# Practical error bounds for properties in plane-wave electronic structure calculations

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

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

CECAM UQ, June 22nd 2022, Lausanne

1 Introduction

2 Mathematical framework
- Structure of the manifold
- Super-operators
- Numerical setting

3 Crude error bounds using linearization
- Linearization in the asymptotic regime
- Error bounds based on operator norms
- Error bounds for the forces

4 Enhanced error bounds based on frequencies splitting

5 Numerical examples

1 Introduction

2 Mathematical framework
- Structure of the manifold
- Super-operators
- Numerical setting

3 Crude error bounds using linearization
- Linearization in the asymptotic regime
- Error bounds based on operator norms
- Error bounds for the forces

4 Enhanced error bounds based on frequencies splitting

5 Numerical examples

# Quantum mechanics of noninteracting electrons

We consider the stationary Schrödinger equation

$$\boxed{\begin{cases} H_0\varphi_i = \varepsilon_i\varphi_i, \ \ \varepsilon_1 \leqslant \cdots \leqslant \varepsilon_N, \\ \|\varphi_i\|_{\mathrm{L}^2} = 1, \end{cases} \qquad H_0 := -\frac{1}{2}\Delta + V}$$

where $\varphi_i$ is the wavefunction associated to electron $i$. Then,

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

# Numerical resolution

$$\boxed{\text{Find } \varphi_i \in \mathbb{C}^{\mathcal{N}}, \text{ s.t } H_0\varphi_i = \varepsilon_i\varphi_i, \quad \varepsilon_1 \leqslant \cdots \leqslant \varepsilon_N}$$

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

$$P := \sum_{i=1}^{N} |\varphi_i\rangle\langle\varphi_i| \in \mathbb{C}^{\mathcal{N}\times\mathcal{N}}_{\text{herm}}.$$

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

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

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

We have two equivalent problems:

$$\boxed{\begin{cases} H_0\varphi_i = \varepsilon_i\varphi_i, \ \varepsilon_1 \leqslant \cdots \leqslant \varepsilon_N, \\ \|\varphi_i\|_{\mathrm{L}^2} = 1, \end{cases} \qquad \Leftrightarrow \qquad \min_{P\in\mathcal{M}_N} \mathrm{Tr}(H_0P)}$$

where

$$\mathcal{M}_N := \left\{ P \in \mathbb{C}^{\mathcal{N}\times\mathcal{N}} \;\middle|\; P = P^*,\ \mathrm{Tr}(P) = N,\ P^2 = P \right\}$$

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

# General framework

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

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

where

- $P \in \mathbb{C}^{\mathcal{N}\times\mathcal{N}}_{\mathrm{herm}}$ is a density matrix;
- $H_0$ is the core Hamiltonian;
- $E_{\mathrm{nl}}$ models the electron-electron interaction depending on the model (Kohn-Sham DFT – local and semi-local functionals –, Hartree-Fock, Gross-Pitaevskii, . . . ).

$$\min_{P\in\mathcal{M}_N} E(P) = \mathrm{Tr}\,(H_0 P) + E_{\mathrm{nl}}(P),$$

$$\mathcal{M}_N := \left\{ P \in \mathbb{C}^{\mathcal{N}\times\mathcal{N}} \;\middle|\; P = P^*,\ \mathrm{Tr}(P) = N,\ P^2 = P \right\}.$$

In practice, the required $\mathcal{N}$ to achieve high precision is way too high. To solve this issue, we use subspaces of smaller dimension to compute a variational approximation of $P_*$, the reference solution in $\mathcal{M}_N$.

$\rightsquigarrow$ we focus on **discretization error**, but there are other sources (models, arithmetics, . . . )

In practice, the required $\mathcal{N}$ to achieve high precision is way too high. To solve this issue, we use subspaces of smaller dimension to compute a variational approximation of $P_*$, the reference solution in $\mathcal{M}_N$.

$\rightsquigarrow$ we focus on **discretization error**, but there are other sources (models, arithmetics, . . . )

**Question :**

How to evaluate the error made on quantities of interest (QoI) ? We focus here on the **energy** and the **forces**.

1 Introduction

2 Mathematical framework

- Structure of the manifold
- Super-operators
- Numerical setting

3 Crude error bounds using linearization

- Linearization in the asymptotic regime
- Error bounds based on operator norms
- Error bounds for the forces

4 Enhanced error bounds based on frequencies splitting

5 Numerical examples

# Assumptions

$$\boxed{\begin{gathered}\min_{P\in\mathcal{M}_N} E(P) = \mathrm{Tr}\,(H_0 P) + E_{\mathrm{nl}}(P),\\[1em] \mathcal{M}_N := \left\{ P \in \mathbb{C}^{\mathcal{N}\times\mathcal{N}} \;\middle|\; P = P^*,\ \mathrm{Tr}(P) = N,\ P^2 = P \right\}.\end{gathered}}$$

Let $\mathcal{H} := \left(\mathbb{C}^{\mathcal{N}\times\mathcal{N}}_{\mathrm{herm}}, \|\cdot\|_{\mathrm{F}}\right)$, endowed with the Frobenius scalar product $\mathrm{Tr}(A^* B)$.

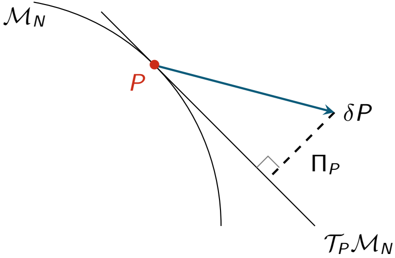
**Assumption 1** $E_{\mathrm{nl}} : \mathcal{H} \to \mathbb{R}$ is twice continuously differentiable, and thus so is $E$.

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

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

# Structure of the manifold: the tangent space

$\mathcal{M}_N$ is a smooth manifold, we can define its tangent space (it is a $\mathbb{R}$ vector space). $\Pi_P$ is the orthogonal projection on $\mathcal{T}_P\mathcal{M}_N$:



# First order condition

$$\min_{P \in \mathcal{M}_N} E(P) = \mathrm{Tr}\,(H_0 P) + E_{\mathrm{nl}}(P)$$

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

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

where $H_* := \nabla E(P_*)$.

In particular, $[H_*, P_*] = 0$.

# Second order condition

$$\min_{P \in \mathcal{M}_N} E(P) = \mathrm{Tr}\,(H_0 P) + E_{\mathrm{nl}}(P)$$

The second order optimality condition reads

$$\boxed{\forall\, X \in \mathcal{T}_{P_*}\mathcal{M}_N,\ \ \langle X, (\boldsymbol{\Omega}_* + \boldsymbol{K}_*)X\rangle_{\mathrm{F}} \geqslant \eta\, \|X\|_{\mathrm{F}}^2}.$$

- $\boldsymbol{K}_* := \Pi_{P_*} \nabla^2 E(P_*) \Pi_{P_*}$;
- the operator $\boldsymbol{\Omega}_* : \mathcal{T}_{P_*}\mathcal{M}_N \to \mathcal{T}_{P_*}\mathcal{M}_N$ is defined by,

$$\forall\, X \in \mathcal{T}_{P_*}\mathcal{M}_N, \quad \boldsymbol{\Omega}_* X := -[P_*, [H_*, X]].$$

# Second order condition

$$\min_{P\in\mathcal{M}_N} E(P) = \mathrm{Tr}\,(H_0 P) + E_{\mathrm{nl}}(P)$$

The second order optimality condition reads

$$\boxed{\forall\, X \in \mathcal{T}_{P_*}\mathcal{M}_N,\ \langle X, (\boldsymbol{\Omega}_* + \boldsymbol{K}_*)X\rangle_{\mathrm{F}} \geqslant \eta\, \|X\|_{\mathrm{F}}^2}.$$

- $\boldsymbol{K}_* := \Pi_{P_*}\nabla^2 E(P_*)\Pi_{P_*}$;
- the operator $\boldsymbol{\Omega}_* : \mathcal{T}_{P_*}\mathcal{M}_N \to \mathcal{T}_{P_*}\mathcal{M}_N$ is defined by,

$$\forall\, X \in \mathcal{T}_{P_*}\mathcal{M}_N, \quad \boldsymbol{\Omega}_* X := -[P_*, [H_*, X]].$$

$\rightsquigarrow$ $\boldsymbol{\Omega}_* + \boldsymbol{K}_*$ can be interpreted as the Hessian of the energy on the manifold, $\boldsymbol{\Omega}_*$ represents the influence of the curvature.

## Plane-wave DFT

Throughout the talk, we perform numerical tests in DFTK[1], a PW DFT tool-kit for Julia. In short:

- we consider a periodic system with lattice $\mathcal{R}$, $\omega$ is the unit cell and $\mathcal{R}^*$ the reciprocal lattice;
- we solve a variational approximation of the KS-DFT equations in the finite dimensional space

$$\mathcal{X}_{E_{\rm cut}} := \left\{ e_{\boldsymbol{G}},\ \boldsymbol{G} \in \mathcal{R}^* \ \middle|\ \frac{1}{2}\left|\boldsymbol{G}\right|^2 \leqslant E_{\rm cut} \right\},$$

where, for $\boldsymbol{G} \in \mathcal{R}^*$,

$$\forall\ \boldsymbol{r} \in \mathbb{R}^3, \quad e_{\boldsymbol{G}}(\boldsymbol{r}) := \frac{1}{\sqrt{|\omega|}} \exp\left(\mathrm{i}\boldsymbol{G} \cdot \boldsymbol{r}\right).$$

[1] `https://dftk.org`, developed by M. F. Herbst and A. Levitt.

# Numerical setting

- FCC phase of the silicon crystal, within LDA approximation and $2 \times 2 \times 2$ Brillouin zone discretization;
- we compute a reference solution for $E_{\mathrm{cut,ref}} = 125$ Ha $\Rightarrow E_{\mathrm{cut,ref}}$ defines $\mathcal{N}$ the size of the reference space and we obtain the reference orbitals $\Phi_*$, the energy $E_*$, density $\rho_*$, the forces $F_*$ on each atoms, etc. . .
- for smaller $E_{\mathrm{cut}}$'s, we compute the associated variational approximation and we measure the error on different quantities:

$$|E - E_*|\,, \quad \|\rho - \rho_*\|_{\mathrm{L}^2}\,, \quad |F - F_*|$$

1 Introduction

2 Mathematical framework

- Structure of the manifold
- Super-operators
- Numerical setting

3 Crude error bounds using linearization

- Linearization in the asymptotic regime
- Error bounds based on operator norms
- Error bounds for the forces

4 Enhanced error bounds based on frequencies splitting

5 Numerical examples

# Linearization: main idea

Assume you want to solve $R(x) = 0$ with $R$ a differentiable quantity, with Jacobian $J_R$. Then, around a solution $x_*$, it holds at first order

$$R(x) = R(x_*) + J_R(x_*)(x - x_*),$$

from which we deduce

$$\boxed{(x - x_*) \approx J_R(x_*)^{-1} R(x)}$$

**Newton's algorithm :**

$$\boxed{x^{k+1} = x^k - J_R(x^k)^{-1} R(x^k)}$$

# Linearization: application to our model

$\mathbf{\Omega}_* + \boldsymbol{K}_*$ is the Jacobian[2] of $P \mapsto \Pi_P H(P) = R(P)$ at $P_*$.

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

[2] Eric Cancès, Gaspard Kemlin, Antoine Levitt. Convergence analysis of direct minimization and self-consistent iterations. SIAM Journal of Matrix Analysis and Applications, 42(1):243–274 (2021).

# Linearization: application to our model

$\boldsymbol{\Omega}_* + \boldsymbol{K}_*$ is the Jacobian[2] of $P \mapsto \Pi_P H(P) = R(P)$ at $P_*$.

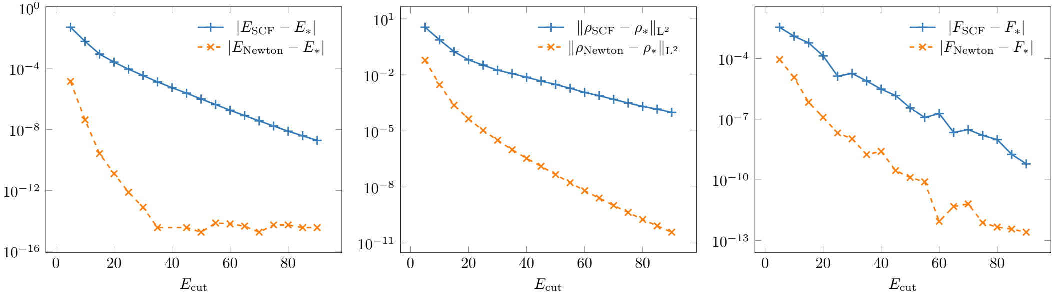
$$\boxed{\Pi_P(P - P_*) = (\boldsymbol{\Omega}_* + \boldsymbol{K}_*)^{-1} R(P)}$$

**Newton's algorithm :** extend the definition of $\boldsymbol{\Omega}$ and $\boldsymbol{K}$ outside of $P_*$ and let $\mathfrak{R}$ be a retraction to the manifold

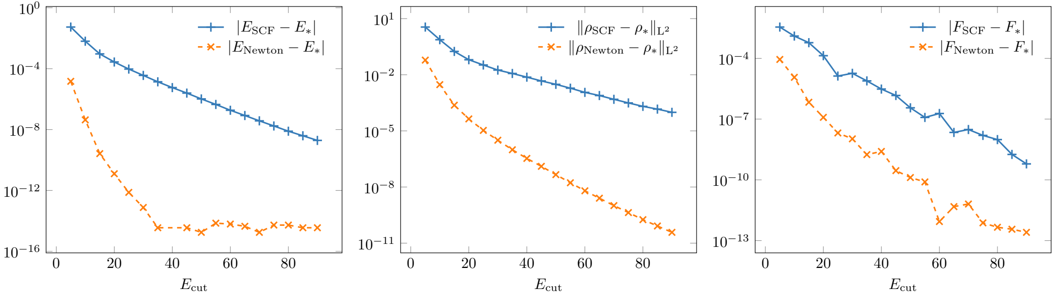
$$\boxed{P^{k+1} = \mathfrak{R}_{P^k}\left(P^k - \left(\boldsymbol{\Omega}(P^k) + \boldsymbol{K}(P^k)\right)^{-1} R(P^k)\right)}$$

[2] Eric Cancès, Gaspard Kemlin, Antoine Levitt. Convergence analysis of direct minimization and self-consistent iterations. SIAM Journal of Matrix Analysis and Applications, 42(1):243–274 (2021).

Compare DFTK QoI for given $E_{\mathrm{cut}} < E_{\mathrm{cut,ref}}$ and the QoI after one Newton step in the reference grid.



Compare DFTK QoI for given $E_{\text{cut}} < E_{\text{cut,ref}}$ and the QoI after one Newton step in the reference grid.



$\rightsquigarrow$ the asymptotic regime is quickly established: $\boxed{\Pi_P(P - P_*) = (\boldsymbol{\Omega}_* + \boldsymbol{K}_*)^{-1} R(P)}$

# Error bounds based on operator norms

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

**First crude bound :** $\|P - P_*\|_{\mathrm{F}}$ and $\|R(P)\|_{\mathrm{F}}$ cannot be directly compared (not the same unit) but we have

$$\begin{aligned} \|P - P_*\|_{\mathrm{F}} &\approx \|\Pi_P(P - P_*)\|_{\mathrm{F}} \\ &\leqslant \left\|(\boldsymbol{\Omega}_* + \boldsymbol{K}_*)^{-1}\right\|_{\mathrm{op}} \|R(P)\|_{\mathrm{F}} . \end{aligned}$$

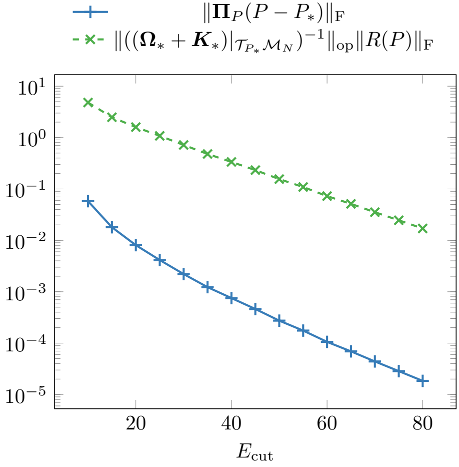
# Error bounds based on operator norms

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

**First crude bound :** $\|P - P_*\|_{\mathrm{F}}$ and $\|R(P)\|_{\mathrm{F}}$ cannot be directly compared (not the same unit) but we have

$$\begin{aligned} \|P - P_*\|_{\mathrm{F}} &\approx \|\Pi_P(P - P_*)\|_{\mathrm{F}} \\ &\leqslant \left\|(\boldsymbol{\Omega}_* + \boldsymbol{K}_*)^{-1}\right\|_{\mathrm{op}} \|R(P)\|_{\mathrm{F}} . \end{aligned}$$

$\rightsquigarrow$ the bounds are several orders of magnitude above the error. . .



# Error bounds based on operator norms

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

One can change the metric with $\boldsymbol{M} \approx 1 - \frac{1}{2}\Delta$

$$\left\| \boldsymbol{M}^{1/2} \Pi_P(P - P_*) \right\|_{\mathrm{F}}$$

$$\leqslant \left\| \boldsymbol{M}^{1/2} (\boldsymbol{\Omega}_* + \boldsymbol{K}_*)^{-1} \boldsymbol{M}^{1/2} \right\|_{\mathrm{op}} \left\| \boldsymbol{M}^{-1/2} R(P) \right\|_{\mathrm{F}} .$$

# Error bounds based on operator norms

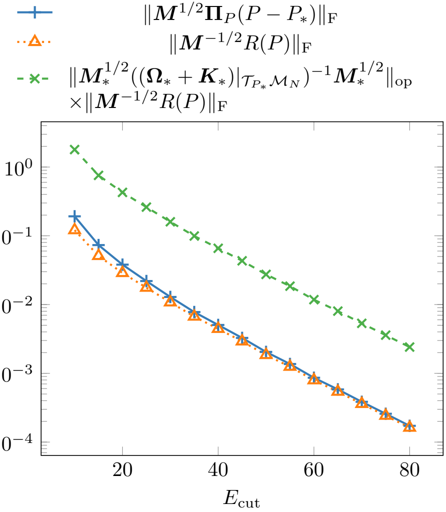
$$\Pi_P(P - P_*) = (\boldsymbol{\Omega}_* + \boldsymbol{K}_*)^{-1} R(P)$$

One can change the metric with $\boldsymbol{M} \approx 1 - \frac{1}{2}\Delta$

$$\left\| \boldsymbol{M}^{1/2} \Pi_P(P - P_*) \right\|_{\mathrm{F}}$$

$$\leqslant \left\| \boldsymbol{M}^{1/2} (\boldsymbol{\Omega}_* + \boldsymbol{K}_*)^{-1} \boldsymbol{M}^{1/2} \right\|_{\mathrm{op}} \left\| \boldsymbol{M}^{-1/2} R(P) \right\|_{\mathrm{F}} .$$

⇝ the bounds are several orders of magnitude above the error. . . but have the same rate
⇝ asymptotically $\left\| \boldsymbol{M}^{-1/2} R(P) \right\|_{\mathrm{F}} \sim \left\| \boldsymbol{M}^{1/2} \Pi_P(P - P_*) \right\|_{\mathrm{F}}$, though not upper bound nor guaranteed. The same holds for $\left\| \boldsymbol{M}^{-1} R(P) \right\|_{\mathrm{F}} \sim \| P - P_* \|_{\mathrm{F}}$.



# Error bounds for the forces

Forces are decomposed into two components (local and non-local)[3].

**Local forces:** Let $F_{j,\alpha}^{\mathrm{loc}}(P)$ be the local forces on atom $j$ in direction $\alpha$. It holds (at first order):

$$F_{j,\alpha}^{\mathrm{loc}}(P) - F_{j,\alpha}^{\mathrm{loc}}(P_*) = \mathrm{d}F_{j,\alpha}^{\mathrm{loc}}(P) \cdot \Pi_P(P - P_*);$$

$$\left|F_{j,\alpha}^{\mathrm{loc}}(P) - F_{j,\alpha}^{\mathrm{loc}}(P_*)\right| \leqslant \left\|\mathrm{d}F_{j,\alpha}^{\mathrm{loc}}(P)\right\|_{\mathcal{T}_P\mathcal{M}_N \to \mathbb{R}} \left\|P - P_*\right\|_{\mathrm{F}}.$$

[3] This comes from the pseudopotentials approximations and Hellmann-Faynman theorem.

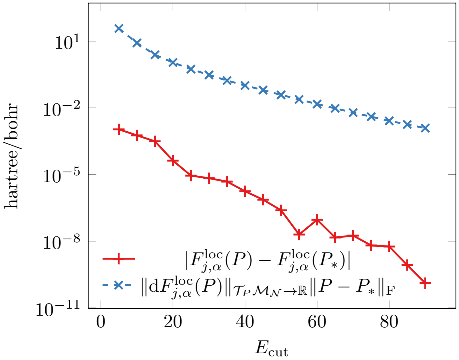
# Error bounds for the forces

Forces are decomposed into two components (local and non-local)[3].

**Local forces:** Let $F^{\mathrm{loc}}_{j,\alpha}(P)$ be the local forces on atom $j$ in direction $\alpha$. It holds (at first order):

$$F^{\mathrm{loc}}_{j,\alpha}(P) - F^{\mathrm{loc}}_{j,\alpha}(P_*) = \mathrm{d}F^{\mathrm{loc}}_{j,\alpha}(P)\cdot\Pi_P(P-P_*);$$

$$\left|F^{\mathrm{loc}}_{j,\alpha}(P) - F^{\mathrm{loc}}_{j,\alpha}(P_*)\right| \leqslant \left\|\mathrm{d}F^{\mathrm{loc}}_{j,\alpha}(P)\right\|_{\mathcal{T}_P\mathcal{M}_N\to\mathbb{R}} \|P-P_*\|_{\mathrm{F}}\,.$$



$\rightsquigarrow$ several orders of magnitude above !

[3] This comes from the pseudopotentials approximations and Hellmann-Faynman theorem.

# Error bounds for the forces

Forces are decomposed into two components (local and non-local)[3].

**Total forces :** Combining local and nonlocal forces on all atoms, we have $F(P) \in \mathbb{R}^{3N_{\#\text{atoms}}}$ and

$$F(P) - F(P_*) = \mathrm{d}F(P) \cdot \Pi_P(P - P_*).$$

$\rightsquigarrow$ What happens if we directly replace $\Pi_P(P - P_*)$ by $\boldsymbol{M}^{-1}R(P)$ in $\mathrm{d}F(P) \cdot \Pi_P(P - P_*)$?

[3] This comes from the pseudopotentials approximations and Hellmann-Faynman theorem.

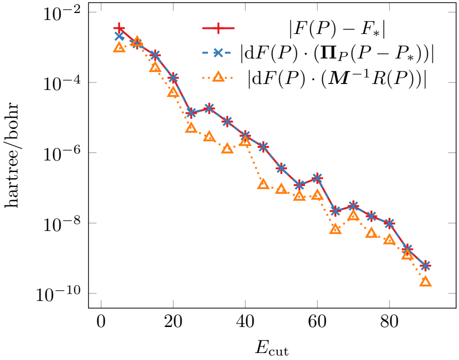
# Error bounds for the forces

Forces are decomposed into two components (local and non-local)[3].

**Total forces :** Combining local and nonlocal forces on all atoms, we have $F(P) \in \mathbb{R}^{3N_{\#\text{atoms}}}$ and

$$F(P) - F(P_*) = \mathrm{d}F(P) \cdot \Pi_P(P - P_*).$$

⇝ What happens if we directly replace $\Pi_P(P - P_*)$ by $\boldsymbol{M}^{-1}R(P)$ in $\mathrm{d}F(P) \cdot \Pi_P(P - P_*)$?



⇝ linearization quickly valid;

⇝ even if $\Pi_P(P - P_*)$ and $\boldsymbol{M}^{-1}R(P)$ are asymptotically equivalent, orange and blue do not match.

[3] This comes from the pseudopotentials approximations and Hellmann-Faynman theorem.

1 Introduction

2 Mathematical framework
- Structure of the manifold
- Super-operators
- Numerical setting

3 Crude error bounds using linearization
- Linearization in the asymptotic regime
- Error bounds based on operator norms
- Error bounds for the forces

4 Enhanced error bounds based on frequencies splitting

5 Numerical examples

# Frequency splitting

Let $P \in \mathcal{M}_N$, then $\mathcal{T}_P\mathcal{M}_N$ can be split into low and high frequencies. More precisely, given $E_{\text{cut}} < E_{\text{cut},ref}$, we have

$$
\begin{array}{ccccc}
\mathcal{T}_P\mathcal{M}_N & = & \Pi_{E_{\text{cut}}}\mathcal{T}_P\mathcal{M}_N & \oplus & \Pi_{E_{\text{cut}}}^{\perp}\mathcal{T}_P\mathcal{M}_N \\
\rotatebox{90}{$\in$} & & \rotatebox{90}{$\in$} & & \rotatebox{90}{$\in$} \\
X & = & X_1 & + & X_2 \\
\updownarrow & & \updownarrow & & \updownarrow \\
\psi & = & \psi_1 & + & \psi_2
\end{array}
$$

with $\psi_1 \in \mathcal{X}_{E_{\text{cut}}}$, $\psi_2 \in \mathcal{X}_{E_{\text{cut}}}^{\perp}$ and $\mathcal{X}_{E_{\text{cut},ref}} = \mathcal{X}_{E_{\text{cut}}} \oplus \mathcal{X}_{E_{\text{cut}}}^{\perp}$.

# Frequency splitting

Let $P \in \mathcal{M}_N$, then $\mathcal{T}_P\mathcal{M}_N$ can be split into low and high frequencies. More precisely, given $E_{\rm cut} < E_{{\rm cut},ref}$, we have

$$\begin{array}{ccccc}
\mathcal{T}_P\mathcal{M}_N & = & \Pi_{E_{\rm cut}}\mathcal{T}_P\mathcal{M}_N & \oplus & \Pi^{\perp}_{E_{\rm cut}}\mathcal{T}_P\mathcal{M}_N \\
\rotatebox{90}{$\in$} & & \rotatebox{90}{$\in$} & & \rotatebox{90}{$\in$} \\
X & = & X_1 & + & X_2 \\
\updownarrow & & \updownarrow & & \updownarrow \\
\psi & = & \psi_1 & + & \psi_2
\end{array}$$

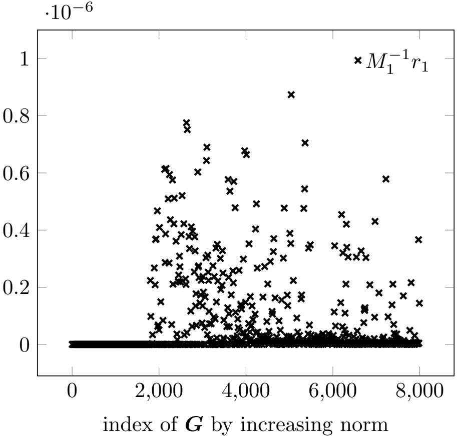
with $\psi_1 \in \mathcal{X}_{E_{\rm cut}}$, $\psi_2 \in \mathcal{X}^{\perp}_{E_{\rm cut}}$ and $\mathcal{X}_{E_{{\rm cut},ref}} = \mathcal{X}_{E_{\rm cut}} \oplus \mathcal{X}^{\perp}_{E_{\rm cut}}$.

If $P$ is a solution of the variational problem for a given $E_{\rm cut}$, then $R(P), \boldsymbol{M}^{-1}R(P) \in \Pi^{\perp}_{E_{\rm cut}}\mathcal{T}_P\mathcal{M}_N$ (not exactly true in practice because of numerical quadrature errors due to exchange-correlation terms.).

Let us analyze in details the computation of $F_{j,\alpha}^{\mathrm{loc}}(P)$: $F_{j,\alpha}^{\mathrm{loc}}(P) = -\operatorname{Tr}\left(\frac{\partial V_{\mathrm{loc}}}{\partial R_{j,\alpha}} P\right)$ so that computing $\mathrm{d}F_{j,\alpha}^{\mathrm{loc}}(P) \cdot X$ for $X \in \mathcal{T}_P\mathcal{M}_N$ reduces to the scalar product of $X$ against $\Pi_P \frac{\partial V_{\mathrm{loc}}}{\partial R_{j,\alpha}}$.

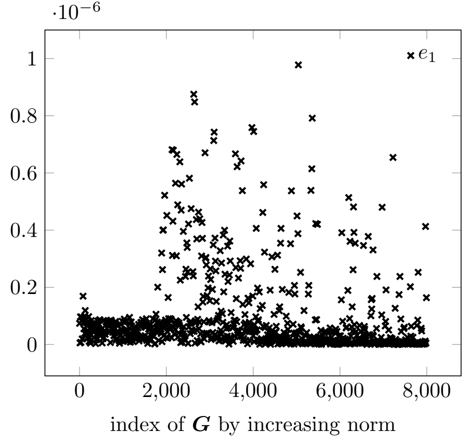
Let us analyze in details the computation of $F_{j,\alpha}^{\mathrm{loc}}(P)$: $F_{j,\alpha}^{\mathrm{loc}}(P) = -\operatorname{Tr}\left(\frac{\partial V_{\mathrm{loc}}}{\partial R_{j,\alpha}} P\right)$ so that computing $\mathrm{d}F_{j,\alpha}^{\mathrm{loc}}(P)\cdot X$ for $X \in \mathcal{T}_P\mathcal{M}_N$ reduces to the scalar product of $X$ against $\Pi_P \frac{\partial V_{\mathrm{loc}}}{\partial R_{j,\alpha}}$.

- $\boldsymbol{M}^{-1}R(P)$ is high frequencies;
- $\Pi_P(P - P_*)$ is mainly high frequencies but with low frequencies components;
- $\Pi_P \frac{\partial V_{\mathrm{loc}}}{\partial R_{j,\alpha}}$ is mainly low frequencies.



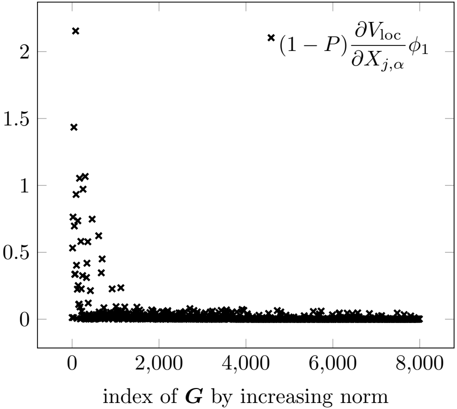
Let us analyze in details the computation of $F^{\mathrm{loc}}_{j,\alpha}(P)$: $F^{\mathrm{loc}}_{j,\alpha}(P) = -\operatorname{Tr}\left(\frac{\partial V_{\mathrm{loc}}}{\partial R_{j,\alpha}} P\right)$ so that computing $\mathrm{d}F^{\mathrm{loc}}_{j,\alpha}(P) \cdot X$ for $X \in \mathcal{T}_P\mathcal{M}_N$ reduces to the scalar product of $X$ against $\Pi_P \frac{\partial V_{\mathrm{loc}}}{\partial R_{j,\alpha}}$.

- $\boldsymbol{M}^{-1}R(P)$ is high frequencies;
- $\Pi_P(P - P_*)$ is mainly high frequencies but with low frequencies components;
- $\Pi_P \frac{\partial V_{\mathrm{loc}}}{\partial R_{j,\alpha}}$ is mainly low frequencies.



Let us analyze in details the computation of $F^{\mathrm{loc}}_{j,\alpha}(P)$: $F^{\mathrm{loc}}_{j,\alpha}(P) = -\operatorname{Tr}\left(\frac{\partial V_{\mathrm{loc}}}{\partial R_{j,\alpha}} P\right)$ so that computing $\mathrm{d}F^{\mathrm{loc}}_{j,\alpha}(P)\cdot X$ for $X \in \mathcal{T}_P\mathcal{M}_N$ reduces to the scalar product of $X$ against $\Pi_P \frac{\partial V_{\mathrm{loc}}}{\partial R_{j,\alpha}}$.

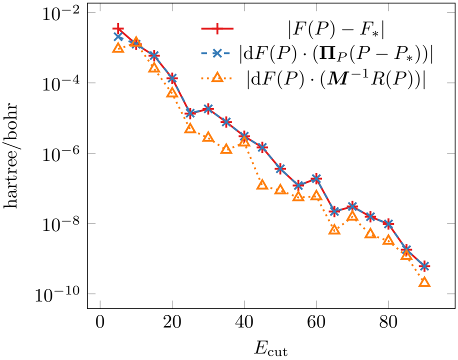
- $\boldsymbol{M}^{-1}R(P)$ is high frequencies;
- $\Pi_P(P - P_*)$ is mainly high frequencies but with low frequencies components;
- $\Pi_P \frac{\partial V_{\mathrm{loc}}}{\partial R_{j,\alpha}}$ is mainly low frequencies.



Let us analyze in details the computation of $F^{\text{loc}}_{j,\alpha}(P)$: $F^{\text{loc}}_{j,\alpha}(P) = -\operatorname{Tr}\left(\frac{\partial V_{\text{loc}}}{\partial R_{j,\alpha}} P\right)$ so that computing $\mathrm{d}F^{\text{loc}}_{j,\alpha}(P) \cdot X$ for $X \in \mathcal{T}_P\mathcal{M}_N$ reduces to the scalar product of $X$ against $\Pi_P \frac{\partial V_{\text{loc}}}{\partial R_{j,\alpha}}$.

- $\boldsymbol{M}^{-1}R(P)$ is high frequencies;
- $\Pi_P(P - P_*)$ is mainly high frequencies but with low frequencies components;
- $\Pi_P \frac{\partial V_{\text{loc}}}{\partial R_{j,\alpha}}$ is mainly low frequencies.

⇝ orange and blue do not match because the error and the residual don't have the same support in frequencies, even if $\left\|\boldsymbol{M}^{-1}R(P)\right\|_{\text{F}} \sim \|\Pi_P(P - P_*)\|_{\text{F}}$ asymptotically.



# Enhanced error bounds

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

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

# Enhanced error bounds

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

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

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

$$(\boldsymbol{\Omega} + \boldsymbol{K})_{21} \approx 0 \quad \text{and} \quad (\boldsymbol{\Omega} + \boldsymbol{K})_{22} \approx \boldsymbol{M}_{22} \approx \left(-\frac{1}{2}\Delta + 1\right)\Big|_{\mathcal{X}_{E_{\rm cut}^\perp}} \quad \text{on the tangent space ,}$$

and thus

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

# Enhanced error bounds

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

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

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

$$(\boldsymbol{\Omega}+\boldsymbol{K})_{21} \approx 0 \quad \text{and} \quad (\boldsymbol{\Omega}+\boldsymbol{K})_{22} \approx \boldsymbol{M}_{22} \approx \left(-\frac{1}{2}\Delta + 1\right)\Big|_{\mathcal{X}_{E_{\rm cut}^\perp}} \quad \text{on the tangent space} ,$$

and thus

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

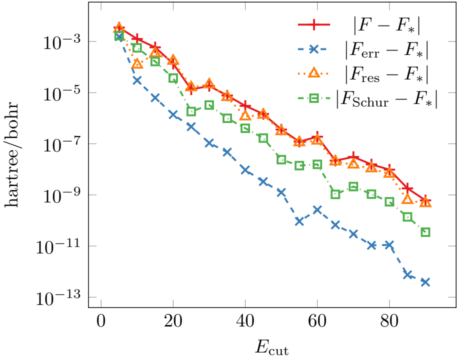
This yields a new residual, which requires only an inversion on the coarse grid $\mathcal{X}_{E_{\rm cut}}$ ($\boldsymbol{M}_{22}$ being easy to invert):

$$\boxed{R_{\rm Schur}(P) = \begin{bmatrix} (\boldsymbol{\Omega}+\boldsymbol{K})_{11}^{-1}\left(R_1 - (\boldsymbol{\Omega}+\boldsymbol{K})_{12}\, \boldsymbol{M}_{22}^{-1} R_2\right) \\ \boldsymbol{M}_{22}^{-1} R_2 \end{bmatrix}}.$$

$$F_{\text{err}} - F_* := F(P) - \mathrm{d}F(P) \cdot (\Pi_P(P - P_*)) - F(P_*),$$

$$F_{\text{res}} - F_* := F(P) - \mathrm{d}F(P) \cdot (\boldsymbol{M}^{-1}R(P)) - F(P_*),$$

$$F_{\text{Schur}} - F_* := F(P) - \mathrm{d}F(P) \cdot (R_{\text{Schur}}(P)) - F(P_*),$$

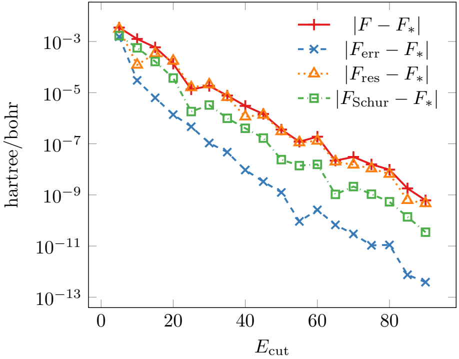


$$F_{\mathrm{err}} - F_* := F(P) - \mathrm{d}F(P) \cdot (\Pi_P(P - P_*)) - F(P_*),$$

$$F_{\mathrm{res}} - F_* := F(P) - \mathrm{d}F(P) \cdot (\boldsymbol{M}^{-1}R(P)) - F(P_*),$$

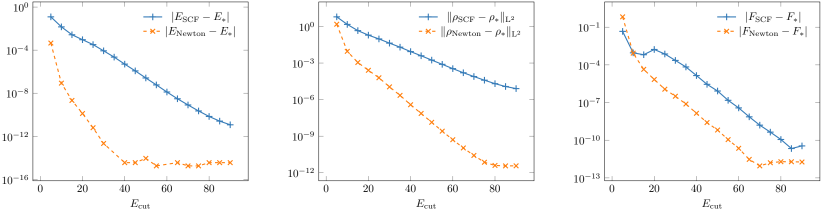
$$F_{\mathrm{Schur}} - F_* := F(P) - \mathrm{d}F(P) \cdot (R_{\mathrm{Schur}}(P)) - F(P_*),$$

⇝ we win about one order of magnitude in the approximation of the error of the forces $F - F_*$.

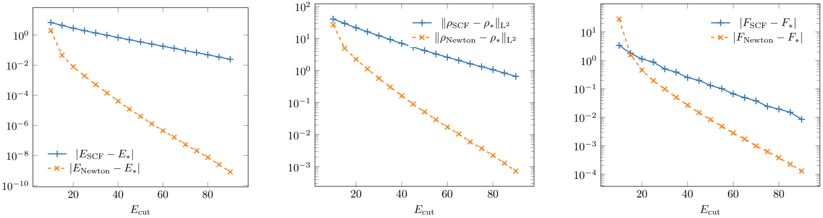


# Numerical examples

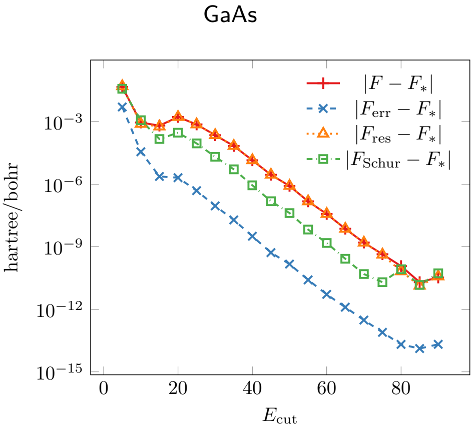
## GaAs

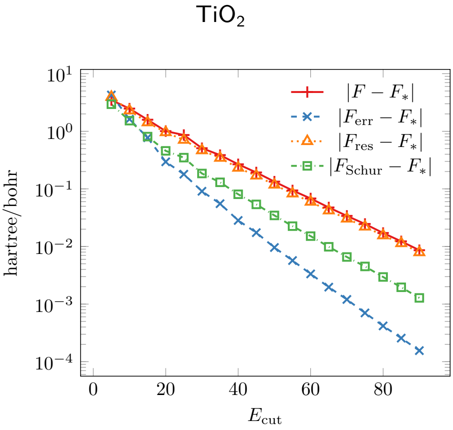


## $\mathrm{TiO_2}$



# Numerical examples





# Conclusion and take-home messages

- The asymptotic regime is quickly established;

# Conclusion and take-home messages

- The asymptotic regime is quickly established;
- error estimates based on operator norms are not good;
- in the PW setting, this come from the high frequencies nature of the residual;
- using a Schur complement to couple high and low frequencies clearly enhances the approximation of the error;

## Conclusion and take-home messages

- The asymptotic regime is quickly established;
- error estimates based on operator norms are not good;
- in the PW setting, this come from the high frequencies nature of the residual;
- 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 precision of the QoI;
- the coupling between high and low frequencies can be pushed further;

# Conclusion and take-home messages

- The asymptotic regime is quickly established;
- error estimates based on operator norms are not good;
- in the PW setting, this come from the high frequencies nature of the residual;
- 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 precision of the QoI;
- the coupling between high and low frequencies can be pushed further;
- **Limits:** we do not have guaranteed bounds, but useful in practice, valid asymptotically and for a cost comparable to a SCF cycle (inverting $\boldsymbol{\Omega} + \boldsymbol{K}$).

# Links

Preprint with more details:
`https://hal.inria.fr/hal-03408321`

Tutorial:
`https://juliamolsim.github.io/DFTK.jl/dev/examples/error_estimates_forces/`

Code:
`https://github.com/gkemlin/paper-forces-estimator`

# Resolution

$$\boxed{\begin{gathered}\min_{P\in\mathcal{M}_N} E(P) = \mathrm{Tr}\,(H_0 P) + E_{\mathrm{nl}}(P),\\[2ex] \mathcal{M}_N := \left\{P \in \mathbb{C}^{\mathcal{N}\times\mathcal{N}} \;\middle|\; P = P^*,\ \mathrm{Tr}(P) = N,\ P^2 = P\right\}.\end{gathered}}$$

direct minimization

↓

projected gradient onto the constraint manifold

Euler-Lagrange equation

↓

SCF formulation

$$\begin{cases}(H_0 + \nabla E_{\mathrm{nl}}(P))\varphi_i = \varepsilon_i\varphi_i,\\ \langle\varphi_i|\varphi_j\rangle = \delta_{ij},\\ P = \displaystyle\sum_{i=1}^{N} |\varphi_i\rangle\langle\varphi_i|.\end{cases}$$

# Tangent space

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

$$P = \begin{bmatrix} 1_N & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathcal{T}_P\mathcal{M}_N := \left\{ X = \begin{bmatrix} 0 & \times \\ \times^* & 0 \end{bmatrix} \right\}.$$

A density matrix $P \in \mathcal{M}_N$ can be described with $N$ orbitals (any orthonormal basis of $\mathrm{Ran}(P)$):

$$P = \sum_{i=1}^{N} |\varphi_i\rangle \langle\varphi_i| \quad \text{with} \quad \langle\varphi_i|\varphi_j\rangle = \delta_{ij}.$$

Given such a $P$, an element $X$ of $\mathcal{T}_P\mathcal{M}_N$ can be described with $N$ vectors that are all orthogonal to the $\varphi_i$'s:

$$X = \sum_{i=1}^{N} |\varphi_i\rangle \langle\psi_i| + |\psi_i\rangle \langle\varphi_i| \quad \text{with} \quad \langle\varphi_i|\psi_j\rangle = 0 \Rightarrow \|X\|_{\mathrm{F}}^2 = 2\sum_{i=1}^{N} \|\psi_i\|^2$$

# Tangent space

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

$$P = \begin{bmatrix} 1_N & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathcal{T}_P\mathcal{M}_N := \left\{ X = \begin{bmatrix} 0 & \times \\ \times^* & 0 \end{bmatrix} \right\}.$$

A density matrix $P \in \mathcal{M}_N$ can be described with $N$ orbitals (any orthonormal basis of $\mathrm{Ran}(P)$):

$$P = \sum_{i=1}^{N} |\varphi_i\rangle \langle\varphi_i| \quad \text{with} \quad \langle\varphi_i|\varphi_j\rangle = \delta_{ij}.$$

Given such a $P$, an element $X$ of $\mathcal{T}_P\mathcal{M}_N$ can be described with $N$ vectors that are all orthogonal to the $\varphi_i$'s:

$$X = \sum_{i=1}^{N} |\varphi_i\rangle \langle\psi_i| + |\psi_i\rangle \langle\varphi_i| \quad \text{with} \quad \langle\varphi_i|\psi_j\rangle = 0 \Rightarrow \|X\|_{\mathrm{F}}^2 = 2\sum_{i=1}^{N} \|\psi_i\|^2$$

$$\boxed{\begin{array}{ccc} P \in \mathcal{M}_N & \leftrightarrow & (\varphi_i)_{1\leqslant i\leqslant N} \in (\mathbb{C}^{\mathcal{N}})^N \text{ spanning } \mathrm{Ran}(P) \\ X \in \mathcal{T}_P\mathcal{M}_N & \leftrightarrow & (\psi_i)_{1\leqslant i\leqslant N} \in (\mathbb{C}^{\mathcal{N}})^N \text{ where } \langle\varphi_i|\psi_j\rangle = 0 \end{array}}$$

**Change of norm :** given $X \in \mathcal{T}_P\mathcal{M}_N$, one might want to compute $\|\boldsymbol{M}X\|_{\mathrm{F}}$ for a metric $\boldsymbol{M}$ on the tangent space. This can be translated in terms of orbitals as

$$\boldsymbol{M}X = \sum_{i=1}^{N} |\varphi_i\rangle \langle M_i\psi_i| + |M_i\psi_i\rangle \langle \varphi_i| \,, \quad \|\boldsymbol{M}X\|_{\mathrm{F}} = 2\sum_{i=1}^{N} \|M_i\psi_i\|$$

where $M_i : \mathrm{Ran}(\{\varphi_j\})^{\perp} \to \mathrm{Ran}(\{\varphi_j\})^{\perp}$ and can eventually depend on the band $i$. In this talk we will use (with $\Pi$ the projection on $\mathrm{Ran}(\{\varphi_j\})^{\perp}$ and $t_i$ the kinetic energy of band $i$):

$$\begin{array}{rcll}
\boldsymbol{M}^{1/2} & \leftrightarrow & \Pi(t_i - \Delta/2)^{1/2}\Pi & \leftrightarrow \quad \mathrm{H}^{1/2} \text{ norm} \\
\boldsymbol{M} & \leftrightarrow & \Pi(t_i - \Delta/2)^{1/2}\Pi(t_i - \Delta/2)^{1/2}\Pi & \leftrightarrow \quad \mathrm{H}^{1} \text{ norm} \\
& & & \\
\boldsymbol{M}^{-1/2} & \leftrightarrow & (\Pi(t_i - \Delta/2)^{1/2}\Pi)^{-1} & \leftrightarrow \quad \mathrm{H}^{-1/2} \text{ norm} \\
\boldsymbol{M}^{-1} & \leftrightarrow & (\Pi(t_i - \Delta/2)^{1/2}\Pi(t_i - \Delta/2)^{1/2}\Pi)^{-1} & \leftrightarrow \quad \mathrm{H}^{-1} \text{ norm}
\end{array}$$

# Extension and computation of super operators

**Computing $\boldsymbol{K}$ :** $\boldsymbol{K}(P) := \Pi_P \nabla^2 E(P) \Pi_P$ can be defined at any $P = \sum_{i=1}^N |\varphi_i\rangle \langle\varphi_i| \in \mathcal{M}_N$. In terms of orbitals, this translates into

$$\forall\ X \in \mathcal{T}_P\mathcal{M}_N, \quad \boldsymbol{K}(P)X = \sum_{i=1}^N |\varphi_i\rangle \langle \delta V \varphi_i| + |\delta V \varphi_i\rangle \langle\varphi_i|\,,$$

where $X$ is described by $(\psi_i)_{1\leqslant i\leqslant N} \in (\mathrm{Ran}(\{\varphi_j\})^\perp)^N$ and

$$(\psi_i)_{1\leqslant i\leqslant N} \mapsto \delta\rho := 2\sum_{i=1}^N \varphi_i\psi_i \mapsto \delta V \mapsto (\delta V \varphi_i)_{1\leqslant i\leqslant N}.$$

**Computing $\boldsymbol{\Omega}$ :** for $P = \sum_{i=1}^N |\varphi_i\rangle\langle\varphi_i| \in \mathcal{M}_N$, we define $\boldsymbol{\Omega}(P) : \mathcal{T}_P\mathcal{M}_N \to \mathcal{T}_P\mathcal{M}_N$ by

$$\forall\ X \in \mathcal{T}_P\mathcal{M}_N, \quad \boldsymbol{\Omega}(P)X = -[P,[H(P),X]],$$

where $H(P) \coloneqq \nabla E(P)$. In terms of orbitals it translates into

$$\boldsymbol{\Omega}(P)X = \sum_{i=1}^N |\varphi_i\rangle \left\langle (1-P)\left(H(P)\psi_i - \sum_{j=1}^N \Lambda_{ij}\psi_j\right)\right| + \mathrm{hc},$$

where $X$ is described by $(\psi_i)_{1\leqslant i\leqslant N} \in (\mathrm{Ran}(\{\varphi_j\})^\perp)^N$ and $\Lambda_{ij} \coloneqq \varphi_j^* H(P)\varphi_i$ (diagonal if $P = P_*$).

| Analysis | | What is used in practice |
|---|---|---|
| $P \in \mathcal{M}_N$ | $\leftrightarrow$ | $\Phi = (\varphi_i)_{1\leqslant i\leqslant N} \in (\mathbb{C}^{\mathcal{N}})^N$ spanning $\mathrm{Ran}(P)$ |
| $X \in \mathcal{T}_P\mathcal{M}_N$ | $\leftrightarrow$ | $\Psi = (\psi_i)_{1\leqslant i\leqslant N} \in (\mathbb{C}^{\mathcal{N}})^N$ s.t. $\langle\varphi_i\|\psi_j\rangle = 0$ |
| $\\|X\\|_{\mathrm{F}}^2$ | $\leftrightarrow$ | $2\sum_{i=1}^{N} \\|\psi_i\\|^2$ |
| $\\|\boldsymbol{M}^s X\\|_{\mathrm{F}}^2$ | $\leftrightarrow$ | $2\sum_{i=1}^{N} \\|M_i^s\psi_i\\|^2$ for $s = -1, -1/2, 1/2, 1$ |
| $\boldsymbol{K}(P)X$ | $\leftrightarrow$ | $K(\Phi)\Psi = (\delta V\varphi_i)_{1\leqslant i\leqslant N}$ |
| $\boldsymbol{\Omega}(P)X$ | $\leftrightarrow$ | $\Omega(\Phi)\Psi = \left((1-P)\left(H(P)\psi_i - \sum_{j=1}^N \Lambda_{ij}\psi_j\right)\right)_{1\leqslant i\leqslant N}$ |

# Mathematical justification for 1D Gross-Pitaevskii

$$\begin{cases} -\Delta\phi_* + V\phi_* + \phi_*^3 = \lambda_*\phi_*, \\ \|\phi_*\|_{\mathrm{L}^2_\#} = 1, \quad \phi_* > 0 \text{ on } \mathbb{R}^d, \end{cases} \qquad \begin{cases} -\Delta\phi_N + \Pi_N\left(V\phi_N - \phi_N^3\right) = \lambda_N\phi_N, \\ \|\phi_N\|_{\mathrm{L}^2_\#} = 1. \end{cases}$$

- $\Pi_{\phi_N}^\perp$ is the orthogonal projector (for the $\mathrm{L}^2_\#$ inner product) onto $\phi_N^\perp$;
- $A_N$ is the self-adjoint operator on $\phi_N^\perp$ defined by $A_N := (\Omega_N + K_N)$ where $\Omega_N$ and $K_N$ represent, in the orbital framework, the super-operators $\boldsymbol{\Omega}(P_N)|_{T_{P_N}\mathcal{M}_\infty}$ and $\boldsymbol{K}(P_N)|_{T_{P_N}\mathcal{M}_\infty}$. We have

$$(1) \qquad \forall\ \psi_N \in \phi_N^\perp, \quad \Omega_N\psi_N = \Pi_{\phi_N}^\perp\left(-\Delta + V + \phi_N^2 - \lambda_N\right)\psi_N,$$

$$(2) \qquad \forall\ \psi_N \in \phi_N^\perp, \quad K_N\psi_N = \Pi_{\phi_N}^\perp\left(2\phi_N^2\psi_N\right);$$

- $M_N^{1/2}$ is the restriction of the operator $\Pi_{\phi_N}^\perp(1-\Delta)^{1/2}\Pi_{\phi_N}^\perp$ to the invariant subspace $\phi_N^\perp$.

## Proposition

*We have*

$$\lim_{N\to\infty}\left\|M_N^{1/2}(\Omega_N + K_N)^{-1}M_N^{1/2} - I_{\mathcal{X}_N^\perp}\right\|_{\mathcal{X}_N^\perp \to L^2_\#} = 0.$$

# Guaranteeing bounds

Solve $R(x) = 0$ with $R : Y \to Z$.

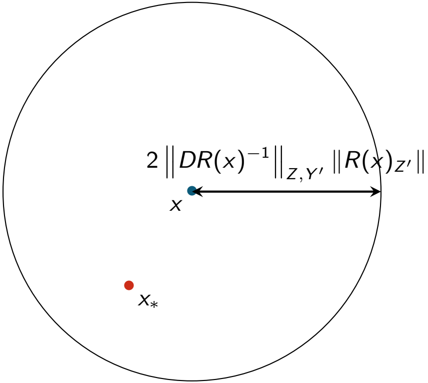
**Theorem (Inverse function theorem – Newton - Kantorovich[a])**

*Assume that*

- $DR(x) \in \mathcal{L}(Y, Z)$ *is an isomorphism*
- $2\left\|DR(x)^{-1}\right\|_{Z,Y'} L\left(2\left\|DR(x)^{-1}\right\|_{Z,Y'}\|R(x)_{Z'}\|\right) \le 1$ *with* $L(\alpha) = \sup_{y\in\bar{B}(x,\alpha)}\|DR(x) - DR(y)\|_{Z,Y'}$.

*Then, the problem* $R(x) = 0$ *has a unique solution* $x_*$ *in the ball* $\bar{B}(x, 2\left\|DR(x)^{-1}\right\|_{Z,Y'}\|R(x)_{Z'}\|)$. *Moreover,*

$$\boxed{\|x - x_*\|_Y \le 2\left\|DR(x)^{-1}\right\|_{Z,Y'}\|R(x)_{Z'}\|\,.}$$



[a] Gabriel Caloz, Jacques Rappaz. Numerical analysis for nonlinear and bifurcation problems. Handbook Numerical Analysis, 5:487-637 (1997).