Mathematical
Inequalities
& Applications
Volume 22, Number 3 (2019), 781–790 doi:10.7153/mia-2019-22-53

EXCESS VERSIONS OF THE MINKOWSKI AND HÖLDER INEQUALITIES

IOSIF PINELIS

(*Communicated by S. Varošanec*)

Abstract. Certain excess versions of the Minkowski and Hölder inequalities are given.

1. Introduction and summary

Let p and q be positive real numbers such that $\frac{1}{p}+\frac{1}{q}=1$; then, of course, $p>1$ and $q>1$. Let X and Y denote nonnegative random variables (r.v.'s), defined on the same probability space. Then one has the Minkowski inequality

$$\|X+Y\|_p \leqslant \|X\|_p + \|Y\|_p$$

and the Hölder inequality

$$\mathsf{E}XY \leqslant \|X\|_p\|Y\|_q,$$

where, as usual, $\|X\|_p := \mathsf{E}^{1/p}|X|^p$; see, e.g., [7]. From now on, to avoid unpleasant trivialities, let us assume that $\|X\|_p+\|Y\|_p+\|Y\|_q<\infty$.

A special case of Hölder's inequality is Lyapunov's inequality, which states that $\mathsf{E}X^\alpha$ is log-convex in real α, with the conventions $0^0:=1$, $0^\alpha:=\infty$ for $\alpha<0$, and $0\cdot\infty:=0$, so that $\mathsf{E}X^0=1$, and $\mathsf{E}X^\alpha=\infty$ if $\alpha<0$ and $\mathsf{P}(X=0)>0$. In particular, we have $\|X\|_1\leqslant\|X\|_p$.

So, we may define the (always nonnegative) p-excess of X by the formula

$$\mathscr{E}_p(X) := \big(\|X\|_p^p - \|X\|_1^p\big)^{1/p}.$$

One may note that $\mathscr{E}_2(X)$ is the standard deviation of the r.v. X. Introduce also the covariance-like expression

$$\mathscr{C}_p(X,Y) := \mathsf{E}X^{p-1}Y - \mathsf{E}^{p-1}X\,\mathsf{E}Y,$$

which is the true covariance, $\mathsf{Cov}(X,Y)$, of the r.v.'s X and Y in the case $p=2$.

As will be shown in this note, the following Minkowski-like and Hölder-like inequalities for the p-excess hold: if $p\leqslant 2$ (so that $1<p\leqslant 2$), then

$$\mathscr{E}_p(X+Y) \leqslant \mathscr{E}_p(X)+\mathscr{E}_p(Y) \tag{1}$$

Mathematics subject classification (2010): 26D15, 60E15.
Keywords and phrases: Minkowski's inequality, Hölder's inequality, p-excess.

and

$$\mathscr{C}_p(X,Y) \leqslant \mathscr{E}_p(X)^{p-1}\mathscr{E}_p(Y). \tag{2}$$

In the case $p = 2$ inequality (2) becomes the covariance inequality, that is, the Cauchy–Schwarz inequality for the centered r.v.'s $X - \mathsf{E}X$ and $Y - \mathsf{E}Y$.

More generally, for $\theta \in [0,1]$ define the (p,θ)-excess of X by the formula

$$\mathscr{E}_{p,\theta}(X) := \big(\|X\|_p^p - \theta^p\|X\|_1^p\big)^{1/p},$$

which interpolates between $\|X\|_p = \mathscr{E}_{p,0}(X)$ and $\mathscr{E}_p(X) = \mathscr{E}_{p,1}(X)$, and then also

$$\mathscr{C}_{p,\theta}(X,Y) := \mathsf{E}X^{p-1}Y - \theta^p\,\mathsf{E}^{p-1}X\,\mathsf{E}Y,$$

which interpolates between $\mathsf{Cov}(X^{p-1},Y) = \mathscr{C}_{p,0}(X,Y)$ and $\mathscr{C}_p(X) = \mathscr{C}_{p,1}(X,Y)$.

Inequalities (1) and (2), along with the Minkowski and Hölder inequalities, can be extended as follows:

THEOREM 1. *Suppose that $p \leqslant 2$ (so that $1 < p \leqslant 2$). Then for all $\theta \in [0,1]$*

$$\mathscr{E}_{p,\theta}(X+Y) \leqslant \mathscr{E}_{p,\theta}(X) + \mathscr{E}_{p,\theta}(Y) \tag{3}$$

and

$$\mathscr{C}_{p,\theta}(X,Y) \leqslant \mathscr{E}_{p,\theta}(X)^{p-1}\mathscr{E}_{p,\theta}(Y). \tag{4}$$

For any real $p > 2$ and any $\theta \in (0,1]$, inequalities (3) and (4) do not hold in general.

Obviously, the Minkowski and Hölder inequalities are the special cases of inequalities (3) and (4), respectively, corresponding to $\theta = 0$, and (1) and (2) are the special cases of (3) and (4) corresponding to $\theta = 1$. Moreover, considerations in Round 1 of the proof of (4), to be given in Section 2, show that inequality (4) is, in a sense, an improvement of Hölder's inequality (for $p \in (1,2)$). Similarly, the derivation of (3) from (4) in the paragraph containing formulas (31) and (32) shows that inequality (3) is an improvement of Minkowski's inequality (again for $p \in (1,2)$). Inequality (1) was conjectured in [3].

2. Proof of Theorem 1

We shall see at the end of this section that inequalities (3) and (4) are easy to obtain from each other, so that it is enough to prove one of them.

Proof of inequality (4) in Theorem 1. This proof is much more difficult than that of Hölder's inequality. It will be done by a number of rounds of reduction of the difficulty of the problem.

Round 1: Reduction to the case $\theta = 1$
Consider the differences

$$\Delta_{p,\theta}(X,Y) := \mathscr{C}_{p,\theta}(X,Y) - \mathscr{E}_{p,\theta}(X)^{p-1}\mathscr{E}_{p,\theta}(Y) \tag{5}$$

and

$$\Delta_p(X,Y) := \Delta_{p,1}(X,Y) := \mathscr{C}_p(X,Y) - \mathscr{E}_p(X)^{p-1}\mathscr{E}_p(Y) \tag{6}$$

between the left and right sides of inequalities (4) and (2), respectively. For nonnegative real numbers A,B,C, consider also

$$\begin{aligned}\Delta_{p;A,B,C}(X,Y) =& A + \mathsf{E}X^{p-1}Y - \mathsf{E}^{p-1}X\,\mathsf{E}Y \\ & - \big(B + \mathsf{E}X^p - \mathsf{E}^p X\big)^{1/q}\big(C + \mathsf{E}Y^p - \mathsf{E}^p Y\big)^{1/p}.\end{aligned} \tag{7}$$

The following lemma will also be used in Round 8 of this proof.

LEMMA 1. *Suppose that the nonnegative real numbers A,B,C are such that $A \leqslant B^{1/q}C^{1/p}$. Then $\Delta_{p;A,B,C}(X,Y) \leqslant \Delta_p(X,Y)$.*

Proof. Since $\Delta_{p;A,B,C}(X,Y)$ is nondecreasing in A, without loss of generality (wlog) $A = B^{1/q}C^{1/p}$. If $B=0$ or $C=0$, then $A=0$, and so, the inequality $\Delta_{p;A,B,C}(X,Y) \leqslant \Delta_p(X,Y)$ is trivial. Hence, wlog $B>0$ and $C>0$, and then we can write $C=\gamma^p B$ and $A=\gamma B$ for some real $\gamma>0$. Let now

$$d(B) := \Delta_{p;\gamma B,B,\gamma^p B}(X,Y). \tag{8}$$

Introduce also

$$c := \big(\gamma^p B + \mathsf{E}Y^p - \mathsf{E}^p Y\big)^{1/p} / \big(B + \mathsf{E}X^p - \mathsf{E}^p X\big)^{1/p}$$

and then $a := \gamma c^{-1/q}$ and $b := c^{1/q}$. Then

$$d'(B) = \gamma - \frac{1}{q}c - \frac{1}{p}\gamma^p c^{-p/q} = ab - \Big(\frac{a^p}{p} + \frac{b^q}{q}\Big) \leqslant 0$$

for all $B>0$, by Young's inequality. So, $\Delta_{p;A,B,C}(X,Y) = \Delta_{p;\gamma B,B,\gamma^p B}(X,Y) = d(B) \leqslant d(0) = \Delta_{p:0,0,0}(X,Y) = \Delta_p(X,Y)$. Lemma 1 is thus proved. □

Now take any $\theta \in [0,1]$ and note that $\Delta_{p,\theta}(X,Y) = \Delta_{p;A,B,C}(\theta X, \theta Y)$ with $A := (1-\theta^p)\mathsf{E}X^{p-1}Y$, $B := (1-\theta^p)\mathsf{E}X^p$, and $C := (1-\theta^p)\mathsf{E}Y^p$, so that, by Hölder's inequality, the condition $A \leqslant B^{1/q}C^{1/p}$ of Lemma 1 holds, which yields $\Delta_{p,\theta}(X,Y) \leqslant \Delta_p(\theta X, \theta Y)$. Thus, to prove inequality (4), it is enough to prove its special case, inequality (2).

Round 2: Removing the case $p=2$

This round is very easy. As was noted, the case $p=2$ of (2) is the Cauchy–Schwarz inequality. So, it is enough to prove (2) for $p \in (1,2)$, which will be henceforth assumed.

Round 3: "Finitization" of the probability space

Wlog the r.v.'s X and Y take only finitely many values (one may approximate X and Y from below by nonnegative simple r.v.'s and then use the monotone convergence theorem). Therefore, wlog X and Y are defined on a finite probability space. For instance, we may assume that the probability space is (I,Σ,μ), where I is the finite set

$\{(x,y)\colon\ \mathsf{P}(X=x,Y=y)>0\}$, Σ is the σ-algebra of all subsets of I, the probability measure μ is defined by the condition $\mu(\{i\})=w_i:=\mathsf{P}(X=x,Y=y)$ for all $i=(x,y)\in I$, and the r.v.'s X and Y are defined by the conditions $X(i)=x$ and $Y(i)=y$ for all $i=(x,y)\in I$. So, the r.v.'s X and Y maybe identified with finite-dimension vectors $(x_i)_{i\in I}$ and $(y_i)_{i\in I}$, respectively.

Round 4: Reduction to an extremal problem

Introducing also the vector $W:=(w_i)_{i\in I}$, we can rewrite inequality (2) as

$$\sup\big\{\Delta_p(X,Y,W)\colon\ (X,Y,W)\in\mathscr{T}_{I;m_{1,1},m_{1,p},m_{2,1},m_{2,p}}\big\}\overset{(?)}{\leqslant}0, \tag{9}$$

where $m_{1,1},m_{1,p},m_{2,1},m_{2,p}$ are any (strictly) positive real numbers,

$$\Delta_p(X,Y,W):=(X^{p-1}Y)\cdot W-(X\cdot W)^{p-1}Y\cdot W \\ -\big(X^p\cdot W-(X\cdot W)^p\big)^{1/q}\big(Y^p\cdot W-(Y\cdot W)^p\big)^{1/p},$$

the symbol $\cdot$ denotes the dot product in $\mathbb{R}^I$, and $\mathscr{T}_{I;m_{1,1},m_{1,p},m_{2,1},m_{2,p}}$ is the set of all triples (X,Y,W) of vectors $X=(x_i)_{i\in I}$, $Y=(y_i)_{i\in I}$, and $W=(w_i)_{i\in I}$ with nonnegative coordinates such that

$$\mathbf{1}\cdot W=\sum_{i\in I}w_i=1,\quad X^p\cdot W=\sum_{i\in I}x_i^p w_i=m_{1,p},\quad Y^p\cdot W=\sum_{i\in I}y_i^p w_i=m_{2,p}, \\ X\cdot W=\sum_{i\in I}x_i w_i=m_{1,1},\quad Y\cdot W=\sum_{i\in I}y_i w_i=m_{2,1};$$

here and in what follows, $\mathbf{1}:=(1)_{i\in I}$, the vector with all coordinates equal 1. One may note that, in view of the standard convention $\sup\emptyset=-\infty$, inequality (9) is trivial whenever $m_{1,1},m_{1,p},m_{2,1},m_{2,p}$ are such that $\mathscr{T}_{I;m_{1,1},m_{1,p},m_{2,1},m_{2,p}}=\emptyset$. A reason for the numbers $m_{1,1},m_{1,p},m_{2,1},m_{2,p}$ to be assumed strictly positive is that, if at least one of them is 0, then for any $(X,Y,W)\in\mathscr{T}_{I;m_{1,1},m_{1,p},m_{2,1},m_{2,p}}$ at least one of the r.v.'s X,Y is almost surely 0, which makes inequality (9) trivial.

Round 5: Compactification, by a change of variables

To solve an extremal problem such as the one stated in Round 4, it is natural to use the method of Lagrange multipliers. To be able to do that, we need to ensure a priori that the supremum in (9) is attained. However, this does not seem easy to do, since the set $\mathscr{T}_{I;m_{1,1},m_{1,p},m_{2,1},m_{2,p}}$ is not bounded and hence not compact in general; indeed, for any real $\beta>0$ and any $i\in I$ such that $w_i=0$, one may take however large $x_i\geqslant 0$ so that the condition $x_i^\beta w_i=0$ hold.

An appropriate way to compactify the set $\mathscr{T}_{I;m_{1,1},m_{1,p},m_{2,1},m_{2,p}}$ is to use the following new variables: for $i\in I$, let

$$u_i:=x_i^p w_i\quad\text{and}\quad v_i:=y_i^p w_i, \tag{10}$$

so that

$$x_i w_i=u_i^{1/p}w_i^{1/q},\quad y_i w_i=v_i^{1/p}w_i^{1/q},\quad x_i^{p-1}y_i w_i=u_i^{1/q}v_i^{1/p}.$$

Then (9) will follow from

$$\sup\big\{\tilde\Delta_p(U,V,W)\colon (U,V,W)\in\tilde{\mathscr T}_{I;m_{1,1},m_{1,p},m_{2,1},m_{2,p}}\big\}\overset{(?)}{\leqslant}0, \tag{11}$$

where $\tilde{\mathscr T}_{I;m_{1,1},m_{1,p},m_{2,1},m_{2,p}}$ is the set of all triples (U,V,W) of vectors $U=(u_i)_{i\in I}$, $V=(v_i)_{i\in I}$, and $W=(w_i)_{i\in I}$ with nonnegative coordinates such that

$$\mathbf 1\cdot W=1,\quad U\cdot\mathbf 1=m_{1,p},\quad V\cdot\mathbf 1=m_{2,p}, \tag{12}$$

$$U^{1/p}\cdot W^{1/q}=m_{1,1},\quad V^{1/p}\cdot W^{1/q}=m_{2,1} \tag{13}$$

and, for $(U,V,W)\in\tilde{\mathscr T}_{I;m_{1,1},m_{1,p},m_{2,1},m_{2,p}}$,

$$\begin{aligned}\tilde\Delta_p(U,V,W):=&U^{1/q}\cdot V^{1/p}-(U^{1/p}\cdot W^{1/q})^{p-1}V^{1/p}\cdot W^{1/q}\\ &-\big(U\cdot\mathbf 1-(U^{1/p}\cdot W^{1/q})^p\big)^{1/q}\big(V\cdot\mathbf 1-(V^{1/p}\cdot W^{1/q})^p\big)^{1/p} \qquad (14)\\ =&U^{1/q}\cdot V^{1/p}-m_{1,1}^{p-1}m_{2,1}-\big(m_{1,p}-m_{1,1}^p\big)^{1/q}\big(m_{2,p}-m_{2,1}^p\big)^{1/p}. \qquad (15)\end{aligned}$$

Indeed, the supremum in (9) is no greater than that in (11); at this point, we can only say "no greater" because the (following by (10)) expressions $x_i=(u_i/w_i)^{1/p}$ and $y_i=(v_i/w_i)^{1/p}$ of x_i and y_i in terms of u_i,v_i,w_i will only be valid if $w_i\ne 0$.

The important point here is that the set $\tilde{\mathscr T}:=\tilde{\mathscr T}_{I;m_{1,1},m_{1,p},m_{2,1},m_{2,p}}$ is compact, and the function $\tilde\Delta_p$ is continuous on it. So, $\tilde\Delta_p$ attains the (global) maximum on the set $\tilde{\mathscr T}$ whenever $\tilde{\mathscr T}\ne\emptyset$, which will be henceforth assumed wlog.

For any vector $R=(r_i)_{i\in I}\in[0,\infty)^I$, let

$$I_R:=\{i\in I\colon r_i>0\}.$$

In view of (13) and the condition (stated below (9)) that $m_{1,1},m_{1,p},m_{2,1},m_{2,p}$ are strictly positive, for any $(U,V,W)\in\tilde{\mathscr T}_{I;m_{1,1},m_{1,p},m_{2,1},m_{2,p}}$ we have

$$I_U\cap I_W\ne\emptyset\quad\text{and}\quad I_V\cap I_W\ne\emptyset. \tag{16}$$

Round 6: Further preparation for Lagrange multipliers

Fix now any triple $(U^*,V^*,W^*)\in\tilde{\mathscr T}_{I;m_{1,1},m_{1,p},m_{2,1},m_{2,p}}$ at which the maximum of $\tilde\Delta_p$ is attained. Then clearly the triple (U^*,V^*,W^*) is a maximizer of $\tilde\Delta_p$ over the set

$$\tilde{\mathscr T}^*_{I;m_{1,1},m_{1,p},m_{2,1},m_{2,p}}:=\{(U,V,W)\in\tilde{\mathscr T}_{I;m_{1,1},m_{1,p},m_{2,1},m_{2,p}}\colon I_U=I_{U^*},I_V=I_{V^*},I_W=I_{W^*}\}.$$

Also, with the triple (U^*,V^*,W^*) fixed, any triple $(U,V,W)\in\tilde{\mathscr T}^*_{I;m_{1,1},m_{1,p},m_{2,1},m_{2,p}}$ may be identified with the triple $(U|_{I_{U^*}},V|_{I_{V^*}},W|_{I_{W^*}})$ of the restrictions of U,V,W to the sets $I_U=I_{U^*},I_V=I_{V^*},I_W=I_{W^*}$, respectively; here, for instance, $U|_{I_{U^*}}=(u_i)_{i\in I_{U^*}}$; so, $\tilde\Delta_p(U,V,W)$ may be considered a function of $(U|_{I_{U^*}},V|_{I_{V^*}},W|_{I_{W^*}})$.

Round 7: Obtaining Lagrange multiplier equations

Now we are ready to apply (say) the Carathéodory–John version of the Lagrange multiplier rule (see e.g. [6, page 441]). In view of (15), there exist some real numbers

$\alpha,\lambda,\mu,\nu,\rho,\omega$ (Lagrange multipliers) – with α corresponding to the minimized $\tilde{\Delta}_p(U,V,W)$, and $\lambda,\mu,\nu,\rho,\omega$ corresponding to the restrictions in (13) and (12) on $U^{1/p}\cdot W^{1/q}, V^{1/p}\cdot W^{1/q}$, $U\cdot\mathbf{1}, V\cdot\mathbf{1}, \mathbf{1}\cdot W$, respectively – such that

$$\alpha^2+\lambda^2+\mu^2+\nu^2+\rho^2+\omega^2>0 \tag{17}$$

and the triple (U^*,V^*,W^*) is a solution to the following system of equations for (U,V,W):

$$\forall\, i\in I_U \quad \alpha(p-1)u_i^{-1/p}v_i^{1/p} = \lambda u_i^{-1/q}w_i^{1/q} + \nu, \tag{18}$$

$$\forall\, i\in I_V \quad \alpha u_i^{1/q}v_i^{-1/q} = \mu v_i^{-1/q}w_i^{1/q} + \rho, \tag{19}$$

$$\forall\, i\in I_W \quad 0 = \lambda u_i^{1/p}w_i^{-1/p} + \mu v_i^{1/p}w_i^{-1/p} + \omega. \tag{20}$$

Multiplying (both sides of) equations (18) and (19) by u_i and v_i, respectively, we have

$$\alpha(p-1)u_i^{1/q}v_i^{1/p} = \lambda u_i^{1/p}w_i^{1/q} + \nu u_i, \tag{21}$$

$$\alpha u_i^{1/q}v_i^{1/p} = \mu v_i^{1/p}w_i^{1/q} + \rho v_i \tag{22}$$

for all $i\in I$.

A difficulty in analyzing these Lagrange multiplier equations is that some of the Lagrange multipliers $\alpha,\lambda,\mu,\nu,\rho,\omega$ may take zero values. In a certain sense, this corresponds to the fact the difference between the left and right sides of inequality (2) can attain its maximum (zero) value in a number of ways, including the cases when $X=Y$ and when Y is a constant. Also, we have to account for cases when some of the values of u_i,v_i,w_i are 0, that is, when i is not in the corresponding sets I_U,I_V,I_W.

In particular, we have to consider the cases when $u_i>0$ or $v_i>0$ while $w_i=0$ $\big($that is, when $i\in(I_U\cup I_V)\setminus I_W\big)$. Recalling (10), we see that, in terms of the "original, pre-compactification" variables x_i,y_i,w_i, these cases reflect the possibility for these variables to vary in such a way that for some $i\in I$ we have $w_i\downarrow 0$ while $x_i\to\infty$ or $y_i\to\infty$ and, moreover, $x_i^p w_i$ or, respectively, $y_i^p w_i$ converges to a finite nonzero limit. This kind of phenomena may be thought of as part of the mass of the "distribution" of U or V running away to ∞. This brings us to the following round.

Round 8: Analysis of Lagrange multipliers, part I: Removing "the masses at ∞"
Take any triple $(U,V,W)\in([0,\infty)^I)^3$ satisfying the Lagrange multiplier equations (18)–(20). On the set I_W, define the probability space by the condition $\mathsf{P}(\{i\})=w_i$ for all $i\in I_W$, and then define r.v.'s X and Y on this probability space by the conditions

$$X(i)=x_i:=(u_i/w_i)^{1/p} \text{ and } Y(i)=y_i:=(v_i/w_i)^{1/p} \text{ for all } i\in I_W. \tag{23}$$

The r.v.'s X and Y are well defined, because $w_i>0$ for all $i\in I_W$ and $\sum_{i\in I_W}w_i=\sum_{i\in I}w_i=1$. Then, by (14) and (7), $\tilde{\Delta}_p(U,V,W)=\Delta_{p;A,B,C}(X,Y)$, where

$$A:=\sum_{i\notin I_W}u_i^{1/q}v_i^{1/p}, \quad B:=\sum_{i\notin I_W}u_i, \quad C:=\sum_{i\notin I_W}v_i, \tag{24}$$

"the masses at ∞". By Hölder's inequality, here the condition $A \leqslant B^{1/q}C^{1/p}$ in Lemma 1 holds. So, $\tilde{\Delta}_p(U,V,W) \leqslant \Delta_p(X,Y)$.

Thus, it remains to show that $\Delta_p(X,Y) \leqslant 0$ for X and Y as in (23), with $(U,V,W) \in ([0,\infty)^I)^3$ satisfying the Lagrange multiplier equations (18)–(20).

Round 9: Analysis of Lagrange multipliers, part II: Reduction to the case $Y = X + t$, $t \in \mathbb{R}$

In terms of the x_i's and y_i's as in (23), for $i \in I_W$ equations (20), (21), (22) can be rewritten as

$$0 = \lambda x_i + \mu y_i + \omega, \tag{25}$$

$$\alpha(p-1)x_i^{p-1}y_i = \lambda x_i + \nu x_i^p, \tag{26}$$

$$\alpha x_i^{p-1}y_i = \mu y_i + \rho y_i^p. \tag{27}$$

LEMMA 2. *Take any pair $(X,Y) \in ([0,\infty)^{I_W})^2$ satisfying equations (25)–(27) with $\mu = 0$. Then $\Delta_p(X,Y) \leqslant 0$.*

Proof. This proof consists in the consideration of a system of simple cases, keeping in mind the condition $\mu = 0$.

Case 1: $\rho = 0$.

Subcase 1.1: $\rho = 0 \neq \alpha$. Then, by (27), $x_i^{p-1}y_i = 0$ for all $i \in I_W$. So, $\mathsf{E}X^{p-1}Y = 0$, and inequality $\Delta_p(X,Y) \leqslant 0$ obviously holds.

Subcase 1.2: $\rho = 0 = \alpha$. Then, by (26), $\lambda x_i + \nu x_i^p = 0$ for all $i \in I_W$.

Subsubcase 1.2.1: $\rho = 0 = \alpha$ and $\lambda = 0 = \nu$. Then, by (25), $\omega = 0$. So, we have a contradiction with (17).

Subsubcase 1.2.2: $\rho = 0 = \alpha$ and $\lambda \neq 0$. Then, by (25), x_i does not depend on $i \in I_W$; that is, the r.v. X is a constant, and hence $\Delta_p(X,Y) = 0$.

Subsubcase 1.2.2: $\rho = 0 = \alpha$ and $\lambda = 0 \neq \nu$. Then, by (26), $x_i = 0$ for all on $i \in I_W$; that is, $X = 0$, and hence $\Delta_p(X,Y) = 0$.

Case 2: $\rho \neq 0$.

Subcase 2.1: $\rho \neq 0 = \lambda$. Then, by (26) and (27), $\nu x_i^p = (p-1)\rho y_i^p$ for all $i \in I_W$. So, $Y = cX$ for some real $c \geqslant 0$, and hence $\Delta_p(X,Y) = 0$.

Subcase 2.2: $\rho \neq 0 \neq \lambda$. Then, by (25), x_i does not depend on $i \in I_W$; that is, the r.v. X is a constant, and hence $\Delta_p(X,Y) = 0$.

Thus, indeed in all cases we have $\Delta_p(X,Y) \leqslant 0$. □

So, by Lemma 2, wlog $\mu \neq 0$. So, in view of (25), $Y = kX + t$ for some real k and t.

Now we need Chebyshev's integral inequality, which states that, if f and g are nondecreasing functions from $\mathbb{R}$ to $\mathbb{R}$, then for any r.v. Z one has $\mathsf{E}f(Z)g(Z) \geqslant \mathsf{E}f(Z)\,\mathsf{E}g(Z)$ whenever all the three expectations here are finite; see e.g. Corollary 2 on page 318 in [2] (with $n=1$, $\phi=1$, and the probability distribution of Z to play the role of the measure λ there). This inequality follows immediately by taking the expectation of both sides of the obvious inequality $(f(Z)-f(Z_1))(g(Z)-g(Z_1))\geqslant 0$, where Z_1 is an independent copy of Z.

By Chebyshev's integral inequality and the mentioned log-convexity of $\mathsf{E}X^\alpha$ in α, for $Y=kX+t$ with $k\leqslant 0$ we have $\mathsf{E}X^{p-1}Y\leqslant \mathsf{E}X^{p-1}\,\mathsf{E}Y\leqslant \mathsf{E}^{p-1}X\,\mathsf{E}Y$, which yields $\Delta_p(X,Y)\leqslant 0$, in view of (6). So, wlog $k>0$, and then, because of the positive homogeneity of $\Delta_p(X,Y)$ in Y, wlog $Y=X+t$.

Round 10: Analysis of the case $Y=X+t$, $t\in\mathbb{R}$
Thus, to finish the proof of (2), it remains to prove

LEMMA 3. *For all real t such that the r.v. $X+t$ is nonnegative, we have*

$$\delta(t):=\Delta_p(X,X+t)\leqslant 0. \tag{28}$$

Proof. In view of (6), $\delta(0)=0=\delta'(0)$. So, it is enough to show that the function δ is concave or, equivalently, that the function f given by the formula

$$f(t):=\big(\mathsf{E}(X+t)^p-(\mathsf{E}X+t)^p\big)^{1/p}$$

for $t\in T:=\{s\in\mathbb{R}\colon X+s\geqslant 0\}$ is convex. The set T is an interval. So, it suffices to show that $f''(t)\geqslant 0$ for all t in the interior $\operatorname{int}T$ of the set T or, equivalently, that

$$H:=(m_p-m_1^p)(m_{p-2}-m_1^{p-2})-(m_1^{p-1}-m_{p-1})^2\geqslant 0, \tag{29}$$

where

$$m_r:=\mathsf{E}Y^r,$$

$Y=X+t$, and $t\in\operatorname{int}T$. Here, by the positive homogeneity, for any fixed $t\in\operatorname{int}T$ wlog

$$m_1=\mathsf{E}Y=1.$$

In principle, inequality (29) can be proved by minimizing the pth moment m_p of the r.v. Y given the moments $m_{p-2},m_0=1,m_{p-1},m_1=1$ of Y of orders $p-2,0,p-1,1$. Using results of, say, [8, 5], we may assume that the support of the distribution of Y consists of at most $\operatorname{card}\{p-2,0,p-1,1\}=4$ points, where card denotes the cardinality. This would reduce (29) to a minimization problem involving 8 variables (not counting p): 4 variables for the points of the support of the distribution and 4 variables for the corresponding masses. In our particular case, the minimization problem can be further simplified by noticing that the moment functions mapping $x\in[0,\infty)$ to $x^{p-2},x^0,x^{p-1},x^1,x^p$ form a Tchebycheff–Markov system and hence we may assume that the support of the distribution of Y consists of at most 2 points; see e.g. [1] or [4, Propositions 1 and 2]. Thus, we would have to deal with 4 variables (not counting p): 2 variables for the points of the support and 2 variables for the masses. The values of the masses could be eliminated by solving the system of equations $m_0=1$ and $m_1=1$, which are linear with respect to the two masses. That would leave us with two variables, one for each of the two support points, plus another variable for p.

Fortunately, again in our particular case, we can actually use a simple trick to reduce the problem to one involving just one variable in addition to p. Indeed, by the mentioned Lyapunov inequality (that is, the log-convexity of m_r in r), $m_{p-1} \leqslant m_1^{p-1} m_0^{2-p} = 1$, $1 = m_1 \leqslant m_{p-1}^{p-1} m_p^{2-p}$, and $1 = m_0 \leqslant m_{p-2}^{p-1} m_{p-1}^{2-p}$, whence

$$1 \geqslant m_{p-1} \geqslant m_* \vee m_{**}, \tag{30}$$

where

$$m_* := m_p^{-(2-p)/(p-1)} \quad \text{and} \quad m_{**} := m_{p-2}^{-(p-1)/(2-p)}.$$

Next, $m_* \geqslant m_{**}$ iff $m_p^{(2-p)^2} \leqslant m_{p-2}^{(p-1)^2}$. So, by (29) and (30),

$$H \geqslant H_* := (m_p - 1)\big(m_p^{(2-p)^2/(p-1)^2} - 1\big) - (1 - m_*)^2 \quad \text{if } m_p^{(2-p)^2} \leqslant m_{p-2}^{(p-1)^2},$$

$$H \geqslant H_{**} := \big(m_{p-2}^{(p-1)^2/(2-p)^2} - 1\big)(m_{p-2} - 1) - (1 - m_{**})^2 \quad \text{if } m_p^{(2-p)^2} \geqslant m_{p-2}^{(p-1)^2}.$$

Note that H_* depends only on p and m_p, whereas H_{**} depends only on p and m_{p-2}.

It suffices to show that $H_* \geqslant 0$ for all $p \in (1,2)$ and real $m_p \geqslant 1$ and that $H_{**} \geqslant 0$ for all $p \in (1,2)$ and real $m_{p-2} \geqslant 1$. At this point, m_p and m_{p-2} may be considered free variables, with the only restriction that they take real values $\geqslant 1$. Then, under the one-to-one correspondence between these free variables given by the formula $m_p^{(2-p)^2} \leftrightarrow m_{p-2}^{(p-1)^2}$, every value of H_{**} turns into the corresponding value of H_*, and vice versa. So, it is enough to show that $H_* \geqslant 0$ for all $p \in (1,2)$ and real $m_p \geqslant 1$.

Making now the substitution $m_p = e^{(p-1)^2 s}$, we can write

$$\frac{H_*}{e^{2(p-2)(p-1)s}} = h(s) := 2e^{(2-p)(p-1)s} - e^{(3-p)(p-1)s} - e^{(2-p)ps} + e^s - 1.$$

So, it suffices to show that $h(s) \geqslant 0$ for all real $s \geqslant 0$ (and all $p \in (1,2)$). Since h is a linear combination of exponential functions, this can be done essentially algorithmically. Indeed, let $h_1(s) := h'(s)e^{(p-2)(p-1)s}$ and $h_2(s) := h_1'(s)e^{-(3-3p+p^2)s}$. Then $h_2'(s)(2-p)^{-2}(p-1)^{-2} = pe^{-(p-1)^2 s} + (3-p)e^{-(2-p)^2 s}$, which is manifestly > 0. So, h_2 is increasing (on the interval $[0,\infty)$), with $h_2(0) = 0$. So, $h_2 \geqslant 0$ and hence h_1 is nondecreasing, with $h_1(0) = 0$. So, $h_1 \geqslant 0$ and hence h is nondecreasing, with $h(0) = 0$. So, indeed $h \geqslant 0$. Thus, Lemma 3 is completely proved. □

This completes the proof of inequality (2) and hence the proof of (4). □

Take now any $\theta \in [0,1]$. For real $t \geqslant 0$, let

$$g(t) := g_{\theta;X,Y}(t) := \mathscr{E}_{p,\theta}(X + tY) - \mathscr{E}_{p,\theta}(X) - t\mathscr{E}_{p,\theta}(Y). \tag{31}$$

If $\mathscr{E}_{p,\theta}(X + tY) = 0$ for some $t \geqslant 0$, then obviously $g(t) \leqslant 0$; otherwise $\big($that is, if $\mathscr{E}_{p,\theta}(X + tY) > 0\big)$, we can write

$$g'(t) = \mathscr{C}_{p,\theta}(X + tY, Y)\mathscr{E}_{p,\theta}(X + tY)^{1-p} - \mathscr{E}_{p,\theta}(Y) \leqslant 0, \tag{32}$$

in view of already proved inequality (4); here, $g'(0)$ is understood as the right derivative of g at 0. So, for each real $t \geqslant 0$ such that $g(t) > 0$, we have $g'(t) \leqslant 0$. Also, $g(0) = 0$ and the function g is continuous. Suppose now that $g(1) > 0$ and let $a := \sup\{t \in [0,1]\colon g(t) = 0\}$. Then $g(a) = 0$ and $0 \leqslant a < 1$; also, $g > 0$ and hence $g' \leqslant 0$ on $(a,1]$. In view of the mean value theorem, this contradicts the conditions $g(a) = 0 < g(1)$. Therefore, $g(1) \leqslant 0$; that is, inequality (3) holds.

To finish the proof of Theorem 1, it remains to show that inequalities (3) and (4) are false in general if $p > 2$ and $\theta \in (0,1]$. To this end, suppose, e.g., that $\mathsf{P}(X=1) = \mathsf{P}(X=0) = 1/2$. Let $\delta_{p,\theta}(t) := \Delta_{p,\theta}(X, X+t)$; cf. (28) and (5). Then $\delta_{p,\theta}(0) = 0 = \delta'_{p,\theta}(0+)$, whereas $\delta''_{p,\theta}(0+) = (p-1)\theta^p/(2^p - 2\theta^p) > 0$, whence $\Delta_{p,\theta}(X, X+c) = \delta_{p,\theta}(c) > 0$ for small enough $c > 0$. Take any such c and let $Y := X + c$, so that $\Delta_{p,\theta}(X,Y) > 0$, that is, inequality (4) is false. So, by (32), $g'(0) > 0$, which implies $g(t) > 0$ for all small enough $t > 0$. Thus, (3) with tY in place of Y is false if $t > 0$ is small enough. $\big($One might note that here $\delta''_{p,\theta}(0+) = -\infty < 0$ if $1 < p < 2$ and $\delta''_{2,\theta}(0) = -(1-\theta^2)/(2-\theta^2) \leqslant 0.\big)$

The entire proof of Theorem 1 is now complete.

REMARK 1. The simple deduction of (3) from (4) in the paragraph containing formulas (31) and (32) is essentially reversible, so that, vice versa, (4) is easy to deduce from (3). Indeed, take again any $\theta \in [0,1]$. If $\mathscr{E}_{p,\theta}(X) = 0$, then $\mathsf{P}(X = a) = 1$ for some real constant $a \geqslant 0$; moreover, if, in addition, $\theta < 1$, then necessarily $a = 0$. So, inequality (4) is trivial if $\mathscr{E}_{p,\theta}(X) = 0$. Therefore, wlog $\mathscr{E}_{p,\theta}(X) > 0$ and hence $g'(0)$ exists (cf. (32)), where g is as in (31). Moreover, (3) with tY in place of Y yields $g(t) \leqslant 0$ for $t \geqslant 0$. Since $g(0) = 0$, we have $g'(0) \leqslant 0$. Now (4) follows by the equality in (32).

REFERENCES

[1] S. KARLIN AND W. J. STUDDEN, *Tchebycheff systems: With applications in analysis and statistics*, Pure and Applied Mathematics, vol. XV. Interscience Publishers John Wiley & Sons, New York-London-Sydney, 1966.

[2] J. H. B. KEMPERMAN, *On the FKG-inequality for measures on a partially ordered space*, Nederl. Akad. Wetensch. Proc. Ser. A 80 = Indag. Math. **39**, 4 (1977), 313–331.

[3] MATHOVERFLOW, *A Minkowski-like inequality*, `https://mathoverflow.net/q/292327` (version: 2018-02-07).

[4] I. PINELIS, *Tchebycheff systems and extremal problems for generalized moments: a brief survey*, `http://arxiv.org/abs/1107.3493`, 2011.

[5] I. PINELIS, *On the extreme points of moments sets*, Math. Methods Oper. Res. **83**, 3 (2016), 325–349.

[6] B. H. POURCIAU, *Modern multiplier rules*, Amer. Math. Monthly **87**, 6 (1980), 433–452.

[7] H. L. ROYDEN, *Real analysis*, Second edition, 1968.

[8] G. WINKLER, *Extreme points of moment sets*, Math. Oper. Res. **13**, 4 (1988), 581–587.

(Received October 7, 2018)

Iosif Pinelis
Department of Mathematical Sciences
Michigan Technological University
Houghton, Michigan 49931, USA
e-mail: ipinelis@mtu.edu