
Compressive Sensing of Signals from a GMM with Sparse Precision Matrices (Supplementary Material)

¹Jianbo Yang ¹Xuejun Liao ²Minhua Chen ¹Lawrence Carin

¹Department of Electrical and Computer Engineering, Duke University

²Department of Statistics & Department of Computer Science, University of Chicago

{jianbo.yang; xjliao; lcarin@duke@duke.edu}, {dukemeeting@gmail.com}

1 Demonstration of GMRF on Image Signals

An image of size 321×481 randomly selected from the Berkeley Segmentation Dataset,¹ is partitioned into a collection of fully-overlapped equal-sized local patches, each of size 8×8 , which are vectorized to constitute the image signals $\{\mathbf{x}_i\}_{i=1}^N$.² We use a standard GMM to fit the signals $\{\mathbf{x}_i\}_{i=1}^N$. The parameters are estimated by the EM algorithm [1], without any prior put on the parameters $\{\pi^{(k)}, \boldsymbol{\mu}^{(k)}, \boldsymbol{\Omega}^{(k)}\}$. The number of the GMM components is set as eight. Since N is fairly large (more than 450,000) relative to the number of the parameters of the GMM (around 528), we mildly assume that this estimated GMM is a good approximation of the ground truth distribution $p(\mathbf{x})$ for the signals $\{\mathbf{x}_i\}_{i=1}^N$. Figure 1 plots the precision matrices of the eight Gaussian components of the estimated GMM.

It is evident from this figure that the precision matrices are sparse with many off-diagonal elements close to zero and each precision matrix has the banding pattern. The bands constituted by the significant nonzero elements are located in the main diagonal and a few lines below or above the main diagonal. As illustrated in Figure 2, those significant nonzero elements in the bands correspond to the pixel-pairs in which the two pixels are neighbors in a 2-D patch. Note that when one chooses other images, uses other patch sizes (e.g., 6×6 , 12×12) or sets a different K (K could be tuned by Akaike information criterion or Bayesian information criterion, or inferred by a Dirichlet process), the resultant plots are similar to those in Figure 1 and 2.

2 Updates of variational EM algorithm

The full update expressions of the variational EM are given below.

Update $q(\mathbf{X}, \mathbf{Z})$: From

$$q^*(\mathbf{X}, \mathbf{Z}) \propto \exp(\mathbb{E}[\ln p(\mathbf{Y}|\mathbf{X}, \kappa)p(\mathbf{X}|\mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\Omega})p(\mathbf{Z}|\boldsymbol{\pi})]),$$

we obtain

$$q^*(\mathbf{X}, \mathbf{Z}) = \prod_{i=1}^N \rho_i^{(z_i)} \mathcal{N}(\mathbf{x}_i | \boldsymbol{\eta}_i^{(z_i)}, \mathbf{C}_i^{(z_i)})$$

¹<http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/resources.html>

²These image signals can be taken as the groundtruth when one performs CS on the image signal $\{\mathbf{x}_i\}_{i=1}^N$.

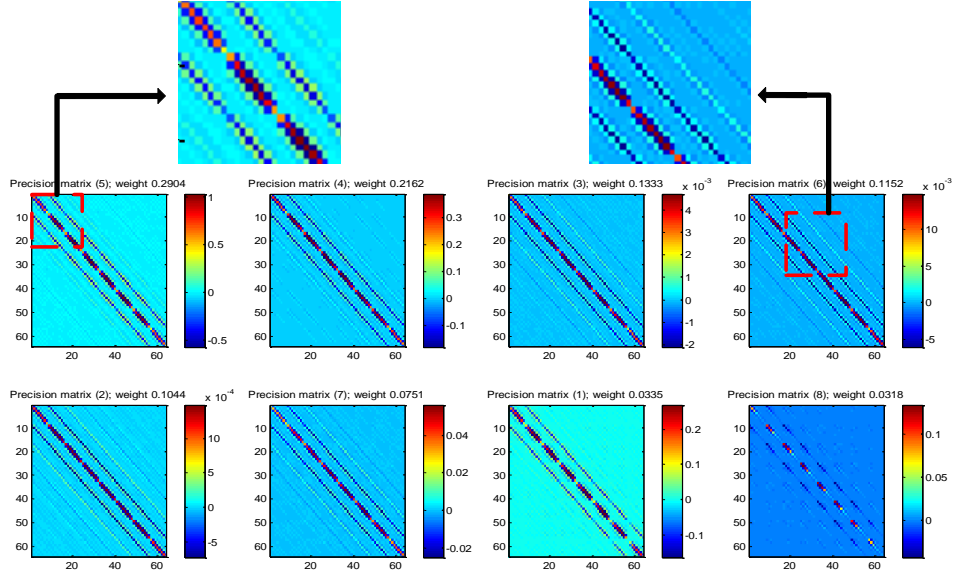


Figure 1: The plots of the eight precision matrices of the GMM estimated from $\{x_i\}_{i=1}^N$ by an EM algorithm. The regions in the red dash-line boxes in the first and fourth plots are enlarged and shown on the top of the plots of the eight precision matrices. In the title of each plot, the index of the component as well as the corresponding weights of the mixture model are shown.

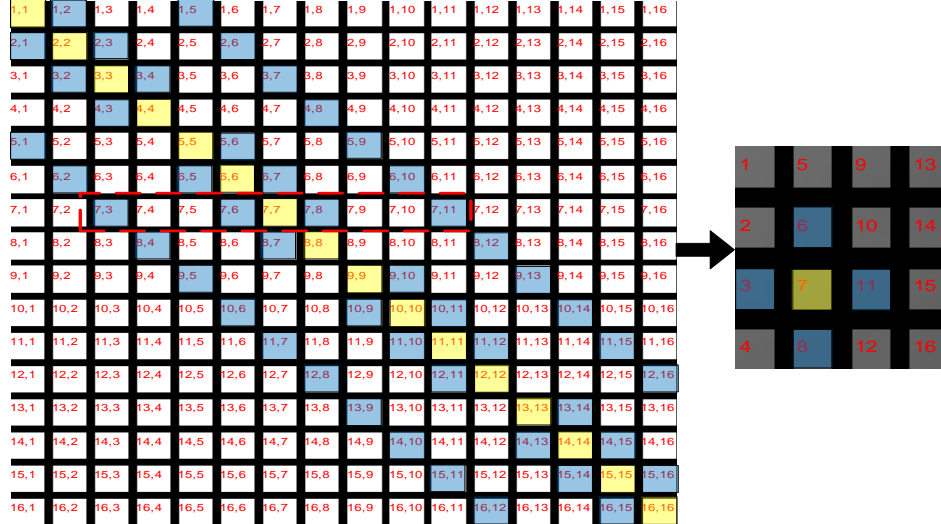


Figure 2: The diagram of the relationship between the 16 pixels in a 4×4 2-D patch (right) and the 196 elements in a 16×16 precision matrix (left). The pairs of between a pixel (in yellow) and each of its four immediate neighbors (in blue), left, right, up and down, in the right figure are considered to have the conditional dependence relationship as highlighted in blue in the left figure. The region in the red dash-line box in the left figure shows the case of Pixel 7 and its four immediate neighbors.

with

$$\begin{aligned}
 C_i^{(z_i)} &= (\kappa^* \Phi_i' \Phi_i + \Omega^{(z_i)*})^{-1} \\
 \eta_i^{(z_i)} &= C_i^{(z_i)} (\kappa^* \Phi_i' y_i + \Omega^{(z_i)*} \mu^{(z_i)*}) \\
 \rho_i^{(z_i)} &= \frac{\pi^{(z_i)} \mathcal{N}(y | \Phi \mu^{(z_i)*}, \kappa^{*-1} \mathbf{I} + \Phi \Omega^{(z_i)*-1} \Phi')}{\sum_{l=1}^K \pi^{(l)} \mathcal{N}(y | \Phi \mu^{(l)*}, \kappa^{*-1} \mathbf{I} + \Phi \Omega^{(l)*-1} \Phi')}.
 \end{aligned}$$

Hence, $\mathbb{E}(\mathbf{x}_i) = \sum_{i=k}^K \rho_i^{(k)} \boldsymbol{\eta}_i^{(k)}$.

Update $q(\boldsymbol{\alpha})$: From

$$q^*(\boldsymbol{\alpha}) \propto \exp(\mathbb{E}[\ln p(\boldsymbol{\Omega}|\boldsymbol{\alpha})p(\boldsymbol{\alpha}|\boldsymbol{\gamma})]),$$

we obtain

$$q^*(\boldsymbol{\alpha}) = \prod_{k=1}^K \prod_{s=1}^n \prod_{t < s} \text{InvGau}(\alpha_{ts}^{(k)} | \sqrt{\frac{\mathbb{E}(\gamma_{ts}^{(k)})}{\omega_{ts}^{(k)*2}}}, \mathbb{E}(\gamma_{ts}^{(k)})),$$

where $\text{InvGau}(x|g, h) = \sqrt{\frac{h}{2\pi x^3}} \exp(-\frac{h(x-g)^2}{2g^2 x})$ denotes an inverse Gaussian distribution. Hence,

$$\mathbb{E}(\alpha_{ts}) = \sqrt{\frac{\mathbb{E}(\gamma_{ts}^{(k)})}{\omega_{ts}^{(k)*2}}} \text{ and } \mathbb{E}(\alpha_{ts}^{-1}) = (\sqrt{\frac{\mathbb{E}(\gamma_{ts}^{(k)})}{\omega_{ts}^{(k)*2}}})^{-1} + \mathbb{E}(\gamma_{ts}^{(k)})^{-1}.$$

Update $q(\boldsymbol{\gamma})$: From

$$q^*(\boldsymbol{\gamma}) \propto \exp(\mathbb{E}[\ln p(\boldsymbol{\alpha}|\boldsymbol{\gamma})p(\boldsymbol{\gamma})]),$$

we obtain

$$q^*(\boldsymbol{\gamma}) = \prod_{k=1}^K \prod_{s=1}^n \prod_{t < s} \text{Ga}(\gamma_{ts}^{(k)} | a_0 + 1, b_0 + \frac{\mathbb{E}(\alpha_{ts}^{(k)-1})}{2}).$$

$$\text{Hence, } \mathbb{E}(\gamma_{ts}) = \frac{a_0 + 1}{b_0 + \frac{\mathbb{E}(\alpha_{ts}^{(k)-1})}{2}}.$$

Update κ :

$$\kappa^* = \arg \max_{\kappa} \{ \mathbb{E}_{\mathbf{X}}[\ln p(\mathbf{Y}|\mathbf{X}, \kappa)p(\kappa)] \}$$

$$\kappa^* = \frac{\frac{Nm}{2} + e_0}{\frac{\sum_{i=1}^N \sum_{k=1}^K \rho_i^{(k)} [(\mathbf{y}_i - \boldsymbol{\Phi}_i \boldsymbol{\eta}_i^{(k)})' (\mathbf{y}_i - \boldsymbol{\Phi}_i \boldsymbol{\eta}_i^{(k)}) + \text{tr}(\boldsymbol{\Phi}_i' \boldsymbol{\Phi}_i (\bar{\boldsymbol{\Omega}}_i^{(k)})^{-1})]}{2}} + f_0}$$

Update $\boldsymbol{\pi}$: Let $\eta^{(k)} = \alpha_0 + N^{(k)}$ with $N^{(k)} = \sum_{i=1}^N \rho_i^{(k)}$, then

$$\boldsymbol{\pi}^* = \arg \max_{\boldsymbol{\pi}} \left\{ \mathbb{E}_{\mathbf{Z}}[\ln p(\mathbf{Z}|\boldsymbol{\pi})p(\boldsymbol{\pi})] + \lambda \left(\sum_{k=1}^K \pi^{(k)} - 1 \right) \right\}$$

where λ is a Lagrange multiplier. We obtain that

$$\pi^{(k)*} = \frac{\eta^{(k)} - 1}{\sum_{i=1}^K \eta^{(i)} - K}.$$

Update $\boldsymbol{\mu}^{(k)}$: Let $\bar{\mathbf{x}}^{(k)} = \frac{1}{N^{(k)}} \sum_{i=1}^N \rho_i^{(k)} \mathbb{E}(\mathbf{x}_i)$, then

$$\boldsymbol{\mu}^{(k)*} = \arg \max_{\boldsymbol{\mu}^{(k)}} \left\{ \mathbb{E}_{\mathbf{X}, \mathbf{Z}}[\ln p(\boldsymbol{\mu}^{(k)} | \mathbf{m}_0, \beta_0, \boldsymbol{\Omega}^{(k)})p(\mathbf{X}|\mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\Omega})] \right\}.$$

We obtain that

$$\boldsymbol{\mu}^{(k)*} = \frac{\beta_0 \mathbf{m}_0 + N^{(k)} \bar{\mathbf{x}}^{(k)}}{\beta_0 + N^{(k)}}.$$

Update $\boldsymbol{\Omega}^{(k)}$:

$$\boldsymbol{\Omega}^{(k)*} = \arg \max_{\boldsymbol{\Omega}^{(k)}} \left\{ \mathbb{E}_{\boldsymbol{\alpha}, \mathbf{Z}, \mathbf{X}}[\ln p(\boldsymbol{\Omega}^{(k)} | \boldsymbol{\alpha})p(\mathbf{X}|\mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\Omega})] \right\}$$

For convenience, we let

$$\mathbf{\Omega}^{(k)} = \begin{bmatrix} \mathbf{\Omega}_{\tilde{s}\tilde{s}}^{(k)} & \mathbf{\Omega}_{\tilde{s}s}^{(k)} \\ \mathbf{\Omega}_{s\tilde{s}}^{(k)} & \mathbf{\Omega}_{ss}^{(k)} \end{bmatrix}$$

Using identity:

$$\mathcal{N}(\mathbf{x}_i | \mathbf{0}, \mathbf{\Omega}^{(k)})^{-1} = \mathcal{N}(\mathbf{x}_{\tilde{s}i} | -(\mathbf{\Omega}_{\tilde{s}\tilde{s}}^{(k)})^{-1} \mathbf{\Omega}_{\tilde{s}s}^{(k)} \mathbf{x}_{si}, \mathbf{\Omega}_{\tilde{s}\tilde{s}}^{(k)})^{-1} \mathcal{N}(\mathbf{x}_{si} | \mathbf{0}, (\mathbf{\Omega}_{ss}^{(k)} - \mathbf{\Omega}_{\tilde{s}s}^{(k)'} \mathbf{\Omega}_{\tilde{s}\tilde{s}}^{(k)-1} \mathbf{\Omega}_{\tilde{s}s}^{(k)})^{-1})$$

To make \mathbf{x}_i have the $\mathbf{0}$ mean, we let $\tilde{\mathbf{x}}_i = \mathbf{x}_i - \boldsymbol{\mu}^{(k)*}$

$$\begin{aligned} & (\mathbf{\Omega}_{\tilde{s}s}^{(k)*}, \mathbf{\Omega}_{ss}^{(k)*}) \\ &= \arg \max_{\mathbf{\Omega}_{ss}^{(k)}, \mathbf{\Omega}_{\tilde{s}s}^{(k)}} \mathbb{E}_{\mathbf{X}, \boldsymbol{\alpha}, \mathbf{Z}} [\log [\prod_{i=1}^N \mathcal{N}(\tilde{\mathbf{x}}_{\tilde{s}i} | -(\mathbf{\Omega}_{\tilde{s}\tilde{s}}^{(k)})^{-1} \mathbf{\Omega}_{\tilde{s}s}^{(k)} \tilde{\mathbf{x}}_{si}, \mathbf{\Omega}_{\tilde{s}\tilde{s}}^{(k)})^{-1} \mathbf{z}_{ik} \\ & \quad \times \mathcal{N}(\tilde{\mathbf{x}}_{si} | \mathbf{0}, (\mathbf{\Omega}_{ss}^{(k)} - \mathbf{\Omega}_{\tilde{s}s}^{(k)'} \mathbf{\Omega}_{\tilde{s}\tilde{s}}^{(k)-1} \mathbf{\Omega}_{\tilde{s}s}^{(k)})^{-1} \mathbf{z}_{ik} \times \mathcal{N}(\mathbf{\Omega}_{\tilde{s}s}^{(k)} | \mathbf{0}, \text{diag}^{-1}(\boldsymbol{\alpha}_{ss}^{(k)})) \times 1(\mathbf{\Omega}^{(k)} \in \mathbf{S}_+)]]] \end{aligned}$$

We can obtain

$$\begin{aligned} \mathbf{\Omega}_{\tilde{s}s}^{(k)*} &= (N^{(k)} \bar{\boldsymbol{\Sigma}}_{ss}^{(k)} (\mathbf{\Omega}_{\tilde{s}\tilde{s}}^{(k)*})^{-1} + \text{diag}(\mathbb{E}(\boldsymbol{\alpha}_{ss}^{(k)})))^{-1} (-N^{(k)} \bar{\boldsymbol{\Sigma}}_{ss}^{(k)}) \\ \mathbf{\Omega}_{ss}^{(k)*} &= ((\bar{\boldsymbol{\Sigma}}_{ss}^{(k)})^{-1} + \mathbf{\Omega}_{\tilde{s}s}^{(k)*'} \mathbf{\Omega}_{\tilde{s}\tilde{s}}^{(k)*-1} \mathbf{\Omega}_{\tilde{s}s}^{(k)*}) \end{aligned}$$

where

$$\bar{\boldsymbol{\Sigma}}^{(k)} = \frac{1}{N^{(k)}} \sum_{i=1}^N \rho_i^{(k)} [(\boldsymbol{\eta}_i^{(k)} - \boldsymbol{\mu}^{(k)*})(\boldsymbol{\eta}_i^{(k)} - \boldsymbol{\mu}^{(k)*})' + \mathbf{C}_i^{(k)}].$$

The optimal $\mathbf{\Omega}^{(k)*}$ computed by the above procedure is guaranteed to satisfied to the positive definite constraint, as proved in [2].

3 Proof of Theorem 1

Proof of Theorem 1: A Taylor expansion of the inverse matrix gives

$$\begin{aligned} & (\mathbf{\Omega}_0 + \boldsymbol{\Delta} + \boldsymbol{\Phi}_i' \mathbf{R}^{-1} \boldsymbol{\Phi}_i)^{-1} \\ &= \left[\mathbf{I} + (\mathbf{\Omega}_0 + \boldsymbol{\Phi}_i' \mathbf{R}^{-1} \boldsymbol{\Phi}_i)^{-1} \boldsymbol{\Delta} \right]^{-1} (\mathbf{\Omega}_0 + \boldsymbol{\Phi}_i' \mathbf{R}^{-1} \boldsymbol{\Phi}_i)^{-1}, \\ &= \sum_{n=0}^{\infty} (-1)^n \left[(\mathbf{\Omega}_0 + \boldsymbol{\Phi}_i' \mathbf{R}^{-1} \boldsymbol{\Phi}_i)^{-1} \boldsymbol{\Delta} \right]^n (\mathbf{\Omega}_0 + \boldsymbol{\Phi}_i' \mathbf{R}^{-1} \boldsymbol{\Phi}_i)^{-1}. \end{aligned} \quad (1)$$

Based on (1), we obtain the signal reconstruction error

$$\tilde{\mathbf{x}}_i - \hat{\mathbf{x}}_i = \sum_{n=1}^{\infty} \left[-(\mathbf{\Omega}_0 + \boldsymbol{\Phi}_i' \mathbf{R}^{-1} \boldsymbol{\Phi}_i)^{-1} \boldsymbol{\Delta} \right]^n (\mathbf{\Omega}_0 + \boldsymbol{\Phi}_i' \mathbf{R}^{-1} \boldsymbol{\Phi}_i)^{-1} \boldsymbol{\Phi}_i' \mathbf{R}^{-1} (\mathbf{y}_i - \boldsymbol{\Phi}_i \boldsymbol{\mu}).$$

Applying some norm, e.g., ℓ_2 norm on vectors and F norm on matrix, on both sides and using the properties of norms, we obtain

$$\begin{aligned} \|\tilde{\mathbf{x}}_i - \hat{\mathbf{x}}_i\|_2 &\leq \sum_{n=1}^{\infty} \left\| (\mathbf{\Omega}_0 + \boldsymbol{\Phi}_i' \mathbf{R}^{-1} \boldsymbol{\Phi}_i)^{-1} \boldsymbol{\Delta} \right\|_F^n \left\| (\mathbf{\Omega}_0 + \boldsymbol{\Phi}_i' \mathbf{R}^{-1} \boldsymbol{\Phi}_i)^{-1} \boldsymbol{\Phi}_i' \mathbf{R}^{-1} (\mathbf{y}_i - \boldsymbol{\Phi}_i \boldsymbol{\mu}) \right\|_2, \\ &= \sum_{n=1}^{\infty} \left\| (\mathbf{\Omega}_0 + \boldsymbol{\Phi}_i' \mathbf{R}^{-1} \boldsymbol{\Phi}_i)^{-1} \boldsymbol{\Delta} \right\|_F^n \|\hat{\mathbf{x}}_i - \boldsymbol{\mu}\|_2, \\ &= \frac{\left\| (\mathbf{\Omega}_0 + \boldsymbol{\Phi}_i' \mathbf{R}^{-1} \boldsymbol{\Phi}_i)^{-1} \boldsymbol{\Delta} \right\|_F}{1 - \left\| (\mathbf{\Omega}_0 + \boldsymbol{\Phi}_i' \mathbf{R}^{-1} \boldsymbol{\Phi}_i)^{-1} \boldsymbol{\Delta} \right\|_F} \|\hat{\mathbf{x}}_i - \boldsymbol{\mu}\|_2, \end{aligned} \quad (2)$$

where the last equation arises from the assumption that $\|(\mathbf{\Omega}_0 + \mathbf{\Phi}'_i \mathbf{R}^{-1} \mathbf{\Phi}_i)^{-1} \mathbf{\Delta}\|_F < 1$.

Because $\left\|(\mathbf{\Omega}_0 + \mathbf{\Phi}'_i \mathbf{R}^{-1} \mathbf{\Phi}_i)^{-1} \mathbf{\Delta}\right\|_F \leq \left\|(\mathbf{\Omega}_0 + \mathbf{\Phi}'_i \mathbf{R}^{-1} \mathbf{\Phi}_i)^{-1}\right\|_F \|\mathbf{\Delta}\|_F$, one has

$$1 - \left\|(\mathbf{\Omega}_0 + \mathbf{\Phi}'_i \mathbf{R}^{-1} \mathbf{\Phi}_i)^{-1} \mathbf{\Delta}\right\|_F \geq 1 - \left\|(\mathbf{\Omega}_0 + \mathbf{\Phi}'_i \mathbf{R}^{-1} \mathbf{\Phi}_i)^{-1}\right\|_F \|\mathbf{\Delta}\|_F.$$

When $\left\|(\mathbf{\Omega}_0 + \mathbf{\Phi}'_i \mathbf{R}^{-1} \mathbf{\Phi}_i)^{-1}\right\|_F \|\mathbf{\Delta}\|_F < 1$, it follows from (2) that

$$\|\tilde{\mathbf{x}}_i - \hat{\mathbf{x}}_i\|_2 \leq \frac{\left\|(\mathbf{\Omega}_0 + \mathbf{\Phi}'_i \mathbf{R}^{-1} \mathbf{\Phi}_i)^{-1}\right\|_F \|\mathbf{\Delta}\|_F}{1 - \left\|(\mathbf{\Omega}_0 + \mathbf{\Phi}'_i \mathbf{R}^{-1} \mathbf{\Phi}_i)^{-1}\right\|_F \|\mathbf{\Delta}\|_F} \|\hat{\mathbf{x}}_i - \boldsymbol{\mu}\|_2. \quad (3)$$

When $\boldsymbol{\mu} = \mathbf{0}$, we have

$$\|\tilde{\mathbf{x}}_i - \hat{\mathbf{x}}_i\|_2 \leq \frac{\left\|(\mathbf{\Omega}_0 + \mathbf{\Phi}'_i \mathbf{R}^{-1} \mathbf{\Phi}_i)^{-1}\right\|_F \|\mathbf{\Delta}\|_F}{1 - \left\|(\mathbf{\Omega}_0 + \mathbf{\Phi}'_i \mathbf{R}^{-1} \mathbf{\Phi}_i)^{-1}\right\|_F \|\mathbf{\Delta}\|_F} \|\hat{\mathbf{x}}_i\|_2. \quad (4)$$

This completes the proof.

4 Proof of Theorem 2

Proof of Theorem 2: The main idea of proof is inspired by [3][4]. Recall that $\mathbf{\Omega}^* = \mathbf{\Omega}_0 + \mathbf{\Delta}$ is the estimate of $\mathbf{\Omega}_0$ that minimizes

$$\mathcal{F}(\mathbf{\Omega}) = \text{tr}(\mathbf{\Omega} \mathbf{\Sigma}_{em}) - \log |\mathbf{\Omega}| + \sum_{(i,j) \in S \cup S^c} \eta_{ij} |\omega_{ij}| \quad (5)$$

where $\eta_{ij} = \frac{\sqrt{\tau \gamma_{ij}}}{N}$. Let $\mathbf{\Omega} = \mathbf{\Omega}_0 + \hat{\mathbf{\Delta}}$, and consider the set

$$\Theta_N(M) = \left\{ \hat{\mathbf{\Delta}} : \hat{\mathbf{\Delta}} = \hat{\mathbf{\Delta}}', \|\hat{\mathbf{\Delta}}\|_F = MR_N \right\}$$

where

$$R_N = \sqrt{\frac{(n+s) \log n}{N}} + \sqrt{n+s} \bar{\sigma}_{de}.$$

Note that $\mathcal{F}(\mathbf{\Omega})$ is a convex function and $\mathcal{F}(\mathbf{\Omega}_0 + \mathbf{\Delta}) \leq \mathcal{F}(\mathbf{\Omega}_0)$. Therefore, if we prove that the following probability tends to 1³

$$p(\inf\{\mathcal{F}(\mathbf{\Omega}_0 + \hat{\mathbf{\Delta}}) | \hat{\mathbf{\Delta}} \in \Theta_N(M)\} > \mathcal{F}(\mathbf{\Omega}_0)). \quad (6)$$

then $\mathbf{\Delta}$ corresponding to the minimizer $\mathbf{\Omega}^*$ are inside the sphere defined by $\Theta_N(M)$ and $\|\mathbf{\Delta}\|_F \leq MR_N$ with high probability, equivalently, $\|\mathbf{\Omega}^* - \mathbf{\Omega}_0\|_F^2 = \mathcal{O}_p\{\sqrt{(n+s) \log n/N} + \sqrt{n+s} \bar{\sigma}_{de}\}$. In the following, we give the proof of expression (6).

According to (5), we obtain that

$$\mathcal{F}(\mathbf{\Omega}) - \mathcal{F}(\mathbf{\Omega}_0) = I_1 + I_2 + I_3,$$

where

$$I_1 = \text{tr}(\mathbf{\Omega} \mathbf{\Sigma}_{em}) - \text{tr}(\mathbf{\Omega}_0 \mathbf{\Sigma}_{em}) - \text{tr}(\hat{\mathbf{\Delta}} \mathbf{\Sigma}_0) = \text{tr}(\hat{\mathbf{\Delta}} (\mathbf{\Sigma}_{em} - \mathbf{\Sigma}_0))$$

$$I_2 = -\log |\mathbf{\Omega}| + \log |\mathbf{\Omega}_0| + \text{tr}(\hat{\mathbf{\Delta}} \mathbf{\Sigma}_0) = \text{vec}(\hat{\mathbf{\Delta}})' \left[\int_0^1 (1-\nu) (\mathbf{\Omega}_0 + \nu \hat{\mathbf{\Delta}})^{-1} \otimes (\mathbf{\Omega}_0 + \nu \hat{\mathbf{\Delta}})^{-1} d\nu \right] \text{vec}(\hat{\mathbf{\Delta}})$$

$$I_3 = \sum_{(i,j) \in S \cup S^c} \eta_{ij} (|\omega_{ij}| - |\omega_{ij}^0|)$$

³More exactly, the claim hold with probability greater than $1 - c_1 \exp(-c_2 \log n)$, where $c_1, c_2 > 0$ are independent of other parameters.

where I_2 is obtained by using Taylor's expansion, and $\text{vec}(\hat{\Delta})$ denotes the vectorized $\hat{\Delta}$ and \otimes represents the Kronecker product.

For I_1 :

Let Δ_{ij} , σ_{ij}^{em} , σ_{ij}^{em0} , σ_{ij}^{de} , σ_{ij}^0 are the elements at the i -th row and j -th column of $\hat{\Delta}$, Σ_{em} , Σ_{em}^0 , Σ_{de} , Σ_0 , respectively. We obtain the following inequality for I_1 as

$$I_1 \geq -|I_1| \geq -\underbrace{\left| \sum_{i \neq j} (\sigma_{ij}^{em} - \sigma_{ij}^0) \hat{\Delta}_{ij} \right|}_{I_1^{(1)}} - \underbrace{\left| \sum_{i=1}^n (\sigma_{ii}^{em} - \sigma_{ii}^0) \hat{\Delta}_{ii} \right|}_{I_1^{(2)}}.$$

Based on the Lemma 1 of [4] and Lemma 2 of [3], we obtain that, with probability tending to 1,

$$\begin{aligned} \max_{i \neq j} |\sigma_{ij}^{em} - \sigma_{ij}^0| &\leq \max_{i \neq j} \{ |\sigma_{ij}^{em0} - \sigma_{ij}^0| + |\sigma_{ij}^{de}| \} \\ &\leq C_1 \sqrt{\frac{\log n}{N}} + \bar{\sigma}_{de} \\ &\leq C_1 \left(\sqrt{\frac{\log n}{N}} + \bar{\sigma}_{de} \right). \end{aligned}$$

where $C_1 \geq 1$. Let $\hat{\Delta}^+$ be a diagonal matrix with the diagonal same as $\hat{\Delta}$ and $\hat{\Delta}^- = \hat{\Delta} - \hat{\Delta}^+$. We obtain

$$I_1^{(1)} \leq C_1 \left(\sqrt{\frac{\log n}{N}} + \bar{\sigma}_{de} \right) |\hat{\Delta}^-|_1$$

where $|\mathbf{X}|$ is the ℓ_1 norm of the vectorized matrix \mathbf{X} .

Similarly, according to Cauchy-Schwartz inequality, we obtain the expression

$$\begin{aligned} I_1^{(2)} &\leq \left[\sum_{i=1}^n (\sigma_{ii}^{em} - \sigma_{ii}^0)^2 \right]^{1/2} \|\hat{\Delta}^+\|_F \\ &\leq \sqrt{n} \max_{1 \leq i \leq n} |\sigma_{ii}^{em} - \sigma_{ii}^0| \|\hat{\Delta}^+\|_F \\ &\leq \sqrt{n} \max_{1 \leq i \leq n} (|\sigma_{ii}^{em0} - \sigma_{ii}^0| + |\sigma_{ii}^{de}|) \|\hat{\Delta}^+\|_F \\ &\leq \left(C_2 \sqrt{\frac{n \log n}{N}} + \sqrt{n} \bar{\sigma}_{de} \right) \|\hat{\Delta}^+\|_F \\ &\leq C_2 \left(\sqrt{\frac{(n+s) \log n}{N}} + \sqrt{n+s} \bar{\sigma}_{de} \right) \|\hat{\Delta}^+\|_F. \end{aligned}$$

holds with also probability tending to 1, where $C_2 \geq 1$.

For I_2

Let $\|\mathbf{X}\|$ be the operator norm defined by $\|\mathbf{X}\| = \sqrt{\psi_{\max}(\mathbf{X}\mathbf{X}^T)}$. We obtain that

$$\begin{aligned} I_2 &\geq \int_0^1 (1-\nu) \min_{0 \leq \nu \leq 1} \psi_{\min}[(\Omega_0 + \nu \hat{\Delta})^{-1} \otimes (\Omega_0 + \nu \hat{\Delta})^{-1}] d\nu \|\hat{\Delta}\|_F^2 \\ &\geq \frac{1}{2} \|\hat{\Delta}\|_F^2 \min_{0 \leq \nu \leq 1} \psi_{\max}^{-2}(\Omega_0 + \nu \hat{\Delta}) \\ &\geq \frac{1}{2} \|\hat{\Delta}\|_F^2 (\|\Omega_0\| + \|\hat{\Delta}\|)^{-2} \\ &\geq \frac{1}{2} \|\hat{\Delta}\|_F^2 (\tau_1^{-1} + o(1))^{-2} \end{aligned}$$

with probability tending to 1, since $\|\hat{\Delta}\| \leq \|\hat{\Delta}\|_F = o(1)$, where the symbols $\psi_{\min}(\mathbf{X})$ and $\psi_{\max}(\mathbf{X})$ denote the smallest and largest eigenvalues of matrix \mathbf{X} respectively.

For I_3

Since

$$|\omega_{ij}| - |\omega_{ij}^0| = \begin{cases} |\omega_{ij}| = |\hat{\Delta}_{ij}| & \text{if } (i, j) \in S^c \\ |\omega_{ij} + \hat{\Delta}_{ij}| - |\omega_{ij}^0| \geq -|\hat{\Delta}_{ij}| & \text{if } (i, j) \in S \end{cases},$$

we obtain that

$$I_3 \geq \sum_{(i,j) \in S^c} \eta_{ij} |\hat{\Delta}_{ij}| - \sum_{(i,j) \in S} \eta_{ij} |\hat{\Delta}_{ij}| \geq \underline{\eta} |\hat{\Delta}_{S^c}^-| - \bar{\eta} |\hat{\Delta}_S^-|.$$

Let

$$\begin{aligned} \underline{\eta} &= \frac{C_1}{\epsilon_1} \left(\sqrt{\frac{\log n}{N}} + \bar{\sigma}_{de} \right) \\ \bar{\eta} &= \frac{C_1}{\epsilon_2} \left(\sqrt{\frac{\log n}{N}} + \bar{\sigma}_{de} \right). \end{aligned}$$

We obtain that

$$\begin{aligned} & I_1 + I_2 + I_3 \\ & \geq \frac{1}{2} \|\hat{\Delta}\|_F^2 (\tau_1^{-1} + o(1))^{-2} - C_1 \left(\sqrt{\frac{\log n}{N}} + \bar{\sigma}_{de} \right) |\hat{\Delta}^-|_1 \\ & \quad - C_2 \left(\sqrt{\frac{(n+s) \log n}{N}} + \sqrt{n+s} \bar{\sigma}_{de} \right) \|\hat{\Delta}^+\|_F + \underline{\eta} |\hat{\Delta}_{S^c}^-| - \bar{\eta} |\hat{\Delta}_S^-| \\ & \geq \frac{1}{2} \|\hat{\Delta}\|_F^2 (\tau_1^{-1} + o(1))^{-2} - C_1 \left(\sqrt{\frac{\log n}{N}} + \bar{\sigma}_{de} \right) \left(1 - \frac{1}{\epsilon_1} \right) |\hat{\Delta}_{S^c}^-|_1 \\ & \quad - C_1 \left(\sqrt{\frac{\log n}{N}} + \bar{\sigma}_{de} \right) \left(1 + \frac{1}{\epsilon_2} \right) |\hat{\Delta}_S^-|_1 - C_2 \left(\sqrt{\frac{(n+s) \log n}{N}} + \sqrt{n+s} \bar{\sigma}_{de} \right) \|\hat{\Delta}^+\|_F \\ & \geq \frac{1}{2} \|\hat{\Delta}\|_F^2 (\tau_1^{-1} + o(1))^{-2} - C_1 \left(\sqrt{\frac{\log n}{N}} + \bar{\sigma}_{de} \right) \left(1 + \frac{1}{\epsilon_2} \right) \sqrt{s+n} \|\hat{\Delta}\|_F \\ & \quad - C_2 \left(\sqrt{\frac{(n+s) \log n}{N}} + \sqrt{n+s} \bar{\sigma}_{de} \right) \|\hat{\Delta}\|_F \\ & \geq \|\hat{\Delta}\|_F^2 \left(\frac{1}{2} (\tau_1^{-1} + o(1))^{-2} - \frac{C_1(1+\epsilon_2)}{\epsilon_2 M} - \frac{C_2}{M} \right) \geq 0 \end{aligned}$$

for sufficiently large M . The third inequality is hold, since the term $-C_1 \left(\sqrt{\frac{\log n}{N}} + \bar{\sigma}_{de} \right) \left(1 - \frac{1}{\epsilon_1} \right) |\hat{\Delta}_{S^c}^-|_1$ is always positive as similarly set in [4] and also

$$|\hat{\Delta}_S^-|_1 \leq \sqrt{s} \|\hat{\Delta}_S^-\|_F \leq \sqrt{s} \|\hat{\Delta}^-\|_F \leq \sqrt{s+n} \|\hat{\Delta}^-\|_F.$$

This completes the proof.

5 Proof of Theorem 3

Proof: According to Theorem 1 and Theorem 2, we can find an appropriate constant M make the inequality $\|\Delta\|_F \leq MC + M\sqrt{n+s}\bar{\sigma}_{de}$ hold with high probability, and further obtain that

$$\begin{aligned}
\frac{1}{N} \sum_{i=1}^N \|\delta_i\|_2 &= \frac{1}{N} \sum_{i=1}^N \|\tilde{\mathbf{x}}_i - \hat{\mathbf{x}}_i\|_2 \\
&\leq \frac{1}{N} \sum_{i=1}^N \left\{ \frac{\left\| (\Omega_0 + \Phi_i' \mathbf{R}^{-1} \Phi_i)^{-1} \right\|_F \|\Delta\|_F}{1 - \left\| (\Omega_0 + \Phi_i' \mathbf{R}^{-1} \Phi_i)^{-1} \right\|_F \|\Delta\|_F} \|\hat{\mathbf{x}}_i - \boldsymbol{\mu}\|_2 \right\} \\
&\leq \frac{\max_i \left\| (\Omega_0 + \Phi_i' \mathbf{R}^{-1} \Phi_i)^{-1} \right\|_F \|\Delta\|_F}{1 - \max_i \left\| (\Omega_0 + \Phi_i' \mathbf{R}^{-1} \Phi_i)^{-1} \right\|_F \|\Delta\|_F} \left\{ \frac{1}{N} \sum_{i=1}^N \|\hat{\mathbf{x}}_i - \boldsymbol{\mu}\|_2 \right\} \\
&\leq \frac{\xi(MC + MD\bar{\sigma}_{de})}{N(1-\zeta)} \sum_{i=1}^N \|\hat{\mathbf{x}}_i - \boldsymbol{\mu}\|_2 \\
&\leq \frac{\xi \left(MC + \frac{MD}{N} \sum_{i=1}^N (2\hat{\mathbf{x}}_{\max} \|\delta_i\|_2 + \|\delta_i\|_2^2 + \text{tr}(\mathbf{C}_i)) \right)}{N(1-\zeta)} \sum_{i=1}^N \|\hat{\mathbf{x}}_i - \boldsymbol{\mu}\|_2 \\
&\leq \frac{\xi}{N(1-\zeta)} \left(MC + \frac{MD}{N} \sum_{i=1}^N (2\hat{\mathbf{x}}_{\max} \|\delta_i\|_2 + \|\delta_i\|_2^2 + \text{tr}(\mathbf{C}_i)) \right) \sum_{i=1}^N \|\hat{\mathbf{x}}_i - \boldsymbol{\mu}\|_2 \\
&\leq \frac{\xi}{N(1-\zeta)} \left(MC + \frac{MD}{N} \sum_{i=1}^N (2\hat{\mathbf{x}}_{\max} \|\delta_i\|_2 + \delta_{\max} \|\delta_i\|_2 + \text{tr}(\mathbf{C}_i)) \right) \sum_{i=1}^N \|\hat{\mathbf{x}}_i - \boldsymbol{\mu}\|_2.
\end{aligned} \tag{7}$$

where $C = \sqrt{(n+s) \log n/N}$, $D = \sqrt{n+s}$. The fourth inequality holds because

$$\begin{aligned}
\bar{\sigma}_{de} &= \|\Sigma_{de}\|_{\max} \leq \text{tr}(\Sigma_{de}) \\
&\leq \frac{2}{N} \sum_{i=1}^N \|\hat{\mathbf{x}}_i\|_2 \|\delta_i\|_2 + \text{tr} \left(\frac{1}{N} \sum_{i=1}^N (\delta_i \delta_i' + \mathbf{C}_i) \right) \\
&= \frac{2}{N} \sum_{i=1}^N \|\hat{\mathbf{x}}_i\|_2 \|\delta_i\|_2 + \frac{1}{N} \sum_{i=1}^N (\|\delta_i\|_2^2 + \text{tr}(\mathbf{C}_i)) \\
&\leq \frac{1}{N} \sum_{i=1}^N (2\hat{\mathbf{x}}_{\max} \|\delta_i\|_2 + \|\delta_i\|_2^2 + \text{tr}(\mathbf{C}_i))
\end{aligned} \tag{8}$$

where the second inequality holds based on Cauchy-Schwarz inequality.

Letting $\epsilon = \frac{1}{N} \sum_{i=1}^N \|\delta_i\|_2$, we write from (7) that

$$\frac{1-\zeta}{\xi \epsilon_0 M} \epsilon - C \leq D((\delta_{\max} + 2\hat{\mathbf{x}}_{\max})\epsilon + v).$$

When $\frac{1-\zeta}{\epsilon_0 MD} > \xi(\delta_{\max} + 2\hat{\mathbf{x}}_{\max})$, one has

$$\begin{aligned}
\epsilon &\leq \frac{C/D + v}{\frac{1-\zeta}{\xi \epsilon_0 MD} - (\delta_{\max} + 2\hat{\mathbf{x}}_{\max})}, \\
&= \frac{\sqrt{(\log n)/N} + v}{\frac{1-\zeta}{\xi \epsilon_0 M \sqrt{n+s}} - (\delta_{\max} + 2\hat{\mathbf{x}}_{\max})} \\
&= \frac{\sqrt{(\log n)/N} + v}{(1-\zeta)/\sqrt{n+s} - M\epsilon_0(\delta_{\max} + 2\hat{\mathbf{x}}_{\max})\xi} M\epsilon_0\xi.
\end{aligned}$$

This completes the proof.

References

- [1] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical Society: Series B*, 1977.
- [2] M. Chen, H. Wang, X. Liao, and L. Carin, “Bayesian learning of sparse Gaussian graphical models.” *submitted (available at: <http://people.ee.duke.edu/~lcarin/gLASSO5.pdf>)*.
- [3] C. Lam and J. Fan, “Sparsistency and rates of convergence in large covariance matrices estimation,” *The Annals of Statistics*, vol. 37, no. 6B, pp. pp. 4254–4278, 2009.
- [4] A. J. Rothman, P. J. Bickel, E. Levina, and J. Zhu, “Sparse permutation invariant covariance estimation,” *Electronic Journal of Statistics*, vol. 2, pp. 494–515, 2008.