Analysis of projection methods for solving large-scale Lyapunov matrix equations

V. Simoncini

Dipartimento di Matematica, Università di Bologna

valeria@dm.unibo.it

Joint work with V. Druskin and work in progress with L. Knizhnerman

The problem

Approximate X in:

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

$A \in \mathbb{R}^{n\times n}$ pos.definite $\quad (x^\top(A + A^\top)x > 0,\ x \neq 0)$

$B \in \mathbb{R}^{n\times s}$ $\quad$ here: $B = b\ (s = 1)$

The problem

Approximate X in:

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

$A \in \mathbb{R}^{n\times n}$ pos.definite $\quad (x^\top(A + A^\top)x > 0,\ x \neq 0)$

$B \in \mathbb{R}^{n\times s}$ $\quad$ here: $B = b$ $(s = 1)$

Applications: signal processing, system and control theory

Time-invariant linear system:

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

see, e.g., Antoulas 2005, Benner 2006

Standard Krylov subspace projection

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

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

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

Standard Krylov subspace projection

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

Galerkin condition: $\quad R := AX_m + X_mA^\top + bb^\top \quad \perp \quad \mathcal{K}$

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

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

Projected Lyapunov equation:

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

$$\Updownarrow$$

$$T_m Y_m + Y_m T_m^\top \quad + \quad e_1 e_1 = 0$$

with $b = V_m e_1$ (Saad, 1990, for $\mathcal{K} = \mathcal{K}_m(A, b)$; Jaimoukha & Kasenally, 1994)

Other related approaches

- Extended projection: Different selection of $\mathcal{K}$, e.g.,

  $\mathcal{K} = \mathcal{K}_m(A, B) \cup \mathcal{K}_m(A^{-1}, B)$ (Druskin-Knizhnerman 1998, S., 2007)

Other related approaches

- Extended projection: Different selection of $\mathcal{K}$, e.g.,

$$\mathcal{K} = \mathcal{K}_m(A, B) \cup \mathcal{K}_m(A^{-1}, B)$$ (Druskin-Knizhnerman 1998, S., 2007)

- "Global" projection: (Jbilou, Messaoudi, Riquet, Sadok, 1999, 2006)

$$\mathrm{range}(\mathcal{V}) = \mathcal{K}_m(A, B), \qquad \mathcal{V} = [V_1, \ldots, V_m]$$

$$\mathrm{trace}(V_i^\top V_j) = 0, i \neq j, \mathrm{trace}(V_i^\top V_i) = 1$$

Other related approaches

- Extended projection: Different selection of $\mathcal{K}$, e.g.,

$$\mathcal{K} = \mathcal{K}_m(A,B) \cup \mathcal{K}_m(A^{-1},B)$$ (Druskin-Knizhnerman 1998, S., 2007)

- "Global" projection: (Jbilou, Messaoudi, Riquet, Sadok, 1999, 2006)

$$\mathrm{range}(\mathcal{V}) = \mathcal{K}_m(A,B), \qquad \mathcal{V} = [V_1, \ldots, V_m]$$

$$\mathrm{trace}(V_i^\top V_j) = 0, i \neq j, \mathrm{trace}(V_i^\top V_i) = 1$$

- Kronecker formulation: (Preconditioning: Hochbruck & Starke, 1995)

$$AX + XA^\top + BB^\top = 0 \;\Leftrightarrow\; (A \otimes I + I \otimes A)\mathrm{vec}(X) + \mathrm{vec}(BB^\top) = 0$$

Other related approaches

- Extended projection: Different selection of $\mathcal{K}$, e.g.,

$$\mathcal{K} = \mathcal{K}_m(A, B) \cup \mathcal{K}_m(A^{-1}, B)$$ (Druskin-Knizhnerman 1998, S., 2007)

- "Global" projection: (Jbilou, Messaoudi, Riquet, Sadok, 1999, 2006)

$$\text{range}(\mathcal{V}) = \mathcal{K}_m(A, B), \qquad \mathcal{V} = [V_1, \ldots, V_m]$$
$$\text{trace}(V_i^\top V_j) = 0, i \neq j, \text{trace}(V_i^\top V_i) = 1$$

- Kronecker formulation: (Preconditioning: Hochbruck & Starke, 1995)

$$AX + XA^\top + BB^\top = 0 \;\Leftrightarrow\; (A \otimes I + I \otimes A)\text{vec}(X) + \text{vec}(BB^\top) = 0$$

- Cyclic low-rank Smith: (see, e.g., Li 2000, Penzl 2000, Benner etal)

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

with $r_\ell(t) = \Pi_{j=1}^{\ell}(t - p_j)$, $\{p_1, \ldots, p_\ell\} = \text{argmin}\max_{t \in \Lambda(A)} |r_\ell(t)/r_\ell(-t)|$

Convergence results and a-priori bounds

- Kronecker formulation: all available results for

$$\mathcal{A}x = f, \qquad \mathcal{A} \in \mathbb{R}^{n^2 \times n^2}$$

- Global projection methods: only a-posteriori estimates (?)
- Cyclic low-rank Smith method: results based on

$$r_\ell(t) = \Pi_{j=1}^{\ell}(t - p_j), \quad \{p_1, \ldots, p_\ell\} = \operatorname{argmin} \max_{t \in \Lambda(A)} |r_\ell(t)/r_\ell(-t)|$$

- Standard Krylov projection: (Robbè & Sadkane, 2002)

$$\|AX_m^g + X_m^g A^\top + BB^\top\|_F \le \left(1 - \frac{d^2}{\|\mathcal{S}\|^2}\right)^{m/2} \|BB^\top\|_F$$

$$d = \operatorname{dist}(\mathcal{F}(A), \mathcal{F}(-A)) > 0, \quad \mathcal{S} : X \mapsto AX + XA^\top$$

(X_m^g Petrov-Galerkin, originally for the Sylvester equation)

The case of A symmetric

$$AX + XA^\top + BB^\top = 0, \qquad X \approx X_m \in \mathcal{K}_m(A, B)$$

$A = A^\top$ symmetric

Let $0 < \hat\lambda_{\min} \le \ldots \le \hat\lambda_{\max}$ eigs of $A + \lambda_{\min} I$, $\qquad \hat\kappa := \hat\lambda_{\max}/\hat\lambda_{\min}$

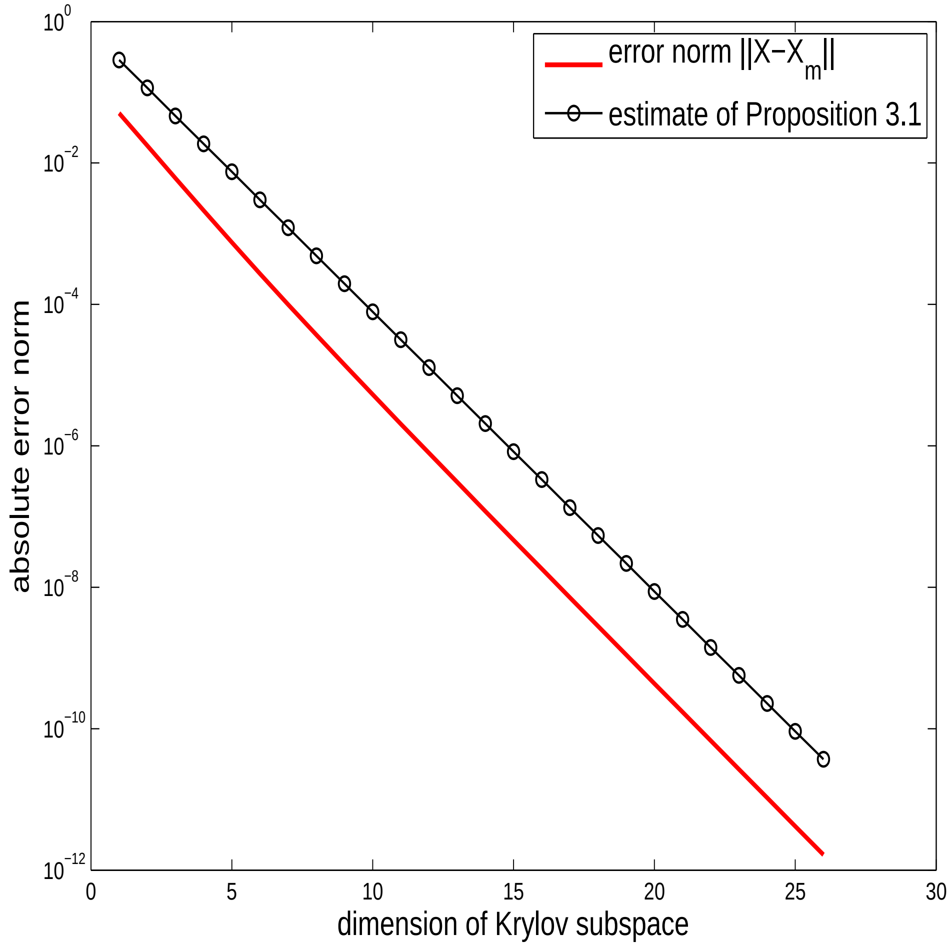
Then

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

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

(S. & Druskin 2007, cf. works of Knizhnerman)

The case of A symmetric. An example



A: 400×400 diagonal with uniformly distributed eigenvalues in $[1, 10]$

The case of $\mathcal{F}(A)$ in an ellipse $E \subset \mathbb{C}^+$

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

Then

$$\|X - X_m\| = \mathcal{O}\left(\left(\frac{r}{r_2}\right)^m\right)$$

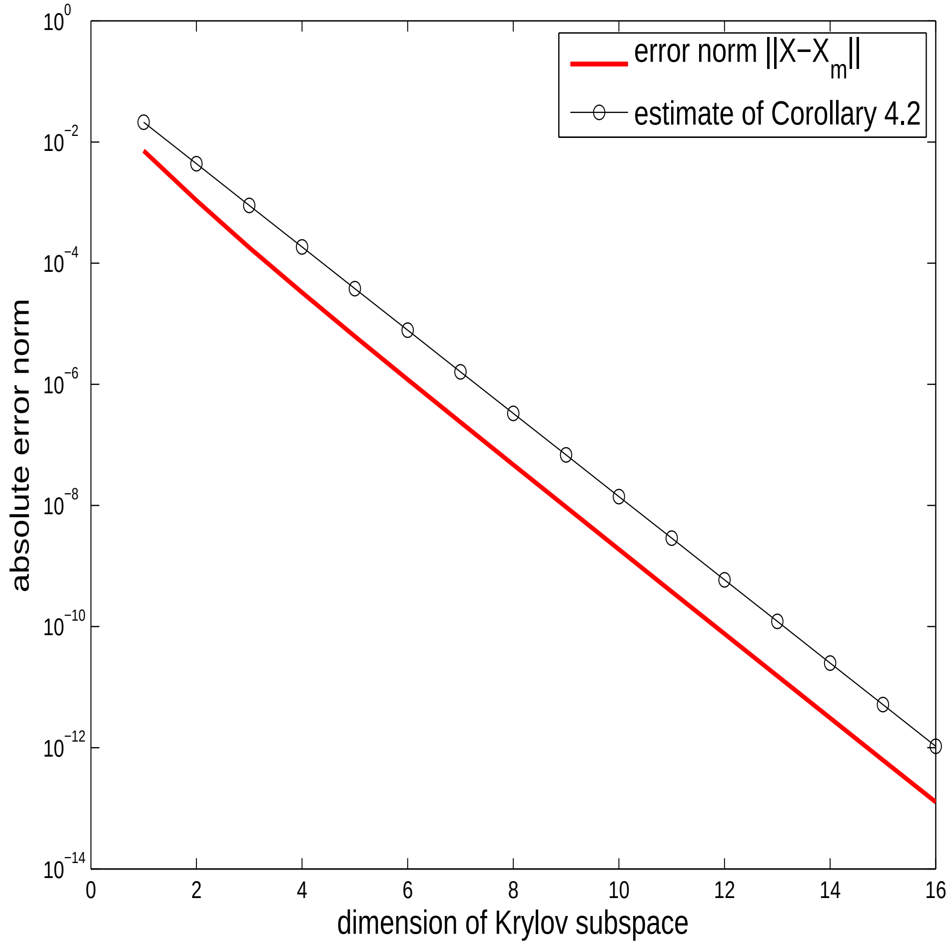
where

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

and $\alpha_{\min} = \lambda_{\min}((A + A^\top)/2) > 0$

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

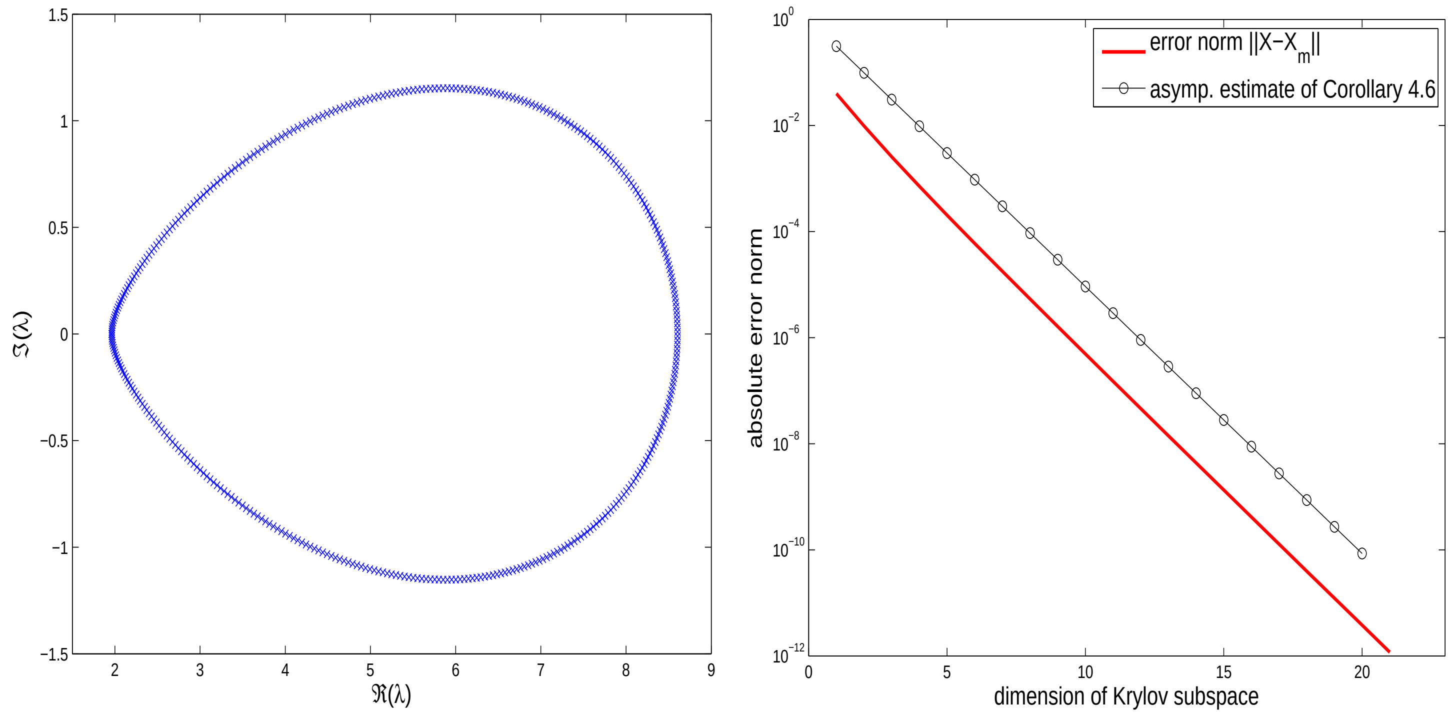
The case of $\mathcal{F}(A)$ in an ellipse. An example



A normal with eigenvalues on an elliptic curve

The case of $\mathcal{F}(A)$ in a wedge-shaped set. An example

Generalization to a wedge-shaped convex set of $\mathbb{C}^+$.

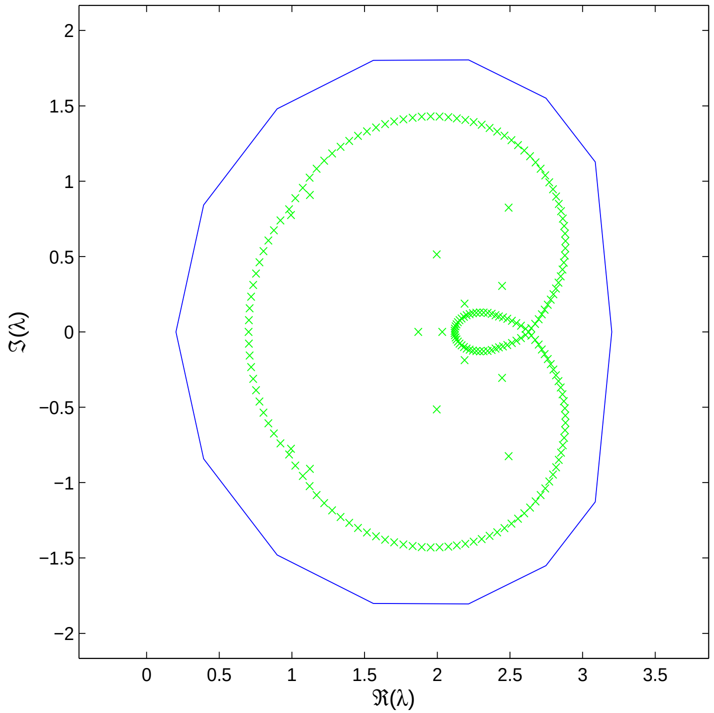


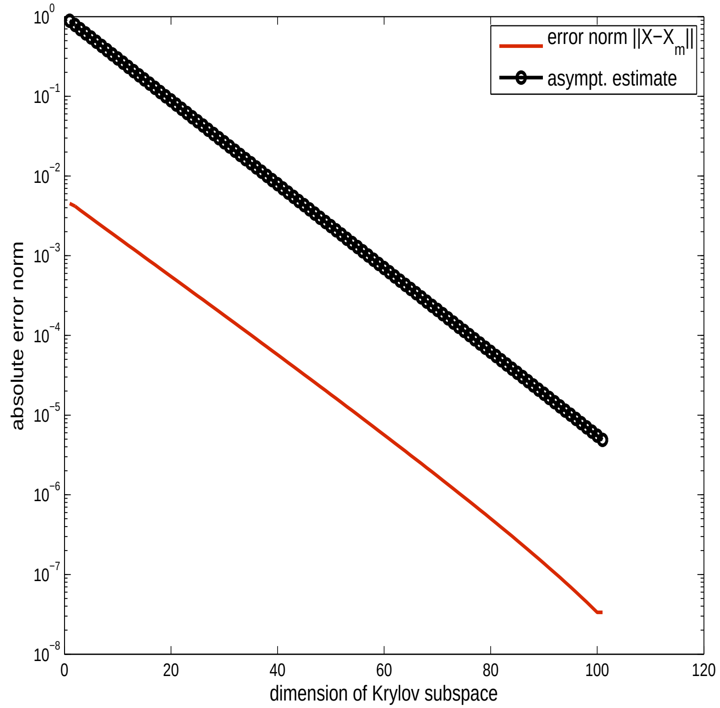
A: diagonal (normal) matrix on the wedge-shaped curve.

(Inclusion set from Hochbruck & Lubich, 1997)

The case of $\mathcal{F}(A)$ in a numerically determined set

Generalization to a Schwarz-Christoffel mapping (Driscoll, Trefethen 2002)





$A:$ Toeplitz$(-1, -1, \underline{2}, 0.1)$

(SC Matlab Toolbox, T. Driscoll 1996)

Extended Krylov subspace method

Galerkin condition: $\mathcal{X}_m \in \mathcal{K}$ s.t.

$$R := A\mathcal{X}_m + \mathcal{X}_m A^\top + bb^\top \quad \perp \quad \mathcal{K}$$

$$\mathcal{K} = \mathcal{K}_m(A, B) \cup \mathcal{K}_m(A^{-1}, B), \qquad \text{range}(\mathcal{V}_m) = \mathcal{K}$$

(Druskin-Knizhnerman 1998, S., 2007)

Projected Lyapunov equation:

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

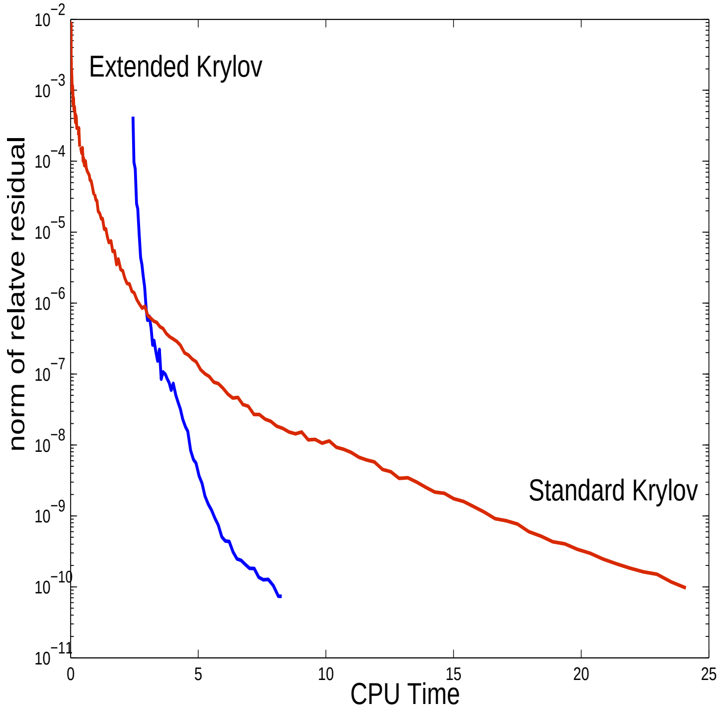
$$\Updownarrow$$

$$\mathcal{T}_m Y_m + Y_m \mathcal{T}_m^\top \quad + \quad e_1 e_1 = 0$$

Performance evaluation

$$\mathbf{x}' = \mathbf{x}_{xx} + \mathbf{x}_{yy} + \mathbf{x}_{zz} - 10x\mathbf{x}_x - 1000y\mathbf{x}_y - 10z\mathbf{x}_z + \mathbf{b}(x,y)\mathbf{u}(t)$$

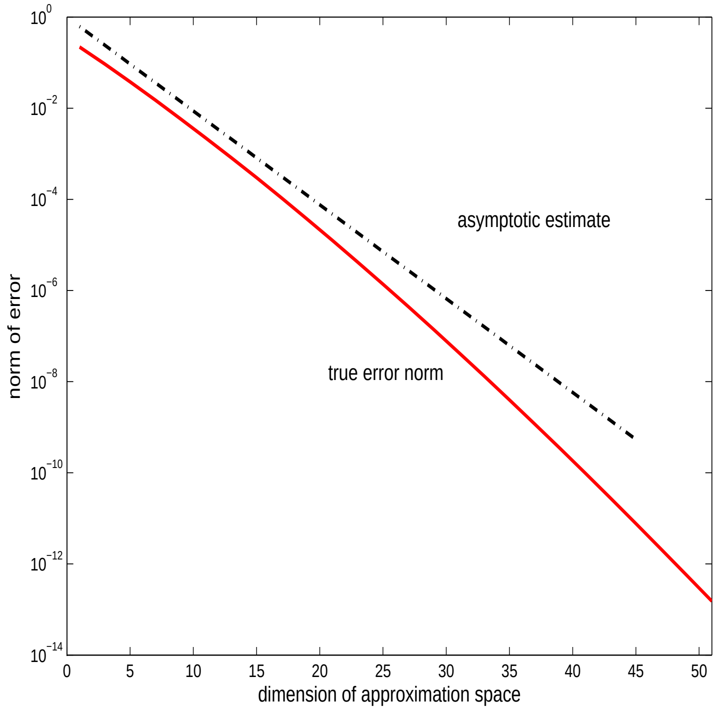
A matrix $\quad 18^3 \times 18^3$



approximation space dim.: 146 (Standard Krylov) 112 (Extended Krylov)

Convergence analysis of Extended Krylov: A symmetric

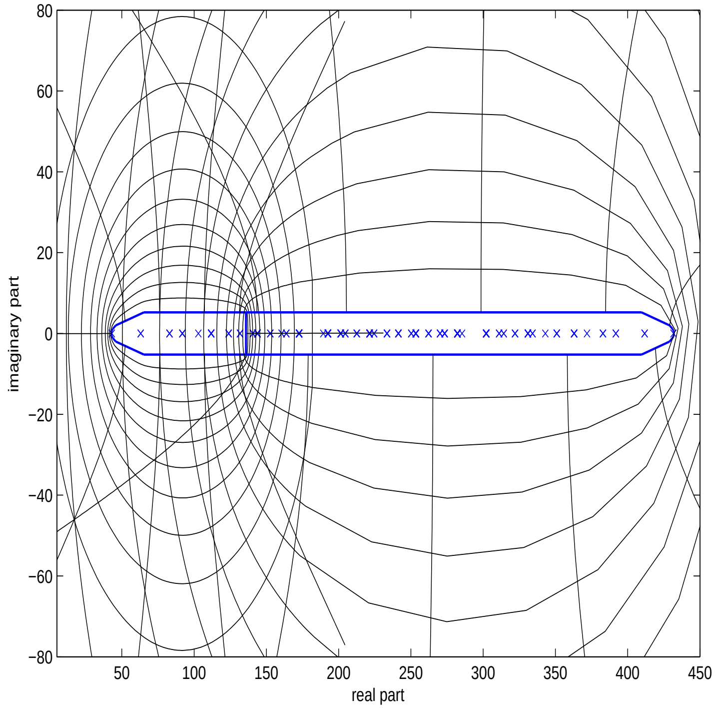
Conjecture: $$\|X - \mathcal{X}_m\| \approx \left(\frac{\hat\kappa^{1/4}-1}{\hat\kappa^{1/4}+1}\right)^m, \ m = 1, 2, \ldots,.$$

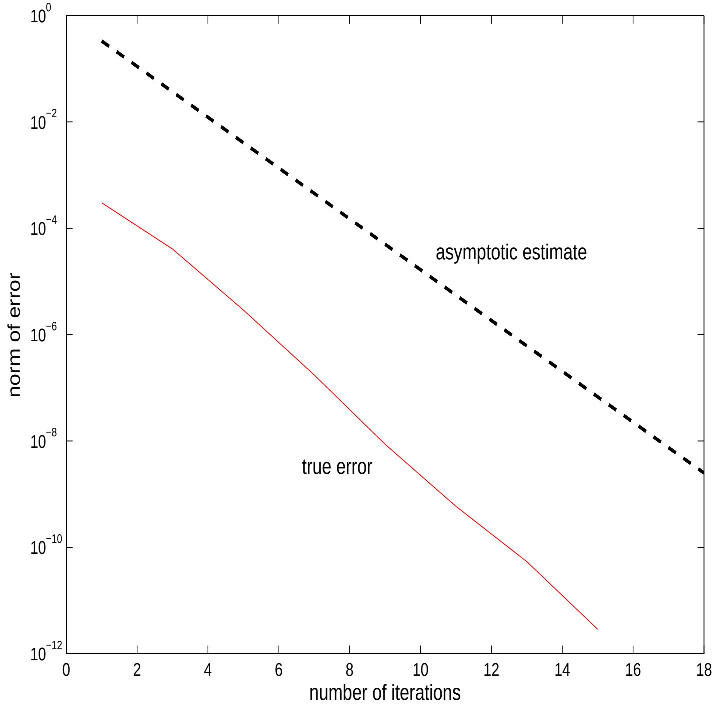


Idea: $$[\hat\lambda_{\min}, \hat\lambda_{\max}] = [\hat\lambda_{\min}, \chi] \cup [\chi, \hat\lambda_{\max}]$$

χ s.t. $$\text{cond}([\hat\lambda_{\min}, \chi]) = \text{cond}([\chi, \hat\lambda_{\max}])$$

Convergence analysis of Extended Krylov: A nonsymmetric



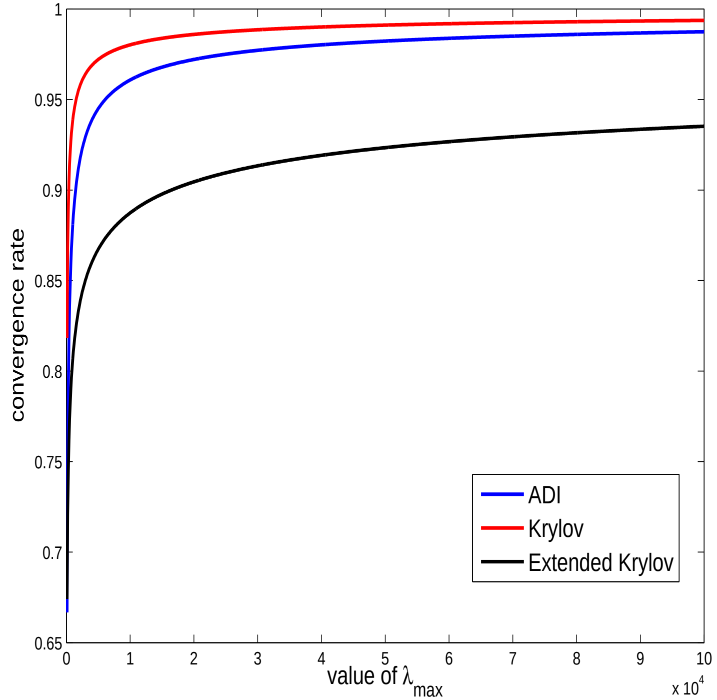


A: discretization of the 3D operator $\mathcal{L}(u) = -\Delta u + u_x + u_y$

Comparison of convergence rates: A symmetric

ADI iteration: $\varepsilon_{adi,j} \approx \left(\frac{\sqrt{\kappa_{adi}}-2}{\sqrt{\kappa_{adi}}+2}\right)^j$ Standard Krylov: $\varepsilon_{kr,j} \approx \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}-1}\right)^j$

Extended Krylov: $\varepsilon_{ek,\ell} \approx \left(\left(\frac{\sqrt[4]{\kappa}-1}{\sqrt[4]{\kappa}-1}\right)^{1/2}\right)^\ell$



$\lambda_{\min} = 1, \ \lambda_{\max} \in [10^2, 10^5]$

Conclusions and future work

- Good understanding of convergence of Standard Krylov method
- Similar results for B tall matrix

- ⋆ Extended Krylov: $\mathcal{K} = \mathcal{K}_m(A, B) \cup \mathcal{K}_m(A^{-1}, B)$ prove conjectures
- ⋆ Connection with the convergence theory of other methods
- ⋆ New acceleration procedures

V.Simoncini, *A new iterative method for solving large-scale Lyapunov matrix equations*. *SIAM J. Sci. Comput.*, 29(3):1268–1288, 2007.

V. Simoncini and V. Druskin, *Convergence analysis of projection methods for the numerical solution of large Lyapunov equations*. August 2007.

Available at `www.dm.unibo.it/~simoncin`