Rank-constrained Optimization: A Riemannian Manifold Approach

Guifang Zhou[1], Wen Huang[2], Kyle A. Gallivan[1],
Paul Van Dooren[2], P.-A. Absil[2]

1- Florida State University   2- Université catholique de Louvain

23 April 2015

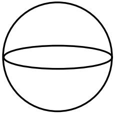
Problem Statements

- Finding an optimum of a rank-constrained problem

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

- $\mathcal{M}_{\leq k} = \{X \in \mathcal{M} | \mathrm{rank}(X) \leq k\}$
- $\mathcal{M}$ is $\mathbb{R}^{m \times n}$ or a submanifold of $\mathbb{R}^{m \times n}$
- Roughly speaking, a manifold is a set endowed with coordinate patches which overlap smoothly, e.g.,

sphere: $\{x \in \mathbb{R}^n | \|x\|_2 = 1\}$.



Motivations

E.g. Compression of an image

- Cost function: $f(X) = \|X - M\|_F^2$, where $M \in \mathbb{R}^{371 \times 600}$ is the matrix representation of the image

Figure : Size : 371×600

Motivations

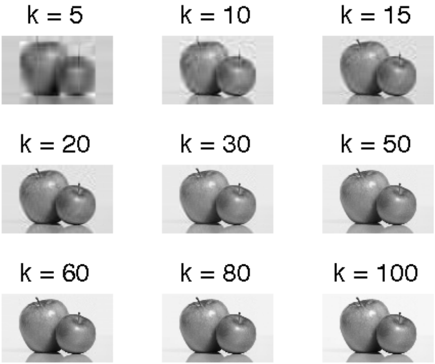


Figure : Low rank approximation of the image

Motivations

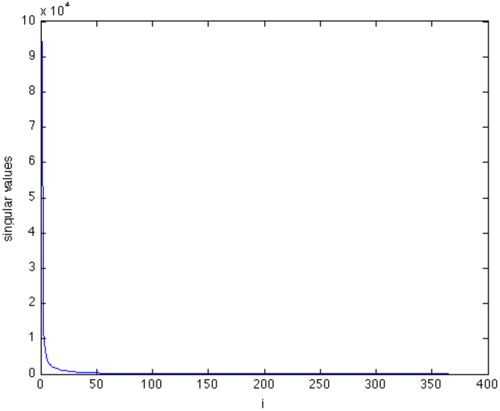


Figure : Singular values of the image

Motivations

Rank-constrained optimization problems have appeared in many areas

- $\mathcal{M} = \mathbb{R}^{m \times n}$
  - Signal and image processing
  - System identification
  - Computational finance
  - Low-dimensional embedding
- $\mathcal{M} \neq \mathbb{R}^{m \times n}$
  - Low-rank approximation of graph similarity matrix
  - Low-rank similarity measure for role model extraction

Existing Methods

- SVD-based algorithms
- Random multistart-type algorithms, e.g. Alternating Projection method, Double Minimization method , EW-TLS , etc.
- Method of alternating between fixed-rank optimization and a simple update to the rank
- Projected line-search method

Challenges

- Efficiency (both time and space)
- Rigorous way for rank updating (explain on the next slide)
- Find the rank independent of k
  - Exact solution
  - Approximate solution

Framework of Modified Riemannian Optimization Method

- $\mathcal{M}$ is a submanifold of $\mathbb{R}^{m \times n}$
- In general, $\mathcal{M}_{\leq k} = \{X \in \mathcal{M} | \mathrm{rank}(X) \leq k\}$ is not a manifold
- $\mathcal{M}_{\leq k} = \mathcal{M}_0 \cup \mathcal{M}_1 \cup \mathcal{M}_2 \cup \cdots \cup \mathcal{M}_k$, where
  $\mathcal{M}_r = \{X \in \mathcal{M} | \mathrm{rank}(X) = r\}$
- Assume $\mathcal{M}_r$ is a manifold
  $\Rightarrow$ General Riemannian optimization algorithms, e.g., RTR-Newton, can be applied on $(\mathcal{M}_r, g)$
- Question: When and how to move from one fixed-rank manifold to another one?

When to increase from one fixed-rank manifold to another one?

Consider two functions

$$f_{\mathrm{F}} : \mathcal{M} \to \mathbb{R} : X \mapsto f_{\mathrm{F}}(X),$$
$$f_r : \mathcal{M}_r \to \mathbb{R} : X \mapsto f_r(X),$$

such that $f = f_{\mathrm{F}}\big|_{\mathcal{M}_{\leq k}}$, $f_r = f|_{\mathcal{M}_r}$.

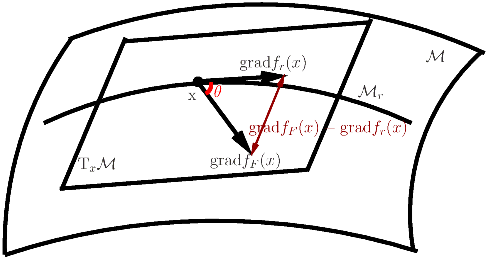
When to increase from one fixed-rank manifold to another one?

- Condition I (angle threshold, θ_0):

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

- Condition II (difference threshold, ϵ):

$$\|\mathrm{grad} f_{\mathrm{F}}(X) - \mathrm{grad} f_r(X)\| \geq \epsilon$$



How to increase from one fixed-rank manifold to another one?

Line-search method

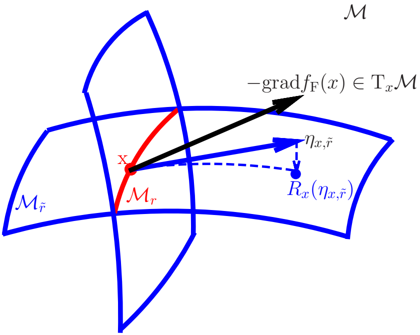
$$X_{i+1} = R_{X_i}(t_i \eta_i),$$

where $t_i \geq 0$ is a step-size, η_i is a search direction and R_{X_i} is a retraction, i.e., $R_{X_i} : \mathrm{T}_{X_i}\mathcal{M} \to \mathcal{M}$ and $\frac{d}{dt} R_{X_i}(t\eta) = \eta$

- How to choose η_i? $\rightarrow$ Rank-related Direction Vector
- How to choose retraction R? $\rightarrow$ Rank-related Retraction

Rank-related Objects

- $\eta_{x,\tilde{r}} \in \mathrm{T}_x\mathcal{M}_{\leq\tilde{r}}$ and $\eta_{x,\tilde{r}} \notin \mathrm{T}_x\mathcal{M}_{\leq\tilde{r}-1}$ is a rank-$\tilde{r}$-related direction vector
- $\eta_{x,\tilde{r}} \in \operatorname{argmin}_{\eta\in\mathrm{T}_x\mathcal{M}_{\leq\tilde{r}}} \|-\mathrm{grad} f_{\mathrm{F}}(x) - \eta\|_F$
- Rank-related retraction $\tilde{R}_x$ satisfies $\tilde{R}_x(t\eta_{x,\tilde{r}}) \in \mathcal{M}_{\tilde{r}}, \forall t \in (0,\delta)$



Rank Reduction

- Truncate rank and retract to $\mathcal{M}$
- May increase the function value
- Do not destroy the progress that is made in the previous rank increasing step, i.e.,

$$f(X_{i+1}) - f(X_s) \leq c(f(X_{s+1}) - f(X_s)),$$

where s is such that the latest rank increase was from X_s to X_{s+1}, $0 < c < 1$.

Main Theoretical Results

- Parameter ϵ is used to adjust the accuracy of the approximate solutions.
- (Global Convergence) Suppose some reasonable assumptions hold. Then the sequence $\{X_i\}$ generated by modified Riemannian Optimization method satisfies

$$\liminf_{i \to \infty} \|P_{\mathrm{T}_{X_i}\mathcal{M}_{\leq k}}(\mathrm{grad} f_F(X_i))\| \leq \left(\sqrt{1 + \frac{1}{\tan(\theta_0)^2}}\right) \epsilon.$$

Weighted Low-rank Approximation

- Problem Formulation

$$\underset{X\in\mathcal{M}_{\le k}}{\operatorname{argmin}} \|B-X\|_W^2, \quad \|B-X\|_W^2 = \operatorname{vec}\{B-X\}^T W \operatorname{vec}\{B-X\}$$

- $\mathcal{M} = \mathbb{R}^{m\times n}$
- $B \in \mathbb{R}^{m\times n}$ is a given data matrix
- $W \in \mathbb{R}^{mn\times mn}$ is a positive definite symmetric weighting matrix
- vec$\{A\}$ denotes the vectorized form of A

Experiment

Comparison of different methods.

- The data matrix $B = B_1 B_2^T$ is a random generated 80×10 matrix with rank 5, $B_1 \in \mathbb{R}^{80\times 5}$, $B_2 \in \mathbb{R}^{10\times 5}$.
- The weighting matrix $W \in \mathbb{R}^{800\times 800}$ is a positive definite symmetric matrix.
- $k = 3, 5, 7$.

Experiment

Four algorithms are compared.

- MROM: modified Riemannian optimization method
- SULS: Projected line-search method (Schneider and Uschmajew [SU14])
- DMM: Double minimization method (Brace and Manton [BM06])
- APM: Alternating projections method (Lu et. al [LPW97] or Wentzell et. al [WAK97])

Experiment

k	method	rank	f	Rel Err($\frac{\|A-X\|_W}{\|A\|_W}$)	t
k = 3	MROM	3	8.846_{+01}	3.513_{-01}	4.823_{-01}
	SULS	3	8.846_{+01}	3.513_{-01}	2.190_{+00}
	DMM	3	8.846_{+01}	3.513_{-01}	4.706_{+00}
	APM	3	8.846_{+01}	3.513_{-01}	5.000_{+00}
k = 5	MROM	5	2.191_{-19}	1.557_{-11}	6.890_{-01}
	SULS	5	2.147_{-12}	4.874_{-08}	1.045_{+00}
	DMM	5	1.606_{-15}	1.324_{-09}	4.351_{+00}
	APM	5	7.611_{-09}	2.895_{-06}	3.585_{+00}
k = 7	MROM	5	1.799_{-21}	1.346_{-12}	4.730_{-01}
	SULS	7(0/100)	1.401_{-12}	3.780_{-08}	2.316_{+00}
	DMM	5	1.915_{-18}	4.407_{-11}	2.182_{+00}
	APM	7(0/100)	2.349_{-10}	4.865_{-07}	7.002_{+00}

Table : The number in the parenthesis indicates the fraction of experiments where the numerical rank (number of singular values greater than 10^{-8}) found by the algorithm equals the true rank. The subscript $\pm k$ indicates a scale of $10^{\pm k}$.

Conclusion

$$\min_{X \in \mathcal{M}_{\leq k}} f(X)$$

- 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;

Thank you!

Thank you for your attention!

Reference

I. Brace and J. H. Manton.
An improved BFGS-on-manifold algorithm for computing low-rank weighted approximations.
In *Proceedings of 17th International Symposium on Mathematical Theory of Networks and Systems*, pages 1735–1738, 2006.

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 SystemsI, 44:650–655, 1997.

R. Schneider and A. Uschmajew.
Convergence results for projected line-search methods on varieties of low-rank matrices via {\L}ojasiewicz inequality.
ArXiv e-prints, pages –1–1, feb 2014.

Peter D. Wentzell, Darren T. Andrews, and Bruce R. Kowalski.
Maximum likelihood multivariate calibration.
Analytical Chemistry, 69(13):2299–2311, 1997.