A Modified Decomposition Covariance Matrix Estimation for Undirected Gaussian Graphical Model

Ridawarni P., Yenny Hermiana A., Saprilina G.
Graduate School of Mathematics,
University of Sumatera Utara
Medan, Indonesia

***Abstract* - A covariance matrix is an undirected graph that associated with a multivariate probability of a given random vector where each vertex represents the different components of the random vector. Graphical model are framework for representing and conditional independence structures with distribution using graph *G*. This paper discussed a distribution estimation in determining decomposable covariance matrix in an undirected Gauss graphical model related to sparsity of invers covariance (concentration matrix). It showed decomposable covariance estimation with lower computational complexity. The result showed the correlation each different components in a given random vector that determined from decomposition covariance matrix estimation.**

Keywords* - *conditional independence, covariance decomposition, Gauss graphical model, concentration graph

I. INTRODUCTION

Graphical models are a framework in determine a conditional independence structures within distributions that represented by a graph. In representing a distribution, undirected graph is used as a common model to describe a distribution problem in high dimension. Some previous researches that related to undirected graph model has been successfully applied to determine a conditional independence structures within a multivariate distribution and computation techniques that implemented using a graph for complexity enhancement problem of a high dimension data [1].

Covariance is an estimation of the two certain variables, *x* and *y*, in *n* sizes of data sample. This estimation is used to determine a variance and linear correlation of a multivariate or multi-dimension data. Estimation of covariance in a Gaussian distribution is basically a common problem in statistical signal processing such as speech recognition [2], [3], image processing [4], [5], sensor networks [6], computer networks [7] and other fields that related to statistical graphical models. Efficient Bayesian inference in Gaussian graphical models is well established [8] – [10].

A conditional independence among some random variables that distributed to some estimated covariance inverse. Estimation of covariance in a high dimension data can be classified to two categories: estimation that based on the sequence of the variables and estimation that based on estimator, permutation of invariance to each available variable index. Some researches of undirected graphical model was studied and developed to determine a conditional independence structure in a multivariate distribution, and extended with a computation method using a representation by a graph, especially for dimension enhancement and complexity problem. The results showed a sum of weighted path of all available paths that connected the two variables in an undirected independence graph [11].

An estimation of decomposition covariance matrix is the main focus topic in this paper. The estimation is focused on an undirected Gaussian graphical model of two random variables that gives correlation of each vertex on a Gaussian graph as a result.

II. REVIEWS

The Gaussian distribution is also referred to as the normal distribution or the bell curve distribution for its bell-shaped density curve. The formula for a *d*-dimension Gaussian probability distribution is

$$p(x) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left(-\frac{(x-\mu)^T \Sigma^{-1} (x-\mu)}{2} \right) \quad (2.1)$$

where *x* is a *d*-element column vector of variables along each dimension, μ is the mean vector, calculated by

$$\mu = E[x] = \int x\rho(x)\, dx \quad (2.2)$$

and Σ is the $d \times d$ covariance matrix, calculated by

$$\Sigma = E[(x-\mu)(x-\mu^T)] = \int (x-\mu)(x-\mu^T)\, p(x)\, dx \quad (2.3)$$

with the following form

$$\begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1d} \\ \sigma_{21} & \cdots & \sigma_{2d} \\ \sigma_{d1} & \cdots & \sigma_{dd} \end{bmatrix} \tag{2.4}$$

The covariance matrix is always symmetric and positive semidefinite, where positive semidefinite means that for all non-zero $x \in R^d, x^T \sum x \geq 0$. We normally only deal with covariance matrices that are positive definite where for all non-zero $x \in R^d, x^T \sum x \geq 0$, such the determinant $|\sum|$ will be strictly positive. The diagonal elements, σ_{ii} are the variances of the respective x_i, i.e., σ_i^2, and the off-diagonal elements, σ_{ij}, are the covariances of x_i and x_j. If the variables along each dimension is statistically independent, then $\sigma_{ij} = 0$, and we would have a diagonal covariance matrix

$$\begin{bmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \sigma_d^2 \end{bmatrix} \tag{2.5}$$

If the covariances along each dimension is the same, then we will have an identify matrix multiplied by a scalar,

$$\sigma^2 I \tag{2.6}$$

by the Eq. (2.6), the determinant of $\sum$ becomes

$$|\sum| = \sigma^{2d} \tag{2.7}$$

and the inverse of $\sum$ becomes

$$\Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sigma^2} \end{bmatrix} \tag{2.8}$$

For 2-*d* Gaussian where *d* = 2, $x = [x_1 \quad x_2]^T$, $|\sum| = \sigma^4$, the formulation becomes

$$p(x_1, x_2) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x_1^2 + x_2^2}{2\sigma^2} \right) \tag{2.9}$$

then we often denote a Gaussian distribution of Eq. (2.1) as $p(x) \sim N(\mu, \Sigma)$.

An undirected graph $G = (V, E)$ is a set of nodes $V = \{1, \ldots, |V|\}$ connected by undirected edges $E = \{(i_1, j_1), \ldots, (i_{|\Sigma|}, j_{|\Sigma|})\}$, where each node is connected to itself, i.e., $(i,i) \in E$ for all $i \in V$. Let $x = [x_1, \ldots, x_p]^T$ be a zero random vector of length $p = |V|$ whose elements are indexed by the nodes in *V*. The vector *x* satisfies the Markov property with respect to *G*, if for any pair of nonadjacent nodes the corresponding pair of elements in *x* are conditionally independent of the remaining elements, i.e., x_i and x_j are conditionally independent of x_r for any $\{i, j\} \notin E$ and $r = \{V \setminus i, j\}$ where

$$p(x_i, x_j \mid x_r) = p(x_i \mid x_r) p(x_j \mid x_r) \tag{2.10}$$

Therefore, the joint distribution satisfies the following factorization:

$$p(x_i, x_j, x_r) = \frac{p(x_i, x_r) p(x_j, x_r)}{p(x_r)} \tag{2.11}$$

In the Gaussian setting, this factorization leads to sparsity in the concentration (inverse covariance) matrix. The multivariate Gaussian distribution is defined as

$$p(x; K) = c \mid K \mid^{1/2} e^{-1/2 x^T K x} \tag{2.12}$$

where *c'* in an appropriate constant and the marginal concentration matrix is

$$\bar{K}_r = ([K^{-1}]_{r,r})^{-1} \tag{2.13}$$

Together with Eq. (2.11) this implies that

$$\bar{K}_r = ([K^{-1}]_{r,r})^{-1}; K = [\bar{K}_{ir}]^0 + [\bar{K}_{jr}]^0 - [\bar{K}_r]^0 \tag{2.14}$$

$$|K| = \frac{|\bar{K}_{ir}||\bar{K}_{jr}|}{|\bar{K}_r|} \tag{2.15}$$

where $\bar{K}_{ir}$, $\bar{K}_{jr}$ and $\bar{K}_r$ are the marginal concentrations of $\{x_i, x_r\}, \{x_j, x_r\}$ and $\{x_r\}$, respectively, and are defined in a similar manner to Eq. (2.13). All the matrices in the right hand side of Eq. (2.14) have a zero value in the $\{i, j\}$ th position, and therefore

$$[K]_{i,j} = 0 \text{ for all } \{i, j\} \notin E$$

This property is the core of Gaussian graphical models: the concentration matrix *K* has a sparsity pattern which represents the topology of the conditional independence graph.

Decomposable models are a special type of graphical model in which the conditional independence graphs satisfy an appealing structure. A decomposable graph can be divided into an ordered sequence of fully connected subgraphs known as cliques and denoted by $C_1,\ldots,C_k$

III. DECOMPOSITION COVARIANCE MATRIX

In our estimation, we use some notations in determine covariance matrix in Gaussian graphical model. Assume $X = X_{V_p}$ where $V_p = \{1,\ldots,p\}$ is a random vector with normal multivariate distribution with p-dimension, $N_p(0,K^{-1})$. Set a graph $G = (V_p, E)$ where for each vertex $i \in V$ is pair to a variable X_i and $E \subset V_p \times V_p$ which is an undirected graph. Set the definition that $(i,j) \in E$ if and only if $(j,i) \in E$. A Gaussian graphical model with conditional independence graph is determined with the limit of diagonal elements of K that not pair to any edge in G. If $(i,h) \notin E$, then X_i and X_j is conditionally independence of a given random variables. Concentration matrix $K = (K_{ij})_{1\le i,j\le p}$ is a limit to a symmetric positive definite matrix with diagonal entry $K_{ij} = 0$ for all $(i,j) \notin E$. Using G-Wishart distribution, $Wis_G(\delta, D)$ with density

$$p(K|G,\delta,D) = \frac{1}{I_{G(\delta,D)}} \det K^{(\delta-2)/2} \exp\{-\frac{1}{2(K,D)}\} \qquad (3.1)$$

based on Lebesgue estimation (see [12]; [13]; [14]). If G is a complete graph $E = (V_p \times V_p)$, then $Wis_G(\delta, D)$ is a G-Wishart distribution $Wis_G(\delta, D)$. We then using Cholesky decomposition for matrix K with $K \in P_G$ is $K \in P_G$ where $Q = (Q_{ij})_{1\le i\le j\le p}$ and $\psi = (\psi_{ij})_{1\le i\le j\le p}$ is an upper triangle where $D^{-1} = Q^T Q$ is the decomposition Cholesky of D^{-1}.

Estimation of covariance matrix is a basic problem of multivariate statistical that related to signal processing, financial mathematics, pattern recognition and convex geometry computation. Set a sample of n points that independent, $X_1, \ldots, X_n$, from distribution. We then have a sample of covariance matrix

$$\Sigma_n = \frac{1}{n}\sum_{i=1}^{n} X_k X_k^T = \frac{1}{n} A^T A \qquad (3.1)$$

with Σ_n is a random matrix. Then we estimated the covariance matrix Σ with rate of accurancy, $\varepsilon = 0.01$ that represented by a norm operation as follows.

$$\|\Sigma_n - \Sigma\| \le \varepsilon \|\Sigma\| \qquad (3.2)$$

Assume $X = (X_1, X_2, \ldots, X_n)$ of a multivariate Gaussian distribution $\mathcal{N}_d(0,\Sigma)$ where Σ is a regular matrix. Then we determined likelihood function

$$L(K) = \frac{(\det K)^{n/2} e^{-tr(Kw)}}{2} \qquad (3.3)$$

where $W = \sum_{v=1}^{n} X^v (X^v)^T$ as a result of sum and multiple matrix. Thus, the likelihood equation of Gaussian matrix can be formulated as

$$E\left(-\frac{W}{2}\right) = -\frac{n\Sigma}{2} = -\frac{w}{2} \qquad (3.4)$$

$$\hat{K}^{-1} = \hat{\Sigma} = \frac{w}{n} \qquad (3.5)$$

$$\log L(K) = \frac{n}{2}\log(\det K) - \frac{tr(Kw)}{2} \qquad (3.6)$$

$$\frac{\partial}{\partial k_{ij}} \log(\det K) = \frac{w_{ij}}{n} \qquad (3.7)$$

$$\frac{\partial}{\partial k_{ij}} \log(\det K) = K^{-1} \qquad (3.8)$$

IV. DECOMPOSITION COVARIANCE MATRIX FOR UNDIRECTED GAUSSIAN GRAPH

Adopted by a basic structure of variance component and time series problem, we suggest definition of linear covariance formula model that represented as

$$\Sigma = a_1 U_1 + \ldots + a_q U_q \qquad (4.1)$$

U_i are symmetric matrices and α_i is an unknown parameter that supposed to be a requirement so that the matrix is positive definite. Eq. (4.1) represents a common formula of time-series covariance model, mixed-linear and graph model. Specifically, high dimension q for any covariance matrix can be denoted as

$$\Sigma = (\sigma_{ij}) = \sum_{i=1}^{p}\sum_{j=1}^{q} \sigma_{ij} U_{ij} \qquad (4.2)$$

with U_{ij} is matrix with dimension $p \times p$ with element 1 at (i,j) and 0 otherwise. For each column and row, we can set variance matrix with these following steps.

Step 1. Set matrix X as a deviation for x where $x = X - 11'X\left(\frac{1}{n}\right)$.

Step 2. Calculate $x'x$ as a result of sum and multiple matrix with dimension $k \times k$ at matrix x.

Step 3. Each element of matrix x is divided by n. Thus, we determined variance-covariance matrix of matrix x, $V = x'x\left(\frac{1}{n}\right)$, where

1 : column vector with element 1 of dimension $n \times 1$
x : deviation matrix of dimension $m \times n$
X : data matrix of dimension $m \times n$
V : covariance matrix of dimension $m \times n$
$x'x$: result of sum and multiple matrix
n : number of trial of matrix X

Thus, we then determined a decomposition covariance matrix that showed correlation partial of the two random variables by

$$\rho_{ij} = -\frac{k_{ij}}{\sqrt{k_{ii}k_{jj}}} \qquad (4.3)$$

V. RESULTS

A. *Matrix* m × m

For illustration, we use a given matrix data of dimension $m \times m$. Assume that there exist matrix $A_{3\times3}$ as follows.

$$A_{3\times3} = \begin{pmatrix} 1 & 4 & 1 \\ 2 & 1 & 4 \\ 3 & 3 & 1 \end{pmatrix}$$

Then, covariance matrix of matrix A is determined by the definition $Cov(A) = A - 11'A$. Thus,

$$\mathrm{cov}(A) = \begin{pmatrix} -5 & -4 & -5 \\ -4 & -7 & -2 \\ -3 & -5 & -5 \end{pmatrix}$$

Because $n = 3$, for each element of covariance matrix we have

$$\mathrm{cov}(A) = \begin{pmatrix} -5 & -4 & -5 \\ -4 & -7 & -2 \\ -3 & -5 & -5 \end{pmatrix}\begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} -1.667 & -1.333 & -1.667 \\ -1.333 & -2.333 & -0.667 \\ -1.000 & -1.667 & -1.667 \end{pmatrix}$$

with the decomposition of inverse matrix is

$$inv(A) = \begin{pmatrix} -1.172 & -0.234 & 1.266 \\ 0.656 & -0.469 & -0.469 \\ 0.047 & 0.609 & -0.890 \end{pmatrix}$$

and for each correlation of the two random variables, we determined

$$\rho_{(12|3)} = \frac{0.2344}{\sqrt{(-1.1719)(-0.4688)}} = \frac{0.2344}{0.7412} = 0.3162$$

$$\rho_{(13|3)} = \frac{-0.0469}{\sqrt{(-1.1719)(-0.8906)}} = \frac{-0.0469}{1.0216} = -0.0459$$

$$\rho_{(23|3)} = \frac{0.4788}{\sqrt{(-0.4688)(-0.8906)}} = \frac{0.4688}{0.6461} = 0.7255$$

B. *Matrix* m × n

Assume that there exist matrix

$$A_{5\times3} = \begin{pmatrix} 4.0 & 2.0 & 0.60 \\ 4.2 & 2.1 & 0.59 \\ 3.9 & 2.0 & 0.58 \\ 4.3 & 2.1 & 0.62 \\ 4.1 & 2.2 & 0.63 \end{pmatrix}$$

which the covariance matrix of matrix A is determined by the definition $Cov(A) = A - 11'A$. Thus,

$$\mathrm{cov}(A) = \begin{pmatrix} -0.1 & -0.08 & -0.004 \\ 0.1 & 0.02 & -0.014 \\ -0.2 & -0.08 & -0.024 \\ 0.2 & 0.02 & 0.016 \\ 0.0 & 0.12 & 0.026 \end{pmatrix}$$

Because $n = 5$, for each element of covariance matrix we have

$$\mathrm{cov}(A) = \begin{pmatrix} -0.1 & -0.08 & -0.004 \\ 0.1 & 0.02 & -0.014 \\ -0.2 & -0.08 & -0.024 \\ 0.2 & 0.02 & 0.016 \\ 0.0 & 0.12 & 0.026 \end{pmatrix}\begin{pmatrix} 1 \\ 5 \end{pmatrix} = \begin{pmatrix} 0.02 & 0.006 & 0.0014 \\ 0.006 & 0.0056 & 0.0011 \\ 0.0014 & 0.0011 & 0.0003 \end{pmatrix}$$

with the decomposition of inverse matrix is

$$inv(A) = \begin{pmatrix} 76.1 & 55.3 & 136.2 \\ 20.8 & 492.8 & 1322.1 \\ -136.2 & 1322.1 & 7612.2 \end{pmatrix}$$

and for each correlation of the two random variables, we determined

$$\rho_{(12|3)} = \frac{55.3}{\sqrt{(76.1)(492.8)}} = \frac{55.3}{193.65} = 0.285$$

$$\rho_{(13|3)} = \frac{55.3}{7612.2} = \frac{55.3}{761.109} = 0.072$$

$$\rho_{(23|3)} = \frac{1322.1}{\sqrt{(492.8)(7612.2)}} = \frac{1322.1}{1936.825} = 0.682$$

VI. CONCLUSIONS

In this paper, we studied the decomposition of covariance for undirected Gaussian graph. Estimation of covariance in a Gaussian distribution is a common problem in statistic. Gaussian graphical model is a method that can be used to represent the structure are independent among different independence random variables in a graph. This paper has examined the results of the covariance estimation for undirected Gaussian by using a likelihood function in order to obtain the concentration of each random variables. The results of the decomposition matrix is decomposed and can be used in solving problems that related to signal transmission, patter recognition and other concentrations matrix problem.

REFERENCES

[1] Dobra, A., Hans, C., Jones, B., Nevins, J.R., Yao, G. and West, M. *Sparse graphical models for exploring gene expression data*, Journal of Multivariate Analysis, special issue on Multivariate Methods in Genomic Data Analysis, Vol. 90 (2004), pp. 196-212.

[2] Bilmes, J.A. *Factored sparse inverse covariance matrices*. In Proc. IEEE Int. Conf. Acoust. Speech Signal Process (ICASSP). (2000), pp. 1009-1012.

[3] Bilmes, J.A., and Bartels, C. *Graphical model architectures for speech recognition.* IEEE Signal Processing Mag., Vol. 22 (2005), pp. 89-100.

[4] Willsky, A. *Multiresolution Markov models for signal and image processing.* Proc. IEEE, Vol. 90 (2002), pp. 1396-1458.

[5] Choi, M.J., and Willsky, A. Multiscale Gaussian graphical models and algorithms for large-scale inference. In Proc. IEEE 14-th Work-Shop Statistical Signal Process (SSP). (2007), pp. 229-233.

[6] Cetin, M., Chen, L., Fisher, W., Ihler, A.T., Moses, R.L., Wainwright, M.J., and Willsky, A.S. *Distributed fusion in sensor networks: A graphical models perspective*. IEEE Signal Process Mag., Vol. 23 (2006), pp. 42-55.

[7] Wiesel, A. and Hero, A.O. *Decomposable principal component analysis.* IEEE Trans. Signal Process, Vol. 59 (2009), pp. 4369-4377.

[8] Sudderth, E.B., Wainwright, M.J. and Willsky, A.S. *Embedded trees: estimation of Gaussian processes on graphs with cycles,* Vol. 52 (2004), pp. 3136-3150.

[9] Malioutov, D.M., Johnson, J.K., and Willsky, A.S. *Walk-sums and belief propagation in Gaussian graphical models.* J. Mach. Learn., Res., Vol. 7 (2006), pp. 2031-2064.

[10] Chandrasekaran, V., Johnson, J.,K. and Willsky, A.S. *Estimation in Gaussian graphical models using tractable subgraphs: A-walk sum analysis.* IEEE Trans. Signal Process, Vol., 56 (2008), pp. 1916-1930.

[11] Wiesel, A. and Hero III, A.O. *Distributed covariance estimation in Gaussian graphical models.*, IEEE Trans. On Signal Processing, Vol. 60 (2012), pp. 211-220.

[12] Roverato, A. *Hyper inverse Wishart distribution for nondecomposable graphs and its applications to Bayesian inference for Gaussian graphical models,* Scandinavian Journal of Statistics, Vol. 29 (2002), pp.391-411.

[13] Atay-Kayis, A. and Massam, H. *A Monte Carlo method for computing the marginal likelihood in nondecomposable Gaussian graphical models*, Biometrica, Vol. 92 (2005), pp. 317-335.

[14] Letac, G. and Massam, H. *Wishart distributions for decomposable graphs,* The Annals of Statistics, Vol. 35 (2007), pp. 1278-1323.