Projection methods for approximating matrix functions

V. Simoncini

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

Approximation problem

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

$$x = f(A)\,v \qquad \text{e.g.} \qquad f(\lambda) = e^{\lambda}$$

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

Problem in context

Wide range of applications. Here we focus on

- 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:

$$x = f(A)\,v \quad \approx \quad \mathcal{R}_{\mu,\nu}(A)v \qquad\qquad \mathcal{R}_{\mu,\nu}(\lambda) = \frac{\Phi_\mu(\lambda)}{\Psi_\nu(\lambda)}$$

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

Approximation using Krylov subspace

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

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

Arnoldi relation

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

A common approach

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

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

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

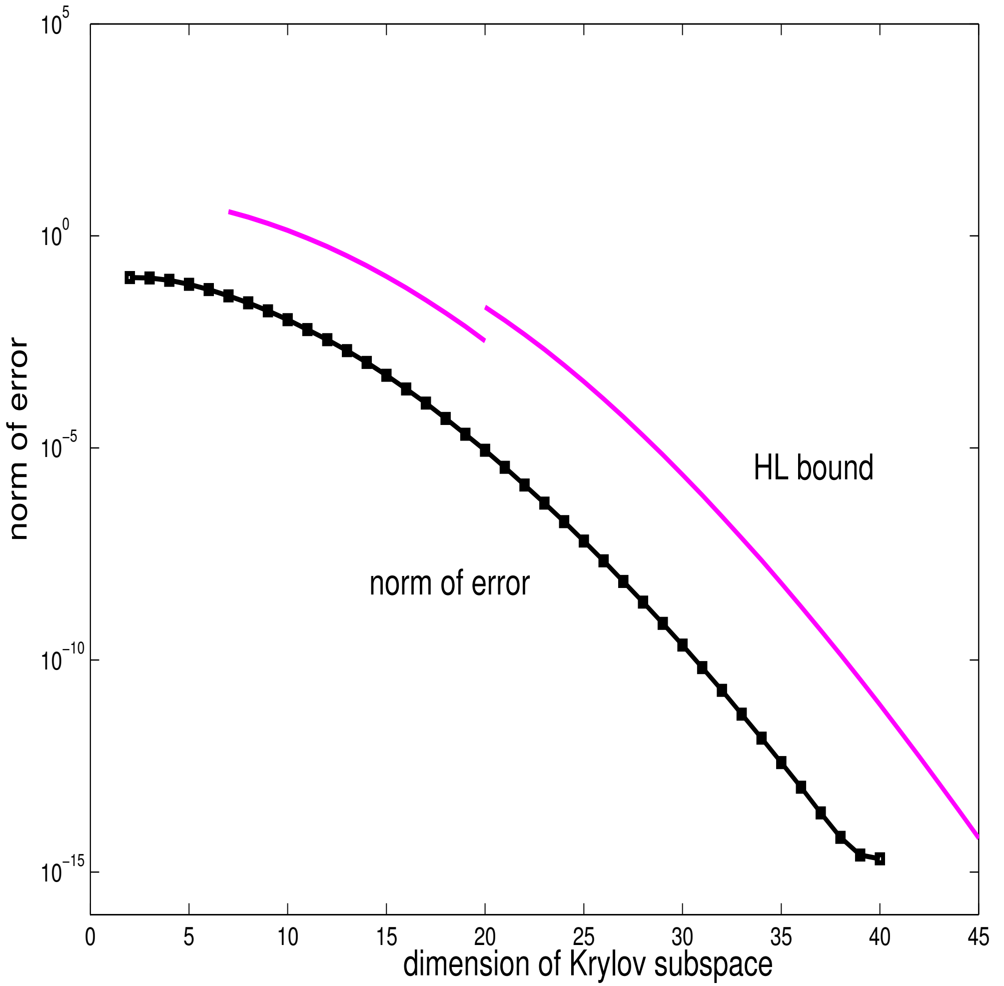
Typical convergence bounds (Hochbruck & Lubich '97)

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

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

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

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)

Exploring Krylov subspace approximation

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

$$\exp(\lambda) \;\approx\; \mathcal{R}_\nu(\lambda) = \frac{\Phi_\nu(\lambda)}{\Psi_\nu(\lambda)} \qquad \text{Rational function approx}$$

- Increase our understanding of approximation in $\mathcal{K}_m(A, v)$
- Set up the stage for acceleration procedures

Mostly taken from: Lopez & S. (to appear in SINUM)

Projection of Rational functions onto Krylov subspaces

Basic fact:

If, for instance, $x_m \approx \mathcal{R}_\nu(A)v$ rational approx. 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:

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

Linear bounds for convergence rate

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

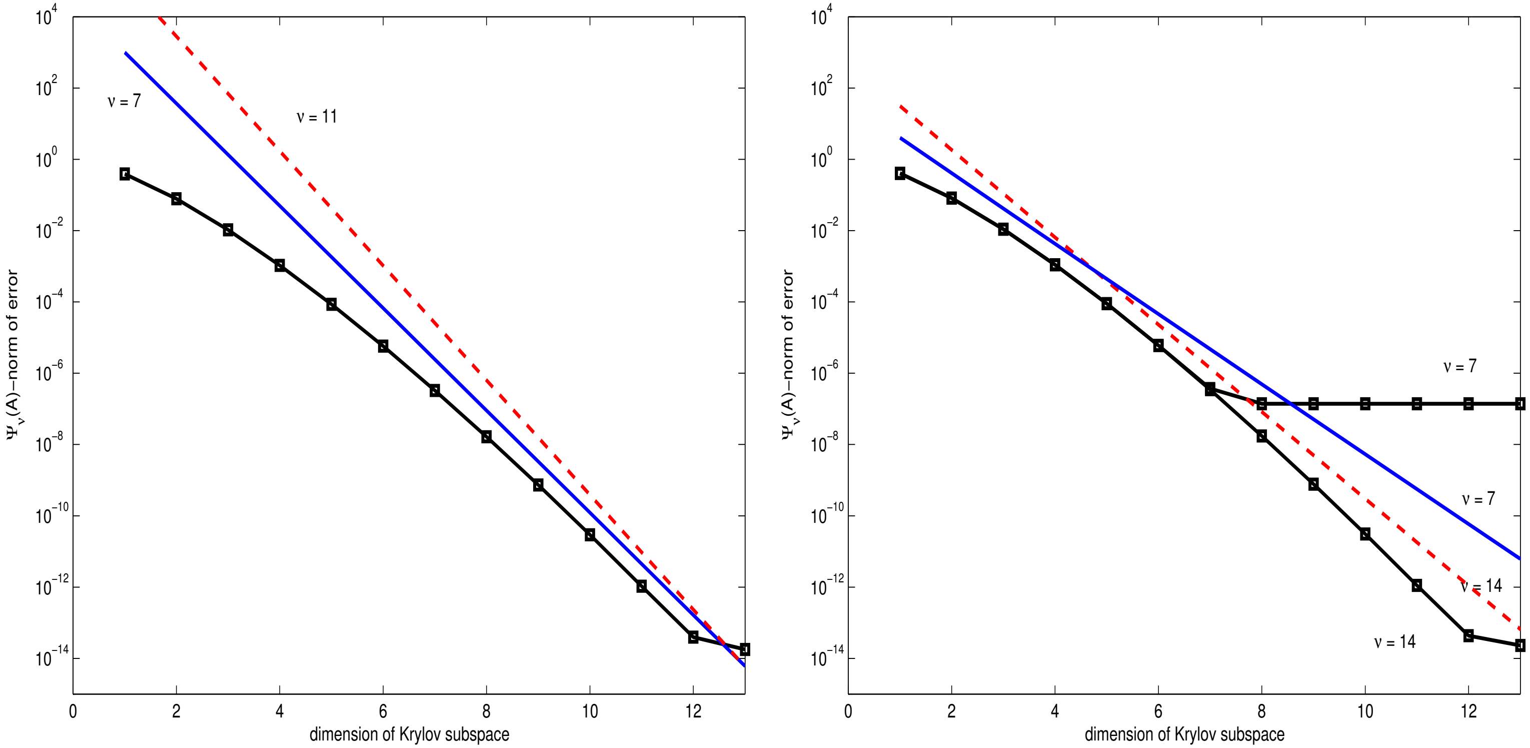
$$x_\star = \Psi_\nu(A)^{-1}\Phi_\nu(A)_\nu v = \tau_0 v + \sum_{j=1}^{\nu} \tau_j (A - \xi_j I)^{-1} v$$

Convergence bound:

$$\frac{\|x_\star - x_m^G\|_{\Psi_\nu(A)}}{\|x_\star - x_0^G\|_{\Psi_\nu(A)}} \leq \sum_{j=1}^{\nu} \eta_j \frac{1}{\rho_j^m + 1/\rho_j^m}$$

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

Galerkin approximation



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

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

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

Krylov approximation

$$x_\star = \exp(A)v \approx \begin{array}{rcl} & & V_m \exp(H_m)e_1 \approx \\ V_m y_m^K & = & V_m \Psi_\nu(H_m)^{-1}\Phi_\nu(H_m)e_1 \end{array}$$

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

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

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

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

Comparison with Galerkin approximation

Galerkin and Krylov solutions "hand-in-hand" convergence:

If $m > \nu$, then

$$\|y_m^G - y_m^K\| \le \gamma \, \|(y_m^K)_{m-\nu+1:m}\|, \qquad \gamma = O(h_{m+1,m}^2)$$

where

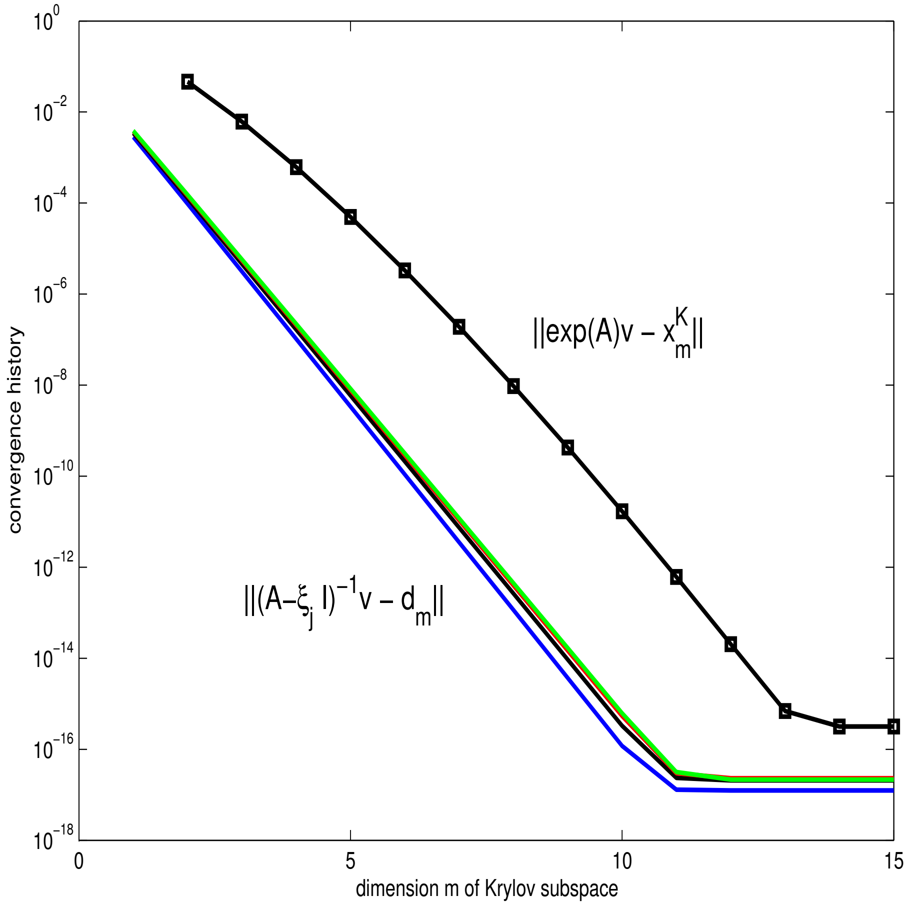
$$|e_k^* y_m^K| \le \sum_{j=1}^{\nu} \frac{|\tau_j|}{\sigma_{\min}(H_m - \xi_j I)} \|r_{k-1}^{(j)}\|, \qquad 1 < k \le m$$

$r_{k-1}^{(j)}$ residual of system $(A - \xi_j I)x = v$ after $k-1$ iterations

τ_j partial fraction coeff's $\qquad$ $\sigma_{\min}(\cdot)$ smallest singular value

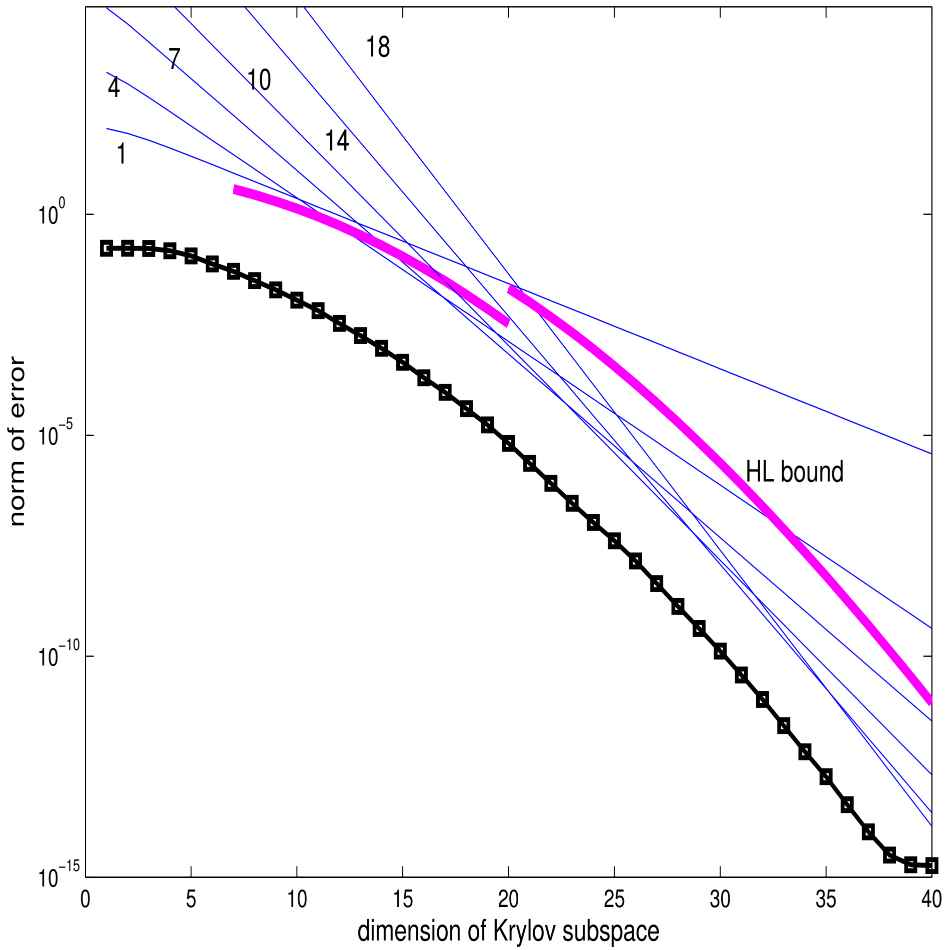
⋆ Similar (linear) convergence estimates as for Galerkin

⋆ 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 I

Hochbruck & van den Eshof ('05)

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

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

However:

If $f(\lambda) = \mathcal{R}_\nu(\lambda)$, x_m corresponds to preconditioning $(A - \xi_j I)d = v$:

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

$(A - \xi_j I)d = v$ preconditioned with $(A - \frac{1}{\gamma} I)$

(Popolizio & S., in preparation)

Acceleration strategies II

Eiermann & Ernst (Tr. '05)

Restarting procedure (small m)

However:

If $f(\lambda) = \mathcal{R}_\nu(\lambda)$, restarted procedure corresponds to restarted FOM on each $(A - \xi_j)d = v$:

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

FOM(m) for $(A - \xi_j I)d = v$

Structure preserving approaches

Motivational problem:

Approximate k largest Lyapunov exponents of

$$x'(t) = \mathcal{A}(t)x, \quad x \in \mathbb{R}^n,$$

This can be accomplished by using the associated system

$$Q_t = A(Q,t)Q, \quad Q \in \mathbb{R}^{n \times k} \qquad A \text{ skew-sym}$$

Q orthonormal columns (Stiefel manifold)

Goal:

numerical method that preserves orthogonality for long time intervals

⋆ A skew-sym. $\Rightarrow$ $\exp(tA)$ unitary, $Q = \exp(tA)Q^{(0)}$ orthogonal

Preserving orthogonality in Krylov subspace

Let $Q^{(0)} = [q_1^{(0)}, \ldots, q_k^{(0)}]$

Regular Krylov subspaces $\mathcal{K}_m(A, q_i^{(0)})$, $i = 1, \ldots, k$

$$A \text{ skew-sym} \quad \Rightarrow \quad H_{m,i} \text{ skew-sym} \quad \Rightarrow \exp(tH_{m,i}) \text{ unitary}$$

This is not enough:

$$\exp(tA)q_i^{(0)} \quad \approx \quad q_i = V_{m,i}\exp(tH_{m,i})e_1$$

$\{q_1, \ldots, q_k\}$ not orthogonal (though unit norm)

Block Krylov methods come to rescue

Block Krylov subspace $\mathcal{K}_m(A, Q^{(0)})$ $\qquad Q^{(0)} = [q_1^{(0)}, \ldots, q_k^{(0)}]$

- $\mathcal{V}_m$ orthonormal columns,

  $\mathcal{H}_m = \mathcal{V}_m^T A \mathcal{V}_m$ skew-sym

- $\mathcal{V}_m \exp(t\mathcal{H}_m) E_1$ orthonormal columns

- $\mathcal{V}_m \mathcal{R}_\nu(t\mathcal{H}_m) E_1$ orthonormal columns (Padé approx)

Further generalizations. I

A skew-symmetric and Hamiltonian

- $\exp(tA)$ ortho-symplectic - w.r.to $J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$

- $Q^{(0)}$ ortho-symplectic then $\exp(tA)Q^{(0)}$ ortho-symplectic

---

Block Krylov approximation:

- Choose *some* of the columns $\widetilde{Q}^{(0)}$ of $Q^{(0)}$,

$$V = \begin{pmatrix} \widetilde{Q}_1^{(0)} & \widetilde{Q}_2^{(0)} \\ \widetilde{Q}_2^{(0)} & -\widetilde{Q}_1^{(0)} \end{pmatrix} \qquad \mathcal{K}_m(A, V)$$

- $\mathcal{V}_m \exp(t\mathcal{H}_m)E_1$ columns of an ortho-symplectic matrix

Further generalizations. II

A Hamiltonian: $Q^{(0)}$ symplectic then $\exp(A)Q^{(0)}$ symplectic

---

Construct symplectic basis $\mathcal{V}_m$ and (logically) Hamiltonian $\mathcal{H}_m$:

Block Lanczos procedure in the block J-inner product:

$$[X, Y]_J = J_2^T X J Y \qquad X, Y \in \mathbb{R}^{2n\times 2}$$

$$J_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

An example

Linear Hamiltonian system:
$$\begin{cases} q' = Aq \\ q(0) = q_0 \end{cases} \qquad A = J^{-1}S$$

with $S \in \mathbb{R}^{400\times 400}$ symmetric (eigs. in [1, 100])

Note. Energy function: $E(q(t)) = q(t)^T S q(t)$, constant for all $t > 0$

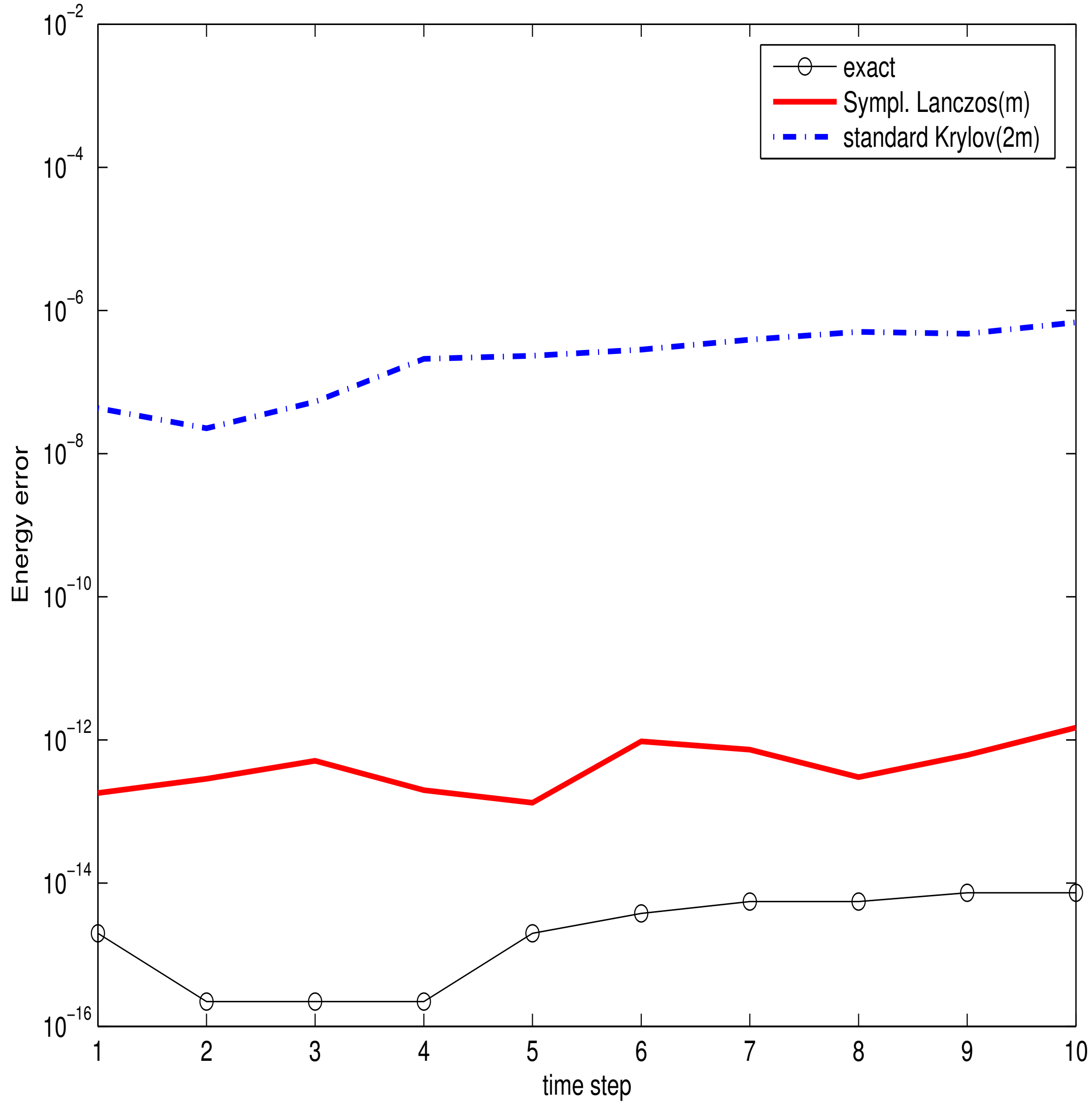
Numerical symplectic integrator: starting with $q^{(0)} = q_0$,

$$q^{(r+1)} = \exp(hA)q^{(r)}, \qquad r \geq 0 \qquad h = \frac{1}{40}$$

⋆ $x_m = \exp(hA)q^{(r)}$ standard Krylov subspace approximation

$\Rightarrow$ energy function is **not** constant, unless x_m is accurate

Conservation of energy. Error: $|E(q^{(r)}) - E(q_0)|$



Conclusions and Outlook

- ⋆ Rational function approximation is insightful framework
- ⋆ Appropriate variants allow structure preservation

- Natural generalizations (A nonsymmetric, other functions, etc.)
- Acceleration procedures