Convergence analysis of adaptive DIIS algorithms with application to electronic ground state calculations

Maxime Chupin, Mi-Song Dupuy, Guillaume Legendre, Eric Séré

Convergence analysis of adaptive DIIS algorithms with application to electronic ground state calculations

Maxime Chupin*, Mi-Song Dupuy†, Guillaume Legendre‡and Éric Séré§

July 19, 2021

Abstract

This paper deals with a general class of algorithms for the solution of fixed-point problems that we refer to as *Anderson–Pulay acceleration*. This family includes the DIIS technique and its variant sometimes called commutator-DIIS, both introduced by Pulay in the 1980s to accelerate the convergence of self-consistent field procedures in quantum chemistry, as well as the related Anderson acceleration which dates back to the 1960s, and the wealth of techniques they have inspired. Such methods aim at accelerating the convergence of any fixed-point iteration method by combining several iterates in order to generate the next one at each step. This extrapolation process is characterised by its *depth*, *i.e.* the number of previous iterates stored, which is a crucial parameter for the efficiency of the method. It is generally fixed to an empirical value.

In the present work, we consider two parameter-driven mechanisms to let the depth vary along the iterations. In the first one, the depth grows until a certain nondegeneracy condition is no longer satisfied; then the stored iterates (save for the last one) are discarded and the method "restarts". In the second one, we adapt the depth continuously by eliminating at each step some of the oldest, less relevant, iterates. In an abstract and general setting, we prove under natural assumptions the local convergence and acceleration of these two adaptive Anderson–Pulay methods, and we show that one can theoretically achieve a superlinear convergence rate with each of them. We then investigate their behaviour in quantum chemistry calculations. These numerical experiments show that both adaptive variants exhibit a faster convergence than a standard fixed-depth scheme, and require on average less computational effort per iteration. This study is complemented by a review of known facts on the DIIS, in particular its link with the Anderson acceleration and some multisecant-type quasi-Newton methods.

1 Introduction

The *Direct Inversion in the Iterative Subspace* (DIIS) technique, introduced by Pulay [46] and also known as *Pulay mixing*, is a locally convergent method widely used in computational quantum chemistry for accelerating self-consistent field convergence. As a complement to available globally convergent methods, like the *optimal damping algorithm* (ODA) [10] or its energy-DIIS (EDIIS) variant [38], it remains a method of choice, in a large part due to its simplicity and nearly unparalleled performance once a convergence region has been attained. Due to this success, variants of the technique have been proposed over the years in other types of application, like the GDIIS adaptation [14, 17] for geometry optimization or the *residual minimisation method–direct inversion in the iterative subspace* (RMM-DIIS) for the simultaneous computation of eigenvalues and corresponding eigenvectors, attributed to Bendt and Zunger and described in [62]. It has also been combined with other schemes to improve the rate of convergence of various types of iterative calculations (see [35] for instance).

From a general point of view, the DIIS can be seen as an acceleration technique based on extrapolation, applicable to any fixed-point iterative scheme for which a measure of the error at each step, in the form of a residual for instance, is (numerically) available. It was recently established that this technique is closely related to an older process known as the *Anderson acceleration* [3]. It was also shown that it amounts to a multisecant-type variant of a Broyden method [8] and that, when applied to linear problems, it is (essentially) equivalent to the *generalised minimal residual* (GMRES) method of Saad and Schultz [50]. On this basis, Rohwedder and Schneider [48] (for the DIIS), and later Toth and Kelley [57] (for the Anderson acceleration), analysed the method in an abstract framework and provided convergence results.

*CEREMADE, UMR CNRS 7534, Université Paris-Dauphine, Université PSL, Place du Maréchal De Lattre De Tassigny, 75775 Paris cedex 16, France (chupin@ceremade.dauphine.fr).

†Zentrum Mathematik, Technische Universität München, Boltzmannstraße 3, 85747 Garching, Germany (dupuy@ma.tum.de)

‡CEREMADE, UMR CNRS 7534, Université Paris-Dauphine, Université PSL, Place du Maréchal De Lattre De Tassigny, 75775 Paris cedex 16, France (guillaume.legendre@ceremade.dauphine.fr).

§CEREMADE, UMR CNRS 7534, Université Paris-Dauphine, Université PSL, Place du Maréchal De Lattre De Tassigny, 75775 Paris cedex 16, France (sere@ceremade.dauphine.fr).

In the present paper, we consider a unified family of methods, which we refer to as *Anderson–Pulay acceleration*, encompassing the DIIS, its commutator-DIIS (CDIIS) variant [47], and the Anderson acceleration. In such methods, one keeps a "history" of previous iterates which are combined to generate the next one at each step with the aim of accelerating the convergence of the sequence of iterates. This extrapolation process is characterised by an integer, sometimes called the *depth* (see [57, 1, 18]), which is the number of previous iterates stored and is an important parameter for the efficiency of the method. In most applications, the depth grows up to an empirically fixed value m, then remains constant. One of the main conclusions of our work is that it can be beneficial to let the depth vary adaptively along the iterations.

In order to prove this point, we propose and investigate two parameter-driven procedures to determine the depth at each step of the Anderson–Pulay acceleration method. The first one allows the method to "restart", based on a condition initially introduced by Gay and Schnabel for a quasi-Newton method using multiple secant equations [25] and used by Rohwedder and Schneider in the context of the DIIS [48]. In the second one, we adapt the depth continuously, thanks to a very simple criterion which is new, as far as we know.

In a general framework, we mathematically analyse these two adaptive Anderson–Pulay methods, and prove local convergence and acceleration properties. In contrast with preceding works [48, 57], our results are obtained *without* any assumption on the boundedness of the extrapolation coefficients (which would have to be verified *a posteriori* in practice). Indeed, the built-in mechanism in each of the proposed algorithms prevents linear dependency from occurring in the least-squares problem for the coefficients, allowing us to derive a theoretical *a priori* bound on these coefficients. Applications of both methods to self-consistent field calculations in quantum chemistry and comparisons with their fixed-depth counterparts, demonstrate their good performances, and even suggest that adapting continuously the depth gives the fastest convergence in practice.

The paper is organized as follows. The principle of the Anderson–Pulay acceleration is presented in Section 2 through an overview of the DIIS and its relation with the Anderson acceleration and a class of quasi-Newton methods based on multisecant updating. We also recall convergence results existing in the literature for this class of methods, in both linear and nonlinear cases. in Section 3, a generic abstract problem is set and two variants of the Anderson–Pulay acceleration with adaptive depth are proposed to solve it. For both variants, a local convergence result is proved as well as an acceleration property, and we show that one can theoretically achieve a superlinear convergence rate. In Section 4, the quantum chemistry problems we consider for the application of the methods are recalled and their mathematical properties are discussed in connection with our abstract framework. Finally, numerical experiments, performed on molecules in order to illustrate the convergence behaviour of both methods and to compare them with their "classical" fixed-depth counterpart, are reported in Section 5.

2 Overview of the DIIS

The class of methods named Anderson–Pulay acceleration in the present paper comprises several methods, introduced in various applied contexts with the same goal of accelerating the convergence of fixed-point iterations by means of extrapolation, the most famous of these in the quantum chemistry community probably being the DIIS technique, introduced by Pulay in 1980 [46]. We first describe its basic principle in an abstract context and explore its relation to similar extrapolation methods and to a family of multisecant-type quasi-Newton methods. We conclude this section by recalling a number of previously established results pertaining to the DIIS and the Anderson acceleration.

2.1 Presentation

Consider the numerical computation of the fixed-point x_* of a nonlinear function g from $\mathbb{R}^n$ to $\mathbb{R}^n$ by the fixed-point iterative scheme

$$\forall k \in \mathbb{N},\ x^{(k+1)} = g(x^{(k)}), \tag{1}$$

for which one can compute at each step an *"error"* vector $r^{(k)}$ of the form

$$r^{(k)} = f(x^{(k)}),$$

with f a function from $\mathbb{R}^n$ to $\mathbb{R}^p$ (the integer p being possibly different from n) satisfying

$$f(x_*) = 0.$$

Given a non-negative integer m (the maximal depth), the DIIS paradigm assumes that a good approximation to the solution x_* can be obtained at step $k+1$ by forming a combination involving the $m_k + 1$ previous guess values, with $m_k = \min\{m, k\}$, that is

$$\forall k \in \mathbb{N},\ x^{(k+1)} = \sum_{i=0}^{m_k} c_i^{(k)}\, g(x^{(k-m_k+i)}), \tag{2}$$

while requiring that the coefficients $c_0^{(k)}, \dots, c_{m_k}^{(k)}$ of the combination are such that the norm of the associated linearised error vector, given by $\sum_{i=0}^{m_k} c_i^{(k)}\, r^{(k-m_k+i)}$, is minimal in the least-squares sense under the constraint that

$$\sum_{i=0}^{m_k} c_i^{(k)} = 1. \tag{3}$$

As discussed by Pulay, this minimisation may be achieved through a Lagrange multiplier technique applied to the normal equations associated with the problem. More precisely, one can introduce an undetermined scalar λ and define the Lagrangian function

$$L(c_0, \cdots, c_{m_k}, \lambda) = \frac{1}{2} \left\| \sum_{i=0}^{m_k} c_i\, r^{(k-m_k+i)} \right\|_2^2 - \lambda \left(\sum_{i=0}^{m_k} c_i - 1 \right) = \frac{1}{2} \sum_{i=0}^{m_k} \sum_{j=0}^{m_k} c_i c_j b_{ij}^{(k)} - \lambda \left(\sum_{i=0}^{m_k} c_i - 1 \right),$$

in which the coefficients $b_{ij}^{(k)}$, $i, j = 0, \dots, m_k$, are those of the Gramian matrix associated with the set of error vectors stored at the end of step k, *i.e.* $b_{ij}^{(k)} = \left\langle r^{(k-m_k+j)}, r^{(k-m_k+i)} \right\rangle_2$. One then finds the saddle point $(c_0^{(k)}, \dots, c_{m_k}^{(k)}, \lambda^{(k)})$ of this Lagrangian by solving the associated Euler-Lagrange equations. This amounts to solving the following system of $m_k + 2$ linear equations

$$\begin{pmatrix} b_{00}^{(k)} & \dots & b_{0m_k}^{(k)} & -1 \\ \vdots & & \vdots & \vdots \\ b_{m_k 0}^{(k)} & \dots & b_{m_k m_k}^{(k)} & -1 \\ -1 & \dots & -1 & 0 \end{pmatrix} \begin{pmatrix} c_0^{(k)} \\ \vdots \\ c_{m_k}^{(k)} \\ \lambda^{(k)} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -1 \end{pmatrix}. \tag{4}$$

The DIIS iterations are generally ended once an acceptable accuracy has been reached, for instance when the value of the norm of the current error vector lies below a prescribed tolerance. This procedure is summarised in Algorithm 1.

Algorithm 1: Pulay's DIIS [46] for the acceleration of a fixed-point method based on a function g.

Data: $x^{(0)}$, tol, m
$r^{(0)} = f(x^{(0)})$
$k = 0$
$m_k = 0$
while $\left\| r^{(k)} \right\|_2 > tol$ **do**
  solve the constrained least-squares problem for the coefficients $\{c_i^{(k)}\}_{i=0,\dots,m_k}$ using linear system (4)
  $x^{(k+1)} = \sum_{i=0}^{m_k} c_i^{(k)} g(x^{(k-m_k+i)})$
  $r^{(k+1)} = f(x^{(k+1)})$
  $m_{k+1} = \min(m_k + 1, m)$
  $k = k + 1$
end

In practice and for a large number of applications, the integer m is chosen small, of the order of a few units. Since system (4) is related to normal equations, it may nevertheless happen that it is ill-conditioned if some error vectors are (almost) linearly dependent. In such a situation, a possible cure is to drop the oldest stored vectors one by one, until the condition number of the resulting system becomes acceptable. One should note that the use of an unconstrained equivalent formulation of the least-squares problem, like those derived in the next subsection, is usually advocated, as it is generally observed that it results in a better condition number for the matrix of the underlying linear system. We refer the reader to [52, 58] for more details on this topic.

In [46], Pulay initially suggested using the quantities

$$\forall k \in \mathbb{N},\ r^{(k)} = x^{(k+1)} - x^{(k)} = g(x^{(k)}) - x^{(k)}, \tag{5}$$

as error vectors. Note that with this choice, $p = n$ and there is a direct relation between the error function f and the fixed-point function g :

$$f = g - \mathrm{id}\,. \tag{6}$$

Pulay later proposed in [47] a variant form of the procedure, sometimes known as the commutator-DIIS or simply CDIIS (see [24] for instance), which is more appropriate for self-consistent field (SCF) iterative schemes in quantum chemistry (see Section 4). In such applications, the fixed-point function takes its values in a submanifold Σ (consisting of idempotent matrices), but an extrapolation like (2) would not lie on Σ. As a remedy, the definition of the new iterate is modified in CDIIS, so that it belongs to the submanifold. Moreover, in this variant, *no* relation of the form (6) is assumed between the fixed-point and error functions. For applications in quantum chemistry, Pulay introduced an error function that turns out to be the commutator of two matrices, hence the name given to the method.

2.2 Relation with the Anderson and nonlinear Krylov accelerations

To solve a nonlinear equation of the form

$$h(x) = 0, \tag{7}$$

a classical idea is to perform a fixed-point iteration based on the function

$$g(x) = x + \beta\, h(x), \tag{8}$$

where β is a sufficiently regular, homogeneous operator[1]. One sees that the choice (5) for the error vectors corresponds to setting

$$f(x) = \beta\, h(x).$$

In this particular case, the error vectors are (possibly preconditioned) *residuals*, and it is possible to relate the DIIS to a structurally similar extrapolation method, the Anderson acceleration [3] (sometimes called the *Anderson mixing*), introduced for the numerical solution of discretised nonlinear integral equations and originally formulated as follows. At the $(k+1)$th step, given the $m_k + 1$ most recent iterates $x^{(k-m_k)}, \ldots, x^{(k)}$ and the corresponding residuals $r^{(k-m_k)}, \ldots, r^{(k)}$, define

$$\forall k \in \mathbb{N},\ x^{(k+1)} = g(x^{(k)}) + \sum_{i=1}^{m_k} \theta_i^{(k)} \left(g(x^{(k-m_k-1+i)}) - g(x^{(k)}) \right), \tag{9}$$

the scalars $\theta_i^{(k)}$, $i = 1, \ldots, m_k$ being chosen so as to minimise the norm of the associated linearised residual, *i.e.*

$$\forall k \in \mathbb{N},\ \boldsymbol{\theta}^{(k)} = \underset{(\theta_1,\ldots,\theta_{m_k}) \in \mathbb{R}^{m_k}}{\arg\min} \left\| r^{(k)} + \sum_{i=1}^{m_k} \theta_i \left(r^{(k-m_k-1+i)} - r^{(k)} \right) \right\|_2. \tag{10}$$

Setting

$$\forall k \in \mathbb{N},\ c_i^{(k)} = \theta_{i+1}^{(k)},\ i = 0, \ldots, m_k - 1,\ \text{and}\ c_{m_k}^{(k)} = 1 - \sum_{j=1}^{m_k} \theta_j^{(k)},$$

one observes that relation (9) can be put into the form of relation (2), so that the DIIS applied to the solution of (7) by fixed-point iteration is actually a reformulation of the Anderson acceleration.

This instance of the DIIS is not the only reinvention of the Anderson acceleration. In 1997, Washio and Oosterlee [60] introduced a closely related process dubbed *Krylov subspace acceleration* for the solution of nonlinear partial differential equation problems. The extrapolation within it relies on a nonlinear extension of the GMRES method (see subsection 2.4) and is in all points identical to (9) and (10), save for the definition of the fixed-point function g which corresponds in their setting to the application of a cycle of a nonlinear multigrid scheme. Another method, also based on a nonlinear generalisation of the GMRES method, is the so-called *nonlinear Krylov acceleration* (NKA), introduced by Carlson and Miller around 1990 (see [12]) and used for accelerating modified Newton iterations in a finite element solver with moving mesh. It leads to yet another equivalent formulation of the Anderson acceleration procedure by setting

$$\forall k \in \mathbb{N},\ x^{(k+1)} = g(x^{(k)}) - \sum_{i=1}^{m_k} \alpha_i^{(k)} \left(g(x^{(k-m_k+i)}) - g(x^{(k-m_k+i-1)}) \right), \tag{11}$$

with

$$\forall k \in \mathbb{N},\ \boldsymbol{\alpha}^{(k)} = \underset{(\alpha_1,\ldots,\alpha_{m_k}) \in \mathbb{R}^{m_k}}{\arg\min} \left\| r^{(k)} - \sum_{i=1}^{m_k} \alpha_i \left(r^{(k-m_k+i)} - r^{(k-m_k+i-1)} \right) \right\|_2, \tag{12}$$

which amounts to take $\alpha_i^{(k)} = \sum_{j=1}^{i} \theta_j^{(k)}$, $i = 1, \ldots, m_k$. Another form of the minimisation problem for the norm of the linearised residual is found in [15], where it is combined with a line-search to yield a so-called *nonlinear GMRES* (N-GMRES) optimisation algorithm for computing the sum of R rank-one tensors that has minimal distance to a given tensor in the Frobenius norm.

The fact that the above acceleration schemes are different forms of the same method has been recognised on several occasions in the literature, see [21, 58, 9] for instance or Anderson's own account in [4]. In [7], it is shown that the Anderson acceleration is actually part of a general framework for the so-called Shanks transformations, used for accelerating the convergence of sequences.

[1] This operator can be seen as a preconditioner of some sort, the simplest case being the multiplication by a (nonzero) constant β – a relaxation (or damping) parameter – or, in the context of the fixed-point iteration (1), a sequence $(\beta^{(k)})_{k\in\mathbb{N}}$ of such constants (see [3]). Hereafter, we restrict ourselves to the case where β is a fixed operator, so that the function f remains unchanged along the iteration.

Like the DIIS, the Anderson acceleration has been used to improve the convergence of numerical schemes in a number of contexts, and the recent years have seen a wealth of publications dealing with a large range of applications (see [40, 23, 61, 32, 1, 42, 64, 31] for instance). Variants with restart conditions to prevent stagnation [20], periodic restarts [45] or a periodic use of the extrapolation [5] have also been proposed.

2.3 Interpretation as a multisecant-type quasi-Newton method

An insight on the behaviour of the DIIS may be gained by seeing it as a quasi-Newton method using multiple secant equations. It was indeed observed by Eyert [19] that the Anderson acceleration procedure amounts to a modification of Broyden's *second*[2] method [8], in which a given number of secant equations are satisfied at each step. This relation was further clarified by Fang and Saad [20] as follows (see also the paper by Walker and Ni [58]).

Keeping with the previously introduced notations[3] and considering vectors as column matrices, we introduce the matrices respectively containing the differences of successive iterates and the differences of associated successive error vectors stored at step k,

$$\mathscr{Y}^{(k)} = \left[x^{(k-m_k+1)} - x^{(k-m_k)}, \ldots, x^{(k)} - x^{(k-1)}\right] \text{ and } \mathscr{S}^{(k)} = \left[r^{(k-m_k+1)} - r^{(k-m_k)}, \ldots, r^{(k)} - r^{(k-1)}\right],$$

in order to rewrite the recursive relation (11) as

$$\forall k \in \mathbb{N},\ x^{(k+1)} = x^{(k)} + r^{(k)} - \left(\mathscr{Y}^{(k)} + \mathscr{S}^{(k)}\right) \boldsymbol{\alpha}^{(k)},$$

the minimization problem (12) becoming a simple least-squares problem,

$$\forall k \in \mathbb{N},\ \boldsymbol{\alpha}^{(k)} = \underset{\boldsymbol{\alpha} \in \mathcal{M}_{m_k,1}(\mathbb{R})}{\arg\min} \left\| r^{(k)} - \mathscr{S}^{(k)} \boldsymbol{\alpha} \right\|_2.$$

Assuming that $\mathscr{S}^{(k)}$ is a full-rank matrix, which means that the error vectors are affinely independent, and characterising $\boldsymbol{\alpha}^{(k)}$ as the solution of the associated normal equations, with closed form

$$\boldsymbol{\alpha}^{(k)} = \left({\mathscr{S}^{(k)}}^{\top} \mathscr{S}^{(k)}\right)^{-1} {\mathscr{S}^{(k)}}^{\top} r^{(k)}, \tag{13}$$

one obtains the relation

$$\forall k \in \mathbb{N},\ x^{(k+1)} = x^{(k)} - \left(-I_n + \left(\mathscr{Y}^{(k)} + \mathscr{S}^{(k)}\right) \left({\mathscr{S}^{(k)}}^{\top} \mathscr{S}^{(k)}\right)^{-1} {\mathscr{S}^{(k)}}^{\top}\right) r^{(k)},$$

which can be identified with the update formula of a quasi-Newton method of multisecant type,

$$\forall k \in \mathbb{N},\ x^{(k+1)} = x^{(k)} - G^{(k)} r^{(k)}, \tag{14}$$

with

$$G^{(k)} = -I_n + \left(\mathscr{Y}^{(k)} + \mathscr{S}^{(k)}\right) \left({\mathscr{S}^{(k)}}^{\top} \mathscr{S}^{(k)}\right)^{-1} {\mathscr{S}^{(k)}}^{\top}. \tag{15}$$

Here, the matrix $G^{(k)}$ is regarded as an approximate inverse of the Jacobian of the function f at point $x^{(k)}$, satisfying the inverse multiple secant condition

$$G^{(k)} \mathscr{S}^{(k)} = \mathscr{Y}^{(k)}.$$

It moreover minimises the Frobenius norm $\|G^{(k)} + I_n\|_2$ among all the matrices satisfying this condition. Thus, formula (15) can be viewed as a rank-m_k update of $-I_n$ generalising Broyden's second method, which effectively links the Anderson acceleration to a particular class of quasi-Newton methods. Apparently not aware of this connection, Rohwedder and Schneider [48] derived a similar conclusion in their analysis of the DIIS considering a full history of iterates, showing that it corresponds to a projected variant of Broyden's second method proposed by Gay and Schnabel in 1977 (see Algorithm II' in [25]).

In contrast, the generalisation of Broyden's first method aims at directly approximating the Jacobian of the function f at point $x^{(k)}$ by a matrix $B^{(k)}$ which minimises the norm $\|B^{(k)} + I_n\|_2$ subject to the multiple secant condition $B^{(k)} \mathscr{Y}^{(k)} = \mathscr{S}^{(k)}$. Under the assumption that the matrix $\mathscr{Y}^{(k)}$ is full-rank, this yields

$$B^{(k)} = -I_n + \left(\mathscr{Y}^{(k)} + \mathscr{S}^{(k)}\right) \left({\mathscr{Y}^{(k)}}^{\top} \mathscr{Y}^{(k)}\right)^{-1} {\mathscr{Y}^{(k)}}^{\top}.$$

[2] It is sometimes also called the *bad* Broyden method (see [27]).

[3] We continue to assume that the relaxation parameter β does not vary from one iteration to the next.

Using this fact, Fang and Saad [20] defined a "type-I" Anderson acceleration (the "type-II" Anderson acceleration corresponding to the original one, related to Broyden's second method as seen above). Indeed, assuming that the matrix ${\mathscr{Y}^{(k)}}^\top \mathscr{S}^{(k)}$ is invertible and applying the Sherman–Morrison–Woodbury formula, it follows that

$$(B^{(k)})^{-1} = -I_n + \left(\mathscr{Y}^{(k)} + \mathscr{S}^{(k)}\right) \left({\mathscr{Y}^{(k)}}^\top \mathscr{S}^{(k)}\right)^{-1} {\mathscr{Y}^{(k)}}^\top,$$

and substituting $(B^{(k)})^{-1}$ to $G^{(k)}$ in (14), one is led to a variant of the original method, the coefficients of which satisfy

$$\tilde{\boldsymbol{\alpha}}^{(k)} = \left({\mathscr{Y}^{(k)}}^\top \mathscr{S}^{(k)}\right)^{-1} {\mathscr{Y}^{(k)}}^\top r^{(k)},$$

instead of (13).

To end this subsection, let us mention that a relation between the DIIS applied to ground state calculations and quasi-Newton methods was also suspected, albeit in heuristic ways, in articles originating from the computational chemistry community (see [55, 37] for instance).

2.4 Equivalence with the GMRES method for linear problems

Consider the application of the DIIS to the solution of a system of n linear equations in n unknowns, with solution x_*, by assuming that the function h in (7) is of the form $h(x) = b - Ax$, where A is a nonsingular matrix of order n and b a vector of $\mathbb{R}^n$. In such a case, relation (8) amounts to $g(x) = (I_n - \beta\, A)\, x + \beta\, b$, and the fixed-point iteration method (1) reduces to the stationary Richardson method. In what follows, we assume that all the previous error vectors are kept at each iteration, that is $m = +\infty$, so that $m_k = k$ for each natural integer k.

A well-known iterative method for the solution of the above linear system is the GMRES method [50], in which the approximate solution $x^{(k+1)}$ at the $(k+1)$th step is the unique vector minimising the Euclidean norm of the residual

$$r_{\mathrm{GMRES}}^{(k+1)} = b - A\, x^{(k+1)}$$

in the affine space

$$\mathcal{W}_{k+1} = \left\{ v = x^{(0)} + z \,|\, z \in \mathcal{K}_{k+1}(A, r_{\mathrm{GMRES}}^{(0)}) \right\},$$

where, for any integer j greater than or equal to 1, $\mathcal{K}_j(A, r_{\mathrm{GMRES}}^{(0)}) = \mathrm{span}\left\{ r_{\mathrm{GMRES}}^{(0)}, A r_{\mathrm{GMRES}}^{(0)}, \ldots, A^{j-1} r_{\mathrm{GMRES}}^{(0)} \right\}$ is the *order-j Krylov (linear) subspace* generated by the matrix A and the residual vector $r_{\mathrm{GMRES}}^{(0)} = b - Ax^{(0)}$, the starting vector $x^{(0)}$ being given. Since each of these Krylov subspaces is contained in the following one, the norm of the residual decreases monotonically with the iterations and the exact solution is obtained in at most n iterations (the idea being that an already good approximation to the exact solution is reached after a number of iterations much smaller than n). More precisely, the following result holds (for a proof, see for instance [44]).

Proposition 2.1 *The GMRES method converges in exactly $\nu(A, r_{\mathrm{GMRES}}^{(0)})$ steps, where the integer $\nu(A, r_{\mathrm{GMRES}}^{(0)})$ is the grade[4] of $r_{\mathrm{GMRES}}^{(0)}$ with respect to A.*

In [28], it was shown that the GMRES method could be interpreted as a quasi-Newton method. Similarly, the fact that the DIIS (or the Anderson acceleration) used to solve a linear system is equivalent to the GMRES method in the following sense was proved in [58] and [48] (see also [44] for more refined results).

Theorem 2.2 *Suppose that $x_{\mathrm{DIIS}}^{(0)} = x_{\mathrm{GMRES}}^{(0)} = x^{(0)}$ and that the sequence of residual norms does not stagnate[5] before step k, i.e. $r_{\mathrm{GMRES}}^{(k-1)} \neq 0$ and $\|r_{\mathrm{GMRES}}^{(j-1)}\|_2 > \|r_{\mathrm{GMRES}}^{(j)}\|_2$ for each integer j such that $0 < j < k$. Then*

$$x_{\mathrm{GMRES}}^{(k)} = \sum_{i=0}^{k} c_i^{(k)} x_{\mathrm{DIIS}}^{(i)},$$

and

$$x_{\mathrm{DIIS}}^{(k+1)} = g(x_{\mathrm{GMRES}}^{(k)}).$$

[4] The grade of a non-zero vector x with respect to a matrix A is the smallest integer ℓ for which there is a non-zero polynomial p of degree ℓ such that $p(A)x = 0$.

[5] As shown in [26], any nonincreasing sequence of residual norms can be produced by the GMRES method. It is then possible for the residual norm to stagnate at some point, say at the kth step, k being a nonzero natural integer, with $r_{\mathrm{GMRES}}^{(k)} = r_{\mathrm{GMRES}}^{(k-1)} \neq 0$. In such a case, the equivalence between the methods implies that $x_{\mathrm{DIIS}}^{(k+1)} = x_{\mathrm{DIIS}}^{(k)}$, making the least-square problem associated with the minimisation ill-posed due to the family $\{f(x_{\mathrm{DIIS}}^{(0)}), \ldots, f(x_{\mathrm{DIIS}}^{(k+1)})\}$ not having full rank. The DIIS method then breaks down upon stagnation before the solution has been found, whereas the GMRES method does not. As pointed out in [58], this results in the conditioning of the least-square problem being of utmost importance in the numerical implementation of the method.

Assuming that a full history of iterates is kept and that no stagnation occurs, this equivalence directly provides convergence results in the linear case for the DIIS in view of the well-known theory for the GMRES method. In particular, if g is a contraction, then stagnation is forbidden and the DIIS converges to the exact solution in a finite number of steps. Theorem 2.2 also justifies the inclusion of Krylov's name in some of the rediscoveries of the Anderson acceleration we previously mentioned. Of course, if the number of stored iterates is fixed or if "restarts" are allowed over the course of the computation, in the spirit of the periodically restarted GMRES method, GMRES(m), these results are no longer valid. Nevertheless, the behaviour of GMRES(m) has been studied in a number of particular cases, and various restart or truncation strategies (see [16] and the references therein for instance) have been proposed over the years in order to compensate for the loss of information by selectively retaining some of it from earlier iterations.

Walker and Ni [58] showed in a similar way that the type-I Anderson acceleration, recalled in the preceding subsection, is essentially equivalent to the full orthogonalisation method (FOM) based on the Arnoldi process (see section 2 of [50]) in the linear case.

2.5 Existing convergence theories in the nonlinear case

In light of the previous subsection, it appears that the convergence theory for the DIIS in the linear case is well developed and intimately linked to that of the GMRES method. On the contrary, this theory is far from being established when the method is applied to nonlinear problems. The interpretation of the DIIS as a particular instance of quasi-Newton methods does not provide an answer, as there is no general convergence theory for these methods. Nevertheless, results have emerged in the literature in the recent years.

In [48], Rohwedder and Schneider proved a local q-convergence result[6] for the DIIS applied to a general nonlinear problem of the form (7), with the choice $\beta = -1$ in (8). They assumed that the underlying mapping is locally contractive and that error vectors associated with the stored iterates satisfy an affine independence condition. Their analysis is based on the equivalence between the DIIS and a projected variant of Broyden's second method, the convergence properties and an improvement of estimates appearing in the proof of convergence by Gay and Schnabel in [25]. By interpreting the DIIS as an inexact Newton method, they also obtained a refined estimate for the error vector at a given step that allowed them to discuss the possibility of obtaining a superlinear convergence rate for the method.

More recently, Kelley and Toth [57] studied the use of the Anderson acceleration on a fixed-point iteration in $\mathbb{R}^n$, also based on the function (8) in which one has set $\beta = 1$, the history of error vectors being of fixed depth m. Assuming that the fixed-point mapping is Lipschitz continuously differentiable and is a contraction in a neighbourhood of the solution, and with the requirement that the extrapolation coefficients remain uniformly bounded in the ℓ^1-norm, they proved that the method converges locally r-linearly. When $m = 1$, they showed that the sequence of residuals converges q-linearly if the fixed-point mapping is sufficiently contractive and the initial guess is close enough to the solution. This analysis was later extended to the case where the evaluation of the fixed point map is corrupted with errors in [56]. Chen and Kelley [13] also obtained global and local convergence results for a variant of the method in which a non-negativity constraint on the coefficients is imposed (an idea previously used in the EDIIS method [38]), generalising the results in [57]. Recently, Evans *et al.* [18] studied the convergence improvement given by the Anderson method.

Note, however, that in all these works, for $m > 1$ an assumption is made on the boundedness of the extrapolation coefficients, or on the affine independence of the error vectors. Such an assumption can only be checked a posteriori in the standard DIIS or Anderson algorithm. To address this issue, a stabilised version of the type-I Anderson acceleration is introduced by Zhang *et al.* in [63]. It includes a regularisation ensuring the non-singularity of the approximated Jacobian, a restart strategy guaranteeing a certain linear independence of the differences of successive stored iterates, and a safeguard mechanism guaranteeing a decrease of the residual norm. A global convergence result for the corresponding acceleration applied to a Krasnoselskii–Mann iteration in $\mathbb{R}^n$ is then proved.

3 Restarted and adaptive-depth Anderson–Pulay acceleration

Two peculiarities of the CDIIS variant proposed by Pulay for electronic ground state calculations [47] are that, when interpreted in terms of an abstract fixed-point function g and an abstract error function f, the target set of the function g is possibly a submanifold and the functions g and f do not necessarily satisfy a relation of the form (6) (see Section 4 for more details). As a consequence, the convergence theories found in [48] or in [57] do not apply.

This observation leads us to introduce an abstract framework in which the class of Anderson–Pulay acceleration algorithms is defined. This class encompasses and generalises the DIIS, the Anderson acceleration, the CDIIS and

[6] The notions of q-linear convergence and r-linear convergence are both recalled in Subsection 3.2.

other methods reviewed in the previous section (see Algorithm 2). In addition, two adaptive modifications of this procedure are proposed: the first one, Algorithm 3, includes a restart condition in the spirit of Gay and Schnabel [25]; in the second one, Algorithm 4, the depth is continuously adapted at each step according to a new criterion.

3.1 Description of the algorithms in an abstract framework

We consider an error function f, a fixed-point function g, and a point x_* in $\mathbb{R}^n$ such that $f(x_*) = 0$ and $g(x_*) = x_*$. We do *not* impose that f and g be linked by a relation of the form (6). In Subsection 3.2, two assumptions will be made on these functions in a neighborhood of x_*, ensuring that the norm $\|f(x^{(k)})\|_2$ of the error diminishes along the basic fixed-point iteration (1), and that this norm controls the distance $\|x^{(k)} - x_*\|_2$.

Our precise definition of Anderson–Pulay acceleration is given below in Algorithm 2.

Algorithm 2: Fixed-depth Anderson–Pulay acceleration for a fixed-point method based on a function g.

Data: $x^{(0)}$, tol, m
$r^{(0)} = f(x^{(0)})$
$k = 0$
$m_k = 0$
while $\left\|r^{(k)}\right\|_2 > tol$ **do**
  solve the constrained least-squares problem for the coefficients
  $\{c_i^{(k)}\}_{i=0,\dots,m_k} = \underset{\substack{(c_0,\dots,c_{m_k})\in\mathbb{R}^{m_k+1} \\ \sum_{i=0}^{m_k} c_i = 1}}{\arg\min} \left\|\sum_{i=0}^{m_k} c_i\, r^{(k-m_k+i)}\right\|_2$
  $x^{(k+1)} = \sum_{i=0}^{m_k} c_i^{(k)} g(x^{(k-m_k+i)})$ (**version A**)
  or
  $x^{(k+1)} = g(\widetilde{x}^{(k+1)})$ with $\widetilde{x}^{(k+1)} = \sum_{i=0}^{m_k} c_i^{(k)} x^{(k-m_k+i)}$ (**version P**)
  $r^{(k+1)} = f(x^{(k+1)})$
  $m_{k+1} = \min(m_k + 1, m)$
  $k = k + 1$
end

One can observe that the Anderson–Pulay acceleration comes in two versions, which only differ in the definition of the new iterate. Comparing with Algorithm 1, one sees that version A corresponds to the DIIS [46], and is equivalent to the Anderson acceleration [3] if relation (6) holds. The formula giving the new iterate in version A can be seen as a linearised form of the one in version P, the latter being directly inspired by the CDIIS [47]. One may thus note that the two versions coincide for a linear fixed-point function g.

As already discussed in Subsection 2.1, the numerical solution of the constrained linear least-squares problem for the extrapolation coefficients requires some care. Indeed, the matrix of linear system (4) resulting from the use of a Lagrange multiplier accounting for the constraint has to be well-conditioned for the method to be applicable in floating-point arithmetic. Unconstrained formulations of the problem exist, as employed in the Anderson acceleration (see (10)) or its numerous reinventions (see (12) for instance), each leading to an equivalent Anderson–Pulay acceleration algorithm, but with a possibly different condition number for the matrix of the linear system associated with the least-squares problem.

Note that a bound on the extrapolation coefficients is also needed when investigating the convergence of the method. In [57], this bound is presented as a requirement, following, for instance, from the uniform well-conditioning of the least-squares problem, which cannot be *a priori* guaranteed. Nevertheless, such a condition can be enforced *a posteriori* by diminishing the value of the integer m_k. In practice, this may be achieved by monitoring the condition number of the least-squares coefficient matrix and by dropping as many of its left-most columns as needed to have this condition number below a prescribed threshold, as proposed in [58]. Another possibility consists in ensuring that the error vectors associated with the stored iterates fulfill a kind of linear independence condition like the one found in [25, 48]. To enforce this condition in practice, one can simply choose to "restart" the method, by resetting the value of m_k to zero and discarding the iterates previously stored, as soon as an "almost" linear dependence is detected, as done in [25, 63]. This approach is adopted for the first adaptive modification of Algorithm 2 proposed and analysed in the present work.

More precisely, assuming that $m_k \geq 1$ at step k, set

$$s^{(k-m_k+i)} = r^{(k-m_k+i)} - r^{(k-m_k)},\ i = 1, \dots, m_k,$$

and let Π_k denote the orthogonal projector onto $\operatorname{span}\{s^{(k-m_k+1)}, \dots, s^{(k)}\}$. The algorithm will "restart" at step $k+1$ (meaning that the number m_{k+1} will be set to zero) if the norm of $s^{(k+1)} - \Pi_k s^{(k+1)}$ is "small" compared to the

norm of $s^{(k+1)}$, that is, if the following inequality holds:

$$\tau \, \|r^{(k+1)} - r^{(k-m_k)}\|_2 > \|(\mathrm{id} - \Pi_k)(r^{(k+1)} - r^{(k-m_k)})\|_2 \tag{16}$$

where τ is a real parameter chosen between 0 and 1. Otherwise, the integer m_k is incremented by one unit with the iteration number k. One can view condition (16) as a near-linear dependence criterion for the set of differences of stored error vectors. Indeed, as explained in [25], it *"recognizes that, in general, the projection of $s^{(i)}$ orthogonal to the subspace spanned by $s^{(1)}, \ldots, s^{(i-1)}$ must be the zero vector for some $i \leq n$"*.

Such a restart condition is particularly adapted to a specific unconstrained formulation of the linear least-squares problem for the extrapolation coefficients

$$\boldsymbol{c}^{(k)} = \underset{\substack{(c_0,\ldots,c_{m_k}) \in \mathbb{R}^{m_k+1} \\ \sum_{i=0}^{m_k} c_i = 1}}{\arg\min} \left\| \sum_{i=0}^{m_k} c_i \, r^{(k-m_k+i)} \right\|_2 , \tag{17}$$

to be effectively employed in practice. In this formulation, instead of using the constrained vector of coefficients $\boldsymbol{c} = (c_0, \ldots, c_{m_k})$, one works with the unconstrained vector $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_{m_k})$ with $\gamma_i = c_i$, $1 \leq i \leq m_k$. One then has

$$\sum_{i=0}^{m_k} c_i \, r^{(k-m_k+i)} = r^{(k-m_k)} + \sum_{i=1}^{m_k} \gamma_i \, s^{(k-m_k+i)},$$

resulting in (17) being replaced by

$$\boldsymbol{\gamma}^{(k)} = \underset{(\gamma_1,\ldots,\gamma_{m_k}) \in \mathbb{R}^{m_k}}{\arg\min} \left\| r^{(k-m_k)} + \sum_{i=1}^{m_k} \gamma_i \, s^{(k-m_k+i)} \right\|_2 .$$

The corresponding modification of the Anderson–Pulay acceleration is given in Algorithm 3. The question of the numerical computation of the orthogonal projector Π_k will be addressed in Subsection 5.1.

Algorithm 3: Restarted Anderson–Pulay acceleration for a fixed-point method based on a function g.

Data: $x^{(0)}$, *tol*, τ
$r^{(0)} = f(x^{(0)})$
$x^{(1)} = g(x^{(0)})$
$r^{(1)} = f(x^{(1)})$
$s^{(1)} = r^{(1)} - r^{(0)}$
$k = 1$
$m_k = 1$
while $\left\|r^{(k)}\right\|_2 > tol$ **do**
  solve the unconstrained least-squares problem for the coefficients
  $\{\gamma_i^{(k)}\}_{i=1,\ldots,m_k} = \underset{(\gamma_1,\ldots,\gamma_{m_k}) \in \mathbb{R}^{m_k}}{\arg\min} \left\|r^{(k-m_k)} + \sum_{i=1}^{m_k} \gamma_i \, s^{(k-m_k+i)}\right\|_2$
  $x^{(k+1)} = g(x^{(k-m_k)}) + \sum_{i=1}^{m_k} \gamma_i^{(k)} (g(x^{(k-m_k+i)}) - g(x^{(k-m_k)}))$ (**version A**)
  or
  $x^{(k+1)} = g(\widetilde{x}^{(k+1)})$ with $\widetilde{x}^{(k+1)} = x^{(k-m_k)} + \sum_{i=1}^{m_k} \gamma_i^{(k)} (x^{(k-m_k+i)} - x^{(k-m_k)})$ (**version P**)
  $r^{(k+1)} = f(x^{(k+1)})$
  $s^{(k+1)} = r^{(k+1)} - r^{(k-m_k)}$
  compute $\Pi_k s^{(k+1)}$ (the orthogonal projection of $s^{(k+1)}$ onto $span\{s^{(k-m_k+1)}, \ldots, s^{(k)}\}$)
  if $\tau \left\|s^{(k+1)}\right\|_2 > \left\|(\mathrm{id} - \Pi_k) s^{(k+1)}\right\|_2$ **then**
    $m_{k+1} = 0$
  else
    $m_{k+1} = m_k + 1$
  end
  $k = k + 1$
end

Note that, with this instance of the method, all the stored iterates, except for the last one, are discarded when a restart occurs. In some practical computations, a temporary slowdown of the convergence is observed after the restart (see Section 5). To try to remedy such an inconvenience, we introduce a smoother adaptive modification of Algorithm 2, based on an update of the set of stored iterates at step $k + 1$. The idea is to eliminate the old iterates

that are too far from the fixed point, compared to the most recent one. More precisely, one chooses the largest possible depth $m_{k+1} \leq m_k + 1$ such that the following inequality is satisfied for all $k + 1 - m_{k+1} \leq i \leq k$:

$$\delta \|r^{(i)}\|_2 < \|r^{(k+1)}\|_2.$$

Here, the real number δ is a small positive parameter. To the best of our knowledge, this criterion is new, and the corresponding variant of the Anderson–Pulay acceleration is given below in Algorithm 4.

Algorithm 4: Adaptive-depth Anderson–Pulay acceleration for a fixed-point method based on a function g.

Data: $x^{(0)}$, *tol*, δ
$r^{(0)} = f(x^{(0)})$
$x^{(1)} = g(x^{(0)})$
$r^{(1)} = f(x^{(1)})$
$s^{(1)} = r^{(1)} - r^{(0)}$
$k = 1$
$m_k = 1$
while $\left\|r^{(k)}\right\|_2 > tol$ **do**
  solve the unconstrained least-squares problem for the coefficients
  $\{\alpha_i^{(k)}\}_{i=1,\dots,m_k} = \underset{(\alpha_1,\dots,\alpha_{m_k})\in\mathbb{R}^{m_k}}{\arg\min} \left\|r^{(k)} - \sum_{i=1}^{m_k} \alpha_i \, s^{(k-m_k+i)}\right\|_2$
  $x^{(k+1)} = g(x^{(k)}) - \sum_{i=1}^{m_k} \alpha_i^{(k)} (g(x^{(k-m_k+i)}) - g(x^{(k-m_k+i-1)}))$ (**version A**)
  or
  $x^{(k+1)} = g(\widetilde{x}^{(k+1)})$ with $\widetilde{x}^{(k+1)} = x^{(k)} - \sum_{i=1}^{m_k} \alpha_i^{(k)} (x^{(k-m_k+i)} - x^{(k-m_k+i-1)})$ (**version P**)
  $r^{(k+1)} = f(x^{(k+1)})$
  $s^{(k+1)} = r^{(k+1)} - r^{(k)}$
  set m_{k+1} the largest integer $m \leq m_k + 1$ such that for $k + 1 - m \leq i \leq k$, $\delta \, \|r^{(i)}\|_2 < \|r^{(k+1)}\|_2$
  $k = k + 1$
end

In the last algorithm, one may observe that the unconstrained form of the least-squares problem for the coefficients uses the differences of error vectors associated with *successive* iterates, that is, $s^{(i)} = r^{(i)} - r^{(i-1)}$, $i = k - m_k + 1, \dots, k$, which is a computationally convenient choice when storing a set of iterates whose depth is adapted at each step. One then has $\alpha_i^{(k)} = \sum_{j=0}^{i-1} c_j^{(k)}$, $i = 1, \dots, m_k$.

The rest of this section is devoted to theoretical convergence results for Algorithms 3 and 4. Since we will analyse the behaviour of infinite sequences, the stopping parameter *tol* used in practice will be set to zero. If there exists some natural integer k_{stop} for which $r^{(k_{\text{stop}})} = 0$, we will also adopt the convention that $x^{(k)} = x^{(k_{\text{stop}})}$, $r^{(k)} = 0$ and $m_k = 0$ for all natural integer k greater or equal to k_{stop}.

3.2 Linear convergence

We shall now study under natural assumptions the convergence of the modified Anderson–Pulay acceleration methods devised in the previous subsection. To this end, we consider the following functional setting.

Let n and p be two nonzero natural integers such that $p \leq n$, Σ be a smooth submanifold in $\mathbb{R}^n$, V be an open subset of $\mathbb{R}^n$, f be a function in $\mathscr{C}^2(V, \mathbb{R}^p)$, g be a function in $\mathscr{C}^2(V, \Sigma)$, and x_* in $V \cap \Sigma$ be a fixed point of g satisfying $f(x_*) = 0$. We next make two assumptions.

Assumption 1. There exists a constant K in $(0, 1)$ such that

$$\forall x \in V \cap g^{-1}(V) \cap \Sigma, \ \|(f \circ g)(x)\|_2 \leq K \, \|f(x)\|_2.$$

Assumption 2. There exists a positive constant σ such that

$$\forall x \in V \cap \Sigma, \ \sigma \|x - x_*\|_2 \leq \|f(x)\|_2.$$

Note that these assumptions ensure that x_* is the unique fixed point of g in V and the unique solution of the equation $f(x) = 0$ in $V \cap \Sigma$. There might be other zeroes of f lying outside of Σ, but they would not be fixed points of g.

Since f and g are both of class $\mathscr{C}^2$, we may assume, taking $U \subset V \cap g^{-1}(V)$ a smaller neighbourhood of x_* if necessary, that

$$\forall (x, y) \in U^2, \ \|f(x) - f(y)\|_2 \leq 2 \, \|\mathrm{D}f(x_*)\|_2 \|x - y\|_2$$
$$\text{and } \|(f \circ g)(x) - (f \circ g)(y)\|_2 \leq 2 \, \|\mathrm{D}f(x_*) \circ \mathrm{D}g(x_*)\|_2 \|x - y\|_2, \quad (18)$$

where $\mathrm{D}f(x_*)$ and $\mathrm{D}g(x_*)$ are the respective differentials of f and g at point x_*, and that one also has, for some positive constants L and L',

$$\forall x \in U,\ \|f(x) - \mathrm{D}f(x_*)(x - x_*)\|_2 \leq \frac{L}{2}\,{\|x - x_*\|_2}^2, \tag{19}$$

$$\forall x \in U,\ \|g(x) - x_* - \mathrm{D}g(x_*)(x - x_*)\|_2 \leq \frac{L'}{2}\,{\|x - x_*\|_2}^2. \tag{20}$$

We may also impose that U be a tubular neighbourhood of $\Sigma \cap U$ containing x_* and such that the formula

$$P_\Sigma(x) = \operatorname*{arg\,min}_{y \in \Sigma} \|x - y\|_2$$

defines a smooth nonlinear projection operator P_Σ from U to $\Sigma \cap U$ (see *e.g.* [53]). Let $P_{T_{x_*}\Sigma}$ be the orthogonal projector onto the tangent space $T_{x_*}\Sigma$ to Σ at x_*. For U small enough, there exists a positive constant M such that it holds:

$$\forall x \in U,\ \|P_\Sigma(x) - (x_* + P_{T_{x_*}\Sigma}(x - x_*))\|_2 \leq \frac{M}{2}{\|x - x_*\|_2}^2. \tag{21}$$

Note that the introduction of the submanifold Σ is needed in view of applications to quantum chemistry, as presented in Section 4; the above assumptions will be discussed in such a context.

When the submanifold Σ is not an affine subspace of $\mathbb{R}^n$, one may observe that the iterate $x^{(k+1)}$ generated by version A of the algorithms may in general lie outside of Σ, which does not seem appropriate. This is the main reason for the introduction of version P, starting with the CDIIS in quantum chemistry (see [47]). On the contrary, when Σ is an affine subspace of $\mathbb{R}^n$, it is isometric to $\mathbb{R}^{n'}$ for some integer n' such that $n' \leq n$, and all the iterates of both versions lie in Σ, so there is no difference with the case $\Sigma = \mathbb{R}^n$. The theoretical results of the present paper are thus stated for version P of the algorithms when Σ is arbitrary, and for version A assuming that $\Sigma = \mathbb{R}^n$.

As seen in Section 2, the usual abstract framework for the DIIS and/or the Anderson acceleration is a special case of ours: there is no submanifold (that is, one takes $\Sigma = \mathbb{R}^n$), the integers p and n are the same, and the functions f and g are related by (6). In this setting, Assumptions 1 and 2 are satisfied as soon as the norm of $\mathrm{D}g(x_*)$ is strictly lower than one. The theoretical results of the present paper are, of course, still valid and, as far as we know, new in this more restrictive setting.

Before stating the convergence results, let us first recall the terminology associated with the different types of linear convergence (see also Appendix A of [41]). Let $(x^{(k)})_{k\in\mathbb{N}}$ be a sequence in a normed vector space E, equipped with the norm $\|\cdot\|$, that converges to some x_* in E. The convergence is said to be (at least) *q-linear* with rate μ in $(0,1)$ if

$$\frac{\|x^{(k+1)} - x_*\|}{\|x^{(k)} - x_*\|} \leq \mu, \text{ for all } k \text{ sufficiently large.}$$

In addition, the convergence of the sequence is said to be (at least) *r-linear* with rate μ in $(0,1)$ if there exists a positive constant C such that

$$\|x^{(k)} - x_*\| \leq C\,\mu^k, \text{ for all } k \text{ sufficiently large.}$$

In dealing with the convergence of the acceleration techniques considered in the present work, we shall say that a method is *locally* q-linearly (resp. r-linearly) convergent if, for $\|r^{(0)}\|_2$ small enough, the sequence of error vectors $(r^{(k)})_{k\in\mathbb{N}}$ converges at least q-linearly (resp. r-linearly) to the zero vector in $\mathbb{R}^p$. Note that, from Assumption 2, this implies that the sequence of iterates $(x^{(k)})_{k\in\mathbb{N}}$ converges at least r-linearly to x_* in $\mathbb{R}^n$ with the same rate.

Our first result deals with the convergence of the restarted Anderson–Pulay acceleration method.

Theorem 3.1 *Let Assumptions 1 and 2 hold and let μ be a real number such that $K < \mu < 1$. There exists a positive constant R_μ such that, for any choice of the restart parameter τ in the interval $(0,1)$, if the initial point $x^{(0)}$ in $U \cap \Sigma$ satisfies $0 < \|r^{(0)}\|_2 \leq R_\mu \tau^{2p}$ and one runs version P of Algorithm 3 (or version A in the case $\Sigma = \mathbb{R}^n$), the sequences $(x^{(k)})_{k\in\mathbb{N}}$ and $(r^{(k)})_{k\in\mathbb{N}}$ are well-defined and, as long as $\|r^{(k)}\|_2$ is nonzero, one has*

$$m_k \leq \min(k, p), \tag{22}$$

$$\|c^{(k)}\|_\infty \leq C_{m_k}\left(1 + \frac{1}{(\tau(1-\mu))^{m_k}}\right), \tag{23}$$

where C_{m_k} is a positive constant depending only on m_k, and

$$\|r^{(k+1)}\|_2 \leq \mu^{m_k+1}\|r^{(k-m_k)}\|_2. \tag{24}$$

As a consequence, one has

$$\forall k \in \mathbb{N},\ \|r^{(k)}\|_2 \leq \mu^k \|r^{(0)}\|_2, \tag{25}$$

meaning that the method is locally r-linearly convergent.
Moreover, there exists a positive constant Γ, independent of the parameter τ, such that, if a restart occurs at step $k+1$ (that is, if $m_{k+1} = 0$), one has

$$(1 - K(1+\tau))\,\|r^{(k+1)}\|_2 \leq \left(K\tau + \frac{\Gamma}{\tau^{2m_k}}\|r^{(k-m_k)}\|_2\right)\|r^{(k-m_k)}\|_2. \tag{26}$$

Our second theorem deals with the convergence of the adaptive-depth variant of the Anderson–Pulay acceleration method.

Theorem 3.2 *Let Assumptions 1 and 2 hold and let μ be a real number such that $K < \mu < 1$. There exists a positive constant c_μ such that, for any choice of the parameter δ in the interval $(0, K)$, if the initial point $x^{(0)}$ in $U \cap \Sigma$ satisfies $0 < \|r^{(0)}\|_2 \leq c_\mu \delta^2$ and one runs version P of Algorithm 4 (or version A in the case $\Sigma = \mathbb{R}^n$), the sequences $(x^{(k)})_{k\in\mathbb{N}}$ and $(r^{(k)})_{k\in\mathbb{N}}$ are well-defined and, as long as $\|r^{(k)}\|_2 > 0$, one has*

$$m_k \leq \min(k, p), \tag{27}$$

$$\|c^{(k)}\|_\infty \leq C_{m_k}\left(1 + \left(\frac{\mu}{1-\mu}\right)^{m_k}\right), \tag{28}$$

where C_{m_k} is a positive constant depending only on m_k, and

$$\|r^{(k+1)}\|_2 \leq \mu \|r^{(k)}\|_2. \tag{29}$$

The method is thus locally q-linearly convergent.
Moreover, if $k - m_k \geq 1$, then one has

$$\|r^{(k)}\|_2 \leq \delta \|r^{(k-m_k-1)}\|_2. \tag{30}$$

In both theorems, we have given an upper bound on the extrapolation coefficients $c^{(k)}$ that depends on m_k (see (23) and (28), respectively). Using the fact that m_k is less than or equal to p, one immediately infers from these inequalities that the coefficients $c^{(k)}$ are bounded independently of k, which is sufficient to prove convergence. Nevertheless, this uniform estimate is rough, as it is observed in practical calculations that the depth is generally much smaller than the dimension p.

In addition to the local linear convergence estimates (24) and (29), we have also provided some technical estimates on the residual – inequalities (26) and (30), respectively – which are crucial for the proofs of the acceleration properties in the next subsection.

Theorems 3.1 and 3.2 are proved in Subsection 3.5. Note that for version P, inequality (20) is not used in the proofs and one just needs the function g to be of class $\mathscr{C}^1$. For version A, on the contrary, (20) is needed, but, of course, (21) is not, since it is assumed that $\Sigma = \mathbb{R}^n$.

3.3 Acceleration

In this subsection, we study from a mathematical viewpoint the local acceleration properties of the proposed variants of the Anderson–Pulay acceleration. We prove that, for any choice of the real number λ in $(0, K)$, r-linear convergence with rate λ can be achieved, meaning that $\|r^{(k)}\|_2 = O(\lambda^k)$, with Algorithm 3 (for τ and $\|r^{(0)}\|_2$ small enough) or with Algorithm 4 (for δ and $\|r^{(0)}\|_2$ small enough). These results are direct consequences of Theorems 3.1 and 3.2, respectively.

Corollary 3.3 *Let Assumptions 1 and 2 hold and let λ and μ be two real numbers such that $0 < \lambda < K < \mu < 1$. Let τ in $(0,1)$ satisfy the smallness constraint $\tau \leq \frac{1-K}{3K}\lambda^{p+1}$. Assume that the initial point $x^{(0)}$ in $U \cap \Sigma$ satisfies $0 < \|r^{(0)}\|_2 \leq \min\{R_\mu, \frac{(1-K)\lambda^{p+1}}{3\Gamma}\}\tau^{2p}$, where R_μ and Γ are the positive constants of Theorem 3.1. If one runs version P of Algorithm 3 (or version A, assuming that $\Sigma = \mathbb{R}^n$), then the sequences $(x^{(k)})_{k\in\mathbb{N}}$ and $(r^{(k)})_{k\in\mathbb{N}}$ are well-defined, with $m_k \leq \min(k, p)$ and*

$$\forall k \in \mathbb{N},\ \|r^{(k)}\|_2 \ \leq\ \mu^{m_k}\lambda^{k-m_k}\|r^{(0)}\|_2 \ \leq\ \mu^{\min(k,p)}\lambda^{\max(0,k-p)}\|r^{(0)}\|_2.$$

As a consequence, the sequence $(x^{(k)})_{k\in\mathbb{N}}$ converges at least r-linearly to x_* with rate λ.

PROOF. First of all, since $0 < \|r^{(0)}\|_2 \leq R_\mu \tau^{2p}$, the conclusions of Theorem 3.1 hold. In particular, the sequences $(x^{(k)})_{k\in\mathbb{N}}$ and $(r^{(k)})_{k\in\mathbb{N}}$ are well-defined, with $m_k \leq \min\{k, p\}$ and $\|r^{(k-m_k)}\|_2 \leq \mu^k \|r^{(0)}\|_2$. From the other smallness condition on $\|r^{(0)}\|_2$, we also have $\|r^{(k-m_k)}\|_2 \leq \frac{(1-K)\lambda^{p+1}}{3\Gamma}\tau^{2p} \leq \frac{(1-K)\lambda^{p+1}}{3\Gamma}\tau^{2m_k}$. Finally, the parameter τ is such that $1-K(1+\tau) \geq \frac{2}{3}(1-K)$ and $K\tau \leq \frac{(1-K)\lambda^{p+1}}{3}$. Using these facts, we infer the following property from estimate (26) of Theorem 3.1:

$$\text{if } m_{k+1} = 0, \text{ then } \|r^{(k+1)}\|_2 \leq \lambda^{p+1} \|r^{(k-m_k)}\|_2. \tag{31}$$

Using (24) and (31), we are going to prove by complete induction that, for any natural integer k, there holds

$$\|r^{(k)}\|_2 \leq \mu^{m_k} \lambda^{k-m_k} \|r^{(0)}\|_2.$$

Let us denote by $\mathcal{P}(k)$ the above property at rank k. Obviously, $\mathcal{P}(0)$ is true. Let us assume that the property is satisfied up to rank k, with k a natural integer. We will establish that $\mathcal{P}(k+1)$ holds by distinguishing two cases.

First, if $m_{k+1} \geq 1$, then, using inequality (24), we may write that

$$\|r^{(k+1)}\|_2 \leq \mu^{m_{k+1}} \|r^{(k+1-m_{k+1})}\|_2.$$

The natural integer $k+1-m_{k+1}$ being less than or equal to k, property $\mathcal{P}(k+1-m_{k+1})$ holds and, since $m_i = 0$, it takes the form

$$\|r^{(k+1-m_{k+1})}\|_2 \leq \lambda^{k+1-m_{k+1}} \|r^{(0)}\|_2.$$

We then conclude by combining the two above estimates.

Second, if $m_{k+1} = 0$, remembering that $m_k \leq p$ and $\lambda < 1$, we infer from inequality (31) that

$$\|r^{(k+1)}\|_2 \leq \lambda^{p+1} \|r^{(k-m_k)}\|_2 \leq \lambda^{m_k+1} \|r^{(k-m_k)}\|_2.$$

The integer $k-m_k$ being less than or equal to k, property $\mathcal{P}(k-m_k)$ holds and, since $m_{k-m_k} = 0$, it follows that

$$\|r^{(k-m_k)}\|_2 \leq \lambda^{k-m_k} \|r^{(0)}\|_2,$$

so that one reaches

$$\|r^{(k+1)}\|_2 \leq \lambda^{m_k+1} \lambda^{k-m_k} \|r^{(0)}\|_2 = \lambda^{k+1} \|r^{(0)}\|_2.$$

Estimate $\mathcal{P}(k)$ thus holds for any natural integer k. Moreover, since $m_k \leq \min\{k, p\}$ and $\lambda < \mu$, we immediately see that $\lambda^{k-m_k}\mu^{m_k} < \lambda^{\max(0,k-p)}\mu^{\min(k,p)}$, which ends the proof. □

Note that our theoretical smallness requirements on τ and $\|r^{(0)}\|_2$ are unreasonably restrictive and certainly not representative of what is observed in applications. We will indeed see in Section 5 that acceleration is commonly achieved in practice.

A similar result for acceleration with Algorithm 4 follows from Theorem 3.2.

Corollary 3.4 *Let Assumptions 1 and 2 hold and let λ and μ be two real numbers such that $0 < \lambda < K < \mu < 1$. Choose δ so that $0 < \delta \leq \lambda^{p+1}$. Suppose that the initial point $x^{(0)} \in U \cap \Sigma$ satisfies $0 < \|r^{(0)}\|_2 \leq c_\mu \delta^2$ where c_μ is the same as in Theorem 3.2, and run version P of Algorithm 4 (or version A in the case $\Sigma = \mathbb{R}^n$). Denote $k_{\max} = \max\{k \in \mathbb{N} \,|\, m_k = k\}$. Then one has $k_{\max} \leq p$ and, for all k in $\mathbb{N}$,*

$$\|r^{(k)}\|_2 \leq \mu^{\min(k,k_{\max})} \lambda^{\max(0,k-k_{\max})} \|r^{(0)}\|_2 \leq \mu^{\min(k,p)} \lambda^{\max(0,k-p)} \|r^{(0)}\|_2.$$

As a consequence, $(x^{(k)})_{k\in\mathbb{N}}$ converges at least r-linearly to x_* with rate λ.

PROOF. Since $\delta < \lambda < K$ and $0 < \|r^{(0)}\|_2 \leq c_\mu \delta^2$, the conclusions of Theorem 3.2 hold. In particular, from bound (27), one has $k_{\max} \leq p$, hence $\lambda^{\max(0,k-k_{\max})} \leq \mu^{\min(k,p)} \lambda^{\max(0,k-p)}$ for any natural integer k. As a result, we just need to prove that the following estimate

$$\|r^{(k)}\|_2 \leq \mu^{\min(k,k_{\max})} \lambda^{\max(0,k-k_{\max})} \|r^{(0)}\|_2 \tag{32}$$

holds for any natural integer k. To do this, we argue by complete induction.

On the one hand, if $k \leq k_{\max}$, inequality (32) reduces to an estimate of r-linear convergence with rate μ, which follows immediately from the q-linear convergence property (29) of Theorem 3.2.

On the other hand, let $k \geq k_{\max}$ be such that (32) holds for all the natural integers less than or equal to k. From the definition of $k_{\max}$, one has $k+1-m_{k+1} \geq 1$, and, since $\delta \leq \lambda^{p+1}$, $m_{k+1} \leq p$ and $\lambda < 1$, estimate (30) implies that

$$\|r^{(k+1)}\|_2 \leq \lambda^{m_{k+1}+1} \|r^{(k-m_{k+1})}\|_2.$$

The natural integer $k - m_{k+1}$ being less than or equal to k, one also has

$$\|r^{(k-m_{k+1})}\|_2 \leq \mu^{\min(k-m_{k+1},k_{\max})}\lambda^{\max(0,k-m_{k+1}-k_{\max})}\|r^{(0)}\|_2,$$

and, combining the last two estimates, one gets

$$\|r^{(k+1)}\|_2 \leq \mu^{\min(k-m_{k+1},k_{\max})}\lambda^{\max(m_{k+1}+1,k+1-k_{\max})}\|r^{(0)}\|_2.$$

In order to study the right-hand side of this inequality, we distinguish two cases.

If $k - m_{k+1} > k_{\max}$, then it holds $\mu^{\min(k-m_{k+1},k_{\max})}\lambda^{\max(m_{k+1}+1,k+1-k_{\max})} = \mu^{k_{\max}}\lambda^{k+1-k_{\max}}$, so that estimate (32) holds for $k+1$. Otherwise, if $k - m_{k+1} \leq k_{\max}$, then it holds $\mu^{\min(k-m_{k+1},k_{\max})}\lambda^{\max(m_{k+1}+1,k+1-k_{\max})} = \mu^{k-m_{k+1}}\lambda^{m_{k+1}+1}$. Since $\lambda < \mu$, one has $\mu^{k-m_{k+1}}\lambda^{m_{k+1}+1} \leq \mu^{k_{\max}}\lambda^{k+1-k_{\max}}$ and estimate (32) holds for $k+1$.

We have thus shown that estimate (32) holds for any natural integer k, ending the proof. □

3.4 Reaching superlinear convergence

Let us recall some standard notions of superlinear convergence (see *e.g.* [41, 43]). We shall say that a sequence $(x^{(k)})_{k\in\mathbb{N}}$ in a normed space E, equipped with the norm $\|\cdot\|$, converges *q-superlinearly* to x_* in E if

$$\lim_{k\to+\infty} \frac{\|x^{(k+1)} - x_*\|}{\|x^{(k)} - x_*\|} = 0.$$

Let θ denote a real number greater than one. It is said that the superlinear convergence occurs with *q-order at least* θ if there exists a positive constant a such that

$$\frac{\|x^{(k+1)} - x_*\|}{\|x^{(k)} - x_*\|^\theta} \leq a, \text{ for all } k \text{ sufficiently large.}$$

More generally, the sequence is said to converge *r-superlinearly* to x_* in E if there exists a sequence of positive numbers $(\epsilon_k)_{k\in\mathbb{N}}$ converging q-superlinearly to zero and such that

$$\|x^{(k)} - x_*\| \leq \epsilon_k, \text{ for all } k \text{ sufficiently large.}$$

If one has additionally

$$\forall k \in \mathbb{N},\ \epsilon_k = b\,\eta^{\theta^k}$$

for some positive constant b and η in $(0,1)$, the sequence is said to converge superlinearly with *r-order at least* θ.

In some references, one can find mentions of the DIIS exhibiting a local superlinear convergence behaviour. Rohwedder and Schneider discussed in [48] the circumstances under which this may occur for the DIIS method. We now propose a modification of the restarted (resp. adaptive-depth) Anderson–Pulay acceleration, in which the value of the parameter τ (resp. δ) is slowly decreased along the iteration. We will show that the resulting sequence of approximations locally converges r-superlinearly to the fixed point.

Let us first describe the modification made to Algorithm 3. One may observe that estimate (26) in Theorem 3.1 allows a local r-superlinear convergence result, by carefully changing the value of the parameter τ after each restart. As a consequence, we modify the algorithm as follows: first, in the list of data, the positive constant τ is replaced by two positive constants T_0 and ζ ; then, we replace the boolean test "$\tau \left\|s^{(k+1)}\right\|_2 > \left\|(\mathrm{id} - \Pi_k)s^{(k+1)}\right\|_2$" by "$T_0\left\|r^{(k-m_k)}\right\|_2^{\ \zeta} \left\|s^{(k+1)}\right\|_2 > \left\|(\mathrm{id} - \Pi_k)s^{(k+1)}\right\|_2$". In other words, the fixed parameter τ in the test is replaced by an adaptive one, $\tau^{(k-m_k)} = T_0\left\|r^{(k-m_k)}\right\|_2^{\ \zeta}$, which becomes smaller and smaller along the iteration. Naming this variant Algorithm 3', we have the following result.

Theorem 3.5 *Let Assumptions 1 and 2 hold and let μ be a real number such that $K < \mu < 1$. Choose a positive T_0 and ζ in $(0, \frac{1}{2p})$. Then, there exists a positive constant $\varepsilon(T_0, \zeta, \mu)$ such that, if we run version P of Algorithm 3' (or version A in the case $\Sigma = \mathbb{R}^n$) with an initial point $x^{(0)}$ in $U \cap \Sigma$ satisfying $0 < \|r^{(0)}\|_2 < \varepsilon(T_0, \zeta, \mu)$, then the sequence $(x^{(k)})_{k\in\mathbb{N}}$ is well-defined, satisfies estimates* (22) *and* (25)*, and converges superlinearly to x_* with r-order at least $(1 + \min\{\zeta, 1 - 2p\zeta\})^{\frac{1}{p}}$.*

For Algorithm 4, the changes are the following: first, in the list of data, we replace the positive constant δ by two positive constants D_0 and ξ ; then, we replace the boolean test "$\delta \|r^{(i)}\|_2 < \|r^{(k+1)}\|_2$" by "$D_0 \left\|r^{(i)}\right\|_2^{1+\xi} < \|r^{(k+1)}\|_2$". That is to say, the fixed parameter δ in the test is replaced by the adaptive one, $\delta^{(i)} = D_0 \left\|r^{(i)}\right\|_2^{\xi}$. We name this variant Algorithm 4' and state the following result.

Theorem 3.6 *Let Assumptions 1 and 2 hold and let μ be a real number such that $K < \mu < 1$. Choose a positive constant D_0 and ξ in $(0, \sqrt{2}-1]$. In the limit case $\xi = \sqrt{2}-1$, assume in addition that $D_0 \geq \Delta_\mu$, where Δ_μ is a suitable positive constant. Then, there exists a positive constant $\varepsilon(D_0, \xi, \mu)$ such that, if we run version P of Algorithm 4' (or version A in the case $\Sigma = \mathbb{R}^n$) with an initial point $x^{(0)}$ in $U \cap \Sigma$ satisfying $0 < \|r^{(0)}\|_2 < \varepsilon(D_0, \xi, \mu)$, the sequence $(x^{(k)})_{k\in\mathbb{N}}$ is well-defined, satisfies estimates (27), (28), and (29), and converges superlinearly to x_* with r-order at least $(1+\xi)^{\frac{1}{p+1}}$.*

Both theorems 3.5 and 3.6 are proved at the end of Subsection 3.5.

Note that the best theoretical value of the r-order given by Theorem 3.5 is $\left(1+\frac{1}{2p+1}\right)^{\frac{1}{p}}$. It corresponds to the choice $\zeta = \frac{1}{2p+1}$. For Theorem 3.6, the best theoretical value is $2^{\frac{1}{2p+2}}$, corresponding to $\xi = \sqrt{2}-1$. In both cases, it is very close to one when p is large.

In the numerical experiments reported in the last part of the paper, we only ran Algorithms 3 and 4 with fixed values of τ and δ and obtained excellent acceleration results for self-consistent field iterations. Nevertheless, Theorems 3.5 and 3.6 suggest that it may be beneficial in practical calculations to decrease the value of the parameters τ and δ in an adaptive way along the iteration. We plan to investigate this in the future.

3.5 Proofs of the theorems

For linear problems in finite dimension, if g is a contraction, then it is well-known that the DIIS (or the Anderson acceleration) converges to the exact solution in a finite number of steps (see Subsection 2.4). In the nonlinear case, this is no longer true due to the presence of quadratic terms. Despite these additional terms, convergence at any given linear rate should be allowed, on the condition that the depth is large enough and one starts close enough to the solution. Then, if one allows the depth to grow slowly along the process, superlinear convergence should occur.

To prove these properties, one has to control the quadratic errors and a bound on the size of the extrapolation coefficients at each step is needed. This amounts to quantitatively measuring the affine independence of the error vectors, since the optimal coefficients are solution to a least-squares problem whose solubility is directly related to this independence.

In the existing literature, it is generally[7] assumed that the coefficients remain bounded throughout the iteration. While we do not know how to prove *a priori* such an estimate for the "classical" fixed-depth Anderson–Pulay acceleration, the mechanisms employed in the restarted and adaptive-depth variants we study allow to derive one. A key theoretical ingredient is that the problem for the extrapolation coefficients will become poorly conditioned when the norm of the last stored error vector is much smaller than that of the oldest one. This phenomenon is rigorously described in Lemma 3.7 below, which constitutes the main technical tool in our proofs of both convergence and acceleration of the method.

In the restarted Anderson–Pulay acceleration, the role of the parameter τ is to control the affine independence of the error vectors. Lemma 3.7 shows that, as long as a restart does not occur, the least-squares problem remains well-conditioned and the size of the coefficients is bounded, whereas the norm of the last stored error vector is necessarily much smaller than that of the oldest one when it does. Since there must be a restart after at most $p+1$ consecutive iterations, convergence with acceleration can be established.

In the adaptive-depth version of the Anderson–Pulay acceleration, the parameter δ is used to eliminate the stored iterates which are not "relevant" enough when compared with the most recent one. This criterion does not directly quantify the independence of the error vectors, and it is certainly not good when the initial guess is chosen far from the solution, since a large number of iterates will be kept in that case. However, starting close enough to the solution, Lemma 3.7 allows to inductively prove a bound on the extrapolation coefficients, for reasons similar to those invoked with the restarted variant. Indeed, at each step, either the stored error vectors are affinely independent, and the extrapolation coefficients are bounded, or the norm of the last stored error vector is smaller than δ times that of some of the oldest ones. In the latter case, the criterion discards the corresponding iterates, which results in a restoration of the condition number associated with the least-squares problem at the next step. In addition, observing that a stored iterate is dismissed at least once in every $p+1$ consecutive iterations, it is inferred that accelerated convergence is possible for a parameter δ chosen small enough.

3.5.1 A preliminary lemma

We first introduce some notations. We recall that the set U, introduced in Subsection 3.2, is a small neighborhood of x_* such that the estimates (18), (19), (20), and (21) hold. For any natural integers k and m_k such that $k \geq m_k \geq 1$,

[7] An exception is made in the paper by Zhang *et al.* [63], where a bound on the coefficients is shown for a restarted type-I Anderson acceleration method (the DIIS rather corresponds to a type-II Anderson acceleration). The authors then prove a global linear convergence result when f is the gradient of a convex functional, but the linear rate they obtain is no better than the rate of the basic iteration process.

consider a family $(x^{(k-m_k)},\dots,x^{(k)})$ of vectors in $U\cap\Sigma\setminus\{x_*\}$. For any integer i in $\{0,\dots,m_k\}$, set $r^{(k-m_k+i)}=f(x^{(k-m_k+i)})$, and, for any integer ℓ in $\{0,\dots,m_k\}$, let us denote by $\mathcal{A}_\ell^{(k)}$ the affine span of $\{r^{(k-m_k)},\dots,r^{(k-m_k+\ell)}\}$, that is

$$\mathcal{A}_\ell^{(k)}=\mathrm{Aff}\left\{r^{(k-m_k)},r^{(k-m_k+1)},\dots,r^{(k-m_k+\ell)}\right\}=\left\{r=\sum_{i=0}^{\ell}c_i\,r^{(k-m_k+i)}\mid\sum_{i=0}^{\ell}c_i=1\right\},$$

and, for any integer ℓ in $\{1,\dots,m_k\}$, by $d_\ell^{(k)}$ the distance between $r^{(k-m_k+\ell)}$ and $\mathcal{A}_{\ell-1}^{(k)}$, that is

$$d_\ell^{(k)}=\min_{r\in\mathcal{A}_{\ell-1}^{(k)}}\|r^{(k-m_k+\ell)}-r\|_2.$$

Lemma 3.7 *Let k and m_k be two integers such that $k\geq m_k\geq 1$, let $(x^{(k-m_k)},\dots,x^{(k)})$ be a family of vectors in $U\cap\Sigma\setminus\{x_*\}$ and t be a positive real number. For any integer i in $\{0,\dots,m_k\}$, set $r^{(k-m_k+i)}=f(x^{(k-m_k+i)})$. Assume that*

$$\forall i\in\{1,\dots,m_k\},\ d_i^{(k)}\geq t\max_{j\in\{i,\dots,m_k\}}\|r^{(k-m_k+j)}\|_2. \tag{33}$$

Then, the vectors $r^{(k-m_k)},\dots,r^{(k)}$ are affinely independent, so that one has $m_k\leq p$ and there is a unique $\tilde{c}$ in $\mathbb{R}^{m_k+1}$ such that $\sum_{i=0}^{m_k}\tilde{c}_i=1$ and $\|\sum_{i=0}^{m_k}\tilde{c}_i r^{(k-m_k+i)}\|_2=\mathrm{dist}_2(0,\mathcal{A}_{m_k}^{(k)})$. Moreover, there exists a positive constant C_{m_k} such that

$$\|\tilde{c}\|_\infty\leq C_{m_k}\,(1+t^{-m_k}). \tag{34}$$

Under the additional assumption that $\tilde{x}^{(k+1)}=\sum_{i=0}^{m_k}\tilde{c}_i\,x^{(k-m_k+i)}$ belongs to U, as well as $g(\tilde{x}^{(k+1)})$ and $\sum_{i=0}^{m_k}\tilde{c}_i\,g(x^{(k-m_k+i)})$, define

$$x^{(k+1)}=g(\tilde{x}^{(k+1)})\ \ \textit{(version P)}\quad\textit{or}\quad x^{(k+1)}=\sum_{i=0}^{m_k}\tilde{c}_i\,g(x^{(k-m_k+i)})\ \ \textit{(version A)}$$

and set $r^{(k+1)}=f(x^{(k+1)})$, $d_{m_k+1}^{(k)}=\mathrm{dist}_2(r^{(k+1)},\mathcal{A}_{m_k}^{(k)})$. Then, there exists a positive constant κ such that

$$\|r^{(k+1)}\|_2\leq K\,\mathrm{dist}_2(0,\mathcal{A}_{m_k}^{(k)})+\kappa(1+t^{-2m_k})\max_{i\in\{0,\dots,m_k\}}{\|r^{(k-m_k+i)}\|_2}^2. \tag{35}$$

and

$$(1-K)\,\|r^{(k+1)}\|_2\leq K\,d_{m_k+1}^{(k)}+\kappa(1+t^{-2m_k})\max_{i\in\{0,\dots,m_k\}}{\|r^{(k-m_k+i)}\|_2}^2. \tag{36}$$

Proof. Set $s^{(i)}=r^{(k-m_k+i)}-r^{(k-m_k+i-1)}$, for any integer i in $\{1,\dots,m_k\}$, $q^{(1)}=s^{(1)}$ and $q^{(i)}=(\mathrm{id}-\Pi_{\boldsymbol{\mathcal{A}}_{i-1}^{(k)}})s^{(i)}$, for any integer i in $\{2,\dots,m_k\}$, where $\Pi_{\boldsymbol{\mathcal{A}}_{i-1}^{(k)}}$ denotes the orthogonal projector onto $\boldsymbol{\mathcal{A}}_{i-1}^{(k)}$, the underlying vector space of $\mathcal{A}_{i-1}^{(k)}$. Then, the vectors $q^{(1)},\dots,q^{(m_k)}$ are mutually orthogonal and, for any integer i in $\{1,\dots,m_k\}$, one has $\|q^{(i)}\|_2=d_i^{(k)}$.
We may write

$$\sum_{i=0}^{m_k}\tilde{c}_i\,r^{(k-m_k+i)}=r^{(k)}+\sum_{i=1}^{m_k}\tilde{\zeta}_i s^{(i)}=r^{(k)}+\sum_{i=1}^{m_k}\tilde{\lambda}_i q^{(i)},$$

with $\tilde{\lambda}_i=-\frac{(q^{(i)})^\top r^{(k)}}{(d_i^{(k)})^2}$, so that $\left|\tilde{\lambda}_i\right|\leq\frac{\|r^{(k)}\|_2}{d_i^{(k)}}\leq\frac{1}{t}$, due to lower bound (33).

On the other hand, $\tilde{\zeta}=P\tilde{\lambda}$, where P is the change-of-basis matrix from $\{s^{(i)}\}_{i=1,\dots,m_k}$ to $\{q^{(i)}\}_{i=1,\dots,m_k}$. We need a bound on $\|P\|$. For that purpose, we introduce the matrix factorisation $P=P^{(m_k)}P^{(m_k-1)}\dots P^{(2)}$ where, for any integer j in $\{2,\dots,m_k\}$, $P^{(j)}$ is the change-of-basis matrix from $\{q^{(1)},\dots,q^{(j-1)},s^{(j)},\dots,s^{(m_k)}\}$ to $\{q^{(1)},\dots,q^{(j)},s^{(j+1)},\dots,s^{(m_k)}\}$. Since $q^{(1)}=s^{(1)}$, and $q^{(j)}=s^{(j)}-\sum_{i=1}^{j-1}\frac{(s^{(j)})^\top q^{(i)}}{(q^{(i)})^\top q^{(i)}}\,q^{(i)}$, we have

$$\forall i\in\{1,\dots,m_k\},\ (P^{(j)})_{ii}=1\text{ and, }\forall i\in\{1,\dots,j-1\},\ (P^{(j)})_{ij}=-\frac{(s^{(j)})^\top q^{(i)}}{(q^{(i)})^\top v^{(i)}},$$

all the other coefficients of this matrix being zero. It follows that

$$\forall i\in\{1,\dots,j-1\},\ \left|(P^{(j)})_{ij}\right|\leq\frac{\|s^{(j)}\|_2}{d_i^{(k)}}\leq\frac{\|r^{(k-m_k+j)}\|_2+\|r^{(k-m_k+j-1)}\|_2}{d_i^{(k)}}\leq\frac{2}{t}.$$

As a consequence, there holds an estimate of the form

$$\|P^{(j)}\|_2 \leq C(1+t^{-1})$$

for some positive constant C depending only on m_k, which thus yields

$$\|P\| \leq \|P^{(m_k)}\|\|P^{(m_k-1)}\|\dots\|P^{(2)}\| \leq C^{m_k-1}(1+t^{-1})^{m_k-1}.$$

Hence, it follows that

$$\|\tilde{\zeta}\|_\infty \leq C^{m_k-1}(1+t^{-1})^{m_k-1}\|\tilde{\lambda}\|_\infty \leq C^{m_k-1}(1+t^{-1})^{m_k-1}t^{-1}.$$

Finally, one has $\tilde{c}_0 = -\tilde{\zeta}_1$, $\tilde{c}_i = \tilde{\zeta}_i - \tilde{\zeta}_{i+1}$, $1 \leq i \leq m_k-1$, and $\tilde{c}_{m_k} = 1+\tilde{\zeta}_{m_k}$, so that $\|\tilde{c}\|_\infty \leq C_{m_k}(1+t^{-m_k})$, for some positive constant C_{m_k} depending only on m_k, thus proving bound (34) on the coefficients.

Next, we remark that estimate (36) is an immediate consequence of (35). Indeed, by the triangle inequality, one has

$$\mathrm{dist}_2(0,\mathcal{A}_{m_k}^{(k)}) \leq \|r^{(k+1)}\|_2 + d_{m_k+1}^{(k)}.$$

It only remains to prove the bound (35) on $\|r^{(k+1)}\|_2$. As a first step, we are going to estimate the norm of $\tilde{r}^{(k+1)} = f(\tilde{x}^{(k+1)})$. One has

$$\|\tilde{r}^{(k+1)} - \sum_{i=0}^{m_k}\tilde{c}_i r^{(k-m_k+i)}\|_2 \leq \|f(\tilde{x}^{(k+1)}) - \mathrm{D}f(x_*)(\tilde{x}^{(k+1)}-x_*)\|_2 + \|\sum_{i=0}^{m_k}\tilde{c}_i f(x^{(k-m_k+i)}) - \mathrm{D}f(x_*)(\tilde{x}^{(k+1)}-x_*)\|_2$$

so that, using inequality (19) and the fact that the coefficients $(\tilde{c}_i)_{0\leq i\leq m_k}$ sum to 1,

$$\begin{aligned}\|\tilde{r}^{(k+1)} - \sum_{i=0}^{m_k}\tilde{c}_i r^{(k-m_k+i)}\|_2 &\leq \frac{L}{2}\|\sum_{i=0}^{m_k}\tilde{c}_i(x^{(k-m_k+i)}-x_*)\|_2^2 + \|\sum_{i=0}^{m_k}\tilde{c}_i(f(x^{(k-m_k+i)}) - \mathrm{D}f(x_*)(x^{(k-m_k+i)}-x_*))\|_2 \\ &\leq \frac{L}{2}(m_k+1)\|\tilde{c}\|_\infty((m_k+1)\|\tilde{c}\|_\infty+1)\max_{i\in\{0,\dots,m_k\}}\|x^{(k-m_k+i)}-x_*\|_2{}^2.\end{aligned}$$

It thus follows from the definition of the coefficients $\tilde{c}_i$ that

$$\begin{aligned}\|\tilde{r}^{(k+1)}\|_2 &\leq \|\tilde{r}^{(k+1)} - \sum_{i=0}^{m_k}\tilde{c}_i\, r^{(k-m_k+i)}\|_2 + \mathrm{dist}_2(0,\mathcal{A}_{m_k}^{(k)}) \\ &\leq \frac{L}{2}(m_k+1)\|\tilde{c}\|_\infty((m_k+1)\|\tilde{c}\|_\infty+1)\max_{i\in\{0,\dots,m_k\}}\|x^{(k-m_k+i)}-x_*\|_2{}^2 + \mathrm{dist}_2(0,\mathcal{A}_{m_k}^{(k)}).\end{aligned}$$

Using bound (34), we find a constant $C'_{m_k} > 0$ independent of t, such that

$$(m_k+1)\|\tilde{c}\|_\infty((m_k+1)\|\tilde{c}\|_\infty+1) \leq C'_{m_k}(1+t^{-2m_k}). \tag{37}$$

So, using Assumption 2, we finally obtain

$$\|\tilde{r}^{(k+1)}\|_2 \leq \frac{L}{2\sigma^2}C'_{m_k}(1+t^{-2m_k})\max_{i\in\{0,\dots,m_k\}}\|r^{(k-m_k+i)}\|_2{}^2 + \mathrm{dist}_2(0,\mathcal{A}_{m_k}^{(k)}). \tag{38}$$

In the remaining part of the proof, we treat separately version P and version A.

- **Version P.** Recall that, in this version, the set Σ is an arbitrary smooth submanifold and one takes $x^{(k+1)} = g(\tilde{x}^{(k+1)})$. An estimate on the distance between $\tilde{x}^{(k+1)}$ and Σ is needed. By the triangle inequality, one has

  $$\|\tilde{x}^{(k+1)} - P_\Sigma(\tilde{x}^{(k+1)})\|_2 \leq \|\tilde{x}^{(k+1)} - x_* - P_{T_{x_*}\Sigma}(\tilde{x}^{(k+1)}-x_*)\|_2 + \|x_* + P_{T_{x_*}\Sigma}(\tilde{x}^{(k+1)}-x_*) - P_\Sigma(\tilde{x}^{(k+1)})\|_2,$$

  and we will thus get bounds for both contributions in the right-hand side of this inequality. For the first one,

remembering that $x^{(k-m_k+i)} = P_\Sigma(x^{(k-m_k+i)})$ for any integer i in $\{0,\dots,m_k\}$ and using (21), we reach

$$\begin{aligned}
\|\tilde{x}^{(k+1)} - x_* - P_{T_{x_*}\Sigma}(\tilde{x}^{(k+1)} - x_*)\|_2 &= \|\sum_{i=0}^{m_k} \tilde{c}_i \,(\mathrm{id} - P_{T_{x_*}\Sigma})(x^{(k-m_k+i)} - x_*)\|_2 \\
&\leq \sum_{i=0}^{m_k} |\tilde{c}_i| \,\|(\mathrm{id} - P_{T_{x_*}\Sigma})(x^{(k-m_k+i)} - x_*)\|_2 \\
&= \sum_{i=0}^{m_k} |\tilde{c}_i| \,\|P_\Sigma(x^{(k-m_k+i)}) - x_* - P_{T_{x_*}\Sigma}(x^{(k-m_k+i)} - x_*)\|_2 \\
&\leq \frac{M}{2} \sum_{i=0}^{m_k} |\tilde{c}_i| \,{\|x^{(k-m_k+i)} - x_*\|_2}^2 \\
&\leq \frac{M}{2}(m_k+1)\|\tilde{c}\|_\infty \max_{i\in\{0,\dots,m_k\}} {\|x^{(k-m_k+i)} - x_*\|_2}^2.
\end{aligned}$$

For the second term, we use again (21) to obtain

$$\begin{aligned}
\|x_* + P_{T_{x_*}\Sigma}(\tilde{x}^{(k+1)} - x_*) - P_\Sigma(\tilde{x}^{(k+1)})\|_2 &\leq \frac{M}{2}{\|\tilde{x}^{(k+1)} - x_*\|_2}^2 \\
&\leq \frac{M}{2}(m_k+1)^2\|\tilde{c}\|_\infty^2 \max_{i\in\{0,\dots,m_k\}} {\|x^{(k-m_k+i)} - x_*\|_2}^2.
\end{aligned}$$

Adding these two estimates, using Assumption 2 and bound (37), we find

$$\begin{aligned}
\|\tilde{x}^{(k+1)} - P_\Sigma(\tilde{x}^{(k+1)})\|_2 &\leq \frac{M}{2} C'_{m_k}(1+t^{-2m_k}) \max_{i\in\{0,\dots,m_k\}} {\|x^{(k-m_k+i)} - x_*\|_2}^2 \\
&\leq \frac{M}{2\sigma^2} C'_{m_k}(1+t^{-2m_k}) \max_{i\in\{0,\dots,m_k\}} {\|r^{(k-m_k+i)}\|_2}^2.
\end{aligned}$$

Finally, using Assumption 1 and combining the above estimate with inequalities (18) and (38), we get

$$\begin{aligned}
\|r^{(k+1)}\|_2 &\leq \|f(g(P_\Sigma(\tilde{x}^{(k+1)})))\|_2 + 2\,\|\mathrm{D}f(x_*)\circ\mathrm{D}g(x_*)\|_2\|\tilde{x}^{(k+1)} - P_\Sigma(\tilde{x}^{(k+1)})\|_2 \\
&\leq K\,\|f(P_\Sigma(\tilde{x}^{(k+1)}))\|_2 + 2\,\|\mathrm{D}f(x_*)\circ\mathrm{D}g(x_*)\|_2\|\tilde{x}^{(k+1)} - P_\Sigma(\tilde{x}^{(k+1)})\|_2 \\
&\leq K\,\|\tilde{r}^{(k+1)}\|_2 + 2\,(K\,\|\mathrm{D}f(x_*)\|_2 + \|\mathrm{D}f(x_*)\circ\mathrm{D}g(x_*)\|_2)\,\|\tilde{x}^{(k+1)} - P_\Sigma(\tilde{x}^{(m+1)})\|_2 \\
&\leq K\,\mathrm{dist}_2(0,\mathcal{A}_{m_k}^{(k)}) + \kappa(1+t^{-2m_k}) \max_{i\in\{0,\dots,m_k\}} {\|r^{(k-m_k+i)}\|_2}^2,
\end{aligned}$$

where $\kappa = \frac{1}{\sigma^2}\left(\frac{KL}{2} + KM\,\|\mathrm{D}f(x_*)\|_2 + M\|\mathrm{D}f(x_*)\circ\mathrm{D}g(x_*)\|_2\right) \max_{m\in\{1,\dots,p\}} C'_m$.

This concludes the proof of the Lemma for version P.

- **Version A.** In this version, the set Σ coincides with $\mathbb{R}^n$ and one takes $x^{(k+1)} = \sum_{i=0}^{m_k} \tilde{c}_i g(x^{(k-m_k+i)})$. A bound on the norm of $x^{(k+1)} - g(\tilde{x}^{(k+1)})$ must be obtained, justifying the assumption that g is of class $\mathscr{C}^2$.

  We proceed as for the estimate obtained in the first step, the only difference being the use of (20) instead of (19), ending up getting something very similar, namely

$$\begin{aligned}
\|x^{(k+1)} - g(\tilde{x}^{(k+1)})\|_2 &\leq \frac{L'}{2}(m_k+1)\|\tilde{c}\|_\infty((m_k+1)\|\tilde{c}\|_\infty + 1) \max_{i\in\{0,\dots,m_k\}} {\|x^{(k-m_k+i)} - x_*\|_2}^2 \\
&\leq \frac{L'}{2\sigma^2} C'_{m_k}(1+t^{-2m_k}) \max_{i\in\{0,\dots,m_k\}} {\|r^{(k-m_k+i)}\|_2}^2.
\end{aligned}$$

  Then, using Assumption 1 and combining the above estimate with inequalities (18) and (38) yields

$$\begin{aligned}
\|r^{(k+1)}\|_2 &\leq \|f(g(\tilde{x}^{(k+1)}))\|_2 + 2\,\|\mathrm{D}f(x_*)\|_2\|x^{(k+1)} - g(\tilde{x}^{(k+1)})\|_2 \\
&\leq K\,\|\tilde{r}^{(k+1)}\|_2 + 2\,\|\mathrm{D}f(x_*)\|_2\|\tilde{x}^{(k+1)} - g(\tilde{x}^{(k+1)})\|_2 \\
&\leq K\,\mathrm{dist}_2(0,\mathcal{A}_{m_k}^{(k)}) + \kappa(1+t^{-2m_k}) \max_{i\in\{0,\dots,m_k\}} {\|r^{(k-m_k+i)}\|_2}^2,
\end{aligned}$$

  where $\kappa = \frac{1}{\sigma^2}\left(\frac{KL}{2} + L'\,\|\mathrm{D}f(x_*)\|_2\right) \max_{m\in\{1,\dots,p\}} C'_m$.

  The bound being obtained for version A, the proof is ended.

□

3.5.2 Proof of Theorem 3.1

For any natural integer k, we consider the following properties

$$\text{For any } i \text{ in } \{0,\dots,m_k\},\ x^{(k-m_k+i)} \text{ exists and lies in } U\cap\Sigma\setminus\{x_*\}, \text{ and } \|r^{(k-m_k)}\|_2 \le R_\mu\tau^{2p}. \tag{a_k}$$

$$\text{If } m_k\ge 1, \text{ then, } \forall i\in\{1,\dots,m_k\},\ \|r^{(k-m_k+i)}\|_2 \le \mu^i\|r^{(k-m_k)}\|_2. \tag{b_k}$$

$$\text{If } m_k\ge 1, \text{ then, } \forall \ell\in\{1,\dots,m_k\},\ d_\ell^{(k)} \ge \tau(1-\mu)\|r^{(k-m_k)}\|_2. \tag{c_k}$$

Let us denote by $(\mathcal{P}_k)$ the set of properties (a_k), (b_k), and (c_k) at rank k. Obviously, $(\mathcal{P}_0)$ holds, since $m_0=0$ and $\|r^{(0)}\|_2\le R_\mu\tau^{2p}$. We are going to prove by induction that $(\mathcal{P}_k)$ is true as long as $r^{(k)}$ is nonzero, if R_μ is chosen small enough. Concurrently, bounds (22), (23) and (24) will be proved as consequences of $(\mathcal{P}_k)$.

Let us then assume that $(\mathcal{P}_k)$ holds for some natural integer k. The error vectors $r^{(k-m_k)},\dots,r^{(k)}$ then satisfy condition (33) in Lemma 3.7 for $t=\tau(1-\mu)$. They are affinely independent and bound (22) is satisfied, as well as (23), which is simply bound (34) in Lemma 3.7.
Moreover, Assumptions 1 and 2 together with properties (a_k) and (b_k) imply the bounds $\|x^{(k-m_k+i)}-x_*\|_2\le\sigma^{-1}R_\mu$ and $\|g(x^{(k-m_k+i)})-x_*\|_2\le\sigma^{-1}R_\mu K$ for any integer i in $\{0,\dots,m_k\}$. Combining them with bound (23), we see that, for R_μ small enough, $\tilde{x}^{(k+1)}$ belongs to U as well as $g(\tilde{x}^{(k+1)})$ and $\sum_{i=0}^{m_k}\tilde{c}_i\,g(x^{(k-m_k+i)})$, so that $x^{(k+1)}$ and $r^{(k+1)}$ are well defined (with $x^{(k+1)}$ belonging to $U\cap\Sigma$), and estimates (35) and (36) of Lemma 3.7 hold.

Now, using property (b_k), bound (35) for $t=\tau(1-\mu)$ gives

$$\begin{aligned}\|r^{(k+1)}\|_2 &\le K\,\mathrm{dist}_2(0,\mathcal{A}_k^{(k)}) + \kappa\left(1+\frac{1}{(\tau(1-\mu))^{2m_k}}\right)\|r^{(k-m_k)}\|_2{}^2\\ &\le \left(K\mu^{m_k}+\kappa\left(1+\frac{1}{(\tau(1-\mu))^{2p}}\right)\|r^{(k-m_k)}\|_2\right)\|r^{(k-m_k)}\|_2,\end{aligned}$$

and, to establish (24), we thus need the inequality

$$\frac{K}{\mu}+\kappa\left(1+\frac{1}{(\tau(1-\mu))^{2p}}\right)\frac{R_\mu\tau^{2p}}{\mu^{m_k+1}}\le 1.$$

For this to be true, it suffices to choose the constant R_μ so that

$$\frac{K}{\mu}+\kappa\left(1+\frac{1}{(1-\mu)^{2p}}\right)\frac{R_\mu}{\mu^{p+1}}\le 1.$$

Assuming that $r^{(k+1)}$ is nonzero, we distinguish two cases.

- If $m_{k+1}=0$ (meaning there is a restart at step k), then $k+1-m_{k+1}=k+1$ and, since we have previously shown that (24) holds, it follows from property (a_k) that

  $$\|r^{(k+1-m_{k+1})}\|_2\le\mu^{m_k+1}\|r^{(k-m_k)}\|_2\le\mu^{m_k+1}R_\mu\tau^{2p}\le R_\mu\tau^{2p},$$

  so the set of properties $(\mathcal{P}_{k+1})$ holds.

- Otherwise, one has $k+1-m_{k+1}=k-m_k$. It then follows from property (a_k) that $\|r^{(k+1-m_{k+1})}\|_2=\|r^{(k-m_k)}\|_2\le R_\mu\tau^{2p}$, so that (a_{k+1}) holds. Moreover, since there is no restart at step k, the condition

  $$\tau\|r^{(k+1)}-r^{(k-m_k)}\|_2\le\|(\mathrm{id}-\Pi_k)(r^{(k+1)}-r^{(k-m_k)})\|_2=d_{m_k+1}^{(k)}$$

  is verified. We then have

  $$\begin{aligned}d_{m_k+1}^{(k)} &\ge \tau\left(\|r^{(k-m_k)}\|_2-\|r^{(k+1)}\|_2\right)\\ &\ge \tau\left(\|r^{(k-m_k)}\|_2-\mu^{m_k+1}\|r^{(k-m_k)}\|_2\right)\\ &\ge \tau(1-\mu)\|r^{(k-m_k)}\|_2.\end{aligned}$$

  Using the same notations as in Lemma 3.7, we also have $d_i^{(k+1)}=d_i^{(k)}$ for any integer i in $\{1,\dots,m_k+1\}$. As a consequence, property (c_{k+1}) follows from property (c_k) and the above inequality. Finally, property (b_k) is equivalent to

  $$\forall j\in\{0,\dots,m_{k+1}-1\},\ \|r^{(k+1-m_{k+1}+j)}\|_2\le\mu^j\|r^{(k+1-m_{k+1})}\|_2,$$

  and bound (24) is equivalent to

  $$\|r^{(k+1-m_{k+1}+m_{k+1})}\|_2\le\mu^{m_{k+1}}\|r^{(k+1-m_{k+1})}\|_2,$$

  so that property (b_{k+1}) holds, and so does $(\mathcal{P}_{k+1})$.

The set of properties $(\mathcal{P}_k)$ is thus satisfied as long as $r^{(k)}$ is nonzero, and (22), (23), and (24) follow from it. The r-linear convergence property (25) is next proved by complete induction.
For $k = 0$, it is obvious. Assuming that for a given natural integer k, one has

$$\forall i \in \{0, \ldots, k\},\ \|r^{(i)}\|_2 \leq \mu^i \|r^{(0)}\|_2$$

and, using estimate (24), one finds that

$$\|r^{(k+1)}\|_2 \leq \mu^{m_k+1} \|r^{(k-m_k)}\|_2 \leq \mu^{m_k+1}\mu^{k-m_k} \|r^{(0)}\|_2 = \mu^{k+1}\|r^{(0)}\|_2,$$

ending the proof of (25).
Finally, if a restart occurs at step $k+1$, one has, employing the same notations as in Lemma 3.7,

$$d^{(k)}_{m_k+1} < \tau\|r^{(k+1)} - r^{(k-m_k)}\|.$$

Inequality (26) then follows from estimate (36) in Lemma 3.7 and the triangle inequality, by setting

$$\Gamma = \kappa \left(1 + (1-\mu)^{-2p}\right).$$

3.5.3 Proof of Theorem 3.2

For any natural integer k, we consider the following properties.

$$\text{For any } i \text{ in } \{0, \ldots, k\},\ x^{(i)} \text{ exists and lies in } U \cap \Sigma \setminus \{x_*\}$$
$$\text{and, } \forall i \in \{0, \ldots, k-1\},\ \|r^{(i+1)}\|_2 \leq \mu\|r^{(i)}\|_2. \quad (a_k)$$

$$\text{If } m_k \geq 1, \text{ then, } \forall j \in \{1, \ldots, m_k\},\ \forall i \in \{0, \ldots, j-1\},\ \delta\|r^{(k-m_k+i)}\|_2 \leq \|r^{(k-m_k+j)}\|_2. \quad (b_k)$$

$$\text{If } m_k \geq 1, \text{ then, } \forall \ell \in \{1, \ldots, m_k\},\ d^{(k)}_\ell \geq \frac{1-\mu}{\mu}\|r^{(k-m_k+\ell)}\|_2. \quad (c_k)$$

Let us denote by $(\mathcal{P}_k)$ the set of properties (a_k), (b_k), and (c_k) at rank k. Obviously, $(\mathcal{P}_0)$ holds. Under the assumptions of Theorem 3.2 and for c_μ sufficiently small, we will prove by induction that $(\mathcal{P}_k)$ holds for all k as long as $r^{(k)}$ is nonzero, and that this will imply the statements of the theorem.
Let us then assume that $(\mathcal{P}_k)$ holds for some natural integer k. From property (a_k), one has

$$\forall \ell \in \{0, \ldots, m_k\},\ \max_{i \in \{\ell, \ldots, m_k\}} \|r^{(k-m_k+i)}\|_2 = \|r^{(k-m_k+\ell)}\|_2 \leq c_\mu\, \delta^2, \quad (39)$$

and property (c_k) implies that assumption (33) in Lemma 3.7 is satisfied with $t = \frac{1-\mu}{\mu}$. As a consequence, both (27) and estimate (34) hold, the latter taking the form

$$\|c^{(k)}\|_\infty \leq C_{m_k}\left(1 + \left(\frac{\mu}{1-\mu}\right)^{m_k}\right),$$

which is exactly estimate (28).

Next, using Assumptions 1 and 2, the bounds

$$\max_{i \in \{0, \ldots, m_k\}} \|x^{(k-m_k+i)} - x_*\|_2 \leq \sigma^{-1}c_\mu K^2 \text{ and } \max_{i \in \{0, \ldots, m_k\}} \|g(x^{(k-m_k+i)}) - x_*\|_2 \leq \sigma^{-1}c_\mu K^3$$

follow from (39). Combining them with (28), we see that, for c_μ small enough, $\tilde{x}^{(k+1)}$ belongs to U, as well as $g(\tilde{x}^{(k+1)})$ and $\sum_{i=0}^{m_k} \tilde{c}_i\, g(x^{(k-m_k+i)})$, so that $x^{(k+1)}$ and $r^{(k+1)}$ are both well defined, with $x^{(k+1)}$ belonging to $U \cap \Sigma$, and estimates (35) and (36) of Lemma 3.7 hold.

We now assume that $r^{(k+1)}$ is nonzero and impose the following additional smallness constraint on the constant c_μ,

$$\kappa\left(1 + \left(\frac{\mu}{1-\mu}\right)^{2p}\right) c_\mu < 1 - \frac{K}{\mu}. \quad (40)$$

From property (b_k), one has that $\|r^{(k-m_k)}\|_2 \leq \delta^{-1}\|r^{(k)}\|_2$, so that, using (39) (for $\ell = 0$), we conclude that

$$\max_{i \in \{0, \ldots, m_k\}} \|r^{(k-m_k+i)}\|_2^{\,2} = \|r^{(k-m_k)}\|_2^{\,2} \leq c_\mu\delta\|r^{(k)}\|_2 \leq c_\mu K\|r^{(k)}\|_2.$$

Hence, using (35) and (40), we get

$$\begin{aligned}\|r^{(k+1)}\|_2 &\le K\,\mathrm{dist}_2(0,\mathcal{A}_k^{(k)}) + \kappa(1+t^{-2p}) \max_{i\in\{0,\dots,m_k\}} \|r^{(k-m_k+i)}\|_2{}^2 \\ &\le K\left(1+\kappa(1+t^{-2p})c_\mu\right)\|r^{(k)}\|_2 \\ &\le \mu\|r^{(k)}\|_2,\end{aligned}$$

which, together with property (a_k), establishes that property (a_{k+1}) holds.

If $m_{k+1}=0$, there is nothing else to check, and $(\mathcal{P}_{k+1})$ holds. Otherwise, one has, by design of the algorithm,

$$\forall i\in\{0,\dots,m_{k+1}\},\ \|r^{(k+1)}\|_2 \ge \delta\|r^{(k+1-m_{k+1}+i)}\|.$$

Using property (b_k), property (b_{k+1}) ensues. Now, from (b_k) and (b_{k+1}), one has

$$\|r^{(k-m_k)}\|_2 \le \frac{1}{\delta}\|r^{(k+1-m_{k+1})}\|_2 \le \frac{1}{\delta^2}\|r^{(k+1)}\|_2.$$

Combining this with (39) (for $\ell=0$), one finds

$$\max_{i\in\{0,\dots,m_k\}} \|r^{(k-m_k+i)}\|_2{}^2 = \|r^{(k-m_k)}\|_2{}^2 \le c_\mu\|r^{(k)}\|_2 \le c_\mu\|r^{(k)}\|_2.$$

As a consequence, estimate (36) of Lemma 3.7 gives

$$\|r^{(k+1)}\|_2 \le \frac{K}{1-K}\,d_{m_k+1}^{(k)} + \frac{\kappa}{1-K}\left(1+\left(\frac{\mu}{1-\mu}\right)^{2p}\right)c_\mu\|r^{(k+1)}\|_2. \tag{41}$$

But constraint (40) means exactly that

$$\frac{K}{1-K}\frac{1-\mu}{\mu} + \frac{\kappa}{1-K}\left(1+\left(\frac{\mu}{1-\mu}\right)^{2p}\right)c_\mu < 1,$$

so that $\|r^{(k+1)}\|_2$ is non-positive if $d_{m_k+1}^{(k)} < \frac{1-\mu}{\mu}\|r^{(k+1)}\|_2$, which is absurd. We have thus proved by contradiction that

$$d_{m_k+1}^{(k)} \ge \frac{1-\mu}{\mu}\|r^{(k+1)}\|_2.$$

Together with our assumption (c_k), this implies the new property

$$\forall \ell\in\{1,\dots,m_k+1\},\ d_\ell^{(k)} \ge \frac{1-\mu}{\mu}\|r^{(k-m_k+\ell)}\|_2. \tag{$\hat{c}_k$}$$

Now, it is easily seen that $\mathcal{A}_{\ell-1}^{(k+1)} \subset \mathcal{A}_{\ell+m_k-m_{k+1}}^{(k)}$ for any integer ℓ in $\{1,\dots,m_{k+1}\}$, so that one has $d_\ell^{(k+1)} \ge d_{\ell+1+m_k-m_{k+1}}^{(k)}$. Property $(\hat{c}_k)$ implying that

$$\forall \ell\in\{1,\dots,m_{k+1}\},\ d_{\ell+1+m_k-m_{k+1}}^{(k)} \ge \frac{1-\mu}{\mu}\|r^{(k+1-m_{k+1}+\ell)}\|_2,$$

property (c_{k+1}) holds, and so does $(\mathcal{P}_{k+1})$.

We have thus proved that, for c_μ small enough, $(\mathcal{P}_k)$ holds for any natural integer k such that $r^{(k)}$ is nonzero, and so do statements (27), (28), and (29). It remains to prove statement (30).

Let k be a natural integer such that $k-p\ge 1$ and $\|r^{(k)}\|_2$ is nonzero. Since $m_k\le p$, one has $k-p-1\le k-m_k-1$, so that $\|r^{(k)}\|_2 \le \delta\|r^{(k-m_k-1)}\|_2 \le \delta\|r^{(k-p-1)}\|_2$. This ends the proof.

3.5.4 Proof of Theorem 3.5

We set $\varepsilon(T_0,\zeta,\mu) = \min\left\{T_0^{-1/\zeta}, (R_\mu T_0^{2p})^{\frac{1}{1-2p\zeta}}, \left(\frac{1-K}{2KT_0}\right)^{1/\zeta}\right\}$ and $k_0=0$. Considering a natural integer k_i such that $x^{(k_i)}$ is well-defined, $m_{k_i}=0$ and $\|r^{(k_i)}\|_2 \le \mu^{k_i}\|r^{(0)}\|_2$, we will prove the existence of a strictly larger natural integer k_{i+1} such that $x^{(k_{i+1})}$ is well-defined, $m_{k_{i+1}}=0$, $m_k = k-k_i$ for any integer k in $\{k_i,\dots,k_{i+1}-1\}$, and $\|r^{(k)}\|_2 \le \mu^{k-k_i}\|r^{(k_i)}\|_2$ for any integer k in $\{k_i,\dots,k_{i+1}\}$.

We may suppose that $r^{(k_i)}$ is nonzero (otherwise, one would have $x^{(k_i)} = x^{(k_{\rm stop})}$ and thus $k_{i+1} = k_i + 1$). In such case, by design of Algorithm 3', the parameter τ in the boolean test at step k remains equal to $\tau^{(k_i)} = T_0 \|r^{(k_i)}\|_2^\zeta$ for any integer k greater than or equal to k_i such that $x^{(k)}$ is well-defined and $m_k = k - k_i$. This means that, in this range of steps, the modified algorithm is equivalent to the original one, with fixed τ equal to $\tau^{(k_i)}$ and initial point equal to $x^{(k_i)}$. Since $0 < \|r^{(k_i)}\|_2 \leq \|r^{(k_0)}\|_2 < T_0^{-1/\zeta}$, one has $0 < \tau^{(k_i)} < 1$. Moreover, by definition of $\tau^{(k_i)}$ and since $\|r^{(k_0)}\|_2 \leq (R_\mu T_0^{2p})^{\frac{1}{1-2p\zeta}}$, one has $\|r^{(k_i)}\|_2 \leq R_\mu (T_0 \|r^{(k_i)}\|_2^\zeta)^{2p} = R_\mu (\tau^{(k_i)})^{2p}$, and Theorem 3.1 applies. As a consequence, the integer k_{i+1} exists and is such that $k_{i+1} \leq k_i + p$ and $\|r^{(k_{i+1})}\|_2 \leq \mu^{k_{i+1}-k_i} \|r^{(k_i)}\|_2 \leq \mu^{k_{i+1}} \|r^{(0)}\|_2$.

We have thus shown that the sequence $(x^{(k)})_{k\in\mathbb{N}}$ is well-defined and converges at least r-linearly to x^*, and that the family of steps at which the method restarts forms an infinite sequence $(k_i)_{i\in\mathbb{N}}$ such that the difference between two consecutive terms is at most equal to p. This allows to say more about the convergence of the sequence of approximations. Indeed, if $\|r^{(k_{i+1})}\|_2 > 0$, then estimate (26) holds with $k+1 = k_{i+1}$, $m_k = k_{i+1} - k_i - 1$ and $\tau = \tau^{(k_i)}$. Moreover, since it is assumed that $\|r^{(0)}\|_2 \leq \left(\frac{1-K}{2KT_0}\right)^{1/\zeta}$, it follows that $1 - K(1+\tau^{(k_i)}) \geq \frac{1-K}{2}$ for any natural integer i. Hence, using the formula for $\tau^{(k_i)}$, one gets from (26) the estimate

$$\frac{1-K}{2}\|r^{(k_{i+1})}\|_2 \leq \left(K\tau^{(k_i)} + \Gamma\, T_0^{-\frac{1}{\zeta}} (\tau^{(k_i)})^{\frac{1}{\zeta}-2p}\right)\|r^{(k_i)}\|_2 \leq \left(K + \Gamma\, T_0^{-\frac{1}{\zeta}}\right)(\tau^{(k_i)})^{\min\{1,\frac{1}{\zeta}-2p\}}\|r^{(k_i)}\|_2,$$

from which the inequality

$$\|r^{(k_{i+1})}\|_2 \leq B\, \|r^{(k_i)}\|_2^{\,\Theta}$$

follows, with $B = \frac{2}{1-K}\left(K + \Gamma\, T_0^{-\frac{1}{\zeta}}\right) T_0^{\frac{\Theta-1}{\zeta}}$ and $\Theta = 1 + \min\{\zeta, 1-2p\zeta\}$.

Next, let i_0 be an integer such that $B^{\frac{1}{\Theta-1}} \mu^{k_{i_0}} \|r^{(0)}\|_2 < \frac{1}{2}$. Since $\|r^{(k_{i_0})}\|_2 \leq \mu^{k_{i_0}} \|r^{(0)}\|_2$ and $\|r^{(k_{i+1})}\|_2 \leq B\, \|r^{(k_i)}\|_2^{\,\Theta}$, one can prove by induction that, for any integer i greater than or equal to i_0, $\|r^{(k_i)}\|_2 \leq B^{-\frac{1}{\Theta-1}} 2^{-\Theta^{i-i_0}}$ and, using that $k_i \leq i\, p$, $\|r^{(k_i)}\|_2 \leq B^{-\frac{1}{\Theta-1}} 2^{-\Theta^{(k_i - k_{i_0})/p}}$.

Finally, for any integer k greater than or equal to k_{i_0}, let the integer $i(k)$ be such that $k - m_k = k_{i(k)}$. This implies that $i(k) \geq i_0$ and $k_{i(k)} \geq k - p$, so that

$$\|r^{(k)}\|_2 \leq \|r^{(k_{i(k)})}\|_2 \leq B^{-\frac{1}{\Theta-1}} 2^{-\Theta^{(k_{i(k)} - k_{i_0})/p}} \leq b\, \eta^{\theta^k},$$

where $\eta = 2^{-\Theta^{-p-k_{i_0}}}$, $b = B^{-\frac{1}{\Theta-1}}$, and $\theta = \Theta^{1/p} = (1 + \min\{\zeta, 1-2p\zeta\})^{1/p}$ is the desired r-order of superlinear convergence.

Remark 3.8 *In Algorithm 3', the sequence $(\tau^{(k-m_k)})_{k\in\mathbb{N}}$ tends to 0 and estimate* (23) *on the extrapolation coefficients is no longer uniform: we can only say that $\|c^{(k)}\|_\infty = O\left(\|r^{(k-m_k)}\|_2^{\,-\zeta m_k}\right)$.*

3.5.5 Proof of Theorem 3.6

The proof follows an induction argument very similar to the one in the proof of Theorem 3.2, but with some novelties, since the value of the parameter δ is no longer fixed along the iteration in Algorithm 4'.

The set of properties at rank k, denoted $(\mathcal{P}_k)$, is almost unchanged, the only modification being the replacement of δ by $D_0\, \|r^{(k-m_k+i)}\|_2^{\,\xi}$ in the second property:

$$\text{if } m_k \geq 1, \text{ then, } \forall j \in \{1,\dots,m_k\},\ \forall i \in \{0,\dots,j-1\},\ D_0\, \|r^{(k-m_k+i)}\|_2^{\,1+\xi} \leq \|r^{(k-m_k+j)}\|_2. \qquad (b_k)$$

Assume that $(\mathcal{P}_k)$ holds for some natural integer k. From properties (a_k) and (c_k), one infers that assumption (33) in Lemma 3.7 is satisfied with $t = \frac{1-\mu}{\mu}$, so that both (27) and (28) hold.

Moreover, the bounds

$$\forall i \in \{0,\dots,m_k\},\ \|x^{(k-m_k+i)} - x_*\|_2 \leq \sigma^{-1}\varepsilon(D_0,\xi,\mu) \text{ and } \|g(x^{(k-m_k+i)}) - x_*\|_2 \leq \sigma^{-1} K \varepsilon(D_0,\xi,\mu)$$

follow from using property (a_k) and the use of Assumptions 1 and 2. As a consequence, for $\varepsilon(D_0,\xi,\mu)$ small enough, the linear combination $\tilde{x}^{(k+1)}$ belongs to U, and so do $g(\tilde{x}^{(k+1)})$ and $\sum_{i=0}^{m_k} \tilde{c}_i\, g(x^{(k-m_k+i)})$, resulting in both $x^{(k+1)}$ and $r^{(k+1)}$ being well-defined (with $x^{(k+1)}$ belonging to $U \cap \Sigma$) and estimates (35) and (36) of Lemma 3.7 being valid.

Let us now assume that $r^{(k+1)}$ is nonzero. Property (b_k) implies in particular that $\|r^{(k-m_k)}\|_2^{\,1+\xi} \leq {D_0}^{-1} \|r^{(k)}\|_2$. This, together with property (a_k) and the fact that $\|r^{(k-m_k)}\|_2 < \varepsilon(D_0,\xi,\mu)$ by assumption, gives

$$\max_{i\in\{0,\dots,m_k\}} \|r^{(k-m_k+i)}\|_2^{\,2} = \|r^{(k-m_k)}\|_2^{\,2} \leq {D_0}^{-\frac{2}{1+\xi}} (\varepsilon(D_0,\xi,\mu))^{\frac{1-\xi}{1+\xi}} \|r^{(k)}\|_2.$$

Hence, using estimate (35), we get

$$\begin{aligned}\|r^{(k+1)}\|_2 &\leq K\,\mathrm{dist}_2(0,\mathcal{A}_k^{(k)}) + \kappa(1+t^{-2p}) \max_{i\in\{0,\dots,m_k\}} \|r^{(k-m_k+i)}\|_2{}^2 \\ &\leq \left(K + \kappa(1+t^{-2p}) D_0{}^{-\frac{2}{1+\xi}} (\varepsilon(D_0,\xi,\mu))^{\frac{1-\xi}{1+\xi}}\right) \|r^{(k)}\|_2.\end{aligned}$$

Since $K<\mu$, we may impose that

$$\left(K + \kappa(1+t^{-2p}) D_0{}^{-\frac{2}{1+\xi}} (\varepsilon(D_0,\xi,\mu))^{\frac{1-\xi}{1+\xi}}\right) \leq \mu, \tag{42}$$

which can be achieved by taking $\varepsilon(D_0,\xi,\mu)$ small enough, since ξ belongs to $(0,\sqrt{2}-1]$. Thus, there holds $\|r^{(k+1)}\|_2 \leq \mu\,\|r^{(k)}\|_2$, which, combined with property (a_k), establishes property (a_{k+1}).

If $m_{k+1}=0$, there is nothing else to check, so $(\mathcal{P}_{k+1})$ is true. If it is not the case, one has, by design of Algorithm 4',

$$\forall i\in\{0,\dots,m_{k+1}\},\ \|r^{(k+1)}\|_2 \geq D_0\,\|r^{(k+1-m_{k+1}+i)}\|_2{}^{1+\xi}.$$

Using property (b_k), property (b_{k+1}) ensues, leading to

$$\|r^{(k-m_k)}\|_2 \leq D_0{}^{-\frac{2+\xi}{(1+\xi)^2}} \|r^{(k+1)}\|_2{}^{\frac{1}{(1+\xi)^2}}.$$

As a consequence, estimate (36) of Lemma 3.7 gives

$$\|r^{(k+1)}\|_2 \leq \frac{K}{1-K}\,d_{m_k+1}^{(k)} + \frac{\kappa}{1-K}\left(1+\left(\frac{\mu}{1-\mu}\right)^{2p}\right) D_0{}^{-\frac{2(2+\xi)}{(1+\xi)^2}} \|r^{(k+1)}\|_2{}^{\frac{2}{(1+\xi)^2}}. \tag{43}$$

We now impose the condition

$$\frac{K}{1-K}\frac{1-\mu}{\mu} + \frac{\kappa}{1-K}\left(1+\left(\frac{\mu}{1-\mu}\right)^{2p}\right) D_0{}^{-\frac{2(2+\xi)}{(1+\xi)^2}} \|r^{(k+1)}\|_2{}^{\frac{2}{(1+\xi)^2}-1} < 1,$$

satisfied by taking $\varepsilon(D_0,\xi,\mu)$ small enough when ξ belongs to $(0,\sqrt{2}-1)$, or Δ_μ large enough when ξ is equal to $\sqrt{2}-1$. Inequality (43) then implies that $\|r^{(k+1)}\|_2$ is non-positive if $d_{m_k+1}^{(k)} < \frac{1-\mu}{\mu}\|r^{(k+1)}\|_2$, which is absurd. We have thus proved by contradiction that

$$d_{m_k+1}^{(k)} \geq \frac{1-\mu}{\mu}\|r^{(k+1)}\|_2.$$

The rest of the induction argument is exactly the same as in the proof of Theorem 3.2, with $(\mathcal{P}_k)$ holding for any natural integer k such that $r^{(k)}$ is nonzero, and implying estimates (27), (28) and (29).

It remains to establish that the convergence is superlinear. If $k\geq p+1$, then $k-m_k \geq k-p \geq 1$, so that, by property (a_k), one has

$$\|r^{(k)}\|_2 \leq D_0\,\|r^{(k-m_k-1)}\|_2{}^{1+\xi} \leq D_0\,\|r^{(k-p-1)}\|_2{}^{1+\xi}.$$

It also follows from property (a_k) that there exists a natural integer k_0 such that

$$\forall i\in\{0,\cdots,p\},\ D_0{}^{1/\xi}\|r^{(k_0+i)}\|_2 \leq \frac{1}{2}.$$

As a consequence, for any integer k greater or equal to k_0+p+1, writing $k-k_0=(p+1)q+i$ with i in $\{0,\dots,p\}$ we get, reasoning by induction on q, that

$$\|r^{(k)}\|_2 \leq D_0{}^{-\frac{1}{\xi}} 2^{-(1+\xi)^q} \leq b\,\eta^{\theta^k},$$

where $b = D_0^{-\frac{1}{\xi}}$, $\eta = 2^{-(1+\xi)^{-\frac{k_0+p}{p+1}}} < 1$ and $\theta = (1+\xi)^{\frac{1}{p+1}}$. The proof is complete.

4 Application to electronic ground state calculations

We now show how the Anderson–Pulay acceleration analysed in the previous section can be applied to the computation of the electronic ground state of a molecular system through two of the most commonly used approximations of the quantum many-body problem in non-relativistic quantum chemistry.

Consider an isolated molecular system composed of M atomic nuclei and N electrons. Within the setting of the Born–Oppenheimer approximation, the motion of atomic nuclei and electrons can be separated and the nuclei are classical point-like particles with fixed positions. The state of the electrons is then entirely described by a wavefunction ψ valued in $\mathbb{C}$, only depending on the time variable and on the respective positions and spins of the electrons. The ground state of the molecular system is determined by solving a minimisation problem, which one may try to attack "directly" by working in a finite basis. Unfortunately, due to the high dimension of the position space $\mathbb{R}^{3N}$, this approach is intractable for systems with more than a few electrons and one has to resort to other types of approximations.

On the one hand, the *ab initio Hartree–Fock method* [30, 22] for the computation of the ground-state electronic wavefunction ψ consists in restricting the functional space in the minimisation problem to the set of so-called *Slater determinants*, which are antisymmetrised products of N *monoelectronic wavefunctions* (also called *molecular (or atomic) orbitals*). While this restriction only provides with an upper bound of the exact energy, the main advantage of this approximation is that the problem to be solved remains variational. There is however a price to pay in a loss of the correlation between the positions of the electrons, as an electron will evolve independently of the way the others do in this model.

On the other hand, the *density functional theory* (DFT) of Hohenberg and Kohn [33] follows a completely different approach and aims at including the electronic correlation missing in the Hartree–Fock method. It is based on the fact that the ground-state properties of a many-electron system are uniquely determined by an *electronic density*, which only depends on the three spatial coordinates. Indeed, following arguments by Hohenberg and Kohn, the energy functional defined in terms of the unknown wavefunction ψ can be replaced by one for the unknown density function ρ. Since such a functional is not known explicitly for a system of N interacting electrons, suitable approximations are needed to make the DFT a practical tool for computing electronic ground states. These rely on exact (or very accurate) evaluations of the density functional of a reference system "close" to the real one. In the semi-empirical approach of Kohn and Sham [36], the chosen reference system consists of N *non*-interacting electrons (the associated wavefunction being a single Slater determinant). The exchange-correlation part of the functional must then be approximated, using models like the local-density approximation (LDA) or the generalised gradient approximation (GGA).

While based on different physical principles, the discrete problems associated with these two approximations both take the form of a nonlinear generalised eigenvalue problem. Omitting the spin variables for the sake of simplicity, the monoelectronic wavefunctions are linearly expanded on a given finite basis set[8] (this is the so-called *LCAO approximation*, LCAO being the acronym to *linear combination of atomic orbitals*), spanning a vector space of finite dimension d.

In this setting, the spinless Hartree–Fock method leads to the so-called *Roothaan–Hall equations* [49, 29],

$$\begin{cases} F^{\mathrm{HF}}(D)C = SC\Lambda, \\ C^*SC = I_N, \\ D = CC^*, \end{cases}$$

in which the matrix $F^{\mathrm{HF}}(D)$ denotes, with a slight abuse of notation, the *Fock matrix* associated with the rectangular matrix C of orbital coefficients, the square matrix D is the *discrete density matrix*, the matrix S is the so-called *overlap* matrix, that is the Gram matrix associated with the basis set, the Hermitian matrix Λ, which is by convention chosen diagonal, stands for the Lagrange multipliers attached to an orthonormality constraints on the orbitals, and I_N is the identity matrix of order N. A necessary condition for a matrix D_* to be a "solution" of the above system in the submanifold of pure state density matrices of rank N,

$$\mathcal{P}_N = \{D \in \mathcal{M}_{d,d}(\mathbb{C}),\ D^* = D,\ DSD = D,\ \mathrm{tr}(SD) = N\}, \tag{44}$$

is that it commutes with the associated Fock matrix $F^{\mathrm{HF}}(D_*)$ in the sense that

$$[F^{\mathrm{HF}}(D_*), D_*] = 0,$$

where the *"commutator"* $[A, B]$ between two matrices A and B is defined by $[A, B] = ABS - SBA$.

The discretisation of other Hartree–Fock models or of the Kohn–Sham models can be dealt with in the same manner. For the spinless Kohn-Sham model, assuming differentiability for the approximated exchange-correlation functional used in practice, the Roothaan–Hall equations read

$$\begin{cases} F^{\mathrm{KS}}(D)C = SC\Lambda, \\ C^*SC = I_N, \\ D = CC^*. \end{cases}$$

[8] Due to computational complexity considerations, the basis functions are typically Slater-type or Gaussian-type orbitals, and generally form a non-orthonormal set.

The Roothaan–Hall equations are usually solved *"self-consistently"*, that is using an iterative fixed-point procedure, the most simple and "natural" approach being the algorithm introduced by Roothaan [49]. Given the choice of an initial discrete density matrix $D^{(0)}$, it consists in generating a sequence of matrices $(D^{(k)})_{k\in\mathbb{N}}$ defined by

$$\forall k \in \mathbb{N},\ \begin{cases} F(D^{(k)})C^{(k+1)} = SC^{(k+1)}E^{(k+1)}, \\ (C^{(k+1)})^* SC^{(k+1)} = I_N, \\ D^{(k+1)} = C^{(k+1)}(C^{(k+1)})^*, \end{cases}$$

where $E^{(k+1)}$ is a diagonal matrix such that $(E^{(k+1)})_{ii} = \varepsilon_i^{(k+1)}$, $i = 1, \cdots, N$, the scalars $\varepsilon_1^{(k+1)} \leq \varepsilon_2^{(k+1)} \leq \cdots \leq \varepsilon_N^{(k+1)}$ being the N smallest eigenvalues, counted with multiplicity, of the *linear* generalised eigenproblem

$$F(D^{(k)})V = \varepsilon\, SV,$$

and the columns of the matrix $C^{(k+1)}$ are associated orthonormal (with respect to the scalar product induced by the matrix S) eigenvectors. The procedure of assembling $D^{(k+1)}$ by populating the molecular orbitals starting with those of lowest energy of the current Fock matrix is called the *Aufbau principle*.

Assuming the *uniform well-posedness property* introduced in [11], the matrix $D^{(k+1)}$ is uniquely defined at each step of the procedure and can be characterised as the minimiser of a variational problem, that is

$$D^{(k+1)} = \arg\inf \left\{ \operatorname{tr}\left(F(D^{(k)})D\right),\ D \in \mathcal{P}_N \right\} = g(D^{(k)}),$$

thus defining a fixed-point iteration process. In practice, a convergence criterion is used to end the iterations. For instance, one may compute the norm of the difference of two successive density matrices at each step and compare it to a prescribed tolerance. Another possibility is to use the norm of the commutator between the current density matrix and its associated Fock matrix.

Both of these choices may be employed to accelerate the convergence of the above self-consistent field (SCF) algorithm. Indeed, the first one corresponds to error vectors proposed in conjunction with the original form of the DIIS [46], that is $r^{(k)} = D^{(k)} - D^{(k-1)}$, while the second leads to those used in the CDIIS [47], that is $r^{(k)} = \left[F(D^{(k)}), D^{(k)}\right]$. The latter is widely used in quantum chemistry softwares for electronic structure calculations, notably because of its simplicity with respect to implementation and its usually rapid, but not assured, convergence. Let us describe it in more detail.

Starting from a guess density matrix $D^{(0)}$, one first sets $\widetilde{D}^{(0)} = D^{(0)}$. After k steps of the method, given a set of $m_k + 1$ previous density matrices $D^{(k-m_k)}, \ldots, D^{(k)}$, one assembles the *pseudo*-density matrix[9] $\widetilde{D}^{(k+1)}$ as

$$\widetilde{D}^{(k+1)} = \sum_{i=0}^{m_k} c_i^{(k)} D^{(k-m_k+i)},$$

where

$$\boldsymbol{c}^{(k)} = \underset{\substack{(c_0,\ldots,c_{m_k})\in\mathbb{R}^{m_k+1} \\ \sum_{i=0}^{m_k} c_i = 1}}{\arg\min} \left\| \sum_{i=0}^{m_k} c_i \left[F\left(D^{(k-m_k+i)}\right), D^{(k-m_k+i)} \right] \right\|_2,$$

the matrix norm $\|\cdot\|_2$ being the Frobenius norm. The next density matrix $D^{(k+1)}$, is then obtained by applying the Aufbau principle to $\widetilde{D}^{(k+1)}$, that is, by diagonalising the Fock matrix[10] associated with $\widetilde{D}^{(k+1)}$ and forming $D^{(k+1)}$ from the N eigenvectors associated with its N smallest eigenvalues. This process is then repeated until the numerical convergence criterion is satisfied. Modifications of the procedure have been proposed in order to make it more robust, either by replacing the constraint on the coefficients, like in the C^2-DIIS [51], or by obtaining them by minimisation of an associated energy, like in the EDIIS[11] [38], the ADIIS[12] [34] or the LIST[13] [59].

[9] This denomination stems from the fact that such a construction does not enforce the idempotency property on $\widetilde{D}^{(k+1)}$.

[10] For Hartree–Fock models, the pseudo-Fock matrix $\sum_{i=0}^{m_k} c_i^{(k)} F(D^{(k-m_k+i)})$ may as well be considered, since it holds that $F(\sum_{i=0}^{m_k} c_i^{(k)} D^{(k-m_k+i)}) = \sum_{i=0}^{m_k} c_i^{(k)} F(D^{(k-m_k+i)})$ due to constraint (3) being satisfied by the coefficients $c_i^{(k)}$, $i = 0, \ldots, m_k$.

[11] The difference in this method is that the coefficients of the linear expansion are such that

$$\boldsymbol{c}^{(k)} = \underset{\sum_{i=0}^{m_k} c_i = 1,\ c_i \in [0,1]}{\arg\min} E\left(\sum_{i=0}^{m_k} c_i D^{(k-m_k+i)} \right),$$

where E is the energy functional of the model under consideration.

[12] This other variant is similar to the EDIIS but uses the following augmented energy functional:

$$E^{\mathrm{A}}(D) = E(D^{(k)}) + 2\operatorname{tr}((D - D^{(k)})F(D^{(k)})) + \operatorname{tr}((D - D^{(k)})(F(D) - F(D^{(k)}))).$$

[13] This other variant is also similar to the EDIIS. It uses the so-called corrected Hohenberg–Kohn–Sham functional [65].

Let us end this section by explaining how the CDIIS enters the theoretical framework set in the previous section for the Anderson–Pulay acceleration.

On the one hand, working with self-adjoint (symmetric or Hermitian) matrices of order d (d being the dimension of the vector space spanned by the basis functions used in the Galerkin approximation) with fixed trace, the integer n is equal to $\frac{1}{2}d(d+1)-1$, the fixed-point function g corresponds to the application of the Aufbau principle to such a matrix, and the submanifold Σ is the set $\mathcal{P}_N$ of pure state density matrices of rank N, defined in (44). On the other hand, the error function f is the commutator between a given density matrix and its associated Fock matrix, $f(D) = [F(D), D]$, so that the integer p is equal to $\frac{1}{2}d(d-1)$.

Assumption 1 then simply states that the Roothaan algorithm is locally convergent, which is a common assumption on the base fixed-point iteration method in the analysis of the DIIS or the Anderson acceleration (see [48] or [57] for instance). Assumption 2 amounts to a non-degeneracy assumption, which is equivalent to saying that the differential of the restriction of the error function to the submanifold is invertible at a solution D_*. In the present context, this function is already the gradient of the discretised Hartree–Fock (or Kohn–Sham) energy, and the assumption thus implies that the Hessian of this energy is non-degenerate at D_*.

Concerning regularity assumptions, the fact that the set $\mathcal{P}_N$ is a smooth submanifold is a consequence of the constant rank theorem. For the functions f and g being of class $\mathscr{C}^2$, this is always true for the Hartree–Fock functional, which is smooth. In the case of the Kohn–Sham model, this issue is more delicate, since most of the exchange-correlation functionals used in practice present a lack of regularity at the origin, see *e.g.* [2]. However, it is reasonable to assume that the Kohn–Sham ground state has a non-vanishing density ρ, so that f and g are indeed regular in its neighbourhood.

5 Numerical experiments

In this section, we report on some numerical experiments with the intent of illustrating the performances of our proposed variants of the CDIIS in the context of the electronic ground state calculations considered in Section 4. All the computations presented were performed using tools provided by the PySCF package [54]. The source code of our implementation of the CDIIS variants can be found in the following repository: https://plmlab.math.cnrs.fr/mchupin/restarted-and-adaptive-cdiis/.

5.1 Implementation details

For each variant, the least-squares problem for the extrapolation coefficients is solved using the unconstrained form involving the differences of error vectors which are deemed the most convenient[14], as seen in Algorithms 3 and 4. The solution is achieved via the QR factorisation of a matrix of *tall and skinny* type (since the integer p is in general very large compared to m_k). Such a factorisation can be efficiently updated from step to step by means of a dedicated routine from the SciPy linear algebra library (namely `scipy.linalg.qr_update`), as the least-squares problem matrix is modified through the addition of a column (as in the variant with restarts) or the addition of a column and the possible removal of a set of columns (as in the adaptive-depth variant). A detailed cost analysis of the resulting factorisation algorithm is given in [32]. An added benefit of using the QR decomposition is that it directly provides the matrix of the orthogonal projector Π_k appearing in the restart condition (16), so that testing for this condition at each step in Algorithm 3 only entails a negligible cost.

Finally, we consider that numerical convergence is reached at iteration k if the error vector norm $\|[F(D^{(k)}), D^{(k)}]\|_2$ is below some prescribed tolerance.

5.2 Test cases

Some of the molecular systems used in our experiments are taken from benchmarks found in [24, 34] and are considered as representative of challenging convergence tests for self-consistent field algorithms. Results of some of the tests are presented here, like the cadmium(II)-imidazole complex ($[\mathrm{Cd(Im)}]^{2+}$) for instance, while others, like the acetaldehyde ($\mathrm{C_2H_4O}$), the acetic acid ($\mathrm{C_2H_4O_2}$), or the silane ($\mathrm{SiH_4}$), are available in the online repository given above. The glycine ($\mathrm{C_2H_5NO_2}$) test case comes from an example given in the PySCF library and the geometries for galactonolactone ($\mathrm{C_6H_{10}O_6}$) and dimethylnitramine ($\mathrm{C_2H_6N_2O_2}$) were found on the PubChem website (https://pubchem.ncbi.nlm.nih.gov/).

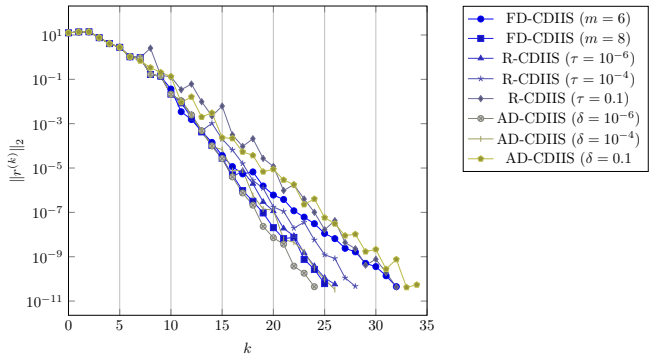
Both the restricted Hartree–Fock (RHF) model and the restricted Kohn–Sham (RKS) model are used in the experiments, the latter in conjunction with the B3LYP approximation of the exchange-correlation functional [6, 39].

[14] Note that, for the fixed-depth CDIIS, the unconstrained formulation we employed uses successive differences.

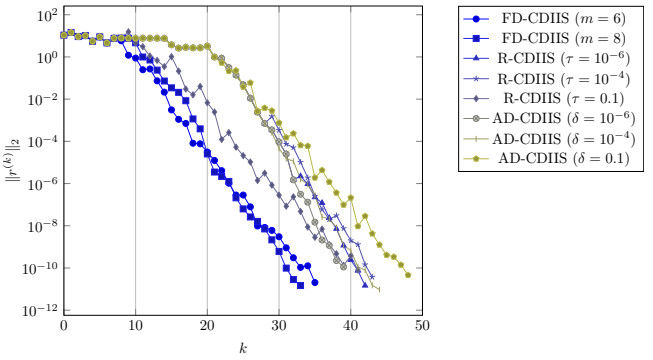
5.2.1 Global convergence behaviour

Acceleration techniques like the CDIIS are locally convergent methods and may sometimes give poor results if a mediocre initial guess is used. In practice, in order to ensure and achieve convergence in a small number of steps, one usually employs a combination of a relaxed constraint algorithm (like the ODA [10], the EDIIS [38] or the ADIIS [34]), for its global convergence properties, and of the CDIIS, for its fast local convergence (see Subsection 5.2.2). Nevertheless, our first experiment is meant to illustrate the convergence behaviour of the methods starting from the *core Hamiltonian guess*, for which the orbital coefficients are simply obtained by diagonalising the matrix representing the kinetic energy and electron-nuclear potential energy parts of the Hamiltonian in the considered discrete basis.

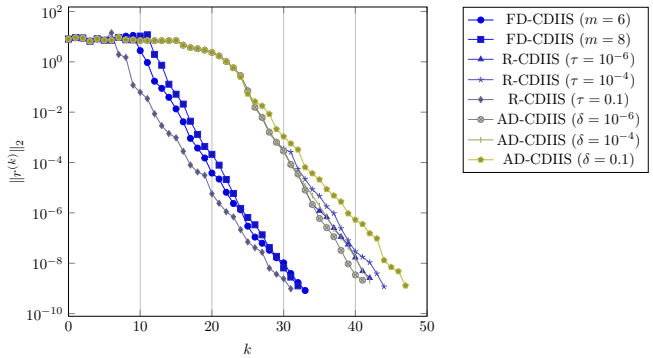
Figure 1 presents the convergence of the norm of the error vector norm for the (classical) fixed-depth, restarted and adaptive-depth variants of the CDIIS with different values of their respective parameters: the fixed maximum depth m, the restart parameter τ and the adaptive-depth parameter δ. It is observed that the restarted and the adaptive-depth algorithms are efficient in most of the cases, but they may reach the convergence regime later than their fixed-depth counterpart (see Figures 1c and 1b). However, when this regime is attained, one can observe that the rate of convergence of both the restarted and adaptive-depth CDIIS is generally better than that of the fixed-depth CDIIS.



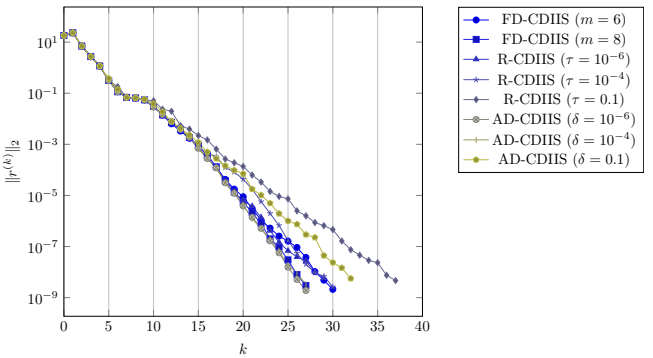
(a) Cadmium-imidazole complex in the RKS/B3LYP model with basis 3-21G.



(b) Glycine molecule in the RKS/B3LYP model with basis 6-31Gs.



(c) Dimethylnitramine molecule in the RHF model with basis 6-31G.



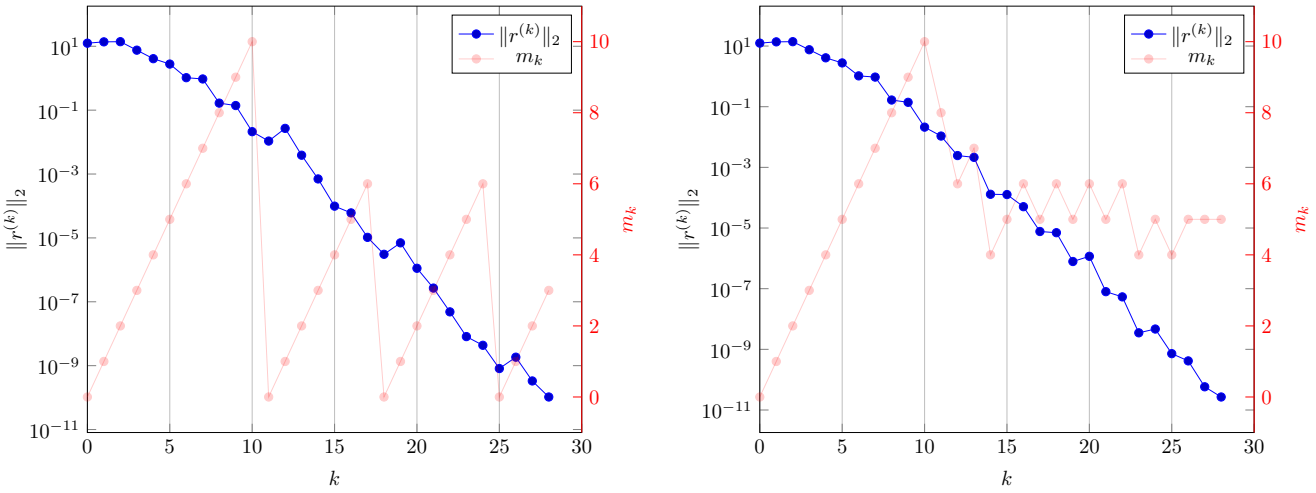
(d) Galactonolactone molecule in the RHF model with basis 6-31G.

Figure 1: Residual norm convergence for the fixed-depth, restarted and adaptive-depth CDIIS on different molecular systems using an initial guess obtained by diagonalising the core Hamiltonian matrix.

An effect of the use of a poor initial guess is the accumulation of many stored iterates in the early stages of the computation by the restarted and adaptive-depth variants, visible in Figure 2. The number of stored iterates clearly decreases as soon as a convergence regime is reached. This behaviour can be observed in all of our numerical experiments.

5.2.2 Local convergence behaviour

In order to properly assess their local convergence properties, the CDIIS variants were combined with a globally convergent method and an improved initial guess, both provided by the PySCF package. More precisely, the EDIIS with a fixed-depth equal to 8 and an initial guess generated from a superposition of atomic density matrices were used for the experiments with the RHF model, while the ADIIS with a fixed-depth equal to 8 and the same type of initial guess were used for the experiments with the RKS model. In both cases, the switch between the "global" method and



(a) Restarted CDIIS with $\tau = 10^{-4}$.

(b) Adaptive-depth CDIIS with $\delta = 10^{-4}$.

Figure 2: Residual norm convergence and corresponding depth value for the restarted and adaptive-depth CDIIS on the cadmium-imidazole complex in the RKS/B3LYP model with basis 3-21G, using an initial guess obtained by diagonalising the core Hamiltonian matrix.

the CDIIS variants was made when the error vector norm was below 10^{-2}. With such an initialisation, convergence was immediately observed in every test case. Figure 3 presents the obtained results for the set of molecules already considered in Figure 1 and a smaller set of parameter values. In such a setting, the restarted and the adaptive-depth CDIIS are shown to be more efficient than the fixed-depth one when adequately chosen values of their respective parameters are used.

Plotting the error vector norm with the corresponding depthduring the course of the numerical experiments, as done in Figure 4 for the dimethylnitramine and the glycine molecules, allows to observe that a restart occurs after a significant decrease of the error vector norm, as predicted by the theory for the restarted Anderson–Pulay acceleration. Unfortunately, one can also notice a slowdown in the convergence (or even a moderate increase of the error vector norm) just after a restart, thus motivating the introduction of an adaptive-depth mechanism which would not suffer from such a defect.

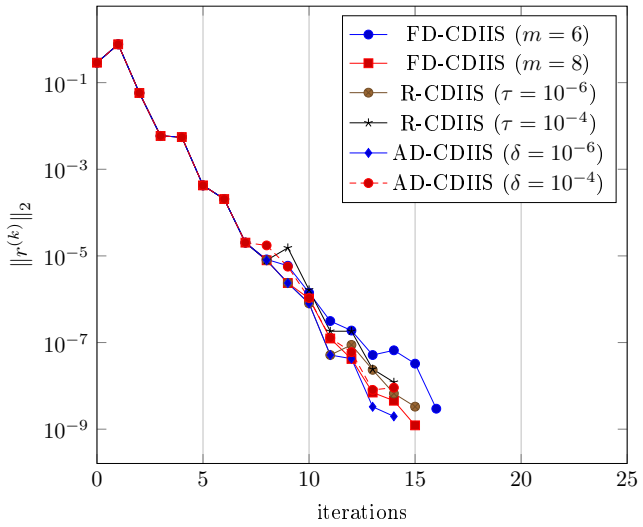
Mean depth. The cost of the CDIIS in terms of storage and computational resource at a given iteration is proportional to the value of the depth at this iteration. As a consequence, to properly compare the restarted and adaptive-depth variants with the classical fixed-depth CDIIS, we have computed the mean depth, denoted by $\bar{m}$, as the average value of m_k during an experiment. Figure 5 presents the evolution of $\bar{m}$ with respect to the values of the restart parameter τ and the adaptive-depth δ for each of the molecular systems we considered. It is seen that $\bar{m}$ is a decreasing function of these parameters and that the two variants have on average lesser costs than their fixed-depth counterpart, while their performances are comparable or better.

Rate of convergence. For each of the molecular systems considered in Figure 3, the practical rate of convergence of the restarted and adaptive-depth CDIIS was computed using a linear regression and plotted against the values of the parameters τ and δ in Figure 6. As expected, it is apparent that this rate increases as the value of the parameter decreases, its evolution for the adaptive-depth variant being noticeably smoother.

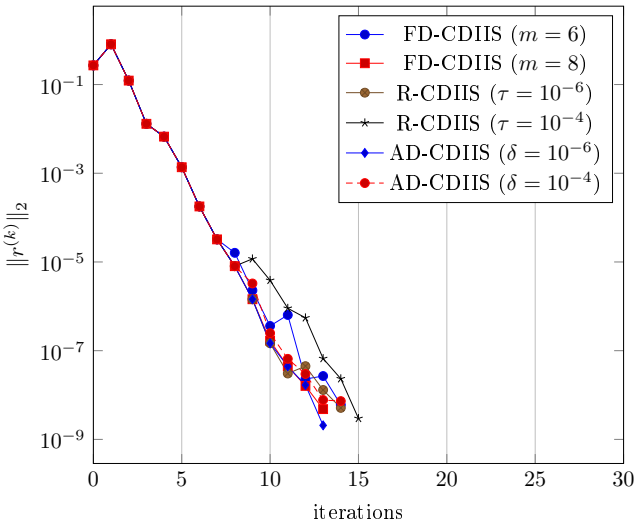
Selecting values for the parameters. As previously mentioned, the decrease of the error vector norm becomes faster, and the average depth $\bar{m}$ increases, as the parameters τ and δ are decreased. However, a closer scrutiny of both Figures 5 and 6 reveals that the convergence rate appears to tend to an asymptotic value while the average depth grows at an almost constant rate. Thus, the gain of convergence by decreasing the parameters may not outweight the added computational cost of keeping a larger history of iterates. In this regard, a satisfying compromise was reached in our numerical tests by setting the values of both parameters at 10^{-4}.

6 Conclusion

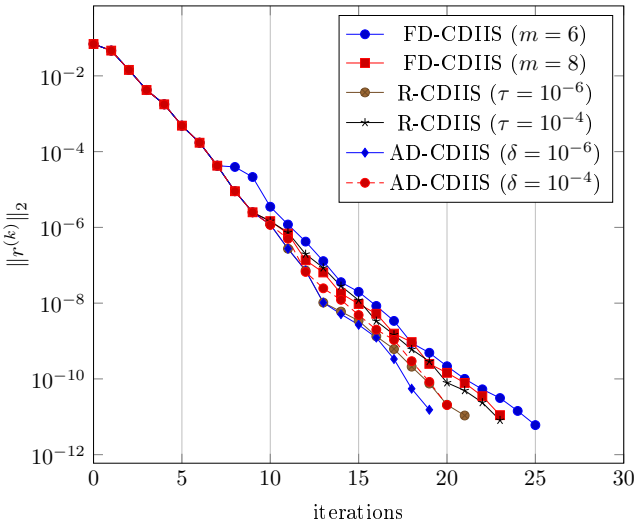
Motivated by the DIIS and CDIIS techniques, respectively introduced by Pulay in 1980 [46] and 1982 [47], and their relation with other extrapolation processes designed to accelerate fixed-point iteration methods – one of them being the well-known Anderson acceleration – we have considered a general and abstract class of acceleration methods and studied, theoretically and numerically, the local convergence properties of two of its instances: one allowing restarts, based on a condition initially introduced for a quasi-Newton using multiple secant equations [25, 48], and another



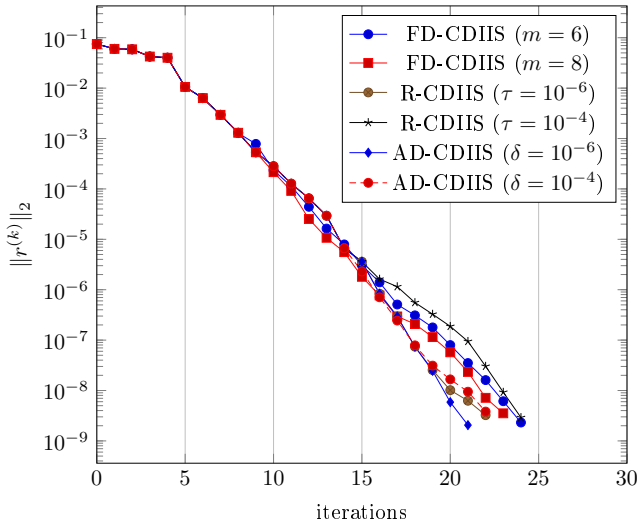
(a) Cadmium-imidazole complex in the RKS/B3LYP model with basis 3-21G.



(b) Glycine molecule in the RKS/B3LYP model with basis 6-31Gs.



(c) Dimethylnitramine molecule in the RHF model with basis 6-31G.



(d) Galactonolactone molecule in the RHF model with basis 6-31G.

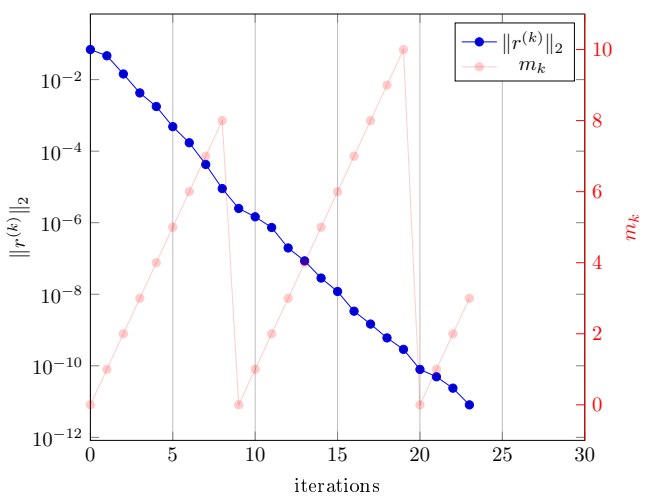
Figure 3: Residual norm convergence for the fixed-depth, restarted and adaptive-depth CDIIS on different molecular systems using an initial guess provided by a globally convergent method.

one whose depth is continuously adapted according to a criterion that appears to be new.

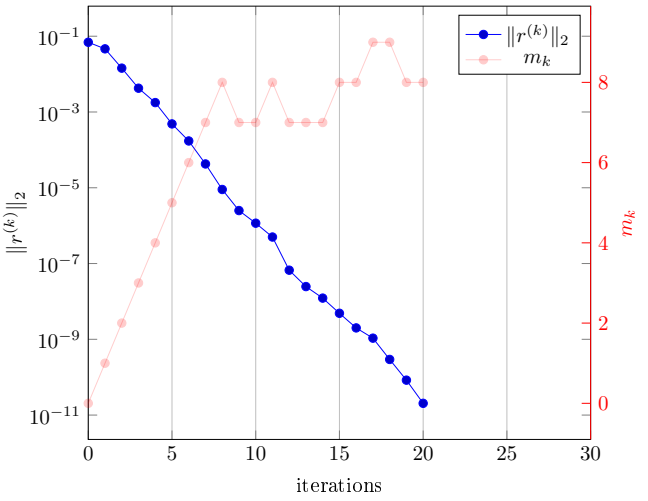
Our main convergence results are obtained in a more general setting and rely on weaker assumptions than those existing in the literature for the the DIIS [48] or the Anderson acceleration [57]. First, we do not impose a direct relation between the function g of the fixed-point iteration used to compute the solution and the error function f used for the extrapolation. Second, the nondegeneracy hypothesis on the solution of the problem is weakened: it only involves the restriction of the function f to a submanifold Σ containing the range of the function g. Such generalisations are necessary in order to deal with the self-consistent field iterations in quantum chemistry, for which the DIIS and the CDIIS were originally introduced. As we already stated before, these results also cover the particular cases of the Anderson acceleration (for which $\Sigma = \mathbb{R}^n$ and $g = \mathrm{id} + f$) and of the DIIS if $\Sigma = \mathbb{R}^n$.

Another novelty of our work is the absence of assumption concerning the uniform boundedness of the extrapolation coefficients. Indeed, the proposed restart and adaptive-depth mechanisms allow us to *a priori* prove such a bound. To our knowledge, the only other work where a similar estimate can be found is [63], in which a global linear convergence analysis for a stabilized variant of the type-I Anderson acceleration is given. As far as we know, the present article thus provides the first *complete* proof of accelerated convergence for a family of adaptive-depth extrapolation algorithms which includes intances of the DIIS, the CDIIS or the Anderson acceleration.

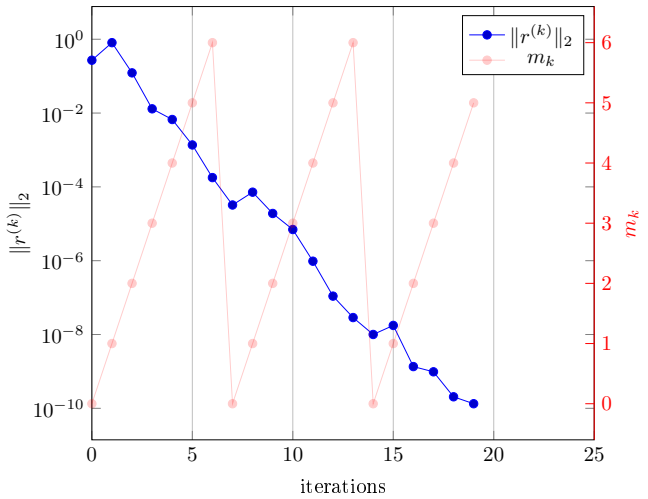
Finally, numerical experiments illustrate the good performances of the restarted and adaptive-depth acceleration algorithms applied to the numerical computation of the electronic ground state of various molecular systems. It



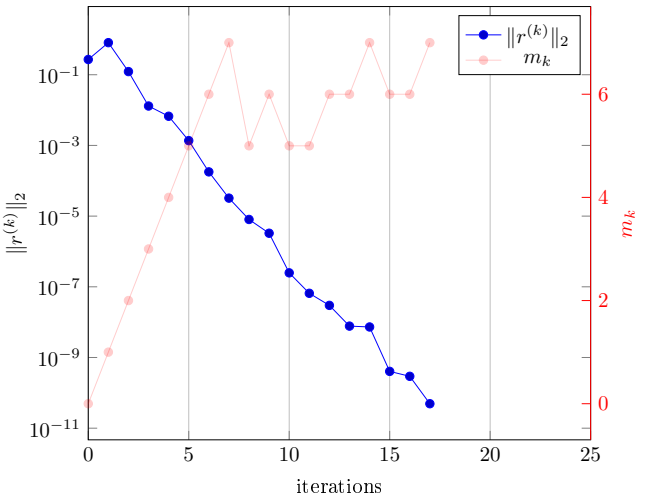
(a) Restarted CDIIS with $\tau = 10^{-4}$ for the dimethylnitramine molecule in the RHF model with basis 6-31G.



(b) Adaptive-depth CDIIS with $\delta = 10^{-4}$ for the dimethylnitramine molecule in the RHF model with basis 6-31G.



(c) Restarted CDIIS with $\tau = 10^{-4}$ for the glycine molecule in the RKS/B3LYP model with basis 6-31Gs.



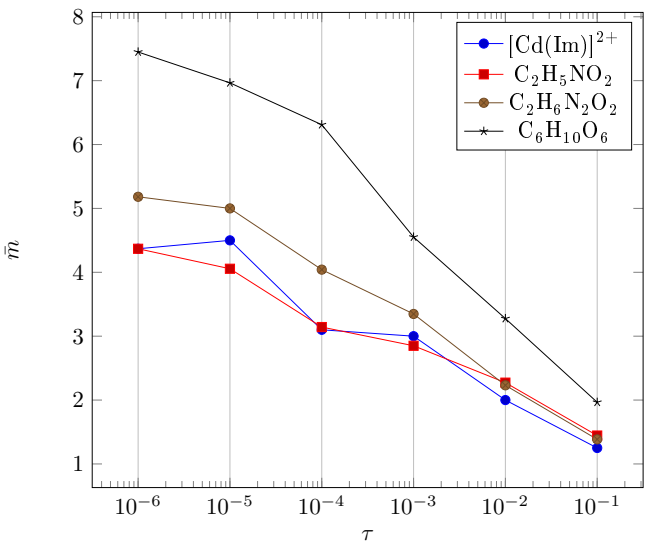
(d) Adaptive-depth CDIIS with $\delta = 10^{-4}$ for the glycine molecule in the RKS/B3LYP model with basis 6-31Gs.

Figure 4: Residual norm convergence and corresponding depth for the restarted and adaptive-depth CDIIS on the dimethylnitramine and glycine molecules.

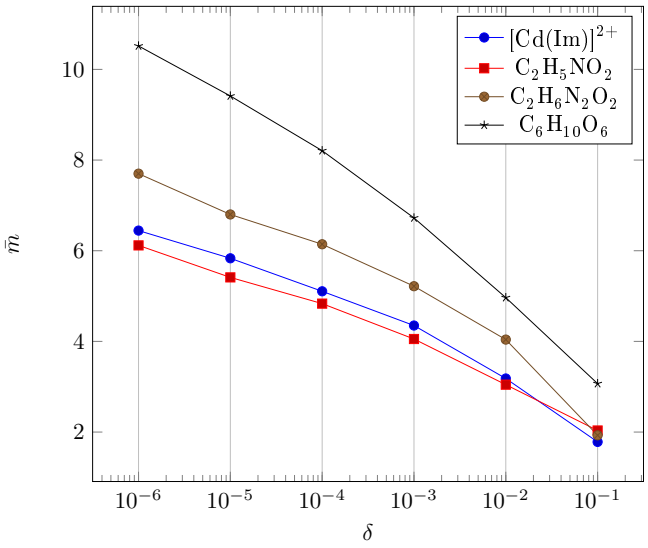
has been observed that, with an adequate choice of their respective parameters, both variants exhibit a better convergence rate than their fixed-depth counterpart. In particular, the adaptive-depth variant shows good promise. It has also been noticed that acceleration occurs for values of the parameters several orders of magnitude larger than the theoretical estimates, and that the size of the set of stored iterates at each step is on average smaller than the fixed "rule of thumb" values generally found in implementations of the CDIIS. Understanding the reasons of this key fact will require further effort.

Acknowledgement

The authors wish to warmly thank Antoine Levitt, for providing them with encouragements, useful references, and insightful remarks on a first draft of the manuscript, and Qiming Sun, for his help with the PySCF package. They also thank the anonymous reviewers whose comments and suggestions helped improve and clarify this manuscript. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement MDFT No 725528 of Mathieu Lewin).



(a) Depth mean $\bar{m}$ as a function of the restart parameter τ.



(b) Depth mean $\bar{m}$ as a function of the adaptive-depth parameter δ.

Figure 5: Evolution of the depth mean for different molecular systems and models.

References

[1] H. An, X. Jia, and H. F. Walker. Anderson acceleration and application to the three-temperature energy equations. *J. Comput. Phys.*, 347:1–19, 2017. DOI: 10.1016/j.jcp.2017.06.031.

[2] A. Anantharaman and E. Cancès. Existence of minimizers for Kohn–Sham models in quantum chemistry. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 26(6):2425–2455, 2009. DOI: 10.1016/j.anihpc.2009.06.003.

[3] D. G. Anderson. Iterative procedures for nonlinear integral equations. *J. ACM*, 12(4):547–560, 1965. DOI: 10.1145/321296.321305.

[4] D. G. M. Anderson. Comments on "Anderson acceleration, mixing and extrapolation". *Numer. Algorithms*, 80(1):135–234, 2019. DOI: 10.1007/s11075-018-0549-4.

[5] A. S. Banerjee, P. Suryanarayana, and J. E. Pask. Periodic Pulay method for robust and efficient convergence acceleration of self-consistent field iterations. *Chem. Phys. Lett.*, 647:31–35, 2016. DOI: 10.1016/j.cplett.2016.01.033.

[6] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38(6):3098–3100, 1988. DOI: 10.1103/PhysRevA.38.3098.

[7] C. Brezinski, M. Redivo-Zaglia, and Y. Saad. Shanks sequence transformations and Anderson acceleration. *SIAM Rev.*, 60(3):646–669, 2018. DOI: 10.1137/17M1120725.

[8] C. G. Broyden. A class of methods for solving nonlinear simultaneous equations. *Math. Comput.*, 19(92):577–593, 1965. DOI: 10.1090/S0025-5718-1965-0198670-6.

[9] M. T. Calef, E. D. Fichtl, J. S. Warsa, M. Berndt, and N. N. Carlson. Nonlinear Krylov acceleration applied to a discrete ordinates formulation of the k-eigenvalue problem. *J. Comput. Phys.*, 238:188–209, 2013. DOI: 10.1016/j.jcp.2012.12.024.

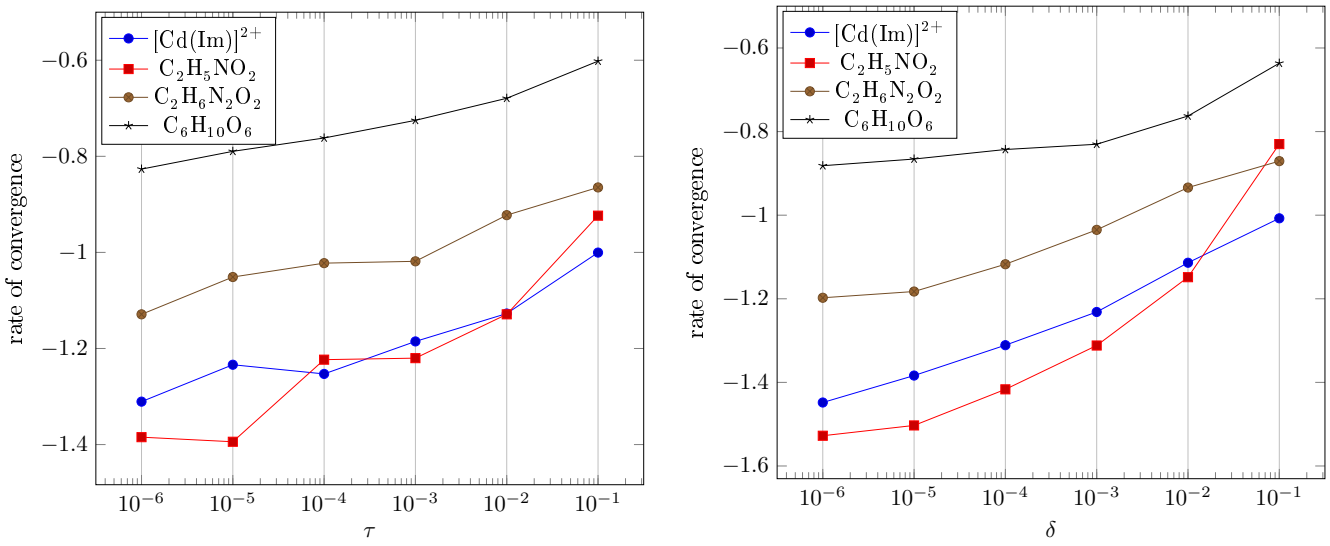
[10] É. Cancès and C. Le Bris. Can we outperform the DIIS approach for electronic structure calculations? *Internat. J. Quantum Chem.*, 79(2):82–90, 2000. DOI: 10.1002/1097-461X(2000)79:2<82::AID-QUA3>3.0.CO;2-I.

[11] É. Cancès and C. Le Bris. On the convergence of SCF algorithms for the Hartree–Fock equations. *ESAIM Math. Model. Numer. Anal.*, 34(4):749–774, 2000. DOI: 10.1051/m2an:2000102.

[12] N. N. Carlson and K. Miller. Design and application of a gradient-weighted moving finite element code I: in one dimension. *SIAM J. Sci. Comput.*, 19(3):728–765, 1998. DOI: 10.1137/S1064827594269955X.

[13] X. Chen and C. T. Kelley. Convergence of the EDIIS algorithm for nonlinear equations. *SIAM J. Sci. Comput.*, 41(1):A365–A379, 2019. DOI: 10.1137/18M1171084.

[14] P. Császár and P. Pulay. Geometry optimization by direct inversion in the iterative subspace. *J. Mol. Struct.*, 114:31–34, 1984. DOI: 10.1016/S0022-2860(84)87198-7.



(a) Rate of convergence as a function of the restart parameter τ.

(b) Rate of convergence as a function of the adaptive-depth parameter δ.

Figure 6: Evolution of the rate of convergence for different molecular systems.

[15] H. De Sterck. A nonlinear GMRES optimization algorithm for canonical tensor decomposition. *SIAM J. Sci. Comput.*, 34(3):A1351–A1379, 2012. DOI: 10.1137/110835530.

[16] E. De Sturler. Truncation strategies for optimal Krylov subspace methods. *SIAM J. Numer. Anal.*, 36(3):864–889, 1999. DOI: 10.1137/S0036142997315950.

[17] V. Eckert, P. Pulay, and H.-J. Werner. *Ab initio* geometry optimization for large molecules. *J. Comput. Chem.*, 18(12):1473–1483, 1997. DOI: 10.1002/(SICI)1096-987X(199709)18:12<1473::AID-JCC5>3.0.CO;2-G.

[18] C. Evans, S. Pollock, L. G. Rebholz, and M. Xiao. A proof that Anderson acceleration increases the convergence rate in linearly converging fixed point methods (but not in quadratically converging ones). *SIAM J. Numer. Anal.*, 58(1):788–810, 2020. DOI: 10.1137/19M1245384.

[19] V. Eyert. A comparative study on methods for convergence acceleration of iterative vector sequences. *J. Comput. Phys.*, 124(2):271–285, 1996. DOI: 10.1006/jcph.1996.0059.

[20] H.-R. Fang and Y. Saad. Two classes of multisecant methods for nonlinear acceleration. *Numer. Linear Algebra Appl.*, 16(3):197–221, 2009. DOI: 10.1002/nla.617.

[21] J.-L. Fattebert. Accelerated block preconditioned gradient method for large scale wave functions calculations in density functional theory. *J. Comput. Phys.*, 229(2):441–452, 2010. DOI: 10.1016/j.jcp.2009.09.035.

[22] V. Fock. Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Z. Phys.*, 61(1-2):126–148, 1930. DOI: 10.1007/BF01340294.

[23] V. Ganine, N. J. Hills, and B. L. Lapworth. Nonlinear acceleration of coupled fluid-structure transient thermal problems by Anderson mixing. *Internat. J. Numer. Methods Fluids*, 71(8):939–959, 2013. DOI: 10.1002/fld.3689.

[24] A. J. Garza and G. E. Scuseria. Comparison of self-consistent field convergence acceleration techniques. *J. Chem. Phys.*, 137(5):054110, 2012. DOI: 10.1063/1.4740249.

[25] D. M. Gay and R. B. Schnabel. Solving systems of nonlinear equations by Broyden's method with projected updates. Working Paper 169, National Bureau of Economic Research, 1977. DOI: 10.3386/w0169.

[26] A. Greenbaum, V. Pták, and Z. Strakoš. Any nonincreasing convergence curve is possible for GMRES. *SIAM. J. Matrix Anal. Appl.*, 17(3):465–469, 1996. DOI: 10.1137/S0895479894275030.

[27] A. Griewank. Broyden updating, the good and the bad! *Documenta Math.*, extra volume: optimization stories:301–315, 2012.

[28] R. Haelterman, J. Degroote, D. Van Heule, and J. Vierendeels. On the similarities between the quasi-Newton inverse least squares method and GMRES. *SIAM J. Numer. Anal.*, 47(6):4660–4679, 2010. DOI: 10.1137/090750354.

[29] G. G. Hall. The molecular orbital theory of chemical valency. VIII. A method of calculating ionization potentials. *Proc. Roy. Soc. London Ser. A*, 205(1083):541–552, 1951. doi: 10.1098/rspa.1951.0048.

[30] D. R. Hartree. The wave mechanics of an atom with a non-Coulomb central field. Part I. Theory and methods. *Math. Proc. Cambridge Philos. Soc.*, 24(1):89–110, 1928. doi: 10.1017/S0305004100011919.

[31] N. C. Henderson and R. Varadhan. Damped Anderson acceleration with restarts and monotonicity control for accelerating EM and EM-like algorithms. *J. Comput. Graph. Statist.*, 28(4):834–846, 2019. doi: 10.1080/10618600.2019.1594835.

[32] N. J. Higham and N. Strabić. Anderson acceleration of the alternating projections method for computing the nearest correlation matrix. *Numer. Algorithms*, 72(1):1021–1042, 2016. doi: 10.1007/s11075-015-0078-3.

[33] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136(3B):B864–B871, 1964. doi: 10.1103/PhysRev.136.B864.

[34] X. Hu and W. Yang. Accelerating self-consistent field convergence with the augmented Roothaan–Hall energy function. *J. Chem. Phys.*, 132(5):054109, 2010. doi: 10.1063/1.3304922.

[35] M. Kawata, C. M. Cortis, and R. A. Friesner. Efficient recursive implementation of the modified Broyden method and the direct inversion in the iterative subspace method: acceleration of self-consistent calculations. *J. Chem. Phys.*, 108(11):4426–4438, 1998. doi: 10.1063/1.475854.

[36] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140(4A):A1133–A1138, 1965. doi: 10.1103/physrev.140.a1133.

[37] K. N. Kudin and G. E. Scuseria. Converging self-consistent field equations in quantum chemistry – Recent achievements and remaining challenges. *ESAIM Math. Model. Numer. Anal.*, 41(2):281–296, 2007. doi: 10.1051/m2an:2007022.

[38] K. N. Kudin, G. E. Scuseria, and E. Cancès. A black-box self-consistent field convergence algorithm: one step closer. *J. Chem. Phys.*, 116(19):8255–8261, 2002. doi: 10.1063/1.1470195.

[39] C. Lee, W. Yang, and R. G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37(2):785–789, 1988. doi: 10.1103/PhysRevB.37.785.

[40] P. A. Lott, H. F. Walker, C. S. Woodward, and U. M. Yang. An accelerated Picard method for nonlinear systems related to variably saturated flow. *Adv. in Water Res.*, 38:92–101, 2012. doi: 10.1016/j.advwatres.2011.12.013.

[41] J. Nocedal and S. J. Wright. *Numerical optimization*. Springer series in operations research and financial engineering. Springer-Verlag New York, second edition, 2006. doi: 10.1007/978-0-387-40065-5.

[42] A. L. Pavlov, G. V. Ovchinnikov, D. Y. Derbyshev, D. Tsetserukou, and I. V. Oseledets. AA-ICP: iterative closest point with Anderson acceleration. arXiv:1709.05479 [cs.RO], 2017.

[43] F. A. Potra. On Q-order and R-order of convergence. *J. Optim. Theory Appl.*, 63(3):415–431, 1989. doi: 10.1007/BF00939805.

[44] F. A. Potra and H. Engler. A characterization of the behavior of the Anderson acceleration on linear problems. *Linear Algebra Appl.*, 438(3):393–398, 2013. doi: 10.1016/j.laa.2012.09.008.

[45] P. P. Pratapa and P. Suryanarayana. Restarted Pulay mixing for efficient and robust acceleration of fixed-point iterations. *Chem. Phys. Lett.*, 635:69–74, 2015. doi: 10.1016/j.cplett.2015.06.029.

[46] P. Pulay. Convergence acceleration of iterative sequences. The case of SCF iteration. *Chem. Phys. Lett.*, 73(2):393–398, 1980. doi: 10.1016/0009-2614(80)80396-4.

[47] P. Pulay. Improved SCF convergence acceleration. *J. Comput. Chem.*, 3(4):556–560, 1982. doi: 10.1002/jcc.540030413.

[48] T. Rohwedder and R. Schneider. An analysis for the DIIS acceleration method used in quantum chemistry calculations. *J. Math. Chem.*, 49(9):1889–1914, 2011. doi: 10.1007/s10910-011-9863-y.

[49] C. C. J. Roothaan. New developments in molecular orbital theory. *Rev. Modern Phys.*, 23(2):69–89, 1951. doi: 10.1103/RevModPhys.23.69.

[50] Y. Saad and M. H. Schultz. GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Statist. Comput.*, 7(3):856–869, 1986. doi: 10.1137/0907058.

[51] H. Sellers. The C^2-DIIS convergence acceleration algorithm. *Internat. J. Quantum Chem.*, 45(1):31–41, 1993. doi: 10.1002/qua.560450106.

[52] H. Shepard and M. Minkoff. Some comments on the DIIS method. *Mol. Phys.*, 105(19-22):2839–2848, 2007. doi: 10.1080/00268970701691611.

[53] M. Spivak. *A comprehensive introduction to differential geometry*, volume one. Publish or Perish, third edition, 1999.

[54] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K.-L. Chan. PySCF: the Python-based simulations of chemistry framework. *WIREs Comput. Mol. Sci.*, 8(1):e1340, 2017. doi: 10.1002/wcms.1340.

[55] L. Thøgersen, J. Olsen, A. Köhn, P. Jørgensen, P. Sałek, and T. Helgaker. The trust-region self-consistent field method in Kohn–Sham density-functional theory. *J. Chem. Phys.*, 123(7):074103, 2005. doi: 10.1063/1.1989311.

[56] A. Toth, J. A. Ellis, T. Evans, S. Hamilton, C. T. Kelley, R. Pawlowski, and S. Slattery. Local improvement results for Anderson acceleration with inaccurate function evaluations. *SIAM J. Sci. Comput.*, 39(5):S47–S65, 2017. doi: 10.1137/16M1080677.

[57] A. Toth and C. T. Kelley. Convergence analysis for Anderson acceleration. *SIAM J. Numer. Anal.*, 53(2):805–819, 2015. doi: 10.1137/130919398.

[58] H. F. Walker and P. Ni. Anderson acceleration for fixed-point iterations. *SIAM J. Numer. Anal.*, 49(4):1715–1735, 2011. doi: 10.1137/10078356X.

[59] Y. A. Wang, C. Y. Yam, Y. K. Chen, and G. Chen. Linear-expansion shooting techniques for accelerating self-consistent field convergence. *J. Chem. Phys.*, 134(24):241103, 2011. doi: 10.1063/1.3609242.

[60] T. Washio and C. W. Oosterlee. Krylov subspace acceleration for nonlinear multigrid schemes. *Electron. Trans. Numer. Anal.*, 6:271–290, 1997.

[61] J. Willert, W. T. Taitano, and D. Knoll. Leveraging Anderson acceleration for improved convergence of iterative solutions to transport systems. *J. Comput. Phys.*, 273:278–286, 2014. doi: 10.1016/j.jcp.2014.05.015.

[62] D. M. Wood and A. Zunger. A new method for diagonalising large matrices. *J. Phys. A Math. Gen.*, 18(9):1343–1359, 1985. doi: 10.1088/0305-4470/18/9/018.

[63] J. Zhang, B. O'Donoghue, and S. Boyd. Globally convergent type-I Anderson acceleration for non-smooth fixed-point iterations. *SIAM J. Optim.*, 30(4):3170–3197, 2020. doi: 10.1137/18M1232772.

[64] J. Zhang, Y. Yao, Y. Peng, H. Yu, and B. Deng. Fast K-Means clustering with Anderson acceleration. arXiv:1805.10638 [cs.LG], 2018.

[65] Y. A. Zhang and Y. A. Wang. Perturbative total energy evaluation in self-consistent field iterations: tests on molecular systems. *J. Chem. Phys.*, 130(14):144116, 2009. doi: 10.1063/1.3104662.