# Recent Advances in Large-Scale Numerical Linear Algebra

Michele Benzi
Department of Mathematics
and Computer Science
Emory University
Atlanta, GA, USA

# Acknowledgements


- Valeria Simoncini (Università di Bologna, Italy)
- Daniel Szyld (Temple University)
- James Warsa (LANL)

# Outline

- Introduction
- A brief review of Krylov subspace methods
- Preconditioning
- Schur complements
- Some examples
- Inexact methods
- References

Some papers can be found at

http://www.mathcs.emory.edu/∼benzi

# Introduction

Large-Scale Numerical Linear Algebra is primarily concerned with the following problems:

- Solution of linear systems $\mathcal{A}x = b$;
- Solution of (generalized) eigenproblems $\mathcal{A}x = \lambda\mathcal{B}x$;
- The approximation of matrix functions $f(\mathcal{A})$; often, one is interested in evaluating $f(\mathcal{A})b$ for a given $b$.

Here the matrices $\mathcal{A}$, $\mathcal{B}$ are very large and not necessarily sparse. These and related problems arise frequently in computational science and engineering, and research on effective algorithms is ongoing.

## Introduction (cont.)

The main difficulties are due to the following features:

1. The ever-increasing size of the problems to be solved, which makes multilevel methods virtually mandatory;

2. The complexity of the problems: strong couplings, multi-physics, multiple scales, etc.

3. Ill-posedness, ill-conditioning, non-normality ...

These challenges call for increasingly sophisticated solution approaches. In this talk I will discuss some recent work that should be of interest to the transport community.

## Introduction (cont.)

We are especially interested in cases where

- The coefficient matrices may not be explicitly available; instead, users have to provide procedures for evaluating the action of an operator on a vector.

- Such matrix-vector products are computationally expensive, and/or difficult to parallelize.

- Two or more nested levels of iteration are present.

- Elaborate preconditioning strategies are needed.

Before giving illustrative examples, a brief review of Krylov subspace methods is in order.

# Krylov Subspace Methods

Suppose $x_0$ is an initial guess for the solution of $\mathcal{A}x = b$, and let $r_0 = b - \mathcal{A}x_0$ be the corresponding residual. Here $\mathcal{A}$ is $N \times N$.

Krylov subspace methods are iterative (approximation) schemes whose $k$th iterate $x_k$ satisfies

$$x_k \in x_0 + \mathcal{K}_k(\mathcal{A}, r_0), \quad k = 1, 2, \ldots$$

where

$$\mathcal{K}_k(\mathcal{A}, r_0) \equiv \text{span}\{r_0, \mathcal{A}r_0, \ldots, \mathcal{A}^k r_0\}$$

denotes the $k$th Krylov subspace generated by $\mathcal{A}$ and $r_0$. The Krylov subspaces form a nested sequence

$$\mathcal{K}_1(\mathcal{A}, r_0) \subset \mathcal{K}_2(\mathcal{A}, r_0) \subset \cdots \subset \mathcal{K}_d(\mathcal{A}, r_0) = \cdots = \mathcal{K}_N(\mathcal{A}, r_0).$$

# Krylov Subspace Methods (cont.)

Since the $k$th Krylov subspace $\mathcal{K}_k(\mathcal{A}, r_0)$ is $k$-dimensional (for $k \leq d$), there are $k$ degrees of freedom in the choice of the iterate $x_k$.

Uniquely defined iterates are obtained by imposing $k$ constraints, in the form of orthogonality of the $k$th residual $r_k$ with respect to a prescribed $k$-dimensional subspace $\mathcal{C}_k$:

$$r_k = b - \mathcal{A}x_k \in r_0 + \mathcal{A}\mathcal{K}_k(\mathcal{A}, r_0), \quad r_k \perp \mathcal{C}_k.$$

Different choices of the constraint subspace $\mathcal{C}_k$ lead to different types of Krylov subspace methods.

# Krylov Subspace Methods (cont.)

**Theorem.** Assume $\dim \mathcal{K}_k(\mathcal{A}, r_0) = k$, and let $x^* = \mathcal{A}^{-1}b$.

1. If $\mathcal{A}$ is SPD and $\mathcal{C}_k = \mathcal{K}_k(\mathcal{A}, r_0)$, then $x_k$ is uniquely defined and satisfies

$$\|e_k\|_{\mathcal{A}} \equiv \|x^* - x_k\|_{\mathcal{A}} = \min_{z \in x_0 + \mathcal{K}_k(\mathcal{A}, r_0)} \|x^* - z\|_{\mathcal{A}} = \min_{p \in \Pi_k} \|p(\mathcal{A})e_0\|_{\mathcal{A}}.$$

2. If $\mathcal{A}$ is nonsingular and $\mathcal{C}_k = \mathcal{A}\mathcal{K}_k(\mathcal{A}, r_0)$, then $x_k$ is uniquely defined and satisfies

$$\|r_k\|_2 \equiv \|b - \mathcal{A}x_k\|_2 = \min_{z \in x_0 + \mathcal{K}_k(\mathcal{A}, r_0)} \|b - \mathcal{A}z\|_2 = \min_{p \in \Pi_k} \|p(\mathcal{A})r_0\|_2.$$

Here $\Pi_k$ denotes the set of all polynomials of degree at most $k$ such that $p(0) = 1$.

The method of conjugate gradients (CG) is of Type 1, while minimal residual methods (like GMRES) are of type 2.

## Krylov Subspace Methods (cont.)

Besides these two basic classes of methods, there exist a number of variations (included restarted and truncated variants) and hybrids. Not all of these methods satisfy an optimality (or "quasi-optimality") property.

However, all Krylov methods share the property that the coefficient matrix $\mathcal{A}$ is accessed only in the form of matrix-vector products. In some older methods the transpose of $\mathcal{A}$ is also needed, but such methods are seldom used nowadays.

This property makes Krylov method attractive, and sometimes the only option, when $\mathcal{A}$ is not explicitly available.

# Krylov Subspace Methods (cont.)

Krylov subspace methods are projection methods: the approximate solution at step $k$ is obtained from the solution of a projection of the original $N$-dimensional problem onto a $k$-dimensional subspace, where $k \ll N$.

The same idea is used to solve eigenvalue problems and to approximate expressions of the form $f(\mathcal{A})b$, where $f$ is a function and $b$ a given vector, for example $\mathrm{e}^{-t\mathcal{A}}b$.

Clearly, it is desirable to keep the dimension $k$ of the approximating Krylov subspace as small as possible: in other words, convergence of the Krylov approximation should be fast.

# Preconditioning

Characterizing the rate of convergence of Krylov subspace methods is a non-trivial problem, still open in the case of general, non-normal matrices.

In the case of linear systems, however, a tightly clustered spectrum of the system matrix $\mathcal{A}$ often results in rapid convergence of the iterates $x_k$ to $x^*$.

A clustering of the spectrum is typically obtained by preconditioning: the original system is replaced by the equivalent left- or right-preconditioned system

$$\mathcal{P}^{-1}\mathcal{A}x = \mathcal{P}^{-1}b, \quad \text{or} \quad \mathcal{A}\mathcal{P}^{-1}y = b, \;\; x = \mathcal{P}^{-1}y.$$

Here the nonsingular matrix $\mathcal{P}$ is the preconditioner, chosen such that $\mathcal{P} \approx \mathcal{A}$, or $\mathcal{P}^{-1} \approx \mathcal{A}^{-1}$.

## Preconditioning (cont.)

The choice of a suitable preconditioner is key to the success of the Krylov subspace solver: for most problems, the Krylov method alone is unable to achieve convergence in a reasonable number of iterations.

A good preconditioner must be effective at reducing the number of iterations to convergence, while at the same time being relatively inexpensive to set-up and to apply.

In the last 15 years or so, a great deal of effort has been put into the development of effective preconditioning techniques. Although the best preconditioners are usually problem-dependent, there has also been considerable interest in general-purpose, algebraic preconditioners.

## Preconditioning (cont.)

Preconditioning usually requires solving a linear system of the form $\mathcal{P}z = v_k$ at the $k$th step of the Krylov subspace method (sometimes more than once) for each $k = 1, 2, \ldots$

In some cases this is accomplished by an inner iteration, where the number of iterations may vary from one outer iteration to the next. This is an example of variable, or non-linear, preconditioning.

In this case the outer iteration is, in general, no longer a true Krylov method: we have instead flexible variants of standard methods like GMRES, QMR, CG. The most popular among these is FGMRES (Saad, 1993).

# Schur Complements

Suppose $\mathcal{A}$ can be partitioned into a $2 \times 2$ block form

$$\mathcal{A} = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

where $A$ is invertible. The Schur complement of $A$ in $\mathcal{A}$ is the matrix

$$S = D - CA^{-1}B\,.$$

Schur complements naturally arise in many contexts.
Note that if $A$ is invertible, $\mathcal{A}$ admits the block factorization

$$\mathcal{A} = \begin{bmatrix} I_n & O \\ CA^{-1} & I_m \end{bmatrix} \begin{bmatrix} A & O \\ O & S \end{bmatrix} \begin{bmatrix} I_n & A^{-1}B \\ O & I_m \end{bmatrix},$$

with $n + m = N$. Hence if $A$ is nonsingular, $\mathcal{A}$ is nonsingular if and only if $S$ is.

# Schur Complements (cont.)

Suppose now we want to solve $\mathcal{A}x = b$; partition $\mathcal{A}$, $x$ and $b$ as

$$\mathcal{A} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \quad x = \begin{bmatrix} y \\ z \end{bmatrix}, \quad b = \begin{bmatrix} c \\ d \end{bmatrix}.$$

Eliminating $y$ from the system leads to the reduced system

$$Sz = h \quad \text{with} \quad h = d - CA^{-1}c\,.$$

Once the reduced system has been solved for $z$, $y$ is recovered as the solution of

$$Ay = c - Bz\,.$$

This approach is often used when $(n \times n)$ linear systems with $A$ can be solved efficiently. The problem then is solving one $(m \times m)$ linear system with $S$. Note that often $m \ll n$. Still, usually $S$ is dense and cannot be formed explicitly.

## Schur Complements (cont.)

The reduced system

$$Sz = h\,, \quad \text{where} \quad S = D - CA^{-1}B\,,$$

can be solved by an iterative method requiring only matrix-vector products with $S$.

Besides the matrix-vector multiplies with $D$, $C$ and $B$, a linear system with coefficient matrix $A$ must be solved at each iteration.

Unless $A$ is "easy to invert" (for instance, diagonal or block diagonal with small blocks), the solution of linear systems with $A$ is also performed iteratively, by an inner Krylov method, leading to a nested iterative scheme.

## Schur Complements (cont.)

Schur complements may arise as a consequence of a natural partitioning induced by the original problem (e.g., the presence of two unknown fields, like velocity and pressure in fluid mechanics).

In other cases, a partitioning may be imposed on the problem to facilitate its solution (e.g., domain decomposition).

In these examples, the Schur complement system arises as a reduced system, upon elimination of part of the variables.

An advantage of this approach is that often, the Schur complement is better behaved than the original, unreduced system: for example it may be definite rather than indefinite, or well-conditioned rather than ill-conditioned.

# Some Examples

The first example is from incompressible fluid mechanics. Consider the Stokes problem: Given a force field $\mathbf{f}$ in $\Omega \subset \mathbf{R}^d$, with $d = 2$ or $d = 3$, find a velocity field $\mathbf{u}$ and a pressure field $p$ such that

$$-\Delta \mathbf{u} + \nabla p = \mathbf{f} \quad \text{in } \Omega$$

$$\text{div } \mathbf{u} = 0 \quad \text{in } \Omega$$

$$\mathcal{B}\mathbf{u} = \mathbf{0} \quad \text{on } \partial\Omega\,.$$

Here $\mathcal{B}$ denotes some kind of boundary operator. To uniquely determine $p$ one may impose the condition

$$\int_\Omega p\,dx = 0\,.$$

## Some Examples (cont.)

LBB-stable discretizations by finite elements or MAC schemes lead to discrete saddle point systems of the form

$$\begin{bmatrix} A & B^T \\ B & O \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ p \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \end{bmatrix}, \quad \text{or} \quad \mathcal{A}x = b\,,$$

where $A$ is SPD and $B$ has full row rank. Thus, $\mathcal{A}$ is symmetric, but indefinite.

Elimination of the velocity vector $\mathbf{u}$ leads to a reduced (Schur complement) system for the pressure:

$$BA^{-1}B^T p = BA^{-1}\mathbf{f}\,, \quad \text{or} \quad \mathcal{S}p = h\,,$$

where $\mathcal{S} = BA^{-1}B^T$ and $h = BA^{-1}\mathbf{f}$.

## Some Examples (cont.)

The pressure Schur complement matrix $\mathcal{S} = BA^{-1}B^T$ is dense and cannot be formed explicitly. However, it is SPD. Also, it can be shown that $\mathcal{S}$ is well-conditioned, and is spectrally equivalent to the pressure mass matrix $M_p$.

Therefore, the system

$$\mathcal{S}p = h$$

can be solved by a rapidly convergent iterative method, like CG, preconditioned with the matrix $\mathcal{D} = \text{diag}(M_p)$.

At each PCG iteration, a matrix-vector multiply with $\mathcal{S} = BA^{-1}B^T$ must be computed. This requires solving linear systems with $A$, which amounts to solving $d$ (= 2 or 3) discrete Poisson equations.

## Some Examples (cont.)

These Poisson problems can be solved efficiently by a multi-grid iteration, leading to an inner-outer iterative scheme.

If these problems are solved inexactly, the resulting method is an inexact Uzawa algorithm (Elman & Golub, 1994).

Similar approaches can be applied to more general problems, including the Oseen problem (Picard linearization of the incompressible Navier–Stokes equations), and also to other types of saddle point problems (see Benzi, Golub & Liesen, 2005).

Also, many block preconditioning strategies require the approximate solution of Schur complement systems. This also leads to inner-outer iterative schemes.

## Some Examples (cont.)

The next example is from domain decomposition or, possibly, from graph partitioning.

For a given integer $p$, we would like to find a reordering (symmetric permutation) of $\mathcal{A}$ into the block form

$$\widehat{\mathcal{A}} = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] = \left[\begin{array}{cccc|c} A_1 & & & & B_1 \\ & A_2 & & & B_2 \\ & & \ddots & & \vdots \\ & & & A_p & B_p \\ \hline C_1 & C_2 & \cdots & C_p & A_S \end{array}\right],$$

with the size $m$ of the block $D = A_S$ as small as possible.

With such a partitioning we have

$$A = \text{diag}\,(A_1, A_2, \ldots, A_p) \quad \text{and} \quad \mathcal{S} = A_S - \sum_{i=1}^{p} C_i A_i^{-1} B_i\,.$$

## Some Exaples (cont.)

Such a reordering of $\mathcal{A}$ can be obtained using graph partitioning by vertex separator (GPVS) techniques.

Given an undirected graph $\mathcal{G} = (V, E)$ and an integer $p$, the $p$-way GPVS problem consists of finding a set of vertices $V_S$ of minimum size whose removal decomposes a graph into $p$ disconnected subgraphs $V_1, V_2, \ldots, V_p$ with balanced sizes. The problem is NP-hard.

Standard graph partitioning software (like METIS) can be applied to the undirected graph $\mathcal{G}$ associated with the symmetrized matrix $\mathcal{A} + \mathcal{A}^T$. Note that

$$|V| = N \quad \text{and} \quad |E| = nnz(\mathcal{A} + \mathcal{A}^T)\,.$$

## Some Examples (cont.)

The matrix $\mathcal{A}$ can be put into the $2 \times 2$ block structure

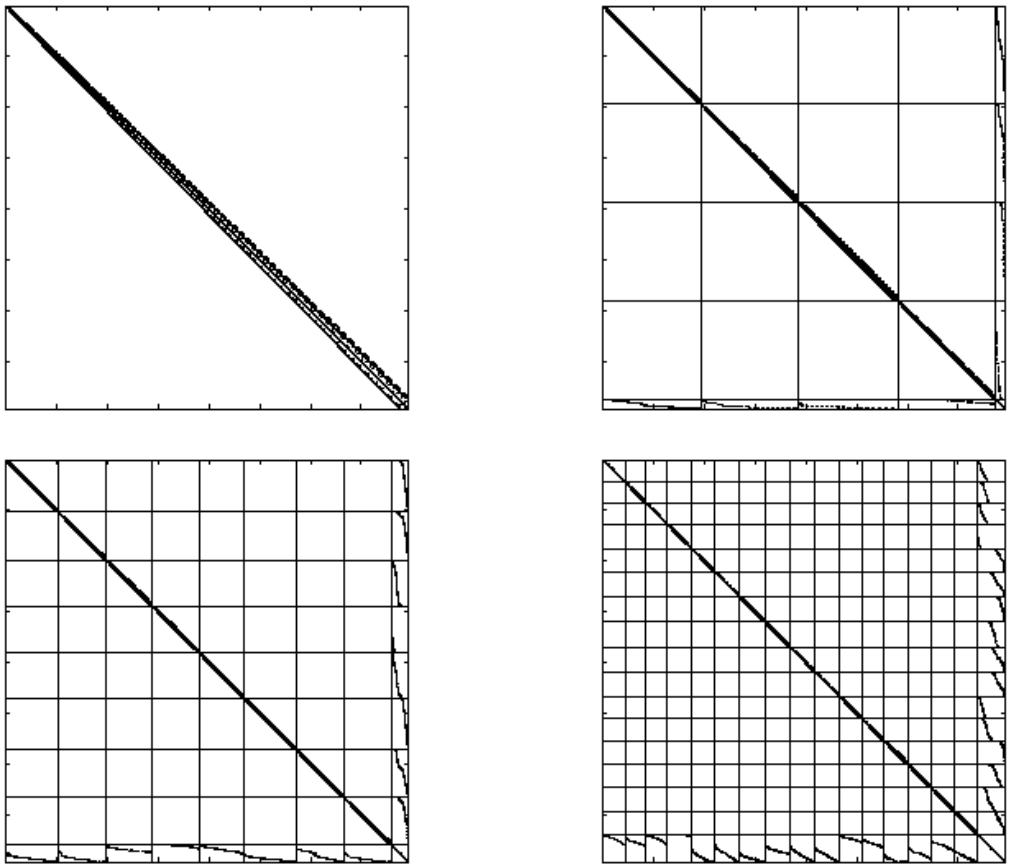
$$\hat{\mathcal{A}} = \left[\begin{array}{cccc|c} A_1 & & & & B_1 \\ & A_2 & & & B_2 \\ & & \ddots & & \vdots \\ & & & A_p & B_p \\ \hline C_1 & C_2 & \cdots & C_p & D \end{array}\right]$$

by permuting the rows and columns associated with the vertices in $\bigcup_k V_k$ before the rows and columns associated with the vertices in the separator set $V_S$. That is, $V_S$ defines the rows and columns of the (2,2) block $D \equiv A_S$.

Hence, the size of the separator set is $|V_S| = m$. This is the number of unknown in the reduced (Schur complement) system,

$$\mathcal{S} x_S = h, \quad \text{where} \quad \mathcal{S} = D - CA^{-1}B.$$

# Some Examples (cont.)



## Some Examples (cont.)

The following important properties hold:

- If $\mathcal{A}$ is SPD, so is $\mathcal{S}$;
- If $\mathcal{A}$ is an $M$-matrix, so is $\mathcal{S}$;
- If $\mathcal{A}$ is a discretization of a 2nd order, elliptic operator, then $\mathcal{S}$ is usually much better conditioned than $\mathcal{A}$; typically

$$\kappa(\mathcal{A}) = O(h^{-2}), \qquad \kappa(\mathcal{S}) = O(h^{-1})$$

  where $h$ denotes the mesh size.

## Some Examples (cont.)

The last example is from transport. Denoting the vector of angular fluxes with $\psi$, the discretized transport equation can be written as

$$L\psi = MSD\psi + q\,,$$

where $L$ is the discrete streaming and removal operator, $M$ maps a vector of scalar flux moments onto angular fluxes, $S$ is a diagonal matrix of scattering cross sections on the cells, $D = M^T W$ where $W$ is a diagonal matrix of quadrature weights, and $q$ is a source.

Note that $M$ and $D$ are rectangular. Furthermore, $\phi = D\psi$ is the vector of scalar fluxes, of size much smaller than $\psi$.

## Some Examples (cont.)

Now the equations $L\psi = MSD\psi + q$ and $\phi = D\psi$ can be written as a $2 \times 2$ block matrix system:

$$\begin{bmatrix} L & -MS \\ -D & I \end{bmatrix} \begin{bmatrix} \psi \\ \phi \end{bmatrix} = \begin{bmatrix} q \\ 0 \end{bmatrix}, \quad \text{or} \quad \mathcal{A}x = b\,.$$

Eliminating the angular fluxes leads to the reduced system for the scalar fluxes:

$$(I - DL^{-1}MS)\phi = DL^{-1}q\,, \quad \text{or} \quad \mathcal{S}\phi = g\,.$$

This is a fixed point problem for the (discretized) compact operator $DL^{-1}MS$, leading to source iteration:

$$\phi_{k+1} = DL^{-1}MS\phi_k + g\,, k = 0, 1, \ldots$$

This iteration is often combined with DSA preconditioning and Krylov acceleration.

## Some Examples (cont.)

At each iteration, the action of the inverse of the discrete streaming and removal operator $L$ on a vector must be evaluated. This is done by sweeping and can be expensive, especially for calculations involving unstructured grids.

On parallel computers, appropriate schedules have to be found; these should be of minimal length. This is a hard problem, although good heuristics exist.

Here we note that computing the action of $L^{-1}$ can be done by an inner iteration and it is of interest to determine how accurate this evaluation needs to be.

# Inexact Methods

Besides the case of Schur complement systems, there are other situations where the exact evaluation of matrix-vector products may be expensive or even impossible.

An important example occurs in computational electromagnetics, where $\mathcal{A}$ represents the discretization of a non-local operator, for example, a boundary integral operator.

In this case the exact evaluation of matrix-vector products with $\mathcal{A}$ is prohibitively expensive, but fast multipole methods can be used to approximate the action of $\mathcal{A}$ on a given vector.

# Inexact Methods (cont.)

Another important example is that of the inverse power method for evaluating the smallest (generalized) eigenvalue of a matrix or, more generally, shift-and-invert methods for approximating interior eigenvalues. These methods take the form

$$x_{k+1} = (\mathcal{A} - \mu\mathcal{B})^{-1} r_k, \quad k = 0, 1, \ldots,$$

followed by a normalization of the iterate.

The above mat-vec requires solving the linear system

$$(\mathcal{A} - \mu\mathcal{B})x = r_k \,,$$

which is usually done by a direct method. When $\mathcal{A}$ and $\mathcal{B}$ are large, however, this becomes impractical and an inner iteration must be used.

# Inexact Methods (cont.)

Around the year 2000, several researchers (Golub, Ye, Bouras, Frayssé, Giraud...) observed that it is often possible to terminate the inner iteration early on, without sacrificing the accuracy and convergence rate of the outer iteration process.

More precisely, at step $k$ of the basic "Krylov" method the mat-vec $z_k = \mathcal{A}v_k$ can be replaced by

$$\widehat{z}_k = (A + E_k)v = z_k + f_k \,,$$

where $E_k$ is an error matrix and $f_k = E_k v$.

Surprisingly, it was experimentally discovered by Bouras et al. that the magnitude $\|E_k\|$ can be quite large. The key is that $\|E_k\|$ must vary in a certain way with $k$.

# Inexact Methods (cont.)

It turns out that the norm of the error matrix $\|E_k\|$ should be small in the first iterations of the Krylov method, after which it can grow.

In the context of minimal residual methods, let $\widehat{r}_m$ denote the computed residual at step $m$; then if

$$\|f_k\| \leq c_k \frac{\varepsilon}{\|\widehat{r}_{k-1}\|} \quad \text{for} \quad k \leq m\,,$$

where the $c_k$'s are suitable problem-dependent constants, then one has $\|r_m - \widehat{r}_m\| \leq \varepsilon$, where $r_m$ is the true residual (with the "exact" $A$). Moreover, the convergence properties are the same as for the exact ($E_k = 0$) method!

The constants $c_k$, however, are not readily computable.

# Inexact Methods (cont.)

Bouras et al. showed that taking $c_k \equiv 1$ for all $k$ resulted in good results in some problems, and poor results (in terms of attained accuracy) in others.

In the context of radiation transport problems, we (Warsa et al.) have used an inner-outer iteration strategy with variable inner residual tolerance given by

$$\varepsilon_{inner} = \tau \frac{\varepsilon_{outer}}{\|\widehat{r}_k\|}, \quad \text{where} \quad \tau \approx 0.1.$$

This is a very conservative strategy, but it is reliable and it results in significant savings in overall solution times. It is now implemented in some of the LANL transport codes.

# Inexact Methods (cont.)

Inexact Krylov methods have been successfully applied for solving linear systems and eigenvalue problems, and for approximating functions of matrices.

A complete theory of these methods has been given in recent papers by Simoncini & Szyld and, independently, by Sleijpen & van den Eshof.

Roughly speaking, the early mat-vec products must be computed to high accuracy so that the appropriate invariant subspace is well-approximated; after that, the mat-vecs are allowed to become more and more inexact. If restarts are used, high accuracy is required at the beginning of each cycle.

# Inexact Methods (cont.)

We now sketch a more formal explanation for the case of inexact GMRES. Instead of the standard Arnoldi relation $AV_m = V_{m+1}H_{m+1,m}$, we now have

$$AV_m = V_{m+1}H_{m+1,m} - [E_1 v_1\,, E_2 v_2\, \ldots, E_m v_m].$$

Let $y_m = [\eta_1^{(m)}\,, \eta_2^{(m)}\, \ldots, \eta_m^{(m)}]^T$ be such that $x_m = V_m y_m$, then

$$\|r_m - \widehat{r}_m\| \leq \sum_{k=1}^{m} |\eta_k^{(m)}| \|E_k\|$$

and

$$\|W_m^T r_m\| \leq \|H_{m+1,m}\| \sum_{k=1}^{m} |\eta_k^{(m)}| \|E_k\|.$$

# Inexact Methods (cont.)

The first inequality is the residual gap, whereas the second measures how far we are from the "exact" case (where we would have $W_m^T r_m = 0$.)

Clearly, if the $|\eta_k^{(m)}|$ are small, then the right-hand sides in the above inequalities can be small even if the norms $\|E_k\|$ are not. All one needs is the products $|\eta_k^{(m)}|\|E_k\|$ to be small.

It turns out that the magnitude of the components of $y_m$ is decreasing: $|\eta_1^{(m)}| \geq |\eta_2^{(m)}| \geq \ldots \geq |\eta_m^{(m)}|$, so that the magnitude of the error matrices can grow as the iteration progresses.

The quantity $|\eta_k^{(m)}|$ is of the order of the residual at the $(k-1)$th step. See Simoncini & Szyld for details.

# Inexact Methods (cont.)

The theory also covers the case of flexible Krylov methods with variable preconditioning: the mat-vecs with the preconditioned matrix $\mathcal{A}\mathcal{P}_k^{-1}$ can be interpreted as inexact mat-vecs.

This is useful, for instance, for preconditioners based on approximate Schur complements, or for DSA-type preconditioners in transport.

An interesting open question is whether the theory can be used to justify the use of non-deterministic (Monte Carlo) solvers as "inner" approximate solvers—for example, for performing sweeps in transport calculations.

# References


1. M. Benzi, *Preconditioning techniques for large linear systems: A survey*, J. Comput. Physics, 182 (2002), pp. 418–477.

2. M. Benzi, G. H. Golub, and J. Liesen, *Numerical solution of saddle point problems*, Acta Numerica 14 (2005), pp. 1–137.

3. V. Simoncini and D. B. Szyld, *Recent developments in Krylov subspace methods for linear systems*, Numer. Linear Algebra Appl., 14 (2007), pp. 1–59.

4. V. Simoncini and D. B. Szyld, *Theory of inexact Krylov subspace methods and applications to scientific computing*, SIAM J. Sci. Comput., 25 (2003), pp. 454–477.

5. J. Warsa, M. Benzi, T. Wareing, and J. Morel, *Preconditioning a mixed discontinuous finite element method for radiation diffusion*, Numer. Linear Algebra Appl., 11 (2004), pp. 795–811.