Time-lag Derivative Convergence for Fixed Point Iterations

Andreas Griewank, Daniel Kressner

Time-lag in Derivative Convergence for Fixed Point Iterations

Le retard en convergence des dérivées pour les calculs itératifs avec point fixe

Andreas Griewank* — Daniel Kressner**

* Institut für Mathematik
Humboldt-Universität zu Berlin
Berlin, Germany
griewank@mathematik.hu-berlin.de

** Department of Mathematics
University of Zagreb
Zagreb, Croatia
kressner@math.hr

RÉSUMÉ. Une étude antérieure a prouvé et vérifié expérimentalement sur un code Euler 2D que les calculs itératifs avec point fixe peuvent être différentiés pour obtenir les dérivées aux premier et deuxième ordres des fonctions implicites définies par des équations d'état. On considérait également que les itérées correspondantes des gradients et Hessiens réduits convergent à la même vitesse que l'itération de point fixe d'origine.
Cette étude plus détaillée révèle néanmoins que ces dérivées convergent avec un certain retard par rapport aux valeurs de la fonction. En effet le rapport des erreurs correspondantes croît vers l'infini proportionnellement au compteur d'itérations ou à son carré. Mathématiquement, cet effet plutôt subtil est causé par l'apparition de blocs de Jordan correspondant à des valeurs propres dégénérées. Nous construisons un modèle théorique de cet effet et nous le validons par des expèriences numériques.

ABSTRACT. In an earlier study it was proven and experimentally confirmed on a 2D Euler code that fixed point iterations can be differentiated to yield first and second order derivatives of implicit functions that are defined by state equations. It was also asserted that the resulting approximations for reduced gradients and Hessians converge with the same R-factor as the underlying fixed point iteration.
A closer look reveals now that nevertheless these derivative values lag behind the functions values in that the ratios of the corresponding errors grow proportional to the iteration counter or its square towards infinity. This rather subtle effect is caused mathematically by the occurrence of nontrivial Jordan blocks associated with degenerate eigenvalues. We elaborate the theory and report its confirmation through numerical experiments.

MOTS-CLÉS : methode iterative de type point fixe, dérivatif, convergence, bloc de Jordan

KEYWORDS : fixed point iteration, derivative, convergence, Jordan block

Novembre 2005, Numéro spécial CARI'04

Revue ARIMA **pages 87 à 102**

1. Introduction and Assumption

The effect to be analyzed arises in the context of design optimization by what has been called piggy-back optimization [5]. Design optimization problems are distinguished from general nonlinear programming problems (NLP) by the fact that the vector of variables x is a priori partitioned into a state vector $y \in Y$ and a set of design variables $u \in U$. For application of this scenario in computational fluid dynamics see for example [10], [8], [9], and [7]. Throughout we assume that the "user" has provided an iteration function

$$G : Y \times U \to Y$$

that is contractive with respect to an inner product norm on Y so that for all $u \in U$ and $y, \tilde{y} \in Y$

$$\|G(y,u) - G(\tilde{y},u)\| \leq \varrho \, \|y - \tilde{y}\|.$$

Here $\varrho < 1$ may vary continuously as a function of the design u and its exact size will usually not be available to a practical algorithm.

As an immediate consequence it follows by the Banach fixed point theorem that for fixed u and any initial $y_0 \in Y$ the sequence $\{y_k\}$ generated by

$$y_{k+1} = G(y_k, u)$$

must converge to the unique fixed point $y_* = y_*(u)$ with $y_* = G(y_*, u)$. In other words, the assumptions made so far ensure that one can obtain for any u a solution $y_*(u)$, a process which one may call "simulate" the underlying system. In a practical simulation the variables u and y will often be restricted to open subsets of the spaces U and Y, respectively.

In order to progress from simulation to design we require more smoothness of G, namely, that it is at least once continuously differentiable in the joint variable vector (y, u). The same assumption will be made for the objective function

$$f : Y \times U \to \mathbb{R},$$

which is meant to be minimized. Provided at least $f \in C^1(Y,U)$, one can obtain in a completely automated fashion the adjoint iteration function

$$\bar{G}(y, \bar{y}, u) \equiv \bar{y} \, G_y(y,u) + f_y(y,u). \tag{1}$$

Here subscripts denote partial differentiation and $\bar{y}$ like the gradient f_y is considered a row-vector belonging to the dual space of Y, which we identify with the Hilbert space Y itself. Then we have in the induced matrix and operator norm

$$\varrho(u) = \max_{y \in Y} \|G_y(y,u)\| \leq \varrho < 1$$

so that also in the dual norm

$$\|\bar{G}(y,\bar{y},u) - \bar{G}(y,\tilde{y},y,u)\| \le \varrho\|\bar{y}-\tilde{y}\|$$

for any two row-vectors $\bar{y},\tilde{y} \in \bar{Y} \equiv Y$.

2. Piggy-Back Convergence of Adjoints

Throughout the remainder of this paper we consider u as constant and may therefore omit it occasionally as an argument in analyzing the simultaneous iteration

$$\begin{bmatrix} y_{k+1} \\ \bar{y}_{k+1} \end{bmatrix} = \begin{bmatrix} G(y_k,u) \\ \bar{G}(y_k,\bar{y}_k,u) \end{bmatrix} \tag{2}$$

Even if G is merely C^1 and thus $G_y(y) = G_y(y,u)$ continuous with respect to y it follows from $y_k \to y_*$ that $G_y(y_k,u) \to G_y(y_*,u)$ and hence the adjoint iterates $\bar{y}_k$ converge to $\bar{y}_*$ the unique solution of the adjoint equation

$$\bar{y}_* = \bar{y}_* G_y(y_*,u) + f_y(y_*,u) \tag{3}$$

The vector $\bar{y}_*$ can be used to compute the so called reduced gradient

$$\bar{u}_* = \bar{y}_* G_u(y_*,u) + f_u(y_*,u) \tag{4}$$

This row vector represents the total derivatives of f with respect to u, after the elimination of the state vector y using the implicit function theorem. In order to be more specific about the rate of convergence we assume that G_y and f_y are Lipschitz continuous with respect to y so that for some $\nu > 0$

$$\|G_y(\tilde{y},u) - G_y(y,u)\| \le \nu\|\tilde{y}-y\| \ge \|f_y(\tilde{y},u) - f_y(y,u)\|.$$

Then we obtain for the discrepancies $\Delta y_k \equiv y_k - y_*$ and $\Delta\bar{y}_k \equiv \bar{y}_k - \bar{y}_*$ the following result.

Lemma 2.1 *The sequences y_k and $\bar{y}_k$ converge R-linearly in that*

$$\limsup_{k\to\infty} \sqrt[k]{\|\Delta\bar{y}_k\|} \;\le\; \varrho \;\ge\; \limsup_{k\to\infty} \sqrt[k]{\|\Delta y_k\|}.$$

Proof. Using the assumed Lipschitz continuity we obtain the estimate

$$\begin{aligned} \|\Delta\bar{y}_{k+1}\| &\le \|\bar{y}_k G_y(y_k,u) - \bar{y}_* G_y(y_*,u)\| + \|f_y(y_k,u) - f_y(y_*,u)\| \\ &= \|\Delta\bar{y}_k\, G_y(y_k,u) + \bar{y}_*(G_y(y_k,u) - G_y(y_*,u))\| + \nu\|y_k - y_0\| \\ &\le \varrho\|\Delta\bar{y}_k\| + \|\Delta y_k\|(\|\bar{y}_*\|+1)\nu. \end{aligned}$$

Consequently we have for any weighted error combination

$$\varepsilon_k \equiv \|\Delta y_k\| + \omega \|\Delta \bar{y}_k\|$$

the recurrence

$$\begin{aligned} \varepsilon_{k+1} &\leq \varrho \|\Delta y_k\| + \omega(\varrho \|\Delta \bar{y}_k\| + \nu(\|\bar{y}_*\| + 1)\|\Delta y_k\|) \\ &= (\varrho + \omega\nu(\|\bar{y}_*\| + 1))\|\Delta y_k\| + \omega\varrho\|\Delta \bar{y}_k\| \\ &\leq (\varrho + \omega\nu(\|\bar{y}_*\| + 1))\varepsilon_k. \end{aligned}$$

This implies for any $\omega < (1 - \varrho)/(\nu(\|\bar{y}_*\| + 1))$ the Q-linear convergence result

$$\limsup_{k\to\infty} \varepsilon_{k+1}/\varepsilon_k \leq \varrho + \omega\nu(\|\bar{y}_*\| + 1) < 1.$$

By standard arguments one derives the R-linear convergence results

$$\limsup_{k\to\infty} \sqrt[k]{\|\Delta \bar{y}_k\|} \leq \lim_{k\to\infty} \sqrt[k]{\varepsilon_k/\omega} \leq \varrho + \omega\nu(\|\bar{y}_*\| + 1) < 1.$$

Taking the infimum over all $\omega > 0$ one finally obtains as in [4] the assertion. The inequality on the right was just added for comparison. □

Since the convergence speed cannot be improved under our assumptions (namely G_y has maximal norm ϱ and is Lipschitz continuous with respect to y) one may arrive at the conclusion that the sequences $\{y_k\}$ and $\{\bar{y}_k\}$ converge essentially at the same speed. In fact this claim has been made repeatedly in the literature and the first author has suffered from the same impression for a long time. On the other hand there has been the persistent notion that the convergence of derivatives is lagging behind those of the underlying fixed point iterates.

3. Relative Convergence Speed of First Adjoints

In the remainder of this paper we require that $Y \equiv \mathbb{R}^n$ and $U \equiv \mathbb{R}^m$ are finite-dimensional Euclidean spaces so that all linear operators can be identified with their matrix presentation. Assuming furthermore, that G and f are twice Lipschitz-continuously differentiable, we may rewrite the recurrence (2) as

$$\begin{bmatrix} y_{k+1} \\ \bar{y}_{k+1} \end{bmatrix} = \begin{bmatrix} G(y_k, u) \\ N_y(y_k, \bar{y}_k, u) \end{bmatrix} \tag{5}$$

Here we have expressed the $\bar{G}$ from (3) as the gradient of the function

$$N(y, \bar{y}, u) \equiv \bar{y}\, G(y, u) + f(y, u)$$

with respect to y. Notice that this function N differs from the familiar Lagrange function L of the optimization problem $\min f(y,u)$ s.t. $G(y,u)-y=0$ by the shift

$$\bar{y}\ y = N(y,\bar{y},u) - L(y,\bar{y},u).$$

Consequently, we have

$$N_y = L_y + \bar{y} \quad \text{and} \quad N_{\bar{y}} = L_{\bar{y}} + y \quad \text{but} \quad N_u = L_u$$

and, for the subsequent analysis more importantly, all second derivatives are identical :

$$N_{yy} = L_{yy},\ N_{yu} = L_{yu},\ N_{uu} = L_{uu}.$$

Differentiating (5) we obtain the block-triangular Jacobian

$$J_k \equiv \frac{\partial(y_{k+1},\bar{y}_{k+1})}{\partial(y_k,\bar{y_k})} = \begin{bmatrix} G_y(y_k,u) & 0 \\ N_{yy}(y_k,\bar{y}_k,u) & G_y^T(y_k,u) \end{bmatrix}.$$

The characteristic polynomial of J_k satisfies

$$\det(J_k - \lambda I) = \det{}^2(G_y(y_k,u) - \lambda I),$$

which implies that J_k has the same eigenvalues as $G_y(y_k,u)$ but each of them with double algebraic multiplicity. Our analyis and in particular the proof of Lemma A.1 reveals that all eigenvalues of J_k are generically defective and generate a Jordan block of dimension two. Another consequence of Lemma A.1 is that one can deduce a linear-geometric decline in the adjoint error as follows.

Linearizing about the fixed point $(y_*,\bar{y}_*)$ we obtain the Taylor expansion

$$\begin{bmatrix} \Delta y_{k+1} \\ \Delta\bar{y}_{k+1} \end{bmatrix} = \begin{bmatrix} A & 0 \\ B & A^T \end{bmatrix} \begin{bmatrix} \Delta y_k \\ \Delta\bar{y}_k \end{bmatrix} + \mathcal{O}(\|\Delta y_k\|^2 + \|\Delta\bar{y}_k\|^2)$$

where $A \equiv G_y(y_*,u)$ and $B \equiv N_{yy}(\bar{y}_*,y_*,u)$. From this it follows by induction using the R-linear convergence of $\|\Delta y_k\| + \|\Delta\bar{y}_k\|$ that for any k and $j>0$

$$\begin{bmatrix} \Delta y_{k+j} \\ \Delta\bar{y}_{k+j} \end{bmatrix} = \begin{bmatrix} A & 0 \\ B & A^T \end{bmatrix}^j \begin{bmatrix} \Delta y_k \\ \Delta\bar{y}_k \end{bmatrix} + O(\|\Delta y_k\|^2 + \|\Delta\bar{y}_k\|^2). \tag{6}$$

Similarly it can be easily verified by induction that

$$J_*^j \quad \equiv \quad \begin{bmatrix} A & 0 \\ B & A^T \end{bmatrix}^j \quad = \quad \begin{bmatrix} A^j & 0 \\ \sum_{i=1}^{j} (A^T)^{i-1} B A^{j-i} & (A^T)^j \end{bmatrix}. \tag{7}$$

To simplify the matrix on the bottom left we assume at first that $A = G_y$ has n distinct real eigenvalues. Then it is certainly diagonalizable so that

$$A = T\Gamma T^{-1} \quad \text{with} \quad \Gamma = \mathrm{diag}(\gamma_\mathrm{j})_{\mathrm{j}=1}^{\mathrm{n}},$$

where

$$\varrho_* \equiv \max_{1\le j\le n} |\gamma_j| \quad \le \varrho < 1$$

Then we can perform a two stage reduction to obtain the Jordan-like representation

$$\begin{aligned}\begin{bmatrix} A & 0 \\ B & A^T \end{bmatrix} &= \begin{bmatrix} T & 0 \\ 0 & T^{-T} \end{bmatrix} \begin{bmatrix} \Gamma & 0 \\ T^T B T & \Gamma \end{bmatrix} \begin{bmatrix} T^{-1} & 0 \\ 0 & T^T \end{bmatrix} \\ &= \begin{bmatrix} T & 0 \\ 0 & T^{-T} \end{bmatrix} \begin{bmatrix} I & 0 \\ C^T & I \end{bmatrix} \begin{bmatrix} \Gamma & 0 \\ D & \Gamma \end{bmatrix} \begin{bmatrix} I & 0 \\ C & I \end{bmatrix} \begin{bmatrix} T^{-1} & 0 \\ 0 & T^T \end{bmatrix}\end{aligned} \tag{8}$$

Here D is the (real) diagonal of $T^T B T$ and $C = -C^T$ is the antisymmetric solution of the Liapunov equation

$$\Gamma C - C\Gamma = T^T B T - D.$$

It is well known that the linear mapping from C to $\Gamma C - C\Gamma$ has the n^2 eigenvalues $\gamma_i - \gamma_j$ and the eigenvectors $e_i e_j^T$ for $1 \le i, j \le n$, so that the Liapunov equation must be solvable since all eigenvalues of A are by assumption distinct.

Then it follows immediately that the j-th power of J_* is given by

$$\begin{bmatrix} A & 0 \\ B & A^T \end{bmatrix}^j = \begin{bmatrix} T & 0 \\ (CT^{-1})^T & T^{-T} \end{bmatrix} \begin{bmatrix} \Gamma^j & 0 \\ j\, D\Gamma^{j-1} & \Gamma^j \end{bmatrix} \begin{bmatrix} T^{-1} & 0 \\ CT^{-1} & T^T \end{bmatrix} \tag{9}$$

Thus we see that unless the diagonal D of B vanishes there might be a pretty strong growth in the adjoint error component $\Delta\bar{y}_k$. When the second order sufficiency conditions for local optimality are satisfied at the limiting fixed point at least some projection of B must be positive definite so that itself and its diagonal cannot vanish. In the following we draw on the analysis in the appendix, which imposes much weaker assumptions on $A = G_y$.

Lemma A.1 shows that unless $x^H B x$ happens to vanish for some possibly complex eigenvector x of A there might be a pretty strong growth in the adjoint error component $\Delta\bar{y}_k$. To compare it to the original error Δy_k itself we firstly have to analyze its recurrence a bit more carefully. Using the Lipschitz constant ν one finds by standard estimates

$$\|\Delta y_{k+1} - A\Delta y_k\| \le \nu \|\Delta y_k\|^2.$$

Let $\mathcal{X}$ be the (right) invariant subspace of A belonging to all eigenvalues of maximal modulus ρ. Then, using the estimate above, one can show that the angle between Δy_k and

$\mathcal{X}$ satisfies a recurrence that has exactly one stable fixed point namely 0, see also [3]. If the columns of Y form a basis for the left invariant subspace belonging to all eigenvalues of maximal modulus, the relation $Y^T x = 0$ for all $x \in \mathcal{X}^\perp$ thus generically implies

$$\lim_{k\to\infty} \frac{\|Y^T \Delta y_k\|}{\|\Delta y_k\|} = 1. \tag{10}$$

Under the assumptions of Lemma A.1, which are generically satisfied, it then follows by (16) that

$$\underline{C} \quad \leq \quad \frac{1}{k}\,\frac{\|\Delta \bar{y}_k\|}{\|\Delta y_k\|} \quad \leq \quad \overline{C} \tag{11}$$

for some constants $\underline{C}, \overline{C} > 0$. Approximately, we have

$$\frac{\|\Delta \bar{y}_k\|}{\|\Delta y_k\|} \quad \sim \quad k.$$

Hence we see that the convergence of the adjoint vectors $\bar{y}_k$ really lags behind that of the underlying iterates y_k even though both sequences have the same R-factor ρ.

4. Convergence of Second Order Adjoints

The above analysis can be extended to second derivatives representing products of the projected Hessian with certain direction vectors. More specifically, after picking a direction $\dot{u} \in U$ we may append (2) by the iterations

$$\dot{y}_{k+1} \equiv \dot{G}(y_k, \dot{y}_k, u, \dot{u}) \equiv G_y(y_k, u)\dot{y}_k + G_u(y_k, u)\dot{u} \tag{12}$$

and

$$\begin{aligned} \dot{\bar{y}}_{k+1} \equiv \dot{\bar{G}}(y_k, \bar{y}_k, \dot{y}_k, \dot{\bar{y}}_k, u, \dot{u}) \quad &\equiv \quad \dot{\bar{y}}_k G_y + \bar{y}_k G_{yy}\dot{y}_k + f_{yy}\dot{y}_k + \bar{y}_k G_{yu}\dot{u} + f_{yu}\dot{u} \\ &= \quad \dot{\bar{y}}_k G_y(y_k, u) + N_{yy}(y_k, \bar{y}_k, u)\dot{y}_k + N_{yu}(y_k, \bar{y}_k, u)\dot{u} \end{aligned} \tag{13}$$

where all derivatives of G and f are evaluated at the current argument (y_k, u). Then an analysis along the lines of Section 3 shows that the $\dot{y}_k$ and $\dot{\bar{y}}_k$ also converge R-linearly to respective fixed points $\dot{y}_*$ and $\dot{\bar{y}}_*$ solving

$$\dot{y}_* = \dot{G}(y_*, \dot{y}_*, u, \dot{u}) \quad \text{and} \quad \dot{\bar{y}}_* \equiv \dot{\bar{G}}(y_*, \bar{y}_*, \dot{y}_*, \dot{\bar{y}}_*, u, \dot{u}).$$

The vector $\dot{y}_*$ represents the feasible direction in state space associated with the variation $\dot{u}$ in the design space. The vector $\dot{\bar{y}}_*$ can be used to compute

$$\dot{\bar{u}}_* \;\equiv\; \dot{\bar{y}}_* G_u(y_*, u) + N_{uy}(y_*, \bar{y}_*, u)\dot{y}_* + N_{uu}(y_*, \bar{y}_*, u)\dot{u} \tag{14}$$

which represents the product of the reduced Hessian with the direction $\dot{u}$. To analyze the speed of convergence more carefully let us consider the extended Jacobian

$$\frac{\partial(y_{k+1}, \bar{y}_{k+1}, \dot{y}_{k+1}, \dot{\bar{y}}_{k+1})}{\partial(y_k, \bar{y}_k, \dot{y}_k, \dot{\bar{y}}_k)} =$$

$$= \begin{bmatrix} G_y(y_k,u) & 0 & 0 & 0 \\ N_{yy}(y_k,\bar{y}_k,u) & G_y^T(y_k,u) & 0 & 0 \\ P(y_k,\dot{y}_k,u,\dot{u}) & 0 & G_y(y_k,u) & 0 \\ H(y_k,\bar{y}_k,\dot{y}_k,\dot{\bar{y}}_k,u,\dot{u}) & P(y_k,\dot{y}_k,u,\dot{u})^T & N_{yy}^T(y_k,\bar{y}_k,u) & G_y^T(y_k,u) \end{bmatrix}$$

where

$$\begin{aligned} P(y,\dot{y},u,\dot{u}) &\equiv G_{yy}(y,u)\dot{y} + G_{yu}(y,u)\dot{u} \\ H(y,\bar{y},\dot{y},\dot{\bar{y}},u,\dot{u}) &\equiv \dot{\bar{y}}G_{yy}(y,u) + N_{yyy}(y,\bar{y},u)\dot{y} + N_{yyu}(y,\bar{y},u)\dot{u}. \end{aligned}$$

We notice that the matrix H is symmetric, while P is general and the values of these two square matrices at the fixed point $(y_*, \bar{y}_*, \dot{y}_*, \dot{\bar{y}}_*)$ are independent of each other as $A \equiv G_y(y_*, u)$ and $B \equiv N_{yy}(\bar{y}_*, y_*, u)$.

We are looking now for estimates of the corresponding discrepancies $\Delta\dot{y}_k = \dot{y}_k - \dot{y}_*$ and $\Delta\dot{\bar{y}}_k \equiv \dot{\bar{y}}_k - \dot{\bar{y}}_*$ in addition to the Δy_k and $\Delta\bar{y}_k$ considered before. Similarly to (6) we obtain the linearization

$$\begin{bmatrix} \Delta y_{k+j} \\ \Delta\bar{y}_{k+j} \\ \Delta\dot{y}_{k+j} \\ \Delta\dot{\bar{y}}_{k+j} \end{bmatrix} = \begin{bmatrix} A & 0 & 0 & 0 \\ B & A^T & 0 & 0 \\ P & 0 & A & 0 \\ H & P^T & B & A^T \end{bmatrix}^j \begin{bmatrix} \Delta y_k \\ \Delta\bar{y}_k \\ \Delta\dot{y}_k \\ \Delta\dot{\bar{y}}_k \end{bmatrix} + \mathcal{O}\begin{bmatrix} \| \Delta y_k\|^2 + \\ \|\Delta\bar{y}_k\|^2 + \\ \|\Delta\dot{y}_k\|^2 + \\ \|\Delta\dot{\bar{y}}_k\|^2 \end{bmatrix}. \tag{15}$$

Assuming at first again that A has distinct real eigenvalues we may use the same transformation as in (8) and the j-th power can be rewritten as follows,

$$\begin{bmatrix} A & 0 & 0 & 0 \\ B & A^T & 0 & 0 \\ P & 0 & A & 0 \\ H & P^T & B & A^T \end{bmatrix}^j = \begin{bmatrix} T & 0 & 0 & 0 \\ T^{-T}C^T & T^{-T} & 0 & 0 \\ 0 & 0 & T & 0 \\ 0 & 0 & T^{-T}C^T & T^{-T} \end{bmatrix} \cdots$$

$$\cdots \begin{bmatrix} \Gamma^j & 0 & 0 & 0 \\ jD\Gamma^{j-1} & \Gamma^j & 0 & 0 \\ \tilde{P}_j & 0 & \Gamma^j & 0 \\ \tilde{H}_j & \tilde{P}_j^T & j\Gamma^{j-1}D & \Gamma^j \end{bmatrix} \begin{bmatrix} T^{-1} & 0 & 0 & 0 \\ CT^{-1} & T^T & 0 & 0 \\ 0 & 0 & T^{-1} & 0 \\ 0 & 0 & CT^{-1} & T^T \end{bmatrix}$$

where with $\tilde{P} \equiv T^{-1}PT$ and $\tilde{H} \equiv T^THT$,

$$\begin{bmatrix} \tilde{P}_j & 0 \\ \tilde{H}_j & \tilde{P}_j^T \end{bmatrix} \sum_{i=1}^{j} \begin{bmatrix} \Gamma^{i-1} & 0 \\ (i-1)\Gamma^{i-2}D & \Gamma^{i-1} \end{bmatrix} \begin{bmatrix} \tilde{P} & 0 \\ \tilde{H} & \tilde{P}^T \end{bmatrix} \begin{bmatrix} \Gamma^{j-i} & 0 \\ (j-i)\Gamma^{j-i-1}D & \Gamma^{j-i} \end{bmatrix}$$

Here we used the relation (7) once again. Hence we have the expressions

$$\tilde{P}_j = \sum_{i=1}^{j} \Gamma^{i-1} \tilde{P} \Gamma^{j-i}$$
$$\tilde{H}_j = \textstyle\sum_{i=1}^{j} (i-1)\Gamma^{i-2} D \tilde{P} \Gamma^{j-i} + \Gamma^{i-1} \tilde{H} \Gamma^{j-i} + (j-i)\Gamma^{i-1} P \Gamma^{j-i-1} D$$

Taking norms we obtain for constants c_1 and c_2

$$\|\tilde{P}_j\| \le c_1\, j\, \rho_*^{j-1} \;\text{ and }\; \|\tilde{H}_j\| \le c_2\, j^2\, \rho_*^{j-2}$$

The later inequality is true because Γ has like A the spectral radius $\rho_* < 1$. Thus we can estimate all four error components as follows.

$$\begin{aligned}
\|\Delta y_{k+j}\| &\le \rho_*^j c_{11} \|\Delta y_k\| + O(\|\Delta y_k\|^2) \\
\|\Delta \bar{y}_{k+j}\| &\le \rho_*^j \left[c_{22}\|\Delta \bar{y}_k\| + c_{21}\, j\, |\Delta y_k\|\right] + O(\|\Delta y_k\|^2 + \|\Delta \bar{y}_k\|^2) \\
\|\Delta \dot{y}_{k+j}\| &\le \rho_*^j \left[c_{33}\|\Delta \dot{y}_k\| + c_{31}\, j\, \|\Delta y_k\|\right] + O(\|\Delta y_k\|^2 + \|\Delta \dot{y}_k\|^2) \\
\|\Delta \dot{\bar{y}}_{k+j}\| &\le \rho_*^j \left[c_{44}\|\Delta \dot{\bar{y}}_k\| + c_{41} j^2 \|\Delta y_k\| + c_{42} j (\|\Delta \bar{y}_k\| + \|\Delta \dot{y}_k\|\right] \\
&+ O(\|\Delta y_k\|^2 + \|\Delta \bar{y}_k\|^2 + \|\Delta \dot{y}_k\|^2 + \|\Delta \dot{\bar{y}}_k\|^2)
\end{aligned}$$

These upper bounds apply in the nonlinear case under the restricted assumption on A. While they suggest that the higher derivatives lag behind, this relation can only been established if we assume linearity and draw on the more detailed analysis in the Appendix. Again, it is critical but reasonable to assume that the relation (10) is satisfied. Then, under the assumptions of Lemma A.2, it follows by (20) that

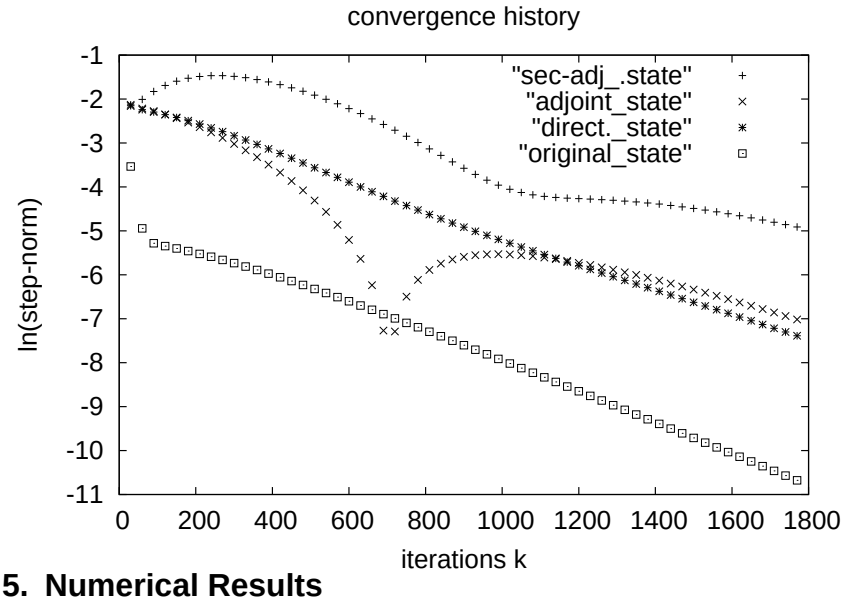
$$\underline{C} \;\le\; \frac{1}{k}\,\frac{\|\Delta \dot{y}_k\|}{\|\Delta y_k\|} \;\le\; \overline{C}, \qquad \underline{D} \;\le\; \frac{1}{k^2}\,\frac{\|\Delta \dot{\bar{y}}_k\|}{\|\Delta y_k\|} \;\le\; \overline{D}$$

for some constants $\underline{C}, \overline{C}, \underline{D}, \overline{D} > 0$.

This implies the proportionality relations

$$\|\Delta \dot{y}_k\| \sim k\, \|\Delta y_k\| \sim k\, \rho^k \quad \text{and} \quad \|\Delta \dot{\bar{y}}_k\| \sim k^2\, \|\Delta y_k\| \sim k^2\, \rho^k,$$

where ρ denotes the spectral radius of A. This means in particular that the second derivatives lag behind the first derivatives by a factor of order k and thus behind the original iteration by a factor of order k^2.



5. Numerical Results

The following results were obtained on the boundary control problem

$$\Delta_x y(x) + e^{y(x)} = 0 \quad \text{for} \quad x = (x_1, x_2) \in [0,1]^2$$

with the periodic and Dirichlet boundary conditions

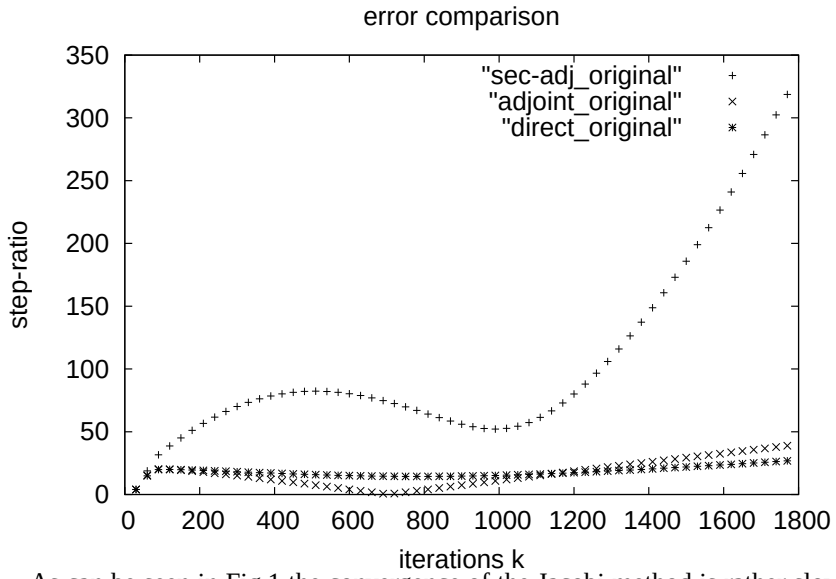
$$y(0,\zeta) = y(1,\zeta), \quad y(\zeta,0) = \sin(2\pi\zeta), \quad y(\zeta,1) = u(\zeta) \quad \text{for} \quad \zeta \in [0,1]$$

The function u is viewed as a boundary control that can be varied to minimize the objective function

$$f(y,u) = \int_0^1 \left[\frac{\partial y(\eta,\zeta)}{\partial \eta}\Big|_{\eta=0} - 4 - \cos(2\pi\zeta) \right]^2 d\zeta + \sigma \int_0^1 \left[u(\zeta)^2 + u'(\zeta)^2 \right] d\zeta$$

In the following calculations we used $\sigma = 0.001$ and set constantly $u(\zeta) = 2.2$. This value is not all that far from the fold point where solutions cease to exist.

We use a central difference discretization with the mesh-width $1/12.0$ so that the resulting algebraic system involves 144 equations in as many variables. Since the nonlinearities occur only on the diagonal one can easily implement Jacobi's method to obtain the basic function $G(y,u)$. For this simple example we also coded by hand the corresponding derived functions $\bar{G}, \dot{G}$ and even $\ddot{\bar{G}}$ as defined in (1, 12) and (13), respectively. The results were later confirmed using the automatic differentiation tool ADOL-C [6].



As can be seen in Fig.1 the convergence of the Jacobi method is rather slow with the common R-factor being about $(1 - 1/300)$. The lowest curve represents the natural logarithms of the Euclidean norm ratios $\|y_{k+1} - y_k\|/\|y_1 - y_0\|$, which provide some indication of the norm ratios $\|\Delta y_k\|/\|\Delta y_0\|$. In view of the very slow convergence this relation need certainly not be very close. Nevertheless the theory is basically confirmed with the first direct and adjoint derivatives $\|\dot{y}_{k+1} - \dot{y}_k\|/\|\dot{y}_1 - \dot{y}_0\|$ and $\|\bar{y}_{k+1} - \bar{y}_k\|/\|\bar{y}_1 - \bar{y}_0\|$ lagging somewhat behind and the second derivatives $\|\ddot{y}_{k+1} - \ddot{y}_k\|/\|\ddot{y}_1 - \ddot{y}_0\|$ coming in last. The ratio between these derivative quantities and the original iterates themselves is plotted in Fig. 2. After an initial transition phase one sees quite clearly a growth proportional to k and k^2 for the first and second derivatives, respectively. While the adjoints were defined as in (3) by the gradient of f, the direct differentiation was performed simultaneously with respect to all components of the discretized u so that the quantity $\dot{u}$ occurring in (12) and (13) was in fact the identity matrix of order 12. Consequently, $\dot{y}_k$ and $\bar{\dot{y}}_k$ had also 12 times as many components as the underlying y_k and $\bar{y}_k$, which are of the same size.

6. Summary, Conclusion and Outlook

We studied the convergence behavior of fixed point iterations for derivatives of implicit functions. These recurrences are generated in a completely mechanical fashion from a user supplied contractive fixed point solver for evaluating the implicit function. While the

contractivity and thus the asymptotic convergence rate is inherited by the derived solvers there is a certain time lag. This is not really surprising since the equations for the adjoints $\bar{y}$ and those for the feasible directions $\dot{y}$ are dependent on y and both in turn impact the second order adjoint equation for $\dot{\bar{y}}$. Mathematically we obtain Jordan blocks of size 2 for the double eigenvalues of the first derivative systems and of size 3 for the quadruple eigenvalues of the second order adjoint system. One does not obtain blocks of size 4 since the $(3, 2)$ sub-block in the big Jacobian system vanishes identically. Otherwise it would connect the two first derivative systems.

Generally if one were to iteratively evaluate derivatives of order d one can expect that the relative errors compared to those of the underlying function iteration grows like k^d, where k is the iteration counter. In the context of constrained optimization one can expect that the correct values of reduced gradients (4) and Hessians (14) are obtained slower than feasibility so that optimality will be arrived at in the tangential fashion that is familiar from SQP calculations [11, 12, 13]. In fact when the state equation only be solved by a slowly convergent fixed point solver as we have assumed throughout it makes little sense to apply an SQP type algorithms. Instead one will prefer a so-called one-shot optimization strategy [14], where feasibility and optimality is achieved at the same time. We are currently investigating a piggy-back optimization scheme, where a third iteration updating the design variables u on the basis of approximate reduced gradient information is appended to (3).

A. Convergence Behavior of Linear Recurrences

In this section, we study the linear recurrences in (6) and (15) in detail. The transition matrices of both recurrences have a very particular structure and the following two lemmas show the convergence behavior that is (generically) induced by these structures.

Lemma A.1 *Consider a linear recurrence of the form*

$$\begin{bmatrix} f_{k+1} \\ \bar{f}_{k+1} \end{bmatrix} = \begin{bmatrix} A & 0 \\ B & A^T \end{bmatrix} \begin{bmatrix} f_k \\ \bar{f}_k \end{bmatrix},$$

where A and B are real $n \times n$ matrices, and assume that ρ, the spectral radius of A, satisfies $0 < \rho < 1$. Let $\lambda_1, \ldots, \lambda_r$ denote the eigenvalues of A with $|\lambda_i| = \rho$. It is assumed that each λ_i is simple and satisfies the following conditions :

1) if λ_i is real then $x_i^T B x_i \neq 0$, where x_i is a right eigenvector belonging to λ_i.

2) if λ_i is complex then $x_{i,R}^T B x_{i,R} \neq x_{i,I}^T B x_I$ or $x_{i,R}^T B x_{i,I} \neq -x_{i,I}^T B x_{i,R}$, where $x_{i,R}$ and $x_{i,I}$ are the real and imaginary parts of a right eigenvector belonging to λ_i.

Let the columns of $Y \in \mathbb{R}^{n\times r}$ form a basis for the space spanned by the left eigenvectors $y_1, \ldots, y_r$ belonging to $\lambda_1, \ldots, \lambda_r$. Then there exist constants $C_1, C_2, C_3, C_4 > 0$ so that

$$C_1\rho^j \le \frac{\|f_{k+j}\|}{\|Y^T f_k\|} \le C_2\rho^j, \quad C_3 j\rho^{j-1} \le \frac{\|\bar{f}_{k+j}\|}{\|Y^T f_k\|} \le C_4 j\rho^{j-1}, \tag{16}$$

provided that $\|Y^T f_k\| \neq 0$.

Proof. If λ_i is real, there is an invertible real matrix T such that the first column of T is x_i and

$$\tilde{J} = \begin{bmatrix} T^{-1} & 0 \\ 0 & T^T \end{bmatrix} \begin{bmatrix} A & 0 \\ B & A^T \end{bmatrix} \begin{bmatrix} T & 0 \\ 0 & T^{-T} \end{bmatrix} = \begin{bmatrix} \lambda_i & 0 & 0 & 0 \\ 0 & A_{22} & 0 & 0 \\ B_{11} & B_{12} & \lambda_i & 0 \\ B_{21} & B_{22} & 0 & A_{22}^T \end{bmatrix},$$

where $B_{11} = x_i^T B x_i$. Since λ_i is simple, the matrix $A_{22} - \lambda_i I$ is invertible. Setting $R_1 = B_{12}(A_{22} - \lambda_i I)^{-1}$ and $R_2 = (\lambda_i I - A_{22}^T)^{-1} B_{21}$ yields

$$\hat{J} = R^{-1}\tilde{J}R = \begin{bmatrix} \lambda_i & 0 & 0 & 0 \\ 0 & A_{22} & 0 & 0 \\ B_{11} & 0 & \lambda_i & 0 \\ 0 & B_{22} & 0 & A_{22}^T \end{bmatrix} \quad \text{with} \quad R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & R_1 & 1 & 0 \\ R_2 & 0 & 0 & I \end{bmatrix}. \tag{17}$$

If y_i denotes the first column of T^{-T} then y_i is a left eigenvector belonging to λ_i. This implies $|y_i^T f_{k+j}| = \rho^j |y_i^T f_k|$ while $|x_i^T \bar{f}_{k+j}| = j\rho^{j-1}|B_{11} y_i^T f_k| + \mathcal{O}(\rho^j)$.

If $\lambda_i = \lambda_{i,R} + \imath\lambda_{i,I}$ is complex there is an invertible real matrix T such that the first two columns of T are $x_{i,R}, x_{i,I}$ and

$$\tilde{J} = \begin{bmatrix} T^{-1} & 0 \\ 0 & T^T \end{bmatrix} \begin{bmatrix} A & 0 \\ B & A^T \end{bmatrix} \begin{bmatrix} T & 0 \\ 0 & T^{-T} \end{bmatrix} = \begin{bmatrix} A_{11} & 0 & 0 & 0 \\ 0 & A_{22} & 0 & 0 \\ B_{11} & B_{12} & A_{11}^T & 0 \\ B_{21} & B_{22} & 0 & A_{22}^T \end{bmatrix}, \tag{18}$$

with

$$A_{11} = \begin{bmatrix} \lambda_{i,R} & \lambda_{i,I} \\ -\lambda_{i,I} & \lambda_{i,R} \end{bmatrix}, \quad B_{11} = \begin{bmatrix} x_{i,R}^T B x_{i,R} & x_{i,R}^T B x_{i,I} \\ x_{i,I}^T B x_{i,R} & x_{i,I}^T B x_{i,I} \end{bmatrix} =: \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}.$$

If R_1 and R_2 denote the solutions of the Sylvester equations $R_1 A_{22} - A_{11} R_1 = B_{12}$ and $R_2 A_{11}^T - A_{22}^T R_2 = B_{12}$, respectively, then the same transformation as in (17) can be used to eliminate the off-diagonal blocks B_{12} and B_{21} in (18), see also [3]. Decompose

$$B_{11} = V + W := \frac{1}{2}\begin{bmatrix} b_{11}+b_{22} & b_{12}-b_{21} \\ b_{21}-b_{12} & b_{11}+b_{22} \end{bmatrix} + \frac{1}{2}\begin{bmatrix} b_{11}-b_{22} & b_{12}+b_{21} \\ b_{12}+b_{21} & b_{22}-b_{11} \end{bmatrix}.$$

Since V is in the range of the Sylvester operator $R \mapsto RA_{11}^T - A_{11}R$, we can eliminate this part by a similarity transformation and set $B_{11} = W$, which is under the given assumptions different from zero. Then we have $B_{11}A_{11} = A_{11}^T B_{11}$ and therefore

$$\begin{bmatrix} A_{11} & 0 \\ B_{11} & A_{11}^T \end{bmatrix}^j = \begin{bmatrix} A_{11}^j & 0 \\ j(A_{11}^T)^{j-1}B_{11} & (A_{11}^T)^j \end{bmatrix}.$$

If $y_{i,R}$ and $y_{i,I}$ denote the first two columns of T^{-T} then $y_i = y_{i,R} + \imath y_{i,I}$ is a left eigenvector belonging to λ_i. This implies $\|[y_{i,R}, y_{i,I}]^T f_{k+j}\| = \rho^j \|[y_{i,R}, y_{i,I}]^T f_k\|$, while

$$\|[x_{i,R}, x_{i,I}]^T \bar{f}_{k+j}\| = j\rho^{j-1}\|B_{11}\| \, \|[y_{i,R}, y_{i,I}]^T f_k\| + \mathcal{O}(\rho^j).$$

Altogether, this shows the existence of constants $\tilde{C}_1, \tilde{C}_2, \tilde{C}_3, \tilde{C}_4 > 0$ such that

$$\begin{array}{rcccl} \tilde{C}_1\rho^j\|Y^T f_k\| & \leq & \|Y^T f_{k+j}\| & \leq & \tilde{C}_2\rho^j\|Y^T f_k\|, \\ \tilde{C}_3 j\rho^{j-1}\|Y^T f_k\| & \leq & \|X^T \bar{f}_{k+j}\| & \leq & \tilde{C}_4 j\rho^{j-1}\|Y^T f_k\|, \end{array} \tag{19}$$

where the columns of $X \in \mathbb{R}^{n\times r}$ form a basis for $x_1, \ldots, x_r$. This concludes the proof as $\|f_{k+j}\| = \|Y^T f_{k+j}\| + \mathcal{O}(\hat{\rho}^j)$ and $\|\bar{f}_{k+j}\| = \|X^T \bar{f}_{k+j}\| + \mathcal{O}(\hat{\rho}^j)$ for some $\hat{\rho} < \rho$. □

Several remarks are in order :

1) The second condition in Lemma A.1 can be written in the more compact form $x_i^H B x_i \neq 0$ with $x_i = x_{i,R} + \imath x_{i,I}$.

2) If B is skew-symmetric then the two conditions in Lemma A.1 are always violated, independent of the eigenvectors of A. Moreover, $\left[\begin{smallmatrix} A & 0 \\ B & A^T \end{smallmatrix}\right]$ is a so called skew-Hamiltonian matrix, which can always be put into block diagonal form $\left[\begin{smallmatrix} A & 0 \\ 0 & A^T \end{smallmatrix}\right]$ by a similarity transformation [1, 2]. Hence, the second inequality in (16) does not hold for this case.

3) In the applications considered in this paper, B is symmetric and it is also reasonable to assume B to be positive definite. In this case, the two conditions in Lemma A.1 are always satisfied, independent of the eigenvectors of A.

Lemma A.2 *Consider a linear recurrence of the form*

$$\begin{bmatrix} f_{k+1} \\ \bar{f}_{k+1} \\ g_{k+1} \\ \bar{g}_{k+1} \end{bmatrix} = \begin{bmatrix} A & 0 & 0 & 0 \\ B & A^T & 0 & 0 \\ P & 0 & A & 0 \\ H & P^T & B & A^T \end{bmatrix} \begin{bmatrix} f_k \\ \bar{f}_k \\ g_k \\ \bar{g}_k \end{bmatrix},$$

where A, B, H, P are real $n \times n$ matrices. Assuming that the spectral radius ρ of A satisfies $0 < \rho < 1$, let $\lambda_1, \ldots, \lambda_r$ denote the eigenvalues of A with $|\lambda_i| = \rho$. Moreover, it

is assumed that each λ_i is simple and satisfies $x_i^H B x_i \neq 0$, $x_i^H H x_i \neq 0$, and $x_i^H P y_i \neq 0$, where x_i and y_i are right and left eigenvectors belonging to λ_i.

Let the columns of $Y \in \mathbb{R}^{n\times r}$ form a basis for the space spanned by the left eigenvectors $y_1, \ldots, y_r$ belonging to $\lambda_1, \ldots, \lambda_r$. Then there exist constants $C_1, \ldots, C_8 > 0$ so that (16) is satisfied, and additionally

$$C_5 j\rho^{j-1} \leq \frac{\|g_{k+j}\|}{\|Y^T f_k\|} \leq C_6 j\rho^{j-1}, \quad C_7 j^2\rho^{j-2} \leq \frac{\|\bar{g}_{k+j}\|}{\|Y^T f_k\|} \leq C_8 j^2\rho^{j-2}, \tag{20}$$

provided that $\|Y^T f_k\| \neq 0$.

Proof. If λ_i is real then by similar arguments as in the proof of Lemma A.1, we may restrict ourselves to the iteration

$$\begin{bmatrix} y_i^T f_{k+1} \\ x_i^T \bar{f}_{k+1} \\ y_i^T g_{k+1} \\ x_i^T \bar{g}_{k+1} \end{bmatrix} = \begin{bmatrix} \lambda_i & 0 & 0 & 0 \\ B_{11} & \lambda_i & 0 & 0 \\ P_{11} & 0 & \lambda_i & 0 \\ H_{11} & P_{11} & B_{11} & \lambda_i \end{bmatrix} \begin{bmatrix} y_i^T f_k \\ x_i^T \bar{f}_k \\ y_i^T g_k \\ x_i^T \bar{g}_k \end{bmatrix},$$

where $B_{11} = x_i^T B x_i$, $H_{11} = x_i^T H x_i$ and $P_{11} = y_i^T P x_i$. Since

$$\begin{bmatrix} \lambda_i & 0 & 0 & 0 \\ B_{11} & \lambda_i & 0 & 0 \\ P_{11} & 0 & \lambda_i & 0 \\ H_{11} & P_{11} & B_{11} & \lambda_i \end{bmatrix}^j =$$

$$= \begin{bmatrix} \lambda_i^j & 0 & 0 & 0 \\ j\lambda^{j-1} B_{11} & \lambda_i^j & 0 & 0 \\ j\lambda^{j-1} P_{11} & 0 & \lambda_i^j & 0 \\ j\lambda^{j-1} H_{11} + (j^2 - j)\lambda^{j-2} H_{11} P_{11} & j\lambda^{j-1} P_{11} & j\lambda^{j-1} B_{11} & \lambda_i^j \end{bmatrix}$$

for $j > 1$, we have

$$\begin{aligned} |y_i^T f_{k+j}| &= \rho^j |y_i^T f_k|, \\ |x_i^T \bar{f}_{k+j}| &= j\rho^{j-1} |B_{11} y_i^T f_k| + \mathcal{O}(\rho^j), \\ |y_i^T g_{k+j}| &= j\rho^{j-1} |P_{11} y_i^T f_k| + \mathcal{O}(\rho^j), \\ |x_i^T \bar{g}_{k+j}| &= j^2\rho^{j-2} |H_{11} P_{11} y_i^T f_k| + \mathcal{O}(j\rho^j), \end{aligned}$$

The complex case is treated analogously. Altogether, there exist constants $\tilde{C}_1, \ldots, \tilde{C}_8 > 0$ so that (19) is satisfied, and additionally

$$\begin{aligned} \tilde{C}_5 j\rho^j \|Y^T f_k\| &\leq \|Y^T g_{k+j}\| \leq \tilde{C}_6 j\rho^j \|Y^T f_k\|, \\ \tilde{C}_7 j^2\rho^{j-2} \|Y^T f_k\| &\leq \|X^T \bar{g}_{k+j}\| \leq \tilde{C}_8 j^2\rho^{j-2} \|Y^T f_k\|, \end{aligned}$$

which concludes the proof using the same argument that concludes the proof of Lemma A.1.
□

B. Bibliographie

[1] P. Benner, D. Kressner, and V. Mehrmann. Skew-Hamiltonian and Hamiltonian eigenvalue problems : Theory, algorithms and applications, 2004. In *Proceedings of the Conference on Applied Mathematics and Scientific Computing, Brijuni (Croatia)*, June 23-27, 2003, pages, 3–39, Springer-Verlag, 2005.

[2] H. Faßbender, D. S. Mackey, N. Mackey, and H. Xu. Hamiltonian square roots of skew-Hamiltonian matrices. *Linear Algebra Appl.*, 287(1-3) :125–159, 1999.

[3] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, third edition, 1996.

[4] A. Griewank, C.H. Bischof, G.F. Corliss, A. Carle, and K. Williamson. Derivative Convergence of Iterative Equation Solvers. *Optimization Methods and Software*, 2 :321–355, 1993.

[5] A. Griewank and C. Faure. Reduced functions, gradients and hessians from fixed point iteration for state equations. *Numerical Algorithms*, 30(2) :113–139, 2002.

[6] A. Griewank, D. Juedes, and J. Utke. ADOL-C, a package for the automatic differentiation of algorithms written in C/C++. *TOMS*, 22(2) :131–167, 1996.

[7] M. Hinze and T. Slawig. Adjoint gradients compared to gradients from algorithmic differentiation in instantaneous control of the navier-stokes equations. *Optimization Methods and Software*, 18(3) :299–315.

[8] A. Jameson. Optimum aerdynamic design using CFD and control theory. In *12th AIAA Computational Fluid Dynamics Conference, AIAA Paper 95-1729*, San Diego, CA, 1995. American Institute of Aeronautics and Astronautics.

[9] B. Mohammadi and O. Pironneau. *Applied Shape Optimization for Fluids*. Numerical Mathematics and Scientific Computation. Cladenen Press, Oxford, 2001.

[10] P.A. Newman, G.J.-W. Hou, H.E. Jones, A.C. Taylor, and V.M. Korivi. Observations on computational methodologies for use in large-scale, gradient-based, multidisciplinary design incorporating advanced CFD codes. Technical Memorandum 104206, NASA Langley Research Center, February 1992. AVSCOM Technical Report 92-B-007.

[11] J. Nocedal and S.J. Wright. *Numerical Optimization*, Springer Series in Operation Research. Springer Verlag, New York,...,Tokyo, 1999.

[12] E. W. Sachs. Control applications of reduced SQP methods. In R. Bulirsch and D. Kraft, editors, *Computational Optimal Control*, volume 115 of *Int. Series Num. Math.*, pages 89–104. Birkhäuser, 1994.

[13] V. H. Schulz. Solving discretized optimization problems by partially reduced SQP methods. *Computing and Visualization in Science*, 1 :83–96, 1998.

[14] S. Ta'asan, G. Kuruvila, and M.D. Salas. Aerdynamic design and optimization in one shot. In *30th AIAA Aerospace Sciences Meeting and Exhibit, AIAA Paper 91-0025*, Reno, Nevada, 1992. American Institute of Aeronautics and Astronautics.