Statistica Sinica **34** (2024), 1-22
doi:https://doi.org/10.5705/ss.202021.0273

A SIMPLE METHOD FOR ESTIMATING GAUSSIAN GRAPHICAL MODELS

Yiyi Yin, Yang Song and Hui Zou*

University of Minnesota

Abstract: The penalized likelihood estimator is the state-of-the-art method for estimating a Gaussian graphical model, because it delivers a symmetric graph and is efficient to compute, owing to the graphical lasso implementation. However, the estimator requires a stringent irrepresentability condition in order to achieve consistent recovery of the underlying graph. Another popular method, neighborhood selection, does not offer a symmetric solution by itself, and also requires a set of irrepresentability conditions for exact recovery. In this paper, we propose a new method, called the simple graph maker, for estimating an underlying graph. The simple graph maker produces a symmetric estimator by using a simple ℓ_1-penalized quadratic problem, which is easily computed by coordinate descent. Furthermore, it is shown to recover the underlying graph with overwhelming probability, without assuming additional structure conditions on the variables. The rates of convergence under various matrix norms are also established. The new method is shown to exhibit excellent performance on simulated and real data.

Key words and phrases: Coordinate descent, exact recovery, Gaussian graphical model, graphical Lasso, irrepresentable conditions, sparsity.

1. Introduction

In this study, we examine the problem of constructing a Gaussian graphical model from n independent and identically distributed observations (i.i.d.) from a multivariate Gaussian distribution. Suppose that $\mathbf{X} = (X_1, X_2, \ldots, X_p)^T$ follows a multivariate Gaussian distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}^*)$. Let $\boldsymbol{\Theta}^* = (\theta^*_{ij})$ and $\boldsymbol{\Sigma}^* = (\boldsymbol{\Theta}^*)^{-1}$ denote the precision matrix and the covariance matrix, respectively. It is known that the (i, j) element of $\boldsymbol{\Theta}^*$ is zero if and only if variables X_i and X_j are conditional independent, given all the other variables (Lauritzen (1996)). Thus, data analysts often use the sparsity pattern of an estimated sparse precision matrix to construct a Gaussian graphical model that describes the dependence relationships between variables. As a result, the problem of estimating a large sparse precision matrix has received increased attention in the past decade, for a comprehensive review, see Chapter 9 of Fan et al. (2020), and the references therein. Currently, the two most popular methods are neighborhood selection (Meinshausen and Bühlmann (2006)) and the penalized likelihood estimator (i.e.,

*Corresponding author.