

Generative Flows

Gabriele Steidl

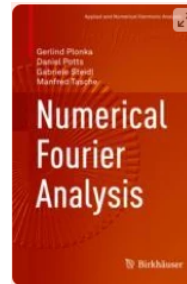
TU Berlin

Gefördert durch:



1	2
3	4
5	6
7	8
9	10
11	12
13	14
15	16
17	18
19	20
21	22
23	24
25	26
27	28
29	30
31	32
33	

0. Preliminaries

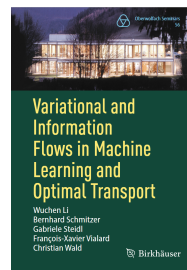


I. Flow Matching for Continuous Distributions

I.1 Motivation

I.2 Flow Matching

I.3 Numerical Examples



II. Flow Matching for Discrete Distributions

II.1 Motivation

II.2 Flow Matching

II.3 Numerical Examples

III. SOL: Measuring Gaps between Text Distributions by Double Sliced Wasserstein Metrics

1	2
3	4
5	6
7	8
9	10
11	12
13	14
15	16
17	18
19	20
21	22
23	24
25	26
27	28
29	30
31	32
33	

0. Preliminaries

- ◆ $\mathcal{M}(\mathbb{R}^d)$ Banach space of (signed) σ -additive Borel measures on $\mathbb{R}^d$ with **total variation norm**

$$\|\mu\|_{TV} := |\mu|(\mathbb{R}^d) \quad \text{where} \quad |\mu|(A) := \sup_{A=\bigcup_k A_k, A_i \cap A_j = \emptyset} \sum_{k=1}^n |\mu(A_k)|$$

- ◆ $\mathcal{M}(\mathbb{R}^d)$ is dual space of $C_0(\mathbb{R}^d)$, i.e. $C_0(\mathbb{R}^d)' = \mathcal{M}(\mathbb{R}^d)$

$$\|\mu\|_{TV} = \sup_{\|\varphi\|_\infty \leq 1} |\langle \varphi, \mu \rangle|, \quad \langle x, \mu \rangle := \int_{\mathbb{R}^d} \varphi(x) \, d\mu(x)$$

and $(\mu_n)_n$ converges weak-* to a measure μ if

$$\lim_{n \rightarrow \infty} \int \varphi \, d\mu_n = \int \varphi \, d\mu \quad \text{for all} \quad \varphi \in C_0(\mathbb{R}^d)$$

- ◆ $\mathcal{P}(\mathbb{R}^d)$ set of probability measures on $\mathbb{R}^d$
 - probability measures $(\mu_n)_n$ do in general not weak-* converge a probability measure
e.g. $\mu_n = \delta_n$ converges weak-* to $\mu \equiv 0$
 - probability measures $(\mu_n)_n$ **converges narrowly/weakly** to a (probability) measure μ if

$$\lim_{n \rightarrow \infty} \int \varphi \, d\mu_n = \int \varphi \, d\mu \quad \text{for all} \quad \varphi \in C_b(\mathbb{R}^d)$$

- TV-distance $\|\mu - \nu\|$ between $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ with disjoint support is always 2
 $\Rightarrow$ **need another distance!**

1	2
3	4
5	6
7	8
9	10
11	12
13	14
15	16
17	18
19	20
21	22
23	24
25	26
27	28
29	30
31	32
33	

Example

1. **Atomic measures:** Let δ_x , $x \in \mathbb{R}^d$ be defined by

$$\delta_x(A) := \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\mu := \sum_{k=1}^n \mu_k \delta_{x_k}, \quad \|\mu\|_{TV} = \sum_{k=1}^n |\mu_k| = \sum_{k=1}^n \mu_k = 1$$



Is the 'cross' measure or the 'triangle' measure nearer to the 'dot' measure?

2. **Absolutely continuous measures:** with respect to the Lebesgue measure $\mu \ll \lambda$, i.e. $\lambda(A) = 0 \Rightarrow \mu(A) = 0$ for all $A \in \mathcal{B}(\mathbb{R}^d)$.

Then the Radon-Nikodym derivative $p \in L_1(\mathbb{R}^d)$ exists, $d\mu = p \, dx$ and

$$\|\mu\|_{TV} = \int_{\mathbb{R}^d} |p_\mu(x)| \, dx, \quad |\mu|(A) = \int_A p_\mu(x) \, dx$$

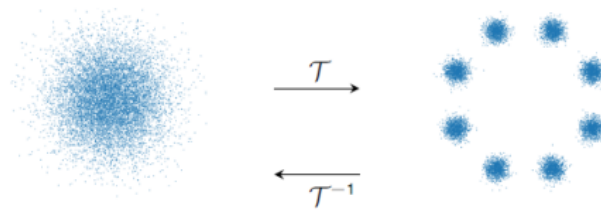
1	2
3	4
5	6
7	8
9	10
11	12
13	14
15	16
17	18
19	20
21	22
23	24
25	26
27	28
29	30
31	32
33	

- ◆ Borel-measurable map $T : \mathbb{R}^d \rightarrow \mathbb{R}^n$, $\mu \in \mathcal{P}(\mathbb{R}^d)$

The **push-forward measure** of μ by T is the measure $\nu \in \mathcal{P}(\mathbb{R}^n)$ defined by

$$\nu = T_{\#}\mu = \mu \circ T^{-1}$$

with the preimage $T^{-1}(A)$ of $A \in \mathcal{B}(\mathbb{R}^n)$



- ◆ $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is integrable with respect to $T_{\#}\mu$ if $f \circ T$ is integrable with respect to μ and

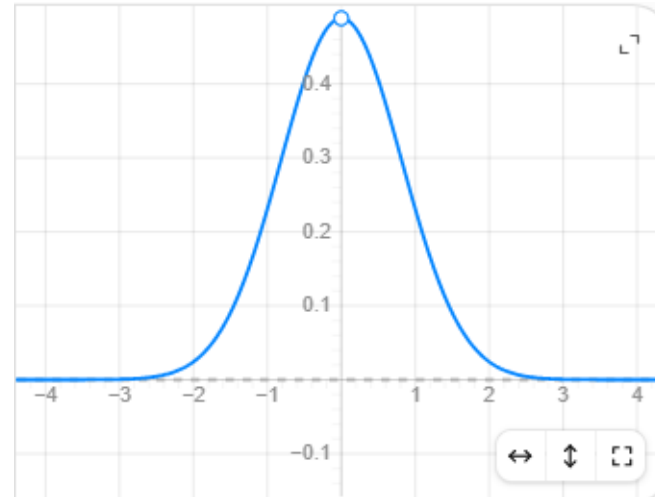
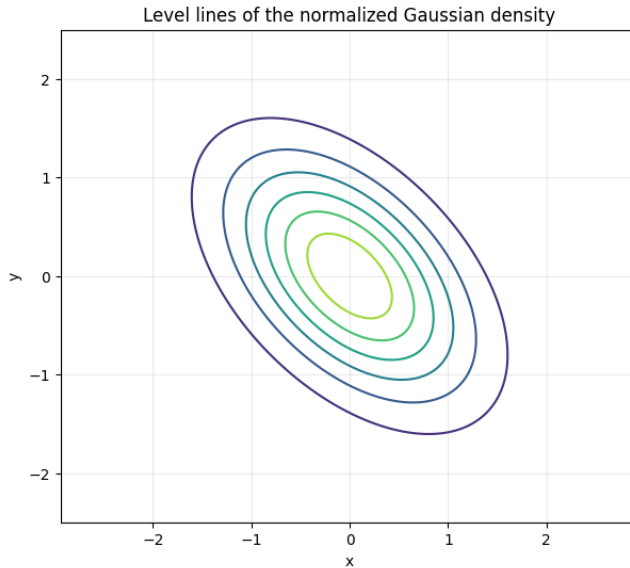
$$\int_{\mathbb{R}^n} f(y) \, d(T_{\#}\mu)(y) = \int_{T^{-1}(\mathbb{R}^n)} f(T(x)) \, d\mu(x)$$

- ◆ if $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a C^1 diffeomorphism and μ has density p , then

$$\begin{aligned} \int f(y) \, d(T_{\#}\mu)(y) &= \int f(T(x))p(x) \, dx \\ &= \int f(y) \underbrace{p(T^{-1}(y))|\det(\nabla T^{-1}(y))|}_{p_{T_{\#}\mu}} \, dy \end{aligned}$$

1	2
3	4
5	6
7	8
9	10
11	12
13	14
15	16
17	18
19	20
21	22
23	24
25	26
27	28
29	30
31	32
33	

Example



Let

$$p(x) = \frac{\sqrt{3}}{2\pi} \exp(x^T A x / 2), \quad A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

and

$$T(x) = \langle x, \xi \rangle, \quad \xi = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Then

$$T_{\#}p \sim \mathcal{N}(0, \frac{2}{3})$$

1 2

3 4

5 6

7 8

9 10

11 12

13 14

15 16

17 18

19 20

21 22

23 24

25 26

27 28

29 30

31 32

33

- ◆ $(\Omega, \mathcal{A}, \mathbb{P})$ probability space, $X : \Omega \rightarrow \mathbb{R}^d$ random variable

law of X : $P_X = X_{\#}\mathbb{P}$

- ◆ Joint distribution

$$\pi = P_{(X,Y)} = P_{X|Y=y} \times_y P_X = P_{Y|X=x} \times_x P_X$$

with **disintegration**

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} f(x, y) \, dP_{(X,Y)}(x, y) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(x, y) \, dP_{X|Y=y}(x) \, dP_Y(y).$$

- ◆ Addition of **independent** random variables

$$X_t = (1 - t)X_0 + tX_1, \quad t \in (0, 1)$$

having laws with densities gives

$$p_{X_t} = \frac{1}{1-t} p_{X_0} * \frac{1}{t} p_{X_1}$$

1	2
3	4
5	6
7	8
9	10
11	12
13	14
15	16
17	18
19	20
21	22
23	24
25	26
27	28
29	30
31	32
33	

0. Preliminaries

I. Flow Matching for Continuous Distributions

I.1 Motivation

I.2 Flow Matching

I.3 Numerical Examples

ADD Diffusion Models

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

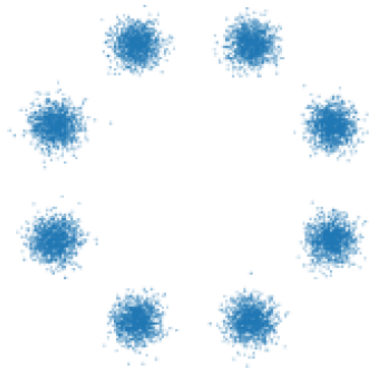
32

33

I.1. Motivation

Aim: Sampling from a probability distribution μ_{target} (**target distribution**) given:

- ◆ some samples (oft many = data hungry)
- ◆ probability density function $p_X(x) = \exp(-\psi(x))/C$ (Boltzmann density) with or without normalizing factor C



$$d = 2$$



$$d = 28^2 = 784$$

It is only easy to sample from μ_{latent} (**latent distribution**)

- ◆ 1d distributions (cdf)
- ◆ $d \gg 1$ uniform or Gaussian "friendly" distributions

Refs: 1. J. Chemseddine, C. Wald, R. Duong, St: Neural sampling from Boltzmann densities: Fisher-Rao curves in the Wasserstein geometry, ICLR 2025

2. Ch. Duan, Y. Jiao, St, Ch. Wald, J. Zh. Yang, and R. Zhang: Sampling via stochastic interpolants by Langevin-based velocity and initialization estimation in flow ODEs, ArXiv 2026

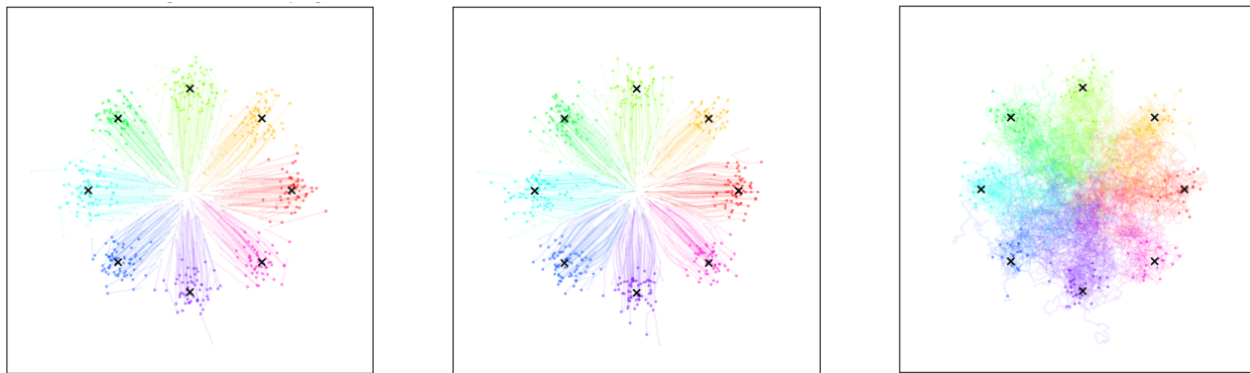
3. Ch. Duan, J. Hertrich, St: Sampling via Wasserstein-Fisher-Rao JKO Scheme to appear

1	2
3	4
5	6
7	8
9	10
11	12
13	14
15	16
17	18
19	20
21	22
23	24
25	26
27	28
29	30
31	32
33	

Idea: Learn a curve $\mu_t : [0, 1] \rightarrow \mathcal{P}_2(\mathbb{R}^d)$ from μ_{latent} to μ_{target}

Basic methods:

- ◆ flow matching **FM**, stochastic interpolants (Lipman et al. 2022/23, Liu et al. 2022/23, Albergo/vanEijnden 2023)
- ◆ score-based diffusion models (Sohl-Dickstein 2015; Song/Ermon et al. 2019)



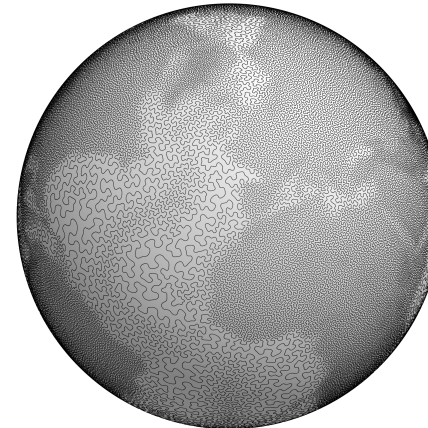
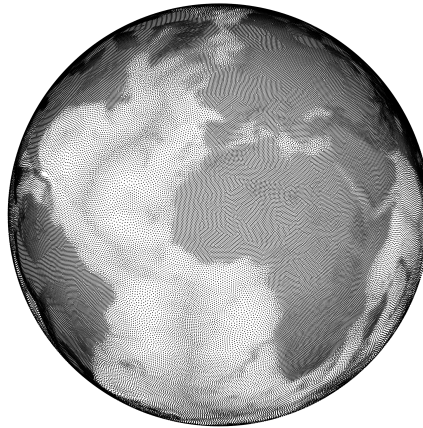
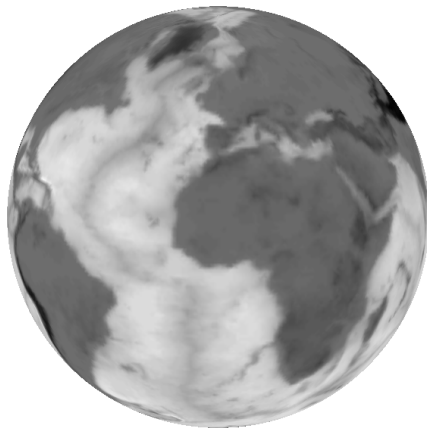
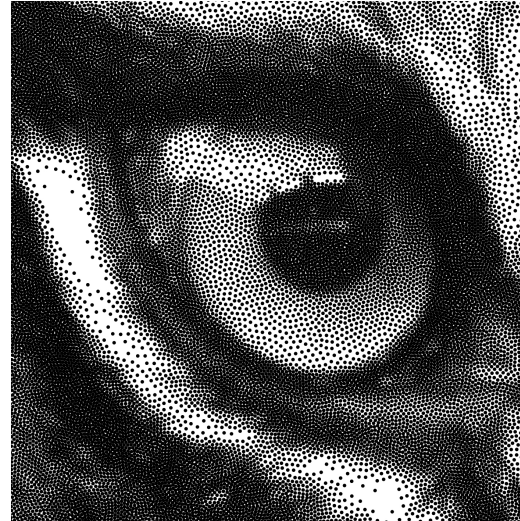
Trajectories of flow matching from OT and independent couplings and diffusion, $d = 2$



Single trajectory of flow matching model trained on cat images, $d = 256^2$

1	2
3	4
5	6
7	8
9	10
11	12
13	14
15	16
17	18
19	20
21	22
23	24
25	26
27	28
29	30
31	32
33	

Work on Generative Models related to Chemnitz



1	2
3	4
5	6
7	8
9	10
11	12
13	14
15	16
17	18
19	20
21	22
23	24
25	26
27	28
29	30
31	32
33	

Refs: 1. Teuber, St., Gwosdek, Schmaltz, Weickert: Dithering by diff. of convex functions, SIIMS 2011,
2. Gräf, Potts, St.: SISC 2012,
3. Ehler, Gräf, Neumayer, St., FoCM 2021

0. Preliminaries

I. Flow Matching for Continuous Distributions

I.1 Motivation

I.2 Flow Matching

I.3 Numerical Examples

ADD Diffusion Models

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

I.2 Absolutely Continuous Curves

- ◆ A function $\gamma : [a, b] = I \rightarrow \mathbb{R}$ is **absolutely continuous** iff there exists $g \in L_1(I)$ such that

$$\gamma(t) = \gamma(s) + \int_s^t g(r) \, dr$$

Such function γ is a.e. differentiable has the **slope** (smallest g)

$$|\gamma'| (t) := \lim_{h \rightarrow 0} \frac{|\gamma(t+h) - \gamma(t)|}{|h|}$$

- ◆ A curve $\gamma = (\gamma_k)_{k=1}^d : I \rightarrow \mathbb{R}^d$ is **absolutely continuous** if each component $\gamma_k : I \rightarrow \mathbb{R}$ is



- ◆ A curve $\gamma : I \rightarrow X$ in a complete **metric space** (X, d) is **absolutely continuous**, iff there exists $g \in L_1(I)$ such that

$$d(\gamma(t), \gamma(s)) \leq \int_s^t g(r) \, dr \quad \text{for } s \leq t, \, s, t \in I.$$

Then there exists a.e. the **metric derivative** (speed)

$$|\gamma'| (t) := \lim_{h \rightarrow 0} \frac{d(\gamma(t+h), \gamma(t))}{|h|}.$$

1 2

3 4

5 6

7 8

9 10

11 12

13 14

15 16

17 18

19 20

21 22

23 24

25 26

27 28

29 30

31 32

33

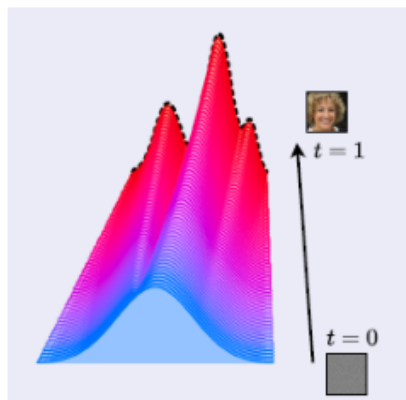
We are interested in the metric space $(X, d) = (\mathcal{P}_2(\mathbb{R}^d), W_2)$.

$$\mathcal{P}_2(\mathbb{R}^d) := \left\{ \mu \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} |x|^2 d\mu(x) < \infty \right\}$$

Wasserstein-2-distance is given by

$$W_2^2(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\pi(x, y)$$

Task: Learn a curve $\mu_t : I \rightarrow \mathcal{P}_2(\mathbb{R}^d)$ from μ_{latent} to μ_{target}



Courtesy: S. Martin

Refs: 1. Ambrosio, Savaré, Gigli. Gradient Flows: In Metric Spaces and in the Space of Probability Measures, 2005

2. Jordan, Kinderlehrer, Otto: The variational formulation of the Fokker-Planck equation, 1998

1	2
3	4
5	6
7	8
9	10
11	12
13	14
15	16
17	18
19	20
21	22
23	24
25	26
27	28
29	30
31	32
33	

Eulerian view: A narrowly continuous curve $\mu_t : [0, 1] \rightarrow \mathcal{P}_2(\mathbb{R}^d)$ is absolutely continuous, if there exists a Borel measurable vector field $v : I \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ with $\|v_t\|_{L_2(\mu_t)} \in L_1([0, 1])$ such that $(\mu_t, v_t)_t$ satisfy the **continuity equation**

$$\partial_t \mu_t + \operatorname{div}(\mu_t v_t) = 0$$

Lagrangian view: For sufficiently smooth velocity fields, there exists a unique solution $\Phi : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ of the **ODE**

$$\partial_t \Phi(t, x) = v_t(\Phi(t, x)), \quad \Phi(0, x) = x$$

and $\Phi(t, \cdot)_\# \mu_0 = \mu_0 \circ \Phi(t, \cdot)^{-1} = \mu_t$.

All you need is v_t !

- ◆ **Learning:** Flow Matching learns such a velocity field v_t^θ by a neural network
- ◆ **Sampling/Inference:** Solve ODE with v_t^θ by your favorite numerical solver starting in a sample $x \sim \mu_{\text{latent}} = \mu_0$

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

Intensity preserving flow: $I_0(x) = I_t(\Phi_t(x)) = I(t, \Phi_t(x))$

Differentiation $\frac{d}{dt}$ gives the **transport equation**

$$\begin{aligned} 0 &= \frac{d}{dt} I_t(\Phi_t(x)) = \partial_t I(t, \Phi_t(x)) + \nabla_x I(t, \Phi_t(x)) \partial_t \Phi_t(x) \\ &= \partial_t I(t, x) + \nabla I(t, x) \cdot v(t, x) \end{aligned}$$

Measure preserving flow:

$$p_t(y) = p_0(\Phi_t^{-1}(y)) / \det(\Phi_t(\Phi_t^{-1}(y))) \rightarrow p_0(x) = p_t(\Phi_t(x)) \det \nabla \Phi_t(x)$$

Differentiation $\frac{d}{dt}$ gives the **continuity equation**

$$\begin{aligned} 0 &= \frac{d}{dt} \left(p_t(\Phi_t(x)) J(t, x) \right), \quad J(t, x) := \det \nabla \Phi_t(x), \\ 0 &= \frac{d}{dt} \left(p_t(\Phi_t(x)) \right) J(t, x) + p_t(\Phi_t(x)) \frac{d}{dt} J(t, x) \\ &= \left(\partial_t p_t(\Phi_t(x)) + \nabla p_t(\Phi_t(x)) \cdot \partial_t \Phi_t(x) \right) J(t, x) + p_t(\Phi_t(x)) J(t, x) \operatorname{div} v_t(\Phi_t(x)) \\ &= \partial_t p_t(\Phi_t(x)) + \nabla p_t(\Phi_t(x)) \cdot v_t(\Phi_t(x)) + p_t(\Phi_t(x)) \operatorname{div} v_t(\Phi_t(x)) \\ &= \partial_t p_t(\Phi_t(x)) + \operatorname{div}(p_t v_t)(\Phi_t(x)) \\ &= \partial_t p_t + \operatorname{div}(p_t v_t) \end{aligned}$$

1	2
3	4
5	6
7	8
9	10
11	12
13	14
15	16
17	18
19	20
21	22
23	24
25	26
27	28
29	30
31	32
33	

How to Learn a Velocity Field?

Task: Learn/approximate v_t by neural network v_t^θ !

Idea 1: Loss function for learning:

$$\mathcal{L}(\theta) := \mathbb{E}_{t \sim \mathcal{U}(0, T), x \sim \mu_t} \left[\|v_t^\theta(x) - v_t(x)\|^2 \right],$$

not accessible 😞.

Idea 2: Make it **conditional**, i.e. end in δ_w , where $w \sim \mu_{\text{target}} = \mu_1$, since up to a constant

$$\mathcal{L}(\theta) = \mathbb{E}_{t \sim \mathcal{U}(0, T), w \sim \mu_1, x \sim \mu_t(\cdot | w)} \left[\|v_t^\theta(x) - v_t(x | w)\|^2 \right]$$

All you need is conditional $v_t(\cdot | w)$ 😊

1	2
3	4
5	6
7	8
9	10
11	12
13	14
15	16
17	18
19	20
21	22
23	24
25	26
27	28
29	30
31	32
33	

Linear Interpolation: $X_0 \sim \mu_0, X_1 \sim \mu_1, X_t = \mu_t$

$$X_t = (1 - t)X_0 + tX_1, \quad t \in [0, 1]$$

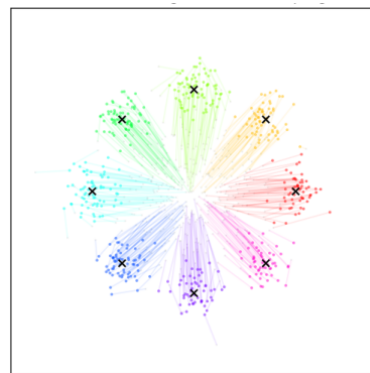
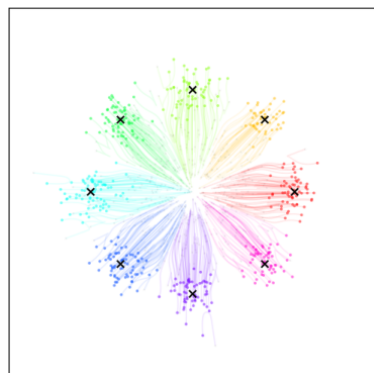
$$X_t \sim \mu_t := e_{t,\#}\pi, \quad \pi := P_{(X_1, X_0)} \in \Pi(\mu_0, \mu_1), \quad e_t(x, y) := (1 - t)x + tw$$

$$(X_t | X_1 = w) = (1 - t)X_0 + tw \quad (\text{conditional})$$

Theorem Let $(X_t | X_1 = w) \sim \mu_t(\cdot | w)$. Then $(\mu_t(\cdot | w), v_t(\cdot | w))$ fulfill continuity equation with

$$v_t(x | w) = \frac{w - x}{1 - t}.$$

Loss function 😊 $\mathcal{L}(\theta) := \mathbb{E}_{t \sim \mathcal{U}(0, T), w \sim \mu_1, x \sim \mu_t(\cdot | w)} \left[\left\| v_t^\theta(x) - \frac{w - x}{t - 1} \right\|^2 \right]$



Trajectories of flow matching from independent and OT couplings, $d = 2$

1 2

3 4

5 6

7 8

9 10

11 12

13 14

15 16

17 18

19 20

21 22

23 24

25 26

27 28

29 30

31 32

33

This becomes interesting later:

Theorem (Averaging) The curve-velocity $(\mu_t, v_t)_t$ with

$$\mu_t := P_{\#}^1 \pi_t = P_{\#}^1 (\mu_t(\cdot|w) \times_w \mu_1), \quad p_t(x) = \int p_t(x|w) \, d\mu_1(w)$$

$$v_t(x) := \int_{\mathbb{R}^d} v_t(x|w) \, dP_{X_1|X_t=x}(w),$$

satisfy the continuity equation.

1 2

3 4

5 6

7 8

9 10

11 12

13 14

15 16

17 18

19 20

21 22

23 24

25 26

27 28

29 30

31 32

33

Example:

$$X_t = (1 - t)X_0 + tX_1 \quad \in [0, 1]$$

$$p_0 = \mathcal{N}(0, I_d)$$

Let X_0, X_1 be independent. Then

$$p_{X_t}(x) = p_t(x) = C_t \int_{\mathbb{R}^d} e^{-\frac{\|x-tw\|^2}{2(1-t)^2}} d\mu_1(w) \quad \text{with} \quad C_t := (2\pi(1-t)^2)^{-\frac{d}{2}}$$

$$v_t(x) = \frac{1-t}{t} \underbrace{\nabla \log p_t}_{\text{score}} + \frac{x}{t}$$

and thus for $\mu_1 = \delta_w$,

$$p_{X_t|X_1=w}(x) = p_t(x|w) = C_t e^{-\frac{\|x-tw\|^2}{2(1-t)^2}}$$

$$v_t(x|w) = \frac{1-t}{t} \cdot \frac{tw-x}{(1-t)^2} + \frac{x}{t} = \frac{w-x}{1-t}$$

1	2
3	4
5	6
7	8
9	10
11	12
13	14
15	16
17	18
19	20
21	22
23	24
25	26
27	28
29	30
31	32
33	

Equivalence between Loss Functions

Technische Universität Berlin	
1	2
3	4
5	6
7	8
9	10
11	12
13	14
15	16
17	18
19	20
21	22
23	24
25	26
27	28
29	30
31	32
33	

Proof (Equivalence of loss functions): We use

$$\mu_t = P_{\#}^1(\mu_t(\cdot|w) \times_w \mu_1), \quad v_t(x) := \int_{\mathbb{R}^d} v_t(x|w) \, dP_{X_1|X_t=x}(w)$$

Then

$$\mathcal{L}(\theta) := \mathbb{E}_{t \sim \mathcal{U}(0,T), x \sim \mu_t} \left[\left\| v_t^\theta(x) - v_t(x) \right\|^2 \right] = \mathbb{E} \left[\|v_t^\theta\|^2 \right] - 2\mathbb{E} \left[\langle v_t, v_t^\theta \rangle \right] + \underbrace{\mathbb{E} \left[\|v_t\|^2 \right]}_{\text{const}}$$

Now we have for the first and second term

$$\begin{aligned} \mathbb{E}_{t \sim \mathcal{U}(0,T), x \sim \mu_t} \left[\|v_t^\theta(x)\|^2 \right] &= \int_0^1 \int_{\mathbb{R}^d} \|v_t^\theta\|^2 \, d\mu_t \, dt \\ &= \int_I \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \|v_t^\theta(x)\|^2 \, dP_{X_t|X_1=w}(x) \, dP_{X_1}(w) \, dt \\ &= \mathbb{E}_{t \sim \mathcal{U}(0,T), w \sim \mu_1, x \sim \mu_t(\cdot|w)} \left[\|v_t^\theta(x)\|^2 \right] \end{aligned}$$

$$\begin{aligned} \mathbb{E}_{t \sim \mathcal{U}(0,T), x \sim \mu_t} \left[\langle v_t^\theta(x), v_t(x) \rangle \right] &= \int_I \int_{\mathbb{R}^d} \left\langle v_t^\theta(x), \int_{\mathbb{R}^d} v_t(x|w) \, dP_{X_1|X_t=x}(w) \right\rangle \, dP_{X_t}(x) \, dt \\ &= \int_I \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle v_t^\theta(x), v_t(x|w) \rangle \, dP_{(X_t, X_1)}(x, w) \, dt \\ &= \mathbb{E}_{t \sim \mathcal{U}(0,T), w \sim \mu_1, x \sim \mu_t(\cdot|w)} \left[\langle v_t^\theta(x), v_t(x|w) \rangle \right] \end{aligned}$$

0. Preliminaries

I. Flow Matching for Continuous Distributions

I.1 Motivation

I.2 Flow Matching

I.3 Numerical Examples

ADD Diffusion Models

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

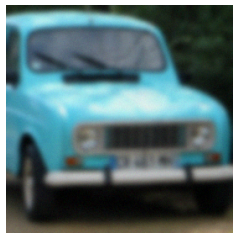
31

32

33

Inverse problems:

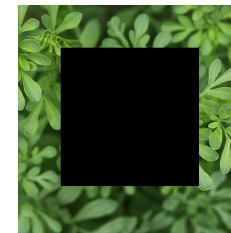
$$y = Ax + \eta, \quad \eta \sim \mathcal{N}(0, \sigma^2 I_d)$$



Deblurring



Superresolution



Inpainting

- ◆ sampling from the posterior $p(x|y)$ knowing some data pairs $(x_i, y_i)_i$
- ◆ learning the regularizer in variational formulations

$$p(x|y) \propto \underbrace{p(y|x)}_{\text{likelihood}} \underbrace{p(x)}_{\text{prior}}$$

knowing samples from the prior distribution and the forward operator A .

1. DeepInverse: A Python package for solving imaging inverse problems with deep learning, Tachella et al. 2025
2. Learning regularization functionals for inverse problems: A comparative study Hertrich et al. 2025
3. A Stability Benchmark of Generative Regularizers for Inverse Problems Denker, Hertrich, Neumayer 2026

1	2
3	4
5	6
7	8
9	10
11	12
13	14
15	16
17	18
19	20
21	22
23	24
25	26
27	28
29	30
31	32
33	

I.3 Numerical Results



Denoising, Deblurring, Superresolution, Inpainting

1. Chemseddine, Hagemann, St., C. Wald, Conditional Wasserstein distances with applications in Bayesian OT flow matching, JMLR, 2026
2. Martin, Gagneux, Hagemann, St. PnP-Flow: Plug-and-play image restoration with flow matching, ICLR, 2025

1	2
3	4
5	6
7	8
9	10
11	12
13	14
15	16
17	18
19	20
21	22
23	24
25	26
27	28
29	30
31	32
33	

0. Preliminaries

I. Flow Matching for Continuous Distributions

I.1 Motivation

I.2 Flow Matching

I.3 Numerical Examples

ADD Diffusion Models

1 2

3 4

5 6

7 8

9 10

11 12

13 14

15 16

17 18

19 20

21 22

23 24

25 26

27 28

29 30

31 32

33

Forward SDE computes $(X_t)_{t \in [0, T]}$ by

$$dX_t = f(t, X_t) dt + g(t) dW_t, \quad X_0 \sim P_{\text{data}},$$

where W_t is a standard Brownian motion.

Example 1: $f(t, X_t) := -\frac{1}{2}\beta_t X_t$ and $g(t) := \sqrt{\beta_t}$, e.g.

$$\beta_t = \beta_{\min} + t(\beta_{\max} - \beta_{\min}), \quad \beta_{\max} = 20, \beta_{\min} = 0.1$$

with increasing "time schedule" $\beta_t > 0$. Then SDE becomes linear

$$dX_t = -\frac{1}{2}\beta_t X_t dt + \sqrt{\beta_t} dW_t,$$

$$X_t = b_t X_0 + \sqrt{1 - b_t^2} Z, \quad b_t = e^{\frac{-h(t)}{2}}, \quad h(t) := \int_0^t \beta_s ds, \quad Z = X_\infty \sim \mathcal{N}(0, I_d)$$

Approximation $X_1 \approx Z$.

Ornstein-Uhlenbeck process (simplest setting): for $\beta_t = 2$

$$dX_t = -X_t dt + \sqrt{2} dW_t$$

$$X_t = X_0 e^{-t} + \sqrt{1 - e^{-2t}} Z, \quad Z \sim \mathcal{N}(0, I_d)$$

$$p_t = p_0(e^t \cdot) * \mathcal{N}(0, (1 - e^{-2t}) I_d)$$

1	2
3	4
5	6
7	8
9	10
11	12
13	14
15	16
17	18
19	20
21	22
23	24
25	26
27	28
29	30
31	32
33	

Example 2: Fokker Planck Equation

Technische Universität Berlin	
1	2
3	4
5	6
7	8
9	10
11	12
13	14
15	16
17	18
19	20
21	22
23	24
25	26
27	28
29	30
31	32
33	

For the forward PDE

$$dX_t = f(t, X_t) dt + g(t) dW_t, \quad X_0 \sim \mu_0,$$

the corresponding densities $p_t = p_{X_t}$ fulfill the **Fokker-Planck equation** (CE)

$$\begin{aligned} \partial_t p_t(x) &= -\nabla \cdot (f(t, x) p_t(x)) + \frac{1}{2} g(t)^2 \Delta p_t(x) \\ &= -\nabla \cdot \left(\underbrace{\left(f(t, x) - \frac{1}{2} g^2(t) \nabla \log p_t(x) \right)}_{v_t} p_t(x) \right) \end{aligned}$$

Example 2.1: Linear case

$$f(t, x) = -\frac{1}{2} \beta_t x, \quad g(t) = \sqrt{\beta_t} \quad \Rightarrow \quad v_t(x) = -\frac{1}{2} \beta_t (x - \nabla \log p_t(x))$$

For $\beta_t = 1/(1-t)$, $g(t)^2 = 2t/(1-t)$ compare to FM with independent coupling starting in Gaussian at slice 20:

$$-v_{1-t}(x) = -\frac{x}{1-t} - \frac{1-t}{t} \nabla \log p_{1-t}(x)$$

Example 2: Fokker Planck Equation

Technische Universität Berlin	
1	2
3	4
5	6
7	8
9	10
11	12
13	14
15	16
17	18
19	20
21	22
23	24
25	26
27	28
29	30
31	32
33	

For the forward PDE

$$dX_t = f(t, X_t) dt + g(t) dW_t, \quad X_0 \sim \mu_0,$$

the corresponding densities $p_t = p_{X_t}$ fulfill the **Fokker-Planck equation** (CE)

$$\begin{aligned} \partial_t p_t(x) &= -\nabla \cdot (f(t, x) p_t(x)) + \frac{1}{2} g(t)^2 \Delta p_t(x) \\ &= -\nabla \cdot \left(\underbrace{\left(f(t, x) - \frac{1}{2} g(t)^2 \nabla_x \log p_t(x) \right)}_{v_t} p_t(x) \right) \end{aligned}$$

Example 2.2: Wasserstein gradient flow for minimizing $\mathcal{F} : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$

Choosing $v_t \in -\partial \mathcal{F}(\mu_t) = -\nabla \delta \mathcal{F}(\mu_t)$ ("Riemannian" gradient of $\mathcal{F}$ wrt W_2) leads to the (CE)

$$\partial_t p_t = \nabla \cdot (p_t \nabla \delta \mathcal{F}(\mu_t))$$

◆ $\mathcal{F}_\nu(\mu) = \text{KL}(\mu, \nu) = \int_{\mathbb{R}^d} \log p(x) p(x) dx - \int_{\mathbb{R}^d} \log q(x) p(x) dx$, $\mu = p\lambda$, $\nu = q\lambda$
Then

$$\delta \mathcal{F}(p_t) = 1 + \log p_t - \log q, \quad \nabla \delta \mathcal{F}(p_t) = \frac{1}{p_t} \nabla p_t - \log q$$

so that

$$\partial_t p_t = \Delta p_t - \nabla \cdot (p_t \nabla \log q)$$

which is the above **Fokker-Planck equation** for $f(t, x) = \nabla \log q(x)$ and $g(t) = \sqrt{2}$.

Under certain condition (log-convace q) it is known that $X_\infty \sim q$.

Example 2: Fokker Planck Equation

For $q(x) = C^{-1}e^{-\beta\Psi(x)}$ we obtain the

Overdamped Langevin SDE:

$$X(0) = X^0 \sim p^0$$

$$dX_t = - \underbrace{\nabla \Psi(X_t)}_{\nabla \log q} dt + \sqrt{2\beta^{-1}} dW_t$$

Euler-Maruyama forward step with step size η :

$$X_k := X_{k-1} - \eta \nabla \Psi(X_{k-1}) + \sqrt{2\beta^{-1}\eta} \xi_k, \quad \xi_k \sim \mathcal{N}(0, I_d)$$

Hint: Use substitution $s = \frac{t}{\beta}$, $\sqrt{\beta^{-1}}W_t = W_{t/\beta}$

1 2

3 4

5 6

7 8

9 10

11 12

13 14

15 16

17 18

19 20

21 22

23 24

25 26

27 28

29 30

31 32

33

Reverse SDE becomes

$$dY_t = \left(-f(1-t, Y_t) + g(1-t)^2 \underbrace{\nabla \log p_{X_{1-t}}(Y_t)}_{\text{score}} \right) dt + g(1-t) dW_t, \quad Y_0 \sim X_T \approx Z.$$

Approximation of score by a neural network s^θ :

$$\begin{aligned} \mathcal{L}(\theta) &:= \mathbb{E}_{t \sim \mathcal{L}_{[0,1]}} \mathbb{E}_{x \sim P_{X_t}} \left[\|s_t^\theta(x) - \nabla \log p_{X_t}(x)\|^2 \right] \\ &= \mathbb{E}_{t \sim \mathcal{L}_{[0,1]}, x_0 \sim X_0, x \sim P_{X_t|X_0=x_0}} \left[\|s_t^\theta(x) - \nabla \log p_{X_t|X_0=x_0}(x)\|^2 \right] + \text{const} \end{aligned}$$

Example: (continued) $dX_t = -\frac{1}{2}\beta_t X_t dt + \sqrt{\beta_t} dW_t$

$$P_{X_t|X_0=x_0} = \mathcal{N}(b_t x_0, (1 - b_t^2) I_d) \quad \text{with} \quad b_t = e^{-\frac{h(t)}{2}}$$

This implies

$$\nabla \log p_{X_t|X_0=x_0}(x) = \nabla \left(-\frac{1}{2(1-b_t^2)} \|x - b_t x_0\|^2 \right) = -\frac{x - b_t x_0}{1 - b_t^2}.$$

Plugging this into the loss function, we get

$$\mathcal{L}(\theta) = \mathbb{E}_{t \sim \mathcal{L}_{[0,1]}, x_0 \sim X_0, x \sim P_{X_t|X_0=x_0}} \left[\left\| s_t^\theta(x) + \frac{x - b_t x_0}{1 - b_t^2} \right\|^2 \right].$$

Disadvantage: explosion of score near the data

1	2
3	4
5	6
7	8
9	10
11	12
13	14
15	16
17	18
19	20
21	22
23	24
25	26
27	28
29	30
31	32
33	

Let $X_t \in L^2(\Omega, \mathbb{R}^d, \mathbb{P})$ be continuously differentiable in every $t \in [0, 1]$. Let

$$\mu_t := X_{t,\#}\mathbb{P} = P_{X_t}, \quad \text{and} \quad v_t := \mathbb{E}[\partial_t X_t | X_t = \cdot].$$

Theorem. Under some assumptions (μ_t, v_t) fulfills the continuity equation

Example: Let $X_t = (1 - t)X_0 + tX_1$ for **independent** random variables X_0, X_1 , where $X_0 \sim \mathcal{N}(0, I_d)$. Then we get with $\mathbb{E}[X_t | X_t] = X_t$ that

$$\begin{aligned} v_t(x) &= \mathbb{E}[\partial_t X_t | X_t = x] = \mathbb{E}[X_1 - X_0 | X_t = x] = \mathbb{E}\left[\frac{X_1}{1-t} - \frac{X_t}{1-t} \middle| X_t = x\right] \\ &= \frac{1}{1-t} \mathbb{E}[X_1 | X_t = x] - \frac{x}{t-1} \\ &= \frac{1-t}{t} \nabla_x \log p_t(x) + \frac{x}{t} \end{aligned}$$

since by **Tweedie's formula** (Miyasawa's)

$$\mathbb{E}[X_1 | X_t = x] = \frac{(1-t)^2}{t} \nabla_x \log p_t(x) + \frac{x}{t}$$

1	2
3	4
5	6
7	8
9	10
11	12
13	14
15	16
17	18
19	20
21	22
23	24
25	26
27	28
29	30
31	32
33	

Tweedie's Formula

Consider

$$y = x + \eta, \quad \eta \sim \mathcal{N}(0, \sigma^2 I)$$

The observation density is related to the prior via Bayes' Theorem:

$$p(y) = \int p(y|x)p(x) \, dx = \int g(y-x)p(x) \, dx, \quad g(z) := \frac{1}{(2\pi\sigma^2)^{d/2}} e^{-\|z\|^2/(2\sigma^2)}.$$

For a given observation y , the minimum mean squared error (MMSE) estimator is given as

$$E[X|Y=y] = \int x p(x|y) \, dx = \int x \frac{p(y|x)p(x)}{p(y)} \, dx.$$

The MMSE can be computed using Tweedie's formula

$$E[X|Y=y] = y + \sigma^2 \nabla \log p(y),$$

which can be seen directly by computing with $p(y, x) = p(y|x)p(x)$,

$$\nabla p(y) = \frac{1}{\sigma^2} \int (x-y)g(y-x)p(x) \, dx = \frac{1}{\sigma^2} \int (x-y)p(y, x) \, dx,$$

and then multiplying both sides by $\sigma^2/p(y)$ gives with $p(y, x)/p(y) = p(x|y)$,

$$\frac{\sigma^2 \nabla_y p(y)}{p(y)} = \int x p(x|y) \, dx - \int y p(x|y) \, dx = E[X|Y=y] - y.$$

1 2

3 4

5 6

7 8

9 10

11 12

13 14

15 16

17 18

19 20

21 22

23 24

25 26

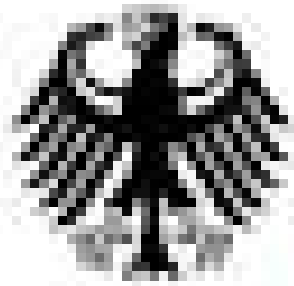
27 28

29 30

31 32

33

Gefördert durch:



Bundesministerium
für Forschung, Technologie
und Raumfahrt



Fellow

Ambassador

Funded under Germany's Excellence Strategy by
DFG Deutsche
Forschungsgemeinschaft

Applied and Numerical Harmonic Analysis



Gertlin Plonka
Daniel Potts
Gabriele Steidl
Manfred Tasche

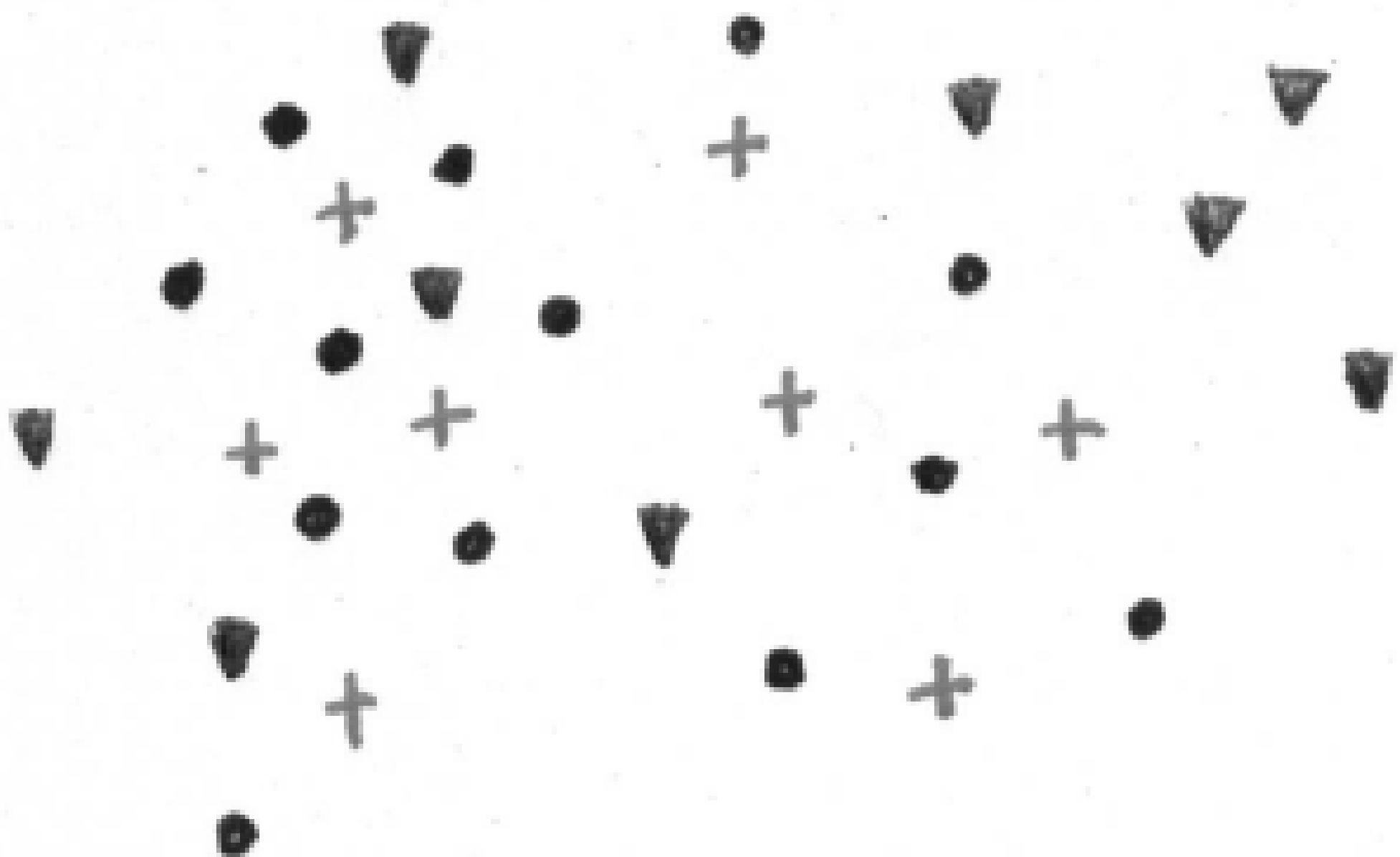
Numerical Fourier Analysis

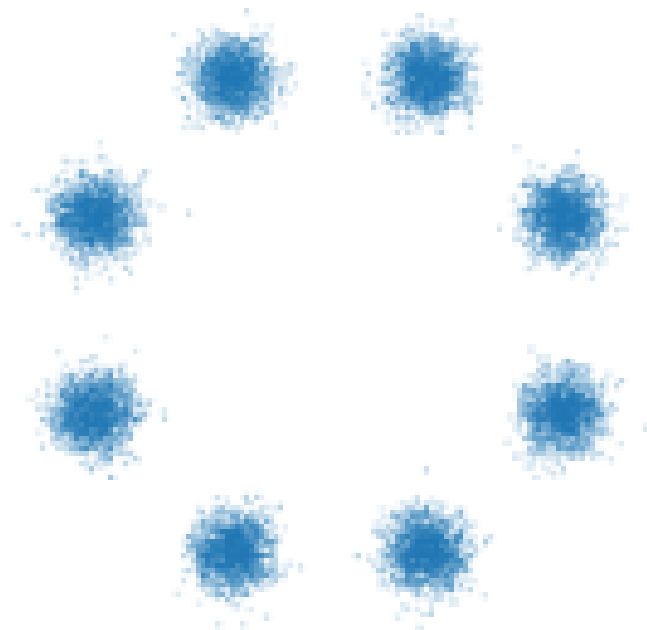
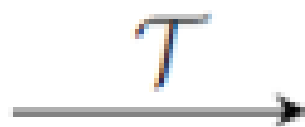
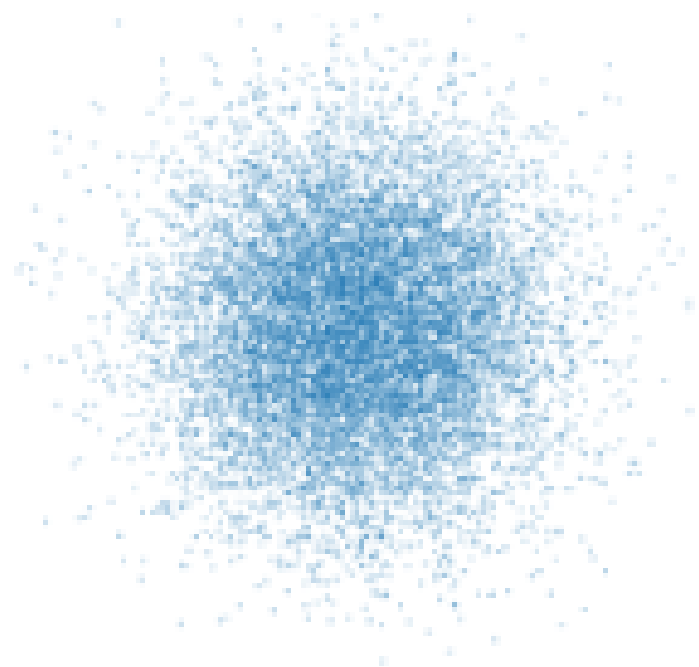
by Birkhäuser



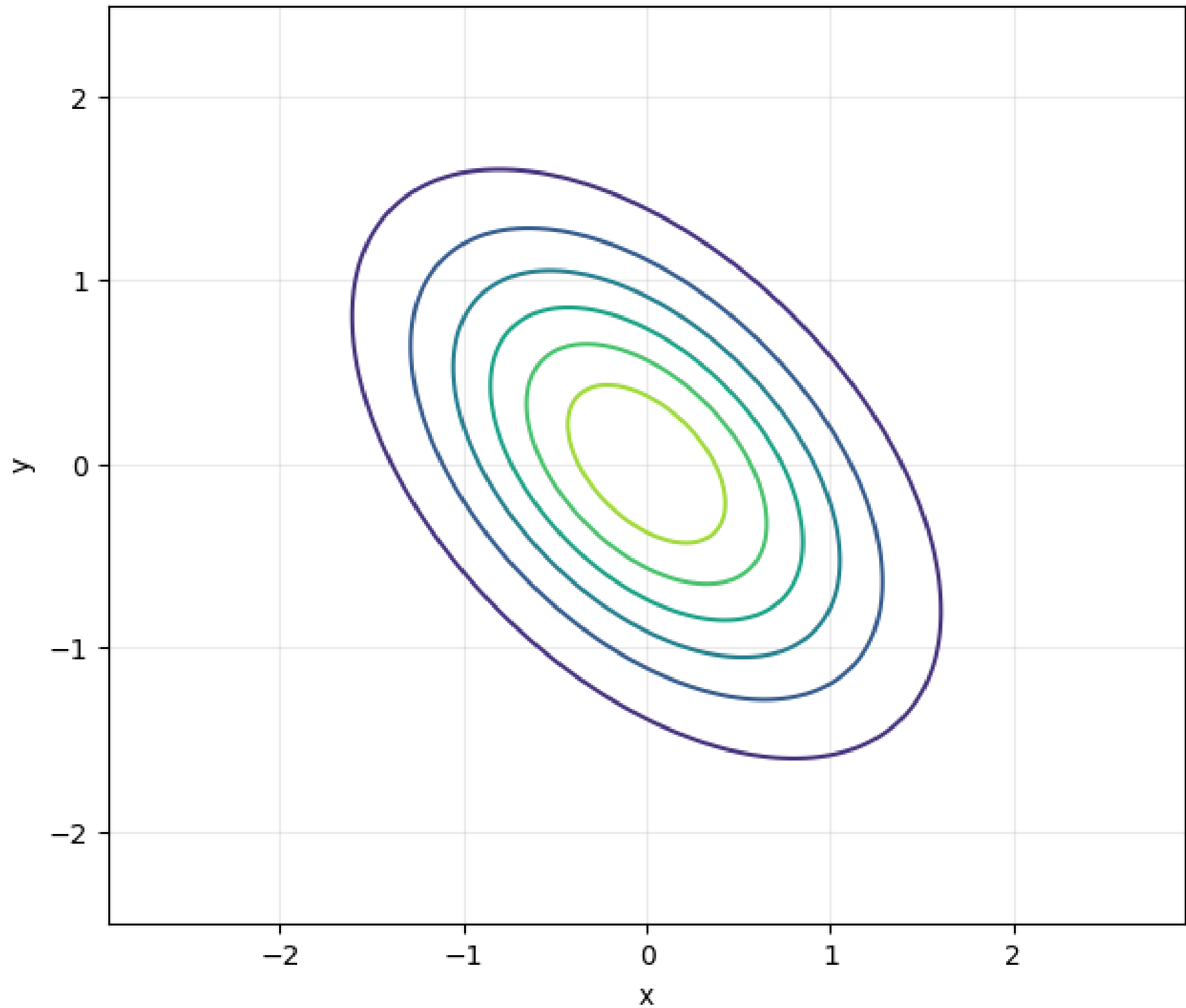
Variational and Information Flows in Machine Learning and Optimal Transport

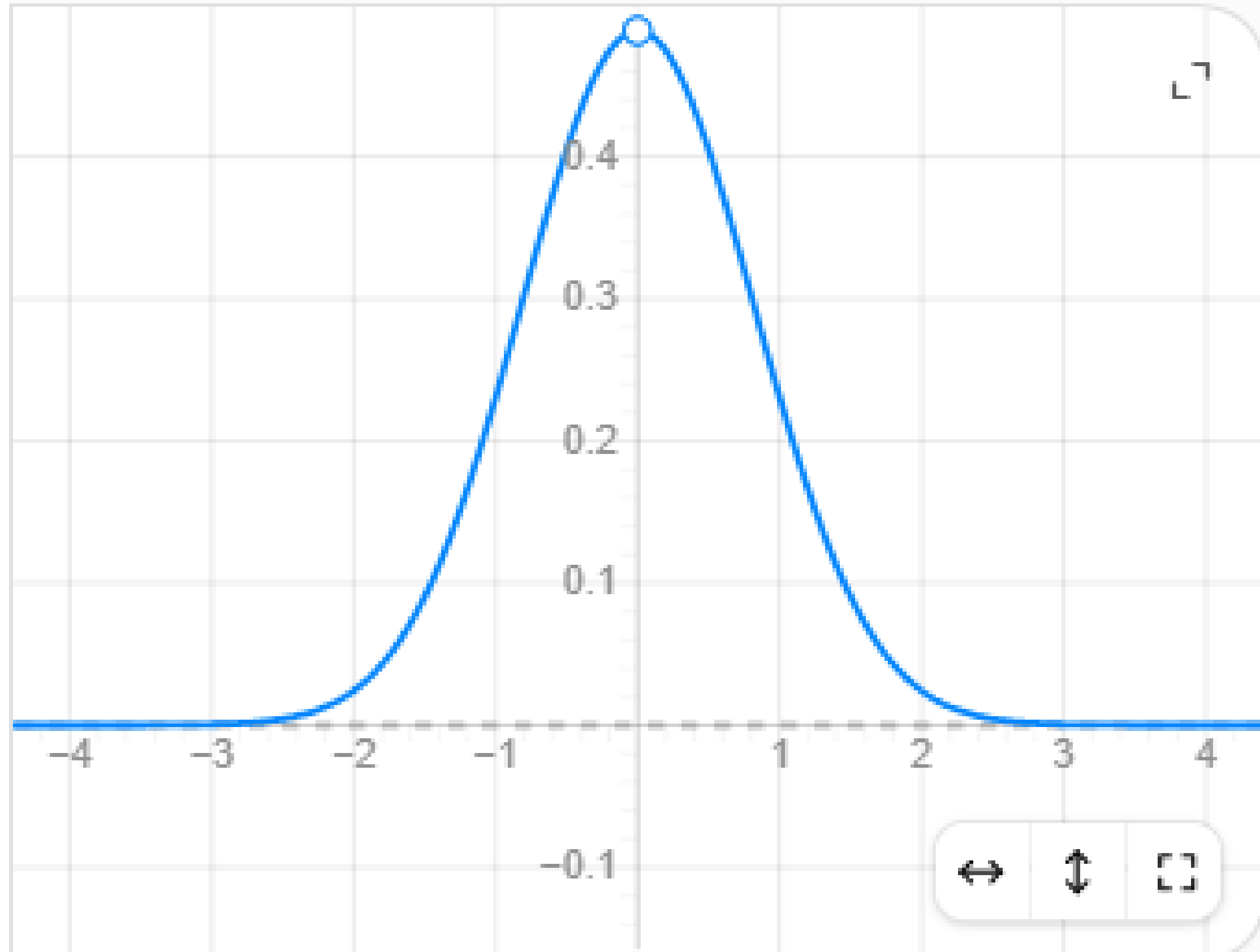
Wuchen Li
Bernhard Schmitzer
Gabriele Steidl
François-Xavier Vialard
Christian Wald

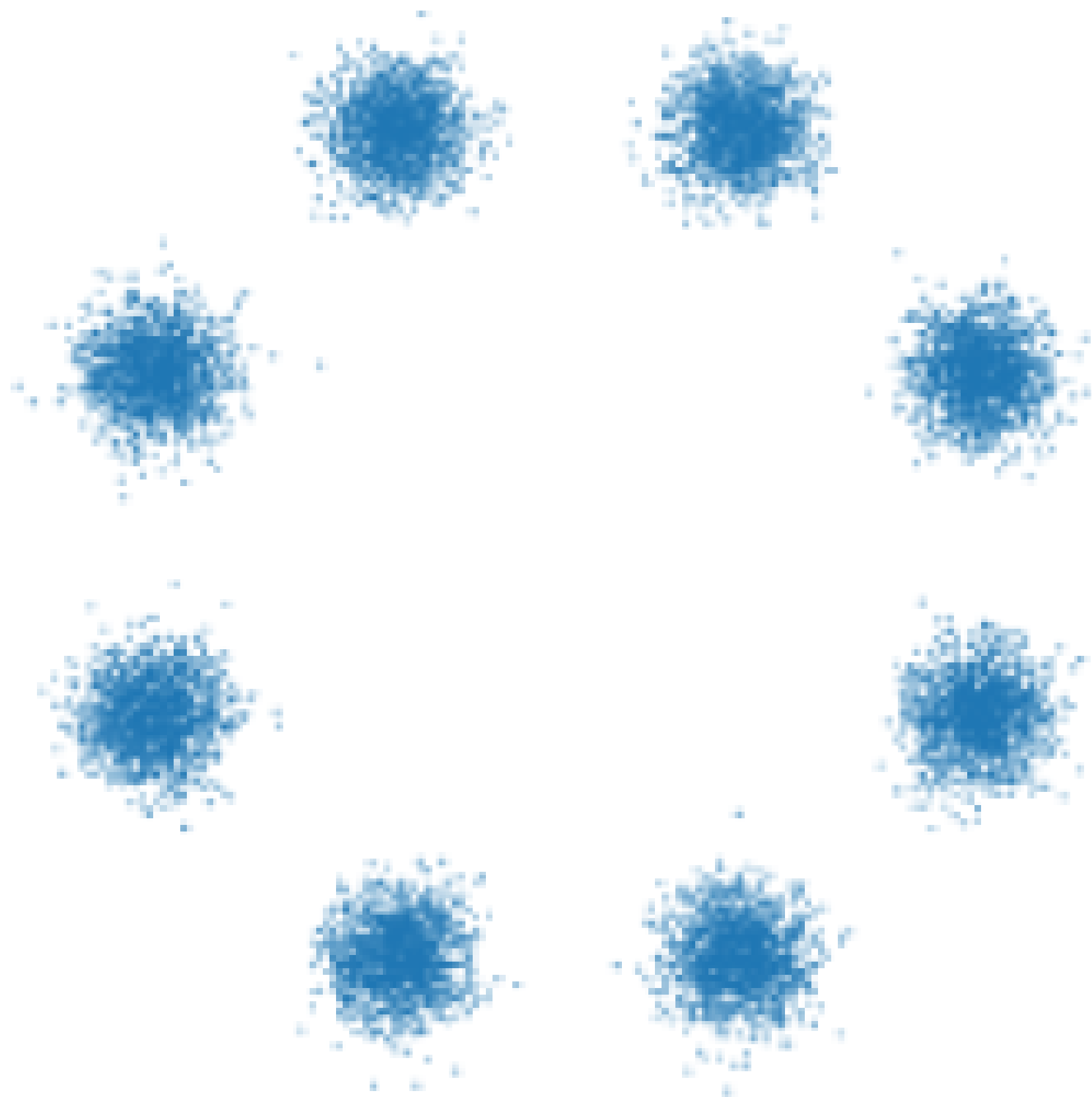




Level lines of the normalized Gaussian density





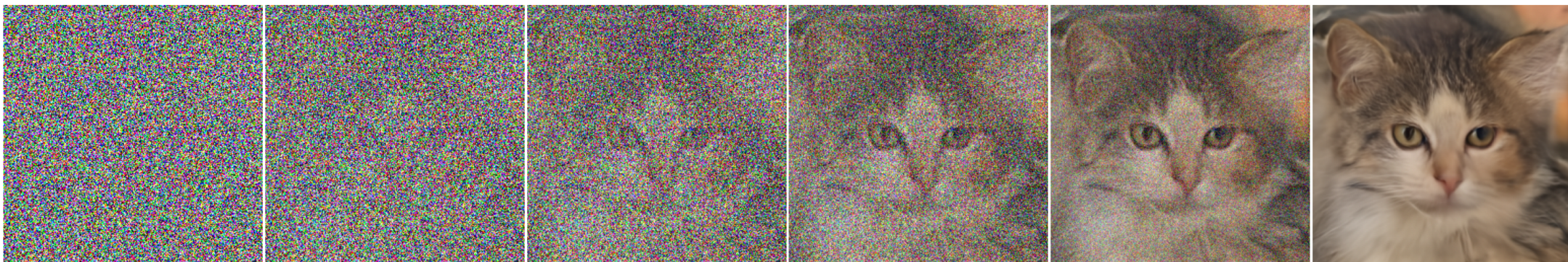


3	4	2	1	9	5	6	2	1	8
8	9	1	2	5	0	0	6	6	4
6	7	0	1	6	3	6	3	7	0
3	7	7	9	4	6	6	1	8	2
2	9	3	4	3	9	8	7	2	5
1	5	9	8	3	6	5	7	2	3
9	3	1	9	1	5	8	0	8	4
5	6	2	6	8	5	8	8	9	9
3	7	7	0	9	4	8	5	4	3
7	9	6	4	7	0	6	9	2	3

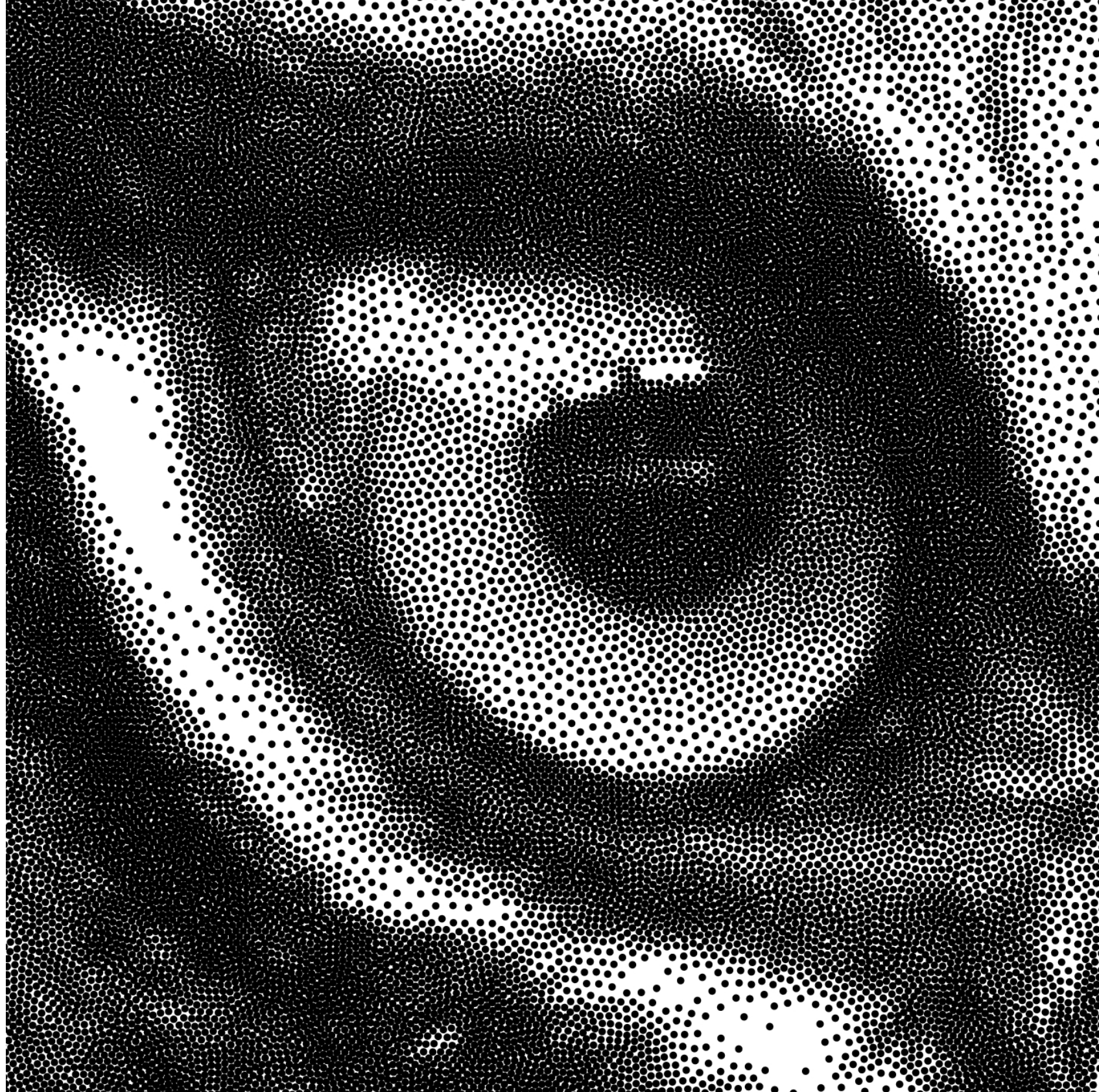


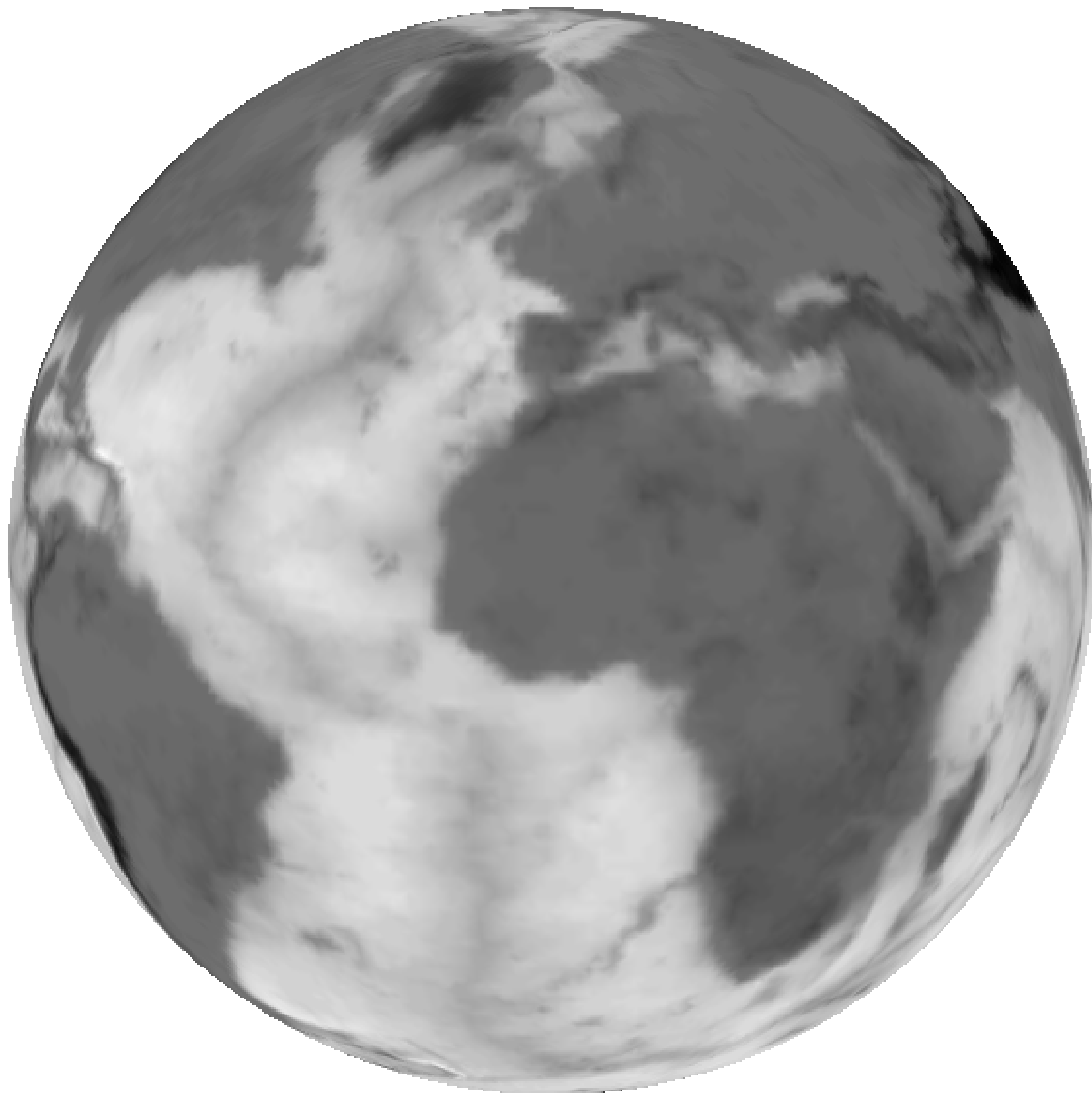


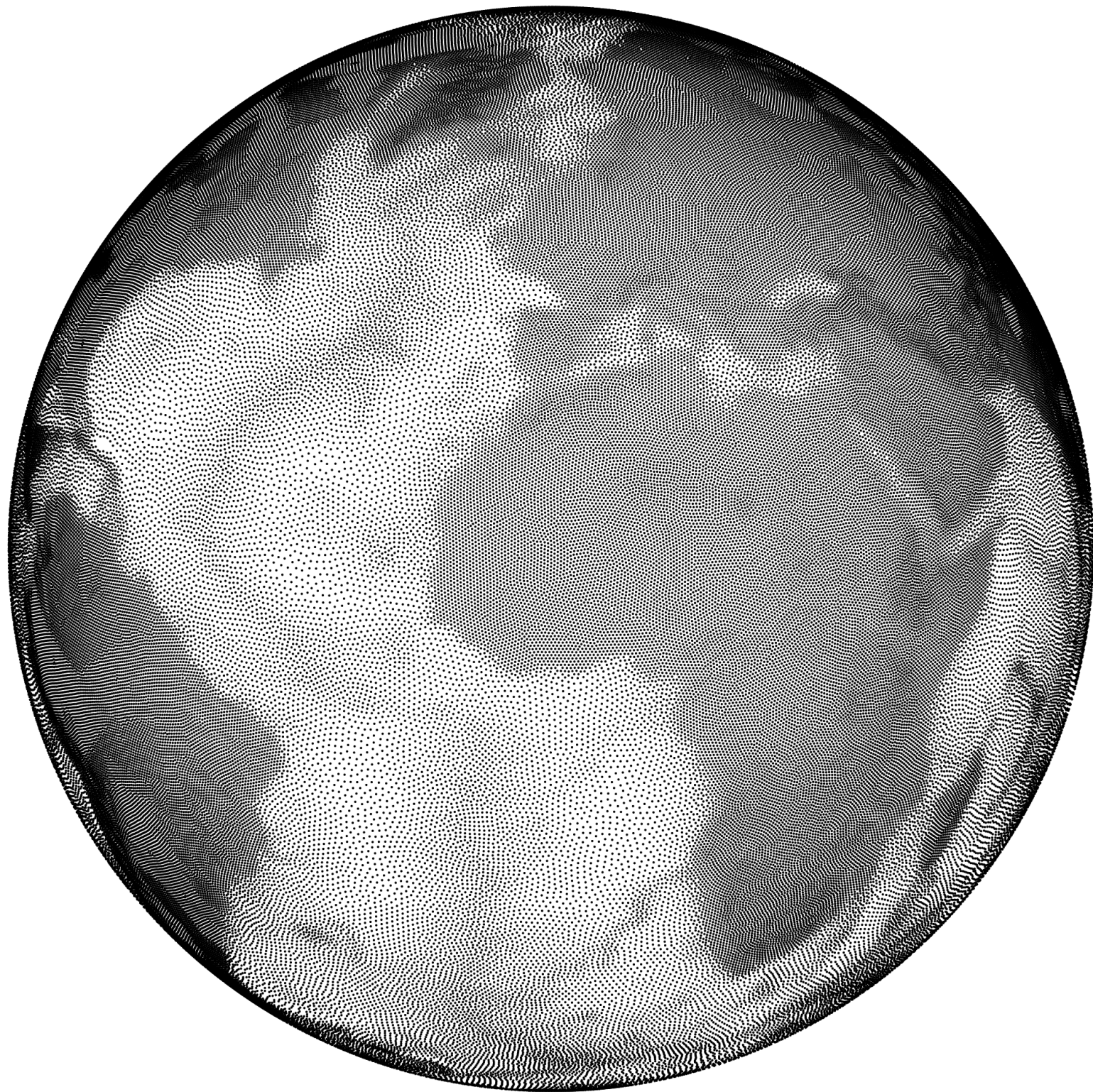


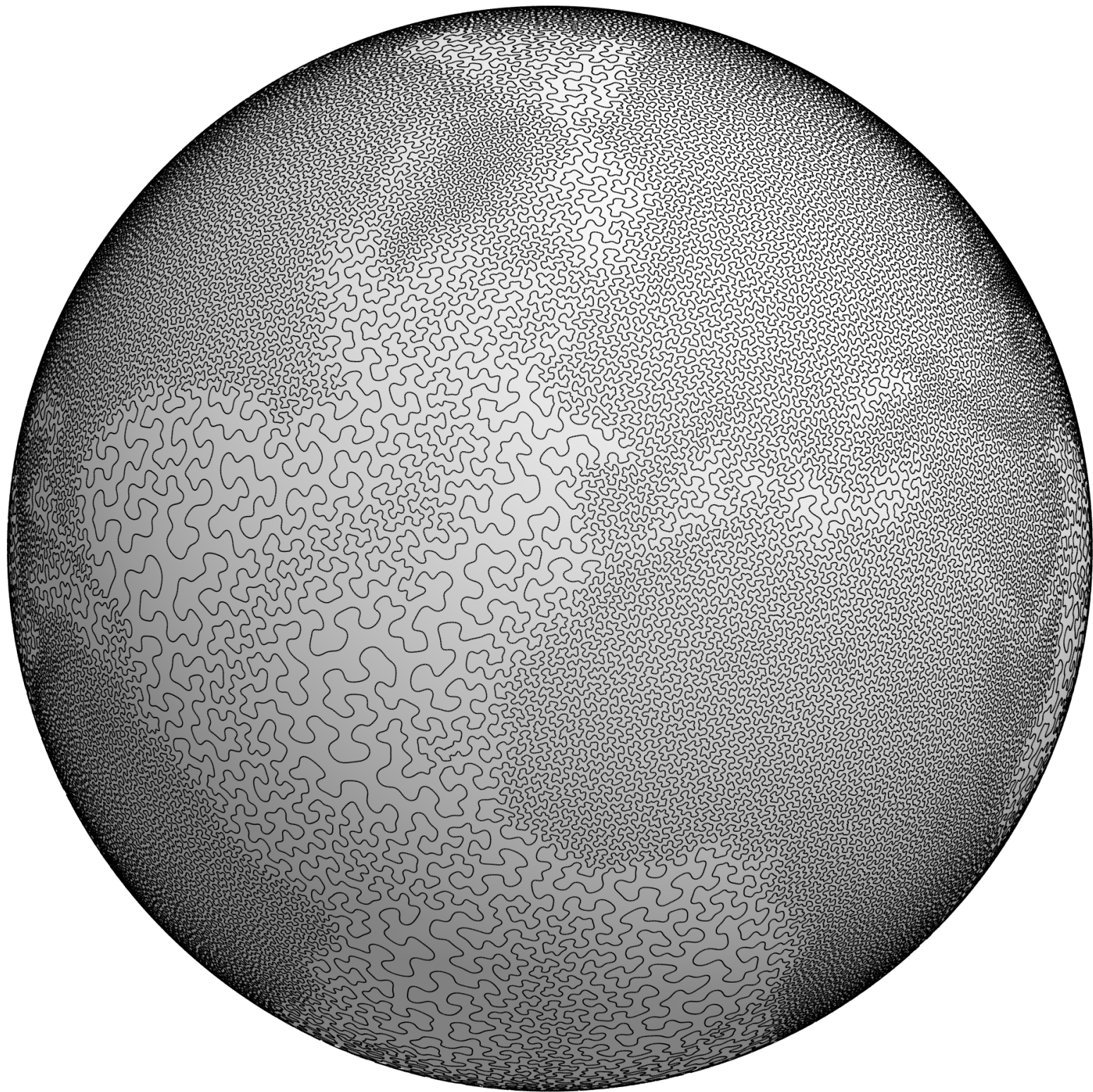


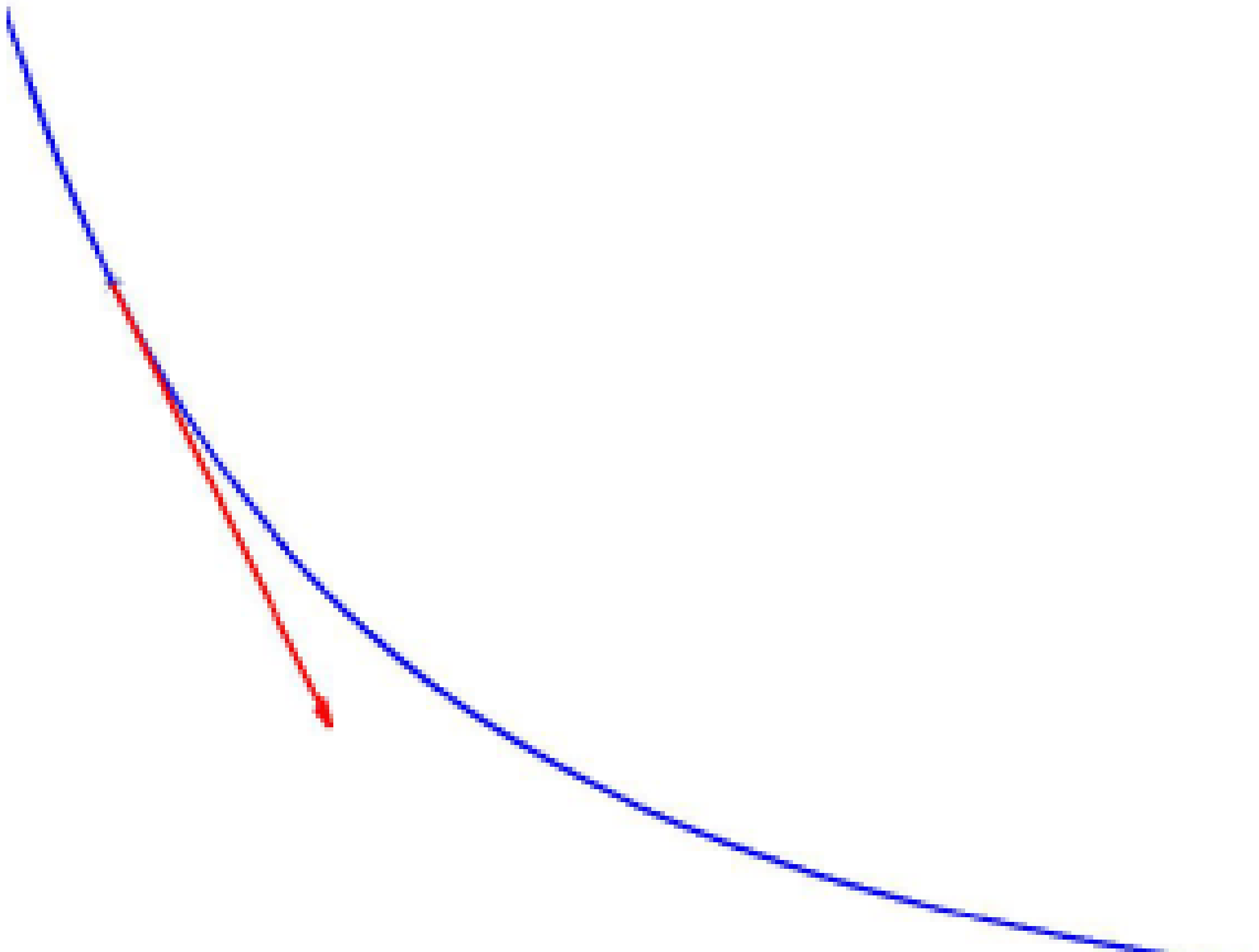


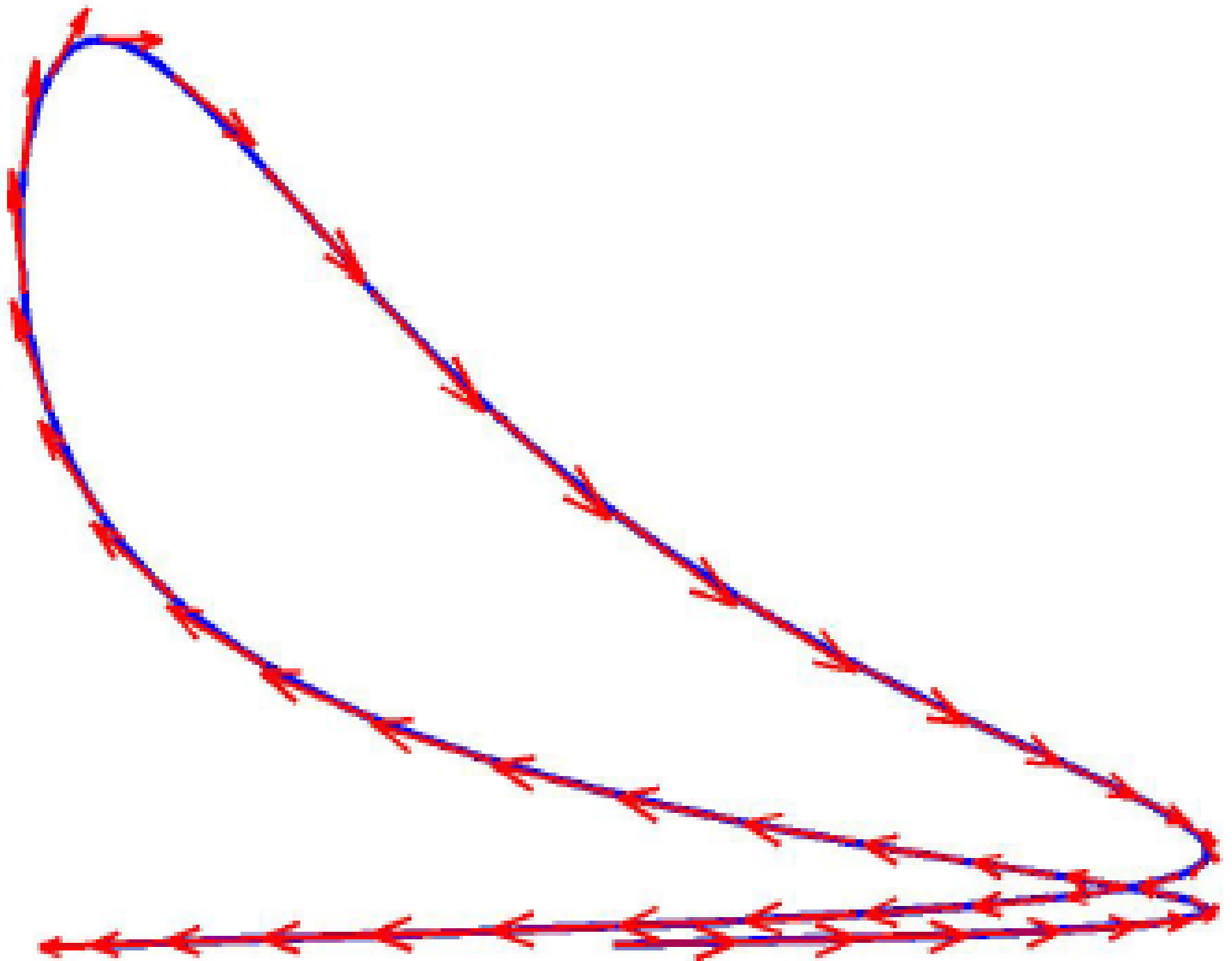


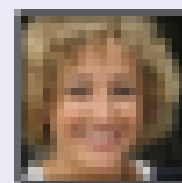
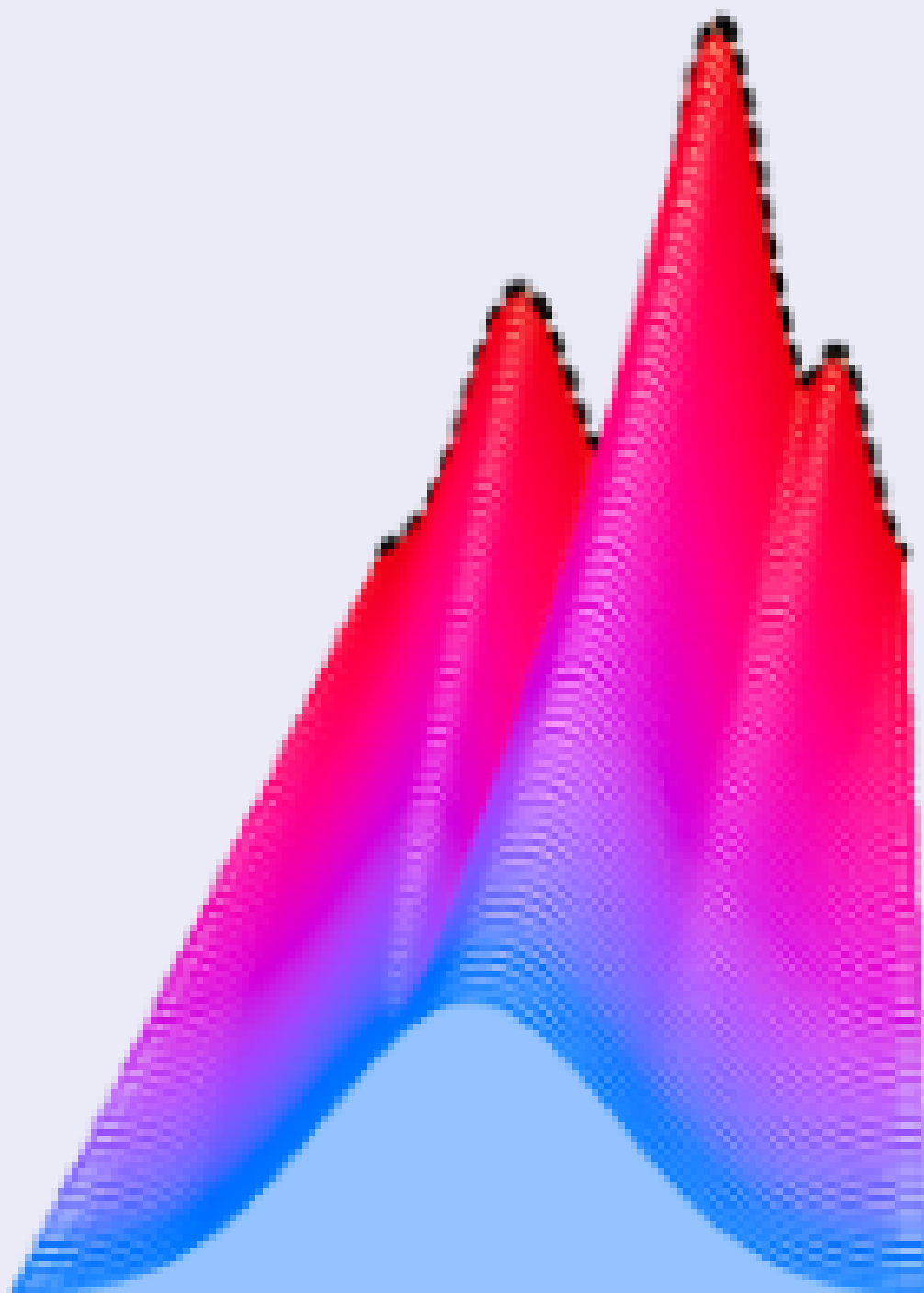












$t = 1$

$t = 0$



