Efficient Krylov methods for linear response in plane-wave electronic structure calculations

Michael F. Herbst[1], Bonan Sun[1,2]

[1]Mathematics for Materials Modelling, Institute of Mathematics & Institute of Materials, EPFL, 1015 Lausanne, Switzerland (michael.herbst@epfl.ch, bonan.sun@epfl.ch)
[2]Max Planck Institute for Dynamics of Complex Technical Systems, 39106 Magdeburg, Germany (bsun@mpi-magdeburg.mpg.de)

Density Functional Perturbation Theory (DFPT)

Kohn-Sham density functional theory (KS-DFT)

• KS-DFT: most widely used electronic structure calculation method for materials simulations.

$\mathcal{F}$: external potential $V_{\text{ext}}(\boldsymbol{r})$ $\xrightarrow[\text{solve a non-linear eigenvalue problem}]{\text{KS-DFT}}$ ground state electron density $\rho(\boldsymbol{r})$

DFT properties are derivatives: DFPT

• Properties of interest: response of system to external changes, i.e., derivatives of ρ w.r.t. V_{ext}:

$$\delta\rho = \mathcal{F}' \cdot \delta V_0. \tag{1}$$

• In plane-wave basis, Eq. (1) leads to the discretized **Dyson equation**:

$$\text{Seek } \boldsymbol{\delta\rho} \in \mathbb{R}^{N_g} \text{ such that } (\boldsymbol{I}_{N_g} - \boldsymbol{\chi}_0\boldsymbol{K})\boldsymbol{\delta\rho} = \boldsymbol{\delta\rho}_0,\ \boldsymbol{I}_{N_g},\ \boldsymbol{\chi}_0,\ \boldsymbol{K} \in \mathbb{R}^{N_g\times N_g}. \tag{2}$$

$\boldsymbol{\mathcal{E}} := \boldsymbol{I}_{N_g} - \boldsymbol{\chi}_0\boldsymbol{K}$: adjoint of the dielectric operator. Each M-v product $\boldsymbol{\mathcal{E}}\boldsymbol{\delta\rho}$ requires solving N_{occ} discretized **Sternheimer equations** (3):

$$\boldsymbol{\mathcal{E}}\boldsymbol{\delta\rho} = \mathcal{L}(\boldsymbol{\delta\phi}_1, \cdots, \boldsymbol{\delta\phi}_{N_{\text{occ}}}),\ \text{where } \boldsymbol{\delta\phi}_n \text{ solves (3)}$$
$$(\widehat{\boldsymbol{H}} - \epsilon_n\boldsymbol{P})\boldsymbol{\delta\phi}_n = -\boldsymbol{PF}\left((\boldsymbol{K}\boldsymbol{\delta\rho}) \odot (\boldsymbol{F}^{-1}\boldsymbol{\phi}_n)\right),\ n = 1, \cdots, N_{\text{occ}}, \tag{3}$$
$$\widehat{\boldsymbol{H}},\ \boldsymbol{P} \in \mathbb{C}^{N_b\times N_b}, \boldsymbol{\delta\phi}_n \in \mathbb{C}^{N_b},\ N_g \approx 16N_b.$$

$\mathcal{L}$ some multilinear map, $H = -\frac{1}{2}\Delta + V_{\text{ext}} + V_{\text{Hxc}}$, $\widehat{H} = PHP$ for some projector P, (ϵ_n, ψ_n) eigenpairs of H

Orbitals, Hamiltonians are represented in $\mathcal{X}^{\bigcirc} := \text{span}\{e_{\boldsymbol{G}} | \boldsymbol{G} \in \mathbb{L}^*, \|\boldsymbol{G}\|_2 \le \sqrt{2E_{\text{cut}}}\}$, E_{cut} cutoff energy

Densities, potentials are represented in $\mathbb{X} := \left\{(i_1\boldsymbol{a}_1 + i_2\boldsymbol{a}_2 + i_3\boldsymbol{a}_3)/\sqrt[3]{N_g} \,|\, 0 \le i_1, i_2, i_3 < \sqrt[3]{N_g}\right\}$.

For technical reasons: $N_g \approx 16N_b$, $N_g := \dim \mathbb{X}$, $N_b := \dim \mathcal{X}^{\bigcirc}$.

Transform from real $\mathbb{X}$ to Fourier $\mathcal{X}^{\bigcirc}$ spaces $\boldsymbol{F} \in \mathbb{C}^{N_b\times N_g}$ & vice versa $\boldsymbol{F}^{-1} \in \mathbb{C}^{N_g\times N_b}$ using FFTs.

$\phi_n, \delta\phi_n, P, \widehat{H} \xrightarrow{\text{discretize in } \mathcal{X}^{\bigcirc}} \boldsymbol{\phi}_n, \boldsymbol{\delta\phi}_n \in \mathbb{C}^{N_b}, P, \widehat{\boldsymbol{H}} \in \mathbb{C}^{N_b\times N_b}$

$\delta\rho, \delta V, \chi_0, K, \mathcal{E} \xrightarrow{\text{discretize in } \mathbb{X}} \boldsymbol{\delta\rho}, \boldsymbol{\delta V} \in \mathbb{R}^{N_g}, \boldsymbol{\chi}_0, \boldsymbol{K}, \boldsymbol{\mathcal{E}} \in \mathbb{R}^{N_g\times N_g}$

Comments on the Dyson equation:

• The Dyson equation (2), real non-symmetric, is solved by GMRES with `tol` τ.
• At i-th GMRES iter, N_{occ} Sternheimer's (3), complex Hermitian, are solved by CG with `tol` $\tau_{i,n}^{\text{CG}}$.
• Dyson + Sternheimer's: **nested iteratively** solved linear systems.
• **Difficulty**: How to choose $\tau_{i,n}^{\text{CG}}$?
• **Ideal** $\tau_{i,n}^{\text{CG}}$'s: (a) **accuracy** τ of Dyson is **guaranteed** with (b) **minimal computational cost** dominated by # Hamiltonian applications $N = \sum_{i,n} N_{i,n}^{\text{CG}}$?

Efficient Krylov subspace methods for Dyson equation

Algorithm: Standard generalized minimal residual method (GMRES)

Require: $\boldsymbol{\mathcal{E}} \in \mathbb{R}^{N_g\times N_g}$, $\boldsymbol{b} \in \mathbb{R}^{N_g}$ as in (2), initial guess $\boldsymbol{x}_0 \in \mathbb{R}^{N_g}$, `tol` $\tau > 0$, restart period $m \in \mathbb{N}$
Ensure: $\boldsymbol{x} \in \mathbb{R}^{N_g}$ s.t. $\|\boldsymbol{r}_m\|_2 \le \tau$ where $\boldsymbol{r}_m := \boldsymbol{b} - \boldsymbol{\mathcal{E}}\boldsymbol{x}_m \in \mathbb{R}^{N_g}$ is the true residual

1: $\boldsymbol{r}_0 = \boldsymbol{b} - \boldsymbol{\mathcal{E}}\boldsymbol{x}_0$, $\beta = \|\boldsymbol{r}_0\|$, $\boldsymbol{v}_1 = \boldsymbol{r}_0/\beta$, $\boldsymbol{V}_1 = \boldsymbol{v}_1$, $\boldsymbol{H}_0 = [\,]$
2: **for** $k = 1, 2, \cdots, m$ **do**
3: Compute $\boldsymbol{w} = \boldsymbol{\mathcal{E}}\boldsymbol{v}_k$ by solving N_{occ} Sternheimer's (3)
4: build orthonormal basis $\boldsymbol{V}_{k+1} \in \mathbb{R}^{N_g\times(k+1)}$ & Hessenberg matrix $\boldsymbol{H}_k \in \mathbb{R}^{(k+1)\times k}$ by Arnoldi
5: Compute the estimated residual $\widetilde{\boldsymbol{r}}_k := \beta\boldsymbol{e}_1 - \boldsymbol{H}_k\boldsymbol{y}_k$ where $\boldsymbol{y}_k = \arg\min_{\boldsymbol{y}\in\mathbb{C}^k} \|\beta\boldsymbol{e}_1 - \boldsymbol{H}_k\boldsymbol{y}\|_2$
6: **if** $\|\widetilde{\boldsymbol{r}}_k\| \le \tau$ **then return** $\boldsymbol{x} = \boldsymbol{x}_0 + \boldsymbol{V}_k\boldsymbol{y}_k$ **end if**
7: **end for**
8: **if** $\|\widetilde{\boldsymbol{r}}_m\| > \tau$ **then** restart: update $\boldsymbol{x}_0 = \boldsymbol{x}_0 + \boldsymbol{V}_m\boldsymbol{y}_m$ and go to Line 1 **end if**

• Consequence of Arnoldi relation $\boldsymbol{\mathcal{E}}\boldsymbol{V}_m = \boldsymbol{V}_{m+1}\boldsymbol{H}_m$: $\|\boldsymbol{r}_m\|_2 = \|\widetilde{\boldsymbol{r}}_m\|_2$. Hence $\|\widetilde{\boldsymbol{r}}_m\|_2 \le \tau \Rightarrow \|\boldsymbol{r}_m\|_2 \le \tau$.

An Inexact GMRES method

• Observation — $\boldsymbol{\mathcal{E}}\boldsymbol{v}_k$ in Line 3 is inexact: the Sternheimer's are solved by CG with `tol` $\tau_{i,n}^{\text{CG}} > 0$.
• Consequence of inexactness: inexact Arnoldi relation: $\left[\widetilde{\boldsymbol{\mathcal{E}}}^{(1)}\boldsymbol{v}_1, \widetilde{\boldsymbol{\mathcal{E}}}^{(2)}\boldsymbol{v}_2, \cdots, \widetilde{\boldsymbol{\mathcal{E}}}^{(m)}\boldsymbol{v}_m\right] = \boldsymbol{V}_{m+1}\boldsymbol{H}_m$
$\Longrightarrow$ $\|\boldsymbol{r}_m\|_2 \ne \|\widetilde{\boldsymbol{r}}_m\|_2$, therefore, the stopping criterion $\|\widetilde{\boldsymbol{r}}_m\|_2 \le \tau \not\Longrightarrow \|\boldsymbol{r}_m\|_2 \le \tau$.
$\Longrightarrow$ the stopping criterion $\|\widetilde{\boldsymbol{r}}_m\|_2 \le \tau$ needs to be modified carefully.

Theorem: $\forall\tau > 0$, the accuracy of the Dyson equation $\|\boldsymbol{r}_m\|_2 \le \tau$ is guaranteed if (a) $\|\widetilde{\boldsymbol{r}}_m\|_2 \le \tau/3$ and (b) for all $i = 1, \cdots, m$

$$\tau_{i,n}^{\text{CG}} \le \frac{|\Omega|^{1/2}(\epsilon_{N_{\text{occ}}+1} - \epsilon_n)}{2f_n\|\boldsymbol{K}\boldsymbol{v}_i\|\,\|\text{Re}(\boldsymbol{F}^{-1}\boldsymbol{\Phi})\|_{2,\infty} N_g^{1/2}N_{\text{occ}}^{1/2}} \frac{\sigma_m(\boldsymbol{H}_m)}{3m} \frac{1}{\|\widetilde{\boldsymbol{r}}_{i-1}\|}\tau \tag{4}$$

Corollary: If the CG tolerances $\tau_{i,n}^{\text{CG}}$ verify (4) and the convergence of GMRES is at least linear, then the true residual decreases at least superlinerly w.r.t. the total number of Hamiltonian applications:

$$\|\boldsymbol{r}_i\| \lesssim C_1^{\sqrt{C_2 - N_{\le i}^{\text{CG}}} + C_3}\|\boldsymbol{r}_0\|. \tag{5}$$

Main features of our result:

• **Guaranteed** accuracy of the Dyson equation with **computable** CG `tol`.
• **Looser** CG `tol` closer to convergence (as $\|\widetilde{\boldsymbol{r}}_{i-1}\| \to 0$, note $\|\widetilde{\boldsymbol{r}}_{i-1}\| \searrow$).
• **Adaptive** CG `tol` w.r.t. the properties of Sternheimer's (eigengap $\epsilon_{N_{\text{occ}}+1} - \epsilon_n$, occupation number f_n).
• **Easy implementation** in DFTK `https://dftk.org`
• **First rigorous error analysis** for the numerical solution of plane-wave Dyson equation.
• Methodology **easy to extend** to perturbation theories of other related mean-field models.

Aggressive modifications of (4) for practical implementations:

• Using Schur trick [2] for Sternheimer's $\Rightarrow$ lower eigenvalue $\gtrsim \mathcal{O}(1) \Rightarrow \epsilon_{N_{\text{occ}}+1} - \epsilon_n$ may be dropped
• Dyson equation is preconditioned $\Rightarrow \|\boldsymbol{K}\boldsymbol{v}_i\| = \mathcal{O}(1) \Rightarrow$ it may be dropped.
• Fourier decay of $\phi_n \Rightarrow \|\boldsymbol{F}^{-1}\boldsymbol{\Phi}\|_{2,\infty}$ is small $\Rightarrow$ it may be replaced by its lower bound $\sqrt{N_{\text{occ}}}/\sqrt{|\Omega|}$.
$\Longrightarrow$ We propose 3 **adaptive strategies** to choose $\tau_{i,n}^{\text{CG}}$ in practice, as specified in the next part.

Numerical simulations

Testing strategies

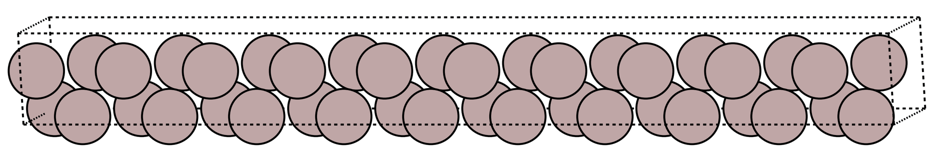
Heuristics to modify (4) for practical usage from last part suggests 3 adaptive strategies as below:

Adaptive strategies	$\tau_{i,n}^{\text{CG}}$
Pagr	$1 \cdot \frac{\sigma_m(\boldsymbol{H}_m)}{3m}\frac{1}{\|\widetilde{\boldsymbol{r}}_{i-1}\|}\tau$
Phdmd	$\frac{\lvert\Omega\rvert}{2f_nN_g^{1/2}N_{\text{occ}}} \cdot \frac{\sigma_m(\boldsymbol{H}_m)}{3m}\frac{1}{\|\widetilde{\boldsymbol{r}}_{i-1}\|}\tau$
Pgrt	$\frac{\lvert\Omega\rvert^{1/2}}{2f_n\|\boldsymbol{K}\boldsymbol{v}_i\|\,\|\text{Re}(\boldsymbol{F}^{-1}\boldsymbol{\Phi})\|_{2,\infty}N_g^{1/2}N_{\text{occ}}^{1/2}} \cdot \frac{\sigma_m(\boldsymbol{H}_m)}{3m}\frac{1}{\|\widetilde{\boldsymbol{r}}_{i-1}\|}\tau$

3 naive strategies: `PD10`, `PD100`, `PD10_n` where $\tau_{i,n}^{\text{CG}} = \tau/10, \tau/100, \tau/(10\|\delta\rho_0\|_2)$, $\forall i, n$, respectively.

Numerical results

Al_{40} **Supercell:** $N_g = 911\,250$, $N_b \approx 54\,200$



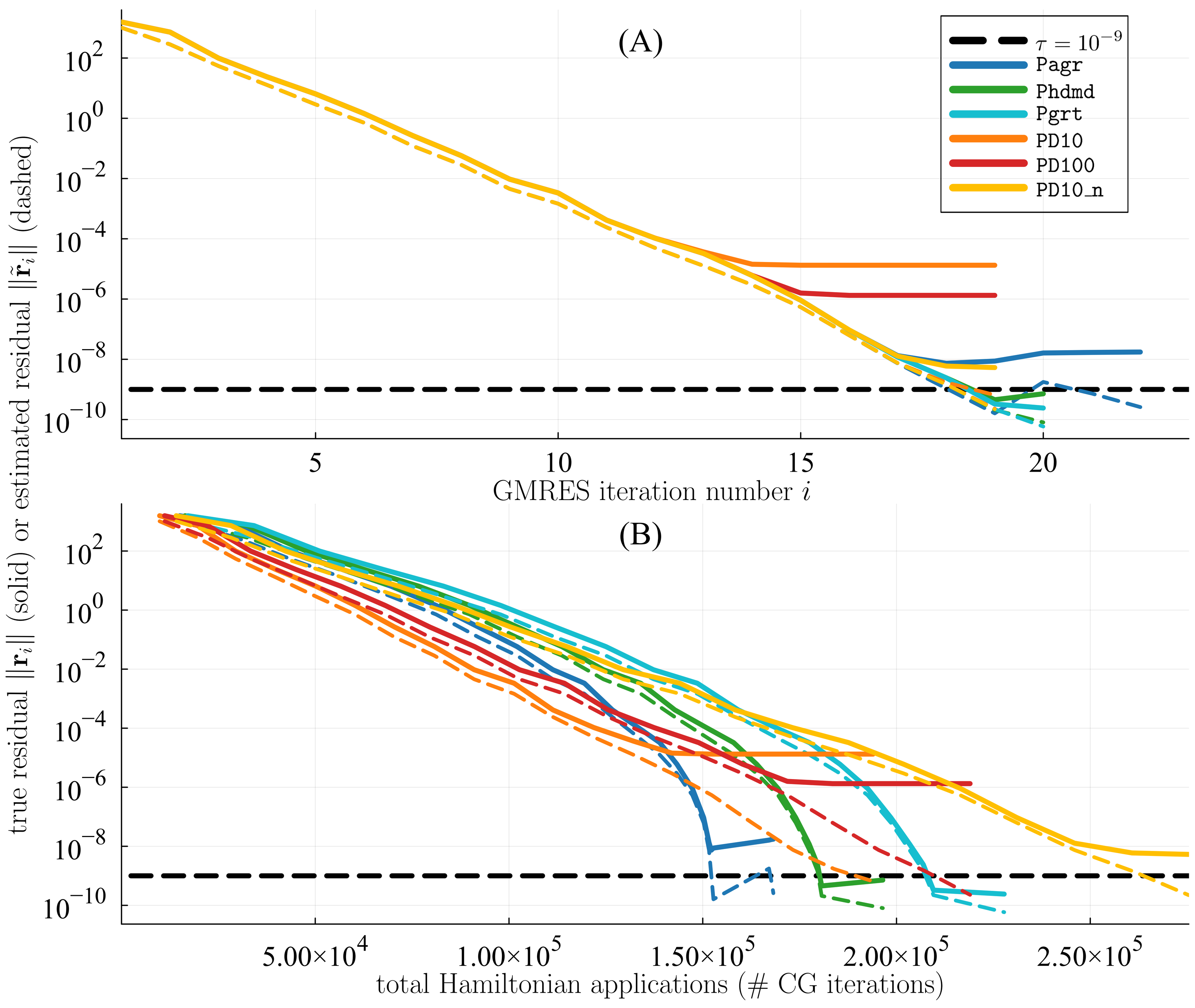


Figure 1: True residual norms $\|\boldsymbol{r}_i\|$ (in solid lines) and estimated residual norms $\|\widetilde{\boldsymbol{r}}_i\|$ (in dashed lines) v.s. (A) GMRES iteration number i and (B) total number of Hamiltonian applications for the Dyson equation of the Al_{40} supercell system. Target tolerance $\tau = 10^{-9}$, GMRES is restarted every $m = 20$ iterations and Kerker preconditioning is applied.

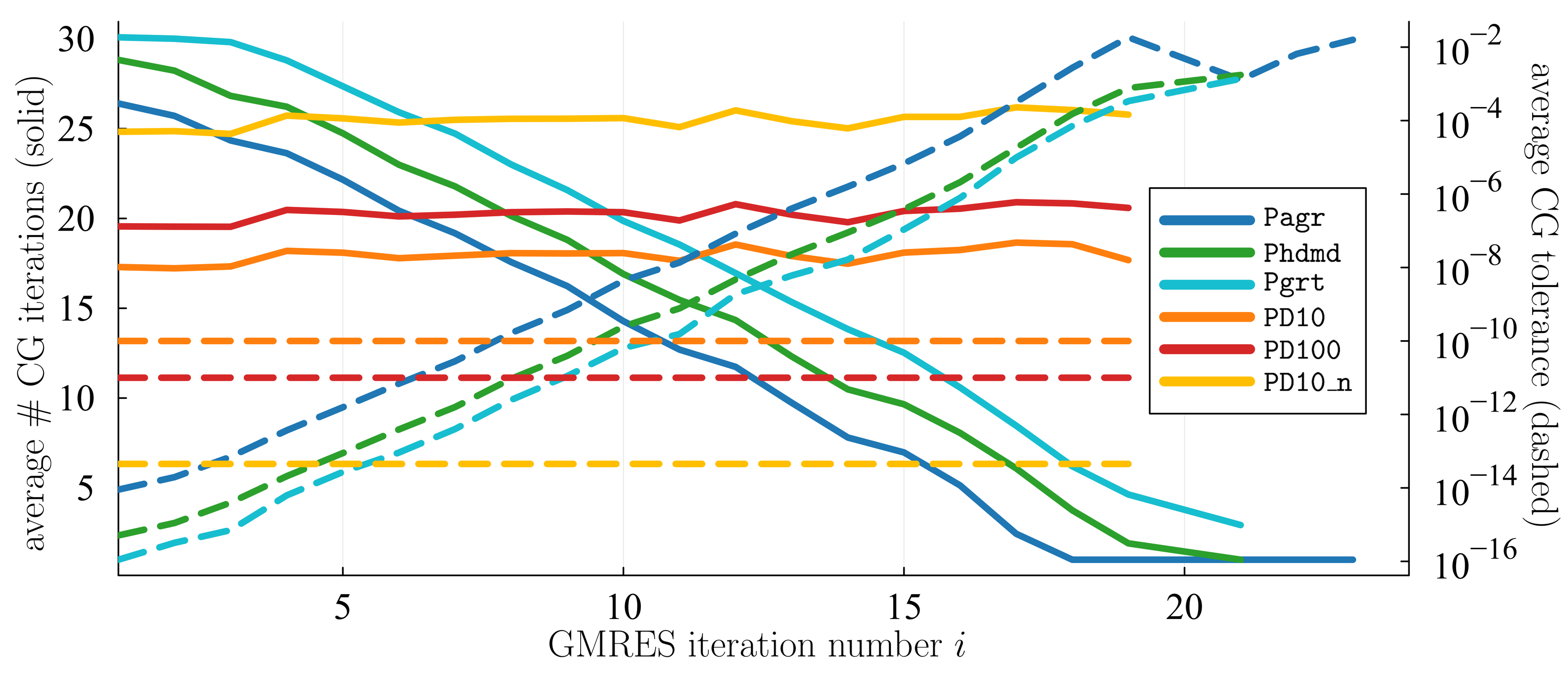


Figure 2: Average number of CG iterations (in solid lines) and geometric average of CG tolerances (in dashed lines) per GMRES iteration for the Dyson equation.

Strategy	Pagr	Phdmd	Pgrt	PD10	PD100	PD10_n
$\|\boldsymbol{r}_{\text{end}}\|$	$1.74\cdot10^{-8}$	$7.17\cdot10^{-10}$	$2.41\cdot10^{-10}$	$1.33\cdot10^{-5}$	$1.33\cdot10^{-6}$	$5.32\cdot10^{-9}$
N^{CG}	168 k	196 k	228 k	194 k	219 k	275 k

Table 1: Returned true residual norm $\|\boldsymbol{r}_{\text{end}}\|$, total number of Hamiltonian applications N^{CG} for different strategies. Our adaptive strategies are highlighted in blue, and the top three strategies for each metric are highlighted in red.

Conclusion

• Reliability: **guaranteed accuracy** of the Dyson equation with **computable** CG tolerances.
• Efficiency: **superlinear convergence** w.r.t. # Hamiltonian applications.
• We achieve **more accurate** results with **fewer** Hamiltonian applications! $\sim 1.5\times$ speedup.

References

[1] M. F. Herbst, B. Sun, *Efficient Krylov methods for linear response in plane-wave electronic structure calculations*, available on *arXiv* soon.
[2] E. Cancès, M. F. Herbst, G. Kemlin, A. Levitt, B. Stamm. Numerical stability and efficiency of response property calculations in density functional theory. *Lett. Math. Phys.*, 113(1), 21.