Deflation based preconditioning of linear systems of equations

Martin H. Gutknecht

Seminar for Applied Mathematics, ETH Zurich
http://www.sam.math.ethz.ch/~mhg

SC2011 — International Conference on Scientific Computing
Santa Margherita di Pula, Sardinia, Italy
October 10–14, 2011

Partly joint work with André Gaul, Jörg Liesen, Reinhard Nabben

Outline

Prerequisites

History

Augmentation and Deflation

Deflated GMRES and MINRES

Oblique projections and truly deflated GMRES

Deflated QMR

Conclusions

Iterative methods based on (Petrov-)Galerkin condition

To solve: $\mathbf{Ax} = \mathbf{b}$ with $\mathbf{A} \in \mathbb{C}^{N \times N}$ nonsingular.

Idea: compute sequence of approximate solutions $\mathbf{x}_n$ such that their residuals $\mathbf{r}_n :\equiv \mathbf{b} - \mathbf{Ax}_n$ approach $\mathbf{o}$ in some norm.

We choose $\mathbf{x}_n$ from an n-dimensional affine search space $\mathbf{x}_0 + \mathcal{S}_n$ such that some Galerkin or Petrov-Galerkin condition is satisfied:

$$\mathbf{x}_n \in \mathbf{x}_0 + \mathcal{S}_n \,, \qquad \mathbf{r}_n = \mathbf{A}(\mathbf{x}_\star - \mathbf{x}_n) \perp \widetilde{\mathcal{S}}_n \,.$$

That is,

$$\mathbf{r}_n \in \mathbf{r}_0 + \mathbf{A}\mathcal{S}_n \,, \qquad \mathbf{r}_n \perp \widetilde{\mathcal{S}}_n \,.$$

This means that $\mathbf{r}_0$ *is approximated from* $\mathbf{A}\mathcal{S}_n$ *such that "error"* $\mathbf{r}_n \perp \widetilde{\mathcal{S}}_n$ *.*

Simplified idea of deflation based preconditioning

Ideal assumption: columns of $\mathbf{U} \in \mathbb{C}^{N \times k}$ span an invariant subspace $\mathcal{U}$ of $\mathbf{A}$ belonging to eigenvalues close to 0.

Let $\mathbf{Z} :\equiv \mathbf{AU}$, $\mathcal{Z} :\equiv \mathbf{A}\mathcal{U} = \mathcal{U}$.

Note: images of the restriction $\mathbf{A}^{-1}\big|_{\mathcal{Z}}$ are trivial to compute:

if $\mathbf{z} = \mathbf{Zc} \in \mathcal{Z}$, then $\mathbf{A}^{-1}\mathbf{z} = \mathbf{Uc}$.

Main idea: split up $\mathbb{C}^N$ into $\mathcal{Z} \oplus \mathcal{Z}^\perp = \mathbb{C}^N$.

Split up $\mathbf{r}_0$ accordingly: $\mathbf{r}_0 = \underbrace{\mathbf{r}_0 - \widehat{\mathbf{r}}_0}_{\in \mathcal{Z}} + \underbrace{\widehat{\mathbf{r}}_0}_{\in \mathcal{Z}^\perp}$.

$\mathbf{A}^{-1}(\mathbf{r}_0 - \widehat{\mathbf{r}}_0)$ is trivial to invert;

$\mathbf{A}^{-1}\widehat{\mathbf{r}}_0$ will be approximated with a Krylov space solver. Essentially, the solver will act on $\mathcal{Z}^\perp$.

Since the (absolutely) small eigenvalues of $\mathbf{A}$ cause trouble in the solver, we want to replace $\mathbf{A}$ by $\widehat{\mathbf{A}}$ on $\mathcal{Z}^\perp$, such that $\widehat{\mathbf{A}}$ will no longer have these small eigenvalues (deflation).

$\widehat{\mathbf{A}}$ will have the form $\widehat{\mathbf{A}} :\equiv \mathbf{PA}$ or $\widehat{\mathbf{A}} :\equiv \mathbf{PAP}$. This looks like preconditioning, but in our case $\mathbf{P}$ will be a projection.

Hopefully, $\mathbf{A}\big|_{\mathcal{Z}^\perp} = \widehat{\mathbf{A}}\big|_{\mathcal{Z}^\perp}$.

Problems:

- Need work out details. E.g., how define/compute $\mathbf{P}$, $\widehat{\mathbf{A}}$.
- We do not want to assume that $\mathcal{Z}$ is exactly **A**–invariant.
- Orthogonal decomposition $\mathcal{Z} \oplus \mathcal{Z}^\perp$ turns out to be incompatible with CG optimality.
- If $\mathbf{A}$ is non-Hermitian, $\mathbf{A}\big|_{\mathcal{Z}^\perp} = \widehat{\mathbf{A}}\big|_{\mathcal{Z}^\perp}$ will not hold, even when $\mathbf{Z}$ is **A**-invariant.
- Need some approximate invariant subspace.

How to find an approximate invariant subspace?

- It may be known from a theoretical analysis of the problem.
- It may result from the solution of previous systems with the same $\mathbf{A}$. ($\rightsquigarrow$ linear system with multiple right-hand sides.)
- It may results from the solution of previous systems with nearby $\mathbf{A}$.
- It may results from previous cycles of the solution process (if the method is restarted).

There are lots of examples in the literature.

Prerequisites: Krylov (sub)space solvers (KSS)

Given: linear system $\mathbf{Ax} = \mathbf{b}$, initial approx. $\mathbf{x}_0 \in \mathbb{C}^N$.
Construct: approximate solutions ("iterates") $\mathbf{x}_n$ and corresponding residuals $\mathbf{r}_n :\equiv \mathbf{b} - \mathbf{A}\mathbf{x}_n$ with

$$\boxed{\mathbf{x}_n \in \mathbf{x}_0 + \mathcal{K}_n(\mathbf{A}, \mathbf{r}_0)\,, \qquad \mathbf{r}_n \in \mathbf{r}_0 + \mathbf{A}\mathcal{K}_n(\mathbf{A}, \mathbf{r}_0)\,,}$$

where $\mathbf{r}_0 :\equiv \mathbf{b} - \mathbf{A}\mathbf{x}_0$ is the initial residual, and

$$\boxed{\mathcal{K}_n :\equiv \mathcal{K}_n(\mathbf{A}, \mathbf{r}_0) :\equiv \operatorname{span}\{\mathbf{r}_0, \mathbf{A}\mathbf{r}_0, \dots, \mathbf{A}^{n-1}\mathbf{r}_0\}}$$

is the *n*th Krylov subspace generated by $\mathbf{A}$ from $\mathbf{r}_0$.

We can, e.g., construct $\mathbf{x}_n$ such that $\|\mathbf{r}_n\|$ is minimal.

$\rightsquigarrow$ *conjugate residual* (CR) *method* (Stiefel, 1955),
$\rightsquigarrow$ MINRES (Paige and Saunders, 1975),
$\rightsquigarrow$ GCR and GMRES.

Prerequisites: preconditioning

In practice, Krylov space solvers often do not work well without *preconditioning*: multiplication of $\mathbf{A}$ by some approximate inverse $\mathbf{P}$, so that $\mathbf{PA}$ or $\mathbf{AP}$ is better conditioned than $\mathbf{A}$.

Normally, $\mathbf{A}$ and $\mathbf{P} \approx \mathbf{A}^{-1}$ are nonsingular.

Here we consider an alternative to preconditioning: *(approximate) spectral deflation*.

Formally, it sometimes looks like preconditioning, but (in most cases) $\mathbf{P}$ is singular.

So, $\mathbf{PA}$ is singular too.

But we apply this *formally preconditioned matrix* or *deflated matrix* only in a suitably chosen invariant subspace.

Buzz words and their meanings

Augmented bases:	$\mathbf{x}_n \in \mathbf{x}_0 + \mathcal{K}_n(\widehat{\mathbf{A}}, \widehat{\mathbf{r}}_0) + \mathcal{U}$, where $\widehat{\mathbf{A}} = \mathbf{A}$ or $\mathrm{spec}(\widehat{\mathbf{A}}) \subset \mathrm{spec}(\mathbf{A}) \cup \{0\}$
(Spectral) deflation:	$\mathbf{A} \rightsquigarrow \widehat{\mathbf{A}} :\equiv \mathbf{PA}$ s.t. small EVals $\rightsquigarrow 0$
EVal translation:	$\mathbf{A} \rightsquigarrow \widehat{\mathbf{A}} :\equiv \mathbf{AP}$ s.t. small EVals $\rightsquigarrow |\lambda_{\max}|$
Krylov space recycling:	choice of $\mathcal{U}$ based on prev. cycles
Flexible KSS:	adaptation of $\mathbf{P}$ at each restart

While (spectral) deflation has been an indispensable tool for eigenvalue computations for at least 55 years, for solving linear systems deflation has become popular in the last 20 years only.

Two basic approaches:

- Augmentation of basis with or without spectral deflation.
- EVal translation by suitable preconditioning.

History

Early contributions (many more papers appeared since):

Nicolaides *'85*/**'87**$_{\text{SINUM}}$: deflated 3-term CG (w/augm. basis)
Dostál *'87*/**'88**$_{\text{IntJCompMath}}$: deflated 2-term CG (w/augm. basis)
Kharchenko / Yeremin *'92*/**'95**$_{\text{NLAA}}$: GMRES with transl. EVals
Morgan *'93*/**'95**$_{\text{SIMAX}}$: GMRES with augmented basis
de Sturler *'93*/**'96**$_{\text{JCAM}}$: inner-outer GMRES/GCR (and, briefly, inner/outer BiCGStab/GCR) with augmented basis
Erhel / Burrage / Pohl *'94*/**'96**$_{\text{JCAM}}$ GMRES with transl. EVals
Chapman / Saad *'95*/**'97**$_{\text{NLAA}}$ GMRES with augmented basis
Saad *'95*/**'97**$_{\text{SIMAX}}$ Analysis of KSS with augmented basis
Burrage, / Erhel / Pohl / Williams *'95*/**'98**$_{\text{SISC}}$ Deflated stationary inner-outer iterations
Baglama / Calvetti / Golub / Reichel *'96*/**'98**$_{\text{SISC}}$ Adaptively preconditioned GMRES

History (contn'd)

More recently, it was discovered by a group of authors that augmentation and deflation (= deflation based preconditioning) is algebraically very similar to

- multigrid,
- balancing Neumann-Neumann preconditioning (see Mandel **'93**$_{\text{CommApplNumMeth}}$).

See, in particular:

Erlangga / Nabben **'08**$_{\text{SIMAX}}$, **'09**$_{\text{SISC}}$
Nabben / Vuik **'08**$_{\text{NLAA}}$
Tang / Nabben / Vuik/ Erlangga **'09**$_{\text{SISC}}$

Augmentation and deflation based on orthogonal projection: the Wang/de Sturler/Paulino (2006) approach

Let $\mathbf{U} \in \mathbb{C}^{N \times k}$ contain approx. EVecs corr. to EVals close to 0. Define

$$\mathcal{U} :\equiv \mathcal{R}(\mathbf{U}), \qquad \mathbf{Z} :\equiv \mathbf{AU}, \qquad \mathcal{Z} :\equiv \mathcal{R}(\mathbf{Z}) = \mathbf{A}\mathcal{U},$$

$$\mathbf{E} :\equiv \mathbf{Z}^{\mathsf{H}}\mathbf{Z}, \qquad \mathbf{Q} :\equiv \mathbf{Z}\mathbf{E}^{-1}\mathbf{Z}^{\mathsf{H}}, \qquad \mathbf{P} :\equiv \mathbf{I} - \mathbf{Q} = \mathbf{I} - \mathbf{Z}\mathbf{E}^{-1}\mathbf{Z}^{\mathsf{H}}.$$

Note that $\mathbf{Q}^2 = \mathbf{Q}$, $\mathbf{P}^2 = \mathbf{P}$, $\mathbf{Q}^{\mathsf{H}} = \mathbf{Q}$, $\mathbf{P}^{\mathsf{H}} = \mathbf{P}$. So,

$\mathbf{Q}$ is the orthogonal projection onto $\mathcal{Z}$; dim $\mathcal{Z} = k$,

$\mathbf{P}$ is the orthogonal projection onto $\mathcal{Z}^\perp$; dim $\mathcal{Z}^\perp = N - k$.

Let $\widehat{\mathbf{r}}_0 :\equiv \mathbf{P}\mathbf{r}_0$, $\widehat{\mathbf{A}} :\equiv \mathbf{PAP}$,

$$\widehat{\mathcal{K}}_n :\equiv \mathcal{K}_n(\widehat{\mathbf{A}}, \widehat{\mathbf{r}}_0) :\equiv \operatorname{span}(\widehat{\mathbf{r}}_0, \widehat{\mathbf{A}}\widehat{\mathbf{r}}_0, \dots, \widehat{\mathbf{A}}^{n-1}\widehat{\mathbf{r}}_0).$$

We choose

$$\mathbf{x}_n \in \mathbf{x}_0 + \widehat{\mathcal{K}}_n + \mathcal{U}, \qquad \mathbf{r}_n :\equiv \mathbf{b} - \mathbf{A}\mathbf{x}_n \in \mathbf{r}_0 + \mathbf{A}\widehat{\mathcal{K}}_n + \mathcal{Z}. \qquad (1)$$

In the inclusions

$$\mathbf{x}_n \in \mathbf{x}_0 + \widehat{\mathcal{K}}_n + \mathcal{U}\,, \qquad \mathbf{r}_n \in \mathbf{r}_0 + \mathbf{A}\widehat{\mathcal{K}}_n + \mathcal{Z}$$

we have $\widehat{\mathcal{K}}_n \subset \mathcal{Z}^\perp$.

So, if $\mathcal{Z}^\perp$ is an invariant subspace, $\mathbf{A}\widehat{\mathcal{K}}_n \subset \mathcal{Z}^\perp$.
Then we could split $\mathbf{r}_0 - \mathbf{r}_n$ into two orthogonal components:

$$\mathbf{r}_0 - \mathbf{r}_n \;\in\; \mathbf{A}\widehat{\mathcal{K}}_n \oplus \mathcal{Z} \;\subset\; \mathcal{Z}^\perp \oplus \mathcal{Z}\,.$$

But, in general, $\mathbf{A}\widehat{\mathcal{K}}_n \cap \mathcal{Z} \neq \{\mathbf{o}\}$.

As mentioned, it is trivial to invert $\mathbf{A}$ on $\mathcal{Z}$.

So, if we split $\mathbf{r}_0$ into $\mathbf{r}_0 = \mathbf{P}\mathbf{r}_0 + \mathbf{Q}\mathbf{r}_0 \in \mathcal{Z}^\perp \oplus \mathcal{Z}$,
we are left with the problem of approximating $\mathbf{A}^{-1}\mathbf{P}\mathbf{r}_0$.

When computing it, we may generate an extra component in $\mathcal{Z}$, which we will avoid by replacing $\mathbf{A}$ by $\widehat{\mathbf{A}}$.

Deflated GMRES

We can compute $\mathbf{x}_n \in \mathbf{x}_0 + \widehat{\mathcal{K}}_n + \mathcal{U}$ with minimum $\|\mathbf{r}_n\|_2$ by a GMRES-like method.

Assume the cols. of $\mathbf{Z}$ are orthonormal, so that $\mathbf{Q} = \mathbf{Z}\mathbf{Z}^{\mathsf{H}}$.
Apply Arnoldi process to get ONBs for spaces $\widehat{\mathcal{K}}_n$:

$$\widehat{\mathbf{A}}\mathbf{V}_n = \mathbf{V}_{n+1}\underline{\mathbf{H}_n}\,, \quad \text{where} \quad \mathbf{v}_0 :\equiv \widehat{\mathbf{r}}_0/\beta\,.$$

Note that here $\widehat{\mathbf{A}}\mathbf{V}_n = \mathbf{PAPV}_n = \mathbf{PAV}_n$.

Using coordinate vectors $\mathbf{k}_n \in \mathbb{C}^n$ and $\mathbf{m}_n \in \mathbb{C}^k$ we write

$$\mathbf{x}_n = \mathbf{x}_0 + \mathbf{V}_n\mathbf{k}_n + \mathbf{U}\mathbf{m}_n\,, \tag{2}$$

so that

$$\mathbf{r}_n = \mathbf{r}_0 - \mathbf{A}\mathbf{V}_n\mathbf{k}_n - \mathbf{Z}\mathbf{m}_n\,. \tag{3}$$

Writing here $\quad \mathbf{r}_0 = \mathbf{P}\mathbf{r}_0 + \mathbf{Q}\mathbf{r}_0 = \widehat{\mathbf{r}}^0 + \mathbf{Q}\mathbf{r}_0 = \mathbf{v}_0\beta + \mathbf{Z}\mathbf{Z}^{\mathsf{H}}\mathbf{r}_0$
and defining $\quad \mathbf{C}_n :\equiv \mathbf{Z}^{\mathsf{H}}\mathbf{A}\mathbf{V}_n \in \mathbb{C}^{k\times n}$, we get

$$\mathbf{r}_n = \mathbf{v}_0\beta + \mathbf{Q}\mathbf{r}_0 - (\mathbf{P}+\mathbf{Q})\mathbf{A}\mathbf{V}_n\mathbf{k}_n - \mathbf{Z}\mathbf{m}_n$$

$$= \begin{bmatrix} \mathbf{Z} & \mathbf{V}_{n+1} \end{bmatrix} \left(\begin{bmatrix} \mathbf{Z}^{\mathsf{H}}\mathbf{r}_0 \\ \mathbf{e}_1\beta \end{bmatrix} - \begin{bmatrix} \mathbf{I}_k & \mathbf{C}_n \\ \mathbf{O} & \underline{\mathbf{H}}_n \end{bmatrix} \begin{bmatrix} \mathbf{m}_n \\ \mathbf{k}_n \end{bmatrix} \right). \qquad (4)$$

Since $\begin{bmatrix} \mathbf{Z} & \mathbf{V}_{n+1} \end{bmatrix}$ has orthonormal columns

$$\|\mathbf{r}_n\|_2 = \left\| \begin{bmatrix} \mathbf{Z}^{\mathsf{H}}\mathbf{r}_0 \\ \mathbf{e}_1\beta \end{bmatrix} - \begin{bmatrix} \mathbf{I}_k & \mathbf{C}_n \\ \mathbf{O} & \underline{\mathbf{H}}_n \end{bmatrix} \begin{bmatrix} \mathbf{m}_n \\ \mathbf{k}_n \end{bmatrix} \right\|_2 . \qquad (5)$$

So, minimizing $\|\mathbf{r}_n\|_2$ becomes an $(n+k+1)\times(n+k)$ least squares problem, but due to its block triangular structure this problem decouples into an $(n+1)\times n$ least squares problem for $\mathbf{k}_n$ and an explicit formula for $\mathbf{m}_n$:

$$\min \|\mathbf{r}_n\|_2 = \min_{\mathbf{k}_n\in\mathbb{C}^n} \|\mathbf{e}_1\beta - \underline{\mathbf{H}}_n\mathbf{k}_n\|_2 , \qquad \mathbf{m}_n := \mathbf{Z}^{\mathsf{H}}\mathbf{r}_0 - \mathbf{C}_n\mathbf{k}_n . \qquad (6)$$

We call the above sketched method deflated GMRES.

It differs from the “deflated GMRES” method of Morgan **’95**$_{\text{SIMAX}}$ and Chapman / Saad **’95**$_{\text{NLAA}}$, which is basically just an *augmented* GMRES method.

Our proposal is analogous to the “recycling MINRES” (RMINRES) method of Wang / de Sturler / Paulino **’06**$_{\text{IJNME}}$.

Difficulties:

1. $\widehat{\mathbf{A}}$ may have rank $< n - k$, which may cause breakdowns. There are ways to avoid such breakdowns, see Gaul et al. **’11**$_{\text{TR-TUB}}$ and Reichel / Ye **’05**$_{\text{SIMAX}}$.
2. If $\mathcal{Z}^\perp$ is $\mathbf{A}$-invariant, there are no breakdowns. But, in general: $\mathcal{Z}$ $\mathbf{A}$-invariant $\not\Longrightarrow$ $\mathcal{Z}^\perp$ $\mathbf{A}$-invariant .

Deflated MINRES

Deflated MINRES, assuming $\mathbf{A}^{\mathsf{H}} = \mathbf{A}$, is essentially a special case of deflated GMRES, but since $\widehat{\mathbf{A}}^{\mathsf{H}} = \widehat{\mathbf{A}}$:

Arnoldi	$\rightsquigarrow$	symmetric Lanczos
extended Hessenberg $\underline{\mathbf{H}}_n$	$\rightsquigarrow$	extended sym. tridiagonal $\underline{\mathbf{T}}_n$
long sum for $\mathbf{x}_n$	$\rightsquigarrow$	short recursion for $\mathbf{x}_n$
need to store $\mathbf{V}_n$	$\rightsquigarrow$	no need to store $\mathbf{V}_n$

Moreover, the following three properties hold when $\mathbf{A}^{\mathsf{H}} = \mathbf{A}$:

(i) $\mathcal{Z}$ **A**-invariant $\Longleftrightarrow$ $\mathcal{Z}^{\perp}$ **A**-invariant

(ii) $\mathcal{Z}$ **A**-invariant $\Longrightarrow$ no breakdowns, $\mathbf{C}_n = \mathbf{O}$

(iii) $\mathcal{Z}$ **A**-invariant $\Longrightarrow$ $\widehat{\mathbf{A}}\big|_{\mathcal{Z}} = \mathbf{O}\big|_{\mathcal{Z}}, \quad \widehat{\mathbf{A}}\big|_{\mathcal{Z}^{\perp}} = \mathbf{A}\big|_{\mathcal{Z}^{\perp}}$

But, in general, breakdowns are still possible, and $\mathbf{C}_n \neq \mathbf{O}$.

Note that (iii) does not hold, in general, if $\mathbf{A}^{\mathsf{H}} \neq \mathbf{A}$.

Deflation based on oblique projections: summary

1st observation: *The orthogonal decomposition* $\mathbb{C}^N = \mathcal{Z} \oplus \mathcal{Z}^\perp$ *used so far is not appropriate if* $\mathbf{A}^H \neq \mathbf{A}$.

Definition: a simple **A**–invariant subspace is an **A**–invariant subspace with the property that for any eigenvector it contains, it also contains all the other eigenvectors and generalized eigenvectors that belong to the same eigenvalue.

2nd observation: Let $\mathcal{Z}$ be a simple **A**–invariant subspace, let $\widetilde{\mathcal{Z}}$ be the complex conjugate of the corresponding left invariant subspace,
let $\mathbf{P}$ be the oblique projection onto $\widetilde{\mathcal{Z}}^\perp$ along $\mathcal{Z}$,
and let $\widehat{\mathbf{A}} :\equiv \mathbf{PAP}$. Then

$$\widehat{\mathbf{A}}\Big|_{\mathcal{Z}} = \mathbf{O}\Big|_{\mathcal{Z}}, \qquad \widehat{\mathbf{A}}\Big|_{\widetilde{\mathcal{Z}}^\perp} = \mathbf{A}\Big|_{\widetilde{\mathcal{Z}}^\perp}. \tag{7}$$

Deflation based on oblique projections: details

Let $\mathbf{U} \in \mathbb{C}^{N \times k}$ and $\widetilde{\mathbf{Z}} \in \mathbb{C}^{N \times k}$ have full rank k, and assume $\mathbf{E} \in \mathbb{C}^{k \times k}$ defined by

$$\mathbf{Z} :\equiv \mathbf{A}\mathbf{U}\,, \qquad \mathbf{E} :\equiv \widetilde{\mathbf{Z}}^{\mathsf{H}}\mathbf{Z}$$

is nonsingular. Then let

$$\mathcal{U} :\equiv \mathcal{R}(\mathbf{U})\,, \qquad \mathcal{Z} :\equiv \mathcal{R}(\mathbf{Z}) = \mathbf{A}\mathcal{U}\,, \qquad \widetilde{\mathcal{Z}} :\equiv \mathcal{R}(\widetilde{\mathbf{Z}})\,,$$

$$\mathbf{Q} :\equiv \mathbf{Z}\mathbf{E}^{-1}\widetilde{\mathbf{Z}}^{\mathsf{H}}\,, \qquad \mathbf{P} :\equiv \mathbf{I} - \mathbf{Q} = \mathbf{I} - \mathbf{Z}\mathbf{E}^{-1}\widetilde{\mathbf{Z}}^{\mathsf{H}}\,.$$

Still $\mathbf{Q}^2 = \mathbf{Q}$ and $\mathbf{P}^2 = \mathbf{P}$, but now

$$\mathbf{Q}\mathcal{Z} = \mathcal{Z}\,, \qquad \mathbf{Q}\widetilde{\mathcal{Z}}^{\perp} = \{\mathbf{o}\}\,, \qquad \mathbf{P}\mathcal{Z} = \{\mathbf{o}\}\,, \qquad \mathbf{P}\widetilde{\mathcal{Z}}^{\perp} = \widetilde{\mathcal{Z}}^{\perp}\,,$$

So, $\mathbf{Q}$ is the oblique projection onto $\mathcal{Z}$ along $\widetilde{\mathcal{Z}}^{\perp}$,
and $\mathbf{P}$ is the oblique projection onto $\widetilde{\mathcal{Z}}^{\perp}$ along $\mathcal{Z}$.

If the columns of $\widetilde{\mathbf{Z}}$ and of $\mathbf{Z}$ are chosen biorthonormal, which means that they form dual bases of $\widetilde{\mathcal{Z}}$ and $\mathcal{Z}$, then $\mathbf{E} = \widetilde{\mathbf{Z}}^{\mathsf{H}}\mathbf{Z} = \mathbf{I}_k$ and simply

$$\mathbf{Q} = \mathbf{Z}\widetilde{\mathbf{Z}}^{\mathsf{H}}, \qquad \mathbf{P} = \mathbf{I} - \mathbf{Q} = \mathbf{I} - \mathbf{Z}\widetilde{\mathbf{Z}}^{\mathsf{H}}. \tag{8}$$

Note that this holds in particular if we choose the columns of $\mathbf{Z}$ as (right-hand side) eigenvectors of $\mathbf{A}$ and those of $\widetilde{\mathbf{Z}}$ as the corresponding left eigenvectors.

As before, we further let

$$\widehat{\mathbf{r}}_0 :\equiv \mathbf{P}\mathbf{r}_0, \qquad \widehat{\mathbf{A}} :\equiv \mathbf{PAP}.$$

Then still

$$\mathcal{N}(\widehat{\mathbf{A}}) \supseteq \mathcal{N}(\mathbf{P}) = \mathcal{Z}, \qquad \mathcal{R}(\widehat{\mathbf{A}}) \subseteq \mathcal{R}(\mathbf{P}) = \widetilde{\mathcal{Z}}^{\perp}, \tag{9}$$

so that $\widehat{\mathbf{A}}\big|_{\widetilde{\mathcal{Z}}^{\perp}}$ is a possibly singular endomorphism of $\widetilde{\mathcal{Z}}^{\perp}$.

Consequently, $\widehat{\mathcal{K}}_n$ is a subset of $\widetilde{\mathcal{Z}}^{\perp}$ since $\widehat{\mathbf{r}}_0 \in \widetilde{\mathcal{Z}}^{\perp}$ too. Therefore, we are able to restrict a Krylov solver to $\widetilde{\mathcal{Z}}^{\perp}$.

With this choice of projections and subspaces holds:

THEOREM

Assume that $\mathcal{Z}$ is a simple **A***–invariant subspace and that $\widetilde{\mathcal{Z}}$ is the corresponding* $\mathbf{A}^{\mathsf{H}}$*–invariant subspace.*
Then $\widetilde{\mathcal{Z}}^{\perp}$ is also **A***–invariant, and the restrictions of* **A**, $\widehat{\mathbf{A}}$, *and* **O** *to $\mathcal{Z}$ and $\widetilde{\mathcal{Z}}^{\perp}$ satisfy*

$$\widehat{\mathbf{A}}\Big|_{\mathcal{Z}} = \mathbf{O}\Big|_{\mathcal{Z}}\,, \qquad \widehat{\mathbf{A}}\Big|_{\widetilde{\mathcal{Z}}^{\perp}} = \mathbf{A}\Big|_{\widetilde{\mathcal{Z}}^{\perp}}\,. \tag{10}$$

So this new setting is based on two non-orthogonal decompositions of $\mathbb{C}^N$:

$$\mathcal{Z} \oplus \widetilde{\mathcal{Z}}^{\perp} = \mathbb{C}^N\,, \qquad \widetilde{\mathcal{Z}} \oplus \mathcal{Z}^{\perp} = \mathbb{C}^N\,.$$

We can use it for a truly deflated GMRES and for deflated QMR.

Truly deflated GMRES

As before we start from the representations

$$\mathbf{x}_n = \mathbf{x}_0 + \mathbf{V}_n \mathbf{k}_n + \mathbf{U}\mathbf{m}_n \,, \qquad \mathbf{r}_n = \mathbf{r}_0 - \mathbf{A}\mathbf{V}_n \mathbf{k}_n - \mathbf{Z}\mathbf{m}_n \,.$$

$\mathbf{Z}$ cannot be expected to have orthogonal columns, but we can construct an orthonormal basis of $\mathcal{Z}$ by QR decomposition:

$$\mathbf{Z} = \mathbf{Z}_{\mathrm{o}} \mathbf{R}_{\mathrm{QR}} \,, \qquad \mathbf{Z}_{\mathrm{o}}^{\mathsf{H}} \mathbf{Z}_{\mathrm{o}} = \mathbf{I}_k \,.$$

A short calculation yields now

$$\mathbf{r}_n = \begin{bmatrix} \mathbf{Z}_{\mathrm{o}} & \mathbf{V}_{n+1} \end{bmatrix} \mathbf{q}_n \,, \tag{11}$$

where

$$\mathbf{q}_n :\equiv \begin{bmatrix} \mathbf{q}_n^{\circ} \\ \mathbf{q}_n^{\perp} \end{bmatrix} :\equiv \left(\begin{bmatrix} \mathbf{R}_{\mathrm{QR}} \widetilde{\mathbf{Z}}^{\mathsf{H}} \mathbf{r}_0 \\ \mathbf{e}_1 \beta \end{bmatrix} - \begin{bmatrix} \mathbf{R}_{\mathrm{QR}} & \mathbf{R}_{\mathrm{QR}} \mathbf{C}_n \\ \mathbf{O} & \underline{\mathbf{H}}_n \end{bmatrix} \begin{bmatrix} \mathbf{m}_n \\ \mathbf{k}_n \end{bmatrix} \right) \tag{12}$$

is the truly deflated GMRES quasi-residual.

The columns of each $\mathbf{Z}_o$ and $\mathbf{V}_{n+1}$ are still orthonormal, but those of $\mathbf{Z}_o$ need not be orthogonal to those of $\mathbf{V}_{n+1}$.

So, in general, $\|\mathbf{r}_n\|_2 \neq \|\mathbf{q}_n\|_2$.
But since

$$\mathbf{r}_n = \mathbf{Z}_o\mathbf{q}_n^\circ + \mathbf{V}_{n+1}\mathbf{q}_n^\perp \quad \text{with } \mathbf{Z}_o\mathbf{q}_n^\circ = \mathbf{Q}\mathbf{r}_n \in \mathcal{Z}\,, \quad \mathbf{V}_{n+1}\mathbf{q}_n^\perp = \mathbf{P}\mathbf{r}_n \in \widetilde{\mathcal{Z}}^\perp$$

we have at least

$$\|\mathbf{q}_n\|_2^2 = \|\mathbf{q}_n^\circ\|_2^2 + \|\mathbf{q}_n^\perp\|_2^2 = \|\mathbf{Q}\mathbf{r}_n\|_2^2 + \|\mathbf{P}\mathbf{r}_n\|_2^2\,.$$

We therefore minimize $\|\mathbf{q}_n\|_2$ instead of $\|\mathbf{r}_n\|_2$.
As before, this reduces to solving an $n \times (n+1)$ least-squares problem with $\underline{\mathbf{H}}_n$ for minimizing $\|\mathbf{q}_n^\perp\|_2$ and finding $\mathbf{k}_n$ and then choosing $\mathbf{m}_n$ such that $\mathbf{q}_n^\circ = \mathbf{o}$:

$$\min \|\mathbf{q}_n\|_2 = \min_{\mathbf{k}_n \in \mathbb{C}^n} \|\mathbf{e}_1\beta - \underline{\mathbf{H}}_n\mathbf{k}_n\|_2\,, \qquad \mathbf{m}_n := \widetilde{\mathbf{Z}}^H\mathbf{r}_0 - \mathbf{C}_n\mathbf{k}_n\,.$$

Deflated QMR

For deflated QMR we apply the nonsymmetric (i.e., two-sided) Lanczos process in the dual spaces $\widetilde{\mathcal{Z}}^\perp$ and $\mathcal{Z}^\perp$, expressed by the Lanczos relations

$$\mathbf{PAV}_n = \mathbf{V}_{n+1}\underline{\mathbf{T}}_n\,, \qquad \mathbf{P}^{\mathsf{H}}\mathbf{A}^{\mathsf{H}}\widetilde{\mathbf{V}}_n = \widetilde{\mathbf{V}}_{n+1}\underline{\widetilde{\mathbf{T}}}_n\,,$$

This leads, as in truly deflated GMRES, to the representation

$$\mathbf{r}_n = \begin{bmatrix} \mathbf{Z}_0 & \mathbf{V}_{n+1} \end{bmatrix} \mathbf{q}_n\,,$$

where

$$\mathbf{q}_n :\equiv \begin{bmatrix} \mathbf{q}_n^\circ \\ \mathbf{q}_n^\perp \end{bmatrix} :\equiv \left(\begin{bmatrix} \mathbf{R}_{\mathrm{QR}}\widetilde{\mathbf{Z}}^{\mathsf{H}}\mathbf{r}_0 \\ \mathbf{e}_1\beta \end{bmatrix} - \begin{bmatrix} \mathbf{R}_{\mathrm{QR}} & \mathbf{R}_{\mathrm{QR}}\mathbf{C}_n \\ \mathbf{O} & \underline{\mathbf{T}}_n \end{bmatrix} \begin{bmatrix} \mathbf{m}_n \\ \mathbf{k}_n \end{bmatrix} \right)$$

is now the deflated QMR quasi-residual.

Formally, all looks the same except that $\underline{\mathbf{H}}_n$ became $\underline{\mathbf{T}}_n$.

But now $\mathbf{V}_{n+1}$ has no longer orthogonal columns. So, in general,

$$\|\mathbf{q}_n^\circ\|_2 = \|\mathbf{Q}\mathbf{r}_n\|_2 \,, \qquad \|\mathbf{q}_n^\perp\|_2 \neq \|\mathbf{P}\mathbf{r}_n\|_2 \,.$$

We end up with an $n \times (n+1)$ least-squares problem with $\underline{\mathbf{T}}_n$ for minimizing $\|\mathbf{q}_n^\perp\|_2$ and finding $\mathbf{k}_n$ and subsequently choosing $\mathbf{m}_n$ such that $\mathbf{q}_n^\circ = \mathbf{o}$:

$$\min \|\mathbf{q}_n^\circ\|_2 = \min_{\mathbf{k}_n \in \mathbb{C}^n} \|\mathbf{e}_1\beta - \underline{\mathbf{T}}_n\mathbf{k}_n\|_2 \,, \qquad \mathbf{m}_n :\!= \widetilde{\mathbf{Z}}^{\mathsf{H}}\mathbf{r}_0 - \mathbf{C}_n\mathbf{k}_n \,.$$

This is all based on the dual oblique decompositions

$$\mathcal{R}\left(\begin{bmatrix} \mathbf{Z} & \mathbf{V}_n \end{bmatrix}\right) = \mathcal{Z} \oplus \widehat{\mathcal{K}}_{n+1} \subseteq \mathcal{Z} \oplus \widetilde{\mathcal{Z}}^\perp = \mathbb{C}^N \,,$$

$$\mathcal{R}\left(\begin{bmatrix} \widetilde{\mathbf{Z}} & \widetilde{\mathbf{V}}_n \end{bmatrix}\right) = \widetilde{\mathcal{Z}} \oplus \widehat{\mathcal{L}}_{n+1} \subseteq \widetilde{\mathcal{Z}} \oplus \mathcal{Z}^\perp = \mathbb{C}^N \,,$$

where

$$\widehat{\mathcal{L}}_n :\equiv \mathcal{K}_n(\widehat{\mathbf{A}}^{\mathsf{H}}, \widetilde{\mathbf{v}}_0) :\equiv \operatorname{span}(\widetilde{\mathbf{v}}_0, \widehat{\mathbf{A}}^{\mathsf{H}}\widetilde{\mathbf{v}}_0, \ldots, (\widehat{\mathbf{A}}^{\mathsf{H}})^{n-1}\widetilde{\mathbf{v}}_0) \subseteq \mathcal{Z}^\perp \,.$$

Conclusions

- Krylov solvers incorporating an augmentation of the bases and a corresp. deflation of **A** have been very successful.
- However, from a theoretical point of view, in most papers addressing nonsymmetric matrices **A** the projections and subspaces have not been chosen the right way.
- We promote oblique decomposition according to

$$\mathcal{R}\left(\begin{bmatrix} \mathbf{Z} & \mathbf{V}_n \end{bmatrix}\right) = \mathcal{Z} \oplus \widehat{\mathcal{K}}_{n+1} \subseteq \mathcal{Z} \oplus \widetilde{\mathcal{Z}}^{\perp} = \mathbb{C}^N ,$$
$$\mathcal{R}\left(\begin{bmatrix} \widetilde{\mathbf{Z}} & \widetilde{\mathbf{V}}_n \end{bmatrix}\right) = \widetilde{\mathcal{Z}} \oplus \widehat{\mathcal{L}}_{n+1} \subseteq \widetilde{\mathcal{Z}} \oplus \mathcal{Z}^{\perp} = \mathbb{C}^N ,$$

so that

$$\text{(i)} \quad \mathcal{Z} \ \ \mathbf{A}\text{-invariant} \quad \Longleftrightarrow \quad \widetilde{\mathcal{Z}}^{\perp} \ \ \mathbf{A}\text{-invariant},$$
$$\text{(ii)} \quad \mathcal{Z} \ \ \mathbf{A}\text{-invariant} \quad \Longrightarrow \quad \widehat{\mathbf{A}}\big|_{\mathcal{Z}} = \mathbf{O}\big|_{\mathcal{Z}}, \quad \widehat{\mathbf{A}}\big|_{\widetilde{\mathcal{Z}}^{\perp}} = \mathbf{A}\big|_{\widetilde{\mathcal{Z}}^{\perp}} .$$

References

1. M.H.G., *Spectral Deflation in Krylov Solvers: a Theory of Coordinate Space Based Methods*. Submitted (2011).
2. A. Gaul, M.H.G., J. Liesen, R. Nabben, *Deflated and Augmented Krylov Subspace Methods: Basic Facts and a Breakdown-free Deflated MINRES*. Submitted (2011).
3. A. Gaul, M.H.G., J. Liesen, R. Nabben, *Deflated and Augmented Krylov Subspace Methods: A Framework for Deflated BiCG and Related Methods*. In preparation.