SIAM J. MATRIX ANAL. APPL.
Vol. 29, No. 4, pp. 1297–1321

MAXIMUM ATTAINABLE ACCURACY OF INEXACT SADDLE POINT SOLVERS*

PAVEL JIRÁNEK[†] AND MIROSLAV ROZLOŽNÍK[‡]

Abstract. In this paper we study numerical behavior of several iterative Krylov subspace solvers applied to the solution of large-scale saddle point problems. Two main representatives of segregated solution approach are analyzed: the Schur complement reduction method based on the elimination of primary unknowns and the null-space projection method, which relies on a basis for the subspace described by the constraints. We show that the choice of the back-substitution formula may considerably influence the maximum attainable accuracy of approximate solutions computed in finite precision arithmetic.

Key words. saddle point problems, Schur complement reduction method, null-space projection method, rounding error analysis

AMS subject classifications. 65F10, 65F20, 65F35

DOI. 10.1137/060659727

1. Introduction.

We want to solve a saddle point system which is in fact the symmetric indefinite system with 2×2 block structure

$$\begin{pmatrix} A & B \\ B^T & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} f \\ 0 \end{pmatrix}, \tag{1.1}$$

where the diagonal $n \times n$ block A is symmetric positive definite and the $n \times m$ off-diagonal block B has full column rank. Saddle point problems have recently attracted a lot of attention and appear to be a time-critical component in the solution of large-scale problems in many applications of computational science and engineering. A large amount of work has been devoted to a wide selection of solution techniques varying from the fully direct approach, through the use of iterative stationary or Krylov subspace methods, up to the combination of direct and iterative techniques including preconditioned iterative schemes. For an excellent survey on applications, methods, and results on numerical solution of saddle point problems, we refer to [5] and numerous references therein (relevant references will be given later in the text). Significantly less attention, however, has been paid so far to the numerical stability aspects. In this paper we concentrate on the numerical behavior of schemes which compute separately the unknown vectors x and y: one of them is first obtained from a reduced system of a smaller dimension and, once it has been computed, the other unknown is obtained by back-substitution solving exactly or inexactly another reduced problem. The main representatives of such a segregated approach are the Schur

[*] Received by the editors May 12, 2006; accepted for publication (in revised form) by M. Benzi May 18, 2007; published electronically January 4, 2008.
http://www.siam.org/journals/simax/29-4/65972.html

[†] Department of Modelings of Processes, Technical University of Liberec, Hálkova 6, CZ-461 17 Liberec, Czech Republic (pavel.jiranek@tul.cz). The work of this author was supported by the MSMT CR under the project 1M0554 "Advanced Remedial Technologies."

[‡] Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 2, CZ-182 07 Prague 8, Czech Republic (miro@cs.cas.cz). The work of this author was supported by the project 1ET400300415 within the National Program of Research "Information Society" and by the Institutional Research Plan AV0Z10300504 "Computer Science for the Information Society: Models, Algorithms, Applications."

complement reduction method and the null-space projection method. In this paper we analyze such algorithms which can be interpreted as iterations for the reduced system but compute the approximate solutions x_k and y_k to both unknown vectors x and y simultaneously.

The Schur complement reduction method uses the block factorization in the form

$$\begin{pmatrix} A & B \\ B^T & 0 \end{pmatrix} = \begin{pmatrix} I & 0 \\ B^T A^{-1} & I \end{pmatrix} \begin{pmatrix} A & B \\ 0 & -B^T A^{-1} B \end{pmatrix},$$

where the matrix $-B^T A^{-1} B$ is the Schur complement of A in (1.1). Such decomposition leads to solving the resulting block triangular system

$$\begin{pmatrix} A & B \\ 0 & -B^T A^{-1} B \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} f \\ -B^T A^{-1} f \end{pmatrix}, \tag{1.2}$$

which is nothing but a block Gaussian elimination applied to the original system (1.1). The block triangular system (1.2) is solved by computing the unknown y from the symmetric positive definite Schur complement system of order m and then by computing the unknown x from a system of order n with the symmetric positive definite matrix A. This approach leads to the explicit formula for the unknown vector $x = A^{-1}(f - By)$. The system (1.1) can be seen as two block equations and we refer to them as the "first block equation in (1.1)" and the "second block equation in (1.1)." The null-space projection method is based on the projection of the first block equation in (1.1) onto the null-space $N(B^T)$ and onto its orthogonal complement $R(B)$, respectively. According to the second block equation of (1.1) the unknown x belongs to $N(B^T)$ and therefore we get the block triangular system

$$\begin{pmatrix} (I-\Pi)A(I-\Pi) & 0 \\ B^T A & B^T B \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} (I-\Pi)f \\ B^T f \end{pmatrix}, \tag{1.3}$$

where $\Pi \equiv B(B^T B)^{-1} B^T$ denotes the orthogonal projector onto $R(B)$. This triangular system is solved by back substitution, where we first compute the unknown x from the projected system of order n with the symmetric positive semidefinite matrix $(I - \Pi)A(I - \Pi)$. Once it has been computed, the unknown y is obtained as $y = B^\dagger(f - Ax)$ by solving the least squares problem

$$\|f - Ax - By\| = \min_{v \in \mathbb{R}^m} \|f - Ax - Bv\|, \tag{1.4}$$

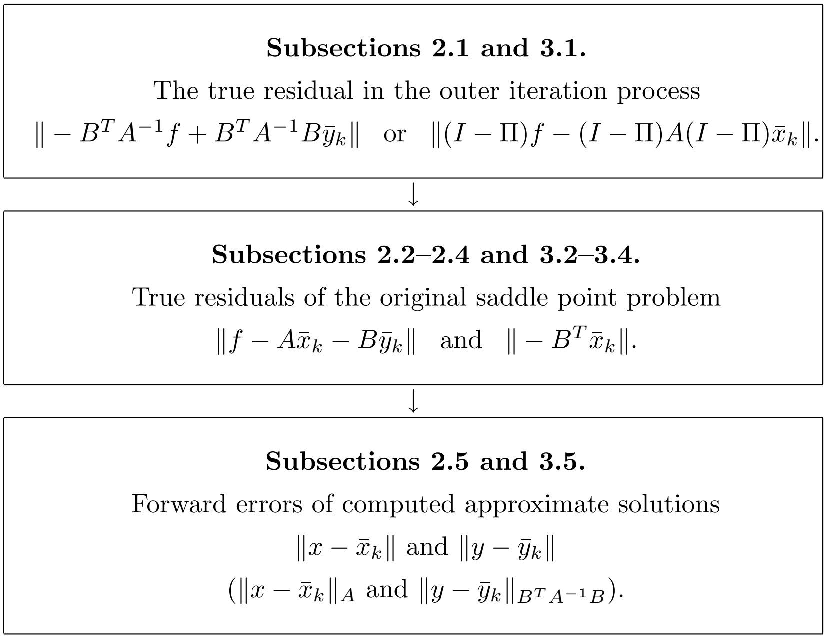
where $B^\dagger$ denotes the Moore–Penrose pseudoinverse of B. The success of algorithms for solving the block triangular system (1.2) or (1.3) depends on the availability of good approximations to the inverse of the block A or to the pseudoinverse of B, respectively. More precisely, one looks for a cheap approximate solution to the inner systems with the matrix A and/or to the associated least squares problems with the matrix B. Numerous inexact schemes have been used and analyzed (see, e.g., the analysis of inexact Uzawa algorithms [15, 11, 12, 4, 37], inexact null-space methods [28, 35, 36], multilevel or multigrid methods [10, 9, 36], domain decomposition methods [8], two-stage iterative processes [27, 16], and inner-outer iterations [19]). These works contain mainly the analysis of a convergence delay caused by the inexact solution of inner systems or least squares problems.

In this paper we concentrate on the question of what is the best accuracy we can get from inexact schemes solving either (1.2) or (1.3) when implemented in finite precision arithmetic. The fact that the inner solution tolerance strongly influences the accuracy of computed iterates is known and was studied in several contexts. The general framework for understanding inexact Krylov subspace methods has been developed in [31] and [33]. Assuming exact arithmetic, Simoncini and Szyld [31] and van den Eshof and Sleijpen [33] investigated the effect of an approximately computed matrix-vector product in every iteration on the ultimate accuracy of several solvers and explained the success of relaxation strategies for the inner accuracy tolerance from [7, 8, 18]. The developed theory strongly exploits the particular properties of an iterative method used for solving the associated system. In the context of saddle point problems, this requires a deep analysis of the outer iteration scheme for solving the reduced Schur complement or projected system (in particular, we refer to [31, section 8]).

The effects of rounding errors in the Schur complement reduction method and the null-space projection method have been studied, e.g., in [1, 2, 14, 26], where the maximum attainable accuracy of computed approximate solutions by means of residuals and errors is estimated depending on the user tolerance specified in the outer iteration. In this paper we analyze the influence of the inexact solution of inner systems/least squares problems on the same quantities. Our approach is based on a standard backward analysis which allows us to take into account both the inexactness of the inner iteration loops as well as the accompanying rounding errors that occur in finite precision arithmetic.

The theory developed for the outer iteration process is similar to the analysis of Greenbaum in [22, 21] who estimated the gap between the true and recursively updated residual for a general class of iterative methods using coupled two-term recursions. The difference here is that every computed approximate solution of an inner problem is interpreted as an exact solution of a perturbed problem induced by the actual stopping criterion, while the theory of [22] considered only the rounding errors associated with a fixed matrix-vector multiplication. In contrast to the theory of inexact Krylov methods [31, 33], the bounds for the true residual in the outer iteration loop are obtained without specifying the solver used for solving the Schur complement or the projected Hessian system. It appears that the maximum attainable accuracy level in the outer process is mainly given by the inexactness of solving the inner problems and it is not further magnified by the associated rounding errors. These results are thus similar to ones which can be obtained in exact arithmetic.

The situation is different when looking at the numerical behavior of residuals associated with the original saddle point system, which describe how accurately the two block equations in (1.1) are satisfied. It is shown that the attainable accuracy of computed approximate solutions then depends significantly on the back-substitution formula used for computing the remaining unknowns. Our results show that, independent of the fact that the inner systems are solved inexactly, some back-substitution schemes lead ultimately to residuals on the roundoff unit level. Indeed, our results confirm that, depending on which back-substitution formula is used, the computed iterates may satisfy either the first or the second block equation to the working accuracy. We believe that such results cannot be obtained using the exact arithmetic considerations and are of importance in applications requiring accurate approximations (see, e.g., [20, 17, 13]). On the other hand, we agree that in many applications the saddle point system comes from a discretization of certain partial differential equa-



Subsections 2.1 and 3.1.

The true residual in the outer iteration process

$\| - B^T A^{-1} f + B^T A^{-1} B \bar{y}_k \|$ or $\|(I - \Pi) f - (I - \Pi) A (I - \Pi) \bar{x}_k \|$.

↓

Subsections 2.2–2.4 and 3.2–3.4.

True residuals of the original saddle point problem

$\|f - A\bar{x}_k - B\bar{y}_k\|$ and $\| - B^T \bar{x}_k\|$.

↓

Subsections 2.5 and 3.5.

Forward errors of computed approximate solutions

$\|x - \bar{x}_k\|$ and $\|y - \bar{y}_k\|$

($\|x - \bar{x}_k\|_A$ and $\|y - \bar{y}_k\|_{B^T A^{-1} B}$).

Fig. 1.1.

tions and much lower accuracy is sufficient. In any case, our paper gives a theoretical explanation for the behavior which was probably observed or is already implicitly known. However, we have not found any explicit references to this issue. The implementations that we point out as optimal are actually those which are widely used and suggested in applications.

The organization of the paper is as follows. Sections 2 and 3 are devoted to the rounding error analysis of the Schur complement reduction method and the null-space projection method, respectively. Each section is divided into five subsections (see the flow-chart in Figure 1.1). In subsections 2.1 and 3.1 we analyze the influence of inexact solution of inner systems or least squares on the maximum attainable accuracy in the outer iteration process for solving (1.2) or (1.3), and we estimate the ultimate norms of the true residuals $-B^T A^{-1} f + B^T A^{-1} B \bar{y}_k$ and $(I - \Pi) f - (I - \Pi) A (I - \Pi) \bar{x}_k$. In the consequent three subsections of sections 2 and 3, we give bounds for the ultimate norm of the true residuals $f - A\bar{x}_k - B\bar{y}_k$ and $-B^T \bar{x}_k$. As we will see in subsections 2.2–2.4 and 3.2–3.4, the limiting accuracy of these residuals may significantly differ for various back-substitution formulas for computing x_k or y_k, respectively. Subsections 2.5 and 3.5 contain forward analysis with the bounds for the errors $x - \bar{x}_k$ and $y - \bar{y}_k$. Throughout this paper our theoretical results are illustrated on the model example taken from [30]: we put $n = 100$, $m = 20$, and

$$A = \mathrm{tridiag}(1, 4, 1) \in \mathbb{R}^{n \times n}, \quad B = \mathrm{rand}(n, m), \quad f = \mathrm{rand}(n, 1).$$

The spectrum of A and singular values of B lie in the interval $[2.0010, 5.9990]$ and $[2.1727, 7.1695]$, respectively. Therefore the conditioning of A or B does not play an important role in our experiments. For further discussion, we refer to subsections 2.5 and 3.5.

For distinction, we denote quantities computed in finite precision arithmetic by bars. We assume that the usual rules of a well-designed floating-point arithmetic hold and use occasionally the notation fl$(\cdot)$ for a computed result of an expression. The roundoff unit is denoted by u. In particular, for a matrix-vector multiplication the bound $\|\mathrm{fl}(Ax)-Ax\| \leq O(u)\|A\|\|x\|$ is used and $\|x\|$ denotes the 2-norm of the vector x; for a general matrix A we make use of the spectral norm $\|A\|$ and the corresponding condition number $\kappa(A) = \|A\|/\sigma_{min}(A)$, where $\sigma_{min}(A)$ is the minimal singular value of A. For a symmetric positive definite matrix A, $\|x\|_A$ denotes the A-norm of the vector x. Finally, we apply the O-notation when suitable.

2. Schur complement reduction method. In this section we will discuss algorithms which compute simultaneously approximations x_k and y_k to the unknowns x and y and ideally fulfill the first block equation in (1.1)

$$Ax_k + By_k = f. \tag{2.1}$$

Our goal here is not to survey all existing schemes based on (2.1) but to analyze the numerical behavior of three implementations which use different back-substitution formulas for computing the approximate solution x_k. More precisely, without specifying any particular method, we assume that we have computed the approximate solution y_{k+1} and the residual vector $r_{k+1}^{(y)}$ using the recursions

$$y_{k+1} = y_k + \alpha_k p_k^{(y)}, \tag{2.2}$$

$$r_{k+1}^{(y)} = r_k^{(y)} + \alpha_k B^T A^{-1} B p_k^{(y)} \tag{2.3}$$

with $r_0^{(y)} = -B^T A^{-1}(f - By_0)$. We will distinguish between the following three mathematically equivalent formulas:

$$x_{k+1} = x_k + \alpha_k(-A^{-1}Bp_k^{(y)}), \tag{2.4}$$

$$x_{k+1} = A^{-1}(f - By_{k+1}), \tag{2.5}$$

$$x_{k+1} = x_k + A^{-1}(f - Ax_k - By_{k+1}). \tag{2.6}$$

The resulting schemes are summarized in Figure 2.1. These schemes have been used and studied in the context of many applications, including various classical Uzawa algorithms, the two-level pressure correction approach, and the inner-outer iteration method for solving (1.1); see, e.g., the schemes with (2.4) in [29, 3], (2.5) in [15], or (2.6) in [11, 12, 4, 37], respectively. Because the solves with matrix A in formulas (2.4)–(2.6) are expensive, these systems are in practice solved only approximately. Our analysis is based on the assumption that every solution of a symmetric positive definite system with the matrix A is replaced by an approximate solution produced by an arbitrary method. The resulting vector is then interpreted as an exact solution of the system with the same right-hand side vector but with a perturbed matrix $A + \Delta A$. We always require that the relative norm of the perturbation is bounded as $\|\Delta A\| \leq \tau\|A\|$, where τ represents a backward error associated with the computed solution vector. We will always assume that the perturbation ΔA does not exceed the limitation given by the distance of A to the nearest singular matrix and put restriction in the form $\tau\kappa(A) \ll 1$. It follows then from the standard perturbation analysis (see, e.g., [23, 6]) that

$$\|(A + \Delta A)^{-1} - A^{-1}\| \leq \frac{\tau\kappa(A)}{1 - \tau\kappa(A)}\|A^{-1}\|.$$

```
outer iteration

$y_0$, solve $Ax_0 = f - By_0$, $r_0^{(y)} = -B^T x_0$

for $k = 0, 1, 2, \dots$
    $y_{k+1} = y_k + \alpha_k p_k^{(y)}$

    inner iteration / back-substitution

        solve $Ap_k^{(x)} = -Bp_k^{(y)}$
        (A) $x_{k+1} = x_k + \alpha_k p_k^{(x)}$
        (B) solve $Ax_{k+1} = f - By_{k+1}$
        (C) solve $Au_k = f - Ax_k - By_{k+1}$, $x_{k+1} = x_k + u_k$

    $r_{k+1}^{(y)} = r_k^{(y)} - \alpha_k B^T p_k^{(x)}$
```

FIG. 2.1. *Schur complement reduction: Three different schemes for computing the approximate solution x_{k+1} (called in the text the updated approximate solution* (A), *the approximate solution computed by a direct substitution* (B), *and the approximate solution computed by a corrected direct substitution* (C), *respectively).*

Note that if $\tau = O(u)$, then we have a backward stable method for solving the positive definite system with A. In our numerical experiments, we solve the systems with A inexactly using the conjugate gradient method or with the Cholesky factorization as indicated by the notation $\tau = O(u)$.

2.1. The attainable accuracy in the Schur complement system. In this subsection we look at the ultimate accuracy in the outer iteration process by means of the true residual $-B^T A^{-1} f + B^T A^{-1} B\bar{y}_k$. It is clear that if we perturb the Schur complement system $-B^T A^{-1} By = -B^T A^{-1} f$ to $-B^T (A + \Delta A)^{-1} B\hat{y} = -B^T A^{-1} f$, where $\|\Delta A\| \leq \tau \|A\|$, then the residual associated with $\hat{y}$ can be bounded as

$$\| - B^T A^{-1} f + B^T A^{-1} B\hat{y}\| \leq \frac{\tau\kappa(A)}{1 - \tau\kappa(A)} \|A^{-1}\| \|B\|^2 \|\hat{y}\|. \tag{2.7}$$

We see from (2.7) that there is a limitation to the accuracy of the residual obtained directly from $\hat{y}$ and its bound is proportional to τ. Note that these considerations were made assuming exact arithmetic. The effects of rounding errors on the same quantity have been studied by Greenbaum [22], who considered a general class of methods for solving the fixed system of linear equations using two-term recursions given by (2.2) and (2.3). Using a similar approach we can extend these results and formulate the following theorem.

THEOREM 2.1. *The gap between the true residual $-B^T A^{-1} f + B^T A^{-1} B\bar{y}_k$ and the updated residual $\bar{r}_k^{(y)}$ can be bounded as*

$$\|-B^T A^{-1} f + B^T A^{-1} B\bar{y}_k - \bar{r}_k^{(y)}\| \leq \frac{[(2k+1)\tau + O(u)]\kappa(A)}{1 - \tau\kappa(A)} \|A^{-1}\| \|B\| (\|f\| + \|B\| \bar{Y}_k),$$

where $\bar{Y}_k$ is defined as a maximum norm over all computed approximate solutions $\bar{Y}_k \equiv \max_{i=0,\ldots,k} \|\bar{y}_i\|$.

Proof. The initial residual $\bar{r}_0^{(y)}$ is computed as $\bar{r}_0^{(y)} = -\mathrm{fl}(B^T\bar{x}_0)$, where $(A + \Delta A_0)\bar{x}_0 = \mathrm{fl}(f - By_0)$, $\|\Delta A_0\| \leq \tau\|A\|$. It is easy to see that the statement holds for $k = 0$. The computed approximate solution $\bar{y}_{k+1}$ and the residual $\bar{r}_{k+1}^{(y)}$ satisfy

$$\text{(2.8)}\ \bar{y}_{k+1} = \bar{y}_k + \bar{\alpha}_k\bar{p}_k^{(y)} + \Delta y_{k+1},\ \|\Delta y_{k+1}\| \leq u\|\bar{y}_k\| + (2u + u^2)\|\bar{\alpha}_k\bar{p}_k^{(y)}\|,$$
$$\text{(2.9)}\ \bar{r}_{k+1}^{(y)} = \bar{r}_k^{(y)} - \bar{\alpha}_k B^T\bar{p}_k^{(x)} + \Delta r_{k+1}^{(y)},\ \|\Delta r_{k+1}^{(y)}\| \leq u\|\bar{r}_k^{(y)}\| + O(u)\|B\|\|\bar{\alpha}_k\bar{p}_k^{(x)}\|,$$

where $\bar{p}_k^{(x)}$ is the exact solution of the perturbed system

$$(A + \Delta A_k)\bar{p}_k^{(x)} = -\mathrm{fl}(B\bar{p}_k^{(y)}),\ \|\Delta A_k\| \leq \tau\|A\|. \tag{2.10}$$

Multiplying (2.8) by $B^TA^{-1}B$, substituting (2.10) into the recurrence (2.9), and subtracting these two equations we get the recurrence

$$-B^TA^{-1}f + B^TA^{-1}B\bar{y}_{k+1} - \bar{r}_{k+1}^{(y)} = -B^TA^{-1}f + B^TA^{-1}B\bar{y}_k - \bar{r}_k^{(y)}$$
$$-\bar{\alpha}_k(B^T\bar{p}_k^{(x)} + B^TA^{-1}B\bar{p}_k^{(y)}) + B^TA^{-1}B\Delta y_k - \Delta r_k^{(y)}.$$

The norm of the vector $\bar{\alpha}_k\bar{p}_k^{(y)}$ can be bounded as $\|\bar{\alpha}_k\bar{p}_k^{(y)}\| \leq \|\bar{y}_{k+1}\| + \|\bar{y}_k\| + \|\Delta y_{k+1}\|$. This bound in combination with (2.8) gives $\|\Delta y_{k+1}\| \leq O(u)\bar{Y}_{k+1}$ and $\|\bar{\alpha}_k\bar{p}_k^{(y)}\| \leq 3\bar{Y}_{k+1}$ which also implies

$$\|\bar{\alpha}_k\bar{p}_k^{(x)}\| \leq \frac{3\|A^{-1}\|}{1 - \tau\kappa(A)}\|B\|\bar{Y}_{k+1}. \tag{2.11}$$

Using (2.10), the bound on $\|\bar{\alpha}_k\bar{p}_k^{(y)}\|$, and some elementary manipulation, we can estimate the term $\bar{\alpha}_k(B^T\bar{p}_k^{(x)} + B^TA^{-1}B\bar{p}_k^{(y)})$

$$\|\bar{\alpha}_k(B^T\bar{p}_k^{(x)} + B^TA^{-1}B\bar{p}_k^{(y)})\| \leq \|\bar{\alpha}_k B^T[(A + \Delta A_k)^{-1} - A^{-1}]\mathrm{fl}(B\bar{p}_k^{(y)})\|$$
$$+\|\bar{\alpha}_k B^TA^{-1}[\mathrm{fl}(B\bar{p}_k^{(y)}) - B\bar{p}_k^{(y)}]\| \leq \frac{[\tau + O(u)]\kappa(A)}{1 - \tau\kappa(A)}\|A^{-1}\|\|B\|^2\bar{Y}_{k+1}.$$

Considering (2.9), (2.11), and the induction assumption on $-B^TA^{-1}f + B^TA^{-1}B\bar{y}_k - \bar{r}_k^{(y)}$ (similar to the one used in [22]), we obtain the bound for the error vector $\Delta r_{k+1}^{(y)}$ in the form

$$\|\Delta r_{k+1}^{(y)}\| \leq \frac{O(u)\kappa(A)}{1 - \tau\kappa(A)}\|A^{-1}\|\|B\|(\|f\| + \|B\|\bar{Y}_{k+1})$$

which proves the statement of the theorem. □

It is a well-known fact that the residual $\bar{r}_k^{(y)}$ computed recursively via (2.3) usually converges far below $O(u)$. Using this assumption we can obtain from the estimate for the gap $-B^TA^{-1}f + B^TA^{-1}B\bar{y}_k - \bar{r}_k^{(y)}$ the estimate for the maximum attainable accuracy of the true residual $-B^TA^{-1}f + B^TA^{-1}B\bar{y}_k$ itself. Summarizing, while the updated residual $\bar{r}_k^{(y)}$ converges to zero the true residual stagnates at the level proportional to τ. This is also illustrated in our numerical example, where the Schur complement system $-B^TA^{-1}By = -B^TA^{-1}f$ is solved using the steepest descent method with the initial approximation y_0 set to zero. In Figure 2.2(a) we show the relative norms of the true residual $-B^TA^{-1}f + B^TA^{-1}B\bar{y}_k$ (solid lines) and the updated residual $\bar{r}_k^{(y)}$ (dashed lines).

Similar to Greenbaum [22], we have shown that the gap between the true and updated residual is proportional to the maximum norm of approximate solutions computed during the whole iteration process. Since the Schur complement system is symmetric negative definite, the norm of the error or residual converges monotonically for the most iterative methods like the steepest descent, the conjugate gradient, the conjugate residual method, or other error/residual minimizing methods or at least becomes orders of magnitude smaller than the initial error/residual without exceeding this limit. In such cases, the quantity $\bar{Y}_k$ does not play an important role in the bound, and it can usually be replaced by $\|y_0\|$ or a small multiple of $\|y\|$. The situation is more complicated when A is nonsingular and nonsymmetric; see [24].

As we already noted, the main difference with respect to the analysis of Greenbaum is that the floating-point multiplication with the fixed A^{-1} is replaced by the step-dependent inexact solution of the system with A such that it can be interpreted as the exact application of the matrix $(A+\Delta A_k)^{-1}$, where the perturbation matrix ΔA_k changes at every step k. This concept is very similar to the notion of inexact Krylov subspace methods (see [31] or [33]), which, on the other hand, do not take into account the effects of rounding errors. The theory of Greenbaum [22] could be directly applied if we only have at each iteration $\|\mathrm{fl}(B^TA^{-1}Bx) - B^TA^{-1}Bx\| \leq O(u)\|A^{-1}\|\|B\|^2\|x\|$. Since in our idealized case $\mathrm{fl}(B^TA^{-1}Bx) = B^T(A+\Delta A_k)^{-1}Bx$ with $\|\Delta A_k\| \leq \tau\|A\|$, we have only

$$\|\mathrm{fl}(B^TA^{-1}Bx) - B^TA^{-1}Bx\| \leq \frac{\tau\kappa(A)}{1-\tau\kappa(A)}\|A^{-1}\|\|B\|^2\|x\|.$$

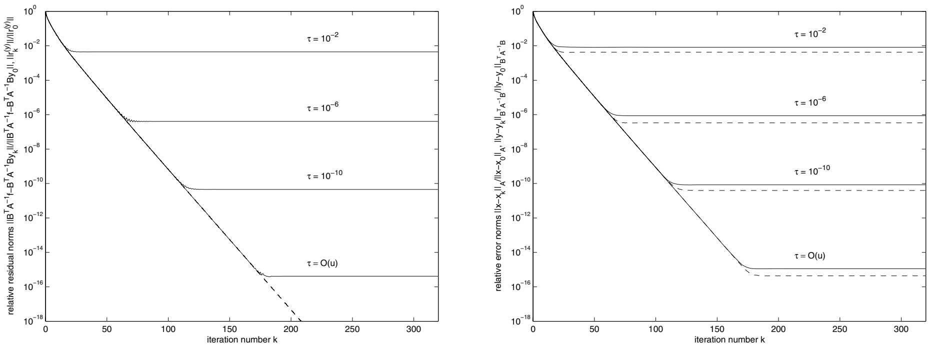
This bound could be improved if we make a restriction and use a variable tolerance for inner systems. If we require that every inner system is solved so that the relative residual of its computed solution needs the tolerance τ, then every inexact application of the matrix $B^TA^{-1}B$ would satisfy the inequality

$$\|\mathrm{fl}(B^TA^{-1}Bx) - B^TA^{-1}Bx\| \leq \tau\|A^{-1}\|\|B\|^2\|x\|. \tag{2.12}$$

Then the whole outer process (2.2) and (2.3) together with (2.12) could be interpreted as a floating-point iteration with the roundoff unit equal to τ. The computation in this "extended" arithmetic would lead to

$$\| - B^TA^{-1}f + B^TA^{-1}B\bar{y}_k - \bar{r}_k^{(y)}\| \leq \frac{O(\tau)}{1-\tau\kappa(A)}\|A^{-1}\|\|B\|^2(\|y\| + \bar{Y}_k).$$

A thorough rounding analysis of the block LU factorization has been given in [14] and further developed in the saddle point context in [26]. The approach was quite converse to the one used in our paper. It is assumed that all inner systems are solved in a backward stable way and the accuracy of computed approximate solutions is estimated in terms of the user prescribed tolerance for the outer Schur complement system. Roughly speaking, the higher tolerance η leads to the higher level of attainable accuracy of the true residuals $f - A\bar{x}_k - B\bar{y}_k$ and $-B^T\bar{x}_k$. This level is magnified by the quantities that play a similar role as the growth factor in the Gaussian elimination with partial pivoting (see, e.g., [23]). On the other hand, the parameter η giving the threshold for the backward error cannot be infinitely small. Theorem 2.1 actually



(a) (b)

FIG. 2.2. *Schur complement reduction method:* (a) *the relative norms of the true residual* $-B^TA^{-1}f + B^TA^{-1}\bar{y}_k$ *(solid lines) and the updated residual* $\bar{r}_k^{(y)}$ *(dashed lines)—the updated solution scheme* (2.4); (b) *the relative error norms* $\|x - \bar{x}_k\|_A/\|x - \bar{x}_0\|_A$ *(solid lines) and* $\|y - \bar{y}_k\|_{B^TA^{-1}B}/\|y - y_0\|_{B^TA^{-1}B}$ *(dashed lines)—the updated solution scheme* (2.4).

gives its lower bound. Dividing the right-hand side by $\|A^{-1}\|\|B\|^2\|\bar{y}\|$ we end up with $\eta \geq O(u)\kappa(A)/(1 - O(u)\kappa(A))$.

In the following we will estimate the residuals $f - A\bar{x}_k - B\bar{y}_k$ and $-B^T\bar{x}_k$. We will show that these quantities depend on the actual implementation of the back-substitution formula for x_k and distinguish between three schemes (2.4), (2.5), and (2.6). No matter how we compute the approximations $\bar{x}_k$ and $\bar{y}_k$ it holds that

$$-B^TA^{-1}f + B^TA^{-1}B\bar{y}_k = -B^T\bar{x}_k - B^TA^{-1}(f - A\bar{x}_k - B\bar{y}_k), \tag{2.13}$$

which gives the mutual relation between the residual $-B^TA^{-1}f + B^TA^{-1}B\bar{y}_k$ in the Schur complement system and the residuals $f - A\bar{x}_k - B\bar{y}_k$ and $-B^T\bar{x}_k$ associated with the saddle point system (1.1). According to Theorem 2.1, $\| - B^TA^{-1}f + B^TA^{-1}B\bar{y}_k\|$ is ultimately $O(\tau)$. Then it is clear from (2.13) that both $f - A\bar{x}_k - B\bar{y}_k$ and $-B^T\bar{x}_k$ cannot be proportional to the roundoff unit u. We will show that, depending on the chosen back-substitution scheme, we can ensure either that $f - A\bar{x}_k - B\bar{y}_k = O(\tau)$ with $-B^T\bar{x}_k = O(u)$ (scheme A (2.4)), or that $f - A\bar{x}_k - B\bar{y}_k = O(u)$ with $-B^T\bar{x}_k = O(\tau)$ (scheme C (2.6)), while the most straightforward scheme B (2.5) leads to both $f - A\bar{x}_k - B\bar{y}_k = O(\tau)$ and $-B^T\bar{x}_k = O(\tau)$.

2.2. Scheme A: The updated approximate solution.

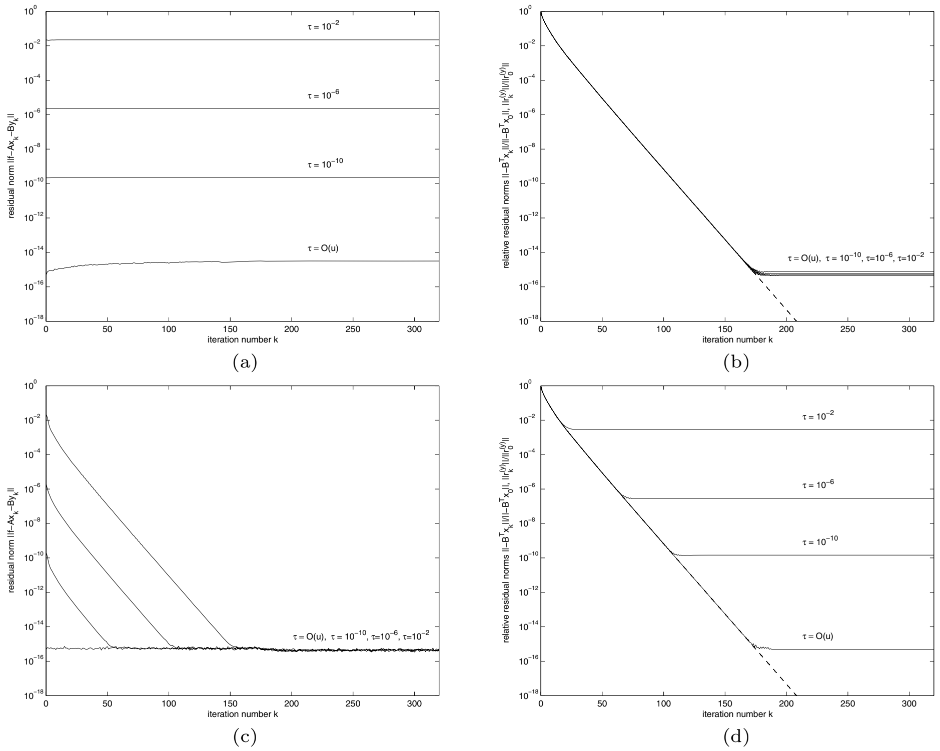
In this subsection we analyze the generic update (2.4). It is clear that this scheme requires only one system solve with A per iteration. Indeed, we compute only the direction vector $p_k^{(x)} = -A^{-1}Bp_k^{(y)}$, which appears in the recurrence $r_{k+1}^{(y)} = r_k^{(y)} - \alpha_k B^T p_k^{(x)}$ anyway. As we will see, in finite precision arithmetic this algorithm guarantees that $-B^T\bar{x}_k$ will ultimately reach $O(u)$. This happens despite the fact that the systems with the matrix block A are computed inexactly with the parameter τ frequently much larger than $O(u)$.

THEOREM 2.2. *The true residual* $f - A\bar{x}_k - B\bar{y}_k$ *satisfies the bound*

$$\|f - A\bar{x}_k - B\bar{y}_k\| \leq O(u)(\|f\| + \|B\|\bar{Y}_k) + [(k+1)\tau + O(u)]\|A\|\bar{X}_k. \tag{2.14}$$

The gap between the residuals $-B^T\bar{x}_k$ *and* $\bar{r}_k^{(y)}$ *can be estimated as*

$$\| - B^T\bar{x}_k - \bar{r}_k^{(y)}\| \leq O(u)\|A^{-1}\|\|B\|(\|f\| + \|A\|\bar{X}_k + \|B\|\bar{Y}_k),$$



(a)

(b)

(c)

(d)

FIG. 2.3. *Schur complement reduction method:* (a) *the norms of the true residual* $f - A\bar{x}_k - B\bar{y}_k$ *and* (b) *the relative norms of the true residual* $-B^T\bar{x}_k$ *(solid lines) and the recursively computed residual* $\bar{r}_k^{(y)}$ *(dashed lines)—the updated solution scheme* (2.4); (c) *the norms of the true residual* $f - A\bar{x}_k - B\bar{y}_k$*—the corrected direct substitution scheme* (2.6); (d) *the relative norms of the true residual* $-B^T\bar{x}_k$ *(solid lines) and the recursively computed residual* $\bar{r}_k^{(y)}$ *(dashed lines)—the direct substitution scheme* (2.5).

where $\bar{X}_k$ *is now defined as a maximum norm over all computed approximate solutions* $\bar{X}_k \equiv \max_{i=0,\dots,k} \|\bar{x}_i\|$.

Proof. The computed approximate solution $\bar{x}_{k+1}$ satisfies

$$\bar{x}_{k+1} = \bar{x}_k + \bar{\alpha}_k \bar{p}_k^{(x)} + \Delta x_{k+1}, \ \|\Delta x_{k+1}\| \leq u\|\bar{x}_k\| + (2u + u^2)\|\bar{\alpha}_k \bar{p}_k^{(x)}\|. \tag{2.15}$$

Substituting recurrently (2.15) and (2.8) into the residual

$$f - A\bar{x}_{k+1} - B\bar{y}_{k+1} = f - A\bar{x}_k - B\bar{y}_k - \bar{\alpha}_k(A\bar{p}_k^{(x)} + B\bar{p}_k^{(y)}) - A\Delta x_{k+1} - B\Delta y_{k+1},$$

we obtain the following bound:

$$\begin{aligned} \|f - A\bar{x}_k - B\bar{y}_k\| &\leq \|f - A\bar{x}_0 - By_0\| \\ &\quad + \sum_{i=0}^{k-1} \left(\|\bar{\alpha}_i(A\bar{p}_i^{(x)} + B\bar{p}_i^{(y)})\| + \|A\|\|\Delta x_{i+1}\| + \|B\|\|\Delta y_{i+1}\| \right). \end{aligned}$$

Here we, in fact, reformulate the main result of Greenbaum [22, Theorem 2.2] and heavily use the fact that the vectors $\bar{p}_k^{(x)}$ satisfy the perturbed system (2.10). From Theorem 2.1 we have bounds $\|\Delta y_{k+1}\| \leq O(u)\bar{Y}_{k+1}$ and $\|\bar{\alpha}_k \bar{p}_k^{(y)}\| \leq 3\bar{Y}_{k+1}$ which also imply the bound (2.11). Using all of these results we get

$$\|\bar{\alpha}_k(A\bar{p}_k^{(x)} + B\bar{p}_k^{(y)})\| \leq \|\bar{\alpha}_k[\mathrm{fl}(B\bar{p}_k^{(y)}) - B\bar{p}_k^{(y)}]\| + \|\Delta A_k\|\|\bar{\alpha}_k\bar{p}_k^{(x)}\|.$$

Further we use $\|\Delta x_{k+1}\| \leq O(u)\bar{X}_{k+1}$ and $\|\bar{\alpha}_k\bar{p}_k^{(x)}\| \leq 3\bar{X}_{k+1}$. Summarizing, we get the first result. The gap between $-B^T\bar{x}_{k+1}$ and $\bar{r}_{k+1}^{(y)}$ is equal to

$$-B^T\bar{x}_{k+1} - \bar{r}_{k+1}^{(y)} = -B^T\bar{x}_k - \bar{r}_k^{(y)} - B^T\Delta x_{k+1} - \Delta r_{k+1}^{(y)}$$

and it leads to the expansion containing just the local errors Δx_{i+1}, Δy_{i+1} and the initial gap $-B^T\bar{x}_0 - \bar{r}_0^{(y)}$

$$-B^T\bar{x}_k - \bar{r}_k^{(y)} = -B^T\bar{x}_0 - \bar{r}_0^{(y)} - \sum_{i=0}^{k-1} B^T\Delta x_{i+1} - \sum_{i=0}^{k-1} \Delta r_{k+1}^{(y)}.$$

Taking norms, considering the bounds on $\|\Delta x_{k+1}\|$, $\|\Delta y_{k+1}\|$, (2.9), and the relation $\bar{r}_0^{(y)} = -\mathrm{fl}(B^T\bar{x}_0)$, we get the second result. □

As we will see in the next subsection, the bound for the gap $-B^T\bar{x}_k - \bar{r}_k^{(y)}$ is considerably better than for the scheme (2.5). In contrast to (2.18), it does not depend on τ. Provided that $\bar{r}_k^{(y)}$ converges to zero, the true residual $-B^T\bar{x}_k$ will stagnate at the level proportional to u and the second block equation in (1.1) will be satisfied to working accuracy.

Figures 2.3(a), (b) show the norms of the true residual $f - A\bar{x}_k - B\bar{y}_k$ and $-B^T\bar{x}_k$ (solid lines), respectively, including the norms of the updated residual $\bar{r}_k^{(y)}$ (dashed lines). The numerical results are in good agreement with Theorem 2.2. The residual $f - A\bar{x}_k - B\bar{y}_k$ is growing slightly due to the accumulation of errors in inner systems $Ap_k^{(x)} = -Bp_k^{(y)}$ but it essentially remains on the level proportional to τ. The residual $-B^T\bar{x}_k$ ultimately stagnates at $O(u)$. The formula (2.4) is suitable whenever the second block equation in (1.1) must be satisfied accurately, no matter how small or big the inner tolerance τ is.

2.3. Scheme B: The approximate solution computed by a direct substitution.

In this subsection we assume that x_k is computed by the direct substitution (2.5). The computed $\bar{x}_k$ then satisfies the equality

$$(A + \Delta A_k)\bar{x}_k = \mathrm{fl}(f - B\bar{y}_k),\ \|\Delta A_k\| \leq \tau\|A\|. \tag{2.16}$$

The perturbation matrices ΔA_k are different from those defined in subsection 2.1, but for simplicity we will keep the same notation. In the following we will show that the residual $\bar{r}_k^{(y)}$ is a good approximation for the residual $-B^T\bar{x}_k$, provided that they are above the level given by the bound for $-B^T\bar{x}_k - \bar{r}_k^{(y)}$. This quantity is now, however, proportional to τ.

THEOREM 2.3. *The true residual $f - A\bar{x}_k - B\bar{y}_k$ satisfies the bound*

$$\|f - A\bar{x}_k - B\bar{y}_k\| \leq O(u)(\|f\| + \|B\|\|\bar{y}_k\|) + \tau\|A\|\|\bar{x}_k\|. \tag{2.17}$$

The gap between the residuals $-B^T\bar{x}_k$ and $\bar{r}_k^{(y)}$ can be bounded as follows:

$$\begin{aligned}\| - B^T\bar{x}_k - \bar{r}_k^{(y)}\| &\leq O(u)\|A^{-1}\|\|B\|(\|f\| + \|B\|\bar{Y}_k)\\ &\quad + [(k+3)\tau + O(u)]\kappa(A)\|B\|\bar{X}_k,\end{aligned} \tag{2.18}$$

where $\bar{X}_k$ is defined as $\bar{X}_k \equiv \max_{i=0,\dots,k-1}\{\|\bar{x}_0\|, \|\bar{x}_k\|, \|\bar{\alpha}_i\bar{p}_i^{(x)}\|\}$.

Proof. The first result follows from (2.16) and the relation for the true residual

$$f - A\bar{x}_k - B\bar{y}_k = f - B\bar{y}_k - \mathrm{fl}(f - B\bar{y}_k) - \Delta A_k\bar{x}_k.$$

For the gap between $-B^T\bar{x}_k$ and $\bar{r}_k^{(y)}$ we have the identity

$$\begin{aligned} -B^T\bar{x}_k - \bar{r}_k^{(y)} = {} & -B^TA^{-1}f + B^TA^{-1}B\bar{y}_k - \bar{r}_k^{(y)} + B^TA^{-1}\Delta A_k\bar{x}_k \\ & + B^TA^{-1}[\mathrm{fl}(f - B\bar{y}_k) - (f - B\bar{y}_k)]. \end{aligned} \tag{2.19}$$

The statement of Theorem 2.1 together with (2.19) gives the second result (2.18). □

Indeed, while the residual $\bar{r}_k^{(y)}$ converges ultimately below $O(u)$, the residual $-B^T\bar{x}_k$ will remain proportional to τ. The norm of $f - A\bar{x}_k - B\bar{y}_k$ is unconditionally bounded by the term proportional to τ dominating other terms in (2.17).

Figure 2.3(d) shows the norms of $-B^T\bar{x}_k$ (solid lines) and $\bar{r}_k^{(y)}$ (dashed lines). The residual $f - A\bar{x}_k - B\bar{y}_k$ behaves similarly to that of the scheme (2.4) shown in plot (a). The residual $f - A\bar{x}_k - B\bar{y}_k$ remains almost constant since it is nothing but the residual of the system $Ax_k = f - By_k$ solved in each iteration with the uniform accuracy.

2.4. Scheme C: The approximate solution computed with a corrected direct substitution. The third back-substitution formula (2.6) can be derived by a correction of the scheme (2.5) and requires two system solves with A. In this subsection we show that its numerical behavior is very similar to the behavior of classical nonstationary iterative methods described and analyzed by Higham [23]. We prove that under certain conditions the true residual $f - A\bar{x}_k - B\bar{y}_k$ ultimately converges to the level proportional to u, which is significantly smaller than those residuals for the previous two schemes.

THEOREM 2.4. *Assume for sufficiently large k with $\|\bar{y}_{k+1} - \bar{y}_k\| \le O(u)\bar{Y}_{k+1}$ that there exists a step k_0 such that the true residual $f - A\bar{x}_k - B\bar{y}_k$ is bounded by*

$$\|f - A\bar{x}_k - B\bar{y}_k\| \le O(u)(\|f\| + \|A\|\bar{X}_k + \|B\|\bar{Y}_k) \tag{2.20}$$

for all steps $k \ge k_0$. The gap between $-B^T\bar{x}_k$ and $\bar{r}_k^{(y)}$ can be estimated as follows:

$$\| - B^T\bar{x}_k - \bar{r}_k^{(y)}\| \le O(u)\|A^{-1}\|\|B\|(\|f\| + \|B\|\bar{Y}_k) + [(k+3)\tau + O(u)]\kappa(A)\|B\|\bar{X}_k.$$

The quantity $\bar{X}_k$ is here defined as $\bar{X}_k \equiv \max_{i=0,\dots,k-1}\{\|\bar{x}_0\|, \|\bar{x}_k\|, \|\bar{\alpha}_i\bar{p}_i^{(x)}\|\}$.

Proof. The computed approximate solution $\bar{x}_{k+1}$ satisfies

$$\bar{x}_{k+1} = \bar{x}_k + \bar{u}_k + \Delta x_{k+1}, \ \|\Delta x_{k+1}\| \le u(\|\bar{x}_k\| + \|\bar{u}_k\|), \tag{2.21}$$

where the vector $\bar{u}_k$ is the exact solution of the system

$$(A + \Delta A_{k+1})\bar{u}_k = \mathrm{fl}(f - A\bar{x}_k - B\bar{y}_{k+1}), \ \|\Delta A_{k+1}\| \le \tau\|A\|. \tag{2.22}$$

The residual $f - A\bar{x}_{k+1} - B\bar{y}_{k+1}$ can be expressed using (2.21) and (2.22) as

$$\begin{aligned} f - A\bar{x}_{k+1} - B\bar{y}_{k+1} &= \Delta A_{k+1}\bar{u}_k - A\Delta x_{k+1} \\ & \quad + \mathrm{fl}(f - A\bar{x}_k - B\bar{y}_{k+1}) - (f - A\bar{x}_k - B\bar{y}_{k+1}) \\ &= G_{k+1}(f - A\bar{x}_k - B\bar{y}_k) - G_{k+1}B(\bar{\alpha}_k\bar{p}_k^{(y)}) + h_{k+1}, \end{aligned} \tag{2.23}$$

where the matrix G_{k+1} and the vector h_{k+1} are defined as $G_{k+1} \equiv \Delta A_{k+1}(A + \Delta A_{k+1})^{-1}$ and $h_{k+1} \equiv (I + G_{k+1})[\mathrm{fl}(f - A\bar{x}_k - B\bar{y}_{k+1}) - (f - A\bar{x}_k - B\bar{y}_{k+1})] - A\Delta x_{k+1} - G_{k+1}B\Delta y_{k+1}$. From a recursive use of the formula (2.23) we obtain

$$f - A\bar{x}_k - B\bar{y}_k = G_k \cdots G_1(f - A\bar{x}_0 - By_0) - \sum_{i=0}^{k-1} G_k \cdots G_{i+2}(G_{i+1}B\bar{\alpha}_i\bar{p}_i^{(y)} - h_{i+1}).$$

Taking norms, using the relation $\|\bar{\alpha}_i\bar{p}_i^{(y)}\| \leq \|\bar{y}_{i+1} - \bar{y}_i\| + \|\Delta y_{i+1}\|$ and $\|\Delta A_i\| \leq \tau\|A\|$ we obtain the uniform bound $\|G_i\| \leq \tau\kappa(A)[1 - \tau\kappa(A)]^{-1} < 1$. This leads to the inequality

$$\begin{aligned}(2.24) \quad \|f - A\bar{x}_k - B\bar{y}_k\| \leq{} & \left(\frac{\tau\kappa(A)}{1 - \tau\kappa(A)}\right)^k \|f - A\bar{x}_0 - By_0\| \\ & + \sum_{i=0}^{k-1} \left(\frac{\tau\kappa(A)}{1 - \tau\kappa(A)}\right)^{k-i} \|B\|\|\bar{y}_{i+1} - \bar{y}_i\| \\ & + k \max_{i=0,\dots,k-1} \|h_{i+1}\| + k \max_{i=0,\dots,k-1} \|B\|\|\Delta y_{i+1}\|.\end{aligned}$$

For the vector h_{k+1} it further follows that

$$\|h_{k+1}\| \leq O(u)[\|f\| + \|A\|(\|\bar{x}_{k+1}\| + \|\bar{x}_k\|) + \|B\|\bar{Y}_{k+1}].$$

It is easy to see that for sufficiently large k the first term on the right-hand side of (2.24) will decrease far below $O(u)$, while the second term will be at most $O(u)\|B\|\bar{Y}_{k+1}$ for all steps k starting from some index k_0. Summarizing, for sufficiently large $k \geq k_0$ we have the bound

$$\|f - A\bar{x}_k - B\bar{y}_k\| \leq O(u)[\|f\| + \|A\|(\|\bar{x}_{k+1}\| + \|\bar{x}_k\|) + \|B\|\bar{Y}_k].$$

The second statement can be proved considering

$$\begin{aligned}-B^T\bar{x}_{k+1} - \bar{r}_{k+1}^{(y)} ={} & -B^TA^{-1}f + B^TA^{-1}B\bar{y}_{k+1} - \bar{r}_{k+1}^{(y)} \\ & - B^T[(A + \Delta A_{k+1})^{-1} - A^{-1}]\mathrm{fl}(f - A\bar{x}_k - B\bar{y}_{k+1}) \\ & - B^TA^{-1}[\mathrm{fl}(f - A\bar{x}_k - B\bar{y}_{k+1}) - (f - A\bar{x}_k - B\bar{y}_{k+1})].\end{aligned}$$

The first term on the right-hand side can be estimated using Theorem 2.1. Based on (2.22) we have

$$\|[(A + \Delta A_{k+1})^{-1} - A^{-1}]\mathrm{fl}(f - A\bar{x}_k - B\bar{y}_{k+1})\| \leq \frac{\tau\kappa(A)}{1 - \tau\kappa(A)}\|\bar{u}_k\|,$$

which together with the bound on $\|\bar{u}_k\|$ completes the proof. □

In Theorem 2.4 we assume that $\bar{y}_k$ ultimately stagnates so that $\|\bar{y}_{k+1} - \bar{y}_k\| \leq O(u)\bar{Y}_{k+1}$ for sufficiently large $k \geq k_0$. It appears that this condition does not represent a serious restriction. Using (2.8) we have $\|\bar{y}_{k+1} - \bar{y}_k\| \leq \|\bar{\alpha}_k\bar{p}_k^{(y)}\| + O(u)\bar{Y}_{k+1}$. We will show that the norm of $\bar{\alpha}_k\bar{p}_k^{(y)}$ is much smaller than u for large k, i.e., we can absorb it into the term $O(u)\bar{Y}_{k+1}$. Denoting $\hat{S}_k \equiv B^T(A + \Delta A_k)^{-1}B$, using (2.9) and (2.10) we have the bound

$$\|\bar{\alpha}_k\bar{p}_k^{(y)}\| \leq 2\|\hat{S}_k^{-1}\|(\|\bar{r}_{k+1}^{(y)}\| + \|\bar{r}_k^{(y)}\|) + O(u)\|\hat{S}_k^{-1}\|\|(A + \Delta A_k)^{-1}\|\|B\|^2\|\bar{\alpha}_k\bar{p}_k^{(y)}\|.$$

Provided that $O(u)\|\hat{S}_k^{-1}\|\|(A+\Delta A_k)^{-1}\|\|B\|^2 < 1$, we obtain

$$\|\bar{\alpha}_k \bar{p}_k^{(y)}\| \leq \frac{2\|\hat{S}_k^{-1}\|(\|\bar{r}_{k+1}^{(y)}\| + \|\bar{r}_k^{(y)}\|)}{1 - O(u)\|\hat{S}_k^{-1}\|\|(A+\Delta A_k)^{-1}\|\|B\|^2}.$$

Since the norms of updated residuals decrease far below the roundoff unit, the assumption on $\|\bar{y}_{k+1} - \bar{y}_k\|$ will be true for sufficiently large k. Note that $O(u)\|\hat{S}_k^{-1}\|\|(A + \Delta A_k)^{-1}\|\|B\|^2 < 1$ is nothing but the restricted assumption of numerical nonsingularity of the Schur complement matrix $B^T A^{-1} B$.

The bound (2.20) is significantly better than its counterparts (2.14) and (2.17). Theorem 2.4 describes that the residual $f - A\bar{x}_k - B\bar{y}_k$ will ultimately reach the roundoff unit level provided that the matrix $G_k G_{k-1} \cdots G_1$ converges to zero for $k \to \infty$. As soon as iterates $\bar{y}_k$ start to stagnate at their limiting accuracy level, the rate of convergence of this nonstationary iteration process is bounded by the factor $\tau\kappa(A)[1-\tau\kappa(A)]^{-1}$. The behavior of $-B^T\bar{x}_k$ is similar to that of scheme (2.5). Indeed, when $\bar{r}_k^{(y)}$ converges ultimately below $O(u)$, the residual $-B^T\bar{x}_k$ remains proportional to τ. Figure 2.3(c) shows the norms of the residual $f - A\bar{x}_k - B\bar{y}_k$. The plot for $-B^T\bar{x}_k$ (not reported) is similar to the plot (d) for the scheme (2.5). It is clear that in our well-conditioned case the stationary method converges very fast and the rate of decrease of $f - A\bar{x}_k - B\bar{y}_k$ is essentially comparable to the convergence rate of the outer iteration.

2.5. Forward error analysis. In this subsection we estimate the maximum attainable accuracy in terms of the errors $x - \bar{x}_k$ and $y - \bar{y}_k$. First we formulate the bounds in the 2-norm, then in the A-norm of the error $x - \bar{x}_k$, and then in the $B^T A^{-1} B$-norm of the error $y - \bar{y}_k$. The errors $x - \bar{x}_k$ and $y - \bar{y}_k$, and the residuals $f - A\bar{x}_k - B\bar{y}_k$ and $-B^T\bar{x}_k$, satisfy

$$\begin{pmatrix} A & B \\ B^T & 0 \end{pmatrix} \begin{pmatrix} x - \bar{x}_k \\ y - \bar{y}_k \end{pmatrix} = \begin{pmatrix} f - A\bar{x}_k - B\bar{y}_k \\ -B^T\bar{x}_k \end{pmatrix}. \tag{2.25}$$

We have the explicit expression for the inverse of the saddle point matrix

$$\begin{pmatrix} A & B \\ B^T & 0 \end{pmatrix}^{-1} = \begin{pmatrix} (I - \Pi)A^{-1} & -\Pi B(B^T B)^{-1} \\ -(B^T B)^{-1} B^T \Pi^T & -(B^T A^{-1} B)^{-1} \end{pmatrix},$$

where $\Pi \equiv A^{-1}B(B^T A^{-1} B)^{-1} B^T$ represents the oblique projector onto a range of $R(B)$ along $N(B^T)$. Considering (2.25), the inequalities $\|(I - \Pi)A^{-1}\| \leq \|A^{-1}\|$, and $\|A^{-1}B(B^T A^{-1}B)^{-1}\| = \|\Pi B(B^T B)^{-1}\| \leq \|(B^T B)^{-1}\|^{1/2}$ we obtain the bounds

$$\|x - \bar{x}_k\| \leq \gamma_1 \|f - A\bar{x}_k - B\bar{y}_k\| + \gamma_2\| - B^T\bar{x}_k\|, \tag{2.26}$$

$$\|y - \bar{y}_k\| \leq \gamma_2 \|f - A\bar{x}_k - B\bar{y}_k\| + \gamma_3\| - B^T\bar{x}_k\|, \tag{2.27}$$

where $\gamma_1 \equiv \sigma_{min}^{-1}(A)$, $\gamma_2 \equiv \sigma_{min}^{-1}(B)$, and $\gamma_3 \equiv \sigma_{min}^{-1}(B^T A^{-1} B)$ are constants independent of the iteration step k. It is clear from (2.26), (2.27) and Theorems 2.2, 2.3, and 2.4 that $\|x - \bar{x}_k\|$ and $\|y - \bar{y}_k\|$ will be $O(\tau)$ for all back-substitution schemes. In contrast to our numerical example, the saddle point systems that arise in practice can be ill-conditioned. In such cases the constants γ_1, γ_2, and γ_3 may play an important role.

In exact arithmetic we have $\|x-x_k\|_A = \|y-y_k\|_{B^TA^{-1}B}$. Since in finite precision arithmetic the residual $f - A\bar{x}_k - B\bar{y}_k$ is no longer zero, instead of this identity we get

$$|\|x - \bar{x}_k\|_A - \|y - \bar{y}_k\|_{B^TA^{-1}B}| \leq \gamma_1^{1/2}\|f - A\bar{x}_k - B\bar{y}_k\|. \tag{2.28}$$

We can also formulate the proposition, which gives bounds for the errors in terms of the residuals $f - A\bar{x}_k - B\bar{y}_k$ and $-B^TA^{-1}f + B^TA^{-1}B\bar{y}_k$.

THEOREM 2.5. *The A-norm of the error $x - \bar{x}_k$ and the $B^TA^{-1}B$-norm of the error $y - \bar{y}_k$ can be bounded as*

$$\|x - \bar{x}_k\|_A \leq \gamma_1^{1/2}\|f - A\bar{x}_k - B\bar{y}_k\| + \gamma_3^{1/2}\| - B^TA^{-1}f + B^TA^{-1}B\bar{y}_k\|, \tag{2.29}$$

$$\|y - \bar{y}_k\|_{B^TA^{-1}B} \leq \gamma_3^{1/2}\| - B^TA^{-1}f + B^TA^{-1}B\bar{y}_k\|. \tag{2.30}$$

Proof. It follows from (2.28) that

$$\begin{aligned} \|x - \bar{x}_k\|_A &\leq \|y - \bar{y}_k\|_{B^TA^{-1}B} + |\|x - \bar{x}_k\|_A - \|y - \bar{y}_k\|_{B^TA^{-1}B}| \\ &\leq \|y - \bar{y}_k\|_{B^TA^{-1}B} + \sigma_{min}^{-1/2}(A)\|f - A\bar{x}_k - B\bar{y}_k\|. \end{aligned} \tag{2.31}$$

For the $B^TA^{-1}B$-norm of the error $y - \bar{y}_k$ we have

$$\|y - \bar{y}_k\|_{B^TA^{-1}B} = \|B^TA^{-1}f - B^TA^{-1}B\bar{y}_k\|_{(B^TA^{-1}B)^{-1}}, \tag{2.32}$$

which completes the proof. ◻

The first term on the right-hand side of (2.29) should be zero in exact arithmetic and it describes how well the computed $\bar{x}_k$ and $\bar{y}_k$ satisfy (2.1). The second term is related to the Schur complement residual which in exact arithmetic should converge to zero. The recursively computed residual $\bar{r}_k^{(y)}$ is a good approximation to $-B^TA^{-1}f + B^TA^{-1}B\bar{y}_k$, provided they are above the level given by Theorem 2.1. Therefore its norm represents an easily computable quantity for the second term on the right-hand side of (2.29). The residual $f - A\bar{x}_k - B\bar{y}_k$ depends on the computed $\bar{x}_k$ and we distinguish between three schemes with (2.4), (2.5), and (2.6), respectively. We can see that, no matter which implementation we use, $-B^TA^{-1}f + B^TA^{-1}B\bar{y}_k$ is a dominating quantity in (2.29). Therefore, $\|x - \bar{x}_k\|_A$ can be thus well approximated during the convergence by the quantity $\gamma_3^{1/2}\|\bar{r}_k^{(y)}\|$ or its estimate. Similar can be said also for $\|y - \bar{y}_k\|_{B^TA^{-1}B}$; see (2.30).

The errors $x - \bar{x}_k$ and $y - \bar{y}_k$ can be estimated with more sophisticated but easily computable bounds (without explicit use of residuals and conditioning). As an example we refer to the rounding error analysis of the conjugate gradient method and various mathematically equivalent formulas for estimating $\|x - \bar{x}_k\|_A$ [32]. It appears that although many existing bounds were developed using exact arithmetic considerations, they estimate successfully the energy error using computed quantities which can be orders of magnitude different from their exact precision counterparts. Therefore, despite that we assume that A^{-1} is performed inexactly, it is feasible to estimate the $B^TA^{-1}B$-norm of the error $y - \bar{y}_k$.

In Figure 2.2(b) we report the relative error norms $\|x - \bar{x}_k\|_A/\|x - \bar{x}_0\|_A$ and $\|y - \bar{y}_k\|_{B^TA^{-1}B}/\|y - y_0\|_{B^TA^{-1}B}$. The inverse of A in the computation of the $B^TA^{-1}B$-norm is computed by a direct solver. In agreement with (2.29) and (2.30) and Theorems 2.2, 2.3, and 2.4 (see also Figure 2.3), the relative A-norm of the error $x - \bar{x}_k$ and also the relative $B^TA^{-1}B$-norm of the error $y - \bar{y}_k$ begin to stagnate at the level proportional to τ. Since the behavior of these quantities for all implementations is similar, we present only the results for the scheme (2.5). The slight difference is visible only in the gap between both error norms given by the estimate (2.28).

3. Null-space projection method. In this section we deal with algorithms which compute approximations x_k and y_k such that x_k satisfies $B^T x_k = 0$ and y_k solves the least squares problem minimizing the residual $f - Ax_k - By_k$, i.e.,

$$\|f - Ax_k - By_k\| = \min_{v \in \mathbb{R}^m} \|f - Ax_k - Bv\|. \tag{3.1}$$

We will denote (3.1) by $By_k \approx f - Ax_k$ and assume that the approximate solution x_{k+1} and the residual vector $r_{k+1}^{(x)}$ are computed using

$$x_{k+1} = x_k + \alpha_k p_k^{(x)}, \tag{3.2}$$

$$r_{k+1}^{(x)} = r_k^{(x)} - \alpha_k A p_k^{(x)} - B p_k^{(y)}, \tag{3.3}$$

where $r_0^{(x)} = B^\dagger(f - Ax_0)$. The vectors x_0 and $p_k^{(x)}$ belong to $N(B^T)$ and $p_k^{(y)}$ solves the problem $Bp_k^{(y)} \approx r_k^{(x)} - \alpha_k A p_k^{(x)}$ minimizing the residual

$$\|r_k^{(x)} - \alpha_k A p_k^{(x)} - B p_k^{(y)}\| = \min_{p \in \mathbb{R}^m} \|r_k^{(x)} - \alpha_k A p_k^{(x)} - Bp\|.$$

This residual update strategy was proposed in [20] (see also [10, 9]) and is used to reduce the roundoff errors in the projection onto $N(B^T)$. Note that the vectors $p_k^{(y)}$ can be, with no additional cost, used as direction vectors for computing the approximate solution y_{k+1}. Again we will distinguish between three back-substitution formulas (the resulting schemes are described in Figure 3.1)

$$y_{k+1} = y_k + p_k^{(y)},\ p_k^{(y)} = B^\dagger(r_k^{(x)} - \alpha_k A p_k^{(x)}), \tag{3.4}$$

$$y_{k+1} = B^\dagger(f - Ax_{k+1}), \tag{3.5}$$

$$y_{k+1} = y_k + B^\dagger(f - Ax_{k+1} - By_k). \tag{3.6}$$

The pseudoinverse $B^\dagger$ in (3.4)–(3.6) is applied by solving the least squares with the matrix B. These problems are solved inexactly. In our considerations we will assume that the computed solution $\bar{v}$ of the least squares problem $Bv \approx c$ is an exact solution of a perturbed problem $(B + \Delta B)\bar{v} \approx c + \Delta c$ with $\|\Delta B\|/\|B\| \leq \tau$ and $\|\Delta c\|/\|c\| \leq \tau$. The parameter τ again represents the measure for the inexact solution of the least squares with B and actually describes the backward error. This can be achieved in many different ways considering the inner iteration loop solving the associated system of normal equations, the augmented system formulation, or solving it directly. Similar inexact schemes have been considered for solving quadratic programming problems [1, 2], multigrid methods [9, 10], or constraint preconditioners [25, 30, 28]. We assume $\tau\kappa(B) \ll 1$ which guarantees $B + \Delta B$ to have a full column rank. This allows the use of the perturbation theory (see [34] or [23, Lemma 19.8]), in particular the inequalities

$$\|(B + \Delta B)^\dagger\| \leq \frac{\|B^\dagger\|}{1 - \tau\kappa(B)},\ \|BB^\dagger - B(B + \Delta B)^\dagger\| \leq \frac{2\tau\kappa(B)}{1 - \tau\kappa(B)}.$$

Note that if $\tau = O(u)$, then we have a backward stable method for solving the least squares problem with B. In our experiments we applied the conjugate gradient least squares (CGLS) method [6] with the stopping criterion based on the corresponding backward error. Notation $\tau = O(u)$ stands for the Householder QR factorization.

```
outer iteration

$x_0$, solve $By_0 \approx f - Ax_0$, $r_0^{(x)} = f - Ax_0 - By_0$

for $k = 0, 1, 2, \dots$
    $x_{k+1} = x_k + \alpha_k p_k^{(x)}$

    inner iteration / back-substitution

        solve $Bp_k^{(y)} \approx r_k^{(x)} - \alpha_k A p_k^{(x)}$
        (A) $y_{k+1} = y_k + p_k^{(y)}$
        (B) solve $By_{k+1} \approx f - Ax_{k+1}$
        (C) solve $Bq_k \approx f - Ax_{k+1} - By_k$, $y_{k+1} = y_k + q_k$

    $r_{k+1}^{(x)} = r_k^{(x)} - \alpha_k A p_k^{(x)} - Bp_k^{(y)}$
```

FIG. 3.1. *Null-space projection method: Three different schemes for computing the approximate solution y_{k+1} (called in the text the updated approximate solution* (A), *the approximate solution computed by a direct substitution* (B), *and the approximate solution computed by a corrected direct substitution* (C), *respectively).*

3.1. The attainable accuracy in the projected system. In this subsection we look at the accuracy in the outer iteration for solving the projected system $(I - \Pi)A(I - \Pi)x = (I - \Pi)f$. We can consider the perturbed system

$$(I - \hat{\Pi})A(I - \hat{\Pi})\hat{x} = (I - \hat{\Pi})f, \tag{3.7}$$

where $\hat{\Pi} = (B + \Delta B)(B + \Delta B)^{\dagger}$ such that $\|\Delta B\| \leq \tau\|B\|$. The residual associated with the solution of (3.7) can be written as

$$(I - \Pi)f - (I - \Pi)A(I - \Pi)\hat{x} = (\hat{\Pi} - \Pi)f + (I - \hat{\Pi})A(\Pi - \hat{\Pi})\hat{x} + (\Pi - \hat{\Pi})A(I - \Pi)\hat{x}$$

and due to $\|\hat{\Pi} - \Pi\| \leq \|\Delta B\| \min\{\|B^{\dagger}\|, \|(B + \Delta B)^{\dagger}\|\}$ [23, Lemma 19.8], we have

$$\|(I - \Pi)f - (I - \Pi)A(I - \Pi)\hat{x}\| \leq \frac{2\tau\kappa(B)}{1 - \tau\kappa(B)}(\|f\| + \|A\|\|\hat{x}\|).$$

Indeed, even if we assume exact arithmetic, the residual obtained directly from $\hat{x}$ is proportional to the parameter τ. In addition, we ideally have $(B + \Delta B)^T\hat{x} = 0$ which implies $\| - B^T\hat{x}\| \leq \tau\|B\|\|\hat{x}\|$. Therefore we can expect that also the residual $-B^T\bar{x}_k$ associated with the computed approximate solution $\bar{x}_k$ will be proportional to τ. Such analysis is dependent on the choice of a particular method with the recurrences (3.2) and (3.3), and therefore we do not give it here. In accordance with [22] it seems reasonable that the bound for $-B^T\bar{x}_k$ is proportional to the factor $\bar{X}_k$. Moreover, the error in the projection of an arbitrary vector is represented in the bounds by $\tau\kappa(B)/[1 - \tau\kappa(B)]$. Therefore $-B^T\bar{x}_k$ and $\Pi\bar{x}_k$ can be expected to have the form

$$\| - B^T\bar{x}_k\| \leq \frac{O(\tau)\|B\|}{1 - \tau\kappa(B)}\bar{X}_k, \quad \|\Pi\bar{x}_k\| \leq \frac{O(\tau)\kappa(B)}{1 - \tau\kappa(B)}\bar{X}_k. \tag{3.8}$$

Theorem 3.1 shows that the true residual $(I-\Pi)f-(I-\Pi)A(I-\Pi)\bar{x}_k$ is ultimately proportional to τ, while its projection onto $N(B^T)$ will finally reach the level $O(u)$ provided that the updated residual $\bar{r}_k^{(x)}$ converges far below that level.

THEOREM 3.1. *The gap between the true residual* $(I-\Pi)f-(I-\Pi)A(I-\Pi)\bar{x}_k$ *and the projection of the updated residual* $(I-\Pi)\bar{r}_k^{(x)}$ *can be bounded by*

$$\|(I-\Pi)f-(I-\Pi)A(I-\Pi)\bar{x}_k-(I-\Pi)\bar{r}_k^{(x)}\| \le \frac{O(\tau)\kappa(B)}{1-\tau\kappa(B)}(\|f\|+\|A\|\bar{X}_k),$$

where $\bar{X}_k \equiv \max_{i=0,\ldots,k}\|\bar{x}_i\|$.

Proof. The computed approximation $\bar{x}_{k+1}$ satisfies the relations

$$\bar{x}_{k+1}=\bar{x}_k+\bar{\alpha}_k\bar{p}_k^{(x)}+\Delta x_{k+1},\ \|\Delta x_{k+1}\|\le u\|\bar{x}_k\|+(2u+u^2)\|\bar{\alpha}_k\bar{p}_k^{(x)}\|. \tag{3.9}$$

The inequality $\|\bar{\alpha}_k\bar{p}_k^{(x)}\|\le\|\bar{x}_{k+1}\|+\|\bar{x}_k\|+\|\Delta x_{k+1}\|$ gives $\|\bar{\alpha}_k\bar{p}_k^{(x)}\|\le 3\bar{X}_{k+1}$ and $\|\Delta x_{k+1}\|\le O(u)\bar{X}_{k+1}$. The vectors $\bar{y}_0$ and $\bar{p}_k^{(y)}$ satisfy $(B+\Delta B_0)\bar{y}_0\approx \mathrm{fl}(f-Ax_0)+\Delta c_0$ with $\|\Delta B_0\|\le\tau\|B\|$, $\|\Delta c_0\|\le\tau\|\mathrm{fl}(f-Ax_0)\|$, and

$$(B+\Delta B_k)\bar{p}_k^{(y)}\approx \mathrm{fl}(\bar{r}_k^{(x)}-\bar{\alpha}_kA\bar{p}_k^{(x)})+\Delta c_k, \tag{3.10}$$

$$\|\Delta B_k\|\le\tau\|B\|,\ \|\Delta c_k\|\le\tau\|\mathrm{fl}(\bar{r}_k^{(x)}-\bar{\alpha}_kA\bar{p}_k^{(x)})\|. \tag{3.11}$$

For updated residuals we have $\bar{r}_0^{(x)}=\mathrm{fl}(f-Ax_0-B\bar{y}_0)$ and

$$\bar{r}_{k+1}^{(x)}=\bar{r}_k^{(x)}-\bar{\alpha}_kA\bar{p}_k^{(x)}-B\bar{p}_k^{(y)}+\Delta r_{k+1}^{(x)}, \tag{3.12}$$

$$\|\Delta r_{k+1}^{(x)}\|\le O(u)(\|\bar{r}_k^{(x)}\|+\|A\|\|\bar{\alpha}_k\bar{p}_k^{(x)}\|+\|B\|\|\bar{p}_k^{(y)}\|). \tag{3.13}$$

The recursive use of (3.9) and (3.12) leads to the expression for the gap between the projections of $f-A\bar{x}_k$ and $\bar{r}_k^{(x)}$

$$(I-\Pi)(f-A\bar{x}_k-\bar{r}_k^{(x)})=(I-\Pi)(f-A\bar{x}_0-\bar{r}_0^{(x)})-\sum_{i=0}^{k-1}(I-\Pi)(A\Delta x_{i+1}+\Delta r_{i+1}^{(x)}).$$

Taking norms and corresponding bounds we get the following after some manipulation:

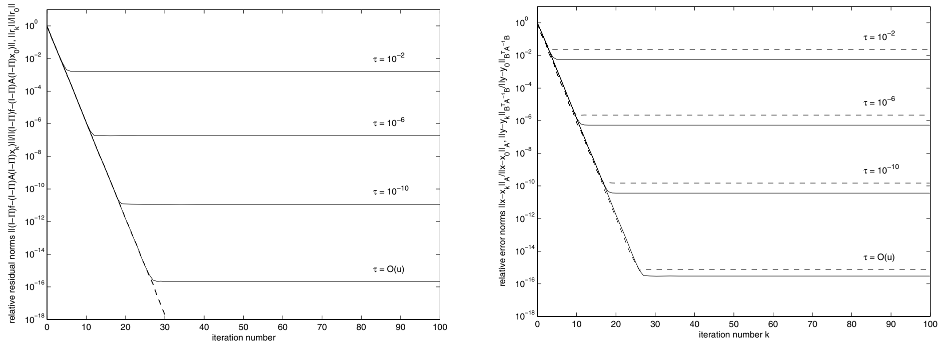
$$\|(I-\Pi)(f-A\bar{x}_k-\bar{r}_k^{(x)})\|\le\frac{O(u)\kappa(B)}{1-\tau\kappa(B)}\left(\|f\|+\|A\|\bar{X}_k\right). \tag{3.14}$$

Here we have used that $\|\bar{r}_k^{(x)}\|\le\|\bar{r}_0^{(x)}\|$ for $k=0,1,\ldots$ which seems reasonable when solving the positive semidefinite problem. For the gap between $(I-\Pi)f-(I-\Pi)A(I-\Pi)\bar{x}_k$ and $(I-\Pi)\bar{r}_k^{(x)}$, we can write

$$\|(I-\Pi)f-(I-\Pi)A(I-\Pi)\bar{x}_k-(I-\Pi)\bar{r}_k^{(x)}\|\le\|(I-\Pi)(f-A\bar{x}_k-\bar{r}_k^{(x)})\|+\|(I-\Pi)A\Pi\bar{x}_k\|.$$

Considering (3.14) and (3.8) we can conclude the proof. □

In Figure 3.2(a) we report the relative norms of the true residual $(I-\Pi)f-(I-\Pi)A(I-\Pi)\bar{x}_k$ (solid lines) and the updated residual $\bar{r}_k^{(x)}$ (dashed lines). The numerical results confirm that the residual $f-A\bar{x}_k$ is within $N(B^T)$ approximated by $\bar{r}_k^{(x)}$ to the working precision u. However, this is not true for the residual $(I-\Pi)f-(I-\Pi)A(I-\Pi)\bar{x}_k$ which is ultimately $O(\tau)$ as it follows from Theorem 3.1.



(a)

(b)

FIG. 3.2. *Null-space projection method:* (a) *the relative norms of the true residual* $(I-\Pi)f - (I-\Pi)A(I-\Pi)\bar{x}_k$ *of the projected system (solid lines) and the updated residual* $\bar{r}_k^{(x)}$ *(dashed lines)—the updated solution scheme* (3.4); *the relative norms of the errors* $\|x-\bar{x}_k\|_A/\|x-x_0\|_A$ *(solid lines) and* $\|y-\bar{y}_k\|_{B^TA^{-1}B}/\|y-\bar{y}_0\|_{B^TA^{-1}B}$ *(dashed lines)—the updated solution scheme* (3.4).

The residual $-B^T\bar{x}_k$ obviously does not depend on the back-substitution scheme; see Figure 3.3(d).

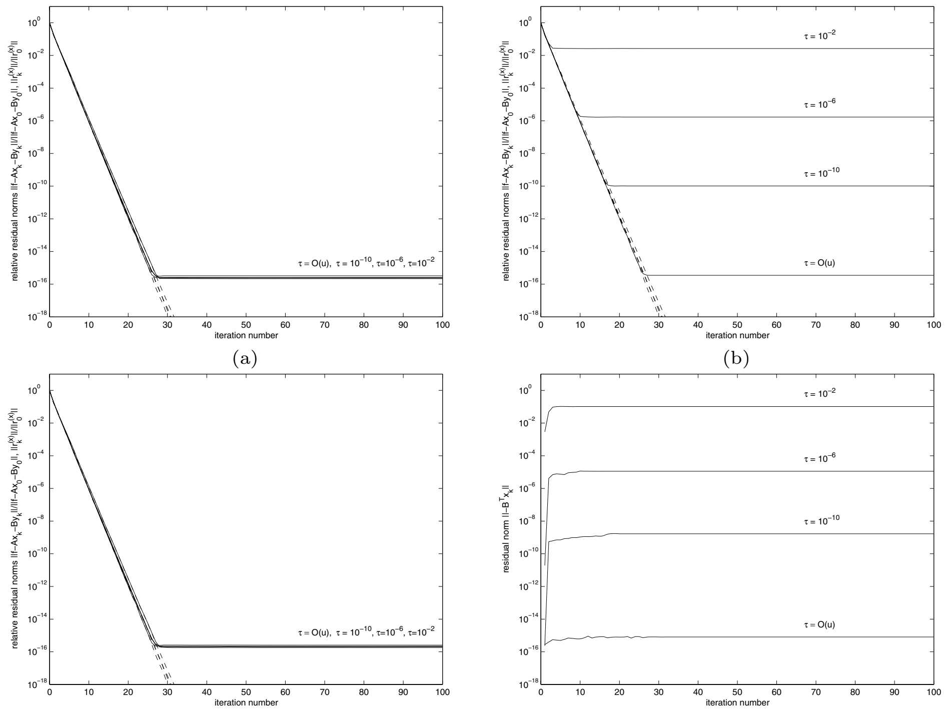
In contrast to the Schur complement reduction method, the inexactness is connected with the matrix B instead of A. In practice, the sequential application of the matrix $(I-\Pi)A(I-\Pi)$ does not represent a symmetric operator. This is also reflected in the fact that we assume a general framework for computing the vector x_k and analyze another projection of residuals $f - A\bar{x}_k - B\bar{y}_k$ and $\bar{r}_k^{(x)}$. Ideally at every iteration step we apply the matrix-vector product with the matrix $(I-\hat{\Pi})A(I-\hat{\Pi})$, where $\hat{\Pi}$ represents the orthogonal projector $\hat{\Pi} = (B+\Delta B)(B+\Delta B)^\dagger$ with $\|\Delta B\| \le \tau\|B\|$. A question similar to one in subsection 2.1 arises as to whether we can apply the results of [22] directly to the system $(I-\hat{\Pi})A(I-\hat{\Pi})\hat{x} = (I-\hat{\Pi})f$. Theorem 3.1 shows that in finite precision arithmetic the residual $(I-\Pi)f-(I-\Pi)A(I-\Pi)\bar{x}_k$ will remain proportional to the parameter τ. The theory of Greenbaum can be directly applied only if the multiplication by $(I-\Pi)A(I-\Pi)$ satisfies $\|\mathrm{fl}[(I-\Pi)A(I-\Pi)x]-(I-\Pi)A(I-\Pi)x\| \le O(u)\|(I-\Pi)A(I-\Pi)\|\|x\|$ which is obviously not the case here. In the idealized case we have $\mathrm{fl}[(I-\Pi)A(I-\Pi)x] = (I-\hat{\Pi})A(I-\hat{\Pi})x$ and hence

$$\|\mathrm{fl}[(I-\Pi)A(I-\Pi)x] - (I-\Pi)A(I-\Pi)x\| \le \frac{O(\tau)\kappa(B)}{1-\tau\kappa(B)}\|A\|\|x\|.$$

If we could improve this bound to satisfy $\|\mathrm{fl}[(I-\Pi)A(I-\Pi)x]-(I-\Pi)A(I-\Pi)x\| \le \tau\|A\|\|x\|$, the outer iteration process could be viewed as an iteration in finite precision arithmetic with the roundoff unit equal to τ and the theory of Greenbaum would lead to the estimate

$$\|(I-\Pi)f - (I-\Pi)A(I-\Pi)\bar{x}_k - \bar{r}_k^{(x)}\| \le \frac{O(\tau)}{1-\tau\kappa(B)}\|A\|(\|x\| + \bar{X}_k).$$

The numerical behavior of the null-space projection method was studied also in [1, 2], where the inner least squares are solved by the QR or LU factorization with $\tau = O(u)$ and the projected system is solved inexactly with the parameter η. Our Theorem 3.1 thus gives an answer to the question of how small the parameter η can be in the outer iteration. Roughly speaking, when using the error or residual minimizing



(a) (b)

(c) (d)

FIG. 3.3. *Null-space projection method: The relative norms of the true residual $f - A\bar{x}_k - B\bar{y}_k$ and the updated residual $\bar{r}_k^{(x)}$ (plots* (a), (b), *and* (c) *for the updated solution scheme* (3.4), *the direct substitution scheme* (3.5), *and the corrected direct substitution scheme* (3.6), *respectively);* (d) *the norms of the residual $-B^T\bar{x}_k$—the updated solution scheme* (3.4).

method for solving the projected Hessian system the backward error associated with the iterate $\bar{x}_k$ cannot be smaller than $O(u)\kappa(B)/[1 - O(u)\kappa(B)]$.

It is clear that no matter how we compute $\bar{x}_k$ and $\bar{y}_k$ we have the following relation between $(I - \Pi)f - (I - \Pi)A(I - \Pi)\bar{x}_k$, $f - A\bar{x}_k - B\bar{y}_k$, and $-B^T\bar{x}_k$:

$$(3.15)\quad (I - \Pi)f - (I - \Pi)A(I - \Pi)\bar{x}_k = (I - \Pi)(f - A\bar{x}_k - B\bar{y}_k) + (I - \Pi)A\Pi\bar{x}_k.$$

Owing to (3.8), $\Pi\bar{x}_k$ (and thus also $-B^T\bar{x}_k$) is $O(\tau)$. From Theorem 3.1 we have that $\|(I - \Pi)f - (I - \Pi)A(I - \Pi)\bar{x}_k\|$ is ultimately $O(\tau)$. Since $(I - \Pi)(f - A\bar{x}_k) = (I - \Pi)(f - A\bar{x}_k - B\bar{y}_k)$ for any $\bar{y}_k$ it also follows from Theorem 3.1 that the projection of $f - A\bar{x}_k - B\bar{y}_k$ onto $N(B^T)$ will ultimately reach $O(u)$. It is not clear from (3.15) whether the whole residual $f - A\bar{x}_k - B\bar{y}_k$ will be ultimately $O(\tau)$ or $O(u)$. It strongly depends on the back-substitution scheme used for computing the approximate solutions y_{k+1}. The following subsections show that the residual $f - A\bar{x}_k - B\bar{y}_k$ for the schemes with (3.4) (scheme A) and with (3.6) (scheme C) will finally reach $O(u)$, while the scheme B using (3.5) leads to the accuracy that is proportional only to τ.

3.2. Scheme A: The updated approximate solution. In this subsection we analyze the generic scheme with the update (3.4). This implementation does not require any additional solution of a least squares problem with the matrix B. Indeed, the computed direction vector $p_k^{(y)}$ is used to update both the iterate y_k and the residual $\bar{r}_k^{(x)}$. As we will see, this algorithm computes the residual $f - A\bar{x}_k - B\bar{y}_k$

which will ultimately reach the level of roundoff unit u independently of the fact that the inner least squares are solved with the accuracy determined by the parameter τ.

THEOREM 3.2. *The gap between the residuals* $f - A\bar{x}_k - B\bar{y}_k$ *and* $\bar{r}_k^{(x)}$ *can be bounded as follows*:

$$\|f - A\bar{x}_k - B\bar{y}_k - \bar{r}_k^{(x)}\| \leq O(u)(\|f\| + \|A\|\bar{X}_k + \|B\|\bar{Y}_k),$$

where $\bar{Y}_k \equiv \max_{i=0,\dots,k} \|\bar{y}_i\|$. *The statement of the theorem remains true if we replace* $\bar{Y}_k$ *by* $\max\{\|y_0\|, \|p_i^{(y)}\|,\ i = 0, 1, \dots, k-1\}$.

Proof. The vector $\bar{x}_{k+1}$ satisfies (3.9) with $\|\Delta x_{k+1}\| \leq O(u)\bar{X}_{k+1}$, and similarly for $\bar{y}_{k+1}$ we have

$$\bar{y}_{k+1} = \bar{y}_k + \bar{p}_k^{(y)} + \Delta y_{k+1},\ \|\Delta y_{k+1}\| \leq u\|\bar{y}_k\| + (2u + u^2)\|\bar{p}_k^{(y)}\|$$

with $\|\Delta y_{k+1}\| \leq O(u)\bar{Y}_{k+1}$. The residual $\bar{r}_{k+1}^{(x)}$ satisfies (3.12) and thus $\|\Delta r_{k+1}^{(x)}\| \leq O(u)(\|\bar{r}_k^{(x)}\| + \|A\|\bar{X}_{k+1} + \|B\|\bar{Y}_{k+1})$. Using the above relations we obtain the recursive formula

$$f - A\bar{x}_{k+1} - B\bar{y}_{k+1} - \bar{r}_{k+1}^{(x)} = f - A\bar{x}_k - B\bar{y}_k - \bar{r}_k^{(x)} - A\Delta x_{k+1} - B\Delta y_{k+1} - \Delta r_{k+1}^{(x)}.$$

Taking the norms we get the following after some manipulation:

$$\|f - A\bar{x}_k - B\bar{y}_k - \bar{r}_k^{(x)}\| \leq O(u)\left(\|f\| + \|A\|\bar{X}_k + \|B\|\bar{Y}_k + \sum_{i=0}^{k-1}\|\bar{r}_i^{(x)}\|\right).$$

The statement can now be proved by induction on k. □

We have shown that $\bar{r}_k^{(x)}$ is a good approximation to $f - A\bar{x}_k - B\bar{y}_k$ independent of the fact that $\bar{p}_k^{(y)}$ are computed inexactly. Note that Theorem 3.1 can be derived using Theorem 3.2 due to $\|(I - \Pi)(f - A\bar{x}_k - \bar{r}_k^{(x)})\| = \|(I - \Pi)(f - A\bar{x}_k - B\bar{y}_k - \bar{r}_k^{(x)})\| \leq \|f - A\bar{x}_k - B\bar{y}_k - \bar{r}_k^{(x)}\|$. In Figure 3.3(a) we show the relative norms of $f - A\bar{x}_k - B\bar{y}_k$ (solid lines) and $\bar{r}_k^{(x)}$ (dashed lines). The results of our numerical experiment are in a good agreement with Theorem 3.2.

3.3. Scheme B: The approximate solution computed by a direct substitution. In this subsection we analyze the scheme (3.5), which uses the directly computed right-hand side vector $f - Ax_k$. The computed $\bar{y}_k$ is then a solution of the perturbed problem

$$(B + \Delta B_k)\bar{y}_k \approx \mathrm{fl}(f - A\bar{x}_k) + \Delta c_k \tag{3.16}$$

with $\|\Delta B_k\| \leq \tau\|B\|$ and $\|\Delta c_k\| \leq \tau\|\mathrm{fl}(f - A\bar{x}_k)\|$. We will show that $(I - \Pi)\bar{r}_k^{(x)}$ is a good approximation of $f - A\bar{x}_k - B\bar{y}_k$ provided that both are above their level of maximum attainable accuracy.

THEOREM 3.3. *The gap between the residuals* $f - A\bar{x}_k - B\bar{y}_k$ *and* $(I - \Pi)\bar{r}_k^{(x)}$ *can be bounded by*

$$\begin{aligned}\|f - A\bar{x}_k - B\bar{y}_k - (I - \Pi)\bar{r}_k^{(x)}\| \leq{}& \tfrac{5\tau\kappa(B)}{1-\tau\kappa(B)}(\|f\| + \|A\|\|\bar{x}_k\|)\\ &+ O(u)(\|f\| + \|A\|\bar{X}_k + \|B\|\bar{Y}_k).\end{aligned}$$

Proof. Considering (3.16) it follows for the true residual that

$$\begin{aligned} f - A\bar{x}_k - B\bar{y}_k &= f - A\bar{x}_k - B(B + \Delta B_k)^\dagger[\mathrm{fl}(f - A\bar{x}_k) + \Delta c_k] \\ &= (I - \Pi)(f - A\bar{x}_k) + B[B^\dagger - (B + \Delta B_k)^\dagger]\mathrm{fl}(f - A\bar{x}_k) \\ &\quad + BB^\dagger[\mathrm{fl}(f - A\bar{x}_k) - (f - A\bar{x}_k)] - B(B + \Delta B_k)^\dagger \Delta c_k. \end{aligned}$$

Taking (3.16), the bounds on $B[B^\dagger - (B + \Delta B_k)^\dagger]$, $(B + \Delta B_k)^\dagger$, and Theorem 3.1 we get the desired result. □

When using the formula (3.5) the residual $f - A\bar{x}_k - B\bar{y}_k$ will not decrease below a level proportional to τ, while $(I - \Pi)\bar{r}_k^{(x)}$ converges beyond the level $O(u)$. This result is illustrated by our numerical experiment. In Figure 3.3(b) we plotted the relative norms of $f - A\bar{x}_k - B\bar{y}_k$ (solid lines) and $\bar{r}_k^{(x)}$ (dashed lines).

3.4. Scheme C: The approximate solution computed with a corrected direct substitution. In this subsection we analyze the scheme (3.6) requiring a solution of two least squares problems with B. We show that its behavior is similar to the algorithm using the update (3.4). We prove that under certain assumptions the true residual $f - A\bar{x}_k - B\bar{y}_k$ converges ultimately to the $O(u)$ level. The difference is that while Theorem 3.2 holds without any additional conditions, here we have a situation analogous to the behavior of nonstationary iterative methods (see [23, Chapter 16]).

THEOREM 3.4. *Provided that for sufficiently large step k the computed vector $\bar{x}_k$ stagnates, i.e., we have $\|\bar{x}_{k+1} - \bar{x}_k\| \leq O(u)\bar{X}_{k+1}$, there exists some iteration step k_0 such that*

$$\|f - A\bar{x}_k - B\bar{y}_k - (I - \Pi)\bar{r}_k^{(x)}\| \leq O(u)(\|f\| + \|A\|\bar{X}_k + \|B\|\bar{Y}_k) \tag{3.17}$$

holds for all $k \geq k_0$.

Proof. The vector $\bar{y}_{k+1}$ satisfies $\bar{y}_{k+1} = \bar{y}_k + \bar{q}_k^{(y)} + \Delta y_{k+1}$ and $\|\Delta y_{k+1}\| \leq O(u)\bar{Y}_{k+1}$, where $\bar{q}_k^{(y)}$ is the solution of the problem $(B + \Delta B_k)\bar{q}_k^{(y)} \approx \mathrm{fl}(f - A\bar{x}_{k+1} - B\bar{y}_k) + \Delta c_k$ with $\|\Delta B_k\| \leq \tau\|B\|$ and $\|\Delta c_k\| \leq \tau\|\mathrm{fl}(f - A\bar{x}_{k+1} - B\bar{y}_k)\|$. For $f - A\bar{x}_{k+1} - B\bar{y}_{k+1}$ we can then write

$$\begin{aligned} f - A\bar{x}_{k+1} - B\bar{y}_{k+1} &= (I - \Pi)(f - A\bar{x}_{k+1}) + G_k(f - A\bar{x}_{k+1} - B\bar{y}_k) \\ &\qquad - B(B + \Delta B_k)^\dagger \Delta c_k + h_k, \end{aligned}$$

where $G_k = B[B^\dagger - (B + \Delta B_k)^\dagger]$ and $h_k = -B(B + \Delta B_k)^\dagger[\mathrm{fl}(f - A\bar{x}_{k+1} - B\bar{y}_k) - (f - A\bar{x}_{k+1} - B\bar{y}_k)] - B\Delta y_{k+1}$. Projecting $f - A\bar{x}_{k+1} - B\bar{y}_{k+1}$ onto $R(B)$ and taking norms, we obtain

$$\begin{aligned} \|\Pi(f - A\bar{x}_{k+1} - B\bar{y}_{k+1})\| &\leq \left[\|G_k\| + \tau\|B(B + \Delta B_k)^\dagger\|\right] \|f - A\bar{x}_{k+1} - B\bar{y}_k\| \\ &\quad + \tau\|B(B + \Delta B_k)^\dagger\|\|\mathrm{fl}(f - A\bar{x}_{k+1} - B\bar{y}_k) - (f - A\bar{x}_{k+1} - B\bar{y}_k)\| + \|h_k\|. \end{aligned}$$

The term $\|f - A\bar{x}_{k+1} - B\bar{y}_k\|$ can be further bounded by

$$\|f - A\bar{x}_{k+1} - B\bar{y}_k\| \leq \|(I - \Pi)(f - A\bar{x}_{k+1})\| + \|\Pi(f - A\bar{x}_k - B\bar{y}_k)\| + \|A(\bar{x}_{k+1} - \bar{x}_k)\|$$

which together with the bound on $\|G_k\|$, $\|h_k\| \le O(u)(\|f\| + \|A\|\bar{X}_{k+1} + \|B\|\bar{Y}_{k+1})$, and $\tau\|B(B+\Delta B_k)^\dagger\| \le \tau\kappa(B)[1-\tau\kappa(B)]^{-1} < 1$ leads to

$$\begin{aligned}\|\Pi(f - A\bar{x}_{k+1} - B\bar{y}_{k+1})\| \\ \le \frac{3\tau\kappa(B)}{1-\tau\kappa(B)}\left[\|\Pi(f - A\bar{x}_k - B\bar{y}_k)\| + \|(I-\Pi)(f - A\bar{x}_{k+1})\| + \|A\|\|\bar{x}_{k+1} - \bar{x}_k\|\right] \\ + O(u)(\|f\| + \|A\|\bar{X}_{k+1} + \|B\|\bar{Y}_{k+1}).\end{aligned}$$

After the recursive use of the previous inequality we obtain

$$\begin{aligned}\|\Pi(f - A\bar{x}_k - B\bar{y}_k)\| \le \left(\frac{3\tau\kappa(B)}{1-\tau\kappa(B)}\right)^k \|f - A\bar{x}_0 - B\bar{y}_0\| \\ + \sum_{i=0}^{k-1}\left(\frac{3\tau\kappa(B)}{1-\tau\kappa(B)}\right)^{k-i}\left[\|(I-\Pi)(f - A\bar{x}_{i+1})\| + \|A\|\|\bar{x}_{i+1} - \bar{x}_i\|\right] \\ + O(u)(\|f\| + \|A\|\bar{X}_k + \|B\|\bar{Y}_k).\end{aligned} \tag{3.18}$$

Under the assumption on the stagnation of iterates there exist some index k_0 such that the second term on the right-hand side of (3.18) will be of order $O(u)(\|f\| + \|A\|\bar{X}_k + \|B\|\bar{Y}_k)$ for all iteration steps $k \ge k_0$. Finally, from Theorem 3.2 we have $\|(I-\Pi)(f - A\bar{x}_k) - (I-\Pi)\bar{r}_k^{(x)}\| \le O(u)(\|f\| + \|A\|\bar{X}_k + \|B\|\bar{Y}_k)$. □

Theorem 3.4 shows that $f - A\bar{x}_k - B\bar{y}_k$ will ultimately reach the $O(u)$ level. As soon as the approximate solutions $\bar{x}_k$ stagnate with $\|\bar{x}_{k+1} - \bar{x}_k\| \le O(u)\bar{X}_{k+1}$, the rate of convergence of this process is roughly given by the factor $3\tau\kappa(B)[1-\tau\kappa(B)]^{-1}$. Note that similar to subsection 2.4 the assumption on the stagnation is not restrictive. The numerical results on a model example are shown in Figure 3.3(c), which reports the relative norms of $f - A\bar{x}_k - B\bar{y}_k$ (solid lines) and $\bar{r}_k^{(x)}$ (dashed lines), and are in good agreement with Theorem 3.4.

3.5. Forward error analysis. In this subsection we look at the maximum attainable accuracy measured by errors $x - \bar{x}_k$ and $y - \bar{y}_k$. The analysis is very similar to the Schur complement reduction method and therefore we focus only on issues particular to the null-space projection method. We recall that relation (2.25) gives the universal bounds (2.26), (2.27), and (2.28). Independent of the back-substitution scheme used for computing $\bar{y}_k$, the terms $\gamma_2\| - B^T\bar{x}_k\|$ and $\gamma_3\| - B^T\bar{x}_k\|$ on the right-hand side of (2.26) and (2.27), respectively, are always proportional to τ. The terms with $f - A\bar{x}_k - B\bar{y}_k$ depend on the back-substitution formula and their final magnitude will be at most $O(\tau)$, leading to similar conclusions on errors as in subsection 2.5. The estimate for $\|x - \bar{x}_k\|_A$ is given in the following theorem.

THEOREM 3.5. *The A-norm of the error $x - \bar{x}_k$ can be bounded as*

$$\|x - \bar{x}_k\|_A \le \delta_1\| - B^T\bar{x}_k\| + \delta_2\|(I-\Pi)(f - A\bar{x}_k)\|, \tag{3.19}$$

where $\delta_1 \equiv \|A\|^{1/2}/\sigma_{min}(B)$ and $\delta_2 \equiv \sigma_{min}^{-1/2}(A)$ are constants independent of the iteration step k.

Proof. Since $(I-\Pi)A(x - \bar{x}_k) = (I-\Pi)(f - A\bar{x}_k)$, $B^Tx = 0$ and $\|B(B^TB)^{-1}\| = \sigma_{min}^{-1}(B)$, $\|x - \bar{x}_k\|_A^2$ can be written as

$$\begin{aligned}\|x - \bar{x}_k\|_A^2 &= (\Pi(x - \bar{x}_k), A(x - \bar{x}_k)) + ((I-\Pi)A(x - \bar{x}_k), x - \bar{x}_k) \\ &\le \|A^{1/2}\|\|x - \bar{x}_k\|_A(\|B(B^TB)^{-1}\|\|B^T(x - \bar{x}_k)\| + \|(I-\Pi)(f - A\bar{x}_k)\|).\end{aligned} \tag{3.20}$$

Dividing both sides by $\|x - \bar{x}_k\|_A$ gives the statement (3.19). □

The first term on the right-hand side of (3.19) should be zero in exact arithmetic. The computed $\bar{x}_k$, however, does not fulfill $-B^T\bar{x}_k = 0$ and its departure from $N(B^T)$ was discussed in (3.8). The second term converges to zero in exact arithmetic and it is related to the projected residual $(I - \Pi)(f - A\bar{x}_k)$; see Theorem 3.14. The result for $y - \bar{y}_k$ can be obtained from (3.19) using (2.28). Provided that $\bar{r}_k^{(x)}$ is larger than $O(\tau)$, $\|x - \bar{x}_k\|_A$ is then well approximated by $\delta_2\|(I - \Pi)\bar{r}_k^{(x)}\|$.

4. Conclusions.

In this paper we have looked at the numerical behavior of certain inexact saddle point solvers. In particular, for several mathematically equivalent implementations we studied the influence of inexact solving of the inner systems and estimated their maximum attainable accuracy. When considering the outer iteration process our rounding error analysis has led to results similar to ones which can be obtained assuming exact arithmetic. The situation was different when we looked at the residuals in the saddle point system. We have shown that some implementations lead ultimately to residuals on the roundoff unit level independently of the fact that the inner systems were solved inexactly on a much higher level τ. Indeed, our results confirmed that the generic and actually the cheapest implementations deliver the approximate solutions which satisfy either the second or the first block equation to the working accuracy. In addition, the schemes with the corrected direct substitution are also very attractive. We gave a theoretical explanation for the behavior which was probably observed or is already tacitly known. The implementations that we pointed out as optimal are actually those which are widely used and suggested in applications. It appears that, when measured in terms of the errors, the maximum attainable accuracy level is similar for all considered implementations and is proportional to the parameter which measures the inexactness in solving the inner systems.

Acknowledgments. The authors would like to thank the anonymous referees for their valuable comments and suggestions and Volker Mehrmann for his discussion during the MAT-TRIAD 2005 meeting, which actually stimulated this work.

REFERENCES

[1] M. Arioli, *The use of QR factorization in sparse quadratic programming and backward error issues*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 825–839.

[2] M. Arioli and L. Baldini, *A backward error analysis of a null space algorithm in sparse quadratic programming*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 425–442.

[3] J. Atanga and D. Silvester, *Iterative methods for stabilized mixed velocity-pressure finite elements*, Internat. J. Numer. Methods Fluids, 14 (1992), pp. 71–81.

[4] C. Bacuta, *A unified approach for Uzawa algorithms*, SIAM J. Numer. Anal., 44 (2006), pp. 2633–2649.

[5] M. Benzi, G. H. Golub, and J. Liesen, *Numerical solution of saddle point problems*, Acta Numer., 14 (2005), pp. 1–137.

[6] A. Björck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996.

[7] A. Bouras and V. Frayssé, *Inexact matrix-vector products in Krylov methods for solving linear systems: A relaxation strategy*, SIAM J. Matrix Anal. Appl., 26 (2005), pp. 660–678.

[8] A. Bouras, V. Frayssé, and L. Giraud, *A Relaxation Strategy for Inner-Outer Linear Solvers in Domain Decomposition Methods*, Technical report TR/PA/00/17, CERFACS, France, 2000.

[9] D. Braess, P. Deuflhard, and K. Lipnikov, *A subspace cascadic multigrid method for mortar elements*, Computing, 69 (2002), pp. 205–225.

[10] D. Braess and R. Sarazin, *An efficient smoother for the Stokes problem*, Appl. Numer. Math., 23 (1997), pp. 3–19.

[11] J. H. Bramble, J. E. Pasciak, and A. T. Vassilev, *Analysis of the inexact Uzawa algorithm for saddle point problems*, SIAM J. Numer. Anal., 34 (1997), pp. 1072–1092.

[12] J. H. Bramble, J. E. Pasciak, and A. T. Vassilev, *Inexact Uzawa algorithms for nonsymmetric saddle point problems*, Math. Comp., 69 (2000), pp. 667–689.

[13] F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods*, Springer Ser. Comput. Math. 15, Springer-Verlag New York, 1991.

[14] J. W. Demmel, N. J. Higham, and R. S. Schreiber, Stability of *block LU factorization*, Numer. Linear Algebra Appl., 2 (1995), pp. 173–190.

[15] H. C. Elman and G. H. Golub, *Inexact and preconditioned Uzawa algorithms for saddle point problems*, SIAM J. Numer. Anal., 31 (1994), pp. 1645–1661.

[16] A. Frommer and D. B. Szyld, *H-splittings and two-stage iterative methods*, Numer. Math., 63 (1992), pp. 345–356.

[17] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*, Academic Press Inc., London, 1981.

[18] L. Giraud, S. Gratton, and J. Langou, *Convergence in backward error of relaxed GMRES*, SIAM J. Sci. Comput., 29 (2007), pp. 710–728.

[19] G. H. Golub and Q. Ye, *Inexact preconditioned conjugate gradient method with inner-outer iteration*, SIAM J. Sci. Comput., 21 (1999), pp. 1305–1320.

[20] N. I. M. Gould, M. E. Hribar, and J. Nocedal, *On the solution of equality constrained quadratic programming problems arising in optimization*, SIAM J. Sci. Comput., 23 (2001), pp. 1376–1395.

[21] A. Greenbaum, *Accuracy of computed solutions from conjugate-gradient-like methods*, in Advances in Numerical Methods for Large Sparse Sets of Linear Systems, vol. 10, M. Natori and T. Nodera, eds., Keio University, Yokohama, Japan, 1994, pp. 126–138.

[22] A. Greenbaum, *Estimating the attainable accuracy of recursively computed residual methods*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 535–551.

[23] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, SIAM, Philadelphia, 1996.

[24] P. Jiránek and M. Rozložník, *Limiting accuracy of segregated solution methods for nonsymmetric saddle point problems*, J. Comput. Appl. Math., to appear.

[25] C. Keller, N. I. M. Gould, and A. J. Wathen, *Constraint preconditioning for indefinite linear systems*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1300–1317.

[26] J. Maryška, M. Rozložník, and M. Tůma, *Schur complement reduction in the mixed-hybrid approximation of Darcy's law: Rounding error analysis*, J. Comput. Appl. Math., 117 (2000), pp. 159–173.

[27] N. K. Nichols, *On the convergence of two-stage iterative processes for solving linear equations*, SIAM J. Numer. Anal., 10 (1973), pp. 460–469.

[28] I. Perugia and V. Simoncini, *Block-diagonal and indefinite symmetric preconditioners for mixed finite element formulations*, Numer. Linear Algebra Appl., 7 (2000), pp. 585–616.

[29] A. Ramage and A. J. Wathen, *Iterative solution techniques for the Stokes and Navier-Stokes equations*, Internat. J. Numer. Methods Fluids, 19 (1994), pp. 67–83.

[30] M. Rozložník and V. Simoncini, *Krylov subspace methods for saddle point problems with indefinite preconditioning*, SIAM J. Matrix Anal. Appl., 24 (2002), pp. 368–391.

[31] V. Simoncini and D. B. Szyld, *Theory of inexact Krylov subspace methods and applications to scientific computing*, SIAM J. Sci. Comput., 25 (2003), pp. 454–477.

[32] Z. Strakoš and P. Tichý, *On error estimation in the conjugate gradient method and why it works in finite precision computations*, Electron. Trans. Numer. Anal., 13 (2002), pp. 56–80.

[33] J. van den Eshof and G. L. G. Sleijpen, *Inexact Krylov subspace methods for linear systems*, SIAM J. Matrix Anal. Appl., 26 (2004), pp. 125–153.

[34] P. A. Wedin, *Perturbation theory for pseudo-inverses*, BIT, 13 (1973), pp. 217–232.

[35] C. Wieners and B. I. Wohlmuth, *Duality estimates and multigrid analysis for saddle point problems arising from mortar discretizations*, SIAM J. Sci. Comput., 24 (2003), pp. 2163–2184.

[36] W. Zulehner, *A class of smoothers for saddle point problems*, Computing, 65 (2000), pp. 227–246.

[37] W. Zulehner, *Analysis of iterative methods for saddle point problems: A unified approach*, Math. Comp., 71 (2002), pp. 479–505.