1. Proofs

1.1. Proof of Theorem 3.1

Theorem 1.1. *Let $p_{data}(x) = \mathcal{N}(0, c^2 I)$ be the data distribution. Then the optimal consistency model $f^*(x_t, t)$ is given by*

$$f^*(x_t, t) = \frac{c \times x_t}{\sqrt{t^2 + (1-t)^2 c^2}}.$$

For any $\epsilon > 0$ consider the following non-optimal consistency model

$$f_\epsilon(x_t, t) = \frac{(c + t \times \epsilon) x_t}{\sqrt{t^2 + (1-t)^2 c^2}}.$$

Then, there exists an integer N for which increasing the number of sampling steps during multistep CM sampling beyond N increases the Wasserstein-2 distance of the generated samples to the ground truth distribution.

Proof. In this Gaussian setting, all intermediate noisy distributions $p(x_t)$ where $x_t = t \times x_1 + (1-t) \times x_0, x_0 \sim p_{data}, x_1 \sim \mathcal{N}(0, I)$ remain Gaussian. Consequently, both the optimal velocity from flow matching, $v^*(x_t, t)$, and the optimal denoiser, $D^*(x_t, t)$, have closed-form solutions.

We first derive the denoiser. Following the framework of (Karras et al., 2022), let $p(x; \sigma)$ denote the distribution obtained by adding independent Gaussian noise with standard deviation σ to the data. Then,

$$p(x;\sigma) = \mathcal{N}(0, (c^2+\sigma^2)I) \Rightarrow \nabla_x \log p(x;\sigma) = \frac{-x}{c^2+\sigma^2} \Rightarrow D^*(x,\sigma) = \frac{c^2}{c^2+\sigma^2} x,$$

where the final step follows from the identity

$$\nabla_x \log p(x;\sigma) = (D(x;\sigma) - x)/\sigma^2$$

from (Karras et al., 2022). Using this, the optimal velocity from flow matching, $v^*(x_t, t)$, can be computed as

$$v^*(x_t, t) = \mathbb{E}[x_1 - x_0 \mid x_t] = \mathbb{E}\left[\frac{x_t - x_0}{t} \mid x_t\right] = \frac{x_t}{t} - \frac{1}{t}\mathbb{E}[x_0 \mid x_t].$$

Substituting $D^*\left(\frac{x_t}{1-t}, \frac{t}{1-t}\right)$ for $\mathbb{E}[x_0 \mid x_t]$, we obtain

$$v^*(x_t, t) = \frac{x_t}{t} - \frac{1}{t} D^*\left(\frac{x_t}{1-t}, \frac{t}{1-t}\right).$$

Using the closed-form expression for $D^*(x, \sigma)$,

$$v^*(x_t, t) = \frac{x_t}{t} - \frac{1}{t} \cdot \frac{c^2}{c^2 + \left(\frac{t}{1-t}\right)^2} \cdot \frac{x_t}{1-t}.$$

Simplifying further,

$$v^*(x_t, t) = \frac{x_t}{t}\left(1 - \frac{c^2(1-t)}{c^2(1-t)^2 + t^2}\right) = \left(\frac{t - c^2(1-t)}{t^2 + (1-t)^2 c^2}\right) x_t.$$

Integrating this velocity along the PF-ODE,

$$dx_t = v^*(x_t, t) dt,$$

from t to 0 yields the optimal consistency model:

$$f^*(x_t, t) = \frac{c \cdot x_t}{\sqrt{t^2 + (1-t)^2 c^2}}.$$

Consider the non-optimal consistency model

$$f_\epsilon(x_t, t) = \frac{(c + t \times \epsilon)x_t}{\sqrt{t^2 + (1-t)^2c^2}}. \tag{1}$$

Note that $f_\epsilon(x, 0) = x$ satisfies the boundary condition and is a valid consistency model.

Let us analyze a single multistep transition from timestep t to timestep s. This process consists of two steps: 1. The noise is removed by predicting the clean data using the consistency model, yielding $x_0' = f_\epsilon(x_t, t)$. 2. The estimated clean data x_0' is then noised back to timestep s using

$$x_s = s \times z + (1-s) \times x_0', \quad z \sim \mathcal{N}(0, I).$$

Assuming that $x_t \sim \mathcal{N}(0, \sigma_t^2 I)$, x_s will also follow an isotropic zero-mean Gaussian distribution and is obtained as follows:

$$x_s = s \times z + (1-s) \times \frac{(c + t\epsilon)x_t}{\sqrt{t^2 + (1-t)^2c^2}}.$$

Therefore, the variance of x_s is given by

$$\mathrm{Var}(x_s) = s^2 + \frac{(1-s)^2(c+t\epsilon)^2}{t^2 + (1-t)^2c^2}\,\mathrm{Var}(x_t). \tag{2}$$

Using this recurrence relation, we can compute the variance of the distribution obtained by running multistep CM sampling on a uniform n-step schedule:

$$[0, \frac{1}{n}, \ldots, \frac{n-1}{n}, 1].$$

Since both the ground truth distribution and the distribution of x_0 derived by n-step sampling are isotropic zero-mean Gaussians, the Wasserstein-2 distance between them has the following closed form solution:

$$W_2(p(x_0); p_{data}(x)) = (\sqrt{\mathrm{Var}(x_0)} - c)^2$$

Let $\mathrm{Var}(s) := \mathrm{Var}(x_s)$ for convenience. We will show that as $n \to \infty$, the variance $\mathrm{Var}(0)$ when computed via the recurrence defined in Eq. (2) on the uniform n step schedule diverges, i.e. $\mathrm{Var}(0) \to \infty$. This means performing multi-step sampling with the consistency model will result in accumulated errors beyond a certain point.

Define

$$h(s) := \frac{\mathrm{Var}(s)}{s^2 + (1-s)^2c^2}.$$

Plugging this into Eq. (2) gives:

$$h(s) = \frac{s^2 + (1-s)^2(c+t\epsilon)^2h(t)}{s^2 + (1-s)^2c^2}. \tag{3}$$

We know $h(1) = 1$ and $h(0) = \mathrm{Var}(0)/c^2$. So, it's enough to show that $h(0) \to \infty$ as $n \to \infty$.

It's easy to see by induction that $h(t) \geq 1$. Define $g(t) := h(t) - 1$ to measure how much it grows. Then:

$$\begin{aligned} g(s) &= \frac{(1-s)^2(c+t\epsilon)^2(g(t)+1) - (1-s)^2c^2}{s^2 + (1-s)^2c^2} \\ &= \frac{(1-s)^2}{s^2 + (1-s)^2c^2}\left(2ct\epsilon + t^2\epsilon^2 + (c+t\epsilon)^2 g(t)\right) \\ &\geq \frac{(1-s)^2}{s^2 + (1-s)^2c^2}(2ct\epsilon + c^2 g(t)). \end{aligned} \tag{4}$$

All terms are positive, so let's lower bound $g(t)$ by defining a simpler sequence:

$$g'(s) := \frac{(1-s)^2}{s^2 + (1-s)^2c^2}(2ct\epsilon + c^2 g'(t)). \tag{5}$$

Clearly $g(s) \geq g'(s)$, so it's enough to show $g'(0) \to \infty$ as $n \to \infty$.

Now define:

$$r(s) := \frac{g'(s)}{2c\epsilon}.$$

Then the recurrence becomes:

$$r(s) = \frac{(1-s)^2}{s^2 + (1-s)^2 c^2}(t + c^2 r(t)), \tag{6}$$

with $r(1) = 0$ and $r(0) = g'(0)/(2c\epsilon)$. So we just need to show $r(0) \to \infty$ as $n \to \infty$.

To analyze this, fix s and t, and consider the function:

$$f_{s,t}(x) := \frac{(1-s)^2}{s^2 + (1-s)^2 c^2}(t + c^2 x). \tag{7}$$

This is an affine map with a unique fixed point:

$$o(s,t) = \frac{t(1-s)^2}{s^2}. \tag{8}$$

Subtracting $o(s,t)$ from both sides, we get:

$$f(x) - o(s,t) = \lambda(s)(x - o(s,t)). \tag{9}$$

where:

$$\lambda(s) := \frac{1}{(\frac{s}{(1-s)c})^2 + 1} < 1.$$

So $f_{s,t}(x)$ pulls every point toward its fixed point $o(s,t)$, with a pull factor of $\lambda(s)$. As $s \to 0$, since $o(s,t) \geq \frac{(1-s)^2}{s}$, the fixed point $o(s,t) \to \infty$, and the pull factor $\lambda(s)$ approaches 1, meaning the pull gets weaker.

This means the recurrence in Eq. (6) is applying a sequence of weaker and weaker pulls toward bigger and bigger targets. To prove $r(0) \to \infty$, we now proceed by contradiction. Assume there exists some $\delta > 0$ and an infinite sequence $n_1 < n_2 < \ldots$ such that $r(0) < \delta$ when using the schedule $[0, \frac{1}{n_i}, \ldots, 1]$.

Since $o(s,t) \geq \frac{(1-s)^2}{s}$ and the function $\frac{(1-x)^2}{x} \to \infty$ as $x \to 0$, we can pick $t^* \in [0,1]$ such that for all $s \in [0, t^*]$:

$$\frac{(1-t^*)^2}{t^*} = 2\delta \Rightarrow o(s,t) \geq \frac{(1-s)^2}{s} \geq 2\delta. \tag{10}$$

So every fixed point in $[0, t^*]$ is at least 2δ.

Now we split the problem into two cases:

Case 1: There exists some $s \in [0, t^*]$ where $r(s) \geq \delta$. Since all future pulls are toward values $> 2\delta$, $r(\cdot)$ will stay above δ, so $r(0) \geq \delta$—a contradiction.

Case 2: For all $s \in [0, t^*]$, we have $r(s) < \delta$. Pick any $s^* \in (0, t^*)$. Let $\ell := \lambda(s^*) < 1$, which is the maximum pull factor (corresponding to the weakest pull) on $[s^*, t^*]$. Then for $s^* \leq s < t \leq t^*$:

$$\begin{aligned}
2\delta - r(s) &= (o(s,t) - r(s)) &&+ (2\delta - o(s,t)) \\
&= (o(s,t) - f_{s,t}(r(t))) &&+ (2\delta - o(s,t)) \\
&= \lambda(s)(o(s,t) - r(t)) &&+ (2\delta - o(s,t)) \\
&\leq \lambda(s^*)(o(s,t) - r(t)) &&+ (2\delta - o(s,t)) \\
&= \ell(o(s,t) - r(t)) &&+ (2\delta - o(s,t)) \\
&= \ell(2\delta - r(t)) &&+ (1-\ell)(2\delta - o(s,t)) \\
&\leq \ell(2\delta - r(t)).
\end{aligned}$$

Applying this inequality M times (where M is the number of steps between s^* and t^* on the schedule), we get:

$$\delta \leq 2\delta - r(s^*) \leq 2\delta\ell^M. \tag{11}$$

If M is large enough so that $2\delta\ell^M < \delta$, we get a contradiction. This happens when $n > M/(t^* - s^*)$.

Since both cases lead to a contradiction, we conclude that $r(0) \to \infty$ as $n \to \infty$, completing the proof.

$\square$

References

Karras, T., Aittala, M., Aila, T., and Laine, S. Elucidating the design space of diffusion-based generative models. *Advances in neural information processing systems*, 35:26565–26577, 2022.