The Annals of Statistics
2012, Vol. 40, No. 4, 1978–1983
DOI: 10.1214/12-AOS981
Main article DOI: 10.1214/11-AOS949

DISCUSSION: LATENT VARIABLE GRAPHICAL MODEL SELECTION VIA CONVEX OPTIMIZATION

BY MARTIN J. WAINWRIGHT

University of California at Berkeley

1. Introduction. It is my pleasure to congratulate the authors for an innovative and inspiring piece of work. Chandrasekaran, Parrilo and Willsky (hereafter CPW) have come up with a novel approach, combining ideas from convex optimization and algebraic geometry, to the long-standing problem of Gaussian graphical model selection with latent variables. Their method is intuitive and simple to implement, based on solving a convex log-determinant program with suitable choices of regularization. In addition, they establish a number of attractive theoretical guarantees that hold under high-dimensional scaling, meaning that the graph size p and sample size n are allowed to grow simultaneously.

1.1. *Background.* Recall that an undirected graphical model (also known as a Markov random field) consists of a family of probability distributions that factorize according to the structure of undirected graph $G = (V, E)$. In the multivariate Gaussian case, the factorization translates into a sparsity assumption on the inverse covariance or precision matrix [9]. In particular, given a multivariate Gaussian random vector $(X_1, \ldots, X_p)$ with covariance matrix Σ, it is said to be Markov with respect to the graph G if its precision matrix $K = \Sigma^{-1}$ has zeroes for each distinct pair of indices (j, k) *not* in the edge set E of the graph. Consequently, the sparsity pattern of the inverse covariance K encodes the edge structure of the graph. The goal of Gaussian graphical model selection is to determine this unknown edge structure, and hence the sparsity pattern of the inverse covariance matrix. It can also be of interest to estimate the matrices K or Σ, for instance, in the Frobenius or ℓ_2-operator norm sense. In recent years, under the assumption that all entries of X are fully observed, a number of practical methods have been proposed and shown to perform well under high-dimensional scaling (e.g., [2, 5–7]).

Chandrasekaran et al. tackle a challenging extension of this problem, in which one observes only p coordinates of a larger $p + h$ dimensional Gaussian random vector. In this case, the $p \times p$ precision matrix K of the observed components need not be sparse, but rather, by an application of the Schur complement formula, can be written as the difference $K = S^* - L^*$. The first matrix S^* is sparse, whereas the second matrix L^* is not sparse (at least in general), but has rank at most h, corresponding to the number of latent or hidden variables. Consequently, the problem

Received February 2012.