An analysis for the DIIS acceleration method used in quantum chemistry calculations

Thorsten Rohwedder and Reinhold Schneider

Abstract. This work features an analysis for the acceleration technique DIIS that is standardly used in most of the important quantum chemistry codes, e.g. in DFT and Hartree-Fock calculations and in the Coupled Cluster method. Taking up results from [23], we show that for the general nonlinear case, DIIS corresponds to a projected quasi-Newton/ secant method. For linear systems, we establish connections to the well-known GMRES solver and transfer according (positive as well as negative) convergence results to DIIS. In particular, we discuss the circumstances under which DIIS exhibits superlinear convergence behaviour. For the general nonlinear case, we then use these results to show that a DIIS step can be interpreted as step of a quasi-Newton method in which the Jacobian used in the Newton step is approximated by finite differences and in which the according linear system is solved by a GMRES procedure, and give according convergence estimates.

Mathematics Subject Classification (2010). 49M15; 65B99; 65Z05.

Keywords. DIIS, quantum chemistry, electronic structure calculation, acceleration, nonlinear equations, Newton's method, Broyden's method, GMRES, SCF algorithms, Hartree-Fock.

1. Introduction

The DIIS (Direct Inversion in the Iterative Subspace) method introduced by Pulay [37, 38] is an acceleration technique for solvers for nonlinear problems that was originally designed to accelerate the self consistent field iteration (SCF, cf. [24]), but has been found to be useful in a much broader context to improve convergence for a variety of algorithms used in electronic structure calculations. Nowadays, DIIS is standardly used on top of the popular algorithms used in the context of density functional theory (DFT) and Hartree-Fock calculations [18, 43], being the two most important methods for quantitative studies of larger electronic systems. In the context of the SCF algorithm, DIIS stands out as the fastest method to get to a minimum

once in the convergence region [30]: it is also successfully applied to the iterative solution of the Coupled Cluster (CC) equations [24], being mostly the method of choice for qualitative description of smaller systems. Variants of DIIS have also proven to be extremely efficient for simultaneous computation of eigenvalues and corresponding eigenvectors (RMM-DIIS, [29]) and when having to deal with the problem of charge sloshing that sometimes appears when DFT is applied to metallic systems [29]. Aside from electronic structure calculation, DIIS is also utilized in molecular dynamics for geometry optimization [13].

In an abstract framework, the DIIS method can be phrased as follows: The above underlying problems of quantum chemistry are (typically nonlinear) equations of the form

$$g(x^*) = 0. \tag{1.1}$$

The basic iterative methods used in the context of, say, DFT or CC, consist in an update step $x_{n+1} := x_n + r_n$, where the residual-like correction term, i.e. $r_n = -g(x_n)$ or a preconditioned, damped or approximate variant of this, is computed from the current iterate x_n. In contrast, DIIS exploits not only the information contained in x_n and r_n but considers a number of previously computed iterates. DIIS defines $\tilde{x}_{n+1} := x_n + r_n$ in an intermediate step, and then computes in a supplementary step improved iterates

$$x_{n+1} = \sum_{i=\ell(n)}^{n} c_i \tilde{x}_{i+1} = \sum_{i=\ell(n)}^{n} c_i (x_i + r_i) \quad \text{with} \quad \sum_{i=\ell(n)}^{n} c_i = 1$$

by minimizing the least square functional

$$J_{DIIS}(y) \;:=\; \frac{1}{2} \Big\| \sum_{i=\ell(n)}^{n} c_i r_i \Big\|^2 \tag{1.2}$$

over the set of all coefficient vectors $(c_i)_{i=\ell(n)}^{n}$ for which $\sum_{i=\ell(n)}^{n} c_i = 1$. Usually, only a short history of previous iterates is considered, i.e. $n - \ell(n) + 1$ is a small number in the above.

Often, the basic algorithm $\tilde{x}_{n+1} = x_n + r(x_n)$ already produces a linearly

	basic iteration	with DIIS
DFT calculation for cinchonidine	43	22
CCSD calculation for N_2, cc-pVTZ	21	12
CCSD calculation for LiH, cc-pVQZ	43	21

Figure 1: Iterations needed to converge some sample sample DFT/CCSD calculations with and without DIIS. DFT calculation performed with bigDFT [5, 18], a part of the ABINIT package [2, 19, 20]), CC calculations performed with NWChem [8, 27].

convergent sequence of iterates, see e.g. [9, 40, 43] for results in the context of quantum chemistry, and if this is the case, DIIS typically approximately halves the number of iteration steps needed to reach a prescribed precision, see the sample calculations in Figure 1. For this reason DIIS is often termed a *convergence acceleration method*; nevertheless, there are even cases e.g. in the SCF iteration for DFT calculations where convergence of the basic algorithm $\tilde{x}_{n+1} := x_n + r(x_n)$ does not have to be guaranteed for the DIIS accelerated procedure to converge. If the actual iterates are close to the solution, there are cases in which DIIS exhibits superlinear convergence. As an alternative to DIIS, Broyden-like quasi-Newton methods [11] have been proposed for use in quantum chemistry, and comparison of DIIS with those methods, e.g. with BFGS, show that the methods behave quite similarly. There are cases in the context of molecular dynamics where BFGS seems to be slightly better if the problem under consideration is not well-conditioned [15]; in the case of charge sloshing, DIIS seems to be superior to Broyden's method [29]. Incorporation of these and other ideas related to DIIS into the various physical applications of quantum chemistry has led to a further improvement and additional variants of DIIS and other acceleration techniques (often also termed "mixing schemes") without adding significant further costs, see e.g. [10, 15, 23, 26, 45, 47]. In particular, a projected version of *Broyden's backward* or *second method* (usually referred to as "bad Broyden method") that was already proposed along with other Broyden type methods in 1978 [17], but not recommended there and scarcely used in practice for its seemingly inferior convergence behaviour, has recently turned out to be a highly efficient method for the numerical treatment of nonlinear equations of quantum chemistry in comparison with other Broyden-type methods [33].

In the present paper we will show that in exact arithmetics, DIIS corresponds exactly to this projected backward Broyden's method mentioned above (see Section 3), thus making the success of its application to quantum chemical problems no surprise at all – a fact that is also implicitly contained in [23, 45]. We first rewrite DIIS as a Broyden type formula and then discuss the relation to some other Broyden type methods and methods proposed in the context of quantum chemistry. In Section 4, we analyze as a model problemn the convergence behaviour of DIIS when applied to linear equations. We establish a relation to the well-known GMRES scheme and use this relation to derive some (positive as well as negative) convergence estimates for DIIS applied to linear equations in Theorem 4.3. In particular, we will find that for the linear case, DIIS quite natually can turn non-convergent iterations into convergent ones. Section 5 then provides some convergence results for the nonlinear case. First of all, we prove in Theorem 5.2 that the DIIS procedure as given in Figure 2 is linearly convergent; in Theorem 5.4, we will use linear convergence and both the relation to Broyden-like methods and to the GMRES procedure to give a second, more refined convergence estimate, showing that DIIS can be interpreted as a quasi-Newton method in which the linearized (Newton) equation is solved approximately by a GMRES step for the linear system and

in which the Jacobian is approximated by finite differences. In practice, "superlinear" convergence behaviour of DIIS is often observed in the sense that the ratio $\|r_{n+1}\|/\|r_n\|$ of successive residual norms decreases, and in the light of the analysis given here, we will along the way discuss the circumstances under which this behaviour can/cannot set in, see Sections 3.3, 4.4 and 5.2.

2. Notations and basic facts about DIIS

Throughout this work, we will denote by V a given Hilbert space with an inner product $\langle \cdot, \cdot \rangle$, and denote the induced norm by $\| \cdot \|$. The dimension of V will be denoted by D, where $D = \infty$ is admitted. We will be concerned with the root problem (1.1), where g is a function that maps the space V to itself. The more general case where W is another Hilbert space and a root of a function $g : V \to W$ has to be computed is included if we can assume that the Jacobian $J^* \in L(V, W)$ of g at x^* is invertible, and that we have a cheaply applicable preconditioner $P \in L(V, W)$ at hand that approximates J^* sufficiently well: Under these circumstances, P is also invertible, and instead of computing the roots of g, we turn our attention to the function $\tilde{g}(x) = P^{-1}g(x)$ that has the same roots as g, but maps $V \to V$. In particular, the case where V is some appropriate Sobolev space $H^s(\Omega)$ and $W = V'$ is also covered. Also, the case where the basic iteration is damped or overrelaxated by a fixed parameter α is included by considering $\tilde{g}(x) = \alpha g(x)$.

The DIIS procedure outlined in the introduction results in the algorithm from [37, 38], displayed in Figure 2. The solution of the constraint minimization problem in step (4) is usually computed from by application of standard Lagrangian calculus to (2.1): this results in the linear system

$$\begin{pmatrix} \mathbf{B} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{c} \\ \lambda \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix} \tag{2.3}$$

with $\mathbf{B}$ determined by the matrix coefficients $b_{j,k} = \langle g(x_j), g(x_k) \rangle$, $\ell(n) \leq j, k \leq n$ and $\mathbf{1} = (1 \dots 1)$ a vector of length $n - \ell(n) + 1$, see [24] for an explicit derivation. In step (3) of the algorithm, $\ell(n)$, determining the number $n-\ell(n)+1$ of previous iterates considered in the computation of x_{n+1}, will generally be fixed unless the system matrix $\mathbf{B}$ becomes ill-conditioned, in which case the number $\ell(n)$ is systematicly reduced, see e.g. [38] for details.

3. Equivalence of DIIS to a projected Broyden's method

3.1. Rewriting DIIS as a Broyden's method.

In the present section, we show that the DIIS algorithm from Figure 2 may be rewritten as a projected variant of a "reverse" Broyden's method,

$$x_{n+1} = x_n - H_n g(x_n),$$

Figure 2: The DIIS algorithm.

Initialization.

Function $g : V \to V$, starting value $x_0 \in V$, $n = 0$ given.

Loop over:

(1) Evaluate the residual $r_n = -g(x_n)$. Let $\tilde{x}_{n+1} := x_n + r_n$.

(2) Terminate if desired precision is reached, e.g. if $\|g(x_n)\| < \epsilon$.

(3) Choose a number of previous iterates $x_{\ell(n)}, \ldots, x_n$ to be considered during the DIIS procedure

such that $g(x_{\ell(n)}), \ldots, g(x_n)$ are linearly independent.

(4) Compute

$$c_i = \text{argmin}\{ \|\sum_{i=\ell(n)}^{n} c_i r_i\|_2 \mid \sum_{i=\ell(n)}^{n} c_i = 1\}. \tag{2.1}$$

(5) Let

$$x_{n+1} = \sum_{i=\ell(n)}^{n} c_i \tilde{x}_{i+1} = \sum_{i=\ell(n)}^{n} c_i x_i + \sum_{i=\ell(n)}^{n} c_i r_i, \tag{2.2}$$

and set $n \leftarrow n + 1$.

End of loop.

---

wherein H_n is a secant approximation of the inverse of the Jacobian matrix of g at x_n, obtained from the previous iterates $x_{\ell(n)}, \ldots, x_n$ and associated function evaluations $g(x_{\ell(n)}), \ldots, g(x_n)$. This fact, also observed informally in [23] and in the context of SCF in [45], will be the basis for the discussion of its relation to other Broyden-type methods in parts (ii) and (iii), and to the analysis given below. We need some preparations, taken care of next.

Definition 3.1. *(Spaces of differences)*

For a given sequence of iterates $x_0, x_1, \ldots, x_n$ produced by DIIS, we define for $i = 0, 1, \ldots, n-1$ the differences

$$s_i := x_{i+1} - x_i, \; y_i := g(x_{i+1}) - g(x_i) \tag{3.1}$$

as well for $n \geq 1$ as the spaces

$$K_n := \text{span}\{s_i \mid i = \ell(n), \ldots, n-1\}, \; Y_n := \text{span}\{y_i \mid i = \ell(n), \ldots, n-1\},$$

in particular, $K_n = Y_n := \emptyset$ if $\ell(n) = n$. We denote the orthogonal projector onto Y_n by Q_n. Finally, we define the projected differences

$$\hat{y}_0 := y_0; \; \hat{y}_n := y_n - \sum_{i=0}^{n-1} \frac{\hat{y_i}^T y_n}{\hat{y_i}^T \hat{y_i}} \; \hat{y}_i. \tag{3.2}$$

□

Theorem 3.2. *(Equivalence of DIIS and a projected Broyden's method)*

The compound iteration steps

$$x_n \to \tilde{x}_{n+1} \overset{DIIS}{\to} x_{n+1} \tag{3.3}$$

can equivalently be computed by the projected Broyden update formula

$$x_{n+1} \;=\; x_n - \big(C_n Q_n + (I - Q_n)\big) g(x_n) \;=:\; x_n - H_n g(x_n), \tag{3.4}$$

with the projector Q_n from Definition 3.1*, and in which C_n is a secant approximation to the inverse of the Jacobian on the space of differences Y^n, fixed by*

$$C_n \;=\; 0 \text{ on } Y_n^{\perp},\; C_n y_i = s_i \text{ for all } i \in \{\ell(n), \dots, n-1\}. \tag{3.5}$$

*If $\ell(n) = 0$ in each step, so that the full history of iterates is considered, the DIIS inverse Jacobian H_n can be calculated from the Jacobian H_{n-1} by the rank-*1 *update formula*

$$H_0 = I,\; H_{n+1} \;=\; H_n \;+\; \frac{(s_n - H_n y_n)\hat{y}_n^T}{\hat{y}_n^T y_n}, \tag{3.6}$$

*with the projected difference $\hat{y}_n$ defined in (*3.2*).*

Before we approach the proof of Theorem 3.2 and discuss the result in Section 3.2, we note at first that for arbitrary $n \in \mathbb{N}$, it is not hard to see that

$$\begin{aligned} \operatorname{span}\{g(x_{\ell(n)}), \dots, g(x_n)\} &= \operatorname{span}\{g(x_n), y_{\ell(n)}, \dots, y_{n-1}\} && (3.7)\\ &= \operatorname{span}\{g(x_n), Y_n\}. && (3.8) \end{aligned}$$

Therefore, the differences $y_{\ell(n)}, \dots, y_{n-1}$ are linearly independent because $g(x_{\ell(n)}), \dots, g(x_n)$ are by definition of the DIIS algorithm; in particular, the update formula (3.6) is well-defined.
We comprise some technical details needed for the proof of Theorem 3.2 in the next lemma. Note that (iii) shows the uniqueness of the solutions of the DIIS minimization task.

Lemma 3.3. *Let $n \in \mathbb{N}$ and a set of iterates $x_1, \dots, x_n$ be fixed.*

(i) *There holds for all $j \in \ell(n), \dots, n-1$ that*

$$\begin{aligned} K_n &= \operatorname{span}\{x_i - x_j \mid j \neq i = \ell(n), \dots, n-1\}, && (3.9)\\ Y_n &= \operatorname{span}\{g(x_i) - g(x_j) \mid j \neq i = \ell(n), \dots, n-1\}. && (3.10) \end{aligned}$$

(ii) *For any $\ell < n \in \mathbb{N}$, any set of vectors $v_\ell, \dots, v_n \in V$ and any set of coefficients $c_\ell, \dots, c_n \in \mathbb{R}$ for which $\sum_{i=\ell}^n c_i = 1$, we have*

$$\sum_{i=\ell}^{n} c_i v_i \;=\; v_j + \sum_{\substack{i=\ell \\ i \neq j}}^{n} c_i (v_i - v_j) \tag{3.11}$$

for all $j \in \{\ell, \dots, n\}$, *in particular;*

$$x_j + K_n \;=\; \Big\{ \sum_{i=\ell}^{n} c_i x_i \;\Big|\; \sum_{i=\ell}^{n} c_i = 1 \Big\}, \tag{3.12}$$

$$g(x_j) + Y_n \;=\; \Big\{ \sum_{i=\ell}^{n} c_i g(x_i) \;\Big|\; \sum_{i=\ell}^{n} c_i = 1 \Big\} \tag{3.13}$$

for all such j.

(iii) *There holds*

$$\min \Big\{ \Big\| \sum_{i=\ell(n)}^{n} c_i g(x_i) \Big\| \;\Big|\; \sum_{i=\ell(n)}^{n} c_i = 1 \Big\} \;=\; \|(I - Q_n) g(x_n)\|. \tag{3.14}$$

The minimizer $(c_i)_{i=\ell(n)}^{n}$ *is unique and fulfils*

$$\sum_{i=\ell(n)}^{n} c_i (g(x_i) - g(x_n)) \;=\; -Q_n g(x_n). \tag{3.15}$$

□

Proof. To prove (i), observe that for all $i \in \{\ell(n), \dots, n-1\}$,

$$s_i \;=\; x_{i+1} - x_i \;=\; x_{i+1} - x_j - (x_i - x_j) \;\in\; \operatorname{span}\{x_i - x_j \,|\, j \neq i = \ell(n), \dots, n-1\}$$

and that vice versa,

$$x_i - x_j = \sum_{k=i}^{j-1} s_k \in K^n \quad \text{if} \quad i < j, \qquad x_i - x_j = -\sum_{k=j}^{i-1} s_k \in K^n \quad \text{if} \quad i > j,$$

from which (3.9) follows. The proof for (3.10) is analogous. Equation (3.11) follows from the constraint condition $\sum_{i=\ell}^{n} c_i = 1$, yielding

$$\sum_{i=\ell}^{n} c_i v_i \;=\; v_j - (1 - c_j) v_j + \sum_{\substack{i=\ell \\ i \neq j}}^{n} c_i v_i \;=\; v_j + \sum_{\substack{i=\ell \\ i \neq j}}^{n} c_i (v_i - v_j)$$

for all $j \in \{\ell, \dots, k\}$. In particular, (3.13) follows from this together with (i), and implies

$$\inf \Big\{ \Big\| \sum_{i=\ell(n)}^{n} c_i g(x_i) \Big\| \;\Big|\; \sum_{i=\ell(n)}^{n} c_i = 1 \Big\} \;=\; \inf \{ \|g(x_n) - y\| \mid y \in Y_n \},$$

from which (3.14), (3.15) can be concluded from the best approximation properties of Hilbert spaces. Finally, (ii) together with (3.8) and the linear independence of $g(x_{\ell(n)}), \dots, g(x_n)$ implies in particular that the vectors $g(x_n) - g(x_i), i = \ell(n), \dots, n-1$ are linearly independent, so that the minimizer $(c_i)_{i=\ell(n)}^{n}$ is unique as coefficient vector of the best approximation of $g(x_n)$ in Y_n.

□

Proof of Theorem 3.2. By linearity, there follows that $C_n(g(x_i)-g(x_n)) = x_i - x_n$ for $i = \ell(n), \dots, n-1$, cf. the proof of Lemma 3.3. Using the definition of the DIIS iterates and Lemma 3.3, we obtain

$$\begin{aligned} x_{n+1} &= \sum_{i=\ell}^{n} c_i \tilde{x}_{i+1} = \sum_{i=\ell(n)}^{n} c_i x_i - \sum_{i=\ell(n)}^{n} c_i g(x_i) \\ &= x_n + \sum_{i=\ell(n)}^{n-1} c_i(x_i - x_n) - \Big(g(x_n) + \sum_{i=\ell(n)}^{n-1} c_i \big(g(x_i) - g(x_n)\big)\Big) \\ &= x_n + C_n \Big(\sum_{i=\ell(n)}^{n-1} c_i \big(g(x_i) - g(x_n)\big)\Big) \\ &\quad - \Big(g(x_n) + \sum_{i=\ell(n)}^{n-1} c_i \big(g(x_i) - g(x_n)\big)\Big) \\ &= x_n - C_n Q_n g(x_n) - (I - Q_n) g(x_n) =: x_n - H_n g(x_n). \end{aligned}$$

This proves (3.4) and (3.5). To show (3.6), we note first of all that for each $n \in \mathbb{N}_0$, H_n is fixed on Y_n by the condition $H_n y_i = s_i$ for all $i = \ell(n), \dots, n-1$, while on $Y_n^\perp$, $H_n = I$. We show by induction that the same holds for (3.6), which we denote by

$$\hat{H}_0 = I, \ \hat{H}_{n+1} = \hat{H}_n + \frac{(s_n - H_n y_n)\hat{y}_n^T}{\hat{y}_n^T y_n}$$

for a moment. For $n = 0$, the assertion holds because $Y_n = \emptyset$ and $\hat{H}_0 = I$ by definition. For $n \in \mathbb{N}$, we have for all $y \in Y_n^\perp$ that

$$\hat{H}_n y = \hat{H}_{n-1} y + \frac{(s_{n-1} - \hat{H}_{n-1} y_{n-1})\hat{y}_{n-1}^T}{\hat{y}_{n-1}^T y_{n-1}} y = y$$

because $\hat{y}_{n-1} \in Y_n$, so using the induction hypothesis, $\hat{H}_n = I$ on $Y_n^\perp$. Moreover, for all $i = 0, \dots, n-2$,

$$\hat{H}_n y_i = \hat{H}_{n-1} y_i + \frac{(s_{n-1} - \hat{H}_{n-1} y_{n-1})\hat{y}_{n-1}^T}{\hat{y}_{n-1}^T y_{n-1}} y_i = s_i + 0,$$

by induction hypothesis and definition of $\hat{y}_{n-1}$. Finally, for y_{n-1},

$$\hat{H}_n y_{n-1} = \hat{H}_{n-1} y_{n-1} + \frac{(s_{n-1} - \hat{H}_{n-1} y_{n-1})\hat{y}_{n-1}^T}{\hat{y}_{n-1}^T y_{n-1}} y_{n-1} = s_{n-1},$$

completing the proof.

□

The next lemma that will be needed later in the Section 4 on linear problems, but also holds in the nonlinear case and is therefore included here.

Lemma 3.4. *If for fixed $n \in \mathbb{N}$, $\ell(i) = 0$ for all $i = 1, \dots, n$, i.e. the full history of previous iterates has been used in every previous step of the DIIS procedure and in particular, $g(x_0), \dots, g(x_{n-1})$ are linearly independent, there holds*

$$K_n \;=\; \mathrm{span}\{g(x_0), \dots, g(x_{n-1})\}. \tag{3.16}$$

Proof. We prove (3.16) by induction on n. For $n = 1$, $g(x_0) = x_1 - x_0$, so the statement holds in this case. For arbitrary $n \in \mathbb{N}$, we exploit (3.8) again, so that to show the assertion for $n+1$, it suffices to show that $x_n - x_{n+1} \in \mathrm{span}\{g(x_n), Y_n\}$ and that $\dim K_{n+1} = n+1$: Using Theorem 3.2, we have

$$x_{n+1} - x_n \quad = C_n Q_n g(x_n) + (I - Q_n) g(x_n) \tag{3.17}$$

and the first term on the right side is an element of

$$K_n \subseteq \mathrm{span}\{g(x_0), \dots, g(x_{n-1})\}$$

by definition of C_n and induction hypothesis, while by the definition of the projector Q_n, the second is in $\mathrm{span}\{g(x_n), Y_n\}$. Because $g(x_0), \dots, g(x_{n-1})$ are linearly independent, the second component on the right hand side of (3.17) (orthogonal to Y_n) is nonzero, implying with (3.8) that $\dim K_{n+1} = n+1$. This completes the proof.

□

3.2. Relation and comparison to other Broyden-type methods.

Theorem 3.2 shows that the DIIS procedure can be interpreted as a Broyden's method, in which the iteration step of a Newton's method, consisting in the usually computationally too expensive solution of the linear system

$$J(x_n) s_n = -g(x_n) \tag{3.18}$$

with $J(x_n)$ denoting the Jacobian of g at x_n, is replaced by solving the equation

$$s_n = -H_n g(x_n). \tag{3.19}$$

Herein, H_n is a rank-$(n - \ell(n) - 1)$-update of the identity, approximating $J^{-1}(x_n)$ by exploiting the information about $J^{-1}(x_n)$ contained in the sequence of former iterates $x_{\ell(n)}, \dots, x_n$ and according sequence of function values $g(x_{\ell(n)}), \dots, g(x_n)$: For all $\ell(n), \dots, n-1$, the directional derivatives $J(x_n)s_n$ are approximated by mapping the corresponding finite differences y_n of function values to s_n, see (3.5). In pursuing the ansatz of using differences of formerly calculated quantities to approximate the Jacobian $J(x_n)$ (or its inverse), DIIS is thus similar to the various variants of Broyden's method (see e.g. [11, 36]), and we will discuss this relation a little deeper in the following. For this comparison, we suppose that $\ell(n) = 0$ for each n-th step of DIIS, so that the full history of iterates is considered in each step until DIIS terminates.

In Broyden's original method [6], starting in our setting with the initial approximate Jacobian $B_0 = I$, the approximate Jacobian B_{n+1} is a rank-1-update of B_n that fulfils the secant condition

$$B_{n+1} s_n = y_n \tag{3.20}$$

and has the additional property that the Frobenius norm[1] $\|B_{n+1} - B_n\|_F$ is minimal among all such possible updates B_{n+1}. The update is given by

$$B_0 = I, \qquad B_{n+1} = B_n + \frac{(y_n - B_n s_n) s_n^T}{s_n^T s_n}.$$

Although Broyden's method does not retain the original quadratic convergence of the (exact) Newton method (3.18), it is q-superlinearly convergent, meaning that the sequence of quotients

$$q_n := \frac{\|x_{n+1} - x^*\|}{\|x_n - x^*\|} \tag{3.21}$$

is not only bounded by a constant $c < 1$ as in the case of (q-)linear convergence, but converges to zero (see [11] for the classical case and [21] for extended results on the operator case). The DIIS-quasi-Newton method (3.4) is a combination of two variants of Broyden's method: The first one is the *reverse Broyden's method* in which the *inverse* $J(x^*)^{-1}$ of the Jacobian is approximated directly by successive rank-1-updates H_{n+1} fulfilling $H_{n+1} y_n = s_n$ and having minimal deviation with respect to the Frobenius norm from H_n, resulting in[2]

$$H_0 = I, \qquad H_{n+1} = H_n + \frac{(s_n - H_n y_n) y_n^T}{y_n^T y_n}.$$

Although this method is also termed as "bad Broyden's method" due to its convergence behaviour in practice, that is inferior to the above "forward" technique, the proof for q-superlinear convergence of the forward method can be modified to show that the reverse Broyden's method also converges q-superlinearly [7].

The second method related to (3.4) is a modification of the "forward" Broyden method, the *Broyden's method with projected updates* [17], developed further in [34]. It consists in the ansatz that the secant condition (3.20) should not only be fulfilled for the latest secant s_n, but by demanding

$$B_{n+1} s_i = y_i \text{ for all } 0 \leq i \leq n, \tag{3.22}$$

while in contrast, the approximations B_{n+1} computed in Broyden's method need not fulfil the condition. This results in the formula

$$B_0 = I, \qquad B_{n+1} = B_n + \frac{(y_n - B_n s_n) \hat{s}_n^T}{\hat{s}_n^T s_n}, \tag{3.23}$$

in which $\hat{s}_n$ is the orthogonalization of s_n against all previous differences s_i. The projected method has the advantage that when applied to linear problems, the exact solution is computed in the $(D+1)$-th step [17], a property

[1] The Frobenius norm is only defined in finite dimensional spaces V; in infinite dimensional spaces, the difference $B_{n+1} - B_n$ has to be a Hilbert-Schmidt operator for a meaningful extension of this concept. See [21] for an alternative, more global characterization of the Broyden update in infinite dimensional spaces V.

[2] This yields a method different from the "forward" Broyden method, for which B_n^{-1} can be computed as a (different) rank-1 update of B_{n-1}^{-1} by the Sherman-Woodbury-Morrison formula, see e.g. [11] for an introduction and a comparison of both methods.

that might also have positive effect on problems that are "close to linear" in the sense that the second order terms in the Taylor expansion are relatively small.[3] Comparison of (3.5) and (3.23) now shows that DIIS (with full history) is the *reverse variant of the projected Broyden's method*, and we already noted that the reverse method (i.e. DIIS) is also introduced in [17], Algorithm II', but not analysed further due to its practical behaviour which – as in the non-projected case – seems to be inferior to the forward method, also see [17] for comments on numerical tests.
This is in agreement with the outcome of [23], in which the forward projected method from [17] is re-introduced as an improvement of DIIS, termed the KAIN (Krylov Accelerated Inexact Newton) solver there and turning out to be superior to DIIS for the test problems attacked there. In contrast to this though is the more recent publication [33], which by testing different Broyden schemes on a range of quantum chemical problems comes to the conclusion that the Multi-Secant Second Broyden method (MSB2, again corresponding to DIIS) is the only one that converges in all of the test cases. Particularly successful is DIIS applied to the self consistent field iteration, where it is found to be superior to comparable methods [30]. The interested reader is also referred to [28] and the references given therein for more related Newton-type algorithms using Krylov spaces spanned by finite differences.

3.3. Superlinear convergence of DIIS? - Part I

We conjectured that as from the "good/forward Broyden's method" to the "bad/reverse Broyden's method", we might transfer theoretical results on q-superlinear convergence on the projected forward method (given in [17]) to the projected reverse variant, i.e. to DIIS. Unfortunately, the proof of q-superlinear convergence given in [17] is erroneous, and this flaw is not straightforward to mend. In order to obtain results like q-superlinear convergence (as a limit process for $n \to \infty$), a replacement/discarding strategy for former iterates that are "almost linearly dependent" will have to be formulated instead of just restarting the algorithm as in [17], and although practical approaches by an SVD of the DIIS inverse Jacobian have been formulated [44], the formulation of a rigurous stategy is to our mind far from obvious, also see the further comments in [39].
We will take a little different approach here and investigate the transient (i.e. "short-term") convergence behaviour of DIIS by treating it as a nonlinear variant of the well-known GMRES procedure. Although sometimes in practical DIIS calculations, "superlinear" convergence behaviour can be observed in the sense that the ratio $\|r(x_{n+1})\|/\|r(x_n)\|$ of the residuals decays with increasing iteration number n, our general experience with DIIS is rather reflected exemplarily in Figure 3, where for the sample calculations from Figure 1, the above ratio has been plotted against the number of iterations.

[3] It can also be shown that the classical Broyden's method computes the exact solution to a linear problem after $2D$ steps, see [16].

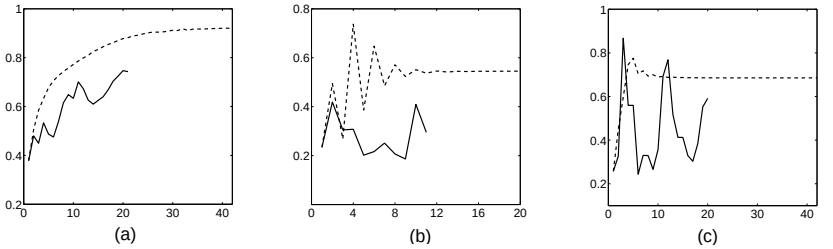

Figure 3: Ratio of the residuals $\|r(x_{n+1})\|/\|r(x_n)\|$ in the course of the iteration for the sample calculations displayed in Figure 1. (a) DFT for cinchonidine, (b) CCSD for N_2, (c) CCSD for LiH. Dashed line: basic iteration only, solid line: with DIIS acceleration.

In our theoretical analysis in Section 5, we will find that the worst-case short-term convergence behaviour of DIIS essentially depends on balancing two opposing error terms associated with the number of previous iterates considered, see the remark after Theorem 4.3 and also Section 4.4. To derive these results, we now first turn to the model case of linear problems.

4. DIIS applied to linear problems

4.1. Viewpoint and assumptions

As a model problem, we will now investigate the special case where DIIS is applied to a linear equation, i.e. for $A : V \to V$ linear and bounded, $b \in V$, an $x^* \in V$ is sought such that

$$g(x^*) = Ax^* - b \;=\; 0. \tag{4.1}$$

This approach will not give insights on the discussion of possible linear convergence of DIIS, but also be used as a basis for the analysis of DIIS applied to nonlinear equations. We use the negative gradient direction as update directions,

$$r(x_n) \;:=\; -g(x_n) \;=\; b - Ax_n. \tag{4.2}$$

By modifying g appropriately (see the remarks in Section 2), preconditioned or damped gradients used in the basic iteration scheme are also included; also, weak equations are covered. If we suppose that the iteration scheme is convergent, it is a well-known problem that convergence of this scheme can be extremely slow, especially if the condition number of A is large (see e.g. [22]). To overcome these problems, procedures like the well-known GMRES or cg solvers were developed, leading to accelerated convergence of the underlying scheme. We will now show that DIIS now inherits some of these properties from GMRES because the minimization problems solved and the subspaces used coincide.

We will assume that A is invertible, so that x^* is unique. Also, we will assume that in each step (2) of the DIIS algorithm displayed in Fig. 2, the full set of previous vectors $x_0, \ldots, x_n, g(x_0), \ldots, g(x_n)$ is used to minimize the least square functional (1.2). Note that this implies that the vectors $g(x_0), \ldots, g(x_n)$ have to be linearly independent, and we will see later that this indeed is the case unless $g(x_n) = 0$, in which case the algorithm terminates.

In the n-th step of a DIIS-accelerated gradient solver, the functional (1.2) is minimized over the space $x_n + K_n$. For the linear problem (4.1), this minimizer is by Lemma 3.3 and Lemma 3.4 given by

$$\begin{aligned} \bar{x} &= \operatorname{argmin}_{\sum_{i=1}^n c_i = 1} \{ \| \sum_{i=0}^n c_i g(x_i) \|^2 \} \\ &= \operatorname{argmin}_{(c_i)_{i=0}^{n-1} \in \mathbb{R}^n} \{ \| A(x_0 + \sum_{i=0}^{n-1} c_i r_i) - b \|^2 \}; \end{aligned}$$

and using linearity once more, it is not hard to see that the next DIIS iterate is given by

$$x_{n+1} := \sum_{i=0}^n c_i \tilde{x}_{i+1} := \sum_{i=0}^n c_i x_i + \sum_{i=0}^n c_i g(x_i) = \bar{x} + r(\bar{x}). \tag{4.3}$$

For a comparison of DIIS with Krylov subspace methods, we now introduce for a given starting value $v_0 \in V$, $n \geq 1$ the well-known Krylov spaces

$$K_n(A, r(v_0)) := \operatorname{span}\{A^i r(v_0) : i = 0, \ldots, n-1\}.. \tag{4.4}$$

We remind the reader that one point of view on Krylov subspace methods for linear systems is that they consist in iterating the two following steps:

(i) Minimize a given error functionals J over the space $v_0 + K_n(A, r(v_0))$ to obtain the next iterate $v_n \in v_0 + K_n(A, r(v_0))$.
(ii) Compute $A^n r(v_0)$ and construct the next Krylov space $K_n(A, r(v_0))$.

If A is symmetric, the well-known method of conjugate gradients ("cg") is an example for such a method, consisting in minimization of the functional

$$J_{\mathrm{cg}}(y) = \frac{1}{2} \langle A(y - x^*), y - x^* \rangle. \tag{4.5}$$

over the respective affine Krylov spaces $v_0 + K_n(A, r(v_0))$. With $\tilde{b} = b - Av_0$, and the minimizer written as $v_n = v_0 + \delta_n$, $\delta_n \in K_n(A, r(v_0))$, the first order condition for (4.5) is the Galerkin (orthogonality) condition $A\delta_n - \tilde{b} \perp K_n$, or more explicitly, $\langle A\delta_n - \tilde{b}, v \rangle = 0$ for all $v \in K_n$. Another example is given by the least-squares functional

$$J_{LS}(y) = \frac{1}{2} \langle A(y - x^*), A(y - x^*) \rangle = \frac{1}{2} \|Ay - b\|^2. \tag{4.6}$$

The first order condition for (4.6) is given by the Petrov-Galerkin condition

$$\langle A\delta_n - \tilde{b}, Av \rangle = 0 \text{ for all } v \in K_n. \tag{4.7}$$

This is an oblique projection method [41] with

$$A\delta_n - \tilde{b} \perp AK_n, \tag{4.8}$$

i.e. the residual $A\delta_n - \tilde{b}$ of the optimal subspace solution v_n is A-orthogonal to K_n, or, in other words, the difference $v_n - x^*$ to the true solution x^* is A^2-conjugate to K_n. The Krylov method associated with (4.6) is the well-known GMRES-method [41], which for symmetric matrices results in the method of conjugate residuals ("cr", see e.g. [22]).
Let us note for later purposes that the Krylov spaces (4.4) allow for the alternative characterization

$$K_n(A, r(v_0)) \;=\; \text{span}\,\{r(v_0), \dots, r(v_{n-1})\}, \tag{4.9}$$

see e.g. [22], Theorem 9.4.2 for a proof.

4.2. Connection between DIIS and GMRES

Comparison of (4.6) and (1.2) shows that the functionals used in GMRES and DIIS coincide. We will now further clarify the relation between DIIS and GMRES, and thus also between GMRES and the projected Broyden's method from Theorem 3.2. Although Broyden-like secant methods have been proposed as an alternative to GMRES to solve large-scale linear equations (see [12] and references therein), we are not aware of literature where the below connection between GMRES and the projected Broyden's method from Theorem 3.2 is made explicit. A similarity between DIIS and GMRES is noted in [30].

Lemma 4.1. *If the starting values of a GMRES procedure and a DIIS procedure applied to the linear system (4.1) coincide, $x_0 = v_0$, there holds*

$$K_n(A, r(v_0)) = K_n \tag{4.10}$$

for any $n \in \mathbb{N}$. The GMRES procedure and the DIIS procedure, applied to linear problems, therefore solve the same minimization problem in each step (only using a different parametrization). The iterates x_n of the DIIS procedure and the iterates v_n of GMRES are related by

$$x_{n+1} = v_n - r(v_n). \tag{4.11}$$

There holds

$$\|r(v_{n+1})\|_2 \;\leq\; \|r(x_{n+1})\|_2 \;\leq\; \|I - A\|_2 \|r(v_n)\|_2. \tag{4.12}$$

□

In Figure 4, the result of Lemma 4.1 is displayed in a flow chart comparing GMRES and DIIS; the iterates of DIIS are denoted by x_n, those of GMRES by v_n.

Proof. We use the representation (4.9) for the Krylov spaces, and the analogous one from Lemma 3.4 for the spaces used in DIIS,

$$K_n = \text{span}\{r(x_0), \dots, r(x_{n-1})\}.$$

Figure 4: The (linear) DIIS procedure vs. the GMRES algorithm

---

Initialization.

▷ Starting value $x_0 = v_0 \in V$, $n = 1$ given. Compute $r_0 = r(x_0)$, let $K_1 := \operatorname{span}\{r_0\}$.

▷ DIIS: Set $x_1 := x_0 + r(x_0)$. Compute $r_1 := r(x_1)$.
GMRES: Compute $r_1 := Ar_0$ from $r(v_0)$.

Loop over:

(1) Add r_n to K_n to obtain K_{n+1}. (The spaces K_n coincide for GMRES and DIIS, Lemma 4.1.)

(2) Calculate $\bar{x} \in x_0 + K_{n+1}$ which minimizes the residual over $x_0 + K_{n+1}$.

(3) DIIS: Let $x_{n+1} = \bar{x} - r(\bar{x})$. Compute $r_{n+1} = r(x_{n+1})$.
GMRES: Let $v_n = \bar{x}$. Compute $r_n := Ar(v_{n-1})$.

(4) Set $n \leftarrow n + 1$.

End of loop.

We proceed by induction. For $n = 1$,

$$K_1(A, r(v_0)) \;=\; \operatorname{span}\{r(v_0)\} = \operatorname{span}\{r(x_0)\} = K_1$$

holds trivially, and $x_1 = v_0 - r(v_0)$ holds by definition of DIIS. Now let the assertion hold for fixed $n \in \mathbb{N}$. We then get that

$$K_{n+1} = \operatorname{span}\{K_n, r(x_n)\},\; K_{n+1}(A, r(v_0)) = \operatorname{span}\{K_n(A, r(v_0)), r(v_n)\},$$

so that by hypothesis, it suffices to show $r(x_n) \in K_{n+1}(A, r(v_0))$ and that $r(v_n) \in K_{n+1}$. Using the hypothesis $x_n = v_{n-1} - r(v_{n-1})$, we have

$$r(x_n) \;=\; A(v_{n-1} + r(v_{n-1})) - b \;=\; r(v_{n-1}) + Ar(v_{n-1});$$

the first term to the right is in $K_n(A, r(v_0))$ according to (4.9), while the second is in $K_{n+1}(A, r(v_0))$ according to (4.4), so $r(x_n) \in K_{n+1}(A, r(v_0))$ follows. Vice versa,

$$Ar(v_{n-1}) = r(x_n) - r(v_{n-1}) \in K_{n+1}$$

because $r(x_n) \in K_{n+1}$ and $r(v_{n-1}) \in K_n(A, r(v_0)) = K_n$. Thus,

$$K_{n+1} = K_{n+1}(A, b),$$

and because the functionals (4.6) and (1.2) also coincide, DIIS and GMRES both compute the same minimizer $\bar{x}$ on $x_0 + K_{n+1}$. While GMRES sets $v_{n+1} = \bar{x}$ by definition, we have $x_n = \bar{x} - g(\bar{x})$ in DIIS, see (4.4). This shows (4.11).

For the left inequality of (4.12), note that $x_{n+1} \in x_0 + K_n$, and that v_n minimizes the 2-norm of the residual over that space. The estimate on the right hand side follows directly from (4.11).

□

Lemma 4.1 shows that we can interpret GMRES as a variant of the DIIS/projected Broyden method for linear problems, exhibiting in the symmetric case the well-known advantages like the shortening of history [22]. While in the linear cases, the Krylov spaces (4.4) and the space (4.9) containing the current residuals coincide, this is not the case anymore in the case of nonlinear problems, and the residuals $g(x_n)$ then have to be evaluated explicitly, leading to the DIIS method. DIIS can thus be interpreted as a globalization of the least square ansatz of GMRES to the nonlinear case. Because in GMRES, only the former two iterates have to be respected to compute the residual minimizer over the whole Krylov space, it will be interesting to investigate explicitly how the omission of former iterates influences the convergence of DIIS when applied to mildly nonlinear problems. As a first corollary, Lemma 4.1 implies the following termination property of "linear DIIS".

Corollary 4.2. *In exact arithmetic, the DIIS procedure, applied to the iteration scheme $\tilde{x}_n = x_n - r(x_n)$ for the linear equation (4.1), terminates after $n \leq D$ steps with the exact solution $x_n = x^*$.*

Proof. Let us note at first that from Lemma 4.1, the vectors (3.16) building the spaces K_n become linearly dependent if and only if the vectors in (4.9) become linearly dependent. It is well-known [22, 41] that for GMRES, there holds $r(v_i) \perp_A K_i(A, b)$ for all $i \in \mathbb{N}$. In particular, the vectors in (4.9) become linearly dependent if and only if $r(v_i) = 0$ and $v_i = x^*$ is the solution of (4.1), and this will happen at latest when $i = D - 1$. For the corresponding DIIS iterate, there then holds $x_{i+1} = v_i + r(v_i) = x^*$ by (4.11), completing the proof.

□

4.3. Convergence of DIIS for linear problems

We will now transfer well-known convergence properties of GMRES [32] to analyze the convergence behaviour of DIIS for the model problem of linear equations. Theorem 4.3 shows that as for GMRES, the worst-case convergence behavior of DIIS applied to normal matrices A is completely determined by the spectrum of A. In the nonnormal case however, the convergence behavior of the GMRES method may not be related to the eigenvalues in any simple way and understanding the convergence of GMRES in the general non-normal case still remains a largely open problem, and this property is thus also inherited by the DIIS procedure. The application of DIIS to nonnormal matrices A also allows for the counterexample (iii), which also has some implications for the discussion of superlinear convergence of DIIS. See the proof and below for more remarks.

Theorem 4.3. *(Convergence of DIIS applied to linear problems)*

(i) *Let $\|I - A\| = \xi$. If A is symmetric positive definite, and*

$$\gamma \|x\|^2 \;\leq\; \langle Ax, x \rangle \;\leq\; \Gamma \|x\|^2$$

holds for all $x \in V$, the residuals of DIIS obey the estimate

$$\|r(x_{n+1})\| \;\leq\; \xi\, \frac{2c^n}{1+c^{2n}}\, \|r(x_0)\|, \tag{4.13}$$

in which c is given by $c = (\sqrt{\kappa}-1)/(\sqrt{\kappa}+1) < 1$, $\kappa := \Gamma/\gamma$.

(ii) *If A is diagonalizable with $A = XDX^{-1}$, where D is a diagonal matrix containing the eigenvalues of A, and if the eigenvalues of A are contained in an ellipse with center c, focal distance d and semimajor axis a which excludes the origin, we let $\kappa(X) = \|X\|_2\|X^{-1}\|_2$ be the condition number of X and there holds*

$$\|r(x_{n+1})\| \;\leq\; \xi \cdot \kappa(X)\, \frac{T_n(\frac{a}{d})}{T_n(\frac{c}{d})}\, \|r(x_0)\|, \tag{4.14}$$

with T_n denoting the n-th Chebyshev polynomial and ξ as in (i). In particular, if A is normal, the estimate (4.14) holds with $\kappa(X) = 1$.

(iii) *Suppose we are given a nonincreasing sequence of D positive numbers*

$$r_0 \;\geq\; \ldots \;\geq\; r_{D-1} \;>\; 0$$

and D complex numbers $\lambda_1, \ldots, \lambda_D$. Then there exists a matrix $A \in \mathbb{C}^{D\times D}$ having the eigenvalues $\lambda_1, \ldots, \lambda_D$, a starting value x_0 and a right hand side b such that DIIS, applied to the tuple (A, b, x_0), gives a sequence of iterates $x_0, x_1, \ldots, x_D = x^*$ for which

$$\|r(x_i)\| \;\geq\; r_i \text{ for all } 0 \;\leq\; i \;\leq\; D-1.$$

Proof. Theorem 4.3 follows together with Lemma 4.1, Eq. (4.12) from the respective properties of the GMRES procedure: Under the assumptions made in (i), there holds

$$\|Ax_n - b\| \;\leq\; \frac{2c^n}{1+c^{2n}}\, \|Ax_0 - b\|, \tag{4.15}$$

for the iterates of GMRES, see e.g. [22], Theorem 9.5.6 for the proof; the results also globalize straightforwardly to the operator case. The analogous estimate for the case (ii) where A may only be diagonalizable is for instance proven in [41], Proposition 6.32 and Corollary 6.33. The counterexample (iii) is a reformulation of the central result of [1], where an analogous statement is proven for GMRES.

□

We remark that Theorem 4.3 gives an insight on how DIIS accelerates convergence in the linear case: While the basic iteration scheme

$$x_n \leftarrow x_{n-1} - r_{n-1},$$

for instance a simple (maybe damped) gradient algorithm, may converge slow or not at all, DIIS optimizes the residual over the whole space K_n, and thus inherits the nice convergence behaviour of GMRES. In particular, for the finite history of length 2, this leads to a line search over the space $x_{n-1} + \alpha r(x_{n-1}), \alpha \in \mathbb{R}$, so that in this case, DIIS may turn non-convergent iterations into convergent ones as a consequence of the convergence of the

Richardson iteration with properly chosen α. We note that this behaviour is also sometimes observed when DIIS is applied to nonlinear systems, see e.g. [33] (where DIIS corresponds to the MSB2-method).

4.4. Superlinear convergence, part II: Conclusions from the linear case

When applied to finite dimensional systems, the DIIS method provides the exact solution after at most $n = D$ steps according to Lemma 4.2. The general notion of (super-)linear convergence (as a limit process for $n \to \infty$, see 3(ii)) is therefore not appropriate for examination of the convergence behaviour in this case. An alternative that is also of more practical interest is the examination of how fast the sequence of DIIS residual norms $\|r(x_n)\|$ decays in the course of a moderate number $n \ll D$ of iterations.

The DIIS scheme essentially reproduces the convergence behaviour of the GMRES scheme, for which in many cases some kind of "superlinear" convergence behaviour can be observed in practice in the sense that the ratio $\|r(x_{n+1})\|/\|r(x_n)\|$ of the residuals decays in the course of the iteration [25], and some results on circumstances under which the GMRES algorithm exhibits in some sense superlinear convergence are available: In [46], it is shown that the decay of the residual norms can be related to how well the outer eigenvalues of A are approximated by the Ritz values of A on the trial subspaces K_n; to the authors' knowledge, there is no analysis available though under which circumstances this approximation property is given. Other approaches relate superlinear convergence behaviour to certain properties of the a priori information provided by the data A, b and x_0, see e.g. [3, 4] for corresponding results for the related [22] cg-method.

Neverless, Theorem 4.3 (iii) also shows that such "superlinear convergence behaviour" cannot always be expected for DIIS/GMRES, also cf. e.g. the last numerical example in [46].

5. Convergence analysis for DIIS

In this final section, we will give two convergence results for DIIS applied to nonlinear problems. We saw that DIIS can be reformulated as a projected Broyden method, and we at first will follow the lines of proof that are generally pursued in this context, and therefore as a first step prove that DIIS is locally linearly convergent if the underlying iteration has this property. Linear convergence is then usually used to prove sharper results like superlinear convergence, see e.g. [11]. For the DIIS/projected reverse Broyden scheme, though, the corresponding proofs do not extend straightforwardly, cf. the remarks at the end of Section 3; moreover, Theorem 4.3 (iii) shows that if superlinear convergence can be shown for DIIS at all, there are cases where the superlinear convergence behaviour sets in after $n > D$ steps, while in the context of quantum chemistry, D is usually much larger than the number of maximal iteration steps.

We will therefore show instead in Theorem 5.4 that DIIS combines the favourable properties of Newton's method with those of a GMRES solver applied to solve the actual linearized Newton's equation, where additional errors only arise from the error made in the finite difference approximation of the Jacobian $J(x^*)$.

5.1. Assumptions and linear convergence

Our analysis will be based on the following assumptions. Additionally to those which are standard in the analysis of quasi-Newton methods, we specify a more precise condition for the linear independence of former differences $y_{\ell(n)}, \ldots, y_{n-1}$ than was stated in the DIIS algorithm in Fig. 2.

Assumptions and Notations 5.1. *We assume the function $g : V \to V$ be differentiable in an open convex set $E \subseteq V$, and that $g(x^*) = 0$ holds for some $x^* \in E$. Denoting for $A \in L(V)$ its operator norm by $\|A\|$, we further assume that for some $K \geq 0$,*

$$\|g'(x) - g'(x^*)\| \;\leq\; K\, \|x - x^*\| \tag{5.1}$$

holds for all $x \in E$, and that the Jacobian $J := g'(x^*)$ is nonsingular. We will denote

$$\gamma \;:=\; \|J^{-1}\| \;=\; \|g'(x^*)^{-1}\|. \tag{5.2}$$

We will also assume that

$$\|I - J^{-1}\| < \delta \tag{5.3}$$

is sufficiently small. If this is not the case, we can use the function $\tilde{g}(x) = P^{-1}g(x)$ instead, where P is an approximation of J, and the above condition is then replaced by the condition that g can be preconditioned sufficiently well such that $\|I - J^{-1}P\| < \delta$.

Finally, we will assume that for the former iterates $x_{\ell(n)}, \ldots, x_n$ considered in the step $n \to n+1$, the corresponding differences of function values fulfil

$$\|P_{j\neq i} y_i\| \;\geq\; \frac{\|y_i\|}{\tau} \text{ for all } i = \ell(n), \ldots, n-1 \tag{5.4}$$

for some $\tau > 1$, where $P_{j\neq i}$ denotes the projector on

$$Y_{n,j\neq i} \;=\; \operatorname{span}\{y_j | i = \ell(n), \ldots, n-1, j \neq i\}.$$

□

Note that results analogous to the ones below also hold if the Lipschitz condition (5.1) is replaced by a more general Hölder condition as used e.g. in [11, 36]. Because the functions used in quantum chemistry are usually locally Lipschitz continuous, we refrained from this generalization here.

The first convergence result we prove is that the DIIS method is ($q-$)linearly convergent for sufficiently good starting values. The according result is stated in the next theorem.

Theorem 5.2. *(Linear convergence of DIIS)*

Let $x_0, x_1, \ldots,$ be a sequence of iterates produced by DIIS update scheme from Fig. 2 – or equivalently, computed from (3.4) – , where in each step n, the number of former iterates $y_{\ell(n)}, \ldots, y_n$ used to build the subspace K_n is chosen such that the linear independence condition (5.4) is fulfilled.

Then, the sequence $x_0, x_1, \ldots,$ is locally linearly ($q-$)convergent for any $0 < q < 1/(2\tau)$, i.e. there are constants

$$\delta = \delta(q, \tau, K), \qquad \epsilon = \epsilon(q, \tau, K) > 0$$

such that if

$$\|I - J^{-1}\| \leq \delta, \qquad \|x_0 - x^*\| \leq \epsilon,$$

we have $x_n \in E$ and there holds

$$\|x_{n+1} - x^*\| \quad \leq \quad q \cdot \|x_n - x^*\| \tag{5.5}$$

for all $n \in \mathbb{N}$.

The quite lengthy and technical proof for Theorem 5.2 is essentially in analogy to the analysis from [17] for the "forward" projected Broyden scheme, and therefore not presented here for sake of brevity; we refer the interested reader to [39], where the full proof is being performed. We only note that some additional technical tricks enable us to bound the according error terms without dependence on the dimension D of the space, in contrast to exponential dependence on D in the estimates from [17] which would render such estimates useless in the context of electronic structure calculations.

5.2. A refined convergence estimate for DIIS

Our second convergence result will show that DIIS can be interpreted as a quasi-Newton method, in which the Newton equation (5.6) is solved approximately by a GMRES/DIIS step for the linear system, and in which the Jacobian J (resp. $J(x_n) = g'(x_n)$) is approximated by finite differences, see also the remarks below. We introduce the necessary notation in the next definition.

Definition 5.3. *Let $n \in \mathbb{N}$ be fixed and let us denote by z^* the exact solution of the linear equation*

$$Jz^* \;=\; Jx_n - g(x_n) \;=:\; b_n \tag{5.6}$$

By z_i, $\ell(n) \leq i \leq n+1$, we denote the iterates of a DIIS procedure applied to the linear equation (5.6) with starting value $z_{\ell(n)} := x_{\ell(n)}$. Thus,

$$z_{i+1} = z_i - G_i r(z_i),$$

in what $r(z_i) = Jz_i - b_n$ is the residual associated with the linear equation (5.6), and G_i is the DIIS inverse Jacobian, fulfilling

$$G_i\big(r(z_i) - r(z_{i+1})\big) = z_i - z_{i+1}$$

for all $\ell(n) \le i \le n$, see Theorem (3.2). We define the associated residual reduction factors,

$$d_{i-\ell(n)} \;:=\; \frac{\|r(z_i)\|}{\|r(z_{\ell(n)})\|}.$$

In the case that $r(z_i) = 0$ for some $i = \ell(n), \dots, \le n+1$, we define $z_{i+j} := z_i$ for all $j \in \mathbb{N}$.

□

Note that the factors $d_{i-\ell(n)}$ are bounded by the statements for linear DIIS given in Theorem 4.3. We now formulate the announced second convergence estimate for DIIS under a little more restrictive assumptions.

Theorem 5.4. *(A refined convergence estimate for DIIS)*

Let the assumptions of Theorem 5.2 hold, so that DIIS is linearly convergent with convergence factor q. Then there are $\delta = \delta(q), \epsilon = \epsilon(q) > 0$ such that if

$$\|I - J^{-1}\| \le \delta, \qquad \|x_0 - x^*\| \le \epsilon$$

and if

$$\ell(j) = \ell(n) \quad \textit{for all} \quad \ell(n) \le j \le n,$$

the "residual error" $\|g(x_{n+1})\|$ can be estimated by

$$\|g(x_{n+1})\| \le \;\; c_1\|g(x_n)\|^2 \;+\; c_2 \cdot \; d_{n-\ell(n)} \; \|g(x_{\ell(n)})\| \;+\; c_3\|g(x_{\ell(n)})\|^2, \tag{5.7}$$

for all $n \in \mathbb{N}$, with $d_{n-\ell(n)}$ is the convergence factor in the $(n-\ell(n))$-th step of the DIIS solution of the linear auxiliary problem defined in 5.3, and where c_1, c_2, c_3 depend on δ, K, and c_2, c_3 additionally on the initial error $\|x^* - x_{\ell(n)})\|$ and q.

Again, we skip the proof here for reasons of brevity, referring the reader to [39] and only giving some general notes. The estimate (5.7) bases on the decomposition

$$\|g(x_{n+1})\| \;\le\; \underbrace{\|g(z^*)\|}_{\text{(I)}} \;+\; \underbrace{\|g(z_{n+1}) - g(z^*)\|}_{\text{(II)}} \;+\; \underbrace{\|g(x_{n+1}) - g(z_{n+1})\|}_{\text{(III)}} \tag{5.8}$$

of the error, with the quantities z_{n+1}, z^* from Definition 5.3. A (quite lengthy) estimation of the single terms gives the three error components of the estimate (5.7). Those three error components have straight-forward interpretations:

- The first term (I) represents the modeling (linearization) error of (the exact) Newton's method, where the correction equation (5.6)[4] is solved exactly, leading to the well-known quadratic error term.

[4]Or alternatively, where the "real" Newton equation $J(x_n)(x_{n+1} - x_n) = F(x_n)$ is solved. Eq. (5.6) was chosen here for convenience, but it is not hard to see that replacing J by $J(x_n)$ only adds anther quadratic error term.

- The second term (II) represents the error made in solving (5.6) approximately by a GMRES/DIIS step on the actual subspace $x_n + K_n$, thus incorporating the convergence rate of the DIIS/GMRES from Theorem 4.3.
- The third error term (III), that can grow large if many older iterates are regarded, is a worst-case estimate for the error made in the finite difference approximation of J^* resp. $J(x_n)$.

We conjecture that the third error term is bounded by $\|f(x_{\ell(n)})\| \cdot \|f(x_n)\|$, so that the result given here is presumably not optimal, but we were not able to show this so far. We also note that the restrictive assumption that $\ell(j) = \ell(n)$ for all $\ell(n) \leq j \leq n$ (meaning that in the DIIS procedure, $K_{\ell(n)} = \emptyset$, and that the used Krylov spaces K_j are constantly increased without discarding iterates; in particular, (5.4) has to be fulfilled in each step) could not be abolished without the error term $\|g(x_{\ell(n)})\|$ in the third term in (5.7) having to be replaced by the less favourable term $\|g(x_{\ell(\ell(n))})\|$.
Note that the second and third error term in (5.7) are opposing perturbations of the quadratic convergence given by the first term: The error term associated with the DIIS procedure for the linear problem (5.6) is reduced with an increasing number of former iterates, according to the well-known theory for the associated GMRES procedure, and thus gives better bounds the longer the history is chosen if the convergence of the GMRES procedure is favourable, e.g. superlinear. On the contrary, the error bound for the finite difference approximation gets worse the more former iterates are taken up in the procedure. In order to obtain the best bounds for convergence rates for the DIIS procedure, the two error terms thus have to be balanced out, and in agreement with this, practical experience with GMRES seems to indicate that the number of iterates has to be kept moderate in order to keep the procedure efficient, especially if the iterates become "almost linearly dependent", i.e. if the constant τ gets large, see [26, 38]. Estimate (5.7) shows that such an inefficiency can solely be due to the effects of nonlinearity, contained in the third error term, so that in principle, if g is "rather nonlinear" in the sense that the constant K in (5.1) is large, it is advisable to discard old iterates more often.
For linear problems, the first and last error terms in (5.7) are zero. By a continuity argument, we can heuristically conclude that if in contrast to the situation discussed in (iii), the nonlinearity, i.e. the constant K in (5.1), is small, the convergence of the DIIS is mainly governed by that of the DIIS/GMRES procedure for the associated linear problem. We note that in the context of electronic structure calculations, similar assumptions entered into our convergence analysis for DFT [43] and CC [40], and they seem to be in good agreement with practice.
In particular, if the Jacobian is symmetric, for instance if (1.1) is the first order condition of a minimization problem as in DFT, the worst-case convergence behaviour of the DIIS procedure is mainly determined by the spectral properties of J, while for nonsymmetric Jacobians, properties of the right

hand side etc. play a role, cf. Section 4. Thus, "superlinear convergence" of the algorithm can be expected if the DIIS/GMRES procedure for the underlying linear problem has this property already for a small number of steps, so that the third error term provoked by the nonlinearity of g and the associated finite difference approximation of J can be kept sufficiently small by discarding old iterates.

6. Concluding remarks

We have identified DIIS with the reverse, projected Broyden's method given by formula (3.4). By this connection between DIIS and the family of Broyden-like methods and Krylov space methods, the development of new problem-adapted variants of DIIS and related convergence accelerators may therefore well profit from the theoretical as well as from the practical experience made with Broyden-like methods and Newton-Krylov type methods. Comparison of practical results for DIIS (corresponding to the reverse, projected Broyden's method, MSB2 [33]) to other Broyden's methods (e.g. the projected forward method KAIN [23]) give an ambivalent picture. In some cases DIIS seems to be inferior to the according "forward" method [17, 23]. On the other hand, DIIS provides a great amount of examples where it is applied to quantum chemical problems with great success, also in direct comparison with Broyden type methods [33], and where it may be preferred due to its simple implementation in the form of DIIS. From this perspective, hybrid schemes combining forward and backward approaches would be a worthwhile a closer investigation, in particular because the according systems to be solved are quite small. The connection between the variants of Broyden's method popular and well-analysed in the context of numerical analysis should now be exploited further to improve the performance of DIIS.

For DIIS applied to linear equations, we found that the convergence is fixed by the mostly favourable convergence behaviour of the according GMRES procedure. DIIS, applied to nonlinear equations, can therefore be viewed as a globalization of GMRES to nonlinear equations, and we have shown that the convergence behaviour is still related to the properties of a linear equation for the Jacobian J if nonlinear effects are small. This is in agreement with similar results for Broyden's method [21]. Our results do not fully explain the extremely good convergence behaviour observed in the convergence regime of the SCF procedure though, and we therefore expect that our results can be sharpened if one considers the particular case of the SCF procedure. If nonlinearities are mild, our results indicate that DIIS still shares the favourable properties of the according GMRES procedure. A further systematic investigation of the influence of increasing nonlinearities on the performance of the DIIS solver would therefore be desirable in the future.

References

[1] M. Arioli, Vlastimil Pták, Zdenek Strakoš, *Krylov sequences of maximal length and convergence of GMRES*, BIT Numerical Mathematics 38, 4, p. 636, 1998.

[2] ABINIT is a common project of the Université Catholique de Louvain, Corning Incorporated, and other contributors, for further details see http://www.abinit.org

[3] B. Beckermann, A. B. J. Kuijlaars, *Superlinear convergence of conjugate gradients*, SIAM J. Numer. Anal. 39, p. 300, 2001.

[4] B. Beckermann, A. B. J. Kuijlaars, *Superlinear CG convergence for special right-hand sides*, Electron. Trans. Numer. Anal. 14, p. 1, 2002.

[5] bigDFT, http://www-drfmc.cea.fr/sp2m/L_Sim/BigDFT/index.en.html

[6] C. G. Broyden, *A class of methods for solving nonlinear simultaneous equations*, Mathematics of Computation 19, p. 577, 1965.

[7] C. G. Broyden, J. E. Dennis, J. J. Moré, *On the local and superlinear convergence of Quasi-Newton methods*, J. Inst. Math. Appl. 12, p. 223, 1973.

[8] E. J. Bylaska et al., NWChem, A Computational Chemistry Package for Parallel Computers, Version 5.1 , Pacific Northwest National Laboratory, Richland, Washington 99352-0999, USA. A modified version, 2007.

[9] E. Cancès, C. Le Bris, *On the convergence of SCF algorithms for the Hartree-Fock equations*, M2AN 34, p. 749, 2000.

[10] P. Császár, P. Pulay, *Geometry optimization by direct inversion in the iterative subspace*, J. of Molecular Structure 114, p. 31, 1984.

[11] J. E. Dennis Jr., R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, SIAM, Philadelphia, 1996.

[12] P. Deuflhard, R. Freund, A. Waltera, *Fast secant methods for the iterative solution of large nonsymmetric linear systems*, Impact of Computing in Science and Engineering 2, 3, p. 2446, 1990.

[13] F. Eckert, P. Pulay, and H.-J. Werner, *Ab Initio Geometry Optimization for Large Molecules*, J. Comp. Chem. 18, p. 1473, 1997.

[14] S. C. Eisenstat, H. C. Elman, M. H. Schultz, *Variational iterative methods for nonsymmetric systems of linear equations*, SIAM J. Numer. Anal. 20, p. 345, 1983.

[15] T. H. Fischer, J. Almlöf, *General methods for geometry and wave function optimization*, J. Phys. Chem. 92, p. 9768, 1992.

[16] D. M. Gay, *Some convergence properties of Broyden's method*, SIAM J. Numer. Anal. 16, p. 623, 1979.

[17] D. M. Gay, R. B. Schnabel, *Solving systems of nonlinear equations by Broyden's method with projected updates*, Nonlinear Programming 3, Academic Press, 1978.

[18] L. Genovese, A. Neelov, S. Goedecker, T. Deutsch, S. A. Ghasemi, A. Willand, D. Caliste, O. Zilberberg, M. Rayson, A. Bergman, R. Schneider, *Daubechies wavelets as a basis set for density functional pseudopotential calculations*, J. Chem. Phys., 129, 014109, 2009.

[19] X. Gonze et al., Comput. Mater. Sci. 25 , p. 478, 2002.

[20] X. Gonze et al., Zeit. Kristallogr. 220, p. 558, 2005.

[21] A. Griewank, *The local convergence of Broyden-like Methods on Lipschitzian problems in Hilbert spaces*, SIAM Num. Anal. 24, 3, p. 684, 1987.

[22] W. Hackbusch, *Iterative solution of large sparse systems of equations*, Springer, 1994.

[23] R. J. Harrison, *Krylov Subspace Accelerated Inexact Newton Method for Linear and Nonlinear Equations*, J. Comp. Chem. 25, p. 328, 2003.

[24] T. Helgaker, P. Jørgensen, J. Olsen, *Molecular Electronic-Structure Theory*, John Wiley & Sons, 2000.

[25] Y. Huang, H. van der Vorst, *Some Observations on the Convergence Behavior of GMRES*, Tech. Rep. 89-09, Faculty of Technical Mathematics and Informatics, Delft University of Technology, The Netherlands, 1989.

[26] M. Kawata, C. M. Kortis, 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, p. 4426, 1998.

[27] R. A. Kendall et al., Comput. Phys. Commun. **128** (2000), 260.

[28] H. M. Klie, *Krylov-secant methods for solving large-scale systems of coupled nonlinear parabolic equations*, PhD thesis, Rice University Houston, TX, USA, 1997.

[29] G. Kresse, J. Furthmüller, *Efficient iterative schemes for* ab initio *total-energy calculations using a plane-wave basis set*, Phys. Rev. B 54, 16, p. 11169, 1996.

[30] K. N. Kudin, G. E. Scuseria, *Converging self-consistent field equations in quantum chemistry - recent achievements and remaining challenges, ESAIM: Mathematical Modelling and Numerical Analysis*, 41, 2, p. 281, 2007.

[31] K. N. Kudin, G. E. Scuseria, E. Cancès, *A black-box self-consistent field iteration convergence algorithm: One step closer*, J. Chem Phys. 116, p. 8255, 2002.

[32] J. Liesen, P. Tichy, *Convergence analysis of Krylov subspace methods*, GAMM-Mitteilungen 27, 2, p. 153, 2004.

[33] L. D. Marks, D. R. Luke, *Robust mixing for ab initio quantum mechanical calculations*, Phys. Rev. B 78, **8**, 075114, 2008.

[34] J. M. Martinez, T. L. Lopez, *Combination of the sequential secant method and Broyden's method with projected updates*, Computing 25, p. 379, 1980.

[35] J. M. Millam, G. E. Scuseria, *Linear scaling conjugate gradient density matrix search as an alternative to diagonalization for first principles electronic structure calculations*, J. Chem. Phys. 106, p. 5569, 1997.

[36] J. M. Ortega, W. C. Rheinboldt, *Iterative Solution of nonlinear equations in several variables*, Acad. Press, 1970.

[37] P. Pulay, *Convergence acceleration of iterative sequences. The case of SCF iteration*, Chem. Phys. Letters 73, 2, p. 393, 1980.

[38] P. Pulay, *Improved SCF Convergence Acceleration*, Journ. Comp. Chem. 3, 4, p. 556, 1982.

[39] T. Rohwedder, *An analysis for some methods and algorithms of quantum chemistry*, PhD thesis, TU berlin, 2010, available at *http://opus.kobv.de/tuberlin/volltexte/2010/2852*.

[40] T. Rohwedder, R. Schneider *The continuous Coupled Cluster method*, in preparation.

[41] Y. Saad, *Iterative Methods for Sparse Linear Systems*, 2nd edition, SIAM, 2003.

[42] Y. Saad, M. H. Schultz, *GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput. 7 , p. 856, 1986.

[43] R. Schneider, T. Rohwedder, J. Blauert, A. Neelov, *Direct minimization for calculating invariant subspaces in density functional computations of the electronic structure*, Journal of Comp. Math., 27, p. 360, 2009.

[44] H. H. B. Sørensen, O. Østerby, *On One-Point Iterations and DIIS*, Numerical Analysis and Applied Mathematics: International Conference on Numerical Analysis and Applied Mathematics 2009: Volume 1 and Volume 2. AIP Conference Proceedings, Volume 1168, p. 468, 2009.

[45] L. Thogersen, J. Olsen, A. Köhn, P. Jorgensen, P. Salek, T. Helgaker, *The trust-region self-consistent field iteration method in Kohn-Sham density functional theory*, J. Chem. Phys. 123, 074103, 2005.

[46] H. A. van der Vorst, C. Vuik, *The superlinear convergence behaviour of GMRES*, Journ. Comp. Appl. Math. 48, 3, p. 327, 1993.

[47] M. Zółkowski, V. Weijo, P. Jørgensen, J. Olsen, *An efficient algorithm for solving nonlinear equations with minimal number of trial vectors: Applications to atomic-orbital based coupled-cluster theory*, J. Chem. Phys. 128, 204105, 2008.

Thorsten Rohwedder
Sekretariat MA 5-3
Institut für Mathematik
TU Berlin
Straße des 17. Juni 136
10623 Berlin, Germany
e-mail: rohwedde@math.tu-berlin.de
URL: http://www.math.tu-berlin.de/∼ rohwedde/

Reinhold Schneider
Sekretariat MA 5-3
Institut für Mathematik
TU Berlin
Straße des 17. Juni 136
10623 Berlin, Germany
e-mail: schneidr@math.tu-berlin.de
URL: http://www.math.tu-berlin.de/∼ schneidr/