Riemannian preconditioning

Bamdev Mishra

Joint work with Rodolphe Sepulchre

University of Liège, Belgium
University of Cambridge, UK

Workshop on low-rank optimization and applications
Hausdorff Center for Mathematics, Bonn, Germany

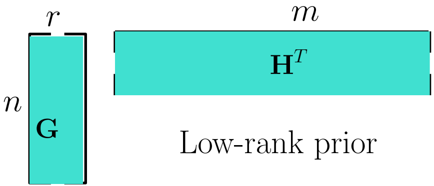
June 2015

Example 1: matrix completion

$$m \text{ Users}$$

$$n \text{ Movies} \begin{bmatrix} ? & ? & * & ? \\ * & * & ? & * \\ ? & * & * & ? \\ * & ? & * & ? \end{bmatrix}$$

$\approx$



$$(n + m - r)r, \; r \ll (m, n)$$

(**G** and **H** are full column rank matrices)

[Netflix Challenge, 2006]

Example 1: matrix completion as a rank constrained error minimization problem

Minimize the error to the known ratings,

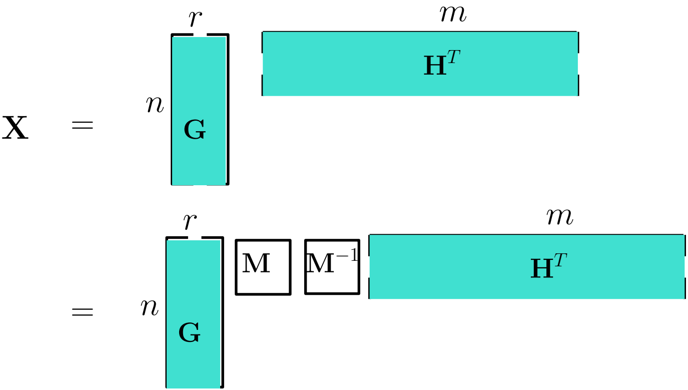
$$\min_{\mathbf{X} \in \mathbb{R}^{n\times m}} \sum_{(i,j)\in\Omega} (\widetilde{\mathbf{X}}_{ij} - (\mathbf{X})_{ij})^2$$
$$\text{subject to} \quad \text{rank}(\mathbf{X}) = r$$

$$\xrightarrow[]{Fixed-rank,\ \mathbf{X}=\mathbf{G}\mathbf{H}^T} \min_{\substack{\mathbf{G} \in \mathbb{R}_*^{n\times r} \\ \mathbf{H} \in \mathbb{R}_*^{m\times r}}} \sum_{(i,j)\in\Omega} (\widetilde{\mathbf{X}}_{ij} - (\mathbf{G}\mathbf{H}^T)_{ij})^2.$$

Ω is the set of known ratings.

$\mathbb{R}_*^{m\times r}$ denotes full rank $m \times r$ matrices.

Example 1: fixed-rank parameterizations have structured symmetries



M is r-by-r non-singular.

Other fixed-rank parameterizations have symmetries too.

Example 1: matrix completion as a least-squares problem with symmetries

$$\min_{\substack{\mathbf{G} \in \mathbb{R}_*^{n \times r} \\ \mathbf{H} \in \mathbb{R}_*^{m \times r}}} \sum_{(i,j) \in \Omega} (\widetilde{\mathbf{X}}_{ij} - (\mathbf{G}\mathbf{H}^T)_{ij})^2 \Longrightarrow \text{Optimization on } [\mathbf{G}, \mathbf{H}]$$

- Equivalence classes:
  $[\mathbf{G}, \mathbf{H}] := \{(\mathbf{G}\mathbf{M}^{-1}, \mathbf{H}\mathbf{M}^T) : \mathbf{M} \text{ non-singular } \in \mathrm{GL}(r)\}$.

- Explicit computations: $(\mathbf{G}, \mathbf{H}) \in \mathbb{R}_*^{n \times r} \times \mathbb{R}_*^{m \times r}$
  Implicit optimization: $[\mathbf{G}, \mathbf{H}] \in \mathbb{R}_*^{n \times r} \times \mathbb{R}_*^{m \times r} / \mathrm{GL}(r)$.

Example 2: eigenvalue problem

$$\max_{x \in \mathbb{R}^n} \quad \frac{x^T \mathbf{A} x}{x^T x}$$

- cost unchanged under map $x \mapsto \alpha x$, α non-zero scalar $\in \mathbb{R}_*$.
- Solutions not isolated, i.e., but are equivalence classes $[x] = \{\alpha x : \alpha \in \mathbb{R}_*\}$ (real projective space)
- An interpretation: direction is important and not length.

Explicit computations: $\mathbb{R}^n$, but
Implicit optimization: $\mathbb{R}^n/\mathbb{R}_*$.

Example 2: eigenvalue problem, the block case

- Generalization: r-dominant eigenvalue-eigenvector pairs:

$$\begin{array}{ll} \max\limits_{\mathbf{X}\in\mathbb{R}^{n\times r}} & \mathrm{Trace}(\mathbf{X}^T\mathbf{A}\mathbf{X}) \\ \text{subject to} & \mathbf{X}^T\mathbf{X}=\mathbf{I}. \end{array}$$

- Symmetry under map $\mathbf{X}\mapsto\mathbf{X}\mathbf{O}$ such that $\mathbf{O}\mathbf{O}^T=\mathbf{O}^T\mathbf{O}=\mathbf{I}$, denoted by $\mathcal{O}(r)$.

- Interpretation: subspace is important and not basis vectors.

- Search space is Grassmann manifold $\mathrm{St}(r,n)/\mathcal{O}(r)$ [Edelman et al., 1998; Absil et al., 2008].

Problems have two fundamental structures: least-squares and symmetries

- $$\begin{array}{lll} \min\limits_{x \in \mathcal{M}} & f(x) & \leftarrow \text{least} - \text{squares cost} \\ \text{subject to} & [x] \in \mathcal{M}/\sim . & \leftarrow \text{equivalence classes on } \mathcal{M} \end{array}$$

Both $\mathcal{M}$ and $\sim$ result from interplay of few matrix manifolds.

$\mathcal{M}/\sim$ has quotient manifold structure.

Rank constraint	**Orthogonality constraints**
$\mathrm{St}(r, n)$	$\mathrm{St}(r, n)$
$\mathbb{R}_*^{n \times r}$	$\mathbb{R}_*^{n \times r}$
$\mathcal{O}(r)$	$\mathcal{O}(r)$
$\mathrm{GL}(r)$	$\mathrm{GL}(r)$
$\mathrm{S}_{++}(r)$	

A number of applications relate to quadratic optimization with orthogonality and rank constraints

$$\begin{array}{ll} \min & \mathrm{Trace}(\mathbf{X}^T\mathbf{A}\mathbf{X}) \\ \text{subject to} & \mathbf{X}^T\mathbf{X} = \mathbf{I} \end{array}$$

Power, inverse, RQI, ...
PCA, Sparse PCA, ...
Max-cut
Procrustes

$$\begin{array}{ll} \min & \|\mathbf{X} - \mathbf{X}^\star\|_F^2 \\ \text{subject to} & \mathrm{rank}(\mathbf{X}) = r \end{array}$$

Linear matrix equations, e.g., Lyanpunov
Matrix completion, Tensor completion

Optimization on manifold framework has gained much attention lately

$$\begin{array}{ll} \min & f(x) \\ \text{subject to} & x \in \mathcal{M}. \end{array}$$

- Nonlinear optimization methods: SQP.

- Geometric methods: Optimization on manifolds [Smith, 1994; Edelman et al., 1998; Absil et al., 2008]

  + Manopt: a Matlab toolbox at Manopt.org [Boumal et al., 2014].

Our contribution

- Exploit cost and constraints with symmetries structures to develop efficient algorithms.

- We propose a notion of preconditioning on manifolds with
  - metric tuning (identifying a good inner product)
  - exploit second-order information in first-order algorithms.

Outline

- Motivation for a geometric framework for constraints with symmetries
- Exploiting structures
  - Metric tuning
  - Quadratic optimization with orthogonality and rank constraints
- Algorithms for low-rank matrix completion with fixed-rank constraint

Outline

- Motivation for a geometric framework for constraints with symmetries
- Exploiting structures
  - Metric tuning
  - Quadratic optimization with orthogonality and rank constraints
- Algorithms for low-rank matrix completion with fixed-rank constraint

Two complementary views of optimization

$$\min_{x \in \mathbb{R}^n} \quad f(x)$$
$$\text{subject to} \quad h(x) = 0$$

Sequential quadratic programming

(constraints are embedded into the cost function)

$$\max_{\lambda \in \mathbb{R}^p} \min_{x \in \mathbb{R}^n} \quad f(x) - \langle \lambda, h(x) \rangle,$$

where λ is the Lagrange multiplier

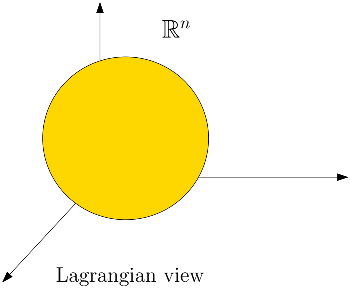
Riemannian framework

(constraints are encoded into the seach space)

$$\min_{x \in \mathcal{M}} \quad f(x),$$

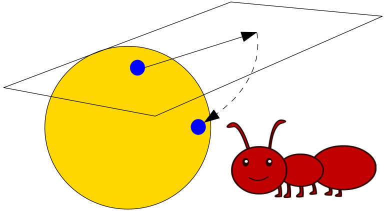
where $\mathcal{M} = \{x : h(x) = 0\}$ has dimension p

Two complementary views of optimization



Lagrangian view

(Constrained viewpoint)



Manifold optimization view

(An ant-on-manifold view: unconstrained viewpoint)

Sequential quadratic programming (SQP)

$$\min_{x \in \mathbb{R}^n} \quad f(x) \quad \text{subject to} \quad h(x) = 0.$$

1. Compute search direction ζ_x^*,

$$\begin{array}{ll} \arg\min_{\zeta_x \in \mathbb{R}^n} & f(x) + \underbrace{\langle f_x(x), \zeta_x \rangle}_{\text{first-order}} + \frac{1}{2} \underbrace{\langle \zeta_x, \mathrm{D}^2 \mathcal{L}(x, \lambda_x)[\zeta_x] \rangle}_{\text{second-order}} \\ \text{subject to} & \underbrace{\mathrm{D}h(x)[\zeta_x] = 0}_{\text{linearization}}. \qquad (\langle \cdot, \cdot \rangle \text{ is scalar product}) \end{array}$$

2. Next x_+ is obtained by projecting $x + s\zeta_x^*$ onto $h(x) = 0$.

Lagrangian function $\mathcal{L}(x, \lambda) = f(x) - \langle \lambda, h(x) \rangle$.

Estimate $\lambda_x = (h_x(x)(h_x(x))^T)^{-1} h_x(x) f_x(x)$ locally [Nocedal and Wright (2006)].

A critical assumption of SQP is not satisfied for constraints with symmetries

- Competitive algorithm near the minimum.
- A well-defined problem with unique solution when $\langle \zeta_x, \mathrm{D}^2\mathcal{L}(x, \lambda_x)[\zeta_x] \rangle > 0$ on $\mathrm{D}h(x)[\zeta_x] = 0$.
- Condition not satisfied for a search space with symmetries $\langle \zeta_x, \mathrm{D}^2\mathcal{L}(x, \lambda_x)[\zeta_x] \rangle \geq 0$.

SQP fails on the eigenvalue problem

$$\min_{x \in \mathbb{R}^n} \quad -\frac{x^T \mathbf{A} x}{x^T x}$$

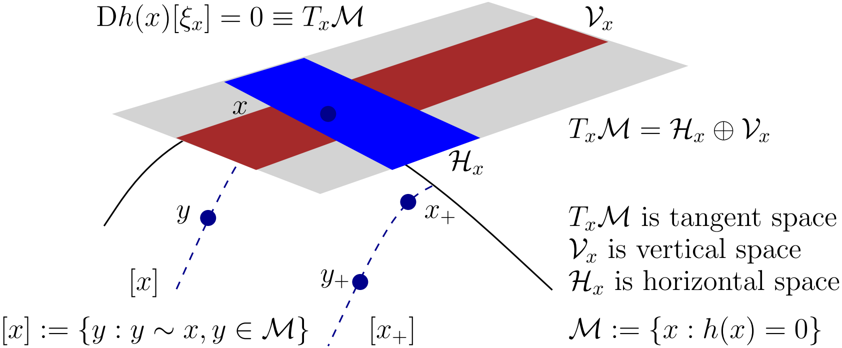
1. Compute the search direction

$$\zeta_x^* = \operatorname*{arg\,min}_{\zeta_x \in \mathbb{R}^n} f(x) + \langle f_x(x), \zeta_x \rangle + \frac{1}{2} \langle \zeta_x, \mathrm{D}^2 \mathcal{L}(x)[\zeta_x] \rangle$$

$$\Rightarrow \quad \zeta_x^* = x. \quad \text{(true for any homogenous function)}$$

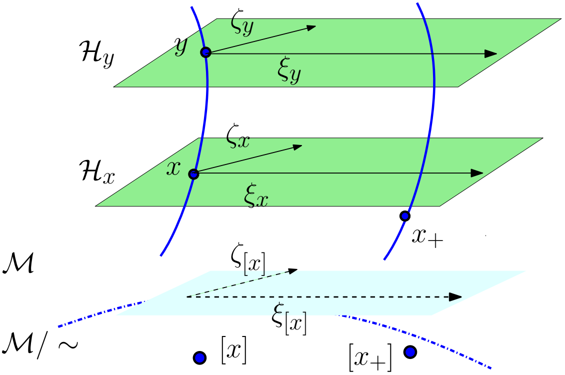
2. Consequently, the next iterate $x_+ = x + s\zeta_x^* = (1+s)x$, i.e., $x_+ = (1+s)x$.

Why SQP fails for constraints with symmetries?



- SQP: $\zeta_x^* = \underset{\zeta_x \in T_x\mathcal{M}}{\arg\min} f(x) + \langle f_x(x), \zeta_x \rangle + \frac{1}{2}\langle \zeta_x, \mathrm{D}^2\mathcal{L}(x)[\zeta_x] \rangle \in \mathcal{V}_x$.
- Resolve: exclude $\mathcal{V}_x$, and only $\mathcal{H}_x$ is relevant.

The Riemannian optimization framework requires a Riemannian metric



- The Riemannian framework enables us to separate $\mathcal{H}_x$ and $\mathcal{V}_x$.
- Riemannian metric that is invariant to $[x]$.

The Riemannian steepest-descent algorithm

$$\min_{x \in \mathcal{M}} \quad f(x)$$

1. Compute the negative Riemannian gradient $\xi_x = -\mathrm{grad}_x f$ w.r.t Riemannian metric g_x, i.e.,

$$\mathrm{grad}_x f = - \underbrace{\underset{\zeta_x \in T_x\mathcal{M}}{\arg\min}}_{\text{linearization}} \quad f(x) + \underbrace{\langle f_x(x), \zeta_x \rangle}_{\text{first}-\text{order}} + \frac{1}{2} \underbrace{g_x(\zeta_x, \zeta_x)}_{\text{second}-\text{order}} .$$

2. Next iterate x_+ is computed using the retraction, equivalent to projection.

- A well-defined scheme on the quotient manifold $\mathcal{M}/\sim$ of $\mathcal{M}$.

Metric profoundly affects performance of Riemannian gradient algorithm

- Conventional: metric g is only motivated by symmetry search space, but ignores cost.

- As a result, performance of the steepest-descent algorithm is profoundly affected for different cost functions [Manton, 2002].

Outline

- Motivation for a geometric framework for constraints with symmetries

- Exploiting structures
    - Metric tuning
    - Quadratic optimization with orthogonality and rank constraints

- Algorithms for low-rank matrix completion with fixed-rank constraint

Connecting SQP to Riemannian gradient descent

$$\text{SQP}: \quad \underset{\zeta_x \in \mathbb{R}^n}{\arg\min} \quad f(x) + \langle f_x(x), \zeta_x \rangle + \frac{1}{2}\langle \zeta_x, \mathrm{D}^2\mathcal{L}(x,\lambda_x)[\zeta_x]\rangle$$

$$\text{subject to} \quad \mathrm{D}h(x)[\zeta_x] = 0$$

$$\text{Riemann}: \quad \underset{\zeta_x \in T_x\mathcal{M}}{\arg\min} \quad f(x) + \langle f_x(x), \zeta_x \rangle + \frac{1}{2} g_x(\zeta_x, \zeta_x)$$

Theorem

If $x^* \in \mathcal{M}$ is a local minimum of $f : \mathcal{M} \to \mathbb{R}$ on $\mathcal{M}/\sim$, then

(i) $\langle \eta_{x^*}, \mathrm{D}^2\mathcal{L}(x^*, \lambda_{x^*})[\eta_{x^*}]\rangle = 0 \quad \text{for all } \eta_{x^*} \in \mathcal{V}_{x^*}$,

(ii) $\langle \xi_{x^*}, \mathrm{D}^2\mathcal{L}(x^*, \lambda_{x^*})[\xi_{x^*}]\rangle$ *captures the second-order information.*

The second-derivative of the Lagrangian induces a valid metric on the quotient space

Metric induced by Lagrangian, i.e., $\langle \zeta_x, \mathrm{D}^2\mathcal{L}(x, \lambda_x)[\eta_x] \rangle$ is only a pseudometric in $T_x\mathcal{M}$.

What we require:

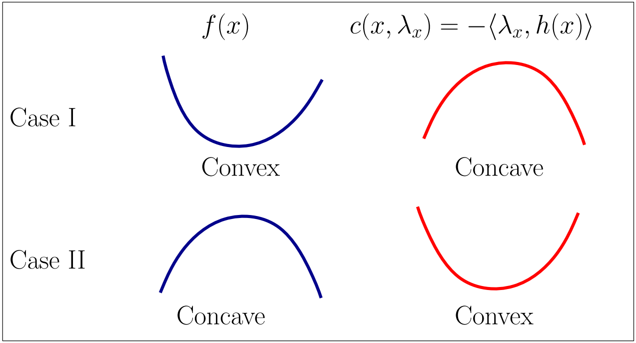
- metric is well-defined in the entire $T_x\mathcal{M}$.
- the metric is a global structure.

Resolve: we exploit the Lagrangian structure further.

Riemannian optimization and local convexity

Lagrangian $\mathcal{L}(x, \lambda) = f(x) - \langle \lambda, h(x) \rangle$, $\langle \cdot, \cdot \rangle$ standard inner product.

$$\begin{aligned} g_x(\xi_x, \eta_x) &= \langle \xi_x, \mathrm{D}^2 \mathcal{L}(x, \lambda_x)[\eta_x] \rangle \\ &= \underbrace{\langle \xi_x, \mathrm{D}^2 f(x)[\eta_x] \rangle}_{\text{cost related}} + \underbrace{\langle \xi_x, \mathrm{D}^2 c(x, \lambda_x)[\eta_x] \rangle}_{\text{constraint related}}. \end{aligned}$$



Quadratic optimization with orthogonality constraints: revisiting the eigenvalue problem

$$\max_{\mathbf{X}\in\mathbb{R}^{n\times r}} \quad \mathrm{Trace}(\mathbf{X}^T\mathbf{A}\mathbf{X}) \qquad \text{subject to} \quad \mathbf{X}^T\mathbf{X} = \mathbf{I}$$

$$\begin{aligned}
\lambda_x &= \mathbf{X}^T\mathbf{A}\mathbf{X} \\
\mathcal{L}(x,\lambda_x) &= \mathrm{Trace}(\mathbf{X}^T\mathbf{A}\mathbf{X})/2 - \langle \lambda_x, \mathbf{X}^T\mathbf{X} - \mathbf{I}\rangle/2 \\
\Rightarrow \qquad \mathcal{L}_x(x,\lambda_x) &= \mathbf{A}\mathbf{X} - \mathbf{X}\lambda_x \\
\Rightarrow \quad \mathrm{D}^2\mathcal{L}(x,\lambda_x)[\xi_x] &= \mathbf{A}\xi_x - \xi_x\lambda_x \\
g_x(\xi_x,\eta_x) &= \underbrace{\langle \xi_x, \mathbf{A}\eta_x\rangle}_{\text{cost related}} - \underbrace{\langle \xi_x, \eta_x\mathbf{X}^T\mathbf{A}\mathbf{X}\rangle}_{\text{constraints related}}.
\end{aligned}$$

- Connects to power, inverse, and Rayleigh quotient iterations [Absil et al., 2002].

Capturing both cost and constraint information in metric leads to a superior performance

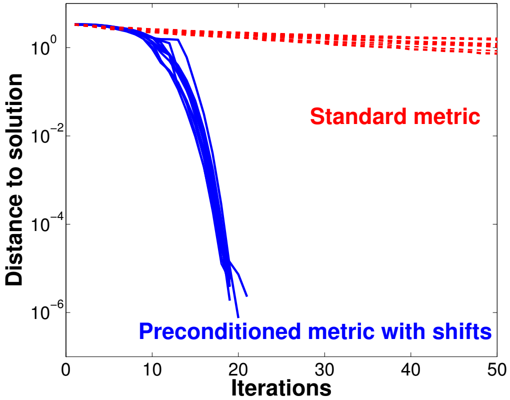


Figure : ω is updated with iterations. metric: $\langle \xi_x, \mathbf{A}\eta_x \rangle - \omega \langle \xi_x, \eta_x \mathbf{X}^T \mathbf{A} \mathbf{X} \rangle$.

Similar story for generalized eigenvalue problem

$$\begin{array}{ll} \min\limits_{\mathbf{X}\in\mathbb{R}^{n\times r}} & \frac{1}{2}\mathrm{Trace}(\mathbf{X}^T\mathbf{A}\mathbf{X}) \\ \text{subject to} & \mathbf{X}^T\mathbf{B}\mathbf{X} = \mathbf{I}, \end{array}$$

$$g_x(\xi_x, \eta_x) = \underbrace{\langle \xi_x, \mathbf{A}\eta_x \rangle}_{\text{cost related}} - \underbrace{\langle \xi_x, \mathbf{B}\eta_x\lambda_x \rangle,}_{\text{constraints related}}$$

$$\lambda_x = \mathrm{Sym}((\mathbf{X}^T\mathbf{B}\mathbf{B}\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{B}\mathbf{A}\mathbf{X})).$$

Quadratic optimization with the rank constraint

$$\begin{array}{ll} \min\limits_{\mathbf{X}\in\mathbb{R}^{n\times m}} & \frac{1}{2}\mathrm{Trace}(\mathbf{X}^T\mathbf{A}\mathbf{X}\mathbf{B}) + \mathrm{Trace}(\mathbf{X}^T\mathbf{C}) \\ \text{subject to} & \mathrm{rank}(\mathbf{X}) = r. \end{array}$$

- We use the parameterization $\mathbf{X} = \mathbf{G}\mathbf{H}^T$, $\mathbf{G} \in \mathbb{R}_*^{n\times r}$, $\mathbf{H} \in \mathbb{R}_*^{m\times r}$.
- The cost is quadratic and convex in arguments $\mathbf{G}, \mathbf{H}$ individually.

$$\begin{array}{rl} g_x(\xi_x, \eta_x) = & \langle \eta_{\mathbf{G}}, 2\mathbf{A}\mathbf{G}\mathrm{Sym}(\mathbf{H}^T\mathbf{B}\xi_{\mathbf{H}}) + \mathbf{C}\xi_{\mathbf{H}}\rangle \\ & +\langle \eta_{\mathbf{H}}, 2\mathbf{B}\mathbf{H}\mathrm{Sym}(\mathbf{G}^T\mathbf{A}\xi_{\mathbf{G}}) + \mathbf{C}^T\xi_{\mathbf{G}}\rangle \\ & +\underbrace{\langle \eta_{\mathbf{G}}, \mathbf{A}\xi_{\mathbf{G}}\mathbf{H}^T\mathbf{B}\mathbf{H}\rangle + \langle \eta_{\mathbf{H}}, \mathbf{B}\xi_{\mathbf{H}}\mathbf{G}^T\mathbf{A}\mathbf{G}\rangle}_{\text{Block diagonal approximation of } \mathcal{L}_{xx}(x)}. \end{array}$$

  $x = (\mathbf{G}, \mathbf{H})$.
- The block-diagonal choice provides a simpler and convenient metric choice.

The case of symmetric fixed-rank constraint is dealt with similarly

The case of symmetric fixed-rank constraint is handled by making $\mathbf{G} = \mathbf{H}$, i.e.,

$$\mathbf{X} = \mathbf{G}\mathbf{G}^T.$$

The metric tuning ideas follow through.

Outline

- Motivation for a geometric framework for constraints with symmetries
- Exploiting structures
  - Metric tuning
  - Quadratic optimization with orthogonality and rank constraints
- Algorithms for low-rank matrix completion with fixed-rank constraint

Revisiting the matrix completion problem: a simpler cost function to infer the metric

$$\begin{array}{ll} \min\limits_{\mathbf{X} \in \mathbb{R}^{n\times m}} & \sum\limits_{(i,j)\in\Omega} (\widetilde{\mathbf{X}}_{ij} - \mathbf{X}_{ij})^2 \\ \text{subject to} & \text{rank}(\mathbf{X}) = r. \end{array}$$

- Second-order derivative of $\sum_{(i,j)\in\Omega}(\widetilde{\mathbf{X}}_{ij} - \mathbf{X}_{ij})^2$ w.r.t $\mathbf{X} = \mathbf{G}\mathbf{H}^T$ is computationally cumbersome.
- Consider simpler cost function

$$\begin{array}{ll} \min\limits_{\mathbf{X} \in \mathbb{R}^{n\times m}} & \frac{1}{2}\text{Trace}(\mathbf{X}^T\mathbf{X}) - \text{Trace}(\mathbf{X}^T\widetilde{\mathbf{X}}) \equiv \|\mathbf{X} - \widetilde{\mathbf{X}}\|_F^2 \\ \text{subject to} & \text{rank}(\mathbf{X}) = r. \end{array}$$

The matrix scaling acts a computationally efficient preconditioner

- Two parameterizations

$$\mathbf{X} = \mathbf{G}\mathbf{H}^T \quad (\text{two} - \text{factor factorization})$$
$$\mathbb{R}_*^{n \times r} \times \mathbb{R}_*^{m \times r}$$
$$\mathbf{X} = \mathbf{U}\mathbf{R}\mathbf{V}^T \quad (\text{SVD} - \text{type factorization}).$$
$$\mathrm{St}(r, n) \times \mathrm{GL}(r) \times \mathrm{St}(r, m)$$

- Novel metrics based on block diagonal approximation

$$g_x(\xi_x, \eta_x) = \langle \eta_{\mathbf{G}}, \xi_{\mathbf{G}} \mathbf{H}^T\mathbf{H} \rangle + \langle \eta_{\mathbf{H}}, \xi_{\mathbf{H}} \mathbf{G}^T\mathbf{G} \rangle$$
$$g_x(\xi_x, \eta_x) = \langle \eta_{\mathbf{U}}, \xi_{\mathbf{U}} \mathbf{R}\mathbf{R}^T \rangle + \langle \eta_{\mathbf{R}}, \xi_{\mathbf{R}} \rangle + \langle \eta_{\mathbf{V}}, \xi_{\mathbf{V}} \mathbf{R}^T\mathbf{R} \rangle$$

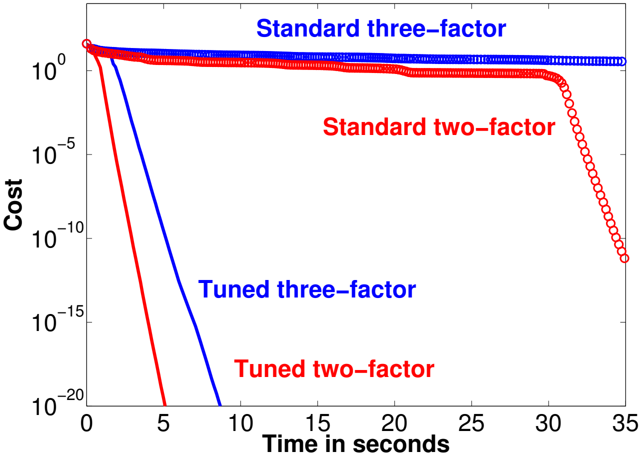
$x = (\mathbf{G}, \mathbf{H})$, $x = (\mathbf{U}, \mathbf{R}, \mathbf{V})$.

We develop conjugate gradient algorithms

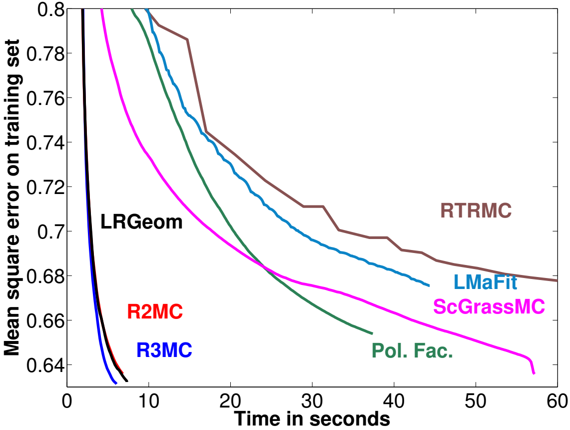
- The matrix scaling connect to various state-of-the-art algorithms, e.g., scaled ALS [Wen et al., 2012] and scaled subspace iteration [Ngo and Saad, 2012] .

- Our algorithms, R2MC and R3MC, have shown competitive performance.

Under high sampling, the performance of metrics is distinctly different

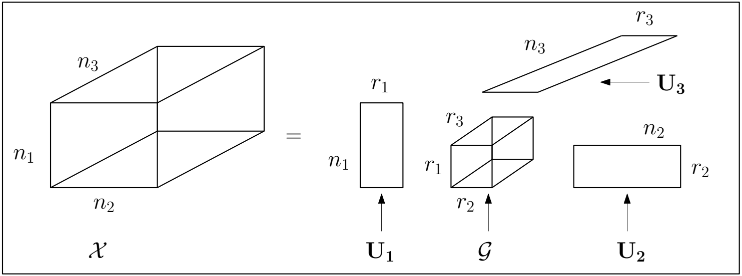


Movielens-1M dataset



(Similar conclusion also obtained on the Netflix and Yahoo datasets.)

Extension to tensors: Tucker decomposition



$$\mathcal{X} = [\![\mathcal{G}; \mathbf{U_1}, \mathbf{U_2}, \mathbf{U_3}]\!] = \mathcal{G} \times_\mathbf{1} \mathbf{U_1} \times_\mathbf{2} \mathbf{U_2} \times_\mathbf{3} \mathbf{U_3},$$

where
(r_1, r_2, r_3) is the multilinear rank of $\mathcal{X}$ and
$\mathbf{U_1}$, $\mathbf{U_2}$, and $\mathbf{U_3}$ are on the Stiefel manifold.

$$(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \mathcal{G}) \mapsto (\mathbf{U}_1\mathbf{O}_1, \mathbf{U}_2\mathbf{O}_2, \mathbf{U}_3\mathbf{O}_3, \mathcal{G} \times_1 \mathbf{O}_1^T \times_2 \mathbf{O}_2^T \times_3 \mathbf{O}_3^T)$$

Riemannian preconditioning for tensor completion

Red is ours. . .

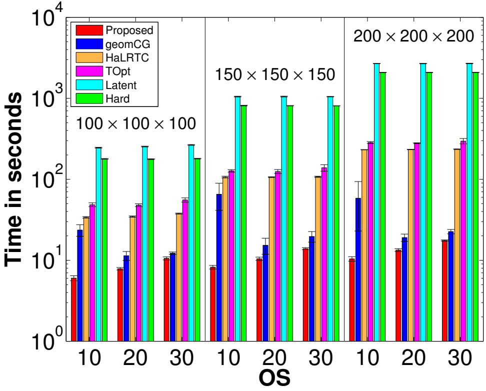
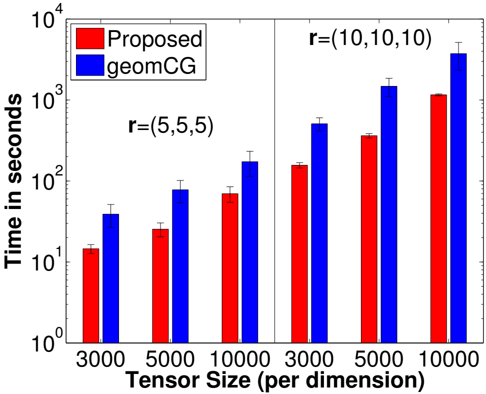


Figure : Small-scale instances of rank $(10, 10, 10)$. OS denotes the over sampling ratio.

Tensor completion: large-scale instances

Red is ours...



Preconditioning:

BM, R Sepulchre (2014) Riemannian preconditioning. Tech. rep., arXiv:1405.6055.

Matrix completion:

BM, R Sepulchre (2014) R3MC: A Riemannian three-factor algorithm for low-rank matrix completion. In: the proceedings of the 53rd IEEE Conference on Decision and Control (CDC).

BM, K Adithya Apuroop, R Sepulchre (2012) A Riemannian geometry for low-rank matrix completion. Tech. rep., arXiv:1211.1550.

Tensor completion:

H Kasai and BM (2015) Riemannian preconditioning for tensor completion. arXiv:1506.02159.

Summary

- Categorized popular problems into quadratic optimization with rank and orthogonality constraints.
- Importance of a geometric framework.
- Question of selecting a metric addressed. Least-squares with rank and/or orthogonality constraints.
- Classical Power, inverse, and Rayleigh quotient iterations interpreted.
- Concrete large-scale algorithms developed for low-rank matrix completion with novel metric.

Riemannian preconditioning

Bamdev Mishra

Joint work with Rodolphe Sepulchre

University of Liège, Belgium
University of Cambridge, UK

Workshop on low-rank optimization and applications
Hausdorff Center for Mathematics, Bonn, Germany

June 2015

References

[Absil et al., 2002] PA Absil, R Mahony, R Sepulchre , P Van Dooren (2002) A Grassmann-Rayleigh quotient iteration for computing invariant subspaces. SIAM Review.

[Absil et al., 2008] PA Absil, R Mahony, R Sepulchre (2008) Optimization Algorithms on Matrix Manifolds. Princeton University Press, Princeton, NJ.

[Boumal and Absil, 2014] N Boumal N, PA Absil (2014) Low-rank matrix completion via preconditioned optimization on the Grassmann manifold. Linear Algebra and its Applications.

[Boumal et al., 2014] N Boumal, BM, PA Absil, R Sepulchre (2014) Manopt, a Matlab toolbox for optimization on manifolds. Journal of Machine Learning Research.

[Edelman et al., 1998] A Edelman, T Arias, S Smith (1998) The geometry of algorithms with orthogonality constraints. SIAM Journal on Matrix Analysis and Applications.

[Manton, 2002] J Manton (2002) Optimization algorithms exploiting unitary constraints. IEEE Transactions on Signal Processing.

[Ngo and Saad, 2012] Ngo TT, Saad Y (2012) Scaled gradients on Grassmann manifolds for matrix completion. NIPS.

[Nocedal and Wright, 2006] J Nocedal, SJ Wright (2006) Numerical Optimization, 2nd edn. Springer, New York, USA.

[Wen et al., 2012] Z Wen, W Yin, Y Zhang (2012) Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm, Mathematical Programming Computation.