J. Math. Anal. Appl. 441 (2016) 744–762

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

www.elsevier.com/locate/jmaa

A fixed-point approach to barycenters in Wasserstein space ☆

Pedro C. Álvarez-Esteban [a], E. del Barrio [a], J.A. Cuesta-Albertos [b], C. Matrán [a,*]

[a] *Departamento de Estadística e Investigación Operativa and IMUVA, Universidad de Valladolid, Spain*
[b] *Departamento de Matemáticas, Estadística y Computación, Universidad de Cantabria, Spain*

A R T I C L E   I N F O

Article history:
Received 12 January 2016
Available online 20 April 2016
Submitted by B. Bongiorno

Keywords:
Mass transportation problem
L_2-Wasserstein distance
Wasserstein barycenter
Fréchet mean
Fixed-point iteration
Location-scatter families

A B S T R A C T

Let $\mathcal{P}_{2,ac}$ be the set of Borel probabilities on $\mathbb{R}^d$ with finite second moment and absolutely continuous with respect to Lebesgue measure. We consider the problem of finding the barycenter (or Fréchet mean) of a finite set of probabilities $\nu_1, \ldots, \nu_k \in \mathcal{P}_{2,ac}$ with respect to the L_2-Wasserstein metric. For this task we introduce an operator on $\mathcal{P}_{2,ac}$ related to the optimal transport maps pushing forward any $\mu \in \mathcal{P}_{2,ac}$ to $\nu_1, \ldots, \nu_k$. Under very general conditions we prove that the barycenter must be a fixed point for this operator and introduce an iterative procedure which consistently approximates the barycenter. The procedure allows effective computation of barycenters in any location-scatter family, including the Gaussian case. In such cases the barycenter must belong to the family, thus it is characterized by its mean and covariance matrix. While its mean is just the weighted mean of the means of the probabilities, the covariance matrix is characterized in terms of their covariance matrices $\Sigma_1, \ldots, \Sigma_k$ through a nonlinear matrix equation. The performance of the iterative procedure in this case is illustrated through numerical simulations, which show fast convergence towards the barycenter.

1. Introduction

Let us consider a set $\{x_1, \ldots, x_k\}$ of elements in a certain space and associated weights, $\lambda_1, \ldots, \lambda_k$, satisfying $\lambda_i > 0$, $\sum_{i=1}^n \lambda_i = 1$, interpretable as a quantification of the relative importance or presence of these elements. The suitable choice of an element in the space to represent that set is an old problem present in many different settings. The weighted mean being the best known choice, it enjoys many nice properties that allow us to consider it a very good representation of elements in an Euclidean space. Yet, it can be highly undesirable for representing shaped objects like functions or matrices with some particular structure. The Fréchet mean or barycenter is a natural extension arising from the consideration of minimum dispersion character of the mean, when the space has a metric structure which, in some cases, may overcome these

☆ Research partially supported by the Spanish Ministerio de Economía y Competitividad, grants MTM2014-56235-C2-1-P and MTM2014-56235-C2-2, and by Consejería de Educación de la Junta de Castilla y León, grant VA212U13.

* Corresponding author.
E-mail address: matran@eio.uva.es (C. Matrán).

http://dx.doi.org/10.1016/j.jmaa.2016.04.045
0022-247X/

difficulties. Like the mean, if d is a distance in the reference space E, a barycenter $\bar{x}$ is determined by the relation

$$\sum_{i=1}^{k} \lambda_i d^2(x_i, \bar{x}) = \min \left\{ \sum_{i=1}^{k} \lambda_i d^2(x_i, y),\ y \in E \right\}.$$

In the last years Wasserstein spaces have focused the interest of researchers from very different fields (see, e.g., the monographs [4,27] or [28]), leading in particular to the natural consideration of Wasserstein barycenters beginning with [1]. This appealing concept shows a high potential for application, already considered in Artificial Intelligence or in Statistics (see, e.g., [16,5,10,20,7] or [3]). The main drawback being the difficulties for its effective computation, several of these papers [16,5,10] are mainly devoted to this hard goal. In fact, the (L_2-)Wasserstein distance between probabilities, which is the framework of this paper, is easily characterized and computed for probabilities on the real line, but there is not a similar, simple closed expression for its computation in higher dimension. A notable exception arises from Gelbrich's bound and some extensions (see [19,13,15]) that allow the computation of the distances between normal distributions or between probabilities in some parametric (location-scatter) families. For multivariate Gaussian distributions in particular, in [1], the barycenter has been characterized in terms of a fixed point equation involving the covariance matrices in a nonlinear way (see (6) below) but, to the best of our knowledge, feasible consistent algorithms to approximate the solution have not been proposed yet.

The approach in [1] for the characterization of the barycenter in the Gaussian setting resorted to duality arguments and to Brouwer's fixed-point theorem. Here we take a different approach which is not constrained to the Gaussian setup. We introduce an operator associated to the transformation of a probability measure through weighted averages of optimal transportations to the set $\{\nu_i\}_{i=1}^k$ of target distributions. This operator is the real core of our approach. We show (see Theorem 3.1 and Proposition 3.3 below) that, in very general situations, barycenters must be fixed points of the above mentioned operator. We also show (Theorem 3.6) how this operator can be used to define a consistent iterative scheme for the approximate computation of barycenters. Of course, the practical usefulness of the iteration will depend on the difficulties arising from the computation of the optimal transportation maps involved in the iteration. The case of Gaussian probabilities is a particularly convenient setup for our iteration. We provide a self-contained approach to barycenters in this Gaussian framework based on first principles of optimal transportation and some elementary matrix analysis. This leads also to the characterization (6) and, furthermore, it yields sharp bounds on the covariance matrix of the barycenter which are of independent interest. We prove (Theorem 4.2) that our iteration provides a consistent approximation to barycenters in this Gaussian setup. We also notice that all the results given for the Gaussian family are automatically extended to location-scatter families (see Definition 2.1 below). Finally, we illustrate the performance of the iteration through numerical simulations. These show fast convergence towards the barycenter, even in problems involving a large number of distributions or high-dimensional spaces.

The remaining sections of this paper are organized as follows. Section 2 gives a succinct account of some basic facts about optimal transportation and Wasserstein metrics and introduces the barycenter problem with respect to these metrics. The section contains pointers to the most relevant references on the topic. Section 3 contains the core of the paper, introducing the operator G in (7), showing the connection between barycenters and fixed points of G and presenting the iterative scheme for approximate computation of barycenters. The Gaussian and location-scatter cases are analyzed in Section 4, while Section 5 presents some numerical simulations. We conclude this Introduction with some words on notation. Throughout the paper our space of reference is the Euclidean space $\mathbb{R}^d$. With $\|x\|$ we denote the usual norm and with $x \cdot y$ the inner product. For a matrix A, A^t will denote the corresponding transpose matrix, $\det(A)$ the determinant and $\mathrm{Tr}(A)$ the trace. Id will be indistinctly used as the identity map and as the $d \times d$ identity matrix, while $\mathcal{M}_{d\times d}^{+}$ will denote the set of $d \times d$ (symmetric) positive definite matrices. The space where we consider