OPERATOR MEANS OF PROBABILITY MEASURES

FUMIO HIAI AND YONGDO LIM

Abstract. Let $\mathbb{P}$ be the complete metric space consisting of positive invertible operators on an infinite-dimensional Hilbert space with the Thompson metric. We introduce the notion of operator means of probability measures on $\mathbb{P}$, in parallel with Kubo and Ando’s definition of two-variable operator means, and show that every operator mean is contractive for the $\infty$-Wasserstein distance. By means of a fixed point method we consider deformation of such operator means, and show that the deformation of any operator mean becomes again an operator mean in our sense. Based on this deformation procedure we prove a number of properties and inequalities for operator means of probability measures.

2010 Mathematics Subject Classification. Primary 47A64; Secondary 47B65, 47L07, 58B20

Key words and phrases. Operator mean, Positive invertible operator, Borel probability measure, Thompson metric, Stochastic order, Deformed operator mean, Arithmetic mean, Harmonic mean, Karcher mean, Power mean, Wasserstein distance, Barycenter, Log-majorization, Minkowski determinant inequality

1. Introduction

A systematic study of two-variable operator means of positive operators on a Hilbert space $\mathcal{H}$ began with the paper of Kubo and Ando [27]. There is one-to-one correspondence between the operator means $\sigma$ in the sense of Kubo-Ando and the positive operator monotone functions $f$ on $[0, \infty)$ with $f(1) = 1$ in such a way that

$$A\sigma B = a^{1/2} f(A^{-1/2}BA^{-1/2})A^{1/2}$$

for positive invertible operators $A, B$ on $\mathcal{H}$. The geometric mean, first introduced by Pusz and Woronowicz [11] and then discussed by Ando [1] in detail, is a two-variable operator mean that has been paid the most attention. It was a long-standing problem to extend the notion of the geometric mean to the case of more than two variables of matrices/operators. A breakthrough came when the definitions of multivariate geometric means of positive definite matrices appeared in the iteration approach by Ando, Li and Mathias [4] and in the Riemannian geometry approach by Moakher [39] and by Bhatia and Holbrook [7]. Since then, the Riemannian multivariate operator means, in particular, the Karcher mean (the generalization of the geometric mean) and the power means, have extensively been developed by many authors, see among others [29, 8, 34]. Furthermore, those multivariate operator means have recently been generalized to probability measures on the positive definite matrices in connection with the Wasserstein distance, see, e.g., [24, 32, 19, 20].
We write $\mathbb{P}$ for the set of positive invertible operators on the Hilbert space $\mathcal{H}$. An important point in the Riemannian geometry approach to operator means when $\dim \mathcal{H} < \infty$ is that $\mathbb{P}$ forms a Riemannian manifold with non-positive curvature (referred to as a global NPC space [43]). Even when $\dim \mathcal{H} = \infty$, $\mathbb{P}$ is a Banach-Finsler manifold with the Thompson metric, although it can no longer have a Riemannian manifold structure. Thus, we can study operator means of probability measures on $\mathbb{P}$ in connection with theory of contractive barycenters with respect to the Wasserstein distance and related stochastic analysis (e.g., ergodic theorems), developed in the framework of complete metric spaces in, e.g., [12, 42, 43, 19]. Moreover, operator means of probability measures, in turn, provide good examples of contractive barycenters.

In recent study of operator means, the fixed point method, apart from the Riemannian geometry method, provides a main technical tool as used in different places in, e.g., [24, 26, 30, 31, 34, 35, 36, 40, 48]. In particular, the Karcher and the power means are defined as the solutions to certain fixed point type equations. In this status of the subject matter, our aim of the present paper is to systematically develop the fixed point method for operator means of Borel probability measures on $\mathbb{P}$ with bounded support. In our approach, we apply the fixed point method based on monotone convergence of Borel probability measures in terms of the strong operator topology, where the stochastic order of probability measures discussed in [18] plays a key role. The idea using monotone convergence is essentially in a similar vein to that of Kubo-Ando’s definition of two-variable operator means. In previous studies of the subject in the fixed point method, a primary tool is the Banach contraction principle, which we never use in the present paper. Thus, the class of operator means of multivariables and probability measures studied in the paper is considerably wider than those in other papers so far.

The paper is organized as follows. We write $\mathcal{P}^\infty(\mathbb{P})$ for the set of Borel probability measures on $\mathbb{P}$ with bounded support (of full measure). In Section 2 we first fix the notion of monotone convergence for a sequence of probability measures in $\mathcal{P}^\infty(\mathbb{P})$ (see Definition 2.3), which plays a primary role in our study as mentioned above. We then give the definition of operator means (see Definition 2.5)

$$M : \mathcal{P}^\infty(\mathbb{P}) \rightarrow \mathbb{P}$$

in parallel with Kubo-Ando’s definition of two-variable operator means, where one important requirement is the monotone continuity that if $\mu_k \nearrow \mu$ or $\mu_k \searrow \mu$ in $\mathcal{P}^\infty(\mathbb{P})$, then $M(\mu_k) \rightarrow M(\mu)$ in the strong operator topology. It is also shown here that every operator means on $\mathcal{P}^\infty(\mathbb{P})$ automatically has the contractivity for the $\infty$-Wasserstein distance with respect to the Thompson metric.

In Section 3 we present the main theorem (Theorem 3.1) of the paper. For an operator mean $M$ on $\mathcal{P}^\infty(\mathbb{P})$ and a two-variable operator mean $\sigma$ (≠ the left trivial mean) and for any $\mu \in \mathcal{P}^\infty(\mathbb{P})$, the theorem says that the fixed point type equation

$$X = M(X \sigma \mu) \quad \text{for } X \in \mathbb{P},$$

where $X \sigma \mu$ is the push-forward of $\mu$ by the map $A \in \mathbb{P} \mapsto X \sigma A \in \mathbb{P}$, has a unique solution, and if we denote the solution by $M_\sigma(\mu)$, then $M_\sigma$ is an operator mean on
Again, we call $M_\sigma$ the deformed operator mean from $M$ by $\sigma$. The notion of deformed operator means is considered in some sense as an extended version of the generalized operator means by Pálfia [25] (see Remark 3.2). The deformation procedure $M \to M_\sigma$ has the order property that
\[
X \leq M(X^{\sigma\mu}) \implies X \leq M_\sigma(\mu), \quad X \geq M(X^{\sigma\mu}) \implies X \geq M_\sigma(\mu). \tag{1.1}
\]
This property is quite useful in later discussions. In Section 4 we prove that several basic properties such as congruence invariance and concavity are preserved under the procedure $M \to M_\sigma$.

In Section 5 we show that all of the arithmetic, the harmonic, the Karcher, and the power means are operator means on $\mathcal{P}^\infty(\mathbb{P})$ in our sense. By starting from those familiar means and by taking deformed operator means again and again, we have a rich class of operator means on $\mathcal{P}^\infty(\mathbb{P})$. By applying the property in (1.1) we can show many inequalities for operator means on $\mathcal{P}^\infty(\mathbb{P})$. For instance, in Section 6, the inequality under positive linear maps and Ando-Hiai type inequalities are obtained for certain classes of operator means on $\mathcal{P}^\infty(\mathbb{P})$. Furthermore in Section 7, when $\mathcal{H}$ is finite-dimensional, a certain norm inequality, eigenvalue majorizations and the Minkowski determinant inequality are obtained.

2. Definitions

Let $\mathcal{H}$ be a separable Hilbert space and $B(\mathcal{H})$ be the algebra of all bounded linear operators on $\mathcal{H}$. Let $B(\mathcal{H})^+$ be the set of positive (not necessarily invertible) operators in $B(\mathcal{H})$, and $\mathbb{P} = \mathbb{P}(\mathcal{H})$ be the set of positive invertible operators in $B(\mathcal{H})$. For self-adjoint $A,B \in B(\mathcal{H})$, $A \preceq B$ means that $B - A \in B(\mathcal{H})^+$. The Thompson metric $d_T$ on $\mathbb{P}$ is defined by
\[
d_T(A,B) := \log \max \{ M(A/B), M(B/A) \} = \| \log A^{-1/2}BA^{-1/2} \|, \quad A,B \in \mathbb{P},
\]
where $\| \cdot \|$ denotes the operator norm on $B(\mathcal{H})$ and $M(A/B) := \inf \{ \alpha > 0 : A \preceq \alpha B \}$. The $d_T$-topology is equivalent to the operator norm topology on $\mathbb{P}$, and $(\mathbb{P}, d_T)$ becomes a complete metric space, see [14]. On the other hand, the strong operator topology is denoted by $SOT$.

Let $\mathcal{P}(\mathbb{P})$ be the set of Borel probability measures $\mu$ on $\mathbb{P}$ with full support, i.e., $\mu(\text{supp}(\mu)) = 1$, where $\text{supp}(\mu)$ denotes the support of $\mu$. We denote by $\mathcal{P}^\infty(\mathbb{P})$ the set of $\mu \in \mathcal{P}(\mathbb{P})$ whose support is bounded in the sense that the support of $\mu$ is included in
\[
\Sigma_\varepsilon := \{ A \in \mathbb{P} : \varepsilon I \leq A \leq \varepsilon^{-1}I \}
\]
for some $\varepsilon \in (0,1)$. For any $\varepsilon \in (0,1)$ the subset $\Sigma_\varepsilon$ of $\mathbb{P}$ with SOT is metrizable by the metric
\[
d_\varepsilon(A,B) := \sum_{n=1}^{\infty} \frac{1}{2^n} \| (A - B)x_n \|, \quad A,B \in \Sigma_\varepsilon, \tag{2.1}
\]
where $\{ x_n \}_{n=1}^{\infty}$ is dense in $\{ x \in \mathcal{H} : \| x \| \leq 1 \}$, see, e.g., [11, p. 262]. As easily verified, the above metric is complete on $\Sigma_\varepsilon$ so that $(\Sigma_\varepsilon, d_\varepsilon)$ becomes a Polish space.
Definition 2.1. A set $\mathcal{U} \subset \mathbb{P}$ is said to be an upper set if $B \in \mathbb{P}$ and $B \geq A$ for some $A \in \mathcal{U}$ then $B \in \mathcal{U}$. Also, a set $\mathcal{L} \subset \mathbb{P}$ is a lower set if $B \in \mathbb{P}$ and $B \leq A$ for some $A \in \mathcal{L}$ then $B \in \mathcal{L}$. For $\mu, \nu \in \mathcal{P}(\mathbb{P})$ we write $\mu \leq \nu$ if $\mu(\mathcal{U}) \leq \nu(\mathcal{U})$ for every upper closed set $\mathcal{U}$, or equivalently, if $\mu(\mathcal{L}) \geq \nu(\mathcal{L})$ for every lower closed set $\mathcal{L}$. It is known [18, Propositions 3.6 and 3.11] that $\mu \leq \nu$ if and only if $\int_{\mathbb{P}} f(A) \, d\mu(A) \leq \int_{\mathbb{P}} f(A) \, d\nu(A)$ for any monotone (bounded) Borel function $f : \mathbb{P} \rightarrow \mathbb{R}^+ := [0, \infty)$, or equivalently, for any monotone (bounded) continuous (in the operator norm) function $f : \mathbb{P} \rightarrow \mathbb{R}^+$. Here, $f$ is monotone if $A \leq B \in \mathbb{P}$ implies $f(A) \leq f(B)$.

Lemma 2.2. Assume that $\mu_1, \mu_2 \in \mathcal{P}^\infty(\mathbb{P})$ and $\mu_1 \leq \mu_2$. Then there exists an $\varepsilon \in (0, 1)$ such that all $\mu \in \mathcal{P}^\infty(\mathbb{P})$ with $\mu_1 \leq \mu \leq \mu_2$ are supported on $\Sigma_\varepsilon$.

Proof. Choose an $\varepsilon > 0$ such that $\mu_1, \mu_2$ are supported on $\Sigma_\varepsilon$. Let $\mu \in \mathcal{P}^\infty(\mathbb{P})$ be such that $\mu_1 \leq \mu \leq \mu_2$. Since $\{A \in \mathbb{P} : A \geq \varepsilon I\}$ is an upper closed set, $\mu(A \geq \varepsilon I) \geq \mu_1(A \geq \varepsilon I) = 1$. Since $\{A \in \mathbb{P} : A \leq \varepsilon^{-1} I\}$ is a lower closed set, $\mu(A \leq \varepsilon^{-1} I) \geq \mu_2(A \leq \varepsilon^{-1} I) = 1$. Hence $\mu(\varepsilon I \leq A \leq \varepsilon^{-1} I) = 1$, so $\mu$ is supported on $\Sigma_\varepsilon$. □

In this paper, the next notion of monotone convergence for a sequence of probability measures in $\mathcal{P}^\infty(\mathbb{P})$ will play an important role.

Definition 2.3. For $\mu, \mu_k \in \mathcal{P}^\infty(\mathbb{P})$ ($k \in \mathbb{N}$) we write

$$\mu_k \nearrow \mu \quad \text{(resp., } \mu_k \searrow \mu)$$

if the following conditions are satisfied:

(a) $\mu_1 \leq \mu_2 \leq \cdots \leq \mu$ (resp., $\mu_1 \geq \mu_2 \geq \cdots \geq \mu$) in the sense of Definition 2.1. In this case, since $\mu_1 \leq \mu_k \leq \mu$ (resp., $\mu_1 \geq \mu_k \geq \mu$) for all $k$, by Lemma 2.2 there is an $\varepsilon \in (0, 1)$ such that $\mu$ and $\mu_k$ are all supported on $\Sigma_\varepsilon$.

(b) For any bounded SOT-continuous real function $f$ on $\Sigma_\varepsilon$ where $\varepsilon$ is chosen in (a),

$$\int_{\mathbb{P}} f(A) \, d\mu_k(A) \rightarrow \int_{\mathbb{P}} f(A) \, d\mu(A) \quad \text{as } k \rightarrow \infty.$$  

Note that condition (b) is independent of the choice of $\varepsilon \in (0, 1)$ in (a); in fact, when $0 < \varepsilon' < \varepsilon$, any bounded SOT-continuous real function on $\Sigma_\varepsilon$ can extend to a bounded SOT-continuous function on $\Sigma_{\varepsilon'}$. The convergence $\mu_k \rightarrow \mu$ in (b) reduces to the usual weak convergence as Borel probability measures on the Polish space $(\Sigma_\varepsilon, d_\varepsilon)$ with $d_\varepsilon$ in (2.1).

One can define a variant of monotone convergence for probability measures in $\mathcal{P}^\infty(\mathbb{P})$ by replacing the SOT-continuity for $f$ with operator norm continuity. When $\mathcal{H}$ is infinite-dimensional, the monotone convergence for probability measures in the SOT sense is strictly weaker than that in the norm sense; in fact, for point measures $\delta_A, \delta_{A_k}$ with $A, A_k \in \mathbb{P}$, $\delta_{A_k} \nearrow \delta_A$ in Definition 2.3 means that $A_k \rightarrow A$ in SOT, while $\delta_{A_k} \nearrow \delta_A$ in the norm sense implies that $A_k \rightarrow A$ in the operator norm. The monotone convergence in the SOT sense adopted in Definition 2.3 is essential in this paper.
Remark 2.4. The assumption of the Hilbert space $\mathcal{H}$ being separable is not essential in the paper. Indeed, when $\mathcal{H}$ is a general Hilbert space, it is known \[28\] Lemma 2.1 that any probability measure on $(\mathbb{P}, \mathcal{B}(\mathbb{P}))$ has the separable support. All of our results in the paper are concerned with an at most countable set $\{\mu_k\}$ in $\mathcal{P}^\infty(\mathbb{P})$. For such $\mu_k$'s there exists a separable closed subset $\mathcal{X}$ of $\mathbb{P}$ such that all $\mu_k$ are supported on $\mathcal{X}$. The $C^*$-algebra generated by $\mathcal{X}$ is faithfully represented on a separable Hilbert space $\mathcal{H}_0$, so that we may regard $\mu_k$'s as probability measures on $\mathbb{P}(\mathcal{H}_0)$. Thus we can reduce all our arguments to the separable Hilbert space case.

The notion of two-variable operator means was introduced by Kubo and Ando \[27\] in an axiomatic way. A map $\sigma : \mathcal{B}(\mathcal{H})^+ \times \mathcal{B}(\mathcal{H})^+ \rightarrow \mathcal{B}(\mathcal{H})^+$ is called an operator mean if it satisfies the following properties:

(I) **Monotonicity:** $A \leq C$, $B \leq D \Rightarrow A\sigma B \leq C\sigma D$.

(II) **Transformer inequality:** $C(A\sigma B)C \leq (CAC)\sigma(CBC)$ for every $C \in \mathcal{B}(\mathcal{H})^+$.

(III) **Downward continuity in SOT:** $A_k \searrow A$, $B_k \searrow B \Rightarrow A_k\sigma B_k \searrow A\sigma B$.

(IV) **Normalized condition:** $I\sigma I = I$.

Each operator mean $\sigma$ is associated with a positive operator monotone function $f$ on $(0, \infty)$ with $f(1) = 1$ in such a way that

$$A\sigma B = A^{1/2}f(A^{-1/2}BA^{-1/2})A^{1/2}, \quad A, B \in \mathbb{P},$$

which extends to general $A, B \in \mathcal{B}(\mathcal{H})^+$ as $A\sigma B = \lim_{\varepsilon \searrow 0}(A + \varepsilon I)\sigma(B + \varepsilon I)$ in SOT. Here, a function $f$ on $(0, \infty)$ is operator monotone if $A \leq B \Rightarrow f(A) \leq f(B)$ for $A, B \in \mathbb{P}$. The above operator monotone function $f$ on $(0, \infty)$ corresponding to $\sigma$ is denoted by $f_\sigma$ and called the representing function of $\sigma$. Note that $f_\sigma$ is analytic on $(0, \infty)$ with $f'_\sigma(1) \in [0, 1]$ and $f'_\sigma(1) = 0$ only when $f_\sigma \equiv 1$. A concise exposition on two-variable operator means is found in \[16\] Chapter 3.

In this paper we shall consider a certain extension of operator means of two variables to those of probability measures in $\mathcal{P}^\infty(\mathbb{P})$. The following is the definition of such operator means of probability measures with minimally required properties, while we shall add further properties accordingly when needed.

**Definition 2.5.** We say that a map

$$M : \mathcal{P}^\infty(\mathbb{P}) \rightarrow \mathbb{P}$$

is an operator mean on $\mathcal{P}^\infty(\mathbb{P})$ if it satisfies the following properties:

(i) **Monotonicity:** If $\mu, \nu \in \mathcal{P}^\infty(\mathbb{P})$ and $\mu \leq \nu$ in the sense of Definition 2.1 then $M(\mu) \leq M(\nu)$.

(ii) **Positive homogeneity:** $M(\alpha \cdot \mu) = \alpha M(\mu)$ for every $\mu \in \mathcal{P}^\infty(\mathbb{P})$ and $\alpha > 0$, where $\alpha \cdot \mu$ is the push-forward of $\mu$ by the map $A \in \mathbb{P} \mapsto \alpha A$.

(iii) **Monotone continuity:** For $\mu, \mu_k \in \mathcal{P}^\infty(\mathbb{P})$ ($k \in \mathbb{N}$), if either $\mu_k \searrow \mu$ or $\mu_k \nearrow \mu$ (in the sense of Definition 2.3), then $M(\mu_k) \rightarrow M(\mu)$ in SOT.

(iv) **Normalized condition:** $M(\delta_1) = I$.
Resemblances of properties (i)–(iv) to (I)–(IV) for Kubo-Ando’s two-variable operator means are apparent, but there are also slight differences between those. For one thing, operator means in Definition 2.5 are maps on $\mathcal{P}^\infty(\mathbb{P})$ while two-variable operator means are on $B(\mathcal{H})^+ \times B(\mathcal{H})^+$ permitting non-invertible operators. For another, (ii) is weaker than (II). Here we note [27] that congruence invariance $S^*(A\sigma B)S = (S^*AS)\sigma(S^*BS)$ for invertible $S \in B(\mathcal{H})$ is automatic when $\sigma$ is a two-variable operator mean. Moreover, we assume continuity both downward and upward in (iii) while only downward is assumed in (III). Continuity from both directions seems natural when we take care of transformation under $A \in \mathbb{P} \mapsto A^{-1} \in \mathbb{P}$ for operator means on $\mathcal{P}^\infty(\mathbb{P})$. Note also that $\sigma$ is upward continuous when restricted to $\mathbb{P} \times \mathbb{P}$, while it is not necessarily SOT-continuous on $\mathbb{P} \times \mathbb{P}$. Since these facts do not seem widely known, we supply the details in Appendix A for the convenience of the reader.

In the rest of the section, we will show that every operator mean on $\mathcal{P}^\infty(\mathbb{P})$ is contractive for the $\infty$-Wasserstein distance on $\mathcal{P}^\infty(\mathbb{P})$. To do so, we first recall some relevant notions in the setting of a general complete metric space $(X,d)$. Let $\mathcal{P}(X)$ be the set of Borel probability measures $\mu$ on $X$ with full support. For $1 \leq p < \infty$ let $\mathcal{P}^p(X)$ be the set of $\mu \in \mathcal{P}(X)$ with finite $p$th moment, i.e., $\int_X d^p(x,y) \, d\mu(x) < \infty$ for some (hence for all) $y \in X$. Moreover, let $\mathcal{P}^\infty(X)$ be the set of $\mu \in \mathcal{P}(X)$ with bounded support, i.e., $\text{supp}(\mu) \subseteq \{x \in X : d(x,y) \leq R\}$ for some $y \in X$ and some $R > 0$. Obviously,

$$\mathcal{P}^\infty(X) \subseteq \mathcal{P}^q(X) \subseteq \mathcal{P}^p(X) \subseteq \mathcal{P}^1(X) \quad (1 < p < q < \infty).$$

For $1 \leq p < \infty$, the $p$-Wasserstein distance $d^W_p$ on the set $\mathcal{P}^p(X)$ is defined as

$$d^W_p(\mu_1,\mu_2) := \left[ \inf_{\pi \in \Pi(\mu_1,\mu_2)} \int_{X \times X} d^p(x,y) \, d\pi(x,y) \right]^{1/p}, \quad \mu_1,\mu_2 \in \mathcal{P}^p(\mathbb{P}),$$

(2.2)

where $\Pi(\mu_1,\mu_2)$ is the set of all couplings for $\mu_1,\mu_2$ (i.e., Borel probability measures on $\mathbb{P} \times \mathbb{P}$ whose marginals are $\mu_1,\mu_2$). It is well-known that $d^W_p$ is a complete metric on $\mathcal{P}^p(X)$. See, e.g., [43, 45] for more details on $d^W_p$. The $\infty$-version of $d^W_p$ is the $\infty$-Wasserstein distance $d^W_\infty$ on $\mathcal{P}^\infty(X)$ defined as

$$d^W_\infty(\mu_1,\mu_2) := \inf_{\pi \in \Pi(\mu_1,\mu_2)} \sup \{d(x,y) : (x,y) \in \text{supp} (\pi)\}, \quad \mu_1,\mu_2 \in \mathcal{P}^\infty(X).$$

It is also known that $d^W_\infty$ is a complete metric on $\mathcal{P}^\infty(X)$ and for every $\mu,\nu \in \mathcal{P}^\infty(X)$,

$$d^W_\infty(\mu,\nu) = \lim_{p \to \infty} d^W_p(\mu,\nu) \quad \text{increasingly.}$$

To prove these facts on $d^W_\infty$, one can assume that $(X,d)$ is a Polish space; then the proof is found in [9, Theorem 2.8].

**Definition 2.6.** Let $1 \leq p \leq \infty$. A map $\beta : \mathcal{P}^p(X) \to X$ is called a barycentric map or a barycenter on $\mathcal{P}^p(X)$ if $\beta(\delta_x) = x$ for all $x \in X$. We say that the map $\beta$ is $d^W_p$-contractive if

$$d(\beta(\mu),\beta(\nu)) \leq d^W_p(\mu,\nu)$$

for all $\mu,\nu \in \mathcal{P}^p(X)$. 
Next, we consider a more specialized situation of an ordered metric space with the Thompson metric. Let $E$ be a Banach space including an open convex cone $C$ such that its closure $\overline{C}$ is a proper cone, i.e., $\overline{C} \cap (-\overline{C}) = \{0\}$. The cone $\overline{C}$ defines a closed partial order on $E$ by $x \leq y$ if $y - x \in \overline{C}$. The cone $\overline{C}$ is said to be normal if there is a constant $K$ such that $0 \leq x \leq y$ implies $\|x\| \leq K\|y\|$. A typical case of $(E, C)$ is $(B(\mathcal{H}), \mathbb{P}(\mathcal{H}))$, which is our setting of this paper. The Thompson metric $d_T$ on $C$ is defined by

$$d_T(x, y) := \log \max\{M(x/y), M(y/x)\} = \min\{r \geq 0 : e^{-r}y \leq x \leq e^r y\},$$

where $M(x/y) := \inf\{\alpha > 0 : x \leq \alpha y\}$. As is well-known \[18\], $d_T$ is a complete metric on $C$ and the $d_T$-topology agrees with the relative norm topology on $C$. For $\mu, \nu \in \mathcal{P}(C)$, the stochastic order $\mu \leq \nu$ is defined as in Definition 2.1. Note \[18\] Theorem 4.3] that $\mu \leq \nu$ is a partial order on $\mathcal{P}(C)$. For every $\mu, \nu \in \mathcal{P}^\infty(C)$, there is an $\alpha \in [1, \infty)$ such that $\alpha^{-1}\nu \leq \mu \leq \alpha\mu$, where $\alpha\nu$ is the push-forward of $\nu$ by the map $x \in C \mapsto \alpha x \in C$. Hence one can define the Thompson metric-like function as

$$\delta_T(\mu, \nu) := \inf\{r \geq 0 : e^{-r}\nu \leq \mu \leq e^r \nu\}, \quad \mu, \nu \in \mathcal{P}^\infty(C).$$

**Proposition 2.7.** $\delta_T(\mu, \nu)$ is a metric on $\mathcal{P}^\infty(C)$ and for every $\mu, \nu \in \mathcal{P}^\infty(C),$

$$\delta_T(\mu, \nu) \leq d_{\infty}^W(\mu, \nu). \tag{2.3}$$

**Proof.** Let $\mu, \nu, \lambda \in \mathcal{P}^\infty(C)$. It is obvious that $\delta_T \geq 0$ and $\delta_T(\mu, \nu) = \delta_T(\nu, \mu)$. If $\delta_T(\mu, \nu) = 0$, then $\nu \leq \mu \leq \nu$, which implies $\mu = \nu$ by \[18\] Theorem 4.3]. To prove the triangle inequality, let $r := \delta_T(\mu, \nu)$ and $t := \delta_T(\nu, \lambda)$. Since $e^{-r}\nu \leq \mu \leq e^r \nu$ and $e^{-t}\lambda \leq \nu \leq e^t \lambda$, we have $\mu \leq e^{-r}(e^t \lambda) = e^{r+t} \lambda$ and $\mu \geq e^{-r}(e^{-t} \lambda) = e^{-(r+t)} \lambda$, so $\delta_T(\mu, \lambda) \leq r + t$.

Next, we prove inequality (2.3). For any $\rho > d_{\infty}^W(\mu, \nu)$, one can choose a $\pi \in \Pi(\mu, \nu)$ such that $\sup\{d_T(x, y) : (x, y) \in \text{supp}(\pi)\} < \rho$. We prove that $e^{-\rho}\nu \leq \mu \leq e^\rho \nu$. For every upper closed set $U$ in $C$, note that

$$\mu(U) = \pi((U \times C) \cap \text{supp}(\pi)),
\quad (e^\rho \nu)(U) = \nu(e^{-\rho}U) = \pi(C \times (e^{-\rho}U)).$$

Assume that $(x, y) \in (U \times C) \cap \text{supp}(\pi)$. Since $x \in U$ and $d_T(x, y) < \rho$ so that $x \leq e^{\rho} y$, we have $e^{\rho} y \in U$, implying $(x, y) \in C \times (e^{-\rho} U)$. Hence $(U \times C) \cap \text{supp}(\pi) \subset C \times (e^{-\rho} U)$, which implies that $\mu(U) \leq (e^\rho \nu)(U)$. This means that $\mu \leq e^\rho \nu$, and similarly $\nu \leq e^\rho \mu$. Therefore, we have $\delta_T(\mu, \nu) \leq \rho$, giving (2.3). \qed

Assume that a map $\beta : \mathcal{P}^\infty(C) \to C$ satisfies monotonicity and positive homogeneity as in (i) and (ii) of Definition 2.5. If $\mu, \nu \in \mathcal{P}^\infty(C)$ and $e^{-r}\nu \leq \mu \leq e^r \nu$ with $r \geq 0$, then $e^{-r}\beta(\nu) \leq \beta(\mu) \leq e^r \beta(\nu)$ so that $d_T(\beta(\nu) \beta(\mu)) \leq r$. Hence $d_T(\beta(\mu), \beta(\nu)) \leq \delta_T(\mu, \nu).$ From this and Proposition 2.7, we have

**Theorem 2.8.** An operator mean $M$ on $\mathcal{P}^\infty(\mathbb{P})$ is $(d_T)^W$-contractive; in fact, for every $\mu, \nu \in \mathcal{P}^\infty(\mathbb{P})$,

$$d_T(M(\mu), M(\nu)) \leq \delta_T(\mu, \nu) \leq (d_T)^W(\mu, \nu).$$
3. Deformed operator means

Throughout the section, let $M$ be an operator mean on $\mathcal{P}^\infty(\mathbb{P})$ as introduced in Definition 2.3. For any two-variable operator mean $\sigma$ (in the Kubo-Ando sense) and any $\mu \in \mathcal{P}^\infty(\mathbb{P})$, we consider the fixed point type equation

$$X = M(X\sigma\mu), \quad X \in \mathbb{P},$$

where $X\sigma\mu$ is the push-forward of $\mu$ by the map $A \in \mathbb{P} \mapsto X\sigma A \in \mathbb{P}$. Note that it is easily seen that if $\mu \in \mathcal{P}^\infty(\mathbb{P})$ and $X \in \mathbb{P}$, then $X\sigma\mu \in \mathcal{P}^\infty(\mathbb{P})$, so the above equation makes sense. In the rest of this section we shall prove the following:

**Theorem 3.1.** Assume that $\sigma \neq I$, where $I$ is the left trivial two-variable operator mean, i.e., $X Y = X$ for all $X, Y \in \mathbb{P}$.

1. For every $\mu \in \mathcal{P}^\infty(\mathbb{P})$ there exists a unique $X_0 \in \mathbb{P}$ satisfying (3.1).
2. If $Y \in \mathbb{P}$ satisfies $Y \geq M(Y\sigma\mu)$, then $Y \geq X_0$, and if $Y' \in \mathbb{P}$ satisfies $Y' \leq M(Y'\sigma\mu)$, then $Y' \leq X_0$.
3. Write $M_\sigma(\mu)$ for the solution $X_0$. Then the map $M_\sigma : \mathcal{P}^\infty(\mathbb{P}) \rightarrow \mathbb{P}$ satisfies (i)–(iv), that is, $M_\sigma$ is an operator mean on $\mathcal{P}^\infty(\mathbb{P})$ again.

We call $M_\sigma : \mathcal{P}^\infty(\mathbb{P}) \rightarrow \mathbb{P}$ given in the theorem the deformed operator mean from $M$ by $\sigma$.

**Remark 3.2.** It is easy to verify that the arithmetic mean $A(\mu) := \int_\mathbb{P} A d\mu(A)$ on $\mathcal{P}^\infty(\mathbb{P})$ satisfies (i)–(iv). When $M = A$, equation (3.1) is equivalently written as

$$X = \int_\mathbb{P} X\sigma A d\mu(A), \quad \text{i.e.,} \quad I = \int_\mathbb{P} f_\sigma(X^{-1/2}AX^{-1/2}) d\mu(A),$$

where $f_\sigma$ is the representing function of $\sigma$ so that $X\sigma A = X^{1/2}f_\sigma(X^{-1/2}AX^{-1/2})X^{1/2}$ for $X, A \in \mathbb{P}$. Here we remark that the functions $A \in \mathbb{P} \mapsto X\sigma A, f_\sigma(X^{-1/2}AX^{-1/2}) \in \mathbb{P}$ are continuous in the operator norm and are operator norm bounded on the support of $\mu$. Hence the above integrals $\int_\mathbb{P} A d\mu(A), \int_\mathbb{P} X\sigma A d\mu(A)$ and $\int_\mathbb{P} f_\sigma(X^{-1/2}AX^{-1/2}) d\mu(A)$ are well defined as Bochner integrals.

Assume that $\sigma \neq I$ so that $f_\sigma'(1) > 0$, and set $g_\sigma(x) := (f_\sigma(x) - 1)/f_\sigma'(1)$. Then $g_\sigma$ is an operator monotone function with $g_\sigma(1) = 0$ and $g_\sigma'(1) = 1$, and (3.1) or (3.2) is equivalent to

$$\int_\mathbb{P} g_\sigma(X^{-1/2}AX^{-1/2}) d\mu(A) = 0.$$

Hence, equation (3.1) in this case reduces to the generalized Karcher equation in [40, Definition 2.2]. The method in [40] is relied on the Banach contraction principle, while our proof of the theorem will be done by a simple argument of monotone convergence. Hence our proof is applicable to any operator mean $M$ satisfying (i)–(iv).

**Remark 3.3.** As discussed in [17, 22] the deformed operator means can be also considered in the restricted setting of two-variable operator means (in the Kubo-Ando sense).
and in that of \( n \)-variable operator means. As for the two-variable case, when \( M = \tau \) is a two-variable operator mean, the reduced equation of (3.1) is, for \( \sigma \neq I \) and \( A, B \in \mathbb{P} \),
\[
X = (X \sigma A) \tau (X \sigma B), \quad X \in \mathbb{P}.
\]
which has a unique solution \( X_0 \in \mathbb{P} \) as in Theorem 3.1. If we write \( A \tau_\sigma B \) for the solution \( X_0 \), then \( \tau_\sigma \) becomes a two-variable operator mean again and the representing function of \( \tau_\sigma \) is exactly determined by those of \( \tau \) and \( \sigma \), see [17]. The restriction to two-variable operator means was also discussed in [10, 49] for the generalized Karcher equation mentioned in Remark 3.2.

The following proof of the theorem is essentially the same as that of [22] Theorem 2.1] (also [17]), where a similar theorem was shown for multivariate operator means \( M : \mathbb{P}^n \to \mathbb{P} \). Now, let \( M \) and \( \sigma \) be as in the theorem. We first give some lemmas.

**Lemma 3.4.** Let \( \varphi, \psi : \mathbb{P} \to \mathbb{P} \) be monotone and Borel measurable. Assume that \( \varphi(A) \leq \psi(A) \) for all \( A \in \mathbb{P} \). If \( \mu, \nu \in \mathcal{P}^\infty(\mathbb{P}) \) and \( \mu \leq \nu \), then \( \varphi_\ast \mu, \psi_\ast \nu \in \mathcal{P}^\infty(\mathbb{P}) \) and \( \varphi_\ast \mu \leq \psi_\ast \nu \).

**Proof.** Let \( \varphi, \psi \) and \( \mu, \nu \) be as stated. That \( \varphi_\ast \mu, \psi_\ast \nu \in \mathcal{P}^\infty(\mathbb{P}) \) follows immediately from \( \varphi, \psi \) being monotone. For any monotone bounded Borel function \( f : \mathbb{P} \to \mathbb{R}^+ \), \( f(\varphi(A)) \leq f(\psi(A)) \) for all \( A \in \mathbb{P} \) and \( f \circ \psi \) is monotone. Hence it follows from [18, Proposition 3.6] that
\[
\int_\mathbb{P} f(A) d(\varphi_\ast \mu)(A) = \int_\mathbb{P} f(\varphi(A)) d\mu(A) \leq \int_\mathbb{P} f(\psi(A)) d\mu(A) \leq \int_\mathbb{P} f(\psi(A)) d\nu(A) = \int_\mathbb{P} f(A) d(\psi_\ast \nu)(A),
\]
which gives \( \varphi_\ast \mu \leq \psi_\ast \nu \) by [18, Proposition 3.6] again. \( \square \)

**Lemma 3.5.** Let \( \sigma \) be arbitrary.

1. Let \( X, X_k \in \mathbb{P} \) \( (k \in \mathbb{N}) \), and assume that \( X_1 \geq X_2 \geq \cdots \) \( \text{resp.,} \ X_1 \leq X_2 \leq \cdots \) and \( X_k \to X \) in SOT. Then for every \( \mu \in \mathcal{P}^\infty(\mathbb{P}) \), \( X_k \sigma \mu \searrow X \sigma \mu \) \( \text{resp.,} \ X_k \sigma \mu \nearrow X \sigma \mu \) in the sense of Definition 2.3.

2. Let \( \mu, \mu_k \in \mathcal{P}^\infty(\mathbb{P}) \) \( (k \in \mathbb{N}) \), and assume that \( \mu_k \searrow \mu \) \( \text{resp.,} \ \mu_k \nearrow \mu \). Then for every \( X \in \mathbb{P} \), \( X \sigma \mu_k \searrow X \sigma \mu \) \( \text{resp.,} \ X \sigma \mu_k \nearrow X \sigma \mu \).

**Proof.** (1) From Lemma 3.4 it is immediate that \( X_1 \sigma \mu \geq X_2 \sigma \mu \geq \cdots \geq X \sigma \mu \) \( \text{resp.,} \ X_1 \sigma \mu \leq X_2 \sigma \mu \leq \cdots \leq X \sigma \mu \). Choose an \( \varepsilon \in (0, 1) \) such that all of \( X, X_k \) and \( \text{supp}(\mu) \) are in \( \Sigma_\varepsilon \). Since \( Y \sigma A \in \Sigma_\varepsilon \) for all \( Y, A \in \Sigma_\varepsilon \), note that all \( X \sigma \mu \) and \( X_k \sigma \mu \) are supported on \( \Sigma_\varepsilon \). Since \( X_k \sigma A \to X \sigma A \) in SOT as \( k \to \infty \) for any \( A \in \mathbb{P} \), for every bounded SOT-continuous real function \( f \) on \( \Sigma_\varepsilon \), it follows from the bounded convergence theorem that
\[
\int_\mathbb{P} f(A) d(X_k \sigma \mu)(A) = \int_\mathbb{P} f(X_k \sigma A) d\mu(A) \to \int_\mathbb{P} f(X \sigma A) d\mu(A) = \int_\mathbb{P} f(A) d(X \sigma \mu)(A).
\]
This implies the assertion.

(2) From Lemma 3.1 again, \( X_\sigma \mu_1 \geq X_\sigma \mu_2 \geq \cdots \geq X_\sigma \mu \) (resp., \( X_\sigma \mu_1 \leq X_\sigma \mu_2 \leq \cdots \leq X_\sigma \mu \)). Choose an \( \varepsilon \in (0, 1) \) such that all \( \mu, \mu_k \) are supported on \( \Sigma_\varepsilon \) and \( X \in \Sigma_\varepsilon \). Then all \( X_\sigma \mu \) and \( X_\sigma \mu_k \) are supported on \( \Sigma_\varepsilon \). For every bounded SOT-continuous real function \( f \) on \( \Sigma_\varepsilon \), since \( A \in \Sigma_\varepsilon \mapsto f(X_\sigma A) \) is SOT-continuous, we have

\[
\int_P f(A) d(X_\sigma \mu_k)(A) = \int_P f(X_\sigma A) d\mu_k(A)
\]

\[
\rightarrow \int_P f(X_\sigma A) d\mu(A) = \int_P f(A) d(X_\sigma \mu)(A).
\]

Hence the assertion follows. \( \square \)

The next lemma is crucial to obtain the uniqueness of a solution to (3.1).

**Lemma 3.6.** Let \( \sigma \neq 1 \). If \( X, Y \in \mathbb{P} \) and \( X \neq Y \), then

\[
d_T(M(X_\sigma \mu), M(Y_\sigma \mu)) < d_T(X, Y)
\]

for every \( \mu \in \mathcal{P}^\infty(\mathbb{P}) \).

**Proof.** Let \( X, Y \in \mathbb{P} \) be such that \( X \neq Y \), and let \( \alpha := d_T(X, Y) > 0 \). Choose an \( \varepsilon \in (0, 1) \) such that \( X \) and \( \text{supp}(\mu) \) are in \( \Sigma_\varepsilon \). For every \( A \in \mathbb{P} \) note that

\[
Y_\sigma A \leq (e^\alpha X) \sigma A = e^\alpha X^{1/2} f_\sigma(e^{-\alpha} X^{-1/2} \sigma A X^{-1/2}) X^{1/2},
\]

(3.3)

\[
Y_\sigma A \geq (e^{-\alpha} X) \sigma A = e^{-\alpha} X^{1/2} f_\sigma(e^{\alpha} X^{-1/2} \sigma A X^{-1/2}) X^{1/2}.
\]

(3.4)

For every \( A \in \Sigma_\varepsilon \), since \( \varepsilon^2 I \leq X^{-1/2} A X^{-1/2} \leq \varepsilon^{-2} I \), we have

\[
f_\sigma(X^{-1/2} A X^{-1/2}) - f_\sigma(e^{-\alpha} X^{-1/2} A X^{-1/2}) \geq \left( \min_{t \in [e^\varepsilon, e^{-\varepsilon}]} \{ f_\sigma(t) - f_\sigma(e^{-\alpha} t) \} \right) I,
\]

\[
f_\sigma(e^\alpha X^{-1/2} A X^{-1/2}) - f_\sigma(X^{-1/2} A X^{-1/2}) \geq \left( \min_{t \in [e^\varepsilon, e^{-\varepsilon}]} \{ f_\sigma(e^\alpha t) - f_\sigma(t) \} \right) I.
\]

Since \( \sigma \neq 1 \), \( f_\sigma \) is strictly increasing (and analytic) on \( (0, \infty) \), the minima in the above two expressions are strictly positive. Hence there exists a \( \rho \in (0, 1) \) such that, for every \( A \in \Sigma_\varepsilon \),

\[
f_\sigma(X^{-1/2} A X^{-1/2}) - f_\sigma(e^{-\alpha} X^{-1/2} A X^{-1/2}) \geq \rho I,
\]

(3.5)

\[
f_\sigma(e^\alpha X^{-1/2} A X^{-1/2}) - f_\sigma(X^{-1/2} A X^{-1/2}) \geq \rho I.
\]

(3.6)

Therefore,

\[
Y_\sigma A \leq e^\alpha (X_\sigma A - \rho X) \leq e^\alpha (1 - \rho \varepsilon^2)(X_\sigma A),
\]

\[
Y_\sigma A \geq e^{-\alpha} (X_\sigma A + \rho X) \geq e^{-\alpha} (1 + \rho \varepsilon^2)(X_\sigma A),
\]

since \( \varepsilon I \leq X \leq \varepsilon^{-1} I \) and \( \varepsilon I \leq X_\sigma A \leq \varepsilon^{-1} I \) so that \( \varepsilon^2 (X_\sigma A) \leq X \leq \varepsilon^{-2} (X_\sigma A) \).

Choosing a \( \beta \in (0, \alpha) \) such that \( e^{\alpha - \beta} \leq 1 + \rho \varepsilon^2 \), we have

\[
e^{-\beta} (X_\sigma A) \leq Y_\sigma A \leq e^\beta (X_\sigma A), \quad A \in \Sigma_\varepsilon.
\]

(3.7)
Now, write $\psi_X(A) := X\sigma A$ and $\psi_Y(A) := Y\sigma A$ for $A \in \mathbb{P}$. The claim given in (3.4) means that $e^{-\beta}\psi_X(A) \leq \psi_Y(A) \leq e^{\beta}\psi_X(A)$ for all $A \in \Sigma_\epsilon$. Since $(e^{\pm\beta}\psi_X)_\epsilon = e^{\pm\beta}(X\sigma \mu)$ and $(\psi_Y)_\epsilon = Y\sigma \mu$, Lemma 3.4 gives

$$e^{-\beta}(X\sigma \mu) \leq Y\sigma \mu \leq e^{\beta}(X\sigma \mu).$$

By (i) and (ii) this implies that

$$e^{-\beta}M(X\sigma \mu) \leq M(Y\sigma \mu) \leq e^{\beta}M(X\sigma \mu)$$

so that $d_T(M(X\sigma \mu), M(Y\sigma \mu)) \leq \beta < d_T(X, Y)$. \hfill \Box

We are now in a position to prove the theorem.

**Proof of Theorem 3.1** (1) For any fixed $\mu \in \mathcal{P}^\infty(\mathbb{P})$ define a map $F : \mathbb{P} \to \mathbb{P}$ by

$$F(X) := M(X\sigma \mu), \quad X \in \mathbb{P},$$

which is monotone, i.e., $X \leq Y$ implies $F(X) \leq F(Y)$, by Lemma 3.4 and (i) of Definition 2.5. Choose an $\epsilon \in (0, 1)$ such that $\mu$ is supported on $\Sigma_\epsilon$, and let $Z := \epsilon^{-1}I$. Since $\mu \leq \epsilon^{-1}\delta_1$ and so $Z\sigma \mu \leq \epsilon^{-1}\delta_1$, we have $F(Z) \leq \epsilon^{-1}I = Z$ by (i) and (iv), and iterating this implies that $Z \geq F(Z) \geq F^2(Z) \geq \cdots$. Moreover, since $(\epsilon I)\sigma \mu \geq \epsilon I$, $F(Z) \geq F(\epsilon I) \geq \epsilon I$, and by iterating this we have $F^k(Z) \geq \epsilon I$ for all $k$. Therefore, $F^k(Z) \searrow X_0 \in \mathbb{P}$ for some $X_0 \in \mathbb{P}$, and hence $F^k(Z)\sigma \mu \searrow X_0\sigma \mu$ by Lemma 3.5(1). From the monotone continuity of $M$ in (iii) it follows that

$$F^{k+1}(Z) = M(F^k(Z)\sigma \mu) \searrow M(X_0\sigma \mu),$$

which yields that $X_0 = M(X_0\sigma \mu)$.

To prove the uniqueness of the solution, assume that $X_1 \in \mathbb{P}$ satisfies $X_1 = M(X_1\sigma \mu)$ and $X_1 \neq X_0$. Then by Lemma 3.6 we have

$$d_T(X_0, X_1) = d_T(M(X_0\sigma \mu), M(X_1\sigma \mu)) < d_T(X_0, X_1),$$

a contradiction.

(2) Let $F$ be as in the proof of (1). If $Y \geq M(Y\sigma \mu)$, then $Y \geq F(Y) \geq F^2(Y) \geq \cdots$. Choose an $\epsilon \in (0, 1)$ such that $Y$ and $\text{supp}(\mu)$ are in $\Sigma_\epsilon$. Then as in the proof of (1), $F^k(Y) \geq \epsilon I$ for all $k$, and hence $F^k(Y) \searrow X'$ for some $X' \in \mathbb{P}$. As in the proof of (1) again, $X'$ is a solution to (3.1) so that $Y \geq X' = X_0$. The proof of the second assertion is similar, where we have $Y' \leq F(Y') \leq F^2(Y') \leq \cdots$ and use the upward continuity in Lemma 3.5(1) and (iii).

(3) (i) Let $\mu, \nu \in \mathcal{P}^\infty(\mathbb{P})$ and $\mu \leq \nu$. Let $X_0 := M_\sigma(\mu)$ and $Y_0 := M_\sigma(\nu)$, i.e., $X_0 = M(X_0\sigma \mu)$ and $Y_0 = M(Y_0\sigma \mu)$. Since $X_0 \leq M(X_0\sigma \nu)$ by Lemma 3.4 we have $X_0 \leq Y_0$ by (2).

(ii) One can easily see that $\alpha(X\sigma \mu) = \alpha(X)\sigma(\alpha, \mu)$ for every $X \in \mathbb{P}$, $\mu \in \mathcal{P}^\infty(\mathbb{P})$ and $\alpha > 0$. Hence, if $X_0 := M_\sigma(\mu)$, then it follows from (ii) for $M$ that

$$\alpha X_0 = M(\alpha(X_0\sigma \mu)) = M((\alpha X_0)\sigma(\alpha, \mu)),$$

which gives $\alpha X_0 = M_\sigma(\alpha \mu)$. 

Hence, Lemma 3.4 again, we have

\[ \mu \text{ that all} \]

For any fixed \( k \), since \( X_k \sigma \mu_k \geq X_0 \sigma \mu \) by Lemma 3.4, we have \( X_k = M(X_k \sigma \mu_k) \geq M(X_0 \sigma \mu) \). Hence \( X_0 \geq M(X_0 \sigma \mu) \) follows. On the other hand, when \( k < l \), since \( X_l \sigma \mu_l \leq X_l \sigma \mu_k \) by Lemma 3.4 again, we have

\[ X_l = M(X_l \sigma \mu_l) \leq M(X_l \sigma \mu_k). \tag{3.8} \]

For any fixed \( k \), since \( X_l \sigma \mu_k \downarrow X_0 \sigma \mu_k \) as \( k \to \infty \) by Lemma 3.5 (1), it follows from (iii) for \( M \) that \( M(X_l \sigma \mu_k) \to M(X_0 \sigma \mu_k) \) in SOT. Hence from (3.8) we have \( X_0 \leq M(X_0 \sigma \mu_k) \) for every \( k \). Since \( X_0 \sigma \mu_k \downarrow X_0 \sigma \mu \) by Lemma 3.4 (2), we furthermore have \( M(X_0 \sigma \mu_k) \to M(X_0 \sigma \mu) \) in SOT, so that \( X_0 \leq M(X_0 \sigma \mu) \) follows. Therefore, we have shown that \( X_0 = M(X_0 \sigma \mu) \), that is, \( X_0 = M_\sigma(\mu) \).

When \( \mu_k \nless \mu \), the proof is analogous, so we may omit the details.

(iv) is obvious since \( I = M(\delta_I) = M(I \sigma \delta_I) \).

\[ \square \]

4. Basic properties

In this section, as in Section 3, let \( M \) be an operator mean on \( \mathcal{P}^\infty(\mathbb{P}) \) and \( \sigma \) be a two-variable operator mean with \( \sigma \neq I \). The next properties of the deformed operator mean \( M_\sigma \) can easily been verified by using Theorem 3.1 whose proofs are left to the reader.

Proposition 4.1. \( M_I = M \), where \( I \) is the right trivial two-variable operator mean \( XrY = Y \).

(2) Let \( \tilde{M} : \mathcal{P}^\infty(\mathbb{P}) \to \mathbb{P} \) be an operator mean satisfying (i)–(iv) and \( \tilde{\sigma} \) be a two-variable operator mean with \( \tilde{\sigma} \neq I \). If \( M \leq \tilde{M} \) and \( \sigma \leq \tilde{\sigma} \), then \( M_\sigma \leq \tilde{M}_\tilde{\sigma} \).

(3) Define the adjoint \( M^* \) of \( M \) by \( M^*(\mu) := M(\mu^{-1})^{-1} \) for \( \mu \in \mathcal{P}^\infty(\mathbb{P}) \), where \( \mu^{-1} \) is the push-forward of \( \mu \) by \( A \mapsto A^{-1}, A \in \mathbb{P} \). Let \( \sigma^* \) be the adjoint of \( \sigma \), i.e., \( A \sigma^* B = (A^{-1} \sigma B^{-1})^{-1} \). Then \( M^* \) satisfies (i)–(iv) again and \( (M_\sigma)^* = (M^*)_{\sigma^*} \).

In addition to properties (i)–(iv) of \( M \) that are essential for Theorem 3.1 we consider the following properties:

(v) Barycentric identity: \( M(\delta_A) = A \) for every \( A \in \mathbb{P} \). This contains (iv).

(vi) Congruence invariance: For every \( \mu \in \mathcal{P}^\infty(\mathbb{P}) \) and every invertible \( S \in B(\mathcal{H}) \),

\[ SM(\mu)S^* = M(S\mu S^*), \]

where \( S\mu S^* \) is the push-forward of \( \mu \) by \( A \in \mathbb{P} \mapsto SAS^* \in \mathbb{P} \). This property contains (ii).
(vii) **Concavity:** For every \( \mu_j, \nu_j \in \mathcal{P}^\infty(\mathbb{P}) \) \((1 \leq j \leq n)\), any weight \((w_1, \ldots, w_n)\) with any \( n \in \mathbb{N} \), and \( 0 < t < 1 \),

\[
M\left( \sum_{j=1}^{n} w_j (\mu_j \nabla_t \nu_j) \right) \geq (1-t) M\left( \sum_{j=1}^{n} w_j \mu_j \right) + t M\left( \sum_{j=1}^{n} w_j \nu_j \right),
\]

where \( \mu_j \nabla_t \nu_j \) is the push-forward of \( \mu_j \times \nu_j \) by \( \nabla_t : \mathbb{P} \times \mathbb{P} \to \mathbb{P}, \nabla_t(A, B) := (1-t)A + tB \), the \( t \)-weighted arithmetic mean. The following two particular cases may be worth noting separately. The first one is the joint concavity when restricted to the weighted \( n \)-variable situation.

(vii-1) For every \( A_j, B_j \in \mathbb{P} \) \((1 \leq j \leq n)\), and \( 0 < t < 1 \),

\[
M\left( \sum_{j=1}^{n} w_j \delta_{(1-t)A_j + tB_j} \right) \geq (1-t) M\left( \sum_{j=1}^{n} w_j \delta_{A_j} \right) + t M\left( \sum_{j=1}^{n} w_j \delta_{B_j} \right).
\]

(vii-2) For every \( \mu, \nu \in \mathcal{P}^\infty(\mathbb{P}) \) and \( 0 < t < 1 \),

\[
M(\mu \nabla_t \nu) \geq (1-t) M(\mu) + t M(\nu).
\]

(viii) **Arithmetic-M-harmonic mean inequality:** For every \( \mu \in \mathcal{P}^\infty(\mathbb{P}) \),

\[
\mathcal{H}(\mu) \leq M(\mu) \leq A(\mu),
\]

where

\[
A(\mu) := \int_{\mathbb{P}} A \, d\mu(A), \quad \mathcal{H}(\mu) := \left[ \int_{\mathbb{P}} A^{-1} \, d\mu(A) \right]^{-1}
\]

are the arithmetic and the harmonic means.

**Theorem 4.2.** If \( M \) satisfies each of (v), (vi), (vii), (vii-1), (vii-2), and (viii) (in addition to (i)–(iv)), then \( M_\sigma \) does the same.

**Proof.** (v) This is obvious since \( A \sigma_\delta A = \delta_A \).

(vi) Assume that \( M \) satisfies (vi). For every \( \mu \in \mathcal{P}^\infty(\mathbb{P}) \) let \( X_0 := M_\sigma(\mu) \). Then for any invertible \( S \in B(\mathcal{H}) \),

\[
SX_0S^* = SM(X_0\sigma\mu)S^* = M(SX_0\sigma\mu S^*) = M((SX_0S^*)\sigma(S\mu S^*))
\]

so that \( SX_0S^* = M_\sigma(S\mu S^*) \).

(vii) Assume that \( M \) satisfies (vii). Let \( \mu_j, \nu_j \in \mathcal{P}^\infty(\mathbb{P}) \) \((1 \leq j \leq n)\) and \((w_1, \ldots, w_n)\) be any weight, and let \( 0 < t < 1 \). Set \( \mu := \sum_{j=1}^{n} w_j \mu_j, \nu := \sum_{j=1}^{n} w_j \nu_j, X_0 := M_\sigma(\mu) \) and \( Y_0 := M_\sigma(\nu) \). Since \( X_0\sigma\mu = \sum_{j=1}^{n} w_j (X_0\sigma\mu_j) \) and \( Y_0\sigma\nu = \sum_{j=1}^{n} w_j (Y_0\sigma\nu_j) \), we have

\[
X_0 \nabla_t Y_0 = M(X_0\sigma\mu) \nabla_t M(Y_0\sigma\nu) \leq M\left( \sum_{j=1}^{n} w_j ((X_0\sigma\mu_j) \nabla_t (Y_0\sigma\nu_j)) \right)
\]

thanks to (vii) for \( M \). We now show that, for every \( \mu, \nu \in \mathcal{P}^\infty(\mathbb{P}) \),

\[
(X_0\sigma\mu) \nabla_t (Y_0\sigma\nu) \leq (X_0\nabla_t Y_0)\sigma(\mu \nabla_t \nu).
\]
For $X \in \mathcal{P}$ define $\psi_X : \mathcal{P} \to \mathcal{P}$ by $\psi_X(A) := X\sigma A$. The left-hand side of \((4.3)\) is

$$
(\nabla_t)_*((\psi_{X_0})_*\mu \times (\psi_{Y_0})_*\nu) = (\nabla_t \circ (\psi_{X_0} \times \psi_{Y_0}))_*(\mu \times \nu).
$$

The right-hand side of \((4.3)\) is

$$
(\psi_{X_0\nabla_t Y_0} \circ \nabla_t)_*(\mu \times \nu).
$$

Set $\varphi_1 := \nabla_t \circ (\psi_{X_0} \times \psi_{Y_0})$ and $\varphi_2 := \psi_{X_0\nabla_t Y_0} \circ \nabla_t$. To prove \((4.3)\), it suffices to show that $\varphi_1(A, B) \leq \varphi_2(A, B)$ for all $A, B \in \mathcal{P}$. In fact, when this holds, we have $(\varphi_1)_*(\mu \times \nu) \leq (\varphi_2)_*(\mu \times \nu)$ similarly to the proof of Lemma 3.4. For every $A, B \in \mathcal{P}$ we have

$$
\varphi_1(A, B) = \psi_{X_0}(A) \nabla_t \psi_{Y_0}(B) = (X_0\sigma A) \nabla_t (Y_0\sigma B)
$$

$$
\leq (X_0 \nabla_t Y_0) \sigma (A \nabla_t B) = \varphi_2(A, B),
$$

where inequality follows from [27, Theorem 3.5]. Hence \((4.3)\) has been shown, from which we have

$$
\sum_{j=1}^{n} w_j(\nabla_t (X_0\sigma \mu_j) \nabla_t (Y_0\sigma \nu_j)) \leq \sum_{j=1}^{n} w_j(\nabla_t (X_0 \nabla_t Y_0) \sigma (\mu_j \nabla_t \nu_j))
$$

$$
= (X_0 \nabla_t Y_0) \sigma \left(\sum_{j=1}^{n} w_j(\mu_j \nabla_t \nu_j)\right).
$$

Applying monotonicity of $M$ to this and combining with \((4.2)\) we have

$$
X_0 \nabla_t Y_0 \leq M \left( (X_0 \nabla_t Y_0) \sigma \left(\sum_{j=1}^{n} w_j(\mu_j \nabla_t \nu_j)\right) \right),
$$

which implies that $X_0 \nabla_t Y_0 \leq M\sigma(\sum_{j=1}^{n} w_j(\mu_j \nabla_t \nu_j))$ by Theorem 3.1 (2).

The proofs of the assertions for (vii-1) and for (vii-2) are similar to the above proof for (vii), so we omit the details.

(viii) Assume that $M$ satisfies (viii). Let $\alpha := f'_{\sigma}(1)$; then $0 < \alpha \leq 1$ since $\sigma \neq 1$. It is well-known [10, (3.3.2)] that $!_{\alpha} \leq \sigma \leq \nabla_{\alpha}$, where $!_{\alpha}$ is the $\alpha$-weighted harmonic mean $A!_{\alpha}B := ((1 - \alpha)A^{-1} + \alpha B)^{-1}$. By Proposition 4.1 (2) we have

$$
\mathcal{H}!_{\alpha} \leq M\sigma \leq \mathcal{A}\nabla_{\alpha}.
$$

Hence it suffices to show that $\mathcal{H}!_{\alpha} = \mathcal{H}$ and $\mathcal{A}\nabla_{\alpha} = \mathcal{A}$. But it is immediate to find that the solutions of the equations $X = \mathcal{A}(X \nabla_{\alpha} \mu)$ and $X = \mathcal{H}(X !_{\alpha} \mu)$ are $\mathcal{A}(\mu)$ and $\mathcal{H}(\mu)$, respectively.

By Theorems 2.8, 3.1 and 4.2 we have

**Corollary 4.3.** If $M$ is a barycenter (i.e., it satisfies (v)), then $M\sigma$ is a $(d_{T})^W$-contractive barycenter on $\mathcal{P}^\infty(\mathcal{P})$ for any $\sigma \neq 1$. 

5. Examples

In this section we provide typical examples of operator means on $\mathcal{P}^\infty(\mathbb{P})$ satisfying (i)–(viii) and their deformed operator means. We note from Corollary 4.3 that all of those operator means are $(d_T)_{\infty}^W$-contractive barycenters on $\mathcal{P}^\infty(\mathbb{P})$.

5.1. Arithmetic and harmonic means. The arithmetic mean $\mathcal{A}$ and the harmonic mean $\mathcal{H}$ on $\mathcal{P}^\infty(\mathbb{P})$ are given in (4.1). It is straightforward to see that $\mathcal{A}$ satisfies all the properties (i)–(viii). It is also easy to see $\mathcal{H}$ satisfies the properties (i)–(viii) except (vii) (including (vii-1) and (vii-2)). Since it does not seem easy to show (vii) directly for $\mathcal{H}$, we take a detour by giving the following proposition, which may be of independent interest.

**Proposition 5.1.** For every $\mu \in \mathcal{P}^\infty(\mathbb{P})$, $\mathcal{A}_{1/s}(\mu) \leq \mathcal{A}_{1}(\mu)$ for $0 < s' < s \leq 1$ and

$$\mathcal{H}(\mu) = \lim_{s \searrow 0} \mathcal{A}_{1/s}(\mu) \quad \text{in SOT}.$$ 

**Proof.** Let $\mu \in \mathcal{P}^\infty(\mathbb{P})$ and set $X_s := \mathcal{A}_{1/s}(\mu)$ for each $s \in (0, 1]$. Assume that $0 < s' < s \leq 1$. Since $X_s = \mathcal{A}(X_{s!s}\mu)$, we have

$$I = \mathcal{A}(I!_{s}(X_s^{-1/2}\mu X_s^{-1/2}))$$

$$= \int_{\mathbb{P}} [(1 - s)I + s(X_s^{-1/2}AX_s^{-1/2})^{-1}]^{-1} d\mu(A)$$

$$= \int_{\mathbb{P}} [I + s(X_s^{1/2}A^{-1}X_s^{1/2} - I)]^{-1} d\mu(A). \quad (5.1)$$

Here note that $s \in (0, 1] \mapsto (1 + s(t - 1))^{1/s}$ is a decreasing function for any $t > 0$. Hence we have

$$[I + s'(X_s^{1/2}A^{-1}X_s^{1/2} - I)]^{1/s'} \leq [I + s'X_s^{1/2}A^{-1}X_s^{1/2} - I]^{1/s'}$$

so that

$$[I + s(X_s^{1/2}A^{-1}X_s^{1/2} - I)]^{-s'/s} \geq [I + s'(X_s^{1/2}A^{-1}X_s^{1/2} - I)]^{-1}. \quad (5.2)$$

Applying the operator concavity of $x^{s'/s}$ on $(0, \infty)$ to (5.1) and using (5.2) we have

$$I \geq \int_{\mathbb{P}} [I + s(X_s^{1/2}A^{-1}X_s^{1/2} - I)]^{-s'/s} d\mu(A)$$

$$\geq \int_{\mathbb{P}} [I + s'(X_s^{1/2}A^{-1}X_s^{1/2} - I)]^{-1} d\mu(A)$$

$$= \mathcal{A}(I!_{s'}(X_s^{-1/2}\mu X_s^{-1/2}))$$

so that $X_s \geq \mathcal{A}(X_{s!s'}\mu)$. Hence $X_s \geq X_{s'}$ by Theorem 3.1(2).

Choose an $\varepsilon \in (0, 1)$ such that $\mu$ is supported on $\Sigma_\varepsilon$. Since $\varepsilon\delta = \mu \leq \varepsilon^{-1}\delta$, we have $\varepsilon I \leq X_s \leq \varepsilon^{-1}I$ for all $s \in (0, 1]$. Hence $X_s \searrow X_0$ for some $X_0 \in \mathbb{P}$. It remains to show that $X_0 = \mathcal{H}(\mu)$. For every $s \in (0, 1]$ and every $A \in \text{supp}(\mu)$, since $X_s^{1/2}A^{-1}X_s^{1/2} \leq \varepsilon^{-1}X_s \leq \varepsilon^{-2}I, \|X_s^{1/2}A^{-1}X_s^{1/2} - I\| \leq \varepsilon^{-2}$. Hence we write, as $s \searrow 0$,

$$[I + s(X_s^{1/2}A^{-1}X_s^{1/2} - I)]^{-1} = I - s(X_s^{1/2}A^{-1}X_s^{1/2} - I) + o(s),$$
where \( o(s)/s \to 0 \) in the operator norm as \( s \searrow 0 \) uniformly for \( A \in \text{supp}(\mu) \). Therefore, from (5.1) we find that
\[
I = (1 + s)I - s \int_{\mathcal{P}} X_s^{1/2} A^{-1} X_s^{1/2} d\mu(A) + o(s) \quad \text{as} \quad s \searrow 0,
\]
which yields that
\[
I = \lim_{s \searrow 0} \int_{\mathcal{P}} X_s^{1/2} A^{-1} X_s^{1/2} d\mu(A)
\]
in the operator norm. On the other hand, for every \( \xi \in \mathcal{H} \), the bounded convergence theorem gives
\[
\lim_{s \searrow 0} \int_{\mathcal{P}} \langle \xi, X_s^{1/2} A^{-1} X_s^{1/2} \xi \rangle d\mu(A) = \int_{\mathcal{P}} \langle \xi, X_0^{1/2} A^{-1} X_0^{1/2} \xi \rangle d\mu(A).
\]
Therefore, \( I = \int_{\mathcal{P}} X_0^{1/2} A^{-1} X_0^{1/2} d\mu(A) \) so that \( X_0^{-1} = \int_{\mathcal{P}} A^{-1} d\mu(A) \), i.e., \( X_0 = \mathcal{H}(\mu) \). \( \square \)

Since \( \mathcal{A}_s \) satisfies (vii) by Theorem 4.2, it follows from Proposition 5.1 that \( \mathcal{H} \) satisfies (vii) as well as all other properties in (i)–(viii).

### 5.2. Power means.
For each \( r \in [-1,1] \setminus \{0\} \) the power mean \( P_r \) on \( \mathcal{P}^\infty(\mathbb{P}) \) is introduced as the solution to the equation for \( X \in \mathcal{P} \)
\[
\begin{cases}
  X = \mathcal{A}(X_{\#}^r \mu) & \text{when } r \in (0,1], \\
  X = \mathcal{H}(X_{\#}^{-r} \mu) & \text{when } r \in [-1,0),
\end{cases}
\]
that is, in our notation, \( P_r = \mathcal{A}_{\#}^r \) and \( P_r = \mathcal{H}_{\#}^r \) for \( r \in (0,1] \). As mentioned in Remark 3.2, (5.3) is rewritten as a typical case of the generalized Karcher equation introduced by Pálfia [40]. Among the properties in (i)–(viii), the only properties not well-known for \( P_r \) are (iii) and (vii) (including (vii-1) and (vii-2)); the other properties are included in [40, Theorem 6.4, Proposition 6.15]. But, all the properties in (i)–(viii) for \( P_r \) are immediate consequences of Theorems 3.1 (3) and 4.2 applied to \( M = \mathcal{A} \) or \( \mathcal{H} \), since \( \mathcal{A} \) and \( \mathcal{H} \) satisfies them, as shown in Section 5.1.

**Remark 5.2.** Note that the \( (d_T)_\infty^W \)-contractivity of the power means on \( \mathcal{P}_{cp}(\mathbb{P}) \), the set of \( \mu \in \mathcal{P}(\mathbb{P}) \) with compact support, was given in [23, Proposition 6.7]. But it follows from Theorem 2.8 that the power means are \( d_\infty^W \)-contractive on \( \mathcal{P}^\infty(\mathbb{P}) \) bigger than \( \mathcal{P}_{cp}(\mathbb{P}) \).

### 5.3. Karcher mean.
The Karcher mean (or the Cartan barycenter) \( G \) on \( \mathcal{P}^\infty(\mathbb{P}) \) is introduced as the solution to the Karcher equation
\[
\int_{\mathcal{P}} \log X^{-1/2} AX^{-1/2} d\mu(A) = 0
\]
for given \( \mu \in \mathcal{P}^\infty(\mathbb{P}) \), which is the original case of the generalized Karcher equation in [40]. So the properties (i)–(viii) for \( G \), except (iii) and (vii), are known in [40]. Below we will prove (iii) and (vii) for \( G \) based on the convergence \( P_r \to G \) as \( r \to 0 \).
For the $n$-variable weighted case with a weight $w = (w_1, \ldots, w_n)$, the convergence $P_{w,r}(A_1, \ldots, A_n) \to G_w(A_1, \ldots, A_n)$ as $r \to 0$ was first established in [33] when $\dim \mathcal{H} < \infty$ and then extended in [31] to the SOT-convergence when $\dim \mathcal{H} = \infty$. Here note that $G_w(A_1, \ldots, A_n) = G(\sum_{j=1}^n w_j \delta_{A_j})$ and $P_{w,r}(A_1, \ldots, A_n) = P_r(\sum_{j=1}^n w_j \delta_{A_j})$. For the probability measure case, the convergence $P_r \to G$ was shown in [24] when $\dim \mathcal{H} < \infty$, and that for compactly supported probability measures when $\dim \mathcal{H} = \infty$ was in [23] Theorem 7.4. In the following we give the convergence for probability measures in $\mathcal{P}^\infty(\mathbb{P})$ when $\dim \mathcal{H} = \infty$. (Even an operator norm convergence of $P_r \to G$ is given in [36].)

**Proposition 5.3.** For every $\mu \in \mathcal{P}^\infty(\mathbb{P})$,

$$P_{-r}(\mu) \leq P_{-r'}(\mu) \leq G(\mu) \leq P_{r'}(\mu) \leq P_r(\mu) \quad \text{for } 0 < r' < r \leq 1,$$

and

$$G(\mu) = \lim_{r \to 0} P_r(\mu) \quad \text{in SOT}.$$

**Proof.** Let $\mu \in \mathcal{P}^\infty(\mathbb{P})$ and set $X_r := P_r(\mu)$ for each $r \in (0,1]$. Assume that $0 < r' < r \leq 1$. Since $X_r = A(X_r \#_{r'} \mu)$, we have

$$I = A(I \#_{r'}(X_r^{-1/2} \mu X_r^{-1/2})) = \int_\mathbb{P} (X_r^{-1/2}AX_r^{-1/2})^r d\mu(A). \quad (5.4)$$

By the operator concavity of $x^{r'/r}$ we have

$$I \geq \int_\mathbb{P} [(X_r^{-1/2}AX_r^{-1/2})^r]^{r'/r} d\mu(A) = \int_\mathbb{P} (X_r^{-1/2}AX_r^{-1/2})^{r'} d\mu(A)$$

so that $X_r \geq A(X_r \#_{r'} \mu)$ and so $X_r \geq X_{r'}$ by Theorem 3.1 (2). Therefore, $P_{r'}(\mu) \leq P_r(\mu)$, which also implies that $P_{-r}(\mu) \leq P_{-r'}(\mu)$ since $P_{-r}(\mu) = P_r(\mu^{-1})$.

Choose an $\varepsilon \in (0,1)$ as in the proof of Proposition 5.1. Since $X_r \geq \varepsilon I$, $X_r \searrow X_0$ for some $X_0 \in \mathbb{P}$. Since $\|X_r^{-1/2}AX_r^{-1/2}\| \leq \varepsilon^{-2}$ for every $r \in (0,1]$ and every $A \in \text{supp}(\mu)$, we have, as $r \searrow 0$,

$$(X_r^{-1/2}AX_r^{-1/2})^r = \exp(r \log X_r^{-1/2}AX_r^{-1/2}) = I + r \log X_r^{-1/2}AX_r^{-1/2} + o(r),$$

where $o(r)/r \to 0$ in the operator norm as $r \searrow 0$ uniformly for $A \in \text{supp}(\mu)$. Therefore, from (5.4) we find that

$$I = I + r \int_\mathbb{P} \log X_r^{-1/2}AX_r^{-1/2} d\mu(A) + o(r)$$

so that

$$\lim_{r \searrow 0} \int_\mathbb{P} \log X_r^{-1/2}AX_r^{-1/2} d\mu(A) = 0 \quad (5.5)$$

in the operator norm. On the other hand, note that $X_r^{-1/2}AX_r^{-1/2} \to X_0^{-1/2}AX_0^{-1/2}$ in SOT and hence $\log X_r^{-1/2}AX_r^{-1/2} \to \log X_0^{-1/2}AX_0^{-1/2}$ in SOT as $r \searrow 0$. For every $\xi \in \mathcal{H}$, the bounded convergence theorem gives

$$\lim_{r \searrow 0} \int_\mathbb{P} \langle \xi, (\log X_r^{-1/2}AX_r^{-1/2}) \xi \rangle d\mu(A) = \int_\mathbb{P} \langle \xi, (\log X_0^{-1/2}AX_0^{-1/2}) \xi \rangle d\mu(A).$$
Therefore, \( \int_{\mathbb{P}} \log X_0^{-1/2}AX_0^{-1/2} \, d\mu(A) = 0 \) so that \( X_0 = G(\mu) \). Hence we have \( P_r(\mu) \downarrow G(\mu) \) as \( r \downarrow 0 \), which also implies that \( P_{-r}(\mu) = P_r(\mu^{-1})^{-1} \searrow G(\mu^{-1})^{-1} = G(\mu) \) as \( r \downarrow 0 \), so that \( P_r(\mu) \to G(\mu) \) in SOT as \( r \to 0 \). \hfill \Box

Since \( P_r \) satisfies (vii) as shown in Section 4.2, we see by Proposition 5.3 that \( G \) satisfies the same. Moreover, we have

**Proposition 5.4.** The Karcher mean \( G \) satisfies (iii).

*Proof.* By Proposition 5.3

\[
G(\mu) = \inf_{0 < r \leq 1} P_r(\mu) = \sup_{0 < r \leq 1} P_{-r}(\mu), \quad \mu \in \mathcal{P}^\infty(\mathbb{P}).
\]

Since \( P_r \) satisfies (iii) as shown in Section 4.2, we find that if \( \mu_k \downarrow \mu \) then for any \( \xi \in \mathcal{H} \),

\[
\langle \xi, G(\mu) \xi \rangle = \inf_{0 < r \leq 1} \langle \xi, P_r(\mu) \xi \rangle = \inf_{0 < r \leq 1} \inf_{k \geq 1} \langle \xi, P_r(\mu_k) \xi \rangle
\]

\[
= \inf_{k \geq 1} \inf_{0 < r \leq 1} \langle \xi, P_r(\mu_k) \xi \rangle = \inf_{k \geq 1} \langle \xi, G(\mu_k) \xi \rangle.
\]

Therefore, \( G(\mu_k) \downarrow G(\mu) \). If \( \mu_k \nearrow \mu \), then we have \( G(\mu_k) \nearrow G(\mu) \) similarly; or since \( \mu_k^{-1} \downarrow \mu^{-1} \), we have \( G(\mu_k) = G(\mu_k^{-1})^{-1} \searrow G(\mu^{-1})^{-1} = G(\mu) \). \hfill \Box

In this way, we have seen that all of \( \mathcal{A}, \mathcal{H}, G \) and \( P_r \) for \( r \in [-1, 1] \setminus \{0\} \) satisfy all the properties in (i)–(viii).

We end the section with an open problem.

**Problem 5.5.** Assume that \( \mu, \mu_k \in \mathcal{P}^\infty(\mathbb{P}) \) \( (k \in \mathbb{N}) \) are supported on \( \Sigma_\varepsilon \) for some \( \varepsilon > 0 \) and \( \mu_k \to \mu \) weakly on \( \Sigma_\varepsilon \) with SOT, i.e., \( \int_{\mathbb{P}} f(A) \, d\mu_k(A) \to \int_{\mathbb{P}} f(A) \, d\mu(A) \) for every bounded SOT-continuous real function \( f \) on \( \Sigma_\varepsilon \). Can we have \( M(\mu_k) \to M(\mu) \) in SOT, for instance, when \( M = G \)? This SOT-continuity property is a modification of (iii) without monotonicity assumption \( \mu_k \nearrow \mu_k \searrow \). The problem was raised in [31, Section 8] for the \( n \)-variable Karcher mean \( G_\mathbb{w} \): Is the map

\[
(A_1, \ldots, A_n) \in (\Sigma_\varepsilon)^n \mapsto G_\mathbb{w}(A_1, \ldots, A_n) = G\left( \sum_{j=1}^n w_j \delta_{A_j} \right)
\]

continuous in SOT? Note here that if \( A_j, A_{j,k} \in \Sigma_\varepsilon \) \( (1 \leq j \leq n, k \in \mathbb{N}) \) and \( A_{j,k} \to A_j \) in SOT as \( k \to \infty \), then \( \sum_{j=1}^n w_j \delta_{A_{j,k}} \to \sum_{j=1}^n w_j \delta_{A_j} \) weakly on \( \Sigma_\varepsilon \) with SOT. Note that the problem is true for any two-variable operator mean \( \sigma \) (in the Kubo-Ando sense), see Proposition 3.1 in Appendix A. For the probability measure case, we note that the SOT-continuity property stated above holds for \( M = \mathcal{A} \) and \( \mathcal{H} \), whose proof is not so easy and given in Appendix A for completeness.

6. Applications

In this section we apply the fixed point method presented in Theorem 3.1 to some important inequalities. To do so, it is convenient to introduce some classes of operator means on \( \mathcal{P}^\infty(\mathbb{P}) \).
6.1. Derived classes of operator means. To define some classes operator means on $\mathcal{P}^\infty(\mathbb{P})$, we consider the following two procedures:

(A) Deformation: from an operator mean $M$ on $\mathcal{P}^\infty(\mathbb{P})$ (satisfying (i)–(iv)) and a two-variable operator mean (in the Kubo-Ando sense) $\sigma \neq 1$, define the deformed operator mean $M_\sigma$. Then $M_\sigma$ is an operator mean on $\mathcal{P}^\infty(\mathbb{P})$ again by Theorem 3.1.

(B) Composition: from operator means $M_0, M_1, \ldots, M_n$ on $\mathcal{P}^\infty(\mathbb{P})$ and a weight $(w_1, \ldots, w_n)$ with any $n \in \mathbb{N}$, define $M(\mu) := M_0(\sum_{j=1}^n w_j \delta_{M_j(\mu)})$ for $\mu \in \mathcal{P}^\infty(\mathbb{P})$. Then it is immediate to see that $M$ is an operator mean on $\mathcal{P}^\infty(\mathbb{P})$ again.

**Definition 6.1.** We denote by $\mathfrak{M}(\mathcal{H})$, or simply $\mathfrak{M}$, the class of operator means on $\mathcal{P}^\infty(\mathbb{P}) = \mathcal{P}^\infty(\mathbb{P}(\mathcal{H}))$ obtained by starting from $\mathcal{A}, \mathcal{H}, G$ (see Sections 4.1 and 4.3) and applying procedures (A) and (B) finitely many times. We refer to an operator mean $M$ in the class $\mathfrak{M}$ as a derived operator mean on $\mathcal{P}^\infty(\mathbb{P})$.

The next proposition says that the derived operator means on $\mathcal{P}^\infty(\mathbb{P}(\mathcal{H}))$ are defined, in a sense, independently of the choice of (separable) $\mathcal{H}$.

**Proposition 6.2.** (1) Assume that $\mathcal{H}$ is isomorphic to another Hilbert space $\tilde{\mathcal{H}}$ with a unitary $U : \mathcal{H} \to \tilde{\mathcal{H}}$. For each $M \in \mathfrak{M}(\mathcal{H})$ let $\tilde{M}$ be the derived operator mean on $\mathcal{P}^\infty(\mathbb{P}(\tilde{\mathcal{H}}))$ defined by applying (A) and (B) in the same way as defining $M$. Then

$$\tilde{M}(U \mu U^*) = UM(\mu)U^*, \quad \mu \in \mathcal{P}^\infty(\mathbb{P}(\mathcal{H})), \tag{6.1}$$

where $U \mu U^*$ is the push-forward of $\mu$ by the unitary conjugation $U \cdot U^* : \mathbb{P}(\mathcal{H}) \to \mathbb{P}(\tilde{\mathcal{H}})$.

(2) Assume that $\mathcal{H} = H_1 \oplus H_2$ with Hilbert spaces $H_1, H_2$. For each $M \in \mathfrak{M}(\mathcal{H})$ let $M^{(i)}$ be the derived operator mean on $\mathcal{P}^\infty(\mathbb{P}(H_i))$, $i = 1, 2$, defined by applying (A) and (B) in the same way as defining $M$. Then

$$M(\mu_1 \oplus \mu_2) = M^{(1)}(\mu_1) \oplus M^{(2)}(\mu_2), \quad \mu_i \in \mathcal{P}^\infty(\mathbb{P}(H_i)), \quad i = 1, 2, \tag{6.2}$$

where $\mu_1 \oplus \mu_2$ is the push-forward of $\mu_1 \times \mu_2$ by the map $(A, B) \in \mathbb{P}(H_1) \times \mathbb{P}(H_2) \mapsto A \oplus B \in \mathbb{P}(\mathcal{H})$. In particular,

$$M(\mu_1 \oplus I_2) = M^{(1)}(\mu_1) \oplus I_2, \quad \mu_1 \in \mathcal{P}^\infty(\mathbb{P}(H_1)), \tag{6.3}$$

where $I_2$ is the identity operator on $H_2$.

**Proof.** (1) By their definitions (see Sections 4.1 and 4.3) it is immediate to see that $\mathcal{A}, \mathcal{H}$ and $G$ satisfy (6.1). Hence it suffices to show that property (6.1) is preserved by procedures (A) and (B), which is easily verified and left to the reader.

(2) It is convenient for us to consider, in addition to (6.2), the following property for $M \in \mathfrak{M}(\mathcal{H})$: for any weight $(w_1, \ldots, w_m)$, $A_k \in \mathbb{P}(H_1)$ and $B_k \in \mathbb{P}(H_2)$ ($1 \leq k \leq m$),

$$M \left( \sum_{k=1}^m w_k \delta_{A_k \oplus B_k} \right) = M^{(1)} \left( \sum_{k=1}^m w_k \delta_{A_k} \right) \oplus M^{(2)} \left( \sum_{k=1}^m w_k \delta_{B_k} \right). \tag{6.3}$$
We now show that (6.2) and (6.3) are preserved by procedures (A) and (B). Assume that \( M \in \mathcal{M}(\mathcal{H}) \) satisfies (6.2) and (6.3), and prove that \( M_\sigma \) does the same for any two-variable operator mean \( \sigma \neq 1 \). For \( \mu_i \in \mathcal{P}_\infty(\mathcal{P}(\mathcal{H}_i)) \) \( (i = 1, 2) \), let \( X_i := M_\sigma^{(i)}(\mu_i) \) so that \( X_i = M^{(i)}(X_i, \sigma \mu_i) \). Then

\[
X_1 \oplus X_2 = M^{(1)}(X_1, \sigma \mu_1) \oplus M^{(2)}(X_2, \sigma \mu_2)
= M((X_1, \sigma \mu_1) \oplus (X_2, \sigma \mu_2))
= M((X_1 \oplus X_2) \sigma(\mu_1 \oplus \mu_2)),
\]

where the last equality follows from the well-known property of two-variable operator means

\[
(X_1 \sigma A_1) \oplus (X_2 \sigma A_2) = (X_1 \oplus X_2) \sigma (A_1 \oplus A_2), \quad A_i \in \mathcal{P}(\mathcal{H}_i).
\]

Hence one has \( X_1 \oplus X_2 = M_\sigma(\mu_1 \oplus \mu_2) \) so that \( M_\sigma \) satisfies (6.2). To prove (6.3) for \( M_\sigma \), let \( Y_1 := M_\sigma^{(1)}(\sum_{k=1}^{m} w_k \delta A_k) \) and \( Y_2 := M_\sigma^{(1)}(\sum_{k=1}^{m} w_k \delta B_k) \). Then, by using (6.3) for \( M \) and (6.4), one has

\[
Y_1 \oplus Y_2 = M^{(1)}(Y_1 \sigma (\sum_{k=1}^{m} w_k \delta A_k)) \oplus M^{(2)}(Y_2 \sigma (\sum_{k=1}^{m} w_k \delta B_k))
= M^{(1)}(\sum_{k=1}^{m} w_k \delta Y_1 \sigma A_k) \oplus M^{(2)}(\sum_{k=1}^{m} w_k \delta Y_2 \sigma B_k)
= M\left((Y_1 \oplus Y_2) \sigma (\sum_{k=1}^{m} w_k \delta A_k \oplus B_k)\right),
\]

which implies that \( Y_1 \oplus Y_2 = M_\sigma(\sum_{k=1}^{m} w_k \delta A_k \oplus B_k) \), i.e., \( M_\sigma \) satisfies (6.3).

Next, to prove that (6.2) and (6.3) are preserved by procedure (B), assume that \( M_0, M_1, \ldots, M_n \in \mathcal{M}(\mathcal{H}) \) satisfy them, and let \( M \) be given as in (B) with a weight \((w_1, \ldots, w_n)\). For \( \mu_i \in \mathcal{P}_\infty(\mathcal{P}(\mathcal{H}_i)) \), by using (6.2) for \( M_j \) and (6.3) for \( M_0 \), one has

\[
M(\mu_1 \oplus \mu_2) = M_0\left(\sum_{j=1}^{m} w_j \delta M_j^{(1)}(\mu_1) \oplus M_j^{(2)}(\mu_2)\right)
= M_0^{(1)}\left(\sum_{j=1}^{n} w_j \delta M_j^{(1)}(\mu_1)\right) \oplus M_0^{(2)}\left(\sum_{j=1}^{n} w_j \delta M_j^{(2)}(\mu_2)\right)
= M^{(1)}(\mu_1) \oplus M^{(2)}(\mu_2).
\]

implying (6.2) for \( M \). Moreover, for any weight \((w'_1, \ldots, w'_m)\), by using (6.3) for \( M_j \) and \( M_0 \), one has

\[
M\left(\sum_{k=1}^{m} w'_k \delta A_k \oplus B_k\right)
= M_0\left(\sum_{j=1}^{n} w_j \delta M_j^{(1)}(\sum_{k=1}^{m} w'_k \delta A_k) \oplus M_j^{(2)}(\sum_{k=1}^{m} w'_k \delta B_k)\right).
\]
= M_0^{(1)} \left( \frac{1}{n} \sum_{j=1}^{n} w_j \delta_{M_j}^{(1)} \left( \sum_{k=1}^{m} w_k' \delta_{A_k} \right) \right) \oplus M_0^{(2)} \left( \frac{1}{n} \sum_{j=1}^{n} w_j \delta_{M_j}^{(2)} \left( \sum_{k=1}^{m} w_k' \delta_{B_k} \right) \right)

= M^{(1)} \left( \frac{1}{m} \sum_{k=1}^{m} w_k' \delta_{A_k} \right) \oplus M^{(2)} \left( \frac{1}{m} \sum_{k=1}^{m} w_k' \delta_{B_k} \right),

implying (6.3) for M. Thus, (6.2) has been shown for all $M \in \mathcal{M}(\mathcal{H})$.

In what follows, in view of Proposition 6.2, we write a derived operator mean $M \in \mathcal{M}$ in common for any choice of the underlying Hilbert space $\mathcal{H}$.

**Proposition 6.3.** Every operator mean $M \in \mathcal{M}$ satisfies all the properties in (i)–(viii), and $M \in \mathcal{M}$ implies $M^* \in \mathcal{M}$, where $M^*$ is the adjoint of $M$ (see Proposition 4.1 (3)).

**Proof.** First, note that $A, \mathcal{H}, G$ satisfies all (i)–(viii). From definition and Theorems 3.1 and 4.2, to prove the first assertion, it remains to show that the properties in (v)–(viii) are also preserved under procedure (B). It is immediate to see this for (v), (vi) and (viii). As for (vii), assume that operator means $M_0, M_1, \ldots, M_n$ satisfy (vii), and let $M$ be defined as in (B) with a weight $(w_1, \ldots, w_n)$. Let $\mu_k, \nu_k \in \mathcal{P}^\infty(\mathbb{P})$ ($1 \leq k \leq m$), and $(w_1', \ldots, w_m')$ be a weight, and let $0 < t < 1$. Then

$$M \left( \sum_{k=1}^{m} w_k' (\mu_k \nabla t \nu_k) \right) = M_0 \left( \sum_{j=1}^{n} w_j \delta_{M_j} \left( \sum_{k=1}^{m} w_k' \mu_k \right) \right)$$

$$\geq M_0 \left( \sum_{j=1}^{n} w_j \delta_{(1-t)M_j} \left( \sum_{k=1}^{m} w_k' \mu_k \right) + t M_j \left( \sum_{k=1}^{m} w_k' \nu_k \right) \right)$$

$$= M_0 \left( \sum_{j=1}^{k} w_j \delta_{M_j} \left( \sum_{k=1}^{m} w_k' \mu_k \right) \nabla t \delta_{M_j} \left( \sum_{k=1}^{m} w_k' \nu_k \right) \right)$$

$$\geq (1-t) M_0 \left( \sum_{j=1}^{n} w_j \delta_{M_j} \left( \sum_{k=1}^{m} w_k' \mu_k \right) \right)$$

$$+ t M_0 \left( \sum_{j=1}^{n} w_j \delta_{M_j} \left( \sum_{k=1}^{m} w_k' \nu_k \right) \right)$$

$$= (1-t) M \left( \sum_{k=1}^{m} w_k' \mu_k \right) + t M \left( \sum_{k=1}^{m} w_k' \nu_k \right),$$

so that $M$ satisfies (vii) again.

Next, note that the adjoint of $M$ defined in (B) is

$$M^* (\mu) = M_0^* \left( \sum_{j=1}^{n} w_j \delta_{M_j^*(\mu)} \right), \quad \mu \in \mathcal{P}^\infty(\mathbb{P}).$$

Since $A^* = \mathcal{H}$ and $G^* = G$, it follows from Proposition 4.1 (3) and the above expression that $M \in \mathcal{M}$ implies $M^* \in \mathcal{M}$.

□
Let $\sigma$ be a two-variable operator mean in the Kubo-Ando sense with the representing function $f_\sigma$. Following [10] we say that $\sigma$ is power monotone increasing (p.m.i. for short) if $f_\sigma(x^r) \geq f_\sigma(x)^r$ for all $x > 0$ and $r \geq 1$, and power monotone decreasing (p.m.d.) if $f_\sigma(x^r) \leq f_\sigma(x)^r$ for all $x > 0$ and $r \geq 1$. We say also that $\sigma$ is g.c.v. (resp., g.c.c) if $f_\sigma$ is geometrically convex (resp., geometrically concave), i.e., $f_\sigma(\sqrt{xy}) \leq \sqrt{f_\sigma(x)f_\sigma(y)}$ (resp., $f_\sigma(\sqrt{xy}) \geq \sqrt{f_\sigma(x)f_\sigma(y)}$) for all $x, y > 0$, that is, $\log f(e^t)$ is convex (resp., concave) on $t \in \mathbb{R}$. It is clear that $\sigma$ is p.m.i. if and only if $\sigma^*$ is p.m.d., and $\sigma$ is g.c.v. if and only if $\sigma^*$ is g.c.c. Note that $\forall_\alpha$ is g.c.v. and $!_\alpha$ is g.c.c. for any $\alpha \in [0, 1]$.

Moreover, the two-variable operator means that are simultaneously p.m.i. and p.m.d. are only $\#_\alpha$ ($0 \leq \alpha \leq 1$). It is easy to see that g.c.v. implies p.m.i. and g.c.c. implies p.m.d. for $\sigma$. But in [17] Wada recently proved that the converse is not true, that is, there is a p.m.i. $\sigma$ that is not g.c.v.

**Definition 6.4.**

1. We denote by $\mathcal{M}^+$ (resp., $\mathcal{M}^-$) the subclass of $\mathcal{M}$ obtained by starting from $\mathcal{A}, G$ (resp., $\mathcal{H}, G$) and applying finitely many times procedure (A) with $\sigma$ restricted to g.c.v. (resp., g.c.c.) and procedure (B). Note that $M \in \mathcal{M}^+$ if and only if $M^* \in \mathcal{M}^-$, and $P_r \in \mathcal{M}^+$ and $P_{-r} \in \mathcal{M}^-$ for $0 < r \leq 1$ (see Section 4.2).

2. We denote by $\mathcal{M}^+_0$ (resp., $\mathcal{M}^-_0$) the subclass of $\mathcal{M}$ obtained by starting from $\mathcal{A}, G$ (resp., $\mathcal{H}, G$) and applying finitely many times procedure (A) with $\sigma$ restricted to p.m.i. (resp., p.m.d.), where procedure (B) is not applied. Note that $M \in \mathcal{M}^+_0$ if and only if $M^* \in \mathcal{M}^-_0$, and $P_r \in \mathcal{M}^+_0$ and $P_{-r} \in \mathcal{M}^-_0$ for $0 < r \leq 1$.

### 6.2. Inequality under positive linear maps.

Let $\mathcal{H}$ and $\mathcal{K}$ be separable Hilbert spaces, and let $\Phi : B(\mathcal{H}) \to B(\mathcal{K})$ be a normal positive linear map. Here, $\Phi$ is normal if $A_k \not\succ A$ in $B(\mathcal{H})^+$ implies $\Phi(A_k) \not\succ \Phi(A)$ in $B(\mathcal{K})^+$. Assume that $\Phi(I_\mathcal{H})$ is invertible, where $I_\mathcal{H}$ is the identity operator on $\mathcal{H}$. Then $\Phi$ maps $\mathbb{P}(\mathcal{H})$ into $\mathbb{P}(\mathcal{K})$. In the case where $\mathcal{H}$ is finite-dimensional, the normality assumption is automatic and the invertibility assumption of $\Phi(I_\mathcal{H})$ is not essential. In fact, let $P_0$ be the support projection of $\Phi(I_\mathcal{H})$; then $\Phi$ may be considered as a map from $B(\mathcal{H})$ to $B(P_0\mathcal{K})$ and we may replace $\mathcal{K}$ with $P_0\mathcal{K}$.

For any two-variable operator mean $\sigma$ and any positive linear map $\Phi : B(\mathcal{H}) \to B(\mathcal{K})$, the following inequality is well-known:

$$\Phi(A\sigma B) \leq \Phi(A)\sigma\Phi(B), \quad A, B \in \mathbb{P}(\mathcal{H}),$$

which is essentially due to Ando [1] while proved only for the geometric and the harmonic means.

**Theorem 6.5.** Let $\Phi$ be as stated above. The for every derived operator mean $M \in \mathcal{M}$,

$$\Phi(M(\mu)) \leq M(\Phi_*\mu), \quad \mu \in \mathcal{P}^\infty(\mathbb{P}(\mathcal{H})), \quad (6.6)$$

where $\Phi_*\mu$ is the push-forward of $\mu$ by the map $A \in \mathbb{P}(\mathcal{H}) \mapsto \Phi(A) \in \mathbb{P}(\mathcal{K})$.

**Proof.** Define $\Psi : B(\mathcal{H}) \to B(\mathcal{K})$ by $\Psi(A) := \Phi(I_\mathcal{H})^{-1/2}\Phi(A)\Phi(I_\mathcal{H})^{-1/2}$; then $\Psi$ is a unital positive map. By congruence invariance (vi) (see Section 4), inequality (6.6) is equivalent to $\Psi(M(\mu)) \leq M(\Psi_*\mu)$. Hence we may assume that $\Phi$ is unital. We
prove that (6.6) is preserved under procedure (A), that is, if \( M \) satisfies (6.6), then the deformed operator mean \( M_\sigma \) does the same for any operator mean \( \sigma \neq 1 \). For every \( \mu \in \mathcal{P}(\mathcal{H}) \) let \( X_0 := M_\sigma(\mu) \). Then
\[
\Phi(X_0) = \Phi(M(X_0\sigma\mu)) \leq M(\Phi_*(X_0\sigma\mu)). \tag{6.7}
\]
Let \( \psi_{X_0}(A) := X_0\sigma A \) for \( A \in \mathcal{P}(\mathcal{H}) \) and \( \psi_{\Phi(X_0)}(B) := \Phi(X_0)\sigma B \) for \( B \in \mathcal{P}(\mathcal{K}) \). Note that
\[
\Phi_*(X_0\sigma\mu) = (\Phi \circ \psi_{X_0})_\mu, \quad \Phi(X_0)\sigma(\Phi_\mu) = (\psi_{\Phi(X_0)} \circ \Phi)_\mu.
\]
By inequality (6.5) we find that
\[
(\Phi \circ \psi_{X_0})(A) = \Phi(X_0\sigma A) \leq \Phi(X_0)\sigma(\Phi(A) = (\psi_{\Phi(X_0)} \circ \Phi)(A)
\]
for all \( A \in \mathcal{P}(\mathcal{H}) \). Hence it follows from Lemma 3.4 that
\[
\Phi_*(X_0\sigma\mu) \leq \Phi(X_0)\sigma(\Phi_\mu).
\]
By monotonicity of \( M \) this implies that
\[
M(\Phi_*(X_0\sigma\mu)) \leq M(\Phi(X_0)\sigma(\Phi_\mu)). \tag{6.8}
\]
By (6.7) and (6.8), \( \Phi(X_0) \leq M(\Phi(X_0)\sigma(\Phi_\mu)) \), which implies by Theorem 3.1(2) that \( \Phi(X_0) \leq M_\sigma(\Phi_\mu) \).

Next, it is immediate to see that (6.6) is preserved under procedure (B) as follows:
\[
\Phi(M(\mu)) \leq M_0 \left( \Phi_* \left( \sum_{j=1}^n w_j \delta_{M_j(\mu)} \right) \right) = M_0 \left( \sum_{j=1}^n w_j \delta_{\Phi(M_j(\mu))} \right) \\
\leq M_0 \left( \sum_{j=1}^n w_j \delta_{\Phi_\mu} \right) = M(\Phi_\mu).
\]

Therefore, to prove the theorem, it remains to show that \( \mathcal{A}, \mathcal{H} \) and \( \mathcal{G} \) satisfy (6.6). This is trivial for \( \mathcal{A} \). Apply procedure (A) to \( M = \mathcal{A} \) and \( \sigma = 1 \), for \( 0 < s \leq 1 \); then we have \( \Phi(\mathcal{A}_{1_s}(\mu)) \leq \mathcal{A}_{1_s}(\Phi_\mu) \). Letting \( s \searrow 0 \) gives \( \Phi(\mathcal{H}(\mu)) \leq \mathcal{H}(\Phi_\mu) \) thanks to Proposition 3.1 since \( \Phi \) is normal. Also, apply (A) to \( M = \mathcal{A} \) and \( \sigma = \#_r \) for \( 0 < r \leq 1 \); then \( \Phi(P_r(\mu)) \leq P_r(\Phi_\mu) \). Letting \( r \searrow 0 \) gives \( \Phi(G(\mu)) \leq G(\Phi_\mu) \) thanks to Proposition 5.3.

\[\square\]

**Remark 6.6.** The normality of \( \Phi \) has been used only to prove (6.6) for \( \mathcal{H} \) and \( \mathcal{G} \) in the last part of the above proof. So, once \( \mathcal{H} \) and \( \mathcal{G} \) satisfy (6.6) without the normality assumption of \( \Phi \), we can remove this assumption from Theorem 6.5. For instance, in view of definition of the power means \( P_r \) in (5.3), note that \( P_r \) for \( r \in (0, 1] \) satisfies (6.6) without the normality of \( \Phi \). But it is unknown to us whether this is also the case for \( P_r \) for \( r \in [-1, 0] \), in particular, for \( \mathcal{H} \). In a different approach in [10] Theorem 6.4 it was shown that (6.6) holds for a certain wide class of operator means on \( \mathcal{P}_\infty(\mathbb{P}) \) under unital positive linear maps \( \Phi \), but the normality of \( \Phi \) seems necessary there.
6.3. Ando-Hiai’s inequality. In [46] Wada proved the extended version of Ando-Hiai’s inequality [3] in such a way that, for a two-variable operator mean $\sigma$ (in the Kubo-Ando sense), the following conditions are equivalent:

(i) $\sigma$ is p.m.i. (resp., p.m.d.) (see the paragraph before Definition 6.4);
(ii) for every $A, B \in \mathbb{P}$, $A\sigma B \geq I \implies A^r\sigma B^r \geq I$ (resp., $A\sigma B \leq I \implies A^r\sigma B^r \leq I$) for all $r \geq 1$.

Ando-Hiai’s inequality for the $n$-variable Karcher mean was proved by Yamazaki [48], which was extended to the case of probability measures in [25, 19] when $\dim \mathcal{H} < \infty$. In [37] Lim and Yamazaki discussed the Ando-Hiai type inequalities for the $n$-variable power means when $\dim \mathcal{H} < \infty$. Furthermore, in a recent paper [22] a comprehensive study on the Ando-Hiai type inequalities for $n$-variable operator means has been made by a similar fixed point method to this paper. Similar Ando-Hiai type inequalities have been independently shown by Yamazaki [49] based on the generalized Karcher equations in [10].

The next theorem is the extension of Ando-Hiai’s inequality in [25, 19] to the infinite-dimensional case and to a wider class of operator means on $\mathcal{P}^\infty(\mathbb{P})$ including power means, as well as the extension of [22, Theorem 3.1] from $n$-variable operator means to operator means on $\mathcal{P}^\infty(\mathbb{P})$.

**Theorem 6.7.** If $M \in \mathcal{M}_0^+$, then for every $\mu \in \mathcal{P}^\infty(\mathbb{P})$,

$$M(\mu) \geq I \implies M(\mu^r) \geq I, \quad r \geq 1. \quad (6.9)$$

Also, if $M \in \mathcal{M}_0^-$, then for every $\mu \in \mathcal{P}^\infty(\mathbb{P})$,

$$M(\mu) \leq I \implies M(\mu^r) \leq I, \quad r \geq 1. \quad (6.10)$$

**Proof.** We show that if $M \in \mathcal{M}$ satisfies (6.9), then $M_\sigma$ does the same for every p.m.i. two-variable operator mean $\sigma$. For every $\mu \in \mathcal{P}^\infty(\mathbb{P})$ assume that $X_0 := M_\sigma(\mu) \geq I$. We first prove the case where $1 \leq r \leq 2$. Since $X_0 = M(X_0\sigma\mu)$ so that $I = M(I\sigma(X_0^{-1/2}\mu X_0^{-1/2}))$, we have $I \leq M((I\sigma(X_0^{-1/2}\mu X_0^{-1/2}))^r)$ for all $r \geq 1$. Note that

$$\begin{align*}
(I\sigma(X_0^{-1/2}\mu X_0^{-1/2}))^r &= (\pi_r \circ f_\sigma \circ \Gamma_{X_0^{-1/2}})^r \mu, \\
I\sigma(X_0^{-1/2} \mu X_0^{-1/2}) &= (\pi_r \circ \Gamma_{X_0^{-1/2}})^r \mu,
\end{align*} \quad (6.11)\quad (6.12)$$

where $\Gamma_{X_0^{-1/2}}(A) := X_0^{-1/2}AX_0^{-1/2}$, $\pi_r(A) := A^r$, and $f_\sigma(A)$ is the functional calculus of $A$ by the representing function $f_\sigma$. Since $f_\sigma(x)^r \leq f_\sigma(x^r)$ for $x > 0$, we find that

$$\begin{align*}
(\pi_r \circ f_\sigma \circ \Gamma_{X_0^{-1/2}})(A) &= f_\sigma(X_0^{-1/2}AX_0^{-1/2})^r \\
&\leq f_\sigma((X_0^{-1/2}AX_0^{-1/2})^r) \\
&\leq f_\sigma(X_0^{-1/2}A^rX_0^{-1/2}) \\
&= (f_\sigma \circ \Gamma_{X_0^{-1/2}} \circ \pi_r)(A), \quad A \in \mathbb{P},
\end{align*}$$

where $\Gamma_{X_0^{-1/2}}(A) := X_0^{-1/2}AX_0^{-1/2}$, $\pi_r(A) := A^r$, and $f_\sigma(A)$ is the functional calculus of $A$ by the representing function $f_\sigma$. Since $f_\sigma(x)^r \leq f_\sigma(x^r)$ for $x > 0$, we find that
where we have used Hansen-Pedersen’s inequality \[15\] Theorem 2.1 for the above latter
inequality thanks to \(X_0^{-1/2} \leq I\). Applying Lemma 3.4 to (6.11) and (6.12) implies that
\[(I\sigma(X_0^{-1/2} \mu X_0^{-1/2}))^r \leq I\sigma(X_0^{-1/2} \mu^r X_0^{-1/2}).\]
The monotonicity of \(M\) gives
\[I \leq M((I\sigma(X_0^{-1/2} \mu X_0^{-1/2}))^r) \leq M(I\sigma(X_0^{-1/2} \mu^r X_0^{-1/2}))\]
so that \(X_0 \leq M(X_0 \sigma \mu^r)\), which implies by Theorem 3.1 (2) that \(X_0 \leq M_\sigma(\mu^r)\) and
hence \(M_\sigma(\mu^r) \geq I\). For general \(r \geq 1\) write \(r = 2^k r_0\) where \(k \in \mathbb{N}\) and \(1 \leq r_0 < 2\), and
iterate the case \(1 \leq r \leq 2\) to obtain (6.9) for \(M_\sigma\).

In view of Definition 6.4 to prove (6.9) for any \(M \in M_0^+\), it remains to show that
\(A\) and \(G\) satisfy (6.9). For \(A\), when \(1 \leq r \leq 2\), the operator convexity of \(x^r\) on
\((0, \infty)\) gives \(A(\mu)^r \leq A(\mu^r)\), so that (6.9) for \(A\) holds in this case. The general case
\(r \geq 1\) follows by iteration as in the last of the first part of the proof. Next, apply the
procedure proved in the first part to \(M = A\) and \(\sigma = \#_\alpha\) for \(0 < \alpha \leq 1\); then (6.9)
holds for \(P_\alpha\) for \(0 < \alpha \leq 1\). Now assume that \(G(\mu) \geq I\). For every \(\alpha \in (0, 1]\), since
\(P_\alpha(\mu) \geq G(\mu) \geq I\), we have \(P_\alpha(\mu^r) \geq I\) for all \(r \geq 1\). Letting \(\alpha \searrow 0\) gives \(G(\mu^r) \geq I\)
for all \(r \geq 1\) thanks to Proposition 5.3.

The latter assertion immediately follows from the first, since (6.9) for \(M\) is equivalent
to (6.10) for \(M^*\), and \(M \in M_0^+ \iff M^* \in M_0^-\). \(\square\)

**Remark 6.8.** Note that the proof of Theorem 6.7 indeed verifies a slightly stronger
result that, for every \(\mu \in \mathcal{P}^\infty(\mathbb{P})\), if \(M \in M_0^+\) then
\[M(\mu) \geq I \implies M(\mu^r) \geq M(\mu), \quad r \geq 1,\]
and if \(M \in M_0^-\) then
\[M(\mu) \leq I \implies M(\mu^r) \leq M(\mu), \quad r \geq 1.\]
These are equivalently stated in such a way that, for every \(\mu \in \mathcal{P}^\infty(\mathbb{P})\), if \(M \in M_0^+\) then
\[M(\mu^r) \geq \lambda_{\min}(M(\mu))^{r-1} M(\mu), \quad r \geq 1,\]
and if \(M \in M_0^-\) then
\[M(\mu^r) \leq \|M(\mu)\|^{r-1} M(\mu), \quad r \geq 1.\]
In the above, \(\lambda_{\min}(A)\) is the minimum of the spectrum of \(A \in \mathbb{P}\).

**Corollary 6.9.** Let \(\alpha \in (0, 1]\). Then for every \(\mu \in \mathcal{P}^\infty(\mathbb{P})\),
\[P_\alpha(\mu) \geq I \implies P_\alpha(\mu^r) \geq I, \quad r \geq 1, \quad (6.13)\]
\[P_{-\alpha}(\mu) \leq I \implies P_{-\alpha}(\mu^r) \leq I, \quad r \geq 1. \quad (6.14)\]

**Remark 6.10.** Corollary 6.9 has been shown based on Theorem 3.1 applied to the
simple case \(M = A\). Since \(P_\alpha \searrow G\) and \(P_{-\alpha} \nearrow G\) as \(r \searrow 0\) by Proposition 5.3 the
corollary in turn gives Ando-Hiai’s inequality for \(G\). This is a new proof of Ando-Hiai’s
inequality even for the two-variable geometric mean.
6.4. Modified Ando-Hiai’s inequalities. We here present two more Ando-Hiai type inequalities, which are weaker than Theorem 6.7 (also Remark 6.8) in the sense that inequalities are between two deformed operator means \( M_\sigma \) and \( M_{\sigma_r} \) on \( \mathcal{P}^\infty(\mathbb{P}) \), where \( 0 < r \leq 1 \) and \( \sigma_r \) is the modified two-variable operator mean with the representing function \( f_\sigma(x^r) \). But instead, those have an advantage since there are no restrictions on \( M \) (except congruence invariance) and \( \sigma \) unlike in Theorem 6.7.

**Theorem 6.11.** Let \( M \) be an operator mean on \( \mathcal{P}^\infty(\mathbb{P}) \) satisfying (vi) (congruence invariance) in Section 4, and \( \sigma \) be any two-variable operator mean with \( \sigma \neq 1 \). Then for every \( \mu \in \mathcal{P}^\infty(\mathbb{P}) \),

\[
M_\sigma(\mu) \geq I \implies M_{\sigma_{1/r}}(\mu^r) \geq M_\sigma(\mu), \quad r \geq 1,
\]

\[
M_\sigma(\mu) \leq I \implies M_{\sigma_{1/r}}(\mu^r) \leq M_\sigma(\mu), \quad r \geq 1.
\]

**Proof.** Let \( \mu \in \mathcal{P}^\infty(\mathbb{P}) \) and \( r \geq 1 \). The assertion in (6.16) follows from (6.15) by replacing \( M_\sigma \), \( \sigma \) and \( \mu \) in (6.15) with \( M^*, \sigma^* \) and \( \mu^{-1} \). Indeed, note that \( M^* \) satisfies (vi) too and \( (\sigma^*)_{1/r} = (\sigma_{1/r})^* \) so that \( (M^*)_{(\sigma^*)_{1/r}} = (M_{\sigma_{1/r}})^* \) as well as \( M_{\sigma^*} = (M_\sigma)^* \) by Proposition 4.4 (3). Hence (6.16) is equivalent to (6.15) for \( M^*, \sigma^* \) and \( \mu^{-1} \). So we may prove (6.15) only. Assume that \( X_0 := M_\sigma(\mu) \geq I \); then \( I = M(I\sigma(X_0^{-1/2}X_0^{-1/2})) \). Using the same notations as in the proof of Theorem 6.7, we have \( I\sigma(X_0^{-1/2}\mu X_0^{-1/2}) = (f_\sigma \circ \Gamma_{X_0^{-1/2}})_* \mu \) and

\[
I\sigma_{1/r}(X_0^{-1/2}\mu^r X_0^{-1/2}) = (f_\sigma \circ \pi_{1/r} \circ \Gamma_{X_0^{-1/2}} \circ \pi_r)_* \mu.
\]

We have

\[
(f_\sigma \circ \pi_{1/r} \circ \Gamma_{X_0^{-1/2}} \circ \pi_r)(A) = f_\sigma((X_0^{-1/2}AX_0^{-1/2})^{1/r})
\]

\[
\geq f_\sigma(X_0^{-1/2}AX_0^{-1/2})
\]

\[
= (f_\sigma \circ \Gamma_{X_0^{-1/2}})(A), \quad A \in \mathbb{P},
\]

where Hansen’s inequality [14] has been used for the above inequality since \( X_0^{-1/2} \leq I \) and \( 0 < 1/r \leq 1 \). Therefore, Lemma 3.1 gives

\[
I\sigma(X_0^{-1/2}\mu X_0^{-1/2}) \leq I\sigma_{1/r}(X_0^{-1/2}\mu^r X_0^{-1/2}),
\]

so that

\[
I = M(I\sigma(X_0^{-1/2}\mu X_0^{-1/2})) \leq M(I\sigma_{1/r}(X_0^{-1/2}\mu^r X_0^{-1/2})).
\]

This implies that \( X_0 \leq M(X_0 \sigma_{1/r} \mu^r) \) so that \( X_0 \leq M_{\sigma_{1/r}}(\mu^r) \) by Theorem 3.1 (2). \( \Box \)

**Remark 6.12.** Similarly to Remark 6.8 the assertions (6.15) and (6.16) together are equivalently stated as

\[
\lambda_{\min}^{-1}(M_\sigma(\mu))M_\sigma(\mu) \leq M_{\sigma_{1/r}}(\mu^r) \leq \|M_\sigma(\mu)\|^{-1}M_\sigma(\mu), \quad r \geq 1.
\]

This is the extension of [22, Theorem 4.1] from the \( n \)-variable case to the probability measure case. Note that the special case of (6.17) for \( n \)-variable power means was first shown in [37, Corollary 3.2] when \( \dim \mathcal{H} < \infty \).
The next result is the complementary version of [6.17], which is the extension of [22, Theorem 4.2] from the $n$-variable case to the probability measure case.

**Theorem 6.13.** Let $M$ and $\sigma$ be as in Theorem 6.11. For every $\mu \in \mathcal{P}^\infty(\mathbb{P})$,
\[
\|M_\sigma(\mu)\|^{r-1}M_\sigma(\mu) \leq M_\sigma(\mu^r) \leq \lambda_{\min}^{-1}(M_\sigma(\mu))M_\sigma(\mu), \quad 0 < r \leq 1.
\] (6.18)

**Proof.** As in the proof of Theorem 6.11, the first and the second inequalities in (6.18) are equivalent by replacing $M$, $\sigma$ and $\mu$ with $M^*$, $\sigma^*$ and $\mu^{-1}$, so we may prove the second only. Let $\mu \in \mathcal{P}^\infty(\mathbb{P})$ and $0 < r \leq 1$. Moreover, let $X_0 := M_\sigma(\mu)$ and $\lambda := \lambda_{\min}(X_0)$. We have $I = M(I\sigma_r(X_0^{-1/2} \mu X_0^{-1/2}))$ and
\[
I\sigma_r(X_0^{-1/2} \mu X_0^{-1/2}) = (f_\sigma \circ \pi_r \circ \Gamma_{X_0^{-1/2}})_* \mu.
\]
By Theorem 3.1(2) what we need to prove is that $\lambda^{r-1}X_0 \geq M((\lambda^{r-1}X_0)\sigma \mu^r)$, or equivalently, $I \geq M(I\sigma(\lambda^{1-r}(X_0^{-1/2} \mu X_0^{-1/2})))$. Note that
\[
I\sigma(\lambda^{1-r}.X_0^{-1/2} \mu X_0^{-1/2}) = (f_\sigma \circ m_{\lambda^{1-r}} \circ \Gamma_{X_0^{-1/2}} \circ \pi_r)_* \mu,
\]
where $m_{\lambda^{1-r}}(A) := \lambda^{1-r}A$. Since $\lambda^{1/2}X_0^{-1/2} \leq I$, Hansen’s inequality [14] gives
\[
\lambda^{1-r}X_0^{-1/2} A^r X_0^{-1/2} \leq (X_0^{-1/2} A X_0^{-1/2})^r,
\]
which implies that
\[
(f_\sigma \circ m_{\lambda^{1-r}} \circ \Gamma_{X_0^{-1/2}} \circ \pi_r)(A) = f_\sigma((\lambda^{1-r}X_0^{-1/2} A^r X_0^{-1/2})^r)
\leq f_\sigma((X_0^{-1/2} A X_0^{-1/2})^r) = (f_\sigma \circ \pi_r \circ \Gamma_{X_0^{-1/2}})(A), \quad A \in \mathbb{P}.
\]
Therefore, by Lemma 3.4 we have
\[
I = M(I\sigma_r(X_0^{-1/2} \mu X_0^{-1/2})) \geq M(I\sigma(\lambda^{1-r}(X_0^{-1/2} \mu X_0^{-1/2}))),
\]
as required. \qed

**Remark 6.14.** Similarly to the case $\dim \mathcal{H} < \infty$ in [20, Theorem 3.4], it is easy to verify that the Karcher mean $G$ satisfies $G(\mu) = G_{\#\alpha}(\mu)$ for all $\mu \in \mathcal{P}^\infty(\mathbb{P})$ and any $\alpha \in (0, 1]$. Since $(\#\alpha)_r = \#_{\alpha r}$ for all $r \in (0, 1]$, either Theorem 6.11 or 6.13 gives Ando-Hiai’s inequality for $G$, in addition to an demonstration in Remark 6.10.

## 7. Further Applications

In this section we present more applications of our method to different inequalities. Throughout the section, for some theoretical and technical reasons, we assume that $\mathcal{H}$ is finite-dimensional.
7.1. Norm inequality. For every unitarily invariant norm $||| \cdot |||$ and any two-variable operator mean $\sigma$, it is well-known (see, e.g., [2, (3.13)]) that

$$|||A \sigma B||| \leq |||A|||\sigma|||B|||, \quad A, B \in B(\mathcal{H})^+. \quad (7.1)$$

In fact, this norm inequality can be extended to more general norms $||| \cdot |||$ on $B(\mathcal{H})$ that is monotone in the sense that if $A \geq B \geq 0$ implies $|||A||| \geq |||B|||$. There are many examples of monotone norms on $B(\mathcal{H})$ that are not unitarily invariant; for example, the numerical radius and $|||A||| = \|A\| + |\text{tr}A|$, where $\text{tr}A$ is the trace of $A$, are such cases. (These are examples of weakly unitarily invariant norms introduced in [6].) Although the extension of (7.1) to monotone norms may be folklore to experts, there seems no literature, so we prove it as a lemma.

Lemma 7.1. If $||| \cdot |||$ is a monotone norm on $B(\mathcal{H})$, then inequality (7.1) holds for every two-variable operator mean $\sigma$.

Proof. By continuity we may assume that $A, B \in B(\mathcal{H})^+$ are invertible. Since

$$(I + A^{1/2}B^{-1}A^{1/2})^{-1} \leq t^2 I + (1 - t)^2 A^{-1/2}BA^{-1/2},$$

we have $A : B \leq t^2 A + (1 - t)^2 B$ for all $t \in \mathbb{R}$, where $A : B := (A^{-1} + B^{-1})^{-1}$, the parallel sum of $A, B$. Therefore,

$$|||A : B||| \leq t^2 |||A||| + (1 - t)^2 |||B|||.$$ 

Minimizing the right-hand side above gives $|||A : B||| \leq |||A||| : |||B|||$. Recall the integral expression [27]

$$A \sigma B = aA + bB + \int_{(0, \infty)} \frac{1 + t}{t} \{(tA) : B\} \, dm(t) \quad (7.2)$$

with a probability measure $m$ on $[0, \infty]$ where $a := m(\{0\}), b := m(\{\infty\})$. From this integral expression we have inequality (7.1) as

$$|||A \sigma B||| \leq a|||A||| + b|||B||| + \int_{(0, \infty)} \frac{1 + t}{t} \{(t|||A|||) : |||B|||\} \, dm(t)$$

$$= |||A|||\sigma|||B|||.$$

□

The following is the extension of (7.1) to derived operator means $M \in \mathfrak{M}$.

Proposition 7.2. For every derived operator mean $M \in \mathfrak{M}$ and every monotone norm $||| \cdot |||$ on $B(\mathcal{H})$,

$$|||M(\mu)||| \leq M(||| \cdot |||\mu), \quad \mu \in \mathcal{P}^\infty(\mathbb{P}), \quad (7.3)$$

where $||| \cdot |||\mu$ is the push-forward of $\mu$ by $A \in \mathbb{P} \mapsto |||A||| \in (0, \infty)$ (so $||| \cdot |||\mu$ is a Borel probability measure on $(0, \infty)$).
Proof. We show that (7.3) is preserved under procedure (A). Assume that \( M \in \mathcal{M} \) satisfies (7.3) and \( \sigma \) is a two-variable operator mean with \( \sigma \neq I \). For every \( \mu \in \mathcal{P}^\infty (\mathbb{P}) \) let \( X_0 := M_{\sigma} (\mu) \). Then
\[
\|\| X_0 \|\| = \|\| M (X_0 \sigma \mu) \|\| \leq M (\|\| \cdot \|\| \sigma (X_0 \sigma \mu)) \tag{7.4}
\]
Note that
\[
\|\| \cdot \|\| \sigma (X_0 \sigma \mu) = (\|\| \cdot \|\| \sigma (X_0 \sigma \mu)) = (\psi_{\|\| X_0 \|\|} \circ \|\| \cdot \|\| \sigma (X_0 \sigma \mu))
\]
where \( \psi_{\|\| X_0 \|\|} \) is as in the proof of Theorem 6.5 and \( \psi_{\|\| X_0 \|\|} \) is similarly defined on \((0, \infty)\). It follows from Lemma 7.1 that
\[
(\|\| \cdot \|\| \sigma (X_0 \sigma \mu)) = (\|\| X_0 \sigma A \|\| \leq \|\| X_0 \|\| \sigma (\|\| \cdot \|\| \sigma (A) \|\|)) = (\psi_{\|\| X_0 \|\|} \circ \|\| \cdot \|\| \sigma (A) \|\|) (A)
\]
for all \( A \in \mathbb{P} \), so that \( \|\| \cdot \|\| \sigma (X_0 \sigma \mu) \leq \|\| X_0 \|\| \sigma (\|\| \cdot \|\| \sigma (A) \|\|) \) by Lemma 3.4. From this and (7.4) one has
\[
\|\| X_0 \|\| \leq M (\|\| X_0 \|\| \sigma (\|\| \cdot \|\| \sigma (A) \|\|)).
\]
By Theorem 6.5(2) (for the case \( \dim \mathcal{H} = 1 \)) this implies that \( \|\| X_0 \|\| \leq M_{\sigma} (\|\| \cdot \|\| \sigma (A) \|\|). \) The remaining proof is similar to the second and the third paragraphs of the proof of Theorem 6.5, so the details may be left to the reader. \( \square \)

In particular, for the Karcher mean \( \Gamma \) and for any monotone norm \( \|\| \cdot \|\| \) on \( B(\mathcal{H}) \), we have
\[
\|\| \Gamma (\mu) \|\| \leq \Gamma (\|\| \cdot \|\| \sigma (A) \|\|) = \exp \int_{\mathbb{P}} \log \|\| A \|\| \ d\mu (A), \ \ \mu \in \mathcal{P}^\infty (\mathbb{P}), \tag{7.5}
\]
as verified in [33], where the last equality is readily seen.

7.2. Eigenvalue majorizations. Assume that \( N = \dim \mathcal{H} \). Let us first recall the notion of majorization. Let \( a = (a_1, \ldots, a_N), \ b = (b_1, \ldots, b_N) \in (\mathbb{R}^+)^N \). We write \( a^k = (a_1^k, \ldots, a_N^k) \) for the decreasing rearrangement of \( a \) and \( a^k = (a_1^k, \ldots, a_N^k) \) for its increasing counterpart. The weak majorization (or submajorization) \( a \prec_w b \) means that
\[
\sum_{i=1}^k a_i^k \leq \sum_{i=1}^k b_i^k, \ \ 1 \leq k \leq N.
\]
We call the majorization \( a \prec b \) if \( a \prec_w b \) and equality holds for \( k = N \) above. The log-majorization \( a \prec_{\log} b \) is defined as
\[
\prod_{i=1}^k a_i^k \leq \prod_{i=1}^k b_i^k, \ \ 1 \leq k \leq N-1, \ \ \text{and} \ \ \prod_{i=1}^N a_i^k = \prod_{i=1}^N b_i^k.
\]
Note that \( a \prec_{\log} b \implies a \prec_w b \).

Let \( A, B \in \mathbb{P} \) and \( \lambda (A) = (\lambda_1 (A), \ldots, \lambda_N (A)) \) be the eigenvalues of \( A \) in decreasing order counting multiplicities. In the following, we identify \( \lambda (A) \) in \((0, \infty)^N \) with a diagonal matrix in \( \mathbb{P} \) with \( \lambda_1 (A), \ldots, \lambda_N (A) \) on its diagonal. We write \( A \prec_w B, A \prec B \) and \( A \prec_{\log} B \) if each majorization for \( \lambda (A) \) and \( \lambda (B) \) holds, respectively. It is well-known that if \( A \prec_w B \) then \( \|\| f (A) \|\| \leq \|\| f (B) \|\| \) for every unitarily invariant norm.
and every non-negative and non-decreasing convex function $f$ on $(0, \infty)$. See, e.g., [2, 16, 38] for more about majorizations for matrices.

The Log-Euclidean mean of $\mu \in \mathcal{P}^\infty(\mathbb{P})$ is given as

$$LE(\mu) := \exp \int_\mathbb{P} \log A \, d\mu(A).$$

Recall the Lie-Trotter formula given in [19, Theorem 5.7]

$$\lim_{r \searrow 0} G(\mu^r)^{1/r} = LE(\mu). \quad (7.6)$$

For $\mu \in \mathcal{P}^\infty(\mathbb{P})$ let $\lambda_\ast \mu$ denote the push-forward of $\mu$ by the continuous map $A \in \mathbb{P} \mapsto \lambda(A) \in (0, \infty)^N$, that is,

$$\lambda_\ast \mu = ((\lambda_1)_\ast \mu, \ldots, (\lambda_N)_\ast \mu),$$

where $(\lambda_i)_\ast \mu$ is the push-forward of $\mu$ by $A \in \mathbb{P} \mapsto \lambda_i(A) \in (0, \infty)$. It is readily seen that

$$G(\lambda_\ast \mu) = LE(\lambda_\ast \mu) = \left( \exp \int_\mathbb{P} \log \lambda_i(A) \, d\mu(A) \right)^N_{i=1}. \quad (7.7)$$

**Proposition 7.3.** For every $\mu \in \mathcal{P}^\infty(\mathbb{P})$ and every $r \in (0, 1)$,

$$G(\mu) \prec_{\log} G(\mu^r)^{1/r} \prec_{\log} LE(\mu) \prec_{\log} G(\lambda_\ast \mu).$$

**Proof.** It was proved in [19, Theorem 4.4] that

$$G(\mu) \prec_{\log} G(\mu^{r'})^{1/r} \prec_{\log} G(\mu^{r''})^{1/r''} \quad \text{for } 0 < r' < r < 1.$$

Combining with (7.6) we have

$$G(\mu) \prec_{\log} G(\mu^r)^{1/r} \prec_{\log} LE(\mu) \quad \text{for } 0 < r < 1.$$

It remains to prove that $G(\mu^r)^{1/r} \prec_{\log} G(\lambda_\ast \mu)$ for any $r \in (0, 1)$. Apply Proposition 7.2 to $M = G$ and $\| \cdot \| = \lambda_1(\cdot)$; then we have

$$\lambda_1(G(\mu)) \leq G((\lambda_1)_\ast \mu). \quad (7.8)$$

Let $(\wedge^k)_\ast \mu$ be the push-forward of $\mu$ by the antisymmetric tensor power map $A \in \mathbb{P}(\mathcal{H}) \mapsto \wedge^k A \in \mathbb{P}(\wedge^k \mathcal{H})$, see [5, 38]. For each $k = 1, \ldots, N$, use [19, Theorem 4.2] and apply (7.8) to $\wedge^k \mu$ to obtain

$$\prod_{i=1}^k \lambda_i(G(\mu)) = \lambda_1(\wedge^k G(\mu)) = \lambda_1(G((\wedge^k)_\ast \mu))$$

$$\leq G((\lambda_1)_\ast((\wedge^k)_\ast \mu)) = G\left( \left( \prod_{i=1}^k \lambda_i \right)_\ast \mu \right)$$

$$= \exp \int_\mathbb{P} \log \prod_{i=1}^k \lambda_i(A) \, d\mu(A) = \prod_{i=1}^k G((\lambda_i)_\ast \mu).$$

Moreover,

$$\prod_{i=1}^N \lambda_i(G(\mu)) = \det G(\mu) = \exp \int_\mathbb{R} \log \det A \, d\mu(A) = \prod_{i=1}^N G((\lambda_i)_\ast \mu).$$
We therefore have

\[ G(\mu) \prec_{\log} G(\lambda \ast \mu). \]  

(7.9)

By replacing \( \mu \) with \( \mu^r \) for \( 0 < r < 1 \) we have

\[ G(\mu^r)^{1/r} \prec_{\log} G(\lambda \ast (\mu^r)^{1/r}) = G(\lambda \ast \mu), \]

where the last equality immediately follows from (7.7).

\[ \square \]

**Corollary 7.4.** Assume that \( f \) is a non-negative and non-decreasing function \( f \) on \((0, \infty)\) such that \( f(e^x) \) is convex on \( \mathbb{R} \). For every \( \mu \in \mathcal{P}_\infty(\mathbb{P}) \) and any unitarily invariant norm \( \| \cdot \| \),

\[ \| f(G(\mu)) \| \leq \| f(LE(\mu)) \| \leq \| f(G(\lambda \ast \mu)) \|. \]

(7.10)

In particular,

\[ \| G(\mu) \| \leq \| LE(\mu) \| \leq \| G(\lambda \ast \mu) \| \leq \exp \int_{\mathbb{P}} \log \| A \| \, d\mu(A). \]

(7.11)

**Proof.** Proposition 7.3 implies (7.10) by [16, Proposition 4.1.6]. For the last inequality of (7.11), since \( \| \lambda(A) \| = \| A \| \) for all \( A \in \mathbb{P} \), we have by Proposition 7.2

\[ \| G(\lambda \ast \mu) \| \leq G(\| \cdot \| \ast (\lambda \ast \mu)) = G(\| \cdot \| \circ \lambda \ast \mu) \]

\[ = G(\| \cdot \| \ast \lambda \ast \mu) = \exp \int_{\mathbb{P}} \log \| A \| \, d\mu(A). \]

\[ \square \]

Note that (7.11) contains (7.8) for unitarily invariant norms \( \| \cdot \| \), while the latter holds for more general monotone norms.

We have the integral version of the Ky Fan majorization as

\[ A(\mu) \prec A(\lambda \ast \mu), \]

(7.12)

i.e.,

\[ \sum_{i=1}^{k} \lambda_i \left( \int_{\mathbb{P}} A \, d\mu(A) \right) \leq \sum_{i=1}^{k} \int_{\mathbb{P}} \lambda_i(A) \, d\mu(A), \quad 1 \leq k \leq N, \]

with equality for \( k = N \). A little argument by replacing \( \mu \) in (7.12) with \( \mu^{-1} \) gives the weak majorization

\[ \mathcal{H}(\mu) \prec_w \mathcal{H}(\lambda \ast \mu). \]

(7.13)

For any two-variable operator mean \( \sigma \) and every \( A, B \in \mathbb{P} \) we have

\[ \lambda(\lambda(\sigma)B) \prec_w \lambda(A)\sigma \lambda(B). \]

(7.14)

Since we find no literature on this, we give the proof in Appendix B.1 for completeness, together with the proofs of (7.12) and (7.13) in Appendix B.2. In view of (7.11)–(7.13) as well as (7.14), one may expect the weak majorization \( \lambda(M(\mu)) \prec_w M(\lambda \ast \mu) \) for other operator means \( M \) on \( \mathcal{P}_\infty(\mathbb{P}) \). In the next proposition we prove that this holds true for the power means \( P_r \) with \( 0 < r \leq 1 \).
Proposition 7.5. For every $r \in (0, 1]$ and every $\mu \in \mathcal{P}^\infty(\mathbb{P})$,

$$P_r(\mu) \prec_w P_r(\lambda_*\mu). \quad (7.15)$$

Proof. Since $P_1 = \mathcal{A}$, the case $\alpha = 1$ holds by (7.12). So we may assume that $0 < \alpha < 1$. Since a simple calculation gives

$$P_r(\lambda_*\mu) = \left( \left[ \int_\mathbb{P} \lambda_i(A)^\alpha d\mu(A) \right]^{1/\alpha} \right)^N,$$

what we need to prove is

$$\sum_{i=1}^{k} \lambda_i(P_\alpha(\mu)) \leq \sum_{i=1}^{k} \left[ \int_\mathbb{P} \lambda_i(A)^\alpha d\mu(A) \right]^{1/\alpha}, \quad 1 \leq k \leq N. \quad (7.16)$$

Set $X_0 := P_\alpha(\mu)$. Since $X_0 = \mathcal{A}(X_0 \#_*\mu)$, for $1 \leq k \leq N$ we have by (7.12)

$$\sum_{i=1}^{k} \lambda_i(X_0) \leq \sum_{i=1}^{k} \mathcal{A}(\lambda_i(X_0 \#_*\mu)) = \int_\mathbb{P} \sum_{i=1}^{k} \lambda_i(X_0 \#_* A) d\mu(A).$$

Since $\lambda(X_0) \#_* \lambda(A) = (\lambda_i(X_0)^{1-\alpha}\lambda_i(A)\lambda_i)^N$, it follows from (7.14) for $\sigma = \#_\alpha$ that

$$\sum_{i=1}^{k} \lambda_i(X_0 \#_* A) \leq \sum_{i=1}^{k} \lambda_i(X_0)^{1-\alpha}\lambda_i(A)^\alpha, \quad A \in \mathbb{P}.$$

Therefore,

$$\sum_{i=1}^{k} \lambda_i(X_0) \leq \sum_{i=1}^{k} \lambda_i(X_0)^{1-\alpha} \int_\mathbb{P} \lambda_i(A)^\alpha d\mu(A).$$

Thanks to Hölder’s inequality we find that

$$\sum_{i=1}^{k} \lambda_i(X_0) \leq \left[ \sum_{i=1}^{k} \lambda_i(X_0) \right]^{1-\alpha} \left( \sum_{i=1}^{k} \left[ \int_\mathbb{P} \lambda_i(A)^\alpha d\mu(A) \right]^{1/\alpha} \right)^{\alpha},$$

which implies (7.16). \qed

Corollary 7.6. Assume that $f$ is a non-negative and non-decreasing convex function on $(0, \infty)$. Let $\alpha \in (0, 1]$. For every $\mu \in \mathcal{P}^\infty(\mathbb{P})$ and any unitarily invariant norm $||| \cdot |||$, $|||f(P_\alpha(\mu))||| \leq |||f(P_\alpha(\lambda_*\mu))|||$. In particular,

$$|||P_\alpha(\mu)||| \leq |||P_\alpha(\lambda_*\mu)||| \leq P_\alpha(||| \cdot |||_*\mu) = \left[ \int_\mathbb{P} |||A|||\alpha d\mu(A) \right]^{1/\alpha}.$$

By letting $\alpha \searrow 0$ in (7.15) we have

$$G(\mu) \prec_w G(\lambda_*\mu),$$

which is however weaker than (7.9).
Problem 7.7. When $\alpha = -1$, (7.15) reduces to (7.13). But it is unknown whether the weak majorization in (7.15) holds even for $\alpha \in (-1, 0)$ or not. The problem is to prove (7.16) for $\alpha \in (-1, 0)$ and $\mu \in P^\infty(\mathbb{F})$. Note that a weaker version
\[
\sum_{i=1}^{k} \lambda_i(P_\alpha(\mu)) \leq \left[ \int_{A} \left( \sum_{i=1}^{k} \lambda_i(A) \right)^{\alpha} d\mu(A) \right]^{1/\alpha}
\]
is the special case of the norm inequality in (7.3) for the Ky Fan $k$-norm. Now, let us try to apply (7.15) to $-\alpha \in (0, 1)$ and $\mu^{-1}$; then we have
\[
P_{-\alpha}(\mu^{-1}) \prec_w P_{-\alpha}(\lambda^*(\mu^{-1})).
\]
Note that $P_{-\alpha}(\mu^{-1}) = P_{\alpha}(\mu)^{-1}$ and
\[
P_{-\alpha}(\lambda^*(\mu^{-1})) = P_{-\alpha}((\lambda^\dagger)_\alpha^{-1}) = P_{\alpha}(\lambda^\dagger)_\alpha^{-1}
\]
\[
= (P_{\alpha}(\lambda^\dagger)_\alpha)^{-1} = (P_{\alpha}(\lambda^\dagger)_\alpha^{-1})^\dagger,
\]
where $\lambda^\dagger(A) := \lambda(A)^\dagger$. Therefore,
\[
P_{\alpha}(\mu)^{-1} \prec_w P_{\alpha}(\lambda^\dagger)_\alpha^{-1},
\]
which is a complementary version of (7.15). Another majorization that is stronger than (7.17) is the supermajorization
\[
P_{\alpha}(\mu) \prec^w P_{\alpha}(\lambda^\dagger)_\alpha,
\]
i.e.,
\[
\sum_{i=1}^{k} \lambda_{N+1-i}(P_{\alpha}(\mu)) \geq \sum_{i=1}^{k} P_{\alpha}(\lambda_{N+1-i}^\dagger)_\alpha, \quad 1 \leq k \leq N,
\]
which however cannot hold. Indeed, if the above supermajorization holds for $P_{-1} = \mathcal{H}$, then together with (7.13) we have $\lambda(\mathcal{H}(\mu)) \prec \mathcal{H}(\lambda^\dagger)_\alpha$, which is impossible.

7.3. Minkowski determinant inequality. The famous Minkowski determinant inequality says that, for every $A, B \in B(\mathcal{H})^+$,
\[
det^{1/N}(A + B) \geq \det^{1/N} A + \det^{1/N} B,
\]
or equivalently, $A \in B(\mathcal{H})^+ \mapsto \det^{1/N} A$ is concave:
\[
det^{1/N}(A \triangledown_t B) \geq (\det^{1/N} A) \triangledown_t (\det^{1/N} B), \quad 0 \leq t \leq 1.
\]
The following extension was given in [10, Corollary 3.2].

Lemma 7.8. If a two-variable operator mean $\sigma$ (in the Kubo-Ando sense) is g.c.v. (see Section 5.1), then
\[
det^{1/N} (A \sigma B) \geq (\det^{1/N} A) \sigma (\det^{1/N} B), \quad A, B \in B(\mathcal{H})^+.
\]
The reverse inequality holds if $\sigma$ is g.c.c.

The following is the further extension to derived operator means $M \in \mathcal{M}^+$ or $\mathcal{M}^-$. 

Proposition 7.9. For every derived operator mean \( M \in \mathcal{M}^+ \),
\[
\det^{1/N} M(\mu) \geq M((\det^{1/N})_\# \mu), \quad \mu \in \mathcal{P}^\infty(\mathbb{P}),
\]
where \((\det^{1/N})_\# \mu\) is the push-forward of \( \mu \) by \( A \in \mathbb{P} \mapsto \det^{1/N} A \in (0, \infty) \). The reversed inequality holds if \( M \in \mathcal{M}^- \).

Proof. The proof is again similar to that of Theorem 6.5. Here we only show that if \( M \in \mathcal{M} \) satisfies (7.18) and \( \sigma \) is a g.c.v. two-variable operator mean with \( \sigma \neq I \), then the deformed \( M_\sigma \) does the same. For every \( \mu \in \mathcal{P}^\infty(\mathbb{P}) \) let \( X_0 := M_\sigma(\mu) \). Then
\[
\det^{1/N} X_0 = \det^{1/N} M(X_0 \sigma \mu) \geq M((\det^{1/N})_\# (X_0 \sigma \mu))
\geq M((\det^{1/N} X_0) \sigma((\det^{1/N})_\# \mu)),
\]
where the last inequality can be shown by use of Lemma 7.8 as in the proof of Proposition 7.2. By Theorem 3.1 (2) (for the case \( \dim \mathcal{H} = 1 \)) this implies that \( \det^{1/N} X_0 \geq M_\sigma((\det^{1/N})_\# \mu) \).

The latter assertion follows from the first since \( \det^{1/N} M(\mu^{-1}) = \det^{-1/N} M^*(\mu) \) and \( M \in \mathcal{M}^+ \iff M^* \in \mathcal{M}^- \). \( \square \)

In particular, since \( G \in \mathcal{M}^+ \cap \mathcal{M}^- \), we have
\[
\det^{1/N} G(\mu) = G((\det^{1/N})_\# \mu) = \exp \int_\mathbb{P} \frac{1}{N} \text{tr}(\log A) \, d\mu(A)
\]
for every \( \mu \in \mathcal{P}^\infty(\mathbb{P}) \), as verified in [33]. For power means Proposition 7.9 gives:

Corollary 7.10. For every \( \mu \in \mathcal{P}^\infty(\mathbb{P}) \),
\[
\det^{1/N} P_\alpha(\mu) \geq P_\alpha((\det^{1/N})_\# \mu) \quad \text{for } 0 < \alpha \leq 1,
\]
\[
\det^{1/N} P_\alpha(\mu) \leq P_\alpha((\det^{1/N})_\# \mu) \quad \text{for } -1 \leq \alpha < 0.
\]

In the case of the \( n \)-variable power means, the above inequality means that, for any weight vector \( w \) and for every \( A_1, \ldots, A_n \in \mathbb{P} \),
\[
\det^{1/N} P_{w,\alpha}(A_1, \ldots, A_n) \geq P_{w,\alpha}((\det^{1/N} A_1, \ldots, \det^{1/N} A_n) \quad \text{for } 0 < \alpha \leq 1,
\]
\[
\det^{1/N} P_{w,\alpha}(A_1, \ldots, A_n) \geq P_{w,\alpha}((\det^{1/N} A_1, \ldots, \det^{1/N} A_n) \quad \text{for } -1 \leq \alpha < 0.
\]

Problem 7.11. It seems interesting to extend (7.18) to the Fuglede-Kadison determinant in a finite von Neumann algebra. Let \( \mathcal{N} \) be a von Neumann algebra on \( \mathcal{H} \) with a faithful normal tracial state \( \tau \). For every invertible operator \( X \in \mathcal{N} \) the Fuglede-Kadison determinant of \( X \) is given as
\[
\Delta(X) := \exp \tau(\log |X|).
\]
Let \( M \in \mathcal{M} \) and \( \mu \in \mathcal{P}^\infty(\mathbb{P}) \). Assume that \( \mu \) is supported on \( \mathbb{P}(\mathcal{N}) \), the set of positive invertible operators in \( \mathcal{N} \). Then, for every unitary \( U \in \mathcal{N}' \), the commutant of \( \mathcal{N} \), we have \( UM(\mu)U^* = M(U\mu U^*) = M(\mu) \). This says that \( M(\mu) \in \mathcal{N} \) and so \( M(\mu) \in \mathbb{P}(\mathcal{N}) \). We may conjecture that the extension of Proposition 7.9 holds as
\[
\Delta(M(\mu)) \geq M(\Delta_\# \mu) \quad \text{for } M \in \mathcal{M}^+.
\]
\[ \Delta(M(\mu)) \leq M(\Delta_\mu) \quad \text{for } M \in \mathfrak{M}. \]

Acknowledgments

The work of F. Hiai was supported in part by JSPS KAKENHI Grant Number JP15H01056. The work of Y. Lim was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) No. 2015R1A5A2031159 and 2016R1A5A1008055.

Appendix A. SOT-continuity of certain operator means

The next proposition is concerned with the SOT-continuity of two-variable operator means.

**Proposition A.1.** Any two-variable operator mean \( \sigma \) (in the Kubo-Ando sense) is SOT-continuous on \( \{ A \in \mathbb{P} : A \geq \varepsilon I \} \times \mathbb{P} \) and on \( \mathbb{P} \times \{ A \in \mathbb{P} : A \geq \varepsilon I \} \) for every \( \varepsilon > 0 \).

**Proof.** Let \( A, A_k, B, B_k \in \mathbb{P} \) \( (k \in \mathbb{N}) \), where \( A, A_k \geq \varepsilon I \) for some \( \varepsilon > 0 \), and assume that \( A_k \to A \) and \( B_k \to B \) in SOT. Note that \( A_k \)'s and \( B_k \)'s are \( \| \cdot \| \)-bounded by the uniform boundedness theorem. Then \( A_k^{1/2} \to A^{1/2} \) and \( A_k^{-1/2} \to A^{-1/2} \) in SOT, so that \( A_k^{-1/2}B_kA_k^{-1/2} \to A^{-1/2}BA^{-1/2} \) in SOT. Therefore, \( f_\sigma(A_k^{-1/2}B_kA_k^{-1/2}) \to f_\sigma(A^{-1/2}BA^{-1/2}) \) in SOT, which implies that \( A_k\sigma B_k = A_k^{1/2}f_\sigma(A_k^{-1/2}B_kA_k^{-1/2})A_k^{1/2} \to A\sigma B \) in SOT. Let \( \sigma' \) be the transpose of \( \sigma \) [27]. We have \( B_k\sigma A_k = A_k\sigma' B_k \to A\sigma' B = B\sigma A \) in SOT. \( \square \)

**Remark A.2.** When \( \dim \mathcal{H} = \infty \), the operator mean \( \sigma \) is not necessarily SOT-continuous on the whole \( \mathbb{P} \times \mathbb{P} \). For this, recall the well-known fact that there are dense subspaces \( \mathcal{K} \) and \( \mathcal{L} \) of \( \mathcal{H} \) such that \( \mathcal{K} \cap \mathcal{L} = \{ 0 \} \). This follows from a classical result of von Neumann, who proved that, for any unbounded self-adjoint operator \( T \), there exists a unitary operator \( U \) such that the domains of \( U \) and \( U^*TU \) have the zero intersection. A readable exposition on this matter is found in [13]. One can then construct two orthonormal bases \( \{|e_j\}_{j=1}^\infty \) and \( \{|f_j\}_{j=1}^\infty \) of \( \mathcal{H} \) which are in \( \mathcal{K} \) and \( \mathcal{L} \), respectively. Let \( P_k, Q_k \) \( (k \in \mathbb{N}) \) be the orthogonal projections onto the spans of \( \{|e_j\}_{j=1}^k \) and of \( \{|f_j\}_{j=1}^k \), respectively. Now let \( \sigma \) be, for instance, the geometric mean or the harmonic mean, and choose a unit vector \( \xi \in \mathcal{H} \). For each \( k \), since \( P_k\sigma Q_k = P_k \wedge Q_k = 0 \) by [27], one can choose an \( \varepsilon_k > 0 \) such that \( \langle \xi, (P_k + \varepsilon_k I)\sigma(Q_k + \varepsilon_k I)\xi \rangle < 1/2 \). Here a sequence \( \varepsilon_k \) can be chosen as \( \varepsilon_k \searrow 0 \). Set \( A_k := P_k + \varepsilon_k I \) and \( B_k := Q_k + \varepsilon_k I \). Then \( A_k \to I \) and \( B_k \to I \) in SOT, but \( \langle \xi, (A_k\sigma B_k)\xi \rangle < 1/2 \) for all \( k \), so that \( A_k\sigma B_k \not\to I \) in SOT. Since \( P_k \not\to I \) and \( Q_k \not\to I \), it is also seen that \( \sigma \) is not upward continuous on \( B(\mathcal{H})^+ \times B(\mathcal{H})^+ \).

Next, to give a proof of the next proposition mentioned in Problem [55], we recall a few more basic facts on the Wasserstein distance \( d^W_p \) on a complete metric space \( (X, d) \), in addition to an account in Section 2. Let \( \mathcal{P}_0(X) \) be the set of finitely
supported uniform probability measures on $X$, i.e., probability measures of the form

$$
\mu = (1/n) \sum_{i=1}^{n} \delta_{x_i} \quad (n \in \mathbb{N}, x_i \in X).$

It is well-known (see [43]) that $\mathcal{P}_0(X)$ is $d_p$-dense in $\mathcal{P}^p(X)$ for any $p \in [1, \infty)$. When $\mu, \nu \in \mathcal{P}_0(X)$ where $\mu = (1/n) \sum_{i=1}^{n} \delta_{x_i}$ and $\nu = (1/n) \sum_{i=1}^{n} \delta_{y_i}$, we have (see [45], p. 5)

$$
d_p^{W}(\mu, \nu) = \min_{\sigma \in S_n} \left[ \frac{1}{n} \sum_{i=1}^{n} d_p^{F}(x_i, y_{\sigma(i)}) \right]^{1/p}, \quad 1 \leq p < \infty, \quad (A.1)
$$

where $S_n$ is the permutation group on $\{1, \ldots, n\}$.

**Proposition A.3.** If $\mu, \mu_k \in \mathcal{P}^{\infty}(\mathbb{P}) \ (k \in \mathbb{N})$ are supported on $\Sigma_{\epsilon}$ with $\epsilon \in (0, 1)$ and $\mu_k \rightarrow \mu$ weakly on $\Sigma_{\epsilon}$ with SOT, i.e., $\int_{\mathbb{P}} f(A) \, d\mu_k(A) \rightarrow \int_{\mathbb{P}} f(A) \, d\mu(A)$ for every bounded SOT-continuous real function $f$ on $\Sigma_{\epsilon}$, then $\mathcal{A}(\mu_k) \rightarrow \mathcal{A}(\mu)$ and $\mathcal{H}(\mu_k) \rightarrow \mathcal{H}(\mu)$ in SOT.

**Proof.** As mentioned before Definition 2.1, $\Sigma_{\epsilon}$ with SOT is a Polish space with the metric $d_{\epsilon}$ in (2.1). Let $\mathcal{P}(\Sigma_{\epsilon}) (\subset \mathcal{P}^{\infty}(\mathbb{P}))$ be the set of Borel probability measures supported on $\Sigma_{\epsilon}$, and consider the 1-Wasserstein distance $(d_{\epsilon})_{1}^{W}$ on $\mathcal{P}(\Sigma_{\epsilon})$ with respect to $d_{\epsilon}$, as in (2.2) with $d_{\epsilon}$ in place of $d_T$. Since $\sup \{d_{\epsilon}(A, B) : A, B \in \Sigma_{\epsilon} \} < \infty$, it follows from [45], Theorem 7.12 that $\mu_k \rightarrow \mu$ weakly on $(\Sigma_{\epsilon}, d_{\epsilon})$ if and only if $(d_{\epsilon})_{1}^{W}(\mu_k, \mu) \rightarrow 0$.

For $\mu = (1/N) \sum_{i=1}^{N} \delta_{A_i}$ and $\nu = (1/N) \sum_{i=1}^{N} \delta_{B_i}$ in $\mathcal{P}_0(\Sigma_{\epsilon})$, we have by (A.1)

$$
(d_{\epsilon})_{1}^{W}(\mu, \nu) = \min_{\sigma \in S_N} \frac{1}{N} \sum_{i=1}^{N} d_{\epsilon}(A_i, B_{\sigma(i)}). \quad (A.1)
$$

Since, for any $\sigma \in S_N$,

$$
d_{\epsilon}(\mathcal{A}(\mu), \mathcal{A}(\nu)) = \sum_{n=1}^{\infty} \frac{1}{2^n} \left\| \left( \frac{1}{N} \sum_{i=1}^{N} A_i - \frac{1}{N} \sum_{i=1}^{N} B_{\sigma(i)} \right) \right\| 
\leq \frac{1}{N} \sum_{n=1}^{\infty} \sum_{i=1}^{N} \frac{1}{2^n} \| (A_i - B_{\sigma(i)}) x_n \| \n = \frac{1}{N} \sum_{i=1}^{N} d_{\epsilon}(A_i, B_{\sigma(i)}),
$$

we have

$$
d_{\epsilon}(\mathcal{A}(\mu), \mathcal{A}(\nu)) \leq (d_{\epsilon})_{1}^{W}(\mu, \nu), \quad \mu, \nu \in \mathcal{P}_0(\Sigma_{\epsilon}). \quad (A.2)
$$

For any $\mu, \nu \in \mathcal{P}(\Sigma_{\epsilon})$ choose sequences $\mu_k, \nu_k \in \mathcal{P}_0(\Sigma_{\epsilon})$ such that $(d_{\epsilon})_{1}^{W}(\mu_k, \mu) \rightarrow 0$ and $(d_{\epsilon})_{1}^{W}(\nu_k, \nu) \rightarrow 0$. Since $d_{\epsilon}(\mathcal{A}(\mu_k), \mathcal{A}(\mu_l)) \leq (d_{\epsilon})_{1}^{W}(\mu_k, \mu_l) \rightarrow 0$ as $k, l \rightarrow \infty$, it follows that $\mathcal{A}(\mu_k)$ converges to $A_0 \in \Sigma_{\epsilon}$ in SOT. Since $(d_{\epsilon})_{1}^{W}(\mu_k, \mu) \rightarrow 0$ implies that $\int_{\Sigma_{\epsilon}} (x, Ay) \, d\mu_k(A) \rightarrow \int_{\Sigma_{\epsilon}} (x, Ay) \, d\mu(A)$, i.e., $\langle x, \mathcal{A}(\mu_k)y \rangle \rightarrow \langle x, \mathcal{A}(\mu)y \rangle$ for any $x, y \in \mathcal{H}$, so $\mathcal{A}(\mu_k) \rightarrow \mathcal{A}(\mu)$ in the weak operator topology. Hence we find that $A_0 = \mathcal{A}(\mu)$ so that $\mathcal{A}(\mu_k) \rightarrow \mathcal{A}(\mu)$ in SOT, and similarly $\mathcal{A}(\nu_k) \rightarrow \mathcal{A}(\nu)$ in SOT. Therefore, by applying (A.2) to $\mu_k, \nu_k$ and taking the limit, inequality (A.2) can extend to

$$
d_{\epsilon}(\mathcal{A}(\mu), \mathcal{A}(\nu)) \leq (d_{\epsilon})_{1}^{W}(\mu, \nu), \quad \mu, \nu \in \mathcal{P}(\Sigma_{\epsilon}),
$$

from which we have $\mathcal{A}(\mu_k) \rightarrow \mathcal{A}(\mu)$ in SOT if $\mu, \mu_k \in \mathcal{P}(\Sigma_{\epsilon})$ and $\mu_k \rightarrow \mu$ weakly on $(\Sigma_{\epsilon}, d_{\epsilon})$, i.e., $(d_{\epsilon})_{1}^{W}(\mu_k, \mu) \rightarrow 0$. 


The assertion for $\mathcal{H}$ immediately follows from that for $\mathcal{A}$ since $A \mapsto A^{-1}$ is a homeomorphic self-map on $\Sigma_{\infty}$.

\section*{Appendix B. Some proofs}

\subsection*{B.1. Proof of (7.14)} Let $A, B \in \mathbb{P}$. The well-known Ky Fan majorization (see [16, 38]) says that

$$\lambda(A + B) \prec_{w} \lambda(A) + \lambda(B).$$

Replacing $A, B$ with $A^{-1}, B^{-1}$ and applying the convex function $x^{-1}$ on $(0, \infty)$, we have

$$\lambda(A^{-1} + B^{-1})^{-1} \prec_{w} (\lambda(A^{-1}) + \lambda(B^{-1}))^{-1}.$$

Note that $\lambda(A^{-1} + B^{-1})^{-1} = \lambda^{\dagger}(A : B)$ and $\lambda(A^{-1}) = \lambda^{\dagger}(A)^{-1}$, where $A : B := (A^{-1} + B^{-1})^{-1}$, the parallel sum of $A, B$ and $\lambda^{\dagger}(A)$ is the eigenvalues of $A$ arranged in increasing order. Therefore,

$$\lambda^{\dagger}(A : B) \prec_{w} \lambda^{\dagger}(A) : \lambda^{\dagger}(B),$$

Since $(\lambda^{\dagger}(A) : \lambda^{\dagger}(B))^{\dagger} = \lambda(A) : \lambda(B)$, the above means that

$$\lambda(A : B) \prec_{w} \lambda(A) : \lambda(B). \quad (B.1)$$

Now consider the integral expression in (7.2). Note that

$$\frac{1 + t}{t} \| (tA) : B \| \leq \frac{1 + t}{t} \{ (t\|A\|) : \|B\| \} = \frac{(1 + t)\|A\| \|B\|}{t\|A\| + \|B\|} \leq \max\{\|A\|, \|B\|\}, \quad t \in (0, \infty).$$

Hence, by approximating the integral with Riemann sums, we have

$$\int_{(0, \infty)} \frac{1 + t}{t} \{ (tA) : B \} \, dm(t) = \lim_{n \to \infty} \sum_{j=1}^{n^2} m\left(\left(\frac{j - 1}{n}, \frac{j}{n}\right)\right) \frac{1 + \frac{j}{n}}{\frac{1}{n}} \left\{ \left(\frac{j}{n} A\right) : B \right\}$$

in the operator norm, and similarly

$$\int_{(0, \infty)} \frac{1 + t}{t} \{ (t\lambda_{i}(A)) : \lambda_{i}(B) \} \, dm(t) = \lim_{n \to \infty} \sum_{j=1}^{n^2} m\left(\left(\frac{j - 1}{n}, \frac{j}{n}\right)\right) \frac{1 + \frac{j}{n}}{\frac{1}{n}} \left\{ \left(\frac{j}{n} \lambda_{i}(A)\right) : \lambda_{i}(B) \right\}, \quad 1 \leq i \leq N.$$

Hence, from (7.2) and the Ky Fan majorization, we have for $1 \leq k \leq N$,

$$\sum_{i=1}^{k} \lambda_{i}(A\sigma B) \leq \lim_{n \to \infty} \sum_{i=1}^{k} \left[ a \lambda_{i}(A) + b \lambda_{i}(B) + \sum_{j=1}^{n^2} m\left(\left(\frac{j - 1}{n}, \frac{j}{n}\right)\right) \frac{1 + \frac{j}{n}}{\frac{1}{n}} \lambda_{i} \left(\left(\frac{j}{n} A\right) : B\right) \right]$$

\begin{align*}
& \leq \lim_{n \to \infty} \sum_{i=1}^{k} \left[ a \lambda_{i}(A) + b \lambda_{i}(B) + \sum_{j=1}^{n^2} m\left(\left(\frac{j - 1}{n}, \frac{j}{n}\right)\right) \frac{1 + \frac{j}{n}}{\frac{1}{n}} \left\{ \left(\frac{j}{n} \lambda_{i}(A)\right) : \lambda_{i}(B) \right\} \right]
\end{align*}
\[ = \sum_{i=1}^{k} \{\lambda_i(A)\sigma_i(B)\}, \]

where we have used (B.1) for the above latter inequality. Therefore, \( \lambda(A\sigma B) \prec_w \lambda(A)\sigma\lambda(B) \).

**B.2. Proofs of (7.12) and (7.13).** Let \( \mu \in \mathcal{P}^\infty(\mathbb{P}) \) and choose an \( \varepsilon > 0 \) such that \( \mu \) is supported on \( \Sigma_\varepsilon \). Since \( \Sigma_\varepsilon \) is compact in the operator norm thanks to \( \dim \mathcal{H} < \infty \), one can choose, for any \( n \in \mathbb{N} \), a finite set \( \{A_1^{(n)}, \ldots, A_m^{(n)}\} \) in \( \mathbb{P} \) such that \( \Sigma_\varepsilon \subset \bigcup_{j=1}^{m_n} \{A \in \mathbb{P} : \|A - A_j^{(n)}\| < 1/n\} \). So one can define disjoint Borel sets \( \mathcal{O}_j^{(n)} \) \((1 \leq j \leq m_n)\) such that \( \mathcal{O}_j^{(n)} \subset \{A : \|A - A_j^{(n)}\| < 1/n\} \) for \( 1 \leq j \leq m_n \) and \( \Sigma_\varepsilon = \bigcup_{j=1}^{m_n} \mathcal{O}_j^{(n)} \). Then it is obvious that

\[ \lim_{n \to \infty} \left\| A(\mu) - \sum_{j=1}^{m_n} \mu(\mathcal{O}_j^{(n)})A_j^{(n)} \right\| = 0, \quad \lim_{n \to \infty} \left\| \mathcal{H}(\mu) - \left[ \sum_{j=1}^{m_n} \mu(\mathcal{O}_j^{(n)})A_j^{(n)-1} \right]^{-1} \right\| = 0. \]

Moreover, for every \( i = 1, \ldots, N \),

\[ \int_{\mathbb{P}} \lambda_i(A) \, d\mu(A) = \lim_{n \to \infty} \sum_{j=1}^{m_n} \mu(\mathcal{O}_j^{(n)})\lambda_i(A_j^{(n)}), \]

\[ \left[ \int_{\mathbb{P}} \lambda_i(A)^{-1} \, d\mu(A) \right]^{-1} = \lim_{n \to \infty} \left[ \sum_{j=1}^{m_n} \mu(\mathcal{O}_j^{(n)})\lambda_i(A_j^{(n)})^{-1} \right]^{-1}. \]

Note that

\[ \mathcal{A}(\lambda_{*}\mu) = \left( \int_{\mathbb{P}} \lambda_i(A) \, d\mu(A) \right)^{N}_{i=1}, \quad \mathcal{H}(\lambda_{*}\mu) = \left( \left[ \int_{\mathbb{P}} \lambda_i(A)^{-1} \, d\mu(A) \right]^{-1} \right)^{N}_{i=1}. \]

By the Ky Fan majorization we have

\[ \lambda \left( \sum_{j=1}^{m_n} \mu(\mathcal{O}_j^{(n)})A_j^{(n)} \right) \prec \sum_{j=1}^{m_n} \mu(\mathcal{O}_j^{(n)})\lambda(A_j^{(n)}), \]

\[ \lambda \left( \left[ \sum_{j=1}^{m_n} \mu(\mathcal{O}_j^{(n)})A_j^{(n)} \right]^{-1} \right) \prec \left[ \sum_{j=1}^{m_n} \mu(\mathcal{O}_j^{(n)})\lambda(A_j^{(n)})^{-1} \right]^{-1}. \]

where the last weak majorization is verified similarly to (B.1). Letting \( n \to \infty \) gives \( \mathcal{A}(\mu) \prec \mathcal{A}(\lambda_{*}\mu) \) and \( \mathcal{H}(\mu) \prec \mathcal{H}(\lambda_{*}\mu) \). \( \square \)

**References**

[1] T. Ando, Concavity of certain maps on positive definite matrices and applications to Hadamard products, *Linear Algebra Appl.* **26** (1979) 203–241.

[2] T. Ando, Majorizations and Inequalities in Matrix Theory, *Linear Algebra Appl.* **199** (1994), 17–67.

[3] T. Ando and F. Hiai, Log majorization and complementary Golden-Thompson type inequality, *Linear Algebra Appl.* **197** (1994), 113–131.

[4] T. Ando, C.-K. Li and R. Mathias, Geometric means, *Linear Algebra Appl.* **385** (2004), 305–334.

[5] R. Bhatia, *Matrix Analysis*, Springer-Verlag, New York, 1996.
[6] R. Bhatia and J. A. R. Holbrook, Unitary invariance and spectral variation, *Linear Algebra Appl.* **95** (1987), 43–68.
[7] R. Bhatia and J. Holbrook, Riemannian geometry and matrix geometric means, *Linear Algebra Appl.* **413** (2006), 594–618.
[8] R. Bhatia and R. L. Karandikar, Monotonicity of the matrix geometric mean, *Math. Ann.* **353** (2012), 1453–1467.
[9] S. Bobkov and M. Ledoux, One-dimensional empirical measures, order statistics, and Kantorovich transport distances, *Mem. Amer. Math. Soc.*, to appear.
[10] J.-C. Bourin and F. Hiai, Jensen and Minkowski inequalities for operator means and anti-norms, *Linear Algebra Appl.* **456** (2014), 22–53.
[11] J. B. Conway, *A Course in Functional Analysis*, Springer-Verlag, New York, 1985.
[12] A. Es-Sahib and H. Heinich, Barycentre canonique pour un espace métrique à courbure négative, *Séminaire de Probabilités*, Lecture Notes in Math. **1709**, Springer-Verlag, Berlin, 1999, pp. 355–370.
[13] P. Fillmore and J. Williams, On operator ranges, *Adv. Math.* **7** (1971), 254–281.
[14] F. Hansen, An operator inequality, *Math Ann.* **246** (1980), 249–250.
[15] F. Hansen and G. K. Pedersen, Jensen's inequality for operators and Löwner's theorem, *Math. Ann.* **258** (1982), 229–241.
[16] F. Hiai, Matrix Analysis: Matrix Monotone Functions, Matrix Means, and Majorization, *Interdisciplinary Information Sciences* **16** (2010), 139–248.
[17] F. Hiai, Operator means deformed by a fixed point method, Preprint, 2017, [arXiv:1711.10170](https://arxiv.org/abs/1711.10170) [math.FA].
[18] F. Hiai, J. Lawson and Y. Lim, The stochastic order of probability measures on ordered metric spaces, *J. Math. Anal. Appl.* **464** (2018), 707–724.
[19] F. Hiai and Y. Lim, Log-majorization and Lie-Trotter formula for the Cartan barycenter on probability measure spaces, *J. Math. Anal. Appl.* **453** (2017), 195–211.
[20] F. Hiai and Y. Lim, Geometric mean flows and the Cartan barycenter on the Wasserstein space over positive definite matrices, *Linear Algebra Appl.* **533** (2017) 118–131.
[21] F. Hiai and Y. Lim, Convergence theorems for barycentric maps, Preprint, 2018, [arXiv:1805.08558](https://arxiv.org/abs/1805.08558) [math.PR].
[22] F. Hiai, Y. Seo and S. Wada, Ando-Hiai inequalities for multivariate operator means, *Linear Multilinear Algebra*, DOI: 10.1080/03081087.2018.1488938, to appear.
[23] S. Kim, J. Lawson and Y. Lim, Barycentric maps for compactly supported measures, *J. Math. Anal. Appl.* **458** (2018), 1009–1026.
[24] S. Kim and H. Lee, The power mean and the least squares mean of probability measures on the space of positive definite matrices, *Linear Algebra Appl.* **465** (2015), 325–346.
[25] S. Kim, H. Lee and Y. Lim, An order inequality characterizing invariant barycenters on symmetric cones, *J. Math. Anal. Appl.* **442** (2016), 1–16.
[26] S. Kim, H. Lee and Y. Lim, A fixed point mean approximation to the Cartan barycenter of positive definite matrices, *Linear Algebra Appl.* **496** (2016), 420–437.
[27] F. Kubo and T. Ando, Means of positive linear operators, *Math. Ann.* **246** (1980), 205–224.
[28] J. Lawson, Ordered probability spaces, *J. Math. Anal. Appl.* **455** (2017), 167–179.
[29] J. Lawson and Y. Lim, Monotonic properties of the least squares mean, *Math. Ann.* **351** (2011) 267–279.
[30] J. Lawson and Y. Lim, Weighted means and Karcher equations of positive operators, *Proc. Natl. Acad. Sci. USA.* **110** (2013), 15626–15632.
[31] J. Lawson and Y. Lim, Karcher means and Karcher equations of positive definite operators, *Trans. Amer. Math. Soc. Series B* **1** (2014), 1–22.
[32] J. Lawson and Y. Lim, Contractive barycentric maps, *J. Operator Theory* **77** (2017), 87–107.
[33] Y. Lim, Matrix analysis of probability measures, *Proc. Natl. Acad. Sci. USA.* **110** (2013), 15626–15632.
[36] Y. Lim and M. Pálfia, Existence and uniqueness of the $L^1$-Karcher mean, Preprint, 2018; \texttt{arXiv:1703.04292v2} [math.FA].

[37] Y. Lim and T. Yamazaki, On some inequalities for the matrix power and Karcher means, \textit{Linear Algebra Appl.} \textbf{438} (2013), 1293–1304.

[38] A.W. Marshall, I. Olkin, B.C. Arnold, \textit{Inequalities: Theory of Majorization and Its Applications}, Second ed., Springer-Verlag, New York, 2011.

[39] M. Moakher, A differential geometric approach to the geometric mean of symmetric positive-definite matrices, \textit{SIAM J. Matrix Anal. Appl.} \textbf{26} (2005), 735–747.

[40] M. Pálfia, Operator means of probability measures and generalized Karcher equations, \textit{Adv. Math.} \textbf{289} (2016), 951–1007.

[41] W. Pusz and S. L. Woronowicz, Functional calculus for sesquilinear forms and the purification map, \textit{Rep. Math. Phys.} \textbf{8} (1975), 159–170.

[42] K.-T. Sturm, Nonlinear martingale theory for processes with values in metric spaces of nonpositive curvature, \textit{Ann. Probab.} \textbf{30} (2002), 1195–1222.

[43] K.-T. Sturm, Probability measures on metric spaces of nonpositive curvature. Heat kernels and analysis on manifolds, graphs, and metric spaces (Paris, 2002), 357-390, \textit{Contemp. Math.} \textbf{338}, Amer. Math. Soc., Providence, RI, 2003.

[44] A. C. Thompson, On certain contraction mappings in a partially ordered vector space, \textit{Proc. Amer. Math. Soc.} \textbf{14} (1963), 438–443.

[45] C. Villani, \textit{Topics in Optimal Transportation}, Graduate Studies in Mathematics, Vol. 58, American Mathematical Society, Providence, RI, 2003.

[46] S. Wada, Some ways of constructing Furuta-type inequalities, \textit{Linear Algebra Appl.} \textbf{457} (2014), 276–286.

[47] S. Wada, Geometric convexity of an operator mean, Preprint, 2019.

[48] T. Yamazaki, The Riemannian mean and matrix inequalities related to the Ando-Hiai inequality and chaotic order, \textit{Oper. Matrices} \textbf{6} (2012), 577–588.

[49] T. Yamazaki, Generalized Karcher equation, relative operator entropy and the Ando-Hiai inequality, Preprint 2018, \texttt{arXiv:1802.06200} [math.FA].

Graduate School of Information Sciences, Tohoku University, Aoba-ku, Sendai 980-8579, Japan

\textit{E-mail address:} hiai.fumio@gmail.com

Department of Mathematics, Sungkyunkwan University, Suwon 440-746, Korea

\textit{E-mail address:} ylim@skku.edu