The Annals of Statistics
2012, Vol. 40, No. 4, 1968–1972
DOI: 10.1214/12-AOS979
Main article DOI: 10.1214/11-AOS949

DISCUSSION: LATENT VARIABLE GRAPHICAL MODEL SELECTION VIA CONVEX OPTIMIZATION[1]

BY MING YUAN

Georgia Institute of Technology

I want to start by congratulating Professors Chandrasekaran, Parrilo and Willsky for this fine piece of work. Their paper, hereafter referred to as CPW, addresses one of the biggest practical challenges of Gaussian graphical models—how to make inferences for a graphical model in the presence of missing variables. The difficulty comes from the fact that the validity of conditional independence relationships implied by a graphical model relies critically on the assumption that all conditional variables are observed, which of course can be unrealistic. As CPW shows, this is not as hopeless as it might appear to be. They characterize conditions under which a conditional graphical model can be identified, and offer a penalized likelihood method to reconstruct it. CPW notes that with missing variables, the concentration matrix of the observables can be expressed as the difference between a sparse matrix and a low-rank matrix; and suggests to exploit the sparsity using an ℓ_1 penalty and the low-rank structure by a trace norm penalty. In particular, the trace norm penalty or, more generally, nuclear norm penalties, can be viewed as a convex relaxation to the more direct rank constraint. Its use oftentimes comes as a necessity because rank constrained optimization could be computationally prohibitive. Interestingly, as I note here, the current problem actually lends itself to efficient algorithms in dealing with the rank constraint, and therefore allows for an attractive alternative to the approach of CPW.

1. Rank constrained latent variable graphical Lasso. Recall that the penalized likelihood estimate of CPW is defined as

$$(\hat{S}_n, \hat{L}_n) = \mathop{\arg\min}_{L\succeq 0, S-L\succ 0} \bigl\{-\ell(S-L, \Sigma_O^n) + \lambda_n\bigl(\gamma\|S\|_1 + \operatorname{trace}(L)\bigr)\bigr\},$$

where the vector ℓ_1 norm and trace/nuclear norm penalties are designated to induce sparsity among elements of S and low-rank structure of L respectively. Of course, we can attempt a more direct rank penalty as opposed to the nuclear norm penalty on L, leading to

$$(\hat{S}_n, \hat{L}_n) = \mathop{\arg\min}_{L\succeq 0, S-L\succ 0} \bigl\{-\ell(S-L, \Sigma_O^n) + \lambda_n\bigl(\gamma\|S\|_1 + \operatorname{rank}(L)\bigr)\bigr\};$$

Received February 2012.
[1]Supported in part by NSF Career Award DMS-08-46234.