The Extended Krylov subspace method for large-scale Lyapunov equations

Davide Palitta

Dipartimento di Matematica, Università di Bologna
davide.palitta3@unibo.it

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Lyapunov equation: Preliminaries

Lyapunov equation

$$AX + XA^T + C = 0, \qquad A, C, X \in \mathbb{R}^{n\times n}$$

Lyapunov equation: Preliminaries

Lyapunov equation

$$AX + XA^T + C = 0, \qquad A, C, \mathbf{X} \in \mathbb{R}^{n\times n}$$

Well-posedness: $\Lambda(A) = \{\lambda_k\}_{k=1}^n$

$$\lambda_j + \lambda_i \neq 0, \quad \forall\, i, j = 1, \dots, n$$

Usually, A *stable* (from applications): $\mathrm{Re}\,(\lambda_k) < 0, \forall\, k = 1, \dots, n$

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Lyapunov equation: Preliminaries

Lyapunov equation

$$A\mathbf{X} + \mathbf{X}A^T + C = 0, \qquad A, C, \mathbf{X} \in \mathbb{R}^{n\times n}$$

Well-posedness: $\Lambda(A) = \{\lambda_k\}_{k=1}^n$

$$\lambda_j + \lambda_i \neq 0, \quad \forall\, i, j = 1, \ldots, n$$

Usually, A *stable* (from applications): $\mathrm{Re}\,(\lambda_k) < 0, \forall\, k = 1, \ldots, n$

Many others:

- Sylvester equation: $A\mathbf{X} + \mathbf{X}B + C = 0$
- Generalized Lyapunov equation: $A\mathbf{X}E^T + E\mathbf{X}A^T + C = 0$
- Generalized Sylvester equation: $A\mathbf{X}D + E\mathbf{X}B + C = 0$
- ...

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Large-scale linear matrix equation

Large-scale Lyapunov equation

$$A\mathbf{X} + \mathbf{X}A^T + bb^T = 0 \tag{1}$$

where $A \in \mathbb{R}^{n \times n}$ is (very) large, sparse, $A < 0$ and $b \in \mathbb{R}^n$

Large-scale linear matrix equation

Large-scale Lyapunov equation

$$A\mathbf{X} + \mathbf{X}A^T + bb^T = 0 \tag{1}$$

where $A \in \mathbb{R}^{n \times n}$ is (very) large, sparse, $A < 0$ and $b \in \mathbb{R}^n$

$\Rightarrow \exists! \ \mathbf{X} \in \mathbb{R}^{n \times n}$ solution to (1), $\mathbf{X} \geq 0$, $\mathbf{X} = \mathbf{X}^T$

Large-scale linear matrix equation

Large-scale Lyapunov equation

$$A\mathbf{X} + \mathbf{X}A^T + bb^T = 0 \tag{1}$$

where $A \in \mathbb{R}^{n \times n}$ is (very) large, sparse, $A < 0$ and $b \in \mathbb{R}^n$

$\Rightarrow \exists!\ \mathbf{X} \in \mathbb{R}^{n \times n}$ solution to (1), $\mathbf{X} \geq 0$, $\mathbf{X} = \mathbf{X}^T$

Remark 1: Even if A is sparse, $\mathbf{X}$ is, in general, dense

Large-scale linear matrix equation

Large-scale Lyapunov equation

$$A\mathbf{X} + \mathbf{X}A^T + bb^T = 0 \tag{1}$$

where $A \in \mathbb{R}^{n\times n}$ is (very) large, sparse, $A < 0$ and $b \in \mathbb{R}^n$

$\Rightarrow \exists!$ $\mathbf{X} \in \mathbb{R}^{n\times n}$ solution to (1), $\mathbf{X} \geq 0$, $\mathbf{X} = \mathbf{X}^T$

Remark 1: Even if A is sparse, $\mathbf{X}$ is, in general, dense

Remark 2: Under certain hypotheses on $\Lambda(A)$, $\mathbf{X}$ presents a fast eigenvalue decay
$\Rightarrow$ Look for a low-rank approximation: $\mathbf{X} \approx \widetilde{\mathbf{X}} = \mathbf{Z}\mathbf{Z}^T$, $\mathbf{Z}$ "tall"

Projection methods

- Look for $\widetilde{\mathbf{X}} = \mathbf{X}_m = V_m \mathbf{Y}_m V_m^T$ where $\mathcal{K}_m = \text{Range}(V_m)$, $V_m \in \mathbb{R}^{n\times m}$, $m \ll n$, w/ orthonormal columns, and $\mathbf{Y}_m \in \mathbb{R}^{m\times m}$

Projection methods

- Look for $\widetilde{\mathbf{X}} = \mathbf{X}_m = V_m \mathbf{Y}_m V_m^T$ where $\mathcal{K}_m = \text{Range}(V_m)$, $V_m \in \mathbb{R}^{n \times m}$, $m \ll n$, w/ orthonormal columns, and $\mathbf{Y}_m \in \mathbb{R}^{m \times m}$
- To compute $\mathbf{Y}_m$, impose a Galerkin condition on the residual matrix

$$R_m := A\mathbf{X}_m + \mathbf{X}_m A^T + BB^T \perp \mathcal{K}_m \iff V_m^T R_m V_m = 0 \qquad (2)$$

Projection methods

- Look for $\widetilde{\mathbf{X}} = \mathbf{X}_m = V_m \mathbf{Y}_m V_m^T$ where $\mathcal{K}_m = \text{Range}(V_m)$, $V_m \in \mathbb{R}^{n\times m}$, $m \ll n$, w/ orthonormal columns, and $\mathbf{Y}_m \in \mathbb{R}^{m\times m}$
- To compute $\mathbf{Y}_m$, impose a Galerkin condition on the residual matrix

$$R_m := A\mathbf{X}_m + \mathbf{X}_m A^T + BB^T \perp \mathcal{K}_m \iff V_m^T R_m V_m = 0 \qquad (2)$$

- Assuming $\text{Range}(V_1) = \text{Range}(b)$, (2) is equivalent to solving

$$T_m \mathbf{Y}_m + \mathbf{Y}_m T_m^T + \gamma^2 e_1 e_1^T = 0$$

$T_m := V_m^T A V_m$, $\mathbf{Y}_m = V_m^T \mathbf{X}_m V_m$
$V_m^T b = \gamma e_1$, $\gamma = \|b\|$, e_1 first column of I_m

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Projection methods

- Look for $\widetilde{\mathbf{X}} = \mathbf{X}_m = V_m \mathbf{Y}_m V_m^T$ where $\mathcal{K}_m = \text{Range}(V_m)$, $V_m \in \mathbb{R}^{n\times m}$, $m \ll n$, w/ orthonormal columns, and $\mathbf{Y}_m \in \mathbb{R}^{m\times m}$
- To compute $\mathbf{Y}_m$, impose a Galerkin condition on the residual matrix

$$R_m := A\mathbf{X}_m + \mathbf{X}_m A^T + BB^T \perp \mathcal{K}_m \iff V_m^T R_m V_m = 0 \quad (2)$$

- Assuming $\text{Range}(V_1) = \text{Range}(b)$, (2) is equivalent to solving

$$T_m \mathbf{Y}_m + \mathbf{Y}_m T_m^T + \gamma^2 e_1 e_1^T = 0$$

$T_m := V_m^T A V_m$, $\mathbf{Y}_m = V_m^T \mathbf{X}_m V_m$
$V_m^T b = \gamma e_1$, $\gamma = \|b\|$, e_1 first column of I_m

- Check the residual norm $\leftarrow$ compute it cheaply (no details here)

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Projection methods (II)

Is $T_m\mathbf{Y}_m + \mathbf{Y}_m T_m^T = \gamma^2 e_1 e_1^T$ well-posed?

Projection methods (II)

Is $T_m \mathbf{Y}_m + \mathbf{Y}_m T_m^T = \gamma^2 e_1 e_1^T$ well-posed? $A < 0 \Rightarrow T_m$ *stable*!

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Projection methods (II)

Is $T_m\mathbf{Y}_m + \mathbf{Y}_m T_m^T = \gamma^2 e_1 e_1^T$ well-posed? $A < 0 \Rightarrow T_m$ *stable*!

How to solve it?

Projection methods (II)

Is $T_m \mathbf{Y}_m + \mathbf{Y}_m T_m^T = \gamma^2 e_1 e_1^T$ well-posed? $A < 0 \Rightarrow T_m$ *stable*!

How to solve it? **Bartels-Stewart**

1. Compute $T_m = Q_m U_m Q_m^T$, $Q_m Q_m^T = I_m$, U_m upper triangular
2. Solve
$$U_m \tilde{\mathbf{Y}} + \tilde{\mathbf{Y}} U_m^T = \gamma^2 Q_m^T e_1 e_1^T Q_m, \quad \tilde{\mathbf{Y}} := Q_m^T \mathbf{Y}_m Q_m$$
by substitution
3. Set
$$\mathbf{Y}_m = Q_m \tilde{\mathbf{Y}} Q_m^T$$

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Projection methods (III)

At convergence

$$\mathbf{X}_m = V_m \mathbf{Y}_m V_m^T \text{ is never computed!}$$

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Projection methods (III)

At convergence

$$\mathbf{X}_m = V_m \mathbf{Y}_m V_m^T \quad \text{is never computed!}$$

Remark: $\mathbf{Y}_m$ solution to $T_m \mathbf{Y}_m + \mathbf{Y}_m T_m^T + \gamma^2 e_1 e_1^T = 0$

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Projection methods (III)

At convergence

$$\mathbf{X}_m = V_m \mathbf{Y}_m V_m^T \text{ is never computed!}$$

Remark: $\mathbf{Y}_m$ solution to $T_m \mathbf{Y}_m + \mathbf{Y}_m T_m^T + \gamma^2 e_1 e_1^T = 0$

$\Rightarrow$ $\mathbf{Y}_m \geq 0$, $\mathbf{Y}_m = \mathbf{Y}_m^T$ and possibly low-rank

Projection methods (III)

At convergence

$$\mathbf{X}_m = V_m \mathbf{Y}_m V_m^T \text{ is never computed!}$$

Remark: $\mathbf{Y}_m$ solution to $T_m \mathbf{Y}_m + \mathbf{Y}_m T_m^T + \gamma^2 e_1 e_1^T = 0$

$\Rightarrow$ $\mathbf{Y}_m \geq 0$, $\mathbf{Y}_m = \mathbf{Y}_m^T$ and possibly low-rank

- Low-rank factor of $\mathbf{Y}_m$: $\mathbf{Y}_m \approx \widehat{Y}\widehat{Y}^T$, $\widehat{Y} \in \mathbb{R}^{m \times t}$, $t \leq m$

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Projection methods (III)

At convergence

$$\mathbf{X}_m = V_m \mathbf{Y}_m V_m^T \ \text{ is never computed!}$$

Remark: $\mathbf{Y}_m$ solution to $T_m \mathbf{Y}_m + \mathbf{Y}_m T_m^T + \gamma^2 e_1 e_1^T = 0$

$\Rightarrow$ $\mathbf{Y}_m \geq 0$, $\mathbf{Y}_m = \mathbf{Y}_m^T$ and possibly low-rank

- Low-rank factor of $\mathbf{Y}_m$: $\mathbf{Y}_m \approx \widehat{Y}\widehat{Y}^T$, $\widehat{Y} \in \mathbb{R}^{m \times t}$, $t \leq m$
- Low-rank factor of $\mathbf{X}_m$:
$$\mathbf{X}_m = V_m \mathbf{Y}_m V_m^T \approx \left(V_m \widehat{Y}\right)\left(V_m \widehat{Y}\right)^T = \mathbf{Z}_m \mathbf{Z}_m^T,\ \mathbf{Z}_m = V_m \widehat{Y} \in \mathbb{R}^{n \times t}$$

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Projection methods (IV)

Algorithm : Galerkin projection method for the Lyapunov matrix equation
Input: $A \in \mathbb{R}^{n\times n}$, $A < 0$, $b \in \mathbb{R}^n$
Output: $\mathbf{Z}_m \in \mathbb{R}^{n\times t}$, $t \leq m$

1. Set $\beta = \|b\|_2$ and $\mathcal{V}_1 \equiv V_1 = b/\beta$
3. For $m = 2, 3, \dots,$ till convergence, **Do**
4. Compute next basis block $\mathcal{V}_m$ and set $V_m = [V_{m-1}, \mathcal{V}_m]$
5. Update $T_m = V_m^T A V_m$
6. Convergence check:
6.1 Solve $T_m \mathbf{Y}_m + \mathbf{Y}_m T_m^T + \gamma^2 e_1 e_1^T = 0, \quad e_1 \in \mathbb{R}^m$
6.2 Compute $\|R_m\|_F$
6.3 If $\|R_m\|_F/\beta^2$ is small enough **Stop**, otherwise **Continue**
7. EndDo
8. Compute the eigendecomposition of $\mathbf{Y}_m$ and retain $\widehat{Y} \in \mathbb{R}^{m\times t}$, $t \leq m$
9. Set $\mathbf{Z}_m = V_m \widehat{Y}$

A rich space

A good subspace $\mathcal{K}_m$ is crucial in the solution process

A rich space

A good subspace $\mathcal{K}_m$ is crucial in the solution process

$$\mathcal{K}_m = EK_m(A, b) = \text{Range}([b, A^{-1}b, \ldots, A^{m-1}b, A^{-m}b])$$

REMARKS:

- $EK_m(A, b) = K_{2m}(A, A^{-m}b)$
- $\dim\left(EK_m(A, b)\right) \leq 2m$
- More expensive basis construction due to linear system solves

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

The "Extended Arnoldi" (I)

Basis construction

- Compute an economy-size QR of $[b, A^{-1}b]$, that is

$$V_1\theta_{1,1} = [b, A^{-1}b]$$

  where $V_1 \equiv \mathcal{V}_1 = [v_1^{(1)}, v_1^{(2)}] \in \mathbb{R}^{n\times 2}$ is orthogonal and $\theta_1 \in \mathbb{R}^{2\times 2}$ upper triangular

- For $m = 1, \ldots,$ till convergence
  1. Set $\widehat{\mathcal{V}}_{m+1} = [Av_m^{(1)}, A^{-1}v_m^{(2)}]$
  2. Orthogonalize $\widehat{\mathcal{V}}_{m+1}$ wrt the whole $V_m \rightsquigarrow \widetilde{\mathcal{V}}_{m+1}$
  3. Compute economy-size QR of $\widetilde{\mathcal{V}}_{m+1} \rightsquigarrow$ new basis block $\mathcal{V}_{m+1}$
  4. Set $V_{m+1} = [V_m, \mathcal{V}_{m+1}]$

The "Extended Arnoldi" (II)

The orthonormalization coefficients θ's are collected in the block upper Hessenberg matrix $\mathcal{H}_m \in \mathbb{R}^{2m \times 2m}$

The "Extended Arnoldi" (II)

The orthonormalization coefficients θ's are collected in the block upper Hessenberg matrix $\mathcal{H}_m \in \mathbb{R}^{2m\times 2m}$

REMARK: $\mathcal{H}_m \neq V_m^T A V_m$!

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

The "Extended Arnoldi" (II)

The orthonormalization coefficients θ's are collected in the block upper Hessenberg matrix $\mathcal{H}_m \in \mathbb{R}^{2m\times 2m}$

REMARK: $\mathcal{H}_m \neq V_m^T A V_m$!

The (block) Arnoldi relation doesn't hold

$$\mathcal{V}_{m+1}\theta_{m+1,m} \neq A\mathcal{V}_m - V_m\mathcal{H}_m[e_{2m-1}, e_{2m}]$$

but

$$\mathcal{V}_{m+1}\theta_{m+1,m} = [Av_m^{(1)}, A^{-1}v_m^{(2)}] - V_m\mathcal{H}_m[e_{2m-1}, e_{2m}]$$

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

The "Extended Arnoldi" (II)

The orthonormalization coefficients θ's are collected in the block upper Hessenberg matrix $\mathcal{H}_m \in \mathbb{R}^{2m\times 2m}$

REMARK: $\mathcal{H}_m \neq V_m^T A V_m$!

The (block) Arnoldi relation doesn't hold

$$\mathcal{V}_{m+1}\theta_{m+1,m} \neq A\mathcal{V}_m - V_m\mathcal{H}_m[e_{2m-1}, e_{2m}]$$

but

$$\mathcal{V}_{m+1}\theta_{m+1,m} = [Av_m^{(1)}, A^{-1}v_m^{(2)}] - V_m\mathcal{H}_m[e_{2m-1}, e_{2m}]$$

$\Rightarrow$ Recover $T_m = V_m^T A V_m$ from $\mathcal{H}_m$ (SUPER TECHNICAL!)

References

References:

- *A new iterative method for solving large-scale Lyapunov matrix equations*
  V. Simoncini
  SIAM J. Sci. Comput., Vol. 29, No. 3, pp. 1268-1288, 2007

- *Computational methods for linear matrix equations* (Survey article)
  V. Simoncini
  To appear in SIAM Review, Vol. 58, No. 3, 2016