Non-Steepness and Maximum Likelihood Estimation Properties of the Truncated Multivariate Normal Distributions

Michael Levine,* Donald Richards,† and Jianxi Su‡

January 29, 2023

Abstract

We consider the truncated multivariate normal distributions for which every component is one-sided truncated. We show that this family of distributions is an exponential family. We identify $\mathcal{D}$, the corresponding natural parameter space, and deduce that the family of distributions is not regular. We prove that the gradient of the cumulant-generating function of the family of distributions remains bounded near certain boundary points in $\mathcal{D}$, and therefore the family also is not steep. We also consider maximum likelihood estimation for $\boldsymbol{\mu}$, the location vector parameter, and $\boldsymbol{\Sigma}$, the positive definite (symmetric) matrix dispersion parameter, of a truncated non-singular multivariate normal distribution. We prove that each solution to the score equations for $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ satisfies the method-of-moments equations, and we obtain a necessary condition for the existence of solutions to the score equations.

Key words and phrases. Cumulant-generating function, Laplace transform, multivariate normal distribution, regular exponential family, truncated distribution, steep exponential family.

2020 Mathematics Subject Classification. Primary: 62H05, 62H12. Secondary: 62E10.

Running head: Truncated multivariate normal distributions.

1 Introduction

In this article, we continue our investigations of the one-sided truncated multivariate normal distributions, a family of distributions which has been applied in numer-

*Department of Statistics, Purdue University, West Lafayette, IN 47907, U.S.A.

†Department of Statistics, Penn State University, University Park, PA 16802, U.S.A.

‡Department of Statistics, Purdue University, West Lafayette, IN 47907, U.S.A.

*Corresponding author; e-mail address: mlevins@purdue.edu

ous areas, including simultaneous equations inference and multivariate regression [1], economics and econometrics [9, 10], educational studies [12, 14], and inference with censored biomedical data [13, 15]. Here we are motivated, by results of Cohen [6], del Castillo [7], and others, to study maximum likelihood estimation for these distributions.

It is well-known that most exponential families appearing in practice are regular and hence are steep. For instance, the two-sided truncated univariate normal distributions, with support restricted to a finite interval, form a regular, and hence steep, exponential family (Barndorff-Nielsen [4, p. 161]). There are also several exponential families which are steep but non-regular, an example being the family of inverse Gaussian distributions; see Bar-Lev and Enis [3, p. 1516], Barndorff-Nielsen [4, p. 117], Brown [5, p. 72]. However, there are fewer known examples of non-steep exponential families.

In the univariate case, del Castillo [7] formulated the one-sided truncated normal distributions as an exponential family of distributions and proved that the family is not steep and therefore is not regular. That this family of distributions is not steep implies that the maximum likelihood estimators of the parameters of such a distribution possibly may not exist, and therefore del Castillo's results have deeper implications for inference for the parameters of the one-sided truncated normal distributions.

Although nearly three decades have elapsed since the appearance of del Castillo's results, there has been no subsequent investigation of the multivariate truncated normal distributions; it is this observation which initially motivated our research. We consider in this article the family, $\mathcal{P}_{\mathcal{D}}$, of truncated multivariate d-dimensional normal distributions such that every univariate component is one-sided truncated. Our first main result, given in Section 2 is that the family $\mathcal{P}_{\mathcal{D}}$ is a multivariate exponential family. We prove that the natural parameter space, $\mathcal{D}$, consists of a union of $d+1$ subsets, all but one of which are not open; as a consequence, we deduce that $\mathcal{D}$ itself is not open and therefore $\mathcal{P}_{\mathcal{D}}$ is not regular.

In Section 3, we prove that the family $\mathcal{P}_{\mathcal{D}}$ is non-steep. As the marginal distributions of the truncated multivariate normal distributions generally are not truncated normal [14], the non-steepness property in the multivariate case does not follow from the univariate case. By analyzing the cumulant-generating function of the family $\mathcal{P}_{\mathcal{D}}$, we prove that the gradient of that function remains bounded near certain boundary points in $\mathcal{D}$, hence we deduce that $\mathcal{P}_{\mathcal{D}}$ is not steep.

We consider in Section 4 the problem of maximum likelihood estimation for $\boldsymbol{\mu}$, the location vector parameter, and $\boldsymbol{\Sigma}$, the positive definite (symmetric) matrix dispersion parameter, of a truncated non-singular multivariate normal distribution. We prove that the solutions to the score equations also satisfy the method-of-moments equations, thereby generalizing results of various authors in the univariate, bivariate, and trivariate cases (see Cohen [6, Sections 2.3 and 12.2]). As a consequence of these results we obtain a necessary condition for the existence of solutions to the score equations for $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, generalizing a result of del Castillo [7, Theorem 4.2].

2 The non-regularity of the exponential family $\mathcal{P}_{\mathcal{D}}$

The following classical concepts in the theory of exponential families are treated in detail by Barndorff-Nielsen [4], Brown [5], Sundberg [16], and other authors. For a positive integer k, let $\mathcal{X} \subseteq \mathbb{R}^k$ be a measurable space. Also let β be a sigma-finite, positive measure on $\mathcal{X}$ such that β is not concentrated on an affine subspace. Let $T : \mathcal{X} \to \mathbb{R}^k$ be a statistic, and denote by β_T the marginal distribution of T. For a $k \times 1$ column vector $\boldsymbol{t} = (t_1, \ldots, t_k)' \in \mathbb{R}^k$ and a parameter $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_k)' \in \mathbb{R}^k$, define the Laplace transform,

$$L_T(\boldsymbol{\omega}) = \int_{\mathcal{X}} \exp(\boldsymbol{\omega}' T)\,\mathrm{d}\beta(T) = \int_{\mathbb{R}^k} \exp(\boldsymbol{\omega}'\boldsymbol{t})\,\mathrm{d}\beta_T(\boldsymbol{t}). \tag{2.1}$$

The set $\mathcal{D} = \{\boldsymbol{\omega} \in \mathbb{R}^k : L_T(\boldsymbol{\omega}) < \infty\}$ is called the *natural parameter space*. By applying Hölder's inequality, we find that $\mathcal{D}$ is convex. Further, define the *cumulant-generating function*, $K_T(\boldsymbol{\omega}) = \log L_T(\boldsymbol{\omega})$, $\boldsymbol{\omega} \in \mathcal{D}$; then K_T is strictly convex whenever it is non-degenerate.

For each $\boldsymbol{\omega} \in \mathcal{D}$, define $p(\boldsymbol{t};\boldsymbol{\omega}) = \exp\big(\boldsymbol{\omega}' T(\boldsymbol{t}) - K_T(\boldsymbol{\omega})\big)$, $\boldsymbol{t} \in \mathbb{R}^k$; then the measure $p(\boldsymbol{t};\boldsymbol{\omega})\,\mathrm{d}\beta(\boldsymbol{t})$, $\boldsymbol{t} \in \mathbb{R}^k$, is a probability distribution. The family of distributions $\mathcal{P}_{\mathcal{D}} = \{p(\boldsymbol{t};\boldsymbol{\omega})\,\mathrm{d}\beta(\boldsymbol{t}), \boldsymbol{t} \in \mathbb{R}^k : \boldsymbol{\omega} \in \mathcal{D}\}$ is an *exponential family*, and the family is called *full* if the parameter $\boldsymbol{\omega}$ is allowed to vary over the entire space $\mathcal{D}$ [5, p. 2].

The family $\mathcal{P}_{\mathcal{D}}$ is said to be *regular* if $\mathcal{D}$ is an open set. If $\mathcal{D}$ is not open then denote by Ω the set $\mathrm{Int}(\mathcal{D})$, the interior of $\mathcal{D}$. We consider only the non-trivial case in which Ω is non-empty, and then we define $\mathcal{P}_{\Omega} = \{p(\boldsymbol{t};\boldsymbol{\omega})\,\mathrm{d}\beta(\boldsymbol{t}), \boldsymbol{t} \in \mathbb{R}^k : \boldsymbol{\omega} \in \Omega\}$.

Throughout, let $\|\cdot\|$ denote the Euclidean norm on $\mathbb{R}^k$. Let $\partial\Omega$ denote the boundary of Ω, and let $\nabla_{\boldsymbol{\omega}} = (\partial/\partial\omega_1, \ldots, \partial/\partial\omega_k)'$ be the gradient operator on $\mathbb{R}^k$. Then the family $\mathcal{P}_{\mathcal{D}}$ is said to be *steep at the point* $\boldsymbol{\omega}_0 \in \partial\Omega$ if

$$\|\nabla_{\boldsymbol{\omega}} K_T(\boldsymbol{\omega})\| \to \infty \quad \text{as} \quad \boldsymbol{\omega} \to \boldsymbol{\omega}_0. \tag{2.2}$$

Further, the family $\mathcal{P}_{\mathcal{D}}$ is called *steep* if it is steep at every $\boldsymbol{\omega}_0 \in \partial\Omega$ [4, p. 86], [5, p. 71].

The exponential family considered in the present article arises from a random vector $\boldsymbol{X} = (X_1, \ldots, X_d)' \in \mathbb{R}^d$, having a truncated multivariate normal distribution and such that every component of $\boldsymbol{X}$ is singly-truncated, i.e., truncated below or above, at a finite point. By applying sign changes to the components, if necessary, we may assume that each X_j is truncated below at a given value a_j, $j = 1, \ldots, d$. Also, by the change-of-variables $x_j \to x_j + a_j$, $j = 1, \ldots, d$, we may assume, without loss of generality, that $a_j = 0$ for all $j = 1, \ldots, d$, so that the distribution of $\boldsymbol{X}$ is supported on $\mathbb{R}_+^d$, the positive orthant. Here, we are using the property that the class of truncated normal distributions is invariant under changes of sign or location [14].

In the sequel, we write $\boldsymbol{x} > \boldsymbol{0}$ to mean that $\boldsymbol{x} \in \mathbb{R}_+^d$ and write $\boldsymbol{x} < \boldsymbol{0}$ whenever $-\boldsymbol{x} > \boldsymbol{0}$; more generally, we write $\boldsymbol{x} > \boldsymbol{x}_0$ if $\boldsymbol{x} - \boldsymbol{x}_0 > \boldsymbol{0}$.

For a location parameter $\boldsymbol{\mu} \in \mathbb{R}^d$, and a positive definite (symmetric) $d \times d$ matrix dispersion parameter $\boldsymbol{\Sigma}$, let $\boldsymbol{X} = (X_1, \ldots, X_d)' \in \mathbb{R}^d$ be a random vector having the probability density function

$$f(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = [C(\boldsymbol{\mu}, \boldsymbol{\Sigma})]^{-1} \exp\big[-\tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\big], \tag{2.3}$$

$\boldsymbol{x} > \boldsymbol{0}$, where

$$C(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \int_{\boldsymbol{w}>\boldsymbol{0}} \exp\big[-\tfrac{1}{2}(\boldsymbol{w}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{w}-\boldsymbol{\mu})\big]\, \mathrm{d}\boldsymbol{w} \tag{2.4}$$

is the corresponding normalizing constant.

We now formulate the truncated multivariate normal distributions as an exponential family. Starting with (2.1), let $\mathcal{X} = \mathbb{R}_+^d$ and let β be the Lebesgue measure on $\mathbb{R}_+^d$, written $\mathrm{d}\beta(\boldsymbol{t}) = \mathrm{d}\boldsymbol{t}$. Choose as sufficient statistics the collection

$$T(\boldsymbol{t}) = \{t_i : i = 1, \ldots, d\} \cup \{t_i t_j : 1 \le i \le j \le d\}; \tag{2.5}$$

then T is a statistic of dimension $k = d + \frac{1}{2}d(d+1)$. The Laplace transform, L_T, of the resulting distribution β_T is the Laplace transform, with respect to $\mathrm{d}\boldsymbol{t}$, of the set of statistics $T(\boldsymbol{t})$.

To determine L_T explicitly, we denote by $\mathcal{S}^{d\times d}$ the vector space of $d \times d$ symmetric matrices and observe that each linear combination of the sufficient statistics in T is of the form $\boldsymbol{\theta}'\boldsymbol{t} - \boldsymbol{t}'\boldsymbol{\Theta}\boldsymbol{t}$ for some $\boldsymbol{\theta} \in \mathbb{R}^d$ and $\boldsymbol{\Theta} \in \mathcal{S}^{d\times d}$. Therefore L_T is parametrized by the pairs $\boldsymbol{\omega} = (\boldsymbol{\theta}, \boldsymbol{\Theta}) \in \mathbb{R}^d \times \mathcal{S}^{d\times d}$, the Laplace transform of β_T is

$$L_T(\boldsymbol{\theta}, \boldsymbol{\Theta}) = \int_{\boldsymbol{w}>\boldsymbol{0}} \exp(\boldsymbol{\theta}'\boldsymbol{w} - \boldsymbol{w}'\boldsymbol{\Theta}\boldsymbol{w})\, \mathrm{d}\boldsymbol{w}, \tag{2.6}$$

whenever the integral converges, and then we define the cumulant-generating function

$$K_T(\boldsymbol{\theta}, \boldsymbol{\Theta}) = \log L_T(\boldsymbol{\theta}, \boldsymbol{\Theta}) = \log \int_{\boldsymbol{w}>\boldsymbol{0}} \exp(\boldsymbol{\theta}'\boldsymbol{w} - \boldsymbol{w}'\boldsymbol{\Theta}\boldsymbol{w})\, \mathrm{d}\boldsymbol{w}. \tag{2.7}$$

To determine the corresponding exponential family $\mathcal{P}_{\mathcal{D}}$, we need to identify the related natural parameter space $\mathcal{D} = \{(\boldsymbol{\theta}, \boldsymbol{\Theta}) \in \mathbb{R}^d \times \mathcal{S}^{d\times d} : L_T(\boldsymbol{\theta}, \boldsymbol{\Theta}) < \infty\}$. Denote by $\boldsymbol{0}$ any zero vector or matrix, as determined by the context.

For $0 \le r \le d$, we denote by $\mathcal{S}_r^{d\times d}$ the set of $d \times d$ positive semidefinite (symmetric) matrices of rank r. Thus, for $r \le d-1$ the matrices in $\mathcal{S}_r^{d\times d}$ are singular, and the matrices in $\mathcal{S}_d^{d\times d}$ are positive definite and hence non-singular.

Theorem 2.1. *The truncated multivariate normal distributions define a full exponential family, $\mathcal{P}_{\mathcal{D}}$, with natural parameter space*

$$\mathcal{D} = \bigcup_{r=0}^{d} \Omega_r,$$

where the parameter spaces Ω_r are defined in (2.9), (2.11), and (2.12). Moreover, the family $\mathcal{P}_{\mathcal{D}}$ is not regular and, for each $(\boldsymbol{\theta}, \boldsymbol{\Theta}) \in \mathcal{D}$, the corresponding probability density function is

$$p(\boldsymbol{t}; \boldsymbol{\theta}, \boldsymbol{\Theta}) = \exp\big(\boldsymbol{\theta}'\boldsymbol{t} - \boldsymbol{t}'\boldsymbol{\Theta}\boldsymbol{t} - K_T(\boldsymbol{\theta}, \boldsymbol{\Theta})\big), \tag{2.8}$$

$\boldsymbol{t} \in \mathbb{R}^d_+$.

Proof. Suppose that $\boldsymbol{\Theta} \in \mathcal{S}_0^{d\times d}$, i.e., $\boldsymbol{\Theta} = \mathbf{0}$. Then the integral (2.6) converges if and only if $\boldsymbol{\theta} < \mathbf{0}$. Therefore, on defining

$$\Omega_0 := \{(\boldsymbol{\theta}, \boldsymbol{\Theta}) \in \mathbb{R}^d \times \mathcal{S}_0^{d\times d} : \boldsymbol{\theta} < \mathbf{0}\} \equiv -\mathbb{R}^d_+ \times \{\mathbf{0}\} \tag{2.9}$$

we find that $\Omega_0 \subset \mathcal{D}$. Moreover, Ω_0 is not open in $\mathbb{R}^d_+ \times \mathcal{S}^{d\times d}$ since its complement in $\mathcal{S}^{d\times d}$ is not closed.

Suppose that $\boldsymbol{\Theta} \in \mathcal{S}_r^{d\times d}$, where $1 \le r \le d-1$. Then there exists a $d \times d$ orthogonal matrix $\boldsymbol{H}$ such that

$$\boldsymbol{H}'\boldsymbol{\Theta}\boldsymbol{H} = \operatorname{diag}(\lambda_1, \ldots, \lambda_r, 0, \ldots, 0),$$

where $\lambda_1, \ldots, \lambda_r > 0$. Making the change-of-variables $\boldsymbol{w} \to \boldsymbol{H}\boldsymbol{w}$ in (2.6), we obtain

$$L_T(\boldsymbol{\theta}, \boldsymbol{\Theta}) = \int_{\boldsymbol{H}\boldsymbol{w}>\mathbf{0}} \exp(\boldsymbol{\theta}'\boldsymbol{H}\boldsymbol{w} - \boldsymbol{w}'\boldsymbol{H}'\boldsymbol{\Theta}\boldsymbol{H}\boldsymbol{w})\, \mathrm{d}\boldsymbol{w}. \tag{2.10}$$

Let $\boldsymbol{\Lambda}^{-1} = 2\operatorname{diag}(\lambda_1, \ldots, \lambda_r)$, and partition $\boldsymbol{w}$ into sub-vectors $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ such that $\boldsymbol{w}' = (\boldsymbol{w}_1', \boldsymbol{w}_2')$, where $\boldsymbol{w}_1$ is $r \times 1$ and $\boldsymbol{w}_2$ is $(d-r) \times 1$. Also let $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ be the first r and the last $d-r$ columns, respectively, of $\boldsymbol{H}$. Then $\boldsymbol{H} = (\boldsymbol{H}_1 \vdots \boldsymbol{H}_2)$, $\boldsymbol{w}'\boldsymbol{H}'\boldsymbol{\Theta}\boldsymbol{H}\boldsymbol{w} = \frac{1}{2}\boldsymbol{w}_1'\boldsymbol{\Lambda}^{-1}\boldsymbol{w}_1$, $\boldsymbol{H}\boldsymbol{w} = \boldsymbol{H}_1\boldsymbol{w}_1 + \boldsymbol{H}_2\boldsymbol{w}_2$, and it follows from (2.10) that

$$L_T(\boldsymbol{\theta}, \boldsymbol{\Theta}) = \int_{\boldsymbol{H}_1\boldsymbol{w}_1+\boldsymbol{H}_2\boldsymbol{w}_2>\mathbf{0}} \exp(\boldsymbol{\theta}'\boldsymbol{H}_1\boldsymbol{w}_1 - \tfrac{1}{2}\boldsymbol{w}_1'\boldsymbol{\Lambda}^{-1}\boldsymbol{w}_1 + \boldsymbol{\theta}'\boldsymbol{H}_2\boldsymbol{w}_2)\mathrm{d}\boldsymbol{w}_1\mathrm{d}\boldsymbol{w}_2.$$

By the standard device of "completing the square" of the quadratic form, we obtain

$$\boldsymbol{\theta}'\boldsymbol{H}_1\boldsymbol{w}_1 - \tfrac{1}{2}\boldsymbol{w}_1'\boldsymbol{\Lambda}^{-1}\boldsymbol{w} \equiv -\tfrac{1}{2}(\boldsymbol{w}_1 - \boldsymbol{\Lambda}\boldsymbol{H}_1'\boldsymbol{\theta})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{w}_1 - \boldsymbol{\Lambda}\boldsymbol{H}_1'\boldsymbol{\theta}) + \tfrac{1}{2}\boldsymbol{\theta}'\boldsymbol{H}_1\boldsymbol{\Lambda}\boldsymbol{H}_1'\boldsymbol{\theta}.$$

Therefore,

$$\begin{aligned}
&\exp(-\tfrac{1}{2}\boldsymbol{\theta}'\boldsymbol{H}_1\boldsymbol{\Lambda}\boldsymbol{H}_1'\boldsymbol{\theta})L_T(\boldsymbol{\theta}, \boldsymbol{\Theta}) \\
&\quad= \int_{\boldsymbol{H}_1\boldsymbol{w}_1+\boldsymbol{H}_2\boldsymbol{w}_2>\mathbf{0}} \exp[-\tfrac{1}{2}(\boldsymbol{w}_1 - \boldsymbol{\Lambda}\boldsymbol{H}_1'\boldsymbol{\theta})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{w}_1 - \boldsymbol{\Lambda}\boldsymbol{H}_1'\boldsymbol{\theta})]\exp(\boldsymbol{\theta}'\boldsymbol{H}_2\boldsymbol{w}_2)\mathrm{d}\boldsymbol{w}_1\mathrm{d}\boldsymbol{w}_2 \\
&\quad\propto \int_{\boldsymbol{H}_1\boldsymbol{w}_1+\boldsymbol{H}_2\boldsymbol{w}_2>\mathbf{0}} n(\boldsymbol{w}_1; \boldsymbol{\Lambda}\boldsymbol{H}_1'\boldsymbol{\theta}, \boldsymbol{\Lambda})\exp(\boldsymbol{\theta}'\boldsymbol{H}_2\boldsymbol{w}_2)\mathrm{d}\boldsymbol{w}_1\mathrm{d}\boldsymbol{w}_2,
\end{aligned}$$

where $n(\boldsymbol{w}_1; \boldsymbol{\Lambda}\boldsymbol{H}_1'\boldsymbol{\theta}, \boldsymbol{\Lambda})$ denotes the probability density function of the multivariate normal random vector $\boldsymbol{W}_1 \sim N_r(\boldsymbol{\Lambda}\boldsymbol{H}_1'\boldsymbol{\theta}, \boldsymbol{\Lambda})$. Denoting by $\chi(A)$ the indicator function of any interval $A \subset \mathbb{R}$, we obtain

$$\begin{aligned}&\exp(-\tfrac{1}{2}\boldsymbol{\theta}'\boldsymbol{H}_1\boldsymbol{\Lambda}\boldsymbol{H}_1'\boldsymbol{\theta})L_T(\boldsymbol{\theta},\boldsymbol{\Theta})\\ &\qquad \propto \int_{\mathbb{R}^r} n(\boldsymbol{w}_1; \boldsymbol{\Lambda}\boldsymbol{H}_1'\boldsymbol{\theta}, \boldsymbol{\Lambda})\Big[\int_{\mathbb{R}^{d-r}} \chi(\boldsymbol{H}_2\boldsymbol{w}_2 > -\boldsymbol{H}_1\boldsymbol{w}_1)\exp(\boldsymbol{\theta}'\boldsymbol{H}_2\boldsymbol{w}_2)\,\mathrm{d}\boldsymbol{w}_2\Big]\,\mathrm{d}\boldsymbol{w}_1.\end{aligned}$$

It follows that if $\boldsymbol{\Theta} \in \mathcal{S}_r^{d\times d}$ then $(\boldsymbol{\theta},\boldsymbol{\Theta}) \in \mathcal{D}$ if and only if $\boldsymbol{\theta}$ satisfies the condition

$$\mathbb{E}_{\boldsymbol{W}_1} \int_{\mathbb{R}^{d-r}} \chi(\boldsymbol{H}_2\boldsymbol{w}_2 > -\boldsymbol{H}_1\boldsymbol{W}_1)\exp(\boldsymbol{\theta}'\boldsymbol{H}_2\boldsymbol{w}_2)\,\mathrm{d}\boldsymbol{w}_2 < \infty.$$

Consequently, defining

$$\Omega_r := \Big\{(\boldsymbol{\theta},\boldsymbol{\Theta}) \in \mathbb{R}^d \times \mathcal{S}_r^{d\times d} : \mathbb{E}_{\boldsymbol{W}_1} \int_{\mathbb{R}^{d-r}} \chi(\boldsymbol{H}_2\boldsymbol{w}_2 > -\boldsymbol{H}_1\boldsymbol{W}_1)\exp(\boldsymbol{\theta}'\boldsymbol{H}_2\boldsymbol{w}_2)\,\mathrm{d}\boldsymbol{w}_2 < \infty\Big\}, \tag{2.11}$$

where $1 \le r \le d-1$, then $\Omega_r \subset \mathcal{D}$.

Also note that $\mathcal{S}_r^{d\times d}$, for $1 \le r \le d-1$, is not open in $\mathcal{S}^{d\times d}$; this follows from the fact that $\mathcal{S}_r^{d\times d}$ is the inverse image of the set $(0,\infty)^r \times [0,\infty)^{d-r}$ under the continuous mapping which assigns to each $\boldsymbol{\Theta} \in \mathcal{S}^{d\times d}$ the vector of principal minors, $(\det_1(\boldsymbol{\Theta}), \ldots, \det_d(\boldsymbol{\Theta}))$. Therefore, for $1 \le r \le d-1$, Ω_r is not open.

Next, suppose that $\boldsymbol{\Theta} \in \mathcal{S}_d^{d\times d}$. Then the integral (2.6) is a constant multiple of a multivariate normal probability; hence the integral converges for all $\boldsymbol{\theta} \in \mathbb{R}^d$. Therefore if we define

$$\Omega_d := \mathbb{R}^d \times \mathcal{S}_d^{d\times d}, \tag{2.12}$$

then $\Omega_d \subset \mathcal{D}$.

Finally, suppose that $\boldsymbol{\Theta} \notin \bigcup_{r=0}^d \mathcal{S}_r^{d\times d}$. In this case, $\boldsymbol{\Theta}$ has at least one negative eigenvalue, so the eigenvalues of $\boldsymbol{\Theta}$ are of the form $-\lambda_1, \ldots, -\lambda_r, \lambda_{r+1}, \ldots, \lambda_d$ where $r \ge 1$, $\lambda_1, \ldots, \lambda_r > 0$, and $\lambda_{r+1}, \ldots, \lambda_d \ge 0$. Let $\boldsymbol{\Lambda}_1 = \operatorname{diag}(\lambda_1, \ldots, \lambda_r)$, $\boldsymbol{\Lambda}_2 = \operatorname{diag}(\lambda_{r+1}, \ldots, \lambda_d)$, and apply an orthogonal transformation, $\boldsymbol{w} \to \boldsymbol{H}\boldsymbol{w}$, to rewrite the Laplace transform (2.6) as in (2.10). We also partition $\boldsymbol{w}$ and $\boldsymbol{H}$ as before, with $\boldsymbol{w}' = (\boldsymbol{w}_1', \boldsymbol{w}_2')$ and $\boldsymbol{H} = (\boldsymbol{H}_1 \vdots \boldsymbol{H}_2)$, where $\boldsymbol{w}_1$ is $r \times 1$, $\boldsymbol{w}_2$ is $(d-r)\times 1$, $\boldsymbol{H}_1$ is $d \times r$, and $\boldsymbol{H}_2$ is $d \times (d-r)$. Then we obtain

$$L_T(\boldsymbol{\theta},\boldsymbol{\Theta}) = \int_{\boldsymbol{H}_1\boldsymbol{w}_1+\boldsymbol{H}_2\boldsymbol{w}_2>\boldsymbol{0}} \exp(\boldsymbol{\theta}'\boldsymbol{H}_1\boldsymbol{w}_1 + \boldsymbol{w}_1'\boldsymbol{\Lambda}_1\boldsymbol{w}_1 + \boldsymbol{\theta}'\boldsymbol{H}_2\boldsymbol{w}_2 - \boldsymbol{w}_2'\boldsymbol{\Lambda}_2\boldsymbol{w}_2)\,\mathrm{d}\boldsymbol{w}_1\,\mathrm{d}\boldsymbol{w}_2.$$

Since

$$\{(\boldsymbol{w}_1,\boldsymbol{w}_2) : \boldsymbol{H}_1\boldsymbol{w}_1 + \boldsymbol{H}_2\boldsymbol{w}_2 > \boldsymbol{0}\} \supseteq \{(\boldsymbol{w}_1,\boldsymbol{w}_2) : \boldsymbol{H}_1\boldsymbol{w}_1 > \boldsymbol{0}, \boldsymbol{H}_2\boldsymbol{w}_2 > \boldsymbol{0}\}$$

then

$$\begin{aligned} L_T(\boldsymbol{\theta}, \boldsymbol{\Theta}) &\geq \int_{\boldsymbol{H}_1\boldsymbol{w}_1>\mathbf{0},\boldsymbol{H}_2\boldsymbol{w}_2>\mathbf{0}} \exp(\boldsymbol{\theta}'\boldsymbol{H}_1\boldsymbol{w}_1 + \boldsymbol{w}_1'\boldsymbol{\Lambda}_1\boldsymbol{w}_1 + \boldsymbol{\theta}'\boldsymbol{H}_2\boldsymbol{w}_2 - \boldsymbol{w}_2'\boldsymbol{\Lambda}_2\boldsymbol{w}_2)\,\mathrm{d}\boldsymbol{w}_1\,\mathrm{d}\boldsymbol{w}_2 \\ &= \int_{\boldsymbol{H}_1\boldsymbol{w}_1>\mathbf{0}} \exp(\boldsymbol{\theta}'\boldsymbol{H}_1\boldsymbol{w}_1 + \boldsymbol{w}_1'\boldsymbol{\Lambda}_1\boldsymbol{w}_1)\,\mathrm{d}\boldsymbol{w}_1 \cdot \int_{\boldsymbol{H}_2\boldsymbol{w}_2>\mathbf{0}} \exp(\boldsymbol{\theta}'\boldsymbol{H}_2\boldsymbol{w}_2 - \boldsymbol{w}_2'\boldsymbol{\Lambda}_2\boldsymbol{w}_2)\,\mathrm{d}\boldsymbol{w}_2. \end{aligned}$$

Since $\boldsymbol{\Lambda}_1$ is positive definite then $\boldsymbol{\theta}'\boldsymbol{H}_1\boldsymbol{w}_1 + \boldsymbol{w}_1'\boldsymbol{\Lambda}_1\boldsymbol{w}_1 \to \infty$ as $\|\boldsymbol{w}_1\| \to \infty$. Therefore the latter integral with respect to $\boldsymbol{w}_1$ diverges, so the Laplace transform $L_T(\boldsymbol{\theta}, \boldsymbol{\Theta})$ also diverges. Hence,

$$\left\{(\boldsymbol{\theta}, \boldsymbol{\Theta}) \in \mathbb{R}^d \times \mathcal{S}^{d\times d} : \boldsymbol{\Theta} \notin \bigcup_{r=0}^{d} \mathcal{S}_r^{d\times d}\right\} \not\subset \mathcal{D}.$$

Collecting together the above cases, we obtain $\mathcal{D} = \bigcup_{r=0}^{d} \Omega_r$. Also, since Ω_r, $r \leq d-1$, is not open then $\mathcal{D}$ is not open. Therefore the exponential family $\mathcal{P}_{\mathcal{D}}$ is not regular.

Further, the family $\mathcal{P}_{\mathcal{D}}$ is full since $\mathcal{D}$ is of dimension $k = d + \frac{1}{2}d(d+1)$, which is the dimension of the statistic T in (2.5). Finally, the formula (2.8) for the density function $p(\boldsymbol{t}; \boldsymbol{\theta}, \boldsymbol{\Theta})$, when $(\boldsymbol{\theta}, \boldsymbol{\Theta}) \in \Omega_d$, follows from (2.7). $\square$

The family $\mathcal{P}_{\mathcal{D}}$ introduced in this example, after a suitable parametrization, is the family of truncated multivariate normal distributions. Let $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ be the parameters of the distribution in (2.3), then the canonical parameters of the family $\mathcal{P}_{\mathcal{D}}$ are $\boldsymbol{\theta} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ and $\boldsymbol{\Theta} = \frac{1}{2}\boldsymbol{\Sigma}^{-1}$.

3 The non-steepness of the exponential family $\mathcal{P}_{\mathcal{D}}$

In this section we show that the exponential family $\mathcal{P}_{\mathcal{D}}$ of truncated multivariate normal studied is not steep.

Recall that the set of natural parameters of the family $\mathcal{P}_{\mathcal{D}}$ consists of pairs $\boldsymbol{\omega} = (\boldsymbol{\theta}, \boldsymbol{\Theta}) \in \mathcal{D} = \bigcup_{r=0}^{d} \Omega_r$. For $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)' \in \mathbb{R}^d$, the gradient operator is

$$\nabla_{\boldsymbol{\theta}} = \Big(\frac{\partial}{\partial\theta_1}, \ldots, \frac{\partial}{\partial\theta_d}\Big)'.$$

Introduce the usual Kronecker symbol: $\delta_{i,j} = 1$ or 0 according as $i = j$ or $i \neq j$, respectively. Then for $\boldsymbol{\Theta} = (\theta_{i,j}) \in \mathcal{S}^{d\times d}$, the gradient operator on $\mathcal{S}^{d\times d}$ is the $d \times d$ matrix

$$\nabla_{\boldsymbol{\Theta}} = \left(\tfrac{1}{2}(1 + \delta_{i,j})\frac{\partial}{\partial\theta_{i,j}}\right).$$

Therefore, for $\boldsymbol{\omega} = (\boldsymbol{\theta}, \boldsymbol{\Theta})$, we define

$$\nabla_{\boldsymbol{\omega}} = (\nabla_{\boldsymbol{\theta}}, \nabla_{\boldsymbol{\Theta}}).$$

For $\boldsymbol{A} \in \mathcal{S}^{d\times d}$, the *Frobenius norm* of $\boldsymbol{A}$ is defined [11, p. 291] as $\|\boldsymbol{A}\|_F := [\mathrm{tr}(\boldsymbol{A}^2)]^{1/2}$. Then for a differentiable function $K : \mathbb{R}^d \times \mathcal{S}^{d\times d} \to \mathbb{R}$, we define

$$\|\nabla_{\boldsymbol{\omega}} K(\boldsymbol{\theta}, \boldsymbol{\Theta})\|^2 := \|\nabla_{\boldsymbol{\theta}} K(\boldsymbol{\theta}, \boldsymbol{\Theta})\|^2 + \|\nabla_{\boldsymbol{\Theta}} K_T(\boldsymbol{\theta}, \boldsymbol{\Theta})\|_F^2.$$

Theorem 3.1. *The exponential family $\mathcal{P}_{\mathcal{D}}$ is not steep.*

Proof. We proceed according to the definition of steepness, as given in (2.2). Thus, we need to show that there exists $(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_0) \in \partial\Omega$, the boundary of Ω, such that $\|\nabla_{\boldsymbol{\omega}} K_T(\boldsymbol{\theta}, \boldsymbol{\Theta})\|^2 \not\to \infty$ as $(\boldsymbol{\theta}, \boldsymbol{\Theta}) \to (\boldsymbol{\theta}_0, \boldsymbol{\Theta}_0)$.

By (2.7),

$$\nabla_{\boldsymbol{\theta}} K_T(\boldsymbol{\theta}, \boldsymbol{\Theta}) = \frac{\int_{\boldsymbol{t}>\boldsymbol{0}} \boldsymbol{t} \exp(\boldsymbol{\theta}'\boldsymbol{t} - \boldsymbol{t}'\boldsymbol{\Theta}\boldsymbol{t}) \,\mathrm{d}\boldsymbol{t}}{\int_{\boldsymbol{t}>\boldsymbol{0}} \exp(\boldsymbol{\theta}'\boldsymbol{t} - \boldsymbol{t}'\boldsymbol{\Theta}\boldsymbol{t}) \,\mathrm{d}\boldsymbol{t}}$$

and

$$\nabla_{\boldsymbol{\Theta}} K_T(\boldsymbol{\theta}, \boldsymbol{\Theta}) = -\frac{\int_{\boldsymbol{t}>\boldsymbol{0}} \boldsymbol{t}\boldsymbol{t}' \exp(\boldsymbol{\theta}'\boldsymbol{t} - \boldsymbol{t}'\boldsymbol{\Theta}\boldsymbol{t}) \,\mathrm{d}\boldsymbol{t}}{\int_{\boldsymbol{t}>\boldsymbol{0}} \exp(\boldsymbol{\theta}'\boldsymbol{t} - \boldsymbol{t}'\boldsymbol{\Theta}\boldsymbol{t}) \,\mathrm{d}\boldsymbol{t}}$$

Consider a sequence $(\boldsymbol{\theta}, \boldsymbol{\Theta}_n) \in \mathbb{R}^d \times \mathcal{S}_d^{d\times d}$ where $\boldsymbol{\theta}$ is fixed, $\boldsymbol{\theta} < \boldsymbol{0}$, and $\boldsymbol{\Theta}_n \to \boldsymbol{0}$ as $n \to \infty$. Then,

$$\nabla_{\boldsymbol{\theta}} K_T(\boldsymbol{\theta}, \boldsymbol{\Theta}_n) := \nabla_{\boldsymbol{\theta}} K_T(\boldsymbol{\theta}, \boldsymbol{\Theta})\Big|_{\boldsymbol{\Theta}=\boldsymbol{\Theta}_n} = \frac{\int_{\boldsymbol{t}>\boldsymbol{0}} \boldsymbol{t} \exp(\boldsymbol{\theta}'\boldsymbol{t} - \boldsymbol{t}'\boldsymbol{\Theta}_n\boldsymbol{t}) \,\mathrm{d}\boldsymbol{t}}{\int_{\boldsymbol{t}>\boldsymbol{0}} \exp(\boldsymbol{\theta}'\boldsymbol{t} - \boldsymbol{t}'\boldsymbol{\Theta}_n\boldsymbol{t}) \,\mathrm{d}\boldsymbol{t}},$$

and the numerator and denominator integrals both exist since $\boldsymbol{\Theta}_n$ is positive definite. Therefore, as $n \to \infty$,

$$\nabla_{\boldsymbol{\theta}} K_T(\boldsymbol{\theta}, \boldsymbol{\Theta}_n) \to \frac{\int_{\boldsymbol{t}>\boldsymbol{0}} \boldsymbol{t} \exp(\boldsymbol{\theta}'\boldsymbol{t}) \,\mathrm{d}\boldsymbol{t}}{\int_{\boldsymbol{t}>\boldsymbol{0}} \exp(\boldsymbol{\theta}'\boldsymbol{t}) \,\mathrm{d}\boldsymbol{t}},$$

and these integrals also are finite since $\boldsymbol{\theta} < \boldsymbol{0}$. On calculating these integrals we find that $\nabla_{\boldsymbol{\theta}} K_T(\boldsymbol{\theta}, \boldsymbol{\Theta}_n)$ converges to a vector whose jth component is

$$\frac{\int_0^\infty t_j \exp(\theta_j t_j) \,\mathrm{d}t_j}{\int_0^\infty \exp(\theta_j t_j) \,\mathrm{d}t_j} = \frac{(-\theta_j)^{-2}}{(-\theta_j)^{-1}} = -\theta_j^{-1},$$

$j = 1, \ldots, d$, and therefore

$$\|\nabla_{\boldsymbol{\theta}} K_T(\boldsymbol{\theta}, \boldsymbol{\Theta}_n)\|^2 \to \sum_{j=1}^d \theta_j^{-2}. \tag{3.1}$$

Note that we also have

$$\boldsymbol{\theta}'\nabla_{\boldsymbol{\theta}} K_T(\boldsymbol{\theta}, \boldsymbol{\Theta}_n) \to \sum_{j=1}^d \theta_j(-\theta_j^{-1}) = -d$$

as $n \to \infty$, which generalizes a result of del Castillo [7, p. 62].

Next, we have

$$\nabla_{\mathbf{\Theta}} K_T(\boldsymbol{\theta}, \mathbf{\Theta}_n) := \nabla_{\mathbf{\Theta}} K_T(\boldsymbol{\theta}, \mathbf{\Theta})\Big|_{\mathbf{\Theta}=\mathbf{\Theta}_n} = -\frac{\int_{\boldsymbol{t}>\mathbf{0}} \boldsymbol{t}\boldsymbol{t}' \exp(\boldsymbol{\theta}'\boldsymbol{t} - \boldsymbol{t}'\mathbf{\Theta}_n\boldsymbol{t})\, \mathrm{d}\boldsymbol{t}}{\int_{\boldsymbol{t}>\mathbf{0}} \exp(\boldsymbol{\theta}'\boldsymbol{t} - \boldsymbol{t}'\mathbf{\Theta}_n\boldsymbol{t})\, \mathrm{d}\boldsymbol{t}}.$$

Therefore, as $n \to \infty$,

$$\nabla_{\mathbf{\Theta}} K_T(\boldsymbol{\theta}, \mathbf{\Theta}_n) \to -\frac{\int_{\boldsymbol{t}>\mathbf{0}} \boldsymbol{t}\boldsymbol{t}' \exp(\boldsymbol{\theta}'\boldsymbol{t})\, \mathrm{d}\boldsymbol{t}}{\int_{\boldsymbol{t}>\mathbf{0}} \exp(\boldsymbol{\theta}'\boldsymbol{t})\, \mathrm{d}\boldsymbol{t}},$$

and these integrals are finite since $\boldsymbol{\theta} < \mathbf{0}$. On calculating these integrals we find that $\nabla_{\mathbf{\Theta}} K_T(\boldsymbol{\theta}, \mathbf{\Theta}_n)$ converges to $\boldsymbol{M}$, a $d \times d$ matrix whose (i, j)th entry is

$$m_{i,j} = -\frac{\int_{\mathbb{R}_+^2} t_i t_j \exp(\theta_i t_i + \theta_j t_j)\, \mathrm{d}t_i \mathrm{d}t_j}{\int_{\mathbb{R}_+^2} \exp(\theta_i t_i + \theta_j t_j)\, \mathrm{d}t_i \mathrm{d}t_j} = -\frac{(-\theta_i)^{-2}(-\theta_j)^{-2}}{(-\theta_i)^{-1}(-\theta_j)^{-1}} = -\theta_i^{-1}\theta_j^{-1},$$

$i, j = 1, \ldots, d$. Hence, as $n \to \infty$,

$$\|\nabla_{\mathbf{\Theta}} K_T(\boldsymbol{\theta}, \mathbf{\Theta}_n)\|_F^2 \to \|\boldsymbol{M}\|_F^2 = \sum_{i,j=1}^{d} m_{i,j}^2 = \sum_{i,j=1}^{d} (-\theta_i^{-1}\theta_j^{-1})^2 = \Big(\sum_{j=1}^{d} \theta_j^{-2}\Big)^2. \tag{3.2}$$

Consequently it follows from (3.1) and (3.2) that, for $\boldsymbol{\theta} < \mathbf{0}$,

$$\|\nabla_{\boldsymbol{\omega}} K_T(\boldsymbol{\theta}, \mathbf{\Theta}_n)\|^2 \to \sum_{j=1}^{d} \theta_j^{-2} + \Big(\sum_{j=1}^{d} \theta_j^{-2}\Big)^2,$$

a finite limit, as $n \to \infty$. Therefore the family $\mathcal{P}_{\mathcal{D}}$ is not steep. □

4 Maximum likelihood estimation for the truncated multivariate normal distributions

In Section 3, we proved that the family of truncated multivariate normal distributions is not steep, thus the maximum likelihood estimator of the parameters of the distribution may not always be a solution to the likelihood equations. In this section, we consider the case in which $\mathbf{\Sigma}$ is positive definite, and we derive the score equations and obtain a necessary condition for the existence of solutions to the likelihood equations.

Let $\boldsymbol{X}$ denote a random vector having the d-dimensional normal distribution truncated at $\mathbf{0}$, with the probability density function given in (2.3) and with the normalizing constant (2.4). We write $\boldsymbol{X} \sim N_d(\boldsymbol{\mu}, \mathbf{\Sigma}; \mathbf{0})$ whenever $\boldsymbol{X}$ has the density function (2.3).

Unlike the untruncated normal distribution, $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ are not the mean and covariance matrix, respectively, of $\boldsymbol{X}$, so we define

$$\boldsymbol{\nu} = (\nu_1, \ldots, \nu_d) := \mathbb{E}(\boldsymbol{X})$$

and

$$\boldsymbol{\Lambda} = (\lambda_{i,j}) := \operatorname{Cov}(\boldsymbol{X}) \equiv \mathbb{E}[(\boldsymbol{X} - \boldsymbol{\nu})(\boldsymbol{X} - \boldsymbol{\nu})'].$$

To obtain the likelihood equations for calculating $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$, the following auxiliary result is crucial. In stating the result, we use the notation

$$(\nabla C)(\boldsymbol{\mu}, \boldsymbol{\Sigma}) := \nabla_{\boldsymbol{t}}\, C(\boldsymbol{t}, \boldsymbol{\Sigma})\Big|_{\boldsymbol{t}=\boldsymbol{\mu}}$$

for the gradient of $C(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with respect to $\boldsymbol{\mu}$ and

$$(\nabla\nabla' C)(\boldsymbol{\mu}, \boldsymbol{\Sigma}) := \nabla_{\boldsymbol{t}}\, \nabla'_{\boldsymbol{t}} C(\boldsymbol{t}, \boldsymbol{\Sigma})\Big|_{\boldsymbol{t}=\boldsymbol{\mu}}$$

for the corresponding Hessian matrix.

Lemma 4.1. *Suppose that $\boldsymbol{X} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \boldsymbol{0})$. Then*

$$\boldsymbol{\nu} = \boldsymbol{\mu} + [C(\boldsymbol{\mu}, \boldsymbol{\Sigma})]^{-1}\ \boldsymbol{\Sigma}\,(\nabla C)(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{4.1}$$

and

$$\boldsymbol{\Lambda} = \boldsymbol{\Sigma} + [C(\boldsymbol{\mu}, \boldsymbol{\Sigma})]^{-1} \boldsymbol{\Sigma} (\nabla\nabla' C)(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \boldsymbol{\Sigma} - (\boldsymbol{\nu} - \boldsymbol{\mu})(\boldsymbol{\nu} - \boldsymbol{\mu})'. \tag{4.2}$$

Proof. For $\boldsymbol{t} = (t_1, \ldots, t_d)' \in \mathbb{R}^d$, it follows from (2.3) that the moment-generating function of $\boldsymbol{X}$ is

$$\mathbb{E}(e^{\boldsymbol{t}'\boldsymbol{X}}) = \frac{1}{[C(\boldsymbol{\mu}, \boldsymbol{\Sigma})]} \int_{\mathbb{R}^d_+} \exp[\boldsymbol{t}'\boldsymbol{x} - \tfrac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})]\, \mathrm{d}\boldsymbol{x}.$$

Applying the usual approach of completing the square, we obtain the algebraic identity,

$$\boldsymbol{t}'\boldsymbol{x} - \tfrac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \equiv \boldsymbol{t}'\boldsymbol{\mu} + \tfrac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t} - \tfrac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu} - \boldsymbol{\Sigma}\boldsymbol{t})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu} - \boldsymbol{\Sigma}\boldsymbol{t}),$$

hence

$$\begin{aligned}
\mathbb{E}(e^{\boldsymbol{t}'\boldsymbol{X}}) &= \frac{1}{[C(\boldsymbol{\mu}, \boldsymbol{\Sigma})]} \exp(\boldsymbol{t}'\boldsymbol{\mu} + \tfrac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}) \int_{\mathbb{R}^d_+} \exp[-\tfrac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu} - \boldsymbol{\Sigma}\boldsymbol{t})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu} - \boldsymbol{\Sigma}\boldsymbol{t})]\, \mathrm{d}\boldsymbol{x} \\
&= \frac{1}{[C(\boldsymbol{\mu}, \boldsymbol{\Sigma})]} \exp(\boldsymbol{t}'\boldsymbol{\mu} + \tfrac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t})\, C(\boldsymbol{\mu} + \boldsymbol{\Sigma}\boldsymbol{t}, \boldsymbol{\Sigma}).
\end{aligned}$$

Therefore

$$\log \mathbb{E}(e^{\boldsymbol{t}'\boldsymbol{X}}) = K(\boldsymbol{t}) = -\log C(\boldsymbol{\mu}, \boldsymbol{\Sigma}) + \boldsymbol{t}'\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t} + \log C(\boldsymbol{\mu} + \boldsymbol{\Sigma}\boldsymbol{t}, \boldsymbol{\Sigma}). \tag{4.3}$$

By the chain rule,

$$\nabla_{\boldsymbol{t}}[C(\boldsymbol{\mu} + \boldsymbol{\Sigma}\boldsymbol{t}, \boldsymbol{\Sigma})] = \boldsymbol{\Sigma}\,(\nabla C)(\boldsymbol{\mu} + \boldsymbol{\Sigma}\boldsymbol{t}, \boldsymbol{\Sigma});$$

therefore, by differentiating (4.3), we obtain

$$\frac{\mathbb{E}(e^{\boldsymbol{t}'\boldsymbol{X}}\boldsymbol{X})}{\mathbb{E}(e^{\boldsymbol{t}'\boldsymbol{X}})} = \nabla K(\boldsymbol{t}) = \boldsymbol{\mu} + \boldsymbol{\Sigma t} + \frac{1}{C(\boldsymbol{\mu}+\boldsymbol{\Sigma t},\boldsymbol{\Sigma})}\,\boldsymbol{\Sigma}\,(\nabla C)(\boldsymbol{\mu}+\boldsymbol{\Sigma t},\boldsymbol{\Sigma})$$
$$= \boldsymbol{\mu} + \boldsymbol{\Sigma}\Big(\boldsymbol{t} + \frac{1}{C(\boldsymbol{\mu}+\boldsymbol{\Sigma t},\boldsymbol{\Sigma})}\,(\nabla C)(\boldsymbol{\mu}+\boldsymbol{\Sigma t},\boldsymbol{\Sigma})\Big). \tag{4.4}$$

Setting $\boldsymbol{t} = \boldsymbol{0}$ we obtain

$$\boldsymbol{\nu} = \mathbb{E}(\boldsymbol{X}) = \nabla K(\boldsymbol{t})\big|_{\boldsymbol{t}=\boldsymbol{0}} = \boldsymbol{\mu} + \frac{1}{C(\boldsymbol{\mu},\boldsymbol{\Sigma})}\,\boldsymbol{\Sigma}\,(\nabla C)(\boldsymbol{\mu},\boldsymbol{\Sigma}), \tag{4.5}$$

and this establishes (4.1).

Next, it follows from (4.4) that

$$\nabla_{\boldsymbol{t}}\nabla'_{\boldsymbol{t}}K(\boldsymbol{t}) \equiv \nabla_{\boldsymbol{t}}[\nabla_{\boldsymbol{t}}K(\boldsymbol{t})]'$$
$$= \nabla_{\boldsymbol{t}}\Big[\boldsymbol{\mu} + \boldsymbol{\Sigma}\Big(\boldsymbol{t} + \frac{1}{C(\boldsymbol{\mu}+\boldsymbol{\Sigma t},\boldsymbol{\Sigma})}\,(\nabla C)(\boldsymbol{\mu}+\boldsymbol{\Sigma t},\boldsymbol{\Sigma})\Big)\Big]'$$
$$= \nabla_{\boldsymbol{t}}\Big(\boldsymbol{t}' + \frac{1}{C(\boldsymbol{\mu}+\boldsymbol{\Sigma t},\boldsymbol{\Sigma})}\,(\nabla C)'(\boldsymbol{\mu}+\boldsymbol{\Sigma t},\boldsymbol{\Sigma})\Big)\boldsymbol{\Sigma}. \tag{4.6}$$

Also $\nabla_{\boldsymbol{t}}\boldsymbol{t}' = \boldsymbol{I}_d$ and, by the chain rule,

$$\nabla_{\boldsymbol{t}}\Big(\frac{1}{C(\boldsymbol{\mu}+\boldsymbol{\Sigma t},\boldsymbol{\Sigma})}\,(\nabla C)'(\boldsymbol{\mu}+\boldsymbol{\Sigma t},\boldsymbol{\Sigma})\Big)$$
$$= \frac{1}{\big[C(\boldsymbol{\mu}+\boldsymbol{\Sigma t},\boldsymbol{\Sigma})\big]^2}\,\boldsymbol{\Sigma}$$
$$\times\big[C(\boldsymbol{\mu}+\boldsymbol{\Sigma t},\boldsymbol{\Sigma})\,(\nabla\nabla' C)(\boldsymbol{\mu}+\boldsymbol{\Sigma t},\boldsymbol{\Sigma}) - (\nabla C)(\boldsymbol{\mu}+\boldsymbol{\Sigma t},\boldsymbol{\Sigma})\,(\nabla C)'(\boldsymbol{\mu}+\boldsymbol{\Sigma t},\boldsymbol{\Sigma})\big]. \tag{4.7}$$

Substituting (4.7) into (4.6) and evaluating the resulting expression at $\boldsymbol{t} = \boldsymbol{0}$, we obtain

$$\boldsymbol{\Lambda} = \nabla_{\boldsymbol{t}}\nabla'_{\boldsymbol{t}}K(\boldsymbol{t})\Big|_{\boldsymbol{t}=\boldsymbol{0}}$$
$$= \boldsymbol{\Sigma} + [C(\boldsymbol{\mu},\boldsymbol{\Sigma})]^{-2}\boldsymbol{\Sigma}\big[C(\boldsymbol{\mu},\boldsymbol{\Sigma})\cdot(\nabla\nabla' C)(\boldsymbol{\mu},\boldsymbol{\Sigma}) - (\nabla C)(\boldsymbol{\mu},\boldsymbol{\Sigma})\cdot(\nabla C)'(\boldsymbol{\mu},\boldsymbol{\Sigma})\big]\boldsymbol{\Sigma}$$
$$= \boldsymbol{\Sigma} + [C(\boldsymbol{\mu},\boldsymbol{\Sigma})]^{-1}\boldsymbol{\Sigma}(\nabla\nabla' C)(\boldsymbol{\mu},\boldsymbol{\Sigma})\boldsymbol{\Sigma} - [C(\boldsymbol{\mu},\boldsymbol{\Sigma})]^{-2}\boldsymbol{\Sigma}(\nabla C)(\boldsymbol{\mu},\boldsymbol{\Sigma})\cdot(\nabla C)'(\boldsymbol{\mu},\boldsymbol{\Sigma})\boldsymbol{\Sigma}. \tag{4.8}$$

By (4.1),

$$[C(\boldsymbol{\mu},\boldsymbol{\Sigma})]^{-2}\boldsymbol{\Sigma}(\nabla C)(\boldsymbol{\mu},\boldsymbol{\Sigma})\cdot(\nabla C)'(\boldsymbol{\mu},\boldsymbol{\Sigma})\boldsymbol{\Sigma}$$
$$\equiv [C(\boldsymbol{\mu},\boldsymbol{\Sigma})]^{-1}\boldsymbol{\Sigma}(\nabla C)(\boldsymbol{\mu},\boldsymbol{\Sigma})\cdot\big([C(\boldsymbol{\mu},\boldsymbol{\Sigma})]^{-1}\boldsymbol{\Sigma}(\nabla C)(\boldsymbol{\mu},\boldsymbol{\Sigma})]\big)'$$
$$= (\boldsymbol{\nu}-\boldsymbol{\mu})(\boldsymbol{\nu}-\boldsymbol{\mu})',$$

and by substituting this expression into (4.8), we obtain (4.2). □

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be a random sample from $\boldsymbol{X}$. As shown in Section 3, if $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is unknown then the truncated multivariate normal distributions are not steep and therefore not regular. Therefore $(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}})$, the maximum likelihood estimator of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, possibly may not exist.

Suppose, on the other hand, that $(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}})$ exists. By (4.1), (4.2), and the invariance principle of maximum likelihood estimation [2, pp. 70–71], we find that $(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}})$ are related to $(\widehat{\boldsymbol{\nu}}, \widehat{\boldsymbol{\Lambda}})$, the maximum likelihood estimators of $(\boldsymbol{\nu}, \boldsymbol{\Lambda})$, through the equations

$$\widehat{\boldsymbol{\nu}} = \widehat{\boldsymbol{\mu}} + [C(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}})]^{-1} \; \widehat{\boldsymbol{\Sigma}} \, (\nabla C)(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}) \tag{4.9}$$

and

$$\widehat{\boldsymbol{\Lambda}} = \widehat{\boldsymbol{\Sigma}} + [C(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}})]^{-1} \widehat{\boldsymbol{\Sigma}} \, (\nabla\nabla' C)(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}) \, \widehat{\boldsymbol{\Sigma}} - (\widehat{\boldsymbol{\nu}} - \widehat{\boldsymbol{\mu}})(\widehat{\boldsymbol{\nu}} - \widehat{\boldsymbol{\mu}})'. \tag{4.10}$$

In order to apply the latter equations to calculate $(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}})$, we now derive the score equations for $(\widehat{\boldsymbol{\nu}}, \widehat{\boldsymbol{\Lambda}})$.

In the following results, we set $\bar{\boldsymbol{x}} = n^{-1} \sum_{j=1}^n \boldsymbol{x}_j$, the mean of the random sample. Also, for any vector $\boldsymbol{\alpha} \in \mathbb{R}^d$, define

$$\boldsymbol{S}(\boldsymbol{\alpha}) = \frac{1}{n} \sum_{j=1}^n (\boldsymbol{x}_j - \boldsymbol{\alpha})(\boldsymbol{x}_j - \boldsymbol{\alpha})'. \tag{4.11}$$

In particular,

$$\boldsymbol{S}(\bar{\boldsymbol{x}}) = \frac{1}{n} \sum_{j=1}^n (\boldsymbol{x}_j - \bar{\boldsymbol{x}})(\boldsymbol{x}_j - \bar{\boldsymbol{x}})'$$

is the maximum likelihood estimator of the covariance matrix in the classical untruncated setting.

Proposition 4.2. *Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be a random sample from $\boldsymbol{X}$ and suppose that $(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}})$, the maximum likelihood estimator of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, exists and lies in the interior of the parameter space. Then $(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}})$ satisfies the equations*

$$\widehat{\boldsymbol{\mu}} + [C(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}})]^{-1} \; \widehat{\boldsymbol{\Sigma}} \, (\nabla C)(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}) = \bar{\boldsymbol{x}}, \tag{4.12}$$

and

$$\widehat{\boldsymbol{\Sigma}} + [C(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}})]^{-1} \widehat{\boldsymbol{\Sigma}} \, (\nabla\nabla' C)(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}) \, \widehat{\boldsymbol{\Sigma}} - (\bar{\boldsymbol{x}} - \widehat{\boldsymbol{\mu}})(\bar{\boldsymbol{x}} - \widehat{\boldsymbol{\mu}})' = \boldsymbol{S}(\bar{\boldsymbol{x}}). \tag{4.13}$$

Proof. In deriving the score equations, we treat the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ as fixed at their sample values, and $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are viewed temporarily as variables in the derivation of the score equations. Then the log-likelihood function corresponding to the random sample $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ is

$$\begin{aligned} \ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \sum_{j=1}^n \log f(\boldsymbol{x}_j; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &= -n \log C(\boldsymbol{\mu}, \boldsymbol{\Sigma}) - \frac{1}{2} \sum_{j=1}^n (\boldsymbol{x}_j - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_j - \boldsymbol{\mu}), \end{aligned} \tag{4.14}$$

and therefore

$$\nabla_{\boldsymbol{\mu}} \ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = -n \frac{\nabla_{\boldsymbol{\mu}} C(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{C(\boldsymbol{\mu}, \boldsymbol{\Sigma})} + n \, \boldsymbol{\Sigma}^{-1} (\bar{\boldsymbol{x}} - \boldsymbol{\mu}). \tag{4.15}$$

By (4.5),

$$\frac{\nabla_{\boldsymbol{\mu}} C(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{C(\boldsymbol{\mu}, \boldsymbol{\Sigma})} = \boldsymbol{\Sigma}^{-1} (\boldsymbol{\nu} - \boldsymbol{\mu}),$$

and by substituting this result into (4.15) we obtain

$$\begin{aligned} \nabla_{\boldsymbol{\mu}} \ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= -n \, \boldsymbol{\Sigma}^{-1} (\boldsymbol{\nu} - \boldsymbol{\mu}) + n \, \boldsymbol{\Sigma}^{-1} (\bar{\boldsymbol{x}} - \boldsymbol{\mu}) \\ &= -n \, \boldsymbol{\Sigma}^{-1} (\boldsymbol{\nu} - \bar{\boldsymbol{x}}). \end{aligned}$$

As $(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}})$ is a stationary point of $\ell(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, we obtain

$$-n \, \widehat{\boldsymbol{\Sigma}}^{-1} (\widehat{\boldsymbol{\nu}} - \bar{\boldsymbol{x}}) = \nabla_{\boldsymbol{\mu}} \ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Big|_{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}})} = \mathbf{0},$$

hence $\widehat{\boldsymbol{\nu}} = \bar{\boldsymbol{x}}$. Substituting for $\widehat{\boldsymbol{\nu}}$ from (4.9), we obtain (4.12).

Next, we derive the score equations for $\boldsymbol{\Sigma}$ in terms of $\boldsymbol{\Psi} = \boldsymbol{\Sigma}^{-1}$. Denoting $\ell(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ by $\tilde{\ell}(\boldsymbol{\mu}, \boldsymbol{\Psi})$, $C(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ by $\tilde{C}(\boldsymbol{\mu}, \boldsymbol{\Psi})$, and using the definition of $\boldsymbol{S}(\boldsymbol{\mu})$ from (4.11), we find that the log-likelihood function (4.14) equals

$$\tilde{\ell}(\boldsymbol{\mu}, \boldsymbol{\Psi}) = -n \log \tilde{C}(\boldsymbol{\mu}, \boldsymbol{\Psi}) - \frac{1}{2} n \operatorname{tr} \boldsymbol{\Psi} \boldsymbol{S}(\boldsymbol{\mu}). \tag{4.16}$$

Denote by $\psi_{i,j}$ the (i,j)th entry of $\boldsymbol{\Psi}$, $i, j = 1, \ldots, d$. Since

$$\tilde{C}(\boldsymbol{\mu}, \boldsymbol{\Psi}) = \int_{\boldsymbol{w} \in \mathbb{R}_+^d} \exp\big[-\tfrac{1}{2} (\boldsymbol{w} - \boldsymbol{\mu})' \boldsymbol{\Psi} (\boldsymbol{w} - \boldsymbol{\mu}) \big] \, \mathrm{d}\boldsymbol{w}$$

then, by differentiating (4.16), we obtain

$$\frac{\partial}{\partial \psi_{i,j}} \tilde{\ell}(\boldsymbol{\mu}, \boldsymbol{\Psi}) = -\frac{n}{\tilde{C}(\boldsymbol{\mu}, \boldsymbol{\Psi})} \frac{\partial}{\partial \psi_{i,j}} \tilde{C}(\boldsymbol{\mu}, \boldsymbol{\Psi}) - \frac{n}{2} \frac{\partial}{\partial \psi_{i,j}} \operatorname{tr} \boldsymbol{\Psi} \boldsymbol{S}(\boldsymbol{\mu}). \tag{4.17}$$

For $i, j = 1, \ldots, d$, denote by $a_{i,j}$ the (i,j)th element of $\boldsymbol{S}(\boldsymbol{\mu})$. Since $\boldsymbol{S}(\boldsymbol{\mu})$ is symmetric then

$$\frac{\partial}{\partial \psi_{i,j}} \operatorname{tr} \boldsymbol{\Psi} \boldsymbol{S}(\boldsymbol{\mu}) = (2 - \delta_{i,j}) \, a_{i,j}. \tag{4.18}$$

Differentiating under the integral sign we obtain

$$\begin{aligned} \frac{\partial}{\partial \psi_{i,j}} \tilde{C}(\boldsymbol{\mu}, \boldsymbol{\Psi}) &= \int_{\boldsymbol{w} \in \mathbb{R}_+^d} \frac{\partial}{\partial \psi_{i,j}} \exp\big[-\tfrac{1}{2} (\boldsymbol{w} - \boldsymbol{\mu})' \boldsymbol{\Psi} (\boldsymbol{w} - \boldsymbol{\mu}) \big] \, \mathrm{d}\boldsymbol{w} \\ &= -\frac{1}{2} \int_{\boldsymbol{w} \in \mathbb{R}_+^d} (2 - \delta_{i,j}) (w_i - \mu_i)(w_j - \mu_j) \exp\big[-\tfrac{1}{2} (\boldsymbol{w} - \boldsymbol{\mu})' \boldsymbol{\Psi} (\boldsymbol{w} - \boldsymbol{\mu}) \big] \, \mathrm{d}\boldsymbol{w} \\ &= -\frac{1}{2} (2 - \delta_{i,j}) \, \tilde{C}(\boldsymbol{\mu}, \boldsymbol{\Psi}) \, \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)]. \end{aligned} \tag{4.19}$$

Substituting (4.18) and (4.19) into (4.17), we obtain

$$\frac{\partial}{\partial \psi_{i,j}} \tilde{\ell}(\boldsymbol{\mu}, \boldsymbol{\Psi}) = \frac{n}{2} (2 - \delta_{i,j}) \big(\mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)] - a_{i,j} \big),$$

$i, j = 1, \dots, d$. Since

$$\begin{aligned} \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)] &= \mathbb{E}[(X_i - \nu_i + \nu_i - \mu_i)(X_j - \nu_j + \nu_j - \mu_j)] \\ &= \mathbb{E}[(X_i - \nu_i)(X_j - \nu_j) + (\nu_i - \mu_i)(\nu_j - \mu_j) \\ &= \lambda_{i,j} + (\nu_i - \mu_i)(\nu_j - \mu_j) \end{aligned}$$

then

$$\frac{\partial}{\partial \psi_{i,j}} \tilde{\ell}(\boldsymbol{\mu}, \boldsymbol{\Psi}) = \frac{n}{2} (2 - \delta_{i,j}) \big(\lambda_{i,j} + (\nu_i - \mu_i)(\nu_j - \mu_j) - a_{i,j} \big).$$

Next we simultaneously set these derivatives equal to 0, evaluating the expressions at $(\boldsymbol{\mu}, \boldsymbol{\Psi}) = (\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Psi}})$. Denoting by $\widehat{a}_{i,j}$ the (i, j)th entry of $\boldsymbol{S}(\widehat{\boldsymbol{\mu}})$, we obtain the system of equations

$$\widehat{\lambda}_{i,j} + (\widehat{\nu}_i - \widehat{\mu}_i)(\widehat{\nu}_j - \widehat{\mu}_j) - \widehat{a}_{i,j} = 0$$

for all $i, j = 1, \dots, d$. Denoting $\boldsymbol{S}(\widehat{\boldsymbol{\mu}})$ by $\widehat{\boldsymbol{S}}$ and writing these equations in matrix form, we obtain

$$\widehat{\boldsymbol{\Lambda}} + (\widehat{\boldsymbol{\nu}} - \widehat{\boldsymbol{\mu}})(\widehat{\boldsymbol{\nu}} - \widehat{\boldsymbol{\mu}})' - \widehat{\boldsymbol{S}} = \boldsymbol{0}. \tag{4.20}$$

Since $\widehat{\boldsymbol{\nu}} = \bar{\boldsymbol{x}}$ then standard algebraic manipulations yield

$$\begin{aligned} \widehat{\boldsymbol{S}} &= n^{-1} \sum_{j=1}^{n} (\boldsymbol{x}_j - \widehat{\boldsymbol{\mu}})(\boldsymbol{x}_j - \widehat{\boldsymbol{\mu}})' \\ &= n^{-1} \sum_{j=1}^{n} (\boldsymbol{x}_j - \widehat{\boldsymbol{\nu}} + \widehat{\boldsymbol{\nu}} - \widehat{\boldsymbol{\mu}})(\boldsymbol{x}_j - \widehat{\boldsymbol{\nu}} + \widehat{\boldsymbol{\nu}} - \widehat{\boldsymbol{\mu}})' \\ &= n^{-1} \sum_{j=1}^{n} (\boldsymbol{x}_j - \widehat{\boldsymbol{\nu}})(\boldsymbol{x}_j - \widehat{\boldsymbol{\nu}})' + (\widehat{\boldsymbol{\nu}} - \widehat{\boldsymbol{\mu}})(\widehat{\boldsymbol{\nu}} - \widehat{\boldsymbol{\mu}})'. \end{aligned} \tag{4.21}$$

Substituting this result into (4.20), we find that

$$\widehat{\boldsymbol{\Lambda}} = \widehat{\boldsymbol{S}} - (\widehat{\boldsymbol{\nu}} - \widehat{\boldsymbol{\mu}})(\widehat{\boldsymbol{\nu}} - \widehat{\boldsymbol{\mu}})' = n^{-1} \sum_{j=1}^{n} (\boldsymbol{x}_j - \bar{\boldsymbol{x}})(\boldsymbol{x}_j - \bar{\boldsymbol{x}})' = \boldsymbol{S}(\bar{\boldsymbol{x}}).$$

Finally, by substituting for $\widehat{\boldsymbol{\Lambda}}$ from (4.10), we obtain (4.13). □

Remark 4.3. (i) In the untruncated case, the normalizing constant $C(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ does not depend on $\boldsymbol{\mu}$ and therefore $\nabla_{\boldsymbol{\mu}} C(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \boldsymbol{0}$. Then (4.12) and (4.13) reduce to $\widehat{\boldsymbol{\nu}} = \widehat{\boldsymbol{\mu}} = \bar{\boldsymbol{x}}$ and $\widehat{\mathrm{Cov}}(\boldsymbol{X}) = n^{-1} \sum_{j=1}^{n} (\boldsymbol{x}_j - \bar{\boldsymbol{x}})(\boldsymbol{x}_j - \bar{\boldsymbol{x}})'$, which are well-known results for the untruncated normal distributions.

(ii) It also follows from (4.12) and (4.13) that the method-of-moments estimator of $(\boldsymbol{\nu}, \boldsymbol{\Lambda})$ for the truncated normal distribution is the same as in the untruncated case. This result was noted by Cohen [6, Sections 2.3 and 12.2] for the univariate truncated normal distribution and for the multivariate normal distribution with a single truncated component variable, so that Proposition 4.2 extends Cohen's observation to general multivariate settings and to the more general case in which all components are truncated.

Due to the implicit nature of the score equations (4.12) and (4.13), it is a challenging problem to derive sufficient conditions for the existence and uniqueness of solutions to those equations. Nevertheless, to complete the article, we now obtain a necessary condition for the existence of solutions to the score equations.

Proposition 4.4. *Suppose that there exists a solution to the score equations* (4.12) *and* (4.13). *Then,*

$$0 \le (\bar{\boldsymbol{x}} - \widehat{\boldsymbol{\mu}})' \widehat{\boldsymbol{S}}^{-1} (\bar{\boldsymbol{x}} - \widehat{\boldsymbol{\mu}}) < 1.$$

Proof. Let $\boldsymbol{U}$ be a positive definite $d \times d$ matrix and $\boldsymbol{v} \in \mathbb{R}^d$. By Woodbury's theorem [8, p. 428, Eq. (2.25)],

$$(\boldsymbol{U} + \boldsymbol{v}\boldsymbol{v}')^{-1} = \boldsymbol{U}^{-1} - \frac{\boldsymbol{U}^{-1}\boldsymbol{v}\boldsymbol{v}'\boldsymbol{U}^{-1}}{1 + \boldsymbol{v}'\boldsymbol{U}^{-1}\boldsymbol{v}}.$$

Multiplying this identity on the left by $\boldsymbol{v}'$ and on the right by $\boldsymbol{v}$, and simplifying the result, we obtain

$$\begin{aligned} \boldsymbol{v}'(\boldsymbol{U} + \boldsymbol{v}\boldsymbol{v}')^{-1}\boldsymbol{v} &= \boldsymbol{v}'\boldsymbol{U}^{-1}\boldsymbol{v} - \frac{(\boldsymbol{v}'\boldsymbol{U}^{-1}\boldsymbol{v})^2}{1 + \boldsymbol{v}'\boldsymbol{U}^{-1}\boldsymbol{v}} \\ &= \frac{(1 + \boldsymbol{v}'\boldsymbol{U}^{-1}\boldsymbol{v})\boldsymbol{v}'\boldsymbol{U}^{-1}\boldsymbol{v} - (\boldsymbol{v}'\boldsymbol{U}^{-1}\boldsymbol{v})^2}{1 + \boldsymbol{v}'\boldsymbol{U}^{-1}\boldsymbol{v}} = \frac{\boldsymbol{v}'\boldsymbol{U}^{-1}\boldsymbol{v}}{1 + \boldsymbol{v}'\boldsymbol{U}^{-1}\boldsymbol{v}}. \end{aligned}$$

This proves that $\boldsymbol{v}'(\boldsymbol{U} + \boldsymbol{v}\boldsymbol{v}')^{-1}\boldsymbol{v} \in [0, 1)$.

Now suppose that the score equations (4.12) and (4.13) have a solution. Setting $\boldsymbol{U} = n^{-1} \sum_{j=1}^{n} (\boldsymbol{x}_j - \bar{\boldsymbol{x}})(\boldsymbol{x}_j - \bar{\boldsymbol{x}})'$ and $\boldsymbol{v} = \bar{\boldsymbol{x}} - \widehat{\boldsymbol{\mu}}$, it follows from (4.21) that $\boldsymbol{U} + \boldsymbol{v}\boldsymbol{v}' = \widehat{\boldsymbol{S}}$. Therefore $(\bar{\boldsymbol{x}} - \widehat{\boldsymbol{\mu}})' \widehat{\boldsymbol{S}}^{-1} (\bar{\boldsymbol{x}} - \widehat{\boldsymbol{\mu}}) \equiv \boldsymbol{v}'(\boldsymbol{U} + \boldsymbol{v}\boldsymbol{v}')^{-1}\boldsymbol{v} \in [0, 1)$. □

References

[1] Amemiya, T. (1974). Multivariate regression and simultaneous equations models when the dependent variables are truncated normal. *Econometrica*, **42**, 999-1012.

[2] Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, third edition. Wiley, New York, NY.

[3] Bar-Lev, S. K., and Enis, P. (1986). Reproducibility and natural exponential families with power variance functions. *Ann. Statist.*, **14**, 1507–1522.

[4] Barndorff-Nielsen, O. E. (2014). *Information and Exponential Families in Statistical Theory*, second printing. Wiley, Chichester, U.K.

[5] Brown, L. D. (1986). *Fundamentals of Statistical Exponential Families: With Applications in Statistical Decision Theory*, Institute of Mathematical Statistics, Hayward, CA.

[6] Cohen, A. C. (2016). *Truncated and Censored Samples: Theory and Applications*, CRC Press, Boca Raton, FL.

[7] del Castillo, J. (1994). The singly truncated normal distribution: A non-steep exponential family. *Ann. Inst. Statist. Math.*, **46**, 57–66.

[8] Harville, D. A. (2008). *Matrix Algebra From a Statistician's Perspective*, Springer, New York, NY.

[9] Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Ann. Econ. Social Measurement*, **5**, 475–492.

[10] Hong, H., and Shum, M. (2002). Econometric models of asymmetric ascending auctions. *J. Econometrics*, **112**, 327–358.

[11] Horn, R. A., and Johnson, C. R. (1990). *Matrix Analysis*. Cambridge University Press, New York, NY.

[12] Jin, B. S., Han, J. J., Ding, S., and Miao, B. Q. (2018). EM algorithm of the truncated multinormal distribution with linear restriction on the variables. *Acta Math. Appl. Sinica*, **34,** 155–162.

[13] Lachos, V. H., Matos, L. A., Castro, L. M., and Chen, M.-H. (2019). Flexible longitudinal linear mixed models for multiple censored responses data. *Statistics in Medicine*, **38**, 1074–1102.

[14] Levine, M., Richards, D., and Su, J. (2020). Independence properties of the truncated multivariate elliptical distributions. *Statist. Probab. Lett.*, **161**, 108729.

[15] Lin, T. I., and Wang, W. L. (2017). Multivariate-t nonlinear mixed models with application to censored multi-outcome AIDS studies. *Biostatistics*, **18**, 666–681.

[16] Sundberg, R. (2019). *Statistical Modelling by Exponential Families*. Cambridge University Press, New York, NY.