Definition 4.1 (Lipschitz functions). *Let $(\mathbb{X}, d)$ be a metric space. A function $f : \mathbb{X} \to \mathbb{R}$ is called L-*Lipschitz *if $|f(x) - f(y)| \le L\, d(x, y)$ for all $x, y \in \mathbb{X}$. The family of all 1-Lipschitz functions is denoted* $\mathrm{Lip}(\mathbb{X})$.

$$|f(x) - f(y)| \le d(x,y)$$

X is σ^2- subgaussian if

variance proxy

$$\log \mathbb{E}\left[e^{\lambda(X - \mathbb{E}[X])} \right] \le \frac{\lambda^2 \sigma^2}{2} \quad \forall \lambda \in \mathbb{R}$$

$\boxed{\sigma^2}$ - subgaussian

$$\Rightarrow \mathbb{P}(X \ge \mathbb{E}[X] + t) \le e^{-t^2/2\sigma^2}$$

$$\Rightarrow \mathbb{P}(X \le \mathbb{E}[X] - t) \le e^{-t^2/2\sigma^2}$$

Definition 4.6 (Wasserstein distance). *The* Wasserstein distance *between probability measures $\mu, \nu \in \mathcal{P}_1(\mathbb{X}) := \{\rho : \int d(x, \cdot)\rho(dx) < \infty\}$ is defined as*[2]

$$W_1(\mu, \nu) := \sup_{f \in \mathrm{Lip}(\mathbb{X})} \left| \int f\, d\mu - \int f\, d\nu \right|. = \inf_{M} \int d(x,y)\, dM$$

Definition 4.7 (Relative entropy). *The* relative entropy *between probability measures ν and μ on any measurable space is defined as*

$$D(\nu||\mu) := \begin{cases} \mathrm{Ent}_\mu\left[\frac{d\nu}{d\mu}\right] & \text{if } \nu \ll \mu, \\ \infty & \text{otherwise.} \end{cases}$$

$$\mathrm{Ent}(Z) = \mathbb{E}[Z \log Z] - \mathbb{E} Z \cancel{\log \mathbb{E} Z}$$

$$= \int \log\left(\frac{d\nu}{d\mu}\right) d\nu$$

Theorem 4.8 (Bobkov-Götze). *Let $\mu \in \mathcal{P}_1(\mathbb{X})$ be a probability measure on a metric space $(\mathbb{X}, d)$. Then the following are equivalent for $X \sim \mu$:*

1. *$f(X)$ is σ^2-subgaussian for every $f \in \mathrm{Lip}(\mathbb{X})$.*
2. *$W_1(\nu, \mu) \le \sqrt{2\sigma^2 D(\nu||\mu)}$ for all ν.* T_1 inequality

p

T_p

Lemma 4.10 (Gibbs variational principle).

$$\log \mathbf{E}_\mu[e^f] = \sup_\nu \{\mathbf{E}_\nu[f] - D(\nu\|\mu)\}.$$

Suppose $f(X)$ is σ^2-subgaussian $\forall$ f Lip-

$$\log \mathbb{E}_\mu\left[e^{\lambda(f - \mathbb{E}_\mu f)}\right] \le \frac{\lambda^2\sigma^2}{2} \qquad \forall f \text{ Lip.}$$

$$\Leftrightarrow \quad \sup_\nu \left\{ \mathbb{E}_\nu\left[\lambda\left(f - \mathbb{E}_\mu f\right)\right] - D(\nu\|\mu) \right\} \le \frac{\lambda^2\sigma^2}{2} \qquad \begin{matrix}\forall f \\ \text{Lip} \\ \forall \lambda \in \mathbb{R}\end{matrix}$$

$$\Leftrightarrow \quad \sup_\lambda \sup_f \sup_\nu \left\{ \lambda \underbrace{\left(\mathbb{E}_\nu f - \mathbb{E}_\mu f\right)} - D(\nu\|\mu) - \frac{\lambda^2\sigma^2}{2} \right\} \le 0$$

$$\Leftrightarrow \quad \sup_\lambda \sup_\nu \left\{ \frac{\lambda W_1(\mu,\nu) - \lambda^2\sigma^2}{2} - D(\nu\|\mu) \right\} \le 0$$

taking $\frac{d}{d\lambda} = 0$

$$\Leftrightarrow \quad \sup_\nu \left\{ \frac{\left[W_1(\mu,\nu)\right]^2}{2\sigma^2} - D(\nu\|\mu) \right\} \le 0$$

$$\Leftrightarrow \quad W_1(\mu,\nu) \le \sqrt{2\sigma^2 D(\nu\|\mu)} \qquad \forall \nu$$

$$\|\mu - \nu\|_{\mathrm{TV}} \leq \sqrt{\frac{1}{2} D(\nu \| \mu)}$$ Pinsker's inequality

$$d(x,y) = \mathbb{1}_{\{x \neq y\}} \qquad f(x) - f(y) \leq 1 \quad \forall\, x,y$$

$$\sup f - \inf f \leq 1$$

$$W_1(\mu,\nu) = \sup_{0 \leq f \leq 1} \left| \int f\,d\mu - \int f\,d\nu \right| = \|\mu - \nu\|_{TV}$$

$0 \leq f(X) \leq 1$ Hoeffding's lemma $\Rightarrow \frac{1}{4}$-subgaussian

$$\|\mu - \nu\|_{TV} \leq \sqrt{2 \cdot \frac{1}{4} \cdot D(\nu\|\mu)} = \sqrt{\frac{1}{2} D(\nu\|\mu)}$$

Theorem 4.13 (Monge-Kantorovich duality). *We have*

$$W_1(\mu,\nu) = \sup_{f \in \mathrm{Lip}(\mathbb{X})} |\mathbf{E}_\mu f - \mathbf{E}_\nu f| = \inf_{\mathbf{M} \in \mathcal{C}(\mu,\nu)} \mathbf{E}_{\mathbf{M}}[d(X,Y)]$$

for all probability measures $\mu, \nu \in \mathcal{P}_1(\mathbb{X})$ *on a separable metric space* $(\mathbb{X}, d)$.

$f(X)$ is σ^2-subgaussian $\forall$ f lip

$$f(X_1, \ldots, X_n)$$

$$d(\vec{x}, \vec{y}) = \sum_{i=1}^{n} d_i(x_i, y_i)$$

$$d(\vec{x}, \vec{y}) = \left(\sum_{i=1}^{n} d_i(x_i, y_i)^2 \right)^{1/2}$$

Theorem 4.15 (Marton). *Let $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ be a convex function, and let $w_i : \mathbb{X}_i \times \mathbb{X}_i \to \mathbb{R}_+$ be positive weight function. Suppose that for $i = 1, \ldots, n$*

$$\inf_{\mathbf{M} \in \mathcal{C}(\mu_i, \nu)} \varphi(\mathbf{E}_{\mathbf{M}}[w_i(X, Y)]) \leq 2\sigma^2 D(\nu || \mu_i) \quad \text{for all } \nu.$$

Then we have

$$\inf_{\mathbf{M} \in \mathcal{C}(\mu_1 \otimes \cdots \otimes \mu_n, \nu)} \sum_{i=1}^{n} \varphi(\mathbf{E}_{\mathbf{M}}[w_i(X_i, Y_i)]) \leq 2\sigma^2 D(\nu || \mu_1 \otimes \cdots \otimes \mu_n) \quad \text{for all } \nu.$$

T_1-inequality

LHS not W_1 ?

$$\varphi(x) = x^2 \qquad w_i = d_i(x, y) \qquad \inf_{M} \left(\mathbb{E}_M \, d_i(X, Y) \right)^2 \leq 2\sigma^2 D(\cdot || \mu_i)$$

Corollary 4.16. *Suppose that the transportation cost inequality*

$$W_1(\mu_i, \nu) \leq \sqrt{2\sigma^2 D(\nu || \mu_i)} \quad \text{for all } \nu$$

holds for μ_i on $(\mathbb{X}_i, d_i)$ for $i = 1, \ldots, n$. Then the transportation cost inequality

$$W_1(\mu_1 \otimes \cdots \otimes \mu_n, \nu) \leq \sqrt{2\sigma^2 D(\nu || \mu_1 \otimes \cdots \otimes \mu_n)} \quad \text{for all } \nu$$

holds for $\mu_1 \otimes \cdots \otimes \mu_n$ on $(\mathbb{X}_1 \times \cdots \mathbb{X}_n, d_c)$ whenever $\boxed{\sum_{i=1}^n c_i^2 = 1.}$

$$\underline{d_c = \sum_{i=1}^{n} c_i d_i}$$

Cauchy-Schwarz

Theorem 4.20 (Talagrand). *Let $X_1, \ldots, X_n$ be independent, and suppose*

$$f(x) - f(y) \le \sum_{i=1}^{n} c_i(x) \mathbf{1}_{x_i \neq y_i} \quad \textit{for all } x, y.$$

$|f(x) - f(y)| \le d(x,y)$

Then $f(X_1, \ldots, X_n)$ is $\| \sum_{i=1}^{n} c_i^2 \|_\infty$-subgaussian.

x

y

$$\mathbf{P}[f(X) \ge \mathbf{E}f(X) + t] \le e^{-t^2/2\| \sum_i c_i^2 \|_\infty},$$
$$\mathbf{P}[f(X) \le \mathbf{E}f(X) - t] \le e^{-t^2/2\mathbf{E}[\sum_i c_i(X)^2]}$$

Suppose $X_1, \ldots, X_n$ indep, $\in [0,1]$, f convex

$$f(x) - f(y) \le \nabla f(x) \cdot (x-y) \qquad |x_i - y_i| \le 1$$

$$\le \sum_{i=1}^{n} \underbrace{\frac{\partial f(x)}{\partial x_i}}_{c_i(x)} \mathbb{1}_{\{x_i \neq y_i\}}$$

$f(X_1, \ldots X_n)$ is σ^2-subgaussian, $\sigma^2 = \| \sum_{i=1}^{n} c_i^2 \|_\infty$

$$\| \, \|\nabla f\|^2 \, \|_\infty$$

Definition 4.29 (Quadratic Wasserstein metric). *The* quadratic Wasserstein metric *for probability measures* μ, ν *on a metric space* $(\mathbb{X}, d)$ *is*

$$W_2(\mu,\nu) := \inf_{\mathbf{M}\in\mathcal{C}(\mu,\nu)} \sqrt{\mathbf{E}[d(X,Y)^2]}.$$

Corollary 4.30 (T_2-inequality). *Suppose that the probability measures* μ_i *on* $(\mathbb{X}_i, d_i)$ *satisfy the* quadratic transportation cost (T_2) inequality

$$W_2(\mu_i,\nu) \le \sqrt{2\sigma^2 D(\nu||\mu_i)} \quad \textit{for all } \nu.$$

Then we have

$$W_2(\mu_1\otimes\cdots\otimes\mu_n,\nu) \le \sqrt{2\sigma^2 D(\nu|\mu_1\otimes\cdots\otimes\mu_n)} \quad \textit{for all } \nu$$

on $(\mathbb{X}_1\times\cdots\times\mathbb{X}_n, [\sum_{i=1}^n d_i^2]^{1/2})$.

$\varphi(x) = x \qquad w_i = d_i(x_i, y_i)^2$

Theorem 4.31 (Gozlan). *Let* μ *be a probability measure on a Polish space* $(\mathbb{X}, d)$, *and let* $\{X_i\}$ *be i.i.d.* $\sim \mu$. *Denote by* $d_n(x,y) := [\sum_{i=1}^n d(x_i,y_i)^2]^{1/2}$ *the Euclidean metric on* $\mathbb{X}^n$. *Then the following are equivalent:*

1. μ *satisfies the* T_2*-inequality on* $(\mathbb{X}, d)$:

$$W_2(\mu,\nu) \le \sqrt{2\sigma^2 D(\nu||\mu)} \quad \textit{for all } \nu.$$

2. $\mu^{\otimes n}$ *satisfies the* T_1*-inequality on* $(\mathbb{X}^n, d_n)$ *for every* $n \ge 1$:

$$W_1(\mu^{\otimes n},\nu) \le \sqrt{2\sigma^2 D(\nu||\mu^{\otimes n})} \quad \textit{for all } \nu,\ n\ge 1.$$

3. *There is a constant* C *such that*

$$\mathbf{P}[f(X_1,\ldots,X_n) - \mathbf{E}f(X_1,\ldots,X_n) \ge t] \le Ce^{-t^2/2\sigma^2}$$

for every $n \ge 1$, $t \ge 0$ *and* 1*-Lipschitz function* f *on* $(\mathbb{X}^n, d_n)$.

Jensen's ineq.

$W_1(\mu,\nu) \le W_2(\mu,\nu)$

$2 \Rightarrow 3$ with $C=1$ $\qquad$ hard : $3 \Rightarrow 1$