Saddle point problems, iterative solution and preconditioning: a short overview

Miroslav Rozložník*

August 31, 2004

Abstract

In this contribution we attempt to review recent advances in the field of iterative methods for solving large saddle point problems. The main focus is on developments in the area of Krylov subspace methods and block preconditioning techniques for symmetric and nonsymmetric linear systems that arise in the context of solving the saddle point problems.

Keywords: saddle point problems, symmetric indefinite systems, iterative methods, stationary iterative methods, Krylov subspace methods, conjugate gradient method, preconditioning.

AMS Subject Classification: 65F10

1 Introduction

Particular attention has been paid recently to a particular class of linear algebraic equations known as saddle point problems. Saddle point problems arise in many applications such as fluid dynamics modelling via Navier-Stokes equations [17], [19], [46], [53], [54]; potential fluid flow in porous

*Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 2, CZ-182 07 Prague 8, Czech Republic. E-mail: miro@cs.cas.cz. Part of this work was supported by the Grant Agency of the Czech Republic under grant No. 201/02/0595 and by the Grant Agency of the Academy of Sciences of the Czech Republic under grant No. A1030103.

media [6], [34], [57], [61]; magnetostatic problems [44], [45]; structural mechanics [36]; quadratic or nonlinear programming [28], [38] ; optimization [2], [25] or least squares problems [7], [43]. Based on particular application one can distinguish between other names or synonyms, which in principle cover the same class of problems: the term augmented system is used in the context of (generalized) least squares problems [7] or the KKT (Karush-Kuhn-Tucker) system in optimization [25]. We are interested in solving the systems of linear algebraic equations with a particular block structure

$$\begin{pmatrix} A & B \\ B^T & 0 \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}, \tag{1}$$

where A is a nonsingular square matrix and B has a full column rank such that they ensure the existence and uniqueness of the solution. In many cases, the block A is symmetric positive definite, implying the symmetricity of the whole system. It is immediately clear that due to zero diagonal block is the system matrix of (1) indefinite. Although this fact may be considered as a potential drawback recent results indicate that indefinitness of the problem does not represent a principial difficulty and many efficient techniques can be used taking also into account its block structure, which differs the saddle point problem from a fully general indefinite system.

In the context of saddle point problems that arise from a discretization of elliptic partial differential equations [14] via mixed or mixed-hybrid formulation of the finite element method [11], [12], [42], [47], the diagonal block A usually describes physical properties or relations between the unknowns, while the off-diagonal block B describes the geometry of discretized domain. These dependencies can be mutually coupled in other applications. In addition, the choice of a particular right-hand side vector, which represents frequently the boundary conditions leading to "physical" solution, may have a significant influence on the whole solution process. An important aspect of mathematical modelling consists in the requirement to cope both with structural and numerical properties of the system matrix and the right-hand side vector, given by properties of a continuous problem, by its formulation as well as by properties of discretization technique. Therefore a particular attention is paid to the dependence on the discretization, which has an influence not only on the sparsity and

nonzero structure of matrix blocks A and B, but also on the dimension of the system matrix or its numerical properties such as its spectrum or conditioning. At least equally important is the dependence on physical parameters or material constants, and in many cases, it is a dominant indicator for the efficiency of a given solution approach. Therefore the robustness with respect to input parameters, which can be usually estimated or measured experimentally only with a certain precision, plays an important role in contemporary mathematical modelling.

This text is an attempt to give a brief overview of approaches, iterative methods and preconditioning techniques used for solving the saddle point problems. The first section recalls basic approaches and discusses their advantages and drawbacks. The second section is devoted to iterative methods for solving linear systems with a particular structure that arise in one of three basic approaches. These methods are applied either on the whole indefinite system or on a certain subproblem such as Schur complement system or the system projected onto null-space of some off-diagonal block. In the third section we briefly review the block saddle point preconditioners which are currently used for increasing the robustness of iterative methods applied on various levels of the solution process.

2 Basic approaches for saddle point problems

In this section we recall several basic approaches used for solving the saddle point systems. The first and perhaps also the most widely used approach is a reduction of the saddle point system (1) to some Schur complement system. It is clear from the first block equation of the system (1) that the unknown vector u, which is frequently called a vector of primary unknowns, can be expressed in the form

$$u = A^{-1}(f - Bp). \tag{2}$$

Substituting (2) into the second block equation of (1) we obtain the system for the unknown vector p, called a vector of dual unknowns, in the form

$$\left(B^T A^{-1} B\right) p = B^T A^{-1} f - g. \tag{3}$$

The system (3) is called a Schur complement system and the whole approach is due to the initial elimination of primary unknowns known as

a primary approach. In the framework of the finite element method this approach can be found as a static condensation [12], [14].

This approach will be efficient when it is easy to invert the diagonal block A or when it is easy to solve systems with the matrix A. Important for this approach is therefore a structure of its nonzero elements. When one can get directly the inverse A^{-1} (there is no significant fill-in in the matrix A^{-1}), it may be useful to construct also the Schur complement matrix $B^T A^{-1} B$ and to attempt to solve this system with some variant of Gaussian elimination. If the matrix A is symmetric positive definite we can use its Choleski factorization $A = LL^T$. The Schur complement system

$$(L^{-1}B)^T (L^{-1}B)p = (L^{-1}B)^T L^{-1} f - g, \tag{4}$$

is then (symmetric) positive definite and it can be also solved by Choleski factorization. The purely direct approach using sparse data structures for storing (only) nonzero elements during the elimination is feasible in some cases and using the graph theory one can estimate also its computational complexity [40].

If the Schur complement matrix $B^T A^{-1} B$ is a sparse matrix (in practice it can be also in a factorized form as a product of two or three sparse matrices), but the system (3) itself cannot be solved by a direct elimination, it may be useful to combine a direct technique in the first step (2) with an iterative method in the second step (3). In contrast to purely direct approach, where determining role is played by the (nonzero) structure of the Schur complement matrix, more important for iterative solution are numerical values in $B^T A^{-1} B$. Namely, in the next section we will show that in the case of a symmetric system the convergence of iterative Krylov subspace methods is determined by the eigenvalue distribution (or roughly speaking by the condition number) of a system matrix. The system (3) is in practical situations solved by iterative method with preconditioning (by a preconditioned iterative method) which frequently accelerates the rate of convergence of the original (unpreconditioned) iterative method. For examples of such approach we refer to papers [34], [40], [49], [50], [53], [54].

If the inverse of the matrix A itself is a dense matrix and/or the explicit constructing of $B^T A^{-1} B$ is not feasible, a purely iterative approach

can be another alternative. One can solve iteratively not only the system (3) but the iterative method can be applied at each step also for solving systems with the matrix A. This approach is in practice known as a combination of outer and inner iterative method (inner-outer iteration method). An example of such approach is a widely used and theoretically analyzed Uzawa method [18], [61]. Both outer and inner methods can be preconditioned whereas the inner preconditioner accelerates the convergence of the iterative method applied on systems with the matrix A and the outer preconditioner accelerates the convergence of iterative method applied on the Schur complement system (3).

As we have already mentioned the primary approach can be efficient especially when one can compute the inverse of A explicitly or when one can easily solve systems with this matrix block. The most important here seems to be the structure of nonzero elements, less attention is paid to numerical values in A or to its conditioning, although one cannot separate these two things completely. It is no surprise that the block A, describing frequently the material constants with huge jumps in magnitude or the physical dependencies in different units, can be very ill-conditioned for some applications. Then instead of elimination of primary unknowns it may be useful to leave solving the system with the matrix A to the very end of the whole solution process. The second possible approach, which is called a dual approach, is based on the expression of the unknown vector u from the second (underdetermined) block equation in (1) as

$$u = u_1 + Zu_2, \tag{5}$$

where Z is a matrix with columns that build a (possibly orthogonal) basis of the null-space of the matrix B^T, and thus it follows that $B^T Z = 0$. The unknown component u_1 from $Range(B)$ can be computed as some particular solution of the undetermined system $B^T u_1 = g$. The first block equation of the system (1) can be rewritten into $AZu_2 + Bp = f - Au_1$ and its projection onto $Null(B^T)$ leads to the system for the unknown component u_2

$$Z^T AZu_2 = Z^T(f - Au_1). \tag{6}$$

The system (6) is usually called a projected system. The unknown vector p can be then computed by solving the overdetermined system $Bp = f - Au$.

This approach will be efficient when the structure of nonzero elements of the matrix B allows to determine or to compute easily the basis vectors of $Null(B^T)$. Moreover it would be ideal to have basis vectors and so the whole matrix Z sparse, normalized and if possible orthogonal. To achieve this there exists a dozen of direct methods. The most frequent are the orthogonalization algorithms for QR factorization [1], [26], but in general one can also use Gaussian elimination [2]. If we consider the factorization $B = QR$, where Q is orthogonal and R is upper triangular, then the vector u_1 can be obtained as $u_1 = BR^{-1}R^{-T}g$ and the projected system (6) can be reformulated into

$$(I - \Pi)A(I - \Pi)u_2 = (I - \Pi)(f - Au_1), \tag{7}$$

where Π is a matrix of the orthogonal projector onto $Range(B)$ expressed via B or via B and R as $\Pi = B(B^TB)^{-1}B^T = B(R^TR)^{-1}B^T$. If there is no significant fill-in in the matrix Z, then the projected matrix Z^TAZ may in some cases preserve its sparse structure and one can attempt to solve the system (6) by using Gaussian elimination. If A is symmetric positive definite, then the projected matrix Z^TAZ is symmetric positive definite and one may use Choleski decomposition.

When the projected system (6) cannot be solved using direct methods either due to the fact that its matrix is only in a factorized form or due to large fill-in in the Gaussian elimination of Z^TAZ, the combination with the iterative solution can be an efficient alternative. An important role in here will have again numerical values of Z^TAZ. Of course, in practice one uses a preconditioned iterative method. As an example of such approach we refer to solving the saddle point problems which arise in quadratic programming [2].

One can also use a purely iterative approach. For construction orthogonal projections in the iterative method for solving the projected system (7) where $\Pi = B(B^TB)^{-1}B^T$, can be the system with the positive definite matrix B^TB solved also iteratively. Again, we obtain a combination of outer and inner iterative method (inner-outer iteration method). Of course each of them can be preconditioned, where the inner preconditioner accelerates the convergence of iterative solver applied on systems with B^TB and the outer preconditioner accelerates the convergence of the method applied on (7). This case may, e.g., occur when the inverse of

$B^T B$ is a dense matrix or even if it is not suitable to construct the matrix $B^T B$ itself. For an example of such approach we refer to [45].

Besides already mentioned advantage of the dual approach another supporting fact is that in many applications the off-diagonal block B describes the geometry of discretized domain, which is constant while the material constants represented in the block A are time-dependent. Then it is clearly more efficient to compute a basis of $Null(B^T)$ only once and to solve the projected system (7) at each time step with a new block A. The important role in the dual approach is therefore played by numerical values in B. On the other hand, if the off-diagonal block is ill-conditioned this approach may not be competitive with the primary approach. This may happen in the case of a complicated geometry with relatively simple physical dependencies.

Another possible approach for saddle point problems is to solve the system (1) as a whole system. Again one can distinguish between the use of purely direct or purely iterative method. From direct techniques are mainly used symmetric variants of Gaussian elimination, suited for solving systems which are not positive definite, such as Bunch-Parlett decomposition [26]. In contrast to general indefinite systems the particular block structure of saddle point problems has interesting properties which can be useful for implementing the elimination scheme. E.g., it was shown in [56] that there exists a whole class of saddle point problems, which arise from a discretization of partial differential equations, where one does not need 2×2 pivots in the elimination, and thus the whole process can be performed using operations of a sparse Choleski factorization.

When solving the saddle point problems in a purely iterative way the first question is a choice of iterative method which is suitable for solving systems with this particular indefinite structure. Here one may consider either some stationary iterative method, induced by certain splitting of the system matrix, or some nonstationary Krylov subspace method that converges for indefinite systems. It appears that this problem is nowadays well settled especially in the symmetric case, where the convergence of iterative Krylov subspace methods can be described by the spectrum of a given system matrix [29]. This area has been recently a target of extensive research, which resulted into many publications on saddle problems from different applications (see, e.g., [19], [27], [44], [58], [59]). A common

feature of these papers is that the block structure itself has a marginal influence on the choice of the Krylov subspace method, while it may be an important factor for the choice of efficient preconditioner or a matrix splitting, which are frequently strongly dependent on the application where does the saddle point problem come from.

3 Iterative methods for linear algebraic systems

In previous section we have recalled three main approaches for solving the saddle point problems. We have shown that one can also use iterative methods on various levels of the solution process. The use of iterative method can be also well justified due to the fact that in practical applications the problem with inaccurate information or the discretization error may provide a reasonable stopping criterion for its termination, often earlier than the time spent by some direct technique. In this section we give a brief overview of iterative methods used most frequently in the context of saddle point problems. We consider a linear system

$$\mathcal{A}x = b, \tag{8}$$

where $\mathcal{A}$ is a nonsingular matrix and b a right-hand side vector. In our context the system (8) can be the whole system (1), the Schur complement system (3) or the projected system (7).

The class of stationary iterative methods (Hageman, Young [32]) is based on a matrix splitting in the form $\mathcal{A} = \mathcal{M} - \mathcal{N}$, where $\mathcal{M}$ is a nonsingular matrix. The n-th approximate solution x_n is given by the recurrence formula

$$x_n = (I - \mathcal{M}^{-1}\mathcal{N})x_{n-1} + \mathcal{M}^{-1}b, \tag{9}$$

where x_0 is an initial guess. It is well-known that the asymptotic rate of convergence of stationary methods is given by the spectral radius of the matrix $\mathcal{M}^{-1}\mathcal{N}$ [29], [32]. Note that $\mathcal{M}^{-1}\mathcal{N}$ is generally nonsymmetric as well as the matrices $\mathcal{M}$ and $\mathcal{N}$ need not be symmetric. In the saddle point context a common feature of these schemes is that the splitting of the matrix $\mathcal{A}$ is induced by the block structure of (1) keeping usually some blocks while modifying the others. For examples of such approaches we refer to [4], [5], [9] or [27].

Nonstationary iterative methods based on Krylov subspaces (Krylov subspace methods) [29], [51] generate a sequence of approximate solutions in the form

$$x_n \in x_0 + K_n(\mathcal{A}, r_0), \tag{10}$$

where $r_0 = b - \mathcal{A}x_0$ is the initial residual; $K_n(\mathcal{A}, r_0)$ denotes the n-th Krylov subspace associated with $\mathcal{A}$ and r_0 and it is defined as $K_n(\mathcal{A}, r_0) = span\,\{r_0, \mathcal{A}r_0, \ldots, \mathcal{A}^{n-1}r_0\}$. It follows from (10) that for the error $x - x_n$ of the n-th approximate solution x_n and the residual $r_n = b - \mathcal{A}x_n$ we have $x - x_n = P_n(\mathcal{A})(x - x_0)$ and $r_n = P_n(\mathcal{A})r_0$, where P_n stands for some polynomial of degree at most n satisfying $P_n(0) = 1$. The whole class od such polynomials will be denoted as Π_n. In literature one can also find for this class of methods the names like the class of polynomial accelerations of stationary iterative methods (Hageman, Young [32]) or the generalized conjugate gradient methods (Weiss [60]).

The most known and widely used Krylov subspace method is the conjugate gradient method (CG) introduced by Hestenes and Stiefel in 1952 [33]. This method has been proposed for solving systems with symmetric positive definite matrix $\mathcal{A}$ and it generates the approximate solutions (10) which minimize the energy norm of the error, i.e.,

$$\|x - x_n\|_{\mathcal{A}} = \min_{u \in x_0 + K_n(\mathcal{A}, r_0)} \|x - u\|_{\mathcal{A}}. \tag{11}$$

Using $x - x_n = P_n(\mathcal{A})(x - x_0)$ and some expansion of the initial error $x - x_0$ into orthogonal eigenvector basis (the matrix $\mathcal{A}$ is symmetric and positive definite) we obtain a bound for the error in the form

$$\|x - x_n\|_{\mathcal{A}} \leq \min_{P \in \Pi_n} \max_{\lambda_i \in \sigma(\mathcal{A})} |P(\lambda_i)|\, \|x - x_0\|_{\mathcal{A}}. \tag{12}$$

This bound is in a sense sharp; for every n there exists an initial residual r_0 (dependent on n) such that equality holds in (12) [29, pp. 50-51]. The convergence of the CG method thus depends strongly on the eigenvalue distribution of the matrix $\mathcal{A}$. The right-hand side of (12) can be further estimated considering a polynomial approximation problem on the minimal interval $[a, b]$ that covers the spectrum $\sigma(\mathcal{A})$ [29]. Solving this problem we obtain a well-known bound for the relative error in the CG method

$$\frac{\|x - x_n\|_{\mathcal{A}}}{\|x - x_0\|_{\mathcal{A}}} \leq 2 \left(\frac{\sqrt{\kappa(\mathcal{A})} - 1}{\sqrt{\kappa(\mathcal{A})} + 1} \right)^n. \tag{13}$$

This bound can be found almost in every textbook on the CG method and it serves as a basic tool for estimating its convergence rate in a given application.

When the matrix $\mathcal{A}$ is symmetric but not positive definite, the most frequently used method is the minimal residual (MINRES) method [43] proposed by Paige and Saunders in 1975. The MINRES method computes the approximate solutions (10) which minimize the residual norm over a sequence of Krylov subspaces

$$\|b - \mathcal{A}x_n\| = \min_{u \in x_0 + K_n(\mathcal{A}, r_0)} \|b - \mathcal{A}u\|. \tag{14}$$

Using analogous approach as for the CG method we can obtain a bound for the residual norm in the form

$$\|r_n\| \leq \min_{P \in \Pi_n} \max_{\lambda_i \in \sigma(\mathcal{A})} |P(\lambda_i)| \, \|r_0\|. \tag{15}$$

This bound is sharp in the same manner as (12) and so the convergence rate of MINRES is determined by the eigenvalue distribution of the matrix $\mathcal{A}$. Since $\mathcal{A}$ is symmetric but generally indefinite the inclusion set for the spectrum is formed by two disjoint intervals $[-c, -d] \cup [a, b]$ (one interval on the positive and one interval on the negative side of the real axis). The polynomial approximation problem on this set has always a unique solution [21] but the optimal polynomial is analytically known only in special cases such as $[-b, -a] \cup [a, b]$. Therefore it is significantly harder task to get a reasonable practical bound for the MINRES method. Frequently one can obtain only a bound for the asymptotic convergence factor [15], [58], [59] defined as $\lim_{n \to \infty} \left[\min_{P \in \Pi_n} \max_{\lambda_i \in \sigma(\mathcal{A})} |P(\lambda_i)| \right]^{\frac{1}{n}}$.

In the nonsymmetric case is the situation even less transparent. Tens of iterative methods were proposed, in practice, however, only a few of them are really used [29], [51]. The most frequently analyzed is the generalized minimum residual (GMRES) method due to Saad and Schultz which is also based on the residual norm minimization (14). The GMRES method is thus a direct generalization of the MINRES method to nonsymmetric case. If the system matrix $\mathcal{A}$ is diagonalizable $\mathcal{A} = X \Lambda X^{-1}$, then there exists a bound for the residual norm (Elman [16, p.40], [51, pp. 134-135]) in the form

$$\|r_n\| \leq \kappa(X) \min_{P \in \Pi_n} \max_{\lambda_i \in \sigma(\mathcal{A})} |P(\lambda_i)| \, \|r_0\|. \tag{16}$$

Therefore, if the condition number $\kappa(X)$ of the eigenbasis X is reasonably bounded, one can use (16) similarly to (15) as in the symmetric case, where the convergence rate is determined by the eigenvalue distribution of the system matrix.

In the general case the system matrix is non-diagonalizable and arguments about the density of the class of diagonalizable matrices in the whole matrix space are not sufficient for extending the bound (16) for arbitrary non-diagonalizable (or non-normal) system [3], [31], [30]. Frequently used bound in applications is based on the field of values of the matrix $\mathcal{A}$ [16, p.40], [51, p.195]. It follows that

$$\|r_n\| \leq \left[1 - \left(\frac{\min_x(\mathcal{A}x, x)}{\max_x(\mathcal{A}x, x)}\right)^2\right]^{\frac{n}{2}} \|r_0\|. \tag{17}$$

The field of values of the matrix $\mathcal{A}$ can be analyzed for some applications in terms of discretization parameters and consequently the convergence rate of the GMRES can be estimated via (17) [19], [37]. The nonsymmetric GMRES method can be implemented only using full-term recurrences which significantly limit its practical applicability [51], [52]. Therefore nonsymmetric iterative methods based on short-term recurrences are used. They generate the approximate solutions (10) which are, however, not optimal in the sense of error or residual norm minimization. Nevertheless, although these methods may not converge and are difficult to analyze, they usually work on practical problems. The most important methods are the biconjugate gradient (Bi-CG) method [22], the quasi-minimal residual (QMR) method [23], [24] or the Bi-CGSTAB method [22], [51]. An excellent overview of symmetric methods can be found in [55]; while the paper [22] gives a comprehensive review of Krylov subspace methods used for nonsymmetric systems.

In practical situations, up to a very limited number of artificially constructed examples where one method clearly outperforms the others, there is no significant difference between the behavior of different iterative methods. In addition, there are many relations between residuals of various methods. In the symmetric case, e.g., there exists a relation between the CG and MINRES method [29, p.86], [62]. It appears that the norms of

their residuals are mutually connected via formula

$$\frac{\|r_n^{MINRES}\|}{\|r_n^{CG}\|} = \sqrt{1 - \left(\frac{\|r_n^{MINRES}\|}{\|r_{n-1}^{MINRES}\|}\right)^2}. \tag{18}$$

This relation describes the "peak/plateau" behavior for this pair of methods in the case of a general symmetric (indefinite) system, where the CG method may have large residual norm or even some of its approximate solutions may not exist (then $\|r_n^{CG}\|$ is considered as infinitely large and (18) in this sense still holds). Usually such situation is accompanied with the jump (peak) in the convergence curve of CG and we have $\|r_n^{MINRES}\| \ll \|r_n^{CG}\|$. Then (18) implies the stagnation (plateau) of the residual norm in the MINRES method $\|r_n^{MINRES}\| \approx \|r_{n-1}^{MINRES}\|$. On the other hand, when MINRES converges quickly, then is the CG residual comparable to that in the MINRES method. The same "peak/plateau" relation, which was actually derived first in the nonsymmetric case, holds for the FOM and GMRES method [13] and also with some modification for the Bi-CG and QMR method. A detailed analysis of the relation between Bi-CG and QMR residuals can be found in [29], [60] or [62]. The fact that there is no substantial difference in using various iterative schemes is even more profound for preconditioned Krylov subspace methods, where the efficiency of the solver is actually determined not by the choice of a particular method but by the choice of a preconditioner. Clearly, an efficient preconditioner "hides" all local differences between various methods, which essentially show very similar global convergence behavior (usually one observes the termination after several iteration steps).

4 Preconditioning of saddle point problems

The convergence and robustness of iterative methods for solving linear algebraic equations are in practical applications accelerated resp. increased by preconditioning. Preconditioning of the system is actually a transformation of the original system (8) to the new system

$$\tilde{\mathcal{A}}\tilde{x} = \tilde{b}, \tag{19}$$

which is called a preconditioned system and the transformation matrix itself (in the text it will be denoted by $\mathcal{P}$) is called a preconditioning matrix.

The concept of preconditioning itself is not new and many preconditioning techniques have been proposed in the last several decades. They vary in algebraic construction and in dependency on particular applications. A basic overview of preconditioning schemes can be found in [51] or in the corresponding chapter of [29]. In this contribution we focus on the most frequent preconditioning techniques used in the context of saddle point problems.

It seems natural that in the case of the symmetric positive definite system (8) is the preconditioning matrix $\mathcal{P}$ also symmetric positive definite. The preconditioned system (19) can be then rewritten into

$$\left(\mathcal{P}^{-1/2}\mathcal{A}\mathcal{P}^{-T/2}\right)\left(\mathcal{P}^{T/2}x\right) = \mathcal{P}^{-1/2}b, \tag{20}$$

where $\tilde{\mathcal{A}} = \mathcal{P}^{-1/2}\mathcal{A}\mathcal{P}^{-T/2}$, $\tilde{x} = \mathcal{P}^{T/2}x$ and $\tilde{b} = \mathcal{P}^{-1/2}b$. The matrix $\tilde{A}$ is again symmetric positive definite and one can apply the same iterative method (usually the CG method [33]). The direct application of the CG method on (19) would lead to a sequence of approximate solutions to the vector $\tilde{x}$, but we want to compute the solution $x = \mathcal{P}^{-T/2}\tilde{x}$. Using a backward transformation from the "tilde" quantities to original ones in the CG method (for details we refer, e.g., to [51]) one can obtain a sequence of approximate solutions to the solution of (8). Formally, this approach corresponds to solving the system $\mathcal{P}^{-1}\mathcal{A}x = \mathcal{P}^{-1}b$ or $\mathcal{A}\mathcal{P}^{-1}\hat{x} = b$, where $\hat{x} = \mathcal{P}x$ (again, the details can be found in [51]). Note that $\mathcal{P}^{-1}\mathcal{A}$ and $\mathcal{A}\mathcal{P}^{-1}$ are generally nonsymmetric and the equivalence is possible only due to the transformation (20). Ideally we seek a preconditioning matrix $\mathcal{P}$, which is spectrally equivalent to the system matrix $\mathcal{A}$, i.e., there exist positive constants γ and Γ such that for every nonzero vector x we have $\gamma \leq \frac{\|\mathcal{A}x\|}{\|\mathcal{P}x\|} \leq \Gamma$. For the relative error in the preconditioned CG method (measured by the $\mathcal{A}\mathcal{P}^{-1}$-norm) it follows that

$$\frac{\|x - x_n\|_{\mathcal{A}\mathcal{P}^{-1}}}{\|x - x_0\|_{\mathcal{A}\mathcal{P}^{-1}}} \leq 2\left(\frac{\sqrt{\kappa(\mathcal{A}\mathcal{P}^{-1})} - 1}{\sqrt{\kappa(\mathcal{A}\mathcal{P}^{-1})} + 1}\right)^n \leq 2\left(\frac{\Gamma - \gamma}{\Gamma + \gamma}\right)^n. \tag{21}$$

If the constants γ and Γ are independent of the matrix dimension, then also the rate of convergence of CG will not be dependent on the dimension (the condition number of $\mathcal{P}^{-1}\mathcal{A}$ or $\mathcal{A}\mathcal{P}^{-1}$ is less or equal than Γ/γ).

The indefinitness of the saddle point problem (1) brings into the field of preconditioning a completely new element. If the matrix $\mathcal{A}$ is symmetric but indefinite, it is not entirely clear what properties should have the preconditioning matrix $\mathcal{P}$. If $\mathcal{P}$ is symmetric positive definite, then the transformation (20) leads again to a symmetric preconditioned system (19). In this case one can apply on (19) iterative method suitable for solving (symmetric) indefinite systems such as already mentioned MINRES [43]. In addition, a similar backward transformation as in the positive definite case can be used (although it may be more difficult for some implementations). When the matrix $\mathcal{P}$ is indefinite its square root does not exist and the preconditioning of the symmetric system (8) by (general) symmetric matrix $\mathcal{P}$ leads surprisingly to solving the nonsymmetric system

$$\mathcal{A}\mathcal{P}^{-1}\hat{x} = b, \quad \hat{x} = \mathcal{P}x. \tag{22}$$

The same holds if the saddle point problem (1) itself or if the preconditioning matrix $\mathcal{P}$ is nonsymmetric. The matrix $\mathcal{A}\mathcal{P}^{-1}$ is generally nonsymmetric, and thus one has to solve the nonsymmetric preconditioned system (22). The first question is which method should be applied. The situation is even more difficult for some cases where the system matrix (22) is not diagonalizable [19], [35]. On the other hand, this is not a fully general non-diagonalizable (non-normal) case; the matrices have only 2×2 nontrivial Jordan blocks in those applications. It appears that due to particular structure of saddle point problems and also due to particular right-hand side vectors in some applications one can often use also symmetric iterative methods such the CG or MINRES method [39], [48], not mentioning the stationary iterative methods [4], [5], [9] or [27], where the matrix splitting can be also seen as a preconditioning of the Richardson method. The use of such methods must be then always theoretically justified and one should prove their convergence for every particular application.

As we have noticed already if the system (1) is nonsymmetric or the preconditioning matrix $\mathcal{P}$ is not positive definite, the system (22) is nonsymmetric, and one must apply nonsymmetric iterative method. Despite its computational cost the most frequently used and theoretically analyzed is the GMRES method. Its use is justified due to the fact that the method with efficient preconditioning often converges very quickly and its

approximate solution reaches a desired tolerance level much earlier than it becomes infeasible. The majority of the GMRES convergence analyses in applications is based on the field of values of the preconditioned matrix $\mathcal{A}\mathcal{P}^{-1}$ and on using the bound (17) [37]. Ideally one looks for a preconditioning matrix $\mathcal{P}$, which is equivalent to the matrix $\mathcal{A}$ with respect to the field of values, i.e., there exist the constants γ and Γ such that for every nonzero vector x we have $\gamma(x,x) \leq (x, \mathcal{A}\mathcal{P}^{-1}x)$ and $\|\mathcal{A}\mathcal{P}^{-1}x\| \leq \Gamma\|x\|$. The GMRES residual can be then bounded by

$$\|r_n\| \leq \left[1 - \left(\frac{\min_x(\mathcal{A}\mathcal{P}^{-1}x, x)}{\max_x(\mathcal{A}\mathcal{P}^{-1}x, x)}\right)^2\right]^{\frac{n}{2}} \leq \left[1 - \left(\frac{\gamma}{\Gamma}\right)^2\right]^{\frac{n}{2}} \|r_0\|. \tag{23}$$

Again, if γ and Γ are independent of the system matrix dimension, then the convergence rate of GMRES will not be dependent on the system dimension. This approach is mainly used for "inexact" versions of block-diagonal or block-triangular preconditioners, where one of their diagonal blocks is negative definite leading to indefinite preconditioner $\mathcal{P}$ for the whole system (1) (see the comments later). Another possible approach is based on the analysis of a degree of the minimal polynomial for the matrix $\mathcal{A}\mathcal{P}^{-1}$. It appears that this degree can be sometimes very small (it can be even equal to 3!). Then one can expect that every Krylov subspace method will terminate within this (small) number of steps. This property holds for so called "exact" versions of block-diagonal or block-triangular preconditioners [19], or with some modification also for the symmetric indefinite preconditioner [38] which will be discussed later, too.

The construction and analysis of preconditioners for saddle point problems have been a subject of current research in a whole range of applications as a porous media flow [10], [20], [49], [50], [61]; solving of Navier-Stokes equations [19], [27], [46], [53], [54], [59]; magnetostatic problems [44], [45]; optimization or quadratic programming [2], [25], etc. Instead of paying attention to some particular application in the following we give a short review of preconditioners with various block structures. One can distinguish between the block diagonal preconditioners in the form

$$\mathcal{P} = \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix}, \tag{24}$$

where the blocks P_1 and P_2 are either diagonal matrices; in such case

we talk about simple scalings of the system matrix (1) [53]; or they are general matrices, where P_1 is usually a preconditioner for A and P_2 is a preconditioner for $B^T A^{-1} B$ or for its approximation $B^T P_1^{-1} B$ [37], [54]. For arbitrary nonzero vectors u and p with corresponding dimensions we require that $\gamma_1 \|P_1 u\| \leq \|Au\| \leq \Gamma_1 \|P_1 u\|$ and $\gamma_2 \|P_2 p\| \leq \|B^T P_1^{-1} B p\| \leq \Gamma_2 \|P_2 p\|$, where γ_1; γ_2 and Γ_1; Γ_2 are positive constants often independent of the problem dimension. When A is symmetric positive definite, the block P_1 is usually also symmetric positive definite; the block P_2 can be either positive definite (leading to indefinite but still symmetric preconditioned system) or negative definite (then we talk about indefinite preconditioning with all its consequences). It is easy to see that the choice $P_1 = A$ corresponds to solving the system with the Schur complement matrix $B^T A^{-1} B$ using a preconditioner P_2. This gives a connection to the primary approach for solving the saddle point problem described in the first section of this manuscript. In addition, it can be shown (we refer to [41]) that if we use its "exact" variant

$$\mathcal{P} = \begin{pmatrix} A & 0 \\ 0 & \pm B^T A^{-1} B \end{pmatrix}, \tag{25}$$

then the preconditioned matrix $\mathcal{A}\mathcal{P}^{-1}$ is diagonalizable and it has three nonzero eigenvalues ($\{\ 1,\ \frac{1}{2} \pm \frac{\sqrt{5}}{2}\ \}$). Thus every Krylov subspace method, including GMRES, will terminate in atmost three steps. The preconditioner can be also nonsymmetric: the second large group of block preconditioners has the block-triangular form

$$\mathcal{P} = \begin{pmatrix} P_1 & B \\ 0 & P_2 \end{pmatrix}, \tag{26}$$

where again the matrix blocks P_1 and P_2 are preconditioners for A and $B^T A^{-1} B$, respectively (including the possibility that P_2 is negative definite) [19], [37]. They are usually required to satisfy relations $\gamma_1(u,u) \leq (u, AP_1^{-1}u)$, $\|AP_1^{-1}u\| \leq \Gamma_1 \|u\|$ and $\gamma_2(p,p) \leq (p, (B^T P_1^{-1} B) P_2^{-1} p)$, $\|(B^T P_1^{-1} B) P_2^{-1} p\| \leq \Gamma_2 \|p\|$ for some constants γ_1, γ_2 and Γ_1, Γ_2. The analysis of "exact" versions of block triangular preconditioners in [41]

$$\mathcal{P} = \begin{pmatrix} A & B \\ 0 & \pm B^T A^{-1} B \end{pmatrix}, \tag{27}$$

leads in the case of the positive definite $P_2 = B^T A^{-1} B$ to the diagonalizable preconditioned system and in the case of the negative definite $P_2 = -B^T A^{-1} B$ to the preconditioned system which is not diagonalizable, having the form

$$\mathcal{A}\mathcal{P}^{-1} = \begin{pmatrix} I & 0 \\ B^T A^{-1} & \mp I \end{pmatrix}. \tag{28}$$

In the former case has the preconditioned matrix two nonzero eigenvalues ($\{\pm 1\}$), in the latter case this matrix has only unit eigenvalues. A considerable attention has been recently paid to the symmetric indefinite preconditioner

$$\mathcal{P} = \begin{pmatrix} I & B \\ B^T & 0 \end{pmatrix}, \tag{29}$$

which is in a sense equivalent to the dual approach described in the section devoted to basic approaches for saddle point problems (for details we refer to [45]). This preconditioner leads in its "exact" form (29) to the nonsymmetric preconditioned system

$$\mathcal{A}\mathcal{P}^{-1} = \begin{pmatrix} A(I-\Pi)+\Pi & (A-I)B(B^T B)^{-1} \\ 0 & I \end{pmatrix}, \tag{30}$$

which has the upper block-triangular structure ($\Pi = B(B^T B)^{-1} B^T$). This preconditioner has been studied in several publications [8], [28], [35], [38], [48] and it appears that the degree of minimal polynomial is given by the number of mutually different nonzero eigenvalues of the symmetric positive semi-definite matrix $(I-\Pi)A(I-\Pi)$ that plays a significant role in (7). The "inexact" versions of such indefinite preconditioners were tested in [44].

The last but not least aspect we want to discuss here are such practically feasible preconditioners which lead to the number of iterations steps independent of the discretization parameters (and so lead to the independence on a system dimension). In some applications (as a rectangular domain, spatial isotropy, homogeneous boundary conditions etc.) one can have the spectral equivalence of matrices $\mathcal{P}$ and $\mathcal{A}$ for the case of positive definite preconditioning or the field-of-values equivalence for the case

of indefinite preconditioning using relatively simple and robust techniques (see, e.g., the paper [37] and the references therein). Majority of preconditioners, which lead to the independence on the matrix dimension, uses for P_1 and P_2 the multilevel or multigrid techniques [17], [37]. The alternative goal is the independence or a moderate dependence on some physical parameters or material constants of the problem such as viscosity parameter in Navier-Stokes equations or permeability coefficients in porous media flow. This topic is still a subject of recent research.

5 Concluding remarks

The field of solving saddle point problems has made a great progress in the last few years, but there are still many open questions related especially to the analysis and to the implementation of efficient preconditioning techniques, not mentioning more general questions related to the convergence description of Krylov subspace methods for various applications and linear algebra problems.

References

[1] E. Anderson, Z. Bai, C. Bischof, J. Demmel, J. J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, S. Ostrouchov and D. Sorensen. *LAPACK User's Guide*, SIAM, Philadelphia, 1992.

[2] M. Arioli. The use of QR factorization in sparse quadratic programming. *SIAM J. Matrix Anal. Appl.* 21 (2000), 829–839.

[3] M. Arioli, V. Pták and Strakoš, Krylov Sequences of Maximal Length and Convergence of GMRES, *BIT* 38 (1998), pp. 636–643.

[4] M. Benzi and G. H. Golub. A Preconditioner for Generalized Saddle Point Problems, October 2002. Revised, September 2003. 22 pages. To appear in SIAM Journal on Matrix Analysis and Applications.

[5] M. Benzi, M. J. Gander and G. H. Golub. Optimization of the Hermitian and Skew-Hermitian Splitting Iteration for Saddle-Point Problems, October 2002. Revised April and July 2003. 19 pages. To appear in BIT Numerical Mathematics.

[6] L. Bergamaschi, S. Mantica and F. Saleri. Mixed finite element approximation of Darcy's law in porous media, Tech. Rep. CRS4, Cagliari, 1994.

[7] Å. Björck. *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996.

[8] D. Braess, P. Deuflhard and K. Lipikov. A Subspace Cascadic Multigrid Method for Mortar Elements, Preprint SC-99-07, Konrad-Zuse-Zentrum, Berlin, 1999.

[9] D. Braess and R. Sarazin. An efficient smoother for the Stokes problem, *Appl. Numer. Math.*, 23 (1997), pp. 3–20.

[10] J.H. Bramble and J.E. Pasciak. A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems, *Math. Comp.* 50 (1988), 1–17.

[11] F. Brezzi. On the existence, iniqueness and approximation of saddle-point problems arising from Lagrangian multipliers, *Revue Francaise d'Automatique, Informatique et Recherche Operationel* 1(1974), 8–22.

[12] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods.* Springer-Verlag, 1991.

[13] P. N. Brown. A theoretical comparison of the Arnoldi and GMRES algorithms, *SIAM J. Sci. Stat. Comp.*, 12 (1991), pp. 58–78.

[14] P.G. Ciarlet. *The Finite Element Method for Elliptic Problems*, North-Holland Publishing Company, Amsterdam, 1978.

[15] M. Eiermann, X. Li and R. S. Varga. On hybrid semi-iterative methods, *SIAM J. Numer. Anal.*, 26 (1989), pp. 152–168.

[16] H.C. Elman. *Iterative methods for large, sparse, nonsymmetric systems of linear equations.* PhD. Thesis, Research Report No. 229, Yale University, 1982.

[17] H.C. Elman. Multigrid and Krylov subspace methods for the discrete Stokes equations, International Journal for Numerical Methods in Fluids 22 (1996), 755-770.

[18] H.C. Elman and G.H. Golub. Inexact and preconditioned Uzawa algorithms for saddle point problems. *SIAM J. Numer. Anal. 31(1994).*

[19] H. C. Elman, D. J. Silvester and A. J. Wathen. Iterative methods for problems in computational fluid dynamics, *in Iterative Methods in Scientific Computing, R. H. Chan, C. T. Chan, and G. H. Golub, eds.*, Springer-Verlag, Singapore, 1997, pp. 271–327.

[20] R. E. Ewing, R. D. Lazarov, P. Lu and P. S. Vassilevski. Preconditioning indefinite systems arising from mixed finite element discretization of second-order elliptic problems, *in Preconditioned Conjugate Gradient Methods, Lecture Notes in Math. 1457*, Springer-Verlag, Berlin, 1990, pp. 28–43.

[21] B. Fischer. Chebyshev polynomials for disjoint compact sets, *Constr. Approx.* 8 (1992), pp. 309–329.

[22] R. W. Freund, G. H. Golub and N. M. Nachtigal. Iterative Solution of Linear Systems, *Acta Numerica* 1 (1992), pp. 1–44.

[23] R. W. Freund and N. M. Nachtigal. QMR: A quasi-minimal residual method for non-Hermitian linear systems, *Numer. Math.*, 60 (1991), pp. 315–339.

[24] R. W. Freund and N. M. Nachtigal. Software for simplified Lanczos and QMR algorithms, *Appl. Numer. Math.*, 19 (1995), pp. 319–341.

[25] P. E. Gill, W. Murray, D. B. Ponceleón and M. A. Saunders. Preconditioners for indefinite systems arising in optimization, *SIAM J. Matrix Anal. Appl.*, 13 (1992), pp. 292–311.

[26] G. Golub and C. F. Van Loan. *Matrix Computations*, 3rd ed., The Johns Hopkins University Press, Baltimore, MD, 1996.

[27] G. H. Golub and A. J. Wathen. An iteration for indefinite systems and its application to the Navier–Stokes equations, *SIAM J. Sci. Comput.*, 19 (1998), pp. 530–539.

[28] 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.

[29] A. Greenbaum. *Iterative Methods for Solving Linear Systems*, SIAM, Philadelphia, 1997.

[30] A. Greenbaum, V. Pták and Z. Strakoš. Any Convergence Curve is Possible for GMRES, *SIAM Matrix Anal. Appl.* 17 (1996), pp. 465-470.

[31] A. Greenbaum and Z. Strakoš. Z. Matrices that Generate the Same Krylov Varieties, in: Recent Advances in Iterative Methods, G.H.Golub et al. (eds.), *IMA Volumes in Maths and Its Applications, Springer*, 1994, pp. 95-119.

[32] L. A. Hageman and D. M. Young. *Applied Iterative Methods*, Academic Press, New York, 1981.

[33] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems, *J. Res. Nat. Bureau of Standards* 49 (1952), 409–436.

[34] E. F. Kaasschieter and A. J. M. Huijben. Mixed-hybrid finite elements and streamline computation for the potential flow problem, *Numer. Methods Partial Differential Equations*, 8 (1992), pp. 221–266.

[35] 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.

[36] J. Kruis, K. Matouš and Z. Dostál. Solving laminated plates by domain decomposition, *Advances in Engineering Software* 33 (2002), pp. 445–452.

[37] D. Loghin and A. J. Wathen. Analysis of preconditioners for saddle-point problems, Report number 02/13, Oxford University Computing Laboratory, July 2002.

[38] L. Lukšan and J. Vlček. Indefinitely preconditioned inexact Newton method for large sparse equality constrained non-linear programming problems, *Numer. Linear Algebra Appl.*, 5 (1998), pp. 219–247.

[39] L. Lukšan and J. Vlček. Conjugate gradient methods for saddle point systems, *in Proceedings of the 13th Summer School on Software and Algorithms of Numerical Mathematics, I. Marek, ed.*, Nečtiny, Czech Republic, 1999, pp. 223–230.

[40] J. Maryška, M. Rozložník and M. Tůma. Schur complement systems in the mixed-hybrid finite element approximation of the potential fluid flow problem, *SIAM J. Sci. Comput.*, 22 (2000), pp. 704–723.

[41] M. F. Murphy, G. H. Golub and A. J. Wathen. A note on preconditioning for indefinite linear systems, *SIAM J. Sci. Comput.*, 21 (2000), pp. 1969–1972.

[42] J.T. Oden and J.K. Lee. Dual-mixed hybrid finite element method for second-order elliptic problems, *In: Mathematical Aspects of Finite Element Methods, Lecture Notes in Mathematics 606, Ed. I. Galligani, E. Magenes*, Springer-Verlag, Berlin, 1977, 275-291.

[43] C.C. Paige and M.A. Saunders. Solution of sparse indefinite systems of linear equations. *SIAM J. Numer. Anal. 12 (1975), pp. 617-629.*

[44] 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.

[45] I. Perugia, V. Simoncini and M. Arioli. Linear algebra methods in a mixed approximation of magnetostatic problems, *SIAM J. Sci. Comput.*, 21 (1999), pp. 1085–1101.

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

[47] P.A. Raviart and J.M. Thomas. A mixed finite element method for 2-nd order elliptic problems, *in: Mathematical Aspects of Finite Element Methods, Lecture Notes in Mathematics 606, Ed. I.Galliani, E.Magenes*, Springer-Verlag, Berlin, 1977, 292–315.

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

[49] T. Rusten and R. Winther. A preconditioned iterative method for saddlepoint problems, *SIAM J. Matrix Anal. Appl.*, 13 (1992), pp. 887–904.

[50] T. Rusten and R. Winther. Substructure preconditioners for elliptic saddle point problems, *Math. Comp.*, 60 (1993), pp. 23–48.

[51] Y. Saad. *Iterative Methods for Sparse Linear Systems.* PWS Publishing Company, ITP, 1996.

[52] Y. Saad and M. H. Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems, *SIAM J. Sci. Statist. Comput.*, 7 (1986), pp. 856–869.

[53] D. Silvester and A. Wathen. Fast iterative solution of stabilized Stokes systems part I: using diagonal block preconditioners, *SIAM J. Numer. Anal.*, 30 (1993), pp. 630–649.

[54] D. Silvester and A. Wathen. Fast iterative solution of stabilised Stokes systems part II: Using general block preconditioners, *SIAM J. Numer. Anal.*, 31 (1994), pp. 1352–1367.

[55] J. Stoer and R. Freund. On the solution of large indefinite systems of linear equations by conjugate gradients algorithm, *in Computing Methods in Applied Sciences and Engineering V, R. Glowinski, J.L. Lions eds.*, North Holland - INRIA, pp. 35–53, 1982.

[56] M. Tůma. A note on the LDL^T decomposition of matrices from saddle point problems, *SIAM J. Matrix Anal. Appl.* 23 (2002), pp. 903–915.

[57] H.F. Wang and M.P. Anderson. *Introduction to Groundwater Modelling, Finite Difference and Finite Element Methods.* W.H. Freeman and Company, San Francisco, 1982.

[58] A.J. Wathen, B. Fischer and D.J. Silvester. The convergence of iterative solution methods for symmetric and indefinite linear systems. *in Numerical analysis 1997, D.F. Griffiths, D.J. Higham and G.A. Watson (eds.)*, Pitman Research Notes in Mathematics Series 380, Longman, 230-242.

[59] A.J. Wathen, B. Fischer and D.J. Silvester. *The convergence rate of the minimal residual method for the Stokes problem*, *Num. Mathematik* 71 (1995), pp. 121–134.

[60] R. Weiss. *Parameter-Free Iterative Linear Solvers* Mathematical Research Series Vol. 97, Akademie Verlag, Berlin, 1996.

[61] M.F. Wheeler and R. Gonzales. Mixed Finite Element Methods for Petroleum Reservoir Engineering Problems. *Proceedings of Sixth International Conference on Computing Methods in Applied Sciences and Engineering, INRIA, Versailles, France (1983), Computing Methods in Engineering and Applied Sciences VI*, North Holland, Amsterdam, pp. 639-658, 1984.

[62] L. Zhou and H. F. Walker. Residual smoothing techniques for iterative methods, *SIAM J. Sci. Comput.*, 15 (1994), pp. 297–312.