SEEDS: Exponential SDE Solvers for Fast High-Quality Sampling from Diffusion Models

Diffusion Journal Club

Reminder

- Forward diffusion process (described via SDE)

General Isotropic DE Formulation. The evolution of a data sample $\mathbf{x}_0 \in \mathbb{R}^d$ taken from an unknown data distribution p_{data} into standard Gaussian noise can be defined as a forward diffusion process $\{\mathbf{x}_t\}_{t\in[0,T]}$, with $T > 0$, which is a solution to a linear SDE:

$$\boxed{\mathrm{d}\mathbf{x}_t = f(t)\mathbf{x}_t\mathrm{d}t + g(t)\mathrm{d}\boldsymbol{\omega}_t,} \qquad f(t) := \frac{\mathrm{d}\log\alpha_t}{\mathrm{d}t}, \quad g(t) = \alpha_t\sqrt{\frac{\mathrm{d}[\sigma_t^2]}{\mathrm{d}t}}, \tag{1}$$

where $f(t), g(t) \in \mathbb{R}^{d\times d}$ are called the drift and diffusion coefficients respectively and $\boldsymbol{\omega}$ is a d-dimensional standard Wiener process, and $\alpha_t, \sigma_t \in \mathbb{R}^{>0}$ are differentiable functions with bounded derivatives. In practice, when specifying the SDE (1), σ_t acts as a schedule controlling the noise

- Has associated backward process

$$\mathrm{d}\mathbf{x}_t = [f(t)\mathbf{x}_t - \frac{1+\ell^2}{2}g^2(t)\boxed{\nabla_{\mathbf{x}_t}\log p_t(\mathbf{x}_t)}]\mathrm{d}t + \ell g(t)\mathrm{d}\bar{\boldsymbol{\omega}}_t,$$

Hard :(→ learn via ANN

Reminder

- During training learn to approximate score (or equivalently to predict noise)

Training. Denoising score-matching is a technique to train a time-dependent model $D_\theta(\mathbf{x}_t, t)$ to approach the score function $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$ at each time t. Intuitively, as D_θ approaches the score, it produces a sample which maximizes the log-likelihood. As such, this model is coined as a *data prediction* model. However, in practice DPMs can be more efficiently trained by reparameterizing D_θ into a different model $F_\theta(\mathbf{x}_t, t)$ whose objective is to predict the noise to be removed from a sample at time t. This *noise prediction* model is trained by means of the loss

$$\mathbb{E}_{t\sim\mathcal{U}[0,T],\mathbf{x}_0\sim p_{\mathrm{data}},\epsilon\sim\mathcal{N}(\mathbf{0},\mathbf{I}_d)}[\|\epsilon - F_\theta(\mu_t\mathbf{x}_0 + \boldsymbol{K}_t\epsilon, t)\|^2_{\boldsymbol{K}_t^{-1}\gamma_t\boldsymbol{K}_t^{-\top}}],$$

where γ_t is a time dependent weighting parameter and $\boldsymbol{K}_t\boldsymbol{K}_t^\top = \Sigma_t$.

- During sampling use learned score to solve backward SDE

Find good way to do this

Consider

$$
\begin{aligned}
\mathrm{d}\mathbf{x}_t &= [A(t)\mathbf{x}_t + b(t)F_\theta(\mathbf{x}_t, t)]\mathrm{d}t + g(t)\mathrm{d}\boldsymbol{\omega}_t, \\
\Leftrightarrow \quad \mathbf{x}_t &= \mathbf{x}_s + \int_s^t [A(\tau)\mathbf{x}_\tau + b(\tau)F_\theta(\mathbf{x}_\tau, \tau)]\mathrm{d}\tau + \int_s^t g(\tau)\mathrm{d}\boldsymbol{\omega}_\tau.
\end{aligned}
$$

Main Ideas behind SEEDS:

1. The variation-of-parameters formula: representing analytic solutions with linear term extracted from the integrand;
2. Exponentially weighted integrals: extracting the time-varying linear coefficient attached to the network from the integrand by means of a specific choice of change of variables which allows analytic computation of the leading coefficients in the truncated Itô-Taylor expansion associated to $F_\theta(\mathbf{x}_\tau, \tau)$ up to any arbitrary order;
3. Modified Gaussian increments: after replicating such change of variables onto the stochastic integral, analytically computing its variance.
4. Markov-preserving noise decomposition: stochastic integrals need to be dependent on overlapping time intervals and independent on non-overlapping ones.

Step 1: Use exponential integrators

Main Idea: Use analytical solution of linear part of DG

$$\mathrm{d}\mathbf{x}_t = [\boxed{A(t)\mathbf{x}_t} + b(t)F_\theta(\mathbf{x}_t, t)]\mathrm{d}t + g(t)\mathrm{d}\boldsymbol{\omega}_t,$$

$$\mathbf{x}_t = \Phi_A(t,s)\mathbf{x}_s + \int_s^t \Phi_A(t,\tau)b(\tau)F_\theta(\mathbf{x}_\tau,\tau)\mathrm{d}\tau + \int_s^t \Phi_A(t,\tau)g(\tau)\mathrm{d}\boldsymbol{\omega}_\tau,$$

Where $\Phi_A(t,s) = \exp\left(\int_s^t A(\tau)\mathrm{d}\tau\right)$

Proof of this form follows →

Some DG background

1-d timeinvariant linear DG

$$\boxed{\frac{\mathrm{d}x}{\mathrm{d}t} = F\,x, \qquad x(0) = \text{given},}$$

$$x(t) = x(0) + \int_0^t F\,x(\tau)\,\mathrm{d}\tau.$$

$$x(t) = x(0) + \int_0^t F\,x(\tau)\,\mathrm{d}\tau$$

$$x(t) = x(0) + \int_0^t F\,x(\tau)\,\mathrm{d}\tau$$

$$= x(0) + F\,x(0)\,t + \int_0^t \int_0^\tau F^2\,x(\tau')\,\mathrm{d}\tau'\,\mathrm{d}\tau$$

$$= x(0) + F\,x(0)\,t + F^2\,x(0)\,\frac{t^2}{2} + \int_0^t \int_0^\tau \int_0^{\tau'} F^3\,x(\tau'')\,\mathrm{d}\tau''\,\mathrm{d}\tau'\,\mathrm{d}\tau$$

$$= x(0) + F\,x(0)\,t + F^2\,x(0)\,\frac{t^2}{2} + F^3\,x(0)\,\frac{t^3}{6} + \cdots$$

$$= \left(1 + F\,t + \frac{F^2\,t^2}{2!} + \frac{F^3\,t^3}{3!} + \cdots\right) x(0).$$

→ Solution $x(t) = \exp(F\,t)\,x(0).$

Some DG background

more-dimensional timeinvariant linear DG

$$\boxed{\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{F}\,\mathbf{x}, \qquad \mathbf{x}(0) = \text{given},}$$

→ Solution $\mathbf{x}(t) = \exp(\mathbf{F}\,t)\,\mathbf{x}(0).$

Matrix Exponential

$$\exp(\mathbf{F}\,t) = \mathbf{I} + \mathbf{F}\,t + \frac{\mathbf{F}^2\,t^2}{2!} + \frac{\mathbf{F}^3\,t^3}{3!} + \cdots$$

Some DG background

more-dimensional linear DG

$$\boxed{\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{F}(t)\,\mathbf{x}, \quad \mathbf{x}(t_0) = \text{given},}$$

→ Solution $\quad \mathbf{x}(t) = \mathbf{\Psi}(t, t_0)\,\mathbf{x}(t_0),$

where $\mathbf{\Psi}(t, t_0)$ is the *transition matrix* which is defined via the properties

$$\begin{aligned}
\frac{\partial \mathbf{\Psi}(\tau, t)}{\partial \tau} &= \mathbf{F}(\tau)\,\mathbf{\Psi}(\tau, t), \\
\frac{\partial \mathbf{\Psi}(\tau, t)}{\partial t} &= -\mathbf{\Psi}(\tau, t)\,\mathbf{F}(t), \\
\mathbf{\Psi}(\tau, t) &= \mathbf{\Psi}(\tau, s)\,\mathbf{\Psi}(s, t), \\
\mathbf{\Psi}(t, \tau) &= \mathbf{\Psi}^{-1}(\tau, t), \\
\mathbf{\Psi}(t, t) &= \mathbf{I}.
\end{aligned} \tag{2.34}$$

Compare

$$\mathrm{d}\mathbf{x}_\tau = [A(\tau)\mathbf{x}_\tau + \cancel{b(\tau)F_\theta(\mathbf{x}_\tau, \tau)}]\mathrm{d}\tau + \cancel{g(\tau)\mathrm{d}\boldsymbol{\omega}_\tau}$$

$$\Phi_A(t, s) \;=\; \exp\left(\int_s^t A(\tau)\mathrm{d}\tau\right)$$

Get exponential integral

$$\mathrm{d}\mathbf{x}_\tau = [A(\tau)\mathbf{x}_\tau + b(\tau)F_\theta(\mathbf{x}_\tau,\tau)]\mathrm{d}\tau + g(\tau)\mathrm{d}\boldsymbol{\omega}_\tau$$

Backwards SDE

$$\mathrm{d}\mathbf{x}_\tau - A(\tau)\mathbf{x}_\tau\mathrm{d}\tau = b(\tau)F_\theta(\mathbf{x}_\tau,\tau)\mathrm{d}\tau + g(\tau)\mathrm{d}\boldsymbol{\omega}_\tau$$

Bring linear term on other side

$$\underbrace{\Phi_A(t,\tau)[\mathrm{d}\mathbf{x}_\tau - A(\tau)\mathbf{x}_\tau\mathrm{d}\tau]}_{= \Phi_A(t,\tau)\mathrm{d}\mathbf{x}_\tau - \Phi_A(t,\tau)A(\tau)\mathbf{x}_\tau\mathrm{d}\tau} = \Phi_A(t,\tau)b(\tau)F_\theta(\mathbf{x}_\tau,\tau)\mathrm{d}\tau + \Phi_A(t,\tau)g(\tau)\mathrm{d}\boldsymbol{\omega}_\tau$$

Multiply with Transition matrix

$$\frac{\mathrm{d}}{\mathrm{d}\tau}[\Phi_A(t,\tau)\mathbf{x}_\tau\mathrm{d}\tau] = \Phi_A(t,\tau)b(\tau)F_\theta(\mathbf{x}_\tau,\tau)\mathrm{d}\tau + \Phi_A(t,\tau)g(\tau)\mathrm{d}\boldsymbol{\omega}_\tau$$

The left side can be written as a differentiation

$$\underbrace{\int_s^t \frac{\mathrm{d}}{\mathrm{d}\tau}\Phi_A(t,\tau)\mathbf{x}_\tau\mathrm{d}\tau}_{\substack{= \Phi_A(t,t)\mathbf{x}_t - \Phi_A(t,s)\mathbf{x}_s \\ = \mathbf{x}_t - \Phi_A(t,s)\mathbf{x}_s}} = \int_s^t \Phi_A(t,\tau)b(\tau)F_\theta(\mathbf{x}_\tau,\tau)\mathrm{d}\tau + \int_s^t \Phi_A(t,\tau)g(\tau)\mathrm{d}\boldsymbol{\omega}_\tau$$

Integrate from s to t

$$\mathbf{x}_t = \Phi_A(t,s)\mathbf{x}_s + \int_s^t \Phi_A(t,\tau)b(\tau)F_\theta(\mathbf{x}_\tau,\tau)\mathrm{d}\tau + \int_s^t \Phi_A(t,\tau)g(\tau)\mathrm{d}\boldsymbol{\omega}_\tau,$$

Step 2: Use Ito-Taylor Expansion

$$\mathbf{x}_t = \Phi_A(t,s)\mathbf{x}_s + \int_s^t \Phi_A(t,\tau)b(\tau)F_\theta(\mathbf{x}_\tau,\tau)\mathrm{d}\tau + \int_s^t \Phi_A(t,\tau)g(\tau)\mathrm{d}\boldsymbol{\omega}_\tau,$$

$$F_\theta(\mathbf{x}_\tau,\tau) = \sum_{k=0}^{n} \frac{(\tau-s)^k}{k!} F_\theta^{(k)}(\mathbf{x}_s,s) + \mathcal{R}_n,$$

Ito-Taylor Expansion

All stochastic terms are in R_n

$$\int_s^t \Phi_A(t,\tau)b(\tau)F_\theta(\mathbf{x}_\tau,\tau)\mathrm{d}\tau = \sum_{k=0}^{n} F_\theta^{(k)}(\mathbf{x}_s,s)\int_s^t \Phi_A(t,\tau)b(\tau)\frac{(\tau-s)^k}{k!}\mathrm{d}\tau + \tilde{\mathcal{R}}_n,$$

Idea Ito-Taylor Expansion

Theorem 4.2 (Itô formula)**.** *Assume that* $\mathbf{x}(t)$ *is an Itô process, and consider an arbitrary (scalar) function* $\phi(\mathbf{x}(t), t)$ *of the process. Then the Itô differential of* ϕ*, that is, the Itô SDE for* ϕ *is given as*

$$\begin{aligned} \mathrm{d}\phi &= \frac{\partial \phi}{\partial t}\,\mathrm{d}t + \sum_i \frac{\partial \phi}{\partial x_i}\,\mathrm{d}x_i + \frac{1}{2}\sum_{i,j}\left(\frac{\partial^2 \phi}{\partial x_i \partial x_j}\right)\mathrm{d}x_i\,\mathrm{d}x_j \\ &= \frac{\partial \phi}{\partial t}\,\mathrm{d}t + (\nabla\phi)^\mathsf{T}\,\mathrm{d}\mathbf{x} + \frac{1}{2}\operatorname{tr}\left\{\left(\nabla\nabla^\mathsf{T}\phi\right)\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{x}^\mathsf{T}\right\}, \end{aligned} \tag{4.14}$$

provided that the required partial derivatives exist, where the mixed differentials are combined according to the rules

$$\begin{aligned} \mathrm{d}\boldsymbol{\beta}\,\mathrm{d}t &= \mathbf{0}, \\ \mathrm{d}t\,\mathrm{d}\boldsymbol{\beta} &= \mathbf{0}, \\ \mathrm{d}\boldsymbol{\beta}\,\mathrm{d}\boldsymbol{\beta}^\mathsf{T} &= \mathbf{Q}\,\mathrm{d}t. \end{aligned} \tag{4.15}$$

Ito-Taylor-Expansion: Use Ito Formula iteratively, similar to Taylor Expansion

Step 3: Calculate Rest Integral 2

$$\int_s^t \Phi_A(t,\tau)b(\tau)F_\theta(\mathbf{x}_\tau,\tau)\mathrm{d}\tau = \sum_{k=0}^{n} F_\theta^{(k)}(\mathbf{x}_s,s)\int_s^t \Phi_A(t,\tau)b(\tau)\frac{(\tau-s)^k}{k!}\mathrm{d}\tau + \tilde{\mathcal{R}}_n,$$

The third key contribution of our work is to rewrite, for any $k \geqslant 0$, the integral $\int_s^t \Phi_A(t,\tau)b(\tau)\frac{(\tau-s)^k}{k!}\mathrm{d}\tau$ as an integral of the form $\int_{\lambda_s}^{\lambda_t} e^{\lambda}\frac{(\lambda-\lambda_s)^k}{k!}\mathrm{d}\lambda$ since the latter is recursively analytically computed in terms of the φ-functions

$$\varphi_0(t) := e^t, \qquad \varphi_{k+1}(t) := \int_0^1 e^{(1-\tau)t}\frac{\tau^k}{k!}\mathrm{d}\tau = \frac{\varphi_k(t)-\varphi_k(0)}{t}, \qquad k \geqslant 0.$$

This parts makes more sense in the VPSDE case ^^'

Step 4: Calculate Integral 2

4. Markov-preserving noise decomposition: stochastic integrals need to be dependent on overlapping time intervals and independent on non-overlapping ones.

$$\mathbf{x}_t = \Phi_A(t,s)\mathbf{x}_s + \int_s^t \Phi_A(t,\tau)b(\tau)F_\theta(\mathbf{x}_\tau,\tau)\mathrm{d}\tau + \int_s^t \Phi_A(t,\tau)g(\tau)\mathrm{d}\boldsymbol{\omega}_\tau,$$

Modified Gaussian increments. In order for making such change of variables to be consistent on the overall system, one needs to replicate it accordingly in the stochastic integral $\int_s^t \Phi_A(t,\tau)g(\tau)\mathrm{d}\bar{\boldsymbol{\omega}}_\tau$. As such, our last key contribution is to transform it into an exponentially weighted stochastic integral with integration endpoints λ_s, λ_t and apply the Stochastic Exponential Time Differencing (SETD) method [1] to compute its variance analytically, as illustrated in (14) below.

This parts makes more sense in the VPSDE case ^^'

Variance Preserving SDE

Case: this forward SDE $$\mathrm{d}\mathbf{x} = -\frac{1}{2}\beta(t)\mathbf{x}\,\mathrm{d}t + \sqrt{\beta(t)}\,\mathrm{d}\mathbf{w}.$$

Reverse: $$\boxed{\mathrm{d}\mathbf{x}_t = [f(t)\mathbf{x}_t - \frac{1+\ell^2}{2}g^2(t)\nabla_{\mathbf{x}_t}\log p_t(\mathbf{x}_t)]\mathrm{d}t + \ell g(t)\mathrm{d}\bar{\boldsymbol{\omega}}_t,} \tag{2}$$

$$f(t) = \frac{\mathrm{d}\log\alpha_t}{\mathrm{d}t}, \quad g^2(t) = 2\bar{\sigma}_t^2\left(\frac{\mathrm{d}\log\bar{\sigma}_t}{\mathrm{d}t} - \frac{\mathrm{d}\log\alpha_t}{\mathrm{d}t}\right) = -2\bar{\sigma}_t^2\frac{\mathrm{d}\lambda_t}{\mathrm{d}t}, \qquad \frac{\bar{\sigma}_t}{\alpha_t} = e^{-\lambda_t}.$$

The VPSDE case. Let $\tilde{\alpha}_t := \frac{1}{2}\beta_d t^2 + \beta_m t$, where $\beta_d, \beta_m > 0$ and $t \in [0,1]$. Then, by denoting

$$\sigma_t := \sqrt{e^{\tilde{\alpha}_t} - 1}, \qquad \alpha_t := e^{-\frac{1}{2}\tilde{\alpha}_t}, \qquad \bar{\sigma}_t := \alpha_t\sigma_t, \qquad \nabla_{\mathbf{x}_t}\log p_t(\mathbf{x}_t) \simeq \bar{\sigma}_t^{-1}F_\theta(\mathbf{x}_t, t), \tag{7}$$

we obtain the VP SDE framework from [34] and the following result.

Proposition 3.1. *Let $t < s$. The analytic solution at time t of the RSDE* (2) *with coefficients* (7) *and initial value $\mathbf{x}_s$ is*

$$\mathbf{x}_t = \frac{\alpha_t}{\alpha_s}\mathbf{x}_s - 2\alpha_t\int_{\lambda_s}^{\lambda_t} e^{-\lambda}\hat{F}_\theta(\widehat{\mathbf{x}}_\lambda, \lambda)\mathrm{d}\lambda - \sqrt{2}\alpha_t\int_{\lambda_s}^{\lambda_t} e^{-\lambda}\mathrm{d}\bar{\boldsymbol{\omega}}_\lambda, \qquad \lambda_t := -\log(\sigma_t). \tag{8}$$

$$! \ \Phi(t,s) = \frac{\alpha_t}{\alpha_s}$$

$$\mathbf{x}_t = \Phi_A(t,s)\mathbf{x}_s + \int_s^t \Phi_A(t,\tau)b(\tau)F_\theta(\mathbf{x}_\tau, \tau)\mathrm{d}\tau + \int_s^t \Phi_A(t,\tau)g(\tau)\mathrm{d}\boldsymbol{\omega}_\tau,$$

Compare

B.2.2 Proof of Proposition 3.1

First of all, denote $F_\theta(\mathbf{x}_t, (M-1)t) = \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)$. We have

$$f(t) = \frac{\mathrm{d}\log\alpha_t}{\mathrm{d}t}, \quad g^2(t) = 2\bar{\sigma}_t^2\left(\frac{\mathrm{d}\log\bar{\sigma}_t}{\mathrm{d}t} - \frac{\mathrm{d}\log\alpha_t}{\mathrm{d}t}\right) = -2\bar{\sigma}_t^2\frac{\mathrm{d}\lambda_t}{\mathrm{d}t}, \qquad \frac{\bar{\sigma}_t}{\alpha_t} = e^{-\lambda_t}.$$

This way, one can directly relate λ_t with the *signal-to-noise ratio* $\mathrm{SNR}(t) = \alpha_t^2/\bar{\sigma}_t^2$, also being used in [22]. As such, $\mathrm{SNR}(t)$ is strictly monotonically decreasing in time. Thus, the analytic solution to (2) yields

$$\begin{aligned}
\mathbf{x}_t &= e^{\int_s^t f(\tau)\mathrm{d}\tau}\mathbf{x}_s + \int_s^t \left(e^{\int_\tau^t f(r)\mathrm{d}r}\frac{g^2(\tau)}{\bar{\sigma}_\tau}\boldsymbol{\epsilon}_\theta(\mathbf{x}_\tau,\tau)\right)\mathrm{d}\tau + \int_s^t \left(e^{\int_\tau^t f(r)\mathrm{d}r} g(\tau)\right)\mathrm{d}\bar{\boldsymbol{\omega}}(\tau) \\
&= \frac{\alpha_t}{\alpha_s}\mathbf{x}_s + \alpha_t\int_s^t \frac{g^2(\tau)}{\alpha_\tau\bar{\sigma}_\tau}\boldsymbol{\epsilon}_\theta(\mathbf{x}_\tau,\tau)\mathrm{d}\tau + \alpha_t\int_s^t \frac{g(\tau)}{\alpha_\tau}\mathrm{d}\bar{\boldsymbol{\omega}}(\tau) \\
&= \frac{\alpha_t}{\alpha_s}\mathbf{x}_s - \alpha_t\int_s^t \frac{2\sigma_\tau^2}{\alpha_\tau\bar{\sigma}_\tau}\frac{\mathrm{d}\lambda_\tau}{\mathrm{d}\tau}\boldsymbol{\epsilon}_\theta(\mathbf{x}_\tau,\tau)\mathrm{d}\tau + \alpha_t\int_s^t \frac{g(\tau)}{\alpha_\tau}\mathrm{d}\bar{\boldsymbol{\omega}}(\tau) \\
&= \frac{\alpha_t}{\alpha_s}\mathbf{x}_s - 2\alpha_t\int_s^t \frac{\bar{\sigma}_\tau}{\alpha_\tau}\frac{\mathrm{d}\lambda_\tau}{\mathrm{d}\tau}\boldsymbol{\epsilon}_\theta(\mathbf{x}_\tau,\tau)\mathrm{d}\tau + \alpha_t\int_s^t \frac{g(\tau)}{\alpha_\tau}\mathrm{d}\bar{\boldsymbol{\omega}}(\tau) \\
&= \frac{\alpha_t}{\alpha_s}\mathbf{x}_s - 2\alpha_t\int_s^t e^{-\lambda_\tau}\frac{\mathrm{d}\lambda_\tau}{\mathrm{d}\tau}\boldsymbol{\epsilon}_\theta(\mathbf{x}_\tau,\tau)\mathrm{d}\tau - \sqrt{2}\alpha_t\int_s^t e^{-\lambda_\tau}\sqrt{\frac{\mathrm{d}\lambda_\tau}{\mathrm{d}\tau}}\mathrm{d}\bar{\boldsymbol{\omega}}(\tau).
\end{aligned}$$

Step 1

$! \ \Phi(t,s) = \frac{\alpha_t}{\alpha_s}$

Step 2a: Change of Variables

By using the change of variables to $\lambda(t)$, our equation now reads

$$\mathbf{x}_t = \frac{\alpha_t}{\alpha_s}\mathbf{x}_s - 2\alpha_t\int_{\lambda_s}^{\lambda_t} e^{-\lambda}\hat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda,\lambda)\mathrm{d}\lambda - \sqrt{2}\alpha_t\int_{\lambda_s}^{\lambda_t} e^{-\lambda}\mathrm{d}\bar{\boldsymbol{\omega}}(\lambda). \tag{26}$$

Finally, notice that $\alpha_t = \sqrt{\frac{1}{1+e^{-2\lambda_t}}}$ and $\bar{\sigma}_t = \sqrt{\frac{1}{1+e^{2\lambda_t}}}$ so that (26) is equivalent to

$$\widehat{\mathbf{x}}_{\lambda_t} = \frac{\hat{\alpha}_{\lambda_t}}{\hat{\alpha}_{\lambda_s}}\widehat{\mathbf{x}}_{\lambda_s} - 2\hat{\alpha}_{\lambda_t}\int_{\lambda_s}^{\lambda_t} e^{-\lambda}\hat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda,\lambda)\mathrm{d}\lambda - \sqrt{2}\hat{\alpha}_{\lambda_t}\int_{\lambda_s}^{\lambda_t} e^{-\lambda}\mathrm{d}\bar{\boldsymbol{\omega}}(\lambda).$$

This finishes the proof.

The VPSDE case. Let $\tilde{\alpha}_t := \frac{1}{2}\beta_d t^2 + \beta_m t$, where $\beta_d, \beta_m > 0$ and $t \in [0,1]$. Then, by denoting

$$\sigma_t := \sqrt{e^{\tilde{\alpha}_t} - 1}, \qquad \alpha_t := e^{-\frac{1}{2}\tilde{\alpha}_t}, \qquad \bar{\sigma}_t := \alpha_t \sigma_t, \qquad \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) \simeq \bar{\sigma}_t^{-1} F_\theta(\mathbf{x}_t, t), \quad (7)$$

we obtain the VP SDE framework from [34] and the following result.

Proposition 3.1. *Let $t < s$. The analytic solution at time t of the RSDE* (2) *with coefficients* (7) *and initial value $\mathbf{x}_s$ is*

$$\mathbf{x}_t = \frac{\alpha_t}{\alpha_s}\mathbf{x}_s - 2\alpha_t \underbrace{\int_{\lambda_s}^{\lambda_t} e^{-\lambda} \hat{F}_\theta(\widehat{\mathbf{x}}_\lambda, \lambda) \mathrm{d}\lambda}_{} - \sqrt{2}\alpha_t \underbrace{\int_{\lambda_s}^{\lambda_t} e^{-\lambda} \mathrm{d}\bar{\boldsymbol{\omega}}_\lambda}_{}, \qquad \lambda_t := -\log(\sigma_t). \quad (8)$$

Step 3

$$\int_s^t \Phi_A(t,\tau) b(\tau) F_\theta(\mathbf{x}_\tau, \tau) \mathrm{d}\tau = \sum_{k=0}^{n} F_\theta^{(k)}(\mathbf{x}_s, s) \underbrace{\int_s^t \Phi_A(t,\tau) b(\tau) \frac{(\tau - s)^k}{k!} \mathrm{d}\tau}_{} + \tilde{\mathcal{R}}_n,$$

$$\int_{\lambda_s}^{\lambda_t} e^{-\lambda} \frac{(\lambda - \lambda_s)^k}{k!} \mathrm{d}\lambda = \sigma_t h^{k+1} \varphi_{k+1}(h).$$

Step 4 $\underbrace{\int_{\lambda_s}^{\lambda_t} e^{-\lambda} \mathrm{d}\bar{\boldsymbol{\omega}}_\lambda}_{}$

obeys a normal distribution with zero mean, and one can analytically compute its variance:

$$\int_{\lambda_s}^{\lambda_t} e^{-2\lambda} \mathrm{d}\lambda = \frac{\sigma_t^2}{2}(e^{2h} - 1). \quad (10)$$

$$\boxed{\begin{aligned} \varphi_0(t) &:= e^t, \\ \varphi_{k+1}(t) &:= \int_0^1 e^{(1-\tau)t} \frac{\tau^k}{k!} \mathrm{d}\tau = \frac{\varphi_k(t) - \varphi_k(0)}{t}, \qquad k \geqslant 0. \end{aligned}}$$

From Step 3

The EDM case. Denote σ_d^2 the variance of the considered initial dataset and set

$$\sigma_t := t, \alpha_t := 1, \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) \simeq \frac{1}{t^2}\left[\frac{\sigma_d^2 \mathbf{x}_t}{t^2+\sigma_d^2} + \frac{t\sigma_d}{\sqrt{t^2+\sigma_d^2}} F_\theta\left(\frac{\mathbf{x}_t}{\sqrt{t^2+\sigma_d^2}}, \frac{\log(t)}{4}\right)\right]. \quad (11)$$

These parameters correspond to the preconditioned EDM framework introduced in [16, Sec. 5, App. B.6]. The following result is the basis for constructing our customized SEEDS in this case, and for which we report experimental results in Table 1. For simplicity, we will write $F_\theta(\mathbf{x}_t, t)$ for the preconditioned model in (11) and we refer to Appendix B for details.

Proposition 3.2. *Let $t < s$. The analytic solution at time t of* (2) *with coefficients* (11) *and initial value $\mathbf{x}_s$ is, for $\ell = 1$,*

$$\mathbf{x}_t = \frac{t^2+\sigma_d^2}{s^2+\sigma_d^2}\mathbf{x}_s + 2(t^2+\sigma_d^2)\int_{\lambda_s}^{\lambda_t} e^{-\lambda}\hat{F}_\theta(\widehat{\mathbf{x}}_\lambda, \lambda)\mathrm{d}\lambda - \sqrt{2}(t^2+\sigma_d^2)\int_{\lambda_s}^{\lambda_t} e^{-\lambda}\mathrm{d}\overline{\boldsymbol{\omega}}_\lambda, \quad (12)$$

where $\lambda_t := -\log\left[\frac{t}{\sigma_d\sqrt{t^2+\sigma_d^2}}\right]$. In the case when $\ell = 0$, it is given by

$$\mathbf{x}_t = \sqrt{\frac{t^2+\sigma_d^2}{s^2+\sigma_d^2}}\mathbf{x}_s + \sqrt{t^2+\sigma_d^2}\int_{\lambda_s}^{\lambda_t} e^{-\lambda}\hat{F}_\theta(\widehat{\mathbf{x}}_\lambda, \lambda)\mathrm{d}\lambda, \quad \lambda_t := -\log\left[\arctan\left[\frac{t}{\sigma_d}\right]\right]. \quad (13)$$

Remark 3.3. One can wonder about the generality of such change of variables. Our method is very general in that one can always make such change of variables with very mild regularity conditions: for $c : [0, T] \longrightarrow \mathbb{R}^{>0}$ integrable, with primitive $C(t) > 0$, we have $c(t) = e^{\log(c(t))}$. This means we can write $c(t) = \dot{C}(t) = e^{\lambda_t}\dot{\lambda}_t$ with $\lambda_t = \log(C(t))$. In other words, for such c, we have

$$\int_s^t c(\tau)\mathrm{d}\tau = \int_s^t e^{\lambda_\tau}\dot{\lambda}_\tau \mathrm{d}\tau = \int_{\lambda_s}^{\lambda_t} e^{\lambda}\mathrm{d}\lambda.$$

SEEDS Algorithm

In this section we present our SEEDS algorithms by putting together all the ingredients presented in the previous section. Let $t < s$. In all what follows, we consider the analytic solution at time t of the RSDE (2) with coefficients (7), $h = \lambda_t - \lambda_s$ and initial value $\mathbf{x}_s$. Plugging (9) with $k = 0$ and (10) into the exact solution (8) allow us to infer the first SEEDS scheme, given by iterations of the form

$$\tilde{\mathbf{x}}_t = \frac{\alpha_t}{\alpha_s}\tilde{\mathbf{x}}_s - 2\bar{\sigma}_t(e^h - 1)\hat{F}_\theta(\widehat{\mathbf{x}}_{\lambda_s}, \lambda_s) - \bar{\sigma}_t\sqrt{e^{2h} - 1}\epsilon, \qquad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d). \tag{14}$$

SEEDS-1: Use only first term of expansion

Theorem 4.1. *Under Assumption C.1, the numerical solution $\tilde{\mathbf{x}}_t$ produced by the SEEDS-1 method* (14) *converges to the exact solution $\mathbf{x}_t$ of*

$$\mathrm{d}\mathbf{x}_t = [f(t)\mathbf{x}_t + g^2(t)\bar{\sigma}_t^{-1}F_\theta(\mathbf{x}_t, t)]\mathrm{d}t + g(t)\mathrm{d}\boldsymbol{\omega}_t, \qquad (\bar{\sigma}_t^{-1} := 1/\bar{\sigma}_t) \tag{15}$$

with coefficients (7) *in Mean-Square sense with strong order 1.0: there is a constant $C > 0$ such that*

$$\sqrt{\mathbb{E}\left[\sup_{0\leqslant t\leqslant 1}|\tilde{\mathbf{x}}_t - \mathbf{x}_t|^2\right]} \leqslant Ch, \qquad \textit{as } h \longrightarrow 0.$$

SEEDS Algorithm

Algorithm 1 Iterative procedure

Input: initial value $\mathbf{x}_T$, steps $\{t_i\}_{i=0}^{M}$, model F_θ
Initialize $\tilde{\mathbf{x}}_{t_0} \leftarrow \mathbf{x}_T$
for $i = 1$ **to** $M-1$ **do**
  $(t, s) \leftarrow (t_i, t_{i-1}), \quad h \leftarrow \lambda_t - \lambda_s$
  $\tilde{\mathbf{x}}_t = \text{SEEDS-k}(F_\theta, \tilde{\mathbf{x}}_s, s, t)$
end for
Return $\tilde{\mathbf{x}}_{t_M} \leftarrow \text{last-step}(\tilde{\mathbf{x}}_{t_{M-1}}, t_{M-1}, t_M)$

Algorithm 2 SEEDS-1$(F_\theta, \tilde{\mathbf{x}}_s, s, t)$

$z \leftarrow \mathcal{N}(0,1)$
$\tilde{\mathbf{x}}_t \leftarrow \frac{\alpha_t}{\alpha_s}\tilde{\mathbf{x}}_s - 2\bar{\sigma}_t\left(e^h - 1\right)F_\theta(\tilde{\mathbf{x}}_s, s) - \bar{\sigma}_t\sqrt{e^{2h}-1}z$

Algorithm 3 SEEDS-2$(F_\theta, \tilde{\mathbf{x}}_s, s, t)$

$s_1 \leftarrow t_\lambda(\lambda_s + \frac{h}{2}), \quad (z^1, z^2) \leftarrow \mathcal{N}(0, \mathrm{Id}) \otimes \mathcal{N}(0, \mathrm{Id})$
$\mathbf{u} \leftarrow \frac{\alpha_{s_1}}{\alpha_s}\tilde{\mathbf{x}}_s - 2\bar{\sigma}_{s_1}\left(e^{\frac{h}{2}} - 1\right)F_\theta(\tilde{\mathbf{x}}_s, s) - \boldsymbol{A}, \qquad \boldsymbol{A} := \bar{\sigma}_{s_1}\sqrt{e^h - 1}z^1$
$\tilde{\mathbf{x}}_t \leftarrow \frac{\alpha_t}{\alpha_s}\tilde{\mathbf{x}}_s - 2\bar{\sigma}_t(e^h - 1)F_\theta(\mathbf{u}, s_1) - \boldsymbol{B}, \qquad \boldsymbol{B} := \bar{\sigma}_t(\sqrt{e^{2h} - e^h}z^1 + \sqrt{e^h - 1}z^2)$

Algorithm 4 SEEDS-3$(F_\theta, \tilde{\mathbf{x}}_s, s, t)$ with $0 < r_1 < r_2 < 1$

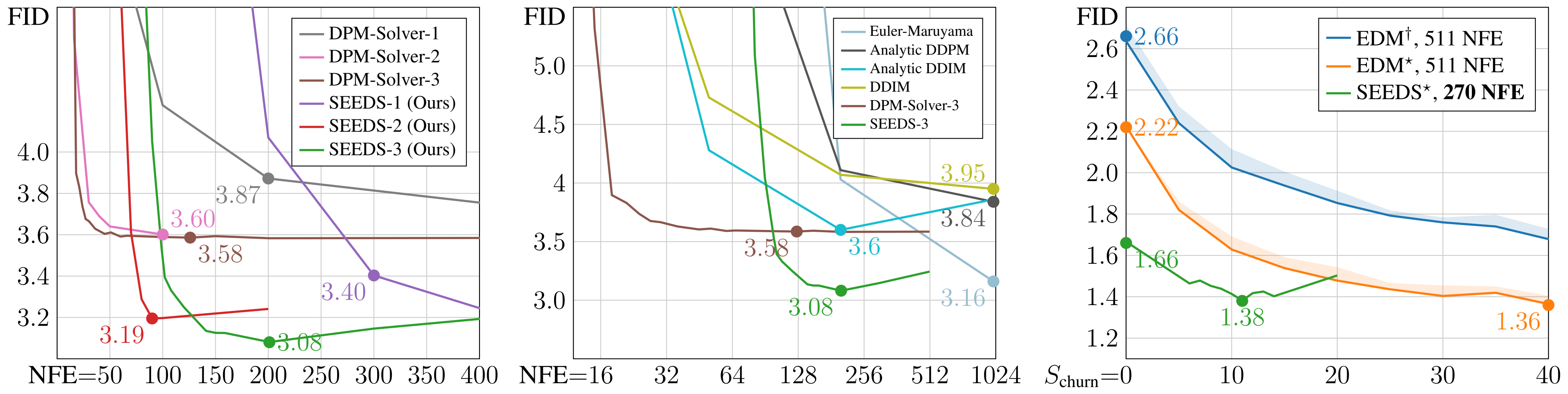
$s_1 \leftarrow t_\lambda\left(\lambda_s + r_1 h\right), \quad s_2 \leftarrow t_\lambda\left(\lambda_s + r_2 h\right), \quad (z^1, z^2, z^3) \leftarrow \mathcal{N}(0, \mathrm{Id})^{\otimes 3}$
$\mathbf{u}_1 \leftarrow \frac{\alpha_{s_1}}{\alpha_s}\tilde{\mathbf{x}}_s - 2\bar{\sigma}_{s_1}\left(e^{r_1 h} - 1\right)F_\theta(\tilde{\mathbf{x}}_s, s) - \bar{\sigma}_{s_1}\sqrt{e^{2r_1 h} - 1}z^1$
$\boldsymbol{A} \leftarrow \bar{\sigma}_{s_2}(\sqrt{e^{2r_2 h} - e^{2r_1 h}}z^1 + \sqrt{e^{2r_1 h} - 1}z^2)$
$\mathbf{u}_2 \leftarrow \frac{\alpha_{s_2}}{\alpha_s}\tilde{\mathbf{x}}_s - 2\bar{\sigma}_{s_2}\left(e^{r_2 h} - 1\right)F_\theta(\tilde{\mathbf{x}}_s, s) - 2\frac{\bar{\sigma}_{s_2} r_2}{r_1}\left(\frac{e^{r_2 h} - 1}{r_2 h} - 1\right)\left(F_\theta(\mathbf{u}_1, s_1) - F_\theta(\tilde{\mathbf{x}}_s, s)\right)) - \boldsymbol{A}$
$\boldsymbol{B} \leftarrow \bar{\sigma}_t(\sqrt{e^{2h} - e^{2r_2 h}}z^1 + \sqrt{e^{2r_2 h} - e^{2r_1 h}}z^2 + \sqrt{e^{2r_1 h} - 1}z^3)$
$\tilde{\mathbf{x}}_t \leftarrow \frac{\alpha_t}{\alpha_s}\tilde{\mathbf{x}}_s - 2\bar{\sigma}_t\left(e^h - 1\right)F_\theta(\tilde{\mathbf{x}}_s, s) - 2\frac{\bar{\sigma}_t}{r_2}\left(\frac{e^h - 1}{h} - 1\right)\left(F_\theta(\mathbf{u}_2, s_2) - F_\theta(\tilde{\mathbf{x}}_s, s)\right)) - \boldsymbol{B}$

Results

Table 1: Sample quality measured by FID↓ on pre-trained DPMs. We report the minimum FID obtained by each model and the NFE at which it was obtained. For CIFAR, CelebA and FFHQ, we use baseline pre-trained models [34, 16]. For ImageNet, we use the optimized pre-trained model from [16]. *discrete-time model, ⋆continuous-time model, †recomputed FID for the non-deep model.

SAMPLING METHOD	FID↓	NFE
CIFAR-10* VP-UNCOND.		
DDIM [33]	3.95	1000
ANALYTIC-DDPM [2]	3.84	1000
GENIE [9]	3.64	25
ANALYTIC-DDIM [2]	3.60	200
F-PNDM (LINEAR) [20]	3.60	250
DPM-SOLVER† [22]	3.48	44
F-PNDM (COSINE) [20]	3.26	1000
DDPM [12]	3.16	1000
SEEDS-3 (OURS)	**3.08**	**201**
CIFAR-10⋆ VP-COND.		
DPM-SOLVER† [22]	3.57	195
EDM ($S_{\text{churn}} = 0$) [16]	2.48	35
SEEDS-3 (OURS)	**2.08**	**129**
CIFAR-10⋆ VP-UNCOND.		
DPM-SOLVER† [22]	2.59	51
GGF [15]	2.59	180
GDDIM [41]	2.56	100
DEIS ρ3KUTTA [40]	2.55	50
EULER-MARUYAMA [34]	2.54	1024
STOCHASTIC EDM [16]	2.54	1534
SEEDS-3 (OURS)	2.39	165
EDM (OPTIMIZED) [16]	**2.27**	**511**

SAMPLING METHOD	FID↓	NFE
CELEBA-64* VP-UNCOND.		
ANALYTIC-DDPM [2]	5.21	1000
DDIM [33]	4.78	200
DDPM [12]	3.50	1000
GDDIM [41]	3.85	50
ANALYTIC-DDIM [2]	3.13	1000
3-DEIS [40]	2.95	50
F-PNDM (LINEAR) [20]	2.71	250
DPM-SOLVER [22]	2.71	36
SEEDS-3 (OURS)	**1.88**	**90**
FFHQ-64⋆ VP-UNCOND.		
DPM-SOLVER† [22]	3.52	90
SEEDS-3 (OURS)	3.40	150
EDM ($S_{\text{churn}} = 0$) [16]	**3.39**	**79**
IMAGENET-64 EDM-COND.		
DPM-SOLVER† [22]	3.01	270
EDM ($S_{\text{churn}} = 0$) [16]	2.22	511
SEEDS-3 (OURS)	1.38	270
EDM (OPTIMIZED) [16]	**1.36**	**511**



(a) Uncond. CIFAR, VP (b) Uncond. CIFAR, VP (c) Stochasticity on ImageNet-64

Figure 1: (a-b) Comparison of sample quality measured by FID ↓ of SEEDS, DPM-Solver and other methods for discretely trained DPMs on CIFAR-10 with varying number of function evaluations. (c) Effect of S_{churn} on SEEDS-3 (at NFE = 270) and EDM method (at NFE = 511) on class-conditional ImageNet-64. †baseline ADM model. ⋆EDM preconditioned model.

Results

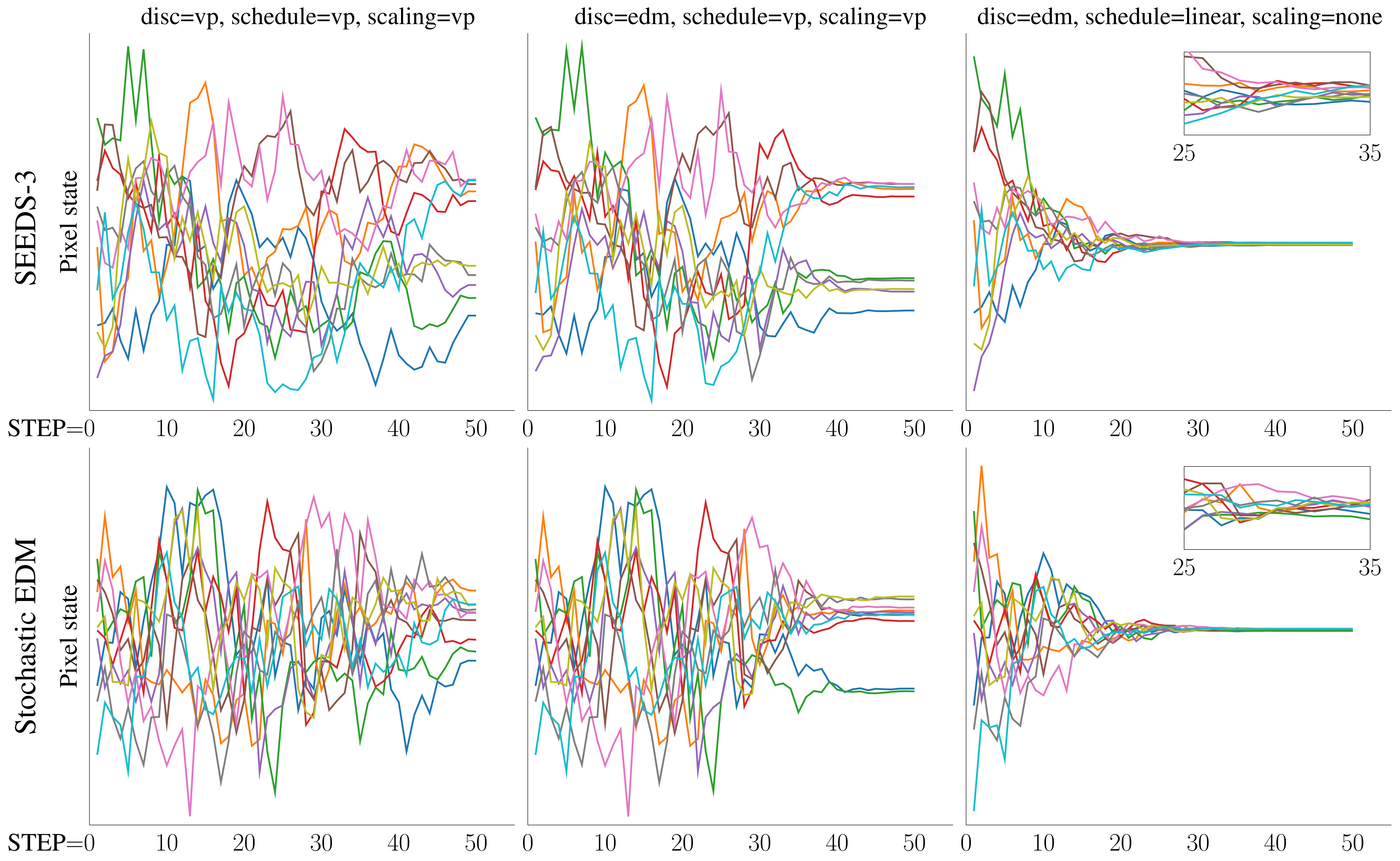


Figure 2: Trajectories of 10 pixels (R channel) sampled from SEEDS (1st line) and Stochastic EDM (2nd line) on the optimized pre-trained model [16] on ImageNet64. Schedule=scaling=vp corresponds to the VP coefficients in (7) and schedule=linear, scaling=none to the EDM coefficients (11). We use the time discretizations disc=vp (linear) and disc=edm given in [16, Tab.1].

(We found this unconvincing)

Stiffness reduction with SEEDS. In Fig. 2, we illustrate the impact of different choices of discretization steps, noise schedule and dynamic scaling on SEEDS and stochastic EDM. We see that choosing the EDM discretization over the linear one has the effect of flattening the pixel trajectories at latest stages of the simulation procedure. Also, choosing the parameters (11) over those in (7) has the effect of greatly changing the distribution variances as the trajectories evolve. Notice that all the SEEDS trajectories seem perceptually more stable than those from EDM. It would be interesting to relate this to the *stiffness* of the semi-linear DE describing these trajectories, and to the magnitude of the parameters involved in the noise injection for EDM solver amplifying this phenomenon.