Approximation of functions of large matrices.
Part I. Computational aspects

V. Simoncini

Dipartimento di Matematica, Università di Bologna

and CIRSA, Ravenna, Italy

valeria@dm.unibo.it

The problem

Given $A \in \mathbb{R}^{n\times n}$, $v \in \mathbb{R}^n$ and f sufficiently smooth function, approximate

$$x = f(A)v$$

$\star$ A large dimension, $\|v\| = 1$

$\star$ A sym. pos. (semi)def., or A positive real

$\star$ $f(A)v$ vs $f(A)$

The problem

Given $A \in \mathbb{R}^{n\times n}$, $v \in \mathbb{R}^n$ and f sufficiently smooth function, approximate

$$x = f(A)v$$

$\star$ A large dimension, $\|v\| = 1$

$\star$ A sym. pos. (semi)def., or A positive real

$\star$ $f(A)v$ vs $f(A)$

Projection-type methods

$\mathcal{K}$ approximation space, $m = \dim(\mathcal{K})$ $V \in \mathbb{R}^{n\times m}$ s.t. $\mathcal{K} = \mathrm{range}(V)$

$$x = f(A)v \quad \approx \quad x_m = Vf(V^\top AV)(V^\top v)$$

Numerical approximation.

$$f(A)v \quad \approx \quad \widetilde{x}$$

Important issues:

- ⋆ Role of f in the approximation quality
- ⋆ Role of A in the approximation quality
- ⋆ Efficiency ?
- ⋆ Measures/Estimates of accuracy?

Standard Krylov subspace approximation

$$\mathcal{K} = K_m(A, v)$$

For $H_m = V^\top AV$, $v = Ve_1$ and $V^\top V = I_m$:

$$x_m = V_m f(H_m) e_1$$

Polynomial approximation: $x_m = p_{m-1}(A)v$

(p_{m-1} interpolates f at eigenvalues of H_m)

(alternatives: other polyns or interpolation points; e.g., Moret & Novati, '01-'04)

Note: Procedure valid for A symmetric and nonsymmetric

⋆ Numerical and theoretical results since mid '80s.

Typical convergence estimates in $\mathcal{K}_m$

Approximation of $f(\lambda) = \exp(-\lambda)$ (Hochbruck & Lubich '97)

A sym. semidef. $\sigma(A) \subseteq [0, 4\rho]$, $\quad \widetilde{x}_m = V_m \exp(-H_m) e_1$,

$$\|f(A)v - \widetilde{x}_m\| \le 10 e^{-m^2/(5\rho)}, \quad \sqrt{4\rho} \le m \le 2\rho$$

$$\|f(A)v - \widetilde{x}_m\| \le \frac{10}{\rho} e^{-\rho} \left(\frac{e\rho}{m}\right)^m, \quad m \ge 2\rho$$

see also Tal-Ezer '89, Druskin & Knizhnerman '89, Stewart & Leyk '96

Typical convergence estimates in $\mathcal{K}_m$

Approximation of $f(\lambda) = \exp(-\lambda)$ (Hochbruck & Lubich '97)

A sym. semidef. $\sigma(A) \subseteq [0, 4\rho]$, $\widetilde{x}_m = V_m \exp(-H_m) e_1$,

$$\|f(A)v - \widetilde{x}_m\| \leq 10 e^{-m^2/(5\rho)}, \quad \sqrt{4\rho} \leq m \leq 2\rho$$

$$\|f(A)v - \widetilde{x}_m\| \leq \frac{10}{\rho} e^{-\rho} \left(\frac{e\rho}{m}\right)^m, \quad m \geq 2\rho$$

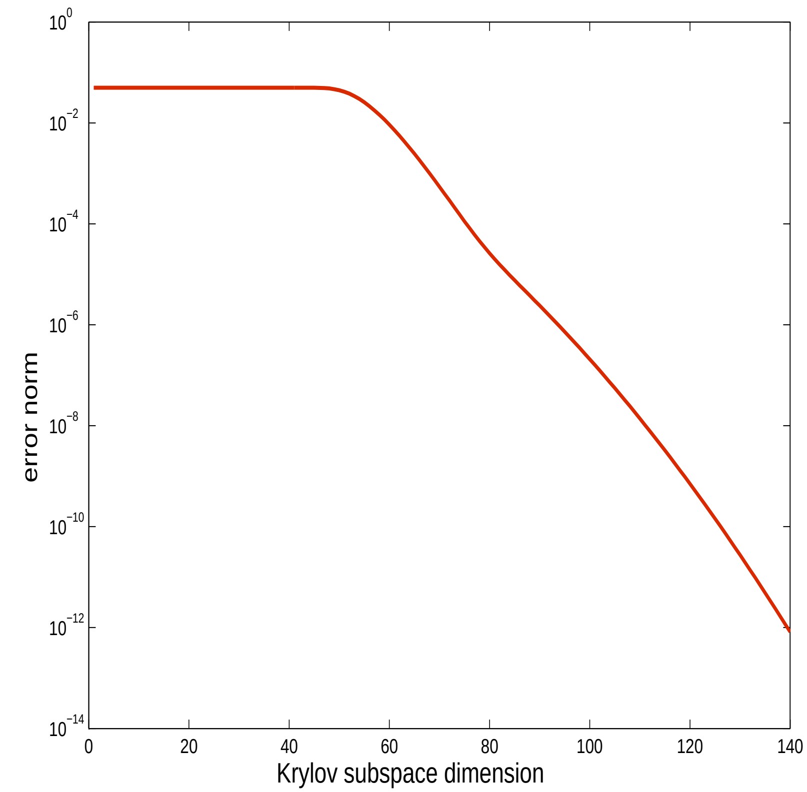
see also Tal-Ezer '89, Druskin & Knizhnerman '89, Stewart & Leyk '96

Approximation of $f(\lambda) = \lambda^{-1/2}, \exp(-\sqrt{\lambda}), \ldots$:

$$\|f(A)v - V_m f(-H_m) e_1\| = \mathcal{O}(\exp(-2m\sqrt{\frac{\lambda_{\min}}{\lambda_{\max}}}))$$

...When things are not so easy

$$\|\exp(-A)v - V_m \exp(-H_m)e_1\| \qquad A \in \mathbb{R}^{400\times 400}, \|A\| = 10^5$$



$$\|\exp(-A)v - V_m \exp(-H_m)e_1\| \leq 10e^{-m^2/(5\rho)}, \qquad \sqrt{4\rho} \leq m \leq 2\rho$$

where $\sigma(A) \subseteq [0, 4\rho]$

Acceleration Procedures: Shift-Invert Lanczos

A symmetric pos. (semi)def.

Choose γ s.t. $(I + \gamma A)$ is invertible, and construct

$\mathcal{K} = K_m((I + \gamma A)^{-1}, v)$, van den Eshof-Hochbruck '06, Moret-Novati '04

with $T_m = V^\top (I + \gamma A)^{-1} V$, $v = V e_1$ and $V^\top V = I_m$

$$x_m = V_m f(\frac{1}{\gamma}(T_m^{-1} - I_m))e_1$$

Rational approximation: $x_m = p_{m-1}((I + \gamma A)^{-1})v$

Choice of γ: $\gamma = 1/\sqrt{\lambda_{\min}\lambda_{\max}}$ (Moret, tr 2005)

Acceleration Procedures: Extended Krylov

For A nonsingular,

$\mathcal{K} = K_{m_1}(A, v) + K_{m_2}(A^{-1}, A^{-1}v)$,   Druskin-Knizhnerman 1998, A sym.

Note: $\mathcal{K} = A^{-(m_2-1)} K_{m_1+m_2-1}(A, v)$

Algorithm (augmentation-style)

- Fix $m_2 \ll m_1$
- Run m_2 steps of Inverted Lanczos
- Run m_1 steps of Standard Lanczos + orth.

Extended Krylov: a new implementation

$m_1 = m_2 = m$ not fixed a priori

$$\mathcal{K} = K_m(A, v) + K_m(A^{-1}, A^{-1}v)$$

$\star$ Arnoldi-type recurrence:

- $U_1 \leftarrow [v, A^{-1}v] + \text{orth}$

- $U_{j+1} \leftarrow [AU_j(:,1), A^{-1}U_j(:,2)] + \text{orth} \qquad j = 1, 2, \ldots$

$\star$ Recurrence to cheaply compute $\mathcal{T}_m = \mathcal{U}_m^\top A \mathcal{U}_m, \quad \mathcal{U}_m = [U_1, \ldots, U_m]$

$\star$ Compute $x_m = \mathcal{U}_m f(\mathcal{T}_m) e_1$

Simoncini, 2007

Extended Krylov: Convergence theory I

f satisfying $\quad f(z) = \int_{-\infty}^{0} \frac{1}{z-\zeta} d\mu(\zeta), \quad z \in \mathbb{C} \backslash]-\infty, 0]$

(with convenient measure $d\mu(\zeta)$)

Druskin-Knizhnerman 1998:

A sym: $\quad \|x - x_m\| = \mathcal{O}(m^2 \exp(-2m \sqrt[4]{\frac{\lambda_{\min}}{\lambda_{\max}}}))$

Extended Krylov: Convergence theory II

A new approximation result for nonsingular A:

Let $f = f_1 + f_2$, $a \in [0, \infty)$

$$\|f_1(z) - \sum_{k=0}^{m-1} \gamma_{1,k} F_{1,k}(z)\| \leq c_1 \Phi_1(-a)^{-m},$$

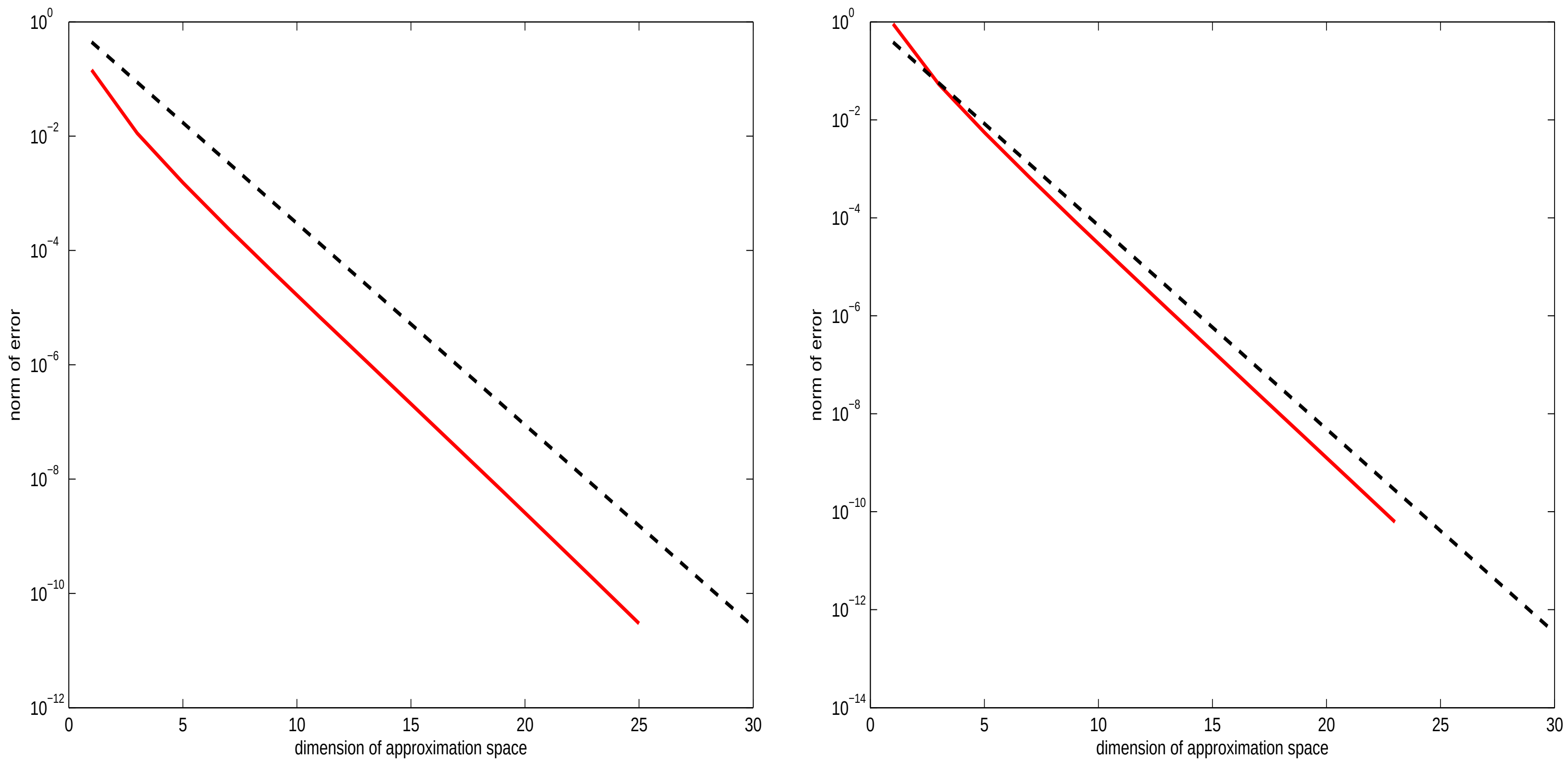
$$\|f_2(z) - \sum_{k=0}^{m-1} \gamma_{2,k} F_{2,k}(z^{-1})\| \leq c_2 \Phi_2(-\frac{1}{a})^{-m},$$

Φ_1, $F_{1,k}$ conformal mapping and Faber Polynomial w.r.to $W(A)$
Φ_2, $F_{2,k}$ conformal mapping and Faber Polynomial w.r.to $W(A^{-1})$

From this, for A symmetric: $\|x - x_m\| = O(\exp(-2m \sqrt[4]{\frac{\lambda_{\min}}{\lambda_{\max}}}))$

⋆ Currently working on A nonsymmetric

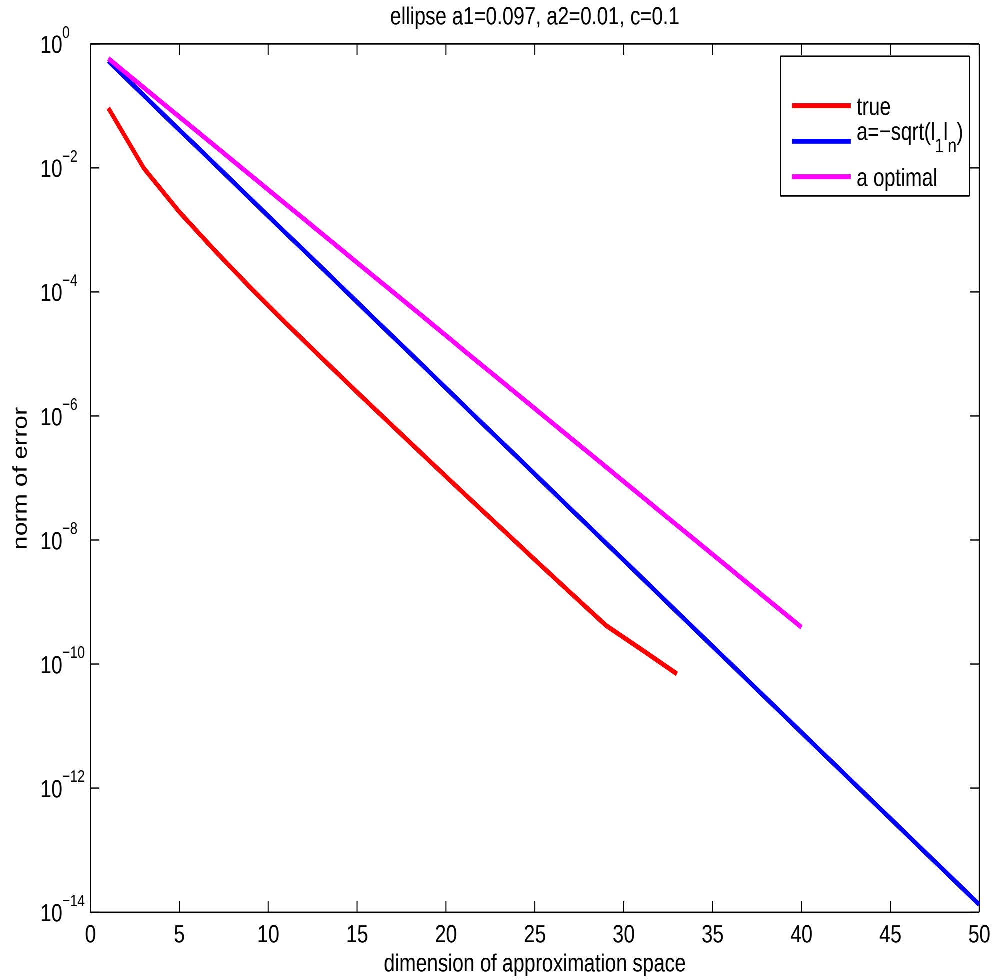
Convergence rate. $A \in \mathbb{R}^{400\times 400}$ symmetric. $f(\lambda) = \lambda^{-1/2}$



$\sigma(A) = [0.01, 0.9]$ $\qquad$ $\sigma(A) = [1, 50]$

Uniform spectral distribution

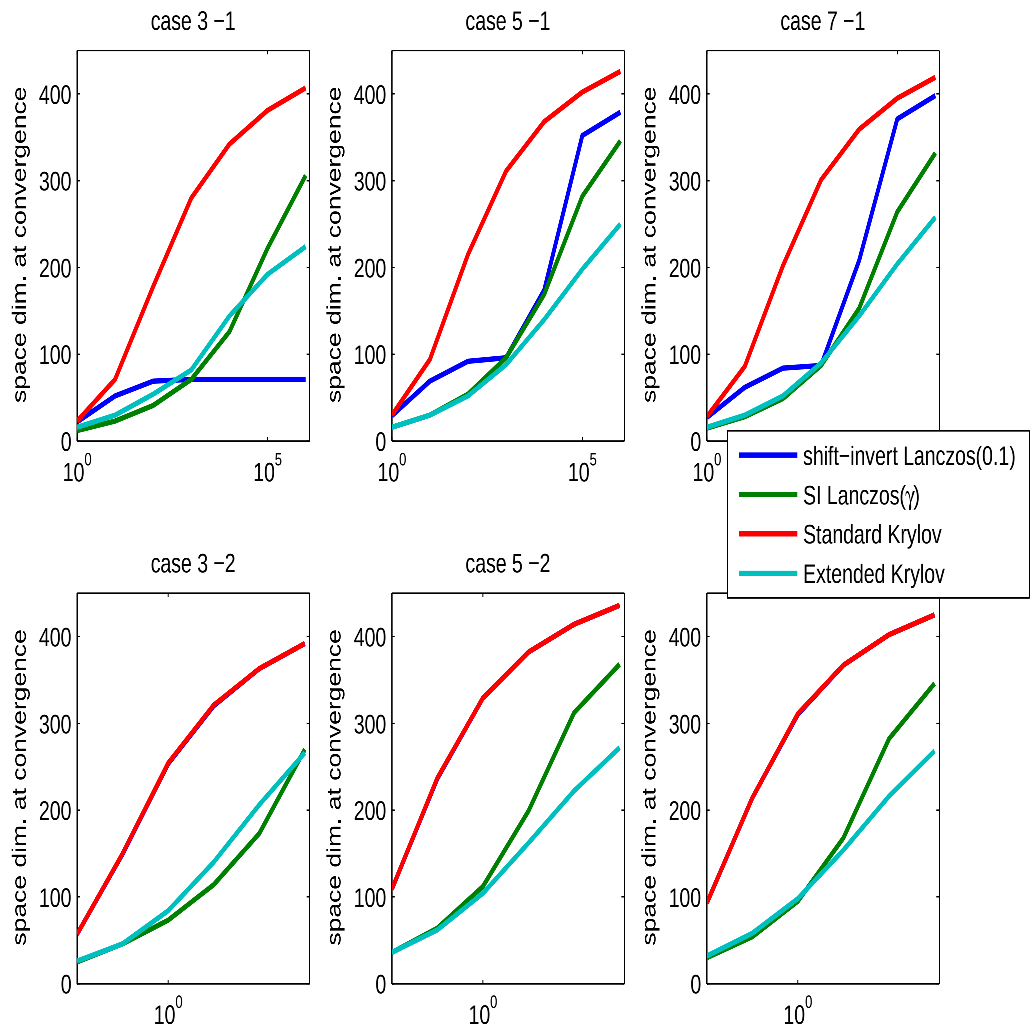
Experiment. $A \in \mathbb{R}^{400\times 400}$ normal. $f(\lambda) = \lambda^{-1/2}$



$\sigma(A)$ on an elliptic curve in $\mathbb{C}^+$ with center on real axis

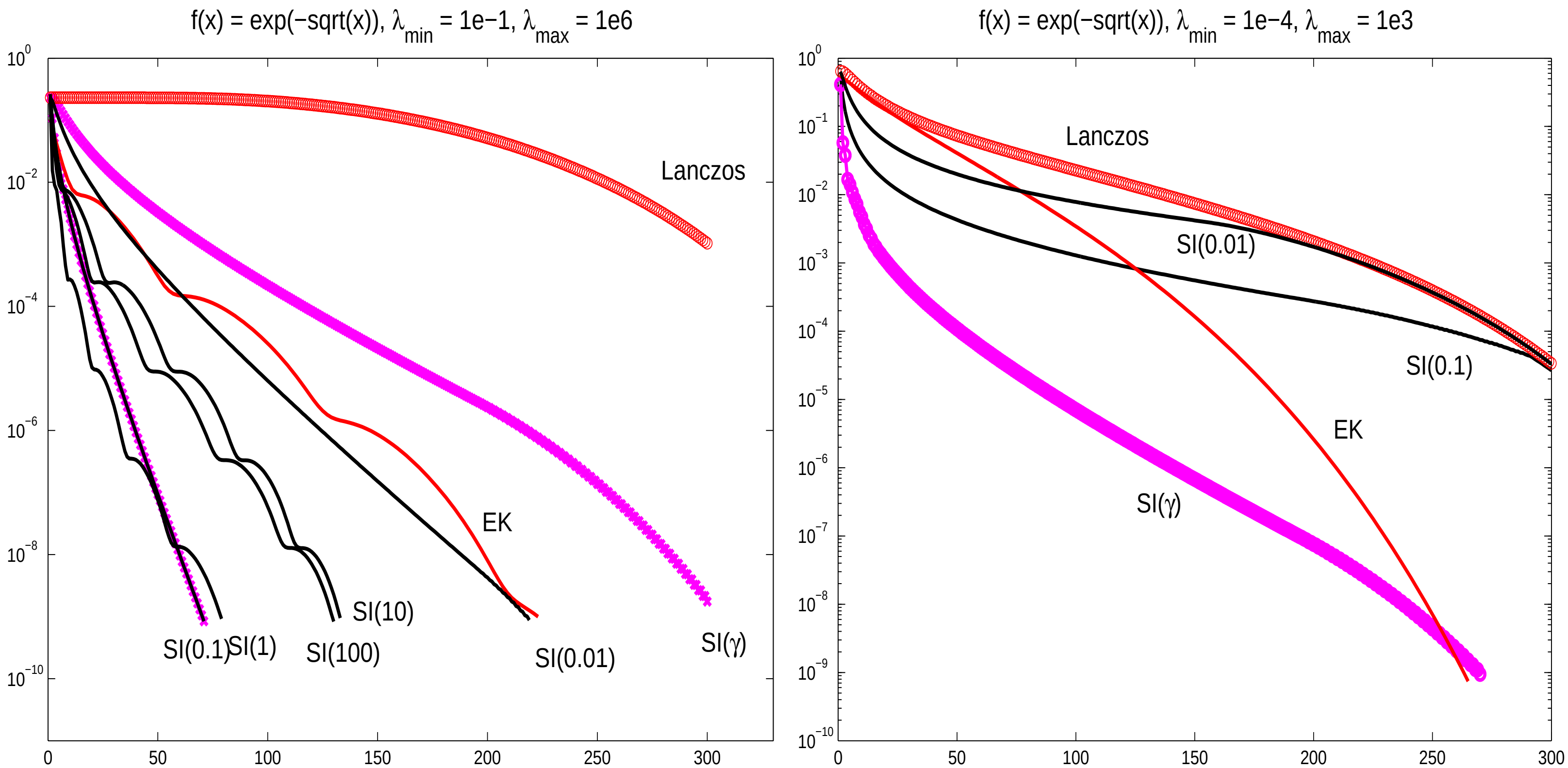
Log-uniform spectral distribution.

$$f_{(3)}(z) = \exp(-\sqrt{z}), \;\; f_{(5)}(z) = z^{-1/2}, \;\; f_{(7)}(z) = z^{-1/4}$$



cases *-1: $\lambda_{\min} = 10^{-1}$ cases *-2: $\lambda_{\min} = 10^{-4}$ x-axis: $\lambda_{\max} = 10^{k}$

SI-Lanczos and dependence on parameter γ



Log-uniform spectral distribution

Comparisons: CPU Time in Matlab

$A \in \mathbb{R}^{4900\times 4900}$: $\mathcal{L}(u) = -\frac{1}{10}u_{xx} - 100u_{yy}$, in $[0,1]^2$, hom.b.c.

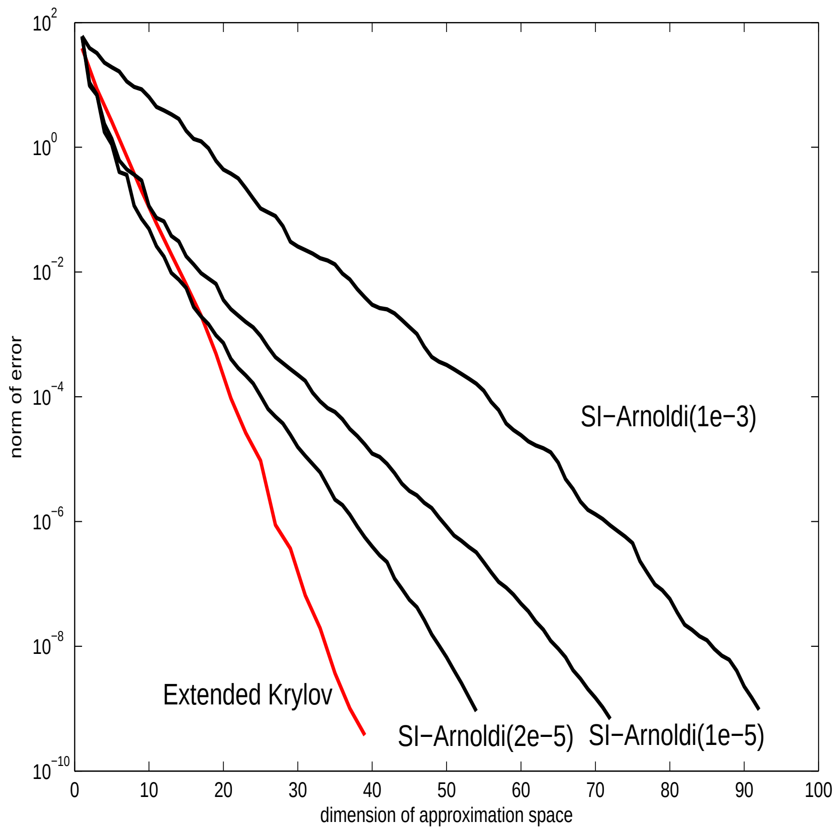
$\sigma(A) \in [9.6 \cdot 10^2,\ 1.96 \cdot 10^6]$, $f(\lambda) = \lambda^{-1/2}$

Method	space dim.	CPU Time
Standard Krylov	185	16.02
Rational (Zolotarev)		0.50
SI-Lanczos(0.001)	62	1.00
SI-Lanczos (1e-5)	49	0.60
SI-Lanczos (γ=2e-5)	**33**	**0.32**
Extended Krylov	32	0.20

No reorthogonalization. Direct solves. Tol $= 10^{-12}$.

A nonsymmetric matrix. $f(z) = \sqrt{z}$

$A \in \mathbb{R}^{900\times 900}$: $\mathcal{L}(u) = -100u_{xx} - u_{yy} + 10xu_x$ in $[0,1]^2$, hom.b.c.



$\sigma(A) \subset \mathbb{R}$, $\quad \lambda_{\min} = 9.2 \cdot 10^2$, $\lambda_{\max} = 3.6 \cdot 10^5$

Full orthogonalization

Conclusions and work in progress

- Great potential of using $f(A)v$ in application problems
- Exploit low cost of using A instead of $f(A)$
- Further developments in acceleration techniques
- The case of A nonsymmetric (preliminary encouraging tests)

Conclusions and work in progress

- Great potential of using $f(A)v$ in application problems
- Exploit low cost of using A instead of $f(A)$
- Further developments in acceleration techniques
- The case of A nonsymmetric (preliminary encouraging tests)

⋆ New implementation of the Extended Krylov method

⋆ Improved convergence bounds for A symmetric

⋆ New convergence results for A nonsymmetric (to be completed)

⋆ Performance:

- Does not depend on parameters
- Competitive for A sym, and nonsym. (preliminary tests)