

Flexible and Multishift Induced Dimension Reduction Algorithms

Martin van Gijzen

Joint work with Gerard Sleijpen and Jens Zemke

NWO-JSPS Seminar, Delft, April 10-13, 2012

Outline

- Introduction
- IDR Hessenberg relation
- Variable preconditioners
 - Flexible QMRIDR(s)
 - Numerical experiments
- Shifted problems
 - Multishift QMRIDR(s)
 - Numerical experiments
- Conclusions

Introduction (1)

Almost all Krylov methods fall in one of the following two classes:

1. Residuals are explicitly generated that satisfy some desirable property. Examples are: CG, CR, BiCG, BiCGSTAB, GCR, IDR(s) etc. The approximate solution is obtained as a side product.
2. A basis for the Krylov subspace is explicitly generated using a Hessenberg relation. The approximate solution is computed using the (partial) Hessenberg decomposition. Examples are: GMRES, FOM, QMR, etc.

Introduction (2)

The distinction is algorithmic rather than mathematical, mathematically equivalent methods exist from both classes (e.g. GMRES and GCR).

For the second class the residual is normally not available, which gives them a less natural, less elegant flavour (personal opinion).

The main advantage, however, is the explicit availability of the partial Hessenberg decomposition, which makes extension for solving other types of problems often quite straightforward.

For example, it is trivial to calculate approximate eigenvalues with GMRES, or to use GMRES to solve a sequence of shifted systems. The same is true for QMR.

Introduction (3)

In this talk we present an IDR-based method of the second type, i.e. IDR is used to make a partial Hessenberg decomposition.

The aim is to derive a flexible IDR method (cf. flexible GMRES), and a multishift IDR method (cf. multishift GMRES).

To compute the 'IDR basis vectors' we use orthogonalisation when possible, and to compute approximate solutions we use quasi-minimization, which results in GMRES-like algorithms *that still use short recurrences*.

IDR Hessenberg relation (1)

IDR Hessenberg relation (1)

- $\text{IDR}(s)$ constructs vectors that are in a sequence of nested subspace \mathcal{G}_j of shrinking dimension.

IDR Hessenberg relation (1)

- IDR(s) constructs vectors that are in a sequence of nested subspace \mathcal{G}_j of shrinking dimension.
- The main steps are as follows:
 - Compute vector $\mathbf{v} \in \mathcal{G}_{j-1} \perp \mathbf{p}_1, \dots, \mathbf{p}_s$;
 - Compute vector $\mathbf{g} \in \mathcal{G}_j : \mathbf{g} = (\mathbf{I} - \omega_j \mathbf{A})\mathbf{v}$.

IDR Hessenberg relation (1)

- IDR(s) constructs vectors that are in a sequence of nested subspace \mathcal{G}_j of shrinking dimension.
- The main steps are as follows:
 - Compute vector $\mathbf{v} \in \mathcal{G}_{j-1} \perp \mathbf{p}_1, \dots, \mathbf{p}_s$;
 - Compute vector $\mathbf{g} \in \mathcal{G}_j : \mathbf{g} = (\mathbf{I} - \omega_j \mathbf{A})\mathbf{v}$.
- To compute a vector $\in \mathcal{G}_j$, $s + 1$ vectors $\in \mathcal{G}_{j-1}$ are needed.

IDR Hessenberg relation (1)

- IDR(s) constructs vectors that are in a sequence of nested subspace \mathcal{G}_j of shrinking dimension.
- The main steps are as follows:
 - Compute vector $v \in \mathcal{G}_{j-1} \perp p_1, \dots, p_s$;
 - Compute vector $g \in \mathcal{G}_j : g = (I - \omega_j A)v$.
- To compute a vector $\in \mathcal{G}_j$, $s + 1$ vectors $\in \mathcal{G}_{j-1}$ are needed.
- Intermediate vectors $\in \mathcal{G}_j$ are not unique. For stability they can be orthonormalised wrt the vectors $\in \mathcal{G}_j$.

IDR Hessenberg relation (2)

The vectors g satisfy a generalised Hessenberg relation.

$$AV_n = G_{n+1}H_n \quad V_n = G_n C_n$$

with

- $G_n = [g_1, \dots, g_n]$;
- $V_n = [v_1, \dots, v_n]$;
- C_n : $n \times n$ upper triangular;
- H_n : $(n + 1) \times n$ extended Hessenberg, with upper bandwidth s .

Solution by quasi-minimisation

For solving a linear system $Ax = b$ we take $g_1 = \frac{1}{\|b\|}b$ and search for a solution

$$x_n = V_n y_n$$

by minimising the residual norm:

$$\min_{y_n} \|b - AV_n y_n\| \Leftrightarrow \min_{y_n} \|G_{n+1}(\|b\|e_1 - H_n y_n)\|$$

We 'quasi-minimise' this expression by solving

$$\min_{y_n} \|\|b\|e_1 - H_n y_n\|$$

Some remarks

- The derivation of the algorithm is analogous to GMRES and QMR.
- As in QMR, the algorithm can be implemented using short recurrences (of length $s + 2$) only.
- Since the initial s vectors g_i form an orthonormal set, our QMRIDR(s) variant is mathematically equivalent to GMRES in the first s iterations.
- An earlier QMRIDR method was proposed by Du, Sogabe, and Zhang (JCAM, 2011), aiming to obtain smoother convergence. Their algorithm is based on the 'prototype' IDR algorithm, and is not mathematically equivalent to GMRES in the first s steps.

Flexible QMRIDR(s) (1)

If right-preconditioning is used, the Hessenberg relation becomes

$$AP^{-1}V_n = G_{n+1}H_n$$

or

$$AZ_n = G_{n+1}H_n, \quad Z_n = P^{-1}V_n .$$

If a variable preconditioner P_n^{-1} is used, this relation still holds.

Z_n is then defined by

$$Z_n = [P_1^{-1}v_1, \dots, P_n^{-1}v_n]$$

Flexible QMRIDR(s) (2)

To find an approximate solution \mathbf{x}_n put

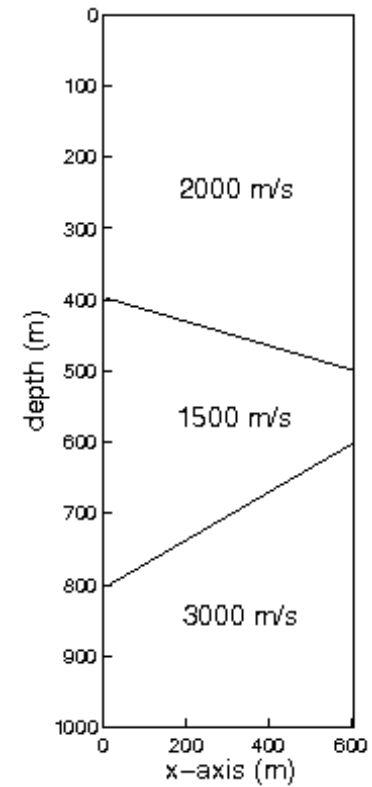
$$\mathbf{x}_n = \mathbf{Z}_n \mathbf{y}_n$$

and follow the 'quasi-minimisation' procedure explained before.
In the first s steps Flexible QMRIDR(s) is mathematically equivalent with FGMRES.

Geophysical example

The test problem models propagation of a sound wave with frequency f in the earth crust. It mimics three layers with a simple heterogeneity.

$$\begin{cases} -\Delta p - \left(\frac{2\pi f}{c(\mathbf{x})}\right)^2 p = s, & \text{in } \Omega = (0, 600) \times (0, 1000) \text{ m}^2 \\ s = \delta(x_1 - 300, x_2), & x_1 = (0, 600), x_2 = (0, 1000) \\ \text{with Neumann conditions } \frac{\partial p}{\partial n} = -\frac{2\pi i f}{c} p & \text{on } \Gamma \equiv \partial\Omega. \end{cases}$$



Discretisation and preconditioning

Discretisation with Finite Element Method gives system

$$(\mathbf{K} - z_1 \mathbf{M})\mathbf{u} = \mathbf{f} .$$

System matrix is indefinite \rightarrow difficult for iterative methods.

We use the Shifted Laplace preconditioner

$$\mathbf{P} = \mathbf{K} - z_2 \mathbf{M}$$

with $z_2 = -i|z_1|$ (Erlangga, Vuik and Oosterlee, 2004, 2006)

Shifted Laplace preconditioner (2)

In practice, a cheap approximation of P^{-1} is used.

The idea behind the shifted Laplace preconditioner is that this matrix is much easier to approximate than the discrete Helmholtz operator.

Erlangga et al. used one geometric multigrid cycle to approximate the inverse of P .

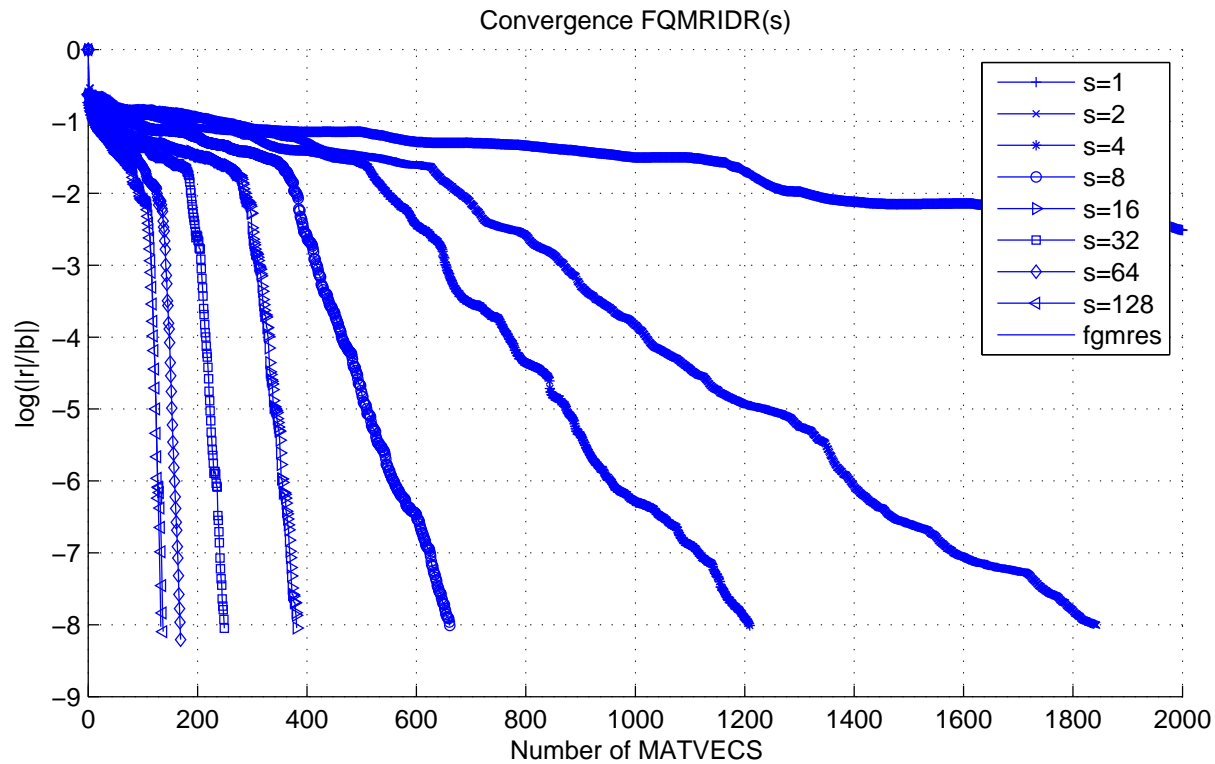
The MG-method we use is AGMG, by Notay (ETNA, 2010). AGMG uses a so-called K-cycle multigrid scheme, which uses a Krylov solver at each level. As a consequence, the outer iteration has to be a *flexible* Krylov method.

Example

In the experiment we use the following parameters:

- $z = (2\pi f)^2$, $f = 8Hz$.
- $h = 12.5m$.
- 3700 equations.

Convergence for increasing s



() Wedge problem

Observations

- Convergence curve for $s = 128$ coincides with FGMRES.
- Convergence improves for larger s (for this example, not true in general if preconditioner is very variable).
- General remark: for more variable preconditioner it is often better to use a small value for s .

The multiple-frequency problem

In practice, often a sequence of systems at different frequencies has to be solved:

$$(\mathbf{A} - z_i \mathbf{I}) \mathbf{x}^{z_i} = \mathbf{b}, \quad i = 1, n_z$$

A number of Krylov methods for this problems exist, e.g.

- Multishift BiCG, Bi-CGSTAB, BiCGstab(ℓ) (Jegerlehner, 1996, Frommer 2002)
- Multishift QMR (Freund, 1994)
- Restarted GMRES (Frommer and Glassner, 1998), FOM (Simoncini, 2003)
- CGLS (Van den Eshof, Sleijpen, 2004)

Multishift QMRIDR(s)

To solve shifted system $(A - z^i I)x^{z_i} = b$ we use again

$$AV_n = G_{n+1}H_n .$$

We search for a solution

$$x_n^{z_i} = V_n y_n^{z_i} .$$

The norm of the residual of the shifted system is 'quasi-minimised':

$$\min_{y_n^{z_i}} \|b - (A - z_i I)V_n y_n^{z_i}\| \Leftrightarrow \min_{y_n^{z_i}} \|\|b\|e_1 - (H_n - z_i \underline{C}_n)y_n^{z_i}\|$$

Here \underline{C}_n is C_n with a zero row appended.

Remarks

The only difference between our QMRIDR(s) and multishift QMRIDR(s) is that the quasi-minimisation is performed on (small) shifted Hessenberg systems. The solutions to the large shifted systems are computed using short recurrences.

- Computation of Hessenberg decomposition is the same as for QMRIDR(s) \rightarrow requires one matvec per iteration plus $3s$ vector operations (inner product or update), and storage for $s + 1$ g vectors.
- Computation of solution vectors for the shifted systems requires *for each shift* $s + 1$ update vectors. Computational cost consists of $n_z(s + 2)$ vector updates plus scalar operations.

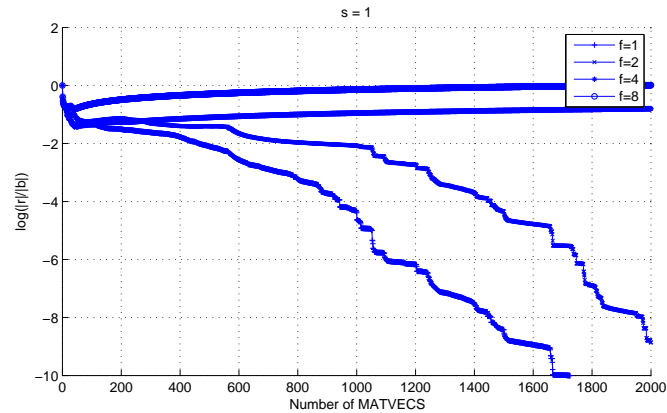
Geophysical example (continued)

We again consider the geophysical test problem, and write it as a sequence of shifted problems.

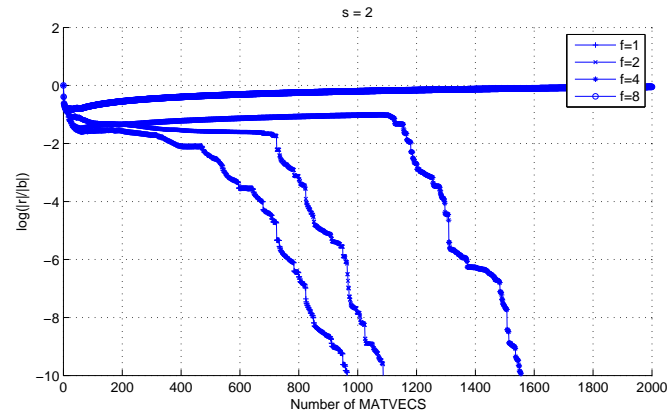
$$(M^{-1}K - z_i I)x^{z_i} = M^{-1}b$$

We consider four shifts $z_i = (2\pi f)^2$, with $f = 1, 2, 4, 8$.

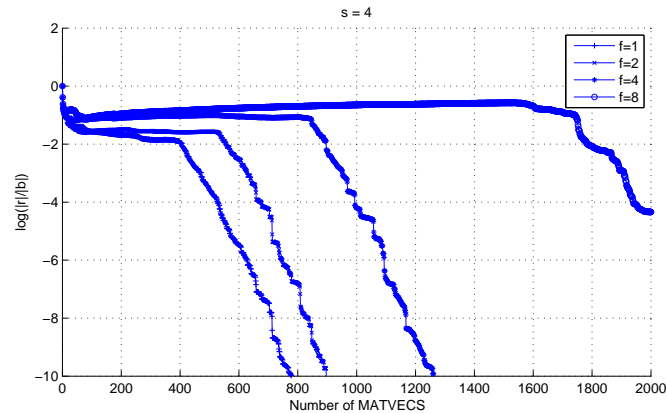
Convergence multi-frequency problem



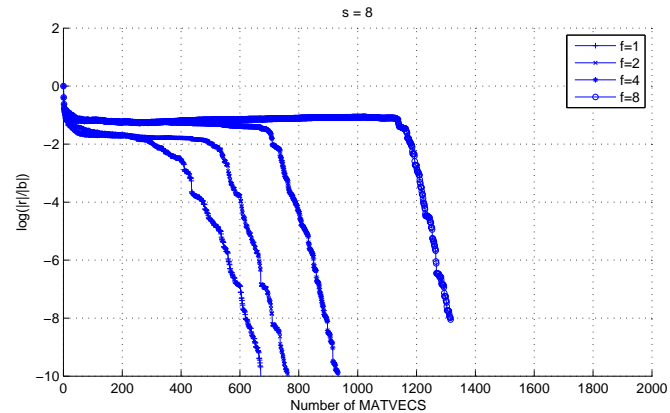
() $s=1$



() $s=2$



() $s=4$



() $s=8$

Observations

- Multishift-QMRIDR works 'okay' for this problem, better for higher s
- Problem too ill-conditioned, needs a preconditioner
- How to apply a preconditioner with a multishift method is an open problem

A convection-diffusion-reaction problem

$$-\epsilon \Delta u + \vec{\beta} \cdot \nabla u - ru = F$$

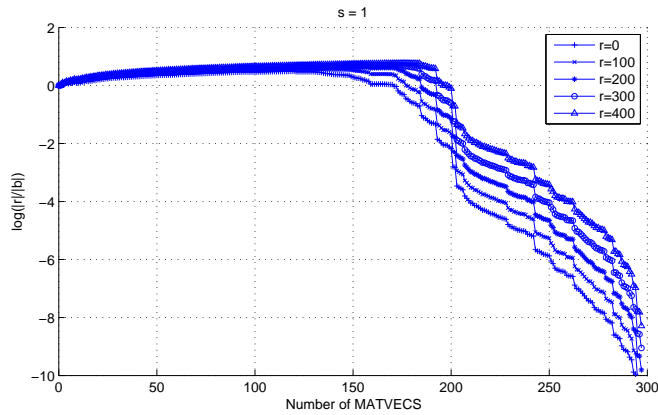
with homogeneous Dirichlet conditions on the unit cube.

F is defined by $u(x, y, z) = x(1 - x)y(1 - y)z(1 - z)$.

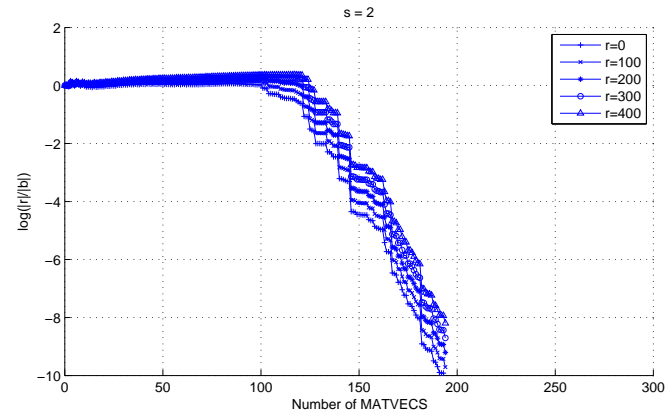
- Central differences discretisation
- $h = 0.025 \rightarrow 60,000$ equations
- $\epsilon = 1$
- $\vec{\beta} = (0/\sqrt{5} \ 250/\sqrt{5} \ 500/\sqrt{5})^T$
- $r = 0, 100, 200, 300, 400$

Matrix is nonsymmetric and indefinite

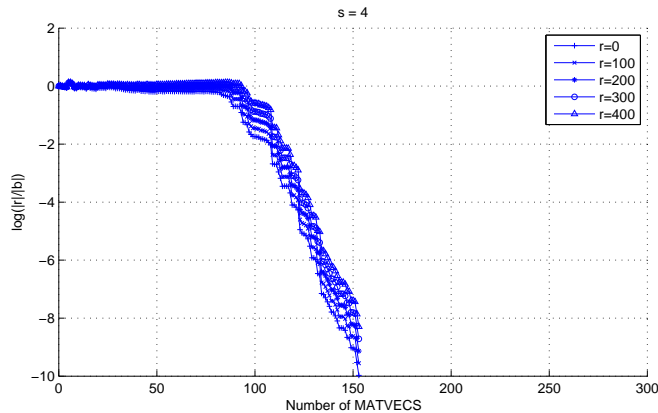
Convergence multishift QMRIDR



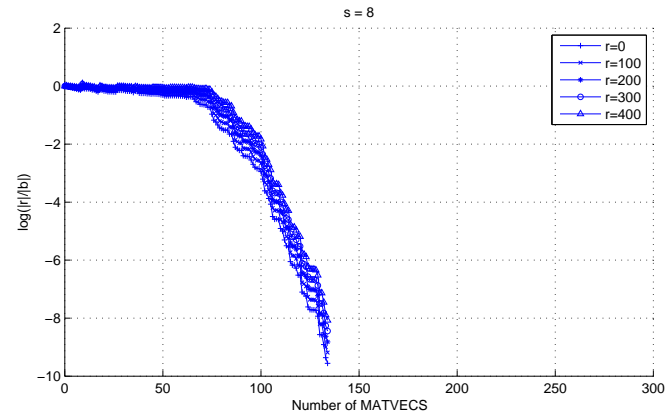
(a) $s=1$.



(b) $s=2$.



(c) $s=4$.



(d) $s=8$.

Concluding remarks

- We have described a flexible and a multishift variant of IDR(s).
- Flexible QMRIDR close to FGMRES, equivalent for s large
- Multishift QMRIDR(s) works quite well for well-conditioned problems

More information (report):

<http://ta.twi.tudelft.nl/nw/users/gijzen/IDR.html>