Flat Metric Minimization with Applications in Generative Modeling

Thomas Möllenhoff Daniel Cremers

Technical University of Munich

Motivation

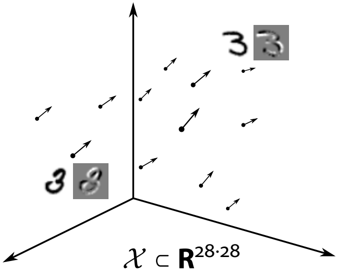
Latent concepts often induce an *orientation* of the data.

Simard et al. 1992, 1998; Rifai et al. 2011

Motivation

Latent concepts often induce an *orientation* of the data.
Tangent vectors to the "data manifold":

- Stroke thickness or shear of MNIST digit.

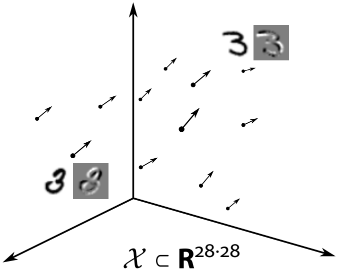


Simard et al. 1992, 1998; Rifai et al. 2011

Motivation

Latent concepts often induce an *orientation* of the data.
Tangent vectors to the “data manifold”:

- Stroke thickness or shear of MNIST digit.
- Camera position, lighting/material in a 3D scene.

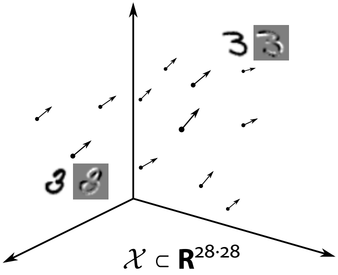


Simard et al. 1992, 1998; Rifai et al. 2011

Motivation

Latent concepts often induce an *orientation* of the data.
Tangent vectors to the “data manifold”:

- Stroke thickness or shear of MNIST digit.
- Camera position, lighting/material in a 3D scene.
- *Arrow of time* (videos, time-series data, ...)

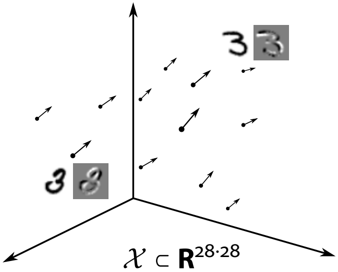


Simard et al. 1992, 1998; Rifai et al. 2011

Motivation

Latent concepts often induce an *orientation* of the data.
Tangent vectors to the "data manifold":

- Stroke thickness or shear of MNIST digit.
- Camera position, lighting/material in a 3D scene.
- *Arrow of time* (videos, time-series data, ...)



Contributions:

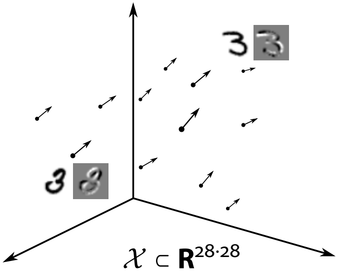
- We propose the novel perspective to represent oriented data with *k-currents* from *geometric measure theory*.

Simard et al. 1992, 1998; Rifai et al. 2011

Motivation

Latent concepts often induce an *orientation* of the data.
Tangent vectors to the “data manifold”:

- Stroke thickness or shear of MNIST digit.
- Camera position, lighting/material in a 3D scene.
- *Arrow of time* (videos, time-series data, ...)

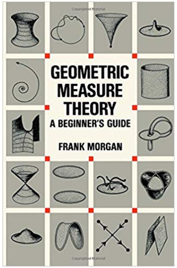


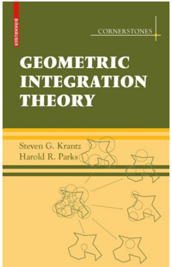
Contributions:

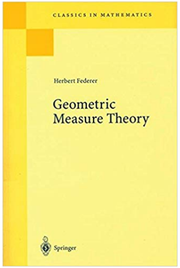
- We propose the novel perspective to represent oriented data with *k-currents* from *geometric measure theory*.
- Using this viewpoint within the context of GANs, we learn a generative model which behaves *equivariantly* to specified tangent vectors.

Simard et al. 1992, 1998; Rifai et al. 2011

An invitation to geometric measure theory (GMT)

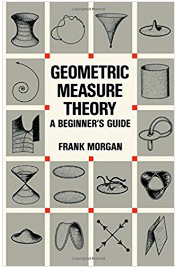


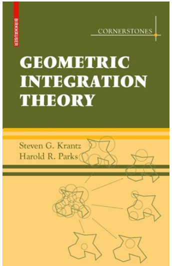


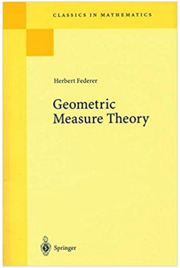


Morgan 2016, Krantz & Parks 2008, Federer 1969

An invitation to geometric measure theory (GMT)



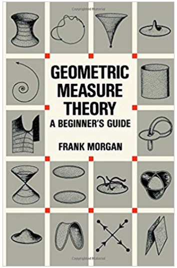


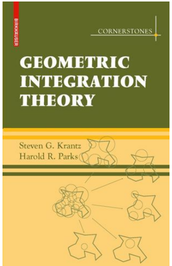


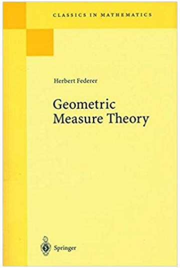
- Differential geometry, generalized through measure theory to deal with surfaces that are not necessarily smooth.

Morgan 2016, Krantz & Parks 2008, Federer 1969

An invitation to geometric measure theory (GMT)



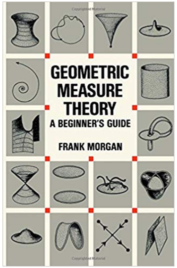


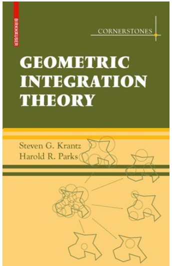


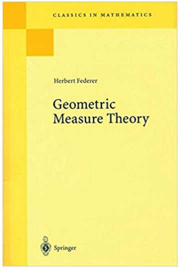
- Differential geometry, generalized through measure theory to deal with surfaces that are not necessarily smooth.
- k-currents $\approx$ generalized (possibly quite irregular) oriented k-dimensional surfaces in d-dimensional space.

Morgan 2016, Krantz & Parks 2008, Federer 1969

An invitation to geometric measure theory (GMT)



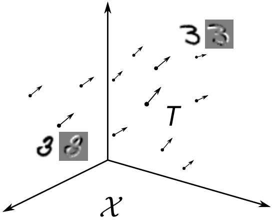


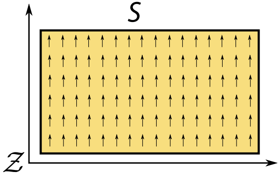


- Differential geometry, generalized through measure theory to deal with surfaces that are not necessarily smooth.
- k-currents ≈ generalized (possibly quite irregular) oriented k-dimensional surfaces in d-dimensional space.
- The class of currents we consider form a *linear space*. It includes oriented k-dimensional surfaces as elements.

Morgan 2016, Krantz & Parks 2008, Federer 1969

Generalizing Wasserstein GANs to *k*-currents

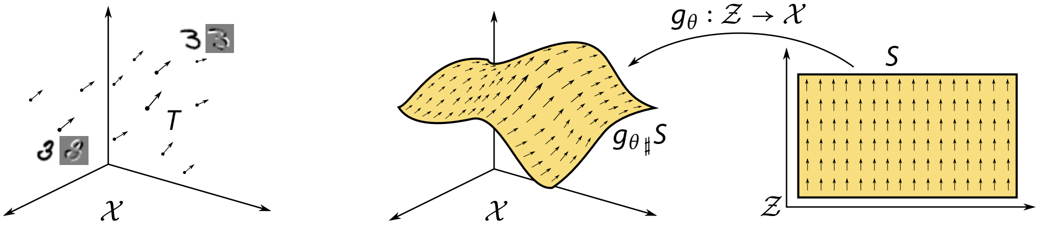




- T and S are 1-currents representing the data and the (partially oriented) latents.

Inspired by the optimal transport perspective on GANs: Bottou et al. 2017, Genevay et al. 2017

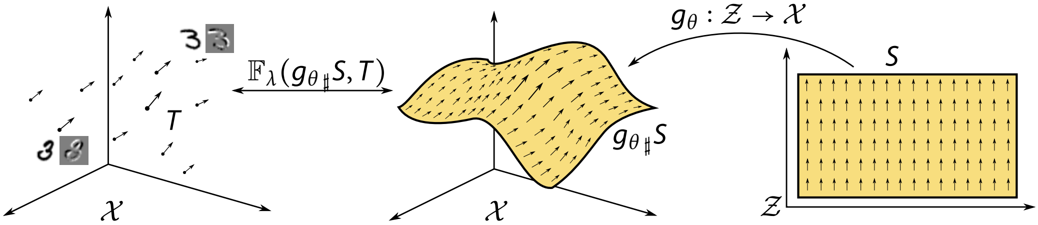
Generalizing Wasserstein GANs to k-currents



- T and S are 1-currents representing the data and the (partially oriented) latents.
- Pushforward operator $g_{\theta\,\sharp}$, yields transformed current $g_{\theta\,\sharp}S$.

Inspired by the optimal transport perspective on GANs: Bottou et al. 2017, Genevay et al. 2017

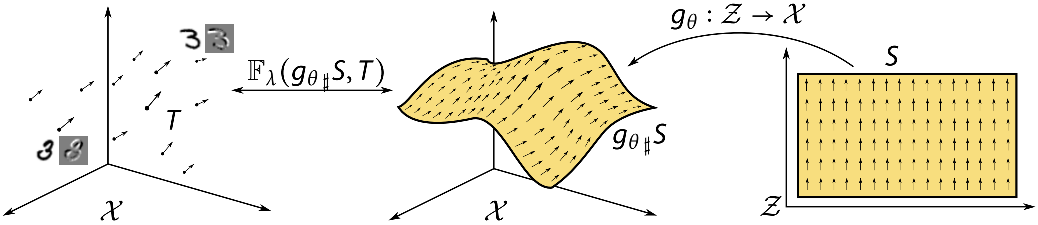
Generalizing Wasserstein GANs to k-currents



- T and S are 1-currents representing the data and the (partially oriented) latents.
- Pushforward operator $g_{\theta\,\sharp}$, yields transformed current $g_{\theta\,\sharp} S$.
- We propose to use the *flat metric* $\mathbb{F}_\lambda$ as a distance between $g_\sharp S$ and T.

Inspired by the optimal transport perspective on GANs: Bottou et al. 2017, Genevay et al. 2017

Generalizing Wasserstein GANs to k-currents

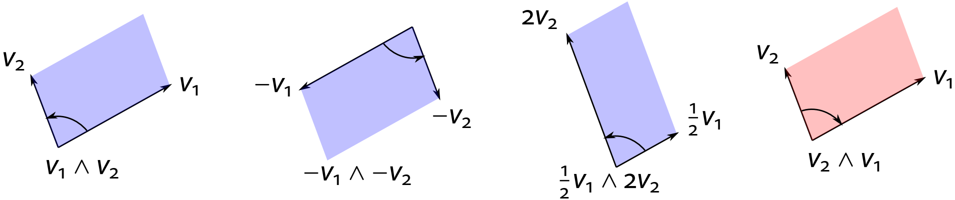


- T and S are 1-currents representing the data and the (partially oriented) latents.
- Pushforward operator $g_{\theta\,\sharp}$, yields transformed current $g_{\theta\,\sharp} S$.
- We propose to use the *flat metric* $\mathbb{F}_\lambda$ as a distance between $g_\sharp S$ and T.
- For $k = 0$ the flat metric is closely related to the Wasserstein−1 distance and positive 0-currents with unit mass are probability distributions.

Inspired by the optimal transport perspective on GANs: Bottou et al. 2017, Genevay et al. 2017

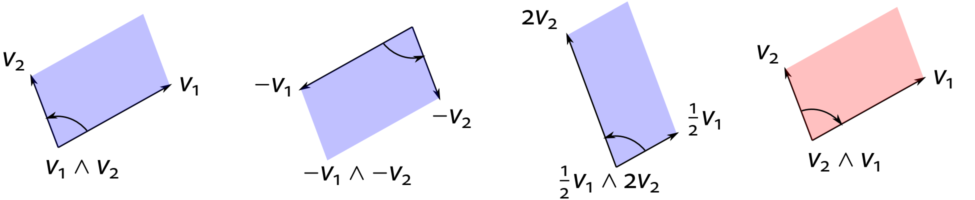
k-dimensional orientation in d-dimensional space

- Simple k-vectors $v = v_1 \wedge \cdots \wedge v_k \in \Lambda_k \mathbf{R}^d$ describe oriented k-dimensional subspaces together with an area in $\mathbf{R}^d$:



k-dimensional orientation in d-dimensional space

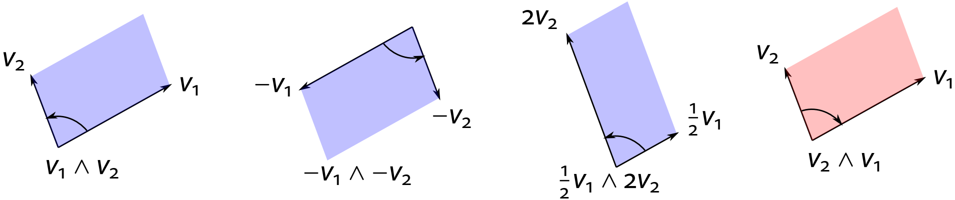
- Simple k-vectors $v = v_1 \wedge \cdots \wedge v_k \in \Lambda_k \mathbf{R}^d$ describe oriented k-dimensional subspaces together with an area in $\mathbf{R}^d$:



- The set of simple k-vectors forms a nonconvex cone in the vector space $\Lambda_k \mathbf{R}^d$.

k-dimensional orientation in d-dimensional space

- Simple k-vectors $v = v_1 \wedge \cdots \wedge v_k \in \Lambda_k \mathbf{R}^d$ describe oriented k-dimensional subspaces together with an area in $\mathbf{R}^d$:



- The set of simple k-vectors forms a nonconvex cone in the vector space $\Lambda_k \mathbf{R}^d$.
- For $v = v_1 \wedge \cdots \wedge v_k$, $w = w_1 \wedge \cdots \wedge w_k$:

$$\langle v, w \rangle = \det(V^\top W),\ |v| = \sqrt{\langle v, v \rangle}.$$

Oriented manifolds, differential forms and currents

- *Orientation* of a k-dimensional manifold $\mathcal{M}$: continuous simple k-vector map $\tau_{\mathcal{M}} : \mathcal{M} \to \Lambda_k \mathbf{R}^d$, $|\tau_{\mathcal{M}}(z)| = 1$ and $T_z\mathcal{M}$ “spanned” by $\tau_{\mathcal{M}}(z)$

Oriented manifolds, differential forms and currents

- *Orientation* of a k-dimensional manifold $\mathcal{M}$: continuous simple k-vector map $\tau_{\mathcal{M}} : \mathcal{M} \to \Lambda_k \mathbf{R}^d$, $|\tau_{\mathcal{M}}(z)| = 1$ and $T_z\mathcal{M}$ "spanned" by $\tau_{\mathcal{M}}(z)$
- *Differential form*: k-covector field $\omega : \mathbf{R}^d \to \Lambda^k \mathbf{R}^d$

Oriented manifolds, differential forms and currents

- *Orientation* of a k-dimensional manifold $\mathcal{M}$: continuous simple k-vector map $\tau_{\mathcal{M}} : \mathcal{M} \to \Lambda_k \mathbf{R}^d$, $|\tau_{\mathcal{M}}(z)| = 1$ and $T_z\mathcal{M}$ "spanned" by $\tau_{\mathcal{M}}(z)$
- *Differential form*: k-covector field $\omega : \mathbf{R}^d \to \Lambda^k \mathbf{R}^d$
- Integration of a k-form over an oriented k-dimensional manifold:

$$\int_{\mathcal{M}} \omega := \int_{\mathcal{M}} \langle \omega(z), \tau_{\mathcal{M}}(z) \rangle \, d\mathcal{H}^k(z) = [\![\mathcal{M}]\!](\omega)$$

Oriented manifolds, differential forms and currents

- *Orientation* of a k-dimensional manifold $\mathcal{M}$: continuous simple k-vector map $\tau_{\mathcal{M}} : \mathcal{M} \to \Lambda_k \mathbf{R}^d$, $|\tau_{\mathcal{M}}(z)| = 1$ and $T_z\mathcal{M}$ "spanned" by $\tau_{\mathcal{M}}(z)$
- *Differential form*: k-covector field $\omega : \mathbf{R}^d \to \Lambda^k \mathbf{R}^d$
- Integration of a k-form over an oriented k-dimensional manifold:

$$\int_{\mathcal{M}} \omega := \int_{\mathcal{M}} \langle \omega(z), \tau_{\mathcal{M}}(z) \rangle \, d\mathcal{H}^k(z) = [\![\mathcal{M}]\!](\omega)$$

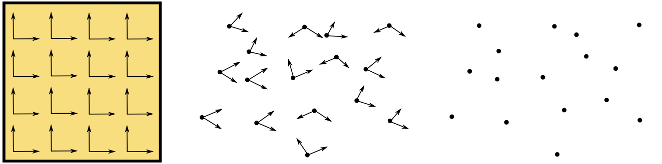
- $[\![\mathcal{M}]\!]$ is a k-current. In general, they are continuous linear functionals acting on compactly supported smooth k-forms

Oriented manifolds, differential forms and currents

- *Orientation* of a k-dimensional manifold $\mathcal{M}$: continuous simple k-vector map $\tau_{\mathcal{M}} : \mathcal{M} \to \Lambda_k \mathbf{R}^d$, $|\tau_{\mathcal{M}}(z)| = 1$ and $T_z\mathcal{M}$ "spanned" by $\tau_{\mathcal{M}}(z)$
- *Differential form*: k-covector field $\omega : \mathbf{R}^d \to \Lambda^k \mathbf{R}^d$
- Integration of a k-form over an oriented k-dimensional manifold:

$$\int_{\mathcal{M}} \omega := \int_{\mathcal{M}} \langle \omega(z), \tau_{\mathcal{M}}(z) \rangle \, d\mathcal{H}^k(z) = [\![\mathcal{M}]\!](\omega)$$

- $[\![\mathcal{M}]\!]$ is a k-current. In general, they are continuous linear functionals acting on compactly supported smooth k-forms



2-current discrete 2-current discrete 0-current

Towards a distance between k-currents

- Mass of a k-current: $\mathbb{M}(T) = \sup_{\|\omega\|^* \leq 1} T(\omega)$

Towards a distance between k-currents

- Mass of a k-current: $\mathbb{M}(T) = \sup_{\|\omega\|^* \leq 1} T(\omega)$
- The boundary operator ∂ maps a k-current to a $(k-1)$-current: $\partial T(\omega) = T(d\omega)$

Towards a distance between k-currents

- Mass of a k-current: $\mathbb{M}(T) = \sup_{\|\omega\|^* \leq 1} T(\omega)$
- The boundary operator ∂ maps a k-current to a $(k-1)$-current: $\partial T(\omega) = T(d\omega)$
- Stokes' theorem:

$$\int_{\partial\mathcal{M}} \omega = \int_{\mathcal{M}} d\omega.$$

Towards a distance between k-currents

- Mass of a k-current: $\mathbb{M}(T) = \sup_{\|\omega\|^* \leq 1} T(\omega)$
- The boundary operator ∂ maps a k-current to a $(k-1)$-current: $\partial T(\omega) = T(d\omega)$
- Stokes' theorem:

$$\int_{\partial \mathcal{M}} \omega = \int_{\mathcal{M}} d\omega.$$

- *Normal currents* $T \in N_{k,\mathcal{X}}(\mathbf{R}^d)$: Finite mass and boundary mass $\mathbb{M}(T) + \mathbb{M}(\partial T) < \infty$

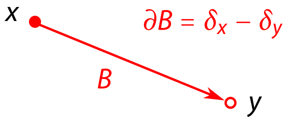
Towards a distance between k-currents

- Mass of a k-current: $\mathbb{M}(T) = \sup_{\|\omega\|^* \leq 1} T(\omega)$
- The boundary operator ∂ maps a k-current to a $(k-1)$-current: $\partial T(\omega) = T(d\omega)$
- Stokes' theorem:

$$\int_{\partial\mathcal{M}} \omega = \int_{\mathcal{M}} d\omega.$$

- *Normal currents* $T \in N_{k,\mathcal{X}}(\mathbf{R}^d)$: Finite mass and boundary mass $\mathbb{M}(T) + \mathbb{M}(\partial T) < \infty$
- A geometric view on the Wasserstein−1 distance:

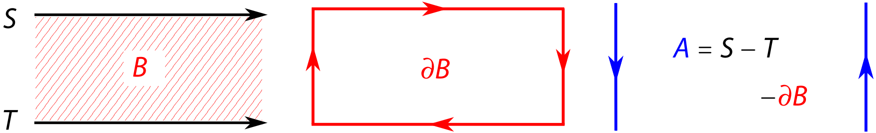
$$\mathcal{W}_1(S,T) = \min_{\partial B = S - T} \mathbb{M}(B).$$ Example: $S = \delta_x$, $T = \delta_y$:



The flat metric

Given two normal k-currents $S \in N_{k,\mathcal{X}}(\mathbf{R}^d)$, $T \in N_{k,\mathcal{X}}(\mathbf{R}^d)$ the flat metric as defined as

$$\mathbb{F}_\lambda(S,T) = \min_{S-T=\partial B+A} \mathbb{M}(B) + \lambda \mathbb{M}(A) = \sup_{\substack{\|\omega\|^* \le \lambda \\ \|d\omega\|^* \le 1}} (S-T)(\omega).$$

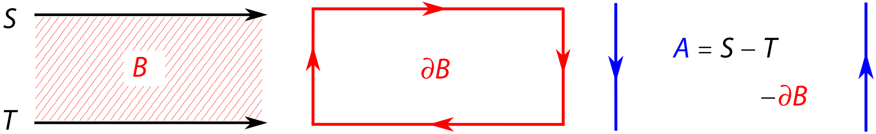


Whitney 1957, Federer & Fleming 1960

The flat metric

Given two normal k-currents $S \in N_{k,\mathcal{X}}(\mathbf{R}^d)$, $T \in N_{k,\mathcal{X}}(\mathbf{R}^d)$ the flat metric as defined as

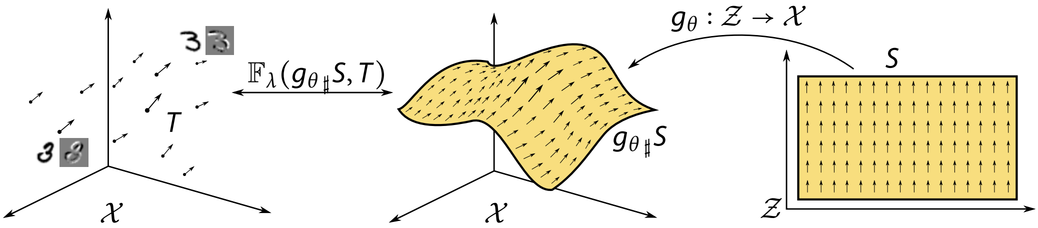
$$\mathbb{F}_\lambda(S,T) = \min_{S-T=\partial B+A} \mathbb{M}(B) + \lambda\mathbb{M}(A) = \sup_{\substack{\|\omega\|^* \leq \lambda \\ \|d\omega\|^* \leq 1}} (S-T)(\omega).$$



Federer & Fleming 1960: The flat metric metrizes the weak* convergence on normal currents with uniformly bounded mass and boundary mass.

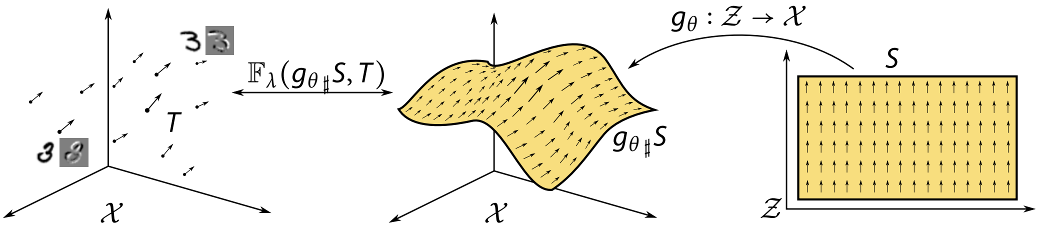
Whitney 1957, Federer & Fleming 1960

Flat metric minimization: our theoretical result



$$\min_{\theta \in \Theta} \mathbb{F}_\lambda(g_{\theta\,\sharp}S, T)$$

Flat metric minimization: our theoretical result

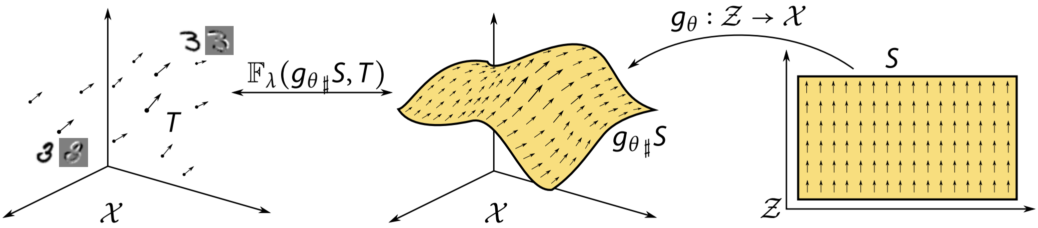


$$\min_{\theta \in \Theta} \mathbb{F}_\lambda(g_{\theta\,\sharp} S, T)$$

Assumptions:

- Normal currents $S \in N_{k,\mathcal{Z}}(\mathbf{R}^l)$, $T \in N_{k,\mathcal{X}}(\mathbf{R}^d)$.
- $g : \mathcal{Z} \times \Theta \to \mathcal{X}$ smooth in z with uniformly bounded derivative, loc. Lipschitz in θ.
- Parameter space Θ is compact.

Flat metric minimization: our theoretical result



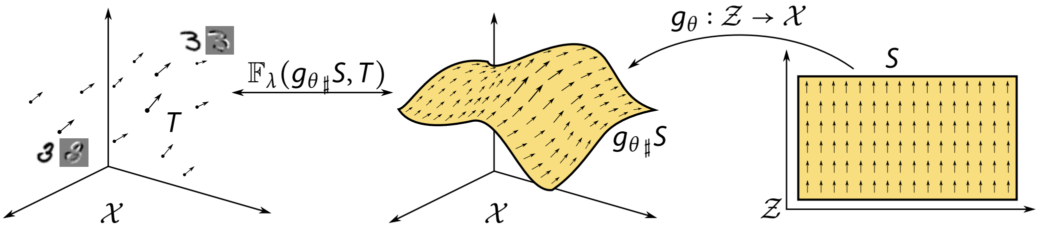
$$\min_{\theta \in \Theta} \mathbb{F}_\lambda(g_{\theta\,\sharp}S, T)$$

Assumptions:

- Normal currents $S \in N_{k,\mathcal{Z}}(\mathbf{R}^l)$, $T \in N_{k,\mathcal{X}}(\mathbf{R}^d)$.
- $g : \mathcal{Z} \times \Theta \to \mathcal{X}$ smooth in z with uniformly bounded derivative, loc. Lipschitz in θ.
- Parameter space Θ is compact.

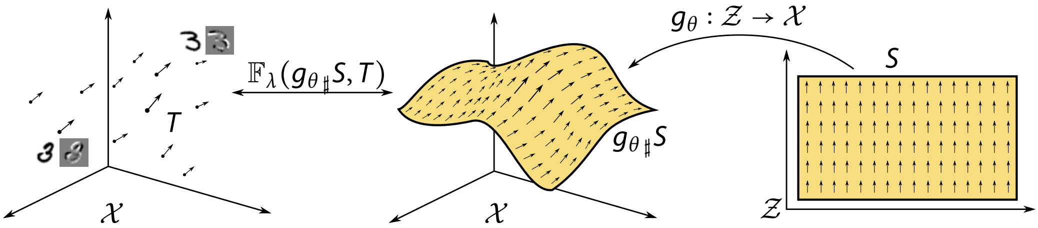
Proposition. The map $\theta \mapsto \mathbb{F}_\lambda(g_{\theta\,\sharp}S, T)$ is Lipschitz continuous.

FlatGAN formulation and implementation



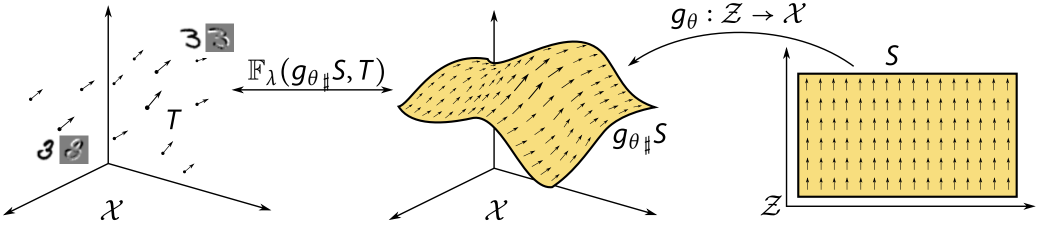
$$\min_{\theta \in \Theta} \mathbb{F}_\lambda(g_{\theta\,\sharp}S, T)$$

FlatGAN formulation and implementation



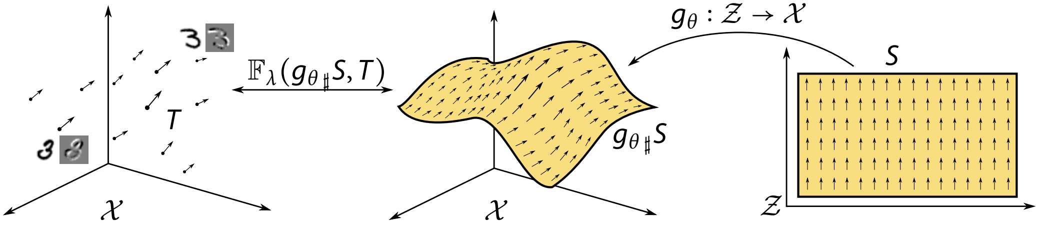
$$\min_{\theta\in\Theta} \sup_{\substack{\|\omega\|^* \leq \lambda \\ \|d\omega\|^* \leq 1}} (g_{\theta\,\sharp}S - T)(\omega)$$

FlatGAN formulation and implementation



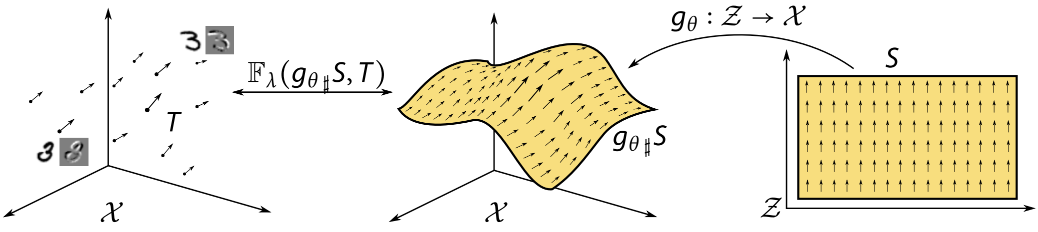
$$\min_{\theta \in \Theta} \sup_{\substack{\|\omega\|^* \leq \lambda \\ \|d\omega\|^* \leq 1}} g_{\theta\sharp}S(\omega) - T(\omega)$$

FlatGAN formulation and implementation



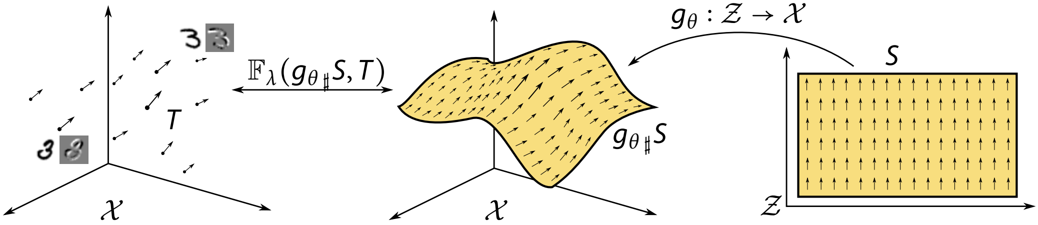
$$\min_{\theta \in \Theta} \sup_{\substack{\|\omega\|^* \le \lambda \\ \|d\omega\|^* \le 1}} S(g_\theta{}^\sharp \omega) - T(\omega)$$

FlatGAN formulation and implementation



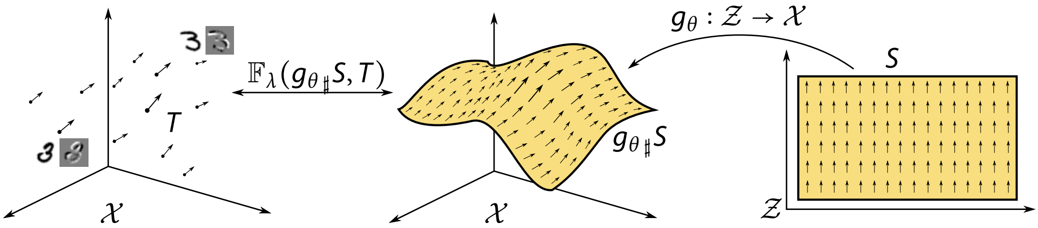
$$\min_{\theta \in \Theta} \sup_{\substack{\|\omega\|^* \leq \lambda \\ \|d\omega\|^* \leq 1}} S({g_\theta}^\sharp \omega) - T(\omega) \qquad T = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i} \wedge (T_{i,1} \wedge \cdots \wedge T_{i,k}), S = \mu \wedge (e_1 \wedge \cdots \wedge e_k)$$

FlatGAN formulation and implementation



$$\min_{\theta \in \Theta} \sup_{\substack{\|\omega\|^* \leq \lambda \\ \|d\omega\|^* \leq 1}} \mathbb{E}_{z \sim \mu} \left[\langle \omega \circ g_\theta, (\nabla_z g_\theta \cdot e_1) \wedge \cdots \wedge (\nabla_z g_\theta \cdot e_k) \rangle \right] - \frac{1}{N} \sum_{i=1}^{N} \langle \omega(x_i), T_{i,1} \wedge \cdots \wedge T_{i,k} \rangle$$

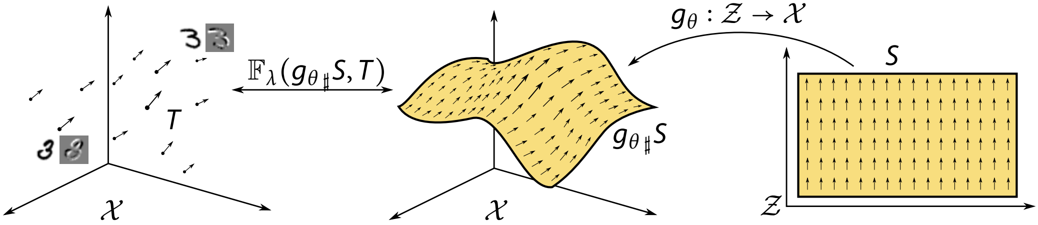
FlatGAN formulation and implementation



$$\min_{\theta \in \Theta} \sup_{\substack{\|\omega\|^* \le \lambda \\ \|d\omega\|^* \le 1}} \mathbb{E}_{z \sim \mu}\left[\langle \omega \circ g_\theta, (\nabla_z g_\theta \cdot e_1) \wedge \cdots \wedge (\nabla_z g_\theta \cdot e_k) \rangle\right] - \frac{1}{N} \sum_{i=1}^{N} \langle \omega(x_i), T_{i,1} \wedge \cdots \wedge T_{i,k} \rangle$$

- Implement $\omega : \mathbf{R}^d \to \Lambda^k \mathbf{R}^d$ and $g_\theta : \mathcal{Z} \to \mathcal{X}$ with deep nets

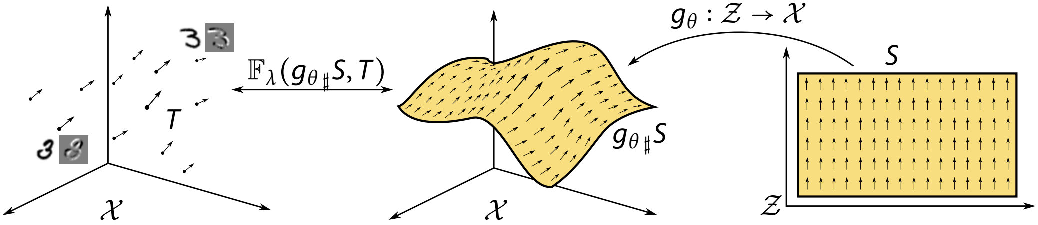
FlatGAN formulation and implementation



$$\min_{\theta \in \Theta} \sup_{\substack{\|\omega\|^* \le \lambda \\ \|d\omega\|^* \le 1}} \mathbb{E}_{z \sim \mu} \left[\langle \omega \circ g_\theta, (\nabla_z g_\theta \cdot e_1) \wedge \cdots \wedge (\nabla_z g_\theta \cdot e_k) \rangle \right] - \frac{1}{N} \sum_{i=1}^{N} \langle \omega(x_i), T_{i,1} \wedge \cdots \wedge T_{i,k} \rangle$$

- Implement $\omega : \mathbf{R}^d \to \Lambda^k \mathbf{R}^d$ and $g_\theta : \mathcal{Z} \to \mathcal{X}$ with deep nets
- Soft penalty for $\|\omega(x)\|^* \le \lambda$, $\|d\omega(x)\|^* \le 1$ (similar to WGAN-GP)

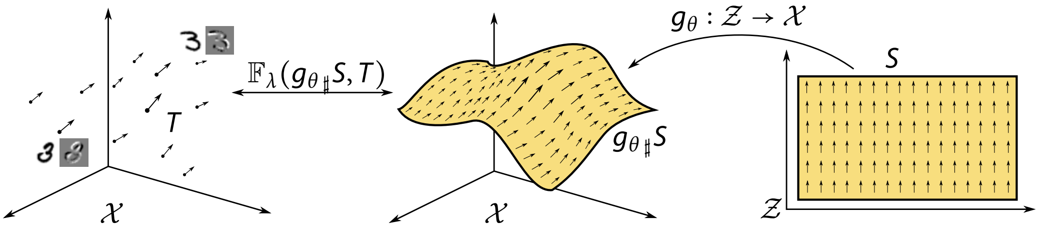
FlatGAN formulation and implementation



$$\min_{\theta\in\Theta} \sup_{\substack{\|\omega\|^*\leq\lambda \\ \|d\omega\|^*\leq 1}} \mathbb{E}_{z\sim\mu}\left[\langle \omega \circ g_\theta, (\nabla_z g_\theta \cdot e_1) \wedge \cdots \wedge (\nabla_z g_\theta \cdot e_k)\rangle\right] - \frac{1}{N}\sum_{i=1}^{N}\langle \omega(x_i), T_{i,1} \wedge \cdots \wedge T_{i,k}\rangle$$

- Implement $\omega : \mathbf{R}^d \to \Lambda^k \mathbf{R}^d$ and $g_\theta : \mathcal{Z} \to \mathcal{X}$ with deep nets
- Soft penalty for $\|\omega(x)\|^* \leq \lambda$, $\|d\omega(x)\|^* \leq 1$ (similar to WGAN-GP)
- Compute $\nabla_z g_\theta \cdot e_i$ with two calls to autograd (`rop`), $\langle\cdot,\cdot\rangle$ by $k \times k$-determinant

FlatGAN formulation and implementation



$$\min_{\theta \in \Theta} \sup_{\substack{\|\omega\|^* \leq \lambda \\ \|d\omega\|^* \leq 1}} \mathbb{E}_{z \sim \mu}\left[\langle \omega \circ g_\theta, (\nabla_z g_\theta \cdot e_1) \wedge \cdots \wedge (\nabla_z g_\theta \cdot e_k)\rangle\right] - \frac{1}{N}\sum_{i=1}^{N} \langle \omega(x_i), T_{i,1} \wedge \cdots \wedge T_{i,k}\rangle$$

- Implement $\omega : \mathbf{R}^d \to \Lambda^k \mathbf{R}^d$ and $g_\theta : \mathcal{Z} \to \mathcal{X}$ with deep nets
- Soft penalty for $\|\omega(x)\|^* \leq \lambda$, $\|d\omega(x)\|^* \leq 1$ (similar to WGAN-GP)
- Compute $\nabla_z g_\theta \cdot e_i$ with two calls to autograd (`rop`), $\langle \cdot, \cdot \rangle$ by $k \times k$-determinant
- Train model by alternating stochastic gradient ascent/descent

Illustration on a 2D toy data set (5 points on a circle)

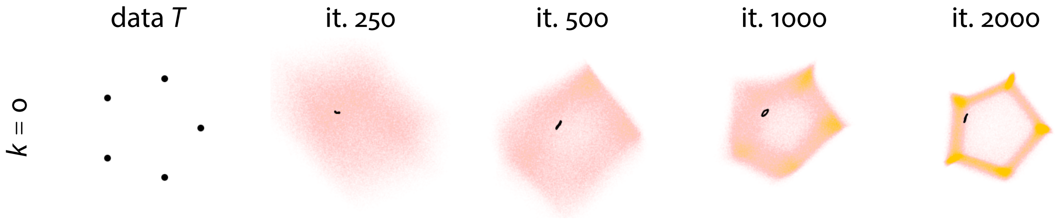
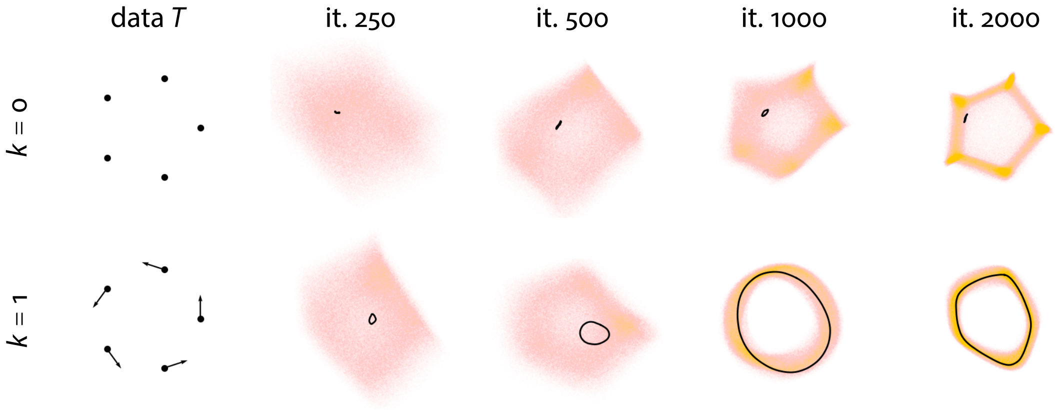


Illustration on a 2D toy data set (5 points on a circle)



Learning equivariant latent representations

MNIST, $k = 2$

varying z_1 (rotation)

varying z_2 (stroke width)

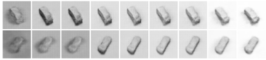
Learning equivariant latent representations

MNIST, $k = 2$

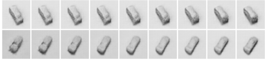
varying z_1 (rotation)

varying z_2 (stroke width)

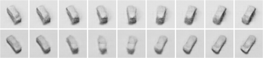
smallNORB, $k = 3$



varying z_1 (lighting)



varying z_2 (elevation)



varying z_3 (azimuth)

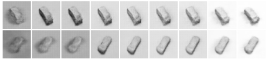
Learning equivariant latent representations

MNIST, $k = 2$

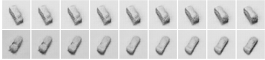
varying z_1 (rotation)

varying z_2 (stroke width)

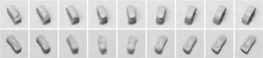
smallNORB, $k = 3$



varying z_1 (lighting)

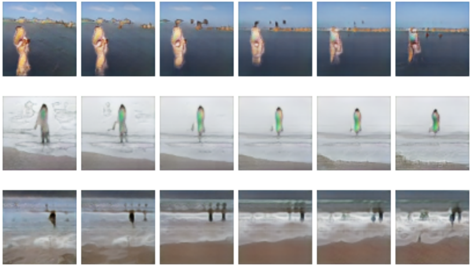


varying z_2 (elevation)



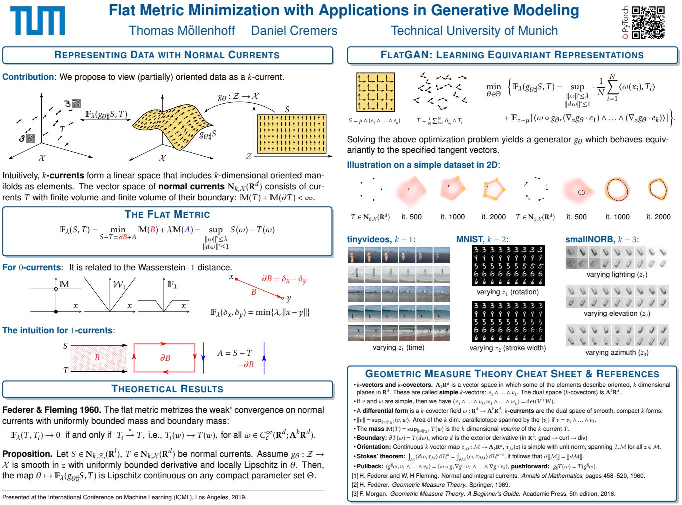
varying z_3 (azimuth)

tinyvideos, $k = 1$



varying z_1 (time)

See you at our poster, Pacific Ballroom #16, 6:30 tonight!



PyTorch implementation: `https://github.com/moellenh/flatgan`