Error estimators for Density Functional Theory

Eric CANCES

Ecole des Ponts et Inria Paris, France

IPAM, Los Angeles, March 27-31, 2023
Motivation

Numerical challenges in DFT

- very large systems with high accuracy
- metallic / heterogeneous systems
- databases for materials and drug discovery and design (machine learning)
  Ex.: open catalyst data set (Chanussot et al. ’20)
    264,890,000 Density Functional Theory (DFT) calculations

Need for fast, reliable, black-box numerical methods with error estimators
**Motivation**

Input $y \rightarrow$ Reference model $\xrightarrow{\text{Modeling error } e_m}$ Approximate model $\xrightarrow{\text{Discretization error } e_d}$ Discretized model $\xrightarrow{\text{Algorithmic error } e_a}$ Numerical scheme $\xrightarrow{\text{Implementation error } e_i}$ Computer code

Output: quantity of interest $s=f(y)$ $\xrightarrow{\text{Error } e}$ Output: computed quantity of interest $s'=F(y,z)$

**Total error:** $e = e_m + e_d + e_a + e_i + e_c$

**Example:** $s = \text{static polarizability of the CO}_2 \text{ molecule}$
$y = (\#\text{C atoms} = 1, \#\text{O atoms} = 2, \text{total charge} = 0)$
(+ **universal parameters:** $\hbar, m_e, e, \varepsilon_0, z_X, m_{X_{\text{iso}}}$)
**Motivation**

Input $y$ → Reference model → Approximate model

- Physical arguments, asymptotic analysis
- Modeling error $e_m$

Discretized model

- Discretization methods and parameters
- Discretization error $e_d$

Numerical scheme

- Algorithms with stopping criteria
- Algorithmic error $e_a$

Computer code

- Programming language and data structures
- Implementation error $e_i$

Hardware and compiler

- Calculation error $e_c$

Output: quantity of interest $s = f(y)$ → Error $e$ → Output: computed quantity of interest $s' = F(y,z)$

**Total error:** $e = e_m + e_d + e_a + e_i + e_c$

**Goal 1:** provide error estimates (certification)
Motivation

Goal 1: provide error estimates (certification)
Goal 2: optimize dynamically the computational resources (error balancing)
(DFT ~ 15% of CPU time available in HPC centers)
Motivation

An example of fully guaranteed error bars
Herbst, Levitt, EC, *Faraday discussions* ’20

Si band diagram
(non-self-consistent DFT-LDA model with GTH pseudopotentials)

fully guaranteed error bars accounting for
• discretization error
• algorithmic error
• floating-point arithmetics (=implementation error assuming bug-free code)
Motivation

Numerical methods implemented in DFTK (Density-Functional ToolKit)

• planewave DFT package in Julia (2018-), MIT license
• outcome of the EMC2 ERC Synergy project
• main developers: Michael Herbst (now at EPFL) and Antoine Levitt (now at Paris-Saclay)

• supports mathematical developments (low entrance barrier, \(\sim 7k\) lines of code) and scale-up to relevant applications (\(\sim 1,000\) electrons)
• fully composable with Julia ecosystem
  – arbitrary precision (32bit, 64bit...)
  – algorithmic differentiation
  – interval arithmetic
  – ...
1. Geometrical structure of the Kohn-Sham model

2. Practical error bounds on interatomic forces in Kohn-Sham DFT

3. Computation of response properties for metals
1 - Geometrical structure of the Kohn-Sham model

Periodic supercell $\Omega = \mathbb{R}^3/L$, $L$: Bravais lattice of $\mathbb{R}^3$

$M$ ionic cores, positions $X := (X_j)_{1 \leq j \leq M} \in \Omega^M$

$N$ valence electron pairs (spin-unpolarized state)
Orbital formulation of the continuous Kohn-Sham model

\[
\min \left\{ E_{\text{KS}}(X, \Phi) \mid \Phi = (\phi_1, \cdots, \phi_N) \in \left( H^1_\#(\Omega; \mathbb{C}) \right)^N, \int_{\Omega} \phi_i^* \phi_j = \delta_{ij} \right\}
\]

\[
E_{\text{KS}}(X, \Phi) = \sum_{i=1}^{N} \int_{\Omega} |\nabla \phi_i|^2 + \int_{\Omega} v(X) \rho_{\Phi} + E_{\text{Hxc}}(\rho_{\Phi}) \quad \text{with} \quad \rho_{\Phi}(r) := 2 \sum_{i=1}^{N} |\phi_i(r)|^2
\]
Orbital formulation of the continuous Kohn-Sham model

\[
\min \left\{ E^{KS}(X, \Phi) \mid \Phi = (\phi_1, \cdots, \phi_N) \in (H^1_\#(\Omega; \mathbb{C}))^N, \int_{\Omega} \phi_i^* \phi_j = \delta_{ij} \right\}
\]

\[
E^{KS}(X, \Phi) = \sum_{i=1}^N \int_{\Omega} |\nabla \phi_i|^2 + \int_{\Omega} v(X) \rho_{\Phi} + E_{\text{Hxc}}(\rho_{\Phi}) \quad \text{with} \quad \rho_{\Phi}(r) := 2 \sum_{i=1}^N |\phi_i(r)|^2
\]

Discretization in a planewave basis set

Fourier modes: \( e_G(r) := |\Omega|^{-1/2} e^{iG \cdot r} \), \( G \in \mathbb{L}^* \) (reciprocal lattice)

Approximation spaces: \( \mathcal{X} = \text{Span} \left( e_G, \ G \in \mathbb{L}^*, \frac{|G|^2}{2} < E_c \right) \)

\[ \uparrow \]
energy cut-off
Orbital formulation of the continuous Kohn-Sham model

\[
\min \left\{ E^{KS}(X, \Phi) \mid \Phi = (\phi_1, \cdots, \phi_N) \in \left( H^1_\#(\Omega; \mathbb{C}) \right)^N, \int_\Omega \phi_i^* \phi_j = \delta_{ij} \right\}
\]

\[
E^{KS}(X, \Phi) = \sum_{i=1}^{N} \int_\Omega |\nabla \phi_i|^2 + \int_\Omega v(X) \rho_\Phi + E_{\text{Hxc}}(\rho_\Phi) \quad \text{with} \quad \rho_\Phi(r) := 2 \sum_{i=1}^{N} |\phi_i(r)|^2
\]

Orbital formulation of the discretized KS model (used in most simulations)

\[
\phi_i(r) = \sum_{G \in \mathbb{L}^* \mid |G| < \sqrt{2E_c}} C_{Gi} e_G(r) \in \mathcal{X}, \quad N_c := \dim(\mathcal{X})
\]

minimization set: \( \text{St}(N, N_c) := \{ C = [C_{Gi}] \in \mathbb{C}^{N_c \times N} \mid C^* C = I_N \} \)

drawback: gauge invariance (\( C \) and \( CU \) have same energy \( \forall U \in U(N_c) \))
Orbital formulation of the continuous Kohn-Sham model

\[
\min \left\{ E_{\text{KS}}(X, \Phi) \mid \Phi = (\phi_1, \ldots, \phi_N) \in (H^1_{\#}(\Omega; \mathbb{C}))^N, \right. \\
\left. \int_{\Omega} \phi_i^* \phi_j = \delta_{ij} \right\}
\]

\[
E_{\text{KS}}(X, \Phi) = \sum_{i=1}^{N} \int_{\Omega} |\nabla \phi_i|^2 + \int_{\Omega} v(X) \rho_\Phi + E_{\text{Hxc}}(\rho_\Phi) \quad \text{with} \quad \rho_\Phi(r) := 2 \sum_{i=1}^{N} |\phi_i(r)|^2
\]

Orbital formulation of the discretized KS model (used in most simulations)

\[
\phi_i(r) = \sum_{\mathbf{G} \in \mathbb{L}^* \mid |\mathbf{G}| < \sqrt{2E_c}} C_{\mathbf{G}i} e_\mathbf{G}(r) \in \mathcal{X}, \quad N_c := \text{dim}(\mathcal{X})
\]

minimization set: \( \text{St}(N, N_c) := \{ C = [C_{\mathbf{G}i}] \in \mathbb{C}^{N_c \times N} \mid C^*C = I_N \} \)

drawback: gauge invariance (\( C \) and \( CU \) have same energy \( \forall U \in U(N_c) \))

Density matrix (1-RDM) formulation (math. analysis, linear scaling methods)

\( P = CC^* \) orthogonal proj. on span(columns of \( C \))

minimization set: \( \text{Gr}(N, N_c) := \{ P \in \mathbb{C}^{N_c \times N_c}_{\text{herm}} \mid P^2 = P, \text{Tr}(P) = N \} \)
Geometrical properties of the Grassmann manifold

The Frobenius inner product on $\mathbb{C}^{N_c \times N_c}_{\text{herm}}$ induces a Riemannian metric on

$$\mathcal{M} := \text{Gr}(N, N_c) = \{ P \in \mathbb{C}^{N_c \times N_c}_{\text{herm}} \mid P^2 = P, \ Tr(P) = N \}$$
Geometrical properties of the Grassmann manifold

The Frobenius inner product on $\mathbb{C}^{N_c \times N_c}_{\text{herm}}$ induces a Riemannian metric on

$$\mathcal{M} := \text{Gr}(N, N_c) = \{ P \in \mathbb{C}_{\text{herm}}^{N_c \times N_c} \mid P^2 = P, \text{Tr}(P) = N \}$$

**Remark:** we consider a hierarchy of 2 approximation spaces ($E_{\text{calc}} \ll E_{\text{ref}}$)

$$\mathcal{X}_{\text{calc}} \subset \mathcal{X}_{\text{ref}}$$

- variational approximation space
- error estimation space

- to study the convergence of numerical algorithms, we take $E_c = E_{\text{calc}}$


- to derive practical error bounds, we take $E_c = E_{\text{ref}}$
Geometrical properties of the Grassmann manifold

The Frobenius inner product on $\mathbb{C}^{N_c \times N_c}_{\text{herm}}$ induces a Riemannian metric on

$$\mathcal{M} := \text{Gr}(N, N_c) = \{ P \in \mathbb{C}^{N_c \times N_c}_{\text{herm}} \mid P^2 = P, \text{Tr}(P) = N \}$$

Tangent space

$$\forall P \in \mathcal{M}, \quad T_P \mathcal{M} = \{ Q \in \mathbb{C}^{N_c \times N_c}_{\text{herm}} \mid PQP = (I - P)Q(I - P) = 0 \}$$

The orthogonal projector $\Pi_P \in \mathcal{L}(\mathbb{C}^{N_c \times N_c}_{\text{herm}})$ on $T_P \mathcal{M}$ if given by

$$\forall M \in \mathbb{C}^{N_c \times N_c}_{\text{herm}}, \quad \Pi_P M = [P, [P, M]]$$
Geometrical properties of the Grassmann manifold

The Frobenius inner product on $\mathbb{C}^{Nc \times Nc}_{\text{herm}}$ induces a Riemannian metric on

$$\mathcal{M} := \text{Gr}(N, Nc) = \{ P \in \mathbb{C}^{Nc \times Nc}_{\text{herm}} \mid P^2 = P, \ Tr(P) = N \}$$

Tangent space

$$\forall P \in \mathcal{M}, \ T_P\mathcal{M} = \{ Q \in \mathbb{C}^{Nc \times Nc}_{\text{herm}} \mid PQP = (I - P)Q(I - P) = 0 \}$$

The orthogonal projector $\Pi_P \in \mathcal{L}(\mathbb{C}^{Nc \times Nc}_{\text{herm}})$ on $T_P\mathcal{M}$ if given by

$$\forall M \in \mathbb{C}^{Nc \times Nc}_{\text{herm}}, \quad \Pi_P M = [P, [P, M]]$$

Riemannian gradient and Hessian

Let $E \in C^1(\mathbb{C}^{Nc \times Nc}_{\text{herm}}; \mathbb{R})$ and $E : \mathcal{M} \ni P \mapsto E(P) := \mathcal{E}(P) \in \mathbb{R}$

- Gradient of $E$ at $P \in \mathcal{M}$:
  $$\nabla_{\mathcal{M}}E(P) = \Pi_P(\nabla \mathcal{E}(P)) \in T_P\mathcal{M}$$
Geometrical properties of the Grassmann manifold

The Frobenius inner product on $\mathbb{C}^{N_c \times N_c}_{\text{herm}}$ induces a Riemannian metric on
$$\mathcal{M} := \text{Gr}(N, N_c) = \{ P \in \mathbb{C}^{N_c \times N_c}_{\text{herm}} \mid P^2 = P, \ \text{Tr}(P) = N \}$$

Tangent space
$$\forall P \in \mathcal{M}, \quad T_P \mathcal{M} = \{ Q \in \mathbb{C}^{N_c \times N_c}_{\text{herm}} \mid PQP = (I - P)Q(I - P) = 0 \}$$

The orthogonal projector $\Pi_P \in \mathcal{L}(\mathbb{C}^{N_c \times N_c}_{\text{herm}})$ on $T_P \mathcal{M}$ if given by
$$\forall M \in \mathbb{C}^{N_c \times N_c}_{\text{herm}}, \quad \Pi_P M = [P, [P, M]]$$

Riemannian gradient and Hessian

Let $\mathcal{E} \in C^1(\mathbb{C}^{N_c \times N_c}_{\text{herm}}; \mathbb{R})$ and $E : \mathcal{M} \ni P \mapsto E(P) := \mathcal{E}(P) \in \mathbb{R}$

- **Gradient of $E$ at $P \in \mathcal{M}$:** $\nabla_{\mathcal{M}} E(P) = \Pi_P (\nabla \mathcal{E}(P)) \in T_P \mathcal{M}$
- **Hessian of $E$ at $P \in \mathcal{M}$:** $D^2_{\mathcal{M}} E(P) = \Omega_P + K_P \in \mathcal{L}(T_P \mathcal{M})$

with $\Omega_P, K_P : T_P \mathcal{M} \to T_P \mathcal{M}$ given by
$$\Omega_P = -[P, [\nabla \mathcal{E}(P), \bullet]] \quad \text{and} \quad K_P = \Pi_P D^2 \mathcal{E}(P) \Pi_P$$
2 - Practical error bounds on interatomic forces

Error bounds on energy and density

- *a priori* error bounds: EC-Chakir-Maday ’12, Zhou et al. ’13, EC-Dusson ’17...
- *a posteriori* error bounds: EC-Dusson-Maday-Stamm-Vohralík ’14-’21, Chen-Dai-Gong-He-Zhou ’14, Chen-Schneider ’15, Kaye-Lin-Yang ’15, Herbst-Levitt-EC ’20 ...

This work: error bounds on properties (non-variational)
Interatomic forces (at the discrete level)

- Manifold of admissible 1-RDM
  \[ \mathcal{M} = \text{Gr}(N, N_{\text{ref}}) = \{ P \in \mathbb{C}^{N_{\text{ref}} \times N_{\text{ref}}} \mid P^2 = P, \ Tr(P) = N \} \]

- Kohn-Sham energy functional \( E^{\text{KS}} : \Omega^M \times \mathcal{M} \to \mathbb{R} \)
  \[ E^{\text{KS}}(X, P) := \text{Tr}(TP) + \text{Tr}(V(X)P) + E_{\text{Hxc}}(P) \]

- Kohn-Sham ground state and interatomic forces (Hellmann-Feynman)
  \[ P_*(X) = \arg\min_{P \in \mathcal{M}} E(X, P) \quad F_j(X) := -\text{Tr} \left( \nabla X_j V(X) P_*(X) \right) + F_{\text{ion}}(X) \]
Force component $f_* = [F_{j\alpha}(X_0)]_\alpha$ for a fixed nuclear configuration $X_0$

$$E(P) := E^{KS}(X_0, P) \quad f(P) := -\text{Tr} \left( \partial_{X_{j\alpha}} V(X_0) P \right)$$

$$P_* := \text{argmin}_{P \in \mathcal{M}} E(P), \quad f_* = f(P_*)$$

First-order optimality condition

$$\nabla_{\mathcal{M}} E(P_*) = 0$$

Residual associated with an approximate solution $P$

$$R(P) := \nabla_{\mathcal{M}} E(P) = [P, [P, H(P)]]$$

where $H(P) := T + V(X_0) + \nabla E_{Hxc}(P)$ is the Kohn-Sham Hamiltonian, \((\nabla E_{Hxc}(P) = \text{Frobenius gradient of } C_{\text{herm}}^{N_{\text{ref}} \times N_{\text{ref}}} \ni P \mapsto E_{Hxc}(P) \in \mathbb{R})\)
General principle:
- $\mathcal{M}$ smooth Riemannian manifold (set of admissible states)
- $R : \mathcal{M} \ni P \mapsto R(P) \in T_P \mathcal{M}$ smooth function (residual)
- $f : \mathcal{M} \to \mathbb{R}$ smooth function (observable)

Equilibrium state: $R(P_*) = 0$
Quantity of interest (QoI): $f_* := f(P_*)$

Error estimators: in the linear regime

$P - P_* \approx R'(P)^{-1}R(P)$ and $f(P) - f_* \approx \langle \nabla \mathcal{M} f(P), R'(P)^{-1}R(P) \rangle$

Two practical questions: if $P \approx \text{DM}$ of the KS ground-state in $\mathcal{X}_{\text{calc}} \subset \mathcal{X}_{\text{ref}}$

1. are we in the linear regime in usual DFT calculations on real materials?
2. if so, how to compute an accurate and computationally efficient bound for $\langle \nabla \mathcal{M} f(P), R'(P)^{-1}R(P) \rangle$?
Are we in the linear regime in practice?

Apparently yes! ... for planewave calculations with nice pseudopotentials

**Numerical test:** if \( P - P_* \approx R'(P)^{-1}R(P) \), then the QoI computed from

\[
P_{\text{Newton}} = \mathcal{R} \left( P - R'(P)^{-1}R(P) \right) \quad (\mathcal{R} \text{ suitable retraction})
\]

should be much better than the ones computed from \( P \)

**Example:** FCC silicon crystal (energies in Hartrees)
Accurate and efficient approximation of $P - P_*$

Let $P$ be the variational approximation of the KS ground state in $X_{\text{cal}}$

\[ P - P_* \approx R'(P)^{-1} R(P) \approx M_P^{-1} R(P) \]  

in energy norm

$M_P^{-1}$ very simple operator deduced from mathematical analysis

Equivalence between these three quantities proved in the asymptotic regime
Accurate and efficient approximation of $P - P^*$

Let $P$ be the variational approximation of the KS ground state in $\mathcal{X}_{\text{cal}}$

\[
P - P^* \approx R'(P)^{-1} R(P) \approx M_P^{-1} R(P) \quad \text{in energy norm}
\]

\[
P = \sum_{i=1}^{N} \phi_i \phi_i^*, \quad \phi_i \in \mathcal{X}_{\text{cal}}, \quad \Pi_{\mathcal{X}_{\text{cal}}} H(P) \Pi_{\mathcal{X}_{\text{cal}}} \phi_i = \varepsilon_i \phi_i, \quad \phi_i \phi_j^* = \delta_{ij}, \quad \varepsilon_1 \leq \varepsilon_2 \leq \cdots
\]

\[
P = \sum_{i=1}^{N} \phi_i \phi_i^* \in \mathcal{M} \iff \Phi = (\phi_1 | \cdots | \phi_N) \in \text{St}(N, N_{\text{ref}}) \quad \text{s.t.} \quad P = \Phi \Phi^*
\]

\[
Q = \sum_{i=1}^{N} \phi_i \xi_i^* + \xi_i \phi_i^* \in \mathcal{T}_P \mathcal{M} \iff \Xi = (\xi_1 | \cdots | \xi_N) \in \mathbb{C}^{N_{\text{ref}} \times N} \quad \text{s.t.} \quad \Phi^* \Xi = 0
\]

\[
R(P) = [P, [P, H(P)]] \in \mathcal{T}_P \mathcal{M} \iff R^{\text{MO}}(\Phi) = (r_1 | \cdots | r_N) = P^\perp H(P) \Phi \in \mathbb{C}^{N_{\text{ref}} \times N}
\]

\[
M_P Q \in \mathcal{T}_P \mathcal{M} \iff M^{\text{MO}}_\Phi \Xi := ((-\Delta - t_i) \xi_i)_{1 \leq i \leq N} \quad t_i := \int_{\Omega} |\nabla \phi_i|^2
\]
First (unsuccessful) attempt to compute error bounds on interatomic forces

\[ f(P) - f_* \approx \langle \nabla_{\mathcal{M}} f(P), P - P_* \rangle \]  

(we are in the linear regime)

\[ \approx \langle \nabla_{\mathcal{M}} f(P), M_P^{-1} R(P) \rangle \]  

\[ (P - P_* \approx M_P^{-1} R(P) \text{ in energy norm}) \]
First (unsuccessful) attempt to compute error bounds on interatomic forces

\[ f(P) - f_* \approx \langle \nabla_M f(P), P - P_* \rangle \]  
  (we are in the linear regime)

\[ \approx \langle \nabla_M f(P), M_P^{-1} R(P) \rangle \]  
  wrong by several orders of magnitude
First (unsuccessful) attempt to compute error bounds on interatomic forces

\[ f(P) - f_* \approx \langle \nabla_M f(P), P - P_* \rangle \]

(we are in the linear regime)

\[ \approx \langle \nabla_M f(P), M_P^{-1} R(P) \rangle \]

wrong by several orders of magnitude

because \( \nabla_M f(P) \) and \( P - P_* \) are almost orthogonal for nonnegative Sobolev norms

**Example:** FCC silicon crystal

Fourier coefficients of \( \nabla_M f(P) \) (left), \( P - P_* \) (center) \( M_P^{-1} R(P) \) (right)
Second (successful) attempt to compute error bounds on interatomic forces

\[
f(P) - f_* \approx \langle \nabla_M f(P), P - P_* \rangle \quad \text{(we are in the linear regime)}
\]

\[
\approx \langle \nabla_M f(P), R'(P)^{-1} R(P) \rangle \quad \text{yes, but too computationally expensive}
\]
Second (successful) attempt to compute error bounds on interatomic forces

\[ f(P) - f_* \approx \langle \nabla_M f(P), P - P_* \rangle \]  

(we are in the linear regime)

\[ \approx \langle \nabla_M f(P), R'(P)^{-1} R(P) \rangle \]  

yes, but too computationally expensive

A Schur complement approach to solve

\[ R'(P) Q = R(P) \]

\[ P = \sum_{i=1}^{N} \phi_i \phi_i^* \in M, \quad Q = \sum_{i=1}^{N} (\phi_i \xi_i^* + \xi_i \phi_i^*) \in T_P M, \]

\[ \mathcal{X}_{\text{ref}} = \mathcal{X}_{\text{calc}} \oplus \mathcal{X}_{\text{calc}}^\perp, \quad \phi_i \in \mathcal{X}_{\text{calc}}, \quad \xi_i = \xi_{i,1} + \xi_{i,2} \in \mathcal{X}_{\text{calc}}^\perp, \quad Q = Q_1 + Q_2 \]

\[
\begin{pmatrix}
R'(P)_{11} & R'(P)_{12} \\
R'(P)_{21} & R'(P)_{22}
\end{pmatrix}
\begin{pmatrix}
Q_1 \\
Q_2
\end{pmatrix}
= 
\begin{pmatrix}
R(P)_{11} \\
R(P)_{22}
\end{pmatrix}
\]
Second (successful) attempt to compute error bounds on interatomic forces

\[ f(P) - f_* \approx \langle \nabla \mathcal{M} f(P), P - P_* \rangle \quad \text{(we are in the linear regime)} \]

\[ \approx \langle \nabla \mathcal{M} f(P), R'(P)^{-1} R(P) \rangle \quad \text{yes, but too computationally expensive} \]

A Schur complement approach to solve \( R'(P) Q = R(P) \)

\[ P = \sum_{i=1}^{N} \phi_i \phi_i^* \in \mathcal{M}, \quad Q = \sum_{i=1}^{N} (\phi_i \xi_i^* + \xi_i \phi_i^*) \in T_P \mathcal{M}, \]

\[ \mathcal{X}_{\text{ref}} = \mathcal{X}_{\text{calc}} \oplus \mathcal{X}_{\text{calc}}^\perp, \quad \phi_i \in \mathcal{X}_{\text{calc}}, \quad \xi_i = \xi_{i,1} + \xi_{i,2}, \quad Q = Q_1 + Q_2 \]

\[ \begin{pmatrix} R'(P)_{11} & R'(P)_{12} \\ R'(P)_{21} & R'(P)_{22} \end{pmatrix} \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} = \begin{pmatrix} R(P)_1 \\ R(P)_2 \end{pmatrix} \]

\( Q_1 \) small (in energy norm) but large impact on the error on interactomic forces
\( Q_2 \) large (in energy norm) but smaller impact on the error on interactomic forces

\[ Q_2 \approx M_P^{-1} R(P)_2 \] (in energy norm), \quad \text{\( R(P)_1 = 0 \) if \( P \) variational solution in} \ \mathcal{X}_{\text{calc}}
Second (successful) attempt to compute error bounds on interatomic forces

\[ f(P) - f_* \approx \langle \nabla_M f(P), P - P_* \rangle \]

(we are in the linear regime)

\[ \approx \langle \nabla_M f(P), R'(P)^{-1} R(P) \rangle \]

yes, but too computationally expensive

A Schur complement approach to solve

\[ R'(P)Q = R(P) \]

\[
P = \sum_{i=1}^{N} \phi_i \phi_i^* \in \mathcal{M}, \quad Q = \sum_{i=1}^{N} (\phi_i \xi_i^* + \xi_i \phi_i^*) \in T_P \mathcal{M},
\]

\[
\chi_{\text{ref}} = \chi_{\text{calc}} \oplus \chi_{\text{calc}}^\perp, \quad \phi_i \in \chi_{\text{calc}}, \quad \xi_i = \xi_{i,1} + \xi_{i,2}, \quad Q = Q_1 + Q_2
\]

\[
\begin{pmatrix}
R'(P)_{11} & R'(P)_{12} \\
0 & M_P
\end{pmatrix}
\begin{pmatrix}
Q_{1}^{\text{app}} \\
Q_{2}^{\text{app}}
\end{pmatrix}
= 
\begin{pmatrix}
R(P)_1 \\
R(P)_2
\end{pmatrix}
\]

\[
Q_{2}^{\text{app}} = M_P^{-1} R(P)_2, \quad (\Omega_P + K_P)_{11} Q_{1}^{\text{app}} = R(P)_1 - (\Omega_P + K_P)_{12} M_P^{-1} R(P)_2
\]
Numerical results: post-processing

\[ f - f_* := f(P) - f(P_*) \]  \hspace{1cm} \text{(red)}

\[ f_{\text{res}} - f_* := f(P) - \langle \nabla_M f(P), M^{-1} R(P) \rangle - f(P_*) \]  \hspace{1cm} \text{(orange)}

\[ f_{\text{Schur}} - f_* := f(P) - \langle \nabla_M f(P), R'_{\text{app}}(P)^{-1} R(P) \rangle - f(P_*) \]  \hspace{1cm} \text{(green)}

\[ f_{\text{err}} - f_* := f(P) - \langle \nabla_M f(P), P - P_* \rangle - f(P_*) \]  \hspace{1cm} \text{(blue)}
3 - Response properties for metals

Linear (noninteracting) Schrödinger-like Hamiltonian $H$

Perturbation $\delta V$ of the potential

$$H \rightarrow H + \delta V \quad \Rightarrow \quad \rho \rightarrow \rho + \delta \rho + o(\delta V) \quad \text{with } \delta \rho \text{ linear in } \delta V$$

$$\delta \rho = \chi_0 \delta V$$

$\chi_0$: non-interacting density-density response function
independent-particle susceptibility
irreducible polarizability

Goal: compute $\delta \rho := \chi_0 \delta V$ for metals
Unperturbed Hamiltonian $H$

$$H \phi_n = \varepsilon_n \phi_n, \quad \int_{\Omega} \phi_m^*(r) \phi_n(r) \, dr = \delta_{mn}, \quad \varepsilon_1 \leq \varepsilon_2 \leq \varepsilon_3 \leq \cdots,$$

Ground-state density

$$\rho(r) = \sum_{n=1}^{+\infty} f_n |\phi_n(r)|^2 \quad \text{with} \quad f_n := f\left(\frac{\varepsilon_n - \varepsilon_F}{T}\right) \quad \text{and} \quad f(x) = \frac{2}{1 + e^x}$$

Occupation numbers $f_n$ for $T = 0$ (left) and $T > 0$ (right)

$\varepsilon_F$ (Fermi level) such that

$$\int_{\Omega} \rho(r) \, dr = \sum_{n=1}^{+\infty} f_n = \sum_{n=1}^{+\infty} f\left(\frac{\varepsilon_n - \varepsilon_F}{T}\right) = N_{\text{el}}$$

Notation: $f'_n := \frac{1}{T} f'\left(\frac{\varepsilon_n - \varepsilon_F}{T}\right) =: \frac{f_n - f_n'}{\varepsilon_n - \varepsilon_n}$
Linear response

\[ H \rightarrow H + \delta V \quad \Rightarrow \quad \rho \rightarrow \rho + \delta \rho + o(\delta V) \quad \text{with } \delta \rho \text{ linear in } \delta V \]

Sum-over-state formula

\[ \delta \rho(r) := (\chi_0 \delta V)(r) = \sum_{n=1}^{+\infty} \sum_{m=1}^{+\infty} \frac{f_n - f_m}{\varepsilon_n - \varepsilon_m} (\delta V_{mn} - \delta \varepsilon_F \delta_{mn}) \phi_n^*(r) \phi_m(r), \]

with \( \delta V_{mn} := \langle \phi_m, \delta V \phi_n \rangle \) and \( \delta \varepsilon_F = \frac{\sum_{n=1}^{+\infty} f'_n \delta V_{nn}}{\sum_{n=1}^{+\infty} f'_n} \) to satisfy \( \int_{\Omega} \delta \rho(r) \, dr = 0 \)
For periodic Schrödinger-like operators $H = \frac{1}{2}\Delta + V$ in 3D,

\[ \varepsilon_n \sim Cn^{2/3} \quad \text{(Weyl formula)} \]

In most applications (always except in Warm Dense Matter),

only a number $N \gtrsim N_{\text{el}}$ of occupation numbers $f_n$ are non-negligible

\[
\rho(r) \simeq \sum_{n=1}^{N} f_n |\phi_n(r)|^2
\]

\[
\delta \rho(r) \simeq \left( \sum_{n=1}^{N} \sum_{m=1}^{+\infty} + \sum_{n=1}^{+\infty} \sum_{m=1}^{N} \right) \frac{f_n - f_m}{\varepsilon_n - \varepsilon_m} (\delta V_{mn} - \delta \varepsilon_F \delta_{mn}) \phi_n^*(r) \phi_m(r)
\]

but still one infinite sum in the sum-over-state formula
Alternatives to the sum-over-state formula

\[ \rho(r) \simeq \sum_{n=1}^{N} f_n |\phi_n(r)|^2 \quad \Rightarrow \quad \delta \rho(r) \simeq \sum_{n=1}^{N} f_n (\phi_n^*(r) \delta \phi_n(r) + \delta \phi_n^*(r) \phi_n(r)) + \delta f_n |\phi_n(r)|^2 \]

the \( \delta \phi_n \)'s and the \( \delta f_n \)'s are not unique (gauge invariance)

Let \( P: \) orthogonal projector on \( \text{Span}(\phi_n)_{1 \leq n \leq N}, \) and \( Q := 1 - P. \) We have

\[ \delta \phi_n = \delta \phi_n^P + \delta \phi_n^Q \]

and

\[ \forall 1 \leq n \leq N, \quad f_n \delta \phi_n = \sum_{m=1}^{N} \Gamma_{mn} \phi_m + f_n \delta \phi_n^Q \quad \text{with} \quad \Gamma_{mn} := f_n \langle \phi_m, \delta \phi_n \rangle \]

Thus

\[ \delta \rho(r) \simeq \sum_{n,m=1}^{N} (\Gamma_{mn} + \Gamma_{nm}) \phi_n^*(r) \phi_m(r) + \sum_{n=1}^{N} 2f_n \text{Re} (\phi_n^*(r) \delta \phi_n^Q(r)) + \sum_{n=1}^{N} \delta f_n |\phi_n(r)|^2 \]
Alternatives to the sum-over-state formula (continued)

\[ \delta \rho(r) \simeq \sum_{n,m=1}^{N} \left( \Gamma_{mn} + \overline{\Gamma_{nm}} \right) \phi_n^*(r) \phi_m(r) + \sum_{n=1}^{N} 2f_n \text{Re} \left( \phi_n^*(r) \delta \phi_n^Q(r) \right) + \sum_{n=1}^{N} \delta f_n |\phi_n(r)|^2 \]

\[ = \left( \sum_{n=1}^{N} \sum_{m=1}^{+\infty} + \sum_{n=1}^{+\infty} \sum_{m=1}^{N} \right) \frac{f_n - f_m}{\varepsilon_n - \varepsilon_m} (\delta V_{mn} - \delta \varepsilon F \delta_{mn}) \phi_n^*(r) \phi_m(r) \]

By identification, we get that

1. the \( \Gamma_{mn} \)'s and the \( \delta f_n \)'s must satisfy

\[
\begin{cases}
2\text{Re}(\Gamma_{nn}) + \delta f_n = f'_n(\delta V_{nn} - \varepsilon_F) & \text{for } m = n \\
\Gamma_{mn} + \overline{\Gamma_{nm}} = \Delta_{mn} := \frac{f_n - f_m}{\varepsilon_n - \varepsilon_m} \delta V_{mn} & \text{for } m \neq n
\end{cases}
\]

\[ \rightarrow \text{ gauge choice needed} \]

2. the \( \delta \phi_n^Q \)'s must satisfy the Sternheimer equations

\[ Q(H - \varepsilon_n)Q \delta \phi_n^Q = b_n := -Q(\delta V \phi_n) \]  \[ \rightarrow \text{ possibly ill-conditioned} \]
Gauge choice for the $\Gamma_{mn}$’s and the $\delta f_n$’s

\[
2\text{Re}(\Gamma_{nn}) + \delta f_n = f'_n(\delta V_{nn} - \varepsilon_F) \quad \text{and} \quad \Gamma_{mn} + \Gamma_{nm} = \Delta_{mn} \quad \text{with} \quad \Delta_{mn} = \Delta_{nm} \quad \text{if} \quad m \neq n
\]
Gauge choice for the $\Gamma_{mn}$’s and the $\delta f_n$’s

$$2\text{Re}(\Gamma_{nn}) + \delta f_n = f_n'(\delta V_{nn} - \varepsilon_F) \quad \text{and} \quad \Gamma_{mn} + \Gamma_{nm} = \Delta_{mn} \quad \text{with} \quad \Delta_{mn} = \overline{\Delta_{nm}} \quad \text{if} \quad m \neq n$$

Natural (and good!) gauge choice for the $\Gamma_{nn}$’s ($m = n$) and the $\delta f_n$’s

$$\Gamma_{nn} = 0 \quad \text{and} \quad \delta f_n = f_n'(\delta V_{nn} - \varepsilon_F)$$
Gauge choice for the $\Gamma_{mn}$’s and the $\delta f_n$’s

$$2\text{Re}(\Gamma_{nn}) + \delta f_n = f_n'(\delta V_{nn} - \varepsilon_F) \quad \text{and} \quad \Gamma_{mn} + \Gamma_{nm} = \Delta_{mn} \text{ with } \Delta_{mn} = \Delta_{nm} \text{ if } m \neq n$$

Natural (and good!) gauge choice for the $\Gamma_{nn}$’s ($m = n$) and the $\delta f_n$’s

$$\Gamma_{nn} = 0 \quad \text{and} \quad \delta f_n = f_n'(\delta V_{nn} - \varepsilon_F)$$

Possible gauge choices for the $\Gamma_{mn}$’s ($m \neq n$)

1. orthonormal gauge: impose $\delta \langle \phi_m, \phi_n \rangle = 0$

$$\Gamma_{mn}^{\text{orth}} = \frac{f_n}{\varepsilon_n - \varepsilon_m} \delta V_{mn} \quad \text{(natural but very bad as the } \Gamma_{mn}^{\text{orth}} \text{ may blow up)}$$
Gauge choice for the $\Gamma_{mn}$’s and the $\delta f_n$’s

$$2 \text{Re}(\Gamma_{nn}) + \delta f_n = f_n' (\delta V_{nn} - \varepsilon_F) \quad \text{and} \quad \Gamma_{mn} + \Gamma_{nm} = \Delta_{mn} \quad \text{with} \quad \Delta_{mn} = \overline{\Delta_{nm}} \quad \text{if} \quad m \neq n$$

Natural (and good!) gauge choice for the $\Gamma_{nn}$’s $(m = n)$ and the $\delta f_n$’s

$$\Gamma_{nn} = 0 \quad \text{and} \quad \delta f_n = f_n' (\delta V_{nn} - \varepsilon_F)$$

Possible gauge choices for the $\Gamma_{mn}$’s $(m \neq n)$

1. orthonormal gauge: impose $\delta \langle \phi_m, \phi_n \rangle = 0$

$$\Gamma_{mn}^{\text{orth}} = \frac{f_n}{\varepsilon_n - \varepsilon_m} \delta V_{mn} \quad (\text{natural but very bad as the } \Gamma_{mn}^{\text{orth}} \text{ may blow up})$$

2. gauges used in Abinit and Quantum Espresso

$$\Gamma_{mn}^{\text{Abinit}} = 1_{f_n > f_m} \Delta_{mn}, \quad \Gamma_{mn}^{\text{QE}} = \frac{\Delta_{mn}}{1 + e^{(\varepsilon_n - \varepsilon_m)/T}}$$
Gauge choice for the $\Gamma_{mn}$’s and the $\delta f_n$’s

\[ 2\text{Re}(\Gamma_{nn}) + \delta f_n = f'_n(\delta V_{nn} - \varepsilon_F) \quad \text{and} \quad \Gamma_{mn} + \overline{\Gamma_{nm}} = \Delta_{mn} \text{ with } \Delta_{mn} = \overline{\Delta_{mn}} \text{ if } m \neq n \]

Natural (and good!) gauge choice for the $\Gamma_{nn}$’s ($m = n$) and the $\delta f_n$’s

\[ \Gamma_{nn} = 0 \quad \text{and} \quad \delta f_n = f'_n(\delta V_{nn} - \varepsilon_F) \]

Possible gauge choices for the $\Gamma_{mn}$’s ($m \neq n$)

1. orthonormal gauge: impose $\delta \langle \phi_m, \phi_n \rangle = 0$

\[ \Gamma_{\text{orth}}_{mn} = \frac{f_n}{\varepsilon_n - \varepsilon_m} \delta V_{mn} \quad \text{(natural but very bad as the } \Gamma_{\text{orth}}_{mn} \text{ may blow up)} \]

2. gauges used in Abinit and Quantum Espresso

\[ \Gamma_{\text{Abinit}}_{mn} = \mathbb{1}_{f_n > f_m} \Delta_{mn}, \quad \Gamma_{\text{QE}}_{mn} = \frac{\Delta_{mn}}{1 + e^{(\varepsilon_n - \varepsilon_m)/T}} \]

3. simple gauge and minimal gauge (used in DFTK)

\[ \Gamma_{\text{simple}}_{mn} = \frac{1}{2} \Delta_{mn} = \arg\min_{\Gamma_{mn} + \overline{\Gamma_{nm}} = \Delta_{mn}} |\Gamma_{mn}|^2 + |\overline{\Gamma_{nm}}|^2, \quad \Gamma_{\text{min}}_{mn} = \frac{f_n^2}{f_m^2 + f_n^2} \Delta_{mn} = \arg\min_{\Gamma_{mn} + \overline{\Gamma_{nm}} = \Delta_{mn}} \frac{1}{f_n^2} |\Gamma_{mn}|^2 + \frac{1}{f_m^2} |\overline{\Gamma_{nm}}|^2 \]
Gauge choice for the $\Gamma_{mn}$’s and the $\delta f_n$’s

For all these gauge choices but the orthogonal one, we have

$$|\Gamma_{mn}| \leq |\Delta_{mn}| \leq \max_{x \in \mathbb{R}} \frac{1}{T} |f'(x)| |\delta V_{mn}| = \frac{1}{2T} |\delta V_{mn}|.$$  \hspace{1cm} (1)

If we make an error on $\delta V$, it is at most amplified by a factor of $\frac{1}{2T}$.

Gauge comparison, $\varepsilon_n = 0$, $\varepsilon_F = 0$, $T = 0.1$, $\delta V_{mn} = 1$
Solving the Sternheimer equation \[ Q(H - \varepsilon_n)Q \delta \phi_n^Q = b_n \]

Output of the numerical diagonalization iterative scheme (ex. LOBPCG) used to compute the density

- **N** fully converged orthonormal eigenmodes \((\varepsilon_n, \phi_n)_{1 \leq n \leq N}\)

  \[ \Phi := (\phi_1, \cdots, \phi_N) \quad \text{orthonormal basis of} \quad \text{Ran}(P) = \text{Ker}(Q) \]

- **N_{ex}** extra, only partially converged, orthonormal eigenvectors

  \[ \Phi_{\text{ex}} := (\phi_{N+1}^\ell, \cdots, \phi_{N+N_{\text{ex}}}^\ell), \quad \ell : \# \text{ of iter. of the diagonalization scheme} \]
Solving the Sternheimer equation

\[ Q(H - \varepsilon_n)Q \delta \phi_n^Q = b_n \]

Output of the numerical diagonalization iterative scheme (ex. LOBPCG) used to compute the density

- \( N \) fully converged orthonormal eigenmodes \((\varepsilon_n, \phi_n)_{1 \leq n \leq N}\)
  \[ \Phi := (\phi_1, \ldots, \phi_N) \text{ orthonormal basis of } \text{Ran}(P) = \text{Ker}(Q) \]

- \( N_{\text{ex}} \) extra, only partially converged, orthonormal eigenvectors
  \[ \Phi_{\text{ex}} := (\phi_{N+1}^\ell, \ldots, \phi_{N+N_{\text{ex}}}^\ell), \quad \ell : \# \text{ of iter. of the diagonalization scheme} \]

We can assume without loss of generality that

- \((\Phi, \Phi_{\text{ex}})\) is an orthonormal family

- \( \Phi_{\text{ex}}^*H\rho\Phi_{\text{ex}} \in \mathbb{C}^{N_{\text{ex}} \times N_{\text{ex}}} \) is a diagonal matrix whose elements are labelled
  \[ \varepsilon_m^\ell := \langle \phi_m^\ell, H\phi_m^\ell \rangle \text{ for } N + 1 \leq m \leq N + N_{\text{ex}} \]

- \((\phi_m^\ell, \varepsilon_m^\ell) \xrightarrow{\ell \to \infty} (\phi_m, \varepsilon_m)\)
Solving the Sternheimer equation \( Q(H - \varepsilon_n)Q \delta \phi_n^Q = b_n \)

Output of the numerical diagonalization iterative scheme (ex. LOBPCG) used to compute the density

- \( N \) fully converged orthonormal eigenmodes \((\varepsilon_n, \phi_n)_{1 \leq n \leq N}\)
  \[ \Phi := (\phi_1, \cdots, \phi_N) \] orthonormal basis of \( \text{Ran}(P) = \text{Ker}(Q) \)

- \( N_{ex} \) extra, only partially converged, orthonormal eigenvectors
  \[ \Phi_{ex} := (\phi_{N+1}^\ell, \cdots, \phi_{N+N_{ex}}^\ell), \quad \ell : \# \text{ of iter. of the diagonalization scheme} \]

We can assume without loss of generality that

- \((\Phi, \Phi_{ex})\) is an orthonormal family

- \( \Phi_{ex}^* H \rho \Phi_{ex} \in \mathbb{C}^{N_{ex} \times N_{ex}} \) is a diagonal matrix whose elements are labelled
  \[ \varepsilon_{m}^\ell := \langle \phi_{m}^\ell, H \phi_{m}^\ell \rangle \] for \( N + 1 \leq m \leq N + N_{ex} \)

- \((\phi_{m}^\ell, \varepsilon_{m}^\ell) \xrightarrow{\ell \to \infty} (\phi_{m}, \varepsilon_{m})\)

Let’s use these extra approximate eigenvectors to improve the computation of response properties!
Solving the Sternheimer equation

\[ Q (H - \varepsilon_n) Q \delta \phi_n^Q = b_n \] (continued)

Schur complement method

\[ \delta \phi_n^Q = \Phi_{\text{ex}} \alpha_n + \delta \phi_n^R, \quad n \leq N \]

coercive, ill-conditioned, but small

\[
\begin{pmatrix}
\Phi_{\text{ex}}^* (H - \varepsilon_n) \Phi_{\text{ex}} \\
R H \Phi_{\text{ex}}
\end{pmatrix}
\begin{pmatrix}
\Phi_{\text{ex}}^* H R \\
R (H - \varepsilon_n) R
\end{pmatrix}
\begin{pmatrix}
\alpha_n \\
\delta \phi_n^R
\end{pmatrix}
= \begin{pmatrix}
\Phi_{\text{ex}}^* b_n \\
R b_n
\end{pmatrix}
\]

coercive, large, but well-conditioned
Solving the Sternheimer equation

\[ Q(H - \varepsilon_n)Q \delta \phi_n^Q = b_n \]  

(continued)

Schur complement method

\[ \delta \phi_n^Q = \Phi_{ex} \alpha_n + \delta \phi_n^R, \quad n \leq N \]

\[ A_n := (\Phi_{ex}^*(H - \varepsilon_n)\Phi_{ex})^{-1}, \quad \alpha_n = A_n (\Phi_{ex}^* b_n - \Phi_{ex}^* (H - \varepsilon_n) \delta \phi_n^R) \]

\[ \left[ R(H - \varepsilon_n) \left( 1 - \Phi_{ex} A_n \Phi_{ex}^* (H - \varepsilon_n) \right) R \right] \delta \phi_n^R = Rb_n - R(H - \varepsilon_n) \Phi_{ex} A_n \Phi_{ex}^* b_n \]
Numerical results: convergence of the Sternheimer solver for Al_{40}

<table>
<thead>
<tr>
<th>$k$-point – coordinate</th>
<th>[0, 0, 0]</th>
<th>[1/3, 0, 0]</th>
<th>[1/3, 1/3, 0]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>69</td>
<td>58</td>
<td>67</td>
</tr>
<tr>
<td>$\varepsilon_{N+1} - \varepsilon_N$</td>
<td>0.0320</td>
<td>0.0134</td>
<td>0.0217</td>
</tr>
<tr>
<td># CG iterations $n = N$ Schur</td>
<td>48</td>
<td>44</td>
<td>41</td>
</tr>
<tr>
<td># CG iterations $n = N$ direct</td>
<td>56</td>
<td>83</td>
<td>58</td>
</tr>
</tbody>
</table>
Conclusion
Error bounds for DFT

- guaranteed, optimal, cheap error bounds on the ground-state energy and density for linear Schrödinger equations
- practical (not guaranteed), optimal, cheap error bounds on the ground-state energy and density for DFT
- practical (not guaranteed), quite accurate, not too expensive error bounds for non-variational quantities such as forces for DFT
Conclusion

Error bounds for DFT

• guaranteed, optimal, cheap error bounds on the ground-state energy and density for linear Schrödinger equations
• practical (not guaranteed), optimal, cheap error bounds on the ground-state energy and density for DFT
• practical (not guaranteed), quite accurate, not too expensive error bounds for non-variational quantities such as forces for DFT

Response properties

• easy for insulators
• harder for metals
  – gauge choice is important
    we have clarified this point and proposed new, very simple gauges
  – the Sternheimer equation is ill-conditioned
    we have improved its resolution using a Schur complement method
Thank you for your attention