Journal of Multivariate Analysis 134 (2015) 61–70

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Covariance matrices associated to general moments of a random vector

Songjun Lv

College of Mathematical Science, Chongqing Normal University, Chongqing, 401331, PR China

ARTICLE INFO

Article history:
Received 24 July 2014
Available online 7 November 2014

AMS subject classifications:
62H05
46A40

Keywords:
Covariance matrix
Gaussian gauge
Power function distribution
Logistic distribution
Characterization

ABSTRACT

It turns out that there exist general covariance matrices associated not only to a random vector itself but also to its general moments. In this paper we introduce and characterize general covariance matrices of a random vector that are associated to some important general moments, which are determined by a specific class of convex functions. As special cases, the original covariance matrices of a random vector, as well as the pth covariance matrices characterized recently, are included. The covariance matrices associated to the p-power function distribution and the logistic distribution are characterized as by-products.

1. Introduction

Two fundamental measures of a random vector X in $\mathbb{R}^n$ are the *expected value* (the mean vector) and the *covariance matrix*. The expected value scales the median vector of a random vector X and usually plays the role of a location parameter in most related mathematical models. The covariance matrix Σ of a random vector X in $\mathbb{R}^n$ is the matrix whose (i, j)-entry is the *covariance* $\Sigma_{ij} = \mathrm{cov}(X_i, X_j) = \mathrm{E}\left[(X_i - \mu_i)(X_j - \mu_j)\right]$ where $\mu_i = \mathrm{E}[X_i]$ is the expected value of the ith entry in the vector X. It is conventional to use the covariance matrix to define a *multivariate Gaussian distribution* Z, which can be expressed by the notation $X \sim G(\mu, \Sigma)$, where $G(\mu, \Sigma)$ is the *standard Gaussian distribution* with mean vector μ and covariance matrix Σ.

We will focus only on an important special situation, in which any random vector X that we consider here has the mean vector $\mu = 0$ and a positive definite covariance matrix Σ. In that case, the covariance matrix Σ can be written as

$$\Sigma = \mathrm{cov}(X) = \mathrm{E}[X \otimes X],$$

and the *second moment of a random vector* X in $\mathbb{R}^n$ is given by

$$\mathrm{E}[|X|^2] = \int_{\mathbb{R}^n} |x|^2 f_X(x)dx, \tag{1}$$

where f_X is the density function of X.

An important characterization of the covariance matrix Σ is the following:

The matrix $\Sigma^{-1/2}$ is the unique positive definite symmetric matrix with maximal determinant among all positive definite symmetric matrices A satisfying $\mathrm{E}[|AX|^2] = n$.

There is a natural extension of the moment and second moment of a random vector. For $p \in \mathbb{R}$, the *pth moment* $\mathrm{E}[|X|^p]$ of a random vector X in $\mathbb{R}^n$ can be defined as

$$\mathrm{E}[|X|^p] = \int_{\mathbb{R}^n} |x|^p f_X(x)dx. \tag{2}$$

E-mail address: lvsongjun@126.com.

http://dx.doi.org/10.1016/j.jmva.2014.10.007

Recently, Lutwak et al. [15] extended the characterization of the covariance matrix to a definition of the pth covariance matrix for all $p \in (0, \infty)$:

For a random vector X in $\mathbb{R}^n$ with positive definite covariance matrix $\mathrm{E}[X \otimes X]$, there exists a unique matrix $A \in S$, the set of positive definite symmetric n-by-n matrices, such that A has maximal determinant among all positive definite symmetric matrices A' satisfying that $\mathrm{E}[|A'X|^p] = n$.

The *pth covariance matrix* is defined to be $\Sigma_p = A^{-p}$, where A is given by the characterization above. Since a pth covariance matrix is defined implicitly, one may wonder what kind of explicit expressions it may have. For some certain invariant pth moments under a linear transformation, an explicit integral formula was found in [11]. The authors of [11] also showed that such linear-transformation-invariant pth moments have found important applications to sharp information theoretic inequalities (cf. [12]).

The progress from the classical covariance matrix to the pth covariance matrix shows that there exist some covariance matrices associated not only to the random vector X but also to some certain functions. In fact, a key observation used in [15] is that the power function t^p/p is naturally associated to the pth moment, and therefore can be used to result in a characterization of the pth covariance matrix. This observation motivates us to investigate more general covariance matrices that are associated to other certain functions.

Another motivation of the investigation on general covariance matrix lies in: Since the general moment $\mathrm{E}[\Phi(X)]$, for a continuous scalar-, vector-, or matrix-valued function Φ, is well-defined and has been widely used in congeneric fields, the corresponding general covariance matrices may find multifarious applications in interrelated areas as well.

To investigate the general covariance matrices associated to a certain function ϕ other than the power functions, a key step is to set the notion of the *ϕ-moment of a random vector*.

Recall that the pth moment of a random vector is associated closely to the power function $\phi(t) = t^p/p$ for $t \in (0, \infty)$. On the other hand, a specific form of the density function of the standard Gaussian distribution can be given by $\alpha e^{-|x|^p/p}$, where $\alpha > 0$ is an appropriate constant (the complete profile of the pth Gaussian density is much more complicated, see [10–17] for details and applications of the generalized pth Gaussian distribution). A Gaussian random vector Z is the normal multivariate distribution whose density function f_Z is given by

$$\begin{aligned} f_Z(y) &= \frac{\alpha}{\sqrt{\det C}} \exp\left\{ -\frac{1}{p} \left(x^t C^{-1} x \right)^{\frac{p}{2}} \right\} \\ &= \frac{\exp\left\{ -\phi\left((x^t C^{-1} x)^{\frac{1}{2}} \right) \right\}}{\int_{\mathbb{R}^n} \exp\left\{ -\phi\left((x^t C^{-1} x)^{\frac{1}{2}} \right) \right\} dx}, \end{aligned} \tag{3}$$

for a positive definite symmetric matrix C and some $y \in \mathbb{R}^n$, where $\det C$ and C^{-1} are the determinant and the inverse matrix of the matrix C, and x^t is the transpose of the vector $x \in \mathbb{R}^n$. The mean of Z is 0 and the covariance matrix of Z is $\frac{1}{n}C$. Here the density function f_Z of the Gaussian distribution is actually defined to be the exponential function of $-\phi$. Therefore we can refer to (3) as the density function of ϕ-Gaussian with $\phi(t) = t^p/p$.

Based on these observations, we define a general moment m by

$$\mathrm{E}\left[\phi\left(\frac{|X|}{m} \right) \right] = \mathrm{E}[\phi(Z)]. \tag{4}$$

We will call the moment m the *ϕ-moment* of X and will denote it by $\mathrm{E}_\phi(|X|)$. There are two advantages over the pth moment (2) for the ϕ-moment, even though both definitions refer to the same measure of X when $\phi(t) = t^p/p$. The first one is in that: no matter whether the function ϕ is given explicitly or not, the definition (4) guarantees that $m = n = \mathrm{E}[|Z|^p]$ if $\phi(t) = t^p/p$ and $X = Z$. The second one is in that: the implicit definition (4) allows us to extend the notion of moment from homogeneous ϕ to inhomogeneous ones, which do distribute in a large quantity of mathematical models.

We follow Cianchi et al. [1] to call the Gaussian density generating function ϕ a Gaussian gauge, and regard m defined in (4) as the ϕ-moment associated to the Gaussian gauge ϕ. Gaussian gauges are usually used to define various general Gaussian density functions directly, which in general act as extremal functions in Sobolev embedding theory (functional analysis) and information theoretic inequalities (information theory), see [1–3,5,7,8,10–15,17] for related topics.

Note that the pth covariance matrix solves the maximal determinant problem in all positive definite symmetric matrices A with given pth moment of AX.

In our context, a certain function like the last function in (3) turns into a proper candidate of the density function of ϕ-Gaussian random vectors for other certain gauges ϕ when we characterize the covariance matrices associated to Gaussian gauges.

The main goal of the present paper is to consider the following maximal determinant problem in all positive definite symmetric matrices A with given ϕ-moment:

Is there a unique positive definite symmetric matrix with maximal determinant among all positive definite symmetric matrices A satisfying that the ϕ-moment of AX equals n for a given random vector X in $\mathbb{R}^n$?

We provide an analytic approach to the above maximal problem and characterize ϕ-covariance matrices associated to Gaussian gauges in an implicit way. We show that under certain assumptions there is a positive definite symmetric matrix, unique up to a scalar multiple, solving the ϕ-covariance matrix problem. Existence and characterization of the pth covariance matrix due to Lutwak et al. [15], as well as the logistic covariance matrix and the ϕ-covariance matrix for the power function distribution, are included as consequences of our main result. For characterizations of covariance matrices in different directions, one can refer to [4,9].

Except for the covariance matrix, there is another fundamental matrix playing crucial roles in a number of applications in statistics, information theory, financial economics, and other related fields, see e.g. [18]. That is the Fisher information matrix, whose inverse matrix derives a lower bound on the covariance of any unbiased estimator of the location parameter, see, for example, [1,10]. In view of the key features of the Fisher information matrix in statistics, information theory, and information geometry, to characterize the general Fisher information matrices associated to Gaussian gauges also is of great interest. We will produce characterizations of ϕ-Fisher information matrices in a separate work.

The present paper is organized as follows: In Section 2, we introduce the notion of general moments of a random vector and set basic notations. Section 3 acts as the main ingredient of this paper. We show how the ϕ-covariance matrix problem can be solved by employing some important analytic ideas and techniques. In this section, we present a unified treat for characterizations of covariance matrices associated to Gaussian gauges. Section 4 is devoted to some applicable examples of characterizations of the ϕ-covariance matrices. Among the large amount of inhomogeneous Gaussian gauges, two important inhomogeneous instances of ϕ-covariance matrices are the ones associated to the power function distribution and the logistic distribution. As consequences of our main theorem, we present the characterizations of the logistic covariance matrix and the general covariance matrix in connection with a Gaussian gauge that is tied to the power function distribution.

2. General moments of a random vector

Let X be a random vector in $\mathbb{R}^n$ with probability density function f_X. If Φ is a continuous scalar-, vector-, or matrix-valued function on $\mathbb{R}^n$, then the *expected value* of the random vector X is given by

$$\mathrm{E}[\Phi(X)] = \int_{\mathbb{R}^n} \Phi(x) f_X(x) dx. \tag{5}$$

We define a *gauge* to be a continuously differentiable function $\phi : (0, b) \to \mathbb{R}$ with $0 < b \le \infty$ such that ϕ' is a strictly increasing function. In particular, ϕ is strictly convex.

A gauge ϕ is said to be *Gaussian*, if

$$\int_0^b e^{-\phi(t)} dt < \infty.$$

For a Gaussian gauge $\phi : (0, b) \to \mathbb{R}$, where $0 < b \le \infty$, the *standard ϕ-Gaussian random vector Z_ϕ* is defined to be the random vector whose density function f_ϕ is given by

$$f_{Z_\phi}(y) = e^{-\phi(|y|)} \mathbf{1}_{\mathrm{supp}\phi}(|y|) \Big/ \left(\int_{\mathbb{R}^n} e^{-\phi(|x|)} \mathbf{1}_{\mathrm{supp}\phi}(|x|) dx \right),$$

for $y \in \mathbb{R}^n$, where $\mathbf{1}_{\mathrm{supp}\phi}$ denotes the *characteristic function* of the support $\mathrm{supp}\phi$ of the Gaussian gauge ϕ.

Any random vector of the form $Z = T(Z_\phi - \mu)$, where T is a nonsingular matrix, is called a *ϕ-Gaussian*. If we let $C = TT^t$, then the density function of Z is given explicitly by

$$f_Z(x) = \frac{a}{\sqrt{\det C}} \exp\left\{ -\phi\left((x^t C^{-1} x)^{\frac{1}{2}} \right) \right\} \mathbf{1}_{\mathrm{supp}\phi}\left((x^t C^{-1} x)^{\frac{1}{2}} \right),$$

where $x \in \mathbb{R}^n$ and a is a constant such that f_Z forms a probability density. In our context the mean vector of Z is $\mu = 0$ and the covariance matrix of Z is $\frac{1}{n}C$.

Let ϕ be a Gaussian gauge. A random vector X in $\mathbb{R}^n$ with density f_X is said to *have finite ϕ-moment*, if there exists $m > 0$ such that

$$\mathrm{P}\left[0 < \frac{|X|}{m} < b \right] = 1 \quad \text{and} \quad -\infty < \int_{\mathbb{R}^n} \phi\left(\frac{|x|}{m} \right) f_X(x) dx < \infty. \tag{6}$$

Let Z_ϕ be the standard ϕ-Gaussian random vector associated to the Gaussian gauge ϕ, and denote

$$\hat{\phi} = \mathrm{E}[\phi(Z_\phi)] = \left(\int_{\mathbb{R}^n} e^{-\phi(|x|)} \mathbf{1}_{\mathrm{supp}\phi}(|x|) dx \right)^{-1} \int_{\mathbb{R}^n} \phi(|x|) e^{-\phi(|x|)} dx.$$

Here we always assume that $\hat{\phi}$ is finite. In that case, in view of (6) we define the *ϕ-moment* $\mathrm{E}_\phi(|X|)$ of X as

$$\mathrm{E}_\phi(|X|) = \inf\left\{ m > 0 : \int_{\mathbb{R}^n} \phi\left(\frac{|x|}{m} \right) f_X(x) dx \le \hat{\phi} \right\}. \tag{7}$$

Observe that the infimum in (7) need not be attained, but

$$\mathrm{E}\left[\phi\left(\frac{|X|}{\mathrm{E}_\phi(|X|)} \right) \right] \le \hat{\phi}, \tag{8}$$

with equality if ϕ satisfies the Δ_2-condition, i.e., there exists $c > 0$ such that

$$\phi(2t) \le c\phi(t) \quad \text{for each } t \in (0, b) \tag{9}$$

(see e.g. [19, Corollary 5, Section 3.4]).

For a Gaussian gauge ϕ that satisfies the Δ_2-condition, we define the ϕ-moment $\mathrm{E}_\phi(|X|)$ of a random vector X in $\mathbb{R}^n$ implicitly by

$$\mathrm{E}\left[\phi\left(\frac{|X|}{\mathrm{E}_\phi(|X|)}\right)\right] = \hat{\phi}. \tag{10}$$

Obviously, if $\phi(t) = t^p/p$ for $p > 0$ then $\mathrm{E}_\phi(|X|)$ is exactly the pth moment $\mathrm{E}\left[|X|^p\right]$ in which the second moment $\mathrm{E}[|X|^2]$ of X corresponds to the case $p = 2$.

If c is a positive real number, then

$$\mathrm{E}_\phi(|cX|) = c\mathrm{E}_\phi(|X|). \tag{11}$$

3. General covariance matrices of a random vector

If $w \in \mathbb{R}^n$, we denote by $w \otimes w$ the n-by-n matrix whose (i, j)th component is $w_i w_j$. We call a random vector X non-degenerate, if the original covariance matrix $\mathrm{cov}(X) = \mathrm{E}[X \otimes X]$ of X is positive definite. Recall that we always assume that the expected value of X equals 0.

Denote by S the set of positive definite symmetric n-by-n matrices. The n-by-n identity matrix is denoted by I_n. For each $A \in S$, define a norm $\|\cdot\|_A$ to be

$$\|x\|_A = |Ax| = \sqrt{Ax \cdot Ax}$$

for each $x \in \mathbb{R}^n$.

Let ϕ be Gaussian gauge, and let X be a non-degenerate random vector in $\mathbb{R}^n$ with finite general moment, an extremal problem asks whether there is a unique positive definite symmetric matrix with maximal determinant among all matrices $A \in S$ satisfying

$$\mathrm{E}_\phi(\|X\|_A) = n.$$

As mentioned above, if $\phi(t) = t^p/p$ this problem was solved by Lutwak et al. [15] and the unique matrix was referred to as p-covariance matrix of X (the case where $p = 2$ is exactly the original covariance matrix of X). The following result gives a unified treat for certain Gaussian gauges ϕ.

Theorem 3.1. *Let a Gaussian gauge ϕ satisfy the Δ_2-condition (9), and let X be a non-degenerate random vector in $\mathbb{R}^n$ with finite ϕ-moment, then there exists $A \in S$, unique up to a scalar multiple, such that*

$$\mathrm{E}_\phi(\|X\|_A) = n,$$

and

$$\det A \ge \det A'$$

for each $A' \in S$ with $\mathrm{E}_\phi(\|X\|_{A'}) = n$. Moreover, if A is such a unique matrix in S, then it satisfies

$$\mathrm{E}\left[\phi'\left(\frac{|AX|}{n}\right)\frac{AX \otimes AX}{|AX|}\right] = \frac{1}{n}\mathrm{E}\left[\phi'\left(\frac{|AX|}{n}\right)|AX|\right]I_n. \tag{12}$$

Proof. We divide the proof into three steps.

Step 1. We first claim the existence of a solution to the extremal problem.

Recall that X is non-degenerate, i.e., $\mathrm{E}[X \otimes X]$ is positive definite, is equivalent to that for any $v \in \mathbb{R}^n\backslash\{0\}$, $\mathrm{E}[|v \cdot X|] > 0$, that is to say, there exists a constant $c > 0$ such that

$$\mathrm{E}[|v \cdot X|] \ge c > 0, \quad \text{for any } v \in \mathbb{R}^n \setminus \{0\}. \tag{13}$$

Denoted by S' the set of positive-definite symmetric matrices B satisfying the constraint

$$\mathrm{E}_\phi(|BX|) = n. \tag{14}$$

Because the determinant function is continuous on the set S', we only need to show that S' is compact. Since this set is closed, it suffices to prove that it is bounded.

For given $B \in S'$, let η be the maximal eigenvalue of B with normalized eigenvector e, then it follows that

$$|Bx| \ge \eta|e \cdot x| \quad \text{for any } x \in \mathbb{R}^n \setminus \{0\}. \tag{15}$$

By (14), (7)–(9), (15), Jensen's inequality, and (13), we have

$$\begin{aligned}\hat{\phi} &= \int_{\mathbb{R}^n} \phi\left(\frac{|Bx|}{n}\right) f_X(x)dx \\ &\geq \int_{\mathbb{R}^n} \phi\left(\frac{\eta|e\cdot x|}{n}\right) f_X(x)dx \\ &\geq \phi\left(\frac{\eta}{n}\int_{\mathbb{R}^n} |e\cdot x| f_X(x)dx\right) \\ &= \phi\left(\frac{\eta}{n}\mathrm{E}[|e\cdot X|]\right) \\ &\geq \phi\left(\frac{c}{n}\eta\right).\end{aligned}$$

This together with the invertibility and monotonicity of the function ϕ gives

$$\eta \leq \frac{n}{c}\phi^{-1}(\hat{\phi}).$$

By the definitions of ϕ and $\hat{\phi}$, it is easily seen that $0 < \phi^{-1}(\hat{\phi}) < \infty$. Thus the eigenvalues of B is uniformly bounded from above, proving that the set S' is bounded.

Step 2. Secondly, we show that if A is the matrix with maximal determinant, then (12) holds.

Note that a solution to the maximal determinant problem exists if and only if a solution to the following minimal ϕ-moment problem exists:

Is there a unique positive definite symmetric matrix with minimal ϕ-moment among all matrices $A \in S$ satisfying $\det A = 1$?

In fact, the solutions to the maximal determinant problem and the minimal ϕ-moment problem only differ by a scale factor.

We also note that the existence of a solution to such an extremal problem guarantees a matrix, say, $A \in S$ to be a solution to the minimal ϕ-moment problem. For every $B \in S$, denote $(BA)^s = [(BA)^t(BA)]^{1/2}$, then $\det(BA) = \det(BA)^s$ and $|(BA)x| = |(BA)^s x|$, for each $x \in \mathbb{R}^n$. Let $Y = AX$, then by the fact that the ϕ-moment is homogeneous of degree 1 we have

$$\begin{aligned}\mathrm{E}_\phi(|\overline{B}Y|) &= (\det B)^{-\frac{1}{n}}\mathrm{E}_\phi(|BY|) \\ &= (\det A)^{\frac{1}{n}}(\det(BA))^{-\frac{1}{n}}\mathrm{E}_\phi(|(BA)X|) \\ &= (\det A)^{\frac{1}{n}}(\det(BA)^s)^{-\frac{1}{n}}\mathrm{E}_\phi(|(BA)^s X|) \\ &\geq (\det A)^{\frac{1}{n}}(\det A)^{-\frac{1}{n}}\mathrm{E}_\phi(|AX|) \\ &= \mathrm{E}_\phi(|Y|),\end{aligned} \tag{16}$$

where $\overline{B} = B/(\det B)^{1/n}$.

Thus by the definition (7), we see that the inequality (16) is equivalent to

$$\int_{\mathbb{R}^n} \phi\left(\frac{|\overline{B}Y|}{\mathrm{E}_\phi(|Y|)}\right) f_Y(y)dy \geq \int_{\mathbb{R}^n} \phi\left(\frac{|Y|}{\mathrm{E}_\phi(|Y|)}\right) f_Y(y)dy = \hat{\phi}. \tag{17}$$

Denoting the first integrand in (17) by $F(\overline{B})$, and setting $\overline{B} = (I_n + \varepsilon B')/(\det(I_n + \varepsilon B'))^{1/n}$ for $B' \in S$, we get

$$F\left(\frac{I_n + \varepsilon B'}{\det(I_n + \varepsilon B')^{\frac{1}{n}}}\right) = \int_{\mathbb{R}^n} \phi\left(\frac{|(I_n + \varepsilon B')Y|}{\det(I_n + \varepsilon B')^{\frac{1}{n}}\mathrm{E}_\phi(|Y|)}\right) f_Y(y)dy \geq \hat{\phi},$$

for every ε near 0. Observe that there is a real $\varepsilon_0 > 0$ such that for every near 0 real ε with $|\varepsilon| \leq \varepsilon_0$, the matrix $I_n + \varepsilon B'$ still belongs to the set S. Thus it follows that

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} \int_{\mathbb{R}^n} \phi\left(\frac{|(I_n + \varepsilon B')Y|}{\det(I_n + \varepsilon B')^{\frac{1}{n}}\mathrm{E}_\phi(|Y|)}\right) f_Y(y)dy = 0. \tag{18}$$

Since

$$\frac{d}{d\varepsilon}\phi\left(\frac{|(I_n + \varepsilon B')y|}{\det(I_n + \varepsilon B')^{\frac{1}{n}}\mathrm{E}_\phi(|Y|)}\right)$$

exists and is bounded for all $\varepsilon \in [-\varepsilon_0, \varepsilon_0]$, and $y \in \operatorname{supp} f_Y$, where $\operatorname{supp} f_Y$ is the support of the density f_Y of Y, by the Dominated Convergence Theorem we get

$$\begin{aligned} 0 &= \int_{\operatorname{supp} f_Y} \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} \phi\left(\frac{|(I_n+\varepsilon B')y|}{\det(I_n+\varepsilon B')^{\frac{1}{n}}\mathrm{E}_\phi(|Y|)}\right) f_Y(y)dy \\ &= \frac{1}{\mathrm{E}_\phi(|Y|)}\int_{\operatorname{supp} f_Y} \phi'\left(\frac{|y|}{\mathrm{E}_\phi(|Y|)}\right)\left[\frac{B'Y\cdot Y}{|Y|} - \frac{1}{n}tr(B')|Y|\right] f_Y(y)dy. \end{aligned}$$

A direct calculation shows that this happens only if

$$\mathrm{E}\left[\phi'\left(\frac{|Y|}{\mathrm{E}_\phi(|Y|)}\right)\frac{Y\otimes Y}{|Y|}\right] = \frac{1}{n}\mathrm{E}\left[\phi'\left(\frac{|Y|}{\mathrm{E}_\phi(|Y|)}\right)|Y|\right] I_n. \tag{19}$$

Now by rescaling the given matrix A in (19) so that $\mathrm{E}_\phi(|Y|) = n$ and substituting $Y = AX$, we obtain (12).

Remark 3.2. We note that a Borel measure μ on $\mathbb{R}^n$ is said to be in isotropic position (see e.g. [15]) if

$$\int_{\mathbb{R}^n} \frac{x\otimes x}{|x|^2} d\mu(x) = \frac{1}{n} I_n. \tag{20}$$

Eq. (19) implies that the probability measure

$$d\mu(y) = \frac{\phi'\left(\frac{|y|}{\mathrm{E}_\phi(|Y|)}\right)|y|}{\mathrm{E}\left[\phi'\left(\frac{|Y|}{\mathrm{E}_\phi(|Y|)}\right)|Y|\right]} f_Y(y)dy \tag{21}$$

is in isotropic position.

Note that (20) is equivalent to

$$\int_{\mathbb{R}^n} \frac{(e_i\cdot x)^2}{|x|^2} d\mu(x) = \frac{1}{n}. \tag{22}$$

Step 3. Finally, we claim that the matrix A with minimal ϕ-moment is unique up to a scalar multiplication. We will show that the equality in the inequality (16), or equivalently, the inequality (17), holds if and only if $A = aI_n$ for some $a > 0$.

In Step 2, we proved that for the normalized $\overline{B} \in S$,

$$F(\overline{B}) \geq F(I_n) \tag{23}$$

holds and entails that the measure μ defined by (21) is in isotropic position. Thus it suffices to show that the equality in (23) holds if and only if $\overline{B} = I_n$.

Let $e_1, \ldots, e_n$ be an orthonormal basis of eigenvectors of $\overline{B}$ with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Then

$$\begin{aligned} F(\overline{B}) &= \int_{\mathbb{R}^n} \phi\left(\frac{\left(\sum_{i=1}^n \lambda_i^2 (e_i\cdot y)^2\right)^{1/2}}{\mathrm{E}_\phi(|Y|)}\right) f_Y(y)dy \\ &= \int_{\mathbb{R}^n} \phi\left(\frac{|\operatorname{diag}(\lambda_1,\ldots,\lambda_n)y|}{\mathrm{E}_\phi(|Y|)}\right) f_Y(y)dy \\ &= F(\operatorname{diag}(\lambda_1,\ldots,\lambda_n)). \end{aligned}$$

If we associate to the diagonal matrix $\operatorname{diag}(\lambda_1,\ldots,\lambda_n)$ the vector $\lambda = (\lambda_1,\ldots,\lambda_n)^t \in \mathbb{R}^n$, then we may identify the matrix $\operatorname{diag}(\lambda_1,\ldots,\lambda_n)$ with the vector $\lambda \in \mathbb{R}^n$ and view F as to a function from $\mathbb{R}^n_+$ (a conic subset of $\mathbb{R}^n$ with nonnegative n tubes) to $\mathbb{R}$.

By the continuity and boundedness on $\operatorname{supp} f_Y$ of the function ϕ, it follows from the Dominated Convergence Theorem that F is continuous on $\mathbb{R}^n_+$. Meanwhile, by the subadditivity of the Euclidean norm, the monotonicity and convexity of the function ϕ, we see that $F(\lambda)$ is convex in λ. Also by the strictly increasing monotonicity of ϕ, we know that F is radially monotone increasing, i.e., $F(c\lambda)$ is strictly increasing in $c > 0$.

Let e be the vector $(1,\ldots,1) \in \mathbb{R}^n$, and $F^{-1}([0, F(e)]) = \{\lambda \in \mathbb{R}^n_+ : F(\lambda) \leq F(e)\}$ be the preimage of $[0, F(e)]$, then $F^{-1}([0, F(e)])$ can be shown to be compact, convex and with non-empty interior. To this end, suppose that $x, y \in F^{-1}([0, F(e)])$, then by the convexity of the function F we get that, for any $\alpha \in [0, 1]$,

$$0 \leq F((1-\alpha)x + \alpha y) \leq (1-\alpha)F(x) + \alpha F(y) \leq F(e),$$

which means that $(1-\alpha)x + \alpha y \in F^{-1}([0, F(e)])$. Thus $F^{-1}([0, F(e)])$ is convex. From the continuity of F, it is trivial to prove that the set $F^{-1}([0, F(e)])$ is compact and has nonempty interior.

The radially monotonicity of the function F and convexity of $F^{-1}([0, F(e)])$ implies that the set $\{\lambda \in \mathbb{R}^n_+ : F(\lambda) = F(e)\}$ belongs to the boundary of $F^{-1}([0, F(e)])$. In particular, $e \in \partial F^{-1}([0, F(e)])$. Choosing an open neighborhood U of e such that for any $\lambda = (\lambda_1, \ldots, \lambda_n) \in U$, $\prod_{i=1}^n \lambda_i \gneqq 0$, we claim that F is smooth on the surface piece $U \cap F^{-1}([0, F(e)])$. By the convexity of F, it is sufficient to show that all partial derivatives of F exist on U (see e.g. [6, Theorem 2.5]). To this end, choose an appropriate $h \in \mathbb{R} \setminus \{0\}$ and for $i \in \{1, \ldots, n\}$ denote Λ_h the diagonal matrix with the same entries as $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ except for that the ith entry λ_i replaced by $\lambda_i + h$. Then the triangle inequality gives

$$\big|\, |\Lambda_h y| - |\Lambda y| \,\big| \le |(\Lambda_h - \Lambda)y| \le |h|\,|y| \le |h| \max_{\operatorname{supp} f_Y} |y|.$$

Observing that the convex function ϕ is Lipschitzian on any compact subset of U, from the boundedness of $\mathrm{E}_\phi(|Y|)$ and $\max_{\operatorname{supp} f_Y} |y|$, we see that

$$\frac{\phi(|\Lambda_h y|/\mathrm{E}_\phi(|Y|)) - \phi(|\Lambda y|/\mathrm{E}_\phi(|Y|))}{h}$$

is bounded for all $y \in \operatorname{supp} f_Y$. By applying the Dominated Convergence Theorem, we get that

$$\begin{aligned}
\frac{\partial}{\partial \lambda_i} F(\lambda) &= \int_{\mathbb{R}^n} \frac{\partial}{\partial \lambda_i} \phi \left(\frac{\left(\sum_{i=1}^n \lambda_i^2 (e_i \cdot y)^2\right)^{1/2}}{\mathrm{E}_\phi(|Y|)} \right) f_Y(y)dy \\
&= \int_{\mathbb{R}^n} \phi' \left(\frac{\left(\sum_{i=1}^n \lambda_i^2 (e_i \cdot y)^2\right)^{1/2}}{\mathrm{E}_\phi(|Y|)} \right) \frac{\lambda_i (e_i \cdot y)^2}{\mathrm{E}_\phi(|Y|) \left(\sum_{i=1}^n \lambda_i^2 (e_i \cdot y)^2\right)^{1/2}} f_Y(y)dy
\end{aligned}$$

holds for almost every $y \in \operatorname{supp} f_Y$. Thus all of the partial derivatives of F exist on U and then F is smooth on the surface piece $U \cap F^{-1}([0, F(e)])$. In particular, this observation together with (21) and (22) gives

$$\begin{aligned}
\left.\frac{\partial}{\partial \lambda_i} F(\lambda)\right|_{\lambda = e} &= \int_{\mathbb{R}^n} \frac{(e_i \cdot y)^2}{|y|^2} \frac{\phi'\left(\frac{|y|}{\mathrm{E}_\phi(|Y|)}\right)|y|}{\mathrm{E}_\phi(|Y|)} f_Y(y)dy \\
&= \frac{\mathrm{E}\left[\phi'\left(\frac{|Y|}{\mathrm{E}_\phi(|Y|)}\right)|Y|\right]}{\mathrm{E}_\phi(|Y|)} \int_{\mathbb{R}^n} \frac{(e_i \cdot y)^2}{|y|^2} d\mu(y) \\
&= \frac{\mathrm{E}\left[\phi'\left(\frac{|Y|}{\mathrm{E}_\phi(|Y|)}\right)|Y|\right]}{n\mathrm{E}_\phi(|Y|)},
\end{aligned}$$

which is a constant independent of the index i. Thus the gradient of F at the point e is given by

$$\nabla F(e) = \frac{\mathrm{E}\left[\phi'\left(\frac{|Y|}{\mathrm{E}_\phi(|Y|)}\right)|Y|\right]}{n\mathrm{E}_\phi(|Y|)} e,$$

which implies that the vector e is an outer normal of $F^{-1}([0, F(e)])$ at the point $e \in \partial F^{-1}([0, F(e)])$. Since $F^{-1}([0, F(e)])$ is convex, it is contained in the half-space $\{x \in \mathbb{R}^n : x \cdot e \le e \cdot e\}$. That is, for each $\lambda \in \mathbb{R}^n_+$,

$$F(\lambda) \le F(e) \quad \Rightarrow \quad \lambda \cdot e \le n. \tag{24}$$

On the other hand, for $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n_+$ with $\prod_{i=1}^n \lambda_i = 1$, the inequality of arithmetic and geometric means gives

$$\lambda \cdot e \ge n, \tag{25}$$

with equality if and only if $\lambda = e$.

Combining (24) and (25) we see that, for all $\lambda \in \mathbb{R}^n_+$ with $\prod_{i=1}^n \lambda_i = 1$, it follows that

$$F(\lambda) \ge F(e),$$

with the equality holding if and only if $\lambda = e$. That confirms the desired claim. □

4. Specific ϕ-covariance matrices of a random vector

In this section we present some specific examples of general covariance matrices implied by the unified covariance matrices given by Theorem 3.1.

4.1. The pth covariance matrix

The following is the pth covariance matrix of a random vector, which was found by Lutwak et al. [15].

Corollary 4.1. *If $p \in (1, \infty)$ and X is a non-degenerate random vector in $\mathbb{R}^n$ with finite pth moment, then there exists a unique matrix $A \in S$ such that*

$$\mathrm{E}[\|X\|_A^p] = n$$

and

$$\det A \geq \det A'$$

for each $A' \in S$ such that $\mathrm{E}[\|X\|_{A'}^p] = n$. Moreover, the matrix A is the unique matrix in S satisfying

$$I_n = \mathrm{E}[|AX|^{p-2}(AX) \otimes (AX)].$$

Proof. For $t \geq 0$, let $\phi(t) = t^p/p$ in Theorem 3.1. Then ϕ is a Gaussian gauge whenever $p > 1$.

It is easily seen that $\phi(t) = t^p/p$ satisfies the Δ_2-condition.

Now a straightforward computation shows that the corollary is an immediate consequence of Theorem 3.1. □

4.2. The logistic covariance matrix

If a Gaussian gauge $\varphi : (0, \infty) \to (\ln 2, \infty)$ is given by

$$\varphi(t) = t + 2\ln\frac{1 + e^{-t}}{\sqrt{2}}, \tag{26}$$

then the probability distribution for any constant multiple of the φ-Gaussian Z_φ is known as a logistic distribution (see e.g. [20]). In fact,

$$\int_0^\infty e^{-\varphi(t)}dt = 2\int_0^\infty \frac{e^{-t}}{(1 + e^{-t})^2}dt = 1,$$

which implies that $f_{Z_\varphi}(x) = e^{-\varphi(|x|)} = 2e^{-|x|}/(1+e^{-|x|})^2$ is the density function of a logistic distribution, where $x \in \mathbb{R}^n \setminus \{0\}$.

Thanks to the fact that $(1 + e^{-2t}) \leq (1 + e^{-t})^2$ and that φ is strictly increasing, we have

$$\begin{aligned}
\varphi(2t) &= 2(t + \ln(1 + e^{-2t})) - \ln 2 \\
&\leq 2(t + 2\ln(1 + e^{-t})) - \ln 2 \\
&= 2\varphi(t) + \ln 2 \\
&\leq 3\varphi(t).
\end{aligned}$$

Thus the Gaussian gauge $\varphi : (0, \infty) \to (\ln 2, \infty)$ defined by (26) satisfies the Δ_2-condition.

A straightforward calculation shows that $\hat{\varphi} = \mathrm{E}[\varphi(Z_\varphi)] = 2 - \ln 2$.

Let X be a random vector in $\mathbb{R}^n$. We define the logistic moment $\mathrm{E}_\varphi(|X|)$ of X implicitly by

$$\mathrm{E}\left[\frac{|X|}{\mathrm{E}_\varphi(|X|)} + 2\ln\left(1 + e^{-\frac{|X|}{\mathrm{E}_\varphi(|X|)}}\right)\right] = 2.$$

The above discussion together with Theorem 3.1 characterizes the logistic covariance matrix of a random vector in $\mathbb{R}^n$:

Corollary 4.2. *Suppose that $\varphi(t) = t + 2\ln\frac{1+e^{-t}}{\sqrt{2}}$ for $t \geq 0$, and X is a non-degenerate random vector in $\mathbb{R}^n$ with finite logistic moment, then there exists $A \in S$, unique up to a scalar multiple, such that*

$$\mathrm{E}_\varphi(\|X\|_A) = n$$

and

$$\det A \geq \det A'$$

for each $A' \in S$ with $\mathrm{E}_\varphi(\|X\|_{A'}) = n$. Moreover, if A is such a unique matrix in S, then it satisfies

$$\mathrm{E}\left[\frac{1 - e^{-|AX|/n}}{1 + e^{-|AX|/n}}\frac{AX \otimes AX}{|AX|}\right] = \frac{1}{n}\mathrm{E}\left[\frac{1 - e^{-|AX|/n}}{1 + e^{-|AX|/n}}|AX|\right]I_n.$$

4.3. The ϕ-covariance matrix for the power function distribution

We say a random vector X in $\mathbb{R}^n$ *fulfilling the power function distribution criterion*, if $|X|$ is a continuous random variable with density g and satisfying that:

$$P[0 < |X| < 1] = 1;$$
$$g : [0, 1] \to [0, \infty) \quad \text{is continuous;}$$
$$g \quad \text{is continuously differentiable on } (0, 1);$$
$$g' > 0 \quad \text{on } (0, 1);$$
$$\mathrm{E}[\ln |X|] > -1;$$
$$-\infty < \mathrm{E}[\ln(g'(|X|)/g(|X|))] < \infty;$$
$$\ln\left(-1 - \frac{1}{\mathrm{E}[\ln |X|]}\right) \leq \mathrm{E}[\ln |X|] + \mathrm{E}[\ln(g'(|X|)/g(|X|))];$$
$$\inf_{0<t<1} \ln(g'(t)/g(t)) \geq \mathrm{E}[\ln |X|] + \mathrm{E}[\ln(g'(|X|)/g(|X|))].$$

For a random vector X fulfilling the power function distribution criterion, Cianchi et al. [1] established the following Cramér–Rao inequality for the power function distribution:

$$\mathrm{E}[\ln |X|] + \mathrm{E}[\ln(g'(|X|)/g(|X|))] \leq \ln(g(1) - 1),$$

with equality if and only if $g(t) = (p + 1)t^p$, where $p > 0$ satisfies

$$p + 1 = -\frac{1}{\mathrm{E}[\ln |X|]}. \tag{27}$$

Given a random vector X in $\mathbb{R}^n$ satisfying the above assumptions, let $p > 0$ be given by (27). We follow Cianchi et al. [1] to define a gauge $\phi : (0, 1) \to \mathbb{R}$ by

$$\phi(t) = -p \ln t. \tag{28}$$

The ϕ-Gaussian Z_ϕ has density function

$$f_{Z_\phi}(x) = \begin{cases} (p + 1)|x|^p, & \text{if } x \in \{y \in \mathbb{R}^n : 0 < |y| < 1\}; \\ 0, & \text{otherwise.} \end{cases}$$

A straightforward calculation gives that $\hat{\phi} = p/(p + 1)$ and that if $\mathrm{E}[\ln |X|] > -1$, then X has a finite ϕ-moment $\mathrm{E}_\phi(|X|)$, where $\mathrm{E}_\phi(|X|)$ is determined by

$$\ln \mathrm{E}_\phi(|X|) - \mathrm{E}[\ln |X|] = \frac{1}{p + 1}.$$

In particular, this together with (27) deduces that if $\mathrm{E}[\ln |X|] > -1$, then

$$\mathrm{E}_\phi(|X|) = 1.$$

Observing that $\phi(2t) = -p \ln 2 - p \ln t < 0 < \phi(t)$, i.e., the Gaussian gauge ϕ satisfies the Δ_2-condition, from Theorem 3.1 we obtain that

Corollary 4.3. *Suppose that the Gaussian gauge $\phi : (0, 1) \to \mathbb{R}$ is given by $\phi(t) = -p \ln t$, and X is a non-degenerate random vector in $\mathbb{R}^n$ fulfilling the power function distribution criterion, then there exists a unique $A \in S$, such that AX fulfils the power function distribution criterion and*

$$\mathrm{E}_\phi(\|X\|_A) = 1;$$
$$\det A \geq \det A',$$

for each $A' \in S$ satisfying that $A'X$ fulfils the power function distribution criterion and $\mathrm{E}_\phi(\|X\|_{A'}) = 1$. Moreover, if A is such a unique matrix in S, then it satisfies

$$\mathrm{E}\left[\frac{AX \otimes AX}{|AX|^2}\right] = \frac{1}{n} I_n.$$

References

[1] A. Cianchi, E. Lutwak, D. Yang, G. Zhang, A unified approach to Cramér–Rao inequalities, IEEE Trans. Inform. Theory 60 (2014) 643–650.
[2] D. Cordero-Erausquin, B. Nazaret, C. Villani, A mass-transportation approach to sharp Sobolev and Gagliardo–Nirenberg inequalities, Adv. Math. 182 (2004) 307–332.
[3] J.A. Costa, A.O. Hero, C. Vignat, A characterization of the multivariate distributions maximizing Renyi entropy, in: Proc. 2002 IEEE Int. Symp. Information Theory, Lausanne, Switzerland, June/July, 2002, p. 263.
[4] J. Decani, R. Stine, A note on deriving the information matrix for a logistic distribution, Amer. Statist. 40 (1986) 220–222.

[5] M. Del Pino, J. Dolbeault, Best constants for Gagliardo–Nirenberg inequalities and applications to nonlinear diffusions, J. Math. Pures Appl. 81 (2002) 847–875.
[6] P.M. Gruber, Convex and Discrete Geometry, in: Grundlehren Math. Wiss., vol. 336, Springer, Berlin, 2007.
[7] O. Johnson, C. Vignat, Some results concerning maximum Rényi entropy distributions, Ann. Inst. H. Poincaré Probab. Statist. 43 (3) (2007) 339–351.
[8] C.P. Kitsos, N.K. Tavoularis, Logarithmic Sobolev inequalities for information measures, IEEE Trans. Inform. Theory 55 (6) (2009) 2554–2561.
[9] M. Ludwig, Covariance matrices and valuations, Adv. Appl. Math. 51 (2013) 359–366.
[10] E. Lutwak, S. Lv, D. Yang, G. Zhang, Extenstions of Fisher information and Stam's inequality, IEEE Trans. Inform. Theory 58 (2012) 1319–1327.
[11] E. Lutwak, S. Lv, D. Yang, G. Zhang, Affine moments of a random vector, IEEE Trans. Inform. Theory 59 (2013) 5592–5599.
[12] E. Lutwak, D. Yang, G. Zhang, Moment–entropy inequalities, Ann. Probab. 32 (2004) 757–774.
[13] E. Lutwak, D. Yang, G. Zhang, Cramér–Rao and moment-entropy inequalities for Renyi entropy and generalized Fisher information, IEEE Trans. Inform. Theory 51 (2005) 473–478.
[14] E. Lutwak, D. Yang, G. Zhang, Optimal Sobolev norms and the L_p Minkowski problem, Int. Math. Res. Not. 2006 (2006) 1–21.
[15] E. Lutwak, D. Yang, G. Zhang, Moment–entropy inequalities for a random vector, IEEE Trans. Inform. Theory 53 (2007) 1603–1607.
[16] S. Lv, Affine isoperimetry and information theoritic inequalities (Ph.D. thesis), Polytechnic Institute of New York University, 2012.
[17] S. Lv, X. Lv, Affine Fisher information inequalities, J. Math. Anal. Appl. 371 (2010) 347–354.
[18] H. Nicholas, Computing the nearest correlation matrix—a problem from finance, IMA J. Numer. Anal. 22 (2002) 329–343.
[19] M.M. Rao, Z.D. Ren, Theory of Orlicz spaces, in: Monographs and Textbooks in Pure and Applied Mathematics, vol. 146, Marcel Dekker, New York, 1991.
[20] Wikipedia. Logistic distribution. Available at: http://en.wikipedia.org/wiki/Logistic_distribution.