On the numerical solution of large-scale linear matrix equations

V. Simoncini

Dipartimento di Matematica, Università di Bologna (Italy)

valeria.simoncini@unibo.it

Some matrix equations

- Sylvester matrix equation

$$A\mathbf{X} + \mathbf{X}B + D = 0$$

Eigenvalue pbs and tracking, Control, MOR, Assignment pbs, Riccati eqn

Some matrix equations

- Sylvester matrix equation

$$A\mathbf{X} + \mathbf{X}B + D = 0$$

Eigenvalue pbs and tracking, Control, MOR, Assignment pbs, Riccati eqn

- Lyapunov matrix equation

$$A\mathbf{X} + \mathbf{X}A^\top + D = 0, \qquad D = D^\top$$

Stability analysis in Control and Dynamical systems, Signal processing, eigenvalue computations

Some matrix equations

- Sylvester matrix equation

$$A\mathbf{X} + \mathbf{X}B + D = 0$$

Eigenvalue pbs and tracking, Control, MOR, Assignment pbs, Riccati eqn

- Lyapunov matrix equation

$$A\mathbf{X} + \mathbf{X}A^\top + D = 0, \qquad D = D^\top$$

Stability analysis in Control and Dynamical systems, Signal processing, eigenvalue computations

- Algebraic Riccati equation

$$A\mathbf{X} + \mathbf{X}A^\top - \mathbf{X}BB^\top\mathbf{X} + D = 0, \qquad D = D^\top$$

Lancaster-Rodman '95, Konstantinov-Gu-Mehrmann-Petkov, '02, Bini-Iannazzo-Meini '12

Some matrix equations

- Sylvester matrix equation

$$A\mathbf{X} + \mathbf{X}B + D = 0$$

Eigenvalue pbs and tracking, Control, MOR, Assignment pbs, Riccati eqn

- Lyapunov matrix equation

$$A\mathbf{X} + \mathbf{X}A^\top + D = 0, \qquad D = D^\top$$

Stability analysis in Control and Dynamical systems, Signal processing, eigenvalue computations

- Algebraic Riccati equation

$$A\mathbf{X} + \mathbf{X}A^\top - \mathbf{X}BB^\top\mathbf{X} + D = 0, \qquad D = D^\top$$

Lancaster-Rodman '95, Konstantinov-Gu-Mehrmann-Petkov, '02, Bini-Iannazzo-Meini '12

Focus: All or some of the matrices are large (and possibly sparse)

Solving the Lyapunov equation. The problem

Approximate X in:

$$AX + XA^\top + BB^\top = 0$$

$A \in \mathbb{R}^{n\times n}$ neg.real $\qquad B \in \mathbb{R}^{n\times p}, \qquad 1 \le p \ll n$

Solving the Lyapunov equation. The problem

Approximate X in:

$$AX + XA^\top + BB^\top = 0$$

$A \in \mathbb{R}^{n\times n}$ neg.real $\qquad B \in \mathbb{R}^{n\times p}, \qquad 1 \le p \ll n$

Time-invariant linear system:

$$\mathbf{x}'(t) = A\mathbf{x}(t) + B\mathbf{u}(t), \qquad \mathbf{x}(0) = x_0$$

Closed form solution:

$$X = \int_0^\infty e^{-tA} BB^\top e^{-tA^\top} dt$$

$\Rightarrow \qquad X$ symmetric semidef.

see, e.g., Antoulas '05, Benner '06

Large linear systems:

$$Ax = b, \quad A \in \mathbb{R}^{n \times n}$$

- Krylov subspace methods (CG, MINRES, GMRES, BiCGSTAB, etc.)
- Preconditioners: find P such that

$$AP^{-1}\widetilde{x} = b \qquad x = P^{-1}\widetilde{x}$$

is easier and fast to solve

Linear systems vs linear matrix equations

Large linear systems:

$$Ax = b, \quad A \in \mathbb{R}^{n\times n}$$

- Krylov subspace methods (CG, MINRES, GMRES, BiCGSTAB, etc.)
- Preconditioners: find P such that

$$AP^{-1}\widetilde{x} = b \qquad x = P^{-1}\widetilde{x}$$

  is easier and fast to solve

Large linear matrix equations:

$$AX + XA^\top + BB^\top = 0$$

- No preconditioning - to preserve symmetry
- X is a large, dense matrix $\Rightarrow$ low rank approximation

$$X \approx \widetilde{X} = ZZ^\top, \quad Z \text{ tall}$$

Linear systems vs linear matrix equations

Large linear systems:

$$Ax = b, \quad A \in \mathbb{R}^{n \times n}$$

- Krylov subspace methods (CG, MINRES, GMRES, BiCGSTAB, etc.)
- Preconditioners: find P such that

$$AP^{-1}\widetilde{x} = b \qquad x = P^{-1}\widetilde{x}$$

  is easier and fast to solve

Large linear matrix equations:

$$AX + XA^\top + BB^\top = 0$$

Kronecker formulation:

$$(A \otimes I + I \otimes A)x = b \qquad x = \mathrm{vec}(X)$$

Projection-type methods

Given an approximation space $\mathcal{K}$,

$$X \approx X_m \qquad \mathrm{col}(X_m) \in \mathcal{K}$$

Galerkin condition: $R := AX_m + X_m A^\top + BB^\top \quad \perp \quad \mathcal{K}$

$$V_m^\top R V_m = 0 \qquad\qquad \mathcal{K} = \mathrm{Range}(V_m)$$

Projection-type methods

Given an approximation space $\mathcal{K}$,

$$X \approx X_m \qquad \mathrm{col}(X_m) \in \mathcal{K}$$

Galerkin condition: $\quad R := AX_m + X_m A^\top + BB^\top \quad \perp \quad \mathcal{K}$

$$V_m^\top R V_m = 0 \qquad\qquad \mathcal{K} = \mathrm{Range}(V_m)$$

Assume $V_m^\top V_m = I_m$ and let $X_m := V_m Y_m V_m^\top$.

Projected Lyapunov equation:

$$V_m^\top (A V_m Y_m V_m^\top + V_m Y_m V_m^\top A^\top + BB^\top) V_m \quad = \quad 0$$

Projection-type methods

Given an approximation space $\mathcal{K}$,

$$X \approx X_m \qquad \mathrm{col}(X_m) \in \mathcal{K}$$

Galerkin condition: $\quad R := AX_m + X_m A^\top + BB^\top \quad \perp \quad \mathcal{K}$

$$V_m^\top R V_m = 0 \qquad\qquad \mathcal{K} = \mathrm{Range}(V_m)$$

Assume $V_m^\top V_m = I_m$ and let $X_m := V_m Y_m V_m^\top$.

Projected Lyapunov equation:

$$\begin{aligned} V_m^\top (A V_m Y_m V_m^\top + V_m Y_m V_m^\top A^\top + BB^\top) V_m &= 0 \\ (V_m^\top A V_m) Y_m + Y_m (V_m^\top A^\top V_m) + V_m^\top B B^\top V_m &= 0 \end{aligned}$$

Early contributions: Saad '90, Jaimoukha & Kasenally '94, for $\mathcal{K} = \mathcal{K}_m(A, B) = \mathrm{Range}([B, AB, \ldots, A^{m-1}B])$

More recent options as approximation space

Enrich space to decrease space dimension

- Extended Krylov subspace

$$\mathcal{K} = \mathcal{K}_m(A, B) + \mathcal{K}_m(A^{-1}, A^{-1}B),$$

that is, $\mathcal{K} = \mathrm{Range}([B, A^{-1}B, AB, A^{-2}B, A^2, A^{-3}B, \ldots,])$

(Druskin & Knizhnerman '98, Simoncini '07)

More recent options as approximation space

Enrich space to decrease space dimension

- Extended Krylov subspace

$$\mathcal{K} = \mathcal{K}_m(A, B) + \mathcal{K}_m(A^{-1}, A^{-1}B),$$

that is, $\mathcal{K} = \mathrm{Range}([B, A^{-1}B, AB, A^{-2}B, A^2, A^{-3}B, \ldots,])$

(Druskin & Knizhnerman '98, Simoncini '07)

- Rational Krylov subspace

$$\mathcal{K} = \mathrm{Range}([B, (A - s_1 I)^{-1}B, \ldots, (A - s_m I)^{-1}B])$$

usually, $\{s_1, \ldots, s_m\} \subset \mathbb{C}^+$ chosen a-priori

More recent options as approximation space

Enrich space to decrease space dimension

- Extended Krylov subspace

$$\mathcal{K} = \mathcal{K}_m(A, B) + \mathcal{K}_m(A^{-1}, A^{-1}B),$$

that is, $\mathcal{K} = \mathrm{Range}([B, A^{-1}B, AB, A^{-2}B, A^2, A^{-3}B, \ldots,])$

(Druskin & Knizhnerman '98, Simoncini '07)

- Rational Krylov subspace

$$\mathcal{K} = \mathrm{Range}([B, (A - s_1 I)^{-1}B, \ldots, (A - s_m I)^{-1}B])$$

usually, $\{s_1, \ldots, s_m\} \subset \mathbb{C}^+$ chosen a-priori

In both cases, for Range$(V_m) = \mathcal{K}$, projected Lyapunov equation:

$$(V_m^\top A V_m)Y_m + Y_m(V_m^\top A^\top V_m) + V_m^\top B B^\top V_m = 0$$

$X_m = V_m Y_m V_m^\top$

Rational Krylov Subspaces. A long tradition...

In general,

$$K_m(A, B, \mathbf{s}) = \mathrm{Range}([(A - s_1 I)^{-1} B, (A - s_2 I)^{-1} B, \ldots, (A - s_m I)^{-1} B])$$

- Eigenvalue problems (Ruhe, 1984)
- Model Order Reduction (transfer function evaluation)
- In Alternating Direction Implicit iteration (ADI) for linear matrix equations

Rational Krylov Subspaces in MOR. Choice of poles.

$$K_m(A, B, \mathbf{s}) = \mathrm{Range}([(A - s_1 I)^{-1}B, (A - s_2 I)^{-1}B, \ldots, (A - s_m I)^{-1}B])$$

cf. General discussion in Antoulas, 2005.

Many contributions:

- Gallivan, Grimme, Van Dooren (1996–, ad-hoc poles)
- Penzl (1999-2000, ADI shifts - preprocessing, Ritz values)
- ....
- Sabino (2006 - tuning within preprocessing)
- IRKA – Gugercin, Antoulas, Beattie (2008)
- Druskin, Lieberman, Simoncini, Zaslavski (adaptive greedy procedure)
- Güttel, Knizhnerman (black-box for matrix functions)
- ....

Alternating Direction Implicit iteration (ADI) - Wachspress

(see, e.g., Li 2000, Penzl 2000)

$$X_0 = 0,\ X_j = -2p_j(A + p_jI)^{-1}BB^\top(A + p_jI)^{-\top} \quad j = 1, \ldots, \ell$$
$$+(A + p_jI)^{-1}(A - p_jI)X_{j-1}(A - p_jI)^\top(A + p_jI)^{-\top}$$

with

$$\phi_\ell(t) = \prod_{j=1}^{\ell}(t - p_j), \quad \{p_1, \ldots, p_\ell\} = \text{argmin} \max_{t\in\Lambda(A)} \left|\frac{\phi_\ell(t)}{\phi_\ell(-t)}\right|$$

Implementation aspects: Benner, Saak, Quintana-Ortì[2], ...

Convergence depends on choice of poles $\{p_j\}$

More advanced approach: Galerkin-Projection Accelerated ADI (Benner, Saak, tr 2010)

ADI and Rational Krylov subspaces

Let $B = b$ (vector). Main consideration (see, e.g., Li, Wright 2000)

$$\mathrm{col}(X_m^{(ADI)}) \in K_m(A, b, \mathbf{s})$$

and also, for $U_m = [(A - s_1 I)^{-1} b, \ldots, (A - s_m I)^{-1} b]$,

$$X_m^{(ADI)} = U_m \boldsymbol{\alpha}^{-1} U_m^*$$

with $\boldsymbol{\alpha}$ Cauchy matrix

ADI and Rational Krylov subspaces

Let $B = b$ (vector). Main consideration (see, e.g., Li, Wright 2000)

$$\mathrm{col}(X_m^{(ADI)}) \in K_m(A, b, \mathbf{s})$$

and also, for $U_m = [(A - s_1 I)^{-1}b, \ldots, (A - s_m I)^{-1}b]$,

$$X_m^{(ADI)} = U_m \boldsymbol{\alpha}^{-1} U_m^*$$

with $\boldsymbol{\alpha}$ Cauchy matrix

Equivalence between ADI and RKSM:

ADI coincides with the Galerkin solution X_m in Rational Krylov space if and only if

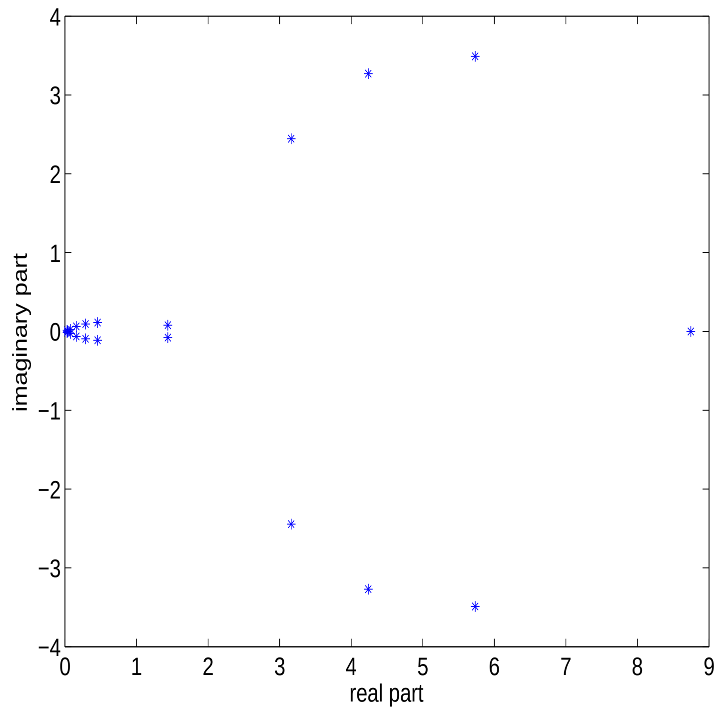
$$s_j = -\bar{\lambda}_j$$

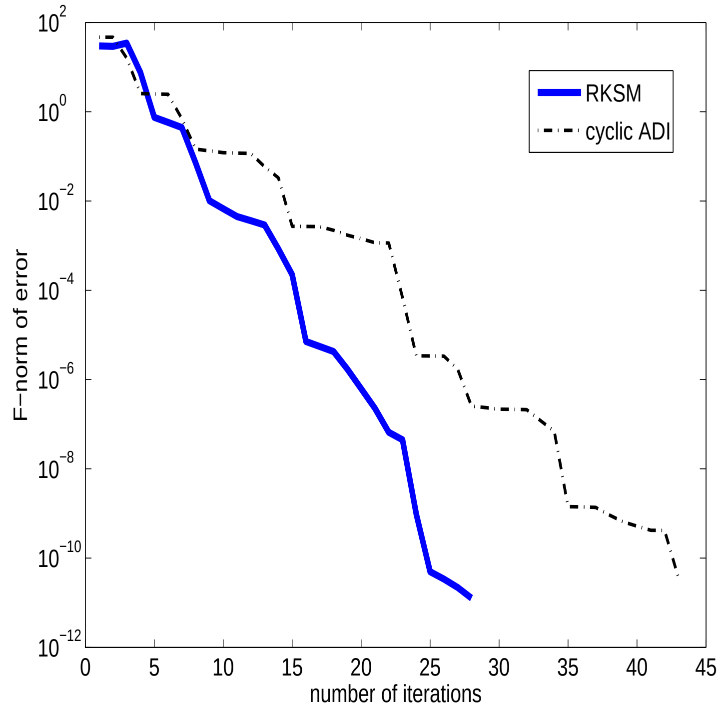
where $\lambda_j = \mathrm{eigs}(V_m^* A V_m)$ Ritz values (suitably ordered)

Druskin, Knizhnerman, S. '11, Beckermann '11, Flagg '09, Gugercin, Flagg '12

Typical behavior of ADI and generic RKSM for the same poles

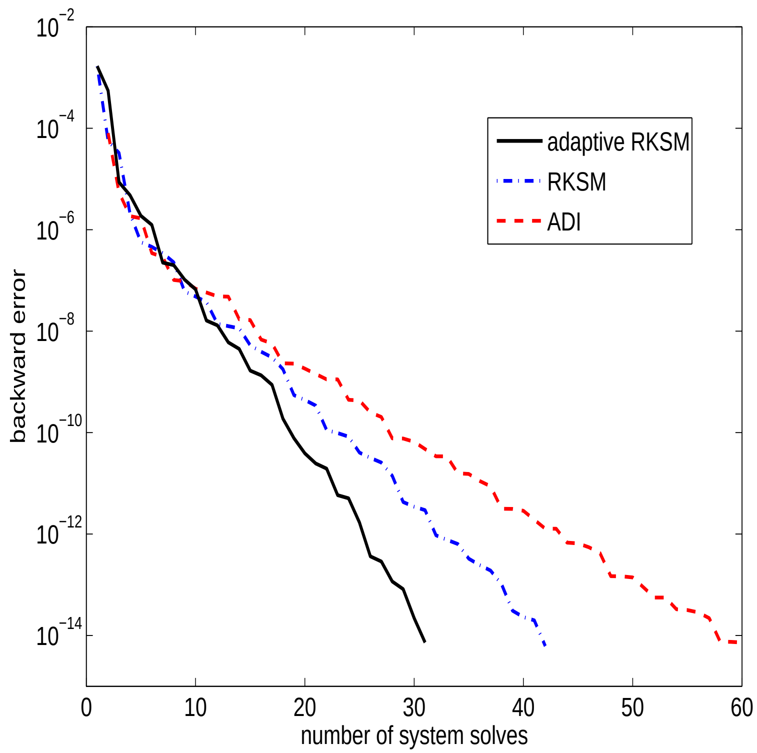
Operator: $L(u) = -\Delta u + (50xu_x)_x + (50yu_y)_y$ on $[0,1]^2$

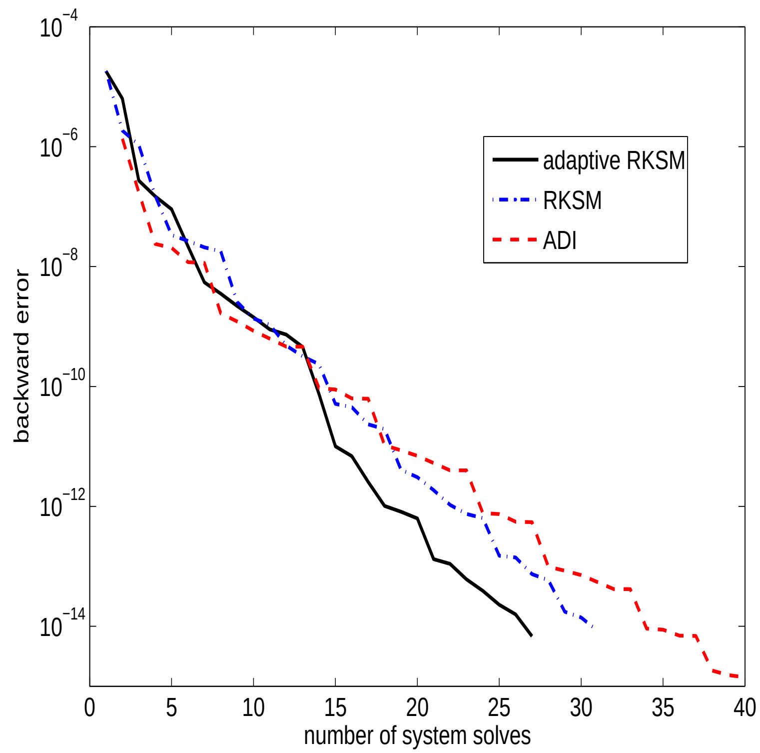




Same non-optimal 20 poles, repeated cyclically.

Expected performance (from Oberwolfach Collection)





Left: `rail` problem, A symmetric.

Right: `flow_meter_model_v0.5` problem, A nonsymmetric.

ADI and RKSM use 10 non-optimal poles cyclically (computed a-priori with `lyapack`, Penzl 2000)

Multiterm linear matrix equation

$$A_1XB_1 + A_2XB_2 + \ldots + A_\ell XB_\ell = C$$

Applications:

- Matrix least squares
- Control
- Stochastic PDEs
- ...

Multiterm linear matrix equation

$$A_1 X B_1 + A_2 X B_2 + \ldots + A_\ell X B_\ell = C$$

Applications:

- Matrix least squares
- Control
- Stochastic PDEs
- ...

Main device: Kronecker formulation

$$\left(B_1^\top \otimes A_1 + \ldots + B_\ell^\top \otimes A_\ell\right) x = c$$

Iterative methods: matrix-matrix multiplications and rank truncation

(Benner, Breiten, Bouhamidi, Chehab, Damm, Grasedyck, Jbilou, Kressner, Matthies, Onwunta, Raydan, Stoll, Tobler, Zander, ...)

Other related matrix equations

More "exotic" linear matrix equations

- Sylvester-like

$$BX + f(X)A = C$$

typically (but not only!)

$$f(X) = \bar{X}, \quad f(X) = X^\top, \quad \text{or} \quad f(X) = X^*$$

(Bevis, Braden, Byers, Chiang, De Terán, Dopico, Duan, Feng, Guillery, Hall, Hartwig, Ikramov, Kressner, Montealegre, Reyes, Schröder, Vorntsov, Watkins, Wu, ...)

The $\top$-Sylvester matrix equations

Solve for X:

$$AX + X^\top B = C, \qquad (*)$$

⇒ A unique solution exists for any $C \in \mathbb{R}^{n\times n}$ iff $A - \lambda B^\top$ is regular and spec$(A, B^\top) \setminus \{1\}$ is reciprocal free (with 1 having at most algebraic multiplicity 1)

⇒ Small scale: Bartel-Stewart type algorithm

(De Teran, Dopico, 2011)

⇒ If X_0 is the unique solution to the *Sylvester* eqn

$$AXA^\top - B^\top XB = C - C^\top A^{-1} B$$

then X_0 is the unique solution to $(*)$

The large scale $\top$-Sylvester matrix equations

$$AX + X^\top B = C_1 C_2^\top, \qquad C_1, C_2 \in \mathbb{R}^{n\times r}, \; r \ll n$$

Find:

$$X \approx X_m = \mathcal{V}_m Y_m \mathcal{W}_m^\top \in \mathbb{R}^{n\times n}$$

Orthogonality (*Petrov-Galerkin*) condition:

$$\mathcal{W}_m^\top (AX_m + X_m^\top B - C_1 C_2^\top)\mathcal{W}_m = 0$$

(the orthogonality space is different from the approximation space)

Reduced T-Sylvester equation:

$$(\mathcal{W}_m^\top A \mathcal{V}_m) Y_m + Y_m^\top (\mathcal{V}_m^\top B \mathcal{W}_m) = (\mathcal{W}_m^\top C_1)(\mathcal{W}_m^\top C_2)^\top$$

Key issue: Choice of $\mathcal{V}_m, \mathcal{W}_m$

The selection of $\mathcal{V}_m, \mathcal{W}_m$

Exploit the generalized Schur decomposition:

$$A = W T_A V^\top \quad \text{and} \quad B^\top = W T_B V^\top$$

(W, V orthogonal) from which

$$B^{-\top} A = V\, T_B^{-1} T_A\, V^\top \quad \text{and} \quad B^\top V = W T_B$$

$$B^{-\top} A V = V\, T_B^{-1} T_A \quad \text{and} \quad B^\top V = W T_B$$

Therefore:

$\mathrm{Range}(\mathcal{V}_m)$ $\leftarrow$ good approx to invariant subspaces of $B^{-\top} A$

$\mathrm{Range}(\mathcal{W}_m) = B^\top \mathrm{Range}(\mathcal{V}_m)$

The selection of $\mathcal{V}_m, \mathcal{W}_m$

Exploit the generalized Schur decomposition:

$$A = W T_A V^\top \quad \text{and} \quad B^\top = W T_B V^\top$$

(W, V orthogonal) from which

$$B^{-\top} A = V\, T_B^{-1} T_A\, V^\top \quad \text{and} \quad B^\top V = W T_B$$

$$B^{-\top} A V = V\, T_B^{-1} T_A \quad \text{and} \quad B^\top V = W T_B$$

Therefore:

$\mathrm{Range}(\mathcal{V}_m)$ $\leftarrow$ good approx to invariant subspaces of $B^{-\top} A$

$\mathrm{Range}(\mathcal{W}_m) = B^\top \mathrm{Range}(\mathcal{V}_m)$

$\mathrm{Range}(\mathcal{V}_m) = \mathcal{K}_m(B^{-\top} A, B^{-\top}[C_1, C_2]), \qquad \mathrm{Range}(\mathcal{W}_m) = B^\top \mathrm{Range}(\mathcal{V}_m)$

The selection of $\mathcal{V}_m, \mathcal{W}_m$

$$\mathrm{Range}(\mathcal{V}_m) = \mathcal{K}_m(B^{-\top}A, B^{-\top}[C_1, C_2]), \qquad \mathrm{Range}(\mathcal{W}_m) = B^{\top}\mathrm{Range}(\mathcal{V}_m)$$

Algorithmic considerations:

- $\mathrm{Range}(\mathcal{W}_m) = \mathcal{K}_m(AB^{-\top}, [C_1, C_2])$ so that

$$\mathrm{Range}(C_1) \cup \mathrm{Range}(C_2) \subset \mathrm{Range}(\mathcal{W}_m)$$

- If $C_1 = C_2$ then

$$\mathrm{Range}(\mathcal{V}_m) = \mathcal{K}_m(B^{-\top}A\,, B^{-\top}C_1)$$

- The role of A and B can be reversed

  $(A \to B^{\top},\ B \to A^{\top},\ C_1 \leftrightarrow C_2)$

Remark: Enriched spaces can be used...

Computational considerations

$n = 10^4$. A and B: finite difference discretizations in $[0,1]^2$ of

$$a(u) = (-\mathrm{exp}(-xy)\, u_x)_x + (-\mathrm{exp}(xy)\, u_y)_y + 100\, x\, u_x + \gamma\, u$$

$$b(u) = -u_{xx} - u_{yy}, \qquad \gamma = 5 \cdot 10^4$$

$\texttt{tol} = 10^{-10}$	EK	BK	BK-TR	EK-SYLV
iterations	8	83	8	8
dim. approx. space	32	166	16	32
time (seconds)	1.7	58.1	0.7	2.4

BK-TR: Standard Krylov subspace, roles of A and B reversed

All eigenvalues of $B^{-\top}A$ are well outside the unit circle

Computational considerations

$n = 10^4$. A and B: finite difference discretizations in $[0,1]^2$ of

$$a(u) = (-\text{exp}(-xy)\, u_x)_x + (-\text{exp}(xy)\, u_y)_y + 100\, x\, u_x + \gamma u\,,$$

$$b(u) = -u_{xx} - u_{yy} + 100\, x\, u_x\,, \qquad \gamma = 5 \cdot 10^4$$

$\texttt{tol} = 10^{-10}$	EK	BK*	BK-TR*	EK-SYLV*
iterations	29	100	100	100
dim. approx. space	116	200	200	400
time (seconds)	10.9	70.7	63.8	521.2

eigenvalues of $B^{-\top}A$ are now located inside *and* outside the unit circle

Conclusions

- Large advances in solving really large linear matrix equations
- Matrix equation challenges rely on strength and maturity of linear system solvers

Some references

⋆ V. S.,
Computational methods for linear matrix equations
SIAM Review, v.58, n.3, 2016

⋆ Froilan Dopico , Javier Gonzalez, Daniel Kressner and V. S.
Projection methods for large T-Sylvester equations
Mathematics of Computations, v.85, n.301, 2016

⋆ Stephen D. Shank , V. S. and Daniel B. Szyld
Efficient low-rank solutions of Generalized Lyapunov equations
Numerische Mathematik, v.134, n.2, 2016