Multidimensional Hermite–Hadamard inequalities and the convex order

J. de la Cal a,∗, J. Cárcamo b

a Departamento de Matemática Aplicada y Estadística e Investigación Operativa, Facultad de Ciencia y Tecnología, Universidad del País Vasco, Apartado 644, 48080 Bilbao, Spain
b Departamento de Matemáticas, Facultad de Ciencias, Universidad Autónoma de Madrid, 28049 Madrid, Spain

Received 8 September 2005
Available online 10 January 2006
Submitted by William F. Ames

Abstract

The problem of establishing inequalities of the Hermite–Hadamard type for convex functions on n-dimensional convex bodies translates into the problem of finding appropriate majorants of the involved random vector for the usual convex order. We present two results of partial generality which unify and extend the most part of the multidimensional Hermite–Hadamard inequalities existing in the literature, at the same time that lead to new specific results. The first one fairly applies to the most familiar kinds of polytopes. The second one applies to symmetric random vectors taking values in a closed ball for a given (but arbitrary) norm on R n. Related questions, such as estimates of approximation and extensions to signed measures, also are briefly discussed.

© 2005 Elsevier Inc. All rights reserved.

Keywords: Hermite–Hadamard inequality; Convex order; H-majorant; Choquet theory; Jensen’s inequality; Convex body; Polytope; Representation system; Uniform distribution

1. Introduction

The Hermite–Hadamard (double) inequality for convex functions on an interval of the real line is usually stated as follows.
Theorem 1. Let $f$ be a real continuous convex function on the finite interval $[a, b]$. Then

$$f\left(\frac{a+b}{2}\right) \leq \frac{1}{b-a} \int_a^b f(x) \, dx \leq \frac{f(a) + f(b)}{2}. \quad (1)$$

There is an extensive literature devoted to develop applications of this inequality, as well as to discuss its extensions, by considering other measures, other kinds of convexity, or higher dimensions (see, for instance, [2–7,9,12,13,17] and the references therein). An account of many of such realizations is given in [6].

In the present paper, we are concerned with analogues of Theorem 1 for convex functions on $n$-dimensional convex bodies. We start by observing that the problem allows for a probabilistic approach. Actually, in probabilistic terms, (1) says that

$$f(\xi) \leq \mathbb{E} f(\xi) \leq \mathbb{E} f(\xi^*),$$

where $\mathbb{E}$ denotes mathematical expectation, $\xi$ (respectively $\xi^*$) is a random variable having the uniform distribution on $[a, b]$ (respectively on $\{a, b\}$), and it should be observed that $\mathbb{E} \xi = \mathbb{E} \xi^* = (a + b)/2$. Also, Theorem 1 means that $\mathbb{E} \xi \leq_{cx} \mathbb{E} \xi^*$, where $\leq_{cx}$ denotes the convex order for random variables (see [10]). This leads us to introduce the following terminology and notations.

Throughout the paper, unless otherwise specified, $K$ denotes a (nonempty) compact convex subset of $\mathbb{R}^n$, and $K^*$ is the set of its extreme points; this means that $K$ is the convex hull of $K^*$, and each $v \in K^*$ does not belong to the convex hull of $K^* - \{v\}$. The dimension of $K$ is defined to be the dimension of the affine hull of $K$ (see [18]).

Definition 1. An Hermite–Hadamard majorant (H-majorant, for short) of a given $K$-valued random vector $\xi$ is a $K^*$-valued random vector $\xi^*$ such that $\xi \leq_{cx} \xi^*$, i.e.,

$$\mathbb{E} f(\xi) \leq \mathbb{E} f(\xi^*), \quad f \in \mathcal{C}_K, \quad (2)$$

where $\mathcal{C}_K$ stands for the set of all real continuous convex functions on $K$. Alternatively, we also say that the distribution of $\xi^*$ is an H-majorant of the distribution of $\xi$.

Remark 1. Observe that (2) implies that $\mathbb{E} f(\xi) = \mathbb{E} f(\xi^*)$, whenever $f$ is an affine function, because such a function is concave and convex at the same time. Therefore, (2) entails that $\xi$ and $\xi^*$ have the same barycenter or expectation, and we also have, by Jensen’s inequality, $\mathbb{E} \xi \leq_{cx} \xi$. Thus, the problem of establishing Hermite–Hadamard inequalities translates into the problem of finding H-majorants. In particular, when $K^*$ is a finite set, and the H-majorant $\xi^*$ is uniform on $K^*$, i.e.,

$$P(\xi^* = v) = |K^*|^{-1}, \quad v \in K^*,$$

where $|K^*|$ is the number of elements of $K^*$, the Hermite–Hadamard inequality takes the form

$$f\left(|K^*|^{-1} \sum_{v \in K^*} v\right) \leq \mathbb{E} f(\xi) \leq |K^*|^{-1} \sum_{v \in K^*} f(v).$$

The following theorem is the finite-dimensional version of a more general result established by Niculescu [13] on the basis of Choquet’s theory [15].
Theorem 2. Every $K$-valued random vector $\xi$ always has at least one $H$-majorant.

This interesting result leaves open the problem of finding explicit $H$-majorants, a necessary task in order to achieve concrete inequalities of the Hermite–Hadamard type. When $K$ is a simplex (in particular, the interval $[a, b]$), there is a unique probability measure on $K^*$ having a given barycenter, which implies the uniqueness of the $H$-majorant (see Section 3.1), but this fact is no longer true for other convex bodies (the Euclidean closed unit ball, for instance), and the problem of finding concrete versions of $\xi^*$ seems to require specific techniques depending heavily both on the geometric structure of $K$ and on (the probability distribution of) $\xi$. In other words, we can hardly expect methods of full generality.

In this paper, we present methods and results of partial generality which unify and extend the most part of the concrete multidimensional Hermite–Hadamard inequalities existing in the literature, at the same time that lead to new specific results.

When $K^*$ is a finite set, we propose a method to find $H$-majorants based upon the construction of a suitable family of functions. This is done in Section 2. The method, which is obviously inspired in the barycentric coordinates, fairly applies to familiar convex bodies such as simplices, hyperrectangles, and crosspolytopes. Moreover, in such cases, it allows us to give simple sufficient conditions guaranteeing that the $H$-majorant is uniform on $K^*$ (see Section 3).

In Section 4, we give explicit $H$-majorants for symmetric random vectors taking values in the closed unit ball for a given norm on $\mathbb{R}^n$. To appreciate the scope of the result, it should be recalled that, if $K$ is a symmetric compact convex subset of $\mathbb{R}^n$ having $0$ as an interior point for the usual topology on $\mathbb{R}^n$, then $K$ is the closed unit ball for the norm on $\mathbb{R}^n$ given by

$$
\|x\| := \inf\{t > 0: t^{-1}x \in K\}, \quad x \in \mathbb{R}^n
$$

(the Minkowski functional of $K$) (see [8, Chapter 1]).

In Section 5, we discuss extensions of previous results to Cartesian products of convex bodies.

In Section 6, we consider two different questions closely related to the problem under consideration, namely, the representation of $Ef(\xi)$ in terms of the $H$-majorant $\xi^*$, and the estimation of the differences between the middle term and the extreme terms in some Hermite–Hadamard inequalities.

Finally, in the last section, we briefly discuss the extension of Theorem 2 to signed measures, and we show by a simple counterexample that a recent result of Niculescu fails to be true even in the one-dimensional case.

2. Polytopes: representation systems

A polytope in $\mathbb{R}^n$ is a compact convex set $K \subset \mathbb{R}^n$ having a finite set $K^*$ of extreme points (customarily called vertices). For this kind of convex body, we introduce the notion of a representation system, which is inspired in the barycentric coordinates.

Definition 2. A representation system on a polytope $K$ is a family $\mathcal{H} := \{h_v: v \in K^*\}$ of real measurable functions on $K$ fulfilling the following three assumptions:

(a) $h_v \geq 0$, for all $v \in K^*$;

(b) $\sum_{v \in K^*} h_v = 1$;

(c) $\sum_{v \in K^*} h_v(x)v = x$, for all $x \in K$.

Theorem 3. If $K \subset \mathbb{R}^n$ is a given polytope, then there is at least one representation system on $K$. 
Proof. We proceed by induction on $|K^*|$. The conclusion being obviously true when $|K^*| = 1$, assume that $|K^*| > 1$, let $w$ be a fixed element of $K^*$, and (by induction hypothesis) let $\mathcal{H}' := \{h'_v: v \in K^* := K^* - \{w\}\}$ be a representation system on $K_1 := \text{the convex hull of } K^*$. We construct a family of real functions on $K$, $H := \{h_v: v \in K^*\}$, in the following way. First, we set

$$
h_w(w) := 1, \quad h_v(w) := 0 \quad (v \neq w).
$$

Now, fix $x \in K - \{w\}$. Then, the set

$$
K_1(x) := \{ y \in K_1: y - w = a(x - w) \text{ for some } a \geq 1 \}
$$

is a nonempty closed segment. Let $T(x)$ be the unique element in $K_1(x)$ such that

$$
\|w - T(x)\| = \inf\{ \|y - w\|: y \in K_1(x) \},
$$

where $\|\cdot\|$ is a fixed norm on $\mathbb{R}^n$, and set

$$
h_w(x) := \frac{\|x - T(x)\|}{\|w - T(x)\|}, \quad h_v(x) := \frac{\|x - w\|}{\|w - T(x)\|} h'_v(T(x)) \quad (v \neq w).
$$

It is readily checked that $\mathcal{H}$ fulfills the conditions to be a representation system on $K$. \qed

If the polytope is a simplex the representation system is unique, but such a fact is no longer true for other polytopes (see the next section).

**Theorem 4.** Let $K$ be a polytope, and let $H := \{h_v: v \in K^*\}$ be a representation system on $K$. If $\xi$ is a $K$-valued random vector, then the random vector $\xi^*$ having the distribution

$$
P(\xi^* = v) := E[h_v(\xi)], \quad v \in K^*,
$$

(3)

is an $H$-majorant of $\xi$.

**Proof.** Assumptions (a) and (b) on $\mathcal{H}$ guarantee that (3) actually defines a probability distribution on $K^*$. On the other hand, we have by assumptions (a)–(c) and the convexity of $f \in C_K$,

$$
f(\xi) \leq \sum_{v \in K^*} h_v(\xi) f(v),
$$

and (2) follows on taking expectations. \qed

**Remark 2.** Theorems 3 and 4 provide a new proof of Theorem 2, in the case that $K$ is a polytope.

The applicability of the preceding theorem requires the construction of explicit examples of representation systems. In the next section, we do such a construction for some familiar kinds of polytopes. In Section 5, the construction is extended to Cartesian products. The following (isomorphism-type) lemma shows that we only need to consider prototypical cases.

**Lemma 1.** Let $K$ and $\xi$ be the same as in Theorem 2. Let $K_0$ be another compact convex subset of $\mathbb{R}^n$, and assume that there is a bijective affine transformation $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$ such that $T(K) = K_0$. Then, we have:

(a) $T(K^*) = K_0^*$. In particular, $K_0$ is a polytope if $K$ is.

(b) If $\xi^*$ is an $H$-majorant of $\xi$, then $\xi_0^* := T(\xi^*)$ is an $H$-majorant of $\xi_0 := T(\xi)$.
(c) If $K$ is a polytope, and $\mathcal{H} := \{h_v: v \in K^*\}$ is a representation system on $K$, then the family $\mathcal{H}^* := \{h_w^* : w \in K_0^*\}$ defined by $h_w^* := h_{T^{-1}(w)} \circ T^{-1}$ is a representation system on $K_0$.

**Proof.** Assertions (a) and (c) readily follow from the fact that $T$ is bijective and we have

$$T \left( \sum_{i=1}^{m} a_i x_i \right) = \sum_{i=1}^{m} a_i T(x_i),$$

for all $m \geq 2$, all $x_1, \ldots, x_m \in \mathbb{R}^n$, and all real numbers $a_1, \ldots, a_m$ such that $\sum_{i=1}^{m} a_i = 1$ (see [18, p. 23]). Finally, (b) follows from (a) and the fact that, for each $f \in C_{K_0}$, we have

$$f(\xi_0) = g(\xi) \quad \text{and} \quad f(\xi_0^*) = g(\xi^*),$$

where $g := f \circ T \in C_K$. □

**Remark 3.** In the setting of the preceding lemma, it is immediate that both $K$ and $K_0$ have the same dimension. Also, if $K$ is $n$-dimensional, and $\xi$ (respectively $\xi^*$) has the uniform distribution on $K$ (respectively $K^*$), then $\xi_0$ (respectively $\xi_0^*$) has the uniform distribution on $K_0$ (respectively $K_0^*$). We recall that, given a Borel set $B \subset \mathbb{R}^n$ having nonzero finite $n$-dimensional volume, the uniform distribution on $B$ is the normalized $n$-dimensional Lebesgue measure on $B$, $[\text{Vol}(B)]^{-1} \, dx$, and we can write

$$Ef(\xi) = \frac{1}{\text{Vol}(B)} \int_B f(x) \, dx,$$

whenever $\xi$ is uniform on $B$ and $f(\xi)$ is integrable.

### 3. Examples

#### 3.1. Simplices

Let $A := \{a_0, a_1, \ldots, a_n\}$ be an affine basis for $\mathbb{R}^n$. This means that $\{a_1 - a_0, \ldots, a_n - a_0\}$ is a linear basis for $\mathbb{R}^n$ and, then, each point $x \in \mathbb{R}^n$ can be expressed uniquely in the form $x = \sum_{i=0}^{n} h_{a_i}(x) a_i$, where $\sum_{i=0}^{n} h_{a_i}(x) = 1$. The scalars $h_{a_i}(x)$ ($i = 0, \ldots, n$) are called the barycentric coordinates of $x$ relative to $A$. The $n$-simplex of basis $A$, to be denoted by $K$, is defined to be the convex hull of $A$, that is, $K := \{x \in \mathbb{R}^n: h_{a_i}(x) \geq 0, \quad i = 0, \ldots, n\}$.

Note that $K$ can also be described as the set of barycenters (expectations) of all possible probability distributions on $A$, and the uniqueness of barycentric coordinates means that each probability distribution on $A$ is determined by its barycenter. It is also clear that $K$ is a polytope in $\mathbb{R}^n$, with $K^* = A$, and that $\mathcal{H} := \{h_{a_i}: i = 0, 1, \ldots, n\}$ is a representation system on $K$.

In particular, when

$$a_0 := 0 =: e_0, \quad a_i := e_i \quad (i = 1, \ldots, n), \quad (4)$$

where $\{e_1, \ldots, e_n\}$ is the canonical basis for $\mathbb{R}^n$, the $n$-simplex is called standard $n$-simplex, and denoted by $K_0$. In this case, the barycentric coordinates of the point $x := (x_1, \ldots, x_n) \in \mathbb{R}^n$ are given by

$$h_{e_0}(x) = 1 - \sum_{i=1}^{n} x_i, \quad h_{e_i}(x) = x_i \quad (i = 1, \ldots, n).$$
We also observe that, if \( T \) is the (unique) affine transformation determined by the conditions
\[
T(a_i) = e_i, \quad i = 0, \ldots, n,
\]
then \( T \) is bijective, and we have \( T(K) = K_0 \), as well as
\[
h_{a_i}(x) = h_{e_i}(T(x)), \quad i = 0, \ldots, n, \quad x \in \mathbb{R}^n,
\]
showing that each \( h_{a_i} \) is a real affine function on \( \mathbb{R}^n \).

With the preceding notations, we assert the following.

**Corollary 1.** If \( \xi \) is a \( K \)-valued random vector, then the random vector \( \xi^\ast \) having the distribution given by
\[
P(\xi^\ast = a_i) := E[h_{a_i}(\xi)] = h_{a_i}(E(\xi)) = h_{e_i}(E(T(\xi))), \quad i = 0, \ldots, n,
\]
is the unique (in the sense of distribution) \( H \)-majorant of \( \xi \).

**Proof.** The result directly follows from Theorem 4 and the fact that the barycenter determines the distribution of \( \xi^\ast \). \( \square \)

**Remark 4.** Corollary 1 was early obtained by Niculescu [13] using Theorem 2 instead of Theorem 4. The version for the interval \([a, b]\) saying that
\[
f(E\xi) \leq Ef(\xi) \leq \frac{b - E\xi}{b - a} f(a) + \frac{E\xi - a}{b - a} f(b), \quad f \in C[a, b],
\]
was first obtained by Fink [7] (see also [10, Example 1.10.5]).

**Remark 5.** In the setting of Corollary 1, \( \xi^\ast \) has the uniform distribution on \( K^\ast \) if and only if each component of \( T(\xi) \) has expectation \( 1/(n + 1) \). This applies, for instance, in the following cases:

(a) When \( nT(\xi) \) has the multinomial distribution with parameters \( n \) and \( (1/(n + 1), \ldots, 1/(n + 1)) \) [16, Chapter 28].

(b) When \( T(\xi) \) has the Dirichlet distribution with (multi)parameter \( (\alpha, \ldots, \alpha) \in (0, \infty)^{n+1} \) [16, Chapter 38]; in particular, when \( T(\xi) \) has the uniform distribution on \( K_0 \) (\( \alpha = 1 \)), which amounts to saying that \( \xi \) has the uniform distribution on \( K \).

### 3.2. Hyperrectangles

In this subsection, \( K \) denotes the hyperrectangle \([a_1, b_1] \times \cdots \times [a_n, b_n]\), where \( a_i < b_i \) \((i = 1, \ldots, n)\). The set of its \( 2^n \) extreme points (or vertices) is
\[
K^\ast := \{ v := (v_1, \ldots, v_n) \in \mathbb{R}^n : v_i = a_i or b_i, \quad i = 1, \ldots, n \}.
\]
We also denote by \( K_0 \) the hypercube \([-1, 1]^n\), i.e., the closed unit ball for the \( l_\infty \)-norm on \( \mathbb{R}^n \), and by \( T \) the bijective affine transformation from \( \mathbb{R}^n \) into itself (uniquely) determined by the relations
\[
T(a_0) = 0, \quad T\left( a_0 + \frac{b_i - a_i}{2} e_i \right) = e_i \quad (i = 1, \ldots, n),
\]
where \( a_0 \) is the center of \( K \), i.e., the point whose \( i \)th coordinate is \((a_i + b_i)/2 (i = 1, \ldots, n)\), and \( \{e_1, \ldots, e_n\} \) is again the canonical basis for \( \mathbb{R}^n \). It is immediately checked that \( T(K) = K_0 \).

Finally, let \( \mathcal{H} := \{h_v: v \in K^*\} \) be the family of real functions on \( K \) defined by

\[
h_v(x) := \prod_{i=1}^n h_{v_i}(x_i), \quad x \in K,
\]

where \( v_i \) (respectively \( x_i \)) stands for the \( i \)th coordinate of \( v \) (respectively \( x \)) and

\[
h_{v_i}(x_i) := \frac{b_i - x_i}{b_i - a_i} \quad \text{or} \quad \frac{x_i - a_i}{b_i - a_i}, \quad a_i \leq x_i \leq b_i,
\]

according to \( v_i = a_i \) or \( v_i = b_i \).

By induction on \( n \), it is easy to see that \( \mathcal{H} \) is a representation system on \( K \). From Theorem 4, we can therefore assert the following.

**Corollary 2.** If \( \xi \) is a \( K \)-valued random vector, then the random vector \( \xi^\ast \) having the probability distribution given by

\[
P(\xi^\ast = v) := E[h_v(\xi)], \quad v \in K^*,
\]

is an \( H \)-majorant of \( \xi \).

The following corollary gives conditions guaranteeing that \( \xi^\ast \) is uniformly distributed on \( K^* \). We say that an \( n \)-dimensional random vector \( \xi := (\xi_1, \ldots, \xi_n) \) is **multisymmetric** (or that it has a multisymmetric distribution), if, for each choice of signs \( \varepsilon := (\varepsilon_1, \ldots, \varepsilon_n) \in \{-1, 1\}^n \), the random vector \( \xi_{\varepsilon} := (\varepsilon_1\xi_1, \ldots, \varepsilon_n\xi_n) \) has the same probability distribution as \( \xi \).

**Corollary 3.** In the setting of the preceding corollary, the random vector \( \xi^\ast \) is uniformly distributed on \( K^* \), if one of the following two conditions is fulfilled:

(a) The components of \( \xi \) are independent and \( E\xi = a_0 \).
(b) The \( K_0 \)-valued random vector \( T(\xi) \) is multisymmetric.

**Proof.** In case (a), we have, for each \( v \in K^* \),

\[
E[h_v(\xi)] = \prod_{i=1}^n E h_{v_i}(\xi_i) = h_v(E\xi) = h_v(a_0) = 1/2^n,
\]

the first equality by (8) and the independence assumption. The conclusion in case (b) follows from the fact that we have, for each \( v \in K^* \),

\[
E[h_v(\xi)] = E[h_{w}(T(\xi))], \quad w := T(v), \quad w_0 := (1, 1, \ldots, 1).
\]

**Remark 6.** The preceding corollary applies, in particular, when \( \xi \) has the uniform distribution on \( K \) (actually, both (a) and (b) are fulfilled in this case). For the two-dimensional rectangle \([a, b] \times [c, d]\), this result was first obtained by Dragomir [4] by using a different method.

Hyperrectangles are nothing but Cartesian products of 1-simplices. For a more general discussion, we refer to Section 5 below.
3.3. Crosspolytopes

In this subsection, \( K \) denotes an \( n \)-crosspolytope, that is, the convex hull of \( n \) linearly independent line segments in \( \mathbb{R}^n \) whose midpoints coincide. More precisely, \( K \) is the convex hull of the set of \( 2n \) points \( K^* := \{ a_0 \pm a_1, \ldots, a_0 \pm a_n \} \), where \( \{ a_1, \ldots, a_n \} \) is a linear basis for \( \mathbb{R}^n \). The particular case described by (4), to be denoted by \( K_0 \), is just the closed unit ball for the \( l_1 \)-norm \( \| \cdot \|_1 \) on \( \mathbb{R}^n \). We also denote by

\[
\Theta := \left\{ \theta := (\theta_1, \ldots, \theta_n) \in [0, 1]^n : \sum_{i=1}^{n} \theta_i = 1 \right\},
\]

and we set, for \( \theta \in \Theta \) and \( x := (x_1, \ldots, x_n) \in K_0 \),

\[
h_{\theta \pm e_i}^\theta (x) := \frac{|x_i| \pm x_i}{2} + \theta_i \frac{1 - \|x\|_1}{2}, \quad i = 1, \ldots, n. \tag{10}
\]

**Corollary 4.** Fix \( \theta \in \Theta \), and let \( T : \mathbb{R}^n \to \mathbb{R}^n \) be the bijective affine transformation (uniquely) determined by (5). If \( \xi \) is a \( K \)-valued random vector, then the random vector \( \xi^* \) having the probability distribution given by

\[
P(\xi^* = a_0 \pm a_i) := E[h_{\theta \pm e_i}^\theta (T(\xi))], \quad i = 1, \ldots, n, \tag{11}
\]

is an \( H \)-majorant of \( \xi \).

**Proof.** When \( K = K_0 \) (and \( T \) is the identity), the family of \( 2n \) functions defined by (10) is a representation system on \( K_0 \), and Theorem 4 yields the conclusion. Then, the general case follows by Lemma 1. \( \square \)

**Corollary 5.** Let \( \xi := (\xi_1, \ldots, \xi_n) \) be a \( K_0 \)-valued random vector such that \( E\xi = 0 \) and \( E|\xi_1| = \cdots = E|\xi_n| \). Then, the random vector having the uniform distribution on \( K^*_0 \) is an \( H \)-majorant of \( \xi \).

**Proof.** On taking \( \theta = (1/n, \ldots, 1/n) \) in Corollary 4, we obtain \( P(\xi^* = \pm e_i) = 1/2n \) \( i = 1, \ldots, n \). \( \square \)

**Remark 7.** The preceding corollary applies, in particular, when \( \xi \) has the uniform distribution on \( K_0 \).

**Remark 8.** When \( n \geq 2 \), (10) defines a multiparameter infinite family of representation systems on \( K_0 \). In the two-dimensional case, another different representation system on \( K_0 \) is given by

\[
h_{\pm e_1}'(x, y) := \frac{1 \pm (x + y)}{2} \frac{1 \pm (x - y)}{2},
\]

\[
h_{\pm e_2}'(x, y) := \frac{1 \pm (x + y)}{2} \frac{1 \mp (x - y)}{2}, \tag{12}
\]

which is nothing but the representation system \( H' \) supplied by Lemma 1(c), when we take \( K := [-1, 1] \times [-1, 1] \), \( T \) is given by \( T(x, y) := ((x + y)/2, (x - y)/2) \), and \( H \) is the representation system on \( K \) introduced in Section 3.2. Therefore, Theorem 4 (or the combination of Lemma 1 and Corollary 3) yields the following result.
Corollary 6. Let $\xi := (\xi_1, \xi_2)$ be a $K_0$-valued random vector ($n = 2$) fulfilling one of the following two conditions:

(a) $\xi_1 + \xi_2$ and $\xi_1 - \xi_2$ are independent and $\mathbb{E}\xi = 0$.
(b) The random vector $(\xi_1 + \xi_2, \xi_1 - \xi_2)$ is multisymmetric.

Then, the random vector having the uniform distribution on $K_0^*$ is an $H$-majorant of $\xi$.

Remark 9. Analogously, in the two-dimensional case, Lemma 1 together with (10) and the transformation $T^{-1}(x, y) := (x + y, x - y)$ (the inverse of the one given in Remark 8) supply with a (one-parameter) infinite family of representation systems on $[-1, 1] \times [-1, 1]$ different from that considered in the preceding subsection. Details are left to the reader.

4. Closed balls: symmetric random vectors

In the following theorem, $K := \{x \in \mathbb{R}^n : \|x\| \leq 1\}$ is the closed unit ball for a given norm $\| \cdot \|$ on $\mathbb{R}^n$, and $K^* := \{x \in \mathbb{R}^n : \|x\| = 1\}$ is the corresponding unit sphere.

It should be observed that $K^*$ always contains the set $K^*$ of extreme points of $K$. We also have $K^* = K^*$ for many norms, including all the $l_p$-norms with $1 < p < \infty$, but such an equality fails to be true for the $l_1$ and the $l_\infty$-norms, among other ones.

Our main result in this section is stated as follows. We recall that an $n$-dimensional random vector $\xi$ is said to be symmetric if $-\xi$ has the same probability distribution as $\xi$; in such a case, we obviously have $\mathbb{E}\xi = 0$.

Theorem 5. Let $\xi$ be a symmetric $K$-valued random vector, and let $\xi^*$ be the symmetric $K^*$-valued random vector defined by

$$\xi^* := \frac{\xi}{\|\xi\|}1_{\{\xi \neq 0\}} + \eta 1_{\{\xi = 0\}},$$

where $1_A$ stands for the indicator function of the event $A$, and $\eta$ is a symmetric $K^*$-valued random vector independent of $\xi$. Then, we have $\mathbb{E}f(\xi) \leq \mathbb{E}f(\xi^*)$, for all $f \in \mathcal{C}_K$. In particular, when $K^* = K^*$, $\xi^*$ is an $H$-majorant of $\xi$.

Proof. We can write

$$\xi = \frac{1 + \|\xi\|}{2} \xi^* + \frac{1 - \|\xi\|}{2} (-\xi^*),$$

and we have by the convexity of $f \in \mathcal{C}_K$

$$\mathbb{E}f(\xi) \leq \frac{1}{2} \mathbb{E}[\left(1 + \|\xi\|\right) f(\xi^*)] + \frac{1}{2} \mathbb{E}\left[(1 - \|\xi\|) f(-\xi^*)\right]$$

$$= \frac{1}{2} \mathbb{E}f(\xi^*) + \frac{1}{2} \mathbb{E}f(-\xi^*) + \frac{1}{2} \mathbb{E}[\|\xi\| f(\xi^*)] - \frac{1}{2} \mathbb{E}[\|\xi\| f(-\xi^*)].$$

Since the symmetry of $\xi$ (and $\eta$) entails

$$\mathbb{E}f(\xi^*) = \mathbb{E}f(-\xi^*) \quad \text{and} \quad \mathbb{E}[\|\xi\| f(\xi^*)] = \mathbb{E}[\|\xi\| f(-\xi^*)],$$

the conclusion follows. $\square$
Remark 10. In the setting of the preceding theorem, if $P(\xi = 0) \neq 0$, different choices of (the distribution of) $\eta$ give different versions of $\xi^*$. 

Remark 11. Assume that $P(\xi = 0) = 0$. Then, the probability distribution of $\xi^*$ is described by 

$$P(\xi^* \in B_*) = \frac{\text{Vol}(B)}{\text{Vol}(K)} =: \mu(B_*),$$

where $B_*$ is the $\sigma$-field of Borel subsets of $K_*$, and 

$$B := \{ rx : r \in [0, 1], x \in B_* \}$$

is the cone with base $B_*$ and cusp 0. In particular, if $\xi$ has the uniform distribution on $K$, we have 

$$P(\xi^* \in B_*) = \frac{\text{Vol}(B)}{\text{Vol}(K)} \mu(B_*),$$

that is, the distribution of $\xi^*$ is the so called normalized cone measure $\mu$ on $K_*$ (see, for instance, [11]). This measure must not be confused with the normalized surface measure $\sigma$ on $K_*$ given by 

$$\sigma(B_* \in B) = \frac{\text{Area}(B_*)}{\text{Area}(K_*)}, \quad B \in \mathcal{B}_*$$

(\text{where, in the case } n = 1, \text{Area must be understood as counting measure}). Actually, it is well known that, when $n \geq 2$, $\sigma$ is absolutely continuous with respect to $\mu$, and its density is given for almost every $x \in K_*$ by 

$$\frac{d\sigma}{d\mu}(x) = \frac{n \text{Vol}(K)}{\text{Area}(K_*)} \left\| \nabla \left( \| \cdot \| \right) (x) \right\|_2,$$

where $\| \cdot \|_2$ is the (Euclidean) $l_2$-norm (see [11, Lemma 1]). Thus, $\sigma$ coincides with $\mu$ if and only if $\| \nabla (\| \cdot \|)(x) \|_2$ is constant a.e. In the case of $l_p$-norms, this holds true only for $p = 1, 2, \infty$. In particular, for the Euclidean closed unit ball in $\mathbb{R}^n$, the Hermite–Hadamard inequality supplied by Theorem 5 is 

$$f(0) \leq \frac{1}{\text{Vol}(K)} \int_K f(x) \, dx \leq \int_{K^*} f(x) \, d\sigma(x), \quad f \in C_K,$$

which, for $n = 2, 3$, was early obtained by Dragomir [2,3] with arguments based on Calculus.

In the following corollary, $K$ denotes the closed ball of center $a \in \mathbb{R}^n$ and radius $r > 0$, and $K_*$ is the corresponding sphere. It directly follows from Theorem 5 by using Lemma 1 and the affine bijective function $T(x) := a + rx$.

Corollary 7. Let $\xi$ be a $K$-valued random vector such that $\xi - a$ is symmetric, and let $\xi^*$ be the $K_*$-valued random vector defined by 

$$\xi^* := \left( a + \frac{r(\xi - a)}{\| \xi - a \|} \right) 1_{\{ \xi \neq a \}} + \eta 1_{\{ \xi = a \}},$$

where $\eta$ is a $K^*$-valued random vector independent of $\xi$, and such that $\eta - a$ is symmetric. Then, we have $Ef(\xi) \leq Ef(\xi^*)$, for all $f \in C_K$. In particular, when $K_* = K^*$, $\xi^*$ is an $H$-majorant of $\xi$. 


5. Products

For \( i = 1, \ldots, n \), let \( K_i \subseteq \mathbb{R}^{m_i} \) be an \( m_i \)-dimensional compact convex set whose set of extreme points is \( K_i^* \). Then, it is easy to see that the Cartesian product \( K := K_1 \times \cdots \times K_n \) is an \( m \)-dimensional compact convex set \( (m := m_1 + \cdots + m_n) \) whose set of extreme points is \( K^* := K_1^* \times \cdots \times K_n^* \).

We recall that a real function \( f \) on \( K \) is said to be componentwise convex if, for each \( i = 1, \ldots, n \), and for arbitrarily fixed \( x_j \in K_j \) \( (j \neq i) \), the real function on \( K_i \) \( f(x_1, \ldots, x_{i-1}, \ast, x_{i+1}, \ldots, x_n) \) is convex. We obviously have \( C_K \subseteq C_K^* := \{ \text{all real componentwise convex functions on } K \} \).

The following theorem generalizes a result of Dragomir for two-dimensional rectangles [4, Theorem 1]. We omit the proof, which is readily achieved by using induction on \( n \) and Fubini’s theorem (see, also, [10, p. 104]).

**Theorem 6.** Let \( \xi := (\xi_1, \ldots, \xi_n) \) be a \( K \)-valued random vector having independent components, and such that \( \xi_i \) is \( K_i \)-valued \( (i = 1, \ldots, n) \), and let \( \xi^* := (\xi_1^*, \ldots, \xi_n^*) \) a \( K^* \)-valued random vector with independent components, and such that \( \xi_i^* \) in an \( H \)-majorant of \( \xi_i \) \( (i = 1, \ldots, n) \). Then, we have \( f(E\xi) \leq Ef(\xi) \leq Ef(\xi^*) \), for all \( f \in C_K^* \). In particular, \( \xi^* \) is an \( H \)-majorant of \( \xi \).

**Remark 12.** In particular, if \( K_i \) is a polytope and \( \xi_i^* \) is uniform on \( K_i^* \) \( (i = 1, \ldots, n) \), then \( \xi^* \) is uniform on \( K^* \).

In the case of polytopes, we can construct representations systems on the product from representations systems on the factors. The following result is immediately proved by induction on \( n \).

**Lemma 2.** Assume that, for \( i = 1, \ldots, n \), \( K_i \) is a polytope and \( \mathcal{H}_i := \{ h_v: v \in K_i^* \} \) is a representation system on \( K_i \). Let \( \mathcal{H} := \{ h_v: v \in K^* \} \) be the family of functions on \( K \) given by

\[
h_v(x) = \prod_{i=1}^n h_{v_i}(x_i),
\]

where \( v := (v_1, \ldots, v_n) \in K^* \) and \( x := (x_1, \ldots, x_n) \in K \). Then, \( \mathcal{H} \) is a representation system on \( K \).

**Corollary 8.** In the setting of the preceding lemma, if \( \xi := (\xi_1, \ldots, \xi_n) \) is a \( K \)-valued random vector such that \( \xi_i \) is \( K_i \)-valued \( (i = 1, \ldots, n) \), then the \( K^* \)-valued random vector \( \xi^* := (\xi_1^*, \ldots, \xi_n^*) \) with probability distribution given by

\[
P(\xi^* = v) := E[h_v(\xi)], \quad v \in K^*,
\]

is an \( H \)-majorant of \( \xi \).

**Remark 13.** In particular, if the random subvectors \( \xi_1, \ldots, \xi_n \) are independent, then we have, for every \( v := (v_1, \ldots, v_n) \in K^* \),

\[
P(\xi^* = v) = \prod_{i=1}^n E[h_{v_i}(\xi_i)] = \prod_{i=1}^n P(\xi_i^* = v_i),
\]

showing that \( \xi_1^*, \ldots, \xi_n^* \) are independent as well.
6. Related results

The following theorem establishes that \( Ef(\xi) \) can be represented in terms of any H-majorant of \( \xi \).

**Theorem 7.** In the setting of Theorem 2, let \( \xi^* \) be an H-majorant of \( \xi \), and let \( x^* \) be the common barycenter of \( \xi \) and \( \xi^* \). Then, for each \( f \in C_K \) there is some \( t \in [0, 1] \) (depending upon \( f \)) such that

\[
Ef(\xi) = Ef\left(x^* + t(\xi^* - x^*)\right).
\]  

(14)

**Proof.** By the continuity of \( f \) and the dominated convergence theorem, the function \( H \) given by

\[
H(t) := Ef\left(x^* + t(\xi^* - x^*)\right), \quad t \in [0, 1],
\]  

(15)

is continuous. Since \( H(0) \leq Ef(\xi) \leq H(1) \), the conclusion follows. \( \Box \)

**Remark 14.** In particular, when \( K \) is a polytope and \( \xi^* \) is uniform on \( K^* \), Eq. (14) says that \( Ef(\xi) \) is the arithmetic mean of the values of \( f \) at the extreme points of the polytope (homothetic to \( K \)) \( K_t := x^* + t(K - x^*) \). Also, when \( K \) is the Euclidean closed unit ball in \( \mathbb{R}^n \) (and \( \xi \) is uniform on it), (14) becomes

\[
\frac{1}{\text{Vol}(K)} \int_K f(x) \, dx = \int_{K^*} f(tx) \, d\sigma(x),
\]

where \( \sigma \) is the normalized surface measure on the sphere.

The following result gives estimates for the differences between the middle term and the extreme terms in some previous Hermite–Hadamard inequalities, when the function \( f \) is \( M \)-Lipschitz with respect to the norm \( \| \cdot \| \), i.e.,

\[
|f(x) - f(y)| \leq M\|x - y\|, \quad x, y \in K.
\]  

(16)

**Theorem 8.** Let \( K, \xi \) and \( \xi^* \) be the same as in Corollary 7. Then, for every real function \( f \) on \( K \) fulfilling (16), we have

\[
|Ef(\xi) - Ef(\xi^*)| \leq ME\|\xi - \xi^*\| \quad \text{and} \quad |Ef(\xi) - f(a)| \leq ME\|\xi - a\|.
\]

**Proof.** It is clear that

\[
|Ef(\xi) - f(a)| \leq E|f(\xi) - f(a)| \leq ME\|\xi - a\|.
\]

On the other hand, it is immediately checked that \( \|\xi - \xi^*\| = r - \|\xi - a\| \) and, therefore,

\[
|Ef(\xi) - Ef(\xi^*)| \leq E|f(\xi) - f(\xi^*)| \leq ME\|\xi - \xi^*\| \leq M\left(r - E\|\xi - a\|\right),
\]

finishing the proof. \( \Box \)

**Corollary 9.** In the setting of the preceding theorem, assume that \( \xi \) has the uniform distribution on \( K \). Then, for every real function \( f \) on \( K \) fulfilling (16), we have

\[
|Ef(\xi) - Ef(\xi^*)| \leq \frac{Mr}{n+1} \quad \text{and} \quad |Ef(\xi) - f(a)| \leq \frac{Mnr}{n+1}.
\]
Proof. It is readily shown that
\[ P(\|\xi - a\| \leq t) = \frac{r^n}{t^n}, \quad 0 \leq t \leq r, \]
and we therefore have
\[ E\|\xi - a\| = \int_0^r P(\|\xi - a\| > t) \, dt = \frac{nr}{n+1}. \]
Thus, the preceding theorem yields the conclusion. \(\square\)

Remark 15. In particular, when \(K\) is the interval \([a, b]\) (and \(\|\cdot\|\) is the absolute value), both the differences
\[ \left| f\left(\frac{a+b}{2}\right) - \frac{1}{b-a} \int_a^b f(x) \, dx \right| \quad \text{and} \quad \left| \frac{1}{b-a} \int_a^b f(x) \, dx - \frac{f(a) + f(b)}{2} \right| \]
are bounded above by \(M(b - a)/4\), when \(f\) is a real \(M\)-Lipschitz function on \([a, b]\). These estimates were early obtained by Dragomir et al. [5]. As observed by several authors, they can also be derived from the classical Ostrowski’s inequality [14] (see also [1]).

7. Concluding remarks

In terms of measures, Theorem 2 can be restated in the following way: For each finite positive measure \(\mu\) on a (nonempty) compact convex set \(K \subset \mathbb{R}^n\), there is a probability measure \(\mu^*\) on \(K^*\) such that
\[ f(x^*) \leq \frac{1}{\mu(K)} \int_K f \, d\mu \leq \int_{K^*} f \, d\mu^*, \quad f \in C_K, \tag{17} \]
where \(x^*\) is the barycenter of \(\mu\) (and \(\mu^*\)).

Several authors have discussed the problem of extending this result to signed measures. As for the first inequality in (17) (i.e., Jensen’s inequality), it is shown in [12, Theorem 1] the following theorem which generalizes an earlier one-dimensional result of Fink [7, Theorem 1].

Theorem 9. Let \(\mu\) be a finite Borel measure on \(K\), with \(\mu(K) > 0\) and barycenter \(x^* \in K\). Then, the following assertions are equivalent:

(a) The first inequality in (17) holds, for all \(f \in C_K\).
(b) \(\int_K f^+ \, d\mu \geq 0\), for all \(f \in C_K\).

Measures fulfilling such conditions are called Popoviciu measures. The extension of the upper Hadamard inequality seems to be a much more delicate problem. In [12, Theorem 4], it is also asserted that, for each Popoviciu measure on \(K\), there is a probability measure \(\mu^*\) on \(K^*\) such that the upper inequality in (17) holds for all \(f \in C_K\). However, such an assertion fails to be true even in the one-dimensional case: Actually, if \(K := [-1, 1]\), \(\mu := \delta_{-1} - \delta_0 + \delta_1\) (where \(\delta_x\) is Dirac’s delta at \(x\)), and \(f(x) := |x|\), then \(\mu\) is a Popoviciu measure on \(K\), and we have \(\int_K f \, d\mu = 2 > 1 = \int_{K^*} f \, d\mu^*\), for every probability measure \(\mu^*\) on \(K^* = \{-1, 1\}\).
References