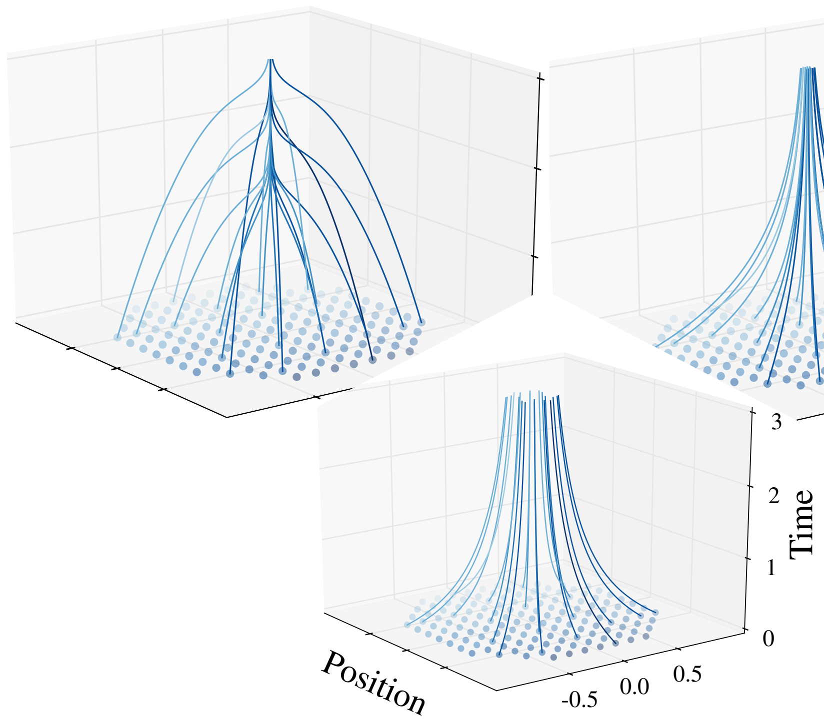
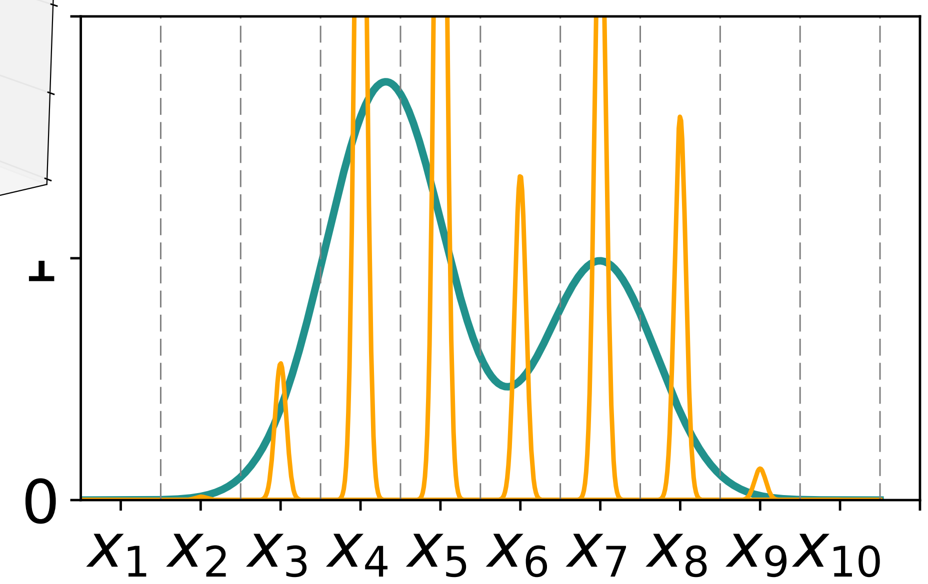




Gradient Flow in the Wasserstein Metric

Katy Craig
University of California, Santa Barbara

NIPS, Optimal Transport & Machine Learning
December 9th, 2017

gradient flow in finite dimensions

A curve x(t): [0,T] → $\mathbb{R}^d$ is the gradient flow of an energy E: $\mathbb{R}^d$ → $\mathbb{R}$ if

$$\frac{d}{dt}x(t) = -\nabla E(x(t))$$

- "x(t) evolves in the direction of steepest descent of E"
- initial value problem: given x(0), find the gradient flow x(t)

Example:

metric	energy functional	gradient flow
$(\mathbb{R}^d, \lvert \cdot \rvert)$	$E(x) = \frac{1}{2}x^2$	$\frac{d}{dt}x(t) = -x(t)$

Given x(0) ∈ $\mathbb{R}^d$, x(t) = x(0)e^{-t} is unique solution of the gradient flow.

gradient flow in finite dimensions

Gradient flows often arise when solving optimization problems:

$$\min_{x \in \mathbb{R}^d} E(x)$$

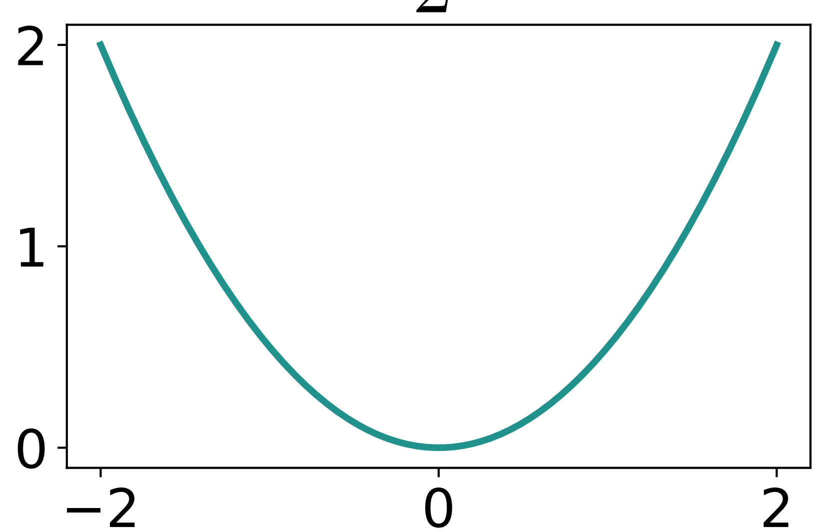
Convexity of the energy determines stability and long time behavior.

Def: An energy E is λ-convex if $D^2 E \geq \lambda I_{d \times d}$ or, equivalently, if

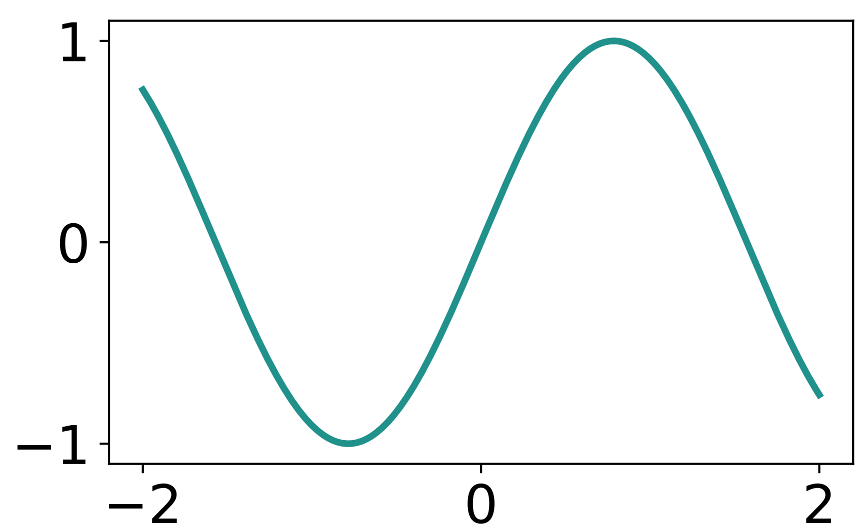
$$E((1-t)x + ty) \leq (1-t)E(x) + tE(y) - t(1-t)\frac{\lambda}{2}|x-y|^2$$

for all x,y ∊ $\mathbb{R}$, t ∊ [0,1].

$$f(x) = \frac{x^2}{2}, \quad \lambda = 1$$



$$f(x) = \sin(x), \quad \lambda = -1$$



gradient flow in finite dimensions

If E(x) is λ-convex, then…

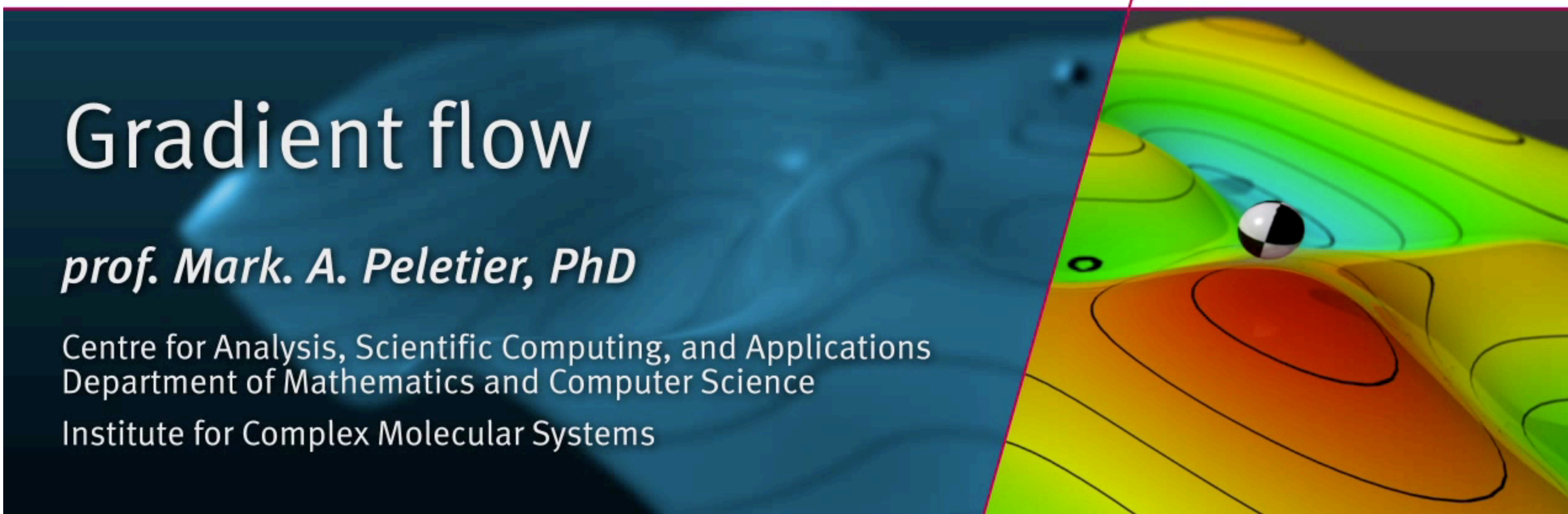
1) Stability: for any gradient flows x(t) and y(t),

$$|x(t) - y(t)| \leq e^{-\lambda t} |x(0) - y(0)|$$

2) long time behavior: if λ>0, there is a unique solution $\bar{x}$ of $\min_{x \in \mathbb{R}^d} E(x)$ and any gradient flow x(t) converges to $\bar{x}$ as t → +∞:

$$|x(t) - \bar{x}| \leq e^{-\lambda t} |x(0) - \bar{x}|$$

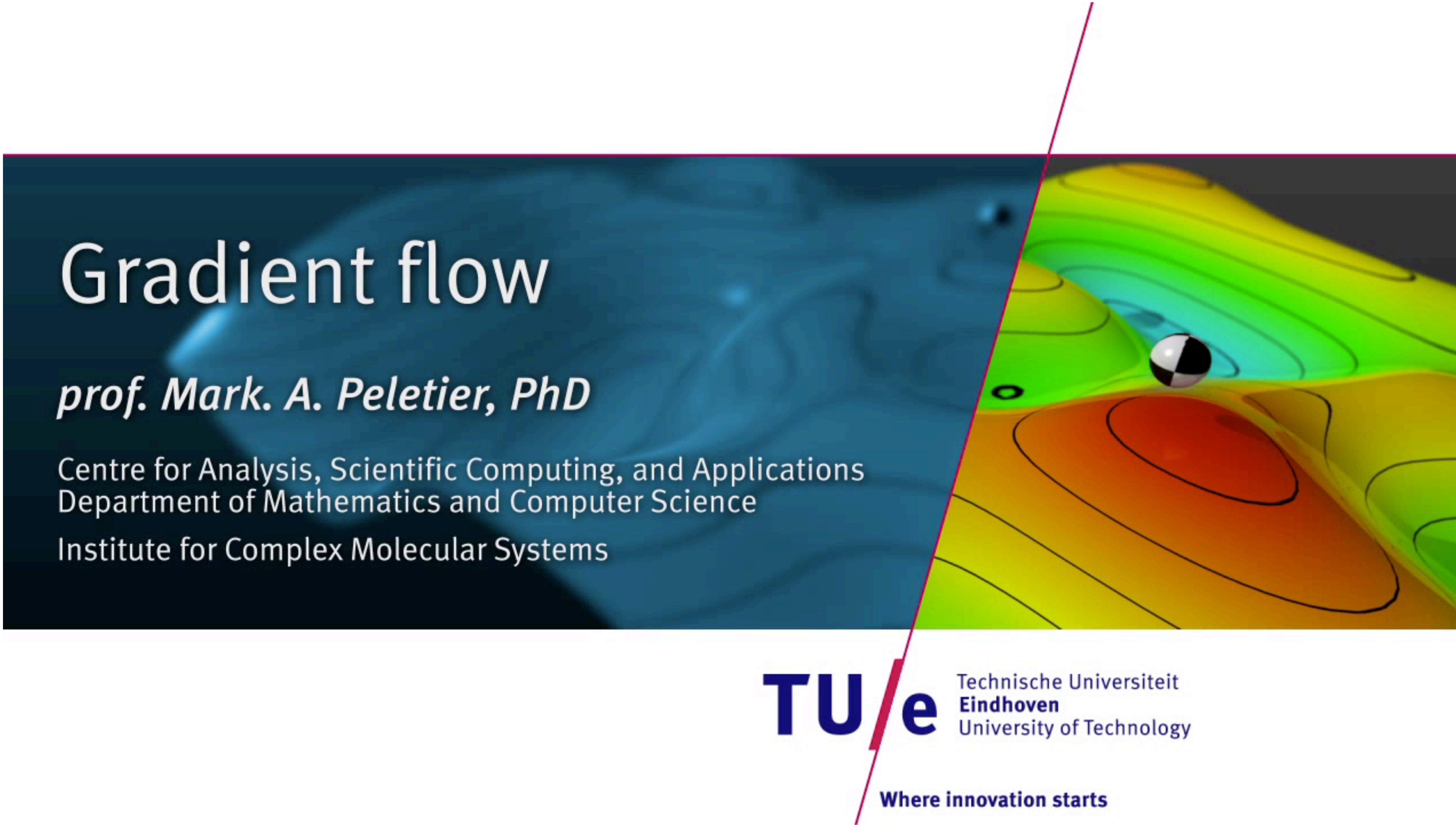
gradient flow



TU/e Technische Universiteit Eindhoven University of Technology

Where innovation starts

gradient flow



gradient flow with different metrics

In general, given a complete metric space (X,d), a curve x(t): $\mathbb{R}$ → X is the gradient flow of an energy E: X → $\mathbb{R}$ if

“ $$\frac{d}{dt}x(t) = -\nabla_X E(x(t))$$ ”

Examples:

	Euclidean	**L^2**
metric (X,d)	$(\mathbb{R}^d, \lvert \cdot \rvert)$	$(L^2(\mathbb{R}^d), \lVert \cdot \rVert_{L^2})$
def of ∇x	$\langle \nabla E(x), v \rangle = \lim_{h \to 0} \frac{E(x+hv) - E(x)}{h}$	$\langle \nabla E(f), g \rangle = \lim_{h \to 0} \frac{E(f+hg) - E(f)}{h}$
formula for ∇_X	$\nabla_{\mathbb{R}^d} E(x) = \nabla E(x)$	$\nabla_{L^2(\mathbb{R}^d)} E(f) = \frac{\partial E}{\partial f}$
energy	$E(x) = \frac{1}{2}x^2$	$E(f) = \frac{1}{2}\int \lvert f \rvert^2$
gradient flow	$\frac{d}{dt}x(t) = -x(t)$	$\frac{d}{dt}f(x,t) = -f(x,t)$

gradient flow with different metrics

In general, given a complete metric space (X,d), a curve x(t): $\mathbb{R} \rightarrow X$ is the gradient flow of an energy E: $X \rightarrow \mathbb{R}$ if

“ $$\frac{d}{dt}x(t) = -\nabla_X E(x(t))$$ ”

Examples:	Euclidean	L^2
metric (X,d)	$(\mathbb{R}^d, \|\cdot\|)$	$(L^2(\mathbb{R}^d), \\|\cdot\\|_{L^2})$
def of ∇_X	$\langle \nabla E(x), v \rangle = \lim_{h \to 0} \frac{E(x+hv) - E(x)}{h}$	$\langle \nabla E(f), g \rangle = \lim_{h \to 0} \frac{E(f+hg) - E(f)}{h}$
formula for ∇_X	$\nabla_{\mathbb{R}^d} E(x) = \nabla E(x)$	$\nabla_{L^2(\mathbb{R}^d)} E(f) = \frac{\partial E}{\partial f}$
energy	$E(x) = \frac{1}{2}x^2$	$E(f) = \frac{1}{2}\int \|\nabla f\|^2$
gradient flow	$\frac{d}{dt}x(t) = -x(t)$	$\frac{d}{dt}f(x,t) = -f(x,t)$

gradient flow with different metrics

In general, given a complete metric space (X,d), a curve x(t): $\mathbb{R} \to X$ is the gradient flow of an energy E: $X \to \mathbb{R}$ if

$$\text{“}\ \frac{d}{dt}x(t) = -\nabla_X E(x(t))\ \text{”}$$

Examples:

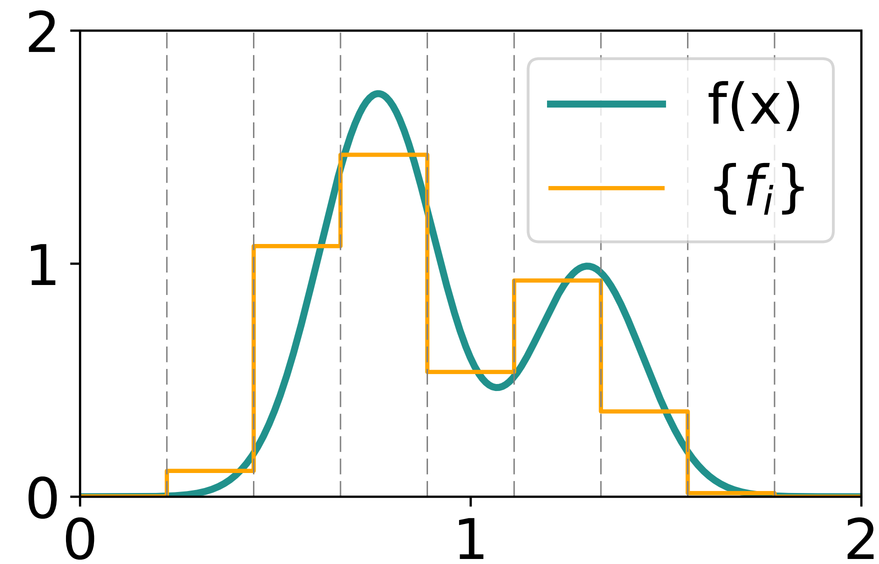
	Euclidean	L^2
metric (X,d)	$(\mathbb{R}^d, \lvert\cdot\rvert)$	$(L^2(\mathbb{R}^d), \lVert\cdot\rVert_{L^2})$
def of ∇_X	$\langle \nabla E(x), v\rangle = \lim_{h\to 0} \frac{E(x+hv)-E(x)}{h}$	$\langle \nabla E(f), g\rangle = \lim_{h\to 0} \frac{E(f+hg)-E(f)}{h}$
formula for ∇_X	$\nabla_{\mathbb{R}^d} E(x) = \nabla E(x)$	$\nabla_{L^2(\mathbb{R}^d)} E(f) = \frac{\partial E}{\partial f}$
energy	$E(x) = \frac{1}{2}x^2$	$E(f) = \frac{1}{2}\int \lvert\nabla f\rvert^2$
gradient flow	$\frac{d}{dt}x(t) = -x(t)$	$\frac{d}{dt}f(x,t) = \Delta f(x,t)$

gradient flow with different metrics

finite difference approximation

$f : \mathbb{R}^d \to \mathbb{R}$ approximated by $\{f_i\}_{i \in h\mathbb{Z}^d}$

approximate values of function



Examples:	Euclidean	L^2
metric (X,d)	$(\mathbb{R}^d, \lvert\cdot\rvert)$	$(L^2(\mathbb{R}^d), \lVert\cdot\rVert_{L^2})$
def of ∇x	$\langle \nabla E(x), v\rangle = \lim_{h\to 0} \frac{E(x+hv)-E(x)}{h}$	$\langle \nabla E(f), g\rangle = \lim_{h\to 0} \frac{E(f+hg)-E(f)}{h}$
formula for ∇_X	$\nabla_{\mathbb{R}^d} E(x) = \nabla E(x)$	$\nabla_{L^2(\mathbb{R}^d)} E(f) = \frac{\partial E}{\partial f}$
energy	$E(x) = \frac{1}{2}x^2$	$E(f) = \frac{1}{2}\int \lvert\nabla f\rvert^2$
gradient flow	$\frac{d}{dt}x(t) = -x(t)$	$\frac{d}{dt}f(x,t) = \Delta f(x,t)$

gradient flow with different metrics

Examples:	Euclidean	W_2
metric (X,d)	$(\mathbb{R}^d, \lvert \cdot \rvert)$	$(\mathcal{P}_2(\mathbb{R}^d), W_2)$
def of ∇x	$\langle \nabla E(x), v \rangle = \lim_{h \to 0} \frac{E(x+hv) - E(x)}{h}$	
formula for ∇_X	$\nabla_{\mathbb{R}^d} E(x) = \nabla E(x)$	$\nabla_{W_2} E(\rho) = -\nabla \cdot \left(\rho \nabla \frac{\partial E}{\partial \rho} \right)$
energy	$E(x) = \frac{1}{2} x^2$	$E(\rho) = \frac{1}{2} \int x^2 \rho(x) dx$
gradient flow	$\frac{d}{dt} x(t) = -x(t)$	$\frac{d}{dt} \rho(x,t) = \nabla \cdot (x \rho(x,t))$

gradient flow with different metrics

$$\langle \nabla E(\mu), -\nabla \cdot (\xi \mu) \rangle_{\mathrm{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d)} = \lim_{h \to 0} \frac{E((\mathrm{id} + h\xi)\#\mu) - E(\mu)}{h}$$

Examples:	Euclidean	W_2
metric (X,d)	$(\mathbb{R}^d, \lvert \cdot \rvert)$	$(\mathcal{P}_2(\mathbb{R}^d), W_2)$
def of ∇_X	$\langle \nabla E(x), v \rangle = \lim_{h \to 0} \frac{E(x+hv) - E(x)}{h}$	
formula for ∇_X	$\nabla_{\mathbb{R}^d} E(x) = \nabla E(x)$	$\nabla_{W_2} E(\rho) = -\nabla \cdot \left(\rho \nabla \frac{\partial E}{\partial \rho} \right)$
energy	$E(x) = \frac{1}{2} x^2$	$E(\rho) = \frac{1}{2} \int x^2 \rho(x) dx$
gradient flow	$\frac{d}{dt} x(t) = -x(t)$	$\frac{d}{dt} \rho(x,t) = \nabla \cdot (x \rho(x,t))$

gradient flow with different metrics

Examples:	Euclidean	W_2
metric (X,d)	$(\mathbb{R}^d, \lvert\cdot\rvert)$	$(\mathcal{P}_2(\mathbb{R}^d), W_2)$
def of ∇_X	$\langle \nabla E(x), v \rangle = \lim_{h \to 0} \frac{E(x+hv) - E(x)}{h}$	
formula for ∇_X	$\nabla_{\mathbb{R}^d} E(x) = \nabla E(x)$	$\nabla_{W_2} E(\rho) = -\nabla \cdot \left(\rho \nabla \frac{\partial E}{\partial \rho} \right)$
energy	$E(x) = \frac{1}{2} x^2$	$E(\rho) = \frac{1}{2} \int x^2 \rho(x) dx$
gradient flow	$\frac{d}{dt} x(t) = -x(t)$	$\frac{d}{dt} \rho(x,t) = \nabla \cdot (x \rho(x,t))$

gradient flow with different metrics

Examples:	Euclidean	W_2
metric (X,d)	$(\mathbb{R}^d, \lvert\cdot\rvert)$	$(\mathcal{P}_2(\mathbb{R}^d), W_2)$
def of ∇x	$\langle \nabla E(x), v\rangle = \lim_{h\to 0} \frac{E(x+hv)-E(x)}{h}$	
formula for ∇_X	$\nabla_{\mathbb{R}^d} E(x) = \nabla E(x)$	$\nabla_{W_2} E(\rho) = -\nabla \cdot \left(\rho \nabla \frac{\partial E}{\partial \rho}\right)$
energy	$E(x) = \frac{1}{2}x^2$	$E(\rho) = \int \rho(x)\log(\rho(x))dx$
gradient flow	$\frac{d}{dt}x(t) = -x(t)$	$\frac{d}{dt}\rho(x,t) = \nabla\cdot(x\rho(x,t))$

gradient flow with different metrics

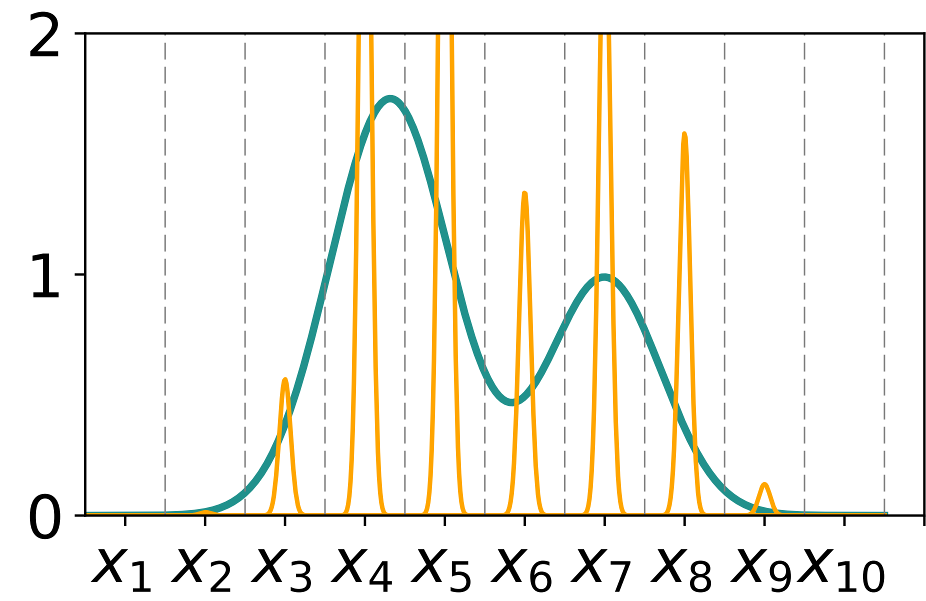
Examples:	Euclidean	W_2
metric (X,d)	$(\mathbb{R}^d, \lvert \cdot \rvert)$	$(\mathcal{P}_2(\mathbb{R}^d), W_2)$
def of ∇_X	$\langle \nabla E(x), v \rangle = \lim_{h \to 0} \frac{E(x+hv) - E(x)}{h}$	
formula for ∇_X	$\nabla_{\mathbb{R}^d} E(x) = \nabla E(x)$	$\nabla_{W_2} E(\rho) = -\nabla \cdot \left(\rho \nabla \frac{\partial E}{\partial \rho} \right)$
energy	$E(x) = \frac{1}{2} x^2$	$E(\rho) = \int \rho(x) \log(\rho(x)) dx$
gradient flow	$\frac{d}{dt} x(t) = -x(t)$	$\frac{d}{dt} \rho(x,t) = \Delta \rho(x,t)$

gradient flow with different metrics

particle approximation

$f : \mathbb{R}^d \to \mathbb{R}$ approximated by $\sum_{i=1}^{N} \delta_{x_i} m_i$

approximate mass of function



Examples:	Euclidean	W_2
metric (X,d)	$(\mathbb{R}^d, \lvert \cdot \rvert)$	$(\mathcal{P}_2(\mathbb{R}^d), W_2)$
def of ∇_X	$\langle \nabla E(x), v \rangle = \lim_{h \to 0} \frac{E(x+hv) - E(x)}{h}$	
formula for ∇_X	$\nabla_{\mathbb{R}^d} E(x) = \nabla E(x)$	$\nabla_{W_2} E(\rho) = -\nabla \cdot \left(\rho \nabla \frac{\partial E}{\partial \rho} \right)$
energy	$E(x) = \frac{1}{2} x^2$	$E(\rho) = \int \rho(x) \log(\rho(x)) dx$
gradient flow	$\frac{d}{dt} x(t) = -x(t)$	$\frac{d}{dt} \rho(x,t) = \Delta \rho(x,t)$

interpolating with different metrics

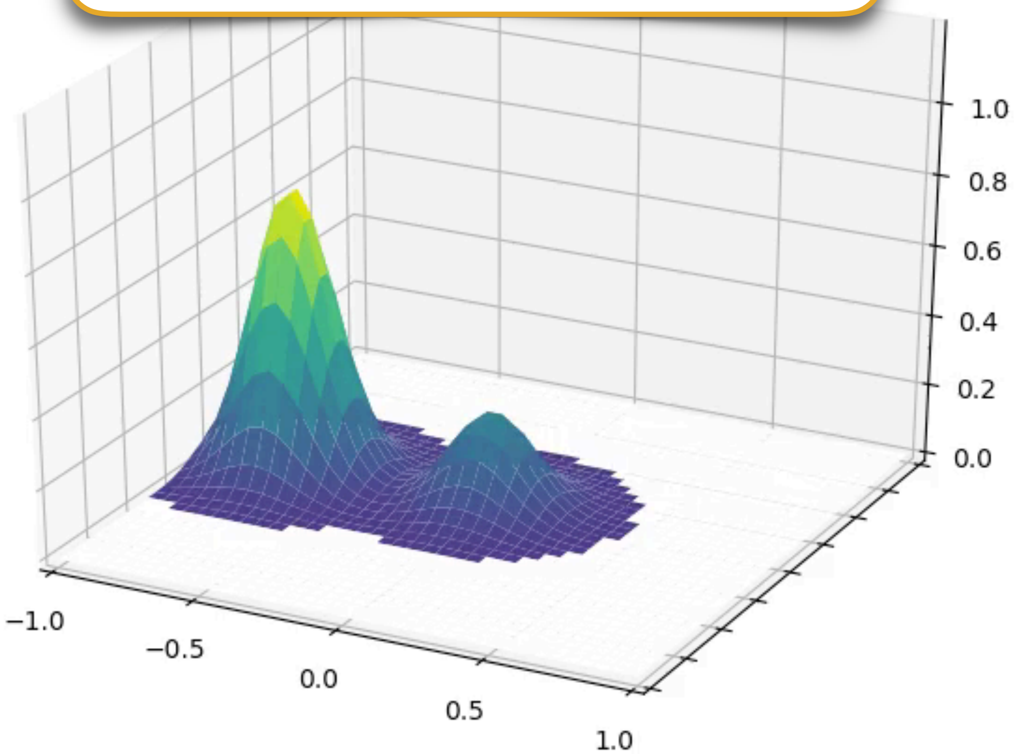
The same dichotomy between **values of a function** and **mass of a function** is also present in the geodesics.

Def: A constant speed geodesic between two points ρ_0 and ρ_1 in a metric space (X,d) is any curve $\rho:[0,1]\to X$ s.t.

$$\rho(0) = \rho_0, \;\; \rho(1) = \rho_1, \;\; d(\rho(t), \rho(s)) = |t - s| d(\rho_0, \rho_1)$$

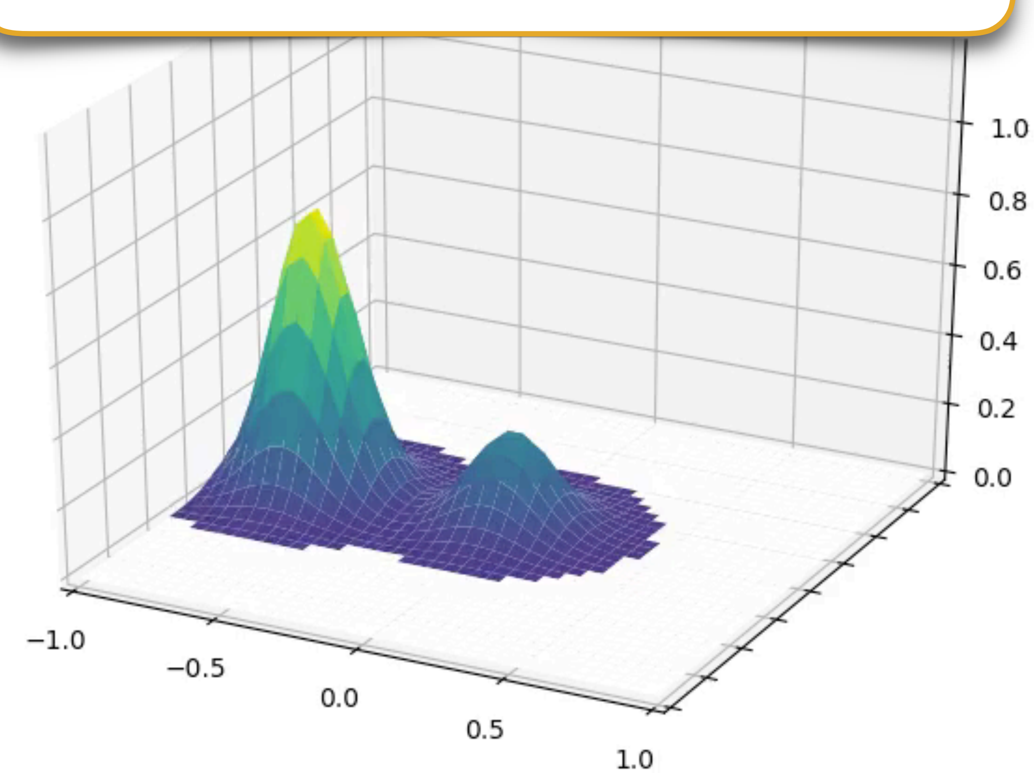
L^2 geodesic

$$\rho(t) = (1-t)\rho_0 + t\rho_1$$



W$_2$ geodesic

$$\rho(t) = ((1-t)\mathrm{id} + tT_{\rho_0}^{\rho_1})\#\rho_0$$



interpolating with different metrics

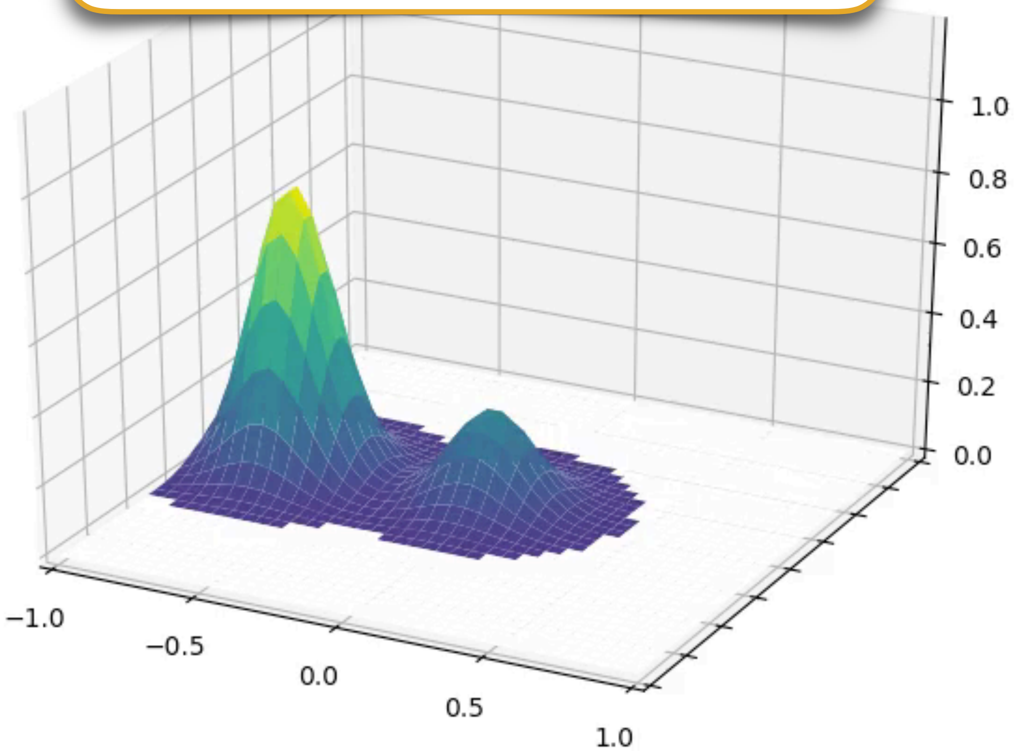
The same dichotomy between **values of a function** and **mass of a function** is also present in the geodesics.

Def: A constant speed geodesic between two points ρ_0 and ρ_1 in a metric space (X,d) is any curve $\rho:[0,1]\to X$ s.t.

$$\rho(0) = \rho_0,\ \ \rho(1) = \rho_1,\ \ d(\rho(t),\rho(s)) = |t-s|d(\rho_0,\rho_1)$$

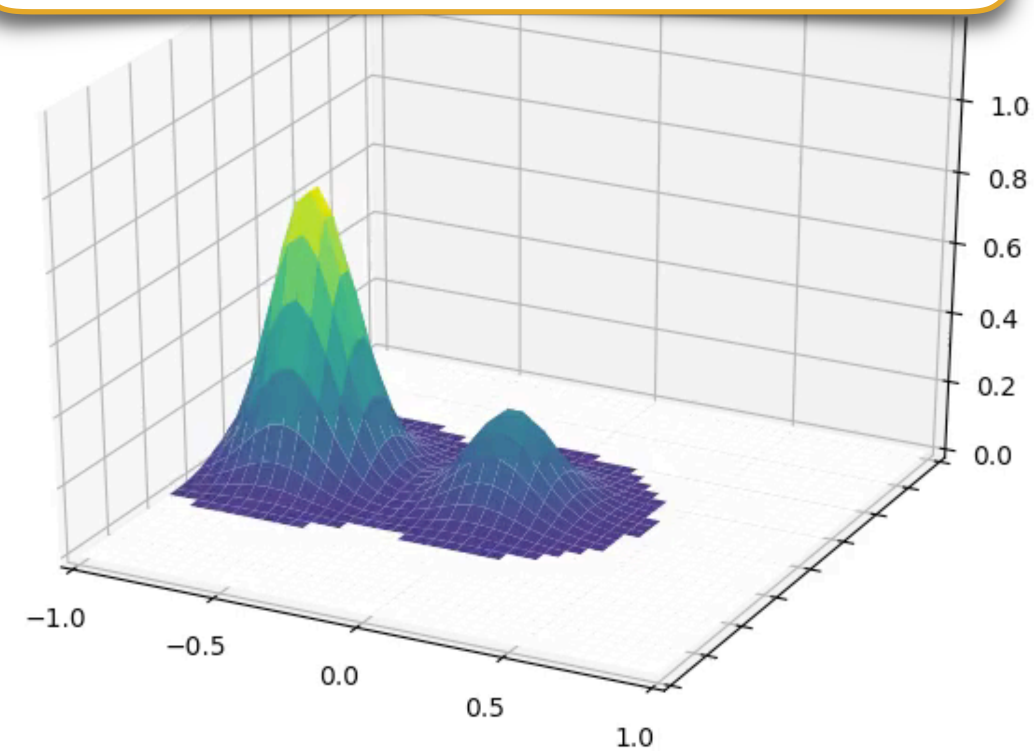
L^2 geodesic

$$\rho(t) = (1-t)\rho_0 + t\rho_1$$



W_2 geodesic

$$\rho(t) = ((1-t)\mathrm{id} + tT_{\rho_0}^{\rho_1})\#\rho_0$$



interpolating with different metrics

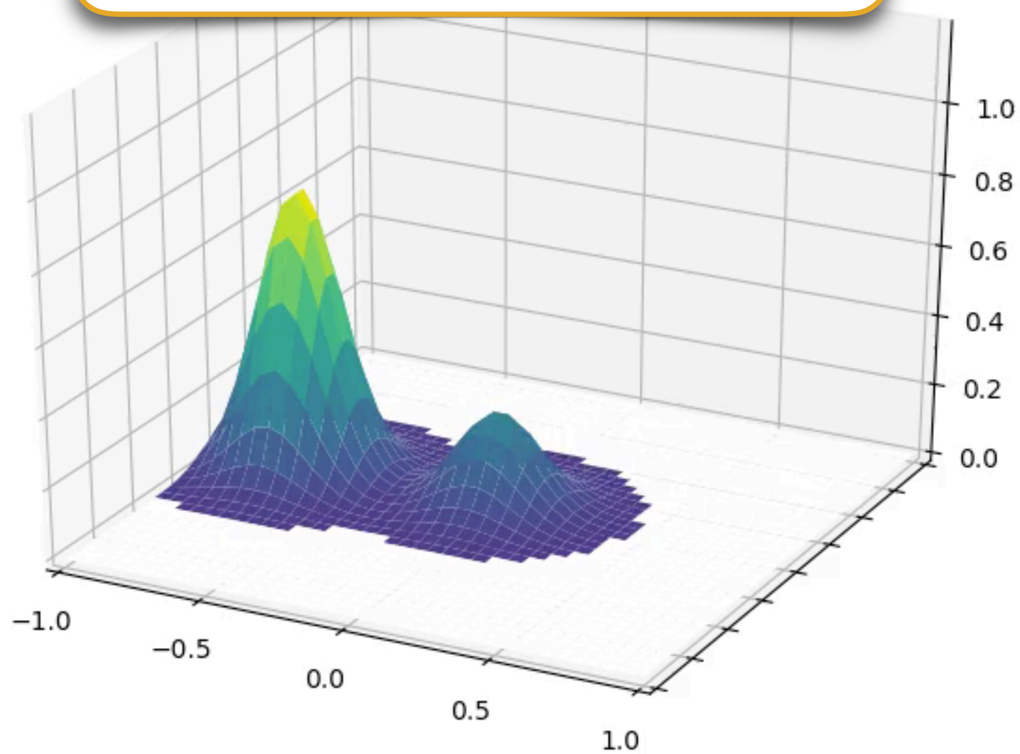
The same dichotomy between **values of a function** and **mass of a function** is also present in the geodesics.

Def: A constant speed geodesic between two points ρ_0 and ρ_1 in a metric space (X,d) is any curve $\rho:[0,1]\rightarrow X$ s.t.

$$\rho(0) = \rho_0,\ \ \rho(1) = \rho_1,\ \ d(\rho(t), \rho(s)) = |t - s| d(\rho_0, \rho_1)$$

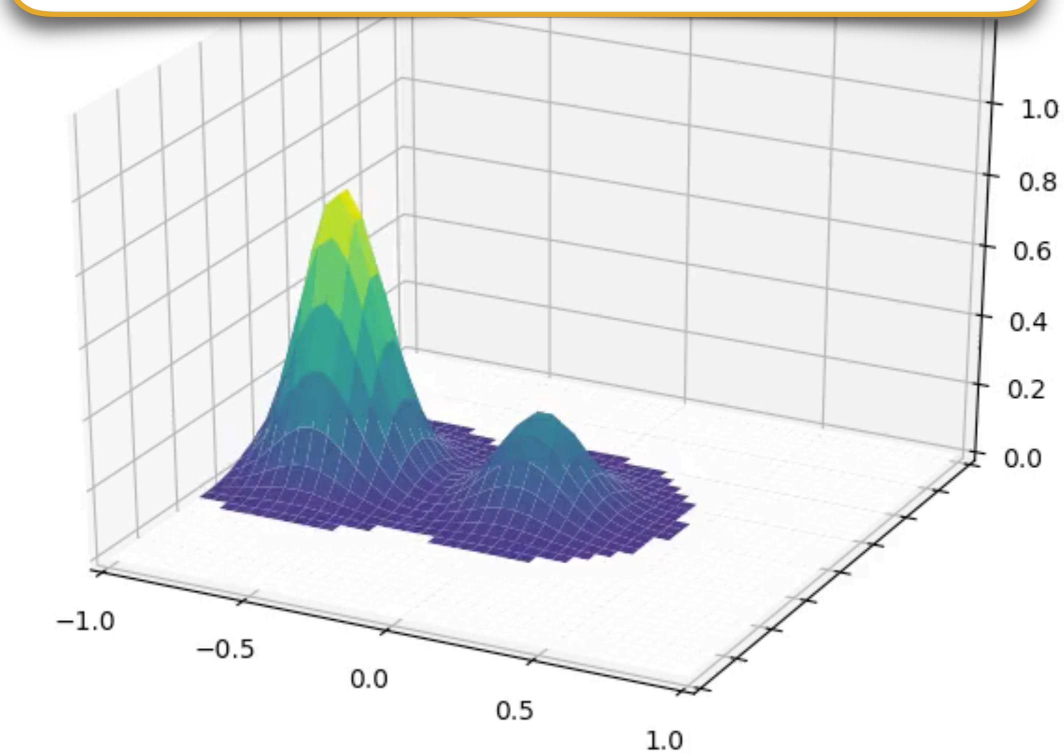
L^2 geodesic

$$\rho(t) = (1-t)\rho_0 + t\rho_1$$



W_2 geodesic

$$\rho(t) = ((1-t)\mathrm{id} + tT_{\rho_0}^{\rho_1})\#\rho_0$$



gradient flow in the Wasserstein metric

Examples:

energy functional	gradient flow
$E(\rho) = \int \rho \log \rho$	$\frac{d}{dt}\rho = \Delta \rho$
$E(\rho) = \frac{1}{m-1} \int \rho^m$	$\frac{d}{dt}\rho = \Delta \rho^m$
$E(\rho) = \int V \rho$	$\frac{d}{dt}\rho = \nabla \cdot (\nabla V \rho)$
$E(\rho) = \int (K * \rho)\rho$	$\frac{d}{dt}\rho = \nabla \cdot (\nabla (K * \rho)\rho)$

All Wasserstein gradient flows are of the form

$$\frac{d}{dt}\rho + \nabla \cdot (v\rho) = 0$$

continuity equation

gradient flow in the Wasserstein metric

Examples:

energy functional	gradient flow
$E(\rho) = \int \rho \log \rho$	$\frac{d}{dt}\rho = \Delta\rho$ $v = -\frac{\nabla\rho}{\rho}$
$E(\rho) = \frac{1}{m-1}\int \rho^m$	$\frac{d}{dt}\rho = \Delta\rho^m$
$E(\rho) = \int V\rho$	$\frac{d}{dt}\rho = \nabla \cdot (\nabla V \rho)$
$E(\rho) = \int (K * \rho)\rho$	$\frac{d}{dt}\rho = \nabla \cdot (\nabla(K * \rho)\rho)$

All Wasserstein gradient flows are of the form

$$\frac{d}{dt}\rho + \nabla \cdot (v\rho) = 0$$

continuity equation

gradient flow in the Wasserstein metric

Examples:

energy functional	gradient flow
$E(\rho) = \int \rho \log \rho$	$\frac{d}{dt}\rho = \Delta \rho$ $v = -\frac{\nabla \rho}{\rho}$
$E(\rho) = \frac{1}{m-1} \int \rho^m$	$\frac{d}{dt}\rho = \Delta \rho^m$ $v = -m\rho^{m-2}\nabla \rho$
$E(\rho) = \int V \rho$	$\frac{d}{dt}\rho = \nabla \cdot (\nabla V \rho)$
$E(\rho) = \int (K * \rho)\rho$	$\frac{d}{dt}\rho = \nabla \cdot (\nabla (K * \rho)\rho)$

All Wasserstein gradient flows are of the form

$$\frac{d}{dt}\rho + \nabla \cdot (v\rho) = 0$$

continuity equation

gradient flow in the Wasserstein metric

Examples:

energy functional	gradient flow	
$E(\rho) = \int \rho \log \rho$	$\frac{d}{dt}\rho = \Delta\rho$	$v = -\frac{\nabla\rho}{\rho}$
$E(\rho) = \frac{1}{m-1}\int \rho^m$	$\frac{d}{dt}\rho = \Delta\rho^m$	$v = -m\rho^{m-2}\nabla\rho$
$E(\rho) = \int V\rho$	$\frac{d}{dt}\rho = \nabla\cdot(\nabla V\rho)$	$v = -\nabla V$
$E(\rho) = \int (K*\rho)\rho$	$\frac{d}{dt}\rho = \nabla\cdot(\nabla(K*\rho)\rho)$	

All Wasserstein gradient flows are of the form

$$\frac{d}{dt}\rho + \nabla\cdot(v\rho) = 0$$

continuity equation

gradient flow in the Wasserstein metric

Examples:

energy functional	gradient flow	
$E(\rho) = \int \rho \log \rho$	$\frac{d}{dt}\rho = \Delta\rho$	$v = -\frac{\nabla\rho}{\rho}$
$E(\rho) = \frac{1}{m-1}\int \rho^m$	$\frac{d}{dt}\rho = \Delta\rho^m$	$v = -m\rho^{m-2}\nabla\rho$
$E(\rho) = \int V\rho$	$\frac{d}{dt}\rho = \nabla\cdot(\nabla V\rho)$	$v = -\nabla V$
$E(\rho) = \int (K * \rho)\rho$	$\frac{d}{dt}\rho = \nabla\cdot(\nabla(K*\rho)\rho)$	$v = -\nabla(K*\rho)$

All Wasserstein gradient flows are of the form

$$\frac{d}{dt}\rho + \nabla\cdot(v\rho) = 0$$

continuity equation

gradient flow in the Wasserstein metric

Examples:

energy functional	gradient flow	
$E(\rho) = \int \rho \log \rho$	$\frac{d}{dt}\rho = \Delta \rho$	$v = -\frac{\nabla \rho}{\rho}$
$E(\rho) = \frac{1}{m-1} \int \rho^m$	$\frac{d}{dt}\rho = \Delta \rho^m$	$v = -m\rho^{m-2}\nabla \rho$
$E(\rho) = \int V \rho$	$\frac{d}{dt}\rho = \nabla \cdot (\nabla V \rho)$	$v = -\nabla V$
$E(\rho) = \int (K * \rho)\rho$	$\frac{d}{dt}\rho = \nabla \cdot (\nabla (K * \rho)\rho)$	$v = -\nabla (K * \rho)$

All Wasserstein gradient flows are of the form

$$\frac{d}{dt}\rho + \nabla \cdot (v\rho) = 0$$

continuity equation

$$v = -\nabla \frac{\partial E}{\partial \rho}$$

gradient flow in the Wasserstein metric

aggregation, drift, and degenerate diffusion:

$$\frac{d}{dt}\rho = \nabla \cdot ((\nabla K * \rho)\rho) + \nabla \cdot (\nabla V \rho) + \Delta \rho^m$$

$K, V : \mathbb{R}^d \to \mathbb{R}$, and $m \geq 1$

self interaction drift diffusion

$$E(\rho) = \frac{1}{2}\int K * \rho d\rho + \int V d\rho + \frac{1}{m-1}\int \rho^m$$

interaction kernels:

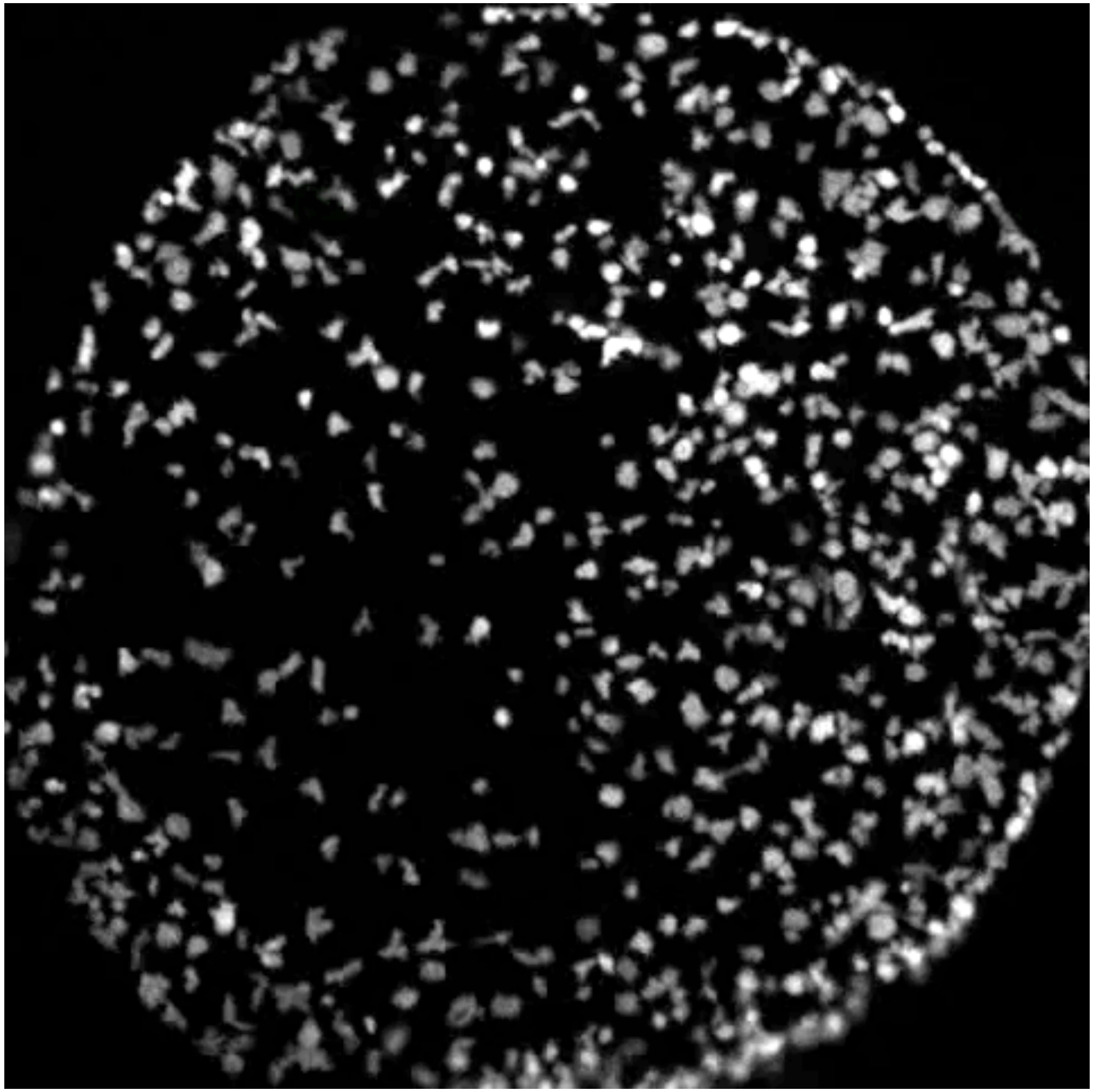
- granular media: $K(x) = |x|^3$
- swarming: $K(x) = |x|^a/a - |x|^b/b$, $-d<b<a$
- chemotaxis: $K(x) = \begin{cases} \frac{1}{2\pi}\log|x| & \text{if } d=2, \\ C_d|x|^{2-d} & \text{otherwise.} \end{cases}$

degenerate diffusion:

- $\Delta \rho^m = \nabla \cdot (\underbrace{m\rho^{m-1}}_{D} \nabla \rho)$

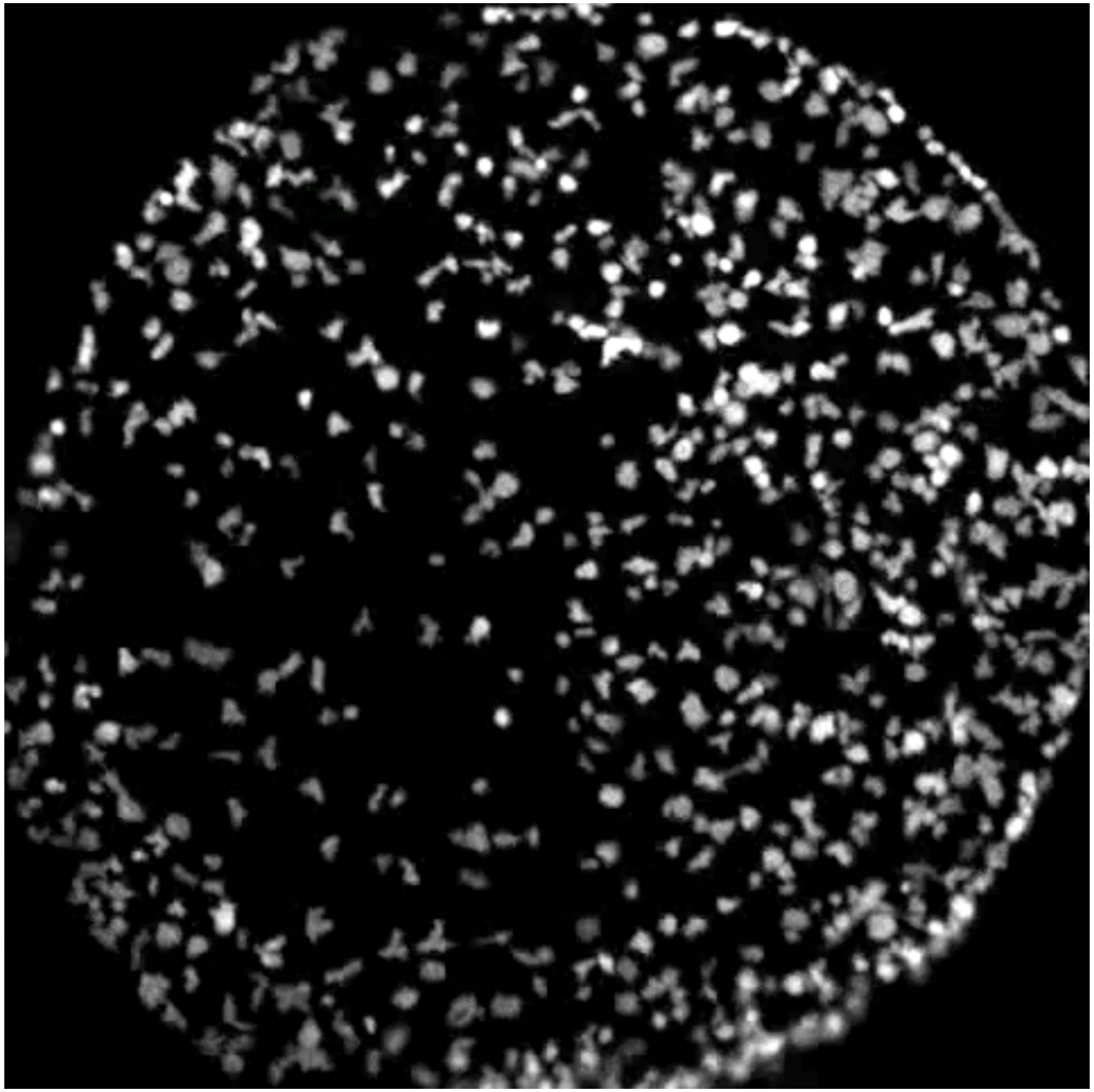
biological chemotaxis

a colony of slime mold [Gregor, et. al]



biological chemotaxis

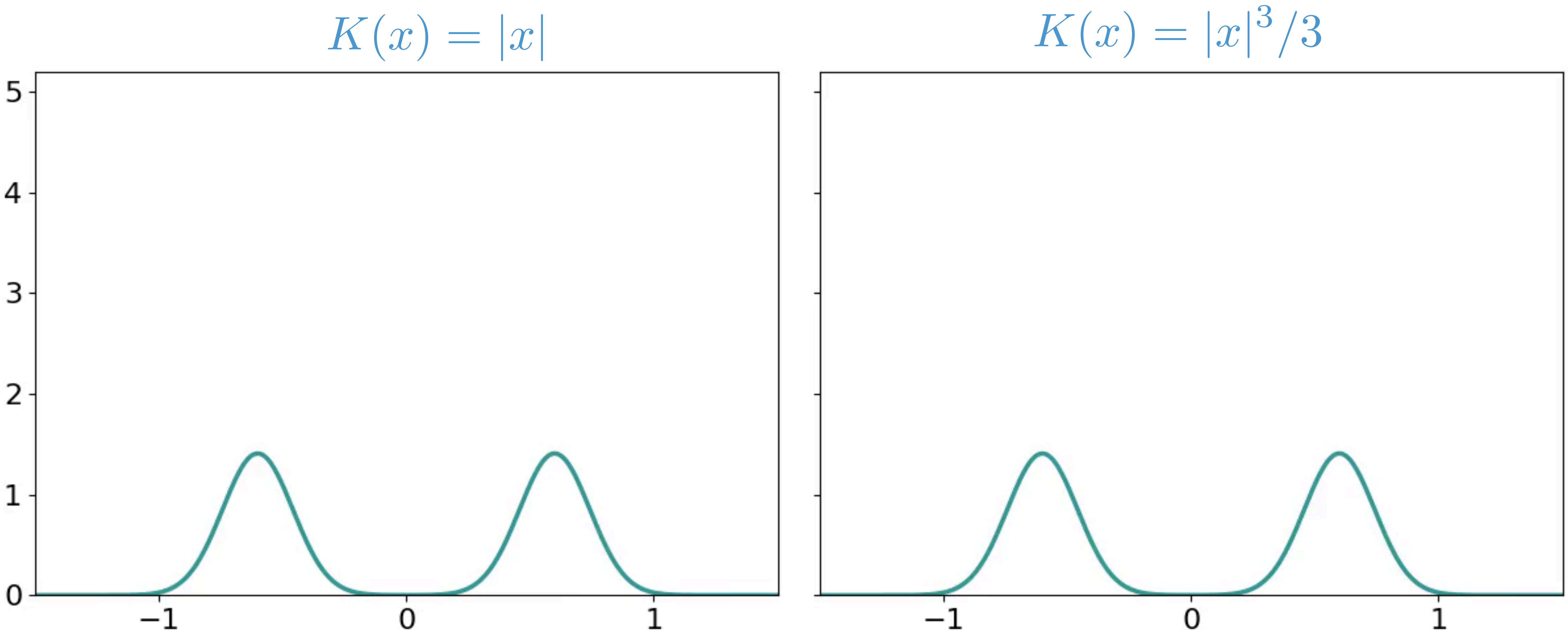
a colony of slime mold [Gregor, et. al]



gradient flow in the Wasserstein metric

aggregation, drift, and degenerate diffusion:

$$\frac{d}{dt}\rho = \nabla \cdot ((\nabla K * \rho)\rho) + \nabla \cdot (\nabla V \rho) + \Delta \rho^m$$

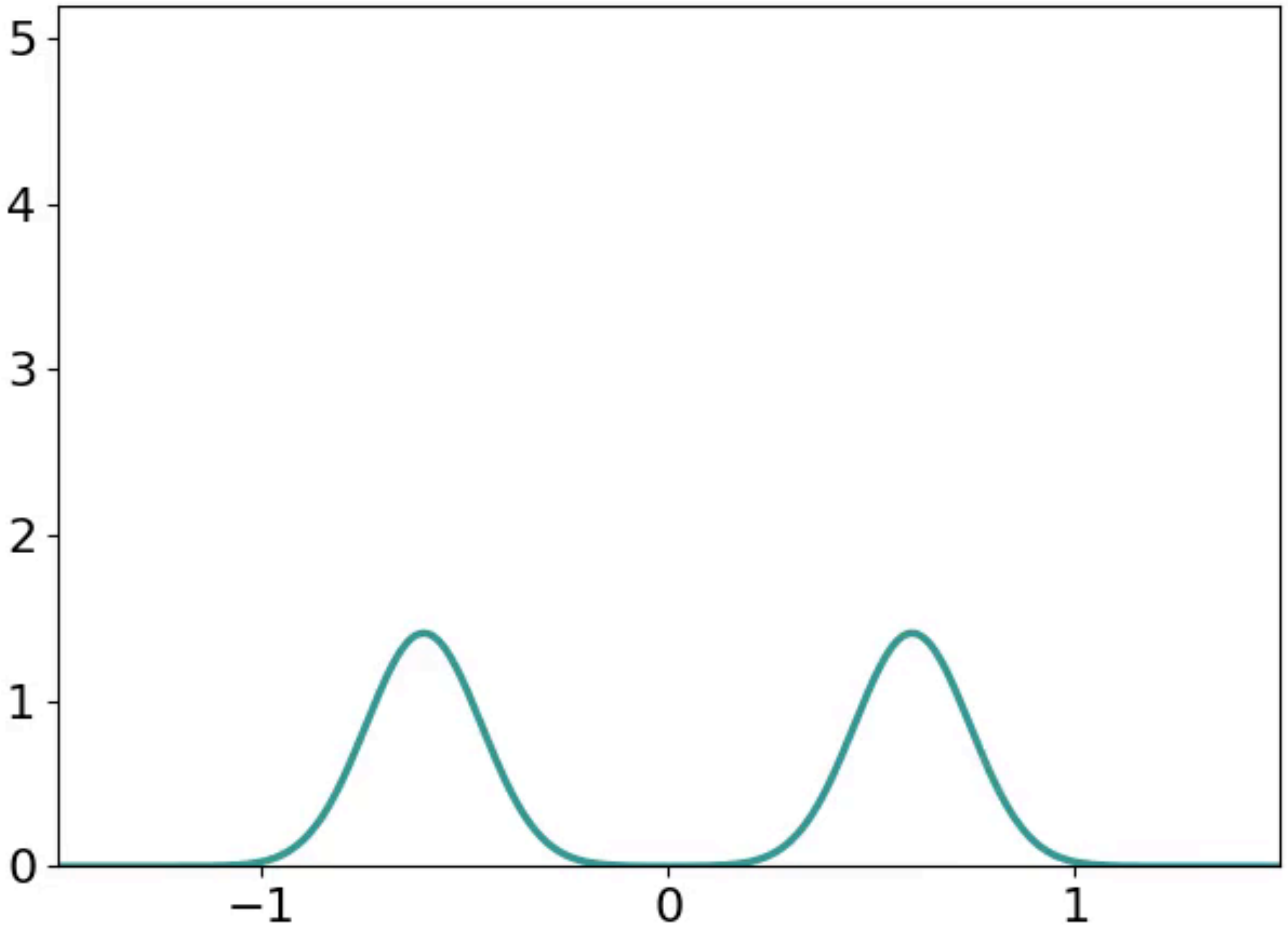


gradient flow in the Wasserstein metric

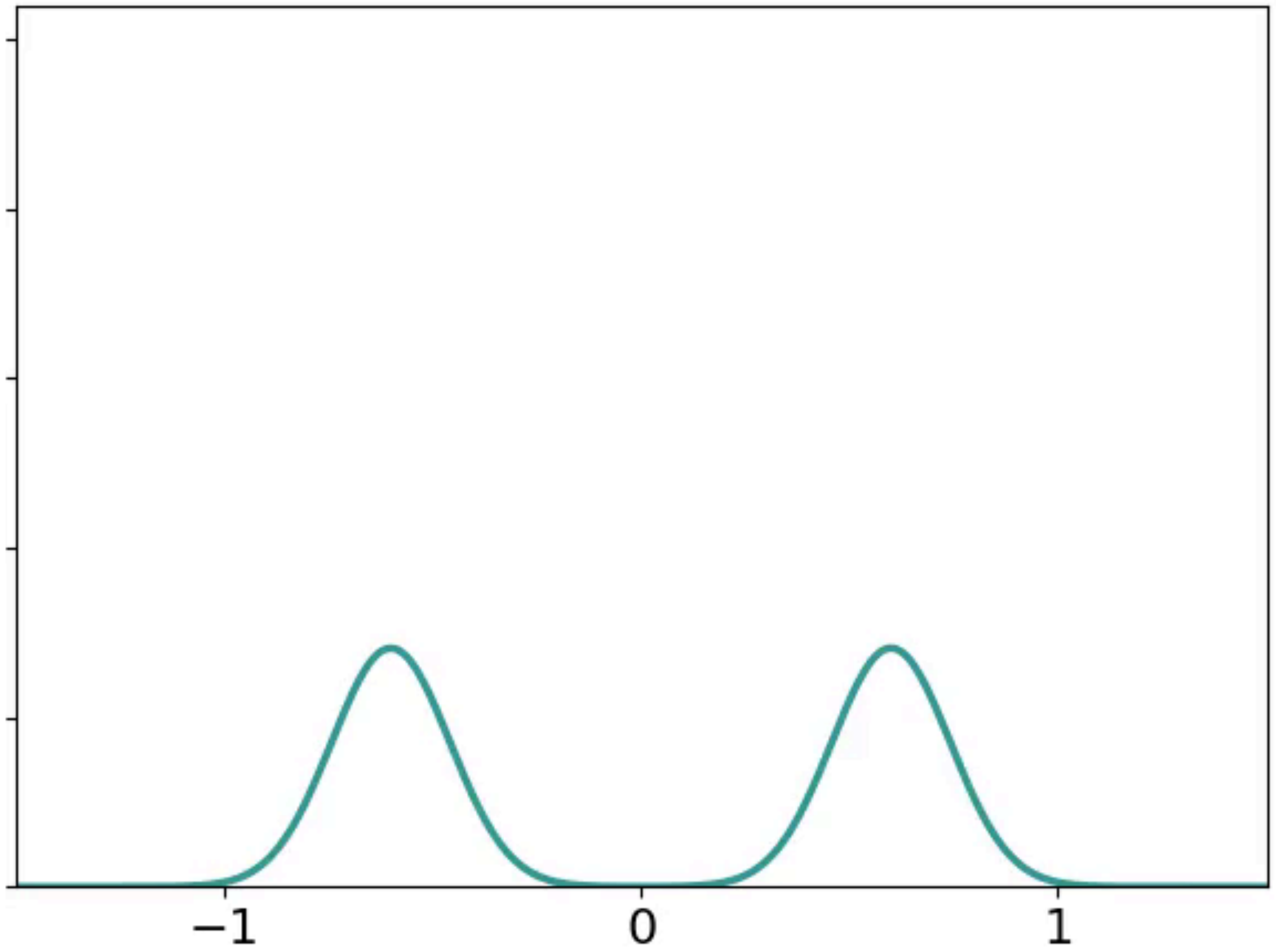
aggregation, drift, and degenerate diffusion:

$$\frac{d}{dt}\rho = \nabla \cdot ((\nabla K * \rho)\rho)$$

$K(x) = |x|$



$K(x) = |x|^3/3$



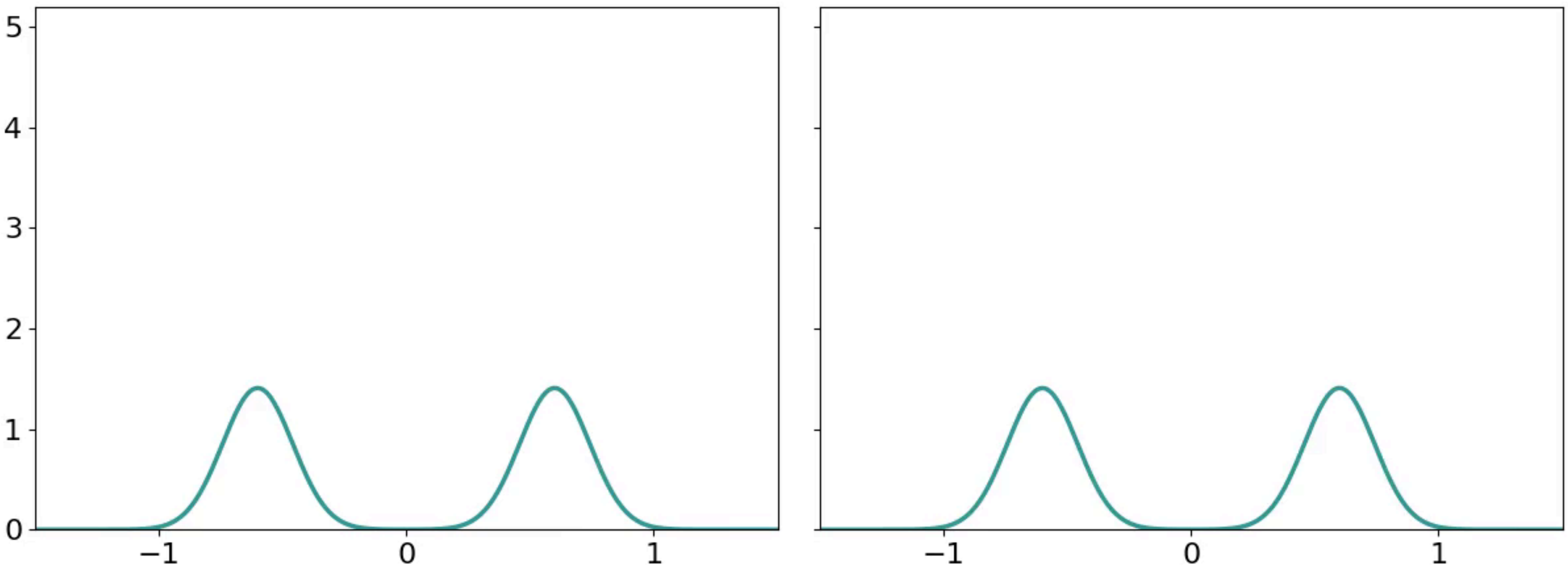
gradient flow in the Wasserstein metric

aggregation, drift, and degenerate diffusion:

$$\frac{d}{dt}\rho = \nabla \cdot ((\nabla K * \rho)\rho)$$

$K(x) = |x|$

$K(x) = |x|^3/3$

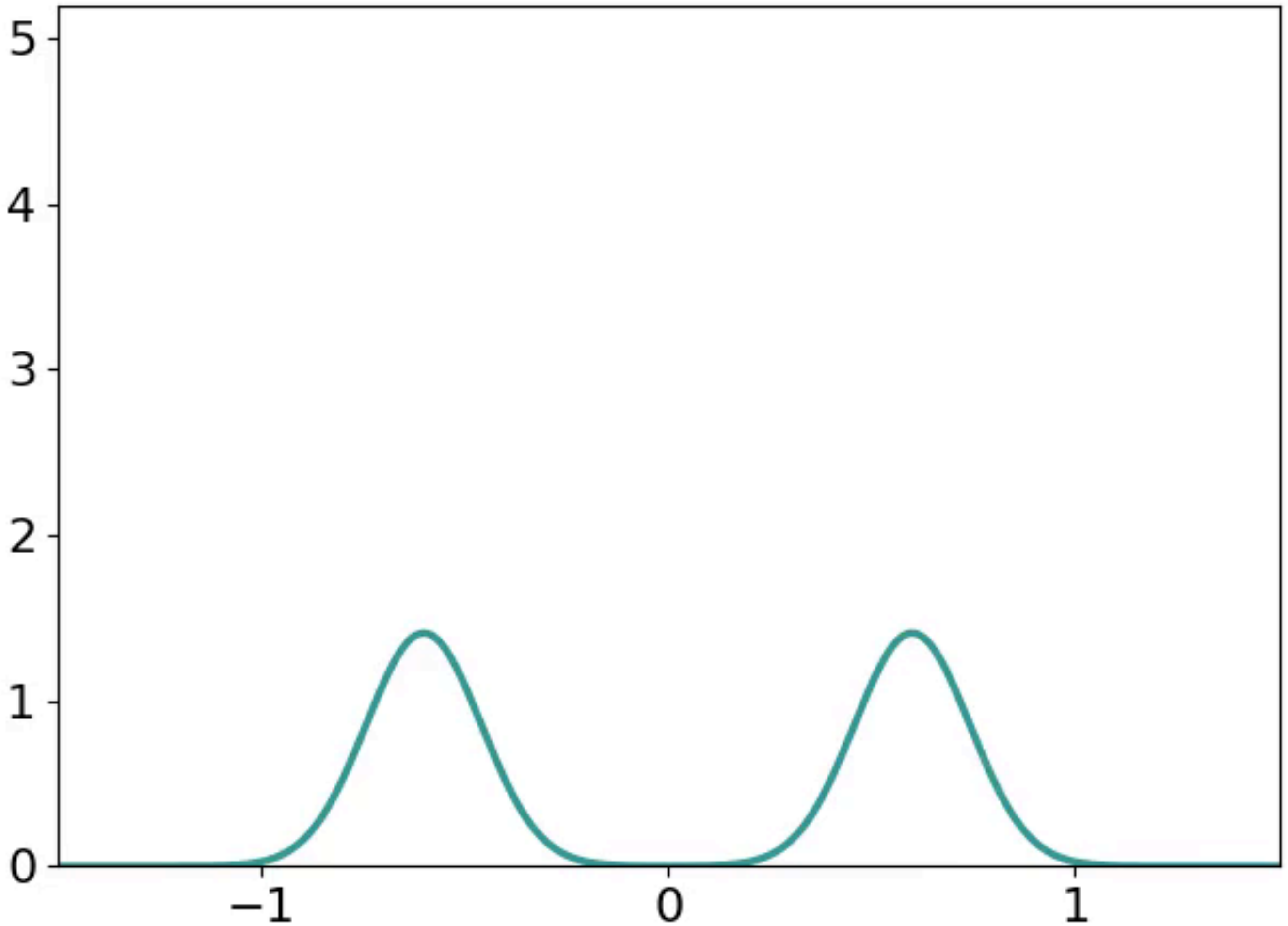


gradient flow in the Wasserstein metric

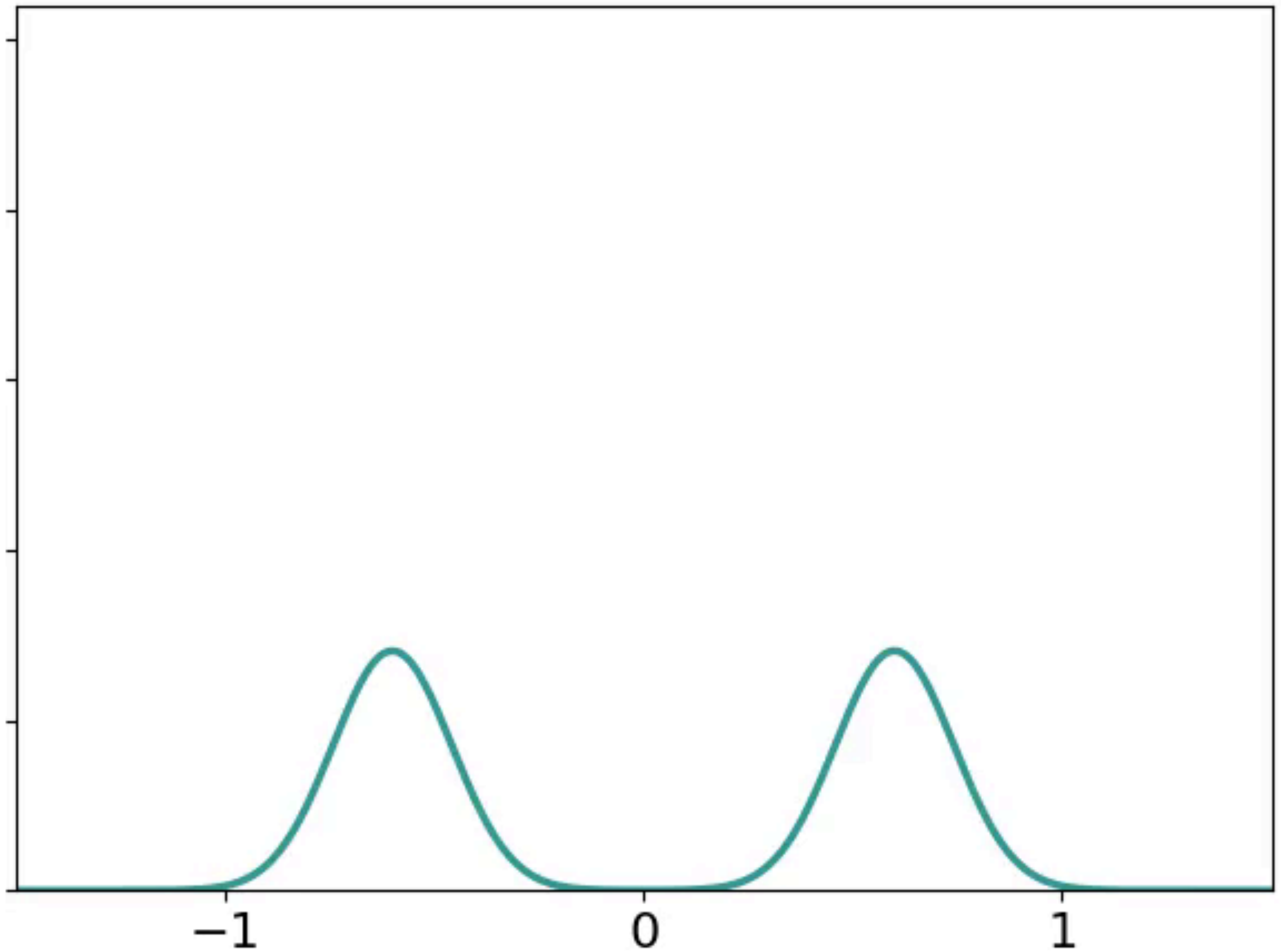
aggregation, drift, and degenerate diffusion:

$$\frac{d}{dt}\rho = \nabla \cdot ((\nabla K * \rho)\rho)$$

$K(x) = |x|$



$K(x) = |x|^3/3$

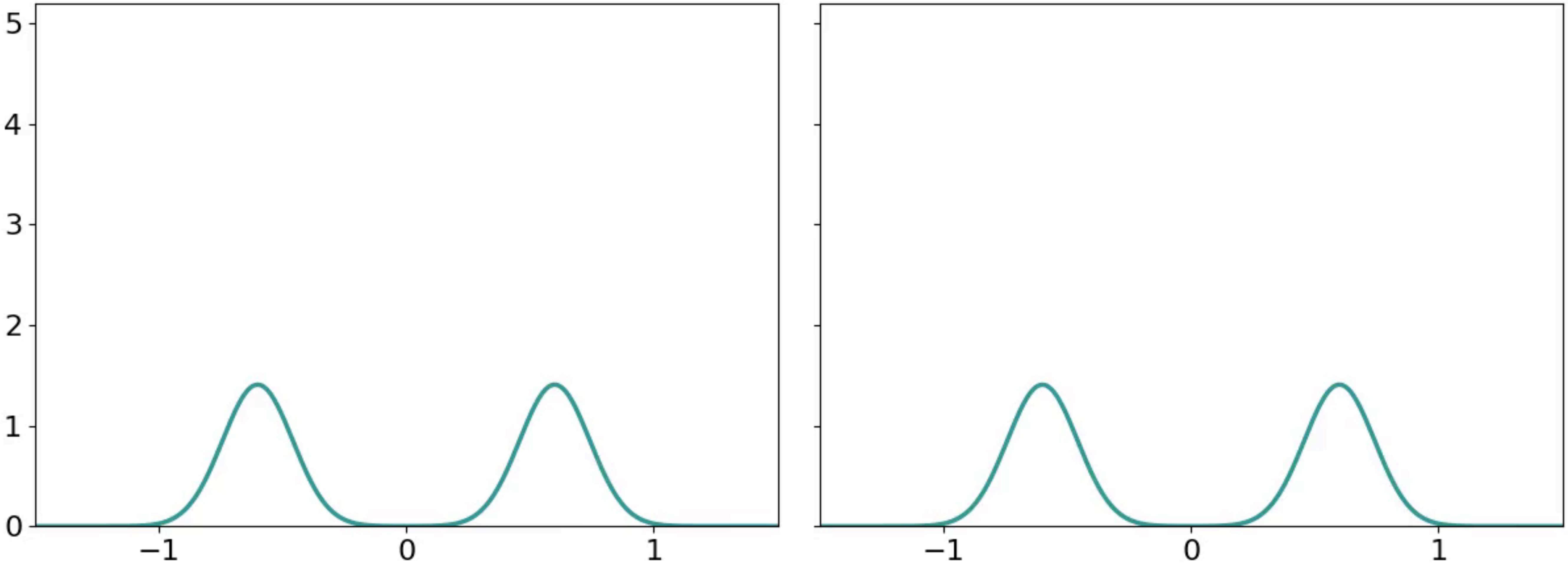


gradient flow in the Wasserstein metric

aggregation, drift, and degenerate diffusion:

$$\frac{d}{dt}\rho = \nabla \cdot ((\nabla K * \rho)\rho)$$

$K(x) = |x|^3/3$

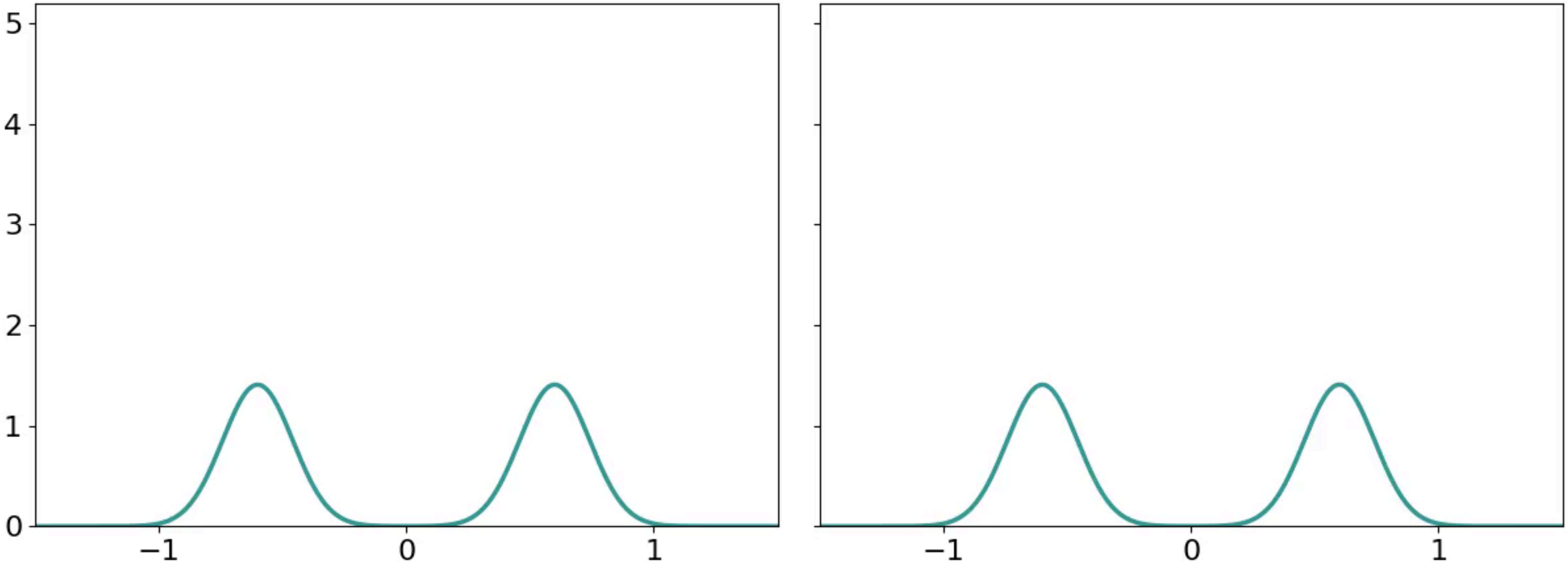


gradient flow in the Wasserstein metric

aggregation, drift, and degenerate diffusion:

$$\frac{d}{dt}\rho = \nabla \cdot ((\nabla K * \rho)\rho)$$

$K(x) = |x|^3/3$

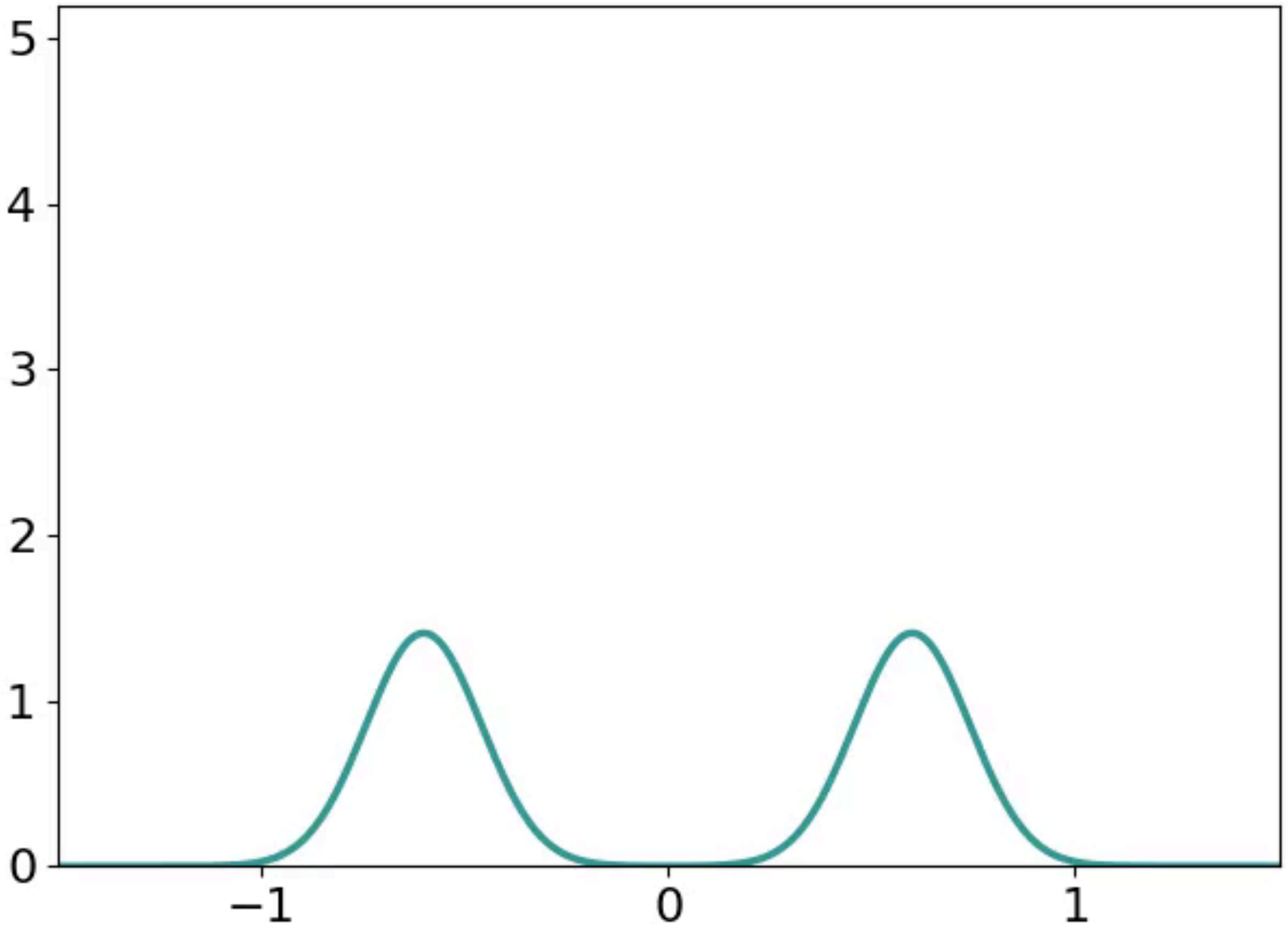


gradient flow in the Wasserstein metric

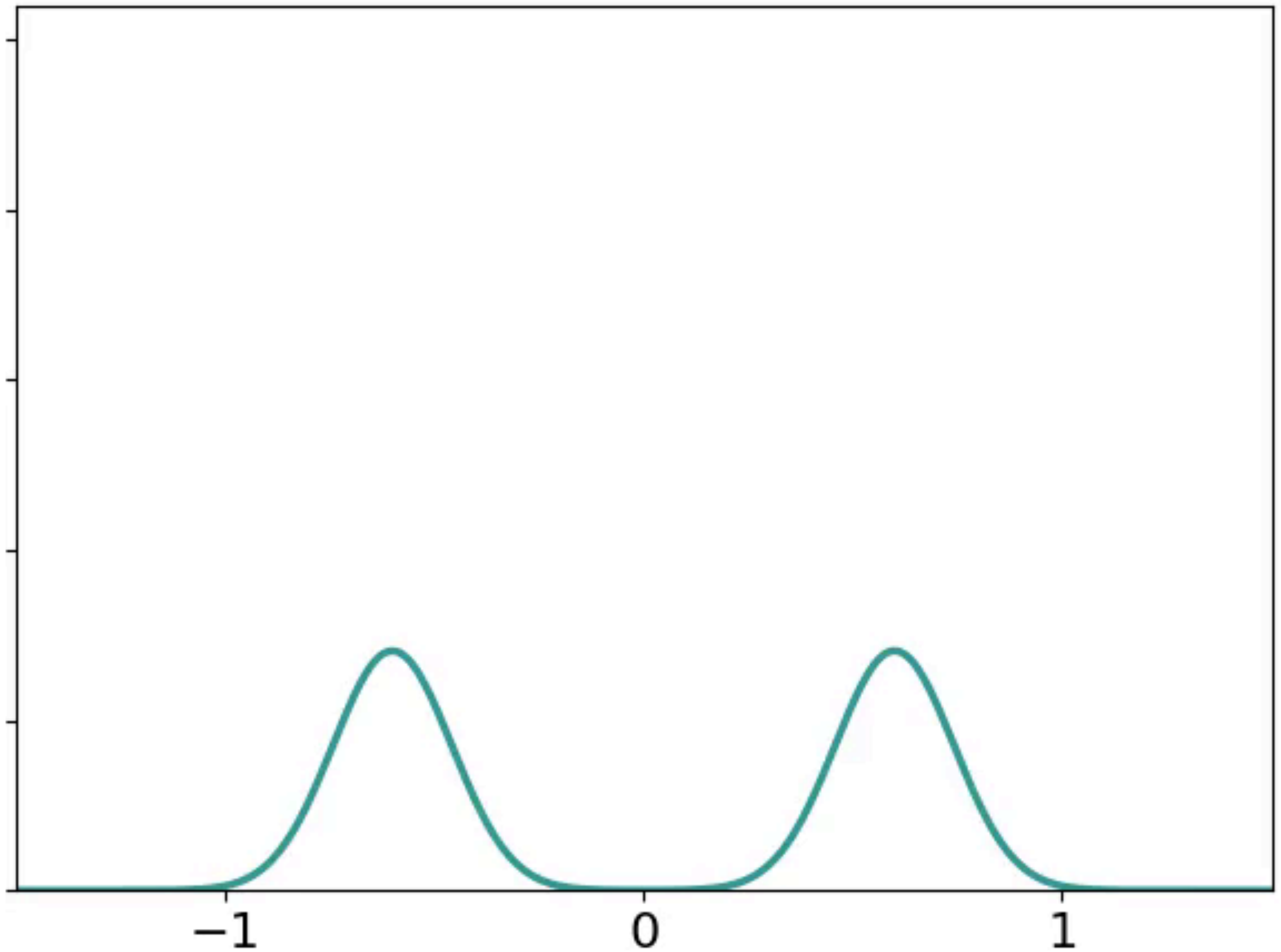
aggregation, drift, and degenerate diffusion:

$$\frac{d}{dt}\rho = \nabla \cdot ((\nabla K * \rho)\rho)$$

$K(x) = |x|^3/3 - |x|$



$K(x) = |x|^3/3$

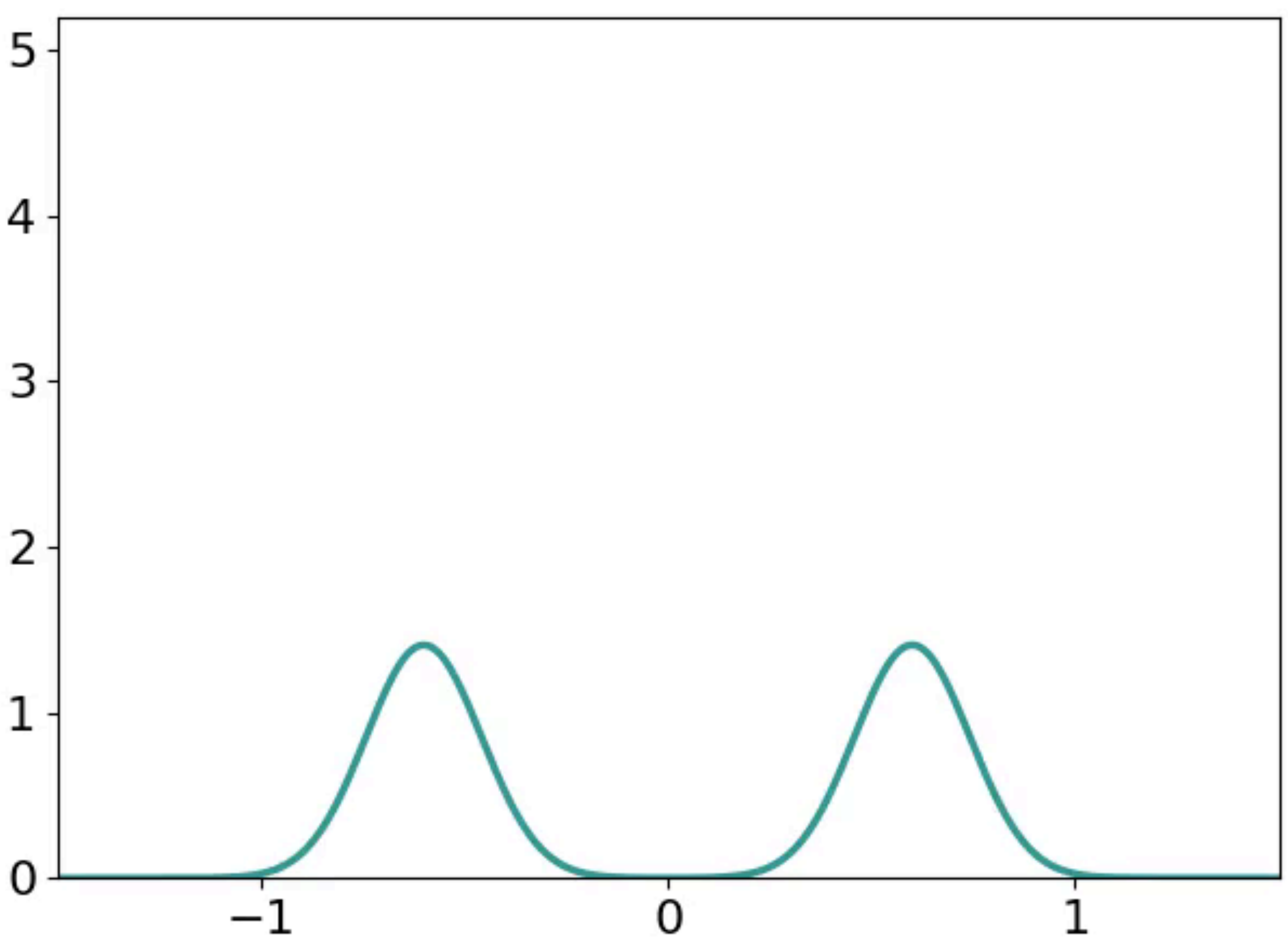


gradient flow in the Wasserstein metric

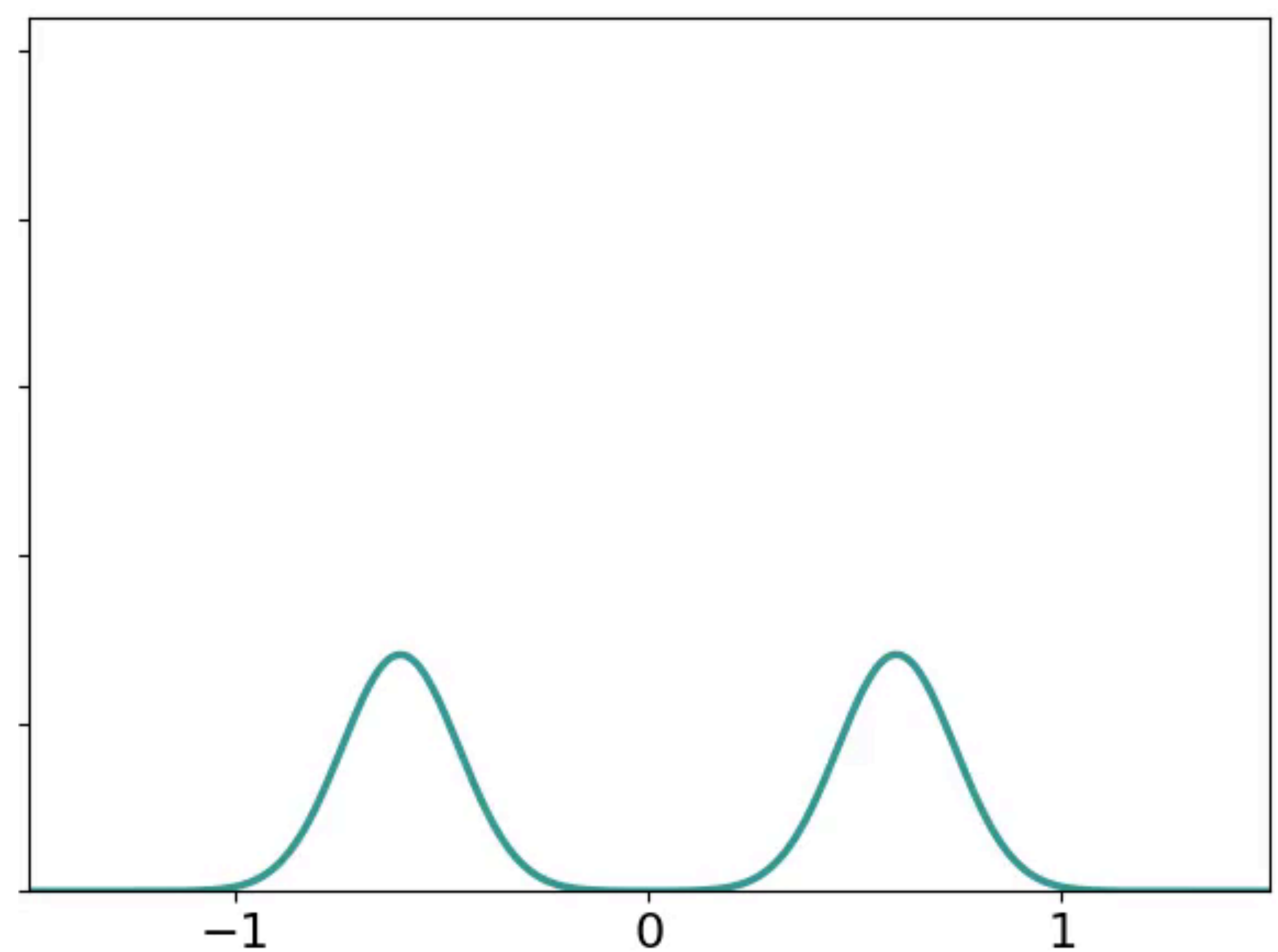
aggregation, drift, and degenerate diffusion:

$$\frac{d}{dt}\rho = \nabla \cdot ((\nabla K * \rho)\rho)$$

$K(x) = |x|^3/3 - |x|$



$K(x) = |x|^3/3$

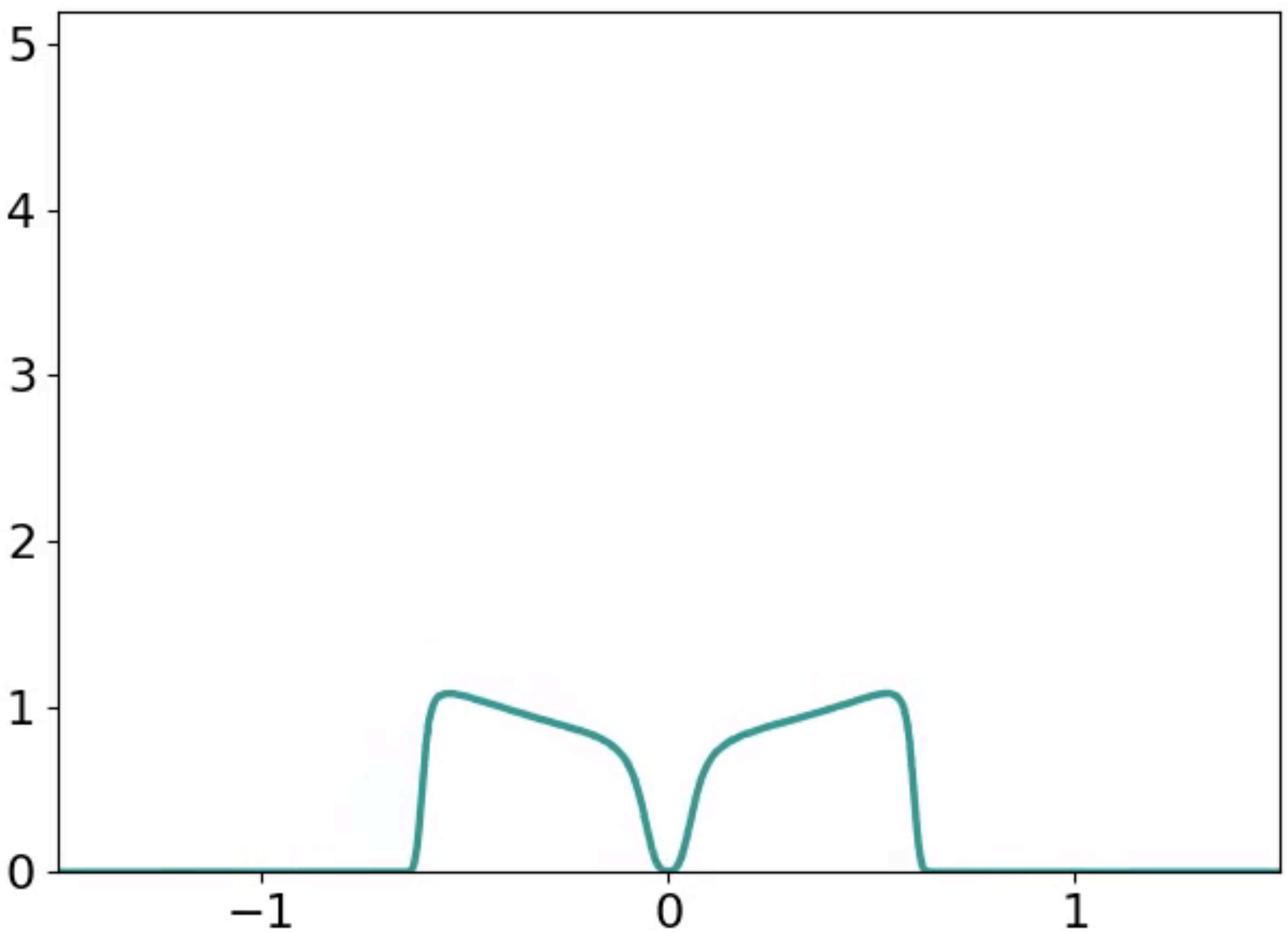


gradient flow in the Wasserstein metric

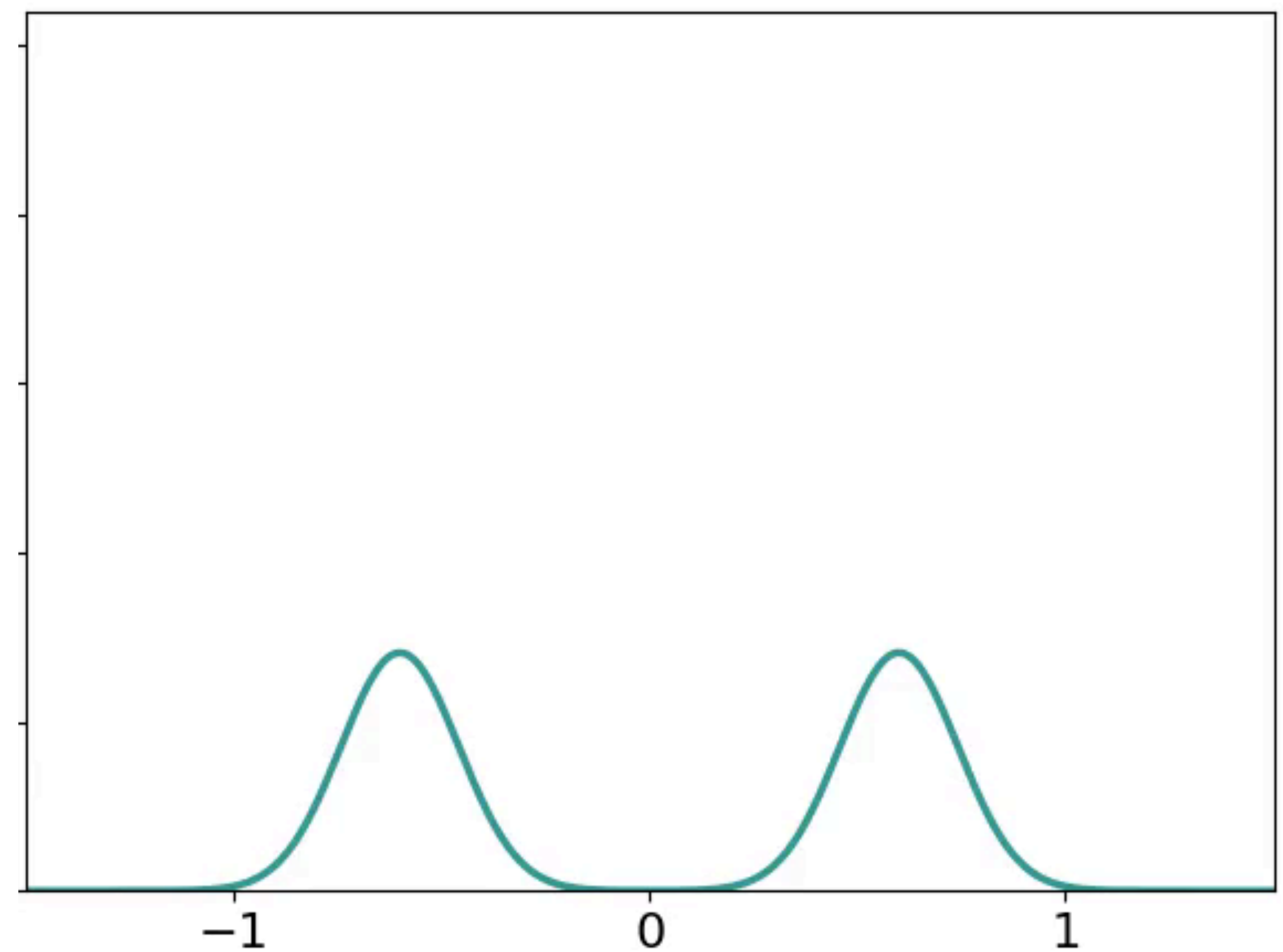
aggregation, drift, and degenerate diffusion:

$$\frac{d}{dt}\rho = \nabla \cdot ((\nabla K * \rho)\rho) + \nabla \cdot (\nabla V \rho)$$

$K(x) = |x|^3/3 - |x|$



$K(x) = |x|^3/3 - |x|$, $V(x) = -x$

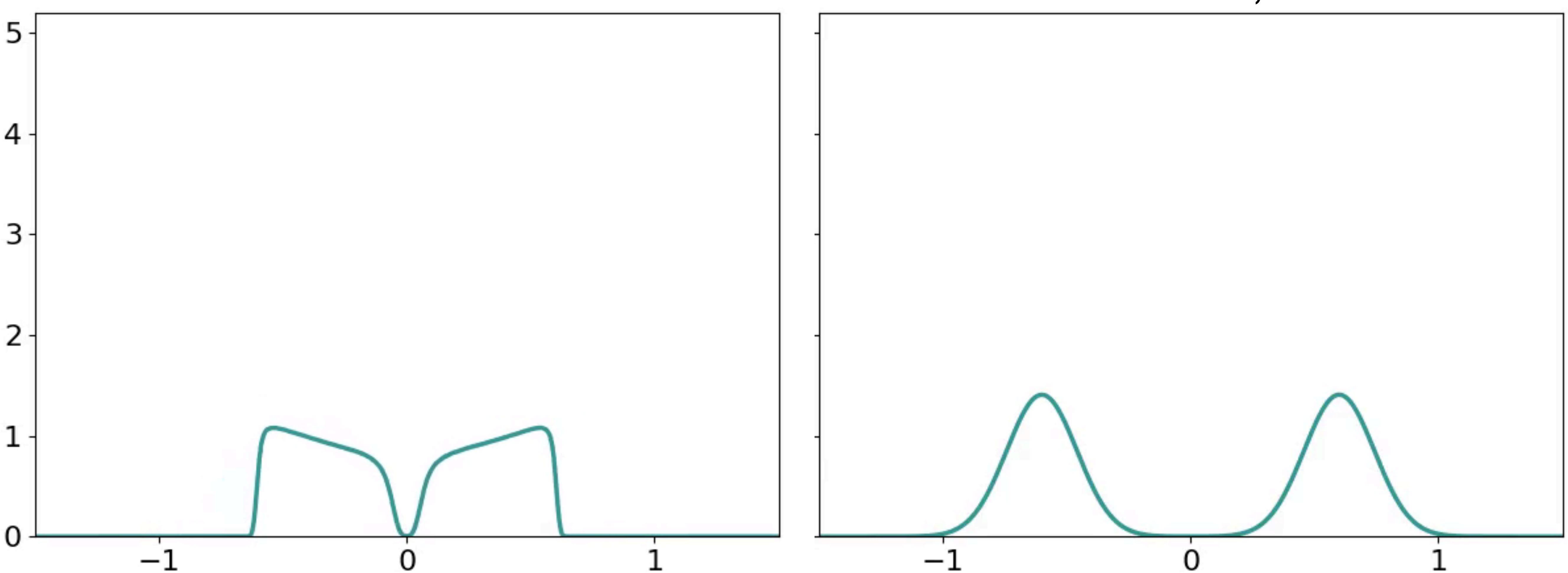


gradient flow in the Wasserstein metric

aggregation, drift, and degenerate diffusion:

$$\frac{d}{dt}\rho = \nabla \cdot ((\nabla K * \rho)\rho) + \nabla \cdot (\nabla V \rho)$$

$K(x) = |x|^3/3 - |x|$ $\qquad$ $K(x) = |x|^3/3 - |x|, \quad V(x) = -x$

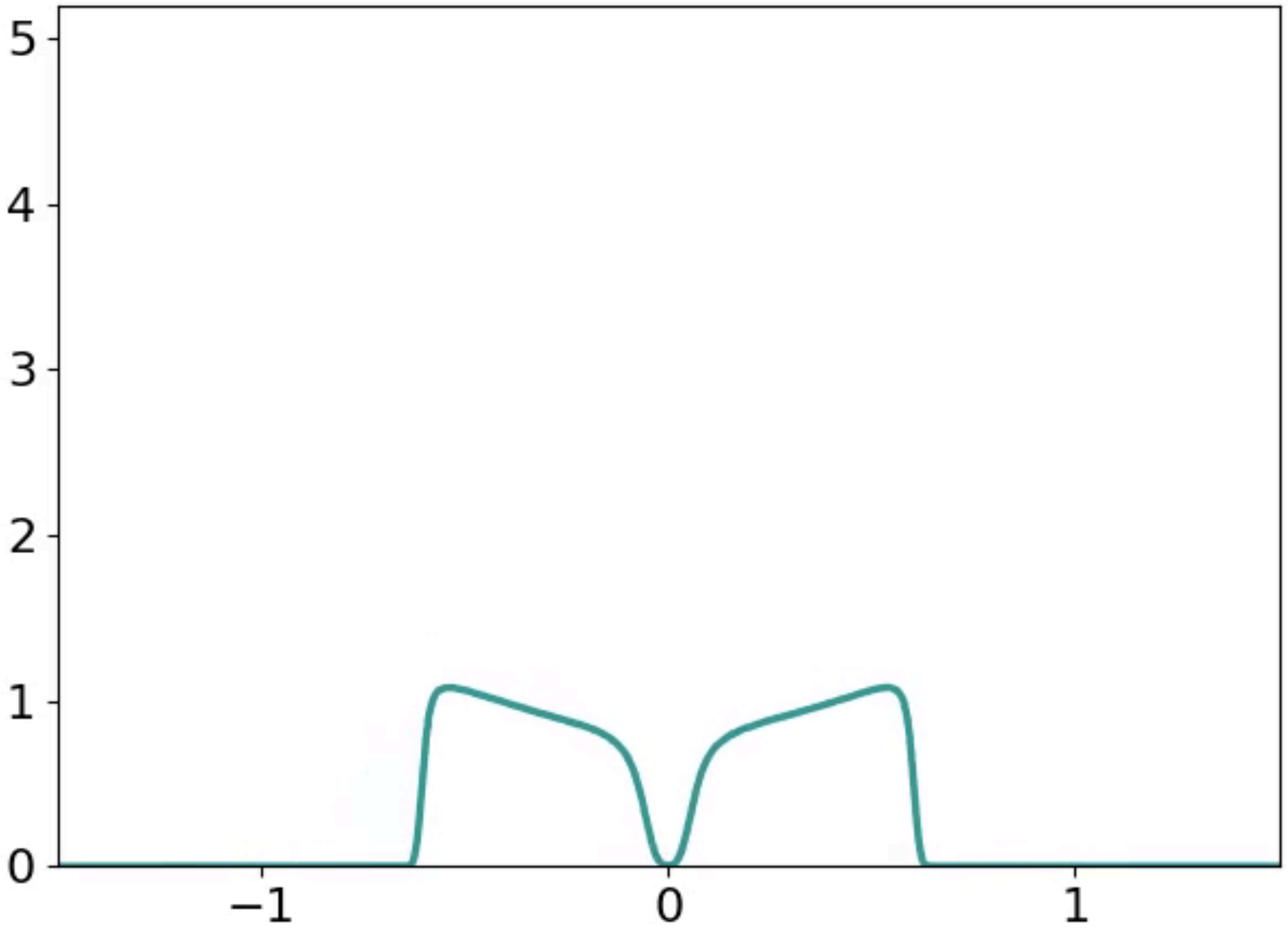


gradient flow in the Wasserstein metric

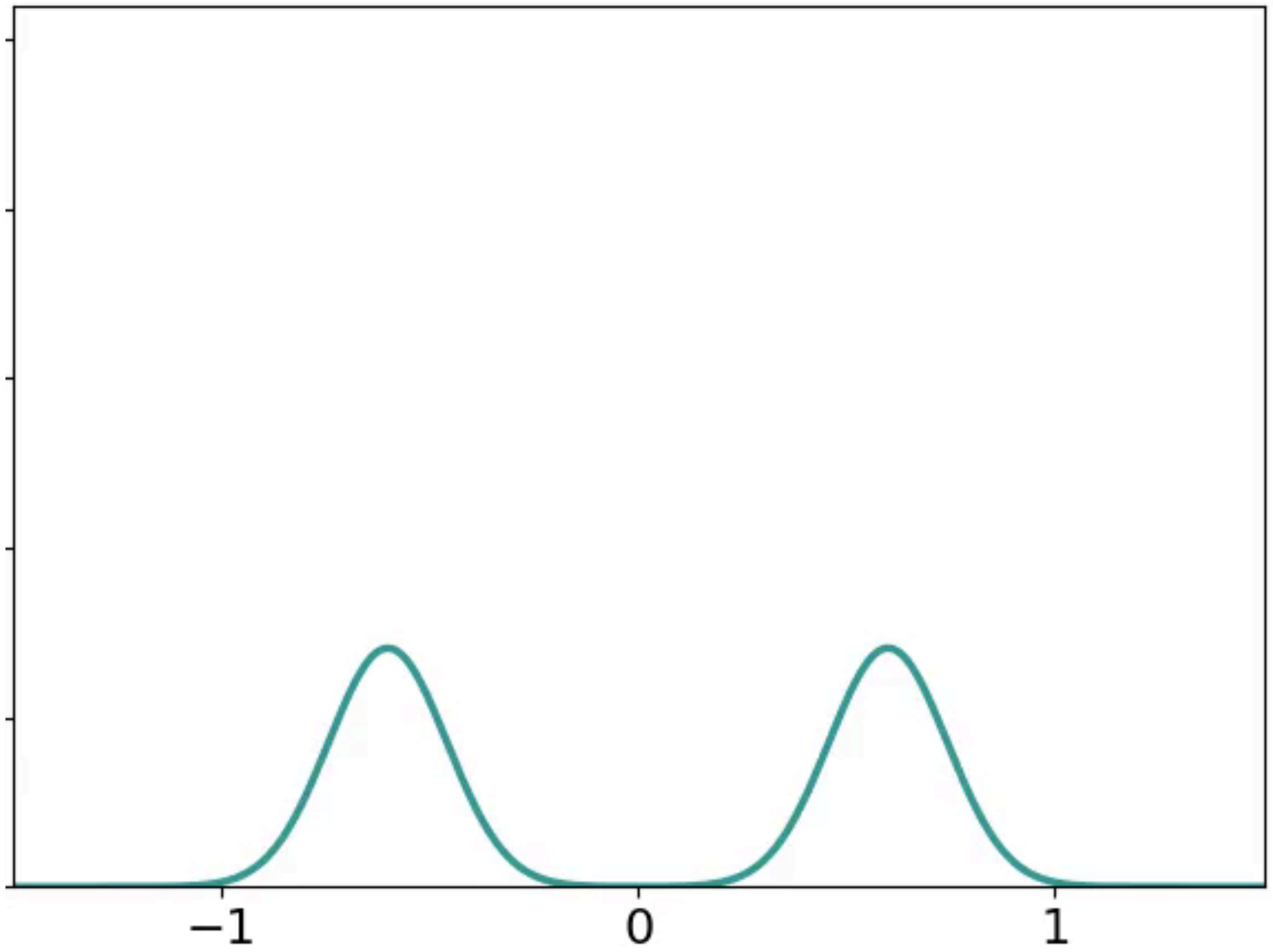
aggregation, drift, and degenerate diffusion:

$$\frac{d}{dt}\rho = \nabla \cdot ((\nabla K * \rho)\rho) + \nabla \cdot (\nabla V \rho)$$

$K(x) = |x|^3/3 - |x|$



$K(x) = |x|^3/3 - |x|, \quad V(x) = -x$

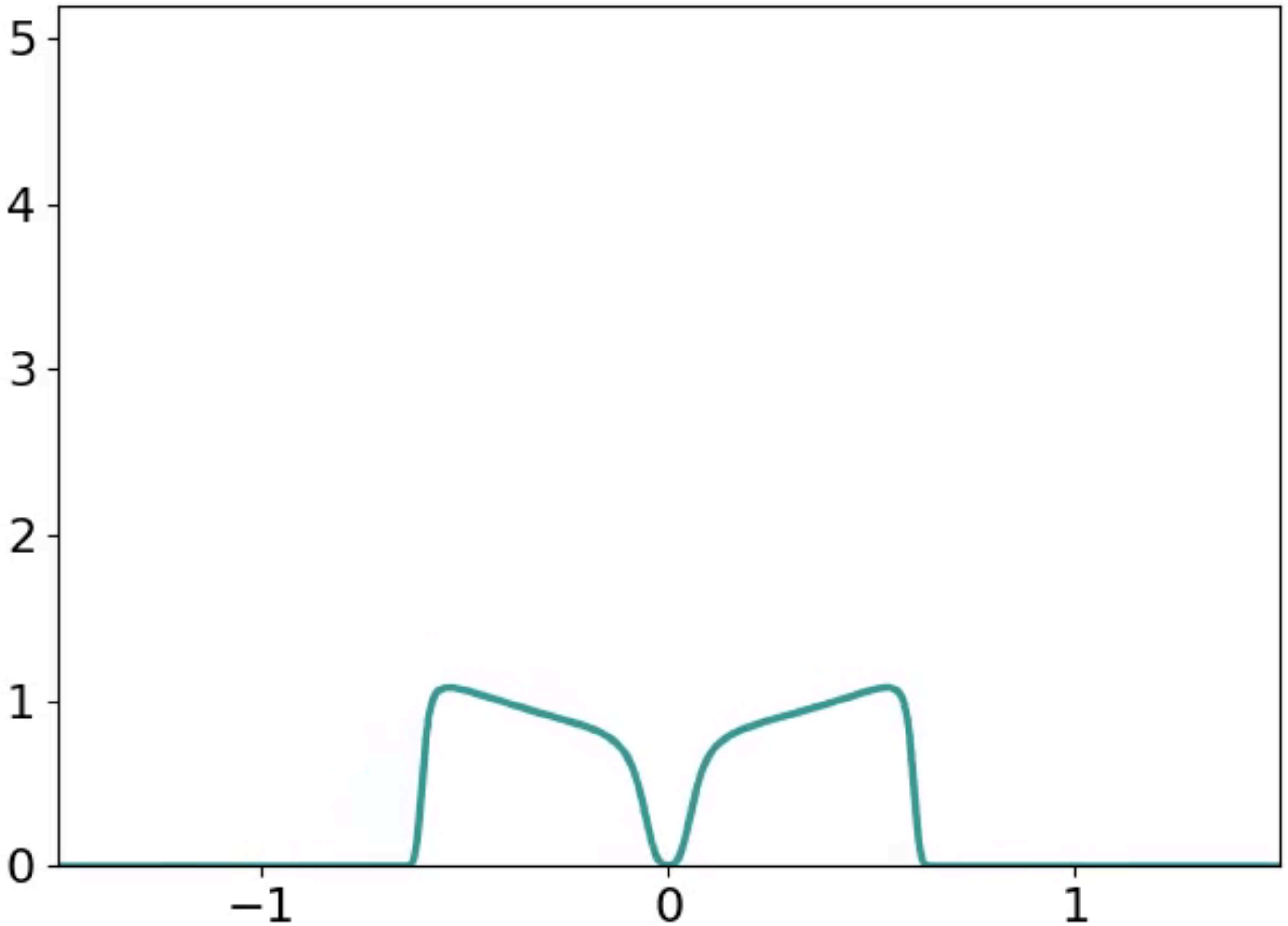


gradient flow in the Wasserstein metric

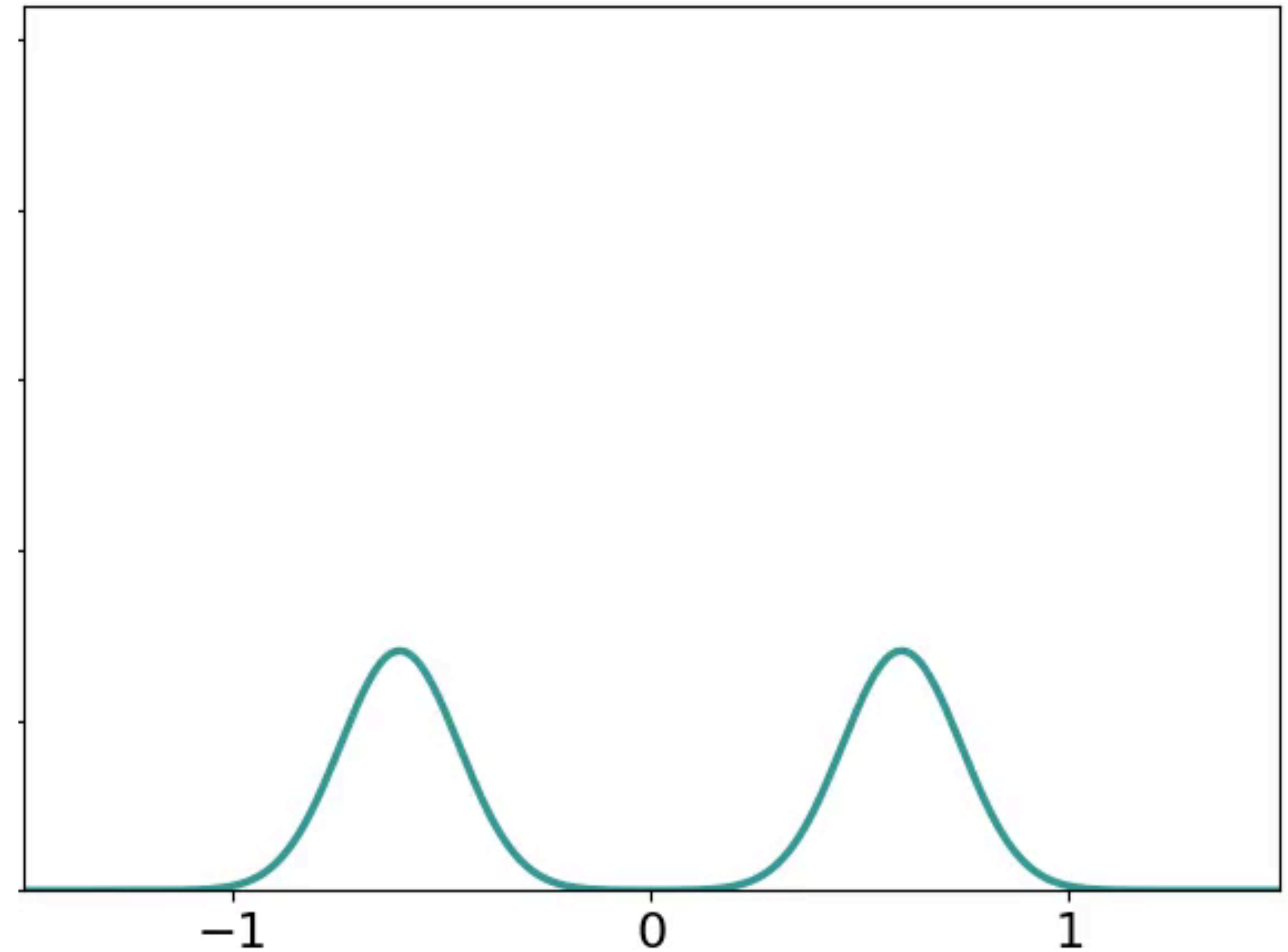
aggregation, drift, and degenerate diffusion:

$$\frac{d}{dt}\rho = \nabla \cdot ((\nabla K * \rho)\rho) \qquad + \Delta \rho^m$$

$K(x) = |x|^3/3 - |x|$



$K(x) = |x|^3/3 - |x|$, $m = 1$

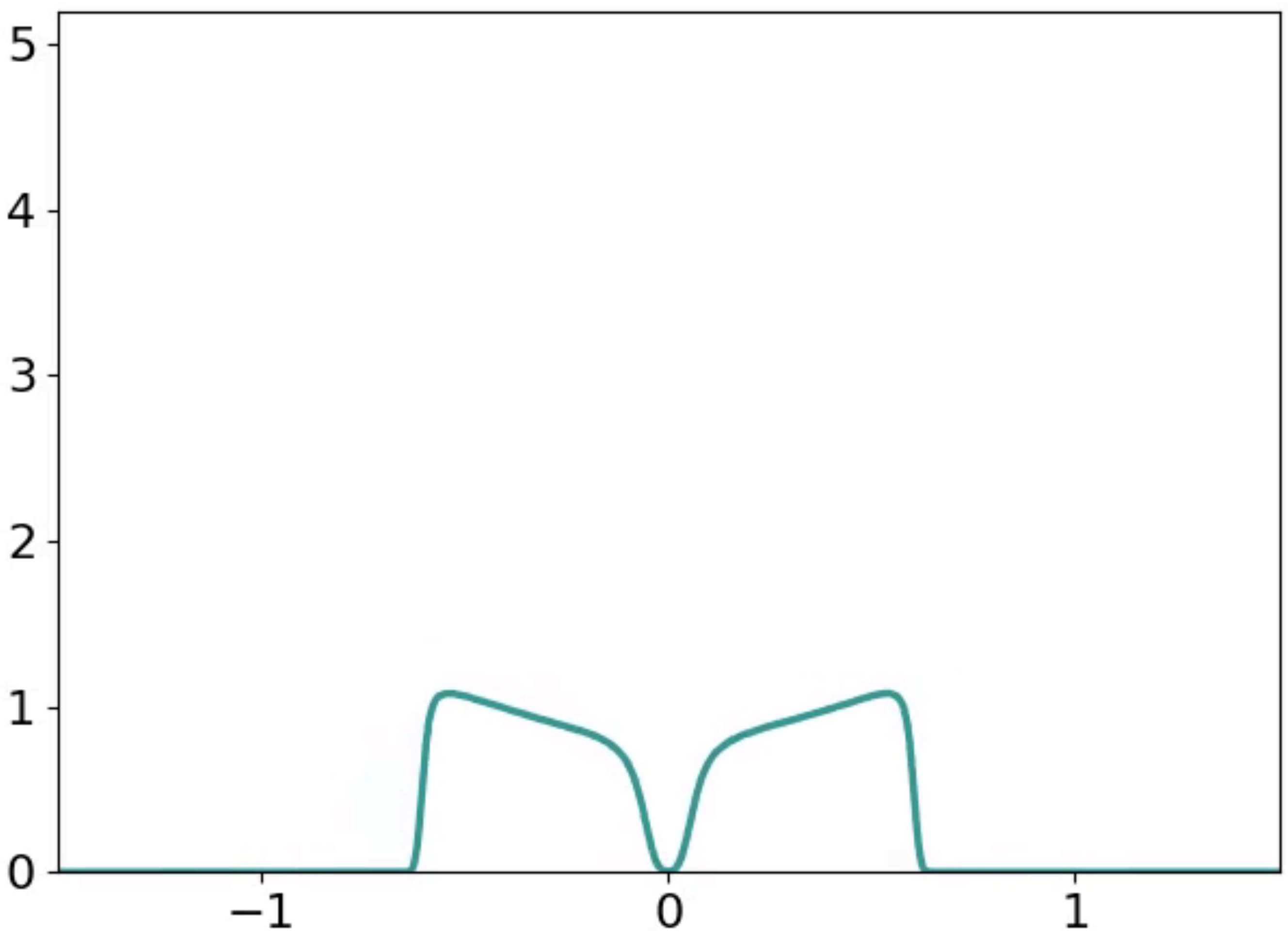


gradient flow in the Wasserstein metric

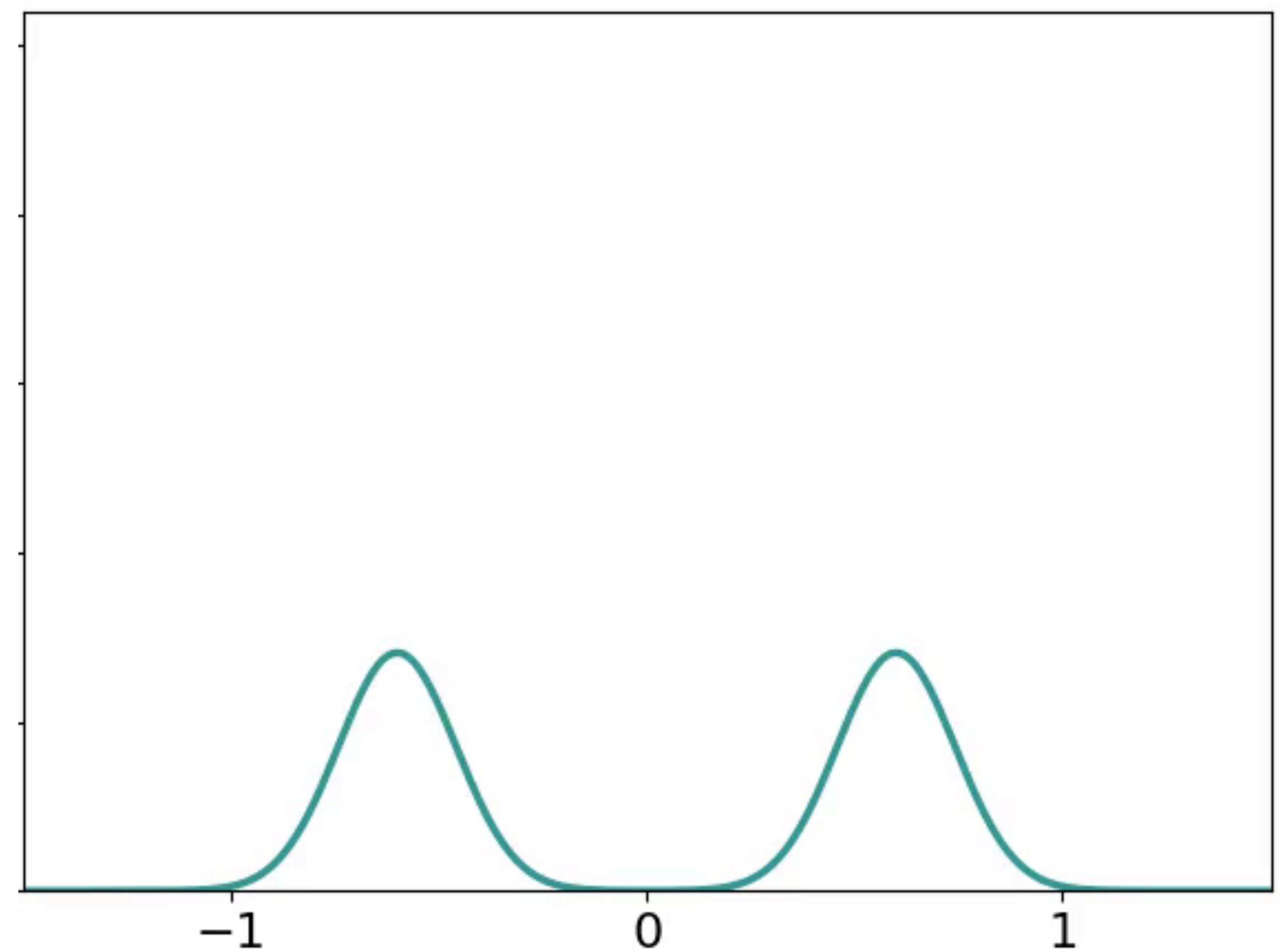
aggregation, drift, and degenerate diffusion:

$$\frac{d}{dt}\rho = \nabla \cdot ((\nabla K * \rho)\rho) \qquad + \Delta \rho^m$$

$K(x) = |x|^3/3 - |x|$



$K(x) = |x|^3/3 - |x|$, $m = 1$

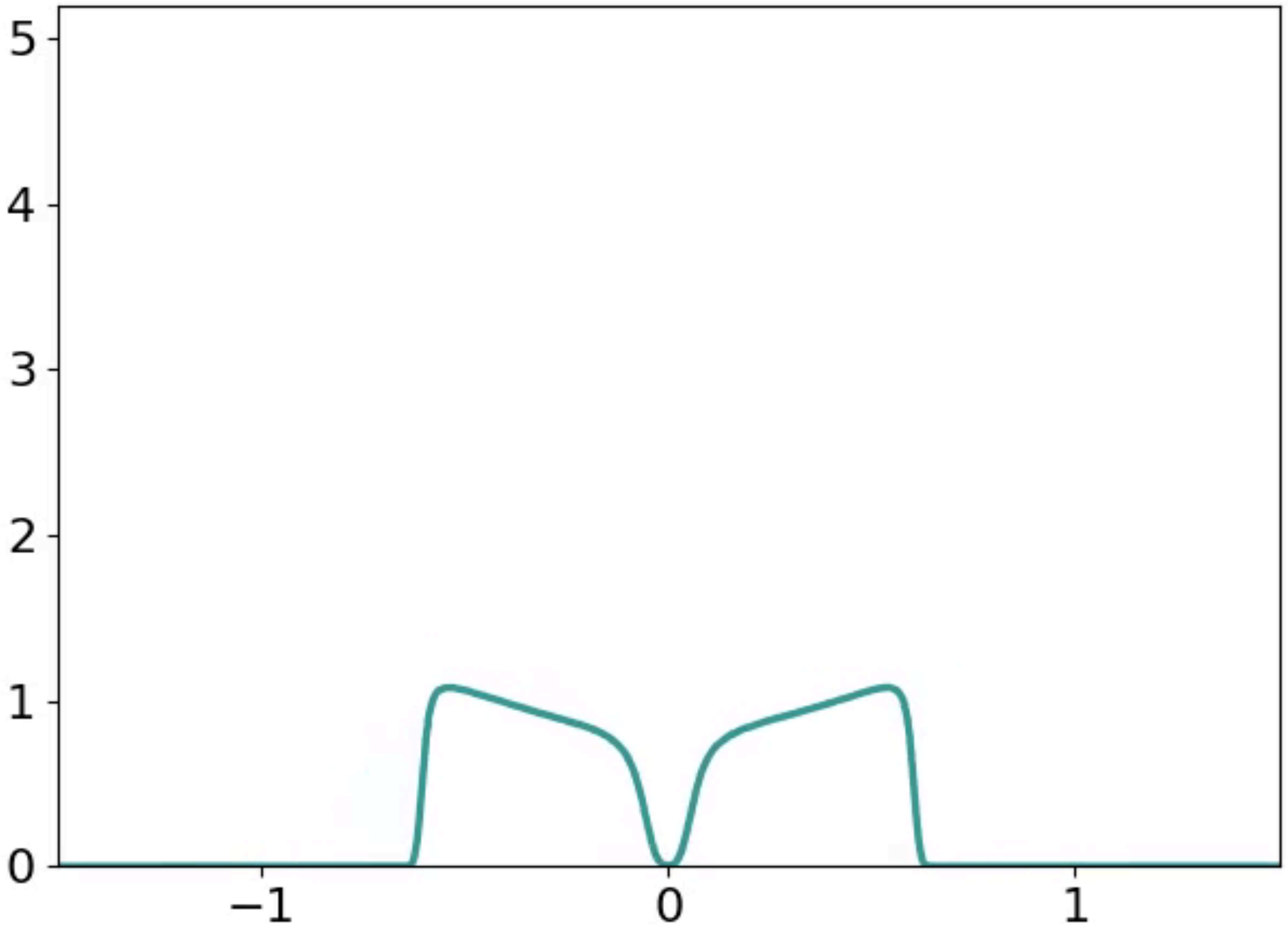


gradient flow in the Wasserstein metric

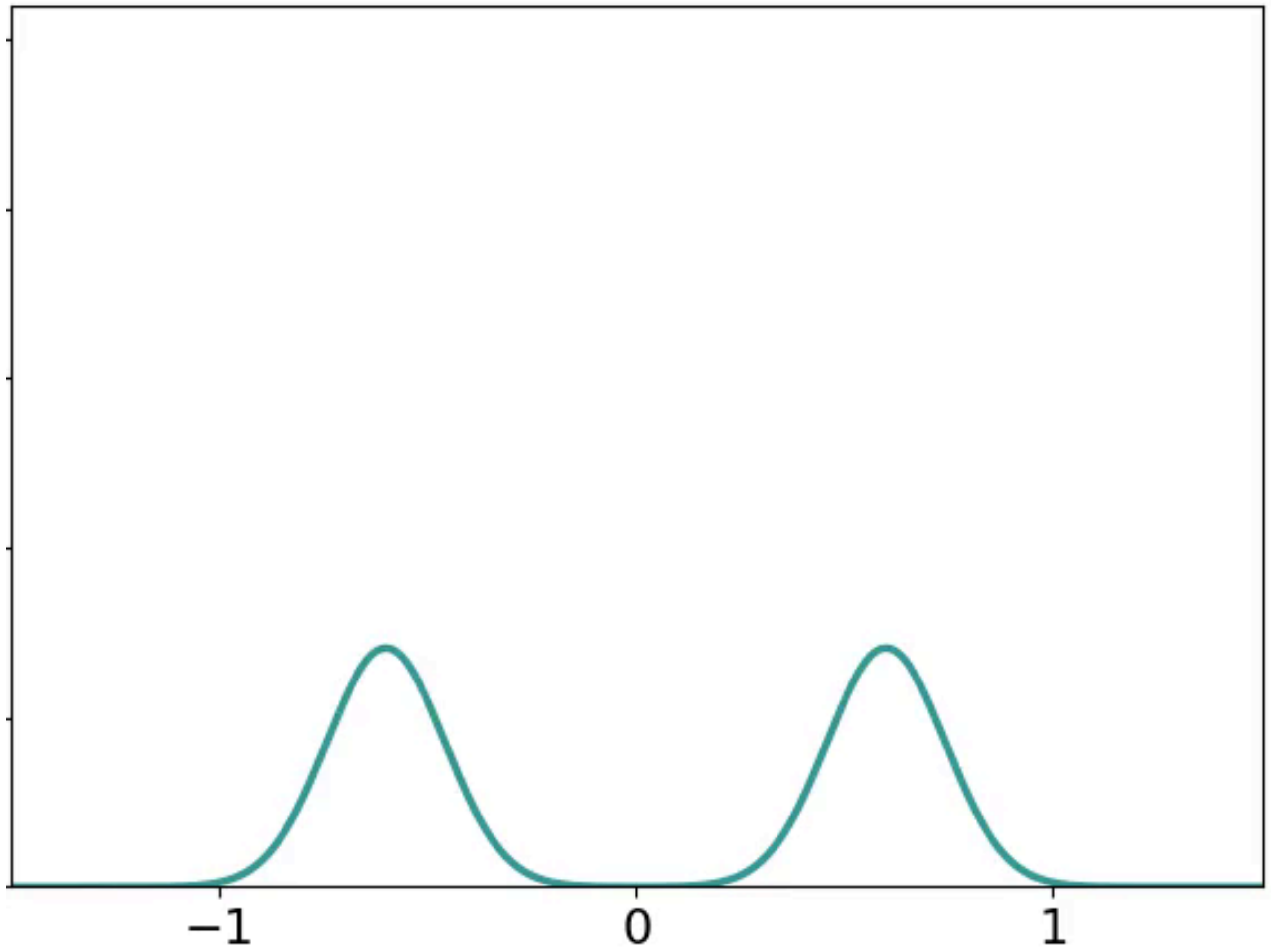
aggregation, drift, and degenerate diffusion:

$$\frac{d}{dt}\rho = \nabla \cdot ((\nabla K * \rho)\rho) \qquad + \Delta \rho^m$$

$K(x) = |x|^3/3 - |x|$



$K(x) = |x|^3/3 - |x|$, $m = 1$



numerical simulation of W_2 grad flow

JKO scheme

- Pros: reduces simulating grad flow to solving a sequence of optimization problems involving W_2 distance; leverages state of the art W_2 solvers.
- Cons: current methods lose convexity/stability properties of gradient flow

Finite volume methods, finite element methods…

- Pros: adapts numerical approaches inspired by classical fluid mechanics to gradient flow setting
- Cons: current methods lose convexity/stability properties of gradient flow

Without stability, we can't prove general results on convergence.

Particle methods

particle methods

Goal: Approximate a solution to $\frac{d}{dt}\rho(x,t) + \nabla \cdot (v(x,t)\rho(x,t)) = 0$

Assume: v(x,t) comes from a Wasserstein gradient flow and is "nice"

A general recipe for a particle methods:

(1) approximate $\rho_0(x)$ as a sum of Dirac masses on a grid of spacing h

$$\rho_0 \approx \sum_{i=1}^{N} \delta_{x_i} m_i$$

(2) evolve the locations of the Dirac masses by

$$\frac{d}{dt}x_i(t) = v(x_i(t),t) \quad \forall i$$

(3) $\rho_N(x,t) = \sum_{i=1}^{N} \delta_{x_i(t)} m_i$ is a gradient flow of the original energy; it inherits all convexity/stability properties, hence $\rho_N(x,t) \to \rho(x,t)$

particle methods

Goal: Approximate a solution to $$\frac{d}{dt}\rho(x,t) + \nabla \cdot (v(x,t)\rho(x,t)) = 0$$

Benefits of particle methods:

(1) positivity preserving

(2) inherently adaptive

(3) energy decreasing

(4) preserves convexity/stability properties of gradient flow at discrete level

but what about when v(x,t) is not “nice”?

v(x,t) is often not nice

aggregation, drift, and degenerate diffusion:

$$\frac{d}{dt}\rho = \nabla \cdot ((\nabla K * \rho)\rho) + \nabla \cdot (\nabla V \rho) + \Delta \rho^m$$

$K, V : \mathbb{R}^d \to \mathbb{R}$, and $m \geq 1$

$$v = \nabla K * \rho + \nabla V + m\rho^{m-2}\nabla \rho$$

- Diffusion term is worst: particles do not remain particles
- But even the interaction term can slow down convergence if it has a strong singularity at the origin…

v(x,t) is often not nice

aggregation, drift, and degenerate diffusion:

$$\frac{d}{dt}\rho = \nabla \cdot ((\nabla K * \rho)\rho) + \nabla \cdot (\nabla V \rho) + \Delta \rho^m$$

$K, V : \mathbb{R}^d \to \mathbb{R}, \text{ and } m \geq 1$

$$v = \nabla K * \rho + \nabla V + m\rho^{m-2}\nabla\rho$$

- Diffusion term is worst: particles do not remain particles
- But even the interaction term can slow down convergence if it has a strong singularity at the origin…

"Blob Method": regularize the velocity field to make it nice!

v(x,t) is often not nice

aggregation, drift, and degenerate diffusion:

$$\frac{d}{dt}\rho = \nabla \cdot ((\nabla K * \rho)\rho)$$

$K, V : \mathbb{R}^d \to \mathbb{R}, \text{ and } m \geq 1$

$$v = \nabla K * \rho$$

- Diffusion term is worst: particles do not remain particles
- But even the interaction term can slow down convergence if it has a strong singularity at the origin…

"Blob Method": regularize the velocity field to make it nice!

a blob method for aggregation

Goal: Approximate a solution to $\frac{d}{dt}\rho = \nabla \cdot ((\nabla K * \rho)\rho)$ $\quad v = \nabla K * \rho$

A <u>regularized</u> particle method:

(0) regularize the interaction kernel via convolution with a mollifier

$$K_\epsilon(x) = K * \varphi_\epsilon(x), \quad \varphi_\epsilon(x) = \varphi(x/\epsilon)/\epsilon^d$$

(1) approximate $\rho_0(x)$ as a sum of Dirac masses on a grid of spacing h

$$\rho_0 \approx \sum_{i=1}^N \delta_{x_i} m_i$$

(2) evolve the locations of the Dirac masses by

$$\frac{d}{dt} x_i(t) = -\sum_{j=1}^N \nabla K_\epsilon(x_i(t) - x_j(t)) m_j$$

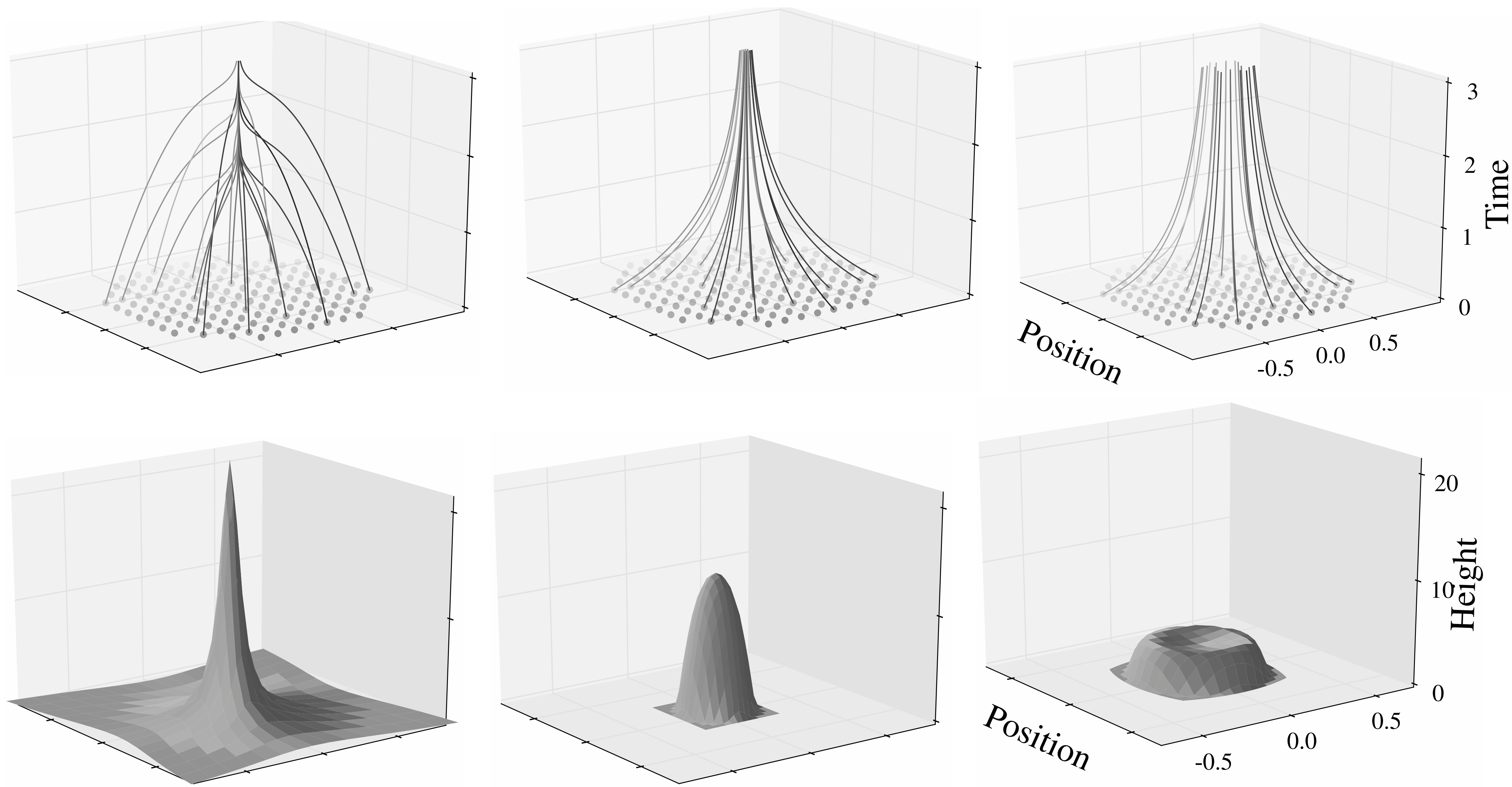
(3) $\rho_N(x,t) = \sum_{i=1}^N \delta_{x_i(t)} m_i$ is a gradient flow of the regularized energy

$$E_\epsilon(\rho) = \int (K_\epsilon * \rho)\rho$$

a blob method for aggregation

Theorem [C., Bertozzi 2014]: If ε = h^q, 0<q<1, the blob method converges as h →0 .

$K(x) = \log|x|/2\pi$ $\quad K(x) = |x|^2/2$ $\quad K(x) = |x|^3/3$



$h = 0.04,\ q = 0.9,\ m = 4$

aggregation + ?

aggregation, drift, and diffusion:

$$\frac{d}{dt}\rho = \nabla \cdot ((\nabla K * \rho)\rho) + \nabla \cdot (\nabla V \rho) + \Delta \rho^m \quad K, V : \mathbb{R}^d \to \mathbb{R}, \text{ and } m \geq 1$$

- Adding a drift term is straightforward: just do a particle method with
$$v = \nabla K_\epsilon * \rho + \nabla V$$
- How can we add diffusion?
- Previous work: **stochastic** [Liu, Yang 2017], [Huang, Liu 2015], **deterministic** [Carrillo, Huang, Patacchini, Wolansky 2016]
- Our idea: regularize by convolution with a mollifier.

a blob method for degenerate diffusion

diffusion equation:

$$\frac{d}{dt}\rho = \Delta\rho^m \qquad m \geq 1$$

Solutions of diffusion equation are gradient flows of $E(\rho) = \frac{1}{m-1}\int \rho^m$

Let's consider gradient flows of $E_\epsilon(\rho) = \frac{1}{m-1}\int (\rho * \varphi_\epsilon)^{m-1}\rho$

- Previous work (m=2):[Lions, Mas-Gallic 2000], [P.E. Jabin, in progress]
- For ε > 0, particles remain particles, so can do a particle method for

$$v = \nabla\varphi_\epsilon * \left((\varphi_\epsilon * \rho)^{m-2}\rho\right) + (\varphi_\epsilon * \rho)^{m-2}(\nabla\varphi_\epsilon * \rho)$$

$\Longrightarrow$ a blob method for diffusion.

a blob method for degenerate diffusion

diffusion equation:

$$\frac{d}{dt}\rho = \Delta\rho^m \qquad m \geq 1$$

Solutions of diffusion equation are gradient flows of

Let's consider gradient flows of $E_\epsilon(\rho) = \frac{1}{m-1}\int (\rho * \varphi_\epsilon)^{m-1}\rho$

- Previous work (m=2):[Lions, Mas-Gallic 2000], [P.E. Jabin, in progress]
- For ε > 0, particles remain particles, so can do a particle method for

$$v = \nabla\varphi_\epsilon * \left((\varphi_\epsilon * \rho)^{m-2}\rho\right) + (\varphi_\epsilon * \rho)^{m-2}(\nabla\varphi_\epsilon * \rho)$$

$\Longrightarrow$ a blob method for diffusion.

a blob method for degenerate diffusion

diffusion equation:

$$\frac{d}{dt}\rho = \Delta\rho^m \qquad m \geq 1$$

Solutions of diffusion equation are gradient flows of

Let's consider gradient flows of

- Previous work (m=2):[Lions, Mas-Gallic 2000], [P.E. Jabin, in progress]
- For ε > 0, particles remain particles, so can do a particle method for

$$v = \nabla\varphi_\epsilon * \left((\varphi_\epsilon * \rho)^{m-2}\rho\right) + (\varphi_\epsilon * \rho)^{m-2}(\nabla\varphi_\epsilon * \rho)$$

$\Longrightarrow$ a blob method for diffusion.

a blob method for degenerate diffusion

diffusion equation:

$$\frac{d}{dt}\rho = \Delta\rho^m \quad m \geq 1$$

Solutions of diffusion equation are gradient flows of $E(\rho) = \int \log(\rho)\rho$

Let's consider gradient flows of

- Previous work (m=2):[Lions, Mas-Gallic 2000], [P.E. Jabin, in progress]
- For ε > 0, particles remain particles, so can do a particle method for

$$v = \nabla\varphi_\epsilon * \left((\varphi_\epsilon * \rho)^{m-2}\rho\right) + (\varphi_\epsilon * \rho)^{m-2}(\nabla\varphi_\epsilon * \rho)$$

$\implies$ a blob method for diffusion.

a blob method for degenerate diffusion

diffusion equation:

$$\frac{d}{dt}\rho = \Delta\rho^m \quad m \geq 1$$

Solutions of diffusion equation are gradient flows of $E(\rho) = \int \log(\rho)\rho$

Let's consider gradient flows of $E_\epsilon(\rho) = \int \log(\rho * \varphi_\epsilon)\rho$

- Previous work (m=2):[Lions, Mas-Gallic 2000], [P.E. Jabin, in progress]
- For ε > 0, particles remain particles, so can do a particle method for

$$v = \nabla\varphi_\epsilon * \left((\varphi_\epsilon * \rho)^{m-2}\rho\right) + (\varphi_\epsilon * \rho)^{m-2}(\nabla\varphi_\epsilon * \rho)$$

$\implies$ a blob method for diffusion.

a blob method for degenerate diffusion

Theorem [Carrillo, C., Patacchini 2017]: Consider

$$E_\epsilon(\rho) = \int (K * \rho)\rho + \int V\rho + \frac{1}{m-1}\int (\rho * \varphi_\epsilon)^{m-1}\rho$$

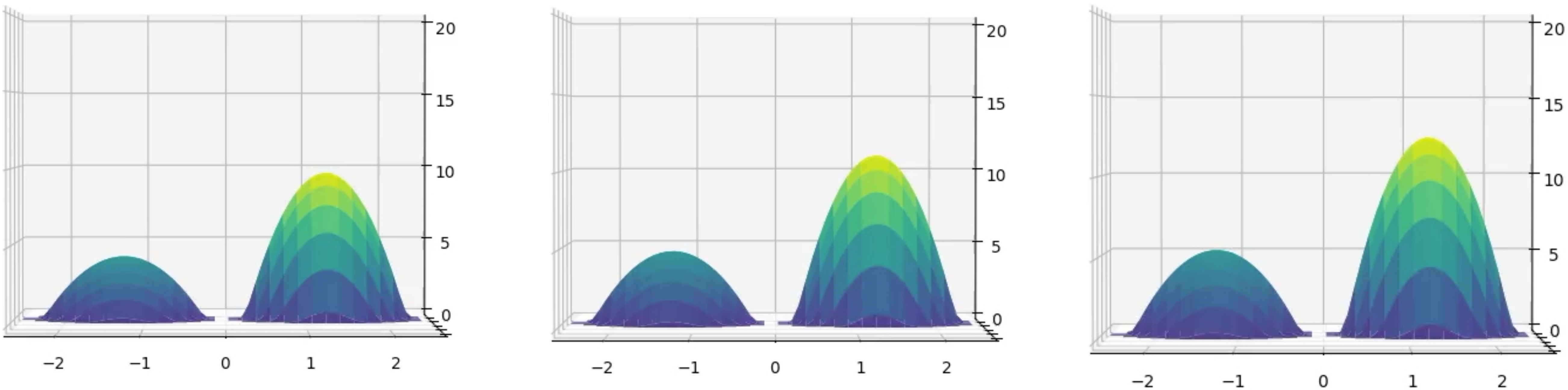
As $\varepsilon \to 0$,

- For all $m \geq 1$, E_ε Γ-converge to E.
- For $m = 2$ and initial data with bounded entropy, gradient flows of E_ε converge to gradient flows of E.
- For $m \geq 2$ and particle initial data with $\varepsilon = h^q$, $0<q<1$, if a priori estimates hold, gradient flows of E_ε converge to gradient flows of E.

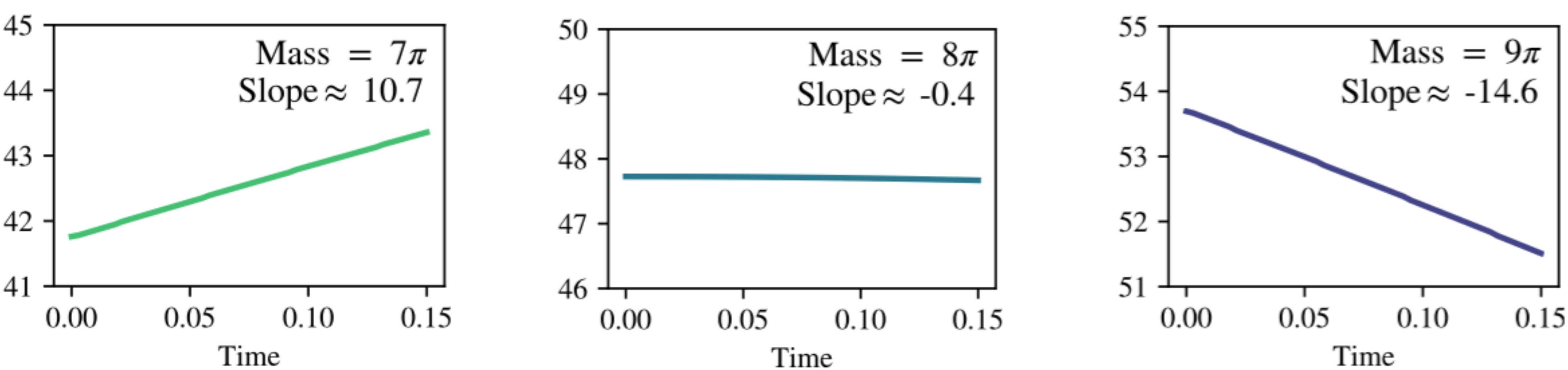
numerics: Keller-Segel (d=2)

subcritical mass	**critical mass**	**supercritical mass**

Evolution of Density



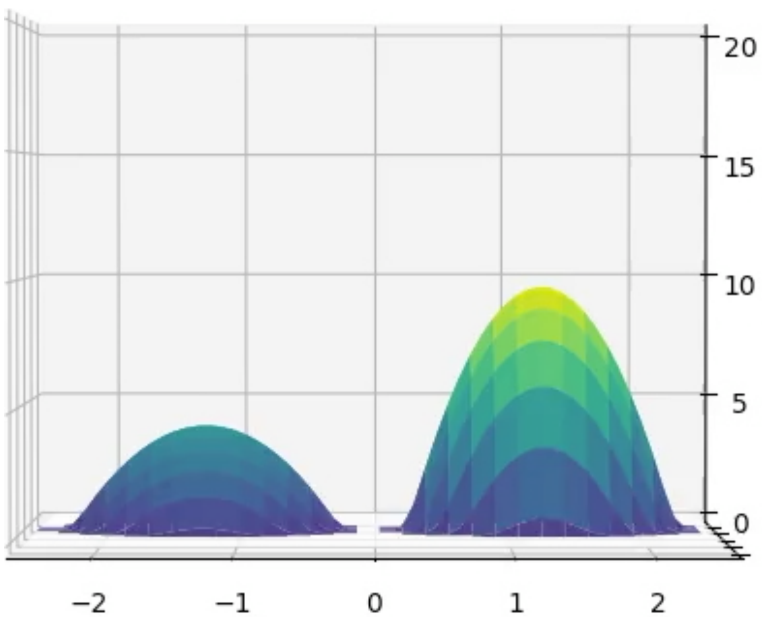
Evolution of Second Moment

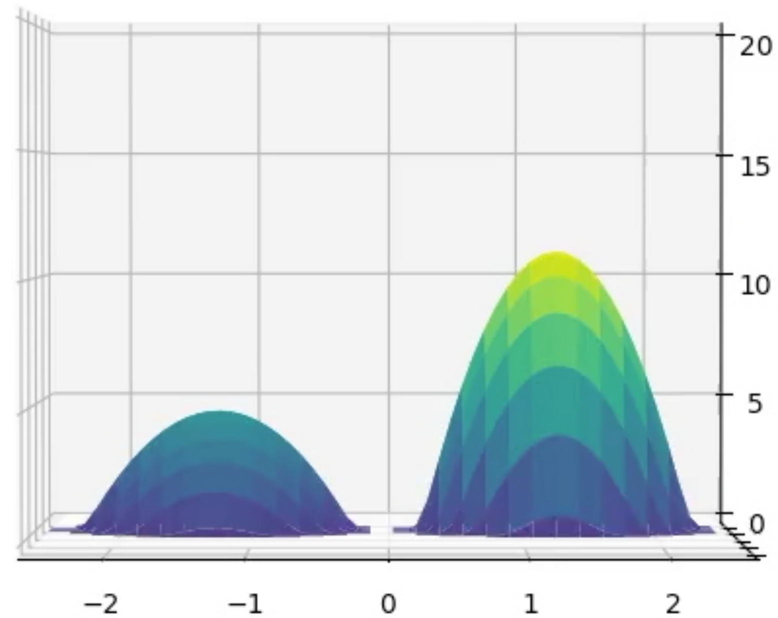


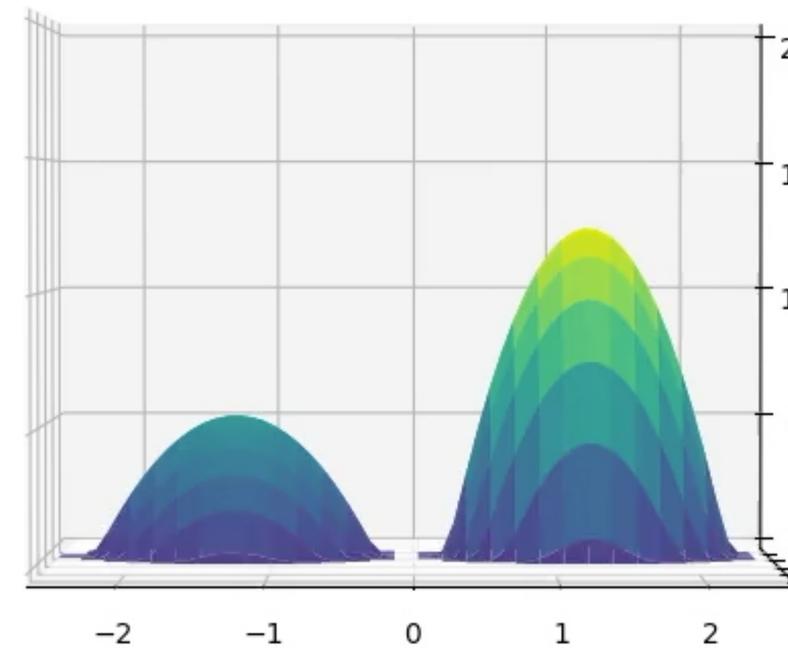
numerics: Keller-Segel (d=2)

subcritical mass | **critical mass** | **supercritical mass**

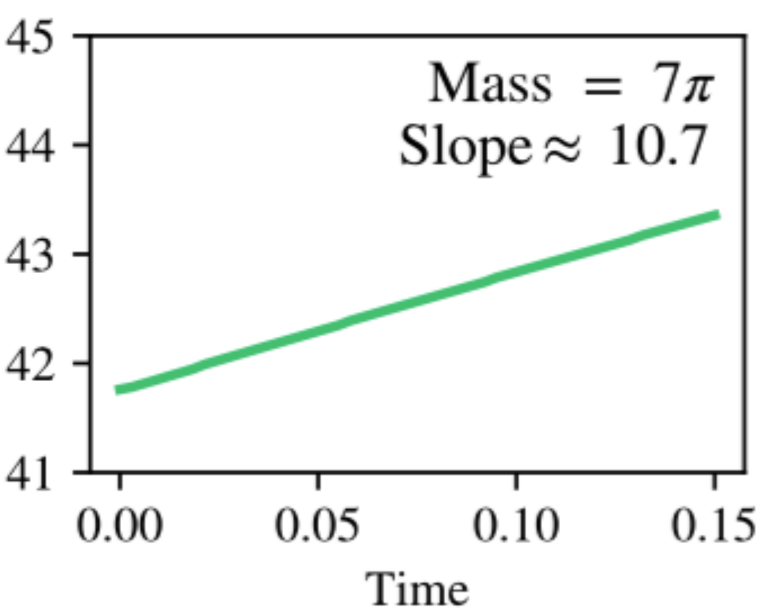
Evolution of Density

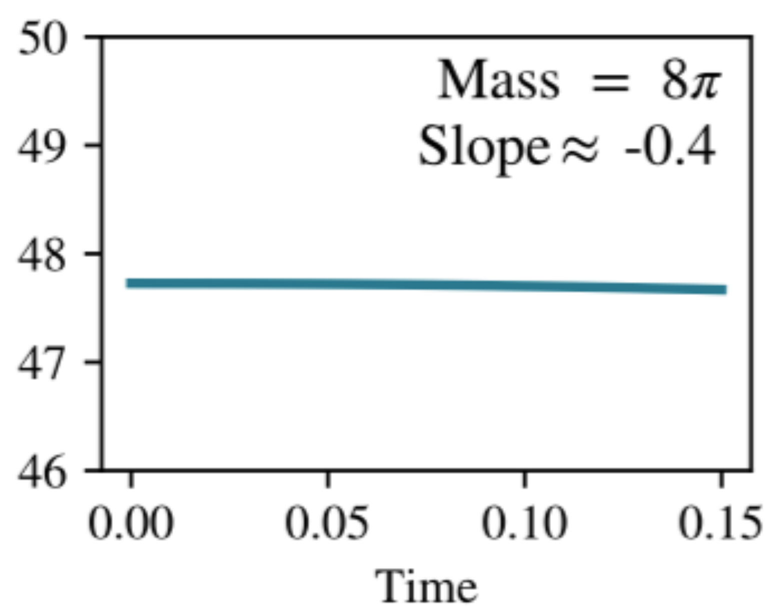


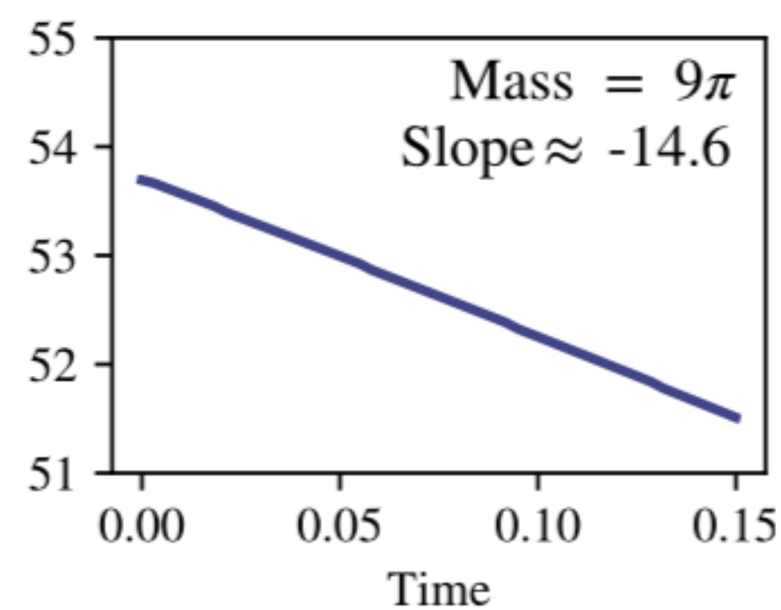




Evolution of Second Moment



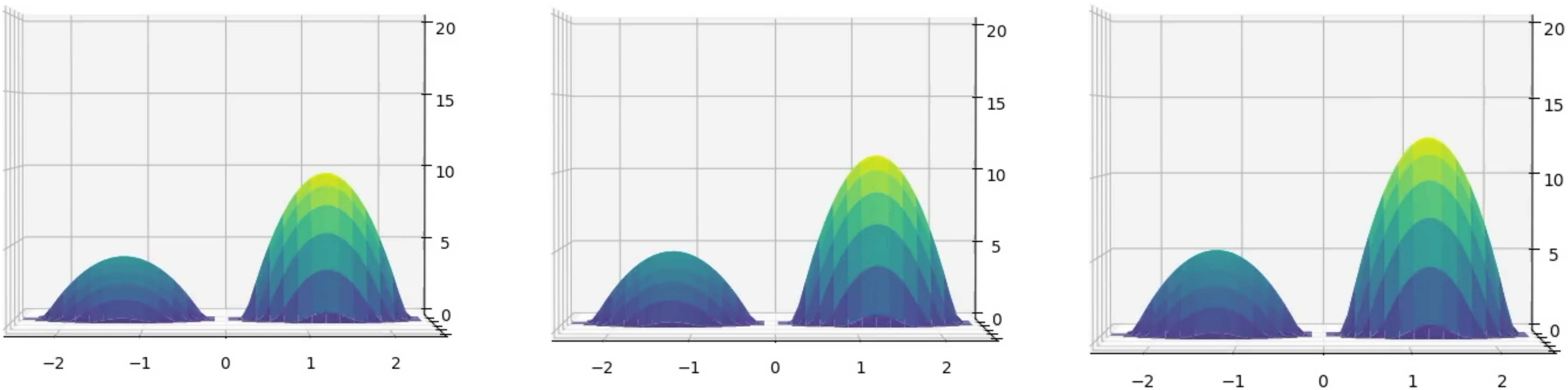




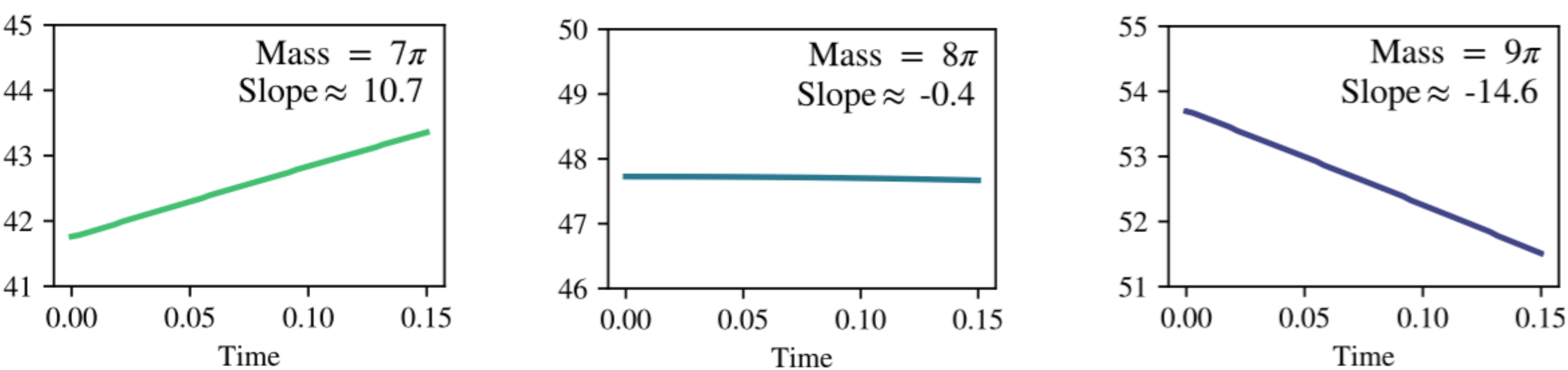
numerics: Keller-Segel (d=2)

subcritical mass | **critical mass** | **supercritical mass**

Evolution of Density



Evolution of Second Moment



Future work

- Convergence for $1 \leq m < 2$?
- Quantitative estimates for $m \geq 2$?
- Utility in related fluids and kinetic equations?

Thank you!