A GPM-based algorithm for solving regularized Wasserstein barycenter problems in some spaces of probability measures

Kum, S.; Duong, Manh Hong; Lim, Y.; Yun, S.

DOI:
10.1016/j.cam.2022.114588

Journal of Computational and Applied Mathematics 416 (2022) 114588

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

A GPM-based algorithm for solving regularized Wasserstein barycenter problems in some spaces of probability measures

S. Kum [a], M.H. Duong [b,*], Y. Lim [c], S. Yun [d]

[a] Department of Mathematics Education, Chungbuk National University, Cheongju 28644, Republic of Korea
[b] School of Mathematics, University of Birmingham, B15 2TT, UK
[c] Department of Mathematics, Sungkyunkwan University, Suwon 440-746, Republic of Korea
[d] Department of Mathematics Education, Sungkyunkwan University, Seoul 03063, Republic of Korea

ARTICLE INFO

Article history:
Received 15 March 2021
Received in revised form 30 June 2022

Keywords:
Wasserstein barycenter
q-Gaussian measures
Gradient projection method
Optimization

ABSTRACT

In this paper, we focus on the analysis of the regularized Wasserstein barycenter problem. We provide uniqueness and a characterization of the barycenter for two important classes of probability measures, each regularized by a particular entropy functional: (i) Gaussian distributions and (ii) q-Gaussian distributions. We propose an algorithm based on gradient projection method (GPM) in the space of matrices in order to compute these regularized barycenters. Finally, we numerically show the influence of parameters and stability of the algorithm under small perturbation of data and compare the gradient projection method with Riemannian gradient method.

1. Introduction

1.1. Regularization of barycenters in the Wasserstein space

In this paper we are interested in *the regularization of barycenters in the Wasserstein space*, which is a minimization problem of the form

$$\min_{\mu\in\mathcal{P}_2(\mathbb{R}^d)}\sum_{i=1}^{n}\frac{1}{2}\lambda_i W_2^2(\mu,\mu_i)+\gamma F(\mu), \tag{1.1}$$

where $\mathcal{P}_2(\mathbb{R}^d)$ is the Wasserstein space of probability measures on $\mathbb{R}^d$ with finite second moments; $\{\mu_i\}_{i=1}^n$ are n given probability measures in $\mathcal{P}_2(\mathbb{R}^d)$; W_2 is the L^2-Wasserstein distance between two probability measures in $\mathcal{P}_2(\mathbb{R}^d)$ (cf. Section 2), and $F:\mathcal{P}_2(\mathbb{R}^d)\to\mathbb{R}$ is an entropy functional. Finally $\gamma\geq 0$ is a given regularization parameter; $\lambda_1,\ldots,\lambda_n$ are given non-negative numbers (weights) satisfying $\sum_{i=1}^n\lambda_i=1$.

1.2. Literature review

Problem (1.1) for $\gamma=0$ has been studied intensively in the literature. It was first studied by Knott and Smith [1] for Gaussian measures. In [2], Agueh and Carlier studied the general case proving, among other things, the existence and

* Corresponding author.
E-mail addresses: shkum@chungbuk.ac.kr (S. Kum), h.duong@bham.ac.uk (M.H. Duong), ylim@skku.edu (Y. Lim), yswmathedu@skku.edu (S. Yun).

https://doi.org/10.1016/j.cam.2022.114588

uniqueness of a minimizer provided that one of μ_i's vanishes on small sets (i.e. sets whose Hausdorff dimension is at most $d-1$). Examples of such measures include those that are absolutely continuous with respect to the Lebesgue measure. The minimizer is called the barycenter of the measures μ_i with weights λ_i extending a classical characterization of the Euclidean barycenter. The article [2] has sparked off many research activities from both theoretical and computational aspects over the last years. Wasserstein barycenters in different settings, such as over compact Riemannian manifolds [3] and over discrete data [4] have been investigated. In the compact Riemannian setting, the condition to vanish on small sets ensuring uniqueness is replaced by absolute continuity with respect to the volume measure [3]. However, in the discrete setting, the uniqueness and absolute continuity of the barycenter is lost [4]. Connections between Wasserstein barycenters and optimal transports have been explored [5,6]. Several computational methods for the computation of the barycenter have been developed [7–10]. Recently Wasserstein barycenters has found many applications in statistics, image processing and machine learning [11–13]. We refer the reader to the mentioned papers and references therein for a more detailed account of the topic.

The case $\gamma > 0$ has been studied in recent papers [14,15] where the existence, uniqueness and stability of a minimizer, which is called the regularized barycenter, has been established. In particular, [14] shows that if the regularizing function is a proper and lower semicontinuous function (for the Wasserstein distance) and is strictly convex on its domain, then there exists a unique regularized barycenter even in the case of discrete measures. In addition, the regularization parameter γ was proved to provide smooth barycenters especially when the input probability measures are irregular which is useful for data analysis [16,17]. In addition, the regularized barycenter problem also resembles the discretization formulation of Wasserstein gradient flows for dissipative evolution equations [18–20] and the fractional heat equation [21] at a given time step where $\{\mu_i\}$ represent discretized solutions at the previous steps and γ is proportional to the time-step parameter.

Gaussian measures play an important role in the study of Wasserstein barycenter problem since in this case a useful characterization of the barycenter exists [2,22] which gives rise to efficient computational algorithms such as the fixed point approach [8] and the gradient projection method [9]. Our aim in this paper is to seek for a large class of probability measures so that the regularized barycenter can be explicitly characterized and computed similarly to the case of Gaussian measures. It is worth mentioning that many papers in the literature study a related problem of entropic regularization of optimal transports where the Wasserstein distance is regularized by an entropic term. The problem of finding a closed form solution for such problems in the case of Gaussian distributions has increasingly attracted interest in the community of computational optimal transport and machine learning [23,24]. The problem that we study in this paper is different from these papers since the entropy term is added outside of the Wasserstein distance.

We will study the regularization problem (1.1) for two important classes of probability measures, namely Gaussian and q-Gaussian measures, where the entropy functional is the negative Boltzmann entropy and the Tsallis entropy, respectively. The two classes are both special cases of a more general class of probability measures, namely φ-exponential measures. To state our main results, we now briefly recall the definition of φ-exponential measures; more details will be given in the Appendix.

1.3. φ-exponential distributions

Let φ be an increasing, positive, continuous function on $(0,\infty)$, the φ-logarithmic function is defined by [25]

$$\ln_\varphi(t) := \int_1^t \frac{1}{\varphi(s)}\,ds, \tag{1.2}$$

which is increasing, concave and C^1 on $(0,\infty)$. Let l_φ and L_φ be respectively the infimum and the supremum of $\ln_\varphi$, that is

$$l_\varphi := \inf_{t>0} \ln_\varphi(t) = \lim_{t\downarrow 0} \ln_\varphi(t) \in [-\infty, 0),$$
$$L_\varphi := \sup_{t>0} \ln_\varphi(t) = \lim_{t\uparrow \infty} \ln_\varphi(t) \in (0, +\infty).$$

The function $\ln_\varphi$ has the inverse function, which is called the φ-exponential function, and is defined on (l_φ, L_φ). This inverse function can be extended to the whole $\mathbb{R}$ as

$$\exp_\varphi(s) := \begin{cases} 0 & \text{for } s \le l_\varphi, \\ \ln_\varphi^{-1}(s) & \text{for } s \in (l_\varphi, L_\varphi), \\ \infty & \text{for } s \ge L_\varphi, \end{cases} \tag{1.3}$$

which is C^1 on (l_φ, L_φ).

Let $\mathbb{S}(d,\mathbb{R})_+$ be the set of symmetric positive definite matrices of order d. Let $v \in \mathbb{R}^d$ be a given vector and $V \in \mathbb{S}(d,\mathbb{R})_+$ be a given symmetric positive definite matrix. The φ-exponential measure with mean v and covariance matrix V, denoted by $G_\varphi(v,V)$, is the probability measure on $\mathbb{R}^d$ with Lebesgue density

$$g_\varphi(v,V)(x) := \exp_\varphi(\lambda_\varphi - c_\varphi |x-v|_V^2)\Big(\det(V)\Big)^{-\frac{1}{2}}, \tag{1.4}$$

where $|x|_V^2 := \langle x, V^{-1}x\rangle$, λ_φ and c_φ are normalization constants. Two important examples of φ-exponential measures include Gaussian measures and q-Gaussian measures corresponding to $\varphi(s) = s$ and $\varphi(s) = s^q$ respectively. The φ-exponential measures play an important role in statistical physics, information geometry and in the analysis of nonlinear diffusion equations [26–29]. More information about φ-exponential measures will be reviewed in the Appendix.

1.4. Main results of the paper

As already mentioned, in this paper we study the regularized problem (1.1) for Gaussian measures and q-Gaussian measures, where the entropy functional is the (negative) Boltzmann entropy functional and the Tsallis entropy functional respectively. Main results of the present paper are explicit characterizations of the minimizer of (1.1) and properties of the objective functions that can be summarized as follows.

Theorem 1.1. *Suppose that for each $i = 1, \dots, n$, μ_i is a q-Gaussian measure (Gaussian measure when $q = 1$) with mean zero and covariance matrix $A_i \in \mathbb{S}(d, \mathbb{R})_+$. Then the regularized barycenter problem (1.1), with F being the Tsallis entropy functional (the negative Boltzmann entropy functional when $q = 1$), has a unique minimizer, which is also a q-Gaussian measure with mean zero and covariance matrix X satisfying*

$$X - \gamma m(q,d)(\det X)^{\frac{q-1}{2}} I = \sum_{i=1}^{n} \lambda_i \left(X^{\frac{1}{2}} A_i X^{\frac{1}{2}}\right)^{\frac{1}{2}},$$

where $m(q, d)$ is a constant depending on q and d (see Theorem 4.1 for its explicit formula, in particular $m = 1$ when $q = 1$).

Theorem 1.2. *Suppose that $\{\mu_i\}$ are all Gaussian measures or all q-Gaussian measures with mean zero. Then the gradient of the objective function in the minimization problem (1.1) is Lipschitz continuous, where the Lipschitz constant in each case can be found explicitly (see Theorem 5.2 and Theorem 5.3 respectively).*

Theorem 1.1 summarizes Proposition 2.3, Theorem 3.1 (for Gaussian measures), Theorem 4.1 (for q-Gaussian measures). Theorem 1.2 summarizes Theorem 5.2 (for Gaussian measures) and Theorem 5.3 (for q-Gaussian measures).

The key to the analysis of the present paper is that the spaces of φ-exponential measures and Gaussian measures are isometric in the sense of Wasserstein geometry [28,29], that is

$$W_2(G_\varphi(v, V), G_\varphi(u, U)) = W_2(\mathcal{N}(v, V), \mathcal{N}(u, U)),$$

where $\mathcal{N}(v, V)$ denotes a Gaussian measure with mean v and covariance matrix V. Therefore, since the Wassertein distance between Gaussian measures can be computed explicitly, the objective functional in (1.1) can also be computed explicitly in terms of the covariance matrices and (1.1) becomes a minimization problem over the space of symmetric positive definite matrices. We then prove the strict convexity of the objective function and the existence of solutions to the optimality equation using matrix analysis tools as in [22]. Theorems 3.1 and 4.1 establish the existence and uniqueness of a minimizer and provide an explicit characterization of the minimizer in terms of nonlinear matrix equations for the covariance matrix generalizing the characterization of the Wasserstein barycenter for Gaussian measures in [2,22] to the regularized Wasserstein barycenter for Gaussian measures and q-Gaussian measures. Theorems 5.2 and 5.3 prove the Lipschitz continuity of the gradient of the objective function providing an explicit upper bound for the Lipschitz constant generalizing the results of [9] for the barycenter for Gaussian measures to our setting. We also perform numerical experiments to show the affect of the parameter q and a stability property of the algorithm under small perturbation of the data, and compare our proposed method with the existing state-of-art Riemannian gradient method [30–32], cf. Section 6.

1.5. Organization of the paper

The rest of the paper is organized as follows. In Section 2 we review relevant knowledge that will be used in subsequent sections on the Wasserstein distance and the Wasserstein geometry of Gaussian and φ-exponential distributions. Then we study the regularization of barycenters for Gaussian measures in Section 3 and extend these results to q-Gaussian measures in Section 4. In Section 5 we describe a gradient projection method for the computation of the minimizer and prove that the gradient function is Lipschitz continuous. In Section 6, we numerically show effect of parameters to the minimizer and stability of the algorithm under small perturbation of data. Comparison of our proposed algorithm, gradient projection method, with Riemannian gradient method [30–32] is also presented. We provide a summary of the paper and possible directions for future work in Section 7. Finally, we recall some detailed knowledge about φ-exponential measures in the Appendix.

2. Wasserstein distance, Gaussian measures and φ-exponential measures

In this section, we summarize relevant knowledge that will be used in subsequent sections on the Wasserstein distance and the Wasserstein geometry of Gaussian and φ-exponential distributions.

2.1. Wasserstein distance

We recall that $\mathcal{P}_2(\mathbb{R}^d)$ is the space of probability measures μ on $\mathbb{R}^d$ with finite second moment, namely

$$\int_{\mathbb{R}^d} |x|^2 \mu(dx) < \infty.$$

Let μ and ν be two probability measures belonging to $\mathcal{P}_2(\mathbb{R}^d)$. The L^2-Wasserstein distance, $W_2(\mu, \nu)$, between μ and ν is defined via

$$W_2^2(\mu, \nu) := \inf_{\gamma \in \Gamma(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \, \gamma(dx, dy), \tag{2.1}$$

where $\Gamma(\mu, \nu)$ denotes the set of transport plans between μ and ν, i.e., the set of all probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ having μ and ν as the first and the second marginals respectively. More precisely,

$$\Gamma(\mu, \nu) := \{\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) : \gamma(A \times \mathbb{R}^d) = \mu(A) \text{ and } \gamma(\mathbb{R}^d \times A) = \nu(A)\},$$

for all Borel measurable sets $A \subset \mathbb{R}^d$. It has been proved that, under rather general conditions (e.g., when μ and ν are absolutely continuous with respect to the Lesbegue measure), an optimal transport plan in (2.1) uniquely exists and is of the form $\gamma = [\mathrm{id} \times \nabla\psi]_{\#}\mu$ for some convex function ψ where # denotes the push forward [33,34].

The Wasserstein distance is an instance of a Monge–Kantorovich optimal transportation cost functional and plays a key role in many branches of mathematics such as optimal transportation, partial differential equations, geometric analysis and has been found many applications in other fields such as economics, statistical physics and recently in machine learning. We refer the reader to the celebrated monograph [35] for a great exposition of the topic.

We now consider two important classes of probability measures, namely Gaussian measures and φ-exponential measures, for which there is an explicit expression for the Wasserstein distance between two members of the same class. Although Gaussian measures are special cases of φ-exponential measures, we consider them separately since many proofs for the former are much simpler than those for the latter.

2.2. Wasserstein distance of Gaussian measures

Given any $X \in \mathbb{S}(d, \mathbb{R})_+$, we define a symmetric positive definite matrix $X^{1/2}$ such that $X^{1/2}X^{1/2} = X$. Throughout the paper, we denote by I the identity matrix of order d. The Wasserstein distance between two Gaussian measures is well-known [36], see also e.g., [28]:

$$W_2(\mathcal{N}(u, U), \mathcal{N}(v, V))^2 = |u - v|^2 + \mathrm{tr}U + \mathrm{tr}V - 2\mathrm{tr}\sqrt{V^{\frac{1}{2}} U V^{\frac{1}{2}}}. \tag{2.2}$$

Furthermore, $[\mathrm{id} \times \nabla\mathcal{T}]_{\#}\mathcal{N}(u, U)$ is the optimal plan between them, where

$$\mathcal{T}(x) = \frac{1}{2}\langle x - u, T(x - u)\rangle + \langle x, v\rangle, \quad T = V^{\frac{1}{2}} \left(V^{\frac{1}{2}} U V^{\frac{1}{2}}\right)^{-\frac{1}{2}} V^{\frac{1}{2}}. \tag{2.3}$$

2.3. The entropy of Gaussian measures

The (negative) Boltzmann entropy of a probability measure $\mu = \mu(x)dx$ on $\mathbb{R}^d$ is defined by

$$F(\mu) := \int_{\mathbb{R}^d} \mu(x) \log \mu(x) \, dx. \tag{2.4}$$

Using Gaussian integral, the (negative) Boltzmann entropy of a Gaussian measure can be computed explicitly [37, Theorem 9.4.1]:

$$F(\mathcal{N}(u, U)) = -\frac{d}{2} \ln(2\pi e) - \frac{1}{2} \ln \det(U). \tag{2.5}$$

2.4. q-Gaussian measures and Wassertein distance

In the Appendix we review an important class $\mathcal{G}_\varphi$ of probability measures, namely φ-exponential measures where φ is a given function satisfying certain conditions. The class contain two special cases:

(i) $\varphi = s$, $\mathcal{G}_\varphi$ reduces to the class of Gaussian measures.

(ii) In the case $\varphi = s^q$, $\mathcal{G}_\varphi$ becomes the class of all q-Gaussian measures

$$\mathcal{G}_q = \left\{G_q(v, V) \middle| (v, V) \in \mathbb{R}^d \times \mathbb{S}(d, \mathbb{R})_+\right\}$$

where

$$G_q(v,V)=C_0(q,d)(\det V)^{-\frac{1}{2}}\exp_q\Big(-\frac{1}{2}C_1(q,d)\langle x-v,V^{-1}(x-v)\rangle\Big)\mathcal{L}^d,$$

and $C_0(q,d)$, $C_1(q,d)$ are given by

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

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

Note that $C_1(1,d)=1$ and $C_0(q,d)\to(2\pi)^{-d/2}$ as $q\to1$, which follows from Stirling's formula. Thus Gaussian measures are special cases of q-Gaussian measures.

The following result explains why q-Gaussian measures and φ-exponential measures are of special interest. It will play a key role in the analysis of this paper.

Proposition 2.1. *The following statements hold [28,29]*

1. *For any $q\in(0,1)\cup\left(1,\frac{d+4}{d+2}\right)$, the space of q-Gaussian measures is convex and isometric to the space of Gaussian measures with respect to the Wasserstein metric.*
2. *For any $\varphi\in\mathcal{O}(2/(d+2))$ with $d\geq2$, the space $\mathcal{G}_\varphi$ is convex and isometric to the space of Gaussian measures with respect to the Wasserstein metric.*
3. *Let $G_\varphi(v,V)$ and $G_\varphi(u,U)$ be two φ-exponential distributions. Then $[\mathrm{id}\times\nabla\mathcal{T}]_\# G_\varphi(u,U)$, where $\mathcal{T}$ is defined in (2.3), is the optimal plan in the definition of $W_2^2(G_q(v,V),G_\varphi(u,U))$.*
4. *We have*

$$\begin{aligned}W_2(G_\varphi(\mu,U),G_\varphi(\nu,V))^2&=W_2(G_q(\mu,U),G_q(\nu,V))^2\\&=W_2(\mathcal{N}(\mu,U),\mathcal{N}(\nu,V))^2\\&=|\mu-\nu|^2+\mathrm{tr}U+\mathrm{tr}V-2\mathrm{tr}\sqrt{V^{\frac{1}{2}}UV^{\frac{1}{2}}}.\end{aligned}\tag{2.6}$$

2.5. The Tsallis entropy of a q-Gaussian measure

The Tsallis entropy of a probability measure $\mu=\mu(x)dx$ on $\mathbb{R}^d$ is defined by

$$F_q(\mu):=\int_{\mathbb{R}^d}\mu(x)\ln_q\mu(x)\,dx=\frac{1}{1-q}\int_{\mathbb{R}^d}[\mu(x)^{1-q}-1]\mu(x)\,dx.\tag{2.7}$$

The Tsallis entropy of a q-Gaussian can also be computed explicitly using the property (A.1) and similar computations as in the Gaussian case.

Lemma 2.2. *It holds that [38]*

$$F_q(G_q(\mu,U))=-\frac{d}{2}C_1(q,d)+\left[1-(1-q)\frac{d}{2}C_1(q,d)\right]\ln_q\frac{C_0(q,d)}{(\det U)^{\frac{1}{2}}}.$$

The first result of the present paper is the following proposition.

Proposition 2.3. *Suppose that $\mu_i=G_q(0,A_i)$. Then the regularized barycenter problem (1.1) has a unique minimizer, which is also a q-Gaussian measure with mean 0. This statement holds also for $q=1$ and in this case, the minimizer is a Gaussian measure with mean 0. Similarly, when $\{\mu_i\}$ are all φ-exponential distributions with mean 0, then the unregularized barycenter problem has a unique minimizer which is also a φ-exponential distribution with mean 0.*

Proof. Since each of $\{\mu_i\}_{i=1}^n$ is a q-Gaussian measure with mean zero, then there exists a unique minimizer $\mu_*\in\mathcal{P}_2(\mathbb{R}^d)$, which is absolutely continuous with respect to the d-dimensional Lebesgue measure [14]. Let v and V be the mean and covariance matrix of μ_*. Let $G_q(v,V)$ be the q-Gaussian measure with the same mean v and covariance matrix V. Next we will show that

$$\mu_*=G_q(v,V)\quad\text{and}\quad v=0\quad(\text{thus }\mu_*=G_q(0,V)).$$

Since $G_q(v, V)$ minimizes the Tsallis entropy F_q among all probability measures μ which are absolutely continuous with the d-dimensional Lebesgue measure having mean v and covariance matrix V (see for instance [28]), we have

$$F_q(\mu_*) \geq F_q(G_q(v, V)). \tag{2.8}$$

We recall the following equivalent, Monge and Kantorovich duality, characterizations of the Wassertein distance between two probability measures $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ (see [39, Theorem 5.10])

$$\begin{aligned} W_2(\mu, \nu)^2 &= \inf_{T_\# \mu = \nu} \int_{\mathbb{R}^d} |x - T(x)|^2 \, d\mu(x) \\ &= \sup_{\phi \in L^1(\mu)} \left\{ \int_{\mathbb{R}^d} \phi(y)^c \, d\nu(y) - \int_{\mathbb{R}^d} \phi(x) \, d\mu(x) \right\}, \end{aligned}$$

where $\phi^c(y) = \inf_{x \in \mathbb{R}^d} \{\phi(x) + |x - y|^2\}$. In addition, the optimal transport map T^* and the optimal Kantorovich potential ϕ^* in the above problems satisfy

$$x - T^*(x) = \frac{1}{2} \nabla \phi^*(x).$$

Let T_i and ϕ_i, $i = 1, \ldots, n$ be the optimal transport map and the optimal Kantorovich potential for $W_2(\mu_i, G_q(v, V))$, that is

$$\begin{aligned} W_2(\mu_i, G_q(v, V))^2 &= \int_{\mathbb{R}^d} |x - T_i(x)|^2 \, d\mu_i(x) \\ &= \int_{\mathbb{R}^d} \phi_i(y)^c \, dG_q(v, V)(y) - \int_{\mathbb{R}^d} \phi(x) \, d\mu_i(x). \end{aligned}$$

According to [28, Theorem A], T_i is given by $T_i = \nabla \mathcal{T}_i(x)$ where

$$\mathcal{T}_i(x) = \frac{1}{2} \langle x, \bar{T}_i x \rangle + \langle x, v \rangle, \quad \bar{T}_i = V^{1/2} \left(V^{1/2} A_i V^{1/2} \right)^{-1/2} V^{1/2}.$$

It follows that

$$\phi_i(x) = |x|^2 - 2\mathcal{T}_i(x) = |x|^2 - \langle x, \bar{T}_i x \rangle - 2\langle x, v \rangle.$$

Therefore,

$$\phi_i(y)^c = \phi_i(\bar{x}) + \frac{1}{4} |\nabla \phi_i(\bar{x})|^2 \quad \text{where} \quad \nabla \phi_i(\bar{x}) + 2(\bar{x} - y) = 0.$$

It follows that (using the symmetry of $\bar{T}_i$)

$$y = y(\bar{x}) = \bar{x} + \frac{1}{2} \nabla \phi_i(\bar{x}) = 2\bar{x} - \bar{T}_i \bar{x} - 2v.$$

Therefore, the Jacobian matrix J_i when changing the variable from y to $\bar{x}$ is constant, which is given by

$$J_i = D_{\bar{x}} y = 2I - \bar{T}_i.$$

We have

$$\begin{aligned} \int_{\mathbb{R}^d} \phi_i^c(y) d\mu_*(y) &= \int_{\mathbb{R}^d} \left(\phi_i(\bar{x}) + \frac{1}{4} |\nabla \phi_i(\bar{x})|^2 \right) d\mu_*(y) \\ &\overset{(*)}{=} |J_i| \int_{\mathbb{R}^d} \left(\phi_i(\bar{x}) + \frac{1}{4} |\nabla \phi_i(\bar{x})|^2 \right) d\mu_*(\bar{x}) \\ &\overset{(**)}{=} |J_i| \int_{\mathbb{R}^d} \left(\phi_i(\bar{x}) + \frac{1}{4} |\nabla \phi_i(\bar{x})|^2 \right) dG_q(v, V)(\bar{x}) \\ &= \int_{\mathbb{R}^d} \phi_i^c(y) dG_q(v, V)(y), \end{aligned}$$

where $(**)$ follows from $(*)$ since $(*)$ depends only on the mean and covariance of μ_* (because $\phi_i(\bar{x}) + \frac{1}{4} |\nabla \phi_i(\bar{x})|^2$ is a quadratic function of $\bar{x}$), which is the same as $G_q(v, V)$. Therefore

$$\begin{aligned} W_2(\mu_i, \mu_*)^2 &\geq \int_{\mathbb{R}^d} \phi_i(y)^c \, d\mu_*(y) - \int_{\mathbb{R}^d} \phi_i(x) \, d\mu_i(x) \\ &= \int_{\mathbb{R}^d} \phi_i(y)^c dG_q(v, V)(y) - \int_{\mathbb{R}^d} \phi_i(x) \, d\mu_i(x) \\ &= W_2(\mu_i, G_q(v, V))^2. \end{aligned} \tag{2.9}$$

From (2.8) and (2.9) we get

$$\sum_{i=1}^{n} \lambda_i W_2(\mu_i, \mu_*)^2 + F_q(\mu_*) \geq \sum_{i=1}^{n} \lambda_i W_2(\mu_i, G_q(v, V))^2 + F_q(G_q(v, V)).$$

By the uniqueness of minimizers, we deduce that $\mu_* = G_q(v, V)$. Moreover, the facts that

$$F_q(G_q(v, V)) = F_q(G_q(0, V)), \quad W_2(\mu_i, G_q(0, V)) \leq W_2(\mu_i, G_q(v, V))$$

ensure $v = 0$. Note that this proof also holds true for $q = 1$ where q-Gaussian measures and the Tsallis entropy are respectively replaced by Gaussian measures and the Boltzmann entropy. Similarly, using the third part of Proposition 2.1, we can show that the minimizer of the unregularized barycenter is again a φ-exponential distribution if all the μ_i are φ-exponential distributions. This completes the proof of this proposition. □

3. Regularization of barycenters for Gaussian measures

In this section we study the following regularization of barycenters in the space of Gaussian measures

$$\min_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \sum_{i=1}^{n} \frac{1}{2} \lambda_i W_2^2(\mu, \mu_i) + \gamma F(\mu), \tag{3.1}$$

where $\mu_i \sim \mathcal{N}(0, A_i)$ $(i = 1, \ldots, n)$, F is the (negative) Boltzmann entropy functional of a probability measure defined in (2.4) and $\gamma > 0$ is a regularization parameter.

According to Proposition 2.3, we only need to seek for the minimizer μ among Gaussian measures with mean zero, that is $\mu \sim \mathcal{N}(0, X)$ for some covariance matrix X. We note that we consider here Gaussian measures with zero mean just for simplicity, see Remark 3.3 for further discussion on this assumption. The main results of the paper can be easily extended to the case of non-zero mean. From now on, we equip $\mathbb{S}(d, \mathbb{R})_+$ with the Frobenius inner product $\langle X, Y\rangle := \mathrm{tr}(X^T Y)$. The Frobenius norm is defined by $\|X\|_F = \left(\mathrm{tr}(X^T X)\right)^{\frac{1}{2}}$. For $X, Y \in \mathbb{S}(d, \mathbb{R})$, we write $X \leq Y$ if $Y - X$ is positive semidefinite, and $X < Y$ if $Y - X$ is positive definite. Note that $X \leq Y$ if and only if $\langle x, Xx\rangle \leq \langle x, Yx\rangle$ for all $x \in \mathbb{R}^d$. We denote $[X, Y]$ by the Löwner order interval $[X, Y] := \{Z : X \leq Z \leq Y\}$.

Theorem 3.1. *Assume that $\{\mu_i\}$ are Gaussian distributions with mean zero and covariance matrix A_i, $\mu_i = \mathcal{N}(0, A_i)$ for $i = 1, \ldots, n$. The regularization of barycenters problem (1.1) has a unique solution $\mu = \mathcal{N}(0, X)$ where the covariance matrix X solves the following nonlinear matrix equation*

$$X - \gamma I = \sum_{i=1}^{n} \lambda_i (X^{\frac{1}{2}} A_i X^{1/2})^{\frac{1}{2}}. \tag{3.2}$$

In particular, in the scalar case ($d = 1$), we obtain

$$X = \frac{\left[\sum_{i=1}^{n} \lambda_i A_i^{\frac{1}{2}} + \left(\left(\sum_{i=1}^{n} \lambda_i A_i^{\frac{1}{2}}\right)^2 + 4\gamma\right)^{\frac{1}{2}}\right]^2}{4}. \tag{3.3}$$

Before proving this theorem, we show the existence of solutions to Eq. (3.2).

Lemma 3.2. *Eq. (3.2) has a positive definite solution.*

Proof. Pick $0 < \alpha_0 < \beta_0$ so that $\alpha_0 I \leq A_i \leq \beta_0 I$ for all $i = 1, \ldots, n$. Set

$$\alpha_* := \left(\frac{\sqrt{\alpha_0} + \sqrt{\alpha_0 + 4\gamma}}{2}\right)^2, \qquad \beta_* := \left(\frac{\sqrt{\beta_0} + \sqrt{\beta_0 + 4\gamma}}{2}\right)^2.$$

Then for matrices X satisfying $\alpha_* I \leq X \leq \beta_* I$ we have,

$$\alpha_0 X \leq X^{1/2} A_i X^{1/2} \leq \beta_0 I, \qquad i = 1, \ldots, n$$

and hence

$$\sqrt{\alpha_0}\sqrt{\alpha_*} I \leq \sqrt{\alpha_0} X^{1/2} \leq (X^{1/2} A_i X^{1/2})^{1/2} \leq \sqrt{\beta_0} X^{1/2} \leq \sqrt{\beta_0}\sqrt{\beta_*} I.$$

By definition of α_* and β_*,

$$\begin{aligned}\alpha_* I &= \sqrt{\alpha_0}\sqrt{\alpha_*} I + \gamma I \leq \sum_{i=1}^{n} \lambda_i (X^{1/2} A_i X^{1/2})^{1/2} + \gamma I \\ &\leq \sqrt{\beta_0}\sqrt{\beta_*} I + \gamma I = \beta_* I\end{aligned}$$

for every $X \in [\alpha_* I, \beta_* I] := \{Z : \alpha_* I \le Z \le \beta_* I\}$. This shows that the map

$$f(X) := \sum_{i=1}^{n} \lambda_i (X^{1/2} A_i X^{1/2})^{1/2} + \gamma I$$

is a continuous self map on the Löwner order interval $[\alpha_* I, \beta_* I]$. By Brouwer's fixed point theorem, it has a fixed point. □

We are now ready to prove Theorem 3.1

Proof of Theorem 3.1. According to (2.2) and (2.5) we have

$$\begin{aligned} W_2^2(\mu_i, \mu) &= \mathrm{tr}X + \mathrm{tr}A_i - 2\mathrm{tr}\Big(A_i^{\frac{1}{2}} X A_i^{\frac{1}{2}}\Big)^{\frac{1}{2}}, \\ F(\mu) &= -\frac{d}{2}\ln(2\pi e) - \frac{1}{2}\ln(\det X). \end{aligned}$$

Thus we can write (1.1) as a minimization problem in the space of symmetric positive definite matrices

$$\min_{X \in \mathbb{S}(d,\mathbb{R})_+} \frac{1}{2} f(X) \tag{3.4}$$

where

$$\begin{aligned} f(X) &:= \sum_{i=1}^{n} \lambda_i \mathrm{tr}A_i + \sum_{i=1}^{n} \lambda_i \mathrm{tr}\Big(X - 2\big(A_i^{\frac{1}{2}} X A_i^{\frac{1}{2}}\big)^{\frac{1}{2}}\Big) - \gamma \ln\det(X) - \gamma d \ln(2\pi e) \\ &:= f_1(X) + \gamma f_2(X), \end{aligned} \tag{3.5}$$

where

$$\begin{aligned} f_1(X) &= \sum_{i=1}^{n} \lambda_i \mathrm{tr}A_i + \sum_{i=1}^{n} \lambda_i \mathrm{tr}\Big(X - 2\big(A_i^{\frac{1}{2}} X A_i^{\frac{1}{2}}\big)^{\frac{1}{2}}\Big), \\ f_2(X) &= -\ln\det(X) - d\ln(2\pi e). \end{aligned}$$

It has been proved [22] that

(i) $X \mapsto f_1(X)$ is strictly convex,
(ii) $Df_1(X)(Y) = \mathrm{tr}\Big(I - \sum_{i=1}^{n} \lambda_i (A_i \sharp X^{-1})\Big) Y$,

where $A \sharp B$ denotes the geometric mean between A and B defined by

$$A \sharp B = A^{1/2} (A^{-1/2} B A^{-1/2})^{1/2} A^{1/2}, \tag{3.6}$$

which is symmetric in A and B. According to [40, Proof of Theorem 8, Chapter 10] $X \mapsto -\ln\det(X)$ is strictly convex. Using Jacobi's formula for the derivative of the determinant and the chain rule, we get

$$Df_2(X)(Y) = -\frac{d}{dt}\ln\det(X + \varepsilon Y)\Big|_{t=0} = -\frac{1}{\det X} \cdot \det X \cdot \mathrm{tr}(X^{-1}Y) = -\mathrm{tr}(X^{-1}Y).$$

It follows that $X \mapsto f(X)$ is strictly convex. Furthermore, we have

$$Df(X)(Y) = \mathrm{tr}\Big(I - \gamma X^{-1} - \sum_{i=1}^{n} \lambda_i (A_i \sharp X^{-1})\Big) Y.$$

From this we deduce that

$$\nabla f(X) = I - \gamma X^{-1} - \sum_{i=1}^{n} \lambda_i (A_i \sharp X^{-1}),$$

where the gradient is with respect to the Frobenius inner product. Hence $\nabla f(X) = 0$ if and only if

$$I - \gamma X^{-1} = \sum_{i=1}^{n} \lambda_i (A_i \# X^{-1}).$$

Using the definition (3.6) of the geometric mean, the above equation can be written as

$$X-\gamma I=\sum_{i=1}^n \lambda_i (X^{\frac{1}{2}}A_iX^{1/2})^{\frac{1}{2}},$$

which is Eq. (3.2). By Lemma 3.2 this equation has a positive definite solution. This together with the strict convexity of f imply that f has a unique minimizer which is a Gaussian measure $\mathcal{N}(0,X)$ where X solves (3.2). In the one dimensional case this equation reads

$$X-\gamma=\sqrt{X}\sum_{i=1}^n \lambda_i\sqrt{a_i},$$

which results in

$$X=\frac{\left[\sum_{i=1}^n \lambda_i a_i^{\frac{1}{2}}+\left(\left(\sum_{i=1}^n \lambda_i a_i^{\frac{1}{2}}\right)^2+4\gamma\right)^{\frac{1}{2}}\right]^2}{4}.$$

This completes the proof of the theorem. □

Remark 3.3 (*The Case of Non-Zero Mean Distributions*)**.** Assume that $\{\mu_i\}$ are Gaussian distributions with means $\{m_i\}$ and covariance matrices $\{A_i\}$, that is $\mu_i\sim\mathcal{N}(m_i,A_i)$. Using the following formulas of the Wasserstein distances

$$W_2^2(\mu_i,\mu)=\|m-m_i\|^2+\mathrm{tr}X+\mathrm{tr}A_i-2\mathrm{tr}\left(A_i^{\frac{1}{2}}XA_i^{\frac{1}{2}}\right)^{\frac{1}{2}},$$

and the formula of the entropy functional (2.5) (noting that the entropy of a normal distribution is independent of its mean), we deduce that the minimizer $\mu\sim\mathcal{N}(m,X)$ where the mean m is given by

$$m=\sum_{i=1}^n \lambda_i m_i,$$

and the covariance matrix X satisfies the nonlinear matrix Eq. (3.2). The above statement about the mean is also true for the case of q-Gaussian measures and φ-exponential measures in the subsequent sections.

4. Regularization of barycenters for *q*-Gaussian measures

In this section we study the following regularization of barycenters in the space of q-Gaussian measures

$$\min_{\mu\in\mathcal{P}_2(\mathbb{R}^d)}\sum_{i=1}^n \frac{1}{2}\lambda_i W_2^2(\mu,\mu_i)+\gamma F_q(\mu), \tag{4.1}$$

where $\mu_i=G_q(0,A_i)$ $(i=1,\dots,n)$, F_q is the Tsallis entropy for a probability measure $\mu=\mu(x)dx$ on $\mathbb{R}^d$ defined by

$$F_q(\mu):=\int_{\mathbb{R}^d}\mu(x)\log_q\mu(x)\,dx.$$

According to Proposition 2.3, we only need to seek for the minimizer μ among q-Gaussian measures with mean zero, that is $\mu=G_q(0,X)$ for some covariance matrix X.

Theorem 4.1. *Assume that $\mu_i=G_q(0,A_i)$. Suppose that $\alpha I\le A_i\le\beta I$ for all $i=1,\dots,n$. The regularization of barycenters problem (4.1) has a unique solution $\mu=G_q(0,X)$ for all $\gamma\ge0$ if either $0<q\le1$ or $1<q\le 1+\frac{2\alpha^2}{d\beta^2}$ and for γ sufficiently small if $1+\frac{2\alpha^2}{d\beta^2}<q<\frac{d+4}{d+2}$. The covariance matrix X solves the following nonlinear matrix equation*

$$X-\gamma m(q,d)(\det X)^{\frac{q-1}{2}}I=\sum_{i=1}^n\lambda_i\left(X^{\frac{1}{2}}A_iX^{\frac{1}{2}}\right)^{\frac{1}{2}}, \tag{4.2}$$

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

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

The following proposition shows that Eq. (4.2) possesses a positive definite solution.

Proposition 4.2. *Eq. (4.2) has a positive definite solution.*

Proof. Similarly as the proof of Lemma 3.2 we will also apply Brouwer's fixed point theorem. We will show that

$$\psi(X) := \sum_{i=1}^{n} \lambda_i (X^{1/2} A_i X^{1/2})^{1/2} + \gamma m(q,d)(\det X)^{\frac{q-1}{2}} I$$

has a fixed point which is a positive definite matrix. Due to the appearance of the second term on the left-hand side of (4.2) the proof of this proposition is significantly involved than that of Lemma 3.2. Suppose that $\alpha_0 I \le A_i \le \beta_0 I$ for all $i = 1, \dots, n$. Then similarly as in the proof of Lemma 3.2, for $\alpha_* I \le X \le \beta_* I$ (with α_*, β_* chosen later), we have

$$\sqrt{\alpha_0}\sqrt{\alpha_*} I \le \sqrt{\alpha_0} X^{1/2} \le (X^{1/2} A_i X^{1/2})^{1/2} \le \sqrt{\beta_0} X^{1/2} \le \sqrt{\beta_0}\sqrt{\beta_*} I, \quad i = 1, \dots, n,$$

so that

$$\sqrt{\alpha_0}\sqrt{\alpha_*} I \le (X^{1/2} A_i X^{1/2})^{1/2} \le \sqrt{\beta_0}\sqrt{\beta_*} I.$$

Multiplying this inequality with λ_i then adding them together, noting that $\sum \lambda_i = 1$, we obtain

$$\sqrt{\alpha_0}\sqrt{\alpha_*} I \le \sum_{i=1}^{n} \lambda_i (X^{1/2} A_i X^{1/2})^{1/2} \le \sqrt{\beta_0}\sqrt{\beta_*} I,$$

from which it follows that

$$\begin{aligned}\sqrt{\alpha_0}\sqrt{\alpha_*} I + \gamma m(q,d)(\det X)^{\frac{q-1}{2}} I &\le \sum_{i=1}^{n} \lambda_i (X^{1/2} A_i X^{1/2})^{1/2} + \gamma m(q,d)(\det X)^{\frac{q-1}{2}} I \\ &\le \sqrt{\beta_0}\sqrt{\beta_*} I + \gamma m(q,d)(\det X)^{\frac{q-1}{2}} I. \end{aligned} \tag{4.3}$$

To continue we consider two cases.

Case 1: $1 < q < \frac{d+4}{d+2}$. It follows from (4.3) that

$$\begin{aligned} &\sqrt{\alpha_0}\sqrt{\alpha_*} I + \gamma m(q,d)\alpha_*^{\frac{d(q-1)}{2}} I \le \sqrt{\alpha_0}\sqrt{\alpha_*} I + \gamma m(q,d)(\det X)^{\frac{q-1}{2}} I \\ &\le \gamma m(q,d)(\det X)^{\frac{q-1}{2}} I + \sum_{i=1}^{n} \lambda_i (X^{1/2} A_i X^{1/2})^{1/2} \\ &\le \sqrt{\beta_0}\sqrt{\beta_*} I + \gamma m(q,d)(\det X)^{\frac{q-1}{2}} I \le \sqrt{\beta_0}\sqrt{\beta_*} I + \gamma m(q,d)\beta_*^{\frac{d(q-1)}{2}} I. \end{aligned} \tag{4.4}$$

Since $1 < q < \frac{d+4}{d+2}$, we have $0 < (q-1)d < \frac{2d}{d+2} < 2$.

Case 1.1: $d(q-1) \le 1$. Consider the following equation

$$g_1(t) := t^{1-\frac{q(d-1)}{2}} - \sqrt{\alpha_0} t^{\frac{1-d(q-1)}{2}} - \gamma m(q,d) = 0.$$

We have $\lim_{t\to 0} g_1(t) = -\gamma m(q,d) < 0$ and $\lim_{t\to+\infty} g_1(t) = +\infty$. Since g_1 is continuous, it follows that there exists $\alpha_* \in (0,\infty)$ such that $g_1(\alpha_*) = 0$, that is

$$\alpha_*^{1-\frac{q(d-1)}{2}} = \sqrt{\alpha_0}\alpha * ^{\frac{1-d(q-1)}{2}} + \gamma m(q,d), \quad \text{i.e.,} \quad \alpha_* = \sqrt{\alpha_0}\sqrt{\alpha_*} + \gamma m(q,d)\alpha_*^{\frac{d(q-1)}{2}}.$$

Similarly by considering the function $g_2(t) := t^{1-\frac{q(d-1)}{2}} - \sqrt{\beta_0} t^{\frac{1-d(q-1)}{2}} - \gamma m(q,d)$, we deduce that there exists $\beta_* \in (0,\infty)$ such that

$$\beta_* = \sqrt{\beta_0}\sqrt{\beta_*} + \gamma m(q,d)\beta_*^{\frac{d(q-1)}{2}}.$$

Case 1.2: $d(q-1) > 1$. Using the same argument as in the previous case for

$$g_3(t) = t^{1/2} - \sqrt{\alpha_0} - \gamma m(q,d) t^{\frac{d(q-1)-1}{2}} \text{ and } g_4(t) = t^{1/2} - \sqrt{\beta_0} - \gamma m(q,d) t^{\frac{d(q-1)-1}{2}}$$

we can show that there exist $\alpha_*, \beta_* \in (0,\infty)$ such that

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

Therefore in both Cases 1.1 and 1.2, there exist $\alpha_*, \beta_* \in (0,\infty)$ such that

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

Substituting these quantities into (4.4) we obtain

$$\alpha_* I = \sqrt{\alpha_0}\sqrt{\alpha_*} I + \gamma m(q,d)\alpha_*^{\frac{d(q-1)}{2}} I$$
$$\le \gamma m(q,d)(\det X)^{\frac{q-1}{2}} I + \sum_{i=1}^{n} \lambda_i (X^{1/2} A_i X^{1/2})^{1/2}$$
$$\le \sqrt{\beta_0}\sqrt{\beta_*} I + \gamma m(q,d)\beta_*^{\frac{d(q-1)}{2}} I = \beta_* I.$$

Thus $\alpha_* I \le \psi(X) \le \beta_* I$. By Brouwer's fixed point theorem, $\psi(X)$ has a fixed point in $[\alpha_* I, \beta_* I]$ as desired.

Case 2. $0 < q < 1$.

It follows from (4.3) that

$$\sqrt{\alpha_0}\sqrt{\alpha_*} I + \gamma m(q,d)\beta_*^{\frac{d(q-1)}{2}} I \le \sqrt{\alpha_0}\sqrt{\alpha_*} I + \gamma m(q,d)(\det X)^{\frac{q-1}{2}} I$$
$$\le \gamma m(q,d)(\det X)^{\frac{q-1}{2}} I + \sum_{i=1}^{n} \lambda_i (X^{1/2} A_i X^{1/2})^{1/2}$$
$$\le \sqrt{\beta_0}\sqrt{\beta_*} I + \gamma m(q,d)(\det X)^{\frac{q-1}{2}} I \le \sqrt{\beta_0}\sqrt{\beta_*} I + \gamma m(q,d)\alpha_*^{\frac{d(q-1)}{2}} I \tag{4.5}$$

Next we will show that following system has positive solutions $0 < \alpha_* < \beta_* < \infty$:

$$\begin{cases} \alpha_* = \sqrt{\alpha_0}\sqrt{\alpha_*} + \gamma m(q,d)\beta_*^{\frac{d(q-1)}{2}} \\ \beta_* = \sqrt{\beta_0}\sqrt{\beta_*} + \gamma m(q,d)\alpha_*^{\frac{d(q-1)}{2}}. \end{cases} \tag{4.6}$$

Define $f : (0,\infty)^2 \to (0,\infty)^2$ by

$$f\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} \sqrt{\alpha_0}\sqrt{x} + \gamma m(q,d) y^{\frac{d(q-1)}{2}} \\ \sqrt{\beta_0}\sqrt{y} + \gamma m(q,d) x^{\frac{d(q-1)}{2}} \end{pmatrix}$$

Set

$$a_* = \left(\frac{\sqrt{\alpha_0} + \sqrt{\alpha_0 + 4\gamma m(q,d)\beta_0^{(q-1)d/2}}}{2}\right)^2,$$
$$b_* = \left(\frac{\sqrt{\beta_0} + \sqrt{\beta_0 + 4\gamma m(q,d)\alpha_0^{(q-1)d/2}}}{2}\right)^2.$$

Thus a_* and b_* satisfy

$$a_* = \sqrt{\alpha_0}\sqrt{a_*} + \gamma m(q,d)\beta_0^{(q-1)d/2}, \quad b_* = \sqrt{\beta_0}\sqrt{b_*} + \gamma m(q,d)\alpha_0^{(q-1)d/2}.$$

We now show that $f : [\alpha_0, a_*] \times [\beta_0, b_*] \to [\alpha_0, a_*] \times [\beta_0, b_*]$. In fact, consider $\alpha_0 \le x \le a_*$ and $\beta_0 \le y \le b_*$. We have

$$\alpha_0 \le \sqrt{\alpha_0}\sqrt{x} \le \sqrt{\alpha_0}\sqrt{x} + \gamma m(q,d) y^{\frac{d(q-1)}{2}} \le \sqrt{\alpha_0}\sqrt{x} + \gamma m(q,d)\beta_0^{\frac{d(q-1)}{2}} = a_*,$$
$$\beta_0 \le \sqrt{\beta_0}\sqrt{y} \le \sqrt{\beta_0}\sqrt{y} + \gamma m(q,d) x^{\frac{d(q-1)}{2}} \le \sqrt{\beta_0}\sqrt{y} + \gamma m(q,d)\alpha_0^{\frac{d(q-1)}{2}} = b_*.$$

Thus $f((x,y)^T) \in [\alpha_0, a_*] \times [\beta_0, b_*]$. By Brouwer's fixed point theorem, f has a fixed point in $[\alpha_0, a_*] \times [\beta_0, b_*]$, which means that system (4.6) has a positive solution (α_*, β_*). Using this solution in (4.5) we obtain

$$\alpha_* I = \sqrt{\alpha_0}\sqrt{\alpha_*} I + \gamma m(q,d)\beta_*^{\frac{d(q-1)}{2}} I$$
$$\le \gamma m(q,d)(\det X)^{\frac{q-1}{2}} I + \sum_{i=1}^{n} \lambda_i (X^{1/2} A_i X^{1/2})^{1/2}$$
$$\le \sqrt{\beta_0}\sqrt{\beta_*} I + \gamma m(q,d)\alpha_*^{\frac{d(q-1)}{2}} I = \beta_* I.$$

Hence by Brouwer's fixed point theorem again, ψ has a fixed point in $[\alpha_* I, \beta_* I]$ as desired. This finishes the proof of the proposition. $\square$

Next we will show that the functional that we wish to minimize in (4.1) is strictly convex under rather general conditions. According to Propositions 2.1 and Lemma 2.2 we have

$$W_2^2(\mu_i,\mu)=\mathrm{tr}X+\mathrm{tr}A_i-2\mathrm{tr}\Big(A_i^{\frac12}XA_i^{\frac12}\Big)^{\frac12},$$
$$F_q(\mu)=-\frac{d}{2}C_1(q,d)+\Big[1-(1-q)\frac{d}{2}C_1(q,d)\Big]\ln_q\frac{C_0(q,d)}{(\det U)^{\frac12}}.$$

Therefore the minimization problem (4.1) can be written as

$$\min_{X\in\mathbb{S}(d,\mathbb{R})_+}\frac12 g(X) \tag{4.7}$$

where

$$\begin{aligned}g(X)&=\sum_{i=1}^n\lambda_i\mathrm{tr}A_i+\sum_{i=1}^n\lambda_i\mathrm{tr}\Big(X-2(A_i^{\frac12}XA_i^{\frac12})^{\frac12}\Big)\\&\quad+\gamma\Big[2-(1-q)dC_1(q,d)\Big]\ln_q\frac{C_0(q,d)}{(\det U)^{\frac12}}-\gamma dC_1(q,d)\\&=f_1(X)+\gamma\Big[2-(1-q)dC_1(q,d)\Big]\ln_q\frac{C_0(q,d)}{(\det U)^{\frac12}}-\gamma dC_1(q,d),\end{aligned} \tag{4.8}$$

with $f_1(X)=\sum_{i=1}^n\lambda_i\mathrm{tr}A_i+\sum_{i=1}^n\lambda_i\mathrm{tr}\Big(X-2(A_i^{\frac12}XA_i^{\frac12})^{\frac12}\Big)$, which appeared in (3.5). Note that by definition of the q-logarithmic function we have

$$\ln_q\frac{C_0(q,d)}{(\det U)^{\frac12}}=\frac{1}{1-q}\Big[C_0(q,d)^{1-q}(\det U)^{-\frac{1-q}{2}}-1\Big].$$

Using explicit formula of $C_1(q,d)$ we get

$$\begin{aligned}2-(1-q)dC_1(q,d)&=2-(1-q)d\frac{2}{2+(d+2)(1-q)}\\&=\frac{4(2-q)}{2+(d+2)(1-q)}.\end{aligned}$$

Substituting these expressions into (4.8) we get

$$\begin{aligned}g(X)&=f_1(X)+\frac{4\gamma(2-q)C_0(q,d)^{1-q}}{(2+(d+2)(1-q))(1-q)}(\det X)^{-\frac{1-q}{2}}\\&\quad-\frac{4(2-q)}{(1-q)(2+(d+2)(1-q))}-\gamma dC_1(q,d).\end{aligned} \tag{4.9}$$

The following proposition studies the convexity of g.

Proposition 4.3. *Suppose that $\alpha I\le A_i,X,\le\beta I$ for all $i=1,\dots,n$. The functional g given in (4.9) is strictly convex for all $\gamma\ge0$ when one of the following condition holds*

1. $0<q<1$,
2. $1<q\le1+\frac{2\alpha^2}{d\beta^2}$.

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

$$\gamma_0=\frac12\frac{\alpha^{1/2}}{\beta^{3/2}}\frac{1}{\frac{1}{\beta^2}-\frac{(q-1)d}{2\alpha^2}}\frac{1}{m(q,d)}\frac{1}{\beta^{d(q-1)/2}}.$$

Proof. We consider two cases.

Case 1. $1<q<\frac{d+4}{d+2}$.

Let $k(X):=\frac{4\gamma(2-q)C_0(q,d)^{1-q}}{(2+(d+2)(1-q))(1-q)}(\det X)^{\frac{q-1}{2}}$. Let $h(X):=(\det X)^{\frac{q-1}{2}}$. Similarly as in the proof of Theorem 3.1, using again Jacobi's formula for the derivative of the determinant and the chain rule, we get

$$Dh(X)(Y)=\frac{q-1}{2}(\det X)^{\frac{q-3}{2}}\cdot\det(X)\cdot\mathrm{tr}(X^{-1}Y)=\frac{q-1}{2}(\det X)^{\frac{q-1}{2}}\mathrm{tr}(X^{-1}Y).$$

Therefore, using the definition of $m(q,d)$, we have

$$\nabla k(X) = -\gamma m(q,d)(\det X)^{\frac{q-1}{2}} X^{-1} = -\gamma m(q,d) h(X) X^{-1}. \tag{4.10}$$

In the computations below the linear operator $P(X)$ is defined to be $P(X)Y = XYX$. This operator is called the quadratic representation in the literature. By the Leibniz rule, we get

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

Thus

$$\begin{aligned}
\langle \nabla^2 k(X)(H), H\rangle &= -\gamma m(q,d)(\det X)^{\frac{q-1}{2}}\left[\frac{q-1}{2}\left\langle X^{-1}, H\right\rangle^2 - \langle X^{-1}H, X^{-1}H\rangle\right]\\
&= -\gamma m(q,d)(\det X)^{\frac{q-1}{2}}\left[\frac{q-1}{2}\mathrm{tr}^2(X^{-1}H) - \|X^{-1}H\|^2\right].
\end{aligned}$$

Furthermore, according to [22], for $\alpha I \le A_i, X \le \beta I$, we have

$$\langle \nabla^2 f_1(X)(H), H\rangle \ge \frac{1}{2}\frac{\alpha^{1/2}}{\beta^{3/2}}\|H\|^2.$$

Thus we get

$$\begin{aligned}
\langle \nabla^2 g(X)(H), H\rangle &= \langle \nabla^2 f_1(X)(H), H\rangle + \langle \nabla^2 k(X)(H), H\rangle\\
&\ge -\gamma m(q,d)(\det X)^{\frac{q-1}{2}}\left[\frac{q-1}{2}\mathrm{tr}^2(X^{-1}H) - \|X^{-1}H\|^2\right] + \frac{1}{2}\frac{\alpha^{1/2}}{\beta^{3/2}}\|H\|^2\\
&= \gamma m(q,d)(\det X)^{\frac{q-1}{2}}\left[\langle P(X^{-1})H, H\rangle - \frac{q-1}{2}\mathrm{tr}^2(X^{-1}H)\right] + \frac{1}{2}\frac{\alpha^{1/2}}{\beta^{3/2}}\|H\|^2\\
&\ge \gamma m(q,d)(\det X)^{\frac{q-1}{2}}\left[\frac{1}{\beta^2}\|H\|^2 - \frac{q-1}{2}\|X^{-1}\|^2\|H\|^2\right] + \frac{1}{2}\frac{\alpha^{1/2}}{\beta^{3/2}}\|H\|^2\\
&= \left\{\gamma m(q,d)(\det X)^{\frac{q-1}{2}}\left[\frac{1}{\beta^2} - \frac{q-1}{2}\|X^{-1}\|^2\right] + \frac{1}{2}\frac{\alpha^{1/2}}{\beta^{3/2}}\right\}\|H\|^2\\
&\ge \left\{\gamma m(q,d)(\det X)^{\frac{q-1}{2}}\left[\frac{1}{\beta^2} - \frac{q-1}{2}\frac{d}{\alpha^2}\right] + \frac{1}{2}\frac{\alpha^{1/2}}{\beta^{3/2}}\right\}\|H\|^2\\
&\ge \left\{\gamma m(q,d)(\det X)^{\frac{q-1}{2}}\left[\frac{1}{\beta^2} - \frac{q-1}{2}\frac{d}{\alpha^2}\right] + \frac{1}{2}\frac{\alpha^{1/2}}{\beta^{3/2}}\right\}\|H\|^2.
\end{aligned}$$

From this estimate, we deduce the following cases

(i) If

$$1 < q \le 1 + \frac{2\alpha^2}{d\beta^2},$$

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

$$q < 1 + \frac{2}{d+2} \le 1 + \frac{2\alpha^2}{d\beta^2},$$

(ii) If

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

then for

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

the Hessian of g is positive since

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

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

$$\begin{aligned}\langle\nabla^2 k(X)(H),H\rangle &= \gamma m(q,d)(\det X)^{\frac{q-1}{2}}\left[\frac{1-q}{2}\langle X^{-1},H\rangle^2+\langle P(X^{-1})H,H\rangle\right]\\ &\ge \gamma m(q,d)(\det X)^{\frac{q-1}{2}}\frac{1}{\lambda^2_{\max}(X)}\|H\|^2.\end{aligned}$$

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

We are now ready to prove Theorem 4.1.

Proof of Theorem 4.1. Suppose that the hypothesis of the statement of Theorem 4.1 is satisfied, that is either (i) $0 < q \le 1$ or (ii) $1 < q \le 1+\frac{2\alpha^2}{d\beta^2}$ or (iii) $1+\frac{2\alpha^2}{d\beta^2} < q < \frac{d+4}{d+2}$. Suppose that γ is sufficiently small in the last case; in the other cases it can be arbitrarily positive. As has been shown in the paragraph before Proposition 4.3, the minimization problem (4.1) can be written as

$$\min_{X\in\mathbb{H}}\frac{1}{2}g(X),$$

where $g(X)$ is given in (4.9)

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

By Proposition 4.3, $X\mapsto g(X)$ is strictly convex. Now we compute the derivative of $g(X)$. We have

$$\nabla g(X) = \nabla f_1(X)+\nabla k(X), \tag{4.11}$$

According to the proof of Theorem 3.1 we have

$$\nabla f_1(X) = I-\sum_{i=1}^n \lambda_i(A_i\sharp X^{-1}).$$

By (4.10), we have

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

Substituting these computations into (4.11) we obtain

$$\nabla g(X) = \Big(I-\sum_{i=1}^n \lambda_i(A_i\sharp X^{-1})\Big)-\gamma m(q,d)(\det X)^{\frac{q-1}{2}}X^{-1}.$$

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

$$I-\gamma m(q,d)(\det X)^{\frac{q-1}{2}}X^{-1} = \sum_{i=1}^n \lambda_i(A_i\sharp X^{-1}),$$

which, by using the definition of the geometric mean (3.6), is equivalent to

$$X-\gamma m(q,d)(\det X)^{\frac{q-1}{2}}I = \sum_{i=1}^n \lambda_i\Big(X^{\frac{1}{2}}A_iX^{\frac{1}{2}}\Big)^{\frac{1}{2}}.$$

This is precisely Eq. (4.2). By Proposition 4.2, it has a positive definite solution. This, together with the strictly convexity of g, guarantees the existence and uniqueness of a minimizer of g. We complete the proof of the theorem. □

5. GPM-based algorithm

In this section, we propose an GPM-based algorithm for solving regularization problems(3.4) and (4.7), and analyze its convergence properties.

First, we formally describe the algorithmic procedure for the gradient projection method (GPM) below.

Algorithm 1 GPM

Choose $X^0 \in [\hat{\alpha}I, \hat{\beta}I]$. Initialize $k = 0$. Update $X^{(k+1)}$ from $X^{(k)}$ by the following template:

Step 1. Find $\bar{X}^{(k)} = [X^{(k)} - \nabla\psi(X^{(k)})]^+$,

Step 2. Select a stepsize $t^{(k)}$,

Step 3. $X^{(k+1)} = X^{(k)} + t^{(k)}(\bar{X}^{(k)} - X^{(k)})$.

Here $[\cdot]^+$ denotes the projection on the set $[\hat{\alpha}I, \hat{\beta}I]$.

The stepsize is selected by Armijo rule along the feasible direction [41]. It is described below.

Let $t^{(k)}$ be the largest element of $\{\xi^j\}_{j=0,1,\ldots}$ satisfying

$$\psi(X^{(k)} + t^{(k)}D^{(k)}) \leq \psi(X^{(k)}) - \sigma t^{(k)}\langle\nabla\psi(X^{(k)}),\ D^{(k)}\rangle,$$

where $0 < \xi < 1$, $0 < \sigma < 1$, and $D^{(k)} = \bar{X}^{(k)} - X^{(k)}$.

Note that $\psi = f$ for the regularization problem (3.4) and $\psi = g$ for the regularization problem (4.7). The projection of the matrix $S \in \mathcal{S}^d$, where $\mathcal{S}^d$ is the set of $d \times d$ symmetric matrices, onto the set $[\hat{\alpha}I, \hat{\beta}I]$ is to find the solution of the following minimization problem

$$\min_{X \in [\hat{\alpha}I, \hat{\beta}I]} \|X - S\|_F.$$

The solution of the above problem is

$$[S]^+ = U\mathrm{Diag}(\min(\max(\hat{\alpha}, \lambda_1), \hat{\beta}), \ldots, \min(\max(\hat{\alpha}, \lambda_d), \hat{\beta}))U^T,$$

where $\lambda_1 \geq \cdots \geq \lambda_d$ are the eigenvalues of S and U is a corresponding orthogonal matrix of eigenvalues of S.

Now, we establish the global convergence of GPM. For the proof, we refer to [41, Proposition 2.3.1].

Theorem 5.1. *Let $\{X^{(k)}\}$ be the sequence generated by GPM with $t^{(k)}$ chosen by Armijo rule along the feasible direction. Then every limit point of $\{X^{(k)}\}$ is stationary.*

In the following subsections, we show the Lipschitz continuity of the gradient function of the regularization problems. In this case, we can use a constant stepsize for the gradient projection method. That is, $t^{(k)} = \frac{1}{L}$ where L is a Lipschitz constant. Then we have

$$X^{(k+1)} = X^{(k)} + \frac{1}{L}(\bar{X}^{(k)} - X^{(k)}).$$

5.1. Regularization of barycenters for Gaussian measures

We recall that the unique minimizer of the minimization problem (3.1) in the space of Gaussian measures satisfies the following nonlinear matrix equation $\nabla f(X) = 0$ where

$$\nabla f(X) = I - \sum_{i=1}^{n} \lambda_i (A_i \sharp X^{-1}) - \gamma X^{-1} =: F_1(X) - \gamma F_2(X).$$

We establish the following theorem for the Lipschitz continuity of the gradient function.

Theorem 5.2. *Suppose that $A_i \in [\alpha I, \beta I]$ for all $i = 1, \ldots, n$. Then for $\alpha I \leq X \neq Y \leq \beta I$ we have*

$$\frac{\|\nabla f(X) - \nabla f(Y)\|_F}{\|X - Y\|_F} \leq \frac{\beta^2}{2\alpha^3} + \frac{\gamma}{\alpha^2}.$$

Proof. According to [9, Proof of Theorem 3.1] we have

$$\frac{\|F_1(X)-F_1(Y)\|_F}{\|X-Y\|_F} \le \frac{\beta^2}{2\alpha^3} \quad \text{and} \quad \frac{\|F_2(X)-F_2(Y)\|_F}{\|X-Y\|_F} \le \frac{1}{\alpha^2}.$$

Therefore we get

$$\begin{aligned}\frac{\|\nabla f(X)-\nabla f(Y)\|_F}{\|X-Y\|_F} &\le \frac{\|F_1(X)-F_1(Y)\|_F + \gamma\|F_2(X)-F_2(Y)\|_F}{\|X-Y\|_F}\\ &\le \frac{\beta^2}{2\alpha^3}+\frac{\gamma}{\alpha^2}. \quad \square\end{aligned}$$

5.2. Regularization of barycenters for q-Gaussian measures

We recall that the unique minimizer of the minimization problem (4.1) in the space of q-Gaussian measures solves the nonlinear matrix equation $\nabla g(X)$=0 where

$$\nabla g(X) = \Big(I-\sum_{i=1}^n \lambda_i(A_i \sharp X^{-1})\Big) - \gamma m(q,d)(\det X)^{\frac{q-1}{2}}X^{-1} =: F_1(X) - \gamma m(q,d)\tilde{h}(X), \tag{5.1}$$

where $F_1(X) = \Big(I-\sum_{i=1}^n \lambda_i(A_i \sharp X^{-1})\Big)$ as in the previous section and $\tilde{h}(X) = (\det X)^{\frac{q-1}{2}}X^{-1} = h(X)X^{-1}$. The following main theorem of this section proves the Lipschitz continuity of ∇g.

Theorem 5.3. *Suppose that $A_i \in [\alpha I, \beta I]$ for all $i = 1,\dots,n$. Then for $\alpha I \le X \ne Y \le \beta I$, we have*

$$\frac{\|\nabla g(X)-\nabla g(Y)\|_F}{\|X-Y\|_F} \le \begin{cases} \frac{\beta^2}{2\alpha^3}+\frac{\gamma}{\alpha^2}+\frac{\gamma m(q,d)}{\alpha^2}\cdot\beta^{\frac{q-1}{2}d}\Big(1+\frac{q-1}{2}d\Big), & \text{if } 1<q<\frac{d+4}{d+2},\\ \frac{\beta^2}{2\alpha^3}+\frac{\gamma}{\alpha^2}+\gamma m(q,d)\alpha^{-2+\frac{q-1}{2}d}\Big(1+\frac{1-q}{2}d\Big), & \text{if } 0<q<1.\end{cases}$$

Proof. Let $\alpha I \le X,\ Y \le \beta I$. According to the proof of Theorem 5.2, we have

$$\frac{\|F_1(X)-F_1(Y)\|_F}{\|X-Y\|_F} \le \frac{\beta^2}{2\alpha^3}+\frac{\gamma}{\alpha^2}. \tag{5.2}$$

It remains to study the Lipschitz continuity of $\tilde{h}(X) = (\det X)^{\frac{q-1}{2}}X^{-1} = h(X)X^{-1}$.

Case 1. $1<q<\frac{d+4}{d+2}$. First, we have

$$\begin{aligned}|h(X)-h(Y)| &= \Big|\exp(\ln(\det X)^{\frac{q-1}{2}}) - \exp(\ln(\det Y)^{\frac{q-1}{2}})\Big|\\ &= e^{\theta}\Big|\ln(\det X)^{\frac{q-1}{2}} - \ln(\det Y)^{\frac{q-1}{2}}\Big|\\ &\le \beta^{\frac{q-1}{2}d}\Big|\ln(\det X)^{\frac{q-1}{2}} - \ln(\det Y)^{\frac{q-1}{2}}\Big|\\ &= \frac{q-1}{2}\cdot\beta^{\frac{q-1}{2}d}|\ln\det X - \ln\det Y|\\ &\le \frac{q-1}{2}\cdot\beta^{\frac{q-1}{2}d}\Big(\max_{\alpha I\le X\le\beta I}\|X^{-1}\|\Big)\|X-Y\|\\ &\le \frac{q-1}{2}\cdot\beta^{\frac{q-1}{2}d}\cdot\frac{\sqrt{d}}{\alpha}\|X-Y\|\end{aligned}$$

where $\ln\alpha^{\frac{q-1}{2}d} \le \theta \le \ln\beta^{\frac{q-1}{2}d}$ because $\ln\alpha^{\frac{q-1}{2}d} \le \ln(\det X)^{\frac{q-1}{2}} \le \ln\beta^{\frac{q-1}{2}d}$. The second equality and inequality are derived from the mean value theorem. Moreover, we get

$$\begin{aligned}\|\tilde{h}(X)-\tilde{h}(Y)\| &= \|h(X)(X^{-1}-Y^{-1}) + (h(X)-h(Y))Y^{-1}\|\\ &\le h(X)\|X^{-1}-Y^{-1}\| + |h(X)-h(Y)|\,\|Y^{-1}\|\\ &\le \Big(\max_{\alpha I\le X\le\beta I} h(X)\Big)\cdot\frac{1}{\alpha^2}\|X-Y\|\\ &+ \Big(\max_{\alpha I\le Y\le\beta I}\|Y^{-1}\|\Big)\cdot\frac{q-1}{2}\cdot\beta^{\frac{q-1}{2}d}\cdot\frac{\sqrt{d}}{\alpha}\|X-Y\|\end{aligned}$$

$$= \left(\beta^{\frac{q-1}{2}d} \cdot \frac{1}{\alpha^2} + \frac{\sqrt{d}}{\alpha} \cdot \frac{q-1}{2} \cdot \beta^{\frac{q-1}{2}d} \cdot \frac{\sqrt{d}}{\alpha}\right) \|X - Y\|$$

$$= \frac{1}{\alpha^2} \cdot \beta^{\frac{q-1}{2}d}\left(1 + \frac{q-1}{2}d\right) \|X - Y\| \tag{5.3}$$

where the second inequality comes from [9, Proof of Theorem 3.1].

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

$$|h(X) - h(Y)| \leq \frac{1-q}{2} \cdot \alpha^{\frac{q-1}{2}d} \cdot \frac{\sqrt{d}}{\alpha} \|X - Y\|.$$

Hence

$$\|\tilde{h}(X) - \tilde{h}(Y)\| \leq \left(\max_{\alpha I \leq X \leq \beta I} h(X)\right) \cdot \frac{1}{\alpha^2} \|X - Y\|$$

$$+ \left(\max_{\alpha I \leq Y \leq \beta I} \|Y^{-1}\|\right) \cdot \frac{1-q}{2} \cdot \alpha^{\frac{q-1}{2}d} \cdot \frac{\sqrt{d}}{\alpha} \|X - Y\|$$

$$= \left(\alpha^{\frac{q-1}{2}d} \cdot \frac{1}{\alpha^2} + \frac{\sqrt{d}}{\alpha} \cdot \frac{1-q}{2} \cdot \alpha^{\frac{q-1}{2}d} \cdot \frac{\sqrt{d}}{\alpha}\right) \|X - Y\|$$

$$= \alpha^{-2+\frac{q-1}{2}d}\left(1 + \frac{1-q}{2}d\right) \|X - Y\|. \tag{5.4}$$

Substituting the estimates (5.2), (5.3) and (5.4) back into (5.1) we obtain the desired inequality. □

6. Numerical experiments

In this section, we report numerical tests for the regularized model (4.7) of barycenters in the space of q-Gaussian measures on randomly generated data.

6.1. Test on various q and stability

in this subsection, we numerically observe how the solution is affected as $q \to 1$. To see this, we apply the gradient projection method to the regularization of barycenters for q-Gaussian measures (4.7) on n randomly generated matrices of the size $d \times d$. The random matrices we use for our test are generated by matlab code as follows:

$$\begin{aligned} &\textit{for} \quad i = 1 : n \\ &\qquad [Q, \ ] = qr(randn(d)); \\ &\qquad A_i = Q * diag(\text{eiglb} + \text{eigub} * rand(d, 1)) * Q'; \end{aligned}$$

The eigenvalues of generated matrices are randomly distributed in the interval [eiglb, eiglb + eigub]. In our experiments, we set $n = 100$, $d = 10$ if $q < 1$ and $n = 50$, $d = 5$ if $q > 1$. And we set eiglb $= 0.1$ and eigub $= 9.9$.

We set $\xi = 0.5$, $\sigma = 0.1$, $\hat{\alpha} = 10^{-5}$, $\hat{\beta} = 10^5$, $\lambda_i = 1/n$, $i = 1, \dots, n$ for GPM in our experiment. All runs are performed on a Laptop with Intel Core i7-10510U CPU (2.30 GHz) and 16 GB Memory, running 64-bit windows 10 and MATLAB (Version 9.8). Throughout the experiments, we choose the initial iterate to be $X^0 = I$ and stop the algorithm when $\| D^{(k)} \|_F \leq 10^{-8}$.

We report in Table 1 our numerical results, showing the Frobenius norm of the difference between the final estimated solution of the model (4.7) with $q = 0.5$ and that with various given q less than 1. In Table 2, the difference between the final estimated solution of the model (4.7) with $q = 1.25$ and that with various given q greater than 1 is reported.

From Tables 1–2, we see that the difference is increasing as q goes to 1 when the regularization parameter γ is fixed and we observe that the bigger the regularization parameter γ is, the bigger the difference is when q is fixed. In particular, when q is fixed, the position of the first nonzero decimal point of the value is related to the position of the first nonzero decimal point of the regularization parameter γ, i.e., if γ is multiplied by 10, then the position of first nonzero decimal point of the difference is moved one position forward as if multiplied by 10.

In the next experiment, we investigate stability properties for the model (4.7). We perturb the given data, A_i as follows:

$$B_i = A_i + \epsilon I \quad i = 1, \dots, n$$

In Table 3, we can observe that $\|X_B - X_A\|_F \leq 5\epsilon$, where X_A is the final estimated solution of the model (4.7) with data A_i and X_B is that with the perturbed data B_i, for all the cases. We see that if q is fixed, then the value $\|X_B - X_A\|_F/\epsilon$ tends to decrease as the regularization parameter γ is decreasing and, if the regularization parameter γ is fixed, then the value is increasing as q goes to 1 except when $q = 0.99$ and $\gamma = 0.1, 0.01$.

From Table 4, for all the cases, $\|X_B - X_A\|_F \leq 3\epsilon$ is observed. It is observed that the value $\|X_B - X_A\|_F/\epsilon$ is decreasing as q goes away from 1 when the regularization parameter γ is fixed and the value is increasing as the regularization

Table 1
Test results of the value $\|X_{0.5} - X_q\|_F$ where $X_{0.5}$ is the final estimated solution of the model (4.7) with $q = 0.5$ and X_q is that with various given q less than 1 on 5 random data sets.

q	Difference when $\gamma = 1$				
0.6	0.00110	0.00111	0.00119	0.00101	0.00108
0.7	0.01023	0.01029	0.01088	0.00959	0.01012
0.8	0.08709	0.08746	0.09084	0.08338	0.08646
0.9	0.76237	0.76417	0.78058	0.74408	0.75926
0.99	6.0825	6.09018	6.18696	5.97885	6.05433
q	Difference when $\gamma = 0.1$				
0.6	0.00011	0.00011	0.00012	0.00010	0.00011
0.7	0.00102	0.00103	0.00109	0.00096	0.00101
0.8	0.00865	0.00869	0.00902	0.00829	0.00859
0.9	0.07344	0.07360	0.07500	0.07186	0.07318
0.99	0.51458	0.51471	0.51590	0.51321	0.51434
q	Difference when $\gamma = 0.01$				
0.6	0.000011	0.000011	0.000011	0.000009	0.000010
0.7	0.00010	0.00010	0.00011	0.00010	0.00010
0.8	0.00086	0.00087	0.00090	0.00083	0.00086
0.9	0.00732	0.00733	0.00747	0.00716	0.00729
0.99	0.05095	0.05096	0.05105	0.05083	0.05093

Table 2
Test results of the value $\|X_{1.25} - X_q\|_F$ where $X_{1.25}$ is the final estimated solution of the model (4.7) with $q = 1.25$ and X_q is that with various given q greater than 1 on 5 random data sets.

q	Difference when $\gamma = 0.01$				
1.2	0.41764	0.39779	0.38627	0.43071	0.42259
1.1	0.71682	0.68549	0.66727	0.73737	0.72459
1.01	0.80209	0.76872	0.74931	0.82395	0.81036
q	Difference when $\gamma = 0.001$				
1.2	0.04387	0.04184	0.04067	0.04520	0.04437
1.1	0.07421	0.07104	0.06919	0.07630	0.07500
1.01	0.08274	0.07936	0.07740	0.08496	0.08358

Table 3
Test results of the value $\|X_B - X_A\|_F/\epsilon$ where X_A is the final estimated solution of the model (4.7) with data A_i and X_B is that with the perturbed data B_i on 5 random data sets when $q < 1$.

q	$\gamma = 1$ and $\epsilon = 10^{-2}$					$\gamma = 1$ and $\epsilon = 10^{-3}$					$\gamma = 1$ and $\epsilon = 10^{-5}$				
0.6	3.897	3.788	3.877	3.831	3.843	3.905	3.793	3.884	3.837	3.850	3.906	3.793	3.885	3.837	3.850
0.7	3.901	3.792	3.881	3.834	3.847	3.909	3.796	3.888	3.840	3.853	3.910	3.797	3.888	3.841	3.854
0.8	3.921	3.811	3.903	3.852	3.866	3.929	3.816	3.909	3.858	3.873	3.930	3.816	3.910	3.859	3.873
0.9	4.014	3.903	4.002	3.937	3.957	4.023	3.908	4.009	3.947	3.964	4.024	3.908	4.010	3.944	3.965
0.99	4.637	4.510	4.761	4.414	4.508	4.653	4.522	4.777	4.424	4.520	4.654	4.523	4.779	4.426	4.521
	$\gamma = 0.1$ and $\epsilon = 10^{-2}$					$\gamma = 0.1$ and $\epsilon = 10^{-3}$					$\gamma = 0.1$ and $\epsilon = 10^{-5}$				
0.6	3.897	3.788	3.877	3.830	3.843	3.905	3.792	3.883	3.836	3.849	3.906	3.793	3.884	3.837	3.850
0.7	3.897	3.788	3.877	3.830	3.843	3.905	3.793	3.884	3.836	3.849	3.906	3.793	3.884	3.837	3.850
0.8	3.899	3.790	3.879	3.832	3.845	3.907	3.794	3.886	3.838	3.851	3.908	3.795	3.886	3.839	3.852
0.9	3.907	3.797	3.887	3.839	3.852	3.915	3.802	3.894	3.845	3.859	3.916	3.803	3.895	3.846	3.860
0.99	3.906	3.796	3.886	3.838	3.851	3.913	3.801	3.892	3.844	3.858	3.914	3.801	3.893	3.845	3.858
	$\gamma = 0.01$ and $\epsilon = 10^{-2}$					$\gamma = 0.01$ and $\epsilon = 10^{-3}$					$\gamma = 0.01$ and $\epsilon = 10^{-5}$				
0.6	3.897	3.787	3.877	3.830	3.843	3.905	3.792	3.883	3.836	3.849	3.907	3.796	3.885	3.840	3.841
0.7	3.897	3.788	3.877	3.830	3.843	3.905	3.792	3.883	3.836	3.849	3.906	3.793	3.884	3.837	3.850
0.8	3.897	3.788	3.877	3.830	3.843	3.905	3.792	3.883	3.836	3.849	3.906	3.793	3.884	3.837	3.850
0.9	3.898	3.788	3.878	3.831	3.844	3.906	3.793	3.884	3.837	3.850	3.906	3.794	3.885	3.838	3.851
0.99	3.898	3.788	3.877	3.831	3.843	3.905	3.793	3.884	3.837	3.850	3.906	3.793	3.885	3.837	3.851

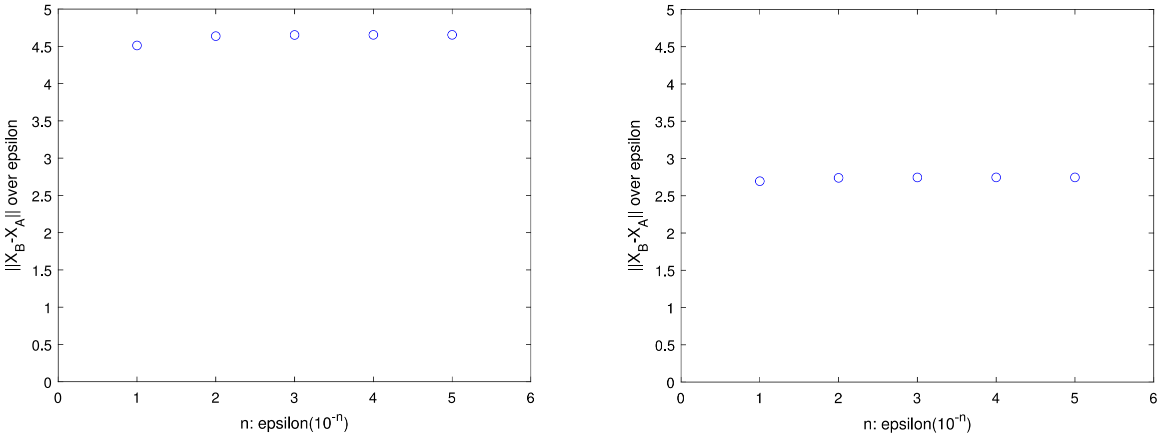
parameter γ is decreasing when q is fixed. To illustrate the aforementioned trends, we present $\|X_B - X_A\|_F/\epsilon$ versus ϵ and $\|X_B - X_A\|_F/\epsilon$ versus q for one random data in Figs. 1 and 2 respectively.

To visualize the effect of γ, we create the following toy example:

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \ A_2 = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}, \ A_3 = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}.$$

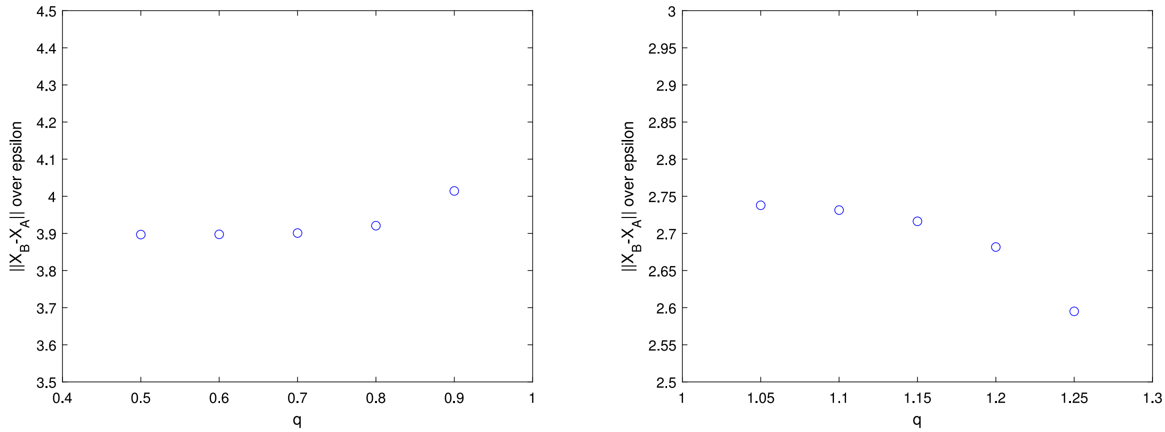
Table 4
Test results of the value $\|X_B - X_A\|_F/\epsilon$ where X_A is the final estimated solution of the model (4.7) with data A_i and X_B is that with the perturbed data B_i on 5 random data sets when $q > 1$.

	$\gamma = 0.01$ and $\epsilon = 10^{-2}$					$\gamma = 0.01$ and $\epsilon = 10^{-3}$					$\gamma = 0.01$ and $\epsilon = 10^{-5}$				
1.2	2.682	2.643	2.761	2.628	2.672	2.687	2.647	2.767	2.632	2.677	2.688	2.647	2.768	2.633	2.677
1.1	2.731	2.694	2.815	2.676	2.722	2.737	2.698	2.821	2.680	2.726	2.738	2.698	2.822	2.681	2.726
1.01	2.740	2.703	2.824	2.684	2.730	2.746	2.707	2.831	2.689	2.735	2.746	2.707	2.832	2.689	2.735
	$\gamma = 0.001$ and $\epsilon = 10^{-2}$					$\gamma = 0.001$ and $\epsilon = 10^{-3}$					$\gamma = 0.001$ and $\epsilon = 10^{-5}$				
1.2	2.735	2.697	2.818	2.679	2.725	2.740	2.701	2.825	2.683	2.730	2.741	2.701	2.826	2.684	2.730
1.1	2.740	2.702	2.824	2.684	2.730	2.745	2.706	2.830	2.688	2.734	2.746	2.706	2.831	2.688	2.734
1.01	2.740	2.703	2.825	2.685	2.731	2.746	2.707	2.831	2.689	2.735	2.747	2.707	2.832	2.689	2.735



(a) $q = 0.99$ with $\gamma = 1$ (b) $q = 1.01$ with $\gamma = 0.01$

Fig. 1. $\frac{\|X_B - X_A\|_F}{\epsilon}$ vs. ϵ.



(a) $\gamma = 1$ with $\epsilon = 10^{-2}$ (b) $\gamma = 0.01$ with $\epsilon = 10^{-2}$

Fig. 2. $\frac{\|X_B - X_A\|_F}{\epsilon}$ vs. q.

In this experiment, we first set $q = 0.5$ and $\epsilon = 10^{-5}$.

$$X_{A,1} = \begin{bmatrix} 4.29505684 & 0 \\ 0 & 4.29505684 \end{bmatrix} X_{B,1} = \begin{bmatrix} 4.29506977 & 0 \\ 0 & 4.29506977 \end{bmatrix}$$
$$X_{A,0.5} = \begin{bmatrix} 4.42471564 & 0 \\ 0 & 4.42471564 \end{bmatrix} X_{B,0.5} = \begin{bmatrix} 4.42472836 & 0 \\ 0 & 4.42472836 \end{bmatrix}$$

Table 5
Test results of algorithms on 5 random data sets when $q = 0.5$ and $q = 1.25$.

	$q = 0.5$ with $\gamma = 1$					$q = 0.5$ with $\gamma = 0.1$				
GPM	333	327	315	343	334	333	321	315	338	327
RGM	222	222	219	226	223	33	33	32	33	33
	$q = 1.25$ with $\gamma = 0.01$					$q = 1.25$ with $\gamma = 0.001$				
GPM	307	285	285	331	310	325	302	300	348	328
RGM	2	2	2	2	2	2	2	2	2	2

$$X_{A,0.1} = \begin{bmatrix} 4.52436136 & 0 \\ 0 & 4.52436136 \end{bmatrix} X_{B,0.1} = \begin{bmatrix} 4.52437393 & 0 \\ 0 & 4.52437393 \end{bmatrix}$$
$$X_{A,0.01} = \begin{bmatrix} 4.54632757 & 0 \\ 0 & 4.546327571 \end{bmatrix} X_{B,0.01} = \begin{bmatrix} 4.54634011 & 0 \\ 0 & 4.54634011 \end{bmatrix}$$
$$X_{A,0} = \begin{bmatrix} 4.54875843 & 0 \\ 0 & 4.54875843 \end{bmatrix} X_{B,0} = \begin{bmatrix} 4.54877096 & 0 \\ 0 & 4.54877096 \end{bmatrix},$$

where $X_{A,\gamma}$ is the final estimated solution with the given γ and data A_i and $X_{B,\gamma}$ is the final estimated solution with the given γ and the perturbed data B_i.

Next, we set $q = 0.5$ and $\epsilon = 10^{-3}$.

$$X_{A,1} = \begin{bmatrix} 4.29505684 & 0 \\ 0 & 4.29505684 \end{bmatrix} X_{B,1} = \begin{bmatrix} 4.29635024 & 0 \\ 0 & 4.29635024 \end{bmatrix}$$
$$X_{A,0.5} = \begin{bmatrix} 4.42471564 & 0 \\ 0 & 4.42471564 \end{bmatrix} X_{B,0.5} = \begin{bmatrix} 4.42598739 & 0 \\ 0 & 4.42598739 \end{bmatrix}$$
$$X_{A,0.1} = \begin{bmatrix} 4.52436136 & 0 \\ 0 & 4.52436136 \end{bmatrix} X_{B,0.1} = \begin{bmatrix} 4.52561833 & 0 \\ 0 & 4.52561833 \end{bmatrix}$$
$$X_{A,0.01} = \begin{bmatrix} 4.54632757 & 0 \\ 0 & 4.54632757 \end{bmatrix} X_{B,0.01} = \begin{bmatrix} 4.54758147 & 0 \\ 0 & 4.54758147 \end{bmatrix}$$
$$X_{A,0} = \begin{bmatrix} 4.54875843 & 0 \\ 0 & 4.54875843 \end{bmatrix} X_{B,0} = \begin{bmatrix} 4.54877096 & 0 \\ 0 & 4.54877096 \end{bmatrix},$$

where $X_{A,\gamma}$ is the final estimated solution with the given γ and data A_i and $X_{B,\gamma}$ is the final estimated solution with the given γ and the perturbed data B_i.

From this experiment, we can see that the bigger the penalty parameter γ is, the smaller the value of each diagonal entry of the solution is. In other words, the regularization term has the effect of reducing the value of diagonal entries. We also observe that diagonal entries of $X_{A,\gamma}$ and $X_{B,\gamma}$ are equal up to the second decimal place when $\epsilon = 10^{-3}$ and diagonal entries of $X_{A,\gamma}$ and $X_{B,\gamma}$ are equal up to the fourth decimal place when $\epsilon = 10^{-5}$ as expected.

6.2. Comparison of gradient projection method and Riemannian gradient method

In this subsection, we compare the performance of the gradient projection method on Euclidean space with that of Riemannian gradient methods (RGM) [30–32]. RGM that we are comparing is the version in [30] and it has been proved in [30, Theorem 4] that RGM has a linear convergence rate and the rate depends on the regularization parameter and the distribution of eigenvalues of given data matrices but does not depends on the dimension of matrices. The algorithmic procedure of this method is briefly described below. We note that the stepsize is selected by the rule in [30, Theorem 4], i.e., $t^k = 1/(1 + 2\gamma\hat{\beta})$ for all k. We refer [30] for detailed description of RGM.

Algorithm 2 RGM

Choose $X^0 \in [\hat{\alpha}I, \hat{\beta}I]$. Initialize $k = 0$. Update $X^{(k+1)}$ from $X^{(k)}$ by the following template:

Step 1. Find $\mathrm{grad}\psi(X^k)$,

Step 2. Select a stepsize $t^{(k)}$,

Step 3. $X^{(k+1)} = \mathrm{Exp}_{X^k}(-t^k \mathrm{grad}\psi(X^k))$.

Here $\mathrm{grad}\psi(X^k)$ denotes the Riemannian gradient at X^k and $\mathrm{Exp}_{X^k}(\cdot)$ denotes the exponential map.

For comparison, we generate random matrices using the same procedure as in Section 6.1 and we first run GPM and then RGM until it reaches the same objective function value reached by GPM. Note that $q = 0.5$ with $\gamma = 1$ and $\gamma = 0.1$

Table 6
Test results of algorithms on 5 random data sets when $q = 1$ with various γ.

	$\gamma = 10$					$\gamma = 1$					$\gamma = 0.1$				
GPM	792	765	745	834	808	349	351	323	361	352	296	298	283	303	299
RGM	2184	2121	2095	2270	2237	290	293	279	294	295	39	40	40	39	40

Table 7
Test results of algorithms on 5 random data sets when $q = 1$ and $\gamma = 1$ with different distribution of eigenvalues.

	eiglb = 0.1 and eigub = 9.9					eiglb = 0.01 and eigub = 99.99				
GPM	349	351	323	361	352	2695	2649	2569	2766	2705
RGM	290	293	279	294	295	2962	2956	2970	2946	2965

Table 8
Test results of algorithms on 5 random data sets when $q = 1$ and $\gamma = 1$ with different dimension of matrices.

	$d = 10$					$d = 100$				
GPM	349	351	323	361	352	363	363	340	339	346
RGM	290	293	279	294	295	299	301	282	283	287

and $q = 1.25$ with $\gamma = 0.01$ and $\gamma = 0.001$ are used. In Table 5, we report the number of iterations of 5 random instances. From Table 5, RGM reaches the final objective value of GPM in a few iterations for the case $q = 1.25$, but, for the case $q = 0.5$ with $\gamma = 1$, it requires many iterations to reach that of GPM and so the difference in the number of iterations between RGM and GPM is not as large as in the other cases. The performance of GPM does not depend significantly on the parameters q and γ but that of RGM depends on the parameter γ since the stepsize for RGM relies on γ.

To further investigate the dependence of the regularization parameter γ, we compare GPM and RGM for the case $q = 1$ with various γ from 10 to 0.1. In Table 6, the dependence of the parameter γ is greater in RGM than in GPM. If the parameter γ is increased a factor of 10, then the number of iterations for RGM is approximately 7 times larger.

In order to investigate the effect of eigenvalue distribution on the performance of algorithms, we run algorithms with eiglb = 0.01 and eigub = 99.99. And to observe the effect of the dimension of matrices on the performance of algorithms, random matrices with $d = 100$ are generated for testing. We note that $q = 1$ is set for observing those effects (see Tables 7 and 8).

The distribution of eigenvalues for randomly generated matrices affects the performance of both GPM and RGM. GPM requires 8 times and RGM needs 10 times more iterations if the distribution range is 10 times wider. The dimension of matrices has little effect on the performance of both GPM and RGM. Neither methods requires much more iterations.

From numerical comparison of GPM with RGM, if the regularization parameter is relatively large, it is recommended to use GPM, otherwise it is recommended to use RGM.

7. Conclusion and outlook

In this paper we have studied the Wasserstein barycenter problem for Gaussian and q-Gaussian measures, each regularized by a particular entropy functional. We have provided the existence and a characterization of the barycenters and proposed an algorithm based on gradient projection method in the space of matrices in order to compute them. We have also numerically shown the influence of parameters and stability of the algorithm under small perturbation of data and compared the gradient projection method with Riemannian gradient method. As has been shown in Proposition 2.1, the space of φ-exponential measures is also isometric to the space of Gaussian measures with respect to the Wasserstein distance. Therefore, the unregularized Wasserstein barycenter problem for φ-exponential measures can be solved exactly the same as in the case of Gaussian measures. However, we presently do not know of an explicit formulation of the entropy in the case of φ-exponential measures and leave the problem of regularized Wasserstein barycenter for them for further study. Another interesting topic for future work would be to generalize the results of this paper to probability measures living in different spaces using unbalanced optimal transport.

Acknowledgments

We thank the Principal Editor and anonymous referees for providing useful comments and suggestions that have helped us to improve the presentation of the paper. The idea of the proof of Proposition 2.3 was suggested by a referee.

The work of M. H. Duong was supported by EPSRC, UK Grants EP/W008041/1 and EP/V038516/1. The work of S. Yun was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2016R1A5A1008055 and No. 2022R1A2C1011503). The work of Y. Lim was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) No. 2015R1A3A2031159.

Appendix. φ-Exponential measures

We recall that for a given increasing, positive and continuous function φ on $(0,\infty)$, the φ-logarithmic function and the φ-exponential function are respectively defined in (1.2) and (1.3). Two important classes of φ-exponential functions are:

(i) $\varphi(s) = s$: the φ-logarithmic function and the φ-exponential function become the traditional logarithmic and exponential functions: $\ln_\varphi(t) = \ln(t), \quad \exp_\varphi(t) = \exp(t)$.

(ii) $\varphi(s) = s^q$ for some $q > 0$: the φ-logarithmic function and the φ-exponential function become the q-logarithmic and q-exponential functions respectively

$$\ln_\varphi(t) = \log_q(t) = \frac{t^{1-q}-1}{1-q} \quad \text{for } t>0, \quad \exp_\varphi(t) = \exp_q(t) = \Big(1+(1-q)t\Big)_+^{\frac{1}{1-q}},$$

where $[x]_+ = \max\{0,x\}$ and by convention $0^a := \infty$. The q-logarithmic function satisfies the following property

$$\ln_q(xy) = \ln_q(x) + \ln_q(y) + (1-q)\ln_q(x)\ln_q(y). \tag{A.1}$$

Definition A.1. For any $a \in \mathbb{R}$, we define $\mathcal{O}(a)$ to be the set of all increasing, positive, continuous functions φ on $(0,\infty)$ such that $\max\{\delta_\varphi, \delta^\varphi\} < a$ where

$$\delta_\varphi := \inf\Big\{\delta\in\mathbb{R}\,\Big|\,\lim_{s\downarrow 0}\frac{s^{1+\delta}}{\varphi(s)} \text{ exists}\Big\}, \quad \delta^\varphi := \inf\Big\{\delta\in\mathbb{R}\,\Big|\,\lim_{s\uparrow\infty}\frac{s^{1+\delta}}{\varphi(s)} = \infty\Big\}.$$

It is proved in [29, Proposition 3.2] that for any $\varphi \in \mathcal{O}(2/(d+2))$ there exist constants λ_φ and c_φ such that (cf. (1.4) in the Introduction)

$$g_\varphi(v,V)(x) := \exp_\varphi(\lambda_\varphi - c_\varphi|x-v|_V^2)\Big(\det(V)\Big)^{-\frac{1}{2}},$$

where $|x|_V^2 := \langle x, V^{-1}x\rangle$, is a probability density on $\mathbb{R}^d$ with mean v and covariance matrix V, which is called a φ-exponential distribution. Note that, in the above expression, λ_φ and c_φ do not depend on the choice of V. We define the space of all φ-exponential distribution measures by

$$\mathcal{G}_\varphi := \Big\{G_\varphi(v,V) := g_\varphi(v,V)\mathcal{L}^d \Big| (v,V)\in\mathbb{R}^d\times\mathbb{S}(d,\mathbb{R})_+\Big\}.$$

Above $\mathcal{L}^d$ is the Lesbesgue measure on $\mathbb{R}^d$. Two important cases:

(i) $\varphi = s$, $\mathcal{G}_\varphi$ reduces to the class of Gaussian measures.

(ii) In the case $\varphi = s^q$, $\mathcal{G}_\varphi$ becomes the class of all q-Gaussian measures

$$\mathcal{G}_q = \Big\{G_q(v,V)\Big|(v,V)\in\mathbb{R}^d\times\mathbb{S}(d,\mathbb{R})_+\Big\}$$

where

$$G_q(v,V) = C_0(q,d)(\det V)^{-\frac{1}{2}}\exp_q\Big(-\frac{1}{2}C_1(q,d)\langle x-v, V^{-1}(x-v)\rangle\Big)\mathcal{L}^d,$$

and $C_0(q,d)$, $C_1(q,d)$ are given by

$$C_1(q,d) = \frac{2}{2+(d+2)(1-q)},$$
$$C_0(q,d) = \begin{cases} \dfrac{\Gamma\Big(\frac{2-q}{1-q}+\frac{d}{2}\Big)}{\Gamma\Big(\frac{2-q}{1-q}\Big)}\Big(\frac{(1-q)C_1(q,d)}{2\pi}\Big)^{\frac{d}{2}} & \text{if } 0<q<1,\\[2ex] \dfrac{\Gamma\Big(\frac{1}{q-1}\Big)}{\Gamma\Big(\frac{1}{q-1}-\frac{d}{2}\Big)}\Big(\frac{(q-1)C_1(q,d)}{2\pi}\Big)^{\frac{d}{2}} & \text{if } 1<q<\frac{d+4}{d+2}. \end{cases}$$

Note that $C_1(1,d) = 1$ and $C_0(q,d) \to (2\pi)^{-d/2}$ as $q \to 1$, which follows from Stirling's formula. Thus Gaussian measures are special cases of q-Gaussian measures.

The φ-exponential measures play an important role in statistical physics, information geometry and in the analysis of nonlinear diffusion equations [26–29]. We refer to [27,28,38] for further details on q-Gaussian measures, φ-exponential measures and their properties.

References

[1] M. Knott, C.S. Smith, On the optimal mapping of distributions, J. Optim. Theory Appl. 43 (1) (1984) 39–49.
[2] M. Agueh, G. Carlier, Barycenters in the wasserstein space, SIAM J. Math. Anal. 43 (2) (2011) 904–924.
[3] Y.-H. Kim, B. Pass, Wasserstein barycenters over Riemannian manifolds, Adv. Math. 307 (2017) 640–683.
[4] E. Anderes, S. Borgwardt, J. Miller, Discrete wasserstein barycenters: optimal transport for discrete data, Math. Methods Oper. Res. 84 (2) (2016) 389–409.
[5] B. Pass, On the local structure of optimal measures in the multi-marginal optimal transportation problem, Calc. Var. Partial Differential Equations 43 (3–4) (2012) 529–536.
[6] Y.-H. Kim, B. Pass, Multi-marginal optimal transport on Riemannian manifolds, Amer. J. Math. 137 (4) (2015) 1045–1060.
[7] M. Cuturi, A. Doucet, Fast computation of wasserstein barycenters, in: E.P. Xing, T. Jebara (Eds.), Proceedings of the 31st International Conference on Machine Learning, 32 of Proceedings of Machine Learning Research, PMLR, Bejing, China, 2014, pp. 685–693.
[8] P.C. Álvarez Esteban, E. del Barrio, J.A. Cuesta-Albertos, C. Matrán, A fixed-point approach to barycenters in wasserstein space, J. Math. Anal. Appl. 441 (2) (2016) 744–762.
[9] S. Kum, S. Yun, Gradient projection methods for the n-coupling problem, J. Korean Math. Soc. 56 (4) (2019) 1001–1016.
[10] G. Peyré, M. Cuturi, Computational optimal transport, Foundations and Trends® in Machine Learning 11 (5–6) (2019) 355–607.
[11] J. Rabin, G. Peyré, J. Delon, M. Bernot, Wasserstein barycenter and its application to texture mixing, in: Proceedings of the Third International Conference on Scale Space and Variational Methods in Computer Vision, in: SSVM'11, Springer-Verlag, Berlin, Heidelberg, 2012, pp. 435–446.
[12] A. Mallasto, A. Feragen, Learning from uncertain curves: The 2-wasserstein metric for Gaussian processes, in: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems 30, Curran Associates, Inc. 2017, pp. 5660–5670.
[13] S. Srivastava, C. Li, D.B. Dunson, Scalable Bayes via barycenter in wasserstein space, J. Mach. Learn. Res. 19 (1) (2018) 312–346.
[14] J. Bigot, E. Cazelles, N. Papadakis, Penalization of barycenters in the wasserstein space, SIAM J. Math. Anal. 51 (3) (2019) 2261–2285.
[15] G. Carlier, K. Eichinger, A. Kroshnin, Entropic-wasserstein barycenters: PDE characterization, regularity, and CLT, SIAM J. Math. Anal. 53 (5) (2021) 5880–5914.
[16] J. Bigot, E. Cazelles, N. Papadakis, Data-Driven Regularization of Wasserstein Barycenters with an Application to Multivariate Density Registration, Vol. 8, (4) 2019, pp. 719–755.
[17] I. Redko, A. Habrard, M. Sebban, On the analysis of adaptability in multi-source domain adaptation, Mach. Learn. 108 (8–9) (2019) 1635–1652.
[18] R. Jordan, D. Kinderlehrer, F. Otto, The variational formulation of the Fokker–Planck equation, SIAM J. Math. Anal. 29 (1) (1998) 1–17.
[19] L. Ambrosio, N. Gigli, G. Savaré, Gradient flows in metric spaces and in the space of probability measures, second ed., in: Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008.
[20] G. Carlier, V. Duval, G. Peyré, B. Schmitzer, Convergence of entropic schemes for optimal transport and gradient flows, SIAM J. Math. Anal. 49 (2) (2017) 1385–1418.
[21] M.H. Duong, B. Jin, Wasserstein gradient flow formulation of the time-fractional Fokker-Planck equation, Commun. Math. Sci. 18 (7) (2020) 1949–1975.
[22] R. Bhatia, T. Jain, Y. Lim, On the Bures–Wasserstein distance between positive definite matrices, Expo. Math. 37 (2) (2018) 165–191.
[23] H. Janati, B. Muzellec, G. Peyré, M. Cuturi, Entropic optimal transport between unbalanced Gaussian measures has a closed form, in: Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems, (3) 2020.
[24] A. Mallasto, A. Gerolin, Q.M. Ha, Entropy-regularized 2-Wasserstein distance between Gaussian measures, Inf. Geom. (2021).
[25] J. Naudts, Generalised Thermostatistics, Springer, London, 2011.
[26] F. Otto, The geometry of dissipative evolution equations: the porous medium equation, Comm. Partial Differential Equations 26 (1–2) (2001) 101–174.
[27] A. Ohara, T. Wada, Information geometry of q-Gaussian densities and behaviors of solutions to related diffusion equations, J. Phys. A 43 (3) (2009) 035002.
[28] A. Takatsu, Wasserstein geometry of porous medium equation, Ann. L'Inst. Henri Poincaré (C) Non Linear Anal. 29 (2) (2012) 217–232.
[29] A. Takatsu, Behaviors of φ-exponential distributions in Wasserstein geometry and an evolution equation, SIAM J. Math. Anal. 45 (4) (2013) 2546–2556.
[30] J. Altschuler, S. Chewi, P.R. Gerber, A. Stromme, Averaging on the Bures-Wasserstein manifold: dimension-free convergence of gradient descent, in: M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, J.W. Vaughan (Eds.), Advances in Neural Information Processing Systems, Vol. 34, Curran Associates, Inc. 2021, pp. 22132–22145.
[31] S. Chewi, T. Maunu, P. Rigollet, A.J. Stromme, Gradient descent algorithms for Bures-Wasserstein barycenters, in: J. Abernethy, S. Agarwal (Eds.), Proceedings of Thirty Third Conference on Learning Theory, 125 of Proceedings of Machine Learning Research, PMLR, 2020, pp. 1276–1304.
[32] Y. Zemel, V.M. Panaretos, Fré, Bernoulli 25 (2) (2019) 932–976.
[33] Y. Brenier, Polar factorization and monotone rearrangement of vector-valued functions, Comm. Pure Appl. Math. 44 (4) (1991) 375–417.
[34] W. Gangbo, R.J. McCann, The geometry of optimal transportation, Acta Math. 177 (2) (1996) 113–161.
[35] C. Villani, Topics in optimal transportation, in: Graduate Studies in Mathematics, American Mathematical Society, 2003.
[36] C.R. Givens, R.M. Shortt, A class of wasserstein metrics for probability distributions, Michigan Math. J. 31 (2) (1984) 231–240.
[37] T.M. Cover, J.A. Thomas, Elements of Information Theory, Wiley-Interscience, New York, NY, USA, 1991.
[38] M.H. Duong, Asymptotic equivalence of the discrete variational functional and a rate-large-deviation-like functional in the Wasserstein gradient flow of the porous medium equation, Asymptot. Anal. 92 (1–2) (2015) 85–106.
[39] C. Villani, Optimal transport: Old and new, in: Grundlehren der mathematischen Wissenschaften, Springer Berlin Heidelberg, 2008.
[40] P. Lax, Linear algebra and its applications, in: Pure and Applied Mathematics: A Wiley Series of Texts, in: Monographs and Tracts, Wiley, 2007.
[41] D. Bertsekas, Nonlinear Programming, Athena Scientific, 1999.