Introduction

Latent Variable Gaussian Graphical Model (LVGGM): $X_0 \in \mathbb{R}^d$ is the observed variables and $X_r \in \mathbb{R}^d$ the latent variables, $X_r \sim N(0, \Sigma^r)$, and sparse precision matrix $\Omega = \Sigma^{-1}$. Then $X_r$ follows a normal distribution with marginal covariance matrix $\Sigma^r = \Sigma_{00}^{-1}$ being the top-left block matrix in $\Sigma$. By Schur complement $\Omega = (\Sigma_{00}^{-1}) - \Sigma_{00} \Omega_{00} \Sigma_{00}$, the precision matrix of LVGGM can be written as $\Omega = \Omega^* + \Omega^L$, where $\|\Omega^*\|_0 = s^* \text{ and } \|\Omega^L\| = r$.

The Proposed Estimator

Suppose that we observe i.i.d. samples $X_1, \ldots, X_n$ from $N(0, \Sigma^r)$. The negative log-likelihood function $p_{\theta}(L, S) = \text{tr} \left[ \Sigma \left( \log(L) + \log(S) \right) \right] - \log \| L + S \|$, where $\Sigma = 1/n \sum_{i=1}^n X_i \tilde{X}_i^T$ is the sample covariance matrix, and $\| \cdot \|$ is the determinant of $\Sigma = L + S$. Let $\Sigma = L + Z^T$, where $Z \in \mathbb{R}^{d \times r}$ and $r > n$ is the number of latent variables.

**Estimator:** We propose a nonconvex estimator using sparsity constrained maximum likelihood minimization:

$$q_{\theta, S} (Z) = \text{tr} \left[ \Sigma (S + ZZ^T) \right] - \log \| S + ZZ^T \|, \quad \text{s.t. } \| S \|_0 \leq s$$

where $s > 0$ is a tuning parameter that controls the sparsity of $S$.

**The Proposed Algorithm**

We present the proposed algorithm here, which consists of two stages: initialization and alternating gradient descent.

### Algorithm 1 Alternating Thresholded Gradient Descent (AltGD) for LVGGM

1. **Input:** i.i.d. samples $X_1, \ldots, X_n$, max number of iterations $T$, and parameters $s, n, r, \epsilon$.
2. **Stage I: Initialization**
   - Compute SVD $\Sigma \approx U \Sigma_r V^T$, where $\Sigma_r$ preserves the $r$ largest singular values of $\Sigma$.
   - Compute $S = U \Sigma_r (U \Sigma_r)^{-1}$.
   - Compute $\Omega^L = U \Sigma_r (U \Sigma_r)^{-1}$.
3. **Stage II: Alternating Gradient Descent**
   - For $t = 1, \ldots, T - 1$ do:
     - $\Omega^L = U \Sigma_r (U \Sigma_r)^{-1}$.
     - $\Omega^S = \Sigma - \epsilon \Omega^L$.
     - $\Omega^L = \text{max} \{ 1 - \eta \epsilon \Omega^L, 0 \}$.
     - $\Omega^S = \text{max} \{ 1 - \eta \epsilon \Omega^S, 0 \}$.

4. **Output:** $(\hat{S}, \hat{Z})$.

**Theoretical Analysis**

- **Assumptions**

  - **A1 (Bounded Eigenvalues):** $\Omega^S$ is bounded, i.e., $0 < \lambda_\min(\Sigma^r) \leq \lambda_\max(\Sigma^r) \leq \epsilon$.
  - **A2 (Spikiness Condition):** the spikiness ratio is defined as $\frac{\| L^* \|_F}{\| L \|_F}$.

5. **FOS (First-Order Stability):** If $\| \Sigma \|_F \leq 2 \gamma_1 \| L \|_F$, then $\| \Sigma - \epsilon \|_F \leq 2 \gamma_2 \| L \|_F$, where $\gamma_1, \gamma_2$ are constants and $\| (L, Z) \| = \min \{ (s, r) \}$.

6. **Validation of Initialization:** Suppose A1 and A2 hold. Assume $n \geq c r^2 s^* \log d / \epsilon R^2$ and $s^* \leq c L^* \epsilon^{-1}$, where $R$ is a constant depending on $\epsilon$. Then with probability at least $1 - 1/d$, we have $\max \{ \| \hat{S} (1) - S^* \|_F, \| \hat{Z} (1) - Z^* \|_F \} \leq \frac{\tau}{\sqrt{1 - \psi^2}} + \frac{\sqrt{d}}{2 r^2 d}$, where $C > 1$ is an absolute constant.

**Main Remarks**

- The initial points returned by the initialization stage of AltGD fall in small neighborhoods of $S^*$ and $Z^*$ if $n = O(s^* \log d)$, which essentially attains the optimal sample complexity for LVGGM estimation. In addition, we require $s^* \leq d^2 / \epsilon^{\rho^2}$, which means the unknown sparse matrix cannot be too dense.

- The statistical error scales as $\max \{ O(\sqrt{s^* \log d}), O(\sqrt{d^2 / \epsilon^{\rho^2}}) \}$, where $O(\sqrt{\log d})$ corresponds to the statistical error, the statistical error, and $O(\sqrt{d^2 / \epsilon^{\rho^2}})$ corresponds to that of $L^*$ (or equivalently $Z^*$).

**Numerical Simulations**

- **Data Generation:** We randomly generated a sparse positive definite matrix $\Omega_0 \in \mathbb{R}^{d \times d}$ with sparsity $s = 0.02d$, where $\Omega^0 \approx \Omega_0$, and $L^* = \Omega_0 (d_1, d_2 \cdot \delta_1, \delta_2) \Omega_0 (d_1, d_2 \cdot \delta_1, \delta_2) \Omega_0 (d_1, d_2 \cdot \delta_1, \delta_2) \Omega_0 (d_1, d_2 \cdot \delta_1, \delta_2) \Omega_0 (d_1, d_2 \cdot \delta_1, \delta_2)$.

- We sampled $X_1, \ldots, X_n \sim N(0, (\Omega^0)^{-1})$, where $\Omega^0 = L^* + L^L$.

- **Validation of Convergence Rate:**

**Experiments on Genomic Datasets**

**Table:** Summary of CPU time on 9 time uniform subtypes breast cancer dataset.

<table>
<thead>
<tr>
<th>Method</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPA</td>
<td>38.63</td>
</tr>
<tr>
<td>ADMM</td>
<td>85.01</td>
</tr>
<tr>
<td>Glasso</td>
<td>7.67</td>
</tr>
</tbody>
</table>

**Figure:** An example of subnetwork in the transcriptional regulatory network of luminal breast cancer. Gray edges are the interactions from the Cancer Genome Database; red edges are the ones inferred by the respective methods; green edges are incorrectly inferred interactions.