912702055
Congyue Deng

Homework 4

Problem 1

(a) We can get an intuitive understanding of $\phi(s,t) = \exp_{\gamma(s)}(t\mathbf{N}(s))$ from Figure 1. If t is too large, the geodesic α such that $\alpha(0) = \gamma(s)$, $\dot{\gamma}(0) = \mathbf{N}(s)$ may not exist on $[0,t]$, and thus the exponential map is not well-defined.

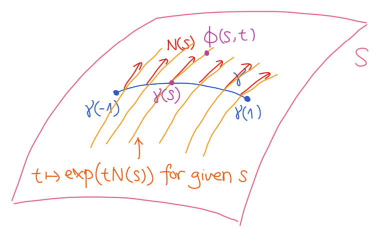


Figure 1: Illustration of the map $\phi(s,t) = \exp_{\gamma(s)}(t\mathbf{N}(s))$.

(b) By the definition of exponential map, for any fixed $s \in (-1,1)$, $t \mapsto \phi(s,t) = \exp_{\gamma(s)}(t\mathbf{N}(s))$ is a geodesic on S starting at $\gamma(s)$ with initial velocity $\mathbf{N}(s)$. Note that $t \mapsto \mathbf{E}_2(s,t)$ is the velocity of this geodesic. Hence by the geodesic equation we have $\nabla_{\mathbf{E}_2}\mathbf{E}_2 = 0$ for all s,t.

(c) By the definition of exponential map, $\forall s \in (-1,1)$, there locally exists a geodesic $\alpha : (-\varepsilon,\varepsilon) \to S$ such that $\alpha(0) = \gamma(s)$, $\dot{\alpha}(0) = \mathbf{N}(s)$ and $\exp_{\gamma(s)}(t\mathbf{N}(s)) = \alpha(t)$. Since α is of constant speed and $\|\dot{\alpha}(0)\| = \|\mathbf{N}(s)\| = 1$, α is parameterized by arc length. Then $\mathbf{E}_2 = \frac{\partial\phi}{\partial t} = \dot{\alpha}(t)$ and $\frac{\partial}{\partial t}\mathbf{E}_2 = \ddot{\alpha}(t)$. We have

$$\frac{\partial}{\partial t}\|\mathbf{E}_2\|^2 = 2\left\langle \mathbf{E}_2, \frac{\partial \mathbf{E}_2}{\partial t}\right\rangle = 2\langle\dot{\alpha}(t), \ddot{\alpha}(t)\rangle = 0.$$

Or equivalently we may notice that $\frac{\partial}{\partial t}\|\mathbf{E}_2\|^2 = \mathbf{E}_2(\langle\mathbf{E}_2,\mathbf{E}_2\rangle) = 2\langle\nabla_{\mathbf{E}_2}\mathbf{E}_2, \mathbf{E}_2\rangle = 0$. Therefore $\|\mathbf{E}_2(s,t)\| \equiv \|\mathbf{E}_2(s,0)\| = \|\dot{\alpha}(0)\| = \|\mathbf{N}(s)\| = 1$.

(d) Since $\mathbf{E}_1, \mathbf{E}_2$ are partial derivatives of ϕ, we have $[\mathbf{E}_1,\mathbf{E}_2] = \nabla_{\mathbf{E}_1}\mathbf{E}_2 - \nabla_{\mathbf{E}_2}\mathbf{E}_1 = 0$. Thus

$$\begin{aligned}\frac{\partial}{\partial t}\langle\mathbf{E}_1,\mathbf{E}_2\rangle &= \mathbf{E}_2(\langle\mathbf{E}_1,\mathbf{E}_2\rangle) = \langle\nabla_{\mathbf{E}_2}\mathbf{E}_1,\mathbf{E}_2\rangle + \langle\mathbf{E}_1,\nabla_{\mathbf{E}_2}\mathbf{E}_2\rangle \\ &= \langle\nabla_{\mathbf{E}_2}\mathbf{E}_1,\mathbf{E}_2\rangle = \langle\nabla_{\mathbf{E}_1}\mathbf{E}_2,\mathbf{E}_2\rangle = \frac{1}{2}\mathbf{E}_1(\langle\mathbf{E}_2,\mathbf{E}_2\rangle) = \frac{1}{2}\frac{\partial}{\partial s}\|\mathbf{E}_2\|^2 = 0,\end{aligned}$$

where the last equality holds because in (c) we have shown that $\|\mathbf{E}_2\| \equiv 1$.

Problem 2

(a) By definition, for all $\mathbf{x} \in \mathbb{R}^3$, $\mathrm{d}f(\mathbf{x}) = \sum_{i=1}^3 f_i x^i = \langle\nabla f, \mathbf{x}\rangle$, which is just $\nabla f = (\mathrm{d}f)^\sharp$.

(b) By definition we have

$$\begin{aligned}\star\mathrm{d}\star(\mathbf{F}^\flat) &= \star\mathrm{d}\star\left(\sum_{i=1}^{3} F^i \mathrm{d}x^i\right) = \star\mathrm{d}\left(F^1\mathrm{d}x^2\wedge\mathrm{d}x^3 + F^2\mathrm{d}x^3\wedge\mathrm{d}x^1 + F^3\mathrm{d}x^1\wedge\mathrm{d}x^2\right)\\ &= *\left(F^1_1\mathrm{d}x^1\wedge\mathrm{d}x^2\wedge\mathrm{d}x^3 + F^2_2\mathrm{d}x^2\wedge\mathrm{d}x^3\wedge\mathrm{d}x^1 + F^3_3\mathrm{d}x^3\wedge\mathrm{d}x^1\wedge\mathrm{d}x^2\right)\\ &= *\left(\left(\sum_{i=1}^{3} F^i_i\right)\mathrm{d}x^1\wedge\mathrm{d}x^2\wedge\mathrm{d}x^3\right) = \sum_{i=1}^{3} F^i_i = \nabla\cdot\mathbf{F}.\end{aligned}$$

(c) By definition we have

$$\begin{aligned}(\star\mathrm{d}(\mathbf{F}^\flat))^\sharp &= \left(\star\mathrm{d}\left(\sum_{i=1}^{3} F^i\mathrm{d}x^i\right)\right)^\sharp\\ &= \left(\star\left((F^2_1 - F^1_2)\mathrm{d}x^1\wedge\mathrm{d}x^2 + (F^3_2 - F^2_3)\mathrm{d}x^2\wedge\mathrm{d}x^3 + (F^1_3 - F^3_1)\mathrm{d}x^3\wedge\mathrm{d}x^1\right)\right)^\sharp\\ &= \left((F^2_1 - F^1_2)\mathrm{d}x^3 + (F^3_2 - F^2_3)\mathrm{d}x^1 + (F^1_3 - F^3_1)\mathrm{d}x^2\right)^\sharp\\ &= \left(F^3_2 - F^2_3, F^1_3 - F^3_1, F^2_1 - F^1_2\right)^t = \nabla\times\mathbf{F}.\end{aligned}$$

(d) By (a) and (b), $\Delta f = \nabla\cdot\nabla f = \star\mathrm{d}\star\left((\mathrm{d}f)^\sharp\right)^\flat = \star\mathrm{d}\star\mathrm{d}f$.

Problem 3

(a) Since $\mathrm{d}x^i\wedge\star\mathrm{d}x^j = \langle\mathrm{d}x^i,\mathrm{d}x^j\rangle\mathrm{d}A = g^{ij}\sqrt{|g|}\mathrm{d}x^1\wedge\mathrm{d}x^2$, we conclude that

$$*\mathrm{d}x^1 = \sqrt{|g|}\left(-g^{12}\mathrm{d}x^1 + g^{11}\mathrm{d}x^2\right) \quad\text{and}\quad *\,\mathrm{d}x^2 = \sqrt{|g|}\left(-g^{22}\mathrm{d}x^1 + g^{12}\mathrm{d}x^2\right).$$

And since

$$\mathrm{d}x^1\wedge\mathrm{d}x^2\wedge\star\left(\mathrm{d}x^1\wedge\mathrm{d}x^2\right) = \left\|\mathrm{d}x^1\wedge\mathrm{d}x^2\right\|^2\mathrm{d}A = |g^{-1}|\sqrt{|g|}\mathrm{d}x^1\wedge\mathrm{d}x^2 = \frac{1}{\sqrt{|g|}}\mathrm{d}x^1\wedge\mathrm{d}x^1.$$

we have $\star(\mathrm{d}x^1\wedge\mathrm{d}x^2) = \frac{1}{\sqrt{|g|}}$. Now suppose $\mathbf{X} = X^i\frac{\partial}{\partial x^i}$ where $\frac{\partial}{\partial x^1}(\mathbf{p}), \frac{\partial}{\partial x^2}(\mathbf{p})$ form an basis of $T_\mathbf{p}S$. Then

$$\begin{aligned}\mathrm{div}(\mathbf{X})\mathrm{d}A &= \star\mathrm{d}\star(\mathbf{X}^\flat)\mathrm{d}A = \star\mathrm{d}\star(g_{ij}X^j\mathrm{d}x^i)\mathrm{d}A\\ &= \star\mathrm{d}\left(g_{1j}X^j\sqrt{|g|}\left(-g^{12}\mathrm{d}x^1 + g^{11}\mathrm{d}x^2\right) + g_{2j}X^j\sqrt{|g|}\left(-g^{22}\mathrm{d}x^1 + g^{12}\mathrm{d}x^2\right)\right)\mathrm{d}A\\ &= \star\mathrm{d}\left(-\sqrt{|g|}\left(X^1g_{1j}g^{j2} + X^2g_{2j}g^{j2}\right)\mathrm{d}x^1 + \sqrt{|g|}\left(X^1g_{1j}g^{j1} + X^2g_{2j}g^{j1}\right)\mathrm{d}x^2\right)\mathrm{d}A\end{aligned}$$

$$
\begin{aligned}
&= \star\mathrm{d}\left(-\sqrt{|g|}X^2\mathrm{d}x^1 + \sqrt{|g|}X^1\mathrm{d}x^2\right)\mathrm{d}A \\
&= \star\left(-\left(\sqrt{|g|}X^2\right)_2 \mathrm{d}x^2 \wedge \mathrm{d}x^1 + \left(\sqrt{|g|}X^1\right)_1 \mathrm{d}x^1 \wedge \mathrm{d}x^2\right)\mathrm{d}A \\
&= \star\left(\left(\sqrt{|g|}X^i\right)_i \mathrm{d}x^1 \wedge \mathrm{d}x^2\right)\mathrm{d}A = \frac{1}{\sqrt{|g|}}\left(\sqrt{|g|}X^i\right)_i \sqrt{|g|}\mathrm{d}x^1 \wedge \mathrm{d}x^2 \\
&= \left(\sqrt{|g|}X^i\right)_i \mathrm{d}x^1 \wedge \mathrm{d}x^2 = \mathrm{d}\left(\sqrt{|g|}X^1\mathrm{d}x^2 - \sqrt{|g|}X^2\mathrm{d}x^1\right).
\end{aligned}
$$

Therefore $\mathrm{div}(\mathbf{X})\mathrm{d}A$ is of the form $\mathrm{d}\omega$ with $\omega = \sqrt{|g|}\left(X^1\mathrm{d}x^2 - X^2\mathrm{d}x^1\right)$, which implies that it is an exact form.

(b) Suppose $\mathbf{N} = N^i\frac{\partial}{\partial x^i}$. Since

$$\star\omega = |g|X^1\left(-g^{22}\mathrm{d}x^1 + g^{12}\mathrm{d}x^2\right) - |g|X^2\left(-g^{12}\mathrm{d}x^1 + g^{11}\mathrm{d}x^2\right),$$

and note that $\mathrm{d}x^i\left(\frac{\partial}{\partial x^j}\right) = \delta_{ij}$, we have

$$
\begin{aligned}
\omega(\mathbf{T}) &= \omega(\mathrm{Rot}_{\pi/2}(-\mathbf{N})) = \star\omega(-\mathbf{N}) = |g|X^1\left(g^{22}N^1 - g^{12}N^2\right) - |g|X^2\left(g^{12}N^1 - g^{11}N^2\right) \\
&= X^1\left(g_{11}N^1 + g_{12}N^2\right) - X^2\left(-g_{12}N^1 - g_{22}N^2\right) \\
&= X^i g_{ij} N^j = \left\langle X^i\frac{\partial}{\partial x^i}, N^j\frac{\partial}{\partial x^j}\right\rangle = \langle\mathbf{X},\mathbf{N}\rangle.
\end{aligned}
$$

Problem 4

(a) $W(\mathbf{p},\mathbf{q})$ measures the minimal "cost" of transporting the measures from $\mathbf{p}$ to $\mathbf{q}$. Compared with KL divergence, $W(\mathbf{p},\mathbf{q})$ measures the displacement between probability distributions, while KL divergence measures the overlap.

(b) Let $K_\alpha = \exp(-D/\alpha)$, *i.e.* $(K_\alpha)_{ij} = \exp(-D_{ij}/\alpha)$ for all i,j. Then

$$
\begin{aligned}
\alpha\cdot\mathrm{KL}(T\|K_\alpha) &= \alpha\sum_{i,j}T_{ij}\left(\log T_{ij} - \log(K_\alpha)_{ij}\right) = \alpha\sum_{i,j}T_{ij}\left(\log T_{ij} + \frac{D_{ij}}{\alpha}\right) \\
&= \sum_{i,j}T_{ij}D_{ij} + \alpha\sum_{i,j}T_{ij}\log T_{ij},
\end{aligned}
$$

which is the objective for $W_\alpha(\mathbf{p},\mathbf{q})$.

(c) Note that for the entropy-regularized EMD, the $T_{ij}\geqslant 0$ constraint is no longer necessary since the term $\alpha\sum_{i,j}T_{ij}\log T_{ij}$ already requires $T_{ij}\geqslant 0$. Ignoring then $T_{ij}\geqslant 0$ constraint, the Lagrange multiplier for $W_\alpha(\mathbf{p},\mathbf{q})$ is

$$\mathcal{L}(T;\lambda,\mu) = \sum_{i,j}T_{ij}D_{ij} + \alpha\sum_{i,j}T_{ij}\log T_{ij} + \sum_i\lambda_i\left(\sum_j T_{ij} - p_i\right) + \sum_j\mu_j\left(\sum_i T_{ij} - q_j\right).$$

Taking derivative on T_{ij}, we get

$$\frac{\partial \mathcal{L}}{\partial T_{ij}}(T;\lambda,\mu) = D_{ij} + \alpha\left(\log T_{ij} + 1\right) + \lambda_i + \mu_j = 0,$$

which implies

$$\log T_{ij} = -\frac{D_{ij}+\gamma_i+\mu_j}{\alpha} - 1 \implies T_{ij} = \exp\left(-\frac{\gamma_i}{\alpha}-\frac{1}{2}\right)\exp\left(-\frac{D_{ij}}{\alpha}\right)\exp\left(-\frac{\mu_j}{\alpha}-\frac{1}{2}\right).$$

Hence $T = \operatorname{diag}(\mathbf{v})K_\alpha\operatorname{diag}(\mathbf{w})$ where $v_i = \exp\left(-\gamma_i/\alpha - 1/2\right)$, $w_i = \exp\left(-\mu_j/\alpha - 1/2\right)$.

(d) Since $D_{ij} = d(x_i,y_j)^2 = \lim_{t\to 0}\left(-2t\log\mathcal{H}_t(x_i,y_j)\right)$, when α is very small we have

$$\begin{aligned}(K_\alpha)_{ij} &= \exp\left(-\frac{D_{ij}}{\alpha}\right) = \lim_{t\to 0}\exp\left(\frac{2t\log\mathcal{H}_t(x_i,y_j)}{\alpha}\right) \approx \exp\left(\frac{2t\log\mathcal{H}_t(x_i,y_j)}{\alpha}\right)\Bigg|_{t=\alpha/2}\\ &= \exp\left(\log\mathcal{H}_{\alpha/2}(x_i,y_j)\right) = \mathcal{H}_{\alpha/2}(x_i,y_j).\end{aligned}$$

(e) If k is odd, the Lagrange multiplier for $T^{(k)}$ is

$$\begin{aligned}\mathcal{L}(T;\lambda) &= \mathrm{KL}\left(T\|T^{(k-1)}\right) + \sum_i \lambda_i\left(\sum_j T_{ij} - p_i\right)\\ &= \sum_{i,j} T_{ij}\left(\log T_{ij} - \log T_{ij}^{(k-1)}\right) + \sum_i \lambda_i\left(\sum_j T_{ij} - p_i\right).\end{aligned}$$

Taking derivative on T_{ij}, we get

$$\frac{\partial \mathcal{L}}{\partial T_{ij}}\left(T^{(k)};\lambda\right) = \log T_{ij}^{(k)} + 1 - \log T_{ij}^{(k-1)} + \lambda_i = 0,$$

which implies

$$T_{ij}^{(k)} = \exp\left(\log T_{ij}^{(k-1)} - \lambda_i - 1\right) = \exp(-\lambda_i - 1)T_{ij}^{(k-1)}.$$

Hence $T^{(k)} = \operatorname{diag}\left(\widetilde{\mathbf{v}}^{(k)}\right)T^{(k-1)}$ where $\widetilde{v}_i^{(k)} = \exp(-\lambda_i - 1)$. If k is even, similarly we have $T^{(k)} = T^{(k-1)}\operatorname{diag}\left(\widetilde{\mathbf{w}}^{(k)}\right)$ where $\widetilde{\mathbf{w}}^{(k)}$ is of the form $\widetilde{w}_j^{(k)} = \exp(-\mu_j - 1)$. Altogether, by induction on k and notice that the product of two diagonal matrices is again a diagonal matrix, we have $T^{(k)} = \operatorname{diag}\left(\mathbf{v}^{(k)}\right)T^{(0)}\operatorname{diag}\left(\mathbf{w}^{(k)}\right) = \operatorname{diag}\left(\mathbf{v}^{(k)}\right)\mathcal{H}_{\alpha/2}\operatorname{diag}\left(\mathbf{w}^{(k)}\right)$ for some vectors $\mathbf{v}^{(k)}, \mathbf{w}^{(k)} \in \mathbb{R}^n$.

Now we deduce the recurrent relations of $\mathbf{v}^{(k)}$ and $\mathbf{w}^{(k)}$. By the KKT condition on $T^{(k)}$ (suppose k is odd) we have

$$\widetilde{v}_i^{(k)} = \exp\left(-\lambda_i^{(k)} - 1\right) = \frac{T_{ij}^{(k)}}{T_{ij}^{(k-1)}} = \frac{\sum_j T_{ij}^{(k)}}{\sum_j T_{ij}^{(k-1)}} = \frac{p_i}{\sum_j T_{ij}^{(k-1)}}.$$

And by the KKT conditions on $T^{(k-1)}, T^{(k-2)}$ we have

$$\widetilde{w}_i^{(k-1)} = \exp\left(-\mu_i^{(k-1)} - 1\right) = \frac{T_{ij}^{(k-1)}}{T_{ij}^{(k-2)}} = \frac{\sum_j T_{ij}^{(k-1)}}{\sum_j T_{ij}^{(k-2)}} = \frac{\sum_j T_{ij}^{(k-1)}}{p_i}.$$

Hence $\widetilde{\mathbf{v}}^{(k)} = \mathbf{1} \oslash \widetilde{\mathbf{w}}^{(k-1)}$, where the division "$\oslash$" operates element-wise. Similarly we have $\widetilde{\mathbf{w}}^{(k)} = \mathbf{1} \oslash \widetilde{\mathbf{v}}^{(k-1)}$ for even k. Notice that at the initial step $T^{(0)} = \mathcal{H}_{\alpha/2}$ and $\widetilde{\mathbf{v}}^{(1)} = \mathbf{v}^{(1)} = \mathbf{p} \oslash \left(\mathcal{H}_{\alpha/2}\mathbf{w}^{(0)}\right)$ with $\mathbf{w}^{(0)} = \mathbf{1}$. By induction we can show that

$$\begin{aligned}
\mathbf{v}^{(2m+1)} &= \widetilde{\mathbf{v}}^{(2m+1)} \otimes \mathbf{v}^{(2m-1)} = \left(\mathbf{1} \oslash \widetilde{\mathbf{w}}^{(2m)}\right) \otimes \left(\mathbf{p} \oslash \left(\mathcal{H}_{\alpha/2}\mathbf{w}^{(2m-2)}\right)\right) \\
&= \mathbf{p} \oslash \left(\mathcal{H}_{\alpha/2}\mathbf{w}^{(2m-2)} \otimes \widetilde{\mathbf{w}}^{(2m)}\right) = \mathbf{p} \oslash \left(\mathcal{H}_{\alpha/2}\mathbf{w}^{(2m)}\right)
\end{aligned}$$

("$\otimes$" stands for element-wise multiplication), where $\mathbf{v}^{(2m-1)} = \mathbf{p} \oslash \left(\mathcal{H}_{\alpha/2}\mathbf{w}^{(2m-2)}\right)$ is the induction hypothesis and the definitions of $\mathbf{v}(2m+1), \mathbf{w}^{(2m)}$ are

$$\begin{aligned}
\mathbf{v}^{(2m+1)} &= \widetilde{\mathbf{v}}^{(2m+1)} \otimes \widetilde{\mathbf{v}}^{(2m-1)} \otimes \cdots \otimes \widetilde{\mathbf{v}}^{(1)} \\
\mathbf{w}^{(2m+1)} &= \widetilde{\mathbf{w}}^{(2m)} \otimes \widetilde{\mathbf{w}}^{(2m-2)} \otimes \cdots \otimes \widetilde{\mathbf{w}}^{(2)}.
\end{aligned}$$

Similarly we can also show that $\mathbf{w}^{(2m)} = \mathbf{p} \oslash \left(\mathcal{H}_{\alpha/2}\mathbf{v}^{(2m-1)}\right)$ (in fact it should be $\mathcal{H}_{\alpha/2}^t$ in the equation, but note that the heat kernel is symmetric). Therefore the Sinkhorn algorithm can be written as in Algorithm 1.

Algorithm 1 Sinkhorn algorithm for computing W_α

Input: Probability distributions $\mathbf{p}, \mathbf{q}$ on the mesh, parameter α
Output: $W_\alpha(\mathbf{p}, \mathbf{q}) = \alpha \cdot \mathbf{1}^t(\mathbf{p} \otimes \log \mathbf{v} + \mathbf{q} \otimes \log \mathbf{w})$

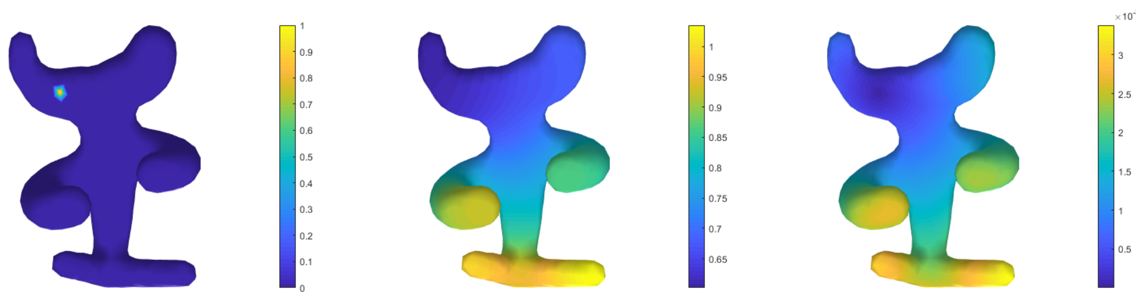
```
Initialize v, w ← 1
for k = 1, 2, 3, ··· (both v, w changes or any other relevant stopping criterion) do
    v ← p ⊘ (H_{α/2} w)
    w ← q ⊘ (H_{α/2} v)
end for
```

After iterating for enough steps, the pair $(\mathbf{v}, \mathbf{w})$ should converge to a fixed point, *i.e.* $\mathbf{v} = \mathbf{p} \oslash (\mathcal{H}_{\alpha/2}\mathbf{w})$, $\mathbf{w} = \mathbf{q} \oslash (\mathcal{H}_{\alpha/2}\mathbf{v})$. Hence the entropy-regularized EMD can be computed as

$$\begin{aligned}
W_\alpha(\mathbf{p}, \mathbf{q}) &= \alpha \cdot \mathrm{KL}(T \| \mathcal{H}_{\alpha/2}) = \alpha \cdot \mathrm{KL}(\mathrm{diag}(\mathbf{v})\mathcal{H}_{\alpha/2}\mathrm{diag}(\mathbf{w}) \| \mathcal{H}_{\alpha/2}) \\
&= \alpha \sum_{i,j} v_i (\mathcal{H}_{\alpha/2})_{ij} w_j \log(v_i w_j) \\
&= \alpha \left(\sum_i v_i \log v_i \sum_j (\mathcal{H}_{\alpha/2})_{ij} w_j + \sum_j w_j \log w_j \sum_i (\mathcal{H}_{\alpha/2})_{ij} v_i \right)
\end{aligned}$$

$$= \alpha \left(\sum_i v_i \log v_i \frac{p_i}{v_i} + \sum_j w_j \log w_j \frac{q_j}{w_j} \right)$$
$$= \alpha \cdot \mathbf{1}^t (\mathbf{p} \otimes \log \mathbf{v} + \mathbf{q} \otimes \log \mathbf{w}).$$

(f) We can take W_α between δ distributions as a vertex-to-vertex distance when α is small. Fix a source point $\mathbf{x}$ (Figure 2a), we compute the distance $W_\alpha(\delta_{\mathbf{x}}, \delta_{\mathbf{y}})$ with any target point $\mathbf{y}$ on the mesh. The results with different choice of α's are shown in Figure 2b, 2c. We see that as α gets smaller, the distance becomes more accurate.



(a) Source $\delta_{\mathbf{x}}$ (b) $W_\alpha(\delta_{\mathbf{x}}, \delta_{\mathbf{y}})$ for $\alpha = 10^{-1}$ (c) $W_\alpha(\delta_{\mathbf{x}}, \delta_{\mathbf{y}})$ for $\alpha = 2 \times 10^{-5}$

Figure 2: W_α between δ distributions as a vertex-to-vertex distance.

Problem 5

Figure 3: Shape interpolation by computing the Wasserstein barycenters.