A Riemannian approach to large-scale constrained least-squares with symmetries

Bamdev Mishra

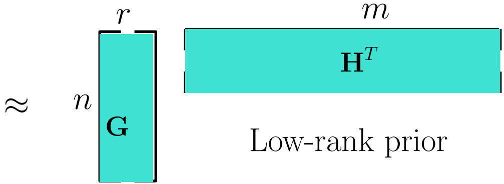
PhD Defense
University of Liège, Belgium

20 October 2014

Example 1: matrix completion

$$m \text{ Users}$$

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



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

(**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,

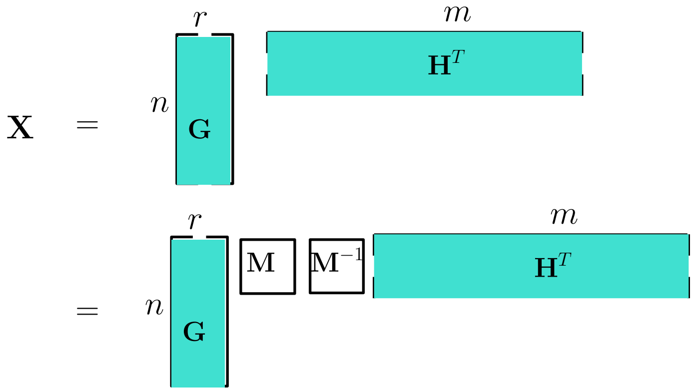
$$\begin{array}{ll} \min\limits_{\mathbf{X} \in \mathbb{R}^{n \times m}} & \sum\limits_{(i,j) \in \Omega} (\widetilde{\mathbf{X}}_{ij} - (\mathbf{X})_{ij})^2 \\ \text{subject to} & \operatorname{rank}(\mathbf{X}) = r \end{array}$$

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

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

- 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 is a non-zero scalar (set denoted by $\mathbb{R}_*$).
- 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 for computing r-dominant eigenvalue-eigenvector pairs:

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

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

- An interpretation: the subspace is important and not the vectors characterizing it.

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

Problems have two fundamental structures of 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**
$\mathrm{St}(r, n)$	$\mathrm{St}(r, n)$
$\mathbb{R}_*^{n \times r}$	$\mathbb{R}_*^{n \times r}$
$\mathcal{O}(r)$	$\mathcal{O}(r)$
$\mathrm{GL}(r)$	$\mathrm{GL}(r)$
$\mathrm{S}_{++}(r)$	

The main contribution of the thesis

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

Our main contribution: to stress the benefit of a Riemannian structure that depends on both constraints and cost.

Conventional way: disregards role of cost in deciding the Riemannian structure on the constraints.

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
- Algorithms for trace norm regularized problems

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
- Algorithms for trace norm regularized problems

Two complementary views of optimization

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & f(x) \\ \text{subject to} & h(x) = 0 \end{array}$$

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

Sequential quadratic programming (SQP)

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

1. Compute search direction ζ_x^*,

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

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

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

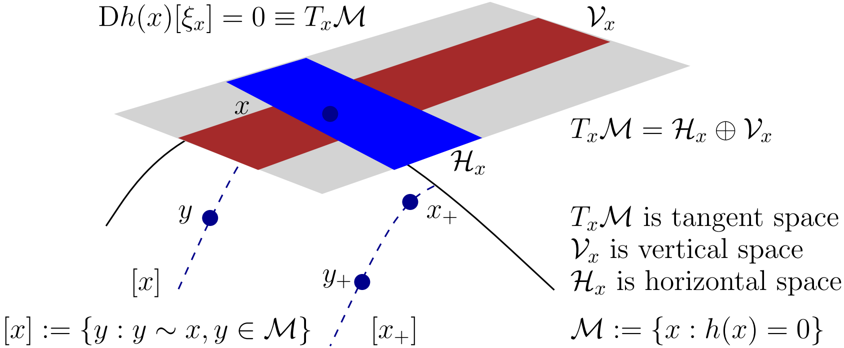
1. Compute the search direction

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

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

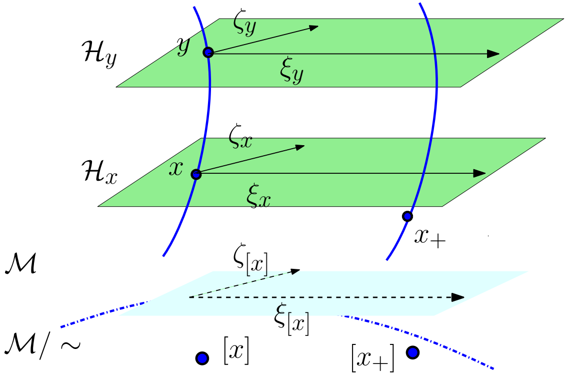
2. Consequently, the next iterate $x_+ = x + s\zeta_x^* = (1+s)x$, i.e., $x_+ = (1+s)x$.

Why SQP fails for constraints with symmetries?



- SQP: $\zeta_x^* = \arg\min_{\zeta_x \in T_x\mathcal{M}} f(x) + \langle f_x(x), \zeta_x \rangle + \frac{1}{2}\langle \zeta_x, \mathrm{D}^2\mathcal{L}(x)[\zeta_x] \rangle \in \mathcal{V}_x$.
- 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 $\mathcal{M}$.

The metric profoundly affects the performance of Riemannian gradient algorithm

- Conventional: the metric g is only motivated by the 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
- Algorithms for trace norm regularized problems

Connecting SQP to Riemannian gradient descent

$$\begin{array}{lll} \mathrm{SQP}: & \underset{\zeta_x \in \mathbb{R}^n}{\arg\min} & f(x) + \langle f_x(x), \zeta_x \rangle + \frac{1}{2}\langle \zeta_x, \mathrm{D}^2\mathcal{L}(x, \lambda_x)[\zeta_x] \rangle \\ & \text{subject to} & \mathrm{D}h(x)[\zeta_x] = 0 \\ \text{Riemann}: & \underset{\zeta_x \in T_x\mathcal{M}}{\arg\min} & f(x) + \langle f_x(x), \zeta_x \rangle + \frac{1}{2} g_x(\zeta_x, \zeta_x) \end{array}$$

Theorem

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

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

$\langle \xi_{x^*}, \mathrm{D}^2\mathcal{L}(x^*, \lambda_{x^*})[\xi_{x^*}] \rangle > 0$ for all $\xi_{x^*} \in \mathcal{H}_{x^*}$, *and*

(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 proper 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 $T_x\mathcal{M}$.

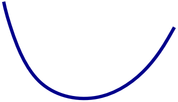
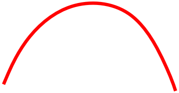
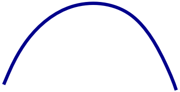
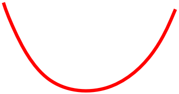
What we require:

- 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

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

	$f(x)$	$c(x, \lambda_x) = -\langle \lambda_x, h(x)\rangle$	Metric candidate
Case I	 Convex	 Concave	$f_{xx}(x) + \omega c_{xx}(x, \lambda_x)$
Case II	 Concave	 Convex	$\omega f_{xx}(x) + c_{xx}(x, \lambda_x)$

weight $\omega \in [0, 1)$ weighs the terms relatively.

Quadratic optimization with orthogonality constraints: revisiting the eigenvalue problem

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

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

- Metric family is parameterized by $\omega \in [0,1)$ for $\mathbf{A}$ is $\succ 0$ or $\prec 0$.
- 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

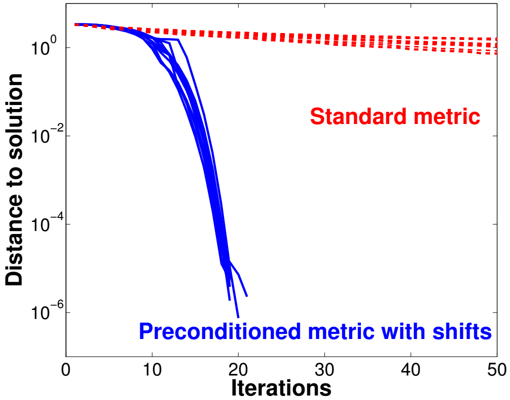


Figure : ω is updated with iterations. metric: $\langle \xi_x, \mathbf{A}\eta_x \rangle - \omega \langle \xi_x, \eta_x \mathbf{X}^T \mathbf{A} \mathbf{X} \rangle$.

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.

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

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

- A family of metrics for $\omega \in [0, 1)$. $\omega = 0$ provides a simpler and a convenient metric choice.

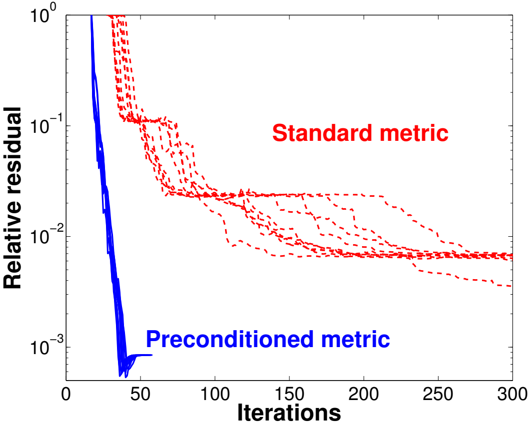
The case of symmetric fixed-rank constraint is dealt with similarly

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.

Lyapunov equation $\mathbf{AXB} + \mathbf{BXA} = \mathbf{C}$, $\mathbf{X}$, the block approximation metric leads a superior performance



Mishra B, Sepulchre R (2014) Riemannian preconditioning. Tech. rep., arXiv:1405.6055 (submitted).

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
- Algorithms for trace norm regularized problems

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

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

- 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

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

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

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

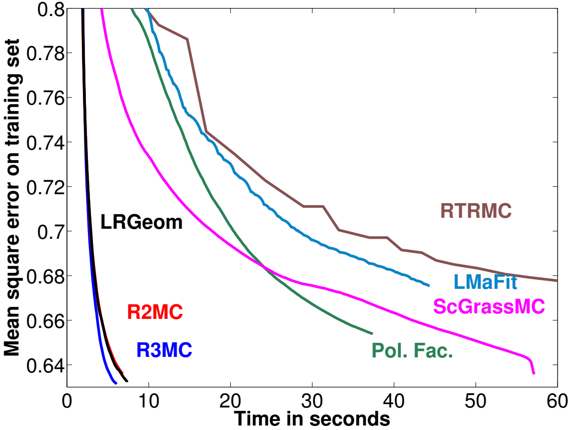
$x = (\mathbf{G}, \mathbf{H})$, $x = (\mathbf{U}, \mathbf{R}, \mathbf{V})$.

We develop conjugate gradient algorithms

- The matrix scaling connect to various state-of-the-art algorithms.

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

Movielens-1M dataset



(Similar conclusion also obtained on the Netflix dataset [Boumal and Absil, 2014].)

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

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

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
- Algorithms for trace norm regularized problems

Trace norm is a convex alternative to rank constraint

$$\min_{\mathbf{X}\in\mathbb{R}^{n\times m}} \underbrace{F(\mathbf{X})}_{\text{smooth convex}} + \lambda \underbrace{\|\mathbf{X}\|_*}_{\text{Non-smooth convex}} .$$

- $\|\mathbf{X}\|_*$ is trace norm, summation of the singular values.
- $\lambda \in \mathbb{R}_+$ enforces low-rank optimal solutions.

Most algorithms rely on singular value thresholding

- Ranks of intermediate iterates are not bounded.
  Resolve: Heuristics like truncation used in practical implementations.

- Conventional warm-restart approach is used to find a regularization path of solutions with different values of λ.

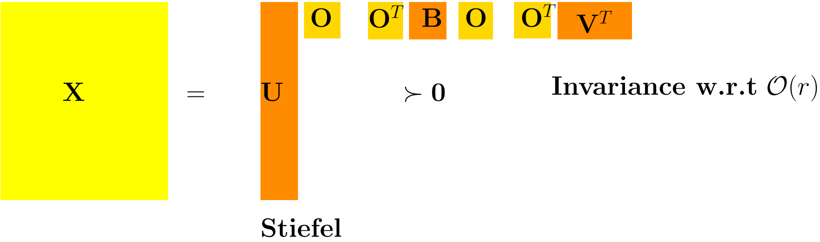
Our approach alternates between fixed-rank optimization and rank-one updating

- Earlier used in computing low-rank solutions to large-scale SDPs [Burer and Monteiro (2003), Journée et al. (2010)].

- Our contribution:
  1. A low-rank factorization that makes $\|\mathbf{X}\|_*$ differentiable.
  2. A trust-region algorithm for rank constraint.
  3. A rank-one descent update.
  4. A duality-gap criterion to monitor convergence.

A fixed-rank parameterization makes the trace norm differentiable

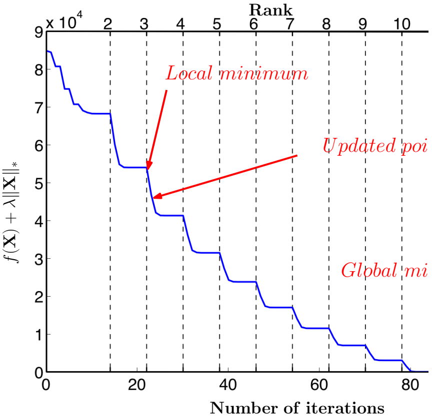
SVD-type decomposition

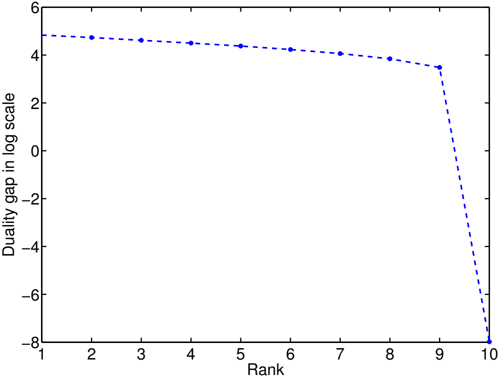


$$\min_{\mathbf{X}\in\mathbb{R}^{n\times m}} F(\mathbf{X})+\lambda \underbrace{\|\mathbf{X}\|_*}_{\text{Non-smooth}} \xrightarrow{Fixed-rank} \min_{[x]=\mathcal{M}/\sim} F(\mathbf{U}\mathbf{B}\mathbf{V}^T)+\lambda \underbrace{\mathrm{Trace}(\mathbf{B})}_{\text{Smooth term}}.$$

$x = (\mathbf{U}, \mathbf{B}, \mathbf{V})$.

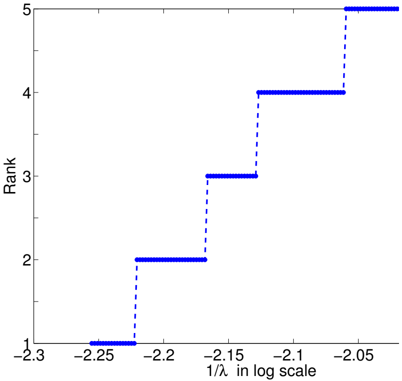
The scheme for a fixed λ

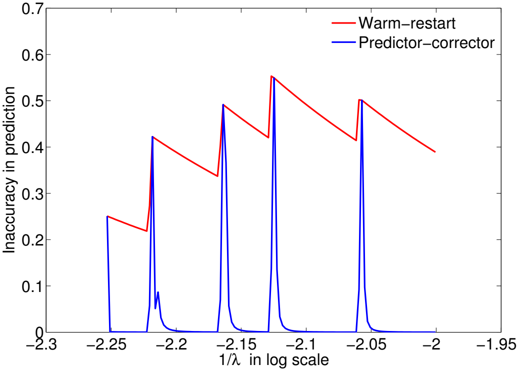




Manifold techniques used to propose a predictor-corrector approach to construct a solution path

$$\mathbf{X}^*(\lambda_i) = \underset{\mathbf{X}\in\mathbb{R}^{n\times m}}{\arg\min}\, F(\mathbf{X}) + \lambda_i\|\mathbf{X}\|_*, i = \{1, 2, \ldots, N\}$$





Mishra B, Meyer G, Bach F, Sepulchre R (2013) Low-rank optimization with trace norm penalty. SIAM Journal on Optimization 23(4):2124–2149.

Summary

- Question of selecting a metric addressed. Least-squares with rank and/or orthogonality constraints.
- Power, inverse, and Rayleigh quotient iterations interpreted.
- Concrete algorithms are developed for ow-rank matrix completion with a novel metric.
- Low-rank optimization for trace norm minimization problem.

Future research directions

We propose three research directions.

1. Affine + manifold constraints? For example, in structured low-rank approximation problems [Markovsky, 2008].
2. Stochastic algorithms on manifolds [Bonnabel, 2013].
3. Block coordinate descents on manifolds? [Shalit and Chechik, 2014].

A Riemannian approach to large-scale constrained least-squares with symmetries

Bamdev Mishra

PhD Defense
University of Liège, Belgium

20 October 2014