Large Deviations Principle for Bures-Wasserstein Barycenters

Adam Quinn Jaffe

With Leonardo Santoro

I. The Bures-Wasserstein Space
II. Bures-Wasserstein Barycenters
III. The Large Deviations Principle
IV. Concentration of Measure
V. Future Work

I. The Bures-Wasserstein Space

Let $\mathbb{K}$ denote the set of real, symmetric, matrices in a fixed dimension.

Many statistical problems feature data or parameters living in $\mathbb{K}$:

- Covariance estimation
- ANOVA
- Diffusion tensor imaging
- Quantum information theory
- Coase grain DNA modeling

It is usually *not* natural to endow $\mathbb{K}$ with the geometry it inherits from the ambient Euclidean space.

Instead, there are many metrics on $\mathbb{K}$ with a more interesting geometry. One can set the distance between Σ and Σ' to be:

- *ℓ_p norm:* $\|\Sigma - \Sigma'\|_p$
- *$\ell_p \to \ell_q$ operator norm:* $\|\Sigma - \Sigma'\|_{p \to q}$
- *log-Euclidean metric:* $\|\log(\Sigma) - \log(\Sigma')\|_2$
- *log-Riemannian metric:* $\|\log(\Sigma^{-1/2}\Sigma'\Sigma^{-1/2})\|_2$
- *Stein's loss:* $\mathrm{tr}(\Sigma'\Sigma^{-1}) - \log\det(\Sigma'\Sigma^{-1}) - m$
- *Bures-Wasserstein metric...*

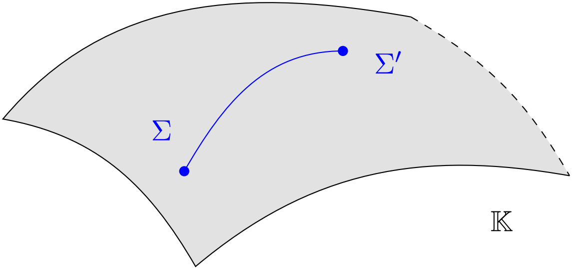
Also many more possiblilities (Dryden-Kolyodenko-Zhou 2009, Pigoli-Aston-Dryden-Secchi 2014).

The *Bures-Wasserstein metric* is a metric Π on $\mathbb{K}$ has many equivalent formulations, and has been independently studied in several application areas (Masarotto-Panaretos-Zemel 2020):

$$\begin{aligned}\Pi(\Sigma, \Sigma') &:= \sqrt{\mathrm{tr}(\Sigma) + \mathrm{tr}(\Sigma') - 2\mathrm{tr}\left(\left(\Sigma^{1/2}\Sigma'\Sigma^{1/2}\right)^{1/2}\right)} \\ &= \min_{U^\top U = I} \|U\Sigma^{1/2} - (\Sigma')^{1/2}\|_2 \\ &= W_2(\mathcal{N}(0,\Sigma), \mathcal{N}(0,\Sigma'))\end{aligned}$$

The *Bures-Wasserstein space* $(\mathbb{K}, \Pi)$ has a rich geometry:

- It is uniquely geodesic,
- It has non-negative (Alexandrov) curvature,
- It is a "stratified space",



II. Bures-Wasserstein Barycenters

For a sufficiently integrable probability measure $P \in \mathcal{P}(\mathbb{K})$, a *Bures-Wasserstein barycenter* is a solution of the optimization problem

$$\begin{cases} \text{minimize} & \int_{\mathbb{K}} \Pi^2(M, \Sigma) \mathrm{d}P(\Sigma) \\ \text{over} & M \in \mathbb{K}. \end{cases}$$

Barycenters are also called *Fréchet means* or *centers of mass*, and they represent a canonical notion of central tendency in $(\mathbb{K}, \Pi)$.

Recall that in the Euclidean space $(\mathbb{R}^d, \|\cdot\|_2)$ when $P \in \mathcal{P}(\mathbb{R}^d)$ has finite variance, the optimization problem

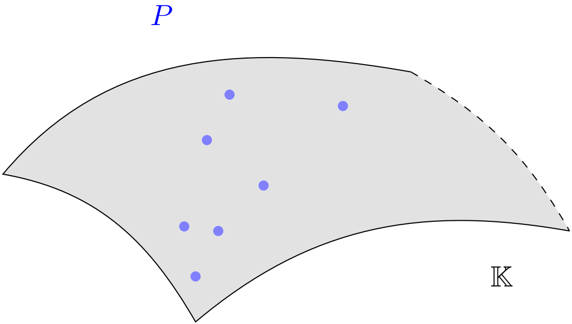
$$\begin{cases} \text{minimize} & \int_{\mathbb{R}^d} \|m - s\|^2 \mathrm{d}P(s) \\ \text{over} & m \in \mathbb{R}^d. \end{cases}$$

is uniquely solved at the mean $m = \int_{\mathbb{R}^d} s \, \mathrm{d}P(s)$.

It is known (Knott-Smith 1984, Agueh-Carlier 2011) that $M \succ 0$ is a Bures-Wasserstein barycenter of P if and only if it satisfies

$$\int_{\mathbb{K}} \left(M^{1/2} \Sigma M^{1/2}\right)^{1/2} \mathrm{d}P(\Sigma) = M,$$

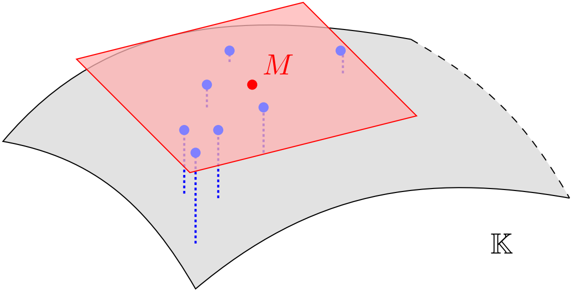
called the *fixed-point equation*.



It is known (Knott-Smith 1984, Agueh-Carlier 2011) that $M \succ 0$ is a Bures-Wasserstein barycenter of P if and only if it satisfies

$$\int_{\mathbb{K}} \left(M^{1/2} \Sigma M^{1/2}\right)^{1/2} \mathrm{d}P(\Sigma) = M,$$

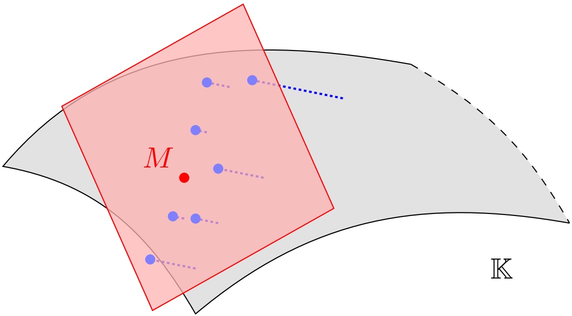
called the *fixed-point equation*.



It is known (Knott-Smith 1984, Agueh-Carlier 2011) that $M \succ 0$ is a Bures-Wasserstein barycenter of P if and only if it satisfies

$$\int_{\mathbb{K}} \left(M^{1/2}\Sigma M^{1/2}\right)^{1/2} \mathrm{d}P(\Sigma) = M,$$

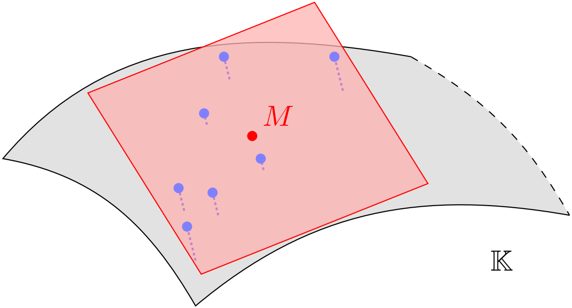
called the *fixed-point equation*.



It is known (Knott-Smith 1984, Agueh-Carlier 2011) that $M \succ 0$ is a Bures-Wasserstein barycenter of P if and only if it satisfies

$$\int_{\mathbb{K}} \left(M^{1/2}\Sigma M^{1/2}\right)^{1/2} \mathrm{d}P(\Sigma) = M,$$

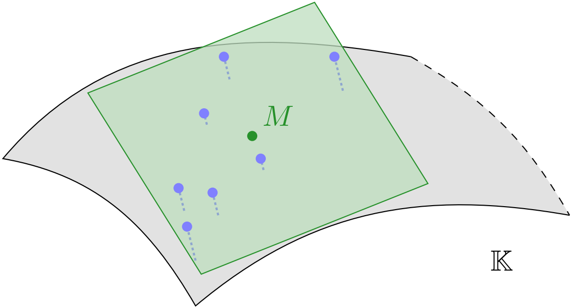
called the *fixed-point equation*.



It is known (Knott-Smith 1984, Agueh-Carlier 2011) that $M \succ 0$ is a Bures-Wasserstein barycenter of P if and only if it satisfies

$$\int_{\mathbb{K}} \left(M^{1/2} \Sigma M^{1/2} \right)^{1/2} \mathrm{d}P(\Sigma) = M,$$

called the *fixed-point equation*.



Let $P \in \mathcal{P}(\mathbb{K})$ be an unknown sufficiently integrable probability measure, and let $\Sigma_1, \Sigma_2, \dots$ be independent, identically distributed (IID) samples from P.

Write M_n^* for the *empirical Bures-Wasserstein barycenter*

$$M_n^* := \arg\min_{M \in \mathbb{K}} \frac{1}{n} \sum_{i=1}^{n} \Pi^2(M, \Sigma_i)$$

and M^* for the *population Bures-Wasserstein barycenter*

$$M^* := \arg\min_{M \in \mathbb{K}} \int_{\mathbb{K}} \Pi^2(M, \Sigma) \mathrm{d}P(\Sigma).$$

How well does M_n^* approximate M^*?

Much is known:

- **SLLN:** $\Pi(M_n^*, M^*) \to 0$ almost surely as $n \to \infty$, from general theory of optimal transport (Le Gouic-Loubes 2017) or general theory of Fréchet means (Evans-Jaffe 2024).
- **CLT:** $\sqrt{n}\left((M^*)^{1/2} M_n^* (M^*)^{1/2}\right)^{1/2} \to \mathcal{N}(M^*, G)$ in distribution as $n \to \infty$ for some G depending on P (Agueh-Carlier 2017).
- **rate of convergence:** $\mathbb{E}\left[\Pi^2(M_n^*, M^*)\right] \lesssim \sigma^2 n^{-1}$ for some σ^2 depending on P (Le Gouic-Paris-Rigollet-Stromme 2023).
- **concentration:** For some $c_1, c_2 > 0$ depending on P,

$$\mathbb{P}\Big(\Pi(M_n^*, M^*) \geq t\Big) \leq c_1 e^{-c_2 n t^2}$$

for all $t \geq 0$ (Kroshnin-Spokoiny-Suvorikova 2021).

III. The Large Deviations Principle

For $P \in \mathcal{P}(\mathbb{K})$ and $M \in \mathbb{K}$, set

$$I_P(M) := \sup_{A \in \mathbb{S}} \left(\mathrm{tr}(AM) - \log \int_{\mathbb{K}} \exp \mathrm{tr} \left(AM\Sigma^{1/2}(\Sigma^{1/2}M\Sigma^{1/2})^{-1/2}\Sigma^{1/2} \right) \mathrm{d}P(\Sigma) \right)$$

where $\mathbb{S}$ denotes the set of all real, symmetric matrices.

Note this is sort of like a Fenchel-Legendre transform, but not quite.

Theorem (AQJ-Santoro, 2024+)

The function $I_P : \mathbb{K} \to [0,\infty]$ enjoys the following properties:

(a) $I_P(M) = 0$ *if and only if* $M = M^*$.

(b) I_P *is lower semi-continuous.*

(c) I_P *is coercive and satisfies* $I_P(M)/\Pi(M,0) \to \infty$ *as* $M \to \infty$.

(d) I_P *is convex along certain continuous paths in* $(\mathbb{K}, \Pi)$.

Theorem (AQJ-Santoro, 2024+)

For all Borel measurable $E \subseteq \mathbb{K}$, we have

$$\begin{aligned} -\inf\{I_P(M) : M \in E^\circ\} \\ &\le \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(M_n^* \in E) \\ &\le \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(M_n^* \in E) \\ &\le -\inf\left\{I_P(M) : M \in \bar{E}\right\}, \end{aligned}$$

where E° and $\bar{E}$ denote the interior and closure of E with respect to Π.

If $E \subseteq \mathbb{K}$ is equal to the closure of its interior, then

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(M_n^* \in E) = -\inf \{I_P(M) : M \in E\},$$

which is like a sort of duality between probability theory and optimization theory. Roughly speaking, the above can be interpreted as

$$\mathbb{P}(M_n^* \in E) \approx \exp(-n \cdot \inf \{I_P(M) : M \in E\}),$$

after ignoring sub-exponential factors.

The core of the proofs is the following duality:

Lemma (AQJ-Santoro, 2024+)

For each $M \in \mathbb{K}$, the optimization problems

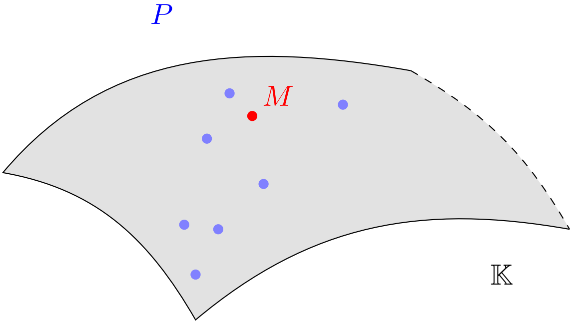
$$\begin{cases} \text{maximize} & \operatorname{tr}(AM) - \log \int_{\mathbb{K}} \exp \operatorname{tr}\left(AM\Sigma^{1/2}(\Sigma^{1/2}M\Sigma^{1/2})^{-1/2}\Sigma^{1/2}\right) \mathrm{d}P(\Sigma) \\ \text{over} & A \in \mathbb{S} \end{cases}$$

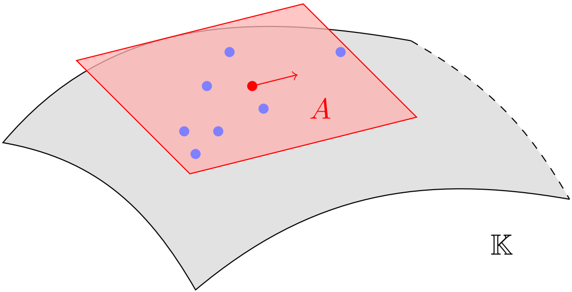
and

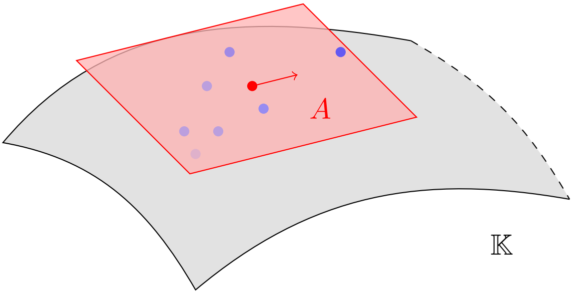
$$\begin{cases} \text{minimize} & H(Q \,|\, P) \\ \text{over} & Q \in \mathcal{P}_2(\mathbb{K}) \\ \text{where} & Q \text{ has barycenter } M \end{cases}$$

have the same value and admit at most one optimizer; furthermore, feasible points $A \in \mathbb{S}$ and $Q \in \mathcal{P}_2(\mathbb{K})$ are optimal if and only if they satisfy $Q = P^{M \to A}$, where

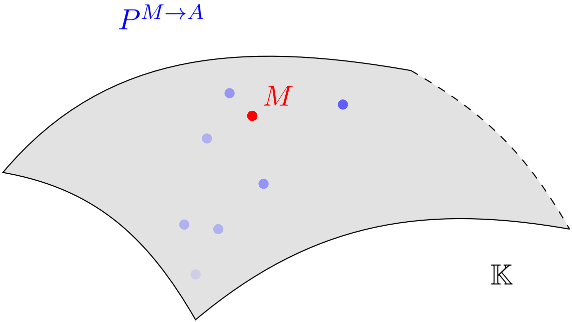
$$\frac{\mathrm{d}P^{M\to A}}{\mathrm{d}P}(\Sigma) \propto \exp\left(AM\Sigma^{1/2}(\Sigma^{1/2}M\Sigma^{1/2})^{-1/2}\Sigma^{1/2}\right). \tag{1}$$


P
M
𝕂


A
𝕂



A
𝕂


$P^{M \to A}$
M
$\mathbb{K}$

Wasserstein barycenters? For $P \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$, let $\nu_1, \nu_2, \ldots$ be independent, identically-distributed from P, and consider the empirical and population Wasserstein barycenters μ_n^* and μ^*. Many limit theorems for $\mu_n^* \to \mu^*$ in W_2 are known (SLLN, rates of convergence, etc.).

Formally, we expect that $\{\mu_n^*\}_{n\in\mathbb{N}}$ satisfies a large deviations principle in $(\mathcal{P}(\mathbb{R}^d), W_2)$ with good rate function

$$I_P(\mu) = \sup_{\phi \in L^2(\mu)} \left(\langle \phi, \mathrm{id} \rangle_{L^2(\mu)} - \log \int_{\mathcal{P}(\mathbb{R}^d)} \exp \left\langle \phi, t_\mu^\nu \right\rangle_{L^2(\mu)} \mathrm{d}P(\nu) \right),$$

where $t_\mu^\nu : \mathbb{R}^d \to \mathbb{R}^d$ denotes the optimal transport map from μ to ν.

Need to use the Otto calculus (Otto 2001) to make this rigorous...

IV. Concentration of Measure

All of stated results in fact hold in infinite dimensions, meaning $\mathbb{K}$ is the space of real, symmetric, *trace-class* operators on an infinite-dimensional Hilbert space $\mathcal{H}$.

So, it is hopeful to try to use these results to study the concentration of measure phenomenon for Bures-Wasserstein barycenters, which aims to develop dimension-free concentration properties of the empirical barycenter around the population barycenter.

Suppose that if Σ is distributed according to P, then $\Pi(\Sigma, 0)$ has a σ^2-sub-Gaussian distribuion.

Corollary

For any $r \geq 0$, we have existence of the limit

$$\Phi_P(r) := -\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\Pi(M_n^*, M^*) \geq r),$$

and the function $\Phi_P : [0,\infty) \to [0,\infty]$ satisfies

$$\liminf_{r\to\infty} \frac{\Phi_P(r)}{r^2} \geq \frac{1}{2\sigma^2}. \tag{2}$$

In other words, we have $\mathbb{P}(\Pi(M_n^*, M^*) \geq r) \lesssim \exp(-\frac{r^2}{2\sigma^2})$ for large $r > 0$, where we ignore sub-exponential factors in $n \in \mathbb{N}$. Note that this is exactly the Hoeffding-style concentration, and that there is no dependence on the dimension!

Corollary

If $r > 0$*, then, we have*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\Pi(M_n^*, M^*) \geq r + \varepsilon \,\middle|\, \Pi(M_n^*, M^*) \geq r\right) < 0$$

for all $\varepsilon > 0$

In other words, if $\Pi(M_n^*, M^*) \geq r$, then $\Pi(M_n^*, M^*) \approx r$ with high probability. This means that M_n^* lies outside the ball $B_r(M^*)$ only if it lies near the boundary of $B_r(M^*)$.

V. Future Work

Applications:

- Bounding the asymptotic relative efficiency (ARE) for hypothesis tests based on Bures-Wasserstein barycenters
- Rare event simulation for Bures-Wasserstein barycenters

Extensions to other spaces:

- The Wasserstein space
- General Riemannian manifolds

Numerical optimization of I_P:

- Geodesic convexity
- Eigenvalues of $\nabla_M^2 I_P$

Thank you!

References

M. Agueh and G. Carlier. Barycenters in the Wasserstein space. *SIAM J. Math. Anal.,* 43:904-924, 2011.

M. Agueh and G. Carlier. Vers un théorème de la limite centrale dans l'espace de Wasserstein? *C. R. Math. Acad. Sci. Paris.,* 355(7):812-818, 2017.

I. L. Dryden, A. Kolyodenko, and D. Zhou. Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Ann. Appl. Statist.,* 3:1102-1123, 2009.

S. N. Evans and A. Q. Jaffe. Limit theorems for Fréchet mean sets. *Bernoulli,* 30(1):419-447, 2024.

M. Knott and C. S. Smith. On the optimal mapping of distributions. *J. Optim. Theory Appl.,* 43:39-49, 1984.

A. Kroshnin, V. Spokoiny, and A. Suvorikova. Statistical inference for Bures-Wasserstein barycenters. *Ann. Appl. Probab.,* 31:1264-1298, 2021.

T. Le Gouic and J.-M. Loubes. Existence and consistency of Wasserstein barycenters. *Probab. Theory Related Fields,* 168:901-917, 2017.

T. Le Gouic, Q. Paris, P. Rigollet, and A. J. Stromme. Fast convergence of empirical barycenters in Alexandrov spaces and the Wasserstein space. *J. Eur. Math. Soc.,* 24:2229-2250, 2023.

V. Masarotto, V. M. Panaretos, and Y. Zemel. Procrustes metrics on covariance operators and optimal transportation of Gaussian processes. *Sankhya A,* 81:172-213.

V. M. Panaretos and Y. Zemel. *Invitation to Statistics in Wasserstein Space*, Springer Briefs in Probability and Mathematical Statistics, 2020.

F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations,* 26(1-2):101-174, 2001.

D. Pigoli, J. A. Aston, I. L. Dryden, and P. Secchi. Distance and inference for covariance operators. *Biometrika,* 101:409-422, 2014.

Y. Zemel and V. M. Panaretos. Fréchet means and Procrustes analysis in the Wasserstein space. *Bernoulli,* 25:932-976, 2019.