E Pluribus Duo

Jeff Hammond

Mazziotti Group
Department of Chemistry
The University of Chicago

DOE-CSGF Conference
June 18, 2007
Washington, D. C.

Abstract

The quantum many-body problem can be reformulated in terms of two particles, avoiding the exponential complexity of N-particle wavefunctions. However, the injection from N particles to two particles introduces a new form of complexity known as the N-representability problem. I will discuss the theoretical formulation of reduced-density-matrix mechanics, specifically the variational approach, as well as the computational method for solving the resulting equations. The computational methods range from the primal-dual method for semidefinite programming to a nonlinear optimization approach which is amenable to problems with millions of constraints and unknowns. Preliminary adventures into parallel algorithms may be discussed.

The life cycle of energy

Energy creation

- nuclear reactions
- combustion
- solar cells

Energy transport

- superconductors
- molecular electronics
- nanoelectronics

Energy consumption

- computer chips

Solution to energy problems

Electrons are quantum mechanical

Solve Schödinger equation and be done. (Isn't everything just a PDE waiting to be solved?)

Sources of angst

Electrons interact repulsively

- Electrons interact repulsively with other electrons
- Electrons interact attractively with protons

Quantum mechanics is nasty

- Uncertainty problem
- Interchange symmetry
- Probabilistic measurements

Solution to energy problems

Electrons are quantum mechanical

Solve Schödinger equation and be done. (Isn't everything just a PDE waiting to be solved?)

Sources of angst

Electrons interact repulsively

- Electrons interact repulsively with other electrons
- Electrons interact attractively with protons

Quantum mechanics is nasty

- Uncertainty problem
- Interchange symmetry
- Probabilistic measurements

The Hamiltonian

Undergraduate version

$$\hat{H} = -\frac{\nabla}{2} + \frac{Z}{r_1} + \frac{1}{r_{12}}$$

Graduate version

$$\hat{H} = \sum_{i,j}^{r} {}^1T_j^i a_i^\dagger a_j + \sum_{i,j,k,l}^{r} {}^2V_{kl}^{ij} a_i^\dagger a_j^\dagger a_l a_k$$

$${}^1T_j^i = \langle \phi_i(1) | -\frac{\nabla_1}{2} + \frac{Z}{r_1} | \phi_j(1) \rangle$$

$${}^2V_{kl}^{ij} = \langle \phi_i(1)\phi_j(2) | \frac{1}{r_{12}} | \phi_k(1)\phi_l(2) \rangle$$

The Schrödinger equation

The theory of everything

$$\hat{H}|\Psi_n\rangle = E_i|\Psi_n\rangle$$

Contraction

$$\hat{H} = \sum_{i,j} {}^1T_j^i a_i^\dagger a_j + \sum_{i,j,k,l} {}^2V_{kl}^{ij} a_i^\dagger a_j^\dagger a_l a_k = \sum_{i,j,k,l} {}^2K_{kl}^{ij} a_i^\dagger a_j^\dagger a_l a_k$$

$$E_0 = \langle\Psi_0|\hat{H}|\Psi_0\rangle = \sum_{i,j,k,l} {}^2K_{kl}^{ij} \langle\Psi_0|a_i^\dagger a_j^\dagger a_l a_k|\Psi_0\rangle = \sum_{i,j,k,l} {}^2K_{kl}^{ij} {}^2D_{kl}^{ij}$$

Reduced-density-matrix method

Twenty years after Schrödinger...

$$E = \langle \Psi | \hat{H} | \Psi \rangle = \sum_{ijkl} {}^2K_{kl}^{ij} \langle \Psi | a_i^\dagger a_j^\dagger a_l a_k | \Psi \rangle = \sum_{ijkl} {}^2K_{kl}^{ij} {}^2D_{kl}^{ij}$$

The fundamental quantity required for evaluating quantum mechanical properties is the distribution of electron pairs, not the wavefunction.

Obvious question

Can we solve the Schrödinger equation using only the pair distribution function ${}^2D_{kl}^{ij}$?

Reduced-density-matrix method

Twenty years after Schrödinger...

$$E = \langle \Psi | \hat{H} | \Psi \rangle = \sum_{ijkl} {}^2K_{kl}^{ij} \langle \Psi | a_i^\dagger a_j^\dagger a_l a_k | \Psi \rangle = \sum_{ijkl} {}^2K_{kl}^{ij} {}^2D_{kl}^{ij}$$

The fundamental quantity required for evaluating quantum mechanical properties is the distribution of electron pairs, not the wavefunction.

Obvious question

Can we solve the Schrödinger equation using only the pair distribution function ${}^2D_{kl}^{ij}$?

The N-representability problem

Variational hypothesis

If the energy is only a functional of the PDF, can we replace the variational minimization over wavefunctions with one over all PDFs?

Answer

$$E \equiv \min_{\Psi} \sum_{ijkl} {}^2K^{ij}_{kl} {}^2D^{ij}_{kl}[\Psi] \neq \min_{{}^2D} \sum_{ijkl} {}^2K^{ij}_{kl} {}^2D^{ij}_{kl}$$

Antisymmetry requires us to consider the wavefunction at all times

Derivations

Variational formulation

Proper minimization requires constraints on the variational space to ensure the electrons behave physically.

Magic

Play Yanni music, wave hands, etc.

Result

Variables are positive semidefinite matrices and the constraints are linear in the matrix elements thereof.

Contrained variational problem

Fundamental variables

Small matrices - ${}^1D$, ${}^1Q$

Big matrices - ${}^2D$, ${}^2Q$, ${}^2G$

Contraints

$${}^2D \geq 0 \qquad {}^2Q \geq 0 \qquad {}^2G \geq 0$$

$${}^2Q^{i,j}_{k,l} = {}^2D^{i,j}_{k,l} - {}^1D^i_l \wedge {}^1D^j_k \qquad {}^2G^{i,k}_{l,j} = {}^1D^i_l \delta^j_k - {}^2D^{i,j}_{k,l}$$

$$\sum_k {}^2D^{i,k}_{j,k} = (N-1){}^1D^i_j \qquad \sum_k {}^1D^k_k = N$$

Semidefinite programming

Interior-point method (Nesterov and Nemirovsky, 1988)

- Very robust convergence
- Fast algorithms when they fit in memory
- Scales terribly
- Cannot handle structured problems

Nonlinear factorized method (Burer, Monteiro and Zhang, 1999)

- Semidefiniteness enforced implicitly through factorization: $P = MM^T \geq 0$ for $M \in \mathbb{R}^{n \times m}$
- Nonlinear equations not fun to solve
- Very efficient and handles structure easily

Semidefinite programming in chemistry

Primal-dual method

- Independently done by Nakatsuji and coworkers and Mazziotti in 2001
- Toy systems using SeDuMi in Matlab
- $\mathcal{O}(10^4)$ variables/constraints

Nonlinear factorized method

- In 2004, Mazziotti implemented Burer's method for molecular calculations
- Significant tuning and optimization for 100-fold performance increase
- $\mathcal{O}(10^7)$ variables/constraints

Constrained minimization

Primal formulation

$$\min_X \left\{ C \bullet X \ s.t. \ A_i \bullet X = b_i, \ i = 1, \ldots, m, \ X \geq 0 \right\},$$

Factorized formulation

$$\min_L \left\{ C \bullet (LL^T) \ s.t. \ A_i \bullet (LL^T) = b_i, \ i = 1, \ldots, m \right\},$$

Augmented Lagrangian

$$\mathcal{L} = C \bullet (LL^T) + \sum_{i=1}^{m} \lambda_i (A_i \bullet (LL^T) - b_i) + \frac{\sigma}{2} \sum_{i=1}^{m} |A_i \bullet (LL^T) - b_i|^2$$

From: S. Burer and R.D.C. Monteiro, *Math. Prog. B* **95**, 329 (2003).

First-order minimization

Function

$$f = \mathcal{L} = C \bullet (LL^T) + \sum_i^m \lambda_i (A_i \bullet (LL^T) - b_i) + \frac{\sigma}{2} \sum_i^m |A_i \bullet (LL^T) - b_i|^2$$

Gradient

$$G = \nabla \mathcal{L} = 2\ CL + 2 \sum_i^m \lambda_i (A_i L) + 2\ \sigma \sum_i^m (A_i \bullet (LL^T) - b_i)(A_i L)$$

From: S. Burer and R.D.C. Monteiro, *Math. Prog. B* **95**, 329 (2003).

Algorithm overview

Computational procedures

- DGEMM (bottleneck)
- BLAS1/2 calls
- SPGEMM (or function call)
- L-BFGS, CG or TN minimization

See: D. A. Mazziotti, *J. Chem. Phys.* **121**, 10957 (2004).

Parallel considerations

First round

- Asymptotically 99% DGEMM ∴ parallel matrix-multiplication
- Eliminate SPGEMM in favor of function call
  $\longrightarrow$ cut memory use by 90%

Distributed memory issues

Implementing classic parallel MMM algorithms works against efficient global memory layout and increases communication

Optimization improvement

Need to improve optimization convergence to decrease number of iterations by working harder (up to DGEMM cost rather than two factors less)

Final Thoughts...

...from this talk

- The majority of computational software is written in the classic matrix-vector framework.
- Significant development for multidimensional algorithms is needed so that implementations reflect the natural representation of the data.

...about CSGF

- Take advantage of unique opportunities like being a PI at NERSC or giving invited talks at national labs (pad that CV).
- If your practicum mentor is excited about your project, good things happen.