Triangular Flows for Generative Modeling

Statistical Consistency, Smoothness Classes, and Fast Rates

Nicholas J. Irons, *UW*

Meyer Scetbon, *CREST-ENSAE*
Soumik Pal, *UW*
Zaid Harchaoui, *UW*

Kantorovich Initiative Retreat 2022/03/18

Overview

- **Triangular flows** based on the **Knöthe-Rosenblatt (KR) map** have been a major building block of **normalizing flows** for **generative modeling**.[1]

[1] Kobyzev et al. Normalizing flows: An introduction and review of current methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

Overview

- **Triangular flows** based on the **Knöthe-Rosenblatt (KR) map** have been a major building block of **normalizing flows** for **generative modeling**.[1]
- We establish **statistical consistency** and **convergence rates** of **triangular flow** estimators. We obtain **novel statistical guarantees** for **normalizing-flow-based generative models** used in practice.

[1] Kobyzev et al. Normalizing flows: An introduction and review of current methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

Overview

- **Triangular flows** based on the **Knöthe-Rosenblatt (KR) map** have been a major building block of **normalizing flows** for **generative modeling**.[1]

- We establish **statistical consistency** and **convergence rates** of **triangular flow** estimators. We obtain **novel statistical guarantees** for **normalizing-flow-based generative models** used in practice.

- Our results identify the function classes at play and shed light on model design.

[1] Kobyzev et al. Normalizing flows: An introduction and review of current methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

Generative Models

Generative Models

Setup Suppose we have an unknown density f on $\mathbb{R}^d$, e.g., an image dataset, that we wish to sample from.

Generative Models

Setup Suppose we have an unknown density f on $\mathbb{R}^d$, e.g., an image dataset, that we wish to sample from.

Suppose Y is a random variable on $\mathbb{R}^d$ that we can sample from easily, e.g., a normal distribution. Let g denote its density.

Generative Models

Setup Suppose we have an unknown density f on $\mathbb{R}^d$, e.g., an image dataset, that we wish to sample from.

Suppose Y is a random variable on $\mathbb{R}^d$ that we can sample from easily, e.g., a normal distribution. Let g denote its density.

Goal Find a **generator** $S : \mathbb{R}^d \to \mathbb{R}^d$ that produces new images distributed according to f:

Generative Models

Setup Suppose we have an unknown density f on $\mathbb{R}^d$, e.g., an image dataset, that we wish to sample from.

Suppose Y is a random variable on $\mathbb{R}^d$ that we can sample from easily, e.g., a normal distribution. Let g denote its density.

Goal Find a **generator** $S : \mathbb{R}^d \to \mathbb{R}^d$ that produces new images distributed according to f:

1. Generate samples $Y_i \sim g$

Generative Models

Setup Suppose we have an unknown density f on $\mathbb{R}^d$, e.g., an image dataset, that we wish to sample from.

Suppose Y is a random variable on $\mathbb{R}^d$ that we can sample from easily, e.g., a normal distribution. Let g denote its density.

Goal Find a **generator** $S : \mathbb{R}^d \to \mathbb{R}^d$ that produces new images distributed according to f:

1. Generate samples $Y_i \sim g$
2. Push forward Y_i under S to produce $X_i = S(Y_i) \sim f$

Generative Models

We refer to S as a **pushforward** or **transport map** from g to f and write $S\#g = f$.

Generative Models

We refer to S as a **pushforward** or **transport map** from g to f and write $S\#g = f$.

By the change of variables formula, $S\#g = f$ implies

$$f(x) = g(S^{-1}(x))|\det(\nabla S^{-1}(x))|. \tag{1}$$

Generative Models

We refer to S as a **pushforward** or **transport map** from g to f and write $S\#g = f$.

By the change of variables formula, $S\#g = f$ implies

$$f(x) = g(S^{-1}(x))|\det(\nabla S^{-1}(x))|. \tag{1}$$

With an estimate of S we can also estimate the unknown density f.

Examples from CelebA

Samples from Real NVP trained on CelebA[2]

[2] Dinh et al. Density estimation using Real NVP. *ICLR* 2017.

Knöthe-Rosenblatt (KR) Rearrangement

KR rearrangement S^* is a transport map between multivariate distributions that exists for *any* pair of Lebesgue densities f, g on $\mathbb{R}^d$.[3]

[3] Carlier et al. From Knothe's transport to Brenier's map and a continuation method for optimal transport. *SIAM Journal on Mathematical Analysis*, 2010

Knöthe-Rosenblatt (KR) Rearrangement

KR rearrangement S^* is a transport map between multivariate distributions that exists for *any* pair of Lebesgue densities f, g on $\mathbb{R}^d$.[3]

The KR map is **triangular** in the sense that

$$S^*(x) = \begin{bmatrix} S_1^*(x_1, \dots, x_d) \\ S_2^*(x_2, \dots, x_d) \\ \vdots \\ S_{d-1}^*(x_{d-1}, x_d) \\ S_d^*(x_d) \end{bmatrix}.$$

[3] Carlier et al. From Knothe's transport to Brenier's map and a continuation method for optimal transport. *SIAM Journal on Mathematical Analysis*, 2010

Knöthe-Rosenblatt (KR) Rearrangement

The components of S^* are **defined recursively** via the **monotone transport** between the univariate conditional densities of f and g.

Knöthe-Rosenblatt (KR) Rearrangement

The components of S^* are **defined recursively** via the **monotone transport** between the univariate conditional densities of f and g.

For $k \in [d]$, let $F_k(x_k|x_{(k+1):d})$ denote the cdf of the conditional density $f_k(x_k|x_{(k+1):d})$ (and similarly for g). We first define

$$S_d^*(x_d) = G_d^{-1}(F_d(x_d)).$$

Knöthe-Rosenblatt (KR) Rearrangement

The components of S^* are **defined recursively** via the **monotone transport** between the univariate conditional densities of f and g.

For $k \in [d]$, let $F_k(x_k | x_{(k+1):d})$ denote the cdf of the conditional density $f_k(x_k | x_{(k+1):d})$ (and similarly for g). We first define

$$S_d^*(x_d) = G_d^{-1}(F_d(x_d)).$$

From here the kth component of S^* is

$$S_k^*(x_k, \ldots, x_d) = G_k^{-1} \left(F_k(x_k | x_{(k+1):d}) \middle| S_{(k+1):d}^*(x_{(k+1):d}) \right).$$

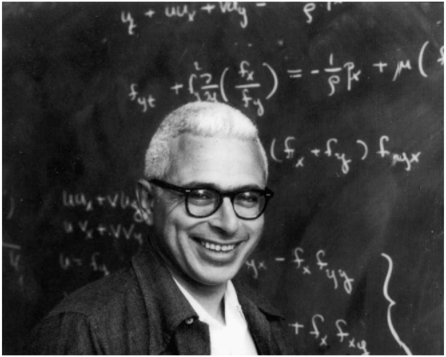
Knöthe-Rosenblatt (KR) Rearrangement

CONTRIBUTIONS TO THE THEORY OF CONVEX BODIES

Herbert Knothe

1. GENERALIZATION OF THE PRINCIPAL THEOREM OF BRUNN AND MINKOWSKI

The Brunn-Minkowski theorem on closed convex bodies in n-dimensional Euclidean space can be extended by introducing a suitably defined logarithmically convex functional $\rho_K(\vec{x})$. In the present paper we give a proof of such an extension



470 MURRAY ROSENBLATT

REMARKS ON A MULTIVARIATE TRANSFORMATION[1]

By Murray Rosenblatt

University of Chicago

The object of this note is to point out and discuss a simple transformation[2] of an absolutely continuous k-variate distribution $F(x_1, \cdots, x_k)$ into the uniform distribution on the k-dimensional hypercube. A discussion of related transformations has been given by P. Lévy [1].

Knöthe-Rosenblatt (KR) Rearrangement

The KR map S^* has the following properties:

Knöthe-Rosenblatt (KR) Rearrangement

The KR map S^* has the following properties:

- S^* is a **transport map**: $S^* \# f = g$.
- The Jacobian matrix ∇S^* is defined a.s. and upper **triangular**.

Knöthe-Rosenblatt (KR) Rearrangement

The KR map S^* has the following properties:

- S^* is a **transport map**: $S^* \# f = g$.
- The Jacobian matrix ∇S^* is defined a.s. and upper **triangular**.
- ∇S^* has non-negative entries a.s. on the diagonal **(monotonicity)**.

Knöthe-Rosenblatt (KR) Rearrangement

The KR map S^* has the following properties:

- S^* is a **transport map**: $S^* \# f = g$.
- The Jacobian matrix ∇S^* is defined a.s. and upper **triangular**.
- ∇S^* has non-negative entries a.s. on the diagonal **(monotonicity)**.
- S^* is the **unique** map (up to null sets) satisfying the above.

Knöthe-Rosenblatt (KR) Rearrangement

The KR map S^* has the following properties:

- S^* is a **transport map**: $S^* \# f = g$.
- The Jacobian matrix ∇S^* is defined a.s. and upper **triangular**.
- ∇S^* has non-negative entries a.s. on the diagonal **(monotonicity)**.
- S^* is the **unique** map (up to null sets) satisfying the above.
- S^* is as **smooth** as the densities f, g.

Knöthe-Rosenblatt (KR) Rearrangement

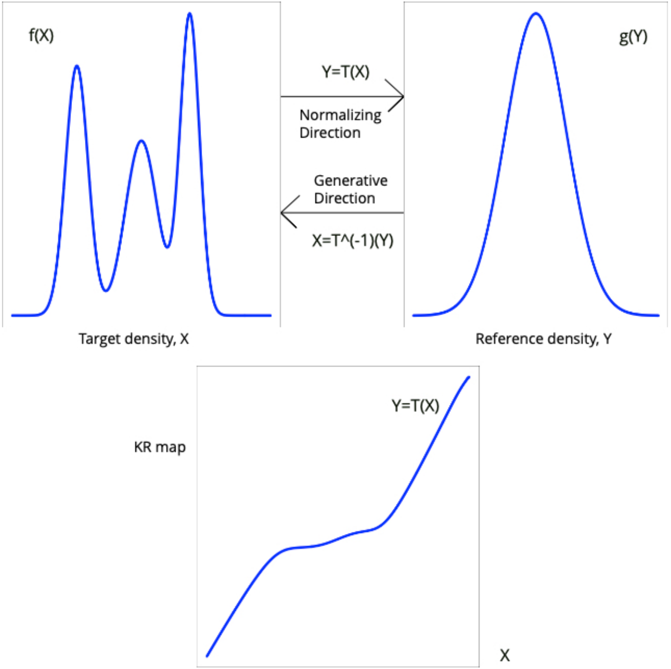
The KR map S^* has the following properties:

- S^* is a **transport map**: $S^* \# f = g$.
- The Jacobian matrix ∇S^* is defined a.s. and upper **triangular**.
- ∇S^* has non-negative entries a.s. on the diagonal **(monotonicity)**.
- S^* is the **unique** map (up to null sets) satisfying the above.
- S^* is as **smooth** as the densities f, g.
- S^* is **explicitly defined** in terms of the **conditional densities** of f and g.

Knöthe-Rosenblatt (KR) Rearrangement

The KR map S^* has the following properties:

- S^* is a **transport map**: $S^* \# f = g$.
- The Jacobian matrix ∇S^* is defined a.s. and upper **triangular**.
- ∇S^* has non-negative entries a.s. on the diagonal **(monotonicity)**.
- S^* is the **unique** map (up to null sets) satisfying the above.
- S^* is as **smooth** as the densities f, g.
- S^* is **explicitly defined** in terms of the **conditional densities** of f and g.
- There are $d!$ ways to build the KR map, depending on the **order** in which we condition the d coordinates.


f(X)
g(Y)
Y=T(X)
Normalizing Direction
Generative Direction
X=T^(-1)(Y)
Target density, X
Reference density, Y
KR map
Y=T(X)
X

Triangular Flows

[4] Kobyzev et al. (2020) Normalizing Flows: An Introduction and Review of Current Methods.

Triangular Flows

Triangular flows can be used to approximate the KR map between a source density and a target density, given samples from the target density.

[4] Kobyzev et al. (2020) Normalizing Flows: An Introduction and Review of Current Methods.

Triangular Flows

Triangular flows can be used to approximate the KR map between a source density and a target density, given samples from the target density.

Due to their desirable computational properties, triangular flows have been proposed and implemented as simple and expressive building blocks of generative models based on normalizing flows.[4]

[4] Kobyzev et al. (2020) Normalizing Flows: An Introduction and Review of Current Methods.

Triangular Flows

Triangular flows can be used to approximate the KR map between a source density and a target density, given samples from the target density.

Due to their desirable computational properties, triangular flows have been proposed and implemented as simple and expressive building blocks of generative models based on normalizing flows.[4]

However, there are few results establishing statistical guarantees for normalizing flow models.

[4] Kobyzev et al. (2020) Normalizing Flows: An Introduction and Review of Current Methods.

Triangular Flow Estimator

Triangular Flow Estimator

By the properties listed above, the KR map can be characterized as the **unique minimizer of the Kullback-Leibler (KL) divergence**

$$\min_{S \in \mathcal{T}} \mathrm{KL}(S \# f \| g),$$

where $\mathcal{T}$ is the convex cone of increasing triangular maps.

Triangular Flow Estimator

By the properties listed above, the KR map can be characterized as the **unique minimizer of the Kullback-Leibler (KL) divergence**

$$\min_{S \in \mathcal{T}} \mathrm{KL}(S \# f \| g),$$

where $\mathcal{T}$ is the convex cone of increasing triangular maps.

By the change of variables formula, the KL objective can be rewritten

$$\mathrm{KL}(S \# f \| g) = \mathbb{E}_{X \sim f}\left[\log f(X) - \log g(S(X)) - \sum_{k=1}^{d} \log D_k S_k(X)\right]. \quad (2)$$

Triangular Flow Estimator

Triangular Flow Estimator

Using our iid data $\{X_i\}_{i=1}^n \overset{iid}{\sim} f$, we study triangular flow estimators $S^n \in \mathcal{T}$ of the KR map derived from minimizing the **sample average approximation** to the KL objective (2) (or, equivalently, the negative log-likelihood):

Triangular Flow Estimator

Using our iid data $\{X_i\}_{i=1}^n \overset{iid}{\sim} f$, we study triangular flow estimators $S^n \in \mathcal{T}$ of the KR map derived from minimizing the **sample average approximation** to the KL objective (2) (or, equivalently, the negative log-likelihood):

$$\widehat{\mathrm{KL}}(S_\# f \| g) := \frac{1}{n} \sum_{i=1}^{n} \left[\log f(X_i) - \log g(S(X_i)) - \sum_{k=1}^{d} \log D_k S_k(X_i) \right]. \quad (3)$$

Triangular Flow Estimator

Using our iid data $\{X_i\}_{i=1}^n \overset{iid}{\sim} f$, we study triangular flow estimators $S^n \in \mathcal{T}$ of the KR map derived from minimizing the **sample average approximation** to the KL objective (2) (or, equivalently, the negative log-likelihood):

$$\widehat{\mathrm{KL}}(S_\# f \| g) := \frac{1}{n}\sum_{i=1}^{n}\left[\log f(X_i) - \log g(S(X_i)) - \sum_{k=1}^{d} \log D_k S_k(X_i)\right]. \quad (3)$$

Goal Study the statistical convergence properties of S^n as an estimator of the KR map S^*.

Triangular Flow Estimator

Using our iid data $\{X_i\}_{i=1}^n \stackrel{iid}{\sim} f$, we study triangular flow estimators $S^n \in \mathcal{T}$ of the KR map derived from minimizing the **sample average approximation** to the KL objective (2) (or, equivalently, the negative log-likelihood):

$$\widehat{\mathrm{KL}}(S_\# f \| g) := \frac{1}{n}\sum_{i=1}^{n}\left[\log f(X_i) - \log g(S(X_i)) - \sum_{k=1}^{d}\log D_k S_k(X_i)\right]. \quad (3)$$

Goal Study the statistical convergence properties of S^n as an estimator of the KR map S^*.

Remark (convexity) Assuming the source density g is **log-concave**, the objective (3) is **convex** in S.

No Free Lunch

Slow rates

Without combining both

- a **tail condition** (e.g., common compact support),
- a **smoothness condition** (e.g., uniformly bounded derivatives)

on the hypothesis function class $\mathcal{F}$ of the target density f, convergence of any estimator S^n of the KR map S^* from f to g can occur at an **arbitrarily slow rate**.

No Free Lunch

Theorem[a] Let $\mathcal{F}$ denote the class of C^∞ Lebesgue densities on $[0,1]^d$ bounded by 2. Let g be any Lebesgue density on $\mathbb{R}^d$.

For any $n \in \mathbb{N}$, the minimax risk in terms of KL divergence is bounded below as

$$\inf_{S^n} \sup_{f \in \mathcal{F}} \mathbb{E}_f[\mathrm{KL}(f \| f_n))] \geq 1/2,$$

where S^n is any estimate of S^*, and $f_n = (S^n)^{-1} \# g$ is the density estimate of f.

[a] Birgé (1986), see also Devroye (1983)

Setup

Exploiting smoothness Based on our "no free lunch" theorem, we restrict our estimator S^n to lying in an s-smooth Sobolev-type ball $\mathcal{T}(s, d, M) \subset \mathcal{T}$.

Setup

Exploiting smoothness Based on our "no free lunch" theorem, we restrict our estimator S^n to lying in an s-smooth Sobolev-type ball $\mathcal{T}(s, d, M) \subset \mathcal{T}$.

Assumptions

1. The target and source densities $f, g > 0$ have **compact, convex supports**.

Setup

Exploiting smoothness Based on our "no free lunch" theorem, we restrict our estimator S^n to lying in an s-smooth Sobolev-type ball $\mathcal{T}(s, d, M) \subset \mathcal{T}$.

Assumptions

1. The target and source densities $f, g > 0$ have **compact, convex supports**.
2. f, g have **continuous derivatives up to order s**.

Setup

Exploiting smoothness Based on our "no free lunch" theorem, we restrict our estimator S^n to lying in an s-smooth Sobolev-type ball $\mathcal{T}(s, d, M) \subset \mathcal{T}$.

Assumptions

1. The target and source densities $f, g > 0$ have **compact, convex supports**.
2. f, g have **continuous derivatives up to order s**.

Under these assumptions, the KR map S^* lies in $\mathcal{T}(s, d, M^*)$ for some $M^* > 0$.

Weak consistency

Theorem (KL consistency) Let $S^n \in \mathcal{T}(s, d, M^*)$ be any near-optimizer of the sample objective (3). Then S^n converges to the true KR map S^* in KL divergence:

$$\mathrm{KL}(S^n \# f \| g) \xrightarrow{p} \mathrm{KL}(S^* \# f \| g) = 0.$$

Weak consistency

Theorem (KL consistency) Let $S^n \in \mathcal{T}(s, d, M^*)$ be any near-optimizer of the sample objective (3). Then S^n converges to the true KR map S^* in KL divergence:

$$\mathrm{KL}(S^n \# f \| g) \xrightarrow{p} \mathrm{KL}(S^* \# f \| g) = 0.$$

Proof idea Use metric entropy bounds on the complexity of the Sobolev-type space $\mathcal{T}(s, d, M)$ to bound the risk of the estimator

$$\mathrm{KL}(S^n \# f \| g) - \mathrm{KL}(S^* \# f \| g) \leq 2 \|\widehat{\mathrm{KL}} - \mathrm{KL}\|_{\mathcal{T}(s,d,M)} + o_P(1).$$

Uniform Consistency

Uniform Consistency

Theorem (uniform consistency) Let $S^n \in \mathcal{T}(s, d, M^*)$ be any near-optimizer of the sample objective (3). Then S^n is a uniformly consistent estimator of S^*:

$$\|S^n - S^*\|_{\infty,d} \xrightarrow{p} 0.$$

Uniform Consistency

Theorem (uniform consistency) Let $S^n \in \mathcal{T}(s, d, M^*)$ be any near-optimizer of the sample objective (3). Then S^n is a uniformly consistent estimator of S^*:

$$\|S^n - S^*\|_{\infty,d} \xrightarrow{p} 0.$$

Proof idea

Using (pre)compactness of $\mathcal{T}(s, d, M^*)$ and lower semicontinuity of KL in $\|\cdot\|_{\infty,d}$, conclude that S^* is a well-separated KL minimizer with respect to $\|\cdot\|_{\infty,d}$.

Uniform Consistency

Theorem (uniform consistency) Let $S^n \in \mathcal{T}(s, d, M^*)$ be any near-optimizer of the sample objective (3). Then S^n is a uniformly consistent estimator of S^*:

$$\|S^n - S^*\|_{\infty,d} \xrightarrow{p} 0.$$

Proof idea

Using (pre)compactness of $\mathcal{T}(s, d, M^*)$ and lower semicontinuity of KL in $\|\cdot\|_{\infty,d}$, conclude that S^* is a well-separated KL minimizer with respect to $\|\cdot\|_{\infty,d}$.

Combine this with the weak consistency theorem above to complete the proof.

Convergence Rates

Convergence Rates

Theorem (KL convergence rate)
Under a technical assumption, the expected KL divergence of S^n is bounded as

$$\mathbb{E}[\mathrm{KL}(S^n \# f \| g)] \lesssim \begin{cases} n^{-1/2}, & d < 2s, \\ n^{-1/2} \log n, & d = 2s, \\ n^{-s/d}, & d > 2s. \end{cases}$$

Convergence Rates

Theorem (KL convergence rate)

Under a technical assumption, the expected KL divergence of S^n is bounded as

$$\mathbb{E}[\mathrm{KL}(S^n \# f \| g)] \lesssim \begin{cases} n^{-1/2}, & d < 2s, \\ n^{-1/2} \log n, & d = 2s, \\ n^{-s/d}, & d > 2s. \end{cases}$$

These rates also hold for:

Convergence Rates

Theorem (KL convergence rate)
Under a technical assumption, the expected KL divergence of S^n is bounded as

$$\mathbb{E}[\mathrm{KL}(S^n \# f \| g)] \lesssim \begin{cases} n^{-1/2}, & d < 2s, \\ n^{-1/2} \log n, & d = 2s, \\ n^{-s/d}, & d > 2s. \end{cases}$$

These rates also hold for:

- convergence $S^n \to S^*$ in a Sobolev-type norm under **strong log-concavity** of the source density g.

Convergence Rates

Theorem (KL convergence rate)
Under a technical assumption, the expected KL divergence of S^n is bounded as

$$\mathbb{E}[\mathrm{KL}(S^n \# f \| g)] \lesssim \begin{cases} n^{-1/2}, & d < 2s, \\ n^{-1/2} \log n, & d = 2s, \\ n^{-s/d}, & d > 2s. \end{cases}$$

These rates also hold for:

- convergence $S^n \to S^*$ in a Sobolev-type norm under **strong log-concavity** of the source density g.
- convergence of **normalizing flows** built from compositions of triangular maps, e.g., **Real NVP**. (Some of the first statistical guarantees for flow models.)

Optimal Coordinate Ordering

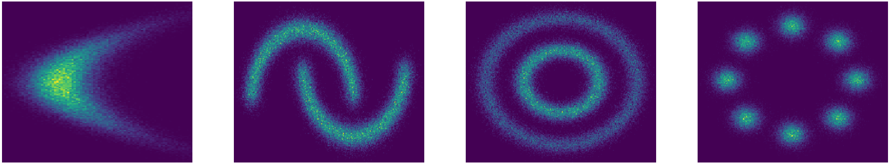
Optimal Coordinate Ordering

Theorem Suppose the target f is **anisotropically smooth**. The upper bound on the rate of convergence is minimized when we first condition on the smoothest coordinate of f, then the second, etc.

Optimal Coordinate Ordering

Theorem Suppose the target f is **anisotropically smooth**. The upper bound on the rate of convergence is minimized when we first condition on the smoothest coordinate of f, then the second, etc.

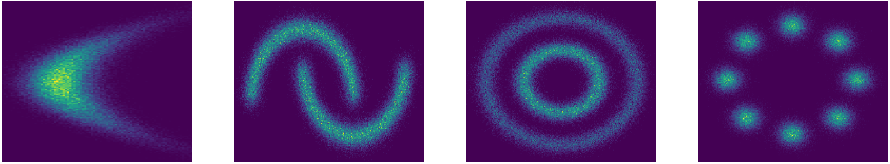
Target density f



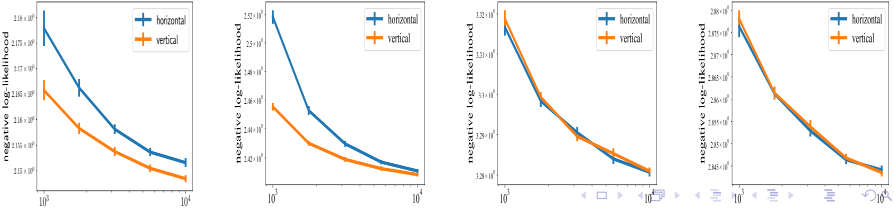
Optimal Coordinate Ordering

Theorem Suppose the target f is **anisotropically smooth**. The upper bound on the rate of convergence is minimized when we first condition on the smoothest coordinate of f, then the second, etc.

Target density f



KL loss vs. sample size for different orderings



Connections to Anisotropic OT

Connections to Anisotropic OT

Our result on the optimal ordering of coordinates complements the following theorem of Carlier et al. adapted to our setup.

Connections to Anisotropic OT

Our result on the optimal ordering of coordinates complements the following theorem of Carlier et al. adapted to our setup.

Theorem (Theorem 2.1; Carlier, Galichon, and Santambrogio (2008))

Let f and g be compactly supported Lebesgue densities on $\mathbb{R}^d$. Let $\epsilon > 0$ and let γ^ϵ be an optimal transport plan between f and g for the cost

$$c_\epsilon(x, y) = \sum_{k=1}^{d} \lambda_k(\epsilon)(x_k - y_k)^2,$$

for some weights $\lambda_k(\epsilon) > 0$. Suppose that for all $k \in \{1, \ldots, d-1\}$, $\lambda_k(\epsilon)/\lambda_{k+1}(\epsilon) \to 0$ as $\epsilon \to 0$. Let S^* be the Knöthe-Rosenblatt map between f and g and $\gamma^* = (id \times S^*)\#f$ the associated transport plan. Then $\gamma^\epsilon \rightsquigarrow \gamma^*$ as $\epsilon \to 0$. Moreover, should the plans γ^ϵ be induced by transport maps S^ϵ, then these maps would converge to S^* in $L^2(f)$ as $\epsilon \to 0$.

Connections to Anisotropic OT

Connections to Anisotropic OT

With this theorem in mind, the KR map S^* can be viewed as a limit of optimal transport maps S^ϵ for which transport in the dth direction is more costly than in the $(d-1)$st, and so on.

Connections to Anisotropic OT

With this theorem in mind, the KR map S^* can be viewed as a limit of optimal transport maps S^ϵ for which transport in the dth direction is more costly than in the $(d-1)$st, and so on.

The anisotropic cost function $c_\epsilon(x, y)$ inherently promotes increasing regularity of S^ϵ in x_k for larger $k \in [d]$. Our dimension ordering theorem establishes the same heuristic for learning triangular flows based on Knöthe-Rosenblatt rearrangement to build generative models.

Further Results

Further Results

- **Finite sample rates of convergence** $S^n \to S^*$ in a Sobolev-type norm under **strong log-concavity** of g.

Further Results

- **Finite sample rates of convergence** $S^n \to S^*$ in a Sobolev-type norm under **strong log-concavity** of g.

- **Uniform consistency** and **Sobolev-type convergence rates** of the inverse map $T^n = (S^n)^{-1} \to T^* = (S^*)^{-1}$, which is used to sample from f.

Further Results

- **Finite sample rates of convergence** $S^n \to S^*$ in a Sobolev-type norm under **strong log-concavity** of g.
- **Uniform consistency** and **Sobolev-type convergence rates** of the inverse map $T^n = (S^n)^{-1} \to T^* = (S^*)^{-1}$, which is used to sample from f.
- **Non-asymptotic convergence rates** of **flows** built from compositions of triangular maps, e.g., **Real NVP**. Some of the first statistical guarantees for flow models.

Further Results

- **Finite sample rates of convergence** $S^n \to S^*$ in a Sobolev-type norm under **strong log-concavity** of g.

- **Uniform consistency** and **Sobolev-type convergence rates** of the inverse map $T^n = (S^n)^{-1} \to T^* = (S^*)^{-1}$, which is used to sample from f.

- **Non-asymptotic convergence rates** of **flows** built from compositions of triangular maps, e.g., **Real NVP**. Some of the first statistical guarantees for flow models.

Check out our paper here:

References

1. I. Kobyzev, S. Prince, and M. Brubaker. Normalizing flows: An introduction and review of current methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.
2. L. Dinh, J. Sohl-Dickstein, S. Bengio. Density estimation using Real NVP. *ICLR*, 2017.
3. G. Carlier, A. Galichon, and F. Santambrogio. From Knothe's transport to Brenier's map and a continuation method for optimal transport. *SIAM Journal on Mathematical Analysis*, 41(6):2554–2576, 2010
4. A. Spantini, D. Bigoni, and Y. Marzouk. Inference via low-dimensional couplings. *The Journal of Machine Learning Research*, 19(1):2639–2709, 2018.
5. L. Birgé. On estimating a density using Hellinger distance and some other strange facts. *Probab. Th. Rel. Fields*, 71:271–291, 1986.