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

- Lyapunov matrix equation

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

Stability analysis in Control and Dynamical systems, Signal processing

- Sylvester matrix equation

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

Eigenvalue problems and tracking, Control, MOR, Assignment problems, Riccati equation

- Algebraic Riccati equation

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

books: Lancaster, Rodman 1995, Bini, Iannazzo, Meini 2012

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

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

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

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

Projection-type methods

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

$$X \approx X_m \qquad X_m \in \mathcal{K}$$

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

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

Projection-type methods

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

$$X \approx X_m \qquad X_m \in \mathcal{K}$$

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 $\mathcal{K}$,

$$X \approx X_m \qquad X_m \in \mathcal{K}$$

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:

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

Early contributions: Saad '90, Jaimoukha & Kasenally '94, for

$\mathcal{K} = \mathcal{K}_m(A, B) = \mathrm{Range}([B, AB, \ldots, A^{m-1}B])$

Standard Krylov projection. In quest of a-priori error bounds

$$AX + XA^\top + BB^\top = 0, \qquad X \approx X_m \in K_m(A,B),\ B \in \mathbb{R}^{n\times 1}$$

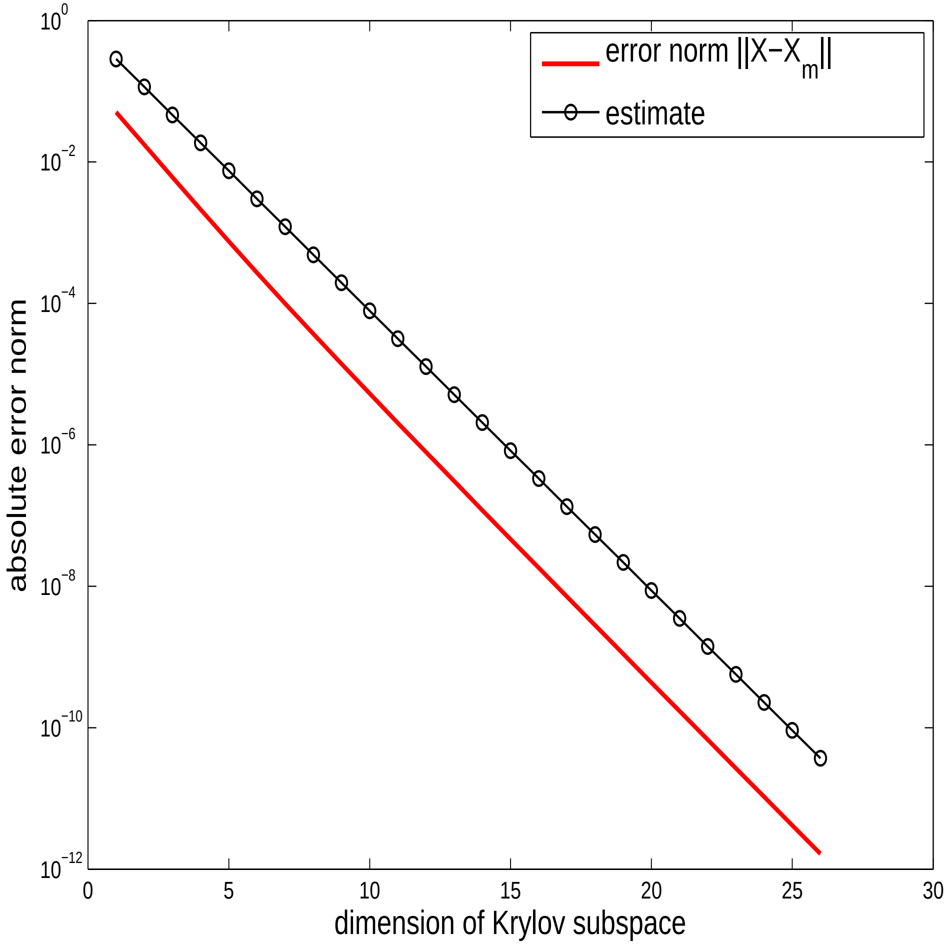
$A < 0$ symmetric, $\quad |\lambda_{\min}| \le \ldots \le |\lambda_{\max}|$ eigs of $A \qquad \|B\| = 1$

Let $\quad \hat\kappa := \frac{\lambda_{\min}+\lambda_{\max}}{2\lambda_{\min}}$. Then

$$\|X - X_m\| \quad \le \quad \frac{\sqrt{\hat\kappa}+1}{\hat\lambda_{\min}\sqrt{\hat\kappa}} \left(\frac{\sqrt{\hat\kappa}-1}{\sqrt{\hat\kappa}+1}\right)^m$$

Note: same rate as CG for $(A + \lambda_{\min} I)z = b$

The case of A symmetric. An example



A: 400×400 diagonal with uniformly distributed eigenvalues in $[1, 10]$ ($\lambda_{\min} = 1$)

The case of Field of Values in an ellipse

Assume Field of Values of A in $\mathbb{C}^-$

(E ellipse of center $(c, 0)$, foci $(c \pm d, 0)$ and major semi-axis a)

Then

$$\|X - X_m\| \le \frac{4}{\alpha_{\min}} \frac{r_2}{r_2 - r} \left(\frac{r}{r_2}\right)^m$$

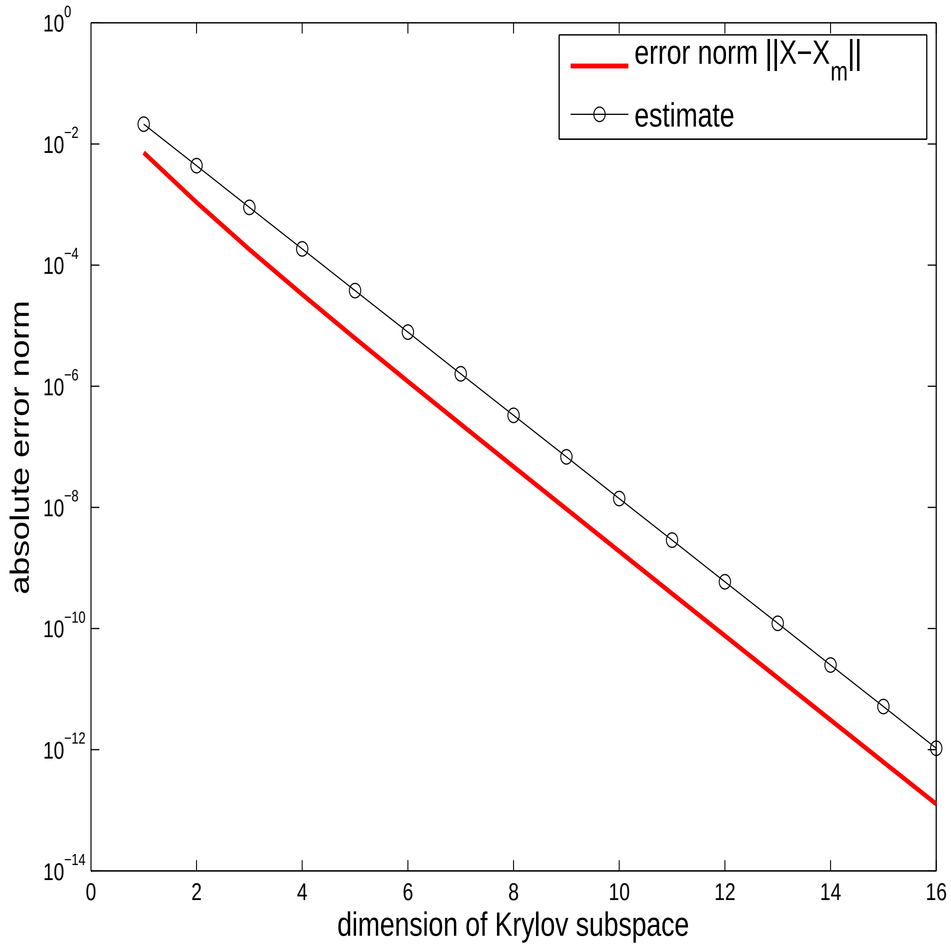
where

$$r = \frac{a}{d} + \sqrt{\left(\frac{a}{d}\right)^2 - 1}, \quad r_2 = \frac{c + \alpha_{\min}}{d} + \sqrt{\left(\frac{c + \alpha_{\min}}{d}\right)^2 - 1}$$

$\alpha_{\min}$ eig. of $\frac{1}{2}(A + A^\top)$ closest to the origin

Note: same rate as FOM for $(A + \alpha_{\min} I)z = b$

The case of Field of values in an ellipse. An example



A normal with eigenvalues on an elliptic curve

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$(\mathcal{V}_m) = \mathcal{K}$, projected Lyapunov equation:

$$(\mathcal{V}_m^\top A \mathcal{V}_m) Y_m + Y_m (\mathcal{V}_m^\top A^\top \mathcal{V}_m) + \mathcal{V}_m^\top B B^\top \mathcal{V}_m = 0$$

$X_m = \mathcal{V}_m Y_m \mathcal{V}_m^\top$

Rational Krylov Subspaces. A long tradition...

In general,

$$K_m(A, B, \mathbf{s}) = \text{span}\{(A - s_1 I)^{-1}B, \ldots, (A - s_m I)^{-1}B\}$$

- 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.

$$K_m(A, B, \mathbf{s}) = \text{span}\{(A - s_1 I)^{-1} B, (A - s_2 I)^{-1} B, \ldots, (A - s_m I)^{-1} B\}$$

cf. General discussion in Antoulas, 2005.

Various attempts:

- 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)

Adaptive choice of poles for Rational Krylov space

$$K_m(A, b, \mathbf{s}) = \text{span}\{(A - s_1 I)^{-1} b, \ldots, (A - s_m I)^{-1} b\}, \quad B = b$$

$\mathbf{s} = [s_1, \ldots, s_m]$ to be chosen sequentially

Greedy procedure. Define:

$$r_m(z) = \prod_{j=1}^{m} \frac{z - \lambda_j}{z - s_j}, \qquad \lambda_j = \text{eigs}(\mathcal{V}_m^* A \mathcal{V}_m)$$

(r_m residual of a related linear system)

Adaptive choice of poles for Rational Krylov space

$$K_m(A, b, \mathbf{s}) = \text{span}\{(A - s_1 I)^{-1} b, \ldots, (A - s_m I)^{-1} b\}, \quad B = b$$

$\mathbf{s} = [s_1, \ldots, s_m]$ to be chosen sequentially

Greedy procedure. Define:

$$r_m(z) = \prod_{j=1}^{m} \frac{z - \lambda_j}{z - s_j}, \qquad \lambda_j = \text{eigs}(\mathcal{V}_m^* A \mathcal{V}_m)$$

(r_m residual of a related linear system)

$$s_{m+1} := \arg\left(\max_{s \in \partial \mathcal{S}_m} \frac{1}{|r_m(s)|} \right)$$

where $\mathcal{S}_m \subset \mathbb{C}^+$ approximately encloses the eigenvalues of $-A$

(Druskin, Lieberman, Zaslavski '10, Druskin, Simoncini '11)

Some numerical experiments

- Adaptive Rational Krylov Subspace method
- Extended Krylov Subspace method (Rational method with fixed poles)

$$\mathbf{EK}_m(A,b) = K_m(A,b) + K_m(A^{-1}, A^{-1}b)$$

Comparison measures:

- Efficiency (CPU time)
- Memory (space dimension)
- Rank of solution

The RAIL (symmetric) data set

n		Rational space direct	Extended space direct		
1357	CPU time (s)	0.84	0.36		
	dim. Approx. Space	21	64		
	Rank of Solution	21	47		
20209	CPU time (s)	11.19	10.97		
	dim. Approx. Space	25	124		
	Rank of Solution	25	75		
79841	CPU time (s)	51.54	73.03		
	dim. Approx. Space	26	168		
	Rank of Solution	26	103		

The RAIL (symmetric) data set

n		Rational space direct	Extended space direct	Rational space iterative	Extended space iterative
1357	CPU time (s)	0.84	0.36	0.96	1.60
	dim. Approx. Space	21	64	21	68
	Rank of Solution	21	47	21	45
20209	CPU time (s)	11.19	10.97	25.31	201.94
	dim. Approx. Space	25	124	25	126
	Rank of Solution	25	75	25	75
79841	CPU time (s)	51.54	73.03	189.48	2779.95
	dim. Approx. Space	26	168	26	170
	Rank of Solution	26	103	26	98

Inner solves: PCG with IC(10^{-2})

More Tests: two nonsymmetric problems

n		Rational space direct	Extended space direct	Rational space iterative	Extended space iterative
9669	CPU time (s)	3.16	3.06	3.01	9.95
	dim. Approx. Space	16	36	16	36
	Rank of Solution	16	24	16	24
20082	CPU time (s)	59.99	45.84	13.01	25.28
	dim. Approx. Space	15	26	15	26
	Rank of Solution	15	22	15	22

Convective thermal flow problems (FLOW, CHIP data sets)

⋆ All real shifts used

Alternating Direction Implicit iteration (ADI) - Wachspress

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

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

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 $\{p_j\}$. For $A < 0$ sym and one pole:

$$\|X - X_\ell\| \approx \left(\frac{\sqrt{\kappa_{adi}} - 2}{\sqrt{\kappa_{adi}} + 2} \right)^\ell, \qquad \kappa_{adi} = \frac{\lambda_{\max}}{\lambda_{\min}}$$

ADI and Rational Krylov subspaces

Main consideration (see, e.g., Li, Wright 2000)

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

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

Main consideration (see, e.g., Li, Wright 2000)

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

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 of ADI with RKSM:

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

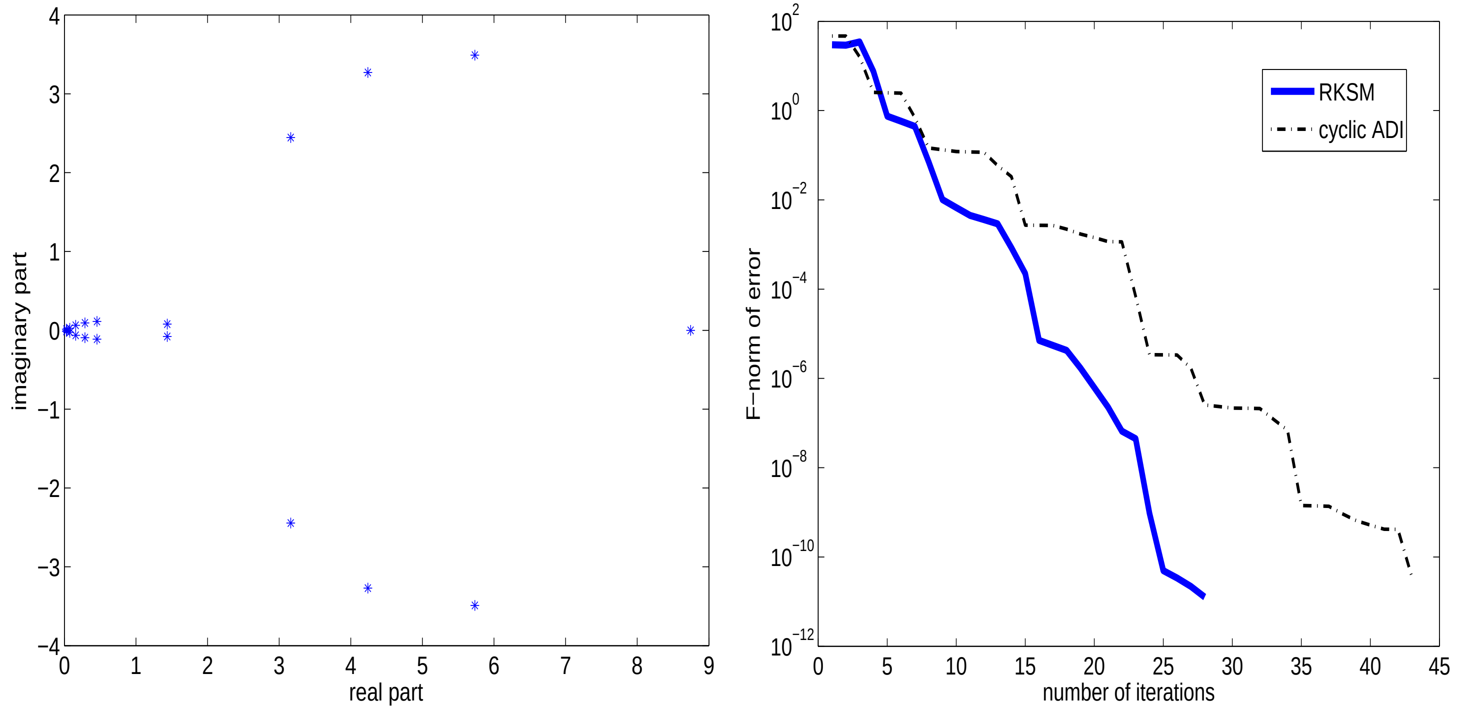
$$s_j = -\bar{\lambda}_j$$

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

Druskin, Knizhnerman, Simoncini '11, Beckermann '11 (and Flagg '09)

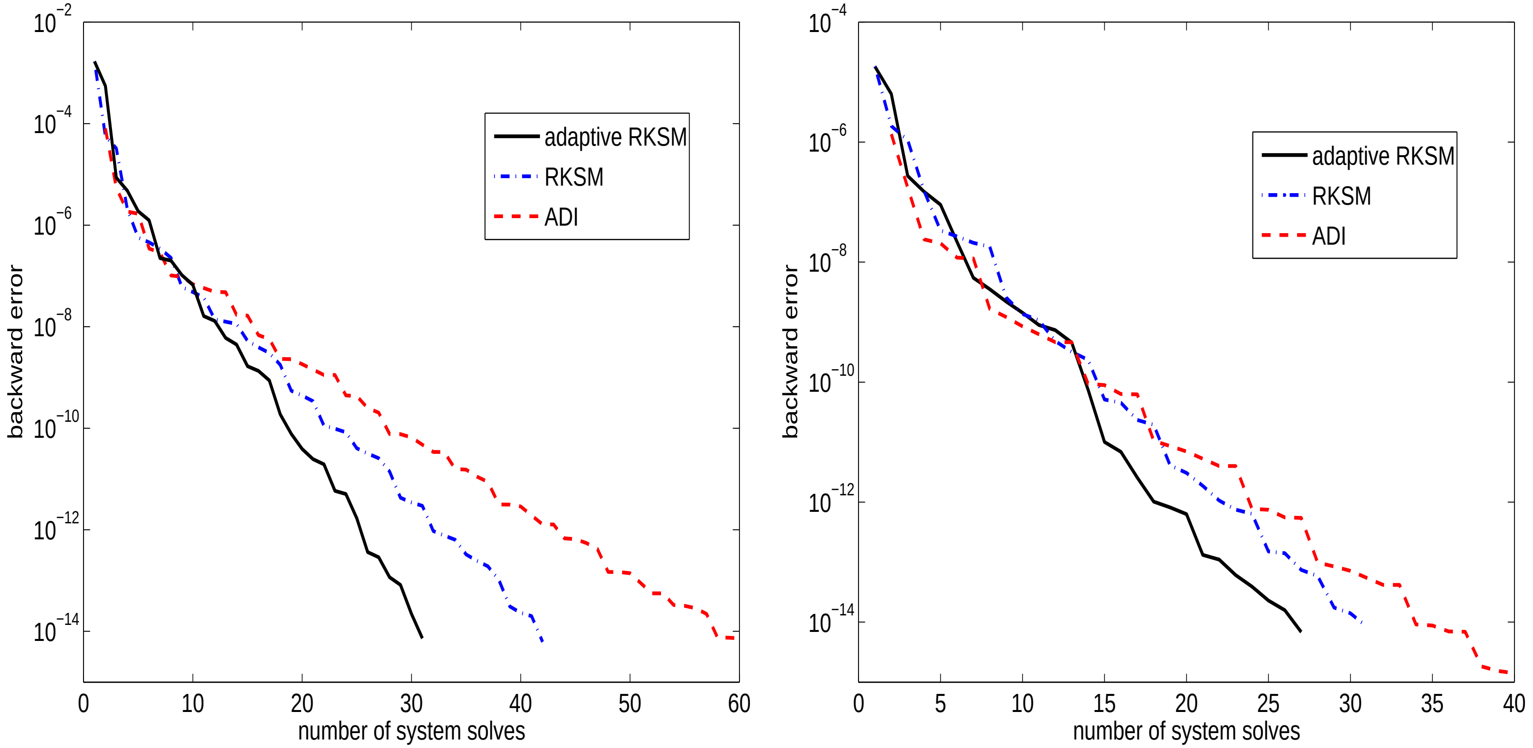
Typical behavior of ADI and generic RKSM for the same poles

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)

Other recent approaches and convergence results

- Kronecker Formulation
- Galerkin-Projection Accelerated ADI (Benner, Saak, tr 2010)

Different aims:

- IRKA (Gugercin, Antoulas, Beattie, 2008)
- Riemann optimization approach (Vandereycken, Vandewalle, 2010)

Other recent approaches and convergence results

- Kronecker Formulation
- Galerkin-Projection Accelerated ADI (Benner, Saak, tr 2010)

Different aims:

- IRKA (Gugercin, Antoulas, Beattie, 2008)
- Riemann optimization approach (Vandereycken, Vandewalle, 2010)

****

Further Convergence results:

- Extended Krylov for Lyapunov eqn (Knizhnerman, Simoncini 2011)
- Rational Krylov for Lyapunov eqn (Druskin, Knizhnerman, Simoncini '11)
- Improved, and for the Sylvester eqn (Beckermann 2011)

Conclusions

General Considerations:

- Large advances in solving really large linear matrix equations
- Second order difficulties exploit strength of linear system solvers

Conclusions

General Considerations:

- Large advances in solving really large linear matrix equations
- Second order difficulties exploit strength of linear system solvers

On-going projects:

- Tangential adapt. RKSM, rank(B)$\gg 1$ (Druskin, S., Zaslavsky, tr'12)
- Minimal residual methods by projection (Lin, S., tr 2012)

$$\min_{\widetilde{X} \in \mathcal{K}} \|A\widetilde{X} + \widetilde{X}A^\top + BB^\top\|_F$$

- Constrained Sylvester equations (Shank, S., in progress)

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

- On Projection methods for (quadratic) Riccati equation

  (Heyouni, Jbilou, 2009, S., Szyld, Monsalve tr. 2012)

Some References

⋆ V. S., *A new iterative method for solving large-scale Lyapunov matrix equations. SISC (2007)*

⋆ V. S. and V. Druskin, *Convergence analysis of projection methods for the numerical solution of large Lyapunov equations. SINUM (2009)*

⋆ L. Knizhnerman and V. S., *Convergence analysis of the Extended Krylov Subspace Method for the Lyapunov equation. Numerische Mathematik (2011)*

⋆ V. Druskin and V. S., *Adaptive rational Krylov subspaces for large-scale dynamical systems. Systems & Control Letters (2011)*

⋆ V. Druskin, L. Knizhnerman and V. S., *Analysis of the rational Krylov subspace and ADI methods for solving the Lyapunov equation. SINUM (2011)*

⋆ V. Druskin, V. S. and M. Zaslavsky, *Adaptive tangential interpolation in rational Krylov subspaces for MIMO model reduction data pp. 1-22, Nov.'12.*