A robust and efficient line search for self-consistent field iterations

Michael F. Herbst*
Applied and Computational Mathematics, RWTH Aachen University, Schinkelstr. 2, 52062 Aachen, Germany.

Antoine Levitt†
Inria Paris and CERMICS, École des Ponts, 6 & 8 avenue Blaise Pascal, 77455 Marne-la-Vallée, France.

We propose a novel adaptive damping algorithm for the self-consistent field (SCF) iterations of Kohn-Sham density-functional theory, using a backtracking line search to automatically adjust the damping in each SCF step. This line search is based on a theoretically sound, accurate and inexpensive model for the energy as a function of the damping parameter. In contrast to usual SCF schemes, the resulting algorithm is fully automatic and does not require the user to select a damping. We successfully apply it to a wide range of challenging systems, including elongated supercells, surfaces and transition-metal alloys.

I. INTRODUCTION

Ab initio simulation methods are standard practice for predicting the chemical and physical properties of molecules and materials. At the level of simulating electronic structure the majority of approaches are either directly based upon Kohn-Sham density-functional theory (DFT) or Hartree-Fock (HF) or use these techniques as starting points for more accurate post-DFT or post-HF developments. Both HF and DFT ground states are commonly found by solving the self-consistent field (SCF) equations, which for both type of methods are very similar in structure. Being thus fundamental to electronic-structure simulations substantial effort has been devoted in the past to develop efficient and widely applicable SCF algorithms. We refer to Woods *et al.* [1] and Lehtola *et al.* [2] for recent reviews on this subject.

However, the advent of both cheap computational power as well as the introduction of data-driven approaches to materials modelling has caused simulation practice to change noticeably. In particular in domains such as catalysis or battery research where experiments are expensive or time-consuming, it is now standard practice to perform systematic computations on thousands to millions of compounds. The aim of such high-throughput calculations is to either (i) generate data for training sophisticated surrogate models or to (ii) directly screen complete design spaces for relevant compounds. The development of such data-driven strategies has already accelerated research in these fields and enabled the discovery of novel semiconductors, electrocatalysts, materials for hydrogen storage or for Li-ion batteries[3–5].

Compared to the early years where the aim was to perform a small number of computations on hand-picked systems, high-throughput screening approaches have much stronger requirements. In particular the key bottleneck is the required human time to set up and supervise computations. To minimize manual effort state-of-the-art high-throughput frameworks[6–8] provide a set of heuristics which automatically select computational parameters based on prior experience. In case of a failing calculation such heuristics may also be employed for parameter adjustment and automatic rescheduling. While this empirical approach manages to take care of the majority of failures automatically, it is far from perfect. First, state-of-the-art heuristic approaches cannot capture all cases, and keeping in mind the large absolute number of calculations already a 1% fraction of cases that require human attention easily equals hundreds to thousands of calculations. This causes idle time and severely limits the overall throughput of a study. Second, any failing calculation, whether automatically caught by a high-throughput framework or not, needs to be redone, implying wasted computational resources that contributes to the already noteworthy environmental footprint of supercomputing[9,10]. The objectives for improving the algorithms employed in high-throughput workflows is therefore to increase the inherent reliability as well as reduce the number of parameters, which need to be chosen. Ideally each building block of a simulation workflow would be entirely black-box and automatically self-adapt to each simulated system. To some extent this amounts to taking the existing empirical wisdom already implemented in existing high-throughput frameworks and converting it into simulation algorithms with convergence guarantees using a mixture of both mathematical and physical arguments.

With this objective in mind, this work will focus on improving the robustness of self-consistent field (SCF) algorithms, as mentioned above one of the most fundamental components of electronic-structure simulations. Our main motivation and application are DFT simulations discretized in plane wave or "large" basis sets, for which it is only feasible to store and compute with orbitals, densities and potentials, and not the full density matrix or Fock matrix. In this setting, the standard SCF approach are damped, preconditioned self-consistent iterations. Using an approach based on potential-mixing the next SCF iterate is found as

$$V_{\text{next}} = V_{\text{in}} + \alpha P^{-1}(V_{\text{out}} - V_{\text{in}}), \tag{1}$$

where V_{in} and V_{out} are the input and output potentials to a simple SCF step, α is a fixed damping parameter and P is a preconditioner. It is well-known that simple SCF

iterations (where P is the identity) can converge poorly for many systems due to a number of instabilities[11]. Examples are the large-wavelength divergence due to the Coulomb operator leading to the "charge-sloshing" behavior in metals or the effect of strongly localized states near the Fermi level, e.g. due to surface states or d- or f-orbitals. To accelerate the convergence of the SCF iteration despite these instabilities, one typically aims to employ a preconditioner P matching the underlying system. Despite some recent progress towards cheap self-adapting preconditioning strategies[11] for the charge-sloshing-type instabilities, choosing a matching preconditioner is still not a straightforward task for other types of instabilities. For example currently no cheap preconditioner is available to treat the instabilities due to strongly localized states near the Fermi level, such that in such systems using a suboptimal preconditioning strategy is unavoidable. While convergence acceleration techniques are usually crucial in such cases, these also complicate the choice of an appropriate damping parameter α to achieve the fastest and most reliable convergence. As we will detail in a series of example calculations on some transition metal systems the interplay of mismatching preconditioner and convergence acceleration can lead to a very unsystematic pattern between the chosen damping parameter α and obtaining a successful or failing calculation. Especially for such cases finding a good combination of preconditioning strategy and damping parameter can require substantial trial and error.

As an alternative approach to a fixed damping selected by a user *a priori* Cancès and Le Bris suggested the optimal damping algorithm (ODA)[12,13]. In this algorithm the damping parameter is obtained automatically by performing a line search along the update suggested by a simple SCF step. Following this strategy, the ODA ensures a monotonic decrease of the energy, which leads to strong convergence guarantees. This can be improved using the history to improve convergence, such as in the EDIIS method[14], or trust-region strategies[15,16]. These approaches are successfully employed for SCF calculations on atom-centered basis sets, where an explicit representation of the density matrix is possible. However, their use with plane-wave DFT methods, where only orbitals, densities and potentials are ever stored, does not appear to be straightforward, in particular in conjunction with accelerated methods.

Another development towards finding an DFT ground state in a mathematically guaranteed fashion are approaches based on a direct minimization of the DFT energy as a function of the orbitals and occupations, not using the self-consistency principle (see Reference 17 for a mathematical comparison). Although direct minimization methods are often quite efficient for gapped systems, their use for metals requires a minimization over occupation numbers[18,19], which is potentially costly and unstable. For this reason such approaches seem to be less used than the SCF schemes in solid-state physics.

In the realm of self-consistent iterations, variable-step methods have been successfully used[20,21] to increase robustness. These methods are based on a minimization of the residual. Although this often proves efficient in practice, this has a number of disadvantages. First, the residual might go up then down on the way to a solution making it rather hard to design a linesearch algorithm. Second, this forces an algorithm to select an appropriate notion of a residual norm, with results potentially sensitive to this choice. Third, there is the possibility of getting stuck in local minima of the residual, or a saddle point of the energy. By contrast, we aim to find a scheme ensuring energy decrease as an important ingredient to ensure robustness. Indeed, under mild conditions, a scheme that decreases the energy monotonically is guaranteed to converge to a solution of the Kohn-Sham equations (see Theorems 1 and 2 below). This is in contrast to residual-based schemes, which afford no such guarantee. The very good practical performance of these schemes, despite the lack of global theoretical guarantees, is an interesting direction for future research.

Our goal in this work is to design a mixing scheme that (a) is applicable to plane-wave DFT, and involves only quantities such as densities and potentials; (b) is based on an energy minimization, to ensure robustness; (c) is based on the self-consistent iterations; (d) is compatible with acceleration and preconditioning. Our scheme is based on a minimal modification of the damped preconditioned iterations (1). Similar to the ODA approach we employ a line search procedure to choose the damping parameter automatically. Our algorithm builds upon ideas of the potential-based algorithm of Gonze [22] to construct an efficient SCF algorithm. In combination with Anderson acceleration on challenging systems we show our adaptive damping scheme to be less sensitive than the approach based on a fixed damping parameter. In contrast to the fixed damping approach the scheme does not require a manual damping selection from the user.

The outline of the paper is as follows. Section II presents the mathematical analysis of the self-consistent field iterations justifying our algorithmic developments. In particular it presents a justification for global convergence of the SCF iterations. The proofs for the results presented in this section are given in the appendix. Section III discusses the adaptive damping algorithm itself followed by numerical tests (Section IV) to illustrate and contrast cost and performance compared to the standard fixed-damping approach. Concluding remarks and some outlook to future work is given in Section V.

II. ANALYSIS

A. Preliminaries

We use similar notation to those in Cancès *et al.* [17], extend the analysis in that paper to the finite-temperature case[23], and introduce the potential mixing algorithm. We work in the *grand-canonical ensemble*: we fix a chemical

potential (or Fermi level) μ and an inverse temperature β. In particular, the number of electrons is not fixed. This is for mathematical convenience: fixing the number of electrons N instead of μ does not change our results. We assume that space has been discretized in a finite-dimensional orthogonal basis (typically, plane-waves) of size N_b, and will not treat either spin or Brillouin zone sampling explicitly for notational simplicity, although of course the formalism can be extended easily. In this section we will work with the formalism of density matrices, self-adjoint operators P satisfying $0 \leq P \leq 1$. Such operators can be diagonalized as

$$P = \sum_{i=1}^{N_\mathrm{b}} f_i |\phi_i\rangle\langle\phi_i|. \tag{2}$$

The numbers $0 \leq f_i \leq 1$ are the occupation numbers, and ϕ_i are the orbitals. Either density matrices or the set of occupation numbers and orbitals can be taken as the primary unknowns in the self-consistency problem. Density matrices are impractical numerically in plane-wave basis sets, since they are $N_\mathrm{b} \times N_\mathrm{b}$; however, they are very convenient to formulate and analyze algorithms. Accordingly, we will use them in this theoretical section, but implement the resulting algorithms using orbitals only.

We work on the sets

$$\mathcal{H} = \{H \in \mathbb{R}^{N_\mathrm{b} \times N_\mathrm{b}}, H^T = H\} \tag{3}$$
$$\mathcal{P} = \{P \in \mathcal{H}, 0 < P < 1\} \tag{4}$$

of Hamiltonians and density matrices, equipped with the standard Frobenius metric. Here and in the following, inequalities between matrices are understood in the sense of symmetric matrices. The closure $\overline{\mathcal{P}} = \{P \in \mathcal{H}, 0 \leq P \leq 1\}$ is compact. Let $\mathcal{E}_0$ be a twice continuously differentiable function on $\overline{\mathcal{P}}$: we aim to solve the problem

$$\min_{P \in \mathcal{P}} \mathcal{E}_0(P). \tag{5}$$

Let

$$H_\mathrm{KS}(P) = \nabla \mathcal{E}_0(P) \tag{6}$$

be its gradient, and

$$\boldsymbol{K}(P) = \boldsymbol{d^2 \mathcal{E}_0}(P) = \boldsymbol{d \nabla \mathcal{E}_0}(P) \tag{7}$$

be its Hessian. We will denote in bold "super-operators" or "four-point operators", operators from $\mathcal{H}$ to $\mathcal{H}$. Let s be the fermionic entropy

$$s(p) = -(p \log p + (1-p)\log(1-p)), \tag{8}$$

with derivatives

$$s'(p) = \log\left(\frac{1-p}{p}\right), \quad s''(p) = -\frac{1}{p(1-p)}. \tag{9}$$

Let

$$\mathcal{E}(P) = \mathcal{E}_0(P) - \frac{1}{\beta}\mathrm{Tr}(s(P)) - \mu \mathrm{Tr} P. \tag{10}$$

be the free energy of a density matrix, where here and in the following we use functional calculus implicitly to define $s(P) \in \mathcal{H}$. $\mathcal{E}$ diverges on the boundary of $\mathcal{P}$, whose closure is compact, and therefore $\mathcal{E}$ has at least one minimizer in $\mathcal{P}$. The first-order optimality condition $\nabla \mathcal{E}(P) = 0$ gives

$$H_\mathrm{KS}(P) - \mu - \frac{1}{\beta} s'(P) = 0, \tag{11}$$

and therefore

$$P = f_\mathrm{FD}(H_\mathrm{KS}(P)), \tag{12}$$

where we define the Fermi-Dirac map f_FD by

$$f_\mathrm{FD}(H) = \frac{1}{1 + e^{\beta(H-\mu)}}. \tag{13}$$

Here we have used the equation $s'(f_\mathrm{FD}(\varepsilon)) = \beta(\varepsilon - \mu)$, which will also be useful in the following. Although we use the Fermi-Dirac smearing function for concreteness, our results apply just as well to Gaussian smearing for instance; however, they don't apply to schemes with non-monotonous occupations such as the Methfessel-Paxton scheme[24].

B. The dual energy

Reformulating the ideas in Gonze [22], we now define a "dual" energy

$$\mathcal{I}(H) = \mathcal{E}(f_\mathrm{FD}(H)). \tag{14}$$

Since the map f_FD is a bijection from $\mathcal{H}$ to $\mathcal{P}$, we have

$$\min_{H \in \mathcal{H}} \mathcal{I}(H) = \min_{P \in \mathcal{P}} \mathcal{E}(P). \tag{15}$$

This is analogous to convex duality since the unknown in this formulation is now $H = \nabla \mathcal{E}(P)$.

We can compute the derivative $\boldsymbol{\chi_0} = df_\mathrm{FD}$ of f_FD (see Lemma 1 in the Appendix for details):

$$\begin{aligned}\boldsymbol{\chi_0}\left(\sum_{i=1}^{N_\mathrm{b}} \varepsilon_i |\phi_i\rangle\langle\phi_i|\right) \cdot \delta H \\ = \sum_{i=1}^{N_\mathrm{b}}\sum_{j=1}^{N_\mathrm{b}} \frac{f_\mathrm{FD}(\varepsilon_i) - f_\mathrm{FD}(\varepsilon_j)}{\varepsilon_i - \varepsilon_j} \langle\phi_i, \delta H \phi_j\rangle |\phi_i\rangle\langle\phi_j|\end{aligned} \tag{16}$$

The linear map $\boldsymbol{\chi_0}$ is a "four-point" generalization of the independent-particle polarizability. It describes the change to the density matrix of a system of independent electrons to a change in Fock matrix.

We then have

$$\begin{aligned}\nabla\mathcal{I}(H) &= \boldsymbol{\chi_0}(H)\nabla\mathcal{E}(f_{\mathrm{FD}}(H)) \\ &= \boldsymbol{\chi_0}(H)\Big(H_{\mathrm{KS}}(f_{\mathrm{FD}}(H)) - \mu - \frac{1}{\beta}s'(f_{\mathrm{FD}}(H))\Big) \\ &= \boldsymbol{\chi_0}(H)(H_{\mathrm{KS}}(f_{\mathrm{FD}}(H)) - H)\end{aligned} \tag{17}$$

where again we used $s'(f_{\mathrm{FD}}(\varepsilon)) = \beta(\varepsilon - \mu)$.

The Hessian of $\mathcal{I}$ is a complicated object due to the derivative of $\boldsymbol{\chi_0}(H)$. However, at a solution of $H_{\mathrm{KS}}(f_{\mathrm{FD}}(H_*)) = H_*$, this term vanishes, and we have the simple result

$$\boldsymbol{d^2\mathcal{I}}(H_*) = -\boldsymbol{\chi_0}(H_*)(1 - \boldsymbol{K}(H_*)\boldsymbol{\chi_0}(H_*)) \tag{18}$$

To better understand this object, we compute the Hessian of $\mathcal{E}$. From $\nabla\mathcal{E}(P) = H_{\mathrm{KS}}(P) - \frac{1}{\beta}s'(P) - \mu$ we get

$$\begin{aligned}\boldsymbol{d^2\mathcal{E}}(P)\cdot\delta P &= \boldsymbol{K}(H_{\mathrm{KS}}(P))\cdot\delta P \\ &\quad - \frac{1}{\beta}\sum_{i=1}^{N_{\mathrm{b}}}\sum_{j=1}^{N_{\mathrm{b}}}\frac{s'(p_i) - s'(p_j)}{p_i - p_j}\langle\phi_i, \delta P\phi_j\rangle|\phi_i\rangle\langle\phi_j|.\end{aligned} \tag{19}$$

Defining

$$\begin{aligned}&\boldsymbol{\Omega}\left(\sum_{i=1}^{N_{\mathrm{b}}}\varepsilon_i|\phi_i\rangle\langle\phi_i|\right)\cdot\delta P \\ &\quad = -\sum_{i=1}^{N_{\mathrm{b}}}\sum_{j=1}^{N_{\mathrm{b}}}\frac{\varepsilon_i - \varepsilon_j}{f_{\mathrm{FD}}(\varepsilon_i) - f_{\mathrm{FD}}(\varepsilon_j)}\langle\phi_i, \delta P\phi_j\rangle|\phi_i\rangle\langle\phi_j|\end{aligned} \tag{20}$$

we get

$$\boldsymbol{d^2\mathcal{E}}(P) = \boldsymbol{K}(H_{\mathrm{KS}}(P)) + \boldsymbol{\Omega}(f_{\mathrm{FD}}^{-1}(P)). \tag{21}$$

The point of this formula is to recognize now that $\boldsymbol{\Omega}(H) = -\boldsymbol{\chi_0}(H)^{-1}$. This links the Hessians of $\mathcal{E}$ and $\mathcal{I}$: at a fixed point $H_* = H_{\mathrm{KS}}(f_{\mathrm{FD}}(H_*))$,

$$\boldsymbol{d^2\mathcal{E}}(f_{\mathrm{FD}}(H_*)) = \boldsymbol{\Omega}(H_*)\boldsymbol{d^2\mathcal{I}}(H_*)\boldsymbol{\Omega}(H_*). \tag{22}$$

Since $\boldsymbol{\Omega}$ is self-adjoint and positive definite, both Hessians have the same inertia (number of negative eigenvalues).

C. Hamiltonian mixing

The very simplest Hamiltonian mixing algorithm is

$$H_{n+1} = H_{\mathrm{KS}}(f_{\mathrm{FD}}(H_n)). \tag{23}$$

As already recognized in Reference 22, (17) makes it possible to reinterpret this simple algorithm in a new light: it is a gradient descent algorithm on $\mathcal{I}$ with step 1, preconditioned by $\boldsymbol{\chi_0}(H_n)^{-1}$. It is natural to use a smaller stepsize to try to ensure convergence, and indeed this is guaranteed to work:

Theorem 1. *Let $H_0 \in \mathcal{H}$. There is $\alpha_0 > 0$ such that, for all $0 < \alpha < \alpha_0$, the algorithm*

$$H_{n+1} = H_n + \alpha(H_{\mathrm{KS}}(f_{\mathrm{FD}}(H_n)) - H_n) \tag{24}$$

satisfies $H_{\mathrm{KS}}(f_{\mathrm{FD}}(H_n)) - H_n \to 0$. If furthermore E is analytic, H_n converges to a solution of the equation $H_{\mathrm{KS}}(f_{\mathrm{FD}}(H)) = H$.

Adaptive-step schemes can also ensure guaranteed convergence:

Theorem 2. *Fix $H_0 \in \mathcal{H}$, and constants $0 < \alpha_{\max} < 1$, $0 < c < 1, 0 < \tau < 1$. Consider the algorithm*

$$H_{n+1} = H_n + \alpha_n(H_{\mathrm{KS}}(f_{\mathrm{FD}}(H_n)) - H_n) \tag{25}$$

where α_n is chosen in the following way: starting from $\alpha_{\max}$, decrease $\alpha_{\max}$ by a factor τ while the Armijo line search condition

$$\begin{aligned}&\mathcal{I}(H_n + \alpha_n(H_{\mathrm{KS}}(f_{\mathrm{FD}}(H_n)) - H_n)) \\ &\quad \leq \mathcal{I}(H_n) - \alpha_n c\langle\boldsymbol{\Omega}(f_{\mathrm{FD}}(H_n))\nabla\mathcal{I}(H_n), \nabla\mathcal{I}(H_n)\rangle\end{aligned} \tag{26}$$

is not verified. Then this algorithm satisfies $H_{\mathrm{KS}}(f_{\mathrm{FD}}(H_n)) - H_n \to 0$. If furthermore E is analytic, H_n converges to a solution of the equation $H_{\mathrm{KS}}(f_{\mathrm{FD}}(H)) = H$.

The proofs of both these statements are found in the Appendix.

The adaptive-step scheme above however suffers from two important drawbacks. First, it is costly (requiring several SCF steps per iteration). Second, it is imcompatible with preconditioned or accelerated schemes because, in contrast to the SCF direction, there is no guarantee in these cases that the chosen direction is a descent direction to the energy. This would make a straightforward implementation of the above algorithm uncompetitive for "easy" systems, and therefore motivates the search for a compromise algorithm that tries to recover some robustness properties while not sacrificing performance.

D. Potential mixing

We now specialize the above discussion to our case of interest of semi-local density-functional theory (DFT) models. We introduce the operators $\mathrm{diag} : \mathbb{R}^{N_{\mathrm{b}}\times N_{\mathrm{b}}} \to \mathbb{R}^{N_{\mathrm{b}}}$ and $\mathrm{diagm} : \mathbb{R}^{N_{\mathrm{b}}} \to \mathbb{R}^{N_{\mathrm{b}}\times N_{\mathrm{b}}}$. The diag operator takes the diagonal (in real space) of a density matrix, yielding a density. The diagm operator constructs a Fock matrix contribution with a given local potential. Both operators are adjoint of each other. With these notations, the energy function takes the form

$$\mathcal{E}_0(P) = \mathrm{Tr}(H_0P) + g\left(\mathrm{diag}(P)\right), \tag{27}$$

where H_0 is a given operator (the core Hamiltonian) and g is a nonlinear function (the Hartree-exchange-correlation energy). For these models the gradient of

$\mathcal{E}_0(P)$ (the Fock matrix) depends on $\mathrm{diag}(P)$ (the density) only:

$$H_{\mathrm{KS}}(P) = H_0 + \mathrm{diagm}(V(\mathrm{diag}(P))), \tag{28}$$

with the potential

$$V(\rho) = \nabla g(\rho) \in \mathbb{R}^{N_\mathrm{b}}. \tag{29}$$

Based on (13) and the definition of the density we define the potential-to-density mapping

$$\rho(V) = \mathrm{diag}(f_{\mathrm{FD}}(H_0 + \mathrm{diagm}(V))), \tag{30}$$

which allows to solve the self-consistency problem (12) by iteration in the potential V only:

$$V_{n+1} = V_n + \alpha \delta V_n, \tag{31}$$

where we defined the search direction

$$\delta V_n = V(\rho(V_n)) - V_n. \tag{32}$$

The corresponding energy functional minimized by this fixed-point problem is

$$\mathcal{I}(V) = \mathcal{E}(f_{\mathrm{FD}}(H_0 + \mathrm{diagm}(V))). \tag{33}$$

Compared to an algorithm based on Kohn-Sham Hamiltonians as suggested in Section II C this formulation has the advantage that only vector-sized potentials V_n instead of matrix-sized quantities need to be handled.

The analysis of the previous sections carries forward straightforwardly to the potential mixing setting. In particular one identifies as the analogue of $\boldsymbol{K}$ the Hessian of g, i.e. the (two-point) Hartree-exchange-correlation kernel K, and as the analogue of $\boldsymbol{\chi_0}$ the derivative of $V(\rho)$, which is the independent-particle susceptibility χ_0. The latter becomes apparent by comparing (16) to the Adler-Wiser formula for χ_0[25,26]

$$\chi_0(V) = \sum_{i=1}^{N_b} \sum_{j=1}^{N_b} \frac{f_{\mathrm{FD}}(\varepsilon_i) - f_{\mathrm{FD}}(\varepsilon_j)}{\varepsilon_i - \varepsilon_j} |\phi_i^* \phi_j\rangle\langle\phi_i^* \phi_j| \tag{34}$$

in which (ε_i, ϕ_i) denotes the eigenpairs of $H_0 + \mathrm{diagm}(V)$. Both K and χ_0 arise naturally when considering the Jacobian matrix

$$J_\alpha = 1 - \alpha\big(1 - K(V_*)\chi_0(V_*)\big) \tag{35}$$

of the potential-mixing SCF iteration (31) near a fixed point V_*. If the eigenvalues of J_α are between -1 and 1 the potential-mixing SCF iterations converge. By analogy with Hamiltonian mixing, Theorem 1 guarantees that global convergence can always be ensured by selecting α small enough. In this respect our results from Section II C strengthen a number of previous results[17,22,27], which established local convergence for sufficiently small α.

E. Improving the search direction δV_n: Preconditioning and acceleration

The Jacobian matrix (35) involves the dielectric matrix $\epsilon(V) = 1 - K(V)\chi_0(V)$, which can become badly conditioned for many systems. In such cases, a very small step must be employed to ensure stability (smallest eigenvalue of J_α larger than -1), which slows down convergence (largest eigenvalue of J_α very close to 1) to a level too slow to be practical. A solution is to improve the search direction δV_n to ensure faster convergence[1]. This is usually achieved by a combination of techniques jointly referred to as "mixing", which amend δV_n using both preconditioning as well as convergence acceleration.

Employing a preconditioned search direction

$$\delta V_n = P^{-1}[V(\rho(V_n)) - V_n] \tag{36}$$

in a damped SCF iteration, the corresponding Jacobian becomes

$$J_\alpha = 1 - \alpha P^{-1}\epsilon(V). \tag{37}$$

Provided that the inverse P^{-1} approximates the inverse dielectric matrix ϵ^{-1} sufficiently well, the spectrum of $P^{-1}\epsilon$ is close to 1, so that a larger damping α and and faster iteration is possible. While suitable cheap preconditioners P are not yet known for all sources of bad conditioning in SCF iterations, a number of successful strategies have been suggested. Examples include Kerker mixing[28] to improve SCF convergence in metals or LDOS-based mixing[11] to tackle heterogeneous metal-vacuum or metal-insulator systems. For a more detailed discussion on this matter we refer the reader to Reference 11.

An additional possibility to speed up convergence is to use black-box convergence accelerators. These techniques build up a history of the previous iterates $V_1, \ldots, V_n$ as well as the previous preconditioned residuals $P^{-1}R_1, \ldots, P^{-1}R_n$ (with $R_n = V(\rho(V_n)) - V_n$) and use this information to obtain the next search direction δV_n. The most frequently used acceleration technique in this context is variously known as Pulay/DIIS/Anderson mixing/acceleration, which we will refer to as Anderson acceleration. This method obtains the search direction as a linear combination

$$\begin{aligned}\delta V_n = {} & P^{-1}R_n \\ & + \frac{1}{\alpha}\sum_{i=1}^{n-1} \beta_i \big(V_i + \alpha P^{-1}R_i - V_n - \alpha P^{-1}R_n\big)\end{aligned} \tag{38}$$

where the expansion coefficients β_i are found by minimizing

$$\left\| P^{-1}R_n + \sum_{i=1}^{n-1} \beta_i \left(P^{-1}R_i - P^{-1}R_n\right) \right\|. \tag{39}$$

In practice, it is impractical to keep a potentially large number of past iterates, and only the last 10 iterates are

taken into account. Furthermore, the associated linear least squares problem can become ill-conditioned[29]. We use the simple strategy of discarding past iterates to ensure a maximal conditioning of 10^6.

This method is known to be equivalent to a multisecant Broyden method. In the linear regime and with infinite history Anderson acceleration is further equivalent to the well-known GMRES method to solve linear equations. For details see Reference[30] and References therein. Provided that nonlinear effects are negligible, Anderson acceleration typically inherits the favorable convergence properties of Krylov methods[31], explaining their frequent use in the DFT context. However, especially at the beginning of the SCF iterations or when treating systems that feature many close SCF minima, nonlinear effects can become important. In such cases the behavior of Anderson is more complex and mathematically not yet fully understood. In particular the dependence of the convergence behavior on numerical parameters such as the chosen damping can become less regular and harder to interpret, as we will see in our numerical examples in Sections IV B and IV C.

III. ADAPTIVE DAMPING ALGORITHM

Up to now we have assumed that the step size α is constant, reflecting common practice in plane-wave DFT computations. We now describe the main contribution of this paper, an algorithm to adapt this step size to increase robustness and minimize user intervention into the convergence process. At step n of the algorithm, given a trial potential V_n, we compute the search direction δV_n through (38), and look for a step α_n to take as

$$V_{n+1} = V_n + \alpha_n \delta V_n \tag{40}$$

Note that the definition of δV_n itself in (38) depends on a stepsize α; since our scheme will adapt α_n to δV_n, we cannot just take $\alpha = \alpha_n$ in (38), and so we use for α a trial damping $\widetilde{\alpha}$ (to be discussed in Section III C).

To select α_n, we could try to minimize $\mathcal{I}(V_{n+1})$, or employ an Armijo line-search strategy. However, each evaluation of $\mathcal{I}$ is very costly, and it is therefore desirable to obtain efficient approximate schemes. The energy $\mathcal{I}(V + \alpha\delta V_n)$, can be expanded as

$$\begin{aligned}\mathcal{I}(V_n + \alpha\delta V_n) = \mathcal{I}(V_n) + \alpha\langle\chi_0(V_n)R_n, \delta V_n\rangle \\ + \frac{1}{2}\alpha^2\langle\delta V_n, d^2\mathcal{I}(V_n)\cdot\delta V_n\rangle + O(\alpha^3\|\delta V_n\|^3),\end{aligned} \tag{41}$$

where we have used $\nabla\mathcal{I}(V_n) = \chi_0(V_n)R_n$. This approximation is good for small dampings α and/or close to the solution, when δV_n is small. The object $d^2\mathcal{I}$ is complicated and expensive to compute in general. However, close to a fixed point, we can use the expression (18) to write $d^2\mathcal{I}(V_n) \approx \chi_0(V_n)(1 - K(V_n))\chi_0(V_n)$. We can then approximate the terms in (41), leading to the model

$$\begin{aligned}\varphi_n(\alpha) = \mathcal{I}(V_n) + \alpha\,\langle R_n, \chi_0(V_n)\delta V_n\rangle \\ - \frac{1}{2}\alpha^2\,\langle\chi_0(V_n)\delta V_n, \big[1 - K(V_n)\chi_0(V_n)\big]\delta V_n\rangle\end{aligned} \tag{42}$$

for the energy, where we have used the self-adjointness of $\chi_0(V_n)$ to make it act only on δV_n. To compute the coefficients in this model, we still need to compute $\chi_0(V_n)\delta V_n$, a costly operation. However, for all α we have to first order

$$\alpha\chi_0(V_n)\delta V_n = \rho(V_n + \alpha\delta V_n) - \rho(V_n) + O(\alpha^2\|\delta V_n\|^2). \tag{43}$$

Note that if we set $V_{n+1} = V_n + \alpha_n\delta V_n$ and then proceed along the iterative algorithm, we will have to compute $\rho(V_{n+1})$ in any case. An approximation to the coefficients of the model φ_n can therefore be constructed without any extra diagonalization.

This is the basis of the adaptive damping scheme described in Algorithm 1. Since $\rho(V_n)$ is already known (it is needed to construct δV_n), the only expensive step in this algorithm is the computation of $\rho(V_{n+1})$, which occurs only once per loop iteration. In particular, when set to always accept V_{n+1}, this algorithm reduces to the standard damped SCF algorithm. Notice that the algorithm only allows α_n to shrink between iterations. As a result (i) the model φ_n provides better and better damping predictions and (ii) keeping in mind our analysis of Section II C the proposed tentative steps V_{n+1} become more likely to be accepted.

Algorithm 1 Adaptive damping algorithm

Input: Current iterate V_n, search direction δV_n, trial damping $\widetilde{\alpha}$
Output: Damping α_n, next iterate V_{n+1}

$\alpha_n \leftarrow \widetilde{\alpha}$
loop
  Make tentative step $V_{n+1} = V_n + \alpha_n\delta V_n$
  Compute $\rho(V_{n+1}), \mathcal{I}(V_{n+1})$ (the expensive step)
  if accept V_{n+1} (see Section III A) **then**
    break
  else
    Build the coefficients of the model φ_n
    if model φ_n is good (see Section III B) **then**
      $\alpha_n \leftarrow \mathrm{argmin}_\alpha\,\varphi_n(\alpha)$
      Scale α_n to ensure $|\alpha_n|$ is strictly decreasing
    else
      $\alpha_n \leftarrow \frac{\alpha_n}{2}$
    end if
  end if
end loop

We complete the description of the algorithm by specifying some practical points: when to accept a step, how to determine whether a model is good, how to select the initial trial step $\widetilde{\alpha}$ and how to integrate adaptive damping with Anderson acceleration.

A. Step acceptance

We accept the step as soon as the proposed next iterate $V_{n+1} = V_n + \alpha_n \delta V_n$ satisfies

$$\mathcal{I}(V_{n+1}) < \mathcal{I}(V_n) \quad \text{or} \quad \left\|P^{-1}R_{n+1}\right\| < \left\|P^{-1}R_n\right\|, \tag{44}$$

i.e. if either the energy or the preconditioned residual decreases. Although accepting steps higher in energy may decrease the robustness of the algorithm, we found in practice that accepting steps that decrease the residual helps keeping the method effective in the later stages of convergence, when the Anderson acceleration is able to take efficient steps that may slightly increase the energy but are not worth reverting.

B. Quality of the model φ_n

Our model φ_n makes various assumptions that might not hold in practice, especially far from convergence. However, by comparing the actual energy $\mathcal{I}(V_n + \alpha_n \delta V_n)$ to the prediction $\varphi_n(\alpha_n)$, we can inexpensively check the quality of the model. We do this by computing the ratio

$$r_n = \frac{|\mathcal{I}(V_n + \alpha_n \delta V_n) - \varphi_n(\alpha_n)|}{|\mathcal{I}(V_n + \alpha_n \delta V_n) - \mathcal{I}(V_n)|}, \tag{45}$$

which should be small if the model is accurate. We deem the model good enough if

$$r_n < 0.1 \qquad \text{and} \qquad \varphi_n \text{ has a minimum.} \tag{46}$$

Notice that the minimizer of φ_n may not necessarily be positive. For particularly accurate models ($r_n < 0.01$) we additionally allow backward steps (i.e. $\alpha_n < 0$), which turned out to overall improve convergence in our tests.

C. Choice of the trial step $\widetilde{\alpha}$

To ensure that as many as possible SCF steps only require a single line search step, we dynamically adjust $\widetilde{\alpha}$ between two subsequent SCF steps. A natural approach is to reuse the adaptively determined damping α_n as the $\widetilde{\alpha}$ in the next line search, which effectively shrinks $\widetilde{\alpha}$ between SCF steps. However, the algorithm may need small values of $\widetilde{\alpha}$ in the initial stages of convergence, and keeping these small values for too long limits the eventual convergence rate. To counteract the decreasing trend, we allow $\widetilde{\alpha}$ to increase if a line search was immediately successful (i.e. $\alpha_n = \widetilde{\alpha}$). In this case we again use the model φ_n. If it is sufficiently good (as described in Section III B), we set

$$\widetilde{\alpha} \leftarrow \max\left(\widetilde{\alpha},\ 1.1 \cdot \operatorname*{argmin}_{\alpha} \varphi_n(\alpha)\right). \tag{47}$$

Otherwise $\widetilde{\alpha}$ is left unchanged.

As an additional measure, to prevent the SCF from stagnating we enforce $\widetilde{\alpha}$ to not undershoot a minimal trial damping $\widetilde{\alpha}_{\min}$. We used mostly $\widetilde{\alpha}_{\min} = 0.2$ as a baseline, and report varying this parameter in the numerical experiments.

With this dynamic adjustment of $\widetilde{\alpha}$, we checked that its initial value $\widetilde{\alpha}_0$, i.e. the value used in the first SCF step, has little influence on the overall convergence behavior. However, in well-behaved cases, too small values for this parameter lead to an unnecessary slowdown of the first few SCF steps. We therefore settled on $\widetilde{\alpha}_0 = 0.8$ similar to standard recommendations for the default damping[32].

IV. NUMERICAL TESTS

The adaptive damping algorithm described in Section III was compared against a conventional preconditioned damped potential-mixing SCF scheme featuring only a fixed damping. For this we employed three kinds of test problems. The first are calculations on aluminium systems of various size including cases with an unsuitable computational setup, i.e. where charge sloshing is not prevented by employing the Kerker preconditioner. These are discussed in more detail in Section IV A. The second, discussed in Section IV B, is a gallium arsenide system which we previously found to feature strongly nonlinear behavior in the initial SCF steps[11]. Lastly in section Section IV C we will consider Heusler systems and other transition-metal compounds, which are generally found to be difficult to converge.

For our tests we used the implementation of the adaptive damping algorithm available in the density-functional toolkit (DFTK)[33,34], a flexible open-source Julia package for plane-wave density-functional theory simulations. For all calculations we used Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional[35] as implemented in the libxc[36] library, and Goedecker-Teter-Hutter pseudopotentials[37]. Depending on the system, a kinetic energy cutoff between 20 and 45 Hartree as well as an unshifted Monkhorst-Pack with a maximal k-point spacing of at most 0.14 inverse Bohrs was used. For the Heusler systems this was reduced to at most 0.08 inverse Bohrs. With the exception of the gallium arsenide system a Gaussian smearing scheme with width of 0.001 Hartree was employed. For the systems containing transition-metal elements collinear spin polarization was allowed and the initial guess was constructed assuming ferromagnetic spin ordering except when otherwise noted. Notice, that this initial guess is generally not close to the final spin ordering, see Section IV C for discussion regarding this choice. The full computational details for each system (including the employed structures) as well as instructions how to reproduce all results of this paper can be found in our repository of supporting information[38].

Table I summarizes the required number of Hamiltonian diagonalizations to converge the SCF energy to an error of 10^{-10} Hartree for various fixed dampings α as

System	Precond.	fixed damping α										adaptive
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	damping
Al_8 supercell	Kerker[a]	×	58	37	27	21	16	13	**11**	12	18	17
Al_8 supercell	None[a]	×	**52**	×	×	×	×	×	×	×	×	24
Al_{40} supercell	Kerker	19	15	14	12	**11**	12	12	12	12	12	12
Al_{40} supercell	None	**38**	40	40	39	44	50	49	×	76	×	44
Al_{40} surface	Kerker	×	×	×	×	×	×	×	×	×	×	×
Al_{40} surface	None	**46**	48	50	49	51	60	61	66	89	×	49
$Ga_{20}As_{20}$ supercell	None	**26**	33	40	42	45	44	70	70	65	76	26
CoFeMnGa	Kerker	×	×	×	×	28	**21**	24	28	22	22	30
Fe_2CrGa	Kerker	×	×	×	27	×	×	**19**	25	×	22	39
Fe_2MnAl	Kerker	×	48	×	×	×	20	21	17	16	**15**	34
$FeNiF_6$	Kerker	×	×	×	×	×	×	×	23	22	**21**	24
Mn_2RuGa	Kerker	×	×	×	×	37	24	23	**22**	23	23	36
Mn_3Si	Kerker	×	×	×	×	26	30	22	**20**	×	×	×
Mn_3Si (AFM)[b]	Kerker	×	×	58	29	31	30	**20**	22	26	28	35
Cr_{19} defect	Kerker	×	×	×	74	46	48	46	**41**	47	53	48
$Fe_{28}W_8$ multilayer	Kerker	**32**	34	37	34	38	43	41	48	×	×	37

[a] without Anderson acceleration
[b] initial guess with antiferromagnetic spin ordering

TABLE I. Number of Hamiltonian diagonalizations required to obtain convergence in the energy to 10^{-10} with a cross (×) denoting a failure to converge within 100 diagonalizations. Except where otherwise noted Anderson acceleration has been employed and for the transition-metal systems (third/fourth group of compounds) a ferromagnetic initial guess has been used. On supercells atomic positions were slightly randomised. Computational details are given in the text. Notice that the transition-metal systems may not converge to the same SCF solution for each calculation.

well as the adaptive damping algorithm. We carefully verified the obtained solutions to be stationary points by monitoring the SCF residual R_n. Note that for the adaptive damping procedure the number of SCF steps is not identical to the number of Hamiltonian diagonalizations, since multiple tentative steps might be required until a step is accepted. Since iterative diagonalization overall dominates the cost of the SCF procedure, the number of diagonalizations provides a better metric to compare between the cost of both damping strategies.

A. Inadequate preconditioning: Aluminum

To investigate the influence of the choice of a suboptimal preconditioner on the convergence for both the fixed damping and adaptive damping strategies we considered three aluminium test systems: two elongated bulk supercells with 8 or 40 atoms as well as a surface with 40 atoms and a portion of vacuum of identical size. For the elongated supercells both the initial guess as well as the atomic positions were slightly rattled.

The results are summarized in the first segment of Table I. For the small Al_8 system, where Anderson acceleration was not used, representative convergence curves are shown in Figure 1. Due to the well-known charge-sloshing behavior, SCF iterations on such metallic systems are ill-conditioned. Without preconditioning small fixed damping values α are thus required to obtain convergence, with only a small window of damping values being able to achieve a convergence within 100 Hamiltonian applications. On the other hand in combination with the matching Kerker preconditioning strategy[28] large fixed damping values generally converge more quickly.

In contrast the adaptive damping strategy is much less sensitive to the choice of the minimal trial damping $\widetilde{\alpha}_{\min}$. Moreover, it leads to a much improved convergence for the case without suitable preconditioning while still maintaining similar costs if Kerker mixing is employed.

These observations carry over to cases including Anderson acceleration and larger aluminium systems, see Figure 2 for a representative computation on an aluminium surface. Notice that Kerker mixing is extremely badly suited for the large aluminium surface, such that convergence is not obtained in 100 Hamiltonian for any of the damping strategies, see Reference 11 for a better preconditioner in such inhomogeneous systems. Overall employing adaptive damping therefore makes the reliability and efficiency of the SCF less dependent on the choice of the preconditioning strategy, while not requiring the user to manually select a damping parameter.

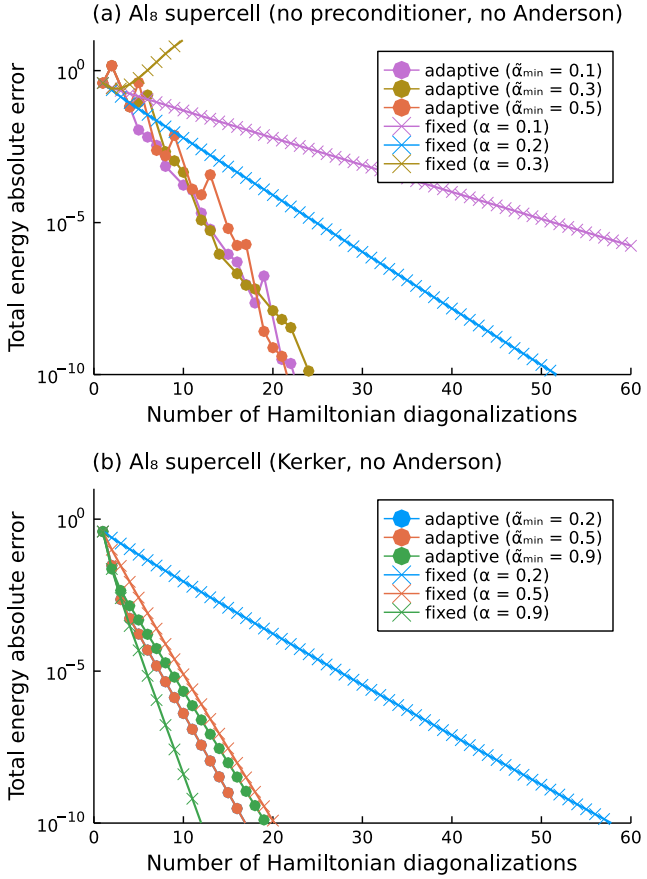

FIG. 1. SCF convergence for a randomized aluminium supercell (8 atoms) without Anderson acceleration and using simple mixing (top) as well as Kerker mixing(bottom). The adaptive scheme converges robustly even in the unpreconditioned case, without requiring the manual selection of a step.

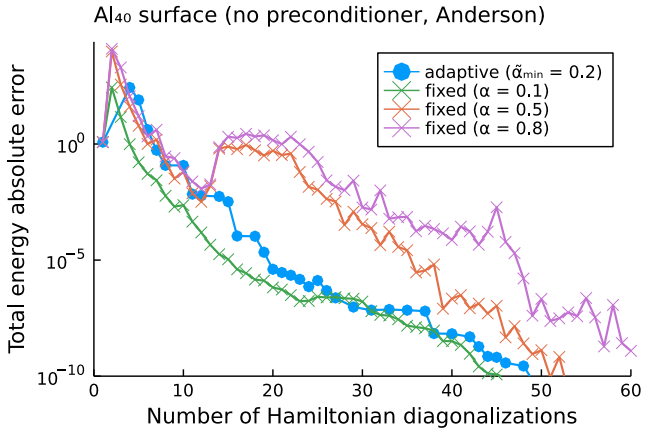

FIG. 2. SCF convergence without preconditioning for an elongated supercell of an aluminium surface with 40 aluminium atom and a vacuum portion of equal size. A fixed step of $\alpha = 0.1$ was optimal here, but the adaptive scheme gets very close performance without manual stepsize selection.

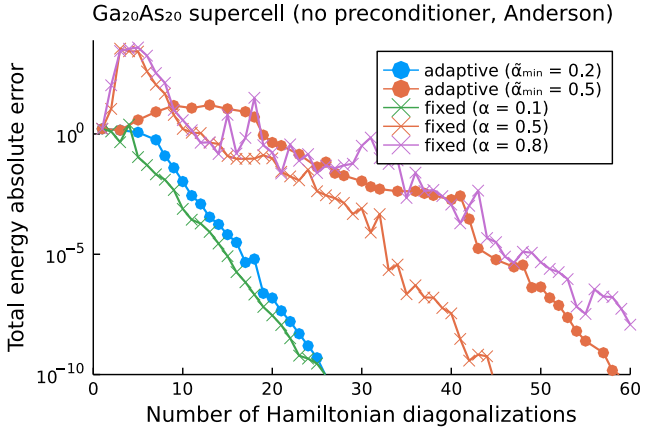

FIG. 3. Elongated gallium arsenide supercell (20 gallium and 20 arsenide atoms) with slightly randomized atomic positions. For all cases, simple mixing and Anderson acceleration are employed.

B. Strong nonlinear effects: Gallium arsenide

In previous work we identified elongated supercells of gallium arsenide with slightly perturbed atomic positions to be a simple system that still exhibits strong nonlinear effects when the SCF is far from convergence[11]. In the convergence profiles of these systems this manifests by the error shooting up abruptly with Anderson failing to quickly recover. In Figure 3, for example, the error increases steeply between Hamiltonian diagonalizations 3 and 6 for the fixed-damping approaches with stepsizes beyond 0.1. It should be noted that this behavior is an artefact of the interplay of Anderson acceleration and damped SCF iterations on these systems, which is not observed in case Anderson acceleration is not employed. For more details see the discussion of the gallium arsenide case in Ref. 11.

For the calculations employing a fixed damping strategy only small damping values of $\alpha = 0.1$ are able to prevent this behavior. Already slightly larger damping values noticably increase the number of Hamiltonian diagonalizations required to reach convergence (compare Table I), and thus a careful selection of the damping value is in order for such systems. In contrast the proposed adaptive damping strategy with our baseline minimal trial damping of $\widetilde{\alpha}_{\min} = 0.2$ automatically detects the unsuitable Anderson steps and downscales them. As a result an optimal or near-optimal cost is obtained without any manual parameter tuning. For comparison, we also display in Figure 3 the results with a large value of $\widetilde{\alpha}_{\min} = 0.5$, which prevents the damping algorithm to avoid the nonlinear effects.

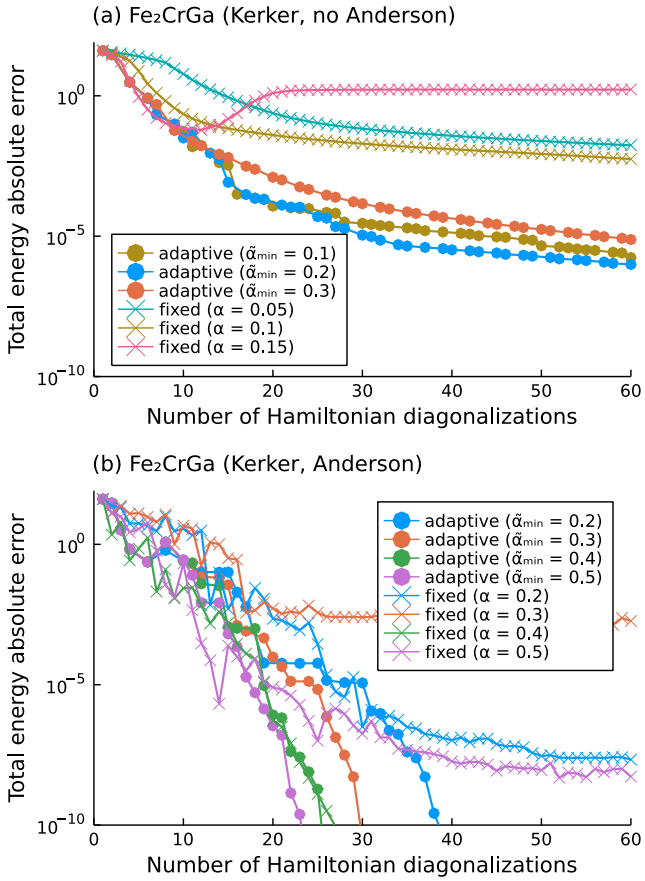

FIG. 4. Convergence of the Fe_2CrGa Heusler alloy with Kerker mixing without Anderson acceleration (top) and with Anderson acceleration (bottom). Notice that the SCF calculations do not necessarily converge to the same local SCF minimum.

C. Challenging transition-metal compounds

In this section we discuss two types of transition-metal systems. First, we consider a selection of smaller primitive unit cells, including the mixed iron-nickel fluoride $FeNiF_6$ as well a number of Heusler-type alloy structures, see the third group of Table I. These structures were found found in the course of high-throughput computations to be difficult to converge[39]. Moving to larger systems we considered an elongated chromium supercell with a single vacancy defect as well as a layered iron-tungsten system, see the fourth group of Table I. Both test cases were taken from previous studies[20,40] on SCF algorithms.

In particular the Heusler compounds are known to exhibit rich and unusual magnetic and electronic properties. From our test set, for example, Fe_2MnAl shows halfmetallic behaviour, i.e. a vanishing density of states at the Fermi level in only the minority spin channel[41]. Other compounds, such as Mn_2RuGa or CoFeMnGa show an involved pattern of ferromagnetic and antiferromagnetic coupling of the neighboring transition-metal sites[42,43]. Such effects are closely linked to the d-orbitals forming localized states near the Fermi level[44,45] and imply that there are a multiple accessible spin configurations, which are close in energy. Unfortunately these two properties also make Heusler compounds difficult to converge using standard methods. First, localized states near the Fermi level are a source of ill-conditioning for the SCF fixed-point problem[11], with no cheap and widely applicable preconditioning strategy being available. Second, the abundance of multiple spin configurations implies a more involved SCF energy landscape with multiple SCF minima. On such a landscape convergence may easily "hesitate" between different local minima or stationary points. Furthermore the setup of an appropriate initial guess, which ideally guides the SCF towards the final spin ordering requires human expertise and is hard to automatise in the high-throughput setting. Albeit not fully appropriate for the systems we consider, we followed the guess setup, which has also been used in the aforementioned high-throughput procedure[39], namely to start the calculations with an initial guess based on ferromagnetic (FM) spin ordering.

As a result in our tests, calculations on Heusler systems without Anderson acceleration require very small fixed damping values below 0.1 even if the Kerker preconditioner is used, see Figure 4 (a). The adaptive damping strategy improves the convergence behavior and in agreement with our previous results partially corrects for the mismatch in preconditioner and initial guess. Still, convergence is extremely slow.

An acceptable convergence is only accessible in combination with Anderson acceleration. However, the Anderson-accelerated fixed-damping SCF is very susceptible to the chosen damping α, see Figure 4 (b). In particular the lowest-energy SCF minimum is only found within 100 Hamiltonian diagonalizations for $\alpha = 0.4$, $\alpha = 0.7$, $\alpha = 0.8$ and $\alpha = 1.0$. Other fixed damping values initially converge, but then convergence stagnates and the error only reduces very slowly beyond around 30 diagonalizations.

We investigated the source of this pathological behavior by restarting the iterations after stagnation. This did not noticeably alter the behavior, eliminating the possibility that the history of the iterates within the Anderson acceleration scheme somehow "jam" the SCF into stagnation in the strongly nonlinear regime — as can happen for instance in nonlinear conjugate gradient methods[46]. Another possibility is that the iterations somehow got into a particularly rough region of SCF energy landscape between multiple stationary points, which is simply hard to escape. However, this is not the case either. For instance on the Fe_2CrGa system with a fixed damping of 0.3, the restarted iterations did converge quickly using an Anderson scheme with a small maximal conditioning of 10^2 for the linear least squares problem. It would therefore appear that this phenomenon is due to inadequate regularization of the least squares problem. We expect more sophisticated techniques for controlling the Ander-

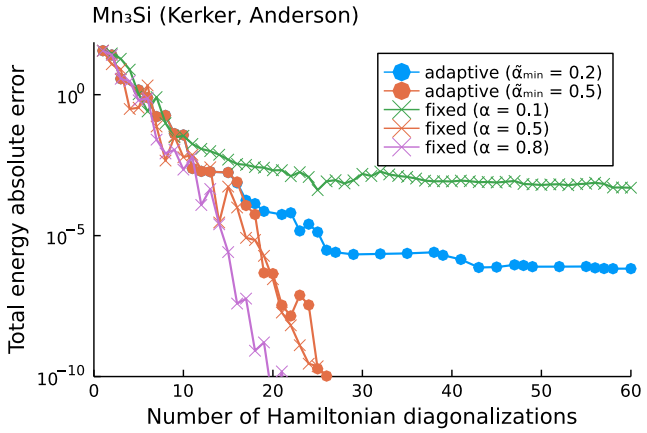

FIG. 5. Convergence of the Mn_3Si Heusler compound with Kerker mixing and Anderson acceleration and starting from a ferromagnetically ordered initial guess. Small dampings are susceptible to stagnation induced by Anderson instabilities.

son history[30] to be worth investigating for such systems in the future.

Because of this stagnation issue we found Anderson-accelerated SCF iterations to become unreliable for our transition-metal test systems: for fixed damping values below $\alpha = 0.5$, hardly any calculation converges. Notably, due to the non-trivial interplay with the Anderson scheme, this result is the exact opposite to our theoretical developments on damped SCF iterations in Section II C, which suggested to reduce the damping to achieve reliable convergence.

Overall the transition-metal cases emphasize the difficulty in manually choosing an appropriate fixed damping. For a number of cases the window of converging damping values is rather narrow, e.g. consider Fe_2CrGa, $FeNiF_6$ or Mn_3Si (with the FM guess) and in our tests only a single damping value of $\alpha = 0.8$ fortitiously manages to converge all systems.

In contrast the proposed adaptive damping strategy with our baseline value of $\widetilde{\alpha}_{\mathrm{min}} = 0.2$ is less susceptible to the stagnation issue. Across the unit cells and extended transition-metal systems we considered we observed only a convergence failure in one test case, namely the Mn_3Si Heusler alloy with the FM guess, see Figure 5. For some test cases, adaptive damping did cause a noticeable computational overhead: in extreme cases (such as Fe_2CrGa or Fe_2MnAl) the number of required Hamiltonian diagonalizations almost doubles. However, it should be emphasized that no user adjustments were needed to obtain these results, even though adaptive damping has been constructed on the here invalid principle that smaller damping increases reliability.

Yet even on cases where Anderson instabilities cause non-convergence of the adaptive scheme, fine-tuning is possible. Increasing the minimal trial damping from $\widetilde{\alpha}_{\mathrm{min}} = 0.2$ to $\widetilde{\alpha}_{\mathrm{min}} = 0.5$, for example, increases the minimal step size and thus lowers the risk of Anderson stagnation. For all transition-metal cases we considered $\widetilde{\alpha}_{\mathrm{min}} = 0.5$ strictly reduces the number of diagonalizations required to reach convergence compared to $\widetilde{\alpha}_{\mathrm{min}} = 0.2$, see for example Figure 4 (b). Moreover this parameter adjustment even resolves the convergence issues of Mn_3Si, see Figure 5. If manual intervention is possible another option is to incorporating prior knowledge of the final ground state electronic structure into the initial guess. For the Mn_3Si case, for example, an improved initial guess based on an antiferromagnetic spin ordering (AFM) between adjacent manganese layers simplifies the SCF problem, such that both a larger range of fixed damping values as well as the adaptive damping strategy give rise to converging calculations.

V. CONCLUSION

We proposed a new linesearch strategy for SCF computations, based on an efficient approximate model for the energy as a function of the damping. Our algorithm follows four general principles: (a) the algorithm should need no manual intervention from the user; (b) it should be combinable with known effective mixing techniques such as preconditioning and Anderson acceleration; (c) in "easy" cases where convergence with a fixed damping is satisfactory it should not slow down too much; and (d) it should be possible to relate it to schemes with proved convergence guarantees. We demonstrated that our proposed scheme fulfills all these objectives. With our default parameter choice of $\widetilde{\alpha}_{\mathrm{min}} = 0.2$ the resulting adaptively damped SCF algorithm is able to converge all of the "easy" cases faster or almost as fast as the fixed-damping method with the best damping. Simultaneously it is more robust than the fixed-damping method on the "hard" cases we considered, such as elongated bulk metals and metal surfaces without proper preconditioning or Heusler-type transition-metal alloys. In particular the latter kind of systems feature a very irregular convergence behavior with respect to the damping parameter, making a robust manual damping selection very challenging. In practice the classification between "easy" and "hard" cases may well depend on the considered system and the details of the computational setup, e.g. the employed mixing and acceleration techniques. However, our scheme makes no assumptions about the details how a proposed SCF step has been obtained. We therefore believe adaptive damping to be a black-box stabilisation technique for SCF iterations, which applies beyond the Anderson-accelerated setting we have considered here.

Still, our results on these "hard" cases also highlight poorly-understood limitations of the commonly used Anderson acceleration process. For example, despite following standard recommendations to increase Anderson robustness, we frequently observe SCF iterations to stagnate. A more thorough understanding of this effect would be an interesting direction for future research.

Our scheme was applied to semilocal density function-

als in a plane-wave basis set. It is not specific to plane-wave basis sets, and we expect it to be similarly efficient in other "large" basis sets frequently used in condensed-matter physics. For atom-centered basis sets, like those common in quantum chemistry, direct mixing of the density matrix is feasible, and likely more efficient. Our scheme does not apply directly to hybrid functionals, where orbitals or Fock matrices have to be mixed also; an extension to this case would be an interesting direction for future research.

ACKNOWLEDGEMENTS

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 810367). We are grateful to Marnik Bercx and Nicola Marzari for pointing us to the challenging transition-metal structures that stimulated most of the presented developments. Fruitful discussions with Eric Cancès, Xavier Gonze and Benjamin Stamm and provided computational time at Ecole des Points and RWTH Aachen University are gratefully acknowledged.

APPENDIX: MATHEMATICAL PROOFS

Lemma 1. *Let*

$$H_* = \sum_{i=1}^{N_\mathrm{b}} \varepsilon_i |\phi_i\rangle\langle\phi_i| \tag{48}$$

with orthonormal ϕ_i and non-decreasing ε_i. Let $f : \mathbb{R} \to \mathbb{R}$ be a real analytic function in a neighborhood of $[\varepsilon_1, \varepsilon_{N_\mathrm{b}}]$. Then the map $H \mapsto f(H)$ is analytic in a neighborhood of H_*, and

$$\mathbf{df}(H_*)\cdot\delta H = \sum_{i=1}^{N_\mathrm{b}}\sum_{j=1}^{N_\mathrm{b}} \frac{f(\varepsilon_i)-f(\varepsilon_j)}{\varepsilon_i-\varepsilon_j}\langle\phi_i,\delta H\phi_j\rangle|\phi_i\rangle\langle\phi_j| \tag{49}$$

with the convention that $\frac{f(\varepsilon_i)-f(\varepsilon_i)}{\varepsilon_i-\varepsilon_i} = f'(\varepsilon_i)$.

Proof of Lemma 1. This is a classical result, known as the Daleckii-Krein theorem in linear algebra; see for instance Higham[47]. To keep this paper self-contained, we follow here the proof in Levitt[48] in the analytic case. Since f is analytic on $[\varepsilon_1, \varepsilon_{N_\mathrm{b}}]$, it is analytic in a complex neighborhood. Let $\mathcal{C}$ be a positively oriented contour enclosing $[\varepsilon_1, \varepsilon_{N_\mathrm{b}}]$. Then, for H close enough to H_*, we have

$$f(H) = \frac{1}{2\pi i}\oint_{\mathcal{C}} f(z)\frac{1}{z-H}dz \tag{50}$$

and analyticity of f follows. For δH small enough,

$$\begin{aligned}
&f(H_*+\delta H)\\
&= \frac{1}{2\pi i}\oint_{\mathcal{C}} f(z)\frac{1}{z-H_*-\delta H}dz\\
&\approx f(H_*) + \frac{1}{2\pi i}\oint_{\mathcal{C}} f(z)\frac{1}{z-H_*}\delta H\frac{1}{z-H_*}dz\\
&= f(H_*) + \frac{1}{2\pi i}\oint_{\mathcal{C}} \sum_{i=1}^{N_\mathrm{b}}\sum_{j=1}^{N_\mathrm{b}} \frac{f(z)\langle\phi_i,\delta H\phi_j\rangle}{(z-\varepsilon_i)(z-\varepsilon_j)}|\phi_i\rangle\langle\phi_j|dz\\
&= f(H_*) + \sum_{i=1}^{N_\mathrm{b}}\sum_{j=1}^{N_\mathrm{b}} \frac{f(\varepsilon_i)-f(\varepsilon_j)}{\varepsilon_i-\varepsilon_j}\langle\phi_i,\delta H\phi_j\rangle|\phi_i\rangle\langle\phi_j|
\end{aligned} \tag{51}$$

where $\approx$ means up to terms of order $O\left(\|\delta H\|^2\right)$. □

Proof of Theorem 1. If $\alpha_0 \le 1$, H_n belongs to the convex hull spanned by H_0 and $\{H_{\mathrm{KS}}(f_{\mathrm{FD}}(H)), H\in\mathcal{H}\}$. On this compact set X, f_{FD}, $\mathcal{I}$ and their derivatives are bounded. We have for all $H \in X$

$$\begin{aligned}
&\mathcal{I}(H+\alpha(H_{\mathrm{KS}}-H))\\
&= \mathcal{I}(H) - \alpha\langle\mathbf{\Omega}^{-1}(H_{\mathrm{KS}}-H),(H_{\mathrm{KS}}-H)\rangle + O(\alpha^2)\\
&= \mathcal{I}(H) - \alpha\langle\mathbf{\Omega}\nabla\mathcal{I}(H),\nabla\mathcal{I}(H)\rangle + O(\alpha^2)
\end{aligned} \tag{52}$$

where in this expression the functions $\mathbf{\Omega}$ and H_{KS} are evaluated at $f_{\mathrm{FD}}(H)$, and the constant in the $O(\alpha^2)$ term is uniform in n. It follows that for α_0 small enough, there is $c > 0$ such that

$$\mathcal{I}(H_{n+1}) \le \mathcal{I}(H_n) - \alpha c\|\nabla\mathcal{I}(H_n)\|^2, \tag{53}$$

and therefore $\nabla\mathcal{I}(H_n)\to 0$, so that $H_{\mathrm{KS}}(f_{\mathrm{FD}}(H_n)) - H_n \to 0$.

We now proceed as in Levitt[49]. Let $\mathcal{I}_* = \lim_{n\to\infty}\mathcal{I}(H_n)$. The set $\Gamma = \{H\in X, \mathcal{I}(H) = \lim_{n\to\infty}\mathcal{I}(H_n)\}$ is non-empty and compact. Furthermore, $d(H_n,\Gamma)\to 0$; if this was not the case, we could extract by compactness of X a subsequence at finite distance of Γ converging to a $H_*\in X$ satisfying $\mathcal{I}(H_*) = \lim_{n\to\infty}\mathcal{I}(H_n)$, which would imply that $H_*\in\Gamma$, a contradiction.

At every point H of Γ, by analyticity there is a neighborhood of H in $\mathcal{H}$ such that the Łojasiewicz inequality

$$|\mathcal{I}(H') - \mathcal{I}_*|^{1-\theta_H} \le \kappa_H\|\nabla\mathcal{I}(H')\| \tag{54}$$

holds for some constants $\theta_H\in(0,1/2]$, $\kappa_H > 0$[49,50]. By compactness, we can extract a finite covering of these neighborhoods, and obtain a Łojasiewicz inequality with universal constants $\theta\in(0,1/2], \kappa > 0$ in a neighborhood of Γ. Therefore, for n large enough, using the concavity inequality $x^\theta \le y^\theta + \theta y^{\theta-1}(x-y)$ with $x = \mathcal{I}(H_{n+1}) - \mathcal{I}_*$,

$y = \mathcal{I}(H_n) - \mathcal{I}_*$, we get

$$\begin{aligned}
\|\nabla\mathcal{I}(H_n)\|^2 &\leq \frac{1}{\alpha c}\mathcal{I}(H_n) - \mathcal{I}(H_{n+1}) \\
&\leq \frac{1}{\theta\alpha c}(\mathcal{I}(H_n) - \mathcal{I}_*)^{1-\theta} \\
&\qquad \cdot \Big[(\mathcal{I}(H_n) - \mathcal{I}_*)^{\theta} - (\mathcal{I}(H_{n+1}) - \mathcal{I}_*)^{\theta} \Big] \\
&\leq \frac{\kappa}{\theta\alpha c}\|\nabla\mathcal{I}(H_n)\| \\
&\qquad \cdot \Big[(\mathcal{I}(H_n) - \mathcal{I}_*)^{\theta} - (\mathcal{I}(H_{n+1}) - \mathcal{I}_*)^{\theta} \Big] \\
\|\nabla\mathcal{I}(H_n)\| &\leq \frac{\kappa}{\theta\alpha c}\Big[(\mathcal{I}(H_n) - \mathcal{I}_*)^{\theta} - (\mathcal{I}(H_{n+1}) - \mathcal{I}_*)^{\theta} \Big]
\end{aligned} \tag{55}$$

It follows that $\|\nabla\mathcal{I}(H_n)\|$ is summable, and therefore that $\|H_{n+1} - H_n\|$ is; this implies convergence of H_n to some H_*. When $\theta = 1/2$ (or, in light of (22), when $\boldsymbol{d^2\mathcal{E}}(f_{\mathrm{FD}}(H_*))$ is positive definite), we can get exponential convergence[49]. □

Note that the bounds used in the proof of the above statement (for instance, on α_0) are extremely pessimistic, since they rely on the fact that the set of possible P is bounded, and therefore all density matrices of the form $f_{\mathrm{FD}}(H_{\mathrm{KS}}(P))$ have occupations bounded away from 0 and 1, which results in bounded derivatives for $\mathrm{Tr}(s(P))$. A more careful analysis is needed to obtain better bounds (for instance, bounds that are better behaved in the zero temperature limit).

Proof of Theorem 2. From (52) it is easily seen that the linesearch process stops in a finite number of iterations, independent on n. This ensures that there is $\alpha_{\min} > 0$ such that $\alpha_{\min} \leq \alpha_n \leq \alpha_{\max}$. From (26) it follows that a similar inequality to (53) holds, and the rest of the proof proceeds as in that of Theorem 1.

* herbst@acom.rwth-aachen.de
† antoine.levitt@inria.fr

[1] N. D. Woods, M. C. Payne, and P. J. Hasnip, Journal of Physics: Condensed Matter **31**, 453001 (2019).
[2] S. Lehtola, F. Blockhuys, and C. Van Alsenoy, Molecules **25**, 1218 (2020).
[3] A. Jain, Y. Shin, and K. A. Persson, Nature Reviews Materials **1** (2016), 10.1038/natrevmats.2015.4.
[4] K. Alberi, M. B. Nardelli, A. Zakutayev, L. Mitas, S. Curtarolo, A. Jain, M. Fornari, N. Marzari, I. Takeuchi, M. L. Green, M. Kanatzidis, M. F. Toney, S. Butenko, B. Meredig, S. Lany, U. Kattner, A. Davydov, E. S. Toberer, V. Stevanovic, A. Walsh, N.-G. Park, A. Aspuru-Guzik, D. P. Tabor, J. Nelson, J. Murphy, A. Setlur, J. Gregoire, H. Li, R. Xiao, A. Ludwig, L. W. Martin, A. M. Rappe, S.-H. Wei, and J. Perkins, Journal of Physics D: Applied Physics **52**, 013001 (2019).
[5] S. Luo, T. Li, X. Wang, M. Faizan, and L. Zhang, WIREs Computational Molecular Science **11** (2021), 10.1002/wcms.1489.
[6] S. Curtarolo, W. Setyawan, G. L. Hart, M. Jahnatek, R. V. Chepulskii, R. H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M. J. Mehl, H. T. Stokes, D. O. Demchenko, and D. Morgan, Computational Materials Science **58**, 218 (2012).
[7] A. Jain, G. Hautier, C. J. Moore, S. Ping Ong, C. C. Fischer, T. Mueller, K. A. Persson, and G. Ceder, Computational Materials Science **50**, 2295 (2011).
[8] S. P. Huber, S. Zoupanos, M. Uhrin, L. Talirz, L. Kahle, R. Häuselmann, D. Gresch, T. Müller, A. V. Yakutovich, C. W. Andersen, F. F. Ramirez, C. S. Adorf, F. Gargiulo, S. Kumbhar, E. Passaro, C. Johnston, A. Merkys, A. Cepellotti, N. Mounet, N. Marzari, B. Kozinsky, and G. Pizzi, Scientific Data **7** (2020), 10.1038/s41597-020-00638-4.
[9] W. Feng and K. Cameron, Computer **40**, 50 (2007).
[10] W. Feng, X. Feng, and R. Ge, IT Professional **10**, 17 (2008).
[11] M. F. Herbst and A. Levitt, Journal of Physics: Condensed Matter (2020), 10.1088/1361-648x/abcbdb.
[12] E. Cancès, "SCF algorithms for HF electronic calculations," (Springer Berlin Heidelberg, 2000) pp. 17–43.
[13] E. Cancès and C. Le Bris, International Journal of Quantum Chemistry **79**, 82 (2000).
[14] K. N. Kudin, G. E. Scuseria, and E. Cancès, The Journal of Chemical Physics **116**, 8255 (2002), http://aip.scitation.org/doi/pdf/10.1063/1.1470195.
[15] J. B. Francisco, J. M. Martınez, and L. Martınez, The Journal of chemical physics **121**, 10863 (2004).
[16] J. B. Francisco, J. M. Martínez, and L. Martínez, Journal of Mathematical Chemistry **40**, 349 (2006).
[17] E. Cancès, G. Kemlin, and A. Levitt, SIAM Journal on Matrix Analysis and Applications **42**, 243 (2021).
[18] N. Marzari, D. Vanderbilt, and M. C. Payne, Physical Review Letters **79**, 1337 (1997).
[19] C. Freysoldt, S. Boeck, and J. Neugebauer, Physical Review B **79**, 241103 (2009).
[20] L. D. Marks, Journal of Chemical Theory and Computation **17**, 5715 (2021).
[21] L. Marks and D. Luke, Physical Review B **78**, 075114 (2008).
[22] X. Gonze, Physical Review B **54**, 4383 (1996).
[23] N. D. Mermin, Physical Review **137**, A1441 (1965).
[24] M. Methfessel and A. Paxton, Physical Review B **40**, 3616 (1989).
[25] S. L. Adler, Physical Review **126**, 413 (1962).
[26] N. Wiser, Physical Review **129**, 62 (1963).
[27] P. H. Dederichs and R. Zeller, Physical Review B **28**, 5462 (1983).
[28] G. P. Kerker, Physical Review B **23**, 3082 (1981).
[29] H. Walker and P. Ni, SIAM Journal on Numerical Analysis **49**, 1715 (2011), https://doi.org/10.1137/10078356X.
[30] M. Chupin, M.-S. Dupuy, G. Legendre, and E. Séré, (2020), 2002.12850v1.
[31] Y. Saad, *Iterative Methods for Sparse Linear Systems*, 2nd ed., edited by Y. Saad (SIAM Publishing, 2003).
[32] G. Kresse and J. Furthmüller, Physical Review B **54**, 11169 (1996).
[33] M. F. Herbst, A. Levitt, and E. Cancès, Proc. JuliaCon Conf. **3**, 69 (2021).

[34] M. F. Herbst and A. Levitt, "Density-functional toolkit (DFTK), version v0.3.10," (2021), `https://dftk.org`.
[35] J. P. Perdew, K. Burke, and M. Ernzerhof, Physical Review Letters **77**, 3865 (1996).
[36] S. Lehtola, C. Steigemann, M. J. Oliveira, and M. A. Marques, SoftwareX **7**, 1 (2018).
[37] S. Goedecker, M. Teter, and J. Hutter, Physical Review B **54**, 1703 (1996).
[38] M. F. Herbst and A. Levitt, "Computational scripts and raw data for the presented numerical tests," (2021), `https://github.com/mfherbst/supporting-adaptive-damping`.
[39] M. Bercx and N. Marzari, Private communication (2020).
[40] M. Winkelmann, E. Di Napoli, D. Wortmann, and S. Blügel, Physical Review B **102** (2020), 10.1103/physrevb.102.195138.
[41] M. Belkhouane, S. Amari, A. Yakoubi, A. Tadjer, S. Méçabih, G. Murtaza, S. Bin Omran, and R. Khenata, Journal of Magnetism and Magnetic Materials **377**, 211 (2015).
[42] L. Wollmann, S. Chadov, J. Kübler, and C. Felser, Physical Review B **92** (2015), 10.1103/physrevb.92.064417.
[43] B. Shi, J. Li, C. Zhang, W. Zhai, S. Jiang, W. Wang, D. Chen, Y. Yan, G. Zhang, and P.-F. Liu, Physical Chemistry Chemical Physics **22**, 23185 (2020).
[44] J. He, S. S. Naghavi, V. I. Hegde, M. Amsler, and C. Wolverton, Chemistry of Materials **30**, 4978 (2018).
[45] S. Jiang and K. Yang, Journal of Alloys and Compounds **867**, 158854 (2021).
[46] W. W. Hager and H. Zhang, Pacific Journal of Optimization **2**, 35 (2006).
[47] N. J. Higham, *Functions of matrices: theory and computation* (SIAM, 2008).
[48] A. Levitt, Archive for Rational Mechanics and Analysis **238**, 901 (2020).
[49] A. Levitt, ESAIM: Mathematical Modelling and Numerical Analysis **46**, 1321 (2012).
[50] S. Lojasiewicz, Lectures Notes IHES (Bures-sur-Yvette) (1965).