Introduction to Probability flows

Samples x are now evolving in time

- $\frac{dx}{dt} = v(t, x)$ ODE in $\mathbb{R}^d$

  Solution called Flow map: $x_0 \to \varphi^t(x_0) := x_t$

  $v(t, x)$ is a vector that gives the instantaneous change of sample path.

  We will write v_t to denote the function $x \to v(t, x)$

- If $x_0 \sim \rho_0$, then $x_t \sim \rho_t$ that satisfies

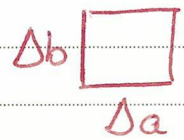
$$\partial_t \rho_t = -\mathrm{div}(\rho_t v_t)$$

This equation is called a continuity equation.

Derivation via mass conservation in 2D

Ex: Convince yourself that the derivation also works in any dimension

(2)

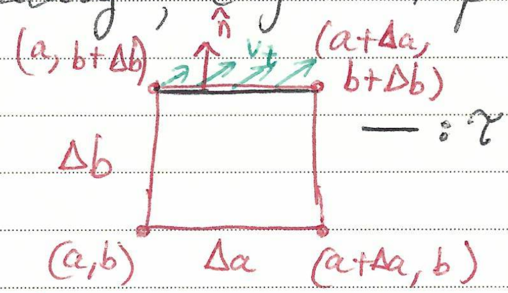


differential volume in 2D (small rectangle)

$x \equiv [a, b]$ in coordinates

probability Mass flowing into differential vol at time t due to evolution $\frac{dx}{dt} = v(t, x) = v_t(x)$

We will add mass flowing into each boundary, e.g. top boundary



$\hat{n}$: normal vector

v_t is shown in green at a few points on the top boundary.

$\hat{n}$ is also a function of x but is fixed or is a const vector at top boundary $= \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

Prob mass flowing in at top boundary in time $[t, t+\Delta t]$ $= -\int_t^{t+\Delta t}\int_\gamma v_t(x) \cdot \hat{n} \, dx \, dt'$

Since $\hat{n} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ at any x on $\mathcal{T}$ (top boundary)

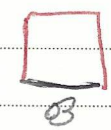
Total mass flow inward in time t to $t+\Delta t$

$$= -\int_t^{t+\Delta t}\int_{\mathcal{T}} v_{t'}^{(b)}(x)\, dx\, dt'$$

Here $v_t^{(b)}$ is the second component "in the b direction" of the vector $v_t(x)$.

There is a minus sign because we are accounting for mass coming in but $\hat{n}$ is pointing out

Similarly on the bottom boundary, $\mathcal{B}$ we will get the total mass flow inward into the box to be



$$\int_t^{t+\Delta t}\int_{\mathcal{B}} v_{t'}^{(b)}(x)\, dx\, dt'$$

④

No minus sign because the outward pointing normal is $\begin{bmatrix} 0 \\ -1 \end{bmatrix}$.

Total mass flowing in to diff vol in diff time Δt is:

$$\int_t^{t+\Delta t} \int_a^{a+\Delta a} \left(v_{t'}^{(b)}([a', b])\, da' - v_{t'}^{(b)}([a', b+\Delta b])\, da' \right) dt'$$

$$+ \int_t^{t+\Delta t} \int_b^{b+\Delta b} \left(v_{t'}^{(a)}([a,b])\, db' - v_{t'}^{(a)}([a+\Delta a, b'])\, db' \right) dt'$$

$$= \left(v_t^{(b)}([a,b]) - v_t^{(b)}([a, b+\Delta b]) \right) \Delta a\, \Delta t$$

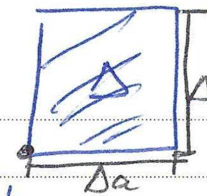
$$+ \left(v_t^{(a)}([a,b]) - v_t^{(a)}([a+\Delta a, b]) \right) \Delta b\, \Delta t$$

$$+ O\left((\Delta a)^2, (\Delta b)^2, (\Delta t)^2 \right)$$

Notice the equality, comes from Assuming smooth derivaties of v_t along t, a, b.

If this is the total amount of prob mass entering diff vol, it should be equal to how much the mass is increasing:

$[a,b] \equiv x$



$$\int_\Delta \rho_{t+\Delta t}(x)\,dx - \int_\Delta \rho_t(x)\,dx$$

$$= \rho_{t+\Delta t}(x)\,\Delta a\,\Delta b - \rho_t(x)\,\Delta a\,\Delta b$$

$$+\ O\left((\Delta a)^2, (\Delta b)^2\right)$$

Assuming that 1st & 2nd order partial derivatives of ρ_t & $\rho_{t+\Delta t}$ wrt a & b are bounded and small

Equating the two mass increases (mass conservation principle), we get

$$\left(\rho_{t+\Delta t}(x)\,\Delta a\,\Delta b - \rho_t(x)\,\Delta a\,\Delta b\right) + O\left((\Delta a)^2, (\Delta b)^2\right)$$

$$= \left(\rho_t v_t^{(b)}([a,b]) - \rho_t v_t^{(b)}([a, b+\Delta b])\,\Delta t\,\Delta a\right)$$

$$+\ O\left((\Delta a)^2, (\Delta b)^2\right) + \left(\rho_t v_t^{(a)}([a,b]) - \rho_t v_t^{(a)}([a+\Delta a, b])\,\Delta t\,\Delta b\right)$$

Dividing by $\Delta a \times \Delta b$ & taking lim $\Delta a, \Delta b \to 0$,

we get
$$\frac{\partial \rho_t(x)}{\partial t} = -\frac{\partial}{\partial b}\left(\rho_t v_t^{(b)}\right)(x) - \frac{\partial}{\partial a}\left(\rho_t v_t^{(a)}\right)(x)$$

Hint:
To derive this in higher dimensions, use the divergence theorem.

Applicability in Generative modeling:
if ρ_0 is an "easy" distribution like a Gaussian, then, we can find a vector field such that $\rho_T \equiv P_{data}$

i.e. $\frac{dx}{dt} = v_t(x)$

can be solved on the samples so that $\varphi^t(x) \sim \rho_t$ follows a desired path

$$\varphi^T(x) \sim P_{data}.$$

Neural ODEs: Idea is to parameterize v_t with a neural network

Then use ~~the~~ any time integration scheme to solve

Change of variables for discrete time evolutions:

$$x_t = F_{t-1} \circ \cdots F_1 \circ F_0 (x_0)$$

$$\rho_t(x_t) = \frac{\rho_0(x_0)}{|\det \nabla(F_{t-1} \circ \cdots F_0)|(x_0)}$$

$$\Rightarrow \log \rho_t(x_t) = \log \rho_0(x_0) - \sum_{k=0}^{t-1} \log |\det \nabla F_k(x_k)|$$

Neural ODEs Chen et al 2018 proposed as continuous-in-time extensions to ResNets.

In a ResNet: outputs from layer t are used to define inputs for next layer $t+1$

But they are important in their connection to sampling.