The Newton-ADI Method for Large-Scale Algebraic Riccati Equations

Peter Benner

Mathematik in Industrie und Technik
Fakultät für Mathematik

TECHNISCHE UNIVERSITÄT
CHEMNITZ

Sonderforschungsbereich 393

NUMERISCHE
PARALLELE
SIMULATION
SFB 393

benner@mathematik.tu-chemnitz.de

Workshop über Matrixgleichungen
MPI Leipzig, 15. April 2005

Outline

- Motivation
- Large-scale algebraic Riccati equations
- The Newton-ADI method
- Numerical results
- Conclusions and open problems

Motivation

Numerical solution of lq optimal control problem for parabolic systems via Galerkin approach, spatial FEM discretization $\rightsquigarrow$ finite-dim. LQR problem

Minimize $\mathcal{J}(u) \;=\; \frac{1}{2}\int\limits_0^\infty \left(y^T Q y + u^T R u\right) dt$ for $u \in \mathcal{L}_2(0,\infty;\mathbb{R}^m)$,
where $M\dot{x} = -Lx + Bu, \quad x(0) = x_0, \quad y = Cx$,
with stiffness matrix $L \in \mathbb{R}^{n\times n}$, mass matrix $M \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$.

Solution of finite-dimensional LQR problem:

$$u_*(t) = -B^T P_* x(t) =: -K_* x(t),$$

where $P_* \geq 0$ is stabilizing solution of the algebraic Riccati equation (ARE)

$$0 = \mathcal{R}(P) := C^T C + A^T P + PA - PBB^T P,$$

with $A := -M^{-1}L$, $B := M^{-1}BR^{-\frac{1}{2}}$, $C := CQ^{-\frac{1}{2}}$.

Large-Scale Algebraic Riccati Equations

General form for $A, G = G^T, Q = Q^T \in \mathbb{R}^{n \times n}$ given and $P \in \mathbb{R}^{n \times n}$ unknown:

$$0 = \mathcal{R}(P) := Q + A^T P + PA - PGP$$

Here, control-theoretic assumptions ensure existence of unique stabilizing solution $P_* = P_*^T \geq 0$, i.e.,

$$\Lambda\left(A - GP_*\right) \subset \mathbb{C}^-.$$

In large scale applications from semi-discretized control problems for PDEs,

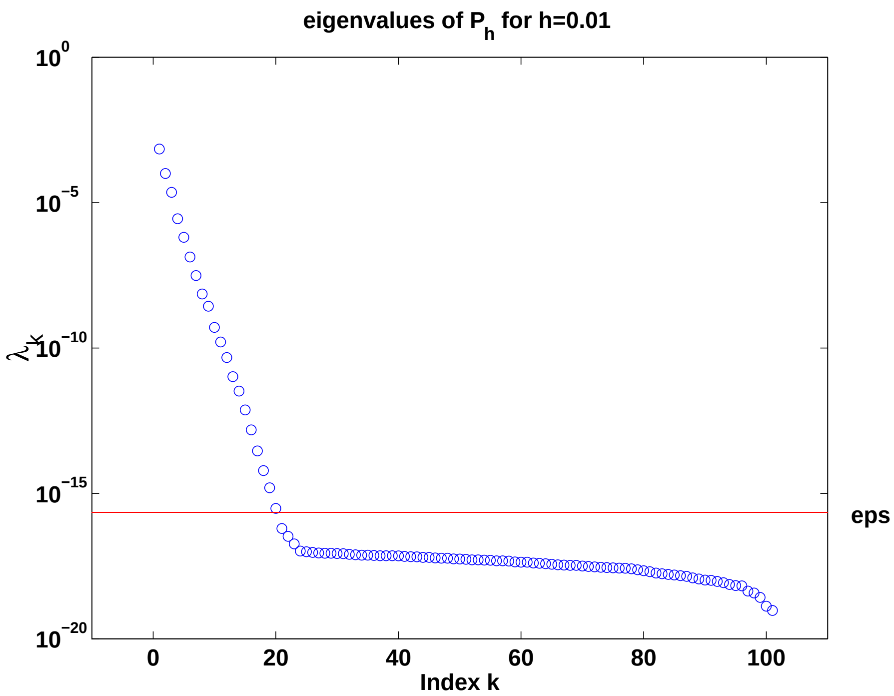
- $n = 10^3$ – 10^6 ($\Longrightarrow$ 10^6 – 10^{12} unknowns!),
- A has sparse representation ($A = -M^{-1}L$),
- G, Q low-rank with
  - $G = BB^T, \quad B \in \mathbb{R}^{n \times m}, \quad m \ll n$,
  - $Q = C^T C, \quad C \in \mathbb{R}^{p \times n}, \quad p \ll n$.
- Standard (eigenproblem-based) $\mathcal{O}(n^3)$ methods are not applicable!

Low-Rank Approximation

Consider spectrum of ARE solution.

Example:

- Linear 1D heat equation with point control,
- $\Omega = [\,0,\ 1\,]$,
- FEM discretization using linear B-splines,
- $h = 1/100 \Longrightarrow n = 101.$



Idea:

$$P = P^T \geq 0 \quad \Longrightarrow \quad P = ZZ^T = \sum_{k=1}^{n} \lambda_k z_k z_k^T \approx Z^{(r)}(Z^{(r)})^T = \sum_{k=1}^{r} \lambda_k z_k z_k^T .$$

Low-Rank Krylov Subspace Methods

Block-Arnoldi method

[Jaimoukha/Kasenally '94]

Consider $0 = \mathcal{R}(P) = CC^T + AP + PA^T - PBB^TP.$

1. Apply (block-)Arnoldi process to A with start (block-)vector C to generate the Krylov space

$$\mathcal{K}_\ell(A, C) = \text{span}\{C, AC, A^2C, \ldots, A^{\ell-1}C\}$$

with orthogonal basis V_ℓ such that

$$AV_\ell = V_\ell A_\ell + W_{\ell+1}A_{\ell+1,\ell}\begin{bmatrix} 0 \\ I_p \end{bmatrix}$$

and $A_\ell = V_\ell^T A V_\ell$ is block upper-Hessenberg.

2. Set $B_\ell := V_\ell^T B$, $C_\ell := V_\ell^T C$.
3. Find stabilizing solution of the ARE

$$0 = \mathcal{R}_\ell(X_\ell) = C_\ell C_\ell^T + A_\ell X_\ell + X_\ell A_\ell^T - X_\ell B_\ell B_\ell^T X_\ell.$$

4. Set $P_\ell := V_\ell X_\ell V_\ell^T$.

Properties:

+ P_ℓ satisfies Galerkin-type condition $\quad V_\ell^T \mathcal{R}(P_\ell) V_\ell = 0$
+ Computable residual error norm

$$\|\mathcal{R}(P_\ell)\|_F = \sqrt{2} \cdot \|A_{\ell+1,\ell} \begin{bmatrix} 0 \\ I_p \end{bmatrix} X_\ell\|_F.$$

– **Block**-Arnoldi, i.e., each step needs p matrix-vector products.

– Stabilizing X_ℓ may not exist as corresponding Hamiltonian matrix

$$H_\ell \;:=\; \begin{bmatrix} A_\ell^T & B_\ell B_\ell^T \\ C_\ell C_\ell^T & -A_\ell \end{bmatrix}$$

may have purely imaginary eigenvalues!

– No stabilization guarantee for P_ℓ!

– No convergence results for residuals.

Hamiltonian Lanczos algorithm

[*Freund/Mehrmann '92, Ferng/Lin/Wang '95, B./Faßbender '95*]

Consider

$$0 = \mathcal{R}(P) = Q + A^T P + PA - PGP.$$

1. Apply symplectic Lanczos method to Hamiltonian matrix $\begin{bmatrix} A & G \\ Q & -A^T \end{bmatrix}$ to generate Krylov space

   $$\mathcal{K}_{2\ell}(H, v_1) = \text{span}\{v_1, Hv_1, H^2 v_1, \ldots, H^{2\ell-1} v_1\},$$

   with symplectic basis

   $$S_\ell = [v_1, w_1, \ldots, v_\ell, w_\ell] \quad \in \quad \mathbb{R}^{2n,2\ell},$$
   $$S_\ell^T \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix} S_\ell \quad = \quad \begin{bmatrix} 0 & I_\ell \\ -I_\ell & 0 \end{bmatrix}$$

   such that

   $$HS_\ell \quad = \quad S_\ell H_\ell + \zeta_{\ell+1} v_{\ell+1} e_{2\ell}^T.$$

2. Compute ARE solution corresponding to H_ℓ, prolongate to $\mathbb{R}^{n \times n}$.

Properties:

+ P_ℓ satisfies Galerkin-type condition

$$V_\ell^T \mathcal{R}(P_\ell) V_\ell = 0$$

– $P_\ell \neq P_\ell^T$ for $\ell < n$.

+ In general less matrix-vector products than for block-Arnoldi.

+ Purely imaginary eigenvalues of small Hamiltonian matrix H_ℓ can be removed by cheap implicit restarts, i.e., can always get stable H_ℓ-invariant subspace.

+ Stabilization property for projected feedback matrix

$$V_\ell^T (A - GP_\ell) V_\ell$$

for sufficiently small Lanczos residual (can be achieved by implicit restarts).

– No stabilization guarantee for P_ℓ.

– No convergence results for residuals.

Related work:

- Two-sided structured Arnoldi method for Hamiltonian matrices based on symplectic URV decomposition/product eigenvalue problem.

  [*Xu '97, Kreßner '04*]

- Improved generation of low-rank approximate solution from r-dimensional ($r \ll n$) H-invariant subspace range (W) via

$$W_1 := W(1:n,:), \qquad W_2 = W(n+1:2n,:)$$
$$[U, \Sigma, V] = \mathrm{svd}(W_1^T W_2),$$
$$\tilde{X} = W_2 U S^{-\frac{1}{2}}.$$

  [*B./Mehrmann/Sorensen '03, Amodei '03*]

- Variants of Arnoldi using Frobenius inner product (global Arnoldi). [*Jbilou '03*]

Newton's Method for AREs

– Consider $0 = \mathcal{R}(P) = C^TC + A^TP + PA - PBB^TP$.

Frechét derivative of $\mathcal{R}(P)$ at P: $\mathcal{R}'_P : Z \to (A - BB^TP)^TZ + Z(A - BB^TP)$.

– Newton-Kantorovich method: $P_{j+1} = P_j - \left(\mathcal{R}'_{P_j}\right)^{-1} \mathcal{R}(P_j), \quad j = 0, 1, 2, \ldots$

$\Longrightarrow$ Newton's method (with line search) for AREs (for given $P_0 = P_0^T$ stabilizing):

FOR $j = 0, 1, \ldots$

1. $A_j \leftarrow A - BB^TP_j =: A - BK_j$.
2. Solve the Lyapunov equation $A_j^TN_j + N_jA_j = -\mathcal{R}(P_j)$.
3. $P_{j+1} \leftarrow P_j + t_jN_j$.

END FOR j

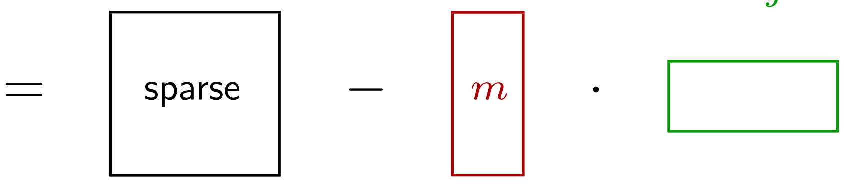
[*Kleinman '68, Mehrmann '91, Lancaster/Rodman '95, B./Byers '94/'98, B. '97, Guo/Laub '99*]

Properties and Implementation

- Convergence for $\Lambda\,(A - BK_0)$ stabilizing:
  - $A_j = A - BK_j = A - BB^T P_j$ is stable $\forall\ j \geq 1$.
  - $\lim_{j\to\infty} \|\mathcal{R}(P_j)\|_F = 0$ (monotonically).
  - $\lim_{j\to\infty} P_j = P_* \geq 0$ (locally quadratic).

- Need large-scale Lyapunov solver; here, ADI iteration [*Wachspress '88*]:
  $P_j = \lim_{k\to\infty} Q_k^{(j)}$, obtain $Q_k^{(j)}$ by solving linear system coefficient matrix A_j:

$$A_j \quad = \quad A \quad - \quad B \quad \cdot \quad K_j$$



  $m \ll n \implies$ efficient "inversion" using Sherman-Morrison-Woodbury formula:

$$(A - BK_j)^{-1} = (I_n + A^{-1}B(I_m - K_j A^{-1}B)^{-1}K_j)A^{-1}.$$

- BUT: $P = P^T \in \mathbb{R}^{n\times n} \implies n(n+1)/2$ unknowns!

Low-rank Newton-ADI for AREs

Re-write Newton's method for AREs [*Kleinman '68*]

$$A_j^T N_j + N_j A_j = -\mathcal{R}(P_j)$$

$$\Longleftrightarrow$$

$$A_j^T \underbrace{(P_j + N_j)}_{=P_{j+1}} + \underbrace{(P_j + N_j)}_{=P_{j+1}} A_j = \underbrace{-C^T C - P_j B B^T P_j}_{=:-W_j W_j^T}$$

Set $P_j = Z_j Z_j^T$ for $\operatorname{rank}(Z_j) \ll n \Longrightarrow$

$$A_j^T \left(Z_{j+1} Z_{j+1}^T \right) + \left(Z_{j+1} Z_{j+1}^T \right) A_j = -W_j W_j^T$$

$$\Downarrow$$

Solve Lyapunov equations for Z_{j+1} directly by factored ADI iteration and use *'sparse + low-rank'* structure of A_j. [*B./Li/Penzl '99–∞*]

ADI Method for Lyapunov Equations

- For $F \in \mathbb{R}^{n \times n}$ stable, $W \in \mathbb{R}^{n \times w}$ $(w \ll n)$, consider Lyapunov equation

$$F^T X + XF = -BB^T.$$

- ADI Iteration: [*Wachspress '88*]

$$\begin{aligned} (F^T + p_k I) X_{(j-1)/2} &= -BB^T - X_{k-1}(F - p_k I) \\ (F^T + \overline{p_k} I) {X_k}^T &= -BB^T - X_{(j-1)/2}(F - \overline{p_k} I) \end{aligned}$$

  with parameters $p_k \in \mathbb{C}^-$ and $p_{k+1} = \overline{p_k}$ if $p_k \notin \mathbb{R}$.

- For $X_0 = 0$ and proper choice of p_k: $\lim_{k \to \infty} X_k = X$ superlinear.

- Re-formulation using $X_k = Y_k Y_k^T$ yields iteration for Y_k...

Factored ADI Iteration

[Penzl '97, Li/Wang/White '99, B./Li/Penzl]

Set $X_k = Y_k Y_k^T$, some algebraic manipulations $\Longrightarrow$

$$V_1 \leftarrow \sqrt{-2\mathrm{Re}\,(p_1)}(F^T + p_1 I)^{-1}B, \qquad Y_1 \leftarrow V_1$$

FOR $j = 2, 3, \ldots$

$$V_k \leftarrow \sqrt{\frac{\mathrm{Re}\,(p_k)}{\mathrm{Re}\,(p_{k-1})}}\left(I - (p_k + \overline{p_{k-1}})(F^T + p_k I)^{-1}\right)V_{k-1}, \qquad Y_k \leftarrow \left[\begin{array}{cc} Y_{k-1} & V_k \end{array}\right]$$

$$\Downarrow$$

$$Y_{k_{\max}} = \left[\begin{array}{ccc} V_1 & \ldots & V_{k_{\max}} \end{array}\right]$$

where

$$V_k = \big[\;\big] \in \mathbb{C}^{n \times w}$$

and

$$Y_{k_{\max}} Y_{k_{\max}}^T \approx X$$

Note: Implementation in real arithmetic possible by combining two steps.

Direct Feedback Iteration

LQR problem: compute feedback matrix directly!

- jth Newton iteration:

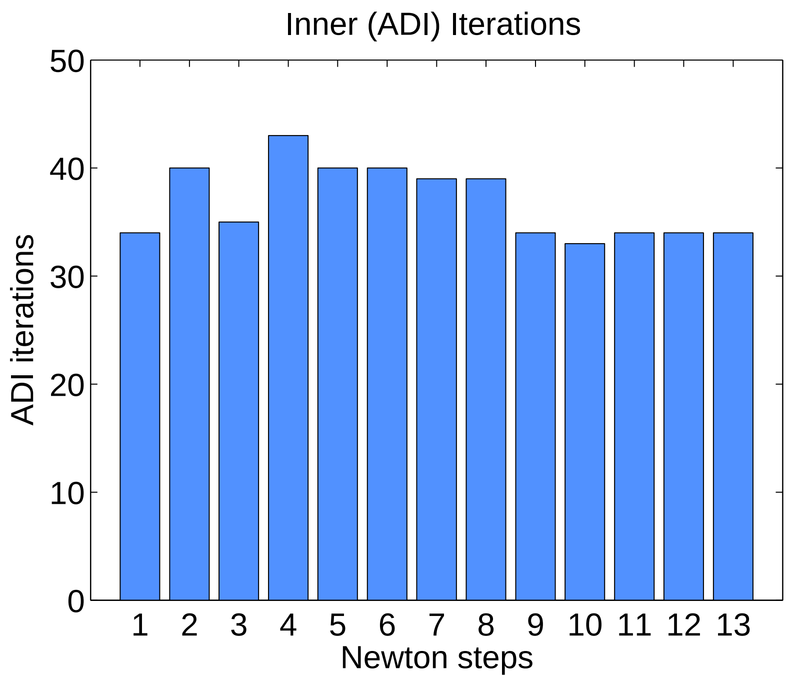
$$-K_j = B^T Z_j Z_j^T = \sum_{k=1}^{k_{\max}} (B^T V_{j,k}) V_{j,k}^T \quad \xrightarrow{j\rightarrow\infty} \quad -K_* = B^T Z_* Z_*^T$$

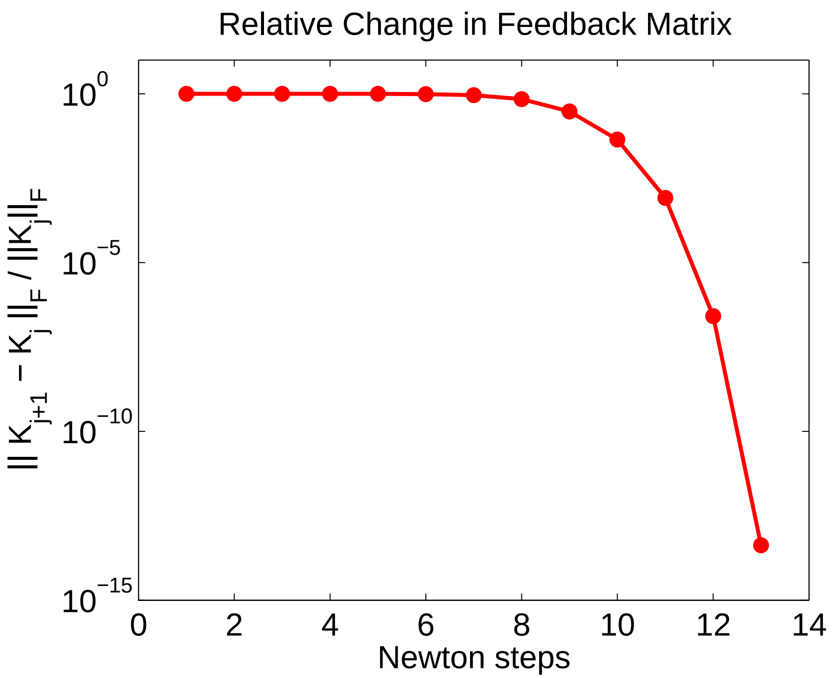
- K_j can be updated in ADI iteration, no need to even form Z_j, need only fixed workspace for $K_j \in \mathbb{R}^{m \times n}$!

$$\text{Cost} \quad \sim \quad (\#\ \text{Newton iterations}) \cdot \text{mean}\ (\#\ \text{ADI iterations}) \cdot (\text{cost for solving the elliptic problem})$$

Numerical Results

- Linear 2D heat equation with homogeneous Dirichlet boundary and distributed control/observation.
- FD discretization on uniform 150×150 grid.
- $n = 22.500$, $m = p = 1$, 10 shifts for ADI iterations.





Conclusions and Open Problems

- Solution of LQR problems for parabolic PDEs via low-rank factor Newton-ADI method is efficient and reliable.
- Riccati-approach applicable to many control problems for linear evolution equations.
- Open Problems:
  - Efficient stopping criteria.
  - Efficient implementation of line search strategy for large-sparse AREs.
  - Newton-ADI method for H_∞ control and other problems with indefinite Hessian possible?