Derivative Convergence for Iterative Equation Solvers*

Andreas Griewank and Christian Bischof, Argonne National Laboratory,
George Corliss, Marquette University,
Karen Williamson, Rice University

March 30, 1999

Abstract When nonlinear equation solvers are applied to parameter-dependent problems, their iterates can be interpreted as functions of these variable parameters. If they exist, the derivatives of these iterated functions can be recursively evaluated by the forward mode of automatic differentiation. Then one may ask whether and how fast these derivative values converge to the derivative of the implicit solution function, which may be needed for parameter identification, sensitivity studies, or design optimization.

It is shown here that derivative convergence is achieved with an R-linear or possibly R-superlinear rate for a large class of memory-less contractions or secant updating methods. For a wider class of multi-step contractions, we obtain R-linear convergence of a simplified derivative updating scheme, which is more economical and can be easily generalized to second higher derivatives. We also formulate a constructive criterion for derivative convergence based on the implicit function theorem. All theoretical results are confirmed by numerical experiments on small test examples.

Keywords. Derivative convergence, automatic differentiation, implicit functions, preconditioning, Newton-like methods, secant updates.

1 Introduction and Assumptions on $F(x,t) = 0$

Many functions of practical interest are defined implicitly as solutions to differential or algebraic equations. The values of these functions are typically evaluated by iterative procedures with a variable number of steps and various, often discontinuous, adjustments. The corresponding computer programs contain branches, and the results are often strictly speaking not everywhere differentiable in the data. Then one may ask if and how automatic differentiation can still be expected to yield derivative values that are reasonable approximations to the underlying implicitly defined derivatives.

Automatic, or computational, differentiation is a chain rule based technique for evaluating the derivatives of functions defined by algorithms, usually in the form of computer programs written in Fortran, C, or some other high level language. If the program can theoretically be unrolled into a finite sequence of arithmetic operations and elementary function calls, then derivatives can be propagated recursively. Exceptions arise when there is a division by zero or one of the elementary functions is evaluated at a point of nondifferentiability. These local contingencies are easily detected

*This work was supported by the Office of Scientific Computing, U.S. Department of Energy, under Contract W-31-109-Eng-38.

and arise only in marginal situations where the undifferentiated evaluation algorithm is already poorly conditioned. For a general review of the theory, implementation and application of automatic differentiation, see [4].

Rather than as a practical problem for automatic differentiation, one can also view the question raised here as a purely theoretical one, namely, whether the iterates generated for parameter-dependent problems converge not only pointwise, but also with respect to some Sobolev norm involving derivatives with respect to the parameters. This theoretical aspect will not be fully explored here, as only pointwise convergence of the derivatives is established. Throughout we will analyze the situation where a nonlinear system

$$F(x,t) \;=\; 0 \quad with \quad F \;: I\!R^n \times I\!R \mapsto I\!R^n$$

is solved for $x(t)$ for fixed t by an iteration of the form

$$x_{k+1} \;=\; \Phi_k(x_k,t) \;\equiv\; x_k \;-\; P_k F(x_k,t)\;. \tag{1}$$

We wish to compute the total derivatives $x'(t) = dx(t)/dt$. Without loss of generality, we have restricted our framework to the case of a single scalar parameter $t \in I\!R$ since multivariate derivatives can always be constructed from families of univariate derivatives [1]. Total derivatives with respect to t will be denoted by primes, and partial derivatives (with x kept constant) by the subscript t.

In this paper, we consider two approaches to computing the desired implicitly defined derivative $x'(t)$. The "simplified" approach treats the P_k as if they were independent of x_k. The "fully differentiated" approach differentiates the entire iterative algorithm.

The rest of this section discusses some of the practical and theoretical pitfalls of using automatic differentiation in the computation of $x'(t)$.

Obviously, any sequence $\{x_k\}_{k\geq 0}$ for which $F(x_k,t)$ never vanishes exactly can be written in the form (1), unless we place some restriction on the $n \times n$ matrices P_k and thus the sequence of iteration functions Φ_k. The assumptions on the P_k that we will make are quite natural and almost necessary for a numerically stable iterative process.

Assumption 1 (Regularity) *For some fixed t the iteration converges to a solution, so that*

$$x_k \;\rightarrow\; x_* = x(t) \quad with \quad F(x(t),t) = 0\;.$$

Moreover, on some ball with radius $\rho > 0$ centered at $(x(t),t)$ the function F is jointly Lipschitz-continuously differentiable and has a nonsingular Jacobian $F_x(x,t) = \partial F(x,t)/\partial x$ with respect to x, so that for two constants c_0, L, and all $\|x - x(t)\| < \rho$

$$\|F_x^{-1}(x,t)\|, \|[F_x(x,t), F_t(x,t)]\| \;\leq\; c_0\;,$$

and

$$\|[F_x(x,t), F_t(x,t)] - [F_x(x_*,t), F_t(x_*,t)]\| \;\leq\; L\|x - x_*\|\;,$$

where we may use l_2 norms without loss of generality.

Under this assumption, local convergence is guaranteed for Newton's method with $P_k = F_x(x_k,t)^{-1}$ or for the Picard iteration with $P_k = I$ if the spectral radius of $(I - F_x)$ is less than one. If this condition is not met by the original system $F = 0$, one might try to find a fixed preconditioner

$P_k = P$ so that $(I - P\,F_x)$ is contracting. Alternatively, one may select P_k as a function of x_k, for example by performing an incomplete triangular decomposition so that we can write

$$P_k = P(x_k, t)\ .$$

Then we will refer to the iteration (1) as a memory-less contraction provided the following condition is met.

Assumption 2 (Contractivity) *The discrepancies*

$$D_k \ = \ [I - P_k\,F_x(x_k, t)]$$

satisfy

$$\delta_k \ \equiv \ \|D_k\| \ \leq \ \delta < 1 \tag{2}$$

with respect to some induced matrix norm so that in the limit

$$\delta_* \ \equiv \ \overline{\lim}_k\,\delta_k \ \leq \delta\ .$$

For the class of methods satisfying this contractivity assumption (which includes Newton's method with analytical Jacobians or divided difference approximations), derivative convergence of the derivatives can be obtained easily. As an immediate consequence of Assumptions 1 and 2, we note that by standard arguments

$$\|P_k\| \ \leq \ c_0(1+_\delta) \quad \textit{and} \quad \|P_k^{-1}\| \leq c_0/(1-\delta)\ . \tag{3}$$

In the case of secant methods [5], the condition (2) is usually imposed for $k = 0$ and deduced for $k > 1$ to guarantee local convergence. If one assumes a certain kind of uniform linear independence for the sequence of the search directions, it can be shown [6] that $\delta_* = 0$. This is a sufficient, but by no means necessary, condition for Q-superlinear convergence. It can be enforced by taking so-called special steps [7] for the sole purpose of reducing the discrepancy D_k. We will see that $\delta_* = 0$ implies R-superlinear rather than just R-linear convergence of the derivatives. Hence, the extra expense of special updating steps might be justified on parameter dependent problems. Secant methods are not memory-less because the preconditioners P_k are computed recursively from step to step. Therefore, they must be considered as functions of all previous points x_k and of the initial choice P_0. Since in formula (1), the matrix P_k must also absorb step multipliers, this functional dependence need not be smooth and may have discontinuities. In that case, the transition from x_k to x_{k+1} may also be nondifferentiable, so that the classical chain rule is not directly applicable.

Even when x_0, P_0, and all subsequent P_k are smooth functions of t, it may be uneconomical to calculate the corresponding derivatives explicitly. For example in the case of Newton's method, the explicit calculation of derivatives would involve the propagation of derivatives through the triangular decomposition of the Jacobian, a process that involves $n^3/3$ arithmetic operations in the dense case. However, we know from the implicit function theorem that

$$F_x(x(t), t)\,x'(t) \ = \ -F_t(x(t), t)\ . \tag{4}$$

In particular, this means that $x'(t)$ is defined in terms of the first derivatives of F alone and does not depend on the second derivatives F_{xx} and F_{xt} . Yet these tensors come implicitly into play if derivatives with respect to x are propagated through the Newton iteration function $\Phi_k(x_k, t) = x_k - F_x(x_k, t)^{-1}F(x_k, t)$. The same applies to any other iteration where the preconditioner P_k depends in some way on derivatives of F with respect to x or t. Therefore, we will examine a

simplified derivative recurrence, where the P_k are considered as (piecewise) constants with respect to the total differentiation of the recurrence (1) with respect to t. We call this the simplifies approach.

On the other hand, it may be difficult to determine which quantities in a complicated nonlinear equation solver need to be differentiated and which can be considered as constants because they belong to the calculation of the preconditioner P_k. This distinction must then be conveyed to the automatic differentiation software by suitably annotating the code or retyping some of its variables. Therefore, one may prefer to adopt a black box approach and differentiate the whole iterative algorithm as though it were a straight line code. This is what we call the fully differentiated approach. Also, the derivative $x'_k = dx_k/dt$ of the iterate x_k that is finally accepted does represent the local tangent of the approximate solution set, which should be close to the exact solution curve if the convergence occurs with some degree of uniformity.

For either the simplified or fully differentiated approach, it seems pretty clear that the derivatives cannot converge faster than the iterates themselves, unless the problem is linear or has some other very special structure. We will show for Newton's method and for secant updating methods that the derivatives converge R-quadratically and R-linearly, respectively. Especially in the case of secant updating methods, we must therefore expect that the derivatives may lag behind the iterates during the final approach to the solution. Fortunately, we can constructively check the accuracy of any derivative approximation so that a premature termination can be avoided if accurate derivative values are required.

Gilbert showed in [3] that the derivatives dx_k/dt converge in the limit to the desired tangent $x' = x'(t)$, provided the spectral radius of $\partial\Phi(x,t)/\partial x$ is less than one in the vicinity of (x_*, t). This fundamental result has removed some serious doubts regarding the general applicability of automatic differentiation. It has been verified on several large codes, including cases where the assumptions of Gilbert's theorem do not appear to be satisfied. Therefore, we wish to relax the hypothesis and avoid derivatives that are not needed either from a theoretical or from a practical point of view. We will also establish rates of convergence, provide a practical stopping criterion, and extend the theory to higher derivatives and multi-step contractions.

The paper is organized as follows. In the next section, we motivate the simplified and fully differentiated derivative recurrences and develop some basic mathematical relations. In Section 3, we establish R-linear derivative convergence for the simplified recurrence under Assumptions 1 and 2 alone and for the fully differentiated recurrence under the additional assumption that the update function of the P_k satisfies a certain differentiability condition. Section 4 contains some generalizations. The paper concludes with a summary and discussion in Section 5.

2 Simplified and Fully Differentiated Recurrences

As we have indicated above, the basic recurrence (1) can be interpreted as one step of a Picard, or Richardson, iteration on the preconditioned nonlinear system

$$F_k(x,t) \;\equiv\; P_k\, F(x,t) \;=\; 0. \tag{5}$$

Provided P_k is nonsingular, as we will assume throughout, the solution set of each $F_k = 0$ is exactly the same as that of the original system $F = 0$. Consequently, the implicitly defined function $x(t)$ and its derivatives are independent from the sequence of preconditioners P_k. Their iterative evaluation certainly need not depend on the derivatives of P_k, which may not even exist.

Differentiating equation (5) with respect to t with P_k considered a constant, one obtains the equation defining $x'(t)$

$$P_k\, F_x(x(t),t)\, x'(t) \;=\; -P_k\, F_t(x(t),t)\,. \tag{6}$$

In the following formulae, we will often suppress the dependence on t, which should be understood. Applying the Richardson iteration to the preconditioned linear system (6) of equations evaluated at the "current" iterate x_k, one obtains the recurrence

$$\tilde{x}'_{k+1} \;=\; \tilde{x}'_k - P_k\,[F_x(x_k,t)\,\tilde{x}'_k \;+\; F_t(x_k,t)]\;. \tag{7}$$

Here the tilde over $\tilde{x}'_k$ indicates that these approximations to the derivative $x'(t)$ are in general not the derivatives of the x_k with respect to t, which may or may not exist. Subtracting the actual implicitly defined derivative

$$x'_* \;\equiv\; -F_x(x_*,t)^{-1}F_t(x_*,t);.$$

from both sides, we find that

$$\tilde{x}'_{k+1} - x'_* \;=\; D_k(\tilde{x}'_k - x'_*) + r'_k\;, \tag{8}$$

where

$$r'_k \;\equiv\; P_k[F_x(x_k,t)x'_* + F_t(x_k,t)] \;=\; \mathcal{O}(\|X_k - x_*\|)\;. \tag{9}$$

Since the perturbation r'_k tends to zero, equation (8) looks very much like a contraction and promises convergence of the x'_k to x'_*.

If the P_k are at least locally smooth functions of t so that the matrices $P'_k = dP_k(x(t))/dt$ are continuous, then the derivatives $x'_k = x'_k(t)$ exist and satisfy the recurrence

$$x'_{k+1} \;=\; x'_k - P_k\;[F_x(x_k,t)x'_k + F_t(x_k,t)] - P'_k\,F(x_k,t)\;, \tag{10}$$

which can be rewritten in the contractive form as

$$x'_{k+1} - x'_* \;=\; D_k(x'_k - x'_*) + r'_k \;-\; P'_kF(x_k,t)\;. \tag{11}$$

We refer to equation (7) as the *simplified* recurrence and to equation (10) as the *fully differentiated* recurrence. The label *update* will be avoided to reduce the danger of confusion with the Jacobian and Hessian update formulas that lie at the heart of secant methods.

Provided the P'_k stay bounded or do not blow up too fast with increasing k, the last term in the linear recurrence (11) becomes more and more negligible as the residual $F(x_k,t)$ approaches zero. In the remainder, we will analyze equation (7) as a special case of equation (10) with P_k considered as constant on some neighborhood of the current t. Obviously the two stage iteration defined by (1) and (10) can only be stationary at the (locally) unique fixed point $(x_k,x'_k) \;=\; (x_*,x'_*)$. In general, the iteration (10) will never reach this fixed point exactly. However, the derivative approximations x'_k can have no limit other than the correct value x'_*, unless the $P'_kF(x_k,t)$ converge by some fluke to a nonzero vector. This possibility would seem rather remote and can only occur if $\|P'_k\|$ tends to infinity exactly at the same rate as the reciprocal $1/\|F(x_k,t\|$. Note, that this cannot happen in the simplified derivative recurrence (7) for which $P'_k \equiv 0$ by definition. In the case of the full recurrence applied to secant methods, the Q-superlinear convergence rate ensures that the perturbation $P'_kF(x_k,t)$ tends to zero R-linearly, as we will show in the proof of Proposition 2 in Section 3.

In general, we expect that the derivatives x'_k exhibit roughly the same convergence behavior as the iterates x_k. To justify this optimism, we note that by Taylor's theorem

$$P_kF(x_k,t) \;=\; P_kF_x(x_k,t)\,(x_k - x_*) - r_k\;,$$

where

$$r_k \;=\; -P_k[F(x_k,t) - F_x(x_k,t)(x_k - x*)] \;=\; \mathcal{O}(\|x_k - x_*\|^2)\;. \tag{12}$$

Consequently, the iterates x_k defined by (1) satisfy the contractive recurrence

$$x_{k+1} - x_* \;=\; D_k(x_k - x_*) + r_k \;. \tag{13}$$

Hence we have essentially the same leading term in (8), (11), and (13). Taking norms, one obtains

$$\|x_{k+1} - x_*\| \;\leq\; \|D_k\|\|x_k - x_*\| + \|r_k\| \;,$$

so that the errors $\|x_k - x_*\|$ converge Q-linearly because of the contractivity assumption:

$$\overline{\lim}_k \frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} \;\leq\; \delta_* \;.$$

No matter how a derivative approximation x'_k was generated, its quality can be checked by evaluating the directional derivative

$$F'(x_k, t, x'_k) \;\equiv\; \left.\frac{\partial F(x_k + \tau\, x'_k, t + \tau)}{\partial \tau}\right|_{\tau=0} \tag{14}$$

$$\;=\; F_x(x_k, t)x'_k + F_t(x_k, t) \quad . \tag{15}$$

This vector can be evaluated cheaply in the forward mode of automatic differentiation, without the need to form the (potentially very large) Jacobian $F_x(x_k, t)$. Note that $P_k F'(x_k, t, x'_*) = r'_k$ as defined in (9). When $F'(x_k, t, x'_k)$ vanishes exactly, x'_k represents the tangent of the perturbed solution set

$$F^{-1}(F_k) \;\equiv\; \{x \in I\!R^n : F(x, t) = F(x_k, t)\} \quad .$$

If $F'(x_k, t, x'_k)$ does not vanish, one can substitute into the right hand side of (7) or (10) to improve the approximation. In general, the x'_k can only be as good approximations to x'_* as the x_k approximate x_*. Abbreviating

$$\rho_k \;\equiv\; \|x_k - x_*\| \quad and \quad \mu_k \;\equiv\; \|x'_k - x'_*\|$$

and setting

$$\eta_k \;\equiv\; (Lc_1 + \|P'_k\|)\rho_k \quad with \quad c_1 \;\equiv\; 2(c_0^2 + 1) \;, \tag{16}$$

one can bound the derivative errors as follows.

Lemma 1 *The regularity and contractivity imposed by Assumptions 1 and 2 imply that*

$$\mu_k \quad\leq\quad \frac{1}{(1-\delta)}\|P_k F'(x_k, t, x'_k)\| + Lc_0c_1\rho_k/2 \;, \tag{17}$$

$$\mu_{k+1} \quad\leq\quad \delta_k\mu_k + c_0\eta_k \;, \quad and \quad \|r'_k\| \leq c_1c_0L\rho_k \;, \tag{18}$$

for all $\rho_k < \rho$.

Proof. First we show that the function $F_x(x,t)^{-1}F_t(x,t) : I\!R^n \to I\!R^n$ with t fixed has the Lipschitz constant $Lc_0c_1/2$ at x_*.

$$\begin{aligned}
& \|F_x(x,t)^{-1}F_t(x,t) - F_x(x_*,t)^{-1}F_t(x_*,t)\| \\
\leq\; & \|F_x(x,t)^{-1}[F_t(x,t) - F_t(x_*,t)]\| + \|[F_x(x,t)^{-1} - F_x(x_*,t)^{-1}]F_t(x_*,t)\| \\
\leq\; & c_0\, L\, \|x - x_*\| + \|F_x(x,t)^{-1}\|\|F_x(x_*,t) - F_x(x,t)\|\|F_x^{-1}(x_*,t)\|\|F_t(x_*,t)\| \\
\leq\; & c_0\, L\, \|x - x_*\| + c_0 L\|x - x_*\|c_0\, c_0 = c_0\, L\, (c_0^2 + 1)\|x - x_*\| \;.
\end{aligned}$$

By definition of $F'(x_k, t, x'_k)$ in (14), we have

$$x'_k - x'_* = F_x^{-1}(x_k, t)F'(x_k, t, x'_k) - [F_x(x_k, t)^{-1}F_t(x_k, t) + x'_*] \ .$$

After taking norms and using the Lipschitz constant just derived, we get

$$\mu_k \leq \|F_x^{-1}(x_k, t)F'(x_k, t, x'_k)\| + c_0 \, L \, c_1 \rho_k / 2 \ .$$

The inverse $F_x^{-1}(x_k, t)$ in the first term on the right hand side can be replaced by P_k noting that by the Banach Perturbation Lemma [8] and the definition of D_k in Assumption 2

$$\|F_x^{-1}(x_k, t)P^{-1}\| = \|(I - D_k)^{-1}\| \leq 1/(1 - \|D_k\|) \ ,$$

which establishes the first assertion.

To prove the third inequality, we derive from (9) by taking norms

$$\begin{aligned} \|r'_k\| &= \|P_k F_x(x_k, t)x'_* + P_k F_t(x_k, t)\| \\ &\leq \|P_k[F_x(x_k, t) - F_x(x_*, t)]\|\|x'_*\| + \|P_k[F_t(x_k, t) - F_t(x_*), t)\| \\ &\leq \|P_k\|L(c_0^2 + 1)\rho_k \leq 2c_0 L(c_0^2 + 1)\rho_k = Lc_0 c_1 \rho_k \ . \end{aligned}$$

Here we have used that $\|x'_*\| \leq \|F_x^{-1}(x_*, t)\|\|F_t(x_*, t)\| \leq c_0^2$ by Assumption 1. The last inequality follows since $\|P_k\| = \|(I - D_k)F_x^{-1}\| \leq (1 + \delta)c_0$ as a consequence of Assumption 2. Finally we derive from (11)

$$\begin{aligned} \mu_{k+1} &\leq \delta_k \mu_k + \|r'_k\| + \|P'_k F(x_k, t)\| \\ &\leq \delta_k \mu_k + (Lc_0 c_1 + \|P'_k\| c_0)\rho_k \\ &\leq \delta_k \mu + c_0 \eta_k \ , \end{aligned}$$

where we have used that c_0 is a bound on the Jacobian F_x and hence a Lipschitz-constant for F, so that $\|F(x, t)\| = \|F(x, t) - F(x_*, t)\| \leq c_0 \rho_k$. ∎

The first equation of Lemma 1 provides us with a constructive stopping criterion for the derivative iteration, provided we can make some reasonable assumption regarding the sizes of L, c_0, and δ, which are also needed to bound $\|x_k - x_*\|$ in terms of $\|F(x_k, t)\|$ or $\|P_k F(x_k, t)\|$. The second inequality is the key to our convergence analysis in the following section.

3 Derivative Convergence for Q-linear Methods

First we will consider memory-less methods, where we may assume that $P_k = P(x_k, t)$ is continuously differentiable near (x, t) so that for some c_2 and all $\rho_k < \rho$

$$\|P'_k\| \ = \ \|P_x x'_k + P_t\| \ \leq \ c_2(\mu_k + 1) \quad . \tag{19}$$

This relation holds trivially with $c_2 = 0$ for simplified iteration (7), where $P'_k = 0$.

Proposition 1 *Under Assumptions 1 and 2, the condition (19) implies* ***R-linear or R-superlinear*** *convergence for the derivative recurrence (10). That is*

$$\overline{\lim}_k \ \|x'_k - x'_*\|^{1/k} \ \leq \ \delta_* . \tag{20}$$

Moreover, for all sufficiently small weights $\omega > 0$, the Sobolev norms

$$\|x_k - x_*\| + \omega\|x'_k - x'_*\|$$

converge ***Q-linearly*** *to zero. If furthermore $\delta_k \leq c\|x_k - x_*\|$, then we have* ***R-quadratic*** *convergence in that*

$$\overline{\lim}_k \, \|x'_k - x'_*\|^{1/2^k} \; < \; 1 \; ,$$

which applies for Newton's method, in particular.

Proof. Substituting (19) into the definition (16), we obtain

$$\eta_k \; \leq \; (Lc_1 + c_2)\rho_k + c_2\mu_k\rho_k \; ,$$

so that by (18)

$$\mu_{k+1} \; \leq \; (\delta_k + c_0c_2\rho_k)\mu_k + c_3\rho_k \; ,$$

where $c_3 = c_0(Lc_1 + c_2)$. Because of (12) and (13), we have by standard arguments

$$\rho_{k+1} \leq \delta_k\rho_k + Lc_0\rho_k^2.$$

Combining the last two inequalities for any ω, one obtains the ratio

$$\begin{aligned} \frac{(\rho_{k+1} + \omega\mu_{k+1})}{(\rho_k + \omega\mu_k)} \quad &\leq \quad \frac{(\delta_k + \omega c_3 + Lc_0\rho_k)\rho_k + \omega(\delta_k + c_0c_2\rho_k)\mu_k}{(\rho_k + \omega\mu_k)} \\ &\leq \quad \delta_k + \omega c_3 + c_0(Lc_0 + c_2)\rho_k \; . \end{aligned}$$

The last bound has a limit superior equal to $\delta_* + \omega c_3$, since we already know that the ρ_k converge to zero. This limiting ratio implies Q-linear convergence of the Sobolev norm, provided we chose $0 < \omega < (1 - \delta_*)/c_3$. Consequently, the linear R-factor of the sequence $\{\mu_k\}_k$ is less than or equal to any $\delta_* + c_3\omega$, and thus is not greater than δ_*,as asserted in (20). With the additional assumption on δ_k, we have for some c_4

$$\mu_{k+1} \leq c_4(\mu_k + 1)\rho_k \; ,$$

which means that the convergent sequence $\{\mu_k\}$ is bounded by a multiple of the Q-quadratically convergent sequence $\{\rho_{k-1}\}$. ∎

Proposition 1 shows that for memory-less contractions, the fully differentiated recurrence (10) yields R-linear convergence and potentially R-superlinear convergence, a possibility which can only occur if the iterates themselves converge superlinearly. The same convergence rates are achieved by the simplified derivative recurrence (7), even when the preconditioners are updated recursively and are not differentiable. In the important case of Newton's method, either derivative recurrence converges R- quadratically, which seems a rather satisfactory result.

Roughly speaking, we can claim in all these cases that the derivatives are converge satisfactorily whenever the iterates x_k converge in a reasonably rapid and stable fashion. The simplest condition under which the x_k, x'_k, and $\tilde{x}'_k$ must all converge linearly to their respective limits is that the shifted Jacobians $D_k \; = \; [I - P_k \, F_x(x_k, t)]$ converge to a limit whose spectral radius is less than one. This condition was implied by the hypothesis of Gilbert's theorem but must be considered quite restrictive. For example, the condition does not hold for Broyden's method nor for other popular quasi- Newton schemes, where $P_k = \alpha_k B_k^{-1}$. Here, α_k is a step multiplier, and B_k is an approximation to the inverse Jacobian $F_x(x_k, t)$, which is not guaranteed to converge to $F_x(f(x), t)$ or to any other limit. However, under the usual assumption for local convergence of secant updating

methods, it can be shown that $\alpha_k \to 1.0$ and that $\|D_k\| < 0.5$ in the l_2 norm for all k. Then it follows from Proposition 1 that the simplified recurrence (7) must converge to the unique limit x'_*. This does not necessarily apply in case of the fully differentiated recurrence (10) because *a priori* nothing is known about the existence or the size of the P'_k.

The differentiability of the secant updates is in question because they contain rank one terms of the form $y_k/\|s_k\|$, where both difference vectors

$$s_k \;\equiv\; x_{k+1} - x_k \quad \textit{and} \quad y_k \;\equiv\; F(x_{k+1},t) - F(x_k,t) \;\approx\; F_x(x_*,t)\,s_k$$

converge to zero. To prove that the matrix derivatives $\|P'_k\|$ do not blow up too fast, we make the observation that all classical updates and many other possible schemes can be written in the form

$$P_{k+1} \;=\; U(P_k, x_k, t, s_k, y_k) \quad , \tag{21}$$

where the update function

$$U \;:\; I\!R^{n\times n} \times I\!R^n \times I\!R \times I\!R^n \times I\!R^n \quad \mapsto \quad I\!R^{n\times n}$$

has the following property.

Assumption 3 (Lipschitzian Update) *There exist constants $c \geq 1$, $\rho < \infty$, $\delta < 1$, and $\gamma < \infty$ such that the domain conditions*

$$\|P\| \;,\; \|P^{-1}\| < c \;,\; \|x - x_*\| \;,\; \|s\| < \rho \;, \quad \textit{and} \quad \|Py - s\| < \delta\|s\| \tag{22}$$

imply that U is differentiable at the point (P, x, t, s, y), and its partial derivatives satisfy

$$\|U_P\| \;,\; \|U_x\| \;,\; \|U_t\| \;\leq \gamma \;, \quad \textit{and} \quad \|U_s\| \;,\; \|U_y\| \leq \gamma/\|s\| \;, \tag{23}$$

where P may be restricted to the open cone of symmetric positive definite matrices in $I\!R^{n\times n}$.

The crucial point here is that the partial derivatives with respect to s and y are only bounded by a multiple of the reciprocal step size $1/\|s\|$, which allows unbounded growth of the matrix derivatives $\|P'_k\|$. The key observation of the following proof is that the Q-superlinear convergence rate

$$\lim_k \|x_{k+1} - x_*\|/\|x_k - x_*\| \;=\; 0 \tag{24}$$

implies that the residuals $\|F_k\|$ decline just a bit faster than the $\|P'_k\|$ may grow. Before we formulate the second major result, let us briefly show that the Broyden update and the DFP formula which do not explicitly depend on (x, t) satisfy the condition above.

Lemma 2 *The Broyden update function*

$$U(P,s,y) \;=\; P + \frac{(s - Py)s^T P}{s^T P y}$$

and the Davidon-Fletcher-Powell (?) formula

$$U(P,s,y) \;=\; P - \frac{Pyy^T P}{y^T P y} + \frac{ss^T}{y^T s}$$

satisfy Assumption 3 with all norms $\|\cdot\|$ induced by the Euclidean vector norm.

Proof. For the nonsymmetric Broyden update, ρ is arbitrary, and δ may be any number between zero and 1. Then we derive from the last domain condition in Assumption 3 that $s \neq 0$ and that

$$\|y\| = \|P^{-1}Py\| \leq c\|Py\| \leq c(1+\delta)\|s\| < 2c\|s\|$$

as well as

$$\|s\|\|y\|c \geq \|s\|\|Py\| \geq s^T Py = s^T(Py - s) + s^T s \geq (1-\delta)\ |s\|^2 \ .$$

In particular, $\|y\| \geq \|s\|(1-\delta)/c$. Now let $P(\tau) \equiv P + \tau\dot{P}$, and compute the derivative $\dot{U}$ of $U(P(\tau), s, y)$ at $\tau = 0$. Then we have by the chain rule with s and y kept constant

$$\dot{U} = \dot{P} - \dot{P}ys^T P + (s - Py)s^T\dot{P}\Big/(s^T Py) - (s - Py)s^T P(s^T\dot{P}y)/(s^T Py)^2 \ ,$$

so that by the triangle inequality in the L_2 norm

$$\begin{aligned}\|\dot{U}\| &\leq \|\dot{P}\| \cdot [1 + (\|y\|\|P^T s\| + \|s\|^2 + \|s\|\|Py\|)/(s^T Py) \\ &\qquad + (\|s\| + \|Py\|)\|P^T s\|\|s\|\|y\|/(s^T Py)^2] \\ &\leq \|\dot{P}\| \cdot [1 + (2c^2 + 1 + 2c^2)/(1-\delta) + (1 + 2c^2)2c^2/(1-\delta)^2] \ .\end{aligned}$$

Since the direction $\dot{P}$ is arbitrary, this shows that the derivative U_P is uniformly bounded as required. Similarly, we find for the differentiation in some direction $\dot{s}$

$$\begin{aligned}\|\dot{U}\| &\leq \|\dot{s}\| \cdot [(\|P^T s\| + \|s - Py\|\|P\|)/(s^T Py) + (\|s\| + \|Py\|)\|P^T s\|\|Py\|/(s^T Py)^2] \\ &\leq (\|\dot{s}\|/\|s\|) \cdot [(1 + 1 + 2c^2)c/(1-\delta) + (2 + 2c^2)2c^3)/(1-\delta)^2] \ ,\end{aligned}$$

which implies that $U_s\|s\|$ is indeed uniformly bounded. Finally, we derive in the direction $\dot{y}$

$$\begin{aligned}\|\dot{U}\| &\leq \|\dot{y}\| \cdot [\|P\|\|P^T s\|/(s^T Py) + (\|s\| + \|Py\|)\|P^T s\|\|P^T s\|/(s^T Py)^2] \\ &\leq (\|\dot{y}\|/\|s\|) \cdot [c^2/(1-\delta) + (c + 2c^2)c^2/(1-\delta)^2] \ ,\end{aligned}$$

which ensures that $U_y\|s\|$ is indeed uniformly bounded.

For the DFP formula, we must impose the restriction $\delta < 0.2\, c^{-2}$. Then we have

$$y^T s = y^T PP^{-1}s \geq s^T P^{-1}s - \|Py - s\|\|P^{-1}\|\|s\| \geq (1/c - c\delta)\|s\|^2 \geq 0.8\|s\|^2/c \ ,$$

where we have used the assumed positive definiteness of P to bound

$$s^T P^{-1}s \geq \|s\|^2/\|P\| \geq \|s\|^2/c \ .$$

As an immediate consequence, we have

$$y^T Py \geq y^T s - y^T(s - Py) \geq 0.8\|s\|^2/c - \delta\|y\|\|s\| \geq (0.8/c - \delta 2c)\|s\|^2 \geq 0.4\|s\|^2/c \ .$$

The rest of the argument is almost the same as the Broyden update. We find by differentiating in some direction $\dot{P}$ with s and y held constant

$$\dot{U} = \dot{P} - [\dot{P}yy^T P + Pyy^T\dot{P}]/(y^T Py) - [Pyy^T P](y^T\dot{P}y)/(y^T Py)^2 \ ,$$

so that after taking norms

$$\begin{aligned}\|\dot{U}\| &\leq \|\dot{P}\|\ [1 + 2\|y\|\|Py\|2.5c/\|s\|^2 + \|Py\|^2\|y\|^2 6.25c^2/\|s\|^4] \\ &\leq \|\dot{P}\|\ [1 + 20c^4 + 100c^8] \leq \|\dot{P}\|(1 + 10c_0^4)^2 \ .\end{aligned}$$

The derivatives with respect to y and s can be bounded by multiples of $\|s\|^{-1}$ in exactly the same fashion. ∎

Since Assumption 3 can also be verified for the BFGS update, it applies for a wide range of methods. Now we obtain for these updating methods almost the same result as in the memory-less case. The rather stringent restriction $\delta \leq 0.2c^{-2}$ used in the proof for the DFP formula could be avoided if other conditions were placed on $y^T s$ and $y^T P y$. This would make perfect sense in the context of convex optimization, but we did not introduce them here because of our primary focus is on the nonlinear equations case.

Proposition 2 *Under Assumptions 1, 2, and 3 with ρ and δ sufficiently small, the fully differentiated recurrence (10) yields R-linear or R-superlinear derivative convergence:*

$$\overline{\lim}_k \ \|x'_k - x'_*\|^{1/k} \ \leq \ \delta_* \ .$$

Moreover,

$$\overline{\lim}_k \ [\|P'_k\|\|x_k - x_*\|]^{\frac{1}{k}} \leq \delta_* \ ,$$

which limits the potential growth the P'_k relative to the decline of the errors $\|x_k - x_*\|$.

Proof. Differentiating (21), we obtain by the chain rule and the triangular inequality using (23)

$$\begin{aligned} \tfrac{1}{\gamma}\|P'_{k+1}\| &\leq \tfrac{1}{\gamma}\|U_P\, P'_k + U_x\, x'_k + U_t + U_s\, s'_k + U_y\, y'_k\| \\ &\leq \|P'_k\| + \mu_k + \|x'_*\| + 1 + (\|s'_k\| + \|y'_k\|)/\|s_k\| \ . \end{aligned}$$

To bound the last two terms, we note that by (18) of Lemma 3

$$\begin{aligned} \|s'_k\| &= \|x'_{k+1} - x'_k\| \ \leq \ \mu_{k+1} + \mu_k \\ &\leq (1+\delta)\mu_k + c_0\eta_k \ \leq \ 2\mu_k + c_0\eta_k \ . \end{aligned}$$

Similarly, we find

$$\begin{aligned} & \|y'_k\| = \|F'(x_{k+1}, t, x'_{k+1}) - F'(x_k, t, x'_k)\| \\ \leq \ & \|F_x(x_{k+1}, t)x'_{k+1} - F_x(x_k, t)x'_k\| + \|F_t(x_{k+1}, t) - F_t(x_k, t)\| \\ \leq \ & \|F_x(x_{k+1}, t)(x'_{k+1} - x'_*)\| + \|F_x(x_k, t)(x'_k - x'_*)\| \\ & \quad + \|[F_x(x_{k+1}, t) - F_x(x_{k+1}, t)]x'_*\| + L(\rho_{k+1} + \rho_k) \\ \leq \ & c_0(\mu_{k+1} + \mu_k) + (c_0^2 + 1)L(\rho_{k+1} + \rho_k) \\ \leq \ & 2c_0\mu_k + c_0^2\eta_k + \eta_k \leq (c_0^2 + 1)(\mu_k + \eta_k) \ . \end{aligned}$$

Adding the last two inequalities and noting that $\|s_k\| \geq \rho_k - \rho_{k+1} \geq 0.9(1-\delta)\rho_k$, we find that for some c_5,

$$(\|s'_k\| + \|y'_k\|)/\|s_k\| \ \leq \ c_5(\mu_k + \eta_k)/\rho_k \ .$$

Now, since ρ_k is bounded, and η_k/ρ_k is bounded away from zero, the first four terms in (25), and an additional Lc_1 can be subsumed into the last bound, with c_5 growing to some c_6 so that

$$Lc_1 + \|P'_{k+1}\| \leq c_6(\mu_k + \eta_k)/\rho_k \quad .$$

After multiplication by ρ_{k+1}, we get

$$\eta_{k+1} \leq q_k(\mu_k + \eta_k) \quad with \quad q_k \equiv c_6\rho_{k+1}/\rho_k \to 0 \ .$$

Adding $\omega > \delta_*$ times this inequality to the bound (18), we find that

$$\frac{(c_0\eta_{k+1} + \omega\mu_{k+1})}{(c_0\eta_k + \omega\mu_k)} \;\leq\; \frac{c_0(q_k + \omega)\eta_k + (c_0 q_k + \omega\delta_k)\mu_k}{(c_0\eta_k + \omega\mu_k)} \;\leq\; \max\left\{\delta_k + \frac{c_0 q_k}{\omega},\; q_k + \omega\right\} .$$

Since the limit superior of the maximum is ω, and one may choose ω arbitrarily close to δ_*, we have shown that the sequences $\{\eta_k\}_k$ and $\{\mu_k\}_k$ both have a linear R-factor no greater than δ_*. The last assertion follows directly from the definition of η_k in (16). ∎

This result applies to all standard classical secant methods and suggests that the rate at which the derivatives x'_k converge is the same whether or not the Jacobian updating procedure is differentiated. That this conclusion is only valid if the line- search eventually becomes inactive, so that all later steps are of unit length. On one hand, this means that the fully differentiated, or black box, approach is reasonably safe. On the other hand, it appears that implicitly defined derivatives can be obtained at a much reduced cost by deactivating the P_k, i.e. treating them as constants as in the simplified updating scheme. Also, the theoretical possibility that the P'_k generated in the fully differentiated update may grow unbounded is numerically worrisome as it may lead to exponent overflows.

4 Numerical Results and Higher Derivative Recurrences

Our very limited numerical experience confirm the theoretical results. We found only a moderate growth of the P'_k for our test case, the Davidon-Fletcher-Powell (DFP) secant method. However, there is clear evidence that the convergence of the first derivatives x'_k lags significantly behind the convergence of the iterates x_k themselves. In the case of secant methods this phenomenon is much more pronounced than for Newton's method, where the $d-$th derivative can be shown to lag roughly d steps behind the functional iterate. We have also propagated higher derivatives for secant methods and found that they converge in a staggered fashion and at about the same rate whether or not P_k is deactivated.

Our numerical experiments were conducted on the test function

$$F(x,t) \;\equiv \nabla_x f(x,t) \quad with \quad f(x,t) \;\equiv\; \frac{1}{2}\left(x^T H x + t\|x\|^4\right) ,$$

where $H = [1/(i+j-1)]$ is the Hilbert matrix of order n, and $\|x\|$ denotes the Euclidean norm. Since the unique solution $x_* = 0$ is independent of the parameter t, all derivatives $x'_*, x''_*, \ldots x^{(j)}_*$ must also vanish, which makes monitoring their errors exceedingly simple. The approximate inverse Hessian was initialized as $P_0 = \mathrm{diag}(i)_{i=1,\ldots,n}$, which is somewhat "smaller" than the exact inverse H^{-1}. Consequently, the inverse form of the DFP update takes a very long time before P_k and the resulting steps $s_k = -P_k F(x_k,t)$ become large enough to achieve superlinear convergence. The starting point was always the vector of ones $x_0 = e$, and the parameter was set to $t = 1$.

Andreas, please check that I translated correctly. The text said: Here xn, gn and del represent the Euclidean (Frobenius) norms of $x_k - x_* = x_k$, **???** $F(x_k,t)$ and $\tilde{D}_k = I - P_k H$, respectively. Note that $\tilde{D}_k$ is not exactly equal to D_k since we have neglected the nonquadratic term. The quantities xpn, gpn, and Apn represent the Euclidean norms of the derivatives x'_k, $F'(x_k,t,x'_k)$ and P'_k. I translated that as:

$\|x_k - x_*\|$	$\|x'_k - x'_*\|$	$\|F(x_k,t)\|$	$\|F'(x_k,t,x'_k)\|$	$\|\tilde{D}_k = I - P_k H\|$	P'_k
xn	xpn	gn	gpn	del	Apn

In this version, I leave these as tables. Work is in progress to make graphs.
First let us consider the fully differentiated iteration without line search for $n = 2$.

$\|x_k - x_*\|$	$\|x'_k - x'_*\|$	$\|F(x_k, t)\|$	$\|F'(x_k, t, x'_k)\|$	$\|\tilde{D}_k = I - P_k H\|$	P'_k
1.41421	0	7.32196	5.65685	1.20185	0
5.91138	5.65685	420.315	1602.43	1.37905	.122897
1.28415	0.322443	5.81274	7.4073	1.37919	.126535
1.23993	0.530983	5.07422	8.79846	1.35374	.353744
1.14152	0.906449	4.05979	9.46657	1.28473	.451704
0.775852	1.39968	1.50164	2.30642	1.24288	.523219
0.583841	1.49961	0.707141	1.90673	1.17944	.352458
0.440021	1.24741	0.275981	1.30797	1.18538	.489282
0.364655	0.93575	0.124256	0.81024	1.32074	1.69699
0.317938	0.668433	0.090391	0.528703	1.51918	2.94176
0.273088	0.362568	0.080089	0.388196	1.51109	3.20875
0.218102	0.151102	0.072478	0.20315	1.30079	5.73189
0.157636	0.304182	0.061558	0.070792	1.11790	8.48362
0.089584	0.502993	0.044328	0.231856	1.61169	10.3928
0.025795	0.439793	0.023258	0.329356	2.56209	16.5102
0.019691	0.222841	0.008383	0.196591	2.73134	24.3488
0.021733	0.342188	0.004685	0.097266	1.06296	28.8397
0.019112	0.357521	0.001751	0.068149	1.72589	23.3437
0.013577	0.263456	0.001214	0.017659	2.71993	85.2334
0.007314	0.201841	0.001688	0.023696	1.62204	37.3540
0.002504	0.075257	0.001454	0.027388	1.07995	28.0788
0.000738	0.022724	0.000829	0.021418	1.98608	41.5908
0.000852	0.015869	0.00029	0.008764	0.638348	16.6442
0.000429	0.012485	4.22336e-05	0.002265	1.00812	39.2452
9.91147e-05	0.003677	1.42962e-05	0.000317	0.515883	24.8241
7.86743e-06	0.00047	4.41258e-06	0.000169	0.338774	8.76305
5.6655e-07	1.87742e-05	4.20846e-07	2.09424e-05	0.193667	2.94516
6.65702e-08	3.48841e-06	1.13056e-08	8.30216e-07	0.093476	5.80960
1.89371e-09	1.24304e-07	1.95901e-10	9.38562e-09	0.054466	2.90143
1.39032e-11	1.24812e-09	6.57715e-12	4.52763e-10	0.024018	0.641633

As we can see, the convergence is pretty sloppy.

Here are the results for the simplified operation.

$\|x_k - x_*\|$	$\|x'_k - x'_*\|$	$\|F(x_k, t)\|$	$\|F'(x_k, t, x'_k)\|$	$\|\tilde{D}_k = I - P_k H\|$	P'_k
1.41421	0	7.32196	5.65685	1.20185	0
5.91138	5.65685	420.315	1602.43	1.37905	0
1.28415	41.5058	5.81274	259.676	1.37919	0
1.23993	79.8816	5.07422	270.158	1.35374	0
1.14152	3.55943	4.05979	10.3051	1.28473	0
0.775852	1.41966	1.50164	2.83181	1.24288	0
0.583841	0.962748	0.707141	1.24517	1.17944	0
0.440021	0.732391	0.275981	0.663272	1.18538	0
0.364655	0.596066	0.124256	0.443747	1.32074	0
0.317938	0.499531	0.090391	0.320956	1.51918	0
0.273088	0.39534	0.080089	0.25424	1.51109	0
0.218102	0.280469	0.072478	0.209113	1.30079	0
0.157636	0.185415	0.061558	0.167374	1.11790	0
0.089584	0.102318	0.044328	0.123206	1.61169	0
0.025795	0.086137	0.023258	0.077214	2.56209	0
0.019691	0.128427	0.008383	0.04098	2.73134	0
0.021733	0.123338	0.004685	0.027106	1.06296	0
0.019112	0.107983	0.001751	0.010115	1.72589	0
0.013577	0.076062	0.001214	0.006931	2.71993	0
0.007314	0.040747	0.001688	0.009568	1.62204	0
0.002504	0.014017	0.001454	0.008199	1.07995	0
0.000738	0.004168	0.000829	0.00468	1.98608	0
0.000852	0.004811	0.00029	0.00164	.638348	0
0.000429	0.002422	4.22336e-05	0.000238	1.00812	0
9.91147e-05	0.000559	1.42962e-05	8.07039e-05	.515883	0
7.86743e-06	4.44117e-05	4.41258e-06	2.49092e-05	.338774	0
5.6655e-07	3.1982e-06	4.20846e-07	2.3757e-06	.193667	0
6.65702e-08	3.75792e-07	1.13056e-08	6.38209e-08	.093476	0
1.89371e-09	1.06901e-08	1.95901e-10	1.10587e-09	.054466	0
1.39032e-11	7.8484e-11	6.57715e-12	3.71284e-11	.024018	0

Suppose we introduce a line search that performs exactly one parabolic interpolation on the function value at each step. Then we get the much more rapid convergence pattern, even when

$n = 3$

$\|x_k - x_*\|$	$\|x'_k - x'_*\|$	$\|F(x_k,t)\|$	$\|F'(x_k,t,x'_k)\|$	$\|\tilde{D}_k = I - P_k H\|$	P'_k
1.73205	0.000000	12.5518	10.3923	1.66700	0.000000
1.67068	0.134673	11.4076	11.7444	1.65588	0.061689
0.882142	0.160312	2.07124	2.12777	1.60236	0.122752
0.546038	0.505207	1.09251	1.76371	1.57634	0.550104
0.189388	0.526785	0.215538	0.670052	1.41861	0.751283
0.132772	0.467438	0.020972	0.118908	1.45088	0.76126
0.048901	0.340127	0.011613	0.075559	1.34129	14.9435
0.026949	0.058005	0.002710	0.030825	1.28867	2.31277
0.026500	0.052905	0.000121	0.001371	1.18469	5.69642
0.026237	0.059376	0.000119	0.001903	1.27383	124.353
0.010553	0.046969	0.000037	0.000273	0.671322	487.070
0.000444	0.006700	0.000005	0.000337	0.315553	233.346
0.00001	0.001317	0.000001	0.000032	0.904784	935.704

Here the P'_k actually show signs of blowing up, at least temporarily. Note that Proposition Again **????** we get pretty much the same for the simplified iteration.

$\|x_k - x_*\|$	$\|x'_k - x'_*\|$	$\|F(x_k,t)\|$	$\|F'(x_k,t,x'_k)\|$	$\|\tilde{D}_k = I - P_k H\|$	P'_k
1.73205	0	12.5518	10.3923	1.667	0
1.67068	0.050909	11.4076	8.41106	1.65588	0
0.882142	0.914877	2.07124	3.35272	1.60236	0
0.546038	1.14635	1.09251	1.81314	1.57634	0
0.189388	0.260913	0.215538	0.081441	1.41861	0
0.132772	0.1987	0.020972	0.041101	1.45088	0
0.048901	0.03018	0.011613	0.026022	1.34129	0
0.026949	0.008106	0.00271	0.001544	1.28867	0
0.0265	0.005859	0.000121	0.000594	1.18469	0
0.026237	0.00603	0.000119	0.000737	1.27383	0
0.010553	0.527305	3.07888e-05	.107082	0.671322	0
0.000444	0.169617	5.43704e-06	.047119	0.315553	0
1.01274e-05	.014605	4.55601e-07	.002665	0.904784	0
2.14482e-08	.005679	1.31503e-08	.000455	0.010328	0

This seems to work significantly worse than the fully differentiated formula.

We can also do higher derivatives! For the simplified recurrence, where P_k is deactivated, the following informal argument establishes the convergence of the higher derivatives $x^{(j)} \equiv d^j x(t)/dt^j$. Differentiating equation (6) $j < m$ times with respect to t, we obtain the following linear system for the $(j+1)$-st derivative from Leibnitz's rule:

$$P_k \, F_x(x(t),t) \, x^{(j+1)} = -P_k \left(\frac{\partial^j}{\partial t^j} \left[F_t(x(t),t) \right] + \sum_{i=1}^{j} \binom{j}{i} \frac{\partial^{j-i}}{\partial t^{j-i}} \left[F_x(x(t),t) \right] x^{(i)}(t) \right) . \tag{25}$$

Here we have assumed that $F(x,t)$ is m times jointly Lipschitz-continuously differentiable. Replacing the $x^{(i)}(t)$ by approximations $\tilde{x}_k^{(i)}$ for $i = 0, 1, \ldots, j$, one may interpret the right hand side as a vector function

$$-P_k R_k^{(j)} \equiv -P_k R_k^{(j)} \left(t, x_k, \tilde{x}'_k, \ldots, \tilde{x}_k^{(j)} \right) .$$

While this may seem a very messy expression, the residual vectors

$$F_x(x_k(t),t)\,x_k^{(i+1)} \;+\; R_k(i), \quad for\ i = 0,1,\ldots,j$$

can be evaluated simultaneously for any given t and $(\tilde{x}_k^{(i)})_{i=0,1,\ldots,j+1}$ by one forward sweep of automatic differentiation [2]. The complexity of this Taylor series propagation is $\mathcal{O}(j^2)$ times that of one function evaluation $F(x,t)$ if ordinary polynomial arithmetic is used. This asymptotic complexity bound can be reduced to $\mathcal{O}(j \log j)$ through the use of the fast Fourier transform, but that is only likely to pay off when j is significantly larger than 10. As a generalization of (7), one may now iterate for $j = 0,1,\ldots,m-1$ and $k = 0,1,\ldots$

$$\tilde{x}_{k+1}^{(j+1)} = \tilde{x}_k^{(j+1)} \;-\; P_k \left[F_x(x_k(t),t)\,\tilde{x}_k^{(j+1)} \;+\; R_k^{(j)} \right] .$$

This family of linear recurrences is again of the form (7) with the same leading linear term. By induction, one sees that if all $\tilde{x}_k^{(i)}$ for $i < j$ converge to the correct values $x_*^{(i)}$, then the $R_k^{(j)}$ converge to the right hand side of (25), and the $\tilde{x}_k^{(j+1)}$ can only converge to the unique fixed point $x_*^{(j+1)}$ of its recurrence. The linear R-factor is again at least δ_*, but the higher derivatives tend to converge in a staggered fashion. This can be seen from the following numerical result listing the first five derivatives of the residual vector $\nabla f(x,t)$ for the fully differentiated case with $n = 3$ and the

parabolic line-search.

7.83333	6	0	0	0	0
7.08333	6	0	0	0	0
6.78333	6	0	0	0	0
twas the gprimes					
7.13467	6.85996	-0.962535	0.98677	-0.916888	0.743149
6.43292	6.76028	-0.812404	0.79177	-0.687363	0.493916
6.15208	6.72086	-0.753236	0.715227	-0.597693	0.397123
twas the gprimes					
0.886147	0.672907	-0.189995	0.217294	-0.24017	0.255908
1.25420	1.29606	-0.117537	0.134092	-0.143875	0.145163
1.38987	1.54753	-0.090792	0.103094	-0.10765	0.103067
twas the gprimes					
0.891133	1.51001	0.164555	-0.195075	-0.941908	0.018713
0.509296	0.764193	0.021875	0.168406	-0.060258	-0.16829
0.374261	0.496541	-0.061992	0.306359	0.276555	-0.224844
twas the gprimes					
0.164546	0.478306	0.089422	-0.085411	-0.59197	-1.26069
0.109936	0.361055	0.299476	0.056043	-0.759126	-1.79916
0.085412	0.29972	0.345063	0.144632	-0.762659	-2.04218
twas the gprimes					
-0.008084	-0.022728	-0.027573	0.088714	0.19331	1.02341
0.012005	0.072649	0.18933	0.069657	-0.562832	-1.28511
0.015177	0.09135	0.257746	0.160079	-0.719024	-2.32696
twas the gprimes					
-0.01141	-0.074535	-0.022343	0.303165	-1.94998	-0.275344
-0.002153	-0.00544	0.06202	0.015314	-1.02746	1.79132
0.000161	0.011139	0.088279	0.011372	-0.747106	1.49802
twas the gprimes					
-0.002072	-0.025347	-0.009128	0.636675	-1.65201	-17.6312
-0.001456	-0.014561	0.026911	0.507209	-1.23151	-10.3558
-0.000964	-0.009782	0.030917	0.401505	-0.97309	-7.31494
twas the gprimes					
6.61736e-05	0.001204	0.005172	-0.155942	-2.27729	-8.48942
-7.21583e-05	-0.000599	-0.010462	-0.137604	-0.792678	1.88576
7.03952e-05	-0.000268	-0.010349	-0.117287	-0.404242	3.53261
twas the gprimes					
-3.33856e-05	0.001891	-0.033072	0.470508	-4.84217	29.4543
-6.47562e-05	-9.33634e-05	-0.003983	0.06335	-0.659946	4.03583
9.38961e-05	0.000187	0.001207	0.022922	-0.684854	12.1638
twas the gprimes					
3.01871e-06	0.00024	-0.0079	0.301536	-10.2752	286.043
-2.06109e-05	-9.00923e-05	1.74332e-05	0.000638	-0.598232	19.6354
2.26722e-05	9.44013e-05	0.000976	-0.033317	0.746131	-21.3334
twas the gprimes					
4.74285e-06	-0.000331	0.007024	0.000395	-6.43432	274.870
-2.22328e-06	-2.8461e-05	0.000697	-0.034716	0.751035	2.71493
-1.45734e-06	5.71076e-05	-0.000316	-0.039082	1.98018	-44.6845
twas the gprimes					
4.3987e-07	-2.90034e-05	0.000314 17	0.064775	-3.00928	10.3549
1.00937e-09	1.87743e-06	0.000254	-0.002493	-0.643611	31.7758
-1.18683e-07	1.24001e-05	-1.18975e-05	-0.016062	0.419384	10.9830
twas the gprimes					

Now we repeat the same calculation but this time with the P_k deactivated. same first components as before left off here

6	0	0	0	0
6	0	0	0	0
6	0	0	0	0
twas the gprimes				
4.82302	-0.476248	0.014832	-0.000152	0
4.86433	-0.477479	0.014851	-0.000152	0
4.88085	-0.477971	0.014858	-0.000152	0
twas the gprimes				
-0.74458	-0.149644	0.243524	-0.033263	0.001018
-2.04741	0.322835	1.43921	-1.08683	0.247745
-2.54841	0.50425	1.91978	-1.50616	0.345771
twas the gprimes				
-1.74185	1.88132	-0.576298	-5.40806	12.0342
-0.498205	0.380387	0.224032	-1.65358	-1.23878
-0.072187	-0.148079	0.573789	-0.360719	-6.28715
twas the gprimes				
0.080149	-0.796362	0.726776	2.13653	-3.35262
0.000382	-0.235109	-0.26597	1.91018	-2.05038
-0.014447	-0.07221	-0.52796	1.72746	-1.75125
twas the gprimes				
0.015076	-0.082498	0.363468	-0.195027	-2.55240
-0.024026	0.102362	-0.240598	0.028066	0.487509
-0.029745	0.141692	-0.393884	0.169606	0.878855
twas the gprimes				
0.023012	-0.026744	-0.646409	0.854433	8.47370
0.010335	-0.019941	-0.280996	0.081682	5.44524
0.006388	-0.014289	-0.171545	-0.105063	4.17312
twas the gprimes				
0.00153	0.001505	-0.044115	0.170653	-0.119117
0.000189	0.002171	-0.009163	0.153992	-0.664591
-7.9969e-05	0.003118	-0.004485	0.121157	-0.631163
twas the gprimes				
-0.000589	0.00128	0.013065	0.020891	-0.373877
-5.30096e-05	-0.000384	0.002598	0.00451	-0.05193
5.60423e-05	0.000492	-0.002736	-0.005069	0.046773
twas the gprimes				
0.000731	-0.001247	-0.017088	-0.027953	0.472549
6.89644e-05	-0.00063	-0.000428	0.000175	0.037797
-6.17383e-05	0.000673	-0.00024	-0.000497	-0.01686
twas the gprimes				
-0.106291	0.15094	2.38140	4.53559	-56.3123
-0.010134	0.017547	0.462909	0.184754	-22.0105
0.008124	0.00277	-0.130307	-0.720322	-1.09736
twas the gprimes				
0.046164	-0.070342	-0.905865	-2.1639	16.6408
0.009385	-0.011191	-0.495743	-0.000487	23.6636
0.000995	-0.008303	-0.174044	0.28883	12.1396
twas the gprimes				
-0.001048	0.001722	0.068507	-0.056817	-4.47033
-0.001811	0.002815	0.056185	-0.027472	-1.00815
-0.00165	0.002772	0.043859	0.204671	-1.29037
twas the gprimes				
0.000322	-0.000601	-0.011861	-0.086866	0.875113

Now let's look at the fully differentiated two-dimensional case without line-search

5.5	4	0	0	0	0
4.83333	4	0	0	0	0
twas the gprimes					
-320.917	-1200.06	-1700.89	-1077.33	-256	0
-271.435	-1061.91	-1569.33	-1034.67	-256	0
twas the gprimes					
4.67786	5.44726	-1.45861	1.11283	-0.395316	-0.623963
3.45045	5.01951	-0.726664	0.177179	0.558952	-1.34465
twas the gprimes					
3.26335	4.44532	-0.889732	-0.053159	0.989838	-1.51802
3.88565	7.59289	0.031227	-1.92695	3.24290	-3.28712
twas the gprimes					
2.47489	6.34817	0.753541	-5.97338	10.1578	-5.59358
3.21819	7.02258	-4.10746	1.08170	12.7960	-42.4326
twas the gprimes					
0.827821	2.23056	5.32941	11.3574	-281.202	-3302.63
1.25285	-0.586674	-6.81129	162.758	823.975	-1904.46
twas the gprimes					
0.356472	1.83922	3.21208	-33.3711	-346.410	-1103.89
0.610718	-0.502897	-2.07225	107.693	303.251	-3036.61
twas the gprimes					
0.084493	1.21808	2.10769	-43.2238	-285.768	553.030
0.262729	-0.476498	-0.920708	66.7940	181.473	-2161.95
twas the gprimes					
-0.031288	0.733451	1.67876	-35.5955	-215.420	776.125
0.120252	-0.344295	-1.19784	40.9095	185.144	-1269.87
twas the gprimes					
-0.067232	0.512487	1.50108	-36.0887	-206.375	1449.61
0.060419	-0.12994	-1.78263	15.5821	205.082	202.420
twas the gprimes					
-0.076023	0.377545	2.81489	-26.4187	-352.445	-242.384
0.025196	0.090309	-0.753557	-6.14754	19.6118	956.143
twas the gprimes					
-0.072478	0.171229	3.15421	-12.3705	-342.268	-1014.17
7.81512e-05	0.109319	0.03917	-10.9744	-44.7461	1129.44
twas the gprimes					
-0.060281	-0.040077	3.26064	7.17298	-292.524	-2593.32
-0.012473	0.058355	1.13385	-6.18375	-159.940	-120.768
twas the gprimes					
-0.04182	-0.224528	2.61024	28.2915	-175.017	-4286.53
-0.014698	-0.057833	1.43948	7.78921	-165.436	-1792.90
twas the gprimes					
-0.021053	-0.30123	0.430609	42.0019	203.048	-4791.62
-0.009885	-0.133176	0.547738	19.9491	34.0736	-2744.82
twas the gprimes					
-0.006728	-0.167877	-1.22758	12.7669	395.634	2268.76
-0.005002	-0.102301	-0.3233	10.7935	171.086	51.4143
twas the gprimes					
-0.003258	-0.07252	-0.439088	4.29992	113.951	663.303
-0.003366	-0.064819	-0.172193	5.57870	80.0809	164.661
twas the gprimes					
-0.000465	-0.044861	-0.752186	4.57146	190.366	281.806
-0.001688	-0.051301	-0.371678	6.53508	115.538	-327.682

Here we have the simplified derivative recurrence on the same problem.

5.5	4	0	0	0	0
4.83333	4	0	0	0	0
twas the gprimes					
-320.917	-1200.06	-1700.89	-1077.33	-256	0
-271.435	-1061.91	-1569.33	-1034.67	-256	0
twas the gprimes					
4.67786	105.528	2936.08	-5961.48	-208775	-1.14961e+06
3.45045	237.267	8009.74	182950	1.31217e+06	4.88335e+06
twas the gprimes					
3.26335	219.285	11658.5	1.37659e+06	1.375e+08	1.02074e+10
3.88565	-157.795	-1809.82	555769	9.64855e+06	-4.95981e+09
twas the gprimes					
2.47489	8.83482	-3272.66	-487306	-7.3468e+07	-6.89382e+09
3.21819	-5.30480	-10171.1	-256851	5.45768e+07	5.90549e+09
twas the gprimes					
0.827821	0.756555	249.520	-26979.6	-1.00104e+07	-1.13448e+09
1.25285	-2.72888	-1564.18	5885.69	1.45899e+07	2.13254e+09
twas the gprimes					
0.356472	0.499834	277.224	-10109.6	-5.27303e+06	-6.78685e+08
0.610718	-1.14045	-530.066	4464.79	5.64325e+06	8.64373e+08
twas the gprimes					
0.084493	0.444293	240.769	-5263.44	-3.53796e+06	-4.98388e+08
0.262729	-0.492477	-200.729	2868.58	2.48968e+06	3.7449e+08
twas the gprimes					
-0.031288	0.357054	175.593	-3243.15	-2.40779e+06	-3.54419e+08
0.120252	-0.263484	-107.532	1804.12	1.41712e+06	2.11234e+08
twas the gprimes					
-0.067232	0.285099	130.544	-2267.58	-1.74953e+06	-2.62736e+08
0.060419	-0.147414	-61.0968	1036.68	811373	1.22627e+08
twas the gprimes					
-0.076023	0.251072	110.741	-1868.17	-1.46968e+06	-2.2365e+08
0.025196	-0.040014	-14.2995	197.947	178782	2.96799e+07
twas the gprimes					
-0.072478	0.206519	88.5443	-1451.07	-1.16432e+06	-1.79635e+08
7.81512e-05	0.032841	17.0677	-315.289	-233430	-3.3957e+07
twas the gprimes					
-0.060281	0.157502	66.2374	-1072.01	-867589	-1.34634e+08
-0.012473	0.056631	26.1979	-440.133	-347220	-5.29505e+07
twas the gprimes					
-0.04182	0.110328	46.5033	-753.296	-609283	-9.45131e+07
-0.014698	0.054841	24.6963	-405.935	-325045	-5.00814e+07
twas the gprimes					
-0.021053	0.066038	28.8677	-471.733	-379256	-5.85937e+07
-0.009885	0.040013	18.2375	-299.520	-239968	-3.69868e+07
twas the gprimes					
-0.006728	0.031365	14.5621	-238.954	-191555	-2.95374e+07
-0.005002	0.026374	12.4127	-206.538	-163995	-2.51191e+07
twas the gprimes					
-0.003258	0.018817	8.93589	-149.884	-118359	-1.80585e+07
-0.003366	0.01951	9.26797	-155.548	-122783	-1.87286e+07
twas the gprimes					
-0.000465	0.002712	1.28907	-21.7026	-17094.8	-2.60365e+06
-0.001688	0.009745	4.62797	-77.1399	-61178.8	-9.36361e+06

Hopefully we have we learned from these numerical examples, bla, bla, bla.

5 Convergence Results for Multi-Step Contractions

Unfortunately, there are many other methods of great practical importance that are not one step contractive in the sense that most or all of the D_k have a spectral radius greater than or equal to one. For example, this is true for any iterative method that keeps some components of x_k fixed at each step, like cyclic reduction or any form of alternating projections. In those cases, one would still hope that over a cycle of iterations, a significant contraction is achieved in the following sense.

Assumption 4 *The preconditioners P_k are chosen uniformly bounded, so that*

$$\|P_k\| + \|P_k^{-1}\| \;\leq\; c_0 \;<\infty \quad for\ allk\ , \tag{26}$$

and there exists an induced matrix norm and a cycle length $m > 0$ such that

$$\delta_m \;\equiv\; \overline{\lim}_j \, \|D_{j+m} \cdot D_{j+m-1} \cdots D_{j+2} \cdot D_{j+1}\|^{\frac{1}{m}} \;<\; 1\ . \tag{27}$$

We will argue at the end of this section that any method for which this condition is not met would appear to be numerically unstable.

Proposition 3 *Under Assumptions 1 and 3, the iterations (1) and (7) converge with a linear R-factor no less than*

$$\delta_* \;=\; \inf_m \delta_m \;<\; 1$$

to their respective limits x_* and x'_*. Thus, we have

$$\overline{\lim}_k \, \|x_k - x_*\|^{1/k} \;\leq\; \delta_*\ , \quad and \quad \overline{\lim}_k \, \|\tilde{x}'_k - x'_*\|^{1/k} \;\leq\; \delta_*\ . \tag{28}$$

Proof. Abbreviating $\widehat{x}_k \equiv x_k - x_*$ and with r_k as defined in (12), we have by (13)

$$\widehat{x}_{k+m} = \left(\prod_{j=1}^{m} D_{k+m-j} \right) \widehat{x}_k \;+\; \sum_{i=1}^{m} \left(\prod_{j=1}^{m-i} D_{k+m-j} \right) r_{k+i-1} \tag{29}$$

over a cycle of m steps. Because of the assumed convergence of the x_k and (26), the D_i are uniformly bounded in norm by $(1 + c_0)^2$, so that by (27) for any ε and sufficiently large k, we have

$$\begin{aligned} \|\widehat{x}_{k+m}\| \;&\leq\; (\delta_m + \varepsilon)^m \, \|\widehat{x}_k\| + \max_{0 \leq i < m} \|r_{k+i}\| \sum_{i=1}^{m} (1 + c_0^2)^{i-1} \\ &\leq\; (\delta_m + \varepsilon)^m \, \|\widehat{x}_k\| + \max_{0 \leq i < m} \|r_{k+i}\| (1 + c_0^2)^m / c_0^2\ . \end{aligned} \tag{30}$$

Because of (13), the assumed convergence, and the uniform boundedness of the D_i, the $\widehat{x}_k$ grow at most linearly. Therefore, for some constant $c_7 = c_7(m)$

$$\|r_{k+j}\| \leq c_7 \|\widehat{x}_k\|^2 \quad for \quad 0 \leq j < m\ . \tag{31}$$

Hence we have by (29) for fixed m

$$\overline{\lim}_k \, \|\widehat{x}_{k+m}\| \, / \|\widehat{m}_k\| \;\leq\; (\delta_m + \varepsilon)^m\ ,$$

which ensures m-step Q-linear convergence with limiting ratio no greater than δ_m^m since ϵ may be chosen arbitrarily small. This implies the R-linear convergence assertion for the x_k by well-known results ([9]) and by taking the infimum of δ_m over m.

For the derivatives, we obtain from (7) for the $\widehat{x}'_k \equiv x'_k - x'_*$ the recurrence

$$\widehat{x}'_{k+m} = \left(\prod_{j=1}^{m} D_{k+m-j} \right) \widehat{x}'_k \; + \; \sum_{i=1}^{m} \left(\prod_{j=1}^{m-i} D_{k+m-j} \right) r'_{k+i-1} \; , \tag{32}$$

where r'_k is as defined in (9). Since the last bound in Lemma 1 was proven without any reference to Assumption 2, it can be used here to derive form the R-linear convergence of the x_k that for some constant $c_8 = c_8(m, \varepsilon)$

$$\max_{0<i<m} \|r'_{k+i}\|(1 + c_0^2)^m / c_0^2 \; \leq \; c_8(\delta_m + \epsilon)^{k+m} \; . \tag{33}$$

Substituting this bound into the "primed" version of (30) and then dividing by $(\delta_m + \varepsilon)^{k+m}$, we obtain the inequality

$$\frac{\|\widehat{x}'_{k+m}\|}{(\delta_m + \varepsilon)^{k+m}} - \frac{\|\widehat{x}'\|}{(\delta_m + \varepsilon)^k} \; \leq \; c_8 \; .$$

Summing for $k = i\,m$ over $i = 0, 1, \ldots, j-1$, we obtain

$$\|\widehat{x}'_{j\,c}\| \, / (\delta_m + \varepsilon)^{j\,c} \; \leq \; \|\widehat{x}'_0\| + j\, c_8 \; .$$

Since the $c\,j-$th root of the right-hand side converges to one, we obtain the asserted result, namely that the x'_k converge with the same linear R-factor δ_* to x'_*. ∎

As we have noticed above, the x_k may converge superlinearly. In those cases, the recurrence for x'_k will soon be almost exactly linear, so that one may seriously consider accelerating the derivative convergence by Richardson extrapolation. Since we have a constructive test on the quality of these extrapolated derivatives, it should be easy to determine the best candidate.

Finally, let us briefly examine the possibility that an iterative method of the general form achieves convergence but that assumption (27) is never satisfied. Then the equation (30) suggests that a small perturbation $\delta x_k = \delta x_k$ of the iterate x_k in the direction of the largest singular value of $D_{k+m} \cdot D_{k+m-1} \cdots D_{k+2} \cdot D_{k+1}$ will not be damped out over an arbitrarily large number m of steps. This would indicate that the method is numerically rather unstable. We can not make this claim rigorously, because the perturbation δx_k might alter the D_{k+j} in such a fortuitous way that it is damped out after all. For example, it is currently not clear whether conjugate direction methods can be interpreted in the form (1) such that (27) is satisfied. Derivative convergence has been observed for the classical conjugate gradient method, but this experimental observation cannot be supported by Proposition 1 and its corollaries.

Acknowledgments

The authors are indebted to John Dennis and Alan Carle for their insistence that the black box differentiation be analyzed.

References

[1] Christian Bischof, George Corliss, and Andreas Griewank. Structured second- and higher-order derivatives through univariate Taylor series. Preprint MCS–P296–0392, Mathematics and

Computer Science Division, Argonne National Laboratory, Argonne, Ill., March 1992. ADIFOR Working Note # 6.

[2] George F. Corliss. Overloading point and interval Taylor operators. In Andreas Griewank and George F. Corliss, editors, *Automatic Differentiation of Algorithms: Theory, Implementation, and Application*, pages 139–146. SIAM, Philadelphia, Penn., 1991.

[3] J. Ch. Gilbert. Automatic differentiation and iterative processes. *Optimization Methods and Software*, 1:13–21, 1992. Also appeared as Preprint, INRIA, Le Chesnay, France, 1991.

[4] Andreas Griewank. Automatic evaluation of first- and higher-derivative vectors. In R. Seydel, F. W. Schneider, T. Küpper, and H. Troger, editors, *Proceedings of the Conference at Würzburg, Aug. 1990, Bifurcation and Chaos: Analysis, Algorithms, Applications*, volume 97, pages 135–148. Birkhäuser Verlag, Basel, Switzerland, 1991.

[5] ??? J. Dennis and Jorge More. Motivation and Theory

[6] ??? Powell. A hybrid method for nonlinear equations, 1970.

[7] ??? special steps.

[8] ??? Banach Perturbation Lemma.

[9] ??? Ortega and Rheinboldt. Iterative Solution.