



The Extended Krylov Subspace for Matrix Function Approximations: Analysis and Applications

V. Simoncini

Dipartimento di Matematica, Università di Bologna

valeria@dm.unibo.it

Partially joint work with Leonid Knizhnerman, Moscow

The Problem

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

$$x = f(A)v$$

★ A large dimensions, $\|v\| = 1$

Applications:

- Numerical solution of evolution PDEs (e.g. $\exp(\lambda)$, $\sqrt{\lambda^{-1}}$, $\cos(\lambda)$, ...)
- Inverse Problems ($\exp(\lambda)$, $\cosh(\lambda)$, ...)
- Fluxes on manifolds
- Problems in Scientific Computing (e.g. QCD, $\text{sign}(\lambda)$)
- (Analysis of) reduced Dynamical System Models (via Gramians)

The Problem

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

$$x = f(A)v$$

★ A large dimensions, $\|v\| = 1$

Projection-type methods

\mathcal{K} approximation space, $m = \dim(\mathcal{K})$

$V \in \mathbb{R}^{n \times m}$ s.t. $\mathcal{K} = \text{range}(V)$

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

The Problem

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

$$x = f(A)v$$

★ A large dimensions, $\|v\| = 1$

Projection-type methods

\mathcal{K} approximation space, $m = \dim(\mathcal{K})$

$V \in \mathbb{R}^{n \times m}$ s.t. $\mathcal{K} = \text{range}(V)$

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

Question: Which \mathcal{K} ?

Some explored alternatives for \mathcal{K}

- Krylov subspace, $\mathcal{K} = K_m(A, v)$
- Shift-Invert Krylov subspace, $\mathcal{K} = K_m((I + \gamma A)^{-1}, v)$ for some γ
- Rational Krylov subspace, for some $\omega_1, \omega_2, \dots$
 $\mathcal{K} = \text{span}\{v, (A - \omega_1 I)^{-1}v, (A - \omega_2 I)^{-1}v, \dots\}$
- Extended Krylov subspace, $\mathcal{K} = K_m(A, v) + K_m(A^{-1}, A^{-1}v)$
- Restarted Krylov subspace

Note: In all cases, A nonsymmetric.

Theory mostly for field of values of A in \mathbb{C}^+

Field of values: $W(A) = \{x^* Ax, x \in \mathbb{C}^n, \|x\| = 1\}$

Krylov subspace approximation

“Classical” approach:

$$\mathcal{K} = K_m(A, v) = \text{span}\{v, Av, \dots, A^{m-1}v\}$$

For $H_m = V_m^\top AV_m$, $v = V_m e_1$ and $V_m^\top V_m = I_m$:

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

Polynomial approximation: $x_m = p_{m-1}(A)v$
(p_{m-1} interpolates f at eigenvalues of H_m)

★ Numerical and theoretical results since mid '80s (Saad '92)

Acceleration Procedures: Shift-Invert Krylov

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

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

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

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

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

Choice of γ : A spd, $\gamma = \frac{1}{\sqrt{\lambda_{\min} \lambda_{\max}}}$ (Moret, 2009)

A nonsym, (Beckermann-Reichel tr2008)

Acceleration Procedures: Restarted Krylov

$$AV_m^{(1)} = V_m^{(1)} H_m^{(1)} + v_{m+1}^{(1)} h_{m+1,m}^{(1)} e_m^\top \quad (V_m^{(1)})^\top V_m^{(1)} = I$$

$$AV_m^{(2)} = V_m^{(2)} H_m^{(2)} + v_{m+1}^{(2)} h_{m+1,m}^{(2)} e_m^\top \quad (V_m^{(2)})^\top V_m^{(2)} = I$$

with $V_m^{(2)} e_1 = v_{m+1}^{(1)}$. Then

$$A[V_m^{(1)}, V_m^{(2)}] = [V_m^{(1)}, V_m^{(2)}] \widehat{H}_{2m} + v_{m+1}^{(2)} h_{m+1,m}^{(2)} e_{2m}^\top,$$

with

$$\widehat{H}_{2m} = \begin{bmatrix} H_m^{(1)} & 0 \\ e_1 h_{m+1,m}^{(1)} e_m^\top & H_m^{(2)} \end{bmatrix}.$$

Therefore (Eiermann-Ernst, '06)

$$\begin{aligned} f(A)v &\approx x_m^{(1)} = V_m^{(1)} f(H_m^{(1)}) \\ &\approx x_m^{(2)} = V_m^{(1)} f(H_m^{(1)}) e_1 + V_m^{(2)} f(\widehat{H}_{2m}) e_1|_{(2)} \\ &x_m^{(2)} = x_m^{(1)} + V_m^{(2)} f(\widehat{H}_{2m}) e_1|_{(2)} \end{aligned}$$

Acceleration Procedures: Extended Krylov

For A nonsingular,

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

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

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: an effective implementation

$m_1 = m_2 = m$ **not** fixed a priori

$$\begin{aligned}\mathcal{K} &= K_m(A, v) + K_m(A^{-1}, A^{-1}v) \\ &= \text{span}\{v, A^{-1}v, Av, A^{-2}v, A^2v, \dots\}\end{aligned}$$

★ *Block* Arnoldi-type recurrence:

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

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

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

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

Simoncini, 2007

Extended Krylov: Convergence theory I

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

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

Druskin-Knizhnerman 1998:

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

Extended Krylov: Convergence theory II

For nonsingular A , with $0 \notin W(A)$, let $f = f_1 + f_2$, $a \in [0, \infty)$. Then

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

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

$\Phi_j, F_{j,k}$, $j = 1, 2$ conformal map and Faber Polyn for $W(A)$ and $W(A)^{-1}$

There exists $a > 0$ s.t. $|\Phi_1(-a)| = |\Phi_2(-\frac{1}{a})|$ so that

$$\|f(A)v - \mathcal{U}_m f(\mathcal{T}_m)e_1\| \leq \frac{c}{|\Phi_1(-a)|^m}$$

Extended Krylov: Convergence theory II

For nonsingular A , with $0 \notin W(A)$, let $f = f_1 + f_2$, $a \in [0, \infty)$. Then

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

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

$\Phi_j, F_{j,k}, j = 1, 2$ conformal map and Faber Polyn for $W(A)$ and $W(A)^{-1}$

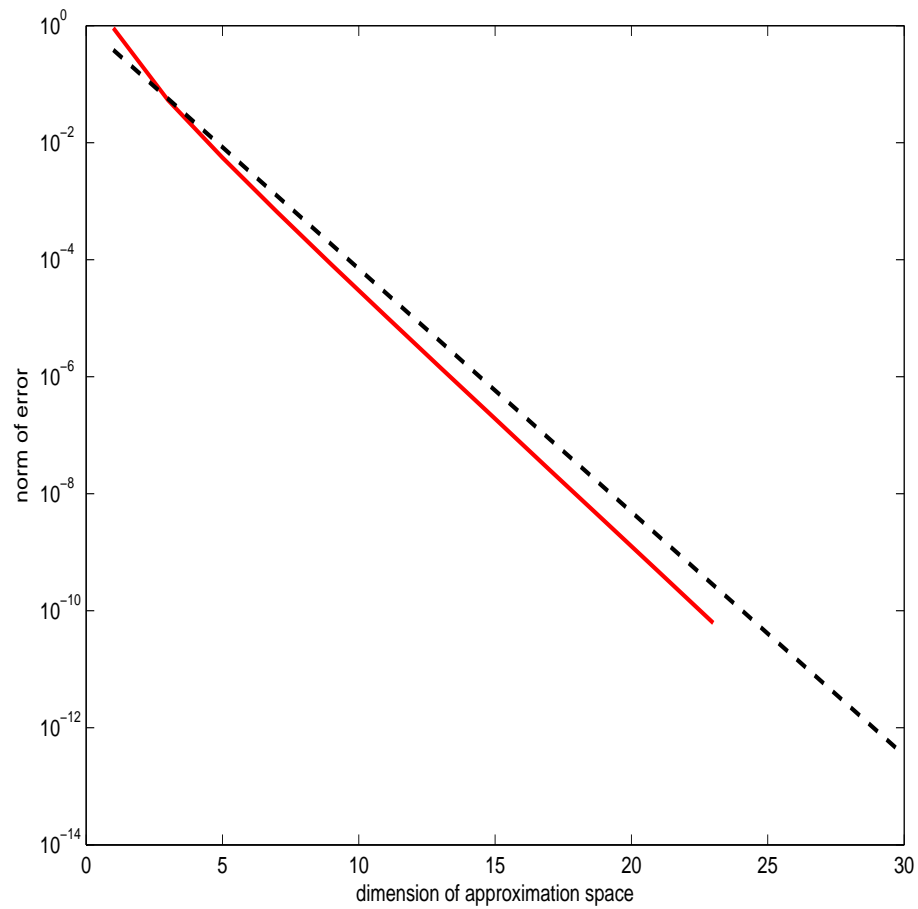
There exists $a > 0$ s.t. $|\Phi_1(-a)| = |\Phi_2(-\frac{1}{a})|$ so that

$$\|f(A)v - \mathcal{U}_m f(\mathcal{T}_m)e_1\| \leq \frac{c}{|\Phi_1(-a)|^m}$$

e.g. for A symmetric (Φ_1, Φ_2 known, $a = \sqrt{\lambda_{\min} \lambda_{\max}}$) :

$$\|x - x_m\| = O(\exp(-2m \sqrt[4]{\frac{\lambda_{\min}}{\lambda_{\max}}}))$$

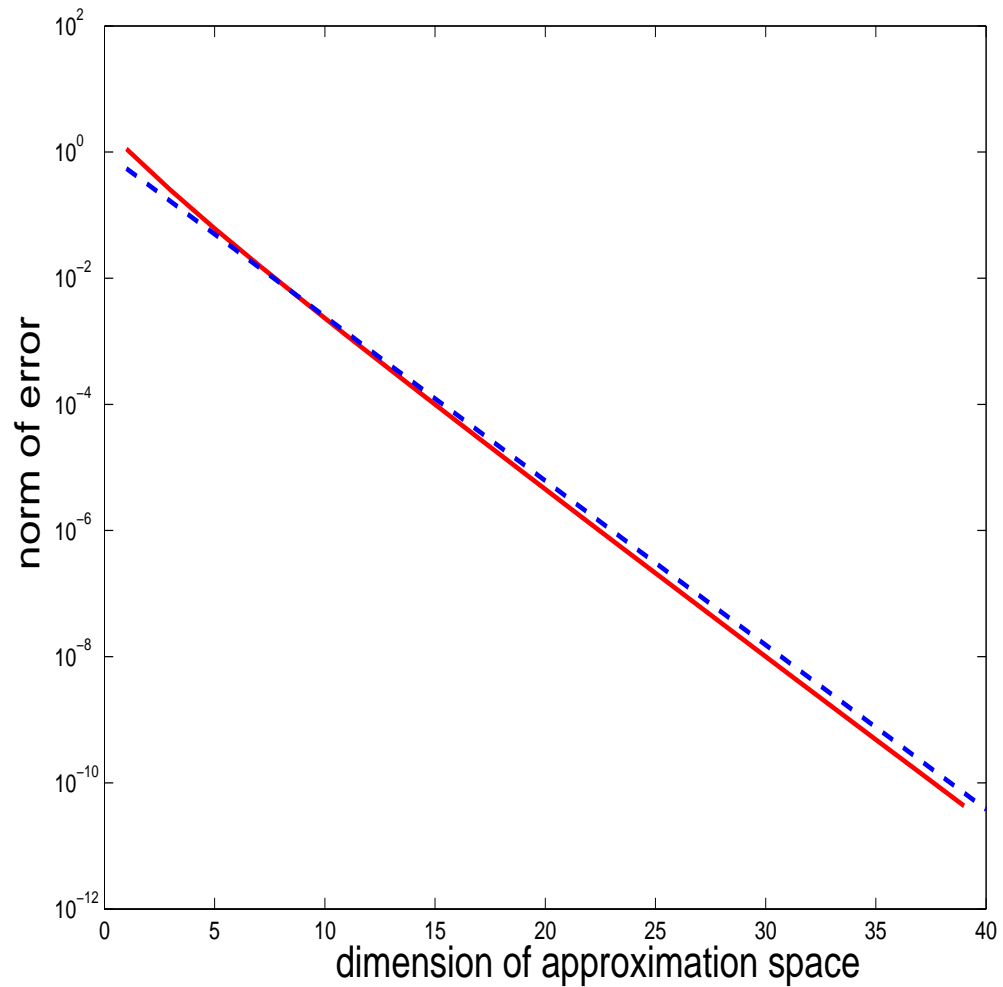
Convergence rate. $A \in \mathbb{R}^{400 \times 400}$ spd. $f(\lambda) = \lambda^{-1/2}$



$$\sigma(A) \subseteq [1, 50]$$

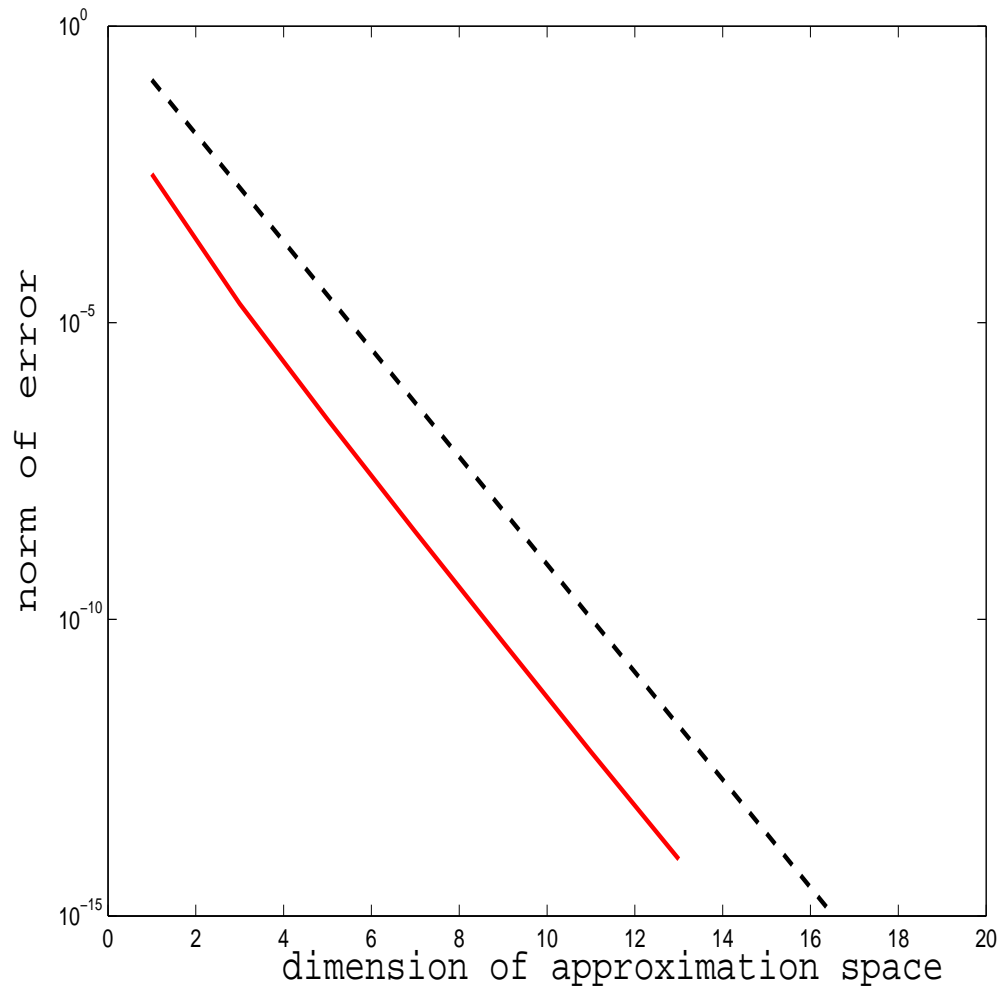
Uniform spectral distribution

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

Experiment. $A \in \mathbb{R}^{200 \times 200}$ Jordan block, $\lambda = 4$. $f(\lambda) = \lambda^{1/2}$



$W(A)$ disk centered at λ and unit radius

Large-scale numerical experiments

A from FD discretization of

$$\mathcal{L}_1(u) = -100u_{x_1x_1} - u_{x_2x_2} + 10x_1u_{x_1},$$

$$\mathcal{L}_2(u) = -100u_{x_1x_1} - u_{x_2x_2} - u_{x_3x_3} + 10x_1u_{x_1},$$

$$\mathcal{L}_3(u) = -e^{-x_1x_2}u_{x_1x_1} - e^{x_1x_2}u_{x_2x_2} + \frac{1}{x_1 + x_2}u_{x_1},$$

$$\mathcal{L}_4(u) = -\operatorname{div}(e^{3x_1x_2}\operatorname{grad}u) + \frac{1}{x_1 + x_2}u_{x_1}$$

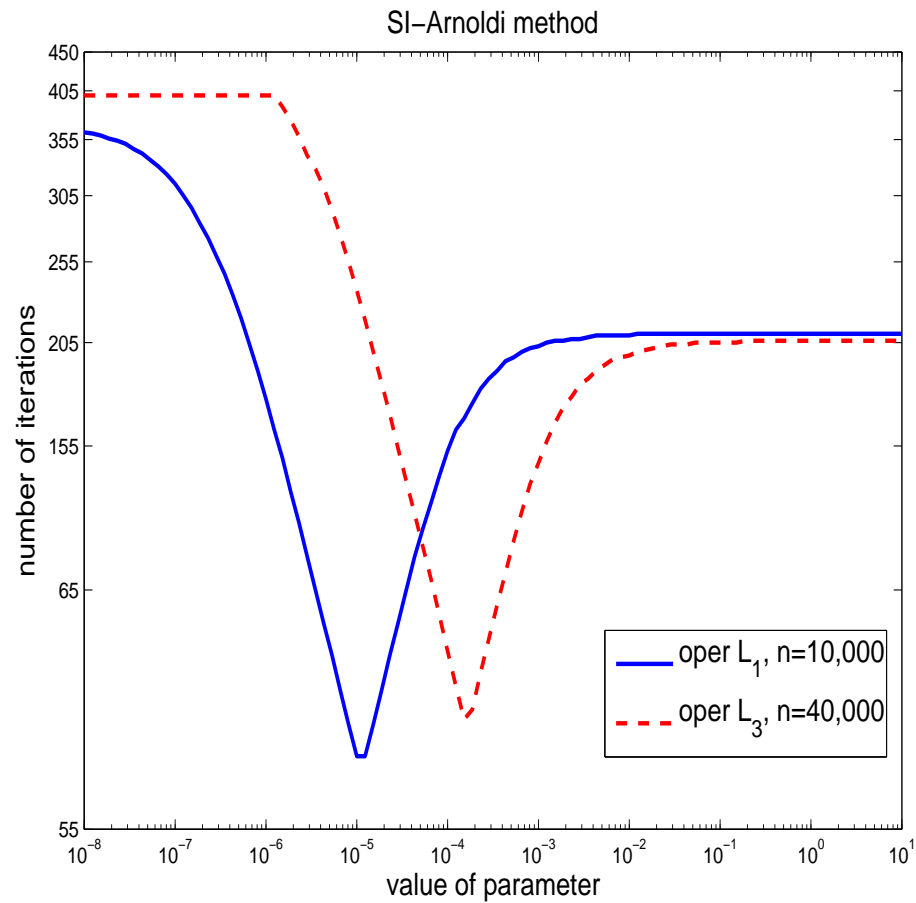
on unit square/cube, Dirichlet hom. bc.

Inner system solves:

- Extended Krylov: systems with A solved with GMRES/AMG
- SI-Arnoldi: systems with $I + \gamma A$ solved with IDR(s)/ILU

An intermezzo

SI-Arnoldi requires getting the parameter γ :



Number of SI-Arnoldi iterations as a function of the parameter for $f(\lambda) = \lambda^{1/2}$

Comparisons: CPU Time in Matlab (space dim.)

f	Oper.	n	SI-Arnoldi	EKSM	Std Krylov
$\lambda^{1/2}$	\mathcal{L}_1	2500	0.9 (59)	0.6 (48)	7 (193)
		10000	4.0 (66)	3.6 (68)	*46 (300)
		160000	642.9(246)	219.7(122)	*458(300)
	\mathcal{L}_2	27000	10.8 (55)	7.4 (40)	6.7(119)
		125000	86.7 (60)	65.3 (52)	138.7(196)
	\mathcal{L}_3	40000	26.3 (75)	21.1 (72)	*87 (300)
		160000	318.5(144)	173.3 (96)	*442(300)
	\mathcal{L}_4	40000	41.1(117)	25.4(106)	*89 (300)
		160000	580.2(442)	231.2(144)	*461 (300)

Comparisons: CPU Time in Matlab (space dim.)

f	Oper.	n	SI-Arnoldi	EKSM	Std Krylov
$\lambda^{-1/3}$	\mathcal{L}_1	2500	0.6 (43)	0.4 (30)	2.2(131)
		10000	2.6 (46)	1.8 (38)	26.2(252)
		160000	79.3 (48)	99.7 (64)	*460(300)
	\mathcal{L}_2	27000	7.8 (41)	4.8 (26)	3.1 (82)
		125000	64.8 (45)	38.9 (32)	67.5(138)
	\mathcal{L}_3	40000	20.7 (61)	13.7 (48)	*88 (300)
		160000	116.5 (62)	105.2 (62)	*460 (300)
	\mathcal{L}_4	40000	35.8(104)	14.2 (66)	*88 (300)
		160000	208.1(104)	112.2 (84)	*461 (300)

Stopping criterion

Unlike linear systems: **no equation \Rightarrow no residual**

Estimate of the error:

(first suggested for $f(\lambda) = e^{-\lambda}$ by van den Eshof-Hochbruck '06)

$$\frac{\|x - x_m\|}{\|x_m\|} \approx \frac{\delta_{m+j}}{1 - \delta_{m+j}}, \quad \delta_{m+j} = \frac{\|x_{m+j} - x_m\|}{\|x_m\|}$$

Stopping criterion:

$$\text{if } \frac{\delta_{m+j}}{1 - \delta_{m+j}} \leq \text{tol then stop}$$

Computational costs awareness: inexact solves in EKSM

systems with A : GMRES with **relaxed** inner tolerance

$$\epsilon_m^{(\text{inner})} = \frac{\text{tolin}}{\|x - x_{m-1}\|}.$$

Final outer error (# outer its / # inner its)

tolin	fixed inner tol	relaxed inner tol
1e-10	6.97e-11 (24/901)	6.58e-11 (24/559)
1e-12	6.48e-11 (24/1052)	6.48e-11 (24/716)

$$\mathcal{L}(u) = -u_{xx} - u_{yy} - u_{zz} + 50(x + y)u_x$$

$$f(\lambda) = \lambda^{-1/3} \quad \epsilon^{(\text{outer})} = 10^{-10}$$

A special case: $f(\lambda) = (\lambda - \sigma)^{-1}$. $x = (A - \sigma I)^{-1}v \equiv f_\sigma(A)v$

All as before - and **new perspective**:

- Many shifts in a wide range (e.g., Structural dynamics, electromagn.)

$$(A - \sigma_j I)x = v, \quad \sigma_j \in [\alpha, \beta], \quad \text{large interval}$$

$$j = 1, \dots, k, \quad k = \mathcal{O}(100)$$

- Few shifts (e.g., quadrature formulas)

$$z = \sum_{j=1}^k \omega_j (A - \sigma_j I)^{-1} v$$

- Transfer function

$$h(\sigma) = c^* (A - i\sigma I)^{-1} b, \quad \sigma \in [\alpha, \beta]$$

Shifted systems

Extended Krylov subspace method:

$$x \approx \mathcal{U}_m f_\sigma(\mathcal{T}_m) e_1 = \mathcal{U}_m (\mathcal{T}_m - \sigma I)^{-1} e_1$$

standard Galerkin-type approximation for shifted systems

(cf. FOM, CG, ...)

Key fact: A single \mathcal{K} for all shifted systems.

Shift invariance:

$$K_m(A, v) = K_m(A + \sigma I, v)$$

Note: Solve systems with A to approximate $(A + \sigma I)^{-1} v$

Added feature: restarting made easy

$$A\mathcal{U}_m = \mathcal{U}_m\mathcal{T}_m + U_{m+1}\boldsymbol{\tau}E_m^\top$$

Proportionality of the residuals:

For $x_m(\sigma) = \mathcal{U}_m(\mathcal{T}_m - \sigma I)^{-1}e_1$, the residual

$$r_m(\sigma) := b - (A - \sigma I)x_m(\sigma), \quad r_m(\sigma) \propto U_{m+1}e_{2m+1} \quad \forall \sigma$$

(typical of Galerkin-type procedures)

Added feature: restarting made easy

$$A\mathcal{U}_m = \mathcal{U}_m\mathcal{T}_m + U_{m+1}\boldsymbol{\tau}E_m^\top$$

Proportionality of the residuals:

For $x_m(\sigma) = \mathcal{U}_m(\mathcal{T}_m - \sigma I)^{-1}e_1$, the residual

$$r_m(\sigma) := b - (A - \sigma I)x_m(\sigma), \quad r_m(\sigma) \propto U_{m+1}e_{2m+1} \quad \forall \sigma$$

(typical of Galerkin-type procedures)

Restarting with a single approximation space:

$$\mathcal{K}^{(2)} = K_m(A, v^{(2)}) + K_m(A^{-1}, A^{-1}v^{(2)}), \quad v^{(2)} = U_{m+1}e_{2m+1}$$

$$x_m^{(2)}(\sigma) = x_m^{(1)}(\sigma) + z_m, \quad z_m \in \mathcal{K}^{(2)}$$

An example from Structural Dynamics

$$(K^* - \sigma^2 M)x = b, \quad \Rightarrow \quad (K^* M^{-1} - \sigma^2 I)\hat{x} = b$$

K^* stiffness + hysteretic damping, M mass $\sigma \in 2\pi[0.1, 60.1]$
frequencies, $n = 3627$

number of restarts (subspace dimension)

restarted EKSM	restarted FOM
3 (20)	- (20)
1 (34)	81(40)
	21(80)

Transfer function approximation (cf. MOR)

$$h(\sigma) = c^*(A - i\sigma I)^{-1}b, \quad \sigma \in [\alpha, \beta]$$

Given space \mathcal{K} and V s.t. $\mathcal{K} = \text{range}(V)$,

$$h(\sigma) \approx (V^*c)^*(V^*AV - \sigma I)^{-1}(V^*b)$$

For $\mathcal{K} = K_m(A, b)$ (standard Krylov):

$$h_m(\sigma) = (V_m^*c)^*(H_m - \sigma I)^{-1}e_1 \|b\|$$

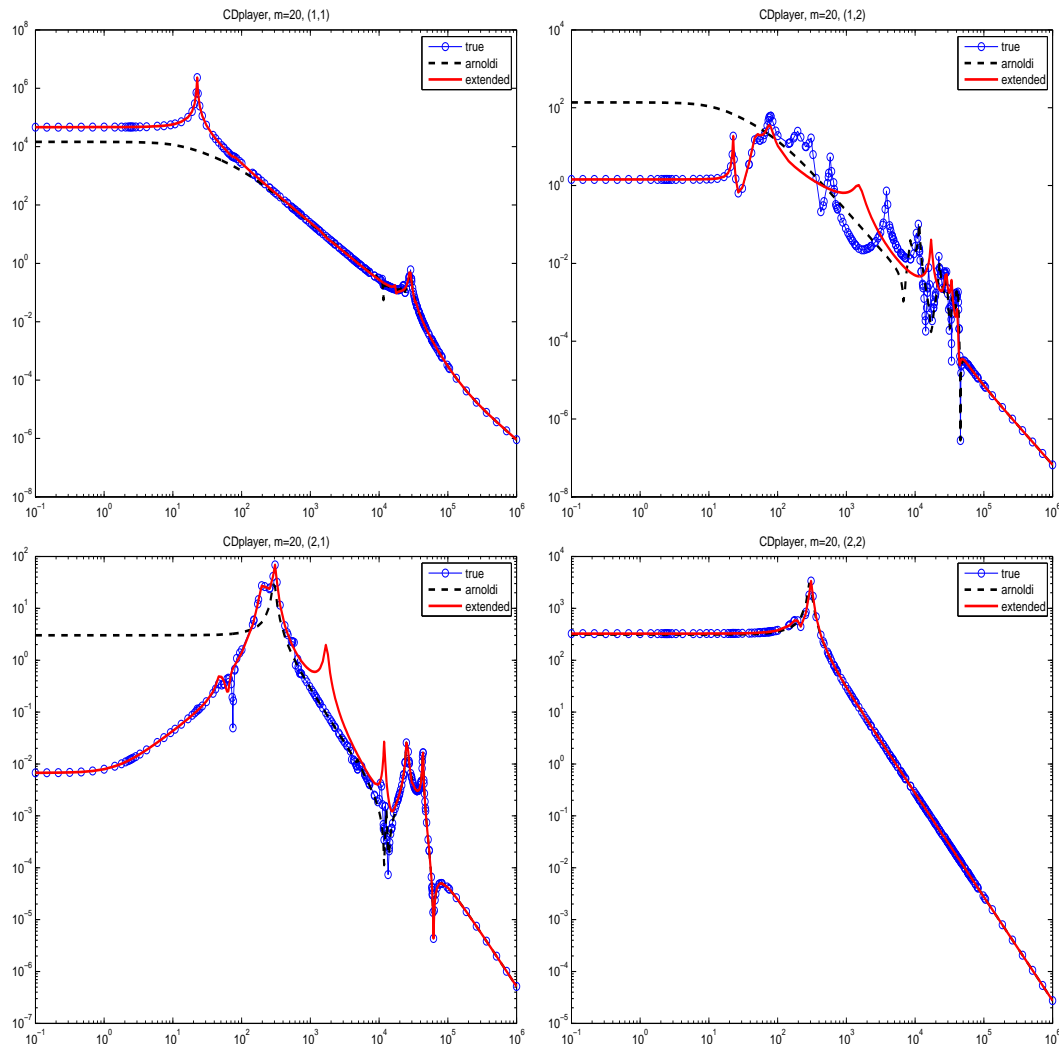
For $\mathcal{K} = K_m(A, b) + K_m(A^{-1}, A^{-1}b)$ (EKSM):

$$h_m(\sigma) = (U_m^*c)^*(\mathcal{T}_m - \sigma I)^{-1}e_1 \|b\|$$

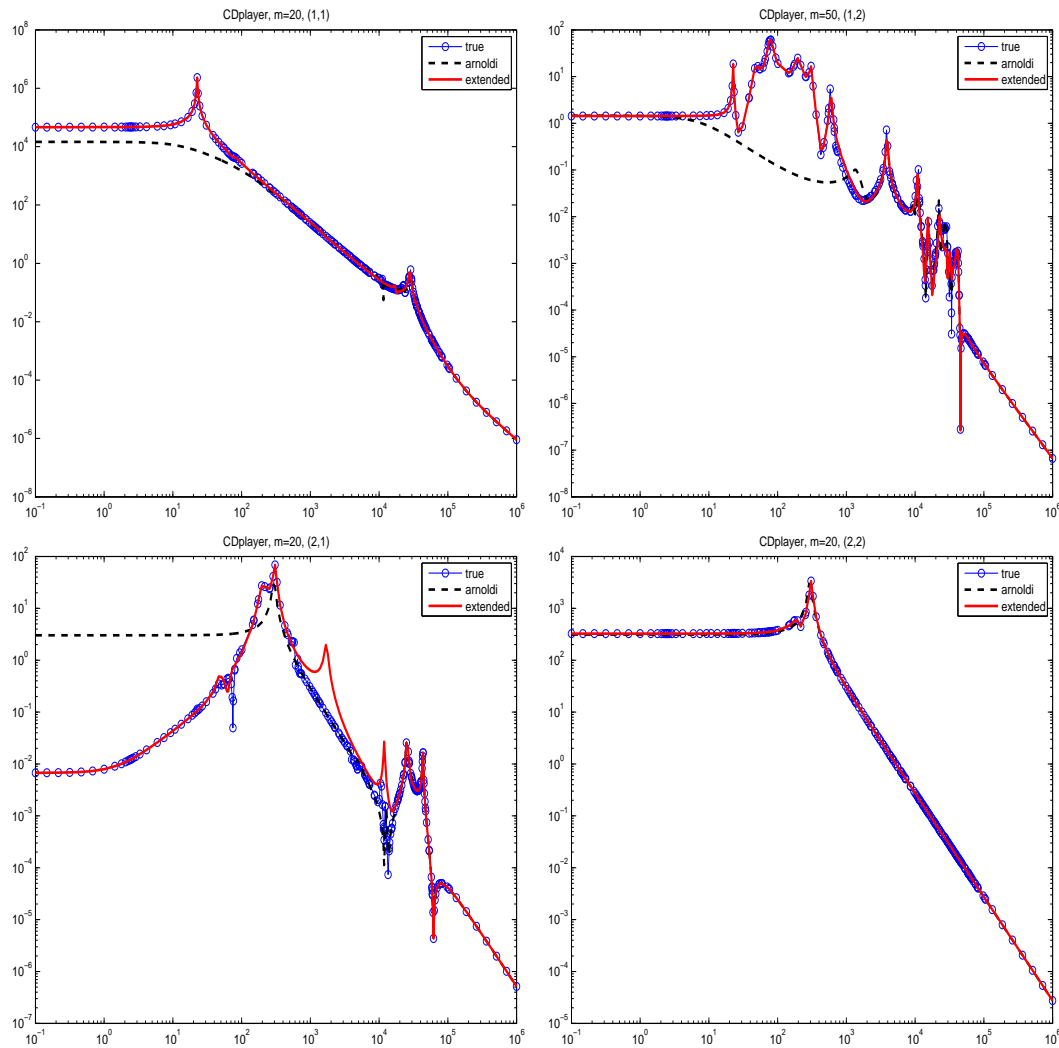
Alternative: Rational Krylov (Grimme-Gallivan-VanDooren etc.)

choosing the poles unresolved issue

An example: CD Player, $|h(\sigma)| = |C_{:,i}^*(A - i\sigma I)^{-1}B_{:,j}|$



An example: CD Player, $|h(\sigma)| = |C_{:,i}^*(A - i\sigma I)^{-1}B_{:,j}|$



Two-parameter linear systems

$$(A + \beta_j B + \alpha_i I)x = v, \quad x = x(\alpha_i, \beta_j)$$

Commonly: $\#\alpha = O(100)$, $\#\beta = O(10)$

Problem: Solve all systems at a cost sublinear in $\#\alpha, \#\beta$

Two-parameter linear systems

$$(A + \beta_j B + \alpha_i I)x = v, \quad x = x(\alpha_i, \beta_j)$$

Commonly: $\#\alpha = O(100)$, $\#\beta = O(10)$

Problem: Solve all systems at a cost sublinear in $\#\alpha, \#\beta$

Currently, Shifted restarted EKSM most efficient strategy:

For each β_j , solve $(A + \beta_j B + \alpha_i I)x = v, \quad \forall \alpha_i$

But: **linear** in β ...

Conclusions

- Efficient generation of the Extended Krylov subspace
- Complete theory for EKSM for a large class of functions
- Performance:
 - Competitive with respect to available methods
(when solving with A can be made cheap)
 - Does not depend on parameters
 - Projection-type method: wide applicability (work in progress)

valeria@dm.unibo.it, <http://www.dm.unibo.it/~simoncin>