

Positive-Definite Sparse Precision Matrix Estimation

Lin Xia¹, Xudong Huang¹, Guanpeng Wang¹, Tao Wu²

¹School of Mathematics and Computer Science, Anhui Normal University, Wuhu, China

²School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan, China

Email: 1491357861@qq.com

How to cite this paper: Xia, L., Huang, X.D., Wang, G.P. and Wu, T. (2017) Positive-Definite Sparse Precision Matrix Estimation. *Advances in Pure Mathematics*, 7, 21-30.

<http://dx.doi.org/10.4236/apm.2017.71002>

Received: December 29, 2016

Accepted: January 19, 2017

Published: January 23, 2017

Copyright © 2017 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The positive-definiteness and sparsity are the most important property of high-dimensional precision matrices. To better achieve those property, this paper uses a sparse lasso penalized D-trace loss under the positive-definiteness constraint to estimate high-dimensional precision matrices. This paper derives an efficient accelerated gradient method to solve the challenging optimization problem and establish its converges rate as $O\left(\frac{1}{k^2}\right)$. The numerical simulations illustrated our method have competitive advantage than other methods.

Keywords

Positive-Definiteness, Sparsity, D-Trace Loss, Accelerated Gradient Method

1. Introduction

In the past twenty years, the most popular direction of statistics is high-dimensional data. In functional magnetic resonance imaging (fMRI), bioinformatics, Web mining, climate research, risk management and social science, it not only has a wide range of applications, but also the main direction of scientific research at present. In theoretical and practical, high-dimensional precision matrix estimation always plays a very important role and has wide applications in many fields.

Thus, estimation of high-dimensional precision matrix is increasingly becoming a crucial question in many field. However, estimation of high-dimensional precision matrix has two difficulty: 1) sparsity of estimator; (ii) the positive-definiteness constraint. Huang *et al.* [1] considered using Cholesky decomposition to estimate the precision matrix. Although the regularized

Cholesky decomposition approach can achieve a positive-semidefiniteness, it can not guarantee sparsity of estimator. Meinshausen *et al.* [2] use a neighbourhood selection scheme in which one can sequentially estimate the support of each row of precision matrix by fitting lasso penalized least squares regression model. Peng *et al.* [3] considered a joint neighbourhood estimator by using the lasso penalization. Yuan [4] considered the Dantzig selector to replace the lasso penalized least squares in the neighbourhood selection scheme. Cai *et al.* [5] considered a constrained ℓ_1 minimization estimator for estimating sparse precision matrices. However, this methods mentioned are not always achieve a positive-definiteness.

To overcome the difficulty (ii), one possible method is using the eigen-decomposition of $\hat{\Theta}$ and designing $\hat{\Theta}$ to satisfy condition $\{\Theta \geq 0\}$. Assume that $\hat{\Theta}$ has the eigen-decomposition $\hat{\Theta} = \sum_{i=1}^p \lambda_i v_i^T v_i$ and then a positive semi-definite estimator was gained by setting $\hat{\Theta} = \sum_{i=1}^p \max(\lambda_i, 0) v_i^T v_i$. However, this strategy destroys the sparsity pattern of $\hat{\Theta}$ for sparse precision matrix estimation. Yuan *et al.* [6] considered the lasso penalized likelihood criterion and used the maxd *et al.* gorithm to compute the estimator. Friedman *et al.* [7] considered the graphical lasso algorithm for solving the lasso penalized Gaussian likelihood estimator. Witten *et al.* [8] optimized the graphical lasso. By the lasso or ℓ_1 penalized Gaussian likelihood estimator, thoses methods simultaneously achieve positive-definiteness and sparsity.

Recently, Zhang *et al.* [9] consider a constrained convex optimization framework for high-dimensional precision matrix. They used lasso penalized D-trace loss replace traditional lasso function, and enforced the positive-definite constraint $\{\Theta \geq \varepsilon \mathbf{I}\}$ for some arbitrarily small $\varepsilon > 0$. In their work, focusing on solving problem as follow:

$$\hat{\Theta}^+ = \arg \min_{\Theta \geq \varepsilon \mathbf{I}} \frac{1}{2} \left\langle \Theta^2, \hat{\Sigma}_n \right\rangle - \text{tr}(\Theta) + \lambda \|\Theta\|_{1,\text{off}} \quad (1)$$

It is important to note that ε is not a tuning parameter like λ . We simply include ε in the procedure to ensure that the smallest eigenvalue of the estimator is at least ε . They developed an efficient alternating direction method of multipliers (ADMM) to solve the challenging optimization problem (1) and establish its convergence properties.

To gain a better estimator for high-dimensional precision matrix and achieve the more optimal convergence rate, this paper mainly propose an effective algorithm, an accelerated gradient method ([10]), with fast global convergence rates to solve problem (1). This method mainly basis on the Nesterov's method for accelerating the gradient method ([11] [12]), showing that by exploiting the special structure of the trace norm, the classical gradient method for smooth problems can be adapted to solve the trace regularized nonsmooth problems. However, for our problem (1), we have not trace norm, instead of is ℓ_1 norm form, but this method have the similar efficiently result for our problem. Numerical results show that this method for our problem (1) not only has

significant computational advantages, but also achieves the optimal convergence rate as $O\left(\frac{1}{k^2}\right)$.

The paper is organized as follows: Section 2 introduces our methodology, including model establishing in Section 2.1; step size estimation in Section 2.2; an accelerate gradient method algorithm in Section 2.3; the convergence analysis results of this algorithm in Section 2.4. Section 3 introduced numerical results for our method in comparing with other methods. And discussion are made in Section 4. All proofs are given in the **Appendix**.

2. An Accelerated Gradient Method

2.1. Model Establishing

According to introduction, our optimization problem D-trace Loss function as follow:

$$\min_{\Theta \geq \varepsilon I} F(\Theta) := \frac{1}{2} \langle \Theta^2, \hat{\Sigma}_n \rangle - \text{tr}(\Theta) + \lambda \|\Theta\|_{1,\text{off}} \tag{2}$$

where λ is a nonnegative penalization parameter, $\hat{\Sigma}_n$ is the sample covariance matrix. $\hat{\Sigma}_n = \frac{1}{n} \sum_{i=1}^n X_i X_i^T$, and $|\Theta|_1 = \|\Theta\|_{1,\text{off}} = \sum_{i \neq j} \|\Theta_{ij}\|$ is the ℓ_1 off-diagonal penalty. Defining $f(\Theta) = \frac{1}{2} \langle \Theta^2, \hat{\Sigma}_n \rangle - \text{tr}(\Theta)$, and $f(\Theta)$ is a continuously differentiable function. Considering the gradient step

$$\Theta_k = \Theta_{k-1} - \frac{1}{t_k} \nabla f(\Theta_{k-1}) \tag{3}$$

where $t_k \geq 0$ is a stepsize, $\nabla f(\Theta) = \frac{1}{2} (\Theta \hat{\Sigma}_n + \hat{\Sigma}_n \Theta) - I$. The smooth part (3) can be reformulated equivalent as a proximal regularization of the linearized function $f(\Theta)$ at Θ_{k-1} :

$$\Theta_k = \arg \min_{\Theta \geq \varepsilon I} \Phi_{t_k}(\Theta, \Theta_{k-1}) \tag{4}$$

where

$$\Phi_{t_k}(\Theta, \Theta_{k-1}) = f(\Theta_{k-1}) + \langle \Theta - \Theta_{k-1}, \nabla f(\Theta_{k-1}) \rangle + \frac{t_k}{2} \|\Theta - \Theta_{k-1}\|_F^2 \tag{5}$$

Based on this equivalence relationship, solving the optimization problem (2) by the following iterative step:

$$\begin{aligned} \Theta_k &= \arg \min_{\Theta \geq \varepsilon I} \Psi_{t_k}(\Theta, \Theta_{k-1}) \triangleq \arg \min_{\Theta \geq \varepsilon I} \Phi_{t_k}(\Theta, \Theta_{k-1}) + \lambda |\Theta|_1 \\ &= \arg \min_{\Theta \geq \varepsilon I} \frac{1}{2} \langle \Theta_{k-1}^2, \hat{\Sigma}_n \rangle - \text{tr}(\Theta_{k-1}) + \left\langle \Theta - \Theta_{k-1}, \frac{1}{2} (\Theta_{k-1} \hat{\Sigma}_n + \hat{\Sigma}_n \Theta_{k-1}) \right. \\ &\quad \left. - I \right\rangle + \frac{t_k}{2} \|\Theta - \Theta_{k-1}\|_F^2 + \lambda |\Theta|_1 \tag{6} \\ &= \arg \min_{\Theta \geq \varepsilon I} \frac{t_k}{2} \left\| \Theta - \left[\Theta_{k-1} - \frac{1}{t_k} \left(\frac{1}{2} (\Theta_{k-1} \hat{\Sigma}_n + \hat{\Sigma}_n \Theta_{k-1}) - I \right) \right] \right\|_F^2 + \lambda |\Theta|_1 \end{aligned}$$

with equality in the last line by ignoring terms that do not depend on Θ .

Defining $(C)_+$ as the projection of a matrix C onto the convex cone $\{C \geq \varepsilon I\}$. Assuming that C has the eigen-decomposition $\sum_{j=1}^p \lambda_j v_j^T v_j$, and then $(C)_+$ can be obtained as $\sum_{j=1}^p \max(\lambda_j, \varepsilon) v_j^T v_j$. Defining an entry-wise soft-thresholding rule for all the off-diagonal elements of a matrix

$$S(Z, \tau) = \{s(z_{j\ell}, \tau)\}_{1 \leq j, \ell \leq p} \quad \text{with}$$

$s(z_{j\ell}, \tau) = \text{sign}(z_{j\ell}) \max(|z_{j\ell}| - \tau, 0) I_{\{j \neq \ell\}} + z_{j\ell} I_{\{j = \ell\}}$. Thus, the above problem can be summarized in the following theorem:

Theorem 1: Let $B \in \mathbb{R}^{n \times n}$, and Θ is symmetric covariance matrix, then:

$$S(B) = \arg \min_{\Theta \geq \varepsilon I} \left\{ \frac{1}{2} \|\Theta - B\|_F^2 + \lambda |\Theta|_1 \right\} \quad (7)$$

is given by $S(B) = (S(B, \lambda))_+$, where $S(B, \lambda) = \{s(B_{j\ell}, \lambda)\}_{1 \leq j, \ell \leq p}$ with $s(B_{j\ell}, \lambda) = \text{sign}(B_{j\ell}) \max(|B_{j\ell}| - \lambda, 0) I_{\{j \neq \ell\}} + B_{j\ell} I_{\{j = \ell\}}$.

The proof of this theorem is easy by applying the soft-thresholding method.

2.2. Step Size Estimation

To guarantee the convergence rate of the resulting iterative sequence, Firstly giving the relationship between our proximal function Ψ_{t_k} and the objection function F at the certain point.

Lemma 1: Let

$$T_\mu(\tilde{\Theta}) = \arg \min_{\Theta \geq \varepsilon I} \Psi_\mu(\Theta, \tilde{\Theta}) \quad (8)$$

where Ψ is defined in Equation (6). Assuming the following inequality holds:

$$F(T_\mu(\tilde{\Theta})) \leq \Psi_\mu(T_\mu(\tilde{\Theta}), \tilde{\Theta}) \quad (9)$$

then for any $\Theta \in \mathbb{R}^{n \times n}$, then:

$$F(\Theta) - F(T_\mu(\tilde{\Theta})) \geq \frac{\mu}{2} \|T_\mu(\tilde{\Theta}) - \tilde{\Theta}\|_F^2 + \mu \langle \tilde{\Theta} - \Theta, T_\mu(\tilde{\Theta}) - \tilde{\Theta} \rangle \quad (10)$$

This lemma is proved in the **Appendix**.

At each iterative step of the algorithm, an appropriate step size μ is needed to satisfy $\Theta_k = T_\mu(\Theta_{k-1})$ and

$$F(\Theta_k) \leq \Psi_\mu(\Theta_k, \Theta_{k-1}) \quad (11)$$

Since the gradient of $f(\cdot)$ satisfies Lipschitz continuous, according to Nesterov et al. [11] work, having follow lemma.

Lemma 2: Supposing that $f(X)$ is a convex function, and the gradient of $f(X)$ denote $\nabla f(X)$ is Lipschitz continuous with constant L , then:

$$f(X) \leq f(Y) + \langle X - Y, \nabla f(Y) \rangle + \frac{L}{2} \|X - Y\|_F^2 \quad \forall X, Y \in \mathbb{R}^n \quad (12)$$

so, we have

$$f(T_L(\tilde{\Theta})) \leq f(\tilde{\Theta}) + \langle T_L(\tilde{\Theta}) - \tilde{\Theta}, \nabla f(\tilde{\Theta}) \rangle + \frac{L}{2} \|T_L(\tilde{\Theta}) - \tilde{\Theta}\|_F^2 \quad (13)$$

Hence, when $\mu \geq L$, then:

$$F(T_\mu(\tilde{\Theta})) \leq \Phi_\mu(T_\mu(\tilde{\Theta}), \tilde{\Theta}) + \lambda \|T_\mu(\tilde{\Theta})\|_1 = \Psi_\mu(T_\mu(\tilde{\Theta}), \tilde{\Theta}) \quad (14)$$

The above results show that the condition in Equation (11) is always satisfied when the update rule

$$\Theta_k = T_L(\Theta_{k-1}) \quad (15)$$

2.3. An Accelerate Gradient Method Algorithm

In practice, L may be unknown or it is expensive to compute. To use the following step size estimation method, usually, giving an initial estimate of L as L_0 and increasing this estimate with a multiplicative factor $\gamma > 1$ repeatedly until the condition in Equation (11) is satisfied. It is well known ([11] [12]) that if the objection function is smooth, then the accelerate gradient method can achieve the optimal convergence rate of $O\left(\frac{1}{k^2}\right)$. Recently, there have other similar methods applying in problems consisted a smooth part and a non-smooth part ([10] [13] [14] [15]). Then giving the accelerate gradient algorithm to solve the optimization problem in Equation (2).

Algorithm1: An accelerate gradient method algorithm for high-dimensional precision matrix

- 1) Initialize: $L_0, \gamma, \tilde{\Theta}_1 = \Theta_0 \in \mathbb{R}^{n \times n}, \alpha_1 = 1$
- 2) Iterate:
- 3) set $\bar{L} = L_{k-1}$
- 4) While $F(T_{\bar{L}}(\tilde{\Theta}_{k-1})) > \Psi_{\bar{L}}(T_{\bar{L}}(\tilde{\Theta}_{k-1}), \tilde{\Theta}_{k-1})$, set $\bar{L} := \gamma \bar{L}$
- 5) Set $L_k = \bar{L}$ and update

$$\begin{aligned} \Theta_k &= T_{L_k}(\tilde{\Theta}_k) \\ \alpha_{k+1} &= \frac{1 + \sqrt{1 + 4\alpha_k^2}}{2} \\ \tilde{\Theta}_{k+1} &= \Theta_k + \left(\frac{\alpha_k - 1}{\alpha_{k+1}} \right) (\Theta_k - \Theta_{k-1}) \end{aligned}$$

2.4. Convergence Analysis

In our method, two sequences Θ_k and $\tilde{\Theta}_k$ are updated recursively. In particular, Θ_k is the approximate solution at the k th step and $\tilde{\Theta}_k$ is called the search point ([11] [12]), which is constructed as a linear combination of the latest two approximate solutions Θ_{k-1} and Θ_{k-2} . In this section, the convergence rate of the method can be showed as $O\left(\frac{1}{k^2}\right)$. This result is summarized in the following theorem.

Theorem 2: Let $\{\Theta_k\}$ and $\{\tilde{\Theta}_k\}$ be the covariance matrices sequence generated by our algorithm. Then for any $k \geq 1$, having

$$F(\Theta_k) - F(\Theta^*) \leq \frac{2\gamma L \|\Theta_0 - \Theta^*\|_F^2}{(k+1)^2} \quad (16)$$

where $\Theta^* = \arg \min_{\Theta \succeq \epsilon I} F(\Theta)$.

3. Simulation

In this section, providing numerical results for our algorithm which will show our algorithmic advantages by three model. In the simulation study, data were generated from $N(0, \Sigma^0)$, where $\Theta^0 = (\Sigma^0)^{-1}$. And the sample size was taken to be $n = 400$ in all models, and let $p = 500$ in Models 1 and 2, and $p = 484$ in Model 3, which is similar to Zhang *et al.* [9]. The numerical results of three models as follow:

Model 1: $\Theta_{i,i}^0 = 1, \Theta_{i,j}^0 = 0.2$ for $1 \leq |i - j| \leq 2$ and $\Theta_{i,j}^0 = 0$ otherwise.

Model 2: $\Theta_{i,i}^0 = 1, \Theta_{i,j}^0 = 0.2$ for $1 \leq |i - j| \leq 4$ and $\Theta_{i,j}^0 = 0$ otherwise.

Model 3: $\Theta_{i,i}^0 = 1, \Theta_{i,i+1}^0 = 0.2$ for $\text{mod}(i, p^{1/2}) \neq 0$, $\Theta_{i,i+p^{1/2}}^0 = 0.2$ and $\Theta_{i,j}^0 = 0$ otherwise; this is the grid model in Ravikumar *et al.* [16] and requires $p^{1/2}$ to be an integer.

Simulation results based on 100 independent replications are showed in **Table 1**. This paper mainly compare the three methods in terms of four quantities: the operator risk $E(\|\hat{\Theta} - \Theta^0\|_2)$, the matrix $\ell_{1,\infty}$ risk $E(\|\hat{\Theta} - \Theta^0\|_{1,\infty})$, and the percentages of correctly estimated nonzeros and zeros (TP and TN), where $\ell_{1,\infty}$ norm $\max_i(\sum_j |X_{i,j}|)$ is written as $\|X\|_{1,\infty}$. In the first two columns smaller numbers are better; in the last two columns larger numbers are better. In general, **Table 1** shows that our estimator performs better than Zhang *et al.*'s method estimator and the lasso penalized Gaussian likelihood estimator.

4. Conclusion

This paper mainly estimate positive-definite sparse precision matrix estimation

Table 1. Comparison of our method with Zhang *et al.*'s method and graphical lasso.

	Operator	$\ell_{1,\infty}$	TP	TN
Model 1				
our method	0.77	0.80	94.91	99.20
Zhang <i>et al.</i> 's method	0.77	1.06	88.80	98.77
Graphical lasso	0.78	1.26	88.12	97.65
Model 2				
our method	1.59	1.60	70.02	98.40
Zhang <i>et al.</i> 's method	1.59	1.92	63.47	98.66
Graphical lasso	1.61	2.11	64.88	97.40
Model 3				
our method	0.56	0.81	1	99.21
Zhang <i>et al.</i> 's method	0.56	0.91	99.41	98.57
Graphical lasso	0.58	1.06	99.76	97.48

via lasso penalized D-trace loss by an efficient accelerated gradient method. The positive-definiteness and sparsity are the most important property of large covariance matrices, our method not only efficiently achieves these property, but also shows an better convergence rate. Numerical results have show that our estimator also have a better performance, comparing to Zhang *et al.*'s method and the Graphical lasso method.

Funding

This project was supported by National Natural Science Foundation of China (71601003) and the National Statistical Scientific Research Projects (2015LZ54).

References

- [1] Huang, J., Liu, N., Pourahmadi, M. and Liu, L. (2006) Covariance Matrix Selection and Estimation via Penalised Normal Likelihood. *Biometrika*, **93**, 85-98. <https://doi.org/10.1093/biomet/93.1.85>
- [2] Meinshausen, N. and Bühlmann, P. (2006) High-Dimensional Graphs and Variable Selection with the Lasso. *Annals of Statist*, **34**, 1436-1462. <https://doi.org/10.1214/009053606000000281>
- [3] Peng, J., Wang, P., Zhou, N. and Zhu, J. (2009) Partial Correlation Estimation by Joint Sparse Regression Models. *Journal of the American Statistical Association*, **104**, 735-746. <https://doi.org/10.1198/jasa.2009.0126>
- [4] Yuan, M. (2010) High Dimensional Inverse Covariance Matrix Estimation via Linear Programming. *Journal of Machine Learning Research*, **11**, 2261-2286.
- [5] Cai, T., Liu, W. and Luo, X. (2011) A Constrained ℓ_1 Minimization Approach to Sparse Precision Matrix Estimation. *Journal of the American Statistical Association*, **106**, 594-607. <https://doi.org/10.1198/jasa.2011.tm10155>
- [6] Yuan, M. and Lin, Y. (2007) Model Selection and Estimation in the Gaussian Graphical Model. *Biometrika*, **94**, 19-35. <https://doi.org/10.1093/biomet/asm018>
- [7] Friedman, J.H., Hastie, T.J. and Tibshirani, R.J. (2008) Sparse Inverse Covariance Estimation with the Graphical Lasso. *Biostatistics*, **9**, 432-441. <https://doi.org/10.1093/biostatistics/kxm045>
- [8] Witten, D., Fienman, J.H. and Simon, N. (2011) New Insights and Faster Computations for the Graphical Lasso. *Journal of Computational and Graphical Statistics*, **20**, 892-900. <https://doi.org/10.1198/jcgs.2011.11051a>
- [9] Zhang, T. and Zou, H. (2014) Sparse Precision Matrix Estimation via Lasso Penalized D-Trace Loss. *Biometrika*, **101**, 103-120. <https://doi.org/10.1093/biomet/ast059>
- [10] Ji, S. and Ye, J. (2009) An Accelerated Gradient Method for Trace Norm Minimization. *International Conference on Machine Learning*, **58**, 457-464. <https://doi.org/10.1145/1553374.1553434>
- [11] Nesterov, Y. (1983) A Method for Solving a Convex Programming Problem with Convergence Rate $O\left(\frac{1}{k^2}\right)$. *Soviet Mathematics Doklady*, **27**, 372-367.
- [12] Nesterov, Y. (2003) Introductory Lectures on Convex Optimization: A Basic Course. *Kluwer Academic*.
- [13] Toh, K.C. and Yun, S. (2010) An Accelerated Proximal Gradient Algorithm for

Nuclear Norm Regularized Linear Least Squares Problems. *Pacific Journal of Optimization*, **6**, 615-640.

- [14] Tseng, P. (2008) On Accelerated Proximal Gradient Methods for Convex-Concave Optimization. *Siam Journal on Optimization*.
- [15] Beck, A. and Teboulle, M. (2009) A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *Siam Journal on Imaging Sciences*, **2**, 183-202. <https://doi.org/10.1137/080716542>
- [16] Ravikumar, P., Wainwright, M., Rasutti, G. and Yu, B. (2011) High-Dimensional Covariance Estimation by Minimizing ℓ_1 -Penalized Log-Determinant Divergence. *Electronic Journal of Statistics*, **5**, 935-980. <https://doi.org/10.1214/11-EJS631>

Appendix: Proff of Theorems and Lemmas

Appendix: Proof of Lemma 1

Since that both the trace function and ℓ_1 norm are all convex function, so

$$\begin{aligned} \frac{1}{2} \langle \Theta^2, \hat{\Sigma}_n \rangle - \text{tr}(\Theta) &= \frac{1}{2} \text{tr}(\Theta^2 \hat{\Sigma}_n) - \text{tr}(\Theta) \\ &\geq \frac{1}{2} \text{tr}(\tilde{\Theta}^2 \hat{\Sigma}_n) - \text{tr}(\tilde{\Theta}) + \left\langle \Theta - \tilde{\Theta}, \frac{1}{2} (\tilde{\Theta} \hat{\Sigma}_n + \hat{\Sigma}_n \tilde{\Theta}) - I \right\rangle \end{aligned} \quad (17)$$

$$\lambda |\Theta|_1 \geq \lambda \|T_L(\tilde{\Theta})\|_1 + \lambda \langle \Theta - T_L(\tilde{\Theta}), g(T_L(\tilde{\Theta})) \rangle \quad (18)$$

where $g(T_L(\tilde{\Theta})) \in \partial \|T_L(\tilde{\Theta})\|_1$, is the sub-gradient of ℓ_1 norm at point $T_L(\tilde{\Theta})$. Since $F(T_\mu(\tilde{\Theta})) \leq \Psi_\mu(T_\mu(\tilde{\Theta}), \tilde{\Theta})$ and combing in Equations (17), (18) then

$$\begin{aligned} F(\Theta) - F(T_L(\tilde{\Theta})) &\geq F(\Theta) - \Psi_L(T_L(\tilde{\Theta}), \tilde{\Theta}) \\ &\geq \left\langle \Theta - \tilde{\Theta}, \frac{1}{2} (\tilde{\Theta} \hat{\Sigma}_n + \hat{\Sigma}_n \tilde{\Theta}) - I \right\rangle + \lambda \langle \Theta - T_L(\tilde{\Theta}), g(T_L(\tilde{\Theta})) \rangle \\ &\quad - \left\langle T_L(\tilde{\Theta}) - \tilde{\Theta}, \frac{1}{2} (\tilde{\Theta} \hat{\Sigma}_n + \hat{\Sigma}_n \tilde{\Theta}) - I \right\rangle - \frac{L}{2} \|T_L(\tilde{\Theta}) - \tilde{\Theta}\|_F^2 \\ &= \left\langle \Theta - T_L(\tilde{\Theta}), \frac{1}{2} (\tilde{\Theta} \hat{\Sigma}_n + \hat{\Sigma}_n \tilde{\Theta}) - I + \lambda g(T_L(\tilde{\Theta})) \right\rangle - \frac{L}{2} \|T_L(\tilde{\Theta}) - \tilde{\Theta}\|_F^2 \end{aligned} \quad (19)$$

Since $T_L(\tilde{\Theta})$ is a minimizer of $\Psi_L(\Theta, \tilde{\Theta})$, thus

$$\frac{1}{2} (\tilde{\Theta} \hat{\Sigma}_n + \hat{\Sigma}_n \tilde{\Theta}) - I + L(T_L(\tilde{\Theta}) - \tilde{\Theta}) + \lambda g(T_L(\tilde{\Theta})) = 0 \quad (20)$$

So the Equation (19) can be simplified as:

$$\begin{aligned} F(\Theta) - F(T_L(\tilde{\Theta})) &\geq \left\langle \Theta - T_L(\tilde{\Theta}), \frac{1}{2} (\tilde{\Theta} \hat{\Sigma}_n + \hat{\Sigma}_n \tilde{\Theta}) - I + \lambda g(T_L(\tilde{\Theta})) \right\rangle - \frac{L}{2} \|T_L(\tilde{\Theta}) - \tilde{\Theta}\|_F^2 \\ &= \left\langle \Theta - T_L(\tilde{\Theta}), -L(T_L(\tilde{\Theta}) - \tilde{\Theta}) \right\rangle - \frac{L}{2} \|T_L(\tilde{\Theta}) - \tilde{\Theta}\|_F^2 \\ &= L \langle \tilde{\Theta} - \Theta, T_L(\tilde{\Theta}) - \tilde{\Theta} \rangle + \frac{L}{2} \|T_L(\tilde{\Theta}) - \tilde{\Theta}\|_F^2 \end{aligned} \quad (21)$$

Appendix: Proof of Theorem 2

Defining $U_k = \alpha_k \Theta_k - (\alpha_k - 1) \Theta_{k-1} - \Theta^*$, $V_k = F(\Theta_k) - F(\Theta^*)$, easily obtaining

$$\frac{2}{L_{k+1}} (\alpha_k^2 V_k^2 - \alpha_{k+1}^2 V_{k+1}) \geq \|U_{k+1}\|_F^2 - \|U_k\|_F^2 \quad (22)$$

since $L_{k+1} \geq L_k$, so

$$\frac{2}{L_k} \alpha_k^2 V_k^2 - \frac{2}{L_{k+1}} \alpha_{k+1}^2 V_{k+1} \geq \|U_{k+1}\|_F^2 - \|U_k\|_F^2 \quad (23)$$

By applying Lemma 1, easily obtaining:

$$\begin{aligned} F(\Theta^*) - F(\Theta_1) &= F(\Theta^*) - F(T_{L_1}(\tilde{\Theta}_1)) \\ &\geq \frac{L_1}{2} \|\Theta_1 - \Theta^*\|_F^2 - \frac{L_1}{2} \|\tilde{\Theta}_1 - \Theta^*\|_F^2 \end{aligned} \quad (24)$$

so:

$$\frac{2V_1}{L_1} \leq \|\tilde{\Theta}_1 - \Theta^*\|_F^2 - \|\Theta_1 - \Theta^*\|_F^2 \quad (25)$$

Applying (23) (25), then:

$$V_{k+1} \leq \frac{L_{k+1}}{2\alpha_{k+1}^2} \|\tilde{\Theta}_1 - \Theta^*\|_F^2 \quad (26)$$

Combining the Equation (26) and the relation $\alpha_k^2 \geq \frac{(k+1)^2}{4}$, easily obtaining:

$$F(\Theta_k) - F(\Theta^*) \leq \frac{2L_k \|\Theta_0 - \Theta^*\|_F^2}{(k+1)^2} \leq \frac{2\gamma L \|\Theta_0 - \Theta^*\|_F^2}{(k+1)^2} \quad (27)$$



Scientific Research Publishing

Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact apm@scirp.org

