An Augmented Lagrangian Method for Optimization Problems with Nonnegative Orthogonality Constraints

Yuchu Yang
Department of Mathematical Informatics
Supervisor: Lecturer Kazuhiro Sato

***Abstract*—In this thesis, we proposed an algorithm based on augmented Lagrangian method combined with proximal alternating minimization approach for optimizing problems with nonnegative and orthogonality constraints. We proved that the algorithm generates at least one convergent subsequence. It is also proved that the subsolver algorithm finds a critical point with a given arbitrary tolerance. Finally, we conduct the numerical experiment and show the performance with a comparison of the existing Riemannian Optimization method.**

***Index Terms*—nonnegative orthogonality constraints, augmented Lagrangian method, nonlinear optimization**

I. Introduction

We consider nonconvex optimization problems with orthogonality and nonnegative constraints:

$$\min_{X\in\mathbb{R}^{n\times m}} \operatorname{Tr}(X^THX)+\operatorname{Tr}(K^TX) \text{ s.t. } X^TX=I_m,\ X_{i,j}\geq 0,$$

where $1\leq m\leq n$, I_m is the m-by-m identity matrix, $H\in\mathbb{R}^{n\times n}$ is a symmetric matrix and $K\in\mathbb{R}^{n\times m}$ is a rectangular matrix.

A wide variety of applications of this problem arises in various field, e.g. sparse principal componenet analysis (SPCA) [1], orthogonal nonnegative matrix factorization (ONMF) [2], etc. This thesis proposed a new augmented Lagrangian based algorithm with proximal alternating minimization (PAM) [3] as its subsolver.

II. Algorithm

Algorithm 1, performing as the outer iteration, minimizes the augmented Lagrangian by utilizing the augmented Lagrangian framework [4]–[6], which updates $(X,U,V),\Lambda_1,\Lambda_2$ and ρ alternately. While Algorithm 2 is the inner iteration which specifically computes the (X^k,U^k,V^k) for each k-th outer iteration based on the proximal alternating minimization method [3].

A. Augmented Lagrangian Scheme (Outer Iteration)

By splitting the constraints into multiple variables, we express the augmented Lagrangian of the original problem as:

$$\begin{aligned}&L(X,U,V,\Lambda_1,\Lambda_2;\rho)\\&:=\operatorname{Tr}\left(X^\top HX\right)+\operatorname{Tr}\left(K^\top X\right)+\langle\Lambda_1,U-X\rangle+\frac{\rho}{2}\|U-X\|_F^2\\&+\langle\Lambda_2,V-X\rangle+\frac{\rho}{2}\|V-X\|_F^2+\delta_{\mathbb{S}^{n\times m}}(U)+\delta_{\mathbb{R}_+^{n\times m}}(V)\end{aligned}$$

where

$$\mathbb{S}^{n\times m}:=\left\{X\in\mathbb{R}^{n\times m}:X^\top X=I_m\right\},$$
$$\mathbb{R}_+^{n\times m}:=\left\{X\in\mathbb{R}^{n\times m}:X_{i,j}\geq 0\right\},$$

and $\delta_S(X)$ is the indicator function defined by

$$\delta_S(X)=\begin{cases}0 & \text{if } X\in S,\\ +\infty & \text{otherwise.}\end{cases}$$

Algorithm 1 is our proposed method.

Algorithm 1 Augmented Lagrangian Scheme

Input: Set $k\leftarrow 1$ and (X^1,U^1,V^1) to be an arbitrary initial point.

Output: Sequence (X^k,U^k,V^k).

1: **Compute** (X^k,U^k,V^k) such that $\Phi^k\in\partial L(X^k,U^k,V^k,\bar{\Lambda_1}^k,\bar{\Lambda_2}^k,\rho)$, satisfying

$$\left\|\Phi^k\right\|_\infty\leq\epsilon^k,\quad \left(U^k\right)^\top U^k=I_m,\quad V^k\geq 0. \quad (1)$$

2: **Estimate multiplier**

$$(\Lambda_1^{k+1},\Lambda_2^{k+1})=([\bar{\Lambda_1^k}]+\rho_k[U_k-X_k],[\bar{\Lambda_2^k}]+\rho_k[V_k-X_k])$$

where $\bar{\Lambda}_p^{k+1}$ is the projection of Λ_p^{k+1} on $\left\{\Lambda_p:\bar{\Lambda}_{p,\min}\leq\Lambda_p\leq\bar{\Lambda}_{p,\max}\right\}$ $p=1,2$.

3: **Update** ρ

$$\rho^{k+1}:=\begin{cases}\rho^k & \text{, if } \left\|(U^k-X^k)\right\|_\infty\leq\tau\left\|(U^{k-1}-X^{k-1})\right\|_\infty\\ & \textit{and} \left\|(V^k-X^k)\right\|_\infty\leq\tau\left\|(V^{k-1}-X^{k-1})\right\|_\infty\\ \gamma\rho^k & \text{, otherwise}\end{cases} \quad (2)$$

4: Set $k\leftarrow k+1$, go to Step 1.

B. Proximal Alternating Minimization Scheme (Inner Iteration)

During each k-th outer iteration, Step 2 of Algorithm 1 can be solved by applying Algorithm 2 based on the PAM algorithm [3].

III. Convergence Analysis

A. Convergence Analysis of Inner Iteration

In this thesis, we proved that Algorithm 2 has global convergence under certain assumption of initial parameters, given in Proposition 1.

Algorithm 2 Proximal Alternating Minimization Scheme

Input:

Set $j \leftarrow 1$. For $k = 1$, set $(X^{1,0}, U^{1,0}, V^{1,0})$ to be an arbitrary initial point. For $k \geq 2$, set $(X^{k,0}, U^{k,0}, V^{k,0}) \leftarrow (X^{k-1}, U^{k-1}, V^{k-1})$

Output:

(X^k, U^k, V^k) that satisfies the constraint in Step 2 of Algorithm 1.

while $||\Phi^{k,j}||_\infty > \epsilon^k$ **do**

Compute

$$X^{k,j} = A^{-1}(\bar{\Lambda}_1^k + \bar{\Lambda}_2^k + \rho^k U^{k,j-1} + \rho^k V^{k,j-1} + c_1^{k,j-1} X^{k,j-1} - K)$$

where $A := A^{k,j-1} = 2H + \left(2\rho^k + c_1^{k,j-1}\right) I_n$

Compute $V_{p,q}^{k,j} = \begin{cases} \tilde{V}_{p,q}^{k,j} & ,\text{where } \tilde{V}_{p,q}^{k,j} \geq 0 \\ 0 & ,otherwise. \end{cases}$, where

$$\tilde{V}^{k,j} = \frac{-\bar{\Lambda}_2^k + \rho^k X^{k,j} + c_2^{k,j-1} V^{k,j-1}}{\rho^k + c_2^{k,j-1}}$$

Compute $U^{k,j} = P I_{n \times m} Q^\top$, where the matrices P, Q are obtained from the SVD of

$$\rho^k X^{k,j} + c_3^{k,j-1} U^{k,j-1} - \bar{\Lambda}_1^k =: P \Sigma Q^\top.$$

Set $(X^k, U^k, V^k) \leftarrow (X^{k,j}, U^{k,j}, V^{k,j})$, $\Phi^k \leftarrow \Phi^{k,j}$ and $j \leftarrow j+1$.

end while

Proposition 1. *Assume the parameters γ and ρ^1 are initialized such that*

$$\gamma > 1, \quad \rho^1 > 0, \quad \rho^1 I_n + 2H \succ 0,$$

then $(X^{k,j}, U^{k,j}, V^{k,j})$ generated by Algorithm 2 converges to a critical point $(\bar{X}^k, \bar{U}^k, \bar{V}^k)$ and the sequence $(X^{k,j}, U^{k,j}, V^{k,j})$ has a finite length. Moreover,

$$||\Phi^{k,j}||_\infty \to 0 \quad as\ j \to \infty.$$

B. Convergence Analysis of Outer Iteration

In this thesis, we proved that the sequence (X^k, U^k, V^k) generated by Algorithm 1 is bounded. That is, there exists a subsequence converging to a limit point.

IV. Numerical Experiment

We compare the performance of the proposed method and the Riemannian augmented Lagrangian method (RALM) [7] on the orthogonal nonnegative matrix approximation problem on matrices of 3 sizes. The results are shown as follows.

TABLE I
Experiment 1. RALM vs. the proposed method on 100×10 random matrix

	RALM	the Proposed Method
Objective Function Residual	314.7610	285.4391
Nonnegativity ($\geq -1e^{-5}$)	968	1000
Orthogonality	5.7334e-16	2.5832e-4

TABLE II
Experiment 2. RALM vs. the proposed method on 1000×10 random matrix

	RALM	the Proposed Method
Objective Function Residual	3256	3161
Nonnegativity ($\geq -1e^{-5}$)	8290	10000
Orthogonality	1.1002e-15	3.8132e-4

TABLE III
Experiment 3. RALM vs. the proposed method on 1000×100 random matrix

	RALM	the Proposed Method
Objective Function Residual	33092	32788
Nonnegativity ($\geq -1e^{-5}$)	98088	100000
Orthogonality	3.3228e-15	3.8657e-4

Fig. 1. shows that the proposed method has better performance on residual.

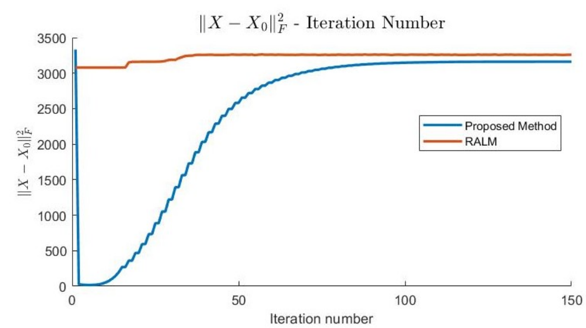

Fig. 1. Residual on the first 150 iterations for problem $\min_{X \in \mathbb{R}^{n \times m}} \|X - X_0\|_F^2$, s.t. $X^\top X = I_n$ and $X \geq 0$, where $(n, m) = (1000, 10)$.

References

[1] L. Berk and D. Bertsimas, "Certifiably optimal sparse principal component analysis," *Mathematical Programming Computation*, vol. 11, no. 3, pp. 381–420, 2019.

[2] C. Ding, T. Li, W. Peng, and H. Park, "Orthogonal nonnegative matrix tri-factorizations for clustering," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2006, pp. 126–135.

[3] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, "Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality," *Mathematics of Operations Research*, vol. 35, no. 2, pp. 438–457, 2010. [Online]. Available: https://doi.org/10.1287/moor.1100.0449

[4] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt, "On augmented Lagrangian methods with general lower-level constraints," *SIAM Journal on Optimization*, vol. 18, no. 4, pp. 1286–1309, 2008. [Online]. Available: https://doi.org/10.1137/060654797

[5] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. New York: Springer, 2006.

[6] A. R. Conn, N. I. M. Gould, and P. L. Toint, *Trust Region Methods*. Philadelphia: Society for Industrial and Applied Mathematics, 2000.

[7] C. Liu and N. Boumal, "Simple algorithms for optimization on Riemannian manifolds with constraints," *Applied Mathematics & Optimization*, vol. 82, pp. 949–981, 2020.