Score, Noise, Denoiser, and Velocity Predictions in Diffusion Models

Beier Zhu

Let $\mathbf{x}_0 \sim p_{\mathsf{data}}$ denote clean data and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, I)$ denote Gaussian noise. The noisy state at timestep t is defined as

$$\mathbf{x}_t = \alpha_t \mathbf{x}_0 + \sigma_t \boldsymbol{\epsilon}, \tag{1}$$

where $\alpha_t = \alpha(t)$ and $\sigma_t = \sigma(t)$ are time-dependent coefficients.

Table 1 summarizes four equivalent denoising parameterizations: score, denoiser, noise, and velocity prediction; Table 2 gives their deterministic linear conversions.

Parameterization	**Definition**	**Training Objective**
Score prediction [1]	$\mathbf{s}_\theta(\mathbf{x}_t, t) = \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$	$\mathbb{E}[\Vert\mathbf{s}_\theta(\mathbf{x}_t, t) + \frac{\boldsymbol{\epsilon}}{\sigma_t}\Vert^2]$
Denoiser prediction [2]	$D_\theta(\mathbf{x}_t, t) = \mathbb{E}[\mathbf{x}_0 \mid \mathbf{x}_t]$	$\mathbb{E}[\Vert D_\theta(\mathbf{x}_t, t) - \mathbf{x}_0\Vert^2]$
Noise prediction [3]	$\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) = \mathbb{E}[\boldsymbol{\epsilon} \mid \mathbf{x}_t]$	$\mathbb{E}[\Vert\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) - \boldsymbol{\epsilon}\Vert^2]$
Velocity prediction [4]	$\mathbf{v}_t(\mathbf{x}_t, t) = \mathbb{E}[\dot{\alpha}_t \mathbf{x}_0 + \dot{\sigma}_t \boldsymbol{\epsilon} \mid \mathbf{x}_t]$	$\mathbb{E}[\Vert\mathbf{v}_\theta(\mathbf{x}_t, t) - (\dot{\alpha}_t \mathbf{x}_0 + \dot{\sigma}_t \boldsymbol{\epsilon})\Vert^2]$

Table 1: Parameterizations for diffusion models. Weighting factors are omitted for simplicity.

	$\mathbf{s}_\theta$	D_θ	$\boldsymbol{\epsilon}_\theta$	$\mathbf{v}_\theta$
$\mathbf{s}_\theta$		$D_\theta = \frac{1}{\alpha_t}\mathbf{x}_t + \frac{\sigma_t^2}{\alpha_t}\mathbf{s}_\theta$	$\boldsymbol{\epsilon}_\theta = -\sigma_t \mathbf{s}_\theta$	$\mathbf{v}_\theta = \frac{\dot{\alpha}_t}{\alpha_t}\mathbf{x}_t + \sigma_t^2(\frac{\dot{\alpha}_t}{\alpha_t} - \frac{\dot{\sigma}_t}{\sigma_t})\mathbf{s}_\theta$
D_θ	$\mathbf{s}_\theta = -\frac{1}{\sigma_t^2}\mathbf{x}_t + \frac{\alpha_t}{\sigma_t^2}D_\theta$		$\boldsymbol{\epsilon}_\theta = \frac{1}{\sigma_t}\mathbf{x}_t - \frac{\alpha_t}{\sigma_t}D_\theta$	$\mathbf{v}_\theta = \frac{\dot{\sigma}_t}{\sigma_t}\mathbf{x}_t + \dot{\alpha}_t(1 - \frac{\dot{\sigma}_t}{\sigma_t}\frac{\alpha_t}{\dot{\alpha}_t})D_\theta$
$\boldsymbol{\epsilon}_\theta$	$\mathbf{s}_\theta = -\frac{1}{\sigma_t}\boldsymbol{\epsilon}_\theta$	$D_\theta = \frac{1}{\alpha_t}\mathbf{x}_t - \frac{\sigma_t}{\alpha_t}\boldsymbol{\epsilon}_\theta$		$\mathbf{v}_\theta = \frac{\dot{\alpha}_t}{\alpha_t}\mathbf{x}_t + \dot{\sigma}_t(1 - \frac{\dot{\alpha}_t}{\alpha_t}\frac{\sigma_t}{\dot{\sigma}_t})\boldsymbol{\epsilon}_\theta$
$\mathbf{v}_\theta$	$\mathbf{s}_\theta = \frac{\dot{\alpha}_t}{\sigma_t\Delta_t}\mathbf{x}_t - \frac{\alpha_t}{\sigma_t\Delta_t}\mathbf{v}_\theta$	$D_\theta = \frac{\dot{\sigma}_t}{\Delta_t}\mathbf{x}_t - \frac{\sigma_t}{\Delta_t}\mathbf{v}_\theta$	$\boldsymbol{\epsilon}_\theta = -\frac{\dot{\alpha}_t}{\Delta_t}\mathbf{x}_t + \frac{\alpha_t}{\Delta_t}\mathbf{v}_\theta$	

Table 2: Conversion formulas among the four parameterizations. Here $\Delta_t = \alpha_t\dot{\sigma}_t - \sigma_t\dot{\alpha}_t$.

These relations show that different prediction targets encode the same denoising information. This is useful because the ODE sampler has a particularly simple form under the velocity parameterization, i.e., $d\mathbf{x} = \mathbf{v}_\theta(\mathbf{x}, t)dt$. In Table 2, for score, denoiser, or noise prediction, we can use the highlighted velocity column to convert their outputs into the corresponding ODE vector fields.

Special cases

We consider three special cases of the general parameterization:

- Flow models [4] with $\alpha_t = 1 - t$ and $\sigma_t = t$, yielding $\dot{\alpha}_t = -1$, $\dot{\sigma}_t = 1$ and $\Delta_t = 1$ (Table 3);
- VE models [1] with $\alpha_t = 1$ and $\sigma_t = \sigma(t)$, yielding $\dot{\alpha}_t = 0$ and $\Delta_t = \dot{\sigma}_t$ (Table 4);
- EDM models [2] with $\alpha_t = 1$ and $\sigma_t = t$, yielding $\dot{\alpha}_t = 0$, $\dot{\sigma}_t = 1$ and $\Delta_t = 1$ (Table 5);

Proof

We derive three pairs of relations among score, noise, denoiser, and velocity under $\mathbf{x}_t = \alpha_t\mathbf{x}_0 + \sigma_t\boldsymbol{\epsilon}$, which together yield Table 2.

(a) Conversion formulas

	$\mathbf{s}_\theta$	D_θ	$\boldsymbol{\epsilon}_\theta$	$\mathbf{v}_\theta$
$\mathbf{s}_\theta$		$D_\theta = \frac{\mathbf{x}_t + t^2\mathbf{s}_\theta}{1-t}$	$\boldsymbol{\epsilon}_\theta = -t\mathbf{s}_\theta$	$\mathbf{v}_\theta = -\frac{\mathbf{x}_t + t\mathbf{s}_\theta}{1-t}$
D_θ	$\mathbf{s}_\theta = \frac{(1-t)D_\theta - \mathbf{x}_t}{t^2}$		$\boldsymbol{\epsilon}_\theta = \frac{\mathbf{x}_t - (1-t)D_\theta}{t}$	$\mathbf{v}_\theta = \frac{\mathbf{x}_t - D_\theta}{t}$
$\boldsymbol{\epsilon}_\theta$	$\mathbf{s}_\theta = -\frac{1}{t}\boldsymbol{\epsilon}_\theta$	$D_\theta = \frac{\mathbf{x}_t - t\boldsymbol{\epsilon}_\theta}{1-t}$		$\mathbf{v}_\theta = \frac{\boldsymbol{\epsilon}_\theta - \mathbf{x}_t}{1-t}$
$\mathbf{v}_\theta$	$\mathbf{s}_\theta = -\frac{\mathbf{x}_t + (1-t)\mathbf{v}_\theta}{t}$	$D_\theta = \mathbf{x}_t - t\mathbf{v}_\theta$	$\boldsymbol{\epsilon}_\theta = \mathbf{x}_t + (1-t)\mathbf{v}_\theta$	

(b) ODE formulations

ODE
$d\mathbf{x}_t = -\frac{\mathbf{x}_t + t\mathbf{s}_\theta}{1-t}\, dt$
$d\mathbf{x}_t = \frac{\mathbf{x}_t - D_\theta}{t}\, dt$
$d\mathbf{x}_t = \frac{\boldsymbol{\epsilon}_\theta - \mathbf{x}_t}{1-t}\, dt$
$d\mathbf{x}_t = \mathbf{v}_\theta\, dt$

Table 3: Conversion formulas and ODE formulations for flow models with $\alpha_t = 1 - t$, $\sigma_t = t$, $\dot{\alpha}_t = -1$, $\dot{\sigma}_t = 1$, and $\Delta_t = 1$. The highlighted column in (a) gives the velocity form used to derive the ODE formulations in (b).

(a) Conversion formulas

	$\mathbf{s}_\theta$	D_θ	$\boldsymbol{\epsilon}_\theta$	$\mathbf{v}_\theta$
$\mathbf{s}_\theta$		$D_\theta = \mathbf{x}_t + \sigma_t^2\mathbf{s}_\theta$	$\boldsymbol{\epsilon}_\theta = -\sigma_t\mathbf{s}_\theta$	$\mathbf{v}_\theta = -\sigma_t\dot{\sigma}_t\mathbf{s}_\theta$
D_θ	$\mathbf{s}_\theta = \frac{D_\theta - \mathbf{x}_t}{\sigma_t^2}$		$\boldsymbol{\epsilon}_\theta = \frac{\mathbf{x}_t - D_\theta}{\sigma_t}$	$\mathbf{v}_\theta = \frac{\dot{\sigma}_t}{\sigma_t}(\mathbf{x}_t - D_\theta)$
$\boldsymbol{\epsilon}_\theta$	$\mathbf{s}_\theta = -\frac{\boldsymbol{\epsilon}_\theta}{\sigma_t}$	$D_\theta = \mathbf{x}_t - \sigma_t\boldsymbol{\epsilon}_\theta$		$\mathbf{v}_\theta = \dot{\sigma}_t\boldsymbol{\epsilon}_\theta$
$\mathbf{v}_\theta$	$\mathbf{s}_\theta = -\frac{\mathbf{v}_\theta}{\sigma_t\dot{\sigma}_t}$	$D_\theta = \mathbf{x}_t - \frac{\sigma_t}{\dot{\sigma}_t}\mathbf{v}_\theta$	$\boldsymbol{\epsilon}_\theta = \frac{\mathbf{v}_\theta}{\dot{\sigma}_t}$	

(b) ODE formulations

ODE
$d\mathbf{x}_t = -\sigma_t\dot{\sigma}_t\mathbf{s}_\theta\, dt$
$d\mathbf{x}_t = \frac{\dot{\sigma}_t}{\sigma_t}(\mathbf{x}_t - D_\theta)\, dt$
$d\mathbf{x}_t = \dot{\sigma}_t\boldsymbol{\epsilon}_\theta\, dt$
$d\mathbf{x}_t = \mathbf{v}_\theta\, dt$

Table 4: Conversion formulas and ODE formulations for VE models with $\alpha_t = 1$, $\dot{\alpha}_t = 0$, and $\Delta_t = \dot{\sigma}_t$. The highlighted column in (a) gives the velocity form used to derive the ODE formulations in (b).

(a) Conversion formulas

	$\mathbf{s}_\theta$	D_θ	$\boldsymbol{\epsilon}_\theta$	$\mathbf{v}_\theta$
$\mathbf{s}_\theta$		$D_\theta = \mathbf{x}_t + t^2\mathbf{s}_\theta$	$\boldsymbol{\epsilon}_\theta = -t\mathbf{s}_\theta$	$\mathbf{v}_\theta = -t\mathbf{s}_\theta$
D_θ	$\mathbf{s}_\theta = \frac{D_\theta - \mathbf{x}_t}{t^2}$		$\boldsymbol{\epsilon}_\theta = \frac{\mathbf{x}_t - D_\theta}{t}$	$\mathbf{v}_\theta = \frac{\mathbf{x}_t - D_\theta}{\sigma_t}$
$\boldsymbol{\epsilon}_\theta$	$\mathbf{s}_\theta = -\frac{\boldsymbol{\epsilon}_\theta}{t}$	$D_\theta = \mathbf{x}_t - t\boldsymbol{\epsilon}_\theta$		$\mathbf{v}_\theta = \boldsymbol{\epsilon}_\theta$
$\mathbf{v}_\theta$	$\mathbf{s}_\theta = -\frac{\mathbf{v}_\theta}{t}$	$D_\theta = \mathbf{x}_t - t\mathbf{v}_\theta$	$\boldsymbol{\epsilon}_\theta = \mathbf{v}_\theta$	

(b) ODE formulations

ODE
$d\mathbf{x}_t = -t\mathbf{s}_\theta\, dt$
$d\mathbf{x}_t = \frac{\mathbf{x}_t - D_\theta}{\sigma_t}\, dt$
$d\mathbf{x}_t = \boldsymbol{\epsilon}_\theta\, dt$
$d\mathbf{x}_t = \mathbf{v}_\theta\, dt$

Table 5: Conversion formulas and ODE formulations for EDM models with $\alpha_t = 1$, $\sigma_t = t$, $\dot{\alpha}_t = 0$, $\dot{\sigma}_t = 1$ and $\Delta_t = 1$ The highlighted column in (a) gives the velocity form used to derive the ODE formulations in (b).

Noise-score relation

The conditional distribution of $\mathbf{x}_t$ given $\mathbf{x}_0$ is

$$q(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t \mid \alpha_t\mathbf{x}_0, \sigma_t^2 I). \tag{2}$$

Using $\mathbf{x}_t - \alpha_t\mathbf{x}_0 = \sigma_t\boldsymbol{\epsilon}$ (Eq. 1), its conditional score is

$$\nabla_{\mathbf{x}_t} \log q(\mathbf{x}_t \mid \mathbf{x}_0) = -\frac{\mathbf{x}_t - \alpha_t\mathbf{x}_0}{\sigma_t^2} = -\frac{\boldsymbol{\epsilon}}{\sigma_t}. \tag{3}$$

The marginal distribution of $\mathbf{x}_t$ is

$$p_t(\mathbf{x}_t) = \int p_{\mathsf{data}}(\mathbf{x}_0) q(\mathbf{x}_t \mid \mathbf{x}_0) d\mathbf{x}_0. \tag{4}$$

We have

$$\mathbf{s}^*(\mathbf{x}_t, t) = \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) = \frac{\nabla_{\mathbf{x}_t} p_t(\mathbf{x}_t)}{p_t(\mathbf{x}_t)} \tag{5}$$

$$= \frac{\int p_{\mathsf{data}}(\mathbf{x}_0) \nabla_{\mathbf{x}_t} q(\mathbf{x}_t \mid \mathbf{x}_0) d\mathbf{x}_0}{p_t(\mathbf{x}_t)} \tag{6}$$

$$\tag{7}$$

Using

$$\nabla_{\mathbf{x}_t} q(\mathbf{x}_t \mid \mathbf{x}_0) = q(\mathbf{x}_t \mid \mathbf{x}_0) \nabla_{\mathbf{x}_t} \log q(\mathbf{x}_t \mid \mathbf{x}_0), \tag{8}$$

we obtain

$$\mathbf{s}^*(\mathbf{x}_t, t) = \int \frac{p_{\mathsf{data}}(\mathbf{x}_0) q(\mathbf{x}_t \mid \mathbf{x}_0)}{p_t(\mathbf{x}_t)} \nabla_{\mathbf{x}_t} \log q(\mathbf{x}_t \mid \mathbf{x}_0) d\mathbf{x}_0 \tag{9}$$

$$= \int p(\mathbf{x}_0 \mid \mathbf{x}_t) \nabla_{\mathbf{x}_t} \log q(\mathbf{x}_t \mid \mathbf{x}_0) d\mathbf{x}_0 \tag{10}$$

$$= \mathbb{E}_{\mathbf{x}_0 | \mathbf{x}_t} \left[\nabla_{\mathbf{x}_t} \log q(\mathbf{x}_t \mid \mathbf{x}_0) \right] \tag{11}$$

$$= -\frac{1}{\sigma_t} \mathbb{E}[\boldsymbol{\epsilon} \mid \mathbf{x}_t]. \tag{12}$$

Recall that the optimal noise predictor is $\boldsymbol{\epsilon}^*(\mathbf{x}_t, t) = \mathbb{E}[\boldsymbol{\epsilon} \mid \mathbf{x}_t]$, then

$$\boxed{\boldsymbol{\epsilon}^*(\mathbf{x}_t, t) = -\sigma_t \mathbf{s}^*(\mathbf{x}_t, t)}. \tag{13}$$

Noise-denoiser relation

Using Eq. 1, we can solve for the noise as

$$\boldsymbol{\epsilon} = \frac{\mathbf{x}_t - \alpha_t \mathbf{x}_0}{\sigma_t}. \tag{14}$$

Taking the conditional expectation given $\mathbf{x}_t$ gives

$$\mathbb{E}[\boldsymbol{\epsilon} \mid \mathbf{x}_t] = \frac{\mathbf{x}_t - \alpha_t \mathbb{E}[\mathbf{x}_0 \mid \mathbf{x}_t]}{\sigma_t}. \tag{15}$$

Using the definitions

$$D^*(\mathbf{x}_t, t) = \mathbb{E}[\mathbf{x}_0 \mid \mathbf{x}_t], \qquad \boldsymbol{\epsilon}^*(\mathbf{x}_t, t) = \mathbb{E}[\boldsymbol{\epsilon} \mid \mathbf{x}_t], \tag{16}$$

we obtain

$$\boxed{\boldsymbol{\epsilon}^*(\mathbf{x}_t, t) = \frac{\mathbf{x}_t - \alpha_t D^*(\mathbf{x}_t, t)}{\sigma_t}}. \tag{17}$$

Noise-velocity relation

Taking conditional expectation given $\mathbf{x}_t$, we define the velocity field

$$\mathbf{v}^*(\mathbf{x}_t, t) = \mathbb{E}[\dot{\alpha}_t \mathbf{x}_0 + \dot{\sigma}_t \boldsymbol{\epsilon} \mid \mathbf{x}_t] = \dot{\alpha}_t D^*(\mathbf{x}_t, t) + \dot{\sigma}_t \boldsymbol{\epsilon}^*(\mathbf{x}_t, t), \tag{18}$$

where

$$D^*(\mathbf{x}_t, t) = \mathbb{E}[\mathbf{x}_0 | \mathbf{x}_t], \qquad \boldsymbol{\epsilon}^*(\mathbf{x}_t, t) = \mathbb{E}[\boldsymbol{\epsilon} | \mathbf{x}_t]. \tag{19}$$

Substituting the noise-denoiser relation

$$\boldsymbol{\epsilon}^*(\mathbf{x}_t, t) = \frac{\mathbf{x}_t - \alpha_t D^*(\mathbf{x}_t, t)}{\sigma_t}, \tag{20}$$

we obtain

$$\boxed{\mathbf{v}^*(\mathbf{x}_t, t) = \frac{\dot{\alpha}_t}{\alpha_t} \mathbf{x}_t + \dot{\sigma}_t (1 - \frac{\dot{\alpha}_t}{\alpha_t} \frac{\sigma_t}{\dot{\sigma}_t}) \boldsymbol{\epsilon}^*(\mathbf{x}_t, t)}. \tag{21}$$

Equivalence between the general EDM ODE Formulation and our α–σ parameterization

We show that the probability flow ODE used in EDM is equivalent to our α–σ parameterization.

α–σ **parameterization.** We parameterize the noise process as

$$\mathbf{x}_t = \alpha_t \mathbf{x}_0 + \sigma_t \boldsymbol{\epsilon}, \tag{22}$$

with the associated probability flow ODE given by

$$\frac{d\mathbf{x}_t}{dt} = \frac{\dot{\alpha}_t}{\alpha_t}\mathbf{x}_t + \sigma_t^2 \left(\frac{\dot{\alpha}_t}{\alpha_t} - \frac{\dot{\sigma}_t}{\sigma_t}\right) \mathbf{s}_\theta(\mathbf{x}_t, t). \tag{23}$$

EDM parameterization. EDM introduces a time-dependent scaling of the noisy variable as

$$\mathbf{x}_t = s(t)\hat{\mathbf{x}}_t, \tag{24}$$

where the underlying noise process is defined in the normalized space as

$$\hat{\mathbf{x}}_t = \mathbf{x}_0 + \sigma(t)\boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, I). \tag{25}$$

The associated probability flow ODE is given by

$$\frac{d\mathbf{x}_t}{dt} = \frac{\dot{s}_t}{s_t}\mathbf{x}_t - s_t^2 \sigma_t \dot{\sigma}_t \nabla_{\mathbf{x}_t} \log p\left(\frac{\mathbf{x}_t}{s_t}; \sigma_t\right). \tag{26}$$

Proof of the equivalence. Applying the chain rule,

$$\frac{d\mathbf{x}_t}{dt} = \frac{\dot{s}(t)}{s(t)}\mathbf{x}_t + s(t)\frac{d\hat{\mathbf{x}}_t}{dt}. \tag{27}$$

In the normalized space, the probability flow ODE becomes (Table 4 (b))

$$\frac{d\hat{\mathbf{x}}_t}{dt} = -\sigma_t \dot{\sigma}_t \nabla_{\hat{\mathbf{x}}_t} \log p(\hat{\mathbf{x}}_t; \sigma_t). \tag{28}$$

Using the change of variables $\nabla_{\hat{\mathbf{x}}_t} = s(t)\nabla_{\mathbf{x}_t}$, we obtain

$$\frac{d\hat{\mathbf{x}}_t}{dt} = -s(t)\sigma_t \dot{\sigma}_t \nabla_{\mathbf{x}_t} \log p\left(\frac{\mathbf{x}_t}{s(t)}; \sigma_t\right). \tag{29}$$

Substituting into the chain rule yields Eq 26.

References

[1] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In *NeurIPS*, 2019.

[2] Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *NeurIPS*, 2022.

[3] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *NeurIPS*, 2020.

[4] Nanye Ma, Mark Goldstein, Michael S Albergo, Nicholas M Boffi, Eric Vanden-Eijnden, and Saining Xie. Sit: Exploring flow and diffusion-based generative models with scalable interpolant transformers. In *ECCV*, 2024.