

# Optimization of Functionals of Orthonormal Functions in the Absence of Unitary Invariance

Peter Klüpfel<sup>1</sup>, Simon Klüpfel<sup>1</sup>,  
Kiril Tsemekhman<sup>2</sup>, and Hannes Jónsson<sup>1,2</sup>

<sup>1</sup> Science Institute, VR-III, University of Iceland, 107 Reykjavík, Iceland

<sup>2</sup> Chemistry Department, University of Washington, Seattle, USA

**Abstract.** We discuss the optimization of a functional with respect to sets of orthonormal functions where unitary invariance does not apply. This problem arises, for example, when density functionals with explicit self-interaction correction are used for systems of electrons. There, unitary invariance cannot be used to reformulate the minimization of the energy with respect to each of the functions as an eigenvalue problem as can be done for the commonly used GGA-DFT and Hartree-Fock theory. By including optimization with respect to unitary transformations as an explicit step in the iterative minimization procedure, fast convergence can, nevertheless, be obtained. Furthermore, by working with two sets of orthonormal functions, the optimal functions and a set of eigenfunctions, the implementation of the extended functional form in existing software becomes easier. The additional computations arising from the lack of unitary invariance can largely be carried out in parallel.

**Keywords:** functional optimization, orthonormal functions, electrons.

## 1 Introduction

The task of optimizing the value of a functional of orthonormal functions arises in many contexts in engineering, physics and chemistry [1]. One example is the description of many-electron systems using density functional theory (DFT) which has become a widely used tool in calculations of the basic properties of solids, liquids and molecules [2]. Various approximations to the exact but unknown energy functional are used, but those that are commonly used and can be applied to large systems have several limitations in terms of the accuracy of the results, as described below. In this article, we discuss a possible extension of the form of energy functionals and the corresponding modifications in the minimization procedure. The extended functional form calls for new numerical methods and software implementations for solving the resulting equations.

In Kohn-Sham (KS) DFT [3] using local (LDA) or semi-local (GGA) functionals, the energy due to Coulomb interaction between the electrons

$$E_{\text{H}}[\rho] = \frac{1}{2} \int d^3\mathbf{r} d^3\mathbf{r}' \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \quad (1)$$

and the energy due to an external potential  $v_{\text{ext}}(\mathbf{r})$

$$E_{\text{ext}}[\rho] = \int d^3\mathbf{r} v_{\text{ext}}(\mathbf{r})\rho(\mathbf{r}) \quad (2)$$

are evaluated directly from the total electron density,  $\rho(\mathbf{r})$ , rather than the much more complicated many-electron wave function. But, in order to get a good enough estimate of the kinetic energy, a set of orthonormal functions  $\varphi^N = \{\varphi_1, \dots, \varphi_N\}$

$$\int d^3\mathbf{r} \varphi_i^*(\mathbf{r})\varphi_j(\mathbf{r}) = \delta_{ij} \quad (3)$$

each depending on the coordinates of just one electron (single-particle functions), are introduced and the kinetic energy minimized with respect to all sets  $\varphi^N$  consistent with the total electron density  $\rho(\mathbf{r}) = \sum_i^N \rho_i(\mathbf{r})$  where  $\rho_i(\mathbf{r}) = |\varphi_i(\mathbf{r})|^2$

$$T^{\text{KS}}[\rho] = \min_{\varphi^N} \sum_i^N \int d^3\mathbf{r} \varphi_i^*(\mathbf{r}) \left( -\frac{1}{2}\nabla^2 \right) \varphi_i(\mathbf{r}) . \quad (4)$$

The remaining contributions to the energy, which include the quantum mechanical exchange and correlation energy as well as correction to the above estimate of the kinetic energy, are denoted by  $E_{\text{xc}}[\rho]$ . They are estimated by comparison with numerically exact calculations of the homogeneous electron gas (when using the LDA approximation) or - as in most calculations today - also include estimates of the effect of local variations by including dependence on the gradient of the density (the GGA approximation) [2]

$$E_{\text{xc}}^{\text{KS}}[\rho] = \int d^3\mathbf{r} \epsilon_{\text{xc}}(\rho, \nabla\rho) . \quad (5)$$

The notation here ignores spin for simplicity. The ground state energy of the system is then obtained by variational minimization of the energy with respect to all electron density distributions,  $\rho$ , integrating to  $N$  electrons

$$E^{\text{KS}}[\rho] = T^{\text{KS}}[\rho] + E_{\text{H}}[\rho] + E_{\text{xc}}^{\text{KS}}[\rho] + \int d^3\mathbf{r} v_{\text{ext}}(\mathbf{r})\rho(\mathbf{r}) . \quad (6)$$

Using Lagrange's method, the orthonormality constraints are incorporated into the objective functional by

$$S^{\text{KS}}[\rho] = E^{\text{KS}}[\rho] - \sum_{i,j} \lambda_{ji} \left[ \int d^3\mathbf{r} \varphi_i^*(\mathbf{r})\varphi_j(\mathbf{r}) - \delta_{ij} \right] \quad (7)$$

where  $\mathbf{\Lambda} = \{\lambda_{ij}\}$  is a matrix of Lagrange multipliers. The variational optimization of  $S^{\text{KS}}$  with respect to the orthonormal, single-particle functions [4] gives

$$\hat{H}\varphi_i(\mathbf{r}) = \sum_{j=1}^N \lambda_{ji}\varphi_j(\mathbf{r}) \quad (8)$$

where  $\hat{H}$  is an operator (the Hamiltonian) defined as

$$\hat{H}\varphi_i(\mathbf{r}) = \frac{\delta E}{\delta\varphi_i^*(\mathbf{r})} \quad (9)$$

and turns out to be the same for all the functions. The functional is invariant under unitary transformations of the functions and variation with respect to  $\varphi_i$  rather than  $\varphi_i^*$  gives the same result. One can choose the particular set of functions for which  $\mathbf{A}$  is diagonal (see section 3). The set of coupled equations for the  $\varphi^N$  functions then reduces to a set of eigenvalue problems

$$\hat{H}\varphi_i(\mathbf{r}) = \epsilon_i\varphi_i(\mathbf{r}) \quad (10)$$

which, however, are still coupled through the total electron density. A solution can be obtained using an iterative procedure starting with a guess and eventually obtaining self-consistency.

Functionals of this type are widely used in the modeling of solids and liquids. Various semi-local approximations to  $E_{xc}[\rho]$  have been proposed and powerful software packages have been developed utilizing highly efficient optimization algorithms to solve the fundamental minimization problem [5]. However, several limitations of these functional approximations have also become apparent: (a) The predicted total energy is generally not accurate enough. Useful estimates of energy differences can still be obtained in many cases because of cancellation of error, but this is problematic when the two systems being compared are qualitatively different. For example, the energy of transition states compared with energy of stable states (i.e. reaction barriers) are typically underestimated [6,7]. (b) Electronic defect states tend to be overly delocalized and even unstable [8]. (c) Neither the functions  $\varphi_i$  nor energy eigenvalues  $\epsilon_i$  have any known, directly observable meaning (but the  $\epsilon_i$  are sometimes used as estimates of ionization energy or band gap, giving generally poor approximations). This list is far from being complete, but illustrates that the deficiencies of GGA functionals are significant. For a more complete discussion, see [9].

One approach to improve the semi-local approximation is to mix in some ‘exact exchange’ in so-called hybrid functionals [10,11] through a linear combination with LDA and GGA. Hybrid functionals can cure some of the deficiencies mentioned above for example improved bond energy and bond length [6,12]. The optimal linear combination coefficients, i.e. mixing parameters, are, however, not the same for all types of systems (for example molecules vs. solids) and this approach should be regarded as semi-empirical and relies on tuned cancellation of errors of different origin. For metallic systems, hybrid functionals in fact give poorer predictions than GGA. Although hybrid functionals are available in most major DFT software packages today, their application to systems with appreciable numbers of electrons is, furthermore, hampered by the expensive evaluation of the non-local, exact exchange. The computational effort scales as  $N^4$  rather than the  $N^3$  scaling for LDA and GGA.

## 2 Orbital Density Dependent Functionals

While the single-particle functions, which frequently are referred to as ‘orbitals’, are in GGA-DFT simply mathematical constructs that represent the total electron density and improve the estimate of the kinetic energy beyond what has been possible from the total electron density alone, these functions can in principle be interpreted as meaningful representations of the electrons and the corresponding probability density,  $\rho_i$ , represent the probability distribution of an electron. This is an assumption, consistent with intuition that is often invoked, but no proof of this has been presented. The form of the energy functional then should include explicit dependence on the orbital densities. We will refer to such an extended functional form as orbital density dependent (ODD) functionals. This can lead to much improved estimates of various properties, but the mathematical task of finding the optimal set of orbitals becomes more challenging.

The ODD functional form can, in particular, be used to correct for the so-called self-interaction error in GGA functionals. The evaluation of the Coulomb energy from the total electron density as in (1) includes interaction of the electrons with themselves, a self-interaction energy. Ideally, the  $E_{xc}$  correction term should remove this error, but in practice the approximations used for  $E_{xc}$ , such as PBE, only partly cancel it out. A better estimate of the Coulomb interaction is the orbital density dependent expression

$$E_H^{\text{ODD}}[\rho^N] = \frac{1}{2} \sum_{i \neq j} \int d^3\mathbf{r} d^3\mathbf{r}' \frac{\rho_i(\mathbf{r})\rho_j(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} = E_H[\rho] - \frac{1}{2} \sum_{i=1}^N \int d^3\mathbf{r} d^3\mathbf{r}' \frac{\rho_i(\mathbf{r})\rho_i(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \quad (11)$$

where the  $i = j$  terms representing self-interaction are excluded. Here,  $\rho^N$  denotes the set of  $N$  orbital densities,  $\rho^N = \{\rho_1 \dots \rho_N\}$ . Revised exchange-correlation functionals are necessary in order to account for this modification in the Coulomb term. The evaluation of this expression for the Coulomb energy requires  $N + 1$  solutions of the Poisson equation and thus is computationally much less demanding than the exact exchange of hybrid functionals and scales as one lower power in  $N$ . Furthermore, the  $N + 1$  Poisson equations can be solved simultaneously on  $N$  nodes or sets of nodes, making parallel implementation easy and efficient. Parallel implementation of hybrid functionals is more difficult [13].

Perdew and Zunger [14] proposed an estimate of the total self-interaction energy for each orbital as

$$E^{\text{SI}}[\rho_i] = \frac{1}{2} \int d^3\mathbf{r}' \frac{\rho_i(\mathbf{r})\rho_i(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3\mathbf{r} - E_{xc}[\rho_i] \quad (12)$$

and an improved estimate of the energy by explicit subtraction

$$E^{\text{KS-SIC}}[\rho^N] = E^{\text{KS}}[\rho] - \sum_i E^{\text{SI}}[\rho_i]. \quad (13)$$

The energy is no longer invariant under unitary transformations of the orbitals. For example, if delocalized Bloch functions are used in a calculation of a crystal, the self-interaction energy is small or even zero, but if localized orbitals - which can be formed by a unitary transformation of the Bloch functions - are used, then  $E^{\text{SI}}$  is finite and can be significant.

The problem is now to optimize  $S^{\text{ODD}}$  with respect to the orbitals, where

$$S^{\text{ODD}}[\rho^N] = E^{\text{ODD}}[\rho^N] - \sum_{i,j} \lambda_{ji} \left[ \int d^3\mathbf{r} \varphi_i^*(\mathbf{r}) \varphi_j - \delta_{ij} \right]. \quad (14)$$

The orbitals  $\varphi^N$  are in general complex functions, yielding two equations for the extremum

$$\left. \begin{aligned} \frac{\delta S^{\text{ODD}}}{\delta \varphi_i^*(\mathbf{r})} = 0 \\ \frac{\delta S^{\text{ODD}}}{\delta \varphi_i(\mathbf{r})} = 0 \end{aligned} \right\} \implies \left\{ \begin{aligned} \hat{H}_i \varphi_i(\mathbf{r}) &= \sum_{j=1}^N \lambda_{ji} \varphi_j(\mathbf{r}) \\ \hat{H}_i \varphi_i(\mathbf{r}) &= \sum_{j=1}^N \lambda_{ij}^* \varphi_j(\mathbf{r}) \end{aligned} \right. \quad (15)$$

where

$$\hat{H}_i \varphi_i(\mathbf{r}) = \frac{\delta E}{\delta \varphi_i^*(\mathbf{r})} \quad (16)$$

is the functional derivative of the energy with respect to the conjugate orbital. Note that both equations in (15) coincide whenever the matrix  $\mathbf{A}$  is Hermitian. Thus, an alternative set of conditions for the optimal set of orbitals is given by

$$\hat{H}_i \varphi_i(\mathbf{r}) = \sum_{j=1}^N \lambda_{ji} \varphi_j(\mathbf{r}) \quad (17a)$$

$$\mathbf{A} = \mathbf{A}^\dagger. \quad (17b)$$

For GGA,  $\mathbf{A}$  is guaranteed to be Hermitian, but not for ODD functionals. The orbitals  $\varphi^N$  obtained from an ODD functional are not arbitrary since the energy is not invariant under unitary transformations.

While there have been few self-consistent calculations using this functional form in the 30 years since the publication of the article by Perdew and Zunger (compared with the very large number of GGA calculations), see references in [9], it is clear that this functional form introduces flexibility that can be used to remove several deficiencies of the semi-local GGA functionals.

In the next section we will give a brief review of the minimization approaches that have previously been used for PZ-SIC. We then present our approach and compare the performance and reliability of various minimization schemes. We emphasize that the algorithms discussed here may be utilized for any functional of the ODD form, the PZ-SIC is used here only as an example. Development of an improved functional of the ODD form is currently ongoing.

### 3 Minimization of Energy Functionals

#### 3.1 Minimization of Unitary Invariant Functionals

For GGA functionals and hybrid functionals, the functional derivative of the energy can be expressed by a single operator,  $\hat{H}$ , the same for all orbitals

$$\hat{H} = \hat{T} + v_{\text{ext}}(\mathbf{r}) + v_{\text{H}}(\mathbf{r}) + \hat{v}_{\text{xc}}(\mathbf{r}) . \quad (18)$$

For GGA,  $\hat{v}_{\text{xc}}$  is a local multiplicative potential while hybrid functionals yield a non-local potential. Projection of (17a) yields an expression for the Lagrange multipliers

$$\lambda_{ij} = \lambda_{ji}^* = \int d^3\mathbf{r} \varphi_i^*(\mathbf{r}) \hat{H} \varphi_j(\mathbf{r}) . \quad (19)$$

which always fulfills also eqn. (17b). As the constraint matrix  $\mathbf{A}$  is Hermitian, it can be diagonalized using a unitary transformation  $W$  giving real eigenvalues  $\epsilon_i$

$$\lambda_{ij} = \sum_{k=1}^N \epsilon_k W_{ki}^* W_{kj} \quad (20)$$

and eigenfunctions  $\psi^N = \{\psi_1, \dots, \psi_N\}$

$$\psi_i(\mathbf{r}) = \sum_{k=1}^N W_{ik}^* \varphi_k(\mathbf{r}) . \quad (21)$$

The total density and energy do not change when the transformation is applied. The functions  $\psi^N$  are commonly taken to represent pseudo-particles of the non-interacting electron reference system. They span the total density  $\rho$  and make it possible to get a good estimate of the kinetic energy. One may be tempted to go beyond this and interpret these orbitals in terms of electrons since the defining equations are structurally equivalent to non-interacting Schrödinger equations. Any unitary transformation of the orbitals is, however, equally justified, but can typically give a range from highly localized to delocalized functions. The introduction of hybrid orbitals, for example  $sp^2$  and  $sp^3$  atomic orbitals, is an example of this flexibility in choosing the unitary transformation.

#### 3.2 Minimization of ODD Functionals

The algorithm for the minimization of an ODD functional with respect to the orbitals needs to be substantially different from the one used to minimize GGA functionals since the functional derivatives give a different operator,  $\hat{H}_i$  for each orbital

$$\hat{H}_i \varphi_i(\mathbf{r}) = \frac{\delta E}{\delta \varphi_i^*(\mathbf{r})} , \quad \hat{H}_i = \hat{H}_0 + \hat{V}_i \quad (22)$$

where  $\hat{H}_0$  is the unitary-invariant part of the operator and  $\hat{V}_i$  an orbital dependent part. From eqn. (17), a projection can be used to evaluate the Lagrange multipliers

$$\lambda_{ij} = \int d^3\mathbf{r} \varphi_i^*(\mathbf{r}) \hat{H}_j \varphi_j(\mathbf{r}) . \quad (23)$$

Note that in contrast to eqn. (19), the constraint matrix is not Hermitian. An asymmetry is introduced by the orbital dependence. The second condition (17b) should be enforced explicitly. This has consequences for the  $S^{\text{ODD}}$  objective functional: Its imaginary part is related to the anti-Hermitian part of  $\mathbf{A}$  and the deviation from orthonormality.

$$\text{Im}\{S^{\text{ODD}}\} = \sum_{i,j=1}^N \frac{\lambda_{ij}^* - \lambda_{ji}}{2i} \left[ \int d^3\mathbf{r} \varphi_i^*(\mathbf{r}) \varphi_j(\mathbf{r}) - \delta_{ij} \right] \quad (24)$$

While any solution of the extremum condition (15) yields a Hermitian  $\mathbf{A}$  matrix, intermediate stages of an iterative procedure typically lead to asymmetric matrices and there is not a unique way to define a proper set of Lagrange multipliers. Various possible choices to deal with this problem are discussed below.

Heaton, Harrison and Lin [16,17] presented an approach which corresponds to solving equations (16) and (23) without considering the symmetry condition for the Lagrange multipliers. An optimization with respect to unitary transformations of the orbitals was not included. Goedecker and Umrigar pointed out problems with this approach [18].

**Asymmetric and symmetric constraint** An alternative approach is to use one of the equations (15) to define the Lagrange multipliers and the other one to define the orbitals, for example

$$\hat{H}_i \varphi_i(\mathbf{r}) = \sum_{j=1}^N \lambda_{ji}^a \varphi_j(\mathbf{r}) , \quad \lambda_{ij}^a = \int d^3\mathbf{r} \varphi_i^*(\mathbf{r}) \hat{H}_i \varphi_j(\mathbf{r}) . \quad (25)$$

The constraint matrix  $\mathbf{A}^a$  is not necessarily Hermitian, but converges to a Hermitian matrix at the end of the iterative procedure.

Goedecker and Umrigar (GU) [18] used the Hermitian average,  $\mathbf{A}^s$ , of the two possible choices for the definition of the Lagrange multipliers

$$\hat{H}_i \varphi_i(\mathbf{r}) = \sum_{j=1}^N \lambda_{ji}^s \varphi_j(\mathbf{r}) , \quad \lambda_{ij}^s = \int d^3\mathbf{r} \varphi_i^*(\mathbf{r}) \frac{\hat{H}_i + \hat{H}_j}{2} \varphi_j(\mathbf{r}) . \quad (26)$$

It can be shown that for a given set of orbitals,  $\varphi^N$ , this choice of the constraint yields the direction of steepest descent for a correction to the orbitals.

**Unitary Optimization** Neither the symmetrized  $\mathbf{A}^s$  nor the asymmetric  $\mathbf{A}^a$  make explicit use of the necessary requirement for a Hermitian constraint matrix. Defining  $\kappa$  as

$$\kappa = (\mathbf{A}^a - \mathbf{A}^{a\dagger})/2 \quad (27)$$

yields an anti-Hermitian matrix with elements

$$\kappa_{ij} = \int d^3\mathbf{r} \varphi_i^*(\mathbf{r}) \frac{\hat{H}_j - \hat{H}_i}{2} \varphi_j(\mathbf{r}) \quad (28)$$

$$= \int d^3\mathbf{r} \varphi_i^*(\mathbf{r}) \frac{\hat{V}_j - \hat{V}_i}{2} \varphi_i(\mathbf{r}). \quad (29)$$

When the two orbital dependent operators are subtracted, the unitary invariant part is cancelled out, leaving only the difference of the orbital dependent parts.

Expressing explicitly the dependency of  $\kappa$  on unitary transformations amongst the orbitals

$$\varphi_i[\mathbf{U}](\mathbf{r}) = \sum_{j=1}^N U_{ji} \varphi_j(\mathbf{r}), \quad \rho_i[\mathbf{U}] = |\varphi_i[\mathbf{U}](\mathbf{r})|^2 \quad (30)$$

the equation defining the optimal unitary transformation  $\mathbf{U}$  is

$$\kappa[\mathbf{U}] = 0. \quad (31)$$

This is referred to as the ‘‘localization condition’’ [19]. The efficiency of unitary variant minimization algorithms depends on fast and reliable solution of this equation. Lagrange multipliers that are evaluated from orbitals satisfying the localization condition are guaranteed to fulfill the symmetry requirement (17b). Furthermore, using unitary optimization as a preconditioner for the constraint matrix, unifies the previously reported constraint definitions.

**Reintroduction of Eigenfunctions** Although the minimization can be carried out by solving the coupled set of equations (15), this is cumbersome and makes it difficult to incorporate ODD functionals into existing software which typically relies on the formulation of the optimization problem as an eigenvalue problem. This has led to the idea of using two sets of functions.

The first set, referred to as the optimal basis, is given by the functions  $\varphi^N$  which should converge to the solutions of the optimization problem (15). The second set, the ‘canonical orbitals’  $\psi^N$ , is introduced to decouple the equations into single-particle eigenvalue equations analogous to the ones obtained for unitary invariant functionals. Both sets of functions span the same total density  $\rho$  and are related to each other by a unitary transformation  $\mathbf{W}$

$$\varphi_i(\mathbf{r}) = \sum_{k=1}^N W_{ki} \psi_k(\mathbf{r}), \quad \psi_i(\mathbf{r}) = \sum_{k=1}^N W_{ik}^* \varphi_k(\mathbf{r}). \quad (32)$$

In order to decouple (15) into single-particle equations,  $\mathbf{W}$  has to diagonalize the constraint matrix

$$\lambda_{ij} = \sum_k \epsilon_k W_{ki}^* W_{kj}, \quad \delta_{ij} = \sum_k W_{ki}^* W_{kj}. \quad (33)$$



This step requires a Hermitian constraint matrix, which is provided either by the symmetrized eqn. (26), or by unitary optimization at each iteration.

Expressing the condition for minimal energy (15) in terms of the  $\psi^N$  yields

$$\sum_{k'=1}^N W_{k'i} \hat{H}_0 \psi_{k'}(\mathbf{r}) + \sum_{k'=1}^N W_{k'i} \hat{V}_i \psi_{k'}(\mathbf{r}) = \sum_{jk'=1}^N \lambda_{ji} W_{k'j} \psi_{k'}(\mathbf{r}). \quad (34)$$

and the equations can be decoupled by forming the linear-combination  $\sum_{i=1}^N W_{ki}^*$

$$\hat{H}_0 \psi_k(\mathbf{r}) + \sum_{ij=1}^N W_{ki}^* \hat{V}_i W_{ji} \psi_j(\mathbf{r}) = \epsilon_k \psi_k(\mathbf{r}). \quad (35)$$

The resulting operator is still orbital dependent but now with respect to the canonical orbitals rather than the optimal ones

$$\left( \hat{H}_0 + \hat{V}_k^c \right) \psi_k(\mathbf{r}) = \epsilon_k \psi_k(\mathbf{r}) \quad (36)$$

where  $\hat{V}_k^c$  is given by

$$\hat{V}_k^c f(\mathbf{r}) = \sum_{i=1}^N W_{ki}^* \hat{V}_i \varphi_i(\mathbf{r}) \int d^3 \mathbf{r}' \psi_k(\mathbf{r}') f(\mathbf{r}'). \quad (37)$$

$\hat{V}_k^c$  is structurally simpler than the previous  $\hat{V}_k$  in eqn. (22). It is invariant under unitary transformations of the functions  $\psi^N$  in a subtle way: The unitary transformation is simply compensated by an inverse change to  $\mathbf{W}$  in eqn. (32) maintaining the same  $\varphi^N$ . The canonical orbitals  $\psi^N$  turn out to converge faster than the  $\varphi^N$  which improve mainly through the unitary optimization. Numerically, the separation into basis set optimization and unitary optimization is advantageous in the electronic structure problem as different energy scales are separated, i.e. the relatively small contribution from  $\hat{V}_k$  is separated from the large contribution from  $\hat{H}_0$ . A similar procedure has been used in time-dependent DFT [20].

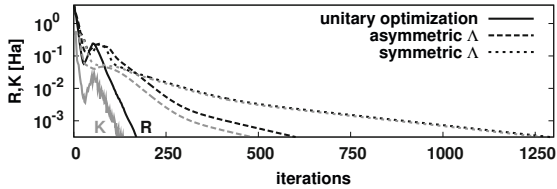
## 4 Performance

The performance of the minimization was benchmarked in all electron Gaussian type orbital based calculations of the  $\text{N}_2$  molecule. The convergence was measured in terms of the residual,  $R$ ,

$$R = \left[ \sum_{i=1}^N \int d^3 \mathbf{r} \left| \hat{H}_i \varphi_i(\mathbf{r}) - \sum_{j=1}^N \lambda_{ji} \varphi_j(\mathbf{r}) \right|^2 \right]^{1/2} \quad (38)$$

and a measure,  $K$ , of the error in the localization condition

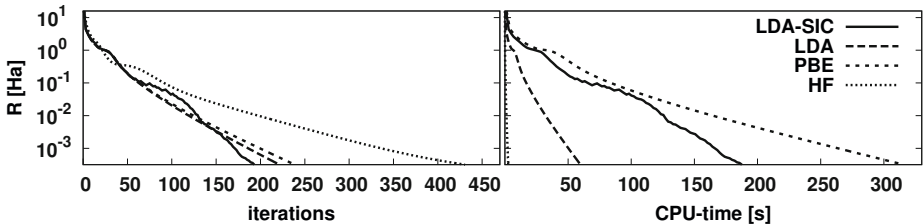
$$K = \|\kappa\| = \left[ \sum_{ij=1}^N |\kappa_{ij}|^2 \right]^{1/2}. \quad (39)$$



**Fig. 1.** Convergence of steepest descent minimization for a  $N_2$  molecule starting from same initial orbitals. The residual  $R$  (black) and  $K$  (grey) are shown for different methods of evaluating the  $\Lambda$  matrix: symmetric, asymmetric and including unitary optimization until  $K < 0.1R$ .

Figure 1 compares different choices for dealing with the  $\Lambda$  matrix. The energy was minimized using the steepest descent method which allows for direct comparison of different functionals and algorithms. Both the symmetric definition (26) and the asymmetric one (25) result in slow convergence rate in the later stage of the minimization. However, the origin of the slow convergence is different in the two cases. For the symmetric definition, the convergence of  $R$  and  $K$  is roughly equally slow, but for the asymmetric definition, which gives faster convergence,  $R$  is slower. In the unitary optimization, which converges much faster,  $K$  is reduced to less than 10% of  $R$  by an intermediate unitary optimization, followed by the use of the symmetric constraint.

The effort involved in minimizing the ODD type LDA-SIC functional is compared with LDA, GGA (using PBE) and Hartree-Fock in Fig. 2. The ODD calculation required similar number of iterations as the LDA and GGA calculations. Hartree-Fock requires many more iterations. The CPU time needed to reach convergence is also shown. The ODD calculation turns out to be faster in this case than the PBE calculation because the gradient dependent terms, which are absent in LDA, involve significant computational effort. The Hartree-Fock calculation is much faster than the others despite the large number of iterations because the integrals can for this small system be stored in memory.



**Fig. 2.** Convergence of a steepest descent minimization of the energy of  $N_2$  using various functionals: LDA, PBE (a GGA functional), LDA-SIC (an ODD functional) and Hartree-Fock (HF).

## 5 Parallelization

The main computational effort when ODD functionals are used is related to the evaluation of the  $N$  orbital dependent potentials. Each of them is as expensive as the GGA potential for the total electron density. But, since the calculations for the  $N$  orbitals are independent, they can readily be distributed over  $N$  processors without the need for significant communication. The real time of an ODD calculation would then be similar to a regular GGA calculation even for a large system. Our results obtained so far indicate that performance similar to GGA functionals can be accomplished for this more general functional form.

**Acknowledgements.** This work was supported by the Icelandic Research Fund, N-Inner Nordic fund, EC Marie Curie network 'Hydrogen' and the US National Science Foundation.

## References

1. Edelman, A., Arias, T.A., Smith, S.T.: *SIAM J. Matrix Anal. Appl.* 20, 303 (1998)
2. Kohn, W.: *Rev. of Mod. Phys.* 71, 1253 (1998)
3. Kohn, W., Sham, L.J.: *Phys. Rev.* 140, A1133 (1965)
4. Traditionally, the variation is carried out with respect to the total density, but for GGA functionals variation with respect to the  $\varphi_i(\mathbf{r})$  gives the same results
5. Davidson, E.R.: *Comp. Phys. Commun.* 53, 49 (1989)
6. Nachtigal, P., Jordan, K.D., Smith, A., Jónsson, H.: *J. Chem. Phys.* 104, 148 (1996)
7. Patchkovskii, S., Ziegler, T.: *J. Chem. Phys.* 116, 7806 (2002)
8. Nuttall, R.H.D., Weil, J.A.: *Can. J. Phys.* 59, 1696 (1981) Pacchioni, G., Frigoli, F., Ricci, D., Weil, J.A.: *Phys. Rev. B* 63, 054102 (2000)
9. Kümmel, S., Kronik, L.: *Reviews of Modern Physics* 80, 3 (2008)
10. Becke, A.D.: *J. Chem. Phys.* 98, 5648 (1993); 104, 1040 (1996); 88, 1053 (1988)
11. Perdew, J.P., Ernzerhof, M., Burke, K.: *J. Chem. Phys.* 105, 9982 (1996)
12. Kurth, S., Perdew, J.P., Blaha, P.: *Int. J. of Quantum Chem.* 75, 889 (1999)
13. Bylaska et al., E., Tsemekhman, K., Baden, S.B., Weare, J.H., Jónsson, H.: *J. Comp. Chem.* 32, 5469 (2011)
14. Perdew, J.P., Zunger, A.: *Phys. Rev. B* 23, 5048 (1981)
15. Perdew, J.P., Burke, K., Ernzerhof, M.: *Phys. Rev. Lett.* 77, 3865 (1996)
16. Heaton, R.A., Harrison, J.G., Lin, C.C.: *Phys. Rev. B* 28, 5992 (1983)
17. Harrison, J.G., Heaton, R.A., Lin, C.C.: *J. Phys. B* 16, 2079 (1983)
18. Goedecker, S., Umrigar, C.J.: *Phys. Rev. A* 55, 1765 (1997)
19. Pederson, M.R., Heaton, R., Lin, C.C.: *J. Chem. Phys.* 80, 1972 (1984)
20. Messud, J., Dinh, P.M., Reinhard, P.-G., Suraud, E.: *Phys. Rev. Lett.* 101, 096404 (2008)