

# Numerical behavior of saddle point solvers

Miro Rozložník, Pavel Jiránek

Institute of Computer Science, Czech Academy of Sciences, Prague  
and Technical University of Liberec, Czech Republic

Advances and perspectives on numerical methods for saddle point problems, Banff international research station, Canada, April 12-17, 2008

## Saddle point problems

We consider a saddle point problem with the symmetric  $2 \times 2$  block form

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

- ▶  $A$  is a square  $n \times n$  nonsingular (symmetric positive definite) matrix,
- ▶  $B$  is a rectangular  $n \times m$  matrix of (full column) rank  $m$ .

Applications: mixed finite element approximations, weighted least squares, constrained optimization etc. [Benzi, Golub, and Liesen, 2005].

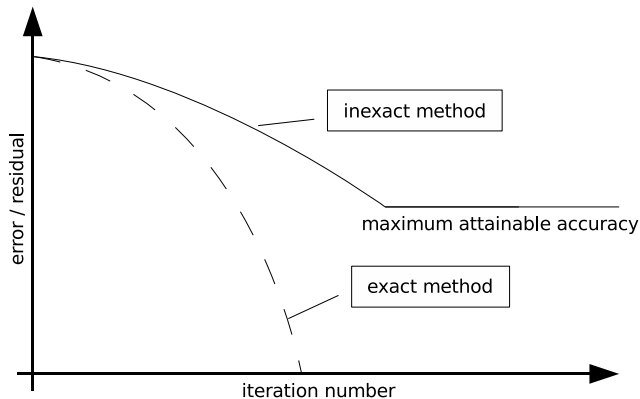
## Inexact saddle point solvers

Solution algorithms for saddle point problems:

1. **the segregated or coupled approach:** outer iteration for solving the reduced system;
2. **the inexact solution of inner systems:** inner iteration loop with appropriate stopping criterion;
3. **the rounding errors:** finite precision arithmetic.

Numerous schemes: inexact Uzawa algorithms, inexact null-space methods, inner-outer iteration methods, two-stage iteration processes, multilevel or multigrid methods, domain decomposition methods  
[Elman and Golub, 1994], [Bramble, Pasciak, and Vassilev, 2000], [Zulehner, 2002], [Braess, Deuffhard, and Lipnikov, 2002],...

## Delay of convergence and limit on the final accuracy



## Schur complement reduction method

- ▶ Compute  $y$  as a solution of the Schur complement system

$$B^T A^{-1} B y = B^T A^{-1} f,$$

- ▶ compute  $x$  as a solution of

$$A x = f - B y.$$

- ▶ Segregated vs. coupled approach:  $x_k$  and  $y_k$  approximate solutions to  $x$  and  $y$ , respectively.

## Iterative solution of the Schur complement system

choose  $y_0$ , solve  $Ax_0 = f - By_0$

compute  $\alpha_k$  and  $p_k^{(y)}$

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

$$\left. \begin{array}{l} \text{solve } Ap_k^{(x)} = -Bp_k^{(y)} \end{array} \right\}$$

**back-substitution:**

$$\mathbf{A}: x_{k+1} = x_k + \alpha_k p_k^{(x)},$$

$$\mathbf{B}: \text{solve } Ax_{k+1} = f - By_{k+1},$$

$$\mathbf{C}: \text{solve } Au_k = f - Ax_k - By_{k+1},$$

$$x_{k+1} = x_k + u_k.$$

inner  
iteration

outer  
iteration

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

## Accuracy in solving the inner systems

- ▶ Inexact solution of systems with  $A$ : **every computed solution  $\hat{u}$  of  $Au = b$  is interpreted an exact solution of a perturbed system**

$$(A + \Delta A)\hat{u} = b + \Delta b, \quad \|\Delta A\| \leq \tau\|A\|, \quad \|\Delta b\| \leq \tau\|b\|, \quad \tau\kappa(A) \ll 1.$$

- ▶ Schur complement system solved with two-term recursion: if  $\|fl(B^T A^{-1} Bx) - B^T A^{-1} Bx\| \leq O(u)\|A^{-1}\| \|B\|^2 \|x\|$  the theory of A. Greenbaum could be applied, but in our 'idealized' case we have  $fl(B^T A^{-1} Bx) = B^T (A + \Delta A)^{-1} Bx$ ,  $\|\Delta A\| \leq \tau\|A\|$  leading to

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

- ▶ Variable tolerance for solving inner systems based on the relative residual  $\|b - A\hat{u}\|/\|b\| \leq \tau$ , where  $b = Bx$  gives

$$\|fl(B^T A^{-1} Bx) - B^T A^{-1} Bx\| \leq \tau \|A^{-1}\| \|B\|^2 \|x\|$$

which can be seen as a **floating point iteration process** for the Schur complement system with the **roundoff unit equal to  $\tau$** .

# Maximum attainable accuracy of inexact Schur complement schemes

The limiting (maximum attainable) accuracy is measured by the ultimate (asymptotic) values of:

1. **the Schur complement residual:**  $B^T A^{-1} f - B^T A^{-1} B y_k$ ;
2. **the residuals in the saddle point system:**  $f - A x_k - B y_k$  and  $-B^T x_k$ ;
3. **the forward errors:**  $x - x_k$  and  $y - y_k$ .

## Numerical experiments: a small model example

$$A = \text{tridiag}(1, 4, 1) \in \mathbb{R}^{100 \times 100}, \quad B = \text{rand}(100, 20), \quad f = \text{rand}(100, 1),$$

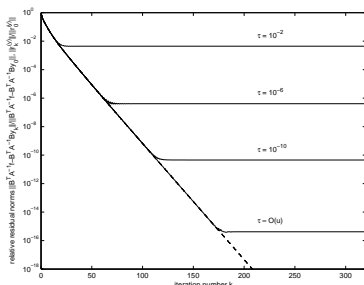
$$\kappa(A) = \|A\| \cdot \|A^{-1}\| = 7.1695 \cdot 0.4603 \approx 3.3001,$$

$$\kappa(B) = \|B\| \cdot \|B^\dagger\| = 5.9990 \cdot 0.4998 \approx 2.9983.$$

## Accuracy in the outer iteration process

$$\| -B^T A^{-1} f + B^T A^{-1} B y_k - r_k^{(y)} \| \leq \frac{O(\tau) \kappa(A)}{1 - \tau \kappa(A)} \|A^{-1}\| \|B\| (\|f\| + \|B\| Y_k).$$

$$Y_k \equiv \max\{\|y_i\| \mid i = 0, 1, \dots, k\}.$$



$$B^T (A + \Delta A)^{-1} B \hat{y} = B^T (A + \Delta A)^{-1} f,$$

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

## Does the final accuracy depend on the method used in the outer iteration?

- ▶ According to A. Greenbaum, the gap between the true and updated residuals for the two-term recurrence methods depends is proportional to the maximum norm of the approximate solutions over the whole iteration process.

$$\| -B^T A^{-1} f + B^T A^{-1} B y_k - r_k^{(y)} \| \leq \frac{O(\tau)\kappa(A)}{1 - \tau\kappa(A)} \|A^{-1}\| \|B\| (\|f\| + \|B\| Y_k).$$

$$Y_k \equiv \max\{\|y_i\| \mid i = 0, 1, \dots, k\}.$$

- ▶ The Schur complement system is negative definite, the norm of the error or residual converges monotonically for the most iterative methods. The quantity  $Y_k$  in the bounds then does not play an important role and it can be replaced by  $\|y_0\|$  or a small multiple of  $\|y\|$ .
- ▶ Our concept is similar to the general framework of inexact Krylov methods due to Simoncini, Szyld, van den Eshof and Sleijpen, but we take into account also the effects of associated rounding errors. For the outer iteration process we consider a general class of iterative methods based on two-term recurrences.

## Accuracy in the saddle point system

$$\|f - Ax_k - By_k\| \leq \frac{O(\alpha_1)\kappa(A)}{1 - \tau\kappa(A)} (\|f\| + \|B\|Y_k),$$

$$\| -B^T x_k - r_k^{(y)} \| \leq \frac{O(\alpha_2)\kappa(A)}{1 - \tau\kappa(A)} \|A^{-1}\| \|B\| (\|f\| + \|B\|Y_k),$$

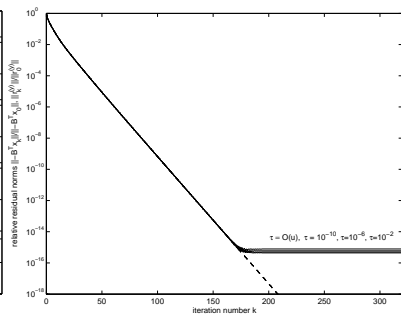
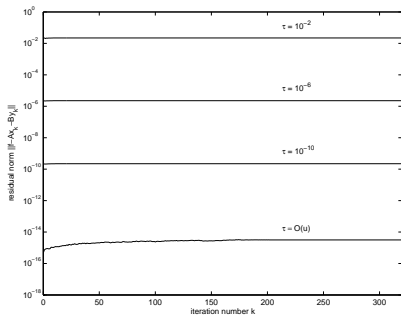
$$Y_k \equiv \max\{\|y_i\| \mid i = 0, 1, \dots, k\}.$$

Back-substitution scheme	$\alpha_1$	$\alpha_2$
<b>A:</b> Generic update $x_{k+1} = x_k + \alpha_k p_k^{(x)}$	$\tau$	$u$
<b>B:</b> Direct substitution $x_{k+1} = A^{-1}(f - By_{k+1})$	$\tau$	$\tau$
<b>C:</b> Corrected dir. subst. $x_{k+1} = x_k + A^{-1}(f - Ax_k - By_{k+1})$	$u$	$\tau$

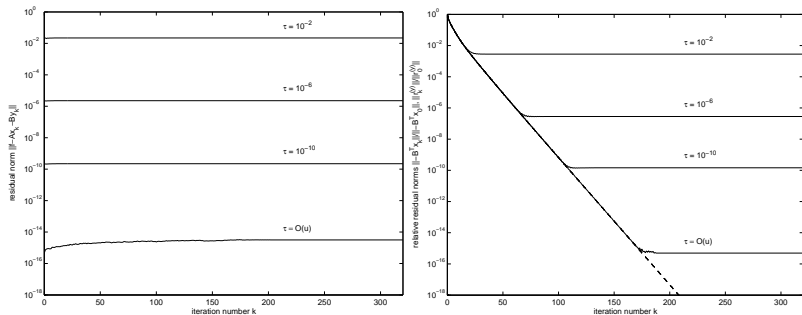
} additional system with A

$$-B^T A^{-1} f + B^T A^{-1} B y_k = -B^T x_k - B^T A^{-1} (f - A x_k - B y_k)$$

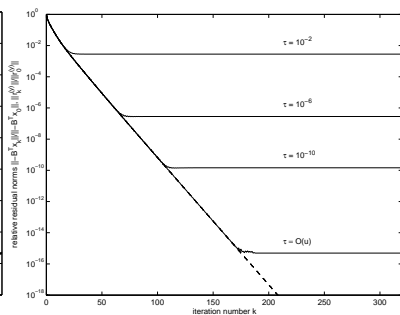
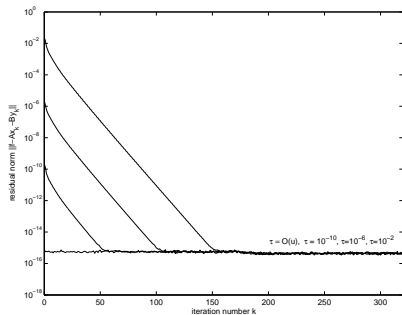
Generic update:  $x_{k+1} = x_k + \alpha_k p_k(x)$



Direct substitution:  $x_{k+1} = A^{-1}(f - By_{k+1})$



Corrected direct substitution:  $x_{k+1} = x_k + A^{-1}(f - Ax_k - By_{k+1})$

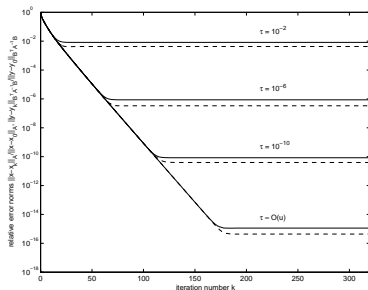


## Forward error of computed approximate solution

$$\|x - x_k\| \leq \gamma_1 \|f - Ax_k - By_k\| + \gamma_2 \| - B^T x_k \|,$$

$$\|y - y_k\| \leq \gamma_2 \|f - Ax_k - By_k\| + \gamma_3 \| - B^T x_k \|,$$

$$\gamma_1 = \sigma_{\min}^{-1}(A), \quad \gamma_2 = \sigma_{\min}^{-1}(B), \quad \gamma_3 = \sigma_{\min}^{-1}(B^T A^{-1} B).$$



## Null-space projection method

Analogous results for schemes, where the least squares with  $B$  are solved inexactly. **Again, every computed approximate solution of a least squares problem with  $B$  is interpreted as an exact solution of a perturbed least squares**

choose  $x_0$ , solve  $By_0 \approx f - Ax_0$

compute  $\alpha_k$  and  $p_k^{(x)} \in N(B^T)$

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

solve  $Bp_k^{(y)} \approx r_k^{(x)} - \alpha_k Ap_k^{(x)}$

**back-substitution:**

**A:**  $y_{k+1} = y_k + p_k^{(y)}$ ,

**B:** solve  $By_{k+1} \approx f - Ax_{k+1}$ ,

**C:** solve  $Bv_k \approx f - Ax_{k+1} - By_k$ ,

$$y_{k+1} = y_k + v_k.$$

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

} inner  
iteration

} outer  
iteration

## Null-space projection method

- ▶ compute  $x \in N(B^T)$  as a solution of the projected system

$$(I - \Pi)A(I - \Pi)x = (I - \Pi)f,$$

- ▶ compute  $y$  as a solution of the least squares problem

$$By \approx f - Ax,$$

$\Pi$  is the orthogonal projector onto  $R(B)$ .

The least squares with  $B$  are solved inexactly, i.e. the computed solution  $\bar{v}$  of  $Bv \approx c$  is an exact solution of a perturbed least squares problem

$$(B + \Delta B)\bar{v} \approx c + \Delta c, \quad \|\Delta B\| \leq \tau\|B\|, \quad \|\Delta c\| \leq \tau\|c\|, \quad \tau\kappa(B) \ll 1.$$

## Accuracy in the saddle point system

$$\|f - Ax_k - By_k - r_k^{(x)}\| \leq \frac{O(\alpha_3)\kappa(B)}{1 - \tau\kappa(B)} (\|f\| + \|A\|X_k),$$

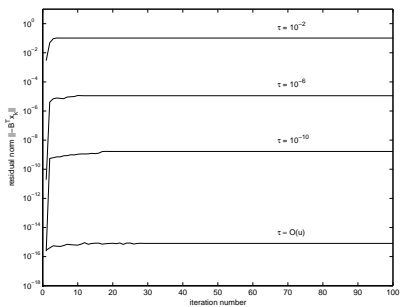
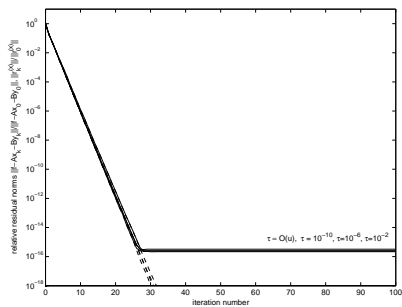
$$\| -B^T x_k \| \leq \frac{O(\tau)\kappa(B)}{1 - \tau\kappa(B)} \|B\|X_k,$$

$$X_k \equiv \max\{\|x_i\| \mid i = 0, 1, \dots, k\}.$$

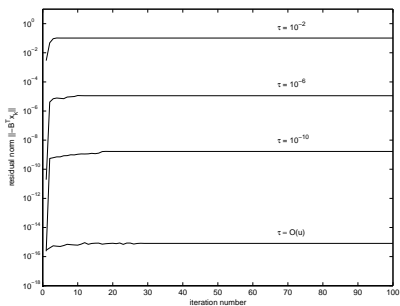
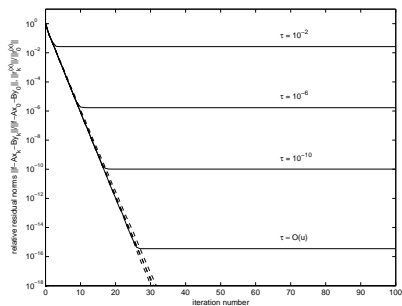
Back-substitution scheme	$\alpha_3$
<b>A:</b> Generic update $y_{k+1} = y_k + p_k^{(y)}$	$u$
<b>B:</b> Direct substitution $y_{k+1} = B^\dagger(f - Ax_{k+1})$	$\tau$
<b>C:</b> Corrected dir. subst. $y_{k+1} = y_k + B^\dagger(f - Ax_{k+1} - By_k)$	$u$

} additional least square with B

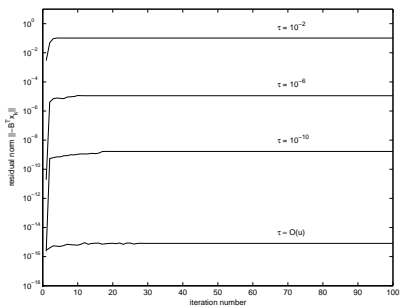
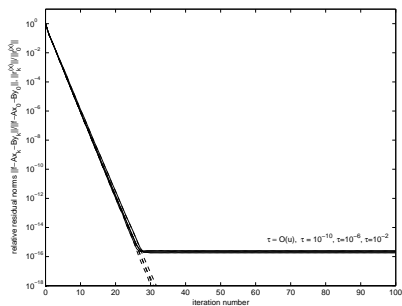
Generic update:  $y_{k+1} = y_k + p_k^{(y)}$



Direct substitution:  $y_{k+1} = B^\dagger(f - Ax_{k+1})$

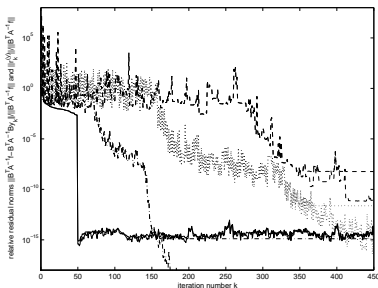


Corrected direct substitution:  $y_{k+1} = y_k + B^\dagger(f - Ax_{k+1} - By_k)$



# Conclusions

- ▶ The accuracy measured by the residuals of the saddle point problem depends on the choice of the back-substitution scheme [Jiránek, R, 2008]. The schemes with (generic or corrected substitution) updates deliver approximate solutions which satisfy either the first or second block equation to working accuracy.
- ▶ Care must be taken when solving nonsymmetric systems [Jiránek, R, 2007], all bounds of the limiting accuracy depend on the maximum norm of computed iterates, cf. [Greenbaum, 1997].



## Related results in the context of saddle-point problems and Krylov subspace methods

- ▶ General framework of inexact Krylov subspace methods: in exact arithmetic the effects of relaxation in matrix-vector multiplication on the ultimate accuracy of several solvers [Simoncini and Szyld, 2003], [van den Eshof and Sleijpen, 2004].
- ▶ The effects of rounding errors in the Schur complement reduction (block LU decomposition) method and the null-space method [Demmel, Higham, and Schreiber, 1995], [Arioli, 2000], the maximum attainable accuracy studied in terms of the user tolerance specified in the outer iteration [Arioli and Baldini, 2001], [Maryška, R, Tůma, 2000].
- ▶ Error analysis in computing the projections into the null-space and constraint preconditioning, limiting accuracy of the preconditioned CG [R, Simoncini, 2002], residual update strategy when solving constrained quadratic programming problems [Gould, Hribar, and Nocedal, 2001], or in cascadic multigrid method for elliptic problems [Braess, Deuffhard, and Lipnikov, 2002].
- ▶ Theory for a general class of iterative methods based on coupled two-term recursions, all bounds of the limiting accuracy depend on the maximum norm of computed iterates, fixed matrix-vector multiplication, cf. [Greenbaum, 1997].

**"new\_value = old\_value + small\_correction"**

- ▶ Fixed-precision iterative refinement for improving the computed solution  $x_{\text{old}}$  to a system  $Ax = b$ : solving update equations  $Az_{\text{corr}} = r$  that have residual  $r = b - Ay_{\text{old}}$  as a right-hand side to obtain  $x_{\text{new}} = x_{\text{old}} + z_{\text{corr}}$ , see [Wilkinson, 1963], [Higham, 1996].
- ▶ Stationary iterative methods for  $Ax = b$  and their maximum attainable accuracy [Higham and Knight, 1993]: assuming splitting  $A = M - N$  and inexact solution of systems with  $M$ , use  $x_{\text{new}} = x_{\text{old}} + M^{-1}(b - Ax_{\text{old}})$  rather than  $x_{\text{new}} = M^{-1}(Nx_{\text{old}} + b)$ , [Higham, 1996].
- ▶ Two-step splitting iteration framework:  $A = M_1 - N_1 = M_2 - N_2$  assuming inexact solution of systems with  $M_1$  and  $M_2$ , reformulation of  $M_1x_{1/2} = N_1x_{\text{old}} + b$ ,  $M_2x_{\text{new}} = N_2x_{1/2} + b$ , Hermitian/skew-Hermitian splitting (HSS) iteration [Bai, Golub, and Ng, 2003].
- ▶ Inexact preconditioners for saddle point problems: SIMPLE and SIMPLE(R) type algorithms [Vuik and Saghiri, 2002] and constraint preconditioners [Benzi and Golub, 2004].

# Thank you for your attention.

<http://www.cs.cas.cz/~miro>

P. Jiránek and M. Rozložník. Maximum attainable accuracy of inexact saddle point solvers. *SIAM J. Matrix Anal. Appl.*, 29(4):1297–1321, 2008.

P. Jiránek and M. Rozložník. Limiting accuracy of segregated solution methods for nonsymmetric saddle point problems. *J. Comput. Appl. Math.* 215 (2008), pp. 28-37.

## References I

- M. Arioli. The use of QR factorization in sparse quadratic programming and backward error issues. *SIAM J. Matrix Anal. Appl.*, 21(3):825–839, 2000.
- M. Arioli and L. Baldini. A backward error analysis of a null space algorithm in sparse quadratic programming. *SIAM J. Matrix Anal. Appl.*, 23(2):425–442, 2001.
- Z. Bai, G. H. Golub, and M. Ng. Hermitian and skew-hermitian splitting methods for non-hermitian positive definite linear systems. *SIAM J. Matrix Anal. Appl.*, 24:603–626, 2003.
- M. Benzi and G. H. Golub. A preconditioner for generalized saddle point problems. *SIAM J. Matrix Anal. Appl.*, 26:20–41, 2004.
- M. Benzi, G. H. Golub, and J. Liesen. Numerical solution of saddle point problems. *Acta Numer.*, 14:1–137, 2005.
- D. Braess, P. Deufllhard, and K. Lipnikov. A subspace cascadic multigrid method for mortar elements. *Computing*, 69(3):205–225, 2002.
- J. H. Bramble, J. E. Pasciak, and A. T. Vassilev. Inexact Uzawa algorithms for nonsymmetric saddle point problems. *Math. Comp.*, 69:667–689, 2000.
- J. W. Demmel, N. J. Higham, and R. S. Schreiber. Stability of the block LU factorization. *Numer. Linear Algebra Appl.*, 2(2):173–190, 1995.

## References II

- H. C. Elman and G. H. Golub. Inexact and preconditioned Uzawa algorithms for saddle point problems. *SIAM J. Numer. Anal.*, 31(6):1645–1661, 1994.
- 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(4):1376–1395, 2001.
- A. Greenbaum. Estimating the attainable accuracy of recursively computed residual methods. *SIAM J. Matrix Anal. Appl.*, 18(3):535–551, 1997.
- N. J. Higham. *Accuracy and Stability of Numerical Algorithms*. SIAM, Philadelphia, 1996.
- N. J. Higham and P. A. Knight. Componentwise error analysis for stationary iterative methods. In C. D. Meyer and R. J. Plemmons, editors, *Linear Algebra, Markov Chains, and Queueing Models*, volume 48 of *IMA Volumes in Mathematics and Its Applications*, pages 29–46, 1993.
- P. Jiránek and M. Rozložník. Maximum attainable accuracy of inexact saddle point solvers. *SIAM J. Matrix Anal. Appl.*, 29(4):1297–1321, 2008a.
- P. Jiránek and M. Rozložník. Limiting accuracy of segregated solution methods for nonsymmetric saddle point problems. *J. Comput. Appl. Math.*, 215: 28–37, 2008b.

## References III

- 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.*, 117:159–173, 2000.
- M. Rozložník and V. Simoncini. Krylov subspace methods for saddle point problems with indefinite preconditioning. *SIAM J. Matrix Anal. Appl.*, 24(2):368–391, 2002.
- V. Simoncini and D. B. Szyld. Theory of inexact Krylov subspace methods and applications to scientific computing. *SIAM J. Sci. Comput.*, 25(2):454–477, 2003.
- J. van den Eshof and G. L. G. Sleijpen. Inexact Krylov subspace methods for linear systems. *SIAM J. Matrix Anal. Appl.*, 26(1):125–153, 2004.
- C. Vuik and A. Saghir. The krylov accelerated simple(r) method for incompressible flow. Technical Report 02-01, Delft University of Technology, 2002.
- J. H. Wilkinson. *Rounding Errors in Algebraic Processes*. Prentice Hall, Inc., New Jersey, 1963.
- W. Zulehner. Analysis of iterative methods for saddle point problems: a unified approach. *Math. Comp.*, 71(238):479–505, 2002.