Rational function approximation to the matrix exponential operator

V. Simoncini

Dipartimento di Matematica, Università di Bologna
and CIRSA, Ravenna, Italy
valeria@dm.unibo.it

Joint works with
L. Lopez (Bari), M. Popolizio (Bari), A. Frommer (Wuppertal)

Approximation problem

Given $v \in \mathbb{R}^n$ and A symmetric and negative semidefinite, approximate

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

- Focus: A large dimension
- General approach: $x_m \in \mathcal{K}_m$ Krylov subspace

Problem in context

Wide range of applications, e.g.

- Numerical solution of Time-dependent PDEs
- (Analysis of) Low dimensional models of dynamical systems: approximate solution to Lyapunov equation

$$AX + XA^T + BB^T = 0$$

- Flows on manifolds

$$Q_t = H(Q,t)Q, \quad Q(t)|_{t=0} = Q_0 \in V_k(\mathbb{R}^n)$$

V_k Stiefel manifold (computation of a few Lyapunov exponents)

Numerical approximation

A large dimension: $\quad x = \exp(A)\, v \quad \approx \quad \mathcal{R}_{\mu,\nu}(A)v$

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

- Polynomial approximation, $\nu = 0$
- Padé (rational function) approximation, e.g., $\mu = \nu$
- Chebyshev (rational function) approximation, $\mu = \nu$
- Restricted Denominator (RD, rational function) approximation
- ...

Approximation using Krylov subspace

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

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

Arnoldi relation

$$AV_m = V_m H_m + h_{m+1,m} v_{m+1} e_m^T$$

A common approach

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

x_m derived from interpolation problem in Hermite sense (Saad '92)

Outline

The tools:

- Krylov space: $\exp(A)v \approx x_m = V_m \exp(H_m) e_1$
- Rational approximation: $\exp(A)\, v \quad \approx \quad \mathcal{R}_{\mu,\nu}(A) v$

The exploration venues:

- Convergence theory
- Stopping criteria
- Acceleration procedures

Approximation of $\exp(A)v$ in Krylov subspace. I

Typical convergence bounds (Hochbruck & Lubich '97)

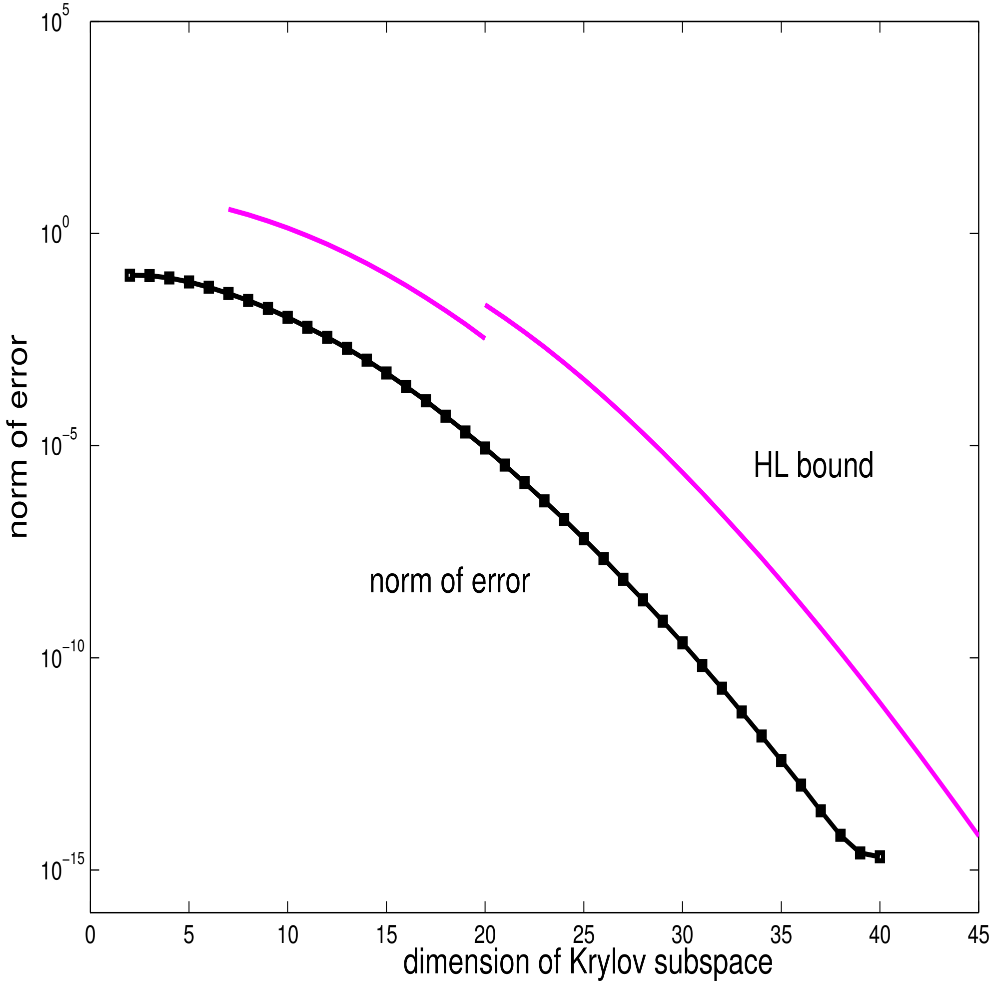
$$\|\exp(A)v - V_m \exp(H_m)e_1\| \quad \le \quad 10e^{-m^2/(5\rho)}, \qquad \sqrt{4\rho} \le m \le 2\rho,$$

$$\|\exp(A)v - V_m \exp(H_m)e_1\| \quad \le \quad \frac{10}{\rho} e^{-\rho} \left(\frac{e\rho}{m}\right)^m, \qquad m \ge 2\rho$$

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

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

A typical picture



Predicts superlinear convergence

Approximation of $\exp(A)v$ in Krylov subspace. II

Typical a-posteriori estimate (see, e.g., Saad '92)

$$\| \exp(A)v - V_m \exp(H_m)e_1\| \quad \approx \quad O(h_{m+1,m}|e_m^T \exp(H_m)e_1|)$$

for m large enough

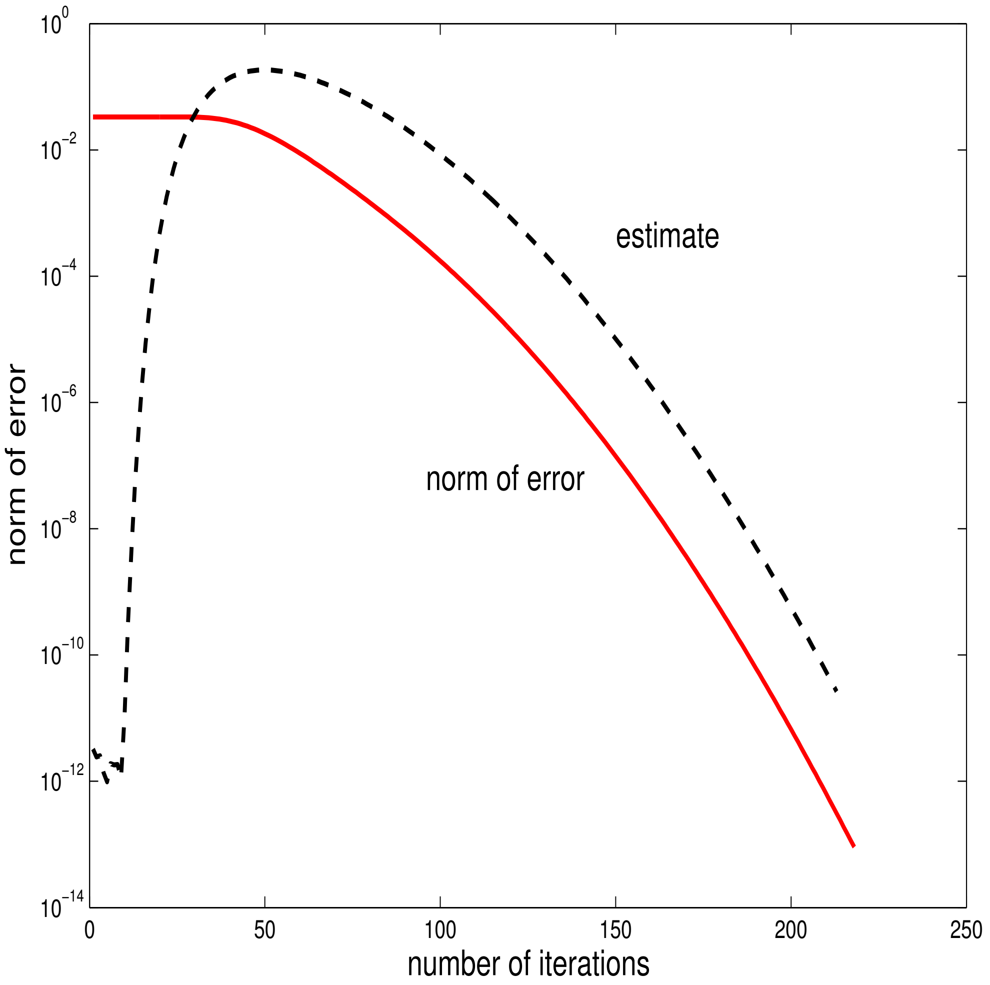
Note: for $Ax(t) - x'(t) = 0$, $x(0) = v$

$$h_{m+1,m}|e_m^* \exp(tH_m)e_1| = \|Ax_m(t) - x'_m(t)\|$$

plays role of residual norm

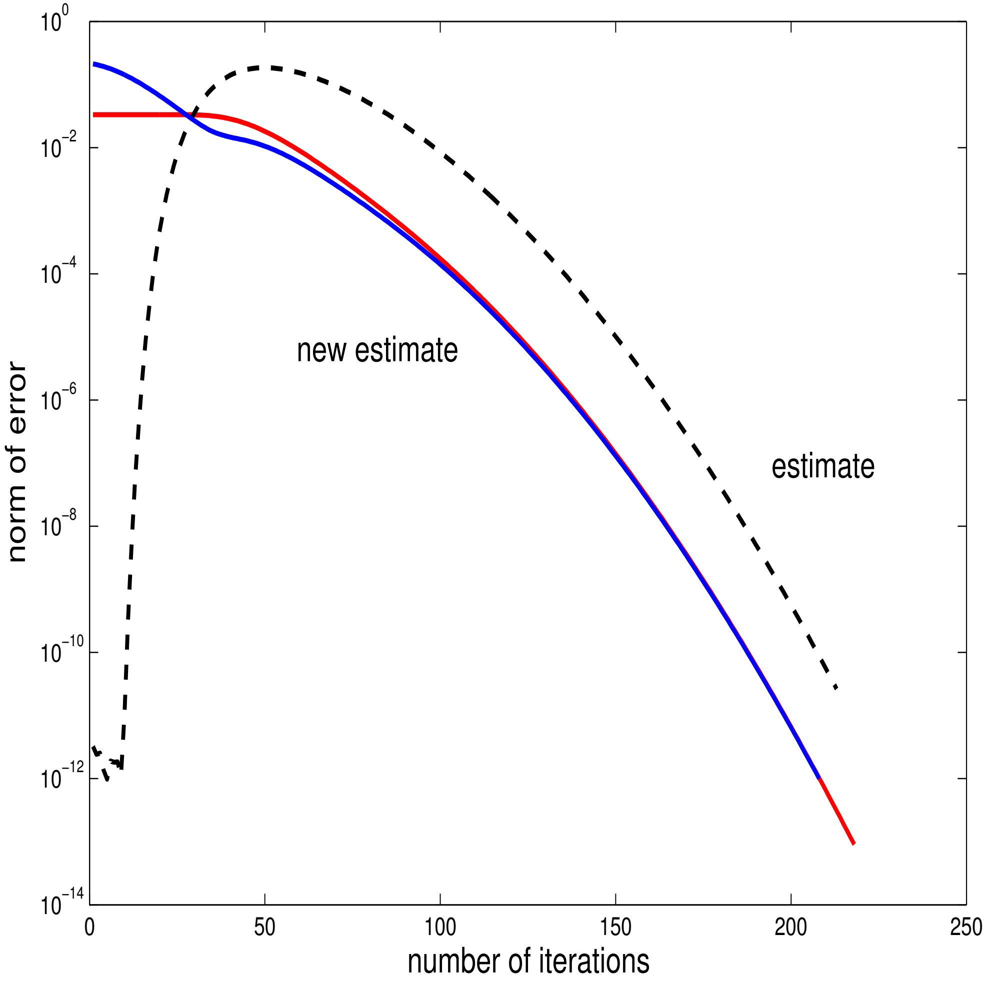
(see, e.g., Druskin & Greenbaum & Knizhnerman '98)

A typical picture



Possible high underestimation for small m

A new bound



Frommer & Simoncini, work in progress.

Projection of Rational functions onto Krylov subspaces

Basic fact: If $x_m \in \mathcal{K}_m(A, v), \quad x_m \approx \mathcal{R}_\nu(A)v$ then

$$\|\exp(A)v - x_m\| \le \|\exp(A)v - \mathcal{R}_\nu(A)v\| + \|\mathcal{R}_\nu(A)v - x_m\|$$

Focus: $\mathcal{R}_\nu = \Phi_\nu / \Psi_\nu$ Padé and Chebyshev approximation

($\Psi_\nu(A)$ positive definite)

Projection onto Krylov subspace

$$x_\star = \mathcal{R}_\nu(A)v = \Psi_\nu(A)^{-1}\Phi_\nu(A)v \quad \Leftrightarrow \quad x_\star \text{ solves } \quad \Psi_\nu(A)x = \Phi_\nu(A)v$$

Galerkin approximation in $\mathcal{K}_m(A, v)$:

$$\text{Solve} \quad V_m^*\Psi_\nu(A)V_m y = V_m^*\Phi_\nu(A)v, \qquad x_m^G = V_m y_m^G$$

Minimization property:

$$\min_{x \in K_m(A,v)} \|x_\star - x\|_{\Psi_\nu(A)} = \|x_\star - x_m^G\|_{\Psi_\nu(A)}$$

But: too expensive

Krylov approximation

$$\mathcal{R}_\nu(A)v \qquad \approx \qquad V_m \mathcal{R}_\nu(H_m)e_1$$

Partial Fraction expansion: $\frac{\Phi_\nu(\lambda)}{\Psi_\nu(\lambda)} = \tau_0 + \sum_{j=1}^{\nu} \frac{\tau_j}{\lambda - \xi_j}$

$$\begin{aligned}
\mathcal{R}_\nu(A)v &= \tau_0 v + \sum_{j=1}^{\nu} \tau_j (A - \xi_j I)^{-1} v \\
&\approx \tau_0 v + \sum_{j=1}^{\nu} \tau_j V_m (H_m - \xi_j I)^{-1} e_1 \\
&= V_m \Psi_\nu(H_m)^{-1} \Phi_\nu(H_m) e_1 \equiv V_m y_m^K
\end{aligned}$$

$V_m y_m^K$ is a term-wise Galerkin projection: (van der Vorst, '87)

Linear bounds for convergence rate

$$x_m^K = V_m \mathcal{R}_\nu(H_m) e_1 \qquad \approx \qquad \mathcal{R}_\nu(A) v = \tau_0 v + \sum_{j=1}^{\nu} \tau_j (A - \xi_j I)^{-1} v$$

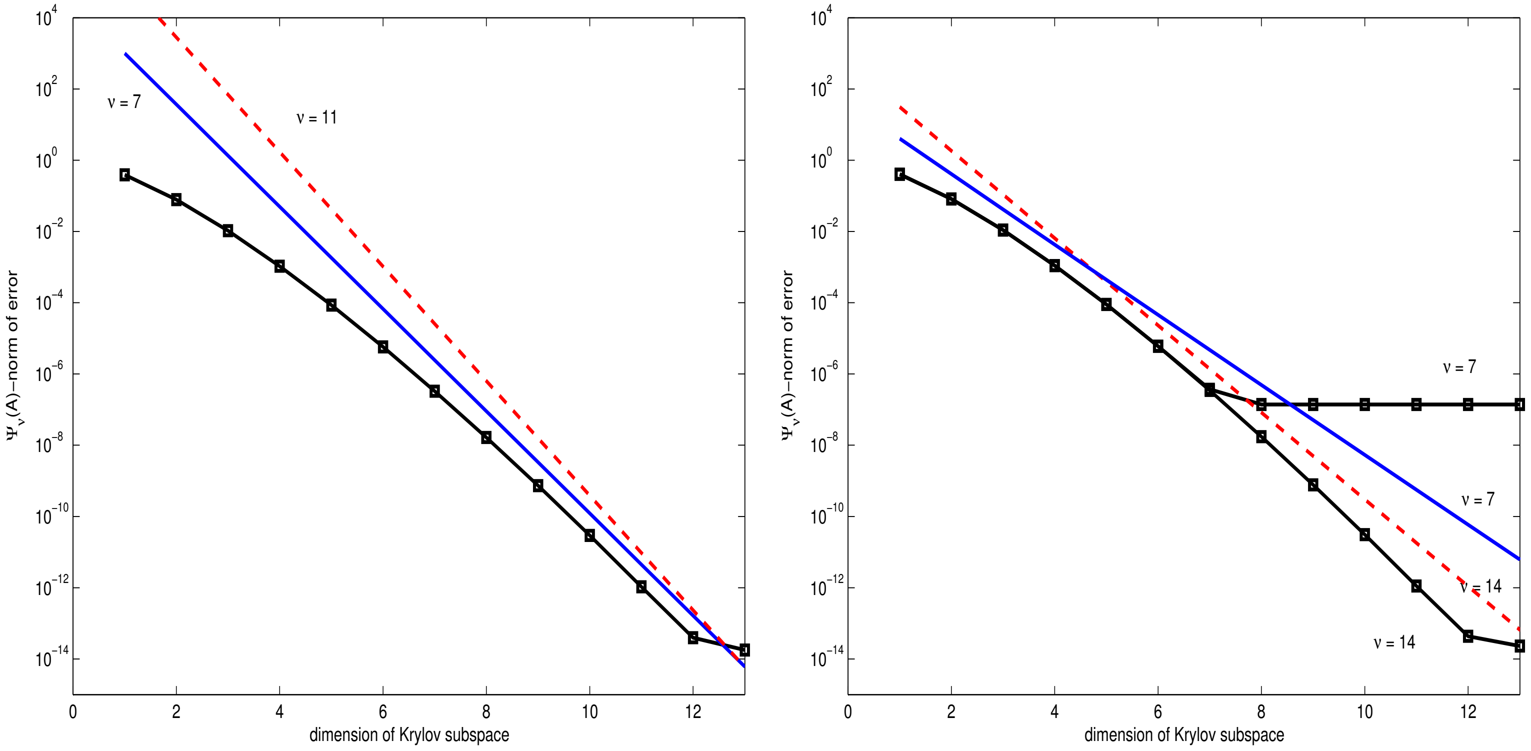
Then:

$$\|\mathcal{R}_\nu(A)v - x_m^K\| \le \sum_{j=1}^{\nu} \eta_j \frac{1}{\rho_j^m + 1/\rho_j^m}$$

$\rho_j = \rho_j(\sigma(A), \xi_j) \qquad \eta_j = \eta_j(\sigma(A), \xi_j)$

Lopez & Simoncini, SINUM '06

Krylov approximation



`A = diag(log(linspace(0.2,0.99,100)))`, $v = \mathbf{1}$

Left: Padé and upper bound for $\nu = 7, 11$ $\qquad$ $\nu = \#$ poles

Right: Chebyshev and upper bounds for $\nu = 7, 14$

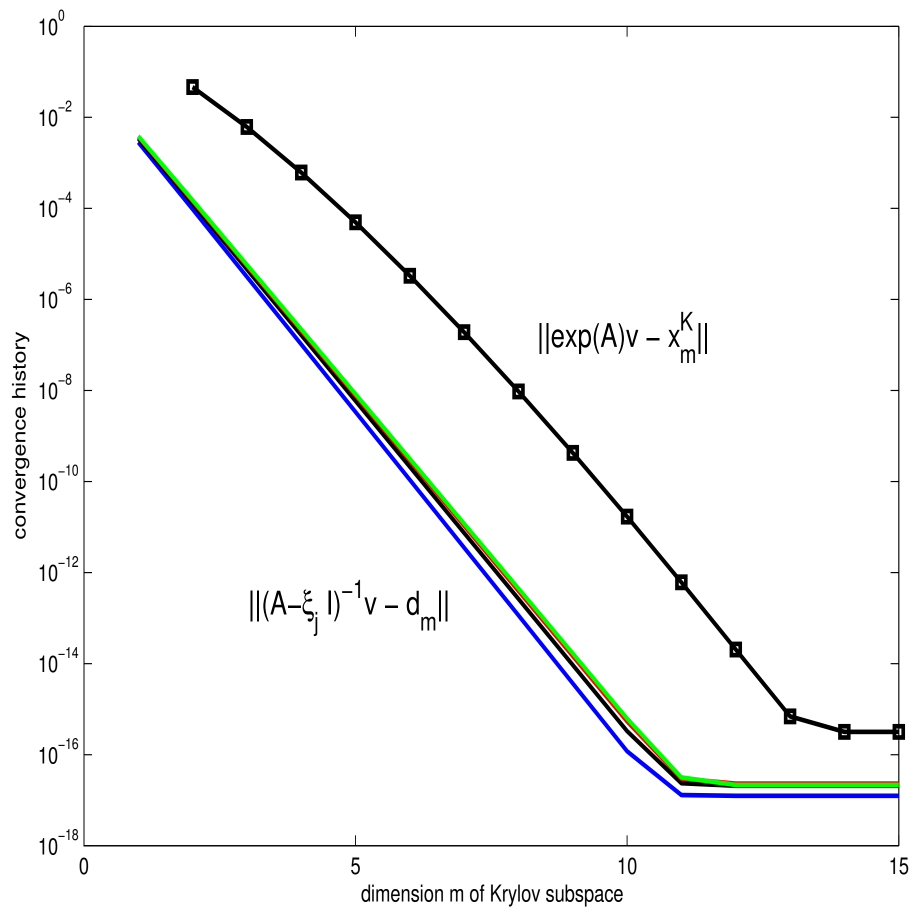
A-posteriori estimate and residual

$$x_\star = \tau_0 v + \sum_{j=1}^{\nu} \tau_j (A - \xi_j I)^{-1} v \approx V_m \left(\tau_0 e_1 + \sum_{j=1}^{\nu} \tau_j (H_m - \xi_j I)^{-1} e_1 \right)$$

Defining $\quad r_m^K := \sum_{j=1}^{\nu} \tau_j r_m^{(j)} \quad$ ($r_m^{(j)}$ single residuals) we have

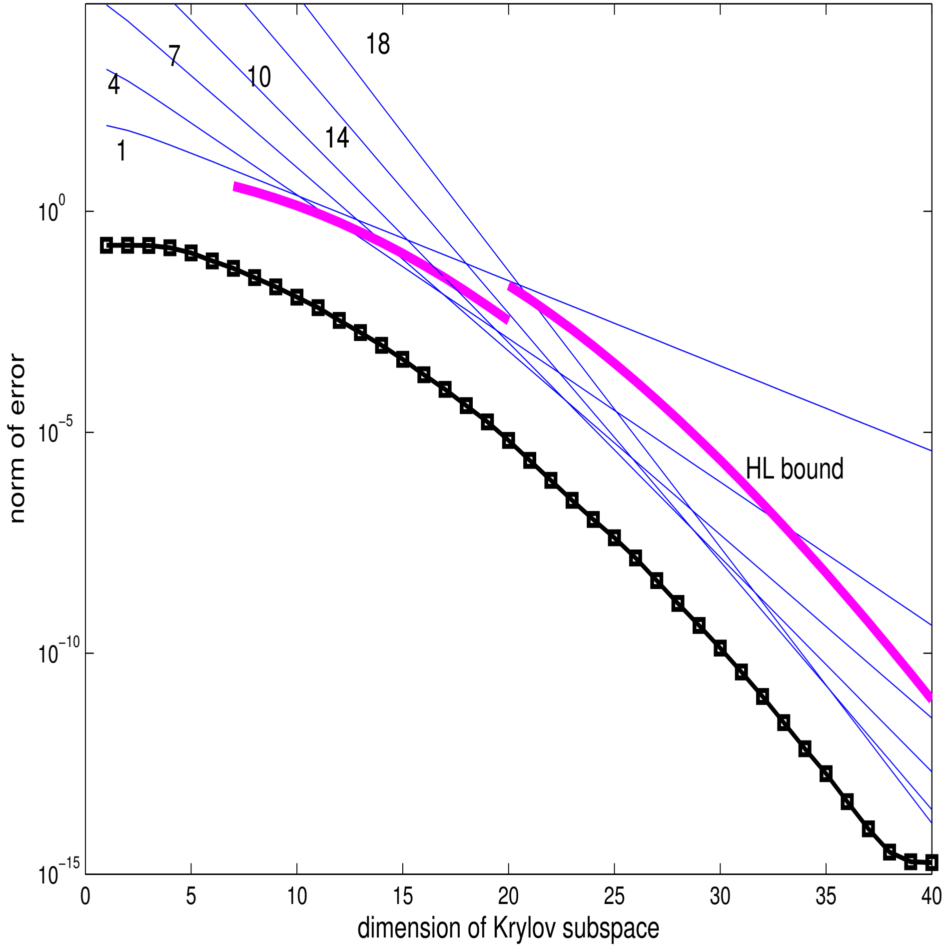
$$h_{m+1,m} |e_m^* y_m^K| = \|r_m^K\|$$

$\star$ Relation to convergence of systems $(A - \xi_j I)x = v, j = 1, \ldots, \nu$



(Padé, $\nu = 7$)

Recovering superlinear convergence



$A \in \mathbb{R}^{1001\times 1001}$, diagonal, uniform distr. in $[-40, 0]$

Acceleration strategies

Hochbruck & van den Eshof (SISC '06): for $f(\lambda) = \exp(\lambda)$

$$x = f(A)v \quad \Rightarrow \quad \begin{array}{l} x_m \in \mathcal{K}_m((I - \gamma A)^{-1}, v), \quad \gamma > 0 \\ \\ x_m = V_m f(\frac{1}{\gamma}(H_m^{-1} - I))e_1 \end{array}$$

If $f(\lambda) = \mathcal{R}_\nu(\lambda)$, $\qquad \mathcal{R}_\nu(A)v = \tau_0 v + \sum_j \tau_j (A - \xi_j I)^{-1} v$

$\Rightarrow \quad x_m$ corresponds to preconditioning $(A - \xi_j I)d = v$:

$$(A - \xi_j I)d = v \;\; \text{preconditioned with} \quad (I - \gamma A)^{-1}$$

Popolizio & Simoncini, tr.2006

Acceleration strategies. Cont'd

Connection to Partial Fraction Expansion used to select "optimal" parameter γ in

$$x_m = V_m f\left(\frac{1}{\gamma}(H_m^{-1} - I)\right)e_1 \;\approx\; \tau_0 v + \sum_j \tau_j (A - \xi_j I)^{-1} v$$

for $\sigma(A) \in [\alpha, 0]$ large, leads to a *discrete* min-max problem:

$$\max_{\substack{\xi_i \\ i=1,\ldots,\nu}} \quad \min_{\substack{\xi_j \\ j=1,\ldots,\nu}} \frac{|\xi_i| + |\xi_j|}{||\xi_i| - \xi_j|}$$

ν	7	8	9	10	11	12	13	14
$\gamma_{\rm opt}^{-1}$	0.1264	0.1062	0.0914	0.0801	0.0711	0.0639	0.0580	0.0530

* Chebyshev approx

Numerical results. 1

- Partial Fraction Expansion (PFE). Explicit numerical solution of

$$\tau_0 v + \sum_j \tau_j (A - \xi_j I)^{-1} v$$

Systems corresponding to conjugate pairs are coupled

- Standard Lanczos: $\exp(A)v \approx V_m \exp(H_m)e_1 \in \mathcal{K}_m(A, v)$

- Shift-Invert Lanczos (SI):

$$x_m \in \mathcal{K}_m((I - \gamma A)^{-1}, v), \qquad x_m = V_m f(\frac{1}{\gamma}(H_m^{-1} - I))e_1$$

Note: PFE and SI require solving *shifted* systems

Numerical results. Direct system solvers.

Ex.1 A = 3D Laplace operator. $n = 125 : \sigma(A) \subset [-179.14, -12.862]$

n	tol	Standard Lanczos	Part.Fract. Expansion	Shift-Invert Lanczos
125	10^{-5}	0.01 (13)	0.01	0.01 (7)
	10^{-8}	0.01 (18)	0.01	0.01 (11)
	10^{-11}	0.01 (22)	0.03	0.01 (14)
	10^{-14}	0.01 (24)	0.03	0.01 (17)
3375	10^{-5}	0.14 (47)	1.32	0.48 (8)
	10^{-8}	0.21 (55)	2.13	0.65 (13)
	10^{-11}	0.35 (67)	2.88	0.85 (19)
	10^{-14}	0.52 (77)	3.70	1.06 (25)
15625	10^{-5}	2.69 (89)	30.35	11.49 (10)
	10^{-8}	2.95 (93)	51.61	11.88 (11)
	10^{-11}	4.76 (113)	69.03	14.22 (17)
	10^{-14}	7.25 (130)	90.20	16.96 (24)

E-Times (# its)

Numerical results. Direct system solvers.

Ex.2 $A \approx \mathcal{L}(u) = ((1 + y - x)u_x)_x + ((1 + x + x^2)u_y)_y$

$n = 2500: \ \sigma(A) \subset [-35424, -25.256]$

n	tol	Standard Lanczos	Part.Fract. Expansion	Shift-Invert Lanczos
2500	10^{-5}	16 (194)	0.22	0.12 (10)
	10^{-8}	18 (200)	0.33	0.13 (11)
	10^{-11}	53 (242)	0.44	0.20 (19)
	10^{-14}	111 (280)	0.53	0.24 (24)
10000	10^{-5}	615 (406)	1.24	0.67 (11)
	10^{-8}	610 (406)	1.87	0.66 (11)
	10^{-11}	1221 (484)	2.55	0.94 (17)
	10^{-14}	- ($>$ 500)	3.20	1.24 (23)

Numerical results. Iterative system solvers.

		Stand.Lanczos	PFE+QMR (avg.its)	SI+PCG (out/avg in)
		Example 1		
	10^{-5}	0.14	0.67 (8)	0.44 (8/7)
3375	10^{-8}	0.21	1.15 (11)	0.81 (13/9)
	10^{-11}	0.35	1.75 (14)	1.27 (19/10)
	10^{-5}	2.69	5.29 (11)	4.05 (10/10)
15625	10^{-8}	2.95	9.36 (17)	5.37 (11/13)
	10^{-11}	4.76	14.29 (22)	8.87 (17/15)
		Example 2		
	10^{-5}	16	0.36 (13)	0.54 (10/12)
2500	10^{-8}	18	0.68 (18)	0.75 (11/16)
	10^{-11}	53	1.09 (22)	1.46 (19/18)
	10^{-5}	615	2.46 (24)	4.4 (11/21)
10000	10^{-8}	610	4.92 (35)	5.5 (11/27)
	10^{-11}	1221	8.17 (43)	9.8 (17/32)

Conclusions and Outlook

- Acceleration procedure effective on tough problem
- ... otherwise use Standard Lanczos !!
- Partial Fraction Expansion very efficient with iterative solves

- Further explore acceleration procedures
- Generalize PFA framework to nonsymmetric problem
- Generalize to other functions ($\log$, $\cos$, ϕ-functions, etc.)

Related references

1. A. Frommer and V. Simoncini, *Matrix functions*, tech. rep., Dipartimento di Matematica, Bologna (I), March 2006.

2. L. Lopez and V. Simoncini, *Analysis of projection methods for rational function approximation to the matrix exponential*, SIAM J. Numer. Anal., 44 (2006), pp. 613–635.

3. M. Popolizio and V. Simoncini, *Acceleration Techniques for Approximating the Matrix Exponential Operator*, October 2006, pp.1-24.

valeria@dm.unibo.it

http://www.dm.unibo.it/~simoncin