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

V. Simoncini

Dipartimento di Matematica, Università di Bologna

valeria@dm.unibo.it

Partially based on joint work with V. Druskin

The problem

Approximate X in:

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

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

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

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

Standard Krylov subspace projection

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

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

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

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

Projected Lyapunov equation:

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

$$\Updownarrow$$

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

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

Other related approaches

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

$$\mathcal{K} = \mathcal{K}_m(A, B) \cup \mathcal{K}_m(A^{-1}, B) \qquad \text{(Simoncini, 2007)}$$

Other related approaches

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

$$\mathcal{K} = \mathcal{K}_m(A, B) \cup \mathcal{K}_m(A^{-1}, B) \qquad \text{(Simoncini, 2007)}$$

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

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

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

Other related approaches

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

$$\mathcal{K} = \mathcal{K}_m(A, B) \cup \mathcal{K}_m(A^{-1}, B) \qquad \text{(Simoncini, 2007)}$$

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

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

- 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

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

$$\mathcal{K} = \mathcal{K}_m(A, B) \cup \mathcal{K}_m(A^{-1}, B) \qquad \text{(Simoncini, 2007)}$$

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

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

- 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 method: (see, e.g., Li 2000, Penzl 2000)

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

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)

Standard Krylov projection. In quest of error bounds

$$AX + XA^\top + BB^\top = 0, \qquad X \approx X_m \in K_m(A, B)$$

$$\|X - X_m\| \le ??$$

Analytic solution:

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

with $x = \exp(-tA)B$.

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

$$\|x\| \le \exp(-t\alpha_{\min})\|B\|$$

First (key) step

Krylov subspace proj.: $X_m = V_m Y_m V_m^\top$, range$(V_m) = K_m(A, b)$

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

Clearly,

$$X_m = V_m \left(\int_0^\infty e^{-tT_m} e_1 e_1^\top e^{-tT_m^\top} dt \right) V_m^\top = \int_0^\infty x_m x_m^\top dt$$

with $x_m = V_m \exp(-tT_m) e_1$, and $\|x_m\| \leq \exp(-t\alpha_{\min})$.

First (key) step

Krylov subspace proj.: $X_m = V_m Y_m V_m^\top$, range$(V_m) = K_m(A, b)$

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

Clearly,

$$X_m = V_m \left(\int_0^\infty e^{-tT_m} e_1 e_1^\top e^{-tT_m^\top} dt \right) V_m^\top = \int_0^\infty x_m x_m^\top dt$$

with $x_m = V_m \exp(-tT_m) e_1$, and $\|x_m\| \le \exp(-t\alpha_{\min})$.

Therefore $\quad \|X - X_m\| = \| \int_0^\infty (xx^\top - x_m x_m^\top) dt\|$, so that

$$\|X - X_m\| \le \int_0^\infty \|xx^\top - x_m x_m^\top\| dt \le 2 \int_0^\infty e^{-t\alpha_{\min}} \|x - x_m\| dt$$

Lyapunov equation and “standard” linear systems

$$\|X - X_m\| \;\leq\; 2\int_0^\infty e^{-t\alpha_{\min}}\,\|x - x_m\| dt$$

with

$$e^{-t\alpha_{\min}}\,\|x - x_m\| = \|e^{-t(A+\alpha_{\min}I)}b - V_m e^{-t(T_m+\alpha_{\min}I)}e_1\|$$

Lyapunov equation and "standard" linear systems

$$\|X - X_m\| \;\le\; 2\int_0^\infty e^{-t\alpha_{\min}}\, \|x - x_m\| dt$$

with

$$e^{-t\alpha_{\min}}\, \|x - x_m\| = \|e^{-t(A+\alpha_{\min} I)} b - V_m e^{-t(T_m+\alpha_{\min} I)} e_1\|$$

From $z^{-1} = \int_0^\infty e^{-tz} dt$,

$$(A + \alpha_{\min} I)^{-1} b \quad - \quad V_m (T_m + \alpha_{\min} I)^{-1} e_1 = \int_0^\infty \left(e^{-t(A+\alpha_{\min} I)} b - V_m e^{-t(T_m+\alpha_{\min} I)} e_1 \right) dt$$

The case of A symmetric

A symmetric $\Rightarrow$ $\alpha_{\min} = \lambda_{\min}(A)$

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

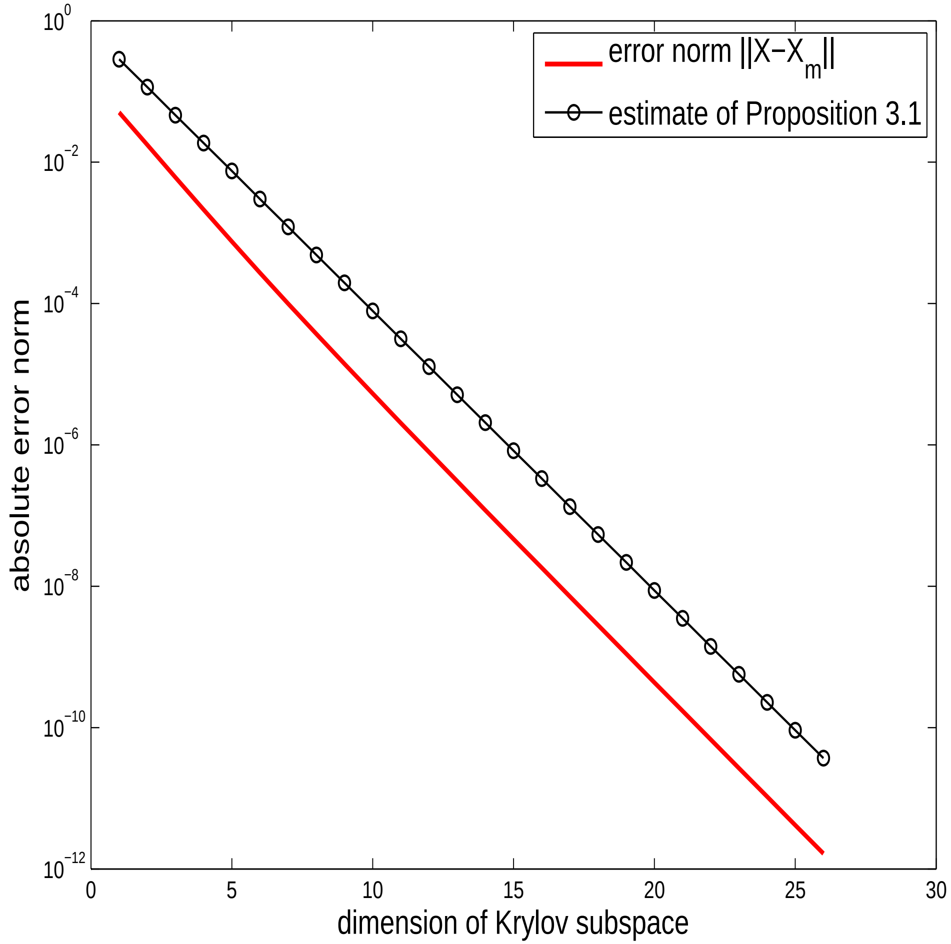
Then

$$\|X - X_m\| \leq \frac{\sqrt{\hat{\kappa}}+1}{\hat{\lambda}_{\min}\sqrt{\hat{\kappa}}} \left(\frac{\sqrt{\hat{\kappa}}-1}{\sqrt{\hat{\kappa}}+1}\right)^m$$

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

(cf. works of Knizhnerman)

The case of A symmetric. An example



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

$(\alpha_{\min} = \lambda_{\min} = 1)$

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

Assume $\mathcal{F}(A) \subseteq E \subset \mathbb{C}^+$

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

Then

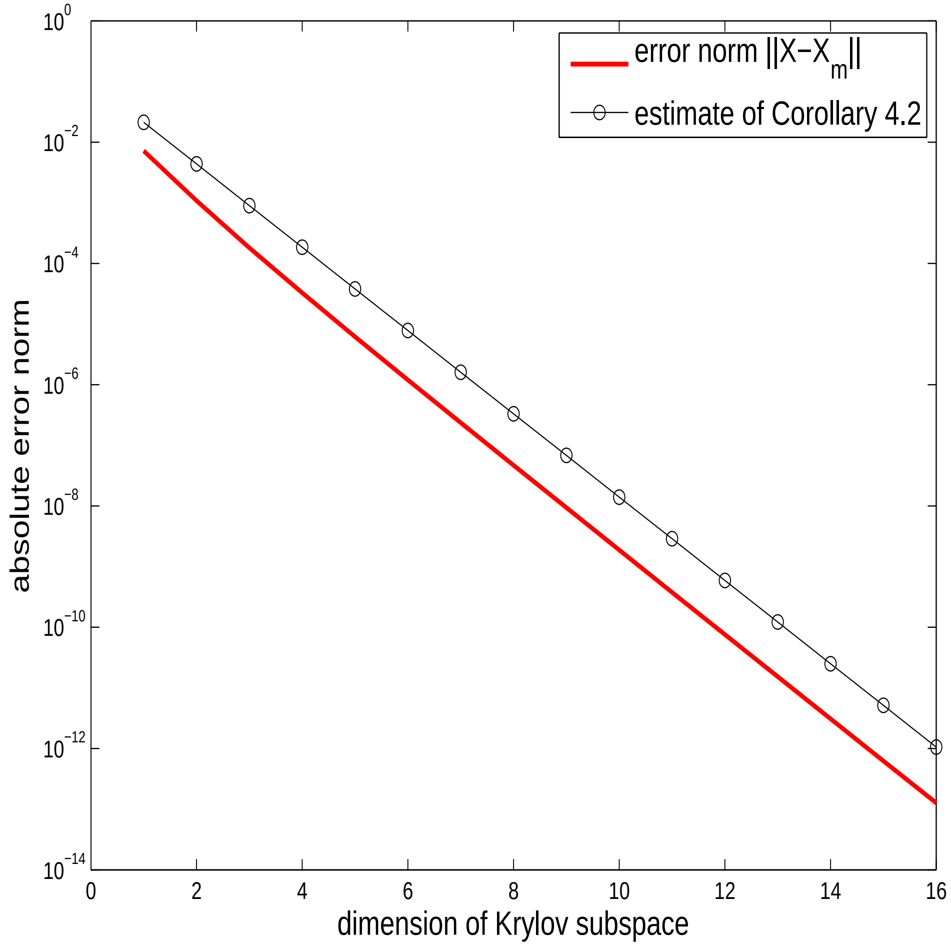
$$\|X - X_m\| \leq \frac{4}{\alpha_{\min}} \frac{r_2}{r_2 - r} \left(\frac{r}{r_2}\right)^m$$

where

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

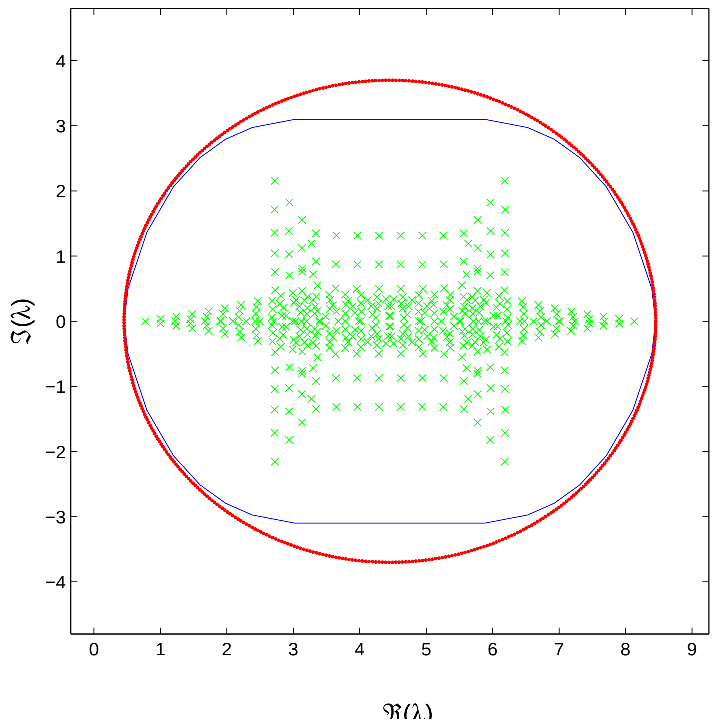
Note: same rate as FOM for $(A + \alpha_{\min} I)z = b$

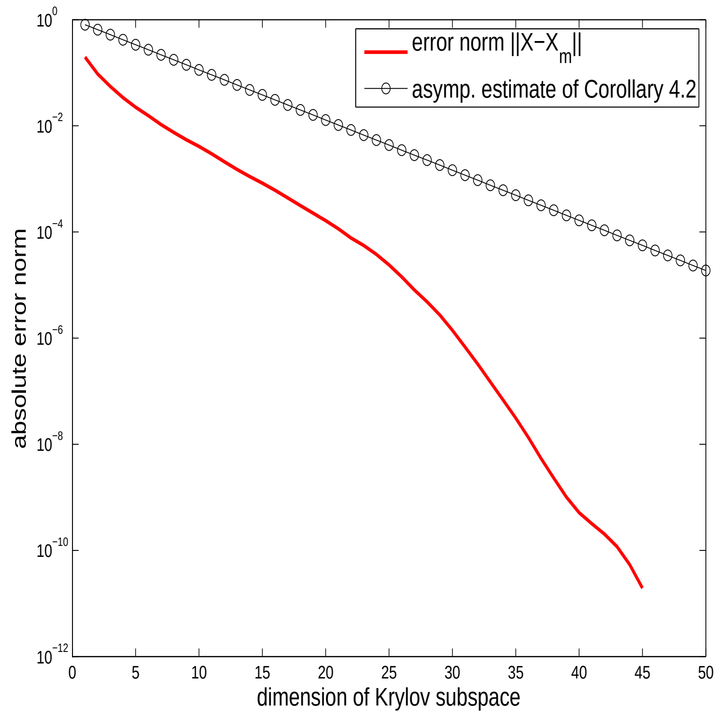
The case of $\mathcal{F}(A)$ in an ellipse. First example



A normal with eigenvalues on an elliptic curve

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

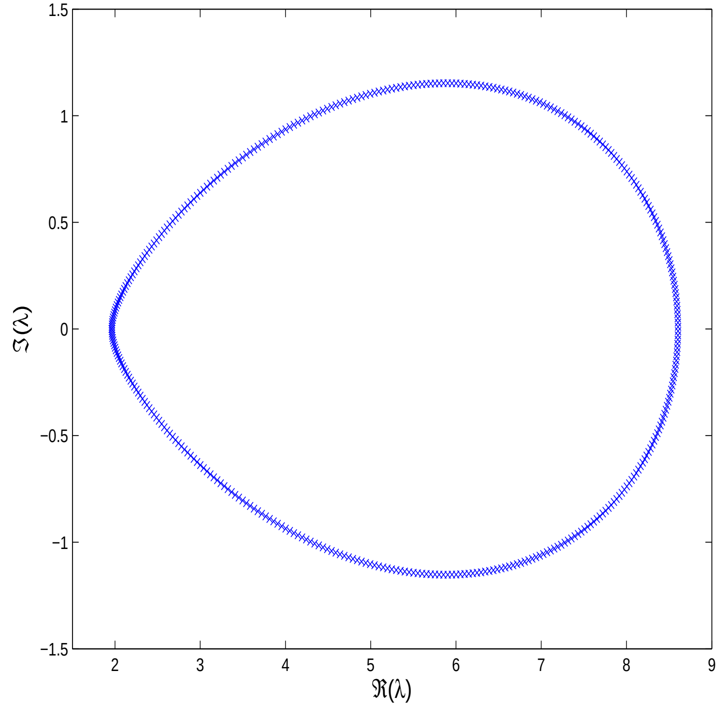


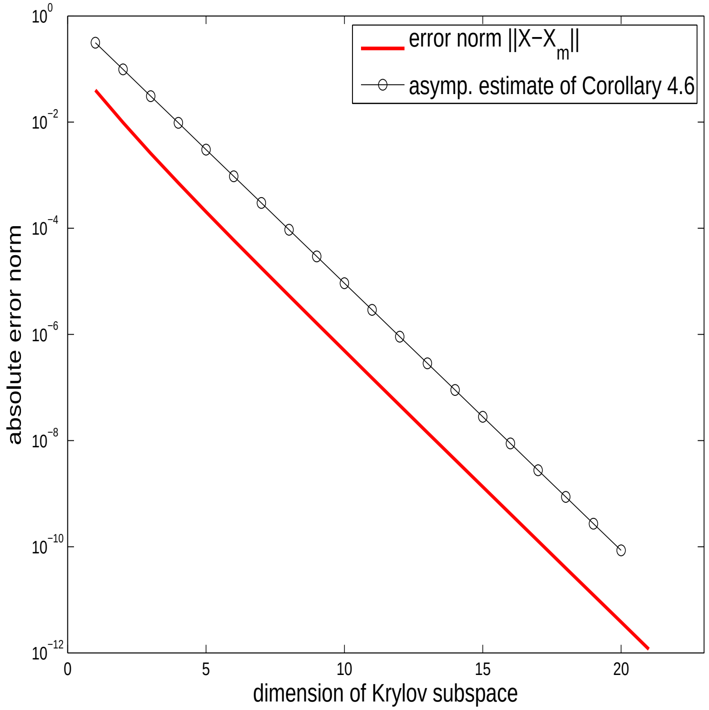


A: from $\mathcal{L}(u) = -\Delta u + 40(x + y)u_x + 200u$ in $\Omega = (0, 1)^2$, homog. Dirichlet b.c.

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

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





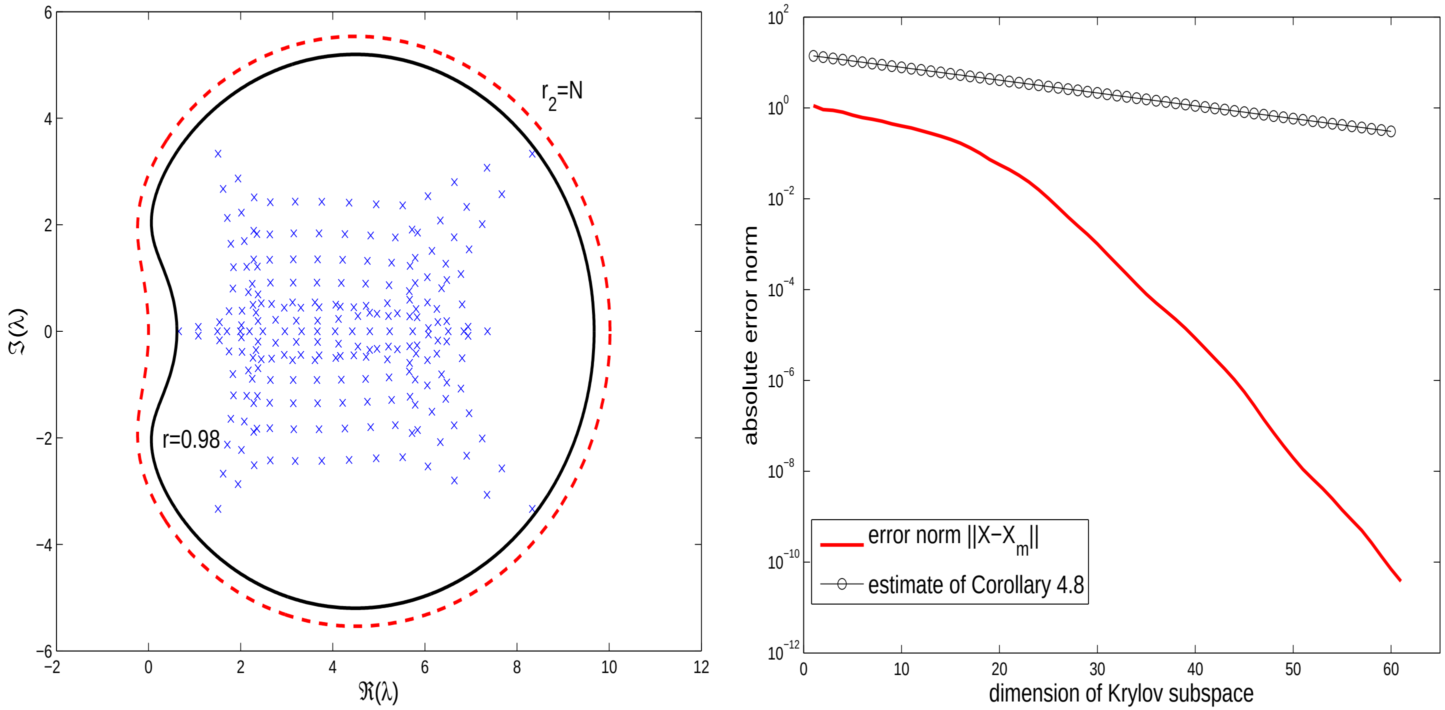
A: diagonal (normal) matrix on the wedge-shaped curve.

(Inclusion set from Hochbruck & Lubich, 1997)

The case of $\Lambda(A)$ in a non-convex, bratwurst set

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

(r, r_2: radii associated with the mapping)



A: matrix PDE225 of the Matrix Market repository

(Inclusion set from Liesen, 1998)

Conclusions and future work

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

- ⋆ Generalize to $\mathcal{K} = \mathcal{K}_m(A, B) \cup \mathcal{K}_m(A^{-1}, B)$
- ⋆ 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`