Approximation of functions of large matrices: computational aspects and applications

V. Simoncini

Dipartimento di Matematica and CIRSA, Bologna

valeria@dm.unibo.it

The Problem

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

$$x = f(A)\,v$$

with f regular function such that $f(A)$ is well defined

Focus:

- A large dimension
- A symmetric pos. (semi)def., or A *positive real*

Context

- A of small dimension:

$$A \text{ symmetric}, \qquad A = X\Lambda X^\top \Rightarrow f(A) = X f(\Lambda) X^\top$$

Similar, but more involved, the definition for A nonsymmetric

- A medium to large dimension:

$$f(A) \qquad \text{vs.} \qquad f(A)v$$

Applications

Among which:

- Numerical solution of evolution PDEs
  (e.g. $\exp(\lambda), \sqrt{\lambda^{-1}}, \cos(\lambda), \varphi_k(\lambda) ...$)
- Inverse Problems $(\exp(\lambda), \cosh(\lambda), ...)$.see Talk by L. Eldén
- Fluxes on manifolds
- Problems in Scientific Computing (e.g. QCD, $\mathrm{sign}(\lambda)$)
- (Analysis of) reduced Dynamical System Models
  (through Grammian Matrices)

$\Rightarrow$ Some examples later on.

Applications

Among which:

- Numerical solution of evolution PDEs (e.g. $\exp(\lambda), \sqrt{\lambda^{-1}}, \cos(\lambda), \varphi_k(\lambda)$...)
- Inverse Problems $(\exp(\lambda), \cosh(\lambda)$, ...).see Talk by L. Eldén
- Fluxes on manifolds
- Problems in Scientific Computing (e.g. QCD, $\mathrm{sign}(\lambda)$)
- (Analysis of) reduced Dynamical System Models (through Grammian Matrices)

$\Rightarrow$ Some examples later on. The idea:

$$\begin{cases} y' = -Ay \\ y(0) = y_0 \end{cases} \quad \Rightarrow \quad y(t) = \exp(-tA)y_0$$

Numerical approximation. I

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

Various alternatives. Among which:

- Substitute f with "simpler" function, $\quad f \approx \mathcal{R}$
  e.g., $\mathcal{R}$ rational function:

$$\|f(A)v - \widetilde{x}\| \le \|f(A)v - \mathcal{R}(A)v\| + \|\mathcal{R}(A)v - \widetilde{x}\|$$

and either $\quad \Rightarrow \quad \widetilde{x} = \mathcal{R}(A)v \quad$ or $\quad \Rightarrow \quad \widetilde{x} \approx \mathcal{R}(A)v$

Numerical approximation. I

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

Various alternatives. Among which:

- Substitute f with "simpler" function, $\quad f \approx \mathcal{R}$
  e.g., $\mathcal{R}$ rational function:

$$\|f(A)v - \widetilde{x}\| \le \|f(A)v - \mathcal{R}(A)v\| + \|\mathcal{R}(A)v - \widetilde{x}\|$$

and either $\quad \Rightarrow \quad \widetilde{x} = \mathcal{R}(A)v \quad$ or $\quad \Rightarrow \quad \widetilde{x} \approx \mathcal{R}(A)v$

- Approximation by projection: Find V and

$$\widetilde{x} \in \mathrm{range}(V), \quad \dim(\mathrm{range}(V)) \ll n$$

Numerical approximation. II

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

Important issues:

- $\star$ Role of f in the approximation quality
- $\star$ Role of A in the approximation quality
- $\star$ Efficiency ?
- $\star$ Measures/Estimates of accuracy? (see Talk by O.Ernst)

First alternative: Rational Approximation

$$x = f(A)\,v \quad \approx \quad \mathcal{R}_{\mu,\nu}(A)v$$

$$\mathcal{R}_{\mu,\nu}(\lambda) = \frac{\Phi_\mu(\lambda)}{\Psi_\nu(\lambda)}, \qquad \Phi_\mu(\lambda),\ \Psi_\nu(\lambda) \quad \text{polynomials}$$

- Polynomial Approx., $\nu = 0$

  (Druskin & Knizhnerman, '89, Bergamaschi & Vianello, '00)

- Rational Approx.: Padé or Chebyshev, e.g. $\mu = \nu$
- Rational Approx w/multiple pole (RD) (Novati & Moret, late 90s)
- Quadrature Methods (see, e.g., Hale, Higham, Trefethen '08)

First alternative: Rational Approximation

$$x = f(A)\, v \quad \approx \quad \mathcal{R}_{\mu,\nu}(A)v$$

$$\mathcal{R}_{\mu,\nu}(\lambda) = \frac{\Phi_\mu(\lambda)}{\Psi_\nu(\lambda)}, \qquad \Phi_\mu(\lambda),\ \Psi_\nu(\lambda) \quad \text{polynomials}$$

- Polynomial Approx., $\nu = 0$

  (Druskin & Knizhnerman, '89, Bergamaschi & Vianello, '00)

- Rational Approx.: Padé or Chebyshev, e.g. $\mu = \nu$
- Rational Approx w/multiple pole (RD) (Novati & Moret, late 90s)
- Quadrature Methods (see, e.g., Hale, Higham, Trefethen '08)

We consider the case of partial fraction expansion:

$$\mathcal{R}_{\mu,\nu}(\lambda) = q(\lambda) + \sum_{k=1}^{\nu} \frac{\omega_k}{\lambda - \xi_k} \qquad\qquad (\mathcal{R}_\nu = \mathcal{R}_{\nu,\nu})$$

Rational Approximation: poles

$f(\lambda) = \exp(-\lambda)$

$\mathcal{R}_\nu$: ℓ_∞ best approx

in $[0, \infty)$, Chebyshev

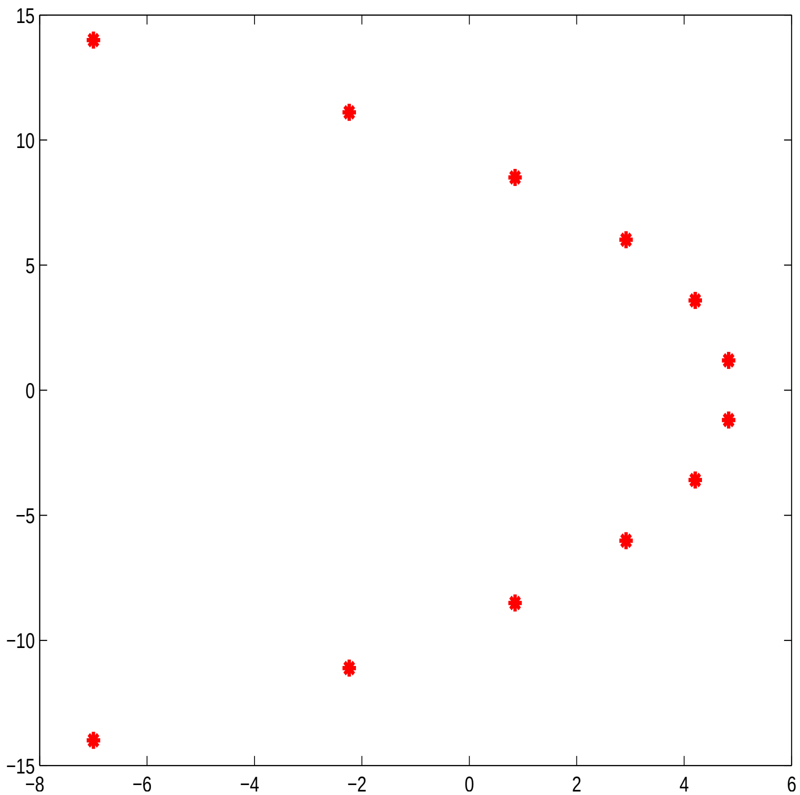
$\|f - \mathcal{R}_\nu\|_\infty \approx 10^{-\nu}$

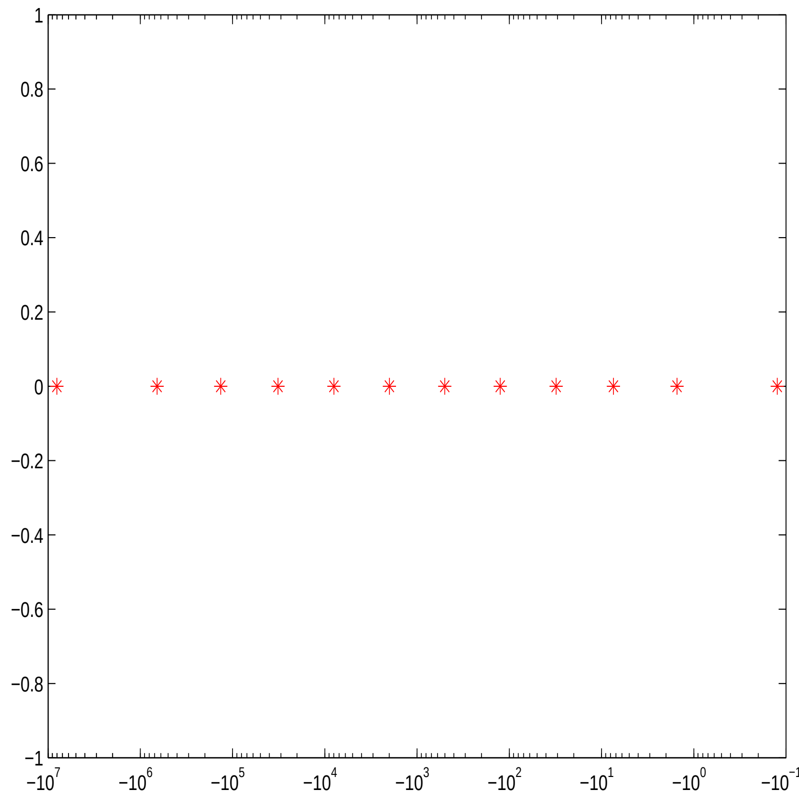
$f(\lambda) = \lambda^{-1/2}$

$\mathcal{R}_\nu$: Zolotarev approx

in $[a, b] \subseteq (0, \infty)$

$\|f - \mathcal{R}_\nu\| \approx e^{-\pi\sqrt{2\nu}}$





Matrix Rational approximation

$$f(A)v \quad \approx \quad \mathcal{R}_\nu(A)v \quad = \quad \sum_{k=1}^{\nu} \omega_k (A - \xi_k I)^{-1} v$$

- $\forall k$, $(A - \xi_k I)$ "Shifted" matrix , $\xi_k \in \mathbb{C}$
- $\xi_{2j-1} = \bar{\xi}_{2j}$, $j = 1, \ldots, \lfloor \nu/2 \rfloor$
- $\forall k$, $x_k = (A - \xi_k I)^{-1} v$ or $\widetilde{x}_k \approx (A - \xi_k I)^{-1} v$

Matrix Rational approximation

$$f(A)v \quad \approx \quad \mathcal{R}_\nu(A)v \quad = \quad \sum_{k=1}^{\nu} \omega_k (A - \xi_k I)^{-1} v$$

- $\forall k$, $(A - \xi_k I)$ "Shifted" matrix , $\xi_k \in \mathbb{C}$
- $\xi_{2j-1} = \bar{\xi}_{2j}$, $j = 1, \ldots, \lfloor \nu/2 \rfloor$
- $\forall k$, $x_k = (A - \xi_k I)^{-1} v$ or $\widetilde{x}_k \approx (A - \xi_k I)^{-1} v$

$\Rightarrow$ Iterative Methods for **shifted** linear systems

(Baldwin & Freund & Gallopoulos '95, Popolizio & S. '08)

$\Rightarrow$ Cheap error estimates (Frommer & S., '08)

Approximation with Krylov subspaces

$$\mathcal{K}_m \equiv \mathcal{K}_m(A, v) = \mathrm{span}\{v, Av, \ldots, A^{m-1}v\}$$

V_m s.t. range(V_m) $= \mathcal{K}_m(A, v)$ and $V_m^\top V_m = I$

$\Rightarrow$ Motivation: $\exists\ p$ polynomial (interpolatory): $f(A) = p(A)$

Approximation with Krylov subspaces

$$\mathcal{K}_m \equiv \mathcal{K}_m(A, v) = \text{span}\{v, Av, \ldots, A^{m-1}v\}$$

V_m s.t. range(V_m) $= \mathcal{K}_m(A, v)$ and $V_m^\top V_m = I$

$\Rightarrow$ Motivation: $\exists\, p$ polynomial (interpolatory): $f(A) = p(A)$

"Classical" approach: (e.g., Gallopoulos & Saad '92, Saad '92)

For $H_m = V_m^\top A V_m$, $v = V_m e_1$

$$f(A)v \;\approx\; x_m = V_m f(H_m) e_1 \qquad \|v\| = 1$$

$\star$ x_m from interpolation pb. in Hermite sense: $V_m f(H_m) e_1 = p_{m-1}(A) v$

Approximation with Krylov subspaces

$$\mathcal{K}_m \equiv \mathcal{K}_m(A,v) = \mathrm{span}\{v, Av, \ldots, A^{m-1}v\}$$

V_m s.t. range$(V_m) = \mathcal{K}_m(A,v)$ and $V_m^\top V_m = I$

$\Rightarrow$ Motivation: $\exists\ p$ polynomial (interpolatory): $f(A) = p(A)$

"Classical" approach: (e.g., Gallopoulos & Saad '92, Saad '92)

For $H_m = V_m^\top A V_m$, $v = V_m e_1$

$$f(A)v \approx x_m = V_m f(H_m) e_1 \qquad \|v\| = 1$$

$\star$ x_m from interpolation pb. in Hermite sense: $V_m f(H_m) e_1 = p_{m-1}(A)v$

For $f = \mathcal{R} \Rightarrow$ Standard Krylov $=$ Rational approx with inexact solves

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

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

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

$$\begin{aligned} \|f(A)v - \widetilde{x}_m\| &\leq 10 e^{-m^2/(5\rho)}, \quad \sqrt{4\rho} \leq m \leq 2\rho \\ \|f(A)v - \widetilde{x}_m\| &\leq \frac{10}{\rho} e^{-\rho} \left(\frac{e\rho}{m}\right)^m, \qquad m \geq 2\rho \end{aligned}$$

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

$$\begin{aligned} \|f(A)v - \widetilde{x}_m\| &\le 10 e^{-m^2/(5\rho)}, \quad \sqrt{4\rho} \le m \le 2\rho \\ \|f(A)v - \widetilde{x}_m\| &\le \frac{10}{\rho} e^{-\rho} \left(\frac{e\rho}{m}\right)^m, \quad m \ge 2\rho \end{aligned}$$

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

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

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

Application. Evolution Problem

$$\begin{cases} \frac{\partial u(x,y,t)}{\partial t} = \Delta u, \quad (x,y) \in (0,1)^2 \quad t \in [0, 0.1] \\ u(x,y,t) = 0, \quad (x,y) \in \partial([0,1]^2) \\ u(x,y,0) = 1, \quad (x,y) \in [0,1]^2 \end{cases}$$

Implicit Euler: $u_{i+1} = (I + \delta t A)^{-1} u_i, \quad i = 0, 1, \ldots$

Exponential Integrator: $u(t) = \exp(-tA) u_0 \quad t = 0.1$

Application. Evolution Problem

$$\begin{cases} \frac{\partial u(x,y,t)}{\partial t} = \Delta u, & (x,y) \in (0,1)^2 \quad t \in [0, 0.1] \\ u(x,y,t) = 0, & (x,y) \in \partial([0,1]^2) \\ u(x,y,0) = 1, & (x,y) \in [0,1]^2 \end{cases}$$

Implicit Euler: $u_{i+1} = (I + \delta t A)^{-1} u_i, \quad i = 0, 1, ...$

Exponential Integrator: $u(t) = \exp(-tA) u_0 \quad t = 0.1$

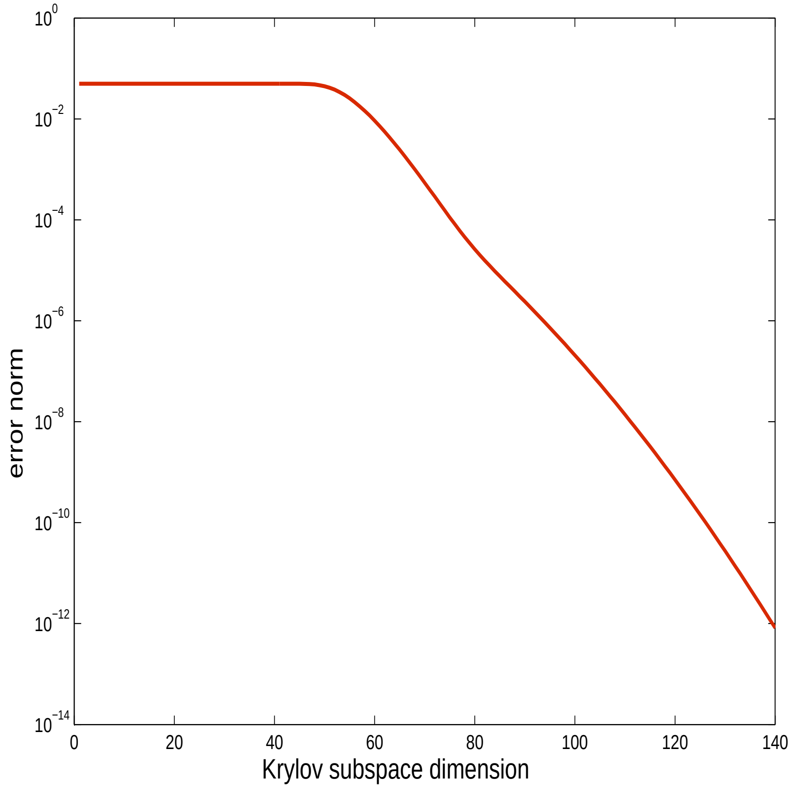
	Euler		Exp	
step δt	CPU	error	CPU	error (#its*)
0.001	1.9	$2 \cdot 10^{-3}$	0.09	$9 \cdot 10^{-4}$(37)
0.005	0.4	$1 \cdot 10^{-2}$	0.07	$4 \cdot 10^{-3}$(28)
0.01	0.2	$2 \cdot 10^{-2}$	0.05	$1 \cdot 10^{-2}$(25)

* : Stopping criterion tolerance related to timestep

$\Rightarrow$ More general exponential integrators (Hochbruck, Lubich, etc.)

...When things are not so easy

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



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

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

Acceleration Techniques

$\star$: Improving approximation space

- Spectral approximation : $\mathcal{K}_m((I+\gamma A)^{-1}, v)$, $\gamma > 0$

$$f(A)v \quad \approx \quad V_m f(\frac{1}{\gamma}(H_m^{-1} - I))e_1$$

(Moret & Novati '04, van den Eshof & Hochbruck, '06, Popolizio & S. '08)

Acceleration Techniques

$\star$: Improving approximation space

- Spectral approximation : $\mathcal{K}_m((I+\gamma A)^{-1}, v)$, $\gamma > 0$

$$f(A)v \quad \approx \quad V_m f(\frac{1}{\gamma}(H_m^{-1} - I))e_1$$

(Moret & Novati '04, van den Eshof & Hochbruck, '06, Popolizio & S. '08)

- "Extended" space: $\mathcal{K}_m(A^{-1}, A^{-1}v) \cup \mathcal{K}_m(A, v)$

$$f(A)v \quad \approx \quad \mathcal{V}_m f(\mathcal{T}_m)e_1, \qquad \mathcal{T}_m = \mathcal{V}_m^\top A \mathcal{V}_m$$

(Druskin & Knizhnerman,'98, S.,'07, Knizhnerman & S., in progress)

Acceleration Techniques

⋆: Improving approximation space

- Spectral approximation : $\mathcal{K}_m((I+\gamma A)^{-1}, v)$, $\gamma > 0$

$$f(A)v \quad \approx \quad V_m f(\frac{1}{\gamma}(H_m^{-1} - I))e_1$$

(Moret & Novati '04, van den Eshof & Hochbruck, '06, Popolizio & S. '08)

- "Extended" space: $\mathcal{K}_m(A^{-1}, A^{-1}v) \cup \mathcal{K}_m(A, v)$

$$f(A)v \quad \approx \quad \mathcal{V}_m f(\mathcal{T}_m)e_1, \qquad \mathcal{T}_m = \mathcal{V}_m^\top A \mathcal{V}_m$$

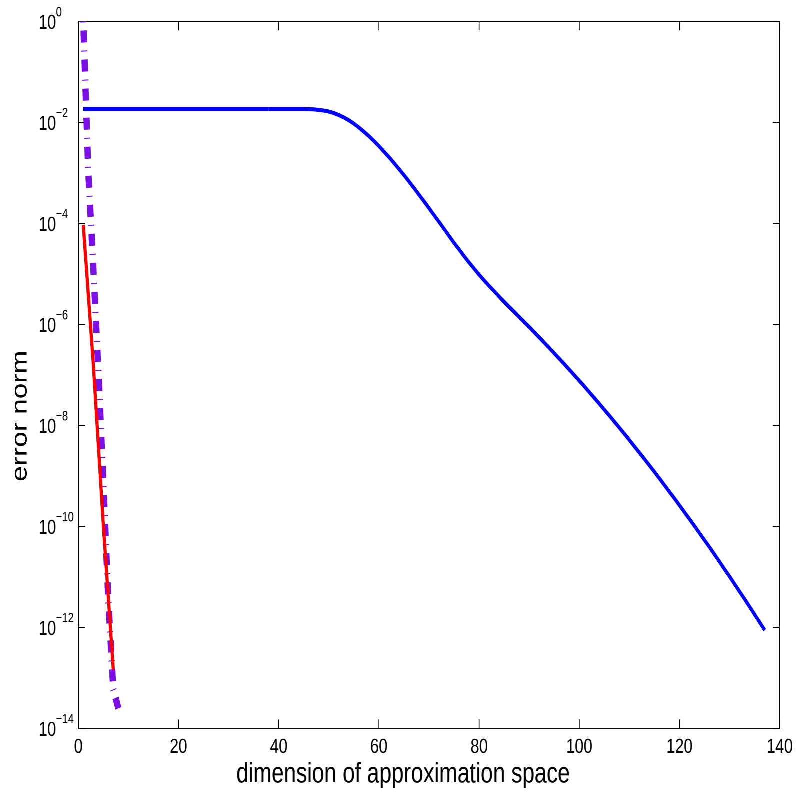
(Druskin & Knizhnerman,'98, S.,'07, Knizhnerman & S., in progress)

⋆: Relaxing optimality properties

- *Local* orthogonality of the basis (Eiermann & Ernst '06)
- Limit costs of rational approx. with $\mathcal{R}_\nu(A)v$ (Popolizio & S. '08)

Acceleration

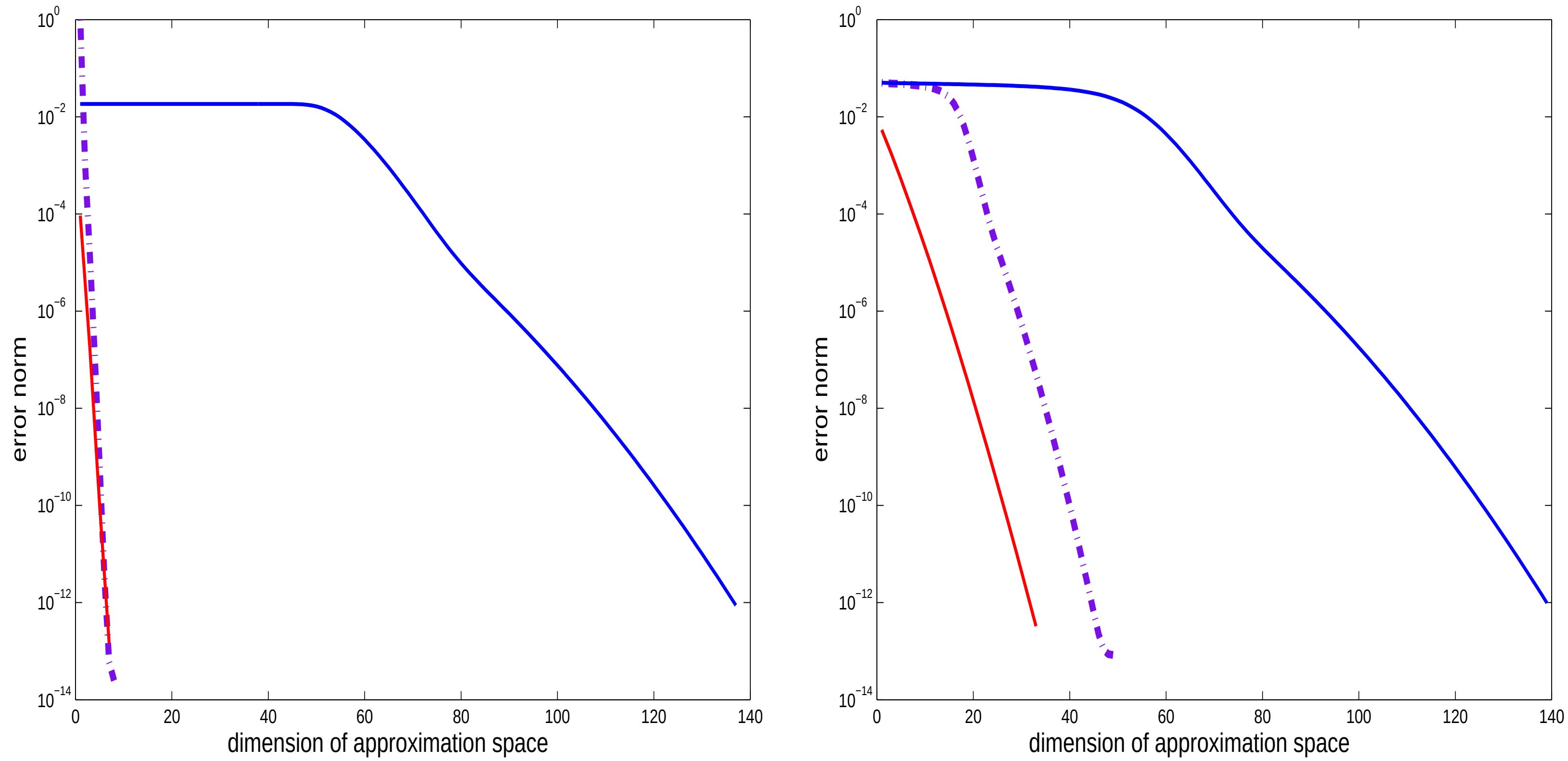
$f(\lambda) = \exp(-\lambda)$



-: std Krylov   -.: Spectral accel.   -: "extended" space

Acceleration

$f(\lambda) = \exp(-\lambda)$ $\qquad$ $f(\lambda) = \lambda^{-1/2}$



-: std Krylov $\qquad$ -.: Spectral accel. $\qquad$ -: "extended" space

Comparisons: CPU Time in Matlab

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

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

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. Exact solves.

Conclusions

- 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)