A Family of Kurtosis Orderings for Multivariate Distributions

Jin Wang[1]

Northern Arizona University

July 2008

Final preprint version for *Journal of Multivariate Analysis*, 2008, to appear

[1]Department of Mathematics and Statistics, Northern Arizona University, Flagstaff, Arizona 86011-5717, USA. E-mail: jin.wang@nau.edu.

Supported by IGP of Northern Arizona University.

Abstract

In this paper, a family of kurtosis orderings for multivariate distributions is proposed and studied. Each ordering characterizes in an affine invariant sense the movement of probability mass from the "shoulders" of a distribution to either the center or the tails or both. All even moments of the Mahalanobis distance of a random vector from its mean (if exists) preserve a subfamily of the orderings. For elliptically symmetric distributions, each ordering determines the distributions up to affine equivalence. As applications, the orderings are used to study elliptically symmetric distributions. Ordering results are established for three important families of elliptically symmetric distributions: Kotz type distributions, Pearson Type VII distributions, and Pearson Type II distributions.

AMS 2000 Subject Classification: Primary 62G05 Secondary 62H05.

Key words and phrases: Kurtosis; Peakedness; Tailweight; Ordering; Elliptically symmetric distributions.

1 Introduction and preliminaries

Up to now, many multivariate kurtosis measures have been proposed (see, e.g., Mardia [10], Oja [13], Srivastava [16], Averous and Meste [1], Liu, Parelius and Singh [9], Serfling [15], and Wang and Serfling [18]). The classical notion of multivariate kurtosis is moment-based, given (Mardia [10]) by the fourth moment of the Mahalanobis distance of a random vector $\boldsymbol{X}$ in $\mathbb{R}^d$ from its mean $\boldsymbol{\mu}$, i.e.,

$$k_d = E[(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})]^2.$$

k_d measures the dispersion of $\boldsymbol{X}$ about the ellipsoid $(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = d$, which defines the "shoulders" of the distribution. Higher kurtosis arises when probability mass is diminished near the shoulders and greater either near $\boldsymbol{\mu}$ (greater peakedness), or greater in the tails (greater tailweight), or both. See Wang and Serfling [18] for detailed discussion. Since the pioneering work of Bickel and Lehmann [3] [4] and Oja [12] about descriptive statistics, it has been commonly admitted that the meaning of a descriptive concept of distributions is given by an ordering and that measures for this concept are meaningful only if they preserve the ordering. While univariate kurtosis orderings and their applications have received considerable attention, kurtosis orderings for multivariate distributions have received relatively little investigation. There has been not even a multivariate kurtosis ordering for the classical multivariate kurtosis measure k_d up to now. Multivariate kurtosis measures are usually developed by intuition. It is necessary to study multivariate kurtosis by the ordering approach. That is the motivation of this work.

For the univariate case, van Zwet [17] defined a kurtosis ordering $\leq_s$ (s-ordering) for univariate symmetric distributions:

$$F_X \leq_s G_Y \textit{ iff } G_Y^{-1}(F_X(x)) \text{ is convex for } x > \mu_F,$$

where μ_F is the point of symmetry of F_X. Using the folded distributions $F_{|X-\mu_X|}$ and $G_{|Y-\mu_Y|}$, Oja [12] gave an equivalent definition of the s-ordering:

$$F_X \leq_k G_Y \textit{ iff } G_{|Y-\mu_Y|}^{-1}(F_{|X-\mu_X|}(z)) \text{ is convex for } z \geq 0, \text{ i.e., } F_{|X-\mu_X|} \leq_c G_{|Y-\mu_Y|},$$

where $\leq_c$ is the van Zwet [17] skewness ordering for univariate distributions. This definition was extended by Balanda and MacGillivray [2] to include the case of univariate asymmetric distributions with finite mean. They called $F_{|X-\mu_X|}$ the moment–based spread function. To allow the use of other location measures instead of the mean μ_X, we will call it the distribution-based spread function. Balanda and MacGillivray [2] also studied various univariate kurtosis orderings by the quantile–based spread function $S_F(p) = F^{-1}(\frac{1}{2} + \frac{p}{2}) - F^{-1}(\frac{1}{2} - \frac{p}{2})$. In fact, the inverse function S_F^{-1} of S_F is a distribution function and can be considered as a distribution-based spread function. Extending S_F^{-1} to the multivariate case, Averous and Meste [1] defined the multivariate kurtosis orderings in L_1-sense.

Generally we should use a standardized version of a random variable or a random vector when we study kurtosis. Then any univariate skewness ordering on the standardized distribution-based spread functions will yield a kurtosis ordering for the underlying distributions. In this paper, we will develop multivariate kurtosis orderings by this approach. A family of kurtosis orderings for multivariate distributions is defined and studied in Section 2. In Section 3, the orderings are used to study elliptically symmetric distributions. Ordering results are established for three important families of elliptically symmetric distributions: Kotz type distributions, Pearson Type VII distributions, and Pearson Type II distributions. Concluding remarks are given in Section 4. Before going to our main topic, we give some preliminaries.

A real-valued function f defined on an interval is called *convex*, if for any two points x_1 and x_2 in its domain and any $\lambda \in (0,1)$,

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2).$$

If the inequality above is strict for all x_1 and x_2 , then $f(x)$ is called *strictly convex.* It is easy to see that $f(x)$ is strictly convex if and only if for any straight line $y = b(x+a)$, $f(x) - b(x+a)$ can have at most two zeros and is negative between these zeros. If $f(x)$ has a second derivative in its domain, then a necessary and sufficient condition for it to be convex is that the second derivative $f^{''}(x) \geq 0$ for all x.

Throughout this paper, we confine attention to continuous distributions. A continuous distribution F in $\mathbb{R}^d$ is called *elliptically symmetric,* denoted by $E_d(h; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, if it has a density of the form

$$f(\boldsymbol{x}) = C|\ \boldsymbol{\Sigma}|^{-1/2} h((\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})), \boldsymbol{x} \in \mathbb{R}^d,$$

for a nonnegative function $h(\cdot)$ with $\int_0^\infty r^{d/2-1}h(r)dr < \infty$ and a positive definite matrix $\boldsymbol{\Sigma}$, where $\boldsymbol{\mu}$ can be viewed as a center and $\boldsymbol{\Sigma}$ as a measure of spread in some sense. If the first moment of F exists, $\boldsymbol{\mu}$ is the mean vector. If the second moment of F exists, then the covariance matrix is $k\boldsymbol{\Sigma}$ for some positive constant k. $f(\boldsymbol{x})$ is unimodal if $h(\cdot)$ is decreasing, uniform if $h(\cdot)$ is constant, and bowl-shaped if $h(\cdot)$ is increasing. Let $R = [(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})]^{1/2}$. Then we have the following result.

Lemma 1.1 *Suppose that* $\boldsymbol{X} \sim E_d(h;\boldsymbol{\mu},\boldsymbol{\Sigma})$. *Then the density of* R^α *is*

$$f_{R^\alpha}(r) = \frac{2C\pi^{d/2}}{\alpha\Gamma(d/2)} r^{d/\alpha-1} h(r^{2/\alpha}), r \geq 0. \tag{1.1}$$

Proof. It is well known that the density of R^2 is

$$f_{R^2}(r) = \tfrac{C\pi^{d/2}}{\Gamma(d/2)} r^{d/2-1} h(r), r \geq 0.$$

Then a transformation leads to the result. ■

2 A family of multivariate kurtosis orderings

In this section, we first propose a family of kurtosis orderings for univariate distributions. Then the multivariate extension of the family is given and studied.

2.1 A family of univariate kurtosis orderings

For any $\alpha > 0$, $|X-\mu_X|^\alpha$ can be interpreted as a spread (or dispersion) of a random variable X. Thus the distribution function $F_{|X-\mu_X|^\alpha}$ of $|X-\mu_X|^\alpha$ can be considered as a distribution-based spread function. If a standardized version $\frac{X-\mu_X}{\sigma_X}$ of X is used, the distribution-based spread function becomes $F_{|\frac{X-\mu_X}{\sigma_X}|^\alpha}$, where μ_X and σ_X can be any corresponding location and spread measures, for example, the moment-based location and spread measures, the quantile-based location and spread measures, and so on. Then applying the van Zwet [17] skewness ordering to the standardized distribution-based spread functions yields the following family of kurtosis orderings for univariate distributions:

$$\text{For any } \alpha > 0, F_X \leqslant_{s_\alpha} G_Y \text{ iff } G^{-1}_{|\frac{Y-\mu_Y}{\sigma_Y}|^\alpha}(F_{|\frac{X-\mu_X}{\sigma_X}|^\alpha}(z)) \text{ is convex for } z \geq 0.$$

When the moment-based location and spread measures are used, the s_1-ordering is equivalent to the Balanda and MacGillivray [2] kurtosis ordering. For univariate symmetric distributions, the s_1-ordering reduces to the van Zwet [17] kurtosis ordering.

2.2 Extension to the multivariate case

For a random vector $\boldsymbol{X}$ in $\mathbb{R}^d$, let $R_{\boldsymbol{X}} = [(\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}})'\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}(\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}})]^{1/2}$, the Mahalanobis distance of $\boldsymbol{X}$ from $\boldsymbol{\mu}_{\boldsymbol{X}}$, where $\boldsymbol{\mu}_{\boldsymbol{X}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{X}}$ are any corresponding location and covariance measures of $\boldsymbol{X}$, for example, the moment-based location and covariance measures, the depth-based location and covariance measures, and so on. Then a natural multivariate extension of the univariate kurtosis orderings $\leqslant_{s_\alpha}$ is based on $F_{R^\alpha_{\boldsymbol{X}}}$ and $G_{R^\alpha_{\boldsymbol{Y}}}$, which are the distribution functions of $R^\alpha_{\boldsymbol{X}} = [(\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}})'\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}(\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}})]^{\alpha/2}$ and $R^\alpha_{\boldsymbol{Y}} = [(\boldsymbol{Y}-\boldsymbol{\mu}_{\boldsymbol{Y}})'\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}(\boldsymbol{Y}-\boldsymbol{\mu}_{\boldsymbol{Y}})]^{\alpha/2}$, respectively.

Definition 2.1 *For $\alpha > 0$, we say that $F_{\boldsymbol{X}}$ is k_α less than or equal to $G_{\boldsymbol{Y}}$ in kurtosis, denoted by $F_{\boldsymbol{X}} \leqslant_{k_\alpha} G_{\boldsymbol{Y}}$, if $G^{-1}_{R^\alpha_{\boldsymbol{Y}}}(F_{R^\alpha_{\boldsymbol{X}}}(r))$ is convex for $r \geq 0$, and $F_{\boldsymbol{X}}$ is k_α less than $G_{\boldsymbol{Y}}$ in kurtosis, denoted by $F_{\boldsymbol{X}} <_{k_\alpha} G_{\boldsymbol{Y}}$, if $G^{-1}_{R^\alpha_{\boldsymbol{Y}}}(F_{R^\alpha_{\boldsymbol{X}}}(r))$ is strictly convex for $r \geq 0$. We say that $F_{\boldsymbol{X}}$ is k_α equal to $G_{\boldsymbol{Y}}$ in kurtosis, denoted by $F_{\boldsymbol{X}} =_{k_\alpha} G_{\boldsymbol{Y}}$, if $F_{\boldsymbol{X}} \leq_{k_\alpha} G_{\boldsymbol{Y}}$ and $G_{\boldsymbol{Y}} \leq_{k_\alpha} F_{\boldsymbol{X}}$.*

Now we study the relationship of the k_α-orderings for different α values. We assume that $G^{-1}_{R^\alpha_{\boldsymbol{Y}}}(F_{R^\alpha_{\boldsymbol{X}}}(r))$ has a second derivative.

Theorem 2.1 *For any $\alpha_2 > \alpha_1 > 0, F_{\boldsymbol{X}} \leqslant_{k_{\alpha_1}} G_{\boldsymbol{Y}}$ implies $F_{\boldsymbol{X}} \leqslant_{k_{\alpha_2}} G_{\boldsymbol{Y}}$. Thus, the k_{α_1}-ordering is stronger than the k_{α_2}-ordering.*

Proof. It is sufficient to show that for any $\lambda > 1, R_{\lambda\alpha_1}(r) = G^{-1}_{R^{\lambda\alpha_1}_{\boldsymbol{Y}}}(F_{R^{\lambda\alpha_1}_{\boldsymbol{X}}}(r))$ is convex if $R_{\alpha_1}(r) = G^{-1}_{R^{\alpha_1}_{\boldsymbol{Y}}}(F_{R^{\alpha_1}_{\boldsymbol{X}}}(r))$ is convex.

By transformations, we see that $F_{R^{\lambda\alpha_1}_{\boldsymbol{X}}}(r) = F_{R^{\alpha_1}_{\boldsymbol{X}}}(r^{1/\lambda})$ and $G_{R^{\lambda\alpha_1}_{\boldsymbol{Y}}}(r) = G_{R^{\alpha_1}_{\boldsymbol{Y}}}(r^{1/\lambda})$. Thus,

$$R_{\lambda\alpha_1}(r) = G^{-1}_{R^{\lambda\alpha_1}_{\boldsymbol{Y}}}(F_{R^{\lambda\alpha_1}_{\boldsymbol{X}}}(r)) = [G^{-1}_{R^{\alpha_1}_{\boldsymbol{Y}}}(F_{R^{\alpha_1}_{\boldsymbol{X}}}(r^{1/\lambda}))]^\lambda = (R_{\alpha_1}(r^{1/\lambda}))^\lambda,$$

$$\frac{d}{dr}R_{\lambda\alpha_1}(r) = (R_{\alpha_1}(r^{1/\lambda}))^{\lambda-1}R'_{\alpha_1}(r^{1/\lambda})r^{1/\lambda-1},$$

$$\begin{aligned}\frac{d^2}{dr^2}R_{\lambda\alpha_1}(r) &= \frac{1}{\lambda}(R_{\alpha_1}(r^{1/\lambda}))^{\lambda-1}R''_{\alpha_1}(r^{1/\lambda})(r^{1/\lambda-1})^2 \\ &\quad +(1-\frac{1}{\lambda})(R_{\alpha_1}(r^{1/\lambda}))^{\lambda-2}R'_{\alpha_1}(r^{1/\lambda})r^{1/\lambda-2}[R'_{\alpha_1}(r^{1/\lambda})r^{1/\lambda} - R_{\alpha_1}(r^{1/\lambda})].\end{aligned}$$

Since $R_{\alpha_1}(0) = 0$, by the mean value theorem, $R_{\alpha_1}(r^{1/\lambda}) = R'_{\alpha_1}(\xi)r^{1/\lambda}$, where $\xi \in [0, r^{1/\lambda}]$. Thus, $R'_{\alpha_1}(r^{1/\lambda})r^{1/\lambda} - R_{\alpha_1}(r^{1/\lambda}) = [R'_{\alpha_1}(r^{1/\lambda}) - R'_{\alpha_1}(\xi)]r^{1/\lambda} \geq 0$ if $R_{\alpha_1}(r)$

is convex. In addition, $R_{\alpha_1}(r) \geq 0$ and $R'_{\alpha_1}(r) = f_{R_{\boldsymbol{X}}^{\alpha_1}}(r)/g_{R_{\boldsymbol{Y}}^{\alpha_1}}(R_{\alpha_1}(r)) \geq 0$, where $f_{R_{\boldsymbol{X}}^{\alpha_1}}$ and $g_{R_{\boldsymbol{Y}}^{\alpha_1}}$ are density functions of $R_{\boldsymbol{X}}^{\alpha_1}$ and $R_{\boldsymbol{Y}}^{\alpha_1}$. Thus, $\frac{d^2}{dr^2}R_{\lambda\alpha_1}(r) \geq 0$ for $r \geq 0$ if $R_{\alpha_1}(r)$ is convex. This completes the proof. ■

Remark 2.1 *From Theorem 2.1, it is seen that the strength of the k_α-ordering decreases as α increases. For a particular application, one can select an ordering with appropriate strength by α. See Section 3 for details.*

2.3 Properties of the k_α-orderings

It is easy to see that for any $\alpha > 0$, the relation $\leqslant_{k_\alpha}$ is symmetrical, reflexive, and transitive. Thus, it is a partial ordering. Some very important properties of the k_α-orderings are established in the following results. For simplicity, we assume that $\boldsymbol{\mu}_{\boldsymbol{X}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{X}}$ are the mean vector and covariance matrix of $\boldsymbol{X}$, and $\boldsymbol{\mu}_{\boldsymbol{Y}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{Y}}$ are the mean vector and covariance matrix of $\boldsymbol{Y}$. However, some results hold generally as long as the location and covariance measures used in the k_α-orderings satisfy some conditions.

Theorem 2.2 *If $F_{\boldsymbol{X}} \leqslant_{k_\alpha} G_{\boldsymbol{Y}}$, then $F_{\mathbf{A}_1\boldsymbol{X}+\boldsymbol{b}_1} \leqslant_{k_\alpha} G_{\mathbf{A}_2\boldsymbol{Y}+\boldsymbol{b}_2}$ for any nonsingular $d \times d$ matrices $\mathbf{A}_1, \mathbf{A}_2$ and vectors $\boldsymbol{b}_1, \boldsymbol{b}_2$ in $\mathbb{R}^d$, that is, each k_α-ordering is affine invariant.*

Proof. Since the Mahalanobis distance of a random vector from its mean vector is affine invariant, the result follows. ■

Our next result establishes the sufficient and necessary condition for $F_{\boldsymbol{X}} =_{k_\alpha} G_{\boldsymbol{Y}}$.

Theorem 2.3 *(1)* $F_{\boldsymbol{X}} =_{k_\alpha} G_{\boldsymbol{Y}}$ *iff* $F_{R^\alpha_{\boldsymbol{X}}}(r) = G_{R^\alpha_{\boldsymbol{Y}}}(cr)$ *for some positive constant* c.

(2) If $F_{\boldsymbol{X}}$ *and* $G_{\boldsymbol{Y}}$ *are continuous elliptically symmetric distributions, then* $F_{\boldsymbol{X}} =_{k_\alpha} G_{\boldsymbol{Y}}$ *iff* $\mathbf{X}$ *and* $\mathbf{Y}$ *are affinely equivalent in distribution:* $\mathbf{Y} \overset{d}{=} \mathbf{A}\mathbf{X} + \boldsymbol{b}$ *for some nonsingular* $d \times d$ *matrix* $\mathbf{A}$ *and some vector* $\boldsymbol{b}$ *in* $\mathbb{R}^d$*. Thus,* $F_{R^\alpha_{\boldsymbol{X}}}(r) = G_{R^\alpha_{\boldsymbol{Y}}}(r)$.

Proof. (1) $F_{\boldsymbol{X}} =_{k_\alpha} G_{\boldsymbol{Y}}$ iff $G^{-1}_{R^\alpha_{\boldsymbol{Y}}}(F_{R^\alpha_{\boldsymbol{X}}}(r)) = cr + a$. Since $G^{-1}_{R^\alpha_{\boldsymbol{Y}}}(F_{R^\alpha_{\boldsymbol{X}}}(r)) \geq 0$ and $G^{-1}_{R^\alpha_{\boldsymbol{Y}}}(F_{R_{\boldsymbol{X}}}(0)) = 0$, $c > 0$ and $a = 0$. Thus, $F_{R^\alpha_{\boldsymbol{X}}}(r) = G_{R^\alpha_{\boldsymbol{Y}}}(cr)$.

(2) Let the densities of $\boldsymbol{X}$ and $\boldsymbol{Y}$ be respectively

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \tfrac{C_1}{|\boldsymbol{\Sigma}_1|^{1/2}} h_1((\boldsymbol{x}-\boldsymbol{\mu}_1)'\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)),\ \boldsymbol{x} \in \mathbb{R}^d,$$

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \tfrac{C_2}{|\boldsymbol{\Sigma}_2|^{1/2}} h_2((\boldsymbol{y}-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_2)),\ \boldsymbol{y} \in \mathbb{R}^d.$$

Then by Lemma 1.1 the densities of $R^\alpha_{\boldsymbol{X}} = [(\boldsymbol{X}-\boldsymbol{\mu}_1)'\boldsymbol{\Sigma}_1^{-1}(\ \boldsymbol{X}-\boldsymbol{\mu}_1)]^{\alpha/2}$ and $R^\alpha_{\boldsymbol{Y}} = [(\boldsymbol{Y}-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_2^{-1}(\ \boldsymbol{Y}-\boldsymbol{\mu}_2)]^{\alpha/2}$ are

$$f_{R^\alpha_{\boldsymbol{X}}}(r) = \tfrac{2C_1\pi^{d/2}}{\alpha\Gamma(d/2)} r^{d/\alpha-1} h_1(r^{2/\alpha}),\ r \geq 0,$$

$$g_{R^\alpha_{\boldsymbol{Y}}}(r) = \tfrac{2C_2\pi^{d/2}}{\alpha\Gamma(d/2)} r^{d/\alpha-1} h_2(r^{2/\alpha}),\ r \geq 0.$$

From (1), $F_{R^\alpha_{\boldsymbol{X}}}(r) = G_{R^\alpha_{\boldsymbol{Y}}}(cr)$ for some $c > 0$, which implies $f_{R^\alpha_{\boldsymbol{X}}}(r) = c\ g_{R^\alpha_{\boldsymbol{Y}}}(cr)$, *i.e.*,

$$\tfrac{2C_1\pi^{d/2}}{\alpha\Gamma(d/2)} r^{d/\alpha-1} h_1(r^{2/\alpha}) = c\tfrac{2C_2\pi^{d/2}}{\alpha\Gamma(d/2)} (cr)^{d/\alpha-1} h_2((cr)^{2/\alpha}).$$

Thus, $h_1(r^{2/\alpha}) = \frac{C_2}{C_1} c^{d/\alpha} h_2((cr)^{2/\alpha}), r \geq 0$, equivalently, $h_1(z) = \frac{C_2}{C_1} c^{d/\alpha} h_2(c^{2/\alpha} z)$, $z \geq 0$. Therefore,

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{C_2}{|\boldsymbol{\Sigma}_1|^{1/2}} c^{d/\alpha} h_2(c^{2/\alpha}(\boldsymbol{x}-\boldsymbol{\mu}_1)'\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)),$$

which leads to

$$\mathbf{\Sigma}_2^{-1/2}(\boldsymbol{Y}-\boldsymbol{\mu}_2) \stackrel{d}{=} c^{1/\alpha}\mathbf{\Sigma}_1^{-1/2}(\boldsymbol{X}-\boldsymbol{\mu}_1),$$

i.e.,

$$\boldsymbol{Y} \stackrel{d}{=} c^{1/\alpha}\mathbf{\Sigma}_2^{1/2}\mathbf{\Sigma}_1^{-1/2}\mathbf{X} + (\boldsymbol{\mu}_2 - c^{1/\alpha}\mathbf{\Sigma}_2^{1/2}\mathbf{\Sigma}_1^{-1/2}\boldsymbol{\mu}_1).$$

By Theorem 2.2 , $F_{R^\alpha_{\boldsymbol{X}}}(r) = G_{R^\alpha_{\boldsymbol{Y}}}(r)$. This completes the proof. ■

Theorem 2.4 *(1) If* $F_{\boldsymbol{X}} \leqslant_{k_\alpha} G_{\boldsymbol{Y}}$*, then there exists* $r_0 > 0$ *such that* $F_{R^\alpha_{\boldsymbol{X}}}(r) \leqslant G_{R^\alpha_{\boldsymbol{Y}}}(r)$ *for* $0 \leqslant r < r_0$ *and* $F_{R^\alpha_{\boldsymbol{X}}}(r) \geq G_{R^\alpha_{\boldsymbol{Y}}}(r)$ *for* $r \geq r_0$.

(2) If $F_{\boldsymbol{X}} \leqslant_{k_\alpha} G_{\boldsymbol{Y}}$ *for* $0 < \alpha \leqslant 2$*, then* $E(R^{2k}_{\boldsymbol{X}}) \leqslant E(R^{2k}_{\boldsymbol{Y}})$ *for* $k = 1, 2, \cdots$ *, as long as they exist.*

Proof. (1) Let $R_\alpha(r) = G^{-1}_{R^\alpha_{\boldsymbol{Y}}}(F_{R^\alpha_{\boldsymbol{X}}}(r))$. Since $G_{R^\alpha_{\boldsymbol{Y}}}(R_\alpha(r)) = F_{R^\alpha_{\boldsymbol{X}}}(r), R^\alpha_{\boldsymbol{Y}}$ has the same distribution as $R_\alpha(R^\alpha_{\boldsymbol{X}})$. If $R_\alpha(r) > r$ for all $r > 0, P(R_\alpha(R^\alpha_{\boldsymbol{X}}) > R^\alpha_{\boldsymbol{X}}) = 1$. Then

$$E(R^2_{\boldsymbol{Y}}) = E((R^\alpha_{\boldsymbol{Y}})^{2/\alpha}) = E((R_\alpha(R^\alpha_{\boldsymbol{X}}))^{2/\alpha}) > E((R^\alpha_{\boldsymbol{X}})^{2/\alpha}) = E(R^2_{\boldsymbol{X}}),$$

which contradicts the fact that $E(R^2_{\boldsymbol{Y}}) = d = E(R^2_{\boldsymbol{X}})$. $R_\alpha(r)$ cannot be less than r for all $r > 0$ either by the same argument. Since $R_\alpha(r)$ is convex, the result follows.

(2) We consider the k_2-ordering first. By (1), there exists $r_0 > 0$ such that $R_2(r) = G^{-1}_{R^2_{\boldsymbol{Y}}}(F_{R^2_{\boldsymbol{X}}}(r)) \leqslant r$ for $0 \leqslant r < r_0$ and $R_2(r) \geq r$ for $r \geq r_0$.

(i) For $r \geqslant r_0$,

$$\sum_{i=0}^{k-1} R_2^i(r) r^{(k-1)-i} \geq \sum_{i=0}^{k-1} r^i r^{(k-1)-i} = kr^{k-1} \geq kr_0^{k-1}.$$

Then,

$$R_2^k(r) - r^k = (R_2(r) - r)(\sum_{i=0}^{k-1} R_2^i(r) r^{(k-1)-i}) \geq kr_0^{k-1}(R_2(r) - r).$$

(ii) For $0 \leqslant r < r_0$,

$$\sum_{i=0}^{k-1} R_2^i(r) r^{(k-1)-i} \leqslant \sum_{i=0}^{k-1} r^i r^{(k-1)-i} = kr^{k-1} < kr_0^{k-1}.$$

Thus,

$$R_2^k(r) - r^k = (R_2(r) - r)(\sum_{i=0}^{k-1} R_2^i(r) r^{(k-1)-i}) \geq kr_0^{k-1}(R_2(r) - r).$$

Combining (i) and (ii), we have

$$R_2^k(r) - r^k \geq kr_0^{k-1}(R_2(r) - r) \text{ for all } r \geq 0.$$

Therefore,

$$\begin{aligned} E(R_{\boldsymbol{Y}}^{2k}) - E(R_{\boldsymbol{X}}^{2k}) &= E((R_2(R_{\boldsymbol{X}}^2))^k) - E((R_{\boldsymbol{X}}^2)^k) \\ &= \int_0^\infty [R_2^k(r) - r^k] dF_{R_{\boldsymbol{X}}^2}(r) \\ &\geq \int_0^\infty kr_0^{k-1}[R_2(r) - r] dF_{R_{\boldsymbol{X}}^2}(r) \\ &= kr_0^{k-1}[E(R_{\boldsymbol{Y}}^2) - E(R_{\boldsymbol{X}}^2)] \\ &= 0, \end{aligned}$$

i.e.,

$$E(R_{\boldsymbol{X}}^{2k}) \leqslant E(R_{\boldsymbol{Y}}^{2k}).$$

For any $0 < \alpha < 2$, the k_α-ordering implies the k_2-ordering by Theorem 2.1. Thus, the result holds. ■

The above result in part (1) leads to the following interpretation of the k_α-ordering. Since the k_α-ordering is affine invariant, without loss of generality we assume that random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ are standardized. Then for any $0 \leqslant r < r_0$,

$$P(\boldsymbol{X}'\boldsymbol{X} \leqslant r^{2/\alpha}) = F_{R^{\alpha}_{\boldsymbol{X}}}(r) \leqslant G_{R^{\alpha}_{\boldsymbol{Y}}}(r) = P(\boldsymbol{Y}'\boldsymbol{Y} \leqslant r^{2/\alpha}),$$

that is, $G_{\boldsymbol{Y}}$ has at least as much probability mass as $F_{\boldsymbol{X}}$ in the region $\{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{x}'\boldsymbol{x} \leqslant r^{2/\alpha}\}$, which means that $G_{\boldsymbol{Y}}$ is at least as peaked as $F_{\boldsymbol{X}}$. For $r \geq r_0$, $F_{R^{\alpha}_{\boldsymbol{X}}}(r) \geq G_{R^{\alpha}_{\boldsymbol{Y}}}(r)$, equivalently, $P(\boldsymbol{X}'\boldsymbol{X} > r^{2/\alpha}) = 1 - F_{R^{\alpha}_{\boldsymbol{X}}}(r) \leqslant 1 - G_{R^{\alpha}_{\boldsymbol{Y}}}(r) = P(\boldsymbol{Y}'\boldsymbol{Y} > r^{2/\alpha})$, i.e., $G_{\boldsymbol{Y}}$ has at least as much probability mass as $F_{\boldsymbol{X}}$ in the region $\{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{x}'\boldsymbol{x} > r^{2/\alpha}\}$, which means that the tails of $G_{\boldsymbol{Y}}$ are at least as heavy as the ones of $F_{\boldsymbol{X}}$. Overall, $F_{\boldsymbol{X}} \leqslant_{k_\alpha} G_{\boldsymbol{Y}}$ implies that $G_{\boldsymbol{Y}}$ has at least as much peakedness and at least as much tailweight as $F_{\boldsymbol{X}}$. From the part (2) of Theorem 2.4, we see that the classical multivariate kurtosis k_d preserves the k_α-orderings with $0 < \alpha \leq 2$.

Remark 2.2 *If we denote the r_0 in Theorem 2.4 by $r_0(\alpha)$, then $r_0(\alpha) = r_0(1)^\alpha$ since $F_{R^{\alpha}_{\boldsymbol{X}}}(r) = F_{R_{\boldsymbol{X}}}(r^{1/\alpha})$. It is reasonable to consider $\{\boldsymbol{x} \in \mathbb{R}^d : [(\boldsymbol{x} - \mu_{\boldsymbol{X}})'\boldsymbol{\Sigma}^{-1}_{\boldsymbol{X}}(\boldsymbol{x} - \mu_{\boldsymbol{X}})]^{\alpha/2} = r_0(\alpha)\} = \{\boldsymbol{x} \in \mathbb{R}^d : (\boldsymbol{x} - \mu_{\boldsymbol{X}})'\boldsymbol{\Sigma}^{-1}_{\boldsymbol{X}}(\boldsymbol{x} - \mu_{\boldsymbol{X}}) = r_0(1)^2\}$ as the "shoulders" of a distribution $F_{\boldsymbol{X}}$. Thus all k_α-orderings use the same shoulders.*

From the above interpretation for the k_α-orderings, it is clear that peakedness, kurtosis, and tailweight should be regarded as distinct features of a distribution, although they are closely interrelated. Peakedness and tailweight orderings for multivariate distributions can also be defined by $F_{R^{\alpha}_{\boldsymbol{X}}}$ and $G_{R^{\alpha}_{\boldsymbol{Y}}}$.

Definition 2.2 *For $\alpha > 0$, we say that $F_{\boldsymbol{X}}$ is p_α less than or equal to $G_{\boldsymbol{Y}}$ in peakedness, denoted by $F_{\boldsymbol{X}} \leqslant_{p_\alpha} G_{\boldsymbol{Y}}$, if there exists a positive r_0 such that $G^{-1}_{R^{\alpha}_{\boldsymbol{Y}}}(F_{R^{\alpha}_{\boldsymbol{X}}}(r))$ is convex for $0 \leq r \leq r_0$, and $F_{\boldsymbol{X}}$ is p_α less than $G_{\boldsymbol{Y}}$ in peakedness, denoted by*

$F_{\boldsymbol{X}} <_{p_\alpha} G_{\boldsymbol{Y}}$, if $G^{-1}_{R^\alpha_{\boldsymbol{Y}}}(F_{R^\alpha_{\boldsymbol{X}}}(r))$ is strictly convex for $0 \le r \le r_0$. We say that $F_{\boldsymbol{X}}$ is p_α equal to $G_{\boldsymbol{Y}}$ in peakedness, denoted by $F_{\boldsymbol{X}} =_{p_\alpha} G_{\boldsymbol{Y}}$, if $F_{\boldsymbol{X}} \le_{p_\alpha} G_{\boldsymbol{Y}}$ and $G_{\boldsymbol{Y}} \le_{p_\alpha} F_{\boldsymbol{X}}$.

Definition 2.3 *For $\alpha > 0$, we say that $F_{\boldsymbol{X}}$ is t_α less than or equal to $G_{\boldsymbol{Y}}$ in tailweight, denoted by $F_{\boldsymbol{X}} \leqslant_{t_\alpha} G_{\boldsymbol{Y}}$, if there exists a positive r_0 such that $G^{-1}_{R^\alpha_{\boldsymbol{Y}}}(F_{R^\alpha_{\boldsymbol{X}}}(r))$ is convex for $r \ge r_0$, and $F_{\boldsymbol{X}}$ is t_α less than $G_{\boldsymbol{Y}}$ in tailweight, denoted by $F_{\boldsymbol{X}} <_{t_\alpha} G_{\boldsymbol{Y}}$, if $G^{-1}_{R^\alpha_{\boldsymbol{Y}}}(F_{R^\alpha_{\boldsymbol{X}}}(r))$ is strictly convex for $r \ge r_0$. We say that $F_{\boldsymbol{X}}$ is t_α equal to $G_{\boldsymbol{Y}}$ in tailweight, denoted by $F_{\boldsymbol{X}} =_{t_\alpha} G_{\boldsymbol{Y}}$, if $F_{\boldsymbol{X}} \le_{t_\alpha} G_{\boldsymbol{Y}}$ and $G_{\boldsymbol{Y}} \le_{t_\alpha} F_{\boldsymbol{X}}$.*

3 Kurtosis orderings of some elliptically symmetric distributions

For univariate distributions, various distribution orderings associated with kurtosis have been established (see, e.g., van Zwet [17] and Rivest [14]). Those distribution orderings are very useful in many statistical analyses. For multivariate distributions, however, we do not have such orderings. Among multivariate distributions, elliptically symmetric distributions are most studied and used. Broad discussion about elliptically symmetric distributions can be found in Fang, Kotz, and Ng [5]. In this section, we will investigate some important families of elliptically symmetric distributions and establish distribution orderings for the families. We use the $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in $E_d(h; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ as the location and covariance measures for the k_α-orderings. Of course, we can also use the mean vector and covariance matrix if they exist. By Theorem

2.2, all results will be the same.

3.1 Ordering of Kotz type distributions

Kotz [7] introduced a family of elliptically symmetric distributions, called Kotz type distributions. The density of a d-dimensional Kotz type distribution is given by

$$f(\boldsymbol{x}) = \frac{s\Gamma(d/2)q^{(2N+d-2)/2s}|\boldsymbol{\Sigma}|^{-1/2}}{\pi^{d/2}\Gamma((2N+d-2)/2s)}[(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})]^{N-1}\exp\{-q[(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})]^s\},$$

$$q>0, s>0, 2N+d>2. \tag{3.1}$$

When $N=1, s=1$, and $q=\frac{1}{2}$, it is a multivariate normal distribution. This family of distributions was found very useful in modeling the data that the multivariate normality assumption is not tenable (see, e.g., Koutras [8]). By Theorem 2.2, any k_α-ordering of Kotz type distributions is not affected by q, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$. We focus on parameters N and s and denote a d-dimensional Kotz type distribution by $MK_d(N,s)$. Without loss of generality we set $q=1$. For Kotz type distributions, we have the following ordering result.

Theorem 3.1 *For any $s>0$, if $N_1<N_2$ then $MK_d(N_2,s) <_{k_{2s}} MK_d(N_1,s)$.*

Proof. Suppose that $\boldsymbol{X}$ has a Kotz type distribution with the density given in (3.1). By Lemma 1.1, the density of $R_{\boldsymbol{X}}^{2s}$ is

$$f_{R_{\boldsymbol{X}}^{2s}}(r) = \frac{1}{\Gamma((N+d/2-1)/s)} r^{(N+d/2-1)/s-1}\exp(-r),\ r\geq 0, \tag{3.2}$$

which is a gamma distribution Gamma$((N+d/2-1)/s,1)$.

Van Zwet [17] showed that for the gamma family $F_\tau = Gamma(\tau, 1), F_{\tau_2}^{-1}(F_{\tau_1}(x))$ is concave for $0 < x < \infty$ if $\tau_1 < \tau_2$. In addition, from his proof, we see that $F_{\tau_2}^{-1}(F_{\tau_1}(x))$ is strictly concave. The result follows immediately. ■

Let $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be a d-dimensional normal distribution. As a special case, we see that $MK_d(N, 1) <_{k_2} N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ if $N > 1$ and $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) <_{k_2} MK_d(N, 1)$ if $N < 1$.

3.2 Ordering of Pearson Type VII distributions

Another important family of elliptically symmetric distributions is the family of Pearson Type VII distributions. A distribution F in $\mathbb{R}^d$ is called a Pearson Type VII distribution if its density is given by

$$f(\boldsymbol{x}) = \tfrac{(\pi v)^{-d/2}\Gamma(K)}{\Gamma(K-d/2)} \left|\boldsymbol{\Sigma}\right|^{-1/2} [1 + (\boldsymbol{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})/v]^{-K}, \ \ K > d/2, \ \ v > 0.$$

When $K = (d + v)/2$ and v is an integer, it is a multivariate t distribution, denoted by $Mt_d(v)$. Further when $v = 1$, it is a multivariate Cauchy distribution, denoted by MC_d. Now we give the ordering result for Pearson Type VII distributions. By Theorem 2.2, v , $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$ do not affect any k_α-ordering of Pearson Type VII distributions. We denote a Pearson Type VII distribution by $MPVII_d(K)$ to emphasize the parameter K. Without loss of generality, we set $v = 1$.

Theorem 3.2 *For Pearson Type VII distributions, if* $K_1 < K_2$ *then* $MPVII_d(K_2) <_{k_2} MPVII_d(K_1)$.

Proof. Suppose that $\boldsymbol{X} \sim MPVII_d(K_1)$ and $\boldsymbol{Y} \sim MPVII_d(K_2)$. By Lemma 1.1, the density functions of $R^2_{\boldsymbol{X}} = (\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}})'\boldsymbol{\Sigma}^{-1}_{\boldsymbol{X}}(\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}})$ and $R^2_{\boldsymbol{Y}} = (\boldsymbol{Y}-\boldsymbol{\mu}_{\boldsymbol{Y}})'\boldsymbol{\Sigma}^{-1}_{\boldsymbol{Y}}(\boldsymbol{Y}-\boldsymbol{\mu}_{\boldsymbol{Y}})$ are respectively

$$f_{R^2_{\boldsymbol{X}},K_1}(r) = \tfrac{1}{B(d/2,K_1-d/2)} r^{d/2-1}(1+r)^{-K_1}, \;\; r \geqslant 0,$$
$$f_{R^2_{\boldsymbol{Y}},K_2}(r) = \tfrac{1}{B(d/2,K_2-d/2)} r^{d/2-1}(1+r)^{-K_2}, \;\; r \geqslant 0.$$

Denote by $F_{R^2_{\boldsymbol{X}},K_1}$ and $F_{R^2_{\boldsymbol{Y}},K_2}$ the cdf's of $R^2_{\boldsymbol{X}}$ and $R^2_{\boldsymbol{Y}}$. We need to prove that $F^{-1}_{R^2_{\boldsymbol{Y}},K_2}(F_{R^2_{\boldsymbol{X}},K_1}(r))$ is strictly concave for $r \geq 0$. We employ the following indirect approach: $F^{-1}_{R^2_{\boldsymbol{Y}},K_2}(F_{R^2_{\boldsymbol{X}},K_1}(r))$ is strictly concave if and only if for any straight line $y = b(r+a)$, $F^{-1}_{R^2_{\boldsymbol{Y}},K_2}(F_{R^2_{\boldsymbol{X}},K_1}(r)) - b(r+a)$ can have at most two distinct zeros and is positive between these zeros. Since $F^{-1}_{R^2_{\boldsymbol{Y}},K_2}(F_{R^2_{\boldsymbol{X}},K_1}(r))$ is strictly increasing, $F^{-1}_{R^2_{\boldsymbol{Y}},K_2}(F_{R^2_{\boldsymbol{X}},K_1}(r)) - b(r+a)$ can have at most one zero for $b \leq 0$. So we need only to consider the case $b > 0$. Since $F^{-1}_{R^2_{\boldsymbol{Y}},K_2}(F_{R^2_{\boldsymbol{X}},K_1}(r)) - b(r+a)$ has the same sign as

$$\phi(r) = F_{R^2_{\boldsymbol{X}},K_1}(r) - F_{R^2_{\boldsymbol{Y}},K_2}(b(r+a)), \text{ and}$$
$$\phi'(r) = f_{R^2_{\boldsymbol{X}},K_1}(r) - b f_{R^2_{\boldsymbol{Y}},K_2}(b(r+a))$$

has the same sign as

$$\psi(r) = \log(f_{R^2_{\boldsymbol{X}},K_1}(r)) - \log(f_{R^2_{\boldsymbol{Y}},K_2}(b(r+a))) - \log(b),$$

the sign pattern of $F^{-1}_{R^2_{\boldsymbol{Y}},K_2}(F_{R^2_{\boldsymbol{X}},K_1}(r)) - b(r+a)$ can be found by the sign pattern of $\psi'(r)$. Here

$$\begin{aligned} \psi'(r) &= \frac{d/2-1}{r} - \frac{K_1}{1+r} - \frac{d/2-1}{r+a} + \frac{K_2 b}{1+b(r+a)} \\ &= \frac{(d/2-1)a}{r(r+a)} + \frac{(K_2-K_1)br + (K_2 b - K_1 - K_1 ab)}{(1+r)[1+b(r+a)]} \end{aligned} \tag{3.3}$$

A study of the sign of $\psi'(r)$ for $r \geq 0$, and the signs of $\psi(r)$ (or $\phi'(r)$) and $\phi(r)$ for $r = 0$ and $r \to \infty$ shows that $\phi(r)$, and hence $F^{-1}_{R^2_{\boldsymbol{Y}},K_2}(F_{R^2_{\boldsymbol{X}},K_1}(r)) - b(r+a)$, can have at most two distinct zeros for $r \geq 0$ and is positive between these zeros. This completes the proof. ■

Corollary 3.1 *For multivariate t distributions, if $v_1 < v_2$ then $Mt_d(v_2) <_{k_2} Mt_d(v_1)$.*

Proof. The result follows from Theorems 2.2 and 3.2 immediately. ■

It is easy to see that $\lim_{v \to \infty} Mt_d(v) = N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. From Corollary 3.1, we have that for any $v > 1$, $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) <_{k_2} Mt_d(v) <_{k_2} MC_d$.

3.3 Ordering of Pearson Type II distributions

A d-dimensional distribution is called a Pearson Type II distribution, denoted by $MPII_d(m)$, if its density function is of the form

$$f(\boldsymbol{x}) = \frac{\Gamma(d/2+m+1)}{\Gamma(m+1)\pi^{d/2}} |\boldsymbol{\Sigma}|^{-1/2} [1 - (\boldsymbol{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})]^m, \;\; 0 \leq (\boldsymbol{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu}) < 1$$
$$\text{and } m > -1.$$

Detailed discussions of this family were given by Johnson [6]. As the parameter m varies, the Pearson Type II distribution takes on many shapes. The density is unimodal when $m > 0$, uniform when $m = 0$, and bowl-shaped when $-1 < m < 0$. About the family of Pearson Type II distributions, we have the following ordering result.

Theorem 3.3 *For Pearson Type II distributions, if* $m_1 < m_2$ *then* $MPII_d(m_1) <_{k_2} MPII_d(m_2)$.

Proof. Suppose that $\boldsymbol{X} \sim MPII_d(m_1)$ and $\boldsymbol{Y} \sim MPII_d(m_2)$. By Lemma 1.1, the density functions of $R^2_{\boldsymbol{X}} = (\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}})'\boldsymbol{\Sigma}^{-1}_{\boldsymbol{X}}(\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}})$ and $R^2_{\boldsymbol{Y}} = (\boldsymbol{Y}-\boldsymbol{\mu}_{\boldsymbol{Y}})'\boldsymbol{\Sigma}^{-1}_{\boldsymbol{Y}}(\boldsymbol{Y}-\boldsymbol{\mu}_{\boldsymbol{Y}})$ are respectively,

$$
\begin{aligned}
f_{R^2_{\boldsymbol{X}},m_1}(r) &= \frac{1}{B(\frac{d}{2}, m_1+1)} r^{d/2-1}(1-r)^{m_1}, \ 0 \leqslant r \leqslant 1, \\
f_{R^2_{\boldsymbol{Y}},m_2}(r) &= \frac{1}{B(\frac{d}{2}, m_2+1)} r^{d/2-1}(1-r)^{m_2}, \ 0 \leqslant r \leqslant 1.
\end{aligned}
$$

Denote by $F_{R^2_{\boldsymbol{X}},m_1}$ and $F_{R^2_{\boldsymbol{Y}},m_2}$ the cdf's of $R^2_{\boldsymbol{X}}$ and $R^2_{\boldsymbol{Y}}$. We need to show that $F^{-1}_{R^2_{\boldsymbol{Y}},m_2}(F_{R^2_{\boldsymbol{X}},m_1}(r))$ is strictly convex for $0 \leq r \leq 1$. As in Section 3.2, we consider the function

$$
\phi(r) = F_{R^2_{\boldsymbol{X}},m_1}(r) - F_{R^2_{\boldsymbol{Y}},m_2}(b(r+a)),
$$

for $b > 0$ and various values of a, which has the same sign as

$$
F^{-1}_{R^2_{\boldsymbol{Y}},m_2}(F_{R^2_{\boldsymbol{X}},m_1}(r)) - b(r+a) \text{ for } r \in [0,1] \cap [-a, -a+\tfrac{1}{b}].
$$

Since $\phi'(r) = f_{R^2_{\boldsymbol{X}},m_1}(r) - b f_{R^2_{\boldsymbol{Y}},m_2}(b(r+a))$ has the same sign as

$$
\psi(r) = \log(f_{R^2_{\boldsymbol{X}},m_1}(r)) - \log(f_{R^2_{\boldsymbol{Y}},m_2}(b(r+a)) - \log(b),
$$

the sign pattern of $\phi'(r)$ can be found by the sign pattern of $\psi'(r)$. Here

$$
\begin{aligned}
\psi'(r) &= \frac{d/2-1}{r} - \frac{m_1}{1-r} - \frac{d/2-1}{r+a} + \frac{m_2 b}{1-b(r+a)} \\
&= \frac{(d/2-1)a}{r(r+a)} + \frac{(m_1-m_2)br + (m_2 b - m_1 + m_1 ab)}{(1-r)[1-b(r+a)]}.
\end{aligned}
\tag{3.4}
$$

A study of the sign of $\psi'(r)$ for $r \in [0,1] \cap [-a, -a+\frac{1}{b}]$, and the signs of $\psi(r)$ (or $\phi'(r)$) and $\phi(r)$ at the end points of [0,1] $\cap$ $[-a, -a+\frac{1}{b}]$ shows that $\phi(r)$, and hence $F^{-1}_{R^2_{\boldsymbol{Y}},m_2}(F_{R^2_{\boldsymbol{X}},m_1}(r)) - b(r+a)$, can have at most two distinct zeros for $r \in [0,1] \cap [-a, -a + \frac{1}{b}]$ and is negative between these zeros.

For $b \leqslant 0$, $F^{-1}_{R^2_{\boldsymbol{Y}},m_2}(F_{R^2_{\boldsymbol{X}},m_1}(r)) - b(r+a)$ can have at most one zero since $F^{-1}_{R^2_{\boldsymbol{Y}},m_2}(F_{R^2_{\boldsymbol{X}},m_1}(r))$ is strictly increasing. Thus, $F^{-1}_{R^2_{\boldsymbol{Y}},m_2}(F_{R^2_{\boldsymbol{X}},m_1}(r))$ is strictly convex for $0 \leq r \leq 1$. ■

From Theorem 3.3, we see that for Pearson Type II distributions, bowl-shaped $<_{k_2}$ uniform $<_{k_2}$ unimodal. Generally, we have the following result.

Theorem 3.4 *For any elliptically symmetric distributions in* $\mathbb{R}^d$*,*

$$\textit{Bowl-shaped} \leqslant_{k_d} \textit{Uniform} \leq_{k_d} \textit{Unimodal}.$$

Proof. Suppose that random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ in $\mathbb{R}^d$ have elliptically symmetric distributions $F_{\boldsymbol{X}}$ and $G_{\boldsymbol{Y}}$ with density functions

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = C_1 |\boldsymbol{\Sigma}_1|^{-1/2} h_1((\boldsymbol{x} - \boldsymbol{\mu}_1)' \boldsymbol{\Sigma}_1^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_1)) \text{ and}$$
$$g_{\boldsymbol{Y}}(\boldsymbol{y}) = C_2 |\boldsymbol{\Sigma}_2|^{-1/2} h_2((\boldsymbol{y} - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}_2^{-1} (\boldsymbol{y} - \boldsymbol{\mu}_2)).$$

Then by Lemma 1.1 the density functions of $R^d_{\boldsymbol{X}} = [(\boldsymbol{X} - \boldsymbol{\mu}_1)' \boldsymbol{\Sigma}_1^{-1} (\boldsymbol{X} - \boldsymbol{\mu}_1)]^{d/2}$ and $R^d_{\boldsymbol{Y}} = [(\boldsymbol{Y} - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}_2^{-1} (\boldsymbol{Y} - \boldsymbol{\mu}_2)]^{d/2}$ are respectively

$$f_{R^d_{\boldsymbol{X}}}(r) = \tfrac{2C_1 \pi^{d/2}}{d\Gamma(d/2)} h_1(r^{2/d}),$$
$$g_{R^d_{\boldsymbol{Y}}}(r) = \tfrac{2C_2 \pi^{d/2}}{d\Gamma(d/2)} h_2(r^{2/d}).$$

Denote by $F_{R^d_{\boldsymbol{X}}}$ and $G_{R^d_{\boldsymbol{Y}}}$ the distribution functions of $R^d_{\boldsymbol{X}}$ and $R^d_{\boldsymbol{Y}}$. If $G_{\boldsymbol{Y}}$ is uniform, $h_2(\cdot) \equiv c$. Then,

$$G_{R^d_{\boldsymbol{Y}}}(r) = \textstyle\int_0^r g_{R^d_{\boldsymbol{Y}}}(t) dt = \tfrac{2C_2 \pi^{d/2} c}{d\Gamma(d/2)} r, \text{ and}$$
$$R_d(r) = G^{-1}_{R^d_{\boldsymbol{Y}}}(F_{R^d_{\boldsymbol{X}}}(r)) = \tfrac{d\Gamma(d/2)}{2C_2 \pi^{d/2} c} F_{R^d_{\boldsymbol{X}}}(r).$$

Thus,

$$R_d'(r) = \frac{C_1}{C_2 c} h_1(r^{2/d}).$$

When $f_{\boldsymbol{X}}(\boldsymbol{x})$ is bowl-shaped, $h_1(\cdot)$ is increasing, which implies that $R_d(r)$ is convex. When $f_{\boldsymbol{X}}(\boldsymbol{x})$ is unimodal, $h_1(\cdot)$ is decreasing, which implies that $R_d(r)$ is concave. This completes the proof. ■

4 Concluding Remarks

Kurtosis of multivariate skewed distributions. For any $\alpha > 0$, the k_α-ordering is defined for all distributions in $\mathbb{R}^d$. It is important to interpret kurtosis of skewed distributions when they are involved in a kurtosis ordering. For the univariate case, Balanda and MacGillivray [2] interpreted kurtosis of an asymmetric distribution by its symmetrized version. Kurtosis of a multivariate skewed distribution with respect to the k_α-ordering can be interpreted as follows. For any skewed distribution $F_{\boldsymbol{X}}$ in $\mathbb{R}^d$, the distributions that are k_α equal to $F_{\boldsymbol{X}}$ consist of an equivalence class in kurtosis $\mathcal{C}_{F_{\boldsymbol{X}}} = \{G_{\boldsymbol{Y}} \text{ in } \mathbb{R}^d : G_{\boldsymbol{Y}} =_{k_\alpha} F_{\boldsymbol{X}}\}$. When we study kurtosis by the k_α-ordering, any distribution in the class can serve as a representative. The important fact is that the class $\mathcal{C}_{F_{\boldsymbol{X}}}$ contains a spherically symmetric distribution with the origin as center, which is unique by Theorem 2.3. In fact, suppose that $f_{R^\alpha_{\boldsymbol{X}}}(r)$ is the density of $R^\alpha_{\boldsymbol{X}}$. The spherically symmetric distribution with the density $f(\boldsymbol{x}) = \frac{\alpha\Gamma(d/2)}{2\pi^{d/2}} \frac{f_{R^\alpha_{\boldsymbol{X}}}((\boldsymbol{x}'\boldsymbol{x})^{\alpha/2})}{(\boldsymbol{x}'\boldsymbol{x})^{(d-\alpha)/2}}$ is in the class. Then kurtosis of $F_{\boldsymbol{X}}$ can be interpreted by the kurtosis of this spherically symmetric distribution. The skewness of $F_{\boldsymbol{X}}$ can be studied by a multivariate skewness ordering with respect to spherical symmetry. By this approach, it can be seen that skewness and kurtosis are distinct components of shape. Kurtosis characterizes the vertical aspect of shape and skewness the horizontal aspect. See MacGillivray and Balanda [11] for the relationship between skewness and kurtosis in the univariate case.

Weaker multivariate kurtosis orderings. For any $\alpha > 0$, weaker multivariate kurtosis orderings can be defined by weakening the convexity condition of $\leq_{k_\alpha}$. For example,

(1) $F_{\boldsymbol{X}} \leq_{k_\alpha^{star}} G_{\boldsymbol{Y}}$ *iff* $G^{-1}_{R^\alpha_{\boldsymbol{Y}}}(F_{R^\alpha_{\boldsymbol{X}}}(r))$ is star-shaped.

(2) $F_{\boldsymbol{X}} \leq_{k_\alpha^{cross}} G_{\boldsymbol{Y}}$ *iff* there is $r_0 > 0$ such that $F_{R^\alpha_{\boldsymbol{X}}}(r) \leq G_{R^\alpha_{\boldsymbol{Y}}}(r)$ for $0 \leq r < r_0$ and $F_{R^\alpha_{\boldsymbol{X}}}(r) \geq G_{R^\alpha_{\boldsymbol{Y}}}(r)$ for $r \geq r_0$.

It can be shown that $F_{\boldsymbol{X}} \leq_{k_\alpha} G_{\boldsymbol{Y}} \Longrightarrow F_{\boldsymbol{X}} \leq_{k_\alpha^{star}} G_{\boldsymbol{Y}} \Longrightarrow F_{\boldsymbol{X}} \leq_{k_\alpha^{cross}} G_{\boldsymbol{Y}}$. All the properties of the k_α-ordering established in Section 2.3 also hold for the k_α^{star}-ordering and the k_α^{cross}-ordering. Since the k_α-ordering is stronger than the k_α^{star}-ordering and k_α^{cross}-ordering, the ordering results established in Section 3 hold for the corresponding k_α^{star}-ordering and the k_α^{cross}-ordering as well.

Conclusion. A multivariate kurtosis ordering should characterize in an affine invariant sense the movement of probability mass from the "shoulders" of a distribution to either the center or the tails or both. Structurally any distribution ordering $\leq_k$ is a multivariate kurtosis ordering if (1) it is affine invariant and (2) for any distributions F and G in $\mathbb{R}^d$, $F \leq_k G$ implies that G has at least as much peakedness and at least as much tailweight as F. Compared with kurtosis, a global feature, peakedness is a local feature on some central region and tailweight is a local feature on some tail region.

Acknowledgments

Very valuable suggestions and insightful comments by an Associate Editor, two anonymous referees, and Roger Berger are greatly appreciated and have been utilized to improve the paper considerably. Special thanks are due to Robert Serfling for stimulating discussions, constructive remarks, and constant encouragement. Also, support by IGP of Northern Arizona University is gratefully acknowledged. The author was supported by IGP of Northern Arizona University.

References

[1] J. Averous, M. Meste, Multivariate kurtosis in L_1-sense, Statist. Probab. lett. 19 (1994) 281-284.

[2] K.P. Balanda, H.L. MacGillivray, Kurtosis and spread, Canad. J. Statist. 18 (1990) 17-30.

[3] P.J. Bickel, E.L. Lehmann, Descriptive statistics for nonparametric models. I. Introduction. II. Location, Ann. Statist. 3 (1975) 1038-1069.

[4] P.J. Bickel, E.L. Lehmann, Descriptive statistics for nonparametric models. III. Dispersion, Ann. Statist. 4 (1976) 1139-1158.

[5] K.T. Fang, S. Kotz, K.W. Ng, Symmetric Multivariate and Related Distributions, Chapman and Hall, London and New York, 1990.

[6] M.E. Johnson, Multivariate Statistical Simulation, Wiley, New York, 1987.

[7] S. Kotz, Multivariate distributions at a cross-road, In: G.P. Patil, S. Kotz, J.K. Ord, D. Reidel (Eds.), in: Statistical Distributions in Scientific Work, vol. 1, Publ. Co., 1975.

[8] M. Koutras, On the generalized noncentral chi-squared distribution induced by an elliptical gamma law. Biometrika 73 (1986) 528-532.

[9] R.Y. Liu, J.M. Parelius, K. Singh, Multivariate analysis by data depth: Descriptive statistics, graphics and inference (with discussion), Ann. Statist. 27 (1999) 783-858.

[10] K.V. Mardia, Measures of multivariate skewness and kurtosis with applications, Biometrika 57 (1970) 519-530.

[11] H.L. MacGillivray, K.P. Balanda, The relationships between skewness and kurtosis, Austral. J. Statist. 30 (1988) 319-337.

[12] H. Oja, On location, scale, skewness and kurtosis of univariate distributions, Scand. J. Stat. 8 (1981) 154-168.

[13] H. Oja, Descriptive statistics for multivariate distributions, Statist. Probab. Lett. 1 (1983) 327-332.

[14] L.P. Rivest, Products of random variables and star-shaped ordering, Canad. J. Statist. 10 (1982) 219-223.

[15] R. Serfling, Nonparametric multivariate descriptive measures based on spatial quantiles, J. Stat. Plan. Inference 123 (2004) 259-278.

[16] M.S. Srivastava, A measure of skewness and kurtosis and a graphical method for assessing multivariate normality, Statist. Probab. Lett. 2 (1984) 263-267.

[17] W.R. van Zwet, Convex Transformations of Random Variables, Mathematish Centrum, Amsterdam, 1964.

[18] J. Wang, R. Serfling, Nonparametric multivariate kurtosis and tailweight measures, J. Nonparamet. Stat. 17 (2005) 441-456.