Analysis, Random Walks and Groups
Spring 2019

Week 4 tutorial

Given two $t, s \in \mathbb{Z}_p$ we can always find $n, m \in \mathbb{N}$ such that

$$t \ominus 1 \ominus 1 \cdots \ominus 1 = s \quad \text{where } \ominus 1 \text{ is taken } n \text{ times}$$

$$t \oplus 1 \oplus 1 \cdots \oplus 1 = s \quad \text{where } \oplus 1 \text{ is taken } m \text{ times}$$

The **word/geodesic distance** $\mathrm{dist}(t, s)$ in $\mathbb{Z}_p$ is then the smallest of these n and m.

1. A function $f : \mathbb{Z}_p \to \mathbb{R}$ is called **Lipschitz** if there is a constant $L \geq 0$ such that

$$|f(t) - f(s)| \leq L \, \mathrm{dist}(t, s), \quad t, s \in \mathbb{Z}_p.$$

The best constant L for which this condition holds is called the **Lipschitz constant** $\mathrm{Lip}(f)$, that is,

$$\mathrm{Lip}(f) = \min\{L \geq 0 : |f(t) - f(s)| \leq L \, \mathrm{dist}(t, s), \quad t, s \in \mathbb{Z}_p\}.$$

Why is any function $f : \mathbb{Z}_p \to \mathbb{R}$ Lipschitz?

Solution. The number

$$0 \leq L := \max_{a \neq b} \frac{|f(a) - f(b)|}{d(a, b)} < \infty$$

as $\mathbb{Z}_p$ is finite since $d(a, b) \geq 1$ when $a \neq b$ and the maximum is over a finite set $\mathbb{Z}_p$. Then

$$|f(t) - f(s)| = \frac{|f(t) - f(s)|}{d(t, s)} \cdot d(t, s) \leq Ld(t, s).$$

Moreover, if $t = s$, then as $d(t, s) \geq 0$ we have $|f(t) - f(s)| = 0 \leq Ld(t, s)$, so f is Lipschitz.

2. Prove the variational formula for the total variation distance between two probability distributions μ, ν in $\mathbb{Z}_p$:

$$d(\mu, \nu) = \frac{1}{2} \max\{|\mu(f) - \nu(f)| : \|f\|_\infty \leq 1, f : \mathbb{Z}_p \to \mathbb{R}\}.$$

Hint: Use the L^1 formula for the total variation distance from the lectures.

Solution. Let us first prove

$$d(\mu, \nu) \leq \frac{1}{2} \max\{|\mu(f) - \nu(f)| : \|f\|_\infty \leq 1, f : \mathbb{Z}_p \to \mathbb{R}\}.$$

To do this, we will construct a function g such that $\frac{1}{2}|\mu(g) - \nu(g)|$ realises the $d(\mu, \nu)$. Write

$$B = \{t \in \mathbb{Z}_p : \mu(t) \geq \nu(t)\}.$$

Define a function

$$g(t) = \begin{cases} +1, & t \in B \\ -1, & t \notin B. \end{cases}$$

Then we have

$$\begin{aligned} \mu(g) - \nu(g) &= \sum_{t \in \mathbb{Z}_p} g(t)\mu(t) - \sum_{t \in \mathbb{Z}_p} g(t)\nu(t) \\ &= \sum_{t \in \mathbb{Z}_p} g(t)(\mu(t) - \nu(t)) \\ &= \sum_{t \in B} g(t)(\mu(t) - \nu(t)) + \sum_{t \notin B} g(t)(\mu(t) - \nu(t)) \end{aligned}$$

Here if $t \in B$ we have by the definition of B that

$$g(t)(\mu(t) - \nu(t)) = \mu(t) - \nu(t) \geq 0$$

and if $t \notin B$ we have by the definition of B that

$$g(t)(\mu(t) - \nu(t)) = -(\mu(t) - \nu(t)) \geq 0.$$

Hence

$$\sum_{t\in B} g(t)(\mu(t)-\nu(t)) + \sum_{t\notin B} g(t)(\mu(t)-\nu(t)) = \sum_{t\in B} |\mu(t)-\nu(t)| + \sum_{t\notin B} |\mu(t)-\nu(t)| = \sum_{t\in \mathbb{Z}_p} |\mu(t)-\nu(t)|.$$

This in particular proves that

$$\mu(g)-\nu(g) = \sum_{t\in\mathbb{Z}_p} |\mu(t)-\nu(t)|,$$

making $\mu(g)-\nu(g)\geq 0$. Thus by the L^1 formula for the total variation distance we obtain

$$d(\mu,\nu) = \frac{1}{2}\sum_{t\in\mathbb{Z}_p} |\mu(t)-\nu(t)| = \frac{1}{2}|\mu(g)-\nu(g)|.$$

Since

$$\|g\|_\infty = \max\{|g(t) : t\in\mathbb{Z}_p\} = 1,$$

we have proved

$$d(\mu,\nu) \leq \frac{1}{2}\max\{|\mu(f)-\nu(f)| : \|f\|_\infty \leq 1, f:\mathbb{Z}_p\to\mathbb{R}\}.$$

We still need to prove the reverse direction:

$$d(\mu,\nu) \geq \frac{1}{2}\max\{|\mu(f)-\nu(f)| : \|f\|_\infty \leq 1, f:\mathbb{Z}_p\to\mathbb{R}\}.$$

To do this, fix any $f:\mathbb{Z}_p\to\mathbb{R}$ with $\|f\|_\infty\leq 1$. Then $f(t)\leq 1$ for all $t\in\mathbb{Z}_p$. Thus we have by the triangle inequality

$$\begin{aligned}
|\mu(f)-\nu(f)| &= \Big|\sum_{t\in\mathbb{Z}_p} f(t)\mu(t) - \sum_{t\in\mathbb{Z}_p} f(t)\nu(t)\Big| \\
&= \Big|\sum_{t\in\mathbb{Z}_p} f(t)(\mu(t)-\nu(t))\Big| \\
&\leq \sum_{t\in\mathbb{Z}_p} |f(t)||\mu(t)-\nu(t)| \\
&\leq \sum_{t\in\mathbb{Z}_p} 1\cdot|\mu(t)-\nu(t)| \\
&= 2d(\mu,\nu),
\end{aligned}$$

where in the last line we used the L^1 identity for the total variation distance. Hence the claim follows as $f:\mathbb{Z}_p\to\mathbb{R}$ is arbitrary.

3. The **earth mover's distance** or also known as the first **Wasserstein distance** $W_1(\mu,\nu)$ of two probability distributions μ,ν in $\mathbb{Z}_p$ is given by

$$W_1(\mu,\nu) = \max\Big\{|\mu(f)-\nu(f)| : \mathrm{Lip}(f)\leq 1, f:\mathbb{Z}_p\to\mathbb{R}\Big\}.$$

Prove that

$$d(\mu,\nu)\leq W_1(\mu,\nu),$$

where $d(\mu,\nu)$ is the total variation distance.

Solution. Take $f:\mathbb{Z}_p\to\mathbb{R}$ such that $\|f\|_\infty\leq 1$. To prove the claim, it will be enough if we construct a Lipschitz function $g:\mathbb{Z}_p\to\mathbb{R}$ with $\mathrm{Lip}(g)\leq 1$ such that

$$|\mu(f)-\nu(f)|\leq 2|\mu(g)-\nu(g)|.$$

For this purpose, let us first define a number

$$L := \max_{t\neq s}\frac{|f(t)-f(s)|}{d(t,s)}.$$

Since for $t\neq s$ the distance $1\leq d(t,s)$ and $|f(t)-f(s)|\leq 2\|f\|_\infty\leq 2$ we have

$$0\leq L\leq 2.$$

Define a function $g:\mathbb{Z}_p\to\mathbb{R}$ by

$$g(t) = f(t)/L, \quad t\in\mathbb{Z}_p.$$

Then for $t \neq s$ we have by the definition of L that

$$|g(t) - g(s)| = \frac{1}{L}|f(t) - f(s)| = \frac{1}{L} \cdot \frac{|f(t) - f(s)|}{d(t,s)} \cdot d(t,s) \leq d(t,s).$$

Moreover, for $t = s$ we have

$$|g(t) - g(s)| = 0 \leq d(t,s).$$

Thus g is Lipschitz with $\mathrm{Lip}(g) \leq 1$. Lastly, since

$$\mu(g) = \sum_{t \in \mathbb{Z}_p} g(t)\mu(t) = \sum_{t \in \mathbb{Z}_p} \frac{f(t)}{L}\mu(t) = L^{-1}\mu(f),$$

and similarly for $\nu(g)$, we see that

$$|\mu(f) - \nu(f)| = L|L^{-1}\mu(f) - L^{-1}\nu(f)| = L|\mu(g) - \nu(g)| \leq 2|\mu(g) - \nu(g)|,$$

which gives the claim.

Remark 0.1. Wasserstein distance appears commonly in a field called **mass transportation theory**, which has many applications throughout economics, physics and mathematics of PDEs. We can state it in our course's language as follows. Given a probability distribution μ on $\mathbb{Z}_p$ and a map $T : \mathbb{Z}_p \to \mathbb{Z}_p$, define the **push-forward** $T_*\mu$ by the formula

$$T_*\mu(t) = \mu(T^{-1}\{t\}), \quad t \in \mathbb{Z}_p,$$

where $T^{-1}\{t\} = \{s \in \mathbb{Z}_p : T(t) = s\}$ is the pre-image of the singleton $\{t\}$. Given two probability distributions μ, ν in $\mathbb{Z}_p$, a mapping $T : \mathbb{Z}_p \to \mathbb{Z}_p$ that maps μ onto ν, $T_*\mu = \nu$, is called an **optimal transportation** if it minimises the "cost":

$$\int \mathrm{dist}(t, T(t))\, d\mu(t). \qquad \Big(= \sum_{t \in \mathbb{Z}_p} \mathrm{dist}(t, T(t))\, \mu(t) \Big)$$

The **Monge-Kantorovich duality theorem** says the minimal cost is the first Wasserstein distance:

$$\min \Big\{ \int \mathrm{dist}(t, T(t))\, d\mu(t) : T_*\mu = \nu \Big\} = W_1(\mu, \nu),$$

the proof can be found from literature on optimal transportation theory.

4. Fix $s \in \mathbb{Z}_p$. Define a transportation map $T : \mathbb{Z}_p \to \mathbb{Z}_p$ by $T(t) = s$, $t \in \mathbb{Z}_p$. Verify that the uniform distribution $\lambda(t) = 1/p$, $t \in \mathbb{Z}_p$, satisfies $T_*\lambda = \delta_s$, where δ_s is the singular distribution at s. What is the cost of transporting the uniform mass λ to a point δ_s? That is, find the cost

$$\int \mathrm{dist}(t, T(t))\, d\lambda(t).$$

Can you transport δ_s to λ?

Solution. Fix $t \in \mathbb{Z}_p$. We have

$$\lambda(T^{-1}\{t\}) = \lambda(\varnothing) = 0;$$

if $t \neq s$ and

$$\lambda(T^{-1}\{t\}) = \lambda(\mathbb{Z}_p) = 1$$

if $t = s$. Therefore, $\lambda(T^{-1}\{t\}) = \delta_s(t)$, for all $t \in \mathbb{Z}_p$, as we claimed.

By the definition of the integral

$$\int \mathrm{dist}(t, T(t))\, d\lambda(t) = \sum_{t \in \mathbb{Z}_p} \mathrm{dist}(t, T(t))\lambda(t) = \sum_{t \in \mathbb{Z}_p} \mathrm{dist}(t, s)\frac{1}{p} = \frac{1}{p} \sum_{t \in \mathbb{Z}_p} \mathrm{dist}(t, s)$$

If p is even, then we see that

$$\sum_{t \in \mathbb{Z}_p} \mathrm{dist}(t, s) = \sum_{k=1}^{p/2} 1 + \sum_{k=1}^{p/2} 1 = \frac{p(p+1)}{2}$$

by taking the both directions from s to either clockwise or counterclockwise. Hence the cost is

$$\int \mathrm{dist}(t, T(t))\, d\lambda(t) = \frac{1}{p}\frac{p(p+1)}{2} = \frac{p+1}{2}.$$

If p is odd, then we have one less distance in the sums, so we have

$$\sum_{t\in\mathbb{Z}_p} \mathrm{dist}(t,s) = \sum_{k=1}^{(p-1)/2} 1 + \sum_{k=1}^{(p-1)/2} 1 = \frac{(p-1)p}{2}.$$

Hence the cost of transporting λ to δ_s is

$$\int \mathrm{dist}(t, T(t))\, d\lambda(t) = \frac{1}{p}\frac{(p-1)p}{2} = \frac{p-1}{2}.$$

It is not possible to transport δ_s to λ since otherwise, if such map $S : \mathbb{Z}_p \to \mathbb{Z}_p$ exists with

$$\lambda = S_*\delta_s.$$

Define $t := S(s) \in \mathbb{Z}_p$. Then $s \in S^{-1}\{t\}$ so we have

$$\delta_s(S^{-1}(t)) = 1.$$

On the other hand, we assumed $\lambda = S_*\delta_s$ so

$$1/p = \lambda(s) = \delta_s(S^{-1}(t)) = 1$$

which is not possible when $p \geq 2$.