CONTINUOUS NORMALIZING FLOW PART II

Table of Contents

$p(x(t), t)$

Take the derivative of $\frac{d}{dt} \log p_t(x_t)$

$$\frac{d}{dt} \log p_t(x_t) = \frac{\partial}{\partial t} \log p_t(x) \Big|_{x=x_t} + \nabla_x \log p_t(x_t, t) \frac{dx_t}{dt}$$

$$= \frac{\partial}{\partial t} \log p_t(x) \Big|_{x=x_t} + \nabla_x \log p_t(x_t, t) \, f_\theta(x_t, t)$$

How does log density of x evolve as we push it through a flow?

Recall $\frac{dx}{dt} = f_\theta(x_t, t)$

Why the Total Derivative?

Because 1) The density field p_t changes wrt time $\left(\frac{\partial}{\partial t}\right)$

2) The sample x_t also moves through space as time evolves $\left(\nabla_x \cdot \frac{dx}{dt}\right)$

Now we have the formula

$$\boxed{\frac{d}{dt} \log p_t(x_t) = \frac{\partial}{\partial t} \log p_t(x) \Big|_{x=x_t} + \nabla_x \log p_t(x_t, t) \, f_\theta(x_t, t)}$$

$$\frac{d}{dt}\log p_t(x_t) = \frac{\partial}{\partial t}\log p_t(x)\Big|_{x=x_t}$$

$$+\ \nabla_x \log p_t(x_t, t)\ f_\theta(x_t, t)$$

continuity equation

$$\frac{\partial}{\partial t}\lg p_t(x) = -\nabla_x f_\theta(x_t, t)$$

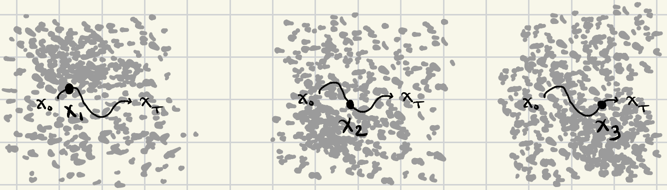
$$- f_\theta(x_t, t)\,\nabla_x \lg p_t(x)$$

$$= \left[-\nabla_x f_\theta(x_t, t) \cancel{- f_\theta(x_t, t)\nabla_x \lg p_t(x)}\right]_{x=x_t} \cancel{+ \nabla_x \lg p_t(x_t, t)\, f_\theta(x_t, t)}$$

$$\therefore\ \boxed{\frac{d}{dt}\log p_t(x_t) = -\nabla_x \cdot f_\theta(x_t, t)}$$

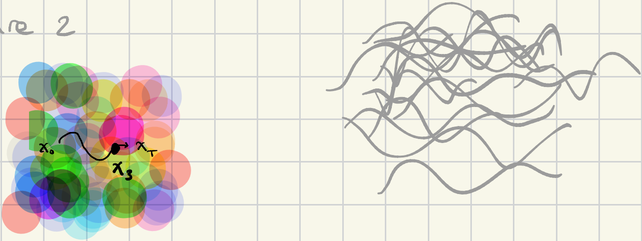
Figure 1

$p_t(x) \neq p_t(x_t)$ $\qquad$ p_t



$t = 1$ $\qquad$ $t = 2$ $\qquad$ $t = 3$

Figure 2



How Do We Get $\frac{\partial}{\partial t} p_t(x)$?

Conservation of Mass (Continuous)

Mass (particles) cannot vanish or appear

Continuity Equation (Fluid Dynamics)

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

local roc of ρ

dot product

divergence of flux

$\rho(x, t)$ density at point x and time t

$v(x, t)$ the particle velocity field

State	mass
+ diverge	out
− converge	in

Flux = flow of mass

Continuity Equation using PDFs

$$\frac{\partial}{\partial t} p_t(x) + \nabla_x \cdot (f_\theta(x_t, t) p_t(x)) = 0$$

$$\Longrightarrow \frac{\partial}{\partial t} p_t(x) = -\nabla_x (f_\theta(x_t, t) p_t(x))$$

by chain rule

log density → $$\frac{\partial}{\partial t} \log p_t(x) = \frac{\partial}{\partial t} \log p_t(x) \frac{\partial p_t(x)}{\partial t}$$

$$= \frac{1}{p_t(x)} \frac{\partial}{\partial t} p_t(x)$$

$$= \frac{1}{p_t(x)} \left[-\nabla_x \cdot \left(f_\theta(x_t, t) p_t(x) \right) \right]$$

$$= -\frac{1}{p_t(x)} \left(\nabla_x f_\theta(x_t, t) p_t(x) + f_\theta(x_t, t) \nabla_x p_t(x) \right)$$

product rule

$$= - \frac{\nabla_x \cdot f_\theta(x_t, t)\, \cancel{p_t(x)}}{\cancel{p_t(x)}} - \frac{f_\theta(x_t, t)\, \nabla_x \cdot p_t(x)}{p_t(x)}$$

$$= - \nabla_x \cdot f_\theta(x_t, t) - \frac{f_\theta(x_t, t)\, \nabla_x \cdot p_t(x)}{p_t(x)}$$

∇_x is a vector of differential operators

$$\nabla_x = \left(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_d}\right)$$

$$\nabla_x g(x) = \sum \frac{\partial g_i(x)}{\partial x_i}$$

$$\frac{\nabla_x p_t(x)}{p_t(x)} = \frac{1}{p_t(x)} \nabla_x p_t(x)$$

$$= \nabla_x \log p_t(x)$$

Multivariate Chain Rule

$$\nabla_x (g \circ h)(x) = g'(h(x)) \nabla_x h(x)$$

$$\frac{\partial}{\partial t} \log p_t(x) = - \nabla_x \cdot f_\theta(x_t, t) - \frac{f_\theta(x_t, t)\, \nabla_x \cdot p_t(x)}{p_t(x)}$$

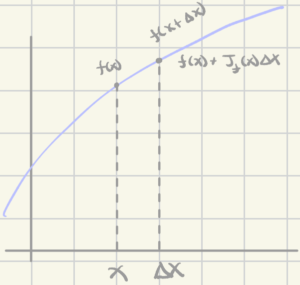
$$\boxed{\frac{\partial}{\partial t} \log p_t(x) = - \nabla_x \cdot f_\theta(x_t, t) - f_\theta(x_t, t)\, \nabla_x \log p_t(x)}$$

Jacobian & Trace

$x \in \mathbb{R}^d \quad f : \mathbb{R}^d \to \mathbb{R}^d$

Jacobian: linear approximation of f around a small neighborhood of x

$$\boxed{f(x + \Delta x) \approx f(x) + J_f(x) \Delta x}$$



entry (i,i)

$$\frac{\partial f_i}{\partial x_i} \quad \begin{matrix} + \\ - \end{matrix}$$

axis i

Jacobian Matrix

$$J_f(x) = \frac{\partial f(x)}{\partial x}$$

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_d} \\ \vdots & \frac{\partial f_2}{\partial x_2} \searrow & \vdots \\ \frac{\partial f_d}{\partial x_1} & \cdots & \frac{\partial f_d}{\partial x_d} \end{pmatrix}$$

$\frac{\partial f_i}{\partial x_i} \rightarrow$ If you nudge x_i, how does the flow in the f_i direction speed up or slow down

Divergence

$$\nabla \cdot f(x) = \sum_{i=1}^{d} \frac{\partial f_i}{\partial x_i}$$

dot product

A Little More to Get Trace

Total Derivative

$$\frac{d}{dt} \log p_t(x_t) = -\nabla_x \cdot f_\theta(x_t, t)$$

$\mathbb{R}^d \cdot \mathbb{R}^d$

$$= -\sum \frac{\partial f_i(x,t)}{\partial x_i} = -\mathrm{Tr}\left(\frac{\partial f_\theta}{\partial x_t}\right)$$

Off diagonals of Jacobian tell us rotation / sheer

The diagonals of the Jacobian are the only entries of the Jacobian that directly tells us how volumes expand/contract along each axis

$$\boxed{\begin{array}{l}\text{Fundamental Theorem of Calculus (FTC)} \\ F'(x) = f(x) \Rightarrow F(b) - F(a) = \int_a^b f(x)\,dx \\ \text{equivalently} \\ \frac{dy}{dt} = g(t) \Rightarrow y(T) - y(0) = \int_0^T g(t)\,dt \end{array}}$$

So now we have

$$\frac{d}{dt} \log p_t(x_t) = -\nabla_x \cdot f_\theta(x_t, t) = -\mathrm{Tr}\left(\frac{\partial f_\theta}{\partial x_t}\right)$$

And

$$\log p(x_T) - \log p(x_0) = -\int_0^T \mathrm{Tr}\left(\frac{\partial f_\theta}{\partial x_t}\right)$$

By FTC

$$\boxed{\begin{array}{l}\text{Instantaneous Change-of-variables} \\ \text{in continuous time} \\ \log p(x_0) = \log p(x_T) - \int_0^T \mathrm{Tr}\left(\frac{\partial f_\theta}{\partial x_t}\right)\end{array}}$$

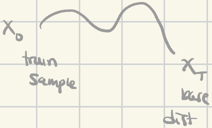
Problem: Trace is the computational bottleneck!

Solution:
Hutchison Estimator → still expensive

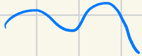
Objective/Loss Function

Let x_T be base density (Gaussian)
x_0 be a training image

$$\ell(x;\theta) = \log p(x_T) - \int_0^T \mathrm{Tr}\left(\frac{\partial f_\theta}{\partial x_t}\right)$$



$$\text{MLE} \quad \max_\theta \mathbb{E}_{x \sim \text{data}} [\log p(x)]$$



Training

- $x_0 \in \mathbb{R}^d$ sample
- Numerically integrate with ODE solver (PyTorch) to get x_T

$$x_T = x_0 + \int_0^T \underbrace{f_\theta}_{\text{NN}}(x_t, t)\, dt$$

FTC

$$\frac{dx}{dt} = f_\theta$$

- Choose simple base PDF & compute

$$\log p(x_T), \text{ where } x_T \sim N(0, I)$$

- Integrate with same ODE solver

$$\int_0^T \mathrm{Tr}\left(\frac{\partial f_\theta}{\partial x_t}\right) dt$$

- Compute as NLL & Back propagate

$$\log p(x_0) = \log p(x_T) - \int_0^T \mathrm{Tr}\left(\frac{\partial f_\theta}{\partial x_t}\right) dt$$