DiffusionNFT as Taylor Approximation

Qiang Liu

Abstract

We show that DiffusionNFT [1] can be viewed as a kind of Taylor approximation to the reward-reweighted distributions with specific wrapper functions.

DiffusionNFT [1] can be viewed as learning a reward-tilted update around a pretrained velocity field. Let $v_t^0(x) = \mathbb{E}[X_1 - X_0 \mid X_t = x]$ be a pretrained velocity field trained on pairs $(X_0, X_1) \sim \pi_0 \times \pi_1$, with the standard linear interpolation $X_t = (1-t)X_0 + tX_1$. Assume we have an additional reward $r(x) \geq 0$, and we are interested in steering v_t^0 to reflect the preference of r.

DiffusionNFT introduces a learnable field $\mu_t(\cdot)$ and forms two symmetric perturbations around v_t^0:

$$\mu_t^+(x) := v_t^0(x) + \beta\big(\mu_t(x) - v_t^0(x)\big), \qquad \mu_t^-(x) := v_t^0(x) - \beta\big(\mu_t(x) - v_t^0(x)\big),$$

where $\beta > 0$. Then it fits μ by a reward-weighted regression:

$$L(\mu) := \mathbb{E}\Big[r(X_1)\left\|\mu_t^+(X_t) - (X_1 - X_0)\right\|^2 + (1 - r(X_1))\left\|\mu_t^-(X_t) - (X_1 - X_0)\right\|^2\Big].$$

Here the expectation is over $(X_0, X_1) \sim \gamma$ (a coupling of π_0 and π_1) and over $t \sim p(t)$ supported on $[0, 1]$, with $X_t = (1-t)X_0 + tX_1$. Intuitively, high-reward samples pull the policy toward v_t^+, while low-reward samples pull it toward v_t^-.

Let v_t^* be the minimum of $L(\mu)$, and $v_t^{*\pm}$ the corresponding perturbations. After training, we may sample from either v_t^* or the positive perturbation v_t^{*+}.

DiffusionNFT as Extrapolation As show in the paper, we can rewrite DiffusionNFT into

$$v_t^*(x) = v_t^0(x) + \frac{2}{\beta}\hat{m}_t^r(x)(\hat{v}_t^r(x) - v_t^0(x)), \qquad v_t^{*\pm}(x) = v_t^0(x) \pm 2\hat{m}_t^r(x)(\hat{v}_t^r(x) - v_t^0(x)),$$

where

$$\hat{v}_t^r(x) = \frac{\mathbb{E}[r(X_1)(X_1 - X_0) \mid X_t = x]}{\mathbb{E}[r(X_1) \mid X_t = x]}, \qquad \hat{m}_t^r(x) = \mathbb{E}[r(X_1) \mid X_t = x].$$

Here $\hat{v}_t^r$ is the rectified flow vector field of the reward-weighted distribution:

$$\hat{\pi}_1(x) = \frac{\pi_1(x)r(x)}{Z}, \qquad Z = \int \pi_1(x)r(x)\mathrm{d}x,$$

where we reweight the density by $r(x)$ (not exponential tilting that we have below).

Here, v_t^* and v_t^{*+} are extrapolation of the original v_t^0 and the $\hat{v}_t^r$, with a magnitude weighted by $\hat{m}_t^r$. Note that $\hat{m}_t^r(x) = \mathbb{E}[r(X_1) \mid X_t = x]$ is the conditional expected reward at the bridge state $X_t = x$.

However, it is unclear what distribution does DiffusionNFT sample from. There is generally no closed-form expression for the resulting distributions, and the learned velocity field does not correspond exactly to the rectified-flow (RF) field of an explicit tilted target distribution.

DiffusionNFT as Taylor Approximation But we can show that it is a kind of Taylor approximation to the exponential-tilted distribution. Let us consider the exponential-tilted distribution:

$$\tilde{\pi}_1^\alpha(x) = \frac{\pi_1(x)\exp(\alpha r(x))}{\tilde{Z}_\alpha}, \qquad \tilde{Z}_\alpha = \int \pi_1(x)\exp(\alpha r(x))\mathrm{d}x.$$

Here $\alpha \in \mathbb{R}$ is a scalar inverse temperature; as α changes from 0 to 1, the distribution $\tilde{\pi}_1^\alpha$ changes from $\tilde{\pi}_1^0 = \pi_1$ to $\tilde{\pi}_1^1$. This exponential tilting differs from the r-reweighted distribution $\hat{\pi}_1 \propto \pi_1 r$.

Let $\tilde{v}_t^\alpha$ be the RF velocity field associated with $\tilde{\pi}_1^\alpha$. As shown in Theorem 1.3 below, we can express the derivative of $\tilde{v}_t^\alpha$ w.r.t. α as a conditional covariance:

$$\begin{aligned}\partial_\alpha \tilde{v}_t^\alpha(x)\mid_{\alpha=0} &= \mathrm{cov}(X_1 - X_0, r(X_1) \mid X_t = x)\\ &= \hat{m}_t^r(x)(\hat{v}_t^r(x) - v_t^0(x)).\end{aligned}$$

Hence, v_t^* and $v_t^{*\pm}$ can be viewed as Taylor approximating $\tilde{v}_t^\alpha$ from $\alpha = 0$:

$$v_t^*(x) = v_t^0(x) + \frac{2}{\beta}\partial_\alpha \tilde{v}_t^\alpha(x)\mid_{\alpha=0} \approx \tilde{v}_t^{2/\beta}(x),$$

$$v_t^{*\pm}(x) = v_t^0(x) \pm 2\partial_\alpha \tilde{v}_t^\alpha(x)\mid_{\alpha=0} \approx \tilde{v}_t^{\pm 2}(x).$$

Assume $r(x)$ is sufficiently close to a constant, when the Taylor approximation is accurate, then the X_1 drawn from $\dot{Z}_t = v_t^*(Z_t)$ would approximately follow $\tilde{\pi}_1^{2/\beta}$, and X_1 drawn from $\dot{Z}_t = v_t^{*\pm}(Z_t)$ would approximately follow $\tilde{\pi}_1^{\pm 2}$.

In addition, note that

$$\exp(\alpha r(x)) = 1 + \alpha r(x) + \frac{\alpha^2}{2} r(x)^2 + \cdots.$$

The DiffusionNFT results should also be close to the distribution reweighted by the polynomial rewards $1 + \alpha r(x) + \frac{1}{2}\alpha^2 r(x)^2 + \cdots$. See Figure 2 for a toy demonstration where we can see the samples from DiffusionNFT approximates closely the predicted distributions. In this toy, it seems that the DiffusionNFT results matches most closely to the result of the second order approximation $\Phi(r(x)) = 1 + \alpha r(x) + \frac{\alpha^2}{2} r(x)^2$.

Given this approximate interpolation, one thing that is unclear is why should not we directly train the velocity field for the corredsponding reward weighted distributions using weighted losses.

In particular, let $\Phi(r(x))$ is a function of the reward. The velocity field of $\pi^\Phi(x) = \pi_1(x)\cdot\Phi(r(x))/Z$ can be written as

$$v_t^\Phi(x) = \frac{\mathbb{E}[\Phi(r(X_1))(X_1 - X_0) \mid X_t = x]}{\mathbb{E}[\Phi(r(X_1)) \mid X_t = x]},$$

which can be solved by

$$\min_v \mathbb{E}[\Phi(r(X_1))\|v_t(X_t) - (X_1 - X_0)\|^2].$$

One question is why not use them directly. One may argue that DiffusionNFT has (slightly?) lower variance? But this may not be the case. It could be that

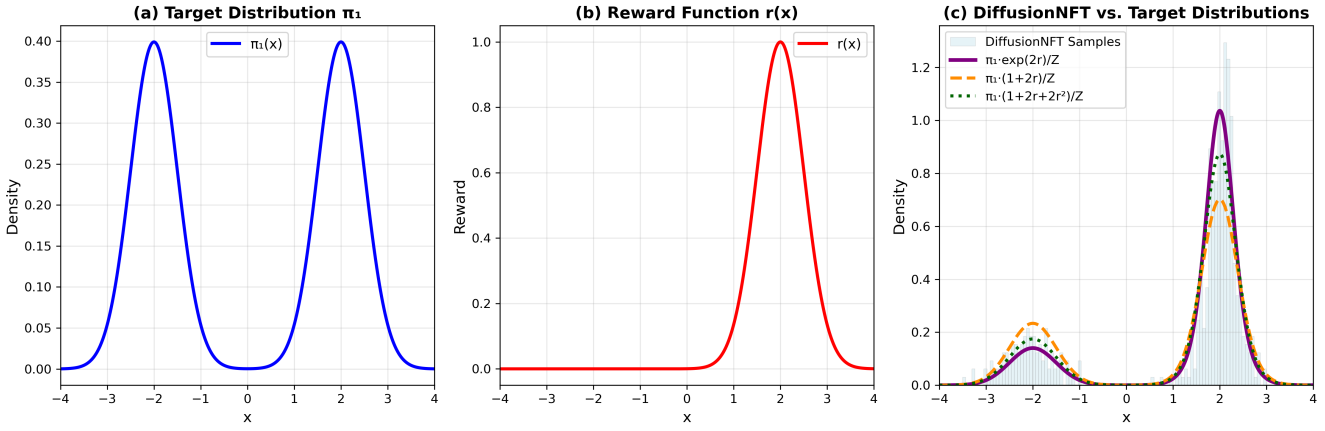

Figure 1: Illustration of Taylor approximation for DiffusionNFT. (a) Original target distribution $\pi_1(x)$ (mixture of two Gaussians). (b) Reward function $r(x)$ (Gaussian centered at $x = 2$). (c) Histogram of DiffusionNFT samples (blue) overlaid with theoretical distributions: $\pi_1 \cdot \exp(2r)/Z$ (purple solid, full exponential), $\pi_1 \cdot (1 + 2r)/Z$ (orange dashed, first-order), and $\pi_1 \cdot (1 + 2r + 2r^2)/Z$ (green dotted, second-order). The DiffusionNFT samples closely match the full exponential distribution, demonstrating the accuracy of the Taylor approximation approach.

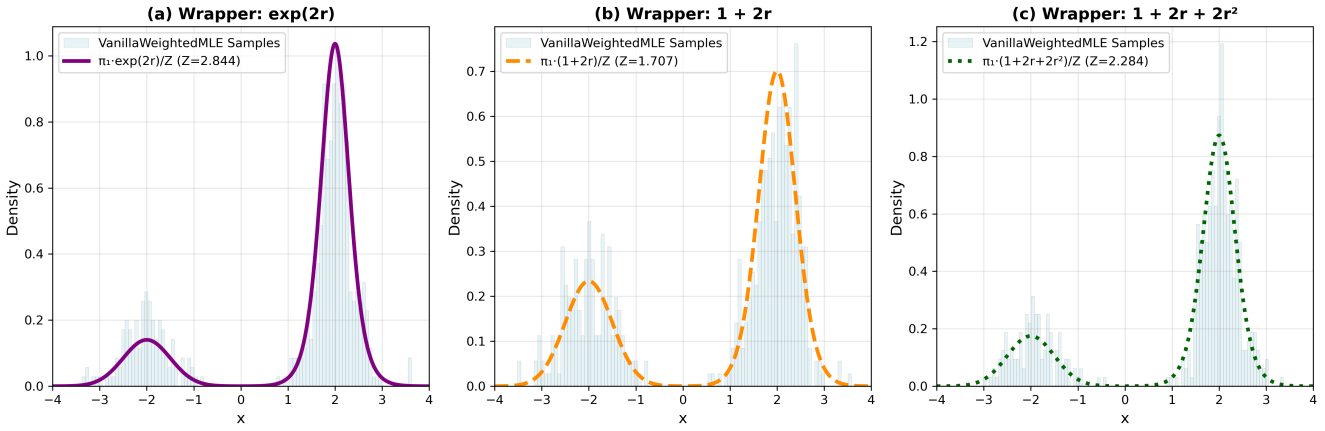

Figure 2: Results from RF of different reward weighted distributions obtained by minimizing the weighted loss (**??**).

1 Proofs

1.1 Formula of DiffusionNFT Solution

Theorem 1.1. *Following the set up above, and assume $v_t^0(x) = \mathbb{E}[X_1 - X_0 | X_t = x]$. We have the following equivalent expressions of $v_t^*(x)$ from DiffusionNFT:*

$$\begin{aligned} v_t^*(x) &= v_t^0 + \frac{2}{\beta}(\hat{m}_t^r(x)(\hat{v}_t^r(x) - v_t^0(x))) \\ &= v_t^0(x) + \frac{2}{\beta}\mathrm{cov}(r(X_1), X_1 - X_0 \mid X_t = x) \\ &= v_t^0(x) + \frac{2}{\beta}\,\partial_\alpha \tilde{v}_t^\alpha(x)\big|_{\alpha=0}, \end{aligned}$$

Similar expression holds for $v_t^{*\pm}(x)$ if $\frac{2}{\beta}$ is replaced by ± 2.

Proof. Using Lemma 1.2 below, we can rewrite the loss $L(\mu)$ as:

$$L(\mu) = \mathbb{E}[\left\|\beta\mu_t(X_t) - \beta v_t^0(X_t) - (2r(X_1) - 1)(X_1 - X_0 - v_t^0(X_t))\right\|^2] + const.$$

Hence, the optimal solution is

$$v_t^*(x) = v_t^0(x) + \frac{1}{\beta}\mathbb{E}[(2r(X_1) - 1)(X_1 - X_0 - v_t^0(X_t)) \mid X_t = x].$$

Plugging $v_t^0(x) = \mathbb{E}[X_1 - X_0 | X_t = x]$ into the above yields

$$\begin{aligned} v_t^*(x) &= v_t^0(x) + \frac{1}{\beta}\mathbb{E}[(2r(X_1) - 1)(X_1 - X_0 - \mathbb{E}[X_1 - X_0 | X_t = x]) \mid X_t = x] \\ &= v_t^0(x) + \frac{2}{\beta}\left(\mathbb{E}[r(X_1)(X_1 - X_0) \mid X_t = x] - \mathbb{E}[r(X_1) \mid X_t = x]\mathbb{E}[X_1 - X_0 | X_t = x]\right) \\ &= v_t^0 + \frac{2}{\beta}(\hat{m}_t^r(x)(\hat{v}_t^r(x) - v_t^0(x))) \\ &= v_t^0(x) + \frac{2}{\beta}\mathrm{cov}(2r(X_1) - 1, X_1 - X_0 \mid X_t = x) \\ &= v_t^0(x) + \frac{2}{\beta}\,\partial_\alpha \tilde{v}_t^\alpha(x)\big|_{\alpha=0}, \end{aligned}$$

where we used Theorem 1.3.

Plugging $v_t^{*\pm}(x) = v_t^0(x) \pm 2\partial_\alpha \tilde{v}_t^\alpha(x)\ |_{\alpha=0}$ into the above yields the formula for $v_t^{*\pm}(x)$.

□

Lemma 1.2. *Let $\mu^\pm = \mu \pm \beta(\mu - v)$, and*

$$\ell(\mu) = r\left\|\mu^+ - y\right\|^2 + (1 - r)\left\|\mu^- - y\right\|^2,$$

where $\mu, v, y \in \mathbb{R}^d$, $r \in \mathbb{R}$. Then, we can rewrite $\ell(\mu)$ into

$$\ell(\mu) = \left\|\beta\mu - \beta v - (2r - 1)(y - v)\right\|^2 + const,$$

where const is a term that does not depend on μ.

Hence, the minimum of $\ell(\mu)$ is achieved at

$$\mu^* = v + \frac{1}{\beta}(2r-1)(y-v).$$

Proof. Expand the loss:

$$\begin{aligned}
\ell(\mu) &= r\left\|\mu^+ - y\right\|^2 + (1-r)\left\|\mu^- - y\right\|^2 \\
&= r\left\|(1-\beta)v + \beta\mu - y\right\|^2 + (1-r)\left\|(1+\beta)v - \beta\mu - y\right\|^2 \\
&= \beta^2\mu^2 + 2\beta\mu(r(1-\beta)v - ry + (1-r)y - (1-r)(1+\beta)v) + const \\
&= \beta^2\mu^2 + 2\beta\mu(-\beta v + (1-2r)(y-v)) + const \\
&= \left\|\beta\mu - \beta v - (2r-1)(y-v)\right\|^2 + const.
\end{aligned}$$

Hence, the minimum is achieved at

$$\mu^* = \frac{1}{\beta}(\beta v - (1-2r)(y-v)) = v + \frac{1}{\beta}(2r-1)(y-v).$$

□

1.2 Derivative of RF Velocity Field

We give the formula for the derivative $\partial_\alpha v_t^\alpha(x)$ of the RF velocity field $v_t^\alpha(x)$ w.r.t. a general parameter α of the target distribution.

Let $X_0 \sim \pi_0$ be a base distribution and $X_1 \sim \pi_1^\alpha$ be a target distribution depending on a scalar parameter α. Define the linear interpolation $X_t = (1-t)X_0 + tX_1$, and the rectified flow velocity field

$$v_t^\alpha(x) = \mathbb{E}_{(X_0,X_1)\sim\gamma^\alpha}[X_1 - X_0 \mid X_t = x], \qquad t \in [0,1),$$

where (X_0, X_1) is a coupling of π_0 and π_1^α with a joint density $\gamma^\alpha(x_0, x_1)$ that is differentiable in α. Typically, (X_0, X_1) is the independent coupling $\gamma(x_0, x_1) = \pi_0(x) \times \pi_1^\alpha(x)$.

Theorem 1.3 (Derivative of the RF velocity via conditional score)**.** *Assume $\log\gamma^\alpha$ is differentiable in α and differentiation can be exchanged with integration and conditioning. Then*

$$\partial_\alpha v_t^\alpha(x) = \mathrm{Cov}_{\gamma^\alpha}\Big(X_1 - X_0,\ \partial_\alpha \log\gamma^\alpha(X_0, X_1)\ \Big|\ X_t = x\Big).$$

In particular, in the case of independent coupling $\gamma^\alpha(x_0, x_1) = \pi_0(x_0)\pi_1^\alpha(x_1)$, we have

$$\partial_\alpha v_t^\alpha(x) = \mathrm{Cov}_{\gamma^\alpha}\Big(X_1 - X_0,\ \partial_\alpha \log\pi_1^\alpha(X_1)\ \Big|\ X_t = x\Big).$$

Further, for the exponential-tilted family $\tilde\pi_1^\alpha(x) = \frac{\pi_1(x)\exp(\alpha r(x))}{\tilde Z_\alpha}$ with $\tilde Z_\alpha = \int \pi_1(x)\exp(\alpha r(x))\,\mathrm{d}x$, and the corresponding RF velocity field $\tilde v_t^\alpha$ under $\tilde\gamma^\alpha(x_0, x_1) = \pi_0(x_0)\tilde\pi_1^\alpha(x_1)$, we have

$$\partial_\alpha \tilde v_t^\alpha(x) = \mathrm{Cov}_{\tilde\gamma^\alpha}\Big(X_1 - X_0,\ r(X_1)\ \Big|\ X_t = x\Big).$$

Proof. Fix $t \in [0,1)$ and x. Note

$$v_t^\alpha(x) = \frac{m_t^\alpha(x) - x}{1-t}, \qquad m_t^\alpha(x) = \mathbb{E}[X_1 \mid X_t = x].$$

So it suffices to differentiate $m_t^\alpha(x)$.

Note that

$$m_t^\alpha(x) = \int x_1 \, p_t^\alpha(x_1 \mid x)\, dx_1,$$

where $p_t^\alpha(x_1 \mid x)$ denotes the conditional density of X_1 given $X_t = x$, defined in terms of the joint density $\gamma^\alpha(x_0, x_1)$. By applying the change of variables $X_0 = \frac{x - t x_1}{1-t}$, we obtain

$$p_t^\alpha(x_1 \mid x) = \frac{\gamma^\alpha\Big(\frac{x-tx_1}{1-t},\, x_1\Big) \cdot \frac{1}{1-t}}{\rho_t^\alpha(x)}, \qquad \rho_t^\alpha(x) = \int \gamma^\alpha\Big(\frac{x - tx_1}{1-t},\, x_1\Big) \cdot \frac{1}{1-t}\, dx_1.$$

Differentiate under the integral:

$$\partial_\alpha m_t^\alpha(x) = \int x_1 \, \partial_\alpha p_t^\alpha(x_1 \mid x)\, dx_1 = \int x_1 \, p_t^\alpha(x_1 \mid x)\, \partial_\alpha \log p_t^\alpha(x_1 \mid x)\, dx_1 = \mathbb{E}\big[X_1 \, \partial_\alpha \log p_t^\alpha(X_1 \mid x) \,\big|\, X_t = x\big].$$

Now expand the conditional score. Up to an α-independent constant,

$$\log p_t^\alpha(x_1 \mid x) = \log \gamma^\alpha\Big(\frac{x - tx_1}{1-t},\, x_1\Big) - \log \rho_t^\alpha(x) + \text{const}.$$

Hence

$$\partial_\alpha \log p_t^\alpha(x_1 \mid x) = \partial_\alpha \log \gamma^\alpha\Big(\frac{x - tx_1}{1-t},\, x_1\Big) - \partial_\alpha \log \rho_t^\alpha(x).$$

Using the Fisher identity for the marginal $\rho_t^\alpha(x)$,

$$\partial_\alpha \log \rho_t^\alpha(x) = \mathbb{E}\Big[\partial_\alpha \log \gamma^\alpha(X_0, X_1) \,\Big|\, X_t = x\Big].$$

So

$$\partial_\alpha \log p_t^\alpha(X_1 \mid x) = \partial_\alpha \log \gamma^\alpha(X_0, X_1) - \mathbb{E}\Big[\partial_\alpha \log \gamma^\alpha(X_0, X_1) \,\Big|\, X_t = x\Big].$$

Plugging back,

$$\begin{aligned}
\partial_\alpha m_t^\alpha(x) &= \mathbb{E}\Big[X_1 \Big(\partial_\alpha \log \gamma^\alpha(X_0, X_1) - \mathbb{E}[\partial_\alpha \log \gamma^\alpha(X_0, X_1) \mid X_t = x]\Big) \,\Big|\, X_t = x\Big] \\
&= \mathrm{Cov}\Big(X_1,\ \partial_\alpha \log \gamma^\alpha(X_0, X_1) \,\Big|\, X_t = x\Big).
\end{aligned}$$

Hence,

$$\begin{aligned}
\partial_\alpha v_t^\alpha(x) &= \frac{1}{1-t}\, \partial_\alpha m_t^\alpha(x) \\
&= \frac{1}{1-t}\, \mathrm{Cov}\Big(X_1,\ \partial_\alpha \log \gamma^\alpha(X_0, X_1) \,\Big|\, X_t = x\Big) \\
&= \mathrm{Cov}\Big(\frac{X_1 - x}{1-t},\ \partial_\alpha \log \gamma^\alpha(X_0, X_1) \,\Big|\, X_t = x\Big) \\
&= \mathrm{Cov}\Big(X_1 - X_0,\ \partial_\alpha \log \gamma^\alpha(X_0, X_1) \,\Big|\, X_t = x\Big),
\end{aligned}$$

where we note that $X_1 - X_0 = \frac{X_1 - x}{1-t}$ conditioned on $X_t = x$.

Finally, for the exponential-tilted family $\tilde{\pi}_1^\alpha$, we have

$$\partial_\alpha \log \tilde{\pi}_1^\alpha(X_1) = r(X_1) - \partial_\alpha \log \tilde{Z}_\alpha.$$

Hence,

$$\begin{aligned}\partial_\alpha \tilde{v}_t^\alpha(x) &= \frac{1}{1-t}\,\mathrm{Cov}\Big(X_1,\ r(X_1) - \partial_\alpha \log \tilde{Z}_\alpha \,\Big|\, X_t = x\Big) \\ &= \mathrm{Cov}\Big(X_1 - X_0,\ r(X_1) \,\Big|\, X_t = x\Big),\end{aligned}$$

where $\partial_\alpha \log \tilde{Z}_\alpha$ is dropped as it is a deterministic constant and does not influence the conditional covariance. □

References

[1] Zheng, K., Chen, H., Ye, H., Wang, H., Zhang, Q., Jiang, K., Su, H., Ermon, S., Zhu, J., and Liu, M.-Y. (2025). Diffusionnft: Online diffusion reinforcement with forward process. *arXiv preprint*.