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@cs.cas.cz

GAMM Workshop on Applied and Numerical Linear Algebra,

Dresden, September 22, 2005

Outline

1. Introduction to preconditioner updates
2. Proposed approximate preconditioner updates

1. Introduction to preconditioner updates

Consider a sequence of linear systems

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

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, non-PDE problems.

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.

Some options to reduce overall costs:

- Modify Newton's method by skipping some Jacobian evaluations: Shamanskii combination of Newton's method and the Newton-chord method. Much weaker nonlinear convergence properties than the standard Newton's method.

- The sequence of linear systems must often be preconditioned; computing preconditioners $\mathbf{M}^{(1)}, \mathbf{M}^{(2)}, \ldots$ 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 linear system $A^{(1)}x = b^{(1)}$. For example:

- In the Newton-Krylov framework: Recycle Krylov subspaces among systems of a sequence, see e.g. [Loghin, Ruiz, Touhami-2004], [Parks, de Sturler, Mackey, Johnson, Maiti - 2004].
- Many interesting algorithms were proposed for exact updates of decompositions. Recent sparse updates [Davis, Hager - 1999, 2001, 2005] replace in some cases classical dense updates from, e.g., [Gill, Murray, Saunders - 1975].

- There is some recent work in approximate updates 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 factorized, and general nonsymmetric preconditioners which may be useful in solving subsequent linear systems. We do not confine ourselves to particular classes of linear solvers (e.g. Krylov subspace 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 $\mathbf{A}$.
The quality of the preconditioner $\mathbf{M}$ can be expressed by

$$\mathbf{A} - \mathbf{M} \tag{1}$$

in some norm or by a norm of one of the matrices

$$\mathbf{I} - \mathbf{M}^{-1}\mathbf{A} \qquad \text{or} \qquad \mathbf{I} - \mathbf{A}\mathbf{M}^{-1} \tag{2}$$

if we consider preconditioning from the left or right, respectively (see, e.g. [Benzi, Bertaccini - 2003]). While the norm of the matrix (1) expresses *accuracy* of the preconditioner, the norms of the matrices (2) relate to its *stability* [Chow, Saad - 1997], see also [Benzi, Haws, Tůma - 2000].

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.

Some of our approximations of the ideal update are based on the assumption that 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

$$\mathbf{L} \approx \mathbf{I} \approx \mathbf{U}.$$

We can approximate $(\mathbf{M}-\mathbf{B})^{-1}$, if it is nonsingular, as

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

provided $\mathbf{D}-\mathbf{B}$ is nonsingular. Denote by $\overline{\mathbf{D}-\mathbf{B}}$ a nonsingular approximation of $\mathbf{D}-\mathbf{B}$ that can be inverted inexpensively. Then define a preconditioner $\mathbf{M}^{+}$ as

$$\mathbf{M}^{+} = \mathbf{L}(\overline{\mathbf{D}-\mathbf{B}})\mathbf{U}.$$

The accuracy of this preconditioner can be significantly higher than the accuracy of the frozen preconditioner $\mathbf{M} = \mathbf{LDU}$ for $\mathbf{A}^{+}$:

Lemma 1. Let $||\mathbf{A}-\mathbf{LDU}|| = \varepsilon||\mathbf{A}|| < ||\mathbf{B}||$. Then

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

with

$$\frac{||\mathbf{L}(\mathbf{D}-\overline{\mathbf{D}-\mathbf{B}})\mathbf{U}-\mathbf{B}||+\varepsilon||\mathbf{A}||}{||\mathbf{B}||-\varepsilon||\mathbf{A}||}\leq$$

$$\frac{\|\mathbf{L}\|\,\|\mathbf{D}-\mathbf{B}-\overline{\mathbf{D}-\mathbf{B}}\|\,\|\mathbf{U}\|+\|\mathbf{L}-\mathbf{I}\|\,\|\mathbf{B}\mathbf{U}\|+\|\mathbf{B}\|\,\|\mathbf{U}-\mathbf{I}\|+\varepsilon||\mathbf{A}||}{||\mathbf{B}||-\varepsilon||\mathbf{A}||}.$$

In the symmetric case, the preconditioner $\mathbf{M}^+$ changes to $\mathbf{M}^+ = \mathbf{L}(\overline{\mathbf{D}-\mathbf{B}})\mathbf{L}^T$, hence symmetry is preserved. In the nonsymmetric case we can assume that only one of the two factors $\mathbf{L}, \mathbf{U}$ is close to the identity matrix, instead of both. We can approximate as

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

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

$$\mathbf{M}^{+} = \mathbf{L}(\overline{\mathbf{DU} - \mathbf{B}}). \tag{3}$$

Lemma 2. Let $||\mathbf{A} - \mathbf{LDU}|| = \varepsilon||\mathbf{A}|| < ||\mathbf{B}||$. Then the preconditioner from (3) satisfies

$$\begin{aligned} ||\mathbf{A}^{+} - \mathbf{M}^{+}|| &\le ||\mathbf{A}^{+} - \mathbf{LDU}|| \frac{||\mathbf{L}(\mathbf{DU} - \overline{\mathbf{DU} - \mathbf{B}}) - \mathbf{B}|| + \varepsilon||\mathbf{A}||}{||\mathbf{B}|| - \varepsilon||\mathbf{A}||} \\ \le ||\mathbf{A}^{+} - \mathbf{LDU}|| &\cdot \frac{\|\mathbf{L}\| \, \|\mathbf{DU} - \mathbf{B} - \overline{\mathbf{DU} - \mathbf{B}}\| + ||\mathbf{L} - \mathbf{I}|| \, \|\mathbf{B}\| + \varepsilon||\mathbf{A}||}{||\mathbf{B}|| - \varepsilon||\mathbf{A}||}. \end{aligned}$$

Next we propose approximations of $\mathbf{DU} - \mathbf{B}$. All techniques we treat can be analogously formulated for updates of the form $(\overline{\mathbf{LD} - \mathbf{B}})\mathbf{U}$. The introduced algorithms can be used to approximate the matrix $\overline{\mathbf{D} - \mathbf{B}}$ as well.

A very simple choice of $\overline{\mathbf{DU}-\mathbf{B}}$ for $\mathbf{M}^+$ in (3) 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/downwind 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 $u_0(x,y) = 0$.

$\mathbf{A}/\mathbf{M}$	$\mathbf{LDU}$	$\mathbf{L} \cdot triu(\mathbf{DU} - \mathbf{B})$
$\mathbf{A}^{(1)}$ / $\mathbf{M}^{(1)}$	21	21
$\mathbf{A}^{(2)}$ / $\mathbf{M}^{(1)}$	29	25
$\mathbf{A}^{(3)}$ / $\mathbf{M}^{(1)}$	39	27
$\mathbf{A}^{(4)}$ / $\mathbf{M}^{(1)}$	52	25
$\mathbf{A}^{(5)}$ / $\mathbf{M}^{(1)}$	77	25
$\mathbf{A}^{(6)}$ / $\mathbf{M}^{(1)}$	80	26
$\mathbf{A}^{(7)}$ / $\mathbf{M}^{(1)}$	102	26
$\mathbf{A}^{(8)}$ / $\mathbf{M}^{(1)}$	102	27
$\mathbf{A}^{(9)}$ / $\mathbf{M}^{(1)}$	98	27
$\mathbf{A}^{(10)}$ / $\mathbf{M}^{(1)}$	101	26
$\mathbf{A}^{(11)}$ / $\mathbf{M}^{(1)}$	99	26
$\mathbf{A}^{(1)-(11)}$ / $\mathbf{M}^{(1)-(11)}$	21 ± 5	—

Numbers of BiCGSTAB iterations for solving preconditioned linear systems of a nonlinear convection-diffusion problem with no updates and triangular updates, respectively. $\mathbf{M}^{(1)} = ILUT(0.1, 5)$.

The presented strategies are strongly based on confining the update to the upper (or, equivalently, lower) triangle. Whereas numerical experiments seem to indicate this makes sense, there may be applications where it is necessary to take into account both triangles of the difference matrix.

Here we introduce a strategy to approximate $\mathbf{DU}-\mathbf{B}$ by a general non-triangular but easily invertible matrix. Denote the matrix $\mathrm{diag}(\overline{\mathbf{DU}-\mathbf{B}})$ by $\tilde{\mathbf{D}}$, and $\tilde{\mathbf{D}}^{-1}(\tilde{\mathbf{D}}-\overline{\mathbf{DU}-\mathbf{B}})$ denote by $\tilde{\mathbf{B}}$. Then

$$\overline{\mathbf{DU}-\mathbf{B}} = \tilde{\mathbf{D}}(\mathbf{I}-\tilde{\mathbf{B}}).$$

First consider the case when $\tilde{\mathbf{B}} = \beta e_i e_j^T$, for some $1 \le i, j \le n, i \neq j$. Then we get

$$(\mathbf{I}-\tilde{\mathbf{B}})^{-1} = \mathbf{I} + \frac{\beta}{1-\beta e_j^T e_i} e_i e_j^T = \mathbf{I} + \beta e_i e_j^T,$$

a Gauss-Jordan transformation [Golub, van Loan - 1996] with fill-in free inverse.

Idea: Approximate $\mathbf{DU}-\mathbf{B}$ by a product of Gauss-Jordan transformations.

We achieve this as follows: $\overline{\mathbf{DU}-\mathbf{B}}$ will consist of the main diagonal plus some rows of $\mathbf{DU}-\mathbf{B}$:

$$\begin{aligned}\overline{\mathbf{DU}-\mathbf{B}} &= \tilde{\mathbf{D}}(\mathbf{I}-\tilde{\mathbf{B}}) = \tilde{\mathbf{D}}(\mathbf{I}-\sum_{j=1}^{K} e_{i_j}\tilde{b}_{i_j*}) \\ &= \tilde{\mathbf{D}}(\mathbf{I}-e_{i_1}\tilde{b}_{i_1*})(\mathbf{I}-e_{i_2}\tilde{b}_{i_2*})\ldots(\mathbf{I}-e_{i_K}\tilde{b}_{i_K*}), \quad (4)\end{aligned}$$

where $\tilde{b}_{j*} = e_j^T\tilde{\mathbf{B}}$.

If $row(i) = \{k|i \neq k \wedge \tilde{b}_{ik} \neq 0\}$, then the number of operations for multiplying a vector by a matrix of the form (4) or its inverse is at most $2\sum_{j=1}^{K}|row(i_j)| + n$.

Clearly,

$$(\mathbf{I} - e_{i_1}\tilde{b}_{i_1*})(\mathbf{I} - e_{i_2}\tilde{b}_{i_2*}) = (\mathbf{I} - e_{i_1}\tilde{b}_{i_1*} - e_{i_2}\tilde{b}_{i_2*} + e_{i_1}\tilde{b}_{i_1*} \cdot e_{i_2}\tilde{b}_{i_2*})$$
$$= (\mathbf{I} - e_{i_1}\tilde{b}_{i_1*} - e_{i_2}\tilde{b}_{i_2*}) \quad \text{if and only if} \quad e_{i_1}\tilde{b}_{i_1*} \cdot e_{i_2}\tilde{b}_{i_2*} = 0.$$

Theorem 1: Let $\mathbf{I} - \tilde{\mathbf{B}} = \mathbf{I} - \sum_{j_l:l=1,\ldots,K} e_{j_l}\tilde{b}_{j_l*}$. Then

$$\mathbf{I} - \tilde{\mathbf{B}} = (\mathbf{I} - e_{i_1}\tilde{b}_{i_1*})(\mathbf{I} - e_{i_2}\tilde{b}_{i_2*})\ldots(\mathbf{I} - e_{i_K}\tilde{b}_{i_K*})$$

if and only if

$$i_l \notin \bigcup_{k=1}^{l-1} row(i_k) \qquad \text{for } 2 \leq l \leq K$$

for all $i_1, \ldots, i_K$ such that $\{j_1, \ldots, j_K\} = \{i_1, \ldots, i_K\}$.

Example: A unit lower triangular matrix $\mathbf{I} - \tilde{\mathbf{B}}$ can be written as

$$(\mathbf{I} - e_2\tilde{b}_{2*})(\mathbf{I} - e_3\tilde{b}_{3*})(\mathbf{I} - e_4\tilde{b}_{4*})\ldots(\mathbf{I} - e_{n-1}\tilde{b}_{n-1}*)(\mathbf{I} - e_n\tilde{b}_n*).$$

In case of a unit lower triangular matrix $\mathbf{I} - \tilde{\mathbf{B}}$ with k additional subsequent subdiagonals starting in the lth column a product of Gauss-Jordan transformations with a fill-in free inverse can cover only $l/(2l + k - 1)$ percent of the rows.

Unfortunately, changes in a sequence of matrices restricted to a couple of diagonals are rather frequent.

A simple greedy procedure based on sparsification and Theorem 1 is the following algorithm.

Algorithm to find rows i_l of $\mathbf{DU} - \mathbf{B}$ such that it is approximated by a product of Gauss-Jordan transformations.

1. set $\mathcal{R} = \{1, \ldots, n\}$, $l = 0$
2. for $k = 1, \ldots, n$ do
3. set $row(k) = \{i | i \neq k \wedge |(\mathbf{DU} - \mathbf{B})_{ki}| > tol\}$
4. set $p_k = \sum_{j \in row(k)} |(\mathbf{DU} - \mathbf{B})_{kj}|$
5. end for
6. while $\mathcal{R} \neq \emptyset$ do
7. choose a row $i \in \mathcal{R}$ maximizing $p_i - \sum_{j \in \mathcal{R} \cap row(i)} p_j$
8. set $l = l + 1$, $i_l = i$
9. set $\mathcal{R} = \mathcal{R} \backslash \{row(i_l) \cup i_l\}$
10. end while

A/M	**LDU**	$\mathbf{L} \cdot GJ(\mathbf{DU} - \mathbf{B})$	$\mathbf{L} \cdot GJ(\mathbf{D} - \mathbf{B}) \cdot \mathbf{U}$
A^1 / M^1	5	5	5
A^2 / M^1	31	17	36
A^3 / M^1	51	18	40
A^4 / M^1	71	21	51
A^5 / M^1	91	21	59
A^6 / M^1	97	23	63
A^7 / M^1	100	21	64
A^8 / M^1	97	23	70
A^9 / M^1	103	22	65
A^{10} / M^1	100	22	76
A^{11} / M^1	99	22	71

Numbers of BiCGSTAB iterations for the preconditioned nonlinear convection-diffusion problem with preconditioner updated by GJ updates applied to approximate $(\mathbf{D} - \mathbf{B})$ *and* $(\mathbf{DU} - \mathbf{B})$, *respectively.*

Same model problem, 50 x 50 grid, $\mathbf{M}^{(1)} = ILU(10^{-3})$,

$$\frac{\|\mathbf{I} - \mathbf{L}\|}{\|\mathbf{L}\|} = 0.434 = \frac{\|\mathbf{I} - \mathbf{U}\|}{\|\mathbf{U}\|}.$$

A / M	$ILU(10^{-1})$			$ILU(10^{-2})$		
	LDU	**L**GJ(**DU** − **B**)	**L**$triu$(**DU** − **B**)	**LDU**	**L**GJ(**DU** − **B**)	**L**$triu$(**DU** − **B**)
A^1 / M^1	24	24	24	13	13	13
A^2 / M^1	27	26	24	32	20	17
A^3 / M^1	38	27	21	58	23	17
A^4 / M^1	47	25	23	89	24	17
A^5 / M^1	52	22	23	127	23	17
A^6 / M^1	58	21	22	131	24	18
A^7 / M^1	68	22	23	182	25	18
A^8 / M^1	91	24	24	172	26	19
A^9 / M^1	70	20	23	157	22	18
A^{10} / M^1	68	22	24	166	24	18
A^{11} / M^1	76	24	25	163	24	19

Numbers of BiCGSTAB iterations for the preconditioned nonlinear convection-diffusion problem with preconditioner updated by Gauss-Jordan updates applied to (**DU**−**B**) *and triangular updates, respectively*

The sizes of the factors **L** and **U** are for the drop tolerances 10^{-3}, 10^{-2} and 10^{-1} equal to approximately 35000, 12000 and 5000, respectively. The sizes of the GJ updates are in the range $< 6700, 7800 >$ for all tolerances.

Future issues:

- An interesting problem is to choose triangular updates which correspond to the sparsity pattern and sizes of entries of the difference matrix *differently* for each system.

- A closely related problem is to find a nonsymmetric permutation which would transform the system matrices into a form which is more suitable for our updates.

More details can be found in ,,Preconditioner updates for solving sequences of large and sparse nonsymmetric linear systems'' [Duintjer Tebbens, Tůma - submitted to SISC in 2005].

Thank you for your attention.

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