MATHEMATICAL ENGINEERING TECHNICAL REPORTS

A Localization Approach to Improve Iterative Proportional Scaling in Gaussian Graphical Models

Hisayuki HARA and Akimichi TAKEMURA

A Localization Approach to Improve Iterative Proportional Scaling in Gaussian Graphical Models

Hisayuki Hara

Department of Geosystem Engineering

University of Tokyo

Akimichi Takemura

Graduate School of Information Science and Technology

University of Tokyo

February 2008

Abstract

We discuss an efficient implementation of the iterative proportional scaling procedure in the multivariate Gaussian graphical models. We show that the computational cost can be reduced by localization of the update procedure in each iterative step by using the structure of a decomposable model obtained by triangulation of the graph associated with the model. Some numerical experiments demonstrate the competitive performance of the proposed algorithm.

1   Introduction

Since Dempster [6] introduced a multivariate Gaussian graphical model, also called a covariance selection model, it has been investigated by many authors from both theoretical and practical viewpoints. On the theory of a Gaussian graphical model, see e.g. Whittaker [20], Lauritzen [15], Cox and Wermuth [2] and Edwards [10]. In recent years much effort has been devoted to application of the Gaussian graphical model to identify sparse large network systems, especially genetic networks (e.g. [8], [17], [9]), and the efficient implementation of the inference in the model has been extensively studied. In this article we discuss an efficient algorithm to compute the maximum likelihood estimator (MLE) of the covariance matrix in the Gaussian graphical models.

When the graph associated with the model is a chordal graph, the model is called a decomposable model. For a decomposable model, the MLE of the covariance matrix is explicitly obtained. For general graphical models other than decomposable models, however, we need some iterative procedure to obtain the MLE. The iterative proportional scaling (IPS) procedure is one of popular algorithms to compute the MLE.

1

The IPS was first introduced by Deming and Stephan [5] to estimate cell probabilities in contingency tables subject to certain fixed marginals. Its convergence and statistical properties have been well studied by many authors (e.g. [12], [11]) and the IPS have been justified in a more general framework ([3]). Speed and Kiiveri [19] first formulated the IPS in a Gaussian graphical model and gave a proof of its convergence.

However, from a practically point of view, a straightforward application of the IPS is often computationally too expensive for larger models. In the contingency tables several techniques have been developed to reduce both storage and computation time of the IPS (e.g. [13], [14]). Badsberg and Malvestuto [1] proposed an efficient implementation of the IPS by using the structure of decomposable models containing the graphical model.

In the present paper we describe a technique for improving the efficiency of the IPS in the Gaussian graphical models. Let Δ be the set of variables which corresponds to the set of vertices of the graph associated with the model. The straightforward implementation of the IPS requires approximately $O(|\Delta|^3)$ time in each iterative step for large models. In the similar way to the technique in Badsberg and Malvestuto [1], we consider to localize the update procedure in each step by using the structure of a decomposable model containing the model. The proposed algorithm is shown to require $O(|\Delta|)$ time for some models.

As pointed out in Dahl et al. [4], the implementation of the IPS requires enumeration of all maximal cliques of the graph and this enumeration has an exponential complexity. Hence the application of the IPS to large models may be limited. However in the case where the model is relatively small or the structure of the model is simple, it may be feasible to enumerate maximal cliques. In this article we consider such situations.

The organization of this paper is as follows. In Section 2 we summarize notations and basic facts on graphs and give a brief review of Gaussian graphical models and the IPS algorithm for covariance matrices. In Section 3 we propose an efficient implementation of the update procedure of the IPS. In Section 4 we perform some numerical experiments to illustrate the effectiveness of the proposed procedure. We end this paper with some concluding remarks in Section 5.

2 Background and preliminaries

2.1 Preliminaries on decompositions of graphs

In this section we summarize some preliminary facts on decompositions of graphs needed in the argument of the following sections according to Leimer [16], Lauritzen [15] and Malvestuto and Moscarini [18].

Let $\mathcal{G} = (\Delta, E)$ be an undirected graph, where Δ denotes the set of vertices and E denotes the set of edges. A subset of Δ which induces a complete subgraph is called a clique of $\mathcal{G}$. Denote by $\mathcal{C}$ the set of maximal cliques of $\mathcal{G}$. For a subset of vertices V, let $\mathcal{G}(V)$ denote the subgraph of $\mathcal{G}$ induced by V. When a graph $\mathcal{G}$ is not connected, we can consider each connected component of $\mathcal{G}$ separately. Therefore we only consider a connected graph from now on.

A subset $S \subset \Delta$ is said to be a separator of $\mathcal{G}$ if $\mathcal{G}(\Delta \setminus S)$ is disconnected. For a separator S, a triple (A, B, S) of disjoint subsets of Δ such that $A \cup B \cup S = \Delta$ is said to form a decomposition of $\mathcal{G}$. A separator S is called a clique separator if S is a clique of $\mathcal{G}$. For two non-adjacent vertices δ and δ', $S \subset \Delta$ is said to be a (δ, δ')-separator if $\delta \in A$ and $\delta' \in B$ for a decomposition (A, B, S). A (δ, δ')-separator which is minimal with respect to inclusion relation is called a minimal (δ, δ')-separator or a minimal vertex separator. Denote by $\mathcal{S}$ the set of minimal vertex separators for all non-adjacent pairs of vertices in $\mathcal{G}$.

A graph $\mathcal{G}$ is called reducible if Δ contains a clique separator and otherwise $\mathcal{G}$ is said to be prime. If $\mathcal{G}(V)$ is prime and $\mathcal{G}(V')$ is reducible for all V' with $V \subsetneq V' \subset \Delta$, $\mathcal{G}(V)$ is called a maximal prime subgraph (mp-subgraph) of $\mathcal{G}$. For any reducible graph, its decomposition into mp-subgraphs is uniquely defined. Denote by $\mathcal{V}$ the set of subsets of Δ which induces mp-subgraphs of $\mathcal{G}$ and let $|\mathcal{V}| = M$. Then there exists a sequence $V_1, \ldots, V_M \in \mathcal{V}$ such that for every $m = 2, \ldots, M$ there exists $m' < m$ with

$$V_{m'} \supset V_m \cap (V_1 \cap \cdots \cap V_{m-1}).$$

Such a sequence is called a D-ordered sequence. Let $S_m := V_m \cap (V_1 \cap \cdots \cap V_{m-1})$ for $m = 2, \ldots, M$. Define $\bar{\mathcal{S}} = \{S_2, \ldots, S_m\}$. Denote by $\mathcal{S}_C$ the set of clique separators of $\mathcal{G}$. Then $\bar{\mathcal{S}}$ satisfy $\bar{\mathcal{S}} = \mathcal{S} \cap \mathcal{S}_C$. So we call elements of $\bar{\mathcal{S}}$ clique minimal vertex separators. Reducible graphs always have a D-ordered sequence with $V_1 = V$ for any $V \in \mathcal{V}$. Hence a D-ordered sequence is not uniquely defined. However $\bar{\mathcal{S}}$ is common for all D-ordered sequences.

Example 1 (A reducible graph). *The graph $\mathcal{G}$ in Figure 1 is an example of reducible graphs. $\mathcal{G}$ has two clique minimal vertex separators $S_2 := \{3, 4\}$ and $S_3 := \{5, 6\}$. Define V_1, V_2 and V_3 by*

$$V_1 := \{1, 2, 3, 4\}, \quad V_2 := \{3, 4, 5, 6\}, \quad V_3 := \{5, 6, 7, 8\}$$

as in Figure 1. Then $\mathcal{V} = \{V_1, V_2, V_3\}$ and the sequence V_1, V_2, V_3 is a D-ordered sequence.
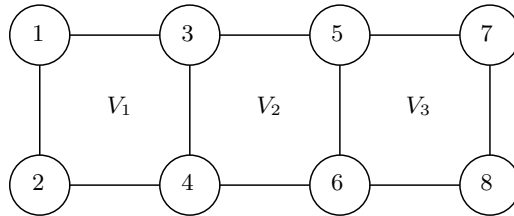

Figure 1: A reducible graph with eight vertices

When $\mathcal{G}$ is a chordal graph, $\mathcal{V}$ is the set of maximal cliques $\mathcal{C}$ and $\mathcal{S} = \bar{\mathcal{S}}$. Hence $|\mathcal{C}| = M$. A D-ordered sequence for a chordal graph is called a perfect sequence of

maximal cliques. There exists a perfect sequence of maximal cliques $C_1, \ldots, C_M$ such that $C_1 = C$ for any $C \in \mathcal{C}$.

For a vertex $\delta \in \Delta$, let adj(δ) denote the set of vertices adjacent to δ. When adj(δ) is a clique, δ is called a simplicial vertex. A simplicial vertex is contained in only one maximal clique C. Hence if δ is simplicial and $\delta \in C$, then adj(δ) $= C \setminus \{\delta\}$. A sequence of vertices $\delta_1, \delta_2, \ldots, \delta_{|\Delta|}$ is called a perfect elimination order of vertices of $\mathcal{G}$ if δ_i is a simplicial vertex in $\mathcal{G}(\bigcup_{j=i}^{|\Delta|}\{\delta_j\})$. It is well known that $\mathcal{G}$ is a chordal graph if and only if $\mathcal{G}$ possesses a perfect elimination order (Dirac[7]). Let $C_1, \ldots, C_M$ be a perfect sequence of maximal cliques of a chordal graph $\mathcal{G}$. Define $R_1 := C_1 \setminus S_2$, $S_m := C_m \cap (C_1 \cap \cdots \cap C_{m-1})$ and $R_m := C_m \setminus S_m$ for $m = 2, \ldots, M$. Let $r_m := |R_m|$. Let $\delta_1^m, \ldots, \delta_{r_m}^m$ be any sequence of vertices in R_m. Then the sequence of vertices

$$\delta_1^M, \ldots, \delta_{r_M}^M, \delta_1^{M-1}, \ldots, \delta_{r_{M-1}}^{M-1}, \ldots, \delta_1^1, \ldots, \delta_{r_1}^1$$

is a perfect elimination order of $\mathcal{G}$. We call it a perfect elimination order induced by the perfect sequence $C_1, \ldots, C_M$.

We introduce some notations and a basic formula for matrices needed in the following sections. Let $A = \{a_{ij}\}$ be a $|\Delta| \times |\Delta|$ matrix. For two subsets Δ_1 and Δ_2 of Δ, we let

$$A_{\Delta_1 \Delta_2} = \{a_{ij}\}_{i \in \Delta_1, j \in \Delta_2}$$

denote a $|\Delta_1| \times |\Delta_2|$ submatrix of A. Define

$$A_{\Delta_1 \Delta_2}^{-1} := (A^{-1})_{\Delta_1 \Delta_2}.$$

We let $[A_{\Delta_1 \Delta_2}]^\Delta$ denote the $|\Delta| \times |\Delta|$ matrix such that

$$([A_{\Delta_1 \Delta_2}]^\Delta)_{ij} = \begin{cases} a_{ij} & \text{if } i \in \Delta_1, j \in \Delta_2 \\ 0 & \text{otherwise .} \end{cases}$$

Let $\Delta_2 = \Delta_1^C$ and decompose a symmetric matrix A into blocks as

$$A = \begin{pmatrix} A_{\Delta_1 \Delta_1} & A_{\Delta_1 \Delta_2} \\ A'_{\Delta_1 \Delta_2} & A_{\Delta_2 \Delta_2} \end{pmatrix}.$$

Here for notational simplicity we displayed A for the case that the elements of Δ_1 are smaller than those of Δ_2. Suppose that $A_{\Delta_2 \Delta_2}$ and $A_{\Delta_1 \Delta_1} - A_{\Delta_1 \Delta_2}(A_{\Delta_2 \Delta_2})^{-1}A'_{\Delta_1 \Delta_2}$ are both positive definite. Then A is positive definite and

$$A_{\Delta_1 \Delta_1}^{-1} = \left(A_{\Delta_1 \Delta_1} - A_{\Delta_1 \Delta_2}(A_{\Delta_2 \Delta_2})^{-1}A'_{\Delta_1 \Delta_2}\right)^{-1}. \tag{1}$$

2.2 Gaussian graphical models

Let $\mathcal{M}^+(\mathcal{G})$ denote the set of $|\Delta| \times |\Delta|$ positive definite matrices $K = \{k_{ij}\}$ such that $k_{ij} = 0$ for all $i, j \in \Delta$ with $(i, j) \notin E$. Then the Gaussian graphical model for $|\Delta|$

dimensional random variable $Y = (Y^{(1)}, \ldots, Y^{(|\Delta|)})'$ associated with a graph $\mathcal{G}$ is defined as

$$Y \sim N_{|\Delta|}(\mu, \Sigma), \quad K := \Sigma^{-1} \in \mathcal{M}^+(\mathcal{G}).$$

$k_{ij} = 0$ indicates the conditional independence between $Y^{(i)}$ and $Y^{(j)}$ given all other variables. In what follows, we identify $\mathcal{M}^+(\mathcal{G})$ with the corresponding graphical model. Let $y_1, \ldots, y_n$ be i.i.d. samples from $\mathcal{M}^+(\mathcal{G})$. Define $\bar{y}$ and W by

$$\bar{y} := n^{-1} \sum_{i=1}^{n} y_i, \quad W := \sum_{i=1}^{n} (y_i - \bar{y})(y_i - \bar{y})',$$

respectively. The likelihood equation is written as

$$L(\mu, K) \propto (\det K)^{n/2} \exp \left\{ -\frac{1}{2} \mathrm{tr} KW - \frac{n}{2} \mathrm{tr} K(\bar{y} - \mu)(\bar{y} - \mu)' \right\}.$$

The MLE of μ is $\bar{y}$. The likelihood equations involving K are expressed as

$$nK_{CC}^{-1} = n\Sigma_{CC} = W_{CC}, \quad \forall C \in \mathcal{C}. \tag{2}$$

For a subset of vertices $V \subset \Delta$, let $\hat{K}_{VV}$ denote the MLE of K in the marginal model associated with the graph $\mathcal{G}(V)$ based on the data in V-marginal sample only. Let S be a clique separator of $\mathcal{G}$ and (A, B, S) be a decomposition of $\mathcal{G}$. Let $V = A \cup S$ and $V' = B \cup S$. Then the MLE $\hat{K}$ is known to satisfy

$$\hat{K} = \left[\hat{K}_{VV} \right]^{\Delta} + \left[\hat{K}_{V'V'} \right]^{\Delta} - n \left[(W_{SS})^{-1} \right]^{\Delta} \tag{3}$$

(e.g. Lauritzen [15]). More generally, for the set of mp-subgraphs $\mathcal{V}$ and the set of clique minimal vertex separators $\bar{\mathcal{S}}$,

$$\hat{K} = \sum_{V \in \mathcal{V}} \left[\hat{K}_{VV} \right]^{\Delta} - n \sum_{S \in \bar{\mathcal{S}}} \left[(W_{SS})^{-1} \right]^{\Delta}. \tag{4}$$

As mentioned in the previous section, when the model is decomposable, $\mathcal{V} = \mathcal{C}$ and $\mathcal{S} = \bar{\mathcal{S}}$. Hence from (2), $\hat{K}$ is explicitly written by

$$\hat{K} = n \sum_{C \in \mathcal{C}} \left[(W_{CC})^{-1} \right]^{\Delta} - n \sum_{S \in \mathcal{S}} \left[(W_{SS})^{-1} \right]^{\Delta}.$$

However for other graphical models, we need some iterative procedure for computing the first term on the right-hand side of (4). The following IPS is commonly used for this purpose. Note that the second term on the right-hand side of (4) needs to be calculated only once and is not involved in the iterative procedure. IPS consists of iteratively and successively adjusting Σ_{CC} for $C \in \mathcal{C}$ as in (2). Let K^t and $\Sigma^t = (K^t)^{-1}$ denote the estimated K and Σ at the t-th step of iteration, respectively. Define $D := \Delta \setminus C$ for

$C \in \mathcal{C}$. Then the t-th iterative step of the IPS is described by the update rule of K^t as follows,

$$(K^t)_{CC} := (W_{CC})^{-1} + (K^{t-1})_{CD}((K^{t-1})_{DD})^{-1}(K^{t-1})_{DC} \tag{5}$$
$$(K^t)_{CD} := (K^{t-1})_{CD}$$
$$(K^t)_{DD} := (K^{t-1})_{DD}.$$

From (1), it is easy to see that

$$(K^t)_{CC}^{-1} = (\Sigma^t)_{CC} = W_{CC}/n.$$

We also see that if the initial estimate K_0 satisfies $K_0 \in \mathcal{M}^+(\mathcal{G})$, then $K^t \in \mathcal{M}^+(\mathcal{G})$ for all t. By using the argument of Csiszár [3], the convergence of the algorithm to the MLE

$$\lim_{n \to \infty} K^t = \hat{K}, \quad \lim_{n \to \infty} \Sigma^t = \hat{\Sigma}$$

is guaranteed (Speed and Kiiveri [19] and Lauritzen [15]).

The fact (3) suggests that the decomposition (A, B, S) for a clique separator S can localize the problem, that is, in order to obtain the MLE $\hat{K}$, it suffices to compute the MLE of submatrix $\hat{K}_{VV}$ and $\hat{K}_{V'V'}$, where $V = A \cup S$ and $V' = B \cup S$. Especially if the decomposition by mp-subgraphs is obtained, we need only to compute $\hat{K}_{VV}$ for each $V \in \mathcal{V}$.

From a complexity theoretic point of view, the t-th iterative step (5) requires $O(|D|^3 + |D|^2|C| + |D||C|^2)$ time. The graphical model with

$$\mathcal{C} = \{\{1, 2\}, \{2, 3\}, \ldots, \{|\Delta| - 1, |\Delta|\}, \{|\Delta|, 1\}\}$$

is called the $|\Delta|$-dimensional cycle model or $|\Delta|$ cycle model. Note that the cycle is prime. In the case of $|\Delta|$ cycle model, $|C| = 2$ and $|D| = |\Delta| - 2$. Hence when $|\Delta| \geq 4$, the iterative step (5) requires $O((|\Delta| - 2)^3)$ time. In the next section we propose a more efficient algorithm for computing (5) by using the structure of a triangulation of a graph.

3   A localized algorithm of IPS

From (1), we note that (5) is rewritten as

$$(K^t)_{CC} = (W_{CC})^{-1} + (K^{t-1})_{CC} - ((\Sigma^{t-1})_{CC})^{-1}$$
$$= (W_{CC})^{-1} + (K^{t-1})_{CC} - ((K^{t-1})_{CC}^{-1})^{-1}. \tag{6}$$

In this section we provide an efficient algorithm to compute $((K^{t-1})_{CC}^{-1})^{-1}$ by using the structure of $\mathcal{G}$. For a graph $\mathcal{G}$, let $\mathcal{G}^*$ be a chordal graph obtained by triangulating $\mathcal{G}$. Figure 2 represents an example of the five cycle model and its triangulation.

Let $C_1^*, \ldots, C_M^*$ be a perfect sequence of the maximal cliques of $\mathcal{G}^*$ with $C_1^* \supset C$. Let $S_m^* := C_m^* \cap (C_1^* \cap \cdots \cap C_{m-1}^*)$ for $m = 2, \ldots, M$ be minimal vertex separators of $\mathcal{G}^*$. We propose the following algorithm to compute $((K^{t-1})_{CC}^{-1})^{-1}$ for each maximal clique $C \in \mathcal{C}$.
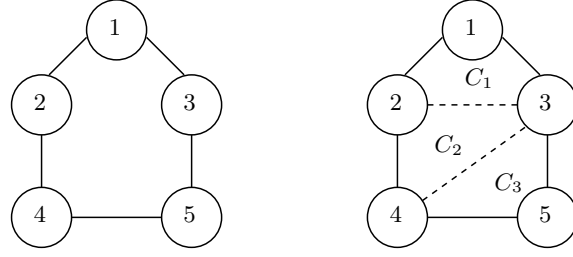
(i) the five cycle model   (ii) a triangulation of (i)

Figure 2: The five cycle model and its triangulation

Algorithm 1 (Computing $((K^{t-1})^{-1}_{CC})^{-1}$).

Step 0 $m \leftarrow M$ and $K^* \leftarrow K^{t-1}$.

Step 1 If $m \neq 1$, select a simplicial vertex $\delta \in C^*_m$ of $\mathcal{G}^*$.
If $m = 1$, select a vertex $\delta \notin C$.
Let $Q = C^*_m \setminus \{\delta\}$.

Step 2 Update K^*_{QQ} by

$$K^*_{QQ} \leftarrow K^*_{QQ} - (k^*_{\delta\delta})^{-1} K^*_{Q\delta} K^*_{\delta Q}. \tag{7}$$

Step 3 Update C^*_m, $\mathcal{G}^*$ and Δ as follows,

$$C^*_m \leftarrow Q, \quad \mathcal{G}^* \leftarrow \mathcal{G}^*(\Delta \setminus \{\delta\}), \quad \Delta \leftarrow \Delta \setminus \{\delta\}.$$

If $C^*_m = S^*_m$, $m \leftarrow m - 1$.
If $C^*_m = C$, return K^*_{CC}. Otherwise, go to Step 1.

Now we state the main theorem of this paper.

Theorem 1. *The output K^*_{CC} of Algorithm 1 is equal to $((K^{t-1})^{-1}_{CC})^{-1}$.*

Proof. Let $\delta \in C^*_M$ be a simplicial vertex in $\mathcal{G}^*$. Define $Q := C^*_M \setminus \{\delta\}$, $Q_1 := \Delta \setminus \{\delta\}$ and $Q_2 := \Delta \setminus C^*_M$. Since $\mathrm{adj}(\delta) \subset C^*_M$ and $K^{(t-1)} \in \mathcal{M}^+(\mathcal{G})$, $(K^{t-1})_{Q_2\delta} = \mathbf{0}$. Noting that $Q \cup Q_2 = Q_1$, we have from (1)

$$((K^{t-1})^{-1}_{Q_1 Q_1})^{-1} = (K^{t-1})_{Q_1 Q_1} - (k^{t-1})^{-1}_{\delta\delta}(K^{t-1})_{Q_1\delta}(K^{t-1})_{\delta Q_1}$$

$$= (K^{t-1})_{Q_1 Q_1} - (k^{t-1})^{-1}_{\delta\delta} \begin{pmatrix} \mathbf{0} \\ (K^{t-1})_{Q\delta} \end{pmatrix} \begin{pmatrix} \mathbf{0} & (K^{t-1})_{\delta Q} \end{pmatrix}$$

$$= (K^{t-1})_{Q_1 Q_1} - \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (k^{t-1})^{-1}_{\delta\delta}(K^{t-1})_{Q\delta}(K^{t-1})_{\delta Q} \end{pmatrix}$$

and $((K^{t-1})^{-1}_{Q_1 Q_1})^{-1} \in \mathcal{M}^+(\mathcal{G}(Q_1))$, where $(k^{t-1})_{\delta\delta}$ is the (δ, δ)-th element of K^{t-1}. By iterating the procedure in accordance with the perfect elimination order induced by the perfect sequence $C^*_1, \ldots, C^*_M$, we complete the proof. $\square$

7

In Algorithm 1, the triangulation $\mathcal{G}^*$ is arbitrary. However for every iterative step of adjusting the C-marginal, we have to use the perfect sequence with $C_1^* \supset C$.

Example 2 (the five cycle model)**.** *Consider the five cycle model in Figure 2-(i). K is expressed by*

$$ K = \begin{pmatrix} k_{11} & k_{12} & k_{13} & 0 & 0 \\ k_{12} & k_{22} & 0 & k_{24} & 0 \\ k_{13} & 0 & k_{33} & 0 & k_{35} \\ 0 & k_{24} & 0 & k_{44} & k_{45} \\ 0 & 0 & k_{35} & k_{45} & k_{55} \end{pmatrix}. $$

By adding the fill-in edges $\{2,3\}$ and $\{3,4\}$, a triangulated graph $\mathcal{G}^*$ can be obtained as in Figure 2-(ii). Consider the case where $C = \{1,2\}$. Define $C_1^* = \{1,2,3\}$, $C_2^* = \{2,3,4\}$ and $C_3^* = \{3,4,5\}$. Then the sequence C_1^*, C_2^*, C_3^* is perfect and it induces a perfect elimination order $5, 4, 3, 2, 1$. The update of K^* in step 2 in accordance with the perfect elimination order is described as follows,

$$ K_{34,34}^* \leftarrow K_{34,34}^* - (k_{55}^*)^{-1} \begin{pmatrix} k_{35}^* \\ k_{45}^* \end{pmatrix} (k_{35}^* \ k_{45}^*), $$

$$ K_{23,23}^* \leftarrow K_{23,23}^* - (k_{44}^*)^{-1} \begin{pmatrix} k_{24}^* \\ k_{34}^* \end{pmatrix} (k_{24}^* \ k_{34}^*), $$

$$ K_{12,12}^* \leftarrow K_{12,12}^* - (k_{33}^*)^{-1} \begin{pmatrix} k_{13}^* \\ k_{23}^* \end{pmatrix} (k_{13}^* \ k_{23}^*). $$

Then $K_{12,12}^* = ((K^{t-1})_{12,12}^{-1})^{-1} = ((K^{t-1})_{CC}^{-1})^{-1}$. $\qquad\square$

We now analyze the computational cost of the proposed algorithm. In Step 2, the running time of the calculation of (7) is as follows,

- $K_1^* := (k_{\delta\delta}^*)^{-1} K_{Q\delta}^*$ requires $|Q|$ divisions ;

- $K_2^* := K_1^* K_{\delta Q}^*$ requires $|Q|^2$ multiplications ;

- $K_{QQ}^* - K_2^*$ requires $|Q|^2$ subtractions.

Define $R_1^* := C_1^* \setminus C$ and $R_m^* := C_m^* \setminus S_m^*$ for $m = 2, \ldots, M$. $|Q|$ ranges over $\{|C_m| - j \mid 1 \le j \le R_m^*, \ 1 \le m \le M\}$. Let μ, γ and σ measure the time units required by a single multiplication, division and subtraction, respectively. Then the running time of Algorithm 1 amounts to

$$ (\mu + \sigma) \sum_{m=1}^{M} \sum_{j=1}^{R_m^*} (|C_m^*| - j)^2 + \delta \sum_{m=1}^{M} \sum_{j=1}^{R_m^*} (|C_m^*| - j) $$

$$ = (\mu + \sigma) \sum_{m=1}^{M} \Big\{ |R_m^*||C_m^*|^2 - |R_m^*||C_m^*| - |R_m^*|^2|C_m^*| $$

$$+ \frac{|R_m^*|(|R_m^*| + 1)(2|R_m^*| + 1)}{6} \Big\}$$
$$+ \delta \sum_{m=1}^{M} \left\{ |R_m^*||C_m^*| - \frac{(1 + |R_m^*|)|R_m^*|}{2} \right\} + 2\sigma|C|^2.$$

Since $|C_m^*| \geq |R_m^*|$, the computational cost of Algorithm 1 is $O\left(\sum_{m=1}^{M} |R_m^*||C_m^*|^2\right)$. Once $((K^{t-1})_{CC}^{-1})^{-1}$ is obtained, $O(|C|^2)$ additions are required to compute (6). Note that we can compute $(W_{CC})^{-1}$ once before the IPS procedure. Hence the computational cost of the t-th iterative step amounts to $O\left(|C|^2 + \sum_{m=1}^{M} |R_m^*||C_m^*|^2\right)$. In the case of cycle models, $|C| = 2$, $M = |\Delta| - 2$, $|C_m^*| = 3$ and $|R_m^*| = 1$. Thus the computational cost is $O(|\Delta|)$. As mentioned in the previous section, the direct computation of $((K^{t-1})_{CC}^{-1})^{-1}$ requires $O((\Delta \setminus C)^3) = O(|D|^3)$ time and in the case of cycle models it requires $O((|\Delta| - 2)^3)$ time. Hence we can see the efficiency of the proposed algorithm.

4 Numerical experiments for cycle models

In this section we compare the localized IPS proposed in the previous section with the direct computation of the IPS by numerical experiments. We consider the $|\Delta|$ cycle models with $|\Delta| = 4, 5, \ldots, 10$. We set $K = I_{|\Delta|}$. We generate 10000 Wishart matrices W with the parameter $I_{|\Delta|}$ and the degrees of freedom 20 and computed the MLE $\hat{K}$ for $|\Delta|$ cycle models by using the proposed algorithm and the direct computation of the IPS. We set the initial estimate $K^0 := I_{|\Delta|}$. As a convergence criterion, we used $\sum_{i,j} |k_{ij}^t| \leq 10^{-6}$. Table 1 presents the average CPU time per one iterative step to update K^{t-1} in (6) for both algorithms. The computation was done on a Pentium 3.2 GHz CPU machine by using R language. We can see the competitive performance of the proposed algorithm.

As seen from Table 1, while the CPU time of the proposed algorithm linearly increases in proportion to $|\Delta|$, that of the direct computation is almost constant, despite a slightly increase. In the update procedure (6), the computation of $((K^{t-1})_{DD})^{-1}$ requires $O(|D|^3)$ time and it is the most computationally expensive. Table 2 presents the average CPU

Table 1: CPU time per one iterative procedure for $|\Delta|$ cycle models

$|\Delta|$	Algorithm 1	direct computation
4	1.775	6.918
5	2.400	6.960
6	3.013	7.138
7	3.567	7.218
8	4.208	7.331
9	4.769	7.375
10	5.454	7.425

$(10^{-4}$ CPU time)

time for computing a $|\Delta| \times |\Delta|$ inverse matrix by using R language on the same machine. As seen from the table, while the CPU time for computing the inverse of a matrix is almost constant for small $|\Delta|$, it increases at the rate $O(|\Delta|^3)$ when $|\Delta| \geq 100$. Therefore the proposed algorithm is expected to be efficient especially for large models. We note that the results of Table 1 suggests that the proposed algorithm is faster than the direct computation of the IPS even for small models.

Table 2: CPU time for calculating $|\Delta| \times |\Delta|$ inverse matrices

$|\Delta|$	CPU time
5	0.09
10	0.09
50	0.13
100	0.56
500	43.25
1000	396.82

$(10^{-2}$ CPU time)

5   Concluding remarks

In this article we discussed the localization of the IPS in two ways. We first showed that the decomposition of the graph by clique separators can localize the IPS. Next we proposed a localized algorithm of the iterative step in the IPS by using the structure of the decomposable model containing the graphical model. We showed that the proposed algorithm is faster than the direct computation of the IPS by numerical experiments.

As mentioned in Section 1, the implementation of the IPS requires enumeration of all maximal cliques of the graph and this enumeration has an exponential complexity. In addition, the proposed algorithm also requires some characteristics of graphs, that is, a triangulated graph, perfect sequences and perfect elimination orders of the triangulated graph. In this sense, the application of the IPS may be limited. However in the case where the model is relatively small or the structure of the model is simple, it may be feasible to obtain characteristics of graphs. In such cases, the proposed algorithm is effective.

References

[1] J. H. Badsberg and F. M. Malvestuto. An implementation of the iterative proportional fitting procecure by propagation trees. *Comput. Statist. Data Anal.*, 37:297–322, 2001.

[2] D. R. Cox and N. Wermuth. *Multivariate Dependencies*. Chapman and Hall, London, 1996.

[3] I. Csiszár. *I*-divergence geometry of probability distributions and minimization problems. *Ann. Probab.*, 3:146–158, 1975.

[4] J. Dahl, L. Vandenberghe, and V. Roychowdhury. Covariance selection for non-chordal graphs via chordal embedding, 2006. To appear in *Optimization Methods and Software*.

[5] W. E. Deming and F. F. Stephan. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Statist*, 11:427–444, 1940.

[6] A. P. Dempster. Covariance selection. *Biometrics*, 28:157–175, 1972.

[7] G. A. Dirac. On rigid circuit graphs. *Abh. Math. Sem. Univ. Hamburg*, 25:71–76, 1961.

[8] A. Dobra, C. Hans, B. Jones, J. R. Nevins, G. Yao, and M. West. Sparse graphical models for exploring gene expression data. *J. Multivariate Anal.*, 90:196–212, 2004.

[9] M. Drton and T. S. Richardson. Graphical methods for efficient likelihood inference in Gaussian covariance models. arXiv:0708.1321, 2007.

[10] D. M. Edwards. *Introduction to Graphical Modelling*. Springer, New York, 2000.

[11] S. E. Fienberg. An iterative procedure for estimation in contingency tables. *Ann. Math. Statist.*, 41:907–917, 1970.

[12] C. T. Ireland and S. Kulback. Contingency table with given marginal. *Biometrika*, 55:179–188, 1968.

[13] R. Jiroušek. Solution of the marginal problem and decomposable solutions. *Kybernetika*, 27:403–412, 1991.

[14] R. Jiroušek and S. Přeučil. On the effective implementation of the iterative proportional fitting procedure. *Comput. Statist. Data. Anal.*, 19:177–189, 1995.

[15] Steffen L. Lauritzen. *Graphical Models*. Oxford University Press, Oxford, 1996.

[16] H. G. Leimer. Optimal decomposition by clique separators. *Discrete Math.*, 113:99–123, 1993.

[17] H. Li and J. Gui. Gradient directed regularization for sparse Gaussian concentration graphs, with applications to inference of genetic network. *Biostatistics*, 7:302–317, 2006.

[18] F. M. Malvestuto and M. Moscarini. Decomposition of a hypergraph by partial-edge separators. *Theoret. Comput. Sci.*, 237:57–79, 2000.

[19] T. P. Speed and H. T. Kiiveri. Gaussian Markov distribution over finite graphs. *Ann. Statist.*, 14:138–150, 1986.

[20] J. Whittaker. *Graphical Models in Applied Multivariate Statistics.* John Wiley and Sons, Chichester, 1990.