Variational optimization of second order density matrices for electronic structure calculation

Helen van Aggelen

Supervisor: Prof. Dr. Patrick Bultinck
Co-supervisor: Prof. Dr. Dimitri Van Neck

Dissertation submitted in fulfillment
of the requirements for the degree of
Doctor (Ph.D.) in Sciences: Chemistry

October 10, 2011

Faculty of Sciences
Department of Inorganic and Physical Chemistry

UNIVERSITEIT GENT

This work was supported by the Special Research Fund (BOF) and the Research Foundation - Flanders (FWO). Computational resources were kindly provided by the Flemish Supercomputer Center (VSC).

Abstract

The exponential growth of the dimension of the exact wavefunction with the size of a chemical system makes it impossible to compute chemical properties of large chemical systems exactly. A myriad of ab initio methods that use simpler mathematical objects to describe the system has thrived on this realization. These methods avoid hitting the exponential wall by using low order densities or density matrices. Density Functional and first order density matrix methods have gained significant popularity, but can only approximate the relationship between the energy and the density or density matrix. Second-order density matrix methods take a special place in the hierarchy of these methods because the second order density matrix (2DM) exactly and explicitly determines the energy.

As was already realized in the 1950's, the most straightforward way to derive a 2DM for a chemical system from scratch simply applies the variational principle to it. In the 1990's, progress in semidefinite optimization techniques revived this idea.

The aim of my thesis has been to evaluate the use of variational second order density matrix (v2DM) methods for chemistry and to identify the major theoretical and computational challenges that need to be overcome to make it successful for chemical applications. This research has led to the following conclusions.

The theoretical challenges that the method faces follow from the need for

the 2DM to be N-representable. After all, even if the method does not make any reference to a wavefunction for the N-electron system, the 2DM still needs to be derivable from an ensemble of N-electron states. Failure of the 2DM to be N-representable is reflected in a too low energy. But even though the variational procedure focuses on the energy, the starting point of our research has been to look at chemical properties other than the energy. This can be motivated by a simple observation: even when the energy is constrained to be exact, variational optimization of the 2DM under approximate N-representability constraints may lead to a wrong 2DM and therefore incorrect chemical properties.

We have identified several problems when commonly used N-representability constraints are applied to chemical problems. First of all, low order positivity conditions generally lead incorrectly to fractionally charged dissociation products. This phenomenon can be explained by the method's failure to represent the ensemble of states from which systems with fractional charges must arise. This finding is numerically illustrated for, for instance, NO^+ in this work. Secondly, commonly used approximate N-representability conditions are not size-consistent. This can be illustrated clearly and explicitly for a system of two-electron non-interacting fragments under the P-condition. Although the P-condition is exact for any two-electron system, it is not exact for a system of several two-electron non-interacting fragments. It is shown that the resulting energy for such a system originates completely from the one-electron terms in the Hamiltonian which leads to an incorrect structure of both the 1DM and 2DM. We have derive constraints on the energy of subspaces of the one-particle basis space that solve these problems, albeit in an ad-hoc manner.

Another topic that needs to be further explored is the description of molecular spin in the v2DM method. Because the 2DM only carries information up to two-electron interactions, ensuring that it represents a proper spin state is a difficult problem. We have applied spin conditions derived from a pure spin state wavefunction and more general conditions that allow the 2DM to describe an ensemble of mixed spin states with a fixed $\hat{S}^2$ eigenvalue. Two major shortcomings of these conditions applied to the v2DM method are false multiplet

splitting and size-inconsistency. These spin conditions are less strong applied to a system of non-interacting fragments than when they are applied to these fragments separately. These problems are not specific to the v2DM method - in fact, they turn up in all methods based on low order densities and density matrices, although they take different forms in different theories. Understanding these problems is therefore of fundamental importance.

The computational challenges that the method faces derive from its formulation as a vast semidefinite optimization problem under generalized inequality constraints on the 2DM. Because the dimension of the 2DM is quadratic in the dimension of the single particle basis set, and typical basis sets used in chemistry include up to a few hundreds of basis functions, the dimension of a typical 2DM surpasses that of a standard semidefinite optimization problem in mathematics. Although much progress has been made in the field of semidefinite programming from the 1990's on, the computational scaling of typical algorithms applied to the 2DM remains prohibitive. We have implemented and compared four different semidefinite optimization algorithms for the v2DM method, in which we exploited the specific structure of the problem. Even so, none of the algorithms performed significantly better than the others. Three of the algorithms we tried were so-called second order methods, related to the barrier method, which employ the gradient and hessian of the problem, and one of them, the boundary point method, was a zeroth order method, which does not use a gradient nor hessian. Remarkably, all of these methods performed more or less similar, with only minor trade-offs between speed, accuracy and robustness. Moreover, the maximal system size our programs can handle is comparable to that of other implementations used in the literature, such as the first-order non-linear method applied by David Mazziotti. This suggests that the origin of the slow convergence of v2DM methods, the singularity of the optimal 2DM, manifests itself in all of these methods, even though it is most explicit in the barrier method. Finding a way to deal with the ill-conditioned equations will be the key factor to making this method workable.

To conclude, significant progress in both the aforementioned theoretical and computational aspects is needed to make the v2DM method competitive to comparable wavefunction based methods. The theoretical challenges that follow from the N-representability problem are fundamentally different to the computational challenges, because in theory, the exact N-representability constraints for each system are available through full configuration interation calculations. The problem is *only* to generalize them to all molecular systems. The computational challenge is less straightforward, because it requires developing new algorithms, which is a rather empirical field of research. Only trial calculations can really prove a new method's success or failure.

Nonetheless, if we find ways to overcome these challenges, the v2DM method will prove a valuable alternative to wavefunction based methods. It is highly complementary to wavefunction based methods, because of its fundamentally different approach to solving the electron correlation problem, independent from any reference system. Herein lies its strength and its future.

Samenvatting

De exponentiële groei van de dimensie van de golffunctie met de grootte van een chemisch systeem maakt het onmogelijk om chemische eigenschappen van grote chemische systemen exact te berekenen. Een verscheidenheid aan ab initio methoden die eenvoudigere wiskundige objecten gebruiken om het systeem te beschrijven gedijen op deze vaststelling. Ze vermijden de 'exponentiële muur' door lage orde densiteiten en densiteitsmatrices te gebruiken. Densiteits Functionaal en eerste orde densiteits matrix methoden hebben aanzienlijke populariteit verworven, maar kunnen de relatie tussen de energie en de densiteit of densiteitsmatrix enkel benaderen. Tweede orde densiteitsmatrix methoden nemen een bijzondere plaats in de hiërarche van deze methoden omdat de tweede orde densiteitsmatrix (2DM) de energie exact en expliciet bepaalt.
Het is al bekend vanaf de jaren 1950 dat de meest voor de hand liggende methode om een 2DM voor een chemisch systeem te bepalen eenvoudigweg het variationele principe erop toepast. Dit idee bloeide opnieuw op in de jaren 1990 door vooruitgang in semidefiniete optimalisatietechnieken.

Het doel van mijn thesis was om het gebruik van variationele tweede orde densiteitsmatrix (v2DM) methoden voor chemische doeleinden te beoordelen en om de belangrijkste theoretische en computationele uitdagingen te identificeren die overwonnen moeten worden om de methode succesvol te maken voor chemische toepassingen. Dit werk heeft tot de volgende conclusies geleid.

De theoretische uitdagingen waar de methode voor staat komen voort uit de

noodzaak dat de 2DM 'N-representabel' moet zijn. Zelfs als de methode geen referentie maakt naar een golffunctie, moet de 2DM immers af te leiden zijn uit een ensemble van N-elektron toestanden. Als de 2DM niet N-representabel is, levert hij een te lage energie op. Hoewel de variationele procedure zich enkel richt op de energie, was het uitgangspunt van ons onderzoek om naar andere chemische eigenschappen dan de energie te kijken. Dit idee kan gemotiveerd worden met een eenvoudige observatie: zelfs als de energie exact opgelegd wordt kan de variationele optimalisatie onder noodzakelijke maar niet voldoende voorwaarden een verkeerde 2DM opleveren, en dus ook verkeerde chemische eigenschappen.

We hebben verschillende problemen met veelgebruikte N-representabiliteitsvoorwaarden toegepast op chemische problemen aangekaart. Ten eerste leiden de lage orde positiviteitsvoorwaarden vaak incorrect tot fractioneel geladen dissociatieprodukten. Dit fenomeen kan verklaard worden door het onvermogen van de methode om het ensemble van toestanden te beschrijven waarvan systemen met een fractionele lading uit ontstaan. Deze vaststelling werd numeriek geïllustreerd voor, bijvoorbeeld, NO^+ in dit werk. Ten tweede zijn veelgebruikte N-representabiliteitsvoorwaarden over het algemeen niet 'size-consistent'. Dit kan expliciet geïllustreerd worden voor een systeem bestaande uit niet-interagerende twee-elektron fragmenten onderhevig aan de P-voorwaarde. Hoewel de P-voorwaarde exact is voor elk twee-elektron systeem, is het niet exact voor een systeem bestaande uit meerdere niet-interagerende twee-elektron fragmenten. We hebben aangetoond dat de energie voor zo'n systeem volledige voortkomt uit de één elektron termen van de Hamiltoniaan, hetgeen tevens leidt tot een 1DM en 2DM met een verkeerde structuur.

Een ander onderwerp dat voorlopig onderbelicht is in dit onderzoeksgebied, is de beschrijving van moleculaire spin in de v2DM methode. Omdat de 2DM enkel informatie bevat over n- en twee-electron interacties, is het een moeilijk probleem om ervoor te zorgen dat hij een correcte spin toestand voorstelt. We hebben verschillende spin voorwaarden afgeleid uit een zuivere spin toestand en meer algemene voorwaarden die de 2DM toelaten om een ensemble van gemengde toestanden met een welbepaalde $\hat{S}^2$ eigenwaarde voor te stellen. Twee belangrijke

nadelen deze voorwaarden toegepast op de v2DM methode zijn oneigenlijke splitsing van multipletten en size-inconsistency. Deze spin voorwaarden zijn minder sterk toegepast op een systeem van niet-interagerende fragmenten dan wanneer ze toegepast worden op deze fragmenten apart. Deze problemen zijn niet specifiek voor de v2DM methode - in feite duiken ze op in alle methoden die gebaseerd zijn op lage orde densiteiten en densiteitsmatrices, hoewel ze een verschillende vorm aannemen in verschillende theorieën. Deze problemen begrijpen is daarom van fundamenteel belang.

De computationele uitdagingen waar deze methode voor staat komen voort uit zijn formulering als een grootschalig semidefiniet optimalisatie probleem onderhevig aan veralgemeende ongelijkheidsvoorwaarden. Omdat de dimensie van de 2DM kwadratisch is in de dimensie van de eendeeltjes basis set, en basis sets in chemische toepassingen typisch een paar honderd basis functies bevatten, overtreft de dimensie van een typische 2DM dat van een standaard semidefiniet optimalisatieprobleem in de wiskunde. Hoewel aanzienlijke vooruitgang is geboekt in het domein van semidefiniete optimalisatie vanaf de jaren 1990 blijft de computationele kost van typische algoritmen toegepast op 2DM onoverkomelijk. We hebben verschillende semidefiniete optimalisatiealgoritmen voor de v2DM methode geïmplementeerd en vergeleken, waarbij we de specifieke structuur van het probleem in acht genomen hebben. Nochtans presteerde geen enkele van de algoritmen significant beter dan de anderen. Drie van deze algoritmen waren zogenaamde tweede orde methoden, verwant aan de barrière methode, die de gradiënt en Hessiaan van het probleem gebruiken. Opvallend genoeg presteerden ze gelijkaardig, met slechts kleine verschillen in snelheid, accuratesse en robuustheid. Bovendien is de maximale systeemgrootte die onze programma's aankunnen vergelijkbaar met die van andere implementaties gebruikt in de literatuur, zoals de eerste orde niet-lineaire methode toegepast door David Mazziotti. Dit suggereert dat de oorzaak voor de trage convergentie van v2DM methoden, de singulariteit van de optimale 2DM, zich in al deze methoden manifesteert, hoewel het het meest expliciet is in de barriere methode. Een manier vinden om

met de slecht geconditioneerde vergelijkingen om te gaan vormt de sleutel tot een praktisch werkbare methode.

Om te besluiten is aanzienlijke vooruitgang in de voorgenoemde theoretische en computationele aspecten nodig opdat de v2DM methode de concurrentie zou kunnen aangaan met vergelijkbare golffunctie gebaseerde methoden. De theoretische uitdagingen die voortvloeien uit de N-representabiliteitsvoorwaarde zijn fundamenteel verschillend van de computationele uitdagingen. In theorie zijn de exacte N-representabiliteitsvoorwaarden voor elk systeem immers toegankelijk via een 'full configuration interaction' berekening. Het probleem is echter om deze te veralgemenen, of ten minste efficiënt te automatiseren, naar alle moleculen toe. De computationele vraagstukken zijn weliswaar minder voor de hand liggend, omdat ze nieuwe algoritmen vereisen. Het ontwikkelen van nieuwe algoritmen is eerder een empirische zoektocht omdat enkel testberekeningen kunnen uitwijzen of de methode een succes is.

Niettemin, als we een manier vinden om deze uitdagingen succesvol aan te gaan, zal de v2DM methode een waardevol alternatief vormen voor golffunctie gebaseerde methoden. De methode is immers sterk complementair met golffunctie gebaseerde methoden, omwille van zijn fundamenteel verschillende aanpak om elektroncorrelatie te beschrijven, onafhankelijk van enig referentiesysteem. Hierin ligt zijn kracht en zijn toekomst.

Acknowledgements

I have had the exceptional privilege of having three supervisors whose experience in chemistry, physics, mathematics and in life has given me the variational freedom to develop a versatile mind. Most importantly, Patrick, you have given me the freedom I needed to find my way in science. Dimitri, I have much enjoyed your patience in explaining me complicated concepts and ideas from physics. Paul, although you get no official credit for being my unofficial co-supervisor, I greatly appreciate and admire how you treated me as one of your own students and welcomed me into your 'scientific family'.
Of course, I have not been alone in producing this work of research and I owe much gratitude to Brecht and Ward. Your enthusiasm and thoroughly different viewpoints on our shared research topic have been a constant motivation and inspiration.

My Ph.D. research has brought me across the ocean and back. Along the way, I have met remarkable people whom I will remember for sharing their viewpoints, hospitality and friendship. In particular, I sustain warm memories of the core theory group at Ghent University I started out with: Sofie, Elke, Stijn and Veerle, the unofficial member of the group, as well as the more recent members of the group. At McMaster, I have enjoyed the pleasant mix of cultures, especially the Latin American charms of Rogelio and Carlos.

I gratefully acknowledge the efforts of all members of the reading committee and jury, in particular David Cooper, in completing the inherently ungrateful

task of reading my thesis.

Mom and dad, thank you for your trust and support even when my scientific interests led me into obscure fields of research that I cannot even adequately describe to you.

Contents

List of abbreviations **i**

List of Figures **iii**

List of Tables **ix**

1 N-representability **1**

1.1 Introduction 1

1.2 Physical importance of the 2DM 2

1.2.1 N^{th} order density matrix 2

1.2.2 2^{nd} order density matrix 4

1.3 N-representability 8

1.3.1 Definition of N-representability 8

1.3.2 Necessary and sufficient conditions for N-representability 10

1.3.3 Necessary conditions implied by N-representability 13

1.4 Practical variational second order density matrix methods 23

1.5 Applications to chemistry 26

1.5.1 Computational details 27

1.5.2 Strengths and failures of 2-index N-representability constraints in chemical applications 27

1.5.3 Additional subspace energy constraints to correct molecular dissociation 47

1.5.4 Application of subspace energy constraints to poly-atomic molecules . . . 57
1.5.5 Size-consistency and separability under 2-index constraints 70

2 S-representability 83

2.1 Introduction . . . 83
2.2 Representation of electronic spin in the 2DM . . . 84
2.3 S-representability conditions: pure spin states . . . 88
2.3.1 Spin symmetry . . . 88
2.3.2 Basic S-representability constraints . . . 92
2.3.3 Relationship between first order density matrix and transition density matrix elements for different spin projections 95
2.3.4 S-representability constraints derived from relations between first order density and transition density matrix elements . . . 98
2.4 S-representability conditions: ensemble spin states . . . 99
2.4.1 Implications of spin symmetry on the structure of the 2DM 99
2.4.2 Basic S-representability constraints . . . 102
2.4.3 S-representability constraints derived from the Gutzwiller projection . . . 103
2.5 Applications . . . 104
2.5.1 Applied S-representability conditions . . . 105
2.5.2 Computational and algorithmic details . . . 106
2.5.3 Results on S-representability calculations . . . 107
2.6 Conclusions on describing spin in v2DM theory . . . 122

3 Semidefinite optimization of the 2DM 125

3.1 Introduction . . . 125
3.2 Basics . . . 126
3.3 Computational aspects . . . 133

3.3.1 Input and data storage 133
3.3.2 Feasible starting points for interior-point algorithms . . . 137
3.4 Barrier method 137
3.4.1 Theoretical background 137
3.4.2 Implementation of a barrier method 140
3.5 Modified barrier method 150
3.5.1 Theoretical background 151
3.5.2 Implementation of a modified barrier method 155
3.6 Primal-dual interior point method 169
3.6.1 Theoretical background 169
3.6.2 Implementation of a primal-dual interior point method . . 172
3.7 Boundary point method 178
3.7.1 Theoretical background 178
3.7.2 Implementation of a boundary point method 182
3.8 Conclusions:
Comparison of selected algorithms 191

A Krylov subspace methods **197**
A.1 Conjugate gradients 197
A.2 Conjugate residuals 198

Bibliography **201**

List of abbreviations

sp	single-particle
tp	two-particle
PES	potential energy surface
2DM	second order density matrix
v2DM	variational second order density matrix
v2DM(PQG)	variational second order density matrix with 2-positivity conditions (P-, Q- and G-condition) imposed
v2DM(PQGs)	variational second order density matrix with 2-positivity conditions and subspace conditions imposed
FCI	full configuration interaction
MRCI	multireference configuration interaction
CASSCF	complete active space self-consistent field
CCSD	coupled clusters, truncated to include single and double exitation operators
DFT	density functional theory
DMFT	density matrix functional theory
FC	frozen core approximation
CB	classical barrier method for semidefinite optimization
MB	modified barrier method for semidefinite optimization
PD	primal-dual method for semidefinite optimization
BP	boundary point method for semidefinite optimization

List of Figures

1.1 Illustration of the separating hyperplane theorem to identify non-N-representable 2DM's 12

1.2 Graphic representation of the P-condition on the 2DM 25

1.3 Graphic representation of the constrained variational optimization of the 2DM 26

1.4 Potential and kinetic energy of Be_2 in the 6-31+G* basis set . . 30

1.5 Virial ratio of of Be_2 in the 6-31+G* basis set 30

1.6 Potential and kinetic energy of BeB^+ in the D95V basis set . . . 31

1.7 Virial ratio of BeB^+ in the D95V basis set 31

1.8 PES of Be_2 calculated with the v2DM(PQG) method in different basis sets 33

1.9 PES of Be_2 calculated with FCI(FC) in different basis sets 34

1.10 Comparison of PES of Be_2 calculated with the v2DM(PQG) method and FCI(FC) in different basis sets with an experimentally determined PES 35

1.11 Comparison of v2DM(PQG) with CASSCF and MRCI PES for several 14-electron diatomic molecules 38

1.12 Energy differences between the v2DM(PQG) and MRCI PES for several 14-electron diatomic molecules 39

1.13 Atomic v2DM(PQG) energies as a function of a fractional number of electrons 41

1.14 Minimum of the v2DM(PQG) energy as a functional of a fractional charge on the O atom in the dissociation limit of NO^+ 42
1.15 Energies of N and O as a function of a fractional number of electrons under P-,Q-,G- and T-conditions 42
1.16 Illustration that the energy of an ensemble with fractional electron number is a linear combination of energies for the nearest integer electron numbers when the pure state energies form a convex set 52
1.17 Influence of subspace constraints on the v2DM(PQG) PES of several 14-electron diatomic molecules 54
1.18 Energy difference between the v2DM(PQG) and MRCI PES upon inclusion of subspace constraints 55
1.19 Numbering of atoms and bond lengths in F_3^- 58
1.20 v2DM(PQG) PES of linear F_3^- 59
1.21 MRCI PES of linear F_3^- 60
1.22 Schematic representation of violated subspace constraints in v2DM-(PQG) calculations of different geometries of linear F_3^- 61
1.23 Influence of subspace constraints on a cut of the v2DM(PQG) PES of F_3^-, showing the two competitive dissociations 62
1.24 Schematic representation of the active subspace constraints in v2DM(PQG) calculations of different geometries of linear F_3^- . . 66
1.25 v2DM(PQG) PES for F_3^- upon inclusion of subspace constraints 67

2.1 Overview of v2DM(PQG) atomic energies for the whole range of $\hat{S}_z$ expectation values under different sets of spin constraints . . . 108
2.2 v2DM(PQG) PES for different spin projections of O_2 under pure state spin conditions 109
2.3 v2DM(PQG) PES for different spin projections of C_2 under pure state spin conditions 110
2.4 Comparison of v2DM(PQG) PES of O_2 under pure spin state conditions with its PES under ensemble spin state conditions . . 112

2.5 Comparison of v2DM(PQG) PES of C_2 under pure spin state conditions with its PES under ensemble spin state conditions . . 113

2.6 v2DM(PQG) PES of O_2 under the pure state zero spin projection conditions 116

2.7 v2DM(PQG) PES of C_2 under the pure state zero spin projection conditions 117

2.8 v2DM(PQG) PES of O_2 under the pure state maximal spin projection conditions 118

2.9 v2DM(PQG) PES of C_2 under the pure state maximal spin projection conditions 119

2.10 v2DM(PQG) PES of O_2 under the pure state maximal spin projection conditions, with and without subspace constraints 121

3.1 Illustration of how the logarithmic barrier function approaches a step function as the barrier parameter decreases to zero 139

3.2 Evolution of the spectrum of the Hessian in the classical barrier method's inner iterations for LiH in a STO-6G basis set for decreasing values of the barrier parameter 143

3.3 Number of inner iterations needed to solve the Newton-Rapshon equations in the classical barrier method applied to LiH in the STO-6G basis set as a function of the barrier parameter 144

3.4 Influence of the barrier parameter update factor on the cumulative number of inner Krylov subspace iterations performed in the classical barrier method, applied to LiH in the STO-6G basis set 149

3.5 Evolution of CPU times needed by the classical barrier method to compuate a half-filled Hubbard model as a function of the sp basis dimension 150

3.6 Shape of the inverse barrier function for several values of the barrier parameter 152

3.7 Evolution of the spectrum of the Hessian in the modified barrier method's inner iterations applied to LiH in the STO-6G basis set with decreasing values of the barrier parameter 158

3.8 Comparison of the spectrum of the Hessian of the Newton equations in the modified barrier method with its spectrum in the classical barrier method, applied to LiH in the STO-6G basis set for a barrier parameter of $t = 10^{-2}$ 159

3.9 Comparison of the spectrum of the Hessian of the Newton equations in the modified barrier method with its spectrum in the classical barrier method, applied to LiH in the STO-6G basis set for a barrier parameter of $t = 10^{-4}$ 160

3.10 Comparison of the spectrum of the Hessian of the Newton equations in the modified barrier method with its spectrum in the classical barrier method, applied to LiH in the STO-6G basis set at convergence of each of the two methods 161

3.11 Comparison of the reduction of the duality gap as a function of the number of outer iterations performed in the classical barrier method and modified barrier method applied to LiH in the STO-6G basis set 162

3.12 Comparison of the reduction of the duality gap as a function of the cumulative number of inner iterations performed in the classical barrier method and modified barrier method applied to LiH in the STO-6G basis set 163

3.13 Comparison of the number of inner Krylov subspace iterations required to solve Newton's equations for different values of the penalty parameter in the classical and modified barrier method applied to LiH in the STO-6G basis set 164

3.14 Influence of the barrrier parameter update factor on the cumulative number of inner iterations needed in the modified barrier method applied to LiH in the STO-6G basis set 167

3.15 Evolution of the CPU time required by the modified barrier method to calculate half-filled Hubbard models as a function of the sp basis dimension 168

3.16 Number of inner predictor and corrector Krylov subspace iterations needed to solve Newton's equations as the duality gap decreases in the primal-dual method 175

3.17 Comparison of the reduction of the duality gap as a function of the number of outer iterations in the classical barrier method, modified barrier method and primal-dual method applied to LiH in the STO-6G basis set 176

3.18 Comparison of the reduction of the duality gap as a function of the cumulative number of inner iterations in the classical barrier method, modified barrier method and primal-dual method applied to LiH in the STO-6G basis set 177

3.19 Number of inner Krylov subspace iterations required to solve Newton's equations in the classical barrier method, modified barrier method and primal-dual method as a function of the duality gap 178

3.20 Evolution of the primal and dual infeasibility as a function of the number of outer iterations in the boundary point method applied to LiH in the STO-6G basis set 184

3.21 Influence of the barrier parameter update factor on the cumulative number of inner iterations needed in the boundary point method applied to LiH in the STO-6G basis set 185

3.22 Comparison of the evolution of the main convergence criterion in the classical barrier method, modified barrier method and boundary point method as a function of the number of outer iterations 186

3.23 Comparison of the evolution of the main convergence criterion in the classical barrier method, modified barrier method and boundary point method as a function of the number of cumulative inner iterations 187
3.24 Evolution of CPU times required by the boundary point method for half-filled Hubbard models as a function of the number of sp orbitals 188
3.25 Comparison of the evolution of CPU times required by the classical barrier method, modified barrier method and boundary point method for half-filled Hubbard models as a function of the number of sp orbitals 189

List of Tables

1.1 v2DM(PQG) dipole moments of NO^+, CN^- and CO in the dissociation limit 37

1.2 v2DM(PQG) Mulliken populations of the dissociation products of NO^+, CN^- and CO 37

1.3 Influence of different positivity constraints around equilibrium and in the dissociation limit of NO^+ 43

1.4 Size-inconsistency of the v2DM(PQG) energy of several 14-electron diatomic molecules 45

1.5 Comparison of v2DM(PQG) dissociation energies, with and without inclusion of subspace constraints, with MRCI dissociation energies 55

1.6 Effect of subspace constraints on v2DM(PQG) dipole moments in the dissociation limit of several 14-electron diatomic molecules . 56

1.7 Effect of subspace constraints on v2DM(PQG) Mulliken populations in the dissociation limit of several 14-electron diatomic molecules 57

1.8 v2DM(PQG) Mulliken populations for the partially dissociated F_3^- under different combinations of subspace constraints 63

1.9 Differences between the v2DM(PQG) energy for the partially dissociated F_3^- and the dissociation products calculated separately under different combinations of subspace constraints 64

1.10 Comparison of dissociation energies for F_3^- calculated with the v2DM(PQG) method, MRCI(FC) and CCSD(FC) 68
1.11 Indicators for non-separability of the cumulant for non-interacting non-entangled singlet states 77

2.1 v2DM(PQG) energies and spin properties of the atoms in the dissociated O_2 under different spin conditions on the molecule . . 114
2.2 v2DM(PQG) energies and spin properties of the oxygen atom under several sets of spin conditions 115

3.1 Influence of an automatic preconditioner on the number of iterations required by the method of conjugate gradients and conjugate residuals to compute one Newton step in the classical barrier method, applied to LiH in the STO-6G basis set 147
3.2 Comparison of CPU times needed by different algirithms for semidefinite optimization applied to LiH in the STO-6G basis set 190

Introduction

The holy quantum chemical grail is to find a method to calculate molecular properties exactly, within the limitations imposed by a finite basis set, without needing an exponentially increasing computation time as the size of the molecule grows. Though utopian, *ab initio* quantum chemists persevere in their quest for methods that provide the best trade-off between computational speed and chemical accuracy. Strongly correlated systems form the main obstacle for wavefunction-based methods: describing their correlation effects well requires a multi-determinantal description, making them inherently expensive to compute. Hence alternative approaches are being pursued, focused on lower-order densities and density matrices. These approaches can beat the curse of exponential scaling that approaches based on the full wavefunction suffer from, as the dimension of their basis descriptors does not grow explicitly with the number of electrons. Such methods include density functional theory, density matrix functional theory, cumulant-based methods and second order density matrix-based methods.

Second order density matrix methods are particularly interesting from a conceptual point of view because the second order density matrix (2DM) determines the energy exactly, and therefore these methods do not require approximate functionals to calculate the energy, unlike density and first order density matrix methods. The importance of this property was already realized by Husimi, Coulson and Löwdin in the 1950's.[1–3] This realization naturally lead to the idea of a variational second order density matrix (v2DM) method as an extension

of the variational principle for wavefunctions to the 2DM.[3] However, they soon realised that practical 2DM based methods suffer from another fundamental problem. Since they avoid making any reference to a wavefunction, they must guarantee that there exists some ensemble of wavefunctions from which the 2DM can be derived such that it represents a physical N-electron system. Such a matrix is 'N-representable'.[4] In contrast to the N-representability problem for the 1DM, for which N-representability can be established in polynomial time, N-representability of the 2DM is QMA-hard.[5,6] Therefore, in practice N-representability can only be imposed approximately, introducing errors in the 2DM.[7] Moreover, the most natural conditions on the 2DM take the form of semidefinite constraints and turn the v2DM method into a difficult semidefinite optimization problem.

The wonderfully simple idea of variational optimization of the 2DM sparked off a lot of enthusiasm in the fifties and sixties, but was halted by the limitations of the semidefinite optimization algorithms available at that time.[8–11] In the nineties, the realization that the highly successful interior-point methods for linear programming could be extended to the field of semidefinite programming by Nesterov, Nemirovski and Alizadeh[12,13] revived interests in the field of variational second order density matrix methods. The increased performance of semidefinite algorithms allowed several interesting applications to chemistry.[14–16] Nevertheless, most of these applications are rooted more in physics than in chemistry.

The object of my research has therefore been first of all to assess the variational second order density matrix method's use for chemical electronic structure calculations and secondly to apply this knowledge to make it more effective. I will establish what I believe to be the major strengths of the v2DM method and the major obstacles that must be overcome in order to apply it successfully to molecular calculations. These insights inspired several ideas to improve on it.

In order to address these questions, my colleagues Brecht Verstichel, Ward Poelmans and I have collaborated to develop several semidefinite programs that

carry out the variational 2DM optimization and apply them to study chemical properties of small test molecules under 2-index constraints for N-representability.

This thesis highlights the two principal aspects of practical v2DM methods: the theoretical *N-representability* problem in chapter 1 - 2 and its formulation as a semidefinite optimization problem in chapter 3.

Chapter 1 introduces the concept of N-representability, which is central to the accuracy of practical v2DM methods, and evaluates the approximate 2-positivity conditions on molecular calculations. Their most severe shortcoming is their size-inconsistency, which is also addressed in this chapter.

Chapter 2 focuses explicitly on the implications of approximate N-representability constraints on molecular spin, the *S-representability* problem. It presents several approaches to describing spin in a second-order 2DM framework and discusses them in the context of non-singlet state molecules.

Chapter 3 addresses the formulation of the v2DM method as a semidefinite program and compares several optimization techniques for molecular calculations.

It has frequently been pointed out that a conventional many-electron wave function tells us more than we need to know. There is an instinctive feeling that matters such as electron correlation should show up in the two-particle density matrix . . . but we still do not know the conditions that must be satisfied by the density matrix.

C. A. Coulson, 1959

1 N-representability

1.1 Introduction

The fundamental quantum chemical problem of describing a many-electron system is replaced by the N-representability problem of the second order density matrix (2DM) in the variational second order density matrix (v2DM) method. The equivalence of both approaches derives from the nature of the electron interaction: since electrons interact pairwise, the 2DM fully characterizes their correlated motion. As a consequence, it also determines the energy exactly. Therefore the variational problem shifts from describing electron correlation by a multideterminantal trial wavefunction to ensuring that the trial 2DM corresponds to a physical N-electron system, i.e., that it is 'N-representable'. However, because the exact necessary and sufficient conditions for N-representability have a worst-case complexity that is practically intractable, only a subset of necessary N-representability conditions is implemented.

This chapter, as well as the remainder of this thesis, will focus on 2-index constraints for N-representability, since the computational scaling of these con-

straints is considerably better than 3- or higher order index constraints and computation time is the most prohibitive bottleneck to v2DM methods. Section 1.2 introduces the concept of reduced density matrices, which naturally raises the question of N-representability in section 1.3. The approximations made in practical applications of this method are explained in section 1.4 and the results of our applications to molecular calculations are discussed in section 1.5. It examines these 2-index constraints from a chemical point of view, and focuses on a major shortcoming that is apparent from these applications: an erroneous description of molecular dissociation which violates size-consistency.

1.2 Physical importance of the 2DM

1.2.1 Nth order density matrix

The N-th order density matrix carries all information about an N-electron system. In practice, the wavefunction for such a system is expressed using an orthonormal K-dimensional basis of single-particle (sp) orbitals $\{\phi_1, \ldots, \phi_K\}$ and will be assumed real throughout. A configuration interaction (CI) expansion for the wavefunction can be written

$$|\Psi\rangle = \sum_{i_1,\ldots,i_N}^{K} c_{i_1\ldots i_N} |i_1(1)\ldots i_N(N)\rangle \tag{1.1}$$

where $|i_1(1)\ldots i_N(N)\rangle$ are antisymmetrized N-particle states composed of the sp basis functions indicated by their indices $i_1, \ldots, i_N$.
Its N-th order DM can be expressed as an operator,

$$\begin{aligned}\Gamma^{(N)} &= |\Psi\rangle\langle\Psi| \\ &= \sum_{i_1,\ldots,i_N}^{K} \sum_{j_1,\ldots,i_N}^{K} c_{i_1\ldots i_N} c_{j_1\ldots j_N} |i_1(1)\ldots i_N(N)\rangle\langle j_1(1')\ldots j_N(N')|\end{aligned}$$

which is normalized to 1. Alternatively, it can be represented as a matrix, which gives its expansion coefficients in terms of the antisymmetrized N-particle states,

for which we will use the normalization $N!$

$$\Gamma^{(N)}_{i_1\ldots i_N\ j_1\ldots j_N} = N! c_{i_1\ldots i_N} c_{j_1\ldots j_N} \tag{1.2}$$

Throughout, we will use the normalization 1 for the N-th order density matrix in first quantization, its wavefunction representation, and the normalization $N!$ in second quantization, its projection onto an antisymmetrized N-particle basis. Such an N-th order density matrix describes a pure state wavefunction, which makes it idempotent. As a consequence of fermion statistics, it must also be antisymmetric under exchange of any two indices i_k, i_l and j_k, j_l. Additionally, it is positive semidefinite and normalized to $tr\ \Gamma^{(N)} = N!$.

A mixed state, on the other hand, is described by a weighted, and normalized, combination of pure state N-th order density matrices. Mixed states provide a natural way of representing statistical ensembles – real systems are rarely well described by a pure state – but also provide the most general representation for an ensemble of degenerate states. They can be represented through an operator,

$$\begin{aligned}\Gamma^{(N)} &= \sum_k w_k |\Psi^k\rangle\langle\Psi^k| \\ &= \sum_k w_k \sum_{i_1,\ldots,i_N}^{K} \sum_{j_1,\ldots,j_N}^{K} c^k_{i_1\ldots i_N} c^k_{j_1\ldots j_N} |i_1(1)\ldots i_N(N)\rangle\langle j_1(1')\ldots j_N(N')|\end{aligned}$$

or a matrix

$$\Gamma^{(N)}_{i_1\ldots i_N\ j_1\ldots j_N} = N! \sum_k w_k c^k_{i_1\ldots i_N} c^k_{j_1\ldots j_N} \tag{1.3}$$

where the weights $0 \leq w_k \leq 1$ with $\sum_k w_k = 1$ may, for instance, describe a Boltzmann distribution in a canonical ensemble. In the following, however, we will always be concerned with the ground state at absolute zero temperature. Admitting a mixed ground state is still relevant, however, as it allows for the most general description of a degenerate system.
An N-th order density matrix for a mixed state is antisymmetric, normalized and positive semidefinite, just like a pure state N-th order DM, but is not idempotent.

The N-th order density matrix, as well as lower order density matrices, are tensor operators since the creation/annihilation operators transform independently

from each other under a basis transformation $a^\dagger_\alpha = \sum_i U_{\alpha i} a^\dagger_i$.

1.2.2 2^{nd} order density matrix

The 2DM is a reduction of the N$^{\text{th}}$ order DM that still contains its information on one- and two-particle interactions. It is derived from the N-th order DM by contraction of all but two of its indices

$$\begin{aligned}\Gamma^{(2)}_{i_1 i_2 j_1 j_2} &= \frac{1}{(N-2)!} \sum_{i_3 \ldots i_N} \Gamma^{(N)}_{i_1 i_2 i_3 \ldots i_N \;\; j_1 j_2 i_3 \ldots i_N} \\ &\equiv \mathcal{L}^2_N(\Gamma^{(N)})_{i_1 i_2 j_1 j_2} \end{aligned} \tag{1.4}$$

It thus inherits the properties of positive semidefiniteness and antisymmetry from the N-th order density matrix, and is normalized to $N(N-1)$. In practice, the 2DM is often represented as a 2-dimensional matrix, by mapping the indices $i_1 i_2$ and $j_1 j_2$ onto two-particle (tp) indices I and J. The partial trace operation that projects an N-th order density matrix onto a 2DM will be denoted $\mathcal{L}^2_N()$, following the notation introduced by Coleman[17] and Kummer.[18] It establishes the essential connection between the N-electron system and its reduced representation in terms of 2-electron interactions only.

The importance of the 2DM lies in its characterization of the electron-electron interaction. Because electrons interact pairwise, the 2DM fully describes electron correlation. It therefore determines the expectation value of any operator involving up to two particle interactions. The expectation value of an operator $\hat{A}$ acting on an (in general mixed) state is given by the inner product of its matrix

representation with the N-th order DM

$$\begin{aligned}
&\sum_k w_k \langle \Psi_k | \hat{A} | \Psi_k \rangle \\
&= \sum_k w_k \sum_{i_1 \ldots i_N} \sum_{j_1 \ldots j_N} c^k_{i_1 \ldots i_N} c^k_{j_1 \ldots j_N} \langle j_1 j_2 \ldots j_N | \hat{A} | i_1 i_2 \ldots i_N \rangle \\
&= \sum_{i_1 \ldots i_N} \sum_{j_1 \ldots j_N} \Gamma^{(N)}_{i_1 \ldots i_N j_1 \ldots j_N} \langle j_1 j_2 \ldots j_N | \frac{1}{N!} \hat{A} | i_1 i_2 \ldots i_N \rangle \\
&= \sum_{i_1 \ldots i_N} \sum_{j_1 \ldots j_N} \Gamma^{(N)}_{i_1 \ldots i_N j_1 \ldots j_N} \; A^{(N)}_{i_1 \ldots i_N j_1 \ldots j_N}
\end{aligned}$$

where the matrix $A^{(N)}$ is the projection of the operator $\hat{A}$ onto the basis.

However, when $\hat{A}$ is a two-electron operator, this can be further simplified. The elements $A^{(N)}_{i_1 \ldots i_N j_1 \ldots j_N} = \langle j_1 j_2 \ldots j_N | \frac{1}{N!} \hat{A} | i_1 i_2 \ldots i_N \rangle$ are nonzero only if the bra and ket states differ in at most two occupied orbitals. Consequently, its matrix representation can be written as the antisymmetrized product of a second order reduced matrix $A^{(2)}$ with elements

$$A^{(2)}_{i_1 i_2 j_1 j_2} = \frac{1}{N(N-1)} \langle j_1 j_2 | \hat{A} | i_2 i_1 \rangle$$

and an $(N-2)$th order identity matrix $I^{(N-2)}$, such that

$$\begin{aligned}
A^{(N)} &= \frac{1}{(N-2)!} A^{(2)} \wedge I^{(N-2)} \\
&\equiv \Gamma^N_2 (A^{(2)})
\end{aligned}$$

where $\wedge$ denotes the antisymmetrized normalized Grassmann product. The operation $\wedge\, I^{(N-2)}$ expands a second order reduced representation to an N-th representation, and will be denoted $\Gamma^N_2()$. This expansion operator is the adjoint under the trace operation to the contraction operator $\mathcal{L}^2_N$, which reduces an N-th order matrix to its second order reduced representation by taking its $(N-2)$-th order partial trace.

$$tr\ [\Gamma^N_2 (A^{(2)})\ \Gamma^{(N)}] = tr\ [A^{(2)}\ \mathcal{L}^2_N (\Gamma^{(N)})]$$

Therefore, if the operator $\hat{A}$ only involves up to 2-electron interactions, its expectation value can be expressed using a second order reduced representation

of the N-th order density matrix,

$$\begin{aligned} tr\ A^{(N)}\Gamma^{(N)} &= \frac{1}{(N-2)!} tr\ (A^{(2)} \wedge I^{(N-2)})\Gamma^{(N)} \\ &= \frac{1}{(N-2)!} \sum_{i_1 i_2} \sum_{j_1 j_2} A^{(2)}_{i_1 i_2 j_1 j_2} \sum_{i_3 \ldots i_N} \Gamma^{(N)}_{i_1 i_2 i_3 \ldots i_N j_1 j_2 i_3 \ldots i_N} \\ &= \sum_{i_1 i_2} \sum_{j_1 j_2} A^{(2)}_{i_1 i_2 j_1 j_2} \mathcal{L}^2_N(\Gamma^{(N)})_{i_1 i_2 j_1 j_2} \\ &\equiv tr\ A^{(2)}\Gamma^{(2)} \end{aligned}$$

where the antisymmetry of the N-th order density matrix produces $N(N-1)$ similar reduced factors that can be written in terms of the second order DM (2DM), defined as

$$\Gamma^{(2)}_{i_1 i_2 j_1 j_2} = \mathcal{L}^2_N(\Gamma^{(N)})_{i_1 i_2 j_1 j_2} = \frac{1}{(N-2)!} \sum_{i_3 \ldots i_N} \Gamma^{(N)}_{i_1 i_2 i_3 \ldots i_N j_1 j_2 i_3 \ldots i_N} \tag{1.5}$$

and normalized to $N(N-1)$.

This concept can be generalized to any p-th order DM ($p \leq N$), normalized to $\frac{N!}{(N-p)!}$

$$\begin{aligned} \Gamma^{(p)}_{i_1 i_2 \ldots i_p j_1 j_2 \ldots j_p} &= \mathcal{L}^p_N(\Gamma^{(N)})_{i_1 i_2 \ldots i_p j_1 j_2 \ldots j_p} \\ &= \frac{1}{(N-p)!} \sum_{i_{p+1} \ldots i_N} \Gamma^{(N)}_{i_1 \ldots i_p i_{p+1} \ldots i_N j_1 \ldots j_p i_{p+1} \ldots i_N} \end{aligned}$$

As a consequence, the 2DM also contains the first order DM (1DM),

$$\Gamma^{(1)}_{i_1 j_1} = \frac{1}{N-1} \sum_{i_2} \Gamma^{(2)}_{i_1 i_2 j_1 i_2} \tag{1.6}$$

which is normalized to N.

Just like the N-th order density matrix is the projection onto the chosen sp basis of its spatial representation, the 2DM's spatial representation, the *pair density matrix*, is often denoted $\rho^{(2)}(x_1, x_2; x'_1, x'_2)$. It can be expanded in the same basis as $\Gamma^{(N)}$,

$$\rho^{(2)}(x_1, x_2; x'_1, x'_2) = \sum_{ijkl} \Gamma^{(2)}_{ijkl} \phi_i(x_1) \phi_j(x_2) \phi_k(x'_1) \phi_l(x'_2)$$

Its 'diagonal form', which has $x_1 = x'_1, x_2 = x'_2$, is called the pair density.

The 2DM thus determines the expectation value of all one and two-electron operators, including the energy. A Hamiltonian composed of a one-particle operator $\hat{h}$ and a two-particle operator $\hat{V}$ in its antisymmetrized second order reduced form has elements

$$\begin{aligned} H^{(2)}_{ijkl} &= \frac{1}{N-1}\langle ij|\hat{h}|kl\rangle + \langle ij|\hat{V}|kl\rangle \\ &= \frac{1}{N-1}(\delta_{jl}h_{ik} + \delta_{ik}h_{jl} - \delta_{il}h_{jk} - \delta_{jk}h_{il}) + V_{ijkl} \end{aligned} \tag{1.7}$$

with $h_{ik} = \langle k|\hat{h}|i\rangle$. The energy of the system is a linear matrix function of the second order density matrix

$$E = \sum_{ijkl} H^{(2)}_{ijkl}\Gamma^{(2)}_{ijkl} = tr\ [H^{(2)}\Gamma^{(2)}] \tag{1.8}$$

The 2DM may thus be used as an alternative to the wavefunction in variational procedures! The very simple idea to apply the variational principle to the 2DM instead of to the wavefunction,

$$\underbrace{min}_{\Gamma^{(2)}\succeq 0,\ tr\ \Gamma^{(2)}=N(N-1)} \qquad E = tr\ [H^{(2)}\Gamma^{(2)}] \tag{1.9}$$

has unchained a whole area of research from the 1950's on, *variational second order density matrix theory*.[17,19,20] The advantage of using the 2DM as descriptor for a chemical system as opposed to the wavefunction has been its major driving force. Whereas the number of variational degrees of freedom in the wavefunction increases exponentially with the number of particles, the size of the 2DM is not directly influenced by the number of particles, only by the dimension of the one-particle basis set. This has invoked such enthusiasm for v2DM theory that a 'quantum mechanics without wavefunctions' was envisaged.[21] However, the first trial calculations that aimed to minimize the energy over a normalized, positive semidefinite 2DM, (1.9), gave energies that were in a sense 'too strongly correlated', as they were well below the exact energy.[22] This finding indicated that the variational space over which the energy was minimized was much too large.[23] It has led to the realization that, using the term Coleman coined to describe this problem, the 2DM must be *N-representable*.[4]

1.3 N-representability

1.3.1 Definition of N-representability

In the foregoing, by defining the 2DM as the second order reduction of an N-th order DM, $\Gamma^{(2)} = \mathcal{L}_N^2(\Gamma^{(N)})$, it has been tacitly assumed N-representable. However, when given a random matrix with the same dimension as the 2DM, does there even exist a physically correct $\Gamma^{(N)}$ from which it can be reduced under the contraction $\mathcal{L}_N^2$? This is the essence of the *N-representability* problem.

When the 2DM $\Gamma^{(2)}$ is derivable from an N-th order DM $\Gamma^{(N)}$ that is antisymmetric, Hermitian, normalized and positive semidefinite under the contraction $\mathcal{L}_N^2$, it represents a physically correct N-electron system. It is thus *ensemble N-representable*.

When the 2DM $\Gamma^{(2)}$ is derivable from an N-th order DM $\Gamma^{(N)}$ that is antisymmetric, Hermitian, normalized and positive semidefinite as well as idempotent under the contraction $\mathcal{L}_N^2$, it is *pure state N-representable*, since an idempotent N-th order density matrix represents a pure state. However, we will not be concerned with pure state N-representability here and will always interpret 'N-representable' to mean 'ensemble N-representable'.

The set of N-representable 2DMs is convex. An ensemble 2DM that derives from a proper N-th order density matrix by contraction is a weighted combination of pure state N-representable 2DM's. Consequently, the set of ensemble N-representable 2DM is the convex hull of all pure state N-representable 2DM. Its convexity plays an important role in v2DM methods.

Given the linear dependence of the energy on the 2DM, the 2DMs on the boundary of the set of N-representable 2DM correspond to the ground state of some Hamiltonian. More specifically, an extreme point on the boundary has a pure state preimage in the set of N-representable N-th order density matrices, although the reverse is not necessarily true.[4]

In fact, an exposed point on the boundary of the set of N-representable 2DM

must correspond uniquely to either a non-degenerate ground state or a degenerate ground state for which the degeneracy cannot not be distinguished through any 2-electron operator. Determining whether the one-to-one correspondence of an exposed point to a ground state can actually be narrowed down to a non-degenerate ground state, has not been solved.[19] A flat or extreme non-exposed point corresponds to a degenerate ground state for which the degeneracy can be removed by adding some infinitesimal operator.[19]

Since any convex set is completely determined by its extreme points by virtue of the Krein-Milman theorem, it would suffice to characterize the extreme points of the set of N-representable 2DM's, which have a pure state preimage in the set of N-representable N-th order density matrices,[4] or even the extreme exposed points. However, even though the extreme points of the 1DM are easily identified, the geometry of the second order N-representable set is much more intricate. The set of N-representable 1DM's is completely described as the convex hull defined by the single Slater determinant 1DM's, which are projectors onto an N-dimensional subspace of the K-dimensional Hilbert space. The set of all these projectors is the set of extreme points of the 1DM N-representable set.[17,24,25]

The correspondence of extreme 1DM's to Slater type 1DM's can be understood as follows. First of all, to show that an extreme 1DM corresponds to a single Slater type 1DM, consider its pure state preimage in the natural orbital basis. Every extreme 1DM's preimage in the set of N-representable N-th order density matrices contains a pure state. The 1DM derived from a pure state in its natural orbital basis is

$$
\begin{aligned}
|\Psi\rangle &= \sum_{i_1 \ldots i_N} c_{i_1 \ldots i_N} |i_1 \ldots i_N\rangle \\
\Gamma^{(N)}_{i_1 \ldots i_N \;\; j_1 \ldots j_N} &= N! \; c_{i_1 \ldots i_N} c_{j_1 \ldots j_N} \\
\Gamma^{(1)}_{kl} &= N \delta_{kl} \sum_{i_2 \ldots i_N} c^2_{k i_2 \ldots i_N}
\end{aligned}
$$

The normalization of the wavefunction implies that $|c_{i_1 \ldots i_N}| \leq \frac{1}{\sqrt{N!}}$, with equality holding only for a single Slater determinant. Therefore, the diagonal elements of

the 1DM lie between 0 and 1

$$\begin{aligned}\Gamma^{(1)}_{kl} &= N\delta_{kl} \sum_{i_2 \ldots i_N} |c_{k i_2 \ldots i_N}|^2 \\ &\leq N\delta_{kl} \frac{(N-1)!}{N!} = \delta_{kl} \end{aligned} \tag{1.10}$$

However, the assumption of extremity only allows 0 and 1 occupations; any other occupation number would imply that the 1DM can be written as a linear combination of 1DM's. As a consequence, any extreme 1DM has a single Slater determinant preimage.
Conversely, a Slater determinant type 1DM has diagonal elements 0 and 1 and is therefore extreme.

However, the set of single determinant 2DM's does not fully characterize the set of extreme points of the set of N-representable 2DM, because their linear combinations do not cover the whole N-representable set. A linear combination of single determinant 2DM's cannot have an eigenvalue larger than one, which can occur for the 2DM (cfr. infra, 1.16). Moreover, because the 2DM in general cannot be diagonalized by a suitable choice of the sp basis, the N-representability conditions for the 2DM are not expressible in terms of its spectrum. This simple argument explains why the N-representability problem for the 2DM is so much harder than that for the 1DM.

In the following discussion on the N-representability problem we focus on the 2DM's, using the notation Γ for $\Gamma^{(2)}$ and H for $H^{(2)}$. To make the distinction with the 1DM, the 1DM will be denoted γ instead of $\Gamma^{(1)}$.

1.3.2 Necessary and sufficient conditions for N-representability

Of course, any method based solely on the 2DM requires a formulation of N-representability in terms of the 2DM itself. In fact, the necessary and sufficient conditions for a 2DM to be N-representable are known in terms of the 2DM, but

impossible to apply to realistic problems. Intuitively, they follow directly from the separating hyperplane theorem, which states that[26]

Separating hyperplane theorem

Suppose C and D are two convex sets that do not intersect, i.e. $C \cap D = \phi$, then there exist $a \neq 0$ and b such that $a^T x \leq b \ \ \forall x \in C$ and $a^T x \geq b \ \ \forall x \in D$. The hyperplane $\left\{x | a^T x = b\right\}$ is called a separating hyperplane for the sets C and D.

Since the set of N-representable 2DM and any single non N-representable 2DM form disjoint convex sets, the separating hyperplane theorem can be applied to it. It implies that for any non-N-representable 2DM, there exists a second order reduced 'Hamiltonian' $H^{(2)}$ that defines a separating hyperplane $tr\ H^{(2)}\Gamma = E_0(H)$ that spatially separates it from the set of N-representable 2DM. Such a separating hyperplane can always be constructed as the hyperplane through the point in the N-representable set that is nearest to the non-N-representable 2DM under consideration and normal to the difference between both. In other words, it supports the N-representable set in the point closest to the non N-representable 2DM under consideration, such that this point is its orthogonal projection onto the plane.

$$\begin{aligned} &\forall\ \tilde{\Gamma}^{(2)} \text{ that are not N-representable} \\ &\exists H^{(2)} :\ tr\ [H^{(2)}\tilde{\Gamma}^{(2)}] \leq E_0(H) \end{aligned}$$

Because this must hold for any non N-representable 2DM, the necessary and sufficient condition for N-representability is that there exists no hyperplane that separates it from the N-representable set

$$\begin{aligned} &\tilde{\Gamma}^{(2)} \text{ is N-representable} \Leftrightarrow \\ &\forall\ H^{(2)} :\ \ tr\ [H^{(2)}\Gamma^{(2)}] \geq E_0(H^{(2)}) \end{aligned} \tag{1.11}$$

with $E_0(H^{(2)})$ the ground state energy for the Hamiltonian. A mathematically more rigorous proof of these conditions was derived by Garrod and Percus,[27] and refined by Kummer.[18]

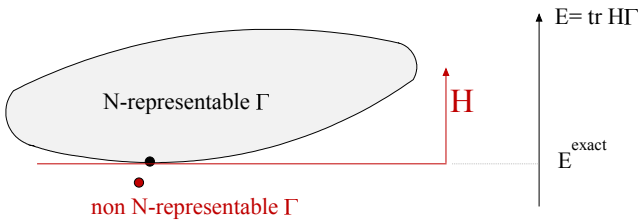

Figure 1.1: For any 2DM that does not lie in the set of N-representable 2DM, there exists a separating hyperplane $tr\ H\Gamma = E_0(H)$ that separates it from the N-representable set.

Clearly, this formulation of the necessary and sufficient conditions for N-representability is hardly practicable, and although it was believed for quite a while[28] that the necessary and sufficient conditions for N-representability of the 2DM could be formulated in a simple manner, this has not been realized up to now. If anything, it seems that the difficulty of this problem has become more apparent.[29] The formulation (1.11) of the necessary and sufficient conditions for N-representability makes it at least as hard as full configuration interaction (FCI).

In fact, the problem of determining whether a given 2DM is N-representable, has been shown to be Quantum Merlin Arthur (QMA)-complete, when the number of electrons is considered the 'size' of the system. The QMA complexity class is the quantum analog of the Nondeterministic Polynomial time complexity class (NP) in a probabilistic setting. A problem is said to be complete in a complexity class if any other problem in this class can be reduced to it, and it is in this class itself. Completeness of a problem is rarely proven directly based on this definition, however. Instead, it is usually proven by showing that the problem lies in the complexity class under consideration and that another problem that has already been proven to be complete in this class can be reduced to it, as this implies that any other problem in the class can be reduced to it. Consequently, proofs of completeness of different problems have been generated

in a tree-like fashion, with branches of proofs of completeness that rely on its predecessor's proof of completeness, and which finally, all depend on one problem at the base of the tree, of which completeness has been proven directly based on its definition. The classical satisfiability (SAT) problem is such a problem for NP, as its NP-completeness has been proven directly by Cook.[30]

Verstraete et al.[6] have proven that deciding whether a 2DM is N-representable is QMA-complete, first of all by showing that the problem is in QMA, and, secondly, that it can be reduced to the problem of finding the ground state energy of the local Hamiltonian problem with 2-body interactions. The local Hamiltonian problem is similar in spirit to the MAX-SAT problem generalized to a probabilistic quantum setting. In fact, the k-local Hamiltonian problem contains the MAX-k-SAT problem. As a consequence, the 2-local Hamiltonian problem is NP-hard because the MAX-2-SAT problem is NP-complete. Kitaev and al. have narrowed this result down by specifying that it is QMA-complete.[5] The 1-local Hamiltonian problem, however, is in P.

The N-representability problem can be linked to the 2-local Hamiltonian problem by realizing that any 2-local Hamiltonian of fermions with an sp basis of dimension 2N can be mapped onto a 2-local Hamiltonian of spins.[6] Although the QMA-completeness of the 2-local Hamiltonian problem implies that the N-representability problem is QMA-complete, it is not necessarily intractable in practice. The specific symmetry present in molecular systems may simplify the problem.

1.3.3 Necessary conditions implied by N-representability

Because the formulation (1.11) of generally holding necessary and sufficient conditions for N-representability is at least as hard as full CI from the complexity point of view, it is useless in practice. Nevertheless, subsets of the constraints (1.11) determined by classes of exactly solvable Hamiltonians provide useful necessary conditions on the 2DM.

A straightforward set of such necessary conditions is provided by the Hamilto-

nians of the form $\hat{H} = \sum_{ij} c_i^A c_j^A \ \hat{A}_i^\dagger \hat{A}_j$, which must have a positive expectation value. Different forms of the operator A lead to different positivity conditions.

$$\sum_{ij} c_i^A c_j^A \langle \Psi | \hat{A}_i^\dagger \hat{A}_j | \Psi \rangle \geq 0 \quad \forall c^A \neq 0$$

$$\Leftrightarrow \langle \Psi | \hat{A}^\dagger \hat{A} | \Psi \rangle \succeq 0$$

where the symbols $\preceq 0$ and $\succeq 0$ denote an ordering with respect to the positive semidefinite cone. The notation $\langle \Psi | \hat{A}^\dagger \hat{A} | \Psi \rangle \succeq 0$ therefore indicates that the matrix $\langle \Psi | \hat{A}^\dagger \hat{A} | \Psi \rangle$ has positive eigenvalues. This type of condition is independent of the choice of basis, because the spectrum of the constraint matrices does not change under a unitary basis transformation. These conditions are called *p-positivity* conditions, where p indicates the order of the creation-/annihilation operator string involved. Using the anticommutation relationships for creation and annihilation operators, the above constraint matrix $\langle \Psi | \hat{A}^\dagger \hat{A} | \Psi \rangle$ can be expressed as a linear function of the 2DM.

1-Positivity

The 1-positivity conditions originate from operators $\hat{A}$ involving one particle/hole operator. The 'p-condition' imposes that the 1DM must be positive semidefinite, whereas the 'q-condition' imposes that the first order hole DM must be positive semidefinite[4]

p-condition

$$p \succeq 0 \quad \text{with} \quad p_{ij} = \langle \Psi | a_j^\dagger a_i | \Psi \rangle \tag{1.12}$$

q-condition

$$q \succeq 0 \quad \text{with} \quad q_{ij} = \langle \Psi | a_j a_i^\dagger | \Psi \rangle \tag{1.13}$$

However, since $q_{ij} = \delta_{ij} - p_{ij}$

$$p \succeq 0$$

$$I - p \succeq 0$$

Therefore all eigenvalues $\lambda^{(1)}$ of the 1DM must lie between 0 and 1

$$0 \leq \lambda^{(1)} \leq 1 \tag{1.14}$$

This condition is necessary but also sufficient for N-representability of the 1DM, which was first proven by Coleman[17] and illustrated above by considering a CI-expansion for the wavefunction (1.10). Since the concept of N-representability is independent of the choice of sp basis and the 1DM can always be made diagonal by choosing a suitable sp basis, N-representability of the 1DM can be formulated completely in terms of its spectrum.

2-Positivity

Unlike N-representability of the 1DM, necessary and sufficient conditions for N-representability of the 2DM cannot simply be expressed in terms of the spectrum of the 2DM.[4] Moreover, contrary to early beliefs, the eigenvalues of the 2DM are not constrained to a value between 0 and 1. As Sasaki, Yang and Coleman have established,[4,31,32] a tight upper bound for the eigenvalues of the 2DM is

$$0 \leq \lambda^{(2)} \leq N \quad \text{for N even} \tag{1.15}$$

$$0 \leq \lambda^{(2)} \leq N-1 \quad \text{for N odd} \tag{1.16}$$

Garrod and Percus have introduced a set of necessary, but in general not sufficient, conditions for N-representability[27] that can be derived by considering an operator $\hat{A}$ with two creation/annihilation operators. This leads to four 2-positivity constraints, only three of which are independent. The so-called P-,Q- and G-condition impose positive semidefiniteness of the particle-particle, hole-hole and particle-hole matrix.

P-condition

$$P \succeq 0 \quad \text{with} \quad P_{ijkl} = \langle \Psi | a_k^\dagger a_l^\dagger a_j a_i | \Psi \rangle \tag{1.17}$$

The P-matrix is simply the 2DM, $\Gamma^{(2)}$.

Q-condition

$$Q \succeq 0 \quad \text{with} \quad Q_{ijkl} = \langle \Psi | a_k a_l a_j^\dagger a_i^\dagger | \Psi \rangle \tag{1.18}$$

The Q-matrix is a linear function of the 2DM,

$$\begin{aligned} Q_{ijkl} &= \delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk} \\ &\quad - \frac{1}{N-1}\sum_m (\delta_{ik}\Gamma_{jmlm} + \delta_{jl}\Gamma_{imkm} - \delta_{il}\Gamma_{jmkm} - \delta_{jk}\Gamma_{imlm}) \qquad (1.19) \\ &\quad + \Gamma_{ijkl} \qquad (1.20) \end{aligned}$$

or written more concisely using an unnormalized Grassmann wedge product

$$Q = \delta \wedge \delta - \delta \wedge \gamma + \Gamma$$

It will be convenient to consider it as a homogeneous linear mapping acting on an antisymmetric two-particle/hole matrix, $Q : \Gamma \to Q(\Gamma)$

$$\begin{aligned} Q(\Gamma)_{ijkl} &= (\delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk})\frac{1}{N(N-1)}\sum_{mn}\Gamma_{mnmn} \\ &\quad - \frac{1}{N-1}\sum_m (\delta_{ik}\Gamma_{jmlm} + \delta_{jl}\Gamma_{imkm} - \delta_{il}\Gamma_{jmkm} - \delta_{jk}\Gamma_{imlm}) \qquad (1.21) \\ &\quad + \Gamma_{ijkl} \qquad (1.22) \end{aligned}$$

G-condition

$$G \succeq 0 \quad \text{with} \quad G_{ijkl} = \langle \Psi | a_k^\dagger a_l a_j^\dagger a_i | \Psi \rangle \tag{1.23}$$

In contrast to the P- and Q-matrices, the G-matrix is not antisymmetric. Nonetheless, the image of the antisymmetric 2DM under the G-map has a specific symmetry that originates from the antisymmetry of the 2DM. Viewed as a linear homogeneous matrix map, $G : \Gamma \to G(\Gamma)$ is

$$G(\Gamma)_{ijkl} = \delta_{jl}\frac{1}{N-1}\sum_m \Gamma_{imkm} - \Gamma_{ilkj} \tag{1.24}$$

A completely analogous map can be applied to the domain of G-like matrices that have the same symmetry as the G-matrix derived from an antisymmetric 2DM, for instance when considering the squared map, $G(G(\Gamma))$. We will use the same notation for both cases, as the distinction will be clear from the context.

$\tilde{G}$-condition

$$\tilde{G} \succeq 0 \quad \text{with} \quad \tilde{G}_{ijkl} = \langle \Psi | a_k a_l^\dagger a_j a_i^\dagger | \Psi \rangle \tag{1.25}$$

This constraint is already implied by the G-condition. Its positive semidefiniteness follows from the positive semidefiniteness of the G-matrix.

$$\begin{aligned}
\tilde{G}_{ijkl} &= G_{jilk} + \delta_{kl}\delta_{ij} - \delta_{kl}\gamma_{ij} - \delta_{ij}\gamma_{kl} \\
x^T \tilde{G} x &= \sum_{ijkl} x_{ij} x_{kl} \sum_{abcd} G_{abcd} (\delta_{aj}\delta_{bi}\delta_{cl}\delta_{dk} + \frac{1}{N^2}\delta_{kl}\delta_{ij}\delta_{ab}\delta_{cd} \\
&\quad - \frac{1}{N}\delta_{kl}\delta_{cd}\delta_{aj}\delta_{bi} - \frac{1}{N}\delta_{ij}\delta_{ab}\delta_{cl}\delta_{dk}) \\
&= \sum_{ijkl} x_{ij} x_{kl} \sum_{abcd} G_{abcd} (\delta_{aj}\delta_{bi} - \frac{1}{N}\delta_{ab}\delta_{ij})(\delta_{cl}\delta_{dk} - \frac{1}{N}\delta_{cd}\delta_{kl}) \\
&\geq 0 \quad \forall x
\end{aligned}$$

Historically, the G-condition was introduced by Garrod and Percus in a nonlinear, but equivalent, form[27]

$$\begin{aligned}
G' &\succeq 0 \\
G'_{ijkl} &= \delta_{jl}\gamma_{ik} - \Gamma_{ilkj} - \gamma_{ij}\gamma_{kl} \\
&= G_{ijkl} - \gamma_{ij}\gamma_{kl}
\end{aligned}$$

This form is equivalent to the definition (1.23) of the G-condition. Positive semidefiniteness of G' trivially implies positive semidefiniteness of G:

$$\begin{aligned}
x^T G x = x^T G' x + (x^T \gamma)^2 &\geq 0 \\
&\geq (x^T \gamma)^2 \\
&\geq 0 \quad \forall x
\end{aligned}$$

Conversely, positive semidefiniteness of G also implies positive semidefiniteness

of G'. Since a positive semidefinite G-matrix can be factored as $G = RR^T$

$$\begin{aligned} x^T G' x &= x^T G x - (x^T \gamma)^2 \\ &= \frac{1}{N^2}(x^T G x)(e^T G e) - \frac{1}{N^2}(x^T G e)^2 \\ &= \frac{1}{N^2}(x^T R)(R^T x)(e^T R)(R^T e) - \frac{1}{N^2}(x^T R)(R^T e)(x^T R)(R^T e) \\ &\geq 0 \quad \forall x \end{aligned} \tag{1.26}$$

The vector e is the vector representation of the identity matrix in the tp basis, $e_{ij} = \delta_{ij}$, such that $(e^T G)_{ij} = \sum_k G_{kkij} = N\gamma_{ij}$ and $e^T G e = \sum_{ij} G_{iijj} = N^2$. In the last line, the Cauchy-Schwarz inequality was invoked. Therefore, the linear form $G(\Gamma)$ and the nonlinear form $G'(\Gamma)$ are equivalent conditions.

In general, p-positivity conditions imply lower order positivity conditions. In particular,

$$\begin{aligned} P \succeq 0 &\Rightarrow p \succeq 0 \\ Q \succeq 0 &\Rightarrow q \succeq 0 \\ G \succeq 0 &\Rightarrow p \succeq 0,\ q \succeq 0 \end{aligned}$$

The G-condition implies both the $p-$ and $q-$condition, because it can contract to either of these, depending on whether it is contracted according to the particle index or the hole index of the particle-hole state.

3-Positivity

When the operator $\hat{A}$ is a string of three creation/annihilation operators, positivity conditions on the third order DM emerge. Although the 3DM is not available in an approach based on the 2DM, lower order conditions can be derived from the 3-positivity conditions. Two-index matrix conditions can be obtained by recognizing that the anticommutators of operators $\hat{A}$ involving three creation-/annihilation operators not only preserve positive semidefiniteness, but also remove their dependence on the 3DM. These conditions were derived

by Erdahl[33] and are referred to as the 'T-conditions'. The anticommutator $\langle\Psi|\hat{A}^\dagger\hat{A}+\hat{A}\hat{A}^\dagger|\Psi\rangle$ for $\hat{A}=\sum_{ijk}c^A_{ijk}a_ia_ja_k$ leads to the 'T^1 condition'

T^1-condition

$$T^1_{ijklmn}\succeq 0 \quad \text{with} \quad T^1_{ijklmn}=\langle\Psi|a_l^\dagger a_m^\dagger a_n^\dagger a_k a_j a_i + a_n a_m a_l a_i^\dagger a_j^\dagger a_k^\dagger|\Psi\rangle \tag{1.27}$$

The T^1-condition depends only on the 2DM because the commutator of both terms yields

$$\begin{aligned}
T^1_{ijklmn} &= \delta_{nk}\Gamma_{mlji}-\delta_{nj}\Gamma_{mlki}+\delta_{in}\Gamma_{mlkj}\\
&-\delta_{mk}\Gamma_{nlji}+\delta_{jm}\Gamma_{nlki}-\delta_{im}\Gamma_{nlkj}\\
&+\delta_{kl}\Gamma_{nmji}-\delta_{jl}\Gamma_{nmki}+\delta_{il}\Gamma_{nmkj}\\
&-(\delta_{nk}\delta_{mj}-\delta_{nj}\delta_{mk})\gamma_{il}+(-\delta_{in}\delta_{mk}+\delta_{nk}\delta_{mi})\gamma_{jl}\\
&-(\delta_{nk}\delta_{il}-\delta_{in}\delta_{lk})\gamma_{mj}+(-\delta_{nj}\delta_{kl}+\delta_{nk}\delta_{jl})\gamma_{mi}\\
&-(\delta_{nj}\delta_{mi}-\delta_{in}\delta_{mj})\gamma_{kl}+(-\delta_{in}\delta_{jl}+\delta_{nj}\delta_{il})\gamma_{mk}\\
&-(\delta_{mk}\delta_{jl}-\delta_{jm}\delta_{kl})\gamma_{in}+(-\delta_{im}\delta_{kl}+\delta_{mk}\delta_{il})\gamma_{nj}\\
&-(\delta_{jm}\delta_{il}-\delta_{im}\delta_{jl})\gamma_{kn}\\
&-\delta_{ni}\delta_{mj}\delta_{kl}-\delta_{nj}\delta_{mi}\delta_{kl}+\delta_{nk}\delta_{mj}\delta_{il}\\
&+\delta_{nj}\delta_{mk}\delta_{il}-\delta_{nk}\delta_{mi}\delta_{jl}+\delta_{ni}\delta_{mk}\delta_{jl}
\end{aligned}$$

which can be written in a very compact manner using the unnormalized Grassmann wedge product

$$T^1=\delta\wedge\Gamma-\delta\wedge\delta\wedge\gamma+\delta\wedge\delta\wedge\delta$$

T^2-condition

Similarly, another 3-index constraint can be derived on the 2DM by considering the anticommutator

$$T^2\succeq 0 \quad \text{with} \quad T^2_{ijklmn}=\langle\Psi|a_l^\dagger a_m^\dagger a_n a_k^\dagger a_j a_i + a_n^\dagger a_m a_l a_i^\dagger a_j^\dagger a_k|\Psi\rangle \tag{1.28}$$

which is independent of the 3DM

$$\begin{aligned} T^2_{ijklmn} = \delta_{kn}\Gamma_{ijkm} + \gamma_{nk}(\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}) \\ - \delta_{il}\Gamma_{kmnj} + \delta_{im}\Gamma_{klnj} + \delta_{jl}\Gamma_{kmni} - \delta_{jm}\Gamma_{klni} \end{aligned}$$

The above T^2-condition arises from $\langle\Psi|\hat{A}^\dagger\hat{A} + \hat{A}\hat{A}^\dagger|\Psi\rangle$ with $\hat{A} = \sum_{ijk} c^A_{ijk} a^\dagger_i a^\dagger_j a_k$. However, changing the relative position of the creation operators and the annihilation operator leads to a different constraint. Instead of imposing all different conditions, the dependence on the arrangement of the creation and annihilation operators can be removed by imposing that

$$\langle\Psi|(\hat{A}^\dagger + \hat{a}^\dagger)(\hat{A} + \hat{a}) + \hat{A}\hat{A}^\dagger|\Psi\rangle \succeq 0$$

with $\hat{a} = \sum_i c^a_i a_i$ a one-electron operator. Equivalently,

$$\forall c^A, c^a : \quad \sum_{ijkk'lmnn'} c^A_{ijk} c^a_{k'} \begin{pmatrix} T^2_{ijklmn} & \Gamma_{ijkn'} \\ \Gamma_{k'lmn} & \gamma_{k'n'} \end{pmatrix} c^A_{lmn} c^a_{n'} \geq 0$$

Although these 3-index constraints only depend on the 2DM, they are still expensive to impose. Being 3-index constraints, their dimension scales as K^6, as opposed to K^4 for the 2-positivity constraints. With current computational and algorithmic means, they are practically unworkable in any chemically relevant basis set. For this reason, I have decided to work primarily with 1- and 2-index constraints. An alternative could be to impose partial 3-positivity conditions,[34] for instance only conditions on the diagonal. That way, one could even attempt to impose partial higher order conditions.[35,36]

P-,Q- and G-type maps and their inverse maps

The P-,Q- and G-map, and other 2-index maps on the 2DM derived from them, have a structure similar to the map $Y^Q()$ or $Y^G()$

$$\begin{aligned} Y^Q(\Gamma)_{ijkl} &= c_0^Q\ \Gamma_{ijkl} + c_1^Q\ \sum_n (\delta_{il}\Gamma_{jnkn} + \delta_{jk}\Gamma_{inln} - \delta_{ik}\Gamma_{jnln} - \delta_{jl}\Gamma_{inkn}) \\ &\quad + c_2^Q(\delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk}) \sum_{mn} \Gamma_{mnmn} \qquad (1.29) \\ Y^G(\Gamma)_{ijkl} &= c_0^G\ \Gamma_{ilkj} + c_1^G\ \sum_n \delta_{jl}\Gamma_{inkn} \qquad (1.30) \end{aligned}$$

where the coefficients c_0^Q, c_1^Q, c_2^Q and c_0^G, c_1^G determine the nature of the map. The map Y^Q is antisymmetric and has a structure similar to the Q-map. The map Y^G has the same symmetry as the G-map. The inverse of such a map $Y()$ is of the same form, because it can be constructed from $Y()$ and its first and second order contractions

$$\begin{aligned} \sum_n Y^Q(\Gamma)_{inkn} &= c_0^Q \sum_n \Gamma_{inkn} + c_1^Q \sum_n (\Gamma_{inkn} + \Gamma_{inkn} - \delta_{ik}\Gamma_{mnmn} - K\Gamma_{inkn}) \\ &\quad + c_2^Q(\delta_{ik}K - \delta_{ik})\Gamma_{mnmn} \\ &= \sum_n \Gamma_{inkn}(c_0^Q + (2-K)c_1^Q) + \sum_{mn} \Gamma_{mnmn}\delta_{ik}(c_2^Q(K-1) - c_1^Q) \\ \sum_{mn} Y^Q(\Gamma)_{mnmn} &= (c_0^Q + 2c_1^Q(1-K) + c_2^Q K(K-1)) \sum_{mn} \Gamma_{mnmn} \\ \sum_n Y^G(\Gamma)_{inkn} &= (c_0^G + c_1^G K) \sum_n \Gamma_{inkn} \end{aligned}$$

The coefficients for the zeroth, first and second order contractions in the inverse maps $Y^{-1}(\Gamma)$ are then chosen to remove any dependence on the contractions of

Γ when they act on $Y(\Gamma)$, such that $Y^{-1}(Y(\Gamma)) = \Gamma$,

$$\begin{aligned}
c_0^{Q,-1} &= \frac{1}{c_0^Q} \\
c_1^{Q,-1} &= \frac{-c_1^Q}{c_0^Q}\frac{1}{c_0^Q + c_1^Q(2-K)} \\
c_2^{Q,-1} &= \frac{-2c_1^Q}{c_0^Q}\frac{c_2^Q(K-1) - c_1^Q}{c_0^Q + 2c_1^Q(1-K) + c_2^Q K(K-1)}\frac{1}{c_0^Q + c_1^Q(2-K)} \\
&\quad - \frac{c_2^Q}{c_0^Q}\frac{1}{c_0^Q + 2c_1^Q(1-K) + c_2^Q K(K-1)} \\
&= \frac{1}{c_0^Q}\frac{2(c_1^Q)^2 - c_0^Q c_2^Q - c_1^Q c_2^Q K}{c_0^Q + 2c_1^Q(1-K) + c_2^Q K(K-1)}\frac{1}{c_0^Q + c_1^Q(2-K)}
\end{aligned} \tag{1.31}$$

and

$$\begin{aligned}
c_0^{G,-1} &= \frac{1}{c_0^G} \\
c_1^{G,-1} &= \frac{-c_1^G}{c_0^G}\frac{1}{c_0^G + c_1^G K}
\end{aligned}$$

Any linear combination of the P-, Q- and G-condition of the form (1.29) or (1.30) is therefore exactly invertible. For instance, the inverse Q- and G-maps are

$$\begin{aligned}
Q^{-1}(\Gamma)_{ijkl} &= \Gamma_{ijkl} + \frac{1}{K-N-1}\sum_n(\delta_{il}\Gamma_{jnkn} + \delta_{jk}\Gamma_{inln} - \delta_{ik}\Gamma_{jnln} - \delta_{jl}\Gamma_{inkn}) \\
&\quad + \frac{1}{(K-N)(K-N-1)}(\delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk})\sum_{mn}\Gamma_{mnmn} \\
G^{-1}(\Gamma)_{ijkl} &= -\Gamma_{ilkj} + \frac{1}{K-N+1}\,\delta_{jl}\sum_n \Gamma_{inkn}
\end{aligned}$$

Moreover, powers of the P-, Q- and G- maps have the same structure as these maps and are thus exactly invertible. In particular, the map $L^\dagger : \Gamma \to \Gamma + Q(Q(\Gamma)) + \mathcal{A}(G(G(\Gamma)))$ with $\mathcal{A}$ an antisymmetrizer, is of the form $Y^Q(\Gamma)$ and has a straightforward inverse, that will be exploited in semidefinite programming applications (see chapter 3). The quadratic P-map is equal to itself; given the

Q-map's contractions

$$\sum_n Q(\Gamma)_{inkn} = \frac{N-K+1}{N-2}\sum_n \Gamma_{inkn} + \delta_{ik}\frac{K-N-1}{N(N-1)}\sum_{mn}\Gamma_{mnmn}$$
$$\sum_{mn} Q(\Gamma)_{mnmn} = \sum_{mn}\Gamma_{mnmn}\frac{(K-N-1)(K-N)}{N(N-1)}$$

the quadratic Q-map is

$$Q(Q(\Gamma))_{ijkl} = \Gamma_{ijkl} + \frac{2N-K}{(N-1)^2}\sum_n(\delta_{il}\Gamma_{jnkn} + \delta_{jk}\Gamma_{inln} - \delta_{ik}\Gamma_{jnln} - \delta_{jl}\Gamma_{inkn}) + \frac{4N^2+K^2-4KN+2N-K}{N^2(N-1)^2}(\delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk})\sum_{mn}\Gamma_{mnmn}$$

Given the G-map's contraction

$$\sum_n G(\Gamma)_{inkn} = \frac{K-1}{N-1}\sum_n \Gamma_{inkn}$$

the quadratic G-map and its antisymmetrization are

$$G(G(\Gamma))_{ijkl} = \Gamma_{ijkl} + \frac{K-N}{N-1}\,\delta_{jl}\sum_n \Gamma_{inkn}$$
$$\mathcal{A}\,(G(G(\Gamma)))_{ijkl} = 4\Gamma_{ijkl} - \frac{K-N}{N-1}\sum_n(\delta_{il}\Gamma_{jnkn} + \delta_{jk}\Gamma_{inln} - \delta_{ik}\Gamma_{jnln} - \delta_{jl}\Gamma_{inkn})$$

Any linear combination of these maps is therefore of the form (1.29) for which (1.31) gives its inverse map.

1.4 Practical variational second order density matrix methods

Because the known necessary and sufficient conditions for N-representability are, in general, intractable, practical v2DM methods attempt to minimize the energy as a function of the 2DM under some set of necessary, but not sufficient, conditions. Given the importance of the basis set dimension for chemically relevant results, we will deal primarily with 2-index conditions, which enable us to express the matrices in a moderately sized basis set.

In addition to Hermiticity, antisymmetry and normalization, 2-positivity conditions are imposed. This leads to the following semidefinite optimization problem, which forms the basis of all applications considered here

$$\begin{aligned} &\underbrace{min}_{\Gamma} \; E = tr\,[H\Gamma] \\ subject\ to \quad &\Gamma = \Gamma^{\dagger} \\ &\Gamma_{ijkl} = -\Gamma_{jikl} \\ &tr\;\Gamma = N(N-1) \\ &\Gamma \succeq 0 \\ &Q(\Gamma) \succeq 0 \\ &G(\Gamma) \succeq 0 \end{aligned} \tag{1.32}$$

Some techniques to solve this type of optimization problem, with semidefinite constraints, are discussed in chapter 3. The properties of Hermiticity (symmetry) and antisymmetry under electron exchange can be imposed by construction. Because the Hamiltonians to be considered are real, the 2DM may be assumed real as well.

Because the feasible set of all 2DM that satisfy the conditions imposed in (1.32) is convex, and the v2DM(PQG) method aims to minimize a linear function $E = tr\,[H\Gamma]$ over this convex set, a global minimum is always found.

The v2DM(PQG) method may be pictured as follows (figure 1.2 illustrates the approach for the P-condition). An unconstrained minimization would follow the direction of the negative Hamiltonian, which is the direction in which the energy decreases most rapidly. However, the optimum must be found within the feasible set, the set of points that satisfy all conditions in (1.32). Each of the positive semidefinite constraints defines an infinite number of constraint hyperplanes on the 2DM. The smallest convex hull of all these constraint planes then defines the set over which the energy is minimized in practice, which includes the set of N-representable 2DM as a subset. Because the dependence of the energy on the 2DM is linear, the N-representability constraints imposed

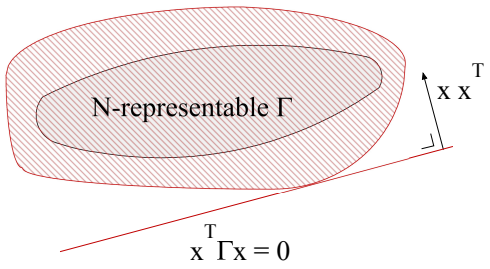

Figure 1.2: The P-condition defines an infinite number of constraint hyperplanes of the form $tr\ [\Gamma\ xx^T] = 0$ on the set of Hermitian, antisymmetric and normalized 2DM, because for any vector x: $x^T\Gamma x \geq 0$ must hold. The resulting set contains, but is bigger than, the set of N-representable 2DM. Although the multidimensionality of the problem cannot be properly depicted in the 2-dimensional plane, the Q- and G-condition similarly define an infinite number of constraint hyperplanes. For most problems, all three constraints impose non-redundant (active) bounds on the 2DM.

determine the optimum, which therefore lies on the boundary of the feasible set (figure 1.3). This implies that in order to obtain the exact energy, the imposed N-representability conditions need to capture the boundary of the N-representable set exactly, at least at the point of the lowest energy under the Hamiltonian under consideration. This realization stresses the importance of the N-representability conditions for practical v2DM methods.

Without the necessary and sufficient conditions to ensure N-representability, the optimum will be found outside of the true N-representable set and will have an energy lower than the exact energy. The method therefore generates a lower bound on the exact energy for the system within the basis set. In a finite basis set, the lower bound provided by the v2DM method may not be a hard lower bound on the exact basis set limit, but it is a lower bound to the exact energy in that basis set. As an increasing number of – active – N-representability constraints are imposed during the optimization, the energy will rise, eventually closing the gap with the exact energy in the basis set upon imposing sufficient

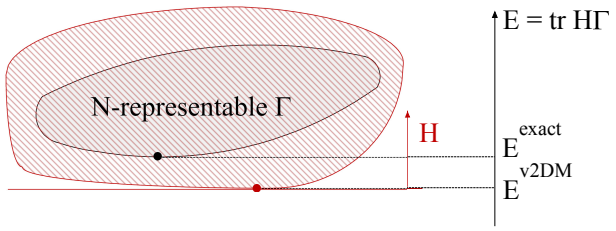

Figure 1.3: In the v2DM(PQG) method, the energy is minimized over the convex hull of all constraint hyperplanes defined by the P-, Q- and G-condition. The set of N-representable 2DM is a subset of this set. As a consequence, minimizing the energy over this set yields an energy lower than the exact energy.

N-representability constraints.[18,27] This idea forms the basis for practical v2DM approaches in which results can be improved by including a bigger number of active N-representability constraints or more stringent active constraints.

1.5 Applications to chemistry

The v2DM method can be seen as a complementary approach to wavefunction based methods, because it approaches the electron correlation problem from a completely different perspective. The increasing accessibility and performance of semidefinite optimization programs has renewed interests in applications of v2DM theory in the past two decades, including applications to chemistry.[14,15,21,37] Nonetheless, most applications on chemical systems have focused on reproducing correct energies. Therefore, we aim to provide chemical insight into the effects of approximate N-representability on chemical systems by evaluating necessary N-representability constraints on the PES of several small molecules, focusing on computationally tractable 2-index constraints.

Section 1.5.2 discusses the numerical accuracy and stability of v2DM theory from a chemical point of view. It demonstrates a fundamental shortcoming, namely molecular dissociation into unphysical fractionally charged atoms or

molecules under 2-positivity conditions. Although this is a serious shortcoming in itself, it signals a more profound problem, which affects many chemical properties. Moreover, it seems to be a persistent problem: even 3-index constraints do not solve it adequately. Section 1.5.3 therefore traces the origin of this problem and proposes additional N-representability constraints to solve it. Section 1.5.4 analyzes the effect of the proposed constraints on several molecular applications. Since molecular dissociation is so badly described, section 1.5.5 examines the concepts of size-consistency and separability in practical v2DM methods.

1.5.1 Computational details

All v2DM calculations are done with the cc-pVDZ basis set, unless otherwise specified. The molecules are constrained to singlet states (conditions 2.5.1 in chapter 2 for the v2DM method), unless stated otherwise.

The PES are generated from single point calculations, using our own classical barrier method (section 3.4 of chapter 3) to carry out the semidefinite optimization (1.32) and Molpro for generating reference CASSCF and MRCI calculations.[38]

In the calculations on the 14-electron isoelectronic series in section 1.5.2, the active space of the full-valence CASSCF comprises 10 electrons and all 8 valence orbitals, and the doubly-occupied inactive orbitals (mostly 1s core) were also optimized. The MRCI calculations were performed subsequently, with the full-valence CASSCF as a reference. The core (1s) orbitals were kept frozen in the MRCI expansion.

1.5.2 Strengths and failures of 2-index N-representability constraints in chemical applications

Ideally, a numerically reliable approximative ab initio method

computes the energy and other chemical properties accurately. If not exact, its potential energy surface (PES) is parallel to the exact PES, such that it still leads to correct spectroscopic constants;

provides the same level of accuracy within each finite basis set;

has stable errors with respect to the number of electrons, the atomic number, spin state and molecular geometry;

is size-extensive and size-consistent.

These aspects only establish its numerical reliability. In reality, there is always a trade-off between numerical accuracy and computational speed. Therefore, the ideal ab initio method has all these desirable properties while requiring a computation time that scales up significantly better than FCI. The computational side of the v2DM method is discussed in chapter 3 and we focus here on its numerical accuracy, which is directly related to the strength of the N-representability conditions imposed.

The PES for a set of N_2 isoelectronic molecules under the P-,Q- and G condition (figure 1.11) and Be_2 calculated in different basis sets (figures 1.8 and 1.9), allow us to make the following observations, which will be a starting point for further examination in sections 1.5.3 to 1.5.5.

For near-equilibrium structures, 2-positivity conditions give a fair, albeit overestimated, description of electron correlation. Because they take electron correlation into account, they capture the basic chemistry in a qualitative manner, even in multireference cases that are difficult to describe using wavefunction based methods. Several examples to illustrate this point will be discussed in upcoming sections. For instance, the 2-positivity conditions are able to describe the most important chemical differences between several 14-electron diatomic molecules, including the kinetically stable pseudo bound state of O_2^{2+}, which is a typical failure case for ab initio methods such as MP2. They capture the van der Waals and covalent bonding in Be_2, although they exaggerate it. They also correctly produce a potential energy well for the near-equilibrium geometries of the F_3^- ion, although the nature of its bonding remains a source of debate.[39–42]

However, the lower bound method's overestimation of correlation is reflected in all chemical properties. Because of this, it is expected to give

too low energies for transition states compared to the equilibrium and therefore underestimate energy barriers

too low dissociation energies

wrong electron affinities and ionization energies

too high polarizabilities

too low band-gaps and overestimated charge-transfer energies

which are problems that other density and density matrix methods face as well. The most important of these consequences will be discussed in more detail in the next sections. Since 2-positivity conditions are only exact for systems with up to two electrons, they cannot provide quantitative accuracy in general, and they are expected to perform best on systems with configurations that resemble a two or three particle or hole system.

The accuracy of the optimal energy under 2-positivity conditions is most improved by the G-condition. The P-condition only gives a very poor approximation (table 1.3), because it basically forces all electron pairs into the lowest energy eigenfunction of the second order reduced Hamiltonian (see chapter 3 section 3.2). The 3-index constraints T_1 and T_2 decrease the error significantly for near-equilibrium structures, up to mHartree precision, which agrees with findings by Nakata et al. and Mazziotti et al.[37,43] However, all of these constraints improve the energy much less in the dissociation limit.

The errors in the v2DM(PQG) energy may originate both from the potential energy and from kinetic energy, and cancel out to some extent. The v2DM(PQG) method may overestimate the potential energy and underestimate the kinetic energy, as illustrated by Be_2 in the 6-31+G* basis set (figure 1.4) or underestimate the potential energy and overestimate the kinetic energy, as illustrated by BeB^+ in the D95V basis set (figure 1.6). Separating the total energy into its kinetic energy and potential energy contributions may therefore produce bigger errors in each of these contributions than in the total energy.

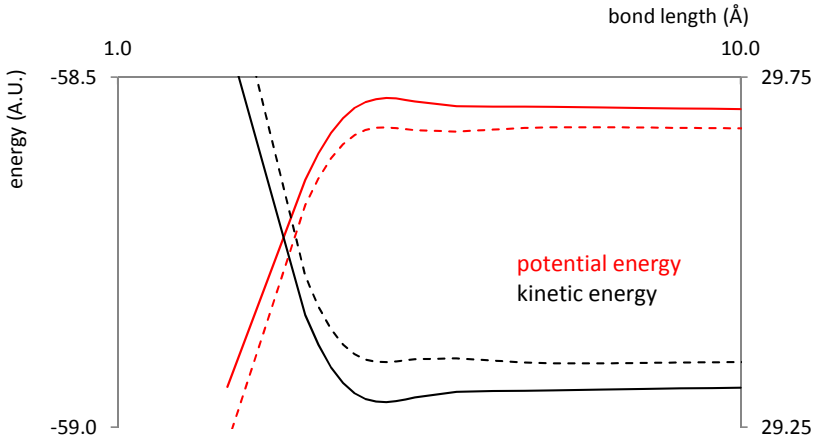

Figure 1.4: The v2DM(PQG) method (solid lines) overestimates the potential energy of Be_2 in the 6-31+G* basis set and overestimates the kinetic energy compared to FCI(FC) calculations (dotted lines), such that the two errors partially cancel out in the total energy. The ratio of potential and kinetic energy is consistently slightly higher than that of FCI(FC) calculations (figure 1.5).

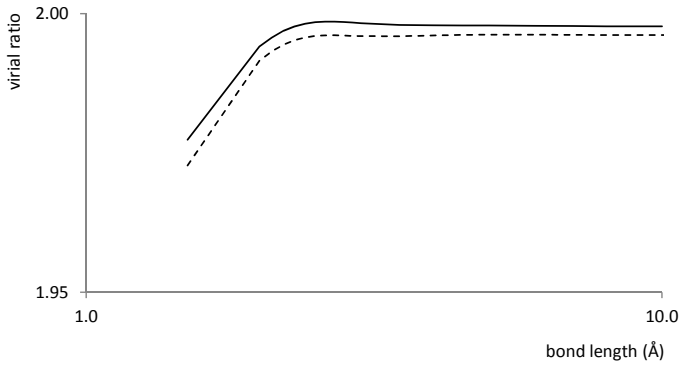

Figure 1.5: The v2DM(PQG) method (solid lines) gives consistently higher virial ratios for Be_2 in the 6-31+G* basis than FCI(FC) calculations (dotted lines).

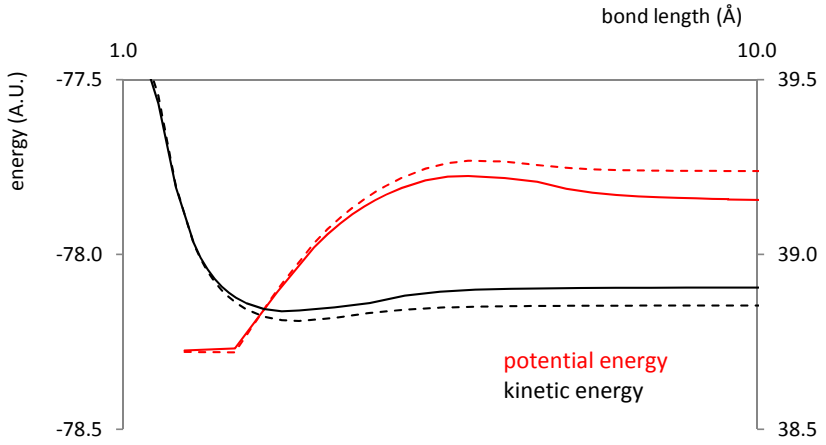

Figure 1.6: The v2DM(PQG) method (solid lines) underestimates the potential energy of BeB^+ in the D95V basis set and overestimates the kinetic energy compared to FCI(FC) calculations (dotted lines), such that the two errors partially cancel out in the total energy. Nonetheless, the ratio of potential and kinetic energy agrees well with FCI(FC) calculations (figure 1.7).

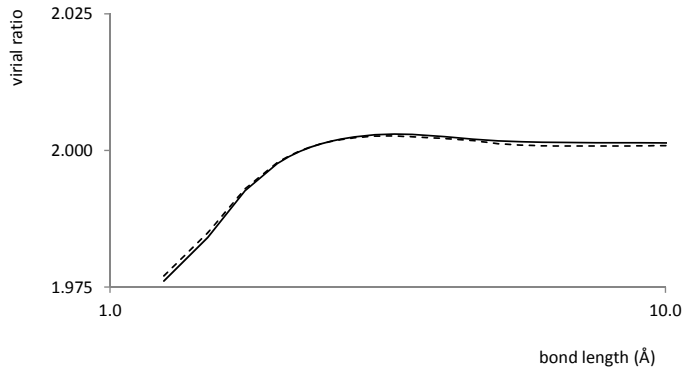

Figure 1.7: The v2DM(PQG) method (solid lines) gives highly similar virial ratios for the BeB^+ in the D95V basis set to FCI(FC) calculations (dotted lines) even though differences in kinetic and potential energy are non-neglible.

The improvement of the v2DM(PQG) method's accuracy with the basis set is not always consistent with that of an exact calculation. Even though the basis set dependence is rather stable for atomic systems,[44] there are important differences between PES for Be_2 in different basis sets calculated with the v2DM(PQG) method and with FCI(FC). Moreover, the differences between the PES for the two methods are not consistent with the choice of basis set (figures 1.8 and 1.9). The v2DM(PQG) method tends to produce a more strongly bound PES than FCI(FC) (figure 1.10), even in basis sets in which FCI(FC) does not describe Be_2 as a stably bound molecule, such as in the 6-31+G* basis set. In this basis set the distance between the two PES is significantly larger than in the other basis sets considered, which may indicate that the v2DM(PQG) method is more sensitive to the inclusion of diffuse functions in the basis set. The distance between the v2DM(PQG) and FCI(FC) PES is smallest for basis sets that do not contain polarization or diffuse functions.

Admittedly, wavefunction based results for the combined van der Waals-covalent bonding present in the Be dimer also depend heavily on the basis set. High-quality PES for the Be dimer not only require a multireference method to describe the bonding attributable to the near degeneracy of the 2s and 2p orbitals, but also a basis set that includes f- or higher angular momentum basis functions.[45–48]

Although in principle the v2DM method may produce an incorrect 2DM with correct energy, we do not observe any obvious inconsistencies between the energy of the variationally optimized 2DM and other chemical properties. In theory, a single constraint on the energy, $tr\ H\Gamma \geq E_0(H)$ with H the Hamiltonian of the system under consideration, would suffice to obtain a 2DM with the exact energy. However, the optimal 2DM under this hypothetical constraint would lie in the intersection between the hyperplane described by $tr\ H\Gamma = E_0(H)$ and the convex hull determined by the other necessary N-representability constraints imposed. This does not necessarily lead to the exact N-representable 2DM, which would lie in the intersection of the same hyperplane with the true N-representable set.

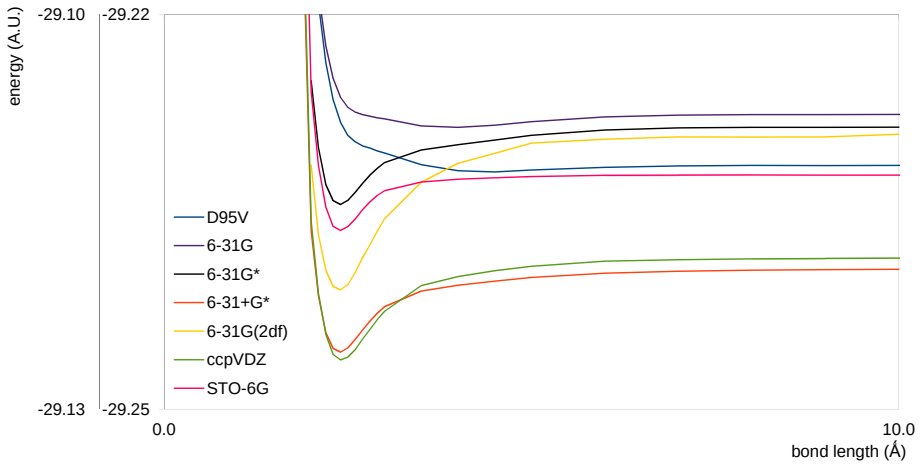

Figure 1.8: The shape of the v2DM(PQG) PES of Be_2 depends heavily on the basis set, but not in the same way as the FCI(FC) PES given in figure 1.9. The v2DM(PQG) PES tends to overestimate the combination of van der Waals attraction and chemical bonding between the Be atoms. The STO-6G PES is depicted on a secondary axis because it is much higher in energy.

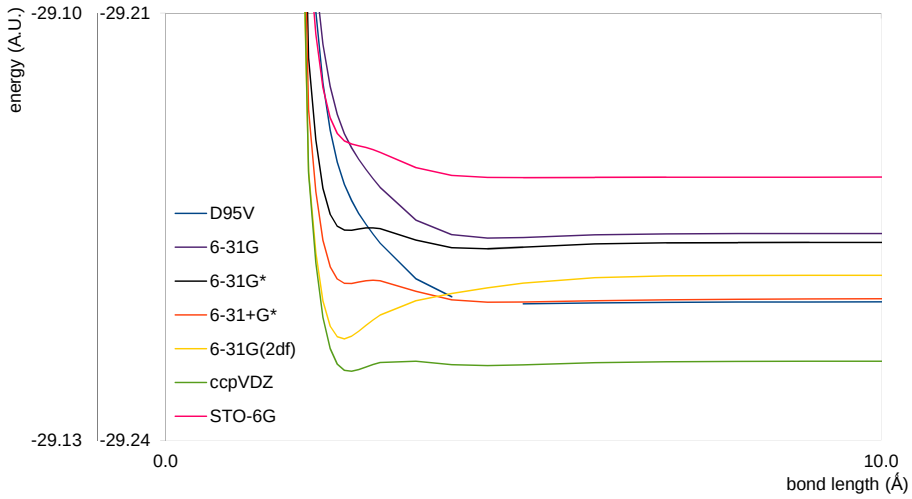

Figure 1.9: The shape of the FCI(FC) PES of Be_2 also depends heavily on the basis set. Some smaller basis sets do not reproduce the potential energy well due to the combination of van der Waals attraction and chemical bonding between the Be atoms. The STO-6G PES is depicted on a secondary axis because it is much higher in energy.

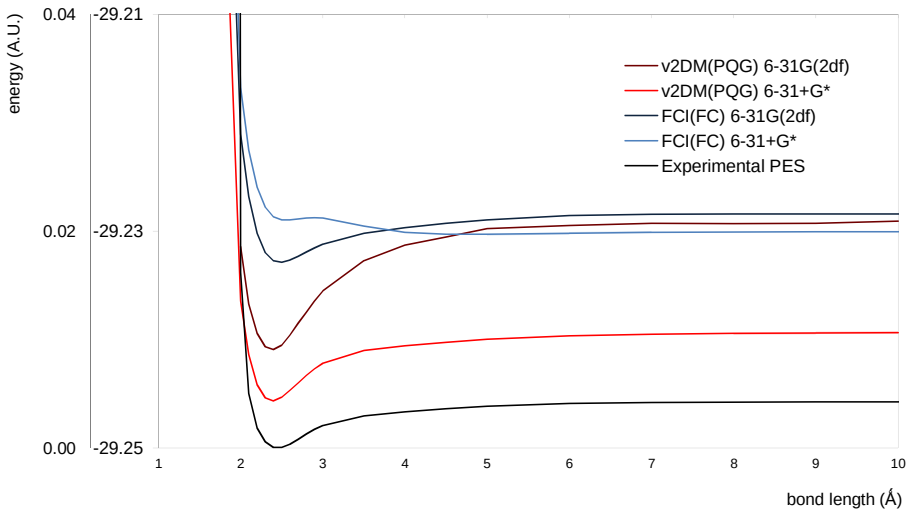

Figure 1.10: Comparison of the Be_2 PES of the v2DM(PQG) method with FCI(FC) in the biggest basis sets considered, and an experimental PES generated from spectroscopic data[47] (only relative energies are shown), confirms that the v2DM(PQG) method exaggerates the bonding. It overestimates the dissociation energy by a factor two in the 6-31G(2df) basis. In the 6-31+G* basis, dissociation energy is considerably smaller, but FCI(FC) does not even yield a stably bound minimum in this basis.

This hypothetical example justifies the need to study properties other than the energy in order to make a full assessment of the v2DM method.

Nonetheless, no such inconsistencies between the energy and other chemical properties appear in our v2DM(PQG) calculations. In fact, shortcomings in our calculated chemical properties are signaled by increasing errors in the energy. In particular, both the energy and chemical properties calculated with the v2DM(PQG) method generally agree fairly well with MRCI properties for near-equilibrium geometries and the increasing errors in the energy towards dissociation prove to be an indication of serious shortcomings in other chemical properties as well (tables 1.1 and 1.2).

The 2-positivity conditions fail dramatically in describing structures far from their equilibrium geometry and lead to unphysical dissociation limits.[49,50] No accounts of this problem have been made in existing v2DM studies of potential energy surfaces, which focus on the energy and mostly study homonuclear molecules up to relatively short bond lengths, 2 or 3 Å.[51,52] Our v2DM(PQG) calculations on the heteronuclear diatomic molecules CO, CN^- and NO^+ nonetheless indicate a serious failure of the v2DM(PQG) method: their energy, atomic charges and dipole moment diverge from their MRCI counterparts as the bond length increases (figure 1.12, tables 1.1 and 1.2). The dipole moments calculated with the v2DM(PQG) method do not correspond to the correct dissociation limit consisting of two neutral or a neutral and a singly charged atom, but to dissociated atoms with a residual non-integer charge. This shortcoming is confirmed by the appearance of non-integer atomic Mulliken charges in the dissociation limit (table 1.2). For instance, NO^+ dissociates into $N^{+0.47} + O^{+0.53}$ instead of $N^{0.0} + O^{+1.0}$ and CN^- into $C^{-0.60} + N^{-0.40}$ instead of $C^{-1.0} + N^{0.0}$. Even the CO molecule carries minor atomic charges in the dissociation limit, which produce a significant dipole moment due to the large spatial separation. The heteronuclear diatomics, on the other hand, dissociate into correct products under 2-positvity because of symmetry.

	MRCI (10 Å)	DM2 (10 Å)	DM2 (20Å)
NO^+	22.39	-0.11	-0.28
CN^-	25.83	7.01	13.38
CO	0.00	-0.90	-1.71

Table 1.1: v2DM(PQG) dipole moments in the dissociation limit (in Debije, origin chosen in the centre of mass), do not correspond to the integer-charged dissociation products, but to dissociated atoms with a fractional residual charge.

	Mulliken population (20 Å)	fitted minimum to atomic sum graph
N_2	7.00/7.00	7.00/7.00
O_2^{2+}	7.00/7.00	7.00/7.00
NO^+	6.53/7.47	6.51/7.49
CN^-	6.60/7.40	6.58/7.42
CO	5.98/8.02	5.98/8.02

Table 1.2: The Mulliken populations of the dissociated molecule (20 Å) are remarkably similar to the minimum populations obtained by fitting a polynomial to the graph of summed atomic energies as a function of the population on one of the atoms (and for a total of 14 electrons)

.

The appearance of such fractional residual charges in the dissociation limit of heteronuclear diatomics indicates a fundamental flaw in the 2-positivity conditions. Fractionally charged dissociation species can occur naturally when several symmetry-equivalent charged species are formed upon dissociation, but they are unjustified in heteronuclear diatomic molecules.

The v2DM(PQG) method does not adequately describe systems with a fractional number of electrons.[50,53] This shortcoming is the origin of its failure to describe structures far from equilibrium geometry. Theoretically, a state with a fractional number of electrons can only arise from an ensemble of states with

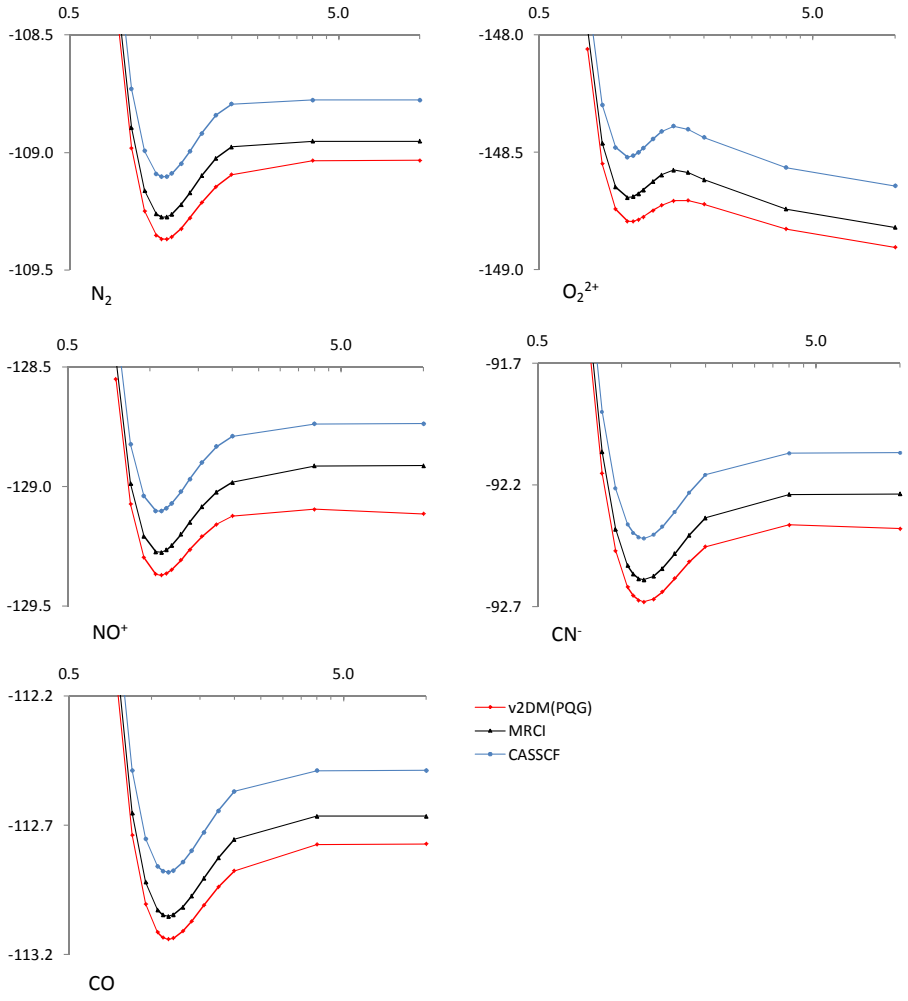

Figure 1.11: The v2DM(PQG) PES has a similar shape to the PES calculated with MRCI and CASSCF for the homonuclear O_2 and N_2, but has a different dissociation limit behavior than the MRCI and CASSCF PES for the heteronuclear molecules NO^+ and CN^-. This is confirmed by the relative energy differences shown in figure 1.12.

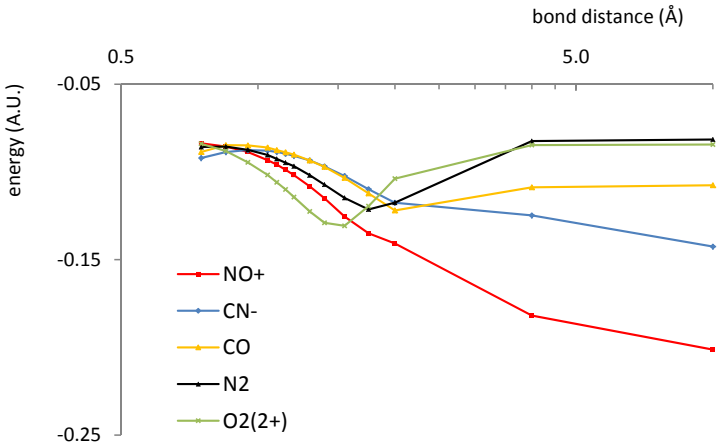

Figure 1.12: The errors between the v2DM(PQG) and MRCI energy increase towards dissociation for the heteronuclear diatomics NO^+, CN^- and CO, yet decrease towards dissociation for the homonuclear diatomics.

different, but integer, numbers of electrons. It thus lies in the convex hull of pure states with integer numbers of electrons. Because the energy is linear in the 2DM, the ground state for an ensemble with a fractional number of electrons must be a linear combination of ground states with integer numbers of electrons. This always leads to a dissociated state with an integer number of electrons, if no degeneracies are present.

The observed fractionally charged dissociation products for CO, NO^+ and CN^- may thus arise from a convex relation between the energy and the number of electrons, when the number of electrons is regarded as a continuous quantity. Indeed, v2DM(PQG) energies for most atoms and molecules are convex functions of the number of electrons (figure 1.13), when the number of atoms is regarded as a continuous quantity in the setup (1.32). The sum of the energies of the constituent atoms of the molecule is therefore also convex in-between consecutive integer numbers of electrons. As a consequence, the molecule may reach a lower energy by dissociating into atoms with a fractional number of electrons. This idea is illustrated in figure 1.14. In fact, we find a remarkable correspondence between the estimated minimum of the sum of the energies of the constituent

atoms and the observed charges in the molecular dissociation limit (table 1.2). Of course, the N-representability conditions imposed here only rigorously hold for ensembles of N-electron states, and do not allow for ensembles of states with other electron numbers as well. In this sense, applying this approach to systems with fractional electron number is physically not justified. However, the numerical data presented here suggest that the system in the dissociation limit nonetheless acts like a combination of systems with fractional electron number under the applied N-representability conditions (the P-, Q- and G-condition on the molecular system imply similar conditions on the subsystems). Ultimately, this shortcoming is a consequence of imposing necessary, but not sufficient, N-representability conditions.

The difficulty of approximate v2DM methods to describe structures far from equilibrium geometry is not only present at the level of 2-positivity conditions. Although several studies on 3-index conditions conclude that they greatly improve accuracy,[37,43] trial calculations on N_2 in a minimal basis set show that 3-positivity conditions much improve the energy over 2-positivity conditions around equilibrium geometries, but much less so in the dissociation limit. Fractionally charged dissociation products still turn up under 3-positivity conditions, as the energy remains a convex function of the number of electrons in between integer occupations. Imposing 3-positivity conditions may lessen the convexity, but does not remove it completely (table 1.3).

The convexity of the energy for systems with fractional numbers of electrons has far-reaching implications for chemical applications, not only for dissociation. Although its effects may not be as obvious as in the dissociation limit, it affects chemical properties at other geometries as well. Because the dissociation limit energy is too low, dissociation energy and therefore spectroscopic constraints are wrong. Similar problems to those occurring in the dissociation limit affect reaction intermediates, which involve partially broken and formed bonds. The reaction barrier can thus be expected to be too low, which leads to an inaccurate prediction of its kinetics. At any geometry, polarizability can be expected to be

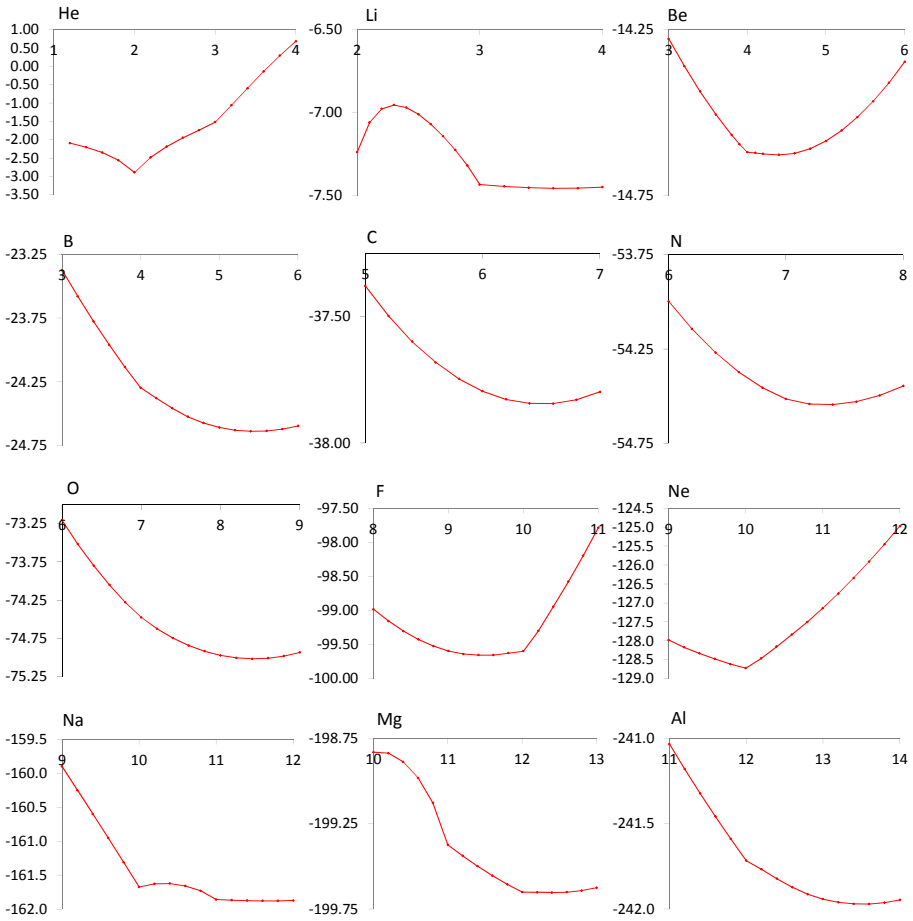

Figure 1.13: The v2DM(PQG) energies are convex functions of the number of electrons for most atoms, except, it seems, for atoms with less than two electrons and similar occupations – that is, atomic configurations with one or more filled shells and less than two electrons in the next shell.

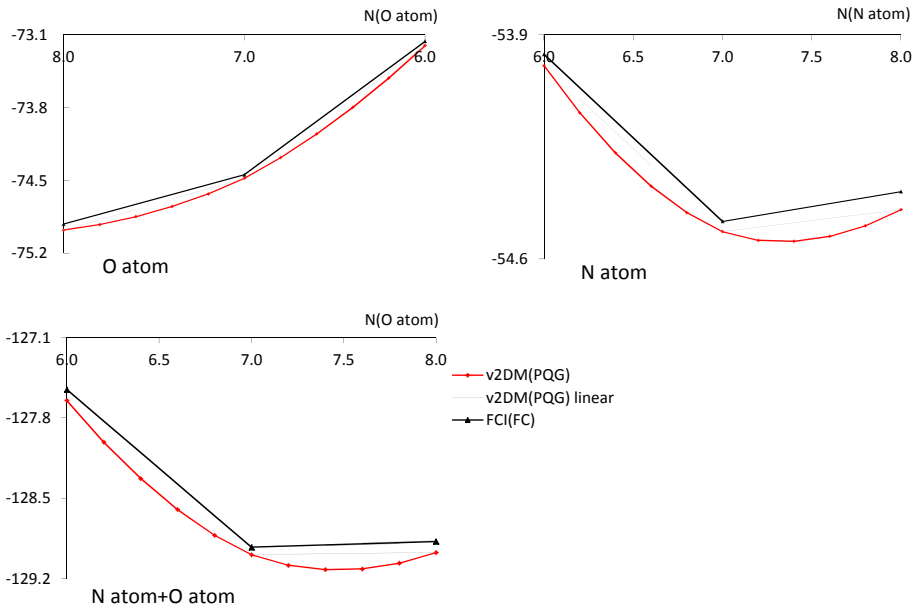

Figure 1.14: Because the energies of both the oxygen and the nitrogen atom are strictly convex functions of the fractional number of electrons, the sum of the atomic energies is also convex, yielding a minimum for fractional occupations on each atom. This explains the unphysical fractional charges on the N and O atom of the dissociated NO^+ molecule.

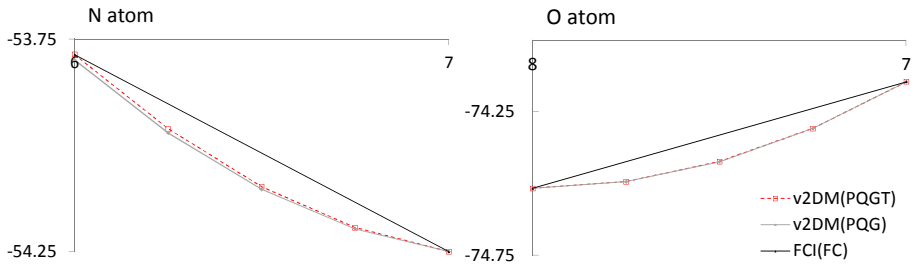

Figure 1.15: Under P-,Q-,G- and T-conditions, the energy is still a convex function of the number of atoms, when the number of atoms is considered as a continuous quantity. In the STO-6G basis, the added T^1 and T^2-constraints only change the energy by a small amount. Consequently, the dissociation problem persists even under the three-index T-constraints.

	R= 1.1365 Å		R= 20.0000 Å	
	singlet	any spin	singlet	any spin
P	-375.1036	-375.1037	-393.5293	-393.5293
PQ	-129.6224	-129.6515	-132.3706	-132.3706
PG	-128.6968	-128.7026	-128.6766	-128.6892
PQG	-128.6421	-128.6440	-128.5190	-128.5221
PQGT		-128.6268		-128.5167
FCI	-128.6256		-128.3950	

Table 1.3: Comparing the different 2-positivity conditions for NO^+ (in the STO-6G basis set), the G-condition affects the energy stronger than the Q-condition. The P-condition only gives energies that are about 3 times too low! Near the equilibrium geometry, the T-conditions (T^1 and T^2, see paragraph 1.3.3) raise the energy significantly, but the difference from FCI remains substantial in the dissociation limit. v2DM(PQGT) energies were only calculated without imposing spin constraints (column 'any spin').

too high, because an additional negative charge produces a larger than usual decrease in energy. The absence of discontinuities in the energy at integer electron numbers produces incorrect electron affinities (EA) and ionization energies (IE). The EA and IE are directly related to the right and left derivative of the energy to the number of electrons for the neutral atom with N_0 electrons:

$$EA = -\left(\frac{\partial E}{\partial N}\right)_{N \gtrdot N_0} \tag{1.33}$$

$$IE = -\left(\frac{\partial E}{\partial N}\right)_{N \lessdot N_0} \tag{1.34}$$

The lack of discontinuities in the energy as a function of the number of electrons in v2DM methods (figure 1.13) implies that the EA equals the IE, which may have a profound influence on chemical properties.

The 2-positivity conditions are not size-consistent, nor size-extensive. Because the 2DM is not an additively separable quantity, it can be expected that necessary but insufficient conditions for N-representability lead to an approach that fails to

be size-consistent and size-extensive.[54] Calculations on the dissociation limits of the isoelectronic 14-electron series estimate the extent to which size-consistency is violated in typical calculations. The differences between the energy of the dissociated NO^+ and CN^- molecule and the dissociation products calculated separately is of the order of 10^{-1} Hartree (1.4). However, part of this difference is due to the fact that it compares the energy of the fractionally charged dissociation products with the integer charged correct dissociation products. But even when atoms with the same fractional charges that arise in the dissociated molecule are taken as a reference, a significant energy difference remains.
Therefore, the method is not size-consistent, which also rules out the possibility of size-extensivity. A more detailed consideration, taking the structure of the 2DM in the dissociation limit into account, is made in section 1.5.5. Nakata et al. have made an independent study of size-extensivity and size-consistency in v2DM(PQG) theory.[51,55] They consider the dissociation of mainly homonuclear diatomics in a minimal STO-6G basis, which dissociate into the correct products by symmetry. Still, they observe violations of size-extensivity and size-consistency, which agrees with our findings.

To conclude, the v2DM(PQG) method captures electron correlation, but overestimates it by a fair amount. The method suffers two main shortcomings: a failure to describe electronic structure for non-equilibrium geometries and a failure to be size-consistent. Near equilibrium, results agree fairly well with accurate wavefunction based methods such as MRCI and CASSCF, and are comparable to CCSD in terms of accuracy, in agreement with findings by Mazziotti et al. and Nakata et al.[14,51,52,56] However, they rapidly deteriorate with bond length. In the dissociation limit, the v2DM(PQG) method tends to produce unphysical fractionally charged dissociation products with much too low an energy, as a consequence of its convex energy-occupancy relationship. This is a profound problem and its convergence to the correct behaviour by including increasingly higher order index constraints may not be as fast as previously believed, since including 3-index constraints still leaves a significantly bigger discrepancy in the

	$E^{AB} - E^A - E^B$	$E^{AB} - E^{A^{+\nu}} - E^{B^{-\nu}}$
N_2	-0.0035	-0.0035
O_2^{2+}	-0.0045	-0.0045
NO^+	-0.1357	-0.0041
CN^-	-0.0797	-0.0028
CO	-0.0036	-0.0034

Table 1.4: The v2DM(PQG) energy differences between the molecular 2DM energy at 20 Å and the sum of the atomic 2DM energies for the correct dissociation products, $E^{AB} - E^A - E^B$, are of the order 10^{-1} Hartree for the heteronuclear diatomics. This is mainly due to the fact that the molecules dissociate into fractionally charged dissociation products with too low energies under 2-positivity constraints. However, the energy differences between the molecular 2DM energy at 20 Å and the sum of the atomic v2DM(PQG) energies with the same populations as the molecular dissociation products, $E^{AB} - E^{A^{+\nu}} - E^{B^{-\nu}}$, is still non-negligible. The method is therefore not size-consistent.

dissociation limit than near equilibrium.

Because the 2DM is not a separable quantity, approximate N-representability conditions lead to size-consistency defects. However, attempts to formulate the v2DM approach in a size-consistent manner based on cumulants are not straightforward, because N-representability manifests itself directly in terms of the 2DM, and not in terms of the cumulant. For this reason, Kutzelniggs size-consistent cumulant based approach relies on a Hartree-Fock reference.[54,57] Although size-consistency will thus be difficult to achieve in any approximate v2DM method, we aim to find additional N-representability constraints that tackle the convex energy-occupancy relationship. This particular problem is a profound and fundamental problem as it arises in very similar forms in any practical density and density matrix based method.

1.5.3 Additional subspace energy constraints to correct molecular dissociation

Heteronuclear diatomic molecules do not dissociate into fractionally charged atoms. As simple as this fact is, it is far less straightforward to establish in reduced density matrix theories. Yet, such fundamental physical properties are needed to make them applicable to geometries other than the equilibrium structure. Non-equilibrium structures like molecules with stretched or partly broken bonds, such as reaction intermediates or dissociation products, play an important role in chemical processes. Despite numerous efforts, they still cause problems in Density Functional Approximations[58–61] and Density Matrix Functional Theory.[62–64] The previous results show that the v2DM(PQG) method also fails in this respect. Although this method cannot be expected to be fully size-consistent because the 2DM is not a separable quantity under a strict subset of N-representability conditions,[65] its failure in describing dissociating chemical systems is dramatic.

But, unlike DFT and DMFT, there is a straightforward approach to solve the problem because the 2DM fully determines the energy in a known manner. Here, we exploit this property and introduce linear constraints on the energy of subspaces of the one-particle basis space for the molecule,[53] defined as the set of basis functions centered on a particular atom, to solve the dissociation problem. These subspace constraints are a physical expression of the notion of separability in chemistry,[66] and can be generalized to subspaces with any other topology.

The following paragraphs give a theoretical background on the subspace constraints and illustrate their effectiveness by applying them to the PES of the 14-electron diatomic molecules considered in previous section. A concept similar to our subspace energy constraints has been adopted by Shenvi et al. to generate active-space constraints on the 2DM.[67]

Theoretical background on the subspace energy constraints

The set of necessary conditions (1.17)-(1.23) can be extended with linear subspace constraints to improve the description of long-distance interactions. As shown in the previous section, the failure of v2DM theory to describe long-range interactions can be attributed to the strictly convex relationship between the energy and the number of electrons on the atom, considered as a continuous quantity, which is ultimately a result of imposing necessary but not sufficient conditions for N-representability. In any exact theory, the relationship between the energy and the number of electrons is piecewise linear.[68] Improper fractionally charged dissociation products cannot occur, because the piecewise linear relationship between energy and electron number ensures the minimum energy always corresponds to an integer electron number for a non-degenerate state.[68] Separability constraints offer a computationally affordable way to impose this behavior. They aim to correct the dissociation problem by forcing the energy of (poly-)atomic subspaces in the molecule to lie above the piece-wise linear graph determined by the v2DM(PQG) energies for integer number of electrons.

The 1DM and 2DM for a subspace A are obtained by projecting onto the subspace A. Suppose the subspace A is spanned by a – not necessarily orthonormal – basis $\{\phi_a, \phi_b, \phi_c, \phi_d, \ldots, \phi_{K^A}\}$. Projecting the creation and annihilation operators in the orthonormal molecular sp basis $\{\phi_i, \phi_j, \phi_k, \phi_l, \ldots, \phi_K\}$ onto the subspace gives an expression for the subspace 1DM and 2DM $\tilde{\gamma}^A$ and $\tilde{\Gamma}^A$ in terms of the 1DM and 2DM γ and Γ in the orthonormal molecular basis:

$$\tilde{\gamma}^A_{ab} = \sum_{ij}^{K} w_{ai} w_{bj} \gamma_{ij} = \frac{1}{N-1} \sum_{ijk}^{K} w_{ia} w_{bj} \Gamma_{ikjk} \tag{1.35}$$

$$\tilde{\Gamma}^A_{abcd} = \sum_{ijkl}^{K} w_{ai} w_{bj} w_{ck} w_{dl} \Gamma_{ijkl} \tag{1.36}$$

The coefficients w_{ai} follow from the projection $\sum_{ab} |a\rangle\langle b| (S^{A^{-1}})_{ab}$ from the K-dimensional orthonormal MO basis onto the K^A-dimensional non-orthogonal

subspace basis

$$w_{ai} \equiv \sum_{b}^{K^A} \sum_{c}^{K} (S^{A^{-1}})_{ab} S_{bc} C_{ic} \tag{1.37}$$

with S^A the K^A dimensional overlap matrix between the non-orthogonal basis functions of subspace A, S the K-dimensional overlap matrix between basis functions of all subspaces that span molecular basis space and C_{ic} the expansion coefficient of the MO i in terms of the basis function c.

The expectation values of a one-electron operator $\hat{h}^A$ and a two-electron operator $\hat{H}^A$ expressed in the subspace A are then

$$\langle \hat{h}^A \rangle = \sum_{ab} \tilde{\gamma}_{ab} \langle b|\hat{h}^A|a\rangle$$
$$\langle \hat{H}^A \rangle = \sum_{abcd} \tilde{\Gamma}_{abcd} \langle cd|\hat{H}^A|ab\rangle$$

For instance, the number of electrons in the subspace A is

$$N^A = \sum_{ab} \tilde{\gamma}^A_{ab} S^A_{ab}$$

These expectation values can also be expressed directly in terms of the full 2DM in the orthonormal MO basis, since (1.35) and (1.36) express the subspace 1DM and 2DM in terms of the full 1DM and 2DM. Therefore

$$\langle \hat{h}^A \rangle = \sum_{ij} \gamma_{ij} h^A_{ij} = \sum_{ij} \gamma_{ij} \langle j|\hat{h}^A|i\rangle$$
$$\langle \hat{H}^A \rangle = \sum_{ijkl} \Gamma_{ijkl} H^A_{ijkl} = \sum_{ijkl} \Gamma_{ijkl} \langle kl|\hat{H}^A|ij\rangle$$

with

$$h^A_{ij} = \sum_{ab}^{K^A} w_{ai} w_{bj} \tilde{h}^A_{ab} \sum_{ab}^{K^A} w_{ai} w_{bj} \langle b|\hat{h}^A|a\rangle$$
$$H^A_{ijkl} = \sum_{abcd}^{K^A} w_{ai} w_{bj} w_{ck} w_{dl} \tilde{H}^A_{abcd} = \sum_{abcd}^{K^A} w_{ai} w_{bj} w_{ck} w_{dl} \langle cd|\hat{H}^A|ab\rangle$$

The subspace dependence can thus be incorporated into the operator in MO space, avoiding the necessity for a transformation of the 2DM to the subspace each time a subspace expectation value needs to be calculated.

In the previous section, it was observed that the v2DM(PQG) method for a system of non-interacting units acts much like separate v2DM(PQG) procedures on each of the non interacting units while allowing them to have a fractional portion of the total number of electrons. Therefore imposing a piecewise linear energy-occupancy relationship on each of those units solves the dissociation problem. We will call this type of constraint a subspace energy constraint. The theoretical justification for such a constraint is built on the property that any pair of subspace 1DM and 2DM (1.35,1.36) must be derivable from an ensemble of N-representable 2DM corresponding to the subspace population N^A

$$\begin{aligned} \tilde{\gamma}_{ab} &= \sum_i x_i \gamma_{ab}^{N_i} \\ \tilde{\Gamma}_{abcd} &= \sum_i x_i \Gamma_{abcd}^{N_i} \end{aligned} \tag{1.38}$$

The weights $\{x_i\}$ represent a physical ensemble corresponding to a fractional number of electrons N^A if they satisfy

$$0 \leq x_i \leq 1 \quad i = 0, \ldots \infty \tag{1.39}$$

$$\sum_{i=0}^{\infty} x_i = 1 \tag{1.40}$$

$$\sum_{i=0}^{\infty} x_i N_i = N^A \tag{1.41}$$

for integer N_i. We shall refer to this property as *fractional N-representability*,[53] which generalizes the concept of integer-N representability.

Because the 2DM that make up the ensemble (1.38) with fractional electron number must be integer-N representable, any Hamiltonian H^A acting on the subspace A, expressed here in MO basis space, imposes a necessary condition on the 2DM expressed in MO basis

$$tr[H^A \Gamma] \geq \underbrace{min}_{\{x_i\}} \sum_{i=0}^{\infty} x_i E_{N_i}^{A,exact} \geq \underbrace{min}_{\{x_i\}} \sum_{i=0}^{\infty} x_i E_{N_i}^{A,2DM} \tag{1.42}$$

where the $E_{N_i}^{exact}$ are exact ground state energies for the Hamiltonian H^A acting on a system with an integer number N_i electrons, which are higher than the

v2DM energies for the same system. The objective will therefore be to minimize the molecular energy subject to (1.32) and (1.39)-(1.41).

Even though the indices in formulae (1.39)-(1.41) run over all positive integers, two low lying states with smaller and bigger integer particle number than the fractional number N_A determine the ground-state energy of an ensemble with N_A electrons. If the set of energies $E_{n_i},\ i = 1, \infty$ with $N_A \in [N, N+1]$ is a convex set, the lowest energy ensemble is a linear combination of the states with N and N+1 electrons (figure 1.16). The assumption of convexity of the set of energies is reasonable; we have never encountered a violation. Consequently, only two indices i=N and i=N+1 in equations (1.39)-(1.41) are practically relevant. All other weights x_i with $i \neq N, N+1$ are zero and the weights x_N and x_{N+1} are completely determined by relation (1.41), which implies that $Nx_N + (N+1)x_{N+1} = N_A$. Since only x_N and x_{N+1} are non-zero, expressions (1.40) and (1.41) are bounded. In summary, under the above assumptions, the constraint reduces to

$$tr[H^A\Gamma] \geq \quad (N+1-N^A)E_N + (N^A - N)E_{N+1} \tag{1.43}$$

To compose the constraint equations (1.43), the v2DM(PQG) energy of the subspace for N, N+1 electrons needs be calculated before the actual molecular v2DM(PQGs) calculation. Additional background on the subspace constraints can be found in Verstichel et al.[49] and van Aggelen et al.[53]

Computational details

The subspace energy constraints are simple linear inequality constraints, which can be incorporated in a barrier method for semidefinite programming (chapter 3, section 3.4) by an additional scalar barrier term for the inequality. In the following numerical application we reconsider the set of 14-electron diatomic molecules in the Cartesian cc-pVDZ basis of section 1.5.2. The v2DM(PQG) subspace energies are calculated under 2-positivity conditions, but no conditions on spin are imposed, whereas the molecular system is constrained to a singlet.

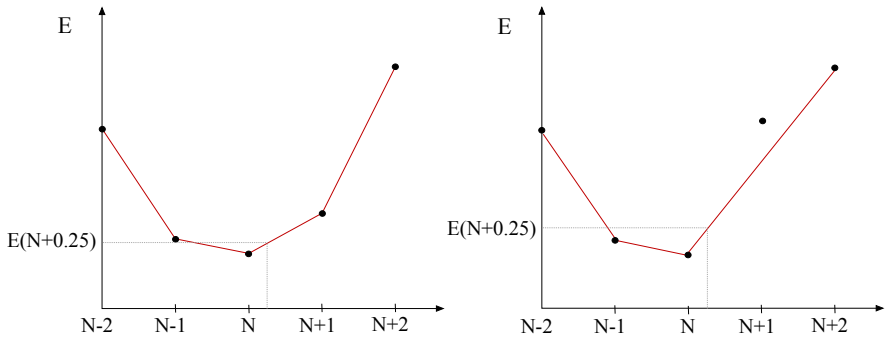

Figure 1.16: An ensemble corresponding to a fractional number of electrons is composed of integer-electron states, each with a positive weight. The lowest energy ensemble is then determined by the two lowest-energy integer occupied states that can form such an ensemble. When the energies for integer electron numbers form a convex set, the states with the nearest smaller and bigger integer electron number determine the ensemble. This situation is pictured on the left, the non-convex case on the right. In practice, we have only encountered convex sets of integer-N energies.

In chapter 2, we come back on this decision and consider the possibility of incorporating spin into the subspace reference calculations.

The subspace energy is calculated by projecting the Hamiltonian expressed in the non-orthogonal subspace onto the orthonormal molecular basis space. The previously introduced coefficients $\{w_{ia}\}$ carry out this projection, such that the subspace Hamiltonian $\tilde{H}^A$ in the subspace basis can be expressed as a Hamiltonian H^A in MO space as

$$H^A_{ijkl} = \sum_{abcd}^{K^A} w_{ia} w_{jb} w_{kc} w_{ld} \tilde{H}^A_{abcd} \tag{1.44}$$

This transformation only needs to be carried out at the start of the semidefinite program, to generate the constraint matrix H^A.

Numerical illustration of the subspace energy constraints

The subspace constraints are applied to the constituent atoms of the 14-electron diatomic molecules considered in section 1.5.2. What is their effect on the energy and other chemical properties?

The subspace energy constraints become active at long bond distances and raise the energy of the heteronuclear diatomics considerably.
They become active between 2 and 4 Å, as the v2DM(PQG) calculation simply does not violate the constraints at shorter bond lengths (figure 1.17). They greatly improve the energy of the heteronuclear diatomics in the dissociation limit (compare figure 1.18 with figure 1.12). Whereas the v2DM(PQG) dissociation energy of NO^+ is underestimated by 0.1 Hartree compared to MRCI, the difference decreases to 0.016 Hartree upon inclusion of subspace constraints (1.5).

The subsystem constraints force the molecule to dissociate into the correct dissociation products in the dissociation limit. This is reflected in the dipole moments and Mulliken populations (tables 1.6 and 1.7). At 20 Å, NO^+ has dissociated into a nitrogen atom and oxygen cation with near-integer occupations.

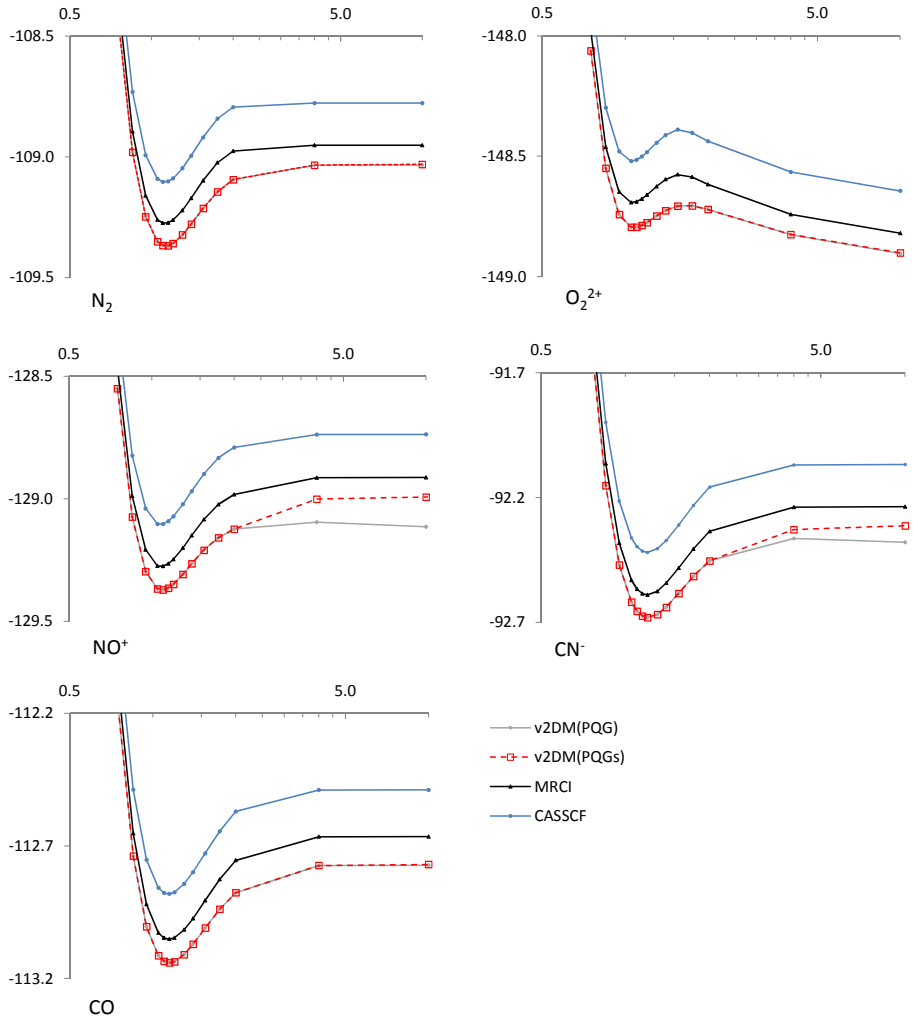

Figure 1.17: Imposing subspace constraints alongside 2-positivity conditions corrects the increasing non-parallelity error of the v2DM(PQG) PES relative to the MRCI PES towards dissociation. The subspace constraints only become active between 2 and 4 Å and affect the heteronuclear diatomics more than the homonuclear diatomics.

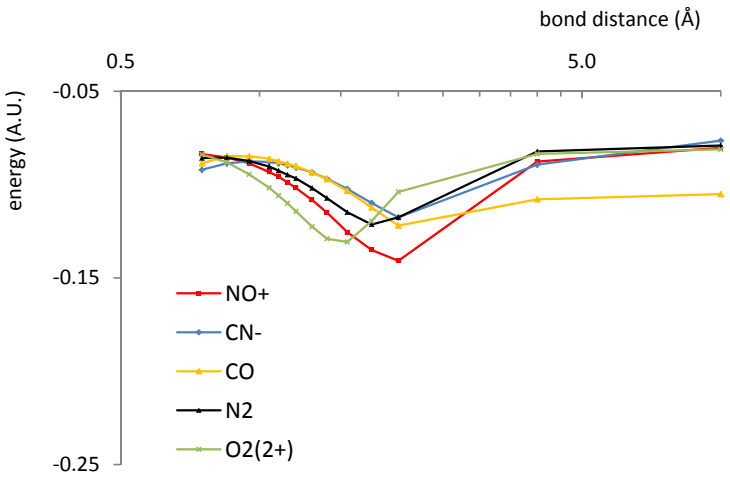

Figure 1.18: Subspace constraints reduce the non-parallelity errors of the v2DM(PQG) method relative to MRCI considerably (compare with figure 1.12). The biggest non-parallelity errors now occur between 1.5 and 2 Å, just before the molecules start to dissociate.

R (Å)	N_2	O_2^{2+}	NO^+	CN^-	CO
v2DM(PQG)	0.335	0.089	0.257	0.301	0.368
v2DM(PQGs)	0.337	0.089	0.378	0.367	0.371
MRCI	0.322	0.116	0.362	0.353	0.387

Table 1.5: Dissociation energies (without correction for zero-point energies) are much improved upon addition of subspace constraints in the v2DM(PQG) method. Without subspace constraints, it tends to underestimate dissociation energies. In case of O_2^{2+}, the barrier height for dissociation is given instead.

	R (Å)	v2DM(PQG)	v2DM(PQGs)	MRCI
NO^+	4.0	0.07	5.68	8.76
	10.0	-0.11	20.97	22.39
	20.0	-0.29	44.18	
CN^-	4.0	3.35	7.63	9.94
	10.0	7.01	25.77	25.83
	20.0	13.38	52.11	
CO	4.0	-0.40	0.01	0.04
	10.0	-0.90	0.00	0.00
	20.0	-1.71	0.00	

Table 1.6: When subspace constraints are imposed in the v2DM(PQG) method (denoted 'v2DM(PQGs)'), dipole moments (in Debije) in the dissociation limit correspond correctly to integer-charged dissociation products. Reference MRCI(FC) calculations were not available for bond lengths beyond 10 Å.

Nonetheless, the convergence to the dissociation limit energy and populations is clearly slower than for the MRCI method, for which the NO^+ molecule has already dissociated into a nitrogen atom and oxygen cation at 4 Å.

The subspace constraints are not only active on heteronuclear molecules. Although the homonuclear molecules N_2 and O_2 dissociate into the correct atomic products due to their symmetry, the energy in the dissociation limit is slightly lower than the sum of the atomic energies calculated separately. The subspace constraints close this gap and raise the energy of N_2 and O_2^{2+} at 20 Å from -148.9305 to -148.9267 Hartree and from -109.0332 to -109.0303 Hartree. In fact, these energies could be constrained even more strongly by including a spin condition in the reference subspace energy, which is explained in section 2.5.3 of chapter 2.

	R (Å)	v2DM(PQG)	v2DM(PQGs)	MRCI
NO^+	4.0	6.54	6.85	7.00
	10.0	6.53	6.97	7.00
	20.0	6.53	6.99	
CN^-	4.0	6.64	6.88	6.99
	10.0	6.61	7.00	7.00
	20.0	6.60	7.00	
CO	4.0	5.98	6.00	6.00
	10.0	5.98	6.00	6.00
	20.0	5.98	6.00	

Table 1.7: When subspace constraints are imposed in the v2DM(PQG) method (denoted 'v2DM(PQGs)'), Mulliken populations in the dissociation limit correspond correctly to integer-charged dissociation products. Reference MRCI(FC) calculations were not available for bond lengths beyond 10 Å.

1.5.4 Application of subspace energy constraints to polyatomic molecules

The subspace energy constraints introduced in previous paragraph, ensure correct dissociation in v2DM based methods. Nonetheless, some questions concerning these constraints remain. The number of possible subspaces that can be composed of all basis functions centered on one or more atoms in an M-atomic system, namely $2^M - 2$, scales exponentially with the size of the molecule. In practice, however, some subspace constraints may not be active. Which of them are active depends on the geometry and nature of the system. How fast does the number of practically relevant, i.e. active, subspace constraints grow with the number of atoms in the molecule? And can these active constraints be predicted beforehand? We clarify these issues by means of a relevant chemical system, the PES of the F_3^- ion.[69]

Figure 1.19: Numbering of the atoms and bond lengths of F_3^- used in section 1.5.4.

Computational details

The F_3^- calculations are done in the D95V basis set, as implemented in Gaussian03,[70] and constrained to linear geometries. Numbering of the atoms is done as follows The reference MRCI calculations were performed with Molpro.[38] The configurations included in the MRCI expansion were determined by a preceding full-valence CASSCF, with an active space of 22 electrons and all 12 valence orbitals, except that the molecular orbitals were taken from an analogous CASSCF calculation for the neutral species with doubly occupied inactive orbitals (mostly 1s core).

Application of subspace constraints to the PES of F_3^-

The shape of the potential energy surface (PES) calculated with the variational v2DM(PQG) method is severely incorrect, especially for molecular geometries with one or more stretched bonds. It is compared to that of an accurate MRCI PES in figures 1.20 and 1.21. Both graphs are composed of non-equidistant data points and show an equally large interval on the energy axis, which is truncated to enhance visibility of the bonding region. There are two striking differences between the two PES. First of all, the 2DM method yields a shallower well corresponding to the formation of the F_3^- anion, with a minimum at a somewhat larger bond length compared to MRCI (1.9 Å for 2DM theory versus 1.8 Å for MRCI). Secondly, in the outer regions of the v2DM(PQG) potential energy surface, describing the dissociation of the F_3^- ion, the energy does not increase but rather decreases. The decrease in energy upon dissociation is so strong that the optimal F_3^- geometry is only a local minimum in the 2DM potential energy surface. The cause of this problem can clearly be traced back to the strictly

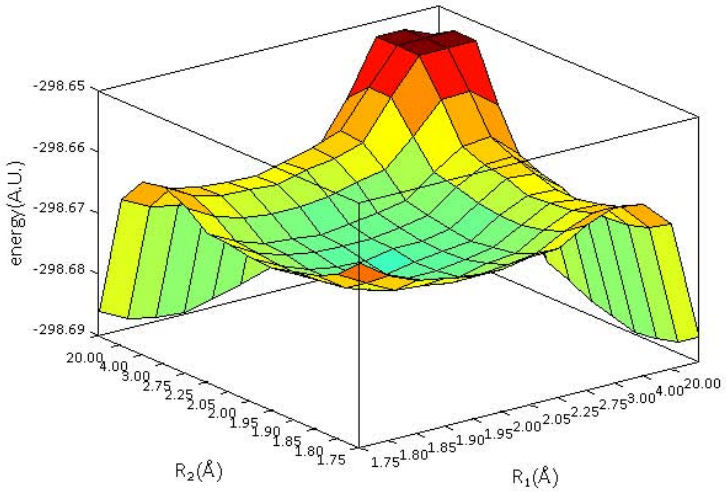

Figure 1.20: The outer regions of the potential energy surface of F_3^- obtained with the v2DM(PQG) method, corresponding to geometries with dissociated bonds, show an erroneous decrease in energy.

convex dependence of the v2DM(PQG) energy on the number of electrons. As a consequence, the dissociating system may reach a lower energy by allowing a fractional number of electrons on both atoms. Unless the decrease in energy caused by allowing a fractional charge on one atom is countered by a bigger increase in energy for the corresponding fractional charge on the other atom, the molecule will incorrectly dissociate into fractionally charged products with too low an energy.

The subspace constraints only affect molecular structures with large bond lengths. They aim to solve the aforementioned dissociation problem by constraining the energy of mono- or diatomic subspaces in the molecule to lie above the lowest ensemble energy for the isolated subspace with the same fractional number of electrons as the subspace in the molecule. The lowest-energy dissociated state will then be automatically obtained at integer occupations on the atoms. These constraints are already satisfied by v2DM(PQG) calculations for geometries with both bond lengths shorter than 2.75 Å (figure 1.22). The bond length of 2.75

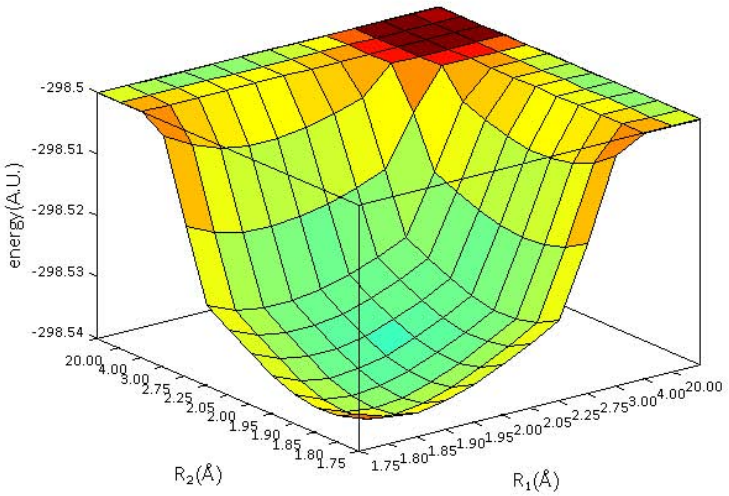

Figure 1.21: A reference potential energy surface of F_3^-, obtained with MRCI, shows the correct shape of the potential energy surface.

Å marks the onset of the 'long-distance behavior'. The constraints that are violated in geometries with dissociated bonds largely obey the following trends. When only one bond (R_1) is dissociated, and the other bond (R_2) is relatively short, the v2DM(PQG) calculation only violates the C_1 and C_{23} constraints, which act on the spatially separated atomic and diatomic unit in the system. When both bonds are dissociated, however, all constraints are violated by the v2DM(PQG) calculations.

The subspace constraints ensure correct dissociation of the F_3^- ion into F_2^- and F. Without subspace constraints, the F_3^- ion dissociates into $F_2^{-0.56} + F^{-0.44}$, which is the minimum energy structure among all structures with one bond dissociated, shown in the lower graph of figure 1.23. In fact, all these structures with one short bond and one dissociated bond (20 Å) should have an electronic structure recognizable as either F_2+F^- or F_2^-+F. However, without subspace constraints, the electronic charge delocalizes over the dissociation species (table 1.8).

When all subspace constraints are imposed, the electronic charge becomes

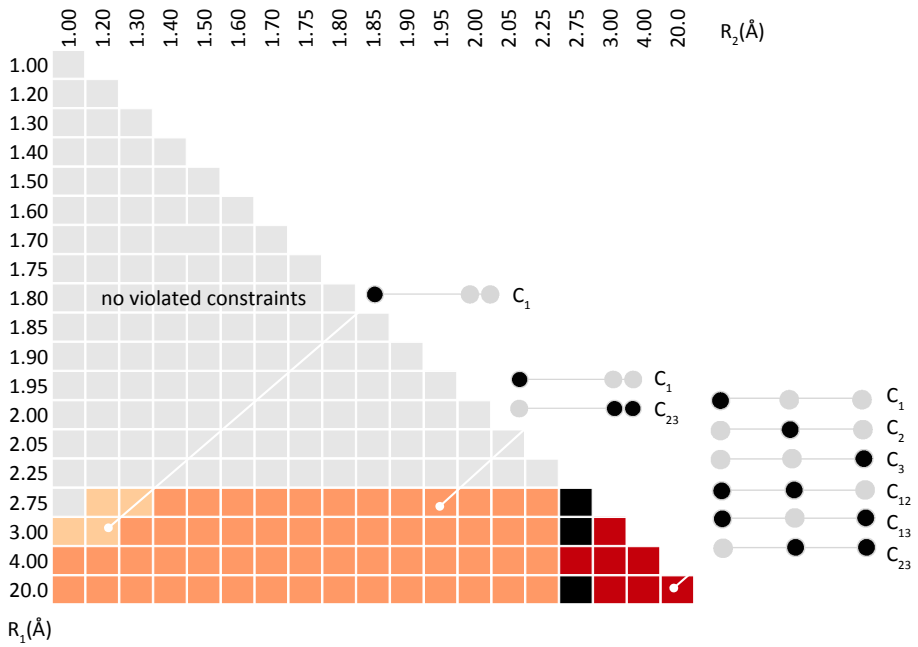

Figure 1.22: Different combinations of subspace constraints are violated by the v2DM(PQG) calculations in different parts of the PES. While no subspace constraints are violated for geometries with only short bond lengths (indicated by light gray squares in the PES), all of them are violated by calculations on fully stretched geometries (indicated by red squares in the PES). A schematic representation indicates with black dots on which atom the basis functions that span the subspace of the violated constraints are centered. Black squares in the overview of the PES indicate geometries at which yet another combination of subspace constraints was violated.

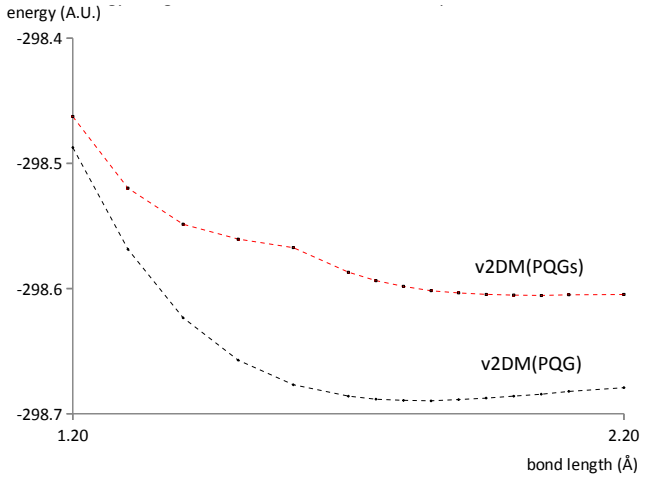

Figure 1.23: The v2DM(PQGs) method applied to the cut of the PES of F_3^- with one bond length fixed at 20 Å reveals the crossing between the two competitive dissociations $F_2 + F^-$ and $F_2^- + F$.

properly localized on the dissociation products (table 1.8). The energies of the structures with one short bond and one dissociated bond are then given by the uppermost graph in figure 1.23, which has a kink between 1.5 and 1.6 Å. At this point, the energy of the dissociation into $F_2^- + F$ becomes lower than that of the alternative dissociation $F_2 + F^-$. The minimum energy dissociation under the subspace constraints is $F_2^- + F$ with a bond length of 2.05 Å and charges -0.50 on both F atoms in the F_2^- molecule. These results agree with MRCI calculations, for which $F_2^- + F$ is also the lowest energy dissociation, with a bond length of 1.95 Å in the F_2^- molecule.

The set of subspace constraints acting on the spatially separate units in the system is the smallest set of subspace constraints that produces the correct dissociation in geometries with either short or dissociated bonds. Not all subspace constraints that are violated in the v2DM(PQG) calculation need be imposed in the v2DM(PQGs) calculation in order to obtain the correct dissociation. Some of them may overrule others, rendering them inactive in the resulting 2DM. For example, in the fully dissociated molecule, with three nuclei with

$R_1 = 20.0$ Å	R_2(Å)	PQG	PQG/C_1, C_2, C_3	PQG/C_1, C_{23}
	1.00	9.98	10.00	10.00
	1.20	9.84	9.98	10.00
	1.30	9.72	9.87	10.00
	1.40	9.62	9.70	10.00
	1.50	9.55	9.58	9.99
	1.60	9.49	9.50	9.13
	1.70	9.46	9.44	9.01
	1.75	9.45	9.42	9.00
	1.80	9.44	9.41	9.00
	1.85	9.44	9.39	9.00
	1.90	9.43	9.38	9.00
	1.95	9.43	9.37	9.00
	2.00	9.43	9.37	9.00
	2.05	9.43	9.36	9.00
	2.10	9.42	9.35	9.00
	2.20	9.42	9.34	9.00

Table 1.8: Subspace constraints on the atomic subspaces C_1, C_2, C_3 are not sufficient to ensure correct atomic (Mulliken, shown are those on F_1) populations on systems with one dissociated bond $R_2 = 20$ Å and one short bond R_1 ranging from 1.00 to 2.20 Å. The set of constraints C_{12}, C_3 on the spatially separated units in the system – one diatomic and one atomic – is the smallest set of constraints that ensures a correct dissociation with integer charges on the dissociated species.

$R_1 = 20.0$ Å	R_2(Å)	PQG	PQG/C_1, C_2, C_3	PQG/C_1, C_{23}	Δ
	1.00	-298.1694	-298.1638	-298.1629	-0.0002
	1.20	-298.4873	-298.4640	-298.4626	-0.0005
	1.30	-298.5687	-298.5278	-298.5198	0.0000
	1.40	-298.6234	-298.5726	-298.5486	0.0004
	1.50	-298.6572	-298.6036	-298.5605	-0.0002
	1.60	-298.6769	-298.6226	-298.5673	-0.0017
	1.70	-298.6860	-298.6317	-298.5870	-0.0001
	1.75	-298.6884	-298.6341	-298.5937	-0.0002
	1.80	-298.6893	-298.6351	-298.5983	-0.0002
	1.85	-298.6897	-298.6354	-298.6017	-0.0003
	1.90	-298.6887	-298.6344	-298.6035	-0.0002
	1.95	-298.6875	-298.6332	-298.6047	0.0002
	2.00	-298.6860	-298.6317	-298.6052	0.0001
	2.05	-298.6844	-298.6302	-298.6055	-0.0003
	2.10	-298.6822	-298.6281	-298.6050	-0.0001
	2.20	-298.6792	-298.6252	-298.6047	-0.0006

Table 1.9: Subspace constraints on the atomic subspaces C_1, C_2, C_3 alone are not sufficient to ensure the energy (in atomic units) of systems with one dissociated bond ($R_2 = 20$ Å) and one short bond (R_1) equals the sum of the energies of the dissociated species. The set of constraints C_{12}, C_3 on the two spatially separated units in the system, one diatomic and one atomic, is the smallest set of constraints that ensures the energy reproduces the sum of the energies of those units calculated separately – the energy difference between these two is given in the last column denoted 'Δ'.

large separations, all six subspace constraints are violated by v2DM(PQG). Nonetheless, the 'diatomic' subspace constraints are unlikely to have a meaningful contribution over the atomic subspace constraints, since their own energy violates the atomic subspace constraints. Indeed, they are made redundant by the atomic subspace constraints (the bright red area in figure 1.24). For all systems with a single dissociated bond, consisting of a diatomic unit and an atomic unit at very large internuclear distance, there are only two active constraints: a constraint on the diatomic unit and a constraint on the atomic unit. Therefore, in systems composed of (poly-) atomic units that are widely separated, the necessary constraints for correct dissociation act on the subspaces associated with the units, i.e. the subspace spanned by all basis functions centered on the atoms in the unit.

Unfortunately, this does not hold for all geometries. In systems with bonds that are stretched but not clearly dissociated, around 2.75 Å, additional diatomic constraints may be active (the dark red area in figure 1.24). As a consequence the number of active constraints does not always increase linearly with the size of the molecule.

The subspace constraints correct the shape of the dissociative regions of the v2DM(PQG) PES, but do not alter results for bound systems with short bonds. They turn the previously observed potential energy wells at long bond lengths into proper potential walls, such that a single well remains, corresponding to the bound F_3^- (with $R_1 = R_2 = 1.9$ Å the global minimum, see figure 1.25). Moreover, they not only correct the energy for geometries with one or more large bond lengths, but correct other chemical properties, such as dipole moments, as well. Nonetheless, the v2DM(PQGs) method still overestimates the bond strength compared to wavefunction-based ab initio methods (table 1.10). The subspace constraints do not alter the equilibrium F_3^- calculation and merely ensure the calculation on the dissociated system is energetically equivalent to separate calculations on the dissociated units. In order to obtain more accurate chemical properties, constraints are needed that improve results for short bond

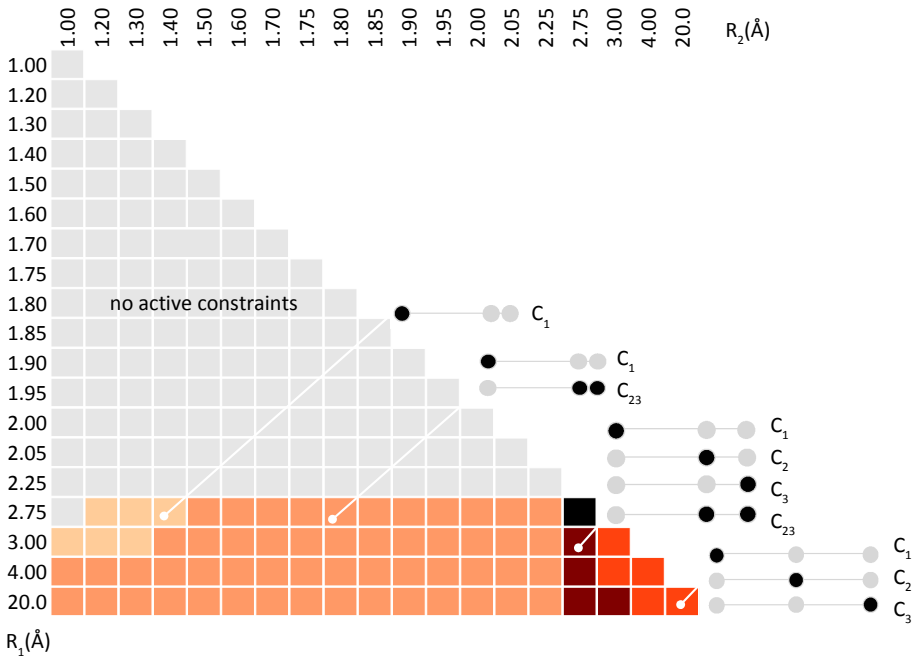

Figure 1.24: Although all subspace constraints are violated by the v2DM(PQG) method for geometries with both bonds dissociated, only the atomic subspace constraints are active when imposed in the calculation (indicated by bright red squares in the PES). A schematic representation indicates with black dots on which atom the basis functions that span the subspace of the active constraints are centered. For geometries with either clearly dissociated or short bonds, the active constraints target subspaces that coincide with the spatially separate units of the system. However, in structures with stretched – but not yet dissociated – bonds more subspace constraints are active (indicated by dark red squares in the PES). Black squares indicate geometries at which yet another combination of subspace constraints was active.

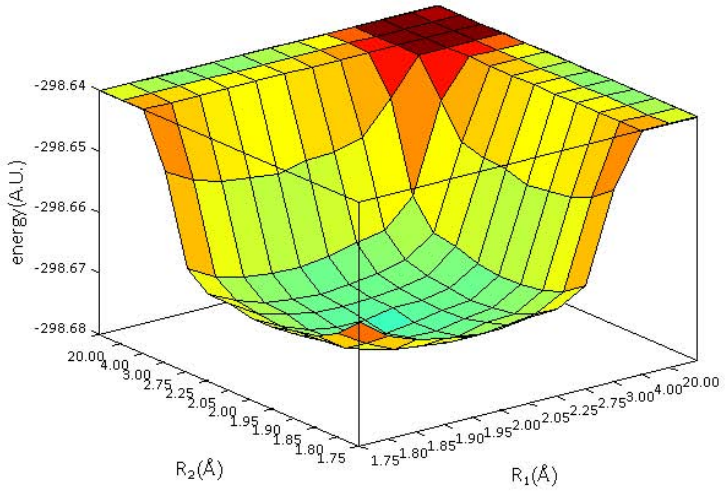

Figure 1.25: Imposing all mono- and diatomic subspace constraints corrects the dissociative regions of the potential energy surface of F_3^- compared to the v2DM(PQG) potential energy surface of figure 1.20

lengths as well. A combination of higher-index constraints to improve the accuracy for near-equilibrium geometries with subspace constraints to improve accuracy for dissociated geometries should improve results for both short and long bond distances, but would be much more costly.

In conclusion, subspace constraints offer a computationally cheap way to obtain correct molecular dissociation of small molecules in variational 2DM theory. Nonetheless, practical difficulties may arise when they are applied to larger systems. First of all, the number of subspace constraints needed to ensure correct dissociation does not always grow linearly with the number of atoms. In geometries with either clearly dissociated or short bonds, the correct dissociation can be obtained using only constraints on the spatially separated units in the system, which may include up to M atoms in an M-atomic molecule, or up to 2 atoms in the F_3^- molecule examined here. However, this does not hold for all geometries. Additional constraints may be active in geometries with stretched, but not fully dissociated, bonds. Secondly, constructing the constraints requires

	MRCI(FC)		CCSD(FC)		v2DM(PQGs)	
	E(A.U.)	R(Å)	E(A.U.)	R(Å)	E(A.U.)	R(Å)
F_3^-	-298.5385	1.80	-298.5792	1.75	-298.6724	1.90
F	-99.4690		-99.4703		-99.4928	
F^-	-99.5307		-99.5350		-99.5441	
F_2	-198.9637	1.60	-198.9652	1.52	-199.0189	1.60
F_2^-	-199.0404	1.95	-199.0557	1.94	-199.1125	2.00
$F_2^- + F$	-298.5094		-298.5260		-298.6053	
$D_e(F_2^-, F)$			0.0532		0.0671	
$D_e(F, F, F^-)$	0.1353		0.1036		0.1426	

Table 1.10: The accuracy of the v2DM(PQGs) method remains poor, as the subspace constraints only correct for improper dissociation. The v2DM(PQGs) dissociation energies D_e for dissociation into $F_2^- + F$ and $F + F + F^-$ are substantially bigger than those obtained with CCSD and MRCI. MRCI calculations on the dissociated system $F_2^- + F$ did not converge properly, hence no value is specified for the dissociation energy. Equilibrium bond lengths R are, in the case of FCI(FC), MRCI(FC) and v2DM(PQGs), determined in steps of 0.05 Å.

separate calculations for each geometry of the multi-atomic subspaces. Multi-atomic subspaces may be needed because the correct dissociation cannot always be realized through constraints on the atomic subspaces only. Assembling the constraint data thus becomes a time-consuming process when applied to large PES. Finally, the subspace constraints merely correct the faulty long-range behavior under the P-,Q- and G-condition, they do not affect the accuracy of the v2DM(PQG) method for geometries with short bonds.

1.5.5 Size-consistency and separability under 2-index constraints

Size-consistency is usually defined in terms of additive separability of the energy. An ab initio method is size-consistent if its energy for a system of non-interacting units equals the sum of its energies for these units considered separately. In previous sections, we have already established the lack of size-consistency in the v2DM(PQG) method and argued that subspace constraints correct energetic size-consistency defects. But what about the relationship between the 2DM for a system of non-interacting units and the 2DM's of each of the units calculated separately: are they consistent? An exact method imposes a relationship between them, which is a much stronger requirement than only consistency of the energy. If the 2DM is consistent with the 2DM's for the fragments calculated separately, not only the energy but also other chemical properties will be size-consistent as well.

Hence we examine to what extent the energy and 2DM are separable for a system of non-interacting units under 2-positivity conditions, both with and without the addition of subspace energy constraints. The first subsection examines the relationships between the concepts of size-consistency, separability and entanglement. The second subsection examines these concepts applied to some simple molecular systems at their dissociation limits.

Theoretical background on size-consistency and separability

A system of non-interacting units can always be described by a separable 2DM, which is the antisymmetrized product of the 1DM's and 2DM's of the non-interacting units. The N-electron Hamiltonian for a system of non-interacting units A and B is the sum of the Hamiltonians for the fragments. To describe each fragment, we consider a K^A-dimensional set of orthonormalized basis functions $\{\psi_i^A\}$ for A and a K^B-dimensional orthonormal basis set $\{\psi_i^B\}$ for B. The orthonormal bases used to describe two different fragments are strongly orthogonal. The number of atoms in fragment A is denoted N^A, the number

of atoms in fragment B is denoted N^B and superscripts are used to specify to which of the two basis sets the indices refer to.

$$\begin{aligned}
\hat{H} &= \hat{H}^A + \hat{H}^B \qquad (1.45)\\
\hat{H}^A &= \sum_{ij}^{K^A} h_{ij}^{AA} a_j^\dagger a_i + \sum_{ijkl}^{K^A} V_{ijkl}^{AAAA} a_k^\dagger a_l^\dagger a_j a_i\\
\hat{H}^B &= \sum_{ij}^{K^B} h_{ij}^{BB} a_j^\dagger a_i + \sum_{ijkl}^{K^B} V_{ijkl}^{BBBB} a_k^\dagger a_l^\dagger a_j a_i\\
\langle \psi_i^A | \psi_j^A \rangle &= \delta_{ij} \quad \text{orthonormalized basis}\\
\langle \psi_i^B | \psi_j^B \rangle &= \delta_{ij} \quad \text{orthonormalized basis}\\
\langle \psi_i^A | \psi_j^B \rangle &= 0 \quad \text{strong orthogonality}
\end{aligned}$$

Given the additive separability of the Hamiltonian (1.45), such a non-interacting system can be described by a multiplicatively separable wavefunction that is the antisymmetrized product of wavefunctions for each of the non-interacting units

$$|\Psi(x_1, \ldots, x_{N^A+N^B})\rangle = |\Psi^A(x_1, \ldots x_{N^A})\rangle \wedge |\Psi^B(x_{N^A+1}, \ldots x_{N^A+N^B})\rangle \quad (1.46)$$

This type of wavefunction covers the whole variational space of the additively separable Hamiltonian (1.45) and is therefore a physically plausible representation of the non-interacting system.

The 1DM and 2DM that correspond to this separable wavefunction follow from integration over all but two variables and have the block structure

$$\gamma = \begin{pmatrix} \gamma^{AA} & 0 \\ 0 & \gamma^{BB} \end{pmatrix} \quad \Gamma = \begin{pmatrix} \Gamma^{AAAA} & 0 & 0 \\ 0 & \Gamma^{BBBB} & 0 \\ 0 & 0 & \Gamma^{ABAB} \end{pmatrix} \qquad (1.47)$$

The blocks of the 1DM correspond to the 1DM's for A and B

$$\begin{aligned}
\gamma_{ik}^{AA} &= \langle \Psi^A | a_k^\dagger a_i | \Psi^A \rangle\\
\gamma_{ik}^{BB} &= \langle \Psi^B | a_k^\dagger a_i | \Psi^B \rangle
\end{aligned}$$

The blocks of the 2DM correspond to the antisymmetrized product of the reduced matrix elements for each of the non-interacting fragments

$$\begin{aligned}
\Gamma^{AAAA}_{ijkl} &= \langle \Psi^A | a^\dagger_k a^\dagger_l a_j a_i | \Psi^A \rangle \\
\Gamma^{BBBB}_{ijkl} &= \langle \Psi^B | a^\dagger_k a^\dagger_l a_j a_i | \Psi^B \rangle \\
\Gamma^{ABAB}_{ijkl} &= \langle \Psi^A | a^\dagger_k a_i | \Psi^A \rangle \langle \Psi^B | a^\dagger_l a_j | \Psi^B \rangle
\end{aligned} \tag{1.48}$$

This implies that

$$\begin{aligned}
(N^A - 1) \sum_j \Gamma^{ABAB}_{ijkj} &= N^B \sum_j \Gamma^{AAAA}_{ijkj} \\
(N^B - 1) \sum_j \Gamma^{BABA}_{ijkj} &= N^A \sum_j \Gamma^{BBBB}_{ijkj}
\end{aligned} \tag{1.49}$$

and since the 1DM is formed as $\sum_j \Gamma^{ABAB}_{ijkj} + \Gamma^{AAAA}_{ijkj} = (N-1)\gamma^{AA}_{ik}$

$$\Gamma^{ABAB}_{ijkl} = \gamma^{AA}_{ik} \gamma^{BB}_{jl} \tag{1.50}$$

This condition is equivalent to the condition that the separable state forms an eigenfunction of the operator $\hat{N}^A$ and $\hat{N}^B$

$$\begin{aligned}
(N\hat{N}^A - N^A \hat{N}) |\Psi^A\rangle \wedge |\Psi^B\rangle = 0 \\
(N\hat{N}^B - N^B \hat{N}) |\Psi^A\rangle \wedge |\Psi^B\rangle = 0
\end{aligned}$$

which requires that the vector corresponding to this operator lies in the nullspace of the G-matrix.

The separable 2DM leads to an additively separable *cumulant*. The 2DM can be separated into first order contributions and contributions that are not expressible in terms of its first order contraction. The part of the 2DM that is not expressible as an antisymmetrized product of its first order contractions is referred to as the cumulant Δ,[17, 65, 71]

$$\begin{aligned}
\Gamma &= \gamma \wedge \gamma + \Delta \\
\Gamma_{ijkl} &= \gamma_{ik}\gamma_{jl} - \gamma_{il}\gamma_{jk} + \Delta_{ijkl}
\end{aligned}$$

Because of the separable structure of the 2DM, $\Gamma^{ABAB}_{ijkl} = \gamma^{AA}_{ik}\gamma^{BB}_{jl}$, the ABAB block of the cumulant is identically zero.

$$\begin{aligned}\Delta^{ABAB}_{ijkl} &= \Gamma^{ABAB}_{ijkl} - \gamma^{AA}_{ik}\Gamma^{BB}_{jl} - \gamma^{AB}_{il}\gamma^{AB}_{kj} \\ &= 0\end{aligned}$$

The structure of the cumulant for the separable system can therefore be written as $\Delta = \Delta^A \oplus \Delta^B$; it is *additively separable*.

$$\Delta = \begin{pmatrix} \Delta^{AAAA} & 0 & 0 \\ 0 & \Delta^{BBBB} & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{1.51}$$

The separable structure of the 2DM guarantees size-consistency. The Γ^{ABAB} block does not contribute to the electron-electron repulsion, but contributes to the energy of the non-interacting system via the one-electron energy, because the one- and two-electron parts of the Hamiltonian have the block structure

$$h = \begin{pmatrix} h^{AA} & 0 \\ 0 & h^{BB} \end{pmatrix} \quad V = \begin{pmatrix} V^{AAAA} & 0 & 0 \\ 0 & V^{BBBB} & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{1.52}$$

so the energy is

$$\begin{aligned} E &= tr\ [V^{AAAA}\Gamma^{AAAA}] + tr\ [V^{BBBB}\Gamma^{BBBB}] + tr\ [h^{AA}\gamma^{AA}] + tr\ [h^{BB}\gamma^{BB}] \\ &= tr\ [V^{AAAA}\Gamma^{AAAA}] + tr\ [V^{BBBB}\Gamma^{BBBB}] \\ &+ \frac{1}{N-1}\sum_{ijk} h^{AA}_{ik}\Gamma^{AAAA}_{ijkj} + \frac{1}{N-1}\sum_{ijk} h^{AA}_{ik}\Gamma^{ABAB}_{ijkj} \\ &+ \frac{1}{N-1}\sum_{ijk} h^{BB}_{ik}\Gamma^{BBBB}_{ijkj} + \frac{1}{N-1}\sum_{ijk} h^{BB}_{ik}\Gamma^{BABA}_{ijkj} \end{aligned} \tag{1.53}$$

where I^{AA}, I^{BB} are identity matrices for the orbitals spanning A and B. Because of the separability of the ABAB block (1.50), the energy expression for the separable wavefunction becomes equivalent to the expression for the energy of

the separate fragments with electron numbers N^A and N^B:

$$E = tr\ [V^{AAAA}\Gamma^{AAAA}] + tr\ [V^{BBBB}\Gamma^{BBBB}] + \frac{1}{N^A - 1}\sum_{ijk} h^{AA}_{ik}\Gamma^{AAAA}_{ijkj} + \frac{1}{N^B - 1}\sum_{ijk} h^{BB}_{ik}\Gamma^{BBBB}_{ijkj} \tag{1.54}$$

Since this kind of separable wavefunction is a valid representation of a noninteracting system composed of units A and B, wavefunction based *ab initio* methods typically generate wavefunctions with this structure for non-interacting systems.

Non-interacting systems with degeneracies can be, but are not necessarily, represented by a separable state. Such degeneracies may include

charge degeneracy: for instance $\mathrm{N} + \mathrm{N}^+ \leftrightarrow \mathrm{N}^+ + \mathrm{N}$

degenerate states in either A or B: for instance, due to spin degeneracy, the singlet dissociated hydrogen dimer may described as $\mathrm{H}^\uparrow + \mathrm{H}^\downarrow \leftrightarrow \mathrm{H}^\downarrow + \mathrm{H}^\uparrow$

When such degeneracies are present, the system need not be described by a pure state, but can be a mixed and/or entangled state. These states do not necessarily lead to a 2DM with the same block diagonal structure (1.48) as a separable pure state. The ground state wavefunction for such a system may be entangled,

$$|\Psi(x_1, \ldots, x_{N^A+N^B})\rangle = \sum_{ab} c^{AB}_{ab} |\Psi^A_a(x_1, \ldots x_{N^A})\rangle \wedge |\Psi^B_b(x_{N^A+1}, \ldots x_{N^A+N^B})\rangle \tag{1.55}$$

where the indices a, b may run over all orthonormal degenerate ground states and the coefficients c^{AB}_{ab} satisfy $\sum_{ab}(c^{AB}_{ab})^2 = 1$ but cannot necessarily be factorized as $c^A_a c^B_b$, as is the case for any separable state. Because these states allow several degenerate representations, the density matrix may be mixed as well as entangled

$$\Gamma^{(N)} = \sum_n w_n \sum_{ab}\sum_{cd} c^{AB,n}_{ab} c^{AB,n}_{cd} |\Psi^A_c\rangle \wedge |\Psi^B_d\rangle\langle\Psi^A_a| \wedge \langle\Psi^B_b|$$

with $\sum_{a,b}(c^{AB,n}_{ab})^2 = 1\ \forall n$. An entangled density matrix for a system of non-interacting fragments clearly is not necessarily separable into density matrices

for these systems calculated separately (1.48). In fact, it may lead to a second order cumulant that is not additively separable

$$\begin{aligned}
\Gamma^{ABAB}_{ijkl} &= \sum_n w_n \sum_{ab}\sum_{cd} c^{AB,n}_{ab} c^{AB,n}_{cd} \langle \Psi^A_a | a^\dagger_k a_i | \Psi^A_c \rangle \langle \Psi^B_b | a^\dagger_l a_j | \Psi^B_d \rangle \\
\gamma^{AA}_{ik} &= \sum_n w_n \sum_{ac}\sum_{b} c^{AB,n}_{ab} c^{AB,n}_{cb} \langle \Psi^A_a | a^\dagger_k a_i | \Psi^A_c \rangle \\
\gamma^{BB}_{jl} &= \sum_n w_n \sum_{bd}\sum_{a} c^{AB,n}_{ab} c^{AB,n}_{ad} \langle \Psi^B_b | a^\dagger_l a_j | \Psi^B_d \rangle \\
\Delta^{ABAB}_{ijkl} &= \Gamma^{ABAB}_{ijkl} - \gamma^{AA}_{ik}\gamma^{BB}_{jl} \\
&\neq 0 \quad \text{in general}
\end{aligned}$$

Even when the state is separable, but mixed, the cumulant need not be additively separable,

$$\begin{aligned}
\Gamma^{ABAB}_{ijkl} &= \sum_n w_n \sum_{ab}\sum_{cd} c^{A,n}_{a} c^{B,n}_{b} c^{A,n}_{c} c^{B,n}_{d} \langle \Psi^A_a | a^\dagger_k a_i | \Psi^A_c \rangle \langle \Psi^B_b | a^\dagger_l a_j | \Psi^B_d \rangle \\
\gamma^{AA}_{ik} &= \sum_n w_n \sum_{a}\sum_{c} c^{A,n}_{a} c^{A,n}_{c} \langle \Psi^A_a | a^\dagger_k a_i | \Psi^A_c \rangle \\
\gamma^{BB}_{jl} &= \sum_n w_n \sum_{b}\sum_{d} c^{B,n}_{b} c^{B,n}_{d} \langle \Psi^B_b | a^\dagger_l a_j | \Psi^B_d \rangle \\
\Delta^{ABAB}_{ijkl} &= \Gamma^{ABAB}_{ijkl} - \gamma^{AA}_{ik}\gamma^{BB}_{jl} \\
&\neq 0 \quad \text{in general, unless } w_n = 1 \quad \text{for some } n
\end{aligned}$$

However, this implies that if either A or B is non-degenerate, the cumulant will still be separable. In order to establish whether the v2DM(PQG) method leads to a structurally correct 2DM for non-interacting states, we will focus on systems that only admit a separable state in the dissociation limit. More specifically, we will consider diatomic molecules that dissociate into non-degenerate closed shell singlet states.

Discussion on size-consistency and separability in v2DM theory

The 2-positivity conditions do not produce the correct structure of the 2DM for a non-interacting system. As explained in previous section, the 2DM for a

dissociated system without any degeneracies must be described by a 2DM that is separable. The 2DM must have a structure that is consistent with the 2DM's calculated separately for the dissociation products (1.47). But even when the reference system is taken as the fragments with the same non-integer occupations that occur in the non-interacting composite system, the 2DM under 2-positivity conditions does not correspond to the 2DM's of those fractionally charged non-interacting fragments (table 1.11). Although the variationally optimized 2DM under 2-positivity conditions has the typical block structure of a system of non-interacting units, given by (1.47), its blocks are not related in the way they would be for a separable system (1.48). The cumulant is not additively separable, indicated by a non-zero block Δ^{ABAB}. V2DM(PQGs) calculations on several non-degenerate dissociated states yield cumulant blocks Δ^{ABAB} with Frobenius norms as big as 10^{-1}, whereas this block would be zero in a separable pure state (table 1.11).

The failure of the 2DM for a system of non-interacting fragments to be consistent with the 2DM's for each of the non-interacting fragments calculated separately also explains why the energy is not size-consistent in homonuclear diatomics, as observed in section 1.5.3.

The failure of the 2-positivity conditions to be size-consistent becomes apparent by considering a system composed of two non-interacting two-electron systems. Although the P-, Q- and G-condition on the whole system imply a P-,Q- and G-condition on each pair of subspace 1DM and 2DM, they are not sufficient for N-representability, because they do not guarantee that the subspace 1DM and 2DM can be derived from the same physical ensemble corresponding to the subspace population.

The 2DM can be projected onto an orthonormal basis for the subspace, for instance by considering a symmetrical orthonormalization of the non-orthogonal subspace basis $\{\phi_a, \phi_b, \ldots\}$ considered before (1.37), leading to the projection coefficients $w_{\alpha i}$ from the orthonormal molecular basis $\{\phi_i, \phi_j, \ldots\}$ onto the

PQG	$\| \Gamma - (1.47) \|_F$	$\| \Delta^{ABAB} \|_F$	E^{AB}	$E^{AB} - (E^A + E^B)$	N^A N^B
Be_2	1.2E-09	8.2E-02	-29.2315	-1.6E-04	4.00 4.00
Be_2^{4+}	3.9E-09	1.1E-05	-27.2217	1.6E-05	2.00 2.00
BeB^+	1.0E-14	1.6E-02	-38.9560	-4.9E-02	4.00 4.00
H_2^{2-}	1.7E-07	1.3E-02	-0.9397	5.0E-07	2.00 2.00
He_2	1.0E-08	9.7E-03	-5.7752	-2.2E-07	2.00 2.00
HeH^-	4.4E-16	1.1E-02	-3.3575	-7.8E-05	2.00 2.00
HeH^+	3.8E-16	3.9E-08	-2.8876	1.3E-06	2.00 0.00
$HeLi^+$	4.8E-15	3.5E-04	-10.1247	3.0E-08	2.00 2.00
Li_2^{2+}	1.2E-09	7.9E-06	-14.4721	1.4E-05	2.00 2.00
Li_2^{2-}	1.3E-10	2.8E-01	-15.6311	-7.4E-01	4.00 4.00

P	$\| \Gamma - (1.47) \|_F$	$\| \Delta^{ABAB} \|_F$	E^{AB}	$E^{AB} - (E^A + E^B)$	N^A N^B
Be_2	4.2E-12	9.6E-01	-63.6285	-3.4E+01	4.00 4.00
Be_2^{4+}	5.4E-12	1.6E-01	-31.8142	-4.6E+00	2.00 2.00
BeB^+	7.1E-15	1.0E+00	-81.6286	-4.3E+01	4.00 4.00
H_2^{2-}	5.5E-13	8.6E-01	-1.9971	-1.1E+00	2.00 2.00
He_2	2.0E-12	8.2E-01	-7.9745	-2.2E00	2.00 2.00
HeH^-	2.2E-15	8.4E-01	-4.9858	-1.6E+00	2.00 2.00
HeH^+	3.8E-16	1.2E-08	-2.8876	1.2E-06	2.00 0.00
$HeLi^+$	5.8E-15	7.0E-01	-12.9159	-2.8E+00	2.00 2.00
Li_2^{2+}	1.2E-09	7.9E-06	-14.4721	1.5E-05	2.00 2.00
Li_2^{2-}	1.1E-12	8.5E-01	-35.5894	-2.1E+01	4.00 4.00

Table 1.11: The 2-positivity conditions do not guarantee that the ABAB block of the cumulant Δ in non-interacting, non-entangled singlet states is identically zero. These tables show the influence of the different 2-positivity conditions (specified in the upper left corner) on the block-diagonal structure of the 2DM, measured by the Frobenius norm of the off-diagonal part $\Gamma - (\Gamma^{AAAA} \oplus \Gamma^{BBBB} \oplus \Gamma^{ABAB})$, cumulant separability measured by the Frobenius norm of its ABAB block $\| \Delta^{ABAB} \|_F$, the energy E^{AB} and the size-consistency error $E^{AB} - (E^A + E^B)$. For the molecules with first-row atoms, a cc-pVDZ basis was used, and a D95V basis for the other molecules. All bond lengths are greater than 10^4 Å.

PQ	$\| \Gamma - (1.47) \|_F$	$\| \Delta^{ABAB} \|_F$	E^{AB}	$E^{AB} - (E^A + E^B)$	N^A N^B
Be_2	1.0E-08	4.7E-01	-29.6119	-3.8E-01	4.00 4.00
Be_2^{4+}	1.0E-08	1.8E-05	-27.2223	-5.7E-04	2.00 2.00
BeB^+	4.0E-15	4.3E-01	-39.7492	-8.4E-01	4.00 4.00
H_2^{2-}	9.4E-10	7.5E-03	-0.9560	-1.6E-02	2.00 2.00
He_2	2.0E-07	7.2E-03	-5.8149	-4.0E-02	2.00 2.00
HeH^-	6.8E-16	7.7E-03	-3.3923	-3.5E-02	2.01 1.99
HeH^+	3.8E-16	1.3E-08	-2.8876	1.3E-06	2.00 0.00
$HeLi^+$	5.8E-15	7.0E-01	-12.9159	-2.8E+00	2.00 2.00
Li_2^{2+}	1.3E-10	2.8E-01	-15.6311	-7.4E-01	2.00 2.00
Li_2^{2-}	5.8E-09	1.5E-05	-14.4723	-1.3E-04	4.00 4.00

PG	$\| \Gamma - (1.47) \|_F$	$\| \Delta^{ABAB} \|_F$	E^{AB}	$E^{AB} - (E^A + E^B)$	N^A N^B
Be_2	2.1E-09	8.0E-02	-29.2318	-5.2E-04	4.00 4.00
Be_2^{4+}	2.0E-07	5.0E-06	-27.2217	1.2E-05	2.00 2.00
BeB^+	1.1E-14	2.8E-02	-38.9694	-6.2E-02	4.00 4.00
H_2^{2-}	2.9E-09	1.4E-02	-0.9397	7.9E-07	2.00 2.00
He_2	4.2E-06	9.4E-03	-5.7752	7.7E-08	2.00 2.00
HeH^-	4.5E-16	3.2E-03	-3.3583	-8.3E-04	2.00 2.00
HeH^+	3.8E-16	4.2E-08	-2.8876	1.4E-06	2.00 0.00
$HeLi^+$	4.8E-15	2.9E-04	-10.1249	-2.0E-04	2.00 2.00
Li_2^{2+}	3.6E-09	7.6E-02	-14.8939	-2.7E-04	2.00 2.00
Li_2^{2-}	1.7E-06	5.1E-06	-14.4721	9.9E-06	4.00 4.00

(continued from table 1.11) Different subsets of 2-positivity conditions are not enough to guarantee separability in a non-entangled non-interacting system. The 2DM does have a block diagonal structure, which is just a consequence of the Hamiltonian's structure. Under P-condition only, all electrons end up in the Γ^{ABAB} block of the 2DM, which produces an energy that corresponds completely to the one-electron Hamiltonian. The G-condition improves the structure of the 2DM and its energy significantly, much more than the Q-condition, but does not make it exact. Even for systems that dissociate into two 2-electron atoms, the 2-positivity conditions are not exact.

orthonormal subspace basis $\{\phi_\alpha, \phi_\beta, \phi_\gamma, \phi_\delta, \ldots\}$

$$w_{\alpha i} \equiv \sum_{b}^{K^A} \sum_{c}^{K} (S^{A^{-1/2}})_{\alpha b} S_{bc} C_{ic}$$

where S^A is the overlap matrix between the non-orthonormal basis functions of subspace A and S is the overlap matrix between the non-orthonormal basis functions of all subspaces. In the dissociation limit, the overlap matrix becomes block diagonal, so these coefficients reduce to

$$w_{\alpha i} \equiv \sum_{bc}^{K^A} (S^{A^{1/2}})_{\alpha c} C_{ic}$$

The creation and annihilation operators in this orthonormal subspace basis define a P-, Q- and G-condition on the subspace, with matrix elements expressed in terms of the subspace 1DM, 2DM pair

$$\begin{aligned}
&\tilde{\Gamma}^A \succeq 0 \quad \text{with} \quad \tilde{\Gamma}^A_{\alpha\beta\gamma\delta} = \quad \langle \Psi | a^\dagger_\gamma a^\dagger_\delta a_\beta a_\alpha | \Psi \rangle \\
&Q(\tilde{\Gamma}^A, \tilde{\gamma}^A) \succeq 0 \quad \text{with} \quad Q(\tilde{\Gamma}^A, \tilde{\gamma}^A)_{\alpha\beta\gamma\delta} = \quad \langle \Psi | a_\gamma a_\delta a^\dagger_\beta a^\dagger_\alpha | \Psi \rangle \\
&G(\tilde{\Gamma}^A, \tilde{\gamma}^A) \succeq 0 \quad \text{with} \quad G(\tilde{\Gamma}^A, \tilde{\gamma}^A)_{\alpha\beta\gamma\delta} = \quad \langle \Psi | a^\dagger_\gamma a_\delta a^\dagger_\beta a_\alpha | \Psi \rangle
\end{aligned}$$

These conditions are implied by the P-,Q- and G-condition on the whole system. Because of the relation between the subspace basis $\{\phi_\alpha\}$ and the molecular basis $\{\phi_i\}$,

$$a^\dagger_\alpha = \sum_i w_{\alpha i} a_i$$

But even though only the P-condition ensures N-representability of a two-electron system, the P-condition on two-electron subspaces is not enough to ensure N-representability. Expression (1.53) for the energy of such a system shows immediately why: putting all electrons into the Γ^{ABAB} block satisfies the P-condition and yields a lower energy than the P-condition applied to separate two-electron systems (1.54) because it only references the one-electron Hamiltonian and therefore no electron-electron repulsion term enters the energy expression (table 1.11). Similarly, the Γ^{ABAB} block leads to have more variational freedom under 2-positivity constraints on the whole system than on the subsystems

considered separately because they do not require that the subspace 1DM and 2DM are derivable from the same fractional N ensemble.

The subspace constraints impose size-consistency but do not impose the exact 2DM structure of a system of non-interacting units in the dissociation limit. As can be expected from their formulation, including the subspace energy constraints in the v2DM(PQG) method makes the energy size-consistent. Calculations on a set of dissociated 14-electron diatomic molecules with bond lengths larger than 10^4 Å in the Cartesian cc-pVDZ basis confirm this numerically. In order to rule out any inconsistencies between the molecular and atomic systems, spin constraints were imposed in neither the molecular nor the subspace calculations. However, even though the energy of the dissociated system is completely consistent with the energies of the dissociation products calculated separately, the structure of the 2DM need not be. Imposing subspace constraints alongside the 2-positivity conditions does not enforce separability of the 2DM. The subspace constraints fall short in correcting the lack of separability because they only act on the energy of the dissociated system. In terms of the notation used here, the subspace constraints in the dissociation limit impose that the energy expression (1.53) must be greater than or equal to the separable energy given by (1.54). This is not enough to guarantee that the 2DM is separable into 2DM's for the dissociation products.

These findings agree with those of Nakata et al.[55] based on a different approach. They studied the separability of the 2DM for systems of non-interacting molecular clusters in a minimal basis set. The 2DM is made block diagonal by construction to describe the non-interacting system. The resulting deviations from zero they observe in the Δ^{ABAB} block are of the same order of magnitude as those observed here.

In conclusion, calculations on a supersystem composed of non-degenerate non-interacting subsystems establish that, even though they allow for an energetically size-consistent description of long-range electronic interactions, the subspace constraints do not ensure separability. The structure of the 2DM for a system

composed of non-interacting fragments obtained under these constraints cannot be separated into 2DM's for non-interacting fragments. Even the 1DM under separability constraints for such a system need not be consistent with the 1DM obtained by separate calculations. This implies that one- and two-electron properties other than the energy are not necessarily size-consistent.

Pure state separability is not a necessary condition for N-representability in a system of non-interacting fragments, as degeneracies in the non-interacting fragments may lead to an ensemble state that may be entangled as well. However, any system of non-interacting units allows a description in terms of a pure separable state. Therefore it can be imposed, but it would only be meaningful if it follows as the dissociation limit behaviour of a more general constraint.

The whole is more than the sum of the parts.

Aristotle

2

S-representability

2.1 Introduction

Electronic spin lies at the heart of chemistry. The surprisingly simple quantum chemical description of spin has helped us to understand the most fundamental properties of matter.[72] However, when we do not wish to work with the full wavefunction, which is an impractical mathematical object, and work with more compact descriptors instead, describing spin is problematic. In Density Functional Approximations (DFA), one often resorts to symmetry breaking.[73] A recent approach by Yang et al. adjusts DFA functionals to correct the origin of the spin problem.[74,75]

Although v2DM theory is typically a ground state method, it can be applied to find the lowest-energy state for a specific spin state. Nonetheless, the problem of describing non-singlet spin states in v2DM theory has received little attention, although Valdemoro and co-workers have made a thorough study of spin purification procedures in the context of the contracted Schrödinger equation.[76,77] Mazziotti has pointed out the advantages of spin and spatial symmetry adap-

tation, providing a framework for singlet and non-singlet state calculations in a spin adapted basis in v2DM theory, but illustrates them only with singlet state calculations.[78] Very little about non-singlet state v2DM calculations has appeared in the literature.[79]

A consistent treatment of non-singlet spin states in v2DM theory is much needed, not only because many important molecules are non-singlet states in their ground state, but also because many singlet molecules dissociate into non-singlet states. Spin may therefore help us understand and solve the size-consistency and dissociation issues discussed in the previous chapter.

For this reason, we make a comparative assessment of several spin constraints in v2DM theory. Section 2.2 discusses the incorporation of electronic spin in the tp basis, followed by an examination of necessary constraints for S-representability in sections 2.3 and 2.4. The constraints in section 2.3 aim to describe a pure spin state, whereas this requirement is lifted in section 2.4 allowing the 2DM to describe an ensemble of spin states. Section 2.5 analyses both approaches by applying them to the PES of non-singlet molecules.

2.2 Representation of electronic spin in the 2DM

Representation of two particle/hole matrices in uncoupled spin basis

The tp states in a general 'uncoupled' spin basis can be described as

$$|i_{m_i} j_{m_j}\rangle = a^\dagger_{j_{m_j}} a^\dagger_{i_{m_i}} |\rangle$$

where $m_i = \frac{1}{2}$ or $-\frac{1}{2}$, denoting the spin projection of the electron in orbital i. In the following we will either specify the spin projection m_i or use the shorthand notation a_i to denote $a_{i_{\frac{1}{2}}}$ and $a_{\bar{i}}$ to denote $a_{i_{-\frac{1}{2}}}$.
The tp states are eigenfunctions of $\hat{S}_z$

$$\hat{S}_z |i_{m_i} j_{m_j}\rangle = (m_i + m_j)|i_{m_i} j_{m_j}\rangle \quad (m_i + m_j) \in \{-1, 0, 1\}$$

but not necessarily of $\hat{S}^2$.

The elements of the 2DM for a pure spin state $|SM\rangle$, which are given by the operators $|k_{m_k} l_{m_l}\rangle\langle i_{m_i} j_{m_j}|$ acting on the state $|SM\rangle$, can only return a non-zero value if they can couple to an overall zero spin projection, $M' = 0$. Equivalently, the only blocks of the 2DM that can be non-zero after spin integration from a pure spin state wavefunction are

$$\begin{pmatrix} \Gamma^{\alpha\alpha\alpha\alpha} & & \\ 0 & \Gamma^{\beta\beta\beta\beta} & \\ 0 & 0 & \Gamma^{\alpha\beta\alpha\beta} \end{pmatrix} \tag{2.1}$$

where the superscripts $\sigma_i\sigma_j\sigma_k\sigma_l$ are used to denote the spin projections of the orbital indices $ijkl$ of the elements Γ_{ijkl} that make up the block. Because of antisymmetry, the blocks $\Gamma^{ABBA}, \Gamma^{BAAB}$ and Γ^{ABAB} are redundant. Likewise, the 1DM has a spin block structure

$$\begin{pmatrix} \gamma^{\alpha\alpha} & 0 \\ 0 & \gamma^{\beta\beta} \end{pmatrix} \tag{2.2}$$

In the uncoupled spin basis, the 1DM and 2DM are thus described in terms of sp and tp states which are eigenfunctions of $\hat{S}_z$ but not necessarily of $\hat{S}^2$.

Representation of two-particle/hole matrices in spin coupled basis

A spin coupled basis consists of tp states which behave like proper spin states under the spin operators. A creation operator $\hat{A}^{SM}_{kl}$ that creates a two particle/hole state with spin S' and spin projection M' must satisfy

$$\begin{aligned} [\hat{S}_z, \hat{A}^{S'M'}_{kl}] &= M' \ \hat{A}^{S'M'}_{kl} \\ [\hat{S}^+, \hat{A}^{S'M'}_{kl}] &= \sqrt{(S'-M')(S'+M'+1)} \ \hat{A}^{S'M'+1}_{kl} \\ [\hat{S}^-, \hat{A}^{S'M'}_{kl}] &= \sqrt{(S'+M')(S'-M'+1)} \ \hat{A}^{S'M'-1}_{kl} \end{aligned}$$

where its spin S' can take the values 0 or 1, and its spin projection M' the values $-S', \ldots, S' \in \{-1, 0, 1\}$. Such spin coupled two-particle, two-hole and particle-hole creation operators to describe the P-, Q- and G-matrix can be expressed as a linear combination of the corresponding operators in the uncoupled spin basis in the following manner.

Spin coupled two-particle states

Spin coupled tp creation operators that form a basis for the 2DM can be composed from two particle creation operators $a_k^\dagger a_l^\dagger$ with suitable normalization through

$$\begin{aligned}
\hat{A}_{kl}^{0\ 0} &= \frac{1}{\sqrt{2(1+\delta_{kl})}}(a_k^\dagger a_{\bar{l}}^\dagger + a_l^\dagger a_{\bar{k}}^\dagger) \\
\hat{A}_{kl}^{1-1} &= a_{\bar{k}}^\dagger a_{\bar{l}}^\dagger \\
\hat{A}_{kl}^{1\ 0} &= \frac{1}{\sqrt{2}}(a_k^\dagger a_{\bar{l}}^\dagger - a_l^\dagger a_{\bar{k}}^\dagger) \\
\hat{A}_{kl}^{1\ 1} &= a_k^\dagger a_l^\dagger
\end{aligned}$$

The operator $\hat{A}_{kl}^{0\ 0}$ generates a singlet pair state and the operators $\hat{A}_{kl}^{1-1}, \hat{A}_{kl}^{1\ 0}$ and $\hat{A}_{kl}^{1\ 1}$ generate a triplet pair state. In general, elements $ijkl$ of the 2DM in the spin coupled basis are therefore

$$\langle\Psi|\hat{A}_{kl}^{S_2\ M_2}\left(\hat{A}_{ij}^{S_1\ M_1}\right)^\dagger|\Psi\rangle$$

although spin symmetry implies that only certain combinations of S_2, S_1 and M_2, M_1 can couple to a non-zero element (section 2.3).

Spin coupled two-hole states

The same spin coupled tp state creation operators may serve as a basis for the spin coupled representation of the Q-matrix, since its representation simply involves the Hermitian adjoint operators from those that express the 2DM

$$\langle\Psi|\left(\hat{A}_{kl}^{S_2\ M_2}\right)^\dagger\hat{A}_{ij}^{S_1\ M_1}|\Psi\rangle$$

Spin coupled particle-hole states

A spin coupled basis for the G-matrix can be generated by means of the following

particle-hole creation operators

$$\hat{A}^{0\ 0}_{kl} = \frac{1}{\sqrt{2}}(a^\dagger_k a_l + a^\dagger_{\bar{k}} a_{\bar{l}})$$
$$\hat{A}^{1-1}_{kl} = a^\dagger_{\bar{k}} a_l$$
$$\hat{A}^{1\ 0}_{kl} = \frac{1}{\sqrt{2}}(a^\dagger_k a_l - a^\dagger_{\bar{k}} a_{\bar{l}})$$
$$\hat{A}^{1\ 1}_{kl} = a^\dagger_k a_{\bar{l}}$$

such that the G-matrix in spin-coupled basis in its most general form has elements

$$\langle\Psi|\hat{A}^{S_2\ M_2}_{kl}\left(\hat{A}^{S_1\ M_1}_{ij}\right)^\dagger|\Psi\rangle$$

The resulting G-map, expressed in terms of a matrix in two particle space, can also be applied to a matrix in particle-hole space, with the only difference being that the latter is not antisymmetrical.
The spin coupled particle-hole creation operators also constitute a basis for the 1DM in spin coupled representation.

To recapitulate, the second order particle-particle, hole-hole and particle-hole matrix, P-, Q- and G-matrix, have elements of the form

$$\langle SM|\left(\hat{A}^{S_2\ M_2}_{kl}\right)^\dagger \hat{A}^{S_1,M_1}_{ij}|SM\rangle \tag{2.3}$$

where the operators $\hat{A}^{S_1 M_1}_{ij}$ are spin coupled two-particle, two-hole or particle-hole generators, respectively. The first order particle and hole density matrix, introduced as the p- and q-matrix in section 1.3.3 of chapter 1, have elements of the form

$$\langle SM|\hat{A}^{S'M'}_{ij}|SM\rangle$$

where the operators $\hat{A}^{S'M'}_{ij}$ are spin coupled particle-hole or hole-particle generators, respectively.

When the state under consideration is a pure spin state $|SM\rangle$ with definite spin eigenvalue $S(S+1)$ and spin projection M, only certain combinations of operators $\left(\hat{A}^{S_2\ M_2}_{kl}\right)^\dagger \hat{A}^{S_1,M_1}_{ij}$ can couple to a symmetry that does not necessarily lead to a zero expectation value. The P-,Q- and G-matrices therefore have a

specific block structure that depends on the state under consideration. The following sections eloborate on the consequences of spin symmetry and other spin conditions on the 2DM.

2.3 S-representability conditions for pure spin states

2.3.1 Spin symmetry

Although correct block structure of the 2DM induced by spin symmetry can be considered an S-representability condition, it does not influence the energy in general, because both the Hamiltonian, which is the driving force behind the optimization, and the N-representability constraints have correct spin symmetry by definition. Except for a few special cases where spin symmetry induces additional symmetry in the 2DM, the variationally optimized 2DM will inherit the right spin symmetry from the Hamiltonian.

Moreover, a spin adapted basis makes it easier to exploit such additional spin symmetries, which occur in systems with zero spin projection ($M = 0$) and zero spin ($S = 0$). These additional symmetries may impose an active constraint on the energy and are much more difficult to exploit in an uncoupled spin basis.

The next paragraphs examine the effect of spin symmetry on the structure of the 2DM, and consider necessary conditions on spin that the 2DM must satisfy if it is derivable from a pure spin state $|SM\rangle$.

Implications of spin symmetry on the structure of the 2DM in a spin coupled basis

The only elements $\langle SM|\hat{A}^{S_2\ M_2}_{kl}\left(\hat{A}^{S_1,M_1}_{ij}\right)^\dagger|SM\rangle$ that are not necessarily zero by spin symmetry considerations, are those in which the tp creation operator $\hat{A}^{S_2M_2}_{kl}$ and annihilation operator $(\hat{A}^{S_1M_1}_{ij})^\dagger$ are coupled to a total spherical tensor operator that may have a nonzero expectation value acting on the state $|SM\rangle$.

Because the Hermitian conjugate of a spherical tensor operator $(\hat{A}^{SM}_{kl})^\dagger$ is only guaranteed to be a spherical tensor operator itself upon inclusion of a phase $(-1)^{S-M}$, we will consider the operator $\hat{B}^{S-M}_{kl} \equiv (-1)^{S-M}(\hat{A}^{SM}_{kl})^\dagger$. The spherical tensor operators $A^{S_2M_2}_{kl}$ and $\hat{B}^{S-M}_{kl}$ can couple to a total spherical tensor operator $[\hat{A}^{S_2}_{kl} \otimes \hat{B}^{S_1}_{ij}]^{S'M'}$ with spin S' and spin projection M'.

Any matrix element in the spin coupled two particle/hole basis can be expressed in terms of matrix elements of a total spherical tensor operator

$$\langle SM|\hat{A}^{S_2M_2}_{kl}\hat{B}^{S_1M_1}_{ij}|SM\rangle = \sum_{S'M'} \langle SM|[\hat{A}^{S_2}_{kl} \otimes \hat{B}^{S_1}_{ij}]^{S'M'}|SM\rangle(S_2\ M_2\ S_1\ M_1\ |S'\ M') \quad (2.4)$$

This is a unitary transformation, given by the Clebsch-Gordan coefficients. Under which conditions can these terms make a non-zero contribution?
The expectation values $\langle SM|[\hat{A}^{S_2}_{kl} \otimes \hat{B}^{S_1}_{ij}]^{S'M'}|SM\rangle$ occurring in the above summation can only be nonzero if

$$0 \leq S' \leq 2S$$
$$M' = 0$$
$$\text{if } M = M' = 0: \ S' \text{ is even}$$

The Clebsch-Gordan coefficients $(S_2\ M_2\ S_1\ M_1|S'\ M')$ occurring in the summation can only be nonzero if

$$|S_1 - S_2| \leq S' \leq |S_1 + S_2|$$
$$M_1 + M_2 = M'$$
$$\text{if } M_1 = M_2 = M' = 0: \ S' + S_1 + S_2 \text{ is even}$$

Since these conditions imply that $M_1 = -M_2$, spin coupled tp states with different spin projection must be orthogonal. This requirement makes the 2DM diagonal in the spin projection of the spin coupled tp basis, which can take the values $M_1 = -M_2 \in \{-1, 0, 1\}$. Hence the P-, Q- and G-matrix of a pure spin state in general have three non-zero spin blocks which can be labeled according

to the spin eigenvalues S_1, S_2 and the mutual spin projection $M_1 = M_2 \equiv M'$ of the spin-coupled tp creation- and annihilation operator, so we introduce the notation

$$\Gamma^{S_1 S_2\ M'}_{ijkl} = \langle\Psi|\hat{A}^{S_2,M'}_{kl}\left(\hat{A}^{S_1\ M'}_{ij}\right)^\dagger|\Psi\rangle \text{ for } M' = -1, 0, 1$$

As for the coupling of the spins S_1 and S_2 of the two particle/hole creation and annihilation operator in formula (2.4), there are several possibilities, depending on the molecular spin state $|SM\rangle$ under consideration.

Singlet spin states

For a zero spin state, $|00\rangle$, the tp creation and annihilation operator must couple to a total spherical tensor operator with zero spin, $S' = 0$. This implies that $S_1 = S_2 \in \{0, 1\}$. Therefore, the 2DM is not only diagonal in the spin projection of the spin coupled tp basis, but also in its spin. Moreover, because a zero spin state must be completely symmetrical in terms of α and β electrons, all triplet blocks with $S_1 = S_2 = 1$ are equivalent, so only one needs to be stored. This follows directly from their coupling (2.4),

$$\begin{aligned}\langle 00|\hat{A}^{S_2 M_2}_{kl}\hat{B}^{S_2 -M_2}_{ij}|00\rangle &= (-1)^{S_2-M_2}\langle 00|\hat{A}^{S_2 M_2}_{kl}\left(\hat{A}^{S_2 M_2}_{ij}\right)^\dagger|00\rangle \\ &= \langle 00|[\hat{A}^{S_2}_{kl}\otimes\hat{B}^{S_2}_{ij}]^{00}|00\rangle(0\ 0\mid S_2\ M_2\ S_2\ -M_2) \\ &= (-1)^{S_2-M_2}\frac{1}{\sqrt{2S_2+1}}\langle 00|[\hat{A}^{S_2}_{kl}\otimes\hat{B}^{S_2}_{ij}]^{00}|00\rangle\end{aligned}$$

Therefore

$$\langle 00|\hat{A}^{S_2 M}_{kl}\left(\hat{A}^{S_2 M}_{ij}\right)^\dagger|00\rangle = \frac{1}{\sqrt{2S_2+1}}\langle 00|[\hat{A}^{S_2}_{kl}\otimes\hat{B}^{S_2}_{ij}]^{00}|00\rangle \tag{2.5}$$

Since the right hand side is independent of the spin projection M_2 of the spin coupled tp basis, all triplet blocks with $S_1 = S_2 = 1$ are equal.

In summary, there are only two linearly independent blocks in the 2DM for a pure spin singlet state, a 'singlet' block with $S_1 = S_2 = 0$ and a 'triplet' block with $S_1 = S_2 = 1$.

$$S = 0,\ M = 0: \quad \begin{pmatrix} \Gamma^{00\ 0} & 0 \\ 0 & \Gamma^{11\ 0} = \Gamma^{11\ 1} = \Gamma^{11\ -1} \end{pmatrix} \tag{2.6}$$

Non-singlet spin states

There are two cases to discern for non-singlet spin states, depending on their spin projection.

For a non-zero spin state with zero spin projection, $|S0\rangle$, the 2DM elements $\langle S0|\hat{A}^{S_2M_2}_{kl}\left(\hat{A}^{S_1M_2}_{ij}\right)^\dagger|S0\rangle$ can only be nonzero if the spins S_1 and S_2 can couple to an even spin S'. Since S_1 and S_2 can only take the values $0, 1$ this also implies that the 2DM is diagonal in the spin of the tp states. The $M_1 = M_2 = 0$ block of the 2DM thus splits into two blocks, $\Gamma^{00\ 0}$ and $\Gamma^{11\ 0}$.

A further reduction in storage and computation requirements can be made by noting that the blocks $\Gamma^{11\ 1}$ and $\Gamma^{11\ -1}$ must be equal.

$$\begin{aligned}\langle S0|\hat{A}^{1M_2}_{kl}\left(\hat{A}^{1M_2}_{ij}\right)^\dagger|S0\rangle = (-1)^{1-M_2}\langle S0|[\hat{A}^1_{kl}\otimes\hat{B}^1_{ij}]^{00}|S0\rangle(0\ 0\mid 1\ M_2\ 1\ -M_2)\\ +(-1)^{1-M_2}\langle S0|[\hat{A}^1_{kl}\otimes\hat{B}^1_{ij}]^{20}|S0\rangle(2\ 0\mid 1\ M_2\ 1\ -M_2)\end{aligned}$$

This leads to the same expression for $M_2 = 1$ and $M_2 = -1$. The $\Gamma^{11\ 0}$ block is different, however.
In summary, the structure of the 2DM in spin coupled tp basis for a zero spin projection system is

$$S \neq 0,\ M = 0: \quad \begin{pmatrix} \Gamma^{00\ 0} & 0 & 0 \\ 0 & \Gamma^{11\ 0} & 0 \\ 0 & 0 & \Gamma^{11\ 1} = \Gamma^{11\ -1} \end{pmatrix} \tag{2.7}$$

For any other nonzero spin state, the 2DM is only diagonal in the spin projection of the tp basis. It thus has a structure

$$S \neq 0,\ M \neq 0: \quad \begin{pmatrix} \Gamma^{00\ 0} & \Gamma^{01\ 0} & 0 & 0 \\ \Gamma^{10\ 0} & \Gamma^{11\ 0} & 0 & 0 \\ 0 & 0 & \Gamma^{11\ 1} & 0 \\ 0 & 0 & 0 & \Gamma^{11\ -1} \end{pmatrix} \tag{2.8}$$

Implications of spin symmetry on the structure of the 2DM in an uncoupled spin basis

The additional spin symmetries for zero spin projection or zero spin states are much more easily exploited in the spin coupled basis. They can be easily imposed by construction in the spin coupled basis through their block diagonal structure but do not take such a simple form in a general uncoupled spin basis.
For a state with zero spin projection $|S0\rangle$, the orthogonality between the symmetrical and antisymmetrical combination of $\alpha\beta$ pairs, the spin coupled two particle/hole states $\hat{A}^{00}|\rangle$ and $\hat{A}^{10}|\rangle$, implies that

$$\Gamma_{i\bar{j}k\bar{l}} = \Gamma_{j\bar{i}l\bar{k}} \quad \text{for } M = 0 \tag{2.9}$$

in an uncoupled spin basis.

Additionally, for a zero spin state $|00\rangle$, the degeneracy between the antisymmetrical combination of opposite spin pairs and same spin pairs, i.e. degeneracy of the three triplet blocks in (2.6), imposes in an uncoupled spin basis that

$$\Gamma_{ijkl} = \Gamma_{\bar{i}\bar{j}\bar{k}\bar{l}} = \Gamma_{i\bar{j}k\bar{l}} - \Gamma_{i\bar{j}l\bar{k}} \quad \text{for } S = 0 \tag{2.10}$$

which would have to be imposed as constraint in an uncoupled spin basis. The symmetry (2.10) which follows from zero spin truly constrains the system, whereas the symmetry (2.9) which follows from its zero spin projection is already satisfied by minimization under a spin independent Hamiltonian and spin constraints that do not alter this symmetry.

2.3.2 Basic S-representability constraints

Because the 2DM only carries information up to two-electron interactions, imposing that it is derivable from an N-electron wavefunction that is a proper eigenfunction of $\hat{S}^2$ and $\hat{S}_z$ is a difficult problem. For 2-electron operators such as $\hat{S}^2$, only their expectation value is directly available.
One-electron operators such as $\hat{S}_z$ and $\hat{S}^+$ can be used to formulate a set of constraints for pure spin states, by demanding that – at least on the level of the 2DM – the state is an eigenstate of the one-electron operator.

$\hat{S}^2$ based constraints

The $\hat{S}^2$ operator can be expressed in two particle space as

$$\begin{aligned}\hat{S}^2 &= \hat{S}_z + \hat{S}_z^2 + \hat{S}^-\hat{S}^+ \\ &= \frac{1}{2}\sum_i a_i^\dagger a_i - a_{\bar{i}}^\dagger a_{\bar{i}} + \frac{1}{4}\sum_{ij}(a_i^\dagger a_i - a_{\bar{i}}^\dagger a_{\bar{i}})(a_j^\dagger a_j - a_{\bar{j}}^\dagger a_{\bar{j}}) + \sum_{ij} a_{\bar{i}}^\dagger a_i a_{\bar{j}}^\dagger a_{\bar{j}} \\ &= \sum_{ij}\frac{1}{4}\Big(a_i^\dagger a_j^\dagger a_j a_i + a_{\bar{i}}^\dagger a_{\bar{j}}^\dagger a_{\bar{j}} a_{\bar{i}} - 2a_i^\dagger a_{\bar{j}}^\dagger a_{\bar{j}} a_i\Big) - a_j^\dagger a_{\bar{i}}^\dagger a_{\bar{j}} a_i + \frac{3}{4}\sum_i a_i^\dagger a_i + a_{\bar{i}}^\dagger a_{\bar{i}}\end{aligned}$$

Taking the normalization of the 2DM into account, its expectation value is thus

$$\begin{aligned}\langle\hat{S}^2\rangle &= \frac{1}{4}\Big(N(N-1) - 4\sum_{ij}\Gamma_{i\bar{j}i\bar{j}}\Big) - \sum_{ij}\Gamma_{i\bar{j}j\bar{i}} + \frac{3}{4}N \\ &= \frac{N}{2}\Big(\frac{N}{2}+1\Big) - \sum_{ij}(\Gamma_{i\bar{j}i\bar{j}} + \Gamma_{i\bar{j}j\bar{i}}) \qquad (2.11)\end{aligned}$$

or in spin coupled basis

$$\langle\hat{S}^2\rangle = \frac{N}{2}\Big(\frac{N}{2}+1\Big) - tr\ \Gamma^{00\ 0} \qquad (2.12)$$

$\hat{S}_z$ based constraints

As the spin projection is a one-electron operator, we can do more than just specify its expectation value

$$\langle\hat{S}_z\rangle = \frac{1}{N-1}(tr\ \Gamma^{\alpha\alpha\alpha\alpha} - tr\ \Gamma^{\beta\beta\beta\beta})$$

for a pure spin state. As the state $|SM\rangle$ must be an eigenfunction of $\hat{S}_z$

$$(N\ \hat{S}_z - M\ \hat{N})|SM\rangle = 0 \qquad (2.13)$$

where $\hat{N}$ is the number operator and N the number of electrons. This condition implies that the vector corresponding to this operator lies in the nullspace of the G^0 block of the G-matrix:

$$\forall k,l: \quad \sum_i N\ G_{iikl}^{10\ 0} - 2M\ G_{iikl}^{00\ 0} = 0 \qquad (2.14)$$

$$\sum_i N\ G_{iikl}^{11\ 0} - 2M\ G_{iikl}^{10\ 0} = 0 \qquad (2.15)$$

This condition ensures that the spin blocks of the 1DM can be derived from the 2DM both by contraction over α orbital indices and by contraction over β orbital indices. For this reason, it is sometimes referred to as the 'contraction condition'. In an uncoupled basis, it expresses that

$$\Big(\frac{N}{2} - M\Big) \sum_i \Gamma_{kili} = \Big(\frac{N}{2} + M - 1\Big) \sum_i \Gamma_{k\bar{i}l\bar{i}} \tag{2.16}$$

$$\Big(\frac{N}{2} + M\Big) \sum_i \Gamma_{\bar{k}\bar{i}l\bar{i}} = \Big(\frac{N}{2} + M - 1\Big) \sum_i \Gamma_{\bar{k}i\bar{l}i} \tag{2.17}$$

As this must hold for all $k \leq l \in \{1, \ldots \frac{K}{2}\}$, it involves $\frac{K}{2}(\frac{K}{2}+1)$ conditions on the 2DM, although the number of practically relevant conditions is less if spatial symmetry is taken into account. As is obvious from formula (2.13), it implies correct $\hat{S}_z$ expectation value, and together with the normalization of the whole 2DM, $tr\Gamma = N(N-1)/2$, it implies normalization of the spin blocks to

$$tr\ \Gamma^{11\ 1} = \frac{1}{2}\left(\frac{N}{2} + M\right)\left(\frac{N}{2} + M - 1\right)$$

$$tr\ \Gamma^{11\ -1} = \frac{1}{2}\left(\frac{N}{2} - M\right)\left(\frac{N}{2} - M - 1\right)$$

$$tr\ \Gamma^{00\ 0} + tr\ \Gamma^{11\ 0} = \left(\frac{N}{2} + M\right)\left(\frac{N}{2} - M\right)$$

$\hat{S}^+$ based constraints

The maximal spin projection for a spin-S state, $M = S$, must satisfy the condition

$$S^+|SS\rangle = 0 \tag{2.18}$$

This condition forces the vector corresponding to the $\hat{S}^+$ operator to lie in the nullspace of the G^1 block of the G-matrix:

$$\forall k, l: \quad \sum_i G^{11\ 1}_{iikl} = 0 \tag{2.19}$$

which implies $\frac{K^2}{4}$ additional conditions on the 2DM. Along with the contraction condition, it imposes the correct $\hat{S}^2$ expectation value, because the contraction condition imposes

$$tr\ \Gamma^{00\ 0} + tr\ \Gamma^{11\ 0} = \Big(\frac{N}{2} - M\Big)\Big(\frac{N}{2} + M\Big)$$

and the maximal spin condition imposes

$$tr\ \Gamma^{00\ 0} - tr\ \Gamma^{11\ 0} = 2\Big(\frac{N}{2} + M\Big)$$

The $\hat{S}^2$ expectation value (2.12) for the maximal spin projection $M = S$ thus becomes

$$\begin{aligned}\langle \hat{S}^2 \rangle &= \frac{N}{2}\Big(\frac{N}{2}+1\Big) - \Big(\frac{N}{2} - S\Big)\Big(\frac{N}{2} + S + 1\Big) \\ &= S(S+1)\end{aligned}$$

Of course, this constraint only holds for maximal spin states – or an equivalent constraint for minimal spin states. A more general form of this constraint, which holds for other pure state spin projections as well, can be derived by considering the relationship between the first order density matrix and transition density matrix elements.

2.3.3 Relationship between first order density matrix and transition density matrix elements for different spin projections

The first order density matrix and transition density matrix elements for different spin projections are related, which is directly expressed by the Wigner-Eckart theorem.[80] It states that the action of any spherical tensor operator $\hat{A}^{S'M'}$ on states with different spin projections is proportional; the proportionality factor is the reduced matrix element, denoted with double lines in the bra-ket notation

Wigner-Eckart theorem

$$\langle \tilde{S}\tilde{M}|\hat{A}^{S'M'}|SM\rangle = (-1)^{\tilde{S}-\tilde{M}} \begin{pmatrix} \tilde{S} & S' & S \\ -\tilde{M} & M' & M \end{pmatrix} \langle \tilde{S}||\hat{A}^{S'M'}||S\rangle \quad (2.20)$$

This theorem directly links the first order density and transition density matrix elements for states with different spin projections to each other.

The first order transition density matrix elements that we are interested in are, with $b_{lm} = (-1)^{\frac{1}{2}+m} a_{l-m}$,

$$\begin{aligned}\langle SM|[a_k^\dagger \otimes b_l]^{11}|SM-1\rangle &= \Big(\frac{1}{2}\ \frac{1}{2}\ \frac{1}{2}\ \frac{1}{2}\ |\ 1\ 1\Big)\langle SM|a^\dagger_{k\,\frac{1}{2}} b_{l\,\frac{1}{2}}|SM-1\rangle \\ &= -\langle SM|a_k^\dagger a_{\bar{l}}|SM-1\rangle \end{aligned} \tag{2.21}$$

According to the Wigner-Eckart theorem, the spin projection dependence can be filtered out

$$\begin{aligned}\langle SM|[a_k^\dagger \otimes b_l]^{11}|SM-1\rangle &= (-1)^{S-M}\begin{pmatrix} S & 1 & S \\ -M & 1 & M-1\end{pmatrix}\langle S||[a_k^\dagger \otimes b_l]^1||S\rangle \\ &= -\frac{1}{\sqrt{2}}\frac{\sqrt{(S+M)(S-M+1)}}{\sqrt{2S+1}\sqrt{S(S+1)}}\langle S||[a_k^\dagger \otimes b_l]^1||S\rangle \end{aligned} \tag{2.22}$$

Since the same reasoning applies to the transition density matrix element for the state with spin projection $M+1$ as to the state with spin projection M

$$\langle SM+1|[a_k^\dagger \otimes b_l]^{11}|SM\rangle = -\frac{1}{\sqrt{2}}\frac{\sqrt{(S+M+1)(S-M)}}{\sqrt{2S+1}\sqrt{S(S+1)}}\langle S||[a_k^\dagger \otimes b_l]^1||S\rangle$$

Therefore these two transition density matrix elements are proportional

$$\begin{aligned}&\sqrt{(S+M)(S-M+1)}\langle SM+1|[a_k^\dagger \otimes b_l]^{11}|SM\rangle = \\ &\qquad\qquad \sqrt{(S-M)(S+M+1)}\langle SM|[a_k^\dagger \otimes b_l]^{11}|SM-1\rangle\end{aligned}$$

or, equivalently,

$$\begin{aligned}&\sqrt{(S+M)(S-M+1)}\langle SM+1|a_k^\dagger a_{\bar{l}}|SM\rangle = \\ &\qquad\qquad \sqrt{(S-M)(S+M+1)}\langle SM|a_k^\dagger a_{\bar{l}}|SM-1\rangle\end{aligned} \tag{2.23}$$

Similarly, the elements of the 1DM in a spin coupled basis for different spin projections are related. First of all, the *total spin density matrix*, composed of the sum of both $\alpha\alpha$ and $\beta\beta$ components of the 1DM in an uncoupled representation,

is independent of the system's spin projection M

$$[a_k^\dagger \otimes b_l]^{00} = \sum_m \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ m & -m & 0 \end{pmatrix} a_{km}^\dagger b_{l-m} = \frac{1}{\sqrt{2}} \left(a_k^\dagger a_l + a_{\bar{k}}^\dagger a_{\bar{l}} \right)$$

since the Wigner-Eckart theorem yields

$$\langle SM|[a_k^\dagger \otimes b_l]^{00}|SM\rangle = (-1)^{S-M} \begin{pmatrix} S & 0 & S \\ -M & 0 & M \end{pmatrix} \langle S||[a_k^\dagger \otimes b_l]^0||S\rangle = \frac{1}{\sqrt{2S+1}} \langle S||[a_k^\dagger \otimes b_l]^0||S\rangle$$

which is independent of the spin projection M. Therefore

$$\langle SM|a_k^\dagger a_l + a_{\bar{k}}^\dagger a_{\bar{l}}|SM\rangle = \langle SM+1|a_k^\dagger a_l + a_{\bar{k}}^\dagger a_{\bar{l}}|SM+1\rangle \tag{2.24}$$

Secondly, the ratio of the elements of the *spin density matrices* for states with different spin projections $M, \tilde{M}$, composed of the difference between the $\alpha\alpha$ and $\beta\beta$ components of the 1DM in an uncoupled representation, is given by the ratio of the spin projections $\frac{M}{\tilde{M}}$,

$$[a_k^\dagger \otimes b_l]^{10} = \sum_m \sqrt{3} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 1 \\ m & -m & 0 \end{pmatrix} a_{km}^\dagger b_{l-m} = \frac{1}{\sqrt{2}} (a_k^\dagger a_l - a_{\bar{k}}^\dagger a_{\bar{l}}) \tag{2.25}$$

because their spin projection dependence is factored out by the Wigner-Eckart theorem as follows

$$\langle SM|[a_k^\dagger \otimes b_l]^{10}|SM\rangle = (-1)^{S-M} \begin{pmatrix} S & 1 & S \\ -M & 0 & M \end{pmatrix} \langle S||[a_k^\dagger \otimes b_l]^1||S\rangle = \frac{M}{\sqrt{2S+1}\sqrt{S(S+1)}} \langle S||[a_k^\dagger \otimes b_l]^1||S\rangle \tag{2.26}$$

This result is directly proportional to the state's spin projection. Equivalently,

$$(M+1)\langle SM|a_k^\dagger a_l - a_{\bar{k}}^\dagger a_{\bar{l}}|SM\rangle = M\langle SM+1|a_k^\dagger a_l - a_{\bar{k}}^\dagger a_{\bar{l}}|SM+1\rangle \tag{2.27}$$

Moreover, the 1DM matrix elements in spin coupled basis are also proportional to the transition density matrix elements

$$\langle SM|a_k^\dagger a_{\bar{l}}|SM-1\rangle = \frac{1}{2M}\sqrt{(S+M)(S-M+1)}\langle SM|a_k^\dagger a_l - a_{\bar{k}}^\dagger a_{\bar{l}}|SM\rangle \quad (2.28)$$

2.3.4 S-representability constraints derived from relations between first order density and transition density matrix elements

The relationship (2.23) between the first order transition density matrices defines additional constraints on the 2DM, since the first order transition density matrix elements for a state $|SM\rangle$ are available through the 2DM

$$\begin{aligned}\sqrt{(S+M)(S-M+1)}\langle SM|a_k a_{\bar{l}}|SM-1\rangle &= \langle SM|a_k a_{\bar{l}} S^-|SM\rangle \\ &= \sum_i G_{i\bar{i}k\bar{l}} \qquad (2.29)\\ \sqrt{(S-M)(S+M+1)}\langle SM|a_{\bar{k}} a_l|SM+1\rangle &= \langle SM|a_{\bar{k}} a_l S^+|SM\rangle \\ &= \sum_i G_{\bar{i}i\bar{k}l} \qquad (2.30)\end{aligned}$$

Because both first order transition density matrices are proportional

$$(S-M)(S+M+1)\sum_i G_{i\bar{i}k\bar{l}} = (S+M)(S-M+1)G_{\bar{i}i\bar{k}l} \quad (2.31)$$

which holds for any index k and l, so it imposes $\frac{K^2}{4}$ conditions on the 2DM.

There are several ways of deriving this condition; another way is to consider

$$\begin{aligned}&\langle SM|S^-[a_k^\dagger a_l, S^+]|SM\rangle \\ &\quad = (S-M)(S+M+1)(\langle \Psi^{SM+1}|a_k^\dagger a_l|\Psi^{SM+1}\rangle - \langle SM|a_k^\dagger a_l|SM\rangle)\end{aligned}$$

and replace its dependence on the 1DM for the state $|SM+1\rangle$ by means of the relationships between the 1DM's for the states $|SM+1\rangle$ and $|SM\rangle$ (2.24 and 2.27).

As a special case of this condition, a state with maximal spin projection must satisfy

$$\sum_i G_{\bar{i}i\bar{k}l} = 0$$
$$\sum_i \Gamma_{l\bar{i}i\bar{k}} = \gamma_{\bar{k}l} \tag{2.32}$$

which imposes on the level of the 2DM that $S^+|\Psi^{SS}\rangle = 0$.

States with zero spin projection are another special case. For these states, the condition simply imposes that

$$\sum_i G_{i\bar{i}k\bar{l}} = G_{\bar{i}i\bar{k}l}$$
$$\gamma_{kl} - \sum_i \Gamma_{i\bar{l}k\bar{i}} = \gamma_{\bar{k}\bar{l}} - \sum_i \Gamma_{l\bar{i}i\bar{k}} \tag{2.33}$$

However, because of spin symmetry (2.7), this constraint does not add a condition in spin coupled basis.

2.4 S-representability conditions for ensemble spin states

Suppose that we allow the 2DM to represent an ensemble composed of different spin projections of the spin state under consideration. What conditions on spin properties must the 2DM then fulfil? Since the composition of the ensemble is not fixed, most of the constraints considered above for pure spin states, cannot be extended to a general spin ensemble without making assumptions about its composition.

2.4.1 Implications of spin symmetry on the structure of the 2DM

In contrast to a pure spin state 2DM, the 2DM for an ensemble spin state does not need to have a block structure by spin symmetry, although in practice any

v2DM calculation on such a system under the usual spin constraints, which do have a spin block structure, will return a 2DM with the same block structure. Because the different spin projections are degenerate, a general wavefunction corresponding to a spin quantum number S may have the form

$$\Psi = \sum_M c_M |SM\rangle \quad \text{with} \sum_M c_M^2 = 1$$

with $\langle \hat{S}_z \rangle = \sum_M c_M^2 M$. The N-th order density matrix may be a mixed state

$$\Gamma^{(N)} = \sum_n w_n \sum_{M\ M'} c_M^n c_{M'}^n |SM\rangle\langle SM'| \quad 0 \leq w_n \leq 1, \sum_n w_n = 1, \sum_M (c_M^n)^2 = 1$$

with $\langle \hat{S}_z \rangle = \sum_n w_n \sum_M (c_M^n)^2 M$. The corresponding 2DM is

$$\Gamma^{(2)}_{ijkl} = \sum_n w_n \sum_{M\ M'} c_M^n c_{M'}^n \langle SM'|a_k^\dagger a_l^\dagger a_j a_i|SM\rangle$$

This form of 2DM in general does not lead to a block diagonal structure, even the off-diagonal blocks in terms of the tp state's spin projection such as $\Gamma^{\alpha\alpha\alpha\beta}$ may be non-zero because of a contribution from the transition density matrix elements $\langle SM'|a_k^\dagger a_l^\dagger a_j a_i|SM\rangle \quad M \neq M'$.

However, in order to reduce computational cost and save memory, the 2DM may be restricted to have the same symmetry as a corresponding pure state would have, since this is a valid, albeit not necessary, representation of the spin state. In the case that a state with zero $\hat{S}_z$ expectation value is considered, a further reduction in computational requirements can be made by assuming an ensemble average over all states of the spin multiplet. Therefore we will distinguish two cases.

Non-zero $\hat{S}_z$ expectation value

Although the ensemble does not need to have a block structure, it can be

constrained to have the same block structure as a pure non-singlet state.

$$\langle \hat{S}_z \rangle \neq 0 : \quad \begin{pmatrix} \Gamma^{00\ 0} & \Gamma^{01\ 0} & 0 & 0 \\ \Gamma^{10\ 0} & \Gamma^{11\ 0} & 0 & 0 \\ 0 & 0 & \Gamma^{11\ 1} & 0 \\ 0 & 0 & 0 & \Gamma^{11\ -1} \end{pmatrix} \tag{2.34}$$

Zero $\hat{S}_z$ expectation value

The dimension of the 2DM for an ensemble with $\langle \hat{S}_z \rangle = 0$ can be significantly reduced by choosing the composition of the ensemble as an average over all spin projections $M = -S, \ldots, S$.

$$\Gamma_{ijkl}^{S_1 S_2\ M_2} = \frac{1}{2S+1} \sum_M \langle SM | \left(A_{kl}^{S_2 M_2} \right) A_{ij}^{S_1 M_2} | SM \rangle \tag{2.35}$$

Taking the spin averaged ensemble makes the α and β spins equivalent, such that the 2DM can be reduced to two non-zero blocks, similar to that of a singlet state (2.6).

$$\begin{aligned}
&\frac{1}{2S+1} \sum_M \langle SM | A_{kl}^{S_2 M_2} \left(A_{ij}^{S_1 M_1} \right)^\dagger | SM \rangle \\
&= \frac{1}{2S+1} \sum_M \sum_{M'S'} \langle SM | [A_{kl}^{S_2} \otimes B_{ij}^{S_1}]^{S'M'} | SM \rangle (-1)^{S_1 - M_1} (-1)^{S_1 - S_2 + M'} \\
&\quad [S'] \begin{pmatrix} S_1 & S_2 & S' \\ -M_1 & M_2 & -M' \end{pmatrix} \\
&= \frac{1}{2S+1} \sum_M \sum_{M'S'} \langle S || [A_{kl}^{S_2} \otimes B_{ij}^{S_1}]^{S'} || S \rangle (-1)^{S-M} (-1)^{S_1 - M_1} (-1)^{S_1 - S_2 + M'} \\
&\quad [S'] \begin{pmatrix} S & S' & S \\ -M & M' & M \end{pmatrix} \begin{pmatrix} S_1 & S_2 & S' \\ -M_1 & M_2 & -M' \end{pmatrix} \\
&= \frac{1}{2S+1} \sum_M \sum_{S'} \langle S || [A_{kl}^{S_2} \otimes B_{ij}^{S_1}]^{S'} | S \rangle \begin{pmatrix} S & S & 0 \\ M & -M & 0 \end{pmatrix} [S] \\
&\quad (-1)^{S_1 - M_2} (-1)^{S_1 - S_2} [S'] \begin{pmatrix} S & S' & S \\ -M & 0 & M \end{pmatrix} \begin{pmatrix} S_1 & S_2 & S' \\ -M_2 & M_2 & 0 \end{pmatrix}
\end{aligned} \tag{2.36}$$

where the factor $(-1)^{S-M}$ in the last line was replaced by the equivalent 3j-symbol, and the formulae were simplified by realizing that only $M_1 = M_2$ and $M' = 0$ can make a non-zero contribution. The notation $[S'] \equiv \sqrt{2S'+1}$. This expression can be further simplified by adding on terms that are zero by symmetry but allow application of the orthogonality property of the 3j-symbols.

$$\begin{aligned}(2.36) = \frac{1}{2S+1}\sum_{M}\sum_{S'}\sum_{M''}&\langle S||[A^{S_2}_{kl}\otimes B^{S_1}_{ij}]^{S'}||S\rangle \begin{pmatrix} S & S & 0 \\ M'' & -M & 0\end{pmatrix}[S] \\ &(-1)^{S_1-M_2}(-1)^{S_1-S_2}[S']\begin{pmatrix} S & S & S' \\ M'' & -M & 0\end{pmatrix}\begin{pmatrix} S_1 & S_2 & S' \\ -M_2 & M_2 & 0\end{pmatrix} \\ = \frac{\delta_{S_1S_2}}{\sqrt{2S+1}}&\langle S||\Big[A^{S_2}_{kl}\otimes B^{S_2}_{ij}\Big]^0|S\rangle\end{aligned}$$

Since $S' = 0$ the elements can be non-zero only if $S_1 = S_2$. Moreover, this expression is independent in the spin projections M_1, M_2 of the tp operators. So the 2DM is diagonal in the spin of the tp state as well as in its spin projection and its blocks corresponding to different spin projections are equal, leading to the structure

$$\langle\hat{S}_z\rangle = 0 \text{ and } (2.35): \quad \begin{pmatrix} \Gamma^{00\ 0} & 0 \\ 0 & \Gamma^{11\ 0} = \Gamma^{11\ 1} = \Gamma^{11\ -1}\end{pmatrix} \tag{2.37}$$

2.4.2 Basic S-representability constraints

Unless a specific composition of the ensemble is assumed, there are few obvious constraints that derive from its corresponding wavefunction. Imposing the total spin of the ensemble is limited to specifying its correct expectation value. Similarly, the $\hat{S}_z$ expectation value of the ensemble can be fixed.

$$\begin{aligned}\langle\hat{S}^2\rangle &= S(S+1) \\ \langle\hat{S}_z\rangle &= \sum_n w_n \sum_M (c^n_M)^2 M\end{aligned}$$

If the $\hat{S}_z$ expectation value is not specified, the symmetry present in the spinless Hamiltonian will lead to a zero expectation value.

2.4.3 S-representability constraints derived from the Gutzwiller projection

A constraint that might be of interest in non-singlet states, is one which acts directly on the space of singly occupied orbitals. Such constraints have proven useful applied to the Hubbard model.[81] Because v2DM(PQG) results for Hubbard models with half-filling are fairly accurate, but those for models with one particle below half-filling are poor, it seems that singly occupied levels are not properly described by the P-,Q- and G-condition.

In the same way that the first order particle and hole density matrix need to be positive-semidefinite, the first-order particle and hole density matrix expressed in the basis of singly occupied space need to be semidefinite as well. The creation and annihilation operators acting only on the space of singly occupied orbitals are

$$g_i = a_i(1 - a_{\bar{i}}^\dagger a_{\bar{i}})$$
$$g_i^\dagger = (1 - a_{\bar{i}}^\dagger a_{\bar{i}})a_i^\dagger$$

The matrices γ^G and q^G are thus positive semi-definite

$$\gamma_{ij}^G = \langle\Psi|g_j^\dagger g_i|\Psi\rangle \qquad \gamma^G \succeq 0$$
$$q_{ij}^G = \langle\Psi|g_j g_i^\dagger|\Psi\rangle \qquad q^G \succeq 0$$

Although both matrices involve terms acting on three-particle/hole space,

$$\gamma_{ij}^G = \gamma_{ij} - \Gamma_{i\bar{j}i\bar{i}} - \Gamma_{i\bar{i}j\bar{i}} + \langle\Psi|a_{\bar{j}}^\dagger a_j^\dagger a_{\bar{j}} a_{\bar{i}}^\dagger a_i a_{\bar{i}}|\Psi\rangle$$
$$q_{ij}^G = \delta_{ij}(1 - 2\gamma_{\bar{i}\bar{i}}) - \gamma_{ij} + \Gamma_{i\bar{j}i\bar{i}} + \Gamma_{i\bar{i}j\bar{i}} + \langle\Psi|a_{\bar{j}}^\dagger a_j a_{\bar{j}} a_{\bar{i}}^\dagger a_i^\dagger a_{\bar{i}}|\Psi\rangle$$

the term $\langle\Psi|\hat{A}_j^\dagger\hat{A}_i|\Psi\rangle$, with $\hat{A}$ a three particle/hole operator, can be eliminated by adding $\langle\Psi|\hat{A}_j\hat{A}_i^\dagger|\Psi\rangle = \langle\Psi|\hat{A}_i\hat{A}_j^\dagger|\Psi\rangle$ to it,

$$\gamma_{ij}^G + \langle\Psi|a_{\bar{j}}^\dagger a_j^\dagger a_{\bar{j}} a_{\bar{i}}^\dagger a_i a_{\bar{i}}|\Psi\rangle$$
$$q_{ij}^G + \langle\Psi|a_{\bar{j}}^\dagger a_j a_{\bar{j}} a_{\bar{i}}^\dagger a_i^\dagger a_{\bar{i}}|\Psi\rangle$$

which preserves positive semidefiniteness and eliminates the dependence on the 3-particle/hole space.

$$\langle\Psi|a^\dagger_{\bar{j}}a^\dagger_j a_{\bar{j}}a^\dagger_{\bar{i}}a_i a_{\bar{i}} + a^\dagger_{\bar{i}}a_i a_{\bar{i}}a^\dagger_{\bar{j}}a^\dagger_j a_{\bar{j}}|\Psi\rangle = \delta_{ij}\gamma_{\bar{i}\bar{i}}$$
$$\langle\Psi|a^\dagger_{\bar{j}}a_j a_{\bar{j}}a^\dagger_{\bar{i}}a^\dagger_i a_{\bar{i}} + a^\dagger_{\bar{i}}a^\dagger_i a_{\bar{i}}a^\dagger_{\bar{j}}a_j a_{\bar{j}}|\Psi\rangle = \delta_{ij}\gamma_{\bar{i}\bar{i}}$$

So the final expressions for the elements of the matrices that must be constrained to be positive semidefinite, are

$$\gamma^G_{ij} = \gamma_{ij} - \Gamma_{i\bar{j}i\bar{i}} - \Gamma_{i\bar{i}j\bar{i}} + \delta_{ij}\gamma_{\bar{i}\bar{i}} \tag{2.38}$$
$$q^G_{ij} = \delta_{ij}(1-\gamma_{\bar{i}\bar{i}}) - \gamma_{ij} + \Gamma_{i\bar{j}i\bar{i}} + \Gamma_{i\bar{i}j\bar{i}} \tag{2.39}$$

Because these constraints specifically target singly occupied orbitals, they might prove useful in open shell, non-singlet systems. Unfortunately, it seems that they are not violated by typical v2DM(PQG) calculations on molecular systems, even when no spin constraints are imposed. The lowest eigenvalue of the γ^G matrix is typically of the order $10^{-3} - 10^{-5}$ for chemical systems, whereas the q^G matrix is much less close to singularity, and becomes even more positive in the dissociation limit in typical diatomic molecular systems under P-,Q- and G-condition.

2.5 Applications

In order to examine the strength of the spin conditions discussed in sections 2.3 and 2.4, we have applied them to the PES for the carbon and oxygen dimer for several spin states and their different spin projections. The singlet, triplet and quintuplet states must become degenerate in the dissociation limit. Because these systems are homonuclear, they do not suffer from unphysical dissociation under 2-positivity conditions (section 1.5.5 chapter 1). These systems therefore make a good test case for examining several issues regarding spin:

Overall, are the proposed spin constraints capable of reproducing the characteristics of the PES for different spin states and spin projections?

How do 'pure spin state' constraints compare to 'ensemble spin state' constraints?

How do different spin projections relate to each other?

How do spin constraints in the dissociation limit compare to those at short bond lengths? Are they equally strong at all bond lengths?

In order to make this assessment, the following S-representability constraints were taken into consideration.

2.5.1 Applied S-representability conditions

Pure spin state constraints

To describe a pure spin state

I the 'contraction' condition

$$\forall k,l: \quad \sum_i N\ G^{10\ 0}_{iikl} - 2M\ G^{00\ 0}_{iikl} = 0 \tag{2.40}$$

$$\sum_i N\ G^{11\ 0}_{iikl} - 2M\ G^{10\ 0}_{iikl} = 0 \tag{2.41}$$

imposes the correct $\hat{S}_z$ expectation value and consistent contraction to the 1DM.

II the maximal spin projection for a spin-S state, $M = S$, must satisfy the condition

$$\forall k,l: \quad \sum_i G^{11\ 1}_{iikl} = 0 \tag{2.42}$$

This condition forces the vector corresponding to the $\hat{S}^+$ operator to lie in the nullspace of the G^1 block of the G-matrix, which implies $\frac{K^2}{4}$ additional conditions on the 2DM. Along with the contraction condition, it imposes the correct $\hat{S}^2$ expectation value.

III a lesser spin projection of a spin-S state, $|M| < S$, must satisfy

$$\forall k,l: \quad (S-M)(S+M+1)\sum_i G^{11\ 1}_{iikl} = (S+M)(S-M+1)\sum_i G^{11\ -1}_{iikl} \tag{2.43}$$

which imposes a correct relation between the first order transition density matrix elements. When combined with the contraction condition, it implies the correct $\hat{S}^2$ expectation value, except for the case of zero spin projection. In that case, the correct $\hat{S}^2$ expectation value must be imposed additionally.

Ensemble spin state constraints

To describe an ensemble spin state of different spin projection of a spin-S state, with a fixed $\hat{S}_z$ expectation value:

I correct $\hat{S}^2$ expectation value is imposed through the generally holding formula

$$tr\ S^2\Gamma = \frac{N}{2}\left(\frac{N}{2}+1\right) - tr\ \Gamma^{00\ 0} = S(S+1) \tag{2.44}$$

II correct $\hat{S}_z$ expectation value is imposed, $tr\ [S_z\Gamma] = \sum_n w_n \sum_M (c_M^n)^2 M$.

Finally, in order to compare molecular dissociation products to the dissociation products calculated separately, we also consider them under correct $\hat{S}_z$ expectation value only, since imposing correct $\hat{S}^2$ expectation value does not guarantee correct $\hat{S}^2$ expectation value for the molecular dissociation products.

2.5.2 Computational and algorithmic details

All calculations are done in the double zeta basis set D95V as specified in Gaussian03.[70] Reference full configuration interaction calculations with core electrons frozen (FCI(FC)) are carried out with Gaussian03. The potential energy surfaces (PES) are composed from single point calculations. To carry out the variational optimization of the 2DM under semidefinite constraints, we have used our implementation of a modified barrier method (see chapter 3 paragraph 3.5).

2.5.3 Results on S-representability calculations

In earlier work on atomic systems,[44] we have reported calculations that assumed an ensemble resulting in a zero $\hat{S}_z$ expectation value for non-singlet spin states, and compared these to calculations for the pure spin state with maximal spin projection. A more detailed comparison of both the pure spin state approach and the ensemble spin state approach over the whole range of possible $\hat{S}_z$ expectation values for several non-singlet atoms (figure 2.1) and for PES of the lowest energy spin states of the oxygen and carbon dimer (figures 2.2, 2.3) reveals some more interesting features.

As a first key observation, no degeneracy for different spin projections of the same spin state is observed. This shortcoming was also noticed by Nakata.[82] In fact, both in the pure spin state approach and in the ensemble spin state approach, the energy is convex with respect to the $\hat{S}_z$ expectation value (figure 2.1). Interestingly, this is exactly the opposite behavior from that observed in DFA.[74] So the v2DM energy is not only convex with respect to the fractional electron number in between two consecutive integer electron numbers, as observed in previous work,[49] but also with respect to fractional $\hat{S}_z$ expectation value in between two consecutive allowed pure state spin projections. The v2DM calculations for the maximal spin projection give the highest energy, both under pure spin state constraints (2.5.1) and ensemble spin state constraints (2.5.1). The pure spin state constraints are especially strong compared to the ensemble spin state constraints near zero spin projection. In fact, as $M \to S$, the v2DM energy under ensemble spin constraints converges practically to that of the pure state maximal spin projection. For near-zero spin projections, however, the ensemble spin state conditions give a much lower energy. Even more so, the $\langle \hat{S}_z \rangle = 0$ condition does not improve the energy at all. Due to the spin independence of the Hamiltonian, the $\langle \hat{S}_z \rangle = 0$ condition without reinforcement by other spin conditions, is equivalent to not imposing any spin condition.

A second key observation is that the spin constraints are not size-consistent:

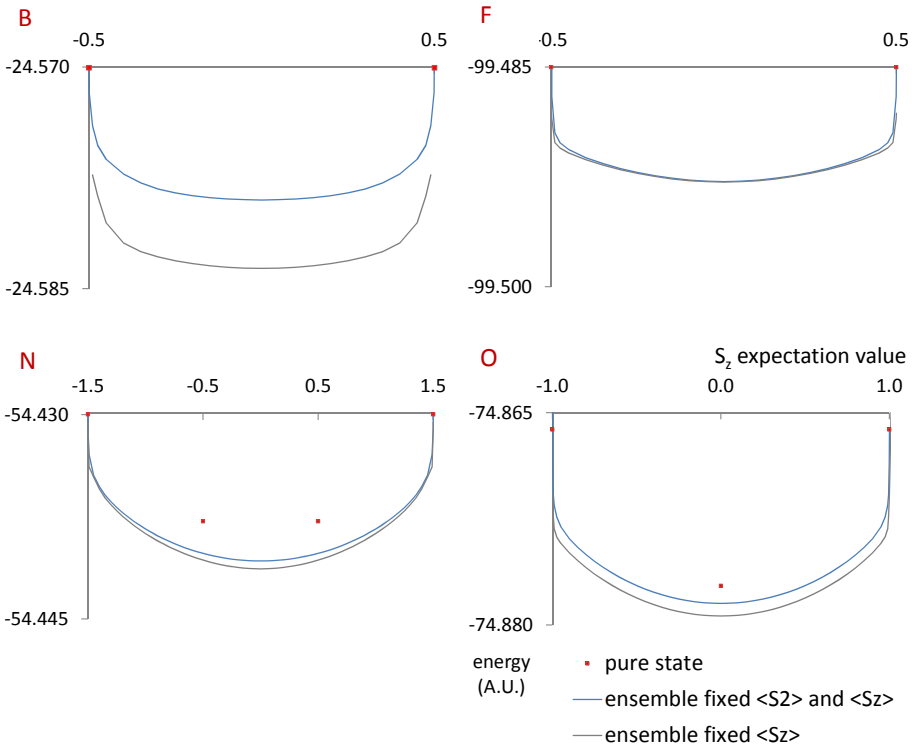

Figure 2.1: The v2DM(PQG) energy is a convex function of the spin projection under both pure spin state and ensemble spin state conditions. The maximal spin projection has the highest energy, even when only the $\hat{S}_z$ expectation value is imposed. The $\langle \hat{S}_z \rangle = 0$ energy without additional spin constraints is equivalent to the energy of a spin unconstrained problem, due to the spin independence of the Hamiltonian.

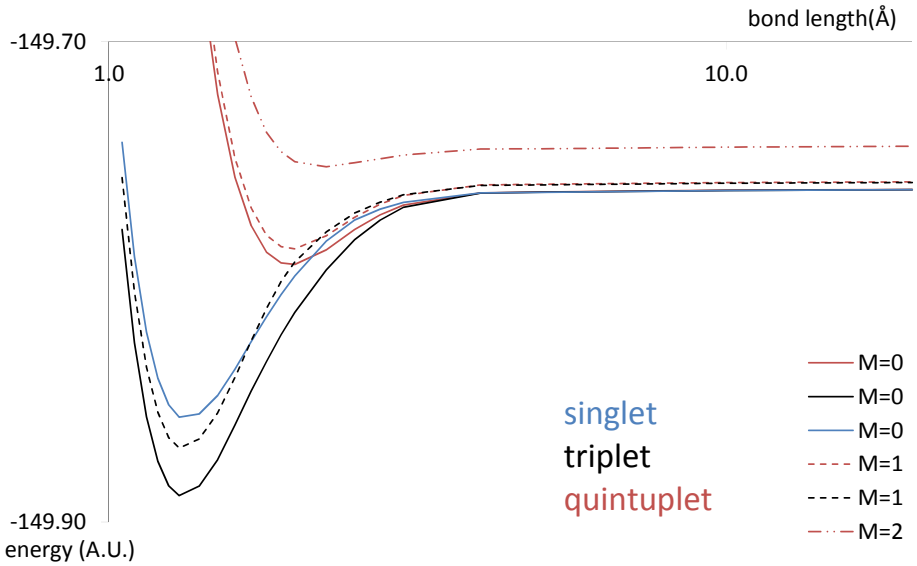

Figure 2.2: The differences between the v2DM PES of the oxygen dimer under pure spin state conditions (2.5.1) for different spin projections M of the same spin state S are remarkable. In the dissociation limit, the difference in the energy for different spin projections seems primarily attributable to the different $\hat{S}_z$ expectation value of the dissociated atoms because their energies are very similar to atomic energies obtained under the same $\hat{S}_z$ expectation value (table 2.1).

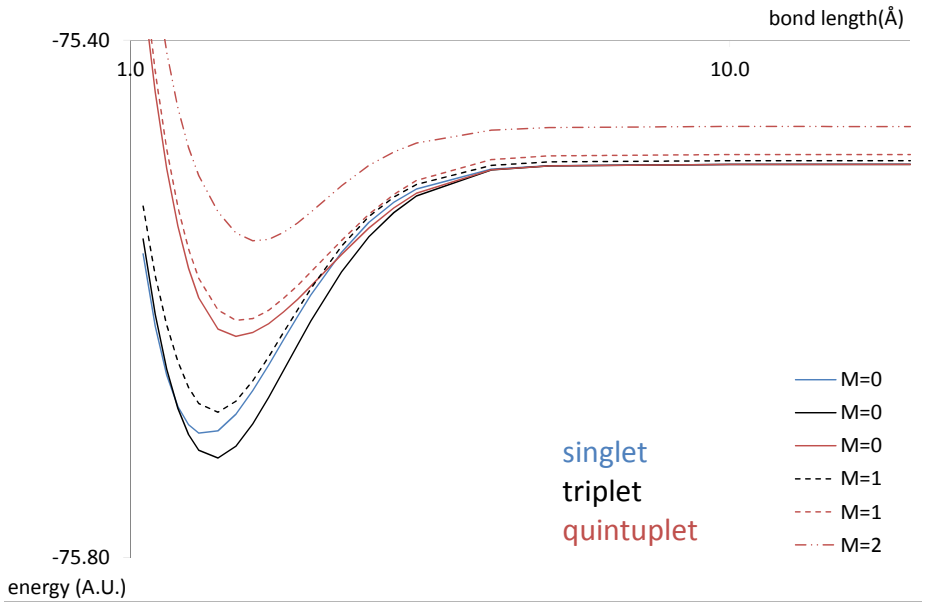

Figure 2.3: Differences between the v2DM PES of the carbon dimer under pure spin state conditions (2.5.1) for different spin projections M of the same spin state S are remarkable. In the dissociation limit, the difference in the energy for different spin projections seems primarily attributable to the different $\hat{S}_z$ expectation value of the dissociated atoms. Only the $S = 2, M = 1$ state gives atomic energies in the dissociation limit that are slightly higher than those obtained under $\langle \hat{S}_z \rangle = 1$.

the constraints on the molecular system do not imply equivalent spin constraints on the dissociation products. The $\hat{S}_z$ expectation values of the dissociation products, being a one-electron property, are fixed to half the homonuclear molecule's $\hat{S}_z$ expectation value, but none of their other spin properties is determined by the spin constraints on the molecule. Of course, even when the molecule is constrained to be a pure spin state, the dissociation products need not be pure spin states, but they need to have a proper $\hat{S}^2$ expectation value. The applied spin constraints, however, lead to dissociated oxygen atoms with $\hat{S}^2$ expectation values around 2.05. Moreover, the effect of imposing spin constraints on the dissociated molecule should be equivalent to imposing them on the dissociation products separately in order to produce size-consistent energies, but this is not true for the applied spin constraints (table 2.2). In fact, the dissociated oxygen atoms have similar energy and $\hat{S}^2$ expectation value under the pure spin state conditions to a calculation constrained only to have the same $\hat{S}_z$ expectation value (figures 2.4, 2.5 and table 2.1). However, when the pure spin state constraints are imposed in separate calculations on the isolated oxygen atoms, they increase the energy significantly compared to a calculation that only imposes $\hat{S}_z$ expectation value (table 2.2).

The absence of degeneracy between different spin projections of the same spin state may have far reaching implications on chemical calculations. Non-interacting states that can couple to different degenerate spin states, such as dissociated molecules, may not be treated on equal footing. Consider for example two triplet oxygen atoms, infinitely far apart. The two states can couple to either a singlet, triplet or quintuplet state. Theoretically, all three spin states, and all of their spin projections, are energetically equivalent. So the singlet, triplet, and quintuplet oxygen dimer should yield the same energy in the dissociation limit. Unfortunately, under the pure state spin constraints only the zero spin projection gives the same dissociation limit for all spin states (figures 2.2 and 2.3) because, for each of the spin states, the zero spin projection leads to dissociated atoms with zero spin projections in a homonuclear molecule. When considering

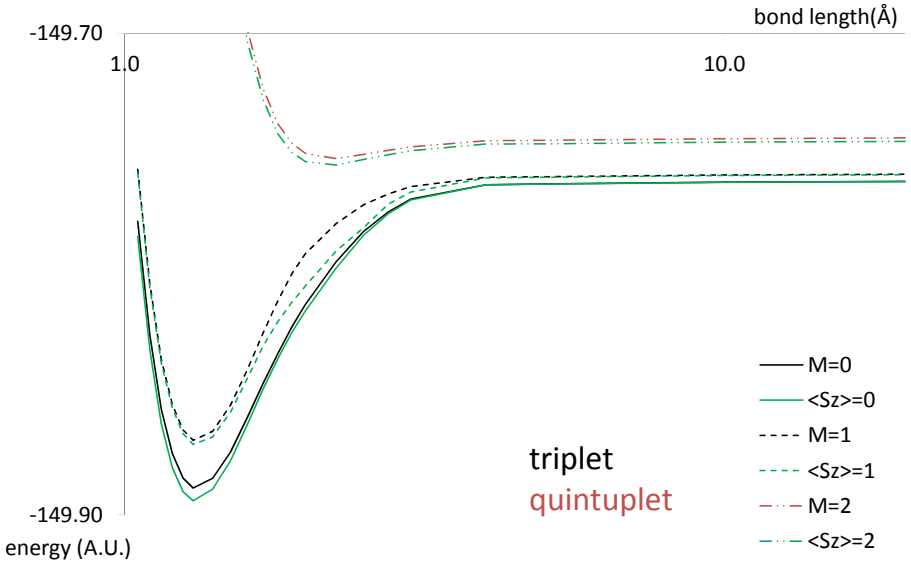

Figure 2.4: Although all of the v2DM(PQG) singlet, triplet and quintuplet PES for the oxygen dimer (black, blue, red) should converge to the same dissociation limit, only the same spin projections converge to a very similar dissociation limit, which practically coincides with the dissociation limit under a constraint on $\hat{S}_z$ expectation value only, $\langle \hat{S}_z \rangle = M$ (gray lines).

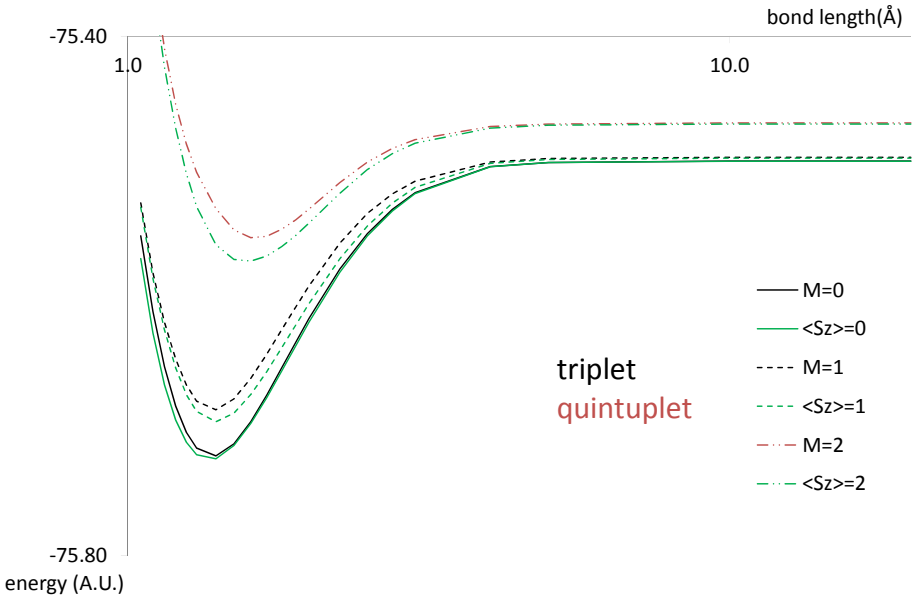

Figure 2.5: Although all of the v2DM(PQG) singlet, triplet and quintuplet PES for the carbon dimer (black, blue, red) should converge to the same dissociation limit, only the same spin projections converge to a very similar dissociation limit, which practically coincides with the dissociation limit under a constraint on $\hat{S}_z$ expectation value only, $\langle \hat{S}_z \rangle = M$ (gray lines).

molecular state	$\|S0\rangle$:			
S	0	1	2	not fixed
E^{atom}	-74.8809	-74.8809	-74.8808	-74.8809
$\langle \hat{S}_z \rangle^{atom}$	0.00	0.00	0.00	0.00
$\langle \hat{S}_z^2 \rangle^{atom}$	0.68	0.45	0.28	0.68
$\langle \hat{S}^- \hat{S}^+ \rangle^{atom}$	1.36	1.59	1.76	1.36
$\langle \hat{S}^2 \rangle^{atom}$	2.04	2.04	2.04	2.04

molecular state	$\|S1\rangle$:		
S	1	2	not fixed
E^{atom}	-74.8794	-74.8793	-74.8795
$\langle \hat{S}_z \rangle^{atom}$	0.50	0.50	0.50
$\langle \hat{S}_z^2 \rangle^{atom}$	0.51	0.48	0.69
$\langle \hat{S}^- \hat{S}^+ \rangle^{atom}$	1.03	1.07	0.85
$\langle \hat{S}^2 \rangle^{atom}$	2.04	2.05	2.04

molecular state	$\|S2\rangle$:	
S	2	not fixed
E^{atom}	-74.8718	-74.8725
$\langle \hat{S}_z \rangle^{atom}$	1.00	1.00
$\langle \hat{S}_z^2 \rangle^{atom}$	1.01	1.01
$\langle \hat{S}^- \hat{S}^+ \rangle^{atom}$	0.05	0.05
$\langle \hat{S}^2 \rangle^{atom}$	2.06	2.06

Table 2.1: The properties of the dissociated atoms in the oxygen dimer in the dissociation limit, denoted by the superscript '*atom*', are remarkably similar for molecular states that lead to dissociated atoms with the same spin projection, both when pure spin state conditions are imposed and when only spin projection is specified (column 'not fixed').

atomic state	$\|S0\rangle$:		atomic state	$\|S1\rangle$:	
S	2	not fixed	S	2	not fixed
E^{atom}	-74.8772	-74.8794	E^{atom}	-74.8662	-74.8706
$\langle\hat{S}_z\rangle^{atom}$	0.00	0.00	$\langle\hat{S}_z\rangle^{atom}$	1.00	1.00
$\langle\hat{S}_z^2\rangle^{atom}$	0.00	0.69	$\langle\hat{S}_z^2\rangle^{atom}$	1.00	1.01
$\langle\hat{S}^-\hat{S}^+\rangle^{atom}$	2.00	1.37	$\langle\hat{S}^-\hat{S}^+\rangle^{atom}$	0.00	1.05
$\langle\hat{S}^2\rangle^{atom}$	2.00	2.06	$\langle\hat{S}^2\rangle^{atom}$	2.00	2.06

Table 2.2: The pure state spin constraints are much stronger when imposed on the triplet atoms separately than when they are imposed on the dissociated singlet, triplet or quintuplet oxygen dimer, even though they should be equivalent (compare with table 2.1).

the maximal spin projections for the singlet, triplet and quintuplet, the different spin projections of the dissociated atoms seem to be the main cause of energy differences in the dissociation limit: the energies of the dissociated molecules under maximal spin projection conditions are very similar to those constrained to the same $\hat{S}_z$ expectation value only (figures 2.4 and 2.5).

None of the spin constraints applied to different spin projections of the lowest-lying spin states of the oxygen and carbon dimer gives a truly satisfying picture of the molecule's properties. The zero spin projection constraints treat all spin states equivalently in the dissociation limit, but fail to reproduce the correct features of the PES for the different spin states. Most remarkably, they produce a triplet PES that is lower than the singlet PES for the carbon dimer, in contradiction with FCI(FC) results (figure 2.7). In case of the oxygen dimer, they make the quintuplet state much too strongly binding (figure 2.6).

The maximal spin projection constraints give the most strongly constrained results, but do not reproduce degeneracy of the different spin states upon dissociation. They do give the lowest equilibrium energy for the singlet state of the carbon dimer (figure 2.9), in agreement with FCI(FC) data, but they severely underestimate the singlet-triplet energy gap, in both the carbon and

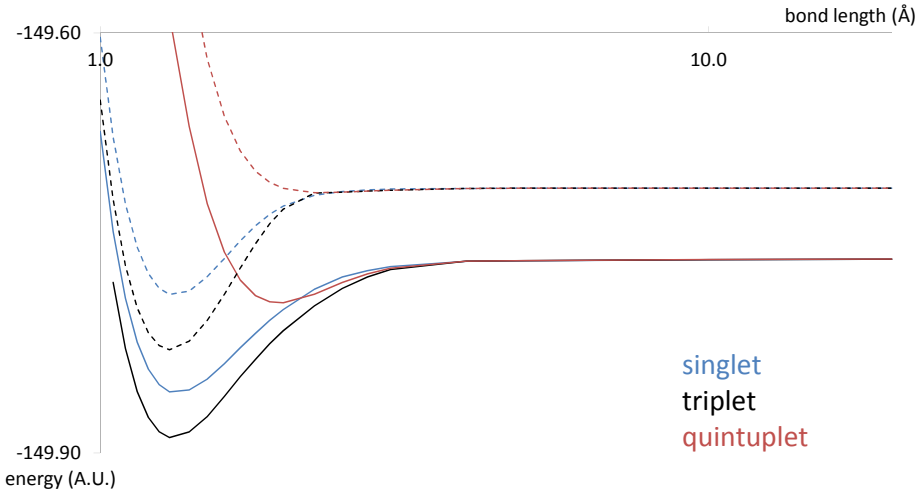

Figure 2.6: The pure spin state conditions for the zero spin projection v2DM(PQG) PES (solid lines) treat all different spin states of the oxygen dimer equivalently in the dissociation limit. Yet they do not give a fully satisfying picture of its properties; the quintuplet state becomes much too strongly binding compared to FCI(FC) calculations (dotted lines).

oxygen dimer (figure 2.8).

The PES of the carbon dimer has also been computed in a 6-31G* basis set under singlet conditions by Gidofalvi and Mazziotti. The conditions they imposed on the singlet 2DM are equivalent to the conditions we use here, except that they did not explicitly impose the equivalence of the three triplet blocks of the 2DM in their early work.[52] The current comparison of the singlet PES with other spin states shows the subtle, but crucial, effect that spin constraints may have. Depending on the spin projection under consideration, the wrong spin state may be obtained as the lowest energy state under approximate spin constraints.

Imposing the conditions that describe a pure state maximal spin projection is theoretically the preferred method for describing spin because these conditions are the most stringent. Computationally, however, these constraints are also the most expensive. The pure state maximal spin projection constraints provide

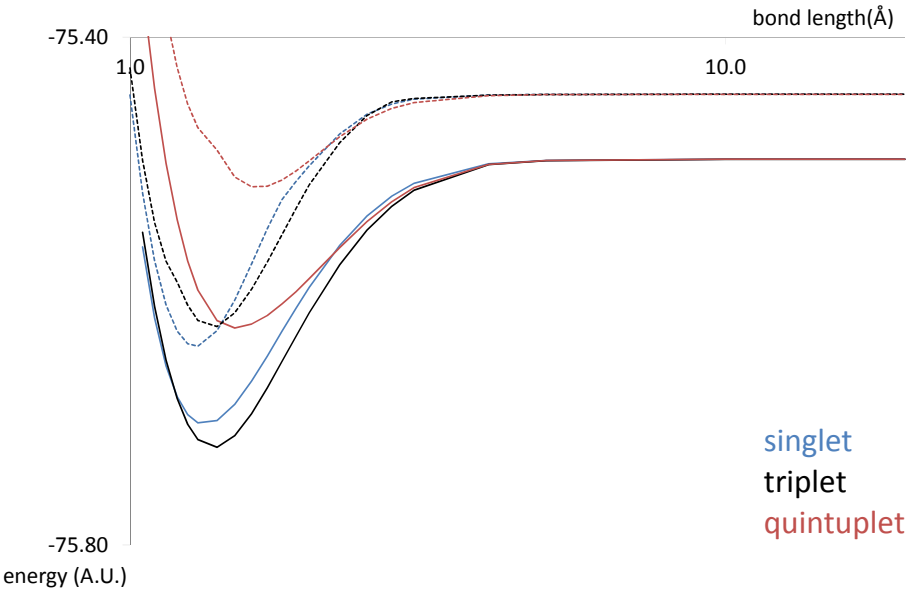

Figure 2.7: The pure spin state conditions for the zero spin projection v2DM(PQG) PES (solid lines) of the carbon dimer singlet yield energies lower than those for the triplet around equilibrium bond length. Yet FCI(FC) calculations (dotted lines) prove that their relative order should be exactly opposite.

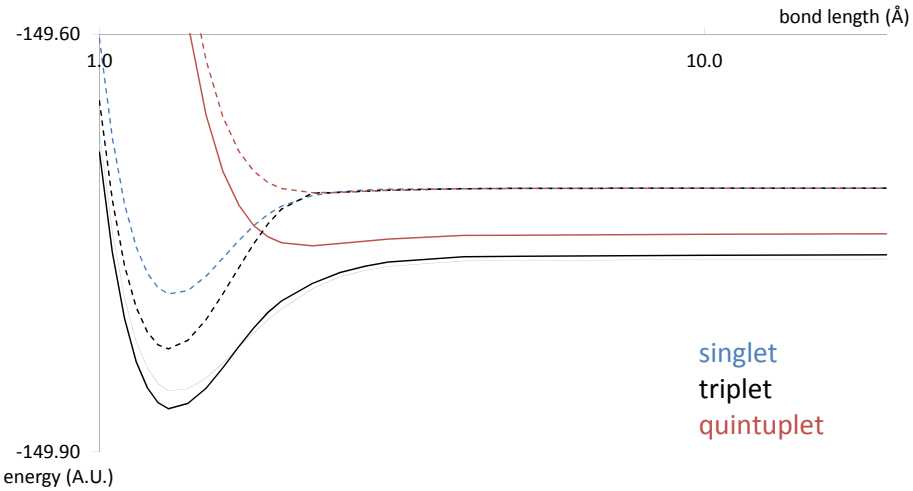

Figure 2.8: The v2DM(PQG) PES (solid lines) for the maximal spin projections of the singlet, triplet and quintuplet of the oxygen dimer under pure spin state constraints (2.5.1) are not consistent: they do not converge to equivalent dissociated states. Moreover, they give a singlet-triplet gap that is much too small compared to FCI(FC) data (dotted lines).

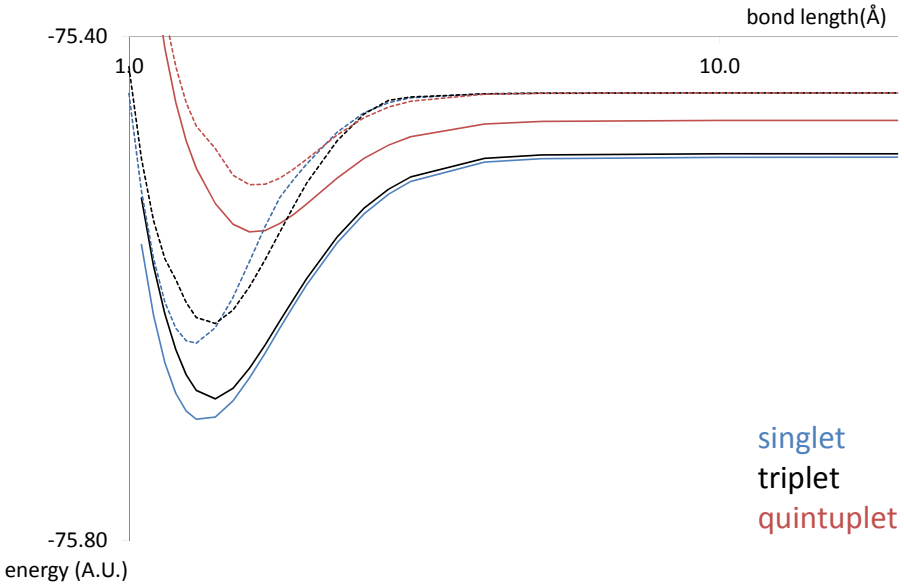

Figure 2.9: The v2DM(PQG) PES (solid lines) for the maximal spin projections of the singlet, triplet and quintuplet of the carbon dimer under pure spin state conditions (2.5.1) are not equivalent in the dissociation limit. Nonetheless, they do give the correct order of singlet and triplet PES, similar to that in FCI(FC) calculations (dotted lines). The zero spin projection PES of the singlet and triplet have exactly the opposite ordering.

the most stringent lower bound on the energy and, moreover, give the correct relative order of the different spin projections for the carbon dimer. Although it does not seem straightforward to derive conditions that directly constrain non-maximal spin projections to the same extent, it is possible to derive a 2DM for a lower spin projection $M < S$ from the maximal spin projection $M = S$ by means of the Wigner-Eckart theorem.[80] By construction, the resulting 2DM for the $M < S$ spin projection will have the exact same energy as the maximal spin projection. This justifies the use of maximal spin projection conditions to get the strictest lower bound on the energy. Additionally, the inconsistencies that arise in the dissociation limit of the maximal spin projections of degenerate spin states, such as those occurring in the dissociation limit of the oxygen and carbon dimer, can be corrected by imposing subspace energy conditions.[50,53,69] At the same time, these constraints will correct size-consistency defects and incorrect dissociation – in contrast to the molecules under consideration here, non-homonuclear molecules generally dissociate into fractionally charged products in practical v2DM methods.[49]

A spin condition can be incorporated indirectly into the subspace constraints by requiring that the energy of the subspace in the molecule must be at least equal to the energy of the lowest-energy spin state of the subspace treated as a separate system. Moreover, because of the lack of degeneracy in multiplets calculated with the v2DM method, the tightest constraint is obtained if the maximal spin projection is considered for the subspace system:

$$tr[H^A\Gamma] \geq E_0(H^A)|_{S=S_0,M=S}$$

with H^A a Hamiltonian matrix for the atomic or molecular subspace A expressed in the molecular basis space, $tr\ [H^A\Gamma]$ the energy of the subspace A in the molecule and $E_0(H^A)|_{S=S_0,M=S}$ the ground state v2DM energy for this atomic or molecular subspace calculated separately in the maximal spin projection of the lowest energy spin state S_0. Because the energy of the reference system A should be degenerate for different spin projections, considering the highest spin

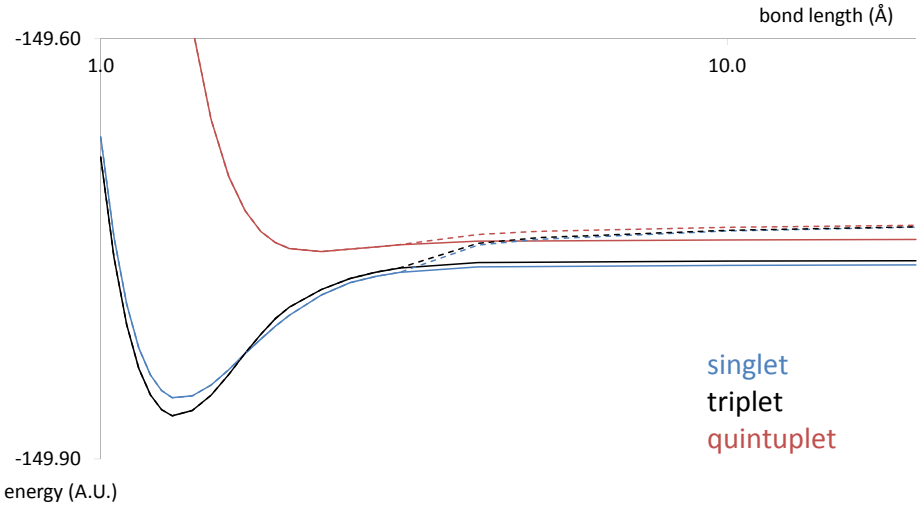

Figure 2.10: When subspace constraints are imposed on the singlet, triplet and quintuplet v2DM(PQG) PES under pure spin state maximal spin projection conditions (2.5.1), both the violation of size-consistency and the absence of degeneracy among dissociated states with different spin projection (solid lines) are corrected in the resulting PES (dotted lines). The shapes of the different spin surfaces remain poor, however.

projection is fully justified. This constraint can be considered an extension of the 'flat plane condition' developed by Yang et al. in DFA[74,75] to v2DM theory. Although this constraint ensures that the energy of maximal spin projections of degenerate spin states effectively becomes degenerate in the dissociation limit, it does not improve the poor relative position of the different spin states around equilibrium bond length, because the subspace constraints only become active upon dissociation (figure 2.10).[53,69]

2.6 Conclusions on describing spin in v2DM theory

Two main shortcomings concerning spin constraints in v2DM theory explain the incorrect features of the PES of the carbon and oxygen dimer for different spin states. First of all, spin constraints on a system composed of non-interacting atoms or molecules do not imply equally strong constraints on the non-interacting atoms or molecules separately. They are therefore a source of size-inconsistency. This shortcoming is inherent to the method, and will be difficult to correct except through a 'quick fix', like the subspace energy constraints introduced in previous work and applied to the PES of the oxygen dimer in figure 2.10.

Secondly, the v2DM energy is a convex function of the $\hat{S}_z$ expectation value, with the highest energy for the maximal spin projection. The pure spin state conditions for the maximal spin projection are therefore the most stringent conditions on the 2DM that one can formulate directly in terms of the spin operators. An equivalent 2DM for a lower spin projection is derivable from the maximal spin projection by application of the Wigner-Eckart theorem. As a consequence of the lack of degeneracy between different spin projections of a multiplet, maximal spin projections of higher spin states are more severely constrained than those of lower spin states, which becomes especially apparent in the dissociation limit, where theoretically degenerate spin states fail to become degenerate in the v2DM method. These differences, together with size-consistency defects and incorrect dissociation, can also be fixed by means of subspace constraints.

Even though the pure spin state maximal spin projection conditions are the strongest, they are also significantly more expensive than the ensemble spin state conditions, because describing them requires about twice as much variables. The ensemble approach allows to consider a spin-averaged ensemble with resulting $\langle \hat{S}_z \rangle = 0$, which has similar spin symmetry to the pure state singlet (2.6). This

reduces the number of variables by about a factor 2. Given the typical scaling of $O(K^6)$ of semidefinite programs with basis set dimension K, this makes the ensemble approach considerably cheaper. Future work may therefore focus on ways of improving on an ensemble approach. However, the inherent lack of information on the composition of such an ensemble makes it much more difficult to find stringent constraints that apply to it.

For a given maximum problem with maximum M, we shall often be able to find an equivalent minimum problem with the same value M as minimum; this is a useful tool for bounding M from above and below.

Courant and Hilbert in *Methods of Mathematical Physics I* (1953)

3

Semidefinite optimization of the 2DM

3.1 Introduction

As well as the theoretical challenges that come with the N-representability problem in practical applications, variational second order density matrix methods also pose a major computational challenge. Since the key N-representability conditions impose positive semidefinite constraints on the 2DM, v2DM methods are semidefinite optimization problems. Originating as an extension of interior point methods for linear programming,[12,13] semidefinite optimization methods have thrived from the 90's on, stimulating applications in engineering, economics, ..., and chemistry. The surge of powerful semidefinite programming algorithms renewed interest in v2DM theory.[29,34]

Typical applications for v2DM methods, however, involve many more variables than standard problems in mathematics. Because the dimension of the 2DM scales as $O(K^2)$, with K the dimension of the sp spin basis, the number of variables involved for typical basis set dimensions (K =100-300) is huge. Although the performance of current semidefinite programs has much improved compared

to early applications from the 70's using cutting plane based methods[7,9] similar to the simplex method for linear programming, applications of the v2DM method are still limited to few-atom systems in modest basis sets.[14,15,83] Therefore, the challenge is to find the semidefinite optimization method that works best for the particular problem of variational optimization of second order density matrices, balancing speed with accuracy. For this reason, we have implemented several optimization algorithms, adjusted to suit atomic and molecular structure calculations, and have made a comparative assessment of these methods.

Section 3.2 gives a concise background to semidefinite optimization. Sections 3.4 to 3.7 introduce some semidefinite programming techniques: the classical barrier method and a modified barrier method, a primal-dual framework and a boundary point method. They give a brief explanation of the working of the method, followed by a report on their application to molecular calculations. Of course, this is by no means an exhaustive coverage of semidefinite programming techniques. Other approaches include the first-order non-linear approach[84] taken by Mazziotti,[15] which replaces the semidefinite constraint matrices by their square root or Cholesky factorization, such that they are positive-semidefinite by construction, but sacrifices linearity; the primal barrier approach taken by Cances et al.;[85] bundle methods,[86,87] and other variants of augmented Lagrangian based algorithms.[88]

3.2 Basics

The dual formulation of the v2DM optimization problem, where the dual variable is the 2DM Γ, is

$$\begin{aligned} \underbrace{min}_{\Gamma} \quad & tr\ H\Gamma \\ subject\ to \quad & L(\Gamma) \succeq 0 \end{aligned} \tag{3.1}$$

The 2DM can be expressed in an n-dimensional basis of traceless matrices $\{F^i\}$ to incorporate the correct normalization by construction

$$\Gamma = F^0 + \sum_{i=1}^{n} \Gamma_i F^i \tag{3.2}$$

Its dependence on the antisymmetrical identity matrix I is then fixed in the basis matrix F^0

$$F^0 = \frac{N(N-1)}{K(K-1)} I$$
$$I_{abcd} = \delta_{ac}\delta_{bd} - \delta_{ad}\delta_{bc}$$

because $tr\ F^i = 0$ for $\quad i = 1, \ldots, n$. N is the number of electrons and K the dimension of the one particle spin basis set. We will assume the basis matrices F^i form an orthonormal set, orthogonal to F^0, and use a projector $\mathcal{P}_{\perp F^0}()$ to denote the part of the matrix in the traceless space spanned by the basis $\{F^i\}$

$$\mathcal{P}_{\perp F^0}(\Gamma) = \sum_i \Gamma_i F^i$$

The positive semidefinite constraints on Γ (chapter 1, section 1.3.3) are generically denoted $L(\Gamma) = L(F^0) + \sum_i \Gamma_i L(F^i) \succeq 0$. L is a homogeneous linear matrix map, which may map the 2DM onto a matrix with a different symmetry or even different dimension. Its adjoint under the trace operation, $L^\dagger$, is defined through

$$tr\ L(X)Y \equiv tr\ XL^\dagger(Y) \tag{3.3}$$

Hence the adjoint of the Q-map is the Q-map itself, $Q^\dagger() = Q()$ and the G-map's adjoint requires an additional antisymmetrizer, $G^\dagger() = \mathcal{A}\{G()\}$. The antisymmetrizer $\mathcal{A}$ is a projector onto the space of antisymmetrical matrices.
In case of the v2DM(PQG) method, the semidefinite constraint matrix $L(\Gamma)$ is the direct sum of the P-, Q- and G-matrix

$$L(\Gamma) = \Gamma \oplus Q(\Gamma) \oplus G(\Gamma) = F^0 \oplus Q(F^0) \oplus G(F^0) + \sum_i \Gamma_i \ \left(F^i \oplus Q(F^i) \oplus G(F^i)\right)$$

The Hermitian adjoint map $L^\dagger$ that maps a block diagonal matrix $X \equiv X^P \oplus X^Q \oplus X^G$, where the blocks $X^P,\ X^Q,\ X^G$ have the same dimension and

symmetry as the P-, Q- and G-matrix, onto an antisymmetrical matrix of the same dimension as the 2DM and satisfies $tr\ L(\Gamma)X = tr\ L^\dagger(X)\Gamma$, is

$$L^\dagger(X) = X^P + Q(X^Q) + G^\dagger(X^G)$$

Every *dual* optimization problem has an associated *primal* problem, which follows from considering the Lagrangian of the problem. The Lagrangian for the dual problem (3.1) is

$$\mathcal{L}(\Gamma, X) = tr\ H\Gamma - tr\ XL(\Gamma) \tag{3.4}$$

where the positive semidefinite Lagrange multiplier X enforces positive semidefiniteness of $L(\Gamma)$, since $XL(\Gamma) \geq 0$ if $L(\Gamma) \succeq 0$. Minimizing the Lagrangian over Γ will provide a lower bound on the optimal value, $tr\ H\Gamma^*$, of the objective function.

$$\begin{aligned}
&\nabla_\Gamma \mathcal{L}(\Gamma, X^*)_i = tr\ HF^i - tr\ X^*L(F^i) = 0 \\
&\mathcal{P}_{\perp F^0}(H) = \mathcal{P}_{\perp F^0}(L^\dagger(X)) \\
&\underbrace{min}_{\Gamma} \quad \mathcal{L}(\Gamma, X) = trHF^0 - tr\ X^*L(F^0) \quad \leq tr\ H\Gamma^*
\end{aligned}$$

Since this gives a lower bound on the optimal value of the objective function, maximizing this expression over the primal variable X will give the tightest lower bound on the optimal value of the objective function. The resulting *primal* optimization problem is a different way of approaching the original, dual, optimization problem.

$$\begin{aligned}
&\underbrace{max}_{X}\ \underbrace{min}_{\Gamma} \quad \mathcal{L}(\Gamma, X) \\
&= \underbrace{max}_{X} \quad tr\ HF^0 - tr\ XL(F^0) \\
\textit{subject to} \quad &X \succeq 0 \\
&tr\ HF^i = tr\ XL(F^i)
\end{aligned} \tag{3.5}$$

The problem of minimizing the trace of $L(X)$ under positive semidefiniteness of X and the equality constraint $tr\ XL(F^i) = tr\ HF^i \quad i = 1, \ldots n$ is the *primal*

to the *dual* problem statement (3.1) that follows directly from the physical formulation of the v2DM problem. It is simply a different perspective on the same problem. Intuitively, both perspectives should be equivalent. This is true for the v2DM problem, but it is not always true in general, since one or both formulations may not have a bounded solution.

As can be expected, the primal problem will give an optimal value which is less than or equal to that of the dual problem, since for any subset $S, S' \in \mathbf{R}^n$

$$\underbrace{sup}_{X \in S'} \; \underbrace{inf}_{\Gamma \in S} \mathcal{L}(\Gamma, X) \;\leq\; \underbrace{inf}_{\Gamma \in S} \; \underbrace{sup}_{X \in S'} \mathcal{L}(\Gamma, X) \tag{3.6}$$

This is the *weak duality* property, which can also be verified by noting that the difference between the dual and the primal objective function is

$$tr\ H\Gamma - tr\ HF^0 + tr\ XL(F^0) = \sum_i tr\ [XL(F^i)]\Gamma_i + tr\ [XL(F^0)] \tag{3.7}$$

$$= tr\ XL(\Gamma) \geq 0 \tag{3.8}$$

The inequality in the last line follows from positive semidefiniteness of both $L(\Gamma)$ and X. Weak duality thus states that the *duality gap* $tr\ XL(\Gamma)$ between the primal and the dual problem formulation must be positive. Hence, optimality is obtained when the primal and dual problem give the same optimal value and the duality gap vanishes

$$tr\ X^* L(\Gamma^*) = 0 \qquad X^*,\ \Gamma^* \text{ optimal}$$

In contrast to linear programming, the existence of feasible dual and primal points does not guarantee a zero duality gap at the optimum. Some problems may have a solvable dual formulation, but an infeasible primal formulation, or the other way around. Strict feasibility of either the dual or the primal problem, however, guarantees the existence of an optimal solution for both and a zero duality gap at the optimum.[86,89] In practical applications, the duality gap plays an important role, as it gives an upper bound on the deviation from optimality of the current primal and dual variable.

Under the assumption of strict feasibility of either the dual or the primal problem, which holds for the v2DM minimization problem, the necessary and sufficient conditions for the linear semidefinite programming problem (3.1) are thus

$$tr\ XL(F^i) = tr\ HF^i \quad i = 1, \ldots, n \tag{3.9}$$

$$L(\Gamma) \succeq 0 \tag{3.10}$$

$$X \succeq 0 \tag{3.11}$$

$$XL(\Gamma) = 0 \tag{3.12}$$

which are equivalent to the Karush-Kuhn-Tucker (KKT)-conditions in linear programming.[26] The condition of zero duality gap, $tr\ XL(\Gamma) = 0$, implies that $XL(\Gamma) = 0$ since both matrices are constrained to be positive semidefinite. Alternatively, this implies that X and $L(\Gamma)$ can be diagonalized simultaneously such that their corresponding eigenvalues satisfy the complementary slackness condition

$$\lambda_i(L(\Gamma))\lambda_i(X) = 0 \quad \forall i = 1, \ldots dim(X)$$

The complementary slackness condition implies that a non-zero eigenvalue in the primal variable X must correspond to a zero eigenvalue in the dual constraint matrix $L(\Gamma)$ and, conversely, that a nonzero eigenvalue of $L(\Gamma)$ must correspond to a zero eigenvalue of X.

The primal formulation thus gives a different, but equivalent, perspective on the v2DM problem. It attempts to find the positive semidefinite X that describes the traceless Hamiltonian by $L^\dagger(X)$, $\mathcal{P}_{\perp F^0}(H) = \mathcal{P}_{\perp F^0}(L^\dagger(X))$, and that minimizes the trace of $L^\dagger(X)$ while remaining positive semidefinite. Because X is block diagonal, with each block corresponding to a block in the constraint matrix $L(\Gamma)$, the primal problem to the dual formulation of the v2DM(PQG)

problem with $L(\Gamma) = \Gamma \oplus Q(\Gamma) \oplus G(\Gamma)$ and $X = X^P \oplus X^Q \oplus X^G$ is

$$\begin{aligned} \underbrace{max}_{X^P, X^Q, X^G} \quad & (tr\ H - tr\ X^P - tr\ Q(X)^Q - tr\ G^\dagger(X^G)) \frac{N(N-1)}{K(K-1)} \\ subject\ to \quad & X^P,\ X^Q,\ X^G \succeq 0 \\ & \mathcal{P}_{\perp F^0}(X^P + Q(X^Q) + G^\dagger(X^G)) = \mathcal{P}_{\perp F^0}(H) \end{aligned}$$

To understand the relation between the primal and dual formulations of the v2DM problem better, it is instructive to consider the P-condition only. The primal formulation to the v2DM(P) problem has a simple interpretation.

$$\begin{aligned} \underbrace{max}_{X} \quad & (tr\ H - tr\ X) \frac{N(N-1)}{K(K-1)} \\ subject\ to \quad & X \succeq 0 \\ & \mathcal{P}_{\perp F^0}(X) = \mathcal{P}_{\perp F^0}(H) \end{aligned} \tag{3.13}$$

Only the trace of X is not fixed by (3.13), but since it must be minimal without violating its positive-semidefiniteness, X must be

$$X = H - \lambda_{min}(H) I$$

which gives a primal optimal value of $-\lambda_{min}(H) N(N-1)$, equal to the dual optimal function value. Therefore, under the P-condition only, the energy simply equals the energy of an $N(N-1)/2$-fold electron pair occupation of the lowest eigenstate of the second-order reduced Hamiltonian. The 2DM follows from the complementary slackness condition $\Gamma X = 0$ and corresponds to this interpretation:

$$\Gamma = N(N-1)\ v_1 v_1^T$$

where v_1 is the eigenvector of the Hamiltonian corresponding to its lowest eigenvalue. The optimal 2DM is thus a rank-1 matrix for an $N(N-1)/2$-fold electron pair occupation of the lowest-energy eigenvector of H. Clearly, this is a very poor approximation to most chemical systems and explains why additional N-representability constraints are needed to obtain sensible results (see chapter 1).

Most interior point methods are based on the *centrality conditions*, which consider a perturbation of the complementary slackness condition, $XL(\Gamma) = tI$, with $t \geq 0$:

$$\begin{aligned} & tr\ XL(F^i) = tr\ HF^i \quad i = 1, \ldots, n \\ & L(\Gamma) \succeq 0 \\ & X \succeq 0 \\ & XL(\Gamma) = tI \end{aligned} \tag{3.14}$$

The optimum of the perturbed optimality conditions defines a *central path*, parametrized by t. The central path converges to the optimum of the original problem as $t \to 0$, since these equations approach the KKT-conditions when $t \to 0$. This path forms a guideline to the optimum for most interior point methods, which converge to the optimum from within the feasible set.

To summarize, dual methods attempt to solve the dual problem

$$\begin{aligned} & \underbrace{min}_{\Gamma} \quad tr\ H\Gamma \\ & subject\ to \quad L(\Gamma) \succeq 0. \end{aligned} \tag{3.15}$$

Primal methods attempt to solve the primal problem

$$\begin{aligned} & \underbrace{max}_{X} \quad tr\ HF^0 - tr\ XL(F^0) \\ & subject\ to \quad X \succeq 0 \\ & \qquad\qquad tr\ HF^i = tr\ XL(F^i) \quad i = 1, \ldots, n \end{aligned} \tag{3.16}$$

Primal-dual methods attempt to solve the primal and the dual problem simulta-

neously, while minimizing the primal-dual duality gap.

$$\begin{aligned} \underbrace{min}_{\Gamma,X} \quad & \sum_i H_i\Gamma_i + trL(X)F^0 = tr\ [XL(\Gamma)] \\ subject\ to \quad & L(\Gamma) \succeq 0 \\ & X \succeq 0 \\ & tr\ HF^i = tr\ XL(F^i) \quad i = 1, \ldots n \end{aligned} \tag{3.17}$$

The primal and dual problem can be cast in an equivalent formulation by introducing an additional variable Z in the dual problem statement which has the same dimension as the primal variable X and must satisfy $Z = L(\Gamma)$. The duality gap then takes the form $tr\ XZ = tr\ XL(\Gamma)$.

Sections 3.4 to 3.7 discuss several semidefinite optimization algorithms and their application to the v2DM method. Some aspects common to all implementations of the v2DM method are discussed in the next section.

3.3 Computational aspects

The beauty of the v2DM method lies in its complete independence of any other method. The only information it requires is the specification of the system, given by the Hamiltonian matrix projected onto a finite-dimensional basis. The choice of basis and the expression of the Hamiltonian in this basis are discussed in the next paragraph.

Interior-point algorithms also require a feasible 'interior' starting point, a problem which is briefly addressed in the subsequent paragraph.

3.3.1 Input and data storage

Storing and handling symmetry in the 2DM

In order to reduce computation and memory requirements, only unique elements of the 2DM are stored and manipulated. The 2DM must obey several different symmetries:

antisymmetry of electrons: only unique antisymmetrical tp-states $|ij\rangle = \frac{1}{\sqrt{2}}(|ij) - |ji))$ with $i < j$ need be taken into account, such that the dimension of the 2-dimensional 2DM is $\frac{K}{2}(K-1)$ with K the dimension of the (spin) sp-basis

Hermiticity of the 2DM: only the upper diagonal part of the 2DM needs to be referenced, which amounts to $\frac{1}{8}K(K-1)(K(K-1)+2)$ antisymmetrical 2DM elements

spin symmetry: spin considerations impose a further block diagonalization of the 2DM, depending on the spin state under consideration (see chapter 2, sections 2.3 and 2.4)

spatial symmetry: spatial symmetry imposes yet another block diagonalization of the 2DM, because the MO's ψ_i in which the 2DM is expressed belong to an irreducible representation χ_i of the point group to which the molecule belongs. The symmetry of the tp creation/annihilation operators corresponds to the direct product of the irreducible representations of the sp states ψ_i, ψ_j involved, $\chi_i \otimes \chi_j$. Hence, they can only give a nonzero result acting on the wavefunction when coupled with another particle/hole operator with symmetry $\chi_k \otimes \chi_l$ such that their direct product $(\chi_k \otimes \chi_l) \otimes (\chi_j \otimes \chi_i)$ contains the fully symmetrical representation.

angular momentum symmetry: in the same way that spin adaptation leads to additional block diagonalization of the 2DM, angular momentum adaptation will impose an additional block structure. However, we have not implemented this symmetry for molecular systems.

Exploiting these symmetries, the dimension of the 2DM is at most $\frac{K}{2}(K-1)$ and the total number of elements is at most $\frac{1}{8}K(K-1)(K(K-1)+2)$. Depending on spin and spatial symmetries, the dimension may be further reduced. Nonetheless, the formally quadratic scaling of its dimension with the basis set size makes the cost of semidefinite programming algorithms grow very quickly with the basis set

dimension and emphasizes the need for fast semidefinite optimization algorithms with a favorable scaling.

Composing the Hamiltonian

In all applications, the electronic Hamiltonian, within the Born-Oppenheimer approximation, is considered in its second-order reduced form (see section 1.2 of chapter 1) form

$$\hat{H} = \frac{1}{N-1}\left(\frac{-1}{2}(\nabla_i^2 + \nabla_2^2) - \sum_n \left(\frac{Z_n}{|r_1 - R_n|} + \frac{Z_n}{|r_2 - R_n|}\right)\right) + \frac{1}{|r_1 - r_2|} \tag{3.18}$$

$$= \frac{1}{N-1}\left(\hat{h}_1 + \hat{h}_2\right) + \hat{V}$$

and then is projected onto an antisymmetrical two-particle basis

$$\begin{aligned} H_{ijkl} &= \langle kl|\hat{H}|ij\rangle \\ &= \frac{1}{N-1}(\delta_{ik}h_{jl} + \delta_{jl}h_{ik} - \delta_{il}h_{jk} - \delta_{jk}h_{il}) + V_{ijkl} \end{aligned}$$

with $h_{ik} = \langle k|\hat{h}|i\rangle$ and $V_{ijkl} = \langle kl|\hat{V}|ij\rangle$, such that the dimension of the Hamiltonian matches that of the 2DM. The elements of the electron-electron repulsion matrix V and the one-electron Hamiltonian h expressed in a basis of atom-centered functions are obtained from Gaussian03[70] and transformed to an orthonormal basis of (Hartree-Fock type) molecular orbitals.

Because most of chemistry is determined by the valence electrons, the fully occupied core shells can be treated as 'uncorrelated' to a good approximation. The electrons of the fully occupied inner shells are considered 'frozen' by only considering their mean-field effect. Especially for small molecules with light nuclei, freezing the core electrons has a minor effect on the energy, but it may considerably reduce the dimension of the 'active' part of the sp-basis.

The mean field effect of the inner shells can be incorporated into the Hamiltonian elements for the valence electrons, such that the dimension of the Hamiltonian and the resulting 2DM may be reduced. In fact, due to the structure of the

wavefunction under the frozen core approximation, the expectation value of any operator can be expressed as a sum of a contribution from the valence electrons and a constant term for the core electrons. This is because the wavefunction for a system with N^{core} of its N electrons frozen can be written as an antisymmetric product of a single-determinant wavefunction for the core electrons $|\Psi^{core}\rangle$ and a correlated wavefunction $|\Psi^{valence}\rangle$ for the valence electrons.

$$|\Psi\rangle = |\Psi^{core}(1,\ldots,N^{core})\ \rangle\ \wedge\ |\Psi^{valence}(N^{core}+1,\ldots,N)\ \rangle$$

such that there is no explicit correlation between the core and valence electrons. The 2DM for such a system then has the following blocks of non-zero elements,

$$\Gamma^{cccc}_{ijkl} = \langle\Psi^{core}|a_k^+ a_l^+ a_j a_i|\Psi^{core}\rangle = \delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk}$$
$$\Gamma^{vvvv}_{ijkl} = \langle\Psi^{valence}|a_k^+ a_l^+ a_j a_i|\Psi^{valence}\rangle$$
$$\Gamma^{cvcv}_{ijkl} = \langle\Psi^{core}|a_k^+ a_i|\Psi^{core}\rangle\langle\Psi^{valence}|a_l^+ a_j|\Psi^{valence}\rangle = \delta_{ik}\langle\Psi^{valence}|a_l^+ a_j|\Psi^{valence}\rangle$$

where the superscripts indicate whether the sp orbitals are frozen core (c) orbitals or valence (v) orbitals. As a consequence, the expectation value for a 2-electron operator $\hat{H}$ can be separated into a valence-electron dependent term and a constant contribution depending only on core electrons.

$$\begin{aligned}
tr\ H\Gamma &= \sum_{i<j,k<l} H^{cccc}_{ijkl}\Gamma^{cccc}_{ijkl} + \sum_{ijkl} H^{cvcv}_{ijkl}\Gamma^{cvcv}_{ijkl} + \sum_{i<j,k<l} H^{vvvv}_{ijkl}\Gamma_{ijkl} \\
&= \sum_{ij} H^{cccc}_{ijij} + \sum_{jl}\sum_{i} H^{cvcv}_{ijil}\gamma^{vv}_{jl} + \sum_{i<j,k<l} H^{vvvv}_{ijkl}\Gamma_{i<j,k<l} \\
&= \sum_{ij} H^{cccc}_{ijij} + \sum_{jkl}\left(\frac{1}{N-1}\sum_{i} H^{cvcv}_{ijil}\right)\Gamma^{vvvv}_{jklk} + \sum_{i<j,k<l} H^{vvvv}_{ijkl}\Gamma_{ijkl} \\
&= \sum_{ij} H^{cccc}_{ijij} + \sum_{i<j,k<l} \tilde{H}^{vvvv}_{ijkl}\Gamma_{ijkl}
\end{aligned}$$

The core electron contribution is thus incorporated into a Hamiltonian $\tilde{H}$ that is reduced to the active orbitals

$$\tilde{H}^{vvvv}_{ijkl} = H^{vvvv}_{ijkl} + \frac{\delta_{ik}}{N-1}\sum_{n} H^{cvcv}_{njnl} + \frac{\delta_{ik}\delta_{jl}}{N(N-1)}\sum_{mn} H^{cccc}_{mnmn}$$

It has dimension $\frac{1}{2}\tilde{K}(\tilde{K}-1)$, where $\tilde{K} = K - N^{core}$ is the number of spin orbitals that are not frozen.

3.3.2 Feasible starting points for interior-point algorithms

Interior point algorithms require initial starting points that are strictly feasible. In general, the primal matrix Γ needs to satisfy

$$\begin{aligned} & L(\Gamma) \succeq 0 \\ & tr\ A_i\Gamma = a_i \qquad i = 1, \ldots, p \end{aligned} \tag{3.19}$$

in order to be feasible. The constraints $tr\ A_i\Gamma = a_i$ can be, for instance, spin constraints like those applied in chapter 2. To find a matrix that satisfies these constraints, an initial optimization algorithm may be used that starts from a starting point which does not satisfy (3.19). The equality constraints can be imposed simply by projection.

$$\begin{aligned} & \underbrace{max}_{\Gamma, s}\ det\ (L(\Gamma) + sI) \\ & subject\ to \quad tr\ A^i\Gamma = a_i \end{aligned}$$

The initial value $s > 0$ is chosen to make the initial matrix $L(\Gamma) + sI \succ 0$ positive-definite. As soon as the algorithm produces a matrix $L(\Gamma) \succ 0$, which is a feasible starting point, it can be stopped.

3.4 Barrier method

The barrier method is the first semidefinite optimization method to be discussed here, because of both its fundamental role in the development of interior points methods and its conceptual simplicity.

3.4.1 Theoretical background

The barrier method is a straightforward extension of the barrier method for linear programming, developed by Fiacco and McCormick in the 1960's,[90] to semidefinite programming. The extension of interior point methods for linear programming to more general convex programming problems was independently

done by Alizadeh[13] and by Nesterov and Nemirovski.[12] They showed that for any conic set that has a self-concordant barrier function, there is an interior point algorithm that optimizes a linear function over this set. Such problems can be solved through a sequence of minimizations in which the conic constraint is incorporated by adding a barrier term to the original objective function. The sequence is parametrized in such a way that the subsequent minima generated in this manner converge to the minimum of the original optimization problem. The $\Phi = log\ det()$ function is a self-concordant barrier for the positive semidefinite cone. Its self-concordance, i.e. $\Phi''' \leq 2\Phi''^{3/2}$, is of theoretical importance because it makes it possible to derive an upper bound to the number of Newton iterations required for convergence, such that the algorithm can be proven to converge in polynomial time, but it does not necessarily imply that the logarithmic barrier is superior to other barrier functions in practice.[26] The barrier function enforces the semidefinite constraints by preventing its arguments from becoming singular because it grows infinitely large upon singularity. The strength of the barrier, mediated by a penalty parameter t, is decreased in subsequent minimization problems of the form (3.20).

Algorithm

initial $\Gamma:\ L(\Gamma) \succ 0$

do while $t\ \geq \epsilon$

$$\text{minimize over}\ \ \Gamma:\ \ f(\Gamma, t) = tr[H\Gamma] - t\ ln\ det\ L(\Gamma)$$

$$\text{update}\ t:\ \ t = \mu t \quad \text{with}\ \ \mu < 1$$

end do (3.20)

Intuitively, the barrier function approaches a step-function as $t \to 0$. As $t \to 0$, it switches on an infinitely high penalty on negative eigenvalues, acting much like an infinitely high step function for which the problem reduces to the original constrained optimization problem (figure 3.1). It can thus be expected to converge to the solution of the original problem in the limit $t \to 0$.

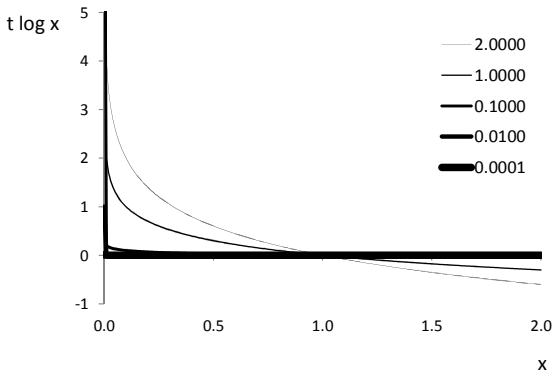

Figure 3.1: As the value of the barrier parameter t, specified in the legend, decreases to zero, the logarithmic barrier approaches an infinitely high step function.

The sequence of optimal points $\Gamma^*(t)$ as a function of t is called the *central path*, because each of these optimal dual feasible points yields a primal feasible point $X^*(t)$ such that the pair $(\Gamma^*(t), X^*(t))$ satisfies the centrality equations (3.14). The matrix $X^*(t) \equiv tL(\Gamma)^{-1}$ is primal feasible, because its positive semidefinitess follows from $X^* \succeq 0$ and because the optimality condition on the inner iteration in (3.20) implies that it satisfies $tr\ X^*L(F^i) = tr\ HF^i$ for $i = 1, \ldots n$

$$\begin{aligned}\nabla f(\Gamma, t)_i &= tr[HF^i] - t\ tr[L(\Gamma)^{-1}L(F^i)] = 0 \\ &= tr[HF^i] - tr[X^*(t)L(F^i)] = 0\end{aligned}$$

Therefore, the matrices $\Gamma^*(t)$ and $X^*(t)$ defined in this way satisfy the centrality conditions (3.14). The duality gap for the pair $\Gamma^*(t), X^*(t)$ is

$$tr[L(\Gamma^*(t))X^*(t)] = t\ dim(X^*)$$

which gives an upper bound on the deviation of the energy, $tr[H\Gamma^*(t)]$, from the optimal energy. This confirms the intuitive idea that solving the barrier equations for $t \to 0$ will lead to the optimal energy.

Referring back to the geometrical picture of the v2DM optimization under

semidefinite constraints sketched in section 1.4 of chapter 1, the central path parameterized by the sequence $\Gamma^*(t)$ approaches the boundary of the feasible set as the barrier parameter t is decreased to zero. As the barrier parameter decreases, the 2DM is allowed to become increasingly close to singular; therefore active constraints can set in as the feasible path approaches the boundary of the feasible set. These singularities mean that the equations become increasingly ill-conditioned as $\Gamma^*(t)$ approaches the boundary.

Even though the method performs relatively well given the severity of its ill-conditioning, concerns about its ill-conditioning have encouraged mathematicians to develop methods which treat the problem in a more stable manner. The most sophisticated way of handling the problem is by simultaneously optimizing the primal and the dual variable, leading to primal-dual methods, discussed in section 3.6. A modified barrier method, which is based on the classical barrier method discussed here but adds an approximate treatment of the primal problem to the purely dual optimization problem, is discussed in section 3.5.

3.4.2 Implementation of a barrier method

The barrier method is a very straightforward and robust way to carry out the constrained semidefinite optimization and has therefore served as a basis for our other implementations. The following paragraphs motivate the choice of inner optimization method and outer iteration updates used in our Fortran implementation of the logarithmic barrier method and discusses its overall performance on molecular systems. Our comparative assessment of the different algorithms focuses on one test system in particular, the LiH molecule with bond length R = 1.5417Å in a STO-6G basis set, because this system's dimensions are small enough to study it in detail. In order to compare the performance of the classical barrier method with the other semidefinite optimization algorithms presented in the next sections, the different algorithms have been compiled and run on the same computer and linked to the same Lapack and BLAS libraries.

Inner iterations

The inner iterations of the barrier method aim to solve Newton's equations for the Newton direction Δ

$$\nabla^2_\Gamma f(\Gamma, t)\ \Delta = -\nabla_\Gamma f(\Gamma, t) \tag{3.21}$$

where $f(\Gamma, t)$ is the log-barrier objective function

$$f(\Gamma, t) = tr\ H\Gamma - t\ ln\ det(L(\Gamma)) \tag{3.22}$$

Solving Newton's equations is the rate-determining step of second-order semidefinite programming methods. Because the dimension of the Hessian depends quadratically on the dimension of the 2DM, its dependence on the basis set dimension is quartic. Storing the Hessian thus becomes impossible except for small basis sets. Instead of factorizing the Hessian to solve the equations exactly, we are forced to use an iterative solver, such as a Krylov subspace method.

Krylov subspace method

Krylov subspace methods are especially advantageous in solving Newton's equations when constructing the full Hessian becomes too expensive, because they only require the Hessian-vector product. The Hessian-vector product can either be approximated by a finite-difference scheme or, in this case, can easily be calculated exactly. The Hessian involved in the dual barrier method allows a simple analytic expression for the Hessian-vector product that is similar in speed to its finite-difference approximation.

$$(\nabla^2 f)_{ij} = t\ tr\ [L(\Gamma)^{-1} L(F^i) L(\Gamma)^{-1} L(F^j)] \tag{3.23}$$

For any traceless update vector $\Delta = \sum_i \Delta_i F^i$, the Hessian-vector product

becomes

$$\begin{aligned}(\nabla^2 f)\Delta &= t\sum_i tr\ [L(\Gamma)^{-1}L(F^i)L(\Gamma)^{-1}\sum_j L(F^j)\Delta_j]F^i \\ &= t\ \sum_i tr\ [L^\dagger\left(L(\Gamma)^{-1}L(\Delta)L(\Gamma)^{-1}\right)F^i]F^i \\ &= t\ \mathcal{P}_{\perp F^0}\Big(L^\dagger\left(L(\Gamma)^{-1}L(\Delta)L(\Gamma)^{-1}\right)\Big)\end{aligned}$$

More specifically, for $L(\Gamma) = \Gamma \oplus Q(\Gamma) \oplus G(\Gamma)$

$$\begin{aligned}(\nabla^2 f)\Delta = t\ \mathcal{P}_{\perp F^0}\Big(&\left(\Gamma^{-1}\Delta\Gamma^{-1}\right) + Q\left(Q(\Gamma)^{-1}Q(\Delta)Q(\Gamma)^{-1}\right) \\ &+ G^\dagger\left(G(\Gamma)^{-1}G(\Delta)G(\Gamma)^{-1}\right)\Big)\end{aligned}$$

Because of the matrix-matrix multiplications, assembling the Hessian-vector product takes $O(K^6)$ flops.

An inherent drawback of Krylov subspace methods is their inefficiency when dealing with ill-conditioned systems of equations. Because the barrier optimization problem (3.22) is convex, the Hessian involved in Newton's equations is positive semidefinite. For this reason, the conjugate gradients (CG) method seems the most suitable approach to solving them in an iterative manner. However, the speed of convergence of the CG method depends heavily on the condition number of the Hessian and on the extent to which its eigenvalues are clustered. The ill-conditioning of the Hessian towards convergence (figure 3.2) severely slows down convergence on the inner iterations of the final outer loops of the barrier program (figure 3.3). The Hessian's spectrum covers a wide range of values, from near-zero to very large eigenvalues, and becomes less and less clustered towards the optimum. This makes it an especially difficult optimization problem. The conjugate gradient method may even need substantially more iterations to converge than necessary to span the whole n-dimensional Krylov space for small values of t.

As an alternative to the CG method, we have tried a minimum residual method adjusted to deal with a positive-semidefinite Hessian to solve the inner Newton's equations, because it may deal better with near-singularities. In

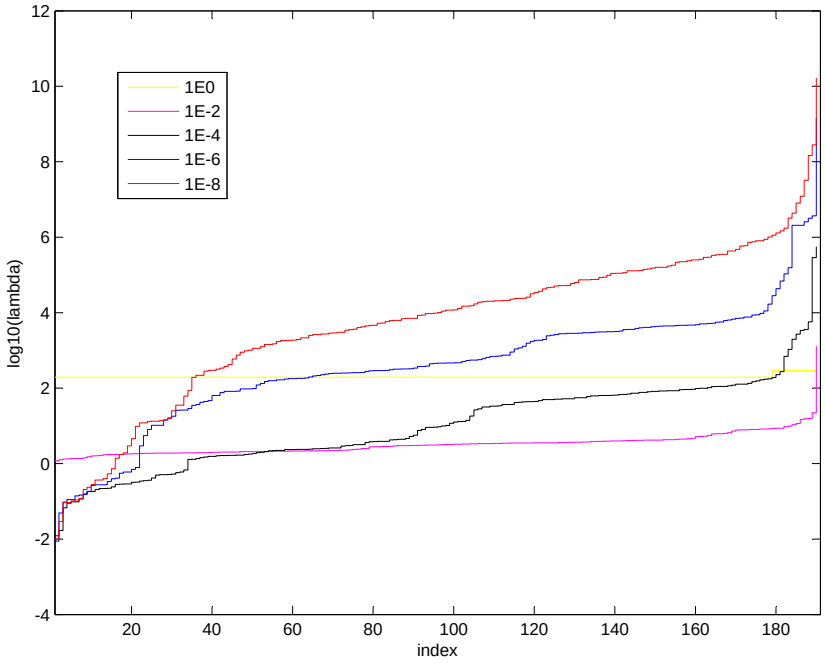

Figure 3.2: The evolution of the spectrum, with its eigenvalues indexed in increasing order, of the projected Hessian for LiH shows its increasing ill-conditioning as the barrier parameter t is decreased. The barrier parameter t is a measure of the duality gap, $dim(L(\Gamma))t = 276\ t$. The Hessian for the LiH molecule in the STO-6G basis was projected onto the space of traceless matrices, $\mathcal{P}(H) = (I - e_0 e_0^T)H(I - e_0 e_0^T)$ with e_0 the vector representation of the identity matrix, and then diagonalized in its reduced-rank form.

general, MINRES needs about half as many iterations as CG to obtain a result with similar accuracy.

Preconditioners

Since the rate of convergence of typical Krylov subspace methods applied to Newton's equations depends so heavily on the spectrum of the Hessian, an equivalent equation could be solved which has a better structured spectrum,

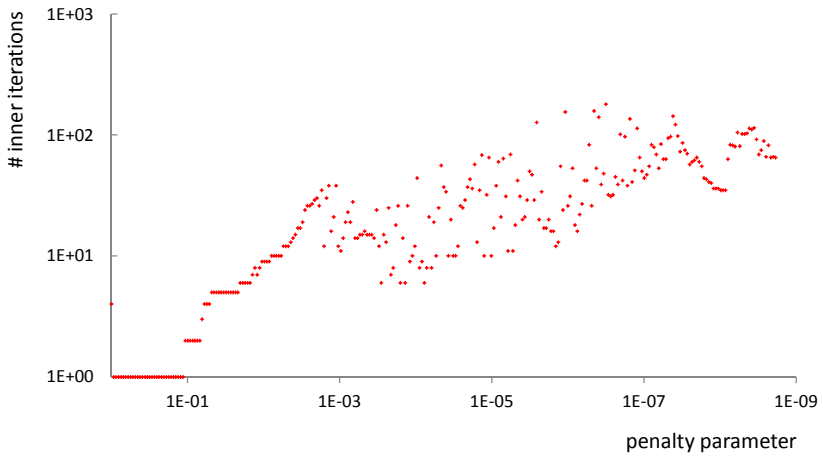

Figure 3.3: Due to the increasing ill-conditioning of the Hessian, the CB method needs an increasing number of inner Krylov subspace iterations to solve Newton's equations as the barrier parameter – and the duality gap – are decreased to zero.

such as

$$M^{-1}\nabla^2 f\Delta = -M^{-1}\nabla f$$

The matrix M^{-1} is then called a preconditioner, because it aims to make the Hessian's spectrum more tightly clustered or better conditioned in order to speed up convergence. Indeed, when M^{-1} is a good approximation to $(\nabla^2)^{-1}f$, the equations may be expected easier to solve. The workings of preconditioners are not always well understood, and their justification may be based pragmatically on the fact that they speed up convergence. However, it is not easy to find a good preconditioner for typical Hessians that arise in the v2DM barrier method. Straightforward preconditioning techniques, such as diagonal preconditioners, were counterproductive. Block diagonal preconditioners were only productive when the dimension of the blocks considered was very large. Incomplete factorization techniques only caused a speed-up at very low drop tolerances ($\leq 10^{-6}$), making them ineffective. These techniques, however, require storage of the Hessian, which is highly undesirable.

The main difficulty in our setup is to find good preconditioners that only require the Hessian-vector product and do not need to access the full Hessian. Most general-purpose preconditioning techniques, such as Jacobi, Gauss-Seidel and incomplete factorization techniques, require at least several rows or columns at the same time. Although we have observed that limited-memory quasi-Newton schemes to approximate the inverse Hessian[91] did not give the desired accuracy, especially in the final outer loops, we might use the concept to construct a preconditioner, since it builds an approximation to the inverse Hessian-vector product. This leads to the idea of an 'automatic preconditioner'[92] that is constructed from the Hessian-vector products generated in an iterative Krylov subspace method.

The quasi-Newton type 'automatic preconditioner' is constructed by storing a number of Krylov subspace vectors m generated in subsequent Krylov subspace iterations, and can then be applied right away to speed up the convergence of the Krylov subspace method. The notation $\tilde{H}$ is used to denote the quasi-Newton type approximation to the inverse Hessian, and is constructed according to

$$\begin{aligned}
\tilde{H}^{(k+1)} &= \left(I - \frac{1}{\Delta x^{(k)^T}\Delta r^{(k)}}\Delta r^{(k)}\Delta x^{(k)^T}\right)\tilde{H}^{(k)}\left(I - \frac{1}{\Delta x^{(k)^T}\Delta r^{(k)}}\Delta r^{(k)}\Delta x^{(k)^T}\right)^T \\
&\quad + \frac{1}{\Delta x^{(k)^T}\Delta r^{(k)}}\Delta x^{(k)}\Delta x^{(k)^T} \\
\Delta x^{(k)} &= x^{(k+1)} - x^{(k)} \\
\Delta r^{(k)} &= r^{(k+1)} - r^{(k)}
\end{aligned}$$

where $x^{(k)}$ are the iterates generated in the Krylov subspace method and $r^{(k)}$ the residuals. This formula guarantees a symmetrical positive definite approximation to the inverse Hessian, as long as $\Delta x^{(k)^T}\Delta r^{(k)} > 0$, which is by definition fulfilled if they are taken as the CG iterates and residuals. Since storing the Hessian requires too much memory, a recursive formula for calculating the inverse Hessian-vector product $\tilde{H}v$ can be used that only requires the set of stored vectors $\left\{\Delta x^{(k)}\right\}, \left\{\Delta r^{(k)}\right\}$.[93] Only a small number m of such vectors can be stored for large optimization problems. These can be chosen from the

first m Krylov subspace iterations or follow a specific distribution within the set of generated Krylov subspace vectors. Constructing the quasi-Newton type preconditioner, however, adds some overhead to each Krylov subspace iteration, and is therefore only advantageous if the reduction in the number of iterations needed to converge is large enough. However, applications of a quasi-Newton type preconditioner constructed from a variable number of stored Krylov subspace vectors indicates that the reduction in the number of Krylov subspace iterations for a small number of stored vectors m is only modest and a more substantial reduction requires storing and manipulating a lot more vectors. Overall this approach did not lead to a significant reduction in CPU time (table 3.1).

Line search

After the Newton direction Δ has been calculated, a line search will prevent the 2DM from leaving the feasible set upon updating Γ to $\Gamma + \alpha\Delta$. A line search procedure based on a generalized eigenvalue decomposition avoids the need to compute the lowest eigenvalue of $L(\Gamma)$ in every iteration.
The eigenvalue decomposition of $L(\Gamma)^{-1/2}L(\Delta)L(\Gamma)^{-1/2}$ is computed

$$L(\Gamma)^{-1/2}L(\Delta)L(\Gamma)^{-1/2} = V\Lambda V^T \tag{3.24}$$

The directional derivative of the objective function can then be expressed in terms of its eigenvalues, making its dependence on the line search coefficient explicit.

$$\begin{aligned}
&\frac{\partial f(\Gamma + \alpha\Delta)}{\partial \alpha} \\
&= tr\ H\Delta - t\ tr\ [L(\Gamma + \alpha\Delta)^{-1}L(\Delta)] \\
&= tr\ H\Delta - t\ tr\ \left[\left(L(\Gamma)^{1/2}(I + \alpha V\Lambda V^T)L(\Gamma)^{1/2}\right)^{-1} L(\Gamma)^{1/2}V\Lambda V^T L(\Gamma)^{1/2}\right] \\
&= tr\ H\Delta - t\ tr\ \left[L(\Gamma)^{-1/2}V(I + \alpha\Lambda)^{-1}V^T L(\Gamma)^{-1/2}L(\Gamma)^{1/2}V\Lambda V^T L(\Gamma)^{1/2}\right] \\
&= tr\ H\Delta - t\ tr\ \left[(I + \alpha\Lambda)^{-1}\Lambda\right] \\
&= tr\ H\Delta - t\ \sum_i \frac{\lambda_i}{1 + \alpha\lambda_i}
\end{aligned}$$

m	# PCG	# PCR
0	231	188
3	359	1146
5	317	995
8	262	587
10	220	973
15	149	394
20	129	654
30	117	375
50	150	359
70	203	342
80	96	251

Table 3.1: The number of conjugate gradient (PCG) and conjugate residuals (PCR) iterations needed to calculate the Newton step for $t = 10^{-6}$ in the LiH (STO-6G) test system can be reduced by about a factor 2 by using an 'automatic' L-BFGS type preconditioner constructed from the first m vectors generated in the Krylov subspace method, using an initial matrix $\tilde{H}^{(0)} = I$. Adjusting the weight of the initial matrix $\tilde{H}^{(0)} = wI$ may greatly influence the number of iterations needed, but even the best choices for w do not reduce the number of iterations by much more than a factor $2 - 3$. At least for this choice of initial $H^{(0)}$, the automatic preconditioner was much more effective in the PCG method than in the PCR method (algorithms A.2 and A.5).

The minimum $\frac{\partial f(\Gamma+\alpha\Delta)}{\partial\alpha} = 0$ can then be found easily by means of a bisection method in the interval $\left[0, \frac{-1}{\lambda^{min}}\right]$, which guarantees positive semidefiniteness of $L(\Gamma + \alpha\Delta)$ after update.

Outer iterations and overall performance

Since the duality gap in the classical barrier method decreases linearly with the barrier parameter t, it needs to be decreased to a very small value. The duality gap $t\ dim(L(\Gamma))$ is substantially bigger than t, as typical dimensions of $L(\Gamma)$ are of the order 10^2 or more. Therefore the outer iterations are repeated until t is decreased to $\epsilon \approx 10^{-9}$. We have chosen to use a static update $t^{(k+1)} = t^{(k)}\mu$ with a constant factor $\mu = 1.075$ (figure 3.4).

The overall scaling of the barrier method with the sp basis dimension is at least $O(K^6)$.
The barrier method algorithm is a four-fold loop:

The outermost iterations optimize $f(\Gamma, t)$ for decreasing values of t. The number of outermost iterations is fully determined by the update scheme for t, and can be kept constant with system size.

On a lower level, solving Newton's equations to minimize $f(\Gamma, t)$ requires several Newton's iterations in which a system $\nabla^2 f(\Gamma, t)\ \Delta = -\nabla f(\Gamma, t)$ is solved. Our calculations indicate that the number of Newton iterations does not change much with system size and is very small (often less than 5).

Calculating each Newton step Δ is done by a Krylov subspace method. Depending on the spectrum of the Hessian, this may take a nearly constant number of iterations, but in the worst case the number of iterations equals the dimension and is therefore $O(K^2)$. Given the spectrum of the Hessian (figure 3.2) for typical systems, calculating the Newton direction for the initial values of t takes a nearly constant number of iterations, whereas for small values of t the number will grow as $O(K^2)$.

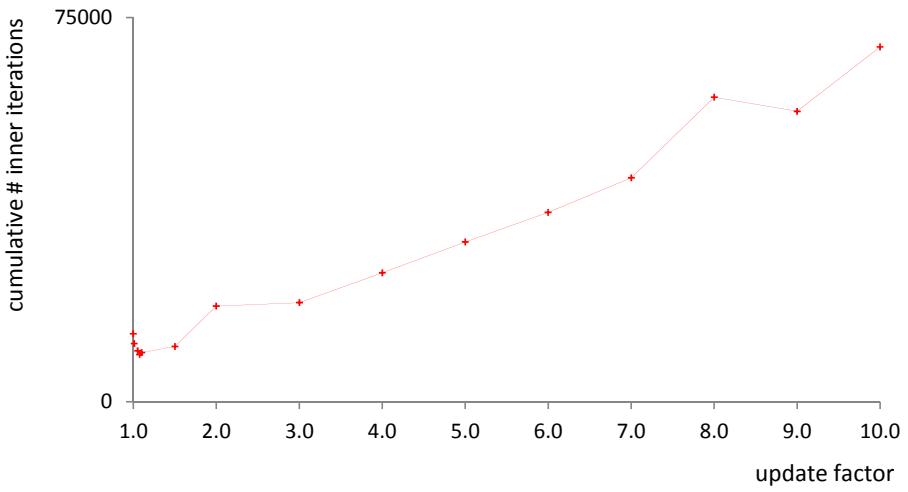

Figure 3.4: The cumulative number of inner Krylov subspace iterations needed by the classical barrier method to converge for the LiH (STO-6G) test system is smallest when an update factor for the penalty parameter around 1.075 is used. More aggressive update schemes make Newton's equations more difficult to solve, resulting in an overall larger number of inner iterations, despite a smaller number of outer iterations. For this reason, we have used an update factor of 1.075 throughout, unless specified otherwise.

The innermost iterations are the Krylov subspace iterations and take $O(K^6)$ flops to assemble the Hessian-vector product.

Therefore, the barrier method's overall scaling may range from $O(K^6)$ to $O(K^{12})$ in the worst-case scenario, though practical calculations suggest it is closer to $O(K^6)$ (figure 3.5).

Because of the method's ill-conditioning and the use of approximate iterative methods to solve the inner Newton iterations, its accuracy is limited to about 10^{-4} Hartree.

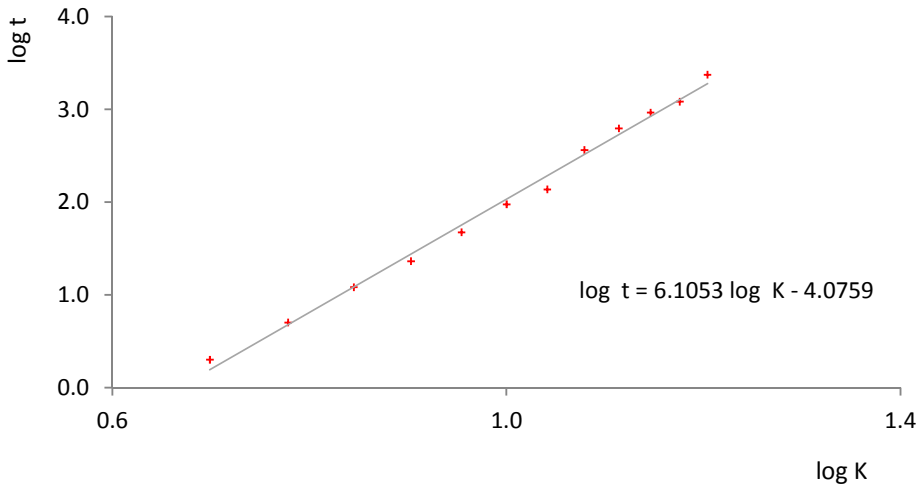

Figure 3.5: The CPU times required by the CB algorithm to optimize the energy of half-filled Hubbard models, with an equal number of spatial orbitals $\frac{K}{2}$ and particles N, $\frac{K}{2} = N$, and interaction strength 1.0, confirms that the algorithm scales roughly as $O(K^6)$ with the dimension of the sp basis.

3.5 Modified barrier method

A modified barrier method has been developed by Polyak for linear and nonlinear programming[94] to overcome the explicit ill-conditioning of the classical barrier method. These algorithms for linear programming have inspired us to adjust our previously discussed classical barrier method to a modified barrier method, which resulted in an extension to semidefinite programming that is very similar to the one made by Stingl and Kocvara.[95,96] Another extension to semidefinite programming, similar to the approach by Zibulevski et al.,[97] in our calculations often resulted in an infeasible update. Therefore, we focus here on the first approach.

3.5.1 Theoretical background

The inherent ill-conditioning of the Newton equations involved in the classical barrier method originates from the non-existence of the classical barrier function at the solution, and its divergence to infinity as it approaches the solution. The barrier function can be modified to exist at the solution by shifting the argument. Instead of the constraint $L(\Gamma) \succeq 0$, the modified barrier method considers a shifted constraint $\frac{1}{t}L(\Gamma) \succeq -I$, which becomes equivalent to the original constraint as the penalty parameter t decreases to zero. However, this will only shift the poles of the barrier function. The main difference from the classical barrier method is that the modified barrier method attempts to make contact with the primal problem by introducing a positive-semidefinite Lagrange multiplier X into the objective function along with a matrix barrier Φ to impose the constraint $\frac{1}{t}L(\Gamma) \succeq -I$

$$f(\Gamma, X, t) = tr\ [H\Gamma] + t\ tr\ \Big[X\Phi\Big(\frac{1}{t}L(\Gamma)\Big)\Big] \tag{3.25}$$

This method is a straightforward extension of the modified barrier method for linear programming, first developed by Polyak,[94] to semidefinite programming. The scalar barrier function ϕ used in linear programming can be extended to semidefinite programming by applying it to the eigenvalues of a semidefinite constraint matrix with eigenvalue decomposition $V\ diag(\lambda_1, \ldots, \lambda_n)\ V^T$ to yield a matrix barrier function Φ

$$\Phi(V\ diag(\lambda_1, \ldots, \lambda_n)\ V^T) = V\ diag(\phi(\lambda_1), \ldots, \phi(\lambda_n))\ V^T$$

Suitable barrier functions $\phi(\lambda)$ with domain $dom\ \phi =]-1, +\infty[$ satisfy

$$\phi \text{ is strictly decreasing, strictly convex, twice differentiable} \tag{3.26}$$

$$\lim_{\lambda \to -1} \phi'(\lambda) = -\infty \tag{3.27}$$

$$\lim_{\lambda \to \infty} \phi'(\lambda) = 0 \tag{3.28}$$

$$\phi(0) = 0 \tag{3.29}$$

$$\phi'(0) = -1 \tag{3.30}$$

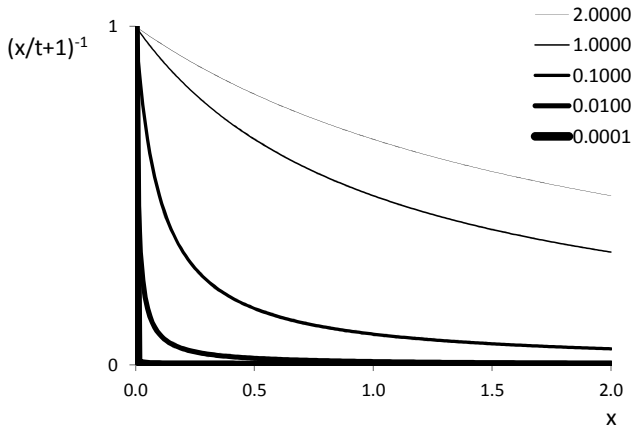

Figure 3.6: The inverse barrier $\phi(\lambda) = (\lambda+1)^{-1}$ is a suitable barrier for the modified barrier method. Shown here are the functions $\phi(\lambda) = (\frac{1}{t}\lambda + 1)^{-1}$ for several values of t, specified in the key.

The barrier function must preserve the convexity of the original optimization problem and penalize negative eigenvalues, but allow inactive constraints to remain inactive – hence the properties (3.26) to (3.28). At the same time, it must return the original objective function at the optimum, which is ensured by the property (3.29). The last requirement (3.30) on the barrier function, $\phi'(0) = -1$, ensures that Lagrange multipliers corresponding to active constraints remain unchanged on update (vide infra). The logarithmic barrier $\phi(\lambda) = -ln\ (\lambda{+}1)$ and the inverse barrier $\phi(\lambda) = (\lambda{+}1)^{-1}$, for instance, fulfil these requirements (figure 3.6). Because the logarithmic barrier is more difficult to extend to a matrix barrier function, we will mainly focus on the inverse barrier $\phi(\lambda) = (\lambda + 1)^{-1}$ which gives a matrix barrier $\Phi(X) = (X + I)^{-1}$ with conveniently closed expressions for the first and second derivatives. The specific objective function considered is thus

$$f(\Gamma, X, t) = tr\ [H\Gamma] + t\ tr\ \left[X\Big(\frac{1}{t}L(\Gamma) + I\Big)^{-1}\right] \tag{3.31}$$

The fundamental idea behind the method is to accelerate convergence to

the optimum by updating a Lagrange multiplier X alongside Γ in the modified barrier function, even if X is only a crude approximation to the primal variable. The modified barrier function $f(\Gamma, X, t)$ shares some key properties with the Lagrangian

$$f(\Gamma^*, X^*, t) = tr[H\Gamma^*] = E^* \tag{3.32}$$

$$\nabla_\Gamma f(\Gamma^*, X^*, t)_i = H_i - tr[X^* L(F^i)] = 0 \tag{3.33}$$

$$f(\Gamma, X, t) \quad \text{is convex in} \quad \Gamma \quad \forall \Gamma : L(\Gamma) \succeq -tI \tag{3.34}$$

Yet it has an advantage over the Lagrangian, as it can be proven (under strict complementarity of X^* and the constraint matrix $L(\Gamma^*)$) that $f(\Gamma, X^*, t)$ is strongly convex for all $t < t_0$ for a certain value $t_0 > 0$, in a neighborhood of Γ^*.[95]

Updating the Lagrange multiplier in each outer loop allows a faster than linear decrease of the duality gap with the penalty parameter. In contrast to a primal-dual approach, the modified barrier method does not optimize the Lagrange multiplier X, it only updates it after each minimization of the modified barrier function over Γ. Key to this method is the choice of update for the Lagrange multiplier X. The update is chosen in such a way that minimizing the modified barrier function (3.31) ensures that the Lagrangian $\mathcal{L}(\Gamma, X)$ for the original problem is minimized over the 2DM at the same time, such that the update defines a primal feasible point,

$$\nabla_\Gamma f(\Gamma^{(k+1)}, X^{(k)}, t^{(k)}) = \nabla_\Gamma \mathcal{L}(\Gamma^{(k+1)}, X^{(k+1)}) = 0 \ .$$

This can be effected by choosing the update

$$X^{(k+1)} = \left(\frac{1}{t^{(k)}} L(\Gamma^{(k+1)}) + I\right)^{-1} X^{(k)} \left(\frac{1}{t^{(k)}} L(\Gamma^{(k+1)}) + I\right)^{-1} \tag{3.35}$$

which preserves positive semidefiniteness and symmetry and is primal feasible if exact optimality holds. The duality gap for the primal-dual feasible pair is

therefore bounded above by

$$\begin{aligned} & tr\ \left[X^{(k+1)} L(\Gamma^{(k+1)})\right] \\ & = tr\ \left[\left(\frac{1}{t^{(k)}} L(\Gamma^{(k+1)}) + I\right)^{-1}\ X^{(k)}\ \left(\frac{1}{t^{(k)}} L(\Gamma^{(k+1)}) + I\right)^{-1} L(\Gamma^{(k+1)})\right] \\ & = \sum_i (v_i^T X^{(k)} v_i) \frac{\lambda_i}{(\frac{1}{t}\lambda_i + 1)^2} \qquad \text{with} \quad L(\Gamma^{(k+1)}) = \sum_i \lambda_i v_i v_i^T \end{aligned}$$

This update for the Lagrange multiplier will help to reduce the duality gap by increasing the barrier function's sensitivity to negative eigenvalues in the dual matrix $L(\Gamma)$ and decreasing its sensitivity to positive eigenvalues of $L(\Gamma)$.

Several alternative update strategies for the penalty parameter t and the Lagrange multiplier X exist. They can be updated simultaneously at each outer iteration or alternatingly, depending on how much progress was made in the previous iteration. The alternating update scheme is advocated by Conn et al. for linear programming,[98] but is more complicated than simultaneous updating. However, convergence can also be proven for a simultaneous update scheme for t and X, where t is updated by a constant factor.[94] We have therefore chosen a simultaneous update scheme (algorithm 3.36).
In contrast to the classical barrier method, for which a limit on the number of iterations needed for convergence can be proven, such a limit has not been proven for the modified barrier problem.

Because this method relaxes the original semidefinite constraint on the 2DM, strictly speaking it is not an interior point method with respect to the original optimization problem. It is only an interior point method with respect to the modified constraint $\frac{1}{t} L(\Gamma) \succeq -I$. Therefore, it does not require a starting point with positive definite P-,Q- and G-map.
More detailed information on the method can be found in work by Stingl[95] and Kocvara et al.[96]

Algorithm

initial $\Gamma:\ L(\Gamma) \succ 0$

initial $X:\ X \equiv I$

do while $tr\ [XL(\Gamma)]\ \geq \epsilon$ or $\Delta E > \epsilon$

$$\text{minimize over } \Gamma:\ f(\Gamma, t) = tr[H\Gamma] + t\ tr\ \left[\left(\frac{1}{t}L(\Gamma) + I\right)^{-1} X\right]$$

$$\text{update } X\quad : X^{(k+1)} = \left(\frac{1}{t^{(k)}}L(\Gamma^{(k+1)}) + I\right)^{-1} X^{(k)}\ \left(\frac{1}{t^{(k)}}L(\Gamma^{(k+1)}) + I\right)^{-1}$$

$$\text{update } t:\ t = \mu t \quad \text{with}\ \ \mu < 1$$

end do (3.36)

3.5.2 Implementation of a modified barrier method

The modified barrier method updates a Lagrange multiplier alongside the 2DM in order to make the duality gap decrease faster than in the barrier method, where it is bound to decrease linearly with the penalty parameter. Although the main incentive behind the modified barrier method is to keep the objective function from becoming singular upon convergence, our applications suggest that its improvement over the classical barrier method is mainly due to its use of an approximate primal variable alongside the dual variable.

Our implementation of the modified barrier (MB) method with an inverse barrier function was done in Fortran. It was based on our implementation of the CB method, so both programs have the same framework and underlying routines.

Inner iterations

In practice, the modified barrier method still bears a strong resemblance to the classical barrier method. Its inner iterations optimize the matrix barrier function over Γ

$$f(\Gamma, X, t) = tr\ H\Gamma - t\ tr\ \left[X\left(\frac{1}{t}L(\Gamma) + I\right)^{-1}\right] \tag{3.37}$$

which is done by solving Newton's equations.

$$\nabla^2 f(\Gamma, X, t)\Delta = -\nabla_\Gamma f(\Gamma, X, t)$$

Krylov subspace method

The inner Newton's equations can be solved iteratively by means of a Krylov subspace method, using an exact expression for the Hessian-vector product.

$$\begin{aligned}
&(\nabla^2 f)_{ij} \\
&= \frac{1}{t} tr \left[X \left(\frac{1}{t} L(\Gamma) + I \right)^{-1} L(F^i) \left(\frac{1}{t} L(\Gamma) + I \right)^{-1} L(F^j) \left(\frac{1}{t} L(\Gamma) + I \right)^{-1} \right] \\
&+ \frac{1}{t} tr \left[X \left(\frac{1}{t} L(\Gamma) + I \right)^{-1} L(F^j) \left(\frac{1}{t} L(\Gamma) + I \right)^{-1} L(F^i) \left(\frac{1}{t} L(\Gamma) + I \right)^{-1} \right]
\end{aligned}$$

For any traceless update vector $\Delta = \sum_i \Delta_i F^i$, the Hessian-vector product becomes

$$\begin{aligned}
&(\nabla^2 f)\Delta \\
&= \frac{1}{t} \sum_i tr \left[\left(\frac{1}{t} L(\Gamma) + I \right)^{-1} X \left(\frac{1}{t} L(\Gamma) + I \right)^{-1} L(\Delta) \left(\frac{1}{t} L(\Gamma) + I \right)^{-1} L(F^i) \right] F^i \\
&+ \frac{1}{t} \sum_i tr \left[\left(\frac{1}{t} L(\Gamma) + I \right)^{-1} L(\Delta) \left(\frac{1}{t} L(\Gamma) + I \right)^{-1} X \left(\frac{1}{t} L(\Gamma) + I \right)^{-1} L(F^i) \right] F^i \\
&= \frac{1}{t} \mathcal{P}_{\perp F^0} \left(L^\dagger \left(\left(\frac{1}{t} L(\Gamma) + I \right)^{-1} X \left(\frac{1}{t} L(\Gamma) + I \right)^{-1} L(\Delta) \left(\frac{1}{t} L(\Gamma) + I \right)^{-1} \right) \right) \\
&+ \frac{1}{t} \mathcal{P}_{\perp F^0} \left(L^\dagger \left(\left(\frac{1}{t} L(\Gamma) + I \right)^{-1} L(\Delta) \left(\frac{1}{t} L(\Gamma) + I \right)^{-1} X \left(\frac{1}{t} L(\Gamma) + I \right)^{-1} \right) \right)
\end{aligned}$$

Although the inner Newton's equations in the modified barrier method are not better conditioned than in the classical barrier method, the modified barrier method converges faster to the optimum. The Hessian is not considerably better conditioned nor significantly more clustered than in the classical barrier method (figure 3.7). It may even be more ill-conditioned for the same penalty parameter (3.8, 3.9). Nonetheless, the modified barrier method needs an overall smaller

number of inner Krylov subspace iterations to converge to an ϵ-suboptimal energy than the classical barrier method because it needs considerably fewer outer iterations. In contrast to the classical barrier method, the penalty parameter does not need to be decreased to 10^{-8} or 10^{-9}, but only to 10^{-5} or 10^{-6}, since the duality gap decreases faster than linearly with the penalty parameter (figure 3.11). Given its smaller penalty parameter upon convergence, it can be expected to be a little better conditioned than the classical barrier method (3.10), although the biggest reduction in computational cost (table 3.2) comes from the reduced number of outer iterations.

Similarly to the classical barrier method, the MINRES method needs fewer inner iterations to solve Newton's equations than the CG method, hence we have chosen to use the MINRES method in all applications.

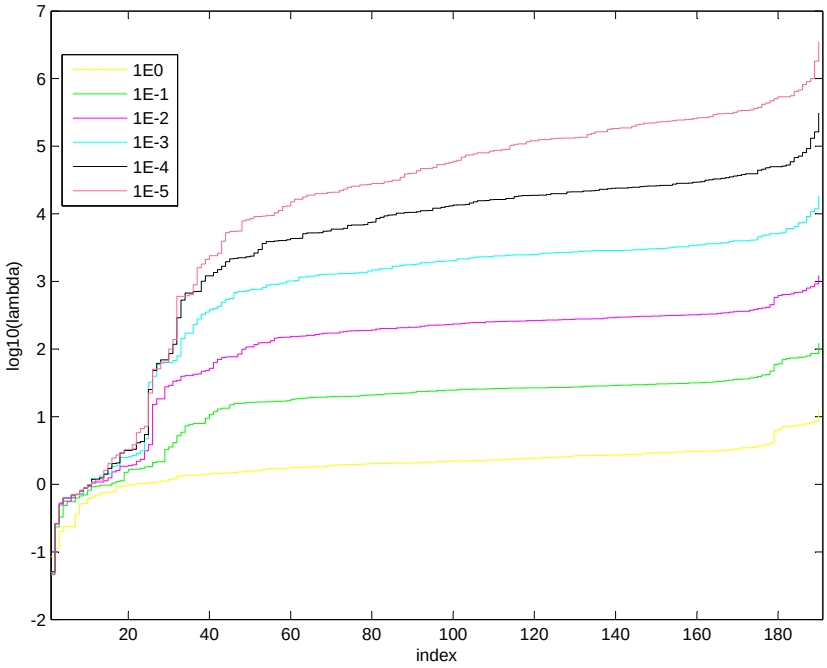

Figure 3.7: The projected Hessian for different values of the penalty parameter t for the system LiH (STO-6G) in the MB method is not better conditioned than the Hessian in the CB method for the same penalty parameter (figure 3.2).

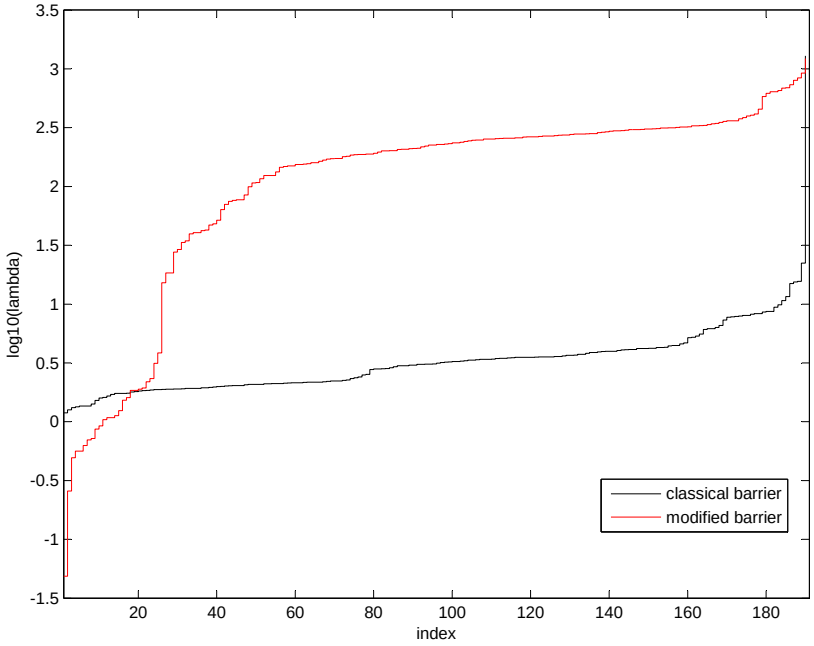

Figure 3.8: The projected Hessian's condition number for LiH may even be worse in the MB method than in the CB method for the same penalty parameter $t = 10^{-2}$.

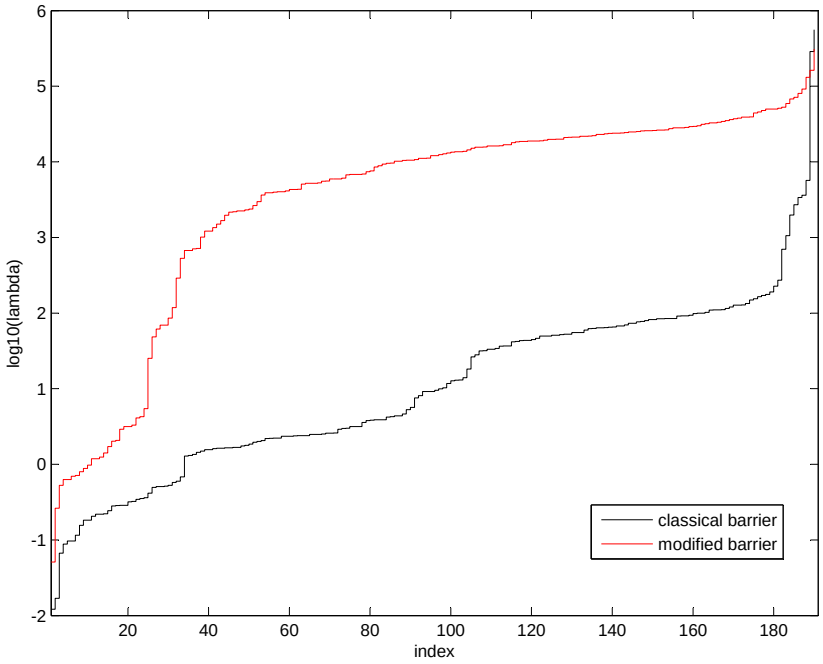

Figure 3.9: The projected Hessian for LiH may even be worse conditioned in the MB method than in the CB method for the same penalty parameter $t = 10^{-4}$.

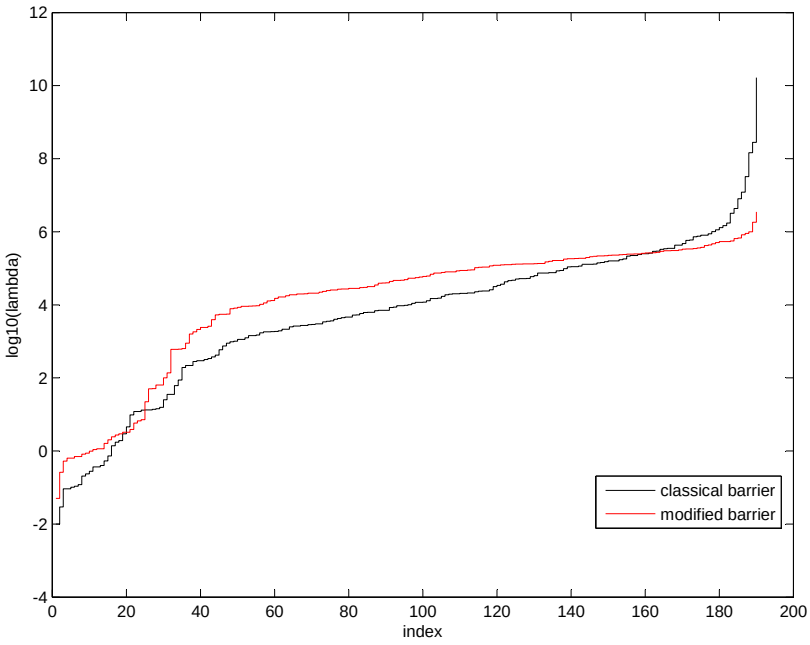

Figure 3.10: At convergence, both the Hessians involved in the MB and CB Newton's equations are highly ill-conditioned. However, because the MB method takes lesser outer iterations to converge ($t = 10^{-5}$ at convergence, as opposed to $t = 10^{-8}$ in the CB method), it is faster for almost all systems.

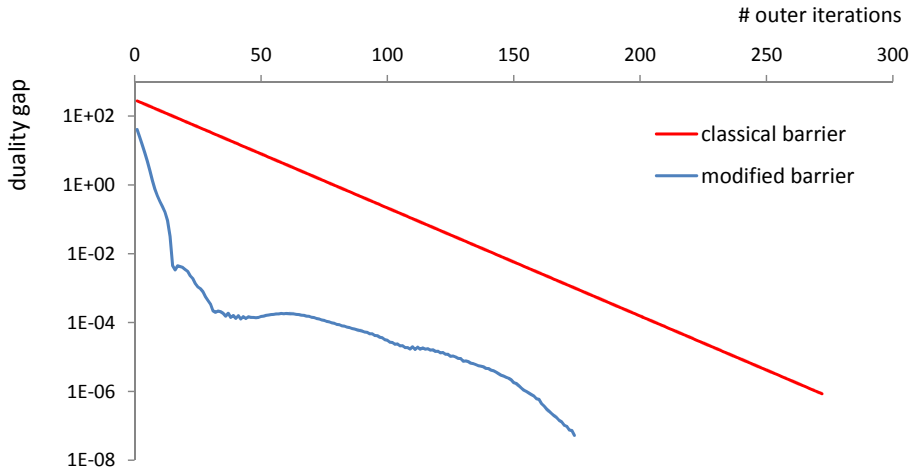

Figure 3.11: The MB method reduces the PD gap faster with the cumulative number of outer iterations performed than the CB method by approximating a primal variable alongside the dual variable. It is thus able to provide faster than linear convergence of the duality gap. As a consequence, the penalty parameter does not need to be reduced as far as in the CB method.

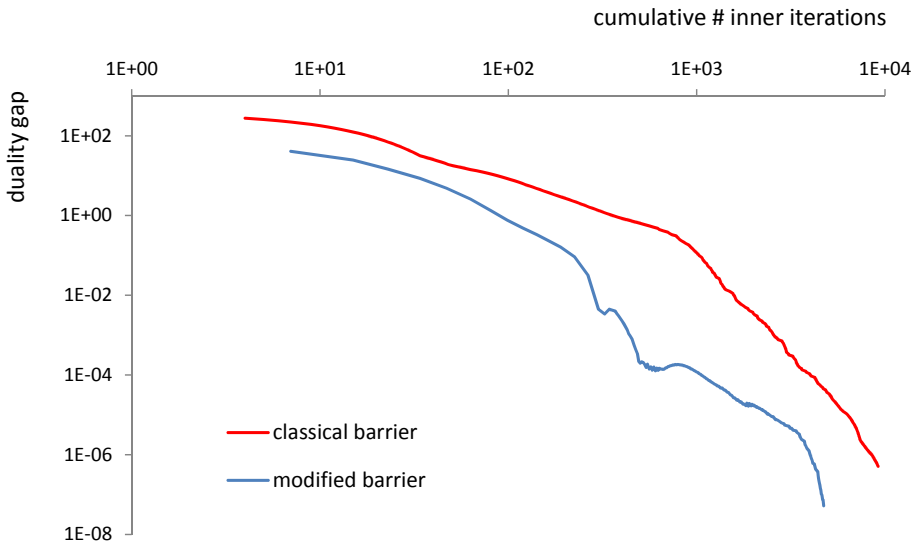

Figure 3.12: The MB method reduces the duality gap faster than linearly with the number of outer iterations, but also requires slightly more inner Krylov subspace iterations to solve Newton's equations. Nonetheless, due to its smaller number of outer iterations needed to converge, it also needs an overall smaller number of Krylov subspace iterations to reach the duality gap shown on the vertical axis.

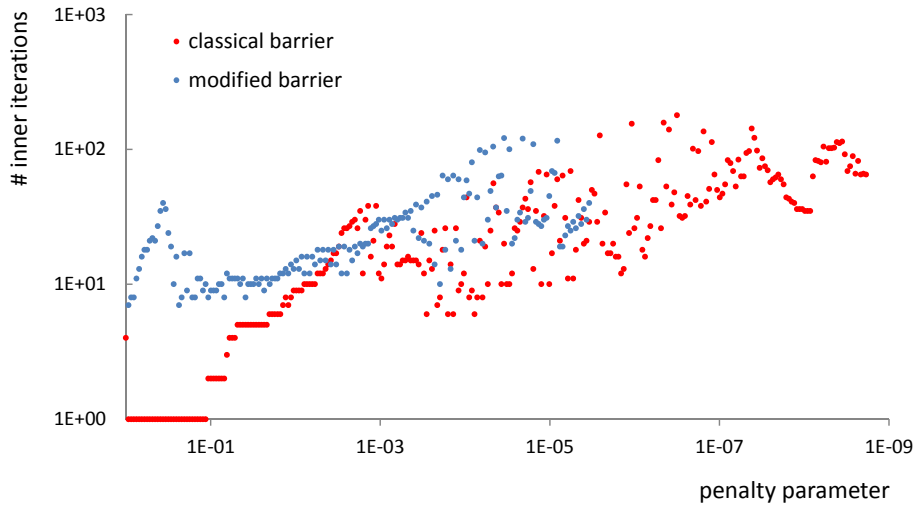

Figure 3.13: Although the idea behind the MB method is to make its inner iterations better conditioned, in fact it needs a few more inner Krylov subspace iterations to solve Newton's equations for the same penalty parameter. This agrees with the finding that its Hessian's condition number is not smaller, or is even higher, than in the CB method for the same penalty parameter.

Line search

A similar line search to that used in the CB method, based on a generalized eigenvalue composition of $L(\Delta)$, can be used for the MB method. Although it requires taking an additional inner product with the matrix X, this can be done beforehand so it is not needed in every step of the bisection method. The eigenvalue decomposition of $\left(\frac{1}{t}L(\Gamma)+I\right)^{-1/2} L(\Delta) \left(\frac{1}{t}L(\Gamma)+I\right)^{-1/2}$ is computed,

$$\left(\frac{1}{t}L(\Gamma)+I\right)^{-1/2} L(\Delta) \left(\frac{1}{t}L(\Gamma)+I\right)^{-1/2} = V\Lambda V^T \tag{3.38}$$

such that the directional derivative of the objective function can be expressed in terms of its eigenvalues and its dependence on the line search coefficient.

$$\begin{aligned}
&\frac{\partial f(\Gamma+\alpha\Delta)}{\partial\alpha} \\
&= tr\ H\Delta - tr\ \left[X\left(\frac{1}{t}L(\Gamma)+\frac{\alpha}{t}\Delta+I\right)^{-1} L(\Delta)\left(\frac{1}{t}L(\Gamma)+\frac{\alpha}{t}\Delta+I\right)^{-1}\right] \\
&= tr\ H\Delta - tr\ \Big[X\left(\frac{1}{t}L(\Gamma)+I\right)^{-1/2}\left(I+\frac{\alpha}{t}V\Lambda V^T\right)^{-1} V\Lambda V^T \\
&\qquad\qquad \left(I+\frac{\alpha}{t}V\Lambda V^T\right)^{-1}\left(\frac{1}{t}L(\Gamma)+I\right)^{-1/2}\Big] \\
&= tr\ H\Delta - \sum_i \frac{\lambda_i}{(1+\frac{\alpha}{t}\lambda_i)^2}\left(V^T\left(\frac{1}{t}L(\Gamma)+I\right)^{-1/2} X\left(\frac{1}{t}L(\Gamma)+I\right)^{-1/2} V\right)_{ii}
\end{aligned}$$

The additional rightmost coefficients required can be stored beforehand, just like the eigenvalues, after which a simple bisection method in the interval $\left[0, \frac{-t}{\lambda^{min}}\right]$ yields the line search coefficient.

Outer iterations and overall performance

We have chosen to update the Lagrange multiplier X and the penalty parameter t simultaneously in each outer iteration. Alternate updating schemes were less successful and more difficult to adapt to work for all cases. At the end of each outer iteration, the Lagrange multiplier is updated to satisfy $\nabla f(\Gamma^{(k+1)}, X^{(k)}, t^{(k)}) = \nabla\mathcal{L}(\Gamma^{(k+1)}, X^{(k+1)}) = 0$, which is fulfilled by the

update

$$X^{(k+1)} = \left(\frac{1}{t^{(k)}} L(\Gamma^{(k+1)}) + I\right)^{-1} X^{(k)} \left(\frac{1}{t^{(k)}} L(\Gamma^{(k+1)}) + I\right)^{-1}$$

Clearly, this update preserves the symmetry and positive semidefiniteness of X. After updating the multiplier X, the penalty parameter t is reduced by a constant factor (figure 3.14), unless reducing it would result in infeasibility, that is, $\frac{1}{t} L(\Gamma) \succeq -I$ would fail to hold. In practice, however, this rarely happens when t is not updated too aggressively.

The modified barrier method converges to the optimum from outside the feasible set, because it replaces the original constraint $L(\Gamma) \succeq 0$ by $\frac{1}{t} L(\Gamma) \succeq -I$. Moreover, it does not require a starting point $\Gamma^{(0)}$ that satisfies $L(\Gamma^{(0)}) \succ 0$ since the initial value of t can be adjusted. Because it converges to the optimum from outside the feasible region, it will have residual negative eigenvalues upon convergence.

The modified barrier method scales similarly to the classical barrier method, at least $O(K^6)$, with the dimension of the sp-basis. The modified barrier algorithm has the same structure as the classical barrier algorithm, essentially a four-fold loop in which each level requires the following number of rate-determining operations

- iterate over t, the barrier parameter, to minimize $f(\Gamma, t)$ by Newton's method: $\sim O(1)$ iterations
- iterate Newton's method until a direction Δ is found that solves $\nabla^2 f \Delta = -\nabla f$: $\sim O(1)$ iterations
- iterate the Krylov subspace method chosen to calculate each step in Newton's method: $O(1)$ to $O(K^2)$ iterations
- compose the Hessian-vector product needed in every Krylov subspace iteration: $O(K^6)$ flops

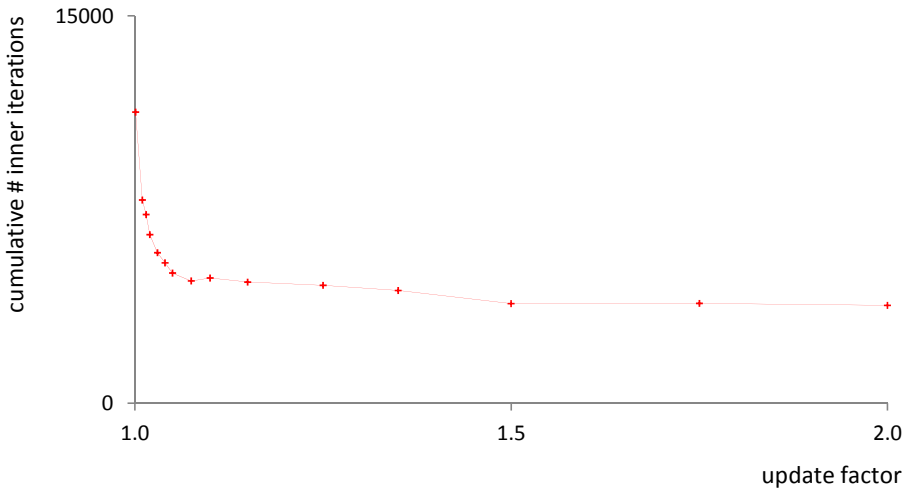

Figure 3.14: The cumulative number of inner iterations needed in the MB method to converge for the LiH (STO-6G) test system is smallest for an update factor for the penalty parameter around 1.075 and for bigger factors. However, because choosing the update factor too big (bigger than 3.0 for this system) results in infeasibility of the constraint $\frac{1}{t}L(\Gamma) \succeq -I$ upon update, we have used an update factor of 1.075 throughout, unless specified otherwise.

The outer iterations are stopped when the energy remains constant upon updating and complementary slackness holds to good precision, $tr\,[XL(\Gamma)] < \epsilon$. The total number of Newton's iterations this requires is roughly the same for different systems. Just like in the CB method, the cost of solving Newton's equations is dominated by assembling the Hessian in each step of the Krylov subspace method, and is therefore at least $O(K^6)$, but with a smaller scaling factor than the CB method due to its faster than linear convergence of the duality gap (figure 3.15).

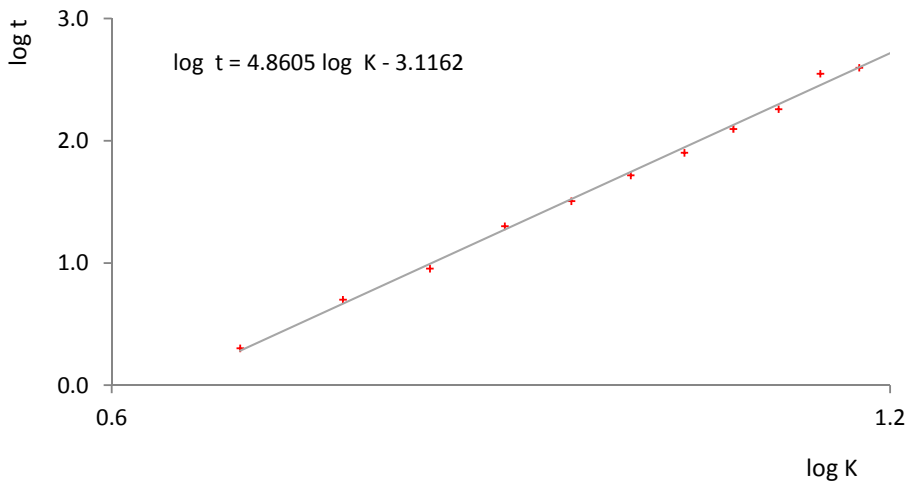

Figure 3.15: The CPU times required by the MB algorithm to optimize the energy of half-filled Hubbard models, with an equal number of spatial orbitals $\frac{K}{2}$ and particles N, $\frac{K}{2} = N$, and interaction strength 1.0, suggests that the algorithm in practice scales a bit better than $O(K^6)$ with the dimension of the sp basis.

3.6 Primal-dual interior point method

3.6.1 Theoretical background

The primal-dual interior point path-following method is in spirit similar to the dual barrier method, as both aim to follow the central path approximately to reach the optimum from within the feasible set. However, whereas the dual barrier method only provides a primal feasible matrix in terms of the dual matrix after each inner minimization, the primal-dual method optimizes the dual and primal problem simultaneously and independently. This allows it to reduce the duality gap in a much faster way than the dual barrier method. A path-following primal-dual method attempts to solve the centrality conditions for both $Z \equiv L(\Gamma)$ and X such that $ZX = tI$ and subsequently reduces t to move closer to the optimum. The 'predictor-corrector' method is a particular instance of this method, which alternates 'corrector' centering steps that force the iterates X and Z to stay close to the central path with 'predictor' steps that reduce the duality gap.

The primal-dual centering equations are overdetermined and nonlinear, hence in practice several different ways to linearize and symmetrize them have been put forward.[86] The primal-dual centering equations that need to be solved are

$$
\begin{aligned}
&tr\ XL(F^i) = tr\ HF^i \quad i = 1, \ldots, n \\
&Z = L(\Gamma) \succeq 0 \\
&X \succeq 0 \\
&XZ = tI
\end{aligned}
$$

where an additional variable $Z \equiv L(\Gamma)$ is used for the dual constraint matrix. The equalities and positive semidefinite inequalities are easily imposed starting from strictly feasible variables. The nonlinear centrality condition, however, is difficult to impose during the optimization. The centrality condition requires

that the updates ΔX and ΔZ satisfy

$$(X + \Delta X)(Z + \Delta Z) = tI$$

which needs to be linearized and symmetrized in order to obtain the updates ΔX and ΔZ by means of a Newton-Raphson approach. Different ways of doing this lead to different directions, among them the frequently used Nesterov-Todd direction.

The Nesterov-Todd direction makes the primal and dual direction symmetric, by applying a scaling matrix D that can be considered the *metric geometric mean* of X and Z^{-1}

$$D = Z^{-1/2}(Z^{1/2}XZ^{1/2})^{1/2}Z^{-1/2}$$

such that

$$D^{-1/2}XD^{-1/2} = D^{1/2}ZD^{1/2} \equiv V$$

Applying this scaling matrix to the centrality condition, the scaled update directions $\Delta V_X = D^{-1/2}\Delta X D^{-1/2}$ and $\Delta V_Z = D^{1/2}\Delta Z D^{1/2}$ must satisfy

$$\begin{aligned} &D^{-1/2}(X + \Delta X)D^{-1/2}D^{1/2}(Z + \Delta Z)D^{1/2} = tI \\ &(V + \Delta V_X)(V + \Delta V_Z) = tI \end{aligned}$$

This expression is linearized by neglecting the second order terms

$$\Delta V_Z V + V\Delta V_X + V^2 = tI$$

and symmetrized

$$\frac{1}{2}(\Delta V_Z + \Delta V_X)V + \frac{1}{2}V(\Delta V_X + \Delta V_Z) = tI - V^2$$

The symmetrical scaled directions $\Delta V_X,\ \Delta V_Z$ that satisfy this requirement are

$$\Delta V_X + \Delta V_Z = tV^{-1} - V$$

Equivalently, the unscaled updates $\Delta X,\ \Delta Z$ satisfy

$$\Delta X + D\Delta Z D = tZ^{-1} - X \tag{3.39}$$

which determines the Nesterov-Todd direction.

Primal-dual methods differ mostly in the update scheme for the parameter t involved in the centering steps and the type of line search to calculate the step length in the obtained Newton direction. Predictor-corrector methods alternate predictor steps that aim to reduce the duality gap with corrector steps that ensure that the iterates stay sufficiently close to the central path (algorithm 3.40). The predictor steps solve the standard Newton equations for KKT-optimality of the primal and dual variable (3.39 with $t = 0$) to reduce the duality gap. The primal and dual variable are then updated as $Z \equiv Z + \alpha\Delta Z,\ X \equiv X + \alpha\Delta X$, where the line search coefficient α is chosen to ensure that the updated variables remain feasible and stay close enough to the central path. The step length therefore also determines the decrease in duality gap upon update. The subsequent corrector step then solves the primal and dual centering equations (3.39) to ensure that the primal and dual variable stay close enough to the central path, but does not change the duality gap. Ensuring that they stay in a neighborhood of the central path guarantees that the following predictor step can bring about a substantial reduction of the duality gap. The predictor-corrector steps are alternated in this way until the primal-dual gap reaches a desired accuracy.

Algorithm

X, Z strictly feasible, sufficiently close to the central path

$t = tr\ XZ$

do while $tr\ XZ > \epsilon$

corrector step

solve for $\Delta X,\ \Delta Z$:

$\Delta X + D\Delta Z D = tZ^{-1} - X$

update $X, Z:\ X = X + \Delta X,\ Z = Z + \Delta Z$ *predictor step*

solve for $\Delta X,\ \Delta Z$:

$\Delta X + D\Delta Z D = -X$

choose $0 \leq \alpha \leq 1$

update $X, Z:\ X = X + \alpha\Delta X,\ Z = Z + \alpha\Delta Z$

calculate duality gap: $t = tr\ XZ$

end do (3.40)

3.6.2 Implementation of a primal-dual interior point method

Just like the classical barrier method, the primal-dual predictor-corrector method follows the central path to the optimum. But unlike the classical barrier method, it simultaneously optimizes the primal and the dual problem. It can therefore make much faster progress towards the optimum, but it comes with a price: the primal-dual method is significantly more expensive.

The implementation of a primal-dual (PD) predictor-corrector method in C++ that was used here was provided by our co-workers Brecht Verstichel and Ward Poelmans.[99,100] It was compiled and run on the same computer and linked to the same Lapack and BLAS libraries as the barrier methods considered before.

Inner iterations

Both the inner iterations of the predictor and the corrector steps can be solved by the conjugate gradients method, which only requires a linear map of the update vector in each inner iteration. Because the primal and dual direction ΔX and ΔZ are orthogonal, projecting the primal-dual equation (3.39) onto their respective spaces gives a separate expression for both directions.

Since the primal variable X must satisfy $tr\ XL(F^i) = tr\ HF^i$, the primal

direction ΔX must satisfy

$$tr\ \Delta X L(F^i) = 0$$

Projecting (3.39) onto the basis $\{L(F^i)\}$ therefore separates out the dependence on the dual direction ΔZ:

$$\begin{aligned}
&tr\ [D\Delta Z D L(F^i)] = t\ tr\ [Z^{-1}L(F^i)] - tr[XL(F^i)] \\
&tr\ [DL(\Delta\Gamma)DL(F^i)] = t\ tr\ [Z^{-1}L(F^i)] - tr[HF^i] \\
&\mathcal{P}_{\perp F^0}(L^\dagger(DL(\Delta\Gamma)D)) = t\ \mathcal{P}_{\perp F^0}(L^\dagger(Z^{-1})) - \mathcal{P}_{\perp F^0}(H)
\end{aligned} \tag{3.41}$$

where the projection $\mathcal{P}_{\perp L(F^0)}$ projects onto the traceless space spanned by $\{F^i\}$.

Since $Z = L(\Gamma) = L(F^0) + \sum_{i=1}^n L(F^i)$, the dual direction can be expanded in the basis of matrices $L(F^i)$

$$Z = L(\Delta\Gamma) = \sum_{i=1}^n \Delta\Gamma_i L(F^i)$$

Suppose the orthogonal complement of the space spanned by $L(F^0)$ and $\{L(F^i)\}$ is spanned by an orthonormal basis $\{C^i\}$, such that the space of all block diagonal matrices of the same dimension as X is spanned by the bases $\{L(F^0)\} \cup \{L(F^i)\} \cup \{C^i\}$. Since $Z = L(F^0) + \sum_i \Gamma_i L(F^i)$ and $\Delta Z = \sum_i \Delta\Gamma_i L(F^i)$ lie in the orthogonal complement of the space spanned by $\{C^i\}$, projecting the primal-dual centering equation (3.39) onto the basis $\{C^i\}$ gives

$$\begin{aligned}
tr\ [D^{-1}\Delta X D^{-1} C^i] &= t\ tr\ [D^{-1}Z^{-1}D^{-1}C^i] - tr\ [D^{-1}XD^{-1}C^i] \\
&= t\ tr\ [X^{-1}C^i] - tr\ [ZC^i] \\
&= t\ tr\ [X^{-1}C^i] \\
\mathcal{P}_{\{C^i\}}\left(D^{-1}\Delta X D^{-1}\right) &= t\ \mathcal{P}_{\{C^i\}}\left(X^{-1}\right)
\end{aligned} \tag{3.42}$$

where the projection $\mathcal{P}_{\{C^i\}}$ onto the orthogonal complement of the space spanned by $\{L(F^0)\} \cup \{L(F^i)\}$ is done by projecting out the part of the matrix that lies in the space spanned by $\{L(F^0)\} \cup \{L(F^i)\}$. Using this mapping requires $O(K^6)$ floating point operations because of the matrix square root and matrix multiplications involved.

Theoretically, solving (3.41) for the dual $\Delta\Gamma$ would give a primal matrix ΔX by substituting it in (3.39). However, this did not provide the desired accuracy for the dual step in our applications. Instead, the resulting ΔX can be used as the initial point for solving the dual equations by conjugate gradients. For this reason, solving the dual equations takes far fewer inner CG iterations (figure 3.16).

To ensure that the updated primal and dual matrices stay feasible and close enough to the central path, a bisection line search is performed. Similarly to the CB method, it uses a logarithmic potential function

$$\begin{aligned}\Phi(X,Z) &= -ln\ det\ XZ\ + ln\ det\ \left(\frac{tr\ [XZ]}{dim(I)}I\right) \\ &= -ln\ det\ XZ + dim(I) ln\ \left(\frac{tr\ [XZ]}{dim(I)}\right)\end{aligned} \tag{3.43}$$

to penalize infeasibility of the updates $X + \alpha\Delta X$ and $Z + \Delta Z$. It considers the difference between the current primal-dual pair and a primal-dual pair with the same duality gap that lies on the central path. Limiting this difference therefore constrains the updates to lie in a neighborhood of the central path, such that the subsequent centering steps will not take too many iterations. Adjusting the size of the neighborhood of the central path will thus adjust the balance between the number of predictor and corrector iterations. In our implementation, the distance from the central path upon update, measured by $\Phi(X + \alpha\Delta X, Z + \alpha\Delta Z)$, is limited to 2.0. This choice of deviation from the central path makes the subsequent predictor and corrector iterations about equally time-consuming to compute (figure 3.16).

Outer iterations and overall performance

Although the inner iterations in the primal-dual method are more expensive than in the classical barrier method, the primal-dual method needs fewer outer iterations to converge to an energy that is ϵ-suboptimal, measured by the duality gap (figure 3.17). Because it optimizes both the primal and the dual variable,

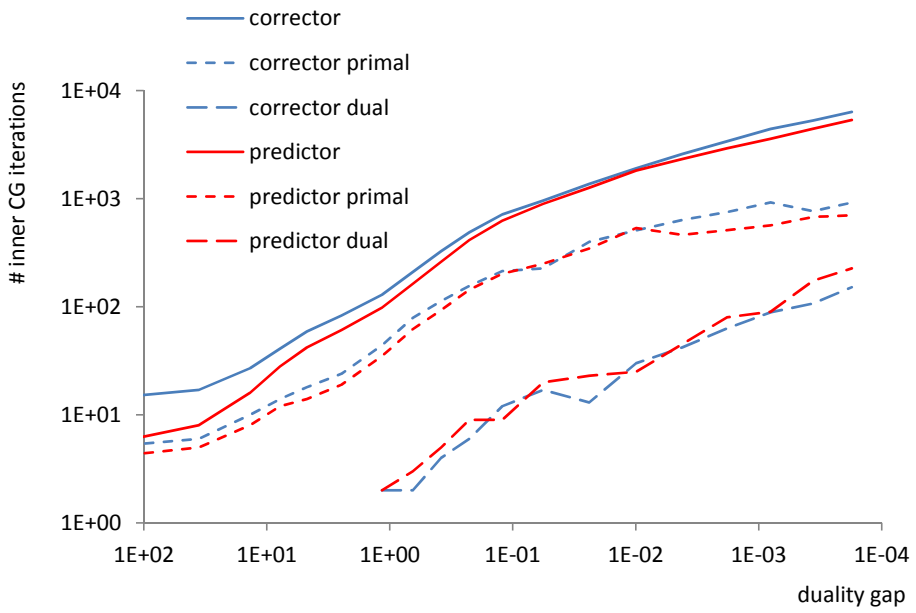

Figure 3.16: The allowed deviation from the central path upon update with the predictor direction is limited to 2.0 (3.43), such that the predictor and corrector steps take a similar number of inner iterations. The primal Newton direction is used to approximate the initial dual direction in the Krylov subspace method, such that solving the dual equations only takes a small number of iterations compared to the primal equations.

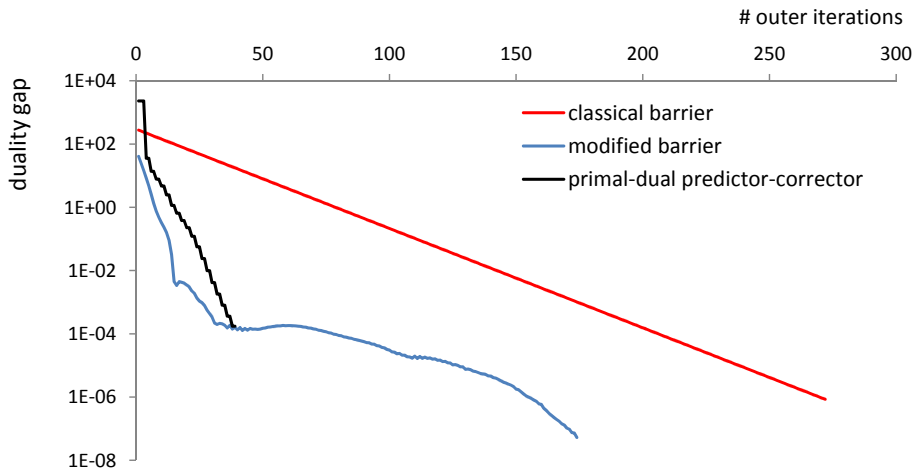

Figure 3.17: The primal-dual gap decreases considerably faster with the number of outer iterations performed in the PD method than in the CB method. The decrease of the duality gap is only linear in the CB method, but the PD method is able to reduce it faster than linearly as it simultaneously optimizes the primal and the dual problem. Surprisingly, the MB method decreases the duality gap at a similar rate.

its progress towards the optimum is much more aggressive than in the barrier method.

In spite of its small number of outer iterations to reach the optimum, explicitly optimizing both the primal and the dual matrix is the most expensive way of solving the optimization problem. Optimizing both the primal and the dual makes it possible to approach the optimal energy both from below and above, yielding an error estimate at all times, but requires more $O(K^6)$ floating-point operations in each inner iteration to do so than the barrier methods because the primal matrix for the P-,Q- and G-condition has three blocks of similar dimension to the 2DM. So even though the use of a mapping for the Hessian-vector product reduces its scaling to $O(K^6)$ compared to $O(K^{12})$ for a factorization of the Hessian it is considerably more expensive than the other methods considered (table 3.2).

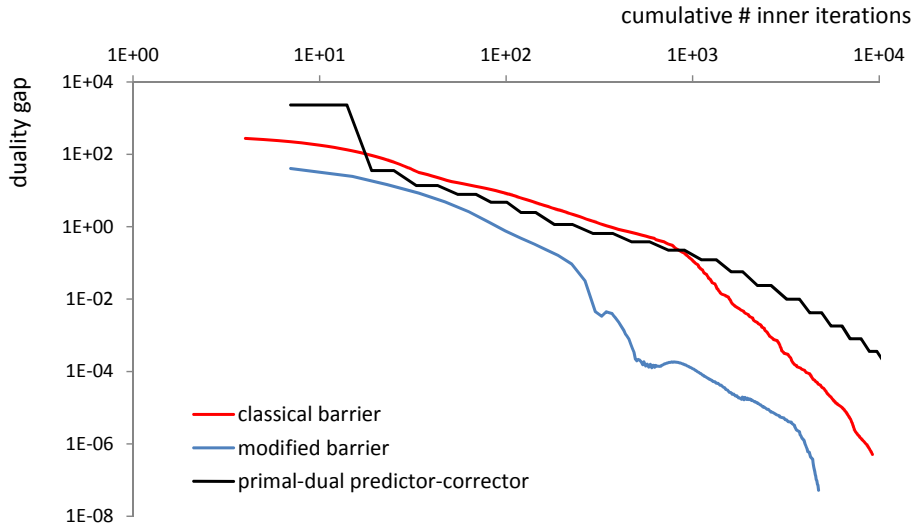

Figure 3.18: Even though the duality gap decreases much faster with the number of outer iterations in the PD method than in the CB method, the PD method needs many more Krylov subspace iterations to solve Newton's equations than the CB and MB method (figure 3.19). As a consequence, the overall number of inner Krylov subspace iterations performed at convergence is bigger for the PD method.

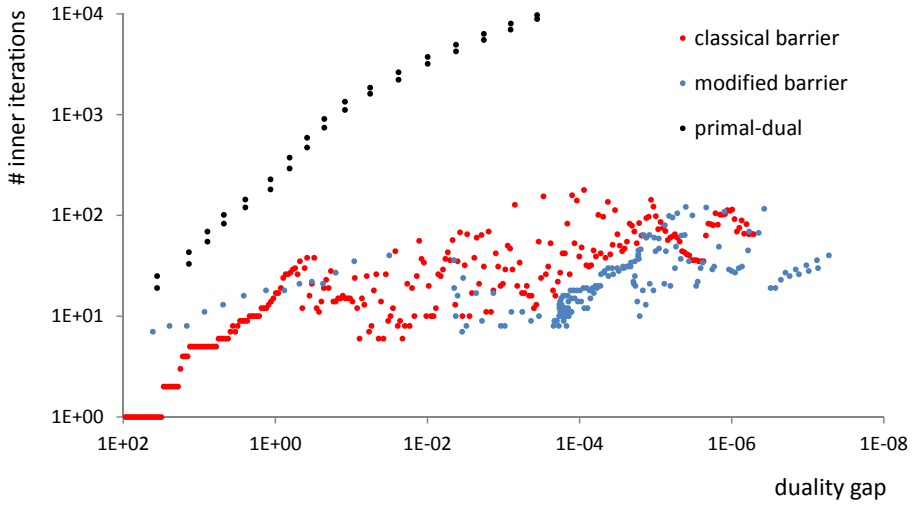

Figure 3.19: The primal-dual method needs considerably more inner Krylov subspace iterations than the CB and MB method to solve Newton's equations resulting in the duality gap shown on the horizontal axis.

3.7 Boundary point method

3.7.1 Theoretical background

In contrast to the interior point methods discussed before, the boundary point method is a zeroth order method: it does not require a gradient or Hessian, it only solves a linear system of equations in its inner iterations. Its application to v2DM theory is particularly interesting because the inner equations can be solved exactly, without needing an iterative solver.

In contrast to primal-dual interior point methods, the boundary point method

aims to converge to an optimal pair X, Z that satisfies

$$tr\ XL(F^i) = tr\ HF^i \quad i = 1, \ldots, n \tag{3.44}$$

$$Z = L(\Gamma) \tag{3.45}$$

$$Z \succeq 0 \tag{3.46}$$

$$X \succeq 0 \tag{3.47}$$

$$XZ = 0 \tag{3.48}$$

by optimizing a primal variable $X \succeq 0$ and dual variable $Z \succeq 0$ that satisfy $ZX = 0$ at any time, but only become primal and dual feasible (3.45 and 3.44) upon convergence to the optimum. The inner minimization over Γ yields a primal feasible $X \succeq 0$ and a matrix $Z \succeq 0$ that satisfies $XZ = 0$ but is not dual feasible, as it does not satisfy $Z = L(\Gamma)$. Optimality is reached for the outer loops when Z becomes dual feasible up to a small deviation. In contrast to interior-point methods, the boundary point method convergences to the optimum from the boundary of the positive semidefinite cone for the primal and the dual variable until they both become feasible upon convergence. Hence its name 'boundary point method'.

The boundary point method is a particular instance of an augmented Lagrangian method.[101,102] It considers an augmented Lagrangian for the dual problem

$$\mathcal{L}(\Gamma, Z, X) = tr\ H\Gamma + tr[X(Z - L(\Gamma))] + \frac{\sigma}{2} \|Z - L(\Gamma)\|^2$$

where X is the Lagrange multiplier. By defining a matrix W

$$W \equiv L(\Gamma) - \frac{1}{\sigma} X$$

the augmented Lagrangian can be written as

$$\mathcal{L}(\Gamma, Z, X) = tr\ H\Gamma + \frac{\sigma}{2} \|Z - W(\Gamma)\|^2 - \frac{1}{2\sigma} \|X\|^2$$

The augmented Lagrangian approach minimizes $f(\Gamma, Z)$ defined as

$$f(\Gamma, Z) \equiv tr\ H\Gamma + \frac{\sigma}{2} \|Z - W(\Gamma)\|^2$$

in its inner iterations to obtain Γ and $Z \succeq 0$, and then updates the Lagrange multiplier X according to

$$X \equiv X + \sigma(Z - L(\Gamma))$$

The inner minimization in the augmented Lagrangian method

$$\underbrace{min}_{\Gamma,Z} f(\Gamma, Z) = \underbrace{min}_{\Gamma,Z}\ tr\ H\Gamma + \frac{\sigma}{2} \|Z - W(\Gamma)\|^2$$

corresponds to a primal problem with primal variable V and Lagrangian $\mathcal{L}'(\Gamma, Z, V)$

$$\mathcal{L}'(\Gamma, Z, V) = tr\ H\Gamma + \frac{\sigma}{2} \|Z - W(\Gamma)\|^2 - tr\ ZV$$

for which the KKT-conditions for optimality are

$$\nabla_\Gamma \mathcal{L}'(\Gamma, Z, V) = \mathcal{P}_{\perp F^0}(H) - \sigma \mathcal{P}_{\perp F^0}\Big(L^\dagger(Z - W(\Gamma))\Big) = 0 \tag{3.49}$$

$$\nabla_Z \mathcal{L}'(\Gamma, Z, V) = \sigma(Z - W(\Gamma)) - V = 0 \tag{3.50}$$

$$V \succeq 0$$

$$Z \succeq 0$$

$$VZ = 0$$

The minimization over both Γ and Z can be uncoupled by alternating minimization of $f(\Gamma, Z)$ over Γ and updating $Z(\Gamma)$ with the obtained Γ. The matrix Z that minimizes $f(\Gamma, Z)$ for fixed Γ is simply $W(\Gamma)_+$,

$$\underbrace{argmin}_{Z \succeq 0} \Big(f(\Gamma, Z) = \quad tr\ H\Gamma + \frac{\sigma}{2} \|Z - W(\Gamma)\|^2\Big) = W(\Gamma)_+,$$

which is the positive semidefinite part of $W(\Gamma)$, as it can be decomposed into a positive semidefinite part and a negative definite part $W(\Gamma) = W(\Gamma)_+ + W(\Gamma)_-$, for example by separating its eigenvalue decomposition. From the KKT-conditions (3.50), an expression for the Lagrange multiplier V in terms of $W(\Gamma)_-$ follows

$$V = \sigma(W(\Gamma)_+ - W(\Gamma)) = -\sigma W(\Gamma)_- \succeq 0$$

which satisfies $VZ = 0$.
The only KKT-condition for optimality that is not satisfied yet is $\nabla_\Gamma \mathcal{L}' = 0$ (3.49).

The inner iterations will thus consist of iteratively solving $\nabla_\Gamma \mathcal{L}' = 0$ (3.49) to obtain Γ, and updating Z and V in terms of the obtained $W(\Gamma)$. The update of the Lagrange multiplier for the augmented Lagrangian of the outer iteration is also determined by $W(\Gamma)$:

$$\begin{aligned} X &\equiv X + \sigma\,(Z - L(\Gamma)) \\ &= X + \sigma\left(Z - W(\Gamma) - \frac{1}{\sigma}X\right) \\ &= \sigma(Z - W(\Gamma)) \\ &= V \\ &= -\sigma W(\Gamma)_- \end{aligned} \tag{3.51}$$

which is clearly positive semidefinite and satisfies $XZ = -\sigma W(\Gamma)_- W(\Gamma)_+ = 0$. Each inner iteration therefore yields a matrix $X = V$, which becomes primal feasible if

$$tr[VL(F^i)] = tr\ HF^i$$

The inner iterations are therefore stopped when this criterion is satisfied to suitable precision. The matrix X is then updated as $X \equiv V$ in the outer loop, after which the inner minimizations are repeated. The outer iterations are stopped when the matrix Z is nearly dual feasible, $Z \approx L(F^0) + \sum_i \Gamma_i L(F^i)$.

Algorithm

do while $\|Z - L(\Gamma)\| \geq \epsilon$
  do while $\left\|\mathcal{P}_{\perp F^0}(L^\dagger(V)) - \mathcal{P}_{\perp F^0}(H)\right\| \geq \sigma\epsilon$
    solve for Γ : $\mathcal{P}_{\perp F^0}(H) - \sigma\mathcal{P}_{\perp F^0}\Big(L^\dagger(Z - W(\Gamma))\Big) = 0$

$$
\begin{aligned}
&W \equiv L(\Gamma) - \frac{1}{\sigma} X \\
&Z \equiv W_+ \\
&V \equiv -\sigma W_- \\
&\text{end do} \\
&X = V \\
&\text{end do}
\end{aligned}
\tag{3.52}
$$

3.7.2 Implementation of a boundary point method

The boundary point method may be an alternative to the computationally expensive second-order methods discussed in previous sections. It is a zeroth order method, as it does not require a gradient or Hessian in its optimization. This property is both its strength and its weakness. On the one hand, it makes its inner iterations much easier to solve. On the other hand, it makes slower progress towards the optimum in each outer iteration than second-order methods like the interior point methods discussed before. Since solving the inner iterations is the major obstacle for second order methods, this algorithm may be a faster alternative. Our calculations suggest that it is faster than the barrier method for most systems, but its slow convergence near the optimal energy can remove its advantage over the barrier method when the parameters involved in the algorithm are not carefully optimized.

The C++ implementation of a boundary point (BP) method used here to study its performance on molecular systems was provided by our co-workers Brecht Verstichel and Ward Poelmans and is based on the algorithms of Povh and Rendl et al.[101,102] A reference wavefunction-based CCSD routine to compare with is provided by the GAMESS package[103] and compiled and run on the same computer as the semidefinite optimization programs.

Inner iterations

The strength of the boundary point algorithm for applications to v2DM theory lies in the ease of solving its inner iterations (3.49). The inner system of equations

$$\begin{aligned}
&\nabla_\Gamma \mathcal{L}'(\Gamma, Z, X) = 0 \\
&\mathcal{P}_{\perp F^0}\Big(\sigma L^\dagger\left(L(\Gamma)\right)\Big) = \mathcal{P}_{\perp F^0}\Big(\sigma L^\dagger(Z) + L^\dagger(X) - H\Big) \\
&\Gamma = \Big(L^\dagger(L)\Big)^{-1}\left(\mathcal{P}_{\perp F^0}\left(L^\dagger(Z) + \frac{1}{\sigma}(L^\dagger(X) - H)\right)\right)
\end{aligned} \tag{3.53}$$

can be solved exactly because the inverse map $\Big(L^\dagger(L)\Big)^{-1}$ can be expressed analytically. The squared P-,Q- and G-map take the same form as the regular maps, but with different normalization coefficients. The sum of any combination of these maps has a straightforward inverse map (chapter 1, section 1.3.3).

In our implementation, the number of inner iterations is limited to one, such that the resulting primal variable is not exactly primal feasible after each inner minimization. However, in the convergence limit, both the primal and the dual variable become feasible (figure 3.20). In order to increase the sensitivity to primal infeasibility, an additional parameter τ is introduced in the inner equations (3.53), following Mazziotti's implementation,[102] which is set to 1.6 for all calculations performed here. So instead of (3.53), the equation

$$\Gamma = \Big(L^\dagger(L)\Big)^{-1}\Big(\mathcal{P}_{\perp F^0}\left(L^\dagger(Z) + \frac{\tau}{\sigma}(L^\dagger(X) - H)\right)\Big)$$

is solved.

Outer iterations and overall performance

The boundary point method needs many more outer iterations to converge to the optimum than any of the second-order methods discussed before, as it does not use any information about the system's gradient or Hessian. Consequently, as the method approaches the optimal energy, subsequent outer iterations do

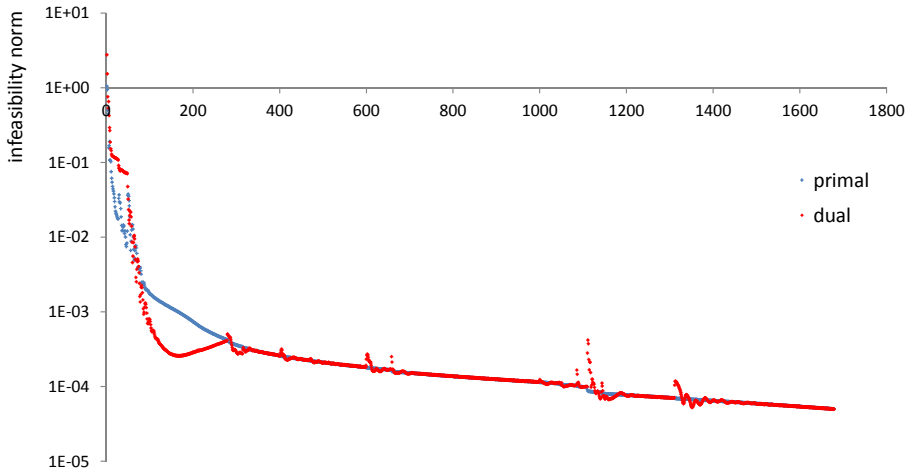

Figure 3.20: The convergence to a primal and dual feasible point with the cumulative number of outer iterations is very slow. The primal and dual infeasibility norm converge to zero at a very similar rate.

not decrease the energy as significantly as in the CB method. However, its inner iterations are much faster to solve than the inner Newton equations in the second order methods, but more expensive than each inner CG iterations involved in solving Newton's equations in the second order methods. This is because the boundary point method's inner iterations require a diagonalization of the matrix $W(\Gamma)$ in order to split it into a positive semidefinite and negative definite part, on top of solving the linear system. Therefore it scales at least as $O(K^6)$ with the sp basis dimension K, if the number of outer iterations is considered more or less constant. Although this is similar to the way the CB method scales with the dimension, it turns out to be faster than the CB method for most systems (table 3.2).

The main difficulty in practical implementations of the boundary point method is therefore to find an update scheme for the parameter σ that minimizes the number of outer iterations required to obtain near-feasibility of the primal and dual variable. Choosing the update factor σ either too large or too small will result in an increase in CPU time (figure 3.21).

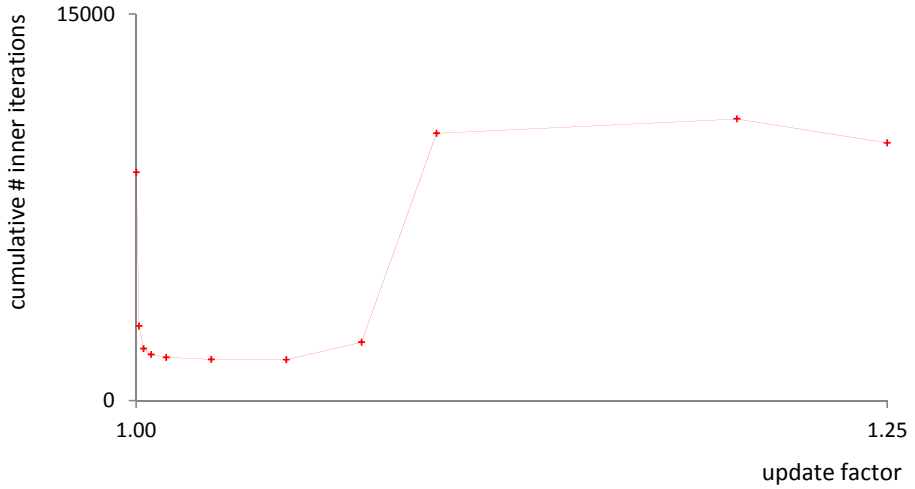

Figure 3.21: The cumulative number of inner iterations needed by the boundary point method to converge for the LiH STO-6G test system is smallest for an update factor close to 1 for the parameter σ that determines the strength of the penalty (algorithm 3.52). For this reason, we have used an update factor of 1.01 in all applications of the boundary point method throughout this chapter.

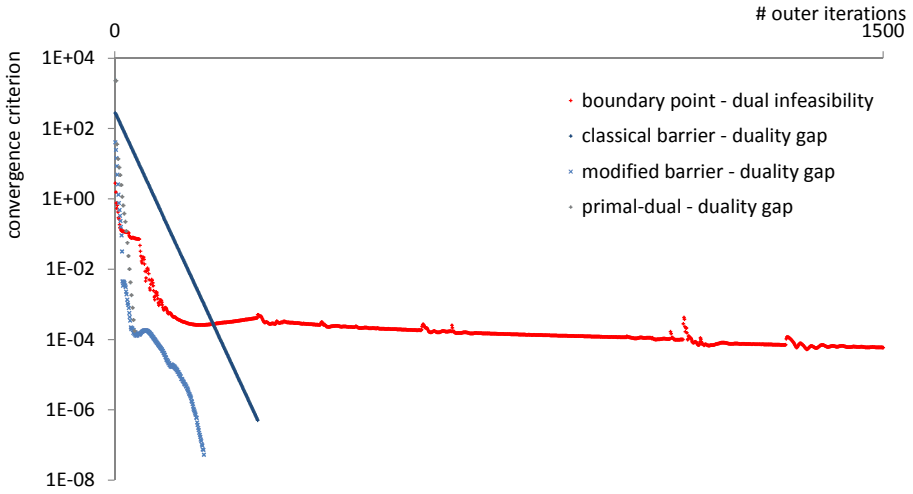

Figure 3.22: The convergence criterion for the second order methods is the duality gap. The convergence criterion for the boundary point method is the dual infeasibility, as measured by its Frobenius norm. The duality gap decreases much faster with the number of outer iterations in second order methods than the dual infeasibility decreases in the zeroth order boundary point method. The number of inner iterations was limited to 1 in the BP method, therefore the total number of inner iterations performed equals the number of outer iterations.

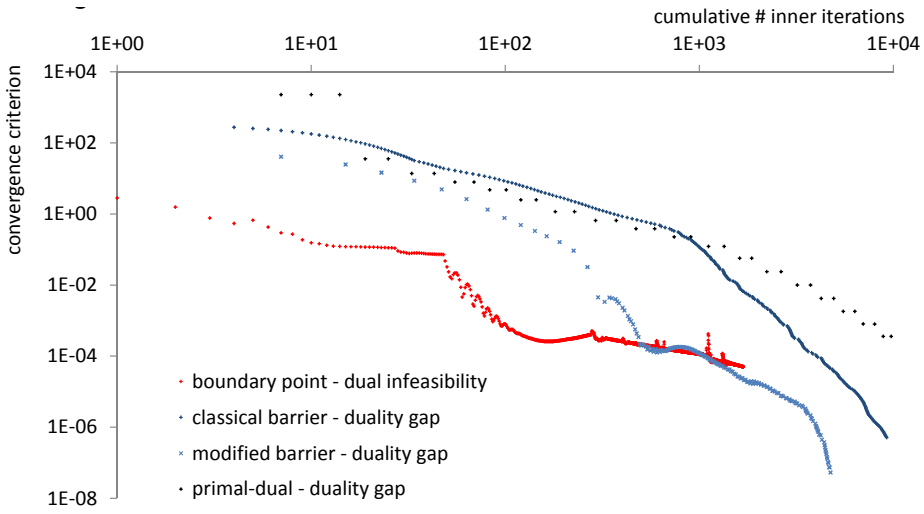

Figure 3.23: Even though the second-order methods reduce the duality gap faster with the number of outer iterations than the zeroth order boundary point method reduces the primal and dual feasibility, they need many more inner Krylov subspace iterations to solve Newton's equations. As a consequence, the second-order methods converge much more slowly in terms of the cumulative number of inner iterations. Nonetheless, their inner iterations are more rapidly solvable than each BP method's inner iteration, such that the two approaches still lead to CPU times of the same order of magnitude (table 3.2).

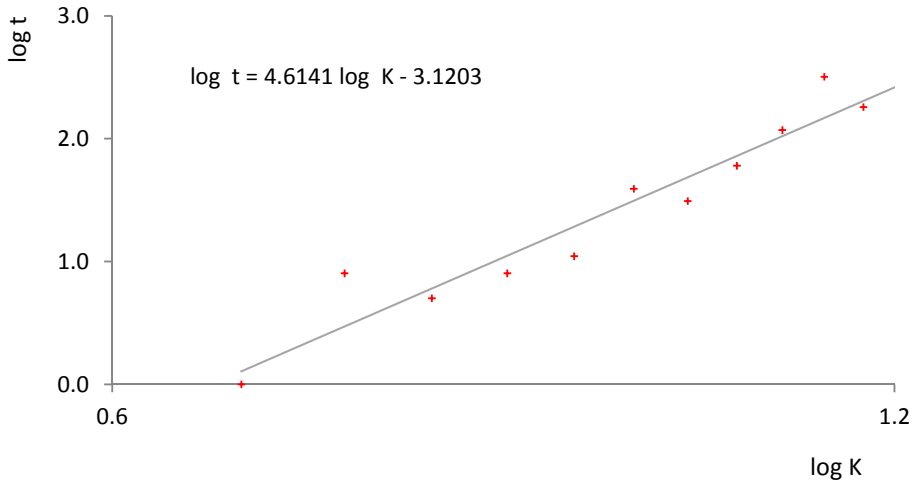

Figure 3.24: The CPU times, t, required by the MB algorithm to optimize the energy of half-filled Hubbard models, with an equal number of spatial orbitals $\frac{K}{2}$ and particles N, $\frac{K}{2} = N$ and interaction strength 1.0, suggests that the boundary point algorithm in practice scales even better than $O(K^6)$ with the dimension of the sp basis.

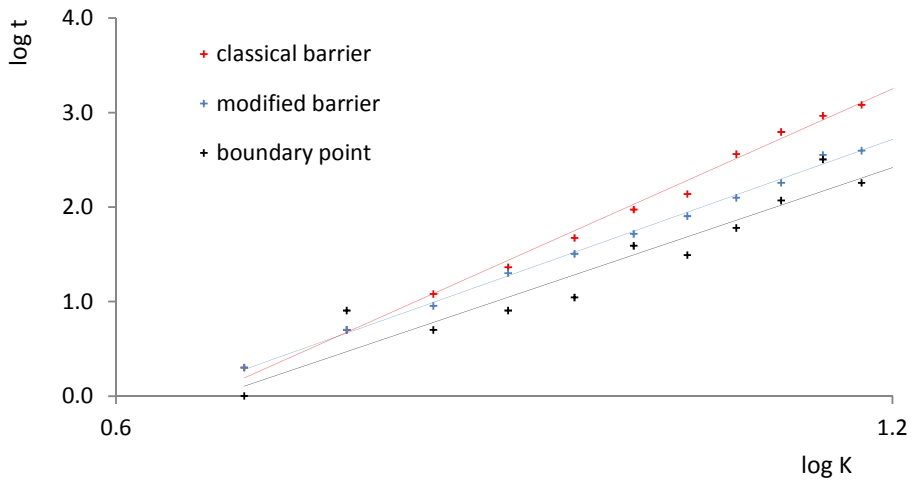

Figure 3.25: The CPU times required by the MB method and BP method to optimize the energy of half-filled Hubbard models, with an equal number of spatial orbitals $\frac{K}{2}$ and particles N, $\frac{K}{2} = N$ and interaction strength 1.0, grow less fast with the basis set dimension than those required by the CB method. The BP method calculates these models faster than the barrier methods, although its superiority on molecular systems is less pronounced (table 3.2).

	STO-6G			D95V		
	MB	PD	BP	MB	BP	CCSD
BH	1.40	0.41	1.17	1.44	0.61	61.35
C_2H_2	1.68	0.12	0.44	3.60	1.88	5660.68
C_2H_4	3.26	0.14	0.68	4.27	2.43	12459.11
CH_2O	4.82	0.22	1.84	5.65	2.01	14216.12
CH_2	1.69	0.42	0.61	2.43	1.01	126.33
CH_4	1.63	0.30	0.59	1.75	1.20	661.69
CH	2.25	0.60	2.25	2.46	0.58	80.18
CO	3.50	0.17	3.24	5.81	1.75	3720.00
F_2	10.83	0.26	5.62	4.05	0.57	2992.18
H_2O_2	5.59	0.16	11.41	4.60	1.14	7724.03
H_2O	3.08	0.43	4.44	3.02	0.53	297.09
HCN	3.72	0.21	1.04	4.32	1.47	4020.82
HF	4.00	0.73	4.00	2.55	0.51	105.20
HNC	2.49	0.17	0.51	5.32	1.95	5018.79
HNO	6.35	0.11	4.87	3.49	0.74	3767.17
HOF	5.42	0.10	7.57	6.29	1.20	7433.02
Li_2	1.68	0.18	1.01	2.15	1.58	858.62
LiF	6.68	0.29	6.00	3.84	0.55	2643.11
LiH	1.40	1.17	1.17	1.79	1.37	50.15
N_2H_2	1.80	0.09	0.27	4.92	1.65	7132.11
NH_3	1.33	0.21	2.50	2.72	1.20	720.06
NH	1.63	0.81	6.50	3.71	1.22	124.62

Table 3.2: The speed-up of our implementations of several semidefinite programming algorithms compared to the classical log-barrier method, measured by the inverse ratio of their wall clock times, t_{CB}/t, shows that the modified barrier method (MB) gives the biggest speed-up for most of the systems considered. The boundary point (BP) method also gives a substantial speed-up compared to the CB method for most, but not all, systems. The primal-dual (PD) predictor-corrector method is more accurate but significantly slower. Hence, we have only applied it to the STO-6G basis set. The molecules considered were at their experimental equilibrium geometries.[104]

3.8 Conclusions: Comparison of selected algorithms

None of the semidefinite algorithms that we have applied to the v2DM(PQG) method approaches the speed of a wavefunction-based method of the same 'level' of theory, such as CCSD. The CCSD method provides a similar level of theory to the v2DM(PQG) method as it is also exact for to 2-electron systems. Just like the v2DM(PQG) algorithms considered, it scales as $O(K^6)$ with the dimension of the sp basis. But even when we take into consideration that our semidefinite programs are not fully optimized, their computation times clearly have a much larger prefactor than the CCSD method (table 3.2 includes CCSD computation times in the D95V basis set).

The second-order methods' strength lies in their substantial progress towards the optimal energy in each outer iteration; their weakness is the difficulty of solving the inner Newton's equations. Although our primal-dual program is by far the slowest, a thoroughly optimized implementation may be very powerful,[29,105] as it is able to make faster than linear progress towards the optimum. The classical barrier method only allows linear convergence of the duality gap with the outer iteration, but it is the most robust. It may be slow, but it will converge, and can be considered a 'black box' approach to the problem. The modified barrier method combines the advantages of the barrier method with the faster than linear convergence of a primal-dual approach. Unfortunately, it is not as robust. A suitable line search procedure, however, may avoid infeasibility upon update of the Lagrange multiplier or penalty parameter.

The zeroth-order boundary point method's strength lies in its rapidly solvable inner iterations, and its weakness in its slow convergence close to the optimum. This drawback meant that we were not able to confirm Mazziotti's promising results of this method's application to molecules in a double-zeta basis set.[16] Mazziotti has reported a 10-20 fold speed-up compared to his implementation of

a first-order non-linear method.[21] Although the computation time may depend heavily on the convergence criterion due to the method's slow convergence near the optimum, we were not able to get the desired accuracy ($\leq 10^{-4}$) after the 1000-1500 outer iterations that Mazziotti's algorithm reportedly needed to converge.[16] Still, the boundary point method is substantially faster than the classical barrier method for the majority of test systems considered, but is slower for others. Moreover, based on the calculations in a STO-6G and D95V basis set, its superiority may be less distinct in larger basis sets.

Conclusions on v2DM methods for chemical calculations

The v2DM method faces two major challenges: the N-representability problem on the theoretical side, which determines the method's accuracy, and the algorithmic challenge of a large-scale semidefinite program on the computational side, which determines its speed. More stringent necessary N-representability constraints may improve the accuracy without increasing computation time significantly, while more efficient algorithms make it feasible to impose higher-order constraints that improve the accuracy. The calculations presented here demonstrate that significant progress in both of these areas is needed to make the method competitive to wavefunction based methods with similar scaling, such as CCSD, for the v2DM(PQG) method. The v2DM(PQG) method is several orders of magnitude slower and is less accurate, especially for molecular geometries with stretched bonds.

On the theoretical side, two major defects of currently applied N-representability methods for chemical applications are size-inconsistency and an inconsistent description of non-singlet spin states. Both shortcomings can be regarded as a failure of approximate positivity conditions to describe subsystems with fractional electron number and spin by a formalism that assumes integer electron number and half-integer spin for the whole system.

Although the v2DM method under sufficient N-representability conditions is size-consistent, the v2DM(PQG) method is by nature not size-consistent, because the positivity constraints are directly formulated in terms of the 2DM, which is not a separable quantity. There is no straightforward way to force it to treat a system composed of non-interacting molecules equivalently to separate calculations on each of those moieties.

Similarly, the formulation of the v2DM(PQG) problem does not allow a straightforward way of treating different spin states on equal footing, resulting in non-degenerate energies for theoretically degenerate spin states as well as size-inconsistency. A pragmatic solution to the non-degeneracy of different spin projections in a multiplet is to apply the most stringent conditions and argue that any other spin projection can be generated from it by applying the Wigner-Eckart theorem.

Any size-inconsistencies can be resolved in an ad-hoc manner using subspace energy constraints. However, these constraints directly tackle the energy and are not strong enough to impose the correct structure on the 2DM. Therefore, in general chemical properties other than the energy are not size-consistent.

On the computational side, the scale-up of current semidefinite algorithms with basis dimension forms a major challenge. In order to be competitive with wavefunction based methods of the same level of theory, such as CISD or CCSD, the v2DM(PQG) method needs to have similar speed. Although current algorithms for the v2DM(PQG) method technically scale up similarly to wavefunction based approaches such as CISD and CCSD, $O(K^6)$, all of the zeroth order, first order, and second order methods examined here and in the literature have a much bigger prefactor, making them several orders of magnitude slower. On the positive side, the v2DM does not suffer from the typical pitfalls of wavefunction-based approaches, such as convergence to a wrong root, and a robust algorithm could make it an ideal 'black box' method for calculating ground states.

On a more personal level, this research has taught me a lot about the fundamental differences between wavefunction-based approaches and density-matrix-based methods, which build up a descriptor for the system 'from scratch' without having any of the self-evident properties of the wavefunction. In the absence of these certainties, I have come to appreciate the complexity of the many-electron problem. Rather than dismiss the method as a 'dead end', as was once suggested to me during a conference, I believe that there must be a

formulation that makes the method competitive to wavefunction based methods, although it may require improvements in the field of semidefinite programming and it may be system-dependent. The N-representability problem is of a different nature from the computational problem, as the exact solution is available to us, whereas the computational problem involves trial and error.

Experience gained in the field of the v2DM may prove useful to other density-matrix-based methods, which hold the ability to break the prohibitive exponential scaling of the wavefunction with the size of the molecule. I currently view the method not as a dead end, but rather as an open end.

A

Krylov subspace methods

We have used two iterative Krylov subspace methods for solving a linear system of equations of the form $Ax = b$, with $A \succeq 0$: the conjugate gradients method and the conjugate residuals method.

A.1 Conjugate gradients

The *conjugate gradients* algorithm minimizes the error $e = x - A^{-1}b$, $A \succeq 0$, and only requires the matrix-vector map $A()$

$$\begin{aligned} p_0 = r_0 &= b - A(x_0) \\ \alpha_i &= \frac{r_i^T r_i}{p_i^T A(p_i)} \\ x_{i+1} &= x_i + \alpha_i p_i \\ r_{i+1} &= r_i + \alpha_i A(p_i) \\ \beta_{i+1} &= \frac{r_{i+1}^T r_{i+1}}{r_i^T r_i} \\ p_{i+1} &= r_{i+1} + \beta_{i+1} p_i \end{aligned} \tag{A.1}$$

The *preconditioned conjugate gradients* method, which only references the preconditioner-vector map $M^{-1}(x)$, can be formulated as follows

$$\begin{aligned} p_0 = r_0 &= b - A(x_0) \\ \alpha_i &= \frac{r_i^T M^{-1}(r_i)}{p_i^T A(p_i)} \\ x_{i+1} &= x_i + \alpha_i p_i \\ r_{i+1} &= r_i + \alpha_i A(p_i) \\ \beta_{i+1} &= \frac{r_{i+1}^T M^{-1}(r_{i+1})}{r_i^T M^{-1}(r_i)} \\ p_{i+1} &= M^{-1}(r_{i+1}) + \beta_{i+1} p_i \end{aligned} \tag{A.2}$$

A.2 Conjugate residuals

The *MINRES* or *conjugate residuals* algorithm minimizes the residual $r = b - A(x)$

$$\begin{aligned} p_0 = r_0 &= b - A(x_0) \\ \alpha_i &= \frac{r_i^T A(r_i)}{(A(p_i))^T A(p_i)} \\ x_{i+1} &= x_i + \alpha_i p_i \\ r_{i+1} &= r_i + \alpha_i A(p_i) \\ \beta_{i+1} &= \frac{r_{i+1}^T A(r_{i+1})}{r_i^T A(r_i)} \\ p_{i+1} &= r_{i+1} + \beta_{i+1} p_i \\ A(p_{i+1}) &= A(r_{i+1}) + \beta_{i+1} A(p_i) \end{aligned} \tag{A.3}$$

The *preconditioned conjugate residuals* method, which only references the

preconditioner-vector map $M^{-1}(x)$, can be formulated as follows

$$
\begin{aligned}
p_0 = r_0 &= b - A(x_0) \\
\alpha_i &= \frac{(M^{-1}(r_i))^T A(M^{-1}(r_i))}{(A(p_i))^T M^{-1}(A(p_i))} \\
x_{i+1} &= x_i + \alpha_i p_i \\
r_{i+1} &= r_i + \alpha_i A(p_i) \qquad \text{(A.4)} \\
\beta_{i+1} &= \frac{M^{-1}(r_{i+1})^T A(M^{-1}(r_{i+1}))}{(M^{-1}(r_i))^T A(M^{-1}(r_i))} \\
p_{i+1} &= M^{-1}(r_{i+1}) + \beta_{i+1} p_i \\
A(p_{i+1}) &= A(M^{-1}(r_{i+1})) + \beta_{i+1} A(p_i) \qquad \text{(A.5)}
\end{aligned}
$$

Bibliography

[1] Löwdin, P. O. *Phys. Rev.* **97**, 1474 (1955).

[2] Husimi, K. *Proc. Phys. Math. Soc. Jpn.* **22**, 262 (1940).

[3] Coulson, C. A. *Reviews of Modern Physics* **32**, 170 (1960).

[4] Coleman, J. *Rev. Mod. Phys.* **35**, 668 (1963).

[5] Kempe, J., Kitaev, A., and Regev, O. *SIAM* **35**, 1070–1097 (2006).

[6] Liu, Y.-K., Christandl, M., and Verstraete, F. *Phys. Rev. Lett.* **98**, 110503 (2007).

[7] Rosina, M. and Garrod, C. *J. Comp. Phys.* **18**, 300 (1975).

[8] Mihailovic, M. V. and Rosina, M. *Nucl. Phys. A* **237**, 221 (1975).

[9] Mihailovic, M. V. and Rosina, M. *Nucl. Phys. A* **237**, 229 (1975).

[10] Garrod, C., Mihailovic, M. V., and Rosina, M. *J. Math. Phys.* **16**, 868 (1975).

[11] Erdahl, R. M. *Reports on Math. Phys.* **15**, 147 (1979).

[12] Nesterov, Y. and Nemirovskii, A. *Interior-point polynomial algorithms in convex programming.* SIAM, (1994).

[13] Alizadeh, F. *SIAM J. Optim.* **5**, 13–51 (1995).

[14] Fukuda, M., Braams, B. J., Nakata, M., Overton, M. L., Percus, J. K., Yamashita, M., and Zhao, Z. *Math. Progr. Ser. B* **109**, 553 (2007).

[15] Mazziotti, D. A. *J. Chem. Phys.* **121**, 10957 (2004).

[16] Mazziotti, D. *Phys. Rev. Lett.* **106**, 083001 (2011).

[17] Coleman, J. and Yukalov, V. I. *Reduced Density Matrices: Coulson's Challenge*. Springer-Verlag, (2000).

[18] Kummer, H. *J. Math. Phys.* **8**, 2063 (1967).

[19] Cioslowski, J., editor. *Many-electron densities and reduced density matrices*. Kluwer Academic, (2000).

[20] Mazziotti, D. A., editor. *Reduced density matrix mechanics: with application to many-electron atoms and molecules*. Adv. Chem. Phys. (2007).

[21] Mazziotti, D. *Phys. Rev. Lett.* **93**, 213001 (2004).

[22] Mayer, J. E. *Phys. Rev.* **100**, 1579–1586 (1955).

[23] Tredgold, R. H. *Phys. Rev.* **105**, 1421–1423 (1957).

[24] Harriman, J. E. *Phys. Rev. A* **17**(4), 1249–1256 Apr (1978).

[25] Harriman, J. E. *Phys. Rev. A* **17**(4), 1257–1268 Apr (1978).

[26] Boyd, S. and Vandenberghe, L. *Convex Optimization*. Cambridge University Press, (2004).

[27] Garrod, C. and Percus, J. K. *J. Math. Phys.* **5**, 1756 (1964).

[28] Coleman, A. J. *Phys. Rev.* **100**, 1579–1586 (1955).

[29] Nakata, M., Nakatsuji, H., and Ehara, M. *J. Chem. Phys.* **114**, 8282 (2001).

[30] Cook, S. A. *STOC ('71 Proceedings of the third annual ACM symposium on Theory of computing)* , 151–158 (1971).

[31] Sasaki, F. *Quantum Chemistry Group, Uppsala* (1962). unpublished.

[32] Yang, C. N. *Rev. Mod. Phys.* **34**, 694–704 (1962).

[33] Erdahl, R. M. *Int. J. Quant. Chem.* **13**, 697–718 (1978).

[34] Mazziotti, D. A. *Phys. Rev. A* **65**, 062511 (2002).

[35] Erdahl, R. M. *Int. J. Quant. Chem.* **13**, 731–736 (1978).

[36] Ayers, P. W. and Davidson, E. *Adv. in Chem. Phys.* **134**, 443 (2007).

[37] Zhao, Z., Braams, B., Fukukda, M., Overton, M. L., and Percus, J. K. *J. Chem. Phys.* **120**, 2095 (2004).

[38] Werner, H.-J., Knowles, P. J., Lindh, R., Manby, F. R., Schütz, M., et al.

[39] Braïda, B. and Hiberty, P. C. *J. Am. Chem. Soc.* **126**, 14890 (2004).

[40] Braïda, B. and Hiberty, P. C. *J. Phys. Chem. A* **112**, 13045 (2008).

[41] Heard, G. L., Marsden, C. J., and Scuseria, G. E. **96**, 4359 (1992).

[42] Ponec, R. and Cooper, D. L. *J. of Mol. Struct. (THEOCHEM)* **727**, 133 (2005).

[43] Mazziotti, D. A. *Phys. Rev. A* **72**, 32510 (2005).

[44] Verstichel, B., van Aggelen, H., Van Neck, D., Ayers, P. W., and Bultinck, P. *Phys. Rev. A* **80**, 32508 (2009).

[45] Stärk, J. and Meyer, W. *Chem. Phys. Lett.* **258**, 421–426 (1996).

[46] Sosa, C., Noga, J., and Bartlett, R. J. *J. Chem. Phys.* **88**, 5974–5976 (1988).

[47] Merritt, J. M., Bondybey, V. E., and Heaven, M. C. *Science* **324**, 1548 (2009).

[48] Gdanitz, R. J. *Chem. Phys. Lett.* **312**, 578–584 (1999).

[49] van Aggelen, H., Bultinck, P., Verstichel, B., Van Neck, D., and Ayers, P. W. *Phys. Chem. Chem. Phys.* **11**, 5558 (2009).

[50] van Aggelen, H., Verstichel, B., Bultinck, P., Van Neck, D., Ayers, P. W., and Cooper, D. L. *J. Chem. Phys.* **132**, 114112 (2010).

[51] Nakata, M., Ehara, M., and Nakatsuji, H. *J. Chem. Phys.* **116**, 5432–5439 (2002).

[52] Gidofalvi, G. and Mazziotti, D. A. *J. Chem. Phys.* **122**, 194104 (2005).

[53] Verstichel, B., van Aggelen, H., Van Neck, D., Ayers, P. W., and Bultinck, P. *J. Chem. Phys.* **132**, 114113 (2010).

[54] Kutzelnigg, W. and Mukherjee, D. *Chem. Phys. Lett.* **317**, 567 (2000).

[55] Nakata, M. and Yusuda, K. *Phys. Rev. A* **80**, 42109 (2009).

[56] Gidofalvi, G. and Mazziotti, D. *J. Chem. Phys.* **125**, 144102 (2006).

[57] Kutzelnigg, W. *Lecture Series on computer and computational sciences* **1**, 1–4 (2006).

[58] Cohen, J., Mori-Sanchez, P., and Yang, W. T. *Science* **321**, 792 (2008).

[59] Mori-Sanchez, P., Cohen, A. J., and Yang, W. T. *J. Chem. Phys.* **125**, 201102 (2006).

[60] Perdew, J. P., Ruzsinszky, A., Csonka, G. I., Vydrov, O. A., Scuseria, G. E., Staroverov, V. N., and Tao, J. M. *Phys. Rev. A* **76**, 40501 (2007).

[61] Savin, A. *Chem. Phys.* **356**(1-3), 91 – 97 (2009).

[62] Lathiotakis, N. N., Gidopoulos, N. I., and Helbig, N. *J. Chem. Phys.* **132**, 084105 (2010).

[63] Rohr, D. R., Pernal, K., Gritsenko, O. V., and Baerends, E. J. *J. Chem. Phys.* **129**(16), 164105 (2008).

[64] Pernal, K. *Phys. Rev. A* **81**(5), 052511 (2010).

[65] Kutzelnigg, W. *Int. J. Quant. Chem.* **95**, 404 (2003).

[66] McWeeny, R. *Adv. Quant. Chem* **31**, 15 (1999).

[67] Shenvi, N. and Izmaylov, A. F. *Phys. Rev. Lett.* **105**, 213003 (2010).

[68] Perdew, J. P., Parr, R. G., Levy, M., and J. L. Balduz, J. *Phys. Rev. Lett.* **49**, 1691 (1982).

[69] van Aggelen, H., Verstichel, B., Bultinck, P., Van Neck, D., Ayers, P. W., and Cooper, D. L. *J. Chem. Phys.* (2010). accepted for publication.

[70] Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Montgomery, Jr., J. A., Vreven, T., Kudin, K. N., Burant, J. C., Millam, J. M., Iyengar, S. S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G. A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J. E., Hratchian, H. P., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Ayala, P. Y., Morokuma, K., Voth, G. A., Salvador, P., Dannenberg, J. J., Zakrzewski, V. G., Dapprich, S., Daniels, A. D., Strain, M. C., Farkas, O., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Ortiz, J. V., Cui, Q., Baboul, A. G., Clifford, S., Cioslowski, J., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Gonzalez, C., and Pople, J. A. Gaussian 03, Revision B.05, Gaussian, Inc., Wallingford, CT, 2004.

[71] Harriman, J. E. *Phys. Rev. A* **75**, 032513 (2007).

[72] McWeeny, R. *Spins in chemistry.* Dover Publications, (1970).

[73] Perdew, J. P., Savin, A., and Burke, K. *Phys. Rev. A* **51**, 4531 (1995).

[74] Cohen, A. J., Mori-Sanchez, P., and Yang, W. *J. Chem. Phys.* **129**, 121104 (2008).

[75] Ess, D., Johnson, E., Hu, X., and Yang, W. *J. Phys. Chem. A* **115**, 76 (2011).

[76] Alcoba, D. R. and Valdemoro, C. *Int. J. Quant. Chem.* **102**, 629 (2005).

[77] Alcoba, D. R., Valdemoro, C., Tel, L. M., and Perez-Romero, E. *Phys. Rev. A* **77**, 042508 (2008).

[78] Gidofalvi, G. and Mazziotti, D. A. *Phys. Rev. A* **72**, 052501 (2005).

[79] Hammond, J. R. and Mazziotti, D. A. *Phys. Rev. A* **73**, 012509 (2006).

[80] Dickhoff, W. H. and Neck, D. V. *Many-body theory exposed! Propagator description of quantum mechanics in many-body systems.* World Scientific Publishing Company, (2005).

[81] Verstichel, B., van Aggelen, H., Van Neck, D., and Bultinck, P. (2010). unpublished work.

[82] Nakata, M. In *Workshop on quantum marginals and density matrices*, (2009). Fields Institute, Toronto.

[83] Nakata, M., Braams, B. J., Fujisawa, K., Fukuda, M., Percus, J., Yamashita, M., and Zhao, Z. *J. Chem. Phys.* **128**, 164113 (2008).

[84] Monteiro, R. D. C. *Math. Prog.* **97**, 209–244 (2003).

[85] Cancés, E. and Stolz, G. *J. Chem. Phys.* **125**, 064101 (2006).

[86] Krishnan, K. and Mitchell, J. E. *Optimization methods and software* **21**, 57–74 (2006).

[87] Helmberg, C. and Rendl, F. *SIAM J. Optim.* **10**, 673–696 (2000).

[88] Toh, K.-C. *SIAM J. Optim.* **14**, 670–698 (2003).

[89] de Klerck, E. *Aspects of semidefinite programming: interior point algorithms and selected applications.* Kluwer Academic Publishers, (2002).

[90] Furrer, L. *Topics in Optimization* (2009).

[91] Liu, J. C. and Nocedal, J. *Math. Programm.* **45**, 503–528 (1989).

[92] Morales, J. L. and Nocedal, J. *SIAM J. Optim.* **10**, 1079–1096 (2000).

[93] Nocedal, J. *Mathematics of computation* **35**, 773–782 (1980).

[94] Polyak, R. *Int. J. Quant. Chem.* **95**, 177 (1992).

[95] Stingl, M. *On the solution of nonlinear semidefinite programs by augmented lagrangian methods.* Shaker Verlag, (2006). Dissertation.

[96] Kocvara, M. and Stingl, M. *Math. Program.* **109**, 413 (2007).

[97] Mosheyev, L. and Zibulevsky, M. *Optimization Meth. and Sofr.* **13**, 235–261 (2000).

[98] Conn, A. R., Gould, N., and Toint, P. L. *Mathematics of computation* **66**, 261–288 (1997).

[99] Verstichel, B., van Aggelen, H., Van Neck, D., and Bultinck, P. *Comp. Phys. Comm.* **182**, 1235–1244 (2011).

[100] Verstichel, B., van Aggelen, H., Van Neck, D., Ayers, P. W., and Bultinck, P. *Comp. Phys. Comm.* **182**, 2025–2028 (2011).

[101] Povh, J., Rendl, F., and Wiegele, A. *Computing* **78**, 277–286 (2006).

[102] Malick, J., Povh, J., and Rendl, F. *SIAM* **20**, 336–356 (2009).

[103] See: "http://www.cfs.dl.ac.uk/gamess-uk/index.shtml", M.F. Guest, I. J. Bush, H.J.J. van Dam, P. Sherwood, J.M.H. Thomas, J.H. van Lenthe, R.W.A Havenith, J. Kendrick, "The GAMESS-UK electronic structure package: algorithms, developments and applications", Molecular Physics, Vol. 103, No. 6-8, 20 March-20 April 2005, 719-747.

[104] McQuarrie, D. A. *Quantum Chemistry.* University Science Books, (1983).

[105] Fukuda, M., Nakata, M., and Yamashita, M. *J. Chem. Phys.* **134**, 103–118 (2007).

The exponential growth of the dimension of the exact wavefunction with the size of a chemical system makes it impossible to compute chemical properties of large chemical systems exactly.

A myriad of ab initio methods that use simpler mathematical objects to describe the system has thrived on this realization, among which the variational second order density matrix method.

The aim of my thesis has been to evaluate the use of variational second order density matrix methods for chemistry and to identify the major theoretical and computational challenges that need to be overcome to make the method successful for chemical applications.

The major theoretical challenges originate from the need for the second order density matrix to be N-representable: it must derive from an ensemble of N-electron states. Our calculations have pointed out major drawbacks of commonly used N-representability conditions in this method, such as incorrect dissociation into fractionally charged molecules and size-inconsistency.

The major computational challenges originate from its formulation as a vast semidefinite optimization problem. We have implemented and compared several algorithms that exploit the specific structure of the problem. Even so, their slow speed remains prohibitive.

If we find ways to overcome these challenges, this method will prove a valuable alternative to wavefunction based methods. It is highly complementary to wavefunction based methods, because of its fundamentally different approach to solving the electron correlation problem. Herein lies its strength and its future.