Numeric behavior of saddle point solvers

Miro Rozložník
joint results with Pavel Jiránek and Valeria Simoncini

Institute of Computer Science, Czech Academy of Sciences,
Prague, Czech Republic

The 4-th IMACS conference on Mathematical modelling and computational methods in applied sciences and engineering - MODELLING 2009,
Rožnov pod Radhoštěm, Czech Republic, June 22-26, 2009

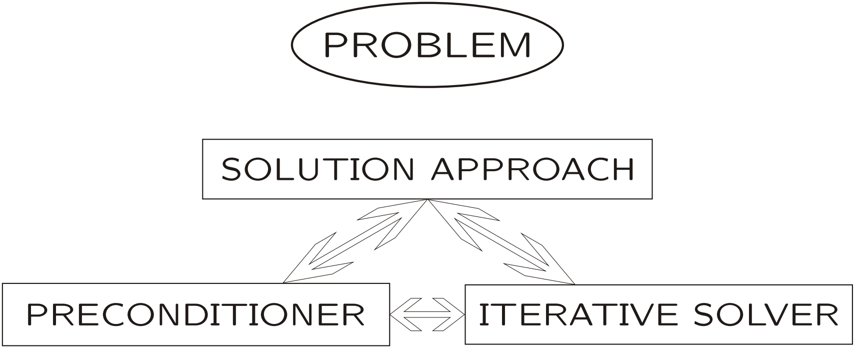
Saddle point problems

We consider a saddle point problem with the symmetric 2×2 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, computational fluid dynamics, electromagnetism etc. [Benzi, Golub and Liesen, 2005]. For the updated list of applications leading to saddle point problems contact [Benzi, 2009].


PROBLEM
SOLUTION APPROACH
PRECONDITIONER
ITERATIVE SOLVER

Segregated or coupled solution approach

1. **Schur complement or null-space projection approach**: outer iteration for solving the reduced system;
2. **inexact solution of inner systems:** inner iteration loop with appropriate stopping criterion;

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

Preconditioning and preconditioners

1. **preconditioning**: iteration scheme for solving the preconditioned system;
2. **approximate or incomplete factorization scheme:** structure-based or with appropriate dropping criterion;

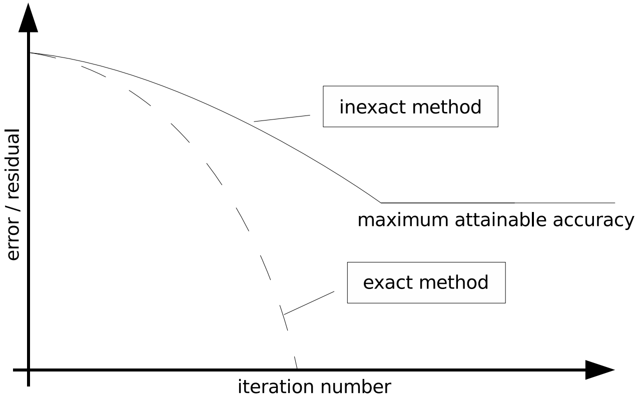
Numerous techniques: block diagonal preconditioners, block triangular preconditioners, constraint preconditioning, Hermitian/skew-Hermitian preconditioning and other splittings, combination preconditioning
References: [Bramble and Pasciak, 1988], [Silvester and Wathen, Wathen and Silvester 1993, 1994], [Elman, Silvester and Wathen, 2002, 2005], [Kay, Loghin and Wathen, 2002], [Perugia, Simoncini, Arioli, 1999], [Keller, Gould and Wathen 2000], [Gould, Hribar and Nocedal, 2001], [Stoll, Wathen, 2008], ...

Exact and finite precision arithmetic

1. **iterative method**: finite termination property, theoretical rate of convergence;
2. **the rounding errors in floating point arithmetic**: numerical stability of the solver

Numerous iterative solvers: conjugate gradient (CG) method, MINRES, GMRES, flexible GMRES, GCR, BiCG, BiCGSTAB, ...
References: [Hestenes and Stiefel, 1952], [Paige and Saunders, 1975], [Saad and Schultz, 1986], [Elman, 1982], [Lanczos 1950], [Fletcher 1976], [van der Vorst 1992], [Paige, 1976], [Greenbaum and Strakoš 1991, 1992], [Greenbaum, Paige, R., Strakoš 1995, 1997, 2006], [Modersitzki, Sleijpen and van der Vorst, 1997], [Gutknecht, Jiránek, R, 2008], ...

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

$$Ax = f - By.$$

- Segregated vs. coupled approach: x_k and y_k approximate solutions to x and y, respectively.
- Inexact solution of systems with A: **every computed solution $\hat{u}$ of $Au = b$ is interpreted an exact solution of a perturbed system**

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

Iterative solution of the Schur complement system

$$\left.\begin{array}{l}
\text{choose } y_0,\ \text{solve } Ax_0 = f - By_0 \\
\text{compute } \alpha_k \text{ and } p_k^{(y)} \\
y_{k+1} = y_k + \alpha_k p_k^{(y)} \\
\qquad \left|\begin{array}{l}
\text{solve } Ap_k^{(x)} = -Bp_k^{(y)} \\
\textcolor{red}{\text{back-substitution:}} \\
\textbf{A: } x_{k+1} = x_k + \alpha_k p_k^{(x)}, \\
\textbf{B: } \text{solve } Ax_{k+1} = f - By_{k+1}, \\
\textbf{C: } \text{solve } Au_k = f - Ax_k - By_{k+1}, \\
\qquad x_{k+1} = x_k + u_k.
\end{array}\right\} \begin{array}{l}\text{inner}\\ \text{iteration}\end{array} \\
r_{k+1}^{(y)} = r_k^{(y)} - \alpha_k B^T p_k^{(x)}
\end{array}\right\} \begin{array}{l}\text{outer}\\ \text{iteration}\end{array}$$

Numerical experiments: a small model example

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

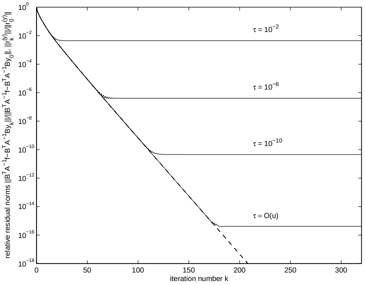
$$\kappa(A) = \|A\| \cdot \|A^{-1}\| = 7.1695 \cdot 0.4603 \approx 3.3001,$$

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

Accuracy in outer iteration

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

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



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

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

Does the final accuracy depend on the outer iteration method?

- Gap between the true and updated residual for any two-term recurrence method depends on the maximum norm of approximate solutions over the whole iteration process. [Greenbaum 1994, 1997]

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

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

- Schur complement system is negative definite, some norm of the error or residual converges monotonically for almost all iterative methods. The quantity Y_k then does not play an important role and it can be replaced by $\|y_0\|$ or a multiple of $\|y\|$.

Accuracy in the saddle point system

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

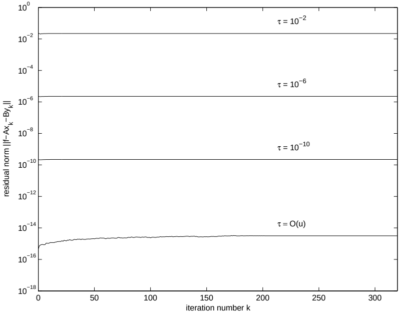
$$\| - B^T x_k - r_k^{(y)}\| \leq \frac{O(\alpha_2)\kappa(A)}{1 - \tau\kappa(A)}\|A^{-1}\|\|B\|(\|f\| + \|B\|Y_k),$$

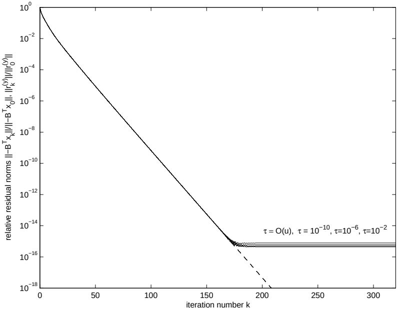
$$Y_k \equiv \max\{\|y_i\| \,|\, i = 0, 1, \dots, k\}.$$

Back-substitution scheme	α_1	α_2	
A: Generic update $x_{k+1} = x_k + \alpha_k p_k^{(x)}$	τ	u	
B: Direct substitution $x_{k+1} = A^{-1}(f - By_{k+1})$	τ	τ	additional system with A
C: Corrected dir. subst. $x_{k+1} = x_k + A^{-1}(f - Ax_k - By_{k+1})$	u	τ	additional system with A

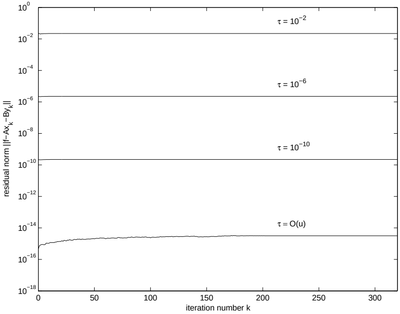
$$-B^T A^{-1} f + B^T A^{-1} B y_k = -B^T x_k - B^T A^{-1}(f - Ax_k - By_k)$$

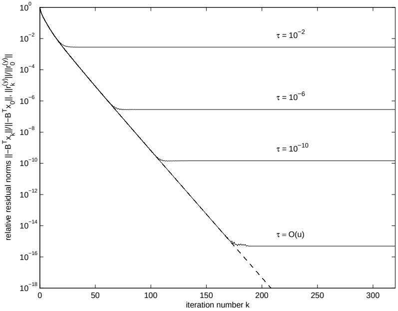
Generic update: $x_{k+1} = x_k + \alpha_k p_k^{(x)}$



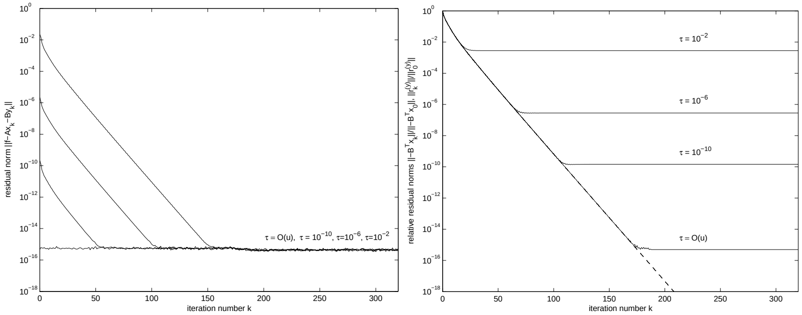


Direct substitution: $x_{k+1} = A^{-1}(f - By_{k+1})$





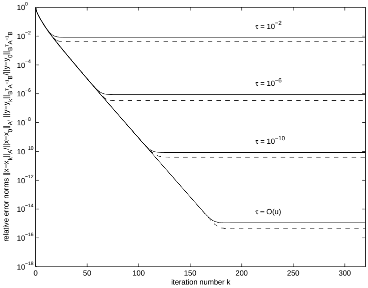
Corrected direct substitution: $x_{k+1} = x_k + A^{-1}(f - Ax_k - By_{k+1})$



Forward error of computed approximate solution

$$\|x - x_k\| \le \gamma_1 \|f - Ax_k - By_k\| + \gamma_2 \| - B^T x_k\|,$$
$$\|y - y_k\| \le \gamma_2 \|f - Ax_k - By_k\| + \gamma_3 \| - B^T x_k\|,$$

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



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

$$By \approx f - Ax,$$

  $\Pi = B(B^T B)^{-1} B^T$ is the orthogonal projector onto $R(B)$.

- Schemes with the inexact solution of least squares with B. Every computed approximate solution $\bar{v}$ of a least squares problem $Bv \approx c$ is interpreted as an exact solution of a perturbed least squares

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

Null-space projection method

$$
\left.\begin{array}{l}
\text{choose } x_0,\ \text{solve } By_0 \approx f - Ax_0\\[4pt]
\text{compute } \alpha_k \text{ and } p_k^{(x)} \in N(B^T)\\[4pt]
x_{k+1} = x_k + \alpha_k p_k^{(x)}\\[4pt]
\qquad \left|\begin{array}{l}
\text{solve } Bp_k^{(y)} \approx r_k^{(x)} - \alpha_k A p_k^{(x)}\\[4pt]
\color{red}{\text{back-substitution:}}\\[4pt]
\textbf{A: } y_{k+1} = y_k + p_k^{(y)},\\[4pt]
\textbf{B: } \text{solve } By_{k+1} \approx f - Ax_{k+1},\\[4pt]
\textbf{C: } \text{solve } Bv_k \approx f - Ax_{k+1} - By_k,\\[4pt]
\qquad y_{k+1} = y_k + v_k.
\end{array}\right\} \begin{array}{l}\text{inner}\\ \text{iteration}\end{array}\\[4pt]
r_{k+1}^{(x)} = r_k^{(x)} - \alpha_k A p_k^{(x)} - Bp_k^{(y)}
\end{array}\right\} \begin{array}{l}\text{outer}\\ \text{iteration}\end{array}
$$

Accuracy in the saddle point system

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

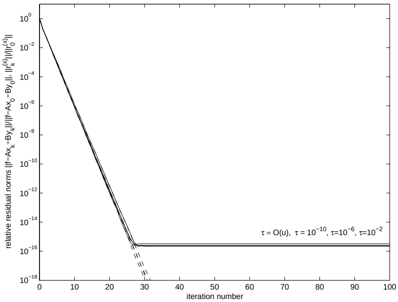
$$\| - B^T x_k\| \leq \frac{O(\tau)\kappa(B)}{1 - \tau\kappa(B)}\|B\|X_k,$$

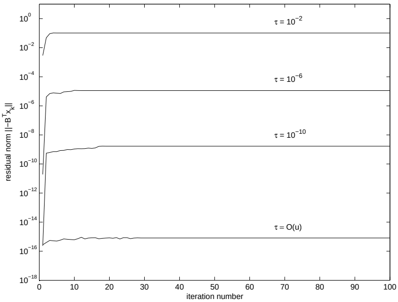
$$X_k \equiv \max\{\|x_i\| \,|\, i = 0, 1, \ldots, k\}.$$

Back-substitution scheme	α_3
A: Generic update $y_{k+1} = y_k + p_k^{(y)}$	u
B: Direct substitution $y_{k+1} = B^\dagger(f - Ax_{k+1})$	τ
C: Corrected dir. subst. $y_{k+1} = y_k + B^\dagger(f - Ax_{k+1} - By_k)$	u

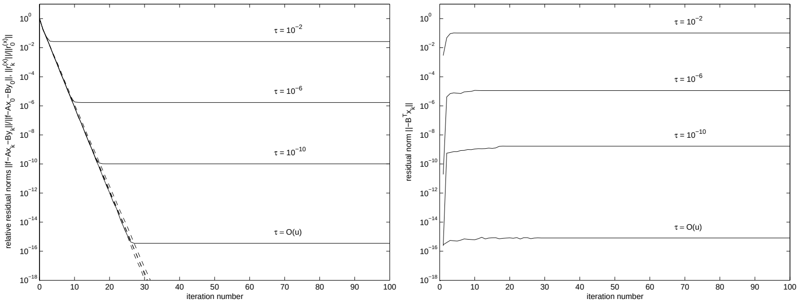
additional least square with B

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

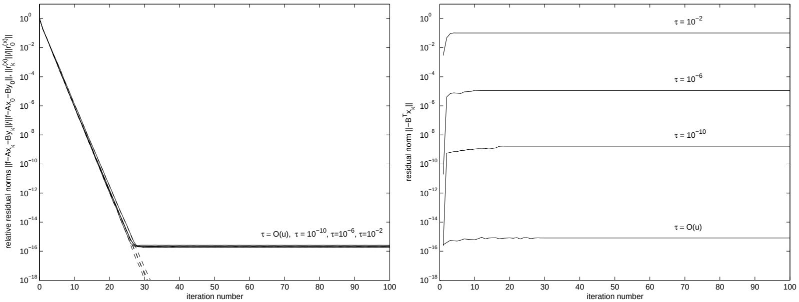




Direct substitution: $y_{k+1} = B^\dagger(f - Ax_{k+1})$



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



$\mathcal{A}$ symmetric indefinite, $\mathcal{P}$ positive definite

$$\mathcal{A} = \begin{pmatrix} A & B \\ B^T & 0 \end{pmatrix} \approx \mathcal{P} = \mathcal{R}^T\mathcal{R}$$

$$\left(\mathcal{R}^{-T}\mathcal{A}\mathcal{R}^{-1}\right)\mathcal{R}\begin{pmatrix} x \\ y \end{pmatrix} = \mathcal{R}^{-T}\begin{pmatrix} f \\ 0 \end{pmatrix}$$

$\mathcal{R}^{-T}\mathcal{A}\mathcal{R}^{-1}$ is symmetric indefinite!

Positive definite preconditioner: iterative solution of indefinite system

- Preconditioned MINRES is the MINRES on $\mathcal{R}^{-T}\mathcal{A}\mathcal{R}^{-1}$, minimizes the $\mathcal{P}^{-1} = \mathcal{R}^{-1}\mathcal{R}^{-T}$-norm of the residual on $K_n(\mathcal{P}^{-1}\mathcal{A}, \mathcal{P}^{-1}r_0)$
- CG applied to indefinite system with $\mathcal{R}^{-T}\mathcal{A}\mathcal{R}^{-1}$:
  CG iterate exists at least at every second step (tridiagonal form T_n is nonsingular at least at every second step)

  [Paige, Saunders, 1975]
- peak/plateau behavior:
  CG converges fast $\rightarrow$ MINRES is not much better than CG
  CG norm increases (peak) $\rightarrow$ MINRES stagnates (plateau)

  [Greenbaum, Cullum, 1996]

Symmetric indefinite or nonsymmetric preconditioner

$\mathcal{P}$ symmetric indefinite or nonsymmetric

$$\mathcal{P}^{-1}\mathcal{A}\begin{pmatrix} x \\ y \end{pmatrix} = \mathcal{P}^{-1}\begin{pmatrix} f \\ 0 \end{pmatrix}$$

$$\left(\mathcal{A}\mathcal{P}^{-1}\right)\mathcal{P}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} f \\ 0 \end{pmatrix}$$

$\mathcal{P}^{-1}\mathcal{A}$ and $\mathcal{A}\mathcal{P}^{-1}$ are nonsymmetric!

Indefinite preconditioner: iterative solution of nonsymmetric system

- The existence of a short-term recurrence solution methods to solve the system with $\mathcal{P}^{-1}\mathcal{A}$ or $\mathcal{A}\mathcal{P}^{-1}$ for arbitrary right-hand side vector
  [Faber, Manteuffel 1984, Liesen, Strakoš, 2006]
- Matrices $\mathcal{P}^{-1}\mathcal{A}$ or $\mathcal{A}\mathcal{P}^{-1}$ can be symmetric (self-adjoint) in a given bilinear form induced by **symmetric indefinite** $\mathcal{H}$ such that
  $\mathcal{H}(\mathcal{P}^{-1}\mathcal{A}) = (\mathcal{P}^{-1}\mathcal{A})^T\mathcal{H}$
  $\mathcal{H}(\mathcal{A}\mathcal{P}^{-1}) = (\mathcal{A}\mathcal{P}^{-1})^T\mathcal{H}$
- **symmetric indefinite preconditioner** $\mathcal{H} = \mathcal{P}^{-1} = (\mathcal{P}^{-1})^T$ so that
  $(\mathcal{P}^{-1})^T(\mathcal{P}^{-1})\mathcal{A} = \mathcal{A}(\mathcal{P}^{-1})^T(\mathcal{P}^{-1})$
  $(\mathcal{P}^{-1})^T\mathcal{A}\mathcal{P}^{-1} = \mathcal{P}^{-1}\mathcal{A}\mathcal{P}^{-1}$
- right vs left preconditioning for symmetric $\mathcal{P}$
  $$\mathcal{P}^{-1}K_n(\mathcal{A}\mathcal{P}^{-1}, r_0) = K_n(\mathcal{P}^{-1}\mathcal{A}, \mathcal{P}^{-1}r_0)$$
  $$(\mathcal{A}\mathcal{P}^{-1})^T = (\mathcal{P}^{-1})^T\mathcal{A} = \mathcal{P}^{-1}\mathcal{A}$$

Iterative solution of a system symmetric in the given bilinear form

- $\mathcal{H}$-symmetric variant of the nonsymmetric Lanczos process:
  $\mathcal{A}\mathcal{P}^{-1}V_n = V_{n+1}T_{n+1,n}$, $(\mathcal{A}\mathcal{P}^{-1})^T W_n = W_{n+1}\tilde{T}_{n+1,n}$
  $W_n^T V_n = I \Longrightarrow W_n = \mathcal{H}V_n$

  [Freund, Nachtigal, 1995]
- $\mathcal{H}$-symmetric variant of Bi-CG
  $\mathcal{H}$-symmetric variant of QMR $\equiv$ ITFQMR

  [Freund, Nachtigal, 1995]
- QMR-from-BiCG:
  $\mathcal{H}$-symmetric Bi-CG $+$ QMR-smoothing
  $\Longrightarrow$ $\mathcal{H}$-symmetric QMR

  [Freund, Nachtigal, 1995, Walker, Zhou 1994]
- peak/plateau behavior:
  QMR does not improve the convergence of Bi-CG (Bi-CG converges fast $\rightarrow$ QMR is not much better, Bi-CG norm increases $\rightarrow$ quasi-residual of QMR stagnates)

  [Greenbaum, Cullum, 1996]

Simplified Bi-CG algorithm is a preconditioned CG algorithm

$\mathcal{H} = \mathcal{P}^{-1}$-symmetric variant of two-term Bi-CG on $\mathcal{A}\mathcal{P}^{-1}$
is the CG algorithm on $\mathcal{A}$ preconditioned with $\mathcal{P}$

$\mathcal{P}^{-1}$-symmetric Bi-CG($\mathcal{A}\mathcal{P}^{-1}$)	PCG($\mathcal{A}$) with $\mathcal{P}^{-1}$
$\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$, $r_0 = b - \mathcal{A}\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$	
$\mathcal{P}^{-1}p_0 = \mathcal{P}^{-1}r_0$, $\tilde{p}_0 = \tilde{r}_0 = \mathcal{P}^{-1}p_0$	$z_0 = \mathcal{P}^{-1}r_0$
$k = 0, 1, \ldots$	
$\alpha_k = (r_k, \tilde{r}_k)/(\mathcal{A}\mathcal{P}^{-1}p_k, \tilde{p}_k)$	$\alpha_k = (r_k, z_k)/(\mathcal{A}\mathcal{P}^{-1}p_k, \mathcal{P}^{-1}p_k)$
$\begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} = \begin{pmatrix} x_k \\ y_k \end{pmatrix} + \alpha_k \mathcal{P}^{-1}p_k$	
$r_{k+1} = r_k - \alpha_k \mathcal{A}\mathcal{P}^{-1}p_k$	
$\tilde{r}_{k+1} = \mathcal{P}^{-1}r_{k+1}$	$z_{k+1} = \mathcal{P}^{-1}r_{k+1}$
$\beta_k = (r_{k+1}, \tilde{r}_{k+1})/(r_k, \tilde{r}_k)$	$\beta_k = (r_{k+1}, z_{k+1})/(r_k, z_k)$
$\mathcal{P}^{-1}p_{k+1} = \mathcal{P}^{-1}r_{k+1} + \beta_k \mathcal{P}^{-1}p_k$	$\mathcal{P}^{-1}p_{k+1} = z_{k+1} + \beta_k \mathcal{P}^{-1}p_k$
$\tilde{p}_{k+1} = \mathcal{P}^{-1}p_{k+1}$	

Saddle point problem and indefinite constraint preconditioner

$$\begin{pmatrix} A & B \\ B^T & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} f \\ 0 \end{pmatrix}, \quad \mathcal{P} = \begin{pmatrix} I & B \\ B^T & 0 \end{pmatrix}$$

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

$\Pi = B(B^T B)^{-1} B^T$ - orth. projector onto $span(B)$

[Lukšan, Vlček, 1998], [Gould, Keller, Wathen 2000]
[Perugia, Simoncini, Arioli, 1999], [R, Simoncini, 2002]

Indefinite constraint preconditioner: spectral properties

$\mathcal{A}\mathcal{P}^{-1}$ **nonsymmetric** and **non-diagonalizable**!
but it has a 'nice' spectrum:

$$\begin{aligned} \sigma(\mathcal{A}\mathcal{P}^{-1}) &\subset \{1\} \cup \sigma(A(I-\Pi)+\Pi) \\ &\subset \{1\} \cup \sigma((I-\Pi)A(I-\Pi)) - \{0\} \end{aligned}$$

and only 2 by 2 Jordan blocks!

[Lukšan, Vlček 1998], [Gould, Wathen, Keller, 1999], [Perugia, Simoncini 1999]

Krylov method with the constraint preconditioner: basic properties

$$\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}, r_0 = \begin{pmatrix} s_0 \\ 0 \end{pmatrix}, e_{k+1} = \begin{pmatrix} x - x_{k+1} \\ y - y_{k+1} \end{pmatrix}$$

$$r_{k+1} = \begin{pmatrix} f \\ 0 \end{pmatrix} - \begin{pmatrix} A & B \\ B^T & 0 \end{pmatrix} \begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix}$$

$$r_0 = \begin{pmatrix} s_0 \\ 0 \end{pmatrix} \Rightarrow r_{k+1} = \begin{pmatrix} s_{k+1} \\ 0 \end{pmatrix}$$
$$\Rightarrow B^T(x - x_{k+1}) = 0$$
$$\Rightarrow x_{k+1} \in Null(B^T)!$$

Preconditioned CG method: error norm

$$r_{k+1}^T \mathcal{P}^{-1} r_j = 0, \; j = 0, \ldots, k$$

x_{k+1} is an iterate from CG applied to

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

satisfying

$$\|x - x_{k+1}\|_A = \min_{u \in x_0 + span\{(I-\Pi)s_j\}} \|x - u\|_A$$

[Lukšan, Vlček 1998], [Gould, Wathen, Keller, 1999]

Preconditioned CG method: residual norm

$$\|x_{k+1} - x\| \to 0$$

but in general

$$y_{k+1} \not\to y$$

which is reflected in

$$\|r_{k+1}\| = \left\| \begin{pmatrix} s_{k+1} \\ 0 \end{pmatrix} \right\| \not\to 0!$$

but under appropriate scaling yes!

Preconditioned CG method: residual norm

$$x_{k+1} \to x$$

$$x - x_{k+1} = \phi_{k+1}((I - \Pi)A(I - \Pi))(x - x_0)$$

$$s_{k+1} = \phi_{k+1}(A(I - \Pi) + \Pi)s_0$$

$$\sigma((I - \Pi)A(I - \Pi)) \sim \sigma(A(I - \Pi) + \Pi)?$$

$$\{1\} \in \sigma((I - \Pi)\alpha A(I - \Pi)) - \{0\}$$
$$\Rightarrow \|r_{k+1}\| = \left\| \begin{pmatrix} s_{k+1} \\ 0 \end{pmatrix} \right\| \to 0!$$

How to avoid misconvergence?

- Scaling by a constant $\alpha > 0$ such that

$$\{1\} \in conv(\sigma((I-\Pi)\alpha A(I-\Pi)) - \{0\})$$

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

$$v : \quad \|(I-\Pi)v\| \neq 0, \quad \alpha = \frac{1}{((I-\Pi)v, A(I-\Pi)v)}!$$

- Scaling by a diagonal $A \rightarrow (diag(A))^{-1/2} A (diag(A))^{-1/2}$ often gives what we want!
- Different direction vector so that $\|r_{k+1}\| = \|s_{k+1}\|$ is locally minimized!

$$y_{k+1} = y_k + (B^T B)^{-1} B^T s_k$$

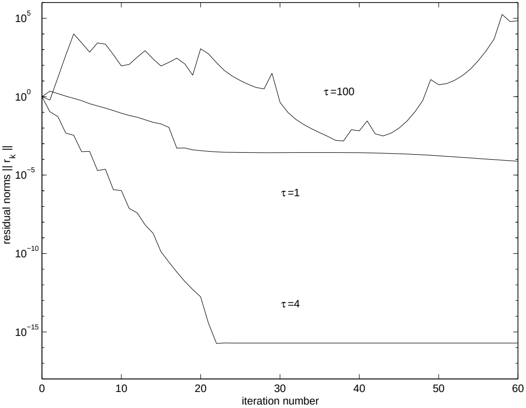
[Braess, Deuflhard,Lipikov 1999], [Hribar, Gould, Nocedal, 1999], [Jiránek, R, 2008]

Numerical experiments: a small model example

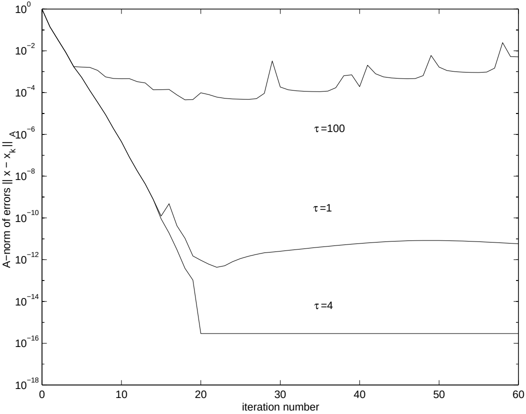
$$A = \text{tridiag}(1,4,1) \in \mathsf{R}^{25,25},\ B = \text{rand}(25,5) \in \mathsf{R}^{25,5}$$
$$f = \text{rand}(25,1) \in \mathsf{R}^{25}$$

$$\sigma(A) \subset [2.0146, 5.9854]$$

$\alpha = 1/\tau$	$\sigma(\begin{pmatrix} \alpha A & B \\ B^T & 0 \end{pmatrix} \begin{pmatrix} I & B \\ B^T & 0 \end{pmatrix}^{-1})$
$1/100$	$[0.0207, 0.0586] \cup \{1\}$
$1/10$	$[0.2067, 0.5856] \cup \{1\}$
$\mathbf{1/4}$	$[\mathbf{0.5170}, \mathbf{1.4641}]$
1	$\{1\} \cup [2.0678, 5.8563]$
4	$\{1\} \cup [8.2712, 23.4252]$



10^5
10^0
10^{-5}
10^{-10}
10^{-15}
residual norms || r_k ||
0
10
20
30
40
50
60
iteration number
τ =100
τ =1
τ =4



10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
10^{-14}
10^{-16}
10^{-18}
A−norm of errors $\| x - x_k \|_A$
$\tau = 100$
$\tau = 1$
$\tau = 4$
0
10
20
30
40
50
60
iteration number

Error norm of the computed approximate solution

Finite precision arithmetic:

$$\begin{pmatrix} \bar{x}_{k+1} \\ \bar{y}_{k+1} \end{pmatrix}, \quad \bar{r}_{k+1} = \begin{pmatrix} \bar{s}^{(1)}_{k+1} \\ \bar{s}^{(2)}_{k+1} \end{pmatrix} \to 0$$

$$\|x - \bar{x}_{k+1}\|_A^2 = (\Pi A(x - \bar{x}_{k+1}), \Pi(x - \bar{x}_{k+1})) + ((I - \Pi)A(x - \bar{x}_{k+1}), (I - \Pi)(x - \bar{x}_{k+1}))$$

$$\|x - \bar{x}_{k+1}\|_A \leq \gamma_1 \|\Pi(x - \bar{x}_{k+1})\| + \gamma_2 \|(I - \Pi)A(I - \Pi)(x - \bar{x}_{k+1})\|$$

Exact arithmetic:

$$\|\Pi(x - x_{k+1})\| = 0$$

$$\|(I - \Pi)A(I - \Pi)(x - x_{k+1})\| \to 0$$

Error norm of the computed approximate solution

departure from the null-space of B^T + projection of the residual onto it

$$\|x - \bar{x}_{k+1}\|_A \leq \gamma_3 \|B^T(x - \bar{x}_{k+1})\| + \gamma_2 \|(I - \Pi)(f - A\bar{x}_{k+1} - B\bar{y}_{k+1})\|$$

can be monitored by easily computable quantities:

$$B^T(x - \bar{x}_{k+1}) \sim \bar{s}_{k+1}^{(2)}$$

$$(I - \Pi)(f - A\bar{x}_{k+1} - B\bar{y}_{k+1}) \sim (I - \Pi)\bar{s}_{k+1}^{(1)}$$

Residuals: maximum attainable accuracy

$$\|(f - A\bar{x}_{k+1} - B\bar{y}_{k+1}) - \bar{s}^{(1)}_{k+1}\|, \; \|B^T(x - \bar{x}_{k+1}) - \bar{s}^{(2)}_{k+1}\| \leq$$

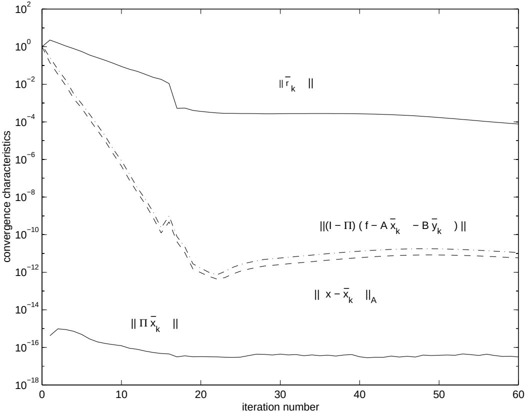
$$\leq \| \begin{pmatrix} f \\ 0 \end{pmatrix} - \begin{pmatrix} A & B \\ B^T & 0 \end{pmatrix} \begin{pmatrix} \bar{x}_{k+1} \\ \bar{y}_{k+1} \end{pmatrix} - \begin{pmatrix} \bar{s}^{(1)}_{k+1} \\ \bar{s}^{(2)}_{k+1} \end{pmatrix} \|$$

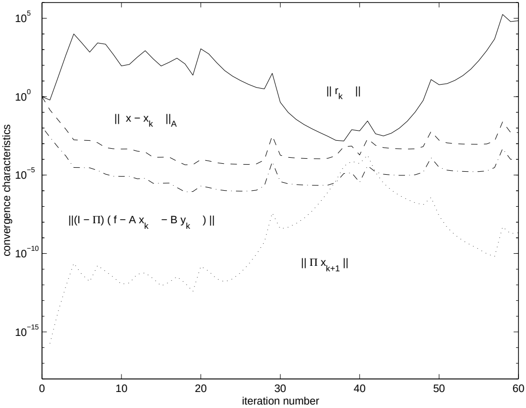
$$\leq c_1 \varepsilon \kappa(\mathcal{A}) \max_{j=0,\ldots,k+1} \|\bar{r}_j\|$$

[Greenbaum 1994,1997], [Sleijpen, et al. 1994]

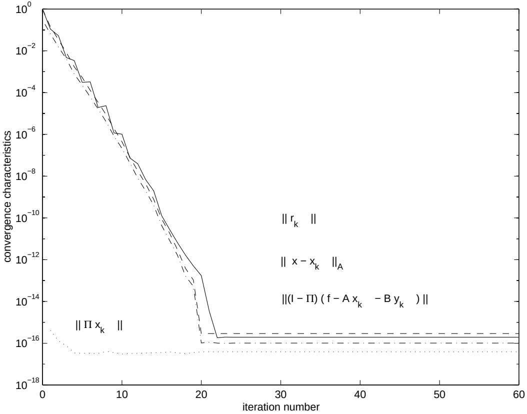
good scaling: $\|\bar{r}_j\| \longrightarrow 0$ nearly monotonically

$\|\bar{r}_0\| \sim \max_{j=0,\ldots,k+1} \|\bar{r}_j\|$

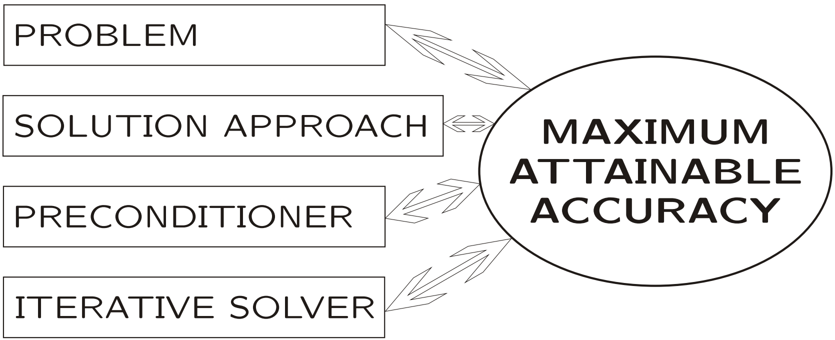

$\|\bar{r}_k\|$
$\|(I-\Pi)(f - A\bar{x}_k - B\bar{y}_k)\|$
$\|x - \bar{x}_k\|_A$
$\|\Pi\bar{x}_k\|$
convergence characteristics
iteration number



10^5
10^0
10^{-5}
10^{-10}
10^{-15}
0
10
20
30
40
50
60
iteration number
convergence characteristics
$\| r_k \|$
$\| x - x_k \|_A$
$\|(I - \Pi)(f - A x_k - B y_k)\|$
$\| \Pi x_{k+1} \|$

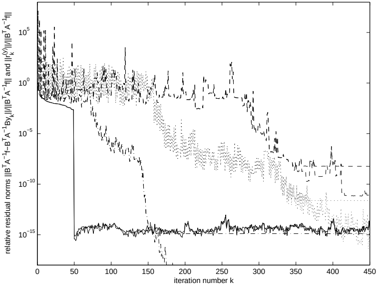


convergence characteristics
iteration number
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
10^{-14}
10^{-16}
10^{-18}
0
10
20
30
40
50
60
$\| r_k \|$
$\| x - x_k \|_A$
$\|(I - \Pi)(f - A x_k - B y_k)\|$
$\| \Pi x_k \|$


PROBLEM
SOLUTION APPROACH
PRECONDITIONER
ITERATIVE SOLVER
MAXIMUM ATTAINABLE ACCURACY

Conclusions: segregated solution approach

- 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, 2008], all bounds of the limiting accuracy depend on the maximum norm of computed iterates, cf. [Greenbaum 1994,1997], [Sleijpen, et al. 1994].



Conclusions: coupled approach with indefinite preconditioner

- Short-term recurrence methods are applicable for saddle point problems with indefinite preconditioning at a cost comparable to that of symmetric solvers. There is a tight connection between the simplified Bi-CG algorithm and the classical CG.
- The convergence of CG applied to saddle point problem with indefinite preconditioner for all right-hand side vectors is not guaranteed. For a particular set of right-hand sides the convergence can be achieved by the appropriate scaling of the saddle point problem.
- Since the maximum attainable accuracy depends heavily on the size of computed residuals, a good scaling of the problems leads to approximate solutions satisfying both two block equations to the working accuracy.

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.

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.*