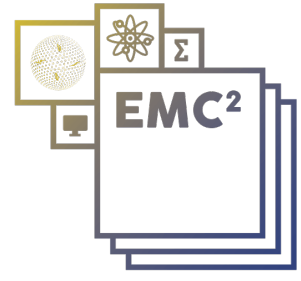
Analysis of SCF and minimization algorithms for electronic structure

Éric Cancès, **Gaspard Kemlin**, Antoine Levitt

gaspard.kemlin@enpc.fr
PhD student with É. Cancès & A. Levitt,
CERMICS, ENPC & Inria Paris, team MATHERIALS

Congrès d'Analyse Numérique pour les Jeunes, Virtual, December 2020



Electronic structure theory

- The behavior of "ordinary" matter is driven by the behavior of nuclei and electrons;
- nuclei are heavier than electrons $\rightsquigarrow$ we consider fixed nuclei and model the electrons with quantum mechanics.

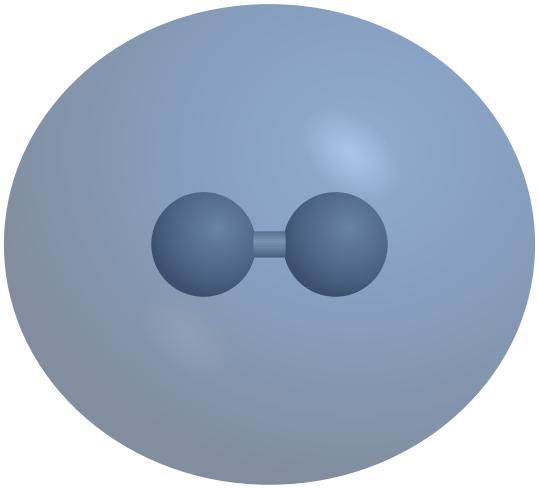
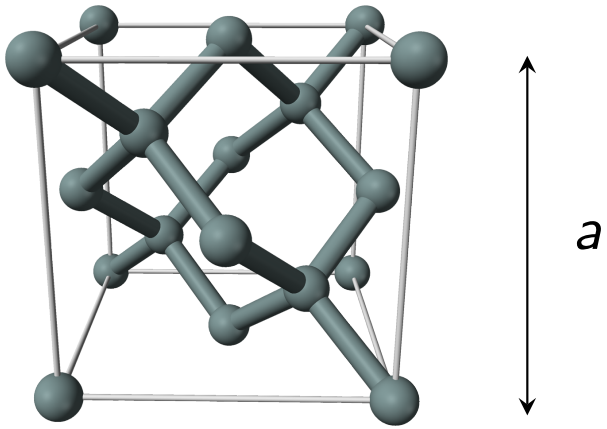


Figure: Electronic density of the dihydrogen molecule H2.

Electronic structure theory for crystals

Electronic structure computations gives the total energy $E(a)$

- Lattice constant: $\operatorname{argmin} E(a)$;
- pressure $\rightsquigarrow \frac{\mathrm{d}E}{\mathrm{d}a}$;
- Young's modulus $\rightsquigarrow \frac{\mathrm{d}^2E}{\mathrm{d}a^2}$;
- anharmonic effects $\rightsquigarrow \frac{\mathrm{d}^3E}{\mathrm{d}a^3}$;
- ...



Quantum mechanics of a single electron

In atomic units, with no spin, we look at the PDE in $\psi(\cdot, t) \in \mathrm{L}^2(\mathbb{R}^3, \mathbb{C})$

$$\mathrm{i}\partial_t \psi(x,t) = \underbrace{-\frac{1}{2}\Delta\psi(x,t)}_{\text{kinetic}} + \underbrace{V(x)\psi(x,t)}_{\text{potential}} =: (H_0\psi)(x,t)$$

- $|\psi(x,t)|^2$ is the probability density of finding the particle at position x at time t: for all time t, $\|\psi(\cdot,t)\|_{\mathrm{L}^2(\mathbb{R}^3)} = 1$;
- we look at stationary states $\psi(x,t) = e^{-\mathrm{i}\varepsilon t}\phi(x)$ where

$$\boxed{H_0\phi = \varepsilon\phi}$$

$\rightsquigarrow$ the linear eigenvalue problem is an essential aspect of the PDE problem.

Quantum mechanics of noninteracting electrons

Consider a system of N noninteracting electrons:

- Pauli exclusion principle $\rightsquigarrow$ two electrons cannot be in the same quantum state;
- ground state $\rightsquigarrow$ electrons fill the first N energy states (*Aufbau* principle).

This leads to solve the linear eigenvalue problem

$$\boxed{H_0\phi_i = \varepsilon_i\phi_i, \quad \varepsilon_1 \leqslant \cdots \leqslant \varepsilon_N}$$

where $H_0 = -\frac{1}{2}\Delta + V$. Then,

- $E = \displaystyle\sum_{i=1}^{N} \varepsilon_i$ is the total energy;
- $\rho(x) = \displaystyle\sum_{i=1}^{N} |\phi_i(x)|^2$ is the total electronic density.

Numerical resolution

$$\text{Find } \phi_i \in \mathbb{R}^{N_b}, \text{ s.t } H_0\phi_i = \varepsilon_i\phi_i, \quad \varepsilon_1 \leqslant \cdots \leqslant \varepsilon_N$$

Orbitals ϕ_i are not unique (degeneracies, phase factor) $\rightsquigarrow$ better to work with the *projectors* onto the space spanned by the $(\phi_i)_{1\leqslant i\leqslant N}$:

$$P := \sum_{i=1}^{N} \phi_i\phi_i^* \in \mathbb{R}^{N_b\times N_b} \quad \text{where} \quad \phi^* = \bar{\phi}^T.$$

- P is a rank N orthogonal projector (*density matrices* for chemists);
- the total energy then writes

$$E = \sum_{i=1}^{N} \varepsilon_i = \sum_{i=1}^{N} \langle\phi_i|H_0\phi_i\rangle = \mathrm{Tr}(H_0P),$$

and is minimal for this P among all rank N orthogonal projectors.

We just showed that we have two equivalent problems:

$$\boxed{H_0\phi_i = \varepsilon_i\phi_i, \quad \varepsilon_1 \leqslant \cdots \leqslant \varepsilon_N \qquad \Leftrightarrow \qquad \min_{P\in\mathcal{M}_N} \mathrm{Tr}(H_0P)}$$

where

$$\mathcal{M}_N := \left\{ P \in \mathbb{R}^{N_b\times N_b} \mid P = P^*,\ \mathrm{Tr}(P) = N,\ P^2 = P \right\}$$

is the set of rank N orthogonal projectors. It is a *Grassmann* manifold.

General mathematical framework

In reality, electrons *do* interact together $\rightsquigarrow$ some models add a nonlinear term $E_{\text{nl}}(P)$ (Kohn-Sham DFT, Hartree-Fock, ...)

$$\inf_{P \in \mathcal{M}_N} E(P) = \operatorname{Tr}(H_0 P) + E_{\text{nl}}(P), \tag{1}$$

where

$$\mathcal{M}_N := \left\{ P \in \mathbb{R}^{N_b \times N_b} \mid P = P^*,\ \operatorname{Tr}(P) = N,\ P^2 = P \right\}.$$

This problem is related to a nonlinear eigenvalue problem as we have the following result:

Proposition

A minimizer P_* of (1) satisfies

$$[H_*, P_*] = 0$$

where $H_* := \nabla E(P_*) = H_0 + \nabla E_{\text{nl}}(P_*)$.

Goals

Main goal: numerically solve the constrained minimization problem

$$\inf_{P \in \mathcal{M}_N} E(P) = \mathrm{Tr}\left(H_0 P\right) + E_{\mathrm{nl}}(P).$$

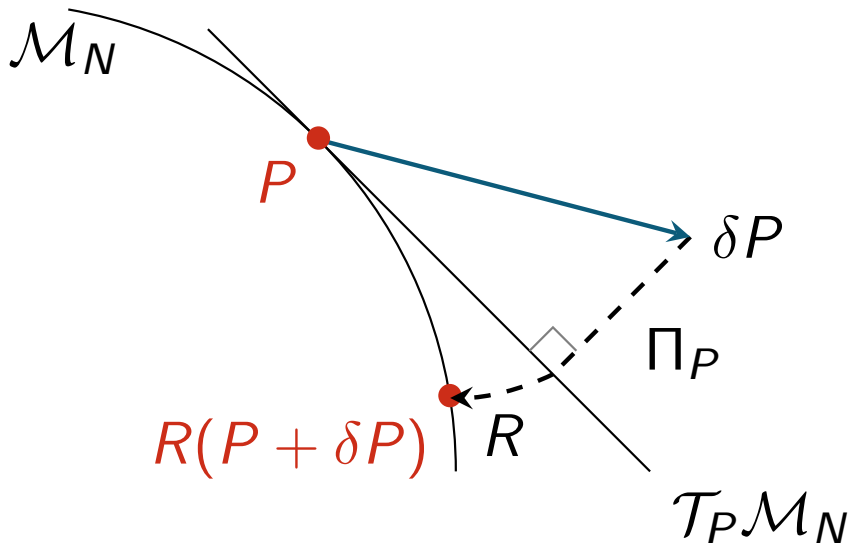
- Optimization on a manifold;
- study/compare convergence and asymptotic behavior;
- justify mathematically known convergence issues.

Some notations...

- $H(P) := \nabla E(P)$;
- Π_P is the orthogonal projection on $\mathcal{T}_P\mathcal{M}_N$, the space tangent to $\mathcal{M}_N$ at P

$$\Pi_P(X) = PX(1-P) + (1-P)XP;$$

- $R : \mathcal{H} \to \mathcal{M}_N$ is a retraction.



Some assumptions...

Assumption 1 $E_{\mathrm{nl}} : \mathcal{H} \to \mathbb{R}$ is twice continuously differentiable, and thus so is E.

Assumption 2 There exists $P_* \in \mathcal{M}_N$ a nondegenerate local minimizer of (1) in the sense that there exists some $\eta > 0$ such that, for $P \in \mathcal{M}_N$ in a neighborhood of P_*, we have

$$E(P) \geqslant E(P_*) + \eta \|P - P_*\|_{\mathrm{F}}^2.$$

Optimality conditions

We have a solution P_* to the constrained minimization problem

$$\inf_{P \in \mathcal{M}_N} E(P) = \mathrm{Tr}\left(H_0 P\right) + E_{\mathrm{nl}}(P).$$

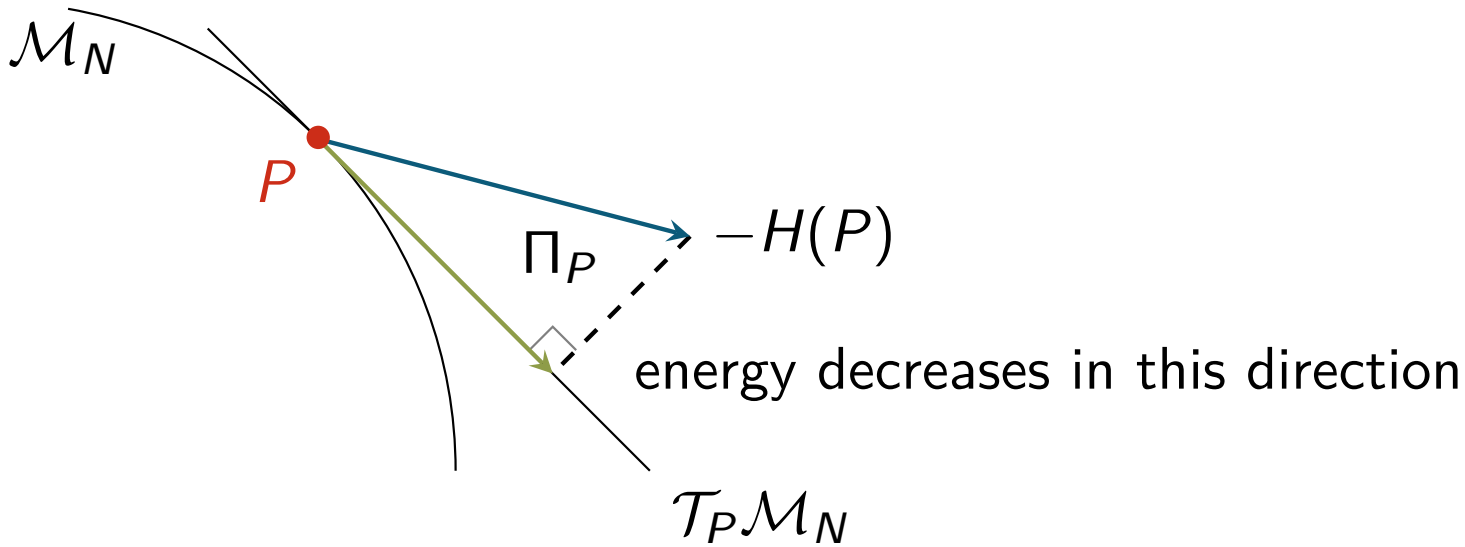
What can we say about P_* and $H_* = \nabla E(P_*)$?

$\rightsquigarrow$ we derive first and second order conditions.

First order condition

The first order optimality condition is $\Pi_{P_*}(H_*) = 0$, which gives

$$\boxed{P_* H_* (1 - P_*) = (1 - P_*) H_* P_* = 0}.$$



First order condition

The first order optimality condition is $\Pi_{P_*}(H_*) = 0$, which gives

$$\boxed{P_* H_* (1 - P_*) = (1 - P_*) H_* P_* = 0}.$$

- $[H_*, P_*] = 0 \Rightarrow H_*$ and P_* can be codiagonalized;
- if $(\phi_i)_{1 \leqslant i \leqslant N_b}$ is an o.n.b. of eigenvectors of H_* ordered by increasing eigenvalues, then $P_* = \sum_{i \in \mathcal{I}} \phi_i \phi_i^*$, with $\mathcal{I}$ the set of *occupied orbitals*;
- $\mathcal{I} \subset \{1, \ldots, N_b\}$ and $|\mathcal{I}| = N$ (rank N projector):
  - $\mathcal{I} = \{1, \ldots, N\}$: *Aufbau* principle;
  - $\mathcal{I} = \{1, \ldots, N\}$ and $\varepsilon_N < \varepsilon_{N+1}$: *strong Aufbau* principle.

$\rightsquigarrow$ in the o.n.b. $(\phi_i)_{1 \leqslant i \leqslant N_b}$ and if $\mathcal{I} = \{1, \ldots, N\}$, P_* and H_* write

$$H_* = \begin{bmatrix} \varepsilon_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \varepsilon_{N_b} \end{bmatrix} \quad \text{and} \quad P_* = \begin{bmatrix} 1_N & 0 \\ 0 & 0 \end{bmatrix}$$

Second order condition

The second order optimality condition reads

$$\boxed{\forall\, X \in \mathcal{T}_{P_*}\mathcal{M}_N,\ \langle X, (\Omega_* + K_*)X\rangle_{\mathsf{F}} \geqslant \eta \|X\|_{\mathsf{F}}^2},$$

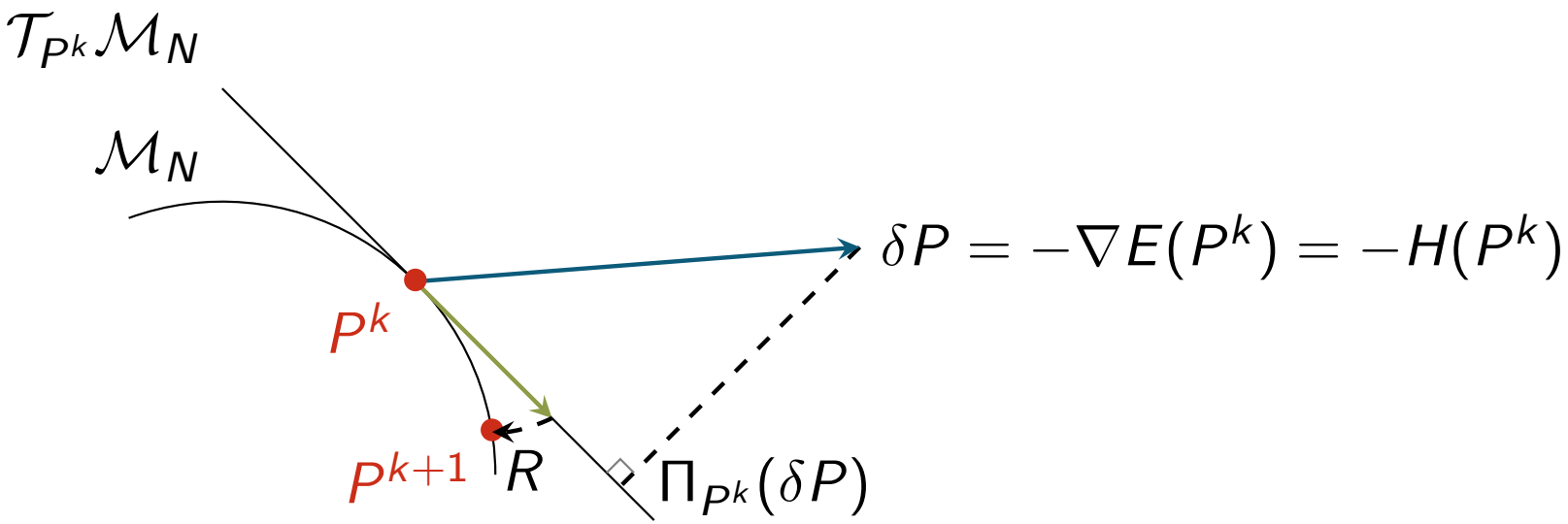
where $K_* := \Pi_{P_*}\nabla^2 E(P_*)\Pi_{P_*}$ and $\Omega_* : \mathcal{T}_{P_*}\mathcal{M}_N \to \mathcal{T}_{P_*}\mathcal{M}_N$.

- $\Omega_* + K_*$ can be interpreted as the Hessian of the energy on the manifold, Ω_* represents the influence of the curvature;
- for $(\phi_i, \varepsilon_i)_{1\leqslant i\leqslant N_b}$ an eigendecomposition of H_*, the gap $\nu = \min_{a\notin\mathcal{I}} \varepsilon_a - \max_{i\in\mathcal{I}} \varepsilon_i$ is the smallest eigenvalue of Ω_*.

Remark: if the *Aufbau* principle is satisfied, the gap is $\varepsilon_{N+1} - \varepsilon_N$.

Projected gradient descent

Solve directly (1) with a projected gradient algorithm:



Data: $P^0 \in \mathcal{M}_N$

while *convergence not reached* **do**

$\quad P^{k+1} := R\left(P^k - \beta \Pi_{P^k} \nabla E(P^k)\right)$;

end

Convergence of projected gradient descent

Theorem (Classical result)

Let $E : \mathcal{H} \to \mathbb{R}$ satisfy Assumption 1 and 2 with P_* a local minimizer of (1). Then, if $P^0 \in \mathcal{M}_N$ is close enough to P_*, the iterations

$$P^{k+1} := R\left(P^k - \beta \Pi_{P^k} \nabla E(P^k)\right)$$

linearly converge to P_* for $\beta > 0$ small enough, with asymptotic rate the spectral radius of $1 - \beta J_{\text{grad}}$ where $J_{\text{grad}} := \Omega_* + K_*$.

$\rightsquigarrow$ in the linear case $K_* = 0$ and the spectral radius depends only on $\|\Omega_*\|_{\text{op}} = \varepsilon_{N_b} - \varepsilon_1 \to \infty$ when $N_b \to \infty$: known preconditioning issues for gradient descent.

Euler-Lagrange equations

Take the constrained minimization problem on $\mathcal{M}_N$

$$\inf_{P\in\mathcal{M}_N} E(P) = \underbrace{\mathrm{Tr}\,(H_0 P)}_{\text{linear}} + \underbrace{E_{\mathrm{nl}}(P)}_{\text{nonlinear}},$$

Recall linear case

$$E(P) = \mathrm{Tr}(H_0 P)$$

$$\downarrow$$

linear eigenvalue problem

$$\begin{cases} H_0\phi_i = \varepsilon_i\phi_i \\ \phi_i^*\phi_j = \delta_{ij}, \\ P = \displaystyle\sum_{i=1}^{N} \phi_i\phi_i^* \end{cases}$$

Nonlinear case

$$E(P) = \mathrm{Tr}(H_0 P) + E_{\mathrm{nl}}(P)$$

$$\downarrow$$

nonlinear eigenvalue problem

$$\begin{cases} (H_0 + \nabla E_{\mathrm{nl}}(P))\phi_i = \varepsilon_i\phi_i, \\ \phi_i^*\phi_j = \delta_{ij}, \\ P = \displaystyle\sum_{i=1}^{N} \phi_i\phi_i^*. \end{cases}$$

Self-consistent field (SCF)

This leads to consider the following iterations:

- Set a starting point $P^0 \in \mathcal{M}_N$;
- solve the linear eigenvalue problem for $H(P^k) = H_0 + \nabla E_{\text{nl}}(P^k)$

$$\begin{cases} H(P^k)\phi_i^k = \varepsilon_i^k \phi_i^k, \\ (\phi_i^k)^* \phi_j^k = \delta_{ij}, \end{cases}$$

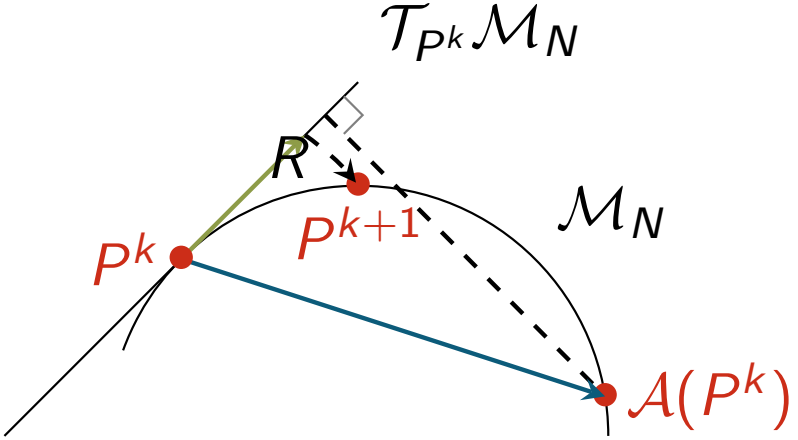
- build the density matrix

$$P^{k+1} = \sum_{i=1}^{N} \phi_i^k (\phi_i^k)^*$$

- solve the linear eigenvalue problem for $H(P^{k+1})$, and so on until convergence.

Damped SCF

Damped SCF algorithm, assuming the *strong Aufbau* principle:



$$\mathcal{A}(P^k) = \sum_{i=1}^{N} \phi_i^k (\phi_i^k)^*$$

Data: $P^0 \in \mathcal{P}_N$

while *convergence not reached* **do**

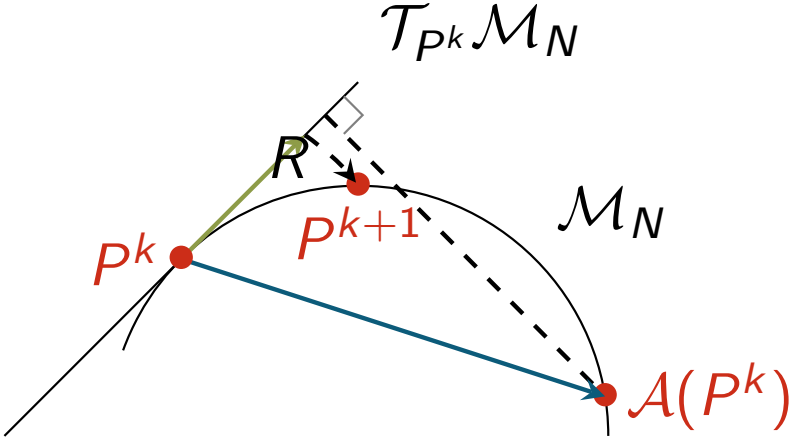
$$\text{solve } \begin{cases} H(P^k)\phi_i^k = \varepsilon_i^k \phi_i^k, \quad \varepsilon_1^k \leqslant \cdots \leqslant \varepsilon_N^k < \varepsilon_{N+1}^k \\ (\phi_i^k)^* \phi_j^k = \delta_{ij}, \end{cases} ;$$

$$P^{k+1} := R\left(P^k + \beta \Pi_{P^k}\left(\mathcal{A}(P^k) - P^k\right)\right);$$

end

Damped SCF

Damped SCF algorithm, assuming the *strong Aufbau* principle:



$$\mathcal{A}(P^k) = \sum_{i=1}^{N} \phi_i^k (\phi_i^k)^*$$

Data: $P^0 \in \mathcal{P}_N$

while *convergence not reached* **do**

$$\text{solve} \begin{cases} H(P^k)\phi_i^k = \varepsilon_i^k \phi_i^k, \quad \varepsilon_1^k \leqslant \cdots \leqslant \varepsilon_N^k < \varepsilon_{N+1}^k \\ (\phi_i^k)^* \phi_j^k = \delta_{ij}, \end{cases} ;$$

$$P^{k+1} := R\left(P^k + \beta \Pi_{P^k}\left(\mathcal{A}(P^k) - P^k\right)\right);$$

end

Convergence of damped SCF

Theorem

Let $E : \mathcal{H} \to \mathbb{R}$ satisfy Assumption 1 and 2 with P_* a local minimizer of (1). Assume that P_* satisfies the strong Aufbau *principle*

$$\mathcal{A}(P_*) = P_* \text{ and } \nu := \varepsilon_{N+1} - \varepsilon_N > 0.$$

Then, for $\beta > 0$ small enough and $P^0 \in \mathcal{M}_N$ close enough to P_*, the iterations

$$P^{k+1} := R\left(P^k + \beta \Pi_{P^k}\left(\mathcal{A}(P^k) - P^k\right)\right)$$

linearly converge to P_*, with asymptotic rate the spectral radius of $1 - \beta J_{\mathrm{SCF}}$ where $J_{\mathrm{SCF}} := 1 + \Omega_*^{-1} K_*$.

$\rightsquigarrow$ consistent with the linear case $K_* = 0$ for which we have a linear eigenvalue problem $H_0 \phi_i = \varepsilon_i \phi_i$: SCF converges in one iteration.

Comparing the Jacobian matrices

Both algorithms have Jacobian matrices of the form $1 - \beta J$ with

- Gradient descent: $J_{\text{grad}} = \Omega_* + K_*$ is self-adjoint for $\langle \cdot, \cdot \rangle_F$;
- SCF: $J_{\text{SCF}} = 1 + \Omega_*^{-1} K_*$ is self-adjoint for $\langle \Omega_* \cdot, \cdot \rangle_F$.

Hence

- in the linear regime, the SCF can be seen as a matrix splitting method for the gradient descent;
- the smaller the gap, the more difficult the convergence of the SCF (known issue for chemists).

More details about the proofs, the models and numerical results available on HAL[1].

[1] Eric Cancès, Gaspard Kemlin, and Antoine Levitt. "Convergence analysis of direct minimization and self-consistent iterations". accepted in SIMAX. Oct. 2020. URL: https://hal.inria.fr/hal-02546060.

Conclusion

Summary

- Initial PDE can be seen as a constrained minimization problem, that can be solved by direct minimization or damped SCF when seen as a nonlinear eigenvalue problem;
- convergence of both algorithms led by Jacobian matrices
  - $\Omega_* + K_*$ for the projected gradient;
  - $1 + \Omega_*^{-1} K_*$ for the damped SCF;
- properties depend on the model $\rightarrow$ mathematical justification of known convergence issues.

Ongoing works (with É. Cancès, G. Dusson and A. Levitt)

- Knowing the asymptotic behavior is a good way of deriving *a posteriori* estimators: how to use this analysis to derive efficient estimators ?

Merci !