Homework 4: Vector Fields and Optimal Transport

Due April 24, 2019

Problem 1 (20 points)**.** Suppose that $\gamma : (-1,1) \to S$ is a geodesic segment on a two-dimensional surface $S \subseteq \mathbb{R}^3$. For every $s \in (0,1)$, let $\mathbf{N}(s)$ be a unit vector in $T_{\gamma(s)}S$ that is orthogonal to $\dot{\gamma}(s)$; you can assume $\mathbf{N}(s)$ is a differentiable function. In this problem, you will prove that the mapping $\phi(s,t) := \exp_{\gamma(s)}(t\mathbf{N}(s))$ for $s \in (-1,1)$ and small t is an orthogonal parameterization of a neighborhood of $\gamma(0)$.

(a) Draw an informative picture. What could go wrong if t is allowed to become too large?
(b) Let $\mathbf{E_1} := \frac{\partial\phi}{\partial s}$ and $\mathbf{E_2} := \frac{\partial\phi}{\partial t}$. Show that $\nabla_{\mathbf{E_2}}\mathbf{E_2} = 0$ for all s, t.
Hint: Recall the connection between geodesics, the geodesic equation, and the exponential map. If it is easier, prove (c) before (b).
(c) Show that $\|\mathbf{E_2}\| = 1$ for all (s,t).
Hint: why is this true when s is arbitrary and $t = 0$? Now hold s fixed and show that $\frac{\partial}{\partial t}\|\mathbf{E_2}\|^2 = 0$ for all t.
(d) Show that $\langle \mathbf{E_1}, \mathbf{E_2} \rangle = 0$ for all (s,t).
Hint: why is this true when s is arbitrary and $t = 0$? Now hold s fixed and show that $\frac{\partial}{\partial t}\langle \mathbf{E_1}, \mathbf{E_2} \rangle = 0$ for all t.

Problem 2 (10 points)**.** Suppose $f : \mathbb{R}^3 \to \mathbb{R}$ is a function and that $F : \mathbb{R}^3 \to \mathbb{R}^3$ is a vector field. Verify the following exterior calculus identities:

(a) $\nabla f = (\mathrm{d}f)^\sharp$
(b) $\nabla \cdot F = \star\mathrm{d}\star(F^\flat)$
(c) $\nabla \times F = (\star\mathrm{d}(F^\flat))^\sharp$
(d) $\Delta f = \star\mathrm{d}\star\mathrm{d}f$

Problem 3 (20 points)**.** The *divergence theorem* says that for any smooth vector field $\mathbf{X}$ on a surface S with boundary ∂S, we have

$$\int_S \mathrm{div}(\mathbf{X})\,\mathrm{d}A = \int_{\partial S} \langle \mathbf{X}, \mathbf{N} \rangle\,\mathrm{d}s\,.$$

where dA is the Riemannian area form, $\mathbf{N}$ is a unit vector tangent to S but an outward pointing normal to ∂S, and we must use an arc-length parametrization for ∂S for this equation to hold. Stokes' Theorem says that for any differential k-form ω and $(k+1)$-dimensional submanifold $c \subseteq S$ we have

$$\int_{\partial c} \omega = \int_c \mathrm{d}\omega\,.$$

In this problem, you will show that Stokes' Theorem implies the divergence theorem, using a well-chosen ω. For this problem, you can assume that the identity you proved in 2(b) holds on surfaces.

(a) Explain why $\operatorname{div}(\mathbf{X})\mathrm{d}A$ can be put in the form $\mathrm{d}\omega$ for some form ω, and what is ω?

(b) Apply Stokes' Theorem to $d\omega$ and S itself. We thus get $\int_S \operatorname{div}(\mathbf{X}) dA = \int_{\partial S} \omega$. To develop the right hand side further, you must know how to evaluate the "line integral" $\int_{\partial S} \omega$. Suppose that we can parametrize the boundary ∂S by arc-length as a curve $\gamma : [0, \ell] \to S$ with tangent vector $\mathbf{T}(s) := \dot{\gamma}(s)$. Now $\int_{\partial S} \omega$ is defined to be $\int_0^\ell \omega(\mathbf{T}(s)) ds$. Show that $\omega(\mathbf{T}) = \langle \mathbf{X}, \mathbf{N} \rangle$ where $\mathbf{N}$ is the vector obtained by rotating $\mathbf{T}$ counterclockwise by $\pi/2$.
Hint: For this problem, you might want to use that in 2 dimensions the Hodge star operates of a one form ω acts on vectors as $\star\omega(\mathbf{v}) = \omega(\mathrm{Rot}_{\pi/2}\mathbf{v})$.

For the coding assignment, we've provided you with helper functions for loading and plotting triangle meshes. Before starting the homework, take a look at `utils/` for `MATLAB` code or `utils.jl` for `Julia` code to familiarize yourself with the syntax.

Problem 4 (40 points)**.** In this problem you will implement the Sinkhorn method for approximating the "earth mover's distance" (EMD) between two probability distributions on a triangle mesh.

(a) Suppose we are given a pairwise squared distance matrix $D \in \mathbb{R}^{n\times n}$. D_{ij} measures the distance between bins i and j of a histogram with n bins. For example, $D_{ij} = \|\mathbf{x}_i - \mathbf{y}_j\|_2^2$ for given point sets $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and $\mathbf{y}_1, \ldots, \mathbf{y}_n$. The EMD between histograms $\mathbf{p}$ and $\mathbf{q}$ is defined as

$$W(\mathbf{p}, \mathbf{q}) = \begin{cases} \min_{T\in\mathbb{R}^{n\times n}} & \sum_{i=1}^n \sum_{j=1}^n T_{ij} D_{ij} \\ \text{subject to} & T_{ij} \geq 0, \quad \forall i, j \in \{1, \ldots, n\} \\ & \sum_j T_{ij} = p_i, \quad \forall i \in \{1, \ldots, n\} \\ & \sum_i T_{ij} = q_j, \quad \forall j \in \{1, \ldots, n\}. \end{cases} \tag{1}$$

Explain what $W(\mathbf{p}, \mathbf{q})$ measures about the difference between $\mathbf{p}$ and $\mathbf{q}$. Compare it to other discrepancy measures between probability distributions such as the Kullback-Leibler divergence.

(b) EMD is difficult to compute when n is large. An alternative is the *entropy-regularized EMD* introduced by Marco Cuturi in **Sinkhorn Distances: Lightspeed Computation of Optimal Transport Distances**. The Sinkhorn distance between $\mathbf{p}$ and $\mathbf{q}$ is given by

$$W_\alpha(\mathbf{p}, \mathbf{q}) = \begin{cases} \min_{T\in\mathbb{R}^{n\times n}} & \sum_{i=1}^n \sum_{j=1}^n T_{ij} D_{ij} + \alpha \left(\sum_{ij} T_{ij} \ln T_{ij} - 1 \right) \\ \text{subject to} & T_{ij} \geq 0, \quad \forall i, j \in \{1, \ldots, n\} \\ & \sum_j T_{ij} = p_i, \quad \forall i \in \{1, \ldots, n\} \\ & \sum_i T_{ij} = q_j, \quad \forall j \in \{1, \ldots, n\}. \end{cases} \tag{2}$$

Define a matrix K_α in terms of D and α so that the objective for computing $W_\alpha(\mathbf{p}, \mathbf{q})$ can be written as $\alpha \cdot \mathrm{KL}(T \| K_\alpha)$, where the KL divergence between $A, B \in \mathbb{R}_+^{n\times n}$ is

$$\mathrm{KL}(A\|B) = \sum_{ij} A_{ij} \ln \frac{A_{ij}}{B_{ij}}$$

(c) Show that the optimal matrix T in the minimization for W_α can be written as $T = \text{diag}(\mathbf{v})K_\alpha\text{diag}(\mathbf{w})$ for some $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$.
Hint: Use Lagrange multipliers; it may be useful to argue that the $T_{ij} \geq 0$ constraint is no longer necessary after entropic regularization.

(d) So far, we have assumed that we have a pairwise distance matrix D. Let's specialized to a triangle mesh, and define $D_{ij} = d(x_i, d_j)^2$ where x_i and x_j are vertices of the mesh and d denotes geodesic distance. Computing the full pairwise geodesic matrix D is very expensive. In **Convolutional Wasserstein Distances: Efficient Optimal Transportation on Geometric Domains**, Solomon et al. propose an alternative solution.

The heat kernel $\mathcal{H}_t(x, y)$ gives the amount of heat diffusion between $x, y \in M$ after time $t > 0$. In particular $\mathcal{H}_t(x, y)$ solves $\partial_t f_t = \Delta f_t$ with initial condition f_0 through the map

$$f_t(x) = \int_M f_0(y)\mathcal{H}_t(x, y)\mathrm{d}y. \tag{3}$$

We have provided an implementation of heat diffusion in the function `heatDiffusion` which you will use in the final part of this problem.

Varadhan's formula states that the distance on the manifold $d(x, y)$ can be recovered by transferring heat from x to y over a short time interval:

$$d(x, y)^2 = \lim_{t \to 0}[-2t \ln \mathcal{H}_t(x, y)]. \tag{4}$$

Argue that K_α can be approximated as $\mathcal{H}_{\alpha/2}$.

(e) The Sinkhorn algorithm for computing W_α proceeds as follows:

1 Initialize $T^0 \equiv H_{\alpha/2}$.

2 For $i = 1, 2, 3, \ldots$

i. If i is odd, compute

$$T^{(i)} \equiv \begin{cases} \arg\min_{T \in \mathbb{R}^{n\times n}} & \mathrm{KL}(T \| T^{(i-1)}) \\ \text{subject to} & \sum_j T_{ij} = p_i \quad \forall i \in \{1, \ldots, n\}. \end{cases} \tag{5}$$

ii. If i is even, compute

$$T^{(i)} \equiv \begin{cases} \arg\min_{T \in \mathbb{R}^{n\times n}} & \mathrm{KL}(T \| T^{(i-1)}) \\ \text{subject to} & \sum_i T_{ij} = q_j \quad \forall j \in \{1, \ldots, n\}. \end{cases} \tag{6}$$

Show that each $T^{(i)}$ can be written $T^{(i)} = \text{diag}(\mathbf{v}^{(i)})H_{\alpha/2}\text{diag}(\mathbf{w}^{(i)})$ for some vectors $\mathbf{v}^{(i)}, \mathbf{w}^{(i)} \in \mathbb{R}^n$. Write the steps of the Sinkhorn algorithm in terms of these vectors. Your algorithm should involve only matrix-vector multiplication and per-element operations on vectors (multiplication/division).

(f) Implement the Sinkhorn algorithm in `emd.m` (`emd.jl`) including reasonable stopping criteria. Try several probability distributions on multiple meshes. The example in the starter code should compute geodesic distances from point 1 to all other points on the mesh assuming you have coded everything correctly.

Problem 5 (10 points extra credit)**.** Reproduce Figure 1 in **Convolutional Wasserstein Distances: Efficient Optimal Transportation on Geometric Domains**.