $2 \times 2$ Slater determinant constructed from the spin orbitals $\phi_i$ and $\phi_j$. According to the bound in Eq. (2.84), this spin-geminal population must approach unity (meaning one pair of electrons) as the populations of the two spin orbitals both approach unity. This utterly reasonable result has to be introduced as a separate condition in density matrix theory; it does not follow, for example, from the Coleman conditions on $D_1$ or the partial trace relations that connects $D_1$ and $D_2$.

An upper bound on the population $2D_{ij;ij}$ is established by the condition $G_{ij;ij} \geq 0$, for $i \neq j$. Actually, the condition $G_{ij;ij}^{\text{GP}} \geq 0$ provides a sharper bound, namely
\[ 2D_{ij;ij} \leq -|D_{ij;ij}|^2 + \min\left\{ D_{i;i}, D_{j;j} \right\}. \tag{2.85} \]
This condition is very interesting, for it implies that the maximum population of any normalized, $2 \times 2$ determinant of spin orbitals equals one electron pair. This is perhaps perplexing in light of the fact\(^{33}\) that the natural spin-geminal populations are bounded above by $[N/2]$. Moreover, eigenvalues arbitrarily close to this upper bound can actually be realized in practice from certain parametrizations of the wavefunction, most notably when $\Psi$ is an antisymmetrized product of $N/2$ copies of a single “generating” geminal\(^{33,85,86}\) [the antisymmetrized geminal power (AGP) model\(^{86-91}\)].
To reconcile these observations, consider the natural expansion of the 2-RDM,

\[ D_2(x_1, x_2; x'_1, x'_2) = \sum_I \lambda_I g_I^*(x'_1, x'_2) g_I(x_1, x_2), \] (2.86)

where \(0 \leq \lambda_I \leq [N/2]\). Suppose that we expand the natural spin geminals \(g_I\) in terms of \(2 \times 2\) determinants of orthonormal spin orbitals,

\[ g_I(x_1, x_2) = \frac{1}{\sqrt{2}} \sum_{jk} g^I_{jk} [\phi_j(x_1) \phi_k(x_2) - \phi_k(x_1) \phi_j(x_2)]. \] (2.87)

Defining \(\tilde{g}^I_{jk} = g^I_{jk} \sqrt{\lambda_I/2}\), one then has

\[ D_2(x_1, x_2; x'_1, x'_2) = \sum_I \left( \sum_{jk} \tilde{g}^I_{jk} [\phi_j(x_1) \phi_k(x_2) - \phi_k(x_1) \phi_j(x_2)] \right) \times \left( \sum_{mn} \tilde{g}^I_{mn} [\phi_m(x'_1) \phi_n(x'_2) - \phi_n(x'_1) \phi_m(x'_2)] \right)^*. \] (2.88)

This equation represents a sum of self-adjoint kernels, indexed by \(I\). Although the individual summands can each be diagonalized, in general the eigenvectors will be different for different \(I\). This is the resolution to our puzzle: in general there is no basis of \(2 \times 2\) determinants that diagonalizes \(D_2\). Stated differently, the natural spin geminals cannot generally be regarded as determinants of spin orbitals. If they could, then according to Eq. (2.85) the spin-geminal populations could not exceed unity, whereas natural spin-geminal populations in excess of unity have been documented for the beryllium atom\(^92-94\) and the carbon isoelectronic sequence.\(^95\) In the latter case, it was found that different wavefunctions consistently afforded a singlet-coupled natural spin geminal with an occupancy of 1.25–1.28. The fact that more of these supergeminals\(^95\) have not been documented probably due more to a lack of investigation than a lack of existence. However, for both the HF wavefunction and the “perfect-pairing” variety of generalized valence-bond wavefunction,\(^96-98\) the natural spin-geminal populations
are bounded above by unity. Hence 2-RDM eigenvalues greatly in excess of unity are not anticipated for finite systems (atoms and molecules).

A slew of other relationships between the 1- and 2-RDM matrix elements can be derived on the basis of \( \hat{Q}_2 \)- and \( \hat{G}_2 \)-positivity. For example, if one takes any \( M \) diagonal elements \( G_{i_1,j_1;i_1,j_1}, G_{i_2,j_2;i_2,j_2}, \ldots, G_{i_M,j_M;i_M,j_M} \), and from them forms a determinant whose \((k,l)\)th element is \( G_{i_k,j_k;i_l,j_l} \), then this determinant must be positive. The reader is referred to Garrod and Percus for elaboration. In addition, Mazziotti and Erdahl have shown that the \( G \)-condition enforces certain uncertainty relations on the Hamiltonian and other two-particle operators.

### 2.4 Intracules and extracules

Mathematically, \( D_2(x_1, x_2; x'_1, x'_2) \) suffices to determine physical observables, but from a conceptual point of view even the two-electron density \( \rho_2(x_1, x_2) \) is difficult to visualize, as it depends upon two space/spin vector coordinates. The decomposition presented below will transform the information content of \( \rho_2 \) into a pair of one-coordinate functions, which are readily visualized but still convey some information about electron correlation.

#### 2.4.1 Definitions

For a pair of electrons with spatial coordinates \( r_1 \) and \( r_2 \), let us first introduce the *intracule* (relative) coordinate \( u \equiv r_1 - r_2 \) and the *extracule* (center-of-mass) coordinate \( R \equiv \frac{1}{2}(r_1 + r_2) \). Next we note that, provided the operators \( \hat{h} \) and \( \hat{g} \) introduced in the last section are independent of spin, the expectation value of the observable \( \hat{W} \) in
Eq. (2.1) can be calculated according to $\langle \hat{W} \rangle = \text{tr}(\hat{W} \hat{D}_2^o)$, where

$$D_2^o(r_1, r_2; r_1', r_2') \overset{\text{def}}{=} \int d\xi_1 d\xi_2 D_2(r_1, \xi_1, r_2, \xi_2; r_1', \xi_1', r_2', \xi_2')$$  \hspace{1cm} (2.89)

is the spin-traced 2-RDM, also known as the two-electron charge density matrix. The kernel $D_2^o$ can equally well be expressed as a function of intracule and extracule coordinates, $D_2^o = D_2^o(u, R; u', R')$. Thus we can define an intracule density $I(u)$

$$I(u) \overset{\text{def}}{=} \int dR D_2^o(u, R; u, R) = \int dr_1 dr_2 \delta(u - (r_1 - r_2)) \rho_2^o(r_1, r_2)$$  \hspace{1cm} (2.90)

and an extracule density

$$X(R) \overset{\text{def}}{=} \int du D_2^o(u, R; u, R) = \int dr_1 dr_2 \delta(R - \frac{1}{2}(r_1 + r_2)) \rho_2^o(r_1, r_2).$$  \hspace{1cm} (2.91)

Alternatively, in terms of the electronic wavefunction

$$I(u) = \sum_{j<k}^N \int dx_1 \cdots dx_N \Psi^*(x_1, \ldots, x_N) \delta(u - r_{jk}) \Psi(x_1, \ldots, x_N),$$  \hspace{1cm} (2.92)

with $r_{jk} \equiv r_j - r_k$. Replacing $\delta(u - r_{jk})$ in this equation with $\delta(R + \frac{1}{2}(r_j + r_k))$ defines $X(R)$ in terms of $\Psi$. The terms “intracule” and “extracule” seem to make their debut in Eddington’s textbook, but it was Coleman who introduced and popularized the terms “intracule density” and “extracule density” in electronic structure theory. This terminology is ubiquitous but unfortunate; in a less esoteric vernacular, $I(u)$ [or more likely its spherical average $I(u)$, defined below] is known simply as the pair distribution function for electrons.

Returning to the observable $\hat{W}$, the most important case for electronic structure theory is when $\hat{g}(x_1, x_2) = |r_1 - r_2|^{-1} = r_{12}^{-1}$. More generally, any time $\hat{g} = g(r_{12}) = \cdots$
$g(u)$ is a multiplicative operator that depends only on $r_{12} = |\mathbf{r}_1 - \mathbf{r}_2|$, one needs only the spherical average of $I(u)$ in order to obtain the two-electron part of $\langle \hat{W} \rangle$. Specifically,

$$\langle \hat{W} \rangle = \text{tr}(\hat{h} \hat{D}_1) + 4\pi \text{ tr}\left(g(u) I(u)\right), \quad (2.93)$$

where

$$I(u) = \frac{1}{4\pi} \int \sin \theta \, d\theta \, d\varphi \, I(u, \theta, \varphi) \quad (2.94)$$

is the spherically-averaged intracule density. The spherical average $X(R)$ of $X(\mathbf{R})$ is similarly defined. The usual Born-Oppenheimer electronic energy is determined by the 1-RDM and $I(u)$ alone.

$I(u)$ provides a quantitative account of electron correlation, but the densities $I(u)$ and $X(R)$ are perhaps even more useful for qualitative or interpretive purposes, since their content is easy to understand and to plot. $I(u)$ monitors how far apart the electron pairs are, while $X(R)$ indicates where the pairs reside in space. The intracule density especially has been widely utilized to understand the shell structure of atoms, but efforts to understand and describe electron correlation in molecules in terms of $I(u)$ and $X(R)$ have been undertaken by only a few research groups. Apart from a strong sociological bias in favor of one-electron models, the main factor that impedes wider use of these quantities is that the two-electron density matrix, which is necessary in order to calculate $I(u)$ and $X(R)$, is not available to casual users of electronic structure codes. Recently, developmental versions of a program to compute intracule densities have been implemented for use with the Gaussian and Q-chem programs, and in the near future it is anticipated that $I(u)$ especially will be more widely exploited as an interpretive tool. In the present work, $I(u)$ and $X(R)$ are used in Chapter 5 to analyze certain reconstructed 2-RDMs.
Of particular significance are the values $I(0)$ and $X(0)$. The former is appropriately termed the *electron-electron coalescence density*,\textsuperscript{121,122} to which only electron pairs with opposite spins can contribute. The coalescence density is the basis of an important (and exceedingly stringent\textsuperscript{111,123–125}) cusp condition,\textsuperscript{123,126}

$$I(0) = \frac{dI}{du}_{u=0},$$

(2.95)

which amounts to a restatement, in density-matrix language, of Kato’s cusp condition\textsuperscript{127} on the electronic wavefunction. The value $I(0)$ also enters as a parameter in relativistic and radiative corrections to the electronic energy. In contrast to $I(0)$, the *electron-electron counterbalance density*, defined as the quantity $X(0)$, depends upon the choice of coordinate origin. For an appropriately chosen coordinate system, however, $X(0)$ can be used to partition electron correlation effects into dynamical versus non-dynamical (or Fermi versus Coulomb) contributions.\textsuperscript{112,128,129}

### 2.4.2 Computational formulas in a Slater-type basis

In Chapter 5 we will examine some atomic intracule and extracule densities. The purpose there is to determine the physical content of certain reconstructive approximations for the 2-RDM, taking as input a 1-RDM derived from an accurate atomic wavefunction, which has the form of a determinantal expansion in a basis of Slater-type orbitals (STOs). We therefore require formulas to evaluate the intracule and extracule densities in such a basis. These are derived in this subsection and the next.

Actually, Koga and Matsuyama\textsuperscript{7,130} have published analytic formulas for the angular parts of the required integrals. For our purposes, however, we require formulas that are expressed directly in terms of 2-RDM matrix elements rather than CI coefficients, so that we may calculate intracule and extracule densities corresponding to
reconstructed 2-RDMs, for which an associated wavefunction (if one exists at all) is unknown. In the case that a wavefunction is known, what follows is formally equivalent to the Koga-Matsuyama formulas, though we have pushed the analytic integration somewhat further, performing part of the radial integration analytically and leaving only one relatively well-behaved spherical Bessel transform that must be evaluated numerically. One ancillary result of this analysis is a novel explanation for an approximate relationship between the intracule and extracule densities that has been noted previously\textsuperscript{4,5} based on computational evidence,\textsuperscript{7-10} but whose origins were a mystery prior to the author’s work.\textsuperscript{11}

We begin from an orbital expansion of the two-electron charge density

$$\rho_2(r_1, r_2) = \sum_{j,k,p,q=1}^\nu \rho_{jk;pq}^0 \chi_j^*(r_1) \chi_k^*(r_2) \chi_p(r_1) \chi_q(r_2)$$

(2.96)

with coefficients [compare Eqs. (2.57)-(2.59)]

$$\rho_{jk;pq}^0 = D_{j;k;p}^\alpha + D_{j;k;p}^\beta + D_{j;k;p}^\alpha + D_{j;k;p}^\beta$$

$$= D_{jk;qp}^\alpha + D_{jk;qp}^\beta + D_{jk;qp}^\alpha + D_{jk;qp}^\beta$$

(2.97)

The tensor $\rho_2^0$ whose elements are defined above is self-adjoint but not antisymmetric.

In developing the formalism that follows, we assume a spin-compensated state, defined as one for which $\rho_1^\alpha \equiv \rho_1^\beta$. (In particular, this includes any singlet state.) In this case $\rho_2^{\alpha\alpha} = \rho_2^{\beta\beta}$.

Employing the orbital expansion of $\rho_2^0$ along with the identity

$$\delta(r) = \frac{1}{8\pi^3} \int ds \ e^{r \cdot s},$$

(2.98)
the intracule density defined in Eq. (2.90) may be rewritten as

\[ I(u) = \frac{1}{8\pi^3} \sum_{jqp}^\nu \rho_{jkpq}^2 \int ds \, e^{iu \cdot s} F^*_{pq}(s) F_{k,q}(s) , \]  

where

\[ F_{jk}(s) \overset{\text{def}}{=} \int dr \, e^{ir \cdot s} \chi_j^*(r) \chi_k(r) . \]  

To simplify \( F_{jk} \), assume that the angular parts of the orbitals are pure spherical harmonics,

\[ \chi_k(r) = R_k(r) \, Y_{l,k,m_k}(\Omega_r) , \]  

with the notation \( r \equiv (r, \Omega_r) \equiv (r, \theta_r, \varphi_r) \). The spherical harmonics

\[ Y_{l,m}(\theta, \varphi) = \Theta_{l,m}(\theta) \left( \frac{1}{\sqrt{2\pi}} e^{im\varphi} \right) \]  

are expressed in terms associated Legendre functions \( \Theta_{l,m} \), normalized to unity, and utilizing Dirac’s phase convention,\(^\text{131}\) wherein \( \Theta_{l,-m} = (-1)^m \Theta_{l,m} \).

Ultimately, we intend to apply our intracule and extracule formulas to atomic calculations, where the indicated form of orbital is used almost exclusively. For molecules, this type of orbital [with \( R_k(r) \) a Gaussian function] is also common but not to the exclusion of Cartesian representations of the angular functions. Requisite formulas to compute \( I(u) \) or \( X(R) \) in a basis of Cartesian Gaussian functions can be found in several places,\(^\text{101,102,118,120,132}\) [See especially the rather sophisticated algorithm proffered by Cioslowski and Liu,\(^\text{118}\) which is currently the preferred way to calculate these quantities in a Gaussian basis set.]

In what follows we will make frequent use of the expansion of a plane wave in terms of spherical waves,\(^\text{133}\)

\[ e^{\pm i \cdot s \cdot r} = 4\pi \sum_{l=0}^\infty \sum_{m=-l}^l (\pm i)^l j_l(s r) Y^*_{l,m}(\Omega_s) Y_{l,m}(\Omega_r) . \]  


The spherical Bessel functions $j_l$ in this expansion are related to the ordinary Bessel functions $J_{l+1/2}$ according to

$$j_l(x) \overset{\text{def}}{=} \sqrt{\frac{\pi}{2x}} J_{l+1/2}(x) = (-x)^l \left( \frac{1}{x} \frac{d}{dx} \right)^l \frac{\sin x}{x}. \quad (2.104)$$

The angular integrals in $I(u)$ will be evaluated in terms of the so-called Gaunt coefficients,\(^{131,134}\)

$$Y_{l_1,m_1,l_2,m_2}^{(k)} \overset{\text{def}}{=} \sqrt{\frac{4\pi}{2k+1}} \int d\Omega \, Y_{k,m_1-m_2}(\Omega) \, Y_{l_1,m_1}^*(\Omega) \, Y_{l_2,m_2}(\Omega) \quad (2.105)$$

The pre-integral factor in Eq. (2.105) is included to match the integrals tabulated by Condon and Shortley.\(^{131}\) The Gaunt coefficients, which are closely related to the Clebsch-Gordan coefficients and the Wigner 3$j$-symbols,\(^{135,136}\) can be evaluated analytically,\(^{134}\) but the general formula is rather complicated and efficient computational implementations of this formula are the subject of ongoing research.\(^{137,138}\) The values tabulated by Condon and Shortley\(^*\) include all nonzero coefficients through $l_1 = 3 = l_2$, which suffices for our purposes. From the general formula for the Gaunt coefficients, it follows that $Y_{l_1,m_1,l_2,m_2}^{(k)} = 0$ unless $k + l_1 + l_2$ is even (the parity condition) and also $|l_1 - l_2| \leq k \leq l_1 + l_2$ (the triangle rule). Also, since $\Theta_{l,m} = (-1)^m \Theta_{l,-m}$, one has

$$Y_{l_1,m_1,l_2,m_2}^{(k)} = (-1)^{m_1-m_2} Y_{l_2,m_2,l_1,m_1}^{(k)} \quad (2.106)$$

Upon substituting the spherical wave expansion for $e^{ir \cdot s}$, the angular integrals appearing in $F_{j,k}(s)$ can by evaluated analytically in terms of the Gaunt coefficients.

\(^*\)Reference 131, Table 16, pg. 178–179.
The result is

\[ F_{j,k}(s) = \sqrt{\frac{4\pi}{i}} \sum_{l=[j-k]}^{l_j+l_k} \psi^{(l)} Y_{l_j, m_j, l_k, m_k} Y_{l, m_j - m_k}^* (\Omega_s) \mathcal{H}_l(R_j R_k)(s), \]  

(2.107)
in which \( \mathcal{H}_l \) denotes the spherical Bessel transform (Hankel transform) of order \( l \),

\[ \mathcal{H}_l[f](s) \overset{\text{def}}{=} \int_0^\infty dr r^2 j_l(sr) f(r). \]  

(2.108)

Next, substitute the above expression for \( F_{j,k} \) in \( I(u) \) into Eq. (2.99) for \( I(u) \), then expand the plane wave \( e^{i\mathbf{u}\cdot\mathbf{s}} \) in the resulting formula using Eq. (2.103) once again. To simplify the resulting expression, observe that \( \rho_{jk;pq}^2 = 0 \) unless \( m_j + m_p = m_k + m_q \), since the bra- and ket-sides of the density operator must have the same \( M_L \) quantum number for an \( \hat{S}_z \) eigenstate. Performing the integration over \( \Omega_s \), one finds that only the \( m = 0 \) spherical harmonics survive in the spherical wave expansion of \( e^{i\mathbf{u}\cdot\mathbf{s}} \). These are functions \( Y_{l,0}(\theta, \varphi) = \Omega_{l,0}(\theta)/\sqrt{2\pi} \), where \( \Omega_{l,0}(\theta) \propto P_l(\cos \theta) \), the \( l \)th Legendre polynomial, with

\[ \int_0^\pi d\theta \sin \theta \left[ \Omega_{l,0}(\theta) \right]^2 = 1. \]  

(2.109)
The final expression for the intracule density is then

\[ I(u, \theta) = \frac{1}{\pi^2 \sqrt{2}} \sum_{jkpq} \rho_{jk;pq}^2 \delta_{m_j + m_p, m_k + m_q} \sum_{l_1, l_2} \psi^{(l_1)} \delta_l^{l_1} (-1)^{l_1 + m_k - m_q} \]

\[ \times \sqrt{(2l + 1)(2l_1 + 1)(2l_2 + 1)} \]

\[ \times Y_{l_1, m_j - m_p, l_2, m_q - m_k}^{(l_1)} Y_{l_1, m_j, l_p, m_p}^{(l_2)} Y_{l_2, m_k, l_q, m_q}^{(l_2)} \]

\[ \times P_l(\cos \theta) \left\{ \mathcal{H}_{l_1}[R_j R_p] \mathcal{H}_{l_2}[R_k R_q]\right\}(u). \]  

(2.110)

All of the functional dependence in this formula is contained in the last line, and we use \( l_k \), for example, to denote the angular momentum quantum number associated with the orbital \( \chi_k \) [see Eqs. (2.101) and (2.111)].
Some remarks concerning this expression are in order. First, the summations over non-negative integers $l$, $l_1$, and $l_2$ in Eq. (2.110) are truncated by the triangle rule so that, for example, $|l_j - l_p| \leq l_1 \leq l_j + l_p$. Furthermore, because $\gamma^{(l)}_{l_1,j,m,j;l_2,m,q-m_k} = 0$ except when $l + l_1 + l_2$, it follows that $i^{l+l_1+l_2} = \pm 1$. (The intracule density is real-valued.) Finally, the fact that atomic intracule densities do not depend upon the azimuthal angle $\varphi$ is easily understood using a coordinate system in which one electron is fixed at the origin and the nucleus lies along the $z$ axis. Then $I(u)$, the conditional probability distribution for a second electron, must possess cylindrical symmetry about the $z$ axis, and is thus independent of $\varphi$. $I(u)$ cannot, however, be independent of $\theta$, as this would lead to the absurd conclusion that the presence of the nucleus does not affect the electron distribution.

When the radial functions $R_k(r)$ are Gaussian- or Slater-type functions, the integral transforms $\mathcal{H}_{l_1} [R_j^* R_p](s)$ and $\mathcal{H}_{l_2} [R_k^* R_q](s)$ in Eq. (2.110) can be evaluated analytically. Let us assume henceforth that the radial functions are linear combinations of primitive Slater-type functions,

$$R_k(r) = \sum_j C_{kj} r^{n_j+l_k} e^{-a_j r}.$$  \hfill (2.111)

This form of radial function is superior to a Gaussian on physical grounds, and is therefore common in atomic calculations, where there are no three- or four-center integrals to evaluate. Here $l_k = 0, 1, 2, \ldots$ is an (integer) orbital angular momentum quantum number, while the exponents $a_j$ and pseudo-shell parameters $n_j$ are positive, real numbers.* The coefficients $C_{kj}$ are constrained so that the functions $R_k$ are

---

*As a matter of semantics, the term “Slater orbital” is sometimes meant to imply that $n_j$ is an integer, namely, the principal quantum number for the $j$th orbital. It is for this reason that the functions in Eq. (2.111) are called Slater-type orbitals. Typically the parameters $a_j$ and $n_j$ in a STO are variationally-optimized for a particular atom.
orthonormal.

To evaluate the requisite integral transforms, let us first write down the spherical Bessel functions explicitly. For even orders they are

\[ j_{2n}(x) = \sum_{k=0}^{n} B_{2k+1}^{(2n)} \left( \frac{\sin x}{x^{2k+1}} \right) + \sum_{k=1}^{n} B_{2k}^{(2n)} \left( \frac{\cos x}{x^{2k}} \right), \tag{2.112} \]

for \( n = 0, 1, 2, \ldots \), while the odd-order functions are

\[ j_{2n+1}(x) = \sum_{k=0}^{n} B_{2k+1}^{(2n+1)} \left( \frac{\cos x}{x^{2k+1}} \right) + \sum_{k=1}^{n+1} B_{2k}^{(2n+1)} \left( \frac{\sin x}{x^{2k}} \right). \tag{2.113} \]

The constants \( B_{k}^{(j)} \) are easily obtained from the definition, Eq. (2.104). Since the product of two Slater-type functions is also a Slater-type function, we content ourselves to evaluate the transform of a generic Slater-type function \( r^n \exp(-ar) \), with \( a \geq 0 \).

From Eq. (2.110) and the triangle rule, it follows that we need calculate the functions \( \mathcal{H}_{l_1}[R_j R_p] \) only for \( l_1 \leq l_j + l_p \). The kernel \( j_{l_1}(sr) \) of the transform \( \mathcal{H}_{l_1} \) contains powers of \( r \) in the denominator up to \( r^{l_1+1} \). However, each term in the function \( r^2 R_j(r) R_p(r) \) contains an equal or larger power of \( r \), so in our generic transform \( \mathcal{H}_{l_1}[r^n \exp(-ar)](s) \) we need only consider \( n \geq 0 \). A similar analysis applied to \( \mathcal{H}_{l_2}[R_k R_q](s) \) leads to the same conclusion. The transform is evaluated using standard integrals.\(^{139}\) For even \( l \) (put \( l = 2p \)), the result is

\[ \mathcal{H}_{2p}[r^n e^{-ar}](s) = \sum_{k=0}^{p} B_{2k+1}^{(2p)} \left( \frac{(n+1-2k)! a^{n+2-2k}}{s^{2k+1}(a^2 + s^2)^{n+2-2k}} \right) \sum_{j=0}^{[(n+1-2k)/2]} (-1)^j \binom{n+2-2k}{2j+1} \frac{s^{2j+1}}{a^{2j+1}} + \sum_{k=1}^{p} B_{2k}^{(2p)} \left( \frac{(n+2-2k)! a^{n+3-2k}}{s^{2k}(a^2 + s^2)^{n+3-2k}} \right) \sum_{j=0}^{[(n+3-2k)/2]} (-1)^j \binom{n+3-2k}{2j} \frac{s^{2j}}{a^{2j}}, \tag{2.114} \]
72

where \([q]\) represents the greatest integer less than or equal to \(q\). For \(l = 2p + 1\), one obtains

\[
\mathcal{H}_{2p+1}[p^n e^{-ar}] \quad (2.115)
\]

\[
= \sum_{k=0}^{p} B_{2k+1}^{(2p+1)} \left( \frac{(n + 1 - 2k)!}{s^{2k+1}(a^2 + s^2)^{n+2-2k}} \right) \sum_{j=0}^{\left\lfloor (n+2-2k)/2 \right\rfloor} (-1)^j \left( \frac{n + 2 - 2k}{2j} \right) \left( \frac{s}{a} \right)^{2j} \\
+ \sum_{k=1}^{p+1} B_{2k}^{(2p+1)} \left( \frac{(n + 2 - 2k)!}{s^{2k}(a^2 + s^2)^{n+3-2k}} \right) \sum_{j=0}^{\left\lfloor (n+2-2k)/2 \right\rfloor} (-1)^j \left( \frac{n + 3 - 2k}{2j + 1} \right) \left( \frac{s}{a} \right)^{2j+1}
\]

That the Hankel transform \(\mathcal{H}_l\) of a Slater-type function can be evaluated analytically does not seem to have been recognized by Koga, Matsuyama, and co-workers in their extensive studies of Hartree-Fock atomic intracule densities.\(^{7-10}\) Wang and Smith,\(^{140}\) however, have given alternative formulas to our Eqs. (2.114) and (2.115) in terms of some recurrence relations. In the calculations presented in Chapter 5, the outermost Hankel transform in Eq. (2.110) is evaluated numerically, using Talman’s algorithm.\(^{141}\)

### 2.4.3 Approximate intracule-extracule isomorphism

The preceding simplifications for \(I(u)\) can be applied equally well to \(X(R)\), starting from the definition of \(X(R)\) in terms of the two-electron charge density, which is given in Eq. (2.91). At the risk of seeming repetitive we present the final, simplified form of \(X(R)\) so that we may compare it to the final \(I(u)\) formula derived above:

\[
X(R, \theta) = \frac{8}{\pi^2 \sqrt{2}} \sum_{j,k,p,q} \rho_{j,k,p,q}^2 \delta_{m_j+m_p,m_k+m_q} \sum_{l_1,l_2} \xi^{l_1+l_2} (-1)^{l_1+l_2+m_k-m_q} \\
\times \sqrt{(2l_1+1)(2l_2+1)} \prod_{k=1}^{\nu} \mathcal{Y}_{l_1,m_j-m_p,l_2,m_q-m_k} \mathcal{Y}_{l_1,m_j,l_2,m_p} \mathcal{Y}_{l_2,m_k,m_q} \\
\times P_l(\cos \theta) \left\{ \mathcal{H}_l \left[ (R_j^*(\rho)) \circ 2 \right] \mathcal{H}_l \left[ (R_k^*(\rho)) \circ 2 \right] \right\}(R) \quad (2.116)
\]
All of the functional dependence in this formula is gathered together in the last line. The notation \( f \circ \omega \), where \( f = f(x) \) and \( \omega \) is a constant, denotes the convolution of \( f \) with the function \( \omega x \),

\[
[f \circ \omega](x) \equiv f(\omega x) .
\] (2.117)

This convolution arises in Eq. (2.116) due to the factor of 1/2 in the \( \delta \)-function integrand of Eq. (2.91).

By changing integration variables one may easily show that for \( \omega \neq 0 \),

\[
\omega^3 \mathcal{H}_i [f](\omega s) = \mathcal{H}_i [f \circ \frac{1}{\omega}](s) ,
\] (2.118)

or in function notation \( \omega^3(\mathcal{H}_i[f] \circ \omega) = \mathcal{H}_i[f \circ \frac{1}{\omega}] \). Consequently

\[
\mathcal{H}_i \left\{ \mathcal{H}_{l_1} [R_{j}^* R_{p}] \mathcal{H}_{l_2} [R_{k}^* R_{q}] \right\} (2R)
\]

\[
= 8 \mathcal{H}_i \left\{ \mathcal{H}_{l_1} [(R_{j}^* R_{p}) \circ 2] \mathcal{H}_{l_2} [(R_{k}^* R_{q}) \circ 2] \right\} (R)
\] (2.119)

is an exact symmetry. The reader will recognize that the expression on the right appears in \( X(R) \), while the expression on the left appears in \( I(2R) \). The expressions for \( X(R) \) and \( I(2R) \) therefore differ only by the factor of \((-1)^{l_2}\) that is present in the former. Of course \( l_2 \) is a summation index in Eqs. (2.110) and (2.116) (it runs from \( |l_k - l_q| \) to \( l_k + l_q \)), but if this index were always even or always odd then we would have \( X(R) \propto I(2R) \). In reality this proportionality is only approximate, as we now show.

Consider the radial intracule and extracule densities \( I(u) \) and \( X(R) \). Since

\[
\int_{0}^{\pi} d\theta \sin \theta \Theta_{l,0}(\theta) = \delta_{l,0} \sqrt{2} ,
\] (2.120)

integration of \( I(u) \) or \( X(R) \) over \( \theta \) eliminates the sum over \( l \) that appears in both Eq. (2.110) and Eq. (2.116). Of the 2-CD expansion coefficients \( \rho_{jk;pq}^0 \) appearing...
in these equations, by far the largest will be the ones that are nonzero in the HF approximation, namely, those elements $\rho_{jk;jk}^0$ for which $\chi_j$ and $\chi_k$ are occupied in the HF determinant. If $k = q$ in $\rho_{jk;pq}^0$ then $l_k = l_q$, which implies that $Y_{l_k,m_k;l_q,m_q}^{(l_2)}(l_2) = 0$ unless $l_2$ is even. The index $l_2$ being strictly even is precisely what is needed to establish a proportionality between the intracule and extracule densities. Thus, if one neglects all matrix elements of $\rho_2^0$ except for the ones that are nonzero in the HF approximation, then using the identity in Eq. (2.119) we obtain the approximate relationship

$$I(2R) \approx X(R).$$

(2.121)

This ought to be a reasonable approximation even for correlated wavefunctions, because the HF-type contributions will still dominate $\rho_2^0$. In fact, we may arrive at the approximate relationship in Eq. (2.121) by retaining all diagonal matrix elements $\rho_{jk;jk}^0$, regardless of whether $\chi_j$ or $\chi_k$ is occupied in the HF determinant. This is because the two-electron density is dominated by its diagonal elements (a property inherited from $D_2$), though the magnitudes of the off-diagonal elements (in the canonical HF basis at least) are likely to be larger at correlated levels of theory than they are at the HF level.

The foregoing arguments provide a sound, mathematical justification for the approximate relation that Koga\textsuperscript{4,5} was able to surmise by examining spherically-averaged Hartree-Fock intracule and extracule densities for each atom from He to Xe.\textsuperscript{7-10} Neglect of the off-diagonal elements of $\rho_2^0$ could be used in similar fashion to establish a proportionality between $I(2R)$ and $X(R)$, but since these quantities contain an additional sum (over $l$) relative to their spherically-averaged counterparts $I(u)$ and
$X(R)$, the necessary approximation neglects even more terms. One therefore expects $I(2R) \approx X(R)$ to be a worse approximation than $I(2R) \approx X(R)$.

### 2.5 Reduced density matrix cumulants

The intracule and extracule densities discussed in the previous section represent a certain decomposition of the information content of the 2-RDM; another such decomposition is based on the notion of RDM cumulants. The cumulant decomposition of $D_2$ naturally partitions the energy or other expectation value into separate contributions from exchange and from correlation, providing a universal definition for these concepts that does not depend upon any single-determinant reference state. The cumulant decompositions of $D_2$, $D_3$, and $D_4$ are also used extensively in Chapters 3 and 4.

#### 2.5.1 Additive versus multiplicative separability

Although the RDMs provide a compact and physically appealing description of electronic structure, this description is unsatisfactory in at least one respect, namely, expectation values calculated from RDMs are not manifestly extensive (additively separable), so any approximate calculation of the RDMs will almost certainly violate extensivity to some extent. This basic flaw ultimately arises because the RDMs, like the electronic wavefunction, are multiplicatively separable rather than additively separable.

To illustrate this point, consider a composite system consisting of two non-interacting subsystems, one with $p$ electrons and the other with $q = N - p$ electrons. The composite system’s Hamiltonian is therefore additively separable in the two sub-
systems:

\[ \hat{H} = \hat{H}_A + \hat{H}_B. \]  

(2.122)

In realistic situations, this is obviously a limiting form of \( \hat{H} \) only; however, within a finite spin-orbital basis set or as a Fock-space operator, such a Hamiltonian is obtained whenever all of the matrix elements connecting subsystem A to subsystem B are negligible. In what follows, let \( B_A \) and \( B_B \) denote disjoint sets of orthonormal spin orbitals, each associated with only one subsystem. By hypothesis, \( \hat{H} \) has no matrix elements connecting the functions in \( B_A \) to those in \( B_B \).

If \( \Psi_A \) and \( \Psi_B \) are eigenfunctions of \( \hat{H}_A \) and \( \hat{H}_B \), each normalized to unity, then the wavefunction for the composite system is

\[ \Psi_{(1,\ldots,N)} = \frac{1}{\sqrt{N!}} \prod_{j=1}^N \langle \Psi_A(1,\ldots,p) | \Psi_B(p+1,\ldots,N) \rangle. \]  

(2.123)

We say that \( \Psi \) is \textit{multiplicatively separable} in the two subsystems, recognizing that in quantum mechanics this separation must include an overall symmetrization or antisymmetrization (according to whether the particles are bosons or fermions) that renders all coordinates equivalent. The separation of the wavefunction in Eq. (2.123) is equivalent, in a necessary and sufficient sense, to the separation (block structure) of the Hamiltonian in Eq. (2.122).\textsuperscript{146,153–155}

Because subsystems A and B do not interact, it must be that \( \Psi_A \) consists of a determinantal expansion in functions taken solely from the set \( B_A \), and similarly \( \Psi_B \) uses only those spin orbitals in \( B_B \). It follows that \( \Psi_A \) and \( \Psi_B \) are strongly orthogonal.\textsuperscript{156} Two functions \( f(x_1,\ldots,x_p) \) and \( g(y_1,\ldots,y_q) \) are said to be \textit{strongly orthogonal} if

\[ \int dz \, f^*(x_1,\ldots,x_{p-1},z) \, g(y_1,\ldots,y_{q-1},z) = 0. \]  

(2.124)
(The functions of interest, $\Psi_A$ and $\Psi_B$, are antisymmetric, so it does not matter which coordinates are chosen for the dummy integration variable z.)

Consider the RDMs that are obtained from the separable wavefunction $\Psi$ in Eq. (2.123).\footnote{The discussion from Eq. (2.125) to Eq. (2.128) is based upon unpublished notes by J. E. Harriman (April, 2003).} It is immediately obvious, by virtue of the strong orthogonality of $\Psi_A$ and $\Psi_B$, that $\langle \Psi | \hat{a}_i^\dagger \hat{a}_j | \Psi \rangle = 0$ unless $\phi_i$ and $\phi_j$ are associated with the same subsystem. Thus the 1-RDM separates into subsystem 1-RDMs,

$$D_1(x; x') = D_A^1(x; x') + D_B^1(x; x').$$

(2.125)

The case $p = 1$ is the unique example for which $D_p$ is additively separable. This is equivalent to the statement (see below) that $D_1$ equals its own cumulant.

To obtain $D_2$, we need to evaluate matrix elements $\langle \Psi_A \Psi_B | \hat{a}_i^\dagger \hat{a}_j \hat{a}_l | \Psi_A \Psi_B \rangle$. (Recall that $| \Psi_A \Psi_B \rangle$, without any direct product symbol, implies an antisymmetrized product state.) For reasons that will become clear, let us introduce the quantity

$$\Delta_{ijkl} \equiv D_{ijkl} - \frac{1}{2} (D_{ik} D_{jl} - D_{il} D_{jk}).$$

(2.126)

The interesting scenario is when two of the four indices $i, j, k,$ and $l$ refer to subsystem A and the other two refer to subsystem B. In the case that $\phi_i, \phi_j \in B$, and $\phi_k, \phi_l \in B$, it is easy to show that $D_{ijkl} = 0$. More interesting is the case when $\phi_i, \phi_k \in B_A$ and $\phi_j, \phi_l \in B_B$. It is not difficult to convince oneself that in this case $D_{ijkl}^g$ is generally nonzero, hence $D_2$ fails to be additively separable, but what about $\Delta_{ijkl}$?

We note first of all that $D_{il} = 0$, since $i$ and $l$ refer to different subsystems. Hence

$$\Delta_{ijkl} = D_{ijkl} = D_{ik} D_{j,l} - \frac{1}{2} D_{ik} D_{j,l}. \text{ Noting furthermore that } i \neq l \text{ and } j \neq k, \text{ some operator}$$
algebra affords

\[ D_{ij;kl} = \frac{1}{2} \langle \Psi_A | \hat{a}_i^\dagger \hat{a}_k \hat{a}_j^\dagger \hat{a}_l | \Psi_A \rangle \]

\[ = \frac{1}{2} \langle \Psi_A | \hat{a}_i^\dagger | \Psi_A \rangle \langle \Psi_B | \hat{a}_j^\dagger \hat{a}_l | \Psi_B \rangle , \]

which we may identify as a product of two 1-RDM elements, one from each subsystem.

It follows that \( \Delta_{ij;kl} = 0 \) for the case in question. Since \( \Delta_{ij;kl} \) as defined in Eq. (2.126) is antisymmetric, it follows that this quantity is zero unless all four indices refer to the same subsystem. In \( \Delta_{ij;kl} \) we have therefore identified the two-particle quantity that, unlike \( D_{ij;kl} \), is additively separable in the two non-interacting subsystems. With obvious notation, and in analogy to Eq. (2.125), we express this as

\[ \Delta_2(x_1; x_2; x'_1, x'_2) = \Delta_2^A(x_1, x_2; x'_1, x'_2) + \Delta_2^B(x_1, x_2; x'_1, x'_2) . \]

In the next subsection we will identify \( \Delta_2 \) as the cumulant of the 2-RDM, which is basically obvious from the foregoing discussion.

Notice that \( D_2 = \Delta_2 + D_1 \wedge D_1 \). Thus, expectation values for the separable \( A + B \) system may be computed using either \( D_2 \) alone, or \( D_1 \) together with \( \Delta_2 \).

From the standpoint of exact quantum mechanics, either method yields exactly the same expectation value; in particular, either method respects the extensivity of the electronic energy. If \( D_2 \) is calculated by means of approximate quantum mechanics, however, one cannot generally expect that extensivity will be preserved, since exchange terms mingle the coordinates on different subsystems, and exact cancellation (unless built in from the start) is too much to ask. Methods that do build in this property are usually said to be size-consistent\textsuperscript{157–159} though extensive is arguably a better term. (The term size-extensive is also encountered sometimes, but the author considers this phrase to be redundant, and it is never used in this work.)
In careful usage, extensivity is actually a more general concept than size consistency. Extensivity implies a complete absence of unconnected terms in one’s working equations,\textsuperscript{159} while size-consistency indicates only that the energy is additive by non-interacting subsystems, $E(A+B) = E(A) + E(B)$, which is but one consequence of extensivity. An additional consequence of extensivity, or rather the lack thereof, is that any approximate electronic structure method that does not preserve this property will tend toward zero correlation energy per particle in an infinite system.\textsuperscript{159} Hence the conventional wisdom is that the use of manifestly extensive methods (coupled-cluster theory being the canonical example) is crucial for “large” systems. It is not entirely clear how large one can go before this becomes a problem, though the effective range of the spinless 1-RDM provides one indication. As discussed in Subsection 5.2.3, for unbranched alkanes the effective range $|\mathbf{r} - \mathbf{r}'|$ over which $D^i_0(\mathbf{r};\mathbf{r}')$ is nonzero is about 15–21 carbon atoms,\textsuperscript{160} and we may judge that for larger systems extensivity violations may have important consequences. For systems more compact than $C_n\text{H}_{2n+2}$, extensivity may be even more important.

In the present context, the way to insure extensivity or size-consistency is to reformulate density matrix theory so that the RDM cumulants are the basic variables, rather than the RDMs themselves. One can always recover the RDMs from the cumulants; however, the additively-separable cumulants satisfy “connected” equations that do not admit the possibility of mixing two non-interacting subsystems. The quantity $\Delta_{ijkl}$, by its very formulation, vanishes automatically whenever the Hamiltonian matrix obtains the block structure characteristic of non-interacting subsystems, so there is no need to rely upon cancellation in order to achieve extensivity.
2.5.2 Cumulant formalism

Having motivated our development of the RDM cumulants using $D_2$ as an example, we now proceed with the general treatment for $D_p$. To develop the theory of cumulants for non-commutative variables (the creation and annihilation operators), we introduce anticommutative probe variables\textsuperscript{17,149} $f(x)$ and $f^\dagger(x)$,

\begin{align}
[f(x), f(x')]_+ &= 0 \quad (2.129a) \\
[f(x), f^\dagger(x')]_+ &= 0 . \quad (2.129b)
\end{align}

The fields $f$ and $f^\dagger$ are “probes” in the sense that they serve as dummy placeholders in the formal differentiations to follow, but do not appear in the final expressions for the cumulants, which are obtained formally in the limit that $f, f^\dagger \to 0$.

The RDMs may be obtained as derivatives of a generating functional $G[f, f^\dagger]$. Letting $\hat{\psi}(x)$ and $\hat{\psi}^\dagger(x)$ represent the usual field operators, this functional is\textsuperscript{17}

\begin{equation}
G[f, f^\dagger] = \langle \Psi | \hat{\mathcal{N}} \exp \left( \int dx \ [f(x) \hat{\psi}^\dagger(x) + f^\dagger(x) \hat{\psi}(x)] \right) | \Psi \rangle . \quad (2.130)
\end{equation}

This is an analogue of the moments generating functional discussed by Kubo.\textsuperscript{12} Upon expanding the exponential as a power series, the effect of the operator $\hat{\mathcal{N}}$ on any particular term is to place the field operators in so-called normal order, creation operator $\hat{\psi}^\dagger$ to the left of all annihilation operators $\hat{\psi}$. By virtue of this ordering (and only by virtue of this ordering) we may write

\begin{equation}
G[f, f^\dagger] = \exp \left( \int dx \ f(x) \hat{\psi}^\dagger(x) \right) \exp \left( \int dx' \ f^\dagger(x') \hat{\psi}(x') \right) \\
= 1 + \mathcal{F}[f, f^\dagger] , \quad (2.131)
\end{equation}
where $\langle \cdots \rangle = \langle \Psi | \cdots | \Psi \rangle$ and

$$
\mathcal{F}[f, f^\dagger] = \sum_{p=1}^{\infty} \frac{1}{p!^2} \left( \int dx_1 \cdots dx_p \, dx'_1 \cdots dx'_p \, f(x_1) \psi^\dagger(x_1) \cdots f(x_p) \psi^\dagger(x_p) \right. \\
\left. \times f^\dagger(x'_p) \, \psi(x'_p) \cdots f^\dagger(x'_1) \, \psi(x'_1) \right). 
$$

(2.132)

Note that the only surviving terms from Eq. (2.131) are those that preserve particle number, hence the two exponentials in Eq. (2.131) produce only a single summation in Eq. (2.132).

The cumulant generating functional is $\ln \mathcal{G}$. That is, the $p$-particle reduced density matrix cumulant (p-RDMC) $\Delta_p$ is defined to be

$$
\Delta_p^{(1, \ldots, p'; 1, \ldots, p')} = \frac{1}{p!} \left( \lim_{f, f^\dagger \to 0} \frac{\delta^{2p}(\ln \mathcal{G})}{\delta f(1) \delta f^\dagger(1') \cdots \delta f(p) \delta f^\dagger(p')} \right). 
$$

(2.133)

The factor of $1/p!$ arises due to our convention that $\text{tr} \, D_p = \binom{N}{p}$. The normalization of the cumulants is more complicated, but some specific examples are given in what follows. In a formal expansion of the logarithm $\ln \mathcal{G} = \ln(1 + \mathcal{F})$, the functional derivatives in Eq. (2.133) serve to select all terms consisting of exactly $p$ creation operators $\psi^\dagger$ and $p$ annihilation operators $\psi$, while at the same time eliminating the integrals and replacing the dummy integration variables by the particle coordinates $x_1, \ldots, x_p$ and $x'_1, \ldots, x'_p$.

The generating functional $\ln \mathcal{G}[f, f^\dagger]$ is a tidy way to generate the cumulant kernels directly, but for those who like indices, Mazziotti\textsuperscript{13,14} has introduced a generating functional for the expansion coefficients $\Delta_{i_1, \ldots, i_p; j_1, \ldots, j_p}$ of $\Delta_p$ in a basis $\{\phi_k\}$ of orthonormal spin orbitals. Mazziotti’s formalism is obtained from the above considerations by expanding the Grassmann fields and the field operators in this basis,

$$
\hat{\psi}(x) = \sum_k \hat{a}_k \phi_k(x) 
$$

(2.134)
and

\[ f(x) = \sum_k J_k \phi_k(x) . \]  

(2.135)

The \( J_k \) are the probe variables in this formulation. The reader is also referred to Kutzelnigg and Mukherjee’s treatment,\textsuperscript{142} which utilizes an antisymmetrized logarithm function, and also some special creation and annihilation operators, to generate the elements \( (\Delta_p)^{(i_1, \ldots, i_p)} \).

Using either Eq. (2.133) or Mazziotti’s adaptation of it, one may derive exact expression for the RDMs in terms of their cumulants. In terms of the antisymmetrized tensor product defined in Eq. (2.46), these formulas are

\[
\begin{align*}
D_1 &= \Delta_1 \quad \text{(2.136a)} \\
D_2 &= \Delta_1^{\wedge 2} + \Delta_2 , \quad \text{(2.136b)} \\
D_3 &= \Delta_1^{\wedge 3} + 3 \Delta_2 \wedge \Delta_1 + \Delta_3 , \quad \text{(2.136c)} \\
D_4 &= \Delta_1^{\wedge 4} + 6 \Delta_2 \wedge \Delta_1^{\wedge 2} + 3 \Delta_2^{\wedge 2} + 4 \Delta_3 \wedge \Delta_1 + \Delta_4 , \quad \text{(2.136d)} \\
D_5 &= \Delta_1^{\wedge 5} + 10 \Delta_2 \wedge \Delta_1^{\wedge 3} + 10 \Delta_3 \wedge \Delta_1^{\wedge 2} + 5 \Delta_4 \wedge \Delta_1 \\
&\quad + 15 \Delta_1 \wedge \Delta_2^{\wedge 2} + 10 \Delta_2 \wedge \Delta_3 . \quad \text{(2.136e)}
\end{align*}
\]

Equation (2.136b) is sometimes called the Levy-Lieb\textsuperscript{161,162} partitioning of the 2-RDM. Formulas equivalent to Eqs. (2.136a)–(2.136e) were known long ago, in the context of time-dependent Green’s functions,\textsuperscript{15,16,163,164} but were recently rediscovered, and brought more to the attention of quantum chemists, by Mazziotti.\textsuperscript{14} For an interesting connection between the RDM cumulant decompositions and the well-known cluster decomposition of the electronic wavefunction, see Subsection 4.6.

Harris\textsuperscript{165} provides the following mnemonic for generating the cumulant decom-
positions with a minimum of algebraic tedium. To obtain the cumulant decomposition of \( D_{p+1} \) from that of \( D_p \), add up—for each term in the former—all possible ways in which the particle number of this term may be increased by one. Particle number can be increased either by changing \( \Delta_n \) to \( \Delta_{n+1} \) or by means of an additional Grassmann product with \( \Delta_1 \). As an example, consider generating \( D_3 \) from \( D_2 \). We can increase the particle number of \( \Delta_1 \wedge \Delta_1 \) by changing either factor of \( \Delta_1 \) to \( \Delta_2 \), or else by appending a third factor of \( \Delta_1 \). From this first term in Eq. (2.136b) we therefore obtain a contribution \( 2 \Delta_1 \wedge \Delta_2 + \Delta_1 \wedge \Delta_1 \wedge \Delta_1 \) to \( D_3 \). The same procedure applied to \( \Delta_2 \) yields \( \Delta_3 + \Delta_2 \wedge \Delta_1 \). Summing the contributions from the first and second terms in Eq. (2.136b), one obtains exactly the right side of Eq. (2.136c).

This mnemonic emphasizes the combinatorial nature of the cumulants; for instance, the term \( \Delta_1 \wedge \Delta_2 \) in \( D_3 \) carries a coefficient of 3 because there are three ways in which to combine a one-particle distribution function and a two-particle distribution function to obtain a three-particle distribution. Up to the use of an antisymmetrized product, the RDM cumulants are identical to cumulant decompositions of the many-particle distribution functions in classical statistical mechanics. The difference \( \Sigma_p \equiv D_p - \Delta_p \) between the actual \( p \)-particle distribution \( D_p \) and its cumulant \( \Delta_p \) equals the three-particle distribution that one would expect were simultaneous \( p \)-particle correlations negligible.

In the present, quantum-mechanical formulation, the antisymmetrized product insures that the distribution of composite \( p \)-particles respects the Pauli exclusion principle and Fermi statistics, even in the case that simultaneous \( p \)-particle correlations vanish. (Quantum mechanically, even uncorrelated particles interact by exchange.) In a manner of speaking, exchange coupling has simply been grafted onto the classical
cumulant formalism, via the Grassmann product, but apart from antisymmetry the cumulants themselves know nothing about spin or spin coupling (for instance, coupling to an $\hat{S}_z^2$ eigenstate). Consequently, although the distribution $\Sigma_p$ properly goes to zero as two same-spin electrons approach one another, this distribution fails to respect other important spin symmetries, as detailed in Subsection 4.3.3.

Let us list a few other salient mathematical properties of $\Delta_p$. First, this quantity is self-adjoint and antisymmetric. For a multiplicatively separable wavefunction like the one in Eq. (2.123), the matrix elements of $\Delta_p$ vanish unless all indices correspond to the same subsystem. More precisely, using the notation introduced in the previous subsection, $\Delta_{i_1,\ldots,i_p;j_1,\ldots,j_p} = 0$ unless $\phi_m \in B_\lambda$ for each index $m$ or else $\phi_m \in B_n$ for each index $m$. This is the essential difference that allows for an extensive formulation of quantum mechanics in terms of the cumulants but not in terms of the RDMs themselves; the ramifications of this difference are explored more thoroughly in the next subsection.

2.5.3 Extensivity

As mentioned previously, the primary reason for working with the RDMCs instead of the RDMs themselves is that the former are extensive quantities, and therefore lend themselves to computational methods that are manifestly size-consistent, even when based on approximate quantum mechanics. We saw in Subsection 2.5.1 that the basic problem with the RDMs was the manner in which the exchange terms mix the coordinates corresponding to non-interacting subsystems. Such terms may be identified by the presence of a Grassmann product, and $\Delta_p$ equals the difference between $D_p$ and these exchange terms. Looking over Eqs. (2.136a)–(2.136e), it is clear that knowledge
of the cumulants $\Delta_1, \ldots, \Delta_n$ is equivalent to knowledge of the RDMs $D_1, \ldots, D_n$ but only the latter contain Grassmann products. In Chapter 3 we will recast the basic equations for direct determination of the 2-RDM into equations for direct determination of the RDMCs $\Delta_1$ and $\Delta_2$. The RDMs do not appear in the reformulated equations.

Let us explore the extensivity of the cumulants a bit further. It is apparent from Eq. (2.136) that the trace of any term containing a Grassmann product will scale asymptotically as $N^n$ for some $n > 1$. For example, the Grassmann product

$$[\Delta_1 \wedge \Delta_1](1;2;1';2') = \frac{1}{2} [\Delta_1(1;1') \Delta_1(2;2') - \Delta_1(1;2') \Delta_1(2;1')]$$

appearing as part of $D_2$ has a trace given by

$$\text{tr}(\Delta_1 \wedge \Delta_1) = N^2 - \text{tr} \Delta_1^2.$$  

Asymptotically (as $N \to \infty$), $\text{tr}(\Delta_1 \wedge \Delta_1) \sim N^2$. One says that $\Delta_1 \wedge \Delta_1$ scales like $N^2$.

This last statement tacitly assumes that $\text{tr} \Delta_1^2$ does not contain any terms with a quadratic or steeper dependence on $N$, which demonstrated below. First introduce the quantities

$$\eta_k \equiv \frac{\text{tr} \Delta_1^k}{N}.$$  

Since the eigenvalues of $\Delta_1 \equiv D_1$ each lie between 0 and 1, one has $0 \leq \eta_k \leq 1$ for each $k$. This is true even for extended systems, where $N \to \infty$. Our claim that $\text{tr}(\Delta_1 \wedge \Delta_1) \sim N^2$ follows since $0 \leq \eta_2 \leq 1$. In fact, without loss of generality one may assume that $\eta_k > 0$, since the $N$-electron wavefunction can always be expressed using only those NSOs with strictly positive occupation numbers. The limiting case
$\eta_k = 1$ corresponds to a single determinant (SD) state, for which $D_1$ is idempotent and\(^{20}\)

$$D_p \overset{\text{SD}}{=} D_1^{\wedge p}.$$  (2.140)

Note that this form of $D_p$ is consistent with the vanishing of $\Delta_p$ for each $p > 1$. This is merely a restatement of the obvious fact that a SD (independent-particle) wavefunction corresponds to a collection of one-electron subsystems that are coupled only by exchange.

Although we will not give a general proof that $\text{tr} \Delta_n \propto N$, for small $n$ this is easy to verify by direct calculation. For instance

$$\text{tr} \Delta_2 = \frac{1}{2} N (\eta_2 - 1),$$  (2.141)

hence $\text{tr} \Delta_2 \sim N$, even as $N \to \infty$. Similarly,

$$\text{tr} \Delta_3 = \frac{1}{3} N (1 - 5\eta_2 + 2\eta_3)$$  (2.142)

scales like $N$. Traces of $\Delta_2$ and $\Delta_3$ were examined for some model wavefunctions in Ref. 142. Whereas $\text{tr} D_2 = \binom{N}{2}$ depends only on the number of electrons, the trace of $\Delta_2$ differs from one $N$-electron system to the next, depending upon how far $D_1$ deviates from idempotency. This is a general feature of $\text{tr} \Delta_n$. Also note that

$$-\frac{1}{2} N \leq \text{tr} \Delta_2 \leq 0.$$  (2.143)

This does not exclude the possibility that $\Delta_2$ can have both positive and negative eigenvalues.

The contractions of the cumulants are also extensive, whereas those of the RDMs are not. Starting from Eq. (2.136c), for example, one may show that

$$\text{tr}_3 \Delta_3 = \frac{2}{3} \left[ -\Delta_2 + \hat{A}_2(\Delta_1 \Delta_2) + \bar{A}_2(\Delta_2 \Delta_1) \right].$$  (2.144)
(Recall that $\tilde{A}_2$ and $\hat{A}_2$ antisymmetrize upper and lower indices, respectively.) We have verified that $\Delta_2$ is extensive, but what about the other two terms on the right side of this equation? These are extensive as well, since $\Delta_1\Delta_2$ and $\Delta_2\Delta_1$ are matrix products of connected factors. One may also verify directly that $\text{tr}(\Delta_1\Delta_2) = N(\eta_3 - \eta_2)/2$.

In general, one can show that $\text{tr}\Delta_n \propto N$ for each $n$, whereas $\text{tr}D_n = \binom{N}{n} \sim N^n$. This is a convenient means to recognize terms that are not extensive, but in some sense overlooks the physical picture behind extensivity. Similarly, insofar as perturbation theory is applicable, the fact that the cumulants scale as $N$ can be viewed as a consequence of the linked-cluster theorem, but the deeper concept of extensivity does not depend upon the validity of perturbation theory. Mathematically, extensivity is a statement about connectivity, in a matrix-product sense, as we saw in Eq. (2.144). The non-perturbative diagrammatic notation for the cumulants that is introduced in Section 3.3 will emphasize this point. In contrast, the expansion of $D_n$ contains unconnected products of diagrams, up to and including the product of $n$ unconnected one-electron diagrams. Energetic contributions from the unconnected diagrams cancel when an exact wavefunction or density matrix is employed, but this feature is not generally preserved by approximations. This is the reason, for example, that FCI is size-consistent but any truncated CI method is not.

2.5.4 Independence of the cumulants

From Eqs. (2.136a) and (2.136b) it is clear that $\Delta_1$ and $\Delta_2$ contain the same information as $D_2$ and can therefore be used to calculate $\langle \hat{W} \rangle$. Unlike the 2-RDM, which contains the 1-RDM and thus affords $\langle \hat{W} \rangle$ with no other information, the 2-RDMC
does not generally determine the 1-RDMC,\textsuperscript{152,167} so both $\Delta_1$ and $\Delta_2$ must be separately determined in order to calculate $\langle \hat{W} \rangle$. In this section, we present a proof\textsuperscript{152} of the independence of $\Delta_1$ and $\Delta_2$. More generally, $\Delta_1, \ldots, \Delta_n$ are independent quantities, whereas the RDMs $D_1, \ldots, D_n$ are interrelated by the partial trace operation. $D_n$ determines all of the lower-order RDMs, and thus all of the lower-order cumulants, but $\Delta_n$ alone is insufficient to specify any of the other cumulants or RDMs.

We know that $\text{tr} D_2$ determines $D_1 \equiv \Delta_1$ and therefore $D_2$ determines $\Delta_2$ via the cumulant decomposition in Eq. (2.136b). The converse is untrue, however. Observe that\textsuperscript{*}

$$\text{tr} \Delta_2 = \frac{1}{2}(\Delta_1^2 - \Delta_1), \quad (2.145)$$

from which it follows that $\Delta_1$ and $\text{tr} \Delta_2$ share a common set of eigenvectors, namely, the NSOs. For each eigenvalue $n_k$ of $\Delta_1$, let $e_k$ be the eigenvalue of $\text{tr} \Delta_2$ associated with the same NSO. These eigenvalues are related according to

$$e_k = \frac{1}{2} n_k(n_k - 1), \quad (2.146)$$

which has solutions

$$n_k = \frac{1}{2} \left( 1 \pm \sqrt{1 + 8 e_k} \right). \quad (2.147)$$

Thus $n_k$ is double-valued as a function of $e_k$, as depicted in Fig. 2.1. Strictly speaking, then, the eigenvalues of $\text{tr} \Delta_2$ do not determine those of $\Delta_1$, and consequently $\Delta_1$ cannot be determined from $\Delta_2$ alone.

For systems that do not exhibit strong multideterminant character (including virtually all singlet ground states), each eigenvalue of $\Delta_1 \equiv D_1$ will lie near 0 or

\textsuperscript{*}We note in passing that the spin-traced analogue of the one-electron matrix $\text{tr} \Delta_2$ has been called the density matrix for effectively unpaired electrons,\textsuperscript{168–171} and forms the basis for a Mulliken-style\textsuperscript{172–175} population analysis that aims to characterize deviations from closed-shell electronic structure,\textsuperscript{170} especially for singlet-coupled diradicals.
1. Given $\Delta_2$ (and thus the $e_k$) for such a system, it may be possible to choose, for each $k$, one of the two solutions $n_k$ in Eq. (2.147), based upon the nature of the $k$th eigenvector and our intuition as to whether this eigenvector should be strongly or weakly occupied.* To see this, suppose that either $n_k = \varepsilon$ or $n_k = 1 - \varepsilon$. Upon calculating $\varepsilon_k$ corresponding to each, and substituting this back into Eq. (2.147), one obtains in either case a choice between the solutions $n_k = 1 - \varepsilon + O(\varepsilon^2)$ and $n_k = \varepsilon + O(\varepsilon^2)$. So long as $\varepsilon^2 \ll \varepsilon$, and assuming that one can ascertain whether each eigenvector of $\text{tr}_2 \Delta_2$ should be strongly or weakly occupied, $\Delta_1$ can be obtained from $\Delta_2$. In this case $\langle \vec{W} \rangle$ can be determined solely from $\Delta_2$.

### 2.6 Contraction of Grassmann products

The algebra of antisymmetrized products plays a crucial role in fermion quantum mechanics, as we have already seen. Within the RDM formalism, the partial trace

---

*This should be immediately evident if one examines the eigenvectors of $\Delta_1$ in the HF basis.
relation [Eq. (2.11)] that connects the RDMs is also important. In this section we evaluate partial traces of various antisymmetrized products of 1- and 2-RDMs. The result is a number of general theorems for manipulating these quantities, that will prove quite useful in subsequent derivations. Our original motivation to pursue this line of research was an interest in the ansatz

\[ D_4^G \overset{\text{def}}{=} a(D_2 \wedge D_2) + b(D_2 \wedge D_1^{\wedge 2}) + c D_1^{\wedge 4} \]  

as an approximate form for the 4-RDM as a functional of the 2-RDM. In this regard, \( a, b, \) and \( c \) were envisioned as parameters of some sort. Presently this line of research does not look promising, but the formal results remain, and by working through their derivations one may obtain great facility with Grassmann algebra that is useful in many other contexts.

We mention in passing several other Grassmann-type relationships for RDMs. Valdemoro and co-workers\(^69\) have derived an explicit formula for the one-index contraction of \( I^{\wedge p} \wedge D_q \) as a linear combination of the Grassmann products \( I^{\wedge p} \wedge D_{q-1} \) and \( I^{\wedge (p-1)} \wedge D_q \). Coleman and Absar\(^38\) and also Casida and Harriman\(^176\) have considered decompositions of RDMs into certain Grassmann products that transform as irreducible representations of the unitary group. None of the aforementioned work provides explicit expressions for the contractions of the Grassmann product 4-RDM \( D_4^G \), however. This problem is considered in Subsection 2.6.3, but first we present some contraction theorems for very general Grassmann products of RDMs.

### 2.6.1 Contraction theorems

In this section we establish several theorems concerning the one-particle contractions of antisymmetrized products of \( D_1 \) and \( D_2 \) that will facilitate contraction of \( D_4^G \). The
techniques used in each proof are similar, so we provide detailed proofs only for the first two theorems in this section. Because the notation used in the theorems (not to mention the proofs) is somewhat involved, following several of the theorems we present examples of their application prior to embarking upon a formal proof.

We first define a deletion operator $\ominus$. Let $S = S_1 \land S_2 \land \cdots \land S_n$ be an antisymmetrized product of a set of tensors $S = \{S_i|i = 1, \ldots, n\}$, and suppose $S' \subseteq S$. Then

$$S \ominus S' \overset{\text{def}}{=} \bigwedge_{S_i \in S \setminus S'} S_i.$$ (2.149)

For example, if $S' = \{S_1, S_2\}$ then $S \ominus S' = S_3 \land \cdots \land S_n$.

**Theorem 2.1.** Let $n$ be a positive integer and $p_1, \ldots, p_n$ nonnegative integers with $p_1 + \cdots + p_n = P$. Define the $P$-particle tensor

$$T = (D_1^{p_1}) \wedge (D_1^{p_2}) \land \cdots \land (D_1^{p_n}).$$ (2.150)

Then the one-particle contraction of $T$ is given by

$$\text{tr}_P(T) = \frac{1}{P^2} \sum_{m=1}^{n} p_m \text{tr} (D_1^{p_m}) \left[ T \ominus \{D_1^{p_m}\} \right]$$

$$- \frac{1}{P^2} \sum_{m=1}^{n} p_m (p_m - 1) \left[ T \ominus \{D_1^{p_m}, D_1^{p_m}\} \right] \land D_1^{2m}$$

$$- \frac{2}{P^2} \sum_{m=1}^{n} \sum_{l=m+1}^{n} p_m p_l \left[ T \ominus \{D_1^{p_m}, D_1^{p_l}\} \right] \land D_1^{l+m}. \quad (2.151)$$

**Remark 2.1.** Deletions such as $T \ominus \{D_1^{p_m}, D_1^{p_m}\}$ make sense only if $p_m \geq 2$. If $p_m < 2$, however, this term does not appear as a result of the multiplicative factor $p_m (p_m - 1)$.

**Example 2.1.** Since $\text{tr} D_1 = N$,

$$\text{tr}_p (D_1^{\lambda p}) = \left( \frac{N}{p} \right) D_1^{\lambda (p-1)} - \left( \frac{p-1}{p} \right) \left[ D_1^{\lambda (p-2)} \and D_1^{2} \right] \quad (2.152)$$

for $p \geq 2$. This illustrates the case $n = 1$ and $p_1 = p$. 
Example 2.2. The case \( n = 3, p_1 = 2, p_2 = 0, \) and \( p_3 = 1 \) \((P = 3)\) corresponds to
\[
\text{tr}_3(D_1 \wedge D_1 \wedge D_1^3) = \frac{1}{9} \left[ 2N(D_1 \wedge D_1^3) + (\text{tr } D_1^3)(D_1 \wedge D_1) \\
- 2(D_1^2 \wedge D_1^3) - 4(D_1 \wedge D_1^1) \right]. 
\] (2.153)

*Proof of Theorem 2.1.* The elements of \( \mathbf{T} \) are given by
\[
T_{j_1, \ldots, j_P} = (P!)^{-2} \sum_{\sigma, \pi \in \mathbb{S}_P} \epsilon(\sigma) \epsilon(\pi) \left( D_1^\omega \right)_{\pi(j_1)}^{\sigma(i_1)} \cdots \left( D_1^\nu \right)_{\pi(j_P)}^{\sigma(i_P)}, 
\] (2.154)
where \( \omega \) and \( \nu \) are, respectively, the smallest and largest values of \( m \) such that \( p_m > 0 \).

For simplicity we define two ordered \( P \)-tuples,
\[
I_P \overset{\text{def}}{=} \{i_1, \ldots, i_P\} \\
J_P \overset{\text{def}}{=} \{j_1, \ldots, j_P\}, 
\] (2.155)
which are re-ordered by permutations \( \sigma, \pi \in \mathbb{S}_P \) according to
\[
\sigma(I_P) = \{\sigma(i_1), \ldots, \sigma(i_P)\} \tag{2.156a} \\
\pi(J_P) = \{\pi(j_1), \ldots, \pi(j_P)\}. \tag{2.156b}
\]

Furthermore, define products
\[
[D_1(\vec{x})]_{I_P}^{I_P} \overset{\text{def}}{=} \left( D_1^\omega \right)_{j_1}^{i_1} \cdots \left( D_1^\nu \right)_{j_P}^{i_P}, 
\] (2.157)
in which \( \vec{x} = \{\omega, \ldots, \nu\} \) is a \( P \)-tuple whose \( r \)th component \( x_r \) is the exponent for the \( r \)th term in the product. We can express the elements of \( \text{tr}_P(\mathbf{T}) \) as
\[
(\text{tr}_P \mathbf{T})_{j_1, \ldots, j_{P-1}}^{i_1, \ldots, i_{P-1}} = (P!)^{-2} \text{tr}_P \left( T_{j_P}^{I_P} \right) 
\] (2.158)
where
\[
T_{j_P}^{I_P} \overset{\text{def}}{=} \sum_{\sigma, \pi \in \mathbb{S}_P} \epsilon(\sigma) \epsilon(\pi) \left[ D_1(\vec{x}) \right]_{\pi(j_P)}^{\sigma(I_P)}. 
\] (2.159)
Because the $P$-trace operator can be written as
\[
\text{tr}_P(\cdots) = \sum_k \delta_{i_P,k} \delta_{j_P,k}(\cdots),
\] (2.160)
we will group the permutations in Eq. (2.159) according to where they place $i_P$ and $j_P$. There are three possibilities: $i_P$ and $j_P$ can index the same matrix, they can index different matrices having the same exponent, or they can index different matrices with different exponents. This leads to a partition
\[
T_{j_P}^{I_P} = \sum_r (T_{j_P}^{I_P})_{rr} + \sum_{r \neq s} (T_{j_P}^{I_P})_{rs} + \sum_{rs} (T_{j_P}^{I_P})'_{rs},
\] (2.161)
where each summand represents a restricted sum over permutations in Eq. (2.159). $(T_{j_P}^{I_P})_{rr}$ includes those $\sigma, \pi \in S_P$ for which $\sigma(r) = P = \pi(r)$. Included in $(T_{j_P}^{I_P})_{rs}$ are permutations for which $x_r = x_s$, $r \neq s$, and $\sigma(r) = P = \pi(s)$. Finally, $(T_{j_P}^{I_P})'_{rs}$ is restricted to $\sigma$ and $\pi$ with $\sigma(r) = P = \pi(s)$ and $x_r \neq x_s$. Together the three summations in Eq. (2.161) exhaust all pairs of permutations in $S_P \otimes S_P$. We evaluate each sum separately, then apply $\text{tr}_P$ to each and thus obtain the three terms in $\text{tr}_P(T)$.

Consider the first summation in Eq. (2.161). To treat terms such as $(T_{j_P}^{I_P})_{rr}$, we introduce permutations $\tau_r^{(P)} \in S_P$ that act on other permutations $\sigma \in S_P$ according to
\[
\tau_r^{(P)} \sigma(I_P) = \left\{ \sigma(i_1), \ldots, \sigma(i_{r-1}), \sigma(i_{r+1}), \ldots, \sigma(i_P), \sigma(i_r) \right\}
= \left\{ \sigma'(I_r^{(P-1)}), \sigma(i_r) \right\}.
\] (2.162)
In this equation we have introduced
\[
I_r^{(P-1)} \defeq I_P \setminus \{\sigma(i_r)\}.
\] (2.163)
Also, $\sigma' \in S_{P-1}$ in Eq. (2.162) is the permutation that places the $P-1$ indices of $I_r^{(P-1)}$ in the same order that they have in $\sigma(I_P)$. With similar notation as in Eq. (2.162),
we write
\[
\tau^{(P)}_r \pi(JP) = \left\{ \pi'(J^{(P-1)}_r), \pi(j_r) \right\}
\] (2.164)
for some \( \pi' \in \mathbb{S}_{P-1} \). Observe that \( \epsilon(\sigma') = \epsilon(\tau^{(P)}_r) \epsilon(\sigma) \) and \( \epsilon(\pi') = \epsilon(\tau^{(P)}_r) \epsilon(\pi) \), so \( \epsilon(\sigma') \epsilon(\pi') = \epsilon(\sigma) \epsilon(\pi) \).

Recall that \( i_r \) and \( j_r \) index the matrix \( D^{x_r}_{i} \) in the product \([D_{1}(\bar{x})]_{JP}^{j_r}\). This implies that \( p_{x_r} > 0 \); in fact there are exactly \( p_{x_r} \) choices of \( r \) with the same exponent \( x_r \). Thus
\[
(T_{JP}^{j_r})_{rr} = \sum_{\sigma, \pi \in \mathbb{S}_{P}} \epsilon(\sigma') \epsilon(\pi') (D^{x_r}_{1})_{j_{P}p}^{j_{P}P} \left( [D_{1}(\bar{x})]_{\pi'}^{P} \right) \left( [D_{1}(\bar{x})]_{p}^{j_{P}(P-1)} \right) = [(P-1)!]^2 \left[ \mathbf{T} \otimes \{D^{x_r}_{1}\}^{j_{P-1}}_{j_{P-1}} \right] \left( D^{x_r}_{1} \right)_{j_{P}}^{j_{P-1}}. 
\] (2.165)
\[ [\mathbf{T} \otimes \{D^{x_r}_{1}\}]^{j_{P-1}}_{j_{P-1}} \] is, by definition, a Grassmann product on \( P-1 \) indices, and includes a factor of \( [(P-1)!]^{-2} \).

Application of \( \text{tr}_P \) to Eq. (2.165) gives \( \text{tr} D^{x_r}_{1} \) in place of the last term \( (D^{x_r}_{1})_{j_{P}}^{j_{P}} \).

There are \( p_{x_r} \) (possibly \( p_{x_r} = 0 \)) values of \( r \) for a given \( x_r \), and it follows that
\[
\text{tr}_P \left[ (P!)^{-2} \sum_{r} (T_{JP}^{j_r})_{rr} \right] = P^{-2} \sum_{m=1}^{n} p_{m}(\text{tr} D^{m}_{1}) \left[ \mathbf{T} \otimes \{D^{m}_{1}\} \right]. 
\] (2.166)
This is precisely the first term in the statement of the theorem.

Next consider the second and third summations in Eq. (2.161). Up to Eq. (2.173) below, the evaluation of the two summands \( (T_{JP}^{j_r})_{rs} \) and \( (T_{JP}^{j_r})_{rs}^{j_{P}} \) proceeds in the same fashion. We first evaluate \( (T_{JP}^{j_r})_{rs} \), where \( r \neq s \), then make appropriate modifications to obtain \( (T_{JP}^{j_r})_{rs}^{j_{P}} \).

Fix \( r \neq s \) and consider the action of permutations \( \tau^{(P)}_r \in \mathbb{S}_{P} \) and \( \tau_{s}^{(P-1)} \in \mathbb{S}_{P-1} \).
as defined in Eq. (2.162):

\[
\tau^{(P-1)}_s \tau^{(P)}_r \sigma(I_P) = \tau^{(P-1)}_s \left\{ \sigma'(I^{(P-1)}_r), \sigma(i_r) \right\} \\
= \left\{ \sigma''(I^{(P-2)}_{r,s}), \sigma(i_s), \sigma(i_r) \right\},
\]

(2.167)

where

\[
\sigma'(I^{(P-1)}_r) = \left\{ \sigma''(I^{(P-2)}_{r,s}), \sigma(i_s) \right\}.
\]

(2.168)

Here \(\sigma'' \in S_{P-2}\) orders the indices in \(I^{(P-2)}_{r,s} \overset{\text{def}}{=} I_P \setminus \{\sigma(i_r), \sigma(i_s)\}\) in the same way that they appear in \(\sigma(I_P)\). Similarly,

\[
\tau^{(P-1)}_s \tau^{(P)}_r \pi(J_P) = \left\{ \pi''(J^{(P-2)}_{r,s}), \pi(j_s), \pi(j_r) \right\}
\]

(2.169)

for some \(\pi'' \in S_{P-2}\). Then \(\epsilon(\sigma'') = \epsilon(\tau^{(P-1)}_s \tau^{(P)}_r) \epsilon(\sigma)\) and \(\epsilon(\pi'') = \epsilon(\tau^{(P-1)}_s \tau^{(P)}_r) \epsilon(\pi)\), consequently \(\epsilon(\sigma'') \epsilon(\pi'') = \epsilon(\sigma) \epsilon(\pi)\).

Thus we have

\[
(T^I_{J_P})_{rs} = \sum'_{\sigma,\pi \in S_P \atop \sigma(r) = P = \pi(s)} \epsilon(\sigma'') \epsilon(\pi'') \left[ D_1(\bar{x}) \setminus \{ D^{x_r}_1, D^{x_s}_1 \} \right]^{\sigma''(I^{(P-2)}_{r,s})}_{\pi''(I^{(P-2)}_{r,s})} \\
\times \left( D^{x_s}_1 \right)_{\pi(j_s)}^{\sigma(i_s)} \left( D^{x_r}_1 \right)_{\pi(j_r)}^{\sigma(i_r)}
\]

(2.170)

so that

\[
\text{tr}_P \left[ \sum_r \sum_{s \neq r} (T^I_{J_P})_{rs} \right] = \sum_r \sum_{s \neq r} \sum'_{\sigma,\pi \in S_P \atop \sigma(r) = P = \pi(s)} \epsilon(\sigma'') \epsilon(\pi'') \left[ D_1(\bar{x}) \setminus \{ D^{x_r}_1, D^{x_s}_1 \} \right]^{\sigma''(I^{(P-2)}_{r,s})}_{\pi''(I^{(P-2)}_{r,s})} \\
\times \left( D^{x_r+x_s}_1 \right)_{\pi(j_r)}^{\sigma(i_s)} \left( D^{x_r+x_s}_1 \right)_{\pi(j_s)}^{\sigma(i_r)}
\]

(2.171)

To cast this as a Grassmann product, consider the permutation \(\sigma' \in S_{P-1}\) defined in Eq. (2.168), along with the analogous permutation \(\pi' \in S_{P-1}\),

\[
\pi'(J^{(P-1)}_s) = \left\{ \pi''(J^{(P-2)}_{r,s}), \pi(j_r) \right\}.
\]

(2.172)
Comparing Eqs. (2.167), (2.168), and (2.172), one finds that 
\[ \epsilon(\sigma') \epsilon(\pi') = -\epsilon(\sigma) \epsilon(\pi). \]

Thus
\[
\text{tr}_P \left[ \sum_r \sum_{s \neq r} (T^I_{J^P})_{rs} \right] = -\sum_r \sum_{s \neq r} \sum_{\sigma', \pi' \in \mathbb{S}_{P-1}} \epsilon(\sigma') \epsilon(\pi') \times \left[ (D^I_1(\bar{x}) \setminus \{D^x_r, D^x_s\}) \cdot D^x_{\sigma' r + x_s} \right]_{\pi'(J^{(P-1)})}^{\sigma'(I^{(P-1)})} 
\]
\[
= -[(P-1)!]^2 \sum_r \sum_{s \neq r} \left[ [T \ominus \{D^x_r, D^x_s\}] \wedge D^2_{x_{r+s}} \right]_{J^{(P-1)}}^{I^{(P-1)}}.
\]

So far we have not used the fact that \( x_r = x_s \); consequently Eq. (2.173) remains valid if we replace \((T^I_{J^P})_{rs}\) with \((T^I_{J^P})'_{rs}\). To evaluate Eq. (2.173) with \( x_r = x_s \), observe that there are \( p_{x_r} \) choices for \( r \) and \( (p_{x_r} - 1) \) choices for \( s \neq r \) such that \( x_r = x_s \). Hence
\[
(P!)^{-2} \text{tr}_P \left[ \sum_r \sum_{s \neq r} (T^I_{J^P})'_{rs} \right] = -P^{-2} \sum_m p_m (p_m - 1) \left[ T \ominus \{D^x_r, D^x_s\} \right] \wedge D^2_{1m}.
\]

To obtain \( \text{tr}_P \sum_{r,s} (T^I_{J^P})'_{rs} \), consider the aforementioned analogue of Eq. (2.173), assuming \( x_r \neq x_s \). Then there are \( p_{x_r} \) choices \( x_r \) for the exponents with \( r > s \), and another \( p_{x_r} \) choices \( x_s \) choices with \( r < s \). Both cases give the same result in Eq. (2.173) so
\[
(P!)^{-2} \text{tr}_P \left[ \sum_{rs} (T^I_{J^P})'_{rs} \right] = -2P^{-2} \sum_{l=1}^n \sum_{m=l+1}^n p_l p_m \left[ T \ominus \{D^x_r, D^x_s\} \right] \wedge D^m_{1l}.
\]

Equations (2.174) and (2.175) are, respectively, the second and third terms in the statement of the theorem.}

In order to state the next theorem succinctly we define one-index products \( D_1D_2 \).
and $D_2D_1$,

$$
(D_1D_2)_{ij}^{rs} \overset{\text{def}}{=} \sum_m (D_1)_m^{rs} (D_2)_m^{ij},
$$

$$
(D_2D_1)_{ij}^{rs} \overset{\text{def}}{=} \sum_m (D_2)_m^{rs} (D_1)_m^{ij}.
$$

(2.176)

If we view $D_2$ as a matrix of matrices, then these tensors result from either right- or left-multiplication of each block of $D_2$ by the matrix $D_1$. Note that $D_2D_1$ is antisymmetric in its lower indices only, $D_1D_2$ is antisymmetric only in its upper indices, and $(D_1D_2)^\dagger = D_2D_1$. Retaining from the preceding proof the notation $I_{p+q}$ and $J_{p+q}$ to denote ordered $(p + q)$-tuples that provide a carrier space for $S_{p+q}$, we further define

$$
[AB]_{I_{p+q}}^{J_{p+q}} \overset{\text{def}}{=} A_{i_1,\ldots,i_p} B_{j_p+1,\ldots,j_{p+q}}^{i_{p+1},\ldots,i_{p+q}}
$$

(2.177)

for tensors $A = (A_{i_1,\ldots,i_p})$ and $B = (B_{j_p,\ldots,j_q})$.

Lemma 2.1. Let $A = (A_{i_1,\ldots,i_p})$ and put

$$
T = \sum_{\sigma,\pi \in S_{p+2}} \epsilon(\sigma) \epsilon(\pi) [AD_2]^{\sigma(I_{p+2})}_{\pi(J_{p+2})}.
$$

(2.178)

Let $\vartheta_1 = (i_{p+1}, i_{p+2})$ and $\vartheta_2 = (j_{p+1}, j_{p+2})$ be transpositions. Then

$$
\vartheta_1 T = T = \vartheta_2 T.
$$

(2.179)

Proof. For any $\sigma \in S_{p+2}$, $\vartheta_1 \sigma(I_{p+2}) = \{\sigma'(I_p), \sigma(i_{p+2}), \sigma(i_{p+1})\}$ for some $\sigma' \in S_p$. Clearly $\epsilon(\vartheta_1 \sigma) = -\epsilon(\sigma)$ so for any pair of permutations $\sigma, \pi \in S_{p+2}$

$$
\epsilon(\vartheta_1 \sigma) \epsilon(\pi) [AD_2]^{\vartheta_1 \sigma(I_{p+2})}_{\pi(J_{p+2})} = -\epsilon(\sigma) \epsilon(\pi) A^{\sigma'(I_p)}_{\pi'(I_p)} (D_2)^{\sigma(i_{p+2}), \sigma(i_{p+1})}_{\pi(j_{p+1}), \pi(j_{p+2})}
$$

$$
= \epsilon(\sigma) \epsilon(\pi) [AD_2]^{\sigma(I_{p+2})}_{\pi(J_{p+2})}.
$$

(2.180)

The last equality follows since $D_2$ is antisymmetric. Summing Eq. (2.180) over all $\sigma, \pi \in S_{p+2}$, we obtain $\vartheta_1 T = T$. The remaining claim is similarly dispatched.
We can now establish the following.

**Theorem 2.2.** For \( p \geq 2, \)

\[
\text{tr}_{p+2}(D_1^{\wedge p} \wedge D_2) = \frac{1}{(p+2)^2} \left[ Np(D_1^{\wedge (p-1)} \wedge D_2) - p(p-1)(D_1^{\wedge (p-2)} \wedge D_2 \wedge D_1^2) \right. \\
+ 2(N-1)(D_1^{\wedge (p+1)}) - 2p(D_1^{\wedge (p-1)} \wedge D_1D_2) \\
- 2p(D_1^{\wedge (p-1)} \wedge D_2D_1) \left. \right].
\]

(2.181)

**Example 2.3.** For \( p = 2, \) Theorem 2.2 reads

\[
\text{tr}_3(D_1 \wedge D_1 \wedge D_2) = \frac{1}{8} \left[ N(D_1 \wedge D_2) - (D_1^2 \wedge D_2) \right. \\
+ (N-1)(D_1 \wedge D_1 \wedge D_1) - 2D_1 \wedge (D_1D_2 + D_2D_1) \left. \right].
\]

(2.182)

**Example 2.4.** On the right side of the equality in Theorem 2.2, the symbol \( \wedge \) stands in for the antisymmetrizer \( \tilde{A}_{p+1} \otimes \tilde{A}_{p+1}, \) since \( \text{tr}_{p+2}(D_1^{\wedge p} \wedge D_2) \) is a \((p+1)\)-particle tensor. In fact the proof of Theorem 2.2 presented below does not require \( p > 1, \) but for \( p = 1 \) we must replace \( D_1D_2 \) and \( D_2D_1 \) by \((\tilde{A}_2 \otimes \tilde{A}_2)D_1D_2\) and \((\tilde{A}_2 \otimes \tilde{A}_2)D_2D_1, \) since in this case there is no place to put the \( \wedge \) symbol. Otherwise, the result stated in Theorem 2.2 holds for \( p = 1 \) and is given explicitly by

\[
\text{tr}_3(D_1 \wedge D_2) = \frac{1}{8} \left[ N(D_1 \wedge D_2) + 2(N-1)(D_1 \wedge D_1) \\
- 2(\tilde{A}_2 \otimes \tilde{A}_2)(D_1D_2 + D_2D_1) \right].
\]

(2.183)

**Proof of Theorem 2.2.** The tensor \((D_1^{\wedge p} \wedge D_2)\) has elements

\[
(D_1^{\wedge p} \wedge D_2)_{i_{p+2}j_{p+2}} = [(p+2)!]^{-2} \sum_{\sigma, \pi \in S_{p+2}} \epsilon(\sigma) \epsilon(\pi) [D_1D_2]^{(i_{p+2})}_{\pi(j_{p+2})}.
\]

(2.184)

Since \( i_{p+2} \) and \( j_{p+2} \) will become summation indices when we apply \( \text{tr}_{p+2}, \) we partition
the above sum according to where \( \sigma \) and \( \pi \) permute these two indices:

\[
(D_1^{\lambda p} \land D_2)_{i_{p+2}}^{j_{p+2}} = [(p + 2)!]^{-2} \left[ (\Upsilon_{2,2})_{i_{p+2}}^{j_{p+2}} + (\Upsilon_{1,2})_{i_{p+2}}^{j_{p+2}} + (\Upsilon_{2,1})_{i_{p+2}}^{j_{p+2}} + \sum_{r,s=1}^{p} (T_{rs})_{i_{p+2}}^{j_{p+2}} \right].
\]

(2.185)

The notation in this equation means the following. The quantity \((T_{rs})_{i_{p+2}}^{j_{p+2}}\), where \( r, s \leq p \), includes those permutations in Eq. (2.184) for which \( \sigma(r) = p + 2 = \pi(s) \), that is, the indices \( i_{p+2} \) and \( j_{p+2} \) both appear in \( D_1^{\lambda p} \). The remaining terms \( \Upsilon_{m,n} \) in Eq. (2.185) are defined as

\[
(\Upsilon_{2,2})_{i_{p+2}}^{j_{p+2}} \overset{\text{def}}{=} \sum_{r,s=1}^{p+2} \sum_{\sigma, \pi \in S_{p+2}}^\prime \epsilon(\sigma) \epsilon(\pi) \left[ D_1^{\lambda p} D_2 \right]_{\sigma(i_{p+2})}^{\pi(j_{p+2})},
\]

(2.186a)

\[
(\Upsilon_{1,2})_{i_{p+2}}^{j_{p+2}} \overset{\text{def}}{=} \sum_{r=1}^{p+2} \sum_{s=p+1}^{p+2} \sum_{\sigma, \pi \in S_{p+2}}^\prime \epsilon(\sigma) \epsilon(\pi) \left[ D_1^{\lambda p} D_2 \right]_{\sigma(i_{p+2})}^{\pi(j_{p+2})},
\]

(2.186b)

and

\[
(\Upsilon_{2,1})_{i_{p+2}}^{j_{p+2}} \overset{\text{def}}{=} \sum_{r=p+1}^{p+2} \sum_{s=1}^{p} \sum_{\sigma, \pi \in S_{p+2}}^\prime \epsilon(\sigma) \epsilon(\pi) \left[ D_1^{\lambda p} D_2 \right]_{\sigma(i_{p+2})}^{\pi(j_{p+2})}.
\]

(2.186c)

The following mnemonic may be helpful. The indices \( m, n \in \{1, 2\} \) in \( \Upsilon_{m,n} \) indicate that this quantity includes all permutations for which \( i_{p+2} \) is an index of \( D_m \) and \( j_{p+2} \) indexes \( D_n \).

The final term in Eq. (2.185), \( \sum_{r,s} (T_{rs})_{i_{p+2}}^{j_{p+2}} \), may itself be partitioned as in the proof of Theorem 2.1 [compare Eq. (2.161)]. Its one-particle trace is then evaluated by the same techniques used to establish Theorem 2.1, and the result is the first two terms in the statement of Theorem 2.2.

Now consider \((\Upsilon_{2,2})_{i_{p+2}}^{j_{p+2}}\). The sums over \( r \) and \( s \) in Eq. (2.186a) give rise to four terms that are related to one another by the transpositions \((i_{p+1}, i_{p+2})\) and \((j_{p+1}, j_{p+2})\).
According to our lemma (with $A = [D_1^{(p)}]$), all four terms are thus identical so we have

$$[p + 2]! - 2 (\text{tr}_{p+2} \ U_{2,2} I_{p+1}^{p+1})$$

$$= 4 [p + 2]! - 2 \sum_k \epsilon(\sigma) \epsilon(\pi) \left[ D_1^{(p)} \right]_{\pi(I_p)} \left[ D_2^{(p+1)} \right]_{\pi'(J_{p+1}), k}$$

$$= 2 [(N - 1)/(p + 1)!] \sum_{\sigma', \pi' \in S_{p+1}} \epsilon(\sigma') \epsilon(\pi') \left[ D_1^{(p+1)} \right]_{\pi'(I_{p+1})}$$

$$= 2 [(N - 1)/(p + 1)!] D_1^{(p+1)} .$$

In the second line we have used $\text{tr}_{2} D_2 = [(N - 1)/2] D_1$ and in the third line we introduce a factor of $[(p + 1)!]^2$ to account for the normalization of the resulting Grassmann product. Equation (2.187) is the third term in the formula for $\text{tr}_{p+2}(D_1^{(p)} \wedge D_2)$ that was given in the statement of Theorem 2.2.

By definition, $(\Upsilon_{1,2})_{I_{p+2}}$ is a sum over permutations $\sigma, \pi \in S_{p+2}$ such that $\sigma(i_r) = i_{p+2}$ for some $r \leq p$ and $\pi(j_s) = j_{p+2}$ for $s = p+1$ or $s = p+2$. For such permutations, consider the action of $\tau_{r}^{(p+2)} \in S_{p+2}$ as defined in Eq. (2.162):

$$\tau_{r}^{(p+2)} \sigma(I_{p+2}) = \left\{ \sigma'(I_{p+2}^{(p+1)}), \sigma(i_r) \right\}$$

$$= \left\{ \sigma'(I_{p+2}^{(p-1)}), \sigma'(i_{p+1}), \sigma'(i_{p+2}), \sigma(i_r) \right\}$$

(2.188)

for $r \leq p$. As usual, $\sigma' \in S_{p+1}$ orders $I_{p+2}^{(p+1)}$ in the same way that $\sigma$ orders the corresponding indices of $I_{p+2}$. We can write an analogous equation for $\pi \in S_{p+2}$, and it should be obvious by now that $\epsilon(\sigma) \epsilon(\pi) = \epsilon(\sigma') \epsilon(\pi')$. Applying the transposition $(p + 1, p + 2)$ of the last two indices to the aforementioned analogue of Eq. (2.188) one obtains

$$(p + 1, p + 2) \tau_{r}^{(p+2)} \pi(J_{p+2}) = \left\{ \pi'(J_{r}^{(p-1)}), \pi'(j_{p+1}), \pi(j_r), \pi'(j_{p+2}) \right\}$$

(2.189)

with $\pi' \in S_{p+1}$. 


We invoke the lemma once again to show, in Eq. (2.186b), that the two terms in the sum over \( s \) are identical. Using Eq. (2.188) and the analogous equation for \( \pi \in S_{p+2} \), we then have

\[
(\Upsilon_{1,2})^{I_{p+2}} = 2 \sum_{r=1}^{p} \sum'_{\sigma, \pi \in S_{p+2}} \epsilon(\sigma') \epsilon(\pi') \left[ D_1^{\wedge(p-1)} \right]_{\pi'} (D_2)^{\sigma'(i_{p+1}), \sigma'(j_{p+2})} (D_1)^{\sigma(j_r)}.
\]

(2.190)

It follows from Eq. (2.189) that

\[
(\text{tr}_{p+2} \Upsilon_{2,2})^{I_{p+2}} = -2 \sum_{r=1}^{p} \sum'_{\sigma, \pi \in S_{p+2}} \epsilon(\sigma') \epsilon(\pi') \left[ D_1^{\wedge(p-1)} \right]_{\pi'} (D_2)^{\sigma'(i_{p+1}), \sigma'(j_{p+2})} (D_1)^{\sigma(j_r)} \times (D_1 D_2)^{\sigma'(i_{p+1}), \sigma'(j_{p+2})}.
\]

(2.191)

The sum over \( r \) is performed trivially, since the summand is the same for each \( r \). Thus

\[
(\text{tr}_{p+2} \Upsilon_{2,2})^{I_{p+2}} = -2p(p + 1)! \left[ D_1^{\wedge(p-1)} \wedge (D_1 D_2) \right]^{I_{p+1}}.
\]

(2.192)

The factor of \((p + 1)!^2\) cancels the corresponding factor introduced by the Grassmann product. We pick up a factor of \((p + 2)!^{-2}\) from Eq. (2.185), and the net result is the fourth term in the statement of Theorem 2.2.

The final term in \( \text{tr}_{p+2}(D_1^{\wedge p} \wedge D_2) \) arises from \( \Upsilon_{2,1} \) in Eq. (2.185), but its evaluation is quite similar to that of \( \Upsilon_{1,2} \) and is omitted.

Many other variations on this theme are possible. We present two additional results without proof, as the proofs are quite similar to the preceding two.

**Theorem 2.3.** For any \( p \geq 2 \),

\[
\text{tr}_{2p} (D_2^{\wedge p}) = [(N - 1)/2p] \left( D_1 \wedge D_2^{\wedge(p-1)} \right) - (1/p)(A_{2p-1} \wedge A_{2p-1}) J^{(p)}.
\]

(2.193)
where $\mathcal{J}^{(p)}$ is the tensor with elements

$$
\left( \mathcal{J}^{(p)} \right)_{j_1, \ldots, j_{2p-1}}^{i_1, \ldots, i_{2p-1}} = \left( \sum_m (D_2)^{i_1, m}_{j_1, j_3} (D_2)^{i_2, i_3}_{j_2, m} \right) (D_2)^{i_4, i_5}_{j_4, j_5} \cdots (D_2)^{i_{2p-2}, i_{2p-1}}_{j_{2p-2}, j_{2p-1}}.
$$

(2.194)

In this paper we shall only require the case $p = 2$, and the elements of $\mathcal{J}^{(2)}$ are given explicitly in the next section. The final result of this subsection is a generalization of Theorem 2.2.

**Theorem 2.4.** For $p \geq 2$ and $k \geq 1$,

$$
\begin{align*}
\text{tr}_{p+2} \left[ (D_1^k)^{\wedge p} \wedge D_2 \right] &= \frac{1}{(p + 2)^2} \left\{ p \left( \text{tr} \ D_1^k \right) [D_1^{(p-1)} \wedge D_2] - p \left( p - 1 \right) \left[ [D_1^k]^{(p-2)} \wedge D_2 \wedge D_1^{2k} \right] \\
&\quad + 2(N - 1)[D_1 \wedge (D_1^k)^{\wedge p}] - 2p \left[ (D_1^k)^{\wedge (p-1)} \wedge (D_1^2 D_2 + D_2 D_1^k) \right] \right\}.
\end{align*}
$$

2.6.2 Generalized contraction theorems

In the previous subsection we obtained explicit, closed-form expressions for one-index partial traces of Grassmann products of $D_1$ and $D_2$. In order to derive certain coupled equations for the RDMCs (Subsection 4.2.2), we need to generalize these results to arbitrary antisymmetric tensors. Rather than proving a long list of contraction theorems, however, we instead present a simple counting algorithm by which many of the results can be derived. This essence of this technique is implicit in the proofs of Subsection 2.6.1.

Throughout this subsection, let $A_n \ (n > 1)$ denote an antisymmetric $n$-particle tensor. $A_1$ will indicate a simple matrix. For any tensor $T = (T_{j_1, \ldots, j_m}^{i_1, \ldots, i_m})$ we define partial trace tensors $\text{tr}_m T, \text{tr}_{m-1, m} T, \text{tr}_{m-2, m-1, m} T, \ldots, \text{tr}_{1, \ldots, m} T$ having elements

$$
[\text{tr}_{m-p, m-p+1, \ldots, m} T]^{i_1, \ldots, i_p}_{j_1, \ldots, j_p} = \sum_{k_{p+1}, \ldots, k_m} T_{j_1, \ldots, j_p, k_{p+1}, \ldots, k_m}^{i_1, \ldots, i_p, k_{p+1}, \ldots, k_m}.
$$

(2.195)
Of course $\text{tr}_{1,\ldots,m}T = \text{tr}T$ is a scalar. Introducing also $S = (T^{i_1,\ldots,i_n}_{j_1,\ldots,j_n})$, we will define a product tensor $TS$, even in the case where $m \neq n$. For $m \geq n$, this product has elements

$$(TS)^{i_1,\ldots,i_m}_{j_1,\ldots,j_m} = \sum_{k_1,\ldots,k_n} T^{i_1,\ldots,i_{m-n}}_{j_1,\ldots,j_m} k_{1,\ldots,n} s_{k_1,\ldots,k_n}^{i_{m-n+1},\ldots,i_m},$$

(2.196)

while for $m < n$ the appropriate tensor elements are

$$(TS)^{i_1,\ldots,i_n}_{j_1,\ldots,j_n} = \sum_{k_1,\ldots,k_m} T^{k_1,\ldots,k_m}_{j_{n-m+1},\ldots,j_n} s_{j_1,\ldots,j_{n-m},k_1,\ldots,k_m}^{i_1,\ldots,i_n}.$$  

(2.197)

When $n = m$, $TS$ is equivalent to the usual product of tensors. Finally, let us also define a $p$-index product $T \stackrel{\ast}{*} S$ according to

$$(T \stackrel{\ast}{*} S)^{i_1,\ldots,i_{m+p-n}}_{j_1,\ldots,j_{m+p-n}} \overset{\text{def}}{=} \sum_{k_1,\ldots,k_p} T^{i_1,\ldots,i_{m-n}}_{j_1,\ldots,j_m} k_{1,\ldots,n} s_{j_{m+1},\ldots,j_{m+p-n},k_1,\ldots,k_p}^{i_{m-n+1},\ldots,i_{m+p-n}}.$$  

(2.198)

When $p = n = m$, $T \stackrel{\ast}{*} S = TS$.

The counting procedure mentioned previously is best illustrated by example. Consider the one-index contraction $\text{tr}_{n+1}(A_1 \wedge A_n)$. Our method is based upon the identity

$$\text{tr}_{n+1} = \sum_k \delta_k,i_{n+1} \delta_k,j_{n+1},$$

(2.199)

which is valid if we use $i_1,\ldots,i_{n+1}$ for the upper indices of $A_1 \wedge A_n$ and $j_1,\ldots,j_{n+1}$ for the lower indices. By definition, $A_1 \wedge A_n$ involves a sum over permutations $\sigma, \pi \in S_{n+1}$. Letting $\sigma$ act on the upper indices and $\pi$ on the lower indices, we need to determine where each pair of permutations moves $i_{n+1}$ and $j_{n+1}$, enumerating all distinct possibilities. The coefficient of each distinct term in the contraction $\text{tr}_{n+1}(A_1 \wedge A_n)$ equals this number of possibilities divided by the normalization factor $(n + 1)!^2$ (from the Grassmann product). The contraction itself is obtained by applying $\text{tr}_{n+1}$ to one member of the class and antisymmetrizing the result on both its upper and
lower indices. There is one additional caveat: whenever the sum over \( k \) in Eq. (2.199) couples two terms in the Grassmann product (\( A_1 \) and \( A_n \) in this case), the resulting coefficient picks up a minus sign. A formal explanation for this is implicit in the detailed proofs of Subsection 2.6.1

Applying this procedure to evaluate \( \text{tr}_{n+1}(A_1 \wedge A_n) \), we make the following observations.

1. There are \( n!^2 \) permutations \((\sigma, \pi) \in S_{n+1} \otimes S_{n+1}\) such that \( \sigma(i_1) = i_{n+1} \) and \( \pi(j) = j_{n+1} \). Because \( A_n \) is antisymmetric, each gives rise to the same contraction. Using Eq. (2.199) for \( \text{tr}_{n+1} \), this contraction is seen to be \((\text{tr} A_1) A_n\), and the appropriate coefficient is \( n!^2/(n+1)!^2 = 1/(n+1)^2 \).

2. There are \( n \cdot n!^2 \) permutations \((\sigma, \pi) \in S_{n+1} \otimes S_{n+1}\) such that \( i_{n+1} \) is an upper index of \( A_n \) and \( j_{n+1} \) is the lower index of \( A_1 \). Application of \( \text{tr}_{n+1} \) to any such term yields \( A_n A_1 \). The actual contraction is obtained by antisymmetrizing this result, and its coefficient is \( -n \cdot n!^2/(n+1)!^2 = -n/(n+1)^2 \).

3. In light of Step 2 it is clear that we also obtain a term \(-n/(n+1)^2\) \((\bar{A}_n \otimes \bar{A}_1)(A_1 A_n)\). It is always the case that the matrix products produced by contraction will occur in both orders.

4. One possibility remains: there are \( n^2 n!^2 \) permutations \((\sigma, \pi) \) that leave both \( i_{n+1} \) and \( j_{n+1} \) on \( A_n \). The appropriate contraction is \( A_1 \wedge (\text{tr}_{n+1} A_{n+1}) \) and its coefficient is \( n^2/(n+1)^2 \).

As a check, note that the total number of permutations enumerated in Steps 1–4 is
\[(n + 1)!^2. \text{ We have just shown that} \]

\[
\text{tr}_{n+1}(A_1 \wedge A_n) = \left[ \frac{\text{tr} A_1}{(n + 1)^2} \right] A_n - \frac{n}{(n + 1)^2} (\hat{A}_n \otimes \hat{\hat{A}}_n)(A_1 A_n + A_n A_1)
\]

\[
+ \frac{n^2}{(n + 1)^2} [A_1 \wedge (\text{tr}_n A_n)]. \tag{2.200}
\]

A slightly more general result, which follows by the same procedure, is

**Theorem 2.5.** For any \(p \geq 2, m \geq 1, \text{ and } n \geq 2,\)

\[
\text{tr}_{p+n}[(A_1^m)^{\wedge p} \wedge A_n]
\]

\[
= \frac{1}{(p + 2)^2} \left\{ p (\text{tr} A_1^m) [A_1 \wedge (A_1^m)^{(p-1)}] - p (p - 1) [A_1 \wedge A_1^{2m} \wedge (A_1^m)^{(p-2)}]
\]

\[- p n [(A_1^m)^{(p-1)} \wedge (A_n A_1 + A_1 A_n)] + n^2 [(A_1^m)^{\wedge p} \wedge \text{tr}_n A_n] \right\}. \tag{2.201}
\]

For \(p = 1\) the result is still valid, with the understanding that \(A_1 \wedge (A_1^m)^{(p-1)} = A_1\) for \(p = 1.\)

The above contraction—and indeed, any one-index contraction of a Grassmann product—consists of two types of terms. There are “pure” contractions, in which the upper and lower summation indices \(k\) for the trace both index the same tensor, and *transvections*, in which the sum over \(k\) serves to couple two different tensors. Following Mazziotti\(^{14}\) (see also Harris\(^{165}\)), let us introduce operators \(\hat{\mathcal{P}}\) and \(\hat{\mathcal{T}}\) that pick out the pure contractions and the transvections, respectively. We write this as

\[
\text{tr}_{n+1}(A_1 \wedge A_n) = \hat{\mathcal{P}}(A_1 \wedge A_n) + \hat{\mathcal{T}}(A_1 \wedge A_n). \tag{2.202}
\]

Inspecting the result in Eq. (2.200), one finds that

\[
\hat{\mathcal{P}}(A_1 \wedge A_n) = \left[ \frac{\text{tr} A_1}{(n + 1)^2} \right] A_n + \frac{n^2}{(n + 1)^2} [A_1 \wedge \text{tr}_n A_n] \tag{2.203}
\]

while

\[
\hat{\mathcal{T}}(A_1 \wedge A_n) = -\frac{n}{(n + 1)^2} (\hat{A}_n \otimes \hat{\hat{A}}_n)(A_1 A_n + A_n A_1). \tag{2.204}
\]
As another example of our counting algorithm we evaluate the one-index contraction \( \text{tr}_{pn} A_n^{\wedge p} \). Focusing on the indices \( i_{pn} \) and \( j_{pn} \), the distinct possibilities are enumerated below.

1. \( i_{pn} \) and \( j_{pn} \) can index the same factor of \( A_n \). There are \( n \) index locations on each \( A_n \) where \( i_{pn} \) could go, and \( p \) factors of \( A_n \), for a total of \( pn \) possible index locations for \( i_{pn} \). Each corresponds to \((pn - 1)!\) different permutations \( \sigma \in S_{pn} \). Since \( j_{pn} \) must appear on the same factor of \( A_n \) as does \( i_{pn} \), there are \( n \) index locations for \( j_{pn} \) and \((pn - 1)!\) permutations \( \pi \in S_{pn} \) that will do the job. The total number of possibilities is thus \( np^2(pn - 1)!^2 \), which implies a coefficient of \( pn^2(pn - 1)!^2/(pn)!^2 = 1/p \). Applying \( \text{tr}_{pn} \) and antisymmetrizing, the contraction relevant to this case is seen to be \( A_n^{\wedge(p-1)} \wedge \text{tr}_n A_n \). This is the only possible pure contraction term, hence

\[
\Psi \left( A_n^{\wedge p} \right) = \frac{1}{p} \left[ A_n^{\wedge(p-1)} \wedge \text{tr}_n A_n \right].
\tag{2.205}
\]

2. There are \( n^2p(p-1) \) joint index locations for \( i_{pn} \) and \( j_{pn} \) on different factors of \( A_n \). Each corresponds to \((pn - 1)!^2 \) different permutations \( (\sigma, \pi) \in S_{pn} \otimes S_{pn} \), so the final coefficient is \( n^2p(p-1)(pn - 1)!^2/(pn)!^2 = (p-1)/p \). Applying \( \text{tr}_{pn} \) and antisymmetrizing, it is not difficult to see that

\[
\hat{\Psi} \left( A_n^{\wedge p} \right) = -\frac{p-1}{p} \left[ A_n^{\wedge(p-2)} \wedge (A_n \wedge A_n) \right]
\tag{2.206}
\]

where the minus sign arises because this is a transvection.

It is left to the reader to verify one final example:

\[
\hat{\Psi} \left( A_n \wedge A_m \right) = -\frac{nm}{(n+m)^2} (\tilde{A}_q \otimes \tilde{A}_q) \left( A_n \wedge A_m + A_m \wedge A_n \right),
\tag{2.207}
\]
where \( q = \max(n,m) \). (\( A_m \ast A_n \) are \( q \)-particle tensors.)

The counting procedure introduced in this subsection can be applied to evaluate a wide range of contractions. In particular, it provides a relatively simple means to derive the theorems proved in Subsection 2.6.1. Other useful contraction formulas obtained by this procedure are compiled in Appendix B. We will use these results extensively—often implicitly—in the next chapter.

2.6.3 Grassmann product ansatz

Now we return to the Grassmann product ansatz \( D_4^G \) defined in Eq. (2.148), and derive formulas for its contractions \( 4 \downarrow D_4^G \), \( 3 \downarrow D_4^G \), and \( 2 \downarrow D_4^G \). These contraction formulas demonstrate that a Grassmann product form for \( D_4 \) is inconsistent with a Grassmann product form for the lower-order RDMs. With the benefit of a well-formulated theory of the density matrix cumulants, this conclusion is obvious, since the \( p \)-RDMC is precisely that part of \( D_p \) that cannot be written as a Grassmann product of lower-order RDMs. Nevertheless, it was speculated at the time\(^{18} \) that perhaps the parameters \( a, b, \) and \( c \) appearing in \( D_4^G \) could be chosen so as to incorporate, in an indirect or semi-empirical way, certain correlation effects properly attributable to the cumulants, much as an uncorrelated Hückel calculation can recover certain correlation effects if the parameters are chosen appropriately.

Applying the theorems in the previous section to \( D_4^G \), one readily obtains

\[
\frac{4}{3} \downarrow D_4^G = \frac{1}{2} \left\{ [2a(N - 1) + bN] (D_1 \wedge D_2) - b [(D_1)^2 \wedge D_2] \\
+ [b(N - 1) + 2cN] (D_1 \wedge D_1 \wedge D_1) - 6c [D_1 \wedge D_1 \wedge (D_1)^2] - 2b [D_1 \wedge (D_1 D_2 + D_2 D_1)] - 4a (\vec{A}_3 \otimes \vec{A}_3) \mathcal{J}^{(2)} \right\} / (N - 3) .
\]

\( \mathcal{J}^{(2)} \) in this equation is the tensor defined in Theorem 2.3. If \( (D_2)_r^t = (D_2)_{rs}^{rs} \) denotes
the matrix comprising the \((i, r)\)th block of \(D_2\), then \((J^{(2)})_{rst}^{ijk}\) is the \((k, t)\)th element of the product matrix \((D_2)^i_r \ (D_2)^j_s\),

\[
(J^{(2)})_{rst}^{ijk} = [(D_2)^i_r \ (D_2)^j_s]^k_t = \sum_m (D_2)^{im}_{rl} \ (D_2)^{jk}_{sm}.
\]  

(2.209)

It will prove convenient to introduce some additional notation. The tensor \(J\),

\[
J_{rs}^{ij} \overset{\text{def}}{=} \text{tr} [(D_2)^i_r \ (D_2)^j_s] = J_{sr}^{ji},
\]

(2.210)

has elements that are traces of matrix products of the \((D_2)^i_j\) blocks introduced above. For tensors \(T = (T_{j_1 \ldots j_m}^{i_1 \ldots i_n})\) and \(S = (S_{j_1 \ldots j_m}^{i_1 \ldots i_n})\), define a tensor \(T \ast S\) with elements [compare Eq. (2.198)]

\[
(T \ast S)_{j_1 \ldots j_n}^{i_1 \ldots i_m} \overset{\text{def}}{=} \sum_{k_1 \ldots k_m} T_{j_1 \ldots j_m}^{k_1 \ldots k_m} S_{k_1 \ldots k_m}^{j_1 \ldots j_n}
\]

(2.211a)

if \(m < n\) and

\[
(T \ast S)_{j_1 \ldots j_m}^{i_1 \ldots i_n} \overset{\text{def}}{=} \sum_{k_1 \ldots k_n} T_{j_1 \ldots j_m, j_{m+1} \ldots j_n}^{k_1 \ldots k_m} S_{k_1 \ldots k_n}^{i_1 \ldots i_n}
\]

(2.211b)

if \(m \geq n\). If \(n = m\), \(T \ast S = TS\). Using this notation along with the results of the previous subsection, one can show that

\[
\frac{4}{3} D_4^Q = \left\{ \frac{1}{6} \left[ 2aN(N - 1) + b[N^2 - \text{tr} D_4^2] \right] D_2 + \frac{2}{3} a D_2^3 - \frac{4}{3} b \left[ D_1 \wedge \text{tr}_2(D_1 D_2) \right] \\
- 2 \left[ b(N - 1) + 2cN \right] \left[ D_1 \wedge (D_1)^2 \right] + \frac{2}{3} b \left[ D_1^2 D_2 + D_2 D_1^2 \right] \\
+ \frac{4}{3} a \left[ 4a(N - 1)^2 + 5bN(N - 1) + 6c(N^2 - \text{tr} D_4^2) \right] \left[ D_1 \wedge D_1 \right] \\
+ 4c \left( D_1 \wedge D_1^3 \right) + 2c(D_1^2 \wedge D_1^2) \\
+ \left( \tilde{A}_2 \otimes \tilde{A}_2 \right) \left( \frac{4}{3} b D_1 D_2 D_1 - \frac{2}{3} \left[ 2a(N - 1) + bN \right] (D_1 D_2 + D_2 D_1) \\
+ \frac{1}{3} b \left[ D_1 \ast D_1 D_2 + D_2 D_1 \ast D_1 + \frac{8}{3} a J \right] \right) / [(N - 3)(N - 2)]. \right\}
\]

(2.212)
We contract once again to obtain

\[ 4 \downarrow \left\langle \begin{array}{c}
4 \\
1
\end{array} \right| \mathbf{D}_4^2 = \left\{ 4a \text{tr}_2(\mathbf{D}_2)^2 + \left\{ aN(N-1)^2 + b(N-1)(N^2 - \frac{3}{2} \text{tr}(\mathbf{D}_1)^2) \\
+ c\left[N^3 - 3N \text{tr}(\mathbf{D}_1)^3 + 2 \text{tr}(\mathbf{D}_1)^3\right]\right\} \mathbf{D}_1 \\
- \left\{ 2a(N-1)^2 + \frac{15}{6}bN(N-1) + 3c[N^2 - \text{tr}\mathbf{D}_1^2]\right\} \mathbf{D}_2^2 \\
+ 3\left\{ b(N-1) + 2cN\right\} (\mathbf{D}_1)^3 - 6c\mathbf{D}_1^4 - \frac{2}{3}b\mathbf{Y} \\
- 2\left\{ 2a(N-1) + bN\right\} \text{tr}_2(\mathbf{D}_1\mathbf{D}_2) + 2b \text{tr}_2[(\mathbf{D}_1)^2\mathbf{D}_2] \\
+ \frac{4}{3}b\left[ \mathbf{D}_1 \cdot (\text{tr}_2 \mathbf{D}_1\mathbf{D}_2) + (\text{tr}_2 \mathbf{D}_1\mathbf{D}_2) \cdot \mathbf{D}_1 \right] \right\}/[(N-3)(N-2)(N-1)]. \tag{2.213} \]

The tensor \( \mathbf{Y} \) in this equation has elements

\[ \mathbf{Y}_r^i \overset{\text{def}}{=} \sum_k \left[ (\mathbf{D}_1\mathbf{D}_2\mathbf{D}_1)^{ik}_{kr} + (\mathbf{D}_1\mathbf{D}_2\mathbf{D}_1)^{ki}_{rk} \right]. \tag{2.214} \]

Although the reconstruction functional \( \mathbf{D}_4^2 \) ostensibly contains three adjustable parameters, if one wants the 4-RDM to exhibit proper normalization then only two of these are independent. The normalization is obtained as a function of \( a, b, \) and \( c \) by contracting Eq. (2.213) once more:

\[ \text{tr} \mathbf{D}_4^2 = \frac{a}{24} \left\{ N^2(N-1)^2 - 4(N-1)^2 \text{tr}\mathbf{D}_1^2 + 4 \text{tr}\mathbf{D}_2^2 \right\} \\
+ \frac{c}{24} \left[ N^4 - 6N^2 \text{tr}\mathbf{D}_1^2 + 8N \text{tr}\mathbf{D}_1^3 + 3(\text{tr}\mathbf{D}_1^2)^2 - 6 \text{tr}\mathbf{D}_1^4 \right] \\
+ \frac{b}{24} \left[ N^3(N-1) - 5N(N-1) \text{tr}\mathbf{D}_1^2 + 4(N-1) \text{tr}\mathbf{D}_1^3 \\
+ 4 \sum_{klmn} (\mathbf{D}_1)_m^l (\mathbf{D}_1)_l^k (\mathbf{D}_2)_l^m (\mathbf{D}_2)_m^k \right\]. \tag{2.215} \]

Setting \( \text{tr} \mathbf{D}_4^2 = \binom{N}{4} \) fixes one of the parameters in terms of the other two.

Up to this point the spin-orbital basis \( \{ \phi_k \} \) has been orthonormal but otherwise arbitrary. We now show how the above expressions simplify in the NSO basis. The
expressions for the reconstructed 1- and 2-RDMs, $\frac{1}{2} \downarrow D_4^G$ and $\frac{1}{4} \downarrow D_4^G$, are especially simple in this basis, specifically,

\[
\left[ (\hat{A}_2 \otimes \hat{A}_2) D_1 D_2 D_1 \right]_{rs}^{ij} \overset{\text{NSO}}{=} \frac{1}{4} (n_i + n_j)(n_r + n_s)(D_2)_{rs}^{ij}, \quad (2.216)
\]

\[
\left\{ (\hat{A}_2 \otimes \hat{A}_2) \left[ D_1^m D_2 + D_2 D_1^m \right] \right\}_{rs}^{ij} \overset{\text{NSO}}{=} \frac{1}{2} (n_i^m + n_j^m + n_r^m + n_s^m)(D_2)_{rs}^{ij}, \quad (2.217)
\]

\[
\left[ (\hat{A}_2 \otimes \hat{A}_2) \left( D_1 \ast D_1 D_2 + D_2 D_1 \ast D_1 \right) \right]_{rs}^{ij} \overset{\text{NSO}}{=} (n_i n_j + n_r n_s)(D_2)_{rs}^{ij}, \quad (2.218)
\]

and

\[
\left[ D_1 \wedge \text{tr}_2(D_1 D_2) \right]_{rs}^{ij} \overset{\text{NSO}}{=} \frac{1}{4} n_r \sum_k n_k \left[ \delta_{jr}(D_2)^{jk}_{sk} - \delta_{jr}(D_2)^{ik}_{sk} \right]
\]

\[
+ \frac{1}{4} n_s \sum_k n_k \left[ \delta_{js}(D_2)^{ik}_{rk} - \delta_{js}(D_2)^{jk}_{rk} \right]. \quad (2.219)
\]

Expressions for the relevant components of $\frac{1}{4} \downarrow D_4^G$ follow readily from these equations. One finds that in general the reconstructed 1-RDM may possess nonvanishing off-diagonal elements, even when the input $D_1 = \frac{2}{3} \downarrow D_2$ is diagonal.

### 2.7 References


[95] R. E. Brown and V. H. Smith, Jr., “Supergeminals and the applicability of the strongly orthogonal geminal ansatz: Be($^1S$) and C($^1S$)”, in Reduced Density Matrices with Applications to Physical and Chemical Systems II, Vol. 40 of Queen’s Papers on Pure and Applied Mathematics, edited by R. Erdahl (Queen’s University Press, Kingston, ON, 1974), pg. 75.


122


Chapter 3

Equations to determine the two-electron density matrix

3.1 Introduction

We have thus far left open the question of how $D_2$ may be determined, in the absence of an electronic wavefunction, except by exceedingly expensive constrained variational methods (Section 1.2). In this chapter we introduce a non-variational methodology for direct determination of the 2-RDM based on the self-consistent solution of a “contracted” Schrödinger equation.

Suppose that $\hat{W}$ [Eq. (2.1)] is any symmetric combination of one- and two-electron operators, for an $N$-electron system. Then for any $p \geq 2$, we will demonstrate in Section 3.2 that the Hilbert-space eigenvalue equation $\hat{W}\ket{\Psi} = w\ket{\Psi}$ is equivalent to a reduced eigenvalue equation involving only the $p$, $(p + 1)$-, and $(p + 2)$-RDMs. (An equation for $p = 1$ also follows from the $N$-electron eigenvalue equation, but is not equivalent to it.) This establishes a hierarchy of coupled equations for the
RDMs that is reminiscent of the coupled equations for the reduced particle distribution functions that are obtained in several other contexts, including the Bogoliubov-Born-Kirkwood-Green-Yvonne (BBKGY) hierarchy\(^1\) for the classical many-particle reduced distribution functions, as well as its multi-time, quantum-mechanical generalization, the so-called\(^2\) Martin-Schwinger-Kadanoff hierarchy\(^2-^8\) for the many-particle Green’s functions. Viewed through this analogy, it is perhaps not surprising that a cumulant formalism and decoupling approximations for the many-particle RDMs play a major role, though we defer these topics to Chapter 4.

The coupled RDM equations differ from these other hierarchies in one crucial respect: in the RDM case, each member \((p \geq 2)\) of the hierarchy is equivalent, in a necessary and sufficient sense, to all of the others, and to the original \(N\)-electron, Hilbert-space eigenvalue equation. Thus for \(p = 2\) and \(\hat{W}\) an electronic Hamiltonian, one has an equation in which the basic quantities are \(\hat{D}_2, \hat{D}_3,\) and \(\hat{D}_4\), rather than \(|\Psi\rangle\), yet is equivalent to the electronic Schrödinger equation (up to \(N\)-representability, as discussed below). This coupled RDM equation was originally termed the density equation,\(^9,^11\) though the term contracted Schrödinger equation (CSE)\(^12\) is presently more popular and conveys the fact that the CSE is derived from the \(N\)-electron Schrödinger equation simply through a partial trace over coordinates \(x_3, \ldots, x_N\). Derivations of the CSE were published nearly simultaneously in 1976\(^*\) by Cohen and Frishberg\(^13\) and by Nakatsuji,\(^9\) though only Nakatsuji proved the converse implication that establishes the equivalence of the Schrödinger equation—an equation in an \(N\)-electron vector space—and the CSE, which is an equation in a two-electron matrix or operator space.

In Section 3.2, we derive the CSE and demonstrate the aforementioned equivalence, which is sometimes called Nakatsuji’s Theorem.  

Several outstanding problems have impeded progress on the CSE as a viable alternative to the Schrödinger equation. The first is the old boundary-condition nemesi-sis, the $N$-representability problem, which appears in a somewhat different guise here. Namely, in order to solve the CSE for $\hat{D}_2$, we must figure out a way to reconstruct the 3- and 4-RDMs from a given 2-RDM. These decoupling approximations for $\hat{D}_3$ and $\hat{D}_4$ are a manifestation of the $N$-representability constraints, in that the exact reconstructions $\hat{D}_3[\hat{D}_2]$ and $\hat{D}_4[\hat{D}_2]$ (whose existence is established by certain Hohenberg-Kohn-like theorems) map an $N$-representable, non-degenerate, ground-state 2-RDM to a unique, $N$-representable 3- or 4-RDM. Accurate density matrix reconstruction is one of the main impediments to the accuracy of electronic structure calculations based on the CSE, but we defer the main part of our discussion on reconstruction until the next chapter (but see Subsection 3.2.2).

The reconstruction and $N$-representability problems aside, the CSE as it stands is still somewhat unsatisfactory in that it contains terms that are not extensive. As a result, the direct solution of this equation does not represent a size-consistent methodology, and will generally fail to preserve extensivity of the electronic energy. Because the CSE is equivalent to the Schrödinger equation, however, one knows in advance that the offending terms in the CSE must cancel exactly, at the level of formal manipulations at least. This exact cancellation will not be preserved in any approximate computational scheme, unless one utilizes explicitly connected working equations. Hence it is of interest to eliminate all unconnected (non-extensive) terms a priori, so as to obtain manifestly extensive (and size-consistent) working equations. This is the topic
of Section 3.3, whose results were first published in Ref. 17.

3.2 The contracted Schrödinger equation

Suppose that the electronic wavefunction $|\Psi\rangle$ satisfies the eigenvalue equation

$$\hat{W}|\Psi\rangle = w|\Psi\rangle,$$  \hspace{1cm} (3.1)

in which $\hat{W}$, defined in Eq. (2.1), has the form of an arbitrary two-electron observable. That is, $\hat{W}$ consists of a symmetric sum of one-electron operators $\hat{h}$ and two-electron operators $\hat{g}$. Equation (3.1) exists in an $N$-electron vector space (Hilbert space); in this section we will consider equivalent, density-matrix formulations of this equation that exist in $p$-electron matrix (or operator) spaces, where $p \leq N$. Originally,\textsuperscript{9} the $p$th-order CSE was derived from the position-space Schrödinger equation in its density-matrix form

$$\Psi^*(t',...,N') \hat{W} \Psi(t,....,N) = w \Psi^*(t',...,N') \Psi(t,....,N),$$ \hspace{1cm} (3.2)

by applying the operator $\text{tr}_{p+1,...,N}$ to both sides. Following some clever relabeling of dummy integration variables, one obtains an equation in $D_p$, $D_{p+1}$, and $D_{p+2}$. The reader is referred to Ref. 9 for this derivation. In Subsection 3.2.1 we present an alternative derivation of the CSE, in second quantization.

3.2.1 Reduced eigenvalue equations

Introducing an arbitrary basis $\{\phi_k\}$ of orthonormal spin orbitals, with associated creation operators $\{\hat{a}_k^\dagger\}$, it follows from Eq. (3.1) that

$$\langle \Psi| \hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_i \hat{a}_k (\hat{W}_{12} - w)|\Psi\rangle = 0$$ \hspace{1cm} (3.3)
for each $i$, $j$, $k$, and $l$. $\tilde{W}_{12}$ is the two-electron reduced operator corresponding to $\tilde{W}$, which was defined in Eq. (2.4). Substituting the second-quantized expansion of $\tilde{W}_{12}$ [Eq. (2.32)] into Eq. (3.3) yields

$$
\sum_{pqmn} W_{mn;pq} \langle \Psi | \hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_l \hat{a}_k \hat{a}_m \hat{a}_n \hat{a}_o | \Psi \rangle = 2w \langle \Psi | \hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_l \hat{a}_k | \Psi \rangle
$$

(3.4)

$$
= 4w D_{ij;kl} .
$$

To express the left side of this equation in terms of density matrix elements, one must place the operator string $\hat{a}_p \cdots \hat{a}_k$ in normal order. This proceeds in straightforward fashion by repeated application of the fermion anticommutation identity [Eq. (2.42)]. In normal order, the operator string $\hat{a}_p \cdots \hat{a}_k$ consists of seven terms. Inserting this into Eq. (3.4) and relabelling some summation indices, one obtains

$$
ED_{ij;kl} = \sum_{mn} W_{mn;kl} D_{ij;mn} + 3 \sum_{mnp} \left[ W_{np;lm} D_{ij;mkn} - W_{np;km} D_{ij;lmnp} \right]
$$

+ $6 \sum_{mnpq} W_{pq;mn} D_{ijmn;klpq}$ .

(3.5)

Since $W$, $D_2$, $D_3$, and $D_4$ each possess some permutational symmetry, various permutations of the indices are possible.

Equation 3.5 is the CSE, though it can be placed in some alternative forms. For instance, each term on the right side of Eq. (3.5) may be rewritten as an element of one of the product tensors

$$
(WD_2)_{ij}^{rs} = \sum_{pq} W_{pq;rs} D_{ij;pq} ,
$$

(3.6a)

$$
(WD_3)_{ijk}^{rst} = \sum_{pq} W_{pq;st} D_{ijk;rpq} ,
$$

(3.6b)

or

$$
(WD_4)_{ijkl}^{rstu} = \sum_{pq} W_{pq;tu} D_{ijkl;rsqp} .
$$

(3.6c)
This yields

\[
ED_{ijkl} = (WD_2)_{ijkl} + 3 \sum_m [(WD_3)^{ilm}_{klm} - (WD_3)^{jlm}_{ilk}] + 6 \sum_{mn} (WD_4)^{imn}_{klmn} \tag{3.7}
\]

Having moved all upper- and lower summation indices to the far right, one can formally perform the summations to obtain

\[
ED_{ijkl} = (WD_2)_{ijkl} + 3 \{ [tr_3(WD_3)]_{ijkl} - [tr_3(WD_3)]_{ilk} \} + 6 [tr_{3,4}(WD_4)]_{ijkl} \tag{3.8}
\]

We emphasize that \(tr_3(WD_3)\) and \(tr_{3,4}(WD_4)\) are tensors. Although \(tr_{3,4}(WD_4) = tr_{3,4}(D_4W)\), in general \(tr_3(WD_3) \neq tr_3(D_3W)\) since \(tr_3\) in this case is a sum over one index, while the partial product involves a sum over two.

Let us examine Eq. (3.8). The temptation is great to combine the two middle terms using some antisymmetry property. While \(WD_3\) is antisymmetric in its upper indices (which come solely from \(D_3\)), this tensor exhibits no special symmetry in its lower indices, which consist of some indices taken from \(W\) and some taken from \(D_3\).* The tensor \(tr_3(WD_3)\) inherits antisymmetry in its upper indices, which we can exploit in order to rewrite the pair of 3-RDM terms in Eq. (3.8):

\[
[tr_3(WD_3)]_{ijkl} - [tr_3(WD_3)]_{iljk} = \frac{1}{2} \left\{ [tr_3(WD_3)]_{ijkl} - [tr_3(WD_3)]_{ilk} \right\} + [tr_3(WD_3)]_{ijkl} \tag{3.9}
\]

This allows us to rewrite Eq. (3.8) in its most compact form,

\[
wD_2 = WD_2 + 6 \left( \tilde{A}_2 \otimes \tilde{A}_2 \right) tr_3(WD_3) + 6 tr_{3,4}(WD_4) \tag{3.10}
\]

Each two-particle tensor in this equation is clearly antisymmetric in both its upper and its lower indices.

*Similarly, \(D_3W\) is antisymmetric in its lower but not its upper indices.
Although we consider only fermion density matrices here, by employing the boson commutation rule $\hat{a}_i^\dagger \hat{a}_j - \hat{a}_j \hat{a}_i^\dagger = \delta_{ij}$, one can derive an analogue of Eq. (3.10) for symmetric boson RDMs. The only difference is that the operators $\overrightarrow{A}_2$ and $\overleftarrow{A}_2$ in the bosonic equation are replaced by *symmetrizing* projection operators. [Equivalently, all the signs in Eq. (3.5) are plus for bosons.]

The formal structure of Eq. (3.10) does not depend upon the identity of the operator $\hat{W}$, and we call Eq. (3.10) a two-electron *reduced eigenvalue equation*\textsuperscript{17,18} corresponding to the Hilbert-space eigenvalue equation $\hat{W}|\Psi\rangle = w|\Psi\rangle$. Clearly the most important example is the reduced electronic Schrödinger equation, that is, the CSE. The only other reduced eigenvalue equation to garner any attention is the reduced equation for total spin angular momentum, $\hat{W} = \hat{S}^2$, which has been employed as a boundary condition for the solution of the CSE.\textsuperscript{19} In this work we refer to Eq. (3.10) as CSE(2), the second-order CSE,\textsuperscript{17} for reasons that will become apparent.

The energy $E$ and the RDMs $D_n$ in CSE(2) are the exact quantities corresponding to the pure state $|\Psi\rangle$. For statistical ensembles (convex linear combinations $\sum_i c_i |\Psi_i\rangle \langle \Psi_i|)$, the RDMs arising from each pure-state component $|\Psi_i\rangle \langle \Psi_i|$ satisfy a CSE of their own, though generally the RDMs for the ensemble do not. An exception is the case when all of the $|\Psi_i\rangle$ are degenerate. Any ensemble of degenerate states satisfies a CSE in which $D_2$, $D_3$, and $D_4$ are RDMs for the ensemble and $E$ is the energy of each $|\Psi_i\rangle$. Degenerate ensembles such as this arise in a spin-free formulation of the RDM cumulants.\textsuperscript{20} We will return to the possibility of ensemble solutions in Section 3.3, but for now we consider only pure states.

Although CSE(2) is commonly called *the* CSE, as our notation indicates this equation is but a special case of a more general $p$th-order CSE, CSE($p$), that couples
the $p$, $(p + 1)$-, and $(p + 2)$-RDMs. [Sometimes \text{CSE}(p) is called the $(p, p + 2)$-\text{CSE}, in order to emphasize this fact.\textsuperscript{21,22}] Starting from
\begin{equation}
\langle \Psi | \hat{a}_{i_1}^\dagger \cdots \hat{a}_{i_p}^\dagger \hat{a}_{j_p} \cdots \hat{a}_{j_1} (\hat{W}_{12} - w) | \Psi \rangle = 0 \tag{3.11}
\end{equation}
in place of Eq. (3.3), \text{CSE}(p) is derived by exactly the same sorts of manipulations used above to derive \text{CSE}(2). As an example, \text{CSE}(3) is found to be
\begin{equation}
w \mathbf{D}_3 = 4 \left( \mathbf{A}_3 \otimes \mathbf{A}_3 \right) (\mathbf{W}\mathbf{D}_3) + 12 \left( \mathbf{A}_3 \otimes \mathbf{A}_3 \right) [\text{tr}_4(\mathbf{WD}_4)] + 10 \text{tr}_{4,5}(\mathbf{WD}_5) \tag{3.12}
\end{equation}
Given Eq. (3.11) along with the fact that $N^{-1} \sum_i \hat{a}_i^\dagger \hat{a}_i = \mathbf{1}$, it is reasonably obvious that \text{CSE}(p - 1) is merely a one-index partial trace of \text{CSE}(p). Thus Eq. (3.12) can be contracted once to obtain \text{CSE}(2), or twice to obtain
\begin{equation}
w \mathbf{D}_1 = 2 \text{tr}_2(\mathbf{WD}_2) + 3 \text{tr}_2(\mathbf{WD}_3) \tag{3.13}
\end{equation}
which is \text{CSE}(1). It is not clear that anything useful is obtained from \text{CSE}(1), though Mazziotti\textsuperscript{22,23} has used this relationship to augment a 2-RDM variational method. There has also been some interest\textsuperscript{21,24} in using \text{CSE}(3) as the basic working equation, rather than \text{CSE}(2), as this allows one to make approximations at the five-particle level, leaving intact all of the cumulants through order four, that is, all of the cumulants appearing in \text{CSE}(2). However, the results obtained with \text{CSE}(3) are not appreciably superior to those obtained from \text{CSE}(2), while the latter is a much simpler equation.

At present, then, it appears that only \text{CSE}(2) is really useful. This is perhaps not surprising given that \text{CSE}(2) is equivalent,\textsuperscript{9} in a necessary and sufficient sense, to \text{CSE}(p) for $2 < p \leq N$, provided that the $N$-representability conditions are satisfied. Here \text{CSE}(N) denotes the uncontracted, Hilbert-space Schrödinger equation. We have just shown that \text{CSE}(N) implies \text{CSE}(2); to demonstrate the equivalence of these two
equations (a result that is sometimes called Nakatsuji’s Theorem\textsuperscript{9,10}) we must prove the converse. This we now do, following the second-quantized proof given by Mazziotti.\textsuperscript{25}

Multiplying CSE(2) in Eq. (3.4) by $\frac{1}{2} W_{kl;ij}$ and then summing over $i$, $j$, $k$, and $l$, one obtains

$$\langle \Psi \left| \left( \frac{1}{2} \sum_{ijkl} W_{kl;ij} \hat{a}^\dagger_i \hat{a}^\dagger_j \hat{a}_l \hat{a}_k \right) \left( \frac{1}{2} \sum_{mnpq} W_{mn;pq} \hat{a}^\dagger_p \hat{a}^\dagger_q \hat{a}_m \hat{a}_n \right) \right| \Psi \rangle$$

$$= w \langle \Psi \left| \frac{1}{2} \sum_{mnpq} W_{mn;pq} \hat{a}^\dagger_p \hat{a}^\dagger_q \hat{a}_m \hat{a}_n \right| \Psi \rangle ,$$

or in other words

$$\langle \hat{W}^2_{12} \rangle = w \langle \hat{W} \rangle .$$  \hspace{1cm} (3.15)

We could just as well write this as $\langle \hat{W}^2 \rangle = w \langle \hat{W} \rangle$. Since $w = \langle \hat{W} \rangle$, we have thus demonstrated that CSE(2) implies the so-called zero-dispersion condition $\langle \hat{W}^2 \rangle - \langle \hat{W} \rangle^2 = 0$. For any self-adjoint operator $\hat{W}$, this is equivalent\textsuperscript{*} to the condition that $\hat{W} | \Psi \rangle = w | \Psi \rangle$. Since $\hat{W}^2$ is a four-electron operator, the 4-RDM is needed in order to calculate $\langle \hat{W}^2 \rangle$. which explains why $\hat{D}_4$ is needed in order to determine $\hat{D}_2$, even though the latter suffices to determine $w$. Although CSE(1) is a necessary consequence of the Schrödinger equation, it contains RDMs only up to $\hat{D}_3$ and is therefore insufficient to guarantee that the stationarity condition $\langle \hat{W}^2 \rangle - \langle \hat{W} \rangle^2 = 0$ is satisfied. This explains why CSE(1) is not equivalent to the higher-order CSE($p$).

In proving that CSE(2) implies CSE($N$), we have glossed over one crucial point, namely, that the RDMs in CSE(2) must be $N$-representable in order to obtain the equivalence of Nakatsuji’s Theorem. The need for $N$-representability constraints arises in using the expectation value, Eq. (2.29), to define the RDM elements, which

\textsuperscript{*}For a proof, see page 360 of Ref. 26.
tacitly assumes the existence of a wavefunction $\Psi$. Absent a complete set of \(N\)-representability constraints, CSE(2) has spurious, non-\(N\)-representable solutions.\(^{27}\) This is discussed more thoroughly in Subsection 3.2.2.

Although the tensor formulation is ultimately what is needed for computational purposes, for certain purposes it is advantageous to express CSE(\(p\)) in a kernel or operator form. To this end, let us introduce the \(p\)-electron energy density matrix\(^9,28\)

\[
\hat{\Omega}_p(1,\ldots,p;1',\ldots,p') = \left[ \sum_{j=1}^{p} \hat{h}(j) + (1 - \delta_{p,1}) \sum_{j<k} \hat{g}(j,k) - \cdots \right] \hat{D}_p \\
+ (p+1) \text{tr}_{p+1} \left\{ \left[ \hat{h}(p+1) + \sum_{j=1}^{p} \hat{g}(j,p+1) \right] \hat{D}_{p+1} \right\} \\
+ \left( \frac{p+2}{2} \right) \text{tr}_{p+1,p+2} \left\{ \hat{g}(p+1,p+2) \hat{D}_{p+2} \right\}.
\]  

Equation (3.16)

Then CSE(\(p\)) may be written as the condition that \(\hat{\Omega}_p\) vanishes,

\[
\hat{\Omega}_p(1,\ldots,p;1',\ldots,p') \equiv 0.
\]  

Equation (3.17)

For \(p = 2\), the above equation has the form

\[
w D_{2,(1,2;1',2')} = \left[ \hat{h}(1) + \hat{h}(2) + \hat{g}(1,2) \right] D_{2,(1,2;1,2)} \\
+ 3 \int dx_3 \left\{ \left[ \hat{h}(3) + \hat{g}(1,3) + \hat{g}(2,3) \right] D_{3,(1,2,3;1',2',3')} \right\}_{3' = 3} \\
+ 6 \int dx_3 \int dx_4 \left\{ \hat{g}(3,4) D_{4,(1,2,3,4;1',2',3',4')} \right\}_{4' = 3}
\]  

Equation (3.18)

in the position representation, while CSE(1) is given by

\[
w D_{1,(1;1')} = \hat{h}(1) D_{1,(1,1')} + 2 \int dx_2 \left\{ \left[ \hat{h}(2) + \hat{g}(1,2) \right] D_{2,(1,2;1',2')} \right\} \\
+ 3 \int dx_2 \int dx_3 \left\{ \hat{g}(2,3) D_{3,(1,2,3;1',2',3')} \right\}_{3' = 3}
\]  

Equation (3.19)

in the position representation, while CSE(1) is given by

Equation (3.18) could also be obtained\(^{18}\) from its tensor analogue, Eq. (3.10), by replacing tensor products with integral kernel products, and extending the definition
of $\hat{W}_{12}$ in Eq. (2.4) to a four-index integral kernel

$$
\hat{W}_{12}(1,2;1',x') = \left[ \hat{g}(1,2) + \frac{\hat{h}(1) + \hat{h}(2)}{N - 1} \right] \delta(x_1 - x'_1) \delta(x_2 - x'_2).
$$

(3.20)

### 3.2.2 Density matrix reconstruction functionals

CSE(2) is the most tractable equation in the CSE hierarchy that is yet equivalent to the Schrödinger equation, and it is CSE(2) that we wish to solve. From one point of view, CSE(2) constitutes a closed equation for $D_4$, since the 2- and 3-RDMs are known functionals of the 4-RDM. In fact, Nakatsuji’s theorem implies that for a fixed energy, CSE(2) possesses a unique, $N$-representable solution $D_4$. One approach to direct determination of the RDMs would thus be to solve CSE(2), as an equation for $D_4$, using approximate 4-RDM $N$-representability conditions. For any realistic atomic or molecular system, however, this approach is doomed from the start, not so much by virtue of the $N$-representability problem but rather because the 4-RDM contains $O(R^8)$ unknowns that must be determined. (See Appendix A for a precise enumeration.) CSE(2) has been solved in this manner only for Lipkin’s quasi-spin model. This extremely simple model belies the overwhelming complexity of a 4-RDM-based solution of CSE(2), since the 4-RDM for Lipkin’s model contains exactly 22 independent parameters, independent of $N$—far fewer than the $O(R^8)$ parameters that are required for an atomic or molecular system in a finite basis of $R$ spin orbitals, where $R$ is roughly proportional to $N$.

In the interest of efficiency, one would therefore like to employ $D_2$ as the fundamental variable in CSE(2), which means that the 3- and 4-RDMs must somehow be represented in terms of the 2-RDM. This is the usual decoupling problem for the higher-order distributions that is encountered also with the BBKGY, Martin-Schwinger-
Kadanoff,\textsuperscript{3-8} and other hierarchial equations for many-particle distribution functions. For the present purpose, one requires reconstruction functionals $D_3 = D_3[D_2]$ and $D_4 = D_4[D_2]$ for the 3- and 4-RDMs.\textsuperscript{*} Such functionals, which are mainly the topic of Chapter 4, convert CSE(2) into a closed set of coupled, nonlinear equations for the matrix elements of $D_2$.

Formally, the existence of reconstruction functionals $D_p[D_2]$ has been proved by Rosina.\textsuperscript{15,16} More precisely, Rosina demonstrated that for non-degenerate ground states there exists a one-to-one map between the set of $N$-representable 2-RDMs and the set of $N$-representable $p$-RDMs, for each $p > 2$. This language evokes the first Hohenberg-Kohn theorem,\textsuperscript{14} which states that for non-degenerate ground states, there exists a one-to-one map between the set of $\nu$-representable\textsuperscript{34,35} one-electron densities and the set of $N$-electron wavefunctions (or energies). Just as the Hohenberg-Kohn theorem lends a formal legitimacy to DFT, Rosina’s theorem provides a formal justification for pursuing RDM reconstruction functionals, though neither theorem provides any indication as to the form of the functional.

The analogy between the aforementioned theorems should not be pushed too far, however. It is tempting, for instance, to regard Rosina’s result as a corollary to that of Hohenberg and Kohn, since the density determines the wavefunction and therefore each of the RDMs. In truth, Rosina’s result is much more general, for it assumes nothing about the nature of the Hamiltonian, save that it must contain only one- and two-electron interactions; hence every such Hamiltonian corresponds to the same RDM reconstruction functionals. The RDM reconstruction functionals for an

\textsuperscript{*}We call the functional $D_3[D_2]$ a reconstruction because $D_2$ is determined by $D_p$, for each $p > 2$, via the functional $D_3[D_p] \equiv \text{tr}_{3,\ldots,p} D_p$. 
interacting system must therefore reduce to the correct non-interacting form, \( D_p = D_1^{\lambda p} \), as the two-electron interaction is turned off. (All of the functionals considered in the next chapter have this property.) In contrast, Hohenberg and Kohn’s result requires a fixed form for the two-electron interaction, so that different interactions may lead to different functionals.

Now we have a formal basis for solving CSE(2) in terms of the 2-RDM. The reconstruction functionals that we employ are responsible for insuring \( N \)-representability of the 3- and 4-RDMs, given an \( N \)-representable input 2-RDM, and (insofar as the exact reconstructions are concerned) we need only enforce \( N \)-representability of the 2-RDM explicitly. Practical methods for enforcing known boundary conditions on the 2-RDM have been developed,\(^{36-38}\) while the solution of CSE(2) using a partial set of \( N \)-representability conditions has been addressed, from a formal perspective, by Harriman.\(^{27}\) Not surprisingly, Harriman found that only in the presence of a complete set of \( N \)-representability conditions is the solution \( D_2 \) of CSE(2) unique. That is, unless \( N \)-representability of \( D_2 \) is enforced exactly, there exist spurious two-electron matrices that satisfy CSE(2) but do not correspond to any physical state of an \( N \)-electron system. (This result has been confirmed computationally by Yasuda.\(^{29}\)) Moreover, Harriman\(^{27}\) showed that the spurious, non-\( N \)-representable solutions are not isolated, but rather span a continuous manifold in the two-electron matrix space. Any neighborhood of the unique, \( N \)-representable solution thus contains infinitely many spurious solutions.

The extent to which spurious solutions to CSE(2) are a problem must be ascertained from actual calculations. In this respect, the news is good: calculations\(^{11,25,39-41}\) indicate that meaningful, almost \( N \)-representable 2-RDMs (as measured by the \( D_\cdot,\)
Q-, and G-conditions of Subsection 2.3.5) can be obtained for ground states, and that these approximate 2-RDMs afford accurate electronic energies. Excited states, on the other hand, remain an open problem. In particular, there is no analogue of Rosina’s theorem for excited states. (A purported generalization of Rosina’s theorem for the case of excited states was later shown to be flawed.) Excited-state CSE(2) calculations using reconstruction functionals that work well for ground states, indicate that existing reconstructions are inadequate for this purpose.

3.3 An explicitly extensive CSE

In introducing the RDM cumulants in Section 2.5, we discussed the fact that the RDMs contain unconnected contributions that tend to destroy the extensivity of the energy and other properties, in any approximate calculation. It is hardly surprising, then, that the fundamental equation for direct determination of the 2-RDM, CSE(2), also contains terms that are manifestly unconnected. Indeed, CSE(2) inherits this unfortunate property from the Schrödinger equation itself. Although unconnected contributions to the energy cancel in an exact treatment, this is generally not the case in any approximate treatment, which is why, for example, full CI calculations are size-consistent but no truncated CI method can preserve this property. The only way to insure extensivity (and therefore size-consistency) is to remove all unconnected terms from the working equations a priori, which provides fully-connected equations that are expressed naturally in terms of cumulants. In wavefunction theory, this is accomplished by virtue of the linked cluster theorem in many-body perturbation theory. Connected equations for the wavefunction cumulants are precisely the coupled-cluster (CC) amplitude equations.
The remainder of this section is devoted to an explicit cancellation of the unconnected terms appearing in CSE(1) and CSE(2). As in Section 3.2, the work below applies equally well to any reduced eigenvalue equation. Ultimately CSE(2) is the equation that we wish to solve, since it is the most tractable reduced equation that is still equivalent to the original Schrödinger equation, but we will see that there is reason to consider CSE(1) as well, because this equation plays an instrumental role in removing unconnected terms from CSE(2). The connected equations that we obtain are formally equivalent to the “irreducible” CSEs introduced, in a second-quantized setting, by Kutzelnigg and Mukherjee.\cite{Kutzelnigg1979, Mukherjee1980, Mukherjee1981} In that work, extensive equations were derived in a manner that does not require the original form of the CSE at all, but instead relies on some alternative creation and annihilation operators. In contrast, we find it easier and more illustrative to start from the CSE itself, explicitly demonstrating the cancellation of all unconnected terms. We have written the algebraic equations in terms of two-electron integral kernels, but the final working equations are easily translated into second quantization. Our derivation does not require the introduction of a basis set, and therefore demonstrates that the final, extensive equations are equivalent to the CSE and to the Schrödinger equation. In addition, our derivation clarifies several important differences between the unconnected and the connected equations: the latter are implicit rather than explicit equations for the cumulants, and furthermore the electronic energy, which is an explicit parameter in the original CSE, is absent from its extensive analogues. The ramifications of these distinctions are discussed in Subsection 3.3.3, where we discuss solution of the connected equations in comparison to existing methods for solving CSE(2).
3.3.1 Connectivity, diagrams, and the CSE

CSE\( (p) \) abounds with unconnected terms. Recalling that this equation may be written as \( \Omega_p(1,\ldots,p;1',\ldots,p') \equiv 0 \), this is immediately evident upon replacing the RDMs in \( \Omega_p \) [Eq. (3.16)] with their cumulant decompositions. As an example, consider the term \( \hat{h} \Delta_2 \) that appears in \( \hat{\Omega}_2 \). In terms of position-space kernels,

\[
\hat{h}(1) D_2(1,2;1',2') = \hat{h}(1) \left[ \Delta_2(1,2;1',2') + \frac{1}{2} \Delta_1(1;1') \Delta_1(2;2') - \frac{1}{2} \Delta_1(1;2') \Delta_1(2;1') \right].
\] (3.21)

The first term on the right is obviously connected. The trace of the second term equals \( N \langle \hat{h} \rangle / 2 \), from which we deduce that this term cannot be extensive, since \( \langle \hat{h} \rangle \) already scales like \( N \). The third term in Eq. (3.21) is actually connected, but differs from the second term only by a coordinate permutation. If the non-extensive second term is removed from CSE(2) then the third term must be removed as well, in order to preserve the antisymmetry of \( \Omega_2 \).

To account for—and remove—all unconnected terms and their exchange counterparts in this way is hopelessly tedious. However, the offending terms are readily identified by means of a diagram technique, which offers the additional advantage of mapping certain permutational symmetries within the equations onto easily-spotted visual (that is, topological) symmetries within the diagrams. As a result, the diagrammatic equations are enormously easier to check for mistakes. Furthermore, several of the reconstruction functionals considered in the next chapter are derived via diagrammatic perturbation theory, and although the diagrams introduced here are not perturbative, in conjunction with perturbative reconstruction formulas they do help to clarify the role that the basic structure of CSE(2) plays in correcting reconstruction approximations and compensating for deficiencies and omissions in the perturbation
theory. Our diagram conventions are conceived with this purpose in mind, and are basically unrelated to the CSE diagrams introduced by Mukherjee and Kutzelnigg.\textsuperscript{20,45}

A basic dictionary of diagrams representing \( D_1 \equiv \Delta_1, \tilde{h}, \tilde{g}, \) and \( \Delta_p \) is depicted in Fig. 3.1. Any term in \( \Omega_p \) can be represented in terms of these quantities, and a diagrammatic representation is constructed by connecting operator diagrams (representing \( \tilde{h} \) and \( \tilde{g} \)) to cumulant diagrams. The cumulant vertices at which an operator diagram is attached represent the coordinates on which the operator acts. For example,

\[
\begin{array}{c}
\text{Fig. 3.1 diagram}
\end{array} = \hat{g}(2,3)\left[ \Delta_1(1:2') \Delta_2(2,3;1',3') \right] \tag{3.22}
\]

is obtained by attaching a \( \hat{g}(2,3) \) diagram at the lower endpoints of a \( \Delta_2(2,3;1',3') \) diagram, because according to the diagram these endpoints stand for coordinates 2 and 3. A factor of \( \Delta_1(1;2') \) is present, as indicated, but because \( \hat{g}(2,3) \) operates on neither \( x_1 \) nor \( x_0 \), this part of the diagram is not connected to the rest. This provides a simple, pictorial means to identify the fact that the algebraic term in question is not connected, and should therefore be excluded from the final, connected version of CSE(2). The algebraic reason why this term must be excluded is seen by permuting coordinates 1 and 2 in Eq. (3.22). The resulting quantity, which is also present in CSE(2), has a trace that scales like \( N^2 \).

A trace over coordinate \( x_n \) is indicated diagrammatically by connecting the line labeled \( n \) to the line labeled \( n' \). The labels \( n \) and \( n' \) are then deleted, since these coordinates become a single dummy integration variable. In the case that both \( x_n \) and \( x_n' \) are arguments of the same cumulant (a pure contraction), this creates a loop.
Applying the $\text{tr}_3$ operator to Eq. (3.22), for example, yields

$$
\left. \begin{array}{c} \hat{g}(1, 2) \end{array} \right| = \Delta_{1(1; 2)} \text{tr}_3 \left\{ \hat{g}(2, 3) \Delta_{2(2, 3; 1, 3)} \right\}.
$$

In the case of a transvection, in which $x_n$ and $x_0'$ are arguments of different cumulants, $\text{tr}_n$ serves to connect two cumulant diagrams. For example, in the diagram

$$
\left. \begin{array}{c} \hat{g}(3, 4) \end{array} \right| = \Delta_{1(1; 2')} \text{tr}_{3, 4} \left\{ \hat{g}(3, 4) \Delta_{1(2, 3')} \Delta_{2(3, 4; 1, 4')} \right\},
$$

the operator $\text{tr}_{3, 4}$ connects a $\Delta_{1(2, 3')}$ diagram to a $\Delta_{2(3, 4; 1, 4')}$$ diagram. Note carefully the difference between the diagrams in Eqs. (3.23) and (3.24). These two examples illustrate that operator vertices *internal* to a diagram (that is, not appearing at the endpoint of a cumulant line) are each associated with a coordinate integration. In contrast, vertices that appear as endpoints of a cumulant line [such as the vertex labeled “2” in Eq. (3.23)] do not indicate coordinate integrations. In our diagrams, only the latter “endpoint” vertices are labeled as $1, 2, \ldots$ or $1', 2', \ldots$.

In close analogy to diagrammatic perturbation theory (although our diagrams are not perturbative), we have transformed the problem of generating terms in $\Omega_p$ into a problem of generating topologically distinct diagrams. Restriction to topologically distinct...
inequivalent diagrams is an easy way to incorporate symmetries such as $\hat{g}_{(j,k)} = \hat{g}_{(k,j)}$ that reduce the number of terms in CSE($p$).

Equations (3.18) and (3.19) give CSE(1) and CSE(2) in forms that are conducive to diagrammatic analysis. Upon substituting cumulant decompositions for the RDMs in these equations, one may identify a set of terms that involve only the one-electron cumulant. These are

$$
6 \text{tr}_3 \left\{ \hat{h}(3) \Delta_1^{\lambda^3_{123;1'2'3'}} \right\} = \hat{P}_2 \left\{ \begin{array}{c}
\hat{P}_2 \\
\hat{P}_2
\end{array} \right\} + \left( \begin{array}{c}
\hat{P}_2 \\
\hat{P}_2
\end{array} \right), \quad (3.25)
$$

$$
6 \text{tr}_3 \left\{ \left[ \hat{g}(1,3) + \hat{g}(2,3) \right] \Delta_1^{\lambda^3_{123;1'2'3'}} \right\} = \hat{P}_2 \hat{P}_2 \left\{ \begin{array}{c}
\hat{P}_2 \\
\hat{P}_2
\end{array} \right\} + \left( \begin{array}{c}
\hat{P}_2 \\
\hat{P}_2
\end{array} \right), \quad (3.26)
$$

$$
6 \text{tr}_{2,3} \left\{ \hat{g}(2,3) \Delta_1^{\lambda^3_{123;1'2'3'}} \right\} = 2 \left\{ \begin{array}{c}
\hat{P}_2 \\
\hat{P}_2
\end{array} \right\} + \left( \begin{array}{c}
\hat{P}_2 \\
\hat{P}_2
\end{array} \right), \quad (3.27)
$$

and

$$
24 \text{tr}_{3,4} \left\{ \hat{g}(3,4) \Delta_1^{\lambda^4_{1234;1'2'3'4'}} \right\} = \hat{P}_2 \left\{ \begin{array}{c}
\hat{P}_2 \\
\hat{P}_2
\end{array} \right\} + \left( \begin{array}{c}
\hat{P}_2 \\
\hat{P}_2
\end{array} \right), \quad (3.28)
$$

along with terms like $\hat{h}(i) \Delta_1^{\lambda_{i(j,k)}}$ that are not worth expressing diagrammatically. (Recall that $\hat{P}_2$ antisymmetrizes coordinates 1 and 2, without any numerical factor, while $\hat{P}'_2$ antisymmetrizes 1' and 2'.) The diagrammatic expressions given above have been simplified by means of certain diagram symmetries, which quickly become obvious upon working through a few examples. Equations (3.25)–(3.28) are highly compact compared to brute-force expansions of the Grassmann products $\Delta_1^{\lambda^3}$ and $\Delta_1^{\lambda^4}$; for example, $\Delta_1^{\lambda^4}$ ostensibly contains $4!^2 = 576$ terms, as compared to the 14 terms given in Eq. (3.28).
In the expressions given above, certain diagrams are seen to lack coordinate
dependence, that is, they are scalars rather than tensors. These scalars are related
to the eigenvalue \( w \). Recall from Eqs. (2.21) and (2.22) the decomposition of \( w \) into
one- and two-electron contributions, \( w = w_1 + w_2 \), with \( w_1 = \langle \hat{W}_1 \rangle = N \langle \hat{h} \rangle \) and
\( w_2 = \langle \hat{W}_2 \rangle = \left( \frac{N}{2} \right) \langle \hat{g} \rangle \). The diagram rules introduced above for tensors are equally
applicable to these scalars, and their application to Eqs. (2.21) and (2.22) affords
diagrammatic representations

\[
w_1 = \text{tr} \left\{ \hat{h}(1) \Delta_1(1;1') \right\} = \ Diagram 1
\]

and

\[
w_2 = \text{tr} \left\{ \hat{g}(1,2) D_2(1,2;1',2') \right\} = \ Diagram 2 + \frac{1}{2} (\ Diagram 3 )
\]

Consider the three terms that make up \( w_2 \) in Eq. (3.30). When \( w \) is the electronic
energy, these three contributions (from left to right) are known as the cumulant cor-
relation energy, the classical electrostatic (Hartree) energy, and the exchange energy.
What we call the “cumulant” correlation energy is sometimes called simply the cor-
relation energy, but we want to distinguish this quantity from the correlation energy
\( E_c = E_{\text{FCI}} - E_{\text{HF}} \) introduced and popularized by Löwdin.\(^46\) Within \textit{ab initio} quantum
chemistry, Löwdin’s is the most common definition of the correlation energy. Since
the HF 1-RDM differs from the FCI one, both \( \Delta_2 \) and \( \Delta_1 \wedge \Delta_1 \) contribute to Löwdin’s
correlation energy \( E_c \). (\( \Delta_2 \equiv 0 \) at the HF level.)

The definition of “exchange” is arguably even more convoluted that that of “cor-
relation,” the DFT community being mostly to blame for the confusion. In the context
of DFT, the term \textit{exact exchange} is used commonly for at least two non-equivalent
quantities. More generally, the extent to which “exchange” and “correlation” energies (and the somewhat related “dynamical” and “non-dynamical” correlation energies) can be rigorously separated remains a matter of some contention. The cumulant decomposition of \( w_2 \), however, provides a universal, extensive, model-independent definition for the exchange and correlation energies. In the same manner, the cumulant decomposition of \( D_2 \) provides a universal definition for the exchange and correlation holes, another source of ill communication between the DFT and \textit{ab initio} quantum chemistry communities. We attempt to alleviate this confusion in our discussion of exchange and correlation holes, which appears in Subsection 5.2.2.

Having elucidated all of the terms in \( \Omega_1 \) and \( \Omega_2 \) that involve only \( \Delta_1 \), we next examine those terms involving higher-order cumulants. In these cases, antisymmetry of \( \Delta_p \) reduces the number of required terms. For example, the identity

\[
\text{(3.31)}
\]

is obtained by exchanging the two lines that enter the top of the \( \Delta_2 \) diagram, which amounts to a permutation of the primed coordinates in this cumulant. Making use of this and similar relationships, one obtains the following simplified expressions for the remaining terms in \( \Omega_2 \):

\[
\text{(3.32)}
\]

\footnote{In the context of “hybrid” density functionals such as Becke’s three-parameter B3 functional, \textit{exact exchange} means the expectation value of the HF exchange operator with respect to the Kohn-Sham determinant. However, in the context of the optimized effective potential (OEP) method and Kohn-Sham perturbation theory, the same term is used to mean a variationally-optimized local approximation to HF exchange.}
9 \text{tr}_3 \left\{ \hat{g}(1,3) \left( \Delta_1 \wedge \Delta_2 \right) (1,2,3;1',2',3') \right\} \\
= \hat{P}_2' \left\{ \begin{array}{c} \\
\end{array} \right\} 
+ \hat{P}_2 \left\{ \begin{array}{c} \\
\end{array} \right\} 
- \hat{P}_2 \left\{ \begin{array}{c} \\
\end{array} \right\} 
- \hat{P}_2 \left\{ \begin{array}{c} \\
\end{array} \right\} 
- \hat{P}_2 \left\{ \begin{array}{c} \\
\end{array} \right\} 
- \hat{P}_2 \left\{ \begin{array}{c} \\
\end{array} \right\} 
\left(3.33\right)

72 \text{tr}_{3,4} \left\{ \hat{g}(3,4) \left( \Delta_1^\wedge 2 \wedge \Delta_2 \right) (1,2,3,4;1',2',3',4') \right\} \\
= \left( \begin{array}{c} \text{tr}_1 - \text{tr}_0 \end{array} \right) 
+ \hat{P}_2 \left\{ \begin{array}{c} \\
\end{array} \right\} 
+ \hat{P}_2 \left\{ \begin{array}{c} \\
\end{array} \right\} 
- \hat{P}_2 \left\{ \begin{array}{c} \\
\end{array} \right\} 
- \hat{P}_2 \left\{ \begin{array}{c} \\
\end{array} \right\} 
\left(3.34\right)

16 \text{tr}_{3,4} \left\{ \hat{g}(3,4) \left( \Delta_1 \wedge \Delta_3 \right) (1,2,3,4;1',2',3',4') \right\} \\
= \hat{P}_2 \hat{P}_2' \left\{ \begin{array}{c} \\
\end{array} \right\} 
- 2 \hat{P}_2 \left\{ \begin{array}{c} \\
\end{array} \right\} 
- 2 \hat{P}_2 \left\{ \begin{array}{c} \\
\end{array} \right\} 
\left(3.35\right)
and

\[ 18 \text{ tr}_{3,4} \left\{ \hat{g}(3,4) \Delta_2^{\wedge 2}(1,2,3,4;1',2',3',4') \right\} \]

\[ = 2 \hat{P}_2 \left\{ \begin{array}{c}
\includegraphics{diagram1.png} \\
\includegraphics{diagram2.png} \\
\includegraphics{diagram3.png}
\end{array} \right\} - 2 \hat{P}_2' \left\{ \begin{array}{c}
\includegraphics{diagram4.png} \\
\includegraphics{diagram5.png} \\
\includegraphics{diagram6.png}
\end{array} \right\} \] (3.36)

In addition to these terms, \( \Omega_2 \) contains

\[ 9 \text{ tr}_3 \left\{ \hat{g}(2,3) \left( \Delta_1 \wedge \Delta_2 \right)(1,2,3;1',2',3') \right\} \]

\[ = \hat{P}_2 \left\{ \begin{array}{c}
\includegraphics{diagram7.png} \\
\includegraphics{diagram8.png} \\
\includegraphics{diagram9.png}
\end{array} \right\} + \left\{ \begin{array}{c}
\includegraphics{diagram10.png} \\
\includegraphics{diagram11.png} \\
\includegraphics{diagram12.png}
\end{array} \right\} + \left\{ \begin{array}{c}
\includegraphics{diagram13.png} \\
\includegraphics{diagram14.png} \\
\includegraphics{diagram15.png}
\end{array} \right\} , \] (3.37)

which can be worked out explicitly or obtained from Eq. (3.33) by swapping the coordinate indices 1 and 2.

We have now identified and simplified all of the permutationally-independent terms in both CSE(1) and CSE(2), and it remains only to remove the unconnected ones by cancellation. This is done in the next subsection. For pedagogical purposes, we conclude this section with another example of the connection between diagrammatic and algebraic formulas, this time working from diagrams to algebra. Equation (3.37), when translated back into algebraic notation, becomes

\[ 9 \text{ tr}_3 \left\{ \hat{g}(2,3) \left( \Delta_1 \wedge \Delta_2 \right)(1,2,3;1',2',3') \right\} \]

\[ = \text{tr}_3 \left\{ \hat{g}(2,3) \left[ \Delta_4(3;3') \Delta_2(1,2;1',2') - \Delta_4(3,2') \Delta_2(1,2;1',3') \right. \right. \]

\[ \left. - \Delta_4(2,3') \Delta_2(1,3;1',2') + \Delta_4(2,3') \Delta_2(1,3;1',3') - \Delta_4(1,3'') \Delta_2(1,3;2',3') \right\} \] (3.38)
3.3.2 Cancellation of unconnected terms

Using a generating functional for $\Omega_p$, Yasuda\textsuperscript{28} has demonstrated that this kernel can be decomposed into a connected part $\Omega_p^c$ and an unconnected part $\Omega_p^u$; our diagrammatic treatment brings this to the forefront, as demonstrated below for the cases $p = 1$ and $p = 2$.

Clearly $\Omega_1$ as defined in Eq. (3.16) [or written out explicitly in Eq. (3.19)] contains unconnected terms. For example, the term $wD_1(1;1')$ is obviously unconnected, since both $w$ and $D_1$ scale like $N$. A diagrammatic analysis of $wD_1(1;1')$ reveals this as well:

$$wD_1(1;1') = \left[ \circ \otimes \bullet + \left( \begin{array}{c} \circ \otimes \circ \end{array} \right) + \frac{1}{2} \left( \circ \otimes \circ - \circ \right) \right]^r.$$  \hspace{1cm} (3.39)

Employing the results in the previous section, however, all unconnected terms in $\Omega_1$ are readily seen to cancel, hence $\Omega_1^u \equiv 0$. The connected part of $\Omega_1$ is found to be

$$\Omega_1^c(1;1') = \left[ \begin{array}{c} \circ \otimes \bullet + \left( \begin{array}{c} \circ \otimes \circ \end{array} \right) + \frac{1}{2} \left( \circ \otimes \circ - \circ \right) \right]^r + 3 \left( \begin{array}{c} \circ \otimes \bullet \end{array} \right)^r + 2 \left\{ \begin{array}{c} \begin{array}{c} \circ \otimes \bullet + \left( \begin{array}{c} \circ \otimes \circ \end{array} \right) - \left( \begin{array}{c} \circ \otimes \circ \end{array} \right) \end{array} \end{array} \right\}.$$  \hspace{1cm} (3.40)
Translated into algebra this reads

\[
\Omega_1^{(1,1')} = \hat{h}(1) \Delta_1(1;1') + \text{tr}_2 \hat{h}(2) \left[ \Delta_2(1,2;1',2') - \Delta_1(1;2') \Delta_1(2;1') \right] \\
+ \text{tr}_2 \hat{g}(1,2) \left[ \Delta_2(1,2;1',2') + \Delta_1(1;1') \Delta_1(2;2') - \Delta_1(1';2) \Delta_1(2;1') \right] \\
+ \text{tr}_2,3 \hat{g}(2,3) \left\{ 3 \Delta_2(1,2,3;1',2',3') + \Delta_2(1,2;1',2') \Delta_1(3;3') - \Delta_2(1,2;1',3') \Delta_1(3;2') - \Delta_2(1,2;3',2') \Delta_1(3;1') - \Delta_2(2,3;1',3') \Delta_1(1;2') + \Delta_1(1;2') \left[ \Delta_1(2;3') \Delta_1(3;1') - \Delta_1(2;1') \Delta_1(3;3') \right] \right\} .
\]

(3.41)

In order to insure that anomalous, unconnected terms do not contribute to the solution of CSE(1), one ought to solve the manifestly extensive equation

\[
\Omega_1^0 \equiv 0 ,
\]

(3.42)

rather than \( \Omega_1 \equiv 0 \). Since the cancellation of unconnected terms in CSE(1) is exact, Eq. (3.42) is equivalent, in a necessary and sufficient sense, to CSE(1). Following Kutzelnigg and Mukherjee, we refer to Eq. (3.42) as the first-order *irreducible* CSE, abbreviated as ICSE(1). ICSE(p) will denote the connected version of CSE(p), which couples the cumulants \( \Delta_1, \ldots, \Delta_p \). To obtain an equation that is equivalent, within a finite basis set, to what we denote as ICSE(p), one must solve the Kutzelnigg-Mukherjee version of this equation simultaneously with its adjoint equation. In contrast, the solution to our ICSE(p) is automatically self-adjoint.

CSE(1) is *not* equivalent to the original \( N \)-electron eigenvalue equation, and therefore neither is ICSE(1); for this we need CSE(2). The unconnected part of \( \Omega_2 \) is

\[
\Omega_2^0 = D_1 \land \Omega_1 \\
= D_1 \land \Omega_1^0 .
\]

(3.43)
This relationship can be verified easily and directly, using the diagrammatic expressions in the previous section. Insofar as CSE(1) is satisfied—a necessary condition in order that CSE(2) be satisfied—all unconnected terms in $\Omega_2$ cancel and we obtain the extensive equation ICSE(2), which is simply

$$\Omega_2^c \equiv 0 \ .$$  \hspace{1cm} (3.44)\]

Carrying out the required cancellations to obtain $\Omega_2^c$ is relatively easy using diagrams.

The result is

$$\Omega_2^c(1,2;1',2') =$$

\begin{align*}
&= \begin{array}{c}
\text{Diagram 1} \\
\text{Diagram 2} \\
\text{Diagram 3} \\
\text{Diagram 4} \\
\text{Diagram 5} \\
\text{Diagram 6} \\
\text{Diagram 7} \\
\text{Diagram 8}
\end{array} + 3 P_2 \left( \begin{array}{c}
\text{Diagram 9} \\
\text{Diagram 10} \\
\text{Diagram 11}
\end{array} \right) \\
&+ 3 \hat{P}_2' \left( \begin{array}{c}
\text{Diagram 12} \\
\text{Diagram 13} \\
\text{Diagram 14}
\end{array} \right) + 3 + 3 \hat{P}_2 \left( \begin{array}{c}
\text{Diagram 15} \\
\text{Diagram 16} \\
\text{Diagram 17} \\
\text{Diagram 18}
\end{array} \right) - 3 \hat{P}_2 \left( \begin{array}{c}
\text{Diagram 19} \\
\text{Diagram 20} \\
\text{Diagram 21}
\end{array} \right) + 6 \hat{P}_2 \left( \begin{array}{c}
\text{Diagram 22}
\end{array} \right) \\
&+ \hat{P}_2 \left( \begin{array}{c}
\text{Diagram 23} \\
\text{Diagram 24} \\
\text{Diagram 25}
\end{array} \right) + \frac{1}{2} \hat{P}_2 \left( \begin{array}{c}
\text{Diagram 26} \\
\text{Diagram 27} \\
\text{Diagram 28}
\end{array} \right) + \frac{1}{2} \hat{P}_2 \left( \begin{array}{c}
\text{Diagram 29} \\
\text{Diagram 30} \\
\text{Diagram 31}
\end{array} \right) - \hat{P}_2 \left( \begin{array}{c}
\text{Diagram 32} \\
\text{Diagram 33} \\
\text{Diagram 34}
\end{array} \right)
\end{align*} \hspace{1cm} (3.45)\]

Including permutations, the above equation consists of 68 terms; lengthy as this may be, it is nevertheless a significant reduction as compared to unsimplified Grassmann products. Equation (3.45) was published in Ref. 17 but its algebraic formulation,
which is not nearly so compact, was omitted from that work. We provide the algebraic version of ICSE(2) in Appendix C.

### 3.3.3 Discussion of the connected equations

Perhaps the most striking feature of ICSE(1) and ICSE(2) is the absence of the eigenvalue \( w \) in these equations. In hindsight, the necessity of its disappearance is obvious, since \( w \) appears in \( \hat{\Omega}_p \) as \( w \hat{D}_p \). The observable \( w \) scales as \( N \), while each term in \( \hat{D}_p \) scales as \( N^k \) for some \( k \geq 1 \), hence no part of the product \( w \hat{D}_p \) exhibits proper scaling in the thermodynamic limit. This entire term must cancel with some other part of CSE(\( p \)). This cancellation is analogous to the fact that the CC amplitude equations, which are extensive, contain the cluster amplitudes but not the electronic energy, even though these equations derive from the Schrödinger equation. \( w \) is specified implicitly by ICSE(1) and ICSE(2), insofar as \( \Delta_1 \) and \( \Delta_2 \) together determine \( D_2 \) and thus \( w = \text{tr} (\hat{W}_{12} \hat{D}_2) \). As shown in Subsection 2.5.4, however, \( \Delta_2 \) generally does not determine \( \Delta_1 \), so that ICSE(1) and ICSE(2) must be solved in tandem in order to obtain the 2-RDM and the electronic energy.

In deriving ICSE(1) and ICSE(2) from the corresponding CSEs, we have merely identified and removed terms that cancel exactly. As such, these two connected equations are together equivalent to CSE(2), and thus equivalent to the original \( N \)-electron eigenvalue equation (Schrödinger equation), given the usual caveat that appropriate \( N \)-representability constraints must be enforced when solving the density-matrix equations. Since these constraints are not completely known, one must in practice contend with an infinite number\(^{27} \) of spurious solutions. Recent calculations\(^{39,41,60,61} \) in which CSE(2) was solved starting from an \( N \)-representable 2-RDM, indicate that for ground
states the solution usually converges to a 2-RDM that is nearly consistent with the $D_-$, $Q_-$, and $G$-conditions\(^\text{62}\) for $N$-representability.\(^*\) As we have remarked in Subsection 2.3.5, these $N$-representability conditions are apparently quite stringent, so we feel safe in concluding that “nearly” $N$-representable solutions are obtained from the CSE, at least in the calculations performed to date.

Even if presented with necessary and sufficient $N$-representability constraints, however, the solution of CSE(2) is only unique provided that $w$ is specified and fixed. This is because CSE(2), being equivalent to the Schrödinger equation, contains both ground- and excited-state solutions. Consequently CSE(2) possesses an infinite number of $N$-representable solutions (in a complete basis), corresponding to different values of $w$.

Because $w$ does not appear in the ICSEs, in order to obtain a unique solution to ICSE(1) and ICSE(2) these equations must in principle be solved subject to $N$-representability constraints and subject to the constraint that $w = \text{tr} (\hat{W}_{12} \hat{D}_2)$ remain fixed. For an auxiliary constraint equation (e.g., $\hat{W} = \hat{S}^2$), the desired value of $\langle \hat{W} \rangle$ is known in advance, so it is possible to enforce such a constraint. However, for the fundamental equations of direct 2-RDM theory, in which $\hat{W}$ is an electronic Hamiltonian, such a constraint is not feasible because the energy is not known in advance. Consequently, the equations ICSE(1) and ICSE(2) in practice possess numerous $N$-representable solutions. These correspond not only to the (pure) ground state and each (pure) excited state, but also to all ensemble states that can be formed from degenerate eigenstates of $\hat{W}$.

\(^*\)For excited states, calculations on Lipkin’s quasi-spin model\(^\text{30–33}\) suggest\(^\text{29}\) that spurious solutions are a much more tenacious problem.
This might at first appear to be a major disadvantage of our connected formulation of the CSE, but in fact the same problem arises in practice with the unconnected CSE(2), insofar as the energy is unknown and therefore $w$ must be updated iteratively during the course of the calculation. This is accomplished at each iteration using the current approximation for $\hat{D}_2$. Viewed in this way, the fact that $w$ does not appear in the irreducible equations may actually be advantageous, as it removes one layer of approximation in the structure of the working equations. Again, there is an analogy to CC theory, wherein the cluster amplitudes are first determined, from equations that do involve the energy, then subsequently the energy is determined using the resulting cluster amplitudes, which define the wavefunction.

We next wish to discuss how ICSE(1) and ICSE(2) can be solved, but first let us briefly summarize how one may solve CSE(2). For $w \neq 0$, CSE(2) may be written in the form

$$\hat{D}_2 = \frac{1}{w} \hat{F}_w[\hat{D}_2, \hat{D}_3, \hat{D}_4], \quad (3.46)$$

where the functional $\hat{F}_w = \hat{\Omega}_2 + w \hat{D}_2$. Assuming that one possesses approximate reconstruction functionals $\hat{D}_3[\hat{D}_2]$ and $\hat{D}_4[\hat{D}_2]$ (see Subsection 3.2.2 and also Chapter 4), Eq. (3.46) can be solved for $\hat{D}_2$ by either of two means. The first option is to substitute the reconstruction functionals directly into $\hat{F}_w$, once and for all, which renders $\hat{F}_w$ a functional of $\hat{D}_2$ only. Upon expanding Eq. (3.46) in a finite basis set, this leads to a closed set of nonlinear equations for 2-RDM elements. These equations can be solved, for example, by a Newton-Raphson procedure. This method has implemented by several groups; see Ref. 24 for a discussion of associated convergence problems. Alternatively, Eq. (3.46) can be solved by self-consistent iteration, employing the re-
construction functionals at each iteration to generate updated 3- and 4-RDMs from the current 2-RDM, and using the latter to estimate \( w \). A variety of rather sophisticated algorithms for carrying out this iteration have been described by Mazziotti\textsuperscript{36,37} and by Valdemoro and co-workers.\textsuperscript{12,38,64,65}

Taking stock of Eqs. (3.40) and (3.45), it does not appear that the ICSEs can be solved by means of self-consistent iteration, since these equations specify \( \hat{\Delta}_1 \) and \( \hat{\Delta}_2 \) only implicitly, whereas the explicit appearance of \( \hat{D}_2 \) in CSE(2) is crucial to one’s ability to iterate this equation. Using cumulant reconstruction functionals \( \hat{\Delta}_3[\hat{\Delta}_1, \hat{\Delta}_2] \) and \( \hat{\Delta}_4[\hat{\Delta}_1, \hat{\Delta}_2] \), however, one could certainly derive closed, nonlinear equations for the cumulant matrix elements, which could be solved using the aforementioned Newton-Raphson procedure. Of the RDM reconstruction functionals derived to date, several\textsuperscript{11,28,39,66} utilize cumulant decompositions (at least implicitly), and thus lend themselves to this approach. In contrast, the reconstruction functionals for \( \hat{D}_3 \) and \( \hat{D}_4 \) introduced by Valdemoro and co-workers\textsuperscript{67,68} (see Subsection 4.2.1) cannot be used to solve the ICSEs because they contain no connected terms whatsoever.

### 3.4 Summary

We have introduced extensive formulations of the CSE and other reduced eigenvalue equations by explicitly cancelling all unconnected terms that appear in the usual formulations of these equations. The resulting equations guarantee that size-consistent solutions are obtained for the one- and two-electron cumulants (and thus the 2-RDM), even for an approximate solution. The extensive formulation is the only one appropriate for extended systems. The derivation is carried out using position-space kernels and Hilbert-space operators, circumventing the introduction of a finite-basis approxi-
information in the formal derivation.

Although the cancellation of unconnected terms is formally exact, this operation is far from trivial, resulting in simultaneous nonlinear equations for the one- and two-particle cumulants that contain these quantities implicitly rather than explicitly. The eigenvalue $w$ of the original Schrödinger equation, which appeared explicitly in the non-extensive form of the CSEs, is absent from the extensive equations (though it can be recovered once the cumulants are known). As a result, within the connected formulation one avoids an additional approximation (namely, iterative approximation of the eigenvalue $w$) that must be introduced in order to solve the unconnected CSEs.

### 3.5 References


Chapter 4

Reconstruction functionals for the three- and four-electron density matrices

4.1 Introduction

The equivalence of CSE(2) and the Schrödinger Equation implies that there is a unique, \(N\)-representable solution \(D_4\) for the latter, for any particular (ground or excited) state of an \(N\)-electron system. As discussed in Subsection 3.2.2, however, it is only practical to seek a solution for the 2-RDM, rather than the 4-RDM, and this requires reconstruction functionals (that is, decoupling approximations) for the 3- and 4-RDMs, by means of which these quantities are expressed in terms of \(D_2\). Formally speaking, for non-degenerate ground states the existence of such functionals has been established,\(^1,2\) in the abstract sense of a one-to-one mapping between the \(N\)-representable 2-RDMs and the \(N\)-representable \(p\)-RDMs, for each \(p > 2\). This result is neither necessary nor sufficient in insure a feasible CSE-based electronic structure method, however. Indeed, it stands to reason that knowledge of the exact reconstruction functionals
for the higher-order RDMs is essentially equivalent to knowledge of necessary and sufficient $N$-representability conditions for $D_2$.* Therefore we will not dwell upon esoteric questions of existence, but instead adopt a pragmatic point of view, inquiring whether reasonable approximations for the functionals $D_3[D_2]$ and $D_4[D_2]$ can be developed in practice. Actually, only $D_4[D_2]$ is essential, as we can always obtain an approximate 3-RDM by contraction of this functional. The examination of certain proposed reconstruction functionals, and the derivation of some new ones, are the primary topics of this chapter.

The first practical approximations to the functionals $D_3[D_2]$ and $D_4[D_2]$ were proposed in 1993 by Valdemoro and co-workers,\textsuperscript{5-7} who then used these functionals to affect an iterative, self-consistent solution of CSE(2). Preliminary results were published for the Be isoelectronic sequence in a double-zeta basis set,\textsuperscript{8} followed later by minimal-basis calculations\textsuperscript{7,9,10} on BeH\textsubscript{2} and H\textsubscript{2}O. In most cases, the resulting energies were quite close to the FCI result, but given that the solution of CSE(2) is non-variational, this observation alone does not constitute convincing evidence of the accuracy of the Valdemoro reconstruction functionals. Ostensibly more compelling is a study by Colmenero and Valdemoro,\textsuperscript{6} in which the authors use FCI 2-RDMs to carry out the reconstruction of $D_3$, then compare the resulting 3-RDM to the FCI one. For Be in a double-zeta basis (4 orbitals) and H\textsubscript{2}O in a minimal basis (7 orbitals), $D_3[D_2^{FCI}]$ is found to reproduce all matrix elements of the FCI 3-RDM to an absolute accuracy of 0.002 or better. Mazziotti\textsuperscript{11} has reported similar results for LiH, BeH\textsubscript{2},

*The author takes a similar view of density functional theory. According to Levy’s constrained search construction\textsuperscript{3} of the universal density functional $E[\rho]$, knowledge of this functional is essentially equivalent to determination of the electronic wavefunction. The Hohenberg-Kohn existence theorem\textsuperscript{4} for $E[\rho]$ is neither necessary nor sufficient to have a successful density functional theory of electronic structure.
BH$_3$, H$_2$O, and CH$_4$ in minimal basis sets, although the information that he presents is more condensed (individual matrix elements are not provided). Each of these studies ought to be viewed skeptically, however, given that only minimal basis sets have been tested to date, with the exception of the Clementi-Roetti double-zeta basis$^*$ for Be, which consists of four $s$ functions; this is hardly better than a minimal basis. It is inadvisable to form an opinion as to the efficacy of any non-variational method based on minimal-basis results alone, as these have a long history of serendipitous error cancellation, several more examples of which are encountered in this chapter and the next. Furthermore, a detailed comparison of the 3-RDM matrix elements reported by Colmenero and Valdemoro$^6$ reveals that the relative accuracy of certain reconstructed matrix elements is quite poor, even if the absolute deviations are small. To wit, some of the largest ($\sim 0.002$) inaccuracies correspond to some of the smallest 3-RDM elements, which are essentially zero at the FCI level but are much larger in the reconstructed RDMs.

We would like to understand this behavior, and also to examine whether it persists in larger basis sets. A first is to examine the mathematical arguments and approximations that led to the Valdemoro reconstruction functionals. This turns out to be a nontrivial exercise, however, as the relevant papers$^5$-$^7$ are rather impenetrable, at least in the opinion of some of the author’s colleagues. We have managed to make sense of these arguments, however, and in Subsection 4.2.1 we present what we feel is the first cogent derivation of the Valdemoro reconstruction functionals, following as closely as possible the original reasoning of Valdemoro and co-workers,$^5$ but clearly exposing assumptions and approximations where they are invoked. In so doing, a num-

$^*$See page 428 of Ref. 12.
A number of peculiar features are revealed, both in the derivation itself and in the final result. First of all, the sequence of mathematical steps leading to these functionals is rife with seemingly inconsistent approximations. There are a number of points where, within a particular equation, the 2-RDM or the 3-RDM is approximated in some but not all of the terms in which it appears. We will show that this subtle balancing of intermediate approximations is essential to the form of the final result; if the approximations are applied consistently, then the result is simply the HF reconstruction $D_p = D_{1}^{p}$. The Valdemoro 3-RDM functional turns out to be exactly the unconnected part $D_{3} - \Delta_{3}$ of the 3-RDM but $D_{4}^{\text{val}} \neq D_{4} - \Delta_{4}$, which is another apparent inconsistency.\footnote{Mazziotti\textsuperscript{13} was the first to point out that $D_{4}^{\text{val}} \neq D_{4} - \Delta_{4}$.}

Other RDM reconstruction functionals have been proposed by Mazziotti\textsuperscript{14} and by Nakatsuji, Yasuda, and co-workers.\textsuperscript{15–18} Each of these is discussed in the present chapter, Mazziotti’s approach in Subsection 4.2.2 and the Nakatsuji-Yasuda (NY) functionals in Subsection 4.5.1 (where we introduce our own generalization of the NY functionals). Though these proposed reconstruction functionals have somewhat different forms, they share a common neglect of the four-particle cumulant $\Delta_{4}$, and in some cases $\Delta_{3}$ as well. Except for the rather limited, minimal-basis studies cited above, there has been no substantive attempt to justify this approximation, either \textit{a priori} or in hindsight, despite recent papers\textsuperscript{19,20} that have raised questions about the validity of truncating the cumulant hierarchy at \textit{any} order, let alone at the four-particle level.

In this chapter we have two goals. The first is to examine the \textit{stand-alone} accuracy (in the absence of the CSE) of proposed 3- and 4-RDM reconstruction functionals, using extended basis sets. This is the subject of Section 4.3. Specifically, in Subsec-
In Subsection 4.3.3 we tackle the question of whether it is reasonable to truncate the cumulant hierarchy, that is, to assume that $\Delta_n \approx 0$ for all $n$ larger than some specified value. We first discuss some formal arguments that cast doubt upon the validity of such an approximation. Next, we calculate exact 3-RDMs and their cumulants for LiH and Be at the FCI level, so that we may determine the extent to which these formal concerns manifest in practice. Remarkably, these calculations are the first of their kind; there are no published data for the cumulants $\Delta_3$ and higher in extended basis sets. (We use 6-31G* and 6-311G for our calculations.) All of the aforementioned theoretical concerns are borne out in these calculations, which demon-
strate conclusively that $\Delta_3$ has many elements that are comparable in magnitude to the corresponding matrix elements of $D_3 - \Delta_3$, the unconnected part of the 3-RDM. Energetic analysis of the FCI cumulants reveals that $\Delta_3$ contributes 10–20 m$E_h$ to the electronic energy of Be and LiH, which is certainly not negligible for chemical applications.

Thus it is of interest to develop reconstruction functionals that do not entirely neglect the higher-order cumulants, and in Sections 4.4 and 4.5 we endeavor to accomplish this by generalizing the perturbative scheme of Nakatsuji and Yasuda.\textsuperscript{15–18} The essence of our approach is based upon the observation that the exact 2-RDM, viewed as the infinite summation of a Rayleigh-Schrödinger perturbation series, in essence defines an effective pairwise interaction potential for electrons, which includes the polarizing effects of electron correlation on the bare Coulomb potential. It is this pair interaction that is seen by pairs of quasiparticles\textsuperscript{23} within the $N$-particle system.* Since the 2-RDM is to be our basic variable in solving the CSE, we have available this renormalized pair interaction to utilize in perturbative expansions of $\Delta_3$ and $\Delta_4$. This is exactly what Nakatsuji and Yasuda have done\textsuperscript{15,16} to obtain an estimate of $\Delta_3$ that may be added to the known, unconnected part of $D_3$, to obtain an approximate 3-RDM. The full generality of this technique, however, was not evident prior to the author’s unification\textsuperscript{28} of the NY functional with more generally known concepts in many-body theory.

Section 4.4 develops the elements of many-body perturbation theory that are necessary in order to derive the new reconstruction functionals. Following an outline of

*This is consistent with Sinanoglu’s observation\textsuperscript{24–27} that pair correlations dominate the correlation energy, and that the shell structure of atoms and molecules is strong evidence that the long-range nature of the Coulomb potential is mitigated by electron correlation.
our basic approach (Subsection 4.4.1), we review the rudiments of diagrammatic many-body perturbation theory in Subsection 4.4.2. As pointed out there, the development of a time-independent perturbation theory for the RDMs proceeds, to a great extent, in exactly the same fashion as development of a time-dependent perturbation theory for the many-particle Green’s functions $\mathcal{D}_p$ that were introduced in Subsection 2.3.2. To the extent possible, we develop both formalisms at once.

Perturbative expansions of $\mathcal{D}_1$ (and to a lesser extent, $\mathcal{D}_2$) are well known, but the formalism is enormously more complex for $\mathcal{D}_p$ with $p > 2$. Actually, the $\mathcal{D}_1$ and $\mathcal{D}_2$ perturbation series are artifactually simple, by virtue of the fact that the Hamiltonian contains only one- and two-particle terms. Much of our efforts thus far have been to elucidate the basic formalism of the many-particle perturbation theory. In particular, beginning at the three-particle level it is absolutely necessary to employ a diagram technique in order to make any progress at all, but the topology of the many-particle diagrams is quite complicated and had not been described in the literature prior to the author's treatment of the subject, which appears here mainly as Subsections 4.4.3–4.4.5.

Subsection 4.4.3 provides a rigorous foundation for the many-particle diagram technique, with special attention paid to the issue of “diagrammatic degeneracy,” in other words, the fact that multiple terms in the perturbation series lead to the same diagrammatic representations. These terms must therefore possess the same algebraic value, and the full power of the diagram technique is only realized if we can enumerate the terms that contribute to any given diagram, because then we can

---

*We will frequently refer to a $p$-particle perturbation series, by which we mean a perturbative expansion of $D_p$ or $\mathcal{D}_p$. An $N$-particle system is still assumed.*
substitute one picture for many algebraic terms. In order to fully divest ourselves of the algebraic perturbation series, we must assign a symmetry weight to each diagram. In Subsections 4.4.4 and Subsection 4.4.5 we derive such weights, for a ladder-type approximation to the $\Delta_3$ and $\Delta_4$ perturbation series. Given the expected dominance of pair correlations, we will argue in Subsection 4.4.4 that the ladder-type terms should dominate the perturbation series. It is shown in Subsection 4.5.4 that self-consistent iteration of the CSE serves to build back in certain diagrams that are absent in the ladder approximation.

Even within the ladder approximation, however, diagrams proliferate rapidly as the order in perturbation theory increases, and brute-force enumeration of them quickly becomes infeasible. Therefore one of the major results of our analysis is a systematization of the derivation of these diagrams. This follows from a constructive proof that each “twisted” ladder-type contribution to the perturbation series is equivalent to some coordinate permutation of a simple, untwisted ladder, the collection of which can be generated by a systematic procedure, without omissions or redundancies. This algorithm is tailor-made for combinatorial analysis, which we exploit in order to derive the diagram weights. Interestingly, for $n$th-order ladder-type diagrams, these weights are the same for both the three- and four-particle cases but differ from the analogous two-particle result, another indication of the relative complexity of the three- and four-particle perturbation theory.

The diagram rules are used in Section 4.5 to derive perturbative expressions for $\Delta_3$ and $\Delta_4$ as functionals of the 2-RDM. In conjunction with the known expressions for the unconnected parts of $D_3$ and $D_4$, which are given in Eqs. (4.10b) and (4.10c) below [see also Eqs. (2.136c) and (2.136d) of Section 2.5], this provides perturbative
realizations of $D_3[D_2]$ and $D_4[D_2]$. The resulting 3-RDM functional recovers the NY one\textsuperscript{15,16} in lowest order. The higher-order functionals have yet to be implemented computationally, but some observations concerning this implementation are presented in Subsection 4.5.3.

4.2 Review of proposed reconstruction functionals

In this section we review two of the three reconstruction schemes proposed prior to the author’s work; the third is introduced and generalized in Section 4.5.

4.2.1 Valdemoro’s unconnected approximations

The business of deriving RDM reconstruction functionals was initiated by Valdemoro and co-workers.\textsuperscript{5,6,29} Valdemoro’s reconstruction functionals are based upon Eqs. (2.51) and (2.52), which furnish exact functionals $D_p[Q_1, \ldots, Q_p]$ and $Q_p[D_1, \ldots, D_p]$ for the particle RDMs in terms of equal- and lower-order hole RDMs, and vice versa. These functionals, which have a simple form reminiscent of the binomial theorem, are essentially a restatement of the fermion anticommutation relations for composite $p$-particles. By positing an approximate separation between particle and hole contributions to these functionals, one may ultimately equate $D_p$ to a sum of Grassmann products of lower-order RDMs. The underlying physics of this sequence of approximations, however, is obfuscated by Valdemoro’s exposition, which utilizes a rather dizzying calliope of graphical relations and other impenetrable notation. Mazzotti\textsuperscript{13,30} at least provides an accurate and readable account of the final results of this derivation, but skips over key details of the original treatment\textsuperscript{5} by Valdemoro and co-workers. What is perhaps most puzzling in all this is that the derivation of Valde-
moro’s reconstruction prescription need not be complicated at all; in this subsection we present a readable derivation, which in our estimation is equivalent to that of Valdemoro et al. Our version reveals a rather inconsistent sequence of approximations leading to the final result.

For the case \( p = 2 \), the HRDM functional \( Q_2[D_1, D_2] \) mentioned above is [see Eq. (2.50)]

\[
Q_2 = I \wedge I - 2D_1 \wedge I + D_2 .
\]  

(4.1)

For simplicity we have assumed the RDMs and HRDMs to be real; otherwise each \( Q_n \) above and in what follows should be replaced by its matrix adjoint. Substituting for the identity operator in the equation above, using \( I = D_1 + Q_1 \) [Eq. (2.43)], one obtains

\[
Q_2 - Q_1 \wedge Q_1 = D_2 - D_1 \wedge D_1 .
\]  

(4.2)

This is an exact equality, but Valdemoro next assumes that the particle and hole matrices separate, so that the right and left sides of this equation may be equated separately to zero. Thus she obtains

\[
D_2 \approx D_1 \wedge D_1 .
\]  

(4.3)

which is exact in the HF case but more generally represents a rather dubious approximation that, among other failings, entirely neglects correlations between opposite-spin electrons. (Consult Chapter 5 for a lengthy discussion of the properties of this approximation for the 2-RDM.) Nevertheless, the final results for the higher-order RDMs in Valdemoro’s reconstruction scheme do contain terms beyond the HF approximation, which foreshadows certain inconsistencies in the arguments to come.
Despite the rather questionable nature of the above result, let us continue un-
daunted, considering next the functional \( Q_3[D_1, D_2, D_3] \) in Eq. (2.51), which is

\[
Q_3 = I^{\delta_3} - 3 D_1 \wedge I^{\delta_2} + 3 D_2 \wedge I - D_3
\]

\[
= Q_1^{\delta_3} - 2 D_1^{\delta_3} + 3 D_2 \wedge Q_1 + 3 D_2 \wedge D_1 - 3 D_1^{\delta_2} \wedge Q_1 - D_3.
\]

The second equality results from the substitution \( I = D_1 + Q_1 \). Now things get dicey. The goal is to obtain a relationship analogous to Eq. (4.2), in which there are no mixed particle/hole terms. To do this, we substitute the approximate 2-RDM in Eq. (4.3) into the term \( D_2 \wedge Q_1 \) in Eq. (4.4). We do not, however, substitute this expression for the other factor of \( D_2 \) appearing in the same equation. This substitution has the effect of cancelling all mixed particle/hole terms, albeit by a rather dubious mechanism. This leaves \( Q_3 \approx Q_1^{\delta_3} - 2 D_1^{\delta_3} + 3 D_2 \wedge D_1 - D_3 \) in place of Eq. (4.4). Positing a separation of the particle and hole RDMs as described above, the particle terms are extracted to obtain

\[
D_3 \approx 3 D_2 \wedge D_1 - 2 D_1^{\delta_3}
\]

\[
\overset{\text{def}}{=} D_3^{\text{Val}},
\]

which defines Valdemoro’s 3-RDM reconstruction functional \( D_3^{\text{Val}} = D_3^{\text{Val}}[D_2] \). [Valde-

moro uses spinless RDMs, so her reconstruction formulas\textsuperscript{5,6} ostensibly differ from Eq. (4.5); the form given above is more easily recovered from Mazziotti’s formulation\textsuperscript{13} of the Valdemoro 3-RDM.]

Derivation of an approximate 4-RDM proceeds in much the same way. Using the 4-HRDM functional in Eq. (2.51) along with the identity \( I = D_1 + Q_1 \), one obtains
the exact relationship

\[ Q_4 - D_4 - Q_1^{\Delta 1} + 4D_3 \wedge D_1 + 3D_1^{\Delta 1} - 6D_2 \wedge D_1^{\Delta 2} \]

\[ = 6(D_2 - D_1^{\Delta 2}) \wedge Q_1^{\Delta 2} + Q_1 \wedge (-8D_1^{\Delta 3} + 12D_1 \wedge D_2 - 4D_3) \]  \hspace{1cm} (4.6)

To eliminate the mixed particle/hole terms, we first substitute \( D_3^{\text{Val}} \) for \( D_3 \) only in the term \( Q_1 \wedge D_3 \). This annihilates all mixed terms that contain only a single factor of \( Q_1 \). To eliminate terms with two factors of \( Q_1 \), substitute \( D_2 \approx D_1 \wedge D_1 \) into the term \( D_2 \wedge Q_1^{\Delta 2} \) only. Finally, assuming the same separation of the unmixed particle/hole terms as above, one obtains Valdemoro’s 4-RDM approximation,\(^5,6,13\)

\[ D_4 \approx 4D_1 \wedge D_3 - 6D_1 \wedge D_2 + 3D_1^{\Delta 4} . \]  \hspace{1cm} (4.7)

If the 3-RDM is known in addition to the 2-RDM, then it can be used directly in approximating \( D_4 \) according to Eq. (4.7); otherwise, substitution of Valdemoro’s approximate 3-RDM into this equation yields \( D_4 \) as a functional of the 1- and 2-RDMs only,

\[ D_4 \approx 6D_1 \wedge D_2 - 5D_1^{\Delta 4} \overset{\text{def}}{=} D_4^{\text{Val}} . \]  \hspace{1cm} (4.8)

However, if we continue down this road and substitute \( D_2 \approx D_1 \wedge D_1 \) into Eq. (4.8), we are left with just the HF approximation, \( D_4 \approx D_1^{\Delta 4} \).

This derivation of Valdemoro’s reconstruction functionals—which follows the arguments originally given by Valdemoro and co-workers\(^5,6,29\)—is more than a little unsatisfying, on account of the inconsistent approximations that are made. This procedure is rendered even more suspicious by the posited particle-hole separation. Such a separation certainly exists at the single-determinant level, where one can partition the usual creation and annihilation operators into disjoints sets, one set for particles
and one set for holes, such that each hole operator anticommutes with every particle operator. However, the presence of multiple determinants in the wavefunction serves to mix particles and holes, which are defined with respect to a particular choice of reference determinant. Valdemoro’s final reconstruction formulas contain terms beyond the HF ones by virtue of a tenuous balance of throwing away certain terms while keeping others.

A more illuminating perspective on Valdemoro’s reconstruction approximations can be obtained by considering their relationship to the RDM cumulants, a connection that was first made by Mazziotti. In Eq. (2.136) we gave formulas for $\hat{D}_p$ ($p \leq 4$) in terms of lower-order cumulants. Upon rearrangement, these relationships afford exact expressions for the unconnected contribution

$$\Sigma_p \overset{\text{def}}{=} D_p - \Delta_p$$

(4.9)

to the $p$-RDM, expressed as a functional $\Sigma_p = \Sigma_p[D_1, \ldots, D_{p-1}]$. Through $p = 4$ these functionals are

$$\Sigma_2 = D_1 \land D_1$$

(4.10a)

$$\Sigma_3 = 3D_1 \land D_2 - 2D_1^\land 3$$

(4.10b)

$$\Sigma_4 = 4D_1 \land D_3 + 3D_2 \land D_2 - 12D_1 \land D_1 \land D_2 + 6D_1^\land 4.$$  

(4.10c)

As discussed in Section 2.5, $\Sigma_p$ is exactly that part of $D_p$ that can be expressed as an unconnected product of lower-order RDMs. The corresponding kernel $\Sigma_p = \Sigma_p(x_1, \ldots, x_p; x_1, \ldots, x_p)$ would represent the quantum-mechanical $p$-particle distribution function, were electrons uncorrelated (up to antisymmetrization) at the $p$-particle

*This is the usual particle-hole formalism; see, for example, §7.5 of Ref. 23.
level. The connected part of the \( p \)-particle distribution, representing genuine correlations between \( p \)-tuples of particles, is of course \( \Delta_p(x_1, \ldots, x_p; x_1, \ldots, x_p) \).

As for Valdemoro’s reconstruction functionals, one sees immediately that \( D_3^{\text{Val}} = \Sigma_3 \), while

\[
D_4^{\text{Val}} = \Sigma_4 - 3 \Delta_2 \wedge \Delta_2 \\
= \Sigma_4 - 3 D_2 \wedge D_2 + 6 D_1 \wedge D_1 \wedge D_2 - 3 D_1^{\text{ad}}.
\]

The inconsistent nature of Valdemoro’s approximation scheme is thus revealed: at the three-particle level, \( D_3^{\text{Val}} \) accounts for all products of lower-order correlation functions but neglects genuine three-particle correlations, while \( D_4^{\text{Val}} \) does not fully account for products of lower-order RDMs. Even if \( D_3 \) is known, \( D_4^{\text{Val}} \) still neglects unconnected products of two-electron correlations. This is potentially a serious flaw, given the importance of two-electron correlations.

The correspondences above also serve to prove the claim by Nakatsuji and Yasuda\(^{15,16}\) to the effect that Valdemoro’s approximations include all terms that contribute to \( D_3 \) through first-order in perturbation theory, since the three-particle term \( \Delta_3 \) that is neglected in \( D_3^{\text{Val}} \) vanishes through first order in perturbation theory.\(^*\) The same is not true of \( D_4^{\text{Val}} \), as the neglected terms \( D_1 \wedge D_1 \wedge D_2 \) and \( D_1^{\text{ad}} \) each contain first-order contributions.

Although we will present some probative calculations using Valdemoro’s reconstruction functionals in Section 4.3, this is done mainly for completeness. (Chronologically, these are among the earliest original results presented here.\(^{32}\)) For future development, Valdemoro’s reconstruction functionals are not recommended. Rather, if

\(^*\)Because there is no way to connect three propagator lines with only one pairwise interaction, \( \Delta_{ijk;rst} \) (which is a sum of connected three-particle diagrams) must vanish through first order in perturbation theory.
one aims to approximate $D_3$ and $D_4$ using only unconnected products, then it is better to use the approximations $D_3 \approx \Sigma_3$ and $D_4 \approx \Sigma_4$. Computationally speaking, these approximations are of the same order of complexity as the Valdemoro functionals, but represent consistent, comprehensible approximations. In Section 4.3 we will discuss our reasons for believing that this level of approximation is probably inadequate for molecular electronic structure applications, though one can view the unconnected approximations $\Sigma_p$ as lowest-order reconstructions, with further approximations for the cumulants $\Delta_3$ and $\Delta_4$ to be made perturbatively. This is the subject of Sections 4.4 and 4.5.

### 4.2.2 Mazziotti’s cumulant-based reconstructions

Mazziotti has suggested an alternative, cumulant-based approximation scheme to obtain $D_3$ and $D_4$, given $D_2$. This method is based on the hierarchy of coupled equations for the RDM cumulants that follows from the one-index contraction relation

$$\text{tr}_{p+1} D_{p+1} = \left( \frac{N - p}{p + 1} \right) D_p.$$  \hspace{1cm} (4.12)

If we replace $D_p$ and $D_{p+1}$ in this equation with their cumulant decompositions, the resulting equation couples all of the cumulants through $(p+1)$st order, and indeed we have a hierarchy of such equations, one for each $p$. Mazziotti’s idea was to close this hierarchy by neglecting the cumulants beyond a certain order.

Notice that Eq. (4.12), when expressed in terms of the cumulants, contains numerous non-extensive pieces on both sides of the equality, since the cumulant expansion of $D_p$ contains terms with up to $p-1$ Grassmann products, and also because $\text{tr}_{p+1}(D_{p+1})$ contains unconnected pure contractions (as defined in Subsection 2.6.2). Just like the CSE hierarchy, however, the hierarchy of contraction relations between...
cumulants must ultimately be equivalent to a set of equations consisting only of extensive terms. All non-extensive terms in Eq. (4.12) must cancel exactly, for each $p$, and this ought to be done at the formal level, in order to derive manifestly connected working equations. The formalism developed in Subsections 2.6.1 and 2.6.2, and also Chapter 3, is precisely what is needed for this task.

By definition, the only connected term on the right side of Eq. (4.12) is $-\frac{p}{(p+1)} \Delta_p$, because all other terms in the cumulant decomposition of $D_p$ are unconnected. Furthermore, the term $\frac{N}{(p+1)} \Delta_p$ on the right cannot be extensive, since $\Delta_p$ already scales like $N$. Now, recall from Subsection 2.6.2 that any one-index contraction may be partitioned into pure contractions ($\mathcal{P}$) and transvections ($\mathcal{T}$), the latter consisting of precisely those terms for which the one-index contraction connects two previously unconnected terms via a matrix product. It follows that the only connected quantities on the left side of Eq. (4.12) are $\text{tr}_{p+1}(\Delta_{p+1})$ along with the transvections of the terms in $D_{p+1}$ that contain exactly one Grassmann product.

The basic pattern is easily seen by example, say, for $p = 2$. Then one has $\text{tr}_{p+1}(\Delta_3 + 3 \Delta_1 \wedge \Delta_2 + \Delta_1^{\wedge 3})$ on the left side of Eq. (4.12). Since the operator $\text{tr}_{p+1}$ cannot connect all three factors in $\Delta_1^{\wedge 3}$, this term must cancel. All pure contractions of $\Delta_1 \wedge \Delta_2$. This leaves

$$3 \mathcal{T}(\Delta_1 \wedge \Delta_2) + \text{tr}_3 \Delta_3 = -\frac{2}{3} \Delta_2$$

(4.13)

as the connected analogue of Eq. (4.12), for $p = 2$. Up to some antisymmetrizers and an overall numerical factor, $\mathcal{T}(\Delta_1 \wedge \Delta_2)$ equals a one-index matrix product of $\Delta_1$ and $\Delta_2$. 
Through \( p = 4 \), the connected analogues of Eq. (4.12) are

\[
\begin{align*}
\text{tr}_1(\Delta_1) &= N \quad (4.14a) \\
2 \text{tr}_2(\Delta_2) &= -\Delta_1 + \Delta_1^2 \quad (4.14b) \\
3 \text{tr}_3(\Delta_3) &= -2\Delta_2 - 9 \hat{x}(\Delta_1 \wedge \Delta_2) \quad (4.14c) \\
4 \text{tr}_4(\Delta_4) &= -3\Delta_3 - 16 \hat{x}(\Delta_1 \wedge \Delta_3) - 12 \hat{x}(\Delta_2 \wedge \Delta_2) \quad (4.14d) \\
5 \text{tr}_5(\Delta_5) &= -4\Delta_4 - 25 \hat{x}(\Delta_1 \wedge \Delta_4) - 50 \hat{x}(\Delta_2 \wedge \Delta_3) \quad (4.14e)
\end{align*}
\]

The transvections appearing in these equations can each be written in terms of matrix products of cumulants, according to the procedure outlined in Subsection 2.6.2.

Mazziotti’s prescription\(^{14}\) involves the iterative solution of Eqs. (4.14a)–(4.14e), starting from given estimates of the cumulants. In conjunction with known formulas for the unconnected parts of the RDMs, this is tantamount to a reconstruction of \( D_3 \) and \( D_4 \) in terms of \( D_2 \). A readable account of this procedure has been given by Harris,\(^{19}\) and will not be repeated here. The gist of the algorithm is an assumption that the cumulants vanish beyond some order, which closes the hierarchy in Eq. (4.14) and allows for a self-consistent solution, starting from input matrices \( \Delta_1 \) and \( \Delta_2 \). In practice, Mazziotti solves for \( \Delta_3 \) by neglecting \( \Delta_4 \).

### 4.3 Assessment of the neglected terms

The validity of both reconstruction methods introduced in Section 4.2 appears to rest upon the assumption that all of the cumulants beyond some given order are negligible, at least in comparison to unconnected products of lower-order cumulants. This tacit assumption has been greeted skeptically by some,\(^{19,20}\) but has never been examined for realistic systems, possibly because the storage and manipulation of three- and
four-particle matrices (that is, six- and eight-index tensors) is highly nontrivial, even for paltry basis sets. In contrast, solution of CSE(2) or its connected equivalent only requires the storage of a two-particle matrix, having about as many elements as there are two-electron integrals. In Subsection 4.3.3 we present the first analysis, for a real physical system and using a realistic wavefunction, of the relative magnitudes of cumulant versus RDM elements. First, however, we present a numerical test of Valdemoro’s reconstruction scheme.

4.3.1 Tests of the Valdemoro 4-RDM functional

As indicated by our derivation in Subsection 4.2.1, the 3- and 4-RDM reconstruction functionals of Valdemoro and co-workers\textsuperscript{5,6} are based upon some rather questionable assumptions and seemingly inconsistent approximations. Serendipitously, the Valdemoro 3-RDM functional $D_3^{\text{val}}$ turns out to be something rational (namely, the unconnected part of the true 3-RDM), but the 4-RDM functional is especially questionable. Nevertheless, Valdemoro’s group has published a number of papers\textsuperscript{7-10} demonstrating accurate results when these functionals are used to solve CSE(2). In particular, electronic energies and 2-RDMs for Be, BeH$_2$, and H$_2$O, in both minimal- and double-zeta-quality basis sets, are comparable to FCI results. Here, we attempt to ascertain the extent to which this reflects the accuracy of the reconstruction functional. Specifically, we will carry out Valdemoro’s reconstruction scheme for $D_4$, taking as input an accurate 2-RDM $D_2^{\text{CI}}$ from a CI calculation\textsuperscript{21} of the ground state of Be.

In addition to the calculations detailed below, which were originally published in Ref. 33, there has been only one other study\textsuperscript{6} of this kind, in which the stand-alone accuracy of Valdemoro’s (or any other) reconstruction procedure was tested against ac-
curate RDMs, without solving the CSE. In that work, Colmenero and Valdemoro examined the 3-RDM reconstruction functional $D_{3}^{\text{Val}}[D_{2}^{	ext{FCI}}]$, taking as input a FCI 2-RDM. This analysis was carried out for $\text{H}_2\text{O}$ in a minimal basis (7 orbitals) and for $\text{Be}$ in a double-zeta basis (4 orbitals), though the latter scarcely deserves to be called an extended basis set.* Although Colmenero and Valdemoro found the reconstructed 3-RDMs to be excellent approximations to the FCI 3-RDMs, one should always be skeptical about drawing conclusions based on minimal-basis or nearly minimal-basis results.

Thus we undertake an examination of Valdemoro’s reconstruction scheme using a good CI calculation as the initial datum. In contrast to the Colmenero-Valdemoro study cited above, we examine $D_{4}^{\text{Val}}$ rather than $D_{3}^{\text{Val}}$; the quality of $D_{4}^{\text{Val}}$ has not previously been tested in a stand-alone capacity. Whereas early CSE(2) calculations by Valdemoro’s group utilized independent 3- and 4-RDM reconstruction functionals, more recently these researchers have abandoned the explicit 3-RDM functional derived in Subsection 4.2.1 in favor of obtaining $D_{3}$ by one-index contraction of the reconstructed 4-RDM.$^{9,10,22,34}$ As such, one facet of $D_{4}^{\text{Val}}$ to be evaluated here is contraction consistency, in other words the extent to which $\frac{1}{2}[D_{4}^{\text{Val}}][D_{2}^{	ext{CI}}]$ reproduces the input 2-RDM, $D_{2}^{	ext{CI}}$.

The starting 2-RDM for our study comes from a CI wavefunction consisting of 180 configurations† (1385 determinants) constructed from HF molecular orbitals,

---

* It is the author’s opinion that the use of basis sets smaller than 6-31G* or cc-pVDZ should be strongly discouraged, except for the purposes of making comparisons to larger basis sets. Minimal basis sets do not adequately reflect the complexity of molecular electronic structure, and there are innumerable examples where conclusions drawn at the minimal-basis level are qualitatively at odds with the results of extended-basis calculations. One of these, we believe, is the analysis of Colmenero and Valdemoro discussed in this subsection, while another such example is encountered in Chapter 5.

†In this context, a configuration is defined as a symmetry-adapted linear combinations of determinants.
which in turn are represented as linear combinations of 48 $s$, $p$, and $d$ STOs. The resulting energy bound $E_{CI} = -14.664193 E_h$ recovers 96.7% of the estimated (non-relativistic) correlation energy of the Be ground state. This wavefunction was chosen mainly due to the large size of the basis and the careful selection of configurations within this basis. Certainly larger—and variationally superior—CI wavefunctions exist for Be (see Ref. 36 for a recent bibliography), though the one that we have selected is more than sufficient to provide a meaningful test of Valdemoro’s reconstruction scheme. It is not clear that the same is true of the tiny basis sets used in the aforementioned work by Colmenero and Valdemoro.

Let us emphasize that the analysis of $D_4^{val}$ carried out here is highly nontrivial, computationally speaking, due to the necessity of storing and manipulating four-electron (eight-index) tensors, containing in the worst-case scenario $O(\nu^8)$ nonzero elements. Some indexing tricks, which exploit permutational symmetries of the RDM spin components, are moderately helpful; these are discussed in Appendix A. In spite of these symmetries, the matrices $D_2^{\alpha\alpha}$ and $D_2^{\alpha\beta}$ obtained from our chosen CI wavefunction still contain, respectively, 5,429 and 33,922 nonzero, permutationally-independent elements, in the canonical HF basis. Transformation to the NO basis increases this to 27,779 and 126,687 nonzero elements, respectively. The 3- and 4-RDMs cannot be stored but must be computed on-the-fly. The 4-RDM, for example, consists of 1,918,215 independent, nonzero elements. Rather than trying to compare the reconstructed 4-RDM to the CI 4-RDM element-by-element, we choose instead to contract

---

*The actual non-relativistic energy (experiment minus an estimate of the relativistic correction) is $(-14.667328 \pm 25) E_h$. See (Ref. 35)

†This number does not seem so bad if one considers that a singlet 4-RDM in a 48-orbital basis can potentially have $\sim 8 \times 10^{11}$ independent, nonzero elements.
these matrices back to the two-electron space, and then compare $D_2^{\text{CI}}$ to the contraction
$\frac{4}{2}D_{4\text{Val}}^{\text{Val}}[D_2^{\text{Cl}}]$ of the reconstructed 4-RDM. This seems like the most logical approach, given our interest in contraction consistency.

The reconstruction $D_2^{\text{Val}}$ will be compared not only to the CI result, but also to two other Grassmann-product ansätze of the type $D_4^{\text{G}}$. $D_4^{\text{G}}$ was introduced in
Eq. (2.148) as a linear combination of the three ways ($D_1^\wedge 4$, $D_2 \wedge D_1 \wedge D_1$, and $D_2 \wedge D_2$) in which the 1- and 2-RDMs can be combined, as unconnected Grassmann products, to yield a four-electron matrix. The “proper” way to approximate the 4-RDM as an unconnected product of lower-order RDMs is to use the unconnected matrix $\Sigma_4$, whose form comes out of the cumulant formalism, in place of $D_4$. However, it is an interesting question whether one can choose the coefficients of the linear combination $D_4^{\text{G}}$ in such a way that $D_4^{\text{G}}$ is a better approximation to $D_4$ than is $\Sigma_4$, according to some specified criterion. As we have already committed to evaluating the contraction consistency of $D_2^{\text{Val}}$, the contraction relations afford an appropriate criterion. Thus we shall make a comparison of Valdemoro’s reconstruction approximation to the $D_4^{\text{G}}$ reconstruction with coefficients $a$, $b$, and $c$ obtained by fitting the appropriate contractions of $D_4^{\text{G}}$ to either $D_2$ or $D_1$. This is not a viable reconstruction scheme in its own right, but it does furnish a benchmark against which we can compare results obtained from Valdemoro’s coefficient set. One set of benchmark values of $a$, $b$, and $c$ are obtained by minimizing the sum of squares of the matrix elements of $\frac{4}{2}D_4^{\text{Val}} - D_2^{\text{Cl}}$, where $D_4^{\text{Val}} = D_4^{\text{Val}}[D_2^{\text{Cl}}]$. A second set of coefficients minimizes the sum of squares of matrix elements of $\frac{4}{2}D_4^{\text{Val}} - D_1^{\text{Cl}}$. The resulting best-fit coefficients are displayed in Table 4.1.

\*The notation $\frac{4}{2}D_4^{\text{Val}}[D_2^{\text{Cl}}]$ means that the 4-RDM is reconstructed from the 2-RDM, using Valdemoro’s functional $D_4^{\text{Val}}[D_2]$, and then the resulting eight-index matrix is contracted over two pairs of indices.
Reconstruction

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>D_{2}^{4} Val</strong>, Eq. (4.7)</td>
<td>Original</td>
<td>0.000000</td>
<td>6.000000</td>
</tr>
<tr>
<td></td>
<td>Rescaled</td>
<td>0.000000</td>
<td>5.840070</td>
</tr>
<tr>
<td><strong>D_{4}^{2} CI</strong></td>
<td>Original</td>
<td>-5.470984</td>
<td>8.709500</td>
</tr>
<tr>
<td></td>
<td>Rescaled</td>
<td>-3.340039</td>
<td>5.317155</td>
</tr>
<tr>
<td><strong>D_{3}^{2} CI</strong></td>
<td>Original</td>
<td>1.171816</td>
<td>5.353079</td>
</tr>
<tr>
<td></td>
<td>Rescaled</td>
<td>-1.016319</td>
<td>4.642740</td>
</tr>
</tbody>
</table>

Table 4.1: Parameters employed in reconstruction of **D_{4}^{G}**. The second set of parameters for each functional have been scaled to obtain proper normalization for the reconstructed 4-RDM.

under the respective headings “fit to **D_{2}^{CI}**” and “fit to **D_{1}^{CI}**”.

The calculations reported here, first published in Ref. 32, represent the first and only test of Valdemoro’s (or any other) reconstruction approximation in a good basis set. In fact, there has been only one other study comparing the reconstructed RDM elements to accurate CI ones. In that work, Colmenero and Valdemoro{6} examined **D_{3}^{Val}(but not D_{4}^{Val})** for Be and H_{2}O, the former in a double-zeta basis and the latter in a minimal basis, and in both cases the authors found that **D_{3}^{Val}(D_{2}^{FCI})** accurately approximates **D_{5}^{FCI}**. Note, however, that a double-zeta for Be consists of only four orbitals, all of s type.

For the present analysis, reconstructions are carried out in the basis of CI NSOs (eigenvectors of **D_{1}^{CI}**). Relative to the canonical HF basis, the CI 2-RDM has fewer large elements in the NSO basis, which facilitates presentation of the data. Nevertheless, each NSO is dominated by a single HF orbital (coefficient greater than 0.95 in magnitude). In Table 4.2 we list all permutationally-independent elements of **D_{2}^{CI}** that are greater than 0.015 in magnitude; juxtaposed with these are the corresponding

*See page 428 of Ref. 12.
elements of $\frac{1}{4}D_{\mathbf{i}j}^2$ for the three sets of parameters given in Table 4.1, one of which correspond to Valdemoro’s 4-RDM reconstruction functional. (In reading the table, recall that “$i$” refers to $|\chi_i\rangle \otimes |\alpha\rangle$ while “$\bar{i}$” refers to $|\chi_i\rangle \otimes |\beta\rangle$.)

Valdemoro’s reconstruction, contracted to the two-electron space, approximates the largest matrix element of $D_{\mathbf{2}}^2$ quite accurately, but for the other elements listed in Table 4.2 the error ranges from 10–40%. These errors are significantly larger than those reported for the Be 3-RDM by Colmenero and Valdemoro, using the aforementioned basis of four $s$-type HF orbitals. In that basis, no element of $D_{\mathbf{3}}^3[D_{\mathbf{2}}^2]_{\text{FCI}}$ differs from the corresponding FCI matrix element by more than 0.0021, prior to application of normalization and positivity corrections. Deviations of $\sim 0.002$ in the Colmenero-Valdemoro study are associated with 3-RDM elements that involve double excitation of the reference configuration; the corresponding FCI 3-RDM elements are zero to five significant digits. Core 3-RDM elements, and also those depending only on single-excitation amplitudes, are all reconstructed to an accuracy of better than 0.0003 in the four-orbital basis. Our results are markedly poorer; already the reconstruction error in the core (1$s$, 2$s$; 1$s$, 2$s$) matrix element of $\frac{3}{4}D_{\mathbf{i}j}^3$ is about 0.0025, comparable to the largest deviation that was found by Colmenero and Valdemoro.
Table 4.2: Largest elements of the CI and reconstructed 2-RDMs, for the Be ground state. The uncorrected traces of the approximate 2-RDMs are $\text{tr} (\frac{1}{2} \mathbf{D}^\text{Val}_4) = 6.164$, $\text{tr} (\frac{3}{2} \mathbf{D}^\text{Core}_4) = 9.828$ for the fit to $\mathbf{D}^\text{CI}_1$, and $\text{tr} (\frac{1}{2} \mathbf{D}^\text{Val}_4) = 6.918$ for the fit to $\mathbf{D}^\text{CI}_2$. The degeneracy $|D_{\frac{3}{2} \uparrow e, \uparrow e \downarrow 2s \downarrow 2s}| \approx |D_{\downarrow 2s, \downarrow 2s \downarrow 2s \downarrow 2s}|$ is accidental and disappears with additional significant digits.
The high accuracy of the Colmenero-Valdemoro results is easy to explain. First, note that if \( n_k = 1 \) exactly, then the corresponding NSO separates from the wavefunction (up to an overall antisymmetrization) when the wavefunction is expressed as an expansion in determinants of NSOs. (This is proved, for example, on page 86 of Ref. 37.) Based upon our discussion of the RDM cumulant formalism in Section 2.5, it is clear in this case that all matrix elements indexed by a unit-occupied NSO are zero, for any \( p > 1 \). (See Ref. 38 for an explicit proof of this last statement.) We have shown that \( \mathbf{D}_{\text{Val}}^{3} = 3 \), hence \( \mathbf{D}_{\text{Val}}^{3} \) is exactly the error in Valdemoro's 3-RDM reconstruction. Thus Valdemoro's 3-RDM functional is exact for all matrix elements indexed by a unit-occupied NSO. Exact unit occupancy is never realized in practice, and the reconstruction errors reported by Colmenero and Valdemoro are a reflection of this fact, but we expect that any matrix element of \( n_k \) with a core spin-orbital index will be small. This explanation has come to light only recently and was not part of the author's original analysis, nor has it ever been mentioned, to the author's knowledge, by Valdemoro and co-workers.

However, for the large matrix elements listed in Table 4.2, Valdemoro's approximation \( \mathbf{D}_{\text{Val}}^{4} \neq 4 \) is no worse than the fit of \( \mathbf{D}_{\text{G}}^{4} \) to \( \mathbf{D}_{\text{CI}}^{2} \). Our results in Table 4.2 demonstrate that \( \mathbf{D}_{\text{Val}}^{4} \) does not share the same accuracy as \( \mathbf{D}_{\text{G}}^{4} \) for core-indexed matrix elements, as one may expect, given that \( \mathbf{D}_{\text{Val}}^{4} \neq 4 \).

The reconstructions listed in Table 4.2 each overestimate the trace of the 2-RDM. Based upon our discussion of the RDM cumulant formalism in Section 2.5, it is clear in this case that all those matrix elements of \( \mathbf{D}_{\text{Val}}^{4} \) containing a core spin-orbital index are zero, for any \( d < n \). We have shown that \( \mathbf{D}_{\text{Val}}^{4} = \mathbf{D}_{\text{G}}^{4} \) for all matrix elements indexed by a spin-occupied NSO. This is proved, for example, on page 86 of Ref. 37. Based upon our discussion of the RDM cumulant formalism in Section 2.5, it is clear in this case that all those matrix elements of \( \mathbf{D}_{\text{Val}}^{4} \) containing a core spin-orbital index are zero, for any \( d < n \). We have shown that \( \mathbf{D}_{\text{Val}}^{4} = \mathbf{D}_{\text{G}}^{4} \) for all matrix elements indexed by a spin-occupied NSO. This is proved, for example, on page 86 of Ref. 37.

The high accuracy of the Colmenero-Valdemoro results is easy to explain.
Table 4.3. It is notable that upon renormalization, Valdemoro’s reconstructed matrix elements show greater deviation from the CI values. The energy calculated using $\frac{1}{2}D_4^{\text{val}}$ also becomes less accurate when we impose proper normalization, increasing from $E = -14.4813 \, E_h$ (not normalized) to $E = -14.0953 \, E_h$ (normalized). In contrast, the energy obtained from the fit to $D_2^{\text{CI}}$ is significantly worse when we do not impose proper normalization: $E = -16.1355 \, E_h$ (not normalized) but $E = -13.9944 \, E_h$ (normalized).

Since $E = \sum_{ijmn} W_{ij;mn} D_{mn;ij}$, an error $\delta D_{mn;ij}$ in $D_{mn;ij}$ induces an error $\delta E_{ij;mn} = W_{ij;mn} \, \delta D_{mn;ij}$ in the energy. The $\delta E_{ij;mn}$ are thus an importance sampling of the errors in the reconstructed 2-RDMs. In Table 4.4 we tabulate $\delta E_{ij;mn}$ values for each of the largest elements of the reconstructed, renormalized 2-RDMs. Again Valdemoro’s approximation fares better than a simple fit to $D_2^{\text{CI}}$. 
| 2-RDM element (NSO basis) | $D_2^{CI}$ | $D_2^{Val}$ | Error | $\frac{1}{2} |D_4^{CI}|$ with fitted parameters $a$, $b$, and $c$ | Fit to $D_1^{CI}$ | Error | Fit to $D_2^{CI}$ | Error |
|--------------------------|-----------|------------|-------|------------------------------------------------|---------------|-------|------------------|-------|
| $D_{1s\uparrow 1s\uparrow}$ | 0.498965  | 0.483228   | -0.015467 | -0.019987 | 0.399078 | -20.0% | 0.439531 | -0.059434 |
| $D_{2s\uparrow 2s\uparrow}$ | 0.455695  | 0.329670   | -0.126025 | -0.093533 | 0.362162 | -20.5% | 0.346983 | -0.108712 |
| $D_{1s\uparrow 1s\uparrow}$ | 0.455013  | 0.405002   | -0.050011 | -0.043813 | 0.411200 | -9.6% | 0.401809 | -0.053204 |
| $D_{1s\uparrow 2s\uparrow}$ | 0.454976  | 0.405001   | -0.049975 | -0.043736 | 0.411240 | -9.6% | 0.401821 | -0.053155 |
| $D_{3s\uparrow 2s\uparrow}$ | 0.080902  | -0.098249  | -0.017347 | 0.032811 | 0.113713 | -140.6% | 0.39999 |
| $D_{4s\uparrow 2s\uparrow}$ | 0.080902  | 0.098249   | 0.017347 | -0.032811 | -0.113713 | -140.6% | 0.39999 |
| $D_{5s\uparrow 2s\uparrow}$ | -0.018827 | -0.024967  | -0.006140 | 0.009120 | 0.027947 | -148.4% | 0.008671 |

Table 4.3: Comparison of CI and reconstructed 2-RDMs for Be, following renormalization.
Table 4.4: Energy errors induced by errors in the largest elements of the Be 2-RDM.

<table>
<thead>
<tr>
<th>Error in Largest Elements of Be 2-RDM</th>
<th>CI 1</th>
<th>CI 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error in Largest Elements of Be 2-RDM</td>
<td>-14.664</td>
<td>-13.994</td>
</tr>
</tbody>
</table>

The total energies calculated with these 2-RDMs are $E_{CI}^1 = -14.664 E_h$ and $E_{CI}^2 = -13.994 E_h$. The fitted to $D_{CI}^1$ gives an energy of $-13.994 E_h$ and the fitted to $D_{CI}^2$ provides an energy of $-14.664 E_h$. The fitted to $D_{CI}^1$ and $D_{CI}^2$ to the largest elements of the Be 2-RDM.

<table>
<thead>
<tr>
<th>Error in Largest Elements of Be 2-RDM</th>
<th>CI 1</th>
<th>CI 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error in Largest Elements of Be 2-RDM</td>
<td>-14.664</td>
<td>-13.994</td>
</tr>
</tbody>
</table>

The total energies calculated with these 2-RDMs are $E_{CI}^1 = -14.664 E_h$ and $E_{CI}^2 = -13.994 E_h$. The fitted to $D_{CI}^1$ gives an energy of $-13.994 E_h$ and the fitted to $D_{CI}^2$ provides an energy of $-14.664 E_h$. The fitted to $D_{CI}^1$ and $D_{CI}^2$ to the largest elements of the Be 2-RDM.

We next examine the eigenvalues of the reconstructed, renormalized matrices and $5a^4b^4$ matrices as a two-index matrix in the NSO basis. In Dirac notation, this two-electron function is $\phi(1)^{1}\phi(2)^{1}$.

\[
\phi(1)^{1}\phi(2)^{1} = f \phi(1)^{1}\phi(2)^{1} - \phi(1)^{1}\phi(2)^{1} \frac{\mathbf{2}}{\mathbf{1}} \equiv \mathbf{2} \phi
\]

where for each $\{\gamma^0 \phi\}$.

In the examples shown in Table 4.4, we may express $\mathbf{2} \phi \equiv \mathbf{2} \Phi \equiv \mathbf{2} \phi$.

\[
\mathbf{2} \phi \equiv \mathbf{2} \Phi \equiv \mathbf{2} \phi
\]

We next examine the eigenvalues of the reconstructed, renormalized matrices.

The table below shows the eigenvalues of the reconstructed, renormalized matrices.

<table>
<thead>
<tr>
<th>Error in Largest Elements of Be 2-RDM</th>
<th>CI 1</th>
<th>CI 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error in Largest Elements of Be 2-RDM</td>
<td>-14.664</td>
<td>-13.994</td>
</tr>
</tbody>
</table>

The total energies calculated with these 2-RDMs are $E_{CI}^1 = -14.664 E_h$ and $E_{CI}^2 = -13.994 E_h$. The fitted to $D_{CI}^1$ gives an energy of $-13.994 E_h$ and the fitted to $D_{CI}^2$ provides an energy of $-14.664 E_h$. The fitted to $D_{CI}^1$ and $D_{CI}^2$ to the largest elements of the Be 2-RDM.

We next examine the eigenvalues of the reconstructed, renormalized matrices.

The table below shows the eigenvalues of the reconstructed, renormalized matrices.

<table>
<thead>
<tr>
<th>Error in Largest Elements of Be 2-RDM</th>
<th>CI 1</th>
<th>CI 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error in Largest Elements of Be 2-RDM</td>
<td>-14.664</td>
<td>-13.994</td>
</tr>
</tbody>
</table>

The total energies calculated with these 2-RDMs are $E_{CI}^1 = -14.664 E_h$ and $E_{CI}^2 = -13.994 E_h$. The fitted to $D_{CI}^1$ gives an energy of $-13.994 E_h$ and the fitted to $D_{CI}^2$ provides an energy of $-14.664 E_h$. The fitted to $D_{CI}^1$ and $D_{CI}^2$ to the largest elements of the Be 2-RDM.

We next examine the eigenvalues of the reconstructed, renormalized matrices.

The table below shows the eigenvalues of the reconstructed, renormalized matrices.

<table>
<thead>
<tr>
<th>Error in Largest Elements of Be 2-RDM</th>
<th>CI 1</th>
<th>CI 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error in Largest Elements of Be 2-RDM</td>
<td>-14.664</td>
<td>-13.994</td>
</tr>
</tbody>
</table>

The total energies calculated with these 2-RDMs are $E_{CI}^1 = -14.664 E_h$ and $E_{CI}^2 = -13.994 E_h$. The fitted to $D_{CI}^1$ gives an energy of $-13.994 E_h$ and the fitted to $D_{CI}^2$ provides an energy of $-14.664 E_h$. The fitted to $D_{CI}^1$ and $D_{CI}^2$ to the largest elements of the Be 2-RDM.

We next examine the eigenvalues of the reconstructed, renormalized matrices.

The table below shows the eigenvalues of the reconstructed, renormalized matrices.

<table>
<thead>
<tr>
<th>Error in Largest Elements of Be 2-RDM</th>
<th>CI 1</th>
<th>CI 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error in Largest Elements of Be 2-RDM</td>
<td>-14.664</td>
<td>-13.994</td>
</tr>
</tbody>
</table>

The total energies calculated with these 2-RDMs are $E_{CI}^1 = -14.664 E_h$ and $E_{CI}^2 = -13.994 E_h$. The fitted to $D_{CI}^1$ gives an energy of $-13.994 E_h$ and the fitted to $D_{CI}^2$ provides an energy of $-14.664 E_h$. The fitted to $D_{CI}^1$ and $D_{CI}^2$ to the largest elements of the Be 2-RDM.
Table 4.5: Spectrum of $D_2^{\alpha\alpha}$ and reconstructed, renormalized approximations thereto, for the Be ground state.

<table>
<thead>
<tr>
<th>$D_2^{\alpha\alpha}$</th>
<th>$\left(\frac{1}{2}D_4^{\text{Val}}\right)^{\alpha\alpha}$</th>
<th>Fit to $D_2^{\alpha\alpha}$</th>
<th>Fit to $D_2^{\beta\beta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.910016</td>
<td>0.810105</td>
<td>0.502110</td>
<td>0.697023</td>
</tr>
<tr>
<td>0.028718</td>
<td>0.047260</td>
<td>0.040686</td>
<td>0.047736</td>
</tr>
<tr>
<td>0.028718</td>
<td>0.047260</td>
<td>0.040686</td>
<td>0.047736</td>
</tr>
<tr>
<td>0.028718</td>
<td>0.047260</td>
<td>0.040686</td>
<td>0.047736</td>
</tr>
<tr>
<td>0.001715</td>
<td>0.007099</td>
<td>0.023537</td>
<td>0.021982</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>0.000000</td>
<td>$-0.003440^a$</td>
<td>$-0.000076$</td>
<td>0.000000</td>
</tr>
</tbody>
</table>

$^a$Not listed are another 41 eigenvalues more negative than $-1 \times 10^{-6}$.

Orbitals, hence there are $R(R-1)/2$ spin geminals. Also note the correspondence $D_{j;K} = 2D_{j_1,j_2;k_1,k_2}$. Natural spin geminals and their occupancies are obtained by diagonalizing the matrix whose elements are the $D_{j;K}$. More precisely, we separately diagonalize the $\alpha\alpha$ and $\alpha\beta$ blocks of this matrix.

The largest and smallest eigenvalues of $D_2$ and the reconstructed, renormalized approximations thereof are displayed in Tables 4.5 and 4.6; eigenvalues of the corresponding 1-RDMs are presented in Table 4.7. For the large eigenvalues we find that Valdemoro’s reconstruction is more accurate than the fitted Grassmann product functionals, but breaks down at the other end of the eigenvalue spectrum. Of the 4560 eigenvalues of $\frac{1}{2}|D_4^{\text{Val}}|$, 195 of them are more negative than $-1 \times 10^{-6}$, though none is more negative than $-0.003440$. This is very much smaller in magnitude than the most negative eigenvalue produced by fitting the parameters $a$, $b$, and $c$.

*The factor of 2 difference in the populations arises because the direct product $|\phi_{k_1} \otimes \phi_{k_2}\rangle$ is not a true two-electron state (since it is not antisymmetric), hence the population of the proper state $|\Phi_K\rangle = |\phi_{k_1} \phi_{k_2}\rangle$ is divided equally amongst $|\phi_{k_1} \otimes \phi_{k_2}\rangle$ and $|\phi_{k_2} \otimes \phi_{k_1}\rangle$, as discussed in Subsection 2.3.1.
Table 4.6: Spectrum of $D_2^{\alpha\beta}$ and reconstructed, renormalized approximations thereto, for the Be ground state.

<table>
<thead>
<tr>
<th>$(D_2^{\alpha\beta})$</th>
<th>$(\frac{1}{2}D_4^{\alpha\beta})$</th>
<th>$(\frac{1}{2}D_4^C)^{\alpha\beta}$</th>
<th>Fit to $D_2^{\alpha\beta}$</th>
<th>Fit to $D_2^C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000664</td>
<td>0.968718</td>
<td>0.502110</td>
<td>0.762930</td>
<td></td>
</tr>
<tr>
<td>0.998484</td>
<td>0.929264</td>
<td>0.502016</td>
<td>0.697030</td>
<td></td>
</tr>
<tr>
<td>0.910307</td>
<td>0.810352</td>
<td>0.487271</td>
<td>0.697030</td>
<td></td>
</tr>
<tr>
<td>0.910016</td>
<td>0.810105</td>
<td>0.450850</td>
<td>0.634978</td>
<td></td>
</tr>
<tr>
<td>0.028718</td>
<td>0.218379</td>
<td>0.040686</td>
<td>0.133495</td>
<td></td>
</tr>
<tr>
<td>0.028718</td>
<td>0.047260</td>
<td>0.040686</td>
<td>0.047736</td>
<td></td>
</tr>
<tr>
<td>0.028718</td>
<td>0.047260</td>
<td>0.040686</td>
<td>0.047736</td>
<td></td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td></td>
</tr>
<tr>
<td>0.000000</td>
<td>-0.003440^a</td>
<td>-0.131858^b</td>
<td>0.000000</td>
<td></td>
</tr>
</tbody>
</table>

^a Not listed are another 110 eigenvalues more negative than $-1 \times 10^{-6}$.

^b Not listed are another 3 eigenvalues more negative than $-1 \times 10^{-6}$.

4.3.2 Iterative improvement via CSE(2)

The results of the previous subsection indicate that Valdemoro’s 4-RDM reconstruction functional, taken by itself, does not preserve the accuracy of an input 2-RDM to any significant extent. In practice, this reconstruction is not employed as a stand-alone approximation but rather is used in an iterative solution of CSE(2). In this subsection, we examine an iterative, CSE-based update of the Valdemoro 4-RDM reconstruction. Starting from the CI 2-RDM used in the previous subsection, we reconstruct $D_3$ and $D_4$, then use these in conjunction with $D_2$ to generate an updated 2-RDM via CSE(2).*

Since our starting 2-RDM is highly accurate, any deviation between this updated

*Because Valdemoro’s 3- and 4-RDM functionals neglect the corresponding cumulants in their entirety, we are forced to use the unconnected equation CSE(2) rather than the ICSEs.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>Fit to $D_1^{CI}$</th>
<th>Fit to $D_2^{CI}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.998136</td>
<td>0.959886</td>
<td>0.952621</td>
<td>0.943006</td>
<td></td>
</tr>
<tr>
<td>0.911596</td>
<td>0.776082</td>
<td>0.871583</td>
<td>0.821229</td>
<td></td>
</tr>
<tr>
<td>0.028778</td>
<td>0.083757</td>
<td>0.055979</td>
<td>0.074902</td>
<td></td>
</tr>
<tr>
<td>0.028778</td>
<td>0.083757</td>
<td>0.055979</td>
<td>0.074902</td>
<td></td>
</tr>
<tr>
<td>0.001856</td>
<td>0.083757</td>
<td>0.055979</td>
<td>0.074902</td>
<td></td>
</tr>
<tr>
<td>0.000484</td>
<td>0.005575</td>
<td>0.003903</td>
<td>0.005039</td>
<td></td>
</tr>
<tr>
<td>0.000352</td>
<td>0.001699</td>
<td>0.000903</td>
<td>0.001414</td>
<td></td>
</tr>
<tr>
<td>0.000352</td>
<td>0.001277</td>
<td>0.000646</td>
<td>0.001043</td>
<td></td>
</tr>
<tr>
<td>0.000055</td>
<td>0.001277</td>
<td>0.000646</td>
<td>0.001043</td>
<td></td>
</tr>
<tr>
<td>0.000055</td>
<td>0.001277</td>
<td>0.000646</td>
<td>0.001043</td>
<td></td>
</tr>
<tr>
<td>0.000055</td>
<td>0.000160</td>
<td>0.000119</td>
<td>0.000148</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>0.000000</td>
<td>0.000001</td>
<td>0.000000</td>
<td>0.000000</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.7: Natural occupation numbers of the CI and reconstructed, renormalized 1-RDMs, for the Be ground state. Each tabulated value represents a set of exactly degenerate values arising from $m_s$- and $|m_l|$-degeneracies. Those degeneracies appearing in the table are accidental.
2-RDM and the input 2-RDM used in the reconstruction ought to be attributable almost entirely to error in the reconstruction functionals.

Let us consider how self-consistent iteration of CSE(2), Eq. (3.10), may be carried out. For this purpose, a more convenient form of CSE(2) is the one given in Eq. (3.46), which for definiteness we rewrite here in matrix form, adding an iteration index \( n \):

\[
D_2^{(n+1)} = F_{E^{(n)}} [D_2^{(n)}, D_3^{(n)}, D_4^{(n)}].
\]  

(4.17)

The matrix \( F_{E^{(n)}} \), which is a functional of the 2-, 3-, and 4-RDMs and also depends parametrically on the energy, can be obtained by comparing this equation to Eq. (3.10). The exact energy \( E \) that defines this functional has been replaced in Eq. (4.17) with

\[
E^{(n)} \overset{\text{def}}{=} \text{tr}(H_{12}D_2^{(n)}),
\]  

(4.18)

where \( H_{12} \) [Eq. (2.35)] is the two-electron reduced Hamiltonian. Given the CI energy and 2-RDMs the \( n = 0 \) iterates, we can use Valdemoro’s reconstruction functionals to evaluate all quantities on the right side of Eq. (4.17) and thereby obtain an updated 2-RDM, \( D_2^{(1)} \).

Actually this update can be accomplished in several ways. Given an approximate 2-RDM \( D_2^{(n)} \) following \( n \) iterations, one might decide to use separate reconstruction functionals (Valdemoro’s or otherwise) for the 3- and 4-RDMs. We indicate this prescription symbolically by writing Eq. (4.17) in the form

\[
D_2^{(n+1)} = \frac{F_{E^{(n)}} [D_2^{(n)}, D_3^{(n)}, D_4^{(n)}]}{E^{(n)}}.
\]  

(4.19)

This is the iteration scheme originally employed by Colmenero and Valdemoro, using the 3- and 4-RDM functionals \( D_3^{\text{val}} \) and \( D_4^{\text{val}} \) that we derived in Subsection 4.2.1.*

*Valdemoro’s procedure also includes adjustments, at each iteration, to enforce various \( N \)-representability constraints, but the iterations themselves are carried out as described here.
Valdemoro’s group has since changed tactics, however, and currently uses an iteration scheme in which the contraction relations are enforced at each iteration.\textsuperscript{9,10,22,34} In this scheme, the 4-RDM is obtained by reconstruction, \( D_4^{(n)} = D_4[D_2^{(n)}] \), but then the lower-order RDMs appearing on the right side of Eq. (4.17) are obtained by contraction of \( D_4^{(n)} \). We indicate this “contraction-consistent” iteration scheme as

\[
D_2^{(n+1)} = \frac{\mathcal{F}_{E(n)}[\frac{4}{3} D_4[D_2^{(n)}], \frac{4}{3} D_4[D_2^{(n)}], D_4[D_2^{(n)}]]}{E(n)},
\]

where \( D_4[\ldots] \) can represent any 4-RDM reconstruction functional. Finally,

\[
D_2^{(n+1)} = \frac{\mathcal{F}_{E(n)}[D_2^{(n)}, \frac{4}{3} D_4[D_2^{(n)}], D_4[D_2^{(n)}]]}{E(n)}
\]

represents an intermediate approach, in which the 3-RDM is obtained by contraction of the 4-RDM, but the 2-RDM is used “as is”. This is the method used by Mazziotti.\textsuperscript{13,39}

Unfortunately, the large basis set used (intentionally) for this example precludes us from carrying out a complete self-consistent solution to the CSE at this time. Without implementing sparse matrix manipulation techniques, we are only able to update a modest number of 2-RDM matrix elements, not the full matrix, hence only a single iteration is possible. Thus, starting from \( D_2^{(0)} = D_2^0 \), we calculate an updated value for each of the largest elements of the 2-RDM using each of the aforementioned iteration schemes. The results are presented in Table 4.8.
<table>
<thead>
<tr>
<th>2-RDM element</th>
<th>Eq. (4.19)</th>
<th>Eq. (4.20)</th>
<th>Eq. (4.21)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$D_2^{(0)}$</td>
<td>$D_2^{(1)}$</td>
<td>$D_2^{(0)} - D_2^{(1)}$</td>
</tr>
<tr>
<td>$D_{1s,1s;1s,1s}$</td>
<td>0.498965</td>
<td>0.496747</td>
<td>-0.002218</td>
</tr>
<tr>
<td>$D_{2s,2s;2s,2s}$</td>
<td>0.455695</td>
<td>0.426295</td>
<td>-0.029400</td>
</tr>
<tr>
<td>$D_{1s,2s;1s,2s}$</td>
<td>0.455013</td>
<td>0.451807</td>
<td>-0.003206</td>
</tr>
<tr>
<td>$D_{1s,2s;1s,2s}$</td>
<td>0.454976</td>
<td>0.451779</td>
<td>-0.003197</td>
</tr>
<tr>
<td>$D_{1s,2s;1s,2s}$</td>
<td>-0.080902</td>
<td>-0.084309</td>
<td>-0.003407</td>
</tr>
<tr>
<td>$D_{1s,2s;1s,2s}$</td>
<td>0.080902</td>
<td>0.084310</td>
<td>0.003408</td>
</tr>
<tr>
<td>$D_{1s,2s;1s,2s}$</td>
<td>-0.018827</td>
<td>-0.020068</td>
<td>-0.001241</td>
</tr>
</tbody>
</table>

Table 4.8: First iterative correction $D_2^{(1)}$ to the CI 2-RDM $D_2^{(0)}$ for the ground state of Be, using Valdemoro’s reconstruction functionals. Three different iteration schemes are used [Eqs. (4.19)–(4.21)], as described in the text.
Among the three iteration schemes, the fully contraction-consistent scheme of Eq. (4.20) is comically bad, while the procedure that employs separate 3- and 4-RDM reconstruction functionals [Eq. (4.19)] stands apart in accuracy. Relative errors from this scheme are an order of magnitude smaller than those obtained using the intermediate iteration strategy [Eq. (4.21)], which employs $D_{2}, D_{4}^{1}, D_{2}^{1}, D_{4}^{1}, D_{2}^{1}$. When separate 3- and 4-RDM functionals are used, the updated 2-RDM matrix elements of the fully contraction-consistent scheme are closer to the CI values than are the elements of $D_{4}^{1}, D_{2}^{1}$. In other words, a single iteration of CSE(2) has improved these 2-RDM elements relative to stand-alone reconstruction. The intermediate iteration scheme of Eq. (4.21) also nudges the reconstructed matrix elements back toward their CI values.

These results may be artifactual. In contrast to the fully contraction-consistent iteration method, the two updating schemes that appear to work well both employ $D_{2}^{1}$ as the 2-RDM that is used to construct $FE$ at the first iteration. The contraction-consistent scheme uses the approximate 2-RDM $D_{4}^{1}$ instead of $D_{2}^{1}$, and the updated 2-RDM suffers as a result.

All told, the results in this subsection demonstrate that the 4-RDM reconstruction functional proposed by Valdemoro and co-workers is not nearly so accurate as those researchers would have one believe. This is especially significant, given that the Valdemoro group currently uses $D_{2}^{1}, D_{4}^{1}, D_{2}^{1}$ in lieu of an explicit functional for the 3-RDM.

As we have indicated, if only an unconnected 4-RDM functional is to be employed (that is, if $D_{4}^{1}$ is to be neglected completely), then the approximation $D_{4}^{1}$ as the 4-RDM currently uses the 2-RDM elements relative to stand-alone reconstruction. The intermediate iteration scheme employs $D_{2}^{1}, D_{4}^{1}, D_{2}^{1}$, which employs $D_{2}^{1}, D_{4}^{1}, D_{2}^{1}$ instead of $D_{2}^{1}$, which employs $D_{2}^{1}$, which employs $D_{2}^{1}$. This scheme is an order of magnitude smaller than those obtained using the intermediate iteration scheme of Eq. (4.19), which stands apart in accuracy. Relative errors from reconstruction functionals [Eq. (4.19)] are consistently bad, while the procedure that employs separate 3- and 4-RDM reconstruction schemes, the fully contraction-consistent scheme of
subsection alone, it is not clear whether the inaccuracy in $D_{4}^{\text{Val}}$ arises solely due to neglect of this term or whether the problem is a more fundamental one, inherent to neglect of $\Delta_{4}$. In the next subsection we try to answer this question by examining the RDM cumulants from FCI wavefunctions for Be.

4.3.3 Can the cumulant hierarchy be truncated?

With the exception of the author’s own proposals (Section 4.5 and Ref. 28), which have yet to be implemented at the computational level, all RDM reconstruction schemes proposed to date rely on a truncation of the cumulant hierarchy at either the three- or the four-particle level. Valdemoro’s scheme\(^5,6\) (Subsection 4.2.1) neglects $\Delta_{3}$ and $\Delta_{4}$ entirely, while Mazziotti’s method\(^14\) (Subsection 4.2.2) neglects $\Delta_{4}$. We have not yet discussed the NY reconstruction functionals,\(^15-17\) which are generalized by the ones that we propose in Section 4.5, but these neglect $\Delta_{4}$ as well. Thus all reconstruction functionals that have been utilized in applications take for granted that higher-order cumulants are negligible relative to unconnected products of lower-order cumulants. This tacit assumption has been greeted with a fair amount of skepticism within the quantum chemistry community, some of it published\(^19,20\) but mostly circulated privately at recent Sanibel Symposia. Some formal arguments indicating the causes for concern are outlined below, but surprisingly there has been no serious attempt to examine the validity of this assumption using realistic atomic or molecular wavefunctions. The paper by Colmenero and Valdemoro,\(^6\) to which we referred several times in the previous subsection, did examine $\Delta_{3}$ for Be (insofar as $\Delta_{3} = D_{3} - D_{3}^{\text{Val}}$ is precisely the error in Valdemoro’s 3-RDM functional), but we have already made clear our reservations concerning these small-basis results. To the author’s knowledge, there are
no other published data that address the magnitudes of $\Delta_3$ and/or $\Delta_4$ using atomic or molecular wavefunctions.

In this subsection, we present some preliminary data that attempt to ascertain whether the cumulants are negligible in comparison to unconnected products of lower-order RDMs or lower-order cumulants. We make our comparison using the RDMs and cumulants derived from FCI wavefunctions for Be and for LiH. We choose Be in order to make comparisons with the work of Colmenero and Valdemoro, and also because Be may represent a pathological case for truncation of the cumulant hierarchy (see below). This potentially difficult case can then be contrasted with calculations for LiH, which by all accounts is a much simpler, weakly-correlated system. Both minimal- and extended-basis results are presented, though of course we are severely limited by basis size (more so than would be the case in direct solution of the CSE), since we must explicitly construct and manipulate three- and four-electron (six- and eight-index) tensors in order to carry out the desired analysis. Due to the complexity of manipulating these large tensors (see Appendix A), we have thus far analyzed only $\Delta_3$, not $\Delta_4$. An upgraded version of our code is under development, which will allow us to examine $\Delta_4$ as well.

Before presenting our numerical results, let us review the formal lines of reasoning that seem to favor a skeptical point of view with regard to the validity of truncating the cumulant hierarchy. One argument against \textit{a priori} truncation has been proffered by Nooijen and co-workers, and centers on the well-known fact that the natural occupation numbers of the $2p$ orbitals in atomic Be are anomalously large, relative to typical occupancies outside of the formal electronic configuration of a closed-shell atom or molecule. In the NSO basis, the exchange terms in the Grassmann product
\( \Delta_1^{np} \) vanish, and consequently the elements of this \( p \)-electron tensor are (up to an overall constant) equal to products of natural occupation numbers. For example,

\[
(\Delta_1 \wedge \Delta_1 \wedge \Delta_1)_{1s,2p,2p}^{1s,2p,2p} \propto n_{1s}^2 n_{2p}^2 \tag{4.22}
\]

and

\[
(\Delta_1 \wedge \Delta_1 \wedge \Delta_1 \wedge \Delta_1)_{2s,2p,2s,2p}^{2s,2p,2s,2p} \propto n_{2s}^2 n_{2p}^2 . \tag{4.23}
\]

Since \( n_{2p} = n_{2p} = n_{2p} \) is large relative to \( n_{3s} \), \( n_{3p} \), etc., the four-electron matrix element in Eq. (4.23) is large, relative to a “typical” four-electron matrix element in which half of the indices lie beyond the formal valence shell. Certainly, the matrix elements listed above are small relative to those in which all indices refer to core spin-orbitals; however, the element in Eq. (4.22), for example, is much larger that the corresponding 4-RDM element \( (D_4)_{2s,2p,2s,2p}^{2s,2p,2s,2p} \). This 4-RDM element equals the population (squared CI coefficient) of a determinant formed by double excitation of the core \( 1s \) electrons.\(^*\) In contrast, \( \Delta_1^{\wedge 4} \) elements like the one in Eq. (4.23) derive their magnitude from determinants involving only a single excitation into the \( 2p \) shell. Elements such as \( (\Delta_2 \wedge \Delta_1 \wedge \Delta_1)_{2s,2p,2s,2p}^{2s,2p,2s,2p} \) are also much larger than \( (D_4)_{2s,2p,2s,2p}^{2s,2p,2s,2p} \), because the former consist of terms like \( n_{2p}^2 (\Delta_2)_{2s,2s}^{2s,2s} \), that involve only the core part of the 2-RDM. The quantity \( n_{2p}^2 (\Delta_2)_{2s,2s}^{2s,2s} \) also picks up contributions from CI coefficients involving only single excitations, and ought therefore to be non-negligible. Barring some stupendous, serendipitous cancellation of these unconnected terms, it is therefore reasonable to anticipate that

\[
\left| (D_4)_{2s,2p,2s,2p}^{2s,2p,2s,2p} \right| \gg \left| (D_4)_{2s,2p,2s,2p}^{2s,2p,2s,2p} \right| , \tag{4.24}
\]

\(^*\) The reader may wish to review the correspondence between RDM elements and CI coefficients, which is worked out in Appendix D.
since the unconnected product on the left contains contributions from single excitations out of the core, whereas the 4-RDM element on the right side of Eq. (4.24) is nearly zero. The anomalous Be occupancy distribution makes this pathological case, but by virtue of the same reasoning (single- versus double-excitations in the CI wavefunction), one may anticipate that

$$\left| (\Delta_4)_{o_1,o_2,v_1,v_2} \right| \gg \left| (D_4)_{o_1,o_2,v_1,v_2} \right|. \quad (4.25)$$

The $o_i$ and the $v_i$ in this equation represent occupied and virtual spin-orbital indices, respectively. According to this argument, then, it is not reasonable to neglect $\Delta_4$ in an attempt to approximate $D_4$.

Next consider any four-electron singlet state, such as the ground state of Be. The wavefunction for such a system consists of determinants with two $\alpha$-spin electrons and two $\beta$-spin electrons apiece, which implies that $D_{3}^{\alpha\alpha\alpha} = 0$ and $D_{4}^{\alpha\alpha\alpha\alpha} = 0$. On the other hand, $D_{1}^{\alpha}$, $D_{2}^{\alpha}$, and $\Delta_{2}^{\alpha\alpha}$ are each nonzero, as are Grassmann products of these matrices. In order to cancel such unconnected products, and thus recover $D_{3}^{\alpha\alpha\alpha} = 0$ and $D_{4}^{\alpha\alpha\alpha\alpha} = 0$, it must be that $\Delta_{3}^{\alpha\alpha}$ and $\Delta_{4}^{\alpha\alpha\alpha}$ are also nonzero. In this case, the three- and four-electron cumulants are infinitely large relative to the 3- and 4-RDMs.

This is a potentially serious indictment of the RDM cumulant formalism, and insinuates that this may not be an appropriate framework for analysis of electron correlation, though it would be premature to dismiss the cumulant decomposition as worthless at this stage. The basic shortcoming, at least with regard to the situation outlined in the preceding paragraph, is that the cumulants know nothing about $\langle \hat{S}^2 \rangle$ or spin-coupling in general. This coupling implies a type of long-range order that arises because the individual electron spins must couple to an angular momentum eigenstate,
and this type of correlation persists even though one expects that $p$-tuples of electrons must be uncorrelated, in a “collisional” sense, for sufficiently large $p$. Thus dynamical correlation is expected to be unimportant beyond some value of $p$, but non-dynamical correlation remains, even in the higher-order cumulants.

Recently, Kutzelnigg and Mukherjee$^{40}$ have introduced a spin-free formulation of the RDM cumulants (see also Ref. 41), wherein the definition of the cumulants depends on the spin quantum numbers $S$ and $M_S$. This formalism deserves further exploration, as a possible way around the difficulties arising from spin coupling.

All of these arguments ought to make a person suspicious of the validity of truncating the cumulant expansion at finite order, although in the absence of concrete, numerical evidence, one must grudgingly concede the possibility of a miraculous cancellation that rescues such a procedure. At the level of formal analysis, Harris$^{19}$ has searched for—and failed to detect—such a miracle. The broader implications of Harris’ work might be questioned, however, given the lack of realism in his model (a one-parameter wavefunction to describe an eight fermion system).

Less questionable are FCI results, so long as extended basis sets are employed. Thus we have computed FCI wavefunctions (using the GAMESS program$^{42}$) for Be and LiH in several double- and triple-zeta basis sets. For comparative purposes, the minimal STO-6G basis is also employed. The 3-RDMs from these FCI wavefunctions are then decomposed as $D_3 = \Sigma_3 + \Delta_3$, and the two components are examined in terms of their matrix elements and their contributions to the electronic energy.

A direct comparison of the matrix elements of $\Sigma_3$ to those of $\Delta_3$ confirms our suspicion that the cumulant is comparable in magnitude to unconnected products of lower-order RDMs. This must be the case for the $alpha$ spin component, as explained
above, but we find it to be true of the $\alpha\alpha\beta$ component as well. This is demonstrated, in a rather coarse-grained fashion, in Fig. 4.1, which shows a histogram of the quantities $\log|\Delta_{ijk;rst}/\Sigma_{ijk;rst}|$. The distribution is strongly peaked around zero, indicating that the ratio of connected (cumulant) to unconnected matrix elements is typically of order unity. The distribution is noticeably skewed toward positive values of the logarithmic ratio, implying that there are also a fair number of cases in which the cumulant matrix element is one or more orders of magnitude larger than the corresponding element of $\Sigma_3$.

Some selected matrix elements of $\Delta_3$ and $\Sigma_3$ are listed in Table 4.9. It is a difficult task, if not a hopeless one, to summarize the information content of the 3-RDM in only a few selected pieces of data, nevertheless we have tried to provide a representative in Table 4.9 for each of what we see as the major trends among elements...
<table>
<thead>
<tr>
<th>Matrix element</th>
<th>$D_{ijkrst}$</th>
<th>$\Sigma_{ijkrst}$</th>
<th>$\Delta_{ijkrst}$</th>
<th>$W_{ijkrst}/E_h$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1s, 2s, 2p$</td>
<td>0.0000(+0)</td>
<td>-1.4754(-7)</td>
<td>1.4754(-7)</td>
<td>-3.0013(+0)</td>
</tr>
<tr>
<td>$2s, 2p, 2p'$</td>
<td>0.0000(+0)</td>
<td>-2.4830(-4)</td>
<td>2.4830(-4)</td>
<td>-8.9888(-1)</td>
</tr>
<tr>
<td>$1s, 2s, \bar{1}s$</td>
<td>1.5044(-1)</td>
<td>1.5044(-1)</td>
<td>-5.4684(-7)</td>
<td>-4.2026(+0)</td>
</tr>
<tr>
<td>$1s, 2s, \bar{2}s$</td>
<td>1.5039(-1)</td>
<td>1.5039(-1)</td>
<td>8.0339(-7)</td>
<td>-3.4079(-2)</td>
</tr>
<tr>
<td>$1s, 2p, \bar{2}p$</td>
<td>4.3160(-3)</td>
<td>4.3160(-3)</td>
<td>-8.0080(-8)</td>
<td>-2.9294(+0)</td>
</tr>
<tr>
<td>$2s, 2p, \bar{2}s$</td>
<td>1.0111(-7)</td>
<td>-3.4825(-3)</td>
<td>3.4826(-3)</td>
<td>-9.9269(-1)</td>
</tr>
<tr>
<td>$2s, 2p, \bar{2}p$</td>
<td>3.2973(-7)</td>
<td>3.6495(-3)</td>
<td>-3.6492(-3)</td>
<td>-8.8221(-1)</td>
</tr>
<tr>
<td>$1s, 2s, \bar{1}s; 2s, 2p, \bar{2}p$</td>
<td>-1.9549(-5)</td>
<td>-1.8570(-6)</td>
<td>-1.7692(-5)</td>
<td>0.9999(+0)</td>
</tr>
<tr>
<td>$1s, 2s, \bar{2}s; 3p, 2p, \bar{3}d$</td>
<td>-8.6374(-5)</td>
<td>3.110(-15)</td>
<td>-8.6374(-5)</td>
<td>0.0000(+0)</td>
</tr>
</tbody>
</table>

*a3d indicates a certain spherically-symmetric combination of $d_{x^2}$, $d_{y^2}$, and $d_{z^2}$ with unequal coefficients.*

Table 4.9: Selected matrix elements of $D_3$, $\Delta_3$, $\Sigma_3$ and the reduced Hamiltonian $W_{123}$ for Be/6-31G*. Where only three spin-orbitals indices are given, a matrix element along the diagonal is implied. Numbers in parentheses indicate power of ten.

of $\Delta_3$ and $\Sigma_3$. For the convenience of the following discussion, we have arranged the matrix elements in Table 4.9 into four groups, each of which represents an important facet of the cumulant decomposition.

From the first two elements in Table 4.9, we note simply that $\Delta_3^{\alpha\alpha} = -\Sigma_3^{\alpha\alpha}$, as required. The largest elements of $\Sigma_3^{\alpha\alpha}$ are of order $10^{-4}$ in magnitude. This is arguably the most troubling feature of the whole cumulant decomposition, for we will discover that $\Sigma_3^{\alpha\alpha}$ contributes a chemically significant amount to the total electronic energy, which must then be annihilated by $\Delta_3^{\alpha\alpha}$.

The second group of matrix elements in Table 4.9 consist of diagonal $\alpha\alpha\beta$ elements for which at least one index refers to a core $1s$ orbital. For each element in this
group, \(|\Sigma_{ijk;rst}| \gg |\Delta_{ijk;rst}|\). This is consistent with our argument in Subsection 4.3.1 to the effect that the unconnected approximation \(D_3 \approx \Sigma_3\) ought to fare best for matrix elements that include a core spin-orbital index, as this approximation becomes exact in the limit \(n_{1s} \to 1\). For the elements listed in Table 4.9, the unconnected approximation for the 3-RDM is accurate to five or six significant digits, comparable to the minimal-basis results of Colmenero and Valdemoro.\(^6\) Recalling Eq. (4.22), we had speculated that perhaps the diagonal \(1s \ 2p \ \overline{2p}\) element of \(\Delta_3\) might be significant, due to the anomalous \(2p\) occupation numbers in Be, but apparently the accuracy of the unconnected approximation for core-indexed matrix elements trumps this effect.

There is yet a class of elements for which the cumulant is far larger than the density matrix, however, and two such elements are listed in the third grouping in Table 4.9. The commonality between these two elements is that neither contains any core indices, which is a general feature of elements for which \(|\Delta_{ijk;rst}| \gg |D_{ijk;rst}|\). We suspect this may be a general feature of the three-particle cumulant, though it is probably enhanced in Be due to the anomalous value of \(n_{2p}\).

Finally, we have listed in Table 4.9 two off-diagonal elements. These should not be construed as indicative of the entire off-diagonal structure of \(D_3\), but are chosen in order to make the point that off-diagonal elements of \(\Delta_3\) can also be comparable to—and in some cases, significantly larger than—elements of \(\Sigma_3\).

The actual values of connected, unconnected, and total 3-RDM elements are only part of the story; arguably, a more important criterion for judging the importance of \(\Delta_3\) is its impact on the energy and other expectation values. Recall from Subsection 2.3.1 that one can define a \(p\)-particle reduced operator \(\hat{W}_{1\ldots p}\) such that \(\langle \Psi | \hat{W} | \Psi \rangle = \text{tr}(\hat{W}_{1\ldots p} \hat{D}_p)\). Specifically, \(\hat{W}_{123}\) is defined in Eqs. (2.23), and its matrix
elements [Eq. (2.36)] provide an element-by-element weighting of how the 3-RDM contributes to $\langle \hat{W} \rangle$. Here we take $\hat{W}$ to be the electronic Hamiltonian. Matrix elements of $\hat{W}_{123}$ are given in Table 4.9.

Using the $p$-electron reduced Hamiltonian, we can decompose the electronic energy into connected and unconnected contributions from $D_p$, which can be further decomposed by spin. For $p = 3$, we thus define

$$E_{\sigma\tau\mu}^c \overset{\text{def}}{=} \text{tr}(\hat{W}_{123} \Delta_3^{\sigma\tau\mu})$$

and

$$E_{\sigma\tau\mu}^u \overset{\text{def}}{=} \text{tr}(\hat{W}_{123} \Sigma_3^{\sigma\tau\mu}) .$$

The total electronic energy for any singlet state is then given by the sum $E = E^c_3 + E^u_3$ of a connected part

$$E_3^c \overset{\text{def}}{=} 2E_{aaa}^c + 6E_{a\alpha\beta}^c$$

and an unconnected part

$$E_3^u \overset{\text{def}}{=} 2E_{aaa}^u + 6E_{a\alpha\beta}^u .$$

For a four-electron singlet state, $E_{aaa}^c = -E_{aaa}^u$.

The energy decomposition described above has been examined for Be and LiH, at the FCI level, in several different basis sets; the results are compiled in Table 4.10. At least for these two systems, it is clear that the energetic contribution from $\Delta_3$ is a small fraction of the total electronic energy, roughly a thousand times smaller in magnitude than the unconnected contribution. (This was expected, as explained below.) Nevertheless $E^c_3 \sim 10^{-2}E_h$, which is by no means negligible. For Be, comparable values of $E^c_3/E^u_3$ are obtained in both minimal and extended basis sets, and because
the 6-31G* basis includes d functions, we feel that the numerical value of this quantity $(E_5^c / E_3^c \approx 1.6)$ is not likely to change much in a larger basis set. In future calculations, we plan to test this presumption, and also to determine how this ratio scales with $N$.

The smallness of $E_3^c$ can be rationalized on several grounds. Introducing the standard notation\(^43\)

$$\Psi_{ijkl...} = \hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_k \hat{a}_l |\Psi_{HF}\rangle$$

(4.30)

to denote excitations from the HF reference wavefunction, we note that $\langle \Psi_{HF}|\hat{H}|\Psi_{HF}\rangle = 0$ (Brillouin’s theorem). This is the usual explanation for why single excitations have only a modest impact on the energy, in comparison to double excitations, since Brillouin’s theorem implies that single excitations can contribute only beyond the first order of perturbation theory, when nonzero elements such as $\langle \Psi_{ij}^\dagger |\hat{H}|\Psi_{ik}\rangle$ enter into the perturbation formulas. Brillouin’s theorem is a more subtle result (see §3.3.2 of Ref. 43 for a proof), whereas it is obvious that $\langle \Psi_{HF}|\hat{H}|\Psi_{ijkl...}\rangle = 0$ for any determinant $|\Psi_{ijkl...}\rangle$ consisting of a triple or higher excitation from $|\Psi_{HF}\rangle$. This is because

<table>
<thead>
<tr>
<th>Basis Set</th>
<th>$\text{tr} \Delta_{3}^{\omega\alpha\alpha}$</th>
<th>$2E_{\omega\alpha\alpha}^c / E_h$</th>
<th>$6E_{\omega\alpha\beta}^c / E_h$</th>
<th>$6E_{\omega\alpha\beta}^u / E_h$</th>
<th>$E_3^c / E_3^u$</th>
</tr>
</thead>
<tbody>
<tr>
<td>STO-6G</td>
<td>0.006993</td>
<td>-16.330</td>
<td>-9.223</td>
<td>-14.546869</td>
<td>1.76</td>
</tr>
<tr>
<td>6-31G</td>
<td>0.006082</td>
<td>-13.334</td>
<td>-8.616</td>
<td>-14.604931</td>
<td>1.50</td>
</tr>
<tr>
<td>6-31G*</td>
<td>0.005798</td>
<td>-12.722</td>
<td>-10.389</td>
<td>-14.606245</td>
<td>1.58</td>
</tr>
<tr>
<td>6-311G</td>
<td>0.005663</td>
<td>-12.255</td>
<td>-13.732</td>
<td>-14.619138</td>
<td>1.78</td>
</tr>
<tr>
<td>STO-6G</td>
<td>0.000060</td>
<td>-0.099</td>
<td>-6.037</td>
<td>-8.961422a</td>
<td>0.68</td>
</tr>
<tr>
<td>6-31G</td>
<td>0.000085</td>
<td>-0.125</td>
<td>-6.181</td>
<td>-8.987230a</td>
<td>0.70</td>
</tr>
<tr>
<td>6-311G</td>
<td>0.000126</td>
<td>-0.223</td>
<td>-10.783</td>
<td>-9.003869a</td>
<td>1.22</td>
</tr>
</tbody>
</table>

\(^a\)Nuclear repulsion energy is not included.
$\hat{H}$ is a two-particle operator and its action on any determinant can modify at most two spin-orbital indices in that determinant. Thus, only double excitations in the CI wavefunction contribute directly to the energy; all other excitation types contribute to the energy only insofar as they modify the double-excitation amplitudes.

In addition to the foregoing argument, one may add the antisymmetry requirement to the list of factors that conspire to reduce the importance of three-electron and higher correlations. Antisymmetry prevents three or more electrons from occupying the same point in space, whereas two electrons can occupy the same position, provided they have different spins.

The arguments given above are the standard ones as to why double excitations dominate the energy, but the smallness of $E_3^C$ is also due to the fact that the unconnected approximation for $D_3^C \approx \Sigma_3^C$ is most accurate for exactly those 3-RDM elements having the largest magnitude, namely, the core-indexed ones. Of the remaining 3-RDM elements, the vast majority contribute only a tiny amount to the electronic energy. This is dramatically illustrated in Fig. 4.2, which presents a histogram the quantities $\log |W_{ijk;rst}^\alpha\alpha\alpha|$, $\log |W_{ijk;rst}^\alpha\alpha\beta|$, and $\log |W_{ijk;rst}^\alpha\alpha\beta\alpha\beta\beta|$ for the ground state of Be. This amounts to an element-by-element accounting of the contributions to $E$ from the spin components of $\Delta_3$ and $\Sigma_3$.

Another important factor that tends to keep $|E_3^C/E_3^U|$ small is that the reduced operators $\hat{W}_{1\ldots p}$ are diagonally dominant for $p > 2$, which is a consequence of the two-electron nature of $\hat{W}$. Although the matrix of $\hat{W}_{12}$ is dense, this is not the case for $\hat{W}_{12\ldots p}$ ($p > 2$). When $\hat{W}_{1\ldots p}$ acts on the state $|\phi_{i_1} \otimes \cdots \otimes \phi_{i_p}\rangle$, at most two of the $p$ spin-orbital indices can change, leading to numerous $\delta$-functions in the $\hat{W}_{1\ldots p}$ matrix elements [see, for example, Eq. (2.36)]. Owing to their sparsity, the higher-order
Figure 4.2: Magnitudes of the energy contributions from the connected (cumulant) and unconnected parts of the 3-RDM spin components, for Be/6-31G*. The unconnected \( \alpha \alpha \alpha \) energy contributions are equal and opposite the connected ones.

reduced Hamiltonians reduce the individual importance of the \( p \)-RDM and \( p \)-RDMC matrix elements, for \( p > 2 \).

This fact is relevant because our analysis of Be and LiH 3-RDMs reveals that \( \hat{\Delta}_3 \) tends to dominate \( \hat{\Sigma}_3 \) only off the diagonal, as for instance in the case of the \((1s, 2s, 2s; 3p, 2p, 3d)\) element listed in Table 4.9. Despite the fact that \(|\Delta_{ijk;rst}| \gg |\Sigma_{ijk;rst}|\), the off-diagonal \( W_{ijk;rst} \) are frequently so small that the cumulant cannot run up \( E^C_3 \) too much relative to \( E^U_3 \).

Most of the analysis in this subsection is expected to generalize to the cumulant decomposition of \( \mathbf{D}_4 \). Our LiH and Be results are probably indicative of the general structure of mixed-spin components (e.g., \( \alpha \alpha / \beta \)) of the RDMs and cumulants. In contrast, it is not possible to extrapolate the \( \alpha \alpha \alpha \) results to other systems, because
generally $D_3^\alpha \neq 0$. Hence in future work we ought to extend this analysis at least to a six-electron singlet system, such as carbon atom; the necessary code for this extension is under development.

### 4.4 Renormalized perturbation theory for many-particle propagators

Our examination of Valdemoro's unconnected 4-RDM functional, and especially our analysis of the magnitude and importance of the 3-RDMC, indicates that the unconnected approximation $D_p \approx \Sigma_p$ is an appropriate zeroth-order reconstruction that recovers the vast majority of the electronic energy. This approximation, which completely neglects $\Delta_3$, recovers 99.9\% of the energy for Be atom and an even larger fraction for the weakly-correlated LiH molecule. Even for these four-electron systems, however, 0.1\% of the electronic energy ($10–20\, mE_h$ or 6–12 kcal/mol) is a chemically relevant amount. It remains to be determined whether the same approximation recovers a comparable fraction of $E$ in larger systems, and whether the error depends strongly on molecular geometry, but in any case we would like to be able to move beyond this approximation in order to achieve greater accuracy.* To remedy this, we have introduced\(^{28}\) a successive approximation scheme for $\Delta_p$, based on a renormalized perturbation theory, with $\Sigma_p$ serving as the zeroth-order approximation to $D_p$. In this section, we lay a foundation for this perturbation theory, which is then used to develop reconstruction functionals in Section 4.5

---

*If the fraction of $E$ recovered by the unconnected approximation does not increase with $N$, then we must move beyond this approximation.
4.4.1 Overview

The starting point for developing a perturbative expansion of $\Delta_{i_1, \ldots, i_p; j_1, \ldots, j_p}$ is the second-quantized expression for $D_{i_1, \ldots, i_p; j_1, \ldots, j_p}$, Eq. (2.29). As such, the theory bears much in common with perturbative expansions of the matrix elements $\bar{D}_{i_1, \ldots, i_p; j_1, \ldots, j_p}$ of the $p$-GF. This quantity was introduced in Subsection 2.3.2 as a time-dependent generalization of the $p$-RDM, and for many years has been a staple of condensed-matter and subatomic physics, at least for the cases $p = 1$ and $p = 2$. Our interest in the Green’s functions is purely incidental, but the topological structure of the perturbation series is the same in the time-dependent case as it is in the time-independent case, so for the sake of generality we start from a time-dependent framework. This close similarity arises because the topological structure of the perturbation series is a consequence of Wick’s theorem\textsuperscript{44,45} and the linked-diagram theorem,\textsuperscript{45,46} versions of which exist in both the time-dependent and the time-independent formalism. It is only in the final step, when the perturbative, diagrammatic equations are translated into algebraic equations, that the two methodologies diverge from one another, leading to certain time integrations in the time-dependent theory that replace the familiar energy denominators of time-independent perturbation theory.\textsuperscript{47,48}

The time-dependent version of the theory could, in principle, be used to decouple various hierarchial equations of motion for the Green’s functions,\textsuperscript{49–56} including the so-called Martin-Schwinger-Kadanoff hierarchy\textsuperscript{53,56} of integro-differential equations. The $n$th Martin-Schwinger-Kadanoff equation couples the multiple-time Green’s functions $\bar{D}_n$, $\bar{D}_{n+1}$, and $\bar{D}_{n+2}$, and therefore this hierarchy has been most widely discussed in the context of relativistic field theories, where the full multiple-time nature of the
Green’s functions is essential for a proper description of causality and retardation effects. Closer to the author’s own interests is the work of Prezhdo and co-workers, who have recently developed a mixed quantum/classical dynamics method (“quantized Hamilton dynamics”), in which expectation values \( \langle \hat{p}^m \hat{q}^n \rangle \) of position (\( \hat{q} \)) and momentum (\( \hat{p} \)) operators are propagated for each quantum-mechanical degree of freedom. The formalism leads to a hierarchy of coupled equations of motion for these quantities, which is ultimately truncated at some level \((m_0, n_0)\) by means of a closure approximation, employing the cumulant decomposition of \( \langle \hat{p}^{m_0} \hat{q}^{n_0} \rangle \) and neglecting the connected part of this expectation value. This leads to a closed set of differential equations for each of the lower-order cumulant averages. Given that the formalism developed in this section is, broadly speaking, a perturbative scheme for estimating the expectation values of connected products of creation and annihilation operators, it is possible that a similar methodology could be used to improve the closure approximations of Prezhdo and co-workers in a systematic fashion. In the present work, however, we have made no attempt to pursue the time-dependent theory, beyond the point at which it diverges from the time-independent formalism.

Let us outline and summarize the remainder of this section. Subsection 4.4.2 summarizes the essential aspects of the perturbative expansion of \( \overline{D}_{j_1, \ldots, j_p; k_1, \ldots, k_p} \), including Wick’s theorem and the linked-diagram theorem, which together furnish an equivalent formulation of many-body perturbation theory in terms of linked diagrams.* A diagram formalism of the same origin exists also for the familiar Möller-Plesset form of many-body perturbation theory, though there are some members of the quan-

*The idea of a connected diagram has already been introduced, in the context of the RDM cumulants; the term linked has a different meaning, which is explained in Subsection 4.4.3.
tum chemistry community who seem to regard the diagram technique as merely a mnemonic, or a parlor trick, that enables facile derivation of otherwise well-known Rayleigh-Schrödinger perturbation formulas. This is patently untrue for the problem at hand. The complexity of the many-particle perturbation theory is such that old-fashioned algebraic manipulation is a practical impossibility, by mortal human beings at least. Computer algebra software holds some promise in this regard, especially with respect to factorization and implementation of the resulting equations (using, say, Mathematica to generate Fortran code automatically). There are tremendous start-up costs associated with such an approach, however, and for our purposes a diagram technique is the best approach.

In Subsections 4.4.3–4.4.5, which were originally published as Ref. 28, we provide a careful, rigorous discussion of the foundations of diagrammatic perturbation theory for the quantities $\Delta_{i_1 \ldots i_p, j_1 \ldots j_p}$ and $\overline{\Delta}_{i_1 \ldots i_p, j_1 \ldots j_p}$ ($p = 3$ or $4$). Much attention is devoted to the fact that each topologically distinct diagram is actually an embedding of numerous separate terms in the perturbation series (“diagrammatic degeneracy”). This is true (and well known) for the $\overline{\Delta}_1$ and $\overline{\Delta}_2$ perturbation series as well; however, three- and four-particle diagrams exhibit a greater degree of degeneracy and are generally more complicated than one- or two-particle diagrams. One must enumerate the extent of this diagrammatic degeneracy, by means of certain symmetry coefficients or diagram weights, if the perturbation series are to be expressed in terms of topologically distinct diagrams alone, which is necessary in order to obtain a useful algorithm that generates the distinct diagrams (only) at a given order in perturbation theory. In Subsections 4.4.4 and 4.4.5, we derive the diagram weights appropriate for “ladder” approximations to the three- and four-particle perturbation series. The gist
of each derivation is none other than the desired algorithm for generating the set of independent diagrams, which can then be systematically enumerated.

The reasonableness of ladder-type approximations to the three- and four-particle cumulants must ultimately be validated computationally, though in Subsection 4.4.4 we make some rough arguments to show that the terms included in these approximations should be the dominant ones. It bears pointing out right away that our ladder approximation is not the same as the eponymous approximation of textbook many-body theory,\textsuperscript{23,45,65} which is typically employed for the low-density electron gas but is not appropriate at typical atomic and molecular valence electron densities. In contrast to ordinary, finite-order perturbation theory (which has been formulated for RDMs\textsuperscript{66,67}), our treatment employs exact one- and two-particle RDMs or GFs in order to construct the expansion of the $\Delta_p$ or $\overline{\Delta}_p$. The one- and two-particle quantities possess ordinary Rayleigh-Schrödinger expansions in their own right, which yield the exact quantity when summed to infinite order. Insofar as $\Delta_1$ and $\Delta_2$ are known in advance, the perturbation series for the higher-order propagators thus contain terms that are formally of infinite order in the perturbation matrix elements. Thus, we end up with a renormalized perturbation series for $\Delta_p$ in which the effective pairwise interaction is, loosely speaking, the two-particle cumulant, which contains terms up to infinite order in the bare perturbation.

Less loosely speaking, it is not yet clear how the exact effective interaction can be extracted from $\Delta_2$, but an approximate method (based on earlier work by Nakatsuji and Yasuda\textsuperscript{15,16}) is introduced in Subsection 4.5.1. The 3-RDM reconstruction functional that is obtained, by supplementing the unconnected approximation $\mathbf{D}_3 \approx \Sigma_3$ with a low (renormalized) order perturbative estimate of $\Delta_3$, recovers in leading order
the 3-RDM reconstruction functional derived previously by Nakatsuji and Yasuda.\textsuperscript{15,16} Our more general technique, however, allows for the construction of an approximate 4-RDM containing a nonzero connected part, something that has not been possible previously, and which is probably necessary in order to improve upon the accuracy of the present functionals.

The renormalized perturbation theory framework is only useful when lower-order cumulants are intended to serve as the basic variables, which is exactly what we want in order to solve the CSE. Closely related to this is the fact that iteration of the CSE serves to incorporate terms that are neglected in the perturbative reconstruction formulas themselves. This is shown in Subsection 4.5.4.

The reformulation of a many-body perturbation theory in terms of a two-particle kernel is not without precedent. Somewhat similar renormalizations have been introduced by Faddeev\textsuperscript{68-71} and others\textsuperscript{72-77} in order to recast the few-particle scattering problem as an expansion in the two-particle scattering matrix. In such applications, the particles are usually assumed to be distinguishable. Along similar lines, Lee and Yang\textsuperscript{78,79} have derived a renormalized perturbation series for the full $N$-particle density matrix in terms of the two-particle scattering matrix, which takes full account of the particle statistics (Bose-Einstein or Fermi-Dirac). For small $N$ this method can be used to calculate canonical partition functions from quantum mechanics, even in the case of a divergent (e.g., hard-sphere) pair potential. In contrast to all of this work, however, we require perturbative expansion of the few-particle propagators $\overline{\Delta}_3$ and $\overline{\Delta}_4$ in a system where the number of particles $N$ is allowed to be arbitrarily large, and furthermore all $N$ particles are assumed to be indistinguishable. Under these circumstances the many-body perturbation theory is significantly more complicated.
and none of the aforementioned techniques is directly applicable.

4.4.2 Many-body perturbation theory for the propagators

In this subsection we provide a cursory overview of many-body perturbation theory for the GFs, providing just enough details to motivate our new developments in the remainder of this section. Additional details can be found in numerous textbooks\cite{bib:213, bib:44, bib:60, bib:61, bib:65, bib:80, bib:81} and monographs\cite{bib:47, bib:48, bib:82, bib:83, bib:84, bib:85} and typically we will not bother to provide individual citations for textbook-level facts about perturbation theory; the material leading up to Eq. (4.38), which summarizes the formal development of the ordinary $\mathcal{D}_p$ perturbation series, is an abbreviated version of the discussion found in §8 of Ref. 45. We have supplemented this discussion with some examples of the correspondence between diagrams and algebraic terms in the perturbation series, for the case of $\mathcal{D}_2$. This includes some examples of the diagrammatic degeneracy mentioned in the previous subsection, along with a discussion of diagram weights for the $\mathcal{D}_2$ perturbation series.

Our starting point is the perturbation series for the time-dependent matrix elements $\mathcal{D}_{j_1,\ldots,j_n;k_1,\ldots,k_n} = \mathcal{D}_{j_1,\ldots,j_n;k_1,\ldots,k_n}(t_1,\ldots,t_n,t'_1,\ldots,t'_n)$ of $\mathbf{D}_n$. Taking $\hat{\mathcal{V}} = \hat{H} - \hat{H}_0$ as the perturbation, where $\hat{H}_0$ is an unspecified one-electron Hamiltonian,\footnote{Nominaly, $\hat{H}_0$ is the Fock operator and $|\Psi_0\rangle$ is the HF wavefunction, though recently there has been interest\cite{bib:86, bib:87} in using the KS determinant as a reference state for various post-HF methodologies.}

$$\hat{H}_0 \phi_k = \epsilon_k \phi_k ,$$

we express $\hat{\mathcal{V}}$ as a time-dependent perturbation in the interaction representation,

$$\hat{\mathcal{V}}(t) = \sum_{pqrs} V_{pq;rs} \hat{p}^\dagger(t) \hat{q}^\dagger(t) \hat{s}(t) \hat{r}(t) .$$ (4.32)
The coefficients \( V_{pqrs} = \langle \phi_p \otimes \phi_q | \hat{V} | \phi_r \otimes \phi_s \rangle \) in this equation are independent of time. The time dependence is carried by the operators \( \hat{k}_1(t) \) and \( \hat{k}(t) \), which are the usual creation and annihilation operators, respectively, expressed in the interaction representation:

\[
\hat{k}(t) \equiv e^{it\hat{H}_0} \hat{a}_\kappa e^{-it\hat{H}_0} .
\]

These operators proliferate so rapidly in what follows that it is inconvenient to use the more conventional notation \( \hat{a}_\kappa(t) \). Given the usual assumption\(^{45,65,88}\) of an adiabatic connection between eigenstates \( |\Psi_0 \rangle \) of \( \hat{H}_0 \) and eigenstates \( |\Psi \rangle \) of \( \hat{H} \), one obtains\(^*\)

\[
\begin{align*}
\overline{D}_{j_1,\ldots,j_n;k_1,\ldots,k_n} = & \sum_{m=0}^{\infty} \frac{(-i)^{m+n}}{m!} \sum_{p_1,\ldots,p_m,q_1,\ldots,q_m} \sum_{r_1,\ldots,r_m,s_1,\ldots,s_m} V_{p_1q_1r_1s_1} \cdots V_{p_mq_mr_ms_m} \int_{-\infty}^{\infty} dt_1 \cdots \int_{-\infty}^{\infty} dt_m \\
& \times \langle \Psi_0 | \hat{T} \left( \hat{k}_1(t_1) \cdots \hat{k}_n(t_n) \hat{p}_1^\dagger(\tau_1) \hat{q}_1^\dagger(\tau_1) \hat{s}_1(\tau_1) \hat{r}_1(\tau_1) \right. \\
& \left. \cdots \hat{p}_m^\dagger(\tau_m) \hat{q}_m^\dagger(\tau_m) \hat{s}_m(\tau_m) \hat{r}_m(\tau_m) \hat{j}_n^\dagger(t_n) \cdots \hat{j}_1^\dagger(t_1) \rangle |\Psi_0 \rangle_{\text{linked}}.
\end{align*}
\]

The notation \( \langle \cdots \rangle_{\text{linked}} \) indicates that the expectation value includes only those Wick contractions (see below) that are “linked” in the sense of the linked cluster theorem. This is the well-known result (generalized to the \( p \)-GF by Kobe\(^{89}\)) that all non-extensive contributions to the perturbation series—which appear in the Rayleigh-Schrödinger perturbation series beginning at third order—ultimately cancel exactly, if all orders of perturbation theory are considered. Hence it is consistent to omit them from the outset. For the case \( p \geq 2 \), there is a subtle distinction between terms that are unlinked in the aforementioned sense, and which cancel exactly, and terms that

\(^*\)As explained in §4.5 of Ref. 65, as part of the adiabatic hypothesis, Eq. (4.34) should be understood to include exponential damping factors where they are needed to rectify divergent integrals. That is, the perturbation series is more properly written in terms of a perturbation \( \hat{V}(t) e^{-i|\tau|t} \), taking the limit \( \varepsilon \to 0^+ \) at the end. Such subtleties will not concern us here.
are unconnected in the cumulant sense, which do not cancel but rather contribute to \( \Sigma_p \). This distinction is explained in the next subsection.

A computationally useful form of the \( \overline{D}_p \) perturbation series is obtained by using Wick’s theorem to express the time-ordered operator product \( \overline{T}(\cdots) \) as a sum of all possible Wick contractions of normal-ordered (\( \overline{N} \)) products. A Wick contraction (not to be confused with the partial trace of a RDM) is defined as the difference between time-ordered and normal-ordered products,

\[
\overline{A}_1 \overline{A}_2 \overset{\text{def}}{=} \overline{T}(\hat{A}_1 \hat{A}_2) - \overline{N}(\hat{A}_1 \hat{A}_2) = -\overline{A}_2 \overline{A}_1 .
\]

(4.35)

Here \( \hat{A}_1, \hat{A}_2 \in \{ \hat{c}(t_1) \} \cup \{ \hat{\mu}^\dagger(t_2) \} \) are creation or annihilation operators, generally with different time arguments. The so-called normal-ordered form \( \overline{N}(\hat{A}_1 \cdots \hat{A}_n) \) of any creation and/or annihilation operator product has all creation operators to the left of each annihilation operator, with a minus sign incurred in case an odd permutation is required to achieve this ordering. For example,

\[
\overline{N}(\hat{i}(t) \hat{j}^\dagger(t')) = -\hat{j}^\dagger(t') \hat{i}(t)
\]

(4.36)

for all \( t \) and \( t' \). For products consisting of more than two operators, contractions are defined only between adjacent pairs, so that for example

\[
\overline{A}_1 \overline{A}_2 \overline{A}_3 \overline{A}_4 = - (\hat{A}_1 \hat{A}_3)(\hat{A}_2 \hat{A}_4) .
\]

(4.37)

A minus sign is incurred because the operator orderings on the two sides of the equality differ by an odd permutation.

Although \( \overline{T}[\hat{A}_1(t_1) \cdots \hat{A}_n(t_n)] \) formally consists of all Wick contractions of the operator product \( \hat{A}_1(t_1) \cdots \hat{A}_n(t_n) \), only the fully-contracted terms appear in the perturbation series because these are the only ones that may have a nonzero expectation.
value with respect to the single-determinant wavefunction $|\Psi_0\rangle$. As an example of a fully-contracted term that arises in the expansion of $D_{j_1j_2;k_1k_2}$ at second order in $\hat{V}$, consider the quantity

$$L_1 \overset{\text{def}}{=} \frac{(-i)^4}{2!} \sum_{p_1,q_1,r_1,s_1} V_{p_1;q_1,r_1,s_1} V_{p_2;q_2,r_2,s_2} \int_{-\infty}^{\infty} d\tau_1 \int_{-\infty}^{\infty} d\tau_2 \langle \Psi_0 | k_1(t_1') k_2(t_1') p_1^\dagger(\tau_1) q_1^\dagger(\tau_1) s_1(\tau_1) p_2(\tau_2) q_2^\dagger(\tau_2) \rangle \times \langle \Psi_0 | \bar{k}_1(\tau_1) \bar{k}_2(\tau_1) \bar{p}_1(\tau_1) \bar{q}_1(\tau_1) \bar{s}_1(\tau_1) \bar{r}_1(\tau_1) \bar{p}_2(\tau_2) \bar{q}_2(\tau_2) \rangle. \quad (4.38)$$

The reader may verify that the indicated Wick contraction is positive in sign. The important fact to note is that

$$\hat{\kappa}(t') \hat{\mu}(t) = i \mathcal{D}^{\text{HF}}_{\mu,\kappa}(t; t) \quad (4.39)$$

is a matrix element of the 1-GF for the unperturbed system,

$$\mathcal{D}^{\text{HF}}_{\mu,\kappa}(t; t') = i \langle \Psi_0 | \mathcal{T}(\hat{\mu}(t) \hat{\kappa}(t')) | \Psi_0 \rangle \begin{cases} i \delta_{\mu,\kappa} \exp[i \epsilon_{\mu}(t - t')] & \text{if } t' \leq t \text{ and } \phi_k \text{ is occupied} \\ -i \delta_{\mu,\kappa} \exp[i \epsilon_{\mu}(t - t')] & \text{if } t' > t \text{ and } \phi_k \text{ is virtual} \\ 0 & \text{otherwise.} \end{cases} \quad (4.40)$$

The notation “HF” is simply a convenient moniker; $|\Psi_0\rangle$ can be any single-determinant reference state; of course, the labels “occupied” and “virtual” are relative to this determinant. This zeroth-order propagator will be represented using the diagram

$$i \mathcal{D}^{\text{HF}}_{\mu,\kappa}(t; t') = \begin{array}{c} \mu, t \\ \kappa, t' \end{array}.$$

Therefore, the zeroth-order propagator in the HF approximation is

$$\mathcal{D}^{\text{HF}}_{\mu,\kappa}(t; t') = i \delta_{\mu,\kappa} \exp[i \epsilon_{\mu}(t - t')] \begin{cases} 1 & \text{if } t' \leq t \text{ and } \phi_k \text{ is occupied} \\ -1 & \text{if } t' > t \text{ and } \phi_k \text{ is virtual} \\ 0 & \text{otherwise.} \end{cases} \quad (4.41)$$
The integrand in $L_1$ [Eq. (4.38)] thus consists of a product of $D_1^{\text{HF}}$ matrix elements.

A diagrammatic representation of $L_1$ is obtained from the template shown in Fig. 4.3, which consists of labeled points corresponding to the vertices $j_1, j_2, k_1,$ and $k_2$. Between these are located two diagram elements $\bullet\cdots\bullet$, each representing one factor of the perturbation and located at different vertical positions, so as to indicate different times $\tau_1$ and $\tau_2$ [see Eq. (4.38)]. Each $\bullet\cdots\bullet$ corresponds to a matrix element of $-i\hat{V}$ and is thus labeled by four dummy indices, representing the indices $\{p_1, q_1, r_1, s_1\}$ and $\{p_2, q_2, r_2, s_2\}$ from Eq. (4.38):

$$j_1, t_1 \quad j_2, t_2 \quad k_1, t'_1 \quad k_2, t'_2$$

$$= -i V_{j_1 j_2; k_1 k_2} \delta(t_1 - t'_1) \delta(t_2 - t'_2) \delta(t_1 - t_2). \quad (4.42)$$

The $\delta$-functions in this equation codify the fact that $\hat{V}$ is instantaneous.

Construction of the diagram corresponding to $L_1$ proceeds by drawing a directed $D_1^{\text{HF}}$ line, which represents a factor of $iD_1^{\text{HF}}$, between each pair of indices that are connected by a Wick contraction in Eq. (4.38). By convention, dummy summation indices $p_1, q_1, \cdots, r_m, s_m \cdots$ and dummy integration variables $\tau_1, \cdots, \tau_m$ are omitted from the diagram, the requisite summations and integrations being implied. Applying this prescription to Eq. (4.38) yields the diagrammatic representation

$$L_1 = \begin{array}{c}
\begin{array}{cc}
j_1 & j_2 \\
k_1 & k_2
\end{array}
\end{array}. \quad (4.43)$$
Consider another example:

\[
\mathcal{L}_2 \overset{\text{def}}{=} -\frac{(-i)^4}{2!} \sum_{p_1,q_1,r_1,s_1} V_{p_1 q_1 \tau_1 s_1} V_{p_2 q_2 \tau_2 s_2} \int_{-\infty}^{\infty} d\tau_1 \int_{-\infty}^{\infty} d\tau_2 \\
\times \langle \Psi_0 | k_1(t'_1) k_2(t'_2) \bar{p}_1(\tau_1) \bar{q}_1(\tau_1) \bar{s}_1(\tau_1) \bar{r}_1(\tau_1) \bar{p}_2(\tau_2) \bar{q}_2(\tau_2) \bar{s}_2(\tau_2) \bar{r}_2(\tau_2) j_2(t_2) j_1(t_1) | \Psi_0 \rangle .
\] (4.44)

Note that the overall sign of this collection of contractions is odd. Evidently, \( \mathcal{L}_1 \) and \( \mathcal{L}_2 \) represent distinct terms in the expansion of \( \overline{D}_2 \). Starting again from the framework in Fig. 4.3, one obtains

\[
\mathcal{L}_2 = \left. \right| k_1 \rightarrow \quad j_1 \ \quad j_2 \rightarrow \quad k_2 \right. (4.45)
\]

This diagram has the opposite time order as \( \mathcal{L}_1 \) but is topologically identical. The concept of topological distinctness is properly defined and discussed in the next subsection, but basically we mean that \( \mathcal{L}_1 \) and \( \mathcal{L}_2 \) have the same connectivity. Working backwards from diagrams to algebra, this implies that the diagrams in Eqs. (4.43) and (4.45) yield the same algebraic expression. This is an example of the diagrammatic
degeneracy that was mentioned in Subsection 4.4.1, as two separate terms in the perturbation series have produced equivalent diagrams, which implies that the two terms in question are numerically identical. In order to insure that the perturbation theory is performed consistently at any given order, one of these diagrams ought to be eliminated in favor of a weight factor multiplying the other, but first we must ascertain whether there are other terms in the $D_2$ perturbation series that lead to the same diagram.

In contrast to these examples, the diagram

\[ L_{3}^{\text{def}} = k_1 \quad j_1 \quad j_2 \quad k_2 \]

(4.46)

is topologically inequivalent to the ones given above. (It has a different connectivity.) Hence the corresponding algebraic expression does not reduce to $L_1$.

This argument can be generalized to an arbitrary propagator $D_n$. For a given $m$th-order term in the expansion, various deformations of the corresponding diagram lead to $m!$ topologically (hence algebraically) equivalent diagrams, one for each time order of the interactions. Each of these diagrams, however, arises from a different set of Wick contractions and therefore each constitutes a distinct contribution to $D_n$.

In practice, one includes only a single time order in the diagrammatic perturbation series (the Feynman convention), thus eliminating the factor of $1/m!$ appearing in Eq. (4.34). The result is a set of $m$th-order diagrams that each contain $m$ factors of the perturbation, but otherwise do not depend upon $m$ in any essential way. (This, in contrast to the explicit dependence on $m$ that was present before we exploited the
time-order symmetry to eliminate the factor of $1/m!$.) Removing this $m$-dependence allows us to express higher-order terms as products of lower-order ones, which is crucial to the renormalization of the one-particle propagator lines (Subsection 4.4.4).

Cancellation of the $1/m!$ factor can be worked out rigorously and algebraically, starting from Eq. (4.34), but the proof is lengthy and tedious. Using diagrams, this result is comparatively obvious. The corresponding result for the $\mathcal{D}_1$ perturbation series is well known, but the above generalization to arbitrary $\mathcal{D}_n$ has not previously been seen by the author.

The Feynman time-ordering convention amounts to a symmetry factor of $m!$ for $m$th-order diagrams. To illustrate another important symmetry factor for one-particle diagrams, we consider the four diagrams in Fig. 4.4, which may be generated from one another by twisting one or both interactions about the vertical propagator line. This operation represents a homeomorphism in three-dimensional space, hence each diagram in Fig. 4.4 has the same algebraic value. Using the framework suggested in Fig. 4.3 to assign indices to the interactions $\bullet\cdots\bullet$, however, each is readily found to arise from a different set of Wick contractions, and therefore each represents a separate term in the perturbation series. According to the usual convention, all four of these terms can be encoded into any one of them by associating a factor of $-2i\hat{V}$ with $\bullet\cdots\bullet$, rather than the usual factor of $-i\hat{V}$. An equivalent way of incorporating the appropriate symmetry factor, which proves useful in the more general case of $\mathcal{D}_n$, is to multiply each $m$th-order diagram by a factor of $2^m$, with the caveat that diagrams related by a $180^\circ$ twist of one or more interactions are not considered to be distinct terms in the perturbation series.

This last point is worth emphasizing. When they are considered to be embedded
in the plane of the page, each of the four diagrams Fig. 4.4 is topologically *inequivalent* to the other three. Referring to the framework introduced in Fig. 4.3, it is clear that this inequivalence in two-dimensional space is the topological manifestation of the algebraic labeling of the indices of $\hat{V}$ matrix elements, and therefore allows us to distinguish separate terms in the perturbation series. Diagrams that are distinct in a plane may or may not have the same algebraic value, depending upon whether they are topologically distinct in three-dimensional space.

We next introduce a renormalization of the many-particle perturbation theory that employs the exact one-particle propagator $\overline{D}_1$ in place of $D^{\text{HF}}_1$. Again the validity of this substitution can be verified algebraically, by isolating within the expansion of $\overline{D}_n$ [Eq. (4.34)] the complete, infinite-order expansion of $\overline{D}_1$, and then demonstrating that the expansion of $\overline{D}_n$ consists of all possible contractions of this infinite-order expansion. This is possible because of a generalized version\textsuperscript{82} of Wick’s theorem, in
which the operators $\hat{A}_1$ and $\hat{A}_2$ in Eq. (4.35) represent normal-ordered products of creation and annihilation operators.* Insofar as the diagram technique is simply a mnemonic for recalling Wick’s theorem, the result is obvious from the diagrams.

Having performed the indicated renormalization, $n$-particle perturbation theory diagrams are henceforth constructed using exact $D_1$ lines. These we denote as

$$i \tilde{D}_{\mu,\kappa}(t; t') = \begin{array}{c}
\mu, t \\
\kappa, t'
\end{array}.$$  

(4.47)

We will refer to the diagram element in Eq. (4.47) as a \textit{quasiparticle line}. A term in the expansion of $\tilde{D}_{j_1, \ldots, j_p; k_1, \ldots, k_p}$ thus consists of $p$ directed quasiparticle lines connected by certain pairwise interactions.

In its most primitive form, the pair interaction between quasiparticles is simply the “bare” interaction $\tilde{V}(t)$ whose diagrammatic representation was defined in Eq. (4.42). However, the formalism developed in the next two subsections is equally applicable to a renormalized (synonymously, “polarized” or “dressed”) interaction, and therefore we shall represent the pair interaction using the symbol $\bullet \cdots \bullet$ rather than $\bullet \cdots \bullet$, to indicate the possibility of renormalization. At present we need not specify the extent of this renormalization (see Subsection 4.5.1 for that), but an illustrative example is the well-known random phase approximation (RPA)\textsuperscript{23,45} for the high-density electron gas, in which the effective interaction is given by the sum of

*Obviously if $\hat{A}_1$ and $\hat{A}_2$ are normal-ordered then there is no need to consider contractions within the products $\hat{A}_i$. The important fact is that the generalized Wick’s theorem otherwise has exactly the same form as the original version, hence one may replace each $D_{HF}$ in the original perturbation series with $\tilde{D}_1$ to obtain a renormalized expansion.
so-called pair bubbles,

\[(\Updownarrow\uparrow\downarrow\downarrow)_{\text{RPA}} = \Updownarrow\uparrow\downarrow\downarrow + \Updownarrow\uparrow\downarrow\downarrow + \cdots. \tag{4.48}\]

More generally, a fully dressed interaction would include the sum of all possible self-energy diagrams that can be inserted into the bare interaction \(\bullet\cdots\bullet\):

\[\Updownarrow\uparrow\downarrow\downarrow = \Updownarrow\uparrow\downarrow\downarrow - \Updownarrow\uparrow\downarrow\downarrow + \Updownarrow\uparrow\downarrow\downarrow + \cdots. \tag{4.49}\]

Algebraically, each higher-order term on the right side of Eq. (4.48) or (4.49) arises from a Wick contraction that connects \(\hat{V}(\tau_1)\) to \(\hat{V}(\tau_2)\), that is, a contraction between some operator within the product \(\hat{p}_{1}^\dagger(\tau_1) \hat{q}_{1}^\dagger(\tau_1) \hat{s}_1(\tau_1) \hat{r}_1(\tau_1)\) and some operator within the product \(\hat{p}_{2}^\dagger(\tau_2) \hat{q}_{2}^\dagger(\tau_2) \hat{s}_2(\tau_2) \hat{r}_2(\tau_2)\). Physically, a single factor of \(\bullet\cdots\bullet\) accounts for intermediate particle-hole pairs that can be created within the medium by a cascade of virtual scattering events; \(\bullet\cdots\bullet\) thus accounts for the fact that the interaction \(\bullet\cdots\bullet\) is screened by the many-particle system.

An important point is that virtual particle-hole pairs created and subsequently annihilated by the dressed interaction, as in higher-order diagrams on the right side of Eq. (4.48) or Eq. (4.49), must propagate through some time interval \(t_2 - t_1\). In contrast to the bare interaction, which is characterized by a single time, the dressed interaction is thus associated with two separate time variables. As can be seen from Eqs. (4.48) and (4.49), however, each intermediate process originates and terminates with a factor of the bare interaction, which is instantaneous. Therefore

\[\frac{\delta(t_1 - t'_1) \delta(t_2 - t'_2)}{(t_1 - t'_1) (t_2 - t'_2)}. \tag{4.50}\]

The dressed interaction furthermore depends only on \(t_2 - t_1\), which follows because \(\overline{D}_{j,k}(t; t')\) depends only on \(t - t'\), for a Hamiltonian with no explicit time dependence.\(^{45}\)
As a matter of notation, both $\delta$-functions in Eq. (4.50), along with two of the $\delta$-functions in Eq. (4.42), can be omitted if we agree to associate only a single time variable with each vertex. In any case, each vertex retains two spin-orbital indices.

In the rest of this section, we will consider perturbative expansions of the $p$-GF in which each $p$-particle diagram consists of $p$ quasiparticle ($D_1$) lines connected by factors of a generic (possibly renormalized) interaction $\cdots$. As usual, unlinked diagrams are excluded from the expansion, as a result of the Kobe’s $p$-particle generalization$^{89}$ of the familiar linked-cluster theorem. On this point, let us be very clear with our language. We have already introduced the term connected to mean additively separable, which in the context of cumulant theory is a standard usage of the word. Where many-body perturbation theory is concerned, however, the term connected is sometimes used as a synonym for linked, in the sense of the linked-cluster theorem. We will use only the term linked to indicate this latter meaning. In truth we shall have little use for the latter term, since unlinked diagrams do not contribute to the perturbation theory. For $p \geq 2$, unconnected terms certainly do contribute to the perturbation series, specifically to $\Sigma_p$. Following Lindgren and Morrison,$^{81}$ we formalize these remarks in the following definition.

**Definition 4.1.** A diagram is said to be connected if it cannot be separated into pieces without severing an interaction or a propagator line. A diagram is closed if it does not contain any incoming or outgoing propagator lines. A diagram that contains a closed, unconnected part is said to be unlinked.

The distinction between linked and connected diagrams is illustrated in Fig. 4.5. Unlinked diagrams such as (a) and (b) are absent from the perturbation theory. Linked
Figure 4.5: Two-particle diagrams at second order in $\hat{V}$. Diagrams (a) and (b) are unlinked, diagram (c) is linked but not connected, and (d) is both linked and connected.

but unconnected terms such as the one in Fig. 4.5(c) are present and contribute to $\Sigma_2$. Linked, connected terms like the one in Fig. 4.5(d) contribute to $\Delta_2$. By definition, the sum of all linked, unconnected diagrams equals $\Sigma_p$, for which we possess an exact formula in terms of lower-order propagators. In what follows we consider only linked, connected $p$-particle diagrams, whose sum is $\Delta_p$.

In our previous discussion of the linked, connected diagrams $L_1$ and $L_2$, we alluded—somewhat imprecisely—to two important ideas. First is the fact that diagrams with distinct “connectivities” (of dummy internal vertex labels) must correspond to different terms in the perturbation series. Secondly, we pointed out that upon removing the internal labels, two diagrams correspond to the same algebraic expression if there exists a continuous deformation that transforms one diagram into another. (Diagrams that are topologically equivalent in three-dimensional space are
also algebraically equivalent.) The second statement should be obvious to the reader who has studied the rules for translating diagrams into algebra, but the identification of distinct terms in the perturbation series requires that we nail down the notion of connectivity a bit more precisely. In the case of the \( p \)-particle perturbation series for \( p > 2 \), the need for precision is even more urgent. Lacking a systematic plan of attack, one is faced with a bewildering panorama of complicated diagrams, whose equivalence may not be obvious. Thus, as a prelude to our systematic algorithm for generating ladder-type diagrams, the next subsection is devoted to precise definitions of connectivity, diagrams, topological equivalence, and related concepts.

### 4.4.3 Connectivity and topology of diagrams

Having completed our whirlwind survey of many-body perturbation theory for Green’s functions, we are now ready to tackle the main objective of this section: establishing a rigorous, systematic foundation for diagrammatic perturbation theory as applied to the three- and four-particle Green’s functions. Specifically, we require a procedure that will generate all unique diagrams (within a given, well-defined approximation to the perturbation series), without redundancies. A symmetry weight for each of these diagrams must also be provided. This program is carried out for \( \Delta_3 \) in Subsection 4.4.4 and for \( \Delta_4 \) in Subsection 4.4.5, for a particular approximation to the perturbation series. Before that, however, we need to clarify some issues related to diagrammatic presentations of the three- and four-particle perturbation series, which are considerably more complicated than either the one- or the two-particle case. Such is the business of the present subsection.

Generalizing the two-particle diagram template introduced in Fig. 4.3 to the case
of \( p \) particles, \( n \)th (renormalized) order terms in the expansion of \( \overline{\mathcal{D}}_{j_1, \ldots, j_p; k_1, \ldots, k_p} \) are generated from the framework of labeled vertices shown in Fig. 4.6 by connecting these vertices using a total of \( p \) directed quasiparticle lines. These lines should originate on the vertices labeled \( k_1, \ldots, k_p \) and terminate on those labeled \( j_1, \ldots, j_p \), with a single quasiparticle line passing through each vertex. Different quasiparticle lines may pass through different numbers of vertices, subject only to the constraint that the resulting diagram must be connected. For brevity, we will typically suppress the time variables \( t_m \) and \( t'_m \) that are indicated in Fig. 4.6.

The “connect-the-dots” paradigm suggested by Fig. 4.6 provides one method

\[ j_1, t_1 \quad j_2, t_2 \quad \cdots \quad j_p, t_p \]
\[ \bullet \quad \bullet \quad \cdots \quad \bullet \]
\[ \begin{array}{c}
\eta_{j_1}^{R} \quad \cdots \quad \eta_{j_p}^{R} \\
\vdots \\
\eta_{j_1}^{L} \quad \cdots \quad \eta_{j_p}^{L}
\end{array} \]
\[ \text{Internal part} \]
\[ \bullet \quad \bullet \quad \cdots \quad \bullet \]
\[ k_1, t'_1 \quad k_2, t'_2 \quad \cdots \quad k_p, t'_p \]

\[ \begin{array}{c}
\eta_{k_1}^{L} \quad \cdots \quad \eta_{k_p}^{L} \\
\vdots \\
\eta_{k_1}^{R} \quad \cdots \quad \eta_{k_p}^{R}
\end{array} \]

\[ j_1, t_1 \quad j_2, t_2 \quad \cdots \quad j_p, t_p \]

\[ \begin{array}{c}
\eta_{j_1}^{R} \quad \cdots \quad \eta_{j_p}^{R} \\
\vdots \\
\eta_{j_1}^{L} \quad \cdots \quad \eta_{j_p}^{L}
\end{array} \]

\[ \text{Internal part} \]

\[ \bullet \quad \bullet \quad \cdots \quad \bullet \]

\[ k_1, t'_1 \quad k_2, t'_2 \quad \cdots \quad k_p, t'_p \]

Figure 4.6: Framework for constructing \( n \)th-(renormalized)-order terms in the expansion of \( \overline{\mathcal{D}}_{j_1, \ldots, j_p; k_1, \ldots, k_p} \).

\[ \begin{array}{c}
\eta_{j_1}^{R} \quad \cdots \quad \eta_{j_p}^{R} \\
\vdots \\
\eta_{j_1}^{L} \quad \cdots \quad \eta_{j_p}^{L}
\end{array} \]

\[ \text{Internal part} \]

\[ \bullet \quad \bullet \quad \cdots \quad \bullet \]

\[ k_1, t'_1 \quad k_2, t'_2 \quad \cdots \quad k_p, t'_p \]

\[ \begin{array}{c}
\eta_{k_1}^{L} \quad \cdots \quad \eta_{k_p}^{L} \\
\vdots \\
\eta_{k_1}^{R} \quad \cdots \quad \eta_{k_p}^{R}
\end{array} \]

\[ \text{Internal part} \]

\[ \bullet \quad \bullet \quad \cdots \quad \bullet \]

\[ j_1, t_1 \quad j_2, t_2 \quad \cdots \quad j_p, t_p \]

\[ \begin{array}{c}
\eta_{j_1}^{R} \quad \cdots \quad \eta_{j_p}^{R} \\
\vdots \\
\eta_{j_1}^{L} \quad \cdots \quad \eta_{j_p}^{L}
\end{array} \]

\[ \text{Internal part} \]

\[ \bullet \quad \bullet \quad \cdots \quad \bullet \]

\[ k_1, t'_1 \quad k_2, t'_2 \quad \cdots \quad k_p, t'_p \]

\[ \begin{array}{c}
\eta_{k_1}^{L} \quad \cdots \quad \eta_{k_p}^{L} \\
\vdots \\
\eta_{k_1}^{R} \quad \cdots \quad \eta_{k_p}^{R}
\end{array} \]

\[ \text{Internal part} \]

\[ \bullet \quad \bullet \quad \cdots \quad \bullet \]

\[ j_1, t_1 \quad j_2, t_2 \quad \cdots \quad j_p, t_p \]

\[ \begin{array}{c}
\eta_{j_1}^{R} \quad \cdots \quad \eta_{j_p}^{R} \\
\vdots \\
\eta_{j_1}^{L} \quad \cdots \quad \eta_{j_p}^{L}
\end{array} \]

\[ \text{Internal part} \]

\[ \bullet \quad \bullet \quad \cdots \quad \bullet \]

\[ k_1, t'_1 \quad k_2, t'_2 \quad \cdots \quad k_p, t'_p \]

\[ \begin{array}{c}
\eta_{k_1}^{L} \quad \cdots \quad \eta_{k_p}^{L} \\
\vdots \\
\eta_{k_1}^{R} \quad \cdots \quad \eta_{k_p}^{R}
\end{array} \]

\[ \text{Internal part} \]

\[ \bullet \quad \bullet \quad \cdots \quad \bullet \]
for obtaining all distinct $p$-particle connectivities. For the sake of completeness we mention that an alternative procedure has been advanced by Lande and Smith,\textsuperscript{90–93} based upon enumeration of the possible ways in which a given pair of $p$-particle connectivities can be connected together in order to yield a new $p$-particle connectivity. While this procedure generates all of the diagrams, it does not do so uniquely, and we shall not employ the Lande-Smith formalism.

\textbf{Definition 4.2.} Consider an $n$th-order term in the expansion of $\overline{D}_{j_1,\ldots,j_p;k_1,\ldots,k_p}$, constructed from the framework in Fig. 4.6 as described above. Let $S_m = (j_\lambda, \ldots, k_\mu)$ denote the ordered sequence of vertices through which the $m$th quasiparticle line passes. Then the set of sequences $\{S_1, \ldots, S_p\}$ defines the connectivity of this particular term in the perturbation series. The order of the $S_m$ within this set is irrelevant; connectivity is defined only up to permutations of these sequences.

\textbf{Definition 4.3.} Two $p$-particle connectivities $\{S_1, \ldots, S_p\}$ and $\{S'_1, \ldots, S'_p\}$ are said to be \textit{distinct} if $\{S_1, \ldots, S_p\} \neq \{S'_1, \ldots, S'_p\}$ as sets.

The fact that connectivity is defined only up to a permutation of the sequences $S_m$ is a consequence of the antisymmetry of $p$-GF, whence $\overline{D}_{j_1,\ldots,j_p;k_1,\ldots,k_p}$ is unchanged if we permute both $j_\lambda \leftrightarrow j_\mu$ and $k_\lambda \leftrightarrow k_\mu$. Within the perturbation series, this symmetry manifests itself as the fact that any fully-contracted term from Wick’s theorem is unchanged following two such permutations. Renormalization of the perturbation series in terms of exact one-particle propagators does not alter this symmetry.

Nothing in the definition of connectivity specifies the paths that the quasiparticle lines must take between vertices, and as a result there exist infinitely many diagrammatic representations of the perturbation series. In the one- and two-particle
cases, this ambiguity is removed via simple conventions for diagram construction, but things are more complicated in the many-particle case. To sort through these various representations it is useful to discuss connectivities as topological objects.

**Definition 4.4.** A *diagrammatic presentation* of a connectivity $C = \{S_1, \ldots, S_p\}$ (equivalently, a diagram having connectivity $C$) is an embedding of $p$ lines, with labeled endpoints, as a topological object in three-dimensional space, in which the $m$th line passes in sequence through the vertices $S_m$. Once this embedding has been made, all internal vertex labels $v_n^L$ and $v_n^R$ are deleted.

A connectivity, as defined above, is not a topological object, but topological equivalence of connectivities can be defined by means of their diagrammatic presentations.

**Definition 4.5.** Two $p$-particle connectivities $C_1$ and $C_2$ are said to be *diagrammatically equivalent*, denoted $C_1 \simeq C_2$, if there exist diagrammatic presentations of $C_1$ and $C_2$ that are topologically identical (can be deformed into one another in a continuous fashion, without severing any propagator lines or interactions) in three-dimensional space.

Although there are infinitely many diagrams (embeddings) sharing a specified connectivity, in practice the ones constructed using the shortest possible lines to connect the appropriate vertices are usually best for demonstrating topological (in)equivalence. As an example, consider the two diagrams shown in Fig. 4.7, both of which are embeddings of the connectivity $\{S_1, S_2, S_3\}$, with $S_1 = (k_1, v_1^L, v_2^L, v_3^L, j_1)$, $S_2 = (k_2, v_3^R, j_2)$, and $S_3 = (k_3, v_1^L, v_2^R, j_3)$. Clearly the diagram on the right in Fig. 4.7 is the preferred choice of embedding. The reader may verify that $\{S_1', S_2', S_3'\}$, where
Figure 4.7: Two topologically inequivalent embeddings of the same three-particle connectivity.

\[ S_0^1 = (k_2, v_3^1, j_2), \quad S_0^2 = (k_1, v_1^2, v_3^1, j_2), \quad \text{and} \quad S_0^3 = (k_3, v_1^3, v_2^3, j_3), \]

is distinct from \( \{S_1, S_2, S_3\} \) above, but is nevertheless diagrammatically equivalent. The embedding of \( \{S_1', S_2', S_3'\} \) shown in Fig. 4.8 makes this clear.

Given the rules for translating diagrams into algebraic expressions\(^{23,45}\) (either in the time-dependent or the time-independent theory), it is obvious that if \( C_1 \simeq C_2 \) then these two connectivities must yield the same algebraic expression. While there may be a large number of distinct connectivities at a given order in perturbation theory, in general each is equivalent to one of a much smaller set of diagrammatically inequivalent connectivities. Furthermore, many inequivalent connectivities will differ only by a permutation of the endpoint labels \( j_1, \ldots, j_p \) and \( k_1, \ldots, k_p \).

**Definition 4.6.** Two distinct connectivities \( C = \{S_1, \ldots, S_p\} \) and \( C' = \{S_1', \ldots, S_p'\} \) are said to be independent if the sequences \( \{S_m'\} \) cannot be obtained from the sequences \( \{S_m\} \) via permutations of the vertex labels \( j_1, \ldots, j_p \) and \( k_1, \ldots, k_p \).

An equivalent statement of this definition is noteworthy, though hardly substantial enough to warrant designation as a theorem.
Proposition 4.1. Two connectivities $C$ and $C'$ are not independent if there exist diagrammatic presentations of $C$ and $C'$ that are topologically equivalent when the endpoint labels are deleted.

When connected together, the interaction vertices $\{v^n_R\}$ and $\{v^n_L\}$ in Fig. 4.6—without the endpoint vertices—make up what we term the internal part of the vertex framework (so labeled in Fig. 4.6). Connectivities are independent if their internal parts are topologically inequivalent. If the internal parts are topologically equivalent, then the connectivities differ only by a permutation of the endpoint labels. Clearly we need enumerate only the independent connectivities.

Finally, note that explicit self-energy insertions within the quasiparticle lines are forbidden in the renormalized perturbation theory. (Because we employ exact one-particle propagators, inclusion of such self-energy parts would overcount terms in the perturbation series.) For example, the diagram in Fig. 4.9(a) does not appear in the renormalized perturbation series. This does not imply that a single quasiparticle line cannot connect both vertices of the same interaction, but in such cases these...
vertices must be separated by one or more additional interactions that couple two
different quasiparticle lines. An example of this sort is the vertex correction shown in
Fig. 4.9(b), which is an allowed diagram. More complicated variants on this theme
are of course possible, but the general rule is clear. We state this rule in the form of
the following proposition.

**Proposition 4.2.** Consider a connectivity \( \{S_1, S_2, \ldots \} \) in which the vertices \( v_{n}^{L} \) and
\( v_{n}^{R} \) appear within the same sequence, \( S_m \) say. This connectivity is forbidden (excluded
from the perturbation series) unless \( v_{n}^{L} \) and \( v_{n}^{R} \) are separated within \( S_m \) by another
vertex \( v_{n}^{L'} (v_{n}^{R'}) \), such that \( v_{n}^{R'} (v_{n}^{L'}) \) is not a member of the sequence \( S_m \).

### 4.4.4 Ladder-type expansions for the three-particle propagator

With the basic formalism in place, we are now ready to derive specific diagram rules for
the three- and four-particle perturbation series. The general case is quite complicated
and we thus introduce an approximation that neglects certain categories of diagrams.
It is well known that electron correlation is dominated by pair correlations,\(^{26,27,94–98}\)
despite the long-range nature of the Coulomb interaction, and thus an appropriate
starting point for decoupling the propagator or density matrix hierarchies is a *ladder*
approximation for the $p$-GF or $p$-RDM. One example of a three-particle ladder, which clearly lives up to its name, is shown in Fig. 4.8; the general definition of a ladder diagram is given below. It is evident that the diagram in Fig. 4.8 accounts for multiple scattering events between pairs of particles but neglects simultaneous three-particle correlations, which is a general feature of the ladder approximation.

We have demonstrated that merely connecting the dots in Fig. 4.6 to generate distinct $n$th-order connectivities results in numerous connectivities that are topologically (and therefore algebraically) equivalent. We wish to eliminate this redundancy, for three-particle ladder diagrams, by means of a well-defined algorithm that accounts for all $n$th-order connectivities possessing a ladder-type embedding. In addition, this algorithm automatically accounts for connectivities such as the one shown in Fig. 4.10, which might not immediately appear to be ladder diagrams, but are nonetheless equivalent to more traditional-looking three-particle ladders. First let us define precisely what is meant by a ladder connectivity.
**Definition 4.7.** A $p$-particle connectivity is called a *$p$-particle ladder* if it possesses a connected diagrammatic presentation in which (i) the interactions do not overlap with one another in time (that is, in the vertical direction), and (ii) no quasiparticle lines run backward in time.

The presentations in Figs. 4.8 and 4.10 clearly satisfy this definition. Diagrams that do not satisfy this definition include vertex corrections such as

\[
\begin{align*}
& k_1 j_1 k_2 j_2 k_3 j_3 \\
& \equiv \\
& k_1 j_1 k_2 j_2 k_3 j_3
\end{align*}
\]  \hspace{1cm} (4.51)

and so-called particle-hole scattering diagrams such as

\[
\begin{align*}
& k_1 j_1 k_2 j_2 k_3 j_3 \\
& \equiv \\
& k_1 j_1 k_2 j_2 k_3 j_3
\end{align*}
\]  \hspace{1cm} (4.52)

A slight modification of the vertex framework introduced previously yields a counting algorithm for the three-particle ladders. This modified framework is shown in Fig. 4.11(a) and consists again of $n$ time-ordered interactions, drawn horizontally so that they do not overlap in time, consistent with **Definition 4.7.** For clarity we have not included the vertex labels $\{v^R_n\}$ and $\{v^L_n\}$ in this figure, but we take the vertices to be labeled as in Fig. 4.6. Let $t^R_\Lambda$ and $t^L_\Lambda$ be the time arguments associated with the vertices $v^R_\Lambda$ and $v^L_\Lambda$, respectively. Ladders constructed from Fig. 4.11 are assumed to
Figure 4.11: Diagrammatic representations of various steps in the algorithmic construction of \( n \)th-order, three-particle ladders. See the text for an explanation of each step.

have the time order

\[
t_1^{A_1} < t_1^{A_1'} < t_2^{A_2} < t_2^{A_2'} < \cdots < t_n^{A_n} < t_n^{A_n'},
\]

where each \( A_\lambda, A_\lambda' \in \{\text{R, L}\} \), with \( A_\lambda \neq A_\lambda' \). The necessity of this time order follows immediately from Definition 4.7: while the relative time order of \( v_R^\lambda \) and \( v_L^\lambda \) is unspecified, these vertices must appear consecutively in time, else the result is not a ladder diagram. As a practical matter, this means that we consider only topological deformations that preserve the time order of the interactions. We shall return to this point at the conclusion of Subsection 4.4.5.

Construction of three-particle ladder diagrams begins, as shown in Fig. 4.11(a), by drawing three quasiparticle lines extending from the vertices \( k_1, k_2, \) and \( k_3 \) to three of the interaction vertices. For generality, the quasiparticle line originating at \( k_1 \) is attached to \( v_m^R \), for arbitrary \( m \) such that \( 2 \leq m \leq n \). The case where this line attaches to \( v_m^L \) will be considered separately. We have less freedom, within the
ladder approximation, to attach the remaining two quasiparticle lines, because given
the arrangement in Fig. 4.11 no quasiparticle line may run backward in time, else the
resulting diagram will not be a ladder. Hence the remaining two quasiparticle lines
must attach to $v_1^L$ and $v_1^R$.

Partially-connected frameworks, such as those shown in Figs. 4.11(a)–(c), will be
called nascent connectivities. Starting from the nascent connectivity in Fig. 4.11(a),
we intend to build up the internal part of a ladder diagram in a systematic manner,
accounting for all topologically equivalent possibilities that arise each time another
interaction is appended to the nascent connectivity. This procedure is carried out
graphically, meaning that we have made an implicit choice of embedding for each
connectivity. Consequently, within the constructive algorithms presented in this sub-
section and the next, the concept of diagrammatic equivalence is the same as the
notion of topological equivalence in three-dimensional space.

For each $n$th-order internal part constructed by our algorithm, every permutation
of the labels $k_1$, $k_2$, and $k_3$ yields a distinct connectivity. We account for these six
non-independent connectivities with an operator $\hat{P}_3$, as indicated in Fig. 4.11. The
presence of this operator compensates for our arbitrary labeling of the three incoming
quasiparticle lines in Fig. 4.11.

The next step in the counting algorithm is to extend the $k_2$ and $k_3$ quasiparticle
lines in order to incorporate interactions 2 through $m - 1$. Each of these interactions
must be appended to the nascent connectivity in order, otherwise there is no way
to complete the connectivity without backward-propagating quasiparticle lines that
would invalidate the nascent connectivity as a candidate to become a ladder diagram.
Each interaction can, however, be connected to the one beneath it in two different
Figure 4.12: Two possibilities for connecting the second interaction in Fig. 4.11(a): (a) untwisted connection, and (b) twisted connection.

ways, one twisted and one untwisted. These possibilities are illustrated, for the second interaction, in Fig. 4.12.

The untwisted nascent connectivity, Fig. 4.12(a), is a diagrammatic presentation of \( \{S_1, S_2, S_3\} \) with

\[
S_1 = (k_1, v^l_m, \ldots) \quad (4.54a)
\]
\[
S_2 = (k_2, v^l_1, v^l_2, \ldots) \quad (4.54b)
\]
\[
S_3 = (k_3, v^r_1, v^r_2, \ldots) . \quad (4.54c)
\]

The twisted variant, Fig. 4.12(b), has nascent connectivity \( \{S'_1, S'_2, S'_3\} \), with

\[
S'_1 = (k_1, v^l_m, \ldots) \quad (4.55a)
\]
\[
S'_2 = (k_2, v^r_1, v^r_2, \ldots) \quad (4.55b)
\]
\[
S'_3 = (k_3, v^r_1, v^l_2, \ldots) . \quad (4.55c)
\]

For any connectivity, the interchange \( v^l_m \leftrightarrow v^r_m \) leads to a distinct but diagrammatically (that is, topologically) equivalent connectivity. Consequently, if one completes
each pair of sequences $S_m$ and $S'_m$ above in exactly the same fashion, the resulting connectivities will be diagrammatically equivalent up to a permutation of $k_2$ and $k_3$. That is,

$$\{S_1, S_2, S_3\} \simeq (k_2 \ k_3) \ \{S'_1, S'_2, S'_3\},$$

(4.56)

where $(k_2 \ k_3)$ denotes the operator that transposes $k_2$ and $k_3$.

Taking into account all non-independent connectivities that can be generated from $\{S_1, S_2, S_3\}$ and $\{S'_1, S'_2, S'_3\}$ via permutations of the endpoint labels, one sees that these connectivities appear in the perturbation series as $\overline{P}_3 \hat{P}_3 \{S_1, S_2, S_3\}$ and $\overline{P}_3 \hat{P}_3 \{S'_1, S'_2, S'_3\}$. Since

$$\overline{P}_3 \hat{P}_3 \{j_2 \ j_3\} \{S'_1, S'_2, S'_3\} \simeq \overline{P}_3 \hat{P}_3 \{S'_1, S'_2, S'_3\},$$

(4.57)

each complete connectivity generated from the nascent one in Fig. 4.12(b) is topologically equivalent to a connectivity obtained from Fig. 4.12(a). [We emphasize, however, that connectivities generated from 4.12(a) and 4.12(b) are distinct, in the sense of Definition 4.3.] As a result of this equivalence, we may discard connectivities generated from Fig. 4.12(b), substituting instead a multiplicative symmetry factor of 2 associated with Fig. 4.12(a).

This result clearly generalizes to interactions $3, \ldots, m - 1$ in Fig. 4.11, since each must be attached to the $k_2$ and $k_3$ quasiparticle lines. Thus the untwisted ladder framework, obtained by connecting these interactions and depicted in Fig. 4.11(b), carries an overall symmetry factor of $2^{m-2}$.

Next we must attach the $m$th interaction to the $k_2$-$k_3$ ladder. The free vertex $v^m$ of the $m$th interaction may be attached to either the left or the right side of this ladder, with the former choice shown in Fig. 4.11(c). In the case that this interaction
is attached to the right side of the ladder—which generates an entirely distinct set of connectivities—the resulting framework is topologically equivalent to Fig. 4.11(c), following a permutation of $k_2$ and $k_3$. Arguing as before, a factor of 2 accounts for the possibility of attaching $v^m_k$ to the right side of the $k_2$-$k_3$ ladder. Hence the symmetry factor associated with the nascent connectivity in Fig. 4.11(c) is $2^{m-1}$.

It is convenient to backtrack at this point, in order to consider the case where the $k_1$ quasiparticle line is connected to $v^m_R$ rather than $v^m_L$. Obviously this does not affect construction of the $k_2$-$k_3$ ladder in Fig. 4.11(b), yet a distinct set of connectivities is obtained because $S_1 = (k_1, v^m_R, \ldots)$ in this case, rather than $S_1 = (k_1, v^m_L, \ldots)$ [cf. Eqs. (4.54a) and (4.55a)]. Again, though, the resulting nascent connectivity is topologically equivalent, through $m$ interactions, to the one in Fig. 4.11(c), and can be neglected simply by increasing the symmetry factor for Fig. 4.11(c) to $2^m$.

It remains to connect the final $n - m$ interactions, starting with the $(m + 1)$st. There are \( \binom{3}{2} \) pairs of vertices to which we might attach each given interaction, and for each pair of vertices the connection can be made in either a twisted or an untwisted fashion. By now it should be clear that the twisted variants can be neglected if we multiply their untwisted counterparts by a factor of 2 for each interaction. The set of untwisted diagrams is, by construction, precisely the set of all diagrammatically distinct three-particle ladders. The remaining $n - m$ interactions contribute a factor of $2^{n-m}$ to the symmetry coefficient, resulting in an overall symmetry factor of $2^m \times 2^{n-m} = 2^n$ for each diagrammatically distinct, $n$th-order ladder. Note that this result is independent of the index $m$ where the $k_1$ quasiparticle line was originally attached.

Finally, an antisymmetrizer $\overline{P}_3$ must be included to permute $j_1$, $j_2$, and $j_3$, since the definition of a ladder involves only the internal part of the connectivity and says
nothing about the manner in which this internal part connects to the endpoint vertices. Hence the sum of all \( n \)th-order ladder diagrams in the expansion of \( \Delta_3 \) equals the action of \( 2^n \hat{\mathcal{P}}_3 \hat{\mathcal{P}}_3 \) on the sum of independent three-particle ladders. Our desired symmetry coefficient is \( 2^n \).

An ancillary result that follows from the above algorithm is the precise number of independent three-particle ladders. Observe that \( \binom{3}{2}^{n-m} \) distinct internal connectivities can be generated from the nascent connectivity in Fig. 4.11(c). Summation over possible values of \( m \) yields the number of independent ladders at \( n \)th order:

\[
\sum_{m=2}^{n} \binom{3}{2}^{n-m} = \frac{1}{2} \left( 3^{n-1} - 1 \right),
\]

which is valid for \( n \geq 2 \). (There are no first-order, three-particle ladders because a ladder must be connected, by definition.) In a sense, the precise number of ladder diagrams is not so important since we have an algorithm for the construction of such diagrams, but the total number of anticipated diagrams provides a usual check of the application of this algorithm.

For the reader’s convenience, the following theorem recapitulates the main results derived in this subsection.

**Theorem 4.1.** At \( n \)th order in the dressed interaction, there exist \( (3^{n-1} - 1)/2 \) independent three-particle ladder connectivities (that is, ladder diagrams that are diagrammatically inequivalent when unlabeled). Each is diagrammatically equivalent to \( 2^n \) distinct ladder connectivities in the \( \Delta_3 \) perturbation series. The total contribution to \( \Delta_3 \) from \( n \)th-order ladder diagrams equals \( 2^n \hat{\mathcal{P}}_3 \hat{\mathcal{P}}_3 \) acting on the sum of independent three-particle ladders.

As an illustration, we display in Figs. 4.13 and 4.14 all of the independent three-
Figure 4.13: All independent third-order, three-particle ladders.

Figure 4.14: All independent fourth-order, three-particle ladders.

particle ladders at third and fourth order, respectively. These are listed in the order that they are generated by the algorithm described above, and for conciseness they are drawn using only single lines (e.g., •¬¬¬¬ rather than •¬¬¬¬), which clearly does not affect their topology.

With the machinery developed in this subsection, we can also derive the symmetry coefficient for two-particle ladders. Actually this is very simple because there are only two possible ways to append each successive interaction, one twisted and another untwisted. The result is

**Proposition 4.3.** The total contribution to $\overline{\mathbf{\Delta}}_2$ from $n$th-order ladder diagrams is equal to the operator $2^{n-1} \overline{P}_2 \overline{P}_2$ acting on the lone diagrammatically inequivalent two-particle ladder diagram.
4.4.5 Ladder-type expansions for the four-particle propagator

Systematic construction of the four-particle ladder diagrams proceeds along similar lines, but now there are two separate cases to consider. Starting from a sequence of $n$ time-ordered interactions (Fig. 4.15), one must again attach two quasiparticle lines (labeled $k_3$ and $k_4$, say) to the first interaction, while the remaining two quasiparticle lines (labeled $k_1$ and $k_2$) may be attached arbitrarily. We first consider the case where these two lines are attached to the same interaction, say the $m$th, with $2 \leq m < n$. The corresponding nascent connectivity is shown in Fig. 4.15(a). Note that $m$ must be strictly less than $n$ in this case, because when $m = n$ there is no way to build a connected diagram without backward-propagating quasiparticle lines. Analogous to the three-particle case, we indicate a sum over signed permutations of $k_1, \ldots, k_4$ with the operator $\hat{P}_4$; this antisymmetrization justifies our arbitrary choice of lower endpoint labels.
The next step is to build up a ladder sequence from the first \( m - 1 \) interactions, just as we did in the three-particle construction. The resulting framework is shown in Fig. 4.15(b) and carries a symmetry factor of \( 2^{m-2} \) to account for twists within the ladder.

We now come to the \( m \)th interaction. There are \( \binom{4}{2} \) pairs of vertices to which we may attach this interaction; as usual we include a factor of 2 in each case and thereby consider only connectivities that are untwisted through the first \( m + 1 \) interactions. There are six such, and they are depicted in Fig. 4.16. The first two are unconnected, while the latter four are seen to be permutations of a single nascent connectivity. Through the first \( m + 1 \) interactions, then, all distinct four-particle ladder connectivities are diagrammatically equivalent to either (a), (b), or (c) in Fig. 4.16. Of these, (a) and (b) each carry a symmetry factor of \( 2^{m-1} \), while the coefficient for (c) is \( 2^{m+1} \).
The extra factor of four in the latter case accounts for the topologically equivalent connectivities in Figs. 4.16(d)–(f).

Any diagram generated from Fig. 4.16(c) is necessarily connected. Untwisted connection of the remaining \( n - m - 1 \) interactions exhausts all inequivalent connectivities that can be built upon Fig. 4.16(c), while a factor of \( 2^{n-m-1} \) accounts for all of the equivalent, twisted connectivities emanating from this nascent connectivity. The overall symmetry factor for independent ladders constructed from Fig. 4.16(c) is therefore \( 2^{n-m-1} \times 2^{m+1} = 2^n \), independent of \( m \).

The situation is somewhat different for the as-yet-unconnected possibilities in Figs. 4.16(a) and 4.16(b). Appending the \((m + 2)\)nd interaction to either of these nascent connectivities results in a set of six diagrams analogous to the ones in Fig. 4.16. Two of these are unconnected, while the other four are connected and also topologically equivalent to one another. The four connected diagrams can be treated as above, the result being a symmetry factor of \( 2^n \) once the connectivity has been completed. For the two unconnected diagrams, attaching the \((m + 3)\)rd interaction leads to another six possibilities, and so on, until finally we come to attach the \( n \)th interaction. This must be done in such a way that the entire diagram becomes connected, so we have in this final case only the four topologically equivalent, connected possibilities analogous to Figs. 4.16(c)–(f). As with the other cases, the result is a diagram weight of \( 2^n \) for each completed connectivity.

To obtain the number of independent four-particle ladders, in the case that all four propagator lines initially connect to only two interaction rungs, let \( q \) index the interaction at which the \( k_1-k_2 \) and \( k_3-k_4 \) halves of the diagram become connected \((m + 1 \leq q \leq n)\). For example, \( q = m + 1 \) in Fig. 4.16(c), but connectivities built
upon Figs. 4.16(a) and 4.16(b) do not become connected until some $q > m + 1$. A
diagram that becomes connected at the $q$th interaction ultimately generates $\binom{4}{2}^{n-q}$
topologically inequivalent ladder diagrams.

Two of the nascent connectivities in Fig. 4.16 are unconnected through the first $m$
interactions, and the number of unconnected nascent connectivities doubles each time
another interaction is appended. Hence there are $2^{q-m-1}$ topologically inequivalent
nascent connectivities that are unconnected below the $q$th interaction. Summation
over the possible values for $q$ and $m$ yields the number of independent four-particle
ladders, for the first of the two cases that we must consider. The result is

$$\sum_{m=2}^{n-1} \sum_{q=m+1}^{n} 2^{q-m-1} \binom{4}{2}^{n-q} = \frac{4}{10}(6^{n-2} - 1) - \frac{1}{2}(2^{n-2} - 1)$$  \hspace{1cm} (4.59)

for each $n \geq 3$.

We have not yet considered the case in which the $k_1$ and $k_2$ quasiparticle lines
initially connect to different interactions. (The $k_3$ and $k_4$ interactions must still attach
to $v_1^L$ and $v_1^R$.) A framework for discussing this scenario is depicted in Fig. 4.17. We
attach the $k_1$ line to $v_r^L$ and the $k_2$ line to $v_m^L$, for $2 \leq m < r \leq n$. Cases where $m > r$
amount to a permutation of $k_1$ and $k_2$, and the operator $\hat{P}_4$ indicated in Fig. 4.17
accounts for all such permutations. Cases in which the $k_1$ line, the $k_2$ line, or both
attach at the right side of an interaction, rather than the left, will be incorporated as
we proceed.

Through $r - 1$ interactions, the four-particle nascent connectivity in Fig. 4.17 is
the same as a three-particle nascent connectivity that was considered previously, in
Fig. 4.11(a). (The fact that the endpoint labels differ is unimportant since we con-
sider all permutations of these labels.) The results of Subsection 4.4.4 thus furnish
the appropriate symmetry coefficient through $r - 1$ interactions, namely $2^{r-1}$. By construction, this factor accounts for all diagrammatically equivalent nascent connectivities, so we may assume that the first $r - 1$ interactions in Fig. 4.17 are connected without topological redundancies. As a result, each of the $\binom{3}{2}$ ways in which the $r$th interaction can be connected to the ones below it generates a topologically distinct nascent connectivity. An additional factor of 2 accounts for the fact that we might have attached the $k_1$ propagator line to $v_r^h$, which generates exactly the same set of diagrams as when this propagator attaches to $v_r^l$. Finally, the remaining $n - r$ interactions can each be attached in $\binom{3}{2}$ ways, with the usual factor of 2 at each interaction in order to account for the twisted connectivities. The total symmetry coefficient is thus $2^{r-1} \times 2 \times 2^{n-r} = 2^n$, just as it was in the other four-particle case.

The total number of ladder diagrams for this case is obtained as follows. We know that nascent four-particle ladders generated from Fig. 4.17 look just like three-
particle ladders, through the \((r-1)\)st interaction, so Eq. (4.58) tells us that there are \((3^{r-2} - 1)/2\) nascent four-particle ladders through the first \(r-1\) factors of \(\bullet \bullet \bullet \bullet \). [In using Eq. (4.58), we have summed over possible values of \(m\) from 2 to \(r-1\), hence this result is independent of \(m\).] Appending the \(r\)th factor of \(\bullet \bullet \bullet \bullet \) increases the number of diagrams by a factor of \(\binom{3}{2}\), but following this each additional interaction increases the number of topologically inequivalent diagrams by a factor of \(\binom{4}{3}\). The number of independent four-particle ladders in this second case is therefore

\[
\sum_{r=3}^{n} \frac{1}{2} (3^{r-2} - 1) \binom{3}{2} \binom{4}{n-r} = \left( \frac{3^{n-1}}{2} \right) (2^{n-2} - 1) - \frac{3}{10} (6^{n-2} - 1) \tag{4.60}
\]

for each \(n \geq 3\). Combining this with Eq. (4.59), we summarize our results for four-particle ladders in the following theorem.

**Theorem 4.2.** At \(n\)th order there exist \((2^{n-2} - 1)(3^{n-1} - 1)/2\) independent, four-particle ladders. The total contribution to \(\Xi_4\) from \(n\)th-order ladders equals \(2^n \bar{P}_4 \cdot \bar{P}_4\) acting on the sum of these independent ladder diagrams.

As an illustration of these results, the independent four-particle ladders at third order are depicted in Fig. 4.18, in the order that they are generated by the procedure described above. The first of these diagrams belongs to Case 1 while the other three arise from Case 2. At fourth order there are 39 independent ladder connectivities; those arising from Case 1 are depicted in Fig. 4.19 while the ones from Case 2 are shown in Fig. 4.20.

There is one subtle feature of the four-particle ladder diagrams that does not arise until fourth order. As indicated by Eq. (4.53), our ladder diagrams are constructed with an intrinsic time order, since we have not considered topological deformations
that alter the time order of the interactions. Consequently, the procedure described in this subsection generates certain diagrams that differ only by a non-essential time ordering relative to other diagrams generated by the same procedure, as exemplified by diagrams (1) and (2) in Fig. 4.19. In evaluating (1), it is understood that the time arguments $t_3^R$ and $t_3^L$ for the penultimate interaction are restricted to be greater than either $t_2^R$ or $t_2^L$, the time arguments for the interaction below. (That is, $t_3^A > t_3^{A'}$, for each $A, A' \in \{R, L\}$.) This ordering is reversed in diagram (2). This is no problem in practice, and in fact both of these diagrams can be combined into one, simply by combining the two integration ranges. Our restriction on time order merely simplifies the algorithm for constructing the ladders and does not affect the diagram weights.
Figure 4.20: Fourth-order, four-particle ladder diagrams arising from Case 2 of the generating algorithm.
4.5 Reconstruction functionals from renormalized perturbation theory

We now utilize the formalism developed in the previous section to derive expressions for the RDMCs $\Delta_3$ and $\Delta_4$ as functionals of the 2-RDM, within a renormalized ladder-type approximation. As we have indicated, the validity of such an approximation requires that simultaneous three-body correlations are small, although we shall show that self-consistent iteration of the CSE serves to rebuild three-body correlations that are neglected by ladder-type decoupling approximations.

4.5.1 The quasi-pair ladder approximation

For the time being we continue to develop the perturbation theory in a format suitable to either $\Delta_p$ or $\overline{\Delta}_p$. Keeping roughly to the diagrammatic notation introduced for the 2-RDMC in Fig. 3.1, let us depict the cumulant of the 2-GF as

$$i^2 \overline{\Delta}_{j_1,j_2;k_1,k_2}(t_1,t_2; t'_1,t'_2) = \quad .$$

(4.61)

This quantity possesses a perturbative expansion that consists of connected two-particle diagrams, several examples of which are shown in Fig. 4.21. Likewise, $\overline{\Delta}_p$ can be expanded in terms of connected $p$-particle diagrams. To develop decoupling approximations for $\overline{\Delta}_3$ and $\overline{\Delta}_4$ in terms of $\overline{\Delta}_2$, we take the latter to be a known quantity and recast the perturbation series for $\overline{\Delta}_3$ and $\overline{\Delta}_4$ as expansions in an effective interaction equal to the sum of all connected two-particle diagrams. This sum is precisely the interaction that is shown diagrammatically in Eq. (4.61).
Figure 4.21: A few diagrams in the expansion of $\mathbf{\Delta}_2$, at second order in the dressed interaction: (a) two-particle ladder; (b) particle-hole scattering diagram; (c) polarization pair bubble; (d) vertex correction.
Consider the selection of connected three-particle diagrams in Fig. 4.22. (End-point labels have been omitted in this figure, for clarity; it is assumed that all diagrams in this figure are indexed in the same way.) Up to permutations of the endpoint labels, each is contained within a single renormalized diagram

\[ Q \overset{\text{def}}{=} k_1,t_1' j_1,t_1' k_2,t_2' j_2,t_2' k_3,t_3' j_3,t_3' ; \]  

(4.62)

in which the hatched square is the same effective interaction that appears in \( \Delta_2 \). The veracity of this claim may be confirmed by substituting various two-particle diagrams (Fig. 4.21) in place of the effective interaction. Equivalently, one can recognize that within each diagram in Fig. 4.22, the interactions can be partitioned into three groups that do not overlap in time; each group corresponds to a different factor of the effective interaction in \( Q \).

We refer to \( Q \) as a quasi-pair ladder (QPL) diagram.\(^{28}\) This language reflects the fact that \( \overline{D}_2(x_1,t_1,x_2,t_2;x'_1,t'_1,x'_2,t'_2) \) is the time-dependent probability amplitude for a pair of particles—a quasi-pair—within the many-particle system, just as \( \overline{D}_1(x,t;x',t') \) represents the distribution of what have come to called quasiparticles.\(^{23}\) The effective interaction appearing in Eq. (4.61) represents the effective pairwise interaction that is felt by a quasi-pair, under the screening influence of the \( N \)-particle system.
The diagram rules and symmetry coefficients derived in the last section must be modified slightly in order to account for this latest renormalization. In contrast to the dressed interaction utilized throughout Section 4.4, the effective interaction in Eq. (4.61) is nonlocal, having a nontrivial dependence on all four of its time arguments. There is no analogue of Eq. (4.50) for this interaction. As a result, one additional caveat is required when constructing diagrams using the quasi-pair interaction in Eq. (4.61). The extra rule is that consecutive factors of the new effective interaction between the same two quasiparticle lines are forbidden. This eliminates certain of the diagrams generated in the last section, and depicted, for example, in Figs. 4.13, 4.14, and 4.18.

The diagram weights derived previously are also modified following the renormalization introduced here. Because $\Delta_2$ is antisymmetric in both its upper and lower indices, and since no index permutations are included in Eq. (4.61), the effective interaction must carry this antisymmetry. In Subsections 4.4.4 and 4.4.5 we did not assume that the pair interaction was antisymmetric, and this adaptation modifies the diagram...
Figure 4.23: Three-particle QPL diagrams, through fourth order.

weights. Specifically, the $n$th-order symmetry coefficient of $2^n$ must be divided by $4^n$ in order to account for the four distinct coordinate permutations (think of them as twisted diagrams) contained within each factor of the effective interaction.

Eliminating the forbidden diagrams from Figs. 4.13 and 4.14, and supplementing the lone second-order ladder diagram that we never bothered to draw, one obtains the three-particle QPL diagrams shown in Fig. 4.23. In light of the remarks above, the appropriate symmetry factor for each is $(\frac{1}{2})^n$, where $n$ is the number of effective interactions. (This factor could, if desired, be subsumed into the effective interaction itself, though this is not done here.) When augmented by this factor and also permutations of the endpoint labels, the diagrams in Fig. 4.23 provide either a second-, third-, or fourth-order QPL approximation to $\Delta_3$.

As noted previously, time-independent many-body perturbation theory shares
a common diagrammatic representation with the time-dependent theory (from which the time-independent theory is obtained, as a limiting case), although the rules for evaluating the diagrams (time integration versus energy denominators, etc.) obviously differ. The derivations in Subsections 4.4.4 and 4.4.5, however, utilized only the topological properties of the diagrams and, implicitly, their relation to Wick’s theorem, so all of these results are equally valid for time-independent diagrams. In particular, the QPL approximation to the 3-RDMC utilizes exactly the same diagrams as those appearing in Fig. 4.23, where now the effective quasi-pair interaction derives from the 2-RDM. For convenience, we have labeled the second-, third-, and fourth-order diagrams appearing in this figure as six-index tensors $S$, $T^{(n)}$, and $F^{(n)}$, respectively. The matrix elements of each tensor are obtained from the corresponding diagram by labeling the indices in the following fashion (compare Fig. 4.23):

$$T^{(1)}_{j_1,j_2,j_3;k_1,k_2,k_3} \overset{\text{def}}{=} j_1 k_1 j_2 k_2 j_3 k_3.$$  \hspace{1cm} (4.63)

Using this notation and the diagram weights discussed above, one obtains explicit formulas for the QPL summations through any finite order. For example, the sum of second- and third-order, three-particle QPL diagrams yields the approximation

$$\Delta_{j_1,j_2,j_3;k_1,k_2,k_3} \approx \mathcal{P}_3 \mathcal{P}_3 \left( \frac{1}{4} S_{j_1,j_2,j_3;k_1,k_2,k_3} + \frac{1}{8} \sum_{n=1}^{2} T^{(n)}_{j_1,j_2,j_3;k_1,k_2,k_3} \right)$$  \hspace{1cm} (4.64)

for the matrix elements of the 3-RDMC.
4.5.2 Factorization approximations for QPL diagrams

Equation (4.64), or any other QPL approximation for $\Delta_p$ or $\overline{\Delta}_p$, neglects non-ladder contributions that represent simultaneous three-electron and higher correlations, but otherwise amounts to nothing more than a regrouping of terms in the perturbation series. As such, Eq. (4.64) is not yet a useful approximation, only a suggestive partitioning of an infinite sum of ladder-type diagrams. Even supposing that a finite sum of QPL diagrams is a decent approximation to $\Delta_3$ and/or $\Delta_4$, we have left open the question of how the QPL diagrams are to be evaluated, without having to perform an infinite summation of ordinary (non-renormalized) perturbation theory diagrams. To obtain a useful result from the renormalized perturbation series, one must develop a method for determining the effective quasi-pair interaction directly from $\Delta_2$.

The author has a strong sense that there ought to be a formally exact solution to this problem, in light of the fact that $\Delta_2$ is—thinking now in terms of ordinary Rayleigh-Schrödinger perturbation theory—exactly equal to the infinite sum of all connected two-electron diagrams, with any number of pairwise interactions, and including any number of polarization parts such as the pair bubble in Fig. 4.21(c). One avenue worthy of exploration is Faddeev theory,$^{68-71,76,77}$ wherein the solution of the $n$-particle scattering problem, for small $n$ (usually $n = 3$) is written as an expansion in the two-particle scattering matrix, which is very much like what we want to do here.

Unfortunately, however, no exact scheme has yet been developed for extracting the exact quasi-pair interaction from the 2-RDMC. However, Nakatsuji and Yasuda$^{15,16,18}$ have developed an approximate method that at least affords a workable solution to this problem. Nakatsuji and Yasuda$^{16}$ have justified their approach based
on time-dependent perturbation theory, but here we simply posit the time-independent version of the NY decomposition approximation, and explore its ramifications. First, denote by $\mathbf{D}^{\text{HF}}_1$ the idempotent 1-RDM corresponding to the reference state $|\Psi_0\rangle$, then define a one-electron matrix

$$
\mathbf{P} \overset{\text{def}}{=} 2 \mathbf{D}^{\text{HF}}_1 - \mathbf{I} .
$$

(4.65)

The matrix $\mathbf{P}$ is diagonal, with diagonal elements $+1$ in the occupied-occupied block and $-1$ in the virtual-virtual block. Also define

$$
\gamma \overset{\text{def}}{=} \mathbf{D}_1 - \mathbf{D}^{\text{HF}}_1 .
$$

(4.66)

The diagram that we call $\mathbf{S}$ (Fig. 4.23) is decomposed in the NY method as

$$
S_{j_1,j_2,j_3;k_1,k_2,k_3} = \gamma + \mathbf{D}^{\text{HF}}_1 \mathbf{P} \gamma .
$$

(4.67)

This is an exact equality, obtained by substituting $\gamma + \mathbf{D}^{\text{HF}}_1 \mathbf{P}$ for $\mathbf{D}_1$ in the definition of $\mathbf{S}$. In the diagrams above, and in the ones that follow, a circled symbol indicates a one-index matrix product with the matrix indicated within the circle.

The approximation in the NY method is $\gamma \approx \mathbf{D}^{\text{HF}}_1 \mathbf{P} \gamma$. Noting that

$$
\mathbf{D}^{\text{HF}}_1 \mathbf{P} = \mathbf{D}^{\text{HF}}_1 = \mathbf{P} \mathbf{D}^{\text{HF}}_1
$$

(4.68)

and therefore

$$
\mathbf{D}^{\text{HF}}_1 \mathbf{P} \gamma = \mathbf{D}^{\text{HF}}_1 \mathbf{P} \mathbf{D}_1 - \mathbf{D}^{\text{HF}}_1 ,
$$

(4.69)
when $D_1^{HF} P \gamma$ is substituted for $\gamma$ in Eq. (4.67), the second diagram in this equation cancels. This leads to an approximate decomposition of the second-order QPL diagram:

\[
S_{j_1,j_2,j_3;k_1,k_2,k_3} \approx \sum_m B_{j_1,j_2;k_1,k_2}^{m,m} P_{m;m} \Delta_{j_2,j_3;m,k_3}, \tag{4.70}
\]

where

\[
B_{j_1,j_2;k_1,k_2}^{m,m} \overset{\text{def}}{=} = \sum_m B_{j_1,j_2;k_1,k_2}^{m,m} P_{m;m} \Delta_{j_2,j_3;m,k_3}, \tag{4.71}
\]

Actually this is not quite satisfactory, as the resulting approximation for $S$ (hence for $\Delta_3$) is not self-adjoint, which can be recognized immediately since the diagram in Eq. (4.70) lacks a horizontal plane of symmetry. Nakatsuji and Yasuda circumvent this by solving CSE(2) simultaneously with its adjoint equation, which leads to a self-adjoint 2-RDM at convergence, but perhaps a simpler solution is to average the approximated tensor $S$ defined by Eq. (4.70) with its own adjoint, that is, to use $(S + S^\dagger)/2$ in place of $S$ in the QPL approximation of Eq. (4.64).

We will comment below as to the validity of the approximation $\gamma \approx D_1^{HF} P \gamma$, but first let us finalize the decomposition procedure by indicating how the last missing ingredient, the tensor $B^{**\bullet\bullet}$, is determined. An equation to determine this tensor
follows from the same two steps used to decompose the $S$ diagram:

\[
\begin{array}{c}
\includegraphics[width=0.5\textwidth]{diagram}
\end{array}
\]

(4.72)

The first, exact equality follows from the identity $D_{1}^{\text{HF}} = D_{1} - \gamma$, while the approximation in the second line is $\gamma \approx D_{1}^{\text{HF}} P \gamma$, the same approximation that was employed in decomposing $S$. Equation (4.72) defines a linear system to be solved for the variables $B_{j_{1}j_{2};k_{1}k_{2}}^{***}$. Introducing the resolvent matrix

\[
R^{(1)} \overset{\text{def}}{=} (I + P \gamma)^{-1} = \sum_{n=0}^{\infty} (-P \gamma)^{n},
\]

(4.73)

the linear system corresponding to Eq. (4.72) is

\[
B_{j_{1}j_{2};k_{1}k_{2}}^{***} = \sum_{m} \Delta_{j_{1};m;k_{1}k_{2}} R_{j_{2};m}^{(1)}.
\]

(4.74)

The tensor $B^{***}$ is determined as follows. We can calculate $R^{(1)}$, once and for all, given the 1-RDM. Then given the 2-RDM (hence $\Delta_{2}$), we solve the linear system above for a fixed triple $(j_{1}, k_{1}, k_{2})$ of indices, which determines the elements $\{B_{j_{1}1;k_{1}k_{2}}^{***}, B_{j_{1}2;k_{1}k_{2}}, \ldots, B_{j_{1}R;k_{1}k_{2}}^{***}\}$. We must then solve this linear system for each such triple of indices, using the same resolvent each time.

The NY 3-RDM reconstruction functional$^{15-17}$ is $D_{3} \approx \Sigma_{3} + \Delta_{3}^{QPL(2)/NY}$. The NY estimate $\Delta_{3}^{QPL(2)/NY}$ of the cumulant involves two separate approximations, as the notation is meant to indicate. The first is truncation of the renormalized perturbation series at the level of second-order QPL diagrams [the $S$-only approximation to
Eq. (4.64)], while the second approximation is the NY decomposition scheme for $S$ that we have just introduced. The symmetry coefficients over which we toiled in Section 4.4 were apparently derived by Nakatsuji and Yasuda using brute-force enumeration of diagrams, as no mention of a systematic procedure is made in the relevant papers.\textsuperscript{15,16} Our derivation allows this approach to be extended to higher orders, and also affords a QPL approximation for $\Delta_4$.

We will illustrate some of this extended formalism below, but first let us return to the issue of the NY decomposition procedure. Recall that the essential approximation in that procedure was

$$\gamma \approx D_1^{\text{HF}} P \gamma.$$  \hspace{1cm} (4.75)

One can judge the error in this approximation by means of the difference matrix

$$\mathcal{E} \overset{\text{def}}{=} \gamma - D_1^{\text{HF}} P \gamma $$  \hspace{1cm} (4.76)

$$= (I - D_1^{\text{HF}}) D_1.$$  

The matrix $I - D_1^{\text{HF}}$ is zero except within the virtual-virtual block, where it is equal to an identity matrix. Partitioning $D_1$ into occupied-occupied, occupied-virtual, and virtual-virtual blocks,

$$D_1 = \begin{pmatrix} d_{oo} & d_{vo}^\dagger \\ d_{vo} & d_{vv} \end{pmatrix},$$  \hspace{1cm} (4.77)

one obtains

$$\mathcal{E} = \begin{pmatrix} 0 & 0 \\ d_{vo} & d_{vv} \end{pmatrix}.$$  \hspace{1cm} (4.78)

Thus the largest elements of $\mathcal{E}$ ought to be the diagonal elements of $d_{vv}$, in particular, the populations (from the exact 1-RDM) of low-lying virtual orbitals. Mazziotti\textsuperscript{11} has reported that 3-RDM matrix elements of the type $D_{\omega_1,\omega_2,\omega_3;\nu_1,\nu_2,\nu_3}$, for occupied indices
and virtual indices \(v_i\), are actually slightly less accurate in the NY approximation than they are if \(\Delta_3\) is neglected altogether, whereas other categories of 3-RDM matrix elements are generally improved by the NY estimate of the cumulant. The \(S\)-decomposition approximation is the likely source of the problem, especially in light of the fact that the first non-vanishing correction to the NY decomposition approximation involves precisely the elements \(D_{o_1,o_2,o_3,v_1,v_2,v_3}\). This correction is second-order in the bare (non-renormalized) interaction.

4.5.3 Higher-order QPL approximations

Lacking a superior alternative to the NY factorization scheme for the QPL diagrams, we proceed to generalize this approximation so that it can be used to decompose QPL diagrams besides the three-particle, second-order one considered by Nakatsuji and Yasuda. The technique is rather straightforward, and the complicated equations that result can mostly be read directly from the diagrams. Thus we limit our discussion to two examples. Rather than choosing a higher-order three-particle QPL diagram (one of the \(T^{(n)}\), for instance), let us focus on four-particle examples, since to date no one else has proposed any 4-RDM reconstruction functional containing connected terms.

Third- and fourth-order, four-particle ladder diagrams were enumerated previously (Figs. 4.18–4.20). However, this collection contains some diagrams that differ by a non-essential time order (see the discussion at the end of Subsection 4.4.5), as well as others that contain consecutive interactions between the same pair of propagator lines, which disqualifies them in the QPL renormalization. After removing all of these offending categories of diagrams, we obtain the diagrams depicted in Fig. 4.24. For conciseness, we have used a pared-down diagrammatic notation that simply illustrates...
the topology; the diagrams in Fig. 4.24 should be understood to bear the full QPL regalia, just like the ones in Fig. 4.23. When the four-particle diagrams are expressed in terms of the quasi-pair interaction (like the ones in Fig. 4.23), the diagram weight is the same as for the three-particle QPL diagrams, namely, \((\frac{1}{2})^n\) at \(n\)th order.

Taking as an example the first diagram in Fig. 4.24, there are potentially several ways in which the NY factorization could be applied. These are depicted in Fig. 4.25. The reader should recall the decomposition of \(S\) considered in the previous subsection, where naïve application of the NY decomposition afforded a non-Hermitian approximation to \(\Delta_3\). The same is true of the present decomposition, and we ought to consider the average of all four decompositions shown in Fig. 4.25. For brevity, we
write out a formula for only one of these:

\[ P = \sum_m S_{j_1,j_2,j_3,k_1,k_2,m} P_{m,m}^* B_{j_4,k_3,k_4}, \]  \hspace{1cm} (4.79)

with

\[ B_{j_1,j_2;k_1,k_2}^* \overset{\text{def}}{=} B_{j_2,j_1;k_1,k_2}. \]  \hspace{1cm} (4.80)

Proceeding as in the last subsection, one may derive a system of linear equations to determine the matrix elements of \( B^{**} \). In so doing, one discovers that

\[ B_{j_1,j_2;k_1,k_2}^* = -B_{j_2,j_1;k_1,k_2}, \]  \hspace{1cm} (4.81)

hence the decomposition in Eq. (4.79) does not introduce any new tensors.

At fourth order, certain of the four-particle diagrams cannot be decomposed by a single application of the NY factorization approximation. Two such factorizations will do the trick, however, as for example in the approximate factorization

\[ \approx \]  \hspace{1cm} (4.82)
Figure 4.25: Alternative ways to apply the NY factorization to the first diagram in Fig. 4.24. None is preferred \textit{a priori}, so the average of all four should be used.
Evaluation of the diagram on the right will involve the quantities

\[ B_{j_1,j_2;k_1,k_2}^{\infty \infty} \quad \text{def} = \begin{array}{c|c} j_1 \quad j_2 \hline k_1 \quad k_2 \end{array}. \tag{4.83} \]

Unlike \( B^{\infty \infty} \), which was determined by \( B^{\infty \infty} \), the tensor \( B^{\infty \infty} \) is independent, obtained as the solution of another linear system involving \( \Delta_2 \). This system of equations can be derived by the same techniques used to arrive at the linear system for \( B^{\infty \infty} \) in Eq. (4.74).

### 4.5.4 Iteration and higher-order correlations

Finally, we present some observations concerning the interconnection between the reconstruction process and the solution of either CSE(\( p \)) or ICSE(\( p \)). We have already pointed out that ladder-type reconstructions—even if extended to infinite order in the effective pair interaction—neglect simultaneous three-electron correlations. The CSEs (or ICSEs), however, help to build these correlations back into the cumulants. This becomes clear upon examination of the diagrammatic representations of these equations, together with diagrammatic representations of the reconstruction functionals.

In Fig. 4.26(a) we show a typical diagram in the expansion of \( \Delta_3 \), which includes terms of third and higher order in the bare interaction, but which cannot be incorporated into any ladder-type diagram because it involves simultaneous correlation between three particles. Because CSE(2) and ICSE(2) are equations in a two-electron matrix space, when \( \Delta_3 \) appears in these equations it is always traced over coordinate \( x_3 \). In Fig. 4.26(b) we show the effect of the \( \text{tr}_3 \) operation on the diagram in 4.26(a). Diagram 4.26(b) is included in the partial trace of a third-order QPL diagram, namely
Figure 4.26: Diagrams illustrating the connection between reconstruction and solution of the CSEs or ICSEs. See the text for an explanation.

the one shown in Fig. 4.26(c). Thus, the presence of $\text{tr}_3$ in the two-particle equations allows us to incorporate three- and higher-body effects that would not otherwise be present in a ladder approximation for the three- and four-electron cumulants.

Actually three-electron correlations such as the one in Fig. 4.26(a) are introduced by the CSEs and ICSEs, even within a second-order ladder approximation. To understand why, consider the diagram in Fig. 4.26(d), representing one of the terms in $\Omega^5_3$, the kernel of the equation ICSE(2). Within a second-order QPL approximation to $\Delta_3$, the diagram in Fig. 4.26(b) is included within 4.26(d). Thus three- and higher-body effects are incorporated into the cumulants $\Delta_3$ and $\Delta_4$ by the CSEs or ICSEs, even when these effects are absent from approximate reconstruction functionals. In effect, solution of these equations corresponds to a partial summation of the perturbation
series for $D_2$ [in the case of CSE(2)] or $\Delta_1$ and $\Delta_2$ [in the case that ICSE(1) and ICSE(2) are solved simultaneously]. The connection between reconstruction and solution of the coupled Green’s function equations of motion, which are time-dependent hierarchies analogous to the CSE($p$) hierarchy, has been explored in some detail.$^{53,56}$ A more thorough exploration of this connection in the present context would be welcome.

4.6 Summary, outlook, and future directions

In this chapter we have discussed several reconstruction schemes for determining the 3- and 4-RDMs as functionals of the 2-RDM. A clear derivation of Valdemoro’s proposed reconstruction functionals$^{5,6}$ reveals that the final form of the Valdemoro 4-RDM functional (which is used by the Valdemoro group not only for $D_4$ but also for $D_3$, via the contraction relation) is based upon a sequence of rather tenuous and inconsistent approximations. These result in a functional that does not recover even the entire unconnected part of the 4-RDM. The cumulant formalism for the RDMs had not been so widely popularized at the time that this functional was introduced, but with the clarity of this new analysis it would seem that Valdemoro’s functionals are obsolete, with $\Sigma_4$ the unconnected approximation of choice for $D_4$.

All other reconstruction functionals proposed to date do exploit the known formulas for $\Sigma_3$ and $\Sigma_4$ to recover the entire unconnected parts of $D_3$ and $D_4$. In some cases, an attempt is made to estimate $\Delta_3$, but $\Delta_4$ has thus far been completely neglected. It is not clear that this is a physically reasonable approximation, however. As we demonstrated in Subsection 4.3.3, the cumulants of all orders appear to be necessary in order to cancel certain unconnected products that are anomalously large, relative to the corresponding RDM elements. Indeed, for Be atom we showed that the
spin component $\Delta_3^{aaa}$ contributes 12–16 m$E_h$ to the electronic energy (depending on the basis set), a contribution equal and opposite that of $\Sigma_3^{aaa}$. (For any four-electron singlet state, $\Sigma_3^{aaa} = -\Delta_3^{aaa}$, since $D_3^{aaa} = 0$.) The other 3-RDMC spin component, $\Delta_3^{a\alpha\beta}$, contributes roughly the same, non-negligible amount to the energy.

For Be, $D_4^{aaa} = 0$ and $D_4^{a\alpha\beta\beta} = 0$. Therefore, if the 4-RDMC is neglected (which is always done at present), $\hat{S}^2$ symmetry is necessarily broken. This is a very troubling observation, suggesting that perhaps the simplistic cumulant formalism—ported directly from classical statistical mechanics, with only an antisymmetrized product to reflect the quantum-mechanical nature of the distribution $D_p(x_1, \ldots, x_p; x'_1, \ldots, x'_p)$—is not appropriate in this context, as it knows nothing about spin. Specifically with regard to the CSE, Valdemoro’s group has gone to considerable effort to spin-adapt their algorithms, which may be a solution to this problem, though it is not clear how this fits in with the perturbative reconstruction schemes developed in this chapter. Alternatively, a spin-free definition of the cumulants should be examined. At this point, the deeper significance (if any) of the RDM cumulants, and their interpretative usefulness as tools to understand electron correlation, is dubious.

Despite these misgivings, we have endeavored to develop our own reconstruction functionals, specifically for $D_4$ and incorporating a nonzero estimate of the 4-RDMC. This is accomplished by means of a renormalized many-body perturbation theory, the rudiments of which were developed in Section 4.4, including an algorithm to generate all independent three- and four-particle ladder diagrams, without omissions or redundancies. To the extent possible, the diagrammatic formalism for this theory was established in a manner that does not discriminate between the time-independent and the time-dependent versions of the theory. The author hopes that the time-dependent
version may find eventual application as a means to improve standard closure approximations for the coupled Green’s function equations of motion, perhaps in relativistic field theories, where the many-time nature of the many-particle Green’s functions is essential.

Equipped with the diagram rules, symmetry factors, and generating algorithms developed in Section 4.4, we showed in Section 4.5 how the renormalized perturbation theory provides a very general scheme for deriving RDM reconstruction functionals. The reconstruction functionals of Nakatsuji and Yasuda$^{15-17}$ are recovered as a special case. Reconstruction functionals derived from renormalized many-body perturbation theory neglect simultaneous many-particle correlations beyond a certain number of particles, though we have shown that the structure of the CSE plays an instrumental role in putting these correlations back into the 2-RDM.

In spite of these developments, the author is pessimistic about the future of the CSE as a method for direct computation of the 2-RDM. The problem is not so much that the method cannot be made to work—varyingly successful calculations coming out of several research groups$^{8-11,13-18,31,99-102}$ have demonstrated that it does work, at least as well as singles-doubles CI and at least in small basis sets. $N$-representability is still a problem, indicating that existing reconstruction functionals are not sufficiently restrictive, which should have been anticipated anyway. As a result, naïve implementation of the CSE without any explicit, auxiliary $N$-representability constraints yields results that are inferior to the ones cited above.$^{20}$ Both the Valdemoro$^{8-10,22,99,103,104}$ and Mazziotti$^{102,105}$ groups utilize rather sophisticated algorithms to accomplish the iterative solution of CSE(2), in which certain auxiliary $N$-representability constraints are applied at each iteration, mostly to remove negative eigenvalues of $\hat{D}_2$, $\hat{Q}_2$, and/or
The outstanding question is whether anything is gained, relative to *ab initio* wavefunction theory, by CSE-based methodologies. To date, the author believes that this question must be answered in the negative. Although the perturbative reconstruction schemes proffered in Section 4.5 are perhaps aesthetically appealing in their generality, their computational expense with respect to basis size quickly spirals out of control. For instance, calculation of the NY correction for the 3-RDM cumulant, Eq. (4.70), scales\(^{11,20}\) as \(R^7\). This is ostensibly the same as that of the CCSD(T) method,\(^{106}\) which would be bad enough, but in fact the RDM method is worse because in that case the scaling is a true \((N_{\text{occupied}} + N_{\text{virtual}})^7\), whereas the scaling for CCSD(T) is really only \(N_{\text{occupied}}^3 N_{\text{virtual}}^4\). Evaluation of the lowest-order corrections to \(\Delta_4\), such as the diagram given in Eq. (4.79), scales as \(R^9\). An active-space implementation of the reconstruction procedure could potentially reduce the scaling dramatically, though of course the same can also be said of active-space CC theory,\(^{107}\) so again it is unclear whether there is a comparative advantage to the CSE method.

A larger question—one that cuts to the heart of 2-RDM theory—is whether the usual many-body formalism can ever be used to develop a 2-RDM-based methodology that offers a comparative advantage over traditional many-body wavefunction theory. Back in Chapter 1 we listed as an advantage of 2-RDM theory the fact that the 2-RDM depends upon four, rather than \(N\), electronic coordinates. Within the context of variational calculations, this may indeed prove to be an important and useful advantage. However, the use of second quantization effectively converts \(N\)-electron Hilbert-space quantum mechanics to a two-electron problem in Fock space, hence the real killer in ordinary many-body theory is not the number of electrons *per se* but rather the num-
ber of summations over spin-orbital indices. (Perturbative or CC approximations to
the wavefunction convert the exponentially-complicated FCI problem to one having
only polynomial complexity.) Sums over spin-orbital indices are precisely the problem
that one encounters in our perturbative approach to density matrix reconstruction.
Rather than co-opting the methods of wavefunction quantum mechanics, in order to
make 2-RDM theory comparatively advantageous it seems that what one really needs
to do is exploit the properties of \( D_2(x_1, x_2; x'_1, x'_2) \) as a position-space kernel. Only
then does the lack of \( N \)-dependence in the 2-RDM truly offer a significant advantage
over \( \Psi(x_1, \ldots, x_N) \).

Whether or not the CSE ultimately matures into a viable electronic structure
method, exploration of the cumulant structure of the RDMs is still an interesting
topic, especially with regard to what these cumulants do or do not tell us about elec-
tron correlation. Reservations concerning the cumulant formalism have already been
detailed above, and the spin-free cumulants\(^{40,41} \) suggested as a possible alternative
formulation. Here we point out another unresolved issue worthy of exploration.

The relationships between \( D_n \) and \( \Delta_n \) that are specified in Eqs. (2.136a)–(2.136e)
amount to expectation values (in principle, with respect to the exact wavefunction)
of certain \( n \)-electron replacement operators \( \hat{a}^{\dagger}_{jk} \cdots \hat{a}^{\dagger}_{kn} \hat{a}_{i_1} \cdots \hat{a}_{i_n} \). These same operators
define the relationship between the CI and CC excitation operators,

\[
\hat{C}_n = \frac{1}{p!^2} \sum_{\text{occupied}} \sum_{\text{virtual}} c_{i_1 \ldots i_n}^{j_1 \ldots j_n} \hat{a}^{\dagger}_{j_1} \cdots \hat{a}^{\dagger}_{j_n} \hat{a}_{i_1} \cdots \hat{a}_{i_n} \quad (4.84a)
\]

and

\[
\hat{T}_n = \frac{1}{p!^2} \sum_{\text{occupied}} \sum_{\text{virtual}} t_{i_1 \ldots i_n}^{j_1 \ldots j_n} \hat{a}^{\dagger}_{j_1} \cdots \hat{a}^{\dagger}_{j_n} \hat{a}_{i_1} \cdots \hat{a}_{i_n} \quad (4.84b)
\]
respectively. Given a determinantal reference state $|\Psi_0\rangle$, the exact wavefunction can
be expressed either in the form of a linear CI expansion,

$$|\Psi\rangle = |\Psi_0\rangle + \sum_{i=1}^{N} \hat{C}_i |\Psi_0\rangle ,$$

or an exponential CC expansion

$$|\Psi\rangle = \exp \left( \sum_{i=1}^{N} \hat{T}_i \right) |\Psi_0\rangle .$$

A CI or CC calculation entails determination of the CI coefficients $c_{j_1,\ldots,j_n}$ or the cluster
amplitudes $t_{j_1,\ldots,j_n}$.

Truncation of the sums over $i$ in Eqs. (4.85) and Eq. (4.86) defines either a
tuncated CI or a truncated CC wavefunction, which differ from one another order-
by-order, beginning at the level of double excitations. When all excitations up to
$N$-fold are included, however, both expansions recover the exact wavefunction. In
that case, the relationship between the CI and CC excitation operators is

$$\hat{C}_1 = \hat{T}_1$$

$$\hat{C}_2 = \hat{T}_2 + \frac{1}{2} \hat{T}_1^2$$

$$\hat{C}_3 = \hat{T}_3 + \hat{T}_1 \hat{T}_2 + \frac{1}{6} \hat{T}_1^3$$

$$\hat{C}_4 = \hat{T}_4 + \frac{1}{2} \hat{T}_2^2 + \frac{1}{2} \hat{T}_1 \hat{T}_2 + \hat{T}_1 \hat{T}_3 + \frac{1}{24} \hat{T}_1^4 .$$

The functional relationships $\hat{C}_n = \hat{C}_n[\hat{T}_1,\ldots,\hat{T}_n]$ defined in these equations are exactly
the same as the functionals $D_n = D_n[\Delta_1,\ldots,\Delta_n]$ given in Eqs. (2.136a)–(2.136d).
The appearance of different coefficients is merely an artifact of different normalization
conventions. It can be verified, for example, that upon combining equivalent terms
in $\Delta_4^N$ (consult Appendix B for assistance), this quantity contains 24 distinct terms,
which should be compared with the factor of $\frac{1}{24}$ that multiplies $\hat{T}_1^4$ in Eq. (4.87d). One may also verify that the cluster operators $\hat{T}_n$ commute, hence there is no ambiguity in the order of the cluster operators.

Equations (4.87a)–(4.87d) seem to imply that the cluster amplitudes $t_{i_1, \ldots, i_n}^{j_1, \ldots, j_n}$ $(n \geq 2)$ measure irreducible electron-pair correlations, three-electron correlations, etc., with respect to the reference determinant $|\Psi_0\rangle$, whereas the cumulant matrix elements $\Delta_{i_1, \ldots, i_n; j_1, \ldots, j_n}$ measure the same type of correlations, with respect to the exact wavefunction. One consequence is that $\Delta_n$ is dense, whereas the upper indices of $t_{i_1, \ldots, i_n}^{j_1, \ldots, j_n}$ refer exclusively to virtual spin orbitals and the lower indices refer exclusively to occupied spin orbitals. A more detailed theoretical and computational study of the relationship between the RDM cumulants and the CC amplitudes should be carried out. Along this line, we have already begun to extend the preliminary numerical calculations presented in Subsection 4.3.3 to larger systems, by means of a more efficient version of our original code.

4.7 References


Chapter 5

(One-electron) density matrix functional theory

5.1 Introduction

The methodology developed in this chapter represents a complete departure from that pursued in Chapters 3 and 4, yet the goal is the same: to determine RDMs—and from them, electronic energies—directly, without the aid of a wavefunction. As the title suggests, in this chapter we explore electronic structure methods in which the basic variable is the 1-RDM rather than the 2-RDM; this is the essence of what has come to be called density matrix functional theory (DMFT). The term natural orbital functional theory is also used, to emphasize that what is sought is an expression for the electronic energy as an explicit functional of the physical (as opposed to reference-state) natural orbitals (NOs) and their (generally fractional) occupation numbers. The NOs as central variables seems quite natural from the standpoint of “chemical intuition,” which is mostly based upon orbital models anyway, while the use of frac-
tional occupation numbers frees us from the restriction of a non-interacting reference system and allows us (at least in principle) to access the actual, quantum-mechanical one-particle description of an atom or molecule, namely, the 1-RDM, which in practice is never idempotent.

Because this approach to electronic structure differs radically from that developed in the preceding two chapters, it warrants a substantive introduction of its own, which we develop in the rest of this Introduction. Subsection 5.1.1 consists of a phenomenological discussion of the motivating principles behind DMFT, while Subsection 5.1.2 describes our particular “reconstructive” approach, and outlines the work that occupies the remainder of this chapter.

5.1.1 Why DMFT?

The term density functional theory (DFT) persists, somewhat anachronistically, to refer to the modern Kohn-Sham (KS) formalism, but in truth KS-DFT ought to be regarded as an orbital functional theory. In fact, it was the infamously difficult problem of constructing an accurate density functional $T[\rho_1]$ for the kinetic energy that originally led Kohn and Sham to formulate DFT in terms of a non-interacting reference state (determinant of KS orbitals) whose kinetic energy is readily calculated. The KS determinant is sometimes called a “fictitious” reference state, which strikes the author as somewhat unfair, given that the same pejorative is seldom applied to the HF determinant, and moreover the KS orbitals nowadays play an active role in both the interpretation and the development of DFT. The latter point is exemplified by the growing popularity of so-called hybrid density functionals, the ubiquitous B3LYP being the most familiar example. These hybrids incorporate a certain
fraction of HF exchange into what is ostensibly a density functional for the exchange-correlation energy. In this context, “HF exchange” means the expectation value of the HF exchange operator \([\text{Eq. (5.6b)}]\) with respect to the determinant of KS orbitals. Since this exchange operator is nonlocal, its expectation value cannot be evaluated in terms of the density and density gradients; rather, the 1-RDM corresponding to the KS determinant is required.

It has been said\(^2\) that hybrid functionals constitute the second generation of DFT, with “pure” density functionals such as local density approximation (LDA) and its gradient corrections comprising the first generation. More recently, work has begun in earnest on a third generation of functionals, in which the KS orbitals and eigenvalues are utilized in a systematic, perturbative construction of the exchange-correlation potential,\(^{22-34}\) making use of the optimized effective potential (OEP) method\(^2,35-47\) to map old-school many-body quantum mechanics onto an effective, local potential to be used in the KS equations. To describe these approaches would take us far afield; the point is that there is substantial precedent to develop explicitly orbital-dependent energy functionals.

But which orbitals are the most appropriate variables for electronic structure theory? On physical grounds, the natural candidates for a one-electron description of a many-electron system are the aptly-named natural orbitals. In this basis the 1-RDM assumes a diagonal form interpretable as a statistical mixture of one-electron states, wherein one-electron properties (kinetic energy, dipole moment, \(\langle r^n \rangle\), various hyperfine interactions, etc.) are calculated according to the simple sum-over-states
formula

$$\langle \hat{h} \rangle = \sum_i n_i h_{ii} .$$

(5.1)

Contrast this with the situation in DFT. Given an accurate energy functional $E[\rho_1]$, the orbitals obtained by solving the KS equations reproduce the one-electron density—and therefore the energy—but these orbitals do not reproduce the 1-RDM of the physical system.* In DFT there is no sum-over-states formula analogous to Eq. (5.1), even for one-electron properties and—because the KS orbitals have integer occupancies—even in cases where the KS orbitals might happen to resemble NOs.

The HF molecular orbitals also have integer occupancies, but the idempotent HF 1-RDM and the single-determinant HF wavefunction can be understood as the leading terms in perturbative expansions of the exact 1-RDM and wavefunction. Higher-order corrections to the NOs and their occupancies thus emerge from post-HF methods. The precise meaning of the KS orbitals is comparatively difficult to write down in precise, mathematical terms. Based upon a relationship between the (exact) KS effective local potential and the nonlocal one-electron Dyson self-energy, Casida has suggested that the KS orbitals may be good approximations to Dyson orbitals, which has been confirmed in a some applications. More speculatively, it has also been suggested that exact KS orbitals are essentially Brueckner orbitals, which reproduce one-electron properties quite accurately, indicating that they are good approximations to the true NOs.

*It is true that exact KS orbitals, obtained by inverting the KS equations for a known density, without introducing any approximation for $E[\rho_1]$, are frequently quite similar to HF orbitals. Noting this, Zhao and Parr declare that “the problem of finding a physically meaningful wave function from an electron density is thereby solved.” This is anathemtic nonsense. The determinant $\Psi_{KS}$ of KS orbitals cannot be a solution to the electronic Schrödinger equation, no matter how good the functional $E[\rho_1]$ may be. This is obvious, if for no other reason, from the fact that $\langle \Psi_{KS} | H | \Psi_{KS} \rangle$ does not equal the energy that is obtained from $E[\rho_1]$. 
In practice, the KS orbitals are often qualitatively similar to the HF ones (though the orbital energies are typically quite different). This close resemblance is bad news for DFT, since the well-known delocalization of the HF orbitals (especially the virtual ones) is an affront to the chemist’s conception of electron pairs in localized chemical bonds. In part, the trouble arises in both HF and KS theory due to the use of strict integer occupancies, which introduces a rigid, unphysical distinction between occupied versus virtual orbitals (whereas only “strongly occupied” versus “weakly occupied” makes physical sense).

To pursue this point further, we posit the existence of a one-electron effective Hamiltonian $\hat{F}$, which we call a generalized Fock operator, whose expectation value affords the electronic energy. Given a 1-RDM energy functional $E[\hat{D}]$, this operator is defined implicitly through the equation

$$E[\hat{D}] = \text{tr}(\hat{F}\hat{D}) ,$$

or explicitly as $\hat{F} = \delta E/\delta \hat{D}$. Equivalently,

$$\hat{F}\phi_k(x) = \lim_{\varepsilon \to 0^+} \left[ (n_k + \varepsilon)^{-1} \frac{\delta E}{\delta \phi_k^*(x)} \right] .$$

This allows us to associate an effective potential with any orbital- and/or occupancy-dependent energy functional, including KS density functionals. HF theory provides one such functional, namely

$$E_{\text{HF}}[\{\phi_k\}] = \sum_i n_i \langle \phi_i | \hat{h} | \phi_i \rangle + \frac{1}{2} \sum_{ij} n_i n_j \left( \langle \phi_i | \phi_j | r_{12}^{-1} | \phi_i \phi_j \rangle - \langle \phi_i | \phi_j | r_{12}^{-1} | \phi_j \phi_i \rangle \right) ,$$

which is really just the formula for the energy of a single Slater determinant, with $\hat{h} = -(1/2)\nabla^2 - \sum_\mu Z_\mu / r_\mu$ in atomic units. The summations in Eq. (5.4) run over
spin-orbital indices, with occupancies \( n_i = 0 \text{ or } 1 \). Usually the \( n_i \) are omitted in favor of restricted summations, but we retain them here in order to connect with Eq. (5.3).

To obtain the effective Hamiltonian \( \hat{F} \), via functional differentiation of \( E_{HF} \), is essentially the same exercise as to derive the HF equations. The result may be written as \( \hat{F} = \hat{h} + \hat{\nu}^{\text{xc}}_{HF} \), where exchange-correlation potential \( \hat{\nu}_{xc} \) is defined according to Eq. (5.3) with \( E_{xc} \) replacing the total energy \( E \). In the HF case, this potential is said to consist of “exchange only,” but we write it as \( \hat{\nu}^{\text{HF}}_{xc} \) in order to connect with the more general effective potential formalism. By functional differentiation of Eq. (5.4), the HF exchange-correlation potential is found to be

\[
\hat{\nu}^{\text{HF}}_{xc} \phi_k(x) = \lim_{\varepsilon \to 0^+} \left[ (n_k + \varepsilon)^{-1} \frac{\delta E^{\text{HF}}_{xc}}{\delta \phi_k^*(x)} \right]
= \sum_{i \neq k} n_i (\hat{J}_i - \hat{K}_i) \phi_k(x),
\]

in which the Coulomb (\( \hat{J}_i \)) and exchange (\( \hat{K}_i \)) operators are defined in the usual way,

\[
\hat{J}_i \phi_j(x) = \left[ \int dx' \phi_i^*(x') r_{12}^{-1} \phi_i(x') \right] \phi_j(x) \quad (5.6a)
\]
\[
\hat{K}_i \phi_j(x) = \left[ \int dx' \phi_i^*(x') r_{12}^{-1} \phi_j(x') \right] \phi_i(x). \quad (5.6b)
\]

Note that \( \hat{\nu}^{\text{HF}}_{xc} \) in Eq. (5.5) operates on spin orbitals; tracing over spin coordinates yields the familiar “2\( J - K \)” form of the spinless HF potential.

Now consider what happens as one moves along a potential surface, through a point where the HOMO and LUMO are degenerate, that is, \( \epsilon_N = \epsilon_{N+1} \). At this point, the nominal electron configuration and the occupation numbers change abruptly, going from \( n_N = 1 \) and \( n_{N+1} = 0 \) on one side of the anomaly to \( n_N = 0 \) and \( n_{N+1} = 1 \) on the other side. According to Eqs. (5.4) and (5.5), this results in a discontinuous change in the HF energy and effective potential at the point of degeneracy. On the
other hand, if the occupancies $n_i$ were parameters that could be varied smoothly, the
theory might be able to interpolate between two or more electronic configurations.\(^*\)
In other words, such a functional might be able to account for the so-called non-
dynamical correlation that arises due to configurational near-degeneracy. Within KS-
DFT, strong non-dynamical correlation requires an ensemble treatment\(^†\) involving
multiple KS determinants.\(^{71-81}\) In particular, the usual issues of symmetry breaking
and single-determinant malfeasance, well-known from HF theory are present in KS-
DFT as well (though they may be less severe\(^\circ\)).

Non-dynamical correlation is not limited to exotic molecules or isolated points
on the potential surface. It is pervasive in excited states, but even in the ground
state, non-dynamical correlation is qualitatively important whenever chemical bonds
are formed or broken. With its rigid, conceptual separation of occupied and virtual
orbitals, the closed-shell determinantal formalism is always inadequate at sufficient
distance from the equilibrium geometry on the potential energy surface. Consider, for
example, the H$_2$ molecule. Both spin-restricted HF (RHF) and spin-restricted KS-
DFT (RKS) fail to describe H$_2$ dissociation correctly, as shown in Fig. 5.1. (The RKS
potential in this figure was obtained using the B3LYP functional,\(^{21}\) but the qualitative
result is independent of this choice.\(^{83}\) Although the RKS result is somewhat better
than RHF is the asymptotic region, both potentials are fundamentally flawed at large

\(^*\)However, the functional $E_{HF}$ in Eq. (5.4) is minimized by an idempotent 1-RDM,\(^{59}\) so it is not
enough simply to allow fractional occupation numbers. A completely new functional is required.

\(^†\)It has been demonstrated recently that there are real atomic\(^{60}\) and molecular\(^{61}\) systems, characterized
by configurational near-degeneracy, for which the interacting ground state density is not equal to the
ground state density of any KS determinant. Rather, starting from an accurate density and inverting
the KS equations, the interacting ground state is found to correspond to an excited state of the non-
interacting system. This represents a failure of non-interacting pure-state $\nu$-representability\(^{62-68}\)
and unequivocally demonstrates that the (single-reference) KS formalism is not universal, unlike the
Hohenberg-Kohn-Levy\(^{69,70}\) density functional formalism itself.
Figure 5.1: Spin-restricted STO-6G potential curves and CI natural occupation numbers for H$_2$. The RKS result was obtained with the B3LYP$^{21}$ functional.

This is a characteristic feature of bond-breaking potentials within any determinantal formalism, as explained below. On the other hand, there is no reason why a spin-restricted wavefunction cannot correctly describe both the near-equilibrium and the dissociative part of the H$_2$ potential energy curve. For instance, a simple two-term CI wavefunction gives a qualitatively correct potential, which is also shown in Fig. 5.1.

The defect in both the RHF and the RKS potentials at large $R$ is the unrealistically large weight given to ionic configurations. Within any determinantal formalism, the two-electron, singlet-coupled wavefunction $\Psi$ is formed by doubly occupying a single molecular orbital $\sigma_g$,

$$\Psi_{(1,2)} = \frac{1}{\sqrt{2}} \sigma_g(1) \sigma_g(2) \left[ \alpha(1) \beta(2) - \beta(1) \alpha(2) \right]. \quad (5.7)$$

The orbital $\sigma_g(1) \equiv \sigma_g(r)$ consists of a linear combination of atomic orbitals $\psi_A$ and
\( \psi_B \) centered on the constituent atoms,

\[
\sigma_g = [2(S_{AB} + 1)]^{-1/2}(\psi_A + \psi_B),
\]

(5.8)

where \( S_{AB} = \langle \psi_A | \psi_B \rangle \). We see that

\[
\Psi_{(1,2)} = \frac{1}{\sqrt{2}} \left[ \alpha(1) \beta(2) - \beta(1) \alpha(2) \right] \\
\times \left[ \psi_A(1) \psi_B(2) + \psi_B(1) \psi_A(2) + \psi_A(1) \psi_A(2) + \psi_B(1) \psi_B(2) \right].
\]

(5.9)

The first two spatial terms (\( \psi_A \psi_B + \psi_B \psi_A \)) are of the homopolar or covalent type, in the sense that each of the two H atoms gets one electron, but the other two terms (\( \psi_A \psi_A + \psi_B \psi_B \)) are the dreaded ionic configurations, in which both electrons reside on a single H atom. Physically, ionic contributions ought to have vanishing importance for homolytic dissociation of \( \text{H}_2 \), but the use of a determinantal wavefunction \( \Psi \) gives the ionic terms equal weight with the covalent ones, irrespective of nuclear separation. This results in an unrealistically steep rise in the potential at large \( R \), which is evident in Fig. 5.1 and is a characteristic feature of spin-restricted potentials for bond breaking, both at the HF and the KS level.

Partial occupation of the \( \sigma_u^* = (\psi_A - \psi_B)/[2(1 - S_{AB})]^{1/2} \) antibonding orbital is sufficient to correct the problem. Thus, if one mixes the \( (\sigma_u^*)^2 \) configuration with the \( (\sigma_g)^2 \) ground state, via a two-configuration CI expansion, a qualitatively correct potential is obtained, as depicted again in Fig. 5.1. Also shown in the figure are the CI natural occupation numbers for the \( \sigma_g \) and \( \sigma_u^* \) orbitals.\(^*\) At the equilibrium bond length, the configuration \( (\sigma_u^*)^2 \) contains only 0.026 electrons, but approaches double occupancy as \( R \rightarrow \infty \).

\(^*\)In Fig. 5.1 only, we have reverted to the common convention \( 0 \leq n_i \leq 2 \). Elsewhere in this chapter, \( n_i \) indicates the occupancy of the natural spin-orbital \( \chi_i(\mathbf{r}) \alpha(\xi) \), hence \( 0 \leq n_i \leq 1 \).
For the minimal STO-6G basis used in this example, the two-configuration CI calculation is actually equivalent to FCI, because single excitations do not possess the proper symmetry to mix with the $(\sigma_g)^2$ reference configuration. Nevertheless, this example demonstrates that in order to obtain qualitatively correct bond-breaking potentials, one must resort to the use of virtual orbitals. Furthermore, it illustrates that by mixing the occupied and virtual orbitals, one may calculate a qualitatively correct potential energy curve without breaking the spin symmetry.

Another source of both conceptual and computational difficulty is the fact that occupied and virtual HF orbitals do not experience the same potential, owing to the summation restriction in Eq. (5.5). A virtual spin orbital $\phi_j$ feels the entirety of the Coulomb potential minus the exchange potential, $\sum_i n_i (J_i - K_i)$, while an occupied spin orbital $\phi_k$ feels this potential sans the self-interaction term $J_k - K_k$. From the point of view of the virtual orbitals, it is as if an extra electron is present in the molecule, and this leads to virtual orbitals that are significantly more diffuse and delocalized than are the occupied orbitals. In addition to the fact that these orbitals typically do not describe bound excitations in a realistic fashion, their delocalized nature makes them a poor choice for subsequent, post-HF calculations, because they do not have enough amplitude where the “action” (that is, correlation) is, which means that relatively lengthy CI expansions are necessary in order to obtain a desired accuracy. In fact, it was suggested long ago,\textsuperscript{84} and recently verified,\textsuperscript{85,86} that approximate natural orbitals are the best possible basis in which to carry out CI expansions, in terms of the most correlation energy recovered with the fewest number of determinants.

Although we have used the HF method to motivate the remarks above, the same criticisms apply to KS-DFT, which also makes a rigid distinction between occupied
and virtual orbitals. In the DFT case, the one-electron effective potential is

$$\hat{\phi}_k^{KS}(\mathbf{x}) \overset{\text{def}}{=} \lim_{\varepsilon \to 0^+} \left( (n_k + \varepsilon)^{-1} \frac{\delta E_{xc}[\rho_1]}{\delta \phi_k^*(\mathbf{x})} \right).$$

(5.10)

As mentioned above, the functional $E_{xc}$ may have some explicitly orbital-dependent parts, the potentials associated with which will closely resemble the HF potential in their basic form. $E_{xc}$ may also have a part depending on the density and its gradients but not the orbitals per se; however, even this part inherits the troubles of integer occupancies through the density

$$\rho_1(\mathbf{x}) = \sum_k n_k |\phi_k(\mathbf{x})|^2,$$

(5.11)

with $n_k = 1$ or 0.

Thus there are compelling reasons to develop orbital- and occupancy-dependent energy functionals $E\{\{n_i\}, \{\phi_i\}\}$, that is, energy functionals $E[\hat{D}_1]$ of the 1-RDM. Given such a functional, one would determine the energy by variational minimization of $E[\hat{D}_1]$ with respect to orbital rotations $|\phi_k \rangle \to \hat{U}|\phi_k \rangle$ and with respect to variations $n_k \to n_k + \delta n_k$ in the occupancies of these orbitals. Because this procedure does not introduce any reference state, it may provide the necessary infrastructure for an even-handed description of both dynamical and non-dynamical correlation. In particular, by eliminating the rigid distinction between occupied and virtual orbitals, DMFT ought to provide a means to generate localized, weakly-occupied pseudo-natural orbitals to replace HF or KS virtual orbitals in compact expansions of correlated wavefunctions.

The existence of a functional $E[\hat{D}_1]$ for non-degenerate ground states was proved by Gilbert, and the formal properties of this functional received much attention during the period 1975–1985. At first glance, Gilbert’s existence theorem might
appear to be a corollary to the first Hohenberg-Kohn theorem, which establishes
the existence of a density functional $E[\rho^n]$ for non-degenerate ground states, but actu-
al Gilbert’s work generalizes the first and second Hohenberg-Kohn theorems to
the case of nonlocal external potentials.* According to the usual vernacular, the ex-
ternal potential is the static one-electron potential generated by the clamped nuclei.
Since this potential, $-\sum_{\mu} Z_{\mu}/r_{\mu}$, is generated by a local one-electron operator, the
additional generality gained by Gilbert’s existence theorem for $E[\hat{D}_1]$ might seem su-
perfluous. However, one should bear in mind that heavy-atom pseudopotentials— which are an essential feature of solid-state DFT and also DFT-based Car-Parrinello
simulations— often must include nonlocality in order to achieve realistic ac-
curacy.†

To calculate the expectation value of a nonlocal pseudopotential within KS-
DFT, one has no choice but to use the idempotent KS 1-RDM. However, this approach
appears to lack theoretical justification, since the KS 1-RDM is not the true 1-RDM,
even in the limit of the exact density functional $E[\rho^n]$.

This is as good a place as any for a diatribe about locality. The terms local and nonlocal have precise mathematical (and physical) meanings but are widely mis-
used by the DFT community. Specifically, within the DFT literature a functional
$E[\rho_1, \nabla \rho_1, \ldots]$, with explicit dependence on the density gradients is often said to be

---

*The issue of $v$-representability also generalizes from DFT to DMFT. As in DFT, however, the existence of the basic energy functional can be demonstrated by a constrained-search formalism that circumvents $v$-representability altogether. However, any attempt to formulate effective one-electron eigenvalue equations of the KS variety must implicitly assume $v$-representability, since the exact energy functional is differentiable only at $v$-representable densities.

†There is a good physical reason for this. A valence electron in an $s$ orbital, for example, interacts differently with the core electrons than does a valence electron in a $p$ orbital. When explicit core electrons are replaced by a pseudopotential, this orbital dependence can be included via certain projection operators that are not equivalent to local potentials. If a local pseudopotential is employed, then a valence electron’s spatial coordinates alone determines its interaction with the core, hence there can be no explicit orbital character in the core-valence interaction.
nonlocal, whereas the term “local” is assumed to imply the LDA,\textsuperscript{97} which has no gradient dependence because it is based on the uniform electron gas model. This terminology, however, is abusive to both the physical and the mathematical concepts of locality. Even generalized-gradient-type exchange-correlation functionals $E_{xc}[\rho_1, \nabla \rho_1, \ldots]$ utilize only local information—such as the density $\rho_1(x)$ and its gradient $(\nabla \rho_1)(x)$ at the particular point $x$—to determine the exchange-correlation energy density

\[
\varepsilon_{xc}(x) \overset{\text{def}}{=} \frac{\delta E_{xc}}{\rho_1(x)}
\]  

(5.12)

at the same point $x$. By definition, this energy density is the kernel of a local operator, and its expectation value

\[
E_{xc} = \int d\mathbf{x} \varepsilon_{xc}(x) \rho_1(x)
\]  

(5.13)

affords the exchange-correlation energy.

Generally, any operator $\hat{\gamma}$ that acts on one-electron functions is associated with a kernel $\gamma(\mathbf{r}; \mathbf{r'})$ in the position representation,

\[
\hat{\gamma} f(\mathbf{r}) = \int d\mathbf{r'} \gamma(\mathbf{r}; \mathbf{r'}) f(\mathbf{r'}) .
\]  

(5.14)

(Spin dependence is neglected for simplicity.) If it happens that $\gamma(\mathbf{r}; \mathbf{r'}) = \gamma(\mathbf{r}) \delta(\mathbf{r} - \mathbf{r'})$, then the operator $\hat{\gamma}$ is said to be local. In this case, $(\hat{\gamma} f)(\mathbf{r}) = \gamma(\mathbf{r}) f(\mathbf{r})$, so $\hat{\gamma}$ is also said to be equivalent to a local (that is, multiplicative) potential function $\gamma(\mathbf{r})$. For example, the electron-nuclear interaction operator $\hat{v} = -\sum_\mu Z_\mu / r_\mu$ is equivalent to the local potential $v(\mathbf{r}) = -\sum_\mu Z_\mu / |\mathbf{r} - r_\mu|$, the nuclear positions $r_\mu$ being fixed parameters. The expectation value of any local, one-electron operator is a simple functional of the density, $\langle \hat{\gamma} \rangle = \int d\mathbf{r} \gamma(\mathbf{r}) \rho_1^2(\mathbf{r})$.

If $\gamma(\mathbf{r}; \mathbf{r'})$ retains a nontrivial dependence on both $\mathbf{r}$ and $\mathbf{r'}$, then $\hat{\gamma}$ is said to be nonlocal and is not equivalent to any multiplicative potential. The one-electron
kinetic energy operator $\hat{t} = -\frac{1}{2} \nabla^2$, for example, is nonlocal.\footnote{The essential difference between KS-DFT and earlier density functional methods such as Thomas-Fermi theory\cite{97,105,106} is that the former uses the non-interacting 1-RDM to evaluate the kinetic energy, while the latter attempts to develop an approximate, local kernel $t(\mathbf{r})$ whose expectation value $\int d\mathbf{r} t(\mathbf{r}) \rho^2_1(\mathbf{r})$ equals the electronic kinetic energy. The failure of Thomas-Fermi theory to bind molecules\cite{105,107} is an indication of the difficulty in finding a decent local approximation to $\hat{t}$ that can be implemented at the level of self-consistent calculations.\cite{9,108,109}} Nonlocality implies that $\langle \hat{\gamma} \rangle(\mathbf{r})$, the value of the function $\hat{\gamma} f$ at the point $\mathbf{r}$, depends upon the value of $f$ at all points $\mathbf{r}'$, so to compute $\langle \hat{\gamma} \rangle$ requires the full 1-RDM, rather than simply the density. On physical grounds, one ought to expect a one-electron effective Hamiltonian to be nonlocal, as the existence of such an effective Hamiltonian requires that the electron-electron repulsion operator $|\mathbf{r}_1 - \mathbf{r}_2|^{-1}$ (which is multiplicative in a many-electron Hilbert space) be mapped onto a one-electron effective potential. Indeed, the generalized Fock operator introduced in Eq. (5.3) has a generally nonlocal kernel $F(\mathbf{x}, \mathbf{x}') = \delta E/\delta D_1(\mathbf{x}; \mathbf{x}')$. In contrast, the KS effective potential $\nu_{xc}(\mathbf{r}) = \delta E_{xc}/\delta \rho_1^2(\mathbf{r})$ is, by construction, the kernel of a local operator.

This is not to say that reasonable local approximations to the proper one-electron effective potential (essentially the Dyson self-energy\cite{24,49,110}) do not exist. To wit, the expectation values $\langle \hat{t} \rangle$ and $\langle \hat{K} \rangle$ can often be approximated with $>90\%$ accuracy as expectation values of local operators.\cite{111,112,113} In addition, the exchange-only OEP method can reproduce HF results, to within a few m$E_h$\cite{43,45,46,114,118} using a variationally-optimized local approximation to $\hat{\epsilon}_{xc}^{\text{HF}}$. However, the effective local potential approximation should not be granted an intellectual monopoly in the field of quantum chemistry.

If one takes the 1-RDM as the basic variable, then the possibility of nonlocality is introduced from the start, and also the exact kinetic energy functional is known.
Consider the usual spin-independent, Born-Oppenheimer electronic Hamiltonian $\hat{H}$. In this case, $E = \text{tr}(\hat{H}\hat{D}_2)$ can be computed using only the spin-traced 1-RDM $D^\circ_1(r; r')$ along with the two-electron charge density $\rho^\circ_2(r_1, r_2) = D^\circ_2(r_1, r_2; r_1, r_2)$. The energy functional in terms of these quantities is

$$E[D^\circ_1, \rho^\circ_2] = - \int d\mathbf{r} \left( \sum_{\mu} \frac{Z_\mu \rho^\circ_1(\mathbf{r})}{|\mathbf{r}_1 - \mathbf{r}_\mu|} + \frac{1}{2} \int d\mathbf{r}' \delta(\mathbf{r} - \mathbf{r}') \nabla^2 D^\circ_1(\mathbf{r}, \mathbf{r}') \right) + \int d\mathbf{r}_1 \int d\mathbf{r}_2 \frac{\rho^\circ_2(\mathbf{r}_1, \mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}$$

in atomic units. Only the electron-repulsion energy functional $V_{ee}$ is known as an explicit functional of the 1-RDM, though it can be expressed as an explicit functional of either $D_2$, $\rho^\circ_2$, or the intracule density $I(u)$, the latter functional being

$$V_{ee}[I] = 4\pi \int du \ u \ I(u) ,$$

including a volume factor of $u^2$. Here $V_{ee}[I]$ means $V_{ee}[I(u)]$ and the factor of $4\pi$ is an artifact of the definition of $I(u)$, Eq. (2.94).

Knowledge of the exact kinetic energy functional is frequently cited as the primary incentive to consider DMFT as an alternative to DFT. While the author personally considers the natural incorporation of fractional occupation numbers to be a more compelling rationale, the benefits of an exact kinetic energy functional should not be underestimated. Consider that an unfortunate side effect of the KS construction is that the exchange-correlation energy $E_{xc} \equiv T - T_{KS} + V_{ee} - V_H$ contains some residual kinetic energy, namely the difference $T - T_{KS}$ between the true kinetic energy and that of the KS determinant. It is believed, based on atomic calculations,\(^{119}\) that the kinetic component of $E_{xc}$ is small. However, the intermingling of kinetic and potential energies implies that the functional $E_{xc}$ does not satisfy simple homogeneity relations,\(^{120}\)
in contrast to the purely kinetic or purely potential energy functionals $T, V_H,$ etc. On the other hand, only the functional $V_{ee}$ is unknown in DMFT, and consists of potential energy only. Thus the unknown functional in DMFT satisfies a homogeneous scaling requirement\textsuperscript{96} that severely limits the allowed forms for this functional. We shall make use of this restriction in deriving our own functionals for DMFT later in this chapter.

5.1.2 Reconstructive approach

Although DMFT commanded a great deal of attention in the period 1975–1985, out of this work emerged only one explicit proposal\textsuperscript{121} for calculating electronic energies as functionals of the 1-RDM. Even this paper contained no actual calculations and went unnoticed until recently (when the same proposal was advanced independently by others\textsuperscript{122–124}). Over the last several years, there have appeared several other explicit proposals for density matrix functionals,\textsuperscript{122–132} along with the first calculations. Many of these proposals\textsuperscript{122–128} constitute, at least implicitly, a sort of generalized HF theory in which the 2-RDM is expressed as a reconstruction functional $\hat{D}_2[\hat{D}_1]$ of the 1-RDM. Given any approximate reconstruction, the electronic energy is then determined by variational minimization of the functional $E[\hat{D}_1] = \text{tr}(\hat{H} \hat{D}_2[\hat{D}_1])$, subject to $N$-representability constraints on the variation of $\hat{D}_1$.

This brand of DMFT complements the reconstructive approaches to 2-RDM determination that were considered in previous chapters. In our view,\textsuperscript{133} reconstructive DMFT amounts to a variational 2-RDM theory where—in an effort to reduce the exorbitant computational expense\textsuperscript{134} inherent to such calculations—the 2-RDM is parametrized in terms of the 1-RDM. Ensemble $N$-representability constraints\textsuperscript{135} for acceptable 1-RDMs are easy to implement, but are insufficient to guarantee that the
reconstructed 2-RDM is $N$-representable. Hence the burden of insuring that $D_2$ is $N$-representable falls instead on the reconstruction functional itself. If one is prepared to be pragmatic, and accept that not all known $N$-representability constraints will be satisfied in practice, then these conditions provide a trove of exact constraints for the functional $E[\hat{D}_1]$.

The reconstructive perspective on DMFT has interesting parallels in DFT, where parametrizations of the exchange-correlation hole, as a functional of the electron density, figure prominently in the development of new energy functionals.\textsuperscript{19,136-140} Only for an exact reconstruction of the hole is one assured that $E \leq E[\rho_{1,\text{trial}}]$, whereas $E \leq E[\hat{D}_{1,\text{trial}}]$ would be guaranteed in DMFT only if the exact reconstruction $\hat{D}_2[\hat{D}_1]$ could be employed. For practical approximations there is no lower bound, at the theorem level, on either $E[\rho_{1,\text{trial}}]$ or $E[\hat{D}_{1,\text{trial}}]$. DFT survives as a viable methodology because, empirically, the energy does converge with respect to enlargement of the variational space.\textsuperscript{141-143} Thus, while the variationally-optimized energy in approximate DFT may be higher or lower than the exact energy, the model exchange-correlation hole is at least associated with a finite energy in the complete-basis limit. Whether the model 2-RDMs corresponding to proposed functionals $\hat{D}_2[\hat{D}_1]$ exhibit this property—dubbed \textit{variational stability}—was questioned in a recent study,\textsuperscript{144} where it was determined that optimized DMFT energies for diatomic molecules decrease precipitously as the number of active orbitals is increased.

The aforementioned result is intriguing but not conclusive, because no more than six active orbitals were used for any molecule other than $\text{H}_2$. For $\text{H}_2$, extrapolation to the limit of a complete basis of $s$- and $p$-functions yields a reasonable result,\textsuperscript{126} but there has been no systematic study of the basis-set dependence of reconstructive
DMFT for molecules beyond $H_2$. The effect of higher angular momentum basis functions also has not been scrutinized in a serious way. In addition, because existing reconstruction functionals fail to preserve antisymmetry of the 2-RDM, their description of same-spin electrons is seriously flawed, but this anomaly does not manifest in $H_2$.

Let us now outline the remainder of this chapter. Proposed reconstruction functionals that were cited above are defined and derived in Subsections 5.2.1 and 5.2.2. None of these originates with the author’s work, though in Subsection 5.2.3 we present an original analysis (to be published in Ref. 133), at the formal level, of certain $N$-representability violations that are manifested by these functionals. In particular, we give a general proof that half of the eigenvalues of the parallel-spin component $\hat{D}_2^{2\alpha}$ of the 2-RDM are necessarily negative, for each functional examined here.

Given their lack of positivity and other $N$-representability violations, one may ask whether the given reconstruction approximations are reasonable at all. To address this, in Section 5.3 we examine various reconstructed pair densities for the ground state of Be in Section 5.3, where positivity violations are indeed observed in regions where two parallel-spin electrons are close together. While this problem is not to be trivialized, the spurious part of the same-spin pair density is small in magnitude as compared to the (strictly positive) opposite-spin electron-pair density, and a more thorough analysis ought to examine the energetic ramifications of the positivity violations (see below). Originally, our analysis of Be pair densities was carried out using CI natural occupation numbers to affect the reconstruction; however, the results in Section 5.3 are augmented with some energy-optimized pair densities. (That is, reconstructed pair densities obtained from the 1-RDM that minimizes $E[\hat{D}_1]$.) Our
procedure for minimizing the functional $E[\hat{\mathcal{D}}_1]$ with respect to variations in the 1-RDM is introduced in Section 5.4.

In Section 5.5 we present the first thorough examination of the basis-set dependence of DMFT with an eye toward variational stability. Two questions are addressed in that section. First, does the variationally-optimized electronic energy ultimately converge with respect to enlargement of the variational space? To answer this question, we study the basis-set dependence of the variationally-optimized DMFT energy for Be and LiH in a variety of standard Gaussian basis sets. LiH and Be were chosen for this study because they incorporate both same-spin electrons and non-hydrogen atoms, but are small enough so that large basis sets impart upon the 1-RDM a great deal of variational flexibility. In our most extensive calculations, the two $\alpha$-spin electrons in Be are distributed among 35 orbitals, including polarization functions through the $f$ shell. Similar quality calculations have been carried out for LiH across the entire potential energy curve. If the energy is indeed variationally unbounded, our calculations (which do not employ any frozen orbitals) ought to detect this.

The results in Subsection 5.5.1 unequivocally demonstrate that the proposed density matrix functionals converge to a finite, reasonable energy in the complete-basis limit. In fact, the decrease in DMFT energy with respect to basis-set enlargement is roughly comparable to that of FCI, although polarization functions in DMFT prove to be more effective in lowering the energy than they are in FCI. What appeared to be an instability in previous (very limited) active-space calculations is instead the manifestation of a pronounced, qualitative discrepancy between the potential curves obtained using a minimal basis set and those calculated with extended basis sets. All
of the functionals examined here yield qualitatively correct LiH potential curves in a minimal basis set, but in several cases these potentials become unrealistically shallow when extended basis sets are employed.

A second issue addressed by our LiH and Be calculations is that of $N$-representability. In Subsection 5.5.2, the $N$-representability violations manifested by the variationally optimized 2-RDMs are examined in detail. The extent of these violations is staggering: in certain applications, the magnitude of the most negative eigenvalue approaches 30\% of the largest positive eigenvalue, and in all cases the negative eigenvalues contribute substantially to the electronic energy. This behavior is inherent to the ansatz, but we shall argue that explicitly antisymmetric reconstruction functionals are likely to assuage this positivity problem to a great extent.

Based on our results for Be and LiH, in conjunction with previous studies of the same functionals,\textsuperscript{124,146} we identify one particular functional [SIC-CH(1), in the notation introduced in Subsection 5.2.1] that gives accurate results in both minimal and extended basis sets. In particular, diatomic potential curves are qualitatively correct at all values of the internuclear distance. The aforementioned $N$-representability analysis demonstrates that this success is attributable to a cancellation of errors, insofar as spurious negative eigenvalues of the 2-RDM annihilate an enormous amount of the electron repulsion energy that is incurred by the positive eigenvalues. However, this cancellation is not only surprisingly accurate, but also quite robust. This we demonstrate in Section 5.6 by applying the SIC-CH(1) functional to calculate potential energy curves of two strongly correlated systems, C$_2$ and FH (hydrogen fluoride, not to be confused with Hartree-Fock).

Finally, in Section 5.7 we derive a functional of our own that is intended to
correct certain of the \( N \)-representability violations manifested by existing functionals. Preliminary results indicate that this is achieved.

5.2 Reconstruction functionals for the 2-RDM

Separate reconstruction functionals are employed for each independent spin component of \( \hat{D}_2 \). Reconstruction of these components \( \hat{D}^{\sigma\mu}_2 \) is carried out exclusively in the NO basis. Existing reconstruction functionals for DMFT express the matrix elements \( D^{\sigma\mu}_{ij;kl} \) as functions of the natural occupation numbers, neglecting any explicit dependence on the NOs themselves. This is driven partly by convenience, but also because HF theory leads to one such reconstruction, and because the energy functional already inherits a strong dependence on the NOs via the one- and two-electron integrals. This assumption may limit the present methods to spin-compensated states, because for a spin-polarized state the spatial parts of \( \alpha \)-spin NSOs differ from those of \( \beta \)-spin NSOs. In any case, a spin-restricted formalism is adopted in this chapter. This is not essential, and in the near future we plan to test spin-unrestricted versions of these functionals.

5.2.1 Specification of the functionals

The reconstructions examined here each utilize the simple Hartree product

\[
D^{\alpha\beta}_{ij;kl} = \frac{1}{2} n_i n_j \delta_{ik} \delta_{jl} \tag{5.17}
\]

for the opposite-spin component of the 2-RDM. For the parallel-spin component, the functionals of interest have the form

\[
D^{\alpha\alpha}_{ij;kl} = \frac{1}{2} \left[ n_i n_j \delta_{ik} \delta_{jl} - f(n_i, n_j) \delta_{il} \delta_{jk} \right] , \tag{5.18}
\]
which is constitute a modification of the HF exchange term, recovering the RHF reconstruction for \( f(n_i, n_j) = n_i n_j \). Substituting this reconstruction of \( \hat{D}_2 \) into the energy functional

\[
E = \text{tr}(\hat{h} \hat{D}_1) + \text{tr}(r_{12}^{-3} \hat{D}_2),
\]

one obtains a NO functional for the electronic energy:

\[
E[f(n_k), \{ |\chi_k\rangle \}] = 2 \sum_i n_i h_{ii} + \sum_{ij} \left( 2 n_i n_j \langle ij | ij \rangle - f(n_i, n_j) \langle ij | ji \rangle \right).
\]

In this equation, \( h_{ij} \) and \( \langle ij | kl \rangle \) denote one- and two-electron integrals in the NO basis,

\[
h_{ij} \overset{\text{def}}{=} \int dr \chi_i^*(r) \left[ -\frac{1}{2} \nabla^2 + \sum_{\mu} \frac{Z_{\mu}}{|r - r_{\mu}|} \right] \chi_j(r)
\]

and

\[
\langle ij | kl \rangle \overset{\text{def}}{=} \int dr_1 dr_2 \chi_i^*(r_1) \chi_j^*(r_2) \frac{1}{r_{12}} \chi_k(r_1) \chi_l(r_2).
\]

We use a Dirac-style notation for the two-electron integrals; Mulliken notation \( (ij|kl) \) \( \equiv \langle ik|jl \rangle \) is also common.

It is significant that in Eq. (5.19) we have expressed the energy in terms of both \( \hat{D}_1 \) and \( \hat{D}_2 \), rather than in terms of \( \hat{D}_2 \), using the two-electron reduced Hamiltonian formalism of Section 2.2. The latter form, \( E = \text{tr}(\hat{H}_{12} \hat{D}_2) \), gives a different energy than Eq. (5.19) when the partial relation that connects \( \hat{D}_1 \) and \( \hat{D}_2 \) is not satisfied. In using Eq. (5.19) to define the energy, any error in the partial trace relation is isolated in the electron repulsion energy.

Listed in Table 5.1 are the functions \( f \) that define the corrected Hartree\textsuperscript{121–123,128} (CH), corrected Hartree-Fock\textsuperscript{122,128} (CHF), and modified CHF\textsuperscript{128} (MCHF) functionals.
Table 5.1: Specification of reconstruction functionals, according to Eq. (5.18).

<table>
<thead>
<tr>
<th>Functional</th>
<th>Refs.</th>
<th>$f(n_i, n_j)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RHF</td>
<td></td>
<td>$n_i n_j$</td>
</tr>
<tr>
<td>CH($\zeta$)</td>
<td>121–123,127</td>
<td>$(n_i n_j)^{\zeta/2}$</td>
</tr>
<tr>
<td>SIC-CH($\zeta$)</td>
<td>124–126</td>
<td>$(n_i n_j)^{\zeta/2} + (n_i^2 - n_i^\zeta)\delta_{ij}$</td>
</tr>
<tr>
<td>CHF($\zeta$)</td>
<td>122,147</td>
<td>$n_i n_j + \zeta \sqrt{\Lambda_{ii} \Lambda_{jj}}$</td>
</tr>
<tr>
<td>MCHF</td>
<td>128</td>
<td>$(n_i n_j + \sqrt{\Gamma_{ii} \Gamma_{jj}})/2$</td>
</tr>
</tbody>
</table>

The quantities

$$\Lambda_{ij} \overset{\text{def}}{=} n_i (1 - n_j) \quad (5.23a)$$

$$\Gamma_{ij} \overset{\text{def}}{=} n_i (2 - n_j) \quad (5.23b)$$

have been introduced for succinctness. As advertised, the CH functional adds an exchange correction to $(\hat{D}_1 \otimes \hat{D}_1)/2$, the Hartree 2-RDM, albeit one that does not fully annihilate self-interaction in the Hartree potential. The CHF functional adds a correction to the HF 2-RDM $\hat{D}_1 \wedge \hat{D}_1$, in order to account for the two-electron cumulant. It bears mention that within the spin-restricted formalism used here, HF means RHF; unless stated otherwise, we will use these terms interchangeably.

The functionals introduced in Table 5.1 appear in the literature under a litany of pseudonyms, and the names used here were introduced by the author in an attempt to unify this nomenclature. The functional CH($\zeta$), which was introduced by Holas, parametrizes and generalizes CH(1), which was derived independently by several authors based on various criteria that are discussed in the next subsection. The functional CHF($\zeta$) includes the scaling parameter $\zeta$ suggested by Staroverov and Scuseria (with an opposite sign convention), and recovers the original CHF functional when $\zeta = 1$. SIC-CH($\zeta$) is a partially self-interaction-corrected version of
CH(\(\zeta\)), introduced\(^{126,127,148}\) as a generalization of the functional SIC-CH(1) proposed by Goedecker and Umrigar.\(^{124,125}\) [In the author’s previous work,\(^{145}\) the functional SIC-CH(\(\zeta\)) was called CP(\(\zeta\)), while CH(\(\zeta\)) was denoted SICP(\(\zeta\)).] The origins of these functionals are discussed in more detail in the next subsection, where each of them is derived.

### 5.2.2 Origins of the reconstruction functionals

The functional CH(1) was first proposed by Müllner,\(^{121}\) who considered the ansatz

\[
D_2(1;2;1';2') = c_1 D_1(1;1') D_1(2;2') + c_2 D_1^{\lambda+1/2}(1;2') D_1^{-\lambda+1/2}(2;1'),
\]

(5.24)

ostensibly having parameters \(c_1, c_2,\) and \(\lambda\), though it is immediately apparent that one must take \(c_1 = 1 = -c_2\) in order that \(D_2\) reduce to the correct form in the limit of an idempotent 1-RDM. With this choice, the one-parameter ansatz above satisfies the partial trace relation

\[
\text{tr}_2 \hat{D}_2 = \frac{1}{2} (N - 1) \hat{D}_1
\]

(5.25)

for all values of \(\lambda\). Unfortunately, \(\rho_2(x,x) \leq 0\) for this ansatz, with equality obtained only for an idempotent 1-RDM. This is a consequence of the fact that \(D_2\) in Eq. (5.24) is antisymmetric only when \(D_1\) is idempotent. Müllner\(^{121}\) determined that for any fixed value of \(x\), the choice \(\lambda = 0\) minimizes \(|\rho_2(x,x)|\); with \(\lambda = 0\) and \(c_1 = 1 = -c_2\), Eq. (5.24) is equivalent to the CH(1) reconstruction. The functional CH(\(\zeta\)) is merely a parametrized version of CH(1) that has also received attention.\(^{127,145}\) Equation (5.25) is satisfied only for \(\zeta = 1\), however.

SIC-CH(\(\zeta\)) is obtained from CH(\(\zeta\)) by deleting all “orbital” self-interactions, that is, all \(D^{\alpha\alpha}_{\gamma\gamma}\) matrix elements.\(^{124,127}\) This was first suggested by Goedecker and
Umrigar,\textsuperscript{124} in whose view the self-interaction correction was more important that satisfaction of the partial trace relation in Eq. (5.25), which is violated by all of the SIC-CH(\(\zeta\)) functionals. Despite the orbital self-interaction correction, SIC-CH(\(\zeta\)) retains some residual electron self-interaction, insofar as this functional is not antisymmetric. The only antisymmetric reconstruction consistent with Eq. (5.18) is the RHF choice \(f(n_i, n_j) = n_i n_j\), which is equivalent to both CH(2) and SIC-CH(2).

From Eq. (5.18) and Table 5.1, the orbital self-interaction terms have the form \(D_{ii;ii}^{\alpha\alpha} = (n_i^2 - n_i^\xi)/2\). This implies that the SIC-CH(1) 2-RDM has trace \({N \choose 2} + \sum_i n_i(n_i - 1)\). On physical grounds, one expects the quantities \(n_i(n_i - 1)\) to be small, for a spin-compensated state, and we find that the SIC-CH(\(\zeta\)) energy functional is indeed minimized by a near-idempotent 1-RDM (see Section 5.5). Observe also that \(\text{tr} \hat{D}_2 \leq {N \choose 2}\) for SIC-CH(\(\zeta\)), that is, the SIC-CH(\(\zeta\)) 2-RDM contains too few electron pairs. As \(\hat{D}_2\) is used to compute only the electron-repulsion part of \(E\), the SIC-CH(\(\zeta\)) 2-RDM appears likely to overstabilize the energy.

Given that CH(1) serves as the inspiration for both the CH(\(\xi\)) and SIC-CH(\(\zeta\)) functionals, we will discuss some alternative considerations that lead to this same functional and thus provide additional insight into its properties. One alternative derivation, which will lead us as well to the CHF(\(\zeta\)) functional, has been given by Csányi and Arias.\textsuperscript{122} These authors consider expansion of \(D_{2(1,2;1',2')}\) by tensor products of one-particle kernels. One can imagine three different forms of tensor product between two one-particle kernels, differing by permutations of the coordinate labels; following Csányi and Arias,\textsuperscript{122} we denote these three possibilities as \(\otimes_1\), \(\otimes_2\), and \(\otimes_3\). For
one-particle kernels $f_1$ and $f_2$, the three products are

$$\left[f_1 \otimes f_2\right]_{(1,2;1',2')} \overset{\text{def}}{=} f_1(1;1') \ f_2(2;2') \quad (5.26a)$$

$$\left[f_1 \otimes f_2\right]_{(1,2;1',2')} \overset{\text{def}}{=} f_1(1;2) \ f_2(2';1') \quad (5.26b)$$

$$\left[f_1 \otimes f_2\right]_{(1,2;1',2')} \overset{\text{def}}{=} f_1(1;2') \ f_2(2;1') \quad (5.26c)$$

Clearly $\otimes \equiv \otimes_i$ is our ordinary direct product while $\wedge \equiv \left(\otimes - \otimes\right)/2$ is the Grassmann product. The product $\otimes$ was immediately discounted by Csányi and Arias, as all such terms must vanish for extended systems, otherwise the 2-RDM cannot satisfy the requirement that $D_2(x_1, x_2; x'_1, x'_2) \to 0$ as the points $(x_1, x_2)$ and $(x'_1, x'_2)$ become infinitely separated in two-particle space. This property of the 2-RDM follows immediately from the definition of $D_2$ as the zero-time limit of the two-particle propagator, Eq. (2.38), from which we see that $|D_2(x_1, x_2; x'_1, x'_2)|^2$ represents the probability for injecting two electrons into the $N$-electron system, at points $x'_1$ and $x'_2$, and simultaneously removing two electrons from points $x_1$ and $x_2$.

On this basis, we seek an approximate 2-RDM as an expansion in Type I and Type III products,

$$\hat{D}_2 \approx \hat{D}_0 + \hat{D}_{xc}$$

$$= \hat{D}_0 + \sum_{r \in \{I, III\}} \sum_i c_i (\hat{f}_{i,1} \otimes \hat{f}_{i,2}) \quad (5.27)$$

There are basically two reasonable choices for the leading term $\hat{D}_0$: the Hartree 2-RDM $\hat{D}_1 \otimes \hat{D}_1$ or the HF 2-RDM $\hat{D}_1 \wedge \hat{D}_1$. For finite systems, one’s inclination is to prefer the latter, since it is antisymmetric, but whether the HF starting point is truly superior depends entirely on how the correction term is constructed. The simple derivation below, which is based upon satisfaction of the partial trace relation in Eq. (5.25), leads
to fundamentally different correction terms depending on the choice of \( \hat{D}_0 \). Neither of these functionals, in the author’s opinion, can be deemed superior to the other \textit{a priori}.

It was recognized by Csányi and Arias\textsuperscript{122} that Hermitian symmetry, along with the partial permutational symmetry relation

\[
D_{2(1,2;1',2')} = D_{2(2,1;2',1')} ,
\]  

(5.28)

imposes a strong constraint between the one-particle operators \( \hat{f}_{i,1} \) and \( \hat{f}_{i,2} \) used to construct \( \hat{D}_{xc} \). Furthermore, if \( \hat{D}_{xc} \) consists of only a single Type I or Type III tensor product, then the partial trace relation can be inverted and solved.

Suppose that \( \hat{D}_{xc} \) consists of a single Type III product. Then the only self-adjoint reconstructions that satisfy both Eqs. (5.25) and (5.28) are the functionals CH(1) and CHF(1).\textsuperscript{122} As their names indicate, for CH(1) we take \( \hat{D}_0 = \hat{D}_1 \otimes \hat{D}_1 \) while CHF(1) uses \( \hat{D}_0 = \hat{D}_1 \wedge \hat{D}_1/2 \). The correction terms obtained from the sum rule are

\[
\hat{D}_{xc}^{\text{CH}(1)} = \hat{D}_1^{1/2} \otimes \hat{D}_1^{1/2} 
\]  

(5.29)

and

\[
\hat{D}_{xc}^{\text{CHF}(1)} = [\hat{D}_1(\hat{1} - \hat{D}_1)]^{1/2} \otimes [\hat{D}_1(\hat{1} - \hat{D}_1)]^{1/2} .
\]  

(5.30)

In contrast to these simple Type III corrections, any one-term Type I ansatz for \( \hat{D}_{xc} \) that is consistent with Eq. (5.25) must vanish in the thermodynamic limit.\textsuperscript{122} As indicated in the previous section, the more general functionals CH(\( \zeta \)) and CHF(\( \zeta \)) were obtained simply by parametrizing CH(1) and CHF(1), respectively; see Table 5.1. Only for \( \zeta = 1 \) do these functionals satisfy the partial trace relationship in Eq. (5.25).
By construction, CH(\(\zeta\)) and CHF(\(\zeta\)) satisfy the partial symmetry requirement in Eq. (5.28), for all \(\zeta\), which insures that \(\rho_{2(1,2)} = \rho_{2(2,1)}\), but neither complies with the antisymmetry requirement for individual \(1 \leftrightarrow 2\) or \(1' \leftrightarrow 2'\) exchange. Antisymmetry demands an expansion of \(\hat{D}_{xc}\) by Grassmann products, viz., consistent pairs of Type I and Type III tensor products. The simplest approximation of this form is

\[
\hat{D}_2 \approx \hat{D}_1 \wedge \hat{D}_1 + \hat{\gamma} \wedge \hat{\gamma}
\]

(5.31)

for some one-particle operator \(\hat{\gamma}\). The only choice consistent with the sum rule and also valid in the thermodynamic limit is\(^{122}\) \(\hat{\gamma} = \pm [\hat{D}_1 (\hat{1} - \hat{D}_1)]^{1/2}\), with one interesting caveat. Note first that the NSOs diagonalize \(\hat{\gamma}\), whose eigenvalues are therefore \(\gamma_{i,\sigma} = \pm [n_{i,\sigma} (1 - n_{i,\sigma})]^{1/2}\), for \(\sigma \in \{\alpha, \beta\}\). The caveat is that in order to obtain proper scaling in the thermodynamic limit, the signs of the \(\gamma_{i,\sigma}\) must be chosen so that they sum to zero (\(\text{tr} \hat{\gamma} = 0\)). Thus different \(\gamma_{i,\sigma}\) have different signs.

What is interesting is that the quantity \(\pm [n_{i,\sigma} (1 - n_{i,\sigma})]^{1/2}\) appears also\(^{149,150}\) in the variational equations for the AGP wavefunction\(^*\) as well as those for the Hartree-Fock-Bogoliubov (HFB) quasiparticle Hamiltonian.\(^{147,151}\) Within these methodologies both the \(n_{i,\sigma}\) as well as the signs of the \(\gamma_{i,\sigma}\) are in effect variational parameters. More precisely, in the AGP case the coefficients of the generating geminal, which determine the AGP wavefunction and energy, are determined by the \(n_{i,\sigma}\) along with an equal number of complex phase factors;\(^{149}\) the latter parameters determine the signs of the \(\gamma_{i,\sigma}\). Similarly, off-diagonal elements of the HFB effective Hamiltonian are complex, with phase factors that give rise to signed quantities \(\gamma_{i,\sigma}\) in the HFB energy expression.\(^{147}\) Both the HFB and AGP methods contain HF as a special case, and

\*The AGP ansatz was introduced on page 60.
one discovers that in the limit $\Psi_{\text{AGP}} \to \Psi_{\text{HF}}$ or $\Psi_{\text{HFB}} \to \Psi_{\text{HF}}$, the dependence on phase disappears$^{147,149}$ as the occupied orbitals all tend to the same phase. Perhaps due to this complexity, Csányi and Arias$^{122}$ did not pursue such an antisymmetric ansatz for $\hat{D}_2$.

Finally there is the MCHF functional. For the homogeneous electron gas, it is found$^{122,128}$ that CH(1) is substantially overcorrelated at high densities, while CHF(1) tends to zero correlation energy per particle in the same limit. (In the low-density limit, both functionals tend toward a finite, nonzero correlation energy per particle.$^{122,126,128,148}$) At high densities the exact electron-gas result lies between the CHF(1) and the CH(1) correlation energies, so a new functional was suggested$^{128}$ in which the coefficient of the HF exchange term in $\hat{D}_0$ equals $-\frac{1}{4}$, intermediate between the CHF(1) value of $-\frac{1}{2}$ and the CH(1) value of zero. The corresponding ansatz for the 2-RDM, which we have termed the MCHF reconstruction, is $\hat{D}_{\text{MCHF}}(1;1')\hat{D}_{\text{MCHF}}(2;2') - \frac{1}{4}\hat{D}_1(1;2')\hat{D}_1(2;1') + \hat{D}_{\text{MCHF}}(1;1',1;2)$. The requirement that the partial trace relation be satisfied then determines $\hat{D}_{\text{MCHF}}$.$^{128}$

$$\hat{D}_{\text{MCHF}} \overset{\text{def}}{=} -\frac{1}{4}\left[\hat{D}_1(\hat{2} - \hat{D}_1)\right]^{1/2} \otimes \left[\hat{D}_1(\hat{2} - \hat{D}_1)\right]^{1/2}.$$ (5.32)

($\hat{2}$ in this equation means twice the unit operator.)

Having now derived each of the functionals introduced in Subsection 5.2.1, we at last return to CH(1) and show that this functional can be envisaged as a model for the exchange-correlation hole, a concept that we now introduce. Observe that the cumulant partitioning of $D_2$ leads to a partitioning

$$\rho_{2(1,2)} = \frac{1}{2}\rho_{1(1)}\rho_{1(2)} - \frac{1}{2}\left|D_{1(1;2)}\right|^2 + \Delta_{2(1,2;1,2)}$$ (5.33)
of the spin-dependent two-electron density. Introducing the quantities
\[ \rho_{x(1/2)} \overset{\text{def}}{=} -\frac{|D_{1(1;2)}|^2}{\rho_{1(1)}} \]  
(5.34)

and
\[ \rho_{c(1/2)} \overset{\text{def}}{=} \frac{2 \Delta_{2(1,2;1,2)}}{\rho_{1(1)}} , \]  
(5.35)

Eq. (5.33) becomes
\[ \rho_{2(1,2)} = \frac{1}{2} \rho_{1(1)} \left[ \rho_{1(2)} + \rho_{x(1/2)} + \rho_{c(1/2)} \right] \]
\[ = \frac{1}{2} \rho_{1(1)} \left[ \rho_{1(2)} + \rho_{xc(1/2)} \right] . \]  
(5.36)

The notation “(1|2)” is intended to indicate a function of \( x_2 \) that depends parametrically on the value of \( x_1 \), that is, a conditional probability distribution for electron 2. The picture that emerges from Eq. (5.36) is that of an exchange-correlation “hole,” \( \rho_{xc(1/2)} \), for the electron at \( x_2 \), which depends parametrically on the position \( x_1 \) of a “reference” electron. The exchange-correlation hole modifies the probability distribution of electron 2 in order to account for the presence of electron 1. The hole must deplete from \( \rho_1(x_2) \) a density corresponding to exactly one electron,
\[ \int dx_2 \rho_{xc}(x_1|x_2) = -1 , \]
(5.37)
irrespective of the value of \( x_1 \). This normalization condition is a consequence of the partial trace relation in Eq. (5.25).

We have also introduced in Eq. (5.36) the common partitioning of the total exchange-correlation hole into an exchange hole \( \rho_x \), also known as the Fermi hole, and a Coulomb hole* \( \rho_c \). Within the HF approximation \( \rho_c \equiv 0 \), but more generally

* The quantity \( \rho_c \) is sometimes called the correlation hole, but we avoid this terminology because it
the Coulomb hole correlates both same-spin and opposite-spin electrons, whereas $\rho_x$ always vanishes for electrons of the same spin. It is on this basis that we identify the latter as the exchange or Fermi contribution to the overall hole.

The definitions of the Fermi and Coulomb holes (and exchange and correlation, more generally) are a frequent source of miscommunication between \textit{ab initio} quantum chemists and practitioners of KS-DFT, so let us clarify what each term implies in either context. Our definition of the exchange hole is consistent with the definition commonly used in DFT,\textsuperscript{11} provided that one employs the KS 1-RDM for $D_1$ in Eq. (5.34). This is straightforward enough, but potentially confusing for two reasons. First, the “exact” exchange hole in DFT has a determinantal form, by definition, which gives it some additional properties that are shared by the HF exchange hole but are not generally true of the “exact” exchange hole [$\rho_x$ in Eq. (5.34)] as defined in a wavefunction or 2-RDM context. Second, given an approximate wavefunction or 2-RDM, one can use the corresponding two-particle density in place of the exact $\rho_2$ in Eq. (5.33) to define Fermi and Coulomb holes that are exact within the model chemistry defined by some approximate computational scheme. In DFT, however, there is no directly accessible two-particle information, since the KS determinantal 2-RDM, for example, contains the same information as the determinantal 1-RDM. Thus DFT has no direct analogue of the Coulomb hole in Eq. (5.35). Instead, the Coulomb hole $\rho_c$ in DFT is defined by difference,$\textsuperscript{11}$ using the first line of Eq. (5.36), with $\rho_x$ representing the aforementioned KS exchange hole and $\rho_2$ taken to be the exact pair density for the

\textsuperscript{11}conflicts with the usage of “correlation” to mean “exact minus HF.” Insofar as the position-space HF 1-RDM is a good approximation to the exact 1-RDM, the Coulomb hole approximates the difference between $\rho_{xc}$ and the corresponding HF quantity. Indeed, it is found\textsuperscript{123} that the HF hole is essentially indistinguishable from the exact Fermi hole in cases where HF gives a good density, $\rho_{xc,\text{HF}}^\alpha(\mathbf{r}) \approx \rho_{\alpha}^\text{HF}(\mathbf{r})$, which is a weaker condition that $D_{\alpha,\text{HF}}^\epsilon(\mathbf{r}; \mathbf{r}') \approx D_{\alpha}^\epsilon(\mathbf{r}; \mathbf{r}')$. 
actual physical system. Thus, given an ansatz for the wavefunction or 2-RDM (a model chemistry), there is a well-defined exchange-correlation hole $\rho_{xc}$ associated with that model, whereas in DFT there is only one exchange-correlation hole for any given system, namely, the one obtained from the exact two-particle density. Insofar as DFT provides a high-quality density and therefore a good exchange hole (see the footnote on page 311), and assuming that the wavefunction or 2-RDM provides a reasonable approximation to the exact $\rho_2$, the two definitions of the Coulomb hole should be quite similar.

Let us take note of a few salient features of the Fermi and Coulomb holes. (More thorough discussions, with examples, can be found in the literature.) First of all, it is seen from Eq. (5.34) that $\rho^{\sigma_1 \sigma_2}(r_1|r_1) = 0$ if $\sigma_1 \neq \sigma_2$, while $\rho^{\alpha \sigma}(r_1|r_2) \leq 0$. Hence the Fermi hole always depletes the conditional density of the second electron. Furthermore, $\rho^{\sigma \sigma}_x(r_1|r_2) \rightarrow -\rho^\tau_1(r_2)$ as $r_2 \rightarrow r_1$. In contrast to $\rho_x$, the density $\rho_c$ can be positive or negative. When $|r_1 - r_2|$ is small, the Coulomb hole is large in magnitude but negative in sign, $\rho_c(x_1|x_2) \ll 0$. Owing to the presence of $\rho_1$ in the denominator of Eq. (5.35), and also because same-spin electrons are already strongly correlated by $\rho_x$, the behavior of $\rho_c$ in other regions of space is not so easy to predict. It is found that the locations of the nuclei have more of an impact on the shape and position of the Coulomb hole than does the position of the reference electron. When $|r_1 - r_2|$ is large, the Coulomb hole is positive, corresponding to an enhancement of the second electron’s density relative to the independent-particle result. This positive tail can be quite large, as for example in the dissociation of H₂, if the reference electron is stationed near one of the nuclei (call it Hₐ). In this case the Fermi hole has substantial magnitude on both nuclei, for all internuclear distances $R$. On physical grounds, the total
hole ought to become localized around the reference electron (that is, around $H_A$) as $R \to \infty$, hence when $R$ is large, the Coulomb hole around $H_B$ must be large and positive in order to cancel the negative Fermi hole. Plots of $\rho_x$, $\rho_c$, and $\rho_{xc}$ for $H_2$ dissociation can be found in various papers by Baerends and co-workers.\textsuperscript{11,123,155}

The importance of the nuclei in determining the exchange-correlation hole is especially significant in light of the fact that many parametrizations of this hole,\textsuperscript{136–138} which are used to develop density functionals, rely on electron-gas data, which account for the reference electron only. It is hoped that this deficiency can be corrected by parametrizing the hole in terms of the 1-RDM, thus including the effects of the nuclei as the NOs are determined self-consistently.

One such parametrization has been given by Buijse and Baerends,\textsuperscript{123,156} and leads to our final derivation of the CH(1) functional. Their development is motivated by the notion of an orbital amplitude $\vartheta$ for the Fermi hole,\textsuperscript{157,158}

$$\rho_{x(1|2)} = -|\vartheta(1|2)|^2.$$  \hspace{1cm} (5.38)

The amplitude

$$\vartheta(x_1|x_2) = \sum_i a_i(x_1) \phi_i(x_2)$$  \hspace{1cm} (5.39)

is envisaged as a function of $x_2$ that is expanded in spin orbitals $\phi_i(x_2)$, with expansion coefficients $a_i$ that depend only on the position of the reference electron. For a SD wavefunction, it is readily verified that $\rho_x = \rho_{xc}$ can be expressed in this form, with

$$a_i(x_1) \overset{\text{SD}}{=} \phi_i^*(x_1) \sqrt{\frac{n_i}{\rho_i(x_1)}}.$$  \hspace{1cm} (5.40)

(This explains, for example, why the Fermi hole for an atom has only a weak dependence on the position of the reference electron, except as this reference crosses
the boundary between shells.\cite{123,154} For the SD case, the expansion orbitals $\phi_i$ in Eq. (5.39) are the NSOs.

To extend this concept to non-idempotent 1-RDMs, we first note that Eqs. (5.38) and (5.39) are completely general, since the Fermi hole is negative definite for correlated 2-RDMs as well. Buijse and Baerends,\cite{123} however, make a negative definite ansatz for the entire hole, which is an approximation. Following these authors, let us assume that $\rho_{xc}(1|2) = -|\vartheta(1|2)|^2$, with $\vartheta$ expanded in terms of NSOs, as in Eq. (5.39). Then the partial trace relation in Eq. (5.25) requires that

$$a_i(x_1) = \pm \phi_i^*(x_1) \sqrt{\frac{n_i}{\rho_i(x_1)}}. \quad (5.41)$$

Here we have another case of a functional that is determined only up to a phase; by choosing all of the signs to be positive, one obtains the CH(1) functional. Buijse and Baerends\cite{123} rationalize this choice because it leads to the correct limit as $\hat{D}_1$ becomes idempotent, but especially in light of our remarks above concerning phase factors in the AGP and HFB wavefunctions, there is no reason why each $a_i$ must have the same sign. In any case, it is reassuring that the two derivations of CH(1) given previously, which were more mathematical in nature, have been connected here to a physical model.

### 5.2.3 $N$-representability

The RHF functional is the only antisymmetric reconstruction consistent with the ansatz in Eq. (5.18), therefore none of the other functionals affords an $N$-representable 2-RDM. We must concede that Csányi and Arias,\cite{122,128} and also Müller,\cite{121} developed the functionals CH(1), CHF(1), and MCHF for use in extended (condensed-matter) systems, where the antisymmetry requirement becomes exponentially unimportant for
distant regions of the system. Similarly, antisymmetry should be unimportant between
distant parts of a large (but finite) molecular system, and it is an interesting question
as to the length scale on which antisymmetry remains important. To obtain a rough
answer, consider that the exchange hole density \( \rho_{x}^{\sigma}(\mathbf{r}; \mathbf{r}') \to 0 \) as \( D_{1}^{\sigma}(\mathbf{r}; \mathbf{r}') \to 0 \). This,
in turn, happens as \( D_{1}^{\sigma}(\mathbf{r}; \mathbf{r}') \to 0 \). For an insulator, characterized by a HOMO-LUMO
gap (band gap) \( \Delta \epsilon \neq 0 \), the spinless 1-RDM vanishes asymptotically as\(^{159}\)

\[
D_{1}^{\sigma}(\mathbf{r}; \mathbf{r}') \sim \exp(\sqrt{\Delta \epsilon} \left| \mathbf{r} - \mathbf{r}' \right|),
\]

(5.42)

whence derives our prior statement about the exponentially vanishing importance of
antisymmetry. For unbranched alkanes \( \text{C}_{m}\text{H}_{2m+2} \), it is found\(^{159}\) that the effective range
of \( D_{1}^{\sigma} \) is 15–21 carbon atoms, assuming a cutoff threshold of \( 10^{-6}–10^{-8} \) times the value
of the largest matrix element. This verifies what one should have expected all along,
namely, that antisymmetry is important for essentially all molecular systems that are
presently treatable by \textit{ab initio} quantum chemistry.

The functionals introduced in this section surely violate other \( N \)-representability
constraints as well, and next we analyze the \( D \)-, \( Q \)-, and \( G \)-conditions\(^{160-162}\) for each
of the reconstructed 2-RDMs. These conditions were introduced in Subsection 2.3.5
as positivity requirements for the operators \( \hat{D}_{2} \), \( \hat{Q}_{2} \), and \( \hat{G}_{2} \) (or, equivalently, their
spin components), each of which is a functional of \( \hat{D}_{2} \). We will also investigate the
value of \( \langle \hat{S}^{2} \rangle \), the total spin angular momentum, as calculated from the reconstructed
2-RDMs. For an \( N \)-representable 2-RDM we ought to obtain \( \langle \hat{S}^{2} \rangle = 0 \), since we utilize
a spin-restricted formalism, but we will discover than the reconstructions in question
reduce to this limit only for an idempotent 1-RDM.

Consider the \( \alpha \beta \) spin components of \( \hat{D}_{2} \), \( \hat{Q}_{2} \), and \( \hat{G}_{2} \). Assuming a Hartree-
product form for $\hat{D}_2^{\alpha\beta}$, and introducing the abbreviations

$$H_{ij} = n_i n_j$$  \hspace{1cm} (5.43a)

$$\Xi_{ij} = 1 - n_i - n_j$$  \hspace{1cm} (5.43b)

one obtains

$$Q_{ij,kl}^{\alpha\beta} = \frac{1}{2} \delta_{ik} \delta_{jl} (H_{ij} + \Xi_{ij})$$  \hspace{1cm} (5.44)

and

$$G_{ij,kl}^{\alpha\beta} = \frac{1}{2} \delta_{ik} \delta_{jl} \Lambda_{ij}.$$  \hspace{1cm} (5.45)

These matrices are diagonal, with eigenvalues $Q_{ij,ij}^{\alpha\beta}$ and $G_{ij,ij}^{\alpha\beta}$ that are non-negative for $0 \leq n_i, n_j \leq 1$. Thus the Hartree-product reconstruction of $\hat{D}_2^{\alpha\beta}$ satisfies each of the $D$-, $Q$-, and $G$-conditions. This ought to have been obvious, since the HF wavefunction corresponds to a Hartree product reconstruction of $\hat{D}_2^{\alpha\beta}$.

With $\hat{D}_2^{\alpha\alpha}$ reconstructed according to Eq. (5.18), one obtains

$$Q_{ij,kl}^{\alpha\alpha} = \frac{1}{2} \delta_{ik} \delta_{jl} (H_{ij} + \Xi_{ij}) - \frac{1}{2} \delta_{il} \delta_{jk} [\Xi_{ij} + f(n_i, n_j)]$$  \hspace{1cm} (5.46)

and

$$G_{ij,kl}^{\alpha\alpha} = \frac{1}{2} \delta_{ik} \delta_{jl} [n_i - f(n_i, n_j)] + \frac{1}{2} H_{ik} \delta_{ij} \delta_{kl}.$$  \hspace{1cm} (5.47)

These matrices, along with $\hat{D}_2^{\alpha\alpha}$, exhibit a simple block structure that allows us to obtain most of their eigenvalues analytically.

Consider $\hat{D}_2^{\alpha\alpha}$ first. Its only nonzero elements have the form $D_{ij,ij}^{\alpha\alpha}$ and $D_{ij,jj}^{\alpha\alpha}$, hence $\hat{D}_2^{\alpha\alpha}$ consists entirely of $1 \times 1$ blocks ($D_{ii,ii}^{\alpha\alpha}$) and $2 \times 2$ blocks

$$B_{ij} = \begin{pmatrix}
D_{ij,ij}^{\alpha\alpha} & D_{ij,jj}^{\alpha\alpha} \\
D_{ji,ij}^{\alpha\alpha} & D_{ji,jj}^{\alpha\alpha}
\end{pmatrix}.$$  \hspace{1cm} (5.48)
The $1 \times 1$ blocks contribute eigenvalues

$$d_i = D_{ii;ii}^{aa} = \frac{1}{2} \left[ n_i^2 - f(n_i, n_i) \right]$$

while $B_{ij}$ has eigenvalues

$$d_{ij}^\pm = \frac{1}{2} \left[ n_i, n_j \mp f(n_i, n_j) \right].$$

The complete set of $\hat{D}_2^{aa}$ eigenvalues is $\{d_i\} \cup \{d_{ij}^\pm | i < j\}$. The eigenvalue $d_i$ is associated with the eigenvector $|\chi_i\chi_i\rangle$, while $d_{ij}^\pm$ is associated with the eigenvector $(|\chi_i\chi_j\rangle \pm |\chi_j\chi_i\rangle)/\sqrt{2}$.

$\hat{Q}^{aa}$ has the same block structure as $\hat{D}_2^{aa}$. Eigenvalues arising from the $1 \times 1$ blocks are the same as for $\hat{D}_2^{aa}$, $q_i = d_i$, while the $2 \times 2$ blocks yield eigenvalues $q_{ij}^+ = d_{ij}^+$ and

$$q_{ij}^- = \Xi_{ij} + \frac{1}{2} \left[ n_i, n_j \mp f(n_i, n_j) \right].$$

Lastly, $\hat{G}^{aa}$ consists of $1 \times 1$ blocks, along with a single $\nu \times \nu$ block, where $\nu$ is the number of orbitals. The $\nu \times \nu$ block consists of the elements

$$G_{ii,jj}^{aa} = \frac{1}{2} \left\{ n_i n_j + \delta_{ij} \left[ n_i - f(n_i, n_j) \right] \right\}.$$ 

The $1 \times 1$ blocks have elements $G_{ij,ij}^{aa}$, for $i \neq j$, and thus yield eigenvalues

$$g_{ij} = \frac{1}{2} \left[ n_i - f(n_i, n_j) \right].$$

To summarize, we have derived analytic expressions for all eigenvalues of $\hat{D}_2$, $\hat{Q}_2$, and $\hat{G}_2$ except those arising from a single $\nu \times \nu$ block of $\hat{G}_2^{aa}$. Functional-specific expressions for $d_i$, $d_{ij}^\pm$, $q_{ij}$, and $g_{ij}$ are compiled in Tables 5.2 and 5.3. It follows from these expressions that $d_{ij}^\pm \geq 0$ for each functional, since $f$, $\Lambda_{ij}$, and $\Gamma_{ij}$ are each non-negative on the domain of allowed occupation numbers. Because $\Xi_{ij} + H_{ij}/2 \geq 0$, it
also follows that $q_{ij}^\pm \geq 0$ for each functional. Consequently, these functionals satisfy the $Q$-condition if and only if they satisfy the $D$-condition.

Unfortunately, all of the functionals listed in Table 5.1 violate the $D$-condition, except for the special cases CH(2) and SIC-CH(2), which are equivalent to the HF functional. For any $\zeta > 0$, both CHF($\zeta$) and MCHF have $d_i \leq 0$ and $d_{ij}^+ \leq 0$ for all allowed values of $n_i$ and $n_j$. Moreover, these inequalities are strict except in the case of integer occupancies. For CH($\zeta$), $d_i \leq 0$ and $d_{ij}^+ \leq 0$ for any $\zeta \leq 2$, again with strict inequalities for fractional occupancies. [In the case of CH($\zeta$), exact relations for the homogeneous electron gas already limit acceptable parameters to the interval $\frac{12}{3},$ none of which yields a positive 2-RDM.] Finally, SIC-CH($\zeta$) has the same eigenvalues as CH($\zeta$), except that $d_i \equiv 0$ for the former (by construction). Thus for each of these functionals, half of the eigenvalues of $\hat{D}_2^{\alpha\alpha}$ will always be negative.
The analytic formulas in Tables 5.2 and 5.3 can be used to place rigorous bounds on the magnitudes of both negative and positive eigenvalues of $\hat{D}_2$, $\hat{Q}_2$, and $\hat{G}_2$. However, the general formulas are not too informative, so we defer this exercise until Section 5.5, when definite values of $\zeta$ are chosen for calculations. Let us foreshadow one important point, however. It is readily verified that all eigenvalues of $\hat{D}_2$, $\hat{Q}_2$, and $\hat{G}_2$ are positive when $\hat{D}_2$ is reconstructed from an idempotent 1-RDM, and it is tempting to conclude that positivity violations will therefore be small for spin-compensated states. In practice this is often not the case, however. Upon variational optimization, several of the functionals studied here are found to yield artificially high occupancies for what ought to be weakly-occupied orbitals (see Section 5.5), so that the optimized 1-RDMs are by no means nearly idempotent. This anomalous behavior is explained in Subsection 5.5.2.

Finally in this subsection, let us calculate $\langle \hat{S}^2 \rangle$ for each of the reconstructions. In order to do this, we first express the $N$-electron operator $\hat{S}^2$ as

\[
\hat{S}^2 \overset{\text{def}}{=} \sum_{j,k=1}^{N} \vec{\tau}(j) \cdot \vec{\tau}(k) = \sum_{k=1}^{N} \sum_{j<k}^{N} \left( \vec{\tau}(j) + \vec{\tau}(k) \right)^2 - (N-2) \sum_{k=1}^{N} \left( \vec{\tau}(k) \right)^2 ,
\]

which has the form of a symmetrized sum of one- and two-electron operators. [Compare the generic operator $\hat{W}$ introduced in Eq. (2.1)]. In Eq. (5.54), $\vec{\tau}(k)$ represents the one-electron total spin operator, acting on the $k$th electron. Since $\vec{\tau}^2 |\sigma\rangle = \frac{3}{4} |\sigma\rangle$, the one-electron part of $\hat{S}^2$ is proportional to the unit operator, hence

\[
\left\langle -(N-2) \sum_{k=1}^{N} (\vec{\tau}_k)^2 \right\rangle = -\frac{3}{4} N(N-2) .
\]
Denote the basic two-electron operator in Eq. (5.54) as

\[ \hat{g}_{(j,k)} = \left( \hat{s}_j + \hat{s}_k \right)^2. \]  

(5.56)

Then \( \langle \hat{S}^2 \rangle = -\frac{3}{4} N(N - 2) + (\hat{g}) \), where \( \langle \hat{g} \rangle = \text{tr}(\hat{g} \hat{D}_2) \). To obtain this expectation value, observe that \( \hat{g} \) is precisely the total \( \hat{S}^2 \) operator for a two-electron system. In terms of two-electron spin dyads, this operator is\(^{163}\)

\[ \hat{g} = 2 |\alpha \otimes \alpha \rangle \langle \alpha \otimes \alpha| + 2 |\beta \otimes \beta \rangle \langle \beta \otimes \beta| + |\alpha \otimes \beta \rangle \langle \alpha \otimes \beta| + |\beta \otimes \alpha \rangle \langle \beta \otimes \alpha| + |\beta \otimes \alpha \rangle \langle \beta \otimes \alpha| + |\alpha \otimes \beta \rangle \langle \alpha \otimes \beta|. \]  

(5.57)

This form of \( \hat{g} \) makes it especially easy to evaluate \( \text{tr}(\hat{g} \hat{D}_2) \), since each dyad in Eq. (5.57) merely picks out the corresponding spin component of \( \hat{D}_2 \). (Recall that RDM spin components are precisely the spatial coefficients of these spin dyads in the expansion of \( \hat{D}_2 \).) Using the permutational symmetry of the spin components, the result can be written in terms of the three independent 2-RDM spin components:

\[ \langle \hat{g} \rangle = 2 \sum_{i,j=1}^{\nu} \left[ D^{a\alpha}_{ij;ij} + D^{b\beta}_{ij;ij} + D^{a\beta}_{ij;ij} - D^{a\beta}_{ij;ij} \right]. \]  

(5.58)

This affords

\[ \langle \hat{S}^2 \rangle = -\frac{3}{4} N(N - 2) + 2 \text{tr} \left( \hat{D}^{a\alpha}_{2} + \hat{D}^{a\beta}_{2} + \hat{D}^{a\beta}_{2} \right) - 2 \sum_{i,j=1}^{\nu} D^{a\beta}_{ij;ij}, \]  

(5.59)

which is an exact expression for evaluating \( \langle \hat{S}^2 \rangle \) from the 2-RDM, for arbitrary spin multiplets.

\( \langle \hat{S}^2 \rangle \) expressions appropriate to the reconstruction functionals studied here are listed in Table 5.4; the matrices \( \Lambda \) and \( \Gamma \) in this table have elements that were defined in Eq. (5.23). The RHF \( \langle \hat{S}^2 \rangle \) formula given in the table yields exactly zero for integer occupancies (any closed-shell determinant is necessarily a pure singlet), but we have
Table 5.4: Total spin expectation values from reconstructed 2-RDMs, within a spin-restricted formalism.

<table>
<thead>
<tr>
<th>Functional</th>
<th>$\langle S^2 \rangle$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RHF</td>
<td>$3N/2 - 3\text{tr}(\hat{D}_1^\alpha)^2$</td>
</tr>
<tr>
<td>CH($\zeta$)</td>
<td>$3N/2 - \text{tr}(\hat{D}_1^\alpha)^2 - \text{tr}(\hat{D}_1^\alpha)^\zeta$</td>
</tr>
<tr>
<td>SIC-CH($\zeta$)</td>
<td>$3N/2 - 3\text{tr}(\hat{D}_1^\alpha)^2$</td>
</tr>
<tr>
<td>CHF($\zeta$)</td>
<td>$3N/2 - 3\text{tr}(\hat{D}_1^\alpha)^2 - 2\zeta\text{tr}\mathbf{A}$</td>
</tr>
<tr>
<td>MCHF</td>
<td>$3N/2 - 2\text{tr}(\hat{D}_1^\alpha)^2 - \text{tr}\mathbf{\Gamma}$</td>
</tr>
</tbody>
</table>

left the result unsimplified for comparison with the remaining $\langle S^2 \rangle$ formulas. None of these other functionals yields a pure singlet, except in the limit of an idempotent 1-RDM.

5.3 Qualitative analysis of reconstructed pair densities

That the eigenvalues of the reconstructed 2-RDMs are not all positive implies that the corresponding pair densities can be negative as well, so in this section we set out to analyze the behavior of these pair densities. This analysis is also important in light of the fact that the proposed reconstruction functionals are not antisymmetric. The antisymmetry requirement, which prevents parallel-spin electrons from coalescing, forces the parallel-spin component of the intracule density to zero at its origin, a property that is not respected by the functionals introduced in Section 5.2.

For numerical convenience, this analysis is carried out mainly using the 1-RDM from an accurate CI calculation of the Be ground state,\textsuperscript{164} which is used to reconstruct the 2-RDM according to the various prescriptions introduced in the last section. Pair densities from these reconstructed 2-RDMs are then compared to the CI result. The CI wavefunction that we employ is the same one used in Subsection 4.3.1 to test
the validity of Valdemoro’s reconstruction approximation, and the reader is referred to that subsection for the details of this wavefunction. Given that we are presently interested in the qualitative features of the electron distribution, it is notable that the aforementioned Be wavefunction utilizes a STO basis rather than a Gaussian-type orbital basis, as the latter provides a better description of the electron-nuclear cusp.* 

Requisite one-center, one- and two-electron integrals over STOs can be evaluated analytically. Although Pitzer has published a recursive scheme by which to evaluate the two-electron integrals, owing to some possible typographical errors in that paper, we have chosen to implement our own formulas, which are given in Appendix E. Not all of the reconstructed densities are properly normalized, since not all of the reconstruction functionals in Section 5.2 satisfy the partial trace relation, so in this section the improperly-normalized densities have been scaled so that all plotted densities share a common normalization, \( \int d\mathbf{x}_1 \, d\mathbf{x}_2 \, \rho_2(\mathbf{x}_1, \mathbf{x}_2) = 6 \) for Be.

Our analysis based on the aforementioned CI 1-RDM, which was published originally in Ref. 145, represents a legitimate test of the proposed reconstruction functionals using “typical” NOs and occupation numbers. However, one could argue that we ought properly to have optimized both the orbitals and the occupancies by variational minimization of \( E[\hat{D}_1] \). To deflect such criticism, in Subsection 5.3.4 we present a few unpublished examples of pair densities reconstructed using optimized 1-RDMs.

The two-electron density \( \rho_2(\mathbf{x}_1, \mathbf{x}_2) \) is a spin-indexed, six-dimensional function, so an immediate concern is how one may best summarize this information. We reduce the dimensionality in several ways. First, the spin variables are eliminated by considering

\[ * \text{Unfortunately, it can be shown}^{165} \text{ that no finite CI expansion can satisfy the Kato-Smith-Thakkar}^{165,166} \text{ cusp condition on } I(u), \text{ Eq. (2.95)}. \text{ To properly describe this electron-electron coalescence cusp, one needs explicitly correlated basis functions.} \]
either charge densities such as $\rho_2^e$ or spin-resolved densities $\rho_2^{e\mu}$. (These quantities were introduced in Subsection 2.3.4.) For fixed $r_1$, we can easily enough examine conditional probability distributions for the second electron, such as $\rho_2^e(r_1|r_2, \theta_2)$ for example. (There is no dependence on the azimuthal angle $\varphi_2$ because the system is spherically symmetric.) The notation $f(x|y)$ is used here in the same sense as in Subsection 5.2.2, namely, $x$ is a fixed parameter and $y$ represents the independent variable on which $f$ depends. In addition, we will examine the spherically-averaged intracule and extracule densities, $I(u)$ and $X(R)$, obtained from the reconstructed 2-RDMs. (These quantities were defined in Subsection 2.4.)

A few values of the adjustable parameter $\zeta$ are chosen for the calculations in this and subsequent sections. For CH($\zeta$) and SIC-CH($\zeta$), we choose $\zeta = 1$, $\zeta = \frac{4}{3}$, and $\zeta = \frac{4}{5}$. The former is an obvious choice, while the latter two values represent upper\textsuperscript{126} and lower\textsuperscript{148} bounds on the range of acceptable $\zeta$ values for SIC-CH($\zeta$), as determined by analysis of the homogeneous electron gas.* For CHF($\zeta$), we consider $\zeta = 0.7$, $\zeta = 1$, and $\zeta = 1.12$. When $\zeta \approx 0.7$, CHF($\zeta$) yields diatomic potential curves with correct shape, while CHF(1.12) produces accurate energies for the Be isoelectronic sequence.\textsuperscript{147}

Although the proposed reconstructions generally produce qualitatively (and, in some cases, quantitatively) correct pair densities for electrons of opposite spin, none produces a physically reasonable result for electrons with parallel spins. As a benchmark, we will compare these results to those predicted by the “exchange-only” reconstruction $D_1 \wedge D_1$, corresponding to a complete neglect of the two-electron cumulant.

*However, the same authors who derived these bounds on $\zeta$ later concluded\textsuperscript{168} that SIC-CH($\zeta$) does not provide a completely satisfactory description of the electron gas, for any value of $\zeta$. 
(Insofar as \(D_1\) is a correlated 1-RDM obtained from a CI calculation, \(D_1 \wedge D_1\) does not represent the HF approximation, and in fact is superior to the HF result.) With the exception of the antisymmetric \(D_1 \wedge D_1\) reconstruction, each of the aforementioned functionals leads to negative values of various same-spin pair densities, not only at small values of the electron-electron distance, but in some cases all the way out to a separation of about 1.0 \(a_0\), by which point the density has decayed nearly to zero. This unphysical behavior is ultimately related to the fact that only \(D_1 \wedge D_1\) corresponds to an antisymmetric 2-RDM, and suggests that future development of 1-RDM functional theories should be couched in terms of explicitly antisymmetric reconstructions of the 2-RDM. This last point is a recurring theme in the present chapter.

5.3.1 Two-electron conditional densities

In Fig. 5.2 we plot the radial distribution function \(4\pi r^2 \rho_1^\circ(r)\) for the approximations and the CI calculation. The CHF(1) and CH(\(\zeta\)) 2-RDMs satisfy the partial trace relation exactly, so each affords the same 1-CD \(\rho_1^\circ(r)\) as the input CI 1-RDM. The SIC-CH(\(\zeta\)) and \(D_1 \wedge D_1\) densities are identical to one another but deviate somewhat from the CI density. They are superior to HF in the second electron shell.

To compare 2-CDs \(\rho_2^\circ\), we fix one electron at a distance \(r_1\) from the nucleus and use this electron to define the \(z\) axis (\(\theta_1 = 0 = \varphi_1\)). We then examine \(\rho_2^\circ\) as a function of \(r_2\) and \(\theta_2\). We find that the 2-CD from SIC-CH(4/5) becomes negative as the two electrons approach one another, as shown in Fig. 5.3 for \(r_1 = 0.1 a_0\). This negativity is most severe when \(r_1\) and \(\theta_2\) are small (the case shown in Fig. 5.3), but for \(\theta_2 \approx 0\) persists all the way out to \(r_1 \approx 1.0 a_0\). In addition, when \(\theta_2 \gtrsim 108^\circ\), the 2-CD for the SIC-CH(4/5) reconstruction exhibits a local maximum (at fixed \(\theta_2\)) away from
Figure 5.2: Radial distribution function for the Be ground state. The CI result is indistinguishable from the RHF density within the first electron shell.

$r_2 = 0$. This maximum is not present in the corresponding CI density, which increases monotonically as $r_2 \to 0$, for each fixed value of $\theta_2$.

Also plotted in Fig. 5.3 for the SIC-CH(4/5) approximation is the function $\langle \rho_2^{\alpha} \rangle_{\Omega_2}$, in which the notation $\langle \cdots \rangle_{\Omega_2}$ denotes integration over the angular coordinates $\Omega_2$ for the second electron. Again $r_1$ is fixed. Notably, the common practice of spherical averaging disguises the negativity of the density as well as its unrealistic deviations from monotonicity.

Based upon a thorough numerical examination at many fixed values of $r_1$, we conclude that the remaining reconstruction functionals considered here produce positive 2-CDs that display the correct monotone behavior as $r_2 \to 0$. However, analysis of the spin-resolved pair densities $\rho_2^{\alpha\alpha}$ and $\rho_2^{\alpha\beta}$ reveals that this is somewhat misleading. In Fig. 5.4 we plot $\langle \rho_2^{\alpha\alpha} \rangle_{\Omega_2}$, the spherically-averaged pair density for parallel-spin elec-
Figure 5.3: Two-electron charge density for the SIC-CH(4/5) functional with $r_1 = 0.1\ a_0$ and $\theta_1 = 0 = \varphi_1$. The solid line represents the spherical average over the second electron.

electrons. The data in the figure are for $\theta_1 = 0 = \varphi_1$ and $r_1 = 0.2655\ a_0$, the latter being the most probable radial distance for a single electron (see Fig. 5.2). The same-spin pair density from $D_1 \wedge D_1$ is positive, as we showed in Subsection 5.2.3 that it must be, but for all of the other functionals this density has both positive and negative regions. The negative regions are most pronounced when $\theta_2$ is small, but even for the maximum value $\theta_2 = 180^\circ$, the same-spin pair densities are qualitatively incorrect, as illustrated in Fig. 5.5 for two of the reconstruction functionals.

The corresponding CI and $D_1 \wedge D_1$ plots are shown in Fig. 5.6. The CI parallel-spin electron-pair density increases monotonically as $r_2 \to 0$, achieving a maximum value of 0.246 at $r_2 = 0$. The $D_1 \wedge D_1$ reconstruction is at least qualitatively correct in this respect, although for small $r_2$ this density is too large (Fig. 5.4) and for large $\theta_2$ it does not correctly locate the local minimum (Fig. 5.6). Because $D_1 \wedge D_1$ is anti-
Figure 5.4: Pair densities for same-spin electrons, integrated over $\theta_2$, with $\theta_1 = 0 = \varphi_1$ and $r_1 = 0.2655 \, a_0$.

Figure 5.5: Pair densities for same-spin electrons, with $\theta_1 = 0 = \varphi_1$ and $r_1 = 0.2655 \, a_0$. The corresponding CI plots (Fig. 5.6) increase monotonically as $r_2 \to 0$, reaching a maximum value of 0.246 at $r_2 = 0$. 
symmetric, the corresponding pair density correctly reflects a depletion in probability centered at \( r_2 = 0.2655 \, a_0 \) and \( \theta_2 = 0 \), due to Pauli repulsion.

As anticipated, the reconstructed parallel-spin pair densities are fundamentally incorrect, except for the lone antisymmetric reconstruction \( D_1 \wedge D_1 \), while the reconstructed opposite-spin pair densities are good approximations to the CI result. In some sense this latter observation is artifactual, since all of the reconstructions considered here utilize a Hartree-product form for \( D_2^{\alpha\beta} \). Stated differently, the quality of the opposite-spin densities testifies that the Hartree-product form for \( D_2^{\alpha\beta} \) is a good approximation for both idempotent and non-idempotent 1-RDMs.

Because the opposite-spin density is so much larger than the same-spin density, the total, reconstructed 2-CDs are quite close to the CI result, in spite of the unphysical nature of the same-spin distribution. In order to distinguish between the reconstructed
2-CDs we must subtract out the HF contribution to this density. The difference

\[ \rho_2^{\text{corr}}(1|2) \overset{\text{def}}{=} \rho_2(1|2) - \rho_2^{\text{HF}}(1|2), \]

viewed as a function of \( x_2 \) that depends parametrically on \( x_1 \), is sometimes called the correlation hole. In more recent literature, however, what constitutes a hole usually involves a factor of \( 1/\rho_1(1) \), as for example the definitions of the Fermi hole [Eq. (5.34)] and the Coulomb hole [Eq. (5.35)] that were introduced in Subsection 5.2.2. To avoid confusion with these quantities, we refer to \( \rho_2^{\text{corr}} \) as the correlation density, borrowing Löwdin’s usage\(^{169} \) of “correlation” to mean “exact minus HF,” and we take the CI calculation to define “exact” in the present case. Parallel- and antiparallel-spin components of \( \rho_2^{\text{corr}} \) are defined in the usual way.

In Fig. 5.7 we plot the spherical average of \( \rho_2^{\text{corr}}(1|2) \) for each reconstruction functional. As before, the reference electron is located at \( \theta_1 = 0 = \varphi_1 \) and \( r_1 = 0.2655 \ a_0 \). Several trends, present at each fixed value of \( r_1 \), are apparent from this figure. The \( D_1 \wedge D_1 \), SIC-CH(4/3), and CHF(1) reconstructions each produce a correlation hole that is too shallow. In fact, considering that \( D_1 \wedge D_1 \) equals SIC-CH(2), we see that the depth of the correlation hole increases as \( \zeta \) increases. Inclusion of the self-interaction contribution renders the hole more shallow for each \( \zeta \), and because the SIC-CH(1) correlation hole is too deep, this translates into a very good 2-CD produced with CH(1). On the other hand, for \( \zeta = \frac{4}{3} \) the approximation is degraded by the inclusion of self-interaction. As \( r_1 \to 0 \), the absolute deviation from the CI result increases but these trends remain, whereas when \( r_1 \to \infty \) all of the reconstructions approach the CI result.
5.3.2 Intracule densities

The spherically-averaged intracule density $I(u)$ is defined in Eq. (2.94), and in Fig. 5.8 we plot a portion of the function radial intracule density, $4\pi u^2 I(u)$. (The complete plot from a CI calculation can be found in Ref. 170, but the approximations considered here are accurate outside the region plotted in Fig. 5.8.) Shown are the HF and CI results along with two of the approximate reconstructions: CH(1), which is an extremely close match to $I(u)$ over the entire range of $u$; and $D_1 \wedge D_1$, which surprisingly gives the least accurate values for the local minimum and local maxima of $I(u)$, of all the reconstruction functionals considered here. The other reconstructions each lie between the $D_1 \wedge D_1$ and CI values at these extrema.

To better compare the reconstructions, we plot the correlation contribution to the radial intracule, $4\pi u^2[I(u) - I_{HF}(u)]$, in Figs. 5.9–5.11. The $\zeta = 1$ functionals (Fig. 5.10) reproduce the shape of this correlation intracule quite well, meaning that
the exact and the approximate densities \( I(u) \) have their extrema at the same locations. As we saw with \( \rho^{corr}_2 \), the SIC-CH(1) correlation intracule is too deep, but when shifted upwards by the self-interaction gives good agreement with the CI result. The remaining reconstructions (Figs. 5.9 and 5.11) do not locate the extrema in \( I(u) \) as accurately and this leads to correlation intracules that deviate significantly from the CI result in the neighborhood of the second maximum (\( u \approx 2.2 \ a_0 \)).

We next decompose the intracule density into spin components using the spin-resolution of \( \rho^2 \) in Eqs. (2.57) and (2.58). The same-spin intracule \( I_{\alpha\alpha}(u) \) is plotted in Fig. 5.12. With the exception of \( D_1 \wedge D_1 \), which correctly satisfies \( I_{\alpha\alpha}(0) = 0 \), each of the reconstructions has \( I_{\alpha\alpha}(0) < 0 \). This unphysical behavior stems directly from the lack of antisymmetry in these reconstructions. The \( D_1 \wedge D_1 \) result, on the other hand, is indistinguishable from the CI calculation on the scale in Fig. 5.12. For larger values of \( \zeta \) the violation of the Pauli Principle is less severe [recall that SIC-CH(2) is
Figure 5.9: Radial correlation intracule for Be.

Figure 5.10: Radial correlation intracule for Be.
equivalent to $D_1 \wedge D_1$, but the presence or absence of the self-interaction makes little difference.

As explained above, $D_2^{α3}$ is the same for each functional considered here, and the corresponding intracule density is expected to approximate the CI result reasonably well. This is shown in Fig. 5.13. Quantitatively, this density is essentially the same as HF (which, for clarity, is not shown in the figure) when $u \lesssim 3 a_0$. For $u > 3 a_0$, the HF density lies between the reconstruction and the CI result.

5.3.3 Extracule densities

We now focus on the spherically-averaged center-of-mass distribution for electron pairs, $X(R)$, and immediately turn our attention to its same-spin and opposite-spin components. For the former it is found that all approximate reconstructions produce qualitatively incorrect forms for $X_{aa}(R)$, as shown in Fig. 5.14. To avoid clutter the CH($ζ$) plots are not included in this figure since the self-interaction terms do not change the extracule density significantly. The same-spin extracule density from CHF(1) is almost
Figure 5.12: Radial intracule densities for parallel-spin electrons. The SIC-CH(\(\zeta\)) results do not differ appreciably from their self-interacting counterparts so the latter are not shown. On this scale, the \(D_1 \wedge D_1\) density is indistinguishable from the CI result, which is small but nonzero.

Figure 5.13: Radial intracule densities for two electrons of opposite spin. All of the reconstructions give the same result when CI occupation numbers are used.
identical to that of the SIC-CH(4/3) functional and is also excluded from Fig. 5.14. It is worth pointing out that a plot of the function $4\pi R^2 X_{aa}(R)$ washes out all of the small-$R$ features that distinguish among the reconstructions in Fig. 5.14.

5.3.4 Pair densities from optimized density matrices

In Table 5.5 we list the Be ground state energies obtained from the reconstructed 2-RDMs. These are not proper DMFT energies as they are not obtained by minimizing the DMFT functional in Eq. (5.20), but instead employ HF orbitals and CI occupation numbers. Although the functionals CHF(1), CH(1), SIC-CH(4/3), and $D_1 \land D_1$ yield energies that are closer to the CI energy than is HF, many of the energies differ from the CI result by a large amount. In particular, the SIC-CH(1) energy differs from CI by $0.120E_h$, while Goedecker and Umrigar\textsuperscript{124} report an essentially correct energy of $-14.667E_h$ obtained by minimizing the SIC-CH(1) energy functional with respect to
Table 5.5: Unoptimized energies for the Be ground state, from 2-RDMs reconstructed using CI\textsuperscript{164} occupation numbers.

<table>
<thead>
<tr>
<th>Reconstruction</th>
<th>$E/E_h$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_1 \wedge D_1$</td>
<td>-14.60418</td>
</tr>
<tr>
<td>CHF</td>
<td>-14.72130</td>
</tr>
<tr>
<td>SIC-CH(4/5)</td>
<td>-14.98061</td>
</tr>
<tr>
<td>CH(4/5)</td>
<td>-14.87729</td>
</tr>
<tr>
<td>SIC-CH(1)</td>
<td>-14.78407</td>
</tr>
<tr>
<td>CH(1)</td>
<td>-14.68075</td>
</tr>
<tr>
<td>SIC-CH(4/3)</td>
<td>-14.66343</td>
</tr>
<tr>
<td>CH(4/3)</td>
<td>-14.60963</td>
</tr>
<tr>
<td>RHF</td>
<td>-14.57299</td>
</tr>
<tr>
<td>CI (Ref. 164)</td>
<td>-14.66419</td>
</tr>
</tbody>
</table>

variations in the 1-RDM. (Up to the choice of basis set, this accuracy is reproduced in our own calculations; see Section 5.5.) This indicates that the energy-optimized occupation numbers, and especially the energy-optimized orbitals, differ substantially from the CI NOs and natural occupation numbers. (The energy-optimized orbitals also differ substantially from the RHF orbitals, as we shall discover in Sections 5.5 and 5.6.)

Given this discrepancy, one might question the validity of the comparisons in this section. Thus we close this section by presenting a few Be pair densities obtained from self-consistently-optimized 2-RDMs. These are obtained by minimizing the functional $E[\{n_i\}, \{\chi_i\}]$ with respect to variations in both the orbitals (which are represented in the same STO basis as the CI calculation) and the occupation numbers. Discussion of the actual computational procedure for minimizing the density matrix functionals is deferred until the next section.

For the SIC-CH(1) and CHF(1) functionals, Figs. 5.15 and 5.16 compare the...
same-spin and opposite-spin intracule densities reconstructed from CI occupation numbers to those reconstructed using energy-optimized occupation numbers. Only the occupation numbers, not the orbitals, contribute directly to the reconstruction, though optimization of the $n_i$ is certainly coupled to the orbital optimization, and both are optimized in the results reported here. Compared to using CI occupancies, the optimized results change the anomalous value of $I_{\alpha\alpha}(0)$ somewhat—increasing it, in the case of CHF(1), but rendering it more negative, in the case of SIC-CH(1)—and also the optimized CHF(1) result moves the opposite-spin electrons closer together, on average. On the whole, however, the results are not too different from those discussed in Subsection 5.3.2. The same can be said of the opposite-spin extracule densities, which are plotted in Fig. 5.17.

The same-spin extracule densities are a different story. As shown in Fig. 5.18 for
Figure 5.16: Be opposite-spin intracule densities calculated using CI\textsuperscript{164} and DMFT occupation numbers. The volume factor $4\pi u^2$ is omitted in the inset. SIC-CH(1) with CI occupation numbers is nearly indistinguishable from the same functional using optimized occupation numbers, so the former data are not shown.

Figure 5.17: Be opposite-spin extracule densities calculated using optimized DMFT occupation numbers.
the CHF(1) and SIC-CH(1) functionals, use of optimized occupation numbers produces a qualitative change in the distribution. For one thing, the energy-optimized $X_{\alpha\alpha}(R)$ densities are positive—quite positive, in fact, with maxima at $R = 0$. Positivity of $X_{\alpha\alpha}(R)$ is not guaranteed by the form of the reconstruction, as reflected in the fact that use of CI occupation numbers leads to same-spin extracule densities that are relatively small in magnitude, and negative at the origin.

The actual CI extracule density $X_{\alpha\alpha}(R)$ is also shown in Fig. 5.18. Like the optimized reconstructions, it has a maximum at $R = 0$ and decays rapidly, but not quite monotonically; as $R$ increases. Mathematically, a large value of $X_{\alpha\alpha}(0)$ can imply one or both of two distinct phenomena. On the one hand, for pairs of same-spin electrons located near the nucleus, $r_1 \approx 0 \approx r_2$, hence a high probability of

*Just barely evident in Fig. 5.18 is a very small local maximum in $X_{\alpha\alpha}(R)$ at $R \approx 0.5 a_0$. This local maximum is present in the reconstructed extracule densities as well.
such pairs corresponds to a large value of $X_{\alpha\alpha}(R)$ at small $R$. Alternatively, if the
two electrons are far from the nucleus, they can still contribute to $X_{\alpha\alpha}(0)$ so long as
$r_1 = -r_2$. The latter phenomenon, which is a form of angular correlation, accounts
for the fact that $X_{\alpha\alpha}(0) > 0$ for the CI wavefunction, since Pauli repulsion excludes
the former explanation. For non-antisymmetric reconstructions, however, things are
not so clear. Angular correlation seems like a less plausible explanation in the case of
the density matrix functionals, given that the dramatic change in $X_{\alpha\alpha}(0)$ upon using
optimized occupation numbers is not accompanied by a large change in $I_{\alpha\alpha}(u)$, which
would indicated that electron pairs are moving farther apart, to locations $r_1$ and $-r_2$.
Probably the behavior of $X_{\alpha\alpha}(R)$ is related to the overall deficiency in the description
of same-spin electrons.

5.4 Variational procedure

Before presenting numerical results from DMFT, let us first describe how the calcu-
lations are carried out. In contrast to the HF method (restricted or unrestricted),
minimization of a more general density matrix functional with fractional occupation
numbers cannot generally be cast as a self-consistent eigenvalue problem. Instead, di-
rect minimization of $E[n_i, \{\chi_i\}]$ is required, subject to the constraints $0 \leq n_i \leq 1$,
$\sum_i n_i = N/2$, and $\langle \chi_i | \chi_j \rangle = \delta_{ij}$. This section describes our algorithm\textsuperscript{133} for performing
this optimization, which differs from other procedures\textsuperscript{144,146,171,172} in several important
respects. These differences are pointed out in the description that follows.

\textsuperscript{*}The CHF($\zeta$) functionals are an exception. Changing only some signs in the working equations, opti-
mization of the CHF($\zeta$) energy expression is equivalent\textsuperscript{147} to self-consistent solution of the Hartree-
Fock-Bogoliubov\textsuperscript{151} one-particle Hamiltonian. It is not yet clear whether this approach can be gen-
eralized to other density matrix functionals.
Bounds on the occupation numbers are enforced by setting

\[ n_i = (\cos \omega_i)^2 \]  

(5.61)

and varying the \( \omega_i \) without constraint. The trace constraint is implemented by minimizing the penalized energy function

\[ \tilde{E} \overset{\text{def}}{=} E + K(N - 2\sum n_i)^2, \]  

(5.62)

in which the penalty parameter \( K \) is a large, positive constant. Having minimized \( \tilde{E} \) for a particular value of \( K \), the energy can be re-optimized rapidly using a larger value of \( K \), in order to improve the trace, since this requires very little rate-limiting orbital re-optimization.

An alternative to a penalty function is to parametrize the vector of occupation numbers in a manner that automatically preserves the trace.\(^{171} \) In our experience, however, freedom to violate the trace constraint accelerates convergence by enabling relatively large changes in the occupation numbers at early stages in the optimization. Furthermore, relaxation of the trace constraint can be crucial as a means to extricate the optimization from a local minimum that is not the global minimum.

Analytic derivatives

\[ \frac{\partial E}{\partial \omega_i} = -2(\sin 2\omega_i) \frac{\partial E}{\partial n_i} \]  

(5.63)

are easily computed, but we find them unnecessary and not cost effective. Note that \( n_i = 0 \) and \( n_i = 1 \) correspond, respectively, to \( \omega_i = (k + \frac{1}{2})\pi \) and \( \omega_i = 0 \). In both cases \( \sin(2\omega_i) = 0 \), hence \( \partial E/\partial \omega_i = 0 \) if \( n_i \) is an integer. We typically initiate the optimization using RHF orbitals but using occupation numbers that are perturbed slightly from their RHF values.
Orbital optimization, subject to orthonormality constraints, is carried out in a manner similar to that used to perform diagonalization-free self-consistent field calculations.\cite{173-182} The natural orbitals $|\chi_i\rangle$ are represented in terms of some fixed, orthonormal basis set,

$$|\chi_i\rangle = \sum_j C_{ij} |\psi_j\rangle,$$

with initial guess coefficients that form an orthogonal matrix $C$. Any variation $\tilde{C} = C + \delta C$ that preserves orthonormality can be parametrized as

$$\tilde{C} = \exp(A) C \approx (I + A + \frac{1}{2}A^2) C,$$

where $A = -A^\top$ is a skew-symmetric matrix of auxiliary variational parameters. We vary the $A_{ij}$, for fixed $C$ and fixed occupation numbers, so as to minimize $\tilde{E}$, which is evaluated at each optimization step using the orbital coefficients $\tilde{C}_{ij}$. Only when $i < j$ does $A_{ij}$ constitute an independent variational parameter, which we denote by $a_{ij}$. Using the fact that $\partial A_{mn}/\partial a_{kl} = \delta_{mk} \delta_{nl} - \delta_{ml} \delta_{nk}$, one obtains orbital gradients

$$\frac{\partial \tilde{E}}{\partial a_{ij}} = \sum_{mn} \frac{\partial \tilde{E}}{\partial \tilde{C}_{mn}} \left( \frac{\partial \tilde{C}_{mn}}{\partial A_{ij}} - \frac{\partial \tilde{C}_{mn}}{\partial A_{ji}} \right),$$

where $i < j$ is assumed but the summation indices are unrestricted.

As indicated in Eq. (5.65), we truncate $\exp(A)$ at second order, which allows us to evaluate the derivatives $\partial \tilde{C}_{mn}/\partial A_{ij}$ in closed form. This approximation notwithstanding, $\tilde{C}$ is orthogonal and therefore the derivatives $\partial \tilde{E}/\partial \tilde{C}_{mn}$ can be evaluated directly from the energy functional, whereas the derivatives $\partial \tilde{E}/\partial C_{mn}$ cannot. In actuality $\tilde{C}$ is only orthogonal through third order in $A$,

$$\tilde{C}^\top \tilde{C} = I + \frac{1}{4} A^4 + O(A^5),$$
so the algorithm described below includes a procedure to re-orthogonalize \( \mathbf{C} \) as necessary.

Let \( (\partial \tilde{C}_{kl}/\partial A_{ij})^{(n)} = (1/n!)[\partial(\mathbf{A}^n \mathbf{C})_{kl}/\partial A_{ij}] \) be the derivative of the \( n \)th term in the expansion of \( \mathbf{C} \). This induces an order-by-order partition of \( \partial E/\partial a_{ij} \), according to Eq. (5.66). Denote the \( n \)th term in this expansion by \( (\partial E/\partial a_{ij})^{(n)} \), so that

\[
\frac{\partial E}{\partial a_{ij}} = \sum_{n=1}^{\infty} \left( \frac{\partial E}{\partial a_{ij}} \right)^{(n)}. \tag{5.68}
\]

The first few terms in the gradient expansion are

\[
\left( \frac{\partial E}{\partial a_{ij}} \right)^{(1)} = \sum_{k=1}^{\nu} \left[ \frac{\partial E}{\partial C_{ik}} C_{jk} - \frac{\partial E}{\partial C_{jn}} C_{ik} \right], \tag{5.69a}
\]

\[
\left( \frac{\partial E}{\partial a_{ij}} \right)^{(2)} = \frac{1}{2} \sum_{k=1}^{\nu} \left[ \frac{\partial E}{\partial C_{ik}} (\mathbf{AC})_{jk} - \frac{\partial E}{\partial C_{jk}} (\mathbf{AC})_{ik} \right. \right.

\left. \left. + \sum_{l=1}^{\nu} \frac{\partial E}{\partial C_{kl}} (C_{jl} A_{ki} - C_{il} A_{kj}) \right] \tag{5.69b}
\]

\[
\left( \frac{\partial E}{\partial a_{ij}} \right)^{(3)} = \frac{1}{6} \sum_{k=1}^{\nu} \left[ \frac{\partial E}{\partial C_{ik}} (\mathbf{A}^2 \mathbf{C})_{jk} - \frac{\partial E}{\partial C_{jk}} (\mathbf{A}^2 \mathbf{C})_{ik} \right. \right.

\left. \left. + \sum_{l=1}^{\nu} \frac{\partial E}{\partial C_{kl}} [A_{ki} (\mathbf{AC})_{jl} - A_{kj} (\mathbf{AC})_{il} + C_{jl} (\mathbf{A}^2)_{ki} - C_{il} (\mathbf{A}^2)_{kj}] \right] \right\}. \tag{5.69c}
\]

In practice we take

\[
\frac{\partial E}{\partial a_{ij}} \approx \left( \frac{\partial E}{\partial a_{ij}} \right)^{(1)} + \left( \frac{\partial E}{\partial a_{ij}} \right)^{(2)}, \tag{5.70}
\]

consistent with Eq. (5.65).

Let us digress briefly to compare this method to an alternative optimization technique\(^{125,144,183}\) in which the expansion coefficients \( C_{ij} \) are varied directly. As a result \( \mathbf{C} \) is not orthogonal during the course of the optimization, but for sufficiently small variations an orthogonal matrix \( \tilde{\mathbf{C}} \) can be obtained by a first-order approximation.
to Löwdin’s symmetric orthogonalization formula,\textsuperscript{184}

\[ \tilde{C} = C(C^\top C)^{-1/2} \approx \frac{3}{2} C - \frac{1}{2} CC^\top C. \] (5.71)

This relationship is then used to encode the orthogonality constraints into the gradient, via the chain rule

\[ \frac{\partial \tilde{E}}{\partial C_{ij}} = \sum_{mn} \frac{\partial \tilde{E}}{\partial \tilde{C}_{mn}} \frac{\partial \tilde{C}_{mn}}{\partial C_{ij}}. \] (5.72)

As before, the derivatives \( \partial \tilde{E}/\partial \tilde{C}_{mn} \) are obtained directly from then energy functional, while \( \partial \tilde{C}_{mn}/\partial C_{ij} \) is gleaned from Eq. (5.71).

For \( \nu \) orbitals, the symmetric orthogonalization scheme employs \( \nu^2 \) variational parameters for the orbital optimization (which is the rate-limiting step), compared to \( \nu(\nu - 1)/2 \) for the exponential parametrization. Both methods preserve orthogonality only in the limit of infinitesimal steps, though we find this restriction to be much less severe in the case of the exponential parametrization, presumably because we use a second-order approximation to \( \exp(A) \) as compared to the first-order approximation to \( (C^\top C)^{-1/2} \) in Eq. (5.71). As corroborated by Cohen and Baerends,\textsuperscript{144} when the first-order symmetric orthogonalization scheme is used, we find that the orbital coefficient matrix must be re-orthogonalized, via iterative application of Eq. (5.71), each time the \( C_{ij} \) are updated. Using the exponential parametrization, 10–30 updates to \( A \) can be made, at early stages in the optimization, before the matrix \( \tilde{C} = (I + A + \frac{1}{2}A^2)C \) deviates from orthogonality by more than \( 10^{-6} \), as judged by the magnitudes of the matrix elements of \( \tilde{C}^\top \tilde{C} - I \). In the vicinity of the minimum, \( A \) can be updated 100–200 times before this threshold is breached. It is only at this point that we re-orthogonalize, using a few iterations of the first-order orthogonalization formula in Eq. (5.71). These intermediate orthogonalizations are deleterious to the orbital optimization, as they
degrade the quality of the iteratively-constructed inverse Hessian and thus adversely affect the selection of a line search direction. It is therefore significant that the second-order exponential parametrization can maintain orthogonality over a large number of optimization steps. This robustness explains why we find this parametrization to be more stable than first-order methods based on either symmetric orthogonalization or the exponential parametrization.

The choice of orthogonalization method is also important. Arguably, the most widely-disseminated numerical orthogonalization technique is singular value decomposition, which in most implementations utilizes row and column permutations for greater numerical stability. This kind of pivoting, however, produces an orthogonal matrix that frequently bears little resemblance to the original matrix, even if the latter was nearly orthogonal. This has devastating consequences within an optimization algorithm. In contrast, the first-order scheme in Eq. (5.71) is more benign and does not alter the coefficient matrix too much, provided that this matrix is almost orthogonal. This is a well-known property of the symmetric orthogonalization technique.

For definiteness, we list our complete optimization algorithm for $E[\hat{D}_1]$.

1. Minimize $\tilde{E}(\omega_1, \ldots, \omega_m)$ for a fixed set of orbitals, using Powell’s quadratically convergent algorithm, typically without analytic first derivatives.

2. For fixed occupation numbers (including any residual trace error), optimize the orbitals using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) variable metric algorithm.

   (a) Pass $A = 0$ and an orthogonal matrix $C$ to the BFGS subroutine. $C$ is not changed within this routine.
(b) Update the parameters \( a_{ij} \) by a single BFGS step, using the derivatives in Eq. (5.66), and thereby update the matrix \( \bar{C} \) according to Eq. (5.65).

(c) If the BFGS procedure has converged, or if the magnitude of any element of \( \bar{C}^T \bar{C} - I \) exceeds a specified tolerance, proceed to Step 3. Otherwise repeat Step 2b.

3. Set \( C = \bar{C} \) and \( A = 0 \), then return to the main program. Orthogonalize \( C \) by iterative application of Eq. (5.71), then return to Step 2. Repeat this cycle until the orbitals are converged.

4. If either the orbital or the occupancy optimization resulted in a significant energy change, then transform the integrals to the current natural orbital basis, set \( C = I \), and return to Step 1.

5. Once \( \tilde{E} \) is converged with respect to all variational parameters, check the value of the penalty function \( K(N - 2 \sum_i n_i)^2 \). If its value is greater than or comparable to the energy convergence criterion, increase the value of \( K \) (typically by a factor of 100) and return to Step 1.

The sole purpose of the integral transformation in Step 4 is to accelerate optimization of the \( n_i \). By virtue of this transformation, \( C \) is always equal to an identity matrix within the occupancy-optimization subroutine. Formally, the integral transformation (which employs our own implementation of Bender’s algorithm\textsuperscript{186,187}) increases the method’s scaling to \( O(\nu^5) \), but this is not the true bottleneck. With RHF orbitals as the initial guess, the integral transformation is facile relative to the orbital optimization, and thus provides a compelling incentive to decouple the orbital- and
occupancy-optimization steps, which more than compensates for the fact that these two optimizations must be iterated when performed separately. In fact, it is sometimes advantageous to abort the orbital optimization (at Step 3) prior to convergence, then transform the integrals, re-optimize the occupation numbers, and finally begin the orbital optimization anew. In our implementation, this is done whenever the orbitals have been re-orthogonalized 5–10 times (typically corresponding to several hundred BFGS updates) without locating a minimum. There is no mention of an intermediate integral transformation in the existing literature on orbital optimization in DMFT,124,144,146,147,171,172 probably because previous studies have focused on smaller basis sets than the ones employed here.

Note that optimized occupation numbers are passed to the BFGS routine in Step 2 without correcting for any trace error. This is not a problem since the penalty function $K(N - 2\sum_i n_i)^2$ has a vanishing derivative with respect to the orbital parameters $a_{ij}$. Moreover, within the penalty-function approach there is no good way to rectify a trace error except by increasing the value of $K$. Figure 5.19 illustrates the trouble that arises if one attempts to correct the trace between Steps 1 and 2, via uniform scaling of the occupation numbers. The occupancy optimization is responsible for the deep spikes in the penalized energy $\tilde{E}$, which are more pronounced at smaller values of $K$. These features indicate that rescaling of the occupation numbers undoes most of the occupancy optimization. Rescaling eliminates the positive-valued penalty function, but it also shifts occupancy out of the core orbitals and into high-energy virtual orbitals; the net effect is to increase the energy by a large amount.

There are several other points worth noting from Fig. 5.19. First, the data set labeled “variable $K$” in the figure corresponds to the algorithm outlined above (i.e.,
sans rescaling), where $K$ is increased from $10^2$ to $10^8$ in steps of $10^2$, as described in Step 5. Second, note that when $K$ is initially too large (e.g., $K = 10^8$ in Fig. 5.19), the optimization does not converge to the global minimum. In this case, an unfortunately large value of the penalty function has precluded the large changes in the $n_i$ that are necessary in order to move from HF occupancies to the true minimum-energy occupancies. Lastly, the reader ought to take away from Fig. 5.19 just how much longer the orbital optimization takes (in terms of the sheer number of steps) relative to the occupancy optimization.

It bears pointing out that in the special case of the HF functional (restricted or unrestricted), one may set to zero all of the $a_{ij}$ except those that couple occupied and virtual orbitals. This reduces the number of orbital parameters from $\nu(\nu - 1)/2$ to $N(\nu - N)$ and makes HF calculations facile relative to DMFT. The optimizations
performed here bear more similarity, at the algorithm level, to multiconfigurational self-consistent field (MCSCF) calculations,\textsuperscript{181} although we have experienced neither the convergence difficulties, nor the exquisite sensitivity to the choice of starting orbitals, that frequently plague such calculations.

This is not to say that the present incarnation of DMFT is computationally facile, and we close this section with some remarks concerning the practical efficacy of reconstructive DMFT. The analogy to MCSCF theory is highly evocative, and we expect that this analogy will guide the development of improved optimization algorithms. In the immediate future, however, it is likely that this form of DMFT, like MCSCF methods, will be limited to a few tens of orbitals. With our present algorithm, which has hardly been optimized for the problem at hand, we can optimize $\sim 35$ orbitals routinely on a workstation; our largest calculations to date are the ones used to calculate Be pair densities using Bunge’s 48-orbital STO basis\textsuperscript{164} (Subsection 5.3.4). Each of the 48-orbital optimizations required a few days of computer time, starting from HF orbitals. This illustrates the pressing need for starting orbitals that are closer to the optimized ones than are the HF orbitals, which (in the virtual space at least) are exceedingly poor initial guesses and lead to lengthy orbital optimizations. One possibility is to use DMFT itself to optimize atomic orbitals for subsequent use in molecular calculations with the same functional. In this case one would probably want to generalize the orbital optimization algorithm for use with nonorthogonal basis functions.

Certain improvements to our first-generation optimization algorithm also suggest themselves immediately. For instance, orbital optimizations are typically accelerated significantly if the BFGS inverse Hessian (which represents an approximation to the
true Hessian, constructed using only first derivative information) is augmented with analytic formulas for the diagonal of the true orbital rotation Hessian.\textsuperscript{182,188} In other words, one should utilize analytic derivatives $\partial^2 E / \partial (a_{ij})^2$ in the search algorithm. Computation of the off-diagonal second derivatives is usually not cost effective. It is also worth investigating whether the Staroverov-Scuseria method,\textsuperscript{147} in which optimization of the CHF($\zeta$) is written as the solution of a one-electron eigenvalue equation with an orbital- and occupancy-dependent Hamiltonian, can be generalized to other functionals.

Active-space implementations are also a possibility. Suppose that the orbital basis is partitioned into a frozen core $\{ |\chi_1\rangle, \ldots, |\chi_{\nu_c}\rangle \}$ and an active set $\{ |\chi_{\nu_c+1}\rangle, \ldots, |\chi_{\nu}\rangle \}$, with $\nu_c$ frozen orbitals and $\nu_a = \nu - \nu_c$ active orbitals. The orbital optimization scheme introduced in this section is then converted to an active-space procedure by partitioning the matrix $A$ as

$$A = \begin{pmatrix} 0 & 0 \\ 0 & A' \end{pmatrix},$$

in which $A' = -(A')^\top$ is a $\nu_a \times \nu_a$ block of variational parameters, and the rest of $A$ is zero. Partitioning the coefficient matrix $C$ in the same way,

$$C = \begin{pmatrix} C_{cc} & C_{ca} \\ C_{ac} & C_{aa} \end{pmatrix},$$

where $C_{ac}$, for example, has dimension $\nu_a \times \nu_c$, one obtains

$$\tilde{C} = \exp(A) = \begin{pmatrix} 0 & 0 \\ \exp(A') C_{ac} & \exp(A') C_{aa} \end{pmatrix}.$$

It follows that $\partial \tilde{C}_{mn} / \partial A_{ij} = 0$ unless $|\chi_i\rangle$, $|\chi_j\rangle$, and $|\chi_m\rangle$ are active orbitals. ($|\chi_n\rangle$ can be active or frozen.) This leaves only $\nu_a(\nu_a - 1)/2$ variational parameters for the orbital optimization. In addition, one of the summations in Eq. (5.66) that defines the orbital gradients $\partial E / \partial a_{ij}$ can be restricted to active orbitals only.
We have not yet implemented an active-space version of our code, preferring instead to first perform all-electron calculations in order to assess variational stability. If the present functionals prove to be sufficiently accurate, then an active-space procedure will eventually be implemented. We will, however, compare our all-electron LiH results with the minimal active-space calculations of Cohen and Baerends. This comparison, as well as an all-electron study of basis-set dependence, indicates that very small active spaces may be inadequate, and also that great care—and extensive benchmarking to all-electron results—must be exercised when resorting to active-space DMFT. In particular, we show that for certain functionals, active-space and minimal-basis results are qualitatively different from the all-electron results obtained using extended basis sets.

5.5 \( N \)-representability, variational stability, and basis-set dependence

Now that we are in a position to carry out DMFT calculations, our foremost goal is to determine whether existing density matrix functionals are variationally stable, that is, whether the computed electronic energy converges to some finite (hopefully reasonable) result in the complete-basis limit. The reader will recall that this stability has been questioned on the basis of small active-space calculations (three to six active orbitals) on diatomic molecules. To this end, we present in Subsection 5.5.1 the first thorough examination of basis-set dependence in reconstructive DMFT. Optimized DMFT correlation energies \( E_c = E - E_{HF} \) are obtained for LiH and Be, in
a sequence of Pople-type contracted Gaussian basis sets,* and these are compared to
the FCI correlation energies in the same basis sets.

Intimately related to variational stability (or lack thereof) is the issue of \( N \)-
representability. We have demonstrated certain \( N \)-representability violations that are
inherent to the reconstruction functionals of interest, and in Subsection 5.5.2 we quan-
tify these violations using optimized, reconstructed 2-RDMs. These \( N \)-representability
violations (as well as others, not addressed here but undoubtedly present) are respon-
sible for the lack of a variational upper bound in DMFT. Quite generally, we may
state that if a density matrix functional is found to be variational unstable, then \( N \)-
representability is to blame, while if such a functional demonstrates some empirical
stability, then we must puzzle over why certain \( N \)-representability conditions have no
empirical importance. At the risk of spoiling the punch line, let us announce from the
start that the latter scenario is the case, for the functionals considered here.

Before presenting our results, let us say a few words about how the calculations
are carried out. For standard electronic structure calculations† (MP2, CCSD, B3LYP,
etc.) we use the GAMESS\textsuperscript{191} and GAUSSIAN98\textsuperscript{192} programs; only the former has FCI
capability. For political reasons, the author prefers the open-source, non-corporate,
zero-cost GAMESS program, which is also used in this work as an engine to generate one-
and two-electron integrals. However, in the presence of configurational degeneracy,

*Literature references for the standard Gaussian basis sets (both Pople- and Dunning-type) can be
found on pages 414–415 of Ref. 189.
†Let us get the acronyms out of the way all at once. MP2 means second-order perturbation theory
with a RHF reference (that is, Møller-Plesset perturbation theory). CISD and CCSD mean single
and double substitutions within the CI and CC frameworks, respectively. CCSD(T) means CCSD
plus a perturbative, non-iterative estimate of the triples amplitudes. The recent textbook by Hel-
gaker, Jørgensen, and Olsen\textsuperscript{190} provides a modern perspective on these and other standard \textit{ab initio}
methods.
SCF convergence remains something of a black art, and for in such cases GAMESS and GAUSSIAN98 sometimes exhibit very different convergence characteristics, in which case we use whichever program seems to work best. The basis sets employed in our study of Be and LiH are listed in Table 5.6 along with the FCI energy corresponding to each; each of these basis sets is available by keyword in both GAMESS and GAUSSIAN98.

For the $d$ and $f$ shells we use Cartesian Gaussians. For both Be and LiH, the largest basis set consists of 35 orbitals, among which we distribute two electrons, within our spin-restricted formalism. In contrast to several previous studies, we do not freeze any orbitals or occupation numbers.

---

Thus the $d$ shell consists of six functions ($d_{xx}$, $d_{yy}$, $d_{zz}$, $d_{xy}$, $d_{xz}$, and $d_{yz}$) rather than five ($d_{-2}$, $d_{-1}$, $d_0$, $d_{+1}$, and $d_{+2}$). The $f$ shell consists of ten functions rather than seven. Cartesian Gaussians are the default choice in GAMESS; they are selected in GAUSSIAN with the keyword 6d.

---

<table>
<thead>
<tr>
<th>Basis set</th>
<th>$E/E_h$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Be</td>
<td>LiH</td>
<td></td>
</tr>
<tr>
<td>STO-6G</td>
<td>-14.541904</td>
<td>-7.972331</td>
<td></td>
</tr>
<tr>
<td>6-31G</td>
<td>-14.613545</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-31G*</td>
<td>-14.616635</td>
<td>-8.003126</td>
<td></td>
</tr>
<tr>
<td>6-31+G**</td>
<td></td>
<td>-8.009185</td>
<td></td>
</tr>
<tr>
<td>6-311G</td>
<td>-14.632864</td>
<td>-8.019522</td>
<td></td>
</tr>
<tr>
<td>6-311G(′,3p)$^a$</td>
<td></td>
<td>-8.035533</td>
<td></td>
</tr>
<tr>
<td>6-311G(d,2p)</td>
<td></td>
<td>-8.035875</td>
<td></td>
</tr>
<tr>
<td>6-311G+(d,3p)</td>
<td></td>
<td>-8.036749</td>
<td></td>
</tr>
<tr>
<td>6-311G(2d)</td>
<td>-14.633952</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-311G(2df)</td>
<td>-14.640254</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$Contains polarization functions on the H atom only.

Table 5.6: FCI energies for basis sets used in this section.
Table 5.7: Correlation energies for Be, in selected basis sets.

<table>
<thead>
<tr>
<th>Method</th>
<th>$E/E_h$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>STO-6G</td>
</tr>
<tr>
<td>CH(1)</td>
<td>0.074745</td>
</tr>
<tr>
<td>SIC-CH(1)</td>
<td>0.019645</td>
</tr>
<tr>
<td>CH(4/3)</td>
<td>0.004417</td>
</tr>
<tr>
<td>SIC-CH(4/3)</td>
<td>0.002078</td>
</tr>
<tr>
<td>CHF(1)</td>
<td>0.021884</td>
</tr>
<tr>
<td>CHF(1.12)</td>
<td>0.052405</td>
</tr>
<tr>
<td>MCHF</td>
<td>0.061122</td>
</tr>
<tr>
<td>FCI</td>
<td>0.038543</td>
</tr>
</tbody>
</table>

5.5.1 Variational stability

Optimized correlation energies for Be, and for LiH at its experimental\textsuperscript{193} bond length, are listed in Tables 5.7 and 5.8, while Figs. 5.20–5.23 compare the DMFT energies to the FCI ones. Consistent with other results,\textsuperscript{146} the CHF(0.7) functional converges to the RHF solution in each case. (This is not true for LiH at stretched bond lengths, as discussed below.) The CH(4/3) and SIC-CH(4/3) functionals recover very little correlation energy (see Figs. 5.22 and 5.23), yield near-integer occupancies, and are generally quite similar to RHF, except at very large Li–H distance.
<table>
<thead>
<tr>
<th>Method</th>
<th>( E/E_h )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>STO-6G</td>
</tr>
<tr>
<td>CH(1)</td>
<td>0.059132</td>
</tr>
<tr>
<td>SIC-CH(1)</td>
<td>0.016119</td>
</tr>
<tr>
<td>CH(4/3)</td>
<td>0.002369</td>
</tr>
<tr>
<td>SIC-CH(4/3)</td>
<td>0.001285</td>
</tr>
<tr>
<td>CHF(1)</td>
<td>0.016983</td>
</tr>
<tr>
<td>CHF(1.12)</td>
<td>0.039186</td>
</tr>
<tr>
<td>MCHF</td>
<td>0.045783</td>
</tr>
<tr>
<td>FCI</td>
<td>0.020369</td>
</tr>
</tbody>
</table>

Table 5.8: LiH(\( R = 1.5953 \, \text{Å} \)) correlation energies in selected basis sets.
The results in Figs. 5.20 and 5.21 are a bit of good news. There, absolute energies obtained from DMFT are plotted against the corresponding FCI energies, which provide a measure of basis-set completeness. For the CH(4/3) and SIC-CH(4/3) functionals, the DMFT energy exhibits a torpid, RHF-like decrease relative to the FCI energy. Remarkably, for the remaining functionals the DMFT energy decreases at a rate that is not greatly different from that of FCI. There is no reason to expect this a priori, indeed, in view of our results concerning N-representability, the mere absence of any precipitous drops in the DMFT energy with respect to increasing completeness is reassuring.

Figures 5.20 and 5.21 also highlight the fact that certain density matrix functionals are quite sensitive to the presence of polarization functions in the basis set, to the line $E_{DMFT} = E_{FCI}$. The results in Figs. 5.20 and 5.21 are a bit of good news. They are drawn parallel to the line $E_{DMFT} = E_{FCI}$.
Figure 5.21: DMFT versus FCI energies for LiH (at $R = 1.5953 \, \text{Å}$). Dashed guide lines are drawn parallel to the line $E_{\text{DMFT}} = E_{\text{FCI}}$.

Figure 5.22: Be correlation energies, as fractions of the FCI correlation energy.
which has been noted previously.\textsuperscript{146} For both LiH and Be, the FCI energy is (coincidentally) 0.016 $E_h$ lower in the 6-311G basis than in the 6-31G* basis. In contrast, within DMFT the 6-31G* basis set frequently yields a lower energy, and if not then the difference is small. Figure 5.22, and to a lesser extent Fig. 5.23, emphasize the same point. In these figure we illustrate the fraction of FCI correlation energy that is recovered by DMFT, that is, the quantity $100 \times \left( \frac{E_{\text{DMFT}} - E_{\text{RHF}}}{E_{\text{FCI}} - E_{\text{RHF}}} \right)$. One observes—especially in the case of Be, which is more strongly correlated than LiH—that as polarization functions are added to a double-zeta basis, DMFT recovers an increasingly large fraction of the FCI correlation energy. This is to say polarization functions are more effective for introducing correlation in DMFT than they are in FCI. Upon passing from a polarized, double-zeta basis (6-31G*) to a non-polarized, triple-
zeta basis (6-311G), the functionals CH(1), CHF(1), CHF(1.12), and MCHF each exhibit a steep drop in the fraction of the FCI correlation energy that is recovered. In contrast, SIC-CH(1) is far less sensitive to the presence of polarization functions.

Optimized occupation numbers for Be are listed in Table 5.9. Atomic Be is an interesting test case on account of a near-degeneracy between the 2s and 2p manifolds, which leads to a FCI value of $n_{2p}$ that is anomalously large, relative to the occupancies of other weakly-occupied NOs. According to Table 5.9, this “large” occupation is only $\sim 0.03$, and because there are three degenerate $2p$ orbitals, the more significant effect of the 2s-2p quasi-degeneracy is to depress the $2s$ occupation number, which at 0.91 is especially low for an orbital that is occupied in the canonical molecular orbital picture. At a qualitative level, both the anomalously small value of $n_{2s}$ and the anomalously large value of $n_{2p}$ are reproduced by the SIC-CH(1), CH(4/3), and SIC-CH(4/3) functionals. However, the CHF(\(\zeta\)) and MCHF functionals shift far too much occupancy into the virtual orbitals, in both Be and LiH. Occupation numbers for LiH are plotted as a function of internuclear distance in Fig. 5.24. The strong deviations from integer occupancies, in the case of CHF(\(\zeta\)) and MCHF, are explained in the next subsection.

The plots of DMFT versus FCI energy for LiH (Fig. 5.21) are smoother than the corresponding plots for Be (Fig. 5.20), and the three LiH data points representing polarized, triple-zeta basis sets are nearly coincident. However, full potential energy curves reveal important differences among basis sets. Figures 5.25 and 5.26 compare LiH potential curves obtained with the STO-6G, 6-31G*, and 6-311G(\(\cdot,3p\)) basis sets. In the minimal basis set, each functional produces a reasonable potential curve, though certain differences are evident at large internuclear distances, as detailed below.
<table>
<thead>
<tr>
<th>Method</th>
<th>STO-6G</th>
<th>6-31G*</th>
<th>6-311G</th>
<th>6-311G(2d)</th>
<th>6-311G(2df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH(1)</td>
<td>0.7932</td>
<td>0.7039</td>
<td>0.7275</td>
<td>0.6726</td>
<td>0.6578</td>
</tr>
<tr>
<td></td>
<td>0.1034</td>
<td>0.0882</td>
<td>0.0803</td>
<td>0.0903</td>
<td>0.0927</td>
</tr>
<tr>
<td></td>
<td>0.1034</td>
<td>0.0882</td>
<td>0.0802</td>
<td>0.0904</td>
<td>0.0927</td>
</tr>
<tr>
<td>SIC-CH(1)</td>
<td>0.9816</td>
<td>0.9609</td>
<td>0.9629</td>
<td>0.9521</td>
<td>0.9480</td>
</tr>
<tr>
<td></td>
<td>0.0092</td>
<td>0.0101</td>
<td>0.0093</td>
<td>0.0107</td>
<td>0.0112</td>
</tr>
<tr>
<td></td>
<td>0.0092</td>
<td>0.0101</td>
<td>0.0093</td>
<td>0.0107</td>
<td>0.0112</td>
</tr>
<tr>
<td>CH(4/3)</td>
<td>0.9872</td>
<td>0.9836</td>
<td>0.9837</td>
<td>0.9834</td>
<td>0.9833</td>
</tr>
<tr>
<td></td>
<td>0.0064</td>
<td>0.0052</td>
<td>0.0049</td>
<td>0.0050</td>
<td>0.0050</td>
</tr>
<tr>
<td></td>
<td>0.0064</td>
<td>0.0052</td>
<td>0.0049</td>
<td>0.0050</td>
<td>0.0050</td>
</tr>
<tr>
<td>SIC-CH(4/3)</td>
<td>0.9963</td>
<td>0.9946</td>
<td>0.9945</td>
<td>0.9942</td>
<td>0.9940</td>
</tr>
<tr>
<td></td>
<td>0.0019</td>
<td>0.0016</td>
<td>0.0016</td>
<td>0.0016</td>
<td>0.0016</td>
</tr>
<tr>
<td></td>
<td>0.0019</td>
<td>0.0016</td>
<td>0.0016</td>
<td>0.0016</td>
<td>0.0016</td>
</tr>
<tr>
<td>CHF(1)</td>
<td>0.7618</td>
<td>0.6305</td>
<td>0.6666</td>
<td>0.6049</td>
<td>0.5897</td>
</tr>
<tr>
<td></td>
<td>0.1188</td>
<td>0.1173</td>
<td>0.1069</td>
<td>0.1187</td>
<td>0.1215</td>
</tr>
<tr>
<td></td>
<td>0.1195</td>
<td>0.1171</td>
<td>0.1069</td>
<td>0.1188</td>
<td>0.1214</td>
</tr>
<tr>
<td>CHF(1.12)</td>
<td>0.6725</td>
<td>0.5509</td>
<td>0.5820</td>
<td>0.5257</td>
<td>0.5118</td>
</tr>
<tr>
<td></td>
<td>0.1640</td>
<td>0.1413</td>
<td>0.1328</td>
<td>0.1397</td>
<td>0.1409</td>
</tr>
<tr>
<td></td>
<td>0.1640</td>
<td>0.1404</td>
<td>0.1326</td>
<td>0.1396</td>
<td>0.1409</td>
</tr>
<tr>
<td>MCHF</td>
<td>0.7719</td>
<td>0.6692</td>
<td>0.6973</td>
<td>0.6371</td>
<td>0.6212</td>
</tr>
<tr>
<td></td>
<td>0.1141</td>
<td>0.1014</td>
<td>0.0919</td>
<td>0.1035</td>
<td>0.1062</td>
</tr>
<tr>
<td></td>
<td>0.1141</td>
<td>0.1007</td>
<td>0.0917</td>
<td>0.1033</td>
<td>0.1064</td>
</tr>
<tr>
<td>FCI</td>
<td>0.9116</td>
<td>0.9031</td>
<td>0.9046</td>
<td>0.9081</td>
<td>0.9094</td>
</tr>
<tr>
<td></td>
<td>0.0442</td>
<td>0.0319</td>
<td>0.0312</td>
<td>0.0299</td>
<td>0.0295</td>
</tr>
<tr>
<td></td>
<td>0.0442</td>
<td>0.0319</td>
<td>0.0312</td>
<td>0.0299</td>
<td>0.0295</td>
</tr>
</tbody>
</table>

Table 5.9: Largest $\alpha$-spin occupations numbers for Be, excluding a unit-occupied core orbital. Degeneracy of the $p$ orbitals is not enforced during the optimization.
Figure 5.24: Largest LiH/6-311G(\(d, 2p\)) occupation numbers, excluding a unit-occupied core orbital. Also shown are potential curves for SIC-CH(1) (solid line) and CHF(1) (broken line), each scaled by the same factor.

For larger basis sets, however, the CH(1), CHF(\(\zeta\)), and MCHF potentials are far too shallow. [CHF(0.7) and CHF(1.12) results are not shown, but are similar to CHF(1) in this respect.]

With the beneficial hindsight of these all-electron calculations, the same defect is just barely perceptible in the active-space LiH/CHF(1) potential curves calculated by Cohen and Baerends.\(^{144}\) These authors observe that the CHF(1) potential becomes increasingly shallow as the number of active orbitals is increased from three to five, which they interpret as possible evidence of variational instability. However, we observe no further qualitative change upon incorporating additional polarization or diffuse functions [e.g., enlarging the basis from 6-311G(\(\cdot, 3p\)) to 6-311+G(\(d, 3p\))], and thus we conclude that the changes observed by Cohen and Baerends are artifacts of a minimal basis (or minimal active space). The shallow potential curves illustrated
Figure 5.25: LiH potential curves obtained with (a) the STO-6G basis, (b) the 6-31G* basis, and (c) the 6-311G(·, 3p) basis.

In Figs. 5.25 and 5.26 for extended basis sets are similar to those obtained for other diatomic molecules using the same functionals, and we believe that this is the true nature of diatomic potentials for these functionals.

In contrast to the ill behavior of CH(1), CHF(ζ), and MCHF, the functionals SIC-CH(1), CH(4/3), and SIC-CH(4/3) produce reasonable potential curves in each basis set. The latter two functionals yield near-RHF potentials, except at large internuclear separation, while the SIC-CH(1) potentials (Fig. 5.25) are actually quite close to the FCI ones at moderate Li–H distance. The dissociative asymptote for SIC-CH(1) is much too high in the minimal basis but is significantly improved in the other basis sets.

LiH/6-311G(·, 3p) potential curves for SIC-CH(1), CHF(0.7), and CHF(1) are compared in Fig. 5.27. The behavior of the CHF functionals here is especially inter-
Figure 5.26: LiH potential curves obtained with (a) the STO-6G basis, (b) the 6-31G* basis, and (c) the 6-311G(\cdot,3p) basis.

interesting. For $R \lesssim 1.5 R_e$, CHF(0.7) converges to the RHF energy, but at large $R$ the CHF(0.7) potential curve is vastly superior to the RHF potential. Not only is the asymptotic energy lower than RHF, but the inflection point in the potential occurs at a shorter bond length, leading to a more realistic shape. (There is nothing unique about $\zeta = 0.7$ in this regard; $\zeta = 0.65$ and $\zeta = 0.75$, for example, yield similar diatomic potential curves.\textsuperscript{147} CHF(0.7) is also superior to SIC-CH(1) in this respect, for although the SIC-CH(1) potential is quite close to FCI for $R \lesssim 2 R_e$, SIC-CH(1) turns over too slowly in the asymptotic region. For a sufficiently large basis, CHF(1) produces a potential curve that is almost purely repulsive, but whose asymptote yields an essentially exact\textsuperscript{194,195} dissociation energy; the latter fact is probably a coincidence.

Confronted with these facts, it is natural to inquire whether there is some hybrid of SIC-CH(1) and CHF($\zeta$) that looks mostly like the former at modest bond lengths,
yet incorporates the early turnover of CHF(\(\zeta \approx 0.7\)) in the dissociative regime. No such hybrid has been examined to date, though a hybrid functional

\[
E_{\text{hybrid}} = \lambda E_{\text{CH(1)}} + (1 - \lambda) E_{\text{CHF(1)}},
\]

with parameter \(\lambda\) optimized for He atom, has been tested.\(^{146}\) We have demonstrated, however, that for a sufficiently large basis neither of these two constituent functionals generates a proper potential energy curve, so it is not surprising that the above hybrid also yields qualitatively incorrect diatomic potentials.

### 5.5.2 \(N\)-representability

The spectra of \(\tilde{D}_2^{\alpha\alpha}\), \(\tilde{Q}_2^{\alpha\alpha}\), and \(\tilde{G}_2^{\alpha\alpha}\) have been determined, for each variationally-optimized 1-RDM. For all of the functionals except CHF(1.12), we find that the \(\nu \times \nu\) block \((G^{\alpha\alpha}_{\nu;\nu})\) of \(\tilde{G}_2^{\alpha\alpha}\) is positive. For all remaining eigenvalues we possess analytic
<table>
<thead>
<tr>
<th>Functional</th>
<th>Matrix</th>
<th>Largest eigenvalue</th>
<th>Smallest eigenvalue</th>
<th>No. $&lt; -10^{-n}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH(1)</td>
<td>$D_2^{aa}$</td>
<td>0.7344</td>
<td>-0.1126</td>
<td>629 602 65</td>
</tr>
<tr>
<td></td>
<td>$G_2^{aa}$</td>
<td>0.7294</td>
<td>-0.1059</td>
<td>544 433 49</td>
</tr>
<tr>
<td>SIC-CH(1)</td>
<td>$D_2^{aa}$</td>
<td>0.9608</td>
<td>-0.0473</td>
<td>595 438 27</td>
</tr>
<tr>
<td></td>
<td>$G_2^{aa}$</td>
<td>0.9614</td>
<td>-0.0473</td>
<td>542 291 27</td>
</tr>
<tr>
<td>CH(4/3)</td>
<td>$D_2^{aa}$</td>
<td>0.9860</td>
<td>-0.0121</td>
<td>190 45 6</td>
</tr>
<tr>
<td></td>
<td>$G_2^{aa}$</td>
<td>0.9848</td>
<td>-0.0121</td>
<td>100 35 6</td>
</tr>
<tr>
<td>SIC-CH(4/3)</td>
<td>$D_2^{aa}$</td>
<td>0.9950</td>
<td>-0.0062</td>
<td>130 33 0</td>
</tr>
<tr>
<td></td>
<td>$G_2^{aa}$</td>
<td>0.9955</td>
<td>-0.0062</td>
<td>67 33 0</td>
</tr>
<tr>
<td>CHF(1)</td>
<td>$D_2^{aa}$</td>
<td>0.5908</td>
<td>-0.1210</td>
<td>629 492 37</td>
</tr>
<tr>
<td></td>
<td>$G_2^{aa}$</td>
<td>0.6960</td>
<td>-0.0554</td>
<td>528 312 15</td>
</tr>
<tr>
<td>CHF(1.12)</td>
<td>$D_2^{aa}$</td>
<td>0.5137</td>
<td>-0.1399</td>
<td>630 504 46</td>
</tr>
<tr>
<td></td>
<td>$G_2^{aa}$</td>
<td>0.6577</td>
<td>-0.0630</td>
<td>672 371 39</td>
</tr>
<tr>
<td>MCHF</td>
<td>$D_2^{aa}$</td>
<td>0.6973</td>
<td>-0.1178</td>
<td>629 602 63</td>
</tr>
<tr>
<td></td>
<td>$G_2^{aa}$</td>
<td>0.7101</td>
<td>-0.0856</td>
<td>547 390 32</td>
</tr>
</tbody>
</table>

Table 5.10: Spectrum of $\hat{D}_2^{aa}$ and $\hat{G}_2^{aa}$, for Be/6-311G(2df).

formulas, given previously in Tables 5.2 and 5.3. Of these, $d_i$ and $d_{ij}^+$ are necessarily negative (or zero) and $g_{ij}$ can also be negative.

Table 5.10 summarizes the spectra of $\hat{D}_2^{aa}$ and $\hat{G}_2^{aa}$ for Be in the best basis set utilized here, 6-311G(2df). This basis consists of 35 orbitals, so there are 595 $d_{ij}^+$ eigenvalues and 1,190 $g_{ij}$ eigenvalues. As indicated in the table, most of the $d_{ij}^+$ are more negative than $-10^{-4}$, except in the case of the CH(4/3) and SIC-CH(4/3) functionals. Results for the LiH/6-311G(d,2p) 2-RDM (Tables 5.11 and 5.12) are similar. In this case there are 28 orbitals, hence 378 $d_{ij}^+$ eigenvalues and 756 $g_{ij}$ eigenvalues.

The negative eigenvalues of $\hat{D}_2^{aa}$ that are listed in Tables 5.10 and 5.11 should
<table>
<thead>
<tr>
<th>Functional</th>
<th>$R$</th>
<th>Largest eigenvalue</th>
<th>Smallest eigenvalue</th>
<th>No. $&lt; -10^{-n}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH(1)</td>
<td>$R_e$ 0.8411</td>
<td>$-0.1075$</td>
<td>405</td>
<td>364</td>
</tr>
<tr>
<td></td>
<td>$1.5 R_e$ 0.7854</td>
<td>$-0.1176$</td>
<td>405</td>
<td>369</td>
</tr>
<tr>
<td></td>
<td>$3 R_e$ 0.6748</td>
<td>$-0.1249$</td>
<td>405</td>
<td>378</td>
</tr>
<tr>
<td>SIC-CH(1)</td>
<td>$R_e$ 0.9820</td>
<td>$-0.0396$</td>
<td>378</td>
<td>189</td>
</tr>
<tr>
<td></td>
<td>$1.5 R_e$ 0.9753</td>
<td>$-0.0530$</td>
<td>378</td>
<td>230</td>
</tr>
<tr>
<td></td>
<td>$3 R_e$ 0.9271</td>
<td>$-0.0939$</td>
<td>378</td>
<td>307</td>
</tr>
<tr>
<td>CH(4/3)</td>
<td>$R_e$ 0.9950</td>
<td>$-0.0093$</td>
<td>116</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>$1.5 R_e$ 0.9878</td>
<td>$-0.0201$</td>
<td>136</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>$3 R_e$ 0.7673</td>
<td>$-0.0738$</td>
<td>178</td>
<td>51</td>
</tr>
<tr>
<td>SIC-CH(4/3)</td>
<td>$R_e$ 0.9981</td>
<td>$-0.0047$</td>
<td>83</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>$1.5 R_e$ 0.9969</td>
<td>$-0.0082$</td>
<td>90</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>$3 R_e$ 0.9811</td>
<td>$-0.0275$</td>
<td>134</td>
<td>34</td>
</tr>
<tr>
<td>CHF(1)</td>
<td>$R_e$ 0.7771</td>
<td>$-0.0869$</td>
<td>406</td>
<td>283</td>
</tr>
<tr>
<td></td>
<td>$1.5 R_e$ 0.6939</td>
<td>$-0.1064$</td>
<td>406</td>
<td>295</td>
</tr>
<tr>
<td></td>
<td>$3 R_e$ 0.5768</td>
<td>$-0.1222$</td>
<td>406</td>
<td>335</td>
</tr>
<tr>
<td>CHF(1.12)</td>
<td>$R_e$ 0.7048</td>
<td>$-0.1169$</td>
<td>406</td>
<td>337</td>
</tr>
<tr>
<td></td>
<td>$1.5 R_e$ 0.6348</td>
<td>$-0.1301$</td>
<td>406</td>
<td>345</td>
</tr>
<tr>
<td></td>
<td>$3 R_e$ 0.5489</td>
<td>$-0.1388$</td>
<td>406</td>
<td>352</td>
</tr>
<tr>
<td>MCHF</td>
<td>$R_e$ 0.8292</td>
<td>$-0.0911$</td>
<td>405</td>
<td>356</td>
</tr>
<tr>
<td></td>
<td>$1.5 R_e$ 0.7683</td>
<td>$-0.1038$</td>
<td>405</td>
<td>369</td>
</tr>
<tr>
<td></td>
<td>$3 R_e$ 0.6613</td>
<td>$-0.1219$</td>
<td>405</td>
<td>378</td>
</tr>
</tbody>
</table>

Table 5.11: Spectrum of $\hat{D}^\alpha_2$ for LiH/6-311G(d,2p), at several values of $R$. 
<table>
<thead>
<tr>
<th>Functional</th>
<th>$R$</th>
<th>Largest eigenvalue</th>
<th>Smallest eigenvalue</th>
<th>No. $&lt; -10^{-n}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$n = 6$</td>
</tr>
<tr>
<td>CH(1)</td>
<td>$R_e$</td>
<td>0.8196</td>
<td>-0.1075</td>
<td>370</td>
</tr>
<tr>
<td></td>
<td>1.5 $R_e$</td>
<td>0.7722</td>
<td>-0.1176</td>
<td>368</td>
</tr>
<tr>
<td></td>
<td>3 $R_e$</td>
<td>0.7022</td>
<td>-0.1249</td>
<td>370</td>
</tr>
<tr>
<td>SIC-CH(1)</td>
<td>$R_e$</td>
<td>0.9822</td>
<td>-0.0396</td>
<td>367</td>
</tr>
<tr>
<td></td>
<td>1.5 $R_e$</td>
<td>0.9755</td>
<td>-0.0530</td>
<td>367</td>
</tr>
<tr>
<td></td>
<td>3 $R_e$</td>
<td>0.9304</td>
<td>-0.0939</td>
<td>367</td>
</tr>
<tr>
<td>CH(4/3)</td>
<td>$R_e$</td>
<td>0.9946</td>
<td>-0.0093</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>1.5 $R_e$</td>
<td>0.9868</td>
<td>-0.0201</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>3 $R_e$</td>
<td>0.8141</td>
<td>-0.0738</td>
<td>78</td>
</tr>
<tr>
<td>SIC-CH(4/3)</td>
<td>$R_e$</td>
<td>0.9983</td>
<td>-0.0047</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>1.5 $R_e$</td>
<td>0.9972</td>
<td>-0.0082</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>3 $R_e$</td>
<td>0.9797</td>
<td>-0.0275</td>
<td>73</td>
</tr>
<tr>
<td>CHF(1)</td>
<td>$R_e$</td>
<td>0.8120</td>
<td>-0.0567</td>
<td>368</td>
</tr>
<tr>
<td></td>
<td>1.5 $R_e$</td>
<td>0.7591</td>
<td>-0.0612</td>
<td>370</td>
</tr>
<tr>
<td></td>
<td>3 $R_e$</td>
<td>0.7026</td>
<td>-0.0532</td>
<td>368</td>
</tr>
<tr>
<td>CHF(1.12)</td>
<td>$R_e$</td>
<td>0.7580</td>
<td>-0.0700</td>
<td>427</td>
</tr>
<tr>
<td></td>
<td>1.5 $R_e$</td>
<td>0.7185</td>
<td>-0.0711</td>
<td>428</td>
</tr>
<tr>
<td></td>
<td>3 $R_e$</td>
<td>0.6811</td>
<td>-0.0643</td>
<td>452</td>
</tr>
<tr>
<td>MCHF</td>
<td>$R_e$</td>
<td>0.8121</td>
<td>-0.0876</td>
<td>370</td>
</tr>
<tr>
<td></td>
<td>1.5 $R_e$</td>
<td>0.7638</td>
<td>-0.0957</td>
<td>368</td>
</tr>
<tr>
<td></td>
<td>3 $R_e$</td>
<td>0.7009</td>
<td>-0.1027</td>
<td>370</td>
</tr>
</tbody>
</table>

Table 5.12: Spectrum of $\hat{G}_2^{\alpha\alpha}$ for LiH/6-311G($d,2p$), at several values of $R$. 
Table 5.13: Theoretical bounds on the reconstructed eigenvalues of $\hat{D}_{2}^{\alpha}$. The expressions corresponding to CH($\zeta$) and SIC-CH($\zeta$) are identical, except that $d_{i} = 0$ for the latter.

<table>
<thead>
<tr>
<th>Functional</th>
<th>Minimum of $d_{i}(n_{i})$</th>
<th>Minimum of $d_{ij}^{+}(n_{i},n_{j})$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Value Location</td>
<td>Value Location</td>
</tr>
<tr>
<td>CH(1)</td>
<td>-0.1250 $n_{i} = 0.50$</td>
<td>-0.1250 $n_{i} = 1/4n_{j}$</td>
</tr>
<tr>
<td>CH(4/3)</td>
<td>-0.0741 $n_{i} = 0.54$</td>
<td>-0.0741 $n_{i} = 8/27n_{j}$</td>
</tr>
<tr>
<td>CHF(1)</td>
<td>-0.1250 $n_{i} = 0.50$</td>
<td>-0.1250 $n_{i} = n_{j} = 1/2$</td>
</tr>
<tr>
<td>CHF(1.12)</td>
<td>-0.1400 $n_{i} = 0.50$</td>
<td>-0.1400 $n_{i} = n_{j} = 1/2$</td>
</tr>
<tr>
<td>MCHF</td>
<td>-0.1250 $n_{i} = 0.50$</td>
<td>-0.1250 $n_{i} = n_{j} = 1/2$</td>
</tr>
</tbody>
</table>

be compared with the theoretical lower bounds on $d_{i}$ and $d_{ij}^{+}$, which are given in Table 5.13. These bounds are exact, representing for example the values of $n_{i}$ and $n_{j}$ for which $\partial d_{ij}^{+}/\partial n_{i} = 0 = \partial d_{ij}^{+}/\partial n_{j}$, at fixed $\zeta$. For CH(1), CHF(1), and MCHF, the most negative eigenvalue that is obtained in practice is frequently quite close to the most negative eigenvalue that is theoretically possible. This appears to be true especially for atoms (Be, and LiH in the separated-atom limit).

Recall from Subsection 5.2.3 that for each of the functionals considered in this work, $\hat{D}_{2}^{\alpha}$ has a simple block structure consisting of some $2 \times 2$ blocks and some (anomalous) $1 \times 1$ blocks. We have shown that each $2 \times 2$ block $B_{ij}$ [Eq. (5.48)] is associated with exactly one negative and one positive eigenvalue. More generally, we suggest that such is likely to be the case in practice (that is, for variationally-optimized density matrices) for any ansatz exhibiting this simple block structure, unless negative eigenvalues are somehow excluded by design. To understand why, consider the natural
expansion of $\hat{D}^{\alpha}_2$, which has the form

$$\hat{D}^{\alpha}_2 = \sum_i d_i |x_i x_i\rangle\langle x_i x_i| + \sum_{i<j} d^+_{ij} \left( |x_i x_j\rangle\langle x_i x_j| + |x_i x_j\rangle\langle x_j x_i| \right)$$

for any reconstruction exhibiting the same block structure as the ones examined here.

$\hat{D}^{\alpha}_2$ contributes $2 E^{\alpha}_2$ to the electron repulsion energy, where $E^{\alpha}_2 = \text{tr}(r_{12}^{-1} \hat{D}^{\alpha}_2)$, which can be decomposed into a sum of natural geminal pair energies; the pair energies for the natural geminals associated with $d_i$, $d^+_{ij}$, and $d^-_{ij}$ are $\langle ii|ii \rangle$, $\langle ij|ij \rangle + \langle ij|ji \rangle$, and $\langle ij|ij \rangle - \langle ij|ji \rangle$, respectively. Whereas $\langle ij|ij \rangle - \langle ij|ji \rangle$ can be positive or negative, the other two pair energies are non-negative, and consequently the electron repulsion energy is always lowered (stabilized) by admitting negative eigenvalues for $\hat{D}^{\alpha}_2$. Moreover, this unphysical energy lowering occurs via a “direct” mechanism, that is, simply by changing $d_i$ or $d^+_{ij}$, without having to modify the orbitals.

Although the final, optimized energy represents a delicate balance between the positive and the negative eigenvalues of $\hat{D}_2$, all of which depend upon the same occupation numbers $n_i$, it is interesting to note from Table 5.13 that $d_i = d_i(n_i)$ always assumes its most negative possible value when $n_i = \frac{1}{2}$, except for the SIC-CH($\zeta$) functionals, which have $d_i = 0$ for all $n_i$. This explains why the CH(1), CHF($\zeta$), and MCHF functionals each shift far too much occupancy into orbitals that ought to be weakly occupied (e.g., the $2p$ shell of Be), a trend that is exemplified by the Be occupation numbers in Table 5.9 as well as the LiH occupation numbers in Fig. 5.24. First, note that the matrix element $h_{1s,1s}$ is always much more negative (more stabilizing) than other matrix elements of $\hat{h}$, since the $1s$ orbital is nodeless. For example, in the HF/6-311G(2df) basis for Be, $h_{1s,1s} = -7.941 E_h$, whereas $h_{2s,2s} = -1.589 E_h$ and
$h_{2p,2p} = -1.073 \, E_h$. In comparison, $\langle 1s \, 1s|1s \, 1s \rangle = 2.273 \, E_h$, $\langle 2s \, 2s|2s \, 2s \rangle = 0.344 \, E_h$ $\langle 2p \, 2p|2p \, 2p \rangle = 0.247 \, E_h$ in the same basis. Thus the 1s orbital gets essentially double occupancy, irrespective of which functional is employed, but the occupancy of the 2s orbital (which ought to be $\sim 0.9$) is anomalously depleted due to the fact that $d_{2s}$ is most negative for $n_{2s} = \frac{1}{2}$. The partial loss of $h_{2s,2s}$ that arises when a large fraction of an electron is shifted out of the 2s orbital is compensated by the annihilation of an amount of electron-repulsion energy equal to $d_{2s} \langle \chi_{2s} \chi_{2s}|r^{-1}_{12}|\chi_{2s} \chi_{2s} \rangle$, which for fixed orbitals is most negative when $n_{2s} = \frac{1}{2}$.

Of all the functionals examined here, only SIC-CH(1), SIC-CH(4/3), and CH(4/3) yield reasonable progressions of occupation numbers. (This is evident from Table 5.9, and many other calculations confirm this trend.) This is consistent with the explanation above, since $d_i \equiv 0$ for the SIC-CH(\(\zeta\)) functionals. The self-interaction correction thus removes a strong but spurious incentive to shift occupancy outside of highly stabilizing core orbitals. That CH(4/3) affords a reasonable occupation number distribution must be a result of the fact that the minimum value of $d_i(n_i = \frac{1}{2})$, is only 60% as large for this functional as it is for CHF(1) and MCHF. By this same logic, the negative $d_{ij}^+$ eigenvalues for CHF(\(\zeta\)) and MCHF will also tend to move electrons out of the core, since in these cases $d_{ij}^+$ is most negative when $n_i = \frac{1}{2} = n_j$. For the CH(\(\zeta\)) and SIC-CH(\(\zeta\)) functionals there is still some incentive in this direction, but perhaps not so strong because for these functionals $d_{ij}^+$ is minimized when $n_i n_j = \lambda_{\text{min}}$ for $\lambda_{\text{min}} \approx 0.25$ (see Table 5.13).

To quantify the energy lowering attributable to negative eigenvalues of $\hat{D}_2^{\alpha\alpha}$, we have decomposed each optimized electronic energy into a one-electron part, $2 \, E_1^\alpha = 2 \, \text{tr}(\hat{h} \, \hat{D}_1^\alpha)$, and a two-electron part, $2(E_2^{2\alpha} + E_2^{3\beta})$. $E_2^{2\alpha}$ is further decomposed into
natural geminal pair energies, as described above. The resulting energy decompositions are given in Table 5.14. Also tabulated is $|\mathcal{R}|$, where $\mathcal{R}$ is the ratio of negative-to positive-eigenvalue contributions to $E_2^{\alpha\alpha}$. Except for the near-RHF functionals CH(4/3) and SIC-CH(4/3), this ratio ranges from 0.4–0.8, meaning that negative eigenvalues of the 2-RDM annihilate 40–80% of the parallel-spin electron repulsion energy arising from the positive eigenvalues. In absolute terms, this anomalous energy lowering is huge—frequently a hartree or more. Clearly, the negative eigenvalues of $\bar{D}_2^{\alpha\alpha}$ are crucial to determining the optimized electronic energy.

As we have indicated, numerous negative 2-RDM eigenvalues are probably to be expected in reconstructive DMFT, unless specific steps are taken to avoid them. The SIC variant of CH(\zeta) takes an important step in this direction by removing the $D_{ii;ii}^{\alpha\alpha}$ matrix elements in the NO direct product basis. This annihilates most of the self-interaction associated with CH(\zeta),\textsuperscript{124} though some residual self-interaction remains because SIC-CH(\zeta) does not correspond to an antisymmetric 2-RDM and therefore the exchange-correlation energy does not vanish for a one-electron system.

Although the functionals SIC-CH(\zeta) are not antisymmetric, one can judge that the most significant self-interactions have been eliminated, insofar as $|D_{ii;ii}^{\alpha\alpha}|$ tends to be larger than $|D_{ij;ij}^{\alpha\alpha} + D_{ij;ij}^{\alpha\alpha}|$. (Both quantities would be zero, were the reconstruction antisymmetric.) Indeed, we have shown that annihilation of the $D_{ii;ii}^{\alpha\alpha}$ is essential in order to obtain a physically reasonable distribution of natural occupation numbers.

What is more, the remaining negative eigenvalues $d_{ij}^{\alpha\alpha}$, which have the same analytic form for both CH(\zeta) and SIC-CH(\zeta), tend to be closer to zero for the latter functional, and their contribution to the energy is also much reduced in the case of SIC-CH(\zeta).
| Functional | $2E_i^{\alpha}$ | Contributions to $2E_2^{\alpha}$ | $2E_2^{\alpha}$ | $E$ | $|\Re|$ |
|------------|----------------|-------------------------------|-----------------|-----|------|
|            |                | From $\{d_i\}$ | From $\{d_{ij}\}$ | From $\{d_{ij}\}$ |     |
| RHF        | $-19.061601$   | $0.000000$        | $0.000000$       | $0.911416$       | $3.578232$ | $-14.571953$ | $0.00$ |
| CH(1)      | $-18.668248$   | $-0.157949$       | $-1.755649$      | $2.481845$       | $3.429023$ | $-14.670978$ | $0.77$ |
| SIC-CH(1)  | $-18.997737$   | $0.000000$        | $-0.784286$      | $1.606510$       | $3.559236$ | $-14.616277$ | $0.49$ |
| CH(4/3)    | $-19.043014$   | $-0.004414$       | $-0.076264$      | $0.905666$       | $3.571371$ | $-14.573330$ | $0.09$ |
| SIC-CH(4/3)| $-19.054487$   | $0.000000$        | $-0.044568$      | $0.950235$       | $3.576077$ | $-14.572744$ | $0.05$ |
| CHF(1)     | $-18.611223$   | $-0.174828$       | $-0.724415$      | $1.532446$       | $3.393895$ | $-14.584124$ | $0.59$ |
| CHF(1.12)  | $-18.518993$   | $-0.216928$       | $-1.007801$      | $1.767850$       | $3.361121$ | $-14.614750$ | $0.69$ |
| MCHF       | $-18.633732$   | $-0.168191$       | $-1.476371$      | $2.228729$       | $3.411587$ | $-14.637979$ | $0.74$ |
| RHF        | $-12.465925$   | $0.000000$        | $0.000000$       | $0.669819$       | $2.814929$ | $-8.981177$  | $0.00$ |
| CH(1)      | $-12.122610$   | $-0.118276$       | $-1.025775$      | $1.601107$       | $2.657043$ | $-9.008510$  | $0.71$ |
| SIC-CH(1)  | $-12.416512$   | $0.000000$        | $-0.509962$      | $1.114852$       | $2.811963$ | $-8.999659$  | $0.46$ |
| CH(4/3)    | $-12.455077$   | $-0.002228$       | $-0.046952$      | $0.712088$       | $2.813598$ | $-8.978517$  | $0.07$ |
| SIC-CH(4/3)| $-12.460630$   | $0.000000$        | $-0.029530$      | $0.695060$       | $2.814948$ | $-8.980153$  | $0.04$ |
| CHF(1)     | $-12.069959$   | $-0.115591$       | $-0.248405$      | $0.869255$       | $2.605495$ | $-8.959206$  | $0.42$ |
| CHF(1.12)  | $-11.941571$   | $-0.159687$       | $-0.405064$      | $1.000109$       | $2.548057$ | $-8.958156$  | $0.56$ |
| MCHF       | $-12.308129$   | $-0.157320$       | $-1.051491$      | $1.680444$       | $2.774178$ | $-9.062319$  | $0.72$ |

Table 5.14: Decomposition of DMFT electronic energies for Be/6-311G(2df) (upper data set) and LiH($R = 1.5953$ Å)/6-311+G(d,3p) (lower data set). All values are in hartrees.
Unfortunately, the self-interaction correction utilized by the SIC-CH(\') reconstruction is representation-dependent, and consequently the corresponding energy functional is not invariant to unitary transformations among orbitals with degenerate occupancies.

Consequently, the SIC-CH(\') energy will generally change if, for example, real linear combinations are substituted for primitive spherical harmonics in any orbitals with nonzero angular momentum. In contrast to this ad hoc procedure for removing self-interaction, any approximate reconstruction functional that preserves the full antisymmetry of the 2-RDM (not just the symmetrized block structure) will exhibit the extremely sparse block structure. Moreover, generally, antisymmetrically re-

antisymmetry implies positivity, at the theorem level, only for 2-RDMs that

is indeed the case for the HF reconstruction, whereas the populations equal that of the same

Consider an antisymmetric ansatz for \( D_{ij} \) that consists of the same block structure as the functionals studied here. Then the 1 \( D_{ij} \) blocks are zero, and each 2 \( D_{ij} \) block contains only one independent element. The eigenvalues of an antisymmetric \( B_{ij} \) are zero and \( 2D_{ij} \), hence antisymmetry cures the positivity problem, provided that the geminal populations \( 2D_{ij} \) are themselves positive. This is indeed the case for the HF reconstruction, wherein the populations equal

Antisymmetry implies positivity, at the theorem level, only for 2-RDMs that exhibit the extremely sparse block structure. Moreover, generally, antisymmetrically re-

antisymmetry implies positivity, at the theorem level, only for 2-RDMs that

is indeed the case for the HF reconstruction, whereas the populations equal that of the same

Consider an antisymmetric ansatz for \( D_{ij} \) that consists of the same block structure as the functionals studied here. Then the 1 \( D_{ij} \) blocks are zero, and each 2 \( D_{ij} \) block contains only one independent element. The eigenvalues of an antisymmetric \( B_{ij} \) are zero and \( 2D_{ij} \), hence antisymmetry cures the positivity problem, provided that the geminal populations \( 2D_{ij} \) are themselves positive. This is indeed the case for the HF reconstruction, wherein the populations equal

Antisymmetry implies positivity, at the theorem level, only for 2-RDMs that exhibit the extremely sparse block structure. Moreover, generally, antisymmetrically re-

antisymmetry implies positivity, at the theorem level, only for 2-RDMs that
ity provides a rigorous bound on the off-diagonal elements, namely

$$|D_{ij;kl}^{\alpha\alpha}|^2 \leq D_{ij;ij}^{\alpha\alpha} D_{kl;kl}^{\alpha\alpha}, \quad (5.78)$$

though in most cases the coherences $D_{ij;kl}^{\alpha\alpha}$ are significantly smaller than this theoretical upper bound. As such, these elements represent small perturbations to the spectrum determined by the $B_{ij}$ blocks. Secondly, let us rewrite the partial trace relation, Eq. (5.25), as

$$D_{ij}^{\alpha} = \frac{2}{N-1} \sum_k \left[ D_{ik;jk}^{\alpha\alpha} + D_{ik;jk}^{\alpha\beta} \right], \quad (5.79)$$

from which it is clear that matrix elements of $\hat{D}_2^{\alpha\alpha}$ outside of the $B_{ij}$ blocks contribute only to off-diagonal elements of the 1-RDM. If $\hat{D}_2^{\alpha\alpha}$ is reconstructed in the NO product basis, these off-diagonal contributions must sum to zero. Although the off-diagonal elements of $\hat{D}_2^{\alpha\alpha}$ can be positive or negative, one hopes that this sum rule leads them to be small in the NO product basis. The off-diagonal elements of $\hat{D}_2^{\alpha\alpha}$ are exactly zero in the NO product basis, if the natural geminals are $2 \times 2$ determinants of NOs. For Bunge’s Be wavefunction,\textsuperscript{164} which we have used throughout this work to test a number of approximate reconstructions, it was pointed out in Subsection 4.3.1 that the 2-RDM has far fewer large elements in the NSO basis than it does in the basis of HF molecular orbitals. A comprehensive examination of the geminal structure of correlated 2-RDMs for small molecules is sorely needed, in order to determine whether the determinantal approximation for the natural spin geminals is reasonable.

In view of these remarks, the development of antisymmetric reconstruction functionals should be a priority. Otherwise, it is difficult to see how the negative spectrum of $\hat{D}_2^{\alpha\alpha}$ can be eliminated or mitigated, within an ansatz bearing any similarity to the ones examined here. (We mention in this regard that an iterative procedure\textsuperscript{196} for
removing negative eigenvalues of $\hat{D}_2$, $\hat{Q}_2$, and $\hat{G}_2$, by adjusting $\hat{D}_2$ while preserving the underlying 1-RDM, is ineffective in the present case, as there are too few nonzero 2-RDM elements.) In the considerations leading to the CH(1) and CHF(1) functionals (Subsection 5.2.2), antisymmetry was abandoned in order to obtain an alternative to the RHF functional that still satisfies the partial trace relation in Eq. (5.79). Based on experience with sum rules for the exchange-correlation hole in DFT, this partial trace relation is expected to be an extremely powerful $N$-representability requirement. To recover antisymmetry in DMFT, yet still maintain this partial trace relation, it is probably necessary to move beyond one-term tensor product approximations for the exchange-correlation part of $\hat{D}_2$. For more complicated reconstructions, the partial trace relation is not so easy to enforce, though we speculate that perhaps some variant of Perdew’s real-space cutoff procedure$^{136-138}$ might be used in this capacity, much as it is used in DFT to enforce sum rules on the exchange-correlation hole.*

5.6 Application to strongly correlated systems

In the previous section we discovered that SIC-CH(1) provides excellent potential energy curves for the LiH molecule in a variety of basis sets. Given a sufficiently flexible orbital basis, the SIC-CH(1) potential for LiH is qualitatively correct across the entire domain of internuclear distances, in contrast to RHF potentials, which are qualitatively correct near the equilibrium geometry but rise much too steeply at large bond lengths (cf. Fig. 5.1). This is a well-known deficiency of the HF approximation, and to a lesser extent the KS approximation as well; at large internuclear distances

*Use of a cutoff procedure to enforce certain sum rules is the essence of the generalized gradient approximation (GGA) in DFT, and is the single most important reason why GGAs are superior to traditional gradient expansions.$^{138,139}$
these methods always put too much weight into ionic electron configurations, leading to unrealistically high energies. Other SIC-CH(\(\zeta\)) functionals—in particular, SIC-CH(4/3)—are also qualitatively correct across the entire LiH potential surface, but SIC-CH(1) stands apart because it also recovers a substantial amount of dynamical correlation, providing energies that are very close to FCI results near the equilibrium geometry.

In this section we demonstrate that the SIC-CH(1) functional can also recover significant non-dynamical correlation. Any method that can accomplish this reliably would be an important development, especially with regard to transition-state chemistry, bond-breaking, or other phenomena in which small changes in nuclear geometry are associated with qualitative changes in electronic structure, and therefore near-degenerate electronic configurations are anticipated. Granted, our analysis has demonstrated unequivocally that SIC-CH(1) succeeds at the expense of important \(N\)-representability conditions—positivity of the 2-RDM not least among them—but in order to understand and correct these deficiencies it is necessary to collect more data. Thus, we have examined SIC-CH(1) potential energy curves for several other molecules, in order to evaluate the performance of SIC-CH(1) in describing bond-breaking phenomena. Here we present results for two infamously difficult molecules: \(\text{C}_2\) and FH. These are among the most problematic closed-shell species for otherwise reliable single-reference methods, owing to the onset of important multiconfigurational effects (non-dynamical correlation) at relatively modest bond lengths.

We first consider FH, a prototypical example both of RHF configurational degeneracy and also HOMO-LUMO near-degeneracy. (Unless otherwise stated, the degeneracies mentioned in this section are accidental, and should be understood as near
degeneracies.) Within the 6-31+G** basis, the true RHF ground-state configuration is \((1\sigma)^2(2\sigma)^2(3\sigma)^2(1\pi)^4\), but a \((1\sigma)^2(2\sigma)^2(1\pi)^4(3\sigma)^2\) configuration lies only \((0.155 \pm 0.003)\mu E_h\) higher in energy at the experimental bond length, \(R_e = 0.9171 a_0\). The quoted uncertainty in the energy equals the energy difference between the final two SCF cycles, and is pointed out merely to indicate that it is possible to distinguish configurations that differ only by fractions of \(\mu E_h\).\(^*\) At \(R = 3.5 R_e\), the energy difference between the two configurations is still only \((2.326 \pm 0.002)\mu E_h\). Thus the RHF potential energy curves are essentially indistinguishable, although post-HF calculations built upon these states can yield very different results, as illustrated by the MP2 potentials in Fig. 5.28. This demonstrates an important fact: RHF potential curves that appear qualitatively correct can mask instabilities that wreak havoc with post-HF electronic structure methods, especially those based on perturbation theory.

Also shown in Fig. 5.28 are the HOMO and LUMO energy levels, which approach degeneracy as the F–H bond is stretched. This is a common feature of bond-breaking potentials at sufficiently large \(R\), and wreaks great havoc with post-HF \textit{ab initio} methods. Some \textit{ab initio} potential energy curves are depicted in Fig. 5.29; the turnover in the MP2 and CCSD(T) potentials is characteristic of the near-degeneracy of \(\epsilon_{\text{LUMO}}\) and \(\epsilon_{\text{HOMO}}\). The trends exhibited by the potentials in Fig. 5.29 continue unabated as \(R \to \infty\) (see, for example, Ref. 198). As indicated above, we tend to think of FH as being “strongly correlated” because this turnover occurs early in the potential curve, relative to many other systems. However, it is reasonably common to find that \(|\epsilon_{\text{LUMO}} - \epsilon_{\text{HOMO}}| \to 0\) as \(R \to \infty\), so the problem is endemic. Note, however, that the density matrix functional SIC-CH(1) yields a potential energy curve that is

\(^*\text{GAUSSIAN98 converges to the correct ground state only when the keyword SCF = Tight is used.}\)
well-behaved (and also quite close to the CISD result) at all $R$. This is the case even though HF orbitals and occupancies were used as the initial guess for the 1-RDM.

Next we consider C$_2$, one of the most notorious cases of strong configurational mixing, which is prevalent even at the equilibrium geometry. The ground state ($X^1\Sigma_g^+$) has the valence configuration $(2\sigma_g)^2(2\sigma_u)^2(1\pi_u)^4$, although this configuration has a coefficient of only 0.84 in an accurate multireference CISD (MR-CISD) wave-function,$^{199}$ while the doubly-excited $(2\sigma_u)^2 \rightarrow (3\sigma_g)^2$ determinant has a coefficient of 0.34. Several other configurations have coefficients exceeding 0.1.

Owing to near-degeneracy of the HOMO–1, HOMO, and LUMO $(2\sigma_u, 1\pi_u$, and $3\sigma_g$, respectively), there are a plethora of low-lying transitions in C$_2$. (Experimental data for this molecule are reviewed in Ref. 200.) The lowest of these ($X^1\Sigma_g^+ \rightarrow a^3\Pi_u$) is in the infrared, having term value (adiabatic excitation energy) $T_e = 3.27 \text{ mE}_h =$
718 cm$^{-1}$ = 0.0891 eV. The $a$ state arises from a one-electron HOMO $\rightarrow$ LUMO excitation and thus has the configuration $(2\sigma_g)^2(2\sigma_u)^2(1\pi_u)^3(3\sigma_g)^1$; this configuration also gives rise to an $A^1\Pi_u$ term ($T_e = 1.040$ eV). A two-electron HOMO $\rightarrow$ LUMO excitation to the $(2\sigma_g)^2(2\sigma_u)^2(1\pi_u)^2(3\sigma_g)^2$ configuration produces states $b^3\Sigma_g^-$ ($T_e = 0.798$ eV), $B^1\Delta_g$ ($T_e = 1.498$ eV), and $B'^1\Sigma_g^+$ ($T_e = 1.910$ eV). Lastly, excitation into the $(2\sigma_g)^2(2\sigma_u)(1\pi_u)^4(3\sigma_g)^1$ valence configuration [that is, (HOMO $- 1$) $\rightarrow$ LUMO] gives rise to a $c^3\Sigma_u^+$ state with $T_e = 1.131$ eV, along with a $D^1\Sigma_u^+$ state that is much higher in energy. This completes the list of electronic states with $T_e < 2$ eV.

Each of the aforementioned electronic configurations is nearly degenerate at the HF level, and their relative energies change order as a function of internuclear distance. Furthermore, the $X^1\Sigma_g^+$ and $B'^1\Sigma_g^+$ states exhibit an avoided crossing at $R \approx 1.5 R_e$. As such, naïve application of a correlated method on top of the HF reference is a recipe for disaster. The nature of the disaster is illustrated in Fig. 5.30, where we
Figure 5.30: 6-31G* potential curves for the C\textsubscript{2} molecule. The MRCl potential curves are analytic fits\cite{201} for the CISD and SIC-CH(1) potentials, single point calculations have been carried out at the indicated values of \(R/R_e\). All potentials are shifted to a common asymptote.

show the results of a CISD calculation with a RHF/6-31G* reference. The computed HF potential curve (not shown) is deceptively reasonable, but it is apparent from the discontinuity in the CISD potential that the wrong set of orbitals has been obtained in the vicinity of \(R = 1.5R_e\). For \(R < 1.5R_e\) the RHF valence electron configuration is \((2\sigma_g)^2(2\sigma_u)^2(1\pi_u)^4\) but for \(R > 1.5R_e\) it is \((2\sigma_g)^2(2\sigma_u)^2(1\pi_u)^2(3\sigma_g)^2\).

Despite the problems with the HF reference, we continue to use the HF orbitals and occupancies to initialize the DMFT calculations. We wish to compare the SIC-CH(1) potential from DMFT with a reliable calculation, and for C\textsubscript{2} this means a multireference calculation. For our purposes, the analytic potentials published by Boggio-Pasqua and co-workers,\cite{201} which are fit to MR-CISD/cc-p5VZ single-point calculations as well as some empirical data for the dispersion coefficients, may be
considered essentially exact. In Fig. 5.30 we have plotted the multireference potentials for both the ground state and the $B'$ $\Sigma^+_g$ state with which it shares an avoided crossing. Also plotted are converged SIC-CH(1) data points, starting from the RHF 1-RDM. Because the multireference calculations use a much larger basis (polarized quintuple zeta) than the one employed in our calculations, all potentials in Fig. 5.30 are shifted to a common asymptote.

Following the shift to a common asymptote, the SIC-CH(1) potential is a remarkable match to the MR-CISD potential for the $B'$ state, the lowest-energy excited state that has the same symmetry as the ground state. Considering that both the equilibrium bond length and the well depth of the $B'$ state are reproduced by SIC-CH(1), and that these parameters differ considerably between the $X$ and $B'$ states, we feel safe in concluding that SIC-CH(1) has reproduced the excited state potential, rather than providing a poor description of the ground state. However, this assignment remains tentative until such time as we are able to produce a converged SIC-CH(1) potential for the $X$ state. The next logical step toward this end is to optimize the SIC-CH(1) energy starting from an excited state of the RHF determinant, corresponding to a HOMO $\rightarrow$ LUMO $[(1\pi_u)^4 \rightarrow (1\pi_u)^2(3\sigma_g)^2]$ excitation. This excitation leaves two singly-occupied, singlet-coupled $\pi$ orbitals. Hence two occupation numbers must be changed in order to preserve $^1\Sigma$ symmetry, and it may be necessary to use a spin-unrestricted formalism (separate occupation numbers for $\alpha$- and $\beta$-spin electrons) in

---

*For the $X$ and $B'$ states, which both have $^1\Sigma^+_g$ symmetry and exhibit an avoided crossing at $R \approx 1.5 R_e$, the adiabatic potentials plotted in Fig. 5.30 were obtained$^{201}$ by diagonalizing a $2 \times 2$ diabatic Hamiltonian at each value of $R$. The diabatic Hamiltonian has a parametrized form that is fit to the aforementioned MR-CISD single point calculations. The parameter $(r^2) = 3.891815$ $a_0$ for carbon atom, obtained from relativistic Dirac-Fock calculations,$^{202}$ is required; all other parameters relevant to the fit may be found in Ref. 201.*
order to provide a reasonable initial guess. These calculations will be carried out in
the near future, along with calculations for other strongly-correlated systems including
\( \text{Be}_2 \), \( \text{N}_2 \), and symmetric double dissociation of \( \text{H}_2\text{O} \).

5.7 A novel, generalized Hartree-Fock functional

Despite the fact that SIC-CH(1) appears to be highly accurate, we know that its suc-
cess is attributable entirely to a cancellation of errors (albeit an apparently robust one). We are presently trying to understand why this cancellation works as well it
does, but at the same time we would like to develop functionals that do not exhibit the
same egregious \( N \)-representability violations as the ones considered so far. In view of
our comments in Subsection 5.5.2, we consider the 2-RDM antisymmetry requirement
to be of paramount importance, though to date the HF reconstruction is the only
reconstruction functional consistent with this constraint. In this section we propose
our own ansatz for reconstruction of \( \hat{D}_2 \) in terms of the natural occupation numbers,
which is based upon the satisfaction of several exact constraints including positivity,
antisymmetry, proper normalization, and Levy’s homogeneous scaling requirement.\(^\text{96}\)

Our primary motivation is to repair \( N \)-representability problems associated with \( D_2^{\alpha} \)
specifically, but the ansatz lends itself also to a reconstruction of \( D_2^{\alpha\beta} \) that general-
izes the ubiquitous Hartree-product reconstruction. The new functionals for \( D_2^{\alpha} \) and
\( D_2^{\alpha\beta} \) each contain one parameter, and although these functionals are derived from the
aforementioned \( N \)-representability considerations alone, they turn out to be general-
izations of the HF reconstruction, recovering the HF result for a particular choice of
the parameters. This is fortuitous, since it provides a clue as to an appropriate range
for the adjustable parameters.

5.7.1 Derivation

Our goal is to provide formulas for the matrix elements $D_{ij;kl}^{\alpha\alpha}$ and $D_{ij;kl}^{\alpha\beta}$ in the NO direct product basis, under the assumption that these matrix elements are functions of the natural occupation numbers only. Given this assumption, Goedecker and Umrigar\textsuperscript{125} have used an example of two non-interacting subsystems in an attempt to reason that the matrix elements $D_{ij;kl}^{\alpha\alpha}$ and $D_{ij;kl}^{\alpha\beta}$ can depend only on the occupation numbers $n_i$, $n_j$, $n_k$, and $n_l$, lest the 2-RDM for one subsystem depend upon the occupation numbers for a separate subsystem. This analysis is flawed, however, as it attempts to attribute to $\tilde{D}_2$ a separability that is associated instead with $\tilde{\Delta}_2$. For non-interacting subsystems, it is true that $\Delta_{ij;kl}^{\sigma\mu} = 0$ unless $i$, $j$, $k$, and $l$ each refer to orbitals centered on the same subsystem, but the same is not true of the 2-RDM, which contains an exchange part $\hat{D}_1 \wedge \hat{D}_1$ that couples even non-interacting subsystems. Nevertheless, one may argue that it is a reasonable approximation to let $D_{ij;kl}^{\sigma\mu}$ be determined by $n_i$, $n_j$, $n_k$, and $n_l$ alone, since this is indeed the case in the HF approximation. Thus we set

$$D_{ij;kl}^{\alpha\alpha} = \tilde{f}_{aa}(n_i, n_j, n_k, n_l) \quad (5.80a)$$

$$D_{ij;kl}^{\alpha\beta} = f_{\alpha\beta}(n_i, n_j, n_k, n_l). \quad (5.80b)$$

The function $\tilde{f}_{aa}$ is defined by antisymmetrizing another function $f_{aa}$,

$$\tilde{f}_{aa}(n_i, n_j, n_k, n_l) \overset{\text{def}}{=} f_{aa}(n_i, n_j, n_k, n_l) - f_{aa}(n_j, n_i, n_k, n_l) - f_{aa}(n_i, n_j, n_l, n_k) + f_{aa}(n_j, n_i, n_l, n_k), \quad (5.81)$$
which insures that $\hat{D}_2^{\alpha\alpha}$ is antisymmetric. The functions $f_{\alpha\alpha}$ and $f_{\alpha\beta}$ will be chosen in order to incorporate certain other properties of the 2-RDM. Self-adjointness, for example, requires that

$$f_{\sigma\mu}(n_1, n_2, n_3, n_4) = f_{\sigma\mu}^*(n_3, n_4, n_1, n_2).$$

(5.82)

Further restrictions on the $f_{\sigma\mu}$ are imposed by a homogeneous scaling requirement for the energy. Levy\textsuperscript{96} has shown that the unknown functional $V_{ee}[\hat{D}_1]$ for the electron repulsion energy obeys the scaling relationship

$$V_{ee}[\lambda^3 D_1(\lambda r, \xi; \lambda r', \xi')] = \lambda V_{ee}[D_1(r, \xi; r', \xi')] ,$$

(5.83)

for scaling parameters $\lambda \geq 1$. Actually, what Levy proved was that the DMFT correlation energy functional, rather than $V_{ee}$, satisfies this homogeneity constraint, where “correlation energy” is used in the cumulant sense of the term (see Subsection 3.3.1). Within our reconstructive approach, the correlation energy functional is

$$E_c[\hat{D}_1] = \text{tr} \left\{ \hat{H}_{12} \left( \hat{D}_2[\hat{D}_1] - \hat{D}_1 \wedge \hat{D}_1 \right) \right\} .$$

(5.84)

All components of $E = \text{tr}(\hat{H}_{12} \hat{D}_2)$ except $V_{ee}$ are known functionals of $\hat{D}_1$, and each is readily shown to satisfy a homogeneous scaling relationship analogous to that asserted for $V_{ee}$ in Eq. (5.83). The exchange energy $E_x[\hat{D}_1] = \text{tr}\{\hat{H}_{12}(\hat{D}_1 \wedge \hat{D}_1)\}$ exhibits the same homogeneity property, hence Eq. (5.83) is equivalent to Levy’s original condition. Generally

$$V_{ee} = \int d\mathbf{r}_1 d\mathbf{r}_2 \frac{[D_2^{\alpha\alpha} + 2 D_2^{\alpha\beta} + D_2^{\beta\beta}]}{|\mathbf{r}_1 - \mathbf{r}_2|} \left( \mathbf{r}_1, \mathbf{r}_2; \mathbf{r}_1, \mathbf{r}_2 \right),$$

(5.85)

though $\hat{D}_2^{\alpha\alpha} \equiv \hat{D}_2^{\beta\beta}$ for spin-compensated states (assumed throughout this chapter). We next show that Eq. (5.83) implies a homogeneity constraint on the functions $f_{\sigma\mu}$.\vspace{1cm}
This derivation is not difficult, but will be dispatched somewhat pedantically because there have been some published misstatements\textsuperscript{124} concerning the nature of the homogeneity constraint in DMFT.

First of all, consider what is meant by the scaled 1-RDM on the left side of Eq. (5.83). If one defines
\begin{equation}
D^\sigma_{1,[\lambda]}(r; r') \overset{\text{def}}{=} \lambda^3 D_1^\sigma(\lambda r; \lambda r')
\end{equation}
then one may write the homogeneous scaling requirement as $V_{ee}[D_{1,[\lambda]}] = \lambda V_{ee}[D_1]$. Scaling the occupation numbers by $\lambda^3$ in Eq. (5.86) preserves the trace of the 1-RDM when coordinate space is scaled by $\lambda$, $r \rightarrow \lambda r$, In other words, $\text{tr} D^\sigma_{1,[\lambda]} = \text{tr} D^\sigma_1$.

Let us evaluate $V_{ee}[D_{1,[\lambda]}]$, using Eq. (5.85). By expanding the kernels $D^\sigma_{1,[\lambda]}$ in the NO basis, changing integration variables from $r_i$ to $\lambda r_i$ [note that $d(\lambda r_i) = \lambda^3 dr_i = (\lambda dx_i)(\lambda dy_i)(\lambda dz_i)$], and using the fact that $|\lambda r_1 - \lambda r_2| = \lambda |r_1 - r_2|$, one ultimately obtains
\begin{equation}
V_{ee}[D_{1,[\lambda]}] = \frac{1}{\lambda^{5}} \sum_{ijkl} D_{ij\lambda;kl}^\mu (\lambda^3 n_1, \lambda^3 n_2, \ldots) + 2 D_{ij\lambda;kl}^{\mu\beta} (\lambda^3 n_1, \lambda^3 n_2, \ldots) + D_{ij\lambda;kl}^{\beta\beta} (\lambda^3 n_1, \lambda^3 n_2, \ldots) \int dr_1 dr_2 \frac{\chi_i^*(r_1) \chi_j^*(r_2) \chi_k(r_1) \chi_l(r_2)}{|r_1 - r_2|}.
\end{equation}
As indicated by the notation, we have assumed that the matrix elements $D^\mu_{ij;kl}$ are functions of the natural occupation numbers only. Modulo this assumption, Eq. (5.88) is an equivalent form of Eq. (5.85).

Under the same assumption, one obtains for $V_{ee}[D_1]$ the same expression as in Eq. (5.88), except that the overall factor of $\lambda^{-5}$ is absent, and the $D^\mu_{ij;kl}$ are evaluated
at the unscaled occupation numbers \( n_p \) rather than at the scaled occupation numbers \( \lambda^3 n_p \) as in Eq. (5.88). As a result, Levy’s homogeneous scaling requirement demands that

\[
D_{ij;kl}^{\alpha\alpha}(\lambda^3 n_1, \lambda^3 n_2, \ldots) + 2D_{ij;kl}^{\alpha\beta}(\lambda^3 n_1, \lambda^3 n_2, \ldots) + D_{ij;kl}^{\beta\beta}(\lambda^3 n_1, \lambda^3 n_2, \ldots)
= \lambda^6 \left[ D_{ij;kl}^{\alpha\alpha}(n_1, n_2, \ldots) + 2D_{ij;kl}^{\alpha\beta}(n_1, n_2, \ldots) + D_{ij;kl}^{\beta\beta}(n_1, n_2, \ldots) \right]
\]  

for any model in which the 2-RDM elements are expressed as functions of the natural occupation numbers only. For a two-electron singlet state (such as the ground state of \( \text{H}_2 \)), \( \hat{D}_2^{\alpha\alpha} \equiv 0 \) and \( \hat{D}_2^{\beta\beta} \equiv 0 \), and it follows that we must be able to separate the \( D_{ij;kl}^{\alpha\alpha} \), \( D_{ij;kl}^{\alpha\beta} \), and \( D_{ij;kl}^{\beta\beta} \) elements in Eq. (5.89). Individually, each of these matrix elements is required to be homogeneous of degree 2,

\[
D_{ij;kl}^{\sigma\rho}(\lambda n_1, \lambda n_2, \ldots) = \lambda^2 D_{ij;kl}^{\sigma\rho}(n_1, n_2, \ldots)
\]

Observe that the HF reconstruction satisfies Eq. (5.90) because it contains only terms like \( n_i n_j \). None of the other reconstructions introduced in Section 5.2 satisfy this constraint. For example, the CH(\( \zeta \)) and SIC-CH(\( \zeta \)) functionals contain some terms like \( n_i n_j \), which are indeed homogeneous of degree 2, but also terms like \( (n_i n_j)^{\zeta/2} \) that are homogeneous of degree \( \zeta \). This analysis refutes the claim made by Goedecker and Umrigar\(^\text{124}\) that their functional [SIC-CH(1)] satisfies Levy’s homogeneity criterion. The CHF(\( \zeta \)) and MCHF reconstructions involve inhomogeneous functions of the \( n_i \), and thus do not satisfy Levy’s criterion either.

In the case of the reconstruction proposed in Eq. (5.80), the homogeneity requirement implies that both \( f_{\alpha\alpha} \) and \( f_{\alpha\beta} \) must be homogeneous of degree 2. This is most easily satisfied if we take these functions to be homogeneous polynomials in the
natural occupation numbers,
\[
    f_{\sigma'}(n_1, n_2, n_3, n_4) = \sum_\kappa C_\kappa \left( n_1^{x_\kappa} n_2^{y_\kappa} n_3^{z_\kappa} n_4^{2-x_\kappa-y_\kappa-z_\kappa} + n_3^{x_\kappa} n_4^{y_\kappa} n_1^{z_\kappa} n_2^{2-x_\kappa-y_\kappa-z_\kappa} \right),
\]
(5.91)
in which the exponents \(x_\kappa, y_\kappa,\) and \(z_\kappa,\) and coefficients \(C_\kappa,\) are adjustable parameters.

The two terms within the parentheses in Eq. (5.91) insure that the reconstruction is self-adjoint and, in the case of \(D_2^{\alpha\beta}\), that it does not artificially distinguish between \(\alpha\) and \(\beta\) electrons.

Here we will study an even simpler model in which we neglect all off-diagonal elements of \(D_2^{\alpha\beta}\), and also all off-diagonal elements of \(D_2^{\alpha\alpha}\) except for those required by antisymmetry. In the latter case, this amounts to neglecting off-diagonal elements in a basis of \(2 \times 2\) Slater determinants of NOs. In other words, we assume that \(\hat{D}_2^{\alpha\alpha}\) has the \(B_{ij}\) block structure introduced previously, whose validity—especially in the NO product basis—has already been discussed.

Based on these considerations, we choose a definite form for the functions \(f_{\sigma\mu}\), defined in terms of parameters \(x\) and \(y\):
\[
    f_{\sigma \alpha}(n_i, n_j, n_k, n_l) \overset{\text{def}}{=} \frac{1}{4} c_{\alpha \alpha} n_i^x n_j^{2-x} \delta_{ik} \delta_{jl},
\]
(5.92a)
\[
    f_{\sigma \beta}(n_i, n_j, n_k, n_l) \overset{\text{def}}{=} \frac{1}{2} c_{\alpha \beta} \delta_{ik} \delta_{jl} \left( n_i^y n_j^{2-y} + n_j^y n_i^{2-y} \right).
\]
(5.92b)
The functions \(c_{\sigma \mu} = c_{\sigma \mu}(n_1, n_2, \ldots)\) have been introduced in order to insure that each component of \(\hat{D}_2\) is separately and properly normalized; we will determine these functions shortly. Note that \(f_{\sigma \beta}\) must include both of the terms given above, because to omit either would artificially distinguish \(\alpha\) from \(\beta\) spin.

Our choice of \(f_{\alpha \alpha}\) and \(f_{\alpha \beta}\) corresponds to reconstructed 2-RDM elements
\[
    D_{i j ; k l}^{\alpha \alpha} = \frac{1}{2} c_{\alpha \alpha} \left( n_i^x n_j^{2-x} + n_j^x n_i^{2-x} \right) (\mathbf{I} \wedge \mathbf{I})_{i j ; k l}
\]
(5.93)
and

\[ D_{ij;kl}^{\alpha\beta} = \frac{1}{2} c_{\alpha\beta} \left( n_i^u n_j^2 + n_j^u n_i^2 - y \right) (I \otimes I)_{ij;kl}, \tag{5.94} \]

which we regard as the simplest occupancy-based reconstruction functional that incorporates both antisymmetry and homogeneity. Except in the case \( x = y = 1 \), or for an idempotent 1-RDM, this functional does not satisfy the partial trace relation, although \( \frac{1}{2} \hat{D}_2 \) is at least diagonal. However, because this functional satisfies the trace conditions for both \( \hat{D}_2^{\alpha\alpha} \) and \( \hat{D}_2^{\alpha\beta} \), there is hope that violations of the contraction condition will be small.

From Eqs. (5.93) and (5.94) it is clear that the new reconstruction is equivalent to the ansatz \((\hat{D}_1)^{x} \wedge (\hat{D}_1)^{2-x}\) for \(\hat{D}_2^{\alpha\alpha}\) and \((\hat{D}_1)^{y} \otimes (\hat{D}_1)^{2-y}\) for \(\hat{D}_2^{\alpha\beta}\). As such, this functional generalizes the HF reconstruction, and we therefore refer to the functional in Eqs. (5.93) and (5.94) as GHF\((x,y)\). The functional GHF\((1,1)\) is equivalent to RHF, but more generally GHF\((x,y)\) reduces to RHF form in the limit of an idempotent 1-RDM, hence this functional satisfies the criterion of zero cumulant correlation energy in the idempotent limit.

Next we determine the normalizing functions \(c_{\alpha\alpha}\) and \(c_{\alpha\beta}\). From Eqs. (5.93) and (5.94) one obtains

\[ \text{tr} \hat{D}_2^{\alpha\alpha} = \frac{1}{8} c_{\alpha\alpha} \left[ (\text{tr} \hat{D}_1^x)(\text{tr} \hat{D}_1^{2-x}) - 2 \text{tr} \hat{D}_1^2 \right] \tag{5.95} \]

and

\[ \text{tr} \hat{D}_2^{\alpha\beta} = \frac{1}{4} c_{\alpha\beta} (\text{tr} \hat{D}_1^y)(\text{tr} \hat{D}_1^{2-y}) \tag{5.96} \]

for the normalization of GHF\((x,y)\). [Recall that \(\text{tr} D_1^z = 2 \sum_i n_i^z = 2 \text{tr} (D_1^z).\] The correct expressions for \(\text{tr} D_2^{\alpha\alpha}\) and \(\text{tr} D_2^{\alpha\beta}\) are given in Eqs. (2.74) and (2.76) in terms of \(N_\alpha\) and \(N_\beta\), the number of electrons with \(\alpha\)- and \(\beta\)-spin. Setting \(N_\alpha = N_\beta = N/2\),
the constraint that both spin components of our ansatz be properly normalized leads us to put

\[
c_{\alpha\alpha} = \frac{N(N - 2)}{(\text{tr} \hat{D}_1^2)(\text{tr} \hat{D}_1^{2-x}) - 2 \text{tr} \hat{D}_1^2}
\]  

(5.97)

and

\[
c_{\alpha\beta} = \frac{N^2}{2(\text{tr} \hat{D}_1^y)(\text{tr} \hat{D}_1^{2-y})}.
\]  

(5.98)

It is shown below that the denominator in Eq. (5.97) does not vanish; the denominator in Eq. (5.98) is clearly nonzero for all values of \(y\).

We now show that our ansatz preserves positivity. Noting that \(c_{\alpha\beta} > 0\) for all values of \(y\) (and for all \(N\)-representable 1-RDMs), it is immediately obvious that \(\hat{D}_2^{\alpha\beta}\) in Eq. (5.94) is diagonal in the \(|\chi_i \otimes \chi_j\rangle\) basis, with eigenvalues \(\frac{1}{2} c_{\alpha\beta} (n_i^y n_j^{2-y} + n_j^y n_i^{2-y}) \geq 0\). On the other hand, \(\hat{D}_2^{2\alpha}\) in Eq. (5.93) is diagonal in the antisymmetrized product basis \(\{ |\chi_i \wedge \chi_j\rangle | i < j \}\),

\[
|\chi_i \wedge \chi_j\rangle \overset{\text{def}}{=} \frac{1}{\sqrt{2}} \left( |\chi_i \otimes \chi_j\rangle - |\chi_j \otimes \chi_i\rangle \right).
\]  

(5.99)

The eigenvalues of \(\hat{D}_2^{2\alpha}\) are \(\{ c_{\alpha\alpha} n_i^x n_j^{2-x} | i < j \}\), whence \(\hat{D}_2^{2\alpha}\) is positive semidefinite if and only if \(c_{\alpha\alpha} \geq 0\). This we now show.

Recall that \(x = 1\) is the HF value, and observe that the reconstruction in Eq. (5.93) is symmetric about \(x = 1\) in parameter space. Therefore subsequent analysis will focus on \(x \in [0, 1]\). For such \(x\), the Schwarz inequality, combined with the fact that \(\text{tr} \hat{D}_1^2 \leq N\), implies that

\[
(\text{tr} \hat{D}_1^2)(\text{tr} \hat{D}_1^{2-x}) - 2 \text{tr} \hat{D}_1^2 \geq N^2 - 2 \text{tr} \hat{D}_1^2
\]  

\[
\geq N(N - 2).
\]  

(5.100)

This establishes that \(0 < c_{\alpha\alpha} \leq 1\), therefore \(\hat{D}_2^{2\alpha} \geq 0\). Similar arguments reveal that \(0 < c_{\alpha\beta} \leq \frac{1}{2}\). These inequalities are strict when \(x \neq 1, y \neq 1\), and \(\hat{D}_1 \neq \hat{D}_1^2\). A
Figure 5.31: Normalizing factors for the GHF($x, y$) functional.

A typical case is shown in Fig. 5.31, where $c_{\alpha\alpha}(x)$ and $c_{\alpha\beta}(y)$ are plotted using natural occupation numbers from Bunge’s CI wavefunction\textsuperscript{164} for Be, which was described in Section 5.3, and also those from a FCI/6-311++G** calculations of the ground states of Be and LiH.

Lastly, in order to verify that our proposed reconstruction provides a reasonable description of the electron-pair distribution, we have calculated $I(u)$ and $X(R)$ using the natural occupation numbers from Bunge’s Be wavefunction.\textsuperscript{164} As illustrated in Fig. 5.32, when $x \gtrsim 0.8$, GHF($x, y$) provides a qualitatively correct description of $I_{\alpha\alpha}(u)$, in contrast to all of the non-antisymmetric reconstructions. [Recall from Section 5.3 that the non-antisymmetric reconstructions fail miserably for $I_{\alpha\alpha}(u)$.] For comparison, we have also plotted in Fig. 5.32 the densities $I_{\alpha\alpha}(u)$ that are obtained by \textit{ad hoc} antisymmetrization of several of the reconstruction functionals discussed previously. As we showed in Subsection 5.5.2, antisymmetrization necessarily repairs the
Figure 5.32: Same-spin intracule densities for Be, reconstructed using natural occupation numbers from Bunge’s CI wavefunction.\textsuperscript{164}

positivity problem, which is reflected in the intracule densities in Fig. 5.32. However, antisymmetrization does not preserve normalization, even when the permutational factor of $\frac{1}{4}$ is included in the antisymmetrizer. Moreover, the resulting intracule densities are still qualitatively incorrect, exhibiting a spurious maximum at small $u$. The same anomaly is observed using $\text{GHF}(x, y)$ for $x \lesssim 0.7$. In the next subsection we shall discover that there are additional reasons, based upon $N$-representability considerations, to prefer values of $x \approx 1$.

For $X_{\alpha\alpha}(R)$, the $\text{GHF}(x, y)$ reconstruction is not as quantitatively accurate as it was for $I_{\alpha\alpha}(u)$, yet the gross features of the extracule density are reproduced, as shown in Fig. 5.33. The main problem with the $\text{GHF}(x, y)$ extracule density lies in the ratio of the values of the two local maxima, one at $R = 0$ and the other at $R \approx 0.7 \ a_0$. For values of $x$ close to unity—which gave us the best agreement with
the CI same-spin intracule density—produce the worst extracule densities, shifting too much density away from $R = 0$. When the non-antisymmetric reconstructions are antisymmetrized, the resulting extracule densities are far too large. None of these density matrix functionals appears to provide a completely satisfactory description of angular correlation.

5.7.2 $Q$- and $G$-conditions for $\text{GHF}(x,y)$

By construction, the functional $\text{GHF}(x,y)$ introduced in the previous subsection satisfies the antisymmetry, positivity, normalization, and homogeneous scaling properties of the 2-RDM. In this subsection we assess the extent to which this functional violates the $Q$- and $G$-conditions. From the matrix elements of the $\text{GHF}(x,y)$ 2-RDM
[Eqs. (5.93) and (5.94)], one obtains

\[ Q_{ij;kl}^{\alpha\alpha} = \frac{1}{2} \left[ \Xi_{ij} + \frac{1}{2} F_{ij}^{\alpha\alpha}(x) \right] (\delta_{ik} \delta_{jl} - \delta_{il} \delta_{jk}) , \]  

(5.101)

\[ Q_{ij;kl}^{\alpha\beta} = \frac{1}{2} \delta_{ik} \delta_{jl} \left[ \Xi_{ij} + F_{ij}^{\alpha\beta}(y) \right] , \]  

(5.102)

\[ C_{ij;kl}^{\alpha\alpha} = \frac{1}{4} \delta_{ik} \delta_{jl} \left[ 2n_i - F_{ij}^{\alpha\alpha}(x) \right] + \frac{1}{2} \delta_{ij} \delta_{kl} F_{ik}^{\alpha\alpha}(x) , \]  

(5.103)

and

\[ C_{ij;kl}^{\alpha\beta} = \frac{1}{2} \delta_{ik} \delta_{jl} \left[ n_i - F_{ij}^{\alpha\beta}(y) \right] \]  

(5.104)

in the \(|\chi_i \otimes \chi_j\rangle\) basis, where

\[ F_{ij}^{\sigma\sigma'}(z) \overset{\text{def}}{=} c_{\sigma\sigma'} \left( n_i^z n_j^{2-z} + n_j^z n_i^{2-z} \right) \]  

(5.105)

and \(\Xi_{ij} = 1 - n_i - n_j\). It follows that \(\hat{Q}_2^{\alpha\alpha}\) is diagonal in the \(|\chi_i \wedge \chi_j\rangle\) basis, with eigenvalues \(\{ q_{ij}^{\alpha\alpha} | i < j \}\),

\[ q_{ij}^{\alpha\alpha} = \Xi_{ij} + \frac{1}{2} F_{ij}^{\alpha\alpha}(x) . \]  

(5.106)

\(\hat{Q}_2^{\alpha\beta}\) and \(\hat{G}_2^{\alpha\beta}\) are each diagonal in the \(|\chi_i \otimes \chi_j\rangle\) basis. Their eigenvalues are

\[ q_{ij}^{\alpha\beta} = \frac{1}{2} \left[ \Xi_{ij} + F_{ij}^{\alpha\beta}(y) \right] \]  

(5.107)

and

\[ g_{ij}^{\alpha\beta} = \frac{1}{2} \left[ n_i - F_{ij}^{\alpha\beta}(y) \right] \]  

(5.108)

respectively, for each \(i\) and \(j\). The eigenvalues of \(\hat{G}_2^{\alpha\alpha}\) cannot be given in closed form.

Judging from the forms of \(q_{ij}^{\alpha\alpha}\) and \(q_{ij}^{\alpha\beta}\), one may anticipate violations of the \(Q\)-condition if both \(n_i \approx 1\) and \(n_j \approx 1\), for if \(n_i = 1 = n_j\) then \(q_{ij}^{\alpha\alpha} \geq 0\) only when \(c_{\alpha\alpha} \geq 1\). However, \(c_{\alpha\alpha} \leq 1\) was a rigorous bound. It therefore appears as if the \(\text{GHF}(x,y)\) reconstruction necessarily violates the \(Q\)-condition, insofar as core orbitals
have nearly unit occupancy. The same problem arises for $Q_{i}^{a\beta}$: when $n_{i} = 1 = n_{j}$, 
$q_{ij}^{a\beta} \geq 0$ if and only if $c_{a\beta} \geq \frac{1}{2}$, whereas we showed previously that $c_{a\beta} \leq \frac{1}{2}$ is a rigorous bound.

Let us digress to consider minimization of the $\text{GHF}(x, y)$ energy functional. At the present time, no successful DMFT calculations have been carried out with this functional because the variational algorithm described in Section 5.4 is inadequate for this purpose. This algorithm satisfies the 1-RDM trace constraint only asymptotically, as $K \to \infty$, while formulas for the normalizing functions $c_{aa}$ and $c_{a\beta}$ were derived under the assumption that $\sum_{i} n_{i} = N/2$ exactly. When we attempt to optimize $\text{GHF}(x, y)$ via a penalty-function approach, the normalizing functions take on unreasonable values and normalization of the 2-RDM is violated by an amount incommensurate to the error in the 1-RDM normalization, which ultimately causes the optimization to fail. To optimize $\text{GHF}(x, y)$ it will therefore be necessary to use a method that satisfies the 1-RDM trace constraint at all intermediate stages of the optimization. One such procedure, described by Cioslowski and Pernal,\textsuperscript{171} employs as variational parameters the deviations from integer occupancies, rather than the occupation numbers themselves. If we write

$$n_{i} = n_{i}^{\text{HF}} + \varsigma_{i} - \frac{1}{\nu} \sum_{k=1}^{\nu} \varsigma_{i},$$

(5.109)

where $n_{i}^{\text{HF}} = 0$ or 1, then the trace constraint is satisfied for all values of the auxiliary variational parameters $\varsigma_{i}$. One drawback to this approach is that we can no longer use the simple parametrization $n_{i} = (\cos \omega_{i})^{2}$ to enforce the Pauli-Coleman constraints. These must instead be enforced by penalty functions, numerical step-size controls,\textsuperscript{171} or a combination of both.
We plan to implement this modified variational procedure in order to test the GHF\((x, y)\) functionals, but this has not been done at the present time. To offer an immediate comparison of this functional’s \(N\)-representability violations with those of the other functionals discussed in this chapter, we will analyze 2-RDMs that are reconstructed from CI natural occupation numbers. Specifically, we employ the same three test sets of occupation numbers that were used to obtain Fig. 5.31, namely, those from Bunge’s accurate CI wavefunction\(^{164}\) for Be, along with those obtained from FCI/6-311G++G** wavefunctions for Be and LiH.

We have plotted in Fig. 5.34 the smallest (most negative) eigenvalues of \(\hat{Q}_2^{\alpha\alpha}\) and \(\hat{Q}_2^{\alpha\beta}\), versus the reconstruction parameters \(x\) and \(y\), using these three sets of occupation numbers. The smallest eigenvalue of \(\hat{G}_2^{\alpha\beta}\) is plotted against \(y\) in Fig. 5.35. Note that the largest eigenvalues of \(\hat{Q}_2^{\alpha\beta}\) and \(\hat{G}_2^{\alpha\beta}\) are 1.0 in the normalization employed here, while the largest eigenvalue of \(\hat{Q}_2^{\alpha\alpha}\) is 0.5. The \(Q\)-condition especially (Fig. 5.34) is a clear indication that the only reasonable values for \(x\) and \(y\) are those lying near unity.

Using our three test sets of natural occupation numbers, we have calculated the spectra of \(\hat{Q}_2^{\alpha\alpha}\), \(\hat{Q}_2^{\alpha\beta}\), \(\hat{G}_2^{\alpha\alpha}\), and \(\hat{G}_2^{\alpha\beta}\) for several values of \(x\) and \(y\) near unity. Tables 5.15–5.18 list the smallest few eigenvalues obtained in each case. In particular, Tables 5.15 and 5.16 include all negative eigenvalues of \(\hat{Q}_2\), while the negative eigenvalues of \(\hat{G}_2\) are more numerous and only the most negative ones are listed in Tables 5.17 and 5.18. The negative eigenvalues in these tables should be compared to those obtained using the functionals described in Section 5.2, which are summarized in Table 5.19.
Figure 5.34: Smallest eigenvalues of $\hat{Q}^{\alpha\alpha}_2$ and $\hat{Q}^{\alpha\beta}_2$, as functions of the reconstruction parameters $x$ and $y$. The inset provides a closer view of the upper right corner.

Figure 5.35: Smallest eigenvalue of $\hat{G}^{\alpha\beta}_2$, as a function of the reconstruction parameter $y$. 
| Occupation numbers | \(x = 0.90\) & \(x = 0.95\) & \(x = 1.00\) |
|-------------------|----------------|----------------|
| \(\hat{D}_1\)     | \(\tilde{\hat{D}}_2\) | \(\hat{D}_1\)     | \(\tilde{\hat{D}}_2\) | \(\hat{D}_1\)     | \(\tilde{\hat{D}}_2\) |
| Be (Bunge; Ref. 164) | -8.06(-2) & -7.55(-2) & -7.35(-2) & -6.92(-2) & -7.11(-2) & -6.72(-2) |
|                   | 9.21(-4)  & 2.80(-2)  & -1.00(-5)  & 2.47(-2)  & -4.41(-4)  & 2.36(-2)  |
|                   | 9.21(-4)  & 2.80(-2)  & -1.00(-5)  & 2.47(-2)  & -4.41(-4)  & 2.36(-2)  |
|                   | 1.87(-3)  & 3.08(-2)  & 1.79(-3)   & 2.80(-2)  & 1.71(-3)   & 2.71(-2)  |
| Be (FCI/6-311++G**) | -8.18(-2) & -7.64(-2) & -7.49(-2) & -7.04(-2) & -7.27(-2) & -6.85(-2) |
|                   | -1.61(-4) & 2.76(-2)  & -1.20(-3)  & 2.44(-2)  & -1.55(-3)  & 2.33(-2)  |
|                   | -1.61(-4) & 2.76(-2)  & -1.20(-3)  & 2.44(-2)  & -1.55(-3)  & 2.33(-2)  |
|                   | 8.75(-4)  & 3.06(-2)  & 8.08(-4)   & 2.80(-2)  & 7.30(-4)   & 2.71(-2)  |
| LiH (FCI/6-311++G**) | -3.60(-2) & -3.51(-2) & -3.06(-2) & -3.00(-2) & -2.88(-2) & -2.83(-2) |
|                   | 1.63(-3)  & 1.35(-2)  & 1.44(-3)   & 1.11(-2)  & 1.15(-3)   & 1.02(-2)  |

Table 5.15: Smallest eigenvalues of \(\tilde{Q}_2^{\alpha\alpha}\) for several values of the reconstruction parameter \(x\). Columns labeled “\(D_1\)” use the exact 1-RDM to reconstruct \(\tilde{Q}_2^{\alpha\alpha}\) while columns labeled “\(\tilde{\hat{D}}_2\)” use in its place the normalized partial trace of the reconstructed 2-RDM. Numbers in parentheses indicate powers of ten.
<table>
<thead>
<tr>
<th>Occupation numbers</th>
<th>$y = 0.90$</th>
<th>$2\downarrow \hat{D}_2$</th>
<th>$y = 0.95$</th>
<th>$2\downarrow \hat{D}_2$</th>
<th>$y = 1.00$</th>
<th>$2\downarrow \hat{D}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Be (Bunge; Ref. 164)</td>
<td>(-3.07(-3))</td>
<td>(-2.42(-3))</td>
<td>(-7.44(-4))</td>
<td>(-3.09(-4))</td>
<td>(2.13(-5))</td>
<td>(3.93(-4))</td>
</tr>
<tr>
<td>Be (FCI/6-311++G**)</td>
<td>(-2.99(-3))</td>
<td>(-2.34(-3))</td>
<td>(-7.41(-4))</td>
<td>(-3.06(-4))</td>
<td>(3.82(-7))</td>
<td>(3.73(-4))</td>
</tr>
<tr>
<td>LiH (FCI/6-311++G**)</td>
<td>(-1.97(-3))</td>
<td>(-1.84(-3))</td>
<td>(-4.84(-4))</td>
<td>(-4.08(-4))</td>
<td>(1.32(-6))</td>
<td>(6.07(-5))</td>
</tr>
</tbody>
</table>

Table 5.16: Smallest eigenvalues of $\hat{Q}_2^{n,\beta}$ for several values of the reconstruction parameter $y$. The notation is the same as in Table 5.15.
Table 5.17: Smallest eigenvalues of $\hat{G}_2^{\alpha\alpha}$ for several values of the reconstruction parameter $x$. The notation is the same as in Table 5.15.
<table>
<thead>
<tr>
<th>Occupation numbers</th>
<th>$y = 0.90$</th>
<th>$y = 0.95$</th>
<th>$y = 1.00$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{D}_1$</td>
<td>$\hat{D}_2$</td>
<td>$\hat{D}_1$</td>
</tr>
<tr>
<td>Be (Bunge; Ref. 164)</td>
<td>-7.92(-4)</td>
<td>-2.67(-4)</td>
<td>-1.78(-4)</td>
</tr>
<tr>
<td></td>
<td>-7.92(-4)</td>
<td>-2.67(-4)</td>
<td>-1.78(-4)</td>
</tr>
<tr>
<td></td>
<td>-7.92(-4)</td>
<td>-2.67(-4)</td>
<td>-1.78(-4)</td>
</tr>
<tr>
<td></td>
<td>-1.81(-4)</td>
<td>-1.86(-4)</td>
<td>-4.30(-5)</td>
</tr>
<tr>
<td>Be (FCI/6-311++G**)</td>
<td>-8.22(-4)</td>
<td>-2.61(-4)</td>
<td>-1.95(-4)</td>
</tr>
<tr>
<td></td>
<td>-8.22(-4)</td>
<td>-2.61(-4)</td>
<td>-1.95(-4)</td>
</tr>
<tr>
<td></td>
<td>-8.22(-4)</td>
<td>-2.61(-4)</td>
<td>-1.95(-4)</td>
</tr>
<tr>
<td></td>
<td>-1.77(-4)</td>
<td>-1.82(-4)</td>
<td>-4.26(-5)</td>
</tr>
<tr>
<td>LiH (FCI/6-311++G**)</td>
<td>-6.29(-4)</td>
<td>-5.96(-4)</td>
<td>-1.47(-4)</td>
</tr>
<tr>
<td></td>
<td>-3.89(-4)</td>
<td>-3.82(-4)</td>
<td>-8.17(-5)</td>
</tr>
<tr>
<td></td>
<td>-3.44(-4)</td>
<td>-3.82(-4)</td>
<td>-8.17(-5)</td>
</tr>
<tr>
<td></td>
<td>-3.44(-4)</td>
<td>-3.69(-4)</td>
<td>-7.20(-5)</td>
</tr>
</tbody>
</table>

Table 5.18: Smallest eigenvalues of $\hat{G}_{2}^{a,b}$ for several values of the reconstruction parameter $y$. The notation is the same as in Table 5.15.
of these involves a reconstruction formula for the 2-RDM in terms of the 1-RDM.
In this chapter we have examined several proposed NO functionals for DMFT. Each

2.8 Summary and conclusion

to have repaired most of the violations in the $D$-, $Q$-, and $G$-conditions.
have to compute optimized 2-RDMs to be certain, but tentatively CHF $(\hat{h} \cdot x)$ appears in Table 5.20. We will
$Q_2$} are listed in Table 5.15. The negative eigenvalues are small. The number of negative eigenvalues of $Q_2$
the same that strict positivity is obtained only for $t = 1$, but for other values of $G_2$.
when was observed for $Q_2$. The results for $Q_2$ (Table 5.18) mimic those for $G_2$. Though they are greater in number
the 2.17 is nearly so large in magnitude as $10^{-2}$. Therefore they are greater in number.
introduced in Section 2.2. The negative eigenvalues of order $-10^{-2}$ to $-10^{-3}$ and $Q_2$ (Table 2.15) is characterized by a single negative eigenvalue of order $-10^{-2}$ and
First we discuss the spectrum of $G_2$ for CHF $(\hat{h} \cdot x)$, $D_2$, and $Q_2$-conditions.
the negative part these results are insensitive
the number of negative eigenvalues of $Q_2$. For each of the latter recon-
positivity of $Q_2$ is not so well preserved, though the situation does seem superior
introduced in Section 2.2. The negative eigenvalues of $Q_2$ are the same as those of
this chapter for which half of the eigenvalues are negative. For the functionals
the few eigenvalues (three or less) of order $-10^{-2}$ or smaller. All other eigenvalues of $Q_2$
section 5.2. For the functionals of Section 5.2, the negative eigenvalues are negative for the functionals
introduced in Section 2.2. The negative eigenvalues of $Q_2$ are the same as those of
} whereas for CHF $(\hat{h} \cdot x)$ this is not the case.

<table>
<thead>
<tr>
<th>Functional numbers</th>
<th>Occupation numbers</th>
<th>Spectra of $D_2^{\alpha\alpha}$ and $Q_2^{\alpha\alpha}$</th>
<th>Spectrum of $G_2^{\alpha\alpha}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Most negative $n = 2$</td>
<td>$n = 4$</td>
</tr>
<tr>
<td>CH(1)</td>
<td>Be (Bunge)</td>
<td>$-0.0704$ 18 199</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Be (FCI)</td>
<td>$-0.0715$ 18 129</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LiH (FCI)</td>
<td>$-0.0540$ 11 100</td>
<td></td>
</tr>
<tr>
<td>SIC-</td>
<td>Be (Bunge)</td>
<td>$-0.0704$ 14 189</td>
<td></td>
</tr>
<tr>
<td>CH(1)</td>
<td>Be (FCI)</td>
<td>$-0.0715$ 14 122</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LiH (FCI)</td>
<td>$-0.0540$ 10 93</td>
<td></td>
</tr>
<tr>
<td>CH(4/3)</td>
<td>Be (Bunge)</td>
<td>$-0.0325$ 8 83</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Be (FCI)</td>
<td>$-0.0332$ 8 60</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LiH (FCI)</td>
<td>$-0.0231$ 8 49</td>
<td></td>
</tr>
<tr>
<td>SIC-</td>
<td>Be (Bunge)</td>
<td>$-0.0325$ 7 77</td>
<td></td>
</tr>
<tr>
<td>CH(4/3)</td>
<td>Be (FCI)</td>
<td>$-0.0332$ 7 54</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LiH (FCI)</td>
<td>$-0.0231$ 8 43</td>
<td></td>
</tr>
<tr>
<td>CHF(1)</td>
<td>Be (Bunge)</td>
<td>$-0.0403$ 10 164</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Be (FCI)</td>
<td>$-0.0417$ 10 119</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LiH (FCI)</td>
<td>$-0.0144$ 2 91</td>
<td></td>
</tr>
<tr>
<td>MCHF</td>
<td>Be (Bunge)</td>
<td>$-0.0528$ 16 186</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Be (FCI)</td>
<td>$-0.0537$ 16 129</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LiH (FCI)</td>
<td>$-0.0397$ 11 100</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.19: Summary of the spectra of $D_2^{\alpha\alpha}$, $Q_2^{\alpha\alpha}$, and $G_2^{\alpha\alpha}$ for the functionals introduced in Section 5.2. Bunge's Be wavefunction comes from Ref. 164; the FCI wavefunctions employed the 6-311++G** basis.
specifically, each reconstructs the 2-RDM in the NO product basis, with matrix elements that are simple functions of the natural occupation numbers. These functionals differ from the HF reconstruction only by a modification of the HF exchange term. The resulting energy functionals $E[\hat{D}_1]$ are generally minimized by a non-idempotent 1-RDM, and are capable of recovering a substantial amount of both dynamical and non-dynamical correlation energy. Unfortunately, none of these functionals (save HF) is $N$-representable, as can be seen immediately by noting that none of the ansätze fully respect the antisymmetry requirement for the reconstructed 2-RDM. We have also given a formal proof that half of the eigenvalues $\hat{D}_2^{\alpha\alpha}$ are necessarily negative, for each of the proposed reconstructions.

To determine whether these prescriptions for reconstructing $\hat{D}_2$ are at all plausible, we examined the behavior of various pair densities reconstructed from an accurate 1-RDM for the ground state of Be. Although the 1-RDM functionals yield qualitatively
and quantitatively accurate descriptions of the distribution of opposite-spin electron pairs, none of them correctly describes the distribution of same-spin electron pairs. In light of the positivity result mentioned above, it is not surprising to discover that these functionals give a negative-valued pair density for parallel-spin electrons, at small values of the pair separation. That is, $I_{\alpha\alpha}(u) < 0$ for small $u$, which is certainly unphysical.

In view of this result, and also on the basis of some small active space calculations for the CHF(1) functional, it has been suggested\textsuperscript{144} that the electronic energy from DMFT might diverge (approach infinite overcorrelation) as the underlying basis set approaches completeness. To test this hypothesis, we carried out the first thorough study of the basis-set dependence of DMFT, using the Be and LiH ground states as test subjects. Despite pervasive $N$-representability violations, we find that the convergence with respect to basis size is quite satisfactory. The apparent instability documented previously\textsuperscript{144} is an artifact of a pronounced difference between the shape of diatomic potential curves calculated in a minimal basis set, and those calculated using extended basis sets. Consistent with previous results,\textsuperscript{146,147} our extended-basis calculations demonstrate that the corrected Hartree-Fock-type functionals\textsuperscript{122,128} CHF(\zeta) and MCHF are inappropriate for molecular applications. The functional CH(1) yields potential curves that are somewhat better, but are substantially overcorrelated. The partial self-interaction correction recommended by Goedecker and Umrigar\textsuperscript{124} eliminates most of this overcorrelation by eliminating those negative eigenvalues of $\hat{D}_2^{\alpha}$ that are largest in magnitude (and thus responsible for the most extensive overcorrelation). Unless eliminated explicitly, optimization of the energy functional will cause these spurious eigenvalues to attain the most negative values that are possible within
the given ansatz. For this reason, self-interaction correction is the single most important factor in achieving a realistic distribution of optimized occupation numbers. Diatomic potential curves calculated with the resulting SIC-CH(1) functional are not only qualitatively correct in all basis sets, but given a sufficiently flexible basis they approach FCI accuracy in some cases. Based on preliminary applications to strongly-correlated diatomic molecules, it appears that this functional recovers both dynamical and non-dynamical correlation. For systems characterized by configurational degeneracy, SIC-CH(1) smoothly interpolates over changes in the nominal electronic configuration, to afford potential energy curves of multireference quality.

It is true that the SIC-CH(1) functional owes its success to a cancellation of errors. In applications to Be and LiH, for example, we find that negative eigenvalues of $\hat{D}_2^{\alpha\alpha}$ reduce the electron repulsion energy by as much as $1-2\ E_h$, a stabilizing effect equal to 40–80% of the positive, destabilizing electron repulsion energy that is incurred by the positive eigenvalues of $\hat{D}_2^{\alpha\alpha}$. That being said, we feel that the SIC-CH(1) functional is simply too accurate and too robust to be succeeding by luck alone; there must be some important physical or mathematical principle lurking beneath this ansatz, which is responsible for its success. A more thorough examination of optimized hole densities may shed light on the physics of the matter, since the CH(1) functional upon which SIC-CH(1) is built can be derived from a particular model for the exchange-correlation hole.\textsuperscript{123} Note that the self-interaction correction did not come out of this model, but had to be grafted on at the end, and in fact this correction has a deleterious effect on the partial trace relation that connects the 1- and 2-RDMs (though the deviations induced by the correction are small). The combination of being nearly free of self-interaction and almost satisfying the partial trace condition
is apparently very powerful. It is the author’s suspicion that any halfway reasonable ansatz for $\hat{D}_2[\hat{D}_1]$ that is free of self-interaction and also satisfies the sum rule is likely to produce good potential energy surfaces, when combined with orbital optimization.*

While the negative 2-RDM eigenvalues manifested by SIC-CH(1) and other functionals are certainly undesirable, it is perhaps inappropriate to demand that reconstruction functionals for DMFT exactly comply with all known $N$-representability requirements for the 2-RDM. Even if this were feasible, it would likely result in a method whose computational cost rivals that of proper variational 2-RDM methods. Instead, we suggest that the more appropriate niche for DMFT is to find models that produce good results (preferably, for identifiable reasons) and are largely free of egregious $N$-representability violations. To this end, we suggest that explicitly antisymmetric reconstruction functionals should help to correct several of the outstanding problems with the current generation of functionals. For instance, any antisymmetric reconstruction is rigorously free of all self-interaction error, since the correlation part (cumulant) of any antisymmetric reconstruction of $\hat{D}_2$ must vanish for a one-electron system. In contrast, the ad hoc correction that defines SIC-CH($\zeta$) removes only the “orbital” self-interactions (spurious diagonal elements in the NO product basis). Furthermore, we have shown that antisymmetry cures the positivity problem in certain important cases, and more generally antisymmetric reconstruction functionals are expected to be less susceptible to the negative 2-RDM eigenvalues. Finally, only by means of an antisymmetric reconstruction can one hope to obtain qualitatively correct pair distributions for parallel-spin electrons.

Based on these ideas, we have begun to pursue the development of novel recon-

*The orbital optimization is crucial; energies obtained using CI 1-RDMs in $E[\hat{D}_1]$ are quite poor.
struction functionals for DMFT. An antisymmetric ansatz that generalizes the HF reconstruction was introduced, which satisfies two other exact relations that are not preserved by any of the other reconstructions considered in this chapter. First, the reconstructed 2-RDM is necessarily positive; and second, the corresponding energy functional reproduces the correct scaling properties for all contributions to the electronic energy (kinetic energy, electron repulsion energy, etc.), which scale differently with respect to a uniform dilation of coordinate space. By construction, both spin components of the reconstructed 2-RDM in this ansatz are properly (and separately) normalized, which we hope will prevent large violations of the partial trace relation that connects $\hat{D}_1$ and $\hat{D}_2$, and which is not enforced explicitly by our ansatz. Technical problems, which should not be insurmountable, have hindered optimization of this new functional. However, we have tested certain $N$-representability conditions for this reconstruction, using CI occupation numbers, and we find that these violations are much less severe than those exhibited by any of the other functionals considered in this chapter.

5.9 References


doubles and singles + doubles + triples + quadruples spaces”, Theor. Chem.
Acc., 107, 220 (2002).


[88] W. Meyer, “PNO-CI studies of electron correlation effects. I. Configuration in-
teraction expansion by means of nonorthogonal orbitals, and application to the

[89] W. Meyer, “Configuration expansion by means of pseudonatural orbitals”, in
Methods of Electronic Structure Theory, Vol. 3 of Modern Theoretical Chemistry,

Rev. B, 12, 2111 (1975).


[93] S. Valone, “Consequences of extending 1 matrix energy functionals from pure-
state representable to all ensemble representable 1 matrices”, J. Chem. Phys.,
73, 1344 (1980).

[94] E. V. Ludeña and A. Sierraalta, “Necessary conditions for the mapping of γ into


[96] M. Levy, “Correlation energy functionals of one-matrices and Hartree-Fock den-
sities”, in Density Matrices and Density Functionals, edited by R. Erdahl and


APPENDICES
Appendix A

Storage and indexing of many-particle tensors

In this Appendix, we derive indexing schemes for efficient storage of many-particle RDMs. Efficient storage schemes for 3- and 4-electron tensors have not appeared in the literature, because for applications one is interested in (at most) the 2-RDM, a four-index tensor with approximately as many elements as there are two-electron integrals. This is true even when solving CSE(2) or its connected equivalents, since the 3- and 4-electron terms in those equations are contracted into two-electron ones, so all 3- and 4-electron information can be incorporated on-the-fly, without explicit construction or storage of anything larger than a four-index tensor. However, in order to evaluate various reconstruction approximations for $D_3$ and $D_4$, it is necessary to construct and store these higher-order tensors. To do this even for small basis sets requires that one take full advantage of Hermitian and permutational symmetries among the density matrix elements. The means to accomplish this are presented here.
A.1 Non-redundant indexing schemes

Suppose that we wish to store a skew-symmetric (one-particle) $\nu \times \nu$ matrix using as little memory as possible. This is accomplished by storing the matrix as a one-dimensional array $\vec{v} = (v_1, \ldots, v_{\nu(\nu-1)/2})$, and we need a formula that assigns a single, unique index for each element in the upper triangle of the matrix in question:

\[
\begin{pmatrix}
1 & 2 & 3 & 4 & 5 & \cdots & \nu - 1 \\
\nu & \nu + 1 & \nu + 2 & \nu + 3 & \cdots & 2\nu - 3 \\
2\nu - 2 & 2\nu - 1 & 2\nu & \cdots & 3\nu - 6 \\
3\nu - 5 & 3\nu - 4 & \cdots & 4\nu - 10 \\
4\nu - 9 & \cdots & 5\nu - 15 \\
\vdots & \vdots & \ddots & & & & \\
\end{pmatrix}
\]

Let $i$ and $j$ index the rows and columns of this matrix, respectively, and let $[i, j]_{aa}$ denote the value of the index for $\vec{v}$. (The “$\alpha\alpha$” foreshadows our eventual application to RDM spin components.) Inspecting the values of $[i, j]_{aa}$ that appear above, we have

\[
\begin{align*}
[1, j]_{aa} &= j - 1 \quad (A.1a) \\
[2, j]_{aa} &= \nu + j - 3 \quad (A.1b) \\
[3, j]_{aa} &= 2\nu + j - 6 \quad (A.1c) \\
[4, j]_{aa} &= 3\nu + j - 10 \quad (A.1d) \\
& \vdots \\
\end{align*}
\]

for various values of $i$. Extrapolating from this pattern leads us to put

\[
[i, j]_{aa} \overset{\text{def}}{=} \nu(i - 1) + j - \sum_{m=1}^{i} m
\]

\[
= \nu(i - 1) + j - i(i + 1)/2
\]

\[
(A.2)
\]
for \( i < j \). As a check of this formula, note that \([\nu, \nu]_{\alpha\alpha} = \nu(\nu + 1)/2\), which is the number of independent elements of an \( \nu \times \nu \) skew-symmetric matrix. An important point is that the definition of the mapping \([i, j]_{\alpha\alpha}\) depends parametrically on \( \nu \), the maximum permissible value of the indices \( i \) and \( j \). Thus when a function such as \([i, j]_{\alpha\alpha}\) is nested with itself or other, similar indexing function, one must replace \( \nu \) in Eq. (A.2) with the appropriate maximal index value.

For a matrix that is symmetric rather than skew-symmetric, the one-dimensional array \( \bar{v} \) should have length \( \nu(\nu + 1)/2 \) in order to store the diagonal plus the upper triangle. An analysis just like the one above demonstrates that for \( i \leq j \),

\[
[i, j]_{\text{sym}} = \nu(i - 1) + j - i(i - 1)/2
\]  

(A.3)
is a unique index for the independent elements of a symmetric matrix.

A similar indexing scheme for the two-electron integrals is well known. \(^1\) Namely, if we put

\[
[i, j]_{\text{int}} \overset{\text{def}}{=} j + i(i - 1)/2,
\]  

(A.4)

for \( i \geq j \), then \([i, j]_{\text{int}}, [k, l]_{\text{int}}\) is a unique index for the two-electron integrals \( \langle ij|kl \rangle \).

This formula arises from the symmetries

\[
\langle ij|kl \rangle = \langle kj|il \rangle = \langle il|kj \rangle = \langle kl|ij \rangle
\]  

(A.5)

\[
= \langle ji|lk \rangle = \langle li|jk \rangle = \langle jk|li \rangle = \langle lk|ji \rangle,
\]

and leads to the formula quoted in Eq. (2.66) for the total number of two-electron integrals.

Now we consider unique, non-redundant indexing schemes for the RDMs. In the interest of efficiency, RDMs are always stored in terms of their spin components, so that
the indices run over orbitals rather than spin orbitals. This obviates the allocation of memory blocks for components that are either identically zero or are related by permutations to other spin components. Next we note that spin components such as $D^{\alpha \alpha}_2$, $D^{3 \beta \beta}_3$, etc., involving only parallel-spin electrons are completely antisymmetric in their orbital indices. For example,

$$D^{\alpha \alpha}_{ijklrst} = -D^{\alpha \alpha}_{jik:rst} = D^{\alpha \alpha}_{jik:rt} = -D^{\alpha \alpha}_{jki:rt} = \cdots . \quad (A.6)$$

In contrast, mixed spin components like $D^{\alpha \beta}_3$ are antisymmetric with respect to permutations of identical spin indices,

$$D^{\alpha \beta}_{ijk:rst} = -D^{\alpha \beta}_{jik:rst} = D^{\alpha \beta}_{jik:srt} = -D^{\alpha \beta}_{jki:srt} ; \quad (A.7)$$

but there is no special symmetry between indices for electrons with different spins.

The upshot is that we ought to store $D^{\alpha \alpha}_{ijklrst}$ only for $i < j < k$ and $r < s < t$, while $D^{\alpha \beta}_{ijklrst}$ should be stored only for $i < j$ and $r < s$ (but for each $k$ and $t$). In addition, since each of these tensors is self-adjoint, we should store only those elements for which the composite upper index is less than or equal to the composite lower index.

The appropriate composite indices for $D^{\alpha \alpha}_{ijkl}$ are $[i,j]_{aa}$ and $[k,l]_{aa}$. (Obviously the $D^{\beta \beta}_{ijkl}$ are indexed in the same way.) This reduces the four-index tensor $D^{\alpha \alpha}_2$ with elements $D^{\alpha \alpha}_{ijkl}$ to a two-index matrix with elements $D^{\alpha \alpha}_{[i,j]_{aa};[k,l]_{aa}}$. This matrix is self-adjoint, and the two indices may be folded into a single index using the indexing function $[p,q]_{\text{sym}}$ introduced in Eq. (A.3). As noted above, $\nu$ in this equation must be replaced by $[\nu-1, \nu]_{aa}$, the maximum value of $[i,j]_{aa}$. The result of all this is a single indexing function

$$\begin{cases}
  i, j \\
  r, s \\
\end{cases}_{\alpha \alpha} \overset{\text{def}}{=} [\nu - 1, \nu]_{aa} ([i,j]_{aa} - 1) + [r,s]_{aa} - \frac{1}{2} [i,j]_{aa} ([i,j]_{aa} - 1) \quad (A.8)
$$
that assigns a unique index when $i < j$, $r < s$, and $[i,j]_{aa} \leq [r,s]_{aa}$. Similar considerations for $D_{2}^{\alpha\beta}$ lead to the indexing function

$$\begin{align*}
\left\{ \begin{array}{c}
i, j \\
r, s
\end{array} \right\}_{\beta\alpha} & \overset{\text{def}}{=} [\nu, \nu]_{\beta\alpha} \left( [i,j]_{\beta\alpha} - 1 \right) + [r,s]_{\beta\alpha} - \frac{1}{2} [i,j]_{\beta\alpha} \left( [i,j]_{\beta\alpha} - 1 \right) \quad (A.9)
\end{align*}$$

where

$$[i,j]_{\beta\alpha} \overset{\text{def}}{=} j + \nu(i - 1) \quad . \quad (A.10)$$

Overall, this indexing scheme for the 2-RDM is more efficient than the one discussed by Nakatsuji and co-workers,\(^2\) which utilizes antisymmetry but apparently does not resolve the 2-RDM into spin components.

Efficient storage of $D_{3}^{\alpha\alpha\alpha}$ and $D_{3}^{\beta\beta\beta}$ requires a three-index analogue of $[i,j]_{aa}$. We shall denote as $[i,j,k]_{\alpha\alpha\alpha}$, which by definition assigns a unique index to each ordered triple $(i,j,k)$ such that $1 \leq i < j < k \leq \nu$. A partial enumeration of such triples is given in in Fig. A.1. Upon studying this list, one may determine that the appropriate indexing function is

$$[i,j,k]_{\alpha\alpha\alpha} \overset{\text{def}}{=} \sum_{i_{0}=1}^{i-1} \sum_{j_{0}=i_{0}+1}^{\nu-1} (\nu - j_{0}) + \sum_{j_{0}=i_{0}+1}^{j-1} (\nu - j_{0}) + k - j \quad . \quad (A.11)$$

The first term in this expression counts the number of ordered triples $(i_{0},j_{0},k_{0})$ whose first index $i_{0}$ is less than the specified value, $i$. The second term counts the number of ordered triples for which $i_{0} = i$ but $j_{0} < j$. Finally, $k - j$ equals the number of terms for which $(i_{0},j_{0}) = (i,j)$ but $k_{0} < k$. This sum brings us to the desired ordered triple $(i,j,k)$. The summations in Eq. (A.11) can be evaluated in closed form, using the fact that

$$\sum_{m=1}^{r} m = r(r + 1)/2 \quad . \quad (A.12)$$
The result is

\[ [i, j, k]_{aaa} = \frac{1}{6} (i-1) [3\nu^2 + (i+1)(i-3\nu)] + \nu(j-i-1) + \frac{1}{2} (i+j)(i-j+1) + k-j . \]  

Whereas the form of \([i, j]_{aa}\) can practically be guessed without elaborate proof, this expression for \([i, j, k]_{aaa}\) is not obvious, and the 4-RDM indexing functions derived below are even more complicated.

We are not quite finished with \(D_3^{aaa}\), but at this point let us consider an appropriate multi-index for the upper and lower indices \(D_{ijk;rst}^{\beta\alpha\alpha}\) of \(D_3^{\beta\alpha\alpha}\). In this case we want to consider ordered triples \((i, j, k)\) such that \(1 \leq j < k \leq \nu\), but \(i\) is not restricted by the values of \(j\) and \(k\), thus \(1 \leq i \leq \nu\). The multi-index for this situation
can be obtained by modification of the summation limits in Eq. (A.11). Thus we put

$$[i, j, k]_{\alpha\alpha\alpha} \overset{\text{def}}{=} \sum_{i_0=1}^{i-1} \sum_{j_0=1}^{j-1} \sum_{k_0=1}^{k-1} (\nu - j_0) + (\nu - j_0) + k - j$$

$$= \nu (j - 1) - \frac{1}{2} j (j - 1) + \frac{1}{2} \nu (\nu - 1) (i - 1) + k - j$$ \hspace{1cm} (A.14)

the three terms in the first line having the same origins as their analogues in Eq. (A.11).

For a singlet state, the tensor $D_{3}^{\beta\alpha}$ is equivalent to $D_{3}^{\alpha\alpha\beta}$, since a singlet cannot distinguish between $\alpha$ and $\beta$ spin, but more generally these two tensors can be indexed in essentially the same fashion.

Summations such as the ones in Eq. (A.14) can be evaluated by repeated application of the identity in Eq. (A.12). To implement these complicated indexing formulas in practice, we use MATHEMATICA to evaluate the summations and to write the result directly to Fortran code. We have also verified independently that these indexing formulas do indeed generate a non-redundant list of all independent density matrix elements.

We have thus reduced storage of the 3-RDM for a singlet state to storage of a pair of two-index matrices, having elements $D_{[i, j, k]_{\alpha\alpha\alpha};[r, s, t]_{\alpha\alpha\alpha}}^{\alpha\alpha\alpha}$ and $D_{[i, j, k]_{\alpha\alpha\alpha};[r, s, t]_{\beta\beta\beta}}^{\beta\beta\beta}$. Two additional matrices are required for non-singlet states but are indexed in the same way. Each of these matrices is self-adjoint and can be folded into a one-dimensional array, again using the $[p, q]_{\text{sym}}$ index for this purpose, but replacing $\nu$ in Eq. (A.3) with the maximum value of either $[i, j, k]_{\alpha\alpha\alpha}$ or $[i, j, k]_{\beta\beta\beta}$. These maximum values are

$$[\nu - 2, \nu - 1, \nu]_{\alpha\alpha\alpha} = \frac{1}{6} \nu (\nu^2 - 3\nu + 2)$$ \hspace{1cm} (A.15a)

$$[\nu, \nu - 1, \nu]_{\beta\beta\beta} = \frac{1}{2} \nu^2 (\nu - 1)$$ \hspace{1cm} (A.15b)
The function that assigns a single unique index to $D_{ijk;rst}^{aaa}$ is thus found to be
\[
\begin{array}{l}
\{ i, j, k \\ r, s, t \}^{def}_{aaa} = [\nu - 2, \nu - 1, \nu]_{aaa}([i, j, k]_{aaa} - 1) \\
+ [r, s, t]_{aaa} - \frac{1}{2} [i, j, k]_{aaa}([i, j, k]_{aaa} - 1)
\end{array}
\] (A.16)
where it is assumed that $i < j < k$, $r < s < t$, and $[i, j, k]_{aaa} \leq [r, s, t]_{aaa}$. The corresponding index function for the $D_{ijk;rst}^{\beta\alpha\alpha}$ is
\[
\begin{array}{l}
\{ i, j, k \\ r, s, t \}^{def}_{\beta\alpha\alpha} = [\nu, \nu - 1, \nu]_{\beta\alpha\alpha}([i, j, k]_{\beta\alpha\alpha} - 1) \\
+ [r, s, t]_{\beta\alpha\alpha} - \frac{1}{2} [i, j, k]_{\beta\alpha\alpha}([i, j, k]_{\beta\alpha\alpha} - 1)
\end{array}
\] (A.17)
whose use is subject to the constraints $j < k$, $s < t$, and $[i, j, k]_{\beta\alpha\alpha} \leq [r, s, t]_{\beta\alpha\alpha}$.

We now consider storage of $D_4$ in terms of its spin components $D_4^{\alpha\alpha\alpha\alpha}$, $D_4^{\beta\alpha\alpha\alpha}$, and $D_4^{\beta\beta\alpha\alpha}$. Except for singlet states, the spin components $D_4^{\beta\beta\beta}$ and $D_4^{\beta\beta\beta\beta}$ are also required, but these are indexed in essentially the same manner as other components, and need not be treated explicitly here. If one enumerates ordered quadruples $(i, j, k, l)$, with $i < j < k < l < \nu$, in essentially the manner that is shown for the triples in Fig. A.1, the multi-index appropriate for four antisymmetric indices is seen to be
\[
[i, j, k, l]_{\alpha\alpha\alpha\alpha}^{def} = \sum_{k_0=1}^{i-1} \sum_{j_0=1}^{i-2} \sum_{k_0=1}^{i-1} (\nu - k_0) + \sum_{j_0=i+1}^{j-1} \sum_{k_0=j_0+1}^{i} (\nu - k_0) \\
+ \sum_{k_0=j+1}^{j-1} (\nu - k_0) + l - k
\] (A.18)
\[= \frac{1}{24} \left\{ -i^4 + i^3 (4 \nu - 6) - i^2 (6 \nu^2 - 18 \nu + 11) + i (4 \nu^3 - 18 \nu^2 + 22 \nu - 6) \\
+ 4 \left[ j^3 - 3k(k+1) + 6l - 3j^2(n-1) - 5\nu + 6k\nu - \nu^3 \\
+ j(3\nu^2 - 6\nu + 2) \right] \right\}. \]
In the first line of this equation, the first term counts non-redundant quadruples with \(i_0 < i\); the second term enumerates those with \(i_0 = i\) and \(j_0 < j\); while the third term equals the number of quadruples with \(i_0 = i, j_0 = j\), and \(k_0 < k\).

Index functions \([i, j, k, l]_{\beta\alpha\alpha\alpha}\) and \([i, j, k, l]_{\beta\beta\alpha\alpha}\) for the other spin components are obtained by similar considerations. In the case of \([i, j, k, l]_{\beta\beta\alpha\alpha}\), for example, we need quadruples \((i, j, k, l)\) such that \(1 < i < j\) and \(1 < k < l\). Counting the number of such quadruples leads to expressions similar to that in the first line of Eq. (A.18) but with different summation limits. Taking all this into consideration, one obtains

\[
[i, j, k, l]_{\beta\alpha\alpha\alpha} \stackrel{\text{def}}{=} \sum_{i_0=1}^{i-1} \sum_{j_0=1}^{\nu-2} \sum_{k_0=1}^{\nu-1} (\nu - k_0) + \sum_{j_0=1}^{j-1} \sum_{k_0=1}^{\nu-1} (\nu - k_0) + \sum_{k_0=j+1}^{k-1} (\nu - k_0) + l - k \tag{A.19}
\]

\[
= l - k + \frac{1}{2}(j + k)(j - k + 1) + \nu(k - j + 1) + \frac{1}{6} [j(j^2 - 1) + 3\nu - 3j^2\nu - 3\nu^2 + 3j\nu^2 + (i - 1)(\nu^3 - 3\nu^2 + 2\nu)]
\]

and

\[
[i, j, k, l]_{\beta\beta\alpha\alpha} \stackrel{\text{def}}{=} \sum_{i_0=1}^{i-1} \sum_{j_0=i_0+1}^{\nu} \sum_{k_0=1}^{\nu-1} (\nu - k_0) + \sum_{j_0=i+1}^{j-1} \sum_{k_0=1}^{\nu-1} (\nu - k_0) + \sum_{k_0=1}^{k-1} (\nu - k_0) + l - k \tag{A.20}
\]

\[
= l - k - \frac{1}{2} k(k - 1) + \nu(k - 1) + \frac{1}{2} \nu(\nu - 1)(j - i - 1) + \frac{1}{2} \nu(\nu + 1) \left[\nu(i - 1) - \frac{1}{2} i(i - 1)\right].
\]

As in the case of the 2- and 3-RDMs, these multi-indices convert the 4-RDM spin components into two-index, self-adjoint matrices, to be indexed using \([p, q]_{\text{sym}}\). The
final vector indices are
\[
\begin{align*}
\{ i,j,k,l \}_{aaaa} & \overset{\text{def}}{=} [\nu - 3, \nu - 2, \nu - 1, \nu]_{aaaa} ([i,j,k,l]_{aaaa} - 1) \\
& + [r,s,t,u]_{aaaa} - \frac{1}{2} [i,j,k,l]_{aaaa} ([i,j,k,l]_{aaaa} - 1) , \\
(A.21)
\end{align*}
\]
and
\[
\begin{align*}
\{ i,j,k,l \}_{\beta\alpha\alpha\alpha} & \overset{\text{def}}{=} [\nu - 3, \nu - 2, \nu - 1, \nu]_{\beta\alpha\alpha\alpha} ([i,j,k,l]_{\beta\alpha\alpha\alpha} - 1) \\
& + [r,s,t,u]_{\beta\alpha\alpha\alpha} - \frac{1}{2} [i,j,k,l]_{\beta\alpha\alpha\alpha} ([i,j,k,l]_{\beta\alpha\alpha\alpha} - 1) , \\
(A.22)
\end{align*}
\]
In Eq. (A.21) it is assumed that \( i < j < k < l \), \( r < s < t < u \), and \([i,j,k,l]_{aaaa} \leq [r,s,t,u]_{aaaa}\). Use of Eq. (A.22) is predicated upon having \( j < k < l \), \( s < t < u \), and \([i,j,k,l]_{\beta\alpha\alpha\alpha} \leq [r,s,t,u]_{\beta\alpha\alpha\alpha}\). Finally, Eq. (A.23) requires that \( i < j \), \( k < l \), \( r < s \), \( t < u \), and \([i,j,k,l]_{\beta\beta\alpha\alpha} \leq [r,s,t,u]_{\beta\beta\alpha\alpha}\).

From the indexing formulas above one can anticipate the size of the array \( \bar{\nu} \) that is required in order to store the 3- and 4-RDMs. For example,
\[
\begin{align*}
\{ \nu - 3, \nu - 2, \nu - 1, \nu \}_{aaaa} & = \frac{\nu}{1152} (\nu^7 - 12\nu^6 + 58\nu^5 - 144\nu^4 + 217\nu^3 - 276\nu^2 + 300\nu - 144) \\
(A.24)
\end{align*}
\]
is the number of independent elements of \( D_{4}^{aaaa} \). For this spin component, intelligent indexing reduces memory requirements by greater than a factor of one thousand. (This is the best-case scenario, since this component has more symmetry than other
<table>
<thead>
<tr>
<th>Component</th>
<th>No. independent elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha\alpha$</td>
<td>$\nu^3 - 2\nu^2 + 3\nu - 2)/8$</td>
</tr>
<tr>
<td>$\beta\alpha$</td>
<td>$\nu^2(\nu^2 + 1)/2$</td>
</tr>
<tr>
<td>$\alpha\alpha\alpha$</td>
<td>$\nu(\nu^5 - 6\nu^4 + 13\nu^3 - 6\nu^2 - 14\nu + 12)/72$</td>
</tr>
<tr>
<td>$\beta\alpha\alpha$</td>
<td>$\nu^2(\nu^4 - 2\nu^3 + \nu^2 + 2\nu - 2)/8$</td>
</tr>
<tr>
<td>$\alpha\alpha\alpha\alpha$</td>
<td>$\nu(\nu^7 - 12\nu^6 + 58\nu^5 - 144\nu^4 + 217\nu^3 - 276\nu^2 + 300\nu - 144)/1152$</td>
</tr>
<tr>
<td>$\beta\alpha\alpha\alpha$</td>
<td>$\nu^2(\nu^5 - 6\nu^4 + 13\nu^3 - 12\nu^2 + 10\nu^2 - 18\nu + 12)/72$</td>
</tr>
<tr>
<td>$\beta\beta\alpha\alpha$</td>
<td>$\nu^2(\nu - 1)^2(\nu^4 - 2\nu^3 + \nu^2 + 4)/32$</td>
</tr>
</tbody>
</table>

Table A.1: Storage requirements for spin components of the 2-, 3-, and 4-RDMs, as functions of the number of orbitals $\nu$.

Spin components of the 4-RDM.) The total storage requirements for the various spin components of $D_2$, $D_3$, and $D_4$ are listed in Table A.1. Spin components such as $D_3^{3\beta\alpha}$ are not listed in the table, as they clearly have the same number of elements as other components ($D_3^{3\alpha\alpha\alpha}$ in this case) that are listed.

The storage and indexing formulas described in this Appendix correspond to the most optimized RDM manipulation schemes that can be achieved for dense RDMs. Further reductions in storage and floating point operations (which could be significant) requires the introduction of special techniques for manipulation of sparse matrices.

### A.2 References


Appendix B

Additional formulas for Grassmann products

Suppose that $A_{p(1,\ldots,p,1',\ldots,p')}$ and $B_{q(1,\ldots,q,1',\ldots,q')}$ are antisymmetric $p$-particle and $q$-particle kernels, respectively. The Grassmann product $A_p \wedge B_q$ is a $(p+q)$-particle kernel, and nominally the antisymmetrization entails $(p+q)!^2$ separate terms. Since $A_p$ and $B_q$ are antisymmetric, however, many of these terms are equivalent and the result can be simplified considerably. The special case $A_p \wedge A_p$, which is needed quite frequently, also leads to much simplification. In this Appendix we give formulas for various Grassmann products of antisymmetric tensors. Certain of these formulas are required at various places in the main text, perhaps most notably in Chapter 3, in deriving a connected formulation of the contracted Schrödinger equation. The formulas in this Appendix were derived by a sequence of algebraic manipulations in Mathematica and written directly to \LaTeX output, to avoid typographical errors.
\[ (A_1 \wedge A_1 \wedge A_1)(1,2,3,1',2',3') \] (B.1) 

\[ = \frac{1}{6} \left[ - A_1(1,3') A_1(2,2') A_1(3,1') + A_1(1,2') A_1(2,3') A_1(3,1') + A_1(1,3') A_1(2,1') A_1(3,2') \right. 

\[ - A_1(1,1') A_1(2,3') A_1(3,2') - A_1(1,2') A_1(2,1') A_1(3,3') + A_1(1,1') A_1(2,2') A_1(3,3') \left. \right] . \]

\[ (A_2 \wedge A_2)(1,2,3,4,1',2',3',4') \]

\[ = \frac{1}{15} \left[ A_2(1,4,3',4') A_2(3,4,1',2') - A_2(1,4,2',4') A_2(2,3,1',4') + A_2(1,4,2',3') A_2(2,3,1',4') \right. 

\[ + A_2(1,4,1',4') A_2(2,3,2',3') - A_2(1,4,1',3') A_2(2,3,2',4') + A_2(1,4,1',2') A_2(2,3,3',4') \]

\[ - A_2(1,3,3',4') A_2(2,4,1',2') + A_2(1,3,2',4') A_2(2,4,1',3') - A_2(1,3,2',3') A_2(2,4,1',4') \]

\[ - A_2(1,3,1',4') A_2(2,4,2',3') + A_2(1,3,1',3') A_2(2,4,2',4') - A_2(1,3,1',2') A_2(2,4,3',4') \]

\[ + A_2(1,2,3',4') A_2(3,4,1',2') - A_2(1,2,2',4') A_2(3,4,1',3') + A_2(1,2,2',3') A_2(3,4,1',4') \]

\[ + A_2(1,2,1',4') A_2(3,4,2',3') - A_2(1,2,1',3') A_2(3,4,2',4') + A_2(1,2,1',2') A_2(3,4,3',4') \right] . \]

(B.2)
\[
(A_1^{A_4})^{(1,2,3,4;1',2',3',4')} = \frac{1}{24} \left\{ A_{1(4;1')} A_{1(1;4')} A_{1(2;3')} A_{1(3;3')} - A_{1(2;4')} A_{1(1;2')} A_{1(3;3')} - A_{1(1;3')} A_{1(3;2')} \right.
+ A_{1(3;4')} A_{1(1;3')} A_{1(2;2')} - A_{1(1;2')} A_{1(3;3')} \bigg[ A_{1(1;4')} A_{1(2;1')} A_{1(3;3')} - A_{1(2;4')} A_{1(1;2')} A_{1(3;3')} - A_{1(1;3')} A_{1(3;2')} \bigg] + A_{1(3;4')} A_{1(1;3')} A_{1(2;2')} - A_{1(1;2')} A_{1(3;3')} \left. \bigg[ A_{1(1;4')} A_{1(2;1')} A_{1(3;3')} - A_{1(2;4')} A_{1(1;2')} A_{1(3;3')} - A_{1(1;3')} A_{1(3;2')} \bigg] \right. 
\]
\[
+ A_{1(3;4')} A_{1(1;3')} A_{1(2;2')} - A_{1(1;2')} A_{1(3;3')} \left. \bigg[ A_{1(1;4')} A_{1(2;1')} A_{1(3;3')} - A_{1(2;4')} A_{1(1;2')} A_{1(3;3')} - A_{1(1;3')} A_{1(3;2')} \bigg] \right. 
\]
\[
+ \frac{1}{18} \left\{ A_{1(4;3)} A_{1(1;4')} A_{1(2;3')} A_{1(3;3')} - A_{1(2;4')} A_{1(1;2')} A_{1(3;3')} - A_{1(1;3')} A_{1(3;2')} \right. 
+ A_{1(3;4')} A_{1(1;3')} A_{1(2;2')} - A_{1(1;2')} A_{1(3;3')} \bigg[ A_{1(1;3')} A_{1(2;2')} A_{1(3;3')} - A_{1(2;4')} A_{1(1;2')} A_{1(3;3')} - A_{1(1;3')} A_{1(3;2')} \bigg] + A_{1(3;4')} A_{1(1;3')} A_{1(2;2')} - A_{1(1;2')} A_{1(3;3')} \left. \bigg[ A_{1(1;3')} A_{1(2;2')} A_{1(3;3')} - A_{1(2;4')} A_{1(1;2')} A_{1(3;3')} - A_{1(1;3')} A_{1(3;2')} \bigg] \right. 
\]
\[
+ A_{1(3;4')} A_{1(1;3')} A_{1(2;2')} - A_{1(1;2')} A_{1(3;3')} \left. \bigg[ A_{1(1;3')} A_{1(2;2')} A_{1(3;3')} - A_{1(2;4')} A_{1(1;2')} A_{1(3;3')} - A_{1(1;3')} A_{1(3;2')} \bigg] \right. 
\]
\[
+ A_{1(3;4')} A_{1(1;3')} A_{1(2;2')} - A_{1(1;2')} A_{1(3;3')} \left. \bigg[ A_{1(1;3')} A_{1(2;2')} A_{1(3;3')} - A_{1(2;4')} A_{1(1;2')} A_{1(3;3')} - A_{1(1;3')} A_{1(3;2')} \bigg] \right. 
\]
\[
+ A_{1(3;4')} A_{1(1;3')} A_{1(2;2')} - A_{1(1;2')} A_{1(3;3')} \left. \bigg[ A_{1(1;3')} A_{1(2;2')} A_{1(3;3')} - A_{1(2;4')} A_{1(1;2')} A_{1(3;3')} - A_{1(1;3')} A_{1(3;2')} \bigg] \right. 
\]
\begin{equation}
(A_1 \land B_2)_{(1,2,3;1',2',3')}
= \frac{1}{5}
\left[
A_1(3,3') B_2(1,2;1',2') - A_1(3,2') B_2(1,2;1',3') + A_1(3,1') B_2(1,2;2',3')
- A_1(2,3') B_2(1,3;1',2') + A_1(2,2') B_2(1,3;1',3') - A_1(2,1') B_2(1,3;2',3')
+ A_1(1,3') B_2(2,3;1',2') - A_1(1,2') B_2(2,3;1',3') + A_1(1,1') B_2(2,3;2',3')
\right].
\tag{B.5}
\end{equation}

\begin{equation}
(A_1 \land B_3)_{(1,2,3,4;1',2',3',4)}
= \frac{1}{15}
\left[
A_1(4,4') B_3(1,2,3;1',2',3') - A_1(4,3') B_3(1,2,3;1',2',4') + A_1(4,2') B_3(1,2,3;1',3',4')
- A_1(4,1') B_3(1,2,3;2',3',4') - A_1(3,4') B_3(1,2,4;1',2',3') + A_1(3,3') B_3(1,2,4;1',2',4')
- A_1(3,2') B_3(1,2,4,3;1',2',4') + A_1(3,1') B_3(1,2,4,2;1',2',3') + A_1(2,4') B_3(1,2,4,1;1',2',3')
- A_1(2,3') B_3(1,2,4,1,3;1',2',4') + A_1(2,2') B_3(1,2,4,1,2;1',2',3') - A_1(2,1') B_3(1,2,4,1,3;2',3')
+ A_1(1,4') B_3(2,3,4;1',2',3') + A_1(1,3') B_3(2,3,4;1',2',4') + A_1(1,2') B_3(2,3,4;1',3',4')
+ A_1(1,1') B_3(2,3,4;2',3',4')
\right].
\tag{B.6}
\end{equation}

Finally we have

\begin{equation}
(A_1 \land A_1 \land B_2)_{(1,2,3,4;1',2',3',4)} = X + Y ,
\tag{B.7}
\end{equation}
where

\[
X = \frac{1}{72} \left\{ A_{1(1;1')} \left[ A_{1(2;2')} B_{2(3,4,3',4')} - A_{1(3;3')} B_{2(3,4,2',4')} + A_{1(2;4')} B_{2(3,4,2',3')} \\
- A_{1(3;2')} B_{2(2,4,3',4')} + A_{1(3;3')} B_{2(2,4,2',4')} - A_{1(3;4')} B_{2(2,4,2',3')} \\
+ A_{1(4;2')} B_{2(2,3,3',4')} - A_{1(4;3')} B_{2(2,3,2',4')} + A_{1(4;4')} B_{2(2,3,2',3')} \right] \\
+ A_{1(1;2')} \left[ A_{1(2;3')} B_{2(3,4,1',4')} - A_{1(2;1')} B_{2(3,4,3',4')} + A_{1(3;1')} B_{2(2,4,3',4')} \\
- A_{1(2;4')} B_{2(3,4,1',3')} + A_{1(3;4')} B_{2(2,4,1',3')} - A_{1(3;3')} B_{2(2,4,1',4')} \\
+ A_{1(3;3')} B_{2(2,3,1',4')} - A_{1(4;4')} B_{2(2,3,1',3')} - A_{1(4;1')} B_{2(2,3,3',4')} \right] \\
+ A_{1(1;3')} \left[ A_{1(2;1')} B_{2(3,4,2',4')} - A_{1(2;2')} B_{2(3,4,1',4')} + A_{1(2;4')} B_{2(3,4,1',2')} \\
- A_{1(2;2')} B_{2(3,4,1',4')} + A_{1(3;2')} B_{2(2,4,1',4')} - A_{1(3;1')} B_{2(2,4,2',4')} \\
+ A_{1(4;1')} B_{2(2,3,2',4')} - A_{1(4;2')} B_{2(2,3,1',4')} - A_{1(3;4')} B_{2(2,4,1',2')} \right] \\
+ A_{1(1;4')} \left[ A_{1(2;2')} B_{2(3,4,1',3')} - A_{1(2;3')} B_{2(3,4,1',2')} + A_{1(3;1')} B_{2(2,4,3',3')} \right. \\
- A_{1(3;2')} B_{2(2,4,1',3')} + A_{1(3;3')} B_{2(2,4,1',2')} - A_{1(4;1')} B_{2(2,3,3',3')} \\
+ A_{1(4;2')} B_{2(2,3,1',3')} - A_{1(4;3')} B_{2(2,3,1',2')} - A_{1(2;1')} B_{2(3,4,2',3')} \right\} \right. \\
+ \left. A_{1(1;2')} \left[ A_{1(3;3')} B_{2(2,4,3',4')} - A_{1(3;3')} B_{2(1,4,2',4')} + A_{1(3;4')} B_{2(1,4,2',3')} \\
- A_{1(4;2')} B_{2(1,3,3',4')} + A_{1(4;3')} B_{2(1,3,2',4')} - A_{1(4;4')} B_{2(1,3,2',3')} \right] \\
+ \left. A_{1(2;2')} \left[ A_{1(3;3')} B_{2(2,4,1',4')} - A_{1(3;1')} B_{2(1,4,3',4')} + A_{1(4;4')} B_{2(1,3,1',3')} \\
- A_{1(3;4')} B_{2(1,4,1',3')} + A_{1(4;1')} B_{2(1,3,3',4')} - A_{1(4;3')} B_{2(1,3,3',4')} \right] \\
+ \left. A_{1(2;3')} \left[ A_{1(3;3')} B_{2(1,4,2',4')} - A_{1(3;2')} B_{2(1,4,3',4')} + A_{1(4;2')} B_{2(1,3,1',4')} \\
- A_{1(4;1')} B_{2(1,3,2',4')} + A_{1(3;4')} B_{2(1,4,1',3')} - A_{1(4;4')} B_{2(1,3,1',2')} \right] \\
+ \left. A_{1(1;4')} \left[ A_{1(3;2')} B_{2(1,4,1',3')} - A_{1(3;3')} B_{2(1,4,2',3')} + A_{1(4;1')} B_{2(1,3,2',3')} \\
- A_{1(3;3')} B_{2(1,4,1',2')} + A_{1(4;3')} B_{2(1,3,1',2')} - A_{1(4;2')} B_{2(1,3,1',3')} \right] \right\}
\]

\(440\)
and

\[ Y = A_1(3;1') \left[ A_1(4;2') B_2(1;2;3',4') - A_1(4;3') B_2(1;2;2',4') + A_1(4;4') B_2(1;2;2',3') \right] \]

\[ + A_1(3;2') \left[ A_1(4;3') B_2(1;2;1',4') - A_1(4;4') B_2(1;2;2',4') - A_1(4;1') B_2(1;2;1',3') \right] \]

\[ + A_1(3;3') \left[ A_1(4;1') B_2(1;2;2',4') - A_1(4;2') B_2(1;2;1',4') + A_1(4;4') B_2(1;2;1',2') \right] \]

\[ + A_1(3;4') \left[ A_1(4;2') B_2(1;2;1',3') - A_1(4;3') B_2(1;2;1',2') - A_1(4;1') B_2(1;2;1',3') \right]. \]  

(B.9)

Next assume that \( C_p(1,...,p;1',...,p') \) and \( C'_p(1,...,p;1',...,p') \) are antisymmetric in only their unprimed and primed coordinates, respectively.

\[ (A_1 \wedge C_2)(1,2,3,1',2',3') \]

\[ = \frac{1}{12} \left\{ A_2(1;1') \left[ C_2(2,3,1',2') - C_2(2,2,3,1') \right] + A_2(2;1) \left[ C_2(2,3,1',2') - C_2(2,3,2,1') \right] \right\} + \frac{1}{12} \{(B.10)\} \]

\[ + A_1(3;1') \left[ C_2(1,2,2',3') - C_2(1,2,3',2') \right] + A_1(3;2') \left[ C_2(1,2,3',1') - C_2(1,2,2',1') \right] \]

\[ + A_1(3;3') \left[ C_2(1,2,1',2') - C_2(1,2,2',1') \right]. \]

\[ (A_1 \wedge C'_2)(1,2,3,1',2',3') \]

\[ = \frac{1}{12} \left\{ A_2(1;1') \left[ C'_2(2,3,1',2') - C'_2(3,2,2,1') \right] + A_2(2;1) \left[ C'_2(3,2,1',2') - C'_2(2,3,1',3') \right] \right\} + \frac{1}{12} \{(B.11)\} \]

\[ + A_1(3;1') \left[ C'_2(1,2,1',2') - C'_2(1,2,2',1') \right] + A_1(3;2') \left[ C'_2(2,1,1',3') - C'_2(2,1,1',2') \right] \]

\[ + A_1(3;3') \left[ C'_2(1,2,1',2') - C'_2(2,1,1',2') \right]. \]
Appendix C

Algebraic form of ICSE(2)

The connected equation ICSE(2) has the form $\Omega^c_{2}(1,2;1',2') \equiv 0$, where $\Omega^c_2$ is the connected part of the second-order energy density matrix. Equation (3.45) gives this kernel in a compact, diagrammatic form that arises conveniently from the diagrammatic manipulations in Sections 3.3.1 and 3.3.2. This is the only form of ICSE(2) that appears in our original, published version of the connected equations.¹ Some time after the appearance of that paper, a connected version of CSE(2) was also published by Nooijen and co-workers.²

In this Appendix we present an algebraic formula for the same equation, obtained from Eq. (3.45) according to the diagram rules introduced in Section 3.3.1. For pedagogical purposes, and owing also to the complexity of the result, we have broken up Eq. (3.45) line-by-line, making each line a separate equation below. Moreover, in what follows certain obvious factorizations are bypassed in the algebraic formulation, in order that diagrams on the left side of the equality match up with algebraic expressions on the right side in the same order. This allows the reader more easily to inspect the connection between the diagrammatic and the algebraic equations, and to verify
the veracity of the latter. Altogether, the kernel $\Omega_2^{(1,2;1',2')}$ is the sum of the terms given in the equations below. Recall that $\hat{P}_2$ equals the identity operator minus the transposition of 1 and 2, with $\hat{P}_2'$ being the analogous operation on the coordinates $1'$ and $2'$.

C.1 Algebraic translation of the diagrammatic equation

\[
\begin{align*}
3 \hat{P}_2' \left\{ & -2 \begin{array}{c} \includegraphics{figure1} \\ \includegraphics{figure2} \end{array} + \frac{1}{2} \begin{array}{c} \includegraphics{figure3} \\ \includegraphics{figure4} \end{array} + \frac{1}{2} \begin{array}{c} \includegraphics{figure5} \\ \includegraphics{figure6} \end{array} + \begin{array}{c} \includegraphics{figure7} \\ \includegraphics{figure8} \end{array} \right\} \\
= & \hat{P}_2' \left\{ -2 \begin{array}{c} \includegraphics{figure1} \\ \includegraphics{figure2} \end{array} + \frac{1}{2} \begin{array}{c} \includegraphics{figure3} \\ \includegraphics{figure4} \end{array} + \frac{1}{2} \begin{array}{c} \includegraphics{figure5} \\ \includegraphics{figure6} \end{array} + \begin{array}{c} \includegraphics{figure7} \\ \includegraphics{figure8} \end{array} \right\} \\
= & \frac{1}{2} \hat{g}(1,2) \Delta_4^{(1;1')} \Delta_1^{(2;2')} - 2 \text{tr}_{3,4} \left[ \hat{g}(3,4) \Delta_2^{(1,2';3')} \Delta_2^{(3,4;1')} \right] \\
+ & \frac{1}{2} \text{tr}_{3,4} \left[ \hat{g}(3,4) \Delta_2^{(1,2';3')} \Delta_1^{(3;1')} \Delta_1^{(4;2')} \right] \\
+ & \text{tr}_{3,4} \left[ \hat{g}(3,4) \Delta_2^{(1,2';3')} \left( \Delta_1^{(3;4')} \Delta_1^{(4;2')} - \Delta_1^{(3;2')} \Delta_1^{(4;4')} \right) \right].
\end{align*}
\]
\[
\hat{P}_2 \hat{P}_2 \left\{ \begin{array}{c}
\begin{array}{c}
\begin{array}{c}
\psi^r \rangle
\end{array}
\end{array}
- \begin{array}{c}
\begin{array}{c}
\begin{array}{c}
\psi^r \langle
\end{array}
\end{array}
\end{array}
+ \begin{array}{c}
\begin{array}{c}
\begin{array}{c}
\psi^r \langle
\end{array}
\end{array}
\end{array}
\end{array}
\right\}
\right. \\
= \hat{P}_2 \hat{P}_2 \left\{ \begin{array}{c}
\begin{array}{c}
\begin{array}{c}
\psi^r \rangle
\end{array}
\end{array}
- \begin{array}{c}
\begin{array}{c}
\begin{array}{c}
\psi^r \langle
\end{array}
\end{array}
\end{array}
+ \begin{array}{c}
\begin{array}{c}
\begin{array}{c}
\psi^r \langle
\end{array}
\end{array}
\end{array}
\end{array}
\right\}
\right. \\
\text{tr}_3 \left[ \hat{g}_{(1,3)} \Delta_2(2,3;2') \Delta_1(1;2') \right] - \text{tr}_3 \left[ \hat{g}_{(1,3)} \Delta_2(1,3';2') \Delta_1(3;1') \right]
\right.
\left. - \text{tr}_3 \left[ \hat{g}_{(3,4)} \Delta_2(1,3;3') \Delta_1(2;4') \Delta_1(4;2') \right]
\right.
\left. + \text{tr}_3 \left[ \hat{g}_{(3,4)} \Delta_2(1,3;3') \Delta_1(2;3') \Delta_1(4;2') \right]
\right.
\left. - \frac{1}{2} \text{tr}_3 \left[ \Delta_1(1;1') \Delta_1(2;3') \Delta_1(3;2') \right] \right\}.
\]

\[
\hat{P}_2 \left\{ \begin{array}{c}
\begin{array}{c}
\begin{array}{c}
\psi^r \rangle
\end{array}
\end{array}
\right. \\
\right. \\
= \hat{P}_2 \left\{ \begin{array}{c}
\begin{array}{c}
\begin{array}{c}
\psi^r \rangle
\end{array}
\end{array}
\right. \\
\right. \\
\text{tr}_3 \left[ \hat{g}_{(3,4)} \Delta_2(1,2;3',4') \Delta_1(3;1') \Delta_1(4;2') \right]
\right.
\left. + \text{tr}_3 \left[ \hat{g}_{(3,4)} \Delta_2(1,3;1',2') \left( \Delta_1(4;3') \Delta_1(2;4') - \Delta_1(2;3') \Delta_1(4;2') \right) \right]
\right.
\left. - \text{tr}_3 \left[ \hat{h}(3) \Delta_2(1,3;1',2') \Delta_1(2;3') \right] - \text{tr}_3 \left[ \hat{h}(3) \Delta_2(1,2;1',3') \Delta_1(3;2') \right]
\right.
\left. - \text{tr}_3 \left[ \hat{g}(2,3) \Delta_2(1,3;1',2') \Delta_1(2;3') \right] \right\}.
\]

(C.4)

(C.5)
\[
\hat{P}_2 \left\{ - \frac{\hat{r}_2}{r_2} + \frac{1}{2} \frac{\hat{r}_2}{r_2} + 2 \frac{\hat{r}_2}{r_2} \hat{g}(2,1) - 2 \frac{\hat{r}_2}{r_2} \right\} \\
= \hat{P}_2 \left\{ - \text{tr}_3 \left[ \hat{g}(1,3) \Delta_2(1,3;1',2') \Delta_4(2,3') \right] \\
+ \frac{1}{2} \text{tr}_{3,4} \left[ \hat{g}(3,4) \Delta_4(1,1') \Delta_1(2,4') \Delta_1(3,1') \Delta_1(4,2') \right] \\
+ 2 \text{tr}_{3,4} \left[ \hat{g}(3,4) \Delta_2(1,3;1',3') \Delta_2(4,2;4',3') \right] \\
- 2 \text{tr}_{3,4} \left[ \hat{g}(3,4) \Delta_2(1,3;1',2) \Delta_2(2,4,3';4') \right] \\
- 2 \text{tr}_{3,4} \left[ \hat{g}(3,4) \Delta_2(1,3;1',4') \Delta_2(4,2;3',2') \right] \right\}.
\]

C.2 References


Appendix D

Reduced density matrices from multideterminant wavefunctions

Although our stated purpose is to calculate RDMs (especially $D_2$) in the absence of an electronic wavefunction, a discussion of RDMs would be incomplete in some sense without some comments on how these are obtained, at a practical level, from correlated (that is, multi-determinantal) wavefunctions. Furthermore, in order to test certain approximation schemes for direct determination of the RDMs, we will want to utilize wavefunction data. Thus in this Appendix we work out expressions that relate the RDM matrix elements to the coefficients in a determinantal expansion of the wavefunction. This is a textbook-level exercise—although, remarkably, the author has not seen these expressions in any textbook (but see Ref. 1)—and is similar to the derivation of the Slater rules for evaluating determinantal matrix elements of one- and two-electron operators. The Slater rules are derived, for example, in §2.3.4 of Szabo and Ostlund’s textbook.2
D.1 Density matrix formulas

Our starting point is the electronic wavefunction

$$|\Psi\rangle = \sum_K C_K|\Phi_K\rangle = \sum_{k_1 < k_2 < \cdots < k_N} C_{k_1,\ldots,k_N} |\phi_{k_1}\phi_{k_2} \cdots \phi_{k_N}\rangle \quad (D.1)$$

expanded by normalized, $N$-electron determinants $|\Phi_K\rangle = |\phi_{k_1}\phi_{k_2} \cdots \phi_{k_N}\rangle$ of orthonormal spin orbitals. According to this notation, $K = (k_1,\ldots,k_N)$ is an ordered $N$-tuple of spin-orbital indices arranged in ascending order. We will write $n \in K$ to indicate that $n$ is equal to one of the $k_i$, that is, $\phi_n$ is occupied in $|\Phi_K\rangle$. In contrast to much of the rest of this text, in this Appendix we employ antisymmetrized product kets rather than direct product kets (see the discussion in Section 2.3.1). In other words

$$\Phi_K^{(1,\ldots,N)} = \frac{1}{\sqrt{N!}} \hat{P}_N\left\{\phi_{k_1}^{(1)} \cdots \phi_{k_N}^{(N)}\right\}. \quad (D.2)$$

The RDMs are obtained as contractions of $|\Psi\rangle\langle\Psi|$. In particular,

$$D_2^{(1,2;1',2')} = \binom{N}{2} \sum_{JK} C_J C_K^* \int dx_3 \cdots dx_N dx_3' \cdots dx_N' \delta(x_3 - x_3') \cdots \delta(x_N - x_N')$$

$$\times \frac{1}{N!} \hat{P}_N\left\{\phi_{j_1}^{(1)} \cdots \phi_{j_N}^{(N)}\right\} \hat{P}_N'\left\{\phi_{k_1}^{(1')} \cdots \phi_{k_N}^{(N')}\right\}. \quad (D.3)$$

The integral in this expression, including the factor of $N!^{-1}$, will be abbreviated as $\text{tr}_{3,\ldots,N}(\Phi_J^*\Phi_K)$. It is helpful to write out this integral in terms of permutations of the spin-orbital indices:

$$\text{tr}_{3,\ldots,N}(\Phi_J^*\Phi_K) = \frac{1}{N!} \int dx_3 \cdots dx_N \sum_{\sigma,\pi \in S_N} \epsilon(\sigma) \epsilon(\pi) \tilde{\sigma} \left\{\phi_{j_1}^{(1)} \cdots \phi_{j_N}^{(N)}\right\}$$

$$\times \hat{\pi} \left\{\phi_{k_1}^{(1')} \phi_{k_2}^{(2')} \phi_{k_3}^{(3')} \cdots \phi_{k_N}^{(N')}\right\}. \quad (D.4)$$

We will evaluate the above integral in detail to obtain the expansion coefficients for $D_2$ in terms of those for $|\Psi\rangle$. This derivation will suggest how to obtain the other
RDMs. Since the $\phi_k$ are orthonormal, the integral in Eq. (D.3) will vanish if the two determinants in the integrand differ by more than two spin-orbitals. (What we mean by “differ” is clarified below; the term is used in the sense of the Slater rules.) Thus we must consider cases where $|\Phi_J\rangle$ and $|\Phi_K\rangle$ differ by zero, one, or two spin-orbitals.

Consider first the case where the two determinants are identical. Two indices from each determinant survive the integration; the remaining $N - 2$ indices of each must have the same order, else the integral vanishes. There are $(N - 2)!$ permutations in $S_N \otimes S_N$ for which this is the case. Following integration, we obtain all permutations of the four surviving indices, each of which may equal any of the indices $k_1, \ldots, k_N$ (though no two are equal). Thus

$$\text{tr}_{3, \ldots, N} (\Phi_K \Phi_K^*) = \frac{(N - 2)!}{N!} \sum_{m, n \in K} \left[ \phi_m(1) \phi_m(2) \phi_n^*(1') \phi_n^*(2') - \phi_m(1) \phi_n(2) \phi_n^*(1') \phi_n^*(2') \\
- \phi_n(1) \phi_m(2) \phi_{m'}^*(1') \phi_n^*(2') + \phi_n(1) \phi_m(2) \phi_{m'}^*(1') \phi_n^*(2') \right]. \quad (D.5)$$

Up to a normalization factor, the term in brackets is a product of two $2 \times 2$ Slater determinants, one for the primed coordinates and one for the unprimed coordinates. In Dirac notation, a normalized $2 \times 2$ determinant of $\phi_m$ and $\phi_n$ is written $|\phi_m \phi_n\rangle$. Thus we have

$$\text{tr}_{3, \ldots, N} (|\Phi_K\rangle\langle\Phi_K|) = \frac{2}{N(N - 1)} \sum_{m, n \in K} |\phi_m \phi_n\rangle\langle\phi_m \phi_n|, \quad (D.6)$$

which is a diagonal element of $D_2$. When this expression is combined with Eq. (D.3), one finds that the coefficient of the two-electron determinantal dyad $|\phi_m \phi_n\rangle\langle\phi_m \phi_n|$ equals $C_J C_K^*$, which is precisely the coefficient of the $N$-electron determinantal dyad $|\Phi_J\rangle\langle\Phi_K|$. 


Next is the case where $|\Phi_J\rangle$ and $|\Phi_K\rangle$ differ by exactly one spin-orbital, say,

$$|\Phi_J\rangle = \hat{a}_m^\dagger \hat{a}_n |\Phi_K\rangle .$$  (D.7)

We denote the indicated relationship between $|\Phi_J\rangle$ and $|\Phi_K\rangle$ by means of the set-theoretic notation $J \setminus \{m\} = K \setminus \{n\}$. We may also write

$$|\Phi_J\rangle = \theta_m |\phi_m \phi_a \phi_b \cdots \rangle \quad \text{(D.8a)}$$

$$|\Phi_K\rangle = \theta_n |\phi_n \phi_a \phi_b \cdots \rangle ,$$  (D.8b)

in which $\theta_m = \pm 1$, for example, equals the sign of the permutation necessary to move $\phi_m$ to the left of the index list $a < b < \cdots < m < \cdots$ that characterizes the determinant $|\Phi_J\rangle$. The determinants in Eq. (D.4) are $\hat{\sigma}\{\phi_{m(1)} \phi_{a(2)} \phi_{b(3)} \cdots \}$ and $\hat{\pi}\{\phi_{n(1)} \phi_{a(2)} \phi_{b(3)} \cdots \}$. $\phi_m$ and $\phi_n$ must survive the integration in $\text{tr}_{3 \ldots N}(\Phi_K \Phi_J)$. Following the same reasoning as in the first case, what remains following integration is a sum of $2 \times 2$ determinants, composed in this case of $\phi_m$, $\phi_n$, and one other spin-orbital $\phi_a$, where $a \in J \cap K$. Thus

$$\text{tr}_{3 \ldots N}(\langle |\Phi_J\rangle |\Phi_K\rangle)_{J \setminus \{m\} = K \setminus \{n\}} = \frac{2 \theta_m \theta_n}{N(N - 1)} \sum_{a \in J \cap K} |\phi_m \phi_a \rangle \langle \phi_m \phi_a | .$$  (D.9)

The subscript “$J \setminus \{m\} = K \setminus \{n\}$” that is attached to $\text{tr}_{3 \ldots N}(\langle |\Phi_J\rangle |\Phi_K\rangle)$ in this equation indicates that this result is valid when $|\Phi_J\rangle$ and $|\Phi_K\rangle$ are related as in Eq. (D.7). ($|\Phi_J\rangle$ is obtained from $|\Phi_K\rangle$ by annihilating an electron in $\phi_n$ and creating an electron in $\phi_m$.) The forms of these determinants fix the values of $m$ and $n$, whereas $a$ is a dummy index. As in the previous case, the coefficient that multiplies the contribution from Eq. (D.9) in the expansion of $\hat{D}_2$ is $C_J C_K^*$. Lastly there is the case where $|\Phi_J\rangle$ and $|\Phi_K\rangle$ differ by exactly two spin orbitals,
say

\[ |\Phi_J\rangle = \theta_{mn}|\phi_m \phi_n \phi_a \phi_b \cdots \rangle \quad (D.10a) \]
\[ |\Phi_K\rangle = \theta_{pq}|\phi_p \phi_q \phi_a \phi_b \cdots \rangle , \quad (D.10b) \]

which we abbreviate as \( J \setminus \{m, n\} = K \setminus \{p, q\} \). It is assumed that \( m, n \in J \) and \( p, q \in K \), but \( m, n \notin K \) and \( p, q \notin J \) (hence \( |\Phi_J\rangle \) and \( |\Phi_K\rangle \) really do differ by exactly two spin orbitals). Again \( \theta_{mn} = \pm 1 \) is a phase factor. Examining Eq. (D.4) for this case, we see that \( \phi_m, \phi_n, \phi_p, \) and \( \phi_q \) must all survive the integration, which leaves us with no free indices for the resulting \( 2 \times 2 \) determinants. As such,

\[ \text{tr}_{3, \ldots, N} \langle |\Phi_J\rangle|\Phi_K\rangle_{J \setminus \{m, n\} = K \setminus \{p, q\}} = \frac{2 \theta_{mn} \theta_{pq}}{N(N-1)} |\phi_m \phi_n \rangle \langle \phi_p \phi_q| . \quad (D.11) \]

Equations (D.6), (D.9), and (D.11) specify all the nonvanishing terms in the expansion of the 2-RDM, Eq. (D.3). The techniques used in this argument readily generalizes to the case of the \( p \)-RDM, with the partial integration \( \text{tr}_{p+1, \ldots, N} (\Phi_K^* \Phi_J) \) contributing dyads of \( p \)-electron determinants whose form can be inferred by examining how the determinants \( |\Phi_J\rangle \) and \( |\Phi_K\rangle \) differ. Let us provide just one example of this generalization, for the case that \( J = K \setminus \{m\} \), meaning that \( |\Phi_J\rangle \) and \( |\Phi_K\rangle \) differ by exactly one spin orbital, as in Eq. (D.7). In this case the contribution to \( \hat{D}_3 \) is

\[ \text{tr}_{4, \ldots, N} (\Phi_K^* \Phi_J)_{J \setminus \{m\} = K \setminus \{n\}} = \frac{6 \theta_m \theta_n}{N(N-1)(N-2)} \sum_{a, b \in J \setminus K} |\phi_m \phi_a \phi_b \rangle \langle \phi_n \phi_a \phi_b| . \quad (D.12) \]

In analogy to the 2-RDM case, the combinatorial factor here equals \( \binom{N}{3}^{-1} \), which cancels the factor of \( \binom{N}{3} \) included in \( D_3 \), when the normalization \( \text{tr} D_3 = \binom{N}{3} \) is employed.


D.2 References


Appendix E

Atomic integrals over Slater-type functions

This appendix provides formulas for one-center, one- and two-electron integrals over primitive Slater-type functions

\[ S_{jl}(r, \theta, \varphi) \equiv C_{jl} \ r^{l+n_j} e^{-\zeta_l} Y_{lm}(\theta, \phi). \]  

(E.1)

The factor

\[ C_{jl} = \left[ \frac{(2\zeta_j)^{l+2n_j+3}}{(2l+2n_j+2)!} \right]^{1/2} \]  

(E.2)

makes these functions unit-normalized.

E.1 One-electron integrals

In the case of the kinetic energy integrals, one may avoid performing any explicit angular integration by recognizing that the kinetic energy operator \( \hat{T} = (-1/2\mu)\nabla^2 \) can be expressed in terms of the angular momentum operator \( \hat{L}^2 \):

\[ \hat{T} = -\frac{1}{2\mu r^2} \left[ \frac{\partial}{\partial r} \left( r^2 \frac{\partial}{\partial r} \right) - \hat{L}^2 \right]. \]  

(E.3)

This simplifies matters considerably. Since \( \hat{L}^2 Y_{lm} = l(l+1)Y_{lm} \), the kinetic energy integrals are diagonal in the angular quantum numbers \( l \) and \( m \). The nonzero integrals
are found to be

\[ \langle S_{klm} | \hat{T} | S_{jlm} \rangle = \frac{C_{jl} C_{kl} (2l + n_j + n_k)!}{2\mu(\zeta_{jl} + \zeta_{kl})^{2l+n_j+n_k+1}} \left[ -n_j(2l + n_j + 1) \\
+ \frac{2\zeta_{jl}(l + n_j + 1)(2l + n_j + n_k + 1)}{\zeta_{jl} + \zeta_{kl}} \\
- \frac{\zeta_{jl}^2 (2l + n_j + n_k + 1) (2l + n_j + n_k + 2)}{(\zeta_{jl} + \zeta_{kl})^2} \right]. \tag{E.4} \]

A similar exercise yields the electron-nucleus potential energy integrals,

\[ \langle S_{klm} | -Z/r | S_{jlm} \rangle = \frac{-ZC_{jl} C_{kl} (2l + n_j + n_k + 1)!}{(\zeta_{jl} + \zeta_{kl})^{2l+n_j+n_k+2}}. \tag{E.5} \]

### E.2 Two-electron integrals

To develop the two-electron integrals we make use of the identities\(^1\)

\[ \frac{1}{r_{12}} = \frac{1}{\sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos \omega}} = \sum_{k=0}^{\infty} \frac{r_k^k}{r_{k+1}^k} P_k(\cos \omega) \tag{E.6} \]

and

\[ P_l(\cos \omega) = \frac{4\pi}{2l+1} \sum_{m=-l}^{l} Y_{l,m}^*(\theta_1, \varphi_1) \ Y_{l,m}(\theta_2, \varphi_2). \tag{E.7} \]

Here \(\omega\) is the angle between the electron position vectors \(\vec{r}_1\) and \(\vec{r}_2\), while

\[ r_\leq \overset{\text{def}}{=} \min\{r_1, r_2\} \tag{E.8a} \]

\[ r_\geq \overset{\text{def}}{=} \max\{r_1, r_2\} \tag{E.8b} \]

Upon expanding \(r_{12}^{-1}\) using these two identities, the requisite two-electron integrals can be written as

\[ \langle S_{i_1'i_1'm_1'}(\vec{r}_1) \ S_{i_2'i_2'm_2'}(\vec{r}_2) | r_{12}^{-1} | S_{i_1i_1m_1}(\vec{r}_1) \ S_{i_2i_2m_2}(\vec{r}_2) \rangle \]

\[ = C \sum_{k=0}^{k_{\text{max}}} \gamma_{l_1,m_1,i_1,m_1'}^{(k)} \gamma_{l_2,m_2,i_2,m_2'}^{(k)} R^{(k)}(\alpha_1 + 2, \alpha_2 + 2, \beta_1, \beta_2) \delta_{m_1+m_2,m_1'+m_2'}. \tag{E.9} \]
in which we have introduced the abbreviations

\[ k_{\text{max}} \overset{\text{def}}{=} \min\{ l_1 + l'_1, l_2 + l'_2 \} , \]  
\[ C \overset{\text{def}}{=} C_{i l_1} C_{i l_2} C_{i' l_1} C_{i' l_2} , \]  
\[ \alpha_\kappa \overset{\text{def}}{=} n_\kappa + n_{\kappa'} + l_\kappa + l_{\kappa'} , \]

and

\[ \beta_\kappa \overset{\text{def}}{=} \zeta_{\kappa l_\kappa} + \zeta_{\kappa' l_{\kappa'}} . \]

In Eq. (E.9), the results of the angular integration are contained in the Gaunt coefficients \( \gamma_{l_1, l'_1; l_2, l'_2}^{(k)} \), which were defined in Eq. (2.105). The triangle rule for these coefficients is responsible for truncating the sum over \( k \) in Eq. (E.9).

What remains to be evaluated in Eq. (E.9) are the Slater-Condon parameters

\[ R^{(k)}(n_1, n_2, \zeta_1, \zeta_2) \overset{\text{def}}{=} \int_0^\infty dr_1 \int_0^\infty dr_2 \left( \frac{r_<^k}{r_>^{k+1}} \right) r_1^{n_1} r_2^{n_2} e^{-\zeta_1 r_1 - \zeta_2 r_2} . \]  
\[ (E.14) \]

(The volume factors \( r_1^2 \) and \( r_2^2 \) are absorbed into the exponents \( n_1 \) and \( n_2 \).) These, too, can be expressed analytically:

\[ R^{(k)}(n_1, n_2, \zeta_1, \zeta_2) = \frac{(n_1 + n_2 - 1)!}{\zeta_1 \zeta_2 (\zeta_1 + \zeta_2)^{n_1 + n_2 - 1}} \left[ 1 + \mathcal{E}_{n_1 + n_2 - 1}^{n_1} (\zeta_1 / \zeta_2) + \mathcal{E}_{n_1 + n_2 - 1}^{n_1} (\zeta_2 / \zeta_1) \right] , \]  
\[ (E.15) \]

where

\[ \mathcal{E}_k^n(x) \overset{\text{def}}{=} \frac{\sum_{j=0}^{k-1} \binom{n}{j} x^j}{\binom{n}{k} x^k} = k \frac{1 + \mathcal{E}_{k-1}^{n-1}(x)}{(n - k + 1)x} . \]  
\[ (E.16) \]

E.3 References
