Model Selection in Gaussian Graphical Models from High-Dimensional Missing Data

Yuxin Chen  Email: yxchen@stanford.edu
EE364b: Convex Optimization II Class Project

Background

- Graphical Models
  - Encode conditional independence among random variables
- Structure Learning
  - Challenging: recover underlying graph structure from a few samples
  - More challenging: what if the samples contain incomplete entries
- Existing Approaches
  (1) Convex Optimization [ChaParWil2010] - a single pattern w/ very few latent variables
  (2) Alternating EM [StaBue2010] - no theoretical guarantee

Motivation

- Consider the following scenario...
  (1) Multiple missing data patterns
  (2) Each pattern may contain only a few entries
- Questions:
  - Is the graphical structure recoverable?
  - If so, under what conditions? (#samples, special structures on graphs...)

Problem Setup

- Underlying Structure: Gaussian Graphical Models $G$
  - $p$-dimensional: $(X_1, \ldots, X_p)$ (high-dimensional regime)
  - Covariance (unknown): $\Sigma$ ($K = \Sigma^{-1}$)
- Samples
  - $n$ samples i.i.d. drawn from $G$
  - $1$ different missing data patterns: each sample belongs to each missing pattern equally likely
- Tasks
  - Given the sample collection, how to recover the edge set of $G$

Exploiting Sparsity

- Sparsity
  - Oftentimes, the associated graph is sparse
  - Gaussian case $\Rightarrow K = \Sigma^{-1}$ is sparse!
- Matrix Completion
  - Index set $\Omega$: $(i,j) = 1$ iff $(i,j) \in \Omega$
  - Estimate $(\hat{\Sigma})_{ij} = K_{ij}$ (for large $n$)
  - minimize $\|\Sigma^{-1}\|_0$ subject to $|\Sigma_{ij} - (\hat{\Sigma})_{ij}| < \epsilon, \forall (i,j) \in \Omega$, Non-convex!

Algorithm: Convex Relaxation

$\Sigma = \begin{bmatrix}
\Sigma_0 & \Phi \\
\Phi & H
\end{bmatrix}$

$\Sigma_0$ is sparse matrix $\Sigma_0^*$ support known $H$ residual

- Applying Taylor expansion:
  $\Sigma_0^* = \Sigma_0 + \sum_{|\Omega|} H_{\Omega} K_{\Omega} H_{\Omega}^T$
  - Treat $W$ as noise $\iff W$ small $\iff H^TKH^T$ and $\Sigma_0 - \Sigma_0^*$ small
- $\ell_1$ Relaxation – sparse representation
  minimize $f(K,H) = \lambda_0 \|K\|_1 + \frac{1}{2} \|\Sigma_0 K \Sigma_0^* - H - \Sigma_0^*\|_F^2$
  subject to $\text{supp}(H) \subseteq \Omega$

Theoretical Guarantee: (Main Result)

- The algorithm recovers the true support if the following assumptions hold:
  (1) $\Sigma_0$ invertible (non-degenerate)
  (2) $\Sigma_0$ and the support set $\hat{S}$ is incoherent (preserves signal energy)
  (3) Each non-zero entry of $K$ contains sufficiently large energy (avoid being buried by noise)
  (4) The residual $\text{HWH}$ is small
  (5) Sufficiently many samples: $n > (p\|\Sigma_0\|_2^2$

Roadmap of Theorem Proofs

- Optimality and Uniqueness Condition
  - $\exists$ a primal-dual pair $(K,Z)$ s.t.
    (1) $Z \in \text{relint}(\partial \|K\|_1)$
    (2) $(K,Z)$ satisfies zero subgradient condition.
- Construction of a Primal-dual Pair
  - Primal-dual witness method
    (1) Study instead the optimizer $\hat{K}$ of another partial convex program
      minimize $\lambda_0 \|K_{\Omega} + \frac{1}{2} \|\Sigma_0 K \Sigma_0^* + H - \Sigma_0^*\|_F^2$
      subject to $\text{supp}(K) \subseteq \Omega$, $\text{supp}(H) \subseteq \Omega$
    - $\hat{K}$ has the correct support.
    (2) Use the zero-subgradient condition to find $\hat{Z}$.
    (3) Test whether $\hat{Z} \in \text{relint}(\partial \|K\|_1)$.

Numerical Example

- Consider Pairwise Missing Patterns
  - Each sample contains only 2 entries
  - Define density $\rho$ as the portion of observed pairs

The Way Ahead

- More general matrix completion problem
  minimize $\|\Sigma^{-1}\|_O$
  subject to $P_1(\Sigma) = P_1(\Sigma_0)$
- Using empirical estimate on covariance is not efficient in sample complexity
  - may consider log-likelihood instead

In general, this is NP-hard $\Rightarrow$ need to exploit structures