

A. Comparing the metric and the subgaussian diameters

Proof of Inequality (8). Let $\Xi = \Xi(\mathcal{X})$ be the symmetrized distance. By (5), we have $\mathbb{E}[\Xi] = 0$ and certainly $|\Xi| \leq \text{diam}(\mathcal{X})$. Hence,

$$\mathbb{E}e^{\lambda\Xi} \leq \exp((2 \text{diam}(\mathcal{X})\lambda)^2/8) = \exp(\text{diam}(\mathcal{X})^2\lambda^2/2),$$

where the inequality follows from Hoeffding's Lemma. \square

Proof of Inequality (9). Let \mathcal{X} be an N -point space with the uniform distribution and $\rho(x, x') = 1$ for all distinct $x, x' \in \mathcal{X}$. Then, for independently drawn $X, X' \in \mathcal{X}$, we have $\rho(X, X') = 0$ with probability $1/N$ and $\rho(X, X') = 1$ with probability $1 - 1/N$. Hence, $\Xi(\mathcal{X}) = \pm 1$ with probability $\frac{1}{2}(1 - 1/N)$ and 0 with probability $1/N$, and

$$\mathbb{E} \exp(\lambda\Xi(\mathcal{X})) = \frac{1}{N}e^1 + \left(\frac{1}{2} - \frac{1}{2N}\right)(e^\lambda + e^{-\lambda}),$$

which approaches $\cosh(\lambda)$ as $N \rightarrow \infty$. Since $\cosh(\lambda) \leq e^{\lambda^2/2}$, taking N sufficiently large makes $\Delta_{\text{SG}}(\mathcal{X})$ arbitrarily close to $\text{diam}(\mathcal{X}) = 1$. \square

A referee has suggested achieving $\text{diam}(\mathcal{X}) = \Delta_{\text{SG}}(\mathcal{X})$ by considering the same metric as above on a continuous space with a continuous distribution. This indeed works, but the induced triple (\mathcal{X}, ρ, μ) is not a metric probability space since the Borel σ -algebra is not compatible with the metric. We do not know whether $\text{diam}(\mathcal{X}) = \Delta_{\text{SG}}(\mathcal{X})$ is possible in a metric probability space.

B. The equivalence of (1) and (16)

Suppose that $f : \mathcal{X}^n \rightarrow \mathbb{R}$ satisfies the Lipschitz condition (2). This is equivalent to saying that f is 1-Lipschitz with respect to the weighted Hamming metric

$$\rho_w(x, x') = \sum_{i=1}^n w_i \mathbf{1}_{\{x_i \neq x'_i\}}, \quad x, x' \in \mathcal{X}^n.$$

But ρ_w is nothing but an ℓ_1 product metric (in the sense of (3)), with $\rho_i(x_i, x'_i) = w_i \mathbf{1}_{\{x_i \neq x'_i\}}$. Since $\text{diam}(\mathcal{X}_i, \rho_i) = w_i$, we have that indeed the formulations (1) and (16) are equivalent.

C. Stability proofs

Proof of Lemma 1. It follows from (Bousquet & Elisseeff, 2002, Lemma 7), that for all $i \in [n]$,

$$\begin{aligned} \mathbb{E}[R(\mathcal{A}, S) - \hat{R}_n(\mathcal{A}, S)] &= \\ \mathbb{E}_{Z_1^n, \tilde{Z}_1^n}[L(\mathcal{A}_{Z_1^n}, \tilde{Z}_i) - L(\mathcal{A}_{\tilde{Z}_1^n}, \tilde{Z}_i)], \end{aligned} \quad (29)$$

where $Z_1^n \sim \mu^n$ and \tilde{Z} is generated from Z via the process defined in (11). For fixed $i \in [n]$ and Z_1^{i-1}, Z_{i+1}^n , define

$$W_i(Z_i, Z'_i) = L(\mathcal{A}_{Z_1^n}, Z'_i) - L(\mathcal{A}_{Z_1^{i-1}Z'_iZ_{i+1}^n}, Z'_i)$$

and note that (23) implies that $|W_i(Z_i, Z'_i)| \leq \beta\rho(Z_i, Z'_i)$. Now rewrite (29) as

$$\begin{aligned} \mathbb{E}[R(\mathcal{A}, S) - \hat{R}_n(\mathcal{A}, S)] &= \\ \sum_{z_1^{i-1}, z_{i+1}^n} \mathbb{P}(z_1^{i-1})\mathbb{P}(z_{i+1}^n) \sum_{z_i, z'_i} \mathbb{P}(z_i)\mathbb{P}(z'_i)W_i(z_i, z'_i). \end{aligned} \quad (30)$$

Invoking Jensen's inequality and the argument in (17),

$$\begin{aligned} \exp\left(\sum_{z_i, z'_i} \mathbb{P}(z_i)\mathbb{P}(z'_i)W_i(z_i, z'_i)\right) &\leq \sum_{z_i, z'_i} \mathbb{P}(z_i)\mathbb{P}(z'_i)e^{W_i(z_i, z'_i)} \\ &= \frac{1}{2} \left[\sum_{z_i, z'_i} \mathbb{P}(z_i)\mathbb{P}(z'_i)e^{W_i(z_i, z'_i)} + \sum_{z_i, z'_i} \mathbb{P}(z_i)\mathbb{P}(z'_i)e^{-W_i(z_i, z'_i)} \right] \\ &\leq \frac{1}{2} \left[\sum_{z_i, z'_i} \mathbb{P}(z_i)\mathbb{P}(z'_i)e^{\beta\rho(z_i, z'_i)} + \sum_{z_i, z'_i} \mathbb{P}(z_i)\mathbb{P}(z'_i)e^{-\beta\rho(z_i, z'_i)} \right] \\ &\leq \exp\left(\frac{1}{2}\beta^2\Delta_{\text{SG}}^2(\mathcal{Z})\right). \end{aligned}$$

Taking logarithms yields the estimate

$$\sum_{z_i, z'_i} \mathbb{P}(z_i)\mathbb{P}(z'_i)W_i(z_i, z'_i) \leq \frac{1}{2}\beta^2\Delta_{\text{SG}}^2(\mathcal{Z}), \quad (31)$$

which, after substituting (31) into (30), proves the claim. \square

Proof of Lemma 2. We examine the two summands separately. The definition (22) of $R(\mathcal{A}, \cdot)$ implies that the latter is β -Lipschitz since it is a convex combination of β -Lipschitz functions. Now $\hat{R}_n(\mathcal{A}, \cdot)$ defined in (21) is also a convex combination of β -Lipschitz functions, but because z_i appears twice in $L(\mathcal{A}_{z_1^n}, z_i)$, changing z_i to z'_i could incur a difference of up to $2\beta\rho(z_i, z'_i)$. Hence, $\hat{R}_n(\mathcal{A}, \cdot)$ is 2β -Lipschitz. Since the Lipschitz seminorm is subadditive, the claim is proved. \square

D. Non-iid proof

Proof of Theorem 3. We begin by examining the martingale difference

$$V_i = \mathbb{E}[\varphi | X_1^i = x_1^i] - \mathbb{E}[\varphi | X_1^{i-1} = x_1^{i-1}]$$

as in the proof of Theorem 1. More explicitly,

$$\begin{aligned} V_i &= \sum_{x_{i+1}^n} \mathbb{P}(x_{i+1}^n | x_i^i) \varphi(x_1^i x_{i+1}^n) - \sum_{x_i^n} \mathbb{P}(x_i^n | x_1^{i-1}) \varphi(x_1^{i-1} x_i^n) \\ &= \sum_{x_i'} \mathbb{P}(x_i' | x_1^{i-1}) \cdot \sum_{x_{i+1}^n} [\mathbb{P}(x_{i+1}^n | x_i^i) \varphi(x_1^i x_{i+1}^n) - \\ &\quad \mathbb{P}(x_{i+1}^n | x_1^{i-1} x_i') \varphi(x_1^{i-1} x_i' x_{i+1}^n)]. \end{aligned}$$

Write \tilde{V}_i to denote V_i as a function of X_1^{i-1} with X_i integrated out:

$$\begin{aligned} \tilde{V}_i &= \sum_{x_i, x_i'} \mathbb{P}(x_i | X_1^{i-1}) \mathbb{P}(x_i' | X_1^{i-1}) \cdot \quad (32) \\ &\quad \sum_{x_{i+1}^n} [\mathbb{P}(x_{i+1}^n | X_1^{i-1} x_i) \varphi(X_1^{i-1} x_i x_{i+1}^n) \\ &\quad - \mathbb{P}(x_{i+1}^n | X_1^{i-1} x_i') \varphi(X_1^{i-1} x_i' x_{i+1}^n)]. \end{aligned}$$

Let π be an optimal coupling realizing the infimum in the transportation cost distance $T_{\rho_{i+1}^n}$ used to define $\tau_i(x_i^i, x_i')$. Recalling (25), we have

$$\begin{aligned} &\sum_{x_{i+1}^n} [\mathbb{P}(x_{i+1}^n | X_1^{i-1} x_i') \varphi(X_1^{i-1} x_i' x_{i+1}^n) \\ &- \mathbb{P}(x_{i+1}^n | X_1^{i-1} x_i) \varphi(X_1^{i-1} x_i x_{i+1}^n)] \\ &= \mathbb{E}_{(\dot{X}_{i+1}^n, \ddot{X}_{i+1}^n) \sim \pi} [\varphi(X_1^{i-1} x_i \dot{X}_{i+1}^n) - \varphi(X_1^{i-1} x_i' \ddot{X}_{i+1}^n)] \\ &\leq \mathbb{E}_\pi [\varphi(X_1^{i-1} x_i \dot{X}_{i+1}^n) - \varphi(X_1^{i-1} x_i' \ddot{X}_{i+1}^n) + \sum_{j=i+1}^n \rho_j(\dot{X}_j, \ddot{X}_j)] \\ &\leq \mathbb{E}_{\dot{X}_{i+1}^n \sim \mathbb{P}(\cdot | X_1^{i-1} x_i)} [\varphi(X_1^{i-1} x_i \dot{X}_{i+1}^n) - \varphi(X_1^{i-1} x_i' \dot{X}_{i+1}^n)] + \bar{\tau}_i \\ &= F(x_i) - F(x_i') + \bar{\tau}_i, \quad (33) \end{aligned}$$

where the first inequality holds by the Lipschitz property and the second by definition of $\bar{\tau}_i$, and $F : \mathcal{X}_i \rightarrow \mathbb{R}$ is defined by

$$F(y) = \sum_{x_{i+1}^n} \mathbb{P}(x_{i+1}^n | X_1^{i-1} x_i) \varphi(X_1^{i-1} y x_{i+1}^n).$$

Let us substitute (33) into (32):

$$\tilde{V}_i \leq \bar{\tau}_i + \sum_{x_i, x_i'} \mathbb{P}(x_i | X_1^{i-1}) \mathbb{P}(x_i' | X_1^{i-1}) (F(x_i) - F(x_i')).$$

Observe that F is 1-Lipschitz under ρ_i and apply Jensen's inequality:

$$\begin{aligned} &\mathbb{E}[e^{\lambda V_i} | X_1^{i-1}] \\ &\leq e^{\lambda \bar{\tau}_i} \sum_{x_i, x_i'} \mathbb{P}(x_i | X_1^{i-1}) \mathbb{P}(x_i' | X_1^{i-1}) e^{\lambda(F(x_i) - F(x_i'))} \\ &\leq e^{\lambda \bar{\tau}_i} \sum_{x_i, x_i'} \mathbb{P}(x_i | X_1^{i-1}) \mathbb{P}(x_i' | X_1^{i-1}) \cosh(\lambda \rho(x_i, x_i')) \\ &\leq \exp\left(\lambda \bar{\tau}_i + \frac{1}{2} \bar{\Delta}_{\text{SG}}^2(\mathcal{X}_i) \lambda^2\right), \end{aligned}$$

where the second inequality follows from the argument in (17) and the third from the definition of $\bar{\Delta}_{\text{SG}}(\mathcal{X}_i)$. Repeating the standard martingale argument in (19) yields

$$\begin{aligned} \mathbb{P}(\varphi - \mathbb{E}\varphi > t) &= \mathbb{P}\left(\sum_{i=1}^n V_i > t\right) \\ &\leq \exp\left(\frac{1}{2} \lambda^2 \sum_{i=1}^n \bar{\Delta}_{\text{SG}}^2(\mathcal{X}_i) - \lambda t + \lambda \sum_{i=1}^n \bar{\tau}_i\right). \end{aligned}$$

Optimizing over λ yields the claim. \square