Preconditioning of sequences of large, sparse and nonsymmetric linear systems

Jurjen Duintjer Tebbens

jointly with **Miroslav Tůma**

Institute of Computer Science

Academy of Sciences of the Czech Republic

email: tebbens,tuma@cs.cas.cz

GAMM annual meeting, Berlin, March 28, 2006

1. Introduction to preconditioner updates

Consider a sequence of linear systems

$$\mathbf{A}^{(i)}x = b^{(i)},\ i = 0, 1, \ldots,$$

where $\mathbf{A}^{(i)} \in \mathbb{R}^{n\times n}$ are nonsingular sparse matrices; $b^{(i)} \in \mathbb{R}^n$.

Applications: Computational fluid dynamics, structural mechanics, numerical optimization, etc . . . .

Classical example: A system of nonlinear equations $F(x) = 0$ for $F : \mathbb{R}^n \to \mathbb{R}^n$ solved by a Newton or Broyden-type method. In case of the classical Newton method

$$\mathbf{J}(x_i)(x_{i+1} - x_i) = -F(x_i),\ i = 1, \ldots,$$

where $\mathbf{J}(x_i)$ is the Jacobian evaluated in the current iteration x_i or its approximation.

There is a strong need for reduction of costs by sharing some of the computational effort among the subsequent linear systems. In the context of preconditioning:

Computing preconditioners $\mathbf{M}^{(0)}, \mathbf{M}^{(1)}, \ldots$ for individual systems separately, may be very expensive. A remedy is freezing the preconditioner: Using the same preconditioner for a sequence of linear systems (see, e.g [Brown, Saad - 1990]). This approach is very natural in the context of a matrix-free environment, where the system matrices $\mathbf{A}^{(i)}$ may be available only in the form of matrix-vector products, see also [Knoll, Keyes - 2004].

Freezing the preconditioner need not be enough. We may reuse some *additional* information from the previous linear systems. For example:

In the Newton-Krylov framework: Define preconditioners that incorporate recycled Krylov subspaces, see e.g. [Loghin, Ruiz, Touhami- 2004], [Parks, de Sturler, Mackey, Johnson, Maiti - 2004].

There is some recent work in approximate preconditioner updates with respect to the changes in the system matrix as well:

- Approximate diagonal updates of approximate inverse preconditioners for solving parabolic PDEs were proposed in [Benzi, Bertaccini - 2003], see also [Bertaccini - 2004].

- A straightforward approximate rank one update for a quasi-Newton method in the SPD case is described in [Morales, Nocedal - 2000], [Bergamaschi, Bru, Martinez, Putti - 2001].

We present new approaches to approximate updates of general nonsymmetric preconditioners which may be useful in solving subsequent linear systems.

We do not confine ourselves to particular classes of linear or nonlinear solvers (e.g. Krylov subspace methods, Quasi-Newton methods). We address the following 2 problems:

- How can we update, in theory, a preconditioner in such a way that the updated preconditioner is likely to be as powerful as the original one?

- How can we approximate, in practice, such an update in order to obtain a preconditioner that is inexpensive to compute and to apply?

Consider two linear systems denoted by

$$\mathbf{A}x = b \qquad \text{and} \qquad \mathbf{A}^+ x^+ = b^+.$$

Denote the difference matrix $\mathbf{A} - \mathbf{A}^+$ by $\mathbf{B}$.
Let $\mathbf{M}$ be a preconditioner approximating $\mathbf{A}$. We have

$$\|\mathbf{A} - \mathbf{M}\| = \|\mathbf{A} - \mathbf{M} + \mathbf{A}^+ - \mathbf{A}^+\| = \|\mathbf{A}^+ - (\mathbf{M} - \mathbf{B})\|,$$

hence $\mathbf{M}^+ \equiv \mathbf{M} - \mathbf{B}$ is an updated preconditioner for $\mathbf{A}^+$ of the same "level" of accuracy as $\mathbf{M}$ is for $\mathbf{A}$.

This "ideal" updated preconditioner cannot be used, in general, in practice since multiplication of vectors with $(\mathbf{M} - \mathbf{B})^{-1}$ may be too expensive.

There are ways, however, to approximate multiplication with $(\mathbf{M} - \mathbf{B})^{-1}$, as we will now show.

2. Proposed sparse preconditioner updates

Assume $\mathbf{M} = \mathbf{LDU} \approx \mathbf{A}$, where $\mathbf{L}$ and $\mathbf{U}$ are lower, resp. upper triangular and have unit main diagonal.

Our approximations of the ideal update are based on the idea that often the entries of $\mathbf{L}$ and $\mathbf{U}$ decay when moving away from the main diagonal, see e.g. [Benzi, Tůma - 2000], [Benzi, Bertaccini - 2003]. Sufficient diagonal dominance may also be imposed if $\mathbf{A}$ contains a strong transversal [Olschowka, Neumaier - 1996], [Duff, Koster - 1999, 2001] such that its entries can be permuted to the main diagonal. Thus we assume more or less

$$\mathbf{L} \approx \mathbf{I} \qquad \text{or} \qquad \mathbf{U} \approx \mathbf{I}.$$

Then we can approximate, for example, as

$$(\mathbf{M}-\mathbf{B})^{-1} = \left(\mathbf{L}(\mathbf{DU}-\mathbf{L}^{-1}\mathbf{B})\right)^{-1} \approx (\mathbf{DU}-\mathbf{B})^{-1}\mathbf{L}^{-1},$$

if $\mathbf{DU}-\mathbf{B}$ is nonsingular. If $\overline{\mathbf{DU}-\mathbf{B}}$ denotes a nonsingular and easily invertible approximation of $\mathbf{DU}-\mathbf{B}$, then we define $\mathbf{M}^{+}$ by

$$\mathbf{M}^{+} = \mathbf{L}(\overline{\mathbf{DU}-\mathbf{B}}). \tag{1}$$

Lemma 2. Let $||\mathbf{A}-\mathbf{LDU}|| = \varepsilon||\mathbf{A}|| < ||\mathbf{B}||$. Then the preconditioner from (1) satisfies

$$||\mathbf{A}^{+}-\mathbf{M}^{+}|| \;\leq\; ||\mathbf{A}^{+}-\mathbf{LDU}||\frac{||\mathbf{L}(\mathbf{DU}-\overline{\mathbf{DU}-\mathbf{B}})-\mathbf{B}||+\varepsilon||\mathbf{A}||}{||\mathbf{B}||-\varepsilon||\mathbf{A}||}$$

$$\leq ||\mathbf{A}^{+}-\mathbf{LDU}|| \quad\cdot\quad \frac{||\mathbf{L}||\;||\mathbf{DU}-\mathbf{B}-\overline{\mathbf{DU}-\mathbf{B}}||+||\mathbf{L}-\mathbf{I}||\;||\mathbf{B}||+\varepsilon||\mathbf{A}||}{||\mathbf{B}||-\varepsilon||\mathbf{A}||}.$$

Next we propose approximations of $\mathbf{DU}-\mathbf{B}$.

A very simple choice of $\overline{\mathbf{DU}-\mathbf{B}}$ for $\mathbf{M}^+$ in (1) is

$$\overline{\mathbf{DU}-\mathbf{B}} \equiv triu(\mathbf{DU}-\mathbf{B}), \qquad \mathbf{M}^+ = \mathbf{L}\cdot triu(\mathbf{DU}-\mathbf{B}),$$

where $triu$ denotes the upper triangle (including the main diagonal). From Lemma 2, assuming $\mathbf{L} \approx \mathbf{I}$, $\mathbf{M}^+$ is accurate if the upper triangle of $\mathbf{B}$ contains an important part of the whole difference matrix $\mathbf{B}$. This seems to be the case if the difference matrix is rather nonsymmetric as in upwind perturbations in nonlinear convection-diffusion problems.

Model problem: The two-dimensional nonlinear convection-diffusion problem [Kelley - 1995]

$$\Delta u - Ru\nabla u = 2000x(1-x)y(1-y),\ R = 50,$$

on the unit square, discretized by 5-point finite differences on a uniform 70x70 grid with as initial approximation the discretization of $u_0(x,y) = 0$.

$\mathbf{A}/\mathbf{M}$	$\mathbf{LDU}$	$\mathbf{L} \cdot triu(\mathbf{DU} - \mathbf{B})$
$\mathbf{A}^{(0)}$ / $\mathbf{M}^{(0)}$	21	21
$\mathbf{A}^{(1)}$ / $\mathbf{M}^{(0)}$	29	25
$\mathbf{A}^{(2)}$ / $\mathbf{M}^{(0)}$	39	27
$\mathbf{A}^{(3)}$ / $\mathbf{M}^{(0)}$	52	25
$\mathbf{A}^{(4)}$ / $\mathbf{M}^{(0)}$	77	25
$\mathbf{A}^{(5)}$ / $\mathbf{M}^{(0)}$	80	26
$\mathbf{A}^{(6)}$ / $\mathbf{M}^{(0)}$	102	26
$\mathbf{A}^{(7)}$ / $\mathbf{M}^{(0)}$	102	27
$\mathbf{A}^{(8)}$ / $\mathbf{M}^{(0)}$	98	27
$\mathbf{A}^{(9)}$ / $\mathbf{M}^{(0)}$	101	26
$\mathbf{A}^{(10)}$ / $\mathbf{M}^{(0)}$	99	26
$\mathbf{A}^{(0)-(10)}$ / $\mathbf{M}^{(0)-(10)}$	21 ± 5	—

Numbers of BiCGSTAB iterations for solving preconditioned linear systems of a nonlinear convection-diffusion problem with no updates and triangular updates, respectively. $\mathbf{M}^{(0)} = ILUT(0.1, 5)$ with $\frac{\|\mathbf{I}-\mathbf{L}\|}{\|\mathbf{L}\|} \approx 0.4$.

We considered two improvements of the update $\mathbf{L}\cdot triu(\mathbf{DU}-\mathbf{B})$:

1. Adaptive choice of $\mathbf{L}\cdot triu(\mathbf{DU}-\mathbf{B})$ or $tril(\mathbf{LD}-\mathbf{B})\cdot\mathbf{U}$ based on $\|triu(\mathbf{B})\|$ and $\|tril(\mathbf{B})\|$ (and $\|\mathbf{I}-\mathbf{L}\|$ and $\|\mathbf{I}-\mathbf{U}\|$);

2. A strategy to approximate $\mathbf{DU}-\mathbf{B}$ by a generally non-triangular but easily invertible matrix. We developed an algorithm to find rows $i_1,\ldots,i_K$ with

$$\begin{aligned}\overline{\mathbf{DU}-\mathbf{B}} &= \tilde{\mathbf{D}}(\mathbf{I}-\tilde{\mathbf{B}}) = \tilde{\mathbf{D}}(\mathbf{I}-\sum_{j=1}^{K} e_{i_j}\tilde{b}_{i_j*}) \\ &= \tilde{\mathbf{D}}(\mathbf{I}-e_{i_1}\tilde{b}_{i_1*})(\mathbf{I}-e_{i_2}\tilde{b}_{i_2*})\ldots(\mathbf{I}-e_{i_K}\tilde{b}_{i_K*}), \qquad (2)\end{aligned}$$

where $\mathrm{diag}\overline{(\mathbf{DU}-\mathbf{B})}\equiv\tilde{\mathbf{D}}$, and $\tilde{\mathbf{D}}^{-1}(\tilde{\mathbf{D}}-\overline{\mathbf{DU}-\mathbf{B}})\equiv\tilde{\mathbf{B}}$.

Example with adaptive choice of triangular updates:

Simulation of air flow in a tunnel at a low Mach number (provided by P. Birken):

- Von Neumann boundary conditions, Lax-Friedrichs flux
- Finite volume discretization (first order)
- ODE solved with the implicit Euler method
- Every time step yields a nonlinear system of equations that is solved with a Newton-type method
- Linear systems have dimension 4800 and have maximally 20 nonzero entries per row.

nz=138024, ILUT(0.001/5), timep=0.05, psize=135798						
A / M	Self prec		Freeze		Update	
	Its	Time	Its	Time	Its	Time
$A^{(1)}$ / $M^{(0)}$	24	0.44	17	0.34	17	0.31
$A^{(5)}$ / $M^{(0)}$	29	0.52	19	0.33	19	0.34
$A^{(10)}$ / $M^{(0)}$	30	0.50	17	0.27	17	0.27
$A^{(15)}$ / $M^{(0)}$	33	0.59	21	0.39	19	0.34
$A^{(20)}$ / $M^{(0)}$	32	0.59	19	0.34	27	0.31
$A^{(25)}$ / $M^{(0)}$	33	0.51	20	0.33	19	0.33
$A^{(30)}$ / $M^{(0)}$	34	0.61	24	0.44	21	0.34
$A^{(35)}$ / $M^{(0)}$	33	0.61	23	0.42	19	0.36
$A^{(40)}$ / $M^{(0)}$	39	0.67	31	0.52	24	0.39
$A^{(45)}$ / $M^{(0)}$	44	0.73	33	0.55	27	0.45
$A^{(50)}$ / $M^{(0)}$	40	0.70	39	0.63	24	0.44
$A^{(55)}$ / $M^{(0)}$	40	0.69	47	0.78	25	0.42
$A^{(60)}$ / $M^{(0)}$	47	0.80	80	1.41	31	0.56
$A^{(65)}$ / $M^{(0)}$	47	0.75	107	1.64	27	0.42
$A^{(70)}$ / $M^{(0)}$	38	0.70	72	1.28	28	0.51
$A^{(75)}$ / $M^{(0)}$	114	1.98	230	4.06	105	1.96

3. A future issue:

Question: Can we permute the whole sequence in such a way that we know where the dominating entries of the difference matrices **B** are located ?

For example: We may permute the first system matrix $\mathbf{A}^{(0)}$ to block triangular form (see, e.g. [Pothen, Fan - 1990]):

$$\mathbf{P}\mathbf{A}^{(0)} = \left|\begin{array}{c|c|c} \mathbf{A}_{11} & \mathbf{A}_{12} & \mathbf{A}_{13} \\ \hline 0 & \mathbf{A}_{22} & \mathbf{A}_{23} \\ \hline 0 & 0 & \mathbf{A}_{33} \end{array}\right|$$

If the other system-matrices have the same sparsity pattern, we can permute the whole sequence as

$$\mathbf{PA}^{(i)}x = \mathbf{P}b^{(i)},\ i = 0, 1, \dots,$$

and all matrices will have upper triangular form.

Then clearly we can expect dominating entries of the difference matrices lay in the upper triangle and the preconditioner updates

$$\mathbf{L}(\mathbf{DU} - triu(\mathbf{PB}))$$

may be stronger than without the permutation **P**.

In addition, the computation of the upper triangular form can be modified to take into account the magnitude of the entries.

More details can be found in ,,Preconditioner updates for solving sequences of large and sparse nonsymmetric linear systems" [Duintjer Tebbens, Tůma - to be submitted in 2006], see webpages of the authors.

Thank you for your attention.

This work is supported by the Program Information Society under project 1ET400300415.