REGULARIZED MEDIANS ON SYMMETRIC CONES

Sangho Kum[1] and Jen-Chih Yao[2]

We are concerned with an extension of main results of [6] *into a general symmetric cone* Ω *from the convex cone of positive definite matrices* $\mathbb{P}$. *To be more specific, two regularized median optimization problems are introduced and the existence and uniqueness of solutions are studied on* Ω. *Moreover the Lipschitz continuity of the gradient of objective functions of the regularized median optimizations are provided for a possible design of gradient-based methods of finding the unique minimizer. Based on some results of* [7], *we present purely Jordan-algebraic techniques of proof in comparison with matrix-analytic ones in* [6].

Keywords: symmetric cone, Euclidean Jordan algebra, Wasserstein barycenter, regularized median optimization.
MSC2020: 90C25, 49J27, 17C20.

1. Introduction

Recently, Kum et al. [7] studied divergences on a symmetric cone Ω, and considered a median optimization problem. The motivation is from the problem of finding Wasserstein barycenter for Gaussian measures in the theory of optimal transport where the symmetric cone Ω is the convex cone of positive definite matrices $\mathbb{P}$. Indeed, for Gaussian measures μ and ν with mean 0 (without loss of generality) and covariance matrices A and B respectively, the Wasserstein distance d_W is expressed by

$$d_W(\mu,\nu) = d_W(A,B) = \sqrt{\mathrm{tr}(A+B) - 2\mathrm{tr}(A^{\frac{1}{2}}BA^{\frac{1}{2}})^{\frac{1}{2}}}. \quad (1)$$

Then the problem of finding Wasserstein barycenter of Gaussian measures ν_j with zero mean and with positive definite covariance matrices $A_j,\ j = 1, \cdots, n$ respectively, is formulated as the least squares problem minimizing the averaged sum of squared d_W:

$$\Omega(\omega; \mathbb{A}) := \underset{X\in\mathbb{P}}{\arg\min} \sum_{j=1}^{n} w_j d_W^2(X, A_j). \quad (2)$$

Here $\mathbb{A} = (A_1, \cdots, A_n)$, and $\omega = (w_1, \ldots, w_n)$ is a positive probability vector. Hence the problem is immediately converted to a problem of matrix analysis in the special symmetric cone $\mathbb{P}$. Then a natural question arises from a theoretical perspective: How about the general symmetric cone case Ω? That is, is it possible to extend the results in $\mathbb{P}$ into Ω? In this respect, the paper [7] can be regarded as an answer to the question.

[1] Department of Mathematics Education, Chungbuk National University, Cheongju 28644, Korea, e-mail: shkum@cbnu.ac.kr

[2] Corresponding author, Center for General Education, China Medical University, Taichung, Taiwan 40402, Republic of China, e-mail: yaojc@mail.cmu.edu.tw

On the other hand, very recently, Kum et al. [6] also introduced a gradient projection method for solving regularized Wasserstein barycenter problems in some probability measure spaces including Gaussian and q-Gaussian measures. Its mathematical analysis is based on the identification of Gaussian and q-Gaussian measures with the corresponding covariance (positive definite) matrices. This enabled them to adopt matrix analysis for dealing with the problem. Under this circumstances, the same question can be cast in the same view as [7]: What happens if a general symmetric cone Ω is considered instead of $\mathbb{P}$? Is it possible to generalize the results in $\mathbb{P}$ to Ω?

The present paper is a trial to give an answer to the question. It is the aim of this work. Thus our work stands on the same line as [7]. We are concerned with an extension of main results of [6] into the general symmetric cone Ω from the convex cone of positive definite matrices $\mathbb{P}$. To be more specific, two regularized median optimization problems are introduced and the existence and uniqueness of solutions are studied on Ω. Moreover the Lipschitz continuity of the gradient of objective functions of the regularized median optimizations are provided for a possible design of gradient-based methods of finding the unique minimizer. Based on some results of [7], we present a purely Jordan-algebraic framework in comparison with the matrix-analytic one in [6]. This may be thought of as a main contribution of our paper.

This work is organized as follows. In Section 2, we take a brief look at basic facts regarding Euclidean Jordan algebras and symmetric cones. In section 3, we give a description of the regularized median optimizations and show the existence and uniqueness of solutions of the optimization problems. In section 4, the Lipschitz continuity of the gradients of objective functions is provided.

2. **Euclidean Jordan algebras and symmetric cones**

As in [7], in this section, we briefly describe (following mostly [4]) some Jordan-algebraic concepts pertinent to our purpose. A Jordan algebra V over $\mathbb{R}$ is a commutative algebra satisfying $x^2(xy) = x(x^2y)$ for all $x, y \in V$. For $x \in V$, let $L(x)$ be the linear operator defined by $L(x)y = xy$, and let $P(x) = 2L(x)^2 - L(x^2)$. The map P is called the quadratic representation of V. An element $x \in V$ is said to be invertible if there exists an element y (denoted by $y = x^{-1}$) in the subalgebra generated by x and e (the Jordan identity) such that $xy = e$.

An element $c \in V$ is called an idempotent if $c^2 = c \neq 0$. We say that $c_1, \ldots, c_k$ is a complete system of orthogonal idempotents if $c_i^2 = c_i, c_ic_j = 0, i \neq j, c_1 + \cdots + c_k = e$. An idempotent is primitive if it is non-zero and cannot be written as the sum of two non-zero idempotents. A Jordan frame is a complete system of orthogonal primitive idempotents.

A finite-dimensional Jordan algebra V with an identity element e is said to be *Euclidean* if there exists an inner product $\langle \cdot, \cdot \rangle$ such that $\langle xy, z \rangle = \langle y, xz \rangle$ for all $x, y, z \in V$.

Theorem 2.1. (Spectral theorem, first version [4, Theorem III.1.1]) *Let V be a Euclidean Jordan algebra. Given $x \in V$, there exist real numbers $\lambda_1, \ldots, \lambda_k$ all distinct and a unique complete system of orthogonal idempotents $c_1, \ldots, c_k$ such that*

$$x = \sum_{i=1}^{k} \lambda_i c_i. \tag{3}$$

The numbers λ_i are called the eigenvalues and (3) *is called the spectral decomposition of x.*

Theorem 2.2. (Spectral theorem, second version [4, Theorem III.1.2]) *Any two Jordan frames in a Euclidean Jordan algebra V have the same number of elements (called the rank of V, denoted by* $\mathrm{rank}(V)$*). Given $x \in V$, there exists a Jordan frame $c_1, \dots, c_r$ and real numbers $\lambda_1, \dots, \lambda_r$ such that $x = \sum_{i=1}^r \lambda_i c_i$. The numbers λ_i (with their multiplicities) are uniquely determined by x.*

Definition 2.1. *Let V be a Euclidean Jordan algebra of* $\mathrm{rank}(V) = r$*. The spectral mapping $\lambda : V \to \mathbb{R}^r$ is defined by $\lambda(x) = (\lambda_1(x), \dots, \lambda_r(x))$, where $\lambda_i(x)$'s are eigenvalues of x (with multiplicities) as in Theorem* 2.2 *in non-increasing order $\lambda_{\max}(x) = \lambda_1(x) \geq \lambda_2(x) \geq \cdots \geq \lambda_r(x) = \lambda_{\min}(x)$. Furthermore,* $\det(x) = \prod_{i=1}^r \lambda_i(x)$ *and* $\mathrm{tr}(x) = \sum_{i=1}^r \lambda_i(x)$.

Let Q be the set of all square elements of V. It turns out that Q has non-empty interior $\Omega := \mathrm{int}(Q)$, and Ω is a symmetric cone, that is, the group $G(\Omega) = \{g \in \mathrm{GL}(V) \mid g(\Omega) = \Omega\}$ acts transitively on it and Ω is a self-dual cone with respect to the inner product $\langle \cdot, \cdot \rangle$ where $\mathrm{GL}(V)$ is the Lie group of the invertible linear operators on V (see [4]). Furthermore, for any $a \in \Omega$, $P(a) \in G(\Omega)$ and is positive definite so that its operator norm

$$\|P(a)\| = \max\{\lambda_i \lambda_j \mid \lambda_i, \lambda_j \text{ are eigenvalues of } a\} \tag{4}$$

because the eigenvalues of $P(a)$ are of the form $\lambda_i \lambda_j$ [12, Theorem 3.1].

Note that $\overline{\Omega} = \{x \in V \mid \lambda_i(x) \geq 0, i = 1, \dots, r\}$. For $x, y \in V$, we define

$$x \leq y \quad \text{if} \quad y - x \in \overline{\Omega}$$

and $x < y$ if $y - x \in \Omega$. Clearly $\overline{\Omega} = \{x \in V \mid x \geq 0\}$ and $\Omega = \{x \in V \mid x > 0\}$.

On the other hand, the symmetric cone Ω is a Riemannian manifold [4]. In this case, the unique geodesic curve joining a and b [10, Proposition 2.6] is

$$t \mapsto a \#_t b := P(a^{1/2})(P(a^{-1/2})b)^t$$

where $a^t = \sum_{j=1}^r \lambda_j(a)^t c_j$ for the spectral decomposition $a = \sum_{j=1}^r \lambda_j(a) c_j$ in Theorem 2.2. Moreover, the geometric mean $a \# b := a \#_{1/2} b$ is a unique geodesic middle between a and b.

Basically the trace is an inner product on V, and the Jordan algebra V endowed with the trace inner product $\langle x, y \rangle = \mathrm{tr}(xy)$ is still Euclidean [4]. Every Euclidean Jordan algebra admits a unique direct sum decompositions with irreducible (simple) Euclidean Jordan algebras. Since the trace of a product of Euclidean Jordan algebras is the sum of their trace functionals, from now on, we assume that V is a simple Euclidean Jordan algebra of rank r equipped with the trace inner product.

3. Regularized Medians

Let $\mathbb{A} = (a_1, \cdots, a_n) \in \Omega^n$, and $\omega = (w_1, \dots, w_n)$ be a positive probability vector. Then we first consider the following minimization problem:

$$\Omega(\omega; \mathbb{A}) := \operatorname*{arg\,min}_{x \in \Omega} \sum_{j=1}^n w_j \Phi_t(a_j, x) + \gamma F(x) \tag{5}$$

where

$$\Phi_t(a, b) = \mathrm{tr}\,((1-t)a + tb) - \mathrm{tr}\left(P(a^{\frac{1-t}{2t}})b\right)^t, \ 0 < t < 1, \text{ and}$$

$$F(x) = -\frac{r}{2}\ln(2\pi e) - \frac{1}{2}\ln(\det x).$$

Now let us briefly describe the problem setting (5). As mentioned in [7], the real valued function Φ_t involves the t-weighted arithmetic mean of a and b, and the *sandwiched quasi-relative entropy*

$$F_t(a,b) := \operatorname{tr}\left(P(a^{\frac{1-t}{2t}})b\right)^t,$$

which is well-known in the theory of quantum information; for positive semidefinite matrices A and B,

$$F_t(A,B) := \operatorname{tr}\left(A^{\frac{1-t}{2t}} B A^{\frac{1-t}{2t}}\right)^t, \quad t \in (0,1).$$

This is a parameterized version of the *fidelity* $F_{1/2}(A,B) = \operatorname{tr}\left(A^{\frac{1}{2}} B A^{\frac{1}{2}}\right)^{\frac{1}{2}}$. Fidelity and sandwiched quasi-relative entropies play an important role in quantum information theory and quantum computation, and it has deep connections with quantum entanglement, quantum chaos, and quantum phase transitions. See [5, 13, 14, 15]. Moreover, the real-valued function F on Ω is originated from the Boltzmann entropy of a Gaussian measure [6]. The definition of F is formally extended into Ω from $\mathbb{P}$. Also $\gamma > 0$ is a regularization parameter.

When the symmetric cone Ω is the particular case $\mathbb{P}$ with $t = 1/2$, (5) immediately reduces to the problem (3.1) in [6], the regularized Wasserstein barycenter problem for Gaussian measures. In this case, $\Phi_{1/2}(a,b)$ is nothing but the squared Wasserstein distance $d_W^2(a,b)$. However, $\Phi_{1/2}(a,b)^{\frac{1}{2}}$ may not be a distance on the general Ω as noted in [7] even though $\Phi_{1/2}(a,b)$ is a divergence (for definition, see [7]) on Ω. Nonetheless, without the regularized term $\gamma F(x)$, the problem (5) has a unique solution [7], which is called the *ω-weighted Φ-median of* $a_1, \ldots, a_m$. So we name the problem (5) as *a regularized median optimization problem.*

Now we state the first main result.

Theorem 3.1. *The minimization problem* (5) *has a unique solution and it satisfies the following nonlinear equation:*

$$tx - \frac{\gamma}{2}e = t\sum_{j=1}^{n} w_j \left(P(x^{\frac{1}{2}})a_j^{\frac{1-t}{t}}\right)^t. \tag{6}$$

Proof. We first show that the objection function f of (5)

$$f(x) = \sum_{j=1}^{n} w_j \Phi_t(a_j, x) + \gamma F(x) \tag{7}$$

is strictly convex. It suffices to verify that for each $0 < t < 1$ and $a \in \Omega$, the map $\varphi_{a,t} : \Omega \to \mathbb{R}$

$$\varphi_{a,t}(x) := \Phi_t(a,x) = \operatorname{tr}((1-t)a + tx) - \operatorname{tr}\left(P(a^{\frac{1-t}{2t}})x\right)^t$$

and F are strictly convex. Indeed, the strict convexity of $\varphi_{a,t}$ is explained in [7] using [3, Lemma 3.1 and Theorem 3.2]. In addition, that of F is directly from the well-known formula [9]

$$\nabla(-\ln\det)(x) = -x^{-1}, \quad \nabla^2(-\ln\det)(x) = P(x)^{-1} = P(x^{-1}). \tag{8}$$

Hence the hessian of F is positive definite, so F is strictly convex. Thus f is strictly convex so that $\nabla f(x) = 0$ at the unique solution if it exists. From [7, Proposition 4.1] we have

$$\nabla\varphi_{a,t}(x) = t\left(e - \left(a^{\frac{1-t}{t}} \#_{1-t} x^{-1}\right)\right). \tag{9}$$

Hence

$$\nabla f(x) = \sum_{j=1}^{n} w_j \nabla \varphi_{a_j,t}(x) - \frac{\gamma}{2}x^{-1} = t\sum_{j=1}^{n} w_j \left[e - \left(a_j^{\frac{1-t}{t}} \#_{1-t} x^{-1}\right)\right] - \frac{\gamma}{2}x^{-1}.$$

So

$$\nabla f(x) = 0 \iff te - \frac{\gamma}{2}x^{-1} = t\sum_{j=1}^{n} w_j \left(a_j^{\frac{1-t}{t}} \#_{1-t} x^{-1}\right). \tag{10}$$

Note that

$$a_j^{\frac{1-t}{t}} \#_{1-t} x^{-1} \quad = \quad x^{-1} \#_t a_j^{\frac{1-t}{t}} = P(x^{-1/2}) \left(P(x^{1/2}) a_j^{\frac{1-t}{t}}\right)^t.$$

Taking $P(x^{\frac{1}{2}})$ in both sides of (10) yields

$$\nabla f(x) = 0 \iff tx - \frac{\gamma}{2}e = t\sum_{j=1}^{n} w_j \left(P(x^{1/2}) a_j^{\frac{1-t}{t}}\right)^t. \tag{11}$$

Thus the existence of solution of (5) is equivalent to that of a fixed point of the mapping

$$H(x) = \sum_{i=1}^{m} w_j \left(P(x^{1/2}) a_j^{\frac{1-t}{t}}\right)^t + \frac{\gamma}{2t}e. \tag{12}$$

Now we show the existence. Let

$$\alpha := \min\{\lambda_{\min}(a_j) \mid j = 1, \ldots, n\}, \qquad \beta := \max\{\lambda_{\max}(a_j) \mid j = 1, \ldots, n\}$$

where $\lambda_{\min}(a_j)$ and $\lambda_{\max}(a_j)$ denote the minimum and maximum eigenvalue of a_j, respectively. So $\forall j,\ a_j \in [\alpha e, \beta e] := \{x \mid \alpha e \leq x \leq \beta e\}$. Consider the elementary equation for a positive real variable y

$$\alpha^{1-t} y^t - y = -\frac{\gamma}{2t}. \tag{13}$$

Then it is easily checked that the function $f(y) = \alpha^{1-t}y^t - y$ is decreasing on $[\alpha, +\infty)$ and $f(y) \to -\infty$ as $y \to +\infty$. Thus the equation (13) has a unique solution $\alpha_* \in [\alpha, +\infty)$. Similarly we can obtain a positive real $\beta_*(> \alpha_*)$ on $[\beta, +\infty)$ satisfying $\beta^{1-t}y^t - y = -\frac{\gamma}{2t}$. So we have

$$\alpha^{1-t}\alpha_*^t + \frac{\gamma}{2t} = \alpha_*, \quad \beta^{1-t}\beta_*^t + \frac{\gamma}{2t} = \beta_*. \tag{14}$$

Then the mapping H in (12) is a self-map on $[\alpha_* e, \beta_* e]$. To see this, let $x \in [\alpha_* e,\ \beta_* e]$. From $\alpha e \leq a_j \leq \beta e$,

$$\alpha^{\frac{1-t}{t}} e \leq a_j^{\frac{1-t}{t}} \leq \beta^{\frac{1-t}{t}} e$$

and hence

$$\alpha^{\frac{1-t}{t}} x = P(x^{1/2}) \left(\alpha^{\frac{1-t}{t}} e\right) \leq P(x^{1/2}) a_j^{\frac{1-t}{t}} \leq P(x^{1/2}) \left(\beta^{\frac{1-t}{t}} e\right) = \beta^{\frac{1-t}{t}} x.$$

Thus

$$\alpha^{\frac{1-t}{t}} \alpha_* e \leq \alpha^{\frac{1-t}{t}} x \leq P(x^{1/2}) a_j^{\frac{1-t}{t}} \leq \beta^{\frac{1-t}{t}} x \leq \beta^{\frac{1-t}{t}} \beta_* e.$$

By (14)

$$\alpha_* e = \alpha^{1-t}\alpha_*^t e + \frac{\gamma}{2t}e \leq \left(P(x^{1/2}) a_j^{\frac{1-t}{t}}\right)^t + \frac{\gamma}{2t}e \leq \beta^{1-t}\beta_*^t e + \frac{\gamma}{2t}e = \beta_* e.$$

Therefore

$$\alpha_* e \leq H(x) = \sum_{j=1}^{n} w_j \left(P(x^{1/2}) a_j^{\frac{1-t}{t}} \right)^t + \frac{\gamma}{2t} e \leq \beta_* e.$$

By Brouwer's fixed point theorem, there exists a point $x \in [\alpha_* e, \beta_* e]$ such that $x = H(x)$. This completes the proof. □

Remark 3.1. Theorem 3.1 is a generalization of [6, Theorem 3.1] in Ω. In fact, when $t = 1/2$ and $\Omega = \mathbb{P}$ the cone of positive definite matrices, Theorem 3.1 immediately reduces to [6, Theorem 3.1].

For the second minimization problem, the following coefficients [7] are necessary: For r the rank of V (replacing d in [7]), $C_0(q,r), C_1(q,r)$ are given by

$$C_1(q,r) = \frac{2}{2 + (r+2)(1-q)},$$

$$C_0(q,r) = \begin{cases} \dfrac{\Gamma\left(\frac{2-q}{1-q} + \frac{r}{2}\right)}{\Gamma\left(\frac{2-q}{1-q}\right)} \left(\frac{(1-q)C_1(q,r)}{2\pi}\right)^{\frac{r}{2}} & \text{if } 0 < q < 1, \\ \dfrac{\Gamma\left(\frac{1}{q-1}\right)}{\Gamma\left(\frac{1}{q-1} - \frac{r}{2}\right)} \left(\frac{(q-1)C_1(q,r)}{2\pi}\right)^{\frac{r}{2}} & \text{if } 1 < q < \frac{r+4}{r+2}. \end{cases}$$

We set the second optimization problem. For simplicity, only the standard case $t = 1/2$ is considered as follow:

$$\Omega(\omega; \mathbb{A}) := \underset{x \in \Omega}{\arg\min} \sum_{j=1}^{n} w_j \Phi_{\frac{1}{2}}(a_j, x) + \gamma F_q(x) \tag{15}$$

where

$$\Phi_{\frac{1}{2}}(a,b) = \frac{1}{2}\mathrm{tr}(a+b) - \mathrm{tr}\left(P(a^{\frac{1}{2}})b\right)^{\frac{1}{2}},$$

$$F_q(x) = -\frac{r}{2}C_1(q,r) + \left[1 - (1-q)\frac{r}{2}C_1(q,r)\right] \ln_q \frac{C_0(q,r)}{(\det x)^{\frac{1}{2}}} \quad \text{and}$$

$$\ln_q \frac{C_0(q,r)}{(\det x)^{\frac{1}{2}}} = \frac{1}{1-q}\left[C_0(q,r)^{1-q}(\det x)^{-\frac{1-q}{2}} - 1\right].$$

The functional F_q on Ω is deduced from the Tsallis entropy of a q-Gaussian measure [6, Lemma 2.2]. Under these circumstances, the minimization problem (15) can be written as

$$\min_{x \in \Omega} \frac{1}{2} g(x) \tag{16}$$

where

$$\begin{aligned} g(x) &= \sum_{i=1}^{n} w_i \,\mathrm{tr}\, a_i + \sum_{i=1}^{n} w_i \,\mathrm{tr}\left(x - 2\left(P(a_i^{\frac{1}{2}})x\right)^{\frac{1}{2}}\right) \\ &\quad + \gamma\Big[2 - (1-q)rC_1(q,r)\Big] \ln_q \frac{C_0(q,r)}{(\det x)^{\frac{1}{2}}} - \gamma r C_1(q,r) \\ &= f_1(x) + \gamma\Big[2 - (1-q)rC_1(q,r)\Big] \ln_q \frac{C_0(q,r)}{(\det x)^{\frac{1}{2}}} - \gamma r C_1(q,r), \end{aligned} \tag{17}$$

with $f_1(x) = \sum_{i=1}^n w_i \operatorname{tr} a_i + \sum_{i=1}^n w_i \operatorname{tr}\left(x - 2\left(P(a_i^{\frac{1}{2}})x\right)^{\frac{1}{2}}\right)$. Using explicit formula of $C_1(q,r)$ we get

$$\begin{aligned} 2-(1-q)rC_1(q,r) &= 2-(1-q)r\frac{2}{2+(r+2)(1-q)} \\ &= \frac{4(2-q)}{2+(r+2)(1-q)}. \end{aligned}$$

Substituting these expressions into (17) we obtain

$$\begin{aligned} g(x) = f_1(x) &+ \frac{4\gamma(2-q)C_0(q,r)^{1-q}}{(2+(r+2)(1-q))(1-q)}(\det x)^{-\frac{1-q}{2}} \\ &- \frac{4(2-q)}{(1-q)(2+(r+2)(1-q))} - \gamma r C_1(q,r). \end{aligned} \tag{18}$$

Now we state the second main result.

Theorem 3.2. *Suppose that $\alpha e \le a_i \le \beta e$ $(0 < \alpha \le \beta)$ for all $i = 1, \dots, n$. The regularized median optimization problem* (15) *has a unique solution x for all $\gamma \ge 0$ if either $0 < q \le 1$ or $1 < q \le 1 + \frac{2\alpha^2}{r\beta^2}$ and for γ sufficiently small if $1 + \frac{2\alpha^2}{r\beta^2} < q < \frac{r+4}{r+2}$. The solution x solves the following nonlinear equation*

$$x - \gamma m(q,r)(\det x)^{\frac{q-1}{2}} e = \sum_{i=1}^n w_i \Big(P(x^{\frac{1}{2}})a_i\Big)^{\frac{1}{2}}, \tag{19}$$

where $m(q,r)$ is defined by

$$m(q,r) := \frac{2(2-q)C_0(q,r)^{1-q}}{2+(r+2)(1-q)}.$$

Proposition 3.1. *The nonlinear equation* (19) *has a solution.*

Proof. The argument in the proof of [6, Proposition 4.2] is available. We will show that

$$\psi(x) := \sum_{i=1}^n w_i \Big(P(x^{\frac{1}{2}})a_i\Big)^{1/2} + \gamma m(q,r)(\det x)^{\frac{q-1}{2}} e$$

has a fixed point. Since $\alpha e \le a_i \le \beta e$, for $\alpha_* e \le x \le \beta_* e$ (with α_*, β_* chosen later as in [6]), we have

$$\alpha x = P(x^{1/2})(\alpha e) \le P(x^{1/2})a_i \le P(x^{1/2})(\beta e) = \beta x, \text{ and}$$
$$\sqrt{\alpha} x^{\frac{1}{2}} \le \Big(P(x^{1/2})a_i\Big)^{\frac{1}{2}} \le \sqrt{\beta} x^{\frac{1}{2}}.$$

Hence

$$\sqrt{\alpha}\sqrt{\alpha_*} e \le \sqrt{\alpha} x^{\frac{1}{2}} \le \Big(P(x^{1/2})a_i\Big)^{\frac{1}{2}} \le \sqrt{\beta} x^{\frac{1}{2}} \le \sqrt{\beta}\sqrt{\beta_*} e, \text{ and}$$
$$\sqrt{\alpha}\sqrt{\alpha_*} e \le \sum_{i=1}^n w_i \Big(P(x^{\frac{1}{2}})a_i\Big)^{1/2} \le \sqrt{\beta}\sqrt{\beta_*} e.$$

Thus

$$\begin{aligned} \sqrt{\alpha}\sqrt{\alpha_*} e + \gamma m(q,r)(\det x)^{\frac{q-1}{2}} e &\le \sum_{i=1}^n w_i \Big(P(x^{\frac{1}{2}})a_i\Big)^{1/2} + \gamma m(q,d)(\det x)^{\frac{q-1}{2}} e \\ &\le \sqrt{\beta}\sqrt{\beta_*} e + \gamma m(q,r)(\det x)^{\frac{q-1}{2}} e. \end{aligned} \tag{20}$$

Case 1. $1 < q < \frac{r+4}{r+2}$. From (20) and the inequalities

$$\alpha_*{}^r = \det(\alpha_* e) \le \det x \le \det(\beta_* e) = \beta_*{}^r$$

we obtain

$$\sqrt{\alpha}\sqrt{\alpha_*}e + \gamma m(q,r)\alpha_*{}^{\frac{r(q-1)}{2}} e \le \psi(x) = \sum_{i=1}^{n} w_i \Big(P(x^{\frac{1}{2}})a_i\Big)^{1/2} + \gamma m(q,d)(\det x)^{\frac{q-1}{2}} e$$
$$\le \sqrt{\beta}\sqrt{\beta_*}e + \gamma m(q,r)\beta_*{}^{\frac{r(q-1)}{2}} e.$$

For the existence of a fixed point of ψ, it suffices to check that ψ is a self-map on $[\alpha_* e, \beta_* e]$ so that we can use Brouwer's fixed point theorem again. Thus we have only to show that there exist $0 < \alpha_* < \beta_*$ such that

$$\alpha_* = \sqrt{\alpha_0}\sqrt{\alpha_*} + \gamma m(q,r)\alpha_*{}^{\frac{r(q-1)}{2}} \text{ and } \beta_* = \sqrt{\beta_0}\sqrt{\beta_*} + \gamma m(q,r)\beta_*{}^{\frac{r(q-1)}{2}}.$$

But this is straightforward from the argument of [6]. So we omit it.

Case 2. $0 < q < 1$. The same argument of [6] is available, too. Therefore, (19) has a solution. This completes the proof of the proposition. □

Proposition 3.2. *Suppose that $\alpha e \le a_i$, $x \le \beta e$ for all $i = 1, \ldots, n$. The functional g given in* (18) *is strictly convex for all $\gamma \ge 0$ when one of the following condition holds*

(1) $0 < q < 1$,
(2) $1 < q \le 1 + \frac{2\alpha^2}{r\beta^2}$.

In addition, if $1 + \frac{2\alpha^2}{r\beta^2} < q < \frac{r+4}{r+2}$, then g is strictly convex for $0 \le \gamma < \gamma_0$ where

$$\gamma_0 = \frac{1}{2}\frac{\alpha^2}{\beta^3}\frac{1}{\frac{(q-1)d}{2\alpha^2} - \frac{1}{\beta^2}}\frac{1}{m(q,d)}\frac{1}{\beta^{d(q-1)/2}}.$$

Proof. We follow the argument in [6].
Case 1. $1 < q < \frac{r+4}{r+2}$.

Let $k(x) := \frac{4\gamma(2-q)C_0(q,r)^{1-q}}{(2+(r+2)(1-q))(1-q)}(\det x)^{\frac{q-1}{2}}$. Let $h(x) := (\det x)^{\frac{q-1}{2}}$. By (8) and the chain rule, we get

$$\nabla \det(x) = \det(x)x^{-1}. \tag{21}$$

Using the definition of $m(q,r)$, we have

$$\nabla k(x) = -\gamma m(q,r)(\det x)^{\frac{q-1}{2}} x^{-1} = -\gamma m(q,r)h(x)x^{-1}. \tag{22}$$

By the Leibniz rule and (8), we obtain

$$\begin{aligned}
\nabla^2 k(x)(h) &= D(\nabla k)(x)(h) \\
&= -\gamma m(q,r)[Dh(x)(h)x^{-1} + h(x)(-P(x^{-1}))(h)] \\
&= -\gamma m(q,r)[\langle \nabla h(x), h\rangle x^{-1} - h(x)P(x^{-1})(h)] \\
&= -\gamma m(q,r)\left[\left\langle \frac{q-1}{2}(\det x)^{\frac{q-1}{2}} x^{-1}, h\right\rangle x^{-1} - (\det x)^{\frac{q-1}{2}} P(x^{-1})h\right] \\
&= -\gamma m(q,r)(\det x)^{\frac{q-1}{2}}\left[\left\langle \frac{q-1}{2} x^{-1}, h\right\rangle x^{-1} - P(x^{-1})h\right].
\end{aligned}$$

Thus

$$\langle \nabla^2 k(x)(h), h\rangle = -\gamma m(q,r)(\det x)^{\frac{q-1}{2}}\left[\frac{q-1}{2}\left\langle x^{-1},h\right\rangle^2 - \langle P(x^{-1})h,h\rangle\right]$$
$$= \gamma m(q,r)(\det x)^{\frac{q-1}{2}}\left[\langle P(x^{-1})h,h\rangle - \frac{q-1}{2}\langle x^{-1},h\rangle^2\right].$$

Furthermore, according to [7, Corollary 6.3], for $\alpha e \leq a_i, x \leq \beta e$, we obtain from (17)

$$\langle \nabla^2 f_1(x)(h), h\rangle \geq \frac{1}{2}\frac{\alpha^2}{\beta^3}\|h\|^2.$$

Thus we get

$$\begin{aligned}
\langle \nabla^2 g(x)(h), h\rangle &= \langle \nabla^2 f_1(x)(h), h\rangle + \langle \nabla^2 k(x)(h), h\rangle \\
&\geq \gamma m(q,r)(\det x)^{\frac{q-1}{2}}\left[\langle P(x^{-1})h,h\rangle - \frac{q-1}{2}\left\langle x^{-1},h\right\rangle^2\right] + \frac{1}{2}\frac{\alpha^2}{\beta^3}\|h\|^2 \\
&\geq \gamma m(q,r)(\det x)^{\frac{q-1}{2}}\left[\frac{1}{\beta^2}\|h\|^2 - \frac{q-1}{2}\|x^{-1}\|^2\|h\|^2\right] + \frac{1}{2}\frac{\alpha^2}{\beta^3}\|h\|^2 \\
&= \left\{\gamma m(q,r)(\det x)^{\frac{q-1}{2}}\left[\frac{1}{\beta^2} - \frac{q-1}{2}\|x^{-1}\|^2\right] + \frac{1}{2}\frac{\alpha^2}{\beta^3}\right\}\|h\|^2 \\
&\geq \left\{\gamma m(q,r)(\det x)^{\frac{q-1}{2}}\left[\frac{1}{\beta^2} - \frac{q-1}{2}\frac{r}{\alpha^2}\right] + \frac{1}{2}\frac{\alpha^2}{\beta^3}\right\}\|h\|^2
\end{aligned}$$

From this estimate, we deduce the following cases

(i) If

$$1 < q \leq 1 + \frac{2\alpha^2}{r\beta^2},$$

thus $\frac{1}{\beta^2} - \frac{q-1}{2}\frac{r}{\alpha^2} \geq 0$, which implies that the Hessian of g is positive for all γ. Note that the above condition is fulfilled if α and β satisfy $\beta^2 \leq \frac{r+2}{r}\alpha^2$. In fact, we have

$$q < 1 + \frac{2}{r+2} \leq 1 + \frac{2\alpha^2}{r\beta^2}.$$

(ii) If

$$1 + \frac{2\alpha^2}{r\beta^2} < q < \frac{r+4}{r+2}.$$

then for

$$\gamma < \frac{1}{2}\frac{\alpha^2}{\beta^3}\frac{1}{\frac{(q-1)r}{2\alpha^2} - \frac{1}{\beta^2}}\frac{1}{m(q,r)}\frac{1}{\beta^{r(q-1)/2}}$$

the Hessian of g is positive since

$$\begin{aligned}
\gamma &< \frac{1}{2}\frac{\alpha^2}{\beta^3}\frac{1}{\frac{(q-1)r}{2\alpha^2} - \frac{1}{\beta^2}}\frac{1}{m(q,r)}\frac{1}{\beta^{r(q-1)/2}} \\
&\leq \frac{1}{2}\frac{\alpha^2}{\beta^3}\frac{1}{\frac{(q-1)r}{2\alpha^2} - \frac{1}{\beta^2}}\frac{1}{m(q,r)}\frac{1}{(\det x)^{(q-1)/2}}
\end{aligned}$$

Case 2. $0 < q < 1$. Similarly, we obtain

$$\begin{aligned}\langle \nabla^2 k(x)(h), h\rangle &= \gamma m(q,r)(\det x)^{\frac{q-1}{2}} \left[\frac{1-q}{2}\left\langle x^{-1}, h\right\rangle^2 + \langle P(x^{-1})h, h\rangle\right] \\ &\geq \gamma m(q,r)(\det x)^{\frac{q-1}{2}} \frac{1}{\beta^2}\|h\|^2.\end{aligned}$$

Hence the Hessian of g is always positive definite in this case. □

Remark 3.2. Proposition 3.2 may be regarded as an extension of [6, Proposition 4.3] in Ω with a slight change of the coefficient $\frac{\alpha^{1/2}}{\beta^{3/2}}$ by $\frac{\alpha^2}{\beta^3}$. This minor difference is due to the fact that we adopt the Jordan-algebraic technique in [7] instead of the matrix-analytic one in [2]. Actually, the matrix-analytic method in [2] is only available to the special case of $\Omega = \mathbb{P}$ the cone of positive definite matrices.

We are now ready to prove Theorem 3.2.

Proof of Theorem 3.2. Suppose that the hypothesis of the statement of Theorem 3.2 is satisfied, that is, either (i) $0 < q \leq 1$ or (ii) $1 < q \leq 1 + \frac{2\alpha^2}{r\beta^2}$ or (iii) $1 + \frac{2\alpha^2}{r\beta^2} < q < \frac{r+4}{r+2}$. Suppose that γ is sufficiently small in the last case; in the other cases γ can be arbitrarily positive. Recall that $g(x)$ is given in (18)

$$g(x) = f_1(x) + k(x) - \frac{4(2-q)}{(1-q)(2+(r+2)(1-q))} - \gamma r C_1(q,r).$$

By Proposition 3.2, $x \mapsto g(x)$ is strictly convex. Now we compute the derivative of $g(x)$. Obviously

$$\nabla g(x) = \nabla f_1(x) + \nabla k(x). \tag{23}$$

From (9) and (17) we get

$$\nabla f_1(x) = e - \sum_{i=1}^{n} w_i(a_i \# x^{-1}). \tag{24}$$

By (22), we have

$$\nabla k(x) = -\gamma m(q,r)(\det x)^{\frac{q-1}{2}} x^{-1}$$

Substituting these computations into (23) we obtain

$$\nabla g(x) = \Big(e - \sum_{i=1}^{n} w_i(a_i \# x^{-1})\Big) - \gamma m(q,r)(\det x)^{\frac{q-1}{2}} x^{-1}. \tag{25}$$

Thus $\nabla g(x) = 0$ if and only if

$$\begin{aligned}e - \gamma m(q,r)(\det x)^{\frac{q-1}{2}} x^{-1} &= \sum_{i=1}^{n} w_i(a_i \# x^{-1}) \\ &= \sum_{i=1}^{n} w_i(x^{-1} \# a_i) = \sum_{i=1}^{n} w_i P(x^{-1/2})\Big(P(x^{1/2})a_i\Big)^{\frac{1}{2}}.\end{aligned}$$

Taking $P(x^{\frac{1}{2}})$ in both sides of the above equation yields

$$x - \gamma m(q,r)(\det x)^{\frac{q-1}{2}} e = \sum_{i=1}^{n} w_i \Big(P(x^{1/2})a_i\Big)^{\frac{1}{2}},$$

which is precisely equation (19). By Proposition 3.1, it has a solution. This, together with the strict convexity of g, guarantees the existence and uniqueness of a minimizer of g. So the proof of the theorem is completed. □

Remark 3.3. Theorem 3.2 is a purely Euclidean Jordan algebraic version of [6, Theorem 4.1] in Ω. Indeed, [6, Theorem 4.1] is the particular case of Theorem 3.2 when $\Omega = \mathbb{P}$ the cone of positive definite matrices with the minor change mentioned in the previous Remark 3.6.

4. **Lipschitz continuity of the gradient maps**

We establish the Lipschitz continuity of the gradients of the objective functions of (5) and (15). For the first regularized median optimization problem (5), the standard case $t = 1/2$ is considered for simplicity. In this case, we may regard $\tilde{f}(x) = 2f(x)$ as the objective function. Indeed, from (7) we have

$$\begin{aligned}\tilde{f}(x) &= \sum_{i=1}^{n} w_i \operatorname{tr} a_i + \sum_{i=1}^{n} w_i \operatorname{tr}\Big(x - 2\big(P(a_i^{\frac{1}{2}})x\big)^{\frac{1}{2}}\Big) - \gamma \ln\det(x) - \gamma r \ln(2\pi e)\\ &= f_1(x) + \gamma f_2(x), \end{aligned} \tag{26}$$

where

$$\begin{aligned} f_1(x) &= \sum_{i=1}^{n} w_i \operatorname{tr} a_i + \sum_{i=1}^{n} w_i \operatorname{tr}\Big(x - 2\big(P(a_i^{\frac{1}{2}})x\big)^{\frac{1}{2}}\Big),\\ f_2(x) &= -\ln\det(x) - r\ln(2\pi e). \end{aligned}$$

Before going to main results, we need the following:

Proposition 4.1. *Let us consider two functions G and H where $0 < \alpha < \beta$ and*

$$G : [\alpha e, \beta e] \to [\sqrt{\alpha} e, \sqrt{\beta} e],\ G(x) = x^{\frac{1}{2}}, \quad H : [\alpha e, \beta e] \to \Big[\frac{1}{\beta} e, \frac{1}{\alpha} e\Big],\ H(x) = x^{-1}.$$

Then for $x, y \in [\alpha e, \beta e]$, we have

$$\|G(x) - G(y)\| \le \frac{1}{2\sqrt{\alpha}}\|x - y\|, \quad \|H(x) - H(y)\| \le \frac{1}{\alpha^2}\|x - y\|.$$

Proof. In fact, by Sun and Sun [12, Theorem 3.2] G, H are continuously differentiable on the Löwner interval $[\alpha e, \beta e]$ because the corresponding real valued functions $g(t) = \sqrt{t}$, $h(t) = 1/t$ are continuously differentiable on the interval $[\alpha, \beta]$. In addition, $[\alpha e, \beta e]$ is a compact convex set so that G, H are Lipschitz continuous on $[\alpha e, \beta e]$ with the Lipschitz constants $\frac{1}{2\sqrt{\alpha}}$ and $\frac{1}{\alpha^2}$ by [12, Theorem 3.2] and the mean value theorem for operators [11, Proposition 2, p.176]. □

Remark 4.1. Note that for the corresponding matrix case in [8], Theorem X.3.8 of [1] was used to derive the same Lipschitz constants of G and H, which can be applied only to matrix case. In this regard, Proposition 4.1 is a general version of [1, Theorem X.3.8].

Observe that for $a, x \in [\alpha e, \beta e]$,

$$\alpha^2 e \le \alpha a = P(a^{\frac{1}{2}})(\alpha e) \le P(a^{\frac{1}{2}})x \le P(a^{\frac{1}{2}})(\beta e) = \beta a \le \beta^2 e.$$

Using Proposition 4.1 and the above observation, we directly obtain the first Lipschitz continuity:

Theorem 4.1. *Suppose that $a_i \in [\alpha e, \beta e]$ for all $i = 1, \dots, n$. Then for $\alpha e \le x \ne y \le \beta e$ we have*

$$\frac{\|\nabla \tilde{f}(x) - \nabla \tilde{f}(y)\|}{\|x-y\|} \le \frac{\beta^2}{2\alpha^3} + \frac{\gamma}{\alpha^2}.$$

Proof. According to (24) and (4) we have

$$\begin{aligned}
\|\nabla f_1(x) - \nabla f_1(y)\| &= \Big\| \sum_{i=1}^n w_i[(a_i \# x^{-1}) - (a_i \# y^{-1})] \Big\| \\
&= \Big\| \sum_{i=1}^n w_i P(a_i^{\frac{1}{2}}) \big[\big(P(a_i^{-\frac{1}{2}})x^{-1}\big)^{\frac{1}{2}} - \big(P(a_i^{-\frac{1}{2}})y^{-1}\big)^{\frac{1}{2}} \big] \Big\| \\
&\le \sum_{i=1}^n w_i \Big\| P(a_i^{\frac{1}{2}}) \Big\| \Big\| \big(P(a_i^{-\frac{1}{2}})x^{-1}\big)^{\frac{1}{2}} - \big(P(a_i^{-\frac{1}{2}})y^{-1}\big)^{\frac{1}{2}} \Big\| \\
&\le \beta \sum_{i=1}^n w_i \Big\| \big(P(a_i^{\frac{1}{2}})x\big)^{-\frac{1}{2}} - \big(P(a_i^{\frac{1}{2}})y\big)^{-\frac{1}{2}} \Big\| \\
&\le \frac{\beta}{\alpha^2} \sum_{i=1}^n w_i \Big\| \big(P(a_i^{\frac{1}{2}})x\big)^{\frac{1}{2}} - \big(P(a_i^{\frac{1}{2}})y\big)^{\frac{1}{2}} \Big\| \\
&\le \frac{\beta}{2\alpha^3} \sum_{i=1}^n w_i \Big\| P(a_i^{\frac{1}{2}})x - P(a_i^{\frac{1}{2}})y \Big\| \\
&\le \frac{\beta}{2\alpha^3} \sum_{i=1}^n w_i \Big\| P(a_i^{\frac{1}{2}}) \Big\| \|x-y\| \\
&\le \frac{\beta^2}{2\alpha^3} \|x-y\|.
\end{aligned}$$

Therefore we get

$$\begin{aligned}
\frac{\|\nabla \tilde{f}(x) - \nabla \tilde{f}(y)\|}{\|x-y\|} &\le \frac{\|\nabla f_1(x) - \nabla f_1(y)\| + \gamma \|\nabla f_2(x) - \nabla f_2(y)\|}{\|x-y\|} \\
&\le \frac{\beta^2}{2\alpha^3} + \frac{\gamma \|x^{-1} - y^{-1}\|}{\|x-y\|} \le \frac{\beta^2}{2\alpha^3} + \frac{\gamma}{\alpha^2}.
\end{aligned}$$

□

Now we are in a position to provide the Lipschitz continuity of ∇g concerned with the second regularized median optimization (15). By (25), we know

$$\nabla g(x) = \Big(e - \sum_{i=1}^n w_i (a_i \# x^{-1}) \Big) - \gamma m(q,r) (\det x)^{\frac{q-1}{2}} x^{-1} =: \nabla f_1(x) - \gamma m(q,r) \tilde{h}(x),$$

where $\nabla f_1(x) = \Big(e - \sum_{i=1}^n w_i (a_i \# x^{-1}) \Big)$ as in (24) and $\tilde{h}(x) = (\det x)^{\frac{q-1}{2}} x^{-1} = h(x) x^{-1}$. Since the method of proof of the second result below is exactly the same as that of [6, Theorem 5.3], only the statement is made without proof. Readers may refer to [6].

Theorem 4.2. *Suppose that* $a_i \in [\alpha e, \beta e]$ *for all* $i = 1, \dots, n$. *Then for* $\alpha e \leq x \neq y \leq \beta e$, *we have*

$$\frac{\|\nabla g(x) - \nabla g(y)\|}{\|x-y\|} \leq \begin{cases} \frac{\beta^2}{2\alpha^3} + \frac{\gamma}{\alpha^2} + \frac{\gamma m(q,r)}{\alpha^2} \cdot \beta^{\frac{q-1}{2}r}\left(1 + \frac{q-1}{2}r\right), & \text{if } 1 < q < \frac{r+4}{r+2}, \\ \\ \frac{\beta^2}{2\alpha^3} + \frac{\gamma}{\alpha^2} + \gamma m(q,r)\alpha^{-2+\frac{q-1}{2}r}\left(1 + \frac{1-q}{2}r\right), & \text{if } 0 < q < 1. \end{cases}$$

Remark 4.2. Theorem 4.1 and Theorem 4.2 are generalizations of [6, Theorems 5.2 and 5.3] in Ω, respectively.

5. **Conclusions**

In this paper we studied two regularized median optimization problems on a general symmetric cone Ω. Basically the existence and uniqueness of solutions are treated. Moreover the Lipschitz continuity of the gradient of objective functions of the regularized median optimizations are provided. All of these results belong to a development of general Jordan-algebraic frameworks beyond the usual matrix-analytic one. This is a main contribution of our work from a theoretical perspective. However, we did not consider a numerical algorithm to find the unique minimizer and implement it numerically. Besides, we presently do not know of various applications of our results. So we leave them for further study.

Acknowledgements

The work of the corresponding author was supported by MOST (111-2115-M-039-001-MY2).

REFERENCES

[1] R. Bhatia, *Matrix Analysis*, Graduate Texts in Mathematics, 169, Springer-Verlag, New York, 1997.

[2] R. Bhatia, T. Jain and Y. Lim, *On the Bures–Wasserstein distance between positive definite matrices*, Expo. Math. **37** (2018), no 2, 165-–191.

[3] Y. Chang and J. Chen, *Convexity of symmetric cone trace functions in euclidean jordan algebras*, J. Nonlinear Convex Anal. **14** (2013), no. 1, 53–61.

[4] J. Faraut and A. Koranyi, *Analysis on Symmetric Cones*, Oxford mathematical monographs, Clarendon Press, 1994.

[5] R. Frank and E. Lieb, *Monotonicity of a relative rényi entropy*, J. Math. Phys. **54** (2013), no. 12, 122201.

[6] S. Kum, M. Duong, Y. Lim and S. Yun, *A GPM-based algorithm for solving regularized Wasserstein barycenter problems in some spaces of probability measures*, J. Comput. Appl. Math. **416** (2022), 114588.

[7] S. Kum, Y. Lim and S. Yun, *Divergences on Symmetric Cones and Medians*, Taiwan J. Math. **26** (2022), no. 4, 867–886.

[8] S. Kum and S. Yun, *Gradient projection methods for the n-coupling problem*, J. Korean Math. Soc. **56** (2019), no. 4, 1001-–1016.

[9] H. Lee and Y. Lim, *Metric and spectral geometric means on symmetric cones*, Kyungpook Math. J. **47** (2007), 133–150.

[10] Y. Lim, *Geometric means on symmetric cones*, Arch. Math. **75** (2000), no. 1, 39–45.

[11] D. Luenberger, *Optimization by Vector Space Methods*, John Wiley & Sons, Inc., 1969.

[12] D. Sun and J. Sun, *Löwner's operator and spectral functions in Euclidean Jordan algebras*, Math. Oper. Res. **33** (2008), no. 2, 421–445.

[13] A. Uhlmann, *Density operators as an arena for differential geometry*, Rep. Math. Phys. **33** (1993), no. 1-2, 253–263.

[14] A. Uhlmann, *Transition probability (fidelity) and its relatives*, Found. Phys. **41** (2011), no. 3, 288–298.

[15] M. Wilde, A. Winter and D. Yang, *Strong converse for the classical capacity of entanglement-breaking and hadamard channels via a sandwiched rényi relative entropy*, Commun. Math. Phys. **331** (2014), no. 2, 593–622.