Appendix to An Improved Iterative Proportional Scaling Procedure for Gaussian Graphical Models published in the Journal of Computational and Graphical Statistics

Ping-Feng Xu, Jianhua Guo and Xuming He

July 9, 2010

PROOF OF OUR MAIN RESULTS

Before proving Theorem 1 and the validity of the IIPS procedure (Algorithm 2), we first introduce some definitions and useful lemmas.

In a graph $G = (V, E)$, for disjoint subsets $A, B, C \subset V$, we say that C separates A and B in G, if every path in G between $u \in A$ and $v \in B$ contains a vertex in C. Let $\mathcal{C}$ be a collection of subsets of V satisfying that $C_1 \nsubseteq C_2$ for any $C_1, C_2 \in \mathcal{C}$. A tree $T = (\mathcal{C},\ E_T)$ is said to be a separation tree of G if for any edge $(K_1, K_2) \in E_T$, the separator $S = K_1 \cap K_2$ separates, in G, the vertex sets $V_1 \setminus S$ and $V_2 \setminus S$, where V_1 and V_2 are the union of the nodes in two subtrees T_1 and T_2 obtained by removing S. The definition of separation tree is presented by Liu, Guo, and Jing (2010). For a D-ordered sequence $K_1, K_2, \cdots, K_M$, let $S_m = K_m \cap (\cup_{i=1}^{m-1} K_i)$, $R_m = K_m \setminus S_m$, for $m = 2, \cdots, M$.

Lemma 1. *Let $T = (\mathcal{K}(G^t), E_T)$ be a junction tree of some triangulation G^t of a graph G, then T is also a separation tree of G.*

Proof: From the last paragraph in page 53 of Cowell et al. (1999), we have that any separator S in T separates $V_1 \setminus S$ and $V_2 \setminus S$ in G^t, where V_1 and V_2 are the union of the nodes in two subtrees T_1 and T_2 obtained by removing S. Since the edge set of G is a subset of that of G^t, S separates V_1 and V_2 in G too. Hence, T is a separation tree of G. $\square$

Lemma 2. *Let $T = (\mathcal{C}, E_T)$ be a separation tree of G. And $f(y|\mu, \Sigma)$ is a density function of $Y \sim N(\mu, \Sigma)$ in Gaussian graphical model $N(G)$, then we have that*

$$f(y|\mu, \Sigma) = \frac{\prod_{K \in \mathcal{C}} f_K(y_K|\mu_K, \Sigma_{KK})}{\prod_{S \in \mathcal{S}} f_S(y_S|\mu_S, \Sigma_{SS})},$$

Ping-Feng Xu is Ph.D. Student and Jianhua Guo is Professor (E-mail: jhguo@nenu.edu.cn), Key Laboratory for Applied Statistics of MOE and School of Mathematics and Statistics, Northeast Normal University, Changchun 130024, Jilin Province, China. Xuming He is Professor, Department of Statistics, University of Illinois at Urbana-Champaign, 725 South Wright Street, Champaign, IL, 61820, USA.

where $\mathcal{S} = \{K_i \bigcap K_j | (K_i, K_j) \in E_T\}$*,* $f_K(y_K|\mu_K, \Sigma_{KK})$ *and* $f_S(y_S|\mu_S, \Sigma_{SS})$ *are marginal density functions of* $Y_K \sim N(\mu_K, \Sigma_{KK})$ *and* $Y_S \sim N(\mu_S, \Sigma_{SS})$*, respectively.*

Proof: Induction on the number of nodes of T. If $|\mathcal{C}| = 1$, then $\mathcal{C} = \{V\}$ and $\mathcal{S} = \emptyset$. It is obvious that the proposition is true.

Suppose that the proposition is true for any separation tree with $|\mathcal{C}| < n$. If $|\mathcal{C}| = n$, let K_0 be a leaf node in T, which connects with only one node K_1 in T, S_0 be the corresponding separator of (K_0, K_1). Since T is a separation tree, then S_0 separates $(K_0 \setminus S_0)$ and $(K_1 \setminus S_0)$ in G. Furthermore, we have that

$$f_{K_0 \cup K_1}(y_{K_0 \cup K_1}|\mu_{K_0 \cup K_1}, \Sigma_{K_0 \cup K_1 K_0 \cup K_1}) = \frac{f_{K_0}(y_{K_0}|\mu_{K_0}, \Sigma_{K_0 K_0}) f_{K_1}(y_{K_1}|\mu_{K_1}, \Sigma_{K_1 K_1})}{f_{S_0}(y_{S_0}|\mu_{S_0}, \Sigma_{S_0 S_0})}. \quad (1)$$

Deleting S_0 from T and unioning K_0 and K_1 as a new node, we can get a new separation tree T' with node set $\mathcal{C}' = \mathcal{C} \setminus \{K_0, K_1\} \bigcup \{K_0 \bigcup K_1\}$ and separator set $\mathcal{S}' = \mathcal{S} \setminus \{S_0\}$, where $\mathcal{C}$ and $\mathcal{S}$ are the node set and separator set of T, respectively. By the induction assumption, we know that

$$\begin{aligned} f(y|\mu, \Sigma) &= \frac{\prod_{K \in \mathcal{C}'} f_K(y_K|\mu_K, \Sigma_{KK})}{\prod_{S \in \mathcal{S}'} f_S(y_S|\mu_S, \Sigma_{SS})} \\ &= \frac{\prod_{K \in \mathcal{C} \setminus \{K_0, K_1\}} f_K(y_K|\mu_K, \Sigma_{KK}) f_{K_0 \cup K_1}(y_{K_0 \cup K_1}|\mu_{K_0 \cup K_1}, \Sigma_{K_0 \cup K_1 K_0 \cup K_1})}{\prod_{S \in \mathcal{S} \setminus \{S_0\}} f_S(y_S|\mu_S, \Sigma_{SS})}. \quad (2) \end{aligned}$$

According to equations (1) and (2), we get that

$$f(y|\mu, \Sigma) = \frac{\prod_{K \in \mathcal{C}} f_K(y_K|\mu_K, \Sigma_{KK})}{\prod_{S \in \mathcal{S}} f_S(y_S|\mu_S, \Sigma_{SS})}.$$

□

By Lemmas 1 and 2, we get the following corollary.

Corollary 1. *Let* $T = (\mathcal{K}(G^t), E_T)$ *be a junction tree of some triangulation* G^t *of* G *with the corresponding D-ordered sequence* $K_1, K_2, \cdots, K_M$*, and suppose that* $f(y|\mu, \Sigma)$ *is a density function of* $Y \sim N(\mu, \Sigma)$ *in Gaussian graphical model* $N(G)$*, then we have that*

$$f(y|\mu, \Sigma) = f(y_{K_1}|\mu_{K_1}, \Sigma_{K_1 K_1}) \prod_{i=2}^{M} f_{R_i|S_i}(y_{R_i}|\mu_{R_i \cdot S_i}, \Sigma_{R_i R_i \cdot S_i}),$$

where $f_{R_i|S_i}(y_{R_i}|\mu_{R_i \cdot S_i}, \Sigma_{R_i R_i \cdot S_i})$ *is the conditional density, and* $\mu_{R_i \cdot S_i} = \mu_{R_i} + \Sigma_{R_i S_i} \Sigma_{S_i S_i}^{-1}(y_{S_i} - \mu_{S_i})$*,* $\Sigma_{R_i R_i \cdot S_i} = \Sigma_{R_i R_i} - \Sigma_{R_i S_i} \Sigma_{S_i S_i}^{-1} \Sigma_{S_i R_i}$ *for* $i = 2, \cdots, M$*.*

The correctness of Proposition 1 follows from Corollary 1. Now we prove Theorem 1.

Proof of Theorem 1: Suppose $f(y|\hat{\mu}, \Sigma)$ is a density function of $Y \sim N(\hat{\mu}, \Sigma)$, then by Corollary 1 we have that

$$f(y|\hat{\mu}, \Sigma) = f(y_{K_1}|\mu_{K_1}, \Sigma_{K_1 K_1}) \prod_{i=2}^{M} f_{R_i|S_i}(y_{R_i}|\mu_{R_i \cdot S_i}, \Sigma_{R_i R_i \cdot S_i}),$$

where $f_{R_i|S_i}(y_{R_i}|\mu_{R_i\cdot S_i}, \Sigma_{R_iR_i\cdot S_i})$ is the conditional density, and $\mu_{R_i\cdot S_i} = \hat{\mu}_{R_i} + \Sigma_{R_iS_i}\Sigma^{-1}_{S_iS_i}(y_{S_i} - \hat{\mu}_{S_i})$, $\Sigma_{R_iR_i\cdot S_i} = \Sigma_{R_iR_i} - \Sigma_{R_iS_i}\Sigma^{-1}_{S_iS_i}\Sigma_{S_iR_i}$ for $i = 2, \cdots, M$. So we get that

$$\begin{aligned}
f(y|\hat{\mu}, A_c\Sigma) &= f(y|\hat{\mu}, \Sigma)\frac{f(y_c|\hat{\mu}_c, S_{cc})}{f(y_c|\hat{\mu}_c, \Sigma_{cc})} \\
&= f(y_{K_1}|\mu_{K_1}, \Sigma_{K_1K_1})\prod_{i=2}^{M} f_{R_i|S_i}(y_{R_i}|\mu_{R_i\cdot S_i}, \Sigma_{R_iR_i\cdot S_i})\frac{f(y_c|\hat{\mu}_c, S_{cc})}{f(y_c|\hat{\mu}_c, \Sigma_{cc})} \\
&= f(y_{K_1}|\mu_{K_1}, \Sigma_{K_1K_1})\frac{f(y_c|\hat{\mu}_c, S_{cc})}{f(y_c|\hat{\mu}_c, \Sigma_{cc})}\prod_{i=2}^{M} f_{R_i|S_i}(y_{R_i}|\mu_{R_i\cdot S_i}, \Sigma_{R_iR_i\cdot S_i}) \\
&= f(y_{K_1}|\mu_{K_1}, A_c\Sigma_{K_1K_1})\prod_{i=2}^{M} f_{R_i|S_i}(y_{R_i}|\mu_{R_i\cdot S_i}, \Sigma_{R_iR_i\cdot S_i}). \qquad (3)
\end{aligned}$$

Since we update ϕ_{K_1} by $[-, A_c\Sigma_{K_1K_1}](K_1|-)$ and the combination operation of potentials corresponds to ordinary composition of conditional and marginal distributions, thus we get that $[-, A_c\Sigma](V|-) = \otimes_{K\in\mathcal{K}(G^t)}\phi_K$.

Furthermore, for the leaf node K_M without any child in T,

$$\begin{aligned}
& f(y_{V\setminus R_M}|\mu_{V\setminus R_M}, (A_c\Sigma)_{V\setminus R_M V\setminus R_M}) \\
= & \int f(y_{K_1}|\mu_{K_1}, A_c\Sigma_{K_1K_1})\prod_{i=2}^{M} f_{R_i|S_i}(y_{R_i}|\mu_{R_i\cdot S_i}, \Sigma_{R_iR_i\cdot S_i})\,\mathrm{d}y_{R_M} \\
= & f(y_{K_1}|\mu_{K_1}, A_c\Sigma_{K_1K_1})\prod_{i=2}^{M-1} f_{R_i|S_i}(y_{R_i}|\mu_{R_i\cdot S_i}, \Sigma_{R_iR_i\cdot S_i}). \qquad (4)
\end{aligned}$$

By equations (3) and (4), we get that

$$f_{R_M|S_M}(y_{R_M}|\mu_{R_M\cdot S_M}, (A_c\Sigma)_{R_MR_M\cdot S_M}) = f_{R_M|S_M}(y_{R_M}|\mu_{R_M\cdot S_M}, \Sigma_{R_MR_M\cdot S_M}),$$

where $(A_c\Sigma)_{R_MR_M\cdot S_M} = (A_c\Sigma)_{R_MR_M} - (A_c\Sigma)_{R_MS_M}(A_c\Sigma)^{-1}_{S_MS_M}(A_c\Sigma)_{S_MR_M}$. By induction, we similarly have that

$$f_{R_i|S_i}(y_{R_i}|\mu_{R_i\cdot S_i}, (A_c\Sigma)_{R_iR_i\cdot S_i}) = f_{R_i|S_i}(y_{R_i}|\mu_{R_i\cdot S_i}, \Sigma_{R_iR_i\cdot S_i}),$$

for $i = 2, \cdots, M$. Therefore, (ii) is correct. $\square$

Finally, we note that once we call DFA(K_1), each clique marginal is adjusted one time. Therefore we conclude by Proposition 2 and Theorem 1 that the proposed IIPS procedure (Algorithm 2) is valid.

References

Cowell, R.G., David, A.P., Lauritzen, S.L., and Spiegelhalter, D.J. (1999), *Probabilistic Networks and Expert Systems*. New York: Springer.

Liu, B.H., Guo, J.H., and Jing, B.Y. (2010), "A note on minimal d-separation trees for structural learning," *Artificial Intelligence*, 174, 442–448.