I) Direct minimization in Hartree Fock and DFT - II) Convergence of Coupled Cluster

Reinhold Schneider, MATHEON TU Berlin

Computational methods in many electron systems: MP2
Workshop Göteborg 2009

DFG Research Center MATHEON
Mathematics for key technologies
Ab initio computation

Reliable computation of atomistic molecular phenomena will play an important role in modern material science, chemistry, (molecular biology)

*Ab initio computation* is based on *first principles of quantum mechanics*
Electronic Schrödinger equation

\[ H\psi = E\psi \]

The Hamilton operator

\[ H = -\frac{1}{2} \sum_i \Delta_i - \sum_i \sum_{\nu=1}^{K} \frac{Z_\nu}{|x_i - a_\nu|} + \frac{1}{2} \sum_{i \neq j}^{N'} \frac{1}{|x_i - x_j|} \]

acts on \textit{anti-symmetric} wave functions \( \psi \in H^1((\mathbb{R}^3 \times \{ \pm \frac{1}{2} \})^{N'}) \),

\( \psi(x_1, s_1, \ldots, x_{N'}, s_{N'}) \in \mathbb{R} \), \( (x_i, s_i) \in \mathbb{R}^3 \times \{ \pm \frac{1}{2} \} \).
Output: ground-state energy

\[ E_0 = \min_{\langle \psi, \psi \rangle = 1} \langle H\psi, \psi \rangle, \quad \psi = \arg \min_{\langle \psi, \psi \rangle = 1} \langle H\psi, \psi \rangle \]

- most quantities for molecules (chemistry) and crystals (solid state physics) can be derived from \( E_0 \), e.g. atomic forces, molecular geometry, bonding and ionization energies etc.
- these quantities are (small) differences \( E_{0,a} - E_{0,b} \)
- excited states are required for opto-electronic effects
- accuracy is limited due to neglecting relativistic and non-Born-Openheimer effects
- .. is beyond the present presentation
Basic Problem - Curse of dimensions

- linear eigenvalue problem, but extremely high-dimensional
- + anti-symmetry constraints + lack of regularity.

- traditional approximation methods (FEM, Fourier series, polynomials, MRA etc.): approximation error in $\mathbb{R}^1$: $\lesssim n^{-s}$, $s$-regularity $\rightsquigarrow$, $\mathbb{R}^{3N'}$: $\lesssim n^{\frac{-s}{3N'}}$, $(s < \frac{5}{2})$ with $n$ DOFs

For large systems $N' \gg 1$ ($N' > 1$) the electronic Schrödinger equation seems to be intractable! But 70 years of impressive progress has been awarded by the Nobel price 1998 in Chemistry: Kohn, Pople
Approximation by sums of anti-symmetric tensor products:

\[ \psi = \sum_{k=1}^{\infty} c_k \Psi_k \]

\[ \Psi_k(x_1, s_1; \ldots; x_{N'}, s_{N'}) = \varphi_{1,k} \wedge \ldots \wedge \varphi_{N',k} = \frac{1}{\sqrt{N'!}} \det(\varphi_{i,k}(x_j, s_j)) \]

with \( \varphi_{i,k} \in \{\varphi_j : j = 1, \ldots\} \), w.l.o. generality

\[ \langle \varphi_i, \varphi_j \rangle = \sum_{s=\pm \frac{1}{2}} \int_{\mathbb{R}^3} \varphi_i(x, s)\overline{\varphi_j(x, s)}dx = \delta_{i,j} . \]

A Slater determinant: \( \Psi_k \) is an (anti-symmetric) product of \( N' \) orthonormal functions \( \varphi_i \), called spin orbital functions

\[ \varphi_i : \mathbb{R}^3 \times \{\pm \frac{1}{2}\} \to \mathbb{R} , \ \ i = 1, \ldots, N, \]
Spin functions and spatial orbitals

Example of Slater determinant \((N' = 2)\)

\[
\Psi[\phi_1, \phi_2](\mathbf{x}, s_1; \mathbf{y}, s_2) = \frac{1}{\sqrt{2}} \left( \phi_1(\mathbf{x}, s_1)\phi_2(\mathbf{y}, s_2) - \phi_2(\mathbf{x}, s_1)\phi_1(\mathbf{y}, s_2) \right).
\]

**Spin functions** \(\chi\) and **spatial orbital functions** \(\phi\):

\[
\varphi(\mathbf{x}, s) = \phi_\alpha(\mathbf{x})\chi_\alpha(s) + \phi_\beta(\mathbf{x})\chi_\beta(s)
\]

With spin functions \(\chi_\alpha(\frac{1}{2}) = 1, \chi_\alpha(-\frac{1}{2}) = 0, \chi_\beta(s) = 1 - \chi_\alpha(s)\)

**Closed shell RHF** (Restricted Hartree Fock):

\[
\phi_{\alpha,i} = \phi_{\beta,i} = \phi_i, \quad i = 1, \ldots, N = \frac{N'}{2}.
\]
Hartree-Fock- (HF) Approximation

- **Ground state energy** \( E_0 = \min \{ \langle H \psi, \psi \rangle : \langle \psi, \psi \rangle = 1 \} \)

- **approximation of** \( \psi \) **by a single Slater determinant**

\[
\psi^S_L(\mathbf{x}_1, s_1, \ldots, \mathbf{x}_{N'}, s_{N'}) := \frac{1}{\sqrt{N'!}} \det(\phi_i(\mathbf{x}_j, s_j))
\]

- **Closed Shell Restricted HF (RHF):** \( N := \frac{N'}{2} \) electron pairs

- **minimization of the functional** \( J^{\text{HF}}(\Phi) \)

\[
\Phi \mapsto J^{\text{HF}}(\Phi) := \langle \mathcal{H} \psi^S_L, \psi^S_L \rangle = \sum_{i=1}^{N} \left( \int (|\nabla \phi_i(x)|^2 + 2V_{\text{core}}(x)|\phi_i(x)|^2 + 
\right.
\]

\[
+ \sum_{j=1}^{N} \int_{\mathbb{R}^3} \left[ \frac{|\phi_j(y)|^2 |\phi_i(x)|^2}{\|x - y\|} - \frac{1}{2} \frac{|\phi_i(x)\phi_j(y)\phi_i(x)\phi_j(y)|}{\|x - y\|} \right] dy \, dx \left) \right)
\]

- **w.r.t. orthogonality constraints**

\[
\Phi = (\phi_i)_{i=1}^{N} \in (H^1(\mathbb{R}^3))^N \text{ and } \langle \phi_i, \phi_j \rangle = \delta_{i,j}
\]
Kohn-Sham model

Theorem (Kohn-Hohenberg)

The ground state energy $E_0$ is a functional of the electron density $n$.

- $\frac{1}{2}E_0 \approx E^{KS} = \inf \{ J^{KS}(\Phi) : \langle \phi_i, \phi_j \rangle = \delta_{ij} \}$

- minimization of the Kohn Sham energy functional $J^{KS}(\Phi)$

$$J^{KS}(\Phi) = \left\{ \int \frac{1}{2} \sum_{i=1}^{N} |\nabla \phi_i|^2 + \int nV_{\text{core}} + \frac{1}{2} \int \int \frac{n(x)n(y)}{|x-y|} \, dx \, dy - E_{xc}(n) \right\}$$

- $\phi_i \in H^1(\mathbb{R}^3)$, electron density $n(x) := \sum_{i=1}^{N} |\phi_i(x)|^2$

- $E_{xc}(n)$ exchange-correlation-energy (not known explicitly)

  e.g. LDA (local density approximation) $n \mapsto E_{xc}(n): \mathbb{R} \rightarrow \mathbb{R}$
\[ \Phi := (\phi_1, \ldots, \phi_N) \in (H^1(\mathbb{R}^3))^N = V^N = \mathcal{V} \]

**Gelfand triple**

\[ V := H^1(\mathbb{R}^3) \subseteq L^2(\mathbb{R}^3) \subseteq H^{-1}(\mathbb{R}^3) = V' \]

\[ \langle \Phi^T \psi \rangle := (\langle \phi_i, \psi_j \rangle)_{i,j} \in \mathbb{R}^{N \times N} \]

**Scalar product**

\[ \langle \langle \Phi, \Psi \rangle \rangle := tr\langle \Phi^T \Psi \rangle = \sum_{i=1}^{N} \langle \phi_i, \psi_i \rangle \in \mathbb{R} \]

\[ A \Phi := (A\phi_1, \ldots, A\phi_N), \ A : V \to V' \]

---

**Simplified Problem:** minimize

\[ J^{SCF}(\Phi) := \sum_{i=1}^{N} \langle A\phi_i, \phi_i \rangle = tr\langle \Phi^T A \Phi \rangle = \langle \langle \Phi, A \Phi \rangle \rangle \]

w.r.t. to orthogonality constraints \( \langle \Phi^T \Phi \rangle = I \). I.e. finding the invariant subspace for the first \( N \) eigenfunctions.
Definition (Stiefel and Grassman manifolds)

**Stiefel manifold** \( \mathcal{V}_{V,N} := \mathcal{V} := \{ \Phi = (\phi_i)_{i=1}^N | \phi_i \in V, \langle \phi_i, \phi_j \rangle = \delta_{i,j} \} \)

**Grassmann manifold** is a quotient manifold

\( \mathcal{G}_{V,N} := \mathcal{G} := \mathcal{V}_{V,N}/\sim, \quad \Phi \sim \tilde{\Phi} \iff \tilde{\Phi} = \Phi U, \quad U \in \mathcal{U}(N) \)

(identify ONB spanning the same subspace span \( \Phi \))

**Density matrix operator** projects onto span \( \Phi := \text{span}\{\phi_i\}, \)

\[ D_{\Phi} := \sum_{i=1}^{N} \langle \phi_i, \cdot \rangle \phi_i \]

There is a one-to-one correspondence between

\([\Phi] \in \mathcal{G} \iff D_{\Phi} \) (density matrix operator)
Theorem (Edelman, Arias, Smith (98); Blauert, Neelov, Rohwedder, S. (08))

**tangent space** $T\Phi_G = \{\delta\psi \in V^N|\langle(\delta\psi)^T\Phi\rangle = 0 \in \mathbb{R}^{N\times N}\}$

$T\Phi_G = \{((\delta\psi_i)_{i=1}^N : \delta\psi_i \in V, \delta\psi_i \perp \text{span}\{\phi_i : i = 1, \ldots, N\}\}$

$(I - D\Phi) : V^N \rightarrow T\Phi_G$, is an orthogonal projection onto the tangent space $T\Phi_G$

**tangent space** $T\Phi_S = T\Phi_G + \{\Phi A : A^T = -A\}$

$$= \{\Theta \in V^N : \langle\Theta^T\Phi\rangle = -\langle\Phi^T\Theta\rangle\}$$
Definition (Gradient)
\[ \nabla J_{KS}(\Psi) = F_{\Psi}^{KS} \Psi \]

Kohn-Sham Fock operator
\[ F_n^{KS} = F_{\Phi}^{KS} : V \rightarrow V' \]

defined by
\[ F_{\Phi}^{KS} \varphi(x) = -\frac{1}{2} \Delta \varphi(x) + V_{\text{core}}(x) \varphi(x) + \int \frac{n(y)}{|x-y|} \, dy \varphi(x) + v_{xc}(n)(x) \varphi(x) \]

Unitary invariance:
\[ J_{KS}(\Phi) = J_{KS}(\Phi U) \]

Theorem (Necessary 12t order conditions)
If \([\Psi] = \text{argmin} \{ J(\Phi) : [\Phi] \in \mathcal{G} \} \in V_N(V^N_h)\) then

\[ \langle \langle F_{[\psi]} \Psi, \delta \Phi \rangle \rangle = 0 \ \forall \delta \Phi \in T_{[\psi]} \mathcal{G} \subset V_N(V^N_h) \]

\[ \langle \langle (I - D_{\psi}) F_{[\psi]} \Psi, \delta \Phi \rangle \rangle = 0 \ \forall \delta \Phi \in V_N(V^N_h) \]

\[ F_{[\psi]}^{KS} \psi_i^{(k)} - \sum_{j=1}^{N} \psi_j^{(k)} \lambda_{ji} = 0 \ \forall i = 1, \ldots, N \]

SCF-iteration Cances-LeBris (01)
Projected Preconditioned Gradient Step Algorithm

Algorithm.

1. Guess initial $\Phi^{(0)} = (\varphi_1, \ldots, \varphi_N) \in \mathcal{V}$

2. repeat

1. compute $\Lambda^{(k)}$ by $\lambda_{ij}^{(k)} = \langle \varphi_i, F_{[\Phi^{(k)}]}^{KS} \varphi_j^{(k)} \rangle$

2. $\hat{\varphi}_i^{(k+1)} = \varphi_i^{(k)} - B_k^{-1} (F_{[\Phi^{(k)}]}^{KS} \varphi_i^{(k)} - \sum_{j=1}^N \varphi_j^{(k)} \lambda_{ji}^{(k)})$ with appropriate preconditioner $B_k$

3. project $\hat{\Phi}^{(k+1)}$ onto $\mathcal{V}^{glob}$: orthonormalize $\hat{\Phi}^{(k+1)}$ to $\Phi^{(k+1)}$

until convergence reached
Projection $P : V^N \to S$, (resp. $G$), $P\hat{\Phi} =: \Phi \in \mathcal{V}$, s.t.
\[ \text{span}\{\hat{\phi}_i : i = 1, \ldots, N\} = \text{span}\{\phi_i : i = 1, \ldots, N\} \]

- Löwdin transformation $\Phi = L^{-1}\hat{\Phi}$ where $LL^T = \langle \hat{\Phi}^T\hat{\Phi} \rangle$
- Diagonalization of $\Lambda^{(n+1)} = \langle \hat{\Phi}^T F[\Phi] \hat{\Phi} \rangle = (\langle \varphi_i, F[\Phi] \varphi_i \rangle)_{i=1}^N$
  yields the first $N$ eigenvalues $\lambda_1^{(n)} \leq \ldots \leq \lambda_N^{(n)}$ of $F[\Phi]$.

**Lemma**

*At the minimizer $[\Psi]$, if $F[\Psi]$ has a spectral gap $\lambda_N < \Lambda_{N+1}$, then $F[\Psi]$ is $V$-elliptic on $T[\Psi]*$

\[ \langle \phi, F[\Psi] \phi \rangle \sim \|\phi\|^2_V \quad \forall \phi \perp \text{span}\{\psi_i; i = 1, \ldots, N\} \]
Comment on direct minimization

The algorithm computes an ON basis of the invariant subspace of $F_Ψ$ corresponding to the $N$ lowest eigenvalues.

- everything is valid if $V := V_h$ is a finite dimensional subspace (Galerkin approximation)
- improvement by `subspace acceleration`: e.g. DIIS
- gradient directed $→$ convergence with Armijo line search

$B : V → V', \|φ\|_B^2 = \langle φ, φ \rangle_B := \langle Bφ, φ \rangle \sim \|φ\|_{H^1}^2$

e.g.: $B \approx \frac{1}{2} \Delta + C$, e.g. multigrid or convolution by FFT

for a fixed operator $A$ it is a block PINVIT iteration cf. e.g. (Bramble & Knyazev et al.) – use for SCF Iteration
Convergence results

**Definition**

measure of the error between subspaces spanned by

\[ \Phi = (\varphi_1, \ldots, \varphi_N), \Psi \]

\[ \| (I - D_\Psi) \Phi \|_B^2 := \sum_{i=1}^{N} \| \phi_i - D_\Psi \phi_i \|_B^2 \]

**Theorem (Blauert, Neelov, Rohwedder, S. (08))**

If \( \Phi^{(0)} \in U_\delta(\Psi) \), and

\[ \langle \langle (\mathcal{J}''(\Psi) - \Lambda) \Phi, \Phi \rangle \rangle \geq \gamma \| \Phi \|_{H^1 N}^2 \]

for all \( \Phi \in \mathcal{T}[\Psi] \mathcal{G} \), then there exists \( \chi < 1 \) such that

\[ \| (I - D_\Psi) \Phi^{(n+1)} \|_B \leq \chi \cdot \| (I - D_\Psi) \Phi^{(n)} \|_B \]
Theorem (Blauert, Neelov, Rohwedder, S. (08))

Let \( \mathcal{J} \in C^2(V^N, \mathbb{R}) \), then

\[
\mathcal{J}(\Phi^{(n)}) - \mathcal{J}(\Psi) + \mathcal{R}_2 = \\
2|\langle\langle(\Psi - \Phi^{(n)}), (\mathcal{A}_{\Phi^{(n)}}\Phi^{(n)} - \Phi^{(n)}\Lambda^{(n)})\rangle\rangle| \lesssim \| (I - D\Psi)\Phi^{(n)} \|^2_{V^N}
\]

where \( \mathcal{R}_2 = O(\| (I - D\Psi)\Phi^{(n)} \|^2_{V^N}) \).

Corresponding results for \textbf{a priori and a posteriori estimates} for the Galerkin solution together with \textbf{S. Schwinger} and \textbf{W. Hackbusch} (08).
Constrained optimization problem (revisited):

\[ u = \arg\min \{ J(v) : G(v) = 0 \} \]

Lagrangian \( L(x) := L(u, \Lambda) = J(u) - \Lambda G(u) \) \( (x = (u, \Lambda) \in X) \)

**Theorem (Rannacher et al.)**

If \( L'(x)y = 0 \ \forall y \in X \) and \( L'(x^n)x^n = 0 \),

\[
L(x) - L(x^n) = \frac{1}{2} L'(x^n)(x - x^n) + \mathcal{O}(\|x - x^n\|_X^3) 
\]

\[
J(u) - J(u^n) = \frac{1}{2} \left[ J'(u^n)(u - u^n) - \Lambda^n G'(u^n)(u - u^n) \right. 
\]

\[
\left. - (\Lambda - \Lambda^n)G(u^n) \right] 
\]

Here \( \mathcal{L}(\Phi, \Lambda) = \mathcal{J}(\Phi) + tr\Lambda(\langle \Phi^T, \Phi \rangle - I) \)
Orbital based functional

\[ \mathcal{L}(\Phi, \Lambda) = \mathcal{J}^{HF}(\Phi) + \text{tr}\Lambda(\langle \Phi^T, \Phi \rangle - I) \]

error measure on \( \mathcal{G} \):

\[ \| [\Phi] - [\Psi] \| := \inf \{ U \in \mathcal{U}(N) : \| \Phi - \Psi U \|_{\mathcal{V}_N} \} \]

**Theorem**

\[ \mathcal{J}(\Psi_h) - \mathcal{J}(\Psi) + \mathcal{R}_3(\Phi U - \Phi_h) = \]

\[ 2\langle \langle (\Psi U - \tilde{\Phi}_h), \mathcal{A}_\Psi \Psi_h - \Psi h \Lambda \rangle \rangle \forall \tilde{\Phi}_h \in \mathcal{V}_h \]

But \( U = \text{argmin}_{U \in \mathcal{U}(N)} \mathcal{R}_3(\Phi U - \Phi_h), \mathcal{R}_3 = \mathcal{O}(\| \Psi_h - \Psi U \|_{\mathcal{V}_N}^3) \) is unknown.
Interpolation estimates

For local basis functions like Finite Elements or wavelets

\( \Omega_k := \text{supp} \psi_k \subset \tilde{\Omega}_k, \ V_h := \text{span}\{\psi_k\}, \)

\( h_k \sim \text{diam} \Omega_k \sim \text{diam} \tilde{\Omega}_k, \) then there exists an operator

\( P_h : V \rightarrow V_h \) reproducing polynomials such that

\[ \|u - P_h u\|_{L^2(\Omega_k)} \lesssim h_k^2 \|u\|_{H^2(\tilde{\Omega}_k)} \]

e.g. \( P_h \) (Clermont, Scott-Zhang quasi) interpolation operator

Lemma (\( H^2 \) regularity)

The minimizer \( \Phi = (\varphi_i) \) of \( J^{HF}, J^{SCF} \) is in \( (H^2(\mathbb{R}^3))^N \). This is assumed to hold also for \( J^{KS} \)
A posteriori error estimates

Let $R := (R_i)_{i=1}^N := (F[\psi_h] - \Lambda_h)\Phi_h \in V^N$ be the residual.

The local residuals $r_k, \rho_k$ are defined by

$$r^2_{i,k} := \sum_{i=1}^{N} \|R_i\|_{L_2(\Omega_k)}^2, \quad \rho^2_{i,k} := h_k^{-1/2} \sum_{i=1}^{N} \|\partial_n \varphi_{h,i}\|_{L_2(e_k)}^2$$

where $[\partial_n \varphi_{h,i}]_{e_k}$ denotes the jump of the normal derivatives across the edges in $\overline{\Omega}_k$ if $\psi_k \not\in C^1$.

- The error estimator depends on the representation $\Phi \in V$
  of $[\Phi] \in G$.
- It allows an individual discretization of $\varphi_i$
A posteriori error estimates

Theorem (S. & Schwinger)

If $\Phi \in G$, $\Phi \in V^N$, $\Phi_i \in V_h^N$ be the minimizers of $J = J^{HF}, J^{KS}, J^{SCF}$, then

$$J(\Phi_h) - J(\Phi) \lesssim \left( \sum_{i,k} (r_{i,k}^2 + \rho_{i,k}^2) h_k^4 \right)^{1/2} \| \Phi \|_{(H^2)^N}$$

- it reveals the eigenvalue error estimator of Larsen
- The $H^2$-regularity assumption simplifies the proof, but may not be required, see (Heuveline & Rannacher).
BigDFT Project

- T. Deutsch (CEA Grenoble) [http://inac.cea.fr/L_Sim/BigDFT/index.html](http://inac.cea.fr/L_Sim/BigDFT/index.html)
- S. Goedecker (Uni Basel)
- X. Gonze (UC Louvain)
- R. Schneider (U Kiel → TU Berlin)

Aims:

- implementing an electronic structure calculation program with wavelet bases
- usable on massively parallel computer
- embed this code in the existing program package ABINIT ([www.abinit.org](http://www.abinit.org))
- linear scaling code with respect to the number of electrons
Scaling Function $\varphi$ and Wavelet $\psi$

Figure: Daubechies scaling functions $\varphi$ with $p = 2$ and $5$. 
Adaptivity

two computational regions: **coarse** and **fine** region

(sufficient with pseudopotentials)

- fine region: scaling functions and wavelets (**8 basis functions per grid point**)

- coarse region: only scaling functions (**1 basis function per grid point**)
Adaptivity

Atomic positions (H$_2$O)
Adaptivity

Direct minimization in Hartree Fock and DFT - Convergence of Coupled Cluster
Adaptivity

Coarse grid (low resolution)
Numerical results

example: $C_{19}H_{22}N_{2}O$ ($N = 55$)

Figure:
Numerical results

Convergence history for the direct minimization scheme and together with DIIS acceleration

Figure: convergence
Accuracy Depending on $h_{\text{grid}}$

convergence rate: $\mathcal{O}(h_{\text{grid}}^{14})$
Wavelets vs. Plane Waves (Degrees of Freedom)

**Graph:**
- **Y-axis:** Absolute energy precision (Ha)
- **X-axis:** Number of degrees of freedom
- **Curves:**
  - **Plane waves** (blue dashed line)
  - **Wavelets** (red line)
- **Markers:**
  - $E_c = 40 \text{ Ha}$
  - $E_c = 90 \text{ Ha}$
  - $E_c = 125 \text{ Ha}$
  - $\hbar = 0.4 \text{ bohr}$
  - $\hbar = 0.3 \text{ bohr}$

Reinhold Schneider, MATHEON TU Berlin

I) Direct minimization in Hartree Fock and DFT - II) Convergence
Wavelets vs. Plane Waves (Runtime)

Runtimes for Cinchonidine

- Abinit
- CPMD
- BigDFT

Reinhold Schneider, MATHEON TU Berlin

I) Direct minimization in Hartree Fock and DFT
II) Convergence of Coupled Cluster
Computing Times

I) Direct minimization in Hartree Fock and DFT
II) Convergence
Tensor product approximation is the key to treat the high dimensional problem

Hartree Fock is a rank one anti-symmetric tensor product approximation

In theory, Density Functional Theory (DFT) can provide the exact ground state energy and density

Due to the unknown exchange correlation potential there remains an unavoidable modelling error

The high dimensional ($d = 3N$) linear eigenvalue problem is reduced by HF and DFT to a (system of) low dimensional ($d = 3$) nonlinear eigenvalue type problems

relatively large systems can be treated if one accepts the modeling error, complexity is $O(N^2 \dim V_h) \sim O(N^3) \rightarrow O(N)$ for large systems.

for more accurate computations one has to consider the original highdimensional problem

Quantum Monte Carlo Methods or Wave Function - Post Hartree Fock Methods
Full CI Configuration Interaction Method

Approximation space for (spin) orbitals \((x_j, s_j) \rightarrow \varphi(x_j, s_j)\)

\[ \mathcal{X}_h := \text{span} \{ \varphi_i : i = 1, \ldots, N \} \subset H^1(\mathbb{R}^3 \times \{ \pm \frac{1}{2} \}), \quad \langle \varphi_i, \varphi_j \rangle = \delta_{i,j}, \]

Full CI (for benchmark computations \(\leq N = 18\)) is a \textbf{Galerkin method} w.r.t. the subspace

\[ \mathcal{V}_{\text{FCI}} = \bigwedge_{i=1}^{N} \mathcal{X}_h = \text{span}\{ \psi_{SL} = \psi[\nu_1, \ldots, \nu_N] = \frac{1}{\sqrt{N!}} \det(\varphi_{\nu_i}(x_j, s_j))_{i,j=1}^{N} \} \]

\textbf{Galerkin ansatz}: \( \psi = c_0 \psi_0 + \sum_{\nu \in J} c_\nu \psi_\nu \)

\[ H = (\langle \psi_\nu', H \psi_\nu \rangle), \quad Hc = Ec, \quad \text{but} \quad \dim \mathcal{V}_h = \binom{N}{N} ! \]
Let $\Psi_0 = \Psi[1, \ldots, N] = \frac{1}{\sqrt{N!}} \det(\varphi_i(x_j, s_j))_{i,j=1}^N$, be a reference.

These first $N$ orbital functions $\varphi_i$ are called occupied orbitals. The others are called unoccupied orbitals $\varphi_1, \ldots, \varphi_N, \varphi_{N+1}, \ldots, \varphi_N$.

They are eigenfunctions of $F : \mathcal{X} \to \mathcal{X}', \mathcal{X} := H^1(\mathbb{R}^3 \times \{\pm \frac{1}{2}\})$,

$$\langle F \varphi_i - \lambda_i \varphi_i, \phi_h \rangle = 0 \quad \forall \phi_h \in \mathcal{X}_h, \quad f^p_r := \langle F \varphi_r, \varphi_p \rangle = \delta_{r,p}, \quad r, p, \leq N$$

**Lemma**

The Fock operator $\mathcal{F} := \mathcal{F}_h = \sum_{k=1}^N F_k : \mathcal{V}_{FCI} \to \mathcal{V}_{FCI}$ admits

$$\mathcal{F} \Psi[\nu_1, \ldots, \nu_N] = \left( \sum_{i=1}^N \lambda_{\nu_i} \right) \Psi[\nu_1, \ldots, \nu_N]$$
Some remarks about choice of $\varphi$

\[ \mathcal{H} = \mathcal{F} + \mathcal{U}, \mathcal{U} \text{ is a two particle operator} \]

- FCI exponentially scaling wr.t. to $N$, one has to confine to a small subspace
- Usually $N \sim 10N$ but one has to resolve the e-e cusp
- $F$ Fock operator for (HF or KS), precomputation
- H. Yserentant has shown the existence of $F$ where the eigen-functions provide a sparse grid basis (to avoid curse of dimensions) $\rightarrow$ adaptive schemes (Griebel et al., Flad, Rohwedder & S)
- $\varphi_k$ can be chosen optimally (MRSCF), not eigen functions of a single operator, existence proved by Friesecke, Levin but still exponential complexity of FCI
- Localized basis functions reduced complexity of matrix elements $\rightarrow$ linear scaling (?)
Second quantization: \textit{annihilation operators}:

\[ a_j \Psi[j, 1, \ldots, N] := \Psi[1, \ldots, N] \]

and \( := 0 \) if \( j \) not apparent in \( \Psi[] \).

The adjoint of \( a_b \) is a \textit{creation operator} \( \nu \)

\[ a_b^\dagger \Psi[1, \ldots, N] = \Psi[b, 1, \ldots, N] = (-1)^N \Psi[1, \ldots, N, b] \]

\textbf{Theorem (Slater-Condon Rules)}

\( H : \mathcal{V} \rightarrow \mathcal{V} \) resp. \( H : \mathcal{V}_{DFC} \rightarrow \mathcal{V}_{DFC} \) reads as (basis dependent)

\[
\mathcal{H} = \mathcal{F} + \mathcal{U} = \sum_{p,q} f_p^r a_r a_p^\dagger + \sum_{p,q,r,s} u_{rs}^{pq} a_r a_s a_q a_p^\dagger
\]
Excitation operators

Single excitation operator, let \( \Psi_0 = \Psi[1, \ldots, N] \) be a reference determinant then e.g.

\[
X^k_1 \Psi_0 := a_k^\dagger a_1 \Psi_0
\]

\((-1)^{-p} \Psi^k_1 = \Psi[k, 2, \ldots, N] = X^k_1 \Psi_0 = X^k_j \Psi[1, \ldots, \ldots, N] = a_k^\dagger a_1 \Psi_0\)

higher excitation operators

\[
X_\mu := X^{b_1, \ldots, b_k}_{l_1, \ldots, l_k} = \prod_{i=1}^{k} X^{b_i}_{l_i}, \quad 1 \leq l_i < l_{i+1} \leq N, \quad N < b_i < b_{i+1}.
\]

A CI solution \( \Psi_h = c_0 \Psi_0 + \sum_{\mu \in J_h} c_\mu \Psi_\mu \) can be written by

\[
\Psi_h = \left( c_0 + \sum_{\mu \in J_h} c_\mu X_\mu \right) \Psi_0, \quad c_0, c_\mu \in \mathbb{R}, \quad J_h \subset J.
\]
Let $\psi_0$ be a reference Slater determinant, e.g. $\psi_0 = \psi_{HF}$ and $\psi \in \mathcal{V}_{\text{FCI}}$, satisfying

$$\langle \psi, \psi_0 \rangle = 1$$

intermediate normalization.

Then there exists an unique excitation operator ($T_1$ - single-, $T_2$ - double-, ... excitation operators)

$$T = \sum_{i=1}^{N} T_i = \sum_{\mu \in \mathcal{J}} t_{\mu} X_{\mu}$$

such that

$$\psi = e^T \psi_0$$
Let $\psi \in \mathcal{V}_{\text{FCI}}$ satisfying $\mathcal{H}\psi := \mathcal{H}_h\psi = E_0\psi$, then, due to the Slater Condon rules and $\langle \psi, \psi_0 \rangle = 1$

$$E = \langle \psi_0, \mathcal{H}\psi \rangle = \langle \psi_0, \mathcal{H}e^T\psi_0 \rangle = \langle \psi_0, \mathcal{H}(I + T_1 + T_2 + \frac{1}{2} T_1^2)\psi_0 \rangle$$

The Projected Coupled Cluster Method applies the ansatz

$$\psi_h = e^{T_h}\psi_0 , \quad T_h := T := \sum_{k=1}^{l} T_k = \sum_{\mu \in J_h} t_\mu X_\mu , \quad 0 \neq \mu \in J_h \subset J$$

$$\psi_\mu \in \text{subspace of } \mathcal{V}_{\text{FCI}} \text{, CCSD } T = T_1 + T_2 = T(t) \text{ and (Galerkin) projection onto subspace.}$$
Projected Coupled Cluster Method

Let \( T = \sum_{k=1}^{I} T_k = \sum_{\mu \in \mathcal{J}_h} t_{\mu} X_{\mu} \), \( 0 \neq \mu \in \mathcal{J}_h \subset \mathcal{J} \) using

\[
0 = \langle \psi_0, (H - E)\psi \rangle = \langle \psi_0, (H - E(t_h)e^{T(t_h)}\psi_0 \rangle
\]

The unlinked projected Coupled Cluster formulation

\[
0 = \langle \psi_{\mu}, (H - E(t_h)) e^{T(t_h)}\psi_0 \rangle =: g_{\mu}(t), \ t = (t_{\nu})_{\nu \in \mathcal{J}_h}, \ \mu, \nu \in \mathcal{J}_h
\]

The linked projected Coupled Cluster formulation consists in

\[
0 = \langle \psi_{\mu}, e^{-T}He^{T}\psi_0 \rangle =: f_{\mu}(t), \ t = (t_{\nu})_{\nu \in \mathcal{J}_h}, \ \mu, \nu \in \mathcal{J}_h
\]

These are \( L = \# \mathcal{J}_h << \mathcal{N} \) nonlinear equations for \( L \) unknown excitation amplitudes \( t_{\mu} \).

Theorem (Kümmel ∼ 1959, S. 09)

The (projected) CC Method is size consistent:

\[
H_{AB} = H_A + H_B \Rightarrow E_{AB}^{CC} = E_A^{CC} + E_B^{CC}.
\]
We recall the **Baker-Campell-Hausdorff formula**

\[ e^{-T} A e^T = A + [A, T] + \frac{1}{2!} [[A, T], T] + \frac{1}{3!} [[[A, T], T], T] + \ldots = \]

\[ A + \sum_{k=1}^{\infty} \frac{1}{k!} [A, T]_k. \]

For \( \Psi \in \mathcal{V}_h \) the above series terminates,

\[ e^{-T} H e^T = H + [H, T] + \frac{1}{2!} [[H, T], T] + \frac{1}{3!} [[[H, T], T], T] + \frac{1}{4!} [H, T]_4 \]

e.g. for a single particle operator e.g. \( \mathcal{F} \) there holds

\[ e^{-T} \mathcal{F} e^T = \mathcal{F} + [\mathcal{F}, T] + [[\mathcal{F}, T], T] \]
\[ 0 = f_{\mu}(t) = \langle \Psi_{\mu}, e^{-T H e^{T}} \Psi_0 \rangle \]
\[= \langle \Psi_{\mu}, [\mathcal{F}, X_{\mu}] \Psi_0 \rangle + \sum_{k=0}^{4} \frac{1}{k!} \langle \Psi_{\mu}, [U, T]_k \Psi_0 \rangle \]

The nonlinear amplitude equation \( f(t) = 0 \) is solved by

**Algorithm (quasi Newton-scheme)**

1. **Choose** \( t^0 \), e.g. \( t^0 = 0 \).
2. **Compute**
   \[ t^{n+1} = t^n - A^{-1} f(t^n), \]
   
   where \( A = \text{diag}(\epsilon_{\mu})_{\mu \in \mathcal{J}} > 0. \)

\(-\epsilon_{\mu} < \lambda_N - \lambda_{N+1} < 0 \) (Bach-Lieb-Sololev) the matrix \( A > 0 \)
We consider the projected CC as an approximation of the full CI solution!

If $h \to 0$, then $\mathcal{M} \to \infty$ and $\max \epsilon_\mu \to \infty$! We need estimates uniformly w.r.t. $h, N$

**Definition**

Let $\mathcal{M} := \dim \mathcal{V}_{FCI}$ dimensional parameter space $\mathcal{V} = \mathbb{R}^\mathcal{M}$ equipped with the norm

$$\| t \|_V^2 := \left\| \sum_{\mu \in \mathcal{J}} \epsilon_\mu t_\mu \Psi_\mu \right\|_{L^2((\mathbb{R}^3 \times \{\pm \frac{1}{2}\})^N)}^2 = \sum_{\mu \in \mathcal{J}} \epsilon_\mu |t_\mu|^2$$
Lemma (S.06)

There holds

\[ \| t \|_V \sim \| T\psi_0 \|_{H^1((\mathbb{R}^3 \times \{\pm \frac{1}{2}\})^N)} \sim \| T\psi_0 \|_V. \]

Lemma (S.06)

For \( t \in \ell_2(J) \), the operator \( T := \sum_{\nu \in J} t_{\nu} X_{\nu} \) maps

\[ \| T\psi \|_{L_2} \lesssim \| t \|_{\ell_2} \| \psi \|_{L_2} \quad \forall \psi \in \mathcal{V}_{FCI} \subset \bigwedge_{i=1}^{N} L_2(\mathbb{R}^3 \times \{\pm \frac{1}{2}\}). \]
Lemma (S.06)

For \( t \in V \), the operator \( T := \sum_{\nu \in J} t_\nu X_\nu \) maps

\[
\| T \psi \|_{H^1} \lesssim \| t \|_V \psi \|_{H^1} \quad \forall \psi \in \mathcal{V}_{FCI}
\]

Corollary (S06)

The function \( f : V \rightarrow V' \) is differentiable at \( t \in V \) with the Frechet derivative \( f'[t] : V \rightarrow V' \) given by

\[
(f'[t])_{\nu,\mu} = \langle \psi_\nu, e^{-T[H,X_\mu]}e^T\psi_0 \rangle \\
= \epsilon_\nu \delta_{\nu,\mu} + \langle \psi_\nu, e^{-T[U,X_\mu]}e^T\psi_0 \rangle
\]

All Frechet derivatives \( t \mapsto f^{(k)}[t] : V \rightarrow V' \), are Lipschitz continuous. In particular \( f^{(5)} \equiv 0 \).
Lemma

Let $\Psi_h = e^{T_h} \Psi_0$ where

$$\| t - t_h \|_V \lesssim \inf_{v \in \mathbb{R}^{\#J_h}} \| t - v_h \|_V.$$ 

then

$$\| \Psi - \Psi_h \|_{H^1} \lesssim \inf_{v \in \mathbb{R}^L} \| \Psi - e^{\sum_{\mu \in J_h} v_\mu X_\mu} \Psi_0 \|_{H^1}.$$ 

Definition

A function $g$ is called "strongly monotone" at $t$ if

$$\langle g(t) - g(t'), (t - t') \rangle \geq \gamma\| t - t' \|_V^2$$

for some $\gamma > 0$ and all $\| t' - t \|_V < \delta$. 

Reinhold Schneider, MATHEON TU Berlin

1) Direct minimization in Hartree Fock and DFT - II) Convergence of the Coupled Cluster Method
Local existence and quasi-optimal convergence

Strict monotonicity of \( f \) is not known yet! Let \( T(t) := \sum_{\mu} t_{\mu} X_{\mu} \)
we consider \( g : V \to V, g(t)_\nu := \langle \psi_\nu, (H - E(t)e^{T(t)}\psi_0 \rangle . \)

**Theorem (S. 2008)**

Let \( E \) be a simple EV. If \( \| \psi - \psi_0 \|_V < \delta \) sufficiently small, and
\( J_h \) excitation complete, then

1. for \( E = E(t_h) := \langle \psi_0, He^{T(t_h)}\psi_0 \rangle \), there holds
   \( \langle g(t_h), v \rangle = 0 , \forall v \in V_h \iff \langle f(t_h), v \rangle = 0 , \forall v \in V_h \)

2. \( g \) is strongly monontone at \( t \) \( \forall \| t \|_V \leq \delta' \)

3. there ex. \( t_h \in V_h \) with \( \langle g(t_h), v \rangle = \langle f(t_h), v \rangle = 0, \forall v \in V_h, \)
   \( \| t - t_h \|_V \lesssim \inf_{v \in V_h} \| t - v_h \|_V. \)
The error in the energy $|J(t) - J(t_h)|$ can be estimated by

$$|E - E_h| \lesssim \|t - t_h\|_V \|a - a_h\|_V + (\|t - t_h\|_V)^2$$

$$\lesssim \inf_{u_h \in V_h} \|t - u_h\|_V \|a - a_h\|_V$$

$$+ (\inf_{u_h \in V_h} \|t - u_h\|_V)^2.$$

$$|E - E_h| \lesssim \inf_{u_h \in V_h} \|t - u_h\|_V \|a - a_h\|_V$$

$$+ (\inf_{u_h \in V_h} \|t - u_h\|_V)^2$$

$$\lesssim \|t - t_h\|_V \|a - a_h\|_V$$

$$\lesssim \inf_{u_h \in V_h} \|t - u_h\|_V \inf_{b_h \in V} \|a - b_h\|_V$$

All constants involved above are uniform w.r.t. $N \to \infty$. 
The relative difference in the correlation energy between CI and CC for several molecules in bonding configuration is plotted over the total number of electrons $N$ and the number of valence electrons.

All computations were performed with MOLPRO.

The lack of size consistency suggests a behavior $\sqrt{N}$. 
Conclusions - CC is a most powerful wave function method

- If the reference $\Psi_0$ is sufficiently close to the exact solution $\Psi$, granted by the results of Yserentant, then we have quasipotimal convergence w.r.t. the wave function and super-optimal convergence w.r.t. the energy.
- Size consistency
- CCSD and CCSD(T) are standard, CCSDT; CCSDTQ etc. only for extremely accurate computations
- Not good for multi-configurational problems, e.g. (near-) degenerate ground state (where RHF is rather bad)