SOLVING ILL-POSED (NEARLY SINGULAR) LINEAR SYSTEMS WITH GMRES AND APPLICATION TO PRECONDITIONING*

LARS ELDÉN†, VALERIA SIMONCINI‡, AND ZOHREH RANJBAR§

Abstract. Almost singular linear systems arise in discrete ill-posed problems. Either because of the intrinsic structure of the problem or because of preconditioning, the spectrum of the coefficient matrix is often characterized by a sizable gap between a large group of numerically zero eigenvalues and the rest of the spectrum. Correspondingly, the right-hand side has leading eigencomponents associated with the eigenvalues away from zero. In this paper the effect of this setting in the convergence of the Generalized Minimal RESidual (GMRES) method is considered. It is shown that in the initial convergence phase of the iterative algorithm, the residual components corresponding to the large eigenvalues are reduced in norm, and these can be monitored without extra computation. A stopping criterion based on the discrepancy principle is used. Our analysis is supported by numerical experiments on benchmark ill-posed problems and on an ill-posed Cauchy problem for a parabolic equation, where a singular (low-rank) preconditioner is employed.

1. Introduction. Large, sparse nonsymmetric and *singular* linear systems arise when certain partial differential equations (PDE) are discretized. In [6] conditions are given for the convergence without breakdown of the Generalized Minimum Residual algorithm (GMRES) [35] applied to such problems. Since the appearance of [6] many papers have been devoted to the analysis and application of GMRES for exactly singular problems, see [24] for a rather extensive account of the relevant literature.

In this paper we are concerned with *almost singular* (or *numerically singular*) linear systems,

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

where $A \in \mathbb{C}^{n\times n}$. Such systems occur in connection with ill-posed problems, and for some problems GMRES works well, while for others it performs badly, see, e.g., [27, Examples 5.3 and 5.1, respectively]. Recently it has been demonstrated that GMRES gives a good approximate solution in few iterations for certain ill-posed problems for PDE's, when a singular preconditioner is used [32, Part III]. However, so far a deeper analysis of the properties of GMRES applied to almost singular systems is lacking.

The purpose of the present paper is to analyze and explain the convergence behavior of GMRES for linear systems that are almost singular, the way they occur in ill-posed problems. Previous attempts have often focused on information associated with the singular value decomposition of the matrix, see, e.g., [27, 17, 5]. Instead, in agreement with, e.g., [7, 8], we will rely on spectral information of the problem, with the Schur decomposition of the coefficient matrix as the core theoretical tool. Indeed, in some cases the matrix has a cluster of eigenvalues of magnitude $O(1)$ that is well separated from another cluster of eigenvalues of small magnitude. Correspondingly, the right-hand side has large and leading components onto the eigendirections associated with the cluster away from the origin. Assuming that the linear system (1.1) is a (possibly large) perturbation of an exactly singular system of rank m, we will show that:

- in the first iterations GMRES mainly reduces the norm of the residual as if solving the unperturbed system;

*Version of May 2, 2011.

†Department of Mathematics, Linköping University, Sweden, lars.elden@liu.se

‡Dipartimento di Matematica, Università di Bologna, Italy, valeria@dm.unibo.it

§Department of Mathematics, Linköping University, Sweden, zoran@mai.liu.se

- after at most m iterations, but often much earlier, the norm of the residual is of the order of magnitude of the perturbation, and if the GMRES procedure is then stopped, it gives a good approximation of the minimum norm solution of the exactly singular system.

Our theoretical findings generalize and are in agreement with the results discussed in [6], [24] for exactly singular systems. In particular, our analysis specifically explores the case when the condition for obtaining a minimum norm solution is not met, which is usually the setting encountered in ill-posed problems.

We will also consider the case when the eigenvalues are not clustered (when the numerical rank is ill-determined, which is often the case in ill-posed problems, see, e.g., the discussion in [2, 8]), and show theoretically and by examples that GMRES will give a good approximate solution if the iterations are stopped when the residual is of the order of the perturbation.

Numerically singular systems with clustered eigenvalues occur when singular preconditioners are applied to discrete ill-posed linear systems $Ax = b$ [32, Part III]. For such a problem, arising from the discretization of a linear equation with a compact operator, the ill-posedness manifests itself in the blow-up of high frequency components in the numerical solution. In order for the problem to be approximately solvable the solution must be well represented in terms of the low frequency part of the operator. If the preconditioner M gives a good approximation of the low frequency part of the operator, but omits the high frequency part, then the preconditioned problem $AM^{\dagger}y = b$ has the properties above, i.e., $AM^{\dagger}$ is numerically singular, but with a well-conditioned low rank part, the minimum norm solution of which will give a good approximation to the solution of the ill-posed problem.

It is well-known, see, e.g., [15, 29], that iterative methods applied to ill-posed problem exhibit *semi-convergence*: initially the approximate solution converges towards the "true solution", then it deteriorates and finally blows up. Such a convergence behavior occurs also here, and we give a theoretical explanation. Note, however, that in the case of singular preconditioners semi-convergence usually does not apply to the final solution approximation, but only to an intermediate quantity. A stopping criterion based on the *discrepancy principle* will give a solution that is close to optimal.

We will use the following notation. The conjugate transpose of a matrix A is A^*. The Euclidean vector norm is denoted $\|x\| = (x^*x)^{1/2}$, and the corresponding matrix norm is $\|A\| = \max_{\|x\|=1} \|Ax\|$. The Frobenius norm is $\|A\|_F = (\sum_{i,j} |a_{ij}|^2)^{1/2}$. The singular values of a matrix $B \in \mathbb{C}^{m\times n}$, where $m \leq n$, are denoted σ_i, $i = 1, 2, \ldots, m$, and are ordered as $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_m \geq 0$; if $\sigma_m \neq 0$, its condition number is $\kappa_2(B) = \sigma_1/\sigma_m$. $A^{\dagger}$ denotes the Moore-Penrose pseudoinverse of A.

2. Theory for the Exactly Singular Case. Given a starting guess x_0 and the associated residual $r_0 = b - Ax_0$, GMRES determines an approximate solution x_k to (1.1) as $x_k \in x_0 + K_k(A, r_0)$ where $K_k(A, r_0)$ =span$\{r_0, Ar_0, \ldots, A^{k-1}r_0\}$ is the Krylov subspace, by requiring that the corresponding residual $r_k = b - Ax_k$ has minimum norm. The problem of solving a singular linear system $Ax = b$ using GMRES is treated in [6, 24], where the following result is proved.

PROPOSITION 2.1. *GMRES determines a least squares solution x_* of a singular system $Ax = b$, for all b and starting approximations x_0, without breakdown, if and only if $\mathcal{N}(\mathcal{A}) = \mathcal{N}(\mathcal{A}^*)$. Furthermore, if the system is consistent and $x_0 \in \mathcal{R}(\mathcal{A})$, then x_* is a minimum norm solution.*

Assume that the rank of A is equal to m. For the analysis it is no restriction to

assume that the matrix of the linear system has the structure[1]

$$\begin{bmatrix} A_{11} & A_{12} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y^{(1)} \\ y^{(2)} \end{bmatrix} = \begin{bmatrix} c^{(1)} \\ c^{(2)} \end{bmatrix}, \qquad A_{11} \in \mathbb{C}^{m\times m}. \tag{2.1}$$

It is easy to see (cf. [24]) that the condition $\mathcal{N}(\mathcal{A}) = \mathcal{N}(\mathcal{A}^*)$ is equivalent to $A_{12} = 0$. Similarly, the consistency condition is equivalent to $c^{(2)} = 0$.

Obviously, applying GMRES to the linear system

$$\begin{bmatrix} A_{11} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y^{(1)} \\ y^{(2)} \end{bmatrix} = \begin{bmatrix} c^{(1)} \\ 0 \end{bmatrix}, \tag{2.2}$$

is mathematically equivalent to applying GMRES to $A_{11}y^{(1)} = c^{(1)}$. Due to the finite termination property of Krylov methods it will never take more than m steps to obtain the solution of this problem (in exact arithmetic).

Finally in this section, the properties of the Krylov subspace ensure that applying GMRES to

$$\begin{bmatrix} A_{11} & A_{12} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y^{(1)} \\ y^{(2)} \end{bmatrix} = \begin{bmatrix} c^{(1)} \\ 0 \end{bmatrix},$$

with zero starting approximation, is also mathematically equivalent to applying GMRES to $A_{11}y^{(1)} = c^{(1)}$. A more common situation occurs when the (2,2) block of (2.1) is *almost* zero, that is it has small but nonzero entries. In this case, the role of the A_{12} block becomes more relevant. We analyze such a setting in the next section for a general A, by first performing a Schur decomposition.

3. The Almost Singular Case. Let $A = UBU^*$ be the Schur decomposition of A [13, p. 313], where B is upper triangular with diagonal elements ordered by decreasing magnitude. By a change of variables we get the linear system $By = c$, which is equivalent to the original one, and which we partition[2] as

$$\begin{bmatrix} L_1 & G \\ 0 & L_2 \end{bmatrix} \begin{bmatrix} y^{(1)} \\ y^{(2)} \end{bmatrix} = \begin{bmatrix} c^{(1)} \\ c^{(2)} \end{bmatrix}, \tag{3.1}$$

where $L_1 \in \mathbb{C}^{m\times m}$. Here we assume[3]

$$|\lambda_{\min}(L_1)| \gg |\lambda_{\max}(L_2)|, \qquad \|c^{(1)}\| \gg \|c^{(2)}\| = \delta. \tag{3.2}$$

By $\lambda_{\min}(L_1)$ we mean the eigenvalue of smallest modulus. We also assume that L_1 is well conditioned, i.e. $\|L_1^{-1}\|$ is not large. The eigenvalue conditioning is related to an assumption that B is almost singular. Thus L_2 can be considered as a perturbation of zero, either corresponding to floating point "noise" or due to some other type of noise, and the same applies to the lower part $c^{(2)}$ of the right hand side. We shall also assume that $\|G\|$ has a small or moderate value, so that we exclude the occurrence of non-normality influencing the two diagonal blocks. The assumptions in (3.2) also exclude

[1] In [24] a transformation of the system is done by decomposing the space $\mathbb{C}^n$ into $\mathcal{R}(\mathcal{A})$ and $\mathcal{R}(\mathcal{A})^\perp$.

[2] The same notational convention will be used throughout.

[3] The meaning of the "much larger than" sign will depend on the context: in the case of singular preconditioners it can be several orders of magnitude, while in the case when GMRES is applied directly to an ill-posed problem, it may be only two orders of magnitude, see the numerical examples.

the case, for instance, where the given problem is a perturbation of a nonsymmetric matrix with all zero eigenvalues and a single eigenvector; cf., e.g., [27, Example of Section 5.1]. Note that the perturbed eigenvalues will tend to distribute in a small disk around the origin. Our assumption is not restrictive, since it is already known that GMRES will perform very badly in this setting; see, e.g., [28, Example R, p. 787].

If the linear system represents an ill-posed problem[4], then the condition $\|c^{(1)}\| \gg \|c^{(2)}\|$ is reminiscent of the discrete Picard criterion in the singular value setting [19]. In this context it is not meaningful to solve the linear system exactly but only to compute an approximate solution y, whose residual $r = c - By$ satisfies $\|r\| \approx \|c^{(2)}\|$. This is closely related to the use of the *discrepancy principle* in the regularization of ill-posed problems, see, e.g., [12, p. 83],[21, p. 179].

Now, since the linear system (3.1) can be seen as a perturbation of (2.2), we may ask whether it is possible to "solve" (3.1) as efficiently as (2.2). We will show in Section 3.1 that a minimal residual method onto a subspace of dimension m, where m is the size of L_1, may provide a sufficiently small residual, whose size depends on $\|L_2\|$, $\|G\|$, and $\|c^{(2)}\|$, as expected. Other quantities also enter the picture. In Section 3.2 we will describe how the use of a minimal residual method with a Krylov subspace, allows one to derive a better residual, and more accurate bounds for its norm. All our results show that in addition to the size of the matrices involved, also the spectral "distance" between L_1 and L_2 has a role in bounding the residual.

We also remark that the model derived by splitting the spectral domain in a "good" part and in a "bad" part has been used both in eigenvalue computation, see, e.g., [3], as well as in the numerical solution of ill-posed linear systems, see, e.g., [2]. In both cited cases, however, the aim is to computationally exploit an approximate decomposition so as to accelerate the convergence of the employed method. In this paper the *exact* splitting is a theoretical device, that allows us to explain the practical behavior of GMRES in certain circumstances.

To proceed we need to introduce some notation and definitions. Under the eigenvalue assumption in (3.2) we can write

$$B = \begin{bmatrix} L_1 & G \\ 0 & L_2 \end{bmatrix} = [X_1, X_2] \begin{bmatrix} L_1 & 0 \\ 0 & L_2 \end{bmatrix} \begin{bmatrix} Y_1^* \\ Y_2^* \end{bmatrix}, \tag{3.3}$$

where $[Y_1, Y_2]^* = [X_1, X_2]^{-1}$, and

$$[X_1, X_2] = \begin{bmatrix} I & P \\ 0 & I \end{bmatrix}, \qquad [Y_1, Y_2] = \begin{bmatrix} I & 0 \\ -P^* & I \end{bmatrix}, \tag{3.4}$$

and P is the unique solution of the Sylvester equation $L_1P - PL_2 = -G$. Note that

$$\|X_2\| \le 1 + \|P\|, \quad \|Y_1\| \le 1 + \|P\|, \quad \text{where} \quad \|P\| \le \frac{\|G\|}{\text{sep}(L_1, L_2)}, \tag{3.5}$$

and $\text{sep}(L_1, L_2)$ is the separation function[5]. It is known, cf., e.g., [36, Th.V.2.3], that $\text{sep}(L_1, L_2) \le \min_{i,j} |\lambda_i(L_1) - \lambda_j(L_2)|$ where $\lambda_i(X)$ denotes the ith eigenvalue of X.

[4] By a discrete, linear ill-posed problem we mean one whose condition number is of the order of magnitude $1/\mu$, where μ is the unit round off of the floating point system.

[5] The sep function is defined as $\text{sep}(L_1, L_2) = \inf_{\|P\|=1} \|T(P)\|$ where $T : P \mapsto L_1P - PL_2$ (cf., e.g., [36, sec.V.2.1])

DEFINITION 3.1. [37, p. 36] *The grade of a matrix L with respect to a vector v is the degree of the lowest degree monic polynomial p such that $p(L)v = 0$.*

The polynomial giving the grade is unique and it is called in the literature the minimum polynomial; see [14], [26] and references therein. In this paper we shall adopt the term *grade polynomial*, to avoid confusion with the minimum residual GMRES polynomial.

3.1. Using a General Minimal Residual Projection Method. We start by providing an upper bound for the residual norm for a general projection method onto a given subspace, enforcing a minimal residual condition. The bound depends on $\|L_2\|\,\|L_1^{-1}\|$, and therefore it is useful only when this quantity is small. This is often the case in the singular preconditioner application, see Section 4. Sharper bounds for GMRES, applicable in the case of ill-determined numerical rank, will be obtained in Section 3.2.

The case when $G = 0$ and $L_2 \neq 0$ provides the simplest generalization of the singular case described in Proposition 2.1. In this setting, it is possible to estimate the residual of an approximate solution as follows. Let

$$B_0 = \begin{bmatrix} L_1 & 0 \\ 0 & L_2 \end{bmatrix},$$

where $L_1 \in \mathbb{R}^{m\times m}$, and let $V_m \in \mathbb{R}^{n\times m}$,

$$V_m = \begin{bmatrix} V_m^{(1)} \\ V_m^{(2)} \end{bmatrix}$$

be a full column rank matrix such that $V_m^{(1)} \in \mathbb{R}^{m\times m}$ is orthogonal and $\|V_m^{(2)}\| \leq 1$. We can now write

$$\|B_0V_my - c\|^2 = \left\| \begin{bmatrix} L_1V_m^{(1)}y - c^{(1)} \\ L_2V_m^{(2)}y - c^{(2)} \end{bmatrix} \right\|^2 = \|L_1V_m^{(1)}y - c^{(1)}\|^2 + \|L_2V_m^{(2)}y - c^{(2)}\|^2.$$

Choosing $z^{(1)} = V_m^{(1)}y^{(1)} = L_1^{-1}c^{(1)}$, the first term is made equal to zero, and we can estimate

$$\begin{aligned} \min_{x\in\mathrm{range}(V_m)} \|B_0x - c\| = \min_{y\in\mathbb{R}^m} \|B_0V_my - c\| &\leq \|B_0V_my^{(1)} - c\| \\ &= \|L_2V_m^{(2)}y^{(1)} - c^{(2)}\| \leq \|L_2\|\,\|L_1^{-1}\|\|c^{(1)}\| + \|c^{(2)}\|. \end{aligned}$$

Then for any full column rank matrix $V_{m+j} = [V_m\ V_j] \in \mathbb{R}^{n\times(m+j)}$ we have

$$\min_{x\in\mathrm{range}(V_{m+j})} \|B_0x - c\| \leq \min_{x\in\mathrm{range}(V_m)} \|B_0x - c\| \leq \|L_2\|\,\|L_1^{-1}\|\|c^{(1)}\| + \|c^{(2)}\|. \quad (3.6)$$

Next consider the case $G \neq 0$. We recall from (3.3) that B can be block-diagonalized, $B = XB_0X^{-1}$, and we have the following result.

THEOREM 3.2. *Let $\chi = \|G\|/\mathrm{sep}(L_1, L_2)$, and assume that we have chosen a full column rank matrix $W_m \in \mathbb{R}^{n\times m}$ such that the matrix V_m in the decomposition $V_mR_m = X^{-1}W_m$ has an orthogonal leading $m\times m$ block $V_m^{(1)}$, and such that $\|V_m^{(2)}\| \leq 1$ and R_m is nonsingular. Then the residual r_{m+j} of any minimal residual projection*

method onto range(W_{m+j}), where $W_{m+j} = [W_m\ W_j]$ has full column rank, satisfies

$$\begin{aligned}\|r_{m+j}\| &= \min_{x\in\text{range}(W_{m+j})} \|Bx - c\| = \min_y \|BW_{m+j}y - c\| \\ &\le (1+\chi)\left(\|L_1^{-1}\|\|L_2\|(1+\chi)\|c^{(1)}\| + \|c^{(2)}\|\right).\end{aligned}$$

Proof. We have

$$\begin{aligned}\|BW_m y - c\| &= \|XB_0X^{-1}W_m y - c\| = \|X(B_0X^{-1}W_m y - X^{-1}c)\| \\ &= \|X(B_0V_m(R_m y) - X^{-1}c)\| = \|X(B_0V_m z - u)\|,\end{aligned}$$

with $z = R_m y$ and $u = X^{-1}c$. For the quantity in parentheses, we can thus proceed in a way similar to the case $G = 0$, i.e., we choose $z^{(1)} = (V_m^{(1)})^* L_1^{-1} u^{(1)}$. Recalling the bound (3.5) for $\|X_2\|$, we obtain

$$\begin{aligned}\min_y \|BW_m y - c\| &\le \|X_2\|\|L_2V_m^{(2)}z^{(1)} - u^{(2)}\| \\ &\le (1+\|P\|)\|L_2V_m^{(2)}z^{(1)} - u^{(2)}\| \\ &\le (1+\chi)\left(\|L_1^{-1}\|\|L_2\|(1+\chi)\|c^{(1)}\| + \|c^{(2)}\|\right),\end{aligned}$$

where we have used $\|u^{(1)}\| \le (1+\|P\|)\|c^{(1)}\|$. ∎

Remark 3.3. *The procedure above is not restricted to minimal residual methods. Any approach that generates y as the solution to the projected problem restricted to L_1 will give the same bound.*

Example 3.4. The example presented in Section 5.2 has numerical rank 20, and we solve it using GMRES. The relevant quantities of the residual estimate are $\|L_1^{-1}\| = 1.58$, $\|L_2\| = 5.8\cdot 10^{-16}$, $\chi = 0.14$, and the true residual after a few iterations is $6.2\cdot 10^{-3}$. The estimate gives

$$\|r_{m+j}\| \le 7.2\cdot 10^{-3},$$

which slightly overshoots the true value by a factor $1+\chi$, approximately. Note that this residual is attained already after 4 GMRES steps.

Clearly, the bound in Theorem 3.2 does not take into account the information contained in range(V_k) and therefore it holds for any minimal residual projection method. On the other hand, in some situations the bound may be weak, as $\|L_2\|\,\|L_1^{-1}\|\|c^{(1)}\|$ only depends on the singular value gap between L_1 and L_2, and it may be significantly larger than $\|c^{(2)}\|$. Therefore, in some cases the bound may highly overestimate the actual residual, which is expected to be of the order of $\|c^{(2)}\|$ for $j \ge 1$ in (3.6). Sharper bounds that are useful when the gap is small may be obtained by taking into account specific choices of range(V_k), as we will do in Section 3.2 for GMRES.

The following variant of the theorem can be used in a situation when $L_1V_k^{(1)}z = u^{(1)}$ is consistent for $k < m$ (recall that $u = X^{-1}c$).

Theorem 3.5. *Let $V_kR_k = X^{-1}W_k$, and assume the leading $m\times k$ block $V_k^{(1)}$ has full column rank and is such that $L_1V_k^{(1)}z = u^{(1)}$ is consistent. Then*

$$\begin{aligned}\|r_k\| &= \min_y \|BW_k\, y - c\| \\ &\le (1+\chi)\left(\|L_2V_k^{(2)}\|\,\|(L_1V_k^{(1)})^\dagger\|(1+\chi)\|c^{(1)}\| + \|c^{(2)}\|\right).\end{aligned}$$

Proof. We have $\|BW_k y - c\| = \|X(B_0 V_k z - u)\|$. Since $V_k^{(1)}$ has full column rank, so has $L_1 V_k^{(1)}$, and the vector $z^{(1)}$ such that $L_1 V_k^{(1)} z^{(1)} = u^{(1)}$ is unique, and can be written $z^{(1)} = (L_1 V_k^{(1)})^\dagger u^{(1)}$. Then,

$$\|L_2 V_k^{(2)} z^{(1)}\| \le \|L_2 V_k^{(2)}\| \, \|(L_1 V_k^{(1)})^\dagger\| \, \|u^{(1)}\|,$$

and the result is obtained as in the proof of Theorem 3.2. ∎

Consider now the case when the projection matrix W_k is constructed using the Arnoldi method, and the grade m_* of L_1 with respect to $u^{(1)}$ is smaller than m. To apply Theorem 3.5 we must show that $V_{m_*}^{(1)}$ has full column rank. We first consider the linear system $B_0 x = u$. Due to the structure of B_0, we generate the Krylov subspace

$$\mathcal{K}_k(B_0, c) = \operatorname{span}\left\{ \begin{pmatrix} u^{(1)} \\ u^{(2)} \end{pmatrix}, \begin{pmatrix} L_1 u^{(1)} \\ L_2 u^{(2)} \end{pmatrix}, \ldots, \begin{pmatrix} L_1^{m_*-1} u^{(1)} \\ L_2^{m_*-1} u^{(2)} \end{pmatrix} \right\}. \tag{3.7}$$

The following lemma shows that the full column rank assumption is satisfied.

LEMMA 3.6. *Assume that the columns of the matrix*

$$V_{m_*} = \begin{bmatrix} V_{m_*}^{(1)} \\ V_{m_*}^{(2)} \end{bmatrix} \in \mathbb{R}^{n \times m_*}$$

constitute an orthonormal basis of the Krylov subspace (3.7)*. Then the upper* $m \times m_*$ *block* $V_{m_*}^{(1)}$ *has full column rank.*

Proof. Let $K^{(i)} = [u^{(i)}, L_i u^{(i)}, \ldots, L_i^{m_*-1} u^{(i)}]$ and

$$K = \begin{bmatrix} K^{(1)} \\ K^{(2)} \end{bmatrix}.$$

The columns of $K^{(1)}$ are linearly independent, otherwise the zero linear combination would imply the existence of a polynomial p of degree strictly less than m_* such that $p(L^{(1)})u^{(1)} = 0$, which is a contradiction with the definition of grade. Therefore, the matrix $K^\top K = (K^{(1)})^\top K^{(1)} + (K^{(2)})^\top K^{(2)}$ is nonsingular, the columns of $V_{m_*} = K(K^\top K)^{-\frac{1}{2}}$ are orthonormal with first block $V_{m_*}^{(1)}$ having full column rank. Any other orthonormal basis differs from V_{m_*} for a right multiplication by a unitary matrix, leaving the full rank property of the first block unchanged. ∎

Next assume that for any $k \ge 1$ we have generated a Krylov (Arnoldi) factorization of B

$$BW_k = W_{k+1} H_k, \qquad w_1 = \frac{\hat{w}}{\|\hat{w}\|}, \quad \hat{w} = \begin{bmatrix} c^{(1)} \\ c^{(2)} \end{bmatrix}, \tag{3.8}$$

with $H_k \in \mathbb{R}^{(k+1)\times k}$ upper Hessenberg. Using the relation $B = X B_0 X^{-1}$ we get the Krylov (Arnoldi) factorization of B_0

$$B_0 V_k = V_{k+1} \widehat{H}_k, \qquad v_1 = \frac{\hat{v}}{\|\hat{v}\|}, \quad \hat{v} = \begin{bmatrix} c^{(1)} - P c^{(2)} \\ c^{(2)} \end{bmatrix} \tag{3.9}$$

where

$$V_k S_k = X^{-1} W_k \tag{3.10}$$

is the thin QR decomposition, and $\widehat{H}_k = S_{k+1} H_k S_k^{-1}$ is upper Hessenberg. Thus, Arnoldi's method applied to B with starting vector c uniquely defines another sequence of vectors, which can be generated by Arnoldi's method applied to B_0 with starting vector v_1 defined by (3.9). Using this correspondence for $k = m_*$, we can prove the following result.

THEOREM 3.7. *Assume that m_* is the grade of L_1 with respect to $u^{(1)} = c^{(1)} - Pc^{(2)}$. If the projection matrix W_{m_*} is constructed using the Arnoldi method, then*

$$\begin{aligned}\|r_{m_*}\| &= \min_y \|BW_{m_*} y - c\| \\ &\le (1+\chi)\left(\|L_2 V_{m_*}^{(2)}\|\,\|(L_1 V_{m_*}^{(1)})^\dagger\|(1+\chi)\|c^{(1)}\| + \|c^{(2)}\|\right),\end{aligned}$$

where W_{m_*} and V_{m_*} are related by (3.10).

Proof. Due to the equivalence between (3.8) and (3.9), the columns of V_{m_*} are an orthonormal basis of the Krylov subspace (3.7). The result now follows from Lemma 3.6 and Theorem 3.5. □

We expect that for general problems m_* will be close to m. However, in the case of preconditioning of ill-posed equations, L_1 may have extremely close or even multiple eigenvalues (depending on the quality of the preconditioner), so that the method behaves as if L_1 had a grade smaller than m. This phenomenon is largely due to the polynomial nature of Krylov subspaces, and it is further explored in the next section.

3.2. Using optimal GMRES. In this section we show how much one can gain by exploiting the approximation properties of Krylov subspaces, in the context of minimal residual methods. For any polynomial p_m of degree not greater than m we can write

$$p_m(B)c = [X_1, X_2]\begin{bmatrix} p_m(L_1)Y_1^* c \\ p_m(L_2)Y_2^* c \end{bmatrix} = X_1 p_m(L_1)Y_1^* c + X_2 p_m(L_2)Y_2^* c,$$

so that, using $X_1^T X_1 = I$, and $Y_2^* c = c^{(2)}$,

$$\|p_m(B)c\| \le \|p_m(L_1)Y_1^* c\| + \|X_2 p_m(L_2)c^{(2)}\|. \tag{3.11}$$

We denote by $\mathbb{P}_k$ the set of polynomials p of degree not greater than k and such that $p(0) = 1$. We also recall that k iterations of GMRES generate an approximate solution x_k for $Bx = c$ with $x_k \in K_k(B, c)$ (for a zero initial guess) by minimizing the residual $r_k = c - Bx_k$ [34]. In terms of polynomials, this implies that $r_k = p_k(B)c$ where $p_k = \arg\min_{p\in\mathbb{P}_k} \|p(B)c\|$; p_k is called the GMRES residual polynomial.

The following theorem provides a description of the GMRES convergence when the spectra of L_1 and L_2 are well separated, and the magnitude of $c^{(2)}$ is small compared to that of the whole vector c, as is the case in our setting. The proof is in the spirit of that in [9].

THEOREM 3.8. *Let m_* be the grade of L_1 with respect to $Y_1^* c$. Assume k iterations of GMRES have been performed on $Bx = c$, and let r_k be the corresponding residual. Let Δ_2 be a circle centered at the origin and radius ρ, enclosing all eigenvalues of L_2.*

i) If $k < m_*$, let $s_k^{(1)} = \phi_k(L_1)Y_1^* c$ be the GMRES residual associated with $L_1 z = Y_1^* c$, where $\phi_k \in \mathbb{P}_k$. Then

$$\|r_k\| \le \|s_k^{(1)}\| + \|X_2\|\gamma_k \tau, \qquad \tau = \rho \max_{z\in\Delta_2} \|(zI - L_2)^{-1} c^{(2)}\|, \tag{3.12}$$

where $\gamma_k = \max_{z\in\Delta_2} \prod_{i=1}^{k} |\theta_i - z|/|\theta_i|$ *and* θ_i *are the roots of* ϕ_k.

ii) If $k = m_* + j$, $j \geq 0$, *let* $s_j^{(2)} = \varphi_j(L_2)c^{(2)}$ *be the GMRES residual associated with* $L_2 z = c^{(2)}$ *after* j *iterations, where* $\varphi_j \in \mathbb{P}_j$, *so that* $\|s_j^{(2)}\| \leq \|c^{(2)}\|$. *Then*

$$\|r_k\| \leq \rho\gamma_{k_*}\|s_j^{(2)}\|\|X_2\| \max_{z\in\Delta_2} \|(zI - L_2)^{-1}\|, \tag{3.13}$$

where $\gamma_{m_*} = \max_{z\in\Delta_2} \prod_{i=1}^{m_*} |\theta_i - z|/|\theta_i|$ *and* θ_i *are the roots of the grade polynomial of* L_1.

Proof. Let us write $r_k = p_k(B)c$, where p_k is the GMRES residual polynomial.

i) For $k < m_*$, we have $\|r_k\| = \min_{p\in\mathbb{P}_k} \|p(B)c\| \leq \|\phi_k(B)c\|$, where ϕ_k is the GMRES residual polynomial associated with L_1 and $Y_1^* c$. Using (3.11), we have

$$\|\phi_k(B)c\| \leq \|\phi_k(L_1)Y_1^* c\| + \|X_2\phi_k(L_2)c^{(2)}\| \leq \|s_k^{(1)}\| + \|X_2\|\,\|\phi_k(L_2)c^{(2)}\|.$$

To evaluate the last term we use the Cauchy integral representation. ¿From $\phi_k(L_2)c^{(2)} = \frac{1}{2\pi i}\int_{\Delta_2} \phi_k(z)(zI - L_2)^{-1}c^{(2)}dz$, we obtain

$$\|\phi_k(L_2)c^{(2)}\| \leq \rho \max_{z\in\Delta_2} |\phi_k(z)| \max_{z\in\Delta_2} \|(zI - L_2)^{-1}c^{(2)}\|.$$

Using $\phi_k(z) = \prod_{i=1}^{k}(1 - \frac{z}{\theta_i})$, the first result follows.

For $k \geq m_*$, we select the polynomial $p_k(z) = q_{m_*}(z)\varphi_j(z)$, where q_{m_*} is the grade polynomial, namely it satisfies $q_{m_*}(L_1)Y_1^* c = 0$, so that $p_k(L_1)Y_1^* c = 0$; moreover, $\varphi_j(z)$ is the GMRES residual polynomial after j iterations on $L_2 z = c^{(2)}$. Then

$$\begin{aligned}\|r_k\| &\leq \|p_k(B)c\| \leq \|p_k(L_1)Y_1^* c\| + \|X_2 p_k(L_2)c^{(2)}\| \\ &\leq \|X_2\|\,\|p_k(L_2)c^{(2)}\| \leq \|X_2\|\,\|q_{m_*}(L_2)\|\,\|\varphi_j(L_2)c^{(2)}\|.\end{aligned}$$

Once again, using the Cauchy integral representation,

$$\|q_{m_*}(L_2)\| \leq \rho \max_{z\in\Delta_2} |q_{m_*}(z)| \max_{z\in\Delta_2} \|(zI - L_2)^{-1}\|.$$

Since $q_{m_*}(z) = \prod_{i=1}^{m_*}(1 - \frac{z}{\theta_i})$, the result follows. □

A few comments are in order before we proceed with some examples. Assuming that $m_* \ll n$, Theorem 3.8 shows that the behavior of the first few iterations of GMRES is driven by the convergence of the reduced system $L_1 x_1 = Y_1^* c$. During these iterations, the noise-related part of the problem may affect the bound on the overall residual if B is non-normal, otherwise the first term $\|s_k^{(1)}\|$ dominates. Such non-normality reveals itself in two different ways: a) The quantity τ may be large if the second diagonal block L_2 is very non-normal, so that its resolvent norm may be large even for z not too close to the spectrum; b) Due to (3.5), $\|P\|$ and thus $\|X_2\|$ may be large if L_1 and L_2 are not well separated, in terms of sep function, while the norm of the "coupling" matrix G is sizable.

If $G = 0$, then X_2 has orthonormal columns and only the non-normality of L_2 plays a role in the balance between the two terms in (3.12).

For k sufficiently large, we then expect that $\|s_k^{(1)}\|$ will become smaller than the second term in (3.12), so that the second term $\|X_2\|\gamma_k\tau$ will start to dominate. For $k > m_*$, the first term is zero, so that a bound based on the system in L_2 may be obtained, as in (3.13).

We also need to comment on the expected size of τ and γ_k. The quantity τ collects information on the non-normality of L_2, and on the size of the data perturbation. We already mentioned the role of the transfer function norm, which appears as $\|(zI - L_2)^{-1}c^{(2)}\| \leq \|(zI - L_2)^{-1}\|\,\|c^{(2)}\|$. Therefore, the size of the noise-related data, $\|c^{(2)}\|$, may be amplified significantly on a non-normal problem. On the other hand, the radius ρ also plays a role. We recall that $\|(zI - L_2)^{-1}\| \leq \mathrm{dist}(z, \mathcal{F}(L_2))^{-1}$ where $\mathcal{F}(L_2)$ is the field of values[6] of L_2. Therefore, the circle Δ_2 may be set to be sufficiently far from $\mathcal{F}(L_2)$ (see Figure 3.1), so that $\|(zI-L_2)^{-1}\|$ be of moderate size, while maintaining ρ not too large, so as not to influence γ_k (see below). In that case, $\rho\|(zI - L_2)^{-1}\| \ll 1$, implying $\tau \approx \|c^{(2)}\|$. Similar considerations hold for the bound (3.13). The quantity γ_k is the maximum value of the GMRES residual polynomial on the circle Δ_2. If the circle tightly surrounds zero, then γ_k is usually very close to one since the residual polynomial ϕ_k satisfies $\phi_k(0) = 1$. Circles of larger radius may cause γ_k to assume significantly larger values, depending on the location of the polynomial roots θ's. We found that values of the radius ρ within $\|L_1\|$ provided good bounds; in general however we tried to selected values of ρ significantly smaller; see the examples below.

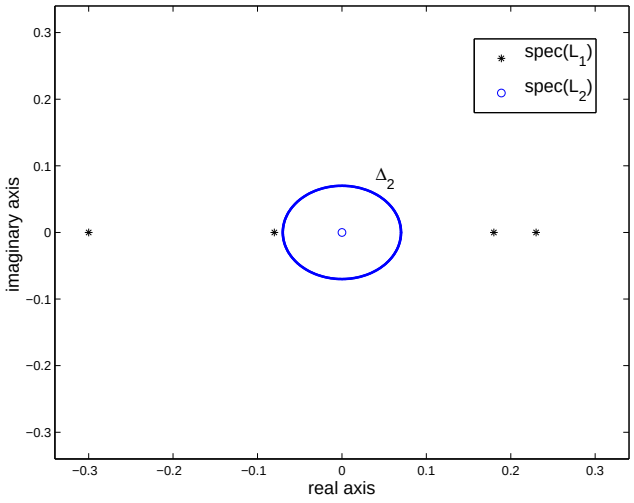

FIG. 3.1. *Location of the spectra of L_1 and L_2, and choice of the circle Δ_2 in Theorem 3.8.*

EXAMPLE 3.9. We consider the `wing` example from the Matlab Regularization Toolbox [20, 22]. We generate the data with dimension $n = 100$, and the largest few eigenvalues of A in absolute value are

```
  3.7471e-01
 -2.5553e-02
  7.6533e-04
```

[6]The field of values of an $n \times n$ matrix L is defined as $\mathcal{F}(L) = \{z^*Lz \,:\, z \in \mathbb{C}^n, \|z\| = 1\}$.

```
-1.4851e-05
 2.1395e-07
-2.4529e-09
 2.3352e-11
-1.8998e-13
 1.3260e-15
```

We perturb the right-hand side b as $\tilde{b} = b + \varepsilon p$, with p having normally distributed random entries and $\|p\| = 1$. With the explicit Schur decomposition of the matrix, we take as L_1 the portion of B corresponding to the largest six eigenvalues in absolute value (that is $m_* = 6$), down to $\lambda_6 = -2.4529 \cdot 10^{-9}$; for this choice we have $\|G\| = 2.29 \cdot 10^{-5}$ and $\|P\| = 10.02$. This choice of L_1 was used to ensure that there is a sufficiently large gap between L_1 and L_2, while still being able to assume that $\|L_2\|$ is mainly noise. Note that since all relevant eigenvalues are simple, $m_* = m$ for this example. We then take a circle of radius $\rho = 2 \cdot 10^{-9} < \mathrm{dist}(\mathrm{spec}(L_1), 0)$. We compute the invariant subspace basis $[X_1, X_2]$ as in (3.4), where P was obtained by solving the associated Sylvester equation.

We note that for $\varepsilon = 10^{-7}$ we have $\|Y_1^* c\| = 1$ and $\|Y_2^* c\| = 6.7 \cdot 10^{-7}$, while for $\varepsilon = 10^{-5}$ we obtain $\|Y_2^* c\| = 6.49 \cdot 10^{-5}$; all these are consistent with the used perturbation ε.

Table 3.1 reports some key quantities in the bound of Theorem 3.8 for a few values of ε at different stages of the GMRES convergence. For $k < m_* = 6$ we see that the two addends of the bound in (3.12) perform as expected: $\|s_k^{(1)}\|$ dominates for the first few iterations, after which the second term leads the bound, providing a quite good estimate of the true residual norm, $\|r_k\|$. A larger perturbation ε makes this dominance effect more visible at an earlier stage.

TABLE 3.1

Example 3.9. Wing data. Key quantities of Theorem 3.8. L_1 of size 6×6 ($m_* = 6$), so that $\|G\| = 2.29 \cdot 10^{-5}$ and $\|P\| = 10.02$. Circle of radius $\rho = 2 \cdot 10^{-9}$.

ε	k	$\|s_k^{(1)}\|$	$\|X_2\|\gamma_k\tau$	Bound (3.12) or (3.13)	$\|r_k\|$
10^{-7}	2	1.640e-03	6.770e-06	1.647e-03	1.640e-03
	3	3.594e-05	6.770e-06	4.271e-05	3.573e-05
	10			6.712e-06	6.311e-07
10^{-5}	2	1.621e-03	6.770e-04	2.298e-03	1.640e-03
	3	6.568e-05	6.770e-04	7.427e-04	7.568e-05
	10			6.442e-04	6.308e-05

EXAMPLE 3.10. We consider the `baart` example from the same toolbox as in the previous example. This example will be considered again in later sections. The leading eigenvalues for the 100×100 matrix are

```
 2.5490e+00
-7.2651e-01
 6.9414e-02
-4.3562e-03
 2.0292e-04
-7.5219e-06
 2.3168e-07
-6.1058e-09
```

TABLE 3.2
Example 3.10. Baart data. Key quantities of Theorem 3.8. L_1 of size 7×7 ($m_* = 7$), so that $\|G\| = 6.4357 \cdot 10^{-3}$ and $\|P\| = 1.48$. Circle of radius $\rho = 2 \cdot 10^{-7}$.

ε	k	$\|s_k^{(1)}\|$	$\|X_2\|\gamma_k\tau$	Bound (3.12) or (3.13)	$\|r_k\|$
10^{-7}	2	1.590e-02	5.851e-08	1.590e-02	1.590e-02
	3	5.105e-06	5.851e-08	5.165e-06	5.105e-06
	10			1.062e-07	3.188e-08
10^{-5}	2	1.590e-02	5.851e-06	1.590e-02	1.590e-02
	3	5.404e-06	5.851e-06	1.125e-05	6.110e-06
	10			1.062e-06	3.188e-06

```
  1.4064e-10
 -2.8770e-12
  5.2962e-14
```

We consider $m_* = 7$, giving $\|G\| = 6.4357 \cdot 10^{-3}$ and $\|P\| = 1.48$, and we chose $\rho = 2 \cdot 10^{-7}$. Also in this case, $m_* = m$ as all involved eigenvalues are simple. For $\varepsilon = 10^{-7}$ we have $\|Y_1^* c\| = 1$ and $\|Y_2^* c\| = 3.26 \cdot 10^{-8}$, while for $\varepsilon = 10^{-5}$ we obtain $\|Y_2^* c\| = 3.26 \cdot 10^{-6}$.

Table 3.2 reports some key quantities in the bound of Theorem 3.8 for a few values of ε at different stages of the GMRES convergence.

The digits in the table fully confirm what we found in the previous example, although here the addend carrying the perturbation is less dominant in the early phase of the convergence history.

Since $\|L_1^{-1}\|\,\|L_2\| \approx 0.061$, we see that Theorem 3.2 gives a much worse residual estimate for this example, where the eigenvalues are not so well separated as in Example 3.4.

3.3. On the Solution Error. The GMRES algorithm delivers a monotonically non-decreasing residual norm $\|r\| = \|c - By\|$, and we have shown that under certain spectral hypotheses on B, this norm can be sufficiently small. We will now discuss the error in that approximation. Assume first that both $\|G\|$ and $\|L_2\|$ are much smaller than $\sigma_m(L_1)$ and can be considered as small perturbations of an underlying noise-free linear system $L_1 x = c^{(1)}$ with solution $x^{(1)}$. The residual of the approximate solution $y^{(1)}$ can be bounded as follows:

$$\|L_1 y^{(1)} - c^{(1)}\| \le \|G\|\,\|y^{(2)}\| + \|r\|, \qquad y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \end{bmatrix}.$$

Using standard perturbation theory for linear systems, see e. g. [25, Section 7.1], and putting $\kappa(L_1) = \sigma_1(L_1)/\sigma_m(L_1)$, we get

$$\frac{\|x^{(1)} - y^{(1)}\|}{\|x^{(1)}\|} \le \frac{2\epsilon\kappa(L_1)}{1 - \epsilon\kappa(L_1)}, \qquad \epsilon = \frac{\|G\|\,\|y^{(2)}\| + \|r\|}{\|L_1\|\,\|y^{(1)}\| + \|c^{(1)}\|}.$$

However, the above argument presupposes that the actual linear system is in Schur form and is a perturbation of an underlying system of smaller dimension. In general this may not be a likely situation. Therefore, if $\sigma_m(L_1) \gg \|G\| \gg \|L_2\|$, then it may be more realistic to consider y as an approximate solution of the underdetermined system

$$B_1 x = \begin{bmatrix} L_1 & G \end{bmatrix} x = c^{(1)}. \tag{3.14}$$

Let x_* be the minimum norm solution of (3.14), and put $r^{(1)} = c^{(1)} - B_1 y$; then, of course, $\|r^{(1)}\| \leq \|By - c\| = \|r\|$. In a numerical example we will show that $\|r^{(1)}\|$ can actually be much smaller than $\|r\|$.

We will now bound $x_* - y$. Using a LQ decomposition of B_1 we have

$$B_1 = \begin{bmatrix} L_1 & G \end{bmatrix} = \begin{bmatrix} S & 0 \end{bmatrix} \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} = S \begin{bmatrix} Z_{11} & Z_{12} \end{bmatrix}, \tag{3.15}$$

where S is lower triangular and Z is unitary. We can now prove the following bound for the approximate solution of (3.14).

THEOREM 3.11. *Let y be an approximate solution solution of* (3.14)*, and denote the residual $r^{(1)} = B_1 y - c^{(1)}$. Then, assuming that $\kappa(B_1)\,\|r^{(1)}\|/\|c^{(1)}\| < 1$, we get the bound*

$$\frac{\|y - x_*\|}{\|x_*\|} \leq 2\kappa(B_1)\,\frac{\|r^{(1)}\|}{\|c^{(1)}\|} + \frac{\|G\|}{\sigma_m(B_1)}\,\frac{\|y^{(1)}\|}{\|x_*\|} + \frac{\|y^{(2)}\|}{\|x_*\|} + O(\|r^{(1)}\|^2). \tag{3.16}$$

Proof. Using the LQ factorization of B_1 we have

$$c^{(1)} = B_1 x_* = \begin{bmatrix} S & 0 \end{bmatrix} Z x_* = \begin{bmatrix} S & 0 \end{bmatrix} \begin{bmatrix} \bar{x}_*^{(1)} \\ 0 \end{bmatrix},$$

where the lower part $\bar{x}_*^{(2)}$ of the vector $\bar{x}_* = Z x_*$ is equal to zero due to the fact that x_* is the minimum norm solution. Defining $\bar{y} = Zy$, we then have

$$\|y - x_*\| \leq \|\bar{y}^{(1)} - \bar{x}_*^{(1)}\| + \|\bar{y}^{(2)}\|.$$

Since $\|B_1 y - c^{(1)}\| = \|S\bar{y}^{(1)} - c^{(1)}\|$, $\bar{y}^{(1)}$ is an approximate solution of the system $Sx = c^{(1)}$, and we can use standard theory for linear systems [25, Section 7.1], and bound

$$\frac{\|\bar{x}_*^{(1)} - \bar{y}^{(1)}\|}{\|\bar{x}_*^{(1)}\|} \leq \frac{2\epsilon\kappa(B_1)}{1 - \epsilon\kappa(B_1)}, \qquad \epsilon = \frac{\|r^{(1)}\|}{\|B_1\|\|\bar{y}^{(1)}\| + \|c^{(1)}\|}.$$

We immediately get

$$\frac{\|\bar{x}_*^{(1)} - \bar{y}^{(1)}\|}{\|\bar{x}_*^{(1)}\|} \leq 2\kappa(B_1)\frac{\|r^{(1)}\|}{\|c^{(1)}\|} + O(\|r^{(1)}\|^2).$$

It remains to bound $\|\bar{y}^{(2)}\| = \|Z_{21} y^{(1)} + Z_{22} y^{(2)}\|$. Using the LQ decomposition (3.15) we have $\begin{bmatrix} Z_{11} & Z_{12} \end{bmatrix} = \begin{bmatrix} S^{-1} L_1 & S^{-1} G \end{bmatrix}$, which, since $\|S^{-1}\| = 1/\sigma_m(B_1)$, gives the inequality $\|Z_{12}\| \leq \|G\|/\sigma_m(B_1)$. The unitarity relations for Z imply that the eigenvalues of $Z_{21}^* Z_{21}$ are the same as those of $Z_{12} Z_{12}^*$, so the singular values of Z_{12} are the same as those of Z_{21}, which means that Z_{21} satisfies

$$\|Z_{21}\| \leq \frac{\|G\|}{\sigma_m(B_1)}.$$

Then, since $\|Z_{22}\| \leq 1$, we get the second and third terms in (3.16). ∎

The first term in the error bound (3.16) depends on the conditioning of the problem and on how small a residual is produced by the solution method (see also the

last paragraph of this section). The second and third terms depend on the null-space component of the approximate solution. The bound indicates that if we want the computed solution to be a good approximation of x_* we should iterate in the GMRES procedure until the residual $By-c$ is of the order δ, so that the first term in the error estimate is small. On the other hand, we do not want to perform too many iterations, because the third term with $\|y^{(2)}\|$ grows as the iteration proceeds. Since in step k we have

$$y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \end{bmatrix} = \begin{bmatrix} W_k^{(1)} \\ W_k^{(2)} \end{bmatrix} z,$$

for some z, it is sufficient to study the growth of $W_k^{(2)}$, with $\|W_1^{(2)}\| \approx \|c^{(2)}\|/\|c^{(1)}\|$, which is assumed to be small. In the following theorem we give a bound for the growth as the GMRES iterations proceed.

THEOREM 3.12. *Let $BW_k = W_{k+1}H_k$ be the Arnoldi factorization after k steps of GMRES for the system $By = c$. Then*

$$\|W_{k+1}^{(2)}\|_F \le \prod_{j=1}^{k} \eta_j \, \frac{\|c^{(2)}\|}{\|c^{(1)}\|}, \qquad \eta_j \le \left(\frac{1}{h_{j+1,j}^2} (\|B\| + \|L_2\|)^2 + 1 \right)^{\frac{1}{2}}. \tag{3.17}$$

Proof. Let $H_k = \widehat{H}_k + h_{k+1,k} e_{k+1} e_k^T$, where $\widehat{H}_k = W_k^* B W_k$. Then the lower part of the Arnoldi factorization can be written $L_2 W_k^{(2)} = W_k^{(2)} \widehat{H}_K + h_{k+1,k} w_{k+1}^{(2)} e_k^T$, which gives

$$\|w_{k+1}^{(2)}\| \le \frac{1}{|h_{k+1,k}|} (\|\widehat{H}_k\| + \|L_2\|) \|W_k^{(2)}\| \le \frac{1}{|h_{k+1,k}|} (\|B\| + \|L_2\|) \|W_k^{(2)}\|_F. \tag{3.18}$$

Squaring this inequality, adding $\|W_k^{(2)}\|_F^2$ to both sides, and taking the square root, we get

$$\|W_{k+1}^{(2)}\|_F \le \eta_k \|W_k^{(2)}\|_F,$$

which leads to (3.17). ∎

REMARK 3.13. *The last inequality in* (3.18) *does not take into account that always $\|W_k^{(2)}\| \le 1$, while $\|W_k^{(2)}\|_F$ can be larger than 1. On the other hand, it is the initial growth of $\|W_k^{(2)}\|$ and $\|W_k^{(2)}\|_F$ that is interesting. Well before $\|W_k^{(2)}\|$ is starting to get close to 1, the iterations should be stopped.*

In order to balance the decreasing and the increasing parts of the error bound, we use the discrepancy principle as stopping criterion: As soon as the residual norm is of the order δ, we stop the iterations, thereby iterating long enough so that the first term in the bound (3.16) becomes small, and at the same time avoiding the null-space component to grow.

Next consider the case when the eigenvalues are not well clustered and separated. This situation accurs frequently when dealing with unpreconditioned ill-posed problems. Our Theorem 3.8 shows that GMRES approximates well the solution components that are associated with the largest eigenvalues, and as the iterations proceed, smaller eigenvalues come into play. Assuming for the moment that the right-hand side is exact, the solution can be written in the form

$$\begin{bmatrix} y^{(1)} \\ y^{(2)} \end{bmatrix} = \begin{bmatrix} L_1^{-1}(c^{(1)} - G y^{(2)}) \\ L_2^{-1} c^{(2)} \end{bmatrix},$$

no matter how we partition the problem, i.e., no matter what dimension we choose for L_1. For the solution to be bounded this requires that there is a fast decay of the components of c, since L_2 has very small eigenvalues. After a few iterations GMRES gives a reasonable approximation of $L_1^{-1}c^{(1)}$. Now, as is customary in the analysis of the stabilizations for ill-posed problems, let δc be a perturbation of the right-hand side, and δy be the corresponding perturbation of the solution. Then we have

$$\begin{bmatrix} \delta y^{(1)} \\ \delta y^{(2)} \end{bmatrix} = \begin{bmatrix} L_1^{-1}(\delta c^{(1)} - G\delta y^{(2)}) \\ L_2^{-1}\delta c^{(2)} \end{bmatrix}.$$

As long as L_1 is well-conditioned and the small eigenvalues do not influence the solution, the perturbation will remain harmless. But it will blow up gradually as the iterations proceed. The optimal number of steps is the one that gives an accurate approximate solution at the same time as the blow-up of the data perturbation is kept low. The natural stopping criterion is the discrepancy principle, see the analysis in [8].

In our numerical experiments we have observed that GMRES applied to (3.1) can produce approximate solutions y such that $\|r^{(1)}\| = \|B_1 y - c^{(1)}\| \ll \|By - c\| = \|r\|$. In actual large-scale computations we do not have access to the Schur decomposition[7], so we cannot obtain $r^{(1)}$. However, consider the the quantity

$$B^* r = B^* \begin{bmatrix} r^{(1)} \\ r^{(2)} \end{bmatrix} = \begin{bmatrix} L_1^* r^{(1)} \\ G^* r^{(1)} + L_2^* r^{(2)} \end{bmatrix}.$$

Since we have assumed that $\|L_2\| \ll \|L_1\|$, we see that the occurrence that $\|B^* r\| \ll \|r\|$ gives an indication that $\|r^{(1)}\|$ is considerably smaller that $\|r\|$. The same is true if $\|A^* s\| \ll \|s\|$, where $s = b - Ax$, since $\|A^* s\| = \|B^* r\|$. This is illustrated in Figure 5.9. In light of these considerations, we would like to encourage monitoring $\|A^* s\|$ during the GMRES iterations as a companion of a stopping criterion based on the discrepancy principle.

4. Singular Preconditioners for Ill-posed Problems. Preconditioners are used routinely for solving linear systems $Ax = b$ using Krylov methods. For the discussion we first assume that the matrix A corresponds to a well-posed problem, by which we mean that its condition number is of moderate magnitude. We will be concerned with right preconditioners. Usually one derives and computes a nonsingular approximation M of A and then solves the equivalent linear system

$$AM^{-1}y = b, \qquad x = M^{-1}y,$$

using the Krylov method. The reason why we use a right preconditioner is that we will apply the *discrepancy principle* [12, p. 83],[21, p. 179], which means that we are not interested in solving the linear system $Ax = b$ exactly, but only determine an approximation $\hat{x}$ with residual $\|A\hat{x} - b\| \approx \delta$. With a right preconditioner we can read off the magnitude of the residual directly in the GMRES iterations, actually without computing $\hat{x}$ explicitly.

Assume, for instance, that A represents a differential operator with variable coefficients. Then M may be a discretization of the corresponding operator with constant coefficients. Apart from being a good approximation, M should be chosen so that it

[7] Either because it is too expensive to compute, or the matrix A is not available explicitly, see Sections 4 and 5.3.

is easy (cheap) to solve a system $Mz = v$. That is so in the above constant coefficient case if the geometry of the problem allows for a fast solver. In other cases, when A has Toeplitz structure, a circulant preconditioner M may be used, for which the system $Mz = v$ can be solved cheaply using the Fast Fourier Transform.

Consider now a linear system of equations $Ax = b$, which represents a discrete, ill-posed problem, i.e. the matrix A is a discretization of an operator equation $\mathcal{A}x = b$, where $\mathcal{A}$ is a compact operator [12]. Then A is extremely ill-conditioned, typically with a gradual decay of singular values and a cluster of singular values at zero.

There are several papers [17, 16, 18, 29, 30, 23] that propose the use of preconditioners for large scale discrete ill-posed problems. The problem with such a preconditioner is that if M is a good approximation of A, then also M is very ill-conditioned. Assume for instance that M is a circulant matrix [17], written as

$$M = F\Lambda F^H,$$

where F is the Fourier matrix, and Λ is a diagonal matrix of eigenvalues. In order to "regularize" the preconditioner, the small eigenvalues are replaced by ones, i.e. the preconditioner is chosen as

$$M_I^\dagger = F \begin{bmatrix} \Lambda_1^{-1} & 0 \\ 0 & I \end{bmatrix} F^H.$$

In this paper, motivated by the application to Cauchy problems for a parabolic PDE in two space dimensions, see [32, Part III] and Section 5.3, we instead choose to use another type of regularized, *singular preconditioner*, defined using a semianalytic expansion of the solution of a corresponding heat equation. If we were to use this idea in the case of a circulant preconditioner we would take

$$M_R^\dagger = F \begin{bmatrix} \Lambda_1^{-1} & 0 \\ 0 & 0 \end{bmatrix} F^H.$$

Thus we solve the singular linear system

$$AM_R^\dagger y = b, \tag{4.1}$$

with the GMRES method, and then compute $x = M_R^\dagger y$. A somehow related approach was proposed in [2], where however the singular preconditioner was generated by means of a projection argument, instead of a pseudo-inverse strategy.

As was shown in Section 3, the distribution of eigenvalues of $AM_R^\dagger$ determines the rate of convergence and the quality of the GMRES solution. In fact, the regularized singular preconditioner also induces regularization on the solution. This is clearly seen in the idealized situation when the preconditioner is a truncated singular value decomposition (SVD) of A. Let $A = U\Sigma V^T$, and $M_R = U_k\Sigma_k V_k^T$, a rank-k approximation. Then the system (4.1) becomes $U_k U_k^T y = b$, which is singular with an eigenvalue of multiplicity k equal to 1. Applied to that system, GMRES will give the solution $y = U_k U_k^T b$, according to Proposition 2.1. Then $x = M_R^\dagger y = V_k \Sigma_k^{-1} U_k^T b$, which is the truncated SVD solution of $Ax = b$. In a more general setting, with M_R a good approximation of the low frequency part of A, the application of $M_R^\dagger$ to the result y of the GMRES iterations will filter away the high frequency and null-space parts and the semi-convergence phenomenon will not be visible in the solution. This also means that in this case the process will not be sensitive to the choice of stopping criterion, as long as the number of iterations is not too small.

5. Numerical Examples. In the literature on the numerical solution of inverse and ill-posed problems the main purpose is to construct stable methods for solving unstable problems. Then one often solves problems constructed from equations given as integral and differential equations and adds perturbations to the data to test the stability of methods. The perturbations are also intended to simulate measurement errors that occur in real applications. We will follow this approach below.

5.1. An Ill-Posed Problem. Our first example is a discretization $Kf = g$ of an integral equation of the first kind [1] (test problem `baart` in [20, 22]),

$$\int_0^\pi \exp(s\,\cos t) f(t) dt = 2\sinh(s)/s, \qquad 0 \le s \le \pi/2,$$

with solution $f(t) = \sin t$. The singular values and the eigenvalues of the matrix K of dimension $n = 200$ are illustrated in Figure 5.1.

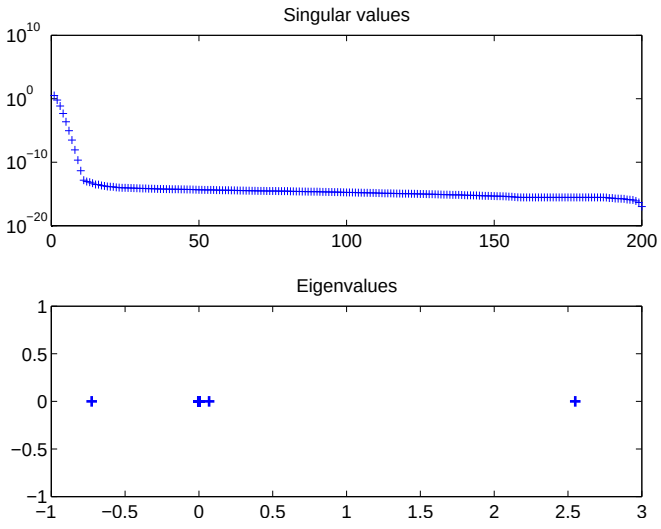

FIG. 5.1. *Singular values and eigenvalues of the matrix K for the* `baart` *problem. Note that all eigenvalues except the three of largest magnitude belong to a cluster at the origin.*

Clearly K is numerically singular. However, it is not easy to decide about its numerical rank. No matter what value, between 2 and 11, of the dimension of L_1 in the ordered Schur decomposition we choose, the smallest singular value of L_1 is much smaller than the norm of G.

We added a normally distributed perturbation of norm 10^{-4} to the right hand side, and performed 10 GMRES steps. In Figures 5.3 and 5.4 we illustrate the approximate solution at iterations 2-5. For comparison we also show the solution using Tikhonov regularization, $\min_f\{\|Kf - g_m\|^2 + \mu^2\|Lf\|^2\}$, where L was a discrete first derivative. The value of the regularization parameter was chosen according to the discrepancy principle: it was successively halved until the least squares residual was smaller than a tolerance, see below.

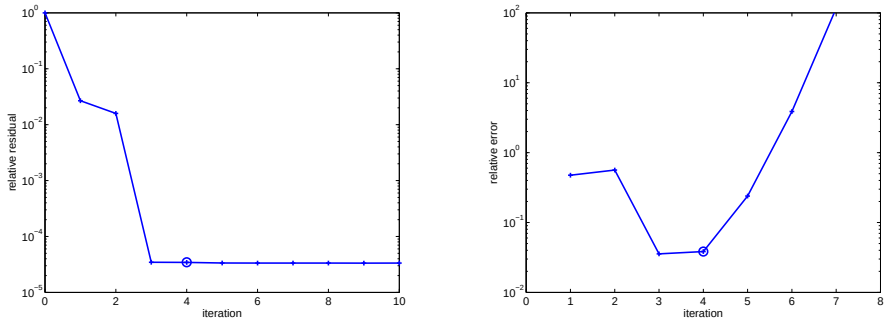

FIG. 5.2. *Baart example: Relative residual (left) and relative error (right) as functions of the GMRES step number. The circle marks when the stopping criterion was first satisfied.*

In Figure 5.2 we give the relative residual and relative error for the GMRES iterations. Clearly the residual stagnates after 3 steps, and the solution starts to diverge after 4. This is also seen in Figures 5.3-5.4.

The discrepancy principle is used as stopping criterion. The data error is $\|g - g_m\| \approx 10^{-4}$. If we choose $m = 4$, then $\|c^{(2)}\| \approx 10^{-4}$. The iterations are stopped when the norm of the residual is smaller than $2 \cdot 10^{-4}$. In Figure 5.2 we mark when the stopping criterion was satisfied. The results agree with those in [27, Example 5.3], and are explained by our theoretical analysis in Sections 3.2 and 3.3.

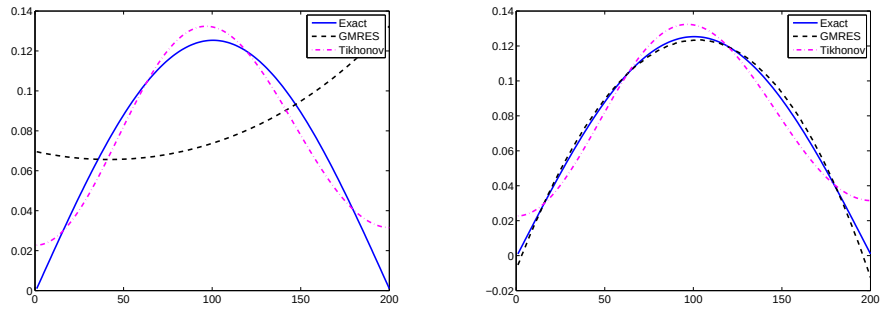

FIG. 5.3. *Baart example: Exact solution (solid), GMRES solution (dashed), and Tikhonov solution for* $\mu = 0.03125$ *(dashed-dotted). Left: after 2 GMRES iterations, right: after 3.*

5.2. A Preconditioned Ill-Posed Problem. In this example we solve numerically a Cauchy problem for a parabolic PDE in the unit square (we will refer to it as Cauchy-1D). The purpose is not to propose a method for solving an ill-posed problem in one space dimension (because there are other, simpler methods for that), but to analyze numerically and illustrate why the preconditioned GMRES method works for the corresponding problem in two space dimensions.

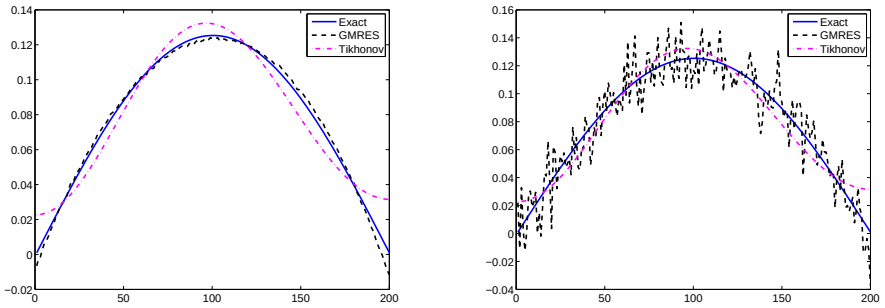

FIG. 5.4. *Baart example: Exact solution (solid), GMRES solution (dashed), and Tikhonov solution (dashed-dotted). Left: after 4 GMRES iterations, right: after 5.*

The Cauchy problem is

$$(a(x)u_x)_x = u_t, \qquad 0 \le x \le 1, \quad 0 \le t \le 1, \tag{5.1}$$
$$u(x,0) = 0, \qquad 0 \le x \le 1, \tag{5.2}$$
$$u_x(1,t) = 0, \qquad 0 \le t \le 1, \tag{5.3}$$
$$u(1,t) = g(t), \qquad 0 \le t \le 1, \tag{5.4}$$

where the parabolic equation has a variable coefficient

$$a(x) = \begin{cases} 1, & 0 \le x \le 0.5, \\ 2, & 0.5 \le x \le 1. \end{cases}$$

The solution $f(t) = u(0,t)$ is sought. This problem, which we call the *sideways heat equation*, is severely ill-posed, see, e.g., [4, 10, 11]. It can be written as a Volterra integral equation of the first kind,

$$\int_0^t k(t-\tau) f(\tau) d\tau = g(t), \qquad 0 \le t \le 1. \tag{5.5}$$

The kernel $k(t)$ is not known explicitly in the case of a variable coefficient $a(x)$. We compute it by solving (using Matlab's stiff solver `ode23s`) a well-posed problem (5.1)-(5.3) and as boundary values at $x = 0$ an approximate Dirac delta function at $t = 0$. The integral equation (5.5) is then discretized giving a linear system of equations

$$Kf = g_m, \tag{5.6}$$

of dimension $n = 200$, where K is a lower triangular Toeplitz matrix, illustrated in Figure 5.5. To construct the data we selected a solution f, solved (5.1)-(5.3) with boundary values $u(0,t) = f(t)$ using Matlab's `ode23s`. The data vector g was then obtained by evaluating the solution at $x = 1$. To simulate measurement errors we added a normally distributed perturbation $g_m = g + \eta$, where $\|\eta\| = 10^{-2}$.

As the diagonal of K is equal to zero, this is an eigenvalue of multiplicity 200, and the assumptions of Section 3 are not satisfied. Therefore it is not surprising that

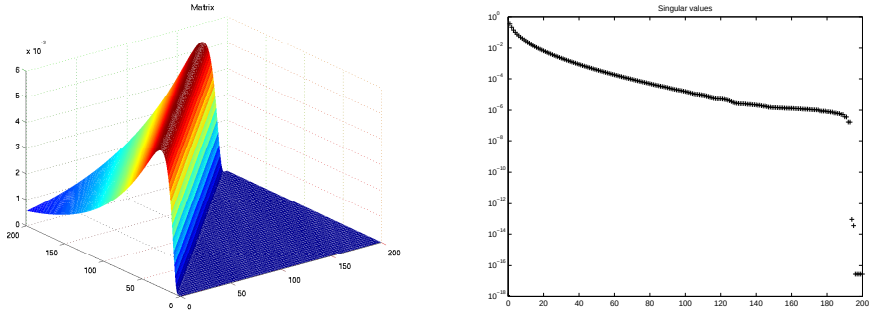

FIG. 5.5. *Cauchy-1D example. Matrix and singular values.*

the linear system (5.6) cannot be solved by GMRES, see [27, Example 5.1] and [8, Example 4.1], where a closely related sideways heat equation is studied.

On the other hand, for this problem the initial decay rate of the singular values is relatively slow, see Figure 5.5, and therefore it should be possible to solve approximately a regularized version of (5.6). To this end we precondition the linear system by a problem with a constant coefficient $a_0 = 1.5$. The kernel functions are given in Figure 5.6.

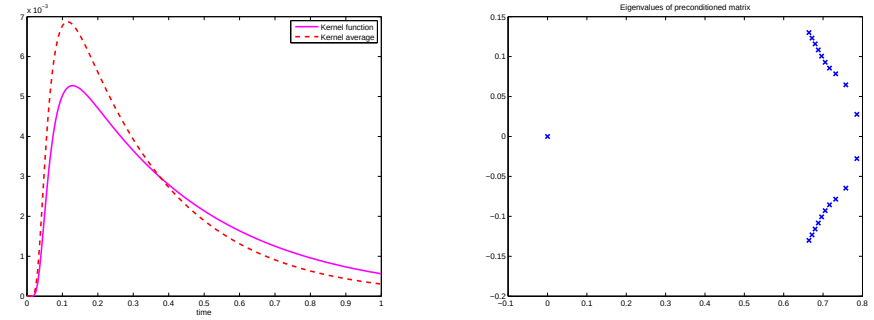

FIG. 5.6. *Cauchy-1D example. Left: Kernel function $k(t)$ for the operator with variable coefficients (solid) and for the constant coefficient (dashed). Right: Eigenvalues of the preconditioned matrix $KM_R^\dagger$.*

For the discretized problem with constant coefficient with matrix K_0 we compute the SVD, $K_0 = U\Sigma V^T$, and define the preconditioner as a truncation to rank $p = 20$ of the pseudoinverse,

$$M_R^\dagger = V_p \Sigma_p^{-1} U_p^T.$$

The eigenvalues of the preconditioned matrix $KM_R^\dagger$ are illustrated in Figure 5.6. Clearly, the numerical rank of $KM_R^\dagger$ is equal to p. We also computed the ordered Schur decomposition (3.1) of $KM_R^\dagger$. The matrix L_1 had condition number $\kappa_2(L_1) =$

$\sigma_1(L_1)/\sigma_k(L_1) = 1.43$, $\|G\| = 0.0962$, and $\|c^{(2)}\| \approx 0.0066$. Thus in this example the data perturbation is larger than $\|c^{(2)}\|$.

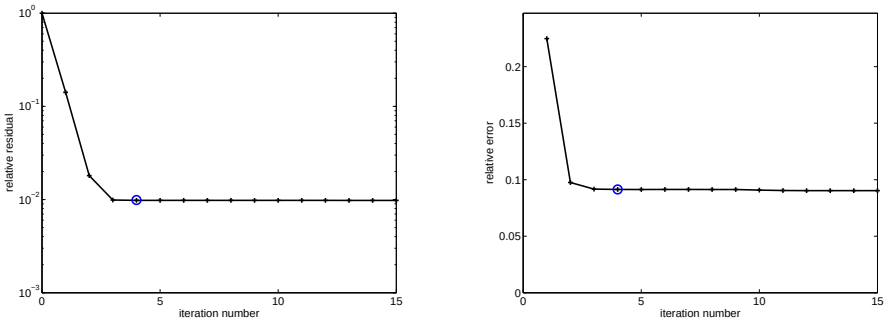

FIG. 5.7. *Cauchy-1D example. Relative residual (left) and relative error (right) as functions of iteration index. The circle marks when the stopping criterion was first satisfied.*

We applied 15 GMRES iterations to the preconditioned system. The relative residual and error are given in Figure 5.7. The numerical solution after 4 steps is illustrated in Figure 5.8, where, for comparison, we also show the solution using Tikhonov regularization, implemented as in the previous example. It is seen that the two approximate solutions have comparable accuracy.

The stopping criterion with $\delta = 0.011$ was satisfied after 4 GMRES steps. From Figure 5.7 we see that the solution accuracy does not deteriorate as the iterations proceed, cf. the last paragraph of Section 4.

Finally, in Figure 5.9 we demonstrate that $\|r^{(1)}\|$ is well approximated by $\|B^*r\|$, and that this part of residual is much smaller than the overall residual $\|r\|$. Here we illustrate 25 GMRES steps to show that after 20 steps the residual for the first part of the system is of the order of the machine precision.

5.3. A Preconditioned 2D Ill-Posed Parabolic Problem. It is in the numerical solution of Cauchy problems for partial differential equations with variable coefficients in two or more space dimensions that the application of a singular preconditioner is particularly interesting. Here we consider the problem

$$\begin{array}{lll} u_t = (a(x)u_x)_x + (b(y)u_y)_y, & 0 < x < 1, \quad 0 < y < 1, & 0 \le t \le 1, \\ u(x,y,0) = 0, & 0 \le x \le 1, \quad 0 \le y \le 1, & \\ u(x,0,t) = u(x,1,t) = 0, & 0 \le x \le 1, \quad 0 \le t \le 1, & \\ u(1,y,t) = g(y,t), & 0 \le y \le 1, \quad 0 \le t \le 1, & \\ u_x(1,y,t) = 0, & 0 \le y \le 1, \quad 0 \le t \le 1, & \end{array} \tag{5.7}$$

where $u(0,y,t) = f(y,t)$ is sought from the Cauchy data at the boundary $x = 1$. The coefficients are

$$a(x) = \begin{cases} 2.5, & 0 \le x \le 0.5, \\ 1.5, & 0.5 < x \le 1, \end{cases}, \qquad b(y) = \begin{cases} 0.75, & 0 \le y \le 0.5, \\ 1.25 & 0.5 < y \le 1. \end{cases}$$

The solution is taken to be

$$f(y,t) = \exp(4 - \frac{1}{y(1-y)}) \exp(4 - \frac{1}{t(1-t)}),$$

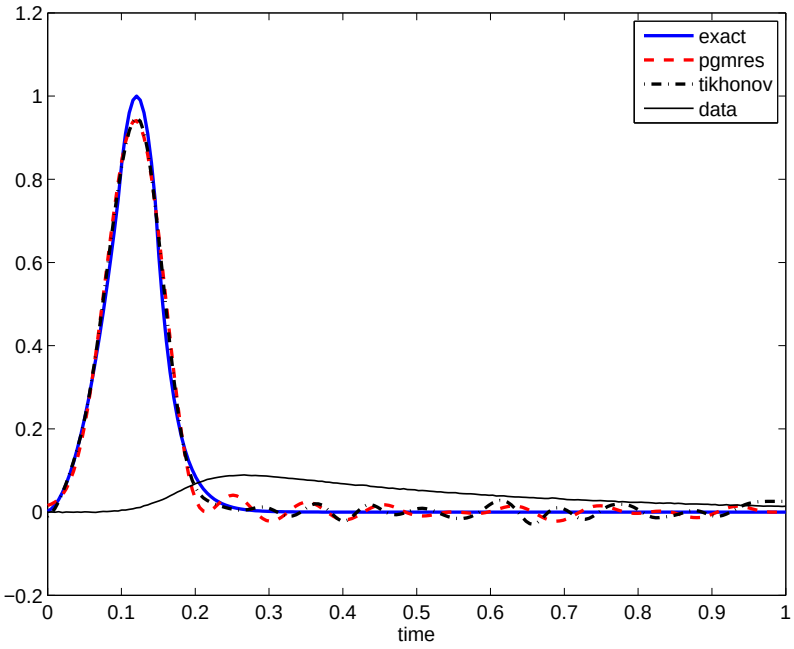

FIG. 5.8. *Cauchy-1D example. Exact solution (solid), approximate solution after 4 iterations of preconditioned GMRES (dashed), and Tikhonov solution with* $\mu = 0.015625$. *The lower solid curve is the right hand side.*

i.e., in order to compute an approximate data function $g(y,t)$ we replace the condition $u(1,y,t) = g(y,t)$ in (5.7) by $u(0,y,t) = f(y,t)$, which gives a well-posed problem. After finite difference discretization with respect to x and y, this problem can be considered as a stiff system of ordinary differential equations of dimension 2500, and is solved using Matlab's `ode23s`. The Cauchy data are then obtained by evaluating the solution at $x = 1$.

A discretization of the problem would give a linear system $Kf = g$. Since we discretize with $n = 50$ equidistant points in both the y and t directions that matrix would have dimension 2500. However, due to the variable coefficients, we cannot compute the matrix; instead, when in GMRES we multiply a vector by K we solve a parabolic equation in the same way as we computed g, described in the previous paragraph.

The preconditioner is based on the approximation of the differential operator by a corresponding one with constant coefficients. Then, since the geometry is rectangular, separation of variables can be applied, and a semi-analytic solution formula can be applied [33] involving an expansion in Fourier series. It is the truncation of this series that leads to a singular preconditioner $M_R^\dagger$, whose rank is equal to nq, where q is the number of terms in the series. Each term in the series involves, in addition, the solution of a 1D ill-posed Cauchy problem, using Tikhonov regularization. The

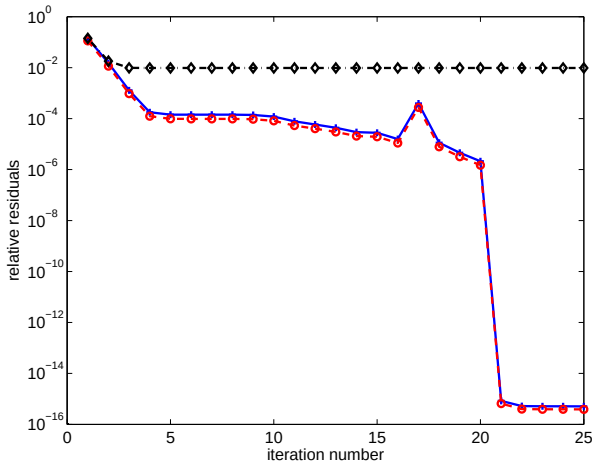

FIG. 5.9. *Cauchy-1D example. Relative residual norm* $\|r\|$ *(diamonds),* $\|r^{(1)}\|$ *(+), and* $\|B^*r\|$ *(o), as functions of iteration index.*

preconditioner is discussed in detail in [31]. In our numerical experiment the data perturbation was equal to $0.5 \cdot 10^{-2}$, the preconditioner regularization parameter was 0.06, and $q = 5$.

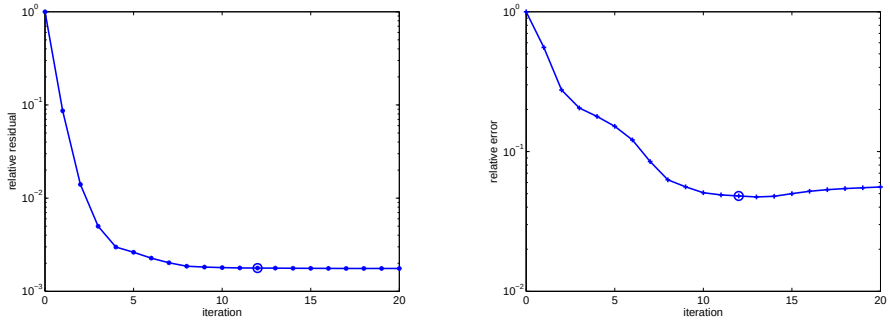

FIG. 5.10. *Relative residual and error as function of the number of iterations. The stopping criterion was satisfied after 12 steps.*

In Figure 5.10 we plot the relative residual and the relative error. Note that also here the solution accuracy is not sensitive to the exact choice of the stopping criterion. The approximate solution after the 12'th iteration, when the residual was first smaller than $0.5 \cdot 10^{-2}$, is shown in Figure 5.11.

6. Conclusions. The main contributions of the present paper are the following. We give an eigenvalue-based analysis of the use of GMRES for almost singular linear

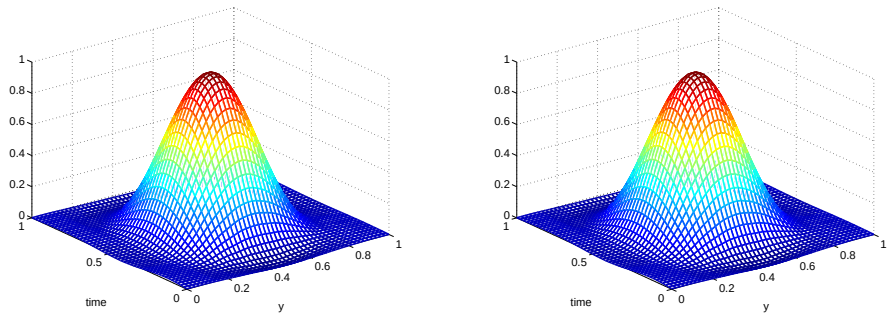

FIG. 5.11. *The solution after 12 iterations (left). To the right the exact solution (solid) and the approximate solution (dashed) at* $t = 0.5$.

systems of equations. The case of "ill-determined numerical rank" (where there is no distinct eigenvalue gap) is treated, which is applicable to the use of GMRES for solving non-selfadjoint ill-posed problems, e.g., certain integral equations of the first kind, and Cauchy problems for PDE's.

We also treat the case of well separated and clustered eigenvalues. This gives a theoretical and algorithmic basis for the use of singular preconditioners for ill-posed problems. Here and in [31] we solve a Cauchy problem for a parabolic equation with variable coefficients, with a singular (low-rank) preconditioner based on a corresponding problem with constant coefficients. It is demonstrated that problems, for which unpreconditioned GMRES does not work, can be solved efficiently using a singular preconditioner.

It is shown that in both cases a stopping criterion based on the discrepancy principle will give a numerical solution that is as good an approximation as it is admissible, given the problem properties and the noise level.

The fact that GMRES with a singular preconditioner can be efficiently applied opens up new possibilities in the numerical solution of ill-posed problems in two and three space dimensions, selfadjoint or non-selfadjoint, linear or non-linear. As soon as it is possible to construct a fast solver for a nearby ill-posed problem, that can be regularized by cutting off high frequencies[8], this can be used as preconditioner. Thus, in each step of the Krylov method a well-posed problem with variable coefficients, need be solved, and a fast solver need be applied as preconditioner. With a good preconditioner only a small number of GMRES (or Conjugate Gradient for selfadjoint and definite problems) steps will have to be performed.

7. Acknowledgement. This work was initiated when the first author was visiting fellow at the Institute of Advanced Study, University of Bologna.

REFERENCES

[1] M. L. Baart. The use of auto-correlation for pseudo-rank determination in noisy ill-conditioned linear least-squares problems. *IMA Journal of Numerical Analysis*, 2(2):241–247, 1982.

[8] This is true, e.g., for a fast Poisson solver for elliptic equations.

[2] James Baglama and Lothar Reichel. Decomposition methods for large linear discrete ill-posed problems. *Journal of Computational and Applied Mathematics*, 198:332–343, 2007.
[3] Christopher A. Beattie, Mark Embree, and D. C. Sorensen. Convergence of polynomial restart krylov methods for eigenvalue computation. *SIAM Review*, 47:492 – 515, 2005.
[4] J. V. Beck, B. Blackwell, and S. R. Clair. *Inverse Heat Conduction. Ill-Posed Problems*. Wiley, New York, 1985.
[5] P. Brianzi, P. Favati, O. Menchi, and F. Romani. A framework for studying the regularizing properties of Krylov subspace methods. *Inverse Problems*, 22:1007–1021, 2006.
[6] Peter N. Brown and Homer F. Walker. GMRES on (nearly) singular systems. *SIAM J. Matrix Anal. Appl.*, 18(1):37 – 51, 1997.
[7] D. Calvetti, B. Lewis, and L. Reichel. GMRES, L-curves, and discrete ill-posed problems. *BIT Numerical Mathematics*, 42(1):44065, 2002.
[8] D. Calvetti, B. Lewis, and L. Reichel. On the regularizing properties of the GMRES method. *Numer. Math.*, 91:605–625, 2002.
[9] S. L. Campbell, I. C. F. Ipsen, C. T. Kelley, and C. D. Meyer. GMRES and the minimal polynomial. *BIT*, 36(4):664–675, 1996.
[10] A. S. Carasso. Determining surface temperatures from interior observations. *SIAM J. Appl. Math.*, 42:558–574, 1982.
[11] L. Eldén, F. Berntsson, and T. Regińska. Wavelet and Fourier methods for solving the sideways heat equation. *SIAM J. Sci. Comput.*, 21(6):2187–2205, 2000.
[12] H. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer Academic Publishers, Dordrecht, the Netherlands, 1996.
[13] G. H. Golub and C. F. Van Loan. *Matrix Computations. 3rd ed.* Johns Hopkins Press, Baltimore, MD., 1996.
[14] Martin H. Gutknecht and Thomas Schmelzer. The block grade of a block Krylov space. *Lin. Alg. Appl.*, 430:174–185, 2009.
[15] M. Hanke. On Lanczos based methods for the regularization of discrete ill-posed problems. *BIT*, 41:1008–1018, 2001.
[16] M. Hanke and J. Nagy. Restoration of atmospherically blurred images by symmetric indefinite conjugate gradient techniques. *Inverse Problems*, 12:157–173, 1996.
[17] M. Hanke, J. Nagy, and R. Plemmons. *Preconditioned Iterative Regularization for Ill-Posed Problems*, pages 141–163. Numerical Linear Algebra, ed. L. Reichel and A. Ruttan and R. S. Varga. de Gruyter, Berlin, New York, 1993.
[18] M. Hanke and J.G. Nagy. Inverse Toeplitz preconditioners for ill-posed problems. *Lin. Alg. Appl.*, 284:137–156, 1998.
[19] P. C. Hansen. The discrete Picard condition for discrete ill-posed problems. *BIT*, 30:658–672, 1990.
[20] P. C. Hansen. Regularization tools: A Matlab package for analysis and solution of discrete ill-posed problems. *Numerical Algorithms*, 6:1–35, 1994. 10.1007/BF02149761.
[21] P. C. Hansen. *Rank-Deficient and Discrete Ill-Posed Problems. Numerical Aspects of Linear Inversion*. Society for Industrial and Applied Mathematics, Philadelphia, 1997.
[22] P. C. Hansen. Regularization tools version 4.0 for Matlab 7.3. *Numer. Algor.*, 46:189–194, 2007.
[23] P. C. Hansen and T. Koldborg Jensen. Smoothing-norm preconditioning for regularizing minimum-residual methods. *SIAM Journal on Matrix Analysis and Applications*, 29(1):1–14, 2006.
[24] K. Hayami and M. Sugihara. A geometric view of Krylov subspace methods on singular systems. *Num. Lin. Alg. Appl.*, 18:449–469, 2011.
[25] N. J. Higham. *Accuracy and Stability of Numerical Algorithms*. SIAM, Philadelphia, PA, USA, second edition, 2002.
[26] M. Ilić and I. W. Turner. Krylov subspaces and the analytic grade. *Num. Lin. Alg. Appl.*, 12:55–76, 2005.
[27] T.K. Jensen and P.C. Hansen. Iterative regularization with minimum-residual methods. *BIT Numerical Mathematics*, 47:103–120, 2007. 10.1007/s10543-006-0109-5.
[28] N. M. Nachtigal, S. C. Reddy, and L. N. Trefethen. How fast are nonsymmetric matrix iterations? *SIAM J. Matrix Anal. Appl.*, 13(3):778–795, 1992.
[29] J. G. Nagy and K. M. Palmer. Steepest descent, CG, and iterative regularization of ill-posed problems. *BIT*, 43:1003–1017, 2003.
[30] J.G. Nagy, R.J. Plemmons, and T.C. Torgersen. Iterative image restoration using approximate inverse preconditioning. *Image Processing, IEEE Transactions on*, 5(7):1151 –1162, 1996.
[31] Z. Ranjbar and L. Eldén. A preconditioned GMRES method for solving a sideways parabolic equation in two space dimensions. Technical Report LiTH-MAT-R-3, Linköping University,

Scientific Computing, The Institute of Technology, 2010.
[32] Zohreh Ranjbar. *Numerical Solution of Ill-posed Cauchy Problems for Parabolic Equations*. Linköping studies in science and technology. Dissertations. no. 1300, Linköping University, 2010.
[33] H.-J. Reinhardt. A numerical method for the solution of two-dimensional inverse heat conduction problems. *International Journal for Numerical Methods in Engineering*, 32:363–383, August 1991.
[34] Y. Saad. *Iterative Methods for Sparse Linear Systems, 2nd ed.* SIAM, Philadelphia, 2003.
[35] Y. Saad and M. H. Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Stat. Comput.*, 7:856–869, 1986.
[36] G. W. Stewart and J.-G. Sun. *Matrix Perturbation Theory*. Academic Press, Boston, 1990.
[37] J. H. Wilkinson. *The Algebraic Eigenvalue Problem*. Oxford University Press, 1965.