

THE HALÁSZ–SZÉKELY BARYCENTER

JAIRO BOCHI¹, GODOFREDO IOMMI²  AND MARIO PONCE²

¹*Department of Mathematics, The Pennsylvania State University, University Park, PA, USA (bochi@psu.edu)*

²*Facultad de Matemáticas, Pontificia Universidad Católica de Chile, Santiago, Chile (giommi@mat.uc.cl; mponcea@mat.uc.cl)*

(Received 18 November 2021; first published online 13 October 2022)

Abstract We introduce a notion of barycenter of a probability measure related to the symmetric mean of a collection of non-negative real numbers. Our definition is inspired by the work of Halász and Székely, who in 1976 proved a law of large numbers for symmetric means. We study the analytic properties of this Halász–Székely barycenter. We establish fundamental inequalities that relate the symmetric mean of a list of non-negative real numbers with the barycenter of the measure uniformly supported on these points. As consequence, we go on to establish an ergodic theorem stating that the symmetric means of a sequence of dynamical observations converge to the Halász–Székely barycenter of the corresponding distribution.

Keywords: symmetric mean; barycenter; ergodic theorem

2020 *Mathematics subject classification:* Primary 26E60
Secondary 26D15; 15A15; 37A30; 60F15

1. Introduction

Means have fascinated man for a long time. Ancient Greeks knew the arithmetic, geometric, and harmonic means of two positive numbers (which they may have learned from the Babylonians); they also studied other types of means that can be defined using proportions: see [17, pp. 85–89]. Newton and Maclaurin encountered the symmetric means (more about them later). Huygens introduced the notion of expected value and Jacob Bernoulli proved the first rigorous version of the law of large numbers, see [20, pp. 51, 73]. Gauss and Lagrange exploited the connection between the arithmetico-geometric mean and elliptic functions: see [6]. Kolmogorov and other authors considered means from an axiomatic point of view and determined when a mean is arithmetic under a change of coordinates (i.e. *quasiarithmetic*), see [16, p. 157–163], [1, Chapter 17]. Means and inequalities between them are the main theme of the classical book [16] by Hardy, Littlewood, and Pólya, and the book [8] by Bullen is a comprehensive account of the

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

subject. Going beyond the real line, there are notions of averaging that relate to the geometric structure of the ambient space, see e.g. [11, 18, 22, 25]. The paper [4] contains an excellent summary of the history of means.

In this paper, we are interested in one of the most classical types of means: the elementary symmetric polynomials means, or symmetric means for short. Let us recall their definition. Given integers $n \geq k \geq 1$, the k th symmetric mean of a list of non-negative numbers x_1, \dots, x_n is:

$$\text{sym}_k(x_1, \dots, x_n) := \left(\frac{E_k^{(n)}(x_1, \dots, x_n)}{\binom{n}{k}} \right)^{\frac{1}{k}}, \quad (1.1)$$

where $E_k^{(n)}(x_1, \dots, x_n) := \sum_{i_1 < \dots < i_k} x_{i_1} \cdots x_{i_k}$ is the elementary symmetric polynomial of degree k in n variables. Note that the extremal cases $k = 1$ and $k = n$ correspond to arithmetic and the geometric means, respectively. The symmetric means are non-increasing as functions of k : this is Maclaurin's inequality: see [16, p. 52] or [8, p. 327]. For much more information on symmetric means and their relatives, see [8, Chapter V].

Let us now turn to Probability Theory. A law of large numbers in terms of symmetric means was obtained by Halász and Székely [15], confirming a conjecture of Székely [26]. Let X_1, X_2, \dots be a sequence of non-negative independent identically distributed random variables, and from them, we form another sequence of random variables:

$$S_n := \text{sym}_k(X_1, \dots, X_n), \quad (1.2)$$

The case of $k = 1$ corresponds to the setting of the usual law of large numbers. The case of constant $k > 1$ is not significantly different from the classical setting. Things become more interesting if k is allowed to depend on n , and it turns out to be advantageous to assume that k/n converges to some number $c \in [0, 1]$. In this case, Halász and Székely [15] have proved that if $X = X_1$ is strictly positive and satisfies some integrability conditions, then S_n converges almost surely to a non-random constant. Furthermore, they gave a formula for this limit, which we call the *Halász–Székely mean* with parameter c of the random variable X . Halász and Székely theorem was extended to the non-negative situation by van Es [28] (with appropriate extra hypotheses). The simplest example consists of a random variable X that takes two non-negative values x and y , each with probability $1/2$, and $c = 1/2$; in this case, the Halász–Székely mean is $(\frac{\sqrt{x} + \sqrt{y}}{2})^2$. But this example is misleadingly simple, and Halász–Székely means are in general unrelated to power means.

Fixed the parameter c , the Halász–Székely mean of a non-negative random variable X only depends on its distribution, which we regard as a probability measure μ on the half-line $[0, +\infty)$. Now we shift our point of view and consider probability measures as the fundamental objects. Instead of speaking of the mean of a probability measure, we prefer the word *barycenter*, reserving the word *mean* for lists of numbers (with or without weights), functions, and random variables. This is more than a lexical change. The space of probability measures has a great deal of structure: it is a convex space and it can be endowed with several topologies. So we arrive at the notion of *Halász–Székely barycenter* (or *HS barycenter*) of a probability measure μ with parameter c , which we denote $[\mu]_c$. This is the subject of this paper. It turns out that HS barycenters can be defined directly, without resorting to symmetric means or laws of large numbers (see Definition 2.3).

Symmetric means are intrinsically discrete objects and do not make sense as barycenters. In [8, Remark, p. 323], Bullen briefly proposes a definition of a weighted symmetric mean, only to conclude that ‘the properties of this weighted mean are not satisfactory’ and therefore, not worthy of further consideration. On the other hand, given a finite list $\underline{x} = (x_1, \dots, x_n)$ of non-negative numbers, we can compare the symmetric means of \underline{x} with the HS barycenter of the associated probability measure $\mu := (\delta_{x_1} + \dots + \delta_{x_n})/n$. It turns out that these quantities obey certain precise inequalities (see Theorem 3.4). In particular, we have:

$$\text{sym}_k(\underline{x}) \geq [\mu]_{k/n}. \quad (1.3)$$

Furthermore, if $\underline{x}^{(m)}$ denotes the nm -tuple obtained by concatenation of m copies of \underline{x} , then

$$[\mu]_{k/n} = \lim_{m \rightarrow \infty} \text{sym}_{km}(\underline{x}^{(m)}), \quad (1.4)$$

and we have precise bounds for the relative error of this approximation, depending only on the parameters and not on the numbers x_i themselves.

Being a natural limit of symmetric means, the HS barycenters deserve to be studied by their own right. One can even argue that they give the ‘right’ notion of weighted symmetric means that Bullen was looking for. HS barycenters have rich theoretical properties. They are also cheap to compute while computing symmetric means involves summing exponentially many terms.

Using our general inequalities and certain continuity properties of the HS barycenters, we are able to obtain in straightforward manner an ergodic theorem that extends the laws of large numbers of Halász–Székely [15] and van Es [28].

A prominent feature of the symmetric mean (1.1) is that it vanishes whenever more than $n - k$ of the numbers x_i vanish. Consequently, the HS barycenter $[\mu]_c$ of a probability measure μ on $[0, +\infty)$ vanishes when $\mu(\{0\}) > 1 - c$. In other words, once the mass of leftmost point 0 exceeds the critical value $1 - c$, then it imposes itself on the whole distribution, and suddenly forces the mean to agree with it. Fortunately, in the subcritical regime, $\mu(\{0\}) < 1 - c$, the HS barycenter turns out to be much better behaved. As it will be seen in § 2, in the critical case $\mu(\{0\}) = 1 - c$ the HS barycenter can be either positive or zero, so the HS barycenter can actually vary discontinuously. Therefore, our regularity results and the ergodic theorem must take this critical phenomenon into account.

This article is organized as follows. In § 2, we define formally the HS barycenters and prove some of their basic properties. In § 3, we state and prove the fundamental inequalities relating HS barycenters to symmetric means. In § 4, we study the problem of continuity of the HS barycenters with respect to appropriate topologies on spaces of probability measures. In § 5, we apply the results of the previous sections and derive a general ergodic theorem (law of large numbers) for symmetric and HS means.

2. Presenting the HS barycenter

Hardy, Littlewood, and Pólya’s axiomatization of (quasiarithmetic) means [16, § 6.19] is formulated in terms of distribution functions, using Stieltjes integrals. Since the first publication of their book in 1934, *measures* became established as fundamental objects in mathematical analysis, probability theory, dynamical systems, etc. Spaces of measures

have been investigated in depth (see e.g. the influential books [23, 29]). The measure-theoretic point of view provides the convenient structure for the analytic study of means or, as we prefer to call them in this case, *barycenters*. The simplest example of barycenter is of course the ‘arithmetic barycenter’ of a probability measure μ on Euclidean space \mathbb{R}^d , defined (under the appropriate integrability condition) as $\int x d\mu(x)$. Another example is the ‘geometric barycenter’ of a probability measure μ on the half-line $(0, +\infty)$, defined as $\exp(\int \log x d\mu(x))$. In this section, we introduce the Halász–Székely barycenters and study some of their basic properties.

2.1. Definitions and basic properties

Throughout this paper we use the following notation:

$$\mathbb{R}_+ := [0, +\infty), \quad \mathbb{R}_{++} := (0, +\infty). \tag{2.1}$$

We routinely work with the extended line $[-\infty, +\infty]$, endowed with the order topology.

Definition 2.1. The *Halász–Székely kernel* (or *HS kernel*) is the following function of three variables $x \in \mathbb{R}_+$, $y \in \mathbb{R}_{++}$, and $c \in [0, 1]$:

$$K(x, y, c) := \begin{cases} \log y + c^{-1} \log(cy^{-1}x + 1 - c) & \text{if } c > 0, \\ \log y + y^{-1}x - 1 & \text{if } c = 0. \end{cases} \tag{2.2}$$

Proposition 2.2. *The HS kernel has the following properties (see also Figure 1):*

- (a) *The function $K : [0, +\infty) \times (0, +\infty) \times [0, 1] \rightarrow [-\infty, +\infty)$ is continuous, attaining the value $-\infty$ only at the points $(0, y, 1)$.*
- (b) *$K(x, y, c)$ is increasing with respect to x .*
- (c) *$K(x, y, c)$ is decreasing with respect to c , and strictly decreasing when $x \neq y$.*
- (d) *$K(x, y, 1) = \log x$ is independent of y .*
- (e) *$K(x, y, c) \geq \log x$, with equality if and only if $x = y$ or $c = 1$.*
- (f) *For each $y > 0$, the function $K(\cdot, y, 0)$ is affine, and its graph is the tangent line to $\log x$ at $x = y$.*
- (g) *$K(\lambda x, \lambda y, c) = K(x, y, c) + \log \lambda$, for all $\lambda > 0$.*

Proof. Most properties are immediate from Definition 2.1. To check monotonicity with respect to c , we compute the partial derivative when $c > 0$:

$$K_c(x, y, c) = \frac{1}{c^2} \left[\frac{c(y^{-1}x - 1)}{cy^{-1}x + 1 - c} + \log \left(\frac{1}{cy^{-1}x + 1 - c} \right) \right]; \tag{2.3}$$

since $\log t \leq t - 1$ (with equality only if $t = 1$), we conclude that $K_c(x, y, c) \leq 0$ (with equality only if $x = y$). Since K is continuous, we obtain property (c). Property (e) is a consequence of properties (c) and (d). □

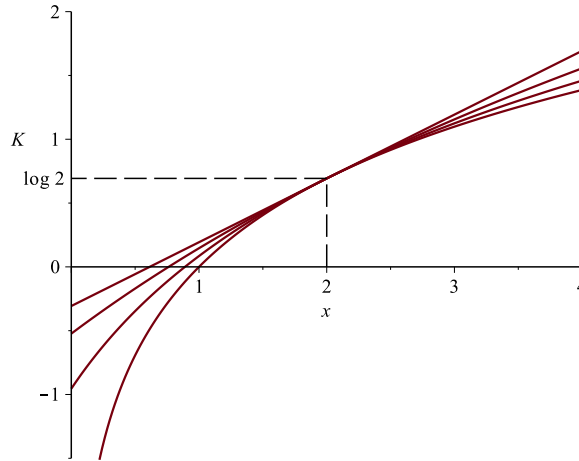


Figure 1. Graphs of the functions $K(\cdot, y, c)$ for $y = 2$ and $c \in \{0, 1/3, 2/3, 1\}$.

Let $\mathcal{P}(\mathbb{R}_+)$ denote the set of all Borel probability measures μ on \mathbb{R}_+ . The following is the central concept of this paper:

Definition 2.3. Let $c \in [0, 1]$ and $\mu \in \mathcal{P}(\mathbb{R}_+)$. If $c = 1$, then we require that the function $\log x$ is semi-integrable* with respect to μ . The *Halász–Székely barycenter* (or *HS barycenter*) with parameter c of the probability measure μ is:

$$[\mu]_c := \exp \inf_{y>0} \int K(x, y, c) d\mu(x), \tag{2.4}$$

where K is the HS kernel (2.2).

First of all, let us see that the definition is meaningful:

- If $c < 1$, then for all $y > 0$, the function $K(\cdot, y, c)$ is bounded from below by $K(0, y, c) > -\infty$, and therefore, it has a well-defined integral (possibly $+\infty$); so $[\mu]_c$ is a well-defined element of the extended half-line $[0, +\infty]$.
- If $c = 1$, then by part (d) of Proposition 2.2, the defining formula (2.4) becomes:

$$[\mu]_1 = \exp \int \log x d\mu(x). \tag{2.5}$$

The integral is a well-defined element of $[-\infty, +\infty]$, so $[\mu]_1$ is well defined in $[0, +\infty]$.

Formula (2.5) means that the HS barycenter with parameter $c = 1$ is the geometric barycenter; let us see that $c = 0$ corresponds to the standard arithmetic barycenter:

Proposition 2.4. For any $\mu \in \mathcal{P}(\mathbb{R}_+)$, we have $[\mu]_0 = \int x d\mu(x)$.

* A function f is called *semi-integrable* if the positive part $f^+ := \max(f, 0)$ is integrable or the negative part $f^- := \max(-f, 0)$ is integrable.

Proof. Let $a := \int x d\mu(x)$. If $a = \infty$, then for every $y > 0$, the non-constant affine function $K(\cdot, y, 0)$ has infinite integral, so definition (2.4) gives $[\mu]_0 = \infty$. On the other hand, if $a < \infty$, then $[\mu]_0$ is defined as $\exp \inf_{y>0} (\log y + a/y - 1) = a$. □

Let $\mathcal{P}_{\text{HS}}(\mathbb{R}_+)$ denote the subset formed by those $\mu \in \mathcal{P}(\mathbb{R}_+)$ such that:

$$\int \log(1+x) d\mu(x) < \infty \tag{2.6}$$

or, equivalently, $\int \log^+ x d\mu < \infty$ (we will sometimes write ‘ $d\mu$ ’ instead of ‘ $d\mu(x)$ ’).

Proposition 2.5. *Let $c \in (0, 1]$ and $\mu \in \mathcal{P}(\mathbb{R}_+)$. Then $[\mu]_c < \infty$ if and only if $\mu \in \mathcal{P}_{\text{HS}}(\mathbb{R}_+)$.*

Proof. The case $c = 1$ being clear, assume that $c \in (0, 1)$. Note that for all $y > 0$, the expression

$$|K(x, y, c) - c^{-1} \log(x + 1)| \tag{2.7}$$

is a bounded function of x , so the integrability of $K(x, y, c)$ and $\log(x + 1)$ are equivalent. □

Next, let us see that the standard properties one might expect for something called a ‘barycenter’ are satisfied. For any $x \geq 0$, we denote by δ_x the probability measure such that $\delta_x(\{x\}) = 1$.

Proposition 2.6. *For all $c \in [0, 1]$ and $\mu \in \mathcal{P}(\mathbb{R}_+)$, the following properties hold:*

- (a) Reflexivity: $[\delta_x]_c = x$, for every $x \geq 0$.
- (b) Monotonicity with respect to the measure: If $\mu_1, \mu_2 \in \mathcal{P}(\mathbb{R}_+)$ have distribution functions F_1, F_2 such that $F_1 \geq F_2^\dagger$, then $[\mu_1]_c \leq [\mu_2]_c$.
- (c) Internality: If $\mu(I) = 1$ for an interval $I \subseteq \mathbb{R}_+$, then $[\mu]_c \in I$.
- (d) Homogeneity: If $\lambda \geq 0$, and $\lambda_*\mu$ denotes the pushforward of μ under the map $x \mapsto \lambda x$, then $[\lambda_*\mu]_c = \lambda[\mu]_c$.
- (e) Monotonicity with respect to the parameter: If $0 \leq c' \leq c \leq 1$, then $[\mu]_{c'} \geq [\mu]_c$.

Proof. The proofs use the properties of the HS kernel listed in Proposition 2.2. Reflexivity is obvious when $c = 1$ or $x = 0$, and in all other cases follows from property (e). Monotonicity with respect to the measure is a consequence of the fact that the HS kernel is increasing in x . The internality property of the HS barycenter follows from reflexivity and monotonicity. Homogeneity follows from property (g) of the HS kernel and the change of variables formula. Finally, monotonicity with respect to the parameter c is a consequence of the corresponding property of the HS kernel. □

[†] I.e., μ_2 is ‘more to the right’ than μ_1 . This defines a partial order, called *usual stochastic ordering* or *first-order stochastic dominance*.

As it will be clear later (see Example 2.13), the internality and the monotonicity properties (w.r.t. μ and w.r.t. c) are not strict.

2.2. Computation and critical phenomenon

In the remaining of this section, we discuss how to actually compute HS barycenters. In view of Proposition 2.5, we may focus on measures in $\mathcal{P}_{\text{HS}}(\mathbb{R}_+)$. The mass of zero plays an important role. Given $c \in (0, 1)$ and $\mu \in \mathcal{P}_{\text{HS}}(\mathbb{R}_+)$, we use the following terminology, where $\mu(0) = \mu(\{0\})$:

$$\begin{cases} \text{subcritical case:} & \mu(0) < 1 - c \\ \text{critical case:} & \mu(0) = 1 - c \\ \text{supercritical case:} & \mu(0) > 1 - c \end{cases} \tag{2.8}$$

The next result establishes a way to compute $[\mu]_c$ in the subcritical case; the remaining cases will be dealt with later in Proposition 2.11.

Proposition 2.7. *If $\mu \in \mathcal{P}_{\text{HS}}(\mathbb{R}_+)$, $c \in (0, 1)$ and $\mu(0) < 1 - c$ (subcritical case), then the equation*

$$\int K_y(x, \eta, c) d\mu(x) = 0 \tag{2.9}$$

has a unique positive and finite solution $\eta = \eta(\mu, c)$, and the inf in formula (2.4) is attained uniquely at $y = \eta$; in particular,

$$[\mu]_c = \exp \int K(x, \eta, c) d\mu(x). \tag{2.10}$$

Proof. Fix μ and c as in the statement. We compute the partial derivative:

$$K_y(x, y, c) = \frac{\Delta(x, y)}{y}, \quad \text{where} \quad \Delta(x, y) := 1 - \frac{x}{cx + (1 - c)y}. \tag{2.11}$$

Since Δ is bounded, we are allowed to differentiate under the integral sign:

$$\frac{d}{dy} \int K(x, y, c) d\mu = \int K_y(x, \eta, c) d\mu = \frac{\psi(y)}{y}, \tag{2.12}$$

where $\psi(y) := \int \Delta(x, y) d\mu$. The partial derivative

$$\Delta_y(x, y) = \frac{(1 - c)x}{(cx + (1 - c)y)^2} \tag{2.13}$$

is positive, except at $x = 0$. Since $\mu \neq \delta_0$, the function ψ is strictly increasing. Furthermore,

$$\lim_{y \rightarrow +\infty} \Delta(x, y) = 1 \quad \text{and} \quad \lim_{y \rightarrow 0^+} \Delta(x, y) = \begin{cases} 1 - 1/c & \text{if } x > 0, \\ 1 & \text{if } x = 0, \end{cases} \tag{2.14}$$

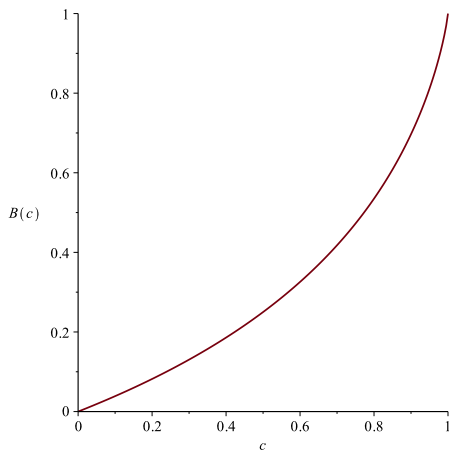


Figure 2. Graph of the function B defined by (2.17).

and so

$$\lim_{y \rightarrow +\infty} \psi(y) = 1 \quad \text{and} \quad \lim_{y \rightarrow 0^+} \psi(y) = 1 - \frac{1 - \mu(0)}{c} < 0, \tag{2.15}$$

using the assumption $\mu(0) < 1 - c$. Therefore, there exists a unique $\eta > 0$ that solves the equation $\psi(\eta) = 0$, or equivalently equation (2.9). By (2.12), the function $y \mapsto \int K(x, y, c) d\mu$ decreases on $(0, \eta]$ and increases on $[\eta, +\infty)$, and so attains its infimum at η . Formula (2.10) follows from the definition of $[\mu]_c$. \square

Let us note that, as a consequence of (2.11), equation (2.9) is equivalent to:

$$\int \frac{x}{cx + (1 - c)\eta} d\mu(x) = 1. \tag{2.16}$$

Remark 2.8. If μ belongs to $\mathcal{P}(\mathbb{R}_+)$ but not to $\mathcal{P}_{\text{HS}}(\mathbb{R}_+)$, and still $c \in (0, 1)$, then equation (2.9) (or its equivalent version (2.16)) still has a unique positive and finite solution $\eta = \eta(\mu, c)$, and formula (2.10) still holds. On the other hand, if $c = 0$ and $\int x d\mu(x) < \infty$, then all conclusions of Proposition 2.7 still hold, with a similar proof.

We introduce the following auxiliary function, plotted in Figure 2:

$$B(c) := \begin{cases} 0 & \text{if } c = 0, \\ c(1 - c)^{\frac{1-c}{c}} & \text{if } 0 < c < 1, \\ 1 & \text{if } c = 1. \end{cases} \tag{2.17}$$

The following alternative formula for the HS barycenter matches the original one from [15], and in some situations is more convenient:

Proposition 2.9. *If $0 < c \leq 1$ and $\mu \in \mathcal{P}_{\text{HS}}(\mathbb{R}_+)$, then:*

$$[\mu]_c = B(c) \inf_{r>0} \left\{ r^{-\frac{1-c}{c}} \exp\left[c^{-1} \int \log(x+r) d\mu(x)\right] \right\}. \tag{2.18}$$

Furthermore, if $\mu(0) < 1 - c$, then the inf is attained at the unique positive finite solution $\rho = \rho(\mu, c)$ of the equation

$$\int \frac{x}{x+\rho} d\mu(x) = c. \tag{2.19}$$

Proof. The formula is obviously correct if $c = 1$. If $0 < c < 1$, we introduce the variable $r := \frac{1-c}{c} y$ in formula (2.4) and manipulate. Similarly, (2.16) becomes (2.19). \square

If μ is a Borel probability measure on \mathbb{R}_+ not entirely concentrated at zero (i.e., $\mu \neq \delta_0$) then we denote by μ^+ the probability measure obtained by conditioning on the event $\mathbb{R}_{++} = (0, \infty)$, that is,

$$\mu^+(U) := \frac{\mu(U \cap \mathbb{R}_{++})}{\mu(\mathbb{R}_{++})}, \quad \text{for every Borel set } U \subseteq \mathbb{R}. \tag{2.20}$$

Obviously, if $\mu \in \mathcal{P}_{\text{HS}}(\mathbb{R}_+)$, then $\mu^+ \in \mathcal{P}_{\text{HS}}(\mathbb{R}_+)$ as well.

Proposition 2.10. *Let $\mu \in \mathcal{P}_{\text{HS}}(\mathbb{R}_+)$ and $p := 1 - \mu(0)$. If $c \in (0, 1]$ and $c \leq p$ (critical or subcritical cases), then*

$$[\mu]_c = \frac{B(c)}{B(c/p)} [\mu^+]_{c/p}. \tag{2.21}$$

Proof. Note that $\mu \neq \delta_0$, so the positive part μ^+ is defined. Using formula (2.18), we have:

$$[\mu]_c = B(c) \inf_{r>0} r^{1-\frac{1}{c}} \exp\left(\frac{1}{c} \int \log(x+r) d\mu\right) \tag{2.22}$$

$$= B(c) \inf_{r>0} r^{1-\frac{1}{c}} \exp\left(\frac{1}{c} (\mu(0) \log r + \int_{\{x>0\}} \log(x+r) d\mu)\right) \tag{2.23}$$

$$= B(c) \inf_{r>0} r^{1-\frac{1}{c} + \frac{1-p}{c}} \exp\left(\frac{1}{c} \int_{\{x>0\}} \log(x+r) d\mu\right) \tag{2.24}$$

$$= B(c) \inf_{r>0} r^{1-\frac{p}{c}} \exp\left(\frac{1}{c/p} \int \log(x+r) d\mu^+\right). \tag{2.25}$$

At this point, the assumption $c \leq p$ guarantees that the barycenter $[\mu^+]_{c/p}$ is well defined, and using (2.18) again we obtain (2.21). \square

Finally, we compute the HS barycenter in the critical and supercritical cases:

Proposition 2.11. *Let $\mu \in \mathcal{P}_{\text{HS}}(\mathbb{R}_+)$ and $c \in (0, 1]$.*

(a) Critical case: *If $\mu(0) = 1 - c$, then $[\mu]_c = B(c)[\mu^+]_1$.*

(b) Supercritical case: *If $\mu(0) > 1 - c$, then $[\mu]_c = 0$.*

In both cases above, the infimum in formula (2.4) is not attained.

Proof. In the critical case, we use (2.21) with $p = c$ and conclude.

In the supercritical case, we can assume that $\mu \neq \delta_0$. Note that $p < c$, thus, $\lim_{r \rightarrow 0^+} r^{1-\frac{p}{c}} = 0$. Moreover, since $\log^+ x \in L^1(\mu)$, we have

$$\lim_{r \rightarrow 0^+} \int_{\{x>0\}} \log(x+r) d\mu = \int_{\{x>0\}} \log x d\mu < +\infty. \tag{2.26}$$

Therefore, using (2.25), we obtain $[\mu]_c = 0$. □

Propositions 2.7 and 2.11 allow us to compute HS barycenters in all cases. For emphasis, let us list explicitly the situations where the barycenter vanishes:

Proposition 2.12. *Let $c \in [0, 1]$ and $\mu \in \mathcal{P}_{\text{HS}}(\mathbb{R}_+)$. Then $[\mu]_c = 0$ if and only if one of the following mutually exclusive situations occur:*

- (a) $c = 0$ and $\mu = \delta_0$.
- (b) $c > 0$, $\mu(0) = 1 - c$, and $\int \log x d\mu^+(x) = -\infty$.
- (c) $c > 0$ and $\mu(0) > 1 - c$.

Proof. The case $c = 0$ being obvious, assume that $c > 0$, and so $B(c) > 0$. In the critical case, part (a) of Proposition 2.11 tells us that $[\mu]_c = 0$ if and only if $[\mu^+]_1 = 0$, which by (2.5) is equivalent to $\int \log x d\mu^+ = -\infty$. In the supercritical case, part (b) of the proposition ensures that $[\mu]_c = 0$. □

Example 2.13. Consider the family of probability measures:

$$\mu_p := (1 - p)\delta_0 + p\delta_1, \quad 0 \leq p \leq 1. \tag{2.27}$$

If $0 < c < 1$, then

$$[\mu_p]_c = \begin{cases} 0 & \text{if } p < c \\ B(c) = c(1 - c)^{\frac{1-c}{c}} & \text{if } p = c \\ \frac{B(c)}{B(c/p)} = (1 - c)^{\frac{1-c}{c}} p^{\frac{p}{c}} (p - c)^{\frac{c-p}{c}} & \text{if } p > c. \end{cases} \tag{2.28}$$

These formulas were first obtained by Székely [26]. It follows that the function

$$(p, c) \in [0, 1] \times [0, 1] \mapsto [\mu_p]_c \tag{2.29}$$

(whose graph is shown on [28, p. 680]) is discontinuous at the points with $p = c > 0$, and only at those points. We will return to the issue of continuity in § 4.

The practical computation of HS barycenters usually requires numerical methods. In any case, it is useful to notice that the function η from Proposition 2.7 satisfies the internality property:

Lemma 2.14. *Let $\mu \in \mathcal{P}_{\text{HS}}(\mathbb{R}_+)$, $c \in (0, 1)$, and suppose that $\mu(0) < 1 - c$. If $\mu(I) = 1$ for an interval $I \subseteq \mathbb{R}_+$, then $\eta(\mu, c) \in I$.*

The proof is left to the reader.

3. Comparison with the symmetric means

3.1. HS means as repetitive symmetric means

The HS barycenter of a probability measure, introduced in the previous section, may now be specialized to the case of discrete equidistributed probabilities. So the *HS mean* of a tuple $\underline{x} = (x_1, \dots, x_n)$ of non-negative numbers with parameter $c \in [0, 1]$ is defined as:

$$\text{hsm}_c(\underline{x}) := \left[\frac{\delta_{x_1} + \dots + \delta_{x_n}}{n} \right]_c. \tag{3.1}$$

Using (2.18), we have more explicitly:

$$\text{hsm}_c(\underline{x}) = B(c) \inf_{r>0} r^{-\frac{1-c}{c}} \prod_{j=1}^n (x_j + r)^{\frac{1}{cn}} \quad (\text{if } c > 0), \tag{3.2}$$

where B is the function (2.17). On the other hand, recall that for $k \in \{1, \dots, n\}$, the k -th symmetric mean of the n -tuple \underline{x} is:

$$\text{sym}_k(\underline{x}) := \left(\frac{E_k^{(n)}(\underline{x})}{\binom{n}{k}} \right)^{\frac{1}{k}}, \tag{3.3}$$

where $E_k^{(n)}$ denotes the elementary symmetric polynomial of degree k in n variables.

Since they originate from a barycenter, the HS means are *repetition invariant*[‡] in the sense that, for any $m > 0$,

$$\text{hsm}_c(\underline{x}^{(m)}) = \text{hsm}_c(\underline{x}), \tag{3.4}$$

where $\underline{x}^{(m)}$ denotes the nm -tuple obtained by concatenation of m copies of the n -tuple \underline{x} . No such property holds for the symmetric means, even allowing for adjustment of the parameter k . Nevertheless, if the number of repetitions tends to infinity, then the symmetric means tend to stabilize, and the limit is a HS mean; more precisely:

Theorem 3.1. *If $\underline{x} = (x_1, \dots, x_n)$, $x_i \geq 0$, and $1 \leq k \leq n$, then:*

$$\lim_{m \rightarrow \infty} \text{sym}_{km}(\underline{x}^{(m)}) = \text{hsm}_{k/n}(\underline{x}). \tag{3.5}$$

Furthermore, the relative error goes to zero uniformly with respect to the x_i 's.

This theorem will be proved in the next subsection.

It is worthwhile to note that the Navas barycenter [22] is obtained as a ‘repetition limit’ similar to (3.5).

[‡] or *intrinsic*, in the terminology of [18, Def. 3.3]

Example 3.2. Using Propositions 2.7 and 2.11, one computes:

$$\text{hsm}_{1/2}(x, y) = \left(\frac{\sqrt{x} + \sqrt{y}}{2} \right)^2 ; \tag{3.6}$$

Therefore,

$$\lim_{m \rightarrow \infty} \text{sym}_m(\underbrace{x, \dots, x}_{m \text{ times}}, \underbrace{y, \dots, y}_{m \text{ times}}) = \left(\frac{\sqrt{x} + \sqrt{y}}{2} \right)^2 . \tag{3.7}$$

The last equality was deduced in [10, p. 31] from the asymptotics of Legendre polynomials.

Repeated symmetric means are particular instances of *Whiteley means*; in fact, in the notation of [8, p. 344],

$$\text{sym}_{km}(\underline{x}^{(m)}) = \mathfrak{M}_n^{[km, m]}(\underline{x}) . \tag{3.8}$$

Let us pose a problem:

Question 3.3. *Is the sequence $m \mapsto \text{sym}_{km}(\underline{x}^{(m)})$ always monotone decreasing?*

There exists a partial result: when $k = 1$ and $n = 2$, [10, Lemma 4.1] establishes *eventual* monotonicity.

3.2. Inequalities between symmetric means and HS means

The following is the first main result of this paper.

Theorem 3.4. *If $\underline{x} = (x_1, \dots, x_n)$, $x_i \geq 0$, and $1 \leq k \leq n$, then*

$$\text{hsm}_{k/n}(\underline{x}) \leq \text{sym}_k(\underline{x}) \leq \frac{\binom{n}{k}^{-1/k}}{B\left(\frac{k}{n}\right)} \text{hsm}_{k/n}(\underline{x}) . \tag{3.9}$$

Let us postpone the proof to the next subsection. The factor at the RHS of (3.9) is asymptotically 1 with respect to k ; indeed:

Lemma 3.5. *For all integers $n \geq k \geq 1$, we have*

$$1 \leq \frac{\binom{n}{k}^{-1/k}}{B\left(\frac{k}{n}\right)} < (9k)^{\frac{1}{2k}} , \tag{3.10}$$

with equality if and only if $k = n$.

Proof. Let $c := k/n$ and

$$b := \binom{n}{k} [B(c)]^k = \binom{n}{k} c^k (1 - c)^{n-k}. \tag{3.11}$$

By the Binomial Theorem, $b \leq 1$, which yields the first part of (3.10). For the lower estimate, we use the following Stirling bounds (see [12, p. 54, (9.15)]), valid for all $n \geq 1$,

$$\sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n} < n! < \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n+\frac{1}{12}}. \tag{3.12}$$

Then, a calculation (cf. [12, p. 184, (3.11)]) gives:

$$b > \frac{e^{-\frac{1}{6}}}{\sqrt{2\pi c(1-c)n}} > \frac{1}{\sqrt{9k}}, \tag{3.13}$$

from which the second part of (3.10) follows. □

Theorem 3.4 and Lemma 3.5 imply that HS means (with rational values of the parameter) can be obtained as repetition limits of symmetric means:

Proof of Theorem 3.1. Applying Theorem 3.4 to the tuple $\underline{x}^{(m)}$, using observation (3.4) and Lemma 3.5, we have:

$$\text{hsm}_{k/n}(\underline{x}) \leq \text{sym}_{km}(\underline{x}^{(m)}) \leq (9km)^{\frac{1}{2km}} \text{hsm}_{k/n}(\underline{x}). \tag{3.14}$$

Making $m \rightarrow \infty$ we obtain (3.5). □

Remark 3.6. If k is fixed, then

$$\lim_{n \rightarrow \infty} \frac{\binom{n}{k}^{-1/k}}{B\left(\frac{k}{n}\right)} = \frac{e(k!)^{1/k}}{k} > 1, \tag{3.15}$$

and therefore, the bound from Theorem 3.7 may be less satisfactory. But in this case, we may use the alternative bound coming from Maclaurin inequality:

$$\text{sym}_k(\underline{x}) \leq \text{sym}_1(\underline{x}) = \text{hsm}_0(\underline{x}). \tag{3.16}$$

3.3. Proof of Theorem 3.4

The two inequalities in (3.9) will be proved independently of each other. They are essentially contained in the papers [5] (the first inequality) and [15] (the second one), though neither was stated explicitly. In the following Theorems 3.8 and 3.7, we also characterize the cases of equality, and in particular show that each inequality is sharp in the sense that the corresponding factors cannot be improved.

Let us begin with the second inequality, which is more elementary. By symmetry, there is no loss of generality in assuming that the numbers x_i are ordered.

Theorem 3.7. *If $\underline{x} = (x_1, \dots, x_n)$ with $x_1 \geq \dots \geq x_n \geq 0$, and $1 \leq k \leq n$, then:*

$$\text{sym}_k(\underline{x}) \leq \frac{\binom{n}{k}^{-1/k}}{B\left(\frac{k}{n}\right)} \text{hsm}_{k/n}(\underline{x}). \tag{3.17}$$

Furthermore, equality holds if and only if $x_{k+1} = 0$ or $k = n$.

Proof. Our starting point is Vieta’s formula:

$$z^n + \sum_{\ell=1}^n E_\ell^{(n)}(x_1, \dots, x_n) z^{n-\ell} = \prod_{j=1}^n (x_j + z). \tag{3.18}$$

Therefore, by Cauchy’s formula, for any $r > 0$:

$$E_k^{(n)}(x_1, \dots, x_n) = \frac{1}{2\pi i} \oint_{|z|=r} \frac{1}{z^{n-k+1}} \prod_{j=1}^n (x_j + z) dz. \tag{3.19}$$

That is,

$$E_k^{(n)}(x_1, \dots, x_n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} (re^{i\theta})^{-n+k} \prod_{j=1}^n (x_j + re^{i\theta}) d\theta, \tag{3.20}$$

Taking absolute values,

$$E_k^{(n)}(x_1, \dots, x_n) \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} r^{-n+k} \prod_{j=1}^n |x_j + re^{i\theta}| d\theta \leq r^{-n+k} \prod_{j=1}^n (x_j + r). \tag{3.21}$$

But these inequalities are valid for all $r > 0$, and therefore:

$$E_k^{(n)}(x_1, \dots, x_n) \leq \inf_{r>0} r^{-n+k} \prod_{j=1}^n (x_j + r). \tag{3.22}$$

So formulas (3.2) and (3.3) imply inequality (3.17).

Now let us investigate the possibility of equality. We consider three mutually exclusive cases, which correspond to the classification (2.8):

$$\left. \begin{array}{l} \text{subcritical case: } x_{k+1} > 0 \\ \text{critical case: } x_k > 0 = x_{k+1} \\ \text{supercritical case: } x_k = 0 \end{array} \right\} \tag{3.23}$$

Using Proposition 2.11, in the critical case, we have:

$$\text{sym}_k(\underline{x}) = \left(\frac{x_1 \cdots x_k}{\binom{n}{k}} \right)^{\frac{1}{k}} \quad \text{and} \quad \text{hsm}_{k/n}(\underline{x}) = \frac{(x_1 \cdots x_k)^{\frac{1}{k}}}{B\left(\frac{k}{n}\right)}, \tag{3.24}$$

while in the supercritical case, the two means vanish together. So, in both cases, inequality (3.17) becomes an equality. Now suppose we are in the subcritical case; then the inf at

the RHS of (3.22) is attained at some $r > 0$: see Proposition 2.9. On the other hand, for this (and actually any) value of r , the second inequality in (3.21) must be strict, because the integrand is non-constant. We conclude that, in the subcritical case, inequality (3.22) is strict, and therefore, (3.17) is strict. \square

The first inequality in (3.9) is a particular case of an inequality between two types of *matrix means* introduced in [5], which we now explain. Let $A = (a_{i,j})_{i,j \in \{1, \dots, n\}}$ be a $n \times n$ matrix with non-negative entries. Recall that the *permanent* of A is the ‘signless determinant’

$$\text{per}(A) := \sum_{\sigma} \prod_{i=1}^n a_{i,\sigma(i)}, \tag{3.25}$$

where σ runs on the permutations of $\{1, \dots, n\}$. Then the *permanental mean* of A is defined as:

$$\text{pm}(A) := \left(\frac{\text{per}(A)}{n!} \right)^{\frac{1}{n}}. \tag{3.26}$$

On the other hand, the *scaling mean* of the matrix A is defined as:

$$\text{sm}(A) := \frac{1}{n^2} \inf_{u,v} \frac{u^\top A v}{\text{gm}(u)\text{gm}(v)}, \tag{3.27}$$

where u and v run on the set of strictly positive column vectors, and $\text{gm}(\cdot)$ denotes the geometric mean of the entries of the vector. Equivalently,

$$\text{sm}(A) = \frac{1}{n} \inf_v \frac{\text{gm}(Av)}{\text{gm}(v)}; \tag{3.28}$$

see [5, Remark 2.6].[§] By [5, Theorem 2.17],

$$\text{sm}(A) \leq \text{pm}(A), \tag{3.29}$$

with equality if and only if A has permanent 0 or rank 1. This inequality is far from trivial. Indeed, if the matrix A is doubly stochastic (i.e. row and column sums are all 1), then an easy calculation (see [5, Prop. 2.4]) shows that $\text{sm}(A) = \frac{1}{n}$, so (3.29) becomes $\text{pm}(A) \geq \frac{1}{n}$, or equivalently,

$$\text{per}(A) \geq \frac{n!}{n^n} \quad (\text{if } A \text{ is doubly stochastic}). \tag{3.30}$$

This lower bound on the permanent of doubly stochastic matrices was conjectured in 1926 by van der Waerden and, after a protracted series of partial results, proved around 1980 independently by Egorichev and Falikman: see [30, Chapter 5] for the exact

[§] Incidentally, formula (3.28) shows that, up to the factor $1/n$, the scaling mean is a *matrix antinorm* in the sense defined by [13].

references and a self-contained proof, and [14] for more recent developments. Our inequality (3.29), despite being a generalization of Egorichev–Falikman’s (3.30), is actually a relatively simple corollary of it: we refer the reader to [5, § 2] for more information.¶

We are now in position to complete the proof of Theorem 3.4, i.e., to prove the second inequality in (3.9). The next result also characterizes the cases of equality.

Theorem 3.8. *If $\underline{x} = (x_1, \dots, x_n)$ with $x_1 \geq \dots \geq x_n \geq 0$, and $1 \leq k \leq n$, then:*

$$\text{hsm}_{k/n}(\underline{x}) \leq \text{sym}_k(\underline{x}). \tag{3.31}$$

Furthermore, equality holds if and only if :

$$k = n \quad \text{or} \quad x_1 = \dots = x_n \quad \text{or} \quad x_k = 0. \tag{3.32}$$

Proof. Consider the non-negative $n \times n$ matrix:

$$A := \begin{pmatrix} x_1 & \dots & x_1 & 1 & \dots & 1 \\ \vdots & & \vdots & \vdots & & \vdots \\ x_n & \dots & x_n & 1 & \dots & 1 \end{pmatrix} \quad \text{with } n - k \text{ columns of 1's.} \tag{3.33}$$

Note that:

$$\text{per}(A) = k! (n - k)! E_k^{(n)}(x_1, \dots, x_n) \tag{3.34}$$

and so

$$\text{pm}(A) = \left(\frac{E_k^{(n)}(x_1, \dots, x_n)}{\binom{n}{k}} \right)^{\frac{1}{n}} \tag{3.35}$$

$$= [\text{sym}_k(\underline{x})]^{k/n}. \tag{3.36}$$

Now let’s compute the scaling mean of A using formula (3.28). Assume that $k < n$. Given a column vector $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$ with positive entries, we have:

$$\text{gm}(Av) = \prod_{i=1}^n (sx_i + r)^{\frac{1}{n}}, \quad \text{where } s := \sum_{i=1}^k v_i \text{ and } r := \sum_{i=k+1}^n v_i. \tag{3.37}$$

On the other hand, by the inequality of arithmetic and geometric means,

$$\text{gm}(v) \leq \left(\frac{s}{k}\right)^{\frac{k}{n}} \left(\frac{r}{n-k}\right)^{\frac{n-k}{n}}, \tag{3.38}$$

with equality if $v_1 = \dots = v_k = \frac{s}{k}$, $v_{k+1} = \dots = v_n = \frac{r}{n-k}$. So, in order to minimize the quotient $\frac{\text{gm}(Av)}{\text{gm}(v)}$, it is sufficient to consider column vectors v satisfying these conditions.

¶ The proof of [5, Theorem 2.17] uses a theorem on the existence of particular type of matrix factorization called Sinkhorn decomposition. The present article only needs the inequality (3.29) for matrices of a specific form (3.33). So the use of the existence theorem could be avoided, since it is possible to explicitly compute the corresponding Sinkhorn decomposition.

We can also normalize s to 1, and (3.28) becomes:

$$\text{sm}(A) = \inf_{r>0} \frac{\prod_{i=1}^n (x_i + r)^{\frac{1}{n}}}{\left(\frac{1}{k}\right)^{\frac{k}{n}} \left(\frac{r}{n-k}\right)^{\frac{n-k}{n}}} \tag{3.39}$$

$$= [\text{hsm}_{k/n}(\underline{x})]^{k/n} \tag{3.40}$$

by (3.2). This formula $\text{sm}(A) = [\text{hsm}_{k/n}(\underline{x})]^{k/n}$ also holds for $k = n$, taking the form $\text{sm}(A) = (x_1 \cdots x_n)^{\frac{1}{n}}$; this can be checked either by adapting the proof above, or more simply by using the homogeneity and reflexivity properties of the scaling mean (see [5]).

In conclusion, the matrix (3.33) has scaling and permanental means given by formulas (3.40) and (3.36), respectively, and the fundamental inequality (3.29) translates into $\text{hsm}_{k/n}(\underline{x}) \leq \text{sym}_k(\underline{x})$, that is, (3.8).

Furthermore, equality holds if and only if the matrix A defined by (3.33) satisfies $\text{sm}(A) = \text{pm}(A)$, by formulas (3.40) and (3.36). As mentioned before, $\text{sm}(A) = \text{pm}(A)$ if and only if A has rank 1 or permanent 0 (see [5, Theorem 2.17]). Note that A has rank 1 if and only if $k = 1$ or $x_1 = \cdots = x_n$. On the other hand, by (3.36), A has permanent 0 if and only if $\text{sym}_k(\underline{x}) = 0$, or equivalently $x_k = 0$. So we have proved that equality $\text{hsm}_{k/n}(\underline{x}) = \text{sym}_k(\underline{x})$ is equivalent to condition (3.32). \square

We close this section with some comments on related results.

Remark 3.9. In [15], the asymptotics of the integral (3.19) are determined using the saddle point method (see e.g. [24, Section 15.4]). However, for this method to work, the saddle must be steep, that is, the second derivative at the saddle must be large in absolute value. Major [21, p. 1987] discusses this situation: if the second derivative vanishes, then ‘a more sophisticated method has to be applied and only weaker results can be obtained in this case. We shall not discuss this question in the present paper’. On the other hand, in the general situation covered by our Theorem 3.4, the saddle can be flat. (It must be noted that the setting considered by Major is different, since he allows random variables to be negative.)

Remark 3.10. Given an arbitrary $n \times n$ non-negative matrix A , the permanental and scaling means satisfy the following inequalities (see [5, Theorem 2.17]),

$$\text{sm}(A) \leq \text{pm}(A) \leq n(n!)^{-1/n} \text{sm}(A). \tag{3.41}$$

The sequence $(n(n!)^{-1/n})$ is increasing and converges to e . In general, as n tends to infinity the permanental mean does not necessarily converge to the scaling mean. However, there are some special classes of matrices for which this is indeed the case: for example, in the repetitive situation covered by the *generalized Friedland limit* [5, Theorem 2.19]. Note that $\text{hsm}_{k/n}(\underline{x})$ and $\text{sym}_k(\underline{x})$ correspond to the n/k -th power of the scaling and permanental mean of the matrix A , respectively. Therefore, (3.9) can be regarded as an improvement of (3.41) for this particular class of matrices.

Remark 3.11. A natural extension of symmetric means is *Muirhead means*, see [16, § 2.18], [8, § V.6] for definition and properties. Accordingly, it should be possible to define a family of barycenters extending the HS barycenters, taking over from [5, § 5.2]. An analogue of inequality (3.31) holds in this extended setting, again as a consequence of the key inequality (3.29) between matrix means. However, we do not know if inequality (3.17) can be extended in a comparable level of generality.

4. Continuity of the HS barycenter

In this section, we study the continuity of the HS barycenter as a two-variable function, $(\mu, c) \mapsto [\mu]_c$, defined in the space $\mathcal{P}(\mathbb{R}_+) \times [0, 1]$. The most natural topology on $\mathcal{P}(\mathbb{R}_+)$ is the weak topology (defined below). The barycenter function is not continuous with respect to this topology, but, on the positive side, it is lower semicontinuous, except in a particular situation. In order to obtain better results, we need to focus on subsets of measures satisfying the natural integrability conditions (usually (2.6), but differently for the extremal parameters $c = 0$ and $c = 1$), and endow these subsets with stronger topologies that are well adapted to the integrability assumptions.

In a preliminary subsection, we collect some general facts on topologies on spaces of measures. In the remaining subsections, we prove several results on continuity of the HS barycenter. And all these results will be used in combination to prove our general ergodic theorem in § 5.

4.1. Convergence of measures

If (X, d) is a separable complete metric space, let $C_b(X)$ be the set of all continuous bounded real functions on X , and let $\mathcal{P}(X)$ denote the set of all Borel probability measures on X . Recall (see e.g. [23]) that the *weak topology* is a metrizable topology on $\mathcal{P}(X)$ according to with a sequence (μ_n) converges to some μ if and only if $\int \phi d\mu_n \rightarrow \int \phi d\mu$ for every test function $\phi \in C_b(X)$; we say that (μ_n) converges *weakly* to μ , and denote this by $\mu_n \rightharpoonup \mu$. The space $\mathcal{P}(X)$ is Polish, and it is compact if and only if X is compact. Despite the space $C_b(X)$ being huge (nonseparable w.r.t. its usual topology if X is noncompact), by [23, Theorem II.6.6] we can nevertheless find a *countable* subset $\mathcal{C} \subseteq C_b(X)$ such that, for all (μ_n) and μ in $\mathcal{P}(X)$,

$$\mu_n \rightharpoonup \mu \iff \forall \phi \in \mathcal{C}, \int \phi d\mu_n \rightarrow \int \phi d\mu. \tag{4.1}$$

The following result deals with sequences of integrals $\int \phi_n d\mu_n$ where not only the measures but also the integrands vary, and bears a resemblance to Fatou’s Lemma:

Proposition 4.1. *Suppose that (μ_n) is a sequence in $\mathcal{P}(X)$ converging weakly to some measure μ , and that (ϕ_n) is a sequence of continuous functions on X converging uniformly on compact subsets to some function ϕ . Furthermore, assume that the functions ϕ_n are bounded from below by a constant $-C$ independent of n . Then, $\liminf_{n \rightarrow \infty} \int \phi_n d\mu_n \geq \int \phi d\mu$.*

Note that, as in Fatou’s Lemma, the integrals in Proposition 4.1 can be infinite.

Proof. Without loss of generality, assume that $C = 0$. Let $\lambda \in \mathbb{R}$ be such that $\lambda < \int \phi \, d\mu$. By the monotone convergence theorem, there exists $m \in \mathbb{N}$ such that $\int \min\{\phi, m\} \, d\mu > \lambda$. For a function $\psi : X \rightarrow \mathbb{R}$, let $\hat{\psi}(x) := \min\{\psi(x), m\}$. Note that

$$\int \phi_n \, d\mu_n \geq \int \hat{\phi}_n \, d\mu_n = \int (\hat{\phi}_n - \hat{\phi}) \, d\mu_n + \int \hat{\phi} \, d\mu_n. \tag{4.2}$$

By Prokhorov’s theorem (see e.g. [23, Theorem 6.7]), the sequence (μ_n) forms a *tight* set, that is, for every $\epsilon > 0$, there exists a compact set $K \subseteq X$ such that $\mu_n(X \setminus K) \leq \epsilon/(2m)$ for all n . Since (ϕ_n) converges uniformly on compact subsets to ϕ , we obtain:

$$\lim_{n \rightarrow \infty} \left| \int_K (\hat{\phi}_n - \hat{\phi}) \, d\mu_n \right| \leq \lim_{n \rightarrow \infty} \sup_{x \in K} |\hat{\phi}_n(x) - \hat{\phi}(x)| = 0. \tag{4.3}$$

We also have:

$$\left| \int_{X \setminus K} (\hat{\phi}_n - \hat{\phi}) \, d\mu_n \right| \leq \mu_n(X \setminus K) \sup_{x \in X \setminus K} |\hat{\phi}_n(x) - \hat{\phi}(x)| < \frac{\epsilon}{2m} 2m = \epsilon. \tag{4.4}$$

Since (μ_n) converges weakly to μ , we have $\lim_{n \rightarrow \infty} \int \hat{\phi} \, d\mu_n = \int \hat{\phi} \, d\mu > \lambda$. Therefore, combining (4.2), (4.3) and (4.4), for sufficiently large values of n , we obtain $\int \phi_n \, d\mu_n > \lambda$. The result now follows. □

The next direct consequence is useful.

Corollary 4.2. *Suppose that (μ_n) is a sequence in $\mathcal{P}(X)$ converging weakly to some measure μ , and that (ϕ_n) is a sequence of continuous functions on X converging uniformly on compact subsets to some function ϕ . Furthermore, assume that the functions $|\phi_n|$ are bounded by a constant C independent of n . Then, $\int \phi_n \, d\mu_n \rightarrow \int \phi \, d\mu$.*

We will also need a slightly stronger notion of convergence. Let $\mathcal{P}_1(X) \subseteq \mathcal{P}(X)$ denote the set of measures μ with finite first moment, that is,

$$\int d(x, x_0) \, d\mu(x) < \infty, \tag{4.5}$$

Here and in what follows, $x_0 \in X$ is a basepoint which we consider as fixed, the particular choice being entirely irrelevant. We metrize $\mathcal{P}_1(X)$ with the *Kantorovich metric* (see e.g. [29, p. 207]):

$$W_1(\mu, \nu) := \sup \left\{ \int \psi \, d(\mu - \nu) ; \psi : X \rightarrow [-1, 1] \text{ is 1-Lipschitz} \right\}. \tag{4.6}$$

Of course, the Kantorovich metric depends on the original metric d on X ; in fact, it ‘remembers’ it, since $W_1(\delta_x, \delta_y) = d(x, y)$. The metric space $(\mathcal{P}_1(X), W_1)$ is called *1-Wasserstein space*; it is separable and complete. Unless X is compact, the topology on $\mathcal{P}_1(X)$ is stronger than the weak topology. In fact, we have the following characterizations of convergence:

Theorem 4.3 (Villani [29, Theorem 7.12]). For all (μ_n) and μ in $\mathcal{P}_1(X)$, the following statements are equivalent:

- (a) $W_1(\mu_n, \mu) \rightarrow 0$.
- (b) if $\phi: X \rightarrow \mathbb{R}$ is a continuous function such that $\frac{|\phi|}{1+d(\cdot, x_0)}$ is bounded, then $\int \phi d\mu_n \rightarrow \int \phi d\mu$.
- (c) $\mu_n \rightharpoonup \mu$ and $\int d(\cdot, x_0) d\mu_n \rightarrow \int d(\cdot, x_0) d\mu$.
- (d) $\mu_n \rightharpoonup \mu$ and the following ‘tightness condition’ holds:

$$\lim_{R \rightarrow \infty} \limsup_{n \rightarrow \infty} \int_{X \setminus B_R(x_0)} [1 + d(\cdot, x_0)] d\mu_n = 0. \tag{4.7}$$

where $B_R(x_0)$ denotes the open ball of centre x_0 and radius R .

The next lemma should be compared to Corollary 4.2:

Lemma 4.4. Suppose that (μ_n) is a sequence in $\mathcal{P}_1(X)$ converging to some measure μ , and that (ϕ_n) is a sequence of continuous functions on X converging uniformly on bounded subsets to some function ϕ . Furthermore, assume that the functions $\frac{|\phi_n|}{1+d(\cdot, x_0)}$ are bounded by a constant C independent of n . Then, $\int \phi_n d\mu_n \rightarrow \int \phi d\mu$.

Proof. Fix $\epsilon > 0$. By part (d) of Theorem 4.3, there exists $R > 0$ such that, for all sufficiently large n ,

$$\int_{X \setminus B_R(x_0)} [1 + d(x, x_0)] d\mu_n(x) \leq \epsilon. \tag{4.8}$$

Then, we write:

$$\begin{aligned} \int \phi_n d\mu_n - \int \phi d\mu &= \int (\phi_n - \phi) d\mu_n + \int \phi d(\mu_n - \mu) \\ &= \underbrace{\int_{X \setminus B_R(x_0)} (\phi_n - \phi) d\mu_n}_{\textcircled{1}} + \underbrace{\int_{B_R(x_0)} (\phi_n - \phi) d\mu_n}_{\textcircled{2}} + \underbrace{\int \phi d(\mu_n - \mu)}_{\textcircled{3}}. \end{aligned} \tag{4.9}$$

By part (b) of Theorem 4.3, the term $\textcircled{3}$ tends to 0 as $n \rightarrow \infty$. By the assumption of uniform convergence on bounded sets, $|\textcircled{2}| \leq \sup_{B_{2R}(x_0)} |\phi_n - \phi|$ tends to 0 as well. Finally,

$$|\textcircled{1}| \leq \int_{X \setminus B_R(x_0)} (|\phi_n| + |\phi|) d\mu_n \leq 2C \int_{X \setminus B_R(x_0)} [1 + d(x, x_0)] d\mu_n(x) \leq 2C\epsilon, \tag{4.10}$$

for all sufficiently large n . Since $\epsilon > 0$ is arbitrary, we conclude that $\int \phi_n d\mu_n \rightarrow \int \phi d\mu$, as claimed. □

Let us now consider the specific case of the metric space $(X, d) = (\mathbb{R}_+, d_{\text{HS}})$, where:

$$d_{\text{HS}}(x, y) := |\log(1 + x) - \log(1 + y)|. \tag{4.11}$$

Then the finite-first-moment condition (4.5) becomes our usual integrability condition (2.6), so the 1-Wasserstein space $\mathcal{P}_1(X)$ becomes $\mathcal{P}_{\text{HS}}(\mathbb{R}_+)$.

4.2. Lower and upper semicontinuity

The HS barycenter is definitely not continuous with respect to the weak topology, since the complement of $\mathcal{P}_{\text{HS}}(\mathbb{R}_+)$ is dense in $\mathcal{P}(\mathbb{R}_+)$, and the barycenter is ∞ there (by Proposition 2.5). Nevertheless, lower semicontinuity holds, except in the critical configuration:

Theorem 4.5. *For every $(\mu, c) \in \mathcal{P}(\mathbb{R}_+) \times [0, 1)$, we have:*

$$\liminf_{(\tilde{\mu}, \tilde{c}) \rightarrow (\mu, c)} [\tilde{\mu}]_{\tilde{c}} \geq [\mu]_c \quad \text{unless} \quad \mu(0) = 1 - c \text{ and } [\mu]_c > 0. \tag{4.12}$$

To be explicit, the inequality above means that for every $\lambda < [\mu]_c$, there exists a neighborhood $\mathcal{U} \subseteq \mathcal{P}(\mathbb{R}_+) \times [0, 1)$ of (μ, c) with respect to the product topology (weak \times standard) such that $[\tilde{\mu}]_{\tilde{c}} > \lambda$ for all $(\tilde{\mu}, \tilde{c}) \in \mathcal{U}$.

Proof. Let us first note that:

$$\mu(0) = 1 - c \text{ and } [\mu]_c > 0 \quad \Rightarrow \quad \liminf_{(\tilde{\mu}, \tilde{c}) \rightarrow (\mu, c)} [\tilde{\mu}]_{\tilde{c}} < [\mu]_c. \tag{4.13}$$

Indeed, if $c_n := c + 1/n$, then $(\mu, c_n) \rightarrow (\mu, c)$. By Proposition 2.11, we have that $[\mu]_{c_n} = 0$. Thus,

$$0 = \lim_{n \rightarrow \infty} [\mu]_{c_n} = \liminf_{(\tilde{\mu}, \tilde{c}) \rightarrow (\mu, c)} [\tilde{\mu}]_{\tilde{c}} < [\mu]_c. \tag{4.14}$$

We now consider the converse implication: given $(\mu, c) \in \mathcal{P}(\mathbb{R}_+) \times [0, 1)$ such that $\mu(0) \neq 1 - c$ or $[\mu]_c = 0$, we want to show that $\liminf_{(\tilde{\mu}, \tilde{c}) \rightarrow (\mu, c)} [\tilde{\mu}]_{\tilde{c}} \geq [\mu]_c$. There are some trivial cases:

- If $[\mu]_c = 0$, then the conclusion is obvious.
- If $\mu(0) > 1 - c$, then $[\mu]_c = 0$ by Proposition 2.11, and again the conclusion is clear.

In what follows, we assume that $\mu(0) < 1 - c$ and $[\mu]_c > 0$. Fix a sequence (μ_n, c_n) converging to (μ, c) . We need to prove that:

$$\liminf_{n \rightarrow \infty} [\mu_n]_{c_n} \geq [\mu]_c. \tag{4.15}$$

We may also assume without loss of generality that $[\mu_n]_{c_n} < \infty$ for each n . We divide the proof in two cases:

Case $c > 0$: We can also assume that $c_n > 0$ for every n . By Proposition 2.5, the hypothesis $[\mu_n]_{c_n} < \infty$ means that $\mu_n \in \mathcal{P}_{\text{HS}}(\mathbb{R}_+)$. By Portmanteau’s Theorem [23, Theorem 6.1(c)], $\limsup_{n \rightarrow \infty} \mu_n(0) \leq \mu(0) < 1 - c$. Thus, for sufficiently large values

of n we have $\mu_n(0) < 1 - c_n$. In this setting, the HS barycenter $[\mu_n]_{c_n}$ can be computed by Proposition 2.7. Recall from (2.16) that $\eta(\tilde{\mu}, \tilde{c})$ denotes the unique positive solution of the equation $\int \frac{x}{\tilde{c}x + (1-\tilde{c})\eta} d\tilde{\mu} = 1$. Note that $\eta(\mu, c)$ is well defined even in the case $[\mu]_c = \infty$ (see Proposition 2.8). We claim that:

$$\lim_{n \rightarrow \infty} \eta(\mu_n, c_n) = \eta(\mu, c). \tag{4.16}$$

□

Proof of the claim. Fix numbers α_0, α_1 with $0 < \alpha_0 < \eta(\mu, c) < \alpha_1$. Then a monotonicity property shown in the proof of Proposition 2.7 gives:

$$\int \frac{x}{cx + (1-c)\alpha_0} d\mu > 1 > \int \frac{x}{cx + (1-c)\alpha_1} d\mu. \tag{4.17}$$

Note the uniform bounds:

$$0 \leq \frac{x}{c_n x + (1 - c_n)\alpha_i} \leq (\inf_n c_n)^{-1}. \tag{4.18}$$

So, using Corollary 4.2, we see that $\int \frac{x}{c_n x + (1 - c_n)\alpha_i} d\mu_n \rightarrow \int \frac{x}{cx + (1 - c)\alpha_i} d\mu$. In particular, for all sufficiently large n ,

$$\int \frac{x}{c_n x + (1 - c_n)\alpha_0} d\mu_n > 1 > \int \frac{x}{c_n x + (1 - c_n)\alpha_1} d\mu_n, \tag{4.19}$$

and thus, $\alpha_0 < \eta(\mu_n, c_n) < \alpha_1$, proving the claim (4.16). □

For simplicity, write $y_n := \eta(\mu_n, c_n)$. By Proposition 2.2.(b),

$$K(x, y_n, c_n) \geq K(0, y_n, c_n) = \log y_n + c_n^{-1} \log(1 - c_n) \geq -C \tag{4.20}$$

for some finite C , since $\sup_n c_n < 1$ and $\inf_n y_n > 0$. This allows us to apply Proposition 4.1 and obtain:

$$\liminf_{n \rightarrow \infty} \int K(x, y_n, c_n) d\mu_n \geq \int K(x, y_\infty, c) d\mu, \tag{4.21}$$

where $y_\infty := \lim_{n \rightarrow \infty} y_n = \eta(\mu, c)$. Using formula (2.10), we obtain (4.15). This completes the proof of Theorem 4.5 in the case $c > 0$.

Case $c = 0$: By Proposition 4.1 we obtain, $\liminf \int x d\mu_n \geq \liminf \int x d\mu$, that is, $\liminf [\mu_n]_0 \geq [\mu]_c$. So we can assume that $c_n > 0$ for every n , like in the previous case.

In order to prove (4.15) in the case $c = 0$, let us fix an arbitrary positive $\lambda < [\mu]_0 = \int x d\mu$, and let us show that $[\mu_n]_{c_n} > \lambda$ for every sufficiently large n . By the monotone convergence theorem, there exists $m \in \mathbb{N}$ such that $\int \min(x, m) d\mu(x) > \lambda$. Let $\hat{\mu}$ (respectively $\hat{\mu}_n$) be the pushforward of the measure $\hat{\mu}$ (respectively μ_n) by the map $x \mapsto \min(x, m)$. Then $[\hat{\mu}]_0 > \lambda$ and, by Proposition 2.6.(c), $[\hat{\mu}_n]_{c_n} \leq [\mu_n]_{c_n}$. Furthermore, we have $\hat{\mu}_n \rightarrow \hat{\mu}$, since for every $f \in C_b(\mathbb{R}_+)$,

$$\int f d\hat{\mu}_n = \int f(\min(x, m)) d\mu_n(x) \rightarrow \int f(\min(x, m)) d\mu(x) = \int f d\hat{\mu}. \tag{4.22}$$

So, to simplify the notation, we remove the hats and assume that the measures μ_n, μ are all supported in the interval $[0, m]$.

The numbers $\eta(\mu_n, c_n)$ and $\eta(\mu, c)$ are well defined, as in the previous case, and furthermore they belong to the interval $[0, m]$, by Lemma 2.14. On the other hand, by Proposition 4.1,

$$\liminf_{n \rightarrow \infty} \int \frac{x}{c_n x + (1 - c_n)\lambda} d\mu_n \geq \int \frac{x}{\lambda} d\mu > 1. \tag{4.23}$$

It follows that $\eta(\mu_n, c_n) > \lambda$ for all sufficiently large n . We claim that $\eta(\mu_n, c_n) \rightarrow \eta(\mu, 0)$, as before in (4.16). The proof is the same, except that the upper bound in (4.18) becomes infinite and must be replaced by the following estimate:

$$\left. \begin{matrix} 0 \leq x \leq m \\ \alpha_i \geq \lambda \end{matrix} \right\} \Rightarrow 0 \leq \frac{x}{c_n x + (1 - c_n)\alpha_i} \leq \frac{m}{c_n m + (1 - c_n)\alpha_i} \leq \frac{m}{\lambda}. \tag{4.24}$$

So a repetition of the previous arguments yields (4.16), then (4.20) and (4.21), and finally (4.15). Therefore, Theorem 4.5 has been proved in both cases $c > 0$ and $c = 0$. \square

Next, let us investigate the behaviour of the HS barycenter on the product space $\mathcal{P}_{\text{HS}}(\mathbb{R}_+) \times [0, 1]$, where $\mathcal{P}_{\text{HS}}(\mathbb{R}_+)$ is endowed with the topology defined in the end of § 4.1.

Theorem 4.6. *For every $(\mu, c) \in \mathcal{P}_{\text{HS}}(\mathbb{R}_+) \times [0, 1]$, we have:*

$$\limsup_{(\tilde{\mu}, \tilde{c}) \rightarrow (\mu, c)} [\tilde{\mu}]_{\tilde{c}} \leq [\mu]_c \text{ unless } c = 0 \text{ and } [\mu]_0 < \infty, \tag{4.25}$$

$$\liminf_{(\tilde{\mu}, \tilde{c}) \rightarrow (\mu, c)} [\tilde{\mu}]_{\tilde{c}} \geq [\mu]_c \text{ unless } \mu(0) = 1 - c \text{ and } [\mu]_c > 0. \tag{4.26}$$

Proof of part (4.25) of Proposition 4.6. Let us start by proving the following implication:

$$\mu \in \mathcal{P}_{\text{HS}}(\mathbb{R}_+), c = 0, \text{ and } [\mu]_0 < \infty \Rightarrow \limsup_{(\tilde{\mu}, \tilde{c}) \rightarrow (\mu, c)} [\tilde{\mu}]_{\tilde{c}} > [\mu]_c. \tag{4.27}$$

Consider measures $\mu_n := \frac{n-1}{n} \mu + \frac{1}{n} \delta_n$. Clearly, $\mu_n \rightarrow \mu$; moreover,

$$\int \log(1+x) d(\mu - \mu_n)(x) = \frac{\log(1+n)}{n} \rightarrow 0. \tag{4.28}$$

So using characterization (c) of Theorem 4.3, we conclude that $\mu_n \rightarrow \mu$ in the topology of $\mathcal{P}_{\text{HS}}(\mathbb{R}_+)$. On the other hand, $[\mu_n]_0 = \frac{n-1}{n} [\mu]_0 + 1 \rightarrow [\mu]_0 + 1$. This proves (4.27).

Next, let us prove the converse implication. So, let us fix (μ, c) such that $c \neq 0$ or $[\mu]_0 = \infty$, and let us show that if (μ_n, c_n) is any sequence in $\mathcal{P}_{\text{HS}}(\mathbb{R}_+) \times [0, 1]$ converging to (μ, c) , then $\limsup_{n \rightarrow \infty} [\mu_n]_{c_n} \leq [\mu]_c$. This is obviously true if $[\mu]_c = \infty$, so let us assume that $[\mu]_c < \infty$. Then our assumption becomes $c > 0$, so by removing finitely many terms from the sequence (μ_n, c_n) , we may assume that $\inf_n c_n > 0$. Fix some finite number $\lambda > [\mu]_c$. By Definition 2.3, there is some $y_0 > 0$ such that $\int K(x, y_0, c) d\mu(x) < \log \lambda$. The sequence of continuous functions $\frac{|K(x, y_0, c_n)|}{1 + \log(1+x)}$ is uniformly bounded, as a direct calculation shows. Furthermore, $K(\cdot, y_0, c_n) \rightarrow K(\cdot, y_0, c)$ uniformly on compact subsets of \mathbb{R}_+ . So Lemma 4.4 ensures that $\int K(x, y_0, c_n) d\mu_n(x) \rightarrow \int K(x, y_0, c) d\mu(x)$. Now it

follows from Definition 2.3 that $[\mu_n]_{c_n} < \lambda$ for all sufficiently large n . Since $\lambda > [\mu]_c$ is arbitrary, we conclude that $\limsup_{n \rightarrow \infty} [\mu_n]_{c_n} \leq [\mu]_c$, as we wanted to show. \square

Proof of part (4.26) of Proposition 4.6. First, we prove that, for all $(\mu, c) \in \mathcal{P}_{\text{HS}}(\mathbb{R}_+) \times [0, 1]$,

$$\mu(0) = 1 - c, \text{ and } [\mu]_c > 0 \implies \liminf_{(\tilde{\mu}, \tilde{c}) \rightarrow (\mu, c)} [\tilde{\mu}]_{\tilde{c}} < [\mu]_c. \tag{4.29}$$

Consider measures $\mu_n := \frac{n-1}{n} \mu + \frac{1}{n} \delta_0$. Clearly, $\mu_n \rightarrow \mu$ in the topology of $\mathcal{P}_{\text{HS}}(\mathbb{R}_+)$. By Proposition 2.11.(b), $[\mu_n]_c = 0$. This proves (4.29). For $c \in [0, 1)$ the converse is a direct consequence of Theorem 4.5, since the topology on $\mathcal{P}_{\text{HS}}(\mathbb{R}_+)$ is stronger. If $c = 1$ and $[\mu]_1 = 0$ then the result is obvious. If $[\mu]_1 > 0$ then, as the example above shows, the result does not hold. \square

4.3. Continuity for extremal values of the parameter

Theorem 4.6 shows that the HS barycenter map on $\mathcal{P}_{\text{HS}}(\mathbb{R}_+) \times [0, 1]$ is not continuous at the pairs $(\mu, 0)$ (except if $[\mu]_0 = \infty$), nor at the pairs $(\mu, 1)$ (except if $[\mu]_1 = 0$). Let us see that continuity can be ‘recovered’ if we impose extra integrability conditions and work with stronger topologies.

If we use the standard distance $d(x, y) := |x - y|$ on the half-line $X = \mathbb{R}_+ = [0, +\infty)$, then the resulting 1-Wasserstein space is denoted $\mathcal{P}_1(\mathbb{R}_+)$. On the other hand, using the distance

$$d_g(x, y) := |\log x - \log y| \text{ on the open half-line } X = \mathbb{R}_{++}, \tag{4.30}$$

the corresponding 1-Wasserstein space will be denoted $\mathcal{P}_g(\mathbb{R}_{++})$. We consider the latter space as a subset of $\mathcal{P}(\mathbb{R}_+)$, since any measure μ on \mathbb{R}_{++} can be extended to \mathbb{R}_+ by setting $\mu(0) = 0$. The topologies we have just defined on the spaces $\mathcal{P}_1(\mathbb{R}_+)$ and $\mathcal{P}_g(\mathbb{R}_{++})$ are stronger than the topologies induced by $\mathcal{P}_{\text{HS}}(\mathbb{R}_+)$; in other words, the inclusion maps below are continuous:

$$\begin{array}{ccc} \mathcal{P}_1(\mathbb{R}_+) & \hookrightarrow & \\ & \searrow & \mathcal{P}_{\text{HS}}(\mathbb{R}_+) \hookrightarrow \mathcal{P}(\mathbb{R}_+) \\ & \swarrow & \\ \mathcal{P}_g(\mathbb{R}_{++}) & \hookrightarrow & \end{array} \tag{4.31}$$

Note that the ‘arithmetic barycenter’ $[\cdot]_0$ is finite on $\mathcal{P}_1(\mathbb{R}_+)$, while the ‘geometric barycenter’ $[\cdot]_1$ is finite and non-zero on $\mathcal{P}_g(\mathbb{R}_{++})$.

Finally, let us establish continuity of the HS barycenter for the extremal values of the parameter with respect to these new topologies:

Proposition 4.7. *Consider a sequence (μ_n, c_n) in $\mathcal{P}_{\text{HS}}(\mathbb{R}_+) \times [0, 1]$.*

- (a) *If $(\mu_n, c_n) \rightarrow (\mu, 0)$ in $\mathcal{P}_1(\mathbb{R}_+) \times [0, 1]$, then $[\mu_n]_{c_n} \rightarrow [\mu]_0$.*
- (b) *If $(\mu_n, c_n) \rightarrow (\mu, 1)$ in $\mathcal{P}_g(\mathbb{R}_{++}) \times (0, 1]$, then $[\mu_n]_{c_n} \rightarrow [\mu]_1$.*

Proof of part (a) of Proposition 4.7. Note that if $c_n = 0$ for every $n \in \mathbb{N}$, the result is direct from the definition of the topology in $\mathcal{P}_1(\mathbb{R}_+)$ (use characterization (c) of Theorem 4.3). We assume now that $c_n > 0$, for every $n \in \mathbb{N}$. It is a consequence of Theorem 4.5 that:

$$\liminf_{(\mu_n, c_n) \rightarrow (\mu, 0)} [\mu_n]_{c_n} \geq [\mu]_0. \tag{4.32}$$

The same proof as that of part (4.25) of Proposition 4.6 can be used to prove,

$$\limsup_{(\mu_n, c_n) \rightarrow (\mu, 0)} [\mu_n]_{c_n} \leq [\mu]_0. \tag{4.33}$$

Indeed, in the topology of $\mathcal{P}_1(\mathbb{R}_+)$ the HS kernels $K(x, y, c_n)$ satisfy the assumptions of Lemma 4.4. For this, it suffices to notice that, for any fixed value $y_0 > 0$, the sequence of continuous functions $\frac{|K(x, y_0, c_n)|}{1+x}$ is uniformly bounded, and that $K(x, y_0, c_n) \rightarrow K(x, y_0, 0)$ uniformly on compact subsets of \mathbb{R}_+ . \square

Proof of part (b) of Proposition 4.7. In the case that $c_n = 1$ for every $n \in \mathbb{N}$, the result is direct from the topology in $\mathcal{P}_g(\mathbb{R}_{++})$ (use characterization (c) of Theorem 4.3). We assume that $c_n < 1$, for every n . It is a consequence of (4.25) of Proposition 4.6 that:

$$\limsup_{(\mu_n, c_n) \rightarrow (\mu, 1)} [\mu_n]_{c_n} \leq [\mu]_1. \tag{4.34}$$

Recall that the HS barycenter is decreasing in the variable c ; see Proposition 2.6.(e). In particular, $[\mu_n]_{c_n} \geq [\mu_n]_1$, for every $n \in \mathbb{N}$. Noticing that $\log x$ is a test function for the convergence in the topology of $\mathcal{P}_g(\mathbb{R}_{++})$, we obtain:

$$\begin{aligned} \liminf_{n \rightarrow \infty} [\mu_n]_{c_n} &\geq \liminf_{n \rightarrow \infty} [\mu_n]_1 \\ &= \exp(\liminf_{n \rightarrow \infty} \int \log x \, d\mu_n) = \exp \int \log x \, d\mu = [\mu]_1. \end{aligned} \tag{4.35}$$

The result follows combining (4.34) and (4.35). \square

The following observation complements part (b) of Proposition 4.7, since it provides a sort of lower semicontinuity property at $c = 1$ under a weaker integrability condition:

Lemma 4.8. *Let $c \in [0, 1)$ and let $\mu \in \mathcal{P}(\mathbb{R}_+)$ be such that $\log^-(x) \in L^1(\mu)$. Then:*

$$[\mu]_c \geq \exp \int (c^{-1} \log^+(x) - \log^-(x)) \, d\mu(x). \tag{4.36}$$

Proof. By definition, $[\mu]_c \geq \int K(x, 1, c) \, d\mu(x)$. Note that if $x \geq 1$, then $K(x, 1, c) = c^{-1} \log(cx + 1 - c) \geq c^{-1} \log x$, while if $x < 1$, then $K(x, 1, c) \geq \log x$ by Proposition 2.2.(e). In any case, $K(x, 1, c) \geq c^{-1} \log^+(x) - \log^-(x)$, and the Lemma follows. \square

5. Ergodic theorems for symmetric and HS means

Symmetric means (3.3) are only defined for lists of non-negative numbers. On the other hand, HS barycenters are defined for probability measures on \mathbb{R}_+ , and are therefore much more flexible objects. In particular, there is an induced concept of HS mean of a list of non-negative numbers, which we have already introduced in (3.1). We can also define the HS mean of a function:

Definition 5.1. If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, $f: \Omega \rightarrow \mathbb{R}_+$ is a measurable non-negative function, and $c \in [0, 1]$, then the *Halász–Székely mean* (or *HS mean*) with parameter c of the function f with respect to the probability measure \mathbb{P} is:

$$[f \mid \mathbb{P}]_c := [f_*\mathbb{P}]_c. \tag{5.1}$$

that is, the HS barycenter with parameter c of the pushforward measure on \mathbb{R}_+ . In the case $c = 1$, we require that $\log f$ is semi-integrable.

For arithmetic means, the classical ergodic theorem of Birkhoff states the equality between limit time averages and spatial averages. From the probabilistic viewpoint, Birkhoff’s theorem is the strong law of large numbers. We prove an ergodic theorem that applies simultaneously to symmetric and HS means and extends the results of [15, 28]:

Theorem 5.2. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $T: \Omega \rightarrow \Omega$ be an ergodic measure-preserving transformation, and let $f: \Omega \rightarrow \mathbb{R}_+$ be a non-negative measurable function. Then there exists a measurable set $R \subseteq \Omega$ with $\mathbb{P}(R) = 1$ with the following properties. For any $\omega \in R$, for any $c \in [0, 1]$ such that*

$$c \neq 1 - \mathbb{P}(f^{-1}(0)) \tag{5.2}$$

and

$$c < 1 \quad \text{unless } \log f \text{ is semi-integrable,} \tag{5.3}$$

and for any sequence (c_n) in $[0, 1]$ tending to c , we have:

$$\lim_{n \rightarrow \infty} \text{hsm}_{c_n}(f(\omega), f(T\omega), \dots, f(T^{n-1}\omega)) = [f \mid \mathbb{P}]_c; \tag{5.4}$$

furthermore, for any sequence (k_n) of integers such that $1 \leq k_n \leq n$ and $k_n/n \rightarrow c$, we have:

$$\lim_{n \rightarrow \infty} \text{sym}_{k_n}(f(\omega), f(T\omega), \dots, f(T^{n-1}\omega)) = [f \mid \mathbb{P}]_c. \tag{5.5}$$

Remark 5.3. Since we allow HS means to take infinity value, we do not need integrability conditions as in [15, 28], except for the unavoidable hypothesis (5.3). In the supercritical case $\mathbb{P}(f^{-1}(0)) > 1 - c$, both limits (5.4) and (5.5) are almost surely attained in finite time. In the critical case $\mathbb{P}(f^{-1}(0)) = 1 - c$, strong convergence does not necessarily hold, and the values $\text{sym}_{k_n}(f(\omega), \dots, f(T^{n-1}\omega))$ may oscillate. However, in the IID setting, van Es proved that the sequence of symmetric means converges *in distribution*, provided that the sequence $(\sqrt{n}(k_n/n - c))$ converges in $[-\infty, \infty]$: see [28, Theorem A1 (b)].

As we will soon see, part (5.4) of Theorem 5.2 is obtained using the results about continuity of the HS barycenter with respect to various topologies proved in § 4, and then part (5.5) follows from the inequalities of Theorem 3.4 and Remark 3.6.

To begin the proof, let us fix $(\Omega, \mathcal{F}, \mathbb{P})$, T , and f as in the statement, and let $\mu := f_*\mathbb{P} \in \mathcal{P}(\mathbb{R}_+)$ denote the pushforward measure. Given $\omega \in \Omega$, we consider the sequence of associated *sample measures*:

$$\mu_n^\omega := \frac{\delta_{f(\omega)} + \delta_{f(T(\omega))} + \cdots + \delta_{f(T^{n-1}(\omega))}}{n}. \tag{5.6}$$

As the next result shows, these sample measures converge almost surely^{||} :

Lemma 5.4. *There exists a measurable set $R \subseteq \Omega$ with $\mathbb{P}(R) = 1$ such that for every $\omega \in R$, the corresponding sample measures converge weakly to μ :*

$$\mu_n^\omega \rightharpoonup \mu; \tag{5.7}$$

furthermore, stronger convergences may hold according to the function f :

- (a) if $\log(1 + f) \in L^1(\mathbb{P})$, then $\mu_n^\omega \rightarrow \mu$ in the topology of $\mathcal{P}_{\text{HS}}(\mathbb{R}_+)$;
- (b) if $f \in L^1(\mathbb{P})$, then $\mu_n^\omega \rightarrow \mu$ in the topology of $\mathcal{P}_1(\mathbb{R}_+)$;
- (c) if $|\log f| \in L^1(\mathbb{P})$, then $\mu_n^\omega \rightarrow \mu$ in the topology of $\mathcal{P}_g(\mathbb{R}_{++})$.

Proof. Let $\mathcal{C} \subset C_b(\mathbb{R}_+)$ be a countable set of bounded continuous functions which is sufficient to test weak convergence, i.e., with property (4.1). For each $\phi \in \mathcal{C}$, applying Birkhoff’s ergodic theorem to the function $\phi \circ f$, we obtain a measurable set $R \subseteq \Omega$ with $\mathbb{P}(R) = 1$ such that for all $\omega \in R$,

$$\lim_{n \rightarrow \infty} \int \phi d\mu_n^\omega = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \phi(f(T^j\omega)) = \int \phi \circ f d\mathbb{P} = \int \phi d\mu. \tag{5.8}$$

Since \mathcal{C} is countable, we can choose a single measurable set R of full probability that works for all $\phi \in \mathcal{C}$. Then we obtain $\mu_n^\omega \rightharpoonup \mu$ for all $\omega \in R$. To obtain stronger convergences, we apply Birkhoff’s theorem to the functions $\log(1 + f)$, f , and $|\log f|$, provided they are integrable, and reduce the set R accordingly. If, for example, f is integrable, then for all $\omega \in R$ we have:

$$\lim_{n \rightarrow \infty} \int x d\mu_n^\omega(x) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j\omega) = \int f d\mathbb{P} = \int x d\mu(x). \tag{5.9}$$

Applying part (c) of Theorem 4.3 with $x_0 = 0$ and $d(x, x_0) = x$, we conclude that μ_n^ω converges to μ in the topology of $\mathcal{P}_1(\mathbb{R}_+)$. The assertions about convergence in $\mathcal{P}_{\text{HS}}(\mathbb{R}_+)$ and $\mathcal{P}_g(\mathbb{R}_{++})$ are proved analogously, using instead the corresponding distances (4.11) and (4.30). □

^{||} [23, Theorem II.7.1] contains a similar result, with essentially the same proof.

Proof of Theorem 5.2. Let R be the set given by Lemma 5.4. By the semi-integrable version of Birkhoff’s theorem (see e.g. [19, p. 15]), we can reduce R if necessary and assume that for all $\omega \in R$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \log^\pm(f(T^j \omega)) = \int \log^\pm(f(\omega)) d\mathbb{P}(\omega). \tag{5.10}$$

Fix a point $\omega \in R$ and a number $c \in [0, 1]$ satisfying conditions (5.2) and (5.3). Consider any sequence (c_n) in $[0, 1]$ converging to c . Let us prove (5.5), or equivalently,

$$[\mu_n^\omega]_{c_n} \rightarrow [\mu]_c. \tag{5.11}$$

There are several cases to be considered, and in all but the last case we will use Lemma 5.4:

- First case: $0 \leq c < 1$ and $[\mu]_c = \infty$. Since $\mu_n^\omega \rightarrow \mu$, (5.11) is a consequence of Theorem 4.5.
- Second case: $0 < c < 1$ and $[\mu]_c < \infty$. Then $\log(1 + f) \in L^1(\mathbb{P})$. Therefore, $\mu_n^\omega \rightarrow \mu$ in the topology of $\mathcal{P}_{\text{HS}}(\mathbb{R}_+)$, and Proposition 4.6 implies (5.11).
- Third case: $c = 0$ and $[\mu]_0 < \infty$. Then $\log f \in L^1(\mathbb{P})$, and hence $\mu_n^\omega \rightarrow \mu$ in the topology of $\mathcal{P}_1(\mathbb{R}_+)$. So (5.11) follows from Proposition 4.7.(a).
- Fourth case: $c = 1$ and $[\mu]_1 = 0$. Then $\log(1 + f) \in L^1(\mathbb{P})$. Thus, $\mu_n^\omega \rightarrow \mu$ in the topology of $\mathcal{P}_{\text{HS}}(\mathbb{R}_+)$, and Proposition 4.6 yields (5.11).
- Fifth case: $c = 1$ and $0 < [\mu]_1 < \infty$. Then $|\log f| \in L^1(\mathbb{P})$. Therefore, $\mu_n^\omega \rightarrow \mu$ in the topology of $\mathcal{P}_g(\mathbb{R}_{++})$, and (5.11) becomes a consequence of Proposition 4.7.(b).
- Sixth case: $c = 1$ and $[\mu]_1 = \infty$. Then $\log^-(f)$ is integrable, but $\log^+(f)$ is not. If n is large enough then $c_n > 0$, so Lemma 4.8 gives:

$$\log[\mu_n^\omega]_{c_n} \geq \int (c_n^{-1} \log^+(x) - \log^-(x)) d\mu_n^\omega(x) \tag{5.12}$$

$$= \frac{1}{c_n} \cdot \frac{1}{n} \sum_{i=0}^{n-1} \log^+(f(T^i \omega)) - \frac{1}{n} \sum_{i=0}^{n-1} \log^-(f(T^i \omega)), \tag{5.13}$$

which by (5.10) tends to ∞ . This proves (5.11) in the last case.

Part (5.4) of the theorem is proved, and now lets use it to prove part (5.5). Consider a sequence (k_n) of integers such that $1 \leq k_n \leq n$ and $c_n := k_n/n$ tends to c . By Theorem 3.4,

$$[\mu_n^\omega]_{c_n} \leq \text{sym}_{k_n}(f(\omega), f(T\omega), \dots, f(T^{n-1}\omega)) \leq \frac{\binom{n}{k_n}^{-1/k_n}}{B(c_n)} [\mu_n^\omega]_{c_n}. \tag{5.14}$$

If $[\mu]_c = \infty$, i.e. $[\mu_n^\omega]_{c_n} \rightarrow \infty$, then the first inequality forces the symmetric means to tend to ∞ as well. So let us assume that $[\mu]_c$ is finite. If $c > 0$ then, by Lemma 3.5, the fraction

on the RHS converges to 1 as $n \rightarrow \infty$, and therefore, we obtain the desired limit (5.5). If $c = 0$, then we appeal to Maclaurin inequality in the form

$$\text{sym}_{k_n}(f(\omega), f(T\omega), \dots, f(T^{n-1}\omega)) \leq \text{sym}_1(f(\omega), f(T\omega), \dots, f(T^{n-1}\omega)). \tag{5.15}$$

So:

$$[\mu_n^\omega]_{c_n} \leq \text{sym}_{k_n}(f(\omega), f(T\omega), \dots, f(T^{n-1}\omega)) \leq [\mu_n^\omega]_0. \tag{5.16}$$

Since (5.11) also holds with $c_n \equiv 0$, we see that all three terms converge together to $[\mu]_c$, thus, proving (5.5) also in the case $c = 0$. \square

Like Birkhoff’s Ergodic Theorem itself, Theorem 5.2 should be possible to generalize in numerous directions. For example, part (5.4) can be easily adapted to flows or semiflows (actions of the group \mathbb{R} or the semigroup \mathbb{R}_+). One can also consider actions of amenable groups, like [2, 22]. We shall not pursue these matters. In another direction, let us note that Central Limit Theorems for symmetric means of i.i.d. random variables have been proved by Székely [27] and van Es [28].

A weaker version of Theorem 5.2, in which the function f is assumed to be bounded away from zero and infinity, was obtained in [5, Theorem 5.1] as a corollary of a fairly general pointwise ergodic theorem: the *Law of Large Permanents* [5, Theorem 4.1]. We now briefly discuss a generalization of that result obtained by Balogh and Nguyen [3, Theorem 1.6]. Suppose that T is an ergodic measure-preserving action of the semigroup \mathbb{N}^2 on the space (X, μ) . Given an observable $g : X \rightarrow \mathbb{R}_{++}$ and a point $x \in X$, we define an infinite matrix whose (i, j) -entry is $g(T^{(i,j)}x)$. Consider square truncations of this matrix and then take the limit of the corresponding permanental means as the size of the square tends to infinity. It is proved that the limit exists μ -almost everywhere. But not only that, it is also possible to identify the limit. It turns out that it is a *functional scaling mean*. This is a far reaching generalization of the matrix scaling mean (3.27): see [5, Section 3.1].

Additional information

The interested reader may find in the ArXiv an extended version of this paper, <https://arxiv.org/abs/2103.05182>. It contains two extra sections, 6 and 7. In Section 6, we study concavity properties of the HS barycenter. Among other results,

- we obtain a version of Newton’s inequality for the HS barycenter;
- we prove that as a function of the measure, the HS barycenter is log-concave but not quasi-affine;
- we provide a formula for the Hessian of the function $\log[\cdot]_c$.

In Section 7, we introduce another barycenter which provides the best quasi-affine approximation of the HS barycenter, and is a deviation mean in a sense akin to Daróczy.

After this paper was written, the first-named author found more situations where HS barycenters arise. The *hypergeometric means* introduced by Carlson in [9] include symmetric means and many others as particular cases: see [7, § 3]. In a forthcoming paper,

the first-named author will show that Halász–Székely barycenters can be defined for all real values of the parameter c , and that they govern the limit behaviour of hypergeometric means. For instance, the fixed parameter t in the convergence result [7, Theorem 3] can be replaced by any asymptotically linear sequence $t(n) \sim -cn$, with the corresponding arithmetic means being replaced by Halász–Székely means with parameter c .

Acknowledgements. We thank Juarez Bochi for helping us with computer experiments at a preliminary stage of this project.

The authors were partially supported by CONICYT PIA ACT172001. J.B. was partially supported by Proyecto Fondecyt 1180371. G.I. was partially supported by Proyecto Fondecyt 1190194. M.P. was partially supported by Proyecto Fondecyt 1180922.

References

1. J. ACZÉL AND J. G. DHOMBRES, *Functional equations in several variables*, Encyclopedia of Mathematics and its Applications, Volume 31 (Cambridge University Press, Cambridge, 1989).
2. T. AUSTIN, A CAT(0)-valued pointwise ergodic theorem, *J. Topol. Anal.* **3**(2) (2011), 145–152.
3. J. BALOGH AND H. NGUYEN, A general law of large permanent, *Discrete Contin. Dyn. Syst.* **37**(10) (2017), 5285–5297.
4. C. D. BENNETT, W. C. HOLLAND AND G. J. SZÉKELY, Integer valued means, *Aequationes Math.* **88**(1–2) (2014), 137–149.
5. J. BOCHI, G. IOMMI AND M. PONCE, The scaling mean and a law of large permanents, *Adv. Math.* **292** (2016), 374–409.
6. J. M. BORWEIN AND P. B. BORWEIN, *Pi and the AGM. A study in analytic number theory and computational complexity*. Reprint of the 1987 original. Canadian Mathematical Society Series of Monographs and Advanced Texts, Volume 4 (John Wiley & Sons Inc., New York, 1998).
7. J. L. BRENNER AND B. C. CARLSON, Homogeneous mean values: weights and asymptotics, *J. Math. Anal. Appl.* **123**(1) (1987), 265–280.
8. P. S. BULLEN, *Handbook of means and their inequalities*, Mathematics and its Applications, Volume 560 (Kluwer Academic Publishers Group, Dordrecht, 2003).
9. B. C. CARLSON, A hypergeometric mean value, *Proc. Am. Math. Soc.* **16** (1965), 759–766.
10. F. CELLAROSI, D. HENSLEY, S. J. MILLER AND J. L. WELLENS, Continued fraction digit averages and Maclaurin’s inequalities, *Exp. Math.* **24**(1) (2015), 23–44.
11. M. ÉMERY AND G. MOKOBODZKI, Sur le barycentre d’une probabilité dans une variété. Séminaire de Probabilités, XXV, 220–233, Lecture Notes in Mathematics, Volume 1485 (Springer, Berlin, 1991).
12. W. FELLER, *An introduction to probability theory and its applications*, Volume I, 3rd edn. (John Wiley & Sons Inc., New York-London-Sydney, 1968).
13. N. GUGLIELMI AND M. ZENNARO, An antinorm theory for sets of matrices: bounds and approximations to the lower spectral radius, *Linear Algebra Appl.* **607** (2020), 89–117.
14. L. GURVITS, Van der Waerden/Schrijver-Valiant like conjectures and stable (aka hyperbolic) homogeneous polynomials: one theorem for all. With a corrigendum, *Electron. J. Combin.* **15**(1) (2008), 66.
15. G. HALÁSZ AND G. J. SZÉKELY, On the elementary symmetric polynomials of independent random variables, *Acta Math. Acad. Sci. Hungar.* **28**(3–4) (1976), 397–400.

16. G. H. HARDY, J. E. LITTLEWOOD AND G. PÓLYA, *Inequalities*. Reprint of the 1952 edition. Cambridge Mathematical Library (Cambridge University Press, Cambridge, 1988).
17. T. HEATH, *A history of Greek mathematics. Vol. I. From Thales to Euclid*. Corrected reprint of the 1921 original (Dover Publications Inc., New York, 1981).
18. S. KIM, J. LAWSON AND Y. LIM, Barycentric maps for compactly supported measures, *J. Math. Anal. Appl.* **458**(2) (2018), 1009–1026.
19. U. KRENGEL, *Ergodic theorems*. With a supplement by Antoine Brunel. De Gruyter Studies in Mathematics, Volume 6. (Walter de Gruyter & Co., Berlin, 1985).
20. L. E. MAISTROV, *Probability theory: a historical sketch*. Translated and edited by Samuel Kotz. Probability and Mathematical Statistics, Volume 23 (Academic Press, New York-London, 1974).
21. P. MAJOR, The limit behaviour of elementary symmetric polynomials of i.i.d. random variables when their order tends to infinity, *Ann. Probab.* **27**(4) (1999), 1980–2010. <https://www.jstor.org/stable/2652852>.
22. A. NAVAS, An L^1 ergodic theorem with values in a non-positively curved space via a canonical barycenter map, *Ergodic Theory Dynam. Syst.* **33**(2) (2013), 609–623.
23. K. R. PARTHASARATHY, *Probability measures on metric spaces*. Probability and Mathematical Statistics, Volume 3 (Academic Press, Inc., New York-London, 1967).
24. B. SIMON, *Advanced complex analysis. A Comprehensive Course in Analysis, Part 2B* (Providence, RI, American Mathematical Society, 2015).
25. M. H. STONE, Postulates for the barycentric calculus, *Ann. Mat. Pura Appl.* **29**(4) (1949), 25–30.
26. G. J. SZÉKELY, On the polynomials of independent random variables. *Limit theorems of probability theory* (Colloq., Keszthely, 1974), pp. 365–371. Colloq. Math. Soc. János Bolyai, Volume 11 (North-Holland, Amsterdam, 1975)
27. G. J. SZÉKELY, A limit theorem for elementary symmetric polynomials of independent random variables, *Z. Wahrsch. Verw. Gebiete* **59**(3) (1982), 355–359.
28. B. VAN ES, On the weak limits of elementary symmetric polynomials, *Ann. Probab.* **14**(2) (1986), 677–695. <https://www.jstor.org/stable/2243933>.
29. C. VILLANI, *Topics in Optimal Transportation*. Graduate Studies in Mathematics, Volume 58 (American Mathematical Society, Providence, RI, 2003).
30. X. ZHAN, *Matrix Inequalities*. Lecture Notes in Mathematics, Volume 1790 (Springer-Verlag, Berlin, 2002).