From fixed-rank optimization to nuclear norm minimization

B. Mishra *, G. Meyer *, F. Bach † and R. Sepulchre *

* Department of Electrical Engineering and Computer Science, University of Liège, 4000 Liège, Belgium (B.Mishra@ulg.ac.be)

† INRIA - Sierra Project-Team Ecole Normale Supérieure Paris, France

Introduction

- Identifying low-rank representation of high dimensional data is a crucial task in machine learning applications
- One approach is to use regularization with nuclear norm which is the sum of singular values
- Solve: a general nuclear norm minimization problem

$$\min_{\mathbf{X}\in\mathbb{R}^{n\times m}} \quad f(\mathbf{X}) + \lambda\|\mathbf{X}\|_*$$

  where f is a smooth function, the nuclear norm $\|\mathbf{X}\|_*$ is non-smooth and $\lambda \in \mathbb{R}_+$ enforces low-rank optimal solutions
- Also construct the path traversed by solutions for different values of λ

Issues with current algorithms

- Most of the algorithms use full matrix $\mathbf{X}$ and try to exploit structure, e.g., *sparse* + *low-rank* structure in the case of low-rank matrix completion
- A numerically demanding step is the computation of the singular value thresholding operator
- Ranks of intermediate iterates are not bounded
- Duality gap guarantees are intractable for large datasets
- Naive warm-restart approach is traditionally used to find the solution path for different values of λ

Contributions

- We use an efficient low-rank factorization to reformulate the problem on a Riemannian quotient manifold
- The objective function is differentiable in the search space and a second-order algorithm proposed
- All intermediate iterates are low-rank
- Duality gap computation is numerically feasible even for large datasets
- The solution path is computed using a combination of an efficient first-order predictor-corrector scheme on the fixed-rank manifold and a few number of warm-restarts

Optimization framework

- Goal: Solve the convex program

$$\min_{\mathbf{X}} \quad f(\mathbf{X}) + \lambda\|\mathbf{X}\|_*$$

- Do:
  1. Solve the non-convex

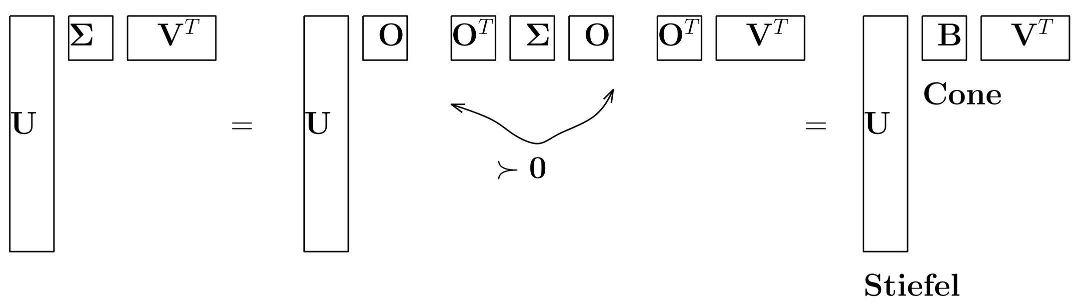
$$\min_{\mathbf{X}} \quad f(\mathbf{X}) + \lambda\|\mathbf{X}\|_* \quad \text{subject to } \operatorname{rank}(\mathbf{X}) = p$$

  2. Relate the local minimum to global minimum of the convex problem
- Since the convex program has a low-rank solution, a few number of smaller non-convex optimization problems are solved

1. Fixed-rank optimization to solve the non-convex problem

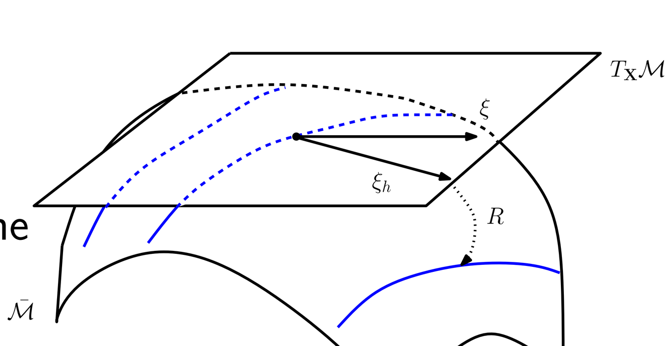
- $p-$rank $\mathbf{X}$ is factorized as $\mathbf{X} = \mathbf{U}\mathbf{B}\mathbf{V}^T$ and hence, the non-convex problem can be written as

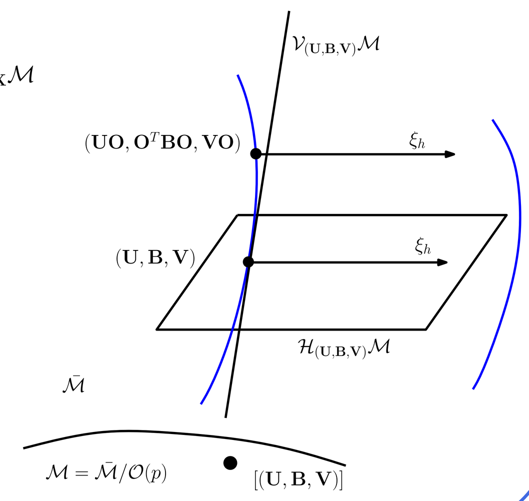
$$\min_{\mathbf{U},\mathbf{B},\mathbf{V}} \quad f(\mathbf{U}\mathbf{B}\mathbf{V}^T) + \lambda \mathrm{Trace}(\mathbf{B}) \quad \text{subject to } (\mathbf{U},\mathbf{B},\mathbf{V}) \in \bar{\mathcal{M}}$$



- Total space: $\bar{\mathcal{M}} = \mathrm{St}(p,n) \times S_{++}(p) \times \mathrm{St}(p,m)$
- Search space is the quotient manifold: $\mathcal{M} = \bar{\mathcal{M}}/\mathcal{O}(p)$
- Horizontal space is representative of the tangent space
- Rotational invariance and scaling invariance with an affine invariant metric g:

$$\text{Stiefel} \rightarrow \mathrm{Trace}(\xi_{\mathbf{U}}^T \eta_{\mathbf{U}}), \quad \text{Cone} \rightarrow \mathrm{Trace}(\mathbf{B}^{-1}\xi_{\mathbf{B}}\mathbf{B}^{-1}\eta_{\mathbf{B}})$$





- A second-order trust-region algorithm is proposed

2. Algorithm to solve the convex problem

- The local minimum of the $p-$rank non-convex problem is embedded into $p+1$ dimensional space by incrementing the rank
- Rank is updated using descent directions i.e., $\mathbf{X}_+ = \mathbf{X} - \beta u v^T$ where u and v are the largest left and right singular vectors of dual variable $\mathbf{S} = \mathrm{Grad}_{\mathbf{X}} f(\mathbf{X})$ and $\beta > 0$
- This update ensures that the saddle point is escaped and the cost is reduced
- Fixed-rank optimization is performed in $p+1$ rank space
- This is continued until the duality gap (stopping criterion) is below a threshold

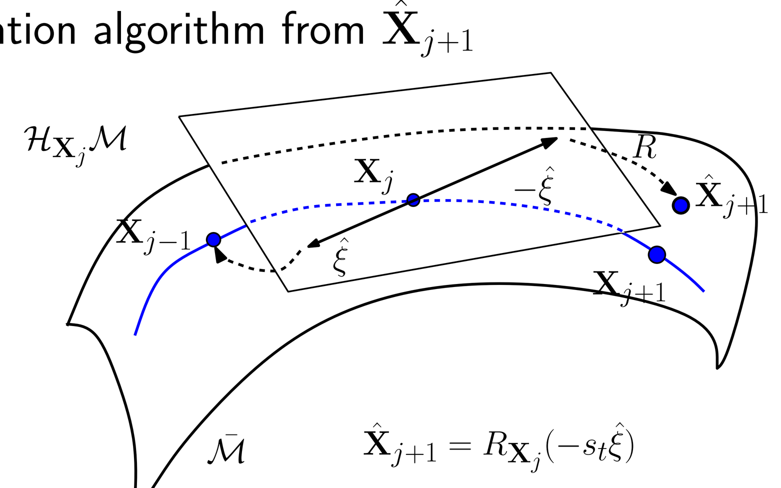
Algorithm to compute solutions for different values of λ

The goal is to efficiently compute the path traversed by solutions $\mathbf{X}_j = \arg\min_{\mathbf{X}} \quad f(\mathbf{X}) + \lambda_j\|\mathbf{X}\|_*$ corresponding to values $\{\lambda_1, \lambda_2, ..., \lambda_N\}$ arranged in decreasing order.

Here, we propose a new first-order predictor-corrector method on the fixed-rank manifold.

- Predictor:
  - If $\mathbf{X}_j$ and $\mathbf{X}_{j-1}$ belong to the same manifold $\mathcal{M}$ then, we define $\hat{\xi} = \mathrm{Project}(R_{\mathbf{X}_j}^{-1}(\mathbf{X}_{j-1}))$ belonging to the horizontal space and $\hat{\mathbf{X}}_{j+1} = R_{\mathbf{X}_j}(-s_t\hat{\xi})$
  - Else $\hat{\mathbf{X}}_{j+1} = \mathbf{X}_j$ (warm-restart)
- Corrector: Initialize nuclear norm minimization algorithm from $\hat{\mathbf{X}}_{j+1}$

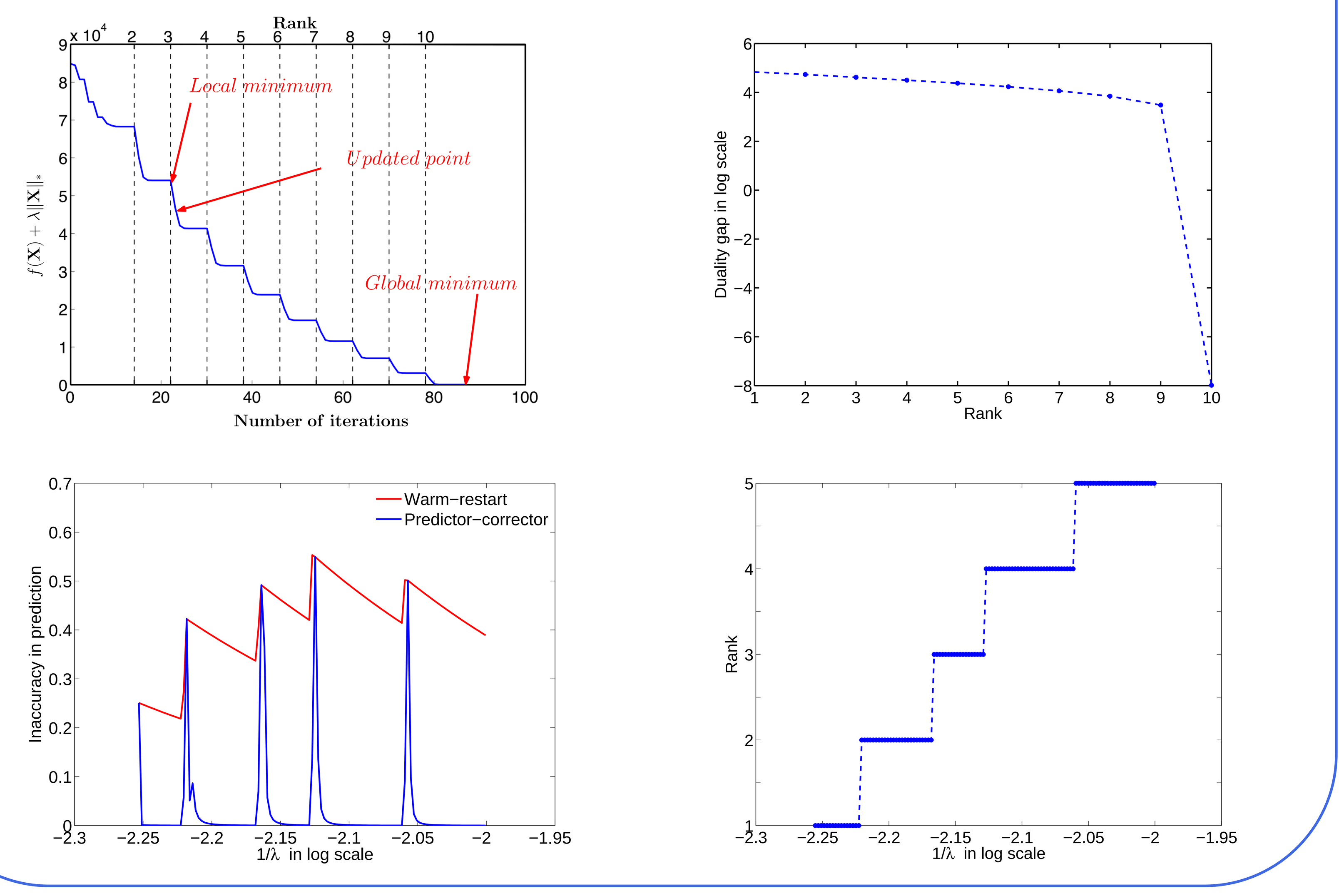
Naive approach uses N restarts where as our algorithm uses r restarts and $N - r$ first-order predictors.



Application to low-rank matrix completion

- $\min_{\mathbf{X}\in\mathbb{R}^{n\times m}} \quad \|\widetilde{\mathbf{X}}_{sampled} - \mathbf{X}_{sampled}\|_F^2 + \lambda\|\mathbf{X}\|_*$
- $\widetilde{\mathbf{X}}_{sampled}$ is the set of known entries
- Dual variable $\mathbf{S} = 2(\mathbf{X}_{sampled} - \widetilde{\mathbf{X}}_{sampled})$
- Numerical complexity per iteration of "Descent-restart + trust-region (TR)" $\propto$ # entries of $\widetilde{\mathbf{X}}$

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



1. P.-A. Absil, R. Mahony, and R. Sepulchre, Optimization Algorithms on Matrix Manifolds, Princeton University Press, 2008
2. S. Bonnabel and R. Sepulchre, Geometric distance and mean for positive semi-definite matrices of fixed rank, SIAM J. Matrix Anal. Appl., 31 (2009), pp. 10551070
3. J.-F. Cai, E. J. Candès, and Z. Shen, A singular value thresholding algorithm for matrix completion, SIAM Journal On Optimization, 20 (2010), pp. 19561982

This poster 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. B. Mishra is supported as an F.R.S.-FNRS research fellow (Belgian Fund for Scientific Research).