

# Numerical analysis for electronic structure calculations

**Gaspard Kemlin**

Under the supervision of Eric Cancès & Antoine Levitt,  
CERMICS, ENPC & Inria Paris, team MATHERIALS

PhD defense, December 15th 2022

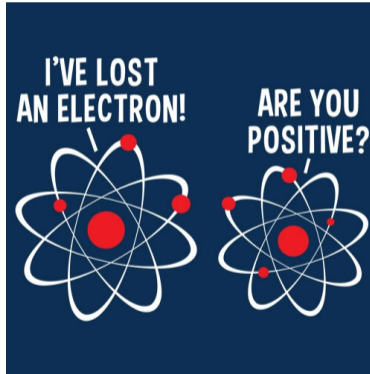


# Numerical analysis for electronic structure calculations

What does it mean ?

# Numerical analysis for **electronic structure calculations**

What does it mean ?



# Molecular simulation in a nutshell

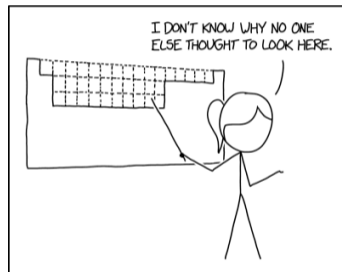
## Wikipedia: Molecular modeling

*Molecular modeling encompasses all methods, theoretical and computational, used to model or mimic the behavior of molecules.*

- Important domain of numerical simulation (1998 and 2013 Chemistry Nobel prizes);
- diversity of physical and mathematical models;
- at the European level,  $1/4 \sim 1/3$  of the computation time on supercomputers is dedicated to molecular simulation.

↪ Electronic structure calculation is part of this field.

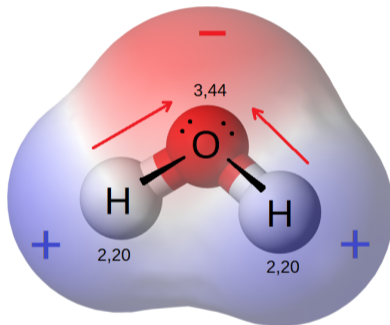
xkcd n°2214



THE 2019 NOBEL PRIZE IN CHEMISTRY WENT TO THE TEAM THAT DISCOVERED THE ELEMENTS IN THE BIG GAP AT THE TOP OF THE PERIODIC TABLE.

# What is electronic structure ?

- The properties of molecules and materials rely on the behavior of their electrons: nuclei are considered as point particles and electrons are modeled with quantum mechanics.
- Electronic structure theory is the study of this behavior:
  - What is the distribution of the electrons ?
  - Which energy levels can they reach ? How do they populate them ?
  - What are the consequences on macroscopic properties ?
- Except for very few systems, modern computers are required to compute (approximate) answers to these questions.



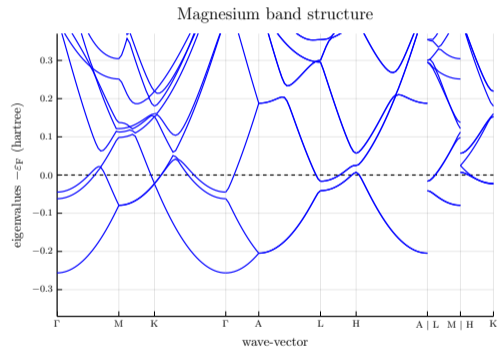
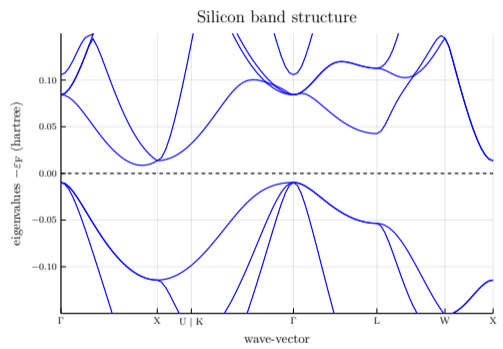
Source: <https://pediaa.com>

Water is a good solvent because of its polarized electron distribution.

# Example: electrical conductivity and band diagrams

## Insulators and semi-conductors

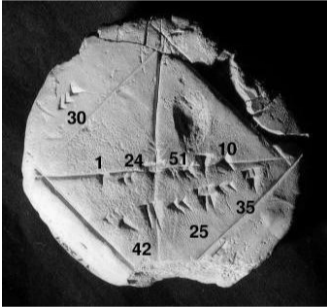
## Metals



Plots generated with DFTK.jl.

# Numerical analysis for electronic structure calculations

What does it mean ?



Babylonian clay tablet YBC 7289 (1800-1600 BC) with annotations to approximate the square root of 2.

Source: Wikipedia Commons.

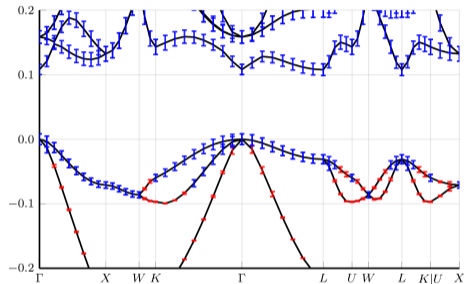


*Hidden Figures* (book & movie, 2016), the story of a team of African-American women mathematicians who played a crucial role at NASA during the early years of the US space program.

# Numerical analysis

As numerical analysts, we:

- analyze the convergence of methods developed by chemists (is the result satisfactory ?);
- estimate the error arising from different sources (models, **discretization**, **solver tolerance**, **finite precision**...);
- try to improve the existing methods (speed, accuracy, robustness...).



Error bars on the band diagram of silicon<sup>a</sup>.






<sup>a</sup> M. F. Herbst, A. Levitt and E. Cancès.

A posteriori error estimation for the non-self-consistent Kohn–Sham equations.

*Faraday Discussions*, 224:227–246, 2020.



## Publications

-  Eric Cancès, **Gaspard Kemplin**, and Antoine Levitt.  
Convergence analysis of direct minimization and self-consistent iterations.  
*SIAM Journal on Matrix Analysis and Applications*, 42(1):243–274, 2021.
-  Eric Cancès, Geneviève Dusson, **Gaspard Kemplin**, and Antoine Levitt.  
Practical error bounds for properties in plane-wave electronic structure calculations.  
*SIAM Journal on Scientific Computing*, 44(5):B1312–B1340, 2022.
-  Eric Cancès, Michael F. Herbst, **Gaspard Kemplin**, Antoine Levitt, and Benjamin Stamm.  
Numerical stability and efficiency of response property calculations in density functional theory.  
*Submitted*, 2022.
-  Eric Cancès, Geneviève Dusson, **Gaspard Kemplin**, and Laurent Vidal.  
On basis set optimisation in quantum chemistry.  
*Accepted in ESAIM Proceedings*, 2022.
-  Eric Cancès, **Gaspard Kemplin**, and Antoine Levitt.  
A priori error analysis of linear and nonlinear periodic Schrödinger equations with analytic potentials.  
*In preparation*, 2022.

- 1 Introduction
- 2 Mathematical framework
- 3 Convergence analysis of direct minimization and SCF iterations – Chapter 2
- 4 Practical error bounds for quantities of interest – Chapter 3
- 5 Numerical stability of response property calculations – Chapter 4
- 6 DFTK and perspectives

- 1 Introduction
- 2 Mathematical framework**
- 3 Convergence analysis of direct minimization and SCF iterations – Chapter 2
- 4 Practical error bounds for quantities of interest – Chapter 3
- 5 Numerical stability of response property calculations – Chapter 4
- 6 DFTK and perspectives

# Quantum mechanics of a single electron

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

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

- $\|\psi(\cdot, t)\|_{L^2(\mathbb{R}^3)} = 1$ ;
- stationary states  $\psi(x, t) = e^{-i\varepsilon t} \varphi(x)$  where

$$\begin{cases} H_0 \varphi = \varepsilon \varphi, \\ \|\varphi\|_{L^2} = 1; \end{cases}$$

- ground-state energy:  $\varepsilon = \min_{\|\varphi\|_{L^2}=1, \varphi \neq 0} \langle \varphi, H_0 \varphi \rangle$ .

# Quantum mechanics of noninteracting electrons

Consider a system of  $N_{\text{el}}$  **noninteracting** electrons:

- Pauli exclusion principle  $\rightsquigarrow$  two electrons cannot be in the same quantum state;
- ground-state  $\rightsquigarrow$  electrons fill the  $N_{\text{el}}$  lowest energy states (*Aufbau* principle).

$$\begin{cases} H_0 \varphi_n = \varepsilon_n \varphi_n, \\ \langle \varphi_n, \varphi_m \rangle_{L^2(\mathbb{R}^3)} = \delta_{nm}, \end{cases} \quad H_0 := -\frac{1}{2} \Delta + V.$$

- $E = \sum_{n=1}^{N_{\text{el}}} \varepsilon_n$  is the ground-state energy;

- $\rho(x) = \sum_{n=1}^{N_{\text{el}}} |\varphi_n(x)|^2$  is the ground-state electronic density, with  $\int_{\mathbb{R}^3} \rho(x) dx = N_{\text{el}}$ .

—  $\varepsilon_{N_{\text{el}}+2}$

—  $\varepsilon_{N_{\text{el}}+1}$

●  $\varepsilon_{N_{\text{el}}}$

⋮

●  $\varepsilon_2$

●  $\varepsilon_1$

## Numerical resolution

Choose your favorite (orthonormal) discretization basis and then:

$$\text{Find } (\varphi_n)_{1 \leq i \leq N_{\text{el}}} \in (\mathbb{R}^{N_b})^{N_{\text{el}}}, \text{ s.t. } H_0 \varphi_n = \varepsilon_n \varphi_n, \quad \varphi_n^T \varphi_m = \delta_{nm}, \quad \varepsilon_1 \leq \dots \leq \varepsilon_{N_{\text{el}}}.$$

Orbitals  $(\varphi_n)_{1 \leq i \leq N_{\text{el}}}$  are not unique (degeneracies)  $\rightsquigarrow$  better to work with the *orthogonal projector* onto the space spanned by the orthonormal family  $(\varphi_n)_{1 \leq i \leq N_{\text{el}}}$ :

$$P_* := \sum_{n=1}^{N_{\text{el}}} |\varphi_n\rangle \langle \varphi_n| = \sum_{n=1}^{N_{\text{el}}} \varphi_n \varphi_n^T \in \mathbb{R}_{\text{sym}}^{N_b \times N_b}.$$

- $P_*$  is a rank  $N_{\text{el}}$  orthogonal projector (*ground-state density matrix*);
- the ground-state energy then reads

$$E = \sum_{n=1}^{N_{\text{el}}} \varepsilon_n = \sum_{n=1}^{N_{\text{el}}} \langle \varphi_n | H_0 \varphi_n \rangle = \text{Tr}(H_0 P_*).$$

$P_*$  minimizes  $\text{Tr}(H_0 P)$  over the set of rank  $N_{\text{el}}$  orthogonal projectors.

We have two equivalent problems:

$$\begin{cases} H_0 \varphi_n = \varepsilon_n \varphi_n, \\ \varphi_n^T \varphi_m = \delta_{nm}, \end{cases} \quad \text{where } \varepsilon_1 \leq \dots \leq \varepsilon_{N_{\text{el}}}, \text{ are the } N_{\text{el}} \text{ lowest eigenvalues} \quad \Leftrightarrow \quad \min_{P \in \mathcal{M}_{N_{\text{el}}}} \text{Tr}(H_0 P)$$

where

$$\mathcal{M}_{N_{\text{el}}} := \{P \in \mathbb{R}^{N_b \times N_b} \mid P = P^T, \text{Tr}(P) = N_{\text{el}}, P^2 = P\}$$

is the set of rank  $N_{\text{el}}$  orthogonal projectors. It is diffeomorphic to the *Grassmann* manifold  $\text{Grass}(N_{\text{el}}, N_b)$ .

## General framework

In reality, electrons *do* interact so that the general form of the energy is

$$E(P) := \underbrace{\text{Tr}(H_0 P)}_{\text{linear term}} + \underbrace{E_{\text{nl}}(P)}_{\text{nonlinear term}}$$

- $P \in \mathbb{R}_{\text{sym}}^{M_b \times N_b}$  is a trial density matrix;
- $H_0 = -\frac{1}{2}\Delta + V$  is the core Hamiltonian;
- $E_{\text{nl}}$  models the electron-electron interaction depending on the model chosen to approximate the  $N$ -body Schrödinger equation (e.g. Kohn–Sham DFT or Hartree–Fock).

### Kohn–Sham equations with LDA

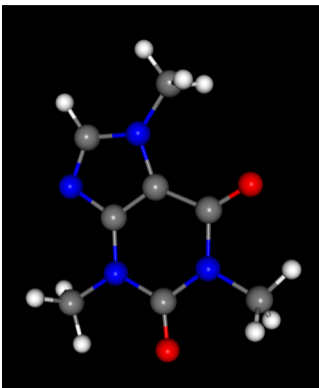
$$\begin{cases} \underbrace{(-\frac{1}{2}\Delta + V_{\text{nuc}})}_{\text{linear term}} \varphi_n + \underbrace{V_{\text{Hxc}}(\rho)}_{\text{nonlinear term}} \varphi_n = \varepsilon_n \varphi_n, \\ \langle \varphi_n, \varphi_m \rangle_{L^2(\mathbb{R}^3)} = \delta_{nm}, \\ \rho = \sum_{n=1}^{N_{\text{el}}} |\varphi_n|^2. \end{cases}$$

(1)

$$\min_{P \in \mathcal{M}_{N_{\text{el}}}} E(P) = \text{Tr}(H_0 P) + E_{\text{nl}}(P),$$

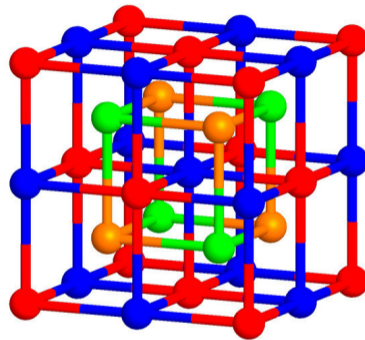
$$\mathcal{M}_{N_{\text{el}}} := \{P \in \mathbb{R}^{M_b \times N_b} \mid P = P^T, \text{Tr}(P) = N_{\text{el}}, P^2 = P\}.$$





The molecular structure of caffeine.

Source: <https://hpc-wiki.info/hpc/Gaussian>



Unit cell of Heusler  $\text{Fe}_2\text{MnAl}$  alloy<sup>1</sup>.

<sup>1</sup>Y. Jirásková, J. Buršík, D. Janičkovič, O. Životský, Influence of Preparation Technology on Microstructural and Magnetic Properties of  $\text{Fe}_2\text{MnSi}$  and  $\text{Fe}_2\text{MnAl}$  Heusler Alloys. *Materials*, 12(5):710-723, 2019).

## The broader picture

Assume that we want to find  $x_*$  such that  $f(x_*) = 0$  for some function  $f : \mathbb{R}^N \rightarrow \mathbb{R}^N$ . Let  $J(x)$  be the Jacobian of  $f$  at  $x$ .

- **Convergence analysis (Chapter 2):** the behavior of  $x^{k+1} = x^k + \beta f(x^k)$  depends on  $1 + \beta J(x_*)$ .
- **Error control (Chapter 3):**  $x - x_* \approx J(x_*)^{-1} f(x)$ .
- **Response calculations (Chapter 4):** if  $f$  depends on a parameter  $\varepsilon$ , then the solution to  $f(x_*(\varepsilon), \varepsilon) = 0$  satisfies

$$\left. \frac{\partial x_*}{\partial \varepsilon} \right|_{\varepsilon=0} = -J(x_*(0))^{-1} \left. \frac{\partial f}{\partial \varepsilon} \right|_{\varepsilon=0}.$$

Here, we have a *constrained* minimization problem:  $f \sim \nabla E$  and we need to define the correct framework to compute the Jacobian  $J(x_*)$  with  $x_*$  on the manifold  $\mathcal{M}_{N_{\text{el}}}$  (**Chapter 2**).

## Some definitions

- $\mathcal{H} := (\mathbb{R}_{\text{sym}}^{N_b \times N_b}, \|\cdot\|_F)$ , endowed with the Frobenius scalar product  $\text{Tr}(A^T B)$ ;
- $\mathcal{M}_{N_{\text{el}}}$  is a smooth manifold, we can define its tangent space  $\mathcal{T}_P \mathcal{M}_{N_{\text{el}}}$  (it is a  $\mathbb{R}$  vector space);
- $\Pi_P$  is the orthogonal projection on  $\mathcal{T}_P \mathcal{M}_{N_{\text{el}}}$ :

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

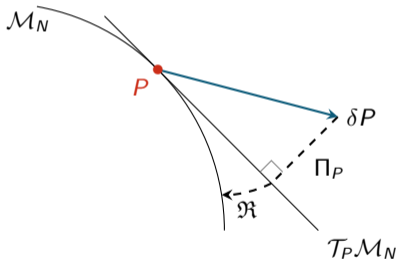
In the decomposition  $\mathcal{H} = \text{Ran}(P) \oplus \text{Ran}(1 - P)$ , we have:

$$P = \begin{bmatrix} \mathbf{1}_{N_{\text{el}}} & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathcal{T}_P \mathcal{M}_{N_{\text{el}}} := \left\{ X = \begin{bmatrix} 0 & \times^T \\ \times & 0 \end{bmatrix} \right\};$$

- $H(P) := \nabla E(P)$  and  $K(P) := \Pi_P \nabla^2 E(P) \Pi_P$ .

- $\mathfrak{R} : \mathcal{H} \rightarrow \mathcal{M}_{N_{el}}$  is a retraction s.t.

$$\mathfrak{R}(P + \delta P) = P + \Pi_P \delta P + O(\delta P^2) \quad \text{for } P \in \mathcal{M}_{N_{el}}.$$



$$\min_{P \in \mathcal{M}_{N_{\text{el}}}} E(P) = \text{Tr}(H_0 P) + E_{\text{nl}}(P),$$

$$\mathcal{M}_{N_{\text{el}}} := \{P \in \mathcal{H} \mid \text{Tr}(P) = N_{\text{el}}, P^2 = P\}.$$

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

**Assumption 2**  $P_* \in \mathcal{M}_{N_{\text{el}}}$  is a nondegenerate local minimizer in the sense that there exists some  $\eta > 0$  such that, for  $P \in \mathcal{M}_{N_{\text{el}}}$  in a neighbourhood of  $P_*$ , we have

$$E(P) \geq E(P_*) + \eta \|P - P_*\|_F^2.$$

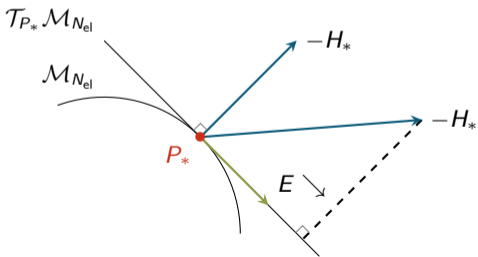
Let  $H_* := H(P_*)$  and  $K_* := K(P_*)$ .

# First-order condition

$$\min_{P \in \mathcal{M}_{N_{\text{el}}}} E(P) = \text{Tr}(H_0 P) + E_{\text{nl}}(P)$$

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

$$P_* H_* (1 - P_*) = (1 - P_*) H_* P_* = 0.$$



# First-order condition

$$\min_{P \in \mathcal{M}_{N_{\text{el}}}} E(P) = \text{Tr}(H_0 P) + E_{\text{nl}}(P)$$

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

$$P_* H_* (1 - P_*) = (1 - P_*) H_* P_* = 0.$$

- $[H_*, P_*] = 0 \Rightarrow H_*$  and  $P_*$  can be codiagonalized;
- if  $(\varphi_i)_{1 \leq i \leq N_b}$  is an o.n.b. of eigenvectors of  $H_*$  ordered by nondecreasing eigenvalues, then  $P_* = \sum_{i \in \text{occ}} \varphi_i \varphi_i^T$ , with  $\text{occ}$  the set of occupied orbitals;
- $\text{occ} \subset \{1, \dots, N_b\}$  and  $|\text{occ}| = N_{\text{el}}$ :
  - $\text{occ} = \{1, \dots, N_{\text{el}}\}$ : *Aufbau* principle;
  - $\text{occ} = \{1, \dots, N_{\text{el}}\}$  and  $\varepsilon_{N_{\text{el}}} < \varepsilon_{N_{\text{el}}+1}$ : strong *Aufbau* principle.

In the decomposition  $\mathcal{H} = \text{Ran}(P_*) \oplus \text{Ran}(1 - P_*)$ , assuming the *Aufbau principle*

$$H_* = \begin{array}{ccc} \leftarrow & \text{occ} & \rightarrow \overline{\text{occ}} \\ \begin{bmatrix} \varepsilon_1 & & & \\ & \ddots & & \\ & & \varepsilon_{N_{\text{el}}} & \\ & & & 0 \\ & 0 & & \ddots \end{bmatrix} & , & P_* = \begin{bmatrix} \text{occ} & \overline{\text{occ}} \\ 1_{N_{\text{el}}} & 0 \\ 0 & 0 \end{bmatrix} \end{array}$$

## Second-order condition

$$\min_{P \in \mathcal{M}_{N_{\text{el}}}} E(P) = \text{Tr}(H_0 P) + E_{\text{nl}}(P)$$

The second-order optimality condition reads

$$\forall X \in \mathcal{T}_{P_*} \mathcal{M}_{N_{\text{el}}}, \langle X, (\Omega_* + K_*) X \rangle_F \geq \eta \|X\|_F^2.$$

- Recall  $K_* = \Pi_{P_*} \nabla^2 E(P_*) \Pi_{P_*}$ ;
- the operator  $\Omega_* : \mathcal{T}_{P_*} \mathcal{M}_{N_{\text{el}}} \rightarrow \mathcal{T}_{P_*} \mathcal{M}_{N_{\text{el}}}$  is defined by,

$$\forall X \in \mathcal{T}_{P_*} \mathcal{M}_{N_{\text{el}}}, \quad \Omega_* X := -[P_*, [H_*, X]].$$

- $\Omega_* + K_*$  can be interpreted as the Hessian of the energy on the manifold,  $\Omega_*$  represents the influence of the curvature. Can also be seen as the Hessian of the Lagrangian.



**Proof:** Let  $X \in \mathcal{T}_{P_*} \mathcal{M}_{N_{el}}$ ,  $I$  be a real interval containing 0 and  $\gamma : I \rightarrow \mathcal{M}_{N_{el}}$  be a smooth path such that  $\gamma(0) = P_*$  and  $\dot{\gamma}(0) = X$ .

$$\begin{aligned} E(\gamma(t)) &= E(P_*) + t \langle H_*, X \rangle_F \\ &\quad + \frac{t^2}{2} \left( \langle H_*, \ddot{\gamma}(0) \rangle_F + \langle X, \nabla^2 E(P_*) X \rangle_F \right) + o(t^2) \\ &= E(P_*) + \frac{t^2}{2} \left( \langle H_*, \ddot{\gamma}(0) \rangle_F + \langle X, K_* X \rangle_F \right) + o(t^2) \end{aligned}$$

$\ddot{\gamma}(0)$  is unknown, but differentiating  $\gamma(t)^2 = \gamma(t)$  at  $t = 0$ , we get

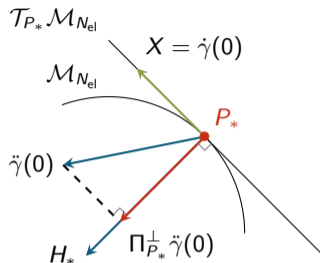
$$P_* \ddot{\gamma}(0) + \ddot{\gamma}(0) P_* + 2X^2 = \ddot{\gamma}(0),$$

from which we obtain

$$\frac{1}{2} P_* \ddot{\gamma}(0) P_* = -P_* X (1 - P_*) X P_*, \quad \frac{1}{2} (1 - P_*) \ddot{\gamma}(0) (1 - P_*) = (1 - P_*) X P_* X (1 - P_*).$$

After some algebra,

$$\langle H_*, \ddot{\gamma}(0) \rangle_F = \text{Tr} \left( X (\Omega_* X) \right) \quad \text{where} \quad \Omega_* X = -[P_*, [H_*, X]]. \quad \square$$



# Structure of $\Omega_*$

Let  $(\varphi_i, \varepsilon_i)_{1 \leq i \leq N_b}$  be an eigendecomposition of  $H_*$ . Then

- for  $i \in \text{occ}$  and  $a \notin \text{occ}$

$$(\Omega_* X)_{ia} = (\varepsilon_a - \varepsilon_i) X_{ia} \quad \text{and} \quad (\Omega_* X)_{ai} = (\varepsilon_a - \varepsilon_i) X_{ai};$$

- the gap  $\min_{a \notin \text{occ}} \varepsilon_a - \max_{i \in \text{occ}} \varepsilon_i$  is the smallest eigenvalue of  $\Omega_*$ .

**Remark:** if the *Aufbau* principle is satisfied, then the gap is  $\varepsilon_{N_{\text{el}}+1} - \varepsilon_{N_{\text{el}}}$ .

## The broader picture

With  $R(P) = \Pi_P \nabla E(P)$  and  $P_*$  such that  $R(P_*) = 0$  we then have:

- **Convergence analysis (Chapter 2):** the behavior of  $P^{k+1} = P^k - \beta R(P^k)$  depends on  $1 - \beta(\mathbf{\Omega}_* + \mathbf{K}_*)$ .
- **Error control (Chapter 3):**  $P - P_* \approx (\mathbf{\Omega}_* + \mathbf{K}_*)^{-1} R(P)$ .
- **Response calculations (Chapter 4):** if  $R$  depends on a parameter  $\varepsilon$ , then the solution to  $R(P_*(\varepsilon), \varepsilon) = 0$  satisfies

$$\left. \frac{\partial P_*}{\partial \varepsilon} \right|_{\varepsilon=0} = -(\mathbf{\Omega}_* + \mathbf{K}_*)^{-1} \left. \frac{\partial R}{\partial \varepsilon} \right|_{\varepsilon=0}.$$

- 1 Introduction
- 2 Mathematical framework
- 3 Convergence analysis of direct minimization and SCF iterations – Chapter 2**
- 4 Practical error bounds for quantities of interest – Chapter 3
- 5 Numerical stability of response property calculations – Chapter 4
- 6 DFTK and perspectives

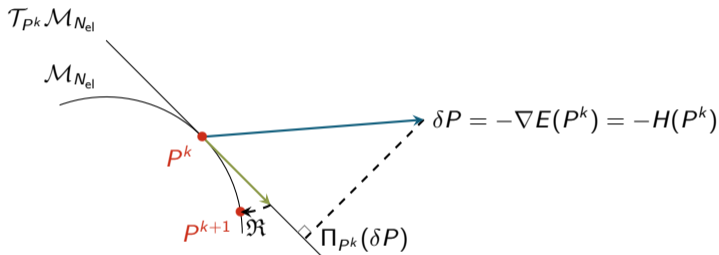
# Solving problem (1)

Problem	Iterative method
Eigenvalue problem	<b>Gradient descent</b> , Lanczos, LOBPCG
SCF (nonlinear eigenvalue problem)	<b>Damping</b> , Anderson acceleration, DIIS
Direct minimization	<b>Projected gradient descent</b> , CG, LBFGS

Each iterative method has already been analyzed in the literature  $\rightsquigarrow$  compare simplest representative of each class.

# Projected gradient descent

Solve directly (1) with a projected gradient algorithm:



**Data:**  $P^0 \in \mathcal{M}_{N_{el}}$

**while** convergence not reached **do**

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

**end**

# Convergence of projected gradient descent

## Theorem (Classical result)

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

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

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

$\rightsquigarrow$  in the linear case  $\mathbf{K}_* = 0$  and the spectral radius depends only on  $\|\mathbf{\Omega}_*\|_{\text{op}} = \varepsilon_{N_b} - \varepsilon_1 \rightarrow \infty$  when  $N_b \rightarrow \infty$ : known conditioning issues for gradient descents.

# Euler–Lagrange equations

Take the constrained minimization problem on  $\mathcal{M}_{N_{\text{el}}}$

$$\inf_{P \in \mathcal{M}_{N_{\text{el}}}} E(P) = \text{Tr}(H_0 P) + E_{\text{nl}}(P).$$

**Recall linear case**

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

↓

linear eigenvalue problem

$$\begin{cases} H_0 \varphi_n = \varepsilon_n \varphi_n \\ \varphi_n^T \varphi_m = \delta_{nm}, \\ P = \sum_{n=1}^{N_{\text{el}}} \varphi_n \varphi_n^T \end{cases}$$

**Nonlinear case**

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

↓

nonlinear eigenvalue problem

$$\begin{cases} (H_0 + \nabla E_{\text{nl}}(P)) \varphi_n = \varepsilon_n \varphi_n, \\ \varphi_n^T \varphi_m = \delta_{nm}, \\ P = \sum_{n=1}^{N_{\text{el}}} \varphi_n \varphi_n^T. \end{cases}$$



# Self-consistent field (SCF)

This leads to consider the following iterations:

- Set a starting point  $P^0 \in \mathcal{M}_{N_{\text{el}}}$ ;
- solve the linear eigenvalue problem for  $H(P^k) = H_0 + \nabla E_{\text{nl}}(P^k)$ : 
$$\begin{cases} H(P^k)\varphi_n^k = \varepsilon_n^k\varphi_n^k, \\ (\varphi_n^k)^T \varphi_m^k = \delta_{nm}, \end{cases}$$
- build the density matrix  $P^{k+1} = \sum_{n=1}^{N_{\text{el}}} \varphi_n^k (\varphi_n^k)^T$ ;
- solve the linear eigenvalue problem for  $H(P^{k+1})$ , and so on until convergence.

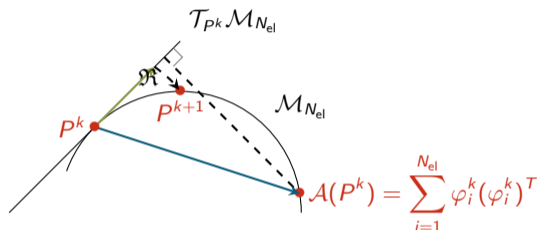
**Theorem (quadratic case, Cancès & Le Bris '00, Levitt '12)**

*The sequence  $(P^k)_{k \in \mathbb{N}}$  generated by this algorithm satisfies one of the two following properties:*

- *either  $(P^k)_{k \in \mathbb{N}}$  converges to an Aufbau solution to the HF equations;*
- *or  $(P^k)_{k \in \mathbb{N}}$  oscillates between two states, none of them being an Aufbau solution to the HF equations.*

# Damped SCF

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



**Data:**  $P^0 \in \mathcal{M}_{N_{el}}$

**while** convergence not reached **do**

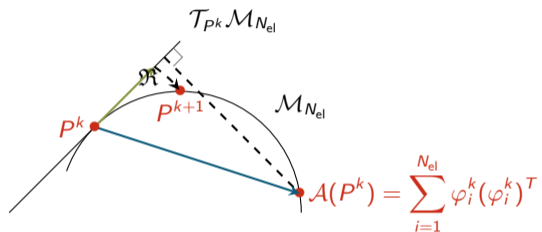
$$\text{solve } \begin{cases} H(P^k) \varphi_n^k = \varepsilon_n^k \varphi_n^k, & \varepsilon_1^k \leq \dots \leq \varepsilon_{N_{el}}^k < \varepsilon_{N_{el}+1}^k \\ (\varphi_n^k)^T \varphi_m^k = \delta_{nm}, \end{cases} ;$$

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

**end**

# Damped SCF

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



**Data:**  $P^0 \in \mathcal{M}_{N_{el}}$

**while** convergence not reached **do**

$$\text{solve } \begin{cases} H(P^k) \varphi_n^k = \varepsilon_n^k \varphi_n^k, & \varepsilon_1^k \leq \dots \leq \varepsilon_{N_{el}}^k < \varepsilon_{N_{el}+1}^k \\ (\varphi_n^k)^T \varphi_m^k = \delta_{nm}, \end{cases} ;$$

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

**end**

# Convergence of damped SCF

## Theorem (Cancès, Kemlin & Levitt '21)

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

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

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

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

linearly converge to  $P_*$ , with asymptotic rate the spectral radius of  $1 - \beta \mathbf{J}_{\text{SCF}}$  where

$$\mathbf{J}_{\text{SCF}} := 1 + \Omega_*^{-1} \mathbf{K}_*.$$

$\rightsquigarrow$  consistent with the linear case  $\mathbf{K}_* = 0$  for which we have a linear eigenvalue problem  $H_0 \varphi_n = \varepsilon_n \varphi_n$ : the SCF converges in one iteration.

## What did we learn ?

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

- Gradient descent:  $\mathbf{J}_{\text{grad}} = \mathbf{\Omega}_* + \mathbf{K}_*$  is sensitive to the spectral radius of  $\mathbf{H}_*$ ;
- SCF:  $\mathbf{J}_{\text{SCF}} = \mathbf{1} + \mathbf{\Omega}_*^{-1} \mathbf{K}_*$  is sensitive to the gap.

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);
- in practice, the choice depends on the convergence rate but also on the cost of each step which depends on the context (quantum chemistry vs condensed matter).

E. Cancès, G. Kемlin, and A. Levitt. Convergence Analysis of Direct Minimization and Self-Consistent Iterations. *SIAM Journal on Matrix Analysis and Applications*, 42(1):243-274, 2021.

Problem	matrix
Linear eigenvalue problem	$\mathbf{\Omega}_*$
Damped SCF	$\mathbf{1} + \mathbf{\Omega}_*^{-1} \mathbf{K}_*$
Gradient Descent	$\mathbf{\Omega}_* + \mathbf{K}_*$

- 1 Introduction
- 2 Mathematical framework
- 3 Convergence analysis of direct minimization and SCF iterations – Chapter 2
- 4 Practical error bounds for quantities of interest – Chapter 3**
- 5 Numerical stability of response property calculations – Chapter 4
- 6 DFTK and perspectives

## Error control in the literature

- Error control for eigenvalues of linear operators is already well established: initially in the 50s (*e.g.* Kato–Temple bound, Forsythe (1954), Weinberger (1956), Bazley and Fox (1961)), then recent progress in the past decades for elliptic operators with the FEM (see *e.g.* Hu, Huang, Lin and Shen (2014), Larson (2000), Liu (2015)).
- Recent progress for the particular case of electronic structure (see works by Cancès, Dusson, Maday, Stamm, Vohralík, Levitt, Herbst. . .).
- For nonlinear models, a few results exist, but mainly for simple models (*e.g.* Gross–Pitaevskii, see Maday and Dusson (2017), see also Chen, He and Zhou (2011)).
- Error control can be used to design adaptive methods (see Dai, Pan, Yang and Zhou (2021) for linear eigenvalue problems with plane-wave discretization or Liu, Chen, Dusson, Fang and Gao (2022) for a recent application to Kohn–Sham models).
- No results on error control for quantities of interest.

↪ We focus here on providing error estimates for generic nonlinear models (*e.g.* Kohn–Sham DFT) and for quantities of interest (*e.g.* forces).

# Linearization

Recall that  $\mathbf{\Omega}_* + \mathbf{K}_*$  is the Jacobian of  $P \mapsto R(P) = \Pi_P H(P)$  at  $P_*$ . Thus, at first order in  $\|P - P_*\|_F^2$ ,

$$\Pi_P H(P) \approx \Pi_{P_*} H_* + (\mathbf{\Omega}_* + \mathbf{K}_*)(P - P_*).$$

As  $\Pi_{P_*} H_* = 0$ , with  $R(P)$  the residual,

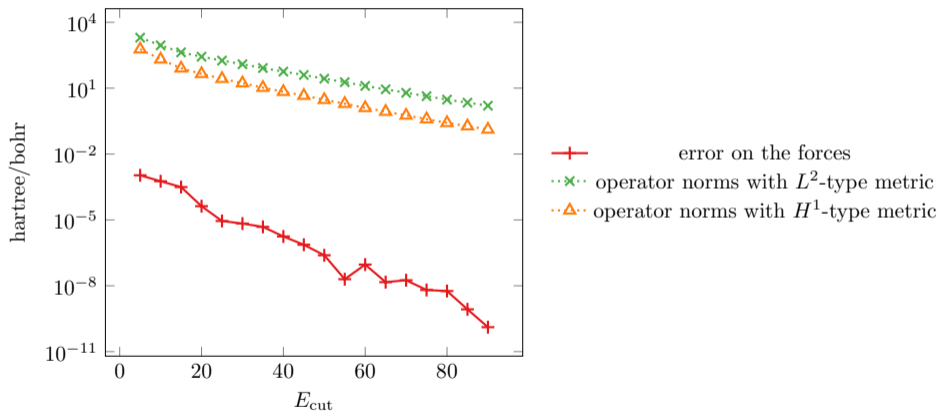
$$\Pi_P(P - P_*) \approx (\mathbf{\Omega}_* + \mathbf{K}_*)^{-1} R(P)$$

For quantity of interest  $F(P)$ :

$$|F(P) - F_*| \leq \|dF(P_*)\|_{\text{op}} \|(\mathbf{\Omega}_* + \mathbf{K}_*)^{-1}\|_{\text{op}} \|R(P)\|_F.$$



Error on the forces for a silicon crystal:  $E_{\text{cut}}$  defines the plane-wave variational approximation space.

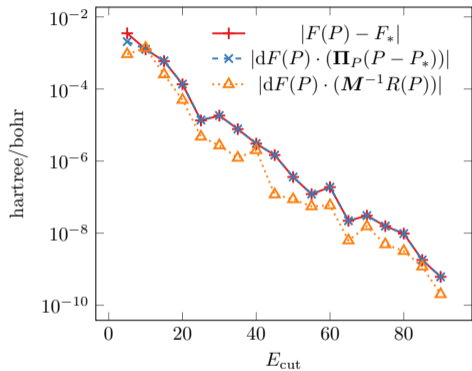


Replace the error  $F(P) - F_*$  by  $dF(P) \cdot (\Pi_P(P - P_*))$ .

↪ Good, but not usable in practice ( $P_*$  is unknown).

Replace  $P - P_*$  by  $M^{-1}R(P)$ , with  $M \sim -\frac{1}{2}\Delta + 1$ .

↪ Better, but still not satisfying.

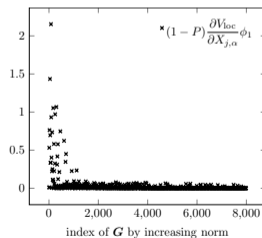
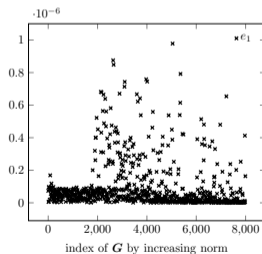
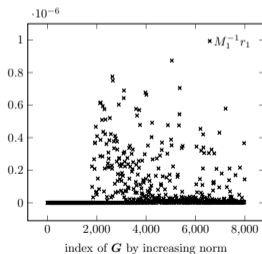


# Frequency splitting

Let  $P \in \mathcal{M}_{N_{\text{el}}}$ , then  $\mathcal{T}_P \mathcal{M}_{N_{\text{el}}}$  can be split into low and high frequencies:

$$\mathcal{T}_P \mathcal{M}_{N_{\text{el}}} = \underbrace{\Pi_{E_{\text{cut}}} \mathcal{T}_P \mathcal{M}_{N_{\text{el}}}}_{\text{low frequencies}} \oplus \underbrace{\Pi_{E_{\text{cut}}}^\perp \mathcal{T}_P \mathcal{M}_{N_{\text{el}}}}_{\text{high frequencies}}.$$

If  $P$  is a solution of the variational problem for a given  $E_{\text{cut}}$ , then  $R(P), M^{-1}R(P) \in \Pi_{E_{\text{cut}}}^\perp \mathcal{T}_P \mathcal{M}_{N_{\text{el}}}$ .



$\rightsquigarrow dF(P)$  is mainly supported on low frequencies.

## Enhanced error bounds

We decompose the error/residual relation onto  $\Pi_{E_{\text{cut}}}\mathcal{T}_P\mathcal{M}_{N_{\text{el}}}\oplus\Pi_{E_{\text{cut}}}^\perp\mathcal{T}_P\mathcal{M}_{N_{\text{el}}}$  to get

$$\begin{bmatrix} (\mathbf{\Omega} + \mathbf{K})_{11} & (\mathbf{\Omega} + \mathbf{K})_{12} \\ (\mathbf{\Omega} + \mathbf{K})_{21} & (\mathbf{\Omega} + \mathbf{K})_{22} \end{bmatrix} \begin{bmatrix} P_1 - P_{*1} \\ P_2 - P_{*2} \end{bmatrix} = \begin{bmatrix} R_1 \\ R_2 \end{bmatrix}.$$

As the kinetic energy is dominating for high-frequencies, we approximate

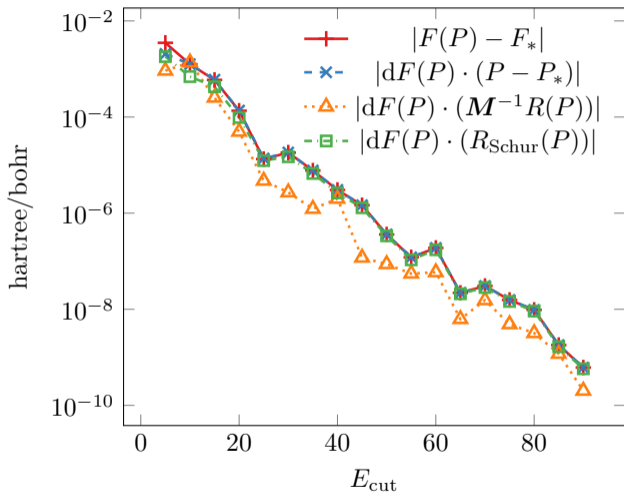
$$(\mathbf{\Omega} + \mathbf{K})_{21} \approx 0 \quad \text{and} \quad (\mathbf{\Omega} + \mathbf{K})_{22} \approx \mathbf{M}_{22},$$

and thus

$$\begin{bmatrix} (\mathbf{\Omega} + \mathbf{K})_{11} & (\mathbf{\Omega} + \mathbf{K})_{12} \\ 0 & \mathbf{M}_{22} \end{bmatrix} \begin{bmatrix} P_1 - P_{*1} \\ P_2 - P_{*2} \end{bmatrix} = \begin{bmatrix} R_1 \\ R_2 \end{bmatrix}.$$

This yields a new residual, which requires only an inversion on the coarse grid ( $\mathbf{M}_{22}$  being easy to invert):

$$R_{\text{Schur}}(P) = \begin{bmatrix} (\mathbf{\Omega} + \mathbf{K})_{11}^{-1} (R_1 - (\mathbf{\Omega} + \mathbf{K})_{12} \mathbf{M}_{22}^{-1} R_2) \\ \mathbf{M}_{22}^{-1} R_2 \end{bmatrix}.$$



## What did we learn ?

- The asymptotic regime is quickly established;
- error estimates based on operator norms are not good;
- using a Schur complement to couple high and low frequencies clearly enhances the approximation of the error;
- we can either compute error bounds or enhance the accuracy of the QoI;
- similar results are observed for more sophisticated systems.
- **Limits:** we do not have guaranteed bounds, but useful in practice, valid asymptotically and for a cost comparable to a SCF cycle (solving  $\Omega + \mathbf{K}$ ).

E. Cancès, G. Dusson, G. Kemplin, and A. Levitt. Practical error bounds for properties in plane-wave electronic structure calculations. *SIAM Journal on Scientific Computing*, 44(5):B1312-B1340, 2022.

- 1 Introduction
- 2 Mathematical framework
- 3 Convergence analysis of direct minimization and SCF iterations – Chapter 2
- 4 Practical error bounds for quantities of interest – Chapter 3
- 5 Numerical stability of response property calculations – Chapter 4**
- 6 DFTK and perspectives

## Response calculations

DFT is useful to compute ground-state properties, but most of quantities of interest depend on the response of the system to an external perturbation (polarizabilities, magnetic susceptibilities, phonons. . . )  $\rightsquigarrow$  DFPT.

Assume  $\delta H$  is an external perturbation and let  $P(\varepsilon)$  solve  $R(P, \varepsilon) := \Pi_P(H(P) + \varepsilon\delta H) = 0$ . Then

$$\left. \frac{\partial P}{\partial \varepsilon} \right|_{\varepsilon=0} = -J(P(0))^{-1} \left. \frac{\partial R}{\partial \varepsilon} \right|_{\varepsilon=0}.$$

Recall that  $J(P(0)) = \mathbf{\Omega}_* + \mathbf{K}_*$  and with  $\delta P = \left. \frac{\partial P}{\partial \varepsilon} \right|_{\varepsilon=0}$ , we obtain

$$\delta P = -(\mathbf{\Omega}_* + \mathbf{K}_*)^{-1} \delta H \quad \Leftrightarrow \quad \boxed{\delta P = (1 - \chi_0 \mathbf{K}_*)^{-1} \chi_0 \delta H}$$

where  $\chi_0 = -\mathbf{\Omega}_*^{-1}$  is the 4 points independent-particle susceptibility operator<sup>2</sup>.

$\rightsquigarrow$  Efficient computations of  $\delta P = \chi_0 \delta H$  are required.

<sup>2</sup>S. Baroni, S. de Gironcoli, A. Dal Corso, and P. Giannozzi. Phonons and related crystal properties from density-functional perturbation theory. *Reviews of Modern Physics*, 73(2):515–562, 2001.



# Insulators and semi-conductors

$P, \delta P$  are not tractable, in practice we use orbitals:

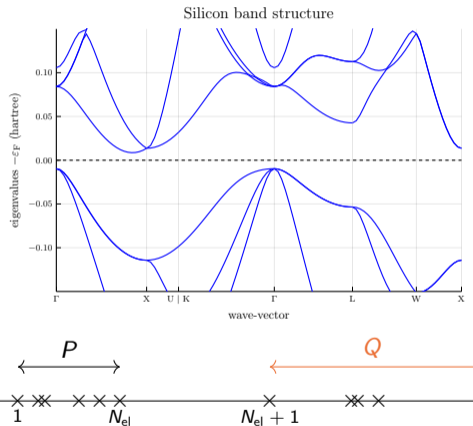
$$P = \sum_{n=1}^{N_{\text{el}}} |\varphi_n\rangle \langle \varphi_n|$$

$$\delta P = \sum_{n=1}^{N_{\text{el}}} |\varphi_n\rangle \langle \delta\varphi_n| + |\delta\varphi_n\rangle \langle \varphi_n|$$

with  $(\delta\varphi_n)_{1 \leq n \leq N_{\text{el}}}$  uniquely defined under the constraint  $\langle \varphi_m, \delta\varphi_n \rangle = 0$  for any  $n, m$ . Then, with  $Q = 1 - P$ , applying  $\chi_0$  leads to the resolution of the Sternheimer equation

$$Q(H_* - \varepsilon_n)Q\delta\varphi_n = -Q\delta H\varphi_n, \quad \forall n = 1, \dots, N_{\text{el}}.$$

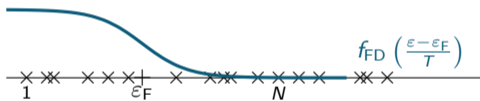
↪ Positive gap makes it easy for insulators and semi-conductors.



## Metals

Different context: introduce numerical temperature to ensure convergence

$$P = \sum_{n=1}^{N_b} f_n |\varphi_n\rangle \langle \varphi_n| \approx \sum_{n=1}^N f_n |\varphi_n\rangle \langle \varphi_n| \quad \text{with} \quad f_n = f_{\text{FD}} \left( \frac{\varepsilon_n - \varepsilon_F}{T} \right) \in [0, 1].$$



Again, in practice we use orbitals:  $\delta P = \sum_{n=1}^N f_n (|\delta\varphi_n\rangle \langle \delta\varphi_n| + |\delta\varphi_n\rangle \langle \varphi_n|) + \delta f_n |\varphi_n\rangle \langle \varphi_n|.$

↪ No uniqueness: gauge choices have to be made.

↪ How to define  $\Omega_*$  in this context? *Via*  $\chi_0!$

First, charge conservation ( $\text{Tr}(\delta P) = 0$ ) helps choosing  $\delta f_n$ . Then, for all  $n = 1, \dots, N$ :

$$f_n \delta \varphi_n = f_n \delta \varphi_n^P + f_n \delta \varphi_n^Q$$

sum-over-states formula

Sternheimer equation

■  $f_n \delta \varphi_n^P = \sum_{m=1}^N \Gamma_{mn} \varphi_m$  where<sup>a</sup>

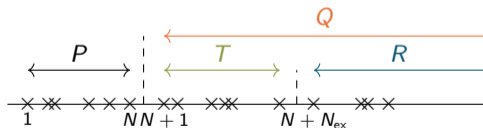
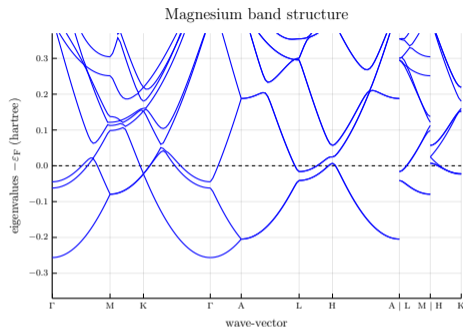
$$\Gamma_{mn} + \overline{\Gamma_{nm}} = \frac{f_n - f_m}{\varepsilon_n - \varepsilon_m} \langle \varphi_m, \delta H \varphi_n \rangle.$$

■  $\delta \varphi_n^Q$  solves

$$Q(H_* - \varepsilon_n)Q\delta\varphi_n^Q = -Q\delta H\varphi_n$$

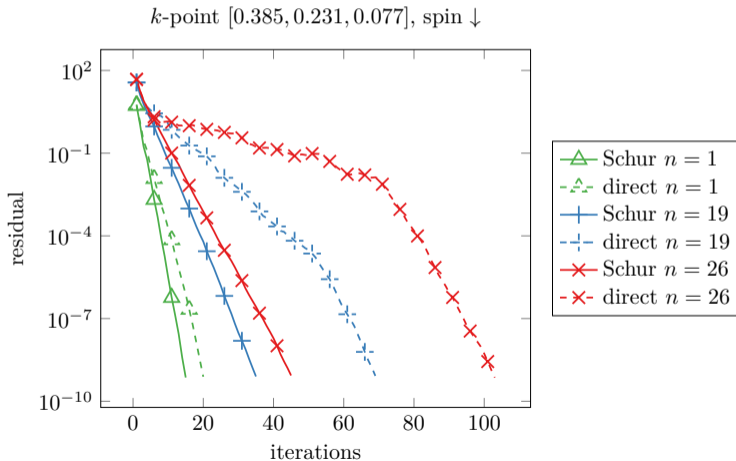
↪ Possibly very ill-conditioned: Schur complement with  $N_{\text{ex}}$  discarded orbitals to solve a better conditioned system.

<sup>a</sup>We use the convention  $(f_n - f_m)/(\varepsilon_n - \varepsilon_m) = \frac{1}{T} f'_{\text{FD}}((\varepsilon_n - \varepsilon_m)/T) =: f'_n$ .



# Resolution of the Sternheimer equation for Heusler compounds.

$\text{Fe}_2\text{MnAl}$ : More than 40% less Hamiltonian applications in total.



## What did we learn ?

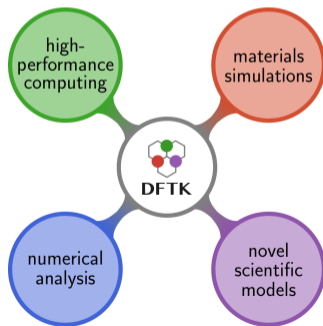
- Insulators are easy:  $\delta\varphi_n \in \text{Span}(\varphi_m)_{N+1 \leq m}$  and the Sternheimer equation is usually well-conditioned;
- metals are more difficult:  $\delta\varphi_n = \delta\varphi_n^P + \delta\varphi_n^Q$ 
  - $\delta\varphi_n^P$  requires a gauge choice and we derived a common framework from the literature which ensures numerical stability (computational time is negligible);
  - $\delta\varphi_n^Q$  solves the ill-conditioned Sternheimer equation in  $\text{Span}(\varphi_m)_{N+1 \leq m}$  and we enhanced its resolution through a Schur complement. Numerical experiments give satisfying results, even for challenging systems.

E. Cancès, M. F. Herbst, G. Kемlin, A. Levitt, and B. Stamm. Numerical stability and efficiency of response property calculations in density functional theory. *arXiv:2210.04512*, 2022.

# DFTK

## Main contributions:

- implementation of a Newton solver thanks to the linearization of the KS equations;
- developments of error estimators for interatomic forces;
- implementation of a framework to perform response calculations, a cornerstone to the use of Automatic Differentiation in DFTK.



## Perspectives

- Insulators are well understood now.
- For metals, the situation is more challenging and possible future works include:
  - comparing direct minimization and SCF;
  - extending error control for forces to finite temperature systems;
  - combining Chapters 3 and 4 to derive error estimates for properties that require response calculations;
  - choosing the appropriate number of extra bands to perform SCF and response calculations with metals.
- Designing adaptive methods using the tools developed for error control.
- More generally, the *a posteriori* numerical methods we proposed require to set different parameters (e.g. two-grids methods) whose choice is mainly empirical at the moment. Understanding how to optimize these parameters would be of high interest for practical applications.

Thanks for your attention !