UNIVERSITÀ DI PAVIA

Preconditioned Low-Rank Riemannian Optimization for Symmetric Positive Definite Linear Matrix Equations

Due Giorni di Algebra Lineare Numerica e Applicazioni @ UNIPI

Ivan Bioli (joint work with Leonardo Robol and Daniel Kressner)

January 20, 2025

UNIVERSITÀ DI PAVIA
Department of Mathematics
“Felice Casorati”

Multiterm Linear Matrix Equations

Goal: Approximately solve multiterm linear matrix equations

$$\mathcal{A}X := A_1 X B_1^\top + A_2 X B_2^\top + \cdots + A_\ell X B_\ell^\top = F,$$

$A_i \in \mathbb{R}^{m\times m}$, $B_i \in \mathbb{R}^{n\times n}$, $F = F_L F_R^\top \in \mathbb{R}^{m\times n}$, and $X \in \mathbb{R}^{m\times n}$ unknown.

Setting of this work

- $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ is symmetric positive definite.
- Large and sparse coefficient matrices A_i, B_i, and low-rank F.
- $X_\star = \mathcal{A}^{-1}F \approx UV^\top$ where $U \in \mathbb{R}^{m\times r}$, $V \in \mathbb{R}^{n\times r}$ and $r \ll m, n$ [Beckermann et al. 2019; Benner and Breiten 2013; Jarlebring et al. 2018].

Low-rank methods for multiterm linear matrix equations

pCG with truncation: Apply pCG to the *vectorized formulation* $\mathcal{K}\,\mathrm{vec}\,(X) = \mathrm{vec}\,(F)$:

1. Iterates and aux. vecs in low-rank format
2. Low-rank truncation

Challenge: Growth of ranks in intermediate iters

↗ memory requirements;

↗ cost of truncation.

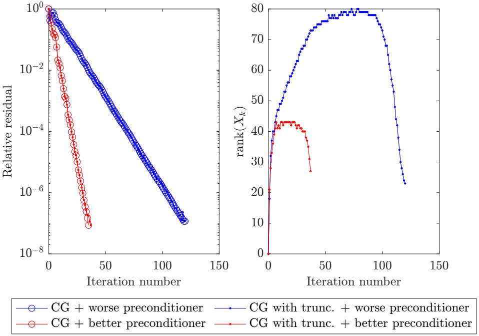

Figure: $\ell = 8$ terms and size $n = m = 1000$

Other methods (often combined): fixed-point iterations [Benner and Breiten 2013]; alternating optimization (ALS/AEM) [Lee et al. 2020]; projective/Galerkin methods [Powell et al. 2017; Kressner et al. 2015]; Sherman-Morrison-Woodbury formula [Hao et al. 2021].

Our approach: low-rank Riemannian optimization

Since $\mathcal{A} \succ 0$

$$\mathcal{A}X = F \iff \min_{X\in\mathbb{R}^{m\times n}} \|X - \mathcal{A}^{-1}F\|_{\mathcal{A}} \iff \min_{X\in\mathbb{R}^{m\times n}} f(X) := \frac{1}{2}\langle \mathcal{A}X, X\rangle - \langle X, F\rangle\,.$$

Low-rank Riemannian optimization approach

1. Restrict to rank-r matrices, $r \ll m, n$

$$\min_{X\in\mathcal{M}_r} f(X) \qquad \text{where} \qquad \mathcal{M}_r = \{X \in \mathbb{R}^{m\times n} \,:\, \operatorname{rank}(X) = r\}.$$

2. $\mathcal{M}_r$ is an embedded submanifold of $\mathbb{R}^{m\times n}$ $\rightsquigarrow$ Riemannian optimization.
3. **Preconditioning** via a change of the Riemannian metric on $\mathcal{M}_r$.

UNIVERSITÀ DI PAVIA

Retraction-based Riemannian optimization

Riemannian gradient w.r.t. the Riemannian metric $\langle \cdot, \cdot \rangle_X$ is $\mathrm{grad} f(X) \in \mathrm{T}_X \mathcal{M}_r$ s.t.

$$\mathrm{D}f(X)[\xi] = \langle \mathrm{grad} f(X), \xi \rangle_X \qquad \forall (X, \xi) \in \mathrm{T}\mathcal{M}_r.$$

If $\mathcal{M}_r$ is a Riemannian submanifold, i.e. $\langle \cdot, \cdot \rangle_X = \langle \cdot, \cdot \rangle$, then

$$\mathrm{grad} f(X) = \mathrm{Proj}_X(\nabla f(X)), \qquad \mathrm{Proj}_X : \mathbb{R}^{m \times n} \to \mathrm{T}_X \mathcal{M} \text{ orthogonal projector.}$$

Euclidean Gradient Descent (GD)

$$X_{k+1} = X_k - \alpha_k \nabla f(X_k)$$

Riemannian Gradient Descent (R-GD)

$$X_{k+1} = \mathrm{R}_{X_k}(-\alpha_k \mathrm{grad} f(X_k))$$

Retraction-based Riemannian optimization

Riemannian gradient w.r.t. the Riemannian metric $\langle \cdot, \cdot \rangle_X$ is $\mathrm{grad} f(X) \in \mathrm{T}_X \mathcal{M}_r$ s.t.

$$\mathrm{D}f(X)[\xi] = \langle \mathrm{grad} f(X), \xi \rangle_X \qquad \forall (X, \xi) \in \mathrm{T}\mathcal{M}_r.$$

If $\mathcal{M}_r$ is a Riemannian submanifold, i.e. $\langle \cdot, \cdot \rangle_X = \langle \cdot, \cdot \rangle$, then

$$\mathrm{grad} f(X) = \mathrm{Proj}_X(\nabla f(X)), \qquad \mathrm{Proj}_X : \mathbb{R}^{m \times n} \to \mathrm{T}_X \mathcal{M} \text{ orthogonal projector}.$$

Euclidean Gradient Descent (GD)

$$X_{k+1} = X_k - \alpha_k \nabla f(X_k)$$

Riemannian Gradient Descent (R-GD)

$$X_{k+1} = \mathrm{R}_{X_k}(-\alpha_k \mathrm{grad} f(X_k))$$

- $\mathrm{R} : \mathrm{T}\mathcal{M}_r \to \mathcal{M}_r$, $\mathrm{R}(X, \xi) = \mathrm{R}_X(\xi)$ is a **retraction** map. For rank-r matrices, the **metric projection retraction**: $\mathrm{R}_X(\xi) = \arg\min_{Y \in \mathcal{M}_r} \|X + \xi - Y\|_\mathrm{F} = \mathrm{TSVD}_r(X + \xi)$.
- Minor modifications for Riemannian Non-Linear Conjugate Gradient (R-NLCG).
- Convergence is slower the larger $\kappa(\mathrm{Hess} f(X_\star))$.

Riemannian preconditioning

$$\mathrm{Hess} f(X)[\xi] = \overbrace{\mathrm{Proj}_X(\mathcal{A}\xi)}^{\text{Projected Euclidean Hessian}} + \overbrace{\mathcal{D}_\xi(\mathrm{Proj}_X^\perp(\mathcal{A}X - F))}^{\text{Curvature term}} \quad \forall (X,\xi) \in \mathrm{T}\mathcal{M},$$

where $\mathcal{D}_\xi$ is the differential of $X \mapsto \mathrm{Proj}_X$ at X along ξ.

Ill-conditioned $\mathcal{A} \rightsquigarrow$ ill-conditioned Riemannian Hessian $\rightsquigarrow$ need preconditioning.

How do we obtain a preconditioner?

Riemannian preconditioning

$$\mathrm{Hess} f(X)[\xi] = \overbrace{\mathrm{Proj}_X(\mathcal{A}\xi)}^{\text{Projected Euclidean Hessian}} + \overbrace{\mathcal{D}_\xi(\mathrm{Proj}_X^\perp(\mathcal{A}X - F))}^{\text{Curvature term}} \quad \forall (X,\xi) \in \mathrm{T}\mathcal{M},$$

where $\mathcal{D}_\xi$ is the differential of $X \mapsto \mathrm{Proj}_X$ at X along ξ.

Ill-conditioned $\mathcal{A} \rightsquigarrow$ ill-conditioned Riemannian Hessian $\rightsquigarrow$ need preconditioning.

How do we obtain a preconditioner? Given $\mathcal{P} \succ 0$ s.t. $\mathcal{P} \approx \mathcal{A}$:

$$\mathrm{Hess} f(X)[\xi] = \mathrm{Proj}_X(\mathcal{A}\xi) + \mathcal{D}_\xi(\mathrm{Proj}_X^\perp(\mathcal{A}X - F))$$

Riemannian preconditioning

$$\mathrm{Hess} f(X)[\xi] = \overbrace{\mathrm{Proj}_X(\mathcal{A}\xi)}^{\text{Projected Euclidean Hessian}} + \overbrace{\mathcal{D}_\xi(\mathrm{Proj}_X^\perp(\mathcal{A}X - F))}^{\text{Curvature term}} \quad \forall (X, \xi) \in \mathrm{T}\mathcal{M},$$

where $\mathcal{D}_\xi$ is the differential of $X \mapsto \mathrm{Proj}_X$ at X along ξ.

Ill-conditioned $\mathcal{A} \rightsquigarrow$ ill-conditioned Riemannian Hessian $\rightsquigarrow$ need preconditioning.

How do we obtain a preconditioner? Given $\mathcal{P} \succ 0$ s.t. $\mathcal{P} \approx \mathcal{A}$:

$$\mathrm{Hess} f(X)[\xi] \approx \mathrm{Proj}_X(\mathcal{A}\xi) + \cancel{\mathcal{D}_\xi(\mathrm{Proj}_X^\perp(\mathcal{A}X - F))}$$

Riemannian preconditioning

$$\mathrm{Hess} f(X)[\xi] = \overbrace{\mathrm{Proj}_X(\mathcal{A}\xi)}^{\text{Projected Euclidean Hessian}} + \overbrace{\mathcal{D}_\xi(\mathrm{Proj}_X^\perp(\mathcal{A}X - F))}^{\text{Curvature term}} \quad \forall (X,\xi) \in \mathrm{T}\mathcal{M},$$

where $\mathcal{D}_\xi$ is the differential of $X \mapsto \mathrm{Proj}_X$ at X along ξ.

Ill-conditioned $\mathcal{A} \rightsquigarrow$ ill-conditioned Riemannian Hessian $\rightsquigarrow$ need preconditioning.

How do we obtain a preconditioner? Given $\mathcal{P} \succ 0$ s.t. $\mathcal{P} \approx \mathcal{A}$:

$$\mathrm{Hess} f(X)[\xi] \approx \mathrm{Proj}_X(\mathcal{P}\xi) + \cancel{\mathcal{D}_\xi(\mathrm{Proj}_X^\perp(\mathcal{A}X - F))}$$

Riemannian preconditioning

$$\mathrm{Hess} f(X)[\xi] = \overbrace{\mathrm{Proj}_X(\mathcal{A}\xi)}^{\text{Projected Euclidean Hessian}} + \overbrace{\mathcal{D}_\xi(\mathrm{Proj}_X^\perp(\mathcal{A}X - F))}^{\text{Curvature term}} \quad \forall (X,\xi) \in \mathrm{T}\mathcal{M},$$

where $\mathcal{D}_\xi$ is the differential of $X \mapsto \mathrm{Proj}_X$ at X along ξ.

Ill-conditioned $\mathcal{A} \rightsquigarrow$ ill-conditioned Riemannian Hessian $\rightsquigarrow$ need preconditioning.

How do we obtain a preconditioner? Given $\mathcal{P} \succ 0$ s.t. $\mathcal{P} \approx \mathcal{A}$:

$$\mathrm{Hess} f(X)[\xi] \approx \mathrm{Proj}_X(\mathcal{P}\xi) + \cancel{\mathcal{D}_\xi(\mathrm{Proj}_X^\perp(\mathcal{A}X - F))} \rightsquigarrow \mathcal{P}_X = \mathrm{Proj}_X \circ \mathcal{P} \circ \mathrm{Proj}_X$$

Preconditioned R-GD $\longleftrightarrow$ R-GD on $\left(\mathcal{M}_r,\ \langle \cdot, \cdot \rangle_{\mathcal{P}_X} = \langle \mathcal{P}_X \cdot, \cdot \rangle\right)$

- For $\eta, \xi \in \mathrm{T}_X\mathcal{M}_r$: $\langle \eta, \xi \rangle_{\mathcal{P}_X} = \langle \mathrm{Proj}_X(\mathcal{P}\xi), \eta \rangle = \langle \mathcal{P}\xi, \eta \rangle = \langle \eta, \xi \rangle_{\mathcal{P}}$

UNIVERSITÀ DI PAVIA

Riemannian preconditioning

$$\operatorname{Hess} f(X)[\xi] = \underbrace{\operatorname{Proj}_X(\mathcal{A}\xi)}_{\text{Projected Euclidean Hessian}} + \underbrace{\mathcal{D}_\xi(\operatorname{Proj}_X^\perp(\mathcal{A}X - F))}_{\text{Curvature term}} \quad \forall (X, \xi) \in \mathrm{T}\mathcal{M},$$

where $\mathcal{D}_\xi$ is the differential of $X \mapsto \operatorname{Proj}_X$ at X along ξ.

Ill-conditioned $\mathcal{A} \rightsquigarrow$ ill-conditioned Riemannian Hessian $\rightsquigarrow$ need preconditioning.

How do we obtain a preconditioner? Given $\mathcal{P} \succ 0$ s.t. $\mathcal{P} \approx \mathcal{A}$:

$$\operatorname{Hess} f(X)[\xi] \approx \operatorname{Proj}_X(\mathcal{P}\xi) + \cancel{\mathcal{D}_\xi(\operatorname{Proj}_X^\perp(\mathcal{A}X - F))} \rightsquigarrow \mathcal{P}_X = \operatorname{Proj}_X \circ \mathcal{P} \circ \operatorname{Proj}_X$$

Preconditioned R-GD $\longleftrightarrow$ R-GD on $(\mathcal{M}_r, \langle \cdot, \cdot \rangle_{\mathcal{P}_X} = \langle \mathcal{P}_X \cdot, \cdot \rangle)$

- For $\eta, \xi \in \mathrm{T}_X\mathcal{M}_r$: $\langle \eta, \xi \rangle_{\mathcal{P}_X} = \langle \operatorname{Proj}_X(\mathcal{P}\xi), \eta \rangle = \langle \mathcal{P}\xi, \eta \rangle = \langle \eta, \xi \rangle_{\mathcal{P}}$, hence

Precond. gradient $\langle \xi, \operatorname{grad} f(X) \rangle = \langle \xi, \mathcal{P}_X^{-1} \operatorname{grad} f(X) \rangle_{\mathcal{P}_X} \implies \operatorname{grad}_{\mathcal{P}_X} f(X) = \mathcal{P}_X^{-1} \operatorname{grad} f(X)$.

Precond. inner product Consider $\mathcal{M}_r \subset (\mathbb{R}^{m \times n}, \langle \cdot, \cdot \rangle_{\mathcal{P}} = \langle \mathcal{P} \cdot, \cdot \rangle)$.

UNIVERSITÀ DI PAVIA

Riemannian preconditioning

$$\mathrm{Hess}f(X)[\xi] = \overbrace{\mathrm{Proj}_X(\mathcal{A}\xi)}^{\text{Projected Euclidean Hessian}} + \overbrace{\mathcal{D}_\xi(\mathrm{Proj}_X^\perp(\mathcal{A}X - F))}^{\text{Curvature term}} \quad \forall (X,\xi) \in \mathrm{T}\mathcal{M},$$

where $\mathcal{D}_\xi$ is the differential of $X \mapsto \mathrm{Proj}_X$ at X along ξ.

Ill-conditioned $\mathcal{A} \rightsquigarrow$ ill-conditioned Riemannian Hessian $\rightsquigarrow$ need preconditioning.

How do we obtain a preconditioner? Given $\mathcal{P} \succ 0$ s.t. $\mathcal{P} \approx \mathcal{A}$:

$$\mathrm{Hess}f(X)[\xi] \approx \mathrm{Proj}_X(\mathcal{P}\xi) + \cancel{\mathcal{D}_\xi(\mathrm{Proj}_X^\perp(\mathcal{A}X - F))} \rightsquigarrow \mathcal{P}_X = \mathrm{Proj}_X \circ \mathcal{P} \circ \mathrm{Proj}_X$$

Preconditioned R-GD $\longleftrightarrow$ R-GD on $\left(\mathcal{M}_r, \langle\cdot,\cdot\rangle_{\mathcal{P}_X} = \langle \mathcal{P}_X \cdot, \cdot\rangle\right)$

- For $\eta, \xi \in \mathrm{T}_X\mathcal{M}_r$: $\langle \eta, \xi\rangle_{\mathcal{P}_X} = \langle \mathrm{Proj}_X(\mathcal{P}\xi), \eta\rangle = \langle \mathcal{P}\xi, \eta\rangle = \langle \eta, \xi\rangle_{\mathcal{P}}$, hence

Precond. gradient $\langle \xi, \mathrm{grad}f(X)\rangle = \langle \xi, \mathcal{P}_X^{-1}\mathrm{grad}f(X)\rangle_{\mathcal{P}_X} \implies \mathrm{grad}_{\mathcal{P}_X} f(X) = \mathcal{P}_X^{-1}\mathrm{grad}f(X)$.

Precond. inner product Consider $\mathcal{M}_r \subset (\mathbb{R}^{m\times n}, \langle\cdot,\cdot\rangle_{\mathcal{P}} = \langle \mathcal{P}\cdot, \cdot\rangle)$.

- We consider $\mathcal{P}X = EXD$ or $\mathcal{P}X = AXD + EXB$, where $A, B, D, E \succ 0$.

UNIVERSITÀ DI PAVIA

$\mathcal{P}X = EXD$ where $E, D \succ 0$

Optimization on $\mathcal{M}_r$ with metric $\langle \cdot, \cdot \rangle_{\mathcal{P}_X} \equiv \langle \cdot, \cdot \rangle_{\mathcal{P}}$, where $\mathcal{P}X = EXD$ with $E, D \succ 0$ and $\mathcal{P}_X = \mathrm{Proj}_X \circ \mathcal{P} \circ \mathrm{Proj}_X$.

$\mathcal{P}X = EXD$ where $E, D \succ 0$

Optimization on $\mathcal{M}_r$ with metric $\langle \cdot, \cdot \rangle_{\mathcal{P}_X} \equiv \langle \cdot, \cdot \rangle_{\mathcal{P}}$, where $\mathcal{P}X = EXD$ with $E, D \succ 0$ and $\mathcal{P}_X = \mathrm{Proj}_X \circ \mathcal{P} \circ \mathrm{Proj}_X$.

Alternative 1: Preconditioned gradient. Through the application of $\mathcal{P}_X^{-1}$

- Given $\eta \in \mathrm{T}_X\mathcal{M}_r$, solve $\mathrm{Proj}_X(E\xi D) = \eta$, $\xi \in \mathrm{T}_X\mathcal{M}_r$.
- **Cost**: r linear systems with E, D + $\mathcal{O}(r^2(m+n))$.

$\mathcal{P}X = EXD$ where $E, D \succ 0$

Optimization on $\mathcal{M}_r$ with metric $\langle\cdot,\cdot\rangle_{\mathcal{P}_X} \equiv \langle\cdot,\cdot\rangle_{\mathcal{P}}$, where $\mathcal{P}X = EXD$ with $E, D \succ 0$ and $\mathcal{P}_X = \mathrm{Proj}_X \circ \mathcal{P} \circ \mathrm{Proj}_X$.

Alternative 1: Preconditioned gradient. Through the application of $\mathcal{P}_X^{-1}$

- Given $\eta \in \mathrm{T}_X\mathcal{M}_r$, solve $\mathrm{Proj}_X(E\xi D) = \eta$, $\xi \in \mathrm{T}_X\mathcal{M}_r$.
- **Cost**: r linear systems with E, D + $\mathcal{O}(r^2(m+n))$.

Alternative 2: Preconditioned inner product. Consider $\mathcal{M}_r \subset (\mathbb{R}^{m\times n}, \langle\cdot,\cdot\rangle_{\mathcal{P}})$

- Idea: Replace the SVD with the E, D-SVD, i.e., $X = \tilde{U}\tilde{\Sigma}\tilde{V}^\top \in \mathcal{M}_r$ where $\tilde{U}^\top E\tilde{U} = \tilde{V}^\top D\tilde{V} = I_r$ and $\tilde{\Sigma} = \mathrm{diag}(\tilde{\sigma}_i)$, $\tilde{\sigma}_1 \geq \cdots \geq \tilde{\sigma}_r > 0$.

$\mathcal{P}X = AXD + EXB$ where $A, B, D, E \succ 0$

Exact application of the preconditioner's inverse

Optimization on $\mathcal{M}_r$ with metric $\langle\cdot,\,\cdot\rangle_{\mathcal{P}_X} \equiv \langle\cdot,\,\cdot\rangle_{\mathcal{P}}$, $\mathcal{P}X = AXD + EXB$ with $A, B, D, E \succ 0$ and $\mathcal{P}_X = \mathrm{Proj}_X \circ \mathcal{P} \circ \mathrm{Proj}_X$.

Case 1 $\mathcal{P}X = AX + XB$. Apply $\mathcal{P}_X^{-1}$, i.e., solve $\mathrm{Proj}_X(A\xi + \xi B) = \eta$ for $\xi \in \mathrm{T}_X\mathcal{M}_r$, given η.

- **Cost:** $(2r^2 + r)$ shifted systems with A, B + $r^2 \times r^2$ system (matvecs $\mathcal{O}(r^3)$).

$\mathcal{P}X = AXD + EXB$ where $A, B, D, E \succ 0$

Exact application of the preconditioner's inverse

Optimization on $\mathcal{M}_r$ with metric $\langle \cdot, \cdot \rangle_{\mathcal{P}_X} \equiv \langle \cdot, \cdot \rangle_{\mathcal{P}}$, $\mathcal{P}X = AXD + EXB$ with $A, B, D, E \succ 0$ and $\mathcal{P}_X = \mathrm{Proj}_X \circ \mathcal{P} \circ \mathrm{Proj}_X$.

Case 1 $\mathcal{P}X = AX + XB$. Apply $\mathcal{P}_X^{-1}$, i.e., solve $\mathrm{Proj}_X(A\xi + \xi B) = \eta$ for $\xi \in \mathrm{T}_X\mathcal{M}_r$, given η.

- **Cost:** $(2r^2 + r)$ shifted systems with A, B + $r^2 \times r^2$ system (matvecs $\mathcal{O}(r^3)$).

Case 2 $\mathcal{P}X = AXD + EXB$. Decompose

$$\mathcal{P} = \mathcal{B}\tilde{\mathcal{P}} \quad \text{where} \quad \mathcal{B}X = EXD, \quad \tilde{\mathcal{P}}X = \mathcal{B}^{-1}\mathcal{P}X = E^{-1}AX + XBD^{-1},$$

and observe that $\langle X, Y \rangle_{\mathcal{P}} = \langle X, \tilde{\mathcal{P}}Y \rangle_{\mathcal{B}}$

$\mathcal{P}X = AXD + EXB$ where $A, B, D, E \succ 0$

Exact application of the preconditioner's inverse

Optimization on $\mathcal{M}_r$ with metric $\langle \cdot, \cdot \rangle_{\mathcal{P}_X} \equiv \langle \cdot, \cdot \rangle_{\mathcal{P}}$, $\mathcal{P}X = AXD + EXB$ with $A, B, D, E \succ 0$ and $\mathcal{P}_X = \mathrm{Proj}_X \circ \mathcal{P} \circ \mathrm{Proj}_X$.

Case 1 $\mathcal{P}X = AX + XB$. Apply $\mathcal{P}_X^{-1}$, i.e., solve $\mathrm{Proj}_X(A\xi + \xi B) = \eta$ for $\xi \in \mathrm{T}_X\mathcal{M}_r$, given η.

- **Cost:** $(2r^2 + r)$ shifted systems with A, B + $r^2 \times r^2$ system (matvecs $\mathcal{O}(r^3)$).

Case 2 $\mathcal{P}X = AXD + EXB$. Decompose

$$\mathcal{P} = \mathcal{B}\tilde{\mathcal{P}} \quad \text{where} \quad \mathcal{B}X = EXD, \quad \tilde{\mathcal{P}}X = \mathcal{B}^{-1}\mathcal{P}X = E^{-1}AX + XBD^{-1},$$

and observe that $\langle X, Y \rangle_{\mathcal{P}} = \langle X, \tilde{\mathcal{P}}Y \rangle_{\mathcal{B}}$:

2.1 Precondition the metric with $\mathcal{B}$;

2.2 Precondition the gradient with $\tilde{\mathcal{P}}$: solve $\mathrm{Proj}_X^{\mathcal{B}}(\tilde{\mathcal{P}}\xi) = \mathrm{Proj}_X^{\mathcal{B}}(E^{-1}A\xi + \xi BD^{-1}) = \eta$.

   - **Cost:** $(2r^2 + r)$ shifted systems with $A + \lambda E, B + \mu D$ + $r^2 \times r^2$ system.

$\mathcal{P}X = AXD + EXB$ where $A, B, D, E \succ 0$

Approximating the preconditioner's inverse via tangADI

Problem: Applying $\mathcal{P}_X^{-1}$ can be expensive as r grows ($2r^2 + r$ linear systems).

Idea: Approximate $\mathcal{P}_X^{-1}$ via an ADI-like iteration on the tangent space (tangADI).

$\mathcal{P}X = AXD + EXB$ where $A, B, D, E \succ 0$

Approximating the preconditioner's inverse via tangADI

Problem: Applying $\mathcal{P}_X^{-1}$ can be expensive as r grows ($2r^2 + r$ linear systems).

Idea: Approximate $\mathcal{P}_X^{-1}$ via an ADI-like iteration on the tangent space (tangADI).

ADI iteration with shifts $(p_j, q_j) = (p, q)$ $\forall j$ is a fixed point iteration from rewriting

$$AXD + EXB = C \iff (A - qE)X(B + pD) - (A - pE)X(B + qD) = (p - q)C.$$

$\mathcal{P}X = AXD + EXB$ where $A, B, D, E \succ 0$

Approximating the preconditioner's inverse via tangADI

Problem: Applying $\mathcal{P}_X^{-1}$ can be expensive as r grows ($2r^2 + r$ linear systems).

Idea: Approximate $\mathcal{P}_X^{-1}$ via an ADI-like iteration on the tangent space (tangADI).

ADI iteration with shifts $(p_j, q_j) = (p, q)$ $\forall j$ is a fixed point iteration from rewriting

$$AXD + EXB = C \iff (A - qE)X(B + pD) - (A - pE)X(B + qD) = (p - q)C.$$

tangADI for $\mathrm{Proj}_X(A\xi D + E\xi B) = \eta \in \mathrm{T}_X\mathcal{M}_r$, where $\xi \in \mathrm{T}_X\mathcal{M}_r$:

$$\mathrm{Proj}_X\Big((A - qE)\xi(B + pD)\Big) = \mathrm{Proj}_X\Big((A - pE)\xi(B + qD)\Big) + (p - q)\eta.$$

UNIVERSITÀ DI PAVIA

$\mathcal{P}X = AXD + EXB$ where $A, B, D, E \succ 0$

Approximating the preconditioner's inverse via tangADI

Problem: Applying $\mathcal{P}_X^{-1}$ can be expensive as r grows ($2r^2 + r$ linear systems).

Idea: Approximate $\mathcal{P}_X^{-1}$ via an ADI-like iteration on the tangent space (tangADI).

ADI iteration with shifts $(p_j, q_j) = (p, q)$ $\forall j$ is a fixed point iteration from rewriting

$$AXD + EXB = C \iff (A - qE)X(B + pD) - (A - pE)X(B + qD) = (p - q)C.$$

tangADI for $\mathrm{Proj}_X(A\xi D + E\xi B) = \eta \in \mathrm{T}_X\mathcal{M}_r$, where $\xi \in \mathrm{T}_X\mathcal{M}_r$:

$$\mathrm{Proj}_X\Big((A - q_jE)\xi^{(j)}(B + p_jD)\Big) = \mathrm{Proj}_X\Big((A - p_jE)\xi^{(j-1)}(B + q_jD)\Big) + (p_j - q_j)\eta.$$

- **Cost (per iter.):** r linear systems with $A - q_jE$, $B + q_jD$ + $\mathcal{O}(r^2(m+n))$.

Numerical Experiments I

Finite difference discretization of 2D PDEs with semiseparable diffusion coefficient

$-\nabla \cdot (k\nabla u) = 0$ in $\Omega = (0,1)^2$ and $u = g$ on $\partial\Omega$, where

$$k(x,y) = \alpha_1 k_{1,x}(x)k_{1,y}(y) + \cdots + \alpha_{\ell_k} k_{\ell_k,x}(x)k_{\ell_k,y}(y).$$

Discretization with Finite Differences (FD) and grid spacing $h = 1/(n+1)$

$$\sum_{j=1}^{\ell_k} \alpha_j \left(A_{j,x} U D_{j,y}^\top + D_{j,x} U A_{j,y}^\top \right) = F \in \mathbb{R}^{n\times n}, \qquad \mathrm{rank}(F) = 4.$$

Data $k(x,y) = \sum_{j=1}^{\ell_k} \frac{10^j}{j!} x^j y^j$, $g(x,y) = e^{-10(x+1)y}$;
We set $\ell_k = 4$, $n = 10000 \rightsquigarrow \ell = 8$ terms and size $n = m = 10000$.

Preconditioning

- $k(x,y) \approx (1 + (\sqrt{\alpha}x)^{\ell_k})(1 + (\sqrt{\alpha}y)^{\ell_k}) \rightsquigarrow \mathcal{P}^{(2)}X = AXD + DXA$;
- Approximate $D \approx I \rightsquigarrow \mathcal{P}^{(1)}X = AX + XA$.

Numerical Experiments I

Finite difference discretization of 2D PDEs with semiseparable diffusion coefficient

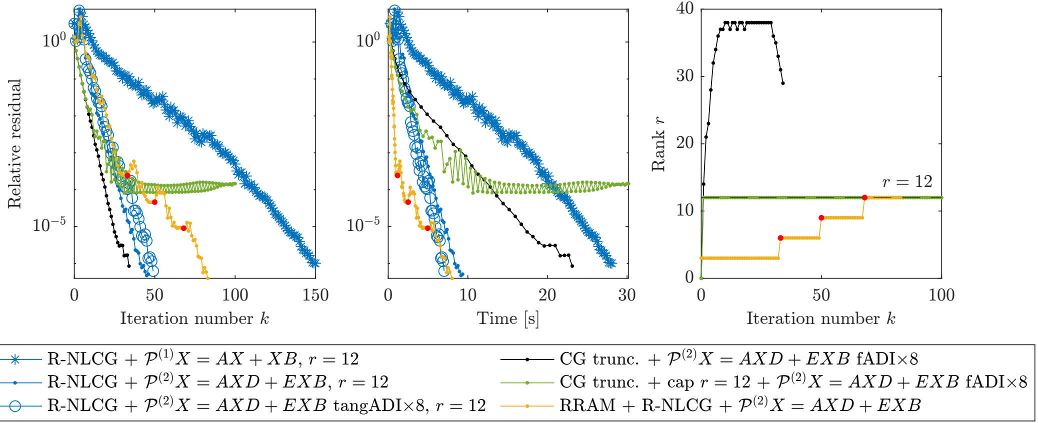

Figure: Comparison of Riemannian optimization and CG with truncation. From left to right: relative residual vs. iterations, relative residual vs. time, and rank of approximate solution vs. iterations.

UNIVERSITÀ DI PAVIA

Numerical Experiments II

Multiterm Lyapunov equation

$$\underbrace{AXM^\top + MXA^\top}_{\mathcal{L}(X)} - \underbrace{\sum_{i=1}^{\ell} N_i X N_i^\top}_{\Pi(X)} = BB^\top,$$

with matrices obtained modifying a bilinear control problem [Benner and Saak 2005]

- $n = 5177$, $\ell = 6$;
- $A \succ 0$ FE stiffness matrix; $M \succ 0$ FE mass matrix, with $\kappa_2(M) \approx 350$;
- $N_i = N_i^\top$ and such that $\rho(\mathcal{L}^{-1}\Pi) < 1$:
  - $\mathcal{A} = \mathcal{L} + \Pi \succ 0$;
  - Solution $X \succeq 0 \rightsquigarrow$ optimization on fixed-rank symmetric positive semidefinite matrices;
- $B \in \mathbb{R}^{n\times 2}$.

Preconditioning

- Dominant term $\mathcal{P}^{(2)}X = \mathcal{L}(X) = AXM + MXA$;
- Approximate $M \approx I \rightsquigarrow \mathcal{P}^{(1)}X = AX + XA$ [Vandereycken 2010].

Numerical Experiments II

Multiterm Lyapunov equation

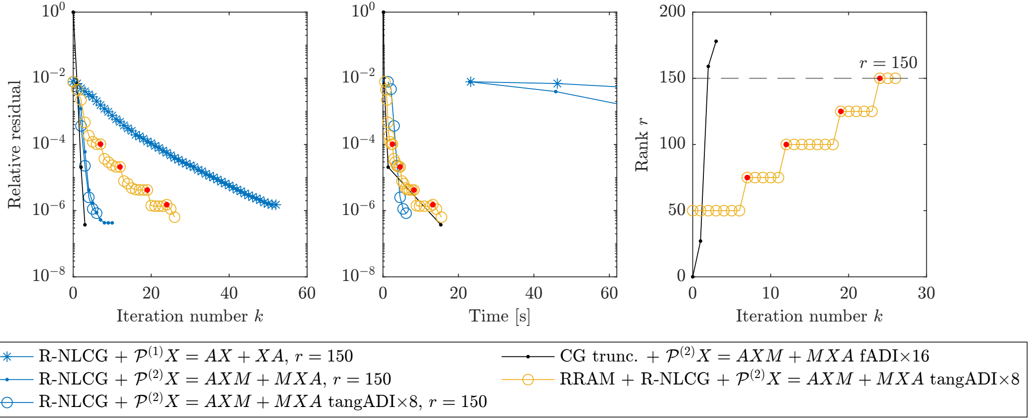

Figure: Multiterm Lyapunov equation ($n = 5177$, $\ell = 6$). Comparison of Riemannian optimization and CG with truncation. From left to right: relative residual vs. iterations, relative residual vs. time, and rank of approximate solution vs. iterations.

Summary

- First-order Riemannian optimization $\rightsquigarrow$ low-rank solver for multiterm matrix eqs.
- Preconditioning is a challenge due to the Riemannian structure. We propose novel preconditioning strategies from preconditioners defined on the ambient space:
  - $\mathcal{P}X = EXD$ with $E, D \succ 0$ (preconditioning the metric or the Riemannian gradient);
  - $\mathcal{P}X = AXD + EXB$ with $A, B, D, E \succ 0$ (exact preconditioning of the Riemannian gradient, or tangADI).
- If rank is not known $\rightsquigarrow$ Riemannian Rank-Adaptive Method (RRAM) $\rightsquigarrow$ a new low-rank solver competitive with CG with truncation.

Thank you for your attention!

UNIVERSITÀ DI PAVIA

Main references

- Beckermann, B. and A. Townsend (2019). "Bounds on the Singular Values of Matrices with Displacement Structure". In: *SIAM Review* 61.2, pp. 319–344.
- Benner, P. and T. Breiten (2013). "Low rank methods for a class of generalized Lyapunov equations and related issues". In: *Numerische Mathematik* 124.3, pp. 441–470.
- Benner, P. and J. Saak (2005). "A Semi-Discretized Heat Transfer Model for Optimal Cooling of Steel Profiles". In: *Dimension Reduction of Large-Scale Systems*. Ed. by P. Benner, D. C. Sorensen, and V. Mehrmann. Lecture Notes in Computational Science and Engineering. Springer, pp. 353–356.
- Bioli, I., D. Kressner, and L. Robol (2024). *Preconditioned Low-Rank Riemannian Optimization for Symmetric Positive Definite Linear Matrix Equations*. arXiv: 2408.16416.
- Hao, Y. and V. Simoncini (2021). "The Sherman–Morrison–Woodbury formula for generalized linear matrix equations and applications". In: *Numerical Linear Algebra with Applications* 28.5, e2384.

Main references

- Jarlebring, E. et al. (2018). "Krylov methods for low-rank commuting generalized Sylvester equations". In: *Numerical Linear Algebra with Applications* 25.6, e2176. arXiv: 1704.02167 [math].
- Kressner, D. and P. Sirković (2015). "Truncated low-rank methods for solving general linear matrix equations". In: *Numerical Linear Algebra with Applications* 22.3, pp. 564–583.
- Lee, K. et al. (2020). *Alternating Energy Minimization Methods for Multi-term Matrix Equations*. arXiv: 2006.08531 [cs, math].
- Powell, C. E., D. Silvester, and V. Simoncini (2017). "An Efficient Reduced Basis Solver for Stochastic Galerkin Matrix Equations". In: *SIAM Journal on Scientific Computing* 39.1, A141–A163.
- Vandereycken, B. (2010). "Riemannian and multilevel optimization for rank-constrained matrix problems". PhD thesis. Department of Computer Science, KU Leuven.

Riemannian optimization on fixed-rank matrices

- Points $X \in \mathcal{M}_r$ are represented by their rank-r TSVD factors:

$$X = U\Sigma V^\top \in \mathbb{R}^{m\times n} \;:\; U^\top U = V^\top V = I_r, \Sigma = \mathrm{diag}(\sigma_i), \sigma_1 \geq \cdots \geq \sigma_r > 0,$$

  hence $X \leftrightarrow (U,\, \Sigma,\, V)$.

- Tangent vectors ξ at $X = U\Sigma V^\top$

$$\mathrm{T}_X\mathcal{M}_r = \left\{ \xi = UMV^\top + U_pV^\top + UV_p^\top \;:\; M \in \mathbb{R}^{r\times r},\, U^\top U_p = V^\top V_p = 0_r \right\},$$

  hence $\xi \leftrightarrow (U_p,\, V_p,\, M)$.

- Flops cost of basic operations
  - $\mathrm{Proj}_X : \mathbb{R}^{m\times n} \to \mathrm{T}_X\mathcal{M}_r \rightsquigarrow \mathcal{O}\left(r^2(n+m)\right)$.
  - $\mathrm{R}_X(\xi) = \mathrm{TSVD}_r(X+\xi) \rightsquigarrow \mathcal{O}\left(r^2(n+m)\right)$.
  - $\mathrm{grad}f(X) \rightsquigarrow \mathcal{O}\left(\ell r^2(n+m)\right)$.

UNIVERSITÀ DI PAVIA

The Riemannian gradient

Euclidean gradient at $X = U\Sigma V^\top \in \mathcal{M}_r$:

$$\nabla f(X) = \mathcal{A}X - F = \begin{bmatrix} A_1 U\Sigma & \cdots & A_\ell U\Sigma & -F_L \end{bmatrix} \begin{bmatrix} B_1 V & \cdots & B_\ell V & F_R \end{bmatrix}^\top.$$

Riemannian gradient: $\mathrm{grad} f(X) = \mathrm{Proj}_X(\nabla f(X)) \leftrightarrow (M, U_p, V_p)$ where

$$M = U^\top \nabla f(X) V, \qquad U_p = \nabla f(X) V - UM, \qquad V_p = \nabla f(X)^\top U - VM^\top.$$

In our case:

$$\nabla f(X) V = \sum_{i=1}^{\ell} (A_i U)\Sigma(V^\top B_i V)^\top - F_L(F_R^\top V),$$

$$\nabla f(X)^\top U = \sum_{i=1}^{\ell} (B_i V)\Sigma(U^\top A_i U)^\top - F_R(F_L^\top U).$$

$\mathcal{P}X = AXD + EXB$ where $A, B, D, E \succ 0$

Exact application of the preconditioner's inverse

Optimization on $\mathcal{M}_r$ with metric $\langle \cdot, \cdot \rangle_{\mathcal{P}_X} \equiv \langle \cdot, \cdot \rangle_{\mathcal{P}}$, $\mathcal{P}X = AXD + EXB$ with $A, B, D, E \succ 0$ and $\mathcal{P}_X = \mathrm{Proj}_X \circ \mathcal{P} \circ \mathrm{Proj}_X$.

Case 1 $\mathcal{P}X = AX + XB$. Apply $\mathcal{P}_X^{-1}$, i.e., solve $\mathrm{Proj}_X(A\xi + \xi B) = \eta$ for $\xi \in \mathrm{T}_X\mathcal{M}_r$, given η:

$$U_\eta = \mathrm{P}_U^\perp (A\xi + \xi B)V, \quad V_\eta = \mathrm{P}_V^\perp (A\xi + \xi B)^\top U, \quad M_\eta = U^\top (A\xi + \xi B)V.$$

- **Cost:** $(2r^2 + r)$ shifted systems with A, B + $r^2 \times r^2$ system (matvecs $\mathcal{O}(r^3)$).

Case 2 $\mathcal{P}X = AXD + EXB$. Decompose

$$\mathcal{P} = \mathcal{B}\tilde{\mathcal{P}} \quad \text{where} \quad \mathcal{B}X = EXD, \quad \tilde{\mathcal{P}}X = \mathcal{B}^{-1}\mathcal{P}X = E^{-1}AX + XBD^{-1},$$

and observe that $\langle X, Y \rangle_{\mathcal{P}} = \langle X, \tilde{\mathcal{P}}Y \rangle_{\mathcal{B}}$:

2.1 Precondition the metric with $\mathcal{B}$;

2.2 Precondition the gradient with $\tilde{\mathcal{P}}$: solve $\mathrm{Proj}_X^{\mathcal{B}}(\tilde{\mathcal{P}}\xi) = \mathrm{Proj}_X^{\mathcal{B}}(E^{-1}A\xi + \xi BD^{-1}) = \eta$.

  - **Cost:** $(2r^2 + r)$ shifted systems with $A + \lambda E, B + \mu D$ + $r^2 \times r^2$ system.