Insights into block rational Krylov methods

Stefan Güttel
(with Steven Elsworth and Kathryn Lund)

The University of Manchester

Block Krylov methods: not a new idea

PAPER NO. THP1.2

A BLOCK LANCZOS ALGORITHM FOR COMPUTING THE q ALGEBRAICALLY LARGEST EIGENVALUES AND A CORRESPONDING EIGENSPACE OF LARGE, SPARSE, REAL SYMMETRIC MATRICES

Jane Cullum
Mathematical Sciences
IBM T. J. Watson Research Center
P. O. Box 218
Yorktown Heights, New York 10598

W. E. Donath
Applied Research
IBM T. J. Watson Research Center
P. O. Box 218
Yorktown Heights, New York 10598

Abstract

Many engineering applications require the computation of the q algebraically largest eigenvalues and a corresponding eigenspace of a large, sparse, real, symmetric matrix. An iterative, block version of the symmetric Lanczos algorithm has been developed for this computation. There are no restrictions on the sparsity pattern within the matrix or on the distribution of the eigenvalues of the matrix. Zero eigenvalues, eigenvalues equal in magnitude but opposite in sign, and multiple eigenvalues can all be handled directly by the procedure.

The iterative block Lanczos procedure described in detail in [6] and summarized in this paper does not require this continuous reorthogonalization. Since it is a Lanczos procedure it can handle nondominant eigenvalues, including zero eigenvalues, and the presence of blocks permits direct treatment of multiple eigenvalues. Nominally, it requires more storage than the simultaneous iteration procedures; however, piecewise methods of solution [7] can sometimes be used to decrease the amount of storage needed. A well-defined convergence test allows the user, using results from [8], to estimate the accuracy of the computed eigenelements. Since it is a descendant

In: *Proceedings of the IEEE Conference on Decision and Control*, pp. 505–509, 1974.

Motivation: why do we need block Krylov methods?

Let $A \in \mathbb{C}^{N\times N}$ be Hermitian and $\mathbf{b} \in \mathbb{C}^N$ nonzero. Let p be the smallest-degree monic polynomial such that

$$p(A)\mathbf{b} = \mathbf{0}.$$

Then in exact arithmetic the non-block Lanczos algorithm will locate the (simple) roots λ of p and find one eigenvector $\mathbf{v}$ such that $A\mathbf{v} = \lambda\mathbf{v}$.

$\Longrightarrow$ multiple eigenvalues or tight clusters will not be found (quickly).

Convergence of non-block Ritz values is slow and erratic

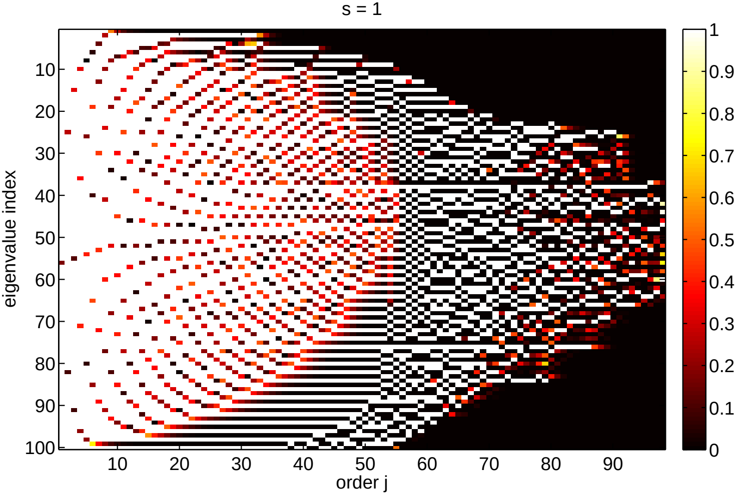
Example: `A = wilkinson(100)` and `b = randn(100,1)`.

Plot the Ritz values of orders $j = 1, 2, \ldots, 100$.

Colour indicates the distance to a closest eigenvalue of A.

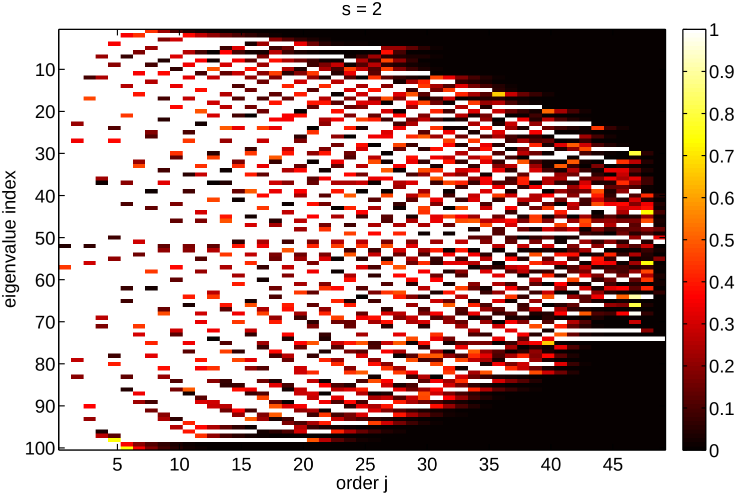
Convergence of non-block Ritz values is slow and erratic

Example: `A = wilkinson(100)` and `b = randn(100,1)`.
Plot the Ritz values of orders $j = 1, 2, \ldots, 100$.
Colour indicates the distance to a closest eigenvalue of A.



Convergence of block Ritz values is more reliable

Example: `A = wilkinson(100)` and `b = randn(100,2)`.
Plot the block Ritz values of orders $j = 1, 2, \ldots, 50$.
Colour indicates the distance to a closest eigenvalue of A.

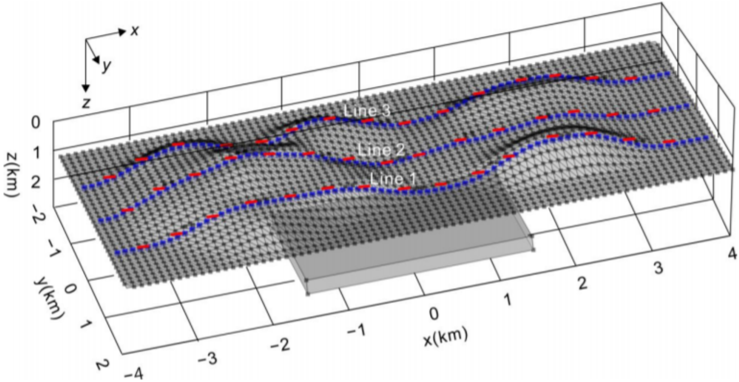


Block methods can speed up computations

Multi-source forward simulation of a synthetic 3D hydrocarbon reservoir requires repeated integrations of large-sparse linear ODEs

$$M\mathbf{e}'(t) = K\mathbf{e}(t), \quad \mathbf{e}(0) = \mathbf{e}_i \in \mathbb{R}^N$$

with multiple initial conditions $\mathbf{e}_1, \ldots, \mathbf{e}_s$. Here, $N = 810\,100$ and $s = 45$.



Block methods can speed up computations ctd.

Define $\mathbf{b} := [\mathbf{e}_1, \ldots, \mathbf{e}_s]$ and compute M-orthonormal basis $\mathbf{V}_m \in \mathbb{C}^{N \times ms}$, $\mathbf{V}_m^T M \mathbf{V}_m = \mathbf{I}_{ms}$, of a block rational Krylov space

$$\mathcal{V}_m := \left\{ \sum_{j=1}^{m} (K - \xi_j M)^{-1} \mathbf{b} C_j \ : \ C_j \in \mathbb{C}^{s \times s} \right\} \subset \mathbb{R}^{N \times s}$$

using (distinct) optimized shifts $\xi_1, \ldots, \xi_m > 0$.

Rational Arnoldi approximation

$$\widehat{\mathbf{e}}(t) := \mathbf{V}_m \exp(t\, \mathbf{V}_m^T K \mathbf{V}_m)(\mathbf{V}_m^T \mathbf{e}_i)$$

approximately solves $M\mathbf{e}'(t) = K\mathbf{e}(t)$, $\mathbf{e}(0) = \mathbf{e}_i$.

We measured 1.5x speedup with block vs non-block

Method		Elapsed time	Relative efficiency
FETD Scheme		922 s	1.00
Single-vector rational Krylov	Solving the initial vectors	4.5 s	7.65
	Construction of rational Krylov basis	114.6 s	
	Evaluation of electric solutions	1.5 s	
Block rational Krylov	Solving the initial vectors	4.5 s	11.33
	Construction of rational Krylov basis	75.2 s	
	Evaluation of electric solutions	1.7 s	

[Qiu/Güttel/Ren/Yin/Liu/Zhang/Egbert, *Geophys. J. Int.*, 2019]

Speedup depends on application and architecture

Some block speedup factors reported in literature:

- Block BFGMRES-DR solver for adjoint problems in aerodynamic design optimization in [Pinel/Montagnac, AIAA J., 2013]: 1.1–1.4
- Block QMR/GMRES solvers for EM modelling in frequency domain in [Puzyrev/Cela, *Geophys. J. Int.*, 2015]: 1.5–2.0
- Block FOM solver for variational data assimilation in [Mercier et al., *Q. J. R. Meteorol. Soc.*, 2018]: about 5 (per iteration)
- Block CG solver on GPUs for lattice QCD in [Clark et al., *Comp. Phys. Comm.*, 2018]: about 2–5 (using mixed precision)

$\Longrightarrow$ **Block speedups of up to 5x appear to be realistic.**

But it's not always about speed!

There might be "structural" requirements to use block methods:

- Eigenvalue problems with clustered eigenvalues:
  Mehrmann/Schröder 2011, Freitag/Kürschner/Pestana 2018, Drineas/Ipsen/Kontopoulou/Magdon-Ismails 2018
- Matrix equations with low-rank right-hand sides:
  El Guennouni/Jbilou/Riquet 2002, Heyouni/Jbilou 2009
- Model order reduction of MIMO systems:
  Freund 2000, Abidi/Hached/Jbilou 2014

But it's not always about speed!

There might be "structural" requirements to use block methods:

- Eigenvalue problems with clustered eigenvalues:
  Mehrmann/Schröder 2011, Freitag/Kürschner/Pestana 2018, Drineas/Ipsen/Kontopoulou/Magdon-Ismails 2018
- Matrix equations with low-rank right-hand sides:
  El Guennouni/Jbilou/Riquet 2002, Heyouni/Jbilou 2009
- Model order reduction of MIMO systems:
  Freund 2000, Abidi/Hached/Jbilou 2014

Block Krylov methods have proven to be useful in practice.

Our recent focus (jointly with Steven Elsworth):

- analyze premature breakdowns of block rational Arnoldi algorithm;
- provide robust implementations of block rational Krylov methods.

The block rational Arnoldi algorithm

Input: $A \in \mathbb{C}^{N \times N}$, $\mathbf{b} \in \mathbb{C}^{N \times s}$, shifts $\xi_1, \ldots, \xi_m \in \mathbb{C} \setminus \Lambda(A)$.

1. $\mathbf{v}_1 := \mathbf{b} R^{-1}$ such that $\mathbf{v}_1^* \mathbf{v}_1 = I_s$
2. **for** $j = 1, \ldots, m$ **do**
3. Choose a *continuation block vector* $\mathbf{t}_j \in \mathbb{C}^{js \times s}$
4. $\mathbf{w} := (A - \xi_j I)^{-1} \mathbf{V}_j \mathbf{t}_j$
5. **for** $i = 1, \ldots, j$ **do**
6. $C_{i,j} := \mathbf{v}_i^* \mathbf{w}$
7. $\mathbf{w} := \mathbf{w} - \mathbf{v}_i C_{i,j}$
8. **end for**
9. $\mathbf{v}_{j+1} := \mathbf{w} C_{j+1,j}^{-1}$ such that $\mathbf{v}_{j+1}^* \mathbf{v}_{j+1} = I_s$
10. $\mathbf{V}_{j+1} := [\mathbf{V}_j, \mathbf{v}_{j+1}]$
11. **end for**

Continuation block vector $\mathbf{t}_j$ affects the numerical rank of $[\mathbf{V}_j, \mathbf{w}]$!

Block rational Arnoldi decompositions

The rational Arnoldi algorithm produces decompositions of the form

$$A\mathbf{V}_{m+1}\underline{\mathbf{K}_m} = \mathbf{V}_{m+1}\underline{\mathbf{H}_m}$$

with block upper-Hessenberg matrices $\underline{\mathbf{K}_m}$ and $\underline{\mathbf{H}_m}$:

$$A \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \\ & K_{32} \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \\ & H_{32} \end{bmatrix}$$

An integral formula for the block basis vectors

Associated with each basis block vector $\mathbf{v}_{j+1}$ is a rational function $R_j : \mathbb{C} \to \mathbb{C}^{s \times s}$ such that

$$\mathbf{v}_{j+1} = R_j(A) \circ \mathbf{v}_1 := \frac{1}{2\pi i} \int_\Gamma (zI - A)^{-1} \mathbf{v}_1 R_j(z)^{-1} dz,$$

with a contour Γ including $\Lambda(A)$ and excluding the poles of R_j.

[Elsworth & G., MIMS Eprint 2019.2, 2019]

A connection with nonlinear eigenvalue problems

Theorem (Elsworth & G., 2019)

The points $\lambda \in \mathbb{C}$ where $R_j(\lambda)$ is singular are contained in $\Lambda(\mathbf{H}_j, \mathbf{K}_j)$.

In order to avoid premature breakdown of the rational Arnoldi algorithm in the step $\mathbf{w} := (A - \xi_j I)^{-1}\mathbf{V}_j\mathbf{t}_j$, the continuation block vector $\mathbf{t}_j$ should be chosen such that the rational block function $\widetilde{R}(z)$ satisfying

$$\mathbf{V}_j\mathbf{t}_j = \widetilde{R}(A) \circ \mathbf{v}_1$$

is nonsingular at ξ_j.

Can show: A block left null vector $\mathbf{t}_j^*$ such that $\mathbf{t}_j^*(\underline{\mathbf{H}_{j-1}} - \xi_j \underline{\mathbf{K}_{j-1}}) = \mathbf{0}^*$ satisfies conditions of the theorem, generalizing a result by Ruhe (1998).

$\Longrightarrow$ **our new recommendation for the continuation block vector $\mathbf{t}_j$.**

INLET test from Oberwolfach Collection

Problem: Compute rational Krylov basis for nonsymmetric $A, E \in \mathbb{R}^{N \times N}$ and a block starting vector $\mathbf{b} \in \mathbb{R}^{N \times 2}$, where $N = 11\,730$.

```
xi1 = 1i*repmat(linspace(0, 40, 4), 1, 6);
xi2 = xi1;
xi2(13) = 0.996 - 0.0762i;
```

The 13th pole is changed to $0.996000 - 0.0762000i$, which is close to an eigenvalue $0.996026 - 0.0762341i$ of the matrix pencil $(\mathbf{H}_{12}, \mathbf{K}_{12})$.

Results:

	`xi1`		`xi2`	
	`'last'`	`'ruhe'`	`'last'`	`'ruhe'`
`cond`	8.7×10^{5}	4.2×10^{3}	1.4×10^{9}	4.8×10^{4}
`space`	5.0×10^{-6}	1.5×10^{-11}	7.3×10^{-3}	1.9×10^{-10}

Summary

- Block (rational) Krylov applications for eigenvalue computations, block linear systems, matrix equations, in model order reduction, etc.
- Block speedups of up to 5x appear to be realistic in practice.
- Use of block methods might be imposed by problem structure.
- Connection rational Arnoldi method $\Longleftrightarrow$ nonlinear eigenproblems.
- New continuation strategy as default in Rational Krylov Toolbox, available at

  `http://rktoolbox.org`

S. Elsworth and S. Güttel. *The block rational Arnoldi method.* Submitted 2019. `http://eprints.maths.manchester.ac.uk/2685/`