Pulay forces in density-functional theory with extended Hubbard functionals: From nonorthogonalized to orthogonalized manifolds

Iurii Timrov [1,*], Francesco Aquilante,[1] Luca Binci,[1] Matteo Cococcioni,[2] and Nicola Marzari[1]

[1]*Theory and Simulation of Materials (THEOS), and National Centre for Computational Design and Discovery of Novel Materials (MARVEL),*
École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland
[2]*Department of Physics, University of Pavia, I-27100 Pavia, Italy*

We present a derivation of the exact expression for Pulay forces in density-functional theory calculations augmented with extended Hubbard functionals and arising from the use of orthogonalized atomic orbitals as projectors for the Hubbard manifold. The derivative of the inverse square root of the orbital overlap matrix is obtained as a closed-form solution of the associated Lyapunov (Sylvester) equation. The expression for the resulting contribution to the forces is presented in the framework of ultrasoft pseudopotentials and the projector-augmented-wave method and using a plane-wave basis set. We have benchmarked the present implementation with respect to finite differences of total energies for the case of NiO, finding excellent agreement. Owing to the accuracy of Hubbard-corrected density-functional theory calculations—provided the Hubbard parameters are computed for the manifold under consideration—the present work paves the way for systematic studies of solid-state and molecular transition-metal and rare-earth compounds.

I. INTRODUCTION

Density-functional theory (DFT) [1,2] with approximate exchange-correlation functionals has been remarkably successful in predicting ground-state properties of a large variety of systems. However, most functionals [such as the local-density approximation (LDA) and generalized-gradient approximation (GGA)] fail in capturing both qualitatively and quantitatively the ground state of systems with strongly localized electrons (typically of d and/or f character), due to large self-interaction errors (SIE) [3,4]. There are different schemes that can be used in DFT to alleviate SIE; in particular, we mention here self-interaction corrections (SIC) [3,5–7], hybrid functionals [8–14], meta-GGA functionals [15–20], and DFT+U [21–27] or its extension DFT+U+V [28–30].

DFT with (extended) Hubbard functionals—DFT+U (DFT+U+V)—is popular due to its simplicity and much improved accuracy in describing structural, electronic, and magnetic ground-state properties of transition-metal and rare-earth compounds [27,31] by removing self-interactions for a subset of electronic states (i.e., states in the Hubbard manifold) [24]. While DFT+U with an onsite Hubbard U correction is effective for many systems with strongly localized electrons, the intersite Hubbard V contribution is crucial in many systems having also a strong covalent hybridization for the same orbitals; case studies have been the evaluation of voltages in Li-ion batteries [32], the determination of formation energies of oxygen vacancies in perovskites [33], and geometries and energetics in molecular systems [26]. Key aspects of these methods are (i) the

choice of the Hubbard parameters (onsite U and intersite V) and (ii) the choice of projector functions that are used to construct the Hubbard manifold. These two fundamental points are not independent. Hubbard parameters that are computed from first principles (e.g., using constrained DFT (cDFT) [34–43], Hartree-Fock-based approaches [29,44–47], constrained random-phase approximation (cRPA) [48–57], or linear-response theory (LRT) [58,59]) vary substantially, depending on the choice of the projector functions for the Hubbard manifold, pseudopotentials and the oxidation state [25,42], exchange-correlation functionals, and chemical environment of the Hubbard atoms [60,61]. Therefore, it is crucial to use the U and V parameters consistently with the Hubbard manifold and other technicalities (pseudopotentials, oxidation states, functionals, etc.) which were used to compute them. In addition, it is important to stress that the final quantities of interest (energies, bond lengths, etc.) are insensitive to large variations (2–3 eV) in interaction parameters computed using pseudopotentials generated in different oxidation states provided that these interaction parameters are computed self-consistently (e.g., using LRT) (see the Appendix of Ref. [25]).

Since the early days [21], DFT+U became widely used and implemented in different electronic-structure codes. However, different projector functions for the Hubbard manifold are used in the available implementations (see below), which makes it difficult to compare results obtained with different codes; in addition, it is still a common practice to choose empirical values of U, disregarding the underlying definition of the Hubbard manifold. In Ref. [62], an effort was made to compare and analyze various types of projector functions for a set of U values, and quite large variations in the results (e.g., density of states, energy differences) were obtained,

---

*e-mail: iurii.timrov@epfl.ch

especially for systems with strong covalent interactions compared to the ones with ionic interactions.

There are quite many possible projector functions to use as a basis for the Hubbard manifold (see, e.g., Ref. [63]). In particular, we highlight here nonorthogonalized atomic orbitals (NAO) [58,64], orthogonalized atomic orbitals (OAO) [32,33,65], nonorthogonalized Wannier functions (NWF) [66], orthogonalized Wannier functions (OWF) [67], linearized augmented plane-wave (LAPW) approaches [68], and projector-augmented-wave (PAW) projector functions [69,70]. A common feature of all these projector functions is that they are spatially localized and depend explicitly on atomic positions; hence, an extra term appears when computing derivatives of the Hubbard corrective energy with respect to atomic displacements (Pulay force) or strain (Pulay stress). The expressions for Pulay forces in the context of Hubbard-corrected DFT were derived for NAO [71,72], NWF [73], OWF [74], and LAPW [75], but no derivation was made so far for the case of OAO. The scope of the present work is to extend the existing expression for Pulay forces using NAO [71,72] to orthogonalized basis sets based on the Löwdin scheme [76,77].

The importance of this development is that NAO and OAO are of special interest due to the simplicity of their implementation and transparency in their practical use. In particular, DFT+U calculations with NAO have proven to be accurate for simulations of various properties in many materials [60,78–81]. However, NAO have also drawbacks dictated by the fact that atomic orbitals can have long tails that extend to a significant spatial range, and as a consequence the formal occupation numbers for these orbitals can be substantially overestimated [62]. In some NAO-based works [64] these tails are truncated (i.e., atomic orbitals are zero outside of "atomic spheres"), and even in LAPW and PAW-based approaches there is some ambiguity in the definition of the cutoff radius at which the projector functions are truncated [43]. Such ambiguity has implications on the final quantities of interest that are computed [43,62]; in addition, in the case of NAO with long tails, the Hubbard correction is essentially applied twice in the overlap regions between atoms, which can further enhance spurious unphysical effects in Hubbard-corrected DFT calculations. These issues can be removed by orthogonalizing the atomic orbitals among all atoms: This insures that Hubbard corrections are applied only once to the respective Hubbard manifolds. Moreover, DFT+U with OAO captures some intersite corrections through orthogonalization of the orbitals of one atom combined with those from neighbor sites, thus making DFT+U closer to the full DFT+U+V. Therefore, OAO is a very attractive alternative to NAO, and in fact it was already observed that the former gives more accurate energetics than the latter [32,33]. Until the present work, though, atomic and cell relaxations with OAO were not possible due to the difficulty in evaluating the derivative of the inverse square root of the orbital overlap matrix, which appears when using the Löwdin scheme [76,77].

In this work, we present a derivation that allows us to calculate Pulay (Hubbard) forces for the case of OAO projector functions, by starting from the expressions for the Hubbard force in the case of NAO [71,72] and using the Hubbard parameters U and V computed using density-functional perturbation theory (DFPT) [59,82]. We present a detailed mathematical formulation of the derivative of the inverse square root of the orbital overlap matrix as a closed-form solution of the Lyapunov (Sylvester) equation, which is the main result of the present formalism, and we compare it with other techniques that were used in literature to compute such a derivative. The formalism is presented in the framework of DFT+U+V, i.e., by taking into account not only Hubbard forces coming from the onsite U term but also from the intersite V term. For the sake of generality, we present the derivation in the case of ultrasoft (US) pseudopotentials (PPs) and PAW. It is worth to note that we have also derived and implemented the formalism for the Hubbard stress using OAO, but this will not be discussed here because this is not the focus of this paper.

The paper is organized as follows: Section II A presents the basics of DFT+U and DFT+U+V in the framework of US and PAW PPs; in Sec. II B we discuss NAO and OAO; Sec. II C discusses the Hellmann-Feynman theorem and its generalization to US and PAW PPs; in Sec. II D and Sec. II E we present the derivation of Hubbard forces in the cases of NAO and OAO, respectively; Sec. III contains technical details of our calculations; in Sec. IV we benchmark the implementation of Hubbard forces using OAO versus the finite difference method, and make comparisons of total and Hubbard forces in the case of OAO and NAO; and finally, in Sec. V we give our conclusions. In Appendix A we give an alternative (approximate) expression for the derivative of the inverse square root of the overlap matrix based on the Taylor series expansion, in Appendix B we present a proof that the solution of the Lyapunov equation can be written in a closed form, and in Appendix C we present the discussion about the computational scaling of forces. Hartree atomic units are used throughout the paper. For the sake of simplicity, the formalism is presented for insulators. We will use notations and definitions similar to those in Ref. [83]. Quantum-mechanical operators will be indicated with a hat on top of capital letters ("$\hat{A}$"), while for matrices we will use a special font ("A").

II. THEORY

In this section we present the formalism for calculation of Hubbard forces starting from the expression for the Hubbard energy in the framework of DFT with extended Hubbard functionals. The main results of this paper can be divided into two parts: (i) generalization of the DFT+U formalism to the framework of DFT+U+V and to the use of US or PAW PPs in the context of NAO and (ii) generalization of "(i)" to OAO. The former is discussed in Secs. II C and II D, while the latter in Sec. II E.

A. DFT+U+V

In this section we briefly recall the basics of the DFT+U+V formalism in the simplified rotationally invariant form, which were presented in Refs. [23,28,58] for norm-conserving (NC) PPs, and later were extended to US and PAW PPs in Refs. [60,82]. This reminder is needed, because it will be our starting point for the derivation of expressions for forces.

Similarly to DFT+U, DFT+U+V is also based on an additive correction to the approximate DFT energy functional [28]:

$$E_{\text{TOT}} = E_{\text{DFT}} + E_{\text{Hub}}, \tag{1}$$

where E_{TOT} is the total energy functional, E_{DFT} is the standard DFT energy functional based on LDA or GGA, and E_{Hub} is the Hubbard energy functional which is the difference between the Hubbard term and its mean-field approximation which is subtracted to avoid the double counting of interactions already included in E_{DFT}. In the present work, this latter term is shaped according to the popular fully localized limit [23,27,84]. The generalized Hubbard corrective energy is defined as [28]:

$$E_{\text{Hub}} = \frac{1}{2} \sum_{I} \sum_{\sigma m_1 m_2} U^I \big(\delta_{m_1 m_2} - n_{m_1 m_2}^{II\sigma} \big) n_{m_2 m_1}^{II\sigma}$$
$$- \frac{1}{2} \sum_{I} \sum_{J(J \neq I)}^{*} \sum_{\sigma m_1 m_2} V^{IJ} n_{m_1 m_2}^{IJ\sigma} n_{m_2 m_1}^{JI\sigma}, \tag{2}$$

where I and J are the atomic site indices, m_1 and m_2 are the magnetic quantum numbers associated with a specific angular momentum (i.e., orbitals of the Hubbard manifold) of atoms I and J, respectively, U^I and V^{IJ} are the onsite and intersite Hubbard parameters, respectively, and the star in the sum on the second line denotes that for each atom I index J covers all its nearest neighbors up to a given distance (or belonging to a given shell). Typically, only nearest neighbors are considered in the intersite term, but the formalism is general and hence allows to include next-nearest neighbors and even further ones when needed.

In Eq. (2), $n_{m_1 m_2}^{IJ\sigma}$ are the generalized occupation matrices that are based on a projection of Kohn-Sham (KS) states $\psi_{i\sigma}(\mathbf{r})$ on the Hubbard manifold specific for each atom $\{\varphi_m^I(\mathbf{r})\}$:

$$n_{m_1 m_2}^{IJ\sigma} = \sum_i \langle \psi_{i\sigma} | \hat{S} | \varphi_{m_2}^J \rangle \langle \varphi_{m_1}^I | \hat{S} | \psi_{i\sigma} \rangle$$
$$= \sum_i \langle \psi_{i\sigma} | \hat{P}_{m_2 m_1}^{JI} | \psi_{i\sigma} \rangle, \tag{3}$$

where

$$\hat{P}_{m_2 m_1}^{JI} = \hat{S} | \varphi_{m_2}^J \rangle \langle \varphi_{m_1}^I | \hat{S}. \tag{4}$$

Here index i labels KS states, $\hat{P}_{m_2 m_1}^{JI}$ is the generalized projector on Hubbard manifolds of atoms J and I, and $\varphi_{m_1}^I(\mathbf{r}) \equiv \varphi_{m_1}^{\gamma(I)}(\mathbf{r} - \mathbf{R}_I)$ are the functions centered on the Ith atom of type $\gamma(I)$ at the position $\mathbf{R}_I$. Depending on the atomic type, functions $\varphi_{m_1}^I(\mathbf{r})$ can be either localized functions (of d or f character like for transition-metal and rare-earth elements) or spread functions (of s or p character for other elements). In the following sections we will discuss in more details these functions and their role in Hubbard-corrected DFT calculations.

In Eq. (4), $\hat{S}$ is the operator of the US or PAW PPs schemes, which reads

$$\hat{S} = 1 + \sum_{I\mu\nu} q_{\mu\nu}^{\gamma(I)} | \beta_\mu^I \rangle \langle \beta_\nu^I |, \tag{5}$$

where

$$q_{\mu\nu}^{\gamma(I)} = \int Q_{\mu\nu}^{\gamma(I)}(\mathbf{r}) \, d\mathbf{r}. \tag{6}$$

Here μ and ν are the indices which label Q and β functions, $Q_{\mu\nu}^{\gamma(I)}(\mathbf{r} - \mathbf{R}_I)$ are the localized augmentation functions pertaining to the pseudopotential of the Ith atom, $\beta_\mu^I(\mathbf{r}) \equiv \beta_\mu^{\gamma(I)}(\mathbf{r})$ are the so-called projector functions of the US or PAW PPs schemes [not to be confused with the projector functions on the Hubbard manifold, $\varphi_{m_1}^I(\mathbf{r})$] that are localized on the Ith atom and vanish outside spheres centered on atoms [85], and the integration in Eq. (6) is performed over a crystal volume.

In DFT+U, only the first line in Eq. (2) is preserved, while the term on the second line is zero because the intersite interactions are neglected, i.e., $V^{IJ} = 0$. In this case, which corrects only the onsite interactions, it is useful to adopt the following notation: $n_{m_1 m_2}^{I\sigma} \equiv n_{m_1 m_2}^{II\sigma}$ and $\hat{P}_{m_1 m_2}^I \equiv \hat{P}_{m_1 m_2}^{II}$. It is easy to see from Eq. (2) that the two terms of the corrective energy functional, proportional to the onsite (U^I) and intersite (V^{IJ}) interactions, counteract each other. In fact, while the onsite term favors localization on atomic sites (by suppressing intersite hybridization of orbitals), the intersite terms restore and stabilize hybridized states in the interstitial regions between neighboring atoms that are characteristic for covalent interactions. Therefore, in systems with predominantly ionic interactions DFT+U is expected to be sufficient and a good level of approximation, while in systems with predominantly covalent interactions DFT+U+V is needed.

For the purpose of this work it is important to discuss the contribution to the KS potential stemming from the extended Hubbard functional [see Eq. (2)]. The action of this term on KS wave functions can be easily obtained by taking a functional derivative of E_{TOT} [see Eq. (1)] with respect to the complex conjugate of KS wave functions [28,86]. The term corresponding to the functional derivative of E_{Hub} [see Eq. (2)] reads:

$$\hat{V}_{\text{Hub},\sigma} = \sum_{I} \sum_{m_1 m_2} U^I \left(\frac{\delta_{m_1 m_2}}{2} - n_{m_1 m_2}^{I\sigma} \right) \hat{P}_{m_1 m_2}^I$$
$$- \sum_{I} \sum_{J(J \neq I)}^{*} \sum_{m_1 m_2} V^{IJ} n_{m_1 m_2}^{IJ\sigma} \hat{P}_{m_1 m_2}^{IJ}. \tag{7}$$

Therefore, the generalized KS equations with the Hubbard corrections can be written as

$$\hat{H}_\sigma | \psi_{i\sigma} \rangle = \varepsilon_{i\sigma} \hat{S} | \psi_{i\sigma} \rangle, \tag{8}$$

where $\varepsilon_{i\sigma}$ are the KS energies, and

$$\hat{H}_\sigma = \hat{H}_{\text{DFT},\sigma} + \hat{V}_{\text{Hub},\sigma}, \tag{9}$$

with $\hat{H}_{\text{DFT},\sigma}$ being the standard DFT Hamiltonian (LDA or GGA), and $\hat{V}_{\text{Hub},\sigma}$ is the Hubbard potential given by Eq. (7). For generalized KS equations, the orthonormality condition reads:

$$\langle \psi_{i\sigma} | \hat{S} | \psi_{i'\sigma'} \rangle = \delta_{ii'} \delta_{\sigma\sigma'}. \tag{10}$$

The DFT Hamiltonian $\hat{H}_{\text{DFT},\sigma}$ contains usual terms [85,87]), among which the Hartree and exchange-correlation potentials

depend on the charge density, which in the US and PAW PPs cases reads:

$$\rho_\sigma(\mathbf{r}) = \sum_i |\psi_{i\sigma}(\mathbf{r})|^2$$
$$+ \sum_i \sum_{I\mu\nu} Q_{\mu\nu}^{\gamma(I)}(\mathbf{r} - \mathbf{R}_I) \langle \psi_{i\sigma} | \beta_\mu^I \rangle \langle \beta_\nu^I | \psi_{i\sigma} \rangle. \tag{11}$$

Finally, Hubbard parameters U^I and V^{IJ}, which are needed for the formalism presented above, can be computed from first principles using, e.g., linear response theory [58] with its recent reformulation based on density-functional perturbation theory [59,82]. As was mentioned in Sec. I, values of U^I and V^{IJ} depend strongly on the choice of projector functions of the Hubbard manifold, $\varphi_{m_1}^I(\mathbf{r})$, as well as on the type of PPs [42], oxidation state [25,61], functional, and chemical composition of the system [60,61].

B. Choosing projector functions for the Hubbard manifold

One of the key aspects of the Hubbard-corrected DFT formalism is the choice of the projector functions for the Hubbard manifold. In other words, we need to choose the basis $\{\varphi_m^I(\mathbf{r})\}$ for the projector $\hat{P}_{m_2 m_1}^{JI}$ that was introduced in Sec. II A. In Sec. I we discussed what are the popular choices in literature for $\{\varphi_m^I(\mathbf{r})\}$. In this work, we focus our discussion only on two types of projector functions, NAO and OAO. Let us comment briefly about each of them.

NAO is one of the most simple projector functions for the Hubbard manifold, which is often a reasonable choice to represent the Hubbard manifold, especially in systems with mostly ionic character of interactions. NAO are provided with pseudopotentials, and these orbitals are orthonormal within each atom (i.e., Hubbard d orbitals are orthonormal to non-

Hubbard s and p orbitals of the same atom) but not between different atoms. However, whenever covalent interactions become important, this type of projector functions is not the best choice (see the discussion in Sec. I), and intersite orthogonalization becomes important.

OAO are obtained by taking atomic orbitals of each atom and then orthogonalizing them to all orbitals of all atoms in the system. In this work, we will use the Löwdin orthogonalization method [76,77]. By doing so, we obtain a new set of orbitals, that are all orthogonalized, which now better represent hybridizations of orbitals between neighboring sites. This choice is particularly good for setting up the Hubbard manifold, because it allows us to avoid counting Hubbard corrections twice in the interstitial regions between atoms, which is especially relevant in the case of DFT+U+V.

The Hubbard-corrected DFT formalism presented in Sec. II A is general and hence it applies both to NAO and OAO. However, the expressions for the Hubbard forces have differences, which we will detail in the following. First, we will briefly recall the formalism for Hubbard forces that are computed using NAO (Sec. II D), and, second, we will present the generalization to OAO and highlight what are the difference with the NAO case (Sec. II E).

C. Total forces in DFT+U+V

In this section we discuss how to evaluate Hubbard forces starting from the expression for the total energy, Eq. (1). The main idea is based on the Hellmann-Feynman theorem which states that in the case of NC PPs the derivative of the total energy with respect to some small perturbation (in this case the perturbation is the atomic displacement) equals to the expectation value of the derivative of the Hamiltonian. However, in the case of US and PAW PPs, there is a contribution coming also from the derivatives of the $\hat{S}$ operator [60,83]. Therefore, the total force acting on the Kth atom upon its displacement is

$$\mathbf{F}_{\text{TOT},K} = -\frac{dE_{\text{TOT}}}{d\mathbf{R}_K} = -\sum_{i\sigma} \int \frac{\delta E_{\text{TOT}}}{\delta \psi_{i\sigma}^*(\mathbf{r})} \frac{d\psi_{i\sigma}^*(\mathbf{r})}{d\mathbf{R}_K} d\mathbf{r} - \sum_{i\sigma} \int \frac{\delta E_{\text{TOT}}}{\delta \psi_{i\sigma}(\mathbf{r})} \frac{d\psi_{i\sigma}(\mathbf{r})}{d\mathbf{R}_K} d\mathbf{r} - \frac{\partial E_{\text{TOT}}}{\partial \mathbf{R}_K}$$

$$= -\sum_{i\sigma} \left\{ \left\langle \frac{d\psi_{i\sigma}}{d\mathbf{R}_K} \middle| \hat{H}_\sigma \middle| \psi_{i\sigma} \right\rangle + \left\langle \psi_{i\sigma} \middle| \hat{H}_\sigma \middle| \frac{d\psi_{i\sigma}}{d\mathbf{R}_K} \right\rangle + \left\langle \psi_{i\sigma} \middle| \frac{\partial \hat{H}_\sigma}{\partial \mathbf{R}_K} \middle| \psi_{i\sigma} \right\rangle \right\}$$

$$= -\sum_{i\sigma} \left\langle \psi_{i\sigma} \middle| \left(\frac{\partial \hat{H}_{\text{DFT},\sigma}}{\partial \mathbf{R}_K} - \varepsilon_{i\sigma} \frac{\partial \hat{S}}{\partial \mathbf{R}_K} \right) \middle| \psi_{i\sigma} \right\rangle - \sum_{i\sigma} \left\langle \psi_{i\sigma} \middle| \frac{\partial \hat{V}_{\text{Hub},\sigma}}{\partial \mathbf{R}_K} \middle| \psi_{i\sigma} \right\rangle, \tag{12}$$

where the first two terms (in the last row) come from the standard DFT in the US or PAW PPs formalism, and the last term is the Hubbard force:

$$\mathbf{F}_{\text{Hub},K} = -\frac{\partial E_{\text{Hub}}}{\partial \mathbf{R}_K} = -\sum_{i\sigma} \left\langle \psi_{i\sigma} \middle| \frac{\partial \hat{V}_{\text{Hub},\sigma}}{\partial \mathbf{R}_K} \middle| \psi_{i\sigma} \right\rangle. \tag{13}$$

In the derivation of Eq. (12) we used Eqs. (8) and (9) and the derivative of Eq. (10) [60,83]. Therefore, the Hubbard contribution to the force can be separated from the standard DFT force, and considered in more detail. It turns out that in practice it is more convenient to work directly with the

derivative $\frac{\partial E_{\text{Hub}}}{\partial \mathbf{R}_K}$ rather than with the matrix element of $\frac{\partial \hat{V}_{\text{Hub},\sigma}}{\partial \mathbf{R}_K}$ [27,71,72], and therefore we will follow this strategy.

D. Hubbard forces: The case of nonorthogonalized atomic orbitals

Let us consider the Hubbard force in the basis of NAO. This derivation was already presented in the case of DFT+U with NC PPs [71,72], and generalized to the case of US PPs in the context of the SIC method of Ref. [88] which has close similarities with DFT+U. Here, we present a generalization

to the case of DFT+U+V, in the general framework of US and PAW PPs, which is the first main result of this paper.

As was mentioned in Sec. II C, we need to evaluate the derivative $\frac{\partial E_{\text{Hub}}}{\partial \mathbf{R}_K}$. Using Eq. (2), we obtain:

$$
\begin{aligned}
\frac{\partial E_{\text{Hub}}}{\partial \mathbf{R}_K} &= \sum_{IJ} \sum_{\sigma m_1 m_2} \frac{\partial E_{\text{Hub}}}{\partial n_{m_1 m_2}^{IJ\sigma}} \frac{\partial n_{m_1 m_2}^{IJ\sigma}}{\partial \mathbf{R}_K} \\
&= \sum_I \sum_{\sigma m_1 m_2} U^I \left(\frac{\delta_{m_1 m_2}}{2} - n_{m_1 m_2}^{II\sigma} \right) \frac{\partial n_{m_2 m_1}^{II\sigma}}{\partial \mathbf{R}_K} \\
&\quad - \sum_I \sum_{J(J \neq I)}^{*} \sum_{\sigma m_1 m_2} V^{IJ} n_{m_1 m_2}^{IJ\sigma} \frac{\partial n_{m_2 m_1}^{II\sigma}}{\partial \mathbf{R}_K}.
\end{aligned}
\tag{14}
$$

Before we proceed, it is important to remark that Eq. (14) neglects derivatives of Hubbard parameters with respect to atomic displacements, namely $\frac{\partial U^I}{\partial \mathbf{R}_K} = 0$ and $\frac{\partial V^{IJ}}{\partial \mathbf{R}_K} = 0$. This is a standard approximation in literature, in part due to the fact that very often empirical values of Hubbard parameters are used and hence it is not possible to evaluate such derivatives (in fact, the error made due to the ambiguity in choosing empirical U^I is likely much larger than the error made due to neglecting changes in U^I due to atomic displacements). However, when Hubbard parameters are computed from first principles, it is in fact possible to compute their derivatives due to atomic displacements [89]. Here, for the sake of simplicity, we neglect derivatives of U^I and V^{IJ}, but this point deserves further considerations in future studies.

Therefore, the problem is reduced to the calculation of the derivative of the generalized occupation matrix with respect to atomic displacements, which can be written as [see Eq. (3)]:

$$
\frac{\partial n_{m_2 m_1}^{IJ\sigma}}{\partial \mathbf{R}_K} = \sum_i \left[\frac{\partial}{\partial \mathbf{R}_K} (\langle \psi_{i\sigma} | \hat{S} | \varphi_{m_1}^I \rangle) \langle \varphi_{m_2}^J | \hat{S} | \psi_{i\sigma} \rangle \right. \\
\left. + \langle \psi_{i\sigma} | \hat{S} | \varphi_{m_1}^I \rangle \frac{\partial}{\partial \mathbf{R}_K} (\langle \varphi_{m_2}^J | \hat{S} | \psi_{i\sigma} \rangle) \right].
\tag{15}
$$

Now the problem is to calculate the object $\frac{\partial}{\partial \mathbf{R}_K}(\langle \psi_{i\sigma} | \hat{S} | \varphi_{m_1}^I \rangle)$ and a similar one appearing in the equation above. Since KS wave functions do not depend explicitly on the atomic positions, the derivative $\frac{\partial \psi_{i\sigma}}{\partial \mathbf{R}_K}$ is zero [27,71,72]. Therefore, we obtain

$$
\begin{aligned}
&\frac{\partial}{\partial \mathbf{R}_K}(\langle \psi_{i\sigma} | \hat{S} | \varphi_{m_1}^I \rangle) \\
&= \left\langle \psi_{i\sigma} \left| \frac{\partial \hat{S}}{\partial \mathbf{R}_K} \right| \varphi_{m_1}^I \right\rangle + \left\langle \psi_{i\sigma} \left| \hat{S} \right| \frac{\partial \varphi_{m_1}^I}{\partial \mathbf{R}_K} \right\rangle.
\end{aligned}
\tag{16}
$$

The derivatives in Eq. (16) were briefly discussed in the case of the SIC method with US PPs in Ref. [88]. Here, we use such a generalization to US (and PAW) PPs in the context of DFT+U+V for the first time. The first term in Eq. (16) is present due to the use of US or PAW PPs and it has no counterpart in the NC PPs case [27,71,72]. By using Eq. (5)

we obtain:

$$
\begin{aligned}
\left\langle \psi_{i\sigma} \left| \frac{\partial \hat{S}}{\partial \mathbf{R}_K} \right| \varphi_{m_1}^I \right\rangle &= \sum_{L\mu\nu} q_{\mu\nu}^{\gamma(L)} \left[\left\langle \psi_{i\sigma} \left| \frac{\partial \beta_\mu^L}{\partial \mathbf{R}_K} \right\rangle \langle \beta_\nu^L | \varphi_{m_1}^I \rangle \right. \right. \\
&\quad \left. \left. + \langle \psi_{i\sigma} | \beta_\mu^L \rangle \left\langle \frac{\partial \beta_\nu^L}{\partial \mathbf{R}_K} \right| \varphi_{m_1}^I \right\rangle \right].
\end{aligned}
\tag{17}
$$

Due to the locality of projector functions β (we recall that these projector functions are different from zero only inside spheres centered on atoms) we have $\frac{\partial \beta_\mu^L}{\partial \mathbf{R}_K} = \delta_{LK} \frac{\partial \beta_\mu^K}{\partial \mathbf{R}_K}$, and, therefore, Eq. (17) becomes

$$
\begin{aligned}
\left\langle \psi_{i\sigma} \left| \frac{\partial \hat{S}}{\partial \mathbf{R}_K} \right| \varphi_{m_1}^I \right\rangle &= \sum_{\mu\nu} q_{\mu\nu}^{\gamma(K)} \left[\left\langle \psi_{i\sigma} \left| \frac{\partial \beta_\mu^K}{\partial \mathbf{R}_K} \right\rangle \langle \beta_\nu^K | \varphi_{m_1}^I \rangle \right. \right. \\
&\quad \left. \left. + \langle \psi_{i\sigma} | \beta_\mu^K \rangle \left\langle \frac{\partial \beta_\nu^K}{\partial \mathbf{R}_K} \right| \varphi_{m_1}^I \right\rangle \right].
\end{aligned}
\tag{18}
$$

It is important to stress that due to the presence of the US or PAW term given by Eq. (18), there are nonzero Hubbard forces even on non-Hubbard atoms (i.e., atoms on which we do not apply the Hubbard correction). Instead, in the case of NC PPs and NAO, Hubbard forces appear only on Hubbard atoms [27,71,72].

Finally, the second term in Eq. (16) requires computing the derivative of NAO with respect to atomic displacements, $\frac{\partial \varphi_{m_1}^I}{\partial \mathbf{R}_K}$, which is similar to the derivative $\frac{\partial \beta_\mu^K}{\partial \mathbf{R}_K}$ from the implementation point of view. These objects can be efficiently computed in the reciprocal space, as was discussed in detail in Refs. [27,71,72,88], and hence it will not be detailed here.

E. Hubbard forces: The case of orthogonalized atomic orbitals

In this section we present a generalization of the formalism discussed in Sec. II D to the case of OAO, which is the second main result of this paper. All the equations of Sec. II D hold also for OAO, by replacing nonorthogonalized atomic orbitals $\varphi_m^I(\mathbf{r})$ by orthogonalized ones $\tilde{\varphi}_m^I(\mathbf{r})$. The main difference with the case discussed so far is how to compute $\frac{\partial \tilde{\varphi}_m^I}{\partial \mathbf{R}_K}$. In the following of this section we present the definition of OAO and then discuss how to compute their derivatives.

1. Orthogonalized atomic orbitals

Using the Löwdin orthogonalization method [76,77], we can define OAO as:

$$
\tilde{\varphi}_{m_1}^I(\mathbf{r}) = \sum_{Jm_2} \left(O^{-\frac{1}{2}} \right)_{m_2 m_1}^{JI} \varphi_{m_2}^J(\mathbf{r}),
\tag{19}
$$

where O is the orbital overlap matrix which is defined as:

$$
(O)_{m_1 m_2}^{IJ} = \langle \varphi_{m_1}^I | \hat{S} | \varphi_{m_2}^J \rangle,
\tag{20}
$$

where $(O)_{m_1 m_2}^{IJ}$ is a matrix element of O. Note, $(O)_{m_1 m_2}^{IJ}$ can be represented as a $N \times N$ matrix by merging indices I with m_1, and J with m_2, where N is the number of all states in the system. Therefore, in the following we will refer to $(O)_{m_1 m_2}^{IJ}$ as a matrix. It is important to note that the following indices must be understood as being in couples, (I, m_1) and (J, m_2), because for different types of atoms the indices m_1 and m_2 run over different number of states. As was already anticipated in

Sec. II B, here we orthogonalize *all* states of *all* atoms in the system. It is important to orthogonalize not only states that belong to the chosen Hubbard manifolds of each atom (e.g., d or f states), but also the remaining states, in order to preserve the onsite orthogonality.

The orbital overlap matrix O is Hermitian and positive definite, therefore we can represent it as

$$\mathsf{O} = \mathsf{U}\mathsf{Z}\mathsf{U}^\dagger, \tag{21}$$

where Z is the diagonal $N \times N$ matrix composed of eigenvalues $\{z_m^I\}$ that are all positive, and U is the unitary $N \times N$ matrix ($\mathsf{U}\mathsf{U}^\dagger = \mathsf{1}$, where $\mathsf{1}$ is the unit matrix) which is formed by the corresponding eigenvectors. Using Eq. (21), we can easily compute $\mathsf{O}^{-\frac{1}{2}}$, which appears in Eq. (19), as:

$$\mathsf{O}^{-\frac{1}{2}} = \mathsf{U}\mathsf{Z}^{-\frac{1}{2}}\mathsf{U}^\dagger, \tag{22}$$

where $\mathsf{Z}^{-\frac{1}{2}}$ is the diagonal $N \times N$ matrix composed of inverse square root of eigenvalues, namely $\{(z_m^I)^{-\frac{1}{2}}\}$.

2. Derivatives of orthogonalized atomic orbitals

When we replace $\varphi_m^I(\mathbf{r})$ by $\tilde{\varphi}_m^I(\mathbf{r})$ in Eq. (16), we need to compute $\frac{\partial \tilde{\varphi}_{m_1}^I}{\partial \mathbf{R}_K}$. According to the definition (19), we obtain:

$$\frac{\partial \tilde{\varphi}_{m_1}^I(\mathbf{r})}{\partial \mathbf{R}_K} = \sum_{Jm_2} \frac{\partial}{\partial \mathbf{R}_K}\left[\left(\mathsf{O}^{-\frac{1}{2}}\right)_{m_2 m_1}^{JI}\right]\varphi_{m_2}^J(\mathbf{r})$$
$$+ \sum_{Jm_2} \left(\mathsf{O}^{-\frac{1}{2}}\right)_{m_2 m_1}^{JI} \frac{\partial \varphi_{m_2}^J(\mathbf{r})}{\partial \mathbf{R}_K}. \tag{23}$$

The second term in Eq. (23) is easy to compute; it is different from zero only when $J = K$ because φ functions are atom centered and hence $\frac{\partial \varphi_{m_2}^J(\mathbf{r})}{\partial \mathbf{R}_K} = \delta_{JK}\frac{\partial \varphi_{m_2}^K(\mathbf{r})}{\partial \mathbf{R}_K}$ [71,72]. It is worth to note that this second term is nonzero even when $K \neq I$ but $K = J$, while in the case of NAO instead of Eq. (23) there is only $\frac{\partial \varphi_{m_1}^I(\mathbf{r})}{\partial \mathbf{R}_K}$ and it equals to zero when $K \neq I$. Such a difference between OAO and NAO is important, because there are nonzero contributions to Hubbard forces for the former but not for the later when $K \neq I$. Now let us consider the first term in Eq. (23). This term is different from zero even when $K \neq J \neq I$ due to the derivative of $\mathsf{O}^{-\frac{1}{2}}$ which has nonzero off-diagonal terms. Therefore, in the case of OAO, both terms in Eq. (23) are responsible for nonzero Hubbard forces on non-Hubbard atoms (even in the case of NC PPs) in addition to another nonzero contribution due to the use of US or PAW PPs [see discussion after Eq. (18)].

The first term in Eq. (23) is the most challenging part and is the main focus of this paper. Indeed, at the first glance it is not trivial how to evaluate *exactly* the derivative of the inverse square root of the orbital overlap matrix

$$\frac{\partial}{\partial \mathbf{R}_K}\left[\left(\mathsf{O}^{-\frac{1}{2}}\right)_{m_2 m_1}^{JI}\right]. \tag{24}$$

In the following we discuss how this object was treated in literature so far, and we present a detailed derivation of the exact solution and its first application in the framework of DFT with extended Hubbard functionals for the evaluation of Hubbard forces (and other first-order derivatives).

3. Derivative of $\mathsf{O}^{-\frac{1}{2}}$ as a solution of the Lyapunov equation

In this section we discuss how to compute the exact derivative of $\mathsf{O}^{-\frac{1}{2}}$. This problem was already addressed in literature, and existing solutions will be briefly discussed in the following.

In Ref. [74], the authors used a representation in which the overlap matrix is close to diagonal, and hence in this case it is straightforward to evaluate the derivative in Eq. (24) (see Eq. (14) in Ref. [74]). Essentially, this approximation neglects off-diagonal matrix elements by assuming that they are small, which makes the calculation of the derivatives of its powers straightforward. Such an approximation turns out to be quite good in the context of Hubbard forces using Wannier functions for systems considered in Ref. [74], however noticeable deviations were observed with respect to forces computed using finite differences. These deviations can become problematic when performing structural optimizations, where the mismatch between forces and gradients of energy can lead to instabilities of the minimization algorithms, such as, e.g., the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [90]. Therefore, more accurate ways of computing Hubbard forces are desired. While systematic improvements of the accuracy of approximate Hubbard forces is possible (see Appendix A), the exact solution is obviously desired.

In Ref. [91], the exact analytical formula for computing the derivative of $\mathsf{O}^{\frac{1}{2}}$ is presented in the context of constrained DFT to explore the diabatic potential energy curves in the Marcus theory of electron transfer (see Eq. (11) in Ref. [91]). Here, we are interested in the derivative of $\mathsf{O}^{-\frac{1}{2}}$, and hence the method of Ref. [91] can be easily adapted to the current problem (the difference is just the sign in the power). The derivation presented in this paper was developed independently from Ref. [91] and, at variance with what is done in that work (that only gives the final formula) will be presented in full detail [92]. Authors are also aware of another independent exact derivation of the same solution [93] as us and Ref. [91]. This work is further motivated by the fact that it provides the first use of the analytical expression of the derivative of $\mathsf{O}^{-\frac{1}{2}}$ in the context of Hubbard-corrected DFT.

In the following, we discuss in detail the derivation of the exact analytical formula for evaluating Eq. (24). Taking the derivative of both members of the identity $\mathsf{O}^{-\frac{1}{2}}\mathsf{O}^{-\frac{1}{2}} = \mathsf{O}^{-1}$, it is easy to find:

$$\mathsf{O}^{-\frac{1}{2}}\frac{\partial \mathsf{O}^{-\frac{1}{2}}}{\partial \mathbf{R}_K} + \frac{\partial \mathsf{O}^{-\frac{1}{2}}}{\partial \mathbf{R}_K}\mathsf{O}^{-\frac{1}{2}} = \mathsf{W}_K, \tag{25}$$

where

$$\mathsf{W}_K \equiv \frac{\partial \mathsf{O}^{-1}}{\partial \mathbf{R}_K} = -\mathsf{O}^{-1}\frac{\partial \mathsf{O}}{\partial \mathbf{R}_K}\mathsf{O}^{-1}. \tag{26}$$

Equation (26) was obtained by taking a derivative of the identity $\mathsf{O}\mathsf{O}^{-1} = \mathsf{1}$, and then by multiplying the resulting equation by O^{-1} on the left-hand side.

Equation (25) can be identified as a type of *Lyapunov equation* [94], or as a particular case of the *Sylvester equation* [95]. Numerical solutions to Eq. (25) can be obtained with the help of the various algorithms as variants of the classical Bartels-Stewart algorithm [95] requiring a QR factorization of

the matrix $O^{-\frac{1}{2}}$. Importantly, these algorithms do not require the diagonalization of $O^{-\frac{1}{2}}$ which can be computationally very expensive for matrices of large size. Unfortunately, all implementations available at the moment as part of standard linear algebra libraries are very limited in terms of size of the matrix $O^{-\frac{1}{2}}$, and therefore of little (if any) use in applications to electronic structure problems. However, for all cases in which one can assume that the diagonalization of the matrices does not introduce a major computational bottleneck, an exact solution to Eq. (25) can be obtained through a closed-form expression [91]. As shown in Appendix B, under conditions fulfilled by the eigenvalues of the matrix O, we can write the formal solution to Eq. (25) as follows:

$$\frac{\partial O^{-\frac{1}{2}}}{\partial \mathbf{R}_K} = \int_0^\infty e^{-t\, O^{-\frac{1}{2}}} W_K\, e^{-t\, O^{-\frac{1}{2}}}\, dt, \qquad (27)$$

where t is an auxiliary (scalar) integration variable. Using Eq. (22), we can rewrite Eq. (27) as:

$$\frac{\partial O^{-\frac{1}{2}}}{\partial \mathbf{R}_K} = U\left(\int_0^\infty e^{-t\, Z^{-\frac{1}{2}}} \tilde{W}_K\, e^{-t\, Z^{-\frac{1}{2}}}\, dt\right)U^\dagger, \qquad (28)$$

where

$$\tilde{W}_K = U^\dagger W_K U. \qquad (29)$$

We can write even more useful result by actually computing the integral in Eq. (28). The matrix element of the term in brackets of Eq. (28) reads:

$$\left(\int_0^\infty e^{-t\, Z^{-\frac{1}{2}}} \tilde{W}_K\, e^{-t\, Z^{-\frac{1}{2}}}\, dt\right)^{IJ}_{m_1 m_2}$$
$$= \int_0^\infty e^{-t\,(z^I_{m_1})^{-\frac{1}{2}}} (\tilde{W}_K)^{IJ}_{m_1 m_2}\, e^{-t\,(z^J_{m_2})^{-\frac{1}{2}}}\, dt$$
$$= \frac{(\tilde{W}_K)^{IJ}_{m_1 m_2}}{\left(z^I_{m_1}\right)^{-\frac{1}{2}} + \left(z^J_{m_2}\right)^{-\frac{1}{2}}}. \qquad (30)$$

We note that a solution given by Eqs. (28)–(30) corresponds to a similar result in Ref. [91]. Using Eqs. (28) and (30), we obtain:

$$\left(\frac{\partial O^{-\frac{1}{2}}}{\partial \mathbf{R}_K}\right)^{IJ}_{m_1 m_2} = \sum_{Lm_3, Mm_4} \frac{(U)^{IL}_{m_1 m_3}(\tilde{W}_K)^{LM}_{m_3 m_4}(U^\dagger)^{MJ}_{m_4 m_2}}{\left(z^L_{m_3}\right)^{-\frac{1}{2}} + \left(z^M_{m_4}\right)^{-\frac{1}{2}}}. \qquad (31)$$

Finally, using Eqs. (26) and (29), and the fact that $O^{-1} = UZ^{-1}U^\dagger$, Eq. (31) can be rewritten in the compact form:

$$\left(\frac{\partial O^{-\frac{1}{2}}}{\partial \mathbf{R}_K}\right)^{IJ}_{m_1 m_2} = -\sum_{Lm_3, Mm_4} \frac{(U)^{IL}_{m_1 m_3}\left(U^\dagger \frac{\partial O}{\partial \mathbf{R}_K} U\right)^{LM}_{m_3 m_4}(U^\dagger)^{MJ}_{m_4 m_2}}{z^L_{m_3}\left(z^M_{m_4}\right)^{-\frac{1}{2}} + z^M_{m_4}\left(z^L_{m_3}\right)^{-\frac{1}{2}}}. \qquad (32)$$

Equation (32) is exact, and it constitutes a closed form of the solution of the Lyapunov equation (25). It can be seen from Eq. (32) that the only ingredients that are needed to compute the derivative of $O^{-\frac{1}{2}}$ are the eigenvalues and eigenvectors of the overlap matrix O, and the derivative of this matrix, $\frac{\partial O}{\partial \mathbf{R}_K}$. The latter can be easily computed using the definition of the

overlap matrix (20):

$$\frac{\partial O}{\partial \mathbf{R}_K} = \left\langle \frac{\partial \varphi^I_{m_1}}{\partial \mathbf{R}_K}\middle|\hat{S}\middle|\varphi^J_{m_2}\right\rangle + \left\langle \varphi^I_{m_1}\middle|\hat{S}\middle|\frac{\partial \varphi^J_{m_2}}{\partial \mathbf{R}_K}\right\rangle + \left\langle \varphi^I_{m_1}\middle|\frac{\partial \hat{S}}{\partial \mathbf{R}_K}\middle|\varphi^J_{m_2}\right\rangle. \qquad (33)$$

The first two terms in Eq. (33) are computed using the localized character of atomic orbitals:

$$\left\langle \frac{\partial \varphi^I_{m_1}}{\partial \mathbf{R}_K}\middle|\hat{S}\middle|\varphi^J_{m_2}\right\rangle + \left\langle \varphi^I_{m_1}\middle|\hat{S}\middle|\frac{\partial \varphi^J_{m_2}}{\partial \mathbf{R}_K}\right\rangle$$
$$= \delta_{IK}\left\langle \frac{\partial \varphi^K_{m_1}}{\partial \mathbf{R}_K}\middle|\hat{S}\middle|\varphi^J_{m_2}\right\rangle + \delta_{JK}\left\langle \varphi^I_{m_1}\middle|\hat{S}\middle|\frac{\partial \varphi^K_{m_2}}{\partial \mathbf{R}_K}\right\rangle, \qquad (34)$$

and the last term in Eq. (33) is computed in a way similar to Eq. (18).

It is worth to point out that the method presented here for the calculation of the first derivative of $O^{-\frac{1}{2}}$ can be generalized to higher-order derivatives. In particular, the second derivative of $O^{-\frac{1}{2}}$ can be computed by differentiating Eq. (25) and by writing a solution of the resulting equation in a closed form similarly to Eq. (27). As an example, such a computation will be useful for the generalization of the density-functional perturbation theory with the Hubbard U correction (the so-called DFPT+U approach) for calculation of phonons, which requires second-order derivatives of the occupation matrix for the calculation of the matrix of interatomic force constants [60].

Finally, the computational scaling of Hubbard forces using OAO is compared to that using NAO and is presented in Appendix C.

III. TECHNICAL DETAILS

The formalism for computing Hubbard forces presented in Sec. II has been implemented in the QUANTUM ESPRESSO distribution [96–98], and it is publicly available to the community [99]. In order to benchmark this implementation, we consider the case of NiO. In this section we review the technical settings of these calculations.

We have used the experimental lattice parameter $a = 4.17$ Å for the rock-salt crystal structure of antiferromagnetic NiO [100]. We have used the GGA for the exchange-correlation functional constructed with the PBEsol prescription [101]. Pseudopotentials were taken from the SSSP library 1.1 efficiency [102,103]: for Ni we have used the US PP from the GBRV library 1.4 [104] (ni_pbesol_v1.4.uspp.F.UPF), and for O we have used the PAW PP from the Pslibrary library 0.3.1 [105] (O.pbesol-n-kjpaw_psl.0.1.UPF). KS wave functions and potentials were expanded in plane waves using the kinetic-energy cutoffs of 80 and 640 Ry, respectively. The Brillouin zone has been sampled with a uniform $8 \times 8 \times 8$ **k** point mesh centered at the Γ point. The accuracy on the computed forces is better than 10^{-5} Ry/bohr.

The Hubbard parameters were computed using the HP code which is based on DFPT [59,82]. We have used the **k** and **q** point meshes of size $8 \times 8 \times 8$ and $5 \times 5 \times 5$, respectively, which give an accuracy of 0.01 eV for the computed values of U and V. Hubbard parameters were computed using

the self-consistent procedure, which is described in detail in Refs. [32,79,82]. In the framework of DFT+U+V, we obtained the following values: using OAO, $U_{OAO} = 7.43$ eV for Ni($3d$) states, and $V_{OAO} = 0.37$ eV between Ni($3d$) and O($2p$) states; using NAO, $U_{NAO} = 6.76$ eV and $V_{NAO} = 1.25$ eV.

For the benchmarking purposes, we evaluated the total force using numerical differentiation (finite differences) of the total energy. We used the symmetric difference quotient, with the atomic displacement parameter of 5×10^{-3} bohr along each Cartesian direction.

The data used to produce the results of this work are available at the Materials Cloud Archive [106].

IV. RESULTS

We now proceed to validate the correctness of implementation of Hubbard forces with OAO by considering NiO as a test case, as it is one of the most studied materials using the DFT + Hubbard scheme, and several theoretical and experimental studies are available in literature. Here we focus on the DFT+U+V case, which is the most general one; similar trends and results were obtained also for the DFT+U case, but they will not be discussed in the following.

Below a Neél temperature of 523 K, NiO has an antiferromagnetic ordering of type II (AFII), where ferromagnetic (111) Ni planes alternate with opposite magnetization along the [111] direction [28]. This magnetic ordering is compatible with rhombohedral symmetry, and hence the crystal structure can be modelled using a primitive cell with four atoms, which we label as Ni_1, Ni_2, O_1, and O_2. To impose the AFII magnetic ordering, we assign to Ni_1 a spin up polarization and for Ni_2 a spin down one, while O_1 and O_2 have zero net spin polarization. The atomic positions are as follows: Ni_1 is at $(0, 0, 0)$, Ni_2 is at $(1/2, 1/2, 0)$, O_1 is at $(1/2, 0, 0)$, and O_2 is at $(1/2, 1/2, 1/2)$ in the Cartesian framework in units of the lattice parameter a (see Ref. [58] for more details). Since all atoms sit in high-symmetry positions, at equilibrium the forces are zero on all atoms and in all directions. This is an ideal test case for the current study, because now we can displace atoms and compute nonzero forces acting on them, and thus we can benchmark the accuracy of our implementation of total and Hubbard forces with OAO.

Before we proceed with the calculation of forces, for the sake of completeness we summarize here basic quantities obtained in our study with DFT+U+V and OAO. Using the self-consistent Hubbard parameters reported in Sec. III, we obtain a band gap of 3.14 eV, and the magnetic moments of Ni of 1.72 μ_B—these values are in good agreement with previous theoretical and experimental studies [28]. We will not elaborate more on this, since this is not the focus of the current study.

A. Benchmark of the total force using OAO

In this section we validate the implementation of the Hubbard forces [Eq. (13)] using OAO by comparing the total force [Eq. (12)] computed using our analytical formulas and by using finite differences of the total energy. Our benchmark procedure is along the same lines as in Refs. [74,75,88].

Figure 1 shows the total energy and total force acting on the Ni_1 atom as a function of the magnitude of its displace-
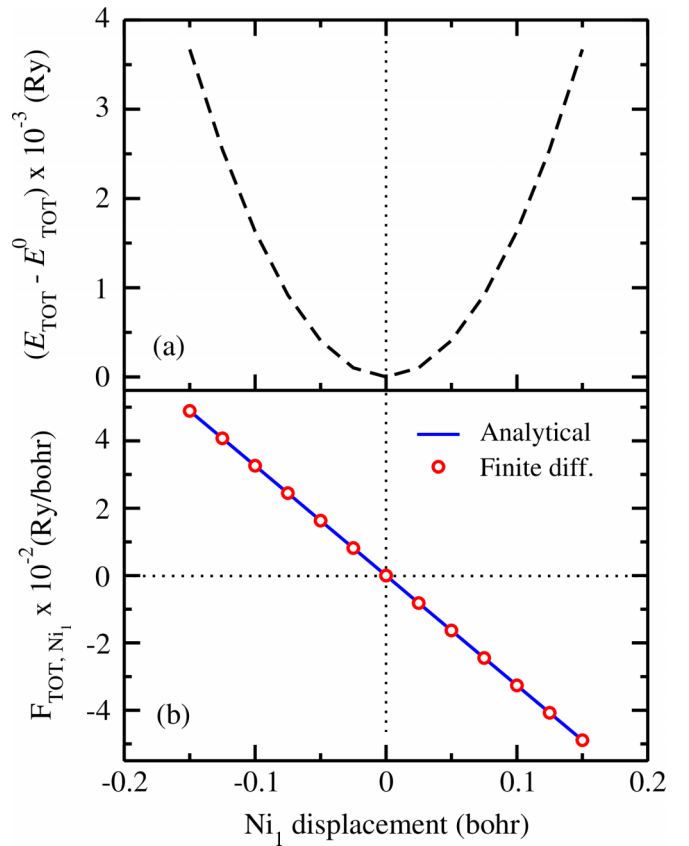

FIG. 1. (a) Total energy E_{TOT} [see Eq. (1)] as a function of the displacement on Ni_1 atom along the [111] direction. E_{TOT}^0 is the total energy computed at zero displacement of the Ni_1 atom. (b) Total force F_{TOT,Ni_1} acting on the Ni_1 atom when it is displaced along the [111] direction from its high-symmetry position (0,0,0). Blue line represents the total force computed using the analytical expression given by Eq. (12), and red empty circles represent the total force computed using finite differences of the total energy shown in panel (a). Data on panels (a) and (b) were obtained using OAO and using Hubbard parameters U_{OAO} and V_{OAO}.

ment along the [111] direction from its equilibrium position (0,0,0), obtained using OAO. The components of the force are equal along three Cartesian directions due to symmetry. The total force is computed as a square root of squared Cartesian components, and we also preserve the sign of the total force in order to highlight its direction. The "analytical" total force (blue solid line) was computed using Eq. (12), i.e., the standard DFT force plus the Hubbard force using OAO as defined in Eq. (13), which is based on the exact calculation of various derivatives and in particular that of the $O^{-\frac{1}{2}}$ matrix as discussed in Sec. II E 3. We also computed the total force using numerical differentiation (finite differences) of the total energy [see Fig. 1(a)] and the result is shown in Fig. 1(b) (red empty circles). It is easy to see from Fig. 1(b) that the analytical and numerical differentiation give exactly the same total force for various magnitudes of the displacement of Ni_1 atom, which validates the correctness of formulas and of their implementation.

From Fig. 1(b) it is difficult to extract information about the precision of agreement between the analytical and numerical

TABLE I. Comparison of the total force (in Ry/bohr) acting on atoms upon a displacement of the Ni_1 atom by $+0.15$ bohr along the [111] direction. The results are presented for the case of finite differences of the total energy and analytical implementation of forces using OAO and using Hubbard parameters U_{OAO} and V_{OAO}.

Atom	$F_{\text{finite diff.}}^{\text{OAO}}$	$F_{\text{analytical}}^{\text{OAO}}$
Ni_1	-0.048891	-0.048891
Ni_2	0.026303	0.026303
O_1	0.012951	0.012950
O_2	0.009639	0.009638

solutions (we just visually see that the results are on top of each other). In fact, in Ref. [74] the authors also considered NiO as a test case to benchmark the implementation of Hubbard forces, and they used an approximate expression for the derivative of $O^{-\frac{1}{2}}$ in the framework of Wannier functions; although the results were very good, the agreement between analytical and numerical forces was not perfect. Here, in order to elaborate more on the accuracy of our formalism, we present a comparison of the analytical and numerical results for the total force for the largest displacement considered here ($+0.15$ bohr). From Table I we can see that the total forces agree with the accuracy of 10^{-6} Ry/bohr, which is remarkable and further validates the correctness and high accuracy of our implementation of analytical formulas for the total and Hubbard forces. As a side note, we see from Table I that the total force acting on Ni_1 is negative, while total forces acting on all other atoms are all positive; this means that all atoms (except Ni_1) counteract to the displaced Ni_1 atom, and the sum of all forces is zero.

B. Hubbard contribution to the total force and the importance of the derivative of $O^{-\frac{1}{2}}$

In this section we discuss the contribution of the Hubbard force to the total force computed using analytical expressions given by Eqs. (12) and (13), using OAO as a localized basis set. Moreover, we investigate the importance of taking into account the derivative of $O^{-\frac{1}{2}}$ in Eq. (23).

Table II presents forces computed using the exact (full) expression for the derivative of OAO given by Eq. (23), and using an approximate expression which neglects entirely the derivative of $O^{-\frac{1}{2}}$ in Eq. (23) (i.e., the first term). In both cases

Hubbard parameters U_{OAO} and V_{OAO} were used. We discuss first the exact forces and then the approximate ones.

As can be seen from Table II (second and third columns), there are nonzero Hubbard forces acting on all atoms when we displace Ni_1. As was already discussed in Sec. II, the nonzero Hubbard forces on non-Hubbard atoms O_1 and O_2 is a consequence of two factors: (i) the use of US and PAW PPs [see Eq. (18)], and (ii) the use of OAO [see Eq. (23) and the discussion after this equation]. Interestingly, the Hubbard force acting on Ni_2 is one order of magnitude larger than the Hubbard force acting on the displaced atom Ni_1. Moreover, the signs of Hubbard forces acting on different atoms are different and it is not trivial to guess them. These two latter observations are a consequence of a complex interplay between different contributions entering in the definition of the Hubbard force [see in particular Eqs. (18) and (23)]. Now let us compare the magnitude of the Hubbard forces with respect to the total forces. For Ni_1 the Hubbard force constitutes 3.4% of the total force, for Ni_2 it is 38.0%, for O_1 it is 1.6%, and for O_2 it is 0.5%. Therefore, while Hubbard forces are quite negligible for some atoms (e.g., O_2), for other atoms the Hubbard contribution to the total force is very large (e.g., Ni_2), and hence Hubbard forces must always be computed and added to the total force.

In this work, large effort was dedicated to the derivation of the exact expression for the derivative of $O^{-\frac{1}{2}}$, and therefore it would be interesting and instructive to investigate the importance of this contribution in Eq. (23), and thus its significance for the Hubbard and total forces. In Table II (fourth and fifth columns) we show the Hubbard and total forces acting on atoms computed by entirely neglecting the derivative of $O^{-\frac{1}{2}}$. We can see that the total force is changed with respect to the exact total force (second column in Table II) by 2.9% for Ni_1, 1.3% for Ni_2, 3.7% for O_1, and 6.1% for O_2. These are rather significant deviations. Moreover, such errors of several percentages on total forces may lead to instabilities of structural optimization algorithms (such as BFGS [90]) which are based on the comparison of the total forces and gradients of total energy. Hence, the contribution from the derivative of $O^{-\frac{1}{2}}$ is relevant and must be always included. The Hubbard forces are affected markedly if we neglect the contribution from the derivative of $O^{-\frac{1}{2}}$: The largest deviation is for O_2 and there is a factor of ~ 12 difference with respect to the exact Hubbard force, while the smallest change is for Ni_2 and it is 3.5%. This finding underlines once more time the importance of including

TABLE II. Total and Hubbard forces (in Ry/bohr) computed using analytical formulas Eqs. (12) and (13) for the Ni_1 atom displacement of $+0.15$ bohr along the [111] direction. The second and third columns correspond to exact forces computed using Eq. (23), while the fourth and fifth columns correspond to approximate forces computed by neglecting the first term in Eq. (23). In all cases OAO were used for the calculation of forces together with the Hubbard parameters U_{OAO} and V_{OAO}.

Atom	$F_{\text{analytical}}^{\text{OAO}}$ (exact)		$F_{\text{analytical}}^{\text{OAO}}$ (approximate)	
	Total force	Hubbard force	Total force	Hubbard force
Ni_1	-0.048891	0.001652	-0.050309	0.000234
Ni_2	0.026303	-0.010003	0.026656	-0.009650
O_1	0.012950	0.000204	0.013427	0.000680
O_2	0.009638	0.000050	0.010226	0.000638

TABLE III. Total and Hubbard forces (in Ry/bohr) computed using analytical formulas Eqs. (12) and (13) for the Ni_1 atom displacement of $+0.15$ bohr along the [111] direction. The second and third columns correspond to forces computed using OAO with U_{OAO} and V_{OAO}, the fourth and fifth columns correspond to forces computed using NAO with U_{NAO} and V_{NAO}, and the sixth and seventh columns correspond to forces computed using NAO with U_{OAO} and V_{OAO}.

Atom	$F_{analytical}^{OAO}$ (U_{OAO} & V_{OAO})		$F_{analytical}^{NAO}$ (U_{NAO} & V_{NAO})		$F_{analytical}^{NAO}$ (U_{OAO} & V_{OAO})	
	Total force	Hubbard force	Total force	Hubbard force	Total force	Hubbard force
Ni_1	-0.048891	0.001652	-0.043358	0.004127	-0.046051	0.004006
Ni_2	0.026303	-0.010003	0.023360	-0.009405	0.024727	-0.010772
O_1	0.012950	0.000204	0.011567	-0.000117	0.012207	-0.000056
O_2	0.009638	0.000050	0.008431	-0.000101	0.009117	-0.000032

the derivative of $O^{-\frac{1}{2}}$ in the formalism. For some applications (e.g., linear scaling algorithms) it might be of interest to use some approximate ways in computing the derivative of $O^{-\frac{1}{2}}$ (such as the one described in Ref. [74] or in Appendix A), but the application of the current formulation for the exact evaluation of $O^{-\frac{1}{2}}$ as a solution of the Lyapunov equation is feasible and convenient for quite large systems (from several tens to a few hundreds of atoms).

C. OAO versus NAO forces

In this section we present a comparison of the total and Hubbard forces computed using OAO and NAO. As was mentioned in the introduction (Sec. I), it is important to keep the consistency of the Hubbard manifolds which are used for the calculation of Hubbard parameters, total energy, forces, and various other properties. Table III presents a comparison for the fully consistent cases of OAO and NAO (i.e., when forces and Hubbard parameters are computed by using consistently the same Hubbard manifold), and for the mixed case when NAO are used to compute forces while Hubbard parameters were obtained using OAO. We stress that this latter case is presented only for demonstrative purposes, while in practice full consistency of Hubbard manifolds should be used.

As can be seen from Table III the differences between fully consistent total forces obtained using OAO and NAO (second and fourth columns) vary in the range from 11% to 14%. This is a rather significant difference, which might be very relevant when optimizing atomic positions for complex transition-metal oxides. As expected, the Hubbard forces in these two cases are very different (see columns three and five in Table III): Not only the absolute values of Hubbard forces differ in OAO and NAO, but even the sign of Hubbard forces for some atoms (O_1 and O_2 in this case) are opposite to each other [108]. This latter finding is obviously related to the extra contributions coming from the $O^{-\frac{1}{2}}$ matrix in Eq. (23) when using OAO.

The origin of the difference between fully consistent total forces is twofold: (i) different Hubbard parameters are used in two calculations, (ii) different Hubbard manifolds are used when computing total energies and forces. In order to disentangle these two effects, we performed an additional calculation: total energies and forces are computed using NAO, while Hubbard parameters are obtained using OAO (see last two columns in Table III). We can see that in this case the difference between OAO and NAO total forces (second and sixth

columns) is reduced down to 5–6%. Therefore, such a residual difference is due to the use of different Hubbard manifolds for computing total energies and forces. Furthermore, we can see that in this mixed case also the Hubbard forces are changed in such a way that they become closer to the fully consistent OAO Hubbard forces (compare columns three, five, and seven in Table III).

Therefore, such a comparative analysis highlights that it is important to keep the full consistency when computing Hubbard parameters, total energies, forces, and other properties of materials. Most importantly, such a consistency must be used from the conceptual point of view: Mixing different Hubbard manifolds is not justified, and ultimately it can lead to unpredictable behavior of DFT+U and DFT+U+V simulations.

Finally, it would be very useful and important to compare various materials' properties using NAO and OAO, and in particular compare atomic positions after structural optimization with available x-ray diffraction data. This task requires systematic study for various types of materials which have different types of interactions (ionic, covalent, or mixed) to see trends, but this is not the objective of this paper and it is beyond its scope. However, we want to stress that such a comparison of NAO and OAO is possible now thanks to the current work, and in fact there are already ongoing efforts the results of which will be presented elsewhere.

V. CONCLUSIONS

We have presented a detailed derivation for the exact expression of the contribution to Pulay forces in the framework of extended Hubbard functionals (DFT+U+V [28]) that originate from the use of orthogonalized atomic orbitals. At variance with the use of nonorthogonalized orbitals, a major difficulty arises in this case due to the need to compute the derivative of the inverse square root of the orbital overlap matrix $O^{-\frac{1}{2}}$. Similarly to Ref. [91], we have developed and presented in full detail the derivation of a closed form expression for the derivative of $O^{-\frac{1}{2}}$ via an integral representation of the solution of the associated Lyapunov equation, where, for the sake of generality, all equations are written for the case of ultrasoft pseudopotentials and projector-augmented-wave method.

The implementation of Hubbard forces using orthogonalized atomic orbitals is benchmarked versus finite differences of total energies for NiO, and excellent agreement between forces is obtained, which validates the correctness of the analytical formulas and of the implementation. In addition, we

show that the contribution from the derivative of $O^{-\frac{1}{2}}$ to the total force is quite significant (from 1% to 6% in the case of NiO), and hence should always be included for the sake of consistency and high accuracy.

Furthermore, we have compared total and Hubbard forces computed using orthogonalized and nonorthogonalized atomic orbitals using Hubbard parameters computed with the respective set of orbitals (i.e., using respective Hubbard manifolds). We have found that the differences in total forces is as large as 11–14% in NiO, which is in part related to the use of different Hubbard parameters and in part due to the use of different Hubbard manifolds when computing Hubbard and total forces. This finding highlights the importance of maintaining the consistency of the Hubbard manifold and of the Hubbard parameters when computing various materials' properties such as total energy, forces, stress, and phonons, to name a few.

Finally, the current formalism for computing Hubbard forces using orthogonalized atomic orbitals has been implemented in the open-source QUANTUM ESPRESSO distribution [96–98] and is freely available to the community at large. The computational scaling of this formalism is higher than that when using nonorthogonalized atomic orbitals, however through the use of standardized mathematical libraries (BLAS and LAPACK [109]) and effective parallelization strategies (across plane waves, **k** points, and bands) large speedups are achieved that make the overhead of no concern. Last, we believe that this work opens up avenues for very accurate geometry optimizations for transition-metal and rare-earth compounds by taking into account complex hybridization effects between neighboring sites via the orbital overlap matrix and subsequently using it in the framework of DFT+U+V which has proven to be very effective in complex materials [26,32,33].

APPENDIX A: DERIVATIVE OF $O^{-\frac{1}{2}}$ VIA TAYLOR SERIES

In this Appendix we present an alternative method for the calculation of the derivative of $O^{-\frac{1}{2}}$, which is based on the expansion of O in Taylor series and then computing its derivative. At variance with the exact solution presented in Sec. II E 3, the use of a Taylor series is by its nature approximate and, thus, it requires convergence checks with respect to the

number of terms to be included in the series. However, this approach is still attractive in cases where the matrix O is very large, as means to avoid the otherwise expensive diagonalization. This is a typical situation for linear-scaling quantum chemistry algorithms [110], as well as DFT codes that exploit extensively sparse linear algebra for calculations on large systems [111]. In this case, the use of localized basis functions leads to diagonally dominant overlap matrices, and therefore the Taylor series can be designed with optimal convergence properties, as explained in what follows. We start by writing the overlap matrix [defined in Eq. (20)] as:

$$O = \mu\left(\frac{O}{\mu}\right) = \mu(1 + A), \tag{A1}$$

where

$$A = \frac{O}{\mu} - 1. \tag{A2}$$

Here $\mu = \max[z_m^I]$ is the maximum eigenvalue of O [see Eq. (21)], 1 is the unit matrix, and A is an auxiliary matrix with a spectral radius (i.e., the magnitude of its largest eigenvalue) strictly smaller than 1. The representation of O given by Eqs. (A1) and (A2) is particularly useful for situations when the orbitals are not normalized to 1.

The Taylor series for $O^{-\frac{1}{2}}$ in terms of A can then be obtained using Eq. (A1), namely

$$\begin{aligned} O^{-\frac{1}{2}} &= \mu^{-\frac{1}{2}}(1 + A)^{-\frac{1}{2}} \\ &= \mu^{-\frac{1}{2}}\left(1 + \sum_{n=1}^{\infty}\prod_{k=1}^{n}\frac{-\frac{1}{2} - k + 1}{k}A^n\right) \\ &= \mu^{-\frac{1}{2}}\left(1 - \frac{1}{2}A + \frac{3}{8}AA + ...\right). \end{aligned} \tag{A3}$$

The desired derivative of $O^{-\frac{1}{2}}$ thus reads:

$$\frac{\partial O^{-\frac{1}{2}}}{\partial \mathbf{R}_K} = \mu^{-\frac{1}{2}}\left(-\frac{1}{2}\frac{\partial A}{\partial \mathbf{R}_K} + \frac{3}{8}\left[\frac{\partial A}{\partial \mathbf{R}_K}A + A\frac{\partial A}{\partial \mathbf{R}_K}\right] + ...\right). \tag{A4}$$

It is important to note that in Eq. (A4) matrices A and $\frac{\partial A}{\partial \mathbf{R}_K}$ may not in general commute (as well as matrices in higher order terms), and therefore it is necessary to keep the correct order of matrices when taking a derivative of Eq. (A3). To do so, Eq. (A4) can be effectively implemented via a loop over Taylor terms and using a recursive formula:

$$\frac{\partial(A^n)}{\partial \mathbf{R}_K} = \frac{\partial(AA^{n-1})}{\partial \mathbf{R}_K} = \frac{\partial A}{\partial \mathbf{R}_K}A^{n-1} + A\frac{\partial(A^{n-1})}{\partial \mathbf{R}_K}. \tag{A5}$$

Finally, the only starting inputs that are needed for the recursive formula (A5) are the matrix A [see Eq. (A2)] and its first derivative

$$\frac{\partial A}{\partial \mathbf{R}_K} = \frac{1}{\mu}\frac{\partial O}{\partial \mathbf{R}_K}, \tag{A6}$$

where $\frac{\partial O}{\partial \mathbf{R}_K}$ is defined in Eq. (33).

In the case of periodic solids, i.e., when Bloch sums of OAO or NAO are used, it turns out that this approximate approach based on the Taylor series is not effective. The main issue is that the Taylor series appearing in Eq. (A4)

converges very slowly, and thus in practice it requires several hundreds of terms for achieving accuracy of Hubbard forces comparable to that obtained when using the exact solution described in Sec. II E 3. This is so because many off-diagonal elements of the overlap matrix O are nonzero in the basis of Bloch-summed OAO or NAO. Possibly, faster convergence of series can be achieved by inverting the order of Taylor series expansion and evaluating Bloch sums. But the correctness of these assumptions requires more detailed investigations; this is beyond the scope of the present work and is left for future studies. However, in the case of finite systems with localized basis sets, or for periodic solids with optimally chosen functions for the Hubbard manifold [74], the overlap matrix O might be diagonally dominant, and therefore the Taylor series can be designed with fast convergence properties.

APPENDIX B: FORMAL SOLUTION OF LYAPUNOV EQUATION

In this Appendix we demonstrate that Eq. (27) represents an exact solution of Eq. (25). This can be verified via a direct substitution of Eq. (27) in Eq. (25), which gives:

$$
\begin{aligned}
\int_0^\infty & dt \left(O^{-\frac{1}{2}} e^{-t\,O^{-\frac{1}{2}}} \, W_K \, e^{-t\,O^{-\frac{1}{2}}} + e^{-t\,O^{-\frac{1}{2}}\,t} \, W_K \, e^{-t\,O^{-\frac{1}{2}}} O^{-\frac{1}{2}} \right) \\
&= \int_0^\infty dt \, \frac{d}{dt} \left[-e^{-t\,O^{-\frac{1}{2}}} \, W_K \, e^{-t\,O^{-\frac{1}{2}}} \right] \\
&= U \left[-e^{-t\,Z^{-\frac{1}{2}}} \, \tilde{W}_K \, e^{-t\,Z^{-\frac{1}{2}}} \right]_0^\infty U^\dagger \\
&= U \tilde{W}_K U^\dagger = W_K,
\end{aligned}
\tag{B1}
$$

where we have used the property that each exponential in Eq. (B1) decays to zero in the limit $t \to \infty$, provided that all eigenvalues of the matrix $O^{-\frac{1}{2}}$ are positive [107]. From this result, we derive the closed form presented in Eq. (30).

APPENDIX C: COMPUTATIONAL SCALING

In this Appendix we compare the scaling (to a leading order) of the calculation of Hubbard forces using NAO and OAO, in the framework of DFT+U+V. Essentially, the goal is to compare the computational cost in evaluating the derivative of occupation matrices given by Eq. (15). Below we present the estimates based on our current implementation of Hubbard forces.

In the case of NAO, the scaling is

$$
T_{\text{NAO}} \propto 3N_{\text{at}}N_{\mathbf{k}} \left[T_{\text{US/PAW}} + T_{\text{NAO}}^{\text{proj}} \right],
\tag{C1}
$$

where the factor of 3 is due to the displacements of all atoms in three Cartesian directions, N_{at} is the number of all atoms in the simulation cell, and $N_{\mathbf{k}}$ is the number of $\mathbf{k}$ point in the irreducible wedge of the Brillouin zone. $T_{\text{US/PAW}}$ is the computational cost of evaluating the US or PAW PP related term given by Eq. (18). This latter term is the same for NAO and OAO, and thus will not be analyzed further. The quantity $T_{\text{NAO}}^{\text{proj}}$ represents the computational cost of evaluating the second term in Eq. (16). The leading scaling of $T_{\text{NAO}}^{\text{proj}}$ can be estimated as:

$$
T_{\text{NAO}}^{\text{proj}} \propto N_{\text{PW}} N_\varphi^U N_{\text{bands}},
\tag{C2}
$$

where N_{PW} is the number of plane waves in the basis set (determined by the kinetic energy cutoff), N_{bands} is the number of electronic bands (KS states), and N_φ^U is the number of states in the Hubbard manifold of each Hubbard atom (for different types of Hubbard atoms the number of states varies depending on which atomic shell is considered as a Hubbard manifold). Generally, $T_{\text{NAO}}^{\text{proj}} \ll T_{\text{US/PAW}}$, and the total computational cost of evaluating Hubbard forces using NAO is negligible with respect to the cost of a self-consistent iterative solution of KS equations.

In the case of OAO, the scaling is

$$
T_{\text{OAO}} \propto 3N_{\text{at}}N_{\mathbf{k}} \left[T_{\text{US/PAW}} + T_{\text{ortho}}^{\text{proj}} + T_{\text{OAO}}^{\text{proj}} \right],
\tag{C3}
$$

where $T_{\text{US/PAW}}$ is the same as in Eq. (C1). In Eq. (C3) there is a new term ($T_{\text{ortho}}^{\text{proj}}$) that is not present in Eq. (C1), and it is related to the evaluation of the derivative of $O^{-\frac{1}{2}}$ (see Sec. II E 3). The scaling of this latter term is

$$
T_{\text{ortho}}^{\text{proj}} \propto 2N_{\text{PW}}N_\varphi^{\text{all}}N_{\text{tot}}^{\text{all}} + T_{\text{US/PAW}}^{\text{ortho}} + 4\left(N_{\text{tot}}^{\text{all}}\right)^3,
\tag{C4}
$$

where N_φ^{all} is the number of all orthogonalized atomic states per each Hubbard atom (obviously, for different atomic types N_φ^{all} is different, but here we use an average value for simplicity), $N_{\text{tot}}^{\text{all}}$ is the total number of all orthogonalized atomic states of all atoms in the system (if all atoms are of the same type then $N_{\text{tot}}^{\text{all}} = N_\varphi^{\text{all}}N_{\text{at}}$). The first term in Eq. (C4) describes the computational cost of computing the first and second terms in Eq. (33). The second term in Eq. (C4), $T_{\text{US/PAW}}^{\text{ortho}}$, is the analog of $T_{\text{US/PAW}}$, and it describes the computational cost of evaluating the last term in Eq. (33). The last term in Eq. (C4) describes the scaling of evaluating Eq. (32), which requires four sequential matrix-matrix multiplications of size $N_{\text{tot}}^{\text{all}} \times N_{\text{tot}}^{\text{all}}$ and one inexpensive linear algebra operation [division by the term in the denominator in Eq. (32)] which we neglected in the estimate of the cost. The general trend is the following: $4(N_{\text{tot}}^{\text{all}})^3 < T_{\text{US/PAW}}^{\text{ortho}} < 2N_{\text{PW}}N_\varphi^{\text{all}}N_{\text{tot}}^{\text{all}}$. Finally, the last term in Eq. (C3), $T_{\text{OAO}}^{\text{proj}}$, describes the computational cost of evaluating Eq. (23) and the second term in Eq. (16) (after replacing $\frac{\partial \varphi_{m_1}^I}{\partial \mathbf{R}_K}$ by $\frac{\partial \tilde{\varphi}_{m_1}^I}{\partial \mathbf{R}_K}$), and it can be written as:

$$
T_{\text{OAO}}^{\text{proj}} = N_{\text{at}}^U \left(1 + N_{\text{neigh}}^V\right) N_{\text{PW}} N_\varphi^U \left(N_{\text{bands}} + N_{\text{tot}}^{\text{all}} + N_\varphi^{\text{all}}\right),
\tag{C5}
$$

where N_{at}^U is the number of Hubbard atoms in the system, and N_{neigh}^V is the number of neighbors for each Hubbard atom for which the intersite Hubbard $V \neq 0$. Note that in the DFT+U case, $N_{\text{neigh}}^V = 0$ and hence the scaling in Eq. (C5) is largely reduced. By comparing Eqs. (C2) and (C5) it is easy to see that in the latter the scaling is largely increased in particular due to the prefactor $N_{\text{at}}^U (1 + N_{\text{neigh}}^V)$. This prefactor is present due to the fact that for OAO the derivative $\frac{\partial \tilde{\varphi}_{m_1}^I}{\partial \mathbf{R}_K}$ is different from zero even when $K \neq I$ [see the discussion after Eq. (23)], while for NAO the derivative $\frac{\partial \varphi_{m_1}^I}{\partial \mathbf{R}_K}$ is different from zero only when $K = I$. Overall, the general trend in Eq. (C3) is the following: $T_{\text{US/PAW}} < T_{\text{ortho}}^{\text{proj}} \ll T_{\text{OAO}}^{\text{proj}}$.

Therefore, from the discussions above it becomes obvious that the computational cost of Hubbard forces using OAO is much more expensive than that using NAO, i.e.,

$T_{NAO} \ll T_{OAO}$. This is not surprising, because the former requires extra computations related to the evaluation of the derivative of $O^{-\frac{1}{2}}$, and due to the fact that there are nonzero contributions to the Hubbard force even for non-Hubbard atoms due to the presence of the overlap matrix O. Despite such an increase in the computational cost when using OAO, still this method can be efficiently implemented using standardized mathematical libraries (BLAS and LAPACK [109]), and effectively parallelized over plane waves, **k** points, and bands.

[1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[2] W. Kohn and L. Sham, Phys. Rev. **140**, A1133 (1965).

[3] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[4] P. Mori-Sánchez, A. Cohen, and W. Yang, J. Chem. Phys. **125**, 201102 (2006).

[5] A. Svane and O. Gunnarsson, Phys. Rev. Lett. **65**, 1148 (1990).

[6] D. Vogel, P. Krüger, and J. Pollmann, Phys. Rev. B **54**, 5495 (1996).

[7] A. Filippetti and N. A. Spaldin, Phys. Rev. B **67**, 125109 (2003).

[8] I. P. R. Moreira, F. Illas, and R. L. Martin, Phys. Rev. B **65**, 155102 (2002).

[9] F. Corà, M. Alfredsson, G. Mallia, D. Middlemiss, W. Mackrodt, R. Dovesi, and R. Orlando, Struct. Bond. (Berlin) **113**, 171 (2004).

[10] X.-B. Feng and N. M. Harrison, Phys. Rev. B **69**, 035114 (2004).

[11] M. Alfredsson, G. D. Price, C. R. A. Catlow, S. C. Parker, R. Orlando, and J. P. Brodholt, Phys. Rev. B **70**, 165111 (2004).

[12] F. Tran, P. Blaha, K. Schwarz, and P. Novák, Phys. Rev. B **74**, 155108 (2006).

[13] V. L. Chevrier, S. P. Ong, R. Armiento, M. K. Y. Chan, and G. Ceder, Phys. Rev. B **82**, 075122 (2010).

[14] D.-H. Seo, A. Urban, and G. Ceder, Phys. Rev. B 92 **92**, 115118 (2015).

[15] J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, Phys. Rev. Lett. **91**, 146401 (2003).

[16] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, L. A. Constantin, and J. Sun, Phys. Rev. Lett. **103**, 026403 (2009).

[17] J. Sun, B. Xiao, and A. Ruzsinszky, J. Chem. Phys. **137**, 051101 (2012).

[18] Y. Zhao and D. Truhlar, J. Chem. Phys. **125**, 194101 (2006).

[19] J. del Campo, J. Gazquez, S. Trickey, and A. Vela, Chem. Phys. Lett. **543**, 179 (2012).

[20] J. Sun, A. Ruzsinszky, and J. P. Perdew, Phys. Rev. Lett. **115**, 036402 (2015).

[21] V. I. Anisimov, J. Zaanen, and O. K. Andersen, Phys. Rev. B **44**, 943 (1991).

[22] A. I. Liechtenstein, V. I. Anisimov, and J. Zaanen, Phys. Rev. B **52**, R5467 (1995).

[23] S. L. Dudarev, G. A. Botton, S. Y. Savrasov, C. J. Humphreys, and A. P. Sutton, Phys. Rev. B **57**, 1505 (1998).

[24] H. J. Kulik, M. Cococcioni, D. A. Scherlis, and N. Marzari, Phys. Rev. Lett. **97**, 103001 (2006).

[25] H. Kulik and N. Marzari, J. Chem. Phys. **129**, 134314 (2008).

[26] H. Kulik and N. Marzari, J. Chem. Phys. **134**, 094103 (2011).

[27] B. Himmetoglu, A. Floris, S. de Gironcoli, and M. Cococcioni, Int. J. Quant. Chem. **114**, 14 (2014).

[28] V. L. Campo Jr and M. Cococcioni, J. Phys.: Condens Matter **22**, 055602 (2010).

[29] N. Tancogne-Dejean and A. Rubio, Phys. Rev. B **102**, 155117 (2020).

[30] S.-H. Lee and Y.-W. Son, arXiv:1911.05967 (2019).

[31] V. Anisimov and Y. Izyumov, *Electronic Structure of Strongly Correlated Materials* (Springer, Berlin, 2010).

[32] M. Cococcioni and N. Marzari, Phys. Rev. Mater, **3**, 033801 (2019).

[33] C. Ricca, I. Timrov, M. Cococcioni, N. Marzari, and U. Aschauer, Phys. Rev. Res. **2**, 023313 (2020).

[34] P. H. Dederichs, S. Blügel, R. Zeller, and H. Akai, Phys. Rev. Lett. **53**, 2512 (1984).

[35] A. K. McMahan, R. M. Martin, and S. Satpathy, Phys. Rev. B **38**, 6650 (1988).

[36] O. Gunnarsson, O. K. Andersen, O. Jepsen, and J. Zaanen, Phys. Rev. B **39**, 1708 (1989).

[37] M. S. Hybertsen, M. Schlüter, and N. E. Christensen, Phys. Rev. B **39**, 9028 (1989).

[38] O. Gunnarsson, Phys. Rev. B **41**, 514 (1990).

[39] W. E. Pickett, S. C. Erwin, and E. C. Ethridge, Phys. Rev. B **58**, 1201 (1998).

[40] I. V. Solovyev and M. Imada, Phys. Rev. B **71**, 045103 (2005).

[41] K. Nakamura, R. Arita, Y. Yoshimoto, and S. Tsuneyuki, Phys. Rev. B **74**, 235113 (2006).

[42] M. Shishkin and H. Sato, Phys. Rev. B **93**, 085135 (2016).

[43] K. Nawa, T. Akiyama, T. Ito, K. Nakamura, T. Oguchi, and M. Weinert, Phys. Rev. B **97**, 035117 (2018).

[44] N. J. Mosey and E. A. Carter, Phys. Rev. B **76**, 155123 (2007).

[45] N. Mosey, P. Liao, and E. Carter, J. Chem. Phys. **129**, 014103 (2008).

[46] A. N. Andriotis, R. M. Sheetz, and M. Menon, Phys. Rev. B **81**, 245103 (2010).

[47] L. A. Agapito, S. Curtarolo, and M. Buongiorno Nardelli, Phys. Rev. X **5**, 011006 (2015).

[48] M. Springer and F. Aryasetiawan, Phys. Rev. B **57**, 4364 (1998).

[49] T. Kotani, J. Phys.: Condens. Matter **12**, 2413 (2000).

[50] F. Aryasetiawan, M. Imada, A. Georges, G. Kotliar, S. Biermann, and A. I. Lichtenstein, Phys. Rev. B **70**, 195104 (2004).

[51] F. Aryasetiawan, K. Karlsson, O. Jepsen, and U. Scönberger, Phys. Rev. B **74**, 125106 (2006).

[52] E. Sasioglu, C. Friedrich, and S. Blügel, Phys. Rev. B **83**, 121101(R) (2011).

[53] L. Vaugier, H. Jiang, and S. Biermann, Phys. Rev. B **86**, 165105 (2012).

[54] B. Amadon, T. Applencourt, and F. Bruneval, Phys. Rev. B **89**, 125110 (2014).

[55] P. Seth, P. Hansmann, A. van Roekeghem, L. Vaugier, and S. Biermann, Phys. Rev. Lett. **119**, 056401 (2017).

[56] S. K. Panda, H. Jiang, and S. Biermann, Phys. Rev. B **96**, 045137 (2017).

[57] K. Nakamura, Y. Yoshimoto, Y. Nomura, T. Tadano, M. Kawamura, T. Kosugi, K. Yoshimi, T. Misawa, and Y. Motoyam, arXiv:2001.02351.

[58] M. Cocciccioni and S. de Gironcoli, Phys. Rev. B **71**, 035105 (2005).

[59] I. Timrov, N. Marzari, and M. Cococcioni, Phys. Rev. B **98**, 085127 (2018).

[60] A. Floris, I. Timrov, B. Himmetoglu, N. Marzari, S. de Gironcoli, and M. Cococcioni, Phys. Rev. B **101**, 064305 (2020).

[61] J. Bennett, B. Hudson, I. Metz, D. Liang, S. Spurgeon, Q. Cui, and S. Mason, Comput. Mater. Sci. **170**, 109137 (2019).

[62] Y.-C. Wang, Z.-H. Chen, and H. Jiang, J. Chem. Phys. **144**, 144106 (2016).

[63] C. Tablero, J. Phys.: Condens. Matter **20**, 325205 (2008).

[64] B. Amadon, F. Jollet, and M. Torrent, Phys. Rev. B **77**, 155104 (2008).

[65] I. Timrov, P. Agrawal, X. Zhang, S. Erat, R. Liu, A. Braun, M. Cococcioni, M. Calandra, N. Marzari, and D. Passerone, Phys. Rev. Res. **2**, 033265 (2020).

[66] D. D. O'Regan, N. D. M. Hine, M. C. Payne, and A. A. Mostofi, Phys. Rev. B **82**, 081102(R) (2010).

[67] D. Korotin, V. Kukolev, A. Kozhevnikov, D. Novoselov, and V. Anisimov, J. Phys.: Condens. Matter **24**, 415603 (2012).

[68] A. B. Shick, and A. I. Liechtenstein, and W. E. Pickett, Phys. Rev. B **60**, 10763 (1999).

[69] O. Bengone, M. Alouani, P. Blöchl, and J. Hugel, Phys. Rev. B **62**, 16392 (2000).

[70] A. Rohrbach, J. Hafner, and G. Kresse, J. Phys.: Condens. Matter **15**, 979 (2003).

[71] M. Cococcioni, Rev. Mineral. Geochem. **71**, 147 (2010).

[72] M. Cococcioni, *The LDA+U Approach: A Simple Hubbard Correction for Correlated Ground States* [chapter 4 in Correlated Electrons: From Models to Materials, Vol. 2 (Verlag des Forschungszentrum, Jülich, 2012)] .

[73] S. Roychoudhury, D. D. O'Regan, and S. Sanvito, Phys. Rev. B **97**, 205120 (2018).

[74] D. Novoselov, D. Korotin, and V. Anisimov, J. Phys.: Condens. Matter **27**, 325602 (2015).

[75] F. Tran, J. Kuneš, P. Novák, P. Blaha, L. Marks, and K. Schwarz, Comput. Phys. Commun. **179**, 784 (2008).

[76] P.-O. Löwdin, J. Chem. Phys. **18**, 365 (1950).

[77] I. Mayer, Int. J. Quant. Chem. **90**, 63 (2002).

[78] F. Zhou, M. Cococcioni, C. A. Marianetti, D. Morgan, and G. Ceder, Phys. Rev. B **70**, 235121 (2004).

[79] H. Hsu, K. Umemoto, M. Cococcioni, and R. Wentzcovitch, Phys. Rev. B **79**, 125124 (2009).

[80] C. Ricca, I. Timrov, M. Cococcioni, N. Marzari, and U. Aschauer, Phys. Rev. B **99**, 094102 (2019).

[81] Y. Sun, M. Cococcioni, and R. M. Wentzcovitch, Phys. Rev. Mater. **4**, 063605 (2020).

[82] I. Timrov, N. Marzari, and M. Cococcioni, arXiv:2011.03271 (2020).

[83] A. Dal Corso, Phys. Rev. B **64**, 235118 (2001).

[84] V. Anisimov, F. Aryasetiawan, and A. Liechtenstein, J. Phys.: Condens. Matter **9**, 767 (1997).

[85] D. Vanderbilt, Phys. Rev. B **41**, 7892 (1990).

[86] Note that all indices in Eq. (7) are correct, while in Ref. [28] in the last term of Eq. (13) there must be $n_{mm'}^{IJ\sigma}$ instead of $n_{m'm}^{IJ\sigma}$ (in the actual implementation in Ref. [28] $n_{mm'}^{IJ\sigma}$ was used).

[87] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).

[88] M. Wierzbowska and J. A. Majewski, Phys. Rev. B **84**, 245129 (2011).

[89] H. Kulik and N. Marzari, J. Chem. Phys. **135**, 194105 (2011).

[90] R. Fletcher, *Practical Methods of Optimization*, 2nd ed. (Wiley, Chichester, UK, 1987).

[91] Q. Wu and T. Voorhis, J. Phys. Chem. A **110**, 9212 (2006).

[92] The authors became aware of Ref. [91] only after working out the solution of Eq. (25) and writing the present paper.

[93] E. Küçükbenli (unpublished).

[94] A. Lyapunov, *Collected Works* (Academy of Sciences of Soviet Union, Moscow, 1948).

[95] R. Bartels and G. Stewart, Commun. ACM **15**, 820 (1972).

[96] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. De Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. Seitsonen, A. Smogunov, P. Umari, and R. Wentzcovitch, J. Phys.: Condens. Matter **21**, 395502 (2009).

[97] P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. Buongiorno Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni, N. Colonna, I. Carnimeo, A. Dal Corso, S. de Gironcoli, P. Delugas, R. A. DiStasio Jr., A. Ferretti, A. Floris, G. Fratesi, G. Fugallo, R. Gebauer, U. Gerstmann, F. Giustino, T. Gorni, J. Jia, M. Kawamura, H.-Y. Ko, A. Kokalj, E. Küçükbenli, M. Lazzeri, M. Marsili, N. Marzari, F. Mauri, N. L. Nguyen, H.-V. Nguyen, A. Otero-de-la Rosa, L. Paulatto, S. Poncé, D. Rocca, R. Sabatini, B. Santra, M. Schlipf, A. Seitsonen, A. Smogunov, I. Timrov, T. Thonhauser, P. Umari, N. Vast, and S. Baroni, J. Phys.: Condens. Matter **29**, 465901 (2017).

[98] P. Giannozzi, O. Baseggio, P. Bonfà, D. Brunato, R. Car, I. Carnimeo, C. Cavazzoni, S. de Gironcoli, P. Delugas, F. Ferrari Ruffino, A. Ferretti, N. Marzari, I. Timrov, A. Urru, and S. Baroni, J. Chem. Phys. **152**, 154105 (2020).

[99] The official release of the QUANTUM ESPRESSO distribution can be downloaded from https://www.quantum-espresso.org.

[100] L. Bartel and B. Morosin, Phys. Rev. B **3**, 1039 (1971).

[101] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou, and K. Burke, Phys. Rev. Lett. **100**, 136406 (2008).

[102] G. Prandini, A. Marrazzo, I. Castelli, N. Mounet, and N. Marzari, npj Comput. Mater. **4**, 72 (2018).

[103] The SSSP library of the Materials Cloud [https://www.materialscloud.org/discover/sssp/table/efficiency].

[104] K. Garrity, J. Bennett, K. Rabe, and D. Vanderbilt, Comput. Mater. Sci. **81**, 446 (2014).

[105] E. Kucukbenli, M. Monni, B. Adetunji, X. Ge, G. Adebayo, N. Marzari, S. de Gironcoli, and A. Dal Corso, arXiv:1404.3015 (2014).

[106] I. Timrov, F. Aquilante, L. Binci, M. Cococcioni, and N. Marzari, Pulay forces in density-functional theory with extended Hubbard functionals: From nonorthogonalized to orthogonalized manifolds, Materials Cloud Archive **2020**, 132 (2020).

[107] Actually, the necessary condition for the integral to converge is less stringent, requiring only the real part of each eigenvalue of $O^{-\frac{1}{2}}$ to be positive.

[108] The Hubbard forces acting on O atoms in NiO are very small, but in cases when the Hubbard U correction is applied also

on O atoms (or to neighbor atoms in other systems) in general these forces can become significant.

[109] E. Anderson, Z. Bai, C. Bischof, L. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, *LAPACK Users' Guide*, 3rd ed. (Society for Industrial and Applied Mathematics, Philadelphia, PA, 1999).

[110] J. M. H. Olsen, S. Reine, O. Vahtras, E. Kjellgren, P. Reinholdt, K. O. Hjorth Dundas, X. Li, J. Cukras, M. Ringholm, E. D. Hedegård, R. Di Remigio, N. H. List, R. Faber, B. N. Cabral Tenorio, R. Bast, T. B. Pedersen, Z. Rinkevicius, S. P. A. Sauer, K. V. Mikkelsen, J. Kongsted, S. Coriani, K. Ruud, T. Helgaker, H. J. A. Jensen, and P. Norman, J. Chem. Phys. **152**, 214115 (2020).

[111] N. Hine, P. Haynes, A. A. Mostofi, C.-K. Skylaris, and M. Payne, Comput. Phys. Commun. **180**, 1041 (2009).