Journal of Computational Mathematics
Vol.39, No.5, 2021, 693–707.
http://www.global-sci.org/jcm
doi:10.4208/jcm.2005-m2019-0151

A GREEDY ALGORITHM FOR SPARSE PRECISION MATRIX APPROXIMATION*

Didi Lv
School of Mathematical Sciences,
Shanghai Jiao Tong University, Shanghai 200240, China
Email: Eric2014_Lv@sjtu.edu.cn
Xiaoqun Zhang[1)]
School of Mathematical Sciences, MOE-LSC and Institute of Natural Sciences,
Shanghai Jiao Tong University, Shanghai 200240, China
Email: xqzhang@sjtu.edu.cn

Abstract

Precision matrix estimation is an important problem in statistical data analysis. This paper proposes a sparse precision matrix estimation approach, based on CLIME estimator and an efficient algorithm GISS$^\rho$ that was originally proposed for l_1 sparse signal recovery in compressed sensing. The asymptotic convergence rate for sparse precision matrix estimation is analyzed with respect to the new stopping criteria of the proposed GISS$^\rho$ algorithm. Finally, numerical comparison of GISS$^\rho$ with other sparse recovery algorithms, such as ADMM and HTP in three settings of precision matrix estimation is provided and the numerical results show the advantages of the proposed algorithm.

Mathematics subject classification: 49J52, 65K05, 62H12.
Key words: Precision matrix estimation, CLIME estimator, Sparse recovery, Inverse scale space method, Greedy methods.

1. Introduction

Covariance matrix and precision matrix estimation are two important problems in statistical analysis and data science. The problems become more challenging for both theoretical analysis and practical computation in high-dimensional setting when the number of variable dimension p is relatively large compared to the sample size n. Therefore, effective estimation model and methods are necessary to achieve a stable, efficient and accurate estimation facing big data challenge.

Denote $\boldsymbol{X} = (X_1, X_2, \cdots, X_p)^T$ by a p variate random vector. The covariance matrix and precision matrix can be traditionally denoted by Σ_0 and $\Omega_0 = \Sigma_0^{-1}$ respectively. Assume an independent and identically distributed n random samples $\{\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n\}$ are from the distribution of $\boldsymbol{X}$. The unbiased sample covariance matrix is the mostly used estimator of covariance matrix, as defined in the following

$$\Sigma_n = \frac{1}{n-1}\sum_{k=1}^{n}(\boldsymbol{X}_k - \bar{\boldsymbol{X}})(\boldsymbol{X}_k - \bar{\boldsymbol{X}})^T,$$

* Received June 27, 2019 / Revised version received December 25, 2019 / Accepted May 25, 2020 / Published online July 6, 2021 /
1) Corresponding author

where $\bar{\boldsymbol{X}} = \frac{1}{n}\sum_{k=1}^{n}\boldsymbol{X}_k$ denotes the sample mean. When p is larger than n, it is obvious that Σ_n is singular and the estimation for Ω_0 naturally becomes unstable and not well defined.

Estimation of precision matrix in high-dimensional setting has been studied for a long time. For example, when the random variable $\boldsymbol{X}$ follows a certain ordering structure, methods based on banding the Cholesky factor of the inverse of sample covariance matrix were studied in [2,33]. Penalized likelihood methods such as l_1-MLE type estimators were studied in [12, 18, 36] and the convergence rate in Frobenius norm was given by [30]. In [35], the authors established the convergence rate for sub-Gaussian distribution cases. For more restrictive conditions, such as mutual incoherence or irrepresentable conditions, [29] showed the convergence rates in elementwise l_∞ norm and spectral norm. To overcome the drawbacks that l_1 penalty inevitably leads to biased estimation, nonconvex penalty such as SCAD penalty [14, 21] was proposed [15, 38], although it often requires high computational cost.

Recently, [8] proposed a new constrained l_1 minimization approach called CLIME for sparse precision estimation. Convergence rates in spectral norm, elementwise l_∞ norm and Frobenius norm were established under weaker assumptions and shown to be faster than those l_1-MLE estimators when the population distributions have polynomial-type tails. In addition, CLIME has a computational advantage that each column of precision matrix estimation can be independently computed in parallel. However, in [8], the columns are obtained by the algorithm l_1-magic [10] based on linear programming, which is still time-consuming in high dimensional setting. More efficient approaches still need to be developed for practical high dimensional applications.

In compressive sensing and sparse optimization community, many algorithms and related theoretical results are developed for l_1 minimization optimization problems [1,6,7,20,27,34,37]. Greedy inverse scale space flows (GISS) [24], originally stems from the adaptive inverse scale (aISS) method [5], is a new sparse recovery approach combining the idea of greedy approach and ℓ_1 minimization. Compared to the aISS method, GISS method enjoys a higher efficiency. GISS$^\rho$ with $\rho \geq 1$ being an acceleration factor, as a variant of GISS, can further accelerate sparse solution recovery by increasing the support of the current iterate by many indices at once.

In this article, we take the advantages of CLIME estimator framework and GISS$^\rho$ algorithm for compressive sensing and propose a new approach for sparse precision matrix estimation. More specifically, we transfer the constraint bound in CLIME estimator to a tuning parameter in the stopping criteria in GISS$^\rho$ method. Convergence results in elementwise l_∞ norm are established under weak assumptions as [8]. Numerical experiments show the competitive advantages of the proposed algorithm in terms of computation time, on obtaining the same level of sparsity and accuracy compared to other existing methods.

The rest of the paper is organized as follows. In Section 2, we firstly introduce basic notations and simply revisit the CLIME estimator. In Section 3, we present our method, which derived from CLIME estimator and GISS$^\rho$ algorithm. In Section 4, we establish the theoretical analysis with assumptions. Section 5 presents the numerical results including simulated experiments and application on real data. The paper is concluded in Section 6 and the proof of the main results can be found in Appendix.