On Cholesky structures on real symmetric matrices and their applications

Hideyuki ISHI
(Osaka City University)

Virtual Conference “Mathematical Methods of Modern Statistics 2,” CIRM, June 2020.

Cholesky structure : a generalization of Fill-in free property of a sparse matrix with respect to the Cholesky decomposition

$\longrightarrow$

Exact calculation for Gaussian Selection model associated to decomposable graph with symmetry of vertex permutation

Plan:

§1. Cholesky structure

§2. Colored graphical model

§3. Gaussian selection model with a quasi-Cholesky structure

Joint with P. Graczyk, B. Kołodziejek and H. Massam

§1. Cholesky structure.

$\mathcal{P}_n := \{\, x \in \mathrm{Sym}(n,\mathbb{R}) \,|\, x \text{ is positive definite} \,\}$,

$\mathfrak{h}_n := \left\{\, T = (T_{ij}) \in \mathrm{Mat}(n,\mathbb{R}) \,|\, T_{ij} = 0 \;\; (i < j) \,\right\}$,

$H_n := \{\, T \in \mathfrak{h}_n \,|\, T_{ii} > 0 \;\; (i = 1, \dots, n) \,\}$.

Fact. One has a bijection $H_n \ni T \mapsto T\,{}^{\mathrm{t}}T \in \mathcal{P}_n$.

In other words,

$$\forall x \in \mathcal{P}_n \, \exists 1 \, T_x \in H_n \text{ s.t. } x = T_x \, {}^{\mathrm{t}}T_x \text{ (Cholesky decomposition).}$$

When x is sparse, T_x is sometimes sparse, too.

For example, if $x = \begin{pmatrix} x_{11} & x_{21} & 0 & 0 \\ x_{21} & x_{22} & x_{32} & 0 \\ 0 & x_{32} & x_{33} & x_{43} \\ 0 & 0 & x_{43} & x_{44} \end{pmatrix} \in \mathcal{P}_4$,

then T_x is of the form $\begin{pmatrix} T_{11} & 0 & 0 & 0 \\ T_{21} & T_{22} & 0 & 0 \\ 0 & T_{32} & T_{33} & 0 \\ 0 & 0 & T_{43} & T_{44} \end{pmatrix}$.

If $x = \begin{pmatrix} x_{11} & x_{21} & 0 & 0 & x_{51} \\ x_{21} & x_{22} & x_{32} & 0 & 0 \\ 0 & x_{32} & x_{33} & x_{43} & 0 \\ 0 & 0 & x_{43} & x_{44} & x_{54} \\ x_{51} & 0 & 0 & x_{54} & x_{55} \end{pmatrix} \in \mathcal{P}_5$, then T_x is

of the form $\begin{pmatrix} T_{11} & 0 & 0 & 0 & 0 \\ T_{21} & T_{22} & 0 & 0 & 0 \\ 0 & T_{32} & T_{33} & 0 & 0 \\ 0 & 0 & T_{43} & T_{44} & 0 \\ T_{51} & T_{52} & T_{53} & T_{54} & T_{55} \end{pmatrix}$.

We have two fill-ins at T_{52} and T_{53}.

Let $\mathcal{Z}_1$ be a vector subspace of $\mathrm{Sym}(5, \mathbb{R})$ consisting of $x = \begin{pmatrix} x_{11} & x_{21} & 0 & 0 & x_{51} \\ x_{21} & x_{22} & x_{32} & 0 & 0 \\ 0 & x_{32} & x_{33} & x_{43} & 0 \\ 0 & 0 & x_{43} & x_{44} & x_{54} \\ x_{51} & 0 & 0 & x_{54} & x_{55} \end{pmatrix}$, and consider a subspace of $\mathfrak{h}_5$ spanned by T_x with $x \in \mathcal{Z}_1 \cap \mathcal{P}_5$. Then we see that

$$\dim \mathrm{span}_{\mathbb{R}} \{ T_x \,|\, x \in \mathcal{Z}_1 \cap \mathcal{P}_5 \} = \dim \mathcal{Z}_1 + 2.$$

On the other hand, if $\mathcal{Z}_2 \subset \mathrm{Sym}(4, \mathbb{R})$ is the space of $x = \begin{pmatrix} x_{11} & x_{21} & 0 & 0 \\ x_{21} & x_{22} & x_{32} & 0 \\ 0 & x_{32} & x_{33} & x_{43} \\ 0 & 0 & x_{43} & x_{44} \end{pmatrix}$, then

$$\dim \mathrm{span}_{\mathbb{R}} \{ T_x \,|\, x \in \mathcal{Z}_2 \cap \mathcal{P}_4 \} = \dim \mathcal{Z}_2.$$

Definition 1. Let $\mathcal{Z}$ be a vector subspace of $\mathrm{Sym}(n,\mathbb{R})$ such that $I_n \in \mathcal{Z}$. We say that $\mathcal{Z}$ has a Cholesky structure if

$$\dim \operatorname{span}\{\, T_x \,|\, x \in \mathcal{Z} \cap \mathcal{P}_n \,\} = \dim \mathcal{Z}.$$

For $x \in \mathrm{Sym}(n,\mathbb{R})$, define $\underset{\vee}{x} \in \mathfrak{h}_n$ by $(\underset{\vee}{x})_{ij} := \begin{cases} 0 & (i<j), \\ x_{ii}/2 & (i=j), \\ x_{ij} & (i>j), \end{cases}$ and $\hat{x} := {}^{\mathrm{t}}(\underset{\vee}{x})$. Then $x = \underset{\vee}{x} + \hat{x}$. Let $\underset{\vee}{\mathcal{Z}}$ denote the space $\left\{ \underset{\vee}{x} \,|\, x \in \mathcal{Z} \right\} \subset \mathfrak{h}_n$. If $I_n \in \mathcal{Z}$, then $\underset{\vee}{\mathcal{Z}}$ equals the tangent space of $\{\, T_x \,|\, x \in \mathcal{Z} \cap \mathcal{P}_n \,\} \subset H_n$ at I_n, so that $\underset{\vee}{\mathcal{Z}} \subset \operatorname{span}\{\, T_x \,|\, x \in \mathcal{Z} \cap \mathcal{P}_n \,\}$.

Theorem 2. Let $\mathcal{Z}$ be a vector subspace of $\mathrm{Sym}(n, \mathbb{R})$ such that $I_n \in \mathcal{Z}$. Then the following are all equivalent:
(i) $\mathcal{Z}$ has a Cholesky structure.
(ii) $\mathrm{span}\,\{\, T_x \,|\, x \in \mathcal{Z} \cap \mathcal{P}_n \,\} = \underset{\vee}{\mathcal{Z}}$.
(iii) $\forall x \in \mathcal{Z}\ \underset{\vee}{x}\hat{x} \in \mathcal{Z}$.
(iv) One has a bijection $\underset{\vee}{\mathcal{Z}} \cap H_n \ni T \mapsto T\,{}^{\mathrm{t}}T \in \mathcal{Z} \cap \mathcal{P}_n$.

(i) $\Leftrightarrow$ (ii) is obvious, (ii) $\Rightarrow$ (iii) is easy, and (iv) $\Rightarrow$ (ii) is trivial. A crucial part is (iii) $\Rightarrow$ (iv).

For $x, y \in \mathrm{Sym}(n, \mathbb{R})$, define $x \diamond y := 2(\underset{\vee}{x}\hat{y} + \underset{\vee}{y}\hat{x}) \in \mathrm{Sym}(n, \mathbb{R})$.
Then $x \diamond I_n = I_n \diamond x = x$, and (iii) is equivalent to $\mathcal{Z} \diamond \mathcal{Z} \subset \mathcal{Z}$ i.e. $\forall x, y \in \mathcal{Z}\ \ x \diamond y \in \mathcal{Z}$.
Temporally, we say that $\mathcal{Z} \subset \mathrm{Sym}(n, \mathbb{R})$ is a Cholesky algebra if $I_n \in \mathcal{Z}$ and $\mathcal{Z} \diamond \mathcal{Z} \subset \mathcal{Z}$.

Let $\mathcal{Z} \subset \mathsf{Sym}(n, \mathbb{R})$ be a Cholesky algebra and $\mathcal{W} \subset \mathsf{Mat}(n, m, \mathbb{R})$ a subspace such that

$$u \in \mathcal{W} \Rightarrow u\,{}^{\mathsf{t}}u \in \mathcal{Z}.$$

Then

$$E(\mathcal{Z}; \mathcal{W}) := \left\{ \begin{pmatrix} cI_m & {}^{\mathsf{t}}u \\ u & x \end{pmatrix} \,\middle|\, c \in \mathbb{R},\ u \in \mathcal{W},\ x \in \mathcal{Z} \right\}$$
$$\subset \mathsf{Sym}(m+n, \mathbb{R})$$

is a Cholesky algebra because

$$\begin{pmatrix} (c/2)I_m & 0 \\ u & \underset{\vee}{x} \end{pmatrix} \begin{pmatrix} (c/2)I_m & {}^{\mathsf{t}}u \\ 0 & \hat{x} \end{pmatrix} = \begin{pmatrix} (c^2/4)I_m & c\,{}^{\mathsf{t}}u/2 \\ cu/2 & \underset{\vee}{x}\,\hat{x} + u\,{}^{\mathsf{t}}u \end{pmatrix} \in E(\mathcal{Z}, \mathcal{W}).$$

Starting from one-dimensional algebra $\mathbb{R}I_n \subset \mathsf{Sym}(n, \mathbb{R})$, we obtain Cholesky algebras by repetition of this extension procedure.

For example, let $\mathcal{Z}$ be the set of symmetric matrices of the form

$$\begin{pmatrix} c_1 & 0 & a & 0 \\ 0 & c_1 & 0 & a \\ a & 0 & c_2 & b \\ 0 & a & b & c_3 \end{pmatrix}$$

Setting $\mathcal{W}_1 := \mathbb{R}I_2$ and $\mathcal{W}_2 := \mathbb{R}$, we have $\mathcal{Z} = E(E(\mathbb{R};\mathcal{W}_2);\mathcal{W}_1)$.

We say that a Cholesky algebra $\mathcal{Z}$ is standard if $\mathcal{Z} = \mathbb{R}I_n$ or $\mathcal{Z} = E(E(\cdots(E(\mathbb{R}I_s;\mathcal{W}_{r-1});\cdots);\mathcal{W}_2);\mathcal{W}_1)$ with appropriate vector spaces $\mathcal{W}_1, \mathcal{W}_2, \ldots, \mathcal{W}_{r-1}$.

Theorem 3. Any Cholesky algebra is isomorphic to a standard one, and the isomorphism is given by an appropriate permutation of rows and columns.

For example, the Cholesky algebra $\mathscr{Z}$ of matrices

$$\begin{pmatrix} a & b & 0 & 0 \\ b & c & 0 & 0 \\ 0 & 0 & a & b \\ 0 & 0 & b & d \end{pmatrix}$$

is isomorphic to the Cholesky algebra $\mathscr{Z}'$ of matrices

$$\begin{pmatrix} a & 0 & b & 0 \\ 0 & a & 0 & b \\ b & 0 & c & 0 \\ 0 & b & 0 & d \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a & b & 0 & 0 \\ b & c & 0 & 0 \\ 0 & 0 & a & b \\ 0 & 0 & b & d \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

by the permutation $(23) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \end{pmatrix}$.

The crucial part of Theorem 2 (i.e. (iii) $\Rightarrow$ (iv)) follows from Theorem 3. Eventually, we conclude that $\mathcal{Z}$ has a Cholesky structure if and only if $\mathcal{Z}$ is a Cholesky algebra.

Definition 4. We say that a subspace $\mathcal{Z}$ of $\mathrm{Sym}(n, \mathbb{R})$ has a quasi-Cholesky structure if there exists an invertible matrix $A \in GL(n, \mathbb{R})$ such that $\mathcal{Z}^A := \left\{ Ax^{\mathrm{t}}A \,\middle|\, x \in \mathcal{Z} \right\}$ has a Cholesky structure.

For example, a vector space $\mathcal{Z} \subset \mathrm{Sym}(4, \mathbb{R})$ consisting of $x = \begin{pmatrix} a & b & 0 & 0 \\ b & c & d & 0 \\ 0 & d & c & b \\ 0 & 0 & b & a \end{pmatrix}$, corresponding to a colored graph $a \overset{b}{-} c \overset{d}{-} c \overset{b}{-} a$, has a quasi-Cholesky structure.

Indeed, putting $A := \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 & -1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}$, we have

$$A \begin{pmatrix} a & b & 0 & 0 \\ b & c & d & 0 \\ 0 & d & c & b \\ 0 & 0 & b & a \end{pmatrix} {}^{\mathrm{t}}A = \begin{pmatrix} a & 0 & b & 0 \\ 0 & a & 0 & b \\ b & 0 & c-d & 0 \\ 0 & b & 0 & c+d \end{pmatrix},$$

so that

$$\mathcal{Z}^A = \left\{ \begin{pmatrix} a & 0 & b & 0 \\ 0 & a & 0 & b \\ b & 0 & c & 0 \\ 0 & b & 0 & d \end{pmatrix} \mid a, b, c, d \in \mathbb{R} \right\},$$

which has a Cholesky structure.

§2. Colored graphical model

Let $G = (V, E)$ be an undirected graph with $V = \{1, \cdots, n\}$ and $E \subset V \times V$. The graph G is said to be decomposable or chordal if any cycle in G of length ≥ 4 has a chord. Let $\mathcal{Z}_G \subset \mathrm{Sym}(n, \mathbb{R})$ be the space of $x = (x_{ij})$ for which $x_{ij} = 0$ if $i \neq j$ and $(i, j) \notin E$. It is known that, if G is decomposable and V is labeled appropriately, then each $x \in \mathcal{Z}_G$ is decomposed as $x = T_x {}^{\mathrm{t}}T_x$ without fill-ins. In our terminology, $\mathcal{Z}_G$ has a Cholesky structure.

For example, when $G = 1 - 2 - 3 - 4$, then $\mathcal{Z}_G$ is the vector space of $x = \begin{pmatrix} x_{11} & x_{21} & 0 & 0 \\ x_{21} & x_{22} & x_{32} & 0 \\ 0 & x_{32} & x_{33} & x_{43} \\ 0 & 0 & x_{43} & x_{44} \end{pmatrix}$.

Let $\mathrm{Aut}(G)$ be the set of permutations $\sigma \in S_n$ such that $(\sigma(i), \sigma(j)) \in E \Leftrightarrow (i,j) \in E$. Let Γ be a subgroup of $\mathrm{Aut}(G)$, and define

$$\mathcal{Z}_G^\Gamma := \left\{ x \in \mathcal{Z}_G \,|\, \forall \sigma \in \Gamma \ \forall i,j \in V \ \ x_{\sigma(i)\sigma(j)} = x_{ij} \right\},$$

which corresponds to the graph G whose vertices and edges are colored so that the objects mapped each other by Γ have the same color.

For example, if $G = 1-2-3-4$ with $\Gamma = \mathrm{Aut}(G) = \{id, (14)(23)\}$. Then we have a colored graph 1–2–3–4 and

$$\mathcal{Z}_G^\Gamma = \left\{ \begin{pmatrix} a & b & 0 & 0 \\ b & c & d & 0 \\ 0 & d & c & b \\ 0 & 0 & b & a \end{pmatrix} \,\Big|\, a,b,c,d \in \mathbb{R} \right\}.$$

Theorem 4. Let G be a decomposable and Γ any subgroup of $\mathrm{Aut}(G)$. Then $\mathcal{Z}_G^{\Gamma} \subset \mathrm{Sym}(n, \mathbb{R})$ has a quasi-Cholesky structure.

A crucial point is how to find $A \in GL(n, \mathbb{R})$ for which $(\mathcal{Z}_G^{\Gamma})^A$ has a Cholesky structure.

Thanks to Theorem 4, we can generalize analysis on $\mathcal{Z}_G$ by Letac-Massam (2007) to $\mathcal{Z}_G^{\Gamma}$.

§3. Gaussian selection model with a quasi-Cholesky structure.

Let $\mathcal{Z}$ be a vector subspace of $\mathrm{Sym}(n, \mathbb{R})$ such that $\mathcal{P}_{\mathcal{Z}} = \mathcal{Z} \cap \mathcal{P}_n$ is non-empty. We consider a statistical model $\mathcal{M} := \left\{ N_n(0, \Sigma) \,|\, \Sigma^{-1} \in \mathcal{P}_{\mathcal{Z}} \right\}$, where $N_n(0, \Sigma)$ stands for the multivariate zero-mean normal law with covariant matrix Σ.

Let $\pi_{\mathcal{Z}} : \mathrm{Sym}(n, \mathbb{R}) \to \mathcal{Z}$ be the orthogonal projection with respect to the trace inner product. Let $X_1, X_2, \ldots, X_s$ be i. i. d. obeying $N_n(0, \Sigma)$ with $\Sigma^{-1} \in \mathcal{P}_{\mathcal{Z}}$. Then a $\mathcal{Z}$-valued random matrix $Y := \pi_{\mathcal{Z}}(X_1{}^{\mathrm{t}}X_1 + \cdots + X_s{}^{\mathrm{t}}X_s)/2$ is a sufficient statistics of the model $\mathcal{M}$. Let $\mathcal{W}_{s,\Sigma}$ denote the law of Y, which we call the Wishart law for the model $\mathcal{M}$.

Let

$$\mathcal{Q}_{\mathcal{Z}} := \left\{ y \in \mathcal{Z} \,|\, \mathrm{tr}(xy) > 0 \text{ for } x \in \overline{\mathcal{P}_{\mathcal{Z}}} \setminus \{0\} \right\}.$$

Then we have a bijection $\mathcal{P}_{\mathcal{Z}} \ni x \mapsto \pi_{\mathcal{Z}}(x^{-1}) \in \mathcal{Q}_{\mathcal{Z}}$. We define $\delta_{\mathcal{Z}} : \mathcal{Q}_{\mathcal{Z}} \to \mathbb{R}$ by

$$\delta_{\mathcal{Z}}(y) := (\det x)^{-1} \quad (y = \pi_{\mathcal{Z}}(x^{-1}) \in \mathcal{Q}_{\mathcal{Z}},\ x \in \mathcal{P}_{\mathcal{Z}}).$$

The log-gradient map $\nabla \log \delta_{\mathcal{Z}} : \mathcal{Q}_{\mathcal{Z}} \to \mathcal{P}_{\mathcal{Z}}$ gives the inverse map of $\mathcal{P}_{\mathcal{Z}} \ni x \mapsto \pi_{\mathcal{Z}}(x^{-1}) \in \mathcal{Q}_{\mathcal{Z}}$. If $x_1, \ldots, x_s \in \mathbb{R}^n$ are samples of the model $\mathcal{M}$, then

$$\hat{\Sigma}^{-1} = \nabla \log \delta_{\mathcal{Z}} \left(\pi_{\mathcal{Z}}\Big(\frac{1}{s} \sum_{k=1}^{s} {x_k}^{\mathrm{t}} x_k\Big) \right) \in \mathcal{P}_{\mathcal{Z}}$$

provided that $\pi_{\mathcal{Z}}(\frac{1}{s}\sum_{k=1}^{s} {x_k}^{\mathrm{t}} x_k) \in \mathcal{Q}_{\mathcal{Z}}$.

In what follows, we assume that $\mathcal{Z}$ has a quasi-Cholesky structure.
Proposition 5. $\delta_{\mathcal{Z}}(y)$ is explicitly expressed as a rational function of $y \in \mathcal{Q}_{\mathcal{Z}}$.

Define $\varphi_{\mathcal{Z}}(y) := \int_{\mathcal{P}_{\mathcal{Z}}} e^{-\mathrm{tr}\,(xy)}\, dx$ for $y \in \mathcal{Q}_{\mathcal{Z}}$.

Theorem 6. One has

$$\int_{\mathcal{Q}_{\mathcal{Z}}} e^{-\mathrm{tr}(xy)}\, \delta_{\mathcal{Z}}(y)^s \varphi_{\mathcal{Z}}(y)\, dy = \Gamma_{\mathcal{Z}}(s)(\det x)^{-s} \quad (x \in \mathcal{P}_{\mathcal{Z}},\ \Re s > s_0),$$

where s_0 is a real number, and $\Gamma_{\mathcal{Z}}(s)$ is a holomorphic function of s with $\Re s > s_0$.
Theorem 7 If $s/2 > s_0$, then the density function of the Wishart law $\mathcal{W}_{s,\Sigma}$ of $Y = \pi_{\mathcal{Z}}(X_1{}^{\mathrm{t}}X_1 + \cdots + X_s{}^{\mathrm{t}}X_s)/2$ equals

$$\Gamma_{\mathcal{Z}}(s/2)^{-1}(\det \Sigma)^{-s/2} e^{-\mathrm{tr}(y\Sigma^{-1})} \delta_{\mathcal{Z}}(y)^{s/2} \varphi_{\mathcal{Z}}(y)\, 1_{\mathcal{Q}_{\mathcal{Z}}}(y).$$

If $\mathcal{Z}$ is the space of matrices $\begin{pmatrix} a & b & 0 & 0 \\ b & c & d & 0 \\ 0 & d & c & b \\ 0 & 0 & b & a \end{pmatrix}$, then

$$\mathcal{Q}_{\mathcal{Z}} = \left\{ y = \begin{pmatrix} a & b & 0 & 0 \\ b & c & d & 0 \\ 0 & d & c & b \\ 0 & 0 & b & a \end{pmatrix} \in \mathcal{Z} \,|\, c - d > 0,\, c + d > 0,\, a - b^2/c > 0 \right\}.$$

Moreover,

$$\begin{aligned} \delta_{\mathcal{Z}}(y) &= (c-d)(c+d)(a-b^2/c)^2, \\ \varphi_{\mathcal{Z}}(y) &= 2^{-1/2}\pi c^{-1/2}(c-d)^{-1}(c+d)^{-1}(a-b^2/c)^{-3/2}, \end{aligned}$$

for $y \in \mathcal{Q}_{\mathcal{Z}}$, and

$$\Gamma_{\mathcal{Z}}(s) = 2^{-3/2}\pi\Gamma(s-1/4)\Gamma(s+1/4)\Gamma(s)^2.$$

The Gamma-type formula becomes

$$\int_{\mathcal{Q}_{\mathcal{Z}}} e^{-\mathrm{tr}\,(xy)}(c-d)^{s-1}(c+d)^{s-1}(a-b^2/c)^{2s-3/2}c^{-1/2}\,dadbdcdd$$
$$= 2^{-7/2}\Gamma(s-1/4)\Gamma(s+1/4)\Gamma(s^2)(\det x)^{-s},$$

for $x \in \mathcal{P}_{\mathcal{Z}}$ and $\Re s > 1/4$. Therefore, for any $s \geq 1$, the density function of $\mathcal{W}_{s,\Sigma}$ equals

$$2^{7/2}\Gamma(s/2-1/4)^{-1}\Gamma(s/2+1/4)^{-1}\Gamma(s/2)^{-2}(\det\Sigma)^{-s/2}$$
$$\times e^{-\mathrm{tr}(y\Sigma^{-1})}(c-d)^{s/2-1}(c+d)^{s/2-1}(a-b^2/c)^{s-3/2}c^{-1/2}1_{\mathcal{Q}_{\mathcal{Z}}}(y).$$