Omitted proofs from "Concentration inequalities for barycenters in metric spaces"

Proof of Lemma 2.2 Let $f : M^{(2)} \to \mathbb{R}$ be a 1-Lipschitz function. Then, for all $\lambda \geq 0$,

$$\mathbb{E}[e^{\lambda(f(\Phi(X))-\mathbb{E}[f(\Phi(X))])}] = \mathbb{E}[e^{\lambda L \frac{(f(\Phi(X))-\mathbb{E}[f(\Phi(X))])}{L}}] = \mathbb{E}[e^{\lambda L(g(X)-\mathbb{E}[g(X)])}]$$

where $g = (1/L)f \circ \Phi$ is a 1-Lipschitz function. Hence, $\mathbb{E}[e^{\lambda(f(\Phi(X))-\mathbb{E}[f(\Phi(X))])}] \leq \Lambda_X(\lambda L)$ and one concludes by taking the supremum over all 1-Lipschitz functions $f : M^{(2)} \to \mathbb{R}$.

Proof of Lemma 2.4 By Definition 2.3, X is K^2-sub-Gaussian if and only if the real-valued random variable $f(X)$ is K^2-sub-Gaussian, for all 1-Lipschitz functions $f : M \to \mathbb{R}$, in the usual sense [2, Section 2.5]. Hence, [2, Proposition 2.5.2] applied to $f(X)$, for any 1-Lipschitz function $f : M \to \mathbb{R}$, yields Lemma 2.4.

Proof of Proposition 2.5 Let X_i be K_i-sub-Gaussian, for each $i = 1, \ldots, n$. Then, $\Lambda_{X_i}(\lambda) \leq e^{\lambda^2 K_i^2/2}$, for all $i = 1, \ldots, n$ and $\lambda \geq 0$. Therefore, by Lemma 2.1,

$$\Lambda_{(X_1,\ldots,X_n)}(\lambda) \leq \Lambda_{X_1}(\lambda) \ldots \Lambda_{X_n}(\lambda) \leq \prod_{i=1}^{n} e^{\lambda^2 K_i^2/2} = e^{\lambda^2(K_1^2+\ldots+K_n^2)/2},$$

for all $\lambda \geq 0$, which yields the result.

Proof of Theorem 3.3 Let T_n be either the inductive or the empirical barycenter of $X_1, \ldots, X_n$. By Proposition 3.1, T_n can be written as $\Phi(X_1, \ldots, X_n)$, for some $(1/n)$-Lipschitz function $\Phi : M^n \to M$, where M^n is equipped with the L^1 product metric d_1. Hence, by Lemma 2.2, for all $\lambda \geq 0$,

$$\Lambda_{T_n}(\lambda) = \Lambda_{\Phi(X_1,\ldots,X_n)}(\lambda) \leq \Lambda_{(X_1,\ldots,X_n)}(\lambda/n)$$

and Lemma 2.1 implies that the latter is bounded by $\prod_{i=1}^{n} \Lambda_{X_i}(\lambda/n)$. Therefore, since each X_i is K_i-sub-Gaussian, we obtain

$$\Lambda_{T_n}(\lambda) \leq \prod_{i=1}^{n} e^{\lambda^2 K_i^2/(2n^2)} = e^{\lambda^2 \bar{K}_n^2/2}$$

where $\bar{K}_n^2 = \dfrac{K_1^2 + \ldots + K_n^2}{n^2}$, which yields the result.

Proof of Corollary 3.5 In Corollary 3.4, K can be replaced with $2C$, by Lemma 2.7, yielding the result.

Comparison with Funano's result [1] Funano [1, Theorem 1.2] proves the following: If M is a p-dimensional Hadamard manifold (i.e., a Riemannian manifold that is complete, and has non-positive curvature), then, for all $\delta \in (0,1)$, it holds with probability at least $1-\delta$ that

$$d(\tilde{B}_n, b^*) \leq A_1 C \sqrt{\frac{p \log(A_2/\delta)}{n}}, \tag{1}$$

where $\tilde{B}_n$ is the inductive barycenter of n i.i.d. random variables in M that bounded within a ball of radius $C > 0$, b^* is the population barycenter and A_1, A_2 are positive universal constants ($4 \leq A_1 \leq 6$ and A_2 is of order $50,000$).

(1) has a few major drawbacks, compared to our result (Corollary 3.5). First, it requires M to be a Riemannian manifold, and in particular, it must have finite dimension p. Our result, on the opposite, is dimension free, but with the same dependence in n and δ, and it applies in a much larger range of NPC spaces (non just Hadamard manifolds). Moreover, even in finite dimension, Funano's bound (1) is clearly sub-optimal, e.g., in a Euclidean space, since, unlike our result, it does not decouple the dimension p from the confidence parameter δ (see our discussion under Corollary 3.5 in the paper, and also in our rebuttal). Finally, our constants are significantly better, and in Euclidean/Hilbert spaces, they are optimal.

References

[1] K. Funano. Rate of convergence of stochastic processes with values in $\mathbb{R}$-trees and Hadamard manifolds. *Osaka J. Math.*, 47(4):911–920, 2010.

[2] R. Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.