A Riemannian Optimization technique for rank inequality constraints

Wen Huang[1], Guifang Zhou[2], Kyle A. Gallivan[2], Paul Van Dooren[1], Pierre-Antoine Absil[1]
[1]Université Catholique de Louvain, [2]Florida State University

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 Science Policy Office.

Problem and Applications

This study considers combining rank inequality constraints with a matrix manifold constraint in a problem of the form

$$\min_{x \in \mathcal{M}_{\leq k}} f(x), \quad (1)$$

where $\mathcal{M}_{\leq k} = \{x \in \mathcal{M} | \text{rank}(x) \leq k\}$ and $\mathcal{M}$ is a submanifold of $\mathbb{R}^{m \times n}$. Numerous applications exist, e.g., [ZW03, FHB04, MLP+06, JHSX11].

Background

Riemannian optimization methods play important roles:

- $\mathcal{M} = \mathbb{R}^{m \times n}$ in most of applications;
- $\mathbb{R}_r^{m \times n} := \{x \in \mathbb{R}^{m \times n} | \text{rank}(x) = r\}$ is a Riemannian manifold.

Existing methods choose the k in (1) a priori. However, it is not easy to choose a suitable k.

- The solution with too small k may be unacceptable;
- The computational time may be unacceptable with too large k.

Contribution

- Generalize the admissible set from $\mathbb{R}_{\leq k}^{m \times n}$ to $\mathcal{M}_{\leq k}$;
- Define an algorithm solving a rank inequality constrained problem while finding a suitable rank for approximation;
- Prove theoretical convergence results;
- Implementations based on Riemannian optimization methods.

Basic Idea

Apply Riemannian optimization methods on a fixed rank manifold $\mathcal{M}_r$ while efficiently and effectively updating the rank r.

Update Rank

Increase rank if next two conditions hold.

- Condition I (angle threshold θ_0):

$$\angle(\text{grad} f_{\text{F}}(x_r), \text{grad} f_r(x_r)) = \theta > \theta_0,$$

- Condition II (difference threshold, ϵ_2):

$$\|\text{grad} f_{\text{F}}(x_r) - \text{grad} f_r(x_r)\| \geq \epsilon_2,$$

where $x_r \in \mathcal{M}_r$, $\text{grad} f_{\text{F}}(x)$ and $\text{grad} f_r(x)$ are the Riemannian gradients with respect to $\mathcal{M}$ and $\mathcal{M}_r$ respectively.

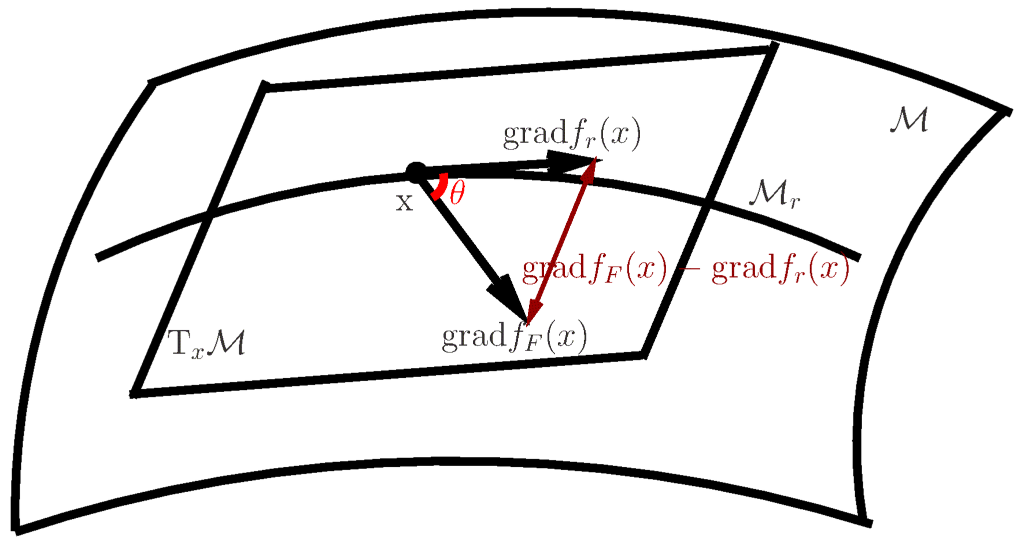

Figure 1. Strategy of increasing the rank.

Rank-Related Objects

The new concepts of rank-related vector and rank-related retraction play an important role in updating the rank and avoiding increasing it excessively.

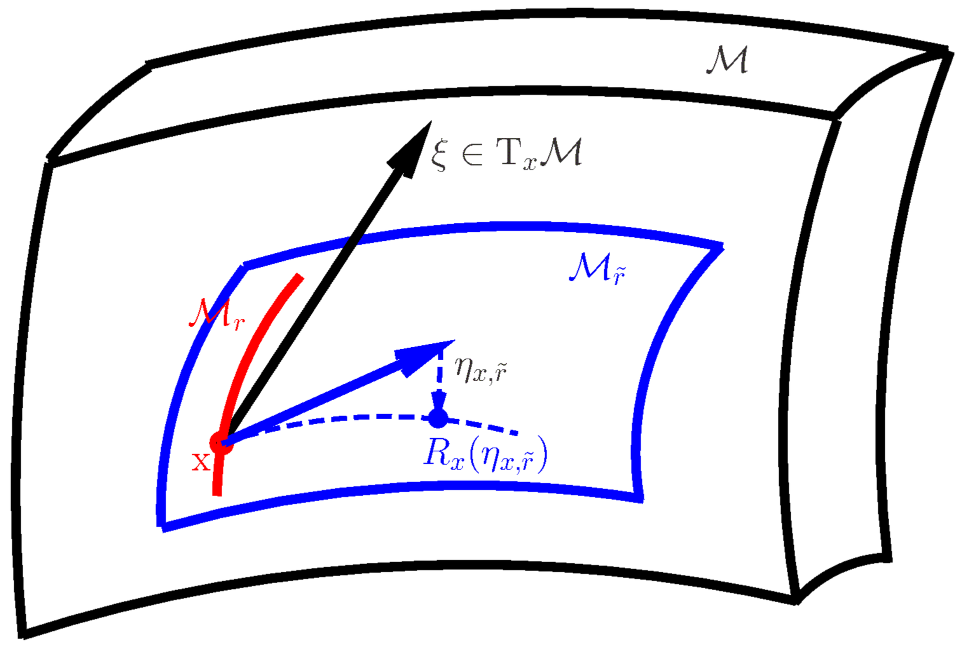

Figure 2. $x \in \mathcal{M}_r$; $r < \tilde{r}$; $\eta_{x,\tilde{r}}$ is a rank-$\tilde{r}$-related vector, i.e., there exists a curve $\gamma(t)$ such that $\gamma(0) = x$, $\dot{\gamma}(0) = \eta_{x,\tilde{r}}$ and $\text{rank}(\gamma(t)) = \tilde{r}$; R is a rank-related retraction, i.e., $\text{rank}(R_x(t\eta_{x,\tilde{r}})) = \tilde{r}$ for $t \in (0, \delta)$, $\delta > 0$.

Algorithm

Algorithm 1

1: **for** n = 0,1, 2,... **do**
2: Approximately optimize f over $\mathcal{M}_r$ with initial point x_n and obtain $\tilde{x}_n$;
3: **if** $\tilde{x}_n$ is not close to a set of lower rank matrices **then**
4: **if** Both Conditions I and II are satisfied **then**
5: Find a $\tilde{r} \in [r, k]$ and obtain a rank-$\tilde{r}$-related vector.
6: Obtain x_{n+1} by applying a line search algorithm along the rank-related vector using a rank-related retraction;
7: **else**
8: If x_{n+1} is accurate enough, stop.
9: **end if**
10: **else**
11: Reduce the rank of $\tilde{x}_n$ if the function value at a lower rank point is nonincreasing; Update r; Obtain next iterate x_{n+1};
12: **end if**
13: **end for**

Main Theoretical Results

Suppose some reasonable assumptions hold:

- (Global Result) The sequence $\{x_n\}$ generated by Algorithm 1 satisfies $\liminf_{n \to \infty} \|P_{\text{T}_{x_n}\mathcal{M}_{\leq k}}(\text{grad} f_{\text{F}}(x_n))\| \leq \left(\sqrt{1 + \frac{1}{\epsilon_1^2}}\right) \epsilon_2$, where $\epsilon_1 = \tan(\theta_0)$.
- (Local Result) The sequence $\{x_n\}$ enters a neighborhood $\mathcal{U}_*$ of a minimizer x_* and remains in $\mathcal{U}_*$. The distance $\text{dist}(x_n, x_*)$ is bounded based on ϵ_1, ϵ_2 and Hess $f_{\text{F}}(x_*)$. The ranks of $\{x_n\}$ are fixed eventually.

Weighted Low Rank Problem

Weighted low rank problem concerns solving

$$\min_{X \in \mathcal{M}_{\leq k}} \|A - X\|_W^2$$

where $\mathcal{M} = \mathbb{R}^{m \times n}$, A is given, $W \in \mathbb{R}^{mn \times mn}$ is symmetric positive definite and $\|A - X\|_W^2 = \text{vec}(A - X)^T W \text{vec}(A - X)$.

Experiments

Algorithm 1 is compared with the state-of-the-art methods for weighted low rank approximation problems.

The matrix A is generated by $A_1 A_2^T \in \mathbb{R}^{10 \times 80}$, where $A_1 \in \mathbb{R}^{10 \times 4}$, $A_2 \in \mathbb{R}^{80 \times 4}$. W is a block diagonal matrix and each block of W is $W_i = U_i \Sigma_i U_i^T \in \mathbb{R}^{10 \times 10}$, where U_i is given by matlab's ORTH and RAND and Σ_i is given by randomly scaling elements from matlab's LOGSPACE.

	k	f	R_err	err	time(s)
(1)	3	2.93_{-01}	8.54_{-02}	2.53_{+00}	1.26_{-1}
	4	3.56_{-29}	9.42_{-16}	2.02_{-14}	1.05_{-2}
	5	7.02_{-29}	1.32_{-15}	2.55_{-14}	9.77_{-3}
(2)	3	2.93_{-01}	8.54_{-02}	2.53_{+00}	8.57_{-1}
	4	3.56_{-29}	9.42_{-16}	2.02_{-14}	2.52_{-2}
	5	4.27_{-29}	1.03_{-15}	2.84_{-14}	2.47_{-2}
(3)	3	2.93_{-01}	8.54_{-02}	2.53_{+00}	8.39_{-1}
	4	3.74_{-26}	3.05_{-14}	1.32_{-12}	3.19_{-2}
	5	1.36_{-22}	1.84_{-12}	1.04_{-10}	3.17_{-2}
(4)	3	2.93_{-01}	8.54_{-02}	2.53_{+00}	6.61_{-1}
	4	4.35_{-29}	1.04_{-15}	2.29_{-14}	5.58_{-2}
	5	6.27_{-29}	1.25_{-15}	2.68_{-14}	7.52_{-2}

Table 1. (1), (2), (3) and (4) denote Algorithm 1, SULS [SU14], EW-TLS [MLP+06] and APM [LPW97] respectively. R_err denotes $\|A - X\|_W / \|A\|_W$ and err denotes $\|A - X\|_F$. The subscript $\pm k$ indicates a scale of $10^{\pm k}$.

References

[FHB04] M. Fazel, H. Hindi, and S. Boyd. Rank minimization and applications in system theory. *Proceedings of American Control Conference*, 2004.
[JHSX11] H. Ji, S.-B. Huang, Z. Shen, and Y. Xu. Robust video restoration by joint sparse and low rank matrix approximation. *SIAM Journal on Imaging Sciences*, pages 1–19, 2011.
[LPW97] W.-S. Lu, S.-C. Pei, and P.-H. Wang. Weighted low-rank approximation of general complex matrices and its application in the design of 2-d digital filters. *IEEE Transactions on Circuits and Systems*, 44(7):650–655, 1997.
[MLP+06] I. Markovsky, M. Luisa Rastello, A. Premoli, A. Kukush, and S. Van Huffel. The element-wise weighted total least-squares problem. *Computational Statistics & Data Analysis*, 50(1):181–209, January 2006. doi:10.1016/j.csda.2004.07.014.
[SU14] R. Schneider and A. Uschmajew. Convergence results for projected line-search methods on varieties of low-rank matrices via lojasiewicz inequality. February 2014. 1402.5284.
[ZW03] Z. Zhang and L. Wu. Optimal low-rank approximation to a correlation matrix. *Linear Algebra and its Applications*, 364:161–187, May 2003. doi:10.1016/S0024-3795(02)00551-7.