FlowSteer: Guiding Few-Step Image Synthesis with Authentic Trajectories

Supplementary Material

1. Proof of Inter-Stage Distribution Mismatch

Here, we provide the detailed proof that a distribution mismatch occurs at stage boundaries unless the teacher model is a perfect Rectified Flow model.

Let the joint distribution of data and noise be $p(z_0, \epsilon) = p_{\text{data}}(z_0)\mathcal{N}(\epsilon|0, I)$. The condition for a seamless transition between stages requires that the distribution of the teacher's evolved state at time t_{k-1}, denoted $\hat{z}_{t_{k-1}}$, matches the prescribed distribution of the ideal starting state $z_{t_{k-1}}$. A necessary condition for this is that their expectations match:

$$\mathbb{E}_{z_0,\epsilon}[\hat{z}_{t_{k-1}}] = \mathbb{E}_{z_0,\epsilon}[z_{t_{k-1}}]. \tag{1}$$

The teacher's output $\hat{z}_{t_{k-1}}$ is obtained by integrating its velocity field v_T from a starting point z_{t_k}. Taking the expectation over the integral form gives:

$$\mathbb{E}[z_{t_k}] + \mathbb{E}\left[\int_{t_k}^{t_{k-1}} v_T(z(s), s)ds\right] = \mathbb{E}[z_{t_{k-1}}]. \tag{2}$$

Using the linearity of expectation and Fubini's theorem to swap the expectation and integral, we get:

$$\int_{t_k}^{t_{k-1}} \mathbb{E}[v_T(z(s), s)]ds = \mathbb{E}[z_{t_{k-1}}] - \mathbb{E}[z_{t_k}]. \tag{3}$$

The expectation of the state $z_t = (1 - \sigma(t))z_0 + \sigma(t)\epsilon$ is $\mathbb{E}[z_t] = (1 - \sigma(t))\mathbb{E}[z_0] = (1 - \sigma(t))\mu_{\text{data}}$, where μ_{data} is the mean of the data distribution and $\mathbb{E}[\epsilon] = 0$. Substituting this into the right-hand side of Eq. 3 yields:

$$\begin{aligned}\mathbb{E}[z_{t_{k-1}}] - \mathbb{E}[z_{t_k}] &= (1 - \sigma(t_{k-1}))\mu_{\text{data}} - (1 - \sigma(t_k))\mu_{\text{data}} \\ &= (\sigma(t_k) - \sigma(t_{k-1}))\mu_{\text{data}}.\end{aligned} \tag{4}$$

By dividing both sides of Eq. 3 by $(t_{k-1} - t_k)$ and taking the limit as $t_{k-1} \to t_k$, the left side becomes the instantaneous expected velocity, and the right side becomes the derivative of $-\sigma(t)$ multiplied by μ_{data}:

$$\mathbb{E}_{z_t \sim p(z,t)}[v_T(z_t, t)] = -\sigma'(t)\mu_{\text{data}}. \tag{5}$$

This equation represents a powerful constraint. It dictates that for a seamless transition, the mean of the teacher's vector field, when averaged over the entire distribution of states $p(z, t)$ at time t, must follow a predetermined path solely defined by the data mean μ_{data} and the noise schedule $\sigma(t)$.

However, a general teacher model v_T (e.g., a pre-trained diffusion model) possesses complex, non-linear dynamics and is not guaranteed to satisfy this condition. This condition is only naturally satisfied if the teacher itself is a perfect Rectified Flow model, where the velocity field is specifically constructed to ensure linear trajectories on average. Forcing a general teacher to satisfy this condition contradicts the premise that it has a complex, non-linear trajectory that requires distillation. Therefore, a distribution mismatch is inevitable.

Algorithm 1 sigma sampling logic. Our method (blue marker) samples from a pre-completed schedule, while the original (red marker) appends $\sigma = 0$ post-sampling.

```
// — Original Method —
indices = linspace(0, 999, N)
s_sampled = sigmas[indices]
return append(s_sampled, 0)

// — Our Improved Method —
s_full = append(sigmas, 0)
indices = linspace(0, 1000, N+1)
return s_full[indices]
```

Table 1. Generated sigma values for N=4 inference steps.

Setting	Generated Sigma Values
Original Method	
Shift=1	[1.000, 0.667, 0.334, 0.001, 0.000]
Shift=3	[1.000, 0.858, 0.602, 0.009, 0.000]
Our Improved Method	
Shift=1	[1.000, 0.750, 0.500, 0.250, 0.000]
Shift=3	[1.000, 0.900, 0.751, 0.502, 0.000]

2. Improved Scheduler Details

As discussed, `FlowMatchEulerDiscreteScheduler` exhibits a structural flaw in the few-step inference. The original implementation samples N sigmas from the full schedule and then appends the terminal state ($\sigma = 0$) separately. This results in a final step size that is not proportional to the preceding steps, which can degrade image quality.

To rectify this, we propose an improved sampling strategy. Our method first augments the full sigma schedule with the terminal $\sigma = 0$ state. Then, it samples $N+1$ points from this complete schedule. This ensures that all step intervals are proportionally scaled.

The unified logic, highlighting the divergence between the original and our improved method, is presented in Algorithm 1. A concrete example of the resulting sigma values for a 4-step inference is shown in Table 1.