Coupling in Dynamical Systems and the Wasserstein Metric

Problem Set

This problem set explores the relationship between *couplings* of probability measures and the *Wasserstein metric*, and how this relationship is used to study convergence and stability of dynamical systems (Markov chains, iterated function systems, and diffusions). Exercises are labelled [Basic], [Intermediate], or [Advanced]; the Advanced problems assume familiarity with graduate-level probability (stochastic calculus, weak convergence) and can be skipped on a first pass.

Preliminaries

Definition 1 (Coupling). *Let μ, ν be Borel probability measures on a Polish space (E, d). A* coupling *of μ and ν is a probability measure π on $E \times E$ whose marginals are μ and ν, i.e. $\pi(A \times E) = \mu(A)$ and $\pi(E \times B) = \nu(B)$ for all Borel sets A, B. Equivalently, a coupling is a pair of random variables (X, Y) on a common probability space with $X \sim \mu$, $Y \sim \nu$. We write $\Pi(\mu, \nu)$ for the set of all couplings of μ and ν.*

Definition 2 (Wasserstein distance). *For $p \geq 1$, let $\mathcal{P}_p(E)$ denote the probability measures on E with finite p-th moment. For $\mu, \nu \in \mathcal{P}_p(E)$, the* Wasserstein-$p$ distance *is*

$$W_p(\mu, \nu) \;=\; \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{E \times E} d(x,y)^p \, d\pi(x,y) \right)^{1/p} \;=\; \left(\inf_{(X,Y):\, X \sim \mu,\, Y \sim \nu} \mathbb{E}\big[d(X,Y)^p\big] \right)^{1/p}.$$

A coupling attaining the infimum is called an optimal coupling.

Theorem 1 (Kantorovich–Rubinstein duality). *For $\mu, \nu \in \mathcal{P}_1(E)$,*

$$W_1(\mu, \nu) \;=\; \sup_{f:\, \mathrm{Lip}(f) \leq 1} \left(\int_E f \, d\mu - \int_E f \, d\nu \right).$$

Remark 1 (Coupling inequality). *If E is countable and $d(x,y) = \mathbf{1}_{x \neq y}$ (or more generally for the total variation distance $d_{TV}(\mu,\nu) = \frac{1}{2} \sum_x |\mu(x) - \nu(x)|$), then for* any *coupling (X, Y) of μ, ν,*

$$d_{TV}(\mu, \nu) \;\leq\; \mathbb{P}(X \neq Y),$$

with equality for the maximal coupling*. This is the discrete case $p = 1$ of W_1 with the discrete metric, and is the starting point for the "coupling method" in Markov chain theory.*

1 Warm-up: Couplings and the Wasserstein Distance

Exercise 1.1 (Point masses). *[Basic] Let $\mu = \delta_a$ and $\nu = \delta_b$ be Dirac masses at $a, b \in \mathbb{R}$, with the usual metric $d(x,y) = |x - y|$.*

(a) Show that $\Pi(\mu, \nu)$ contains exactly one element, and identify it.

(b) *Compute $W_p(\mu, \nu)$ for every $p \geq 1$ directly from the definition.*

Exercise 1.2 (Independent vs. optimal coupling). [Basic] *Let $\mu = \nu = \text{Bernoulli}(1/2)$ on $\{0, 1\}$, with $d(x, y) = |x - y|$.*

(a) *Write down the* independent *coupling $\pi = \mu \otimes \nu$ and compute $\mathbb{E}_\pi |X - Y|$ under it.*

(b) *Exhibit a coupling with $X = Y$ almost surely, and conclude $W_1(\mu, \nu) = 0$. Why is this the expected answer, given that $\mu = \nu$?*

(c) *Now let $\nu_p = \text{Bernoulli}(p)$ for $p \neq 1/2$. Find the coupling of μ and ν_p that maximizes $\mathbb{P}(X = Y)$, and use it to compute $W_1(\mu, \nu_p)$.*

2 Coupling, Total Variation, and W_1

Exercise 2.1 (The coupling inequality). [Basic] *Let μ, ν be probability measures on a countable set E, and let (X, Y) be any coupling of μ, ν.*

(a) *Prove that $d_{TV}(\mu, \nu) \leq \mathbb{P}(X \neq Y)$.*

(b) *Construct explicitly a coupling that attains equality (the* maximal coupling*).* Hint: *let $\rho = \min(\mu, \nu)$ pointwise, put mass $\rho(x)$ on the diagonal at x for every x, and couple the remaining mass $\mu - \rho$ and $\nu - \rho$ independently on the off-diagonal.*

Exercise 2.2 (From total variation to W_1). [Intermediate] *Suppose the metric d on E is bounded, say $d(x, y) \leq D$ for all x, y.*

(a) *Show that $W_1(\mu, \nu) \leq D \cdot d_{TV}(\mu, \nu)$.*

(b) *Give an example of μ, ν on a bounded metric space for which this inequality is strict.*

(c) *Explain, with an example, why boundedness of d cannot be dropped: exhibit μ, ν on an unbounded space with $d_{TV}(\mu, \nu)$ fixed but $W_1(\mu, \nu)$ arbitrarily large.*

Exercise 2.3 (Kantorovich–Rubinstein, easy direction). [Intermediate] *Let (X, Y) be* any *coupling of $\mu, \nu \in \mathcal{P}_1(E)$ and let f be 1-Lipschitz.*

(a) *Show that*

$$\int_E f \, d\mu - \int_E f \, d\nu = \mathbb{E}[f(X) - f(Y)] \leq \mathbb{E}[d(X, Y)].$$

(b) *Taking the infimum over couplings on the right and the supremum over 1-Lipschitz f on the left, deduce*

$$\sup_{\text{Lip}(f) \leq 1} \left(\int f \, d\mu - \int f \, d\nu \right) \; \leq \; W_1(\mu, \nu).$$

(The reverse inequality — the hard direction, requiring the Kantorovich duality theorem for linear programs on Polish spaces — may be cited without proof.)

3 Coupling of Markov Chains and Convergence to Equilibrium

Exercise 3.1 (Coupling time and mixing). *[Intermediate] Let $(X_n)_{n\geq 0}$ and $(Y_n)_{n\geq 0}$ be two copies of the same time-homogeneous Markov chain on a countable state space, run on a common probability space so that once they meet they stay together: $X_n = Y_n \implies X_m = Y_m$ for all $m \geq n$. Let $T = \inf\{n \geq 0 : X_n = Y_n\}$ be the* coupling time.

(a) Prove the Markov chain coupling inequality:

$$d_{TV}\big(\mathrm{Law}(X_n), \mathrm{Law}(Y_n)\big) \;\leq\; \mathbb{P}(T > n).$$

(b) Let (X_n) be simple random walk on $\mathbb{Z}/N\mathbb{Z}$ (nearest-neighbour, lazy with holding probability $1/2$), started from two different points $x_0 \neq y_0$. Construct a coupling (X_n, Y_n) in which the two chains move with independent *increments as long as they have not met, and then move together (sharing all future increments) once they have. Show that the gap $D_n = X_n - Y_n \pmod N$ is itself a lazy random walk on $\mathbb{Z}/N\mathbb{Z}$ absorbed at 0 at time T, use this to bound $\mathbb{E}[T]$, and deduce a bound on $\mathbb{P}(T > n)$ and hence on the mixing time of the chain.* (Note: coupling the two chains with the **same** increment at every step, as one might first try, keeps the gap D_n constant — the chains would never meet unless $x_0 = y_0$. Independence before meeting is essential.)

Exercise 3.2 (Wasserstein contraction for a linear autoregression). *[Intermediate] Consider the AR(1) process on $\mathbb{R}$,*

$$X_{n+1} = aX_n + \xi_{n+1}, \qquad n \geq 0,$$

with $|a| < 1$ fixed and $(\xi_n)_{n\geq 1}$ i.i.d., independent of X_0, with $\mathbb{E}|\xi_1| < \infty$.

(a) Given two initial laws $\mu_0, \nu_0 \in \mathcal{P}_1(\mathbb{R})$, construct the synchronous coupling*: let $X_0 \sim \mu_0$, $Y_0 \sim \nu_0$ (coupled arbitrarily), and run both recursions with the* same *noise sequence (ξ_n).*

(b) Show that under this coupling, $|X_n - Y_n| = |a|^n |X_0 - Y_0|$ for all n, and deduce

$$W_1(\mu_n, \nu_n) \;\leq\; |a|^n\, W_1(\mu_0, \nu_0),$$

where μ_n, ν_n denote the laws of X_n, Y_n.

(c) Conclude that the chain has a unique invariant measure $\pi \in \mathcal{P}_1(\mathbb{R})$, and that for every $\mu_0 \in \mathcal{P}_1(\mathbb{R})$,

$$W_1(\mu_n, \pi) \;\leq\; |a|^n\, W_1(\mu_0, \pi) \xrightarrow[n\to\infty]{} 0.$$

State the exponential rate explicitly in terms of a.

4 Coupling in Dynamical Systems: Iterated Function Systems

Exercise 4.1 (The Hutchinson operator and its contraction rate). *[Intermediate] Let $f_1, \dots, f_k : \mathbb{R}^d \to \mathbb{R}^d$ be contractions with Lipschitz constants $c_1, \dots, c_k \in [0, 1)$, and let $p_1, \dots, p_k > 0$ with $\sum_i p_i = 1$. Define the Markov (transfer) operator on $\mathcal{P}_1(\mathbb{R}^d)$ by*

$$(P\mu)(A) \;=\; \sum_{i=1}^{k} p_i\, \mu\big(f_i^{-1}(A)\big), \qquad A \subseteq \mathbb{R}^d \text{ Borel.}$$

(a) Let (X, Y) be any coupling of $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$, and let I be a random index independent of (X, Y) with $\mathbb{P}(I = i) = p_i$. Show that $(f_I(X), f_I(Y))$ is a coupling of $P\mu$ and $P\nu$.

(b) *Deduce that*

$$W_1(P\mu, P\nu) \;\le\; \bar{c}\, W_1(\mu,\nu), \qquad \bar{c} := \sum_{i=1}^{k} p_i c_i \;<\; 1.$$

(c) *The space* $(\mathcal{P}_1(\mathbb{R}^d), W_1)$ *is a complete metric space. Using the Banach fixed-point theorem, conclude that* P *has a unique invariant (stationary) measure* $\pi \in \mathcal{P}_1(\mathbb{R}^d)$ *with* $P\pi = \pi$*, and that* $P^n\mu \to \pi$ *in* W_1*, at geometric rate* $\bar{c}^n$*, for every* $\mu \in \mathcal{P}_1(\mathbb{R}^d)$.

(d) (Optional computation.) *Take* $d = 1$, $f_1(x) = x/3$, $f_2(x) = x/3 + 2/3$, $p_1 = p_2 = 1/2$ *(the IFS generating the Cantor measure). Compute* $\bar{c}$ *and use part (c) to bound* $W_1(P^n\mu, \pi)$ *after* n *iterations, starting from* $\mu = \delta_0$.

5 Challenge Problems

Exercise 5.1 (Wasserstein convergence versus weak convergence). *[Advanced] Let* $(\mu_n)_{n\ge 1}$ *be probability measures on a Polish space* (E, d)*, and fix* $p \ge 1$, $x_0 \in E$.

(a) *Show that if* $W_p(\mu_n, \mu) \to 0$ *for some* $\mu \in \mathcal{P}_p(E)$*, then* $\mu_n \to \mu$ *weakly.* Hint: *bound* $\left|\int f\, d\mu_n - \int f\, d\mu\right|$ *for bounded Lipschitz* f *using an optimal (or near-optimal) coupling.*

(b) *Suppose in addition* $\sup_n \int d(x, x_0)^p\, d\mu_n(x) < \infty$ *and* $\{d(\cdot, x_0)^p\}$ *is uniformly integrable under* (μ_n)*. Show that weak convergence* $\mu_n \to \mu$ *then implies* $W_p(\mu_n, \mu) \to 0$. Hint: *use a coupling realizing* $\mu_n \to \mu$ *almost surely (Skorokhod representation) together with a truncation argument.*

(c) *Give an explicit example on* $E = \mathbb{R}$ *(with* $p = 1$*) of a sequence* $\mu_n \to \mu$ *weakly for which* $W_1(\mu_n, \mu) \not\to 0$*, showing that uniform integrability in part (b) cannot be dropped.*

Hints

1.1 Both marginals are single points, so π must put all its mass at (a, b); then $W_p(\mu,\nu)^p = |a - b|^p$ for every p, so $W_p(\mu, \nu) = |a - b|$.

1.2 (a) $\frac{1}{2}$. (b) Take $X = Y$; both are Bernoulli(1/2), so this is a valid coupling. (c) The optimal coupling makes $X = Y$ with the largest possible probability, i.e. $\mathbb{P}(X = Y) = \min(p, 1 - p) + \dots$; work out the residual mass explicitly and compute $\mathbb{E}|X - Y| = |p - 1/2| \cdot (\text{something})$ — in fact $W_1(\mu, \nu_p) = |p - 1/2|$.

2.1 For (a), split the sum $\sum_x |\mu(x) - \nu(x)|$ using $\mathbb{P}(X = x, Y = x) \le \min(\mu(x), \nu(x))$. For (b), check the two marginals of the constructed π equal μ, ν and that $\mathbb{P}(X \ne Y) = 1 - \sum_x \min(\mu(x), \nu(x)) = d_{TV}(\mu, \nu)$.

2.2 (a) Multiply the coupling inequality by D and take $d(X, Y) \le D\mathbf{1}_{X\ne Y}$. (c) Try $\mu = \delta_0$, $\nu = \delta_R$ with $R \to \infty$: $d_{TV} = 1$ always, but $W_1 = R \to \infty$.

2.3 Direct computation using $|f(x) - f(y)| \le d(x, y)$; no further hint needed.

3.1 (a) Condition on $\{T \le n\}$ vs. $\{T > n\}$ and use that $X_n = Y_n$ on $\{T \le n\}$. (b) Do *not* give both walks the same increment — that leaves the gap $D_n = X_n - Y_n \pmod N$ constant, so the walks never meet. Instead let X_n and Y_n take *independent* lazy ± 1 steps until they meet; then D_n evolves as the difference of two independent lazy steps, which is again a lazy, symmetric random walk on $\mathbb{Z}/N\mathbb{Z}$. Being an irreducible walk on a finite state space, D_n reaches 0 almost surely, and a variance/optional-stopping computation ($\mathbb{E}[D_n^2]$ grows linearly in n until absorption) gives $\mathbb{E}[T] = O(N^2)$. Once $X_n = Y_n$, switch to a shared increment so the chains stay equal. Combine with part (a) and Markov's inequality, $\mathbb{P}(T > n) \le \mathbb{E}[T]/n$, to bound d_{TV}.

3.2 (b) follows from (a) by taking expectations and using that the constructed pair is *a* coupling, so $W_1(\mu_n, \nu_n) \le \mathbb{E}|X_n - Y_n|$. (c) Existence of π: show (μ_n) is W_1-Cauchy using (b) with $\nu_0 = \mu_1$; the rate is $|a|^n$.

4.1 (a) Check marginals directly: conditioning on I, $f_I(X) \sim P\mu$. (b) $\mathbb{E}d(f_I(X), f_I(Y)) = \sum_i p_i \mathbb{E}[d(f_i(X), f_i(Y))] \le \sum_i p_i c_i \mathbb{E}d(X,Y)$; take the infimum over couplings (X,Y) of μ, ν. (c) Apply the contraction mapping theorem to $P : \mathcal{P}_1(\mathbb{R}^d) \to \mathcal{P}_1(\mathbb{R}^d)$. (d) $\bar{c} = 1/3$.

5.1 (a) Bound bounded-Lipschitz test functions by an optimal coupling and Markov's inequality on $d(X,Y)$. (b) Truncate $d(x, x_0)^p$ at level M, use uniform integrability to control the tail uniformly in n, and weak convergence (via Skorokhod) to control the truncated part. (c) Try $\mu_n = (1 - \frac{1}{n})\delta_0 + \frac{1}{n}\delta_n$: $\mu_n \to \delta_0$ weakly but $W_1(\mu_n, \delta_0) = 1$ for all n.