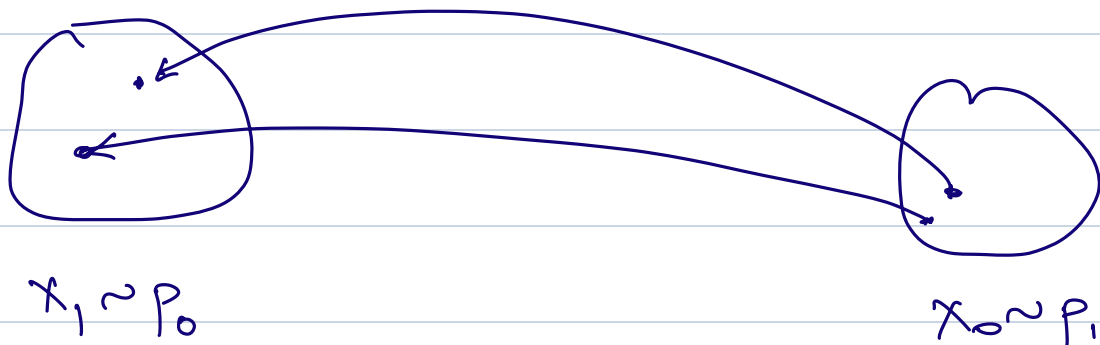


Rectified Flows

References:

- ① Liu et. al., Flowing straight and fast: Learning to generate and transfer data with rectified flow, ICLR'22
- ② Lipman et. al., Flow matching for generative modeling, 2022
- ③ Albergo and Vanden-Eijnden, Building normalizing flows with stochastic interpolants, ICLR'23
- ④ Ront et. al., Semantic image inversion and editing using stochastic rectified differential equations, ICLR'25.

Rectified Flow (RF): An ODE to transport samples from one dist. to another



(typically $P_0 \sim P_{\text{data}}$; dist. over images)

(typically, $P_1 \sim N(0, I)$)

Motivation: The ODE flow is "straighter" than that with DDIM, thus, coarser discretization suffices in implementation.

The ODE: $dX_t = v_t(X_t) dt$, $X_0 \sim p_1$
 $X_1 \sim p_0$
 $t \in [0, 1]$.

reverse process /
generative process

Since this is an ODE, as in DDIM, this can be reversed, with

$$dY_t = u_t(Y_t) dt, \quad Y_0 \sim p_0$$
$$Y_1 \sim p_1$$
$$t \in [0, 1]$$

forward process /
noising process

with $u_t(y) = -v_{1-t}(y)$

Thus, Y_t is the "forward process" that transports an image $Y_0 \sim p_0$ to noise $Y_1 \sim p_1$, that is typically Gaussian (i.e., $p_1 \sim N(0, I)$). This is the RF equivalent of the ODE corresponding to that of the OU process in diffusions.

Deriving the RF vector field for Generation:

If we had access to $X_0 \sim p_1$ ^(noise) and $X_1 \sim p_0$ ^(data), use a straight line path to interpolate between these two samples:

$$\text{i.e. } X_t = (1-t)X_0 + tX_1, \quad t \in [0, 1].$$

$$\Rightarrow \frac{dX_t}{dt} = -X_0 + X_1,$$

However, we need a causal ODE, i.e., we only have access to X_0, X_t and would like to flow toward X_1 , so find a field that does this "on average" over samples and time, i.e., find $v_t(\cdot)$ s.t.

$$\min_{v(\cdot)} E_{(X_0, X_1, T) \sim p_1 \times p_0 \times \text{Unif}[0, 1]} \left[\| (X_1 - X_0) - v_T(X_T) \|^2 \right]$$

$(1-t)X_0 + tX_1$

random variable T

$\rightarrow$ ①

Training RF: The above objective (also called

conditional flow matching in literature) can be used to optimize a flow model. Specifically, parameterize

$$V_t(x) \equiv V_\theta(t, x), \text{ where } \theta \text{ is learned from data.}$$

Thus, a model is trained as:

$$\min_{\theta} E_{(x_0, x, \tau) \sim (p_0 \times p, \times \text{Unif}[0, 1])} \left[\|x_1 - x_0 - V_\theta(\tau, x_\tau)\|^2 \right]$$

data dist $\swarrow$ noise $\searrow$ random time $\nearrow$
dist $(N(0, I))$

As we will see below, this is closely related to the diffusion model!

Deriving the Solution in terms of Score

The formulation in ① is a LMS problem, for

which we know that the conditional expectation is the optimal solution.

Given $T=t$, and $X_T = x_t$, we have:

$$V_t(x_t) = E[X_1 - X_0 | X_t = x_t]$$

The above is the reverse process that transports noise to data.

Similar to the discussion above, we can consider the forward process (transport data to noise).

Let us denote this vector field by $u_t(\cdot)$. Since these are ODEs, $u_t(x) = -V_{1-t}(x)$.

$$(Y_0, Y_1) \sim p_0 \times p_1, \quad Y_t = tY_1 + (1-t)Y_0, \\ \text{(data) (noise)} \quad t \in [0, 1]$$

As in the generative case, a causal approximation using the Least Mean Square Estimator leads to

a forward flow (transporting data to noise):

$$U_t(y_t) = E_{(\gamma_0, \gamma_1) \sim P_0 \times P_1} [\gamma_1 - \gamma_0 \mid \gamma_t = y_t]$$

This has a closed form expression, that is derived from Tweedie's formula:

$$U_t(y_t) = -\frac{y_t}{1-t} - \frac{t}{1-t} \nabla \ln p_t(y_t).$$

To see this, (see Lemma A.1 in reference (4)), recall Tweedie's formula:

Suppose $R_t = R_0 + \sigma Z$ $\xrightarrow{N(0, I)}$, then,

$$E[R_0 \mid R_t = r_t] = r_t + \sigma^2 \nabla \ln p_t(r_t).$$

$\therefore$ if $\gamma_t = \underbrace{(1-t)\gamma_0}_{R_0} + \underbrace{t\gamma_1}_{\substack{\omega \\ \sigma}}$, $\xrightarrow{N(0, I)}$, then

$$E[(1-t)\gamma_0 | \gamma_t = y_t] = y_t + t^2 \cdot \nabla \ln p_t(y_t)$$

$$\therefore E[\gamma_0 | \gamma_t = y_t] = \frac{y_t}{1-t} + \frac{t^2}{1-t} \nabla \ln p_t(y_t)$$

→ ①.

$$\text{Also, } \gamma_t = (1-t)\gamma_0 + t\gamma_1 \Rightarrow \gamma_1 = \frac{\gamma_t}{t} - \frac{(1-t)}{t}\gamma_0.$$

$$\therefore E[\gamma_1 | \gamma_t = y_t] = \frac{y_t}{t} - \left(\frac{1-t}{t}\right) E[\gamma_0 | \gamma_t = y]$$

$$= \frac{y_t}{t} - \frac{1-t}{t} \left(\frac{y_t}{1-t} + \frac{t^2}{1-t} \nabla \ln p_t(y_t) \right)$$

$$= \frac{y_t}{t} - \frac{y_t}{t} - t \nabla \ln p_t(y_t)$$

$$\therefore E[\gamma_1 | \gamma_t = y_t] = -t \nabla \ln p_t(y_t) \rightarrow \textcircled{2}$$

$$\therefore u_t(y_t) = E[\gamma_1 - \gamma_0 | \gamma_t = y_t]$$

$$= -\frac{1}{1-t} y_t - \frac{t}{1-t} \nabla \ln p_t(y_t). \quad \textcircled{3}$$

Now, from the continuity equation (the special case of the Fokker-Planck eqn when $\sigma \equiv 0$), we have the density evolution of this ODE with $Y_0 \sim P_0$ given by:

$$\frac{\partial P_t}{\partial t} =$$

$$\nabla \cdot \left[\left(\frac{1}{1-t} Y_t + \frac{t}{1-t} \nabla \ln P_t(Y_t) \right) P_t(Y_t) \right]$$

→ (3)

Finally, consider the SDE:

$$d\tilde{Y}_t = -\frac{1}{1-t} \tilde{Y}_t dt + \sqrt{\frac{2t}{1-t}} dB_t, \quad t \in [0, 1],$$

$Y_0 \sim P_0.$

→ (4)

Then, the F-P Equation results in the same density evolution as (3).

Thus (4) is the SDE equivalent of the ODE corresponding to the forward process of RF.

Just as the diffusions case, we can now reverse this process, and we get the following triplet for RF generation:

$$d\tilde{x}_t = \frac{1}{t} \tilde{x}_t + \frac{2(1-t)}{t} \nabla \ln p_{1-t}(\tilde{x}_t) + \sqrt{\frac{2(1-t)}{t}} dB_t$$

$$x_0 \sim p_1, \quad t \in [0, 1].$$

$\hookrightarrow$ (5) $\hookrightarrow$ (noise dist.)

$$dx_t = v_t(x_t) dt, \quad x_0 \sim p_1, \quad t \in [0, 1]$$

$\hookrightarrow$ (noise dist.)

where $v_t(\cdot) = -u_{1-t}(\cdot)$

$$= \frac{1}{t} y_t + \left(\frac{1-t}{t}\right) \nabla \ln p_{1-t}(y_t).$$

$\hookrightarrow$ (6)