Research Statement

Martin Molina Fructuoso

My research focuses on using mathematical tools to tackle problems related to data science, machine learning, and numerical optimal transportation. I am particularly interested in deepening mathematical understanding in these areas of application and in developing and adapting the mathematical tools needed for this purpose.

More specifically, I use methods from the theory of partial differential equations (PDE), calculus of variations, convex analysis, non-smooth analysis, and optimal transportation to study problems in which one needs to relate discrete and continuum settings. I have focused on problems from several related areas: non-convex optimization problems related to machine learning (implicit regularization in nonconvex optimization problems [19]), mathematical problems coming from data science (PDE-based variational methods for statistical depths [24],[25], [23]) and applied optimal transportation (regularity of optimal transportation plans for rough measures [11]). Some of the key underlying themes of my work is the use of variational methods and the connection of continuum variational problems on smooth measures and variational problems posed on approximating discrete objects.

Specific problems I have worked on and some directions I plan to pursue in the future include:

1. **Implicit regularization in accelerated descent methods**

   As a Postdoctoral Associate at Brandeis University, I have been working in nonconvex optimization problems coming from machine learning. More specifically, in a collaboration with Prof. Tyler Maunu, we have shown in [19] that there is an implicit regularization effect in the phase retrieval problem for Polyak's Heavy Ball and Nesterov Accelerated Gradient methods when the measuring vectors are chosen from a Gaussian distribution. That is, the method converges to a global minimizer at an accelerated rate even though the loss function is non-convex. This result can be understood by considering than for this type of sampling, the corresponding loss function is strongly convex and smooth in expectation. We have also verified these findings experimentally and observed that the methods do achieve the expected accelerated rate in practice.

   We are currently working in extending this result to more general methods that might depend on the gradient at all previous time steps, and in extending the proof of implicit regularization to the low-rank matrix completion problem. Namely, given a low rank matrix and a subset of observations of its entries we would like to reconstruct the original matrix as closely as possible.

2. **PDE-based variational methods for statistical depths**. As a Postdoctoral Research Scholar at North Carolina State University, in collaboration with Prof. Ryan Murray, one of the projects I worked on was a PDE-based formulation for the study of the Tukey statistical depth. I have recently shown in a paper with Murray [24] that, for reasonably smooth probability densities, the Tukey depth satisfies an equation of the Hamilton-Jacobi type at the points of differentiability. We have shown as well that the viscosity solution of this equation is well-posed in any dimension and that it coincides with the Tukey depth of the distribution for probability distributions with compact and convex supports on the plane. More recently, Prof. Murray and I have been working [23] on an alternative numerical algorithm that relies on geometric characterization of Tukey depths and does not resort to finite difference schemes. This method allows us to approximate the Tukey depth of an arbitrary distribution in any dimension and the approximations agree with closed form solutions for the very few cases where they are available.

On a previous project with Prof. Murray [25] we proposed and studied a novel concept of statistical depth based on optimal control. A strong motivation to introduce this concept is to replace the nonlocal equation for the Tukey depth introduced in [24] with a local equation of the eikonal type. Our approach relies on solving our eikonal equation on graphs numerically, an approach that allows us to directly use this depth for probability distributions on manifolds and in general discrete settings. An example of the notion of centrality that can be defined on the digits of the MNIST set using our scheme is shown in Figure 1.

Future research directions in connection with Tukey depths include proving that, in any dimension, viscosity solutions of the Hamilton-Jacobi equation we derived in [24] correspond to the Tukey depth for log concave measures with convex support. Prof. Murray and have found that this correspondence is true in two dimensions and that a result based on Brunn-Minkowski theorem for convex bodies could be useful to extend this equivalence to an arbitrary dimension. Our alternative method to calculate Tukey depths based on the Hamilton-Jacobi equation introduced in [24] uses derivatives of moments of slices of the measure and relies on approximating this measure as a sum of Gaussian distributions of small bandwidth. Another research direction I would like to pursue is to use Monte-Carlo methods to calculate derivatives of the moments of slices of the measure directly and in this way eliminate the error introduced by the Gaussians' bandwidth.

3. **Regularity of optimal transportation plans for rough measures**. My PhD thesis work at UMD concerned the regularity of optimal transportation plans for measures that are not necessarily absolutely continuous. In [11], a joint work with Jabin and Mellet, we derived quantitative $\mathcal{C}^1$ regularity estimates for any Kantorovich potential between measures that are only absolutely continuous up to a certain scale, for instance discretizations of absolutely continuous measures on grids of given widths. These results extend the classical result of regularity of optimal transportation maps for measures that are absolutely continuous and open the doors to quantifying the relation between the optimal transportation problem posed in terms of absolutely continuous measures and a discrete optimal transportation problem that approximates it, a direction that is of interest for numerical optimal transportation.

   A future research direction that is closely related to both my work on regularity of optimal transportion for rough measures and statistical depths is the study of some analytical properties of the Wasserstein statistical depth [7], defined in terms of Tukey depths and optimal transportation. The Wasserstein statistical depth has a variational structure and I plan to use tools from optimal transportation to find convergence rates between the discrete and continuum scales, and to deepen mathematical understanding of the algorithms used to calculate this statistical depth.

This is not an exhaustive list of my research plans, and I would be excited to develop new colaboration projects with faculty at your department.

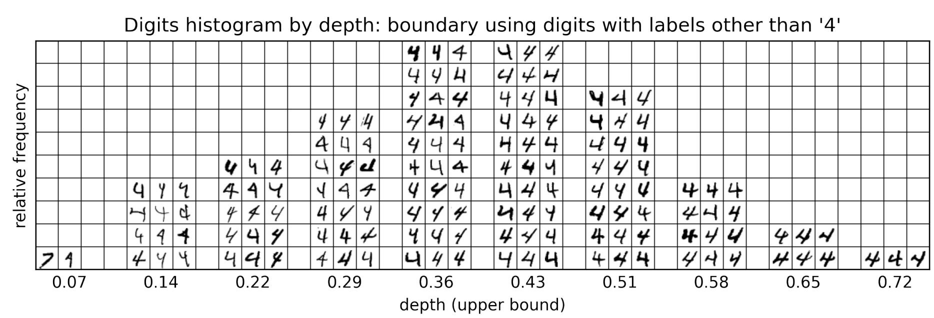


Figure 1: This figure shows an example of the notion of centrality that can be established using statistical depths. The histogram shows the digits labeled '4' in the MNIST classified into groups according to the level sets of the depth we propose in [25]. For each class the number of samples depicted depends on the size of the class. We can observe that the legibility of the handwriting increases with depth.

1 Implicit regularization in accelerated descent methods

The phase retrieval problem consists in finding $x_* \in \mathbb{R}^n$ from a set of measurements $y_i = (a_i^T x_*)$ with respect to a set $\{a_i\}_{i=1}^m \subset \mathbb{R}^d$. For this purpose, we consider the minimization of the following objective function:

$$f(x) = \frac{1}{4m} \sum_{i=1}^{m} ((a_i^T x)^2 - y_i^2)^2. \tag{1}$$

This problem is nonconvex in general which makes it theoretically hard to solve. One strategy to address this issue is to modify the standard gradient descent minimization procedure, namely $x^{t+1} = x^t - \eta_t \nabla f(x^t), t \geq 0$ to explicitely force convergence to a global minimizer. For instance, one can add a penalization term $R(x)$ to the descent update rule: $x^{t+1} = x^t + \eta_t(\nabla f(x^t) + \nabla R(x^t))$.

However, it is known that for certain choices of the vectors $\{a_i\}$, for instance a_i sampled from a normal distribution $\mathcal{N}(0, I)$, using standard gradient descent to minimize the loss function (1) works exceptionally well [17]. This is referred to as a bias in the algorithm or implicit regularization in the algorithm for the nonconvex minimization problem (1).

Recently, in collaboration with Prof. Tyler Maunu [19], we have shown that this phenomenon extends to other accelerated gradient methods for the nonconvex minimization problem (1), more specifically we have obtained accelerated convergence rates with certain probability guarantee for the Polyak's Heavy Ball and Nesterov's Fast Gradient methods when the vectors $\{a_i\}$ are sampled from a Gaussian distribution $\mathcal{N}(0, I)$. These two methods are defined as

$$x^{t+1} = x^t - \eta \nabla f(x^t) + \beta(x^t - x^{t-1}) \tag{2}$$

where $x^1 = x^0$ and

$$x^{t+1} = x^t - \eta \nabla f(x^t + \beta(x^t - x^{t-1})) + \beta(x^t - x^{t-1}). \tag{3}$$

Our result can be stated as:

Theorem 1.1. *Let x_* be a fixed vector and $a_i \overset{\text{iid}}{\sim} \mathcal{N}(0, I)$. Provided that $0 \leq \eta \lesssim rac1\log n \|x_0\|^2$ for some initialization x_0, $m \gtrsim n \log n$, $\beta = \dfrac{C\sqrt{\log n} - \sqrt{1/2}}{C\sqrt{\log n} + \sqrt{1/2}}$, for some sufficiently large constant C, then there exists a constant $0 < \varepsilon < 1$ such that with probability at least $1 - \mathcal{O}(mn^{-5})$, (2) and (3) achieve the following convergence rate*

$$\mathit{dist}(x^t, x_*) \leq \varepsilon(1 - \sqrt{\eta}\|x_*\|^2/2)^t \|x_*\|$$

where $\mathit{dist}(x, x_*) = \min\{\|x - x_*\|, \|x + x_*\|\}$.

The essential part of the proof is to show that the iterations produced by (2) remain, with certain probability, in a nice region in which f can be considered as strongly convex and smooth for vectors $a_i \overset{\text{iid}}{\sim} \mathcal{N}(0.I)$. In fact, using rough bounds, when a_i are sampled from a Gaussian, we have that $I \preceq \nabla^2 f(x) \preceq 10I$ in expectation.

Future directions

- **Implicit regularization in multi-step accelerated methods**. Both Heavy Ball and Nesterovs method are examples of more general accelerated methods that consider the value of the gradient of f at previous timesteps. For instance, Heavy Ball iterations can be written as

$$x^{t+1} - x^t = -\eta \sum_{i}^{t} \beta^{t-i} \nabla f(x^t).$$

As a future research direction, I plan to study the implicity regularization of methods of the form

$$x^{t+1} = x^t - \eta \sum_{i=1}^{t} \alpha_i^t \nabla f(x^i)$$

for some sequence of weights $\{\alpha_i^t\}_{i=1}^{\infty}{}_{t=1}^{\infty}$, which would be a natural generalization of the results described in this section.

- **Implicit regularization in low-rank matrix completion**. I also expect to prove an analogous result to Theorem 1.1 for the problem of low-rank matrix completion, namely where we want to recover a low rank matrix $M_* \in \mathbb{R}^{n\times n}$ from an incomplete subset of its entries. In mathematical terms, if M_* is a positive semidefinite, r-rank matrix (with $r \ll n$), i.e. $M_* = X_* X_*^T$ with $X_* \in \mathbb{R}^{n\times r}$, and we know some entries

$$Y_{i,j} = [M_*]_{ij} = [X_* X_*^T]_{i,j}, \qquad (i,j) \in S$$

for a set S of m indices, we want to solve the minimization problem

$$\min_{X\in\mathbb{R}^{n\times r}} f(X) = \frac{n^2}{4m} \sum_{(i,j)\in S} (Y_{i,j} - e_i^T X X^T e_j)^2,$$

where $\{e_i\}_{i=1}^n$ is the canonical vector basis in $\mathbb{R}^n$.

2 PDE-based Variational Methods for Statistical Depths

2.1 Tukey depth and HJB equations

As a Postdoctoral Research Scholar at NCSU, in colaboration with Prof. Ryan Murray, I worked on PDE-based methods for problems related to machine learning. In particular, I studied PDE-based variational approaches applicable to a wide variety of statistical depths defined in continous and discrete settings.

In simple terms, statistical depths are functions that measure how deep a point is within a given point cloud and are useful for defining medians. The level sets of a statistical depth provide a natural way to establish an order in a dataset. For example, data depths can be used to detect and handle outliers in a data set for classification purposes [28] [12] and for functional data analysis [4]. An instance of the classification result that can be achieved on the MNIST dataset with one of our PDE-based statistical depths is shown in Figure 1. We have used all of the 70000 hand-written digits in the set, considered as points in $\mathbb{R}^{28\times 28}$.

Statistical depths are commonly defined on sets of data points, and one of the most widely used is the Tukey depth. Generalizations of the Tukey depth to a functional setting have also been used to study brain data and handwriting recognition [5]. For a given point cloud $\{x_i\}_{i=1}^N \subset \mathbb{R}^d$ and some discrete measure $\mu_N = \sum_{i=1}^N \alpha_i \delta_{x_i}$ the Tukey depth on the point cloud is defined as

$$d_T(x_i, \mu_N) = \min\{\mu_N(H) | H \text{ is a closed halfspace with } x_i \in H\}. \tag{4}$$

It is reasonable to expect that, when the number of data points increases, this discrete optimization problem approaches the following continuum variational problem. For a general probability distribution μ in $\mathbb{R}^d$ the Tukey depth at the point x is defined as

$$d_T(x, \mu) = \inf\{\mu(H) | H \text{ is a closed halfspace with } x \in H\}.$$

This formulation has already been used [6, 29, 18] in the context of empirical approximations to the continuum problem. The goal of reformulating the depth in this way is to use tools from the calculus of variations and the theory of partial differential equations to gain additional insight on the problem and to provide alternative ways to calculate the Tukey depth. In a recent paper with Murray [24], we have shown that the Tukey depth $d_T(\cdot,\mu)$ for an absolutely continuous measure with a continuous density ρ can be related to viscosity solutions [3] of the following equation of the Hamilton-Jacobi type:

$$H(x,\nabla u(x)) = |\nabla u(x)| - \int_{(y-x)\cdot\frac{\nabla u}{|\nabla u|}=0} \rho(y)dH^{d-1}(y) = 0. \tag{5}$$

One of the main results of our recent paper [24] demonstrates that the more general theory of viscosity solutions can be adapted for this problem in certain situations. We have proved the following result [24]:

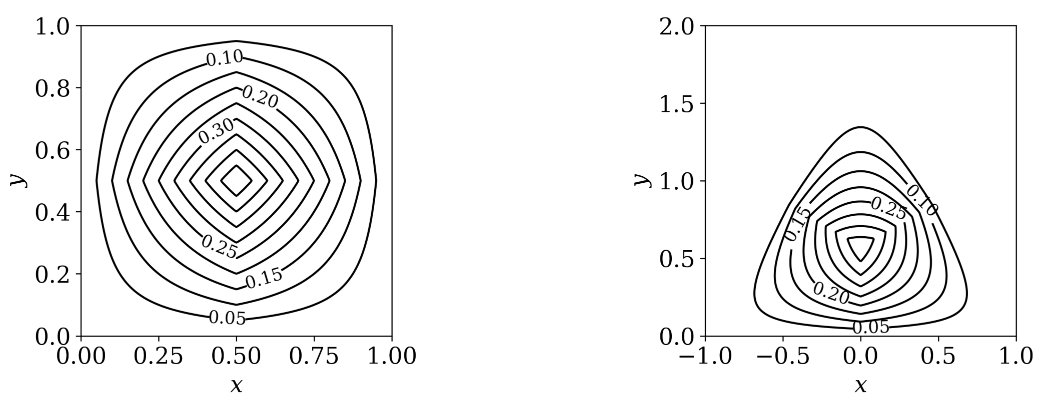
Theorem 2.1. *Assume that μ is an absolutely continuous measure with a continuous density ρ whose support is $\overline{S}$ for an open and bounded set $S \subset \mathbb{R}^d$. Let us define $\Omega = \overline{co}(S)$, the closure of the convex hull of S, and let us assume that $\int_{p\cdot(\xi-x)=0}\rho(\xi)dH^{d-1} \geq \delta(x)$ for any $p \in \mathbb{R}^n$ and any $x \in \Omega$, with $\delta(x) = 0$ only if $x \in \partial\Omega$. Then there exists a unique viscosity solution of equation (5) with $u = 0$ on $\partial\Omega$.*

Additionally, if $S \subset \mathbb{R}^2$ is convex and μ is uniformly distributed, then the Tukey depth $d_T(\cdot,\mu)$ coincides with this viscosity solution. In any other case, including when $\mathbb{R}^d$, the viscosity solution provides an upper bound for the Tukey depth $d_T(\cdot,\mu)$.

The structure of equation (5) bears a close resemblance to that of the Eikonal equation. Recasting the problem of calculating the Tukey depth as a PDE opens the possibility of applying a wide variety of numerical techniques for Hamilton-Jacobi equations and viscosity solutions of nonlinear partial differential equations. For instance, in a recent paper [2] a finite difference scheme to calculate the Tukey depth of measures with continuous densities was constructed using formulation (5) and the theoretical guarantee that the equation is well posed provided by Theorem 2.1. Having alternative approaches to calculate Tukey depths is valuable because the methods used to solve the optimization problem (4) are combinatorial in nature [16, 22] and costs of the order of the order $O(N^{d-1}\log N)$ that does not scale well in the number of data samples [15].

2.2 Eikonal depth

Prof. Murray and I have proposed [25] the following depth based on control theory and the eikonal equation: For probability distributions μ with continuous densities ρ with support in all of $\mathbb{R}^d$ (alternativerly,



(a) Uniform distribution on a square

(b) Uniform distribution on triangle

Figure 2: Shown in the picture are the level sets of the Tukey depth on a square and a triangle for uniform distributions.

with compact support $\overline{S} \subset \mathbb{R}^d$) and a non-decreasing function ϕ that maps $\mathbb{R}^+$ to $\mathbb{R}^+$, we define the eikonal depth $d_{eik}(x, \mu)$ as the viscosity solution of the equation

$$|u(x)| = \phi(\rho(x))$$

with boundary conditions $\lim_{\|x\| \to \infty} u(x) = 0$ (alternatively, with $u = 0$ on ∂S).

We are mainly interested in the cases when $\phi(s) = s^{1/d}$ and $\phi(s) = s$. In the latter and for a density with support in all of $\mathbb{R}^d$, calculating $d_{eik}(x, \mu)$ is equivalent to solving the optimal control problem

$$\inf_{\mathcal{U}_x} \int_0^\infty \rho(x(\tau))|\dot{x}(\tau)|d\tau$$

where $\mathcal{U}_x$ is the set of paths starting at x and such that $\lim_{\tau \to \infty} x(\tau) = \infty$. This depth can therefore be interpreted as the minimum amount of mass a particle needs to go through to escape to infinity. This equivalent formulation of the eikonal depth has allowed us to prove [25] properties such as the existence of local maxima of the depth near local maxima of the density, which is desirable for multimodal distributions.

The study of variational problems on graphs and their continuum limits have received a lot of attention in recent years [33, 34] due to their various applications in machine learning, among which are supervised and unsupervised learning on data points. In our paper [25] we calculate the eikonal depth using a discrete scheme on graphs whose generalization to densities on manifolds is immediate. We consider a set of data points $\{X_i\}_{i=1}^n \subset \mathbb{R}^d$ and construct a weighted graph with weights given by

$$w_{ij} = \frac{\sigma}{h^d n} \eta \left(\frac{|X_i - X_j|}{h} \right)$$

where $\eta : \mathbb{R} \to \mathbb{R}$ is a kernel that describes the degree to which two points are related depending on the distance that separates them, h is a kernel bandwith and σ is a normalizing constant. A heuristic calculation [25] shows that it makes sense to use the following difference scheme

$$\rho_i = \sum_{j \sim i} h^{-2} w_{ij} (0, u_i - u_j)_+^2.$$

2.3 Alternative Geometric Method to Calculate Tukey depth

In a more recent collaboration with Prof. Murray, I have worked [23] on an alternative method to approximate the Tukey depth of an arbitrary distribution that does not resort to finite difference schemes and uses instead purely geometric characterizations of Tukey depth that are related to convex geometry.

We can think of convex sets as uniformly distributed measures with convex support. In convex geometry, the floating body $K_{[\delta]}$ of index $\delta > 0$ of a convex set K is defined as the nonempty convex subset of K such that any supporting hyperplane to $K_{[\delta]}$ cuts off a set of volume δ from K.

When the floating body of K exists, a supporting hyperplane $\mathcal{H}$ to $K_{[\delta]}$ touches the boundary of of $K_{[\delta]}$ at exactly one point x, the center of mass of the slice $\mathcal{H} \cap K$. Using the formula for the Tukey depth that we introduced in [24] we can show that for an arbitary continuous density a necessary condition for a hyperplane $\mathcal{H}$ to realize the minimum in the Tukey depth at a point x is that the centroid of the intersection of the plane and the support of the density is equal to x. Thus, for a uniform density with convex support, when the floating bodies of its support exist for all δ they coincide with the level sets of the Tukey depth.

Floating bodies do not exist in general for any convex set, but it is know [21] that if the set K is symmetric, namely both x and $-x$ are in K, and its boundary is C^2 then its floating bodies exists for all $0 \leq \delta < 1/2$. Prof Murray and I have been able to obtain a similar result in [23] for log concave densities with support in $\mathbb{R}^d$ in the sense that if ρ is a log concave density that is symmetric, then all of the subsets constructed by cutting off an amount of mass equal to δ are convex and correspond to the level sets of the Tukey depth of

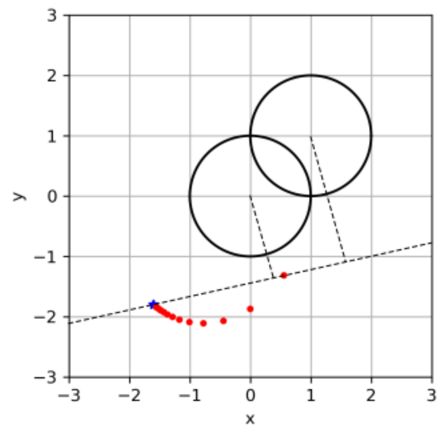


Figure 3: This figure shows x_c for consecutive iterations of our method for the sum of two Gaussian densities with $\sigma = 1$ and means $(0, 0)$ and $(1, 1)$

ρ. Also, to find the hyperplane that realizes the minimum in the definition of Tukey depth at a point x we only need to find $p \in \mathbb{S}^{d-1}$ so that

$$F(x, p) = \int_{p\cdot(y-x)=0} \rho(y)(y - x)d\mathcal{H}^{d-1} = 0.$$

We also have devised an iterative numerical method in [23] to calculate the p that realizes the Tukey depth at the point x based on this result. Starting from an inital guess (x_0, p_0) and calling (x_c, p_c) the current calculated values, we can add the correction $\Delta p = (-D_p F(x_c, p_c))^{-1} D_x F(x_c, p_c)(x - x_c)$ at each time step. The next value of x_c is then calculated by finding the centroid of the slice of ρ with normal vector p_c. This procedure is illustrated in Figure 3 for a sum of two Gaussian densities.

Our result in [23] justifies the use of this method for log concave measures with support in all of $\mathbb{R}^d$. However, we have found that our method approximates the Tukey depth to great accuracy for a much wider class of densities, as long as they are symmetric, by approximating them as

$$\bar{\rho}(x) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{(2\pi)^{d/2}\sigma^d} \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \tag{6}$$

where x_i are samples from ρ. Examples of densities that can be treated using this approximation include uniform densities on convex polygons and arbitrary mixtures of Gaussians.

Future directions

- **Tukey Depths**. I plan to study the stability of solutions of (5) with respect to the input measure μ, and use these results to study the convergence of the discrete Tukey depth to the continuous Tukey depth. I am particularly interested in studying the relation between the Tukey depth for a discrete measure and the Tukey depth for the corresponding mollified measure, and also in deriving the rates of convergence of a minimizer of the discrete problem to a minimizer of the continuous approximation of the problem.
- **Convex Geometry and Tukey depth as a viscosity solution in higher dimensions**. The level sets of the Tukey depth of a uniform distribution with compact and convex support have been studied in convex geometry under the name of floating bodies. A classical result in convex geometry by Buseman, which is a counterpart to Brunn-Minkowski theorem, states that the intersection body of an origin-symmetric

convex body is also convex [8]. Attempting to extend our result on the equivalence of viscosity solutions of our PDE model to the Tukey depth to higher dimensions naturally leads to the need for a similar result to Buseman's. Prof. Murray and I have found that Buseman's result still holds for convex bodies that just satisfy a balanced-moments condition but are not necessarily symmetric with respect to the origin. This result will be the base to show that Tukey depths coincide with the unique viscosity solution of (5) in arbitary dimensions.

- **Geometric Method to calculate Tukey depths**. The expressions for D_xF, D_pF, and those used to calculate the centroids used in our algorithm cannot be solved exactly for densities in general. In our experiments we have instead approximated the measures as sums of Gaussian densities (6). A direction I plan to explore is to calculate these expresions using Monte-Carlo methods to eliminate the intrinsic error that the bandwidths of the Gaussians cause.

  Additionally, I plan to extend the result of Meyer-Reissner [21] to probability densities that are log concave, have convex support, and whose boundary is smooth (C^2). Namely, I would like to prove that the level sets of the Tukey depth for such measures are all smooth and are fully characterized by the fact that supporting hyperplanes touch them exactly at the centroid of the slice obtained as the intersection of the hyperplane and the support of the measure.

3 Regularity of Optimal Transportation Plans for Rough Measures

Another part of my research concerns the regularity of optimal transportation for general measures. Optimal transportation have been used in a variety of applications in recent years. For instance, it is used in biomedical imaging to evaluate the evolution of pictures of an organ over time [9] [10]. Another application of optimal transportation is super resolution [14], where a high resolution image is generated from a low resolution one by solving a minimization problem on a set of transformations between images in a training set.

These applications involve the optimal transportation problem for the quadratic cost and are based on the approximation of absolutely continuous measures by sums of point masses. Given two discrete measures $\mu = \sum_i a_i\delta_{x_i}, \nu = \sum_j b_j\delta_{y_j}$ and the matrix cost $c_{ij} = \|x_i - y_j\|^2$ the optimal transportation problem consists of finding

$$\arg\min \sum_i \sum_j c_{ij}\gamma_{ij} \tag{7}$$

among all matrices γ with $\gamma_{ij} \geq 0$ that satisfy $\sum_j \gamma_{ij} = a_i, \sum_i \gamma_{ij} = b_j$. More generally, given $X, Y \subset \mathbb{R}^d$ and probability measures μ, ν in X, Y the optimal transportation problem for the quadratic cost is to find

$$\arg\inf \int \|x-y\|^2 d\pi \tag{8}$$

among all π probability measures in $X \times Y$ with marginals μ and ν (i.e. $\pi(A\times Y) = \mu(A), \pi(X \times B) = \nu(B)$). For any measures μ and ν, the existence of an optimal plan π that solves Kantorovich problem (8) is guaranteed [35]. Although π is not necessarily unique, its support is contained in the graph of the subdifferential of a convex function ψ (referred to as a Kantorovich potential [35])

$$\operatorname{supp}(\pi) \subset \{(x, y) \in X \times Y | y \in \partial\psi(x)\}.$$

It is a classical result that if μ is absolutely continuous, the Kantorovich potential is unique up to a constant and $\pi = (Id \times \nabla\psi)_\# \mu$ [1] [20] .

The importance of regularity of solutions of this problem for discrete measures is highlighted in a recent work by Cuturi [31]. The solution to the linearly constrained problem (7) might in principle be dense, causing the problem to be too costly to solve, even if we expect that the support of the optimal plan is close to being the graph of a function if the discrete measures are approximations of absolutely continuous measures. Based on the results obtained in [31], using potentials that are strictly convex and C^1-continuous is necessary for efficient numerical calculations for the fully discrete problem.

My contribution to regularity results of the solution to problem (8) has been to prove in [11] that the Kantorovich potential ψ is $\mathcal{C}^1$ even if μ and ν are not necessarily absolutely continuous. As there is no analogous of Monge-Ampere equation for ψ in this general setting, Caffarelli's classical regularity theory for the problem is no longer applicable. However, by using the optimal transportation problem directly I proved the following regularity result. For open, bounded, and convex sets $\Omega_1, \Omega_2 \subset \mathbb{R}^2$ I considered measures μ, ν that satisfy

Assumption 1. *Assume there are constants $h_1, h_2 > 0$ and $\lambda_1, \lambda_2 > 0$ such that μ and ν satisfy*

$$\mu(R) \leq \lambda_2 |R| \text{ and } \frac{|R'|}{\lambda_1} \leq \nu(R')$$

for any rectangles $R \subset \Omega_1, R' \subset \Omega_2$ with dimensions at least h_1 and h_2 in every direction for R and R' respectively.

Assumption 1 can be interpreted as absolute continuity and boundedness for μ and ν up to the resolution of the scale. For instance, for a pointed partition $\{U_i, x_i\}_{i=1,..,N}$ and an absolutely continuous measure $\mu' = f dx \leq \lambda_2'$ the measure $\mu = \sum_{i=1}^{N} \mu'(U_i)\delta_{x_i}$ satisfies Assumption 1 provided h_1 is small enough (for some λ_1, λ_2 depending on λ_1', λ_2' and the geometry of the partition). This is a more general discrete approximation to a measure than the computational method for discrete approximations of measures considered in [30].

Let Ω_i^δ be the set of points in Ω_i at a distance at least δ away from the boundary, then I proved that [11]:

Theorem 3.1 ((C^1 regularity of ψ))**.** *Let ψ be a Kantorovich potential for the problem (8) between measures that satisfy Assumption 1 and let $\Omega_2^\delta = \partial\psi(\Omega_1^\delta)$. There are functions $\rho(\ell), \rho_1(\ell), \rho_2(\ell)$ monotone increasing, with limit 0 when $\ell \to 0^+$, that depend only on $\delta, \lambda_1, \lambda_2, diam(\Omega_2)$ and diam $(\partial\psi(\Omega_1^\delta))$ such that for all $(x, x') \in \Omega_1^\delta \times \Omega_1^\delta$ we have*

$$|y - y'| \leq \max(\rho(|x - x'|), \rho_1(h_1), \rho_2(h_2)) \qquad \text{for all} \qquad y \in \partial\psi(x), y' \in \partial\psi(x')$$

Theorem 3.1 states that, up to scales that depend on h_1, h_2, ψ is C^1. The key concepts in the proof of theorem 3.1 are convex duality and strict convexity. Under Assumption 1 on the measures μ and ν, it can be shown that the conjugate of ψ:

$$\varphi(y) = \sup_{x \in \Omega_1} (x \cdot y - \psi(x))$$

is strictly convex up to scales that depends on h_1, h_2. A classical result [32] states that strict convexity of a convex function f at a point y implies the differentiability of its conjugate at the point $x \in \partial f(y)$. Theorem 3.1 can be seen as an analogous quantitative result for ψ.

Future Directions

- For classification purposes some authors, such as in [26], [26], and [13], have looked into linear embeddings of the space of probability measures into a space of functions. For a reference absolutely continuous measure $\sigma \in \mathcal{P}(\mathbb{R}^d)$ we can look at the embedding $F_\sigma : \mathcal{P}(\mathbb{R}^d) \to L^2(\mathbb{R}^d, \sigma)$ defined as $\mu \in \mathcal{P}(\mathbb{R}^d) \mapsto T_\sigma^\mu$, where T_σ^μ is the unique optimal transport map that pushes σ to μ guaranteed by theory.

  One of the reasons for using this embedding for image classification is that computing distances in $L^2(\mathbb{R}^d, \sigma)$, using $W_2(\mu_i, \mu_j)$, is less computationally expensive than in $\mathcal{P}(\mathbb{R}^d)$, using $\int \|T_\sigma^{\mu_i}(x) - T_\sigma^{\mu_j}(x)\| d\sigma(x)$. The authors in [27] show that if two measures are similar up to small perturbations of shifts and scalings, such as the elements corresponding to a certain digit in MNIST, then the distance between them in $\mathcal{P}(\mathbb{R}^d)$ is close to the distance between their embeddings in $L^2(\mathbb{R}^d, \sigma)$. It is also shown that if two sets of measures in $\mathcal{P}(\mathbb{R}^d)$ are far enough, then their embeddings in $L^2(\mathbb{R}^d, \sigma)$ are linearly separable.

  Although the theoretical results in [27] are solely for absolutely continuous measures in $\mathbb{R}^d$, the authors show, through a series of experiments, that this idea works well in practice for image classification. Their theoretical results rely on Caffarelli's regularity results for the optimal transport map for the quadratic cost. I would like to prove that analogous results might hold if we allow plans for the embedding instead of maps and consider actual discrete measures as in [11] so that the gap between theory and experiments is bridged. Our previous regularity result for optimal plans in [11] would be useful for this purpose. An algorithm that uses plans instead of maps constructed from plans should also offer a much better performance.

- **Wasserstein depth**. For a given measure μ and an absolutely continuous reference measure μ_0 the Monge-Kantorovich depth is defined as $d_{MK}(x, \mu) = d_T(T^{-1}(x), \mu_0)$, where d_T is the Tukey depth for the reference measure μ_0 and T is the optimal transportation map between μ_0 and μ, namely

$$T = \text{argmin}_{Y : Y_{\#}\mu_0 = \mu} \int |Y(x) - x|^2 dx.$$

I plan to study stability properties of the Monge-Kantorovich depth [7] with respect to the input measures and to find rates of convergence of a minimizer for a discrete problem to the minimizer of the corresponding continuum problem using optimal transportation tools, such as cyclical monotonicity and the stability of optimal transportation with respect to input measures. This research direction is closely related to my previous work on regularity of optimal transportation plans for rough measures (of which discrete measures are a particular example) [11].

References

[1] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4):375–417, 1991.

[2] Jeff Calder and Wonjun Lee. Monotone discretizations of levelset convex geometric pdes. *arXiv preprint arXiv:2310.08450*, 2023.

[3] Michael G Crandall and Pierre-Louis Lions. Viscosity solutions of hamilton-jacobi equations. *Transactions of the American mathematical society*, 277(1):1–42, 1983.

[4] Antonio Cuevas, Manuel Febrero, and Ricardo Fraiman. Robust estimation and classification for functional data via projection-based depth notions. *Computational Statistics*, 22(3):481–496, 2007.

[5] Pierre Lafaye de Micheaux, Pavlo Mozharovskyi, and Myriam Vimond. Depth for curve data and applications. *Journal of the American Statistical Association*, pages 1–17, 2020.

[6] David L Donoho, Miriam Gasko, et al. Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *The Annals of Statistics*, 20(4):1803–1827, 1992.

[7] Alfred Galichon and Victor Chernozhukov. Monge-kantorovich depth, quantiles, ranks and signs. *Annals of Statistics*, 2015.

[8] Richard J Gardner. *Geometric tomography*, volume 6. Cambridge University Press Cambridge, 1995.

[9] Eldad Haber, Tauseef Rehman, and Allen Tannenbaum. An efficient numerical method for the solution of the L_2 optimal mass transfer problem. *SIAM Journal on Scientific Computing*, 32(1):197–211, 2010.

[10] Steven Haker, Lei Zhu, Allen Tannenbaum, and Sigurd Angenent. Optimal mass transport for registration and warping. *International Journal of Computer Vision*, 60(3):225–240, 2004.

[11] P-E Jabin, A Mellet, and M Molina-Fructuoso. Local regularity result for an optimal transportation problem with rough measures in the plane. *Journal of Functional Analysis*, 281(2):109041, 2021. ArXiv version: arXiv:1911.00574, 2021.

[12] Rebecka Jörnsten. Clustering and classification based on the l1 data depth. *Journal of Multivariate Analysis*, 90(1):67–89, 2004.

[13] Varun Khurana, Harish Kannan, Alexander Cloninger, and Caroline Moosmüller. Supervised learning of sheared distributions using linearized optimal transport. *Sampling Theory, Signal Processing, and Data Analysis*, 21(1):1, 2023.

[14] Soheil Kolouri and Gustavo K Rohde. Transport-based single frame super resolution of very low resolution face images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4876–4884, 2015.

[15] Xiaohui Liu. Fast implementation of the tukey depth. *Computational Statistics*, 32(4):1395–1410, 2017.

[16] Xiaohui Liu, Karl Mosler, and Pavlo Mozharovskyi. Fast computation of tukey trimmed regions and median in dimension p$>$ 2. *Journal of Computational and Graphical Statistics*, 28(3):682–697, 2019.

[17] Cong Ma, Kaizheng Wang, Yuejie Chi, and Yuxin Chen. Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval and matrix completion. In *International Conference on Machine Learning*, pages 3345–3354. PMLR, 2018.

[18] Jean-Claude Massé et al. Asymptotics for the tukey depth process, with an application to a multivariate trimmed mean. *Bernoulli*, 10(3):397–419, 2004.

[19] Tyler Maunu and Martin Molina-Fructuoso. Acceleration and implicit regularization in phase retrieval. ArXiv preprint: arXiv:2311.12888, 2023.

[20] Robert J McCann et al. Existence and uniqueness of monotone measure-preserving maps. *Duke Mathematical Journal*, 80(2):309–324, 1995.

[21] M Meyer and S Reisner. A geometric property of the boundary of symmetric convex bodies and convexity of flotation surfaces. *Geom. Dedicata*, 37(3):327–337, 1991.

[22] Kim Miller, Suneeta Ramaswami, Peter Rousseeuw, J Antoni Sellares, Diane Souvaine, Ileana Streinu, and Anja Struyf. Efficient computation of location depth contours by methods of computational geometry. *Statistics and Computing*, 13(2):153–162, 2003.

[23] Martin Molina-Fructuoso and Ryan Murray. A characteristics-based approach to computing tukey depths.

[24] Martin Molina-Fructuoso and Ryan Murray. Tukey depths and hamilton–jacobi differential equations. *SIAM Journal on Mathematics of Data Science*, 4(2):604–633, 2022 ArXiv version: arXiv:2104.01648, 2021.

[25] Martin Molina-Fructuoso and Ryan Murray. Eikonal depth: an optimal control approach to statistical depths. ArXiv preprint: arXiv:2201.05274, 2021.

[26] Caroline Moosmüller and Alexander Cloninger. Linear optimal transport embedding: Provable wasserstein classification for certain rigid transformations and perturbations. *arXiv preprint arXiv:2008.09165*, 2020.

[27] Caroline Moosmüller and Alexander Cloninger. Linear optimal transport embedding: provable wasserstein classification for certain rigid transformations and perturbations. *Information and Inference: A Journal of the IMA*, 12(1):363–389, 2023.

[28] Karl Mosler. Depth statistics. In *Robustness and complex data structures*, pages 17–34. Springer, 2013.

[29] Deborah Nolan. Asymptotics for multivariate trimming. *Stochastic processes and their applications*, 42(1):157–169, 1992.

[30] Adam M Oberman and Yuanlong Ruan. An efficient linear programming method for optimal transportation. *arXiv preprint arXiv:1509.03668*, 2015.

[31] François-Pierre Paty, Alexandre d'Aspremont, and Marco Cuturi. Regularity as regularization: Smooth and strongly convex brenier potentials in optimal transport. *arXiv preprint arXiv:1905.10812*, 2019.

[32] R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

[33] Nicolas Garcia Trillos, Moritz Gerlach, Matthias Hein, and Dejan Slepcev. Error estimates for spectral convergence of the graph laplacian on random geometric graphs towards the laplace–beltrami operator. *arXiv preprint arXiv:1801.10108*, 2018.

[34] Nicolas Garcia Trillos, Ryan Murray, and Matthew Thorpe. From graph cuts to isoperimetric inequalities: Convergence rates of cheeger cuts on data clouds. *arXiv preprint arXiv:2004.09304*, 2020.

[35] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.