Chapter 1

Krylov methods for large linear systems

1.1 Motivation

We shall first motivate the title of this chapter, since the people not dealing with numerical analysis may naturally ask: "which is the role of such a study, since the Cramer formulas and the Gauss method are known for such a long time, and may be applied to any nonsingular linear system?".

The answer to this question is dictated by practical considerations, specific to the numerical analysis. More precisely, both of the mentioned methods present major impediments when we try to use them with computers, for large number of unknowns.

Below we present the situation arising if we want to solve with the computer a linear system of dimension $n = 100$ using the Cramer formulas, without optimizing the operations:

> Following the argument from [30, p.311], let us imagine that we have a computer occupying a volume V, formed by cubic elements of side l, which performs parallel operations. Let us admit that the time

required by an element for performing an elementary arithmetic operation is $t = l/c$, $c = 3 \cdot 10^8$ m/s and that there does not appear the problem of transferring the information between the elements of this computer. In such a case, the amount of elementary arithmetic operations performed in a second is

$$\text{No. of op.} = \frac{\text{No. of elem.}}{\text{Time required by an elem.}} = \frac{V}{l^3} / \frac{l}{c} = \frac{cV}{l^4}.$$

For $V = 1$ km^3 and $l = 10^{-8}$ cm (the order of the size of an atom) we get

$$\text{No. of op.} = \frac{3 \cdot 10^8 \cdot 10^9}{10^{-40}} = 3 \cdot 10^{57}.$$

Let us admit that the Cramer formulas for the considered system require the performing of only 100! elementary arithmetic operations. Since $100! = 10^{157,9\ldots}$, this means that this computer would need approximately 10^{94} years to compute the solution of the system.

The Cramer formulas are not used for practical problems even in the case of small numbers of unknowns, because the number of elementary operations (if admitting that are just of order $\mathcal{O}(n!)$) is much higher than of other direct methods, which require only $\mathcal{O}\left(n^3\right)$ operations. As an example, Ciarlet [73, p.81] mentions that for a linear system with $n = 10$ unknowns, the Gauss method requires 700 operations, while the Cramer rule requires 400.000.000 operations. On the other hand, the size of the cumulated errors in the Cramer rule may lead in floating point arithmetic to meaningless solutions.

Let us see now the drawbacks of the Gauss method when used at the computer, being known that the representation of the real numbers and the floating point operations are inherently performed with errors. In the case of the partial pivoting variant for solving the system $Ax = b$, it is known that the relative error of the obtained "solution" $\tilde{x}$ is bounded in the following way (see [108] and [137, ch.9]):

$$\frac{\|x^* - \tilde{x}\|_\infty}{\|x^*\|_\infty} \leq 4n^2 \rho \epsilon \cdot \kappa_\infty(A),$$

where

- n is the dimension of the system;
- ϵ is the machine epsilon (the exact upper bound of the relative errors appearing in the representation of the real numbers and in performing the elementary arithmetic operations in floating point arithmetic);
- $\kappa_\infty(A)$ is the condition number of the matrix A in the Chebyshev norm;
- ρ is a parameter specific to any matrix, its maximum value being 2^{n-1}; Wilkinson has shown in 1965 that this bound is exact and has provided theoretical examples of matrices when this bound is attained, but for the majority of the practical problems the value of ρ is small.[1]

The above relation shows in a clear way that the floating point solution computed by the Gauss method may be far away from the exact one when the number of unknowns is large or when the matrix A is ill conditioned.

On the other hand, the number of elementary arithmetic operations performed by the algorithm makes that the time required to be exaggerated long as n increases.

For $n \geq 100.000$ no linear system with "full" matrix has been solved by direct methods (cf. [226, p.339]); the "record" seems to be attained by a system having dimension $n = 76.800$. Most often, the large linear systems

[1] Wilkinson has further stated that in all his activity he has not found practical applications for which the amplification factor to be larger than 16. This was a challenge for those willing to find counterexamples. Only in 1993 Wright (cf. [137, ch.9]) has found a class of two point boundary value problems for ordinary differential equations which, if solved by the multiple shooting method, leads to linear systems for which the partial pivoting has exponential growth of the error; Foster [108] has shown next in 1994 that applying to a Volterra integral equation often arising in practice a certain quadrature method, we obtain exponential growth in ρ when solving by partial pivoting. Despite of these, Trefethen and Bau [226, p.167] have offered a preliminary argumentation, through statistical considerations, according to which the amplification factor has small values in the majority of the practical situations.

have sparse matrices, coming from different discretizations. The Gauss method does not take into account such structures, and its different variants as well as other direct methods for sparse matrices have not imposed.

The Strassen and resp. Coppersmith and Winograd methods require just $\mathcal{O}\left(n^{2,81\ldots}\right)$ resp. $\mathcal{O}\left(n^{2,37\ldots}\right)$ operations, but they had until present only a more theoretical impact (cf. [226, p.247]). The closeness between $n^{2,81\ldots}$ and n^3 makes that the difference between the efficiency of the Strassen and Gauss methods to become essential for such large values of n, that such systems are not approachable with the computers from today. The methods with exponents smaller than $2,81$ have the constant factors from the asymptotic expressions so large that they are more inefficient that the Gauss method (again for the systems representable in the nowadays computers). On the other hand, the stability of this type of methods is very less understood.

1.2 Krylov methods based on backward error minimization properties

Consider the linear system

$$Ax = b, \tag{1.1}$$

where $A \in \mathbb{R}^{N\times N}$ is nonsingular and $b \in \mathbb{R}^N$. The Krylov methods for solving such systems when the dimension N is large, are methods based on the Krylov subspaces – defined for any initial approximation $x_0 \in \mathbb{R}^N$ and for any value $m \in \{1,\ldots,N\}$ as

$$\mathcal{K}_m = \mathcal{K}_m\left(A, r_0\right) = \operatorname{span}\left\{r_0, Ar_0, \ldots, A^{m-1}r_0\right\},$$

where $r_0 = b - Ax_0$ is the residual of x_0. We shall assume in the following that the matrix A is unstructured. It is known however that the Krylov methods for symmetric and/or positive (semi)definite matrices have a behavior better understood than those for the general case (see, e.g., [121], [89], [27]).

The Krylov are regarded as iterative methods (some authors call them semiiterative methods); though, unlike the Jacobi, Gauss-Seidel and other iterations, the exact solution may be computed (in exact arithmetic) in at most N steps, for any initial approximation.

The study of the iterative methods of this kind has begun with the paper of Hestenes and Stiefel [136] from 1952, who introduced the conjugate gradient method (cf. [27, p.451] and [226, p.341]).[2] Lanczos [154] has introduced the iterations bearing his name two years before, these iterations being connected to the conjugate gradient methods, but Hestenes and Stiefel have independently given the standard formulation of this method (cf. [226, p.341]). Krylov subspaces seem to be associated to the paper [153].[3] Regarding the Arnodi algorithm, the original paper [25] dates since 1951, the intentions from that paper being however far from its present uses (cf. [226, p.340]).

The clear advantages of the Krylov methods were recognized and exploited from the seventies, when the development of the computational tools have permitted the approach of large linear systems, in the present accept of the notion.

The efficiency of the Krylov methods consists in the following aspects:

- *For large values of N, one may obtain in many situations satisfactory approximations, by performing a small amount of steps and computations.* We present in the following a justification of this fact given by Ipsen and Meyer [141].

  The minimal polynomial $P(t)$ of the matrix A is the unique monic polynomial (i.e., having the coefficient of the maximum degree term equal to 1) of minimal degree for which $P(A) = 0$. It can be constructed with the aid of the eigenvalues of A as follows. Denoting the distinct eigenvalues of A by [4] $\lambda_1, \ldots, \lambda_d$ and

[2] Concerning the history of these methods one may also consult [118].

[3] In [134] there are expressed some doubts concerning this.

[4] Ipsen and Meyer do not explicitly mention this, but for the real case it is necessary for the matrix A to have N real eigenvalues (counting the multiplicity orders).

if λ_i has index m_i (the dimension of the largest Jordan block associated to λ_i), then

$$M = \sum_{i=1}^{d} m_i \quad \text{and } P(t) = \prod_{i=1}^{d} (t - \lambda_i)^{m_i}.$$

Writing $P(t) = \sum_{i=0}^{M} \alpha_i t^i$, then the last term is $\alpha_0 = \prod_{i=1}^{d} (-\lambda_i)^{m_i}$ and so $\alpha_0 \neq 0 \Leftrightarrow \exists A^{-1}$.
It follows next that

$$A^{-1} = -\tfrac{1}{\alpha_0} \sum_{i=0}^{M-1} \alpha_{i+1} A^i,$$

and, consequently, the smaller the degree of the minimal polynomial is, the shorter is the description of A^{-1}. The connection with the Krylov subspaces is immediately obtained.

THEOREM 1.1 [141] *If the minimal polynomial of the nonsingular matrix A has degree M, then the solution of the system $Ax = b$ is contained in the subspace $\mathcal{K}_M(A, b)$.*

We remark that $\mathcal{K}_M(A, b)$ is the Krylov subspace corresponding to the initial approximation $x_0 = 0$.

- *The Krylov methods do not require the storage of the matrix A in the computer memory, for these algorithms being necessary only matrix-vector products of the form Av, for certain values $v \in \mathbb{R}^N$.*

- *In the case when the matrix A is rare, the speed of computing Av is considerably increased.*

At present, the solving of the general linear systems by this type of methods constitutes a domain of assiduous research, the literature containing an

abundance of results on this theme. There exists two main classes of such methods (cf. [226, p.305]). The first class is formed by the *Hessenberg orthogonalization methods*; after a modified Gram-Schmidt orthogonalization (the Arnoldi algorithm) one obtains, in the notations below, the factorization $A = V_{m+1}\bar{H}_m V_m^t$ (the matrix $\bar{H}_m$ is upper Hessenberg and the matrix $V_{m+1} = [V_m\ v_{m+1}]$ contains on its columns orthogonal normed vectors). The most representative algorithm of this class is GMRES, introduced by Saad and Schultz [211] in 1986. The second class contains the *tridiagonal biorthogonalization methods*: $A = QTQ^{-1}$, where T is a tridiagonal matrix and Q is a matrix which generally is not orthogonal; the term of biorthogonalization refers to the fact that the columns of Q are orthogonal to the columns of the inverse of the adjoint of Q [226, p.305]. The algorithms from this class are based on three term recurrence formulas; among the most representatives we mention BiCGStab, QMR and TFQMR, introduced by Vorst [232] in 1992, Freund and Nachtigal [110] in 1991 and resp. by Freund [109] in 1993.

In the case of symmetric matrices (or hermitian, in the complex case) the tridiagonal orthogonalization is possible (the conjugate gradient method, the Lanczos method) but in the general case, one must renounce either to orthogonalization or to tridiagonalization.

In this chapter we shall study the GMRES, GMBACK and MINPERT methods. According to the above classification, they belong to the first category, and we shall see that they are based on backward error minimization properties.

1.2.1 The GMRES method

Given an initial approximation $x_0 \in \mathbb{R}^N$ to the solution x^* of the linear system (1.1), the GMRES method uses the Arnoldi process for constructing an orthonormal basis $\{v_1, \ldots, v_m\}$ in the subspace $\mathcal{K}_m$. By the exact solving of a least squares problem in $\mathbb{R}^m$, one determines the approximate solution $x_m^{GM} \in \mathbb{R}^N$ satisfying:

$$\left\|b - Ax_m^{GM}\right\|_2 = \min_{x_m \in x_0 + \mathcal{K}_m} \|b - Ax_m\|_2 = \min_{z \in \mathcal{K}_m} \|r_0 - Az\|_2 . \tag{1.2}$$

Saad and Schultz have considered this problem also in the approximation theory framework, obtaining some error bounds. From this viewpoint, the GMRES method with $x_0 = 0$ solves the following problem: [5]

$$\left\| p_{m,b}^{GM}(A)\, b \right\|_2 = \min_{p_m \in \mathbb{P}_m} \left\| p_m(A)\, b \right\|_2 ,$$

where $\mathbb{P}_m = \{p : p \text{ polynomial of degree} \leq m \text{ with } p(0) = 1\}$; the polynomial $p_{m,b}^{GM}$ always exists, being uniquely determined if the above minimum is nonzero.

"The convergence speed " of GMRES attained when increasing m (see also Proposition 1.3 below) depends both on the matrix A and on the vector b. However, in practical situations, apart of some cases when b has a special structure, it seems that the matrix A is the one which predominantly determines the convergence speed. Greenbaum and Trefethen [124] have considered the "ideal GMRES problem"

$$\left\| p_m^{GM}(A) \right\|_2 = \min_{p_m \in \mathbb{P}_m} \left\| p_m(A) \right\|_2 .$$

The polynomial p_b^{GM} always exists, being uniquely determined if the above minimum is nonzero. The norm of the error from the ideal approximation problem constitutes an upper bound for the error norm corresponding to the "real" case, for any $m = 1, \ldots, N$ (see [124]):

$$\max_{b \in \mathbb{C}^N,\, \|b\|_2 = 1} \left\| p_{m,b}^{GM}(A)\, b \right\|_2 \leq \left\| p_m^{GM}(A) \right\|_2 .$$

It is known that this inequality becomes equality when A belongs to different classes of matrices (normal, triangular Toeplitz, etc.) and also when A is arbitrary and $m = 1$ (see [124], [223], [121] and the references therein). Such results led Greenbaum and Trefethen to the conjecture that in the above relation one has equality for any matrix and iteration step m.

[5] The following relations concerning this aspect remain essentially the same if the initial approximation x_0 is arbitrary (according to the last equality in relation (1.2), b is replaced with r_0).

In 1994, Faber, Joubert, Knill and Manteuffel have presented a counterexample with a "dense" matrix of dimension $N = 4$ for which the inequality is strict at step $m = 3$, namely $\|p_{m,b}^{GM}(A)\, b\| = 0.99988 < 1 = \left\|p_m^{GM}(A)\right\|$ (cf. [223]). Toh presents in [223] a class of bidiagonal matrices depending on a parameter ε, for which

$$\frac{\max\limits_{\|b\|_2=1} \|p_{m,b}^{GM}(A)\, b\|_2}{\|p_m^{GM}(A)\|_2} \to 0 \quad \text{as } \varepsilon \to 0,$$

so that this conjecture is even more categorically invalidated.

Other results concerning the GMRES as a problem in the approximation theory may be found in the references [211], [121], [124], [223], [141] and [48]. We shall also mention the interpretation of the GMRES method as an interpolation problem.

For $x_0 = 0$, if the matrix A is diagonalizable, the following inequalities hold:

$$\begin{aligned}\left\|b - Ax_m^{GM}\right\|_2 &\leq \min_{p_m \in \mathbb{P}_m} \left\|V p_m(\Lambda)\, V^{-1} b\right\|_2 \\ &\leq \kappa_2(V) \min_{p_m \in \mathbb{P}_m} \max_{i=1,\ldots,N} |p_m(\lambda_i)|\, \|b\|_2,\end{aligned}$$

where V is the passing matrix and the matrix Λ is diagonal, containing the eigenvalues λ_i. If the matrix A is normal, then $\kappa_2(V) = 1$ and it is known that the above upper bound is exact (see [121, pp. 54–55] and the references therein). The problem may be interpreted as how well may the zero value is approximated on the nodes λ_i by a polynomial of degree m which has the value 1 at the origin. In this case it can be seen that there may appear unfavorable situations when there exists an eigenvalue of A close to 0, resp. favorable when all the eigenvalues are close to a value far away from zero. Though, when the matrix A is not normal, the algebraic and geometric orders of the eigenvalues remain important, but not the distribution of the eigenvalues on the axis (or on the plane, in the complex case) — see also Theorem 1.4 below.

From the practical viewpoint, the GMRES method is based on the following algorithm:

A. (Arnoldi)

A1. Let $r_0 = b - Ax_0$, $\beta = \|r_0\|_2$ and $v_1 = \frac{1}{\beta} r_0$;

A2. For $j = 1, \ldots, m$ do

$$h_{ij} = (Av_j, v_i), \quad \text{for } i = 1, \ldots, j$$
$$\hat{v}_{j+1} = A\hat{v}_j - \sum_{j=1}^{i} h_{ij} v_i$$
$$h_{j+1,j} = \|\hat{v}_{j+1}\|_2$$
$$v_{j+1} = \frac{1}{h_{j+1,j}} \hat{v}_{j+1}$$

A3. Form the Hessenberg matrix $\bar{H}_m \in \mathbb{R}^{(m+1)\times m}$ with the (possible) nonzero elements h_{ij} determined above, and the matrix $V_m \in \mathbb{R}^{N\times m}$ having as columns the vectors v_j: $V_m = [v_1 \ \ldots \ v_m]$.

GM. (GMRES)

GM1. Determine the exact solution y_m^{GM} to the least squares problem

$$\min_{y_m \in \mathbb{R}^m} \left\| \bar{H}_m y_m - \beta e_1 \right\|_2 \tag{1.3}$$

GM2. Set $x_m^{GM} = x_0 + V_m y_m^{GM}$.

At a certain step j_0 in the Arnoldi algorithm, there may appear the division by zero, when $h_{j_0+1,j_0} = 0$ (or, equivalently, when $\dim \mathcal{K}_{j_0+1} = j_0$). Such situations are called *breakdowns*, but in this case the solving of the problem GM1 with $m = j_0$ leads to the exact determination of x^* using only V_m and $\bar{H}_m$. The terminology is therefore *happy breakdown*.

Saad and Schultz have obtained the following result.

PROPOSITION 1.2 [211] *Consider the linear system* (1.1) *and an initial approximation* $x_0 \in \mathbb{R}^N$. *If the Arnoldi algorithm determines the elements* $h_{j+1,j} \neq 0$ *for* $j = 1, \ldots, m-1$, *then the approximation* x_m^{GM} *is exact* ($x_m^{GM} = x^*$) *if and only if one of the following equivalent relations hold:*

- $h_{m+1,m} = 0$;
- $\hat{v}_{m+1} = 0$;
- *the polynomial of minimal degree of A with respect to r_0 has degree m, i.e.,*

$$\min_{p_m \in \mathbb{P}_m} \|p_m(A)\, r_0\| = 0 \quad \textit{and} \quad \min_{p_i \in \mathbb{P}_i} \|p_i(A)\, r_0\| \neq 0 \quad \textit{for } i < m.$$

In the case when $h_{m+1,m} \neq 0$, the vector y_m^{GM} from step GM1 is uniquely determined.

The solution of the problem (1.2) may be explicitly written as (see [41]):

$$\begin{aligned} x_m^{GM} &= x_0 + V_m \left(\bar{H}_m^t \bar{H}_m\right)^{-1} \bar{H}_m^t V_{m+1}^t r_0 \\ &= x_0 + V_m \left(\bar{H}_m^t \bar{H}_m\right)^{-1} \bar{H}_m^t \beta e_1, \end{aligned} \tag{1.4}$$

but the computations are not performed by this formula. Problem (1.3) has a structure which makes it to be easily solved, and its size is small. For different implementation variants the following references may be consulted: [211], [233], [234], [236], [148], [121] and [36], to mention only a few.

In practice there is usually considered a maximum value $\bar{m}$ for the dimensions of the Krylov subspaces (in many situations, the values are of order $\bar{m} \in \{10, \ldots, 20\}$, or even smaller). The iterations from step A2 in the Arnoldi algorithm are performed for $j = 1, \ldots, \bar{m}$. In case of breakdown one determines the exact solution, and the algorithm is stopped. The algorithm may also be stopped if at a certain step $j < \bar{m}$ the residual of the solution x_j^{GM} is smaller than a given value. It is interesting to notice that this test may be performed without the explicit computation of x_j^{GM} (see [211]), which avoids some costly operations when, as we have assumed, N has large values.

If after $\bar{m}$ steps the determined solution does not have a sufficiently small residual,[6] the whole algorithm is restarted, taking as initial approx-

[6] As a stopping test one may also consider the value of the relative residual, as well as other quantities based on the residual (see [121], [23], [137], [237]).

imation x_0 the previously determined solution $x_{\bar{m}}^{GM}$. The terminology for this variant is *restarted GMRES*, and is the most used variant.

In the case of the restarted variant we shall denote by $x_m^{GM(0)}$, $m = 1, \ldots, \bar{m}$ the first $\bar{m}$ solutions, by $x_m^{GM(1)}$, $m = 1, \ldots, \bar{m}$ the following $\bar{m}$ solutions, and so on. The initial starting value x_0 will be clear from the context. For the nonrestarted variant we shall use the notations x_m^{GM}, while for a generic GMRES solution[7] we shall simply write x^{GM}; the notations which will correspond to the k-th correction from the Newton-GMRES method will be $s_{k,m}^{GM}$ resp. s_k^{GM}.

With these notations, the following result can be immediately obtained by the optimality properties of the GMRES solutions. As we have previously mentioned, the algorithm stops whenever at a certain step the exact solution may be determined. This aspect is implicitly assumed in the following results.

PROPOSITION 1.3 *Consider the linear system* (1.1) *and the initial approximation* $x_0 = 0$. *Then the following statements are true:*

- *For any* $m \in \{1, \ldots, N\}$, *the residual* r_m^{GM} *of the solution* x_m^{GM} *obeys*

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

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

- *The residuals associated to the GMRES solutions obey*

$$0 = \left\| r_N^{GM} \right\|_2 \leq \ldots \leq \left\| r_1^{GM} \right\|_2 \leq \|b\|_2 .$$

  Moreover, the inequalities between the norms of two consecutive residuals are strict if and only if the corresponding GMRES solutions are distinct.

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

- *For any fixed upper bound $\bar{m} \in \{1, \ldots, N-1\}$, the residuals from the restarted variant obey*

$$\ldots \leq \left\|r_1^{GM(l+1)}\right\|_2 \leq \left\|r_{\bar{m}}^{GM(l)}\right\|_2 \leq \ldots \leq \left\|r_1^{GM(l)}\right\|_2 \leq$$
$$\leq \left\|r_{\bar{m}}^{GM(l-1)}\right\|_2 \leq \ldots \leq \left\|r_1^{GM(0)}\right\|_2 \leq \|b\|_2 .$$

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

The above Proposition does not appear in the literature in this form though some of its statements are implicitly in different papers (see [41], [40], [122] and [24]). We enounce only the result obtained by Greenbaum, Pták and Strakoš.

THEOREM 1.4 [122] *For any nonincreasing sequence $f(1) \geq f(2) \geq \ldots \geq f(N-1) > 0$ of positive numbers and any set $\{\lambda_1, \ldots, \lambda_N\}$ of (real, distinct[8]) values there exists a matrix $A \in \mathbb{R}^{N\times N}$ having the eigenvalues $\lambda_1, \ldots, \lambda_N$, and a vector $b \in \mathbb{R}^N$ such that the residuals of the GMRES method with $x_0 = 0$ applied to the system $Ax = b$ satisfy*

$$\left\|r_m^{GM}\right\| = f(m), \quad \text{for } m = 1, \ldots, N-1.$$

Nachtigal, Reddy and Trefethen mentioned in the paper [167] a necessary and sufficient condition for strict monotonicity $\left\|r_{m+1}^{GM}\right\| < {}^{\|}r_m^{GM}$ for $m = 0, \ldots, N-1$ and for any initial approximation x_0 ("the field of values of A should lie in an open half-plane with respect to the origin")

A natural question raises: can one compute the exact solution x^* using the GMRES method considering a small dimension $\bar{m}$, and performing sufficiently restarts? We shall see that the answer is negative.

[8] In the cited paper, the above theorem is enounced for the complex case. The authors do not explicitly mention, but is required that the eigenvalues to be distinct in order to have $r_{N-1}^{GM}{}_2 > 0$.

Even if (in exact arithmetic) the GMRES method always offers the unique solution of the problem (1.2) in at most N steps, in some situations there may appear a stagnation in improving the iterations.[9]

EXAMPLE 1.1. [41], [40] Consider the permutation matrix

$$A = \begin{pmatrix} 0 & & & 1 \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{pmatrix} \in \mathbb{R}^{N\times N}$$

and the vector $b = e_1 \in \mathbb{R}^N$, the system $Ax = b$ having the unique solution $x^* = e_N$. Taking $x_0 = 0$ and performing m steps of the Arnoldi algorithm, for $m < N$ one obtains

$$\bar{H}_m = \begin{pmatrix} 0 & & \\ 1 & \ddots & \\ & \ddots & 0 \\ & & 1 \end{pmatrix} \in \mathbb{R}^{(m+1)\times m} \quad \text{and} \quad V_m = [e_1 \ \dots \ e_m]\,.$$

Using formula (1.4) it easily follows that $x_m^{GM} = 0$; the initial approximation $x_0 = 0$ cannot be improved by increasing m until the final step, $m = N$, when the exact solution is obtained. The situation is the same in the restarted version if $\bar{m}, m \leq N - 1$. ∎

This example of theoretical nature is meant to show that there exist situations when satisfactory approximations may be obtained only for large values of the Krylov subspaces — in which case the method is no longer efficient.

In the above description of GMRES we have supposed that the arithmetic operations are performed exactly. Since the numbers are represented in computers in floating point arithmetic, there inherently appear errors in

[9] This phenomenon is tightly connected to the breakdowns of the Arnoldi-FOM method [210], as shown by the results of Brown [40] and resp. Cullum and Greenbaum [81].

their representation as well as in the elementary arithmetic operations. For certain results regarding the behavior of GMRES in this context one may consult [121], [105], [123] and [22].

* * *

The test based on the magnitude of the residual of an approximate solution is not always indicated. It is known that, in exact arithmetic, "for an approximate solution $\tilde{x}$ of the linear system (1.1), if the error $\tilde{x} - x^*$ is small, then so is the residual $b - A\tilde{x}$"[10] [146] (see also [137], [237]). The converse of this statement is not true: "when the matrix A is ill conditioned, if the residual $b - A\tilde{x}$ is small, it does not necessarily follow that the error $\tilde{x} - x^*$ is small" [146] (see also [237]).

Methods for minimizing other quantities than residuals have been recently proposed by Kasenally, resp. Kasenally and Simoncini in [146] and [147], where they consider the *normwise backward error* corresponding to an approximation $\tilde{x}$ to x^*:

$$\Pi(\tilde{x}) = \min\left\{\varepsilon : (A + \Delta)\,\tilde{x} = b + \delta,\ \|\Delta\|_F \leq \varepsilon \|E\|_F,\ \|\delta\|_2 \leq \varepsilon \|f\|_2\right\},$$

where the parameters $E \in \mathbb{R}^{N\times N}$ and $f \in \mathbb{R}^N$ are arbitrary, but fixed. The value of $\Pi(\tilde{x})$ is known to be given by

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

and the minimum value is attained by the *backward errors*

$$\Delta_A = \frac{\|E\|_F \cdot \|\tilde{x}\|_2}{\|E\|_F \cdot \|\tilde{x}\|_2 + \|f\|_2}\,(b - A\tilde{x})\,\frac{\tilde{x}^t}{\|\tilde{x}\|_2^2}, \quad \text{and}$$

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

[10] It is interesting to mention that if the residual $b - A\tilde{x}$ is computed in floating point arithmetic, then this statement does not hold in general [23].

Kasenally [146] mentions that these results have been obtained by Rigal and Gaches in 1967, in [208].

The problem solved by the mentioned algorithms is that of finding an element $x_m \in x_0 + \mathcal{K}_m$ which minimizes the backward errors.

REMARK 1.2. Kasenally shows in [146] that the property of (1.2) the GMRES solution may be expressed in terms of the backward errors:

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

i.e., x_m^{GM} minimizes the backward error Δ_b, assuming $\Delta_A = 0$. ∎

1.2.2 The GMBACK method

The GMBACK algorithm determines an element $x_m^{GB} \in x_0 + \mathcal{K}_m$ which minimizes the backward error in the matrix A, assuming zero the error in b:

$$\min_{x_m \in x_0+\mathcal{K}_m} \|\Delta_A\|_F, \quad \text{such that } (A - \Delta_A)\, x_m = b. \tag{1.5}$$

The solution x_m^{GB} is obtained in the following way.

A. (Arnoldi)

Determine the elements $\bar{H}_m$ and V_{m+1};

GB. (GMBACK)

GB1. Let $\beta = \|r_0\|_2$,
$\hat{H}_m = \begin{bmatrix} -\beta e_1 & \bar{H}_m \end{bmatrix} \in \mathbb{R}^{(m+1)\times(m+1)}$,
$\hat{G}_m = [x_0 \quad V_m] \in \mathbb{R}^{N\times(m+1)}$,
$P = \hat{H}_m^t \hat{H}_m \in \mathbb{R}^{(m+1)\times(m+1)}$,
$Q = \hat{G}_m^t \hat{G}_m \in \mathbb{R}^{(m+1)\times(m+1)}$;

GB2. Determine an eigenvector u_{m+1} corresponding to the smallest eigenvalue λ_{m+1}^{GB} of the problem $Pu = \lambda Qu$;

GB3. Compute y_m^{GB} such that[11]

$$\begin{bmatrix} 1 \\ y_m^{GB} \end{bmatrix} = \frac{1}{u_{m+1}^{(1)}} u_{m+1};$$

GB4. Set $x_m^{GB} = x_0 + V_m y_m^{GB}$.

REMARK. The matrices P and Q from step GB1 are symmetric, P is positively defined Q is semipositively defined, such that the eigenvalues of the generalized eigenproblem $Pu = \lambda Qu$ are positive: $+\infty \geq \lambda_1^{GB} \geq \lambda_2^{GB} \geq \ldots \geq \lambda_{m+1}^{GB} > 0$. The case $\lambda_1^{GB} = +\infty$ appears when Q is singular (see, e.g., [221]). ∎

Kasenally [146] shows that, for any initial approximation $x_0 \in \mathbb{R}^N$ and for any $m \in \{1, \ldots, N\}$, the backward error $\Delta_{A,m}^{GB}$ corresponding to the GMBACK solution x_m^{GB} is given by

$$\Delta_{A,m}^{GB} = V_{m+1} \left(\bar{H}_m y_m^{GB} - \beta e_1 \right) \frac{\left(x_m^{GB} \right)^t}{\| x_m^{GB} \|_2^2}, \tag{1.6}$$

and its norm is

$$\left\| \Delta_{A,m}^{GB} \right\|_F = \sqrt{\lambda_{m+1}^{GB}}. \tag{1.7}$$

As in the GMRES method, the happy breakdowns from the Arnoldi algorithm lead to the determination of the exact solution[12]. Unlike GMRES, the solution x_m^{GB} of the problem (1.5) may not be uniquely determined if the eigenvalue λ_{m+1}^{GB} is not simple. Moreover, there may appear some undesired situations when the solution x_m^{GB} cannot be determined ("*uncircumventible breakdowns*"). Such situations appear when all the eigenvectors corresponding to λ_{m+1}^{GB} have the first component zero, which leads to divisions by zero.

[11] See the notations from page xiii.

[12] This property is in fact true for all the Krylov methods based on Hessenberg orthogonalization.

EXAMPLE 1.3. [146] Consider again the linear system $Ax = b$ from the previous example. Taking again $x_0 = 0$ and $m \leq N - 1$, one obtains the same matrices $\bar{H}_m$, V_m, and then

$$P = I_{m+1} \text{ and } Q = \begin{pmatrix} 0 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix} = [0\ e_2\ \ldots\ e_{m+1}] \in \mathbb{R}^{(m+1)\times(m+1)}.$$

The eigenvalues of the problem $Pu = \lambda Qu$ are $\lambda_1 = +\infty$ and $\lambda_2 = \ldots = \lambda_{m+1} = 1$, with the eigenvectors $u_1 = e_1 \in \mathbb{R}^{m+1}$ and $u_2 = e_2, \ldots, u_{m+1} = e_{m+1} \in \mathbb{R}^{m+1}$, such that the vector y_m^{GB} cannot be determined. For $m = N$ one obtains the exact solution. ∎

This example is, as the previous one, of theoretical nature. In practical situations, when one cannot choose an eigenvector u_{m+1} with $u_{m+1}^{(1)} \neq 0$, either one considers a further step in the Arnoldi algorithm, or the algorithm is restarted with another initial approximation x_0. In the following results presented in this work we shall assume that in the GMBACK method one does not encounter divisions by zero. The same assumption shall be made on the MINPERT solutions.

Unlike GMRES, the residual of an approximate solution x_m^{GB} cannot be computed without having explicitly the approximation; one may however determine the norm of the backward error in A, by computing the eigenvalue λ_{m+1}^{GB}. Also, the computation of the eigenpair $\left(\lambda_{m+1}^{GB}, u_{m+1}\right)$ constitutes a different problem compared to the linear least squares problem with Hessenberg matrix from GMRES.

Concerning the elements computed by GMBACK, we obtain the following result:

PROPOSITION 1.5 [58] *Consider the initial approximation $x_0 \in \mathbb{R}^N$ and $m \in \{1, \ldots, N\}$. If these elements determine a GMBACK solution x_m^{GB} of the linear system* (1.1)*, then*

$$\left\|\Delta_{A,m}^{GB} \cdot x_m^{GB}\right\|_2 = \left\|r_m^{GB}\right\|_2,$$

where $r_m^{GB} = b - Ax_m^{GB}$.

Proof. The matrices V_{m+1} and $\bar{H}_m$ determined in the Arnoldi algorithm obey the following known relation (see [211]):

$$AV_m = V_{m+1}\bar{H}_m,$$

which shows that

$$\begin{aligned}\left\|V_{m+1}\left(\bar{H}_m y_m^{GB} - \beta e_1\right)\right\|_2 &= \left\|AV_m y_m^{GB} - r_0\right\|_2 \\ &= \left\|AV_m y_m^{GB} + Ax_0 - b\right\|_2 \\ &= \left\|Ax_m^{GB} - b\right\|_2 \\ &= \left\|r_m^{GB}\right\|_2.\end{aligned}$$

Taking into account formula (1.6) we are immediately lead to the stated affirmation. ∎

1.2.3 The MINPERT method

The MINPERT method determines an element $x_m^{MP} \in x_0 + \mathcal{K}_m$ which minimizes the joint backward error $[\Delta_A \quad \Delta_b]$:

$$\min_{x_m \in x_0+\mathcal{K}_m} \|[\Delta_A \quad \Delta_b]\|_F, \quad \text{such that } (A - \Delta_A)\, x_m = b + \Delta_b, \tag{1.8}$$

where the matrix $[\Delta_A \quad \Delta_b] \in \mathbb{R}^{N\times(N+1)}$ contains in its first N columns the matrix Δ_A, and in the $N+1$-th column the vector Δ_b. In other words, the solution x_m^{MP} minimizes the distance from the original system to a perturbed system, that an approximation x_m satisfies exactly. Another interpretation, given by Kasenally and Simoncini [147], shows that the above minimization is tightly connected to the total least squares problem — see [119] and [120].

The algorithm is similar to GMBACK, the only difference appearing in forming the matrix Q:

A. (Arnoldi)

Determine $\bar{H}_m$ and V_{m+1};

MP. (MINPERT)

MP1. Let $\beta = \|r_0\|_2$,
$\hat{H}_m = \left[-\beta e_1 \ \bar{H}_m\right]$,
$G_m = \begin{bmatrix} x_0 & V_m \\ 1 & 0 \end{bmatrix} \in \mathbb{R}^{(N+1)\times(m+1)}$,
$P = \hat{H}_m^t \hat{H}_m \in \mathbb{R}^{(m+1)\times(m+1)}$
$Q = G_m^t G_m \in \mathbb{R}^{(m+1)\times(m+1)}$;

MP2. Determine an eigenvector u_{m+1} corresponding to the smallest eigenvalue λ_{m+1}^{MP} of the problem $Pu = \lambda Qu$;

MP3. Determine y_m^{MP} such that

$$\begin{bmatrix} 1 \\ y_m^{MP} \end{bmatrix} = \frac{1}{u_{m+1}^{(1)}} u_{m+1};$$

MP4. Set $x_m^{MP} = x_0 + V_m y_m^{MP}$.

The same remarks regarding the problem $Pu = \lambda Qu$ hold, as in the GMBACK case.

Kasenally and Simoncini [147] have shown that the elements determined by MINPERT verify for all $x_0 \in \mathbb{R}^N$ and $m \in \{1, \ldots, N\}$ the following relations:

$$x_m^{MP} = x_0 + V_m \left(\bar{H}_m^t \bar{H}_m - \lambda_{m+1}^{MP} I_{m+1}\right)^{-1} \left(\bar{H}_m^t \beta e_1 + \lambda_{m+1}^{MP} V_m^t x_0\right), \tag{1.9}$$

$$\left\| \left[\Delta_{A,m}^{MP} \quad \Delta_{b,m}^{MP}\right] \right\|_F = \sqrt{\lambda_{m+1}^{MP}}, \tag{1.10}$$

$$\Delta_{b,m}^{MP} = \frac{-1}{\left\| \left[(x_m^{MP})^t \quad 1\right]^t \right\|_2^2} r_m^{MP}, \tag{1.11}$$

(1.12)

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

where $\Delta_{A,m}^{MP}$, $\Delta_{b,m}^{MP}$ and resp. r_m^{MP} represent the backward errors, resp. the residual of the approximate solution x_m^{MP}.

We obtain the following results regarding the size of the joint backward error of MINPERT when $x_0 = 0$, which are similar to those from GMRES.

PROPOSITION 1.6 [66] *Consider the initial approximation $x_0 = 0$ to the solution of the linear system (1.1). Then for any $m \in \{1, \ldots, N\}$, the norm of the joint backward error associated to the solution x_m^{MP} given by MINPERT satisfies the inequality:*

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

Proof. As it is also shown in [147], for $x_0 = 0$ the MINPERT algorithm generates the matrix $Q = I_{m+1}$, such that the eigenvalue problem from step MP2 is not generalized. Applying the Rayleigh quotient formula [221] one obtains

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

which, together with (1.10), proves the assertion. ∎

REMARKS. a) In proving the above result one may consider directly the quotient $\frac{z^t P z}{z^t Q z}$ for the generalized eigenvalue problem $Pu = \lambda Q u$, applying then the Fisher theorem [221, Cor. VI.1.16].

b) The vector e_1 cannot be similarly used for bounding the backward error $\Delta_{A,m}^{GB}$ from GMBACK with $x_0 = 0$, since in this case

$$e_1^t Q e_1 = e_1^t \left[0 \; e_2 \; \ldots \; e_{m+1} \right] e_1 = 0$$

(the vector e_1 is an eigenvector corresponding to $\lambda_1^{GB} = +\infty$). Such a result could not be expected to hold since, anticipating, this would imply that there would be enough to use one-dimensional Krylov subspaces in order to obtain local convergence with q-order 2 for the Newton-GMBACK method, for any nonlinear system approachable with the Newton method (see Corollary 2.37). ∎

We obtain the following consequences.

COROLLARY 1.7 [66] *Consider the linear system* (1.1) *and the initial approximation* $x_0 = 0$. *Then the joint backward errors associated to the MINPERT solutions obey*

$$\begin{aligned} 0 &= \left\| \left[\Delta_{A,N}^{MP} \quad \Delta_{b,N}^{MP} \right] \right\|_F \\ &\leq \left\| \left[\Delta_{A,N-1}^{MP} \quad \Delta_{b,N-1}^{MP} \right] \right\|_F \leq \ldots \leq \left\| \left[\Delta_{A,1}^{MP} \quad \Delta_{b,1}^{MP} \right] \right\|_F \leq \|b\|_2 . \end{aligned}$$

Proof. The last inequality was shown in the previous result, and the others are proved in [147]. ∎

REMARK. The optimization problem (1.8) does not always have a unique solution, such that the above inequalities may not be strict even for two different consecutive solutions. ∎

COROLLARY 1.8 [66] *Consider the linear system* (1.1) *and* $\bar{m} \in \{1, \ldots, N-1\}$ *fixed. Then in using the restarted version of the MINPERT method with the initial approximation* $x_0 = 0$, *the obtained elements verify*

$$\begin{aligned} \ldots &\leq \left\| \left[\Delta_{A,1}^{MP(l+1)} \ \Delta_{b,1}^{MP(l+1)} \right] \right\|_F \\ &\leq \left\| \left[\Delta_{A,\bar{m}}^{MP(l)} \ \Delta_{b,\bar{m}}^{MP(l)} \right] \right\|_F \leq \ldots \leq \left\| \left[\Delta_{A,1}^{MP(l)} \ \Delta_{b,1}^{MP(l)} \right] \right\|_F \\ &\leq \left\| \left[\Delta_{A,\bar{m}}^{MP(l-1)} \ \Delta_{b,\bar{m}}^{MP(l-1)} \right] \right\|_F \leq \ldots \leq \left\| \left[\Delta_{A,1}^{MP(0)} \ \Delta_{b,1}^{MP(0)} \right] \right\|_F \leq \|b\|_2 . \end{aligned}$$

The proof is immediately obtained by taking into account the minimum properties of the MINPERT solutions. ∎

As in the GMRES case, there may appear situations when the MINPERT iterations stagnate. On the other hand, similarly to the GMBACK method, the MINPERT method may also lead to uncircumventible breakdowns, when all the eigenvectors corresponding to λ_{m+1}^{MP} have the first component equal to zero. We offer some concrete examples in [66].

EXAMPLE 1.4. [66] Consider the matrix $A \in \mathbb{R}^{N\times N}$ from the preceding examples and the linear system $Ax = b$, where $b = he_1 \in \mathbb{R}^N$, with $h \in \mathbb{R}$, $|h| > 1$.

Taking $x_0 = 0$, for $m \leq N-1$ we obtain the same matrices $\bar{H}_m$ and V_m, and next

$$P = \begin{pmatrix} h^2 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}, \quad Q = I_{m+1}.$$

The eigenvalues are $\lambda_1^{MP} = h^2 > 1$ and $\lambda_2^{MP} = \ldots = \lambda_{m+1}^{MP} = 1$, with the eigenvectors $u_1 = e_1$, $u_2 = e_2, \ldots, u_{m+1} = e_{m+1} \in \mathbb{R}^{m+1}$. For $m \in \{1, \ldots, N-1\}$, the vector x_m^{MP} cannot be determined because of the first component of the eigenvectors corresponding to the eigenvalue 1. For $m = N$ one obtains the exact solution. ∎

EXAMPLE 1.5. [66] Consider the same matrix as above, and take $b = he_1 \in \mathbb{R}^N$, $0 < |h| < 1$. In this case, for any value $m \in \{1, \ldots, N-1\}$ the MINPERT method with $x_0 = 0$ leads to the unique solution $x_m^{MP} = 0$. For $m = N$ one obtains the exact solution. The use of the restarted variant for $\bar{m}, m \leq N-1$ leads to the zero solution too. ∎

The following inequalities are immediately obtained by taking into account the optimum properties of the GMRES, GMBACK and MINPERT solutions.

PROPOSITION 1.9 *Consider the linear system* (1.1) *and the arbitrary elements* $x_0 \in \mathbb{R}^N$ *and* $m \in \{1, \ldots, N\}$. *Then*

$$\max\left\{\left\|\Delta_{A,m}^{MP}\right\|_F, \left\|\Delta_{b,m}^{MP}\right\|_2\right\} \leq \left\|\left[\Delta_{A,m}^{MP} \quad \Delta_{b,m}^{MP}\right]\right\|_F \leq \left\|\Delta_{b,m}^{GM}\right\|_2, \tag{1.13}$$

$$\max\left\{\left\|\Delta_{A,m}^{MP}\right\|_F, \left\|\Delta_{b,m}^{MP}\right\|_2\right\} \leq \left\|\left[\Delta_{A,m}^{MP} \quad \Delta_{b,m}^{MP}\right]\right\|_F \leq \left\|\Delta_{A,m}^{GB}\right\|_F, \tag{1.14}$$

where the solutions x_m^{GM}, x_m^{GB} *and* x_m^{MP} *are determined by the same elements* m *and* x_0.

The inequalities from the right in relations (1.13) and (1.14) have been obtained by Kasenally and Simoncini in [147]. In [66] we notice that the inequalities from the left are obvious. In our opinion, the following inequalities stated in [147, Th.4.4]

$$\left\| \left[\Delta_{A,m}^{MP} \quad \Delta_{b,m}^{MP} \right] \right\|_F \leq \left\| \Delta_{A,m}^{GB} \right\|_2 \leq \left\| \Delta_{A,m}^{MP} \right\|_F .$$

are not true in general.

The sequence in relation (1.13) may be completed by another inequality:

PROPOSITION 1.10 [147], [66] *Under the hypothesis of the above proposition, the following inequalities hold:*

$$\max \left\{ \left\| \Delta_{A,m}^{MP} \right\|_F , \left\| \Delta_{b,m}^{MP} \right\|_2 \right\} \leq \left\| \left[\Delta_{A,m}^{MP} \quad \Delta_{b,m}^{MP} \right] \right\|_F \leq \left\| \Delta_{b,m}^{GM} \right\|_2 \leq \left\| r_m^{MP} \right\|_2 .$$

Proof. [66] When the backward error Δ_A is assumed to be zero for a certain approximate solution $\tilde{x}$ of the system $Ax = b$, the backward error Δ_b coincides with the residual $\tilde{x}$, both the quantities being (apart of their sign) uniquely determined by the expression $b - A\tilde{x}$. This consideration, together with the optimum property of the GMRES solution, justifies the stated affirmation. ∎

The inequality from the above proposition has been obtained first in [147]. Its proof has been reconsidered by us in [66].

The connection between the three Krylov solutions is given by the following results. The first two relate the MINPERT and GMRES solutions.

THEOREM 1.11 [147] *Consider the linear system* (1.1) *and the arbitrary elements* $x_0 \in \mathbb{R}^N$ *and* $m \in \{1, \ldots, N\}$. *Let* $\sigma_1^2 \geq \sigma_2^2 \geq \ldots \geq \sigma_m^2$ *denote the eigenvalues of* $\bar{H}_m^t \bar{H}_m$ *(i.e., the squared singular values of* $\bar{H}_m$*) and assume that* $\sigma_m^2 \neq \lambda_{m+1}^{MP}$. *Then*

$$\left\| x_m^{MP} - x_m^{GM} \right\|_2 \leq \frac{\lambda_{m+1}^{MP}}{\sigma_m^2 - \lambda_{m+1}^{MP}} \left\| x_m^{GM} \right\|_2 ,$$

$$\left\| r_m^{MP} - r_m^{GM} \right\|_2 \leq \frac{\lambda_{m+1}^{MP}}{\sigma_m^2 - \lambda_{m+1}^{MP}} \sigma_1 \left\| V_m^t x_m^{GM} \right\|_2 .$$

REMARK. [147] An auxiliary result (also proved by Kasenally and Simoncini) shows that the eigenvalues determined at the m-th step interlace in the following way: $\lambda_i^{MP} \geq \sigma_i^2 \geq \lambda_{i+1}^{MP}$, $i = 1, \ldots, m$. So, the bounds from the above inequalities are large when σ_m^2 is close to λ_{m+1}^{MP}, and the result does not hold when the two values are identic (such a situation arises for example when λ_{m+1}^{MP} is a multiple eigenvalue).

COROLLARY 1.12 [147] *Under the assumptions of the above proposition, if* $x_0 = 0$ *then*

$$\left\|x_m^{MP} - x_m^{GM}\right\|_2 \leq \frac{\lambda_{m+1}^{MP}\beta}{\sigma_m\left(\sigma_m^2 - \lambda_{m+1}^{MP}\right)}.$$

The third result refers to the connection between the MINPERT and GMBACK solutions.

THEOREM 1.13 [147] *Consider the linear system* (1.1) *and the arbitrary elements* $x_0 \in \mathbb{R}^N$ *and* $m \in \{1, \ldots, N\}$. *Then the following inequality is true:*

$$\left\|x_m^{MP} - x_m^{GB}\right\|_2 \leq \frac{\left|\lambda_{m+1}^{GB} - \lambda_{m+1}^{MP}\right|}{\sigma_m^2 - \lambda_{m+1}^{GB}} \left\|V_m^t x_m^{MP}\right\|_2.$$