J. Math. Anal. Appl. 411 (2014) 119–128

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

www.elsevier.com/locate/jmaa

An approach of majorization in spaces with a curved geometry

Constantin P. Niculescu *, Ionel Rovenţa

University of Craiova, Department of Mathematics, Street A. I. Cuza 13, Craiova 200585, Romania

ARTICLE INFO

Article history:
Received 21 December 2012
Available online 25 September 2013
Submitted by Richard M. Aron

Keywords:
Generalized convex function
Global NPC space
Majorization of measures
Wasserstein distance

ABSTRACT

The Hardy–Littlewood–Pólya majorization theorem is extended to the framework of some spaces with a curved geometry (such as the global NPC spaces and the Wasserstein spaces). We also discuss the connection between our concept of majorization and the subject of Schur convexity. Several applications are included.

1. Introduction

In 1929, G.H. Hardy, J.E. Littlewood and G. Pólya [13,14] have proved an important characterization of convex functions in terms of a partial ordering of vectors $x = (x_1, \dots, x_n)$ in $\mathbb{R}^n$. In order to state it we need a preparation. We denote by $x^{\downarrow}$ the vector with the same entries as x but rearranged in decreasing order,

$$x_1^{\downarrow} \geqslant \cdots \geqslant x_n^{\downarrow}.$$

Then x is *weakly majorized* by y (abbreviated, $x \prec_* y$) if

$$\sum_{i=1}^{k} x_i^{\downarrow} \leqslant \sum_{i=,1}^{k} y_i^{\downarrow} \quad \text{for } k = 1, \dots, n \tag{1}$$

and x is *majorized* by y (abbreviated, $x \prec y$) if in addition

$$\sum_{i=1}^{n} x_i^{\downarrow} = \sum_{i=1}^{n} y_i^{\downarrow}. \tag{2}$$

Intuitively, $x \prec y$ means that the components in x are less spread out than the components in y. As shown in Theorem 1 below, the concept of majorization admits an order-free characterization based on the notion of doubly stochastic matrix. Recall that a matrix $A \in \mathrm{M}_n(\mathbb{R})$ is *doubly stochastic* if it has nonnegative entries and each row and each column sums to unity.

Theorem 1. *(Hardy, Littlewood and Pólya [13, Theorem 8]) Let x and y be two vectors in $\mathbb{R}^n$, whose entries belong to an interval I. Then the following statements are equivalent*:

* Corresponding author.
E-mail addresses: cniculescu47@yahoo.com (C.P. Niculescu), ionelroventa@yahoo.com (I. Rovenţa).

0022-247X/$ – see front matter
http://dx.doi.org/10.1016/j.jmaa.2013.09.038

i) $x \prec y$;
ii) *There is a doubly stochastic matrix* $A = (a_{ij})_{1 \leqslant i,j \leqslant n}$ *such that* $x = Ay$;
iii) *The inequality* $\sum_{i=1}^{n} f(x_i) \leqslant \sum_{i=1}^{n} f(y_i)$ *holds for every continuous convex function* $f : I \to \mathbb{R}$.

The proof of this result is also available in the recent monographs [22] and [25].

Remark 1. M. Tomić [31] and H. Weyl [32] have noticed the following characterization of weak majorization: $x \prec_* y$ if and only if $\sum_{i=1}^{n} f(x_i) \leqslant \sum_{i=1}^{n} f(y_i)$ for every continuous nondecreasing convex function f defined on an interval containing the components of x and y. The reader will find the details in [22, Proposition B2, p. 157].

Nowadays many important applications of majorization to matrix theory, numerical analysis, probability, combinatorics, quantum mechanics etc. are known, see [3,22,25,27,28]. They were made possible by the constant growth of the theory, able to uncover the most diverse situations.

In what follows we will be interested in a simple but basic extension of the concept of majorization as mentioned above: the weighted majorization. Indeed, the entire subject of majorization can be switched from vectors to Borel probability measures by identifying a vector $x = (x_1, \dots, x_n)$ in $\mathbb{R}^n$ with the discrete measure $\frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}$ acting on $\mathbb{R}$. By definition,

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{x_i} \prec \frac{1}{n}\sum_{i=1}^{n} \delta_{y_i}$$

if the conditions (1) and (2) above are fulfilled, and Theorem 1 can be equally seen as a characterization of this instance of majorization.

Choquet's theory made available a very general framework of majorization by allowing the comparison of Borel probability measures whose supports are contained in a compact convex subset of a locally convex separated space. The highlights of this theory are presented in [28] and refer to a concept of majorization based on condition iii) in Theorem 1 above. The particular case of discrete probability measures on the Euclidean space $\mathbb{R}^N$, that admits an alternative approach via condition ii) in the same Theorem 1 is of interest to us. Indeed, in this case one can introduce a relation of the form

$$\sum_{i=1}^{m} \lambda_i \delta_{x_i} \prec \sum_{j=1}^{n} \mu_j \delta_{y_j}, \tag{3}$$

where all coefficients λ_i and μ_j are weights, by asking the existence of an $m \times n$-dimensional matrix $A = (a_{ij})_{i,j}$ such that

$$a_{ij} \geqslant 0, \quad \text{for all } i, j, \tag{4}$$

$$\sum_{j=1}^{n} a_{ij} = 1, \quad i = 1, \dots, m, \tag{5}$$

$$\mu_j = \sum_{i=1}^{m} a_{ij}\lambda_i, \quad j = 1, \dots, n \tag{6}$$

and

$$x_i = \sum_{j=1}^{n} a_{ij} y_j, \quad i = 1, \dots, n. \tag{7}$$

The matrices verifying the conditions (4) and (5) are called *stochastic on rows*. When $m = n$ and all weights λ_i and μ_j are equal, the condition (6) assures the *stochasticity on columns*, so in that case we deal with doubly stochastic matrices.

The fact that (3) implies

$$\sum_{i=1}^{m} \lambda_i f(x_i) \prec \sum_{j=1}^{n} \mu_j f(y_j),$$

for every continuous convex function f defined on a convex set containing all points x_i and y_i, is covered by a general result due to S. Sherman [29]. See also the paper of J. Borcea [7] for a nice proof and important applications.

It is worth noticing that the extended definition of majorization given by (3) is related, via equality (7), to an optimization problem as follows:

$$x_i = \operatorname*{arg\,min}_{z \in \mathbb{R}^N} \frac{1}{2}\sum_{j=1}^{n} a_{ij}\|z - y_j\|^2, \quad \text{for } i = 1, \dots, m.$$

The aim of the present paper is to discuss the analogue of the relation of majorization (3) within certain classes of spaces with curved geometry. We will start with the spaces with global nonpositive curvature (abbreviated, global NPC spaces). The subject of majorization in these spaces was touched in [24] via a different concept of majorization which however does not provide an extension of Theorem 1. This goal is accomplished here via another approach of majorization, inspired by the recent work on least squares mean on Riemannian manifolds, due to Lawson and Lim [19] (and simplified by Bhatia and Karandikar [6]). See Theorem 4. In a recent paper, Lim [21] also obtained an extension of Hardy–Littlewood–Pólya Theorem, using a different argument. However, our extension is more general, allowing the majorization between discrete measures with supports of different cardinality (as in the case of Sherman's aforementioned result). As an application we are able to derive a number of new inequalities involving elements of suitable global NPC spaces and to recover some well known results (such as the matrix form of the arithmetic mean–geometric mean inequality) as well as some of the results proved by Lim.

The important consequence of majorization, Schur's convexity, also works in the context of global NPC spaces. Our Theorem 5 asserts that if $\frac{1}{n}\sum_{i=1}^{n}\delta_{x_i} \prec \frac{1}{n}\sum_{i=1}^{n}\delta_{y_i}$ in a global NPC space M, and $f: M^n \to \mathbb{R}$ is a continuous convex function invariant under the permutation of coordinates, then

$$f(x_1,\dots,x_n) \leqslant f(y_1,\dots,y_n).$$

The argument of Theorem 5 yields a partial generalization of Rado's geometric characterization of majorization: if $\frac{1}{n}\sum_{i=1}^{n}\delta_{x_i} \prec \frac{1}{n}\sum_{i=1}^{n}\delta_{y_i}$ then $(x_1,\dots,x_n)$ is a convex combination of the $n!$ points $(y_{\sigma(1)},\dots,y_{\sigma(n)})$, obtained by permuting the components of $(y_1,\dots,y_n)$. The converse implication also works and this is proved in the paper of Lim [21].

The paper ends with a discussion concerning the case of Wasserstein space $\mathcal{P}_2(\mathbb{R}^N)$, of all Borel probability measures on $\mathbb{R}^N$ having finite second moments. This is a complete metric space when endowed with the Wasserstein metric,

$$\mathcal{W}_2(\mu,\nu) = \inf\Big(\int_{\mathbb{R}^N\times\mathbb{R}^N} \|x-y\|^2\,d\gamma(x,y)\Big)^{1/2},$$

the infimum being taken over all Borel probability measures γ on $\mathbb{R}^N \times \mathbb{R}^N$ with marginals μ and ν. Despite the fact that the Wasserstein space is not a global NPC space, it has nice features that assure the validity of the Hardy–Littlewood–Pólya Theorem and of Schur convexity in its context.

A previous version of our paper has been circulated as a preprint posted on arXiv [26].

2. Global NPC spaces

Definition 1. A global NPC space is a complete metric space $M = (M,d)$ for which the following inequality holds true: for each pair of points $x_0, x_1 \in M$ there exists a point $y \in M$ such that for all points $z \in M$,

$$d^2(z,y) \leqslant \frac{1}{2}d^2(z,x_0) + \frac{1}{2}d^2(z,x_1) - \frac{1}{4}d^2(x_0,x_1). \tag{8}$$

These spaces are also known as CAT(0) spaces or Hadamard spaces. See respectively [9] and [30]. In a global NPC space, each pair of points $x_0, x_1 \in M$ can be connected by a geodesic (that is, by a rectifiable curve $\gamma: [0,1] \to M$ such that the length of $\gamma|_{[s,t]}$ is $d(\gamma(s),\gamma(t))$ for all $0 \leqslant s \leqslant t \leqslant 1$). Moreover, this geodesic is unique.

In a global NPC space, the geodesics play the role of segments. The point y that appears in Definition 1 is the *midpoint* of x_0 and x_1 and has the property

$$d(x_0,y) = d(y,x_1) = \frac{1}{2}d(x_0,x_1).$$

We will denote it as $\frac{1}{2}x_0 \boxplus \frac{1}{2}x_1$. It is worth noticing that

$$\frac{1}{2}x_0 \boxplus \frac{1}{2}x_1 = \arg\min_{z\in M} \frac{1}{2}\big[d^2(x_0,z) + d^2(x_1,z)\big].$$

See [5, Proposition 6.2.8]. All other convex combinations $(1-\lambda)x_0 \boxplus \lambda x_1$ of x_0 and x_1 can be introduced in the same manner. An important role is played by the inequality (8) (in technical terms, the uniform convexity of the square distance).

Every Hilbert space is a global NPC space. Its geodesics are the line segments and $\frac{1}{2}x_0 \boxplus \frac{1}{2}x_1 = \frac{x_0+x_1}{2}$.

A more sophisticated example is provided by the upper half-plane

$$\mathbf{H} = \{z \in \mathbb{C}:\ \operatorname{Im} z > 0\},$$

when endowed with the Poincaré metric,

$$ds^2 = \frac{dx^2+dy^2}{y^2}.$$

In this case the geodesics are the semicircles in **H** perpendicular to the real axis and the straight vertical lines ending on the real axis.

The space $\mathrm{Sym}^{++}(n,\mathbb{R})$, of all positively definite matrices with real coefficients becomes a global NPC space when endowed with the *trace metric*,

$$d_{\mathrm{trace}}(A,B)=\left(\sum_{k=1}^{n}\log^2\lambda_k\right)^{1/2},$$

where $\lambda_1,\dots,\lambda_n$ are the eigenvalues of AB^{-1}. In this case the curve

$$\gamma:[0,1]\to \mathrm{Sym}^{++}(n,\mathbb{R}),\quad \gamma_{AB}(t)=A^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^tA^{1/2}$$

is the unique minimal geodesic (up to parametrization) joining A and B (so that the midpoint of A and B is in this case $\frac{1}{2}A\boxplus\frac{1}{2}B=A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2}A^{1/2}$).

A natural generalization of the last example in the setting of separable Hilbert spaces has recently been worked out by Larotonda [17]. If H is such a space, we denote by $HS(H)$ the bilateral ideal of Hilbert–Schmidt operators of $L(H,H)$. The space $HS(H)$ is a Banach algebra (without unit) with respect to the Hilbertian norm $\|A\|_2=\mathrm{trace}^{1/2}(A^*A)$. The positive part,

$$\widetilde{HS(H)}^{++}=\left\{A+\lambda I>0\colon\ A^*=A,\ A\in HS(H),\ \lambda\in\mathbb{R}\right\},$$

of the extended Hilbert–Schmidt algebra $\widetilde{HS(H)}=HS(H)+\mathbb{R}I$ (the algebra obtained by adjoining the identity), constitutes a global NPC space with respect to the distance

$$d(A,B)=\left\|\log\left(A^{1/2}B^{-1}A^{1/2}\right)\right\|_2.$$

Both spaces $\mathrm{Sym}^{++}(n,\mathbb{R})$ and $\widetilde{HS(H)}^{++}$ are Riemannian manifolds. See respectively [18] and [20]. In general, a Riemannian manifold is a global NPC space if and only if it is complete, simply connected and of nonpositive sectional curvature. Besides manifolds, other important examples of global NPC spaces are the Bruhat–Tits buildings (in particular, the trees). See [9].

The direct product of metric spaces $M_i=(M_i,d_i)$ $(i=1,\dots,n)$ is the metric space $M=(M,d_M)$ defined by $M=\prod_{i=1}^n M_i$ and

$$d_M(x,y)=\left(\sum_{i=1}^{n}d_i(x_i,y_i)^2\right)^{1/2}.$$

It is a global NPC space if all factors are global NPC spaces.

More information on global NPC spaces is available in [2,16,30].

Definition 2. A set $C\subset M$ is called *convex* if $\gamma([0,1])\subset C$ for each geodesic $\gamma:[0,1]\to M$ joining $\gamma(0),\gamma(1)\in C$.

A function $\varphi:C\to\mathbb{R}$ is called convex if C is a convex set and for each geodesic $\gamma:[0,1]\to C$ the composition $\varphi\circ\gamma$ is a convex function in the usual sense, that is,

$$\varphi\left(\gamma(t)\right)\leqslant(1-t)\varphi\left(\gamma(0)\right)+t\varphi\left(\gamma(1)\right)$$

for all $t\in[0,1]$.

The function φ is called concave if $-\varphi$ is convex.

The distance function on a global NPC space $M=(M,d)$ verifies not only the inequality (8), but also the following stronger version of it,

$$d^2(z,x_t)\leqslant(1-t)d^2(z,x_0)+td^2(z,x_1)-t(1-t)d^2(x_0,x_1);$$

here z is any point in C and x_t is any point on the geodesic γ joining $x_0,x_1\in C$. In terms of Definition 2, this shows that all the functions $d^2(\cdot,z)$ are uniformly convex. In particular, they are convex and the balls are convex sets.

In a global NPC space M the distance function d is convex on $M\times M$, while the functions $d^\alpha(\cdot,z)$, with $\alpha\geqslant1$, are convex on M. See [30, Corollary 2.5], for details.

A stronger condition than convexity is log-convexity, that means that the logarithm of the given function is convex. Due to the arithmetic mean–geometric mean inequality, any such function is necessarily convex.

Some examples of log-convex functions defined on the global NPC space of positively defined matrices are mentioned below:

Theorem 2. *Every positive linear functional on $\mathrm{M}_n(\mathbb{R})$ induces a log-convex function on $\mathrm{Sym}^{++}(n,\mathbb{R})$ in the sense of Definition 2.*

In particular, the trace is a log-convex function.

Proof. It suffices to consider the case of norm-1 positive linear functionals. The set of all such functionals constitutes a weak-star compact convex K, whose extreme points are the functionals $\omega_u : A \to \langle Au, u \rangle$, associated to the unit vectors $a \in \mathbb{R}^n$. See [8, Theorem 2.3.15, p. 53]. According to the Krein–Milman theorem, every $\omega \in K$ is the pointwise limit of a net consisting of pure states. Thus the proof of Theorem 1 reduces to the case of pure states, which is covered by Theorem 1 in [33]. □

The determinant function is also log-convex due to the identity

$$\det A^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^t A^{1/2} = (\det A)^{1-t}(\det B)^t.$$

Theorem 2 also works in the case of Larotonda's space $\widetilde{HS(H)}^{++}$, but in this case the convexity of the trace function should be understood in the context of functions taking values in $(-\infty, \infty]$ (when the inequality of convexity in Definition 2 is assumed only for $t \in (0, 1)$).

Theorem 2 makes possible the extension of the Legendre duality to the case of functions defined on certain global NPC spaces such as $\mathrm{Sym}^{++}(n, \mathbb{R})$ and $\widetilde{HS(H)}^{++}$. Though this is outside the scope of our paper we mention here the definition of the conjugate of a function $f : \mathrm{Sym}^{++}(n, \mathbb{R}) \to \mathbb{R}$:

$$f^*(A) = \sup_{B \in \mathrm{Sym}^{++}(n,\mathbb{R})} \left[\mathrm{trace}(AB) - f(B)\right].$$

Notice that the effective domain of f^*,

$$\mathrm{dom}\, f^* = \left\{A\colon\ f^*(A) < \infty\right\},$$

is a convex set and when this set is nonempty, then f^* is a lower semicontinuous convex function on it. The function $\frac{1}{2}\,\mathrm{trace}(A^2)$ is an example of self-conjugated function.

More examples of convex functions defined on the global NPC space $\mathrm{Sym}^{++}(n, \mathbb{R})$ can be obtained by taking into account the following two simple remarks:

- If $f : \mathrm{Sym}^{++}(n, \mathbb{R}) \to \mathbb{R}$ is convex and $g : \mathbb{R} \to \mathbb{R}$ is increasing and convex, then $g \circ f$ is convex;
- If $\Psi : \mathrm{M}_n(\mathbb{R}) \to \mathrm{M}_n(\mathbb{R})$ is a strictly positive linear map and f is increasing and convex, then $f \circ \Psi|_{\mathrm{Sym}^{++}(n,\mathbb{R})}$ is convex.

The concept of a convex/concave function defined on a convex subset of a global NPC space can be extended verbatim to the case of vector-valued functions taking values in a regular ordered Banach space E. See [11] for a short account on these spaces. A nontrivial example is the embedding of $\mathrm{Sym}^{++}(n, \mathbb{R})$ into the regular ordered Banach space $\mathrm{Re}\,\mathrm{M}_n(\mathbb{R})$ of self-adjoint matrices, endowed with the Hilbertian structure associated to the trace norm and the natural ordering,

$$A \leqslant B \quad \Leftrightarrow \quad \langle Ax, x\rangle \leqslant \langle Bx, x\rangle \quad \text{for every } x.$$

The convexity of this function proves to be equivalent to the matrix form of the arithmetic mean–geometric mean inequality. A proof of this inequality is given in the next section.

Most of our results extend easily to this context, by taking into account that a vector-valued function $f : C \to E$ is convex if, and only if, the composition $x' \circ f$ with every positive linear functional $x' \in E'$ is a convex function. Indeed, an element $x \in E$ is in the positive cone if, and only if, $x'(x) \geqslant 0$ for every positive linear functional $x' \in E'$.

3. The extension of Hardy–Littlewood–Pólya Theorem

When $x_1, \dots, x_m, y_1, \dots, y_n$ are points of a global NPC space M, and $\lambda_1, \dots, \lambda_m$ in $[0, 1]$ are weights that sum to 1, we will define the relation of majorization

$$\sum_{i=1}^{m} \lambda_i \delta_{x_i} \prec \sum_{j=1}^{n} \mu_j \delta_{y_j} \tag{9}$$

by asking the existence of an $m \times n$-dimensional matrix $A = (a_{ij})_{i,j}$ that is stochastic on rows and verifies the following two conditions:

$$\mu_j = \sum_{i=1}^{m} a_{ij}\lambda_i, \quad j = 1, \dots, n \tag{10}$$

and

$$x_i = \arg\min_{z \in M} \frac{1}{2}\sum_{j=1}^{n} a_{ij} d^2(z, y_j), \quad i = 1, \dots, m. \tag{11}$$

The existence and uniqueness of the optimization problems (11) is assured by the fact that the objective functions are uniformly convex and positive. See [16, Section 3.1], or [30, Proposition 1.7, p. 3]. In the case where $M = \mathrm{Sym}^{++}(n, \mathbb{R})$, the points x_i are the unique solutions of the *Karcher equations*,

$$\sum_{j=1}^{n} a_{ij} \log\bigl(x^{1/2} y_j^{-1} x^{1/2}\bigr) = 0.$$

This follows by computing the differentials of the objective functions in the optimization problems (11) and by adapting the argument of Theorem 6.3.4 in [5], p. 219, to the presence of weights.

Notice that the above definition agrees with the usual one in the Euclidean case. It is also related to the definition of the barycenter of a Borel probability measure μ defined on a global NPC space M. Precisely, if $\mu \in \mathcal{P}_2(M)$ (the set of those probability measures under which all functions $d^2(\cdot, z)$ are integrable), then its barycenter is defined by the formula

$$\mathrm{bar}(\mu) = \arg\min_{z\in M} \frac{1}{2} \int_M d^2(z, x)\, d\mu(x).$$

This definition, due to E. Cartan [10], is inspired by Gauss's Least Squares Method. A detailed approach of the notion of barycenter is offered by the paper of Sturm [30].

The particular case of discrete probability measures $\lambda = \sum_{i=1}^{n} \lambda_i \delta_{x_i}$ is of special interest because the barycenter of λ provides the true analogue of the arithmetic weighted mean $\lambda_1 x_1 + \cdots + \lambda_n x_n$ in the context of global NPC spaces. Indeed,

$$\mathrm{bar}(\lambda) = \arg\min_{z\in M} \frac{1}{2} \sum_{i=1}^{n} \lambda_i d^2(z, x_i),$$

and the way $\mathrm{bar}(\lambda)$ provides a mean with nice features was recently clarified by Lawson and Lim [19]. As an immediate consequence one obtains the relation

$$\delta_{\mathrm{bar}(\lambda)} \prec \lambda.$$

We will denote $\mathrm{bar}(\lambda)$ as $\lambda_1 x_1 \boxplus \cdots \boxplus \lambda_n x_n$ in order to outline its special meaning and to avoid any confusion with $\lambda_1 x_1 + \cdots + \lambda_n x_n$, when the later makes sense. For example, in the case of the global NPC space $\mathrm{Sym}^{++}(n, \mathbb{R})$, endowed with the trace metric,

$$\frac{1}{2} A \boxplus \frac{1}{2} B = A^{1/2}\bigl(A^{-1/2} B A^{-1/2}\bigr)^{1/2} A^{1/2} \leqslant \frac{1}{2} A + \frac{1}{2} B,$$

a fact which illustrates the noncommutative analogue of the arithmetic mean–geometric mean inequality. See [5, Section 6.3], or [18], for details.

Our extension of the Hardy–Littlewood–Pólya Theorem of majorization is based on Jensen's inequality. In the case of flat spaces the discrete form of this inequality follows immediately from the property of associativity of convex combinations:

$$\sum_{i=1}^{n+1} \lambda_i x_i = (1 - \lambda_{n+1}) \left(\sum_{i=1}^{n} \frac{\lambda_i}{1 - \lambda_{n+1}} x_i \right) + \lambda_{n+1} x_{n+1}.$$

Since this doesn't work in the general context of global NPC spaces, the generalization of Jensen's inequality needed new ideas and was done by J. Jost [15]. We recall it here in the formulation of Eells and Fuglede [12, Proposition 12.3, p. 242]:

Theorem 3 *(Jensen's Inequality). For any lower semicontinuous convex function $f : M \to \mathbb{R}$ and any Borel probability measure $\mu \in \mathcal{P}_2(M)$ we have the inequality*

$$f\bigl(\mathrm{bar}(\mu)\bigr) \leqslant \int_M f(x)\, d\mu(x),$$

provided the right hand side is well-defined.

The proof of Eells and Fuglede is based on the following remark concerning barycenters: If a probability measure μ is supported by a convex closed set K, then its barycenter $\mathrm{bar}(q)$ lies in K. A probabilistic approach of Theorem 3 is due to Sturm [30].

An immediate consequence of Theorem 3 is the weighted form of the matrix form of the arithmetic mean–geometric mean inequality,

$$\lambda_1 A_1 \boxplus \cdots \boxplus \lambda_n A_n \leqslant \lambda_1 A_1 + \cdots + \lambda_n A_n,$$

which follows from the convexity of the functions $A \to \langle Ax, x\rangle$, for $A \in \mathrm{Sym}^{++}(n, \mathbb{R})$ and $x \in \mathbb{R}^n$. See [33], for another short proof of this inequality.

Another consequence of Theorem 3 is the following couple of inequalities that work for any points $z, x_1, \dots, x_n, y_1, \dots, y_n$ in a global NPC space:

$$d^2\left(\frac{1}{n}x_1 \boxplus \cdots \boxplus \frac{1}{n}x_n, z\right) \leqslant \frac{d^2(x_1, z) + \cdots + d^2(x_n, z)}{n}$$

and

$$d\left(\frac{1}{n}x_1 \boxplus \cdots \boxplus \frac{1}{n}x_n, \frac{1}{n}y_1 \boxplus \cdots \boxplus \frac{1}{n}y_n\right) \leqslant \frac{d(x_1, y_1) + \cdots + d(x_n, y_n)}{n}.$$

The next theorem offers a partial extension of Hardy–Littlewood–Pólya Theorem to the context of global NPC spaces.

Theorem 4. *If*

$$\sum_{i=1}^{m} \lambda_i \delta_{x_i} \prec \sum_{j=1}^{n} \mu_j \delta_{y_j}$$

in the global NPC space M, then, for every continuous convex function f defined on a convex subset $U \subset M$ containing all points x_i and y_j we have

$$\sum_{i=1}^{m} \lambda_i f(x_i) \leqslant \sum_{j=1}^{n} \mu_j f(y_j).$$

Proof. By our hypothesis, there is an $m \times n$-dimensional matrix $A = (a_{ij})_{i,j}$ that is stochastic on rows and verifies the conditions (10) and (11). The last condition, shows that each point x_i is the barycenter of the probability measure $\sum_{j=1}^{n} a_{ij}\delta_{y_j}$, so by Jensen's inequality we infer that

$$f(x_i) \leqslant \sum_{j=1}^{n} a_{ij} f(y_j).$$

Multiplying each side by λ_i and then summing up over i from 1 to m, we conclude that

$$\sum_{i=1}^{m} \lambda_i f(x_i) \leqslant \sum_{i=1}^{m} \left(\lambda_i \sum_{j=1}^{n} a_{ij} f(y_j)\right) = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} a_{ij}\lambda_i\right) f(y_j) = \sum_{j=1}^{n} \mu_j f(y_j). \qquad \square$$

In a global NPC space the distance function from a convex set is a convex function. See [30, Corollary 2.5]. Combining this fact with Theorem 4 we infer the following result.

Corollary 1. *If*

$$\sum_{i=1}^{m} \lambda_i \delta_{x_i} \prec \sum_{j=1}^{n} \mu_j \delta_{y_j},$$

and all coefficients λ_i are strictly positive, then $\{x_1, \dots, x_m\}$ is contained in the convex hull of $\{y_1, \dots, y_n\}$.

In particular, the points x_i spread out less than the points y_j.

Another application of Theorem 4 yields a new set of inequalities verified by the functions $d(\cdot, z)$ in a global NPC space M. These functions are convex and the same is true for the functions $f(d(\cdot, z))$ whenever f is a continuous nondecreasing convex function defined on $\mathbb{R}_+$. According to Theorem 4, if $\frac{1}{n}\sum_{i=1}^{n} \delta_{x_i} \prec \frac{1}{n}\sum_{i=1}^{n} \delta_{y_i}$ in M, then $\sum_{i=1}^{n} f(d(x_i, z)) \leqslant \sum_{i=1}^{n} f(d(y_i, z))$. Taking into account Remark 1 we arrive at the following result:

Corollary 2. *If $\frac{1}{n}\sum_{i=1}^{n} \delta_{x_i} \prec \frac{1}{n}\sum_{i=1}^{n} \delta_{y_i}$ in $M = (M, d)$, then for all $z \in M$,*

$$\big(d(x_1, z), \dots, d(x_n, z)\big) \prec_* \big(d(y_1, z), \dots, d(y_n, z)\big).$$

According to a result due to Ando (see [22, Theorem B.3a, p. 158]), the converse of Corollary 2 works at least when $M = \mathbb{R}$.

The entropy function,

$$H(t) = -t \log t,$$

is concave and decreasing for $t \in [1/e, \infty)$, so by Corollary 2 we infer that

$$\prod_{i=1}^{n} d(x_i, z)^{d(x_i,z)} \geqslant \prod\nolimits_{i=1}^{n} d(y_i, z)^{d(y_i,z)},$$

whenever $\frac{1}{n}\sum_{i=1}^{n}\delta_{x_i} \prec \frac{1}{n}\sum_{j=1}^{n}\delta_{y_j}$ and all the points x_i and y_i are at a distance $\geqslant 1/e$ from z.

Many other inequalities involving distances in a global NPC space can be derived from Corollary 2 and the following result due to Fan and Mirsky (detailed in [22, Proposition B6, p. 160]) proves very useful in this respect: if $x, y \in \mathbb{R}_+^n$, then $x \prec_* y$ if, and only if,

$$\Phi(x) \leqslant \Phi(y)$$

for all *symmetric gauges* Φ on $\mathbb{R}^n$. These are the functions $\Phi : \mathbb{R}^n \to \mathbb{R}$ such that:

(1) $\Phi(x) > 0$ when $x \neq 0$;
(2) $\Phi(\alpha x) = |\alpha|\Phi(x)$ for all real α;
(3) $\Phi(x + y) \leqslant \Phi(x) + \Phi(y)$;
(4) $\Phi(x_1, \ldots, x_n) = \Phi(\varepsilon_1 x_{\pi(1)}, \ldots, \varepsilon_n x_{\pi(n)})$ whenever each ε_i belongs to $\{-1, 1\}$ and π is any permutation of $\{1, \ldots, n\}$.

Every symmetric gauge Φ induces a distance d_Φ on the space $\mathrm{Sym}^{++}(n, \mathbb{R})$ given by the formula

$$d_\Phi(A, B) = \Phi\big(\lambda_1(A, B), \ldots, \lambda_n(A, B)\big),$$

where $\lambda_1(A, B), \ldots, \lambda_n(A, B)$ are the eigenvalues of the matrix $\log(A^{-1/2}BA^{-1/2})$. This distance makes $\mathrm{Sym}^{++}(n, \mathbb{R})$ a global NPC space, a fact that was first noticed by Bhatia [4]. According to Theorem 4, if $\frac{1}{n}\sum_{i=1}^{n}\delta_{A_i} \prec \frac{1}{n}\sum_{i=1}^{n}\delta_{B_i}$ in $\mathrm{Sym}^{++}(n, \mathbb{R})$, then for all $C \in M$,

$$\big(d_\Phi(A_1, C), \ldots, d_\Phi(A_n, C)\big) \prec_* \big(d_\Phi(B_1, C), \ldots, d_\Phi(B_n, C)\big),$$

which allows us to retrieve a result due to Lim. See [21, Corollary 7.6].

4. The connection with Schur convexity

It is worth noticing the connection between our definition of majorization and the subject of Schur convexity (as presented in [22]):

Theorem 5. *Suppose that $\frac{1}{n}\sum_{i=1}^{n}\delta_{x_i} \prec \frac{1}{n}\sum_{i=1}^{n}\delta_{y_i}$ in the global NPC space M, and $f : M^n \to \mathbb{R}$ is a continuous convex function invariant under the permutation of coordinates. Then*

$$f(x_1, \ldots, x_n) \leqslant f(y_1, \ldots, y_n).$$

Proof. For the sake of simplicity we will restrict here to the case where $n = 3$. The proof of the general case is similar.

According to the definition of majorization, if $\frac{1}{3}\sum_{i=1}^{3}\delta_{x_i} \prec \frac{1}{3}\sum_{i=1}^{3}\delta_{y_i}$, then there exists a doubly stochastic matrix $A = (a_{ij})_{i,j=1}^{3}$ such that

$$x_i = \mathrm{bar}\left(\sum_{j=1}^{3} a_{ij}\delta_{y_j}\right) \quad \text{for } i = 1, \ldots, 3.$$

As A can be uniquely represented in the form

$$A = \begin{pmatrix} \lambda_1+\lambda_2 & \lambda_3+\lambda_5 & \lambda_4+\lambda_6 \\ \lambda_3+\lambda_4 & \lambda_1+\lambda_6 & \lambda_2+\lambda_5 \\ \lambda_5+\lambda_6 & \lambda_2+\lambda_4 & \lambda_1+\lambda_3 \end{pmatrix},$$

where all λ_k are nonnegative and $\sum_{k=1}^{6}\lambda_k = 1$ (a simple matter of linear algebra) we can represent the elements x_j as

$$x_1 = \mathrm{bar}\big((\lambda_1+\lambda_2)\delta_{y_1} + (\lambda_3+\lambda_4)\delta_{y_2} + (\lambda_5+\lambda_6)\delta_{y_3}\big),$$
$$x_2 = \mathrm{bar}\big((\lambda_3+\lambda_5)\delta_{y_1} + (\lambda_1+\lambda_6)\delta_{y_2} + (\lambda_2+\lambda_4)\delta_{y_3}\big),$$
$$x_3 = \mathrm{bar}\big((\lambda_4+\lambda_6)\delta_{y_1} + (\lambda_2+\lambda_5)\delta_{y_2} + (\lambda_1+\lambda_3)\delta_{y_3}\big).$$

It is easy to see that (x_1, x_2, x_3) is the barycenter of

$$\mu = \lambda_1\delta_{(y_1,y_2,y_3)} + \lambda_2\delta_{(y_1,y_3,y_2)} + \lambda_3\delta_{(y_2,y_1,y_3)} + \lambda_4\delta_{(y_2,y_3,y_1)} + \lambda_5\delta_{(y_3,y_1,y_2)} + \lambda_6\delta_{(y_3,y_2,y_1)},$$

so by Jensen's inequality and the symmetry of f we get

$$\begin{aligned} f(x_1,\ldots,x_n) &\leqslant \lambda_1 f(y_1,y_2,y_3)+\lambda_2 f(y_1,y_3,y_2)+\lambda_3 f(y_2,y_1,y_3) \\ &\quad +\lambda_4 f(y_2,y_3,y_1)+\lambda_5 f(y_3,y_1,y_2)+\lambda_6 f(y_3,y_2,y_1) \\ &=(\lambda_1+\cdots+\lambda_6) f(y_1,y_2,y_3)=f(y_1,y_2,y_3). \qquad \square \end{aligned}$$

The following consequence of Theorem 5 relates the majorization of measures to the dispersion of their supports.

Corollary 3. *If $\frac{1}{n}\sum_{i=1}^n \delta_{x_i} \prec \frac{1}{n}\sum_{i=1}^n \delta_{y_i}$ in the global NPC space $M=(M,d)$, then*

$$\sum_{1\leqslant i<j\leqslant n} d^\alpha(x_i,x_j) \leqslant \sum_{1\leqslant i<j\leqslant n} d^\alpha(y_i,y_j),$$

for every $\alpha \geqslant 1$.

Rado's geometric characterization of majorization in $\mathbb{R}^n$ asserts that $(x_1,\ldots,x_n) \prec (y_1,\ldots,y_n)$ in $\mathbb{R}^n$ if, and only if, $(x_1,\ldots,x_n)$ lies in the convex hull of the $n!$ points $(y_{\sigma(1)},\ldots,y_{\sigma(n)})$, obtained by permuting the components of $(y_1,\ldots,y_n)$ (that is, $(x_1,\ldots,x_n)$ is a convex combination of these $n!$ points). See [22, Corollary B.3, p. 34]. A relation of majorization of this kind can be introduced in the power space M^n (of any global NPC space $M=(M,d)$ as well as of $\mathcal{P}_2(\mathbb{R}^N)$) by putting

$$(x_1,\ldots,x_n) \prec (y_1,\ldots,y_n)$$

if $\frac{1}{n}\sum_{i=1}^n \delta_{x_i} \prec \frac{1}{n}\sum_{i=1}^n \delta_{y_i}$. The proof of Theorem 5 yields immediately the necessity part of Rado's characterization: if $(x_1,\ldots,x_n) \prec (y_1,\ldots,y_n)$ in M^n, then $(x_1,\ldots,x_n)$ is a convex combination of the $n!$ points $(y_{\sigma(1)},\ldots,y_{\sigma(n)})$. During the revision of the initial submission of our paper we learned about the paper of Y. Lim [21], that contains the proof of the full extension of Rado's characterization: $(x_1,\ldots,x_n) \prec (y_1,\ldots,y_n)$ in M^n if, and only if, $(x_1,\ldots,x_n)$ is a convex combination of the points $(y_{\sigma(1)},\ldots,y_{\sigma(n)})$.

5. The case of Wasserstein space

The reader has probably noticed that the occurrence of the following two facts is essential for the theory of majorization presented above:

- the existence of a unique minimizer for the functionals of the form

$$J(x)=\frac{1}{2}\sum_{i=1}^m \lambda_i d^2(x,x_i)$$

(thought of as the barycenter $\mathrm{bar}(\lambda)$ of the discrete probability measure $\lambda=\sum_{i=1}^m \lambda_i\delta_{x_i}$);
- Jensen's inequality: for every convex function $f: C \to \mathbb{R}$ and every discrete probability measure $\lambda=\sum_{i=1}^m \lambda_i\delta_{x_i}$ supported by C,

$$f\big(\mathrm{bar}(\lambda)\big) \leqslant \int_C f\,d\lambda=\sum_{i=1}^m \lambda_i f(x_i).$$

The recent paper of Agueh and Carlier [1] shows that such a framework is available also in the case of certain Borel probability measures, equipped with the Wasserstein metric. More precisely, they consider the space $\mathcal{P}_2(\mathbb{R}^N)$ (of all Borel probability measures on $\mathbb{R}^N$ having finite second moments), endowed with the Wasserstein metric,

$$\mathcal{W}_2(\mu,\nu)=\inf\left(\int_{\mathbb{R}^N\times\mathbb{R}^N} \|x-y\|^2\,d\gamma(x,y)\right)^{1/2},$$

where the infimum is taken over all Borel probability measures γ on $\mathbb{R}^N\times\mathbb{R}^N$ with marginals μ and ν.

The barycenter $\mathrm{bar}(\sum_{i=1}^m \lambda_i\delta_{\nu_i})$, of a discrete probability measure $\sum_{i=1}^m \lambda_i\delta_{\nu_i}$, is defined as the minimizer of the functional

$$J(\nu)=\frac{1}{2}\sum_{i=1}^m \lambda_i \mathcal{W}_2^2(\nu_i,\nu).$$

This minimizer is unique when at least one of the measures ν_i vanishes on every Borel set of Hausdorff dimension $N-1$. See [1, Proposition 2.2 and Proposition 3.5].

The natural class of convex function on the Wasserstein space is that of functions convex along barycenters. According to [1, Definition 7.1], a function $\mathcal{F}\colon \mathcal{P}_2(\mathbb{R}^N) \to \mathbb{R}$ is said to be *convex along barycenters* if for any discrete probability measure $\sum_{i=1}^m \lambda_i \delta_{\nu_i}$ on $\mathcal{P}_2(\mathbb{R}^N)$ we have

$$\mathcal{F}\left(\operatorname{bar}\left(\sum_{i=1}^m \lambda_i \delta_{\nu_i}\right)\right) \leqslant \sum_{i=1}^m \lambda_i \mathcal{F}(\nu_i).$$

This notion of convexity coincides with the notion of displacement convexity introduced by McCann [23] if $N = 1$, and is stronger than this in the general case. However, the main examples of displacement convex functions (such as the internal energy, the potential energy and the interaction energy) are also examples of functions convex along barycenters. See [1, Proposition 7.7].

Theorem 6. *The concept of majorization and all results noticed in the case of global NPC spaces (in particular, Theorem 4 and Theorem 5) remain valid in the context of discrete probability measures on $\mathcal{P}_2(\mathbb{R}^N)$ having unique barycenters and the functions $\mathcal{F}$: $\mathcal{P}_2(\mathbb{R}^N) \to \mathbb{R}$ convex along barycenters.*

Rado's geometric characterization of majorization in $\mathbb{R}^n$ extends also to the case of Wasserstein space, the argument being similar to that provided by Lim [21, Theorem 7.1], in the case of global NPC spaces.

Acknowledgments

The authors thank the referees for their comments and suggestions. This work was supported by a grant of the Romanian National Authority for Scientific Research, CNCS–UEFISCDI, project number PN-II-ID-PCE-2011-3-0257.

References

[1] M. Agueh, G. Carlier, Barycenters in the Wasserstein space, SIAM J. Math. Anal. 43 (2) (2011) 904–924.
[2] W. Ballmann, Lectures on Spaces with Nonpositive Curvature, DMV Seminar, vol. 25, Birkhäuser, Basel, 2005.
[3] R. Bhatia, Matrix Analysis, Springer-Verlag, 1997.
[4] R. Bhatia, On the exponential metric increasing property, Linear Algebra Appl. 375 (2003) 211–220.
[5] R. Bhatia, Positive Definite Matrices, Princeton University Press, 2007.
[6] R. Bhatia, R.L. Karandikar, Monotonicity of the matrix geometric mean, Math. Ann. 353 (2012) 1453–1467.
[7] J. Borcea, Equilibrium points of logarithmic potentials induced by positive charge distributions. I. Generalized de Bruijn–Springer relations, Trans. Amer. Math. Soc. 359 (2007) 3209–3237.
[8] O. Bratteli, D.W. Robinson, Operator Algebras and Quantum Statistical Mechanics, vol. 1, 2nd edition, Springer-Verlag, 2002, 2nd printing.
[9] M.R. Bridson, A. Haefliger, Metric Spaces of Non-Positive Curvature, Grundlehren Math. Wiss., vol. 319, Springer-Verlag, 1999.
[10] E. Cartan, Leçons sur la géométrie des espaces de Riemann, Gauthiers-Villars, Paris, 1928, 2nd edition, 1946.
[11] E.B. Davies, The structure and ideal theory of the predual of a Banach lattice, Trans. Amer. Math. Soc. 131 (1968) 544–555.
[12] J. Eells, B. Fuglede, Harmonic Maps Between Riemannian Polyhedra, Cambridge University Press, 2001.
[13] G.H. Hardy, J.E. Littlewood, G. Pólya, Some simple inequalities satisfied by convex functions, Messenger Math. 58 (1929) 145–152.
[14] G.H. Hardy, J.E. Littlewood, G. Pólya, Inequalities, 2nd edition, Cambridge Math. Lib., 1952. Reprinted 1988.
[15] J. Jost, Equilibrium maps between metric spaces, Calc. Var. 2 (1994) 173–204.
[16] J. Jost, Nonpositive Curvature: Geometric and Analytic Aspects, Lectures Math. ETH Zürich, Birkhäuser Verlag, Basel, 1997.
[17] G. Larotonda, Nonpositive curvature: a geometric approach to Hilbert–Schmidt operators, Differential Geom. Appl. 25 (2007) 679–700.
[18] J.D. Lawson, Y. Lim, The geometric mean, matrices, metrics, and more, Amer. Math. Monthly 108 (9) (2001) 797–812.
[19] J.D. Lawson, Y. Lim, Monotonic properties of the least squares mean, Math. Ann. 351 (2011) 267–279.
[20] J.D. Lawson, Y. Lim, The least squares mean of positive Hilbert–Schmidt operators, J. Math. Anal. Appl. 403 (2013) 365–375.
[21] Y. Lim, Convex geometric means, J. Math. Anal. Appl. 404 (1) (2013) 115–128.
[22] A.W. Marshal, I. Olkin, B.C. Arnold, Inequalities: Theory of Majorization and Its Applications, 2nd edition, Springer-Verlag, 2011.
[23] R.J. McCann, A convexity principle for interacting gases, Adv. Math. 128 (1) (1997) 153–179.
[24] C.P. Niculescu, The Hermite–Hadamard inequality for convex functions on a global NPC space, J. Math. Anal. Appl. 356 (1) (2009) 295–301.
[25] C.P. Niculescu, L.-E. Persson, Convex Functions and Their Applications. A Contemporary Approach, CMS Books in Math., vol. 23, Springer-Verlag, New York, 2006.
[26] C.P. Niculescu, I. Rovenţa, Majorization in spaces with a curved geometry, preprint, available at http://arxiv.org/abs/1204.0929.
[27] M.A. Nielsen, I.L. Chuang, Quantum Computation and Quantum Information, Cambridge University Press, 2000.
[28] R.R. Phelps, Lectures on Choquet's Theorem, 2nd edition, Lecture Notes in Math., vol. 1757, Springer-Verlag, Berlin, 2001.
[29] S. Sherman, On a theorem of Hardy, Littlewood, Pólya, and Blackwell, Proc. Natl. Acad. Sci. USA 37 (1951) 826–831; Proc. Natl. Acad. Sci. USA 38 (1952) 382 (Erratum).
[30] K.T. Sturm, Probability measures on metric spaces of nonpositive curvature, in: Pascal Auscher, et al. (Eds.), Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces, Paris, France, April 16–July 13, 2002, in: Contemp. Math., vol. 338, 2003, pp. 357–390. Lecture notes from a quarter program on heat kernels, random walks, and analysis on manifolds and graphs.
[31] M. Tomić, Théorème de Gauss relatif au centre de gravité et son application, Bull. Soc. Math. Phys. Serbie 1 (1949) 31–40.
[32] H. Weyl, Inequalities between two kinds of eigenvalues of a linear transformation, Proc. Natl. Acad. Sci. USA 35 (1949) 408–411.
[33] T. Yamazaki, An elementary proof of arithmetic–geometric mean inequality of the weighted Riemannian mean of positive definite matrices, Linear Algebra Appl. 438 (4) (2011) 1564–1569.