The Resolvent Average for Positive Semidefinite Matrices

Heinz H. Bauschke[*], Sarah M. Moffat[†] and Xianfu Wang[‡]

November 17, 2009

Abstract

We define a new average — termed the *resolvent average* — for positive semidefinite matrices. For positive definite matrices, the resolvent average enjoys self-duality and it interpolates between the harmonic and the arithmetic averages, which it approaches when taking appropriate limits. We compare the resolvent average to the geometric mean. Some applications to matrix functions are also given.

2000 Mathematics Subject Classification:
Primary 47A64; Secondary 15A45, 15A24, 47H05, 90C25.

Keywords:
Arithmetic average, arithmetic mean, convex function, Fenchel conjugate, geometric mean, harmonic average, harmonic mean, positive semi-definite matrix, proximal average, resolvent average, subdifferential.

1 Introduction

Let $A_i, i = 1, \ldots, n$ be $N \times N$ positive semidefinite matrices, $\lambda_i > 0$ with $\sum_{i=1}^n \lambda_i = 1$ and $\mathrm{Id} : \mathbb{R}^N \to \mathbb{R}^N$ be the identity mapping. For

$$\boldsymbol{A} = (A_1, \ldots, A_n), \quad \boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n),$$

we define

$$\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) = \left[\lambda_1(A_1 + \mu^{-1}\,\mathrm{Id})^{-1} + \cdots + \lambda_n(A_n + \mu^{-1}\,\mathrm{Id})^{-1}\right]^{-1} - \mu^{-1}\,\mathrm{Id}, \tag{1}$$

[*]Mathematics, Irving K. Barber School, University of British Columbia Okanagan, Kelowna, British Columbia V1V 1V7, Canada. E-mail: heinz.bauschke@ubc.ca.
[†]Mathematics, Irving K. Barber School, University of British Columbia Okanagan, Kelowna, British Columbia V1V 1V7, Canada. E-mail: smoffat99@gmail.com.
[‡]Mathematics, Irving K. Barber School, University of British Columbia Okanagan, Kelowna, British Columbia V1V 1V7, Canada. E-mail: shawn.wang@ubc.ca.

and call it the *resolvent average* of $\boldsymbol{A}$. This is motivated from the fact that when $\mu = 1$

$$\big(\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda}) + \mathrm{Id}\big)^{-1} = \lambda_1\big(A_1 + \mathrm{Id}\big)^{-1} + \cdots + \lambda_n\big(A_n + \mathrm{Id}\big)^{-1}, \tag{2}$$

which says that the resolvent of $\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})$ is the (arithmetic) average of resolvents of the A_i, with weight $\boldsymbol{\lambda} = (\lambda_1,\ldots,\lambda_n)$. The resolvent average provides a novel averaging technique, and having the parameter μ in $\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})$ will allow us to take limits later on. We denote the well known *harmonic average* and *arithmetic average* by

$$\mathcal{H}(\boldsymbol{A},\boldsymbol{\lambda}) = (\lambda_1 A_1^{-1} + \cdots + \lambda_n A_n^{-1})^{-1},$$

$$\mathcal{A}(\boldsymbol{A},\boldsymbol{\lambda}) = \lambda_1 A_1 + \cdots + \lambda_n A_n,$$

respectively. In the literature, $(A_1^{-1} + \cdots + A_n^{-1})^{-1}$ is called the *parallel sum* of the matrices $A_1,\ldots,A_n$; see, e.g., [1, 13, 17, 21].

The goal of this note is to study relationships among the resolvent average, the harmonic average and the arithmetic average of matrices. Our proofs are based on convex analytical techniques and on the proximal average, instead of the more commonly employed matrix diagonalizations.

The plan of the paper is as follows. After proving some elementary properties of $\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})$ in Section 2, we gather some basic properties of proximal averages and general convex functions in Section 3. The main results, which are given in Section 4, state that

$$\mathcal{H}(\boldsymbol{A},\boldsymbol{\lambda}) \preceq \mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda}) \preceq \mathcal{A}(\boldsymbol{A},\boldsymbol{\lambda}),$$

$$\lim_{\mu\to 0^+} \mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda}) = \mathcal{A}(\boldsymbol{A},\boldsymbol{\lambda}), \quad \lim_{\mu\to+\infty} \mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda}) = \mathcal{H}(\boldsymbol{A},\boldsymbol{\lambda}),$$

and that $\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})$ enjoys self-duality, namely $\big[\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})\big]^{-1} = \mathcal{R}_{\mu^{-1}}(\boldsymbol{A}^{-1},\boldsymbol{\lambda})$. In Section 5, we show that the resolvent average and geometric mean have strikingly similar properties, even though they are different.

Notation: Throughout, $\mathbb{R}^N$ is the standard N-dimensional Euclidean space. For $\lambda > 0$,

$$J_A = (\mathrm{Id} + A)^{-1}, \quad {}^{\lambda}A = \lambda^{-1}(\mathrm{Id} - J_{\lambda A}), \tag{3}$$

are called the *resolvent* of A and *Yosida λ-regularization* of A. A function $f : \mathbb{R}^N \to\ ]-\infty,+\infty] = \mathbb{R}\cup\{+\infty\}$ is said to be convex if its domain is convex and

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y) \quad \forall\ x,y \in \mathbb{R}^N, 0 < \lambda < 1, \tag{4}$$

with f being strictly convex if (4) becomes a strict inequality whenever $x \neq y$. The function f is proper if $f(x) > -\infty\ \forall x \in \mathbb{R}^N$ and $f(x_0) < +\infty$ for some $x_0 \in \mathbb{R}^N$. The class of proper lower semicontinuous convex functions from $\mathbb{R}^N \to\ ]-\infty,+\infty]$ will be denoted by Γ. For $f \in \Gamma$, ∂f denotes its convex subdifferential: $\partial f(x) = \{x^* \in \mathbb{R}^N : f(y) \ge f(x) + \langle x^*, y - x\rangle\ \forall y \in \mathbb{R}^N\}$. If f is differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$. f^* denotes its *Fenchel conjugate* given by $(\forall x^* \in \mathbb{R}^N)\ f^*(x^*) = \sup_x\{\langle x^*, x\rangle - f(x)\}$. For $\alpha > 0$, $\alpha \star f = \alpha f(\cdot/\alpha)$. If $f,g \in \Gamma$, $f\Box g$ stands for the infimal convolution of f, g given by $(f\Box g)(x) = \inf\{f(x_1)+g(x_2):\ x_1+x_2 = x\}\ \forall x \in \mathbb{R}^N$. When $A : \mathbb{R}^N \to \mathbb{R}^N$ is linear, the quadratic form $q_A : \mathbb{R}^N \to \mathbb{R}$ is defined by

$$q_A(x) = \tfrac{1}{2}\langle Ax, x\rangle\ \forall x \in \mathbb{R}^N,$$

and we also use $q_{\mathrm{Id}} = j$ interchangeably. For convex functions $f_1, \ldots, f_n$, we write

$$\boldsymbol{f} = (f_1, \ldots, f_n), \qquad \boldsymbol{f}^* = (f_1^*, \ldots, f_n^*).$$

In the space $\mathbb{S}^N$ of $N \times N$ real symmetric matrices, $\mathbb{S}^N_+$ (resp. $\mathbb{S}^N_{++}$) denotes the set of $N \times N$ positive semidefinite matrices (resp. positive definite matrices). For $X, Y \in \mathbb{S}^N$, we write $Y \preceq X$ if $X - Y \in \mathbb{S}^N_+$ and $Y \prec X$ if $X - Y \in \mathbb{S}^N_{++}$.

2 Basic properties

In this section, we give some basic properties of $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})$.

Proposition 2.1 *We have*

$$J_{\mu\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})} = \lambda_1 J_{\mu A_1} + \cdots + \lambda_n J_{\mu A_n}, \tag{5}$$

$${}^{\mu}\big(\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})\big) = \lambda_1\, {}^{\mu}A_1 + \cdots + \lambda_n\, {}^{\mu}A_n. \tag{6}$$

Proof. Multiplying (1) both sides by μ gives

$$\mu\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) + \mathrm{Id} = \big[\lambda_1(\mu A_1 + \mathrm{Id})^{-1} + \cdots + \lambda_n(\mu A_n + \mathrm{Id})^{-1}\big]^{-1}. \tag{7}$$

Then (5) follows by taking inverse both sides and using (3).

By (5), we obtain that

$$(\mathrm{Id} - J_{\mu\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})}) = \lambda_1(\mathrm{Id} - J_{\mu A_1}) + \cdots + \lambda_n(\mathrm{Id} - J_{\mu A_n}).$$

Dividing both sides by μ,

$$\mu^{-1}(\mathrm{Id} - J_{\mu\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})}) = \lambda_1\mu^{-1}(\mathrm{Id} - J_{\mu A_1}) + \cdots + \lambda_n\mu^{-1}(\mathrm{Id} - J_{\mu A_n}).$$

It remains to use (3). ■

Proposition 2.2 *Let* $\boldsymbol{A} = (A_1, A_1^{-1}, \ldots, A_m, A_m^{-1})$, $\lambda = (\frac{1}{2m}, \frac{1}{2m}, \ldots, \frac{1}{2m})$, *and* $\mu = 1$. *Then* $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) = \mathrm{Id}$.

Proof. This follows from (2) and the identity $(A + \mathrm{Id})^{-1} + (A^{-1} + \mathrm{Id})^{-1} = \mathrm{Id}$. ■

Proposition 2.3 *Let* $\boldsymbol{A} = (A_1, \ldots, A_1)$. *Then* $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) = A_1$.

Proof. We have

$$\begin{aligned}\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda}) &= \big((\lambda_1+\cdots+\lambda_n)(A_1+\mu^{-1}\operatorname{Id})^{-1}\big)^{-1}-\mu^{-1}\operatorname{Id}\\ &= \big((A_1+\mu^{-1}\operatorname{Id})^{-1}\big)^{-1}-\mu\operatorname{Id} = A_1+\mu^{-1}\operatorname{Id}-\mu^{-1}\operatorname{Id}=A_1,\end{aligned}$$

which proves the result. ■

Note that for $A, B \in \mathbb{S}^N_{++}$, we have

$$A \succeq B \quad \Leftrightarrow \quad A^{-1} \preceq B^{-1} \tag{8}$$

and

$$A \succ B \quad \Leftrightarrow \quad A^{-1} \prec B^{-1}; \tag{9}$$

see, e.g., [14, Corollary 7.7.4.(a)] and [16, Section 16.E] or [11, page 55].

Proposition 2.4 *Assume that* $(\forall\ i)$ $A_i, B_i \in \mathbb{S}^N_+$ *and* $A_i \succeq B_i$. *Then*

$$\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda}) \succeq \mathcal{R}_\mu(\boldsymbol{B},\boldsymbol{\lambda}). \tag{10}$$

Furthermore, if additionally some $A_j \succ B_j$, *then* $\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda}) \succ \mathcal{R}_\mu(\boldsymbol{B},\boldsymbol{\lambda})$.

Proof. Note that $\forall\ \mu > 0$,

$$A_i + \mu^{-1}\operatorname{Id} \succeq B_i + \mu^{-1}\operatorname{Id} \succ 0,$$

so that

$$0 \prec (A_i + \mu^{-1}\operatorname{Id})^{-1} \preceq (B_i + \mu^{-1}\operatorname{Id})^{-1},$$

by (8). As $\mathbb{S}^N_+$ and $\mathbb{S}^N_{++}$ are convex cones, we obtain that

$$0 \prec \sum_{i=1}^n \lambda_i (A_i + \mu^{-1}\operatorname{Id})^{-1} \preceq \sum_{i=1}^n \lambda_i (B_i + \mu^{-1}\operatorname{Id})^{-1}. \tag{11}$$

Using (8) on (11), followed by subtracting $\mu^{-1}\operatorname{Id}$, gives

$$\Big[\sum_{i=1}^n \lambda_i (A_i + \mu^{-1}\operatorname{Id})^{-1}\Big]^{-1} - \mu^{-1}\operatorname{Id} \succeq \Big[\sum_{i=1}^n \lambda_i (B_i + \mu^{-1}\operatorname{Id})^{-1}\Big]^{-1} - \mu^{-1}\operatorname{Id},$$

which establishes (10). The "Furthermore" part follows analogously using (9). ■

Theorem 2.5 *Assume that* $(\forall\ i)$ $A_i \in \mathbb{S}^N_+$. *Then* $\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda}) \in \mathbb{S}^N_+$. *Furthermore, if additionally some* $A_j \in \mathbb{S}^N_{++}$, *then* $\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda}) \in \mathbb{S}^N_{++}$.

Proof. This follows from Proposition 2.4 (with each $B_i = 0$) and Proposition 2.3. ■

The following recursion formula may be verified directly using the definitions.

Proposition 2.6 (recursion) *We have*

$$\mathcal{R}_\mu(A_1,\ldots,A_n;\lambda_1,\ldots,\lambda_n) = \mathcal{R}_\mu\Big(\mathcal{R}_\mu\big(A_1,\ldots,A_{n-1};\tfrac{\lambda_1}{1-\lambda_n},\ldots,\tfrac{\lambda_{n-1}}{1-\lambda_n}\big), A_n; 1-\lambda_n, \lambda_n\Big).$$

The next interesting result is due to an anonymous referee, who also observed that all results in this paper have counterparts for Hermitian matrices.

Proposition 2.7 *For $\boldsymbol{A} = (A_1,\ldots,A_n)$ and $\boldsymbol{\lambda} = (\lambda_1,\ldots,\lambda_n)$, we have*

$$\mathcal{R}_\mu(\boldsymbol{A}_\sigma, \boldsymbol{\lambda}_\sigma) = \mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}), \tag{12}$$

where σ is any permutation of $\{1,\ldots,n\}$ and where $\boldsymbol{A}_\sigma = (A_{\sigma(1)},\ldots,A_{\sigma(n)})$ and $\boldsymbol{\lambda}_\sigma = (\lambda_{\sigma(1)},\ldots,\lambda_{\sigma(n)})$. Furthermore, if U is any orthogonal $N\times N$ matrix, then

$$U\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})U^T = \mathcal{R}_\mu(U\boldsymbol{A}U^T, \boldsymbol{\lambda}), \tag{13}$$

where $U\boldsymbol{A}U^T = (UA_1U^T,\ldots,UA_nU^T)$.

We conclude this section with a variational characterization of the resolvent average. For two symmetric matrices X and Y of the same size, we write $\langle X, Y\rangle$ for the trace of XY.

Proposition 2.8 (variational characterization) *Fix $\boldsymbol{\lambda} = (\lambda_1,\ldots,\lambda_n) \in \mathbb{R}^N_{++}$ such that $\sum_i \lambda_i = 1$ and let $\mu > 0$. Define*

$$f\colon \mathbb{S}^N_+ \to \mathbb{R}\colon A \mapsto -\ln\det\big(A + \mu^{-1}\operatorname{Id}\big)$$

and the corresponding ("Bregman distance" [3]) function

$$D\colon \mathbb{S}^N_+ \times \mathbb{S}^N_+ \to [0,+\infty[\,\colon (X,A) \mapsto f(X) - f(A) - \langle \nabla f(A), X - A\rangle.$$

Then for every $\boldsymbol{A} = \big(A_1,\ldots,A_n\big) \in (\mathbb{S}^N_+)^n$, the unique minimizer of the function

$$F\colon \mathbb{S}^N_+ \to [0,+\infty[\,\colon X \mapsto \lambda_1 D(X,A_1) + \cdots + \lambda_n D(X,A_n) \tag{14}$$

is precisely the resolvent average $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})$.

Proof. Observe that F is convex so that the minimizers of F are precisely the critical points. Using the well known fact that $\nabla f(X) = -(X + \mu^{-1}\operatorname{Id})^{-1}$, we see that the critical point equation $\nabla F(X) = 0$ turns into

$$\big(X + \mu^{-1}\operatorname{Id}\big)^{-1} = \sum_{i=1}^n \lambda_i\big(X + \mu^{-1}\operatorname{Id}\big)^{-1} = \sum_{i=1}^n \lambda_i\big(A_i + \mu^{-1}\operatorname{Id}\big)^{-1}.$$

Therefore, the unique minimizer of F is $X = \mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})$. ■

3 Auxiliary results and facts

The key tool in this note is the *proximal average of convex functions*, which finds its roots in [4, 18, 20], and which has been further systematically studied in [6, 7, 8, 9].

Definition 3.1 (proximal average) *Let* $(\forall i)$ $f_i \in \Gamma$. *The* λ*-weighted proximal average of* $\boldsymbol{f} = (f_1, \ldots, f_n)$ *with parameter* μ *is defined by*

$$p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \left(\lambda_1(f_1 + \tfrac{1}{\mu}j)^* + \lambda_2(f_2 + \tfrac{1}{\mu}j)^* + \cdots + \lambda_n(f_n + \tfrac{1}{\mu}j)^*\right)^* - \tfrac{1}{\mu}j. \tag{15}$$

The function $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ is a proper lower semicontinuous convex function on $\mathbb{R}^N$, and it inherits many desirable properties from each underlying function f_i; see [7, 8]. A fundamental property of proximal average is:

Fact 3.2 ([7, Theorem 5.1]) $\big(p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})\big)^* = p_{\mu^{-1}}(\boldsymbol{f}^*, \boldsymbol{\lambda})$.

To give new proofs of Fact 3.4 and Fact 3.5 below, we shall need reformulations of $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$.

Proposition 3.3 *Let* $f_1, \ldots, f_n \in \Gamma$ *and* $\lambda_1, \ldots, \lambda_n > 0$ *with* $\sum_{i=1}^n \lambda_i = 1$. *Then for every* $x \in \mathbb{R}^n$, $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x)$

$$= \inf_{x_1+\cdots+x_n=x} \left\{\lambda_1(f_1 + \frac{1}{\mu}j)(\frac{x_1}{\lambda_1}) + \cdots + \lambda_n(f_n + \frac{1}{\mu}j)(\frac{x_n}{\lambda_n})\right\} - \frac{1}{\mu}j(x) \tag{16}$$

$$= \inf_{x_1+\cdots+x_n=x} \left\{\lambda_1 f_1(\frac{x_1}{\lambda_1}) + \cdots + \lambda_n f_n(\frac{x_n}{\lambda_n}) + \frac{1}{4\mu}\sum_{i=1}^n\sum_{j=1}^n \lambda_i\lambda_j\|\frac{x_i}{\lambda_i} - \frac{x_j}{\lambda_j}\|^2\right\} \tag{17}$$

$$= \inf_{\lambda_1 y_1+\cdots+\lambda_n y_n=x} \left\{\lambda_1 f_1(y_1) + \cdots + \lambda_n f_n(y_n) + \frac{1}{\mu}[\lambda_1 j(y_1) + \cdots + \lambda_n j(y_n) - j(\lambda_1 y_1 + \cdots + \lambda_n y_n)]\right\} \tag{18}$$

$$= \inf_{\lambda_1 y_1+\cdots+\lambda_n y_n=x} \left\{\lambda_1 f_1(y_1) + \cdots + \lambda_n f_n(y_n) + \frac{1}{4\mu}\sum_{i=1}^n\sum_{j=1}^n \lambda_i\lambda_j\|y_i - y_j\|^2\right\} \tag{19}$$

$$= \inf_{x_1+\cdots+x_n=x} \left\{\lambda_1 f_1(\frac{x_1}{\lambda_1}) + \cdots + \lambda_n f_n(\frac{x_n}{\lambda_n}) + \frac{1}{\mu}[\lambda_1 j(x - \frac{x_1}{\lambda_1})) + \cdots + \lambda_n j(x - \frac{x_n}{\lambda_n})]\right\}. \tag{20}$$

Furthermore, the infimal convolutions in (16)–(20) are exact.

Proof. Indeed, as

$$\left(f_i + \frac{1}{\mu}j\right)^* = f_i^* \Box (\mu j),$$

it is finite-valued everywhere, we write

$$f = \lambda_1 \star (f_1 + \frac{1}{\mu}j)\Box \cdots \Box \lambda_n \star (f_n + \frac{1}{\mu}j) - \frac{1}{\mu}j,$$

by [23, Theorem 16.4]. That is, for every x,

$$f(x) = \inf \left\{ \lambda_1(f_1 + \frac{1}{\mu}j)(\frac{x_1}{\lambda_1}) + \cdots + \lambda_n(f_n + \frac{1}{\mu}j)(\frac{x_n}{\lambda_n}) : \; x_1 + \cdots + x_n = x \right\} - \frac{1}{\mu}j(x),$$

and the infimum is attained. Hence (16) holds.

Now rewrite (16) as

$$\begin{aligned} &\inf_{x_1+\cdots+x_n=x} \left\{ \lambda_1 f_1(\frac{x_1}{\lambda_1}) + \cdots + \lambda_n f_n(\frac{x_n}{\lambda_n}) + \frac{1}{\mu}[\lambda_1 j(\frac{x_1}{\lambda_1}) + \cdots + \lambda_n j(\frac{x_n}{\lambda_n}) - j(x_1 + \cdots + x_n)] \right\}, \\ &= \inf_{\lambda_1 y_1+\cdots+\lambda_n y_n=x} \left\{ \lambda_1 f_1(y_1) + \cdots + \lambda_n f_n(y_n) + \frac{1}{\mu}[\lambda_1 j(y_1) + \cdots + \lambda_n j(y_n) - j(\lambda_1 y_1 + \cdots + \lambda_n y_n)] \right\}. \end{aligned} \tag{21}$$

Thus, (17)–(19) follow by using the identity

$$\sum_{i=1}^{n} \lambda_i j(y_i) - j(\sum_{i=1}^{n} \lambda_i y_i) = \frac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \|y_i - y_j\|^2.$$

Observe that

$$\begin{aligned} &\lambda_1 j(x_1 + \cdots + x_n - \frac{x_1}{\lambda_1}) + \cdots + \lambda_n j(x_1 + \cdots + x_n - \frac{x_n}{\lambda_n}) \\ &= \lambda_1 j(\frac{x_1}{\lambda_1}) + \cdots + \lambda_n j(\frac{x_n}{\lambda_n}) - j(x_1 + \cdots + x_n), \end{aligned}$$

we have (20) by (21). ■

Fact 3.4 ([7, Theorem 5.4]) $(\lambda_1 f_1^* + \cdots + \lambda_n f_n^*)^* \leq p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) \leq \lambda_1 f_1 + \cdots + \lambda_n f_n$.

Proof. This follows from (17) or (19). ■

Fact 3.5 ([7, Theorem 8.5]) *Let $x \in \mathbb{R}^N$. Then the function*

$$]0, +\infty[\; \to \;]-\infty, +\infty] : \mu \mapsto p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) \quad \textit{is decreasing.} \tag{22}$$

Consequently, $\lim_{\mu \to 0^+} p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x)$ *and* $\lim_{\mu \to +\infty} p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x)$ *exist. In fact,*

$$\lim_{\mu \to 0^+} p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) = \sup_{\mu > 0} p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) = \big(\lambda_1 f_1 + \cdots + \lambda_n f_n\big)(x) \tag{23}$$

and

$$\lim_{\mu \to +\infty} p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) = \inf_{\mu > 0} p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) = \big(\lambda_1 \star f_1 \Box \cdots \Box \lambda_n \star f_n\big)(x). \tag{24}$$

Proof. (22) follows from (17). (24) also follows from (17).

To see (23), by (20), $\forall x \in \mathbb{R}^N$,

$$\begin{aligned} p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) &\geq \lambda_1 \inf_{x_1}\left(f_1(x_1/\lambda_1) + \frac{1}{\mu} j(x - x_1/\lambda_1)\right) + \cdots + \lambda_n \inf_{x_n}\left(f_n(x_n/\lambda_n) + \frac{1}{\mu} j(x - x_n/\lambda_n)\right) \\ &= \lambda_1 e_\mu f_1(x) + \cdots + \lambda_n e_\mu f_n(x), \end{aligned}$$

where $e_\mu f_i = f_i \Box (1/\mu j)$. Then

$$\lambda_1 e_\mu f_1 + \cdots + \lambda_n e_\mu f_n \leq p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) \leq \lambda_1 f_1 + \cdots + \lambda_n f_n,$$

so that

$$\lim_{\mu \to 0^+} p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) = \lambda_1 f_1 + \cdots + \lambda_n f_n,$$

since $\lim_{\mu\to 0^+} e_\mu f_i = f_i$ by [24, Theorem 2.26 and Theorem 1.25]. ■

Fact 3.6 ([23, Theorem 25.7]) *Let C be a nonempty open convex subset of $\mathbb{R}^N$, and let f be a convex function which is finite and differentiable on C. Let $f_1, f_2, \ldots,$ be a sequence of convex functions finite and differentiable on C such that $\lim_{i\to\infty} f_i(x) = f(x)$ for every $x \in C$. Then*

$$\lim_i \nabla f_i(x) = \nabla f(x), \quad \forall x \in C.$$

In fact, the sequence of gradients ∇f_i converges to ∇f uniformly on every compact subset of C.

Fact 3.7 ([23, page 108]) *Let $Q \in \mathbb{S}^N_{++}$. Then $(q_Q)^* = q_{Q^{-1}}$.*

Fact 3.8 ([23, Theorem 23.5]) *Let $f : \mathbb{R}^N \to \,]-\infty, +\infty]$ be a proper lower semicontinuous convex function. Then $\partial f^* = (\partial f)^{-1}$.*

4 Main results

We start by computing the proximal average of general linear-quadratic functions thereby extending [7, Example 4.5] and [8, Example 7.4].

Lemma 4.1 *Let $A_i \in \mathbb{S}^N_+$, $b_i \in \mathbb{R}^N$, $r_i \in \mathbb{R}$. If each $f_i = q_{A_i} + \langle b_i, \cdot\rangle + r_i$, i.e., linear-quadratic, then $\forall x^*$,*

$$\begin{aligned} p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x^*) = q_{\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})}(x^*) + \langle x^*, \big(\sum_{i=1}^n \lambda_i (A_i + \mu^{-1}\operatorname{Id})^{-1}\big)^{-1} \sum_{i=1}^n \lambda_i (A_i + \mu^{-1}\operatorname{Id})^{-1} b_i\rangle + \\ q_{\left(\sum_{i=1}^n \lambda_i (A_i+\mu^{-1}\operatorname{Id})^{-1}\right)^{-1}}\big(\sum_{i=1}^n \lambda_i (A_i + \mu^{-1}\operatorname{Id})^{-1} b_i\big) - \sum_{i=1}^n \lambda_i \big(q_{(A_i+\mu^{-1}\operatorname{Id})^{-1}}(b_i) - r_i\big). \end{aligned} \quad (25)$$

In particular, if $(\forall i)$ f_i *is quadratic, i.e.,* $b_i = 0, r_i = 0$*, then* $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ *is quadratic with*

$$p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = q_{\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})};$$

If $(\forall i)$ f_i *is affine, i.e.,* $A_i = 0$*, then* $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ *is affine.*

Proof. We have $f_i + \mu^{-1} j = q_{(A_i+\mu^{-1}\,\mathrm{Id})} + \langle b_i, \cdot\rangle + r_i$ and by Fact 3.7

$$\begin{aligned}(f_i + \mu^{-1}j)^*(x^*) &= q_{(A_i+\mu^{-1}\,\mathrm{Id})^{-1}}(x^* - b_i) - r_i\\ &= q_{(A_i+\mu^{-1}\,\mathrm{Id})^{-1}}(x^*) - \langle x^*, (A_i + \mu^{-1}\,\mathrm{Id})^{-1} b_i\rangle + q_{(A_i+\mu^{-1}\,\mathrm{Id})^{-1}}(b_i) - r_i.\end{aligned}$$

Then $\big(\lambda_1(f_1 + \mu^{-1}j)^* + \cdots + \lambda_n(f_n + \mu^{-1}j)^*\big)(x^*) =$

$$\begin{aligned}&\sum_{i=1}^n \lambda_i\Big(q_{(A_i+\mu^{-1}\,\mathrm{Id})^{-1}}(x^*) - \langle x^*, (A_i + \mu^{-1}\,\mathrm{Id})^{-1} b_i\rangle + q_{(A_i+\mu^{-1}\,\mathrm{Id})^{-1}}(b_i) - r_i\Big)\\ &= q_{\sum_{i=1}^n \lambda_i(A_i+\mu^{-1}\,\mathrm{Id})^{-1}}(x^*) - \langle x^*, \sum_{i=1}^n \lambda_i (A_i + \mu^{-1}\,\mathrm{Id})^{-1} b_i\rangle + \sum_{i=1}^n \lambda_i\big(q_{(A_i+\mu^{-1}\,\mathrm{Id})^{-1}}(b_i) - r_i\big).\end{aligned}$$

It follows that $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x^*) =$

$$q_{[\sum_{i=1}^n \lambda_i(A_i+\mu^{-1}\,\mathrm{Id})^{-1}]^{-1}}(x^* + \sum_{i=1}^n \lambda_i (A_i + \mu^{-1}\,\mathrm{Id})^{-1} b_i) - \sum_{i=1}^n \lambda_i\big(q_{(A_i+\mu^{-1}\,\mathrm{Id})^{-1}}(b_i) - r_i\big) - q_{\mu^{-1}\,\mathrm{Id}}(x^*).$$

As

$$\begin{aligned}&q_{[\sum_{i=1}^n \lambda_i(A_i+\mu^{-1}\,\mathrm{Id})^{-1}]^{-1}}(x^* + \sum_{i=1}^n \lambda_i (A_i + \mu^{-1}\,\mathrm{Id})^{-1} b_i)\\ &= q_{[\sum_{i=1}^n \lambda_i(A_i+\mu^{-1}\,\mathrm{Id})^{-1}]^{-1}}(x^*) + \langle x^*, [\sum_{i=1}^n \lambda_i(A_i + \mu^{-1}\,\mathrm{Id})^{-1}]^{-1} \sum_{i=1}^n \lambda_i (A_i + \mu^{-1}\,\mathrm{Id})^{-1} b_i\rangle +\\ &q_{[\sum_{i=1}^n \lambda_i(A_i+\mu^{-1}\,\mathrm{Id})^{-1}]^{-1}}(\sum_{i=1}^n \lambda_i (A_i + \mu^{-1}\,\mathrm{Id})^{-1} b_i),\end{aligned}$$

we obtain that $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x^*) =$

$$\begin{aligned}&q_{[\sum_{i=1}^n \lambda_i(A_i+\mu^{-1}\,\mathrm{Id})^{-1}]^{-1} - \mu^{-1}\,\mathrm{Id}}(x^*) + \langle x^*, [\sum_{i=1}^n \lambda_i(A_i + \mu^{-1}\,\mathrm{Id})^{-1}]^{-1} \sum_{i=1}^n \lambda_i (A_i + \mu^{-1}\,\mathrm{Id})^{-1} b_i\rangle +\\ &q_{[\sum_{i=1}^n \lambda_i(A_i+\mu^{-1}\,\mathrm{Id})^{-1}]^{-1}}(\sum_{i=1}^n \lambda_i (A_i + \mu^{-1}\,\mathrm{Id})^{-1} b_i) - \sum_{i=1}^n \lambda_i\big(q_{(A_i+\mu^{-1}\,\mathrm{Id})^{-1}}(b_i) - r_i\big),\end{aligned}$$

which is (25). The remaining claims are immediate from (25) and that $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) = 0$ when $(\forall\ i)$ $A_i = 0$ by Proposition 2.3. ■

We are ready for our main result:

Theorem 4.2 (harmonic-resolvent-arithmetic average inequality and limits) *Let $A_1, \ldots, A_n \in \mathbb{S}^N_{++}$. We have*

(i)
$$\mathcal{H}(\boldsymbol{A}, \boldsymbol{\lambda}) \preceq \mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) \preceq \mathcal{A}(\boldsymbol{A}, \boldsymbol{\lambda}); \tag{26}$$
In particular, $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) \in \mathbb{S}^N_{++}$.

(ii) $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) \to \mathcal{A}(\boldsymbol{A}, \boldsymbol{\lambda})$ *when* $\mu \to 0^+$.

(iii) $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) \to \mathcal{H}(\boldsymbol{A}, \boldsymbol{\lambda})$ *when* $\mu \to +\infty$.

Proof. (i). According to Fact 3.4,

$$(\lambda_1 f_1^* + \cdots + \lambda_n f_n^*)^* \leq p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) \leq \lambda_1 f_1 + \cdots + \lambda_n f_n. \tag{27}$$

Let $f_i = q_{A_i}$. Using $(q_{A_i})^* = q_{A_i^{-1}}$ (by Fact 3.7) and Lemma 4.1 we have

$$\begin{aligned}(\lambda_1 f_1^* + \cdots + \lambda_n f_n^*)^* &= (\lambda_1 q_{A_1^{-1}} + \cdots + \lambda_n q_{A_n^{-1}})^* = (q_{\lambda_1 A_1^{-1} + \cdots + \lambda_n A_n^{-1}})^* \\ &= q_{(\lambda_1 A_1^{-1} + \cdots + \lambda_n A_n^{-1})^{-1}} = q_{\mathcal{H}(\boldsymbol{A}, \boldsymbol{\lambda})}. \end{aligned} \tag{28}$$

$$\lambda_1 f_1 + \cdots + \lambda_n f_n = q_{\lambda_1 A_1 + \cdots + \lambda_n A_n} = q_{\mathcal{A}(\boldsymbol{A}, \boldsymbol{\lambda})}, \tag{29}$$

$$p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = q_{\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})}. \tag{30}$$

Then (27) becomes

$$q_{\mathcal{H}(\boldsymbol{A}, \boldsymbol{\lambda})} \leq q_{\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})} \leq q_{\mathcal{A}(\boldsymbol{A}, \boldsymbol{\lambda})}.$$

As $q_X \leq q_Y \Leftrightarrow X \preceq Y$, (26) is established. Since $A_i \in \mathbb{S}^N_{++}$, $A_i^{-1} \in \mathbb{S}^N_{++}$, $\lambda_1 A_1^{-1} + \cdots + \lambda_n A_n^{-1} \in \mathbb{S}^N_{++}$, we have $\mathcal{H}(\boldsymbol{A}, \boldsymbol{\lambda}) = (\lambda_1 A_1^{-1} + \cdots + \lambda_n A_n^{-1})^{-1} \in \mathbb{S}^N_{++}$, thus $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) \in \mathbb{S}^N_{++}$ by (26). (Alternatively, apply Theorem 2.5.)

(ii) and (iii): Observe that $(\forall i)$ $\big(\lambda_i \star f_i\big)^* = \lambda_i f_i^* = \lambda_i q_{A_i^{-1}}$ has full domain, by [23, Theorem 16.4],

$$(\lambda_1 f_1^* + \cdots + \lambda_n f_n^*)^* = \big(\lambda_1 \star f_1 \Box \cdots \Box \lambda_n \star f_n\big).$$

By Fact 3.5, $\forall x \in \mathbb{R}^N$ one has

$$\lim_{\mu \to 0^+} p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) = (\lambda_1 f_1 + \cdots + \lambda_n f_n)(x),$$

$$\lim_{\mu \to +\infty} p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) = (\lambda_1 f_1^* + \cdots + \lambda_n f_n^*)^*(x).$$

Since $(\forall i)$ f_i, f_i^* are differentiable on $\mathbb{R}^N$, so is $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ by [7, Corollary 7.7]. According to Fact 3.6, $\forall x$

$$\lim_{\mu \to 0^+} \nabla p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) = \lambda_1 \nabla f_1(x) + \cdots + \lambda_n \nabla f_n(x), \tag{31}$$

$$\lim_{\mu \to +\infty} \nabla p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) = \nabla(\lambda_1 f_1^* + \cdots + \lambda_n f_n^*)^*(x). \tag{32}$$

Moreover, the convergences in (31)-(32) are uniform on every closed bounded subset of $\mathbb{R}^N$. Now it follows from (28)-(30) that $\nabla p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) = \mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})$, $\nabla(\lambda_1 f_1 + \cdots + \lambda_n f_n) = \mathcal{A}(\boldsymbol{A},\boldsymbol{\lambda})$, $\nabla(\lambda_1 f_1^* + \cdots + \lambda_n f_n^*)^* = \mathcal{H}(\boldsymbol{A},\boldsymbol{\lambda})$. (31)-(32) transpire to

$$\lim_{\mu\to 0^+} \mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})x = \mathcal{A}(\boldsymbol{A},\boldsymbol{\lambda})x, \tag{33}$$

$$\lim_{\mu\to +\infty} \mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})x = \mathcal{H}(\boldsymbol{A},\boldsymbol{\lambda})x, \tag{34}$$

where the convergences are uniform on every closed bounded subset of $\mathbb{R}^N$. Hence (ii) and (iii) follow from (33) and (34). ■

Note that in Theorem 4.2(ii),(iii), there is no ambiguity since all norms in finite dimensional spaces are equivalent. The following remark is due to an anonymous referee.

Remark 4.3 Given $\boldsymbol{A} = (A_1,\ldots,A_n)$, where $A_1,\ldots,A_n$ are in $\mathbb{S}^N_+$, and $\boldsymbol{\lambda} = (\lambda_1,\ldots,\lambda_n) \in \mathbb{R}^n_{++}$, consider the map

$$f\colon \mu \mapsto \Big(\sum_{i=1}^n \lambda_i \big(\mu A_i + \mathrm{Id}\big)^{-1}\Big)^{-1}, \tag{35}$$

which is well defined on a neighborhood of 0. Now (7) implies

$$\begin{aligned}\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda}) &= \frac{1}{\mu}\Big(\Big(\lambda_1(\mu A_1 + \mathrm{Id})^{-1} + \cdots + \lambda_n(\mu A_n + \mathrm{Id})^{-1}\Big)^{-1} - \mathrm{Id}\Big)\\ &= \frac{f(\mu) - f(0)}{\mu}.\end{aligned} \tag{36}$$

Consequently, Theorem 4.2(ii) states that $f'(0) = \mathcal{A}(\boldsymbol{A},\boldsymbol{\lambda})$.

Definition 4.4 *A function $g\colon \mathbb{D} \to \mathbb{S}^N$, where $\mathbb{D}$ is a convex subset of $\mathbb{S}^N$, is matrix convex if $\forall A_1, A_2 \in \mathbb{D}$, $\forall \lambda \in [0,1]$,*

$$g(\lambda A_1 + (1-\lambda)A_2) \preceq \lambda g(A_1) + (1-\lambda) g(A_2).$$

Matrix concave functions are defined similarly.

It is easy to see that a symmetric matrix valued function g is matrix concave (resp. convex) if and only if $\forall x \in \mathbb{R}^N$ the function $A \mapsto q_{g(A)}(x)$ is concave (resp. convex). Some immediate consequences of Theorem 4.2 on matrix-valued functions are:

Corollary 4.5 *Assume that $(\forall i)$ $A_i \in \mathbb{S}^N_{++}$ and $\sum_{i=1}^n \lambda_i = 1$ with $\lambda_i > 0$. Then*

$$(\lambda_1 A_1 + \cdots + \lambda_n A_n)^{-1} \preceq \lambda_1 A_1^{-1} + \cdots + \lambda_n A_n^{-1}.$$

Consequently, the matrix function $X \mapsto X^{-1}$ is matrix convex on $\mathbb{S}^N_{++}$.

Proof. Apply Theorem 4.2 equation (26) for $\boldsymbol{A} = (A_1^{-1}, \cdots, A_n^{-1})$. ■

Corollary 4.6 *For every $\mu > 0$, the resolvent average matrix function $\boldsymbol{A} \mapsto \mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})$ given by*

$$(A_1, \cdots, A_n) \mapsto [\lambda_1(A_1 + \mu^{-1}\,\mathrm{Id})^{-1} + \cdots + \lambda_n(A_n + \mu^{-1}\,\mathrm{Id})^{-1}]^{-1} - \mu^{-1}\,\mathrm{Id} \quad \textit{is matrix concave on } \mathbb{S}^N_{++} \times \cdots \times \mathbb{S}^N_{++}. \tag{37}$$

For each $\lambda = (\lambda_1, \cdots, \lambda_n)$ with $\sum_{i=1}^n \lambda_i = 1$ and $\lambda_i > 0\ \forall i$, the harmonic average matrix function

$$(A_1, \cdots, A_n) \mapsto (\lambda_1 A_1^{-1} + \cdots + \lambda_n A_n^{-1})^{-1} \textit{ is matrix concave} \tag{38}$$

on $\mathbb{S}^N_{++} \times \cdots \times \mathbb{S}^N_{++}$. Consequently, the harmonic average function

$$(x_1, \cdots, x_n) \mapsto \frac{1}{x_1^{-1} + \cdots + x_n^{-1}} \textit{ is concave} \tag{39}$$

on $\mathbb{R}_{++} \times \cdots \times \mathbb{R}_{++}$.

Proof. Set $f_i = q_{A_i}$. Then $\forall x \in \mathbb{R}^N$, we have $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) =$

$$\min_{\lambda_1 x_1 + \cdots + \lambda_n x_n = x} \Big(\big(\lambda_1 q_{A_1}(x_1) + \cdots + \lambda_n q_{A_n}(x_n)\big) + \big(\mu^{-1}\lambda_1 q_{\mathrm{Id}}(x_1) + \cdots + \mu^{-1}\lambda_n q_{\mathrm{Id}}(x_n)\big) \Big) - \mu^{-1} q_{\mathrm{Id}}(x).$$

Since for each fixed $(x_1, \ldots, x_n)$,

$$(A_1, \cdots, A_n) \mapsto \big(\lambda_1 q_{A_1}(x_1) + \cdots + \lambda_n q_{A_n}(x_n)\big) + \big(\mu^{-1} q_{\mathrm{Id}}(x_1) + \cdots + \mu^{-1} q_{\mathrm{Id}}(x_n)\big),$$

is affine, being the infimum of affine functions we have that $\forall x$ the function

$$(A_1, \cdots, A_n) \mapsto p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x),$$

is concave. As $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) = q_{\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})}(x)$ by Lemma 4.1, this shows that $\forall x \in \mathbb{R}^N$ the function

$$\boldsymbol{A} = (A_1, \cdots, A_n) \mapsto q_{\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})}(x) \text{ is concave,}$$

so $\boldsymbol{A} \mapsto \mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda})$ is matrix concave.

Now by Theorem 4.2(iii), $\mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) \to \mathcal{H}(\boldsymbol{A}, \boldsymbol{\lambda})$ when $\mu \to +\infty$. This and (37) implies that

$$\boldsymbol{A} \mapsto \mathcal{H}(\boldsymbol{A}, \boldsymbol{\lambda}),$$

is also matrix concave. (39) follows from (38) by setting $N = 1$ and $\lambda_1 = \cdots = \lambda_n = 1/n$. ■

Remark 4.7 Corollary 4.5 is well-known, cf. [24, Proposition 2.56 on page 73]. Corollary 4.6 (39) is also well-known, cf. [12, Exercise 3.17 on page 116].

We proceed to show that resolvent averages of matrices enjoy self-duality.

Theorem 4.8 (self-duality) *Let* $(\forall\ i)\ A_i \in \mathbb{S}^N_{++}$ *and* $\mu > 0$. *Assume that* $\sum_{i=1}^n \lambda_i = 1$ *with* $\lambda_i > 0$. *Then*

$$\big[\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})\big]^{-1} = \mathcal{R}_{\mu^{-1}}(\boldsymbol{A}^{-1},\boldsymbol{\lambda}),\ \ i.e., \tag{40}$$

$$\left[\Big(\lambda_1(A_1+\mu^{-1}\,\mathrm{Id})^{-1}+\cdots+\lambda_n(A_n+\mu^{-1}\,\mathrm{Id})^{-1}\Big)^{-1}-\mu^{-1}\,\mathrm{Id}\right]^{-1} =$$
$$\Big(\lambda_1(A_1^{-1}+\mu\,\mathrm{Id})^{-1}+\cdots+\lambda_n(A_n^{-1}+\mu\,\mathrm{Id})^{-1}\Big)^{-1}-\mu\,\mathrm{Id}\,.$$

In particular, for $\mu = 1$, $[\mathcal{R}_1(\boldsymbol{A},\boldsymbol{\lambda})]^{-1} = \mathcal{R}_1(\boldsymbol{A}^{-1},\boldsymbol{\lambda})$.

Proof. Let $f_i = q_{A_i}$. By Fact 3.2, $\big(p_\mu(\boldsymbol{f},\boldsymbol{\lambda})\big)^* = p_{\mu^{-1}}(\boldsymbol{f}^*,\boldsymbol{\lambda})$, taking subgradients both sides, followed by using Fact 3.8, we obtain that

$$\partial\big(p_\mu(\boldsymbol{f},\boldsymbol{\lambda})\big)^* = \big(\partial p_\mu(\boldsymbol{f},\boldsymbol{\lambda})\big)^{-1} = \partial\big(p_{\mu^{-1}}(\boldsymbol{f}^*,\boldsymbol{\lambda})\big).$$

By Lemma 4.1, $p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) = q_{\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})}, p_{\mu^{-1}}(\boldsymbol{f}^*,\boldsymbol{\lambda}) = q_{\mathcal{R}_{\mu^{-1}}(\boldsymbol{A}^{-1},\boldsymbol{\lambda})}$, we have

$$\partial p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) = \mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda}),$$
$$\partial p_{\mu^{-1}}(\boldsymbol{f}^*,\boldsymbol{\lambda}) = \mathcal{R}_{\mu^{-1}}(\boldsymbol{A}^{-1},\boldsymbol{\lambda}).$$

Hence

$$\big[\mathcal{R}_\mu(\boldsymbol{A},\boldsymbol{\lambda})\big]^{-1} = \mathcal{R}_{\mu^{-1}}(\boldsymbol{A}^{-1},\boldsymbol{\lambda}),$$

as claimed. ■

Remark 4.9 Although the harmonic and arithmetic average lack self-duality, they are dual to each other:

$$[\mathcal{H}(\boldsymbol{A},\boldsymbol{\lambda})]^{-1} = \lambda_1 A_1^{-1}+\cdots+\lambda_n A_n^{-1} = \mathcal{A}(\boldsymbol{A}^{-1},\boldsymbol{\lambda}),$$
$$[\mathcal{A}(\boldsymbol{A},\boldsymbol{\lambda})]^{-1} = \big[\lambda_1(A_1^{-1})^{-1}+\cdots+\lambda_n(A_n^{-1})^{-1}\big]^{-1} = \mathcal{H}(\boldsymbol{A}^{-1},\boldsymbol{\lambda}).$$

5 A comparison to weighted geometric means

If $A, B \in \mathbb{S}^N_{++}$, the geometric mean is defined by

$$A\#B = A^{1/2}\big(A^{-1/2}BA^{-1/2}\big)^{1/2}A^{1/2}.$$

In general, the geometric mean of $A_1,\ldots,A_n \in \mathbb{S}^N_+$ for $n \geq 3$ is defined either as the limit of an inductive procedure or by the Riemannian distance without a closed form [2, 22, 19, 15].

We are grateful to an anonymous referee for the following remark.

Remark 5.1 The important problem of extending the geometric mean $A\#B$ of positive definite matrices A and B to n positive definite matrices has been done successfully by Ando-Li-Mathias [2] and by Bhatia-Holbrook [10] via the standard "symmetrization procedure" and the "least square" method, respectively, without a closed form. The least square mean is also called the Cartan mean in the context of Riemannian geometry and its monotonicity is still open; however, the Ando-Li-Mathias mean and the resolvent average are monotone.

The λ-weighted geometric mean $A\#_\lambda B$ of A and B is defined by

$$A\#_\lambda B = A^{1/2}(A^{-1/2}BA^{-1/2})^\lambda A^{1/2}$$

and it has the following geometric interpretation: the curve $\lambda \mapsto A\#_\lambda B$ is the unique geodesic line passing from A to B in the Riemannian manifold of positive definite matrices. The problem of extending the weighted geometric mean $A\#_\lambda B$ to n positive definite matrices via the "symmetrization procedure" is still open. However, the Cartan mean has a natural extension to multi-variable weighted mean via the least square method [10], and the resolvent average has also a closed form. If the weighted geometric mean of n positive definite matrices exists through the symmetrization procedure, then it should have the invariance property under congruence transformation $X \mapsto MXM^T$, where M is invertible. The resolvent average, however, has this property only for orthogonal matrices; see (13).

To compare the resolvent average with the well-known geometric mean, we restrict our attention to non-negative real numbers (1×1 matrices). When $\boldsymbol{A} = \boldsymbol{x} = (x_1, \ldots, x_n)$ with $x_i \in \mathbb{R}_+$ and $\mu = 1$, we write

$$\mathcal{R}(\boldsymbol{x}, \boldsymbol{\lambda}) = \mathcal{R}_\mu(\boldsymbol{A}, \boldsymbol{\lambda}) = \big(\lambda_1(x_1+1)^{-1} + \cdots + \lambda_n(x_n+1)^{-1}\big)^{-1} - 1,$$

and $\boldsymbol{x}^{-1} = (1/x_1, \ldots, 1/x_n)$ when $(\forall i)\ x_i \in \mathbb{R}_{++}$.

Proposition 5.2 *Let* $(\forall\ i)\ x_i > 0, y_i > 0$. *We have*

(i) (**harmonic-resolvent-arithmetic mean inequality**)*:*

$$\big(\lambda_1 x_1^{-1} + \cdots + \lambda_n x_n^{-1}\big)^{-1} \leq \mathcal{R}(\boldsymbol{x}, \boldsymbol{\lambda}) \leq \lambda_1 x_1 + \cdots + \lambda_n x_n. \tag{41}$$

Moreover, $\mathcal{R}(\boldsymbol{x}, \boldsymbol{\lambda}) = \lambda_1 x_1 + \cdots + \lambda_n x_n$ *if and only if* $x_1 = \cdots = x_n$.

(ii) (**self-duality**)*:* $[\mathcal{R}(\boldsymbol{x}, \boldsymbol{\lambda})]^{-1} = \mathcal{R}(\boldsymbol{x}^{-1}, \boldsymbol{\lambda})$.

(iii) If $\boldsymbol{x} = (x_1, \ldots, x_1)$, *then* $\mathcal{R}(\boldsymbol{x}, \boldsymbol{\lambda}) = x_1$.

(iv) If $\boldsymbol{x} = (x_1, x_1^{-1}, x_2, x_2^{-1}, \ldots, x_n, x_n^{-1})$ *and* $\boldsymbol{\lambda} = (\frac{1}{2n}, \ldots, \frac{1}{2n})$, *then* $\mathcal{R}(\boldsymbol{x}, \boldsymbol{\lambda}) = 1$.

(v) The function $\boldsymbol{x} \mapsto \mathcal{R}(\boldsymbol{x}, \boldsymbol{\lambda})$ *is concave on* $\mathbb{R}_{++} \times \cdots \times \mathbb{R}_{++}$.

(vi) If $\boldsymbol{x} \succeq \boldsymbol{y}$, *then* $\mathcal{R}(\boldsymbol{x}, \boldsymbol{\lambda}) \geq \mathcal{R}(\boldsymbol{y}, \boldsymbol{\lambda})$.

Proof. (i): For (41), apply Theorem 4.2(i) with $\mu = 1$. Now $\mathcal{R}(\boldsymbol{x}, \boldsymbol{\lambda}) = \lambda_1 x_1 + \cdots + \lambda_n x_n$ is equivalent to

$$\big(\lambda_1(x_1+1)^{-1} + \cdots + \lambda_n(x_n+1)^{-1}\big)^{-1} = \lambda_1 x_1 + \cdots + \lambda_n x_n + 1, \tag{42}$$

As $\sum_{i=1}^n \lambda_i = 1$, (42) is the same as

$$\lambda_1 \frac{1}{(x_1+1)} + \cdots + \lambda_n \frac{1}{(x_n+1)} = \frac{1}{\lambda_1(x_1+1) + \cdots + \lambda_n(x_n+1)}.$$

Since the function $x \mapsto 1/x$ is strictly convex on $\mathbb{R}_{++}$, we must have $x_1 = \cdots = x_n$.

(ii): Theorem 4.8. (iii): Proposition 2.3. (iv): Proposition 2.2. (v): Corollary 4.6. (vi): Proposition 2.4. ∎

Recall the *weighted geometric mean*:

$$\mathcal{G}(\boldsymbol{x}, \boldsymbol{\lambda}) = x_1^{\lambda_1} x_2^{\lambda_2} \cdots x_n^{\lambda_n}.$$

$\mathcal{G}(\boldsymbol{x}, \boldsymbol{\lambda})$ always has the following properties:

Fact 5.3 *Let* $(\forall\ i)\ x_i > 0, y_i > 0$. *We have*

(i) (**harmonic-geometric-arithmetic mean inequality**)*:*

$$\big(\lambda_1 x_1^{-1} + \cdots + \lambda_n x_n^{-1}\big)^{-1} \le \mathcal{G}(\boldsymbol{x}, \boldsymbol{\lambda}) \le \lambda_1 x_1 + \cdots + \lambda_n x_n.$$

Moreover, $\mathcal{G}(\boldsymbol{x}, \boldsymbol{\lambda}) = \lambda_1 x_1 + \cdots + \lambda_n x_n$ *if and only* $x_1 = \cdots = x_n$.

(ii) (**self-duality**)*:* $[\mathcal{G}(\boldsymbol{x}, \boldsymbol{\lambda})]^{-1} = \mathcal{G}(\boldsymbol{x}^{-1}, \boldsymbol{\lambda})$.

(iii) If $\boldsymbol{x} = (x_1, \ldots, x_1)$, *then* $\mathcal{G}(\boldsymbol{x}, \boldsymbol{\lambda}) = x_1$.

(iv) If $\boldsymbol{x} = (x_1, x_1^{-1}, x_2, x_2^{-1}, \ldots, x_n, x_n^{-1})$ *and* $\boldsymbol{\lambda} = (\frac{1}{2n}, \ldots, \frac{1}{2n})$, *then* $\mathcal{G}(\boldsymbol{x}, \boldsymbol{\lambda}) = 1$.

(v) The function $\boldsymbol{x} \mapsto \mathcal{G}(\boldsymbol{x}, \boldsymbol{\lambda})$ *is concave on* $\mathbb{R}_{++} \times \cdots \times \mathbb{R}_{++}$.

(vi) If $\boldsymbol{x} \succeq \boldsymbol{y}$, *then* $\mathcal{G}(\boldsymbol{x}, \boldsymbol{\lambda}) \ge \mathcal{G}(\boldsymbol{y}, \boldsymbol{\lambda})$.

Proof. (i): See [23, page 29]. (ii)-(iv) and (vi) are simple. (v): See [24, Example 2.53]. ∎

The means $\mathcal{R}(\boldsymbol{x}, \boldsymbol{\lambda})$ and $\mathcal{G}(\boldsymbol{x}, \boldsymbol{\lambda})$ have strikingly similar properties. Are they the same?

Example 5.4 (i). Let $\lambda = (\frac{1}{2}, \frac{1}{2})$. When $x = (\frac{1}{4}, 1)$, $\mathcal{G}(\boldsymbol{x}, \boldsymbol{\lambda}) = \frac{1}{2}$ but $\mathcal{R}(\boldsymbol{x}, \boldsymbol{\lambda}) = \frac{7}{13}$, so $\mathcal{G}(\boldsymbol{x}, \boldsymbol{\lambda}) \neq \mathcal{R}(\boldsymbol{x}, \boldsymbol{\lambda})$.

(ii). *Is it right that* $\mathcal{G}(\boldsymbol{x}, \boldsymbol{\lambda}) \le \mathcal{R}(\boldsymbol{x}, \boldsymbol{\lambda})\ \forall\ x \in \mathbb{R}^2_{++}$*?* The answer is also no. Assume to the contrary that $\mathcal{G}(\boldsymbol{x}, \boldsymbol{\lambda}) \le \mathcal{R}(\boldsymbol{x}, \boldsymbol{\lambda})$, $\forall\ \boldsymbol{x} \in \mathbb{R}_{++} \times \mathbb{R}_{++}$. Taking inverse both sides, followed by applying the self-duality of $\mathcal{G}(\boldsymbol{x}, \boldsymbol{\lambda}), \mathcal{R}(\boldsymbol{x}, \boldsymbol{\lambda})$, gives

$$\mathcal{G}(\boldsymbol{x}, \boldsymbol{\lambda})^{-1} \ge \mathcal{R}(\boldsymbol{x}, \boldsymbol{\lambda})^{-1} = \mathcal{R}(\boldsymbol{x}^{-1}, \boldsymbol{\lambda}) \ge \mathcal{G}(\boldsymbol{x}^{-1}, \boldsymbol{\lambda}) = \mathcal{G}(\boldsymbol{x}, \boldsymbol{\lambda})^{-1},$$

and this gives that $\mathcal{G}(\boldsymbol{x}, \boldsymbol{\lambda})^{-1} = \mathcal{R}(\boldsymbol{x}, \boldsymbol{\lambda})^{-1}$ so that $\mathcal{G}(\boldsymbol{x}, \boldsymbol{\lambda}) = \mathcal{R}(\boldsymbol{x}, \boldsymbol{\lambda})$. This is a contradiction to (i).

Finally, we note that the resolvent average can be defined for general monotone operators and that Theorem 4.8 holds even when $A_1, \ldots, A_n$ are monotone operators (not necessarily positive semi-definite matrices), in that situation one needs to use *set-valued inverses*. This and further details on the resolvent average for general monotone operators will appear in the forthcoming paper [5].

Acknowledgment

The authors thank an anonymous referee for his/her insightful comments. Heinz Bauschke was partially supported by the Canada Research Chair program and by the Natural Sciences and Engineering Research Council of Canada. Sarah Moffat was partially supported by the Natural Sciences and Engineering Research Council of Canada. Xianfu Wang was partially supported by the Natural Sciences and Engineering Research Council of Canada.

References

[1] W. N. Anderson Jr. and R. J. Duffin, "Series and parallel addition of matrices", *J. Math. Anal. Appl.* 26, pp. 576–594, 1969.

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

[3] H. H. Bauschke and J. M. Borwein, "Legendre functions and the method of random Bregman projections", *J. Convex Anal.* 4, pp. 27–67, 1997.

[4] H. H. Bauschke, E. Matoušková, and S. Reich, "Projection and proximal point methods: convergence results and counterexamples", *Nonlinear Anal.* 56, pp. 715–738, 2004.

[5] H. H. Bauschke, S. M. Moffat, and X. Wang, "Resolvent averages of monotone operators", preprint.

[6] H. H. Bauschke and X. Wang, "The kernel average for two convex functions and its applications to the extension and representation of monotone operators", *Trans. Amer. Math. Soc.* 361, pp. 5947–5965, 2009.

[7] H. H. Bauschke, R. Goebel, Y. Lucet, and X. Wang, "The proximal average: basic theory", *SIAM J. Optim.* 19, pp. 766–785, 2008.

[8] H. H. Bauschke, Y. Lucet, and M. Trienis, "How to transform one convex function continuously into another", *SIAM Rev.* 50, pp. 115–132, 2008.

[9] H. H. Bauschke, Y. Lucet, and X. Wang, "Primal-dual symmetric intrinsic methods for finding antiderivatives of cyclically monotone operators", *SIAM J. Control Optim.* 46, pp. 2031–2051, 2007.

[10] R. Bhatia and J. Holbrook, "Riemannian geometry and matrix geometric means", *Linear Algebra Appl.* 413, pp. 594–618, 2006.

[11] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization*, Springer-Verlag, New York, 2000.

[12] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.

[13] J.-B. Hiriart-Urruty and M.-L. Mazure, "Formulations variationnelles de l'addition parallèle et de la soustraction parallèlle d'opérateurs semi-définis positifs", *C. R. Acad. Sci. Paris Sér. I Math.* 302, pp. 527–530, 1986.

[14] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.

[15] F. Kubo and T. Ando, "Means of positive linear operators", *Math. Ann.* 246, no. 3, pp. 205–224, 1980.

[16] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, San Diego, CA, 1979.

[17] M.-L. Mazure, "L'addition parallèle d'opérateurs interprétée comme inf-convolution de formes quadratiques convexes", *RAIRO Modél. Math. Anal. Numér.* 20, pp 497–515, 1986.

[18] G. J. Minty, "Monotone (nonlinear) operators in Hilbert space", *Duke Mathematical Journal* 29, pp. 341–346, 1962.

[19] M. Moakher, "A differential geometric approach to the geometric mean of symmetric positive-definite matrices", *SIAM J. Matrix Anal. Appl.* 26, pp. 735–747, 2005.

[20] J.-J. Moreau, "Proximité et dualité dans un espace hilbertien", *Bull. Soc. Math. France* 93, pp. 273–299, 1965.

[21] G. B. Passty, "The parallel sum of nonlinear monotone operators", *Nonlinear Anal.* 10, pp. 215–227, 1986.

[22] D. Petz, "Means of positive matrices: geometry and a conjecture", *Ann. Math. Inform.* 32, pp. 129–139, 2005.

[23] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.

[24] R. T. Rockafellar and R. J-B Wets, *Variational Analysis*, Springer-Verlag, 1998.