Generative AI with Maxwell, Schrödinger, and Radon

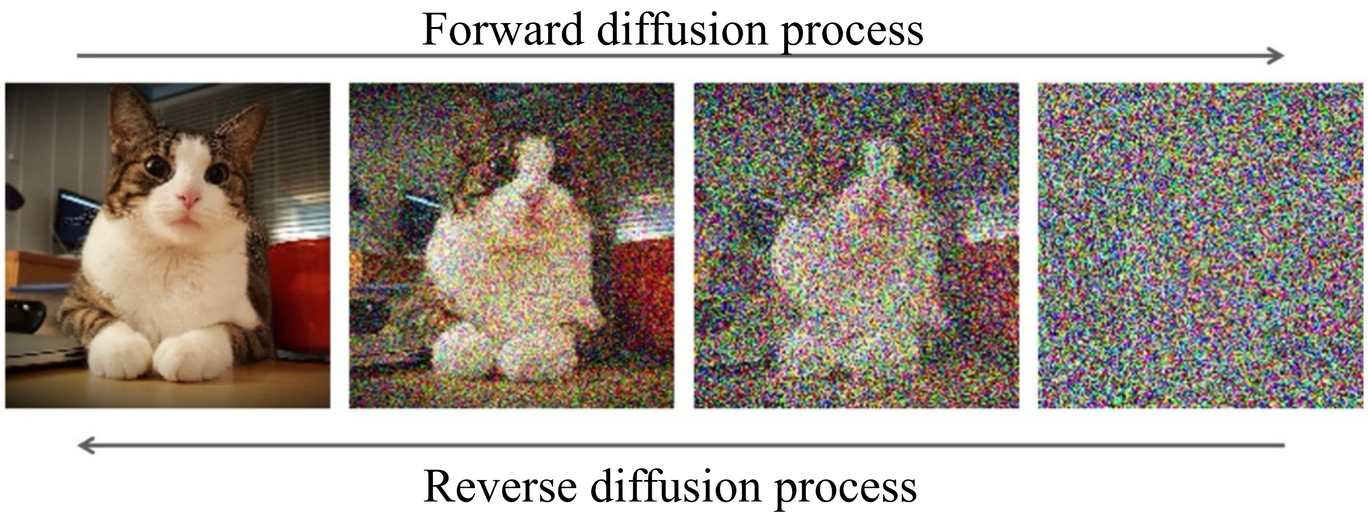
Daniel D. Lee

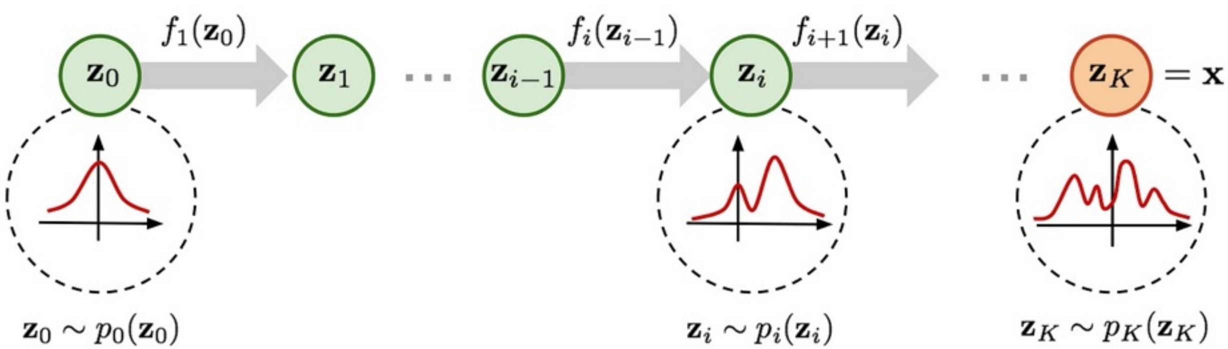
- Thanks to Yung-Kyun Noh for discussions and morning tutorial!

Topics

- Generative AI and probability densities
- Optimal transport formulation
- Flows and score functions
- Maxwell with divergence and curl
- Radon and charge slab model
- Schrödinger gas problem

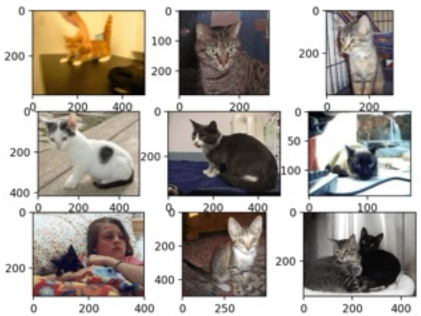
Success of Generative AI

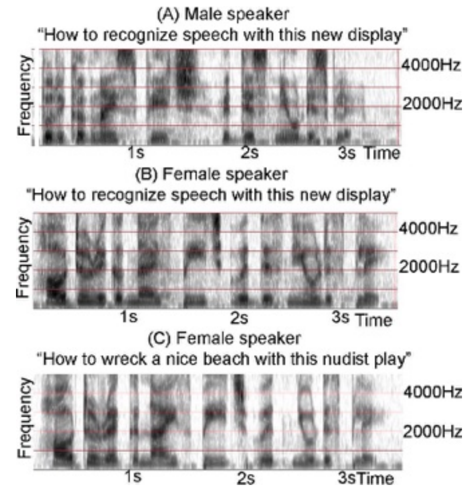


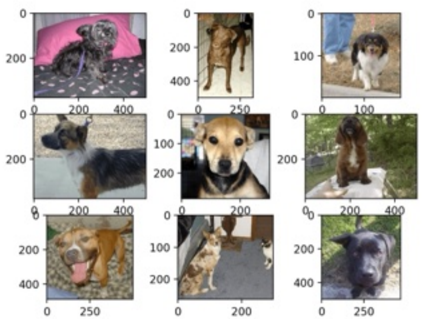


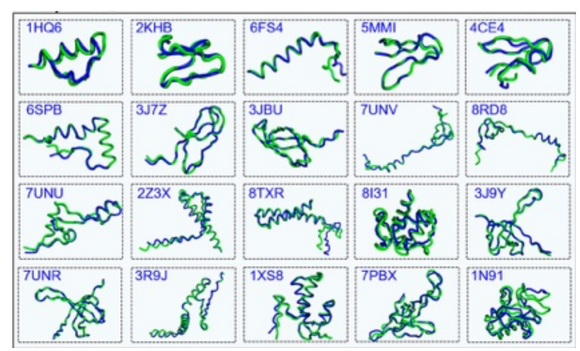
Normalizing flow models

Multimodal applications



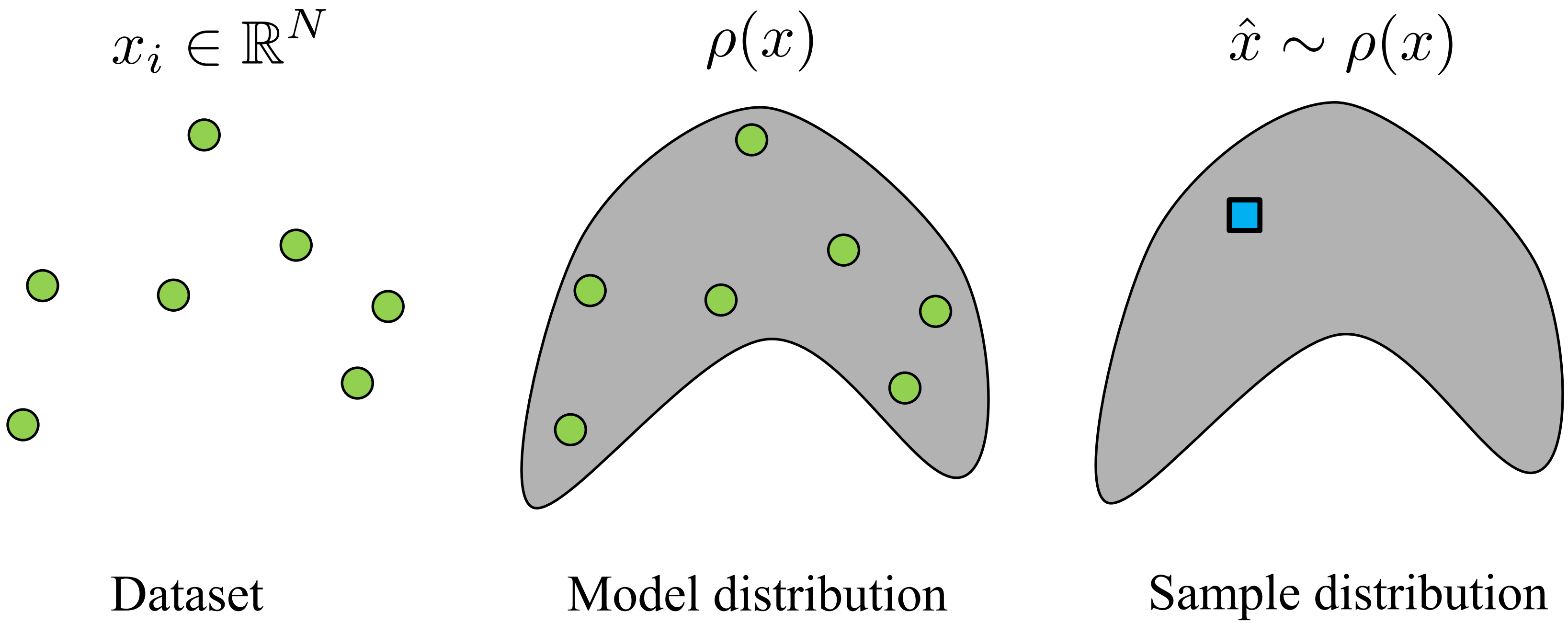






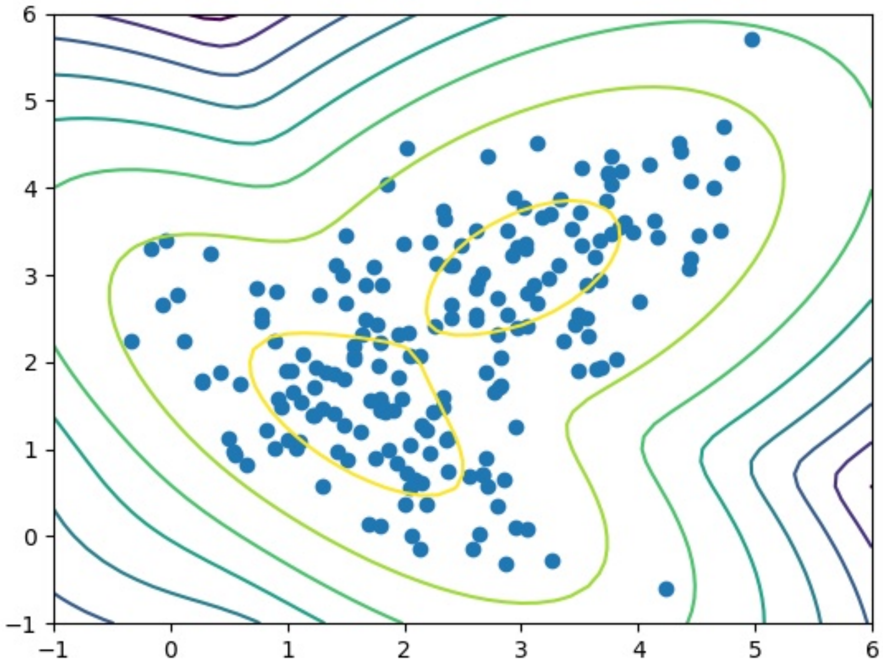
- Generative AI in different domains and data types

Gen AI paradigm



- AI approaches to model data distribution and generate new samples as opposed to discriminative models

Data density distributions



$$x \in \mathbb{R}^N$$

Probability density function

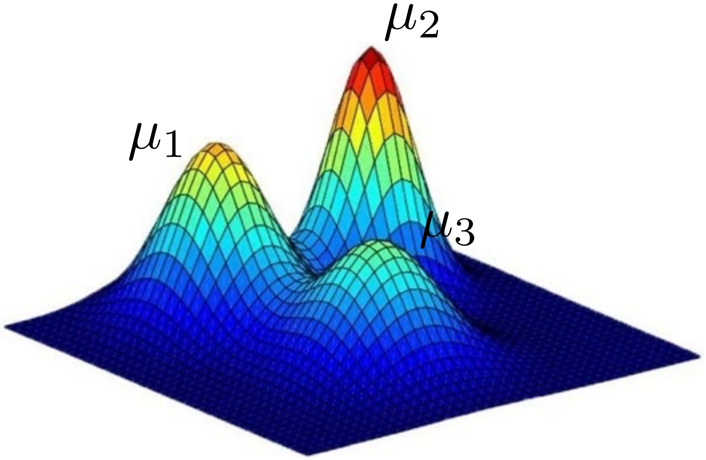
$$\rho(x) \geq 0$$

- Analyzing data distributions in high dimensional feature spaces is critical to AI and ML

Classical density modeling

Parametric density function

$$\rho_\theta(x) = \sum_k \alpha_k \exp\left[-\frac{1}{2}(x-\mu_k)^\top \Sigma_k^{-1}(x-\mu_k)\right]$$



- Model probability density using parametric model such as mixture of Gaussian distributions

Maximum likelihood

Given data samples: $\{x_\alpha\}$

$$L = \prod_\alpha \rho_\theta(x_\alpha)$$

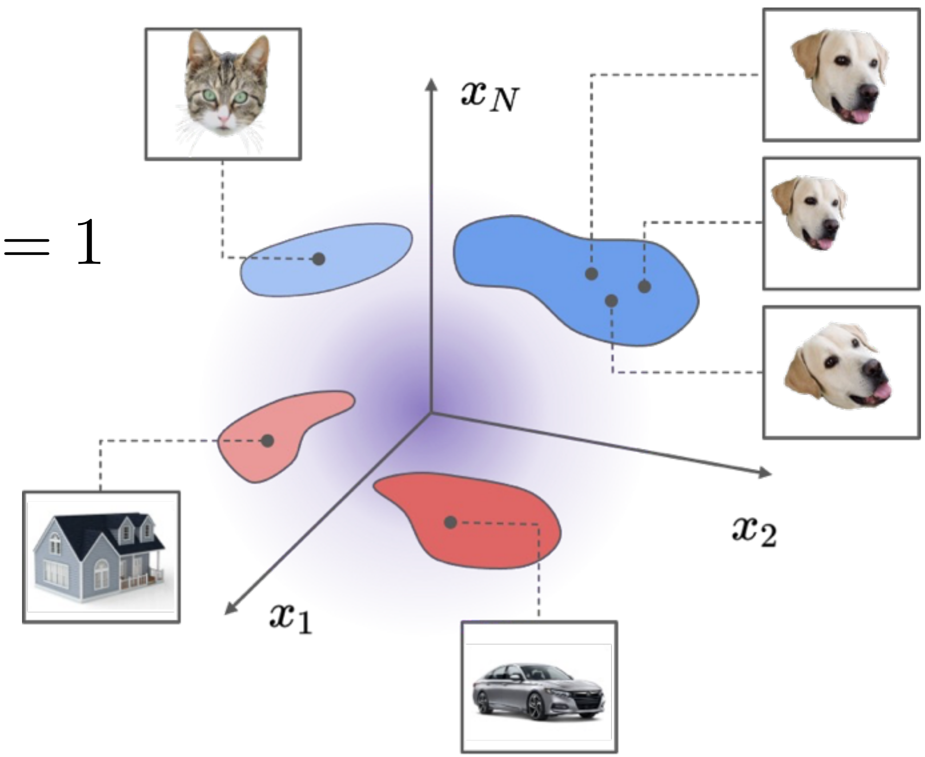
$$\max_\theta \log L = \sum_\alpha \log \rho_\theta(x_\alpha)$$

- Learn model parameters by optimizing log likelihood, e.g. EM algorithm

Issues

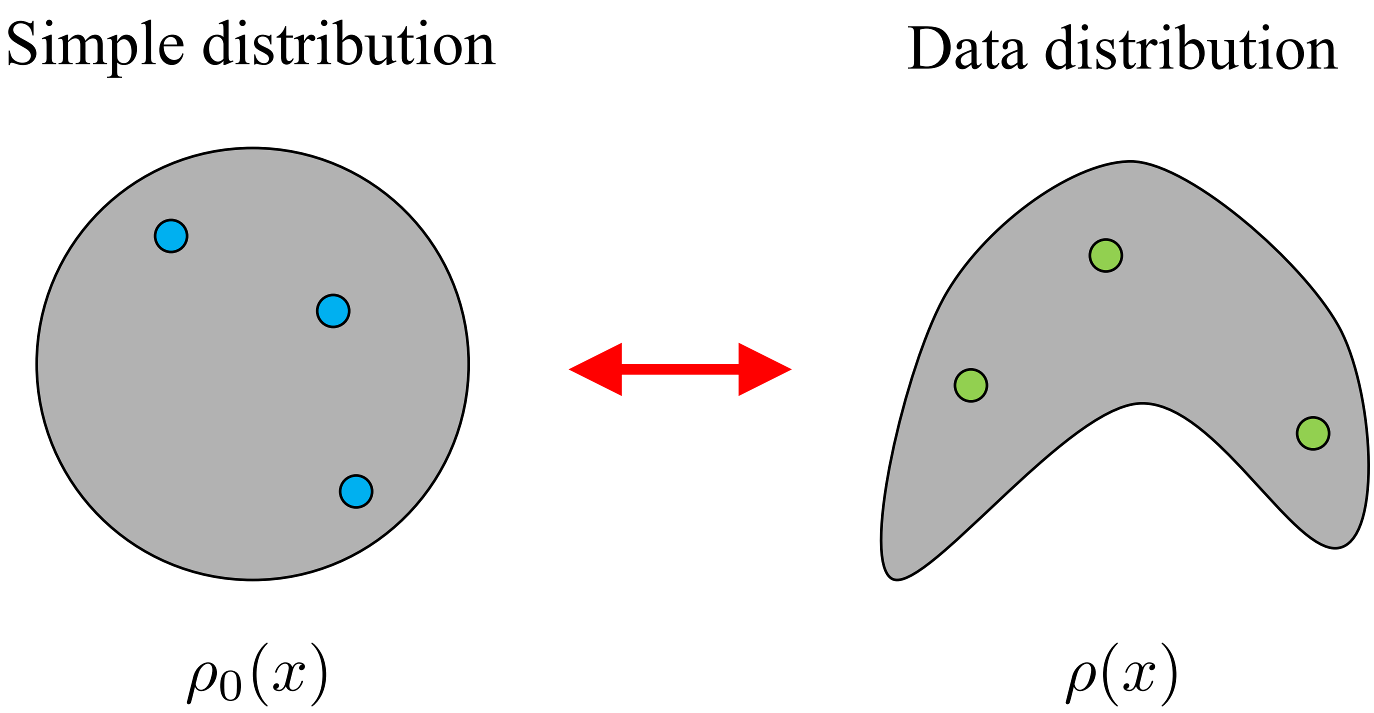
High dimensional data distributions

$$\int dx\, \rho_\theta(x) = 1$$



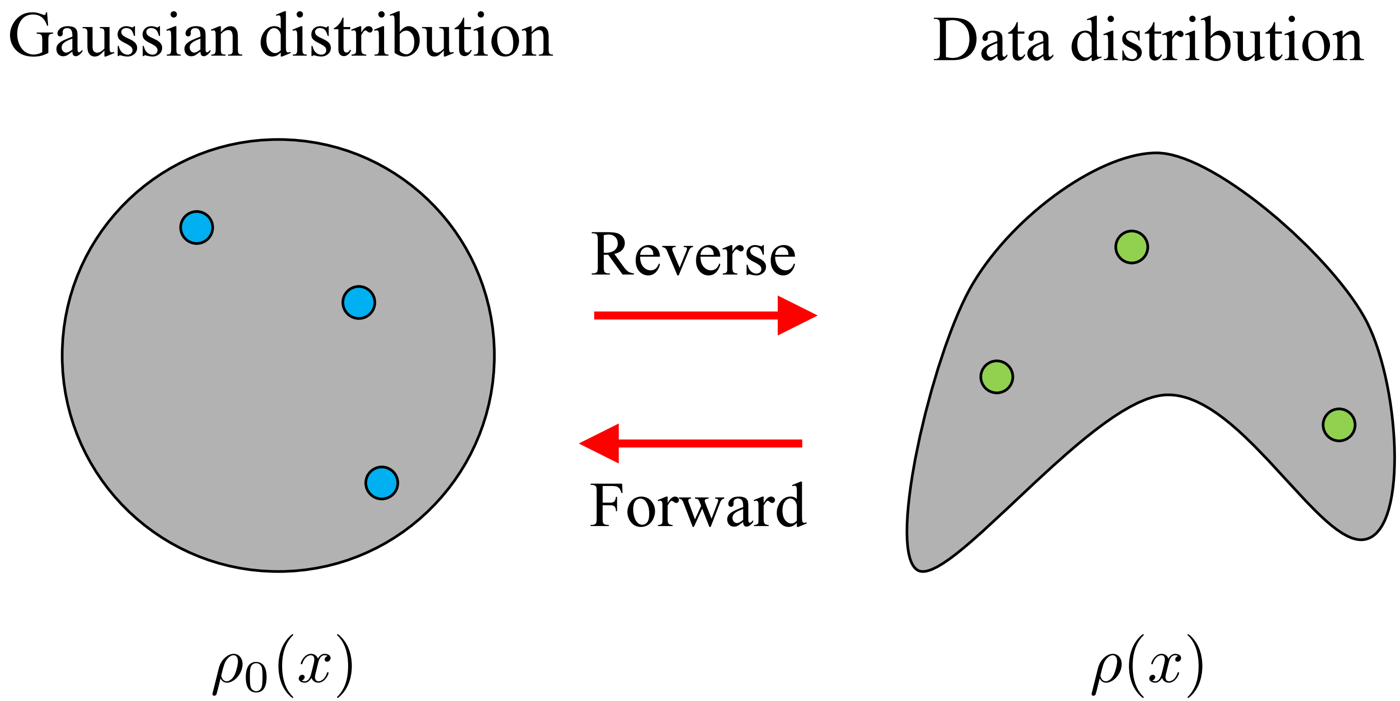
- Normalization constant and difficulties in scaling conventional models

Modern ML approach



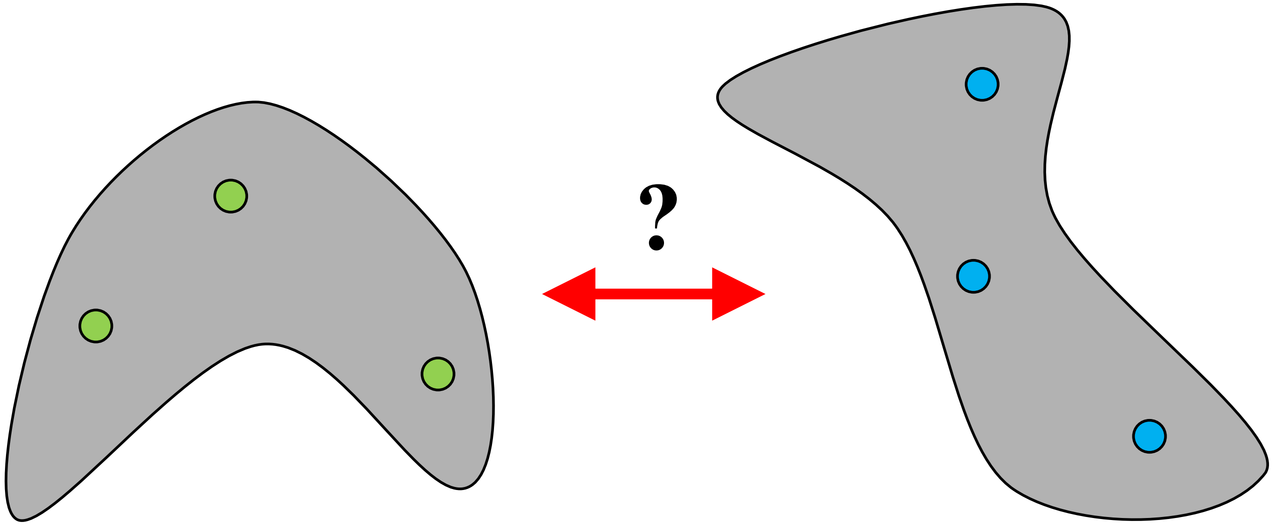
- Transformation to known probability distribution

Reverse diffusion model



- Diffusion models transform data distributions to Gaussian distributions

Transforming distributions

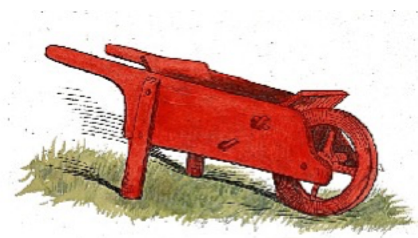


$\rho_0(x)$ $\rho_1(x)$

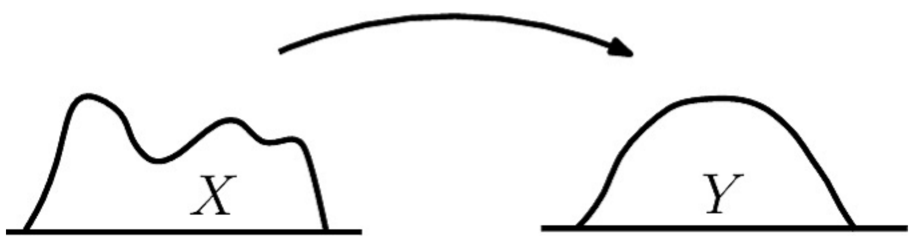
- How to relate two arbitrary distributions, given only the distributions or samples?

Optimal transport

Earth moving problem:



Gaspard Monge
France (1746-1818)

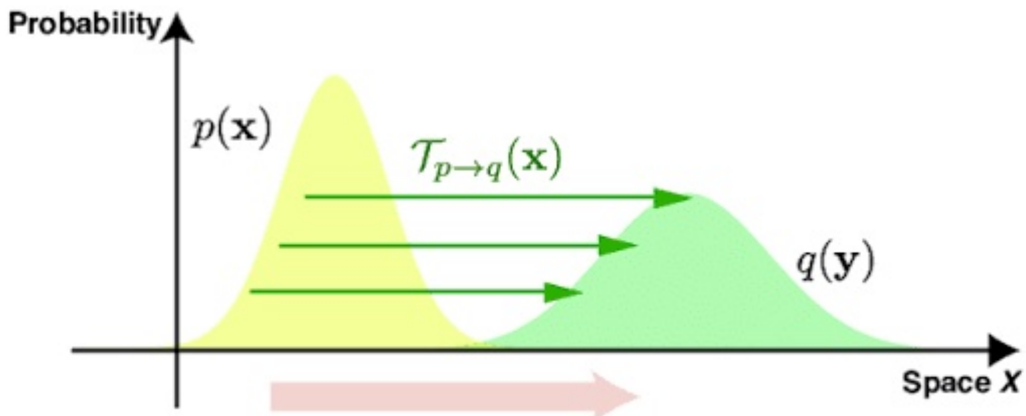


$\rho_0(x)$ $\rho_1(y)$

- Most efficiently move dirt from one distribution to another

Wasserstein distance

Transportation cost: $c(x, y) = \|x - y\|^2$



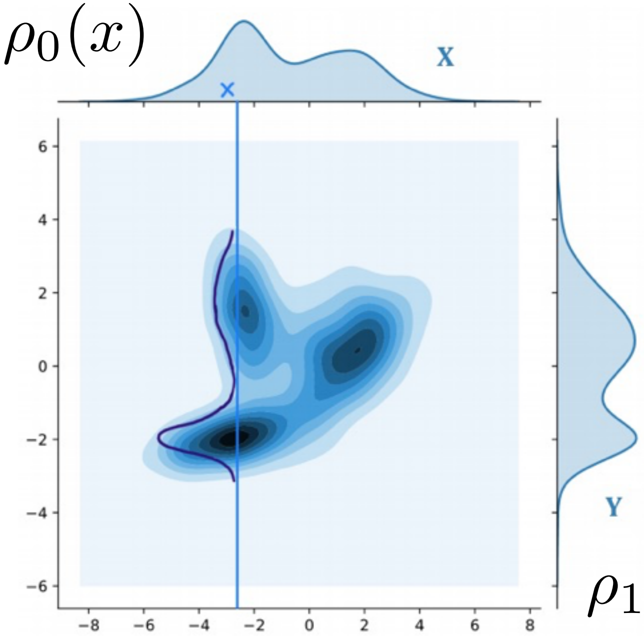
$$W(p,q)^2 = \int dx\, p(x) c(x, T_{p\to q}(x))$$

- Wasserstein distance is optimal cost of moving one distribution to another

Monge-Kantorovich formulation

Consider joint distribution $\rho(x, y)$

$\rho_0(x)$



$\rho_1(y)$

Constrained marginals:

$$\int \rho(x, y)dy = \rho_0(x)$$

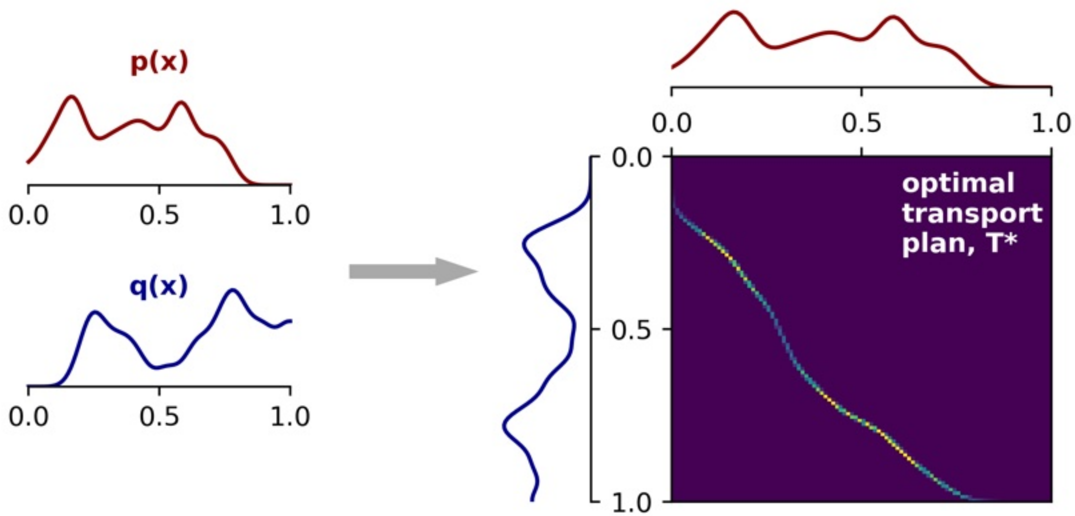
$$\int \rho(x, y)dx = \rho_1(y)$$

- Transportation plan is given by conditional distributions

Monge-Kantorovich optimization

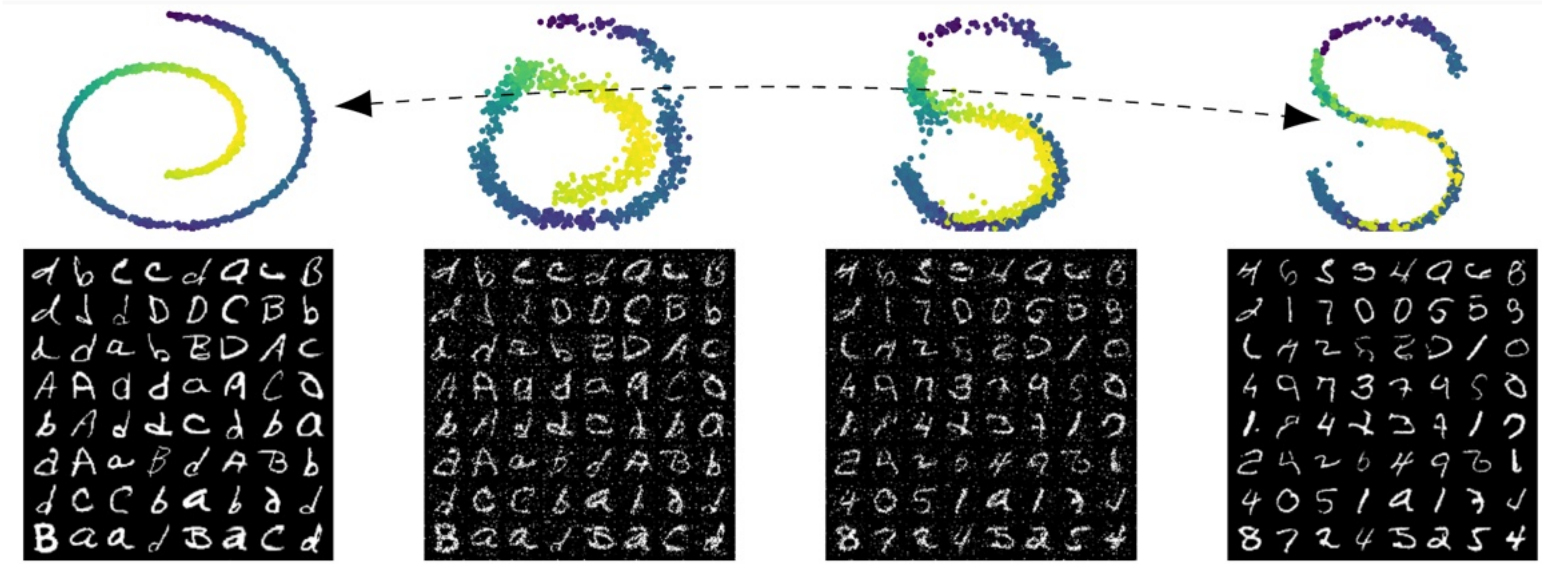
$$\min_{\rho(x,y)} \int dx\, dy\, \rho(x,y) c(x,y)$$

Constraints: $\displaystyle \int \rho(x,y) dy = \rho_0(x) \qquad \int \rho(x,y) dx = \rho_1(y)$



Unpaired data translation

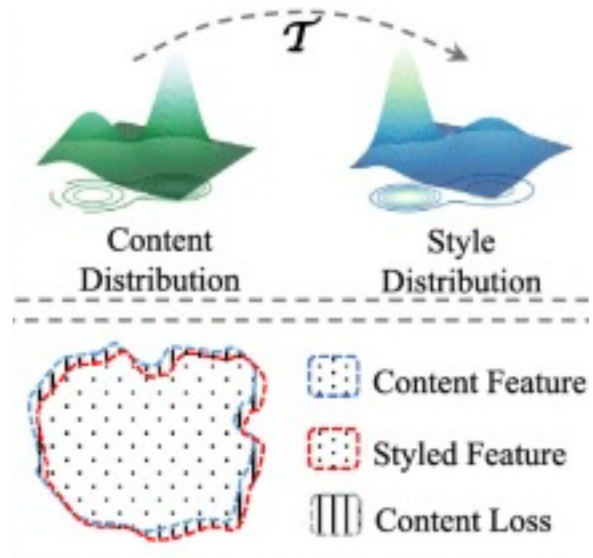
$\rho_0(x)$ $\rho_1(x)$



De Bortoli et. al. (2021)

- Learning correspondences between two data distributions

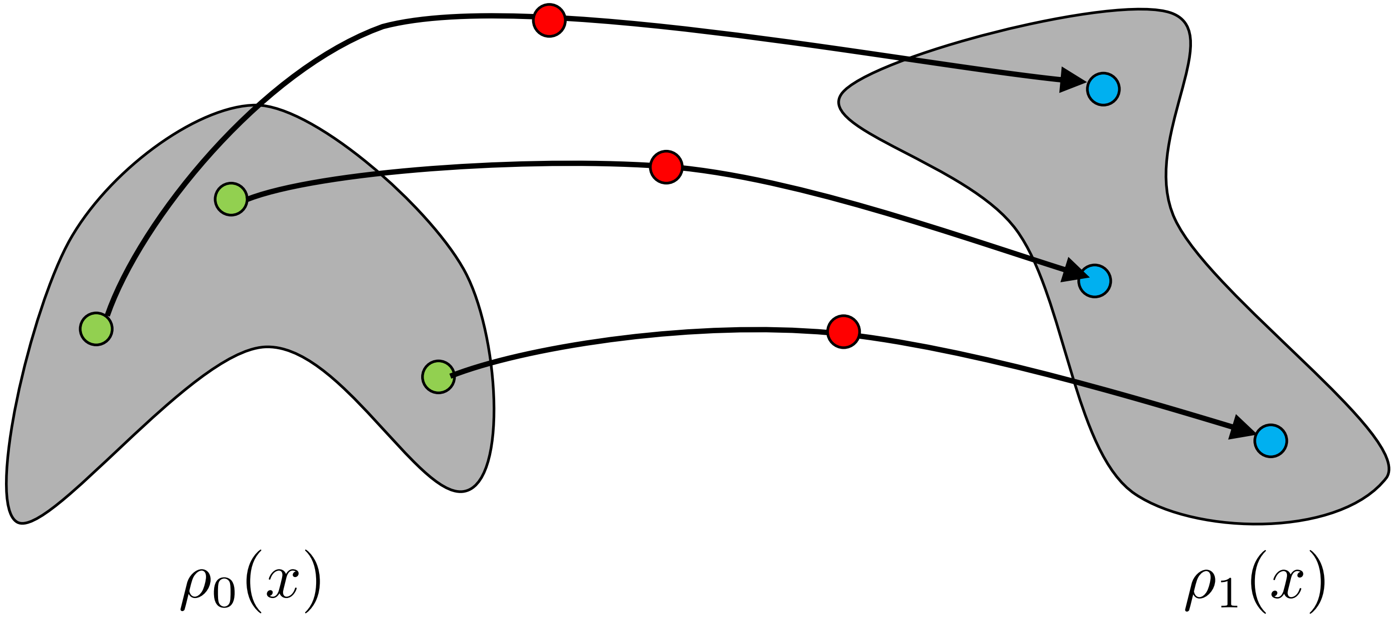
Image style transfer



Kolkin et. al. (2019)

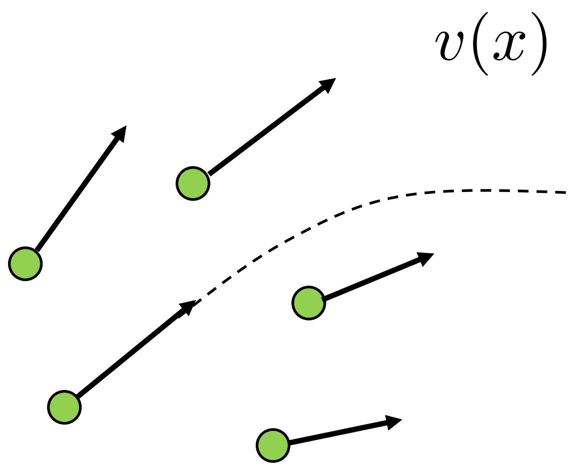
- Map distribution of image content features with distribution of image style features

Dynamic transport



- We also want to understand how samples move (interpolate) between distributions.

Dynamic flows



Velocity flow field

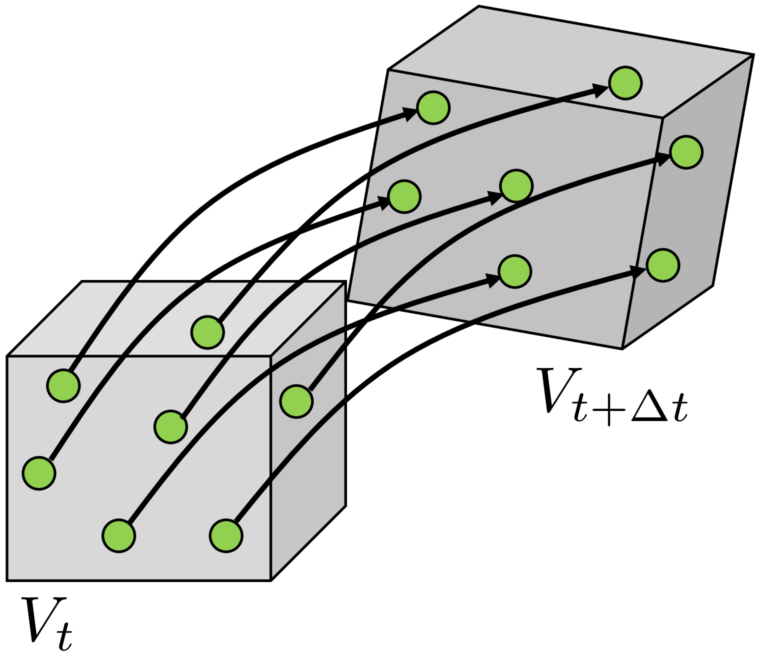
$$v : \mathbb{R}^N \to \mathbb{R}^N$$

Particle motion described by differential equation:

$$dx_t = v(x_t)dt$$

- Dynamical motion of particles described by velocity flows and differential equations

Conservation of mass



Number of particles enclosed in moving volume

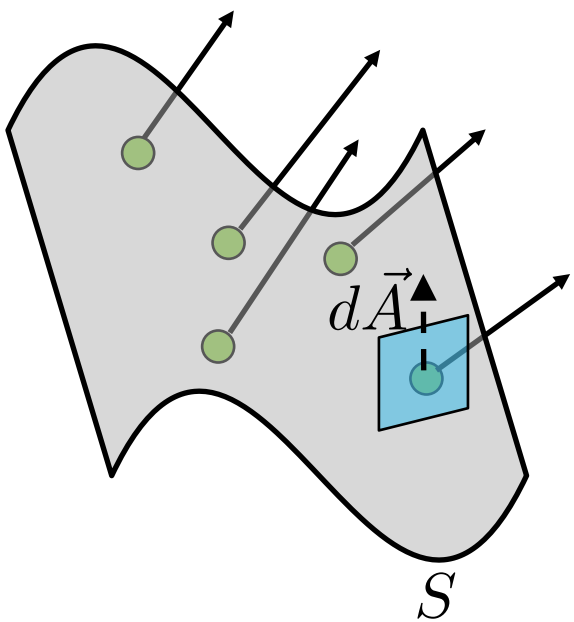
$$\frac{d}{dt}\int_{V_t} p(x)dx = 0$$

Total number of particles is conserved

$$\int_{-\infty}^{+\infty} p(x)dx = 1$$

- With no sources or sinks, particles are neither created nor destroyed

Flux



Flux vector field is
current flow per unit area

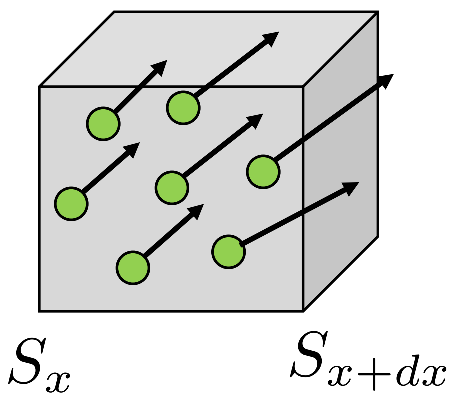
$$\vec{J}(x) = p(x)\vec{v}(x)$$

Particle flow per unit time
across surface

$$\iint_S \vec{J}(x) \cdot d\vec{A}$$

- Flux measures the rate of particle flow across surfaces

Equation of continuity

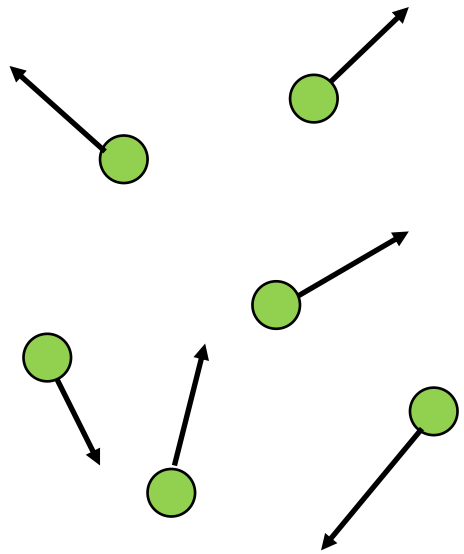


$$\frac{\partial p}{\partial t} + \nabla \cdot \vec{J} = 0$$

$$\nabla \cdot \vec{J} = \frac{J_x(x+dx) - J_x(x)}{dx} + \frac{J_y(y+dy) - J_y(y)}{dy} + \frac{J_z(z+dz) - J_z(z)}{dz}$$

- Green’s theorem relates total flux to temporal change in density within volume

Diffusion



Random white noise $\vec{\eta}_t$

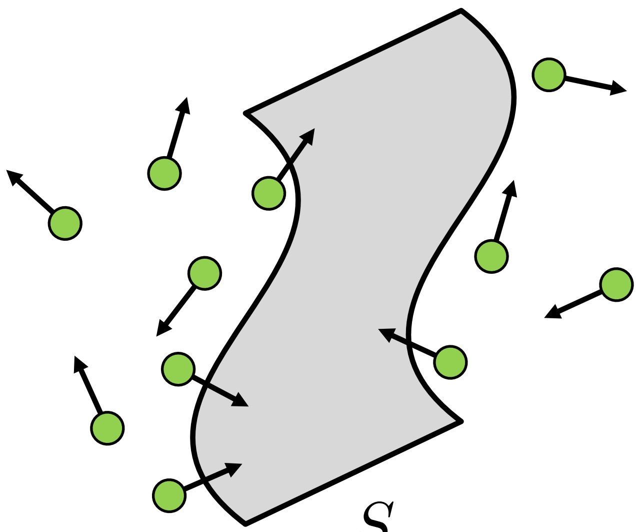
$$\langle \vec{\eta}_t \rangle = 0 \quad \langle \vec{\eta}_t \vec{\eta}_{t'}^\top \rangle = \delta_{tt'} I_{N \times N}$$

Stochastic differential equation:

$$d\vec{x}_t = \nu d\vec{\eta}_t$$

- Adding Brownian random noise to particle motion

Diffusion equation



Flux from diffusion (Fick's law)

$$\vec{J}_D = -\frac{1}{2}\nu\nabla p(\vec{x})$$

Diffusion equation:

$$\frac{\partial p}{\partial t} = -\nabla \cdot \vec{J}_D = \frac{1}{2}\nu\nabla^2 p$$

- Diffusion causes net flow from regions of high density to low density

Fokker-Planck equation

$$\partial_t \rho(t,x) = -\nabla \cdot [\rho(t,x) f(t,x,u)] + \frac{1}{2}\nu_t \nabla^2 \rho(t,x)$$

$$\frac{\partial \rho_t}{\partial t} = -\nabla \cdot \vec{J}$$

Fluxes:

$$\vec{J}_v = \rho_t(x) f(t,x,u)$$

$$\vec{J}_D = -\frac{1}{2}\nu_t \nabla \rho_t(x)$$

- Fokker-Planck equation as generalization of continuity equation with fluxes from both drift and diffusion

Score function

Fokker-Planck:

$$\partial_t \rho(t,x) = -\nabla \cdot [\rho(t,x) f(t,x,u)] + \frac{1}{2}\nu_t \nabla^2 \rho(t,x)$$

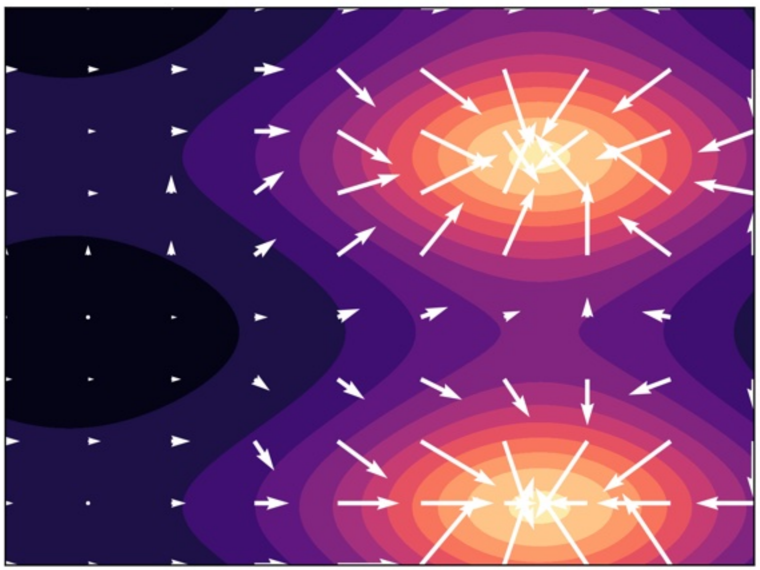
$$\begin{aligned}\nabla^2 \rho(t,x) &= \nabla \cdot [\nabla \rho(t,x)] \\ &= \nabla \cdot \left[\rho(t,x) \frac{1}{\rho(t,x)} \nabla \rho(t,x)\right] \\ &= \nabla \cdot [\rho(t,x) \nabla \log \rho(t,x)]\end{aligned}$$

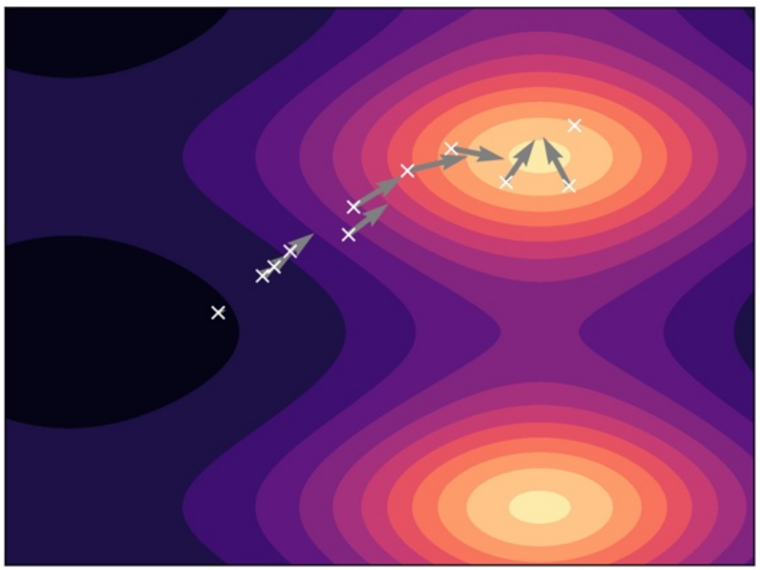
$$\vec{f}_D(t,x) = -\frac{1}{2}\nu_t \nabla \log \rho(t,x)$$

- Effect of diffusion is equivalent to deterministic flow given by negative score function

Score functions

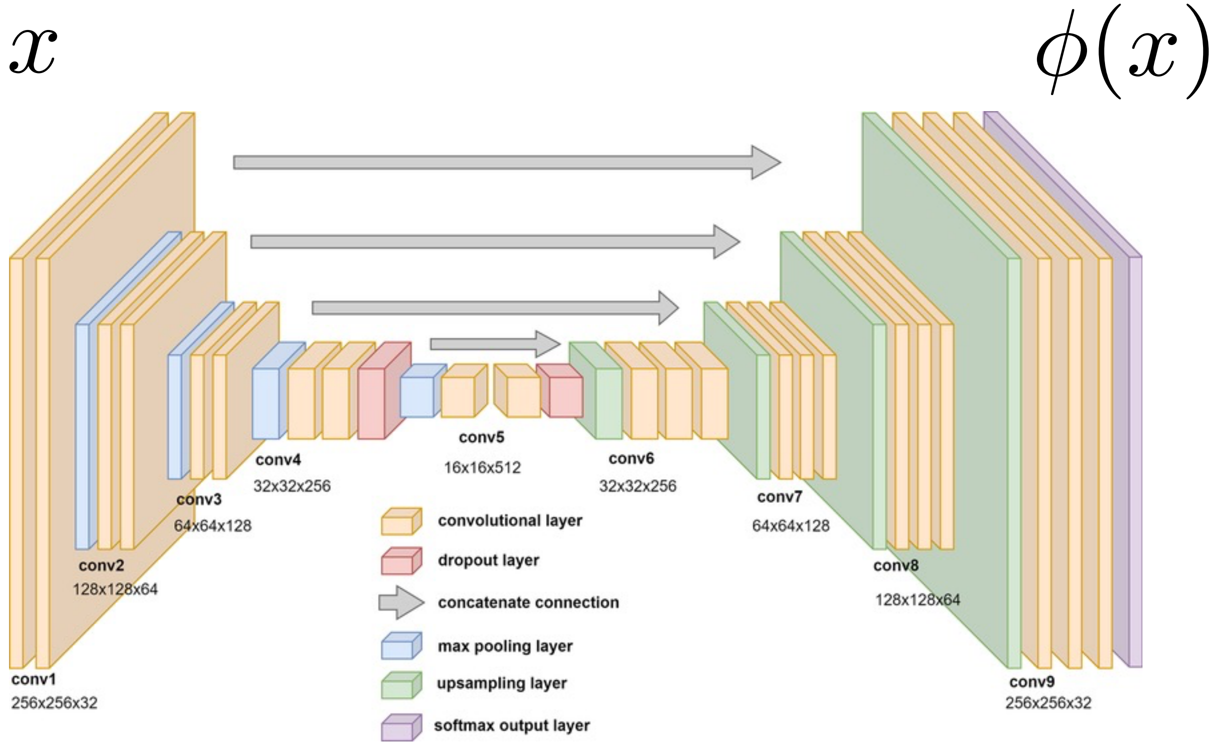
$$\phi(t, x) = \nabla \log \rho(t, x)$$





- Need to model score function to dynamically generate samples from distribution

Score function approximation



- U-net architecture to approximate score function flows

Score matching

Learn score function: $\phi(t,x) = \nabla \log \rho(t,x)$

$$\min_{\phi} \int dx \rho(t,x) \left[\phi(t,x) - \nabla \log \rho(t,x)\right]^2$$

$$= \min_{\phi} \int dx \rho(t,x) \left[\phi^2 - 2\phi \cdot \nabla \log \rho(t,x)\right]$$

$$= \min_{\phi} \int dx \rho(t,x) \left[\phi^2 + 2\nabla \cdot \phi\right]$$

$$\approx \min_{\phi} \frac{1}{N} \sum_{i} \left[\phi^2(x_i) + 2\nabla \cdot \phi(x_i)\right]$$

- Variational Stein identity to learn score function from data samples (Hyvarinen 2005)

Densities and scores

Density with energy-based model:

$$\rho(x) = \frac{1}{Z} \exp[-V(x)]$$

Score function:

$$\phi(x) = \nabla \log \rho(x) = -\nabla V(x)$$

Score matching:

$$\min \left\langle \phi^2(x) + 2\nabla \cdot \phi(x) \right\rangle$$

- Relationship between scores and energy-based density models

ICML workshop

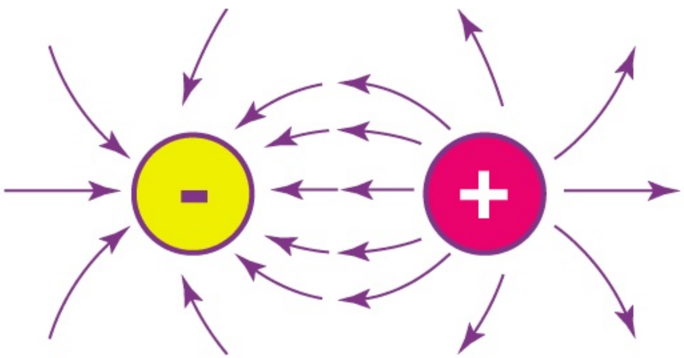
“***Electrostatic models for score matching***”
(ICML SPIGM workshop 2026)

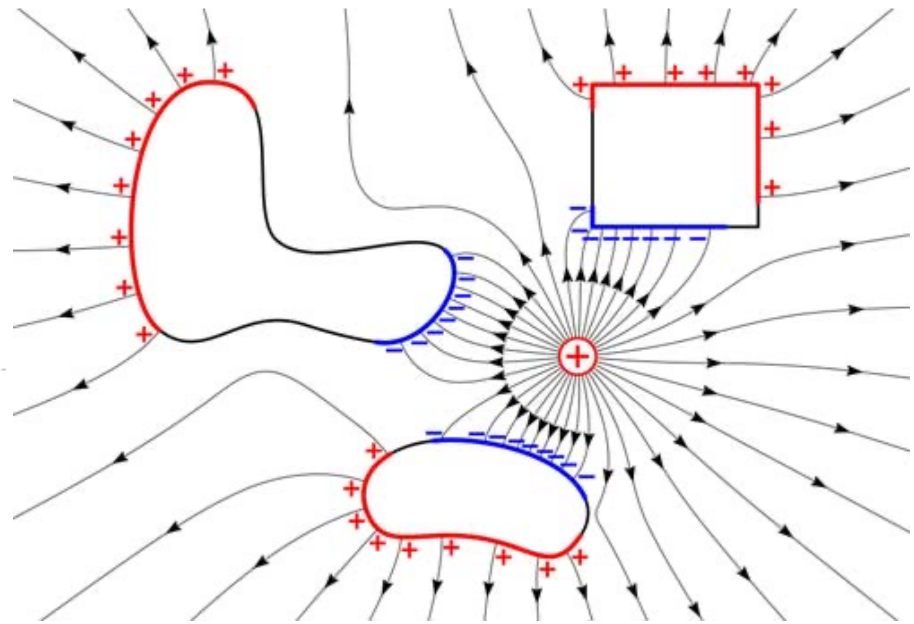
Laura Appleby
Cornell Univ.

Lawrence Saul
Flatiron/Simons

Diana Cai
Flatiron/Simons

Electrostatics





- Describes forces between electric charges

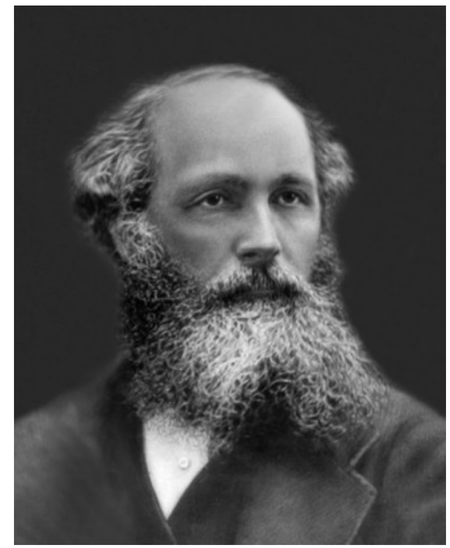
Maxwell equations

$$B(x) = 0$$

$$\nabla \cdot E(x) = q(x)$$

$$\nabla \times E(x) = 0$$

$$E(x) = -\nabla V(x)$$



James Clerk Maxwell
~1860

- Maxwell equations for electrostatics with no moving charges and magnetic field

Electrostatic analogy

$$E(x) \quad \longleftrightarrow \quad \phi(x) = \nabla \log \rho(x)$$

Concept	**Electrostatics**	**Machine Learning**
Potential	$V(\boldsymbol{x})$	$p(\boldsymbol{x}) \propto e^{-V(\boldsymbol{x})}$
Field	$\mathbf{E}(\boldsymbol{x}) = -\nabla V(\boldsymbol{x})$	$\nabla \log p(\boldsymbol{x})$
Charge	$q(\boldsymbol{x}) = \nabla \cdot \mathbf{E}(\boldsymbol{x})$	$\nabla^2 \log p(\boldsymbol{x})$
Superposition	$\mathbf{E}(\boldsymbol{x}) = \sum_k \mathbf{E}_k(\boldsymbol{x})$	$p(\boldsymbol{x}) \propto \prod_k p_k(\boldsymbol{x})$

- Explicit analogy between electric fields and score functions in machine learning

Benefits

$$E(x) \longleftrightarrow \phi(x) = \nabla \log \rho(x)$$

Automatically models a gradient flow (curl-free):

$$E(x) = -\nabla V(x)$$

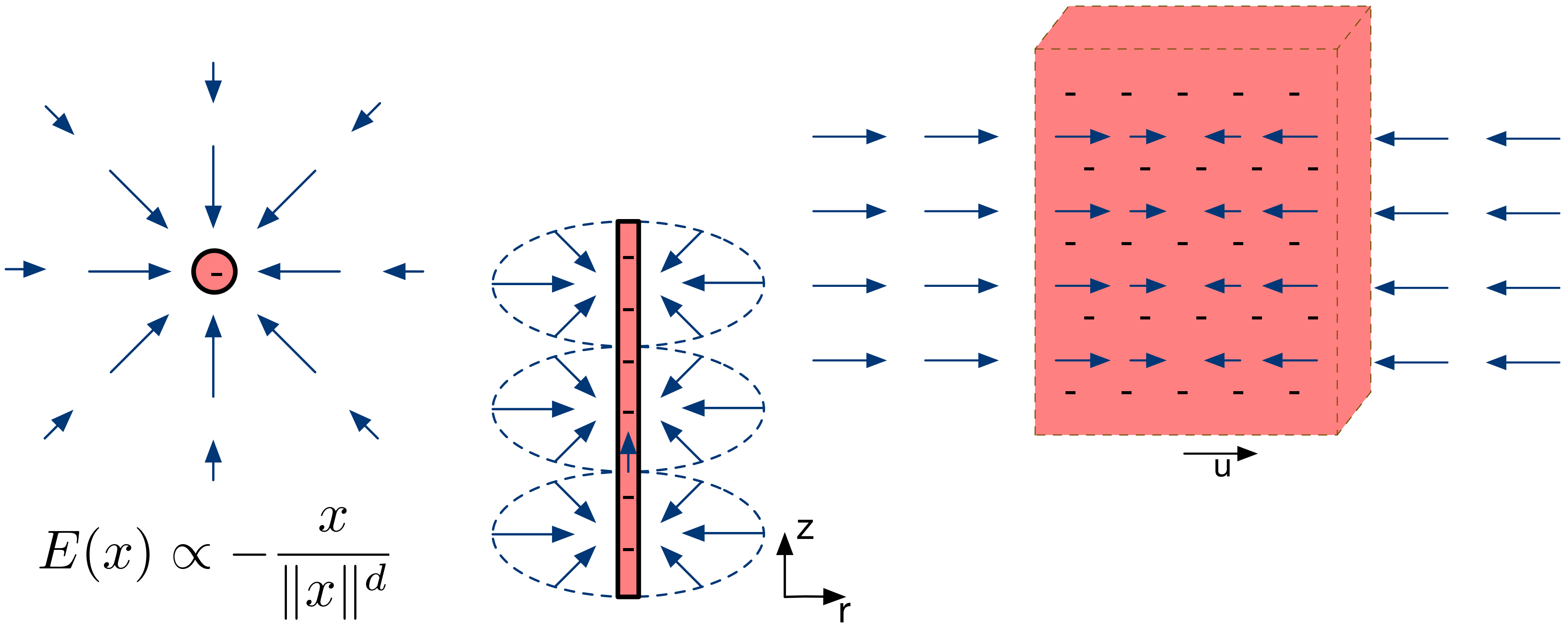
Divergence is easily computed as underlying charge:

$$\nabla \cdot E(x) = q(x)$$

Superposition is equivalent to product of experts model:

$$E(x) = \sum_k E_k(x)$$

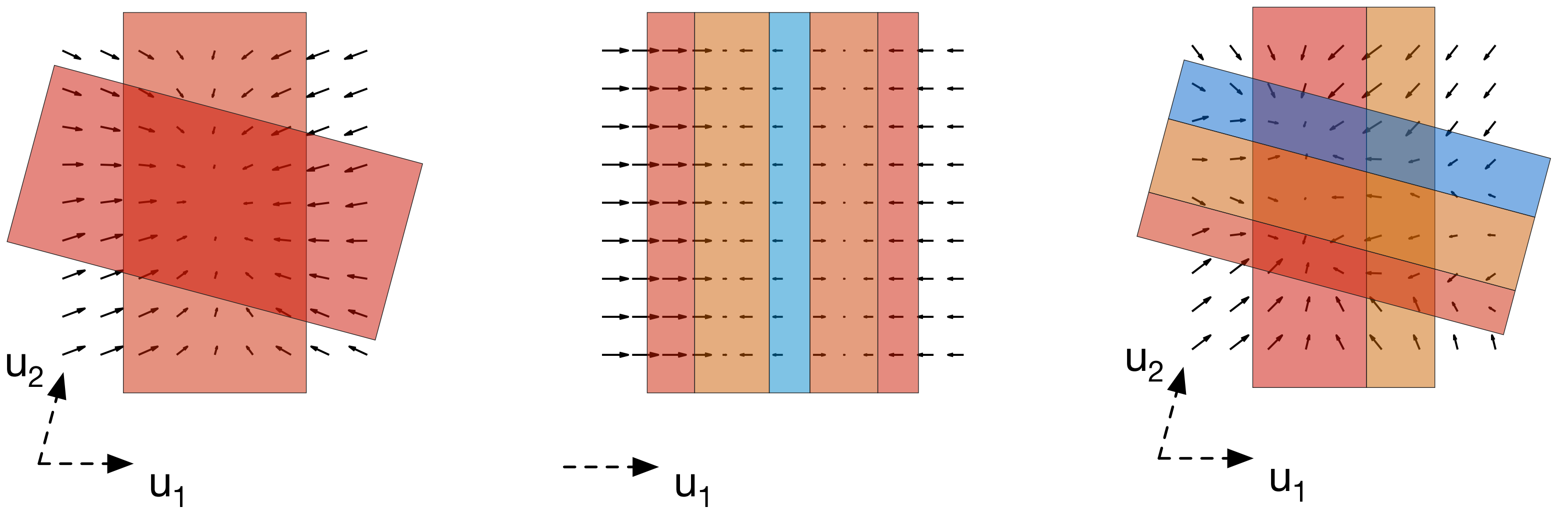
Electric fields



$$E(x) \propto -\frac{x}{\|x\|^d}$$

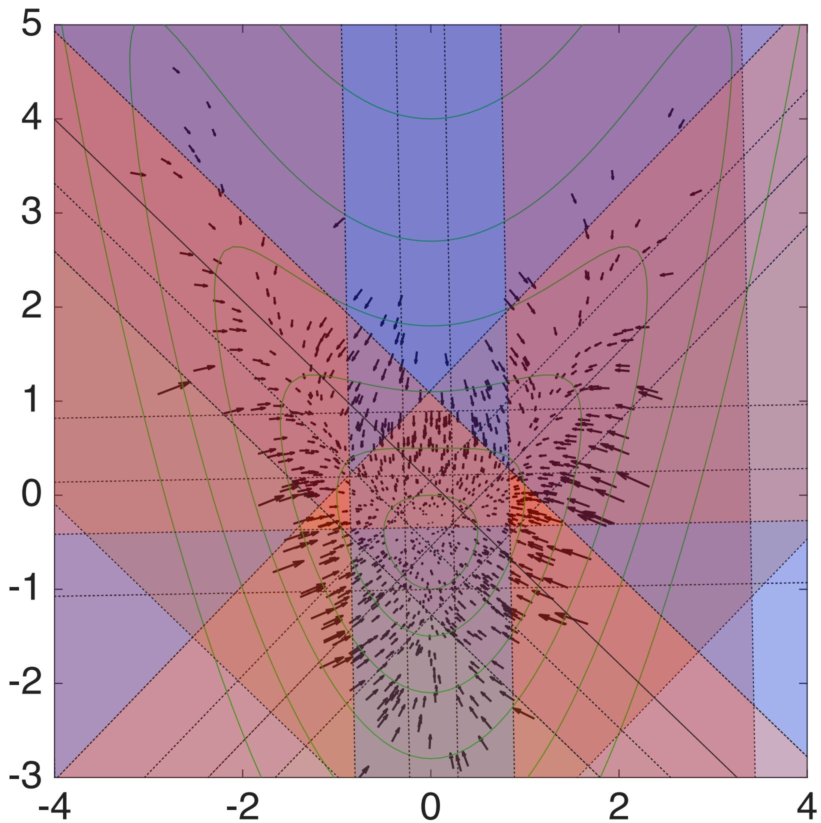
- Relationship between charge distribution and associated electric fields

Superposition



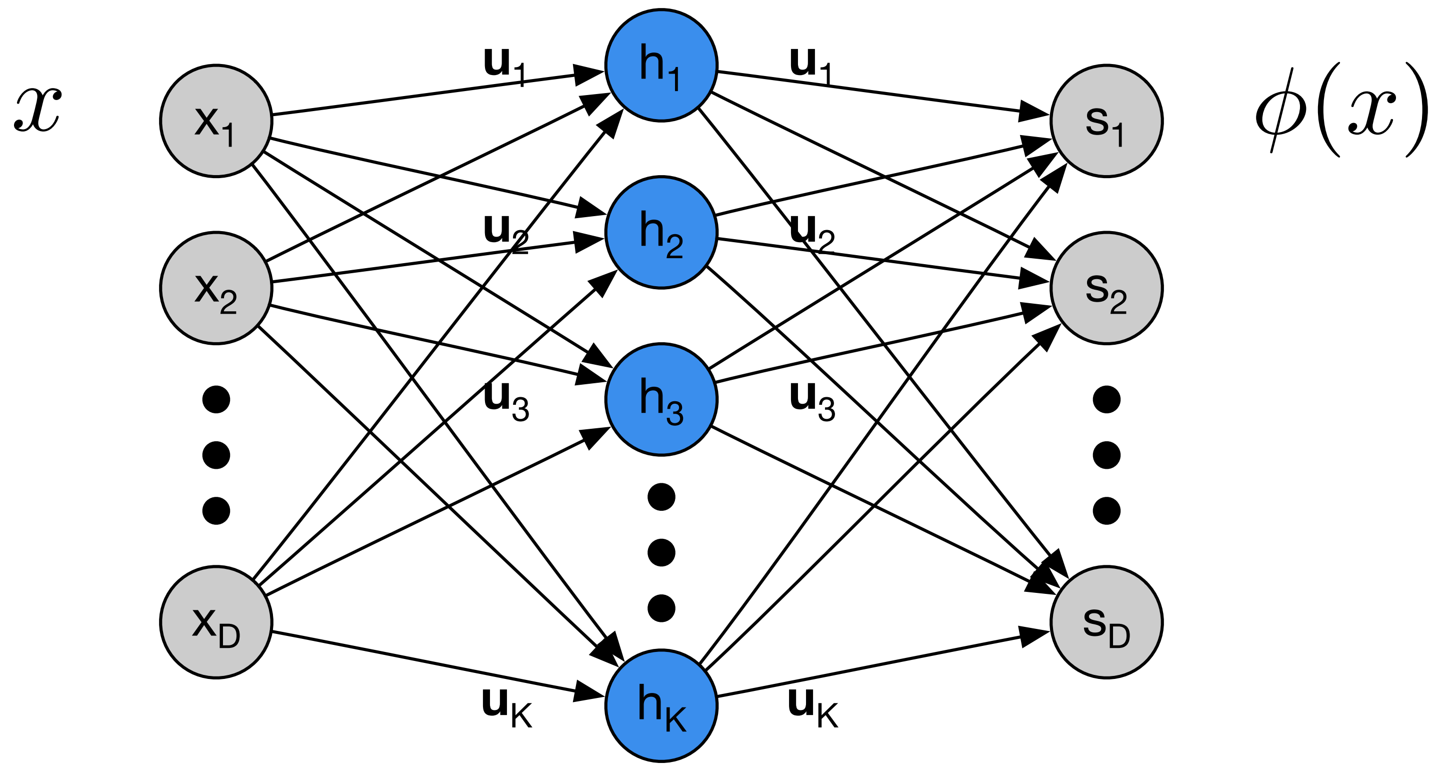
- Model score functions as superposition of electric fields from charged slabs

Example



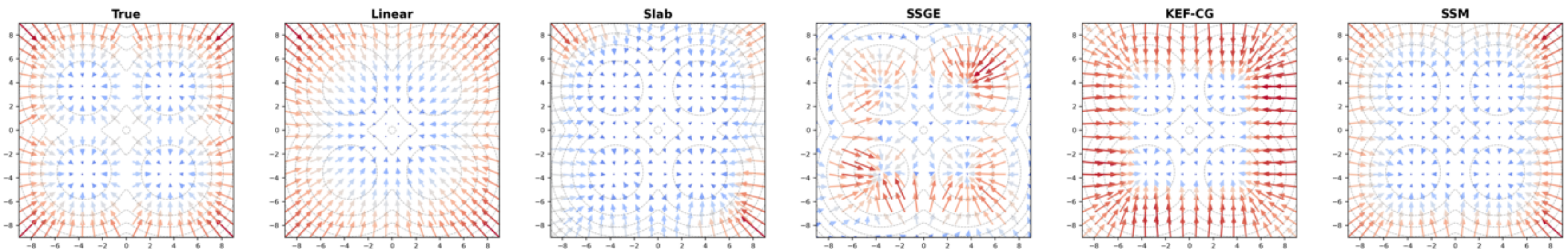
- Approximate score function of complex data distribution with charged slabs

Neural network implementation



- Charged slab model is equivalent to wide ReLU neural network with tied weights

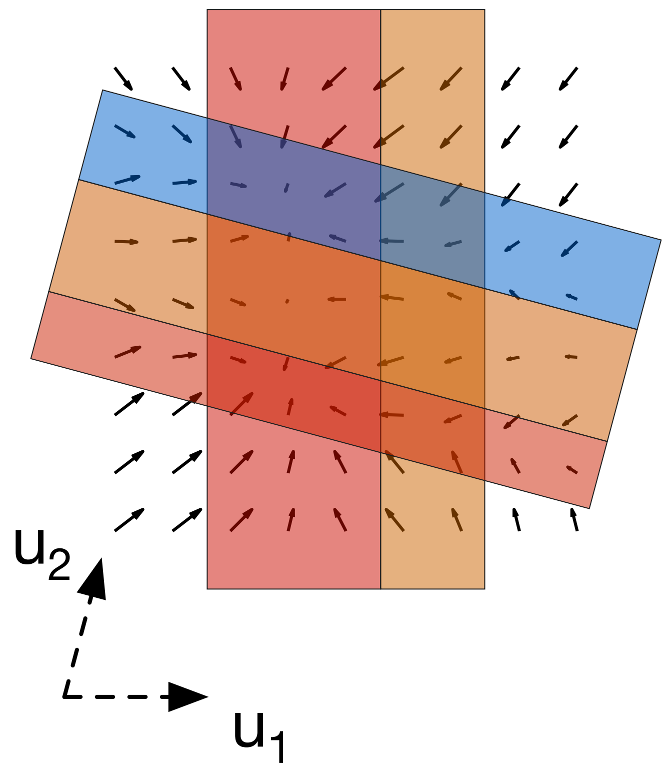
Experimental results



	d=2	d=5	d=10	d=20	d=50
Linear	1.801	1.851	1.901	2.128	3.610
Slab	0.134±.009	0.392±.031	0.684±.026	**1.039**±.035	3.327±.124
SSGE†	0.163±.003	0.459±.019	0.892±.015	2.055±.059	**3.180**±.034
KEF-CG†	0.130±.008	0.326±.009	**0.672**±.010	1.649±.035	OOM
SSM†	**0.074**±.003	**0.239**±.006	1.553±.303	2.732±.034	5.426±.063

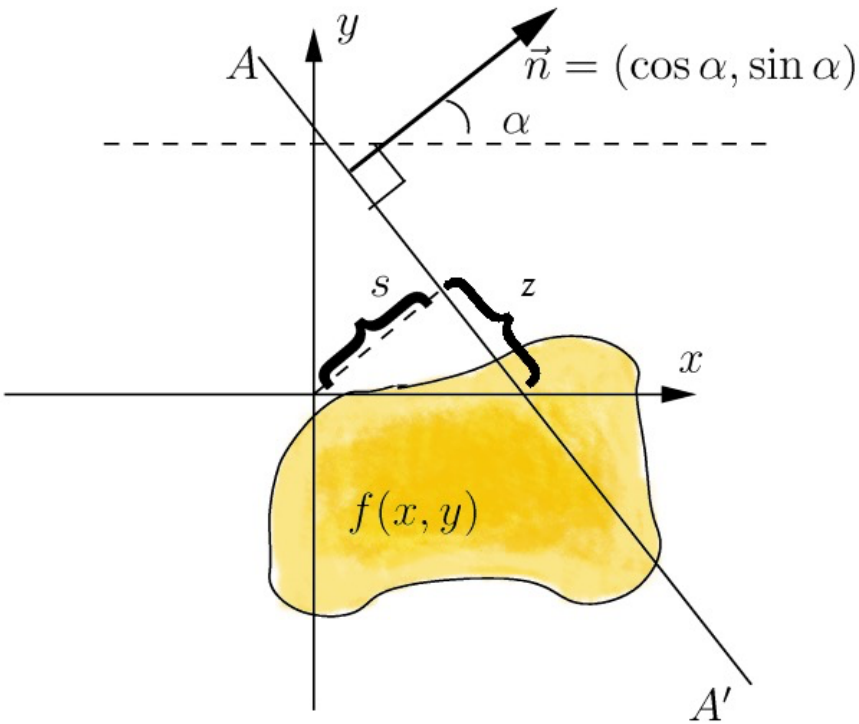
- Simple charged slab model is competitive and more efficient than other score matching methods

Universal approximation



- Can a superposition of 1-d charged slabs approximate all possible score functions?

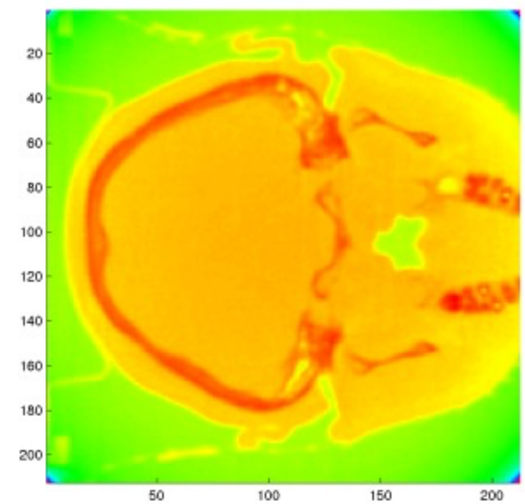
Radon transform



Johann Radon

- Mathematics of projections of high-dimensional densities

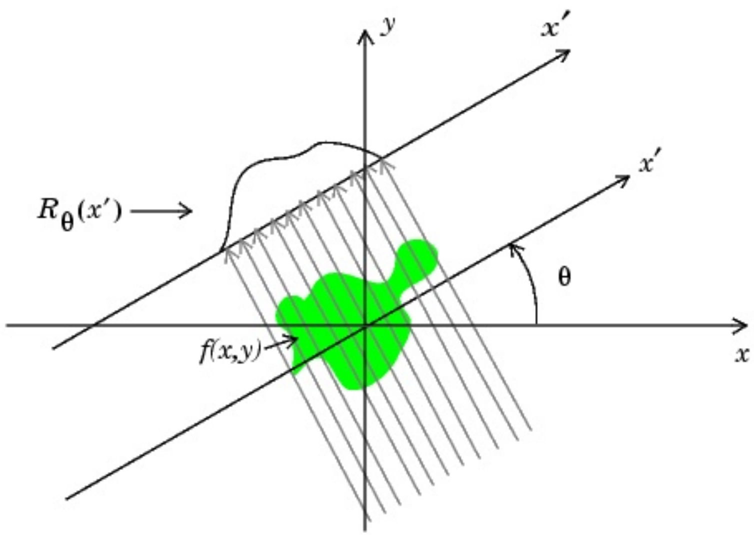
Computed Tomography (CT)



- How to reconstruct volumetric density from X-ray projections?

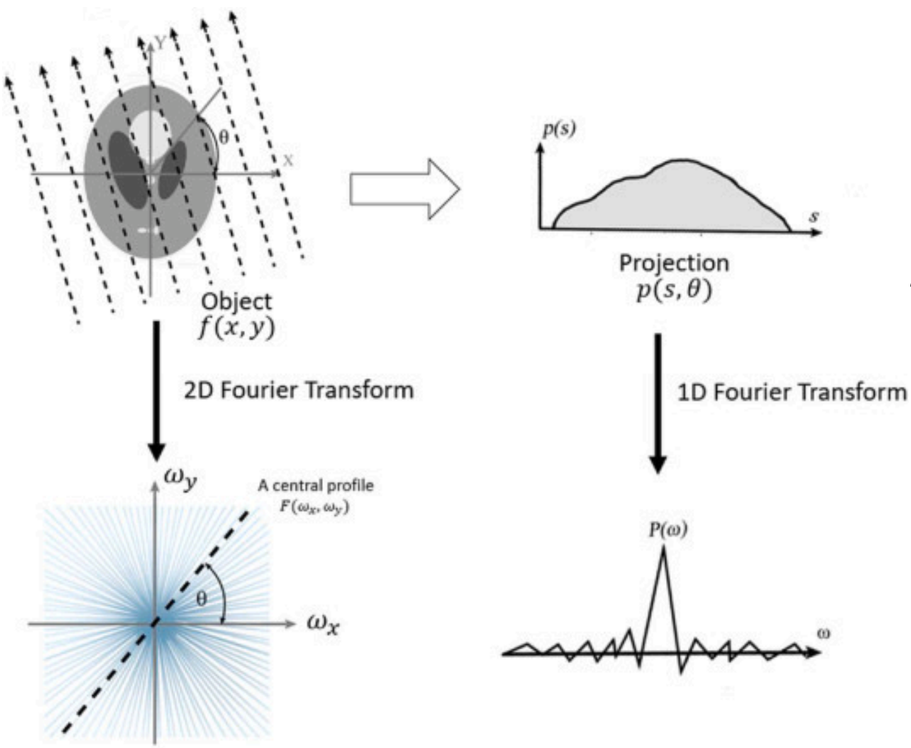
Radon transform

$$\rho_u(x') = \int dx\, \rho(x)\delta(t - u \cdot x)$$



- Convert high-dimensional density into collection of low-dimensional densities

Fourier slice theorem

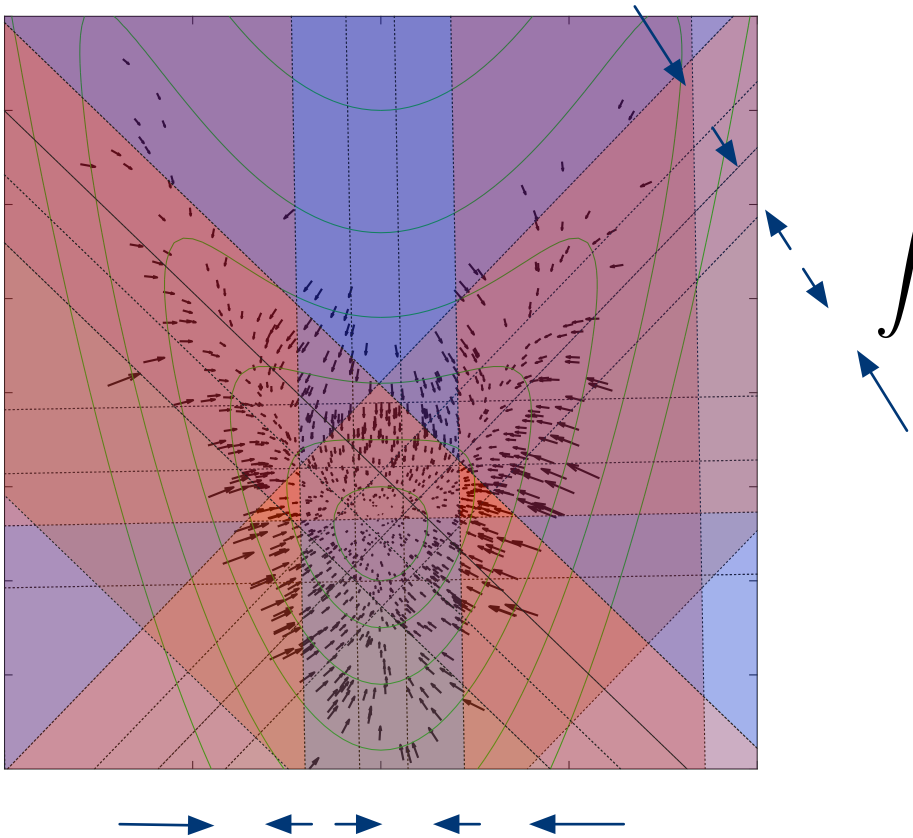


$$\tilde{\rho}_u(k) = \int dx'\, e^{-ikx'} \rho_u(x')$$

$$= \int dx \rho(x) e^{-ik(u\cdot x)}$$

- Fourier transform of Radon transforms are related to axial lines in the Fourier transform of the high-dimensional density

Score slice theorem

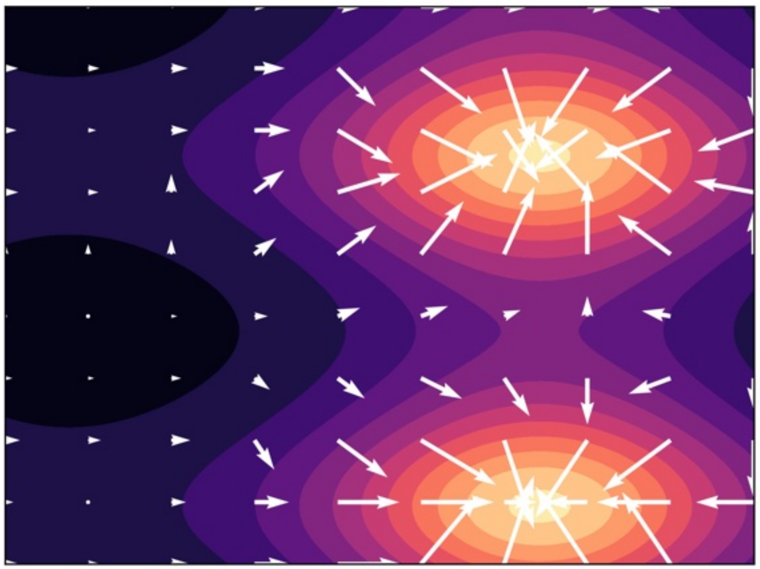


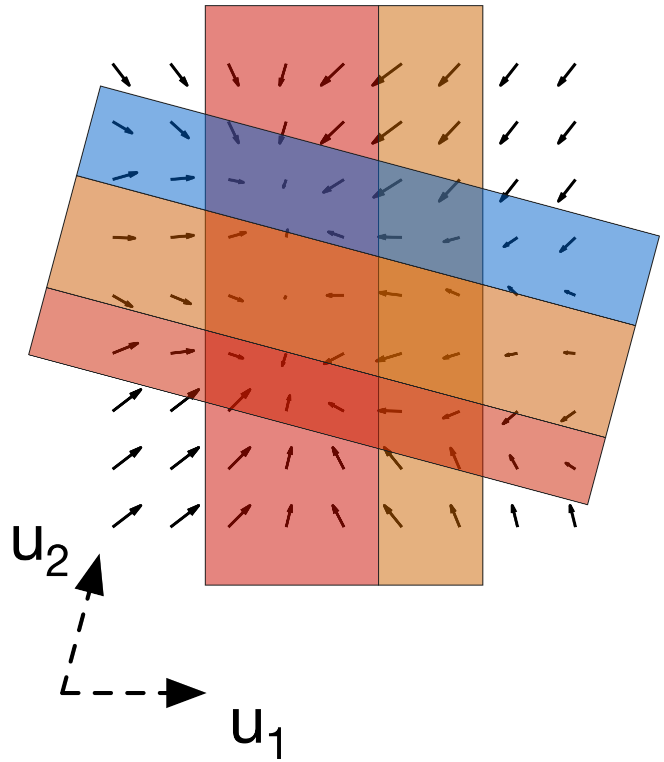
$$\int dx \rho(x)\phi(x)\delta(x' - u \cdot x)$$

$$= u\rho_u(x')\phi_u(x')$$

- Theorem about score functions under Radon transforms

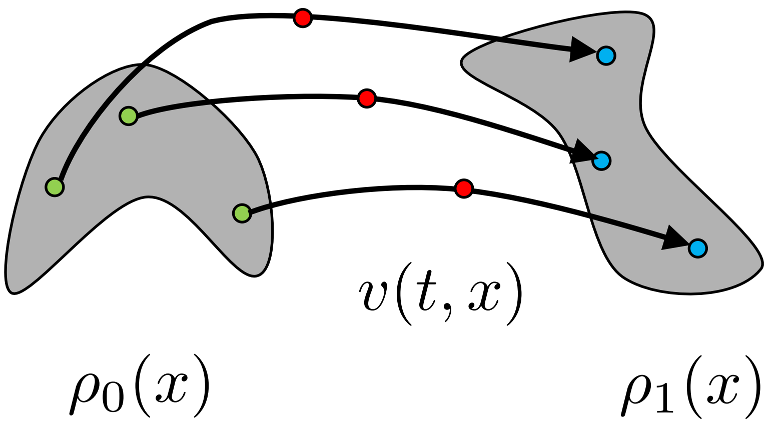
Score approximation

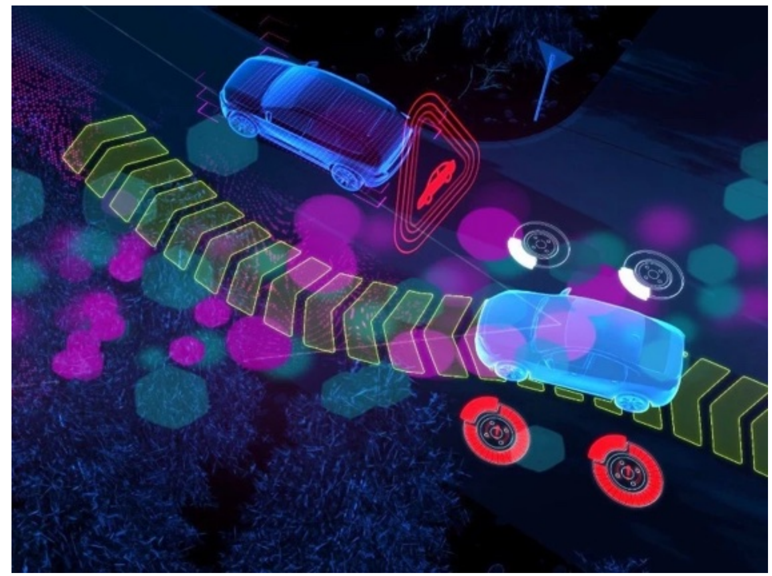




- Any score function can be arbitrarily approximated with superposition of 1-D electric fields

Flow as Control



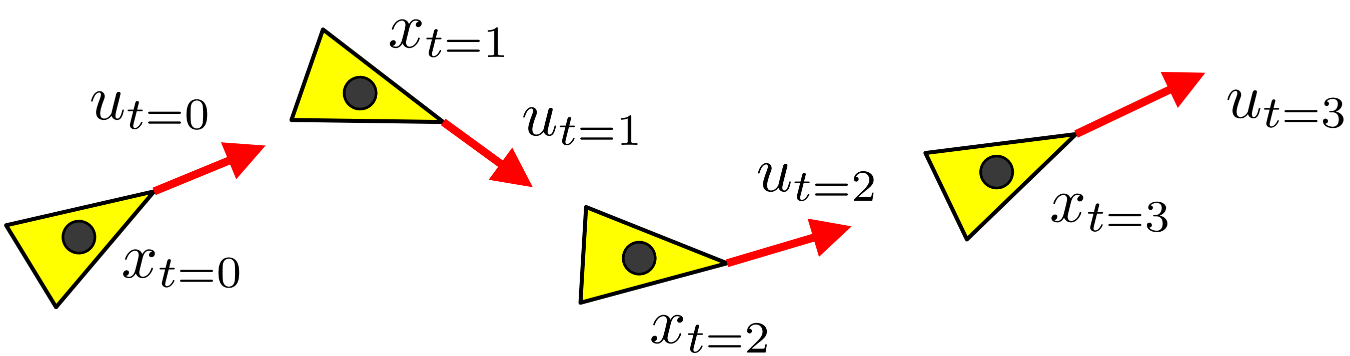


- Controlling flow from one distribution to another is analogous to steering individual particles to goal positions

Discrete time system

Discrete time deterministic dynamical system:

$$x_{t+1} = x_t + f(t, x_t, u_t) \qquad t = 0, 1, \ldots, T-1$$



State vectors: $x_t \in \mathbb{R}^N$ Control actions: $u_t \in \mathbb{R}^M$

- Dynamics describe evolution of state vectors in time

Optimal control

$$x_{t+1} = x_t + f(t, x_t, u_t) \qquad t = 0, 1, \ldots, T-1$$

State vectors: $x_t \in \mathbb{R}^N$ Control actions: $u_t \in \mathbb{R}^M$

Given initial state $x_{t=0}$ and controls $u_{0:T-1} = u_0, u_1, \ldots, u_{T-1}$

we can compute states $x_{1:T}$

Finite time horizon cost:

$$J(x_0, u_{0:T-1}) = \phi(x_T) + \sum_{t=0}^{T-1} R(t, x_t, u_t)$$

- Problem of optimal control is to find sequence $u_{0:T-1}$ that minimizes $J(x_0, u_{0:T-1})$

Naïve algorithm

$$J(x_0, u_{0:T-1}) = \phi(x_T) + \sum_{t=0}^{T-1} R(t, x_t, u_t)$$

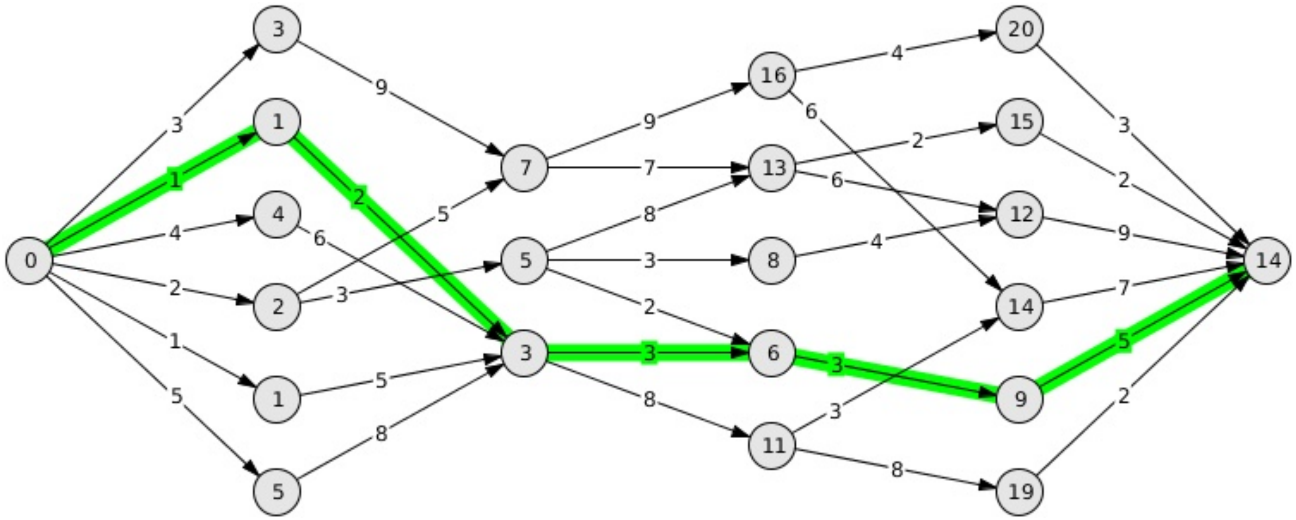
$$\min_{u_{0:T-1}} J(x_0, u_{0:T-1})$$

Evaluate cost function over all possible control vectors: $u_{0:T-1} = u_0, u_1, \ldots, u_{T-1}$

Time complexity: $O(N_u^T)$

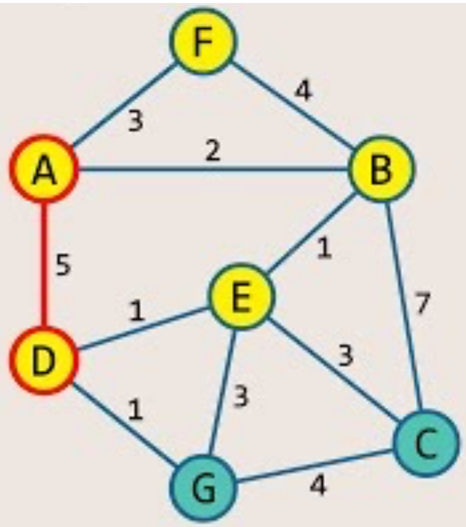
- Exponential time complexity is intractable for most problems

Principle of optimality



- Principle of Optimality: An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision. (Bellman, 1957, Chap. III.3.)

Dynamic programming



Node	Cost	Previous
A	0	
B	2	A
C	6	E
D	4	E
E	3	B
F	3	A
G	6	E

- Dijkstra’s algorithm recursively computes cost-to-go function to determine shortest paths

Optimal cost-to-go

$$J(x_0, u_{0:T-1}) = \phi(x_T) + \sum_{t=0}^{T-1} R(t, x_t, u_t)$$

Optimal cost-to-go (value) function:

$$V(t, x_t) = \min_{u_{t:T-1}} \left[\phi(x_T) + \sum_{s=t}^{T-1} R(s, x_s, u_s) \right]$$

Solves optimal control problem from an intermediate time and states until terminal time T

$$V(T, x) = \phi(x)$$

$$V(0, x) = \min_{u_{0:T-1}} C(x, u_t)$$

Bellman equation

Recursion relation:

$$V(t,x_t) = \min_{u_{t:T-1}} \left[\phi(x_T) + \sum_{s=t}^{T-1} R(s,x_s,u_s)\right]$$

$$= \min_{u_t} \left[R(t,x_t,u_t) + \min_{u_{t+1:T-1}} \left(\phi(x_T) + \sum_{s=t+1}^{T-1} R(s,x_s,u_s)\right)\right]$$

$$= \min_{u_t} \left[R(t,x_t,u_t) + V(t+1,x_{t+1})\right]$$

$$= \min_{u_t} \left[R(t,x_t,u_t) + V(t+1,x_t + f(t,x_t,u_t))\right]$$

- This is called the Bellman equation which can compute the optimal cost-to-go for all intermediate times and states

Optimal control policy

Initialization:

$$V(T,x) = \phi(x)$$

Backwards: $t = T-1,\ldots,0$

$$u_t^\star(x) = \arg\min_u \left[R(t,x,u) + V(t+1, x + f(t,x,u))\right]$$

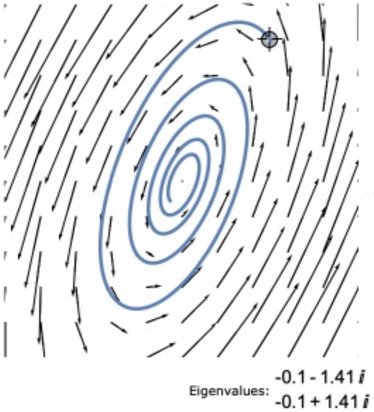
$$V(t,x) = R(t,x,u_t^\star(x)) + V(t+1, x + f(x, u_t^\star(x)))$$

Forwards: $t = 0,\ldots,T-1$

$$x_{t+1}^\star = x_t^\star + f(t, x_t^\star, u_t^\star(x_t^\star))$$

Linear system example

$$x_{t+1} = Ax_t + Bu_t$$



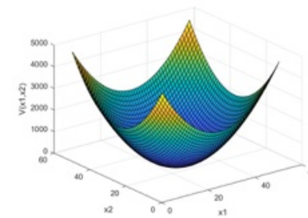
Quadratic cost function:

$$C = x_T^\top Q_T x_T + \sum_{s=0}^{T-1} x_s^\top Q x_s + u_s^\top R u_s$$

- This is called the Linear-Quadratic Regulator (LQR) control problem

Riccati equation

Quadratic cost-to-go function:



$$V(t,x) = x^\top P_t x$$

Initialization: $\quad V(T,x) = x^\top Q_T x \quad P_T = Q_T$

$$V(t,x) = \min_u \left[x^\top Q x + u^\top R u + V(t+1, Ax+Bu)\right]$$

$$P_t = Q + A^\top P_{t+1} A - A^\top P_{t+1} B (R + B^\top P_{t+1} B)^{-1} B^\top P_{t+1} A$$

- Recursive solution to Bellman equation for LQR is known as the Riccati equation

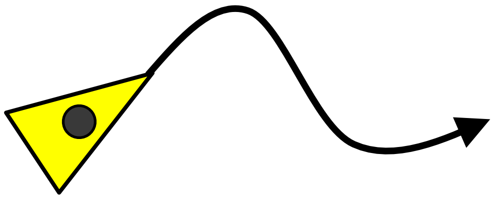
Continuous time

Take the continuous time limit: $\Delta t \to 0$

$$x_{t+\Delta t} = x_t + f(t, x_t, u_t)\Delta t$$

Deterministic ODE:

$$dx_t = f(t, x_t, u_t)dt$$



Integrated cost:

$$J(x_0, u_{0\to T}) = \phi(x_T) + \int_0^T d\tau\, R(\tau, x(\tau), u(\tau))$$

- Control problem is to find optimal controls $u^\star(t)$ that minimize cost

Hamilton-Jacobi

Value function:

$$V(t,x) = \min_u \left[R(t,x,u)dt + V(t+dt, x+f(t,x,u)dt)\right]$$

$$\approx \min_u \left[R(t,x,u)dt + V(t,x) + \partial_t V(t,x)dt + \partial_x V(t,x) f(t,x,u)dt\right]$$

$$\boxed{-\partial_t V(t,x) = \min_u \left[R(t,x,u) + \partial_x V(t,x) f(t,x,u)\right]}$$

Boundary condition: $V(T,x) = \phi(x)$

- Hamilton-Jacobi-Bellman (HJB) is a nonlinear PDE that solves *V(t,x)* backwards for all intermediate time and states

Optimal HJB policy

Initialization:

$$V(T, x) = \phi(x)$$

Backwards PDE: $\quad t = T \to 0$

$$u_t(x) = \arg\min_u \left[R(t, x, u) + \partial_x V(t, x) f(t, x, u)\right]$$

$$-\partial_t V(t, x) = R(t, x, u) + \partial_x V(t, x) f(t, x, u_t^\star(x))$$

Forwards ODE: $\quad t = 0 \to T$

$$x_{t=0}^\star = x_0$$

$$dx_t^\star = f(t, x_t^\star, u_t^\star(x_t^\star))dt$$

Continuous time LQR

Linear ODE: $\dot{x} = Ax_t + Bu_t$

Quadratic cost: $J = x_T^\top Q_T x_T + \int_0^T d\tau \, (x_\tau^\top Q x_\tau + u_\tau^\top R u_\tau)$

Value function: $V(t, x) = x^\top P_t x$

Initialization: $P_T = Q_T$

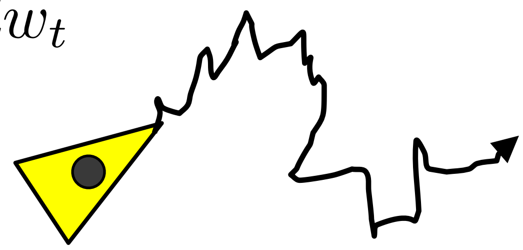
$$-\dot{P}_t = A^\top P_t + P_t A - P_t B R^{-1} B^\top P_t + Q$$

- Continuous-time Riccati equation is matrix ODE running backwards in time

Stochastic dynamics

Stochastic differential equation (SDE) dynamics:

$$dx_t = f(t, x_t, u_t)dt + \nu_t dw_t$$



Wiener process (Brownian noise):

$$dw_t \sim \mathcal{N}(0, I\, dt)$$

Cost becomes an expectation over random variables:

$$J(x_0, u) = \left\langle \phi(x_T) + \int_0^T d\tau\, R(t, x_t, u_t) \right\rangle$$

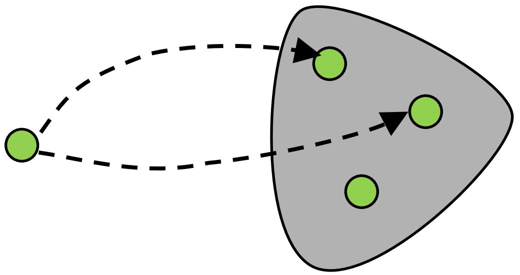
Belief state

SDE dynamics:

$$dx_t = f(t, x_t, u_t)dt + \nu_t dw_t$$

Distribution over states:

$$\rho(t, x)$$



- Because of stochastic dynamics, need to track distribution of states (belief state) over time

HJB equation

Value function recursion:

$$V(t, x_t) = \min_{u_t} R(t, x_t, u_t) + \langle V(t + dt, x_{t+dt}) \rangle$$

$$\langle V(t + dt, x_{t+dt}) \rangle = \int dx_{t+dt} \mathcal{N}(x_{t+dt} | x_t, \nu_t dt) V(t + dt, x_{t+dt})$$

$$= V(t, x_t) + \partial_t V(t, x_t) dt + \langle dx \rangle \partial_x V(t, x_t) + \frac{1}{2} \langle dx^2 \rangle \partial_x^2 V(t, x_t)$$

$$\langle dx \rangle = f(t, x, u) dt \quad \langle dx^2 \rangle = \nu_t dt$$

$$-\partial_t V(t, x) = \min_u \left[R(t, x, u) + f(t, x, u) \partial_x V(t, x) + \frac{1}{2} \nu_t \partial_x^2 V(x, t) \right]$$

Boundary condition: $V(T, x) = \phi(x)$

Effect of noise

Fokker-Planck (forwards in time):

$$\partial_t \rho(t,x) = -\nabla \cdot [\rho(t,x) f(t,x,u)] + \frac{1}{2}\nu_t \nabla^2 \rho(t,x)$$

$$\rho(0,x) = \rho_0(x)$$

Hamilton-Jacobi-Bellman (backwards in time):

$$-\partial_t V(t,x) = \min_u \left[R(t,x,u) + f(t,x,u) \cdot \nabla V(t,x) + \frac{1}{2}\nu_t \nabla^2 V(x,t) \right]$$

$$V(T,x) = \phi(x)$$

- PDE evolution of belief state and value function contain Laplacian terms describe influence of noise

LQG example

System State and Measurement

$$\dot{\mathbf{x}}(t) = \mathbf{F}(t)\mathbf{x}(t) + \mathbf{G}(t)\mathbf{u}(t) + \mathbf{L}(t)\mathbf{w}(t)$$
$$\mathbf{z}(t) = \mathbf{H}\mathbf{x}(t) + \mathbf{n}(t)$$

Control Law

$$\mathbf{u}(t) = -\mathbf{C}(t)\hat{\mathbf{x}}(t) + \mathbf{C}_F(t)\mathbf{y}_C(t)$$

State Estimate

$$\dot{\hat{\mathbf{x}}}(t) = \mathbf{F}(t)\hat{\mathbf{x}}(t) + \mathbf{G}(t)\mathbf{u}(t) + \mathbf{K}(t)\left[\mathbf{z}(t) - \mathbf{H}\hat{\mathbf{x}}(t)\right]$$
$$= \left[\mathbf{F}(t) - \mathbf{G}(t)\mathbf{C}(t) - \mathbf{K}(t)\mathbf{H}\right]\hat{\mathbf{x}}(t) + \mathbf{G}(t)\mathbf{C}_F(t)\mathbf{y}_C(t) + \mathbf{K}(t)\mathbf{z}(t)$$

Estimator Gain and State Covariance Estimate

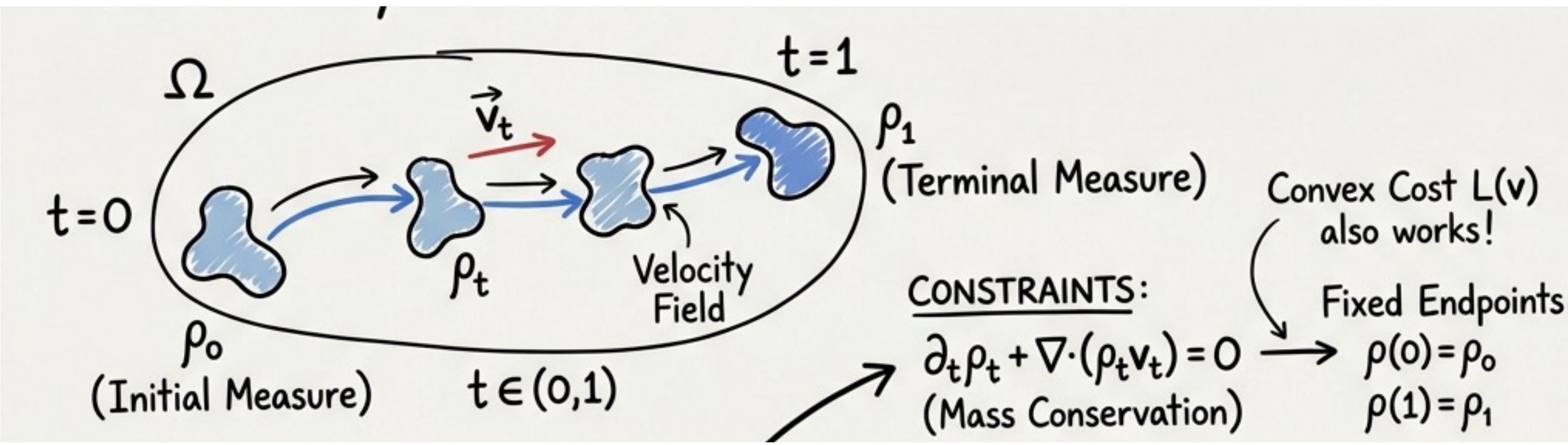
$$\mathbf{K}(t) = \mathbf{P}(t)\mathbf{H}^T\mathbf{N}^{-1}(t)$$
$$\dot{\mathbf{P}}(t) = \mathbf{F}(t)\mathbf{P}(t) + \mathbf{P}(t)\mathbf{F}^T(t) + \mathbf{L}(t)\mathbf{W}(t)\mathbf{L}^T(t) - \mathbf{P}(t)\mathbf{H}^T\mathbf{N}^{-1}(t)\mathbf{H}\mathbf{P}(t)$$

Control Gain and Adjoint Covariance Estimate

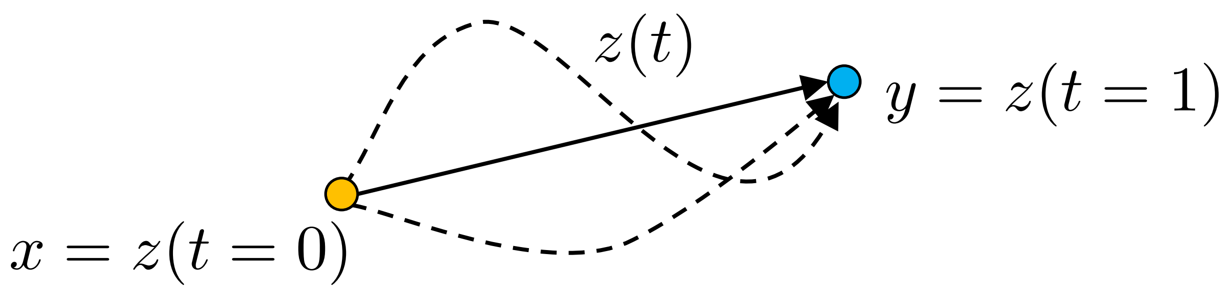
$$\mathbf{C}(t) = \mathbf{R}^{-1}(t)\mathbf{G}^T(t)\mathbf{S}(t)$$
$$\dot{\mathbf{S}}(t) = -\mathbf{Q}(t) - \mathbf{F}(t)^T\mathbf{S}(t) - \mathbf{S}(t)\mathbf{F}(t) + \mathbf{S}(t)\mathbf{G}(t)\mathbf{R}^{-1}(t)\mathbf{G}^T(t)\mathbf{S}(t)$$

- Linear-Quadratic-Gaussian: Linear dynamics, Quadratic cost function, Gaussian (Brownian) noise

Dynamic optimal transport



Dynamic formulation



$$\|y - x\|^2 = \int_0^1 dt \int Dz_t \delta\left(z_t - [ty + (1-t)x]\right) \|\dot{z}_t\|^2$$

$$= \min_{z(t)} \int_0^1 dt \, \|\dot{z}_t\|^2$$

- Rewrite Euclidean cost as minimum over paths with constrained endpoints

Dynamic formulation

$$\|y-x\|^2 = \int_0^1 dt \int Dz_t \delta\left(z_t - [ty + (1-t)x]\right) \|\dot{z}_t\|^2$$

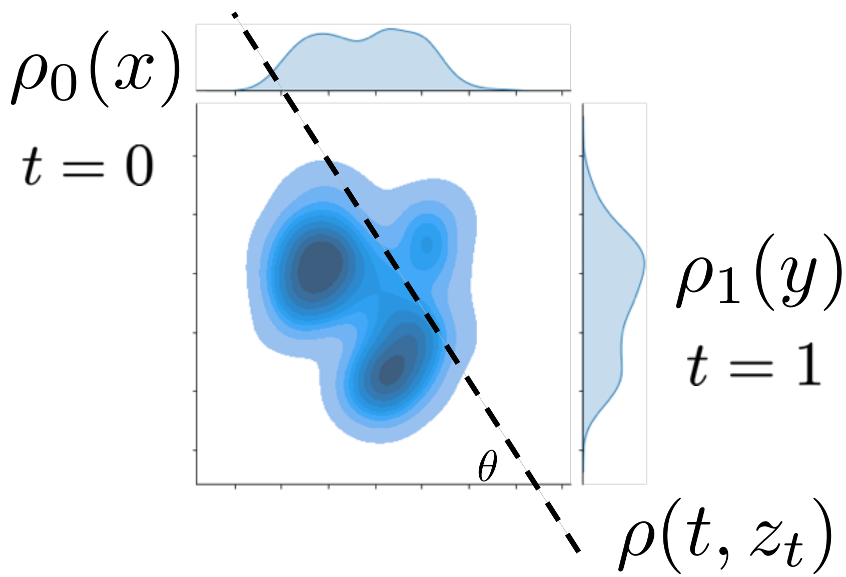
$$\int dx\, dy\, \rho(x,y) \|y-x\|^2 = \int_0^1 dt \int Dz_t\, \rho(t, z_t) \|\dot{z}_t\|^2$$

$$\rho(t, z_t) = \int dx\, dy\, \rho(x,y) \delta\left(z_t - [ty + (1-t)x]\right)$$

- Derive dynamic formulation with delta-function substitution and change order of integration

Dynamic density distribution

$$\rho(t, z_t) = \int dx\, dy\, \rho(x, y)\delta\left(z_t - [ty + (1-t)x]\right)$$



- Can view time-dependent distribution as marginalization across diagonal line in joint distribution function

Benamou-Brenier formulation

$$\min_{(\rho,v)} \int_0^1 dt \int \frac{1}{2} \|v(t,x)\|^2 \rho(t,x) dx$$

Constraints:

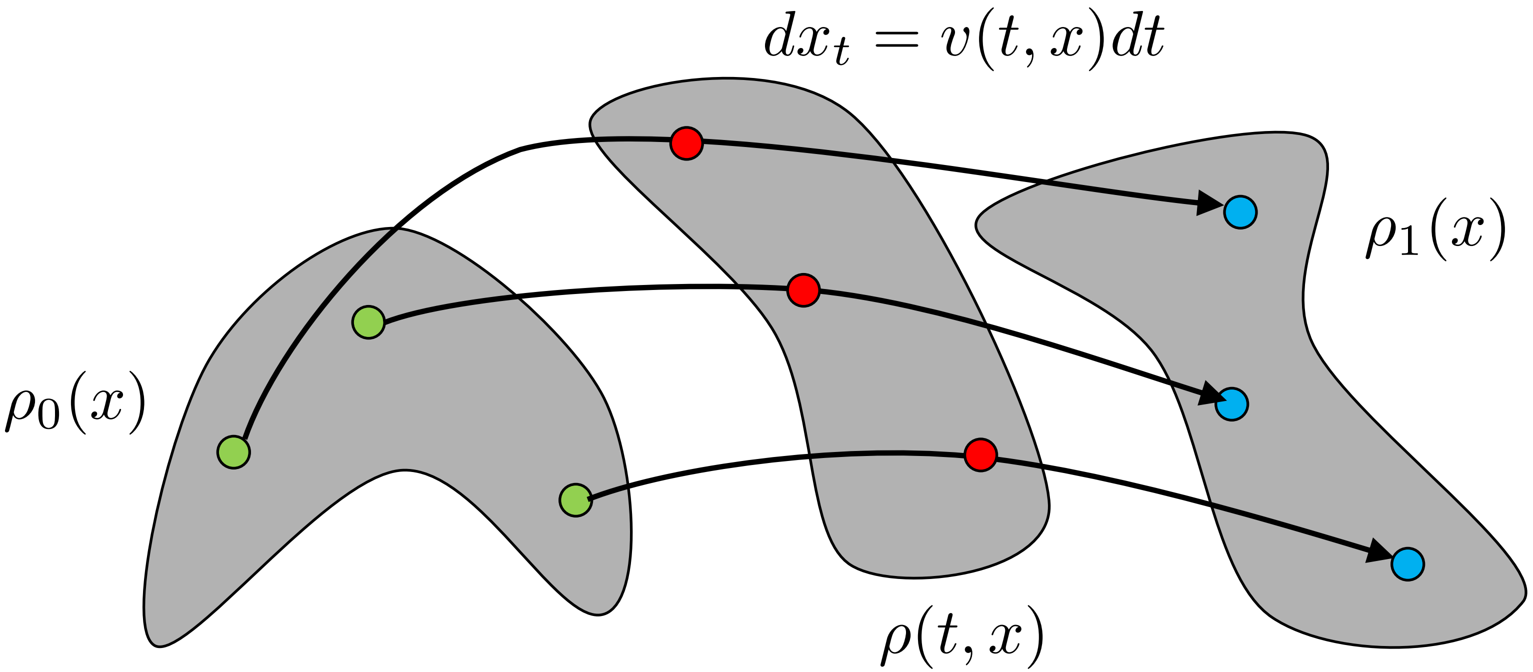
$$\frac{\partial \rho}{\partial t} + \nabla \cdot (v\rho) = 0$$

$$\rho(0,x) = \rho_0(x)$$

$$\rho(1,x) = \rho_1(x)$$

- Solution is equivalent to Monge-Kantorovich static optimal transport problem

Benamou-Brenier dynamics



$$\min_{(\rho,v)} \int_0^1 dt \int \frac{1}{2}\|v(t,x)\|^2 \rho(t,x) dx$$

Optimal control

Lagrangian with constraints:

$$\mathcal{L}(\rho, v) = \int_0^1 dt \int \left[\frac{1}{2} \|v(t,x)\|^2 \rho(t,x) + \lambda(t,x) \left(\frac{\partial \rho}{\partial t} + \nabla \cdot (v\rho) \right) \right] dx$$

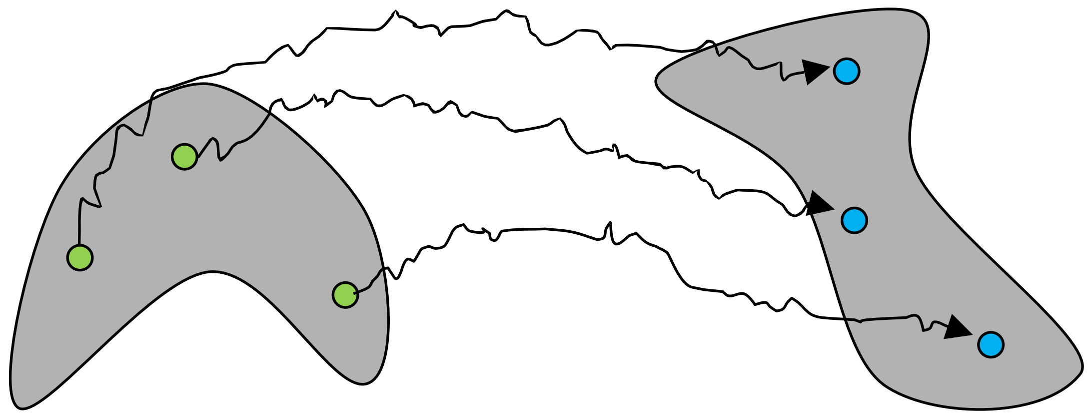
$$= \int_0^1 dt \int \left[\frac{1}{2} \|v(t,x)\|^2 - \left(\frac{\partial \lambda}{\partial t} + \nabla \lambda \cdot v \right) \right] \rho(t,x) dx$$

$$v^\star(t,x) = \nabla \lambda(t,x) \qquad \frac{\partial \lambda}{\partial t} + \frac{1}{2} \|\nabla \lambda\|^2 = 0$$

- Optimal flow field is gradient of value function satisfying HJB equation

Schrödinger problem (1931/32)

Erwin Schrödinger



- How to discover forces on hot gas particles consistent with observed particle distribution?

Schrödinger bridge

SDE dynamics: $dx_t = v(t,x)dt + \epsilon_t dw_t$

$$\min_{(\rho,v)} \int_0^1 dt \int \frac{1}{2}\|v(t,x)\|^2 \rho(t,x) dx$$

Constraints: $\dfrac{\partial \rho}{\partial t} + \nabla \cdot (v\rho) - \dfrac{1}{2}\epsilon_t \nabla^2 \rho = 0$

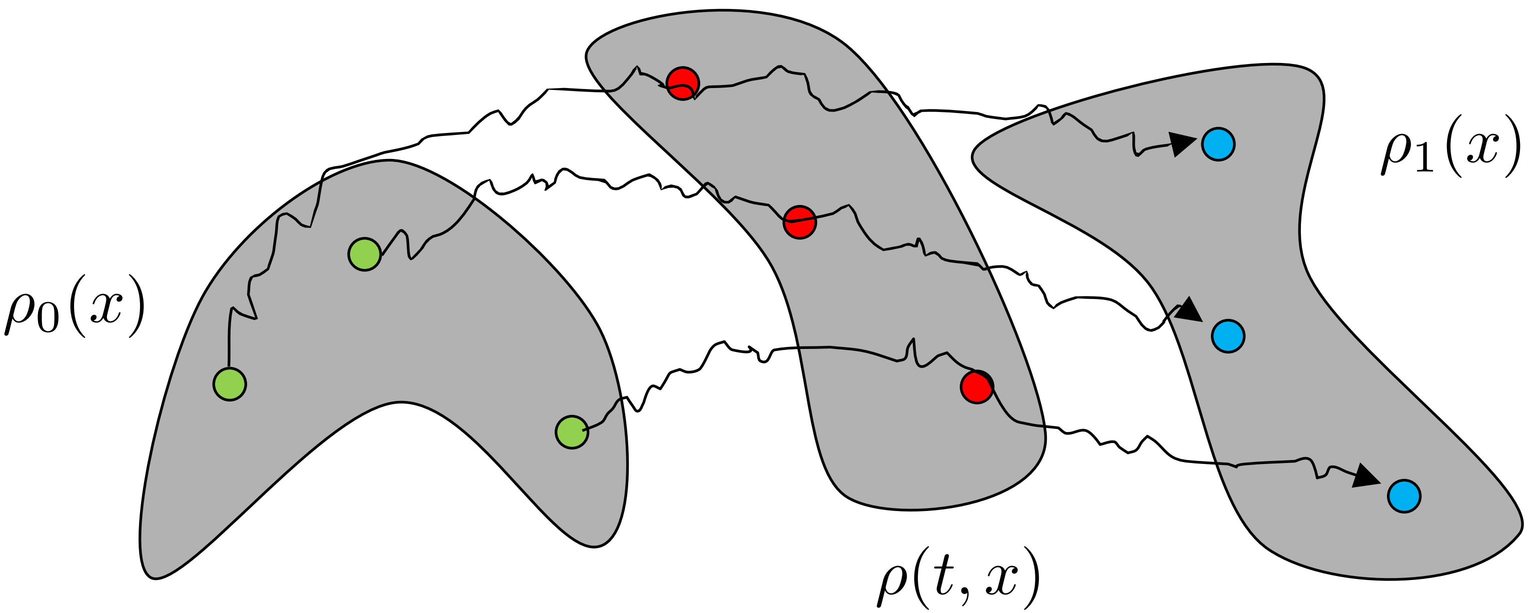
$$\rho(0,x) = \rho_0(x)$$

$$\rho(1,x) = \rho_1(x)$$

- Optimize flow field with Fokker-Planck equation constraints and initial and final distributions

Schrödinger bridge dynamics

$$dx_t = [v_0(t,x) + v(t,x)]dt + \epsilon_t dw_t$$



- Dynamics include prior on flow field and stochastic Brownian noise

Entropy regularized solution

$$\min_{(\rho,v)} \int_0^1 dt \int \frac{1}{2}\|v(t,x)\|^2 \rho(t,x) dx$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot ((v+v_0)\rho) - \frac{1}{2}\epsilon_t \nabla^2 \rho = 0 \quad \longleftrightarrow \quad \min_{\rho(x,y)} \int dx\, dy\, \rho(x,y)\|y-x\|^2 + \varepsilon D_{KL}[\rho(x,y)\|\rho_0(x)\rho_1(y)]$$

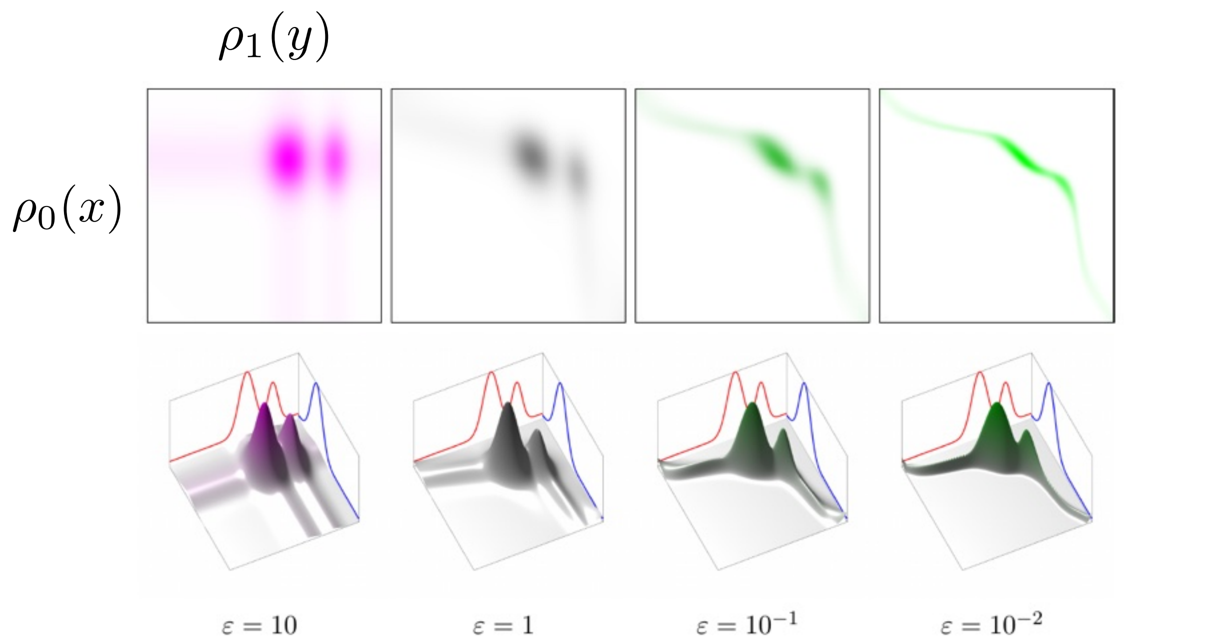
$$\rho(0,x) = \rho_0(x)$$

$$\rho(1,x) = \rho_1(x)$$

- Schrödinger bridge is equivalent to entropy-regularized optimal transport

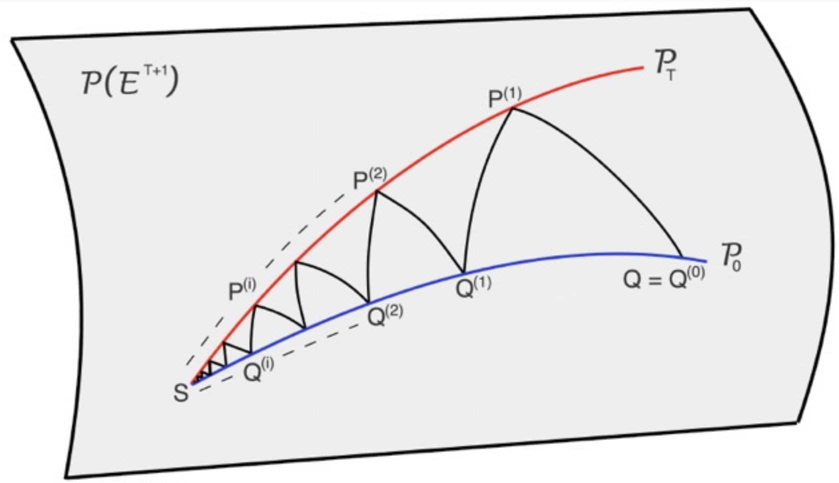
Entropy regularization

$$\min_{\rho(x,y)} \int dx\, dy\, \rho(x,y) \|y-x\|^2 + \varepsilon D_{KL}[\rho(x,y) \| \rho_0(x)\rho_1(y)]$$



- Regularization for smooth solutions to optimization

Iterative projection methods



$$\rho(0, x) = \rho_0(x)$$

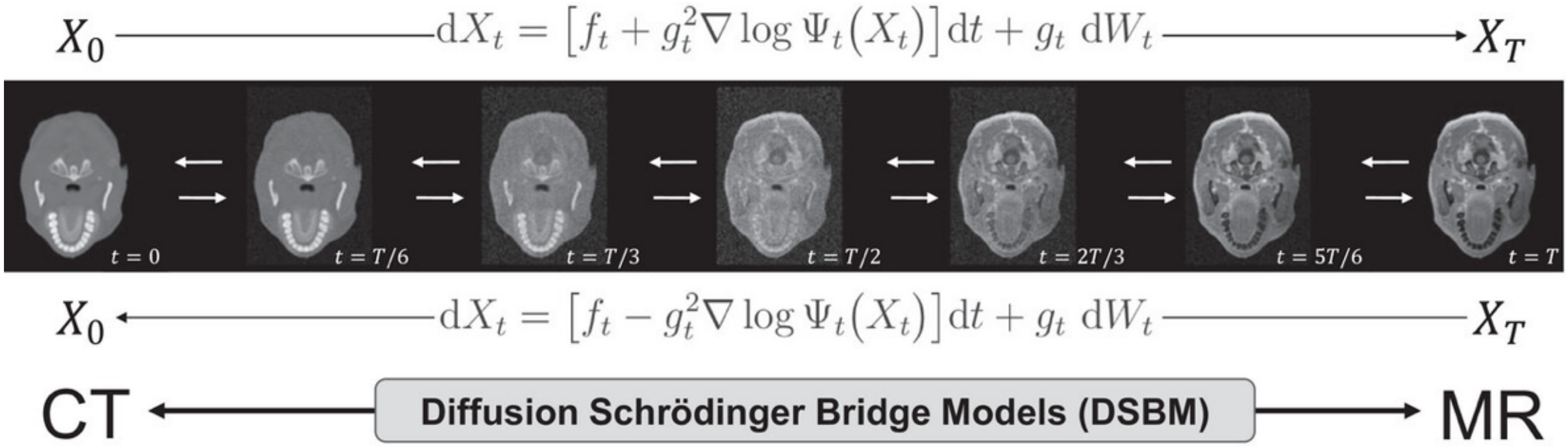
$$\rho(1, x) = \rho_1(x)$$

- Iterative Proportional Fitting (Sinkhorn) algorithm alternatively projects distribution onto initial and final marginal constraints (half-bridge solutions)

Applications

(a) Cat $\rightarrow$ Wild (b) Wild $\rightarrow$ Cat

De Bortoli et. al. (2024)



Li et. al. (2025)

Summary

- Optimal transport to map one distribution to another distribution
- Analogy between score functions and electric fields
- Universal approximation with charged slabs via Radon transform
- Connections to optimal control and Schrodinger gas problem