Data depths meet Hamilton-Jacobi equations

Ryan Murray, NCSU
Joint work with Martin Molina-Fructuoso

IMA Data science seminar, October 2021
Are machine learning algorithms robust?

- Significant advances in recent years in learning problems.

Figure: Neural networks can solve difficult classification problems, for example involving images. Image credit: Caltech 101 images
Are machine learning algorithms robust?

- But algorithms don’t always act in predictable ways!

**Figure:** Neural networks are not (from a human’s viewpoint) robust. Image credit: gradientscience.org
Why do we care about robustness?

- Safety critical applications
- Generalization problem
- Real world is noisy

Figure: Non-robustness realized physically, from paper by Eykholt et al.
What methods can we use to induce robustness into learning?

- Regularize learning methods
- Modify inputs (using randomness or an adversary)
- Construct statistics that are inherently robust

Work on all three, today I’ll talk about the last one.
Classical robust statistics

- Statisticians have thought about robustness for a long time.
- There are many questions about elementary robust statistics methods which are not fully understood.
- Let’s consider the simplest example of a robust statistic: a median.
Mean vs. Median

- **Mean:** $\mu = \int_{\mathbb{R}} z \, d\rho(z)$.
- **Quantile:** $Q(x) = \int_{-\infty}^{x} d\rho(z)$.
- **Depth:** $d_Q(x) = \min(Q(x), 1 - Q(x))$.
- **Median:** $\arg \max d_Q$.

**Figure:** Visual representation of mean and median of a probability distribution, image credit wikipedia.
Mean vs. Median

- Mean: \( \mu = \int_{\mathbb{R}} z \, d\rho(z) \).
- Quantile: \( Q(x) = \int_{-\infty}^{x} d\rho(z) \).
- Depth: \( d_Q(x) = \min(Q(x), 1 - Q(x)) \).
- Median: \( \arg \max d_Q \).

Robustness: quantity cannot be changed significantly if a small amount of probability density is moved. Insensitivity to outliers.
Mean vs. Median

- Mean: \( \mu = \int_{\mathbb{R}} z \, d\rho(z) \).
- Quantile: \( Q(x) = \int_{-\infty}^{x} d\rho(z) \).
- Depth: \( d_Q(x) = \min(Q(x), 1 - Q(x)) \).
- Median: \( \arg \max d_Q \).

**Breakdown point**

A statistic has a breakdown point \( \delta \) if we are able to move the statistic arbitrarily by only moving \( \delta \) of the probability mass.

Means have arbitrarily small breakdown point, median breaks down at \( \delta = 1/2 \).
Mean vs. Median

- Mean: $\mu = \int_{\mathbb{R}} z \, d\rho(z)$.
- Quantile: $Q(x) = \int_{-\infty}^{x} d\rho(z)$.
- Depth: $d_Q(x) = \min(Q(x), 1 - Q(x))$.
- Median: $\arg\max d_Q$.

Medians are the simplest example of a robust statistic, central to many ideas for improving robustness of algorithms.
Mean extends directly: $\int_{\mathbb{R}^d} x \, d\rho(x)$.

Quantiles, depths, and medians can be generalized in many different ways.

$$d_Q(x) = \min \left( \int_{-\infty}^{x} d\rho(y), \int_{x}^{\infty} d\rho(y) \right).$$
Mean extends directly: $\int_{\mathbb{R}^d} x \; d\rho(x)$.

Quantiles, depths, and medians can be generalized in many different ways.

$$d_Q(x) = \min_{a \neq 0} \left( \int_{ay \leq ax} d\rho(y) \right).$$
Mean extends directly: $\int_{\mathbb{R}^d} x \, d\rho(x)$.

Quantiles, depths, and medians can be generalized in many different ways.

$$d_T(x) = \min_{a \in \mathbb{R}^d \neq 0} \left( \int_{a \cdot y \leq a \cdot x} d\rho(y) \right).$$
Mean extends directly: \( \int_{\mathbb{R}^d} x \times d\rho(x) \).

Quantiles, depths, and medians can be generalized in many different ways.

**Definition of halfspace depth**

The halfspace depth (or Tukey depth) is defined by

\[
d_T(x) = \min_{a \in \mathbb{R}^d \neq 0} \int_{a \cdot y \leq a \cdot x} d\rho(y).
\]

A Tukey median is any point in the argmax of the halfspace depth.
Applications for medians and depths

- Inherently robust
- Identifies outliers
- Creates a feature based upon centrality
- Induces ordering (“ranks and signs”)
- Information beyond clustering
Many different methods for generalizing depths/medians

- Mahalanobis depth: Use ellipsoids, oriented by variance.
- Convex peeling depth: iteratively take convex hulls and remove points that are in the boundary. Continuum limit with HJB equations Calder & Smart 2020
- Wasserstein depths: Use the optimal transport mapping to a reference domain $B(0, 1)$, and then compute depths on that domain. Chernozhukov, Hallin, Henry, Galichon 2017
- Many more…
Properties of halfspace depth

\[ d_T(x) = \min_{a \in \mathbb{R}^d \neq 0} \int_{a \cdot y \leq a \cdot x} d \rho(y). \]

- Very classical in robust statistics and computational geometry
- Convex level sets, affine invariant
- Computation is not simple

Figure: Contours of halfspace depth on square
Towards a differential equation

Define $Z(x, \nu) = \int_{(y-x) \cdot \nu \geq 0} d\rho(y)$. By using necessary conditions and (formal) chain rule

$$\nabla d_T = D_x Z = -\nu(x) \int_{(y-x) \cdot \nu = 0} \rho(y) \, dH^{n-1}(y),$$

which we can manipulate to give

$$|\nabla d_T| = \int_{(y-x) \cdot \nabla d_T = 0} \rho(y) \, dH^{n-1}(y).$$

Eikonal equation, with an unusual right hand side.
Infinitely many solutions?

Consider the domain $[0, 1]$ with a uniform distribution. $|\nabla d_T| = 1$.

Figure: Different weak solutions to the Tukey depth equation

How do we make sense of solutions? Infinitely many? How do we select the correct one?
Viscosity solutions to Hamilton Jacobi equations

Figure: Different weak solutions to the Tukey depth equation

- The "correct" solution is the largest one.
- Satisfies a generalized tangent property
A function $u$ is called a supersolution if any line touching $u$ from below with slope $p$ satisfies

$$|p| - \int_{(y-x) \cdot p = 0} \rho(y) \, dH^{n-1}(y) \geq 0.$$ 

A function $u$ is called a sub-solution if any plane touching $u$ from above with slope $p$ satisfies the reverse inequality.
Hamilton-Jacobi equations usually represent value functions in optimal control theory.

Direct connection with dynamic programming.

Established numerical methods: one-sided derivatives, dynamic programming, viscous approximations.

Figure: Different weak solutions to the Tukey depth equation.
Well-posedness of viscosity solution Tukey depth PDE

\[ |\nabla d_T| = \int_{(y-x) \cdot \nabla d_T = 0} \rho(y) \, dH^{n-1}(y). \]

**Theorem**

*As long as \( \rho \) is continuous and not too degenerate then there exists a unique viscosity solution to the PDE.*
Are Tukey depths viscosity solutions?

- Supersolutions are stable under infimum: so the Tukey depth is always a supersolution
- Viscosity solutions always provide a lower bound!
Are Tukey depths viscosity solutions?

- Are Tukey depths generally viscosity solutions? No! Explicit example where this fails.
- Are Tukey depths ever viscosity solutions? We suspect that if $\rho$ is log-concave then this is the case. Only have a proof in two dimensions.

Note non-convexity in $d_T$:

$$|\nabla d_T| = \int_{(y-x) \cdot \nabla d_T = 0} \rho(y) \, dH^{n-1}(y)$$
Sub-solution inequalities

$$|p| - \int_{(y-x) \cdot p=0} \rho(y) \, dH^{n-1}(y) \geq 0$$

Affine invariance reduces to orthogonal case. Balance condition!

*Bonnesen theorem*

**Figure**: Representation of the subsolution inequality
A detour into Convex geometry

Convex floating body

Let $K$ be convex. The convex floating body $K_\delta$ is given by

$$K_\delta = \bigcap_{|K \cap H| = \delta, H \text{ halfspace}} H^c$$

- Convex floating body iff super-level set of Tukey depth.
- One can define affine surface area using floating bodies:
  $$|K_\delta| - |K| \approx \delta^\alpha \kappa(K)$$
- Smoothness of $K_\delta$ relates to empirical to population consistency/limit theorems.
How would one build a numerical method (especially given empirical measures)?

- approximating integral
  \[ \int_{(x-y) \cdot \nu = 0} \rho(y) dH^{n-1}(y) \approx \varepsilon^{-1} \int_{|(x-y) \cdot \nu| \leq \varepsilon} \rho(y) dy \]

- empirical measures: graphs and weight scaling
- consistent derivatives for viscosity solutions
- Iterate non-linear equation to convergence
How would one build a numerical method (especially given empirical measures)?

- approximating integral
- empirical measures: graphs and weight scaling

Given data points \( \{x_i\}_{i=1}^{N} = X \) we construct a graph with weights

\[
w_{ij} = \varepsilon^{-d} \eta \left( \frac{|x_i - x_j|}{\varepsilon} \right)
\]

We can approximate derivatives

\[
C_\eta \rho(x) |\nabla u(x_i)| \approx \varepsilon^{-1} \sum_j w_{ij} |u(x_i) - u(x_j)|.
\]

- consistent derivatives for viscosity solutions
- Iterate non-linear equation to convergence
How would one build a numerical method (especially given empirical measures)?

- approximating integral
- empirical measures: graphs and weight scaling
- consistent derivatives for viscosity solutions

\[ |u'(x)| \approx h^{-1} \max((u(x) - u(x + h))_, (u(x) - u(x - h))_+) \]

Always selects derivative data using places where \( u(x \pm h) < u(x) \): “monotone” “upwind” schemes.
- Iterate non-linear equation to convergence
Numerical examples

(a) Exact halfspace depth

(b) Approximate halfspace depth

Figure: Uniform distribution on a square
Numerical examples

(a) Exact halfspace depth  

(b) Approximate halfspace depth

**Figure:** Uniform distribution on a triangle
Connections, open questions

- Deep connection with convex geometry: floating bodies, affine surface area. Open questions about regularity
- Question about higher dimensions, relates to generalizing Busemann’s theorem.
- Improving algorithms? Consistency results? Empirical measures?
Alternative way to generalize quantile depth

We could have chosen a different way to extend the one dimensional definition:

\[ d_Q(x) = \inf_{y(t): y(0) = x, y \to \infty} \int_0^\infty \rho(y(t)) |\dot{y}| \, dt \]

This definition directly extends to d-dimensions

\[ d_{eik}(x) = \inf_{y(t): y(0) = x, y \to \infty} \int_0^\infty \rho(y(t)) |\dot{y}| \, dt \]
Localizing the Tukey depth

Eikonal depth

We define the eikonal depth by the optimal control problem

\[ d_{eik}(x) = \inf_{y(t): y(0) = x, y \to \infty} \int_0^\infty \rho(y(t)) |\dot{y}| \, dt \]

\[ |\nabla d_{eik}| = \rho(x). \]

- Dynamic programming principle, viscosity solution characterizes depth
- Geodesic under a \( \rho \) weighted metric
- Replaced non-convex, non-local term with much friendlier one!
Properties of eikonal depth

\[ |\nabla d_{eik}| = \rho(x). \]

- Good stability (when \( \rho \) not degenerate)
- Non-convex level sets, mode preservation
- Scale invariance: replace \( \rho \rightarrow \rho^{1/d} \).
- Efficient computation via fast marching methods
Numerical illustrations

(a) Density function

(b) Approximate eikonal depth

Figure: Bimodal density: Gaussian mixture
Numerical illustrations

(a) Approximate eikonal depth

Figure: Uniform distribution on a square
The definition of the eikonal depth required only a few ingredients:

1. What is curve length?
2. What do we mean by data density along curves?
3. What is our boundary condition?

This type of eikonal depth extends readily to manifolds, metric measure spaces, graphs...
Numerical examples

(a) Density function on a cylinder

(b) Approximate eikonal depth on that cylinder

Figure: Eikonal depth for a non-uniform density on a cylinder. Here the density varies periodically, and we use the boundary of the manifold as a natural boundary condition.
MNIST depths

Figure: histogram of eikonal depths for mnist "4" digits, using a weighted neighborhood graph, and using all non-4's as the "boundary"
Depth functions from robust statistics have a direct connection with Hamilton-Jacobi equations.

Many deep connections with convex geometry, control theory.

Many open questions: analytical, computational, statistical.