Measure Estimation in the Barycentric Coding Model

Matthew Werenski, Ruijie Jiang,
Abiy Tasissa, Shuchin Aeron, James M. Murphy

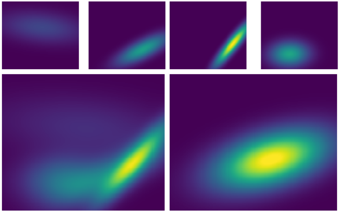
Tufts University

ICML 2022

The Barycentric Coding Model

Given p measures $\mu_1, \dots, \mu_p$
and a coordinate λ with $\lambda_i \geq 0, \sum_i \lambda_i = 1$
the Wasserstein barycenter is defined

$$\nu_\lambda = \text{argmin}_\nu \sum_{i=1}^{p} \lambda_i W_2^2(\mu_i, \nu).$$



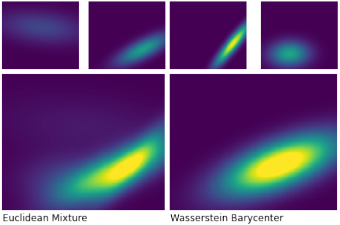
The barycentric coding model
is the set of all minimizers as λ varies:

$$\text{Bary}(\{\mu_i\}_{i=1}^{p}) = \left\{ \nu_\lambda : \lambda_i \geq 0, \sum_i \lambda_i = 1 \right\}.$$

The Barycentric Coding Model

Given p measures $\mu_1, \ldots, \mu_p$
and a coordinate λ with $\lambda_i \geq 0, \sum_i \lambda_i = 1$
the Wasserstein barycenter is defined

$$\nu_\lambda = \text{argmin}_\nu \sum_{i=1}^{p} \lambda_i W_2^2(\mu_i, \nu).$$



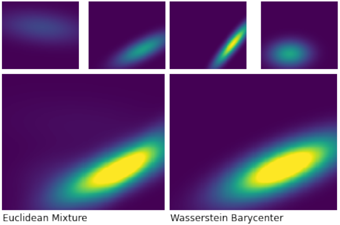
The barycentric coding model
is the set of all minimizers as λ varies:

$$\text{Bary}(\{\mu_i\}_{i=1}^p) = \left\{ \nu_\lambda : \lambda_i \geq 0, \sum_i \lambda_i = 1 \right\}.$$

The Barycentric Coding Model

Given p measures $\mu_1, \ldots, \mu_p$
and a coordinate λ with $\lambda_i \geq 0, \sum_i \lambda_i = 1$
the Wasserstein barycenter is defined

$$\nu_\lambda = \text{argmin}_\nu \sum_{i=1}^{p} \lambda_i W_2^2(\mu_i, \nu).$$



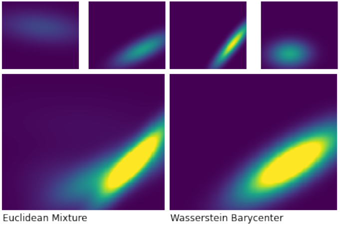
The barycentric coding model
is the set of all minimizers as λ varies:

$$\text{Bary}(\{\mu_i\}_{i=1}^{p}) = \left\{ \nu_\lambda : \lambda_i \geq 0, \sum_i \lambda_i = 1 \right\}.$$

The Barycentric Coding Model

Given p measures $\mu_1, \ldots, \mu_p$
and a coordinate λ with $\lambda_i \geq 0, \sum_i \lambda_i = 1$
the Wasserstein barycenter is defined

$$\nu_\lambda = \text{argmin}_\nu \sum_{i=1}^{p} \lambda_i W_2^2(\mu_i, \nu).$$



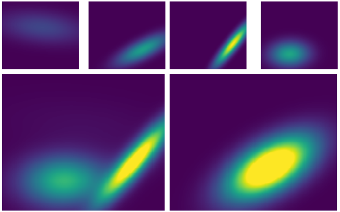
The barycentric coding model
is the set of all minimizers as λ varies:

$$\text{Bary}(\{\mu_i\}_{i=1}^p) = \left\{ \nu_\lambda : \lambda_i \geq 0, \sum_i \lambda_i = 1 \right\}.$$

The Barycentric Coding Model

Given p measures $\mu_1, ..., \mu_p$
and a coordinate λ with $\lambda_i \geq 0, \sum_i \lambda_i = 1$
the Wasserstein barycenter is defined

$$\nu_\lambda = \text{argmin}_\nu \sum_{i=1}^{p} \lambda_i W_2^2(\mu_i, \nu).$$



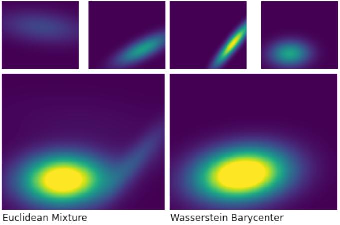
The barycentric coding model
is the set of all minimizers as λ varies:

$$\text{Bary}(\{\mu_i\}_{i=1}^p) = \left\{ \nu_\lambda : \lambda_i \geq 0, \sum_i \lambda_i = 1 \right\}.$$

The Barycentric Coding Model

Given p measures $\mu_1, ..., \mu_p$
and a coordinate λ with $\lambda_i \geq 0, \sum_i \lambda_i = 1$
the Wasserstein barycenter is defined

$$\nu_\lambda = \mathrm{argmin}_\nu \sum_{i=1}^{p} \lambda_i W_2^2(\mu_i, \nu).$$



The barycentric coding model
is the set of all minimizers as λ varies:

$$\mathrm{Bary}(\{\mu_i\}_{i=1}^p) = \left\{ \nu_\lambda : \lambda_i \geq 0, \sum_i \lambda_i = 1 \right\}.$$

Measure Estimation

- Given a new measure μ_0, observed through a noisy $\tilde{\mu}_0$.
- Find the "best approximation" to $\tilde{\mu}_0$ by a member of $\text{Bary}(\{\mu_i\}_{i=1}^p)$.
- Use this measure as an estimate $\hat{\mu}_0 = \nu_\lambda$.

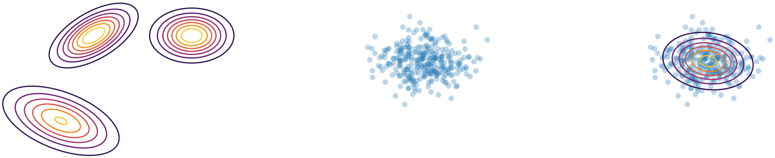


Figure: Left: References $\{\mu_i\}_{i=1}^3$, Center: Noisy observation $\tilde{\mu}_0$, Right: Estimated $\hat{\mu}_0 = \nu_\lambda$

Objective

We would like to solve

$$\min_{\lambda \in \mathbb{R}_+ : \sum_i \lambda_i = 1} W_2^2(\nu_\lambda, \tilde{\mu}_0) \tag{1}$$

which requires solving

$$\min_{\lambda} W_2^2 \left(\operatorname{argmin}_\nu \sum_{i=1}^{p} \lambda_i W_2^2(\mu_i, \nu), \tilde{\mu}_0 \right) \tag{2}$$

Difficult nested optimization problem.

Alternative Characterization

A measure μ_0 is a Karcher mean of $\{\mu_i\}_{i=1}^p$ for coordinate λ if it satisfies

$$\int \left\langle \sum_{i=1}^{p} \lambda_i (T_i(x) - x), \sum_{j=1}^{p} \lambda_j (T_j(x) - x) \right\rangle d\mu_0(x) = 0, \qquad (3)$$

where T_i is the optimal map from μ_0 to μ_i. Under technical conditions barycenters are **equivalent** to Karcher means. Introducing the matrix $A \in \mathbb{R}^{p \times p}$ with entries

$$A_{ij} = \int \langle T_i(x) - x, T_j(x) - x \rangle \, d\mu_0(x).$$

we can write $(3) = \lambda^T A \lambda$.

First Questions

Given a set of reference measures $\mu_1, ..., \mu_p$ and a new measure μ_0, can we

1. Determine if $\mu_0 \in \text{Bary}(\{\mu_i\}_{i=1}^p)$ or not?
2. If so, can we recover the coordinate λ?

First Questions

Given a set of reference measures $\mu_1, \ldots, \mu_p$ and a new measure μ_0, can we

1. Determine if $\mu_0 \in \text{Bary}(\{\mu_i\}_{i=1}^p)$ or not?
2. If so, can we recover the coordinate λ?

Theorem 1

Under suitable conditions, then the value of

$$\min_{\lambda \in \mathbb{R}_+ : \sum_i \lambda_i = 1} \lambda^T A \lambda$$

is 0 if and only if $\mu_0 \in \text{Bary}(\{\mu_i\}_{i=1}^p)$. If $\mu_0 \in \text{Bary}(\{\mu_i\}_{i=1}^p)$ then $\mu_0 = \nu_{\lambda_*}$ where λ_* is the minimizer.

Working with Random Samples

What if we can only draw i.i.d. samples from each $\mu_0, ..., \mu_p$?

Corollary 2

Under the appropriate technical conditions

$$\mathbb{E}[\|\hat{\lambda} - \lambda_*\|_2^2] \lesssim \frac{1}{\sqrt{n}} + n^{-\frac{\alpha+1}{4(d'+\alpha+1)}} \sqrt{\log n}$$

where $\hat{\lambda}$ is the coordinate estimated using samples and λ_* is the true minimizer and n is the number of samples used in the estimate.

Three step procedure

1. Using Entropically regularized OT to estimate $\{T_i\}_{i=1}^p$.
2. Using the estimated maps, estimate the entries A_{ij}.
3. Solve a convex QP.

The most expensive part is estimating $\{T_i\}_{i=1}^p$ - still cheap using Sinkhorn iterations.

Applications

We apply our approach in three settings

1. Covariance Estimation
2. Document Topic Classification
3. Image Denoising