Nonlinear Multigrid for Algebraic Riccati Equations

(with Wolfgang Hackbusch)

Lars Grasedyck

Max-Planck-Institute for
Mathematics in the Sciences
Leipzig

Outline

1. The discrete problem

$$
\begin{aligned}
\text{(Sylvester)} \quad & AX - XB + G = 0, \\
\text{(Riccati)} \quad & AX + XA^T - XFX + G = 0, \quad A, F, G \in \mathbb{R}^{N\times N}
\end{aligned}
$$

2. Structure of the solution X
   - Low rank structure $\rightarrow \mathcal{O}(N)$ data
   - $\mathcal{H}$-matrix structure $\rightarrow \mathcal{O}(N \log N)$ data
3. Multigrid Method
   - Requires: Grid hierarchy + suitable system
   - Sylvester: Linear Multigrid
   - Riccati: Nonlinear Multigrid / Newton
4. Numerical results for systems of size $N = 4190209$
   - Lyapunov
   - Riccati

Model Problem: Control of Heat Flow

Continuous model: minimise $J(u) = \int_0^\infty u(t)^2 + y(t)^2$,

$$\frac{\partial}{\partial_t} x(t) = \Delta x(t) + \mathcal{B}u(t) \qquad y(t) = \mathcal{C}x(t)$$

$$x : [0, \infty) \to H_0^1(\Omega), \quad \Omega \subset \mathbb{R}^d$$

Semidiscrete model (FEM, FD)

$$\dot{x} = Ax + Bu, \quad x(t) \in V_n, \qquad y = Cx$$

Optimal feedback control is $u = -B^T X x$, $\qquad X$ solves

$$AX + XA^T - XBB^T X + C^T C = 0 \qquad \text{(Riccati)}$$

Linearisation (Newton's method) yields equations $i = 1, \ldots$

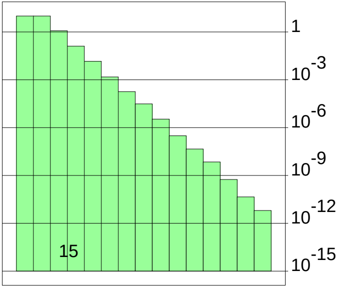
$$A_i X + X A_i^T + G_i = 0 \qquad \text{(Lyapunov)}$$

In the first step $A_1 = A$, otherwise

$$A_i = A - X_{i-1} B B^T.$$

Existence of Low Rank Solutions X

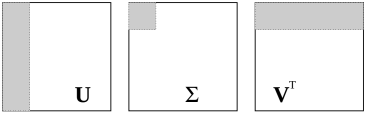
Singular values of X

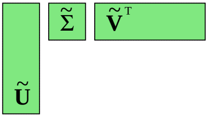


- $AX + XA + G = 0$
- $\mathrm{rank}(G) = 1$, $A < 0$
- singular values decay rapidly
- most singular values ≈ 0
- low rank approximation

$$\underbrace{X}_{N\times N} = U\cdot\Sigma\cdot V^T$$

$$\tilde{X} = \underbrace{\tilde{U}}_{N\times k}\cdot\underbrace{\tilde{\Sigma}}_{k\times k}\cdot\underbrace{\tilde{V}^T}_{k\times N}$$





EXISTENCE OF LOW RANK SOLUTIONS X

Diagonal case: $DX + XD \equiv 1$, $D > 0$

$$X_{ij} = 1/(d_i + d_j)$$

- Taylor expansion in d_i at $\bar{d} := \frac{1}{2}(\min_\ell d_\ell + \max_\ell d_\ell)$:

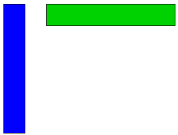
$$1/(d_i + d_j) = \sum_{\nu=1}^{\infty} (\bar{d} - d_i)^{\nu-1} (\bar{d} + d_j)^{-\nu-2}$$

$$\bar{d} + d_j \geq \bar{d} + d_{min}, \qquad |\bar{d} - d_i| \leq \bar{d} - d_{min}$$

$$\left| \frac{1}{d_i + d_j} - \sum_{\nu=1}^{k} u_{i\nu} v_{j\nu} \right| \leq \sum_{\nu=k}^{\infty} \left(\frac{\bar{d} - d_{min}}{\bar{d} + d_{min}} \right)^{\nu} / (\bar{d} + d_{min}) \leq \eta^k, \quad \eta < 1$$

$$X_{ij} \approx \sum_{\nu=1}^{k} u_{i\nu} v_{j\nu} =: \tilde{X}_{ij}$$

$$|X_{ij} - \tilde{X}_{ij}| \leq \eta^k$$

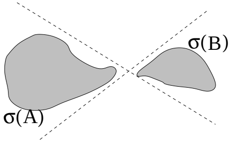


Existence of Low Rank Solutions X

Theorem [G. '01,'02] Low Rank case

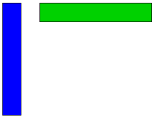
Assumptions:

- X solves $AX - XB + G = 0$
- $\text{conv}(\sigma(A)) \cap \text{conv}(\sigma(B)) = \emptyset$



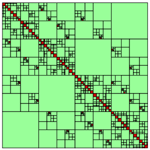
Result: for each ε exists $\tilde{X}$ s.t.

- $\|X - \tilde{X}\| \leq \varepsilon \|X\|$
- $\text{rank}(\tilde{X}) = \mathcal{O}(\log^2(\varepsilon)\text{rank}(G))$



Generalisation

- up to $\mathcal{O}(1)$ eigenvalues arbitrary
- also for Riccati $AX - XB - XFX + G = 0$
- blockwise low rank matrices
  (G., Hackbusch,Khoromskij '03, Computing 70)



Related Results: Symmetric Lyapunov: Penzl '99
Computable bounds: Antoulas,Sorensen,Zhou '02

First Summary

Theory:

- We want to solve Riccati/Sylvester equations

$$AX + XA^T - XFX + G = 0$$

$$AX - XB + G = 0$$

- Solution matrix

$$X \in \mathbb{R}^{N \times N}$$

- We can represent an approximate solution

$$\tilde{X} \in \mathbb{R}^{N \times N} \qquad \text{with } \mathcal{O}(N \log^c \varepsilon) \text{ data}$$

$$\|X - \tilde{X}\| \leq \varepsilon \|X\|$$

Practice:

- How do we compute X ?
- How do we compute $\tilde{X}$ without forming X ?
- Recall: Can easily solve diagonal Sylvester equations

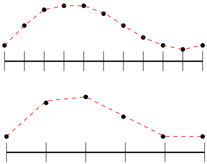
Introduction: Multigrid

Sylvester eq. $AX - XB + G = 0$ is $N^2 \times N^2$ linear system $Mx + g = 0$

$$M := (I \otimes A - B \otimes I), \quad x := \mathrm{vec}(X), \quad g := \mathrm{vec}(G)$$

- Multilevel discretisation $(M_\ell x_\ell + g_\ell = 0)_{\ell=1}^{L}$
- Want to solve $M_L x_L + g_L = 0$
- Approximate solution $\tilde{x}_\ell$, error $e_\ell := x_\ell - \tilde{x}_\ell$, defect $d_\ell := M_\ell \tilde{x}_\ell + g_\ell$

$$M_\ell e_\ell + d_\ell = 0 \qquad \text{(defect equation)}$$



If e_ℓ, d_ℓ smooth: $\rightarrow$ coarse grid approx.

$$M_{\ell-1} e_{\ell-1} + d_{\ell-1} = 0$$

- Need transfer operators

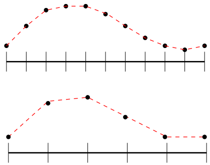
$$R_{\ell-1\leftarrow\ell}: \text{ level } \ell \textit{ to } \text{level } \ell-1, \quad P_{\ell\leftarrow\ell-1}: \text{ level } \ell-1 \textit{ to } \text{level } \ell$$

- Need solver (“smoother”) on level ℓ so that e_ℓ, d_ℓ smooth (not small!)

MULTIGRID SMOOTHER FOR $Mx + g = 0$

Approximation $\tilde{x}$,
error $e := x - \tilde{x}$,
defect $d := M\tilde{x} + g$

$$Me + d = 0 \qquad \text{(defect equation)}$$



Goal: e, d smooth

The two simplest smoothers:

JACOBI:

$$x^{i+1} := x^i - \mathrm{diag}(M)^{-1}(Mx^i + g)$$

RICHARDSON:

$$x^{i+1} := x^i - \theta(Mx^i + g)$$

Convergence: very slow
Smoothing: very good

Multigrid for $Mx + g = 0$

Procedure multigrid_step(M_ℓ, g_ℓ, **var** x_ℓ)
if $\ell = 1$ **then**
  Solve $M_\ell x_\ell + g_\ell = 0$ {solve coarse grid}
else
  $x_\ell := S_\ell(x_\ell, g_\ell)$ {presmooth}
  $g_{\ell-1} := R_{\ell-1\leftarrow\ell}(M_\ell x_\ell + g_\ell)$ {restrict defect to coarse grid}
  $x_{\ell-1} := 0$
  Call **multigrid_step**($M_{\ell-1}$, $g_{\ell-1}$, $x_{\ell-1}$) {recursive call}
  $x_\ell := x_\ell + P_{\ell\leftarrow\ell-1}(x_{\ell-1})$ {coarse grid correction}
  $x_\ell := S_\ell(x_\ell, g_\ell)$ {postsmooth}
end if

Multigrid Convergence

System to solve:

$$M_\ell x_\ell + g_\ell = 0$$

Linear iteration

$$x_\ell{}^{i+1} = x_\ell{}^{i} - W_\ell (M_\ell x_\ell{}^{i} + g_\ell)$$

has iteration matrix $I - W_\ell M_\ell$, convergence rate $\rho(I - W_\ell M_\ell) \overset{!}{<} 1$.

Multigrid Convergence

System to solve:

$$M_\ell x_\ell + g_\ell = 0$$

Linear iteration

$$x_\ell{}^{i+1} = x_\ell{}^i - W_\ell(M_\ell x_\ell{}^i + g_\ell)$$

has iteration matrix $I - W_\ell M_\ell$, convergence rate $\rho(I - W_\ell M_\ell) \overset{!}{<} 1$.

Smoothing iteration:

$$\begin{aligned} \text{(Richardson)} \quad & x_\ell{}^{i+1} = x_\ell{}^i - \theta(M_\ell x_\ell{}^i + g_\ell) \\ \text{(Jacobi)} \quad & x_\ell{}^{i+1} = x_\ell{}^i - \operatorname{diag}(M_\ell)^{-1}(M_\ell x_\ell{}^i + g_\ell) \end{aligned}$$

Iteration matrix of smoothers:

Richardson: $S_\ell := I - \theta M_\ell$, Jacobi: $S_\ell := I - \operatorname{diag}(M_\ell)^{-1} M_\ell$.

Multigrid Convergence

System to solve:

$$M_\ell x_\ell + g_\ell = 0$$

Linear iteration

$$x_\ell{}^{i+1} = x_\ell{}^{i} - W_\ell (M_\ell x_\ell{}^{i} + g_\ell)$$

has iteration matrix $I - W_\ell M_\ell$, convergence rate $\rho(I - W_\ell M_\ell) \overset{!}{<} 1$.

Richardson: $S_\ell := I - \theta M_\ell$, Jacobi: $S_\ell := I - \mathrm{diag}(M_\ell)^{-1} M_\ell$.

Multigrid Convergence

System to solve:

$$M_\ell x_\ell + g_\ell = 0$$

Linear iteration

$$x_\ell{}^{i+1} = x_\ell{}^i - W_\ell(M_\ell x_\ell{}^i + g_\ell)$$

has iteration matrix $I - W_\ell M_\ell$, convergence rate $\rho(I - W_\ell M_\ell) \overset{!}{<} 1$.

Richardson: $S_\ell := I - \theta M_\ell$, Jacobi: $S_\ell := I - \mathrm{diag}(M_\ell)^{-1} M_\ell$.

Coarse grid correction:

$$x_\ell{}^{i+1} = x_\ell{}^i - P_{\ell\leftarrow\ell-1} M_{\ell-1}^{-1} R_{\ell-1\leftarrow\ell}(M_\ell x_\ell{}^i + g_\ell)$$

Iteration matrix of coarse grid correction:

$$M_{cg} := I - P_{\ell\leftarrow\ell-1} M_{\ell-1}^{-1} R_{\ell-1\leftarrow\ell} M_\ell$$

Multigrid Convergence

System to solve:

$$M_\ell x_\ell + g_\ell = 0$$

Linear iteration

$$x_\ell{}^{i+1} = x_\ell{}^{i} - W_\ell(M_\ell x_\ell{}^{i} + g_\ell)$$

has iteration matrix $I - W_\ell M_\ell$, convergence rate $\rho(I - W_\ell M_\ell) \overset{!}{<} 1$.

Richardson: $S_\ell := I - \theta M_\ell$, Jacobi: $S_\ell := I - \mathrm{diag}(M_\ell)^{-1} M_\ell$.

$$M_{cg} := I - P_{\ell \leftarrow \ell-1} M_{\ell-1}^{-1} R_{\ell-1 \leftarrow \ell} M_\ell$$

Multigrid Convergence

System to solve:

$$M_\ell x_\ell + g_\ell = 0$$

Linear iteration

$$x_\ell{}^{i+1} = x_\ell{}^i - W_\ell (M_\ell x_\ell{}^i + g_\ell)$$

has iteration matrix $I - W_\ell M_\ell$, convergence rate $\rho(I - W_\ell M_\ell) \overset{!}{<} 1$.

Richardson: $S_\ell := I - \theta M_\ell$, Jacobi: $S_\ell := I - \mathrm{diag}(M_\ell)^{-1} M_\ell$.

$$M_{cg} := I - P_{\ell \leftarrow \ell-1} M_{\ell-1}^{-1} R_{\ell-1 \leftarrow \ell} M_\ell$$

Iteration matrix of twogrid method with ν smoothing steps:

$$\begin{aligned} M_{cg} S_\ell^\nu &= (I - P_{\ell \leftarrow \ell-1} M_{\ell-1}^{-1} R_{\ell-1 \leftarrow \ell} M_\ell) S_\ell^\nu \\ &= \left(M_\ell^{-1} - P_{\ell \leftarrow \ell-1} M_{\ell-1}^{-1} R_{\ell-1 \leftarrow \ell} M_\ell \right) (M_\ell S_\ell^\nu) \end{aligned}$$

Multigrid Convergence

System to solve:

$$M_\ell x_\ell + g_\ell = 0$$

Linear iteration

$$x_\ell{}^{i+1} = x_\ell{}^{i} - W_\ell(M_\ell x_\ell{}^{i} + g_\ell)$$

has iteration matrix $I - W_\ell M_\ell$, convergence rate $\rho(I - W_\ell M_\ell) \overset{!}{<} 1$.

Richardson: $S_\ell := I - \theta M_\ell$, Jacobi: $S_\ell := I - \mathrm{diag}(M_\ell)^{-1} M_\ell$.

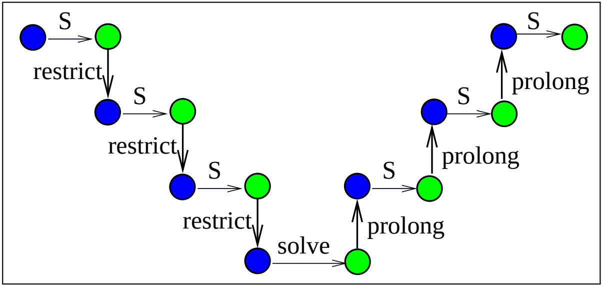
$$M_{cg} := I - P_{\ell\leftarrow\ell-1} M_{\ell-1}^{-1} R_{\ell-1\leftarrow\ell} M_\ell$$

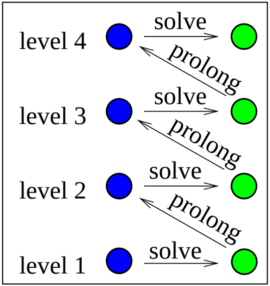
Iteration matrix of twogrid method with ν smoothing steps:

$$M_{cg} S_\ell^\nu = \left(M_\ell^{-1} - P_{\ell\leftarrow\ell-1} M_{\ell-1}^{-1} R_{\ell-1\leftarrow\ell} M_\ell\right)(M_\ell S_\ell^\nu)$$

- Approx. property: $\left\|M_\ell^{-1} - P_{\ell\leftarrow\ell-1} M_{\ell-1}^{-1} R_{\ell-1\leftarrow\ell} M_\ell\right\|_2 < C_A / \|M_\ell\|_2$
- Smoothing property: $\|M_\ell S_\ell^\nu\|_2 < \eta(\nu)\|M_\ell\|_2$
- Together: $\|M_{cg} S_\ell^\nu\|_2 \le C_A \eta(\nu)$, $\lim_{\nu\to\infty} \eta(\nu) = 0$

Multigrid and Nested Iteration





LEFT: One multigrid cycle with smoother S

RIGHT: Nested iteration provides good starting values

Low Rank Multigrid

- Each iterate ${X_\ell}^i$ of low rank
- Have to compute

$$M_\ell {x_\ell}^i = A_\ell {X_\ell}^i + {X_\ell}^i A_\ell^T - {X_\ell}^i F_\ell {X_\ell}^i + G_\ell$$

- Prolongation/Restriction

$$P_{\ell \leftarrow \ell-1} {x_\ell}^i = p_{\ell \leftarrow \ell-1} {X_\ell}^i p^*_{\ell \leftarrow \ell-1}$$

- Truncation for rank reduction:

$$X = UV^T$$

1. $U = Q_U R_U, \quad V = Q_V R_V$
2. $R_U R_V^T = U_R \Sigma {V_R}^T$
3. $X = (Q_U U_R) \Sigma (Q_V V_R)^T$

Multigrid Convergence

What can we prove ?

1. One step of multigrid costs $\mathcal{O}(Nk^2)$, where X_ℓ of rank k
2. If A is the FEM discretisation of a d-dimensional elliptic PDE then $M = A \otimes E + E \otimes A$ is the FEM discretisation of a $2d$-dimensional elliptic PDE
3. Without truncation the convergence rate is
   - level-independent and
   - bounded from 1
4. We need $\mathcal{O}(1)$ multigrid steps on the fine level
5. We observe: Truncation does not influence the convergence, unless we are very close to the best approximation

Numerical Example (One processor, 900 MHz)

We solve

$$AX + XA + G = 0$$

where

- A: 2d-Laplacian on unit square with $N = 127^2, 255^2, 511^2$ dof
- G: rank 6
- Target accuracy: $\varepsilon = 3 \times 10^{-3}, 1 \times 10^{-3}, 3 \times 10^{-4}$
- Algorithm: Low Rank Multigrid
- Complexity: $\sim \mathcal{O}(N \log^2 N)$ (truncation)

Time in seconds for solve:

	$N = 127^2$ $\varepsilon = 0.003$	$N = 255^2$ $\varepsilon = 0.001$	$N = 511^2$ $\varepsilon = 0.0003$
Sec.	2.7	12.3	74.3

Remark Complexity is not optimal: $\mathcal{O}(Nk)$ data $\leftrightarrow$ $\mathcal{O}(Nk^2)$ time

NONLINEAR MULTIGRID

QUESTION What do we have to do for Riccati?

$$R(X) := AX + XA^T - XFX + G, \qquad R(X) \overset{!}{=} 0$$

ANSWER A Newton-Kleinman iteration + linear solves

$$\begin{aligned} X^{i+1} &:= L_{X^i}^{-1}(-G - X^i F X^i), \\ & \quad L_{X^i}(Y) := (A - X^i F)Y + Y(A - X^i F)^T \end{aligned}$$

ANSWER B

- Replace Jacobi iteration by

$$X^{i+1} := X^{i+1} - \theta R(X^i)$$

$\rightarrow$ nonlinear Richardson iteration

- Modify coarse grid correction (operator nonlinear)

Total Complexity $\mathcal{O}(Nk^2)$ as for Sylvester/Lyapunov

Nonlinear Multigrid

Theorem [G., '06]:

- $\|X - \tilde{X}\| \leq \varepsilon$,
- Y^i, $i = 1, \ldots,$ iterates of linear Richardson for linearised Riccati at $\tilde{X}$
- $\|Y^i - \tilde{X}\| \leq \varepsilon$
- X^i, $i = 1, \ldots,$ iterates of nonlinear Richardson for Riccati (start $\tilde{X}$)

Then

$$\|Y^i - X^i\| \leq \varepsilon^2$$

Conclusion :

- Nonlinear Richardson for Riccati behaves like linear Richardson for the linearisation
- Positivity is not preserved $\rightarrow$ projection

Numerical Example (One processor, 900 MHz)

We solve

$$AX + XA - XFX + G = 0$$

where

- A: 2d-Laplacian on unit square with $N = 511^2, 1023^2, 2047^2$ dof
- G, F: rank 1
- Target accuracy: discretisation error
- Algorithm: Low Rank Multigrid
- Complexity: $\sim \mathcal{O}(N \log^2 N)$ (truncation)

	$N = 511^2$ $k = 5$	$N = 1023^2$ $k = 6$	$N = 2047^2$ $k = 7$
Seconds	90	956	6376

Remark Complexity is not optimal: $\mathcal{O}(Nk)$ data $\leftrightarrow$ $\mathcal{O}(Nk^2)$ time

Outlook

$$AX + XA^T - XFX + G = 0$$

Speed: Optimal complexity solver $\mathcal{O}(N)$
$\longrightarrow$ Not possible in low rank format!

Generality: Algebraic multigrid/ robust smoothers
$\longrightarrow$ Bart Vandereycken, KU Leuven

$\mathcal{H}$: Non Low Rank solvers for general G

$$AX - XB + G = 0$$

Robustness: ADI as smoother

Speed: Acceleration by GMRES/cg

Thank You for Your Attention!

L. Grasedyck:
Existence of a low rank or $\mathcal{H}$-matrix approximant to the solution of a Sylvester equation,
Num.Lin.Alg.Appl. 11

L. Grasedyck, W. Hackbusch:
A multigrid method to solve large scale Sylvester equations,
accepted for **SIMAX**

L. Grasedyck:
Nonlinear multigrid for the solution of large scale Riccati equations,
accepted for **Num.Lin.Alg.Appl.**

Please visit `http://www.hmatrix.org`
`http://www.mis.mpg.de`