# Fixed-rank matrix factorizations and the design of invariant optimization algorithms [1]

Bamdev Mishra and R. Sepulchre

University of Liège, Belgium

21st International Symposium on Mathematical Programming, August 2012

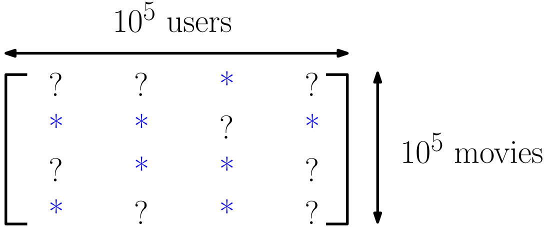
[1] This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its authors. Bamdev Mishra is a research fellow of the Belgian National Fund for Scientific Research (FNRS).

Our interest in this talk is to equip $p - \mathrm{rank}\ \mathbb{R}^{n \times m}$ matrices (denoted by $\mathbb{R}_p^{n \times m}$) with computationally efficient Riemannian geometry

We have in mind:

$$
\begin{aligned}
p &= O(10 \ldots 100) \\
n,\ m &= O(10^4 \ldots 10^8)
\end{aligned}
$$

# Matrix completion: a popular benchmark[2]




$$\text{Sampling} = 0.1\%$$

[2]Netflix Challenge, 2008

# Matrix completion with low-rank assumption

$$\begin{bmatrix} ? & ? & * & ? \\ * & * & ? & * \\ ? & * & * & ? \\ * & ? & * & ? \end{bmatrix}$$

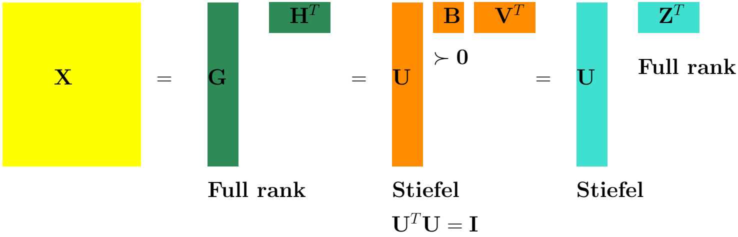
$\approx$

- Riemmanian geometries of fixed-rank matrices
- Geometric optimization
- Invariance
- Numerical examples and summary

Riemmanian geometries of fixed-rank matrices

# Fixed-rank factorizations

$p-\mathrm{rank}$ **X** is factorized[3] [4] as




- Separation of rotational and scaling
- Well-studied individual search spaces

[3] Bonnabel and Sepulchre, Geometric distance and mean for positive semi-definite matrices of fixed rank, 2009

[4] Meyer, Geometric optimization algorithms for linear regression on fixed-rank matrices, Ph.D. thesis, 2011

# Consequences of fixed-rank matrix factorization

- *Symmetry* in the search space
  - Local minima of the cost function are not isolated
  - We require a search space that isolates the minima
- Flexibility in choosing *metrics* on individual search spaces
  - Invariant to rotations of the data?
  - Invariant to scaling of the data?

# Symmetry in the search space

$$\mathbf{X} = \mathbf{G}\,\mathbf{M}^{-1}\,\mathbf{M}^{T}\,\mathbf{H}^{T}$$

**Full rank** ($\mathbf{G}$)

**Invariance w.r.t** $\mathrm{GL}_p$

$$\mathbf{X} = \mathbf{U}\,\mathbf{O}\,\mathbf{O}^{T}\,\mathbf{B}\,\mathbf{O}\,\mathbf{O}^{T}\,\mathbf{V}^{T}$$

$\succ \mathbf{0}$ ($\mathbf{B}$)

**Stiefel** ($\mathbf{U}$)

**Invariance w.r.t** $\mathcal{O}_p$

$$\mathbf{X} = \mathbf{U}\,\mathbf{O}\,\mathbf{O}^{T}\,\mathbf{Z}^{T}$$

**Full rank** ($\mathbf{Z}$)

**Stiefel** ($\mathbf{U}$)

**Invariance w.r.t** $\mathcal{O}_p$

# Symmetry leads to quotient manifolds

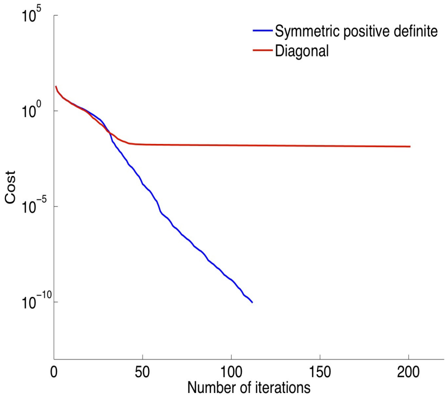
$$\begin{array}{lcl} \mathbb{R}_p^{n\times m} & \sim & \mathbb{R}_p^{n\times p} \times \mathbb{R}_p^{m\times p}/\mathrm{GL}(p) \\ \mathbf{X} = \mathbf{G}\mathbf{H}^T & : & (\mathbf{G},\mathbf{H}) \sim (\mathbf{G}\mathbf{M}^{-1}, \mathbf{H}\mathbf{M}^T) \end{array}$$

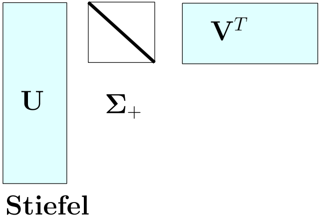
$$\begin{array}{lcl} \mathbb{R}_p^{n\times m} & \sim & \mathrm{St}(p,n) \times S_{++}(p) \times \mathrm{St}(p,m)/\mathcal{O}(p) \\ \mathbf{X} = \mathbf{U}\mathbf{B}\mathbf{V}^T & : & (\mathbf{U},\mathbf{B},\mathbf{V}) \sim (\mathbf{U}\mathbf{O}, \mathbf{O}^T\mathbf{B}\mathbf{O}, \mathbf{V}\mathbf{O}) \end{array}$$

$$\begin{array}{lcl} \mathbb{R}_p^{n\times m} & \sim & \mathrm{St}(p,n) \times \mathbb{R}_p^{m\times p}/\mathcal{O}(p) \\ \mathbf{X} = \mathbf{U}\mathbf{Z}^T & : & (\mathbf{U},\mathbf{Z}) \sim (\mathbf{U}\mathbf{O}, \mathbf{Z}\mathbf{O}) \end{array}$$

# Anything better than SVD?

## SVD as a compact matrix factorization




Completing a generic $1000 \times 1000$ matrix of rank 10 from 6% entries with a gradient search algorithm

# Tuning geometries to problems

- Different factorizations lead to different geometries
- Geometries affect numerical cost and convergence, primarily because of metrics

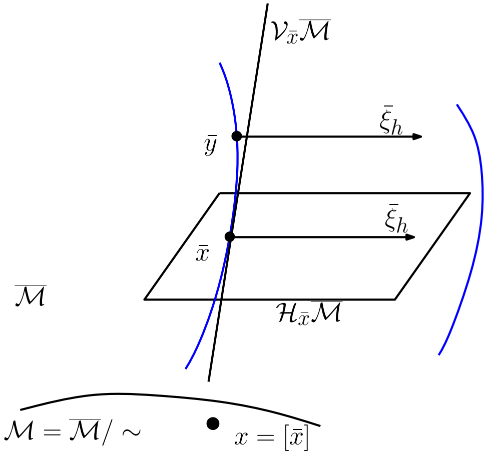
Geometric optimization

# Examples of nice total spaces

$$
\begin{aligned}
\mathbb{R}_p^{n\times m} &\sim \mathbb{R}_p^{n\times p} \times \mathbb{R}_p^{m\times p}/\mathrm{GL}(p) \\
\mathbb{R}_p^{n\times m} &\sim \mathrm{St}(p,n) \times S_{++}(p) \times \mathrm{St}(p,m)/\mathcal{O}(p) \\
\mathbb{R}_p^{n\times m} &\sim \mathrm{St}(p,n) \times \mathbb{R}_p^{m\times p}/\mathcal{O}(p)
\end{aligned}
$$

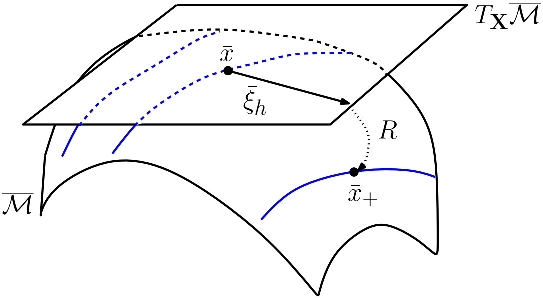
- The three spaces $\mathrm{St}(p,n)$, $S_{++}(p)$ and $\mathbb{R}_p^{n\times p}$ share a common nice structure
- Complete metric spaces for well-chosen metrics

# Geometric view of optimization[5]




[5] Absil et al., Optimization algorithms on matrix manifolds, 2008

# Another view



# The main outcome

- Tangent vectors on the quotient space are represented by horizontal lifts
- Convenient formulas relate the geometric objects in the quotient and in the total space:
  - gradient

$$\overline{\mathrm{grad} f_x} = \mathrm{grad} \overline{f}(\overline{x})$$

  - Riemannian connection

$$\overline{\nabla_{\eta}\xi} = \mathrm{P}^h\left(\overline{\nabla}_{\overline{\eta}}\overline{\xi}\right)$$

  - Hessian

$$\overline{\mathrm{Hess} f(x)} = \mathrm{P}^h\left(\mathrm{Hess}\overline{f}(\overline{x})\right)$$

Invariance

# How to select a metric?

- Should respect symmetry imposed by factorization

- Try to capture invariance in data space (for robustness)

# Invariant metrics along equivalence classes

In the space $\mathrm{St}(p, n)$

$$\langle \eta, \xi \rangle_x = \mathrm{Trace}\left(\eta^T \xi\right)$$

In the space $\mathbb{R}_p^{n \times p}$

$$\langle \eta, \xi \rangle_x = \mathrm{Trace}\left((x^T x)^{-1} \eta^T \xi\right)$$

In the space $S_{++}(p)$

$$\langle \eta, \xi \rangle_x = \mathrm{Trace}\left(x^{-1} \eta x^{-1} \xi\right)$$

- Invariant along the equivalence class (inner invariance)
- Also tries to capture invariance of data matrix $\mathbf{X}$

# Different factoizations capture different invariance of the data space

Consider the transformation $\mathbf{X} \longrightarrow \mathbf{AXB}$ where $\mathbf{X} \in \mathbb{R}_p^{n\times m}$, $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $\mathbf{B} \in \mathbb{R}^{m\times m}$

1. $\mathbf{X} = \mathbf{UBV}^T$

$$\mathbf{A} = \mathbf{R_1 U}^T, \mathbf{B} = \mathbf{R_2 V}^T, \quad \text{where } \mathbf{R_1}, \mathbf{R_2} = \mathbb{R}_p^{p\times p}$$

2. $\mathbf{X} = \mathbf{UZ}^T$

$$\mathbf{A} = \alpha \mathbf{I}, \mathbf{B} = \mathbf{R_2 V}^T, \quad \text{where } \mathbf{R_2} = \mathbb{R}_p^{p\times p} \text{ and } \alpha \in \mathbb{R}_+$$

3. $\mathbf{X} = \mathbf{GH}^T$

$$\mathbf{A} = \alpha \mathbf{I}, \mathbf{B} = \beta \mathbf{I}, \quad \text{where } \alpha, \beta \in \mathbb{R}_+$$

# Affine invariance for optimization

- In addition, the choice should be influenced by cost function
- Specifically, affine invariance for better convergence
- Connection can be made (and improvised) to nonlinear Gauss-Seidel method for rank-constrained minimization

## Numerical examples and summary

# Influence of metric in $\mathbf{G}\mathbf{H}^T$ factorization

$\|\mathbf{G}\|_F \approx 4\|\mathbf{H}\|_F$

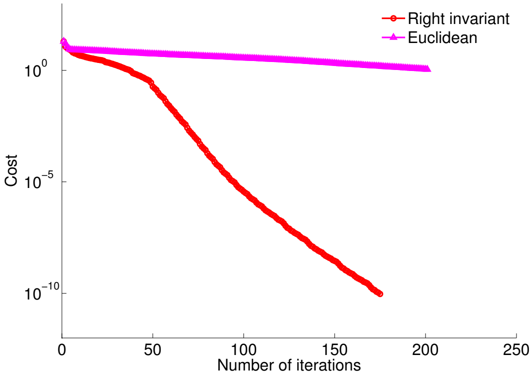
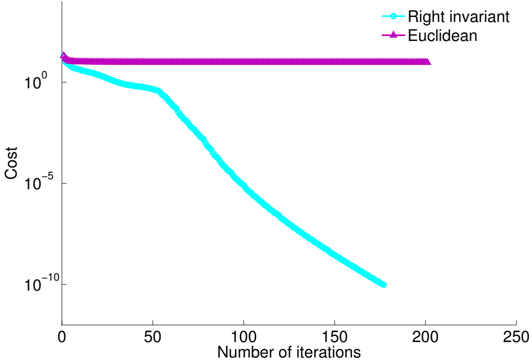


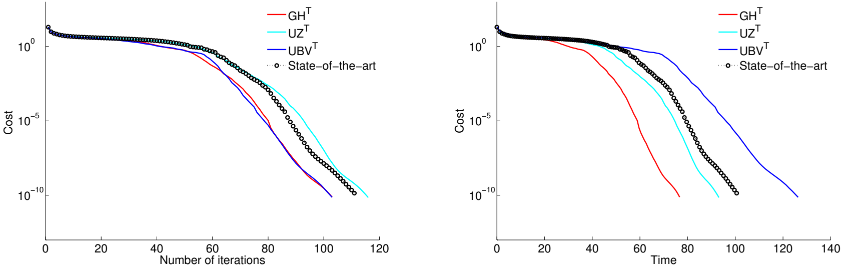

Figure: Completing a generic $5000 \times 5000$ matrix of rank 10 from 1.2% entries with a gradient search algorithm

# In $\mathbf{UZ}^T$ factorization, the factors are alreay misbalanced!



# The designed algorithms compete with state-of-the-art[6]

Geometric conjugate gradient schemes
$30000 \times 30000$ of rank 10 with 0.2% entries




[6] Vandereycken, Low-rank matrix completion by Riemannian optimization, Techreport, 2011

# Finally, ...

- We have studied 3 factorization models, namely, $\mathbf{GH}^T$, $\mathbf{UBV}^T$ and $\mathbf{UZ}^T$
- Different factorizations capture different transformation of the data space
- For specific applications (like matrix completion) it is possible to further tune the metrics
- Quotient geometry present a unified framework

# Thank You.