Block Krylov Space Solvers: a Survey

Martin H. Gutknecht

Seminar for Applied Mathematics
ETH Zurich

Nagoya University 8 Dec. 2005

Partly joint work with Thomas Schmelzer, Oxford University

Systems with multiple RHSs

Given is a nonsingular linear system with s RHSs,

$$\boxed{\mathbf{A}\mathbf{x} = \mathbf{b}} \tag{1}$$

where

$$\mathbf{A} \in \mathbb{C}^{N \times N}, \qquad \mathbf{b} \in \mathbb{C}^{N \times s}, \qquad \mathbf{x} \in \mathbb{C}^{N \times s}. \tag{2}$$

Using Gauss elimination we can solve it much more efficiently than s single linear systems with different matrices, since the LU decomposition of $\mathbf{A}$ is computed only once.

It does not matter if all the RHSs are known at the beginning or are produced one after another while the systems are solved.

Systems with multiple RHSs

Given is a nonsingular linear system with s RHSs,

$$\boxed{\mathbf{A}\mathbf{x} = \mathbf{b}} \tag{1}$$

where

$$\mathbf{A} \in \mathbb{C}^{N \times N}, \qquad \mathbf{b} \in \mathbb{C}^{N \times s}, \qquad \mathbf{x} \in \mathbb{C}^{N \times s}. \tag{2}$$

Using Gauss elimination we can solve it much more efficiently than s single linear systems with different matrices, since the LU decomposition of $\mathbf{A}$ is computed only once.

It does not matter if all the RHSs are known at the beginning or are produced one after another while the systems are solved.

Systems with multiple RHSs (cont'd)

If iterative methods are applied, it is hard to solve (1)–(2) much faster than s systems with single RHS.

Two approaches:

- using the (iterative) solution of a **seed system** for solving subsequently the other systems faster,
- using **block iterations**: treat several RHSs at once.

In the second case, all RHSs are needed at once.

Most iterative methods are generalized easily to block methods, but the stability of block methods requires extra effort.
Block methods may be, but need not be much faster than solving the s systems separately.

Related iterative methods for eigenvalues allow us to find multiple eigenvalues and corresponding eigenspaces.

If iterative methods are applied, it is hard to solve (1)–(2) much faster than s systems with single RHS.

Two approaches:

- using the (iterative) solution of a **seed system** for solving subsequently the other systems faster,
- using **block iterations**: treat several RHSs at once.

In the second case, all RHSs are needed at once.

Most iterative methods are generalized easily to block methods, but the stability of block methods requires extra effort.
Block methods may be, but need not be much faster than solving the s systems separately.

Related iterative methods for eigenvalues allow us to find multiple eigenvalues and corresponding eigenspaces.

Block Krylov space methods

We seek approximate solutions of the form

$$\boxed{\mathbf{x}_n \in \mathbf{x}_0 + \mathcal{B}_n^{\square}(\mathbf{A}, \mathbf{r}_0)\,,} \tag{3}$$

where the **block Krylov space** $\mathcal{B}_n^{\square} :\equiv \mathcal{B}_n^{\square}(\mathbf{A}, \mathbf{r}_0)$ is defined by

$$\mathcal{B}_n^{\square}(\mathbf{A}, \mathbf{r}_0) :\equiv \textbf{block span}\,(\mathbf{r}_0, \mathbf{A}\mathbf{r}_0, \ldots, \mathbf{A}^{n-1}\mathbf{r}_0) \subset \mathbb{C}^{N\times s} \tag{4}$$

$$:\equiv \left\{ \sum_{k=0}^{n-1} \mathbf{A}^k \mathbf{r}_0 \gamma_k \,;\; \gamma_k \in \mathbb{C}^{s\times s}\ (k = 0, \ldots, n-1) \right\}. \tag{5}$$

DEFINITION. A (complex) **block vector** is a matrix $\mathbf{y} \in \mathbb{C}^{N\times s}$.

Hence, the elements of $\mathcal{B}_n^{\square}$ are block vectors.

Block Krylov space methods

We seek approximate solutions of the form

$$\boxed{\mathbf{x}_n \in \mathbf{x}_0 + \mathcal{B}_n^{\square}(\mathbf{A}, \mathbf{r}_0)\,,} \tag{3}$$

where the **block Krylov space** $\mathcal{B}_n^{\square} :\equiv \mathcal{B}_n^{\square}(\mathbf{A}, \mathbf{r}_0)$ is defined by

$$\mathcal{B}_n^{\square}(\mathbf{A}, \mathbf{r}_0) :\equiv \textbf{block span}\,(\mathbf{r}_0, \mathbf{A}\mathbf{r}_0, \ldots, \mathbf{A}^{n-1}\mathbf{r}_0) \subset \mathbb{C}^{N\times s} \tag{4}$$

$$:\equiv \left\{ \sum_{k=0}^{n-1} \mathbf{A}^k \mathbf{r}_0 \gamma_k \,;\; \gamma_k \in \mathbb{C}^{s\times s}\ (k = 0, \ldots, n-1) \right\}. \tag{5}$$

DEFINITION. A (complex) **block vector** is a matrix $\mathbf{y} \in \mathbb{C}^{N\times s}$.

Hence, the elements of $\mathcal{B}_n^{\square}$ are block vectors.

Block Krylov space methods (cont'd)

This means that for an individual approximation $x^{(j)}$ holds

$$x_n^{(j)} \in x_0^{(j)} + \mathcal{B}_n(\mathbf{A}, \mathbf{r}_0)\,, \tag{6}$$

where

$$\mathcal{B}_n :\equiv \mathcal{B}_n(\mathbf{A}, \mathbf{r}_0) :\equiv \mathcal{K}_n^{(1)} + \cdots + \mathcal{K}_n^{(s)}\,, \tag{7}$$

with the s "usual" Krylov spaces for the s systems,

$$\mathcal{K}_n^{(j)} :\equiv \mathcal{K}_n(\mathbf{A}, r_0^{(j)}) :\equiv \left\{ \sum_{k=0}^{n-1} \mathbf{A}^k r_0^{(j)} \beta_{k,j}\,;\; \beta_{k,j} \in \mathbb{C}\ (\forall k) \right\}. \tag{8}$$

In other words, each approximation $x^{(j)}$ is from a space that is as large as all s "usual" Krylov spaces together: $\dim \mathcal{B}_n \leq ns$.

$\mathcal{B}_n^{\square}$ is a Cartesian product of s copies of $\mathcal{B}_n$:

$$\mathcal{B}_n^{\square} = \underbrace{\mathcal{B}_n \times \cdots \times \mathcal{B}_n}_{s \text{ times}}\,.$$

This means that for an individual approximation $x^{(j)}$ holds

$$\boxed{x_n^{(j)} \in x_0^{(j)} + \mathcal{B}_n(\mathbf{A}, \mathbf{r}_0)\,,} \tag{6}$$

where

$$\boxed{\mathcal{B}_n :\equiv \mathcal{B}_n(\mathbf{A}, \mathbf{r}_0) :\equiv \mathcal{K}_n^{(1)} + \cdots + \mathcal{K}_n^{(s)}\,,} \tag{7}$$

with the s "usual" Krylov spaces for the s systems,

$$\mathcal{K}_n^{(j)} :\equiv \mathcal{K}_n(\mathbf{A}, r_0^{(j)}) :\equiv \left\{ \sum_{k=0}^{n-1} \mathbf{A}^k r_0^{(j)} \beta_{k,j}\,;\ \beta_{k,j} \in \mathbb{C}\ (\forall k) \right\}. \tag{8}$$

In other words, each approximation $x^{(j)}$ is from a space that is as large as all s "usual" Krylov spaces together: $\dim \mathcal{B}_n \leq ns$.

$\mathcal{B}_n^{\square}$ is a Cartesian product of s copies of $\mathcal{B}_n$:

$$\boxed{\mathcal{B}_n^{\square} = \underbrace{\mathcal{B}_n \times \cdots \times \mathcal{B}_n}_{s \text{ times}}\,.}$$

This means that for an individual approximation $x^{(j)}$ holds

$$x_n^{(j)} \in x_0^{(j)} + \mathcal{B}_n(\mathbf{A}, \mathbf{r}_0)\,, \tag{6}$$

where

$$\mathcal{B}_n :\equiv \mathcal{B}_n(\mathbf{A}, \mathbf{r}_0) :\equiv \mathcal{K}_n^{(1)} + \cdots + \mathcal{K}_n^{(s)}\,, \tag{7}$$

with the s "usual" Krylov spaces for the s systems,

$$\mathcal{K}_n^{(j)} :\equiv \mathcal{K}_n(\mathbf{A}, r_0^{(j)}) :\equiv \left\{ \sum_{k=0}^{n-1} \mathbf{A}^k r_0^{(j)} \beta_{k,j}\,;\ \beta_{k,j} \in \mathbb{C}\ (\forall k) \right\}. \tag{8}$$

In other words, each approximation $x^{(j)}$ is from a space that is as large as all s "usual" Krylov spaces together: $\dim \mathcal{B}_n \leq ns$.

$\mathcal{B}_n^{\square}$ is a Cartesian product of s copies of $\mathcal{B}_n$:

$$\mathcal{B}_n^{\square} = \underbrace{\mathcal{B}_n \times \cdots \times \mathcal{B}_n}_{s \text{ times}}\,.$$

Block Krylov space methods

Main reasons for using block Krylov spaces:

- The search space for each $x^{(j)}$ is much bigger, namely as big as all s Krylov spaces together.
  But do these extra dimensions really help much?
- In some implementations, s matrix-vector products with $\mathbf{A}$ can be computed at once, and this is much faster than s separate matrix-vector products, even on sequential computers (due to better usage of cached data).

Work on block methods started in the 1970ies with block Lanczos for symmetric EVal problems and block CG.

Nearly all work for nonsymmetric matrices since 1990.

There are on the order of 100 publications on block methods.

Block Krylov space methods (cont'd)

Main reasons for using block Krylov spaces:

- The search space for each $x^{(j)}$ is much bigger, namely as big as all s Krylov spaces together.
  But do these extra dimensions really help much?
- In some implementations, s matrix-vector products with $\mathbf{A}$ can be computed at once, and this is much faster than s separate matrix-vector products, even on sequential computers (due to better usage of cached data).

Work on block methods started in the 1970ies with block Lanczos for symmetric EVal problems and block CG.

Nearly all work for nonsymmetric matrices since 1990.

There are on the order of 100 publications on block methods.

Deflation

The extra challenge comes from the **possible linear dependence of the residuals** (of the s systems).

In most block methods such a dependence **requires** an explicit reduction of the number of RHSs. We call this **deflation**.

(The term "deflation" is also used with different meanings.)

In the literature on block methods deflation is only treated in a few papers, and there are hardly any investigations about its necessity and its effects.

Deflation may be possible at startup or in a later step.

In particular: when "one of the systems converges".
Actually: when **"a linear combination of the s systems converges"**.

Deflation

The extra challenge comes from the **possible linear dependence of the residuals** (of the *s* systems).

In most block methods such a dependence **requires** an explicit reduction of the number of RHSs. We call this **deflation**.

(The term "deflation" is also used with different meanings.)

In the literature on block methods deflation is only treated in a few papers, and there are hardly any investigations about its necessity and its effects.

Deflation may be possible at startup or in a later step.

In particular: when "one of the systems converges".
Actually: when **"a linear combination of the *s* systems converges"**.

EXAMPLES (of extreme cases)

1. $\mathbf{r}_0$ is made up of s identical vectors r,

$$\mathbf{r}_0 := \begin{pmatrix} r & r & r & \dots & r \end{pmatrix} .$$

These might come from different $b^{(i)}$ and suitably chosen $x_0^{(i)}$:

$$r = b^{(i)} - \mathbf{A}x_0^{(i)} \qquad (i = 1, \dots, s)$$

Here, it suffices to solve one system.

2. $$\mathbf{r}_0 := \begin{pmatrix} r & \mathbf{A}r & \mathbf{A}^2 r & \dots & \mathbf{A}^{s-1} r \end{pmatrix} .$$

Here, even if $\operatorname{rank} \mathbf{r}_0 = s$, still
$\operatorname{rank} \begin{pmatrix} \mathbf{r}_0 & \mathbf{A}\mathbf{r}_0 \end{pmatrix} \leq s + 1$.

3. $\mathbf{r}_0$ has s columns that are linear combinations of s eigenvectors of $\mathbf{A}$. Then $\operatorname{rank} \begin{pmatrix} \mathbf{r}_0 & \mathbf{A}\mathbf{r}_0 \end{pmatrix} \leq s$.
Hence, one block iteration is enough to solve all systems.
A non-block solver may require s^2 iterations.

EXAMPLES (of extreme cases)

1. $\mathbf{r}_0$ is made up of s identical vectors r,
$$\mathbf{r}_0 := \begin{pmatrix} r & r & r & \dots & r \end{pmatrix} .$$
These might come from different $b^{(i)}$ and suitably chosen $x_0^{(i)}$:
$$r = b^{(i)} - \mathbf{A}x_0^{(i)} \qquad (i = 1, \dots, s)$$
Here, it suffices to solve one system.

2. $$\mathbf{r}_0 := \begin{pmatrix} r & \mathbf{A}r & \mathbf{A}^2 r & \dots & \mathbf{A}^{s-1} r \end{pmatrix} .$$
Here, even if $\operatorname{rank} \mathbf{r}_0 = s$, still
$\operatorname{rank} \begin{pmatrix} \mathbf{r}_0 & \mathbf{A}\mathbf{r}_0 \end{pmatrix} \leq s + 1$.

3. $\mathbf{r}_0$ has s columns that are linear combinations of s eigenvectors of $\mathbf{A}$. Then $\operatorname{rank} \begin{pmatrix} \mathbf{r}_0 & \mathbf{A}\mathbf{r}_0 \end{pmatrix} \leq s$.
Hence, one block iteration is enough to solve all systems.
A non-block solver may require s^2 iterations.

EXAMPLES (of extreme cases)

1. $\mathbf{r}_0$ is made up of s identical vectors r,

$$\mathbf{r}_0 := \begin{pmatrix} r & r & r & \ldots & r \end{pmatrix}.$$

These might come from different $b^{(i)}$ and suitably chosen $x_0^{(i)}$:

$$r = b^{(i)} - \mathbf{A}x_0^{(i)} \qquad (i = 1, \ldots, s)$$

Here, it suffices to solve one system.

2. $$\mathbf{r}_0 := \begin{pmatrix} r & \mathbf{A}r & \mathbf{A}^2 r & \ldots & \mathbf{A}^{s-1} r \end{pmatrix}.$$

Here, even if $\operatorname{rank} \mathbf{r}_0 = s$, still $\operatorname{rank} \begin{pmatrix} \mathbf{r}_0 & \mathbf{A}\mathbf{r}_0 \end{pmatrix} \leq s + 1$.

3. $\mathbf{r}_0$ has s columns that are linear combinations of s eigenvectors of $\mathbf{A}$. Then $\operatorname{rank} \begin{pmatrix} \mathbf{r}_0 & \mathbf{A}\mathbf{r}_0 \end{pmatrix} \leq s$.
Hence, one block iteration is enough to solve all systems.
A non-block solver may require s^2 iterations.

The grade

Recall from *single RHS case* ($s = 1$):

Characteristic properties of **grade** $\bar{\nu}(\mathbf{y}, \mathbf{A})$ of $\mathbf{y}$ with resp. to $\mathbf{A}$:

- $$\dim \mathcal{K}_n(\mathbf{A}, \mathbf{y}) = \begin{cases} n & \text{if} \quad n \leq \bar{\nu}\,, \\ \bar{\nu} & \text{if} \quad n \geq \bar{\nu}\,; \end{cases}$$
- $$\bar{\nu} = \min \left\{ n \;\middle|\; \dim \mathcal{K}_n(\mathbf{A}, \mathbf{y}) = \dim \mathcal{K}_{n+1}(\mathbf{A}, \mathbf{y}) \right\} \;;$$
- $$\bar{\nu} = \min \left\{ n \;\middle|\; \mathbf{A}^{-1}\mathbf{y} \in \mathcal{K}_n(\mathbf{A}, \mathbf{y}) \right\} \leq \partial\widehat{\chi}_{\mathbf{A}},$$

  where $\partial\widehat{\chi}_{\mathbf{A}} :\equiv$ degree of minimal polynomial of $\mathbf{A}$;
- $$\bar{\nu} = \min \left\{ n \;\middle|\; \mathbf{x}_\star \in \mathbf{x}_0 + \mathcal{K}_n(\mathbf{A}, \mathbf{r}_0) \right\} \;,$$

  where $\mathbf{A}\mathbf{x}_\star = \mathbf{b}$, $\mathbf{r}_0 :\equiv \mathbf{b} - \mathbf{A}\mathbf{x}_0$.

The grade

Recall from *single RHS case* ($s = 1$):

Characteristic properties of **grade** $\bar{\nu}(\mathbf{y}, \mathbf{A})$ of $\mathbf{y}$ with resp. to $\mathbf{A}$:

- $$\dim \mathcal{K}_n(\mathbf{A}, \mathbf{y}) = \begin{cases} n & \text{if} \quad n \leq \bar{\nu}\,, \\ \bar{\nu} & \text{if} \quad n \geq \bar{\nu}\,; \end{cases}$$

- $$\bar{\nu} = \min \left\{ n \;\middle|\; \dim \mathcal{K}_n(\mathbf{A}, \mathbf{y}) = \dim \mathcal{K}_{n+1}(\mathbf{A}, \mathbf{y}) \right\} \;;$$

- $$\bar{\nu} = \min \left\{ n \;\middle|\; \mathbf{A}^{-1}\mathbf{y} \in \mathcal{K}_n(\mathbf{A}, \mathbf{y}) \right\} \leq \partial\widehat{\chi}_{\mathbf{A}},$$

  where $\partial\widehat{\chi}_{\mathbf{A}} :\equiv$ degree of minimal polynomial of $\mathbf{A}$;

- $$\bar{\nu} = \min \left\{ n \;\middle|\; \mathbf{x}_\star \in \mathbf{x}_0 + \mathcal{K}_n(\mathbf{A}, \mathbf{r}_0) \right\} \;,$$

  where $\mathbf{A}\mathbf{x}_\star = \mathbf{b}$, $\mathbf{r}_0 :\equiv \mathbf{b} - \mathbf{A}\mathbf{x}_0$.

The grade

Recall from *single RHS case* ($s = 1$):

Characteristic properties of **grade** $\bar{\nu}(\mathbf{y}, \mathbf{A})$ of $\mathbf{y}$ with resp. to $\mathbf{A}$:

- $$\dim \mathcal{K}_n(\mathbf{A}, \mathbf{y}) = \begin{cases} n & \text{if} \quad n \leq \bar{\nu}\,, \\ \bar{\nu} & \text{if} \quad n \geq \bar{\nu}\,; \end{cases}$$
- $$\bar{\nu} = \min \left\{ n \;\middle|\; \dim \mathcal{K}_n(\mathbf{A}, \mathbf{y}) = \dim \mathcal{K}_{n+1}(\mathbf{A}, \mathbf{y}) \right\} \;;$$
- $$\bar{\nu} = \min \left\{ n \;\middle|\; \mathbf{A}^{-1}\mathbf{y} \in \mathcal{K}_n(\mathbf{A}, \mathbf{y}) \right\} \leq \partial\widehat{\chi}_{\mathbf{A}},$$

  where $\partial\widehat{\chi}_{\mathbf{A}} :\equiv$ degree of minimal polynomial of $\mathbf{A}$;
- $$\bar{\nu} = \min \left\{ n \;\middle|\; \mathbf{x}_\star \in \mathbf{x}_0 + \mathcal{K}_n(\mathbf{A}, \mathbf{r}_0) \right\} \;,$$

  where $\mathbf{A}\mathbf{x}_\star = \mathbf{b}$, $\mathbf{r}_0 :\equiv \mathbf{b} - \mathbf{A}\mathbf{x}_0$.

The grade

Recall from *single RHS case* ($s = 1$):

Characteristic properties of **grade** $\bar{\nu}(\mathbf{y}, \mathbf{A})$ of $\mathbf{y}$ with resp. to $\mathbf{A}$:

- $$\dim \mathcal{K}_n(\mathbf{A}, \mathbf{y}) = \begin{cases} n & \text{if} \quad n \leq \bar{\nu}\,, \\ \bar{\nu} & \text{if} \quad n \geq \bar{\nu}\,; \end{cases}$$
- $$\bar{\nu} = \min \left\{ n \,\middle|\, \dim \mathcal{K}_n(\mathbf{A}, \mathbf{y}) = \dim \mathcal{K}_{n+1}(\mathbf{A}, \mathbf{y}) \right\} \;;$$
- $$\bar{\nu} = \min \left\{ n \,\middle|\, \mathbf{A}^{-1}\mathbf{y} \in \mathcal{K}_n(\mathbf{A}, \mathbf{y}) \right\} \leq \partial\widehat{\chi}_{\mathbf{A}},$$

  where $\partial\widehat{\chi}_{\mathbf{A}} :\equiv$ degree of minimal polynomial of $\mathbf{A}$;
- $$\bar{\nu} = \min \left\{ n \,\middle|\, \mathbf{x}_\star \in \mathbf{x}_0 + \mathcal{K}_n(\mathbf{A}, \mathbf{r}_0) \right\} \;,$$

  where $\mathbf{A}\mathbf{x}_\star = \mathbf{b}$, $\mathbf{r}_0 :\equiv \mathbf{b} - \mathbf{A}\mathbf{x}_0$.

The block grade

In *multiple RHS case* ($s > 1$):

Introduce **block grade** $\bar{\nu}(\mathbf{y}, \mathbf{A})$ of $\mathbf{y}$ with respect to $\mathbf{A}$ with characteristic properties:

- $\bar{\nu} = \min \left\{ n \,\middle|\, \dim \mathcal{B}_n(\mathbf{A}, \mathbf{y}) = \dim \mathcal{B}_{n+1}(\mathbf{A}, \mathbf{y}) \right\}$;
- $\bar{\nu} = \min \left\{ n \,\middle|\, \mathbf{A}^{-1}\mathbf{y} \in \mathcal{B}_n^{\square}(\mathbf{A}, \mathbf{y}) \right\} \leq \partial\widehat{\chi}_{\mathbf{A}}$,

  where $\partial\widehat{\chi}_{\mathbf{A}} :\equiv$ degree of minimal polynomial of $\mathbf{A}$;
- $\bar{\nu} = \min \left\{ n \,\middle|\, \mathbf{x}_\star \in \mathbf{x}_0 + \mathcal{B}_n^{\square}(\mathbf{A}, \mathbf{r}_0) \right\}$,

  where $\mathbf{A}\mathbf{x}_\star = \mathbf{b}$, $\mathbf{r}_0 :\equiv \mathbf{b} - \mathbf{A}\mathbf{x}_0$.

The block grade

In *multiple RHS case* ($s > 1$):

Introduce **block grade** $\bar{\nu}(\mathbf{y}, \mathbf{A})$ of $\mathbf{y}$ with respect to $\mathbf{A}$ with characteristic properties:

- $$\boxed{\bar{\nu} = \min\left\{ n \,\middle|\, \dim\ \mathcal{B}_n(\mathbf{A}, \mathbf{y}) = \dim\ \mathcal{B}_{n+1}(\mathbf{A}, \mathbf{y}) \right\}\ ;}$$
- $$\boxed{\bar{\nu} = \min\left\{ n \,\middle|\, \mathbf{A}^{-1}\mathbf{y} \in \mathcal{B}_n^{\square}(\mathbf{A}, \mathbf{y}) \right\} \leq \partial\widehat{\chi}_{\mathbf{A}},}$$

  where $\partial\widehat{\chi}_{\mathbf{A}} :\equiv$ degree of minimal polynomial of $\mathbf{A}$;
- $$\boxed{\bar{\nu} = \min\left\{ n \,\middle|\, \mathbf{x}_\star \in \mathbf{x}_0 + \mathcal{B}_n^{\square}(\mathbf{A}, \mathbf{r}_0) \right\}\ ,}$$

  where $\mathbf{A}\mathbf{x}_\star = \mathbf{b}$, $\mathbf{r}_0 :\equiv \mathbf{b} - \mathbf{A}\mathbf{x}_0$.

The block grade

In *multiple RHS case* ($s > 1$):

Introduce **block grade** $\bar{\nu}(\mathbf{y}, \mathbf{A})$ of $\mathbf{y}$ with respect to $\mathbf{A}$ with characteristic properties:

- $$\bar{\nu} = \min \left\{ n \,\middle|\, \dim\ \mathcal{B}_n(\mathbf{A}, \mathbf{y}) = \dim\ \mathcal{B}_{n+1}(\mathbf{A}, \mathbf{y}) \right\} \ ;$$
- $$\bar{\nu} = \min \left\{ n \,\middle|\, \mathbf{A}^{-1}\mathbf{y} \in \mathcal{B}_n^{\square}(\mathbf{A}, \mathbf{y}) \right\} \leq \partial\widehat{\chi}_{\mathbf{A}},$$

  where $\partial\widehat{\chi}_{\mathbf{A}} :\equiv$ degree of minimal polynomial of $\mathbf{A}$;
- $$\bar{\nu} = \min \left\{ n \,\middle|\, \mathbf{x}_\star \in \mathbf{x}_0 + \mathcal{B}_n^{\square}(\mathbf{A}, \mathbf{r}_0) \right\} \ ,$$

  where $\mathbf{A}\mathbf{x}_\star = \mathbf{b}$, $\mathbf{r}_0 :\equiv \mathbf{b} - \mathbf{A}\mathbf{x}_0$.

In the single RHS case, in exact arithmetic, computing $\mathbf{x}_\star$ requires

$$\dim\ \mathcal{K}_{\bar{\nu}} = \bar{\nu} \quad \text{MVs.}$$

In the multiple RHS case, in exact arithmetic, computing $\mathbf{x}_\star$ requires

$$\dim\ \mathcal{B}_{\bar{\nu}}\ \in\ [\bar{\nu},\ s \cdot \bar{\nu}] \quad \text{MVs.}$$

This is a big interval!

Block methods are most effective (compared to single RHS methods) if

$$\dim\ \mathcal{B}_{\bar{\nu}} \ll s \cdot \bar{\nu}\,.$$

More exactly: block methods are most effective if

$$\dim\ \mathcal{B}_{\bar{\nu}(\mathbf{r}_0,\mathbf{A})} \ll \sum_{k=1}^{s} \dim\ \mathcal{K}_{\bar{\nu}(r_0^{(k)},\mathbf{A})}\,.$$

In other words: **block methods are most effective (compared to single RHS methods) if deflation is possible and used!**

However, exact deflation is rare, and we need approximate deflation depending on a **deflation tolerance** in RRQR.

Approximate deflation introduces a **deflation error**.

The deflation error may deteriorate the convergence speed and/or the accuracy of the computed solution.

Restarting the iteration can be useful from this point of view.

Symmetric block Lanczos algorithm

In the 1970ies a number of people started around the same time with block Lanczos for symmetric EVal problems.

It is hard to tell now who had the idea first.

Cullum/Donath [[*IEEE Decision Control*/'**74**], [*'74*]
(symmetric, EV)

Kahan/Parlett [*Sparse Matrix Comp.*/'**76**] *(symmetric, EV)*

Underwood ['**75**$_{\text{Diss}}$] *(symmetric, EV + CG)*

Golub/Underwood [*Math. Software*/'**77**] *(symmetric, EV)*

Lewis ['**77**$_{\text{Diss}}$] *(symmetric)*

Cullum ['**78**$_{\text{BIT}}$] *(symmetric, EV)*

Algorithm (Symmetric block Lanczos algorithm)

Start: *Given* $\widetilde{\mathbf{y}}_0 \in \mathbb{C}^{N\times s}$ *let*

$$\mathbf{y}_0\ \rho_0 := \widetilde{\mathbf{y}}_0 \qquad (\text{QR factorization:} \quad \rho_0 \in \mathbb{C}^{s\times s}, \quad \mathbf{y}_0 \in \mathbb{C}^{N\times s})$$

Loop:

for $n = 1, 2, \ldots$ **do**

$$\widetilde{\mathbf{y}} := \mathbf{A}\mathbf{y}_{n-1} \qquad (s \text{ MVs in parallel})$$

$$\widetilde{\mathbf{y}} := \widetilde{\mathbf{y}} - \mathbf{y}_{n-2}\beta^\star_{n-2} \quad \text{if } n > 1 \qquad (s^2 \text{ SAXPYs in parallel})$$

$$\alpha_{n-1} := \mathbf{y}^\star_{n-1}\widetilde{\mathbf{y}} \qquad (s^2 \text{ SDOTs in parallel})$$

$$\widetilde{\mathbf{y}} := \widetilde{\mathbf{y}} - \mathbf{y}_{n-1}\alpha_{n-1} \qquad (s^2 \text{ SAXPYs in parallel})$$

$$\mathbf{y}_n\ \beta_{n-1} := \widetilde{\mathbf{y}} \qquad (\text{QR factorization:} \quad \beta_{n-1} \in \mathbb{C}^{s\times s})$$

end

Need to add stopping criterion and deflation.

Symmetric block Lanczos with deflation

Deflation (not [?] treated in old papers): We apply in both

$$\underbrace{\mathbf{y}_0}_{\mathbf{Q}} \underbrace{\rho_0}_{\mathbf{R}} := \widetilde{\mathbf{y}}_0 \quad \text{and} \quad \underbrace{\mathbf{y}_n}_{\mathbf{Q}} \underbrace{\beta_{n-1}}_{\mathbf{R}} := \widetilde{\mathbf{y}}$$

a *(high) rank-revealing QR factorization (RRQR)*.

Columns in $\mathbf{y}_0$ or $\mathbf{y}_n$ that are multiplied only with small elements of ρ_0 or $\eta_{n,n-1}$, respectively, can be deleted $\rightsquigarrow$ **deflation**.

s is replaced by s_n, where $s \geq s_0 \geq s_1 \geq \ldots$.

Two types: **initial deflation** and **Lanczos deflation**.

ρ_0 and β_{n-1} are upper triangular up to a column permutation.

In case of deflation ρ_0 and β_{n-1} are (nearly) singular.

Symmetric block Lanczos with deflation

Deflation (not [?] treated in old papers): We apply in both

$$\underbrace{\mathbf{y}_0}_{\mathbf{Q}} \underbrace{\rho_0}_{\mathbf{R}} := \widetilde{\mathbf{y}}_0 \qquad \text{and} \qquad \underbrace{\mathbf{y}_n}_{\mathbf{Q}} \underbrace{\boldsymbol{\beta}_{n-1}}_{\mathbf{R}} := \widetilde{\mathbf{y}}$$

a *(high) rank-revealing QR factorization (RRQR)*.

Columns in $\mathbf{y}_0$ or $\mathbf{y}_n$ that are multiplied only with small elements of ρ_0 or $\eta_{n,n-1}$, respectively, can be deleted $\rightsquigarrow$ **deflation**.

s is replaced by s_n, where $s \geq s_0 \geq s_1 \geq \ldots$.

Two types: **initial deflation** and **Lanczos deflation**.

ρ_0 and β_{n-1} are upper triangular up to a column permutation.

In case of deflation ρ_0 and β_{n-1} are (nearly) singular.

Symmetric block Lanczos with deflation

Deflation (not [?] treated in old papers): We apply in both

$$\underbrace{\mathbf{y}_0}_{\mathbf{Q}} \underbrace{\boldsymbol{\rho}_0}_{\mathbf{R}} := \widetilde{\mathbf{y}}_0 \qquad \text{and} \qquad \underbrace{\mathbf{y}_n}_{\mathbf{Q}} \underbrace{\boldsymbol{\beta}_{n-1}}_{\mathbf{R}} := \widetilde{\mathbf{y}}$$

a *(high) rank-revealing QR factorization (RRQR)*.

Columns in $\mathbf{y}_0$ or $\mathbf{y}_n$ that are multiplied only with small elements of $\boldsymbol{\rho}_0$ or $\boldsymbol{\eta}_{n,n-1}$, respectively, can be deleted $\leadsto$ **deflation**.

s is replaced by s_n, where $s \geq s_0 \geq s_1 \geq \ldots$.

Two types: **initial deflation** and **Lanczos deflation**.

$\boldsymbol{\rho}_0$ and $\boldsymbol{\beta}_{n-1}$ are upper triangular up to a column permutation.

In case of deflation $\boldsymbol{\rho}_0$ and $\boldsymbol{\beta}_{n-1}$ are (nearly) singular.

HHQR for Lanczos deflation in detail:

$$\widetilde{\mathbf{y}} =: \begin{pmatrix} \mathbf{y}_n & \mathbf{y}_n^{\Delta} \end{pmatrix} \begin{pmatrix} \rho_n & \rho_n^{\square} \\ \mathbf{o} & \rho_n^{\Delta} \end{pmatrix} \pi_n^{\mathsf{T}} =: \begin{pmatrix} \mathbf{y}_n & \mathbf{y}_n^{\Delta} \end{pmatrix} \begin{pmatrix} \beta_{n-1} \\ \beta_{n-1}^{\Delta} \end{pmatrix},$$

where:

- π_n is an $s_{n-1} \times s_{n-1}$ permutation matrix,
- $\mathbf{y}_n$ is an $N \times s_n$ block vector with full numerical column rank, which goes into the basis,
- $\mathbf{y}_n^{\Delta}$ is an $N \times (s_{n-1} - s_n)$ matrix that will be deflated (deleted),
- ρ_n is an $s_n \times s_n$ upper triangular, nonsingular matrix,
- $\rho_n^{\square}$ is an $s_n \times (s_{n-1} - s_n)$ matrix,
- ρ_n^{Δ} is an upper triangular $(s_{n-1} - s_n) \times (s_{n-1} - s_n)$ matrix with $\|\rho_n^{\Delta}\|_F = O(\sigma_{s_n+1})$, where σ_{s_n+1} is the largest singular value of $\widetilde{\mathbf{y}}$ smaller or equal to tol.

The fundamental block Lanczos relation $\mathbf{AY}_m = \mathbf{Y}_{m+1}\underline{\mathbf{T}}_m$ (with a block tridiagonal matrix $\underline{\mathbf{T}}_m$ extended at the bottom with s_m rows) is in case of inexact deflation replaced by

$$\mathbf{AY}_m = \mathbf{Y}_{m+1}\underline{\mathbf{T}}_m + \mathbf{Y}^{\Delta}_{m+1}\underline{\mathbf{T}}^{\Delta}_m \,,$$

where

$$\underline{\mathbf{T}}^{\Delta}_m :\equiv \begin{pmatrix} \mathbf{o} & \mathbf{o} & \cdots & \mathbf{o} \\ \beta^{\Delta}_0 & \mathbf{o} & \cdots & \mathbf{o} \\ & \beta^{\Delta}_1 & \ddots & \vdots \\ & & \ddots & \mathbf{o} \\ & & & \beta^{\Delta}_{m-1} \end{pmatrix}$$

is $(s - s_m) \times t_{m-1}$, where $t_m :\equiv \sum_{k=0}^{m} s_k$.

Is deflation important? **YES!**

The fundamental block Lanczos relation $\mathbf{AY}_m = \mathbf{Y}_{m+1}\underline{\mathbf{T}}_m$ (with a block tridiagonal matrix $\underline{\mathbf{T}}_m$ extended at the bottom with s_m rows) is in case of inexact deflation replaced by

$$\mathbf{AY}_m = \mathbf{Y}_{m+1}\underline{\mathbf{T}}_m + \mathbf{Y}^{\Delta}_{m+1}\underline{\mathbf{T}}^{\Delta}_m ,$$

where

$$\underline{\mathbf{T}}^{\Delta}_m :\equiv \begin{pmatrix} \mathbf{o} & \mathbf{o} & \cdots & \mathbf{o} \\ \beta^{\Delta}_0 & \mathbf{o} & \cdots & \mathbf{o} \\ & \beta^{\Delta}_1 & \ddots & \vdots \\ & & \ddots & \mathbf{o} \\ & & & \beta^{\Delta}_{m-1} \end{pmatrix}$$

is $(s - s_m) \times t_{m-1}$, where $t_m :\equiv \sum_{k=0}^{m} s_k$.

Is deflation important? **YES!**

Symmetric block Lanczos with deflation (cont'd)

The fundamental block Lanczos relation $\mathbf{A}\mathbf{Y}_m = \mathbf{Y}_{m+1}\underline{\mathbf{T}}_m$ (with a block tridiagonal matrix $\underline{\mathbf{T}}_m$ extended at the bottom with s_m rows) is in case of inexact deflation replaced by

$$\mathbf{A}\mathbf{Y}_m = \mathbf{Y}_{m+1}\underline{\mathbf{T}}_m + \mathbf{Y}_{m+1}^{\Delta}\underline{\mathbf{T}}_m^{\Delta},$$

where

$$\underline{\mathbf{T}}_m^{\Delta} :\equiv \begin{pmatrix} \boldsymbol{o} & \boldsymbol{o} & \cdots & \boldsymbol{o} \\ \beta_0^{\Delta} & \boldsymbol{o} & \cdots & \boldsymbol{o} \\ & \beta_1^{\Delta} & \ddots & \vdots \\ & & \ddots & \boldsymbol{o} \\ & & & \beta_{m-1}^{\Delta} \end{pmatrix}$$

is $(s - s_m) \times t_{m-1}$, where $t_m :\equiv \sum_{k=0}^{m} s_k$.

Is deflation important? **YES!**

Symmetric block Lanczos: numerical experiments

Experiment (1)

$\mathbf{A}$ is a sparse 100×100 random matrix.
In the block vector $\widetilde{\mathbf{y}}_0$ each of the first two columns is a random linear combinations of 20 distinct eigenvectors of $\mathbf{A}$. The third column is a linear combination of 5 other eigenvectors.
Hence these 45 eigenvectors are an orthonormal basis for the $\mathbf{A}$-invariant subspace

$$\mathcal{B}_{20}\left(\mathbf{A}, \widetilde{\mathbf{y}}_0\right) = \mathcal{K}_{20}\left(\mathbf{A}, \widetilde{\mathbf{y}}_0^{(1)}\right) \oplus \mathcal{K}_{20}\left(\mathbf{A}, \widetilde{\mathbf{y}}_0^{(2)}\right) \oplus \mathcal{K}_{5}\left(\mathbf{A}, \widetilde{\mathbf{y}}_0^{(3)}\right).$$

Constructing $\mathbf{y}_0, \ldots, \mathbf{y}_4$ we expect no problems. However, the Krylov subspace $\mathcal{K}_5\left(\mathbf{A}, \widetilde{\mathbf{y}}_0^{(3)}\right)$ is exhausted.
The smallest eigenvalue of β_4 is close to 10^{-10}.
Proceeding without deflation we construct a highly indetermined vector in order to complete the block vector $\mathbf{y}_5$.

One might hope that this vector does not disturb the Lanczos process, and that it does not influence the construction of the Krylov subspaces $\mathcal{K}_n\left(\mathbf{A}, \widetilde{\mathbf{y}}_0^{(1)}\right)$ and $\mathcal{K}_n\left(\mathbf{A}, \widetilde{\mathbf{y}}_0^{(2)}\right)$.

In particular one might hope that the corresponding columns in the block vector $\mathbf{y}_6$ remain orthogonal to all previously constructed vectors.

However, this experiment shows that the orthogonality is lost.

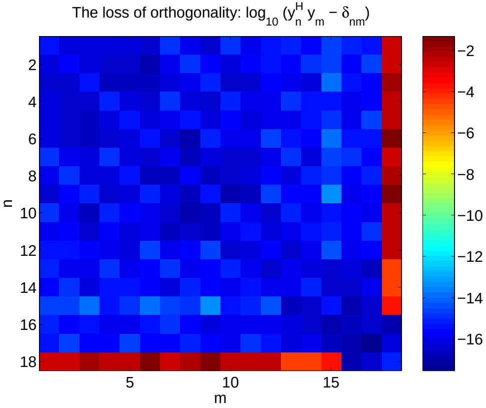

Figure: Experiment 1: The vector corresponding to a singular value of approximately 10^{-10} is highly indetermined. It is not orthogonal to the vectors of the previous blocks. However, it is orthogonal to the two other vectors of the block vector $\mathbf{y}_5$.

Symmetric block Lanczos: experiments

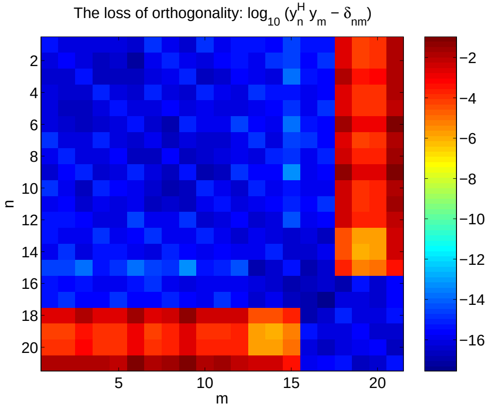

Figure: Experiment 1: The block vector $\mathbf{y}_6$ is far away from being orthogonal to all previous blocks.

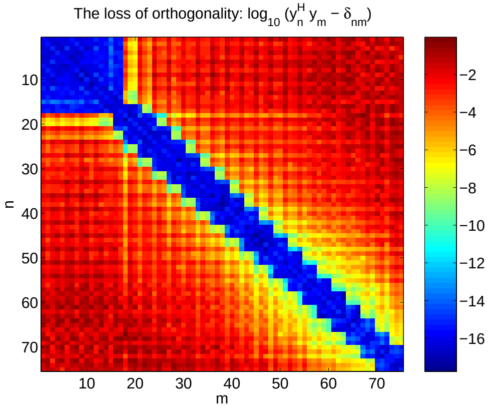

Figure: Experiment 1: Colormap of the matrix $\mathbf{V} = \log |\mathbf{Y}_{20}^{\star}\mathbf{Y}_{20} - \mathbf{I}_{20}|$. Orthogonality is completely lost after ignoring the exhausted Krylov space.

Block Lanczos and block CG: more references

O'Leary [*'78/'***80**$_{\text{LAA}}$] *(nonsym./symmetric: BiCG/CG)*
First statement of block BiCG, but there is only a very short discussion of the added problems in the nonsymmetric case.

Ruhe [**'79**$_{\text{MathComp}}$] *(symmetric, band, EV)*
Ruhe shows that the orthonormal basis can be built up vector by vector. He also discusses reorthogonalization: it suffices to reorthogonalize against $\mathbf{y}_{n-1}$.
However, his alg. does not allow RRQR: no pivoting possible. Therefore less stable than our current implementation.

Parlett [**'80**$_{\text{Book}}$] *(symmetric, block and band, EV)*

Saad [**'80**$_{\text{SINUM}}$] *(symmetric, EV, convergence)*

O'Leary [**'87**$_{\text{Par. Comp.}}$] *(symmetric)*

Block Lanczos and block CG: more references (cont'd)

Boley/Golub ['**91**$_{\text{Syst. Control Lett.}}$] *(nonsymmetric, control)*

Kim/Craig ['**90**$_{\text{Int. J. Num. Meth. Eng.}}$] *(nonsymmetric, EV)*

Broyden [*'92*/'**93**$_{\text{Optim. Methods Softw.}}$]
(sym., indef., nonsym., look-ahead)

Broyden [*'94*/'**95**$_{\text{Optim. Methods Softw.}}$]
(sym., indef., nonsym., look-ahead)

Grimes/Lewis/Simon [*'88*/'**94**$_{\text{SIMAX}}$] *(symmetric, EV)*

Kim/Chronopoulos ['**92**$_{\text{JCAM}}$] *(nonsymmetric)*

Ruiz ['**92**$_{\text{Diss}}$] *(Bl-CG and symmetric block Lanczos)*

Nikishin/Yeremin [*'93*/'**95**$_{\text{SIMAX}}$] *(symmetric, defl.)*
First detailed treatment of deflation for CG.

Block Lanczos and block CG: more references (cont'd)

Aliaga/Hernández/Boley [*Lanczos*/'**94**]
(nonsym., look-ahead (cluster), model red.)
Bai [*5th SIAM ALA*/'**94**] *(nonsym., EV, spectral trafo)*
Cullum [*Lanczos*/'**94**] *(symmetric, EV)*
Cullum [*Lanczos*/'**94**] *(nonsym., EV)*
Freund [*AT VIII*/'**95**] *(nonsym., band, matrix Padé)*

Boyse/Seidl [*'94*/'**96**$_{\text{SISC}}$] *(compl. symmetric, QMR)*
Simoncini [*'94*/'**97**$_{\text{SIMAX}}$]
(nonsym., block-2-term, band, deflation, QMR)
Ye [*'94*/'**96**$_{\text{Num. Alg.}}$] *(symmetric, EV, adapt. block size)*

Block Lanczos and block CG: more references (cont'd)

Aliaga/Boley/Freund/Hernández [*'96/*'**99/00**$_{\text{MathComp}}$]
(nonsym., band, defl., look-ahead, QMR)

Bai/Day/Ye [*'97/*'**99**$_{\text{SIMAX}}$]
(nonsym., EV, adapt. block size, look-ahead, $\rightsquigarrow$ `ABLE`*)*

Freund/Malhotra ['**97**$_{\text{LAA}}$] *(nonsym., band, defl.,* $\rightsquigarrow$ `Bl-QMR`*)*

Malhotra/Freund/Pinsky ['**97**$_{\text{Comp. Meth. Appl. Mech. Eng.}}$]
(appl. to radiation/scattering probs.)

Freund [*Systems, Control 21st Cent./*'**97**]
(nonsym., band, model red.)

Freund [*Appl. Comput. Control, Signals, Circuits/*'**99**]
(nonsym., band, model red.)

Freund [*'99/*'**00**$_{\text{JCAM}}$] *(nonsym., band, model red.)*

Freund [*'99/*'**01**$_{\text{JCAM}}$] *(nonsym., band, block Hankel, FOPs)*

Block Lanczos and block CG: more references (cont'd)

Broyden [**'97**$_{\text{Optim. Methods Softw.}}$] *(indef. sym., nonsym.)*
Broyden [*Alg. large scale lin. sys./***'98**] *(indef. sym., look-ahead)*
Dai [*'98*] *(symmetric)*
Dai [*'98*] *(nonsymmetric)*
El Guennouni/Jbilou/Sadok [*'99*] *(nonsym.)*
El Guennouni/Jbilou/Sadok [*'99*] *(BlBiCGStab)*
El Guennouni [**'00**$_{\text{Diss}}$] *(nonsym.)*
El Guennouni/Jbilou [*'00*]
(nonsym., bl/gl-BiCGStab, deflation, seed BiCGStab)
Jbilou/Sadok [*'97*] *(nonsym., global, Lanczos-based)*
Yeung/Chan [*'97/***'99**$_{\text{SISC}}$] *(nonsym., 1 eq., ML(k)BiCGStab)*

Block Lanczos and block CG: more references (cont'd)

Bai/Freund [*'00/'***01**$_{\text{SISC}}$] *(symmetric, band, EV, model red.)*
Bai/Freund [*'00/'***01**$_{\text{LAA}}$] *(nonsym.?, band, Padé, model red.)*
Baglama/Calvetti/Reichel [preprint]
(nonsym., implic. restarted)
Kilmer/Miller/Rappaport [*'99/'***01**$_{\text{SISC}}$]
(Bl-QMR combined with seeds)
Meerbergen/Scott [*'00*]
(sym., EV, partial reorth., impl. restarts, $\rightsquigarrow$ `EA16`*)*
Hsu [*'03*] *(symmetric, EV, block size choice)*

Block GMRES, block MINRES and related methods

It is seemingly straightforward to define and implement block GMRES (BLGMRES), but some questions come up quickly.

- First, we apply block Arnoldi process to create an orthonormal basis of $\mathcal{B}_n(\mathbf{A}, \mathbf{r}_0)$.
- Then, we determine simultaneously the coordinates of the s systems, *i.e.,* solve them at once in coordinate space.
- This requires to solve a least square problem with s RHSs in every iteration.
- To solve it we update the QR decomposition of a rectangular block Hessenberg matrix to which s columns and rows are added in every iteration.

For block MINRES (BLMINRES) we start instead from the symmetric block Lanczos process.

Block Arnoldi/GMRES with deflation: introduction

Algorithm (m STEPS OF BLOCK ARNOLDI ALGORITHM)

Start: Given $\widetilde{\mathbf{y}}_0 \in \mathbb{C}^{N\times s}$ *let*

$$\mathbf{y}_0\ \rho_0 := \widetilde{\mathbf{y}}_0 \qquad \text{(QR factorization:} \quad \rho_0 \in \mathbb{C}^{s\times s}, \quad \mathbf{y}_0 \in \mathbb{C}^{N\times s})$$

Loop:

for $n = 1$ **to** m **do**

$\widetilde{\mathbf{y}} := \mathbf{A}\mathbf{y}_{n-1}$ (s MVs in parallel)

for $k = 0$ **to** $n-1$ **do** (blockwise MGS)

$\eta_{k,n-1} := \mathbf{y}_k^\star\ \widetilde{\mathbf{y}}$ (s^2 SDOTs in parallel)

$\widetilde{\mathbf{y}} := \widetilde{\mathbf{y}} - \mathbf{y}_k\ \eta_{k,n-1}$ (s^2 SAXPYs in parallel)

end

$\mathbf{y}_n\ \eta_{n,n-1} := \widetilde{\mathbf{y}}$ (QR factorization: $\eta_{n,n-1} \in \mathbb{C}^{s\times s}$)

end

We apply in both

$$\underbrace{\mathbf{y}_0}_{\mathbf{Q}} \underbrace{\rho_0}_{\mathbf{R}} := \widetilde{\mathbf{y}}_0 \quad \text{and} \quad \underbrace{\mathbf{y}_n}_{\mathbf{Q}} \underbrace{\eta_{n,n-1}}_{\mathbf{R}} := \widetilde{\mathbf{y}}$$

a *(high) rank-revealing QR factorization (RRQR)*.

Columns in $\mathbf{y}_0$ or $\mathbf{y}_n$ that are multiplied only with small elements of ρ_0 or $\eta_{n,n-1}$, respectively, can be deleted $\rightsquigarrow$ **deflation**.

s is replaced by s_n, where $s \geq s_0 \geq s_1 \geq \ldots$.

Two types: **initial deflation** and **Arnoldi deflation**.

ρ_0 and $\eta_{n,n-1}$ are upper triangular up to a column permutation.

In case of deflation ρ_0 and $\eta_{n,n-1}$ are (nearly) singular.

We apply in both

$$\underbrace{\mathbf{y}_0}_{\mathbf{Q}} \; \underbrace{\rho_0}_{\mathbf{R}} := \widetilde{\mathbf{y}}_0 \qquad \text{and} \qquad \underbrace{\mathbf{y}_n}_{\mathbf{Q}} \; \underbrace{\eta_{n,n-1}}_{\mathbf{R}} := \widetilde{\mathbf{y}}$$

a *(high) rank-revealing QR factorization (RRQR)*.

Columns in $\mathbf{y}_0$ or $\mathbf{y}_n$ that are multiplied only with small elements of ρ_0 or $\eta_{n,n-1}$, respectively, can be deleted $\rightsquigarrow$ **deflation**.

s is replaced by s_n, where $s \geq s_0 \geq s_1 \geq \ldots$.

Two types: **initial deflation** and **Arnoldi deflation**.

ρ_0 and $\eta_{n,n-1}$ are upper triangular up to a column permutation.

In case of deflation ρ_0 and $\eta_{n,n-1}$ are (nearly) singular.

Block Arnoldi/GMRES with deflation: intro (cont'd)

We apply in both

$$\underbrace{\mathbf{y}_0}_{\mathbf{Q}} \underbrace{\rho_0}_{\mathbf{R}} := \widetilde{\mathbf{y}}_0 \qquad \text{and} \qquad \underbrace{\mathbf{y}_n}_{\mathbf{Q}} \underbrace{\eta_{n,n-1}}_{\mathbf{R}} := \widetilde{\mathbf{y}}$$

a *(high) rank-revealing QR factorization (RRQR)*.

Columns in $\mathbf{y}_0$ or $\mathbf{y}_n$ that are multiplied only with small elements of ρ_0 or $\eta_{n,n-1}$, respectively, can be deleted $\rightsquigarrow$ **deflation**.

s is replaced by s_n, where $s \geq s_0 \geq s_1 \geq \ldots$.

Two types: **initial deflation** and **Arnoldi deflation**.

ρ_0 and $\eta_{n,n-1}$ are upper triangular up to a column permutation.

In case of deflation ρ_0 and $\eta_{n,n-1}$ are (nearly) singular.

$$\mathbf{r}_n = \mathbf{Y}_{n+1} \underbrace{(\underline{\mathbf{e}}_1 \rho_0 - \underline{\mathbf{H}}_n \mathbf{k}_n)}_{\equiv: \mathbf{q}_n}$$

Ass.: $\underline{\mathbf{H}}_n$ *has full rank.*
(This is most likely even when some $\eta_{n,n-1}$ *is singular.)*

(1) **Initial deflation:**

$\mathbf{r}_0$ rank-deficient $\Longrightarrow$ ρ_0, $\mathbf{k}_n$, $\mathbf{q}_n$, $\mathbf{r}_n$, $\mathbf{x}_n - \mathbf{x}_0$ rank-def.

$\rightsquigarrow$ *initial deflation reduces # MVs, but introduces errors if not exact.*

$$\boxed{\mathbf{r}_n = \mathbf{Y}_{n+1} \underbrace{(\underline{\mathbf{e}}_1 \rho_0 - \underline{\mathbf{H}}_n \mathbf{k}_n)}_{\equiv: \, \mathbf{q}_n}}$$

Ass.: $\underline{\mathbf{H}}_n$ *has full rank.*
(This is most likely even when some $\eta_{n,n-1}$ *is singular.)*

(1) **Initial deflation:**

$\mathbf{r}_0$ rank-deficient $\Longrightarrow$ ρ_0, $\mathbf{k}_n$, $\mathbf{q}_n$, $\mathbf{r}_n$, $\mathbf{x}_n - \mathbf{x}_0$ rank-def.

$\rightsquigarrow$ *initial deflation reduces # MVs, but introduces errors if not exact*.

(2) **Arnoldi deflation:** $\widetilde{\mathbf{y}}$ in block Arnoldi rank-deficient

Rather unlikely, because we start from $\mathbf{A}\mathbf{y}_{n-1}$.

Unless we deflate, search space contains extra basis vectors:

$$\mathcal{R}(\mathbf{Y}_n) \supsetneq \mathcal{B}_n$$

But they are unlikely to help much, since the block solution lies in $\mathbf{x}_0 + \mathcal{B}_n$ for some n.

$\rightsquigarrow$ *Arnoldi deflation reduces cost (MVs) too, but is rare; in particular if the restart period m is small. The block Arnoldi matrix relation is valid only with an error term.*

Hence:

- We deflate at startup and each restart if $\mathbf{r}_0$ is rank-deficient.
- We *may* deflate in the Arnoldi process if $\widetilde{\mathbf{y}}$ is rank-deficient.

(2) **Arnoldi deflation:** $\widetilde{\mathbf{y}}$ in block Arnoldi rank-deficient

Rather unlikely, because we start from $\mathbf{A}\mathbf{y}_{n-1}$.

Unless we deflate, search space contains extra basis vectors:

$$\mathcal{R}(\mathbf{Y}_n) \supsetneqq \mathcal{B}_n$$

But they are unlikely to help much, since the block solution lies in $\mathbf{x}_0 + \mathcal{B}_n$ for some n.

$\rightsquigarrow$ *Arnoldi deflation reduces cost (MVs) too, but is rare; in particular if the restart period m is small. The block Arnoldi matrix relation is valid only with an error term.*

Hence:

- We deflate at startup and each restart if $\mathbf{r}_0$ is rank-deficient.
- We *may* deflate in the Arnoldi process if $\widetilde{\mathbf{y}}$ is rank-deficient.

(2) **Arnoldi deflation:** $\widetilde{\mathbf{y}}$ in block Arnoldi rank-deficient

Rather unlikely, because we start from $\mathbf{A}\mathbf{y}_{n-1}$.

Unless we deflate, search space contains extra basis vectors:

$$\mathcal{R}(\mathbf{Y}_n) \supsetneq \mathcal{B}_n$$

But they are unlikely to help much, since the block solution lies in $\mathbf{x}_0 + \mathcal{B}_n$ for some n.

⇝ *Arnoldi deflation reduces cost (MVs) too, but is rare; in particular if the restart period m is small. The block Arnoldi matrix relation is valid only with an error term.*

Hence:

- We deflate at startup and each restart if $\mathbf{r}_0$ is rank-deficient.
- We *may* deflate in the Arnoldi process if $\widetilde{\mathbf{y}}$ is rank-deficient.

Block Arnoldi and block GMRES: references

Vital ['**90**$_{\text{Diss}}$] *(Bl-GMRES)*
Sadkane ['**93**$_{\text{NM}}$] *(nonsym., block Arnoldi-Chebyshev)*
Sadkane ['**93**$_{\text{NM}}$] *(nonsym., block Arnoldi / Davidson, EV)*

Chapman/Saad [*'95/'***97**$_{\text{NLAA}}$] *(Bl-GMRES, FGMRES, ...)*
Jia ['**94**$_{\text{Diss}}$] *(nonsym., EV, "general. Lanczos"* $\supset$ *Bl-Arnoldi)*
Jia [*'94/'***98**$_{\text{NM}}$] *(nonsym., EV, "general. Lanczos"* $\supset$ *Bl-Arnoldi)*
Jia ['**98**$_{\text{LAA}}$] *(nonsym., EV, Bl-Arnoldi)*
Jbilou ['**99**$_{\text{JCAM}}$] *(nonsym., residual smoothing)*
Li ['**97**$_{\text{Par. Comp.}}$] *(parallelization of* BLGMRES*)*
Saad ['**96**$_{\text{Book}}$] *(overview of* BLGMRES *versions)*
Simoncini/Gallopoulos [*'94/'***96**$_{\text{LAA}}$] *(*BLGMRES, *convergence)*

Block Arnoldi and block GMRES: references (cont'd)

Cullum/Zhang [*'98*/**'02**$_{\text{SIMAX}}$]
(two-sided BLGMRES, *deflation, control, rel. to Lanczos)*

El Guennouni/Jbilou/Riquet [*'00*/**'02**$_{\text{NumAlg}}$] *(Sylvester eq.)*
Fattebert [*'98*/**'98**$_{\text{ETNA}}$] *(Rayleigh quot. iter., gen. EV)*
Jbilou [**'99**$_{\text{JCAM}}$] *(nonsym., block smoothing)*
Jbilou/Messaoudi/Sadok [**'99**$_{\text{ApNuM}}$]
(nonsym., global FOM/GMRES)
Langou [**'03**$_{\text{Diss}}$] *(*BLGMRES*)*
Saad [**'03**$_{\text{Book}}$] *(overview of* BLGMRES *versions)*
Robbé/Sadkane [**'02**$_{\text{LAA}}$] *(error bounds for* BLGMRES*)*
Robbé/Sadkane [**'02**$_{\text{Num. Alg.}}$]
*(*BLGMRES, BLFOM *for Sylvester eq.)*
Robbé/Sadkane [*'04*] *(*BLGMRES *and* BLFOM *with deflation)*

Schmelzer [**'04**$_{\text{Dipl}}$] *(*BLMINRES *and* BLSYMMLQ *w/deflation)*

da Cunha/Becker [*'05*] *(*BLGMRES *w/deflation)*

Thanks for listening and come to ...

6th International Congress on Industrial and Applied Mathematics

www.iciam07.ch

Zurich, Switzerland
16 - 20 July 2007

iciam07

References

J. I. Aliaga, D. L. Boley, R. W. Freund and V. Hernández (2000), 'A Lanczos-type method for multiple starting vectors', *Math. Comp.* **69**(232), 1577–1601. , .

J. I. Aliaga, V. Hernandez and D. L. Boley (1994), Using the block clustered nonsymmetric Lanczos algorithm to solve control problems for MIMO linear systems, in *Proceedings of the Cornelius Lanczos International Centenary Conference* (J. D. Brown, M. T. Chu, D. C. Ellison and R. J. Plemmons, eds), SIAM, Philadelphia, PA, pp. 387–389.

Z. Bai (1994), A spectral transformation block nonsymmetric Lanczos algorithm for solving sparse nonsymmetric eigenvalue problems, in *Proceedings Fifth SIAM Conference on Applied Linear Algebra* (J. G. Lewis, ed.), SIAM, pp. 307–311.

Z. Bai and R. W. Freund (2001*a*), A partial Padé-via-Lanczos method for reduced-order modeling, in *Proceedings of the Eighth Conference of the International Linear Algebra Society (Barcelona, 1999)*, Vol. 332/334, pp. 139–164. Jan. 5, 2000.

Z. Bai and R. W. Freund (2001*b*), 'A symmetric band Lanczos