# Semi-Discrete Normalizing Flows through Differentiable Tessellation

Ricky T. Q. Chen, Brandon Amos, Maximilian Nickel

Meta AI

## Main Takeaways

Differentiable tessellation + bijective mapping to construct normalizing flows on bounded supports.

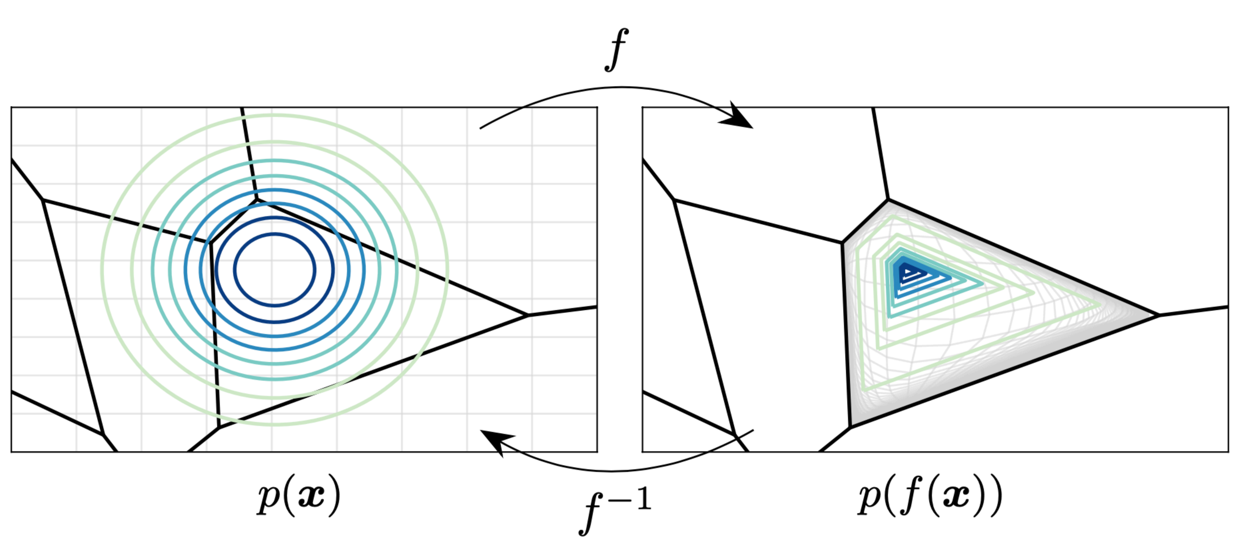
Maps between discrete & continuous distributions.

Generalizes existing dequantization methods.

Disjoint mixture models with O(1) compute cost.

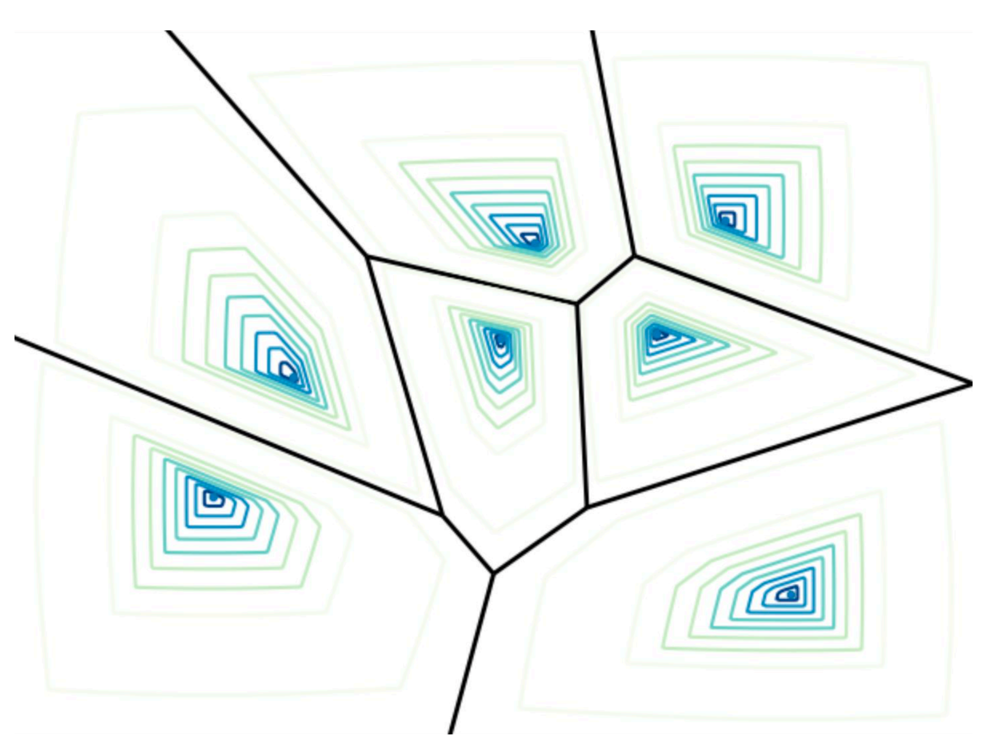
## What?

Distributions with **bounded** support.




## Why?

Combine lots of them with **disjoint** support.



Maps each continuous value to a discrete value.

$$\mathbb{1}_{[\boldsymbol{x} \in A_{\boldsymbol{y}}]}$$

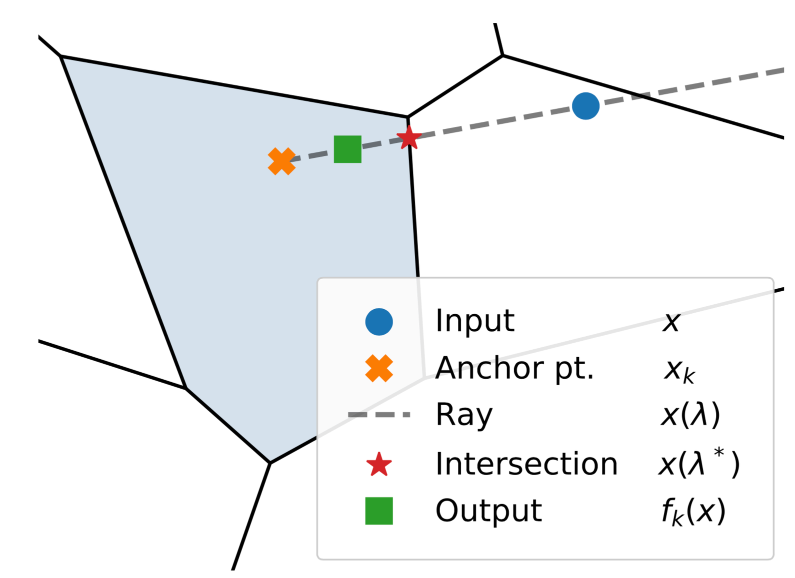
Maps each discrete value to a continuous distribution.

$$\log q(\boldsymbol{x}|\boldsymbol{y})$$

## Bijective Mapping to Convex Supports

Parameterize Voronoi tessellation using anchor points.

Bijective map through 1D transformation:



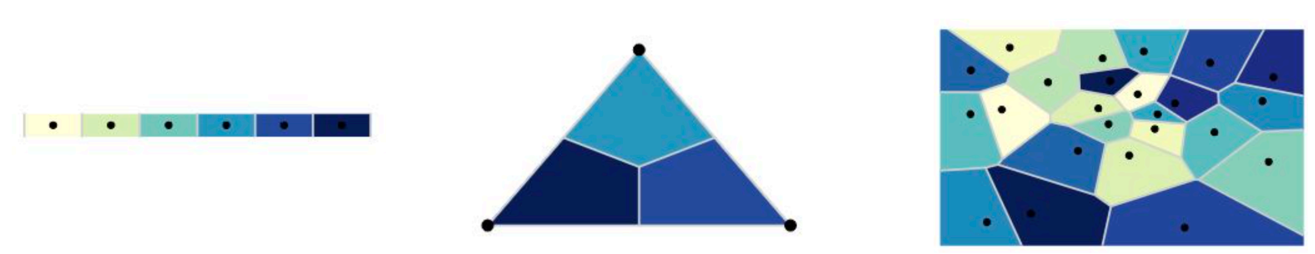

Log probability is easy to compute in closed form.

$$p_z(f(\boldsymbol{x})) = p_x(\boldsymbol{x}) \left| \det \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} \right|^{-1}$$

## Voronoi Dequantization (Discrete Data)

Learns the map from discrete to continuous.

Does not couple dimension with #discrete values.




$$\log p(\boldsymbol{y}) \geq \mathbb{E}_{\boldsymbol{x} \sim q(\boldsymbol{x}|\boldsymbol{y})} \left[ \log(\mathbb{1}_{[\boldsymbol{x} \in A_{\boldsymbol{y}}]} p(\boldsymbol{x})) - \log q(\boldsymbol{x}|\boldsymbol{y}) \right]$$
$$= \mathbb{E}_{\boldsymbol{x} \sim q(\boldsymbol{x}|\boldsymbol{y})} \left[ \log p(\boldsymbol{x}) - \log q(\boldsymbol{x}|\boldsymbol{y}) \right]$$

Discrete model

Density model (continuous)

Dequantization model

## Disjoint Mixture Modeling (Continuous Data)

Mixture models are expensive:

$$p(\boldsymbol{x}) = \sum_{k=1}^{K} p(\boldsymbol{x}|k) p(k)$$

Scales with number of components

But if components are disjoint:

$$p(\boldsymbol{x}) = \sum_{k=1}^{K} \mathbb{1}_{[\boldsymbol{x} \in A_k]} p(\boldsymbol{x}|k) p(k)$$
$$= p(\boldsymbol{x}|k = g(\boldsymbol{x})) p(k = g(\boldsymbol{x}))$$

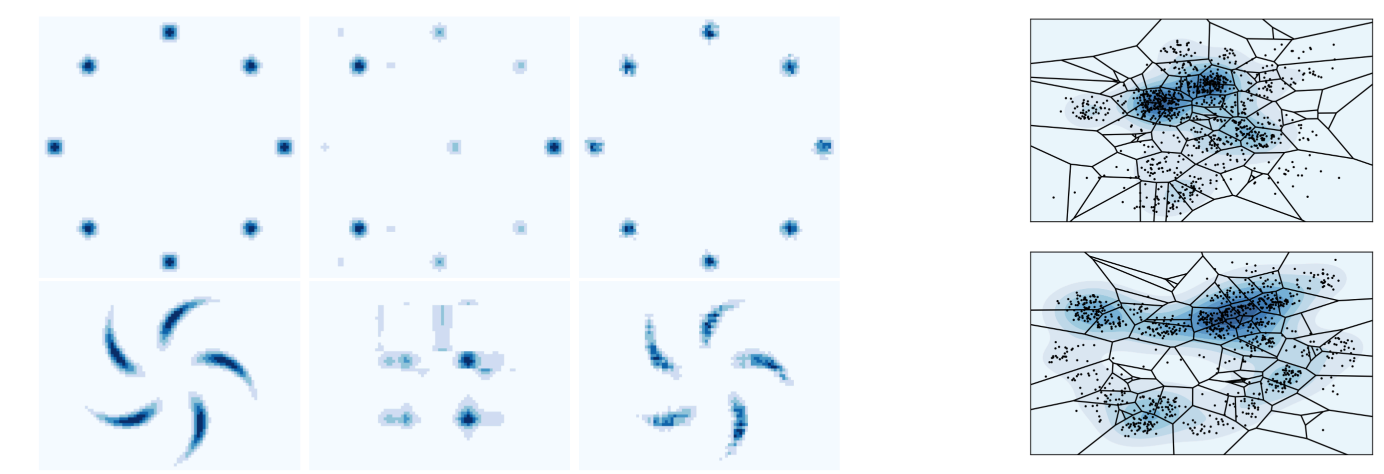
Disjoint subsets

Set identification function

## Experiments

Can model complex relations between discrete data.

Learns to cluster discrete values with similar probabilities.




Beats existing dequantization approaches across many data modalities

Table 1: **Discrete UCI data sets.** Negative log-likelihood results on the test sets in nats.

| Method | Connect4 | Forests | Mushroom | Nursery | PokerHands | USCensus90 |
|---|---|---|---|---|---|---|
| Voronoi Deq. | 12.92±0.07 | 14.20±0.05 | 9.06±0.05 | 9.27±0.04 | 19.86±0.04 | 24.19±0.12 |
| Simplex Deq. | 13.46±0.01 | 16.58±0.01 | 9.26±0.01 | 9.50±0.00 | 19.90±0.00 | 28.09±0.08 |
| BinaryArgmax Deq. | 13.71±0.04 | 16.73±0.17 | 9.53±0.01 | 9.49±0.00 | 19.90±0.01 | 27.23±0.02 |
| Discrete Flow | 19.80±0.01 | 21.91±0.01 | 22.06±0.01 | 9.53±0.01 | 19.82±0.03 | 55.62±0.35 |

Table 2: **Permutation-invariant discrete itemset modeling.**

| Model (Dequantization) | **Retail** (nats) | **Accidents** (nats) |
|---|---|---|
| CNF (Voronoi) | 9.44±2.34 | 7.81±2.84 |
| CNF (Simplex) | 24.16±0.21 | 19.19±0.01 |
| CNF (BinaryArgmax) | 10.47±0.42 | 6.72±0.23 |
| Determinantal Point Process | 20.35±0.05 | 15.78±0.04 |

Table 3: **Language modeling.**

| Dequantization | **text8** (bpc) | **enwik8** (bpc) |
|---|---|---|
| Voronoi ($D$=2) | 1.39±0.01 | 1.46±0.01 |
| Voronoi ($D$=4) | 1.37±0.00 | 1.41±0.00 |
| Voronoi ($D$=6) | 1.37±0.00 | 1.40±0.00 |
| Voronoi ($D$=8) | 1.36±0.00 | 1.39±0.01 |
| BinaryArgmax [18] | 1.38 | 1.42 |
| Ordinal [18] | 1.43 | 1.44 |

Disjoint mixture modeling increases flexibility at no cost.

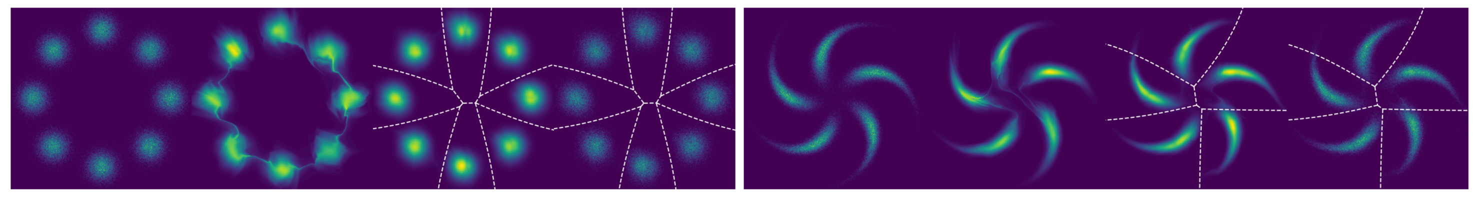



Figure 6: Tessellation is done in a transformed space; nonlinear boundaries are shown.

Table 4: **Disjoint mixture modeling.** NLL on the test sets in nats. *Baseline results from [13, 35].

| Method | POWER | GAS | HEPMASS | MINIBOONE | BSDS300 |
|---|---|---|---|---|---|
| Real NVP* | -0.17±0.01 | -8.33±0.07 | 18.71±0.01 | 13.55±0.26 | -153.28±0.89 |
| MAF* | -0.24±0.01 | -10.08±0.01 | 17.73±0.01 | 12.24±0.22 | -154.93±0.14 |
| FFJORD* | -0.46±0.01 | -8.59±0.12 | 14.92±0.08 | 10.43±0.22 | -157.40±0.19 |
| Base Coupling Flow | -0.44±0.01 | -11.75±0.02 | 16.78±0.08 | 10.87±0.06 | -155.14±0.04 |
| Voronoi Disjoint Mixture | -0.52±0.01 | -12.63±0.05 | 16.16±0.01 | 10.24±0.14 | -156.59±0.14 |