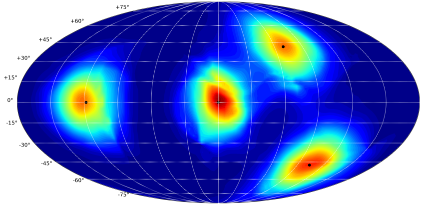
Circle Slice Flows and the Variational Determinant Estimator

Simon Passenheim

University of Amsterdam

20th of January, 2021



Contents

Introduction

Most work concerning Normalizing Flows (Rezende, Mohamed, 2015; Dinh et al., 2016; Kingma, Dhariwal, 2018) so far considered density estimation on Euclidean spaces.

Manifold hypothesis (Fefferman et al., 2016) states that real-world high dimensional data like images often lies on lower-dimensional manifold.

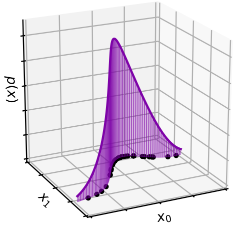


Figure: Data populating a one dimensional manifold embedded in $\mathbb{R}^2$. Plot from Brehmer, Cranmer (2020).

Circle Slice Flows

- Novel method to perform density estimation on $\mathbb{S}^D$ which allows uniform prior on $\mathbb{S}^D$
- *Circle Slices* $(x_i, x_j) \in \mathbb{S}^D$ are transformed with univariate flows and sphere slices with rotations
- Benchmarked against Cylindrical flows by Rezende et al. (2020)

Variational Determinant Estimator

- Novel method based on stochastic estimator by Sohl-Dickstein (2020) which solely depends on matrix vector products
- A sample generator distribution is modeled with spherical flows, reducing the variance of the Monte Carlo estimator
- Work has been accepted at AABI workshop 2021

Euclidean Change of Variables if we go from a *simple* space Z to a *complex* space X with $f\colon Z \to X$, then p on X

$$\log p(x) = \log \pi(z) - \log|\det J_f(z)|$$

Manifold Change of Variables Formula Let f be a flow between D-dim manifolds $\mathcal{M}$ and $\mathcal{N}$

$$f\colon \mathcal{M} \subset \mathbb{R}^m \to \mathcal{N} \subset \mathbb{R}^n$$

and π a distribution on $\mathcal{M}$, p distr. on $\mathcal{N}$, then

$$\log p(x) = \log \pi(z) - \log \sqrt{\det \left(J_f(z)E(z)\right)^\top J_f(z)E(z)} \qquad (1)$$

where E is the $m \times D$ matrix of tangent vectors at $T_z\mathcal{M}$.

Moebius Flow (MF) is a diffeomorphism $h\colon r\mathbb{S}^D \to r\mathbb{S}^D$ and used by Wang, Gelfand (2013); Rezende et al. (2020) for $z \in \mathbb{R}^{D+1}$

$$h_\omega(z) = \frac{r^2 - \|\omega\|^2}{\|z-\omega\|^2}(z-\omega) - \omega \quad \text{with} \quad \|\omega\| < r$$

but can only use it to define flow on $r\mathbb{S}^1$.

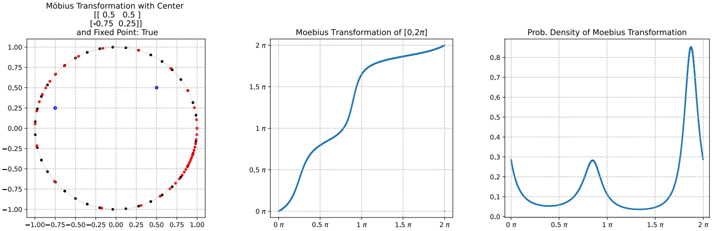
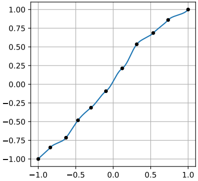


Figure: Left: image of 50 unifrom samples, mid: Moebius Flow in angle space, right: implied density

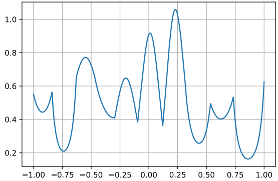
Interval Spline Flow (ISF) (Durkan et al., 2019) diffeomorphism on $[a, b]$ via piecewise rational-quadratic functions on K bins

$$f(x) = \frac{a_{k2}x^2 + a_{k1}x + a_{k0}}{b_{k2}x^2 + b_{k1}x + b_{k0}}$$

Can be easily extended to **Circular Spline Flows (CSF)** on $\mathbb{S}^1$ as seen as $[0, 2\pi]$.



(a) Interval Spline Flow



(b) Implied density

Figure: Neural Spline Flow with randomly initialized parameters.

Unfolding the Sphere via $T_{s\to c}\colon \mathbb{S}^D \to \mathbb{S}^1 \times [-1,1]^{D-1}$ and

$$T_{s\to c}\colon (x_{1:D+1}) \mapsto \left(\frac{x_{1:2}}{\sqrt{1-\sum_{i=3}^{D+1} x_i^2}}, \ldots, \frac{x_k}{\sqrt{1-\sum_{i=k+1}^{D+1} x_i^2}}, \ldots, x_{D+1} \right)$$

The circle $\mathbb{S}^1$ is transformed via MF or CSF and $[-1,1]^{D-1}$ with ISF.

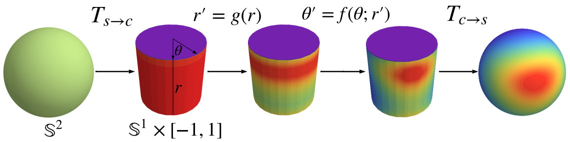


Figure: Cylindrical flow acting on $\mathbb{S}^2$ autoregressively. Plot is taken from Rezende et al. (2020).

Circle Slice Flows

Circle Slice Flows

A Slice of Pie

Sphere Slice Each subset $(x_{i_1}, \ldots, x_{i_n}) \subset \mathbb{S}^D$ lies on $r\mathbb{S}^{n-1}$ with

$$r^2 = 1 - \sum_{k \neq i_1, i_2, \ldots, i_n} x_k^2$$

Circle Slice
In particular, $(x_i, x_j) \subset \mathbb{S}^D$ defines a one dimensional circle $r\mathbb{S}^1$ on the sphere.

Angle Identification
Each pair (x_i, x_j) can be bijectively identified with an angle $\theta \in [0, 2\pi)$ via

$$\theta = \operatorname{arctan2}(x_j, x_i) \mod 2\pi$$

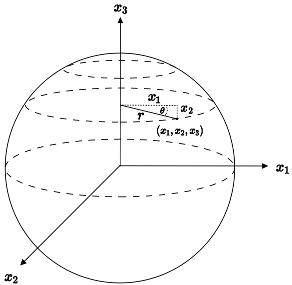


Figure: Dashed lines represent circle slices for varying x_3.

Circle Slice Flows

Architecture

Split of the data can be interpreted as slicing of $\mathbb{S}^D$ into two subspheres $r\mathbb{S}^{d-1} \times r'\mathbb{S}^k$.

Rotations purpose is two-fold.

Circle Slices are transformed with MF or CSF by considering independent pairs of Cartesian coordinates.

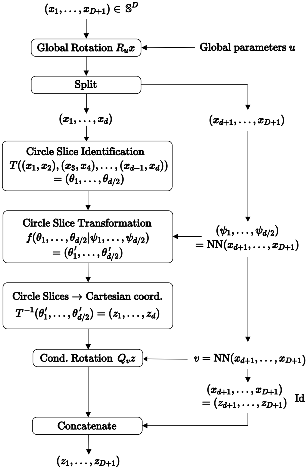


Figure: Architecture of *Circle Slice Flows*

Global Rotation acts as learnable permutation of the data and inspired by 1×1 convolution of Kingma, Dhariwal (2018).

Conditional Rotation additionally acts as arbitrary phase translation.

Orthogonal Transformation Strictly speaking our rotations are orthogonal transformations $R \in \mathrm{O}(D)$ with $|\det J_R| = 1$.

Householder Reflections Orthogonal matrix $R \in \mathrm{O}(D)$ can be expressed as composition of up to D Householder reflections

$$R_i = I - 2\frac{u_i u_i^\top}{\|u_i\|^2} \quad \text{with} \quad u_i \in \mathbb{R}^D$$

University of Amsterdam

Circle Slice Identification

$$T\colon r_1\mathbb{S}^1 \times r_2\mathbb{S}^1 \times \ldots \times r_{d/2}\mathbb{S}^1 \to [0, 2\pi)^{d/2}$$
$$T\colon (x_1, x_2), (x_3, x_4), \ldots, (x_{d-1}, x_d) \mapsto (\theta_1, \ldots, \theta_{d/2})$$

Circle Slice Transformation with either Moebius or Circular Spline flow and parameters $\{\psi_i\}_i$

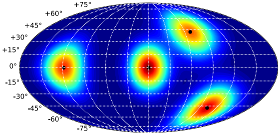
$$f\colon \left(\theta_1, \ldots, \theta_{d/2}\right) \mapsto \left(f\left(\theta_1|\psi_1\right), \ldots, f(\theta_{d/2}|\psi_{d/2})\right)$$

Jacobian of T and T^{-1} cancel out each other since radii are left invariant under MF and CSF and update rule via Equation 1

$$\log\sqrt{\det G_T} = \sum_i \log r_i$$

UNIVERSITY OF AMSTERDAM

Toy Density Consider mixture of Power Spherical Distributions (De Cao, Aziz, 2020). Fast sampling procedure.



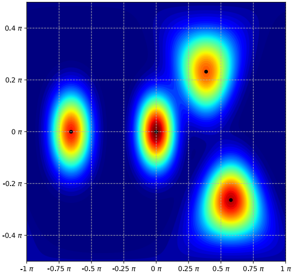


Figure: Synthetic target density on $\mathbb{S}^2$. Parameters can be found on git.

Architecture:
Rezende et al. (2020) used only a low number of dims. Therefore, we chose powers of 2 and $D = 3$.

Training Parameters

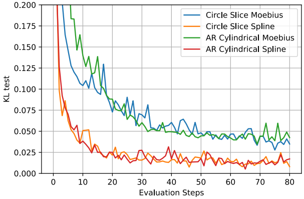
- In total 8k iterations
- Batch 256
- AdamW optimizer
- GTX 1080 Ti 11GB RAM

Dim	N_C	N_B	N_F	h_{dim} MLP
3	8	8	3	128
32	8	12	3	128
64	8	12	4	128
128	8	12	5	128

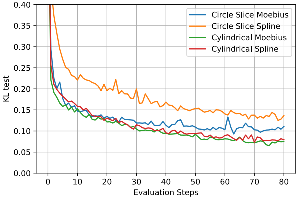
Table: Model parameters, both for Circle Slice and Cylindrical flows.

Circle Slice Flows

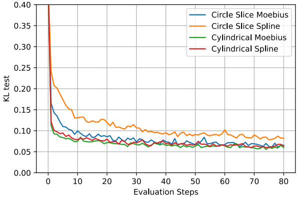
Experiment 1: Results



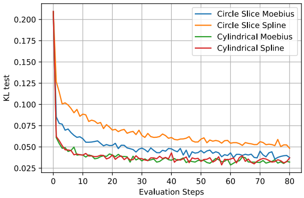
(a) KL in dimension $D = 3$



(b) KL in dimension $D = 32$



(c) KL in dimension $D = 64$



(d) KL in dimension $D = 128$

Circle Slice Flows

Experiment 2: Numerical Instabilities

Unit Norm often appears naturally in high dimensional data Dryden, others (2005). For example, in shape and curve analysis a data set might be recorded at arbitrary scales and only the general shape of the data points is the object of interest.

	Cylindrical Flows			Circle Slice Flows
Dim	NaN in $\mathbb{S}^1$	Heights outside $[-1,1]^{D-1}$	$(z_1, z_2) \in \mathbb{S}^1$ outside $[-1,1]^2$	Num. Instability
64	6	15	3k	**0**
128	14	33	6k	**0**
256	41	85	12k	**0**
512	105	196	24k	**0**

Table: Occurrences of numerical instabilities when unfolding $\mathbb{S}^D \to \mathbb{S}^1 \times [-1,1]^{D-1}$ for $50k$ uniform samples. Numbers are rounded to full thousands in the last column.

The Variational Determinant Estimator

UNIVERSITY OF AMSTERDAM

Original Estimator by Sohl-Dickstein (2020). Let $A \in \mathbb{R}^{D \times D}$ be non-singular matrix:

$$|A|^{-1} = \mathbb{E}_{s \sim \mathcal{U}(\mathbb{S}^{D-1})} \left[\|As\|^{-D} \right]$$

VDE Introduction of variational distribution $q \colon \mathbb{S}^{D-1} \to \mathbb{R}^{+}$ acts as variance reducer:

$$|A|^{-1} = \mathbb{E}_{s \sim q(s)} \left[\frac{\mathcal{U}(s)}{q(s)} \|As\|^{-D} \right]$$

Zero Variance Variational Distribution According to Owen (2013) optimal variational distribution:

$$q^{*}(s) \propto \|As\|^{-D}$$

Our Contribution Model variational distribution $q(s)$ by Cylindrical flow $f\colon \mathbb{S}^{D-1} \to \mathbb{S}^{D-1}$ (Rezende et al., 2020) or our own Circle Slice Flows.

Objective Plugging in change of variables formula and modeling the flow as from *simple* space to *complex* space $f\colon Z \to X$ yields

$$\overline{\mathrm{KL}}(q(s)\,;\,\mathcal{U}(s)\|As\|^{-D}) = \\ \mathbb{E}_{s_0\sim\mathcal{U}(s)}\left[-\log|\det J_f(s_0)| + n\log\|Af(s_0)\|\right]$$

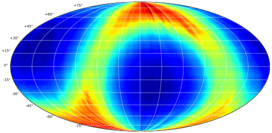
Upper Bound Property

$$\log|A| = -\log\mathbb{E}_{s\sim q(s)}\left[\frac{\mathcal{U}(s)}{q(s)}\|As\|^{-D}\right] \\ \leq -\mathbb{E}_{s\sim q(s)}\left[\log\frac{\mathcal{U}(s)}{q(s)}\|As\|^{-D}\right] = \overline{\mathrm{KL}}(q(s)\,;\,\mathcal{U}(s)\|As\|^{-D})$$

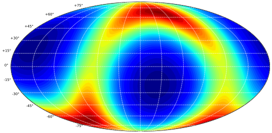
Example in $D = 3$

Matrix A rounded to one decimal:

$$A = \begin{bmatrix} -0.7 & 0.7 & -0.5 \\ 0.9 & 1.1 & 0.1 \\ -1.3 & -0.2 & 1.0 \end{bmatrix}$$



(e) Optimized variational distribution $q(s)$ on $\mathbb{S}^2$.



(f) Optimal variational distribution $q^*(s) = \|As\|^{-3}/Z$ on $\mathbb{S}^2$.

Figure: Learned variational distribution modeled by a Cylindrical flow and optimal distribution.

Architecture Cylindrical flow (Rezende et al., 2020) on $\mathbb{S}^1 \times [-1,1]^{D-1}$. Moebius flow for circle part $\mathbb{S}^1$ and Interval Spline flow for interval $[-1,1]^{D-1}$.

N_C	N_B	N_F	h_{dim} MLP
12	16	8	64

Table: Model parameters

Dense Matrices Five 10×10 unit Gaussian sampled matrices and can be found on git.

Metric Relative mean absolute difference (RMD)

$$\mathrm{RMD} = \frac{1}{5}\sum_{i=1}^{5}\left|\frac{\widehat{|A_i|} - |A_i|}{|A_i|}\right|$$

Architecture and Parameters
Identical except $40k$ iterations.

Matrix of Conv. Operator
Structured, equivalent 16×16 matrix W of convolutional operator with a 3×3 filter w applied to an 4×4 image x. Can be found on git.

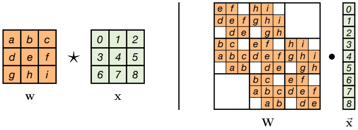


Figure: Illustration of the equivalent matrix W associated to a convolutional operator with a 3×3 filter w applied to an 3×3 image x. Plot from Hoogeboom et al. (2019).

Metric RMD and absolute det estimates.

UNIVERSITY OF AMSTERDAM

Dense Matrices

Nr. of samples	10^2	10^3	10^4	10^5
VDE det. (ours)	**3.4** ± 2.1 %	**1.7** ± 0.6 %	**1.6** ± 1.3 %	**0.3** ± 0.3 %
MC det.	533 ± 660 %	348 ± 262 %	104 ± 30 %	59 ± 43 %

Table: Results of dense matrix experiment in terms of RMD.

Structured Matrix

Nr. of samples	10^2	10^3	10^4	10^5	true determinant
VDE det. (ours)	7.62	7.71	7.64	7.70	7.71
MC det.	481.09	74.37	39.08	13.21	
VDE Rel. diff of det. (ours)	**1.1** %	**0.05** %	**0.9** %	**0.1** %	0 %
MC Rel. diff of det.	6144 %	865 %	407 %	71 %	

Table: Results of structured matrix experiment in terms of RMD and absolute estimates.

Conclusion

Takeaways

- Circle Slice Flows are on par in two out four density estimation experiments
- Still, Cylindrical Flows problematic in high dimensions
- Numerical instabilities solved by clamping and manually replacing NaNs with straight-trough gradients

Future Work

- Transform radius of circle slices by obeying boundary condition that overall radius remains constant
- Compose circle slice transformation with single phase translation

Conclusion

Takeaways

- The VDE yields accurate determinant estimates with low sample sizes
- Relies only on matrix-vector products
- We considered offline setting where variational distribution needs to be optimized first

Future Work

- Train VDE in parallel in an online setting to enable estimation of Jacobian determinant of unconstrained Normalizing Flows
- Could do this within single objective

References I

Brehmer Johann, Cranmer Kyle. Flows for simultaneous manifold learning and density estimation // arXiv preprint arXiv:2003.13913. 2020.

De Cao Nicola, Aziz Wilker. The Power Spherical distribution // arXiv preprint arXiv:2006.04437. 2020.

Dinh Laurent, Sohl-Dickstein Jascha, Bengio Samy. Density estimation using real nvp // arXiv preprint arXiv:1605.08803. 2016.

Dryden Ian L, others . Statistical analysis on high-dimensional spheres and shape spaces // The Annals of Statistics. 2005. 33, 4. 1643–1665.

Durkan Conor, Bekasov Artur, Murray Iain, Papamakarios George. Neural spline flows // Advances in Neural Information Processing Systems. 2019. 7511–7522.

References II

Fefferman Charles, Mitter Sanjoy, Narayanan Hariharan. Testing the manifold hypothesis // Journal of the American Mathematical Society. 2016. 29, 4. 983–1049.

Hoogeboom Emiel, Berg Rianne van den, Welling Max. Emerging convolutions for generative normalizing flows // arXiv preprint arXiv:1901.11137. 2019.

Kingma Durk P, Dhariwal Prafulla. Glow: Generative flow with invertible 1x1 convolutions // Advances in neural information processing systems. 2018. 10215–10224.

Owen Art B. Monte Carlo Theory // Methods and Examples. 2013. 665.

Rezende Danilo Jimenez, Mohamed Shakir. Variational inference with normalizing flows // arXiv preprint arXiv:1505.05770. 2015.

Rezende Danilo Jimenez, Papamakarios George, Racanière Sébastien, Albergo Michael S, Kanwar Gurtej, Shanahan Phiala E, Cranmer Kyle. Normalizing flows on tori and spheres // arXiv preprint arXiv:2002.02428. 2020.

Sohl-Dickstein Jascha. Two equalities expressing the determinant of a matrix in terms of expectations over matrix-vector products // arXiv preprint arXiv:2005.06553. 2020.

Wang Fangpo, Gelfand Alan E. Directional data analysis under the general projected normal distribution // Statistical methodology. 2013. 10, 1. 113–127.