On Gaussian Group Convex Models

Hideyuki Ishi

Citation	GSI 2021: Geometric Science of Information. PP.256-264.
Conference proceedings	5th International Conference, GSI 2021, Paris, France, July 21–23, 2021, Proceedings
book series	Lecture Notes in Computer Science (LNCS, volume 12829)
book sub series	Image Processing, Computer Vision, Pattern Recognition, and Graphics (LNIP)
Published	2021-07-14
Type	Conference Paper
Textversion	Author
Relation	This is an Accepted Manuscript of a book chapter published in Lecture Notes in Computer Science book series (LNIP,volume 12829). The final authenticated version is available online at: https://doi.org/10.1007/978-3-030-80209-7_29.
Rights	See Springer Nature terms of use: https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms
DOI	10.1007/978-3-030-80209-7_29

Ishi, H. (2021). On Gaussian Group Convex Models. In: Nielsen, F., Barbaresco, F. (eds) Geometric Science of Information. GSI 2021. Lecture Notes in Computer Science, vol 12829. Springer, Cham. https://doi.org/10.1007/978-3-030-80209-7_29

On Gaussian Group Convex Models⋆

Hideyuki Ishi

Osaka City University, OCAMI, 3-3-138 Sugimoto, Sumiyoshi-ku, 558-8585 Osaka, Japan
hideyuki@sci.osaka-cu.ac.jp

Abstract. The Gaussian group model is a statistical model consisting of central normal distributions whose precision matrices are of the form $gg^\top$, where g is an element of a matrix group G. When the set of $gg^\top$ is convex in the vector space of real symmetric matrices, the set forms an affine homogeneous convex domain studied by Vinberg. In this case, we give the smallest number of samples such that the maximum likelihood estimator (MLE) of the parameter exists with probability one. Moreover, if the MLE exists, it is explicitly expressed as a rational function of the sample data.

Keywords: Gaussian group model · affine homogeneous convex domain · maximum likelihood estimator · Riesz distribution

1 Introduction

Let G be a subgroup of $GL(N,\mathbb{R})$, and $\mathcal{M}_G$ the set $\{\, gg^\top \,;\, g \in G \,\} \subset \mathrm{Sym}(N,\mathbb{R})$. The Gaussian group model associated to G is a statistical model consisting of central multivariate normal laws whose precision (concentration) matrices belong to $\mathcal{M}_G$ ([1]). It is an example of a transformation family, that is, an exponential family whose parameter space forms a group (see [2]). A fundamental problem is to estimate an unknown precision matrix $\theta = gg^\top \in \mathcal{M}_G$ from samples $X_1, X_2, \dots, X_n \in \mathbb{R}^N$. In [1], the existence and uniqueness of the maximum likelihood estimator (MLE) $\hat{\theta}$ of θ are discussed in connection with Geometric Invariant Theory. In the present paper, under the assumption that $\mathcal{M}_G \neq \{I_N\}$ is a convex set, we compute the number n_0 for which MLE exists uniquely with probability one if and only if $n \geq n_0$ (Theorem 4). In this case, an expression of the MLE as a rational function of the samples $X_1, \dots, X_n$ is given (Theorem 3).

Since $\mathcal{M}_G$ is convex, it is regarded as a convex domain in an affine space $I_N + V$, where V is a linear subspace of the vector space $\mathrm{Sym}(N,\mathbb{R})$ of real symmetric matrices of size N. Moreover $\mathcal{M}_G$ is contained in the cone $\mathrm{Sym}^+(N,\mathbb{R})$ of positive definite symmetric matrices, so that $\mathcal{M}_G$ does not contain any straight line (actually, we shall see that $\mathcal{M}_G$ is exactly the intersection $(I_N+V)\cap\mathrm{Sym}^+(N,\mathbb{R})$, see Proposition 2 and Theorem 2). Thus $\mathcal{M}_G$ is an affine homogeneous convex

⋆ Partially supported by KAKENHI 20K03657 and Osaka City University Advanced Mathematical Institute (MEXT Joint Usage/Research Center on Mathematics and Theoretical Physics JPMXP0619217849)

domain on which G acts transitively as affine transforms. Then we can apply Vinberg's theory [9] to $\mathcal{M}_G$. In particular, using the left-symmetric algebra structure on V explored in [6], we give a specific description of $\mathcal{M}_G$ as in Theorem 2, where the so-called real Siegel domain appears naturally. On the other hand, every homogeneous cone is obtained as $\mathcal{M}_G$ by [6]. In particular, all the symmetric cones discussed in [3] as well as the homogeneous graphical models in [8, Section 3.3] appear in our setting.

The rational expression of MLE is obtained by using the algebraic structure on V, whereas the existence condition is deduced from the previous works [4] and [5] about the Wishart and Riesz distributions on homogeneous cones. It seems feasible to generalize the results of this paper to a wider class of Gaussian models containing decomposable graphical models based on [7] in future.

The author should like to express his gratitude to Professor Fumihiro Sato for the information of the paper [1] as well as the interest to the present work. He is also grateful to Professors Piotr Graczyk, Bartosz Kołodziejek, Yoshihiko Konno, and Satoshi Kuriki for the collaboration about applications of homogeneous cones to mathematical statistics. The discussions with them yield most of the featured ideas and techniques in this paper. Finally, the author sincerely thanks the referees for their valuable comments and suggestions.

2 Structure of the convex parameter set

In what follows, we assume that $\mathcal{M}_G$ is a convex set in the vector space $\mathrm{Sym}(N,\mathbb{R})$ and that $\mathcal{M}_G \neq \{I_N\}$. The affine subspace spanned by elements of $\mathcal{M}_G$ is of the form $I_N + V$, where V is a linear subspace of $\mathrm{Sym}(N,\mathbb{R})$, and $\mathcal{M}_G$ is an open connected set in $I_N + V$. Let $\overline{G}$ be the closure of G in $GL(N,\mathbb{R})$. Then we have $\mathcal{M}_{\overline{G}} = \mathcal{M}_G$. Indeed, for $\tilde{g} \in \overline{G}$, the set $\tilde{g}\mathcal{M}_G\tilde{g}^\top = \left\{\, \tilde{g}\theta\tilde{g}^\top\,;\, \theta \in \mathcal{M}_G \,\right\}$ is contained in the closure $\overline{\mathcal{M}_G}$ of $\mathcal{M}_G$. On the other hand, since $\tilde{g}$ is invertible, the set $\tilde{g}\mathcal{M}_G\tilde{g}^\top$ is open in the affine space $I_N + V$. Thus the point $\tilde{g}\tilde{g}^\top \in \tilde{g}\mathcal{M}_G\tilde{g}^\top \subset \overline{\mathcal{M}_G}$ is an relatively interior point of the convex set $\overline{\mathcal{M}_G}$, so that $\tilde{g}\tilde{g}^\top \in \mathcal{M}_G$, which means that $\mathcal{M}_{\overline{G}} \subset \mathcal{M}_G$. Therefore we can assume that G is a closed linear Lie group without loss of generality. Furthermore, we can assume that G is connected. Following Vinberg's argument [9, Chapter 1, Section 6], we shall show that $\mathcal{M}_{G^0_{\mathrm{alg}}} = \mathcal{M}_G$, where G^0_{alg} is the identity component (in the classical topology) of the real algebraic hull of G in $GL(N,\mathbb{R})$. Let $\mathfrak{g}$ be the Lie algebra of G. For $h \in G^0_{\mathrm{alg}}$, we have $hh^\top \in I_N + V$ and $h\,\mathfrak{g}\,h^{-1} = \mathfrak{g}$ because these are algebraic conditions that are satisfied by all the elements of G. The second condition together with the connectedness of G tells us that $hGh^{-1} = G$. Let $\mathcal{U}$ be a neighborhood of I_N in G^0_{alg} such that $hh^\top \in \mathcal{M}_G$ for all $h \in \mathcal{U}$. Then for any $\theta = gg^\top \in \mathcal{M}_G$ $(g \in G)$, we have $h\theta h^\top = (hgh^{-1})hh^\top(hgh^{-1})^\top \in g_0\mathcal{M}_G g_0^\top = \mathcal{M}_G$, where $g_0 := hgh^{-1} \in G$. Thus, if h is the product $h_1h_2\cdots h_m \in G^0_{\mathrm{alg}}$ with $h_1,\ldots,h_m \in \mathcal{U}$, we see that $hh^\top \in \mathcal{M}_G$ inductively. Therefore we conclude that $\mathcal{M}_{G^0_{\mathrm{alg}}} = \mathcal{M}_G$.

By Vinberg [9], we have the generalized Iwasawa decomposition

$$G^0_{\mathrm{alg}} = \mathcal{T} \cdot (G^0_{\mathrm{alg}} \cap O(N)),$$

where $\mathcal{T}$ is a maximal connected split solvable Lie subgroup of G^0_{alg}. Moreover, $\mathcal{T}$ is triangularized simultaneously by an orthogonal matrix $U \in O(N)$, which means that a group $\mathcal{T}^U = \left\{ U^{-1}hU \,;\, h \in \mathcal{T} \right\}$ is contained in the group $\mathcal{T}_N$ of lower triangular matrices of size N with positive diagonal entries. Let $\mathcal{M}^U_G$ be the set $\left\{ U^\top \theta U \,;\, \theta \in \mathcal{M}_G \right\}$, which is equal to $\mathcal{M}_{\mathcal{T}^U}$. By the uniqueness of the Cholesky decomposition, we have a bijection $\mathcal{T}^U \ni h \mapsto hh^\top \in \mathcal{M}^U_G$. The tangent space of $\mathcal{M}^U_G$ at I_N is naturally identified with $V^U := \left\{ U^\top y U \,;\, y \in V \right\}$.

In general, for $x \in \mathrm{Sym}(N, \mathbb{R})$, we denote by $\underset{\vee}{x}$ the lower triangular matrix for which $x = \underset{\vee}{x} + (\underset{\vee}{x})^\top$. Actually, we have $(\underset{\vee}{x})_{ij} = \begin{cases} x_{ij} & (i > j) \\ x_{ii}/2 & (i = j) \\ 0 & (i < j). \end{cases}$ Then we define a bilinear product $\triangle$ on $\mathrm{Sym}(N, \mathbb{R})$ by

$$x \triangle y := \underset{\vee}{x} y + y (\underset{\vee}{x})^\top \qquad (x, y \in \mathrm{Sym}(N, \mathbb{R})).$$

The algebra $(\mathrm{Sym}(N, \mathbb{R}), \triangle)$ forms a <u>c</u>ompact <u>n</u>ormal <u>l</u>eft-symmetric <u>a</u>lgebra (CLAN), see [9, Chapter 2] and [6].

Lemma 1. *The space V^U is a subalgebral of $(\mathrm{Sym}(N, \mathbb{R}), \triangle)$. Namely, for any $x, y \in V^U$, one has $x \triangle y \in V^U$.*

Proof. Let $\mathfrak{t}^U$ be the Lie algebra of $\mathcal{T}^U$. In view of the Cholesky decomposition mentioned above, we have a linear isomorphism $\mathfrak{t}^U \ni T \mapsto T + T^\top \in V^U$. Thus we obtain

$$\mathfrak{t}^U = \left\{ \underset{\vee}{x} \,;\, x \in V^U \right\}. \tag{1}$$

Let us consider the action of $h \in \mathcal{T}^U$ on the set $\mathcal{M}^U_G$ given by $\mathcal{M}^U_G \ni \theta \mapsto h\theta h^\top \in \mathcal{M}^U_G$. This action is naturally extended to the affine space $I_N + V^U$ as affine transformations. The infinitesimal action of $T = \underset{\vee}{x} \in \mathfrak{t}^U$ on $I_N + y \in I_N + V^U$ $(y \in V^U)$ is equal to $T(I_N + y) + (I_N + y)T^\top = x + x \triangle y$ which must be an element of V^U. Therefore $x \triangle y \in V^U$. $\square$

Proposition 1 ([6, Theorem 2 and Proposition 2]). *If V^U contains the identity matrix I_N, after an appropriate permutation of the rows and columns, V^U becomes the set of symmetric matrices of the form*

$$y = \begin{pmatrix} Y_{11} & Y_{21}^\top & \dots & Y_{r1}^\top \\ Y_{21} & Y_{22} & & Y_{r2}^\top \\ \vdots & & \ddots & \vdots \\ Y_{r1} & Y_{r2} & \dots & Y_{rr} \end{pmatrix} \quad \begin{pmatrix} Y_{kk} = y_{kk} I_{\nu_k}, \ y_{kk} \in \mathbb{R}, \quad (k = 1, \dots, r) \\ Y_{lk} \in V_{lk} \quad (1 \le k < l \le r) \end{pmatrix},$$

where $N = \nu_1 + \cdots + \nu_r$, and V_{lk} are subspaces of $\mathrm{Mat}(\nu_l, \nu_k; \mathbb{R})$ satisfying
(V1) $A \in V_{lk} \Rightarrow AA^\top \in \mathbb{R} I_{\nu_l}$ *for $1 \le k < l \le r$,*

(V2) $A \in V_{lj}, B \in V_{kj} \Rightarrow AB^\top \in V_{lk}$ *for* $1 \le j < k < l \le r$,
(V3) $A \in V_{lk}, B \in V_{kj} \Rightarrow AB \in V_{lj}$ *for* $1 \le j < k < l \le r$.

Clearly $I_N \in V^U$ if and only if $I_N \in V$. In this case, Proposition 1 together with [6, Theorem 3] tells us that $\mathcal{T}^U = \mathfrak{t}^U \cap \mathcal{T}_N$ and $\mathcal{M}_G^U = \mathcal{M}_{\mathcal{T}^U} = V^U \cap \mathrm{Sym}^+(N, \mathbb{R})$. It follows that we obtain:

Proposition 2. *If* $I_N \in V$*, one has* $\mathcal{M}_G = V \cap \mathrm{Sym}^+(N, \mathbb{R})$.

Here we remark that every homogeneous cone is realized as $\mathcal{M}_G$ this way by [6]. For example, if $U = I_N$, $\nu_1 = \cdots = \nu_r = 2$ and $V_{lk} = \left\{ \begin{pmatrix} a & -b \\ b & a \end{pmatrix} ; a, b \in \mathbb{R} \right\}$ $(1 \le k < l \le r)$, then $\mathcal{M}_G = V \cap \mathrm{Sym}^+(N, \mathbb{R})$ is linearly isomorphic to the cone $\mathrm{Herm}^+(r, \mathbb{C})$ of positive definite $r \times r$ Hermitian matrices, which is a realization of the symmetric space $GL(r, \mathbb{C})/U(r)$. See [4] and [6] for other examples.

If V^U does not contain I_N, we consider the direct sum $\tilde{V}^U := \mathbb{R}I_N \oplus V^U$, which is also a subalgebra of $(\mathrm{Sym}(N, \mathbb{R}), \triangle)$. Then we apply Proposition 1 to $\tilde{V}^U$. Since V^U is a two-sided ideal of $\tilde{V}^U$ of codimension one, after some renumbering of indices k and l, we see that V^U equals the subspace of $\tilde{V}^U$ costing of elements y with $y_{11} = 0$. Namely, we have the following.

Theorem 1. *If* V^U *does not contain the identity matrix* I_N*, after an appropriate permutation of the rows and columns,* V^U *becomes the set of symmetric matrices of the form*

$$y = \begin{pmatrix} 0 & Y_{21}^\top & \dots & Y_{r1}^\top \\ Y_{21} & Y_{22} & & Y_{r2}^\top \\ \vdots & & \ddots & \vdots \\ Y_{r1} & Y_{r2} & \dots & Y_{rr} \end{pmatrix} \quad \begin{pmatrix} Y_{kk} = y_{kk} I_{\nu_k},\ y_{kk} \in \mathbb{R}, \quad k = 2, \dots, r \\ Y_{lk} \in V_{lk}, \quad 1 \le k < l \le r \end{pmatrix},$$

where V_{lk} *are the same as in Proposition 1.*

In what follows, we shall consider the case where $I_N \notin \mathcal{M}_G^U$ because our results below for the case $I_N \in \mathcal{M}_G^U$ will be obtained formally by putting $\nu_1 = 0$ and $V_{k1} = \{0\}$, as is understood by comparing Proposition 1 and Theorem 1. Put $N' := \nu_2 + \cdots + \nu_r = N - \nu_1$. Let W and V' be the vector spaces of matrices w and y' respectively of the forms

$$w = \begin{pmatrix} Y_{21} \\ \vdots \\ Y_{r1} \end{pmatrix} \in \mathrm{Mat}(N', \nu_1; \mathbb{R}), \quad y' = \begin{pmatrix} Y_{22} & & Y_{r2}^\top \\ \vdots & \ddots & \\ Y_{r2} & \dots & Y_{rr} \end{pmatrix} \in \mathrm{Sym}(N', \mathbb{R}).$$

Let $\mathcal{P}'$ be the set $V' \cap \mathrm{Sym}^+(N', \mathbb{R})$. Then $\mathcal{P}'$ forms a pointed open convex cone in the vector space V'.

Theorem 2. *Under the assumptions above, one has*

$$\mathcal{M}_G = \left\{ U \begin{pmatrix} I_{\nu_1} & w^\top \\ w & y' \end{pmatrix} U^\top ; w \in W,\ y' \in V',\ y' - ww^\top \in \mathcal{P}' \right\}.$$

Moreover $\mathcal{M}_G$ *equals the intersection of* $\mathrm{Sym}^+(N, \mathbb{R})$ *and the affine space* $I_N + V$.

Proof. Let $\mathfrak{t}'$ be the set of the lower triangular matrices $\underset{\vee}{y'}$ with $y' \in V'$. Then we see from [6] that $\mathfrak{t}'$ is a Lie algebra, and that the corresponding Lie group $\mathcal{T}' := \exp \mathfrak{t}' \subset GL(N', \mathbb{R})$ equals $\mathfrak{t}' \cap \mathcal{T}_{N'}$. Namely, $\mathcal{T}'$ is the set of h' of the form

$$h' = \begin{pmatrix} T_{22} & & \\ \vdots & \ddots & \\ T_{r2} & \dots & T_{rr} \end{pmatrix} \in \mathcal{T}_{N'} \quad \begin{pmatrix} T_{kk} = t_{kk} I_{\nu_k},\ t_{kk} > 0, \quad k = 2, \dots, r \\ T_{lk} \in V_{lk}, \quad 2 \le k < l \le r \end{pmatrix}.$$

Since we have $\mathfrak{t}^U = \left\{ \begin{pmatrix} 0 & \\ L & T' \end{pmatrix} ;\ L \in W,\ T' \in \mathfrak{t}' \right\}$ by (1), the corresponding Lie group $\mathcal{T}^U$ is the set of $\begin{pmatrix} I_{\nu_1} & 0 \\ L & h' \end{pmatrix}$ with $L \in W$ and $h' \in \mathcal{T}'$. Therefore, $\mathcal{M}_G^U = \mathcal{M}_{\mathcal{T}^U}$ is the set of matrices $\begin{pmatrix} I_{\nu_1} & 0 \\ L & h' \end{pmatrix} \begin{pmatrix} I_{\nu_1} & L^\top \\ 0 & (h')^\top \end{pmatrix} = \begin{pmatrix} I_{\nu_1} & L^\top \\ L & LL^\top + a' \end{pmatrix}$ with $a' = h'(h')^\top \in \mathcal{P}'$. On the other hand, as is discussed in [6, Theorem 3], the map $T' \ni h' \mapsto h'(h')^\top \in \mathcal{P}'$ is bijective, which completes the proof. □

3 Existence condition and an explicit expression of MLE

Let $X_1, \dots, X_n$ be independent random vectors obeying the central multivariate normal law $N(0, \Sigma)$ with $\theta := \Sigma^{-1} \in \mathcal{M}_G$. The density function f of $(X_1, \dots, X_n)$ is given by

$$f(x_1, \dots, x_n; \theta) := (2\pi)^{-nN/2} (\det \theta)^{n/2} \prod_{j=1}^{N} e^{-x_j^\top \theta x_j / 2} \quad (x_j \in \mathbb{R}^N,\ j = 1, \dots, n).$$

Let $\pi : \mathrm{Sym}(N, \mathbb{R}) \to V$ be the orthogonal projection with respect to the trace inner product. Putting $y := \pi(\sum_{j=1}^N x_j x_j^\top / 2) \in V$, we have $f(x_1, \dots, x_n; \theta) = c(x) (\det \theta)^{n/2} e^{-\mathrm{tr}\, \theta y}$, where $c(x) := (2\pi)^{-nN/2} \exp(\mathrm{tr}\, \{(I_N - \pi(I_N)) \sum_{j=1}^N x_j x_j^\top / 2\})$, which is independent of θ. Thus, given $Y := \pi(\sum_{k=1}^n X_j X_j^\top / 2) \in V$, the maximum likelihood estimator $\hat{\theta}$ is an element of $\mathcal{M}_G$ at which the log likelihood function $F(\theta; Y) := (n/2) \log \det \theta - \mathrm{tr}(Y\theta)$ attains the maximum value.

In what follows, we shall assume that $U = I_N$. Indeed, a general case is easily reduced to this case. For $k = 2, \dots, r$, let $V_{[k]}$ and W_{k-1} be the vector spaces of matrices $y_{[k]}$ and Z_{k-1} respectively of the forms

$$y_{[k]} = \begin{pmatrix} Y_{kk} & & Y_{rk}^\top \\ \vdots & \ddots & \\ Y_{rk} & \dots & Y_{rr} \end{pmatrix} \in \mathrm{Sym}(N_k, \mathbb{R}), \quad Z_{k-1} = \begin{pmatrix} Y_{k,k-1} \\ \vdots \\ Y_{r,k-1} \end{pmatrix} \in \mathrm{Mat}(N_k, \nu_{k-1}; \mathbb{R}),$$

where $N_k := \nu_k + \dots + \nu_r$. Note that $V' = V_{[2]}$ and $W = W_1$ in the previous notation. We have an inductive expression of $y \in V$ as

$$y = \begin{pmatrix} 0 & Z_1^\top \\ Z_1 & y_{[2]} \end{pmatrix}, \quad y_{[k]} = \begin{pmatrix} y_{kk} I_{\nu_k} & Z_k^\top \\ Z_k & y_{[k+1]} \end{pmatrix} \quad (k = 2, \dots r-1), \quad y_{[r]} = y_{rr} I_{\nu_r}. \quad (2)$$

Let $\mathcal{T}_{[k]} \subset \mathcal{T}_{N_k}$ be the group of lower triangular matrices $y_{[k]}$ $(y_{[k]} \in V_{[k]})$ with positive diagonal entries. Then $\mathcal{T}' = \mathcal{T}_{[2]}$. Any element $h \in \overset{\vee}{\mathcal{T}}$ is expressed as

$$h = \begin{pmatrix} I_{\nu_1} & 0 \\ L_1 & h_{[2]} \end{pmatrix}, \quad h_{[k]} = \begin{pmatrix} t_{kk} I_{\nu_k} & 0 \\ L_k & h_{[k+1]} \end{pmatrix} \in \mathcal{T}_{[k]} \quad (k = 2, \dots, r-1)$$

with $L_k \in W_k$ and $h_{[r]} = t_{rr} I_{\nu_r} \in \mathcal{T}_{[r]}$. We observe that

$$h_{[k]} h_{[k]}^\top = \begin{pmatrix} t_{kk}^2 I_{\nu_k} & t_{kk} L_k^\top \\ t_{kk} L_k & L_k L_k^\top + h_{[k+1]} h_{[k+1]}^\top \end{pmatrix} \in V_{[k]}.$$

We shall regard $V_{[k]}$ as a subspace of $V' = V_{[2]}$ by zero-extension. Define a map $q_k : \mathbb{R}_{>0} \times W_k \to V'$ for $k = 2, \dots, r-1$ by $q_k(t_{kk}, L_k) := \begin{pmatrix} t_{kk}^2 & t_{kk} L_k^\top \\ t_{kk} L_k & L_k L_k^\top \end{pmatrix} \in V_{[k]} \subset V'$, and define also $q_r(t_{rr}) := t_{rr}^2 I_{n_r} \in V_{[r]} \subset V'$. If $\theta = hh^\top$, we have

$$\theta = \begin{pmatrix} I_{\nu_1} & L_1 \\ L_1 & L_1 L_1^\top + \theta' \end{pmatrix}, \quad \theta' = \sum_{k=2}^{r-1} q_k(t_{kk}, L_k) + q_r(t_{rr}). \tag{3}$$

For $y' \in V' = V_{[2]}$ and $k = 2, \dots, r-1$, we have

$$\mathrm{tr}\,(y' q_k(t_{kk}, L_k)) = \nu_k y_{kk} t_{kk}^2 + 2 t_{kk}\, \mathrm{tr}\,(Z_k^\top L_k) + \mathrm{tr}\,(y_{[k+1]} L_k L_k^\top). \tag{4}$$

Let m_k $(k = 1, \dots, r-1)$ be the dimension of the vector space W_k, and take an orthonormal basis $\{e_{k\alpha}\}_{\alpha=1}^{m_k}$ of W_k with respect to the trace inner product. For $L_k \in W_k$, let $\lambda_k := (\lambda_{k1}, \dots, \lambda_{km_k})^\top \in \mathbb{R}^{m_k}$ be the column vector for which $L_k = \sum_{\alpha=1}^{m_k} \lambda_{k\alpha} e_{k\alpha}$. Defining $\zeta_k \in \mathbb{R}^{m_k}$ for $Z_k \in W_k$ similarly, we have $\zeta_k^\top \lambda_k = \mathrm{tr}\,(Z_k^\top L_K)$. Let $\psi_k : V' \to \mathrm{Sym}(m_k, \mathbb{R})$ be a linear map defined in such a way that $\mathrm{tr}\,(y' L_k L_k^\top) = \lambda_k^\top \psi_k(y') \lambda_k$, and define $\phi_k : V' \to \mathrm{Sym}(1 + m_k, \mathbb{R})$ by $\phi_k(y') := \begin{pmatrix} \nu_k y_{kk} & \zeta_k^\top \\ \zeta_k & \psi_k(y) \end{pmatrix}$. In view of (4), we have

$$\mathrm{tr}(y' q_k(t_{kk}, L_k)) = \mathrm{tr}\, \left(\phi_k(y') \begin{pmatrix} t_{kk}^2 & t_{kk} \lambda_k^\top \\ t_{kk} \lambda_k & \lambda_k \lambda_k^\top \end{pmatrix} \right). \tag{5}$$

Let $\phi_k^* : \mathrm{Sym}(1 + m_k, \mathbb{R}) \to V'$ be the adjoint map of ϕ_k, which means that $\mathrm{tr}\, \phi_k^*(S) y' = \mathrm{tr}\, S \phi(y')$ for $S \in \mathrm{Sym}(1 + m_k, \mathbb{R})$. Define also $\phi_r(y') := \nu_r y_{rr} \in \mathbb{R} \equiv \mathrm{Sym}(1, \mathbb{R})$ and $\phi_r^*(c) = c I_{\nu_r} \in V_{[r]} \subset V'$ for $c \in \mathbb{R}$.

Let $\mathcal{Q}' \subset V'$ be the dual cone of $\mathcal{P}' \subset V'$, that is, the set of $y' \in V'$ such that $\mathrm{tr}\,(y' a) > 0$ for all $a \in \overline{\mathrm{P}'} \setminus \{0\}$. If $y' \in \mathcal{Q}'$, then $\phi_k(y')$ and $\psi_{k-1}(y')$ are positive definite for $k = 2, \dots, r$. Moreover, it is known (see [4, Proposition 3.4 (iii)]) that

$$\mathcal{Q}' = \{\, y' \in V' \,;\, \det \phi_k(y') > 0 \text{ for all } k = 2, \dots, r \,\}. \tag{6}$$

If $Y := \pi(\sum_{j=1}^n X_j X_j^\top / 2)$ is expressed as in (2), then we can show that $Y' := Y_{[2]}$ belongs to the closure of $\mathcal{Q}'$, so that $\phi_k(Y')$ and $\psi_{k-1}(Y')$ are positive semidefinite for $k = 2, \dots r$.

Theorem 3. (i) *If $Y' \in \mathcal{Q}'$, then $\hat\theta = \arg\max_{\theta\in\mathcal{M}_G} F(\theta;Y)$ exists uniquely, and it is expressed as $\hat\theta = \begin{pmatrix} I_{\nu_1} & \hat L_1^\top \\ \hat L_1 & \hat L_1 \hat L_1^\top + \hat\theta' \end{pmatrix}$ with $\hat\lambda_1 = -\psi_1(Y')^{-1}\zeta_1 \in \mathbb{R}^{m_1}$, which is the column vector corresponding to $\hat L_1$, and*

$$\hat\theta' = \frac{n\nu_r}{2}\phi_r^*(\phi_r(Y')^{-1}) + \sum_{k=2}^{r-1}\frac{n\nu_k}{2}\phi_k^*\left(\phi_k(Y')^{-1} - \begin{pmatrix} 0 & 0 \\ 0 & \psi_k(Y')^{-1}\end{pmatrix}\right). \tag{7}$$

(ii) *If $Y' \notin \mathcal{Q}'$, then $F(\theta;Y)$ is unbounded, so that $\hat\theta$ does not exist.*

Proof. (i) Keeping (2), (3) and $(\det\theta)^{n/2} = \prod_{k=2}^r (t_{kk})^{n\nu_k}$ in mind, we define

$$\begin{aligned} F_1(L_1;Y) &:= -2\mathrm{tr}(Z_1^\top L_1) - \mathrm{tr}(Y' L_1 L_1^\top), \\ F_k(t_{kk}, L_k; Y) &:= n\nu_k \log t_{kk} - \mathrm{tr}\,(q_k(t_{kk}, L_k)Y') \quad (k = 2,\dots,r-1), \\ F_r(t_{rr};Y) &:= n\nu_r \log t_{rr} - \nu_r y_{rr} t_{rr}^2, \end{aligned}$$

so that $F(\theta;Y) = F_1(L_1;Y) + F_r(t_{rr};Y) + \sum_{k=2}^{r-1} F_k(t_{kk}, L_k;Y)$. It is easy to see that $F_r(t_{rr};Y)$ takes a maximum value at $\hat t_{rr} = \sqrt{\frac{n}{2y_{rr}}}$. Then $q_r(\hat t_{rr}) = \frac{n}{2y_{rr}} I_{\nu_r} = \frac{n\nu_r}{2}\phi_r^*(\phi_r(Y')^{-1})$. On the other hand, $F_1(L_1;Y) = -2\zeta_1^\top\lambda_1 - \lambda_1^\top\psi_1(Y')\lambda_1$ equals

$$\zeta_1^\top\psi_1(Y')^{-1}\zeta_1 - (\lambda_1 + \psi_1(Y')^{-1}\zeta_1)^\top\psi_1(Y')(\lambda_1 + \psi_1(Y')^{-1}\zeta_1),$$

which attains a maximum value when $\lambda_1 = -\psi_1(Y')^{-1}\zeta_1$ because $\psi_1(Y')$ is positive definite. Similarly, we see from (5) that

$$\begin{aligned} F_k(t_{kk}, L_k) =& n\nu_k \log t_{kk} - (\nu_k y_{kk} - \zeta_k^\top \psi_k(Y')^{-1}\zeta_k)t_{kk}^2 \\ &- (\lambda_k + t_{kk}\psi_k(Y')^{-1}\zeta_k)^\top \psi_k(Y')(\lambda_k + t_{kk}\psi_k(Y')^{-1}\zeta_k). \end{aligned} \tag{8}$$

Since $\nu_k y_{kk} - \zeta_k^\top\psi_k(Y')^{-1}\zeta_k = \det\phi_k(Y')/\det\psi_k(Y') > 0$, we see that $F_k(t_{kk}, L_k; Y')$ attains at $(\hat t_{kk}, \hat L_k)$ with $\hat t_{kk} = \sqrt{\frac{n\nu_k}{2(\nu_k y_{kk} - \zeta_k^\top\psi_k(Y')^{-1}\zeta_k)}}$ and $\hat\lambda_k = -\hat t_{kk}\psi_k(Y')^{-1}\zeta_k \in \mathbb{R}^{m_k}$, which is the column vector corresponding to $\hat L_k$. By a straightforward calculation, we have

$$\begin{pmatrix} \hat t_{kk}^2 & \hat t_{kk}\hat\lambda_k^\top \\ \hat t_{kk}\hat\lambda_k & \hat\lambda_k\hat\lambda_k^\top \end{pmatrix} = \frac{n\nu_k}{2}\left(\phi_k(Y')^{-1} - \begin{pmatrix} 0 & 0 \\ 0 & \psi_k(Y')^{-1}\end{pmatrix}\right) \in \mathrm{Sym}(1+m_k, \mathbb{R}),$$

which maps to $q_k(\hat t_{kk}, \hat L_k)$ by $\phi_k^* : \mathrm{Sym}(1+m_k,\mathbb{R}) \to V'$ thanks to (5). Therefore the assertion (i) is verified.

(ii) By (6), there exists k for which $\det\phi_k(Y') = 0$. If $\det\phi_r(Y') = y_{rr}^{\nu_r} = 0$, then $F_r(t_{rr};Y') = n\nu_r \log t_{rr} \to +\infty$ in $t_{rr} \to +\infty$. Let us consider the case where $\det\phi_l(Y') > 0$ for $l = k+1,\dots,r$ and $\det\phi_k(Y') = 0$. Then one can show that $\psi_k(Y')$ is positive definite because in this case $Y_{[k+1]} \in V_{[k+1]}$ belongs to the dual cone of $\mathcal{P}_{[k+1]} := V_{[k+1]} \cap \mathrm{Sym}^+(N_k, \mathbb{R})$ by the same reason as (6). Since $\nu_k y_{kk} - \zeta_k^\top \psi_k(Y')^{-1}\zeta_k = \det \phi_k(Y')/\det \psi_k(Y') = 0$, we see from (8) that

$$\max_{L_k \in W_k} F_k(t_{kk}, L_k; Y') = n\nu_k \log t_{kk} \to +\infty \quad (t_{kk} \to +\infty),$$

which completes the proof. □

We remark that a generalization of the formula (7) is found in [5, Theorem 5.1]. By Theorem 3, the existence of the MLE $\hat{\theta}$ is equivalent to that the random matrix Y' belongs to the cone $\mathcal{Q}'$ with probability one. On the other hand, the distribution of Y' is nothing else but the Wishart distribution on $\mathcal{Q}'$ studied in [4] and [5]. The Wishart distribution is obtained as a natural exponential family generated by the Riesz distribution μ_n on V' characterized by its Laplace transform: $\int_{V'} e^{-\mathrm{tr}(y'a)}\mu_n(dy') = (\det a)^{-n/2}$ for all $a \in \mathcal{P}'$. Note that, if $a \in \mathcal{P}'$ is a diagonal matrix, we have $(\det a)^{-n/2} = \prod_{k=2}^r a_{kk}^{-n\nu_k/2}$. As is seen in [5, Theorem 4.1 (ii)], the support of the Riesz distribution μ_n is determined from the parameter $(n\nu_2/2, \dots, n\nu_r/2) \in \mathbb{R}^{r-1}$. In fact, $\mathrm{supp}\,\mu_n = \overline{\mathcal{Q}'}$ if and only if $n\nu_k/2 > m_k/2$ for $k = 2, \dots, r$, and $Y' \in \mathcal{Q}'$ almost surely in this case. Otherwise, $\mathrm{supp}\,\mu_n$ is contained in the boundary of $\mathcal{Q}'$, so that Y' never belongs to $\mathcal{Q}'$. Therefore, if n_0 is the smallest integer that is greater than $\max\left\{ \frac{m_k}{\nu_k} ;\, k = 2, \dots, r \right\}$, we have the following final result.

Theorem 4. *The maximum likelihood estimator* $\hat{\theta} = \arg\max_{\theta \in \mathcal{M}_G} F(\theta; Y)$ *exists with probability one if and only if* $n \geq n_0$. *If* $n < n_0$, *then the log likelihood function* $F(\theta; Y)$ *of* $\theta \in \mathcal{M}_G$ *is unbounded.*

References

1. C. Améndola, K. Kohn, P. Reichenbach and A. Seigal, *Invariant theory and scaling algorithms for maximum likelihood estimation*, arXiv:2003.13662, 2020.
2. O. Barndorff-Nielsen, P. Blaesild, J. Ledet Jensen, and B. Jørgensen, *Exponential transformation models*, Proc. Roy. Soc. London Ser. A **379** (1982), no. 1776, 41–65.
3. J. Faraut and A. Korányi, "Analysis on symmetric cones", Clarendon Press, 1994.
4. P. Graczyk and H. Ishi, *Riesz measures and Wishart laws associated to quadratic maps*, J. Math. Soc. Japan **66** (2014), 317–348.
5. P. Graczyk, H. Ishi and B. Kołodziejek, *Wishart laws and variance function on homogeneous cones*, Probab. Math. Statist. **39** (2019), 337–360.
6. H. Ishi, *Matrix realization of a homogeneous cone*, Lecture Notes in Comput. Sci. **9389**, pp. 248–256, Springer, 2015.
7. H. Ishi, *Explicit formula of Koszul-Vinberg characteristic functions for a wide class of regular convex cones*, Entropy **18** (2016), Issue 11, 383; doi:10.3390/e18110383.
8. G. Letac and H. Massam, *Wishart distributions for decomposable graphs*, Ann. Statist. **35** (2007), 1278–1323.
9. E. B. Vinberg, *The theory of convex homogeneous cones*, Trans. Mosc. Math. Soc. **12** (1963), 340–403.