

Inexact Perturbed Newton Methods and Applications to a Class of Krylov Solvers¹

E. CĂȚINAȘ²

Communicated by F. A. Potra

Abstract. Inexact Newton methods are variant of the Newton method in which each step satisfies only approximately the linear system (Ref. 1). The local convergence theory given by the authors of Ref. 1 and most of the results based on it consider the error terms as being provided only by the fact that the linear systems are not solved exactly. The few existing results for the general case (when some perturbed linear systems are considered, which in turn are not solved exactly) do not offer explicit formulas in terms of the perturbations and residuals. We extend this local convergence theory to the general case, characterizing the rate of convergence in terms of the perturbations and residuals.

The Newton iterations are then analyzed when, at each step, an approximate solution of the linear system is determined by the following Krylov solvers based on backward error minimization properties: GMRES, GMBACK, MINPERT. We obtain results concerning the following topics: monotone properties of the errors in these Newton–Krylov iterates when the initial guess is taken 0 in the Krylov algorithms; control of the convergence orders of the Newton–Krylov iterations by the magnitude of the backward errors of the approximate steps; similarities of the asymptotical behavior of GMRES and MINPERT when used in a converging Newton method. At the end of the paper, the theoretical results are verified on some numerical examples.

Key Words. Nonlinear systems, (inexact) Newton methods, (inexact) perturbed Newton methods, convergence orders, linear systems, backward errors, Krylov methods (GMRES, GMBACK, MINPERT), Newton–Krylov methods.

¹This research was supported by the Romanian Academy of Sciences under Grant GAR 95/1998. The support and guidance of Dr. I. Păvăloiu, Head of the Numerical Analysis Institute, proved very important to the progress of the research. The author is also grateful to two anonymous referees for their careful reading of previous versions of this paper and helpful criticism and suggestions.

²Research Fellow, T. Popoviciu Institute of Numerical Analysis, Romanian Academy of Sciences, Cluj–Napoca, Romania.

1. Introduction

Consider the system of nonlinear equations

$$F(x) = 0, \quad (1)$$

where $F: \mathbb{R}^N \rightarrow \mathbb{R}^N$ is a nonlinear mapping and suppose that:

(C1) There exists $x^* \in \mathbb{R}^N$ such that $F(x^*) = 0$.

A classical approach for approximating x^* is the Newton method, which results in the following algorithm:

(NM) Choose an initial approximation $x_0 \in \mathbb{R}^N$.

For $k = 0, 1, \dots$ until convergence, do the following steps.

Step 1. Solve $F'(x_k)s_k = -F(x_k)$.

Step 2. Set $x_{k+1} = x_k + s_k$.

At each iteration of the Newton method, a routine for solving the resulting linear system must be called. The direct methods for linear systems cannot offer always the exact solution when used in floating point arithmetic and they may be inefficient for large general systems. Usually, some of the iterative methods cannot offer the exact solution after a finite number of steps even in exact arithmetic. Another fact is that, when x_k is far from x^* , it may be worthless to solve exactly the system. These reasons require a convergence analysis which takes into account some error terms.

Several Newton-type methods have been studied. Some sufficient conditions for different convergence orders have been given in Refs. 2–16 and references therein. A characterization of superlinear convergence and of convergence with orders $1+p$, $p \in (0, 1]$, has been obtained by Dennis and Moré in Refs. 17–18 for sequences given by the following quasi-Newton method:

$$x_{k+1} = x_k - B_k^{-1}F(x_k), \quad k = 0, 1, \dots, \quad x_0 \in \mathbb{R}^N,$$

$(B_k)_{k \geq 0} \subset \mathbb{R}^{N \times N}$ being a sequence of invertible matrices.

A characterization of local superlinear convergence and local convergence with orders $1+p$, $p \in (0, 1]$, of the Newton methods which take into account other error terms has been given by Dembo, Eisenstat, and Steihaug (Ref. 1). They considered in their paper the following inexact Newton method:

(INM) Choose an initial approximation $x_0 \in \mathbb{R}^N$.

For $k = 0, 1, \dots$, until convergence, do the following steps.

Step 1. Find s_k such that $F'(x_k)s_k = -F(x_k) + r_k$.

Step 2. Set $x_{k+1} = x_k + s_k$.

The error terms (residuals) r_k represent the amounts by which the solutions s_k , determined in an unspecified manner, fail to satisfy the exact systems (NM). Their magnitudes are determined by the relative residuals $\|r_k\|/\|F(x_k)\|$, supposed to be bounded by the forcing sequence $(\eta_k)_{k \geq 0}$,

$$\|r_k\|/\|F(x_k)\| \leq \eta_k, \quad k = 0, 1, \dots \quad (2)$$

The local convergence analysis given in Ref. 1 characterizes the convergence orders of $(x_k)_{k \geq 0}$ given by the (INM) in terms of the magnitudes of r_k ; see also Refs. 19–20 for other convergence results. However, in this paper as well as in most other papers using these results, the error terms r_k are considered to appear only because the exact Newton systems (NM) are solved approximately. But in many situations, it is hard to find the exact value of $F'(x)$, and in some cases even of $F(x)$. On the other hand, $F'(x)$ or its approximation and $F(x)$ are both altered when represented in floating point arithmetic.

The question that arises naturally is: What magnitudes can we allow in perturbing the matrices $F'(x_k)$ and the vectors $-F(x_k)$ so that the convergence order of the resulting method does not decrease?

The Newton methods with perturbed linear systems have been considered by several authors (see for instance Refs. 9, 11, 21, and references therein), but these methods were not analyzed with respect to their convergence orders.

Martinez, Parada, and Tapia (Ref. 22) have analyzed the superlinear convergence of sequences given by the following damped and perturbed quasi-Newton method:

$$x_{k+1} = x_k - \alpha_k B_k^{-1} [F(x_k) + r_k],$$

where $0 < \alpha_k \leq 1$, $r_k \in \mathbb{R}^N$, $k = 0, 1, \dots$, and $x_0 \in \mathbb{R}^N$; but their superlinear convergence was not characterized simultaneously in terms of α_k , B_k , r_k .

In Ref. 23, Ypma studied the case when the matrices $F'(x_k)$ and the vectors $-F(x_k)$ are calculated approximately, the resulting linear systems being solved inexactly. However, the residuals were considered to be incorporated in those perturbed linear systems. Consequently, the obtained convergence results do not offer explicit formulas in terms of the magnitude of the perturbations and residuals. Recently, Cores and Tapia (Ref. 24) have considered the exact solving of perturbed linear systems, but they have obtained only sufficient conditions on the perturbations for different convergence orders to be attained.

We consider here the perturbations $(\Delta_k)_{k \geq 0} \subset \mathbb{R}^{N \times N}$ and $(\delta_k)_{k \geq 0} \subset \mathbb{R}^N$, being led to the study of the following inexact perturbed Newton method:

$$(\text{IPNM}) \quad [F'(x_k) + \Delta_k] s_k = [-F(x_k) + \delta_k] + \hat{r}_k,$$

$$x_{k+1} = x_k + s_k, \quad k = 0, 1, \dots, \quad x_0 \in \mathbb{R}^N.$$

The terms \hat{r}_k denote the residuals of the approximate solutions s_k of the linear systems

$$[F'(x_k) + \Delta_k]s = -F(x_k) + \delta_k.$$

When these systems are assumed to be solved exactly [i.e., $\hat{r}_k = 0$, $k = 0, 1, \dots$, in the (IPNM)], we call the resulting method a perturbed Newton method,

$$(\text{PNM}) \quad [F'(x_k) + \Delta_k]s_k = -F(x_k) + \delta_k.$$

This frame will be useful for the study of some Newton–Krylov methods in connection with backward errors.

The paper is structured as follows. In Section 2, we give two convergence results for the (IPNM). In Section 3, we characterize the superlinear convergence and the convergence with orders $1 + p$, $p \in (0, 1]$, of the (IPNM), and we also provide some sufficient conditions. In Section 4, we analyze the Newton methods in which the linear systems from each step are solved by GMRES, GMBACK, or MINPERT. We verify the obtained theoretical results on numerical examples performed on two test problems (Section 5), and we end the paper with concluding remarks.

Throughout the paper, we consider the Euclidean norm (denoted by $\|\cdot\|_2$) and an arbitrary given norm on \mathbb{R}^N (denoted by $\|\cdot\|$), together with their induced operator norms. We use also the Frobenius norm of a matrix $Z \in \mathbb{R}^{M \times N}$, defined as

$$\|Z\|_F = \sqrt{\text{tr}(ZZ^T)}.$$

We use the column notation for vectors; when we write vectors (or matrices) inside square brackets, we consider the matrix containing on columns those vectors (or the columns of those matrices). The same convention is used when joining rows. As usual, the vectors e_i , $i = 1, \dots, n$, form the standard basis in \mathbb{R}^n , n being clear from the context.

2. Convergence Results for Inexact Perturbed Newton Methods

The common conditions for the local convergence analysis of the Newton method are Condition (C1) and the following additional conditions:

- (C2) The mapping F is differentiable on a neighborhood of x^* and F' is continuous at x^* .
- (C3) The Jacobian $F'(x^*)$ is nonsingular.

These conditions ensure that x^* is a point of attraction for the Newton method; i.e., there exists $\epsilon > 0$ such that $(x_k)_{k \geq 0}$ given by the Newton method converges to x^* for any initial approximation $x_0 \in \mathbb{R}^N$ with $\|x_0 - x^*\| < \epsilon$.

Moreover, the convergence is q -superlinear³; see Ref. 2, Theorem 10.2.2 and Ref. 10, Theorem 4.4.

In the convergence analysis of the (IPNM) iterations, we assume that:

- (C4) The perturbations Δ_k are such that the matrices $F'(x_k) + \Delta_k$ are nonsingular for $k = 0, 1, \dots$.

Though different in notation, in fact this condition is similar to the one considered for the quasi-Newton iterates, which requires a sequence of invertible matrices $(B_k)_{k \geq 0}$.

The following sufficient condition for the convergence of the (INM) was proved by Dembo, Eisenstat, and Steihaug (Ref. 1).

Theorem 2.1. See Ref. 1. Assume Conditions (C1)–(C3) and $\eta_k \leq \eta_{\max} < t < 1$, $k = 0, 1, \dots$. There exists $\epsilon > 0$ such that, if $\|x_0 - x^*\| \leq \epsilon$, then the sequence of the (INM) iterates $(x_k)_{k \geq 0}$ satisfying (2) converges to x^* . Moreover, the convergence is linear, in the sense that

$$\|x_{k+1} - x^*\|_* \leq t \|x_k - x^*\|_*, \quad k = 0, 1, \dots,$$

where $\|y\|_* = \|F'(x^*)y\|$.

Using this theorem, we obtain the following convergence results for the (IPNM).

Theorem 2.2. Assume Conditions (C1)–(C4) and $\eta_k \leq \eta_{\max} < t < 1$, $k = 0, 1, \dots$. There exists $\epsilon > 0$ such that, if $\|x_0 - x^*\| \leq \epsilon$ and

$$\|\Delta_k [F'(x_k) + \Delta_k]^{-1} F(x_k) + \{I - \Delta_k [F'(x_k) + \Delta_k]^{-1}\}(\delta_k + \hat{r}_k)\| \leq \eta_k \|F(x_k)\|,$$

for $k = 0, 1, \dots$, then the sequence of the (IPNM) iterates $(x_k)_{k \geq 0}$ converges to x^* , the convergence being linear,

$$\|x_{k+1} - x^*\|_* \leq t \|x_k - x^*\|_*, \quad k = 0, 1, \dots$$

Proof. The (IPNM) can be viewed as an (INM) with

$$s_k = -[F'(x_k) + \Delta_k]^{-1} F(x_k) + [F'(x_k) + \Delta_k]^{-1} (\delta_k + \hat{r}_k),$$

$$F'(x_k)s_k = -\Delta_k s_k - F(x_k) + \delta_k + \hat{r}_k$$

$$= -F(x_k) + \Delta_k [F'(x_k) + \Delta_k]^{-1} F(x_k)$$

$$- \Delta_k [F'(x_k) + \Delta_k]^{-1} (\delta_k + \hat{r}_k) + \delta_k + \hat{r}_k$$

$$= -F(x_k) + \Delta_k [F'(x_k) + \Delta_k]^{-1} F(x_k)$$

$$+ \{I - \Delta_k [F'(x_k) + \Delta_k]^{-1}\}(\delta_k + \hat{r}_k).$$

³For definitions and results concerning convergence orders, see Ref. 2, Chapter 9 and also Refs. 10, 25.

Denoting

$$r_k = \Delta_k [F'(x_k) + \Delta_k]^{-1} F(x_k) + \{I - \Delta_k [F'(x_k) + \Delta_k]^{-1}\} (\delta_k + \hat{r}_k), \quad (3)$$

the conclusion follows from Theorem 2.1. \square

Corollary 2.1. Assume Conditions (C1)–(C4). There exists $\epsilon > 0$ such that, if $\|x_0 - x^*\| \leq \epsilon$ and

$$\|\Delta_k [F'(x_k) + \Delta_k]^{-1}\| \leq q_1 < 1, \quad k = 0, 1, \dots,$$

$$\|\delta_k\| + \|\hat{r}_k\| \leq [\eta_k / (1 + q_1)] \|F(x_k)\|,$$

$$\text{where } \eta_k \leq q_2 < 1 - q_1, \quad k = 0, 1, \dots,$$

then the sequence of the (IPNM) iterates $(x_k)_{k \geq 0}$ converges to x^* . Moreover, the convergence is linear,

$$\|x_{k+1} - x^*\|_* \leq t \|x_k - x^*\|_*, \quad k = 0, 1, \dots,$$

where $t = q_1 + q_2$.

Proof. The proof is obtained easily from the previous result making use of the hypotheses. \square

Remark 2.1. The idea of reducing certain perturbed Newton methods to (INM) iterations, which is used in the proof of Theorem 2.2, can be found in the work of several authors; see for example Refs. 9–10 and 26. The inexact secant methods considered by us in Ref. 27 are in fact instances of the (IPNM) model, but it was Ref. 28 that inspired us to consider the (IPNM).

3. Convergence Orders of Inexact Perturbed Newton Methods

Stronger conditions imposed on the continuity of F' at x^* offer higher convergence orders for the Newton method. Namely, if F' is Hölder continuous at x^* with exponent $p, p \in (0, 1]$, i.e., if there exist $\epsilon > 0$ and $L \geq 0$ such that

$$\|F'(x) - F'(x^*)\| \leq L \|x - x^*\|^p, \quad \text{for } \|x - x^*\| \leq \epsilon,$$

then the Newton method converges locally with q -order at least $1 + p$; see Ref. 2, Theorem 10.2.2 and Ref. 10, Theorem 4.4.

For the inexact Newton methods, Dembo, Eisenstat, and Steihaug proved the following result.

Theorem 3.1. See Ref. 1. Assume that Conditions (C1)–(C3) hold and that the inexact Newton iterates $(x_k)_{k \geq 0}$ converge to x^* . Then, $x_k \rightarrow x^*$ q -superlinearly if and only if

$$\|r_k\| = o(\|F(x_k)\|), \quad \text{as } k \rightarrow \infty.$$

Moreover, if F' is Hölder continuous at x^* with exponent $p, p \in (0, 1]$, then $x_k \rightarrow x^*$ with q -order at least $1 + p$ if and only if

$$\|r_k\| = \mathcal{O}(\|F(x_k)\|^{1+p}), \quad \text{as } k \rightarrow \infty.$$

We obtain the following result for the inexact perturbed Newton methods.

Theorem 3.2. Assume that Conditions (C1)–(C4) hold and that the iterates $(x_k)_{k \geq 0}$ given by the (IPNM) converge to x^* . Then, $x_k \rightarrow x^*$ q -superlinearly if and only if

$$\|\Delta_k[F'(x_k) + \Delta_k]^{-1}F(x_k) + \{I - \Delta_k[F'(x_k) + \Delta_k]^{-1}\}(\delta_k + \hat{r}_k)\| = o(\|F(x_k)\|),$$

as $k \rightarrow \infty$. Moreover, if F' is Hölder continuous at x^* with exponent p , then $x_k \rightarrow x^*$ with q -order at least $1 + p$ if and only if

$$\|\Delta_k[F'(x_k) + \Delta_k]^{-1}F(x_k) + \{I - \Delta_k[F'(x_k) + \Delta_k]^{-1}\}(\delta_k + \hat{r}_k)\| = \mathcal{O}(\|F(x_k)\|^{1+p}),$$

as $k \rightarrow \infty$.

Proof. The proof is obtained from the previous theorem by using (3). \square

In the following result, we characterize the convergence orders of the (IPNM) iterates in terms of the rate of convergence to zero of residuals and perturbations.

Corollary 3.1. Assume that:

- (a) Conditions (C1)–(C4) hold;
- (b) $\Delta_k \rightarrow 0, \delta_k \rightarrow 0, \hat{r}_k \rightarrow 0$, as $k \rightarrow \infty$;
- (c) the sequence $(x_k)_{k \geq 0}$ given by the (IPNM) converges to x^* .

In addition, if

$$\|\delta_k\| = o(\|F(x_k)\|), \quad \|\hat{r}_k\| = o(\|F(x_k)\|), \quad \text{as } k \rightarrow \infty,$$

then, $x_k \rightarrow x^*$ q -superlinearly.

Under the same assumptions (a)–(c), if additionally F' is Hölder continuous at x^* with exponent p , and if

$$\begin{aligned}\|\Delta_k\| &= \mathcal{O}(\|F(x_k)\|^p), & \|\delta_k\| &= \mathcal{O}(\|F(x_k)\|^{1+p}), \\ \|\hat{r}_k\| &= \mathcal{O}(\|F(x_k)\|^{1+p}), & \text{as } k \rightarrow \infty,\end{aligned}$$

then $x_k \rightarrow x^*$ with q -order at least $1 + p$.

The corresponding results for the r -convergence orders of the (IPNM) iterates may be stated in a similar manner, taking into account the existing results for the (INM); see Ref. 1, Theorem 3.4 and Corollary 3.5.

4. Applications to Krylov Solvers Based on Backward Error Minimization Properties

Consider the nonsymmetric linear system

$$Ax = b, \tag{4}$$

with $A \in \mathbb{R}^{N \times N}$ nonsingular and $b \in \mathbb{R}^N$. The Krylov methods for solving such a system when the dimension N is large are based on the Krylov subspaces, defined for any initial approximation $x_0 \in \mathbb{R}^N$ as

$$\mathcal{K}_m = \mathcal{K}_m(A, r_0) = \text{span}\{r_0, Ar_0, \dots, A^{m-1}r_0\},$$

where

$$r_0 = b - Ax_0$$

is the initial residual and $m \in \{1, \dots, N\}$.

The normwise backward error of an approximate solution \tilde{x} of (4) was introduced by Rgal and Gaches (Ref. 29) and is defined by

$$\Pi(\tilde{x}) = \min\{\epsilon: (A + \Delta_A)\tilde{x} = b + \Delta_b, \|\Delta_A\|_F \leq \epsilon\|E\|_F, \|\Delta_b\|_2 \leq \epsilon\|f\|_2\},$$

where the parameters $E \in \mathbb{R}^{N \times N}$ and $f \in \mathbb{R}^N$ are arbitrary. The value of $\Pi(\tilde{x})$ is

$$\Pi(\tilde{x}) = \|b - A\tilde{x}\|_2 / (\|E\|_F \cdot \|\tilde{x}\|_2 + \|f\|_2),$$

and the minimum is attained by the backward errors

$$\Delta_A = [\|E\|_F \cdot \|\tilde{x}\|_2 / (\|E\|_F \cdot \|\tilde{x}\|_2 + \|f\|_2)](b - A\tilde{x})z^t, \quad \text{with } z = (1/\|\tilde{x}\|_2^2)\tilde{x},$$

$$\Delta_b = -[\|f\|_2 / (\|E\|_F \cdot \|\tilde{x}\|_2 + \|f\|_2)](b - A\tilde{x}).$$

We shall analyze the behavior of the following three Krylov solvers when used for the linear systems arising in the Newton methods: GMRES, GMBACK, MINPERT. Each of these solvers is based on backward error minimization properties, as we shall see.

In order to use some existing notations, we prefer to denote hereafter by y^* the solution of the nonlinear system (1) and by y_k the iterates from the Newton methods, using x in the involved linear systems.

4.1. Newton–GMRES Method. Given an initial approximation $x_0 \in \mathbb{R}^N$, the GMRES method of Saad and Schultz (Ref. 30) uses the Arnoldi process (see Ref. 31) to construct an orthonormal basis $\{v_1, \dots, v_m\}$ in the Krylov subspace \mathcal{K}_m . The approximation $x_m^{\text{GM}} \in x_0 + \mathcal{K}_m$ is then determined such that

$$\|b - Ax_m^{\text{GM}}\|_2 = \min_{x_m \in x_0 + \mathcal{K}_m} \|b - Ax_m\|_2. \quad (5)$$

Kasensally has noted in Ref. 32 that the minimizing property of x_m^{GM} may be expressed in terms of the backward errors, namely,

$$\min_{x_m \in x_0 + \mathcal{K}_m} \|b - Ax_m\|_2 = \min_{x_m \in x_0 + \mathcal{K}_m} \{\|\Delta_b\|_2 : Ax_m = b - \Delta_b\};$$

i.e., x_m^{GM} minimizes over $x_0 + \mathcal{K}_m$ the backward error Δ_b , assuming $\Delta_A = 0$.

The solution x_m^{GM} is determined roughly in the following way (see also Ref. 33):

- (i) Arnoldi Process. Determine $V_m = [v_1, \dots, v_m] \in \mathbb{R}^{N \times m}$ and the upper Hessenberg matrix $\tilde{H}_m \in \mathbb{R}^{(m+1) \times m}$.
- (ii) GMRES Solution. Find the exact solution y_m^{GM} of the following least squares problem in \mathbb{R}^m :

$$\min_{y_m \in \mathbb{R}^m} \|\beta e_1 - \tilde{H}_m y_m\|_2,$$

$$\text{where } \beta = \|r_0\|_2.$$

$$\text{Set } x_m^{\text{GM}} = x_0 + V_m y_m^{\text{GM}}.$$

The GMRES method is used in an iterative fashion. Saad and Schultz proved that, at each step, the solution x_m^{GM} is uniquely determined and that the algorithm breaks down in the Arnoldi method only if the exact solution has been reached. Moreover, the process terminates in at most N steps. In the results presented below, we shall assume, as usually, that the Krylov methods do not continue after the exact solution has been determined.

First, we introduce some notations for the restarted GMRES iterations in order to express the relations which they satisfy.

Since only small values of m are attractive in practice, an upper bound $\bar{m} \in \{1, \dots, N-1\}$ is usually fixed for the subspace dimensions. If after \bar{m} steps the computed solution does not have a sufficiently small residual, the GMRES method is restarted, taking for the initial approximation the last computed solution. We denote by $x_m^{\text{GM}(0)}$, $m \in \{1, \dots, \bar{m}\}$, the first \bar{m} solutions, by $x_m^{\text{GM}(1)}$, $m \in \{1, \dots, \bar{m}\}$, the \bar{m} solutions from the first restart, and so on. The value of the initial approximation x_0 will be clear from the context. For the nonrestarted version, we shall use the common notations x_m^{GM} , while for a generic GMRES solution⁴, we shall simply write x^{GM} ; the corresponding notations for the k th correction from a Newton–GMRES method will be $s_{k,m}^{\text{GM}}$ [resp. s_k^{GM}].

The choice $x_0 = 0$ in the Krylov solvers is a popular one when no better guess is known; see e.g. Refs. 26 and 34–35. With the above notation, the following result is obtained immediately from the properties of the GMRES solutions. Certain affirmations from this result are more or less explicitly stated in some papers dealing with GMRES; see for example Refs. 26 and 36–37.

Proposition 4.1. Consider the linear system (4) and the initial approximation $x_0 = 0$. Then, the following statements are true:

- (i) For any $m \in \{1, \dots, N\}$, the residual r_m^{GM} of the GMRES solution satisfies

$$\|r_m^{\text{GM}}\|_2 \leq \|b\|_2.$$

Moreover, the inequality is strict if and only if the solution x_m^{GM} is nonzero.

- (ii) The residuals associated to the N successive solutions satisfy

$$0 = \|r_N^{\text{GM}}\|_2 \leq \|r_{N-1}^{\text{GM}}\|_2 \leq \dots \leq \|r_1^{\text{GM}}\|_2 \leq \|b\|_2.$$

The inequality between the norms of two consecutive residuals is strict if and only if the corresponding GMRES solutions are distinct.

- (iii) For any fixed upper bound $\bar{m} \in \{1, \dots, N-1\}$, the residuals of the restarted GMRES method satisfy

$$\begin{aligned} \dots \leq \|r_1^{\text{GM}(l+1)}\|_2 &\leq \|r_{\bar{m}}^{\text{GM}(l)}\|_2 \leq \dots \leq \|r_1^{\text{GM}(l)}\|_2 \\ &\leq \|r_{\bar{m}}^{\text{GM}(l-1)}\|_2 \leq \dots \leq \|r_1^{\text{GM}(0)}\|_2 \leq \|b\|_2. \end{aligned}$$

⁴In such a case, we assume that the initial approximation $x_0 \in \mathbb{R}^N$, the upper bound $\bar{m} \in \{1, \dots, N-1\}$, the number of (eventual) restarts $l \geq 0$, and the number of (final, if $l \geq 1$) iterations may be arbitrary.

The inequality between the norms of two consecutive residuals is strict if and only if the corresponding GMRES solutions are distinct; the residuals may eventually get to zero or may indefinitely stagnate, depending on the problem.

Considering the linear systems from the Newton method, we get easily the following proposition.

Proposition 4.2. Let $y \in D$ be an element for which the derivative $F'(y)$ is nonsingular. Applying the GMRES method with $x_0 = 0$ to the linear system $F'(y)s = -F(y)$, then for all $m \in \{1, \dots, N\}$, the residual satisfies

$$\|r_m^{\text{GM}}\|_2 \leq \|F(y)\|_2.$$

Moreover, the above inequality is strict if and only if the correction s_m^{GM} is nonzero.

We have chosen to state the result corresponding only to part (i) of Proposition 4.1. The other results are similarly enounced.

Brown (Ref. 26) and Brown and Saad (Ref. 38) considered solving (1) by global minimization,

$$\min_{y \in \mathbb{R}^N} f(y) = \min_{y \in \mathbb{R}^N} F(y)'F(y),$$

when $F: \mathbb{R}^N \rightarrow \mathbb{R}^N$. They obtained that any nonzero correction from a certain step of the Newton-GMRES method with $x_0 = 0$ in GMRES is a descent direction for the above minimization problem. Now, we are able to describe this positive behavior in terms of the distance to the solution.

Theorem 4.1. Assume that the mapping F satisfies Conditions (C1)–(C3), and consider a current approximation $y_c \neq y^*$ for y^* . Let s^{GM} be a nonzero approximate solution of the linear system $F'(y_c)s = -F(y_c)$ provided by GMRES, assuming the initial guess $x_0 = 0$. Let $y_+ = y_c + s^{\text{GM}}$, denote $\eta = \|r^{\text{GM}}\|_2 / \|F(y_c)\|_2$, and take $t \in (\eta, 1)$. If y_c is sufficiently close to y^* , then

$$\|y_+ - y^*\|_* \leq t \|y_c - y^*\|_*,$$

where $\|y\|_* = \|F'(y^*)y\|_2$.

Proof. The thesis can be proved easily with slight modifications of the proof of Theorem 2.1, given in Ref. 1. \square

The above result says that, when using GMRES (with $x_0 = 0$) in the Newton method, any nonzero correction improves the current outer

approximation, provided that the approximation is sufficiently good. Though this theorem guarantees a nonincreasing curve for the errors of the Newton–GMRES iterations starting from a sufficiently good approximation, the local convergence is not guaranteed. It is sufficient to notice that, considering a sequence obeying at each step the hypotheses of the above result, there may appear a situation when the set of forcing terms has 1 as accumulation point, in which case Theorem 2.1 cannot be applied.

The following theoretical example shows that, when the dimensions of the Krylov subspaces are smaller than N , the outer iterations may stagnate at the initial approximation $y_0 \neq y^*$, no matter how close to y^* we choose y_0 .

Example 4.1. Consider

$$F: \mathbb{R}^N \rightarrow \mathbb{R}^N, \quad F(y) = [y^{(N)}, y^{(1)}, y^{(2)}, \dots, y^{(N-1)}]^t, \\ y = [y^{(1)}, \dots, y^{(N)}]^t \in \mathbb{R}^N,$$

for which

$$F'(y) \equiv \begin{bmatrix} 0 & & & 1 \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{bmatrix} = [e_2, e_3, \dots, e_N, e_1] =: A.$$

The Newton method for solving $F(y) = 0$ should yield the unique solution $y^* = 0$ in one iteration for any $y_0 \in \mathbb{R}^N$, whenever the corresponding linear system is solved exactly. However, when the correction is determined approximately, the situation changes, and we may use the (INM) setting.

Taking $y_0 = -he_N$, for some arbitrarily small $h \neq 0$, we must solve

$$F'(y_0)s = -F(y_0),$$

i.e.,

$$Ax = b, \quad \text{with } b = he_1.$$

Applying $m \leq N-1$ steps of the Arnoldi process with $x_0 = 0$, we obtain $V_m = [e_1, \dots, e_m] \in \mathbb{R}^{N \times m}$ and $\bar{H}_m = [e_2, \dots, e_{m+1}] \in \mathbb{R}^{(m+1) \times m}$.

The GMRES solution is given by (see Refs. 36 and 39)

$$x_m^{\text{GM}} = x_0 + V_m(\bar{H}_m^t \bar{H}_m)^{-1} \bar{H}_m^t V_{m+1}^t b,$$

such that

$$V_{m+1}^t b = he_1, \quad \bar{H}_m^t he_1 = 0,$$

and so

$$x_m^{\text{GM}} = s_{0,m}^{\text{GM}} = 0, \quad \text{for all } m = 1, \dots, N-1.$$

We also note that the restarting version of the GMRES yields the same result whenever $x_0 = 0$ and $\bar{m}, m \leq N - 1$.

This theoretical example shows that the dimension of the Krylov subspaces may sometimes be crucial for the efficiency of the Newton–GMRES method, and that the use of the restarted version of GMRES cannot overcome the difficulties. The stagnation of the GMRES algorithm was first reported in Ref. 36 (where A was as above and b stood for e_1), but we are not aware of any extension to a (non)linear mapping in order to show such stagnation of the Newton–GMRES method.

4.2. Newton–GMBACK Method. The GMBACK algorithm for solving the linear system (4) was introduced by Kasenally in Ref. 32. Given $x_0 \in \mathbb{R}^N$, it computes a vector $x_m^{\text{GB}} \in x_0 + \mathcal{K}_m$ which minimizes the backward error in the matrix A , assuming $\Delta_b = 0$,

$$\min_{x_m \in x_0 + \mathcal{K}_m} \|\Delta_A\|_F, \quad \text{s.t. } (A - \Delta_A)x_m = b.$$

The following steps are performed for determining x_m^{GB} .

- (i) Arnoldi Process. Compute the matrices V_m and \tilde{H}_m .
- (ii) GMBACK Solution. Execute Steps (a), (b), (c), (d) below.

- (a) Let $\beta = \|r_0\|_2$,

$$\hat{H}_m = [-\beta e_1, \tilde{H}_m] \in \mathbb{R}^{(m+1) \times (m+1)},$$

$$\hat{G}_m = [x_0, V_m] \in \mathbb{R}^{N \times (m+1)},$$

$$P = \hat{H}_m^t \hat{H}_m \in \mathbb{R}^{(m+1) \times (m+1)},$$

$$Q = \hat{G}_m^t \hat{G}_m \in \mathbb{R}^{(m+1) \times (m+1)}.$$

- (b) Determine an eigenvector u_{m+1} corresponding to the smallest eigenvalue $\lambda_{m+1}^{\text{GB}}$ of the generalized eigenproblem $Pu = \lambda Qu$.
- (c) If the first component $u_{m+1}^{(1)}$ is nonzero, compute the vector $y_m^{\text{GB}} \in \mathbb{R}^m$ by scaling u_{m+1} such that

$$\begin{bmatrix} 1 \\ y_m^{\text{GB}} \end{bmatrix} = [1/u_{m+1}^{(1)}] u_{m+1}.$$

- (d) Set $x_m^{\text{GB}} = x_0 + V_m y_m^{\text{GB}}$.

This algorithm may lead to two possible breakdowns, either in the Arnoldi method or in the scaling of u_{m+1} . As in the case of GMRES, the

first breakdown is a happy breakdown, because the solution may be determined exactly using \bar{H}_m and V_m . The second breakdown appears when all the eigenvectors associated to $\lambda_{m+1}^{\text{GB}}$ have the first component zero, the inevitable divisions by zero leading to uncircumventible breakdowns. In such case, either m is increased or the algorithm is restarted with a different initial approximation x_0 . We shall assume in the following analysis that x_m^{GB} exists.

Kasenally proved that, for any $x_0 \in \mathbb{R}^N$ and $m \in \{1, \dots, N\}$, the backward error $\Delta_{A,m}^{\text{GB}}$ corresponding to the GMBACK solution satisfies

$$\|\Delta_{A,m}^{\text{GB}}\|_F = \sqrt{\lambda_{m+1}^{\text{GB}}}. \quad (6)$$

The Newton–GMBACK iterates may be written in two equivalent ways, taking into account the properties of the Krylov solutions,

$$F'(y_k)s_k^{\text{GB}} = -F(y_k) + r_k^{\text{GB}},$$

$$[F'(y_k) - \Delta_{A_k}^{\text{GB}}]s_k^{\text{GB}} = -F(y_k),$$

$k = 0, 1, \dots$, $y_0 \in D$, where we considered

$$A_k = F'(y_k) \quad \text{and} \quad b_k = -F(y_k).$$

There are three results which may be applied to characterize the high convergence orders of these sequences (the corresponding statements are left to the reader): Theorem 2.1 of Dembo, Eisenstat, and Steihaug; the results of Dennis and Moré for the quasi-Newton methods (see Refs. 17–18); and Theorem 2.2 for the (IPNM) iterates, in which we must take $\delta_k = \hat{r}_k = 0$, $k = 0, 1, \dots$. Naturally, these results must be equivalent, but here we do not analyze this aspect.

Concerning the sufficient conditions, the convergence orders of the Newton–GMBACK method may be controlled by the computed eigenvalues λ_k^{GB} from GMBACK.

Theorem 4.2. Consider the sequence of Newton–GMBACK iterates $y_{k+1} = y_k + s_k^{\text{GB}}$, where s_k^{GB} satisfies

$$[F'(y_k) - \Delta_{A_k}^{\text{GB}}]s_k^{\text{GB}} = -F(y_k), \quad k = 0, 1, \dots$$

Assume the following:

- (a) Conditions (C1)–(C4) hold.
- (b) The derivative F' is Hölder continuous with exponent p at y^* .
- (c) The sequence $(y_k)_{k \geq 0}$ converges to y^* .

Moreover, if

$$\sqrt{\lambda_k^{\text{GB}}} = \mathcal{O}(\|F(y_k)\|^p), \quad \text{as } k \rightarrow \infty,$$

then the Newton–GMBACK iterates converge with q -order at least $1 + p$.

Proof. The proof follows from Corollary 3.1, taking into account formula (6), the inequality $\|Z\|_2 \leq \|Z\|_F$, true for all $Z \in \mathbb{R}^{N \times N}$, and the fact that all the norms are equivalent on a finite-dimensional normed space. \square

We are interested now if the curve of the Newton–GMBACK errors is nonincreasing when starting with a sufficiently good approximation. In the following, we shall construct an example which shows that, unlike the Newton–GMRES case, this property is not generally shared by the Newton–GMBACK method when GMBACK is used in the nonrestarted version.

Example 4.2. Consider the system $F(y) = 0$, with

$$F: \mathbb{R}^N \rightarrow \mathbb{R}^N,$$

$$F(y) = [y^{(1)} + y^{(N)}, y^{(1)}, y^{(2)}, \dots, y^{(N-1)}]^t,$$

$$y = [y^{(1)}, \dots, y^{(N)}]^t \in \mathbb{R}^N,$$

having the unique solution $y^* = 0$. For all $y \in \mathbb{R}^N$, the derivative of F is given by

$$F'(y) \equiv [e_1 + e_2, e_3, \dots, e_N, e_1] =: A.$$

Taking $y_0 = -he_N$ with arbitrarily small $h \neq 0$, we must solve

$$F'(y_0)s = -F(y_0),$$

or equivalently,

$$Ax = b, \quad \text{with } b = he_1.$$

Applying $m \leq N-1$ steps of the Arnoldi process with $x_0 = 0$, we obtain successively

$$V_m = [e_1, \dots, e_m] \in \mathbb{R}^{N \times m},$$

$$\tilde{H}_m = [e_1 + e_2, e_3, \dots, e_{m+1}] \in \mathbb{R}^{(m+1) \times m},$$

$$P = \hat{H}_m^t \hat{H}_m = [h^2 e_1 - he_2, 2e_2 - he_1, e_3, \dots, e_{m+1}],$$

$$Q = \hat{G}_m^t \hat{G}_m = [0, e_2, \dots, e_{m+1}] \in \mathbb{R}^{(m+1) \times (m+1)}.$$

The eigenpair $(u_{m+1}, \lambda_{m+1}^{\text{GB}})$ of

$$Pu = \lambda Qu$$

is determined uniquely by

$$u_{m+1} = [1/h, 1, 0, \dots, 0]^t, \quad \lambda_{m+1}^{\text{GB}} = 1.$$

It follows that

$$y_m^{\text{GB}} = he_1 \in \mathbb{R}^m,$$

and the GMBACK, solution is

$$x_m^{\text{GB}} = he_1,$$

such that, when $N \geq 3$,

$$\|y_1 - y^*\|_2 = \sqrt{2}|h| > |h| = \|y_0 - y^*\|_2.$$

4.3. Newton–MINPERT Method. The MINPERT method for solving (4) was introduced by Kasenally and Simoncini in Ref. 40. Given an initial approximation $x_0 \in \mathbb{R}^N$, it computes an element $x_m^{\text{MP}} \in x_0 + \mathcal{K}_m$ which minimizes the joint backward error,

$$\min_{x_m \in x_0 + \mathcal{K}_m} \|[\Delta_A, \Delta_b]\|_F, \quad \text{s.t. } (A - \Delta_A)x_m = b + \Delta_b. \quad (7)$$

In other words, x_m^{MP} minimizes the distance from the original system to the nearest one that an approximation $x_m \in x_0 + \mathcal{K}_m$ actually satisfies. The above minimization problem may also be viewed as a total least squares problem in the Krylov subspace (see Ref. 40).

The algorithm is similar to GMBACK, the only difference being given by the computation of the matrices from the eigenproblem

$$Pu = \lambda Qu.$$

The following is a sketch of the steps performed by MINPERT.

- (i) Arnoldi Process. Compute the matrices V_m and \tilde{H}_m .
- (ii) MINPERT Solution.
 - (a) Let $\beta = \|r_0\|_2$,

$$\hat{H}_m = [-\beta e_1, \tilde{H}_m] \in \mathbb{R}^{(m+1) \times (m+1)},$$

$$\hat{G}_m = \begin{bmatrix} x_0 & V_m \\ 1 & 0 \end{bmatrix} \in \mathbb{R}^{(N+1) \times (m+1)},$$

$$P = \hat{H}_m^t \hat{H}_m \in \mathbb{R}^{(m+1) \times (m+1)},$$

$$Q = \hat{G}_m^t \hat{G}_m \in \mathbb{R}^{(m+1) \times (m+1)}.$$

- (b) Determine an eigenvector u_{m+1} corresponding to the smallest eigenvalue $\lambda_{m+1}^{\text{MP}}$ of the generalized eigenproblem $Pu = \lambda Qu$.

- (c) If the first component $u_{m+1}^{(1)}$ is nonzero, compute the vector $y_m^{\text{MP}} \in \mathbb{R}^m$ by scaling u_{m+1} such that

$$\begin{bmatrix} 1 \\ y_m^{\text{MP}} \end{bmatrix} = [1/u_{m+1}^{(1)}] u_{m+1}.$$

- (d) Set $x_m^{\text{MP}} = x_0 + V_m y_m^{\text{MP}}$.

The remarks concerning the breakdowns of the GMBACK method hold also for MINPERT.

Kasenally and Simoncini proved that, for any $x_0 \in \mathbb{R}^N$ and $m \in \{1, \dots, N\}$, the following relations hold:

$$\|[\Delta_{A,m}^{\text{MP}}, \Delta_{b,m}^{\text{MP}}]\|_F = \sqrt{\lambda_{m+1}^{\text{MP}}}, \quad (8)$$

$$\|r_m^{\text{MP}}\|_2 = \sqrt{\lambda_{m+1}^{\text{MP}}} \|[(x_m^{\text{MP}})^t, 1]^t\|_2, \quad (9)$$

where $\Delta_{A,m}^{\text{MP}}$, $\Delta_{b,m}^{\text{MP}}$ and r_m^{MP} represent the backward errors and the residual corresponding to the MINPERT solution x_m^{MP} .

We shall prove for MINPERT a result somehow similar to Proposition 4.1.

Proposition 4.3. Consider the linear system (4), the initial approximation $x_0 = 0$, and an arbitrary value $m \in \{1, \dots, N\}$. If there exists a MINPERT solution x_m^{MP} , then its joint backward error satisfies

$$\|[\Delta_{A,m}^{\text{MP}}, \Delta_{b,m}^{\text{MP}}]\|_F \leq \|b\|_2.$$

Proof. When $x_0 = 0$, as noticed in Ref. 40, we are led to a regular eigenproblem in the MINPERT algorithm, since $Q = I_{m+1}$. The boundedness of the Rayleigh quotient then implies

$$\lambda_{m+1}^{\text{MP}} = \min_{z \in \mathbb{R}^{m+1}} z^t P z / z^t z \leq e_1^t P e_1 / e_1^t e_1 = e_1^t \hat{H}_m^t \hat{H}_m e_1 = \beta^2 = \|b\|_2^2;$$

so, by (8), the conclusion is immediate. \square

The other two affirmations similar to those from Proposition 4.1 may be stated correspondingly with the single remark that, since the solution of the minimization problem (7) may not be uniquely determined, the inequality between two consecutive joint backward errors may not be strict even if the consecutive solutions x_m^{MP} and x_{m+1}^{MP} are distinct.

We also note that, when the initial approximation in GMBACK is taken $x_0 = 0$, one cannot similarly use the vector e_1 in the Fisher result [see, e.g., Ref. 41, Corollary VI.1.16, which says that $\lambda_{m+1}^{\text{GB}} = \min_{z \in \mathbb{R}^{m+1}} z^t P z / z^t Q z$] in

order to bound $\lambda_{m+1}^{\text{GB}}$ by β^2 , since in this case

$$e_1' Q e_1 = 0;$$

i.e., e_1 is an eigenvector corresponding to the infinite eigenvalue $\lambda_1 = +\infty$. Such a result could not be expected to hold in general, because by Corollary 3.1 it would imply that, for an arbitrary mapping F satisfying Conditions (C1)–(C3) and with Lipschitz derivative, the Newton–GMBACK iterations with $x_0 = 0$ in GMBACK and obeying Condition (C4) converge locally with q -order 2 even for one-dimensional Krylov subspaces.

The convergence orders of the Newton–MINPERT iterates may be characterized by (9) in terms of the computed eigenvalues λ_k^{MP} .

Theorem 4.3. Assume that Conditions (C1)–(C3) hold, that the derivative F' is Hölder continuous with exponent p at y^* , and that the sequence $(y_k)_{k \geq 0}$ given by the Newton–MINPERT iterations $y_{k+1} = y_k + s_k^{\text{MP}}$, with

$$F'(y_k)s_k^{\text{MP}} = -F(y_k) + r_k^{\text{MP}}, \quad k = 0, 1, \dots, \quad y_0 \in D,$$

converges to y^* . Then, $(y_k)_{k \geq 0}$ converges with q -order at least $1 + p$ if and only if

$$\sqrt{\lambda_k^{\text{MP}}} = \mathcal{O}(\|F(y_k)\|^{1+p}), \quad \text{as } k \rightarrow \infty.$$

The direct application of Proposition 4.3 to the study of the monotonicity of the Newton–MINPERT errors leads to a result similar to Theorem 4.1, but which requires some additional conditions.

Theorem 4.4. Assume that the mapping F satisfies Conditions (C1)–(C3), and consider a current approximation $y_c \neq y^*$ for y^* . Let s^{MP} be a nonzero approximate solution of the linear system $F'(y_c)s = -F(y_c)$ provided by MINPERT with the initial guess $x_0 = 0$. Let

$$y_+ = y_c + s^{\text{MP}}.$$

If

$$\eta = \sqrt{\lambda^{\text{MP}}} \|[(s^{\text{MP}})^t, 1]^t\|_2 / \|F(y_c)\|_2 < 1,$$

and if y_c is sufficiently close to y^* , then

$$\|y_+ - y^*\|_* \leq t \|y_c - y^*\|_*, \quad \text{for all } t \in (\eta, 1),$$

where $\|y\|_* = \|F'(y^*)y\|_2$.

The following example shows that the Newton–MINPERT iterates may stagnate no matter how close to the solution y^* we choose y_0 .

Example 4.3. Consider the mapping F from Example 4.1, which leads us again to solving the linear system $Ax = b$. The MINPERT method with $x_0 = 0$ and $m \in \{1, \dots, N-1\}$ yields the diagonal matrix $P \in \mathbb{R}^{(m+1) \times (m+1)}$, with $\text{diag}(P) = [h^2, 1, \dots, 1]$ and $Q = I_{m+1}$, such that, for any h with $0 < |h| < 1$, we get

$$\lambda_{m+1}^{\text{MP}} = h^2 \quad \text{and} \quad u_{m+1} = e_1.$$

Then, the unique choice $y_m^{\text{MP}} = 0$ follows, and so $s_{0,m}^{\text{MP}} = 0$.

Remark 4.1. This example also shows that, for the linear systems $Ax = b$ above, the MINPERT method behaves identically with GMRES; i.e., for $x_0 = 0$ and $\bar{m}, m \in \{1, \dots, N-1\}$, it cannot improve the accuracy of the approximate solution, even in the restarted version. The systems with the same matrix A , but with $b = he_1$, $|h| > 1$, constitute some examples when the MINPERT method for a linear system leads to uncircumventible breakdowns. The stagnations and the uncircumventible breakdowns of MINPERT were theoretically known to be possible, but we had not encountered concrete examples before.

We end the theoretical considerations by applying a result which relates the GMRES and MINPERT approximations to the Newton iterations. However, we do not tackle here this problem in its general setting nor do we analyze the conditions required in this result.

Kasenally and Simoncini proved that the difference between the GMRES and MINPERT solutions may be bounded under certain circumstances.

Theorem 4.5. See Ref. 40. Consider the same arbitrary elements $m \in \{1, \dots, N\}$ and $x_0 \in \mathbb{R}^N$ in the GMRES and MINPERT methods for solving the linear system (4). Denote by σ_m^2 the smallest eigenvalue of $\tilde{H}_m^t \tilde{H}_m$, i.e., the smallest squared singular value of \tilde{H}_m , and assume that $\sigma_m^2 \neq \lambda_{m+1}^{\text{MP}}$. Then,

$$\|x_m^{\text{MP}} - x_m^{\text{GM}}\|_2 \leq [\lambda_{m+1}^{\text{MP}} / (\sigma_m^2 - \lambda_{m+1}^{\text{MP}})] \|x_m^{\text{GM}}\|_2.$$

An auxiliary result proved by these authors shows that the eigenvalues determined at the step m interlace in the following way:

$$\lambda_i^{\text{MP}} \geq \sigma_i^2 \geq \lambda_{i+1}^{\text{MP}}, \quad \text{for } i = 1, \dots, m.$$

The bound from the above theorem is large whenever σ_m^2 is close to $\lambda_{m+1}^{\text{MP}}$ and the result cannot be applied when $\sigma_m^2 = \lambda_{m+1}^{\text{MP}}$; such a situation arises for example when $\lambda_{m+1}^{\text{MP}}$ is a repeated eigenvalue.

Theorem 4.5 may determine theoretical possibilities when GMRES and MINPERT have the same asymptotical behavior when used in a converging Newton method.

Theorem 4.6. Assume that Conditions (C1)–(C3) hold and that the sequence $(y_k^{\text{GM}})_{k \geq 0}$ given by the Newton–GMRES method converges to y^* , where the elements from the GMRES algorithm are arbitrary at each outer step. Denote by s_k^{GM} the corrections obtained from the linear systems

$$F'(y_k^{\text{GM}})s = -F(y_k^{\text{GM}}),$$

and consider the approximate solutions s_k^{MP} of these linear systems obtained with MINPERT, which is assumed to use the same initial approximations, the same subspace dimensions, and the same number of restarts as GMRES. If $\sigma_k^2 \neq \lambda_k^{\text{MP}}$, for $k = 0, 1, \dots$, and $\lambda_k^{\text{MP}} = o(\sigma_k^2)$, as $k \rightarrow \infty$, then GMRES and MINPERT yield asymptotically the same normalized corrections:

$$[1/\|s_k^{\text{GM}}\|_2]s_k^{\text{GM}} - [1/\|s_k^{\text{MP}}\|_2]s_k^{\text{MP}} \rightarrow 0, \quad \text{as } k \rightarrow \infty.$$

The same result holds inverting the role of GMRES and MINPERT.

Proof. The hypotheses of the theorem imply that

$$\lambda_k^{\text{MP}} \rightarrow 0, \quad \text{as } k \rightarrow \infty.$$

For the nonrestarted version of the Krylov solvers, the affirmation is a straightforward application of Theorem 4.5, while for the restarted version, the conclusion is obtained by an inductive argument on the number of restarts from each outer step. \square

5. Numerical Examples

We shall consider two test problems from Refs. 34 and 42, that will provide some nonlinear systems in \mathbb{R}^N . We shall apply to them the studied Newton–Krylov methods. The Krylov solvers are considered in the non-restarted version and the initial guess is taken 0 in both the inner and outer iterations. We are interested in the behavior of the magnitude of the backward errors of these methods and in verifying theoretical results; therefore, we are not aiming to implement some efficient methods for solving the problems under consideration.

5.1. Bratu Problem. Consider the nonlinear partial differential equation

$$-\Delta u + \alpha u_x + \lambda e^u = f,$$

over the unit square of \mathbb{R}^2 , with Dirichlet boundary conditions. As mentioned in Refs. 34 and 42, this is a standard problem, a simplified form of which is known as the Bratu problem (see Ref. 43). We have discretized by means of 5-point finite differences [respectively, by means of central finite differences] on a uniform mesh, obtaining a system of nonlinear equations of size $N = (n-2)^2$, where n is the number of mesh points in each direction. As in Ref. 34, we took f such that the solution of the discretized problem is the constant unity. We have considered

$$\alpha = 10, \quad \lambda = 1, \quad N = 1024, \quad m = 10.$$

The runs were made on a PC, using MATLAB Version 4.0. The symbol $\|\cdot\|$ denotes either the Euclidean norm or the Frobenius norm, and e_k , F_k stand for $y^* - y_k$, $F(y_k)$.

Table 1 contains the results obtained by using the Newton–GMRES method. We have also considered the corrections obtained with MINPERT at each step, denoting by a_k the norm of the difference $(1/\|s_k^{\text{GM}}\|)s_k^{\text{GM}} - (1/\|s_k^{\text{MP}}\|)s_k^{\text{MP}}$; the corrections s_k^{MP} were computed only for the comparison. It can be seen that the normalized corrections agree the closer the iterations get to the solution.

Table 2 contains the results obtained by using the Newton–GMBACK method. We notice the rather constant magnitudes of $\Delta_{A_k}^{\text{GB}}$, which do not approach zero even when the iterates are close to the solution.

In Table 3, we have denoted

$$b_k = \|s_{k,10}^{\text{GM}} - s_{k,10}^{\text{MP}}\|_2 / \|s_{k,10}^{\text{GM}}\|_2, \quad c_k = \lambda_{k,10}^{\text{MP}} / (\sigma_{k,10}^2 - \lambda_{k,10}^{\text{MP}}).$$

The bounds c_k seem to be tight here for b_k . As soon as the iterations approach the solution, the inequalities from Theorem 4.5 cease to hold numerically; we believe that this is due to the different types of errors which

Table 1. Newton–GMRES applied to the Bratu problem.

k	$\ e_k\ $	$\ F_k\ ^2$	$\ r_k\ $	a_k	k	$\ e_k\ $	$\ F_k\ ^2$	$\ r_k\ $	a_k
0	3e+1	1e+2	9e-1	4e-2	11	3e-4	1e-10	8e-6	1e-9
1	2e+1	9e-1	5e-1	3e-2	12	2e-4	7e-11	2e-6	4e-11
2	1e+1	3e-1	4e-1	2e-2	13	3e-5	8e-12	7e-7	4e-12
3	1e+1	1e-1	2e-1	9e-3	14	1e-5	5e-13	4e-7	3e-12
4	6e+0	8e-2	3e-2	9e-5	15	1e-5	2e-13	2e-7	3e-13
5	1e-1	1e-3	6e-3	2e-4	16	1e-6	4e-14	2e-8	1e-14
6	5e-2	4e-5	2e-3	5e-5	17	5e-7	6e-16	1e-8	5e-14
7	2e-2	5e-6	6e-4	2e-6	18	4e-7	1e-16	6e-9	3e-15
8	1e-2	4e-7	3e-4	2e-6	19	8e-8	4e-17	5e-10	2e-15
9	1e-2	1e-7	1e-4	2e-7	20	9e-9	2e-19	2e-10	4e-14
10	1e-3	2e-8	1e-5	4e-10	21	6e-9	5e-20	9e-11	2e-16

Table 2. Newton–GMBACK applied to the Bratu problem.

k	$\ e_k\ $	$\ F_k\ $	$\ \Delta_{A_k}^{\text{GB}}\ $	r_k	k	$\ e_k\ $	$\ F_k\ $	$\ \Delta_{A_k}^{\text{GB}}\ $	r_k
0	3e+1	1e+1	6e-2	1e+0	11	1e-4	2e-5	4e-2	4e-6
1	2e+1	1e+0	5e-2	7e-1	12	8e-5	4e-6	5e-2	3e-6
2	1e+1	7e-1	6e-2	5e-1	13	3e-5	3e-6	1e-2	5e-7
3	9e+0	5e-1	6e-2	3e-1	14	8e-6	5e-7	6e-2	3e-7
4	4e+0	3e-1	2e-2	7e-2	15	4e-6	3e-7	2e-2	8e-8
5	1e+0	7e-2	4e-2	6e-2	16	1e-6	8e-8	5e-2	5e-8
6	5e-1	6e-2	2e-2	1e-2	17	5e-7	5e-8	2e-2	1e-8
7	2e-1	1e-2	1e-2	4e-3	18	2e-7	1e-8	3e-2	6e-9
8	1e-2	4e-3	5e-2	5e-4	19	3e-8	6e-9	2e-2	7e-10
9	4e-3	5e-4	4e-2	9e-5	20	1e-8	7e-10	2e-2	3e-10
10	2e-3	9e-5	1e-2	2e-5	21	1e-9	3e-10	1e-2	1e-11

Table 3. Newton–MINPERT applied to the Bratu problem.

k	$\ e_k\ $	$\ F_k\ $	$\sqrt{\lambda_{k,10}^{\text{MP}}}$	$\ F_k\ ^2$	$\ r_k\ $	b_k	c_k
0	3e+1	1e+1	6e-2	1e+2	1e+0	8.3594e-2	1.0364e-1
1	2e+1	1e+0	5e-2	1e+0	7e-1	6.6518e-1	6.7180e-1
2	1e+1	7e-1	6e-2	5e-1	5e-1	7.3033e-1	7.3995e-1
3	9e+0	5e-1	6e-2	2e-1	3e-1	6.1259e-1	6.2053e-1
4	4e+0	3e-1	2e-2	1e-1	7e-2	4.4297e-2	4.4403e-2
5	1e+0	7e-2	3e-2	5e-3	5e-2	3.3261e-1	3.3456e-1
6	8e-1	5e-2	1e-2	2e-3	1e-2	3.0693e-2	3.0808e-2
7	3e-1	1e-2	5e-3	2e-4	5e-3	1.5905e-2	1.6004e-2
8	6e-2	5e-3	1e-3	3e-5	1e-3	2.0734e-4	2.0883e-4
9	2e-2	1e-3	7e-4	2e-6	7e-4	1.1074e-4	1.1233e-4
10	1e-2	7e-4	1e-4	5e-7	1e-4	1.0523e-5	1.0526e-5
11	2e-3	1e-4	7e-5	3e-8	7e-5	5.2230e-7	5.3930e-7
12	1e-3	7e-5	6e-6	5e-9	6e-6	2.6605e-8	2.6605e-8
13	4e-5	6e-6	1e-6	4e-11	1e-6	3.1452e-11	3.2821e-11
14	2e-5	1e-6	6e-7	1e-12	6e-7	1.6474e-10	1.6433e-10
15	9e-6	6e-7	1e-7	3e-13	1e-7	4.9880e-12	5.0888e-12
16	2e-6	1e-7	8e-8	2e-14	8e-8	1.6094e-12	1.6504e-12
17	1e-6	8e-8	1e-8	6e-15	1e-8	1.6952e-13	8.3076e-14
18	1e-7	1e-8	4e-9	1e-16	4e-9	1.2156e-14	1.7097e-15
19	1e-7	4e-9	1e-9	2e-17	1e-9	4.7003e-13	9.6014e-16
20	5e-9	1e-9	2e-11	1e-18	2e-11	2.2673e-14	2.2768e-20

have appeared in computing the eigenpairs, the Krylov approximations, and the elements b_k , c_k .

5.2. Driven Cavity Flow Problem. This is a classical problem from incompressible fluid flow. It has the following equations in the stream

function-vorticity formulation:

$$v\Delta\omega + \psi_y\omega_x - \psi_x\omega_y = 0, \quad \text{in } \Omega, \quad (10)$$

$$-\Delta\psi = \omega, \quad \text{in } \Omega, \quad (11)$$

$$\psi = 0, \quad \text{on } \partial\Omega,$$

$$(\partial\psi/\partial n)(x, y) \Big|_{\partial\Omega} = \begin{cases} 1, & \text{if } y = 1, \\ 0, & \text{if } 0 \leq y < 1, \end{cases}$$

where Ω denotes again the interior of the unit square from \mathbb{R}^2 and the viscosity v is the reciprocal of the Reynolds number Re . In terms of ψ alone, (10) and (11) are replaced by

$$v\Delta^2\psi + \psi_y(\Delta\psi)_x - \psi_x(\Delta\psi)_y = 0, \quad \text{in } \Omega.$$

This equation was discretized again on a uniform mesh. We have considered $n = 25$, choosing the boundary conditions to be incorporated in the nonlinear system. We have used the symbolic facilities offered by Maple V Release 3.0 in order to determine the discretized equations. The Reynolds number was taken small ($Re = 10$) and the subspace dimensions of the Krylov methods were considered large ($m = 50$), since the nonrestarted versions of these methods were not efficient for this problem.

Tables 4–5 contain the results of the three Newton–Krylov methods applied to this problem. We notice again that, close to the solution, the computed elements b_k and c_k do not satisfy the inequalities $b_k \leq c_k$.

Table 4. Newton–GMRES (left) and Newton–GMBACK (right) applied to the driven cavity problem.

k	$\ F_k\ $	$\ F_k\ ^2$	$\ r_k\ $	a_k	k	$\ F_k\ $	$\ \Delta_{\mathcal{A}_k}^{\text{GB}}\ $	$\ F_k\ ^2$	r_k
0	4e+0	2e+1	9e-1	2e-2	0	4e+0	6e-2	2e+1	9e-1
1	8e+0	7e+1	1e+0	7e-2	1	9e+0	3e-1	9e+1	1e+0
2	4e+0	2e+1	8e-1	4e-2	2	6e+0	3e-1	3e+1	1e+0
3	1e+0	3e+0	6e-1	4e-2	3	1e+0	3e-1	3e+0	9e-1
4	7e-1	5e-1	5e-1	1e-2	4	9e-1	2e-1	8e-1	8e-1
...					...				
131	9e-5	8e-9	8e-5	6e-9	131	4e-5	6e-1	1e-9	2e-5
132	8e-5	7e-9	8e-5	7e-9	132	2e-5	4e-1	5e-10	3e-5
133	8e-5	6e-9	7e-5	5e-9	133	3e-5	6e-1	1e-9	2e-5
134	7e-5	6e-9	7e-5	5e-9	134	2e-5	4e-1	4e-10	3e-5
135	7e-5	5e-9	6e-5	4e-9	135	3e-5	6e-1	1e-9	1e-5

Table 5. Newton–MINPERT applied to the driven cavity problem.

k	$\ F_k\ $	$\sqrt{\lambda_{k,50}^{\text{MP}}}$	$\ F_k\ ^2$	$\ r_k\ $	b_k	c_k
0	4e+0	6e-2	2e+1	9e-1	7.0262e-2	7.8592e-2
1	9e+0	3e-1	9e+1	1e+0	3.3821e-1	3.5336e-1
2	6e+0	3e-1	3e+1	1e+0	8.0633e-1	8.2988e-1
3	1e+0	2e-1	3e+0	8e-1	1.1452e+0	1.1887e+0
4	8e-1	2e-1	7e-1	7e-1	1.6037e+0	1.6438e+0
...						
131	1e-5	1e-5	2e-10	1e-5	4.0286e-8	4.0292e-8
132	1e-5	1e-5	2e-10	1e-5	3.2414e-9	3.2489e-9
133	1e-5	1e-5	2e-10	1e-5	3.6246e-8	3.6241e-8
134	1e-5	1e-5	1e-10	1e-5	2.4452e-9	2.4576e-9
135	1e-5	1e-5	1e-10	1e-5	2.1203e-8	2.1201e-8

In Fig. 1, we have plotted on a semilog scale the evolution of $\|F_k\|$ for the three methods. The first 150 steps were considered. Notice the monotone behavior of Newton–GMRES and Newton–MINPERT starting from certain steps, and also notice the nonmonotone behavior of Newton–GMBACK.

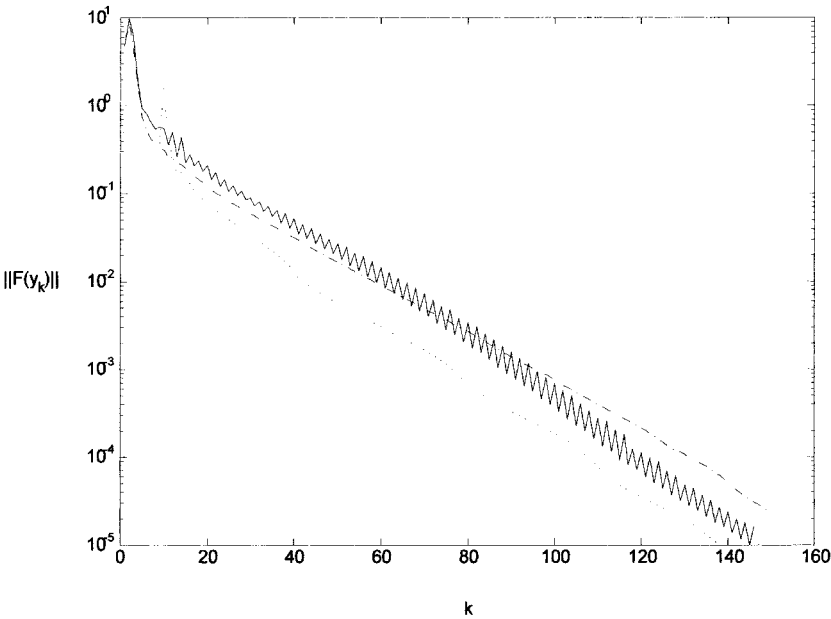


Fig. 1. Newton–GMRES (dash-dot line), Newton–GMBACK (solid line), and Newton–MINPERT (dotted line) applied to the driven cavity problem.

6. Conclusions

The local superlinear convergence and local convergence with orders $1 + p$, $p \in (0, 1]$, of the Newton methods were characterized in a more natural setting, which assumes that the Jacobians are perturbed, the function evaluations are performed approximately for F , and the resulting linear systems are solved inexactly. The results of Dembo, Eisenstat, and Steihaug allowed us to extend their local convergence analysis to this setting.

The sufficient conditions for local convergence with different convergence orders show that the perturbations in the Jacobians may be allowed to have much greater magnitudes than those in the function evaluations and than the magnitudes of the residuals; this may be explained by the fact that the right-hand sides $-F(x_k)$ of the linear systems from the Newton method tend to zero, while the matrices $F'(x_k)$ tend to $F'(x^*)$, assumed here to be nonsingular. It follows that the methods for solving the linear systems from the Newton methods must be analyzed in this respect and also that special care must be taken when the function evaluations may be affected by significant errors. The greater sensitivity of the right-hand sides $-F(x_k)$ has been known for a longer time (see for example Refs. 11, 21, and references therein) such that the evaluation of these vectors in double precision or extended precision is already a standard.

The existing results for the (INM) iterates allowed us to show that the Newton–GMRES iterates have a monotone property. We were able to prove that the Newton–MINPERT iterates share this property only under some additional conditions, whereas from a concrete example we saw that Newton–GMBACK iterates do not generally share this property. The convergence orders of the Newton–GMBACK and Newton–MINPERT methods can be controlled in terms of the magnitude of the backward errors of the approximate steps. The theoretical results were confronted with the performed numerical examples.

Many of the existing results for different Newton-type methods may be reconsidered in the (IPNM) setting. For example, the local convergence orders of some finite-difference Newton–Krylov methods studied by Brown (Ref. 26) now may be analyzed in the (IPNM) setting. On the other hand, new approaches can be developed as well; for instance, we should mention the local convergence and acceleration techniques for the (IPNM) in the case of singular Jacobian at the solution. Other results and directions of research are included in our PhD Thesis (Ref. 44).

References

1. DEMBO, R. S., EISENSTAT, S. C., and STEIHAUG, T., *Inexact Newton Methods*, SIAM Journal on Numerical Analysis, Vol. 19, pp. 400–408, 1982.

2. ORTEGA, J. M., and RHEINBOLDT, W. C., *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, NY, 1970.
3. KANTOROVICH, L. V., and AKILOV, G. P., *Functional Analysis*, Editura Științifică și Enciclopedică, Bucharest, Romania, 1986 (in Romanian).
4. PĂVĂLOIU, I., *Introduction to the Theory of Approximating the Solutions of Equations*, Editura Dacia, Cluj-Napoca, Romania, 1976 (in Romanian).
5. MĂRUȘER, ȘT., *Numerical Methods for Solving Nonlinear Equations*, Editura Tehnică, Bucharest, Romania, 1981 (in Romanian).
6. DENNIS, J. E., JR., and SCHNABEL, R. B., *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Series in Computational Mathematics, Prentice-Hall, Englewood Cliffs, New Jersey, 1983.
7. POTRA, F. A., and PTÁK, V., *Nondiscrete Induction and Iterative Processes*, Pitman, London, England, 1984.
8. ARGYROS, I., and SZIDAROVSKY, F., *The Theory and Applications of Iteration Methods*, CRC Press, Boca Raton, Florida, 1993.
9. KELLEY, C. T., *Iterative Methods for Linear and Nonlinear Equations*, SIAM, Philadelphia, Pennsylvania, 1995.
10. RHEINBOLDT, W. C., *Methods for Solving Systems of Nonlinear Equations*, SIAM, Philadelphia, Pennsylvania, 1998.
11. DENNIS, J. E., JR., and WALKER, H. F., *Inaccuracy in Quasi-Newton Methods: Local Improvement Theorems*, Mathematical Programming Study, Vol. 22, pp. 70–85, 1984.
12. POTRA, F. A., and PTÁK, V., *Sharp Error Bounds for Newton's Process*, Numerische Mathematik, Vol. 34, pp. 63–72, 1980.
13. DEUFLHARD, P., and POTRA, F. A., *Asymptotic Mesh Independence of Newton–Galerkin Methods via a Refined Mysovskii Theorem*, SIAM Journal on Numerical Analysis, Vol. 29, pp. 1395–1412, 1992.
14. ARGYROS, I., *Concerning the Convergence of Inexact Newton Methods*, Journal of Computational and Applied Mathematics, Vol. 79, pp. 235–247, 1997.
15. ARGYROS, I., *On a New Newton–Mysovskii-Type Theorem with Applications to Inexact Newton-Like Methods and Their Discretization*, IMA Journal of Numerical Analysis, Vol. 18, pp. 37–56, 1998.
16. SHERMAN, A. H., *On Newton-Iterative Methods for the Solution of Systems of Nonlinear Equations*, SIAM Journal on Numerical Analysis, Vol. 15, pp. 755–771, 1978.
17. DENNIS, J. E., JR., and MORÉ, J. J., *A Characterization of Superlinear Convergence and Its Application to Quasi-Newton Methods*, Mathematics of Computation, Vol. 28, pp. 549–560, 1974.
18. DENNIS, J. E., JR., and MORÉ, J. J., *Quasi-Newton Methods, Motivation and Theory*, SIAM Review, Vol. 19, pp. 46–89, 1977.
19. EISENSTAT, S. C., and WALKER, H. F., *Choosing the Forcing Terms in an Inexact Newton Method*, SIAM Journal on Statistical and Scientific Computing, Vol. 17, pp. 16–32, 1996.
20. EISENSTAT, S. C., and WALKER, H. F., *Globally Convergent Inexact Newton Methods*, SIAM Journal on Optimization, Vol. 4, pp. 393–422, 1994.

21. YPMA, T. J., *The Effect of Rounding Errors on Newton-Like Methods*, IMA Journal of Numerical Analysis, Vol. 3, pp. 109–118, 1983.
22. MARTINEZ, H. J., PARADA, Z., and TAPIA, R. A., *On the Characterization of Q -Superlinear Convergence of Quasi-Newton Interior-Point Methods for Nonlinear Programming*, Boletín de la Sociedad Matematica Mexicana, Vol. 1, pp. 137–148, 1995.
23. YPMA, T. J., *Local Convergence of Inexact Newton Methods*, SIAM Journal on Numerical Analysis, Vol. 21, pp. 583–590, 1983.
24. CORES, D., and TAPIA, R. A., *Perturbation Lemma for the Newton Method with Application to the SQP Newton Method*, Journal of Optimization Theory and Applications, Vol. 97, pp. 271–280, 1998.
25. POTRA, F. A., *On Q -Order and R -Order of Convergence*, Journal of Optimization Theory and Applications, Vol. 63, pp. 415–431, 1989.
26. BROWN, P. N., *A Local Convergence Theory for Combined Inexact-Newton/Finite-Difference Projection Methods*, SIAM Journal on Numerical Analysis, Vol. 24, pp. 407–434, 1987.
27. CĂȚINAȘ, E., *A Note on Inexact Secant Methods*, Revue d'Analyse Numérique et de Théorie de l'Approximation, Vol. 25, pp. 33–41, 1996.
28. KRÓLIKOWSKA, M., *Nonlinear Mappings with an Almost Sparse Jacobian Matrix*, Annales Universitatis Mariae Curie-Skłodowska, Lublin, Poland, Vol. 49A, pp. 145–157, 1995.
29. RIGAL, J. L., and GACHES, J., *On the Compatibility of a Given Solution with the Data of a Linear System*, Journal of the Association for Computing Machinery, Vol. 14, pp. 543–548, 1967.
30. SAAD, Y., and SCHULTZ, M. H., *GMRES: A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Systems*, SIAM Journal on Scientific and Statistical Computing, Vol. 7, pp. 856–869, 1986.
31. SAAD, Y., *Krylov Subspace Methods for Solving Large Unsymmetric Linear Systems*, Mathematics of Computation, Vol. 37, pp. 105–126, 1981.
32. KASENALLY, E. M., *GMBACK: A Generalized Minimum Backward Error Algorithm for Nonsymmetric Linear Systems*, SIAM Journal on Statistical and Scientific Computing, Vol. 16, pp. 698–719, 1995.
33. SAAD, Y., *Iterative Methods for Sparse Linear Systems*, PWS Publishing Company, Boston, Massachusetts, 1996.
34. BROWN, P. N., and SAAD, Y., *Hybrid Krylov Methods for Nonlinear Systems of Equations*, SIAM Journal on Scientific and Statistical Computing, Vol. 11, pp. 450–481, 1990.
35. TURNER, K., and WALKER, H. F., *Efficient High Accuracy Solutions with GMRES(m)*, SIAM Journal on Scientific and Statistical Computing, Vol. 13, pp. 815–825, 1992.
36. BROWN, P. N., and HINDMARSH, A. C., *Reduced Storage Matrix Methods in Stiff ODE Systems*, Applied Mathematics and Computation, Vol. 31, pp. 40–91, 1989.
37. GREENBAUM, A., PTÁK, V., and STRAKOŠ, Z., *Any Nonincreasing Convergence Curve is Possible for GMRES*, SIAM Journal on Matrix Analysis and Applications, Vol. 17, pp. 465–469, 1996.

38. BROWN, P. N., and SAAD, Y., *Convergence Theory of Nonlinear Newton–Krylov Algorithms*, SIAM Journal on Optimization, Vol. 4, pp. 297–330, 1994.
39. BROWN, P. N., *A Theoretical Comparison of the Arnoldi and GMRES Algorithms*, SIAM Journal on Scientific and Statistical Computing, Vol. 12, pp. 58–78, 1991.
40. KASENALLY, E. M., and SIMONCINI, V., *Analysis of a Minimum Perturbation Algorithm for Nonsymmetric Linear Systems*, SIAM Journal on Numerical Analysis, Vol. 34, pp. 48–66, 1997.
41. STEWART, G. W., and SUN, J. G., *Matrix Perturbation Theory*, Academic Press, New York, NY, 1990.
42. MORÉ, J. J., *A Collection of Nonlinear Model Problems*, Computational Solutions of Nonlinear Systems of Equations, Lectures in Applied Mathematics, Edited by E. L. Allgower and K. Georg, American Mathematical Society, Providence, Rhode Island, Vol. 26, pp. 723–762, 1990.
43. GLOWINSKY, R., KELLER, H. B., and REINHART, L., *Continuation-Conjugate Gradient Methods for the Least-Squares Solution of Nonlinear Boundary-Value Problems*, SIAM Journal on Scientific and Statistical Computing, Vol. 6, pp. 793–832, 1985.
44. CĂȚINAȘ, E., *Newton and Newton–Krylov Methods for Solving Nonlinear Systems in \mathbb{R}^n* , PhD Thesis, Babeș–Bolyai University of Cluj–Napoca, Cluj–Napoca, Romania, 1999.