Contents

A Detailed Derivation of the SEEDS Design Space

In Sections 2 and 3 we proposed a simplified presentation of the design space of diffusion models and of the ingredients that constitute our proposed SEEDS methodology. In this section, we further develop our presentation in a technical manner, making explicit the formalization of our design choices.

A.1 The Isotropic General SDE framework

In Section 2, we presented a parametric family of differential equations (2) driving the generative process for DPMs, based on time-reversing the forward noising diffusion process (1). While doing so, we presented two parameters - the noise schedule σ_t and the scaling α_t - for which the effects on DPMs have been widely studied in [12].

As the shape of the trajectories of (1) and (2) (for $\ell = 0, 1$) are defined by α_t and σ_t, we start by writing down, for the scaling $\mathbf{x}_t = \alpha_t \widehat{\mathbf{x}}_t$, the scaled generalization of the proposed SDEs in [12, Eq. 103] which unifies in a single framework the forward and reverse trajectories:

$$\mathrm{d}\mathbf{x}_t^{\pm} = \left[\frac{\dot\alpha_t}{\alpha_t}\mathbf{x}_t^{\pm} - \alpha_t^2\dot\sigma_t\sigma_t\nabla_{\mathbf{x}_t^{\pm}}\log p(\widehat{\mathbf{x}}_t^{\pm};\sigma_t) \pm \alpha_t^2\dot\sigma_t\sigma_t\nabla_{\mathbf{x}_t^{\pm}}\log p(\widehat{\mathbf{x}}_t^{\pm};\sigma_t)\right]\mathrm{d}t + \alpha_t\sigma_t\sqrt{2\frac{\dot\sigma_t}{\sigma_t}}\mathrm{d}\boldsymbol{\omega}_t^{\pm}. \tag{17}$$

Following [12], the previous VP, VE, iDDPM, DDIM and EDM frameworks all are unified as different choices of α_t, σ_t, among other choices presented in [12, Tab. 1] and we will use this as a basis for all the proofs contained in this Appendix. In particular, forward time means taking $\mathbf{x}_t^+$ for which the score vanishes in this context. Now set $\mathbf{x}_t = \mathbf{x}_t^-$.

The formulation in (2) involves a family of backwards differential equations controlled by a parameter $\ell \in [0, 1]$ which all yield reverse-time processes for (1), a fact that can be obtained by studying the Fokker-Planck equation for marginals $p(\widehat{\mathbf{x}}_t^+;\sigma_t)$ of (17).

When $\ell = 1$, (2) the obtained SDE is known as the reverse SDE (RSDE) and, when $\ell = 0$, we obtain an ODE that is known as the Probability Flow ODE (PFO):

$$\mathrm{d}\mathbf{x}_t = \left[\frac{\dot\alpha_t}{\alpha_t}\mathbf{x}_t - \alpha_t^2\dot\sigma_t\sigma_t\nabla_{\mathbf{x}_t}\log p(\widehat{\mathbf{x}}_t;\sigma_t)\right]\mathrm{d}t. \tag{18}$$

Now, finding a minimum for the loss function in [12, Eq. 51] is formulated as a convex optimization problem. As such, for the ideal model $D(\mathbf{x}_t;\sigma_t) = \arg\min_D \mathcal{L}(D;\mathbf{x}_t,\sigma_t)$, the score function with scaled input is expressed as

$$\nabla_{\mathbf{x}_t} \log p(\widehat{\mathbf{x}}_t;\sigma_t) \quad = \quad \frac{D(\widehat{\mathbf{x}}_t;\sigma_t) - \widehat{\mathbf{x}}_t}{\alpha_t \sigma_t^2}.$$

This ideal model is usually subtracted by a *raw* network F in the form of a time-dependent preconditioning:

$$D(\mathbf{x}_t;\sigma_t) := c_1(t)\mathbf{x}_t + c_2(t)F(c_3(t)\mathbf{x}_t; c_4(t)), \qquad c_i(t) \in \mathbb{R}^d, \quad i = 1,\dots,4.$$

As such, we can express the score function as two parameterizations involving D or F as follows:

$$\nabla_{\mathbf{x}_t} \log p(\widehat{\mathbf{x}}_t;\sigma_t) \quad = \quad \frac{D(\widehat{\mathbf{x}}_t;\sigma_t) - \widehat{\mathbf{x}}_t}{\alpha_t \sigma_t^2} = \frac{(c_1(t)-1)\widehat{\mathbf{x}}_t + c_2(t)F(c_3(t)\widehat{\mathbf{x}}_t; c_4(t))}{\alpha_t \sigma_t^2}. \tag{19}$$

Let us now denote $D^1_{\theta,t} := D_\theta(\widehat{\mathbf{x}}_t;\sigma_t)$ for a pre-trained network approximating the ideal denoiser and let $D^2_{\theta,t} := F_\theta(c_3(t)\widehat{\mathbf{x}}_t; c_4(t))$ be the corresponding raw pre-trained network. Substituting the score function in the RSDE and the PFO with each of these models yields four *different* differential equations with a neural network as one of their components. These are given, for $i = 1, 2$, by

$$\mathrm{d}\mathbf{x}_t \quad = \quad [A^i(t)\mathbf{x}_t + B^i(t)D^i_{\theta,t}]\mathrm{d}t + g(t)\mathrm{d}\bar{\boldsymbol{\omega}}_t, \tag{20}$$

$$\mathrm{d}\mathbf{x}_t \quad = \quad [A^{i+2}(t)\mathbf{x}_t + B^{i+2}(t)D^i_{\theta,t}]\mathrm{d}t. \tag{21}$$

When $D^1_{\theta,t}$ (resp. $D^2_{\theta,t}$) is employed to replace the score function in (17) using (19), the resulting SDE (20) will be called *data (resp. noise) prediction neural SDE*. Proceeding analogously for the PFO (18) yield two ODEs (21) which will be called *data (resp. noise) prediction neural PFO*. The general form of the A^i and B^i coefficients determining each of these DEs is as follows:

$$A^1(t) = \frac{\dot\alpha_t}{\alpha_t} + 2\frac{\dot\sigma_t}{\sigma_t} \qquad B^1(t) = -2\alpha_t\frac{\dot\sigma_t}{\sigma_t} \qquad \text{(DP NRSDE)}$$

$$A^2(t) = \frac{\dot\alpha_t}{\alpha_t} + 2\frac{\dot\sigma_t}{\sigma_t}(1-c_1(t)) \qquad B^2(t) = -2\alpha_t\frac{\dot\sigma_t}{\sigma_t}c_2(t) \qquad \text{(NP NRSDE)}$$

$$A^3(t) = \frac{\dot\alpha_t}{\alpha_t} + \frac{\dot\sigma_t}{\sigma_t} \qquad B^3(t) = -\alpha_t\frac{\dot\sigma_t}{\sigma_t} \qquad \text{(DP NPFO)}$$

$$A^4(t) = \frac{\dot\alpha_t}{\alpha_t} + \frac{\dot\sigma_t}{\sigma_t}(1-c_1(t)) \qquad B^4(t) = -\alpha_t\frac{\dot\sigma_t}{\sigma_t}c_2(t) \qquad \text{(NP NPFO)}$$

Remark A.1. At first glance, it would seem misleading to differentiate four DEs as these essentially correspond to different choices of α_t, σ_t, $c_1(t),\dots,c_4(t)$. But the reason why we do so is that, after applying the *variation of constants* formula, each of these DEs will yield a different representation of their exact solutions (see (22) and (23) below). As we will see below, constructing exponential integrators heavily depends on such representation and will show to lead to four different modes of SEEDS solvers, each one showing different behavior and performance for sampling from pre-trained DPMs. As such, we will articulate this difference already at the DE formulation.

For $t < s$, the variation of constants formulae for NSDEs allows to represent the exact solutions of (20) as

$$\mathbf{x}_t = \Phi^i(t,s)\mathbf{x}_s + \int_s^t \Phi^i(t,\tau)B^i(\tau)D^i_{\theta,\tau}\mathrm{d}\tau + \int_s^t \Phi^i(t,\tau)g(\tau)\mathrm{d}\bar{\boldsymbol{\omega}}_t, \qquad i = 1,2 \tag{22}$$

and those for NPFOs (21) as

$$\mathbf{x}_t = \Phi^i(t,s)\mathbf{x}_s + \int_s^t \Phi^i(t,\tau)B^i(\tau)D^{i-2}_{\theta,\tau}\mathrm{d}\tau, \qquad i = 3,4 \tag{23}$$

where

$$\Phi_{A^i}(t,s) = \exp\left(\int_s^t A^i(\tau)\mathrm{d}\tau\right) \tag{24}$$

is called the transition matrix associated with $A^i(t)$ and is defined as the solution to

$$\frac{\partial}{\partial t}\Phi_{A^i}(t,s) = A^i(t)\Phi_{A^i}(t,s), \qquad \Phi_{A^i}(s,s) = \boldsymbol{I}_d.$$

When $A^i(t)$ is constant and $B^i(t) = 1$, there is a well-established literature on exponential ODE and SDE solvers with explicit *stiff order conditions* and prescribed by different forms of Butcher tableaux. When $A^i(t)$ is not constant, for the expression in (24) to make sense in the usual sense (in terms of exponential series expansion) instead of having to make use of time-ordered exponentials/ Magnus expansions, the $f(t) := A^i(t)$ coefficients must satisfy $[f^{(k)}(t), f^{(l)}(s)] = 0$. This condition is trivially satisfied here as the $A^i(t)$ considered here are d-dimensional diagonal matrices.

Notice that, if $A^i \neq A^j$ for some $i \neq j$, their associated transition matrices will not be equal. In particular, if $A^1 \neq A^2$, then the variances of the stochastic integrals in (22) are different for $i = 1$ and $i = 2$. This is the first step in explaining the statement in Rem. A.1, and we refer the reader to the proof of Proposition 4.4 where we put into evidence its validity.

A.2 Re-framing and Generalizing Previous Exponential Solvers

A.2.1 The VP case

Let $\tilde{\alpha}_t := \int_0^t (\beta_d \tau - \beta_m)\mathrm{d}\tau = \frac{1}{2}\beta_d t^2 + \beta_m t$, where $\beta_d > \beta_m > 0$. Set

$$f(t) := \frac{\mathrm{d}\log\alpha_t}{\mathrm{d}t}, \quad g(t) = \alpha_t\sqrt{\frac{\mathrm{d}[\sigma_t^2]}{\mathrm{d}t}}, \qquad \sigma_t = \sqrt{e^{\tilde{\alpha}_t} - 1}, \qquad \alpha_t = e^{-\frac{1}{2}\tilde{\alpha}_t} = \frac{1}{\sqrt{\sigma_t^2 + 1}}.$$

Recall that in the VP case, and the *noise prediction mode*, [15] construct exponential solvers on the base of the following ODE

$$\mathrm{d}\mathbf{x}_t \quad = \quad \left[f(t)\mathbf{x}_t + \frac{g^2(t)}{2\bar{\sigma}_t}\epsilon_\theta(\mathbf{x}_t; t)\right]\mathrm{d}t, \quad t \in [T, 0], \tag{25}$$

where $\bar{\sigma}_t := \alpha_t\sigma_t$. The ODE (25) identifies with that in [12] for the VP case for which the authors identify the preconditioning

$$c_1(t) = 1, \quad c_2(t) = -\sigma_t, \quad c_3(t) = \frac{1}{\sqrt{\sigma_t^2 + 1}}, \quad c_4(t) = (M-1)\sigma^{-1}(\sigma_t) = (M-1)t.$$

As such, we obtain the following coefficients for the NP NPFO:

$$A^4(t) = \frac{\dot{\alpha}_t}{\alpha_t}, \quad B^4(t) = \alpha_t\dot{\sigma}_t, \quad \Phi^4(t,s) = \frac{\alpha_t}{\alpha_s}$$

and

$$\nabla_{\mathbf{x}_t}\log p(\widehat{\mathbf{x}}_t;\sigma_t) = \frac{D_\theta(\widehat{\mathbf{x}}_t;\sigma_t) - \widehat{\mathbf{x}}_t}{\alpha_t\sigma_t^2} \quad = \quad \frac{(c_1(t)-1)\widehat{\mathbf{x}}_t + c_2(t)F(c_3(t)\widehat{\mathbf{x}}_t; c_4(t))}{\alpha_t\sigma_t^2}$$
$$= \quad \frac{-\sigma_t F_\theta(\mathbf{x}_t, (M-1)t)}{\alpha_t\sigma_t^2}.$$

A.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 [15]. As such, $\mathrm{SNR}(t)$ is strictly monotonically decreasing in time. Thus, the analytic solution to

(2) yields

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

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.

A.2.3 Proof of Proposition 4.2

Recall that the functions φ_k are the integrals

$$\varphi_{k+1}(t) = \int_0^1 e^{(1-\delta)t}\frac{\delta^k}{k!}\mathrm{d}\delta,$$

which satisfy $\varphi_k(0) = \frac{1}{k!}$. The truncated Itô-Taylor expansion of $\widehat{\boldsymbol{\epsilon}}_\theta$ with respect to λ reads

$$\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda,\lambda) = \sum_{k=0}^n \frac{(\lambda-\lambda_s)^k}{k!}\widehat{\boldsymbol{\epsilon}}_\theta^{(k)}(\widehat{\mathbf{x}}_{\lambda_s},\lambda_s) + \mathcal{R}_{n+1}.$$

where here $\widehat{\boldsymbol{\epsilon}}_\theta^{(k)}$ denotes the L_t^k operators defined in (58) applied to $\widehat{\boldsymbol{\epsilon}}_\theta$. On the one hand, since $\int_{\lambda_s}^{\lambda_t} e^{-\lambda}d\lambda = \frac{\bar{\sigma}_t}{\alpha_t}(e^h-1)$, we obtain by iteratively integrating by parts

$$\begin{aligned}
\int_{\lambda_s}^{\lambda_t} e^{-\lambda}\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda,\lambda)\mathrm{d}\lambda &= \sum_{k=0}^n \widehat{\boldsymbol{\epsilon}}_\theta^{(k)}(\widehat{\mathbf{x}}_{\lambda_s},\lambda_s)\int_{\lambda_s}^{\lambda_t} e^{-\lambda}\frac{(\lambda-\lambda_s)^k}{k!}d\lambda + \mathcal{R}_{n+2}\\
&= \frac{\bar{\sigma}_t}{\alpha_t}\sum_{k=0}^n \widehat{\boldsymbol{\epsilon}}_\theta^{(k)}(\widehat{\mathbf{x}}_{\lambda_s},\lambda_s)h^{k+1}\varphi_{k+1}(h) + \mathcal{R}_{n+2}.
\end{aligned}$$

On the other hand, we have $s > t, h = \lambda_t - \lambda_s > 0$. Note that since the stochastic integrals $\int_{\lambda_s}^{\lambda_t} e^{-\lambda}\mathrm{d}\bar{\boldsymbol{\omega}}(\lambda)$ are measurable with respect to $(\bar{\boldsymbol{\omega}}(\lambda)-\bar{\boldsymbol{\omega}}(\lambda_s), 0 \le \lambda \le \lambda_t-\lambda_s)$, they are independent on disjoint time intervals by the independence of increments property of Brownian motion. Thus the random variable $\epsilon := \epsilon_{s,t}$ in our algorithms are independent on disjoint time intervals. We then write

$$\begin{aligned}
\int_{\lambda_s}^{\lambda_t} e^{-\lambda}\mathrm{d}\bar{\boldsymbol{\omega}}(\lambda) &= \mathcal{N}\left(0, \int_{\lambda_s}^{\lambda_t} e^{-2\lambda}\mathrm{d}\lambda\right)\\
&= \frac{1}{\sqrt{2}}\sqrt{e^{-2\lambda_s}-e^{-2\lambda_t}}\epsilon, \quad \epsilon\sim\mathcal{N}(\mathbf{0},\mathbf{I}_d)\\
&= \frac{1}{\sqrt{2}}\sqrt{\left(\frac{\bar{\sigma}_s}{\alpha_s}-\frac{\bar{\sigma}_t}{\alpha_t}\right)\left(\frac{\bar{\sigma}_s}{\alpha_s}+\frac{\bar{\sigma}_t}{\alpha_t}\right)}\epsilon, \quad \epsilon\sim\mathcal{N}(\mathbf{0},\mathbf{I}_d)\\
&= \frac{1}{\sqrt{2}}\frac{\bar{\sigma}_t}{\alpha_t}\sqrt{e^{2h}-1}\epsilon, \quad \epsilon\sim\mathcal{N}(\mathbf{0},\mathbf{I}_d).
\end{aligned}$$

In conclusion, the truncated Itô-Taylor expansion of the analytic expression

$$\mathbf{x}_t = \frac{\alpha_t}{\alpha_s}\mathbf{x}_s - 2\alpha_t \int_{\lambda_s}^{\lambda_t} e^{-\lambda}\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda, \lambda)\mathrm{d}\lambda - \sqrt{2}\alpha_t \int_{\lambda_s}^{\lambda_t} e^{-\lambda}\mathrm{d}\bar{\boldsymbol{\omega}}(\lambda) \tag{27}$$

simplifies to

$$\mathbf{x}_t = \frac{\alpha_t}{\alpha_s}\mathbf{x}_s - 2\bar{\sigma}_t \sum_{k=0}^{n} h^{k+1}\varphi_{k+1}(h)\widehat{\boldsymbol{\epsilon}}_\theta^{(k)}(\widehat{\mathbf{x}}_{\lambda_s}, \lambda_s) - \bar{\sigma}_t\sqrt{e^{2h}-1}\epsilon + \mathcal{R}_{n+2}, \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d).$$

This finishes the proof.

A.2.4 Generalization to the remaining data prediction and deterministic modes

Propositions 3.1 and 4.2 consist on the first steps for crafting SEEDS solvers in the VP case associated to the noise prediction neural RSDE (NP NRSDE). Generalizing the above procedure for crafting SEEDS for the 4 modes associated to (22) and (23) yields the following sets of coefficients

$$A^1(t) = \frac{\dot{\alpha}_t}{\alpha_t} + 2\frac{\dot{\sigma}_t}{\sigma_t} \qquad B^1(t) = -2\alpha_t\frac{\dot{\sigma}_t}{\sigma_t} \qquad \text{(DP NRSDE)}$$

$$A^2(t) = \frac{\dot{\alpha}_t}{\alpha_t} \qquad B^2(t) = 2\alpha_t\dot{\sigma}_t \qquad \text{(NP NRSDE)}$$

$$A^3(t) = \frac{\dot{\alpha}_t}{\alpha_t} + \frac{\dot{\sigma}_t}{\sigma_t} \qquad B^3(t) = -\alpha_t\frac{\dot{\sigma}_t}{\sigma_t} \qquad \text{(DP NPFO)}$$

$$A^4(t) = \frac{\dot{\alpha}_t}{\alpha_t} \qquad B^4(t) = \alpha_t\dot{\sigma}_t \qquad \text{(NP NPFO)}$$

We readily obtain

$$\Phi^2(t,s) = \Phi^4(t,s) = \frac{\alpha_t}{\alpha_s}, \qquad \Phi^3(t,s) = \frac{\bar{\sigma}_t}{\bar{\sigma}_s}, \qquad \Phi^1(t,s) = \frac{\sigma_t^2\alpha_t}{\sigma_s^2\alpha_s}.$$

Then, by setting the simpler change of variables $\lambda_t := -\log(\sigma_t)$, we obtain

$$\int_s^t \Phi^4(t,\tau)B^4(\tau)\mathrm{d}\tau = \alpha_t\int_s^t \frac{1}{\alpha_\tau}\alpha_\tau\dot{\sigma}_\tau\mathrm{d}\tau = \alpha_t\int_s^t \dot{\sigma}_\tau\mathrm{d}\tau = -\alpha_t\int_{\lambda_s}^{\lambda_t} e^{-\lambda}\mathrm{d}\lambda = -\bar{\sigma}_t(e^h-1).$$

By recursion we obtain

$$\int_s^t \Phi^4(t,\tau)B^4(\tau)F_{\theta,\tau}\mathrm{d}\tau = -\bar{\sigma}_t\sum_{k=0}^{n-1} h^{k+1}\varphi_{k+1}(h)F_{\theta,s}^{(k)} + \mathcal{O}(h^{n+1}).$$

In the same way, we obtain

$$\begin{aligned}\int_s^t \Phi^3(t,\tau)B^3(\tau)\mathrm{d}\tau = \sigma_t\alpha_t\int_s^t \frac{-1}{\alpha_\tau\sigma_\tau}\alpha_\tau\frac{\dot{\sigma}_\tau}{\sigma_\tau}\mathrm{d}\tau &= \sigma_t\alpha_t\int_s^t \frac{-\dot{\sigma}_\tau}{\sigma_\tau^2}\mathrm{d}\tau \\ &= \sigma_t\alpha_t\int_{\lambda_s}^{\lambda_t} e^{\lambda}\mathrm{d}\lambda = -\alpha_t(e^{-h}-1).\end{aligned}$$

Next,

$$\int_s^t \Phi^2(t,\tau)B^2(\tau)\mathrm{d}\tau = \alpha_t\int_s^t \frac{2}{\alpha_\tau}\alpha_\tau\dot{\sigma}_\tau\mathrm{d}\tau = -2\bar{\sigma}_t(e^h-1),$$

and finally, as already shown in Propositions 3.1 and 4.2:

$$\begin{aligned}\int_s^t \Phi^1(t,\tau)B^1(\tau)\mathrm{d}\tau = \sigma_t^2\alpha_t\int_s^t \frac{-2\dot{\sigma}_\tau}{\sigma_\tau^3}\mathrm{d}\tau &= \sigma_t^2\alpha_t\int_s^t -2\frac{\mathrm{d}\lambda_\tau}{\mathrm{d}\tau}e^{2\lambda_\tau}\mathrm{d}\tau \\ &= \sigma_t^2\alpha_t\int_{\lambda_s}^{\lambda_t} e^{2\lambda}\mathrm{d}\lambda = -\alpha_t(e^{-2h}-1).\end{aligned}$$

Now, since $g^2(t) = 2\alpha_t^2 \dot{\sigma}_t \sigma_t$, the stochastic integrals $\int_s^t \Phi^i(t,\tau) g(\tau) \mathrm{d}\bar{\boldsymbol{\omega}}_\tau$, for $i = 1, 2$, have zero mean and variances:

$$\begin{aligned}\int_s^t (\Phi^1(t,\tau))^2 g^2(\tau)\mathrm{d}\tau &= \sigma_t^4 \alpha_t^2 \int_s^t \frac{1}{\sigma_\tau^4 \alpha_\tau^2} g^2(\tau)\mathrm{d}\tau = \bar{\sigma}_t^2(1 - e^{-2h}) \\ \int_s^t (\Phi^2(t,\tau))^2 g^2(\tau)\mathrm{d}\tau &= \alpha_t^2 \int_s^t 2\dot{\sigma}_\tau \sigma_\tau \mathrm{d}\tau = -\bar{\sigma}_t^2(e^{2h} - 1).\end{aligned}$$

We deduce from this the SEEDS-1 schemes in all four modes as given by iterates:

$$\widetilde{\mathbf{x}}_t = \frac{\sigma_t^2 \alpha_t}{\sigma_s^2 \alpha_s} \widetilde{\mathbf{x}}_s - \alpha_t(e^{-2h} - 1) D_\theta(\widetilde{\mathbf{x}}_s, s) + \bar{\sigma}_t \sqrt{1 - e^{-2h}} \epsilon \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d) \tag{28}$$

$$\widetilde{\mathbf{x}}_t = \frac{\alpha_t}{\alpha_s} \widetilde{\mathbf{x}}_s - 2\bar{\sigma}_t(e^h - 1) \boldsymbol{\epsilon}_\theta(\widetilde{\mathbf{x}}_s, s) - \bar{\sigma}_t \sqrt{e^{2h} - 1} \epsilon \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d) \tag{29}$$

$$\widetilde{\mathbf{x}}_t = \frac{\bar{\sigma}_t}{\bar{\sigma}_s} \widetilde{\mathbf{x}}_s - \alpha_t(e^{-h} - 1) D_\theta(\widetilde{\mathbf{x}}_s, s) \tag{30}$$

$$\widetilde{\mathbf{x}}_t = \frac{\alpha_t}{\alpha_s} \widetilde{\mathbf{x}}_s - \bar{\sigma}_t(e^h - 1) \boldsymbol{\epsilon}_\theta(\widetilde{\mathbf{x}}_s, s). \tag{31}$$

Notice that the iterates (30) and (31) are exactly the iterates of the first stage solvers in [15] and [16] with $F_\theta(\mathbf{x}_t; (M-1)t) = \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)$. The iterates (29) coincide with the SEEDS-1 method presented in (14) and (28) consist on our SEEDS-1 method in the *data prediction mode*, which we will use in the following section.

A.2.5 Proof of Proposition 4.4

We will write in **bold** the statement to be proven.

If we set $g = 0$ in (15), the resulting SEEDS solvers do not yield DPM-Solver. Indeed, if we set $g = 0$ in (15), then the method (29) does not contain a noise contribution and we readily see that it cannot be equal to (31). As the latter has been shown to be DPM-Solver-1, the conclusion follows.

If we parameterize (15) in terms of the data prediction model D_θ, the resulting SEEDS solvers are not equivalent to their noise prediction counterparts defined in Alg. 1 to 4.

One can check that the SEEDS solver are not the same between the noise and data prediction modes by simply noticing that the noise contributions in (28) and (29) do not equate.

The gDDIM solver [27] Theorem 1, for $\ell = 1$, is equal to SEEDS-1 in the data prediction mode.

As shown in (29), our proposed method SEEDS-1 in the data prediction mode for the VP case has iterates of the form

$$\widetilde{\mathbf{x}}_t = \frac{\sigma_t^2 \alpha_t}{\sigma_s^2 \alpha_s} \widetilde{\mathbf{x}}_s - \alpha_t(e^{-2h} - 1) D_\theta(\widetilde{\mathbf{x}}_s, s) + \bar{\sigma}_t \sqrt{1 - e^{-2h}} \epsilon, \tag{32}$$

where $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$, $\bar{\sigma}_t = \alpha_t \sigma_t$ and $h = \log \dfrac{\sigma_s}{\sigma_t}$. As our notation and that of [27, Theorem 1] overlap, we will use blue color when referring to their notation.

On the one hand, gDDIM constructs iterates over a representation of the exact solution of the following family of neural differential equations:

$$\mathrm{d}\mathbf{u}_t = \left[f(t)\mathbf{u}_t + \frac{1 + \lambda^2}{2} \frac{g^2(t)}{\sqrt{1 - \alpha_t}} \epsilon_\theta(\mathbf{u}_t; t) \right] \mathrm{d}t + \lambda g(t) \mathrm{d}\bar{\boldsymbol{\omega}}_t \tag{33}$$

where α_t decreases from $\alpha_0 = 1$ to $\alpha_T = 0$, and coefficients

$$f(t) := \frac{1}{2} \frac{\mathrm{d} \log \alpha_t}{\mathrm{d}t}, \quad g(t) = \sqrt{-\frac{\mathrm{d} \log \alpha_t}{\mathrm{d}t}}$$

In particular, they choose an approximation, for $\tau \in [t - \Delta t, t]$, given by

$$\boldsymbol{s}_\theta(\mathbf{u}, \tau) = \frac{\epsilon_\theta(\mathbf{u}_\tau, \tau)}{\sqrt{1 - \alpha_t}} \approx \frac{1 - \alpha_t}{1 - \alpha_\tau} \sqrt{\frac{\alpha_\tau}{\alpha_t}} \boldsymbol{s}_\theta(\mathbf{u}(t), t) - \frac{1}{1 - \alpha_\tau} (\mathbf{u} - \sqrt{\frac{\alpha_\tau}{\alpha_t}} \mathbf{u}(t)).$$

The gDDIM iterates, for $\lambda = 1 = \ell$, are then written as follows:

$$\mathbf{u}(t-\Delta t) = \sqrt{\frac{\alpha_{t-\Delta t}}{\alpha_t}}\mathbf{u}(t) + \left[-\sqrt{\frac{\alpha_{t-\Delta t}}{\alpha_t}}\sqrt{1-\alpha_t} + \sqrt{1-\alpha_{t-\Delta t}-\sigma_t^2}\right]\epsilon_\theta(\mathbf{u}(t),t) + \sigma_t\epsilon \tag{34}$$

where $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ and where

$$\sigma_t^2 = (1-\alpha_{t-\Delta t})\left[1 - \left(\frac{1-\alpha_{t-\Delta t}}{1-\alpha_t}\right)\left(\frac{\alpha_t}{\alpha_{t-\Delta t}}\right)\right]. \tag{35}$$

Now set $(s,t) \leftarrow (t, t-\Delta t)$. Then

$$\begin{aligned}
\sigma_s^2 &= (1-\alpha_t)\left[1 - \left(\frac{1-\alpha_t}{1-\alpha_s}\right)\left(\frac{\alpha_s}{\alpha_t}\right)\right], \\
\mathbf{u}(t) &= \sqrt{\frac{\alpha_t}{\alpha_s}}\mathbf{u}(s) + \left[\sqrt{1-\alpha_t-\sigma_s^2} - \sqrt{\frac{\alpha_t}{\alpha_s}}\sqrt{1-\alpha_s}\right]\epsilon_\theta(\mathbf{u}(s),s) + \sigma_s\epsilon.
\end{aligned}$$

Now set $\alpha_t = \sqrt{\alpha_t}$, and $\bar{\sigma}_t = \sqrt{1-\alpha_t}$. Then the variance of the noise in (35) is

$$\begin{aligned}
\sigma_s^2 &= \bar{\sigma}_t^2\left[1 - \left(\frac{\bar{\sigma}_t}{\bar{\sigma}_s}\right)^2\left(\frac{\alpha_s}{\alpha_t}\right)^2\right] = \bar{\sigma}_t^2\left[1 - \left(\frac{\alpha_t\sigma_t}{\alpha_s\sigma_s}\right)^2\left(\frac{\alpha_s}{\alpha_t}\right)^2\right] \\
&= \bar{\sigma}_t^2\left[1 - \left(\frac{\sigma_t}{\sigma_s}\right)^2\right] \\
&= \bar{\sigma}_t^2(1-e^{-2h}).
\end{aligned}$$

Hence, by denoting $\widetilde{\mathbf{x}}_t = \mathbf{u}(t)$, the gDDIM iterate (34) reads, for $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$:

$$\begin{aligned}
\widetilde{\mathbf{x}}_t &= \frac{\alpha_t}{\alpha_s}\widetilde{\mathbf{x}}_s + \left[\sqrt{\bar{\sigma}_t^2 - \bar{\sigma}_t^2(1-e^{-2h})} - \frac{\alpha_t}{\alpha_s}\bar{\sigma}_s\right]\epsilon_\theta(\widetilde{\mathbf{x}}_s, s) + \sqrt{\bar{\sigma}_t^2(1-e^{-2h})}\epsilon \\
&= \frac{\alpha_t}{\alpha_s}\widetilde{\mathbf{x}}_s + \left[\bar{\sigma}_t\sqrt{e^{-2h}} - \alpha_t\sigma_s\right]\epsilon_\theta(\widetilde{\mathbf{x}}_s, s) + \bar{\sigma}_t\sqrt{1-e^{-2h}}\epsilon \\
&= \frac{\alpha_t}{\alpha_s}\widetilde{\mathbf{x}}_s + \bar{\sigma}_t\left[\frac{\sigma_t}{\sigma_s} - \frac{\sigma_s}{\sigma_t}\right]\epsilon_\theta(\widetilde{\mathbf{x}}_s, s) + \bar{\sigma}_t\sqrt{1-e^{-2h}}\epsilon.
\end{aligned} \tag{36}$$

On the other hand, the data and noise prediction models in our case are related by the following equation:

$$D_\theta(\widetilde{\mathbf{x}}_s, s) = c_1(s)\frac{\widetilde{\mathbf{x}}_s}{\alpha_s} + c_2(s)\boldsymbol{\epsilon}_\theta(\widetilde{\mathbf{x}}_s, s) = \frac{\widetilde{\mathbf{x}}_s}{\alpha_s} - \sigma_s\boldsymbol{\epsilon}_\theta(\widetilde{\mathbf{x}}_s, s).$$

As such, and as $h = \lambda_t - \lambda_s = \log\frac{\sigma_s}{\sigma_t}$, one can rewrite (32) in terms of $\boldsymbol{\epsilon}_\theta$ as follows:

$$\begin{aligned}
\widetilde{\mathbf{x}}_t &= \frac{\sigma_t^2\alpha_t}{\sigma_s^2\alpha_s}\widetilde{\mathbf{x}}_s - \alpha_t(e^{-2h}-1)D_\theta(\widetilde{\mathbf{x}}_s, s) + \bar{\sigma}_t\sqrt{1-e^{-2h}}\epsilon \\
&= \frac{\sigma_t^2\alpha_t}{\sigma_s^2\alpha_s}\widetilde{\mathbf{x}}_s - \alpha_t\left(\frac{\sigma_t^2}{\sigma_s^2}-1\right)\left[\frac{\widetilde{\mathbf{x}}_s}{\alpha_s} - \sigma_s\boldsymbol{\epsilon}_\theta(\widetilde{\mathbf{x}}_s, s)\right] + \bar{\sigma}_t\sqrt{1-e^{-2h}}\epsilon \\
&= \left[\frac{\sigma_t^2\alpha_t}{\sigma_s^2\alpha_s} - \frac{\alpha_t}{\alpha_s}\left(\frac{\sigma_t^2}{\sigma_s^2}-1\right)\right]\widetilde{\mathbf{x}}_s + \alpha_t\left(\frac{\sigma_t^2}{\sigma_s^2}-1\right)\sigma_s\boldsymbol{\epsilon}_\theta(\widetilde{\mathbf{x}}_s, s) + \bar{\sigma}_t\sqrt{1-e^{-2h}}\epsilon \\
&= \frac{\alpha_t}{\alpha_s}\widetilde{\mathbf{x}}_s + \bar{\sigma}_t\left(\frac{\sigma_t^2}{\sigma_s^2}-1\right)\frac{\sigma_s}{\sigma_t}\boldsymbol{\epsilon}_\theta(\widetilde{\mathbf{x}}_s, s) + \bar{\sigma}_t\sqrt{1-e^{-2h}}\epsilon \\
&= \frac{\alpha_t}{\alpha_s}\widetilde{\mathbf{x}}_s + \bar{\sigma}_t\left(\frac{\sigma_t}{\sigma_s} - \frac{\sigma_s}{\sigma_t}\right)\boldsymbol{\epsilon}_\theta(\widetilde{\mathbf{x}}_s, s) + \bar{\sigma}_t\sqrt{1-e^{-2h}}\epsilon,
\end{aligned}$$

which coincides with the gDDIM iterate in Equation (36).

A.2.6 The VE, DDIM and iDDPM cases

Following [12, Eq. 217], here $\alpha_t = 1$ and $c_1(t) = 1$ and so the only possibilities incurring into semi-linear differential equations are

$$A^1(t) = 2\frac{\dot\sigma_t}{\sigma_t} \qquad B^1(t) = -2\frac{\dot\sigma_t}{\sigma_t} \qquad \text{(DP SDE)}$$

$$A^3(t) = \frac{\dot\sigma_t}{\sigma_t} \qquad B^3(t) = -\frac{\dot\sigma_t}{\sigma_t}. \qquad \text{(DP PFO)}$$

We obtain, again with the choice $\lambda_t = -\log(\sigma_t)$ and setting $h = \lambda_t - \lambda_s$, the following:

$$\Phi^1(t,s) = \frac{\sigma_t^2}{\sigma_s^2}, \qquad \Phi^3(t,s) = \frac{\sigma_t}{\sigma_s},$$

$$\begin{aligned}\int_s^t \Phi^1(t,\tau)B^1(\tau)\mathrm{d}\tau &= -\sigma_t^2\int_{\lambda_s}^{\lambda_t} e^{2\lambda}\mathrm{d}\lambda = \frac{1}{2}(e^{-2h}-1)\\ \int_s^t \Phi^3(t,\tau)B^3(\tau)\mathrm{d}\tau &= -\sigma_t\int_{\lambda_s}^{\lambda_t} e^{\lambda}\mathrm{d}\lambda = e^{-h}-1.\end{aligned}$$

Next,

$$\int_s^t (\Phi^1(t,\tau))^2 g^2(\tau)\mathrm{d}\tau = \sigma_t^4\int_s^t \frac{1}{\sigma_\tau^4}\sigma_\tau^2 2\frac{\dot\sigma_\tau}{\sigma_\tau}\mathrm{d}\tau = \sigma_t^4\int_s^t \frac{2\dot\sigma_\tau}{\sigma_\tau^3}\mathrm{d}\tau = \sigma_t^2(e^{-2h}-1).$$

We readily see that these cases are identical to the VP case with $\alpha_t = 1$. In particular, the obtained SEEDS-1 iterates are

$$\begin{aligned}\widetilde{\mathbf{x}}_t &= \frac{\sigma_t^2}{\sigma_s^2}\widetilde{\mathbf{x}}_s - (e^{-2h}-1)D_\theta(\widetilde{\mathbf{x}}_s;\sigma_s) + \sigma_t\sqrt{1-e^{-2h}}\epsilon.\\ \widetilde{\mathbf{x}}_t &= \frac{\sigma_t}{\sigma_s}\widetilde{\mathbf{x}}_s - (e^{-h}-1)D_\theta(\widetilde{\mathbf{x}}_s;\sigma_s)\end{aligned}$$

Now denote $s_1 = t_\lambda(\lambda_s + rh)$, for $0 < r \leqslant 1$, where $t_\lambda = e^{-\lambda}$. There are two families of single-step one-parameter two-stage exponential ODE schemes:

$$\begin{aligned}\widetilde{\mathbf{x}}_t &= \frac{\sigma_t}{\sigma_s}\widetilde{\mathbf{x}}_s - (e^{-h}-1)\left[\left(1-\frac{1}{2r}\right)D_\theta(\widetilde{\mathbf{x}}_s;\sigma_s) + \frac{1}{2r}D_\theta(\widetilde{\mathbf{x}}_1;\sigma_{s_1})\right]\\ \widetilde{\mathbf{x}}_t &= \frac{\sigma_t}{\sigma_s}\widetilde{\mathbf{x}}_s - (e^{-h}-1)D_\theta(\widetilde{\mathbf{x}}_s;\sigma_s) + \frac{1}{r}\left(\frac{e^{-h}-1}{h}+1\right)[D_\theta(\widetilde{\mathbf{x}}_1;\sigma_{s_1}) - D_\theta(\widetilde{\mathbf{x}}_s;\sigma_s)],\end{aligned}$$

with same supporting value

$$\widetilde{\mathbf{x}}_1 = \frac{\sigma_{s_1}}{\sigma_s}\widetilde{\mathbf{x}}_s - (e^{-rh}-1)D_\theta(\widetilde{\mathbf{x}}_s;\sigma_s).$$

In the same vein we define a single-step two stage exponential SDE scheme:

$$\begin{aligned}\widetilde{\mathbf{x}}_1 &= \frac{\sigma_{s_1}^2}{\sigma_s^2}\widetilde{\mathbf{x}}_s - (e^{-2rh}-1)D_\theta(\widetilde{\mathbf{x}}_s;\sigma_s) + \sigma_{s_1}\sqrt{e^{-2rh}-1}\epsilon_1\\ \widetilde{\mathbf{x}}_t &= \frac{\sigma_t^2}{\sigma_s^2}\widetilde{\mathbf{x}}_s - (e^{-2h}-1)\left[\left(1-\frac{1}{2r}\right)D_\theta(\widetilde{\mathbf{x}}_s;\sigma_s) + \frac{1}{2r}D_\theta(\widetilde{\mathbf{x}}_1;\sigma_{s_1})\right]\\ &\quad +\sigma_t\left[\sqrt{e^{-2h}-e^{-2rh}}\epsilon_1 + \sqrt{e^{-2rh}-1}\epsilon_2\right].\end{aligned}$$

A.2.7 The EDM case

In the EDM-preconditioned case [12, App. B.6], we set $\sigma_t := t$ and $\alpha_t := 1$. We denote $\sigma_d := \sigma_{\mathrm{data}}$ the variance of the considered initial dataset and we set

$$c_1(t) = \frac{\sigma_d^2}{t^2+\sigma_d^2}, \quad c_2(t) = \frac{t\sigma_d}{\sqrt{t^2+\sigma_d^2}}, \quad c_3(t) = \frac{1}{\sqrt{t^2+\sigma_d^2}}, \quad c_4(t) = \frac{1}{4}\log(t),$$

so we obtain the following coefficients:

$$A^1(t) = \frac{2}{t} \qquad B^1(t) = -\frac{2}{t} \qquad \text{(DP NRSDE)}$$

$$A^2(t) = \frac{2}{t}\left(1 - \frac{\sigma_d^2}{t^2+\sigma_d^2}\right) \qquad B^2(t) = \frac{-2\sigma_d}{\sqrt{t^2+\sigma_d^2}} \qquad \text{(NP NRSDE)}$$

$$A^3(t) = \frac{1}{t} \qquad B^3(t) = -\frac{1}{t} \qquad \text{(DP NPFO)}$$

$$A^4(t) = \frac{1}{t}\left(1 - \frac{\sigma_d^2}{t^2+\sigma_d^2}\right) \qquad B^4(t) = \frac{-\sigma_d}{\sqrt{t^2+\sigma_d^2}} \qquad \text{(NP NPFO)}$$

In particular, the data prediction neural SDE/PFO are identical to those in the VE case with $\sigma_t = t$. So let us concentrate on the noise prediction regime, leading us to prove Proposition 3.2.

A.2.8 Proof of Proposition 3.2

In the Noise Prediction case, we have

$$\Phi^2(t,s) = \frac{t^2+\sigma_d^2}{s^2+\sigma_d^2}, \quad \Phi^4(t,s) = \sqrt{\frac{t^2+\sigma_d^2}{s^2+\sigma_d^2}},$$

so we readily compute:

$$\begin{aligned}
\int_s^t \Phi^2(t,\tau)B^2(\tau)\mathrm{d}\tau &= (t^2+\sigma_d^2)\int_s^t \frac{1}{\tau^2+\sigma_d^2}\cdot\frac{-2\sigma_d}{\sqrt{\tau^2+\sigma_d^2}}\mathrm{d}\tau, \\
\int_s^t \Phi^4(t,\tau)B^4(\tau)\mathrm{d}\tau &= \sqrt{t^2+\sigma_d^2}\int_s^t \frac{\sigma_d}{\sqrt{\tau^2+\sigma_d^2}}\cdot\frac{-1}{\sqrt{\tau^2+\sigma_d^2}}\mathrm{d}\tau.
\end{aligned}$$

Let us consider two different changes of variables:

$$\lambda_t := -\log\left(\arctan\left(\frac{t}{\sigma_d}\right)\right) \qquad \text{and} \qquad \lambda_t := -\log\left(\frac{t}{\sigma_d\sqrt{t^2+\sigma_d^2}}\right), \tag{37}$$

that will be used for the (NP NPFO) and (NP NRSDE), respectively. For the former case, we have

$$e^{-\lambda_t}\mathrm{d}\lambda_t = -\frac{\frac{1}{\sigma_d}\mathrm{d}t}{1+\frac{t^2}{\sigma_d^2}} = -\frac{\sigma_d \mathrm{d}t}{\sigma_d^2+t^2}.$$

Therefore, we can deduce that

$$\begin{aligned}
\int_s^t \Phi^4(t;\tau)B^4(\tau)\mathrm{d}\tau &= \int_s^t \sqrt{\frac{t^2+\sigma_d^2}{\tau^2+\sigma_d^2}}\cdot\frac{-\sigma_d}{\sqrt{\tau^2+\sigma_d^2}}d\tau \\
&= \sqrt{t^2+\sigma_d^2}\int_{\lambda_s}^{\lambda_t}\frac{-\sigma_d}{\tau^2+\sigma_d^2}\mathrm{d}\tau \\
&= \sqrt{t^2+\sigma_d^2}\int_{\lambda_s}^{\lambda_t} e^{-\lambda}\mathrm{d}\lambda \\
&= \sqrt{t^2+\sigma_d^2}\arctan\left(\frac{t}{\sigma_d}\right)(e^h-1).
\end{aligned}$$

For the latter case, we have

$$\begin{aligned}
e^{-\lambda_t}\mathrm{d}\lambda_t &= -\frac{\sigma_d\sqrt{t^2+\sigma_d^2} - t\sigma_d\frac{t}{\sqrt{t^2+\sigma_d^2}}}{\sigma_d^2(t^2+\sigma_d^2)} \\
&= -\frac{\sigma_d^2}{\sigma_d(t^2+\sigma_d^2)\sqrt{t^2+\sigma_d^2}} = -\frac{\sigma_d}{(t^2+\sigma_d^2)\sqrt{t^2+\sigma_d^2}}.
\end{aligned}$$

We then obtain

$$\begin{aligned}\int_s^t \Phi^2(t;\tau)B^2(\tau)\mathrm{d}\tau &= \int_s^t \frac{t^2+\sigma_d^2}{\tau^2+\sigma_d^2}\cdot\frac{-2\sigma_d}{\sqrt{\tau^2+\sigma_d^2}}\mathrm{d}\tau \\ &= 2(t^2+\sigma_d^2)\int_s^t e^{-\lambda}\mathrm{d}\lambda \\ &= \frac{2t\sqrt{t^2+\sigma_d^2}}{\sigma_d}(e^h-1).\end{aligned}$$

The stochastic integral $\int_s^t \Phi^2(t,\tau)g(\tau)\mathrm{d}\overline{\boldsymbol{\omega}}_\tau$ in noise prediction case is a Gaussian random variable with zero mean and whose variance can be computed by Itô isometry as

$$\begin{aligned}\int_t^s [\Phi^2(t;\tau)g^2(\tau)]^2\mathrm{d}\tau &= (t^2+\sigma_d^2)^2\int_t^s \frac{1}{[\tau^2+\sigma_d^2]^2}2\tau\mathrm{d}\tau \\ &= (t^2+\sigma_d^2)^2\int_t^s \frac{1}{[\tau^2+\sigma_d^2]^2}\mathrm{d}\tau^2 \\ &= (t^2+\sigma_d^2)^2\left(-\frac{1}{s^2+\sigma_d^2}+\frac{1}{t^2+\sigma_d^2}\right) \\ &= \frac{(t^2+\sigma_d^2)(s^2-t^2)}{(s^2+\sigma_d^2)}.\end{aligned}$$

Putting everything together, we obtain the analytic solution at time t of (2) with coefficients (11) and initial value $\mathbf{x}_s$ for the (NP NRSDE):

$$\mathbf{x}_t = \frac{t^2+\sigma_d^2}{s^2+\sigma_d^2}\mathbf{x}_s + 2(t^2+\sigma_d^2)\int_{\lambda_s}^{\lambda_t} e^{-\lambda}\hat{F}_\theta(\widehat{\mathbf{x}}_\lambda,\lambda)\mathrm{d}\lambda - \sqrt{2}(t^2+\sigma_d^2)\int_{\lambda_s}^{\lambda_t} e^{-\lambda}\mathrm{d}\overline{\boldsymbol{\omega}}_\lambda, \tag{38}$$

where $\lambda_t := -\log\left[\frac{t}{\sigma_d\sqrt{t^2+\sigma_d^2}}\right]$. For the (NP NPFO), it is given by

$$\mathbf{x}_t = \sqrt{\frac{t^2+\sigma_d^2}{s^2+\sigma_d^2}}\mathbf{x}_s + \sqrt{t^2+\sigma_d^2}\int_{\lambda_s}^{\lambda_t} e^{-\lambda}\hat{F}_\theta(\widehat{\mathbf{x}}_\lambda,\lambda)\mathrm{d}\lambda, \quad \lambda_t := -\log\left[\arctan\left[\frac{t}{\sigma_d}\right]\right]. \tag{39}$$

This finishes the proof.

Remark A.2. From the above proof, we immediately deduce the SEEDS-1 iterates in the EDM-preconditioned noise prediction case. These are given, for the (NP NPFO) and (NP NRSDE) respectively, by

$$\widetilde{\mathbf{x}}_t = \sqrt{\frac{t^2+\sigma_d^2}{s^2+\sigma_d^2}}\widetilde{\mathbf{x}}_s + \sqrt{t^2+\sigma_d^2}\arctan\left(\frac{t}{\sigma_d}\right)(e^h-1)\epsilon_\theta(\widetilde{\mathbf{x}}_s,s), \tag{40}$$

$$\widetilde{\mathbf{x}}_t = \frac{t^2+\sigma_d^2}{s^2+\sigma_d^2}\widetilde{\mathbf{x}}_s + \frac{2t\sqrt{t^2+\sigma_d^2}}{\sigma_d}(e^h-1)\epsilon_\theta(\widetilde{\mathbf{x}}_s,s) + \sqrt{\frac{(t^2+\sigma_d^2)(s^2-t^2)}{(s^2+\sigma_d^2)}}\epsilon, \tag{41}$$

where $\epsilon\sim\mathcal{N}(\mathbf{0},\mathbf{I}_d)$.

B Convergence Proofs

In this Section, we give detailed proofs of Theorem 4.1 and Corollary 4.3 stated in the main part of this paper. Let us start recalling its framework. We start by considering the NP NRSDE (15) with VP

coefficients:

$$\begin{aligned} \mathrm{d}\mathbf{x}_t &= \left[f(t)\mathbf{x}_t + \frac{g^2(t)}{\bar{\sigma}_t}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\right]\mathrm{d}t + g(t)\mathrm{d}\bar{\boldsymbol{\omega}}_t \\ &= \left[f(t)\mathbf{x}_t + \frac{2\alpha_t^2\dot{\sigma}_t\sigma_t}{\bar{\sigma}_t}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\right]\mathrm{d}t + g(t)\mathrm{d}\bar{\boldsymbol{\omega}}_t \\ &= \left[\frac{\mathrm{d}\log\alpha_t}{\mathrm{d}t}\mathbf{x}_t + 2\alpha_t\dot{\sigma}_t\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\right]\mathrm{d}t + \alpha_t\sqrt{\frac{\mathrm{d}[\sigma_t^2]}{\mathrm{d}t}}\mathrm{d}\bar{\boldsymbol{\omega}}_t \end{aligned} \tag{42}$$

Denote t_λ the inverse of $\lambda_t := -\log(\sigma_t)$ (which is a strictly decreasing function of t) and denote $\widehat{\mathbf{x}}_\lambda := \mathbf{x}(t_\lambda(\lambda)), \hat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda, \lambda) := \boldsymbol{\epsilon}_\theta(\mathbf{x}(t_\lambda(\lambda)), t_\lambda(\lambda))$. We consider a time discretization $\{t_i\}_{i=0}^{M+1}$ going backwards in time starting from $t_0 = T$ to $t_{M+1} = 0$ and to ease the notation we will always denote $t < s$ for two consecutive time-steps $t_i < t_{i-1}$. The analytic solution at time t of the RSDE (2) with coefficients (7) and initial value $\mathbf{x}_s$ reads

$$\mathbf{x}_t = \frac{\alpha_t}{\alpha_s}\mathbf{x}_s - 2\alpha_t\int_{\lambda_s}^{\lambda_t} e^{-\lambda}\hat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda, \lambda)\mathrm{d}\lambda - \sqrt{2}\alpha_t\int_{\lambda_s}^{\lambda_t} e^{-\lambda}\mathrm{d}\bar{\boldsymbol{\omega}}(\lambda) \tag{43}$$

Given an initial condition $\widetilde{\mathbf{x}}_{t_0} = \mathbf{x}_T$, the SEEDS-1 iterates read, for $h_i = \lambda_{t_i} - \lambda_{t_{i-1}} = \lambda_t - \lambda_s$,

$$\widetilde{\mathbf{x}}_t = \frac{\alpha_t}{\alpha_s}\widetilde{\mathbf{x}}_s - 2\bar{\sigma}_t(e^{h_i} - 1)\boldsymbol{\epsilon}_\theta(\widetilde{\mathbf{x}}_s, s) - \bar{\sigma}_t\sqrt{e^{2h_i} - 1}\epsilon \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d).$$

Assumption B.1.

1. The function $\boldsymbol{\epsilon}_\theta(\mathbf{x}, t)$ is continuous (and hence bounded) on $[0, T]$, Lipschitz with respect to $\mathbf{x}$ and there is a constant C such that, for $t, s \in [0, T]$ with $t < s$, we have

$$|\boldsymbol{\epsilon}_\theta(\mathbf{x}, t) - \boldsymbol{\epsilon}_\theta(\mathbf{y}, t)|^2 \leqslant L_1|\mathbf{x} - \mathbf{y}|^2 \tag{44}$$

$$|\boldsymbol{\epsilon}_\theta(\mathbf{x}, t)|^2 \vee |g(t)|^2 \leqslant L_2(1 + |\mathbf{x}|^2) \tag{45}$$

$$|\boldsymbol{\epsilon}_\theta(\mathbf{x}, t) - \boldsymbol{\epsilon}_\theta(\mathbf{x}, s)|^2 \leqslant L_3(1 + |\mathbf{x}|^2)|t - s|^2 \tag{46}$$

2. $h = \max_{1\leq i\leq M}|h_i| \sim O(1/M)$, where $h_i = \lambda_{t_i} - \lambda_{t_{i-1}}$.

Let $\mathcal{C}_P^l(\mathbb{R}^d, \mathbb{R})$ denote the family of L times continuously differentiable real-valued functions on $\mathbb{R}^d$ whose partial derivatives of order $\leqslant l$ have polynomial growth and let $\mathcal{C}_P^{k,l}(I \times \mathbb{R}^d, \mathbb{R})$ be the space of functions $g(\cdot, \cdot)$ such that, for all $(t, x) \in I \times \mathbb{R}^d$, $g(\cdot, x) \in \mathcal{C}^k(I, \mathbb{R})$ and $g(t, \cdot) \in \mathcal{C}_P^l(\mathbb{R}^d, \mathbb{R})$.

Assumption B.2. In addition to Assumption B.1, assume that all the components of $\boldsymbol{\epsilon}_\theta$ belong to $\mathcal{C}_P^{4,2}(\mathbb{R}^d \times [0, 1], \mathbb{R})$.

Before going into the proofs, we give some context that lead us to necessitate such assumptions.

B.1 Preliminaries

For an interval $I = [t_0, T]$, let $\mathbf{x} = (\mathbf{x}(t))_I$ the solution of the following SDE

$$\mathrm{d}\mathbf{x}(t) = f(\mathbf{x}(t), t)\mathrm{d}t + g(t)\mathrm{d}\boldsymbol{\omega}(t), \tag{47}$$

where $g(t) = \hat{g}(t) \cdot \mathrm{Id}_d$ is considered here as a diagonal matrix with identical diagonal entries $\hat{g}(t)$. Suppose that f, g are continuous, and satisfy a linear growth and Lipschitz condition so that the conditions of the Existence and Uniqueness Theorem are fulfilled for the SDE (47).

Let $I_h = \{t_0, \ldots, t_M\}$ be a time discretization of I with step sizes $h_n = t_{n+1} - t_n$ for $n = 0, \ldots, M-1$ and let $h = \max_{0\leqslant n<M} h_n$. A time discrete approximation scheme $\widehat{\mathbf{x}} = (\widehat{\mathbf{x}}_n)_{I_h}$ will be defined as a sequence $\widehat{\mathbf{x}}_0 = \mathbf{x}(t_0)$ and

$$\widehat{\mathbf{x}}_{n+1} = \Phi(\widehat{\mathbf{x}}_n, h_n, I_n) \qquad n = 0, \ldots, M-1$$

where I_0 is independent of $\widehat{\mathbf{x}}_0$, with $I_n = \boldsymbol{\omega}(t_{n+1} - t_n)$ Wiener increments drawn from the normal distributions with zero mean and variance h_n and which are independent of $\widehat{\mathbf{x}}_0, \dots, \widehat{\mathbf{x}}_n$ and $I_0, \dots, I_{n-1}$.

A scheme $\widehat{\mathbf{x}}$ converges in the strong (resp. weak) sense, with global order $p > 0$, to the solution $\mathbf{x}$ of the SDE (47) if there is a constant $C > 0$, independent of h and $\delta > 0$, such that, for each $h \in]0, \delta]$, we have

$$(\mathbb{E}[|\mathbf{x}(T) - \widehat{\mathbf{x}}_M|^2])^{1/2} \leq Ch^p, \quad (\text{resp. } |\mathbb{E}[G(\mathbf{x}(T))] - \mathbb{E}[G(\widehat{\mathbf{x}}_M)]| \leq Ch^p, \forall G \in \mathcal{C}_P^{2(p+1)}(\mathbb{R}^d, \mathbb{R})).$$

Notice that if $(\mathbb{E}[|\mathbf{x}(T) - \widehat{\mathbf{x}}_M|^2])^{1/2} = \mathcal{O}(h^p)$ then for every function f satisfying a Lipschitz condition, we have $|\mathbb{E}[f(\mathbf{x}(T))] - \mathbb{E}[f(\widehat{\mathbf{x}}_M)]| = \mathcal{O}(h^p)$. Nevertheless, this is not enough to infer the optimal weak order of convergence of such method.

Strong convergence is concerned with the precision of the path, while the weak convergence is with the precision of the moments. As, for DPMs, the center of attention is the evolution of the probability densities rather than that of the noising process of single data samples, weak convergence is enough to guarantee the well-conditioning of our numerical schemes. Moreover, when the diffusion coefficient vanishes, then both strong and weak convergence (with the choice $G = \mathrm{id}$) reduce to the usual deterministic convergence criterion for ODEs.

Let us now state some useful results that will be used later on.

Assumption B.3. All moments of the initial value $\widehat{\mathbf{x}}_0$ exist, f is continuous, satisfy a linear growth and globally Lipschitz condition.

In particular, since I is a closed finite interval in $\mathbb{R}$, $f(\cdot, x)$ and g are bounded by some constant.

Theorem B.4 ([18]). *In addition to B.3, suppose that*

$$\begin{aligned} |\mathbb{E}[\mathbf{x}(t_1) - \widehat{\mathbf{x}}_1]| &\leqslant Ch^{p_1} \\ (\mathbb{E}[|\mathbf{x}(t_1) - \widehat{\mathbf{x}}_1|^2])^{1/2} &\leqslant Ch^{p_2} \end{aligned}$$

with $p_2 \geqslant 1/2$ and $p_1 \geqslant p_2 + 1/2$. Then $\widehat{\mathbf{x}}$ is of strong global order $p = p_2 - 1/2$.

Assumption B.5. All moments of the initial value $\widehat{\mathbf{x}}_0$ exist, f is continuous, satisfy a linear growth and Lipschitz condition with all their components belonging to $\mathcal{C}_P^{p+1,2(p+1)}(I \times \mathbb{R}^d, \mathbb{R})$ and $g \in \mathcal{C}^{p+1}(I, \mathbb{R})$.

Theorem B.6 ([18]). *In addition to B.5, suppose that*

1. *for large enough $r \in \mathbb{N}$, the moments $\mathbb{E}[|\widehat{\mathbf{x}}_n|^{2r}]$ exist and are uniformingly bounded with respect to M and $n = 0, \dots, M$*
2. *for all $G \in \mathcal{C}_P^{2(p+1)}(\mathbb{R}^d, \mathbb{R})$, if $\widehat{\mathbf{x}}_n = \mathbf{x}(t_n)$, then for some $K \in \mathcal{C}_P^0(\mathbb{R}^d, \mathbb{R})$, we have*

$$|\mathbb{E}[G(\mathbf{x}(t_{n+1}))] - \mathbb{E}[G(\widehat{\mathbf{x}}_{n+1})]| \leq K(\widehat{\mathbf{x}}_n)h^{p+1}.$$

Then $\widehat{\mathbf{x}}$ is of weak global order p.

Lemma B.7. *Suppose that $\widehat{\mathbf{x}}_0$ has moments of all orders and that, for $h < 1$,*

$$\begin{aligned} |\mathbb{E}[\Phi(\widehat{\mathbf{x}}_n, h_n, I_n) - \widehat{\mathbf{x}}_n]| &\leqslant K(1 + |\widehat{\mathbf{x}}_n|)h \\ |\Phi(\widehat{\mathbf{x}}_n, h_n, I_n) - \widehat{\mathbf{x}}_n| &\leqslant X_n(1 + |\widehat{\mathbf{x}}_n|)h^{1/2} \end{aligned}$$

where X_n has moments of all orders. Then Condition 1 in Theorem B.6 is fulfilled.

B.2 Convergence of SEEDS-1

In this section we will prove that that SEEDS-1 as described above is of global strong order 1.0.

B.2.1 Strong Itô-Taylor approximation

Then the truncated Itô-Taylor expansion of the analytic solution $\mathbf{x}_t$ of the VP NP NRSDE starting from $\mathbf{x}_s$ is given, for $\epsilon \sim \mathcal{N}(0,1)$, by

$$\begin{aligned}\mathbf{x}_t &= \frac{\alpha_t}{\alpha_s}\mathbf{x}_s - 2\alpha_t\int_{\lambda_s}^{\lambda_t} e^{-\lambda}\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda,\lambda)\mathrm{d}\lambda - \sqrt{2}\alpha_t\int_{\lambda_s}^{\lambda_t} e^{-\lambda}\mathrm{d}\boldsymbol{\omega}_\lambda \\ &= \frac{\alpha_t}{\alpha_s}\mathbf{x}_s - 2\bar{\sigma}_t(e^h-1)\boldsymbol{\epsilon}_\theta(\mathbf{x}_s,s) - \bar{\sigma}_t\sqrt{e^{2h}-1}\epsilon + \mathcal{R}_1 \\ &= \frac{\alpha_t}{\alpha_s}\mathbf{x}_s - 2\bar{\sigma}_t(e^h-1)\boldsymbol{\epsilon}_\theta(\mathbf{x}_s,s) - \bar{\sigma}_t\sqrt{e^{2h}-1}\epsilon + \mathcal{O}(h)\end{aligned}$$

where the symbol $\mathcal{O}(h^p)$ represent terms $\mathbf{u}$ such that $\|\mathbf{u}\| \leqslant |K(\mathbf{x}_s)|h^p$, for $K \in \mathcal{C}^0_P(\mathbb{R}^d,\mathbb{R})$ and small $h>0$. The SEEDS-1 scheme corresponds to such truncated Itô-Taylor expansion containing only the time and Wiener integrals of multiplicity one. As such, since $G_t g(t) = 0$ and assuming Lipschitz and linear growth conditions on $\boldsymbol{\epsilon}_\theta$ as in Assumption B.1, $\mathbf{x}_t$ can be interpreted as an order 1.0 strong Itô-Taylor approximation [13, Theorem 10.6.3] of the solution to (42).

B.2.2 Continuous approximation of SEEDS-1

Let $\tilde{\alpha}_t := \frac{1}{2}\beta_d t^2 + \beta_m t$, where $\beta_d = \beta_{\max} - \beta_{\min} = 19.9$, $\beta_m = \beta_{\min} = 0.1$. We have

$$\sigma_t = \sqrt{e^{\tilde{\alpha}_t}-1}, \qquad \alpha_t = e^{-\frac{1}{2}\tilde{\alpha}_t} = \frac{1}{\sqrt{\sigma_t^2+1}}$$

so, in particular, as $T=1$, we have $\tilde{\alpha}_{t_0} = \tilde{\alpha}_1 = \frac{1}{2}(\beta_{\max}-\beta_{\min})+\beta_{\min} = \frac{1}{2}(\beta_{\max}+\beta_{\min}) \approx 10.05$ and $\tilde{\alpha}_{t_{M+1}} = \tilde{\alpha}_0 = 0$. We deduce $\alpha_{t_0} = \alpha_1 \approx e^{-\frac{10.05}{2}} < 1$, $\alpha_{t_{M+1}} = \alpha_0 = 1$. Next, $\sigma_{t_0} = \sigma_1 \approx \sqrt{e^{10.05}-1} > 1$ and $\sigma_{t_{M+1}} = \sigma_0 = 0$. As such, $\lambda_{t_0} = -\log(\sigma_1) := -L_0 < 0$ and $\lambda_t \underset{t\to t_{M+1}}{\longrightarrow} +\infty$. As such, we will set $t_M = t_{M+1} + \varepsilon$ the end time so that $\lambda_{t_M} = L_0$ is finite. This implies that, for $\lambda \in [-L_0, +\infty[$, $0 < e^{-L_0} \leqslant e^\lambda < 1 < e^{\lambda_{t_M}}$ and, for $\hat{T} = \lambda_{t_M} - \lambda_{t_0}$, $e^h = e^{\lambda_t - \lambda_s} \leqslant e^{\hat{T}}$. Now set, for $\lambda \in [-L_0, +\infty[$,

$$\hat{\alpha}_\lambda := \sqrt{\frac{1}{1+e^{-2\lambda}}}, \qquad \hat{\sigma}_\lambda := \sqrt{\frac{1}{1+e^{+2\lambda}}}$$

Then, as λ increases, $\hat{\alpha}_\lambda$ increases starting from $0 < \hat{\alpha}_{\lambda_{t_0}} < 1$ and $\hat{\alpha}_\lambda \underset{\lambda\to+\infty}{\longrightarrow} 1$ while at the same time $\hat{\sigma}_\lambda$ decreases starting from $0 < \hat{\sigma}_{\lambda_{t_0}} < 1$ and $\hat{\sigma}_\lambda \underset{\lambda\to+\infty}{\longrightarrow} 0$. As such, we can rewrite the exact solution (43) as

$$\widehat{\mathbf{x}}_{\lambda_t} = \frac{\hat{\alpha}_{\lambda_t}}{\hat{\alpha}_{\lambda_s}}\widehat{\mathbf{x}}_{\lambda_s} - 2\hat{\alpha}_{\lambda_t}\int_{\lambda_s}^{\lambda_t} e^{-\lambda}\hat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda,\lambda)\mathrm{d}\lambda - \sqrt{2}\hat{\alpha}_{\lambda_t}\int_{\lambda_s}^{\lambda_t} e^{-\lambda}\mathrm{d}\bar{\boldsymbol{\omega}}(\lambda). \tag{48}$$

Notice that $\frac{\hat{\alpha}_{\lambda_t}}{\hat{\alpha}_{\lambda_s}}$ is bounded for all t, s by

$$\frac{\hat{\alpha}_{\lambda_t}}{\hat{\alpha}_{\lambda_s}} \leqslant \frac{1}{\hat{\alpha}_{\lambda_{t_0}}}$$

and $0 < \hat{\alpha}_\lambda < 1$ for all $\lambda \in [-L_0, +\infty[$.

Recall that the SEEDS-1 is defined recursively as

$$\mathbf{y}_{t_0} \leftarrow \mathbf{x}_T, \mathbf{y}_{t_i} \leftarrow \frac{\alpha_{t_i}}{\alpha_{t_{i-1}}}\mathbf{y}_{t_{i-1}} - 2\bar{\sigma}_{t_i}(e^{h_i}-1)\boldsymbol{\epsilon}_\theta(\mathbf{y}_{t_{i-1}}, t_{i-1}) - \bar{\sigma}_{t_i}\sqrt{e^{2h_i}-1}\epsilon_i$$

and, for simplicity, we will denote $\mathbf{y}_{\lambda_{t_i}}$ for the iterates (48).

Define a continuous approximation of SEEDS-1 as follows. For $\hat{h} = t - s$, we write $\hat{s} = [s/\hat{h}]\hat{h}$ where $[x]$ denotes the largest integer lesser or equal to x and $I_{[A]}$ is the indicator function associated to a set A. We define the step function:

$$\widehat{\mathbf{y}}(\lambda) := \sum_{k \geqslant 0} I_{[\lambda_{t_k}, \lambda_{t_{k+1}}[} \mathbf{y}_{\lambda_{t_k}}$$

and the continuous approximation

$$\mathbf{y}(t) := \frac{\alpha_t}{\alpha_{t_0}} \mathbf{y}(t_0) - 2\alpha_t \int_{\lambda_{t_0}}^{\lambda_t} e^{-\hat{\lambda}} \hat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{y}}(\lambda), \hat{\lambda}) \mathrm{d}\lambda - \sqrt{2}\alpha_t \int_{\lambda_{t_0}}^{\lambda_t} e^{-\lambda} \mathrm{d}\boldsymbol{\omega}_\lambda,$$

Proposition B.8. *There are two constants C_1, C_2 independent of h such that, for all $t \in [0, T]$, we have*

$$\begin{aligned} \mathbb{E}\left[\sup_{t_0 \leqslant t \leqslant t_M} |\mathbf{y}(t)|^2\right] &\leqslant C_1 \\ \mathbb{E}\left[|\mathbf{y}(t) - \widehat{\mathbf{y}}(t)|^2\right] &\leqslant C_2 h^2. \end{aligned}$$

Proof. Recall the standard inequality $(a+b+c)^2 \leqslant 3(a^2+b^2+c^2)$ for $a, b, c \in \mathbb{R}$. Then:

$$|\mathbf{y}(t)|^2 \leqslant 3\left[\left|\frac{\alpha_t}{\alpha_{t_0}} \mathbf{y}_{t_0}\right|^2 + 4\alpha_t^2 \left|\int_{\lambda_{t_0}}^{\lambda_t} e^{-\hat{\lambda}} \hat{\epsilon}_\theta(\widehat{\mathbf{y}}(\lambda), \hat{\lambda}) \mathrm{d}\lambda\right|^2 + 2\alpha_t^2 \left|\int_{\lambda_{t_0}}^{\lambda_t} e^{-\lambda} \mathrm{d}\boldsymbol{\omega}_\lambda\right|^2\right]$$

Using the fact that $\alpha_t \leqslant 1$, we have, by writing $\hat{T} = \lambda_{t_M} - \lambda_{t_0}$ and taking the expectation:

$$\begin{aligned} &\mathbb{E}\left[\sup_{t_0 \leqslant t \leqslant t_M} |\mathbf{y}(t)|^2\right] \\ \leqslant\ & 3\left[\left|\frac{\alpha_t}{\alpha_{t_0}}\right|^2 \mathbb{E}\left[|\mathbf{y}_{t_0}|^2\right] + 4\mathbb{E}\left[\left|\int_{\lambda_{t_0}}^{\lambda_t} e^{-\hat{\lambda}} \hat{\epsilon}_\theta(\widehat{\mathbf{y}}(\lambda), \hat{\lambda}) \mathrm{d}\lambda\right|^2\right] + 2\alpha_t^2 \mathbb{E}\left[\left|\int_{\lambda_{t_0}}^{\lambda_t} e^{-\lambda} \mathrm{d}\boldsymbol{\omega}_\lambda\right|^2\right]\right] \\ \leqslant\ & 3\left[\left|\frac{\alpha_t}{\alpha_{t_0}}\right|^2 \mathbb{E}\left[|\mathbf{y}_{t_0}|^2\right] + 4\hat{T}\mathbb{E}\left[\int_{\lambda_{t_0}}^{\lambda_t} |e^{-\hat{\lambda}}|^2 \left|\hat{\epsilon}_\theta(\widehat{\mathbf{y}}(\lambda), \hat{\lambda})\right|^2 \mathrm{d}\lambda\right] + 2\alpha_t^2 \mathbb{E}\left[\int_{\lambda_{t_0}}^{\lambda_t} |e^{-\lambda}|^2 \mathrm{d}\lambda\right]\right] \\ \leqslant\ & 3\left[\left|\frac{\alpha_t}{\alpha_{t_0}}\right|^2 \mathbb{E}\left[|\mathbf{y}_{t_0}|^2\right] + 4\hat{T}\int_{\lambda_{t_0}}^{\lambda_t} |e^{-\hat{\lambda}}|^2 \mathbb{E}\left[\left|\hat{\epsilon}_\theta(\widehat{\mathbf{y}}(\lambda), \hat{\lambda})\right|^2\right] \mathrm{d}\lambda + 2\alpha_t^2 \int_{\lambda_{t_0}}^{\lambda_t} e^{-2\lambda} \mathrm{d}\lambda\right] \\ \leqslant\ & 3K\left[\mathbb{E}\left[|\mathbf{y}_{t_0}|^2\right] + 4\hat{T}\int_{\lambda_{t_0}}^{\lambda_t} \mathbb{E}\left[\left|\hat{\epsilon}_\theta(\widehat{\mathbf{y}}(\lambda), \hat{\lambda})\right|^2\right] \mathrm{d}\lambda + \left(e^{2(\lambda_t - \lambda_{t_0})} - 1\right)\right] \\ \leqslant\ & 3K\left[\mathbb{E}\left[|\mathbf{y}_{t_0}|^2\right] + 4\hat{T}L_2 \int_{\lambda_{t_0}}^{\lambda_t} \left(1 + \mathbb{E}\left[|\widehat{\mathbf{y}}(\lambda)|^2\right]\right) \mathrm{d}\lambda + (e^{2\hat{T}} - 1)\right] \\ \leqslant\ & 3K\left[\mathbb{E}\left[|\mathbf{y}_{t_0}|^2\right] + 4\hat{T}L_2\hat{T} + e^{2\hat{T}} - 1 + 4\hat{T}L_2 \int_{\lambda_{t_0}}^{\lambda_t} \mathbb{E}\left[|\widehat{\mathbf{y}}(\lambda)|^2\right] \mathrm{d}\lambda\right] \\ \leqslant\ & 3K\left(\mathbb{E}\left[|\mathbf{y}_{t_0}|^2\right] + 4\hat{T}L_2\hat{T} + e^{2\hat{T}} - 1\right) + 3K4\hat{T}L_2 \int_{\lambda_{t_0}}^{\lambda_t} \mathbb{E}\left[|\widehat{\mathbf{y}}(\lambda)|^2\right] \mathrm{d}\lambda \\ \leqslant\ & 3K\left(\mathbb{E}\left[|\mathbf{x}(t_0)|^2\right] + 4\hat{T}^2L_2 + e^{2\hat{T}} - 1\right) + 12K\hat{T}L_2 \int_{\lambda_{t_0}}^{\lambda_t} \mathbb{E}\left[\sup_{\lambda_0 \leqslant r \leqslant \lambda} |\mathbf{y}(r)|^2\right] \mathrm{d}\lambda \end{aligned}$$

where we used the linearity of expectation, Hölder's inequality, Doob's martingale inequality, Itô isometry, the linear growth condition of $\hat{\epsilon}_\theta$, and we set $K = \max\left(\left|\frac{\alpha_t}{\alpha_{t_0}}\right|^2, |e^{L_0}|^2, \bar{\sigma}_t^2, 1\right)$. As we

know that $\mathbb{E}\left[|\mathbf{x}(t_0)|^2\right] < \infty$, we apply Grönwall's inequality in the last line to obtain

$$\mathbb{E}\left[\sup_{t_0 \leqslant t \leqslant t_M} |\mathbf{y}(t)|^2\right] \leqslant C_1, \qquad C_1 := 3K\left[\mathbb{E}\left[|\mathbf{x}(t_0)|^2\right] + 4\hat{T}^2 L_2 + e^{2\hat{T}} - 1\right] e^{12K\hat{T}L_2}.$$

Second, we have, for $s = t_i$, $u = t_{i+1}$ and $t \in [t_\lambda(u), t_\lambda(s)[$,

$$\mathbf{y}(t) - \widehat{\mathbf{y}}(t) = \left(\frac{\alpha_t}{\alpha_s} - 1\right)\mathbf{y}_s - 2\alpha_t \int_{\lambda_s}^{\lambda_t} e^{-\hat{\lambda}} \hat{\epsilon}_\theta(\mathbf{y}_{\lambda_s}, \lambda_s) \mathrm{d}\lambda,$$

so that, using Hölder's inequality, we get

$$|\mathbf{y}(t) - \widehat{\mathbf{y}}(t)|^2 \leqslant 3\left[\left|\frac{\alpha_t}{\alpha_s} - 1\right|^2 |\mathbf{y}_s|^2 + 4h \int_{\lambda_s}^{\lambda_t} |e^{-\hat{\lambda}}|^2 \left|\hat{\epsilon}_\theta(\mathbf{y}_{\lambda_s}, \lambda_s)\right|^2 \mathrm{d}\lambda\right].$$

Now using Itô isometry we obtain:

$$\mathbb{E}\left[|\mathbf{y}(t) - \widehat{\mathbf{y}}(t)|^2\right] \leqslant 3\left[\left|\frac{\alpha_t}{\alpha_s} - 1\right|^2 \mathbb{E}\left[|\mathbf{y}_s|^2\right] + 4h\mathbb{E}\left[\int_{\lambda_s}^{\lambda_t} |e^{-\hat{\lambda}}|^2 \left|\hat{\epsilon}_\theta(\mathbf{y}_{\lambda_s}, \lambda_s)\right|^2 \mathrm{d}\lambda\right]\right].$$

Now, using the bound $\mathbb{E}\left[\max |\mathbf{y}_t|^2\right] \leqslant C_1$, the fact that ϵ_θ is bounded and the same arguments as above, we get

$$\begin{aligned}
\mathbb{E}\left[|\mathbf{y}(t) - \widehat{\mathbf{y}}(t)|^2\right] &\leqslant 3\left[\left|\frac{\alpha_t}{\alpha_s} - 1\right|^2 C_1 + 4Kh\mathbb{E}\left[\int_{\lambda_s}^{\lambda_t} \left|\hat{\epsilon}_\theta(\mathbf{y}_{\lambda_s}, \lambda_s)\right|^2 \mathrm{d}\lambda\right]\right] \\
&\leqslant 3\left[\left|\frac{\alpha_t}{\alpha_s} - 1\right|^2 C_1 + 4KhL_2 \int_{\lambda_s}^{\lambda_t} (1 + C_1)\mathrm{d}\lambda\right] \\
&\leqslant 3\left[\left|\frac{\alpha_t}{\alpha_s} - 1\right|^2 C_1 + 4Kh^2 L_2(1 + C_1)\right].
\end{aligned}$$

Finally, as we have

$$e^h = \frac{\sigma_s}{\sigma_t} = \frac{\alpha_t}{\alpha_s}\frac{\sqrt{1-\alpha_s^2}}{\sqrt{1-\alpha_t^2}},$$

and $1 < \frac{\sqrt{1-\alpha_s^2}}{\sqrt{1-\alpha_t^2}} \underset{M\to\infty}{\longrightarrow} 1$, we obtain $\left|\frac{\alpha_t}{\alpha_s} - 1\right|^2 \sim |e^h - 1|^2 \sim \mathcal{O}(h^2)$. Now, by denoting $|e^h - 1|^2 \leqslant K_2 h^2$, we conclude that

$$\mathbb{E}\left[|\mathbf{y}(t) - \widehat{\mathbf{y}}(t)|^2\right] \leqslant C_2 h^2,$$

with $C_2 := 3K_2 C_1 + 4KL_2(1 + C_1)$. This finishes the proof. □

B.2.3 Proof of Theorem 4.1

Let's now take a look at the approximation given $\mathbf{y}_{t_0} = \mathbf{x}_{t_0}$. We have

$$\mathbf{y}_t - \mathbf{x}_t = 2\alpha_t \int_{\lambda_{t_0}}^{\lambda_t} \left[e^{-\lambda}\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda, \lambda) - e^{-\hat{\lambda}}\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{y}}_\lambda, \hat{\lambda})\right] \mathrm{d}\lambda.$$

Using the inequality $\alpha_t \leq 1$, the Lipschitz property of $\widehat{\boldsymbol{\epsilon}}_\theta$, Assumption (B.1), and Hölder's inequality we deduce the bound:

$$\begin{aligned}
|\mathbf{y}_t - \mathbf{x}_t|^2 &\leqslant 2\left[4\left|\int_{\lambda_{t_0}}^{\lambda_t} \left[e^{-\lambda}\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda, \lambda) - e^{-\hat{\lambda}}\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{y}}_\lambda, \hat{\lambda})\right] \mathrm{d}\lambda\right|^2\right] \\
&\leqslant 8\hat{T}\int_{\lambda_{t_0}}^{\lambda_t} \left|e^{-\lambda}\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda, \lambda) - e^{-\hat{\lambda}}\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{y}}_\lambda, \hat{\lambda})\right|^2 \mathrm{d}\lambda.
\end{aligned}$$

Taking the expectation yields

$$\begin{aligned}
&\mathbb{E}\left[\sup_{t_0\leqslant s\leqslant t}|\mathbf{y}_s-\mathbf{x}_s|^2\right]\\
\leqslant\;& 8\hat{T}\mathbb{E}\left[\int_{\lambda_{t_0}}^{\lambda_t}\left|e^{-\lambda}\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda,\lambda)-e^{-\hat{\lambda}}\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{y}}_\lambda,\hat{\lambda})\right|^2\mathrm{d}\lambda\right].
\end{aligned}$$

Now, for the first integral, by writing:

$$\begin{aligned}
&e^{-\lambda}\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda,\lambda)-e^{-\hat{\lambda}}\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{y}}_\lambda,\hat{\lambda})\\
=\;& e^{-\lambda}\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda,\lambda)-e^{-\hat{\lambda}}\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda,\lambda)+e^{-\hat{\lambda}}\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda,\lambda)-e^{-\hat{\lambda}}\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda,\hat{\lambda})\\
&+e^{-\hat{\lambda}}\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda,\hat{\lambda})-e^{-\hat{\lambda}}\widehat{\boldsymbol{\epsilon}}_\theta(\mathbf{y}_\lambda,\hat{\lambda})+e^{-\hat{\lambda}}\widehat{\boldsymbol{\epsilon}}_\theta(\mathbf{y}_\lambda,\hat{\lambda})-e^{-\hat{\lambda}}\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{y}}_\lambda,\hat{\lambda})\\
=\;& (e^{-\lambda}-e^{-\hat{\lambda}})\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda,\lambda)+e^{-\hat{\lambda}}\left(\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda,\lambda)-\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda,\hat{\lambda})\right)\\
&+e^{-\hat{\lambda}}\left(\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda,\hat{\lambda})-\widehat{\boldsymbol{\epsilon}}_\theta(\mathbf{y}_\lambda,\hat{\lambda})+\widehat{\boldsymbol{\epsilon}}_\theta(\mathbf{y}_\lambda,\hat{\lambda})-\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{y}}_\lambda,\hat{\lambda})\right),
\end{aligned}$$

we can state the following inequalities:

$$\begin{aligned}
&\mathbb{E}\left[\int_{\lambda_{t_0}}^{\lambda_t}\left|e^{-\lambda}\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda,\lambda)-e^{-\hat{\lambda}}\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{y}}_\lambda,\hat{\lambda})\right|^2\mathrm{d}\lambda\right]\\
\leqslant\;& 3\int_{\lambda_{t_0}}^{\lambda_t}|e^{-\lambda}-e^{-\hat{\lambda}}|^2\mathbb{E}\left[|\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda,\lambda)|^2\right]\mathrm{d}\lambda+3\mathbb{E}\left[\int_{\lambda_{t_0}}^{\lambda_t}|e^{-\hat{\lambda}}|^2\left|\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda,\lambda)-\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda,\hat{\lambda})\right|^2\mathrm{d}\lambda\right]\\
&+3\mathbb{E}\left[\int_{\lambda_{t_0}}^{\lambda_t}|e^{-\hat{\lambda}}|^2\left|\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda,\hat{\lambda})-\widehat{\boldsymbol{\epsilon}}_\theta(\mathbf{y}_\lambda,\hat{\lambda})+\widehat{\boldsymbol{\epsilon}}_\theta(\mathbf{y}_\lambda,\hat{\lambda})-\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{y}}_\lambda,\hat{\lambda})\right|^2\mathrm{d}\lambda\right]\\
\leqslant\;& 3\int_{\lambda_{t_0}}^{\lambda_t}|e^{-\lambda}-e^{-\hat{\lambda}}|^2\mathbb{E}\left[|\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda,\lambda)|^2\right]\mathrm{d}\lambda+3KL_3\mathbb{E}\left[\int_{\lambda_{t_0}}^{\lambda_t}(1+|\widehat{\mathbf{x}}_\lambda|^2)|(\lambda-\hat{\lambda})|^2\mathrm{d}\lambda\right]\\
&+3\mathbb{E}\left[\int_{\lambda_{t_0}}^{\lambda_t}|e^{-\hat{\lambda}}|^2\left|\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda,\hat{\lambda})-\widehat{\boldsymbol{\epsilon}}_\theta(\mathbf{y}_\lambda,\hat{\lambda})+\widehat{\boldsymbol{\epsilon}}_\theta(\mathbf{y}_\lambda,\hat{\lambda})-\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{y}}_\lambda,\hat{\lambda})\right|^2\mathrm{d}\lambda\right]\\
\leqslant\;& 3L_2\int_{\lambda_{t_0}}^{\lambda_t}|e^{-\hat{\lambda}}|^2|e^{\lambda-\hat{\lambda}}-1|^2(1+\mathbb{E}[|\widehat{\mathbf{x}}_\lambda|^2])\mathrm{d}\lambda+3KL_3h^2\mathbb{E}\int_{\lambda_{t_0}}^{\lambda_t}(1+|\widehat{\mathbf{x}}_\lambda|^2)\mathrm{d}\lambda\\
&+3KL_1\int_{\lambda_{t_0}}^{\lambda_t}\mathbb{E}\left[|\widehat{\mathbf{x}}_\lambda-\mathbf{y}_\lambda+\mathbf{y}_\lambda-\widehat{\mathbf{y}}_\lambda|^2\right]\mathrm{d}\lambda.
\end{aligned}$$

Thus, we obtain

$$\begin{aligned}
&\mathbb{E}\left[\int_{\lambda_{t_0}}^{\lambda_t}\left|e^{-\lambda}\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{x}}_\lambda,\lambda)-e^{-\hat{\lambda}}\widehat{\boldsymbol{\epsilon}}_\theta(\widehat{\mathbf{y}}_\lambda,\hat{\lambda})\right|^2\mathrm{d}\lambda\right]\\
\leqslant\;& 3L_2K\hat{T}|e^h-1|^2(1+C_1)+3L_3h^2K(1+C_1)\hat{T}\\
&+3KL_1\int_{\lambda_{t_0}}^{\lambda_t}\mathbb{E}\left[|\widehat{\mathbf{x}}_\lambda-\mathbf{y}_\lambda+\mathbf{y}_\lambda-\widehat{\mathbf{y}}_\lambda|^2\right]\mathrm{d}\lambda\\
\leqslant\;& 3L_2K\hat{T}|e^h-1|^2(1+C_1)+3L_3h^2K(1+C_1)\hat{T}\\
&+6KL_1\int_{\lambda_{t_0}}^{\lambda_t}\mathbb{E}\left[|\widehat{\mathbf{x}}_\lambda-\mathbf{y}_\lambda|^2+|\mathbf{y}_\lambda-\widehat{\mathbf{y}}_\lambda|^2\right]\mathrm{d}\lambda
\end{aligned}$$

$$
\begin{aligned}
&\leqslant 3L_2K\hat{T}|e^h-1|^2(1+C_1)+3L_3h^2K(1+C_1)\hat{T}\\
&\quad +6KL_1\hat{T}C_2h^2+6KL_1\int_{\lambda_{t_0}}^{\lambda_t}\mathbb{E}\left[|\widehat{\mathbf{x}}_\lambda-\mathbf{y}_\lambda|^2\right]\mathrm{d}\lambda\\
&\leqslant 3L_2K\hat{T}|e^h-1|^2(1+C_1)+3L_3h^2K(1+C_1)\hat{T}\\
&\quad +6KL_1\hat{T}C_2h^2+6KL_1\int_{\lambda_{t_0}}^{\lambda_t}\mathbb{E}\left[\sup_{\lambda_0\leqslant r\leqslant\lambda}|\widehat{\mathbf{x}}(r)-\mathbf{y}(r)|^2\right]\mathrm{d}\lambda\\
&\leqslant 3K\hat{T}(1+C_1)[L_2|e^h-1|^2+L_3h^2]+6KL_1\hat{T}C_2h^2\\
&\quad +6KL_1\int_{\lambda_{t_0}}^{\lambda_t}\mathbb{E}\left[\sup_{\lambda_0\leqslant r\leqslant\lambda}|\widehat{\mathbf{x}}(r)-\mathbf{y}(r)|^2\right]\mathrm{d}\lambda.
\end{aligned}
$$

Putting everything together yields

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{t_0\leqslant s\leqslant t}|\mathbf{y}_s-\mathbf{x}_s|^2\right]\\
&\leqslant 8\hat{T}[3K\hat{T}(1+C_1)[L_2|e^h-1|^2+L_3h]+6KL_1\hat{T}C_2h^2]\\
&\quad +8\hat{T}6KL_1\int_{\lambda_{t_0}}^{\lambda_t}\mathbb{E}\left[\sup_{\lambda_0\leqslant r\leqslant\lambda}|\widehat{\mathbf{x}}(r)-\mathbf{y}(r)|^2\right]\mathrm{d}\lambda\\
&\leqslant 24K\hat{T}^2(1+C_1)[L_2|e^h-1|^2+L_3h^2]+48KL_1\hat{T}^2C_2h^2\\
&\quad +48\hat{T}KL_1\int_{\lambda_{t_0}}^{\lambda_t}\mathbb{E}\left[\sup_{\lambda_0\leqslant r\leqslant\lambda}|\widehat{\mathbf{x}}(r)-\mathbf{y}(r)|^2\right]\mathrm{d}\lambda.
\end{aligned}
$$

Now, by denoting $|e^h-1|^2\leqslant K_2h^2$, we apply the continuous version of the Grönwall Lemma to obtain:

$$
\mathbb{E}\left[\sup_{t_0\leqslant s\leqslant t}|\mathbf{y}_t-\mathbf{x}_t|^2\right]\leqslant C_0h^2,
$$

where

$$
C_0=[24K\hat{T}^2(1+C_1)[L_2K_2+L_3]+48K\hat{T}^2L_1C_2]e^{48\hat{T}KL_1}.
$$

Finally, using Lyapunov's inequality we obtain, for $C=\sqrt{C_0}$

$$
\mathbb{E}[|\mathbf{y}_T-\mathbf{x}_T|]\leqslant\left(\mathbb{E}[|\mathbf{y}_T-\mathbf{x}_T|^2]\right)^{1/2}\leqslant Ch.
$$

In other words, for C as stated above, we have the following inequality

$$
\sqrt{\mathbb{E}\left[\sup_{t_0\leqslant t\leqslant t_M}|\widetilde{\mathbf{x}}_t-\mathbf{x}_t|^2\right]}\leqslant Ch,\qquad\text{as } h\longrightarrow 0.
$$

Remark B.9. From the above, it is easy to induce that the following order for the one step error

$$
\mathbb{E}[|\mathbf{x}(t_1)-\mathbf{y}_1|^2]=\mathcal{O}(h^3).
$$

Now, since $G_tg(t)=0$ in this case (additive noise), it is easy to see from the truncated Itô-Taylor expansion of $\mathbf{x}$ that $|\mathbb{E}[\mathbf{x}(t_1)-\mathbf{y}_1]|=\mathcal{O}(h^2)$. As such, we apply Theorem B.4 to conclude that SEEDS-1 has strong convergence of global order 1.0.

B.2.4 Discrete-time approximation

By Theorem 4.1, we know that SEEDS-1, being of strong order 1.0, it is immediately also of weak order 1. Nonetheless, let's give a discrete approach of this statement that we will use for the proofs of convergence for the remaining solvers as stated in Corollary 4.3. Define the discrete time process:

$$
\mathbf{y}_{t_0}\leftarrow\mathbf{x}_T,\mathbf{y}_{t_i}\leftarrow\frac{\alpha_{t_i}}{\alpha_{t_{i-1}}}\mathbf{y}_{t_{i-1}}-2\sigma_{t_i}(e^{h_i}-1)\boldsymbol{\epsilon}(\mathbf{y}_{t_{i-1}},t_{i-1})-\sqrt{2}\alpha_{t_i}\int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}}e^{-s}\mathrm{d}\bar{\boldsymbol{\omega}}(s).
$$

We will prove that $\mathbb{E}[|\mathbf{y}_{t_M} - \mathbf{x}_{t_M}|]$ is of order h as $h \longrightarrow 0$. Note that $(\mathbf{y}_{t_i})_i$ has the same distribution as $(\widetilde{\mathbf{x}}_{t_i})_i$ described in Algorithm 2 since the stochastic integrals $\left(\int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-s} \mathrm{d}\bar{\boldsymbol{\omega}}(s)\right)_i$ are independent and each $\int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-s} \mathrm{d}\bar{\boldsymbol{\omega}}(s)$ is distributed as $\frac{1}{\sqrt{2}} \frac{\sigma_{t_i}}{\alpha_{t_i}} \sqrt{e^{2h_i} - 1}\epsilon$ with $h_i = \lambda_{t_i} - \lambda_{t_{i-1}}, \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$. We have:

$$\mathbf{y}_{t_i} - \mathbf{x}_{t_i} = \frac{\alpha_{t_i}}{\alpha_{t_{i-1}}}(\mathbf{y}_{t_{i-1}} - \mathbf{x}_{t_{i-1}}) + 2\alpha_{t_i} \int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-\lambda}(\widehat{\boldsymbol{\epsilon}}(\widehat{\mathbf{x}}_\lambda, \lambda) - \boldsymbol{\epsilon}(\mathbf{y}_{t_{i-1}}, t_{i-1}))\mathrm{d}\lambda.$$

For simplicity, in what follows C will denote a constant not dependent on the subdivision of $[0, T]$ that may change from one line to the next by systematically denoting the maximum value of the different constants appearing in the line before. Using the inequality $\alpha_t \leq 1$, the Lipschitz property of ϵ, we deduce the bound:

$$|\mathbf{y}_{t_i} - \mathbf{x}_{t_i}| \leq \frac{\alpha_{t_i}}{\alpha_{t_{i-1}}}|\mathbf{y}_{t_{i-1}} - \mathbf{x}_{t_{i-1}}| + C\left[\int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-\lambda}|\widehat{\mathbf{x}}_\lambda - \mathbf{y}_{t_{i-1}}|\mathrm{d}\lambda + \int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-\lambda}|t_\lambda(\lambda) - t_{i-1}|\mathrm{d}\lambda\right].$$

Notice that

$$\begin{aligned}\int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-\lambda}|\widehat{\mathbf{x}}_\lambda - \mathbf{y}_{t_{i-1}}|\mathrm{d}\lambda &\leq \int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-\lambda}|\widehat{\mathbf{x}}_\lambda - \mathbf{x}_{t_{i-1}}|\mathrm{d}\lambda + \int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-\lambda}|\mathbf{x}_{t_{i-1}} - \mathbf{y}_{t_{i-1}}|\mathrm{d}\lambda \\ &\leq \int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-\lambda}|\widehat{\mathbf{x}}_\lambda - \mathbf{x}_{t_{i-1}}|\mathrm{d}\lambda + Ch|\mathbf{x}_{t_{i-1}} - \mathbf{y}_{t_{i-1}}|,\end{aligned}$$

and recall that $\widehat{\mathbf{x}}_u = \mathbf{x}_{t_\lambda(u)}$. Using the fact that t_λ is increasing and Lemma B.11, we have:

$$\mathbb{E}[\int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-\lambda}|\widehat{\mathbf{x}}_\lambda - \mathbf{x}_{t_{i-1}}|\mathrm{d}\lambda] \leq C\int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-u}\sqrt{|t_\lambda - t_{i-1}|}\mathrm{d}\lambda \leq Ch^{3/2}.$$

Introduce $U_i = \mathbb{E}[|\mathbf{y}_{t_i} - \mathbf{x}_{t_i}|]$. Since $\int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-\lambda}|t_\lambda(\lambda) - t_{i-1}|\mathrm{d}\lambda \leq Ch^2$:

$$U_i \leq \left(\frac{\alpha_{t_i}}{\alpha_{t_{i-1}}} + Ch\right) U_{i-1} + Ch^2.$$

Let $a_i = \frac{\alpha_{t_i}}{\alpha_{t_{i-1}}} + Ch$ and $b_i = Ch^2$. By applying Lemma B.12, we have: $U_M \leq A_M U_0 + \sum_{k=1}^M A_{k,n} b_k$ with $A_M = \prod_{k=1}^M a_k$ and $A_{k,M} = A_M / A_k = \prod_{i=k+1}^M a_i$. Note that $U_0 = 0$ since $\mathbf{y}_{t_0} = \mathbf{x}_T$ and so:

$$U_M \leq Ch^2 \sum_{k=0}^{M-1} (\sup_{1 \leq i \leq M} a_i)^k.$$

Using our hypothesis, we can bound:

$$\sum_{k=0}^{M-1} (\sup_{1 \leq i \leq M} a_i)^k \leq \sum_{k=0}^{M-1} (\exp(Ch) + Ch)^k = \frac{(\exp(Ch) + Ch)^{1/h} - 1}{(\exp(Ch) + Ch) - 1}.$$

The quantity on the right is of order C/h. Indeed, as h goes to 0, $\exp(Ch) + Ch - 1$ is equivalent to $2Ch$ and $(\exp(Ch) + Ch)^{1/h}$ converges to a constant. This gives the bound $U_M \leq Ch$, when $h \to 0$ and by using Proposition B.8 and Theorem B.6 we conclude that SEEDS-1 is convergent of weak order 1.

B.2.5 Useful Lemmas

Lemma B.10. *(Continuous Grönwall Lemma) Let $I = [a, b]$ denote a compact interval of the real line with $a < b$. Let α, β, u be continuous real-valued functions defined on I. Assume β is non-negative, α is non-decreasing and u satisfies the integral inequality*

$$u(t) \leq \alpha(t) + \int_a^t \beta(s)u(s)\,\mathrm{d}s, \qquad \forall t \in I.$$

Then

$$u(t) \leq \alpha(t) \exp\left(\int_a^t \beta(s)\,\mathrm{d}s\right), \qquad \forall t \in I.$$

Lemma B.11. *Assume the following forward SDE is satisfied $dX_t = F(t, X_t)dt + G(t)dW_t, t \in [0, T]$, where $T > 0$, $F(t, x)$ is Lipschitz with respect to (t, x), G is continuous and X_0 is an integrable random variable. Then, there exists $C > 0$ such that for all $s < t \in [0, T]$ with $t - s \leq 1$,*

$$\mathbb{E}[|X_t - X_s|] \leq C\sqrt{t-s}.$$

Proof. We take $s = 0$ and apply the triangular inequality:

$$|X_t - X_0| \leq \int_0^t |F(u, X_u) - F(0, X_0))|\mathrm{d}u + |\int_0^t G(u)dW_u| + t|F(0, X_0)|.$$

Setting $u(t) = \mathbb{E}[|X_t - X_0|]$ and taking the expectation, we deduce:

$$u(t) \leq K\frac{t^2}{2} + t\mathbb{E}[|F(0, X_0)|] + \mathbb{E}[(\int_0^t G(u)dW_u)^2]^{\frac{1}{2}} + K\int_0^t u(s)ds,$$

where K is a positive constant. Note that $\mathbb{E}[(\int_0^t G(u)dW_u)^2]^{\frac{1}{2}} = [\int_0^t G^2(s)ds]^{\frac{1}{2}}$ by the Itô isometry property which is less than $C\sqrt{t}$ since G is continuous. Thus, we have proved that:

$$u(t) \leq \alpha(t) + K\int_0^t u(s)ds,$$

where $\alpha(t) = K\frac{t^2}{2} + t\mathbb{E}[|F(0, X_0)|] + C\sqrt{t}$ is non-decreasing. By Lemma B.10: $u(t) \leq \alpha(t)\exp(Kt) \leq \alpha(t)\exp(KT)$. Since $\alpha(t) \leq C\sqrt{t}$ for $t \in [0, 1]$, the lemma holds. □

Lemma B.12. *(Discrete Grönwall Lemma) Consider a real number sequence $(u_n)_n$ such that*

$$u_{n+1} \leq a_{n+1}u_n + b_{n+1}, n \geq 0$$

where (a_n) and $(b_n)_n$ are two given sequences such that b_n is positive. Then

$$u_n \leq A_n u_0 + \sum_{k=1}^{n} A_{k,n}b_k,$$

where $A_n = \prod_{k=1}^n a_k$ and $A_{k,n} = A_n/A_k = \prod_{i=k+1}^n a_i$.

B.3 Proof of Corollary 4.3

B.3.1 Convergence of SEEDS-2:

Let $\{\mathbf{y}_{t_i}\}_i$ be the discrete stochastic process defined as follows:

$$\mathbf{y}_{t_0} \leftarrow \mathbf{x}_T, \mathbf{y}_{t_i} \leftarrow \frac{\alpha_{t_i}}{\alpha_{t_{i-1}}}\mathbf{y}_{t_{i-1}} - 2\alpha_{t_i}\int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-u}\boldsymbol{\epsilon}(\boldsymbol{u}_i, s_i)\mathrm{d}u - \sqrt{2}\alpha_{t_i}\int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-s}\mathrm{d}\bar{\boldsymbol{\omega}}(s),$$

with $s_i \leftarrow t_\lambda\left(\lambda_{t_{i-1}} + \frac{h_i}{2}\right)$ and

$$\boldsymbol{u}_i \leftarrow \frac{\alpha_{s_i}}{\alpha_{t_{i-1}}}\mathbf{y}_{t_{i-1}} - 2\sigma_{s_i}\left(e^{\frac{h_i}{2}} - 1\right)\boldsymbol{\epsilon}(\mathbf{y}_{t_{i-1}}, t_{i-1}) - \sqrt{2}\alpha_{s_i}\int_{\lambda_{t_{i-1}}}^{\lambda_{s_i}} e^{-s}\mathrm{d}\bar{\boldsymbol{\omega}}(s),$$

Then $\{\mathbf{y}_{t_i}\}_i$ has the same distribution as $\{\widetilde{\mathbf{x}}_{t_i}\}_i$ in Algorithm 3. We can compute the difference:

$$\mathbf{y}_{t_i} - \mathbf{x}_{t_i} = \frac{\alpha_{t_i}}{\alpha_{t_{i-1}}}(\mathbf{y}_{t_{i-1}} - \mathbf{x}_{t_{i-1}}) + \Gamma; \Gamma = \Gamma_1 + \Gamma_2$$

and:

$$\begin{aligned}\Gamma_1 &= -2\alpha_{t_i}\int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-u}\left[\boldsymbol{\epsilon}(\boldsymbol{u}_i, s_i) - \boldsymbol{\epsilon}(x_{s_i}, s_i)\right]\mathrm{d}u \\ \Gamma_2 &= -2\alpha_{t_i}\int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-u}\left[\boldsymbol{\epsilon}(x_{s_i}, s_i) - \widehat{\boldsymbol{\epsilon}}(\widehat{\mathbf{x}}_u, u)\right]\mathrm{d}u\end{aligned}$$

Similarly to the case $k = 1$, $\mathbb{E}[|\Gamma_2|] \leq Ch^2$. Note that $\mathbb{E}[|\Gamma_1|] \leq Ch\Delta_i$ with $\Delta_i = \mathbb{E}[|\boldsymbol{u}_i - \mathbf{x}_{s_i}|]$. Introduce: $U_i = \mathbb{E}[|\mathbf{y}_{t_i} - \mathbf{x}_{t_i}|]$. We will bound Δ_i by a function of U_{i-1}. For this, recall:

$$\boldsymbol{u}_i = \frac{\alpha_{s_i}}{\alpha_{t_{i-1}}}\mathbf{y}_{t_{i-1}} - 2\alpha_{s_i}\int_{\lambda_{t_{i-1}}}^{\lambda_{s_i}} e^{-\lambda}\boldsymbol{\epsilon}(\mathbf{y}_{t_{i-1}}, t_{i-1})\mathrm{d}\lambda - \sqrt{2}\alpha_{s_i}\int_{\lambda_{t_{i-1}}}^{\lambda_{s_i}} e^{-\lambda}\mathrm{d}\bar{\boldsymbol{\omega}}(\lambda),$$

and write the difference:

$$\boldsymbol{u}_i - \mathbf{x}_{s_i} = \frac{\alpha_{s_i}}{\alpha_{t_{i-1}}}(\mathbf{y}_{t_{i-1}} - \mathbf{x}_{t_{i-1}}) + 2\alpha_{t_i}\int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-u}(\widehat{\boldsymbol{\epsilon}}(\widehat{\mathbf{x}}_u, u) - \boldsymbol{\epsilon}_\theta(\mathbf{y}_{t_{i-1}}, \lambda_{t_{i-1}}))\mathrm{d}\lambda.$$

By the triangular inequality:

$$|\widehat{\boldsymbol{\epsilon}}(\widehat{\mathbf{x}}_u, u) - \boldsymbol{\epsilon}(\mathbf{y}_{t_{i-1}}, \lambda_{t_{i-1}})| \leq C(|\mathbf{x}_{t_\lambda(u)} - \mathbf{x}_{t_{i-1}}| + |\mathbf{x}_{t_{i-1}} - \mathbf{y}_{t_{i-1}}| + |t_\lambda(u) - t_{i-1}|),$$

and so: $\mathbb{E}[|\widehat{\boldsymbol{\epsilon}}(\widehat{\mathbf{x}}_u, u) - \boldsymbol{\epsilon}(\mathbf{y}_{t_{i-1}}, \lambda_{t_{i-1}})|] \leq C(\sqrt{h} + U_{i-1} + h)$. Finally:

$$\Delta_i \leq \frac{\alpha_{s_i}}{\alpha_{t_{i-1}}}U_{i-1} + Ch\left(\sqrt{h} + U_{i-1} + h\right) \leq U_{i-1} + Ch\left(\sqrt{h} + U_{i-1}\right),$$

and $\mathbb{E}[|\Gamma_1|] \leq Ch^{5/2} + C(h + h^2)U_{i-1}$. This gives the bound:

$$U_i \leq Ch^2 + (\frac{\alpha_{t_i}}{\alpha_{t_{i-1}}} + Ch + Ch^2)U_{i-1}.$$

Now using Theorem B.6, the proof can now be finished following as in proof in Section B.2.4.

B.3.2 Convergence of SEEDS-3:

Continuing with the same notations as before, we will prove by analogy that:

$$U_i \leq Ch^2 + (\frac{\alpha_{t_i}}{\alpha_{t_{i-1}}} + Ch + Ch^2 + Ch^3)U_{i-1}, \tag{49}$$

so that, by Theorem B.6, we obtain the desired result.

Let $\mathbf{y}_{t_i} = \widetilde{\mathbf{x}}_{t_i}$. We have:

$$y_{t_i} = \frac{\alpha_{t_i}}{\alpha_{t_{i-1}}}y_{t_{i-1}} - 2\alpha_{t_i}\int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-\lambda}\boldsymbol{\epsilon}(y_{t_{i-1}}, t_{i-1})d\lambda$$

$$-\frac{2\alpha_{t_i}}{h_i r_2}\int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-\lambda}(\lambda - \lambda_{t_{i-1}})(\boldsymbol{\epsilon}(u_{2i}, s_{2i}) - \boldsymbol{\epsilon}(y_{t_{i-1}}, t_{i-1}))d\lambda + \text{Noise},$$

and $x_{t_i} = \frac{\alpha_{t_i}}{\alpha_{t_{i-1}}}x_{t_{i-1}} - 2\alpha_{t_i}\int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-\lambda}\widehat{\boldsymbol{\epsilon}}(\widehat{\mathbf{x}}_\lambda, \lambda)\mathrm{d}\lambda + \text{Noise}$, where Noise is the same in both equations. From the proof for $k = 1$, it suffices to bound:

$$\Gamma = \frac{1}{h_i}\int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-\lambda}(\lambda - \lambda_{t_{i-1}})(\boldsymbol{\epsilon}(u_{2i}, s_{2i}) - \boldsymbol{\epsilon}(y_{t_{i-1}}, t_{i-1}))d\lambda,$$

in the L^1 norm. By the Lipschitz property of $\boldsymbol{\epsilon}$:

$$\begin{aligned}\mathbb{E}[|\Gamma|] &\leq Ch(\mathbb{E}[|u_{2i} - x_{s_{2i}}|] + \mathbb{E}[|x_{t_{i-1}} - x_{s_{2i}}|] + \mathbb{E}[|y_{t_{i-1}} - x_{t_{i-1}}|]) + Ch^2 \\ &\leq Ch(\mathbb{E}[|u_{2i} - x_{s_{2i}}|] + \sqrt{h} + U_{i-1}) + Ch^2.\end{aligned}$$

Now, let us bound $\mathbb{E}[|u_{2i} - x_{s_{2i}}|]$ and for this, write

$$u_{2i} = \frac{\alpha_{s_{2i}}}{\alpha_{t_{i-1}}} y_{t_{i-1}} - 2\alpha_{s_{2i}} \int_{\lambda_{t_{i-1}}}^{\lambda_{s_{2i}}} e^{-\lambda} \boldsymbol{\epsilon}(y_{t_{i-1}}, t_{i-1}) \mathrm{d}\lambda$$

$$-2\alpha_{s_{2i}} \int_{\lambda_{t_{i-1}}}^{\lambda_{s_{2i}}} e^{-\lambda} (\lambda - \lambda_{t_{i-1}}) \frac{\epsilon(u_{2i-1}, s_{2i-1}) - \epsilon(y_{t_{i-1}}, t_{i-1})}{r_1 h_i} \mathrm{d}\lambda + \text{Noise}$$

and

$$x_{s_{2i}} = \frac{\alpha_{s_{2i}}}{\alpha_{t_{i-1}}} x_{t_{i-1}} - 2\alpha_{s_{2i}} \int_{\lambda_{t_{i-1}}}^{\lambda_{s_{2i}}} e^{-\lambda} \widehat{\boldsymbol{\epsilon}}(\widehat{\mathbf{x}}_\lambda, \lambda) \mathrm{d}\lambda + \text{Noise},$$

where Noise is the same in both equations. So

$$\begin{aligned}u_{2i} - x_{s_{2i}} &= \frac{\alpha_{s_{2i}}}{\alpha_{t_{i-1}}} (y_{t_{i-1}} - x_{t_{i-1}}) - 2\alpha_{s_{2i}} \int_{\lambda_{t_{i-1}}}^{\lambda_{s_{2i}}} e^{-\lambda} (\boldsymbol{\epsilon}(\mathbf{y}_{t_{i-1}}, t_{i-1}) - \widehat{\boldsymbol{\epsilon}}(\widehat{\mathbf{x}}_\lambda, \lambda)) \mathrm{d}\lambda \\ &\quad -2\alpha_{s_{2i}} \int_{\lambda_{t_{i-1}}}^{\lambda_{s_{2i}}} e^{-\lambda} (\lambda - \lambda_{t_{i-1}}) D \mathrm{d}\lambda,\end{aligned}$$

where $D = \frac{\epsilon(u_{2i-1}, s_{2i-1}) - \epsilon(y_{t_{i-1}}, t_{i-1})}{r_1 h_i}$. From the convergence proof of SEEDS-1, and the Lipschitz property of ϵ, we obtain:

$$\mathbb{E}[|u_{2i} - x_{s_{2i}}|] \leq U_{i-1} + ChU_{i-1} + Ch^2 + Ch\mathbb{E}[|u_{2i-1} - y_{t_{i-1}}|].$$

Again, by the triangular inequality:

$$\mathbb{E}[|u_{2i-1} - y_{t_{i-1}}|] \leq \mathbb{E}[|u_{2i-1} - x_{s_{2i-1}}|] + C\sqrt{h} + U_{i-1},$$

and since

$$u_{2i-1} - x_{s_{2i-1}} = \frac{\alpha_{s_{2i-1}}}{\alpha_{t_{i-1}}} (y_{t_{i-1}} - x_{t_{i-1}}) - 2\alpha_{s_{2i-1}} \int_{\lambda_{t_{i-1}}}^{\lambda_{s_{2i-1}}} e^{-\lambda} (\widehat{\boldsymbol{\epsilon}}(\widehat{\mathbf{x}}_\lambda, \lambda) - \boldsymbol{\epsilon}(y_{t_{i-1}}, t_{i-1}) \mathrm{d}\lambda,$$

we have, as before,

$$\mathbb{E}[|u_{2i-1} - x_{s_{2i}}|] \leq U_{i-1} + ChU_{i-1} + Ch^2.$$

Combining the previous inequalities leads to (49). This finishes the proof of Corollary 4.3.

C Implementation Details

The SEEDS solvers used in our experiments are exactly the variants we described in Algorithms 2, 3, 4 in the main part of the paper. In particular, SEEDS-2 and SEEDS-3 solvers are completely determined by one-parameter families of deterministic exponential integrators from [10] of order 2 and 3 respectively and prescribed by the following Butcher tableaux:

$$\begin{array}{c|cc} 0 & & \\ c_2 & c_2\varphi_{1,2} & \\ \hline & \left(1 - \frac{1}{2c_2}\right)\varphi_1 & \frac{1}{2c_2}\varphi_1 \end{array} \qquad \begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ c_2 & c_2\varphi_{1,2} & 0 & 0 \\ \frac{2}{3} & \frac{2}{3}\varphi_{1,3} - \frac{4}{9c_2}\varphi_{2,3} & \frac{4}{9c_2}\varphi_{2,3} & 0 \\ \hline & \varphi_1 - \frac{3}{2}\varphi_2 & 0 & \frac{3}{2}\varphi_2 \end{array}$$

In all experiments we fix the parameter $c_2 = 0.5$ for SEEDS-2 and $c_2 = \frac{1}{3}$ for SEEDS-3. We point out that the Butcher tableau here associated to SEEDS-3 is the result of a weakening of the *stiff order conditions* so it might suffer from order reduction in the deterministic case. We also point out that a full theory of stiff order conditions for stochastic exponential Runge-Kutta methods for semi-linear DEs with a time-varying linear coefficient have not been yet developed to the authors knowledge. As such, this is only an analogy whose purpose is to clarify how our solvers relate to well-known solvers from the literature, but such Butcher tableaux do not rigorously reflect their convergence order.

C.1 Stabilization of the exponential terms

In the proposed algorithms, one subtle detail is to re-arrange all equations in order for them to make use only of the $\texttt{expm1}(h)$ function which computes $e^h - 1$ with great numerical stability, specially for small values of h. We have for values $r_1 = 1/3$, $r_2 = 2/3$, the following identity

$$\begin{aligned}
\sqrt{e^{2h_i} - e^{h_i}} z_i + \sqrt{e^{h_i} - 1} v_i &= \sqrt{e^{h_i} - 1} \left(\sqrt{e^{h_i}} z_i + v_i \right) \\
&= \sqrt{e^{h_i} - 1} \left(e^{\frac{h_i}{2}} z_i + v_i \right) \\
&= \sqrt{\texttt{expm1}(h_i)} \left(\left(\texttt{expm1}\left(\frac{h_i}{2} \right) + 1 \right) z_i + v_i \right) \\
&= \sqrt{\texttt{expm1}(h_i)} \left(\texttt{expm1}\left(\frac{h_i}{2} \right) + 1 \right) z_i + \sqrt{\texttt{expm1}(h_i)} v_i.
\end{aligned}$$

Now by using

$$e^{2(r_2 - r_1)h_i} = e^{2\left(\frac{2}{3} - \frac{1}{3}\right)h_i} = e^{2\frac{1}{3}h_i} = e^{r_2 h_i},$$

we now compute

$$\begin{aligned}
\sqrt{e^{2r_2 h_i} - e^{r_2 h_i}} z_i^1 + \sqrt{e^{r_2 h_i} - 1} z_i^2 &= \sqrt{e^{r_2 h_i} - 1} \left(\sqrt{e^{r_2 h_i}} z_i^1 + z_i^2 \right) \\
&= \sqrt{e^{r_2 h_i} - 1} \left(e^{\frac{h_i}{3}} z_i^1 + z_i^2 \right) \\
&= \sqrt{\texttt{expm1}(r_2 h_i)} ((\texttt{expm1}(r_1 h_i) + 1) z_i^1 + z_i^2)
\end{aligned}$$

$$\begin{aligned}
\sqrt{e^{2h_i} - e^{2r_2 h_i}} z_i^1 &= \sqrt{e^{3r_2 h_i} - e^{2r_2 h_i}} z_i^1 \\
&= e^{r_2 h_i} \sqrt{e^{r_2 h_i} - 1} z_i^1 \\
&= \sqrt{\texttt{expm1}(r_2 h_i)} (\texttt{expm1}(r_2 h_i) + 1) z_i^1.
\end{aligned}$$

Finally, we obtain

$$\begin{aligned}
&\sqrt{e^{2h_i} - e^{2r_2 h_i}} z_i^1 + \sqrt{e^{2r_2 h_i} - e^{r_2 h_i}} z_i^2 + \sqrt{e^{r_2 h_i} - 1} z_i^3 \\
= \; &\sqrt{\texttt{expm1}(r_2 h_i)} ((\texttt{expm1}(r_2 h_i) + 1) z_i^1 + (\texttt{expm1}(r_1 h_i) + 1) z_i^2 + z_i^3).
\end{aligned}$$

C.2 Noise schedules parameterizations

The inverses of $\lambda(t)$ are given in the (VP) linear and cosine schedules respectively by

$$\begin{aligned}
t_\lambda(\lambda) &= \frac{2 \log(e^{-2\lambda} + 1)}{\sqrt{\beta_{\min}^2 + 2(\beta_{\max} - \beta_{\min}) \log(e^{-2\lambda} + 1)} + \beta_{\min}}, \\
t_\lambda(\lambda) &= \frac{2(1+s)}{\pi} \arccos\left(\exp\Big(-\frac{1}{2} \log(e^{-2\lambda} + 1) + \log\cos\big(\frac{\pi s}{2(1+s)}\big) \Big) \right) - s.
\end{aligned}$$

In EDM Noise Prediction case, the inverses of $\lambda(t)$, as given in (37), with respect to the NPFO and NRSDE are respectively given by

$$t_\lambda(\lambda) = \sigma_d \tan(e^{-\lambda}) \quad \text{and} \quad t_\lambda(\lambda) = \frac{\sigma_d}{\sqrt{\frac{1}{\sigma_d^2 e^{-2\lambda}} - 1}}.$$

C.3 EDM discretization

We follow Karras et al. [12] to implement the EDM discretization timesteps $\{t_i\}_{i=0}^{M}$ as $t_i = \sigma^{-1}(\sigma_i)$ such that, for $\rho > 0$,

$$\sigma_{i<M} = \left(\sigma_{\max}^{\frac{1}{\rho}} + \frac{i}{N-1}\left(\sigma_{\min}^{\frac{1}{\rho}} - \sigma_{\max}^{\frac{1}{\rho}}\right)\right)^{\rho} \quad \text{and} \quad \sigma_N = 0.$$

From the definition, we note that $\sigma_0 = \sigma_{\max}$ and $\sigma_{M-1} = \sigma_{\min}$, where $\sigma_{\min}$ and $\sigma_{\max}$ denote the minimum and maximum noise magnitude, respectively. We also keep default value $\rho = 7$ as in [12]. However, we figured out that when using EDM discretization with linear schedule, the noise schedule improvement in iDDPM pre-trained models would result in two consecutive time steps of the same value, i.e. $t_j = t_{j+1}$ for some index $j = 1, \dots, M-1$ and large steps ($M \geq 61$). Thus, for SEEDS-3 and DPM-Solver-3 [15], the usage of the function $\varphi_2(h) = \dfrac{e^h - h - 1}{h^2}$ (see Appendix D for more details) will cause zero division error where $h = \lambda_{t_{j+1}} - \lambda_{t_j} = 0$. Therefore, in our implementation, we ignore the noise schedule improvement of iDDPM models when using solvers of order three.

C.4 Final sampling step

The sampling phase in DPMs using SEEDS follows the RSDE, which requires gradual computing through discretization time steps $\{t_i\}_{i=0}^{M}$ and the latter goes from $t_0 = T$ to $t_M = 0$. In our implementation, to avoid the logarithm of zero error at the last step, i.e. $\log \sigma_{t_M} = \log(0)$, we stop the sampling phase at step $(M-1)$. Hence, the NFE used in a run will be given as

$$\text{NFE} = k \times (M-1),$$

where k represents the order of the solver. We also do not use the "denoising" trick, i.e., ignoring the random noise at the last step and leave it to further research.

D Reminders on Stochastic Exponential Integrators

Let us consider a SDE of the form

$$\mathrm{d}\mathbf{x}(t) = [a(t)\mathbf{x}(t) + c(t)f(\mathbf{x}(t), t)]\mathrm{d}t + g(t)\mathrm{d}\boldsymbol{\omega}(t), \tag{50}$$

where $a, c : [0, T] \to \mathbb{R}$ and $g : [0, T] \to \mathbb{R}^{d\times d}$. In other words we concentrate to high-dimensional semi-linear non autonomous SDEs with additive noise. The objective of this section is to construct explicit stochastic exponential derivative-free methods for the above equation following the Runge-Kutta (RK) approach. These methods ideally should fulfill the following properties:

1. If $f \equiv 0$, then (50) can be solved *exactly*;
2. If $g \equiv 0$, then a SEEDS method for (50) identifies with an exponential RK (ERK) method;
3. If $a \equiv 0$, then a SEEDS method for (50) identifies with a stochastic RK (SRK) method and if moreover $g \equiv 0$ then it identifies with a classical RK ODE method.

Before tackling the aimed SEEDS problem let us rapidly recall elementary constructions of RK, ERK, weak and strong SRK methods. We will not deal with time-adaptive methods here.

D.1 Derivative-free exponential ODE schemes

D.1.1 Runge-Kutta approach

Derivative-free schemes are obtained by comparing the Itô-Taylor expansion of the above paragraph with expressions of $\mathbf{x}(t)$ in terms of its intermediate evaluations between s and $t = s + h$ and the Taylor expansions of such evaluations. As a simple example, in the ODE regime

$$\mathrm{d}\mathbf{x}(t) = f(\mathbf{x}(t), t)\mathrm{d}t. \tag{51}$$

As such, analytic solutions to the above equations are of the form

$$\mathbf{x}(t) \quad = \quad \mathbf{x}(s) + \int_s^t f(\mathbf{x}(\tau), \tau) \mathrm{d}\tau.$$

Now, Taylor expansion gives, up to order 2:

$$\begin{aligned}
\mathbf{x}(t) &= \mathbf{x}(s) + h\mathbf{x}'(s) + \frac{h^2}{2}\mathbf{x}''(s) + \mathcal{O}(h^3) \\
&= \mathbf{x}(s) + hf(\mathbf{x}(s), s) + \frac{h^2}{2}(\mathbf{x}'(s)\partial_{\mathbf{x}} f(\mathbf{x}(s), s) + \partial_t f(\mathbf{x}(s), s)) + \mathcal{O}(h^3) \\
&= \mathbf{x}(s) + hf(\mathbf{x}(s), s) + \frac{h^2}{2}(f(\mathbf{x}(s), s)\partial_{\mathbf{x}} f(\mathbf{x}(s), s) + \partial_t f(\mathbf{x}(s), s)) + \mathcal{O}(h^3).
\end{aligned}$$

A straightforward recursion yields a Taylor expansion

$$\mathbf{x}(t) \quad = \quad \mathbf{x}(s) + \sum_{k=1}^{n} \frac{h^k}{k!} L_t^{k-1} f(\mathbf{x}(t), t) + \int_s^t \cdots \int_s^\tau L_t^n f(\mathbf{x}(\tau), \tau) \mathrm{d}\tau^{n+1},$$

where we denote the generalized infinitesimal operator of the solution $\mathbf{x}$ of (51) by

$$L_t(\bullet) = \partial_t(\bullet) + f(\mathbf{x}(t), t) \cdot \partial_{\mathbf{x}}(\bullet).$$

Derivative-free methods seek to get rid of the derivatives in L_t by computing Taylor expansions of f at well-chosen points. In the explicit one-step case, this amounts of defining

$$\widehat{\mathbf{x}}(t) = \widehat{\mathbf{x}}(s) + h \sum_{k=1}^{n} \Phi(\widehat{\mathbf{x}}(s), s, h),$$

where the Φ function does not contain derivatives of f. The general high order case is given by well-tuned coefficients in the following scheme

$$\begin{aligned}
\widehat{\mathbf{x}}(t) &:= \widehat{\mathbf{x}}(s) + h \sum_{i=1}^{n} \alpha_i f(\mathbf{x}_i, s + c_i h) \\
\mathbf{x}_i &= \widehat{\mathbf{x}}(s) + h \sum_{j=1}^{n} a_{i,j} f(\mathbf{x}_j, s + c_j h).
\end{aligned}$$

By denoting $\alpha = [\alpha_1 \dots \alpha_n]$, $c = [c_1 \dots c_n]^T$ and $A = (a_{i,j})$, these can be represented by a Butcher tableau of the form

$$\begin{array}{c|c} c & A \\ \hline & \alpha \end{array}$$

In the following sections we will present extended version of this tableau for representing more involved numerical schemes. As such, the Euler, midpoint, Heun and general second order explicit methods are respectively:

$$\begin{array}{c|c} 0 & 0 \\ \hline & 1 \end{array} \qquad \begin{array}{c|cc} 0 & 0 & 0 \\ 1/2 & 1/2 & 0 \\ \hline & 0 & 1 \end{array} \qquad \begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1 & 0 \\ \hline & 1/2 & 1/2 \end{array} \qquad \begin{array}{c|cc} 0 & 0 & 0 \\ c_2 & c_2 & 0 \\ \hline & 1 - \frac{1}{2c_2} & \frac{1}{2c_2} \end{array}$$

D.1.2 Exponential Runge-Kutta approach

We now concentrate on a non-autonomous semi-linear ODE of the form

$$\mathrm{d}\mathbf{x}(t) = [f(t)\mathbf{x}(t) + g(\mathbf{x}(t), t)]\mathrm{d}t, \tag{52}$$

where $f(t) = \bar{f}(t) \cdot \mathrm{Id}_d$ and $g(\mathbf{x}(t), t) = \bar{g}(\mathbf{x}(t), t) \cdot \mathrm{Id}_d$ are diagonal time-dependent matrices with identical diagonal coefficients. In particular, we have $[f^{(k)}(t), f^{(l)}(s)] = 0$, a property we will need to facilitate exponential matrix multiplications. As $f(s)$ is a d-dimensional diagonal matrix, the

fundamental matrix for (52), usually given by the Peano-Baker series, simplifies in this case to the form

$$\Phi(t;s) := e^{\int_s^t f(r)\mathrm{d}r} = \sum_{n=0}^{\infty} \frac{1}{n!} \left(\int_s^t f(r)\mathrm{d}r \right)^n ,$$

which satisfies $\Phi'(t;s) = f(t)\Phi(t;s)$ and $\Phi(s;s) = 1$. Then the exact solution for (52) is given, via the variation of constants formula, by

$$\mathbf{x}(t) \quad = \quad \Phi(t;s) \left(\mathbf{x}(s) + \int_s^t \Phi^{-1}(\tau;s) g(\mathbf{x}(\tau),\tau)\mathrm{d}\tau \right) = \Phi(t;s)\mathbf{x}(s) + \int_s^t \Phi(t;\tau) g(\mathbf{x}(\tau),\tau)\mathrm{d}\tau.$$

In light of this integral form, one can formalize a general class of

exponential n-stage RK methods

$$\begin{aligned} \mathbf{x}_i &= \gamma_i(h,f_s)\mathbf{x}(s) + \sum_{j=1}^{n} a_{ij}(h,f_s) g(\mathbf{x}_j, s + c_j h) \\ \mathbf{x}(t) &= \gamma_0(h,f_s)\mathbf{x}(s) + \sum_{i=1}^{n} b_i(h,f_s) g(\mathbf{x}_i, s + c_i h), \end{aligned}$$

where $\gamma_0, \gamma_i, a_{ij}, b_j$ are functions of the step-size h, $f$and $f_s(r) := \int_0^r f(s+\tau)\mathrm{d}\tau$. There are two possible approaches to create exponential integrators, namely the exponential time-differencing (ETD) approach that uses the variation of constants formula and makes use of the φ functions, and the Lawson approach, also know as integrating factor (IF), which makes a change of variables on the above SDE thus avoiding the use of the φ functions but computing exponential factors step-wise.

The Lawson and the ETD approaches for Exponential Euler schemes

There are two choices one can make when computing first order approximations of

$$\mathbf{x}(t) \quad = \quad \Phi(t;s)\mathbf{x}(s) + \int_s^t \Phi(t;\tau) g(\mathbf{x}(\tau),\tau)\mathrm{d}\tau.$$

First, by interpolating $g(\mathbf{x}(\tau),\tau)$ as $g(\mathbf{x}(s),s)$ we obtain

$$\mathbf{x}(t) \quad = \quad \Phi(t;s)\mathbf{x}(s) + g(\mathbf{x}(s),s) \int_s^t \Phi(t;\tau)\mathrm{d}\tau.$$

Now two choices remain. Either $\int_s^t \Phi(t;\tau)\mathrm{d}\tau$ is computed exactly or we again interpolate $\Phi(t;\tau)$ as $\Phi(t;s)$. Taking for simplicity $f \equiv A$ to be constant and by denoting $h = t - s$, the first case yields the ETD Euler method

$$\begin{aligned} \widehat{\mathbf{x}}(t) &= \Phi(t;s)\widehat{\mathbf{x}}(s) + h\gamma_1(f,t,s) g(\widehat{\mathbf{x}}(s),s) \\ &= e^{Ah}\widehat{\mathbf{x}}(s) + h\varphi_1(Ah) g(\widehat{\mathbf{x}}(s),s) \\ &= \widehat{\mathbf{x}}(s) + h\varphi_1(Ah)[g(\widehat{\mathbf{x}}(s),s) - \widehat{\mathbf{x}}(s)], \end{aligned}$$

and the second yields the IF Euler (also called Lawson-Euler) method

$$\begin{aligned} \widehat{\mathbf{x}}(t) &= \Phi(t;s)\widehat{\mathbf{x}}(s) + h\Phi(t;s) g(\widehat{\mathbf{x}}(s),s) \\ &= e^{Ah}(\widehat{\mathbf{x}}(s) + h g(\widehat{\mathbf{x}}(s),s)). \end{aligned}$$

Now, consider a solution

$$\mathbf{x}(t) \quad = \quad \Phi(t;s)\mathbf{x}(s) + \int_0^h \Phi(h;\tau) g(\mathbf{x}(s+\tau), s+\tau)\mathrm{d}\tau.$$

Exponential methods then aim to approximate the term $g(\mathbf{x}(s+\tau), s+\tau)$ by its interpolation polynomial in certain non-confluent quadrature nodes $c_1, \ldots, c_n$.

The ETD approach

In this case, the variation of constants formula yields

$$\begin{aligned}\mathbf{x}(s+ch) &= e^{\int_s^{s+ch} a(\tau)\mathrm{d}\tau}\mathbf{x}(s) + \int_s^{s+ch} e^{\int_\tau^{s+ch} a(r)\mathrm{d}r} f(\mathbf{x}(\tau),\tau)\mathrm{d}\tau. \\ &= e^{\int_0^{ch} a(s+\tau)\mathrm{d}\tau}\mathbf{x}(t) + \int_0^{ch} e^{\int_{\tau-s}^{ch} a(s+r)\mathrm{d}r} f(\mathbf{x}(s+\tau),s+\tau)\mathrm{d}\tau.\end{aligned}$$

Now, as before, the Taylor expansion of f yields

$$\begin{aligned}f(\mathbf{x}(s+\tau),s+\tau) &= \sum_{j=1}^{q}\frac{\tau^{j-1}}{(j-1)!}f^{(j-1)}(\mathbf{x}(s),s) \\ &\quad + \int_0^\tau \frac{(\tau-\tau_1)^{q-1}}{(q-1)!}f^{(q)}(\mathbf{x}(s+\tau_1),s+\tau_1)\mathrm{d}\tau_1.\end{aligned}$$

Recall that the φ functions are given in an integral form as follows

$$\varphi_{k+1}(t) = \int_0^1 e^{(1-\delta)t}\frac{\delta^k}{k!}\mathrm{d}\delta,$$

which satisfy $\varphi_k(0) = \frac{1}{k!}$. Now denote

$$\begin{aligned}\varphi_j(t,a) &:= \frac{1}{t^j}\int_0^t e^{\int_\tau^t a(r)\mathrm{d}r}\frac{\tau^{j-1}}{(j-1)!}\mathrm{d}\tau, \qquad j \geqslant 1. \\ \varphi_1(h,a) &= \frac{1}{h}\int_0^h e^{\int_\tau^h a(r)\mathrm{d}r}\mathrm{d}\tau = \int_0^1 e^{h\int_\theta^1 a(r)\mathrm{d}r}\mathrm{d}\theta.\end{aligned}$$

The exact solution at $s+ch$ now reads

$$\begin{aligned}\mathbf{x}(s+ch) &= e^{\int_0^{ch} a(s+\tau)\mathrm{d}\tau}\mathbf{x}(s) + \sum_{j=1}^{q}(ch)^j\varphi_j(ch,a)f^{(j-1)}(\mathbf{x}(s),s) \\ &\quad + \int_0^{ch} e^{\int_\tau^{ch} a(s+r)\mathrm{d}r}\left(\int_0^\tau \frac{(\tau-\tau_1)^{q-1}}{(q-1)!}f^{(q)}(\mathbf{x}(s+\tau_1),s+\tau_1)\mathrm{d}s\right)\mathrm{d}\tau_1.\end{aligned}$$

Using the left endpoint rule yields

$$\begin{aligned}\mathbf{x}(t) &= e^{\int_0^h a(s+\tau)\mathrm{d}\tau}\mathbf{x}(s) + f(\mathbf{x}(s),s)\int_0^h e^{\int_\tau^h a(s+r)\mathrm{d}r}\mathrm{d}\tau + \mathcal{O}(h^2) \\ &= e^{\int_0^h a(s+\tau)\mathrm{d}\tau}\mathbf{x}(s) + h\varphi_1(h,a)f(\mathbf{x}(s),s) + \mathcal{O}(h^2).\end{aligned}$$

Second order examples:

The condition $b_2(z)c_2 = \varphi_2(z)$ implies $b_2 = \varphi_2(z)/c_2$ and we obtain

$$\begin{aligned}\widehat{\mathbf{x}} &= e^{chA}\mathbf{x}(s) + ch\varphi_1(chA)f(\mathbf{x}(s),s) \\ \mathbf{x}(t) &= e^{hA}\mathbf{x}(s) + h\left(\varphi_1(hA) - \frac{1}{c}\varphi_2(hA)\right)f(\mathbf{x}(s),s) + \frac{h}{c}\varphi_2(hA)f(\widehat{\mathbf{x}},s+ch). \quad (53)\end{aligned}$$

A second method is obtained by weakening the above condition to $b_2(0)c_2 = \varphi_2(0) = 1/2$ giving

$$\begin{aligned}\widehat{\mathbf{x}} &= e^{chA}\mathbf{x}(s) + ch\varphi_1(chA)f(\mathbf{x}(s),s) \\ \mathbf{x}(t) &= e^{hA}\mathbf{x}(s) + h\varphi_1(hA)\left(1 - \frac{1}{2c}\right)f(\mathbf{x}(s),s) + \frac{h}{2c}\varphi_1(hA)f(\widehat{\mathbf{x}},s+ch). \quad (54)\end{aligned}$$

In some cases this method can suffer from order reduction and not reach order 2 of convergence.
Moreover, setting $c = 1/2$ gives the exponential midpoint method:

$$\begin{aligned}\widehat{\mathbf{x}} &= e^{chA}\mathbf{x}(s) + ch\varphi_1(chA)f(\mathbf{x}(s),s) \\ \mathbf{x}(t) &= e^{hA}\mathbf{x}(s) + h\varphi_1(hA)f(\widehat{\mathbf{x}},s+ch). \quad (55)\end{aligned}$$

The above order 1 and 2 exponential methods are represented by the following *exponential* Butcher tableaux

$$\begin{array}{c|c} 0 & 0 \\ \hline & \varphi_1 \end{array} \qquad \begin{array}{c|cc} 0 & & \\ c_2 & c_2\varphi_{1,2} & \\ \hline & \varphi_1 - \frac{1}{c_2}\varphi_2 & \frac{1}{c_2}\varphi_2 \end{array} \qquad \begin{array}{c|cc} 0 & & \\ c_2 & c_2\varphi_{1,2} & \\ \hline & \left(1 - \frac{1}{2c_2}\right)\varphi_1 & \frac{1}{2c_2}\varphi_1 \end{array} \qquad \begin{array}{c|cc} 0 & & \\ c_2 & c_2\varphi_{1,2} & \\ \hline & 0 & \varphi_1 \end{array}$$

We check that when formally setting $A = 0$ then the first two methods are identical and give the generic 2nd order RK method, the choice $c = 1$ gives the Heun method and the choice $c = 1/2$ gives the midpoint method.

Fourth order methods

It can be shown that ERK methods need at least 5 stages to achieve order 4. By setting formally $A = 0$ these methods do not have a non-exponential counterpart to our knowledge.

5-stage sequential We have a fourth-order ERK scheme given by

$$\begin{array}{c|ccccc} 0 & & & & & \\ \frac{1}{2} & \frac{1}{2}\varphi_{1,2} & & & & \\ \frac{1}{2} & \frac{1}{2}\varphi_{1,3} - \varphi_{2,3} & \varphi_{2,3} & & & \\ 1 & \varphi_{1,4} - 2\varphi_{2,4} & \varphi_{2,4} & \varphi_{2,4} & & \\ \frac{1}{2} & \frac{1}{2}\varphi_{1,5} - 2a_{5,2} - a_{5,4} & a_{5,2} & a_{5,2} & \frac{1}{4}\varphi_{2,5} - a_{5,2} & \\ \hline & \varphi_1 - 3\varphi_2 + 4\varphi_3 & 0 & 0 & -\varphi_2 + 4\varphi_3 & 4\varphi_2 - 8\varphi_3 \end{array}$$

with

$$a_{5,2} = c_5\varphi_{2,5} - \varphi_{3,4} + c_5^2\varphi_{2,4} - c_5\varphi_{3,5}.$$

This can also be represented by Rosenbrock-like Butcher tableau

$$\begin{array}{c|ccccc} 0 & & & & & \\ \frac{1}{2} & \frac{1}{2}\varphi_{1,2} & & & & \\ \frac{1}{2} & \frac{1}{2}\varphi_{1,3} & \varphi_{2,3} & & & \\ 1 & \varphi_1 & \varphi_2 & \varphi_2 & & \\ \frac{1}{2} & \frac{1}{2}\varphi_{1,5} & a_{5,2} & a_{5,2} & \frac{1}{4}\varphi_{2,5} - a_{5,2} & \\ \hline & \varphi_1 & 0 & 0 & 4\varphi_3 - \varphi_2 & 4\varphi_2 - 8\varphi_3 \end{array}$$

with

$$a_{5,2} = c_5\varphi_{2,5} - c_3\varphi_{3,3} + c_5^2\varphi_2 - \varphi_3.$$

This traduces into (recall that $D_i = f(\mathbf{x}_i, s + c_i h) - f(\mathbf{x}(s), s)$)

$$\begin{aligned} \mathbf{x}(t) &= \mathbf{x}(s) + h(\varphi_1(hA))F(\mathbf{x}(s), s) + h(4\varphi_3(hA) - \varphi_2(hA))D_4 \\ &\quad + h(4\varphi_2(hA) - 8\varphi_3(hA))D_5 \\ \mathbf{x}_2 &= \mathbf{x}(s) + c_2 h\varphi_1(c_2 hA)F(\mathbf{x}(s), s) \\ \mathbf{x}_3 &= \mathbf{x}(s) + hc_3\varphi_1(c_3 hA)F(\mathbf{x}(s), s) + h\varphi_2(c_3 hA)D_2 \\ \mathbf{x}_4 &= \mathbf{x}(s) + h\varphi_1(hA)F(\mathbf{x}(s), s) + h\varphi_2(hA)(D_2 + D_3) \\ \mathbf{x}_5 &= \mathbf{x}(s) + h\varphi_1(hA)F(\mathbf{x}(s), s) + ha_{5,2}(hA)(D_2 + D_3) + h(c_5^2\varphi_2(c_5 hA) - a_{5,2}(hA))D_4. \end{aligned}$$

Inspired by this, we deduce the following fourth order Algorithm 5 specialized for DPMs in the VP noise prediction regime. We highlight the fact that SEEDS can produce any deterministic exponential method as in the DPM-Solver approach via the specification of a Butcher tableau this way and most of all follow-ups which have been done in [16] can readily be applied for SEEDS such as parallel steps schemes and dynamic thresholding. We leave these directions for future work.

D.2 Derivative-free exponential SDE schemes

Let $\mathbf{x}_t$ be the path at the continuous limit $h \to 0$, and $\{\widehat{\mathbf{x}}_t\}_{t_0}^{t_M}$ be the discretized numerical path, computed by a numerical scheme $\mathcal{S}$ with $M = 1/h$ steps of length $h > 0$. Then, $\mathcal{S}$ has

1. *strong order of convergence* γ if there is $K > 0$ such that

$$\mathbb{E}[|\mathbf{x}_{t_M} - \widehat{\mathbf{x}}_{t_M}|] \leq Kh^\gamma, \tag{56}$$

Algorithm 5 DPM-Solver-4

def DPM-Solver-4($\boldsymbol{\epsilon}_\theta, \tilde{\mathbf{x}}_{t_{i-1}}, t_{i-1}, t_i, r = 0.5$):
  $h_i \leftarrow \lambda_{t_i} - \lambda_{t_{i-1}}$
  $s_2 \leftarrow t_\lambda\left(\lambda_{t_{i-1}} + rh_i\right), \quad s_3 \leftarrow t_\lambda\left(\lambda_{t_{i-1}} + rh_i\right)$
  $s_4 \leftarrow t_\lambda\left(\lambda_{t_{i-1}} + h_i\right), \quad s_5 \leftarrow t_\lambda\left(\lambda_{t_{i-1}} + rh_i\right)$
  $\boldsymbol{k}_1 \leftarrow \boldsymbol{\epsilon}_\theta(\tilde{\mathbf{x}}_{t_{i-1}}, t_{i-1})$
  $\boldsymbol{k}_2 \leftarrow \frac{\alpha_{s_2}}{\alpha_s}\tilde{\mathbf{x}}_{t_{i-1}} - \sigma_{s_2}(e^{rh} - 1)\boldsymbol{k}_1$
  $\boldsymbol{k}_3 \leftarrow \frac{\alpha_{s_3}}{\alpha_s}\tilde{\mathbf{x}}_{t_{i-1}} - \sigma_{s_3}(e^{rh} - 1)\boldsymbol{k}_1 - \sigma_{s_3}\left(4\frac{e^{rh}-1}{h} - 2\right)[\epsilon_\theta(\boldsymbol{k}_2, s_2) - \boldsymbol{k}_1]$
  $\boldsymbol{k}_4 \leftarrow \frac{\alpha_{s_4}}{\alpha_s}\tilde{\mathbf{x}}_{t_{i-1}} - \sigma_{s_4}(e^{h} - 1)\boldsymbol{k}_1 - \sigma_{s_4}\left(\frac{e^{h}-1}{h} - 1\right)[\epsilon_\theta(\boldsymbol{k}_3, s_3) + \epsilon_\theta(\boldsymbol{k}_2, s_2) - 2\boldsymbol{k}_1]$
    $A = \sigma_{s_5}(e^{rh} - 1)\boldsymbol{k}_1 - \frac{1}{4}\sigma_{s_5}\left(\frac{e^{h}-1}{h} - 1\right)[\boldsymbol{k}_1 + \epsilon_\theta(\boldsymbol{k}_2, s_2) + \epsilon_\theta(\boldsymbol{k}_3, s_3)]$
    $B = \sigma_{s_5}\left(\frac{e^{rh}-1}{h} - \frac{1}{2}\right)[\boldsymbol{k}_1 + 4\epsilon_\theta(\boldsymbol{k}_2, s_2) + 4\epsilon_\theta(\boldsymbol{k}_3, s_3) - \epsilon_\theta(\boldsymbol{k}_4, s_4)]$
    $C = \sigma_{s_5}\left(\frac{e^{h}-1+4(e^{rh}-1)-3h}{h^2} - 1\right)[-\boldsymbol{k}_1 - \epsilon_\theta(\boldsymbol{k}_2, s_2) - \epsilon_\theta(\boldsymbol{k}_3, s_3) + \epsilon_\theta(\boldsymbol{k}_4, s_4)]$
  $\boldsymbol{k}_5 \leftarrow \frac{\alpha_{s_5}}{\alpha_{t_{i-1}}}\tilde{\mathbf{x}}_{t_{i-1}} - A - B - C$
    $D = \sigma_t(e^{h} - 1)\boldsymbol{k}_1 - \sigma_t\left(\frac{e^{h}-1}{h} - 1\right)[4\epsilon_\theta(\boldsymbol{k}_5, s_5) - \epsilon_\theta(\boldsymbol{k}_4, s_4) - 3\boldsymbol{k}_1]$
    $E = \sigma_t\left(4\frac{e^{h}-1-h}{h^2} - 2\right)[\boldsymbol{k}_1 + \epsilon_\theta(\boldsymbol{k}_4, s_4) - 2\epsilon_\theta(\boldsymbol{k}_5, s_5)]$
  $\tilde{\mathbf{x}}_{t_i} \leftarrow \frac{\alpha_t}{\alpha_{t_{i-1}}}\tilde{\mathbf{x}}_{t_{i-1}} - D - E$
  Return $\mathbf{x}_{t_i}$

2. *weak order of convergence* β if there is $K > 0$ and a function class $\mathcal{K}$ such that

$$|\mathbb{E}[\phi(\mathbf{x}_{t_M})] - \mathbb{E}[\phi(\widehat{\mathbf{x}}_{t_M})]| \leq Kh^\beta, \qquad \forall \phi(\cdot) \in \mathcal{K}. \tag{57}$$

Strong convergence is concerned with the precision of the path, while the weak convergence is with the precision of the moments. As, for diffusion models, the center of attention is the evolution of the probability densities rather than that of the noising process of single data samples, weak convergence is enough to guarantee the well-conditioning of our numerical schemes. Moreover, when the diffusion coefficient vanishes, then both strong and weak convergence (with the choice $\phi = \mathrm{id}$) reduce to the usual deterministic convergence criterion for ODEs.

D.2.1 Strong and Weak Stochastic Runge-Kutta approach

In all what follows ω is considered a d-dimensional Wiener process (with identity diffusion matrix).

Consider the following SDE

$$\mathrm{d}\mathbf{x}(t) = f(\mathbf{x}(t), t)\mathrm{d}t + g(t)\mathrm{d}\omega(t),$$

where $g(t) = \hat{g}(t) \cdot \mathrm{Id}_d$ is considered here as a diagonal matrix with identical diagonal entries $\hat{g}(t)$. Given an initial value independent of ω, the integral form of $\mathbf{x}(t)$ is given by

$$\mathbf{x}(t) \quad = \quad \mathbf{x}(s) + \int_s^t f(\mathbf{x}(\tau), \tau)d\tau + \int_s^t g(\tau)\mathrm{d}\omega(\tau).$$

The idea underlying stochastic numerical schemes is very similar to the one used in the deterministic approach, that is to take expansions of the terms inside the integrals based at the integral's initial value and replace the obtained derivatives that appear by interpolated approximations. A key difference here is that as we have to consider Itô-Taylor expansions, the infinitesimal operators are different but most importantly most of the stochastic iterated integrals will need to be approximated in an appropriate sense, whenever that is possible. We will develop this expansion up to triple integrals.

D.2.2 Truncated Itô-Taylor expansions

Applying Itô formula to $h = f$ or g yields

$$\begin{aligned} h(\mathbf{x}(t),t) &= h(\mathbf{x}(s),s) + \int_s^t g(\tau)\cdot\partial_{\mathbf{x}} h(\mathbf{x}(\tau),\tau)\mathrm{d}\omega(\tau) \\ &\quad + \int_s^t \left(\partial_t h(\mathbf{x}(\tau),\tau) + f(\mathbf{x}(\tau),\tau)\cdot\partial_{\mathbf{x}} h(\mathbf{x}(\tau),\tau) + \frac{g^2(\tau)}{2}\partial^2_{\mathbf{x}^2} h(\mathbf{x}(\tau),\tau)\right)\mathrm{d}\tau. \end{aligned}$$

This allows us to define two differential operators L, G as

$$L_t = \partial_t + f(\mathbf{x}(t),t)\cdot\partial_{\mathbf{x}} + \frac{g^2(t)}{2}\cdot\partial^2_{\mathbf{x}} \tag{58}$$

$$G_t = g(t)\cdot\partial_{\mathbf{x}}. \tag{59}$$

In particular $L_t g(t) = \partial_t g(t)$ and $G_t g(t) = 0$. With this notation, we have

$$h(\mathbf{x}(t),t) = h(\mathbf{x}(s),s) + \int_s^t L_t h(\mathbf{x}(\tau),\tau)\mathrm{d}\tau + \int_s^t G_t h(\mathbf{x}(\tau),\tau)\mathrm{d}\omega(\tau),$$

so our solution reads

$$\begin{aligned} \mathbf{x}(t) &= \mathbf{x}(s) + \int_s^t f(\mathbf{x}(\tau_1),\tau_1)\mathrm{d}\tau_1 + \int_s^t g(\tau_1)\mathrm{d}\omega(\tau_1) \\ &= \mathbf{x}(s) + \int_s^t \left(f(\mathbf{x}(t),t) + \int_s^{\tau_1} L_t f(\mathbf{x}(\tau_2),\tau_2)\mathrm{d}\tau_2 + \int_s^{\tau_1} G_t f(\mathbf{x}(\tau_2),\tau_2)d\omega(\tau_2)\right)\mathrm{d}\tau_1 \\ &\quad + \int_s^t \left(g(t) + \int_s^{\tau_1} L_t g(\tau_2)\mathrm{d}\tau_2 + \int_s^{\tau} G_t g(\tau_2)\mathrm{d}\omega(\tau_2)\right)\mathrm{d}\omega(\tau_1) \\ &= \mathbf{x}(s) + f(\mathbf{x}(s),s)h + g(t)(\omega(t)-\omega(s)) + R_1. \end{aligned}$$

Now $G_t g(\tau_2) = 0$ and

$$g(s) + \int_s^{\tau_1} L_t g(\tau_2)\mathrm{d}\tau_2 = g(s) + \int_s^{\tau_1} \partial_t g(\tau_2)\mathrm{d}\tau_2 = g(s) + g(\tau_1) - g(s) = g(\tau_1).$$

So we have

$$\int_s^t \left(g(s) + \int_s^{\tau_1} L_t g(\tau_2)\mathrm{d}\tau_2 + \int_s^{\tau} G_t g(\tau_2)\mathrm{d}\omega(\tau_2)\right)\mathrm{d}\omega(\tau_1) = g(s)(\omega(t)-\omega(s)),$$

and now our solution reads

$$\mathbf{x}(t) = \mathbf{x}(s) + f(\mathbf{x}(s),s)h + g(s)(\omega(t)-\omega(s)) + R_1,$$

where

$$R_1 = \int_s^t \int_s^{\tau_1} L_t f(\mathbf{x}(\tau_2),\tau_2)\mathrm{d}\tau_2\mathrm{d}\tau_1 + \int_s^t \int_s^{\tau_1} G_t f(\mathbf{x}(\tau_2),\tau_2)\mathrm{d}\omega(\tau_2)\mathrm{d}\tau_1.$$

Now we have

$$\begin{aligned} & L_t f(\mathbf{x}(t),t) \\ = \; & L_t f(\mathbf{x}(s),s) + \int_s^t g(\mathbf{x}(\tau),\tau)\cdot\partial_{\mathbf{x}} L_t f(\mathbf{x}(\tau),\tau)\mathrm{d}\omega(\tau) \\ & + \int_s^t \left(\partial_t L_t f(\mathbf{x}(\tau),\tau) + f(\mathbf{x}(\tau),\tau)\cdot\partial_{\mathbf{x}} L_t f(\mathbf{x}(\tau),\tau) + \frac{g^2(\mathbf{x}(\tau),\tau)}{2}\partial^2_{\mathbf{x}^2} L_t f(\mathbf{x}(\tau),\tau)\right)\mathrm{d}\tau \\ = \; & L_t f(\mathbf{x}(s),s) + \int_s^t L_t^2 f(\mathbf{x}(\tau),\tau)\mathrm{d}\tau + \int_s^t G_t L_t f(\mathbf{x}(\tau),\tau)\mathrm{d}\omega(\tau), \end{aligned}$$

and

$$\begin{aligned}
& G_t f(\mathbf{x}(t), t) \\
= \; & G_t f(\mathbf{x}(s), s) + \int_s^t g(\mathbf{x}(\tau), \tau) \cdot \partial_{\mathbf{x}} G_t f(\mathbf{x}(\tau), \tau) \mathrm{d}\omega(\tau) \\
& + \int_s^t \left(\partial_t G_t f(\mathbf{x}(\tau), \tau) + f(\mathbf{x}(\tau), \tau) \cdot \partial_{\mathbf{x}} G_t f(\mathbf{x}(\tau), \tau) + \frac{g^2(\mathbf{x}(\tau), \tau)}{2} \partial^2_{\mathbf{x}^2} G_t f(\mathbf{x}(\tau), \tau) \right) \mathrm{d}\tau \\
= \; & G_t f(\mathbf{x}(s), s) + \int_s^t L_t G_t f(\mathbf{x}(\tau), \tau) \mathrm{d}\tau + \int_s^t G_t^2 f(\mathbf{x}(\tau), \tau) \mathrm{d}\omega(\tau).
\end{aligned}$$

Now, if we denote $\mathrm{d}\omega^0(\tau) = \mathrm{d}\tau$, $\mathrm{d}\omega^1(\tau) = \mathrm{d}\omega(\tau)$ and

$$I_{(i)} = \int_s^t \mathrm{d}\omega^i(\tau_1) \qquad I_{(i,j)} = \int_s^t \int_s^{\tau_1} \mathrm{d}\omega^j(\tau_2) \mathrm{d}\omega^i(\tau_1) \qquad i, j \in \{0, 1\},$$

applying the same procedure to R_1 leads to

$$\begin{aligned}
& R_1 \\
= \; & \int_s^t \int_s^{\tau_1} L_t f(\mathbf{x}(\tau_2), \tau_2) \mathrm{d}\tau_2 \mathrm{d}\tau_1 + \int_s^t \int_s^{\tau_1} G_t f(\mathbf{x}(\tau_2), \tau_2) \mathrm{d}\omega(\tau_2) \mathrm{d}\tau_1 \\
= \; & \int_s^t \int_s^{\tau_1} \left[L_t f(\mathbf{x}(t), t) + \int_s^{\tau_2} L_t^2 f(\mathbf{x}(\tau_3), \tau_3) \mathrm{d}\tau_3 + \int_s^{\tau_2} G_t L_t f(\mathbf{x}(\tau_3), \tau_3) \mathrm{d}\omega(\tau_3) \right] \mathrm{d}\tau_2 \tau_1 \\
& + \int_s^t \int_s^{\tau_1} \left[G_t f(\mathbf{x}(t), t) + \int_s^{\tau_2} L_t G_t f(\mathbf{x}(\tau_3), \tau_3) \mathrm{d}\tau + \int_s^{\tau_2} G_t^2 f(\mathbf{x}(\tau_3), \tau_3) \mathrm{d}\omega(\tau_3) \right] \mathrm{d}\omega(\tau_2) \mathrm{d}\tau_1 \\
= \; & L_t f(\mathbf{x}(t), t) \int_s^t \int_s^{\tau_1} \mathrm{d}\tau_2 \mathrm{d}\tau_1 + G_t f(\mathbf{x}(t), t) \int_s^t \int_s^{\tau_1} \mathrm{d}\omega(\tau_2) \mathrm{d}\tau_1 \\
& + \int_s^t \int_s^{\tau_1} \int_s^{\tau_2} L_t^2 f(\mathbf{x}(\tau_3), \tau_3) \mathrm{d}\tau_3 \mathrm{d}\tau_2 \mathrm{d}\tau_1 + \int_s^t \int_s^{\tau_1} \int_s^{\tau_2} G_t L_t f(\mathbf{x}(\tau_3), \tau_3) \mathrm{d}\omega(\tau_3) \mathrm{d}\tau_2 \mathrm{d}\tau_1 \\
& + \int_s^t \int_s^{\tau_1} \int_s^{\tau_2} L_t G_t f(\mathbf{x}(\tau_3), \tau_3) \mathrm{d}\tau_3 \mathrm{d}\omega(\tau_2) \mathrm{d}\tau_1 \\
& + \int_s^t \int_s^{\tau_1} \int_s^{\tau_2} G_t^2 f(\mathbf{x}(\tau_3), \tau_3) \mathrm{d}\omega(\tau_3) \mathrm{d}\omega(\tau_2) \mathrm{d}\tau_1 \\
= \; & L_t f(\mathbf{x}(s), s) I_{(0,0)} + G_t f(\mathbf{x}(s), s) I_{(0,1)} + L_t^2 f(\mathbf{x}(s), s) I_{(0,0,0)} + G_t L_t f(\mathbf{x}(s), s) I_{(0,0,1)} \\
& + L_t G_t f(\mathbf{x}(s), s) I_{(0,1,0)} + G_t^2 f(\mathbf{x}(s), s) I_{(1,1,0)} + R_2,
\end{aligned}$$

and with R_2 consisting on quadruple integrals. As such, our solution now reads

$$\begin{aligned}
\mathbf{x}(t) \; = \; & \mathbf{x}(s) + f(\mathbf{x}(s), s) h + g(s)(\omega(t) - \omega(s)) + L_t f(\mathbf{x}(s), s) I_{(0,0)} + G_t f(\mathbf{x}(s), s) I_{(0,1)} \\
& + L_t^2 f(\mathbf{x}(s), s) I_{(0,0,0)} + G_t L_t f(\mathbf{x}(s), s) I_{(0,0,1)} \\
& + L_t G_t f(\mathbf{x}(s), s) I_{(0,1,0)} + G_t^2 f(\mathbf{x}(s), s) I_{(1,1,0)} + R_2.
\end{aligned}$$

Now, in the SDE regime, one cannot get rid of the iterated Itô integral and so stochastic RK methods cannot be derived as simple extensions of their deterministic counterparts. In order to continue we now take into account the fact that the diffusion SDE has additive and diagonal noise. In this case, both the Itô and the Stratonovich SDE coincide.

Iterated integrals

Now, for simplicity, set $t = 0$. We then have $I_{(0)} = h$, $I_{(0,0)} = \int_0^h \int_0^{\tau_1} \mathrm{d}\tau_2 \mathrm{d}\tau_1 = \frac{h^2}{2}$, $I_{(0,0,0)} = \frac{h^3}{6}$. Now notice that

$$\begin{aligned}
I_{(1)} \; & := \; \hat{w}_h \sim \mathcal{N}(0, h) \\
I_{(0,1)} \; & := \; \hat{z}_h := \int_0^h \int_0^{\tau_1} \mathrm{d}\omega(\tau_2) \mathrm{d}\tau_1 = \lim_{n \to \infty} \frac{h}{n} \sum_{i=0}^{n-1} \sum_{j=1}^{i} \epsilon_j, \qquad \epsilon_j \sim \mathcal{N}\left(0, \frac{h}{n}\right).
\end{aligned}$$

Additionally, $\hat{w}_h$ and $\hat{z}_h$ satisfy $\mathbb{E}[\hat{w}_h^2] = h$,

$$\begin{aligned}
\mathbb{E}[(\hat{w}_h h - \hat{z}_h)^2] &= \mathbb{E}\left[\left(\int_0^h \tau \mathrm{d}\omega(\tau)\right)^2\right] = \frac{1}{3}h^3 \\
\mathbb{E}[\hat{w}_h \hat{z}_h] &= \mathbb{E}\left[\hat{w}_h \int_0^h \tau \mathrm{d}\omega(\tau)\right] = \mathbb{E}\left[\int_0^h \tau \mathrm{d}\tau\right] = \frac{1}{2}h^2 \\
\mathbb{E}[\hat{z}_h^2] &= \mathbb{E}[(\hat{w}_h h - \hat{z}_h)^2 - h^2\hat{w}_h^2 + 2h\hat{w}_h\hat{z}_h] = \frac{1}{3}h^3.
\end{aligned}$$

D.2.3 Integral approximations

Weak Approximations

When crafting weak stochastic approximations to SDEs one may replace multiple Itô integrals by other random variables satisfying the corresponding moment conditions. We will denote $\hat{I}_\alpha$ the approximation of I_α for α a multi-index following [13, Corollary 5.12.1]. Of course, the deterministic integrals $I_{(0,\dots,0)}$ need not to be approximated.

First order approximations

The random variable $\hat{I}_{(1)}$ must satisfy for some constant K:

$$|\mathbb{E}[\hat{I}_{(1)}]| + |\mathbb{E}[(\hat{I}_{(1)})^3]| + |\mathbb{E}[(\hat{I}_{(1)})^2 - h]| \leqslant Kh^2$$

Two possible choices for $\hat{I}_{(1)}$ are either $\hat{I}_{(1)} \sim \mathcal{N}(0,h)$ or $\hat{I}_{(1)}$ is a two-pointed distributed discrete random variable with

$$\mathbb{P}\left[\hat{I}_{(1)} = \pm\sqrt{h}\right] = \frac{1}{2}.$$

Second order approximations

The random variable $\hat{I}_{(1)}$ must satisfy for some constant K:

$$|\mathbb{E}[\hat{I}_{(1)}]| + |\mathbb{E}[(\hat{I}_{(1)})^3]| + |\mathbb{E}[(\hat{I}_{(1)})^5]| + |\mathbb{E}[(\hat{I}_{(1)})^2 - h]| + |\mathbb{E}[(\hat{I}_{(1)})^4 - 3h^2]| \leqslant Kh^2.$$

Two possible choices for $\hat{I}_{(1)}$ are either $\hat{I}_{(1)} \sim \mathcal{N}(0,h)$ or $\hat{I}_{(1)}$ is a three-pointed distributed random variable with

$$\mathbb{P}\left[\hat{I}_{(1)} = \pm\sqrt{3h}\right] = \frac{1}{6}, \qquad \mathbb{P}[\hat{I}_{(1)} = 0] = \frac{2}{3}.$$

The rest follows from the above calculations:

$$\hat{I}_{(0,1)} = \frac{1}{2}h\hat{I}_{(1)}, \qquad i = 0, 1.$$

Third order approximations

One can choose $\hat{I}_{(1)} \sim \mathcal{N}(0,h)$, $\hat{I}_{(0,1)} \sim \mathcal{N}\left(0, \frac{1}{3}h^3\right)$ satisfying $\mathbb{E}[\hat{I}_{(1)}\hat{I}_{(0,1)}] = \frac{1}{2}h^2$.

Then, one can deduce the following:

$$\begin{aligned}
\hat{I}_{(1,0,0)} = \hat{I}_{(0,1,0)} = \hat{I}_{(0,0,1)} &= \frac{1}{6}h^2\hat{I}_{(1)} \\
\hat{I}_{(0,1,1)} = \hat{I}_{(1,0,1)} = \hat{I}_{(1,1,0)} &= \frac{1}{6}h(\hat{I}_{(1)}^2 - h).
\end{aligned}$$

Thus, we can write the solution weak approximation as

$$\begin{aligned}
\mathbf{x}(t) = \; & \mathbf{x}(s) + f(\mathbf{x}(s),s)h + g(s)\hat{I}_{(1)} + L_t f(\mathbf{x}(s),s)\frac{h^2}{2} + R_2 \\
& + G_t f(\mathbf{x}(s),s)\hat{I}_{(0,1)} + L_t^2 f(\mathbf{x}(s),s)\frac{h^3}{6} + G_t L_t f(\mathbf{x}(s),s)\frac{1}{6}h^2\hat{I}_{(1)} \\
& + L_t G_t f(\mathbf{x}(s),s)\frac{1}{6}h^2\hat{I}_{(1)} + G_t^2 f(\mathbf{x}(s),s)\frac{1}{6}h(\hat{I}_{(1)}^2 - h). \qquad (60)
\end{aligned}$$

An example of such a pair $(\hat{I}_{(1)}, \hat{I}_{(0,1)}) = (\hat{w}_h, \hat{z}_h)$ can be easily obtained as follows

$$\begin{bmatrix} \hat{w}_h \\ \hat{z}_h \end{bmatrix} = \begin{bmatrix} \sqrt{h} & 0 \\ \frac{h\sqrt{h}}{2} & \frac{h\sqrt{h}}{2\sqrt{3}} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \qquad u_1, u_2 \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1).$$

Indeed, for such a pair we have

$$\mathbb{E}\left[\begin{bmatrix} \hat{w}_h \\ \hat{z}_h \end{bmatrix} \begin{bmatrix} \hat{w}_h & \hat{z}_h \end{bmatrix} \right] = \begin{bmatrix} h & \frac{h^2}{2} \\ \frac{h^2}{2} & \frac{h^3}{3} \end{bmatrix}.$$

In light of the above expression, the truncated Taylor expansions we refer in the main part of the paper consists on the consideration of only the coefficients in L_t^k. The only noise noise contribution we will consider corresponds to $g(s)\hat{I}_{(1)}$.

E Experiment Details

We evaluate the Fréchet inception distance (FID) after generating 50K samples with each solver, and compare with the statistics of real-data. In our experiments we make use of the code from [12] for continuously trained models as well as their reference FID stats[1] and that of [15] for discretely trained models.

All the experiments of SEEDS for continuous-time models are parameterized within the EDM framework with the discretization of type EDM, linear schedule, and scaling none as described on [12] in noise prediction mode unless explicitly stated. We use the SEEDS-3 method that has 3 NFEs per step and fixed step-size and report FID scores at NFEs divisible by 3.

We leverage the explicit Langevin-like "churn" trick using in [12] to add or remove noise in the sampling phase. Specifically, [12] uses 4 hyper-parameters $S_{\text{churn}}, S_{\text{min}}, S_{\text{max}}$ and S_{noise} in which S_{churn} controls the overall amount of stochasticity added before giving the input to the SEEDS-3 method when the noise level (or time step in EDM configuration) t_i is contained in the noise interval $[S_{\text{min}}, S_{\text{max}}]$. It means that the EDM proposed sampler is stochastic under some conditions of those hyper-parameters and deterministic otherwise, while our method is completely stochastic. In our experiments, we set $S_{\text{churn}} = 0$ except for ImageNet-64 EDM optimized model. We noticed that using the additional stochasticity indeed helps to improve the image quality as in Fig. 1 (c). Moreover, setting S_{noise} slightly above 1 might correct the errors in earlier steps more effectively as indicated in [12].

E.1 Pre-trained model specifications

For producing the CIFAR-10 time-continuous results in Table 1, we use the VP DDPM++ continuous architecture. These models are publicly available in conditional[2] and unconditional[3] versions and were directly derived from [23] under the Apache 2.0 license. On the unconditional mode (Figure 1 (a-b)), we first generate the FID curves of 3 types of DPM-Solver (with orders 1, 2 and 3) using the updates from their official implementation[4] in noise prediction mode. Taking profit of the tuning advancements proposed by [12], we used a linear noise schedule with $\beta_{\text{d}} = 19.1$ and $\beta_{\text{min}} = 0.1$ that slightly differs from the original parameters from [23] but were proven beneficial. We set the end time of sampling ε to 1e-4 as recommended by [15, Appendix D.2]. The values of all benchmark models for Figure 1 (b) were taken directly from tables provided by [15].

In our FFHQ-64 experiments, we employ the unconditional VP pretrained[5] model provided by [12].

[1] https://nvlabs-fi-cdn.nvidia.com/edm/fid-refs/

[2] https://nvlabs-fi-cdn.nvidia.com/edm/pretrained/baseline/baseline-cifar10-32x32-cond-vp.pkl

[3] https://nvlabs-fi-cdn.nvidia.com/edm/pretrained/baseline/baseline-cifar10-32x32-uncond-vp.pkl

[4] https://github.com/LuChengTHU/dpm-solver

[5] https://nvlabs-fi-cdn.nvidia.com/edm/pretrained/baseline/baseline-ffhq-64x64-uncond-vp.pkl

For the CelebA-64 experiments, we use the pre-trained VP unconditional model whose checkpoint[6] is provided by [22]. We use the Type-1 discretization proposed in [15] to ensure compatibility of our method with the prescribed trained steps of such model.

For ImageNet-64, we both use the baseline and the optimized pre-trained models given in [12]. We note that the baseline is trained on the iDDPM class of model (see [12, Tab.1 Col.3] for details), which actually uses different preconditioning and thus the change of variables compared to the optimized model. The Figure 1 (c) was obtained using the EDM-preconditioned checkpoint[7]. The added noise settings of SEEDS-3 solver were not subject to a grid-search optimization procedure. The chosen hyper-parameters were $S_{\text{churn}} = 11$, $S_{\text{noise}} = 1.003$, $S_{\text{min}} = 0.05$, and $S_{\text{max}} = 15$ but we are confident that this configuration can be optimized to further improve our results.

E.2 Noise vs. Data Prediction approaches

In Appendix B of [16], the authors compare DPM-Solver2 and DPM-Solver++(2S), which amounts on comparing in our framework the difference between the obtained exponential integrators for the PFO in the noise and data prediction regimes to detect exactly a coefficient on the non-linear term that is absent in the noise prediction regime. The term they find corresponds exactly to the difference between applying the variation of constants formula before (instead of after) replacing the score function with the desired neural network. In Tab. 2 we report both data and noise prediction SEEDS-3. At low NFEs the DP approach gives better results but stabilizes in high NFEs at a FID score that is worse than the one the NP approach reaches.

Table 2: Comparison between noise prediction $F_{\theta,t}$ and data prediction $D_{\theta,t}$ modes of SEEDS-3 on CIFAR-10 (VP uncond. cont.).

SAMPLING METHOD / NFE	9	30	60	90	150	165	180
SEEDS-3 DATA PREDICTION	60.75	22.42	12.47	2.95	2.51	2.54	2.55
SEEDS-3 NOISE PREDICTION	471.29	288.20	33.92	3.76	2.40	**2.39**	2.48

E.3 Low vs. High stage Solvers

Similar to DPM-Solver [15], the FID scores in Tab. 3 and 4 show that at low NFEs, higher stage methods performs more poorly while at higher NFEs, DPM-Solver-3 and SEEDS-3 are better than their counterparts with 1 and 2 stages.

E.4 Deterministic vs. Stochastic Solver

Deterministic solvers as DPM-Solver [15] are fast and well-adapted to applications in which speed is the most concern. As shown in [15], Table 3 and 4, DPM-Solver converges to a local minimum at early steps and cannot be improved in large NFEs. Moreover, preconditioned deterministic solver in EDM gives optimal quality on unconditional CIFAR-10 [12, Fig. 5 (b)]. However, for more complicated datasets as ImageNet-64, the stochasticity indeed helps improve the samples quality [12, Fig. 5 (c)]. We can consider the random noise as a corrector that approaches a better local or even global minimum. At $S_{\text{churn}} = 0$, our SEEDS-3 with stochasticity gives lower FID score than EDM deterministic Heun on ImageNet-64 (see Fig. 1 (c)). In more detail, for the EDM baseline model, we obtain FID$=2.06$ in 210 NFEs, compared to the reported FID$=2.66$ at 511 NFEs in [12]. For the EDM-optimized model, our method gives FID=1.66 in 270 NFEs, which is superior to the 2.22 FID score in 511 NFEs [12]. Additionally, SEEDS-3 also reaches the best quality prior to the number steps needed in Euler Maruyama and other solvers as in Table 5.

[6] https://drive.google.com/file/d/1R_H-fJYXSH79wfSKs9D-fuKQVan5L-GR/view?usp=sharing

[7] https://nvlabs-fi-cdn.nvidia.com/edm/pretrained/edm-imagenet-64x64-cond-adm.pkl

Table 3: FID comparison between SEEDS (Ours) and DPM-Solver for low NFEs on CIFAR-10 VP uncond. discrete. We recomputed the DPM-Solver score using the "non-deep" model while [15] reports results for the "deep" architecture. The symbol * is used when using 1 NFE more and † when using 1 NFE less because the given NFE cannot be divided by 2 or 3. This corresponds to Figure 1 (a).

METHOD / NFE	10	12	15	20	30	40	50	100
DPM-SOLVER-1	22.90	17.73	13.36	9.78	6.87	5.77	5.17	4.22
DPM-SOLVER-2	**12.22**	**6.52**	***4.55**	–	3.75	3.68	3.64	3.60
DPM-SOLVER-3	†66.92	9.72	5.32	*3.83	**3.66**	–	***3.61**	†3.58
SEEDS-1	303.48	239.79	279.84	192.68	84.78	45.26	28.18	8.24
SEEDS-2	481.09	473.48	*430.98	305.88	223.01	51.41	11.10	**3.19**
SEEDS-3	†483.04	482.19	479.63	*462.61	280.48	†247.44	*62.62	†3.53

Table 4: FID comparison between SEEDS (Ours) and DPM-Solver for high NFEs on CIFAR-10 VP uncond. discrete. We recomputed the DPM-Solver score using the "non-deep" model while [15] reports results for the "deep" architecture. The symbol * is used when using 1 NFE more because the given NFE cannot be divided by 2 or 3. This corresponds to Figure 1 (a).

METHOD / NFE	150	200	300	510
DPM-SOLVER-3	3.59	*3.58	–	3.58
SEEDS-1	–	4.07	3.40	–
SEEDS-3	3.12	*3.08	3.14	3.24

Table 5: FID comparison between SEEDS-3 (Ours) and other solvers on CIFAR-10 VP uncond. discrete. The symbol * is used when using 1 NFE more and † when using 1 NFE less because the given NFE cannot be divided by 2 or 3. This corresponds to Figure 1 (b).

METHOD / NFE	10	12	15	20	50	200	1000
EULER-MARUYAMA	278.67	246.29	197.63	137.34	32.63	4.03	3.16
ANALYTIC DDPM	35.03	27.69	20.82	15.35	7.34	4.11	3.84
ANALYTIC DDIM	14.74	11.68	9.16	7.20	4.28	3.60	3.86
DDIM	13.58	11.02	8.92	6.94	4.73	4.07	3.95
DPM-SOLVER-3	†**6.92**	**9.72**	**5.32**	***3.83**	***3.61**	*3.58	–
SEEDS-3	†483.04	482.19	479.63	*462.61	*62.62	***3.08**	–

E.5 Hardware configuration

During the experiments, we used three Linux-based servers with 60GB memory each and 4 GPUs NVIDIA V100 32GB, 4 GPUs NVIDIA V100 16GB, and 2 GPUs NVIDIA V100 32GB, respectively. Table 6 shows the detail of the configuration utilized for each experiment. We noted that when using the 4 GPUs configuration, the FID results were slightly lower (around 2%), even after using a stacked fixed random seed. We run the experiments multiple times and reported the minimum FID value each time.

E.6 Licences

Pre-trained models:

- CIFAR-10 models by [23]: Apache V2.0 license
- FFHQ-64 model by [12]: Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Table 6: Details of GPUs utilized during the experiments.

EXPERIMENT	MODEL	NUMBER	GPU SIZE
CIFAR-10 CONTINUOUS	NVIDIA V100	4	16 GB
CIFAR-10 DISCRETE	NVIDIA V100	2	32 GB
FFHQ64	NVIDIA V100	4	16 GB
CELEBA64	NVIDIA A100	2	16 GB
IMAGENET64	NVIDIA V100	4	32 GB

- CelebA-64 model by [22]: Apache V2.0 license
- ImageNet-64 model by [12]: Apache V2.0 license
- Inception-v3 model by [24]: Apache V2.0 license

E.7 Supporting samples

In this subsection, we report the image grids supporting our claims in the Experiments section.

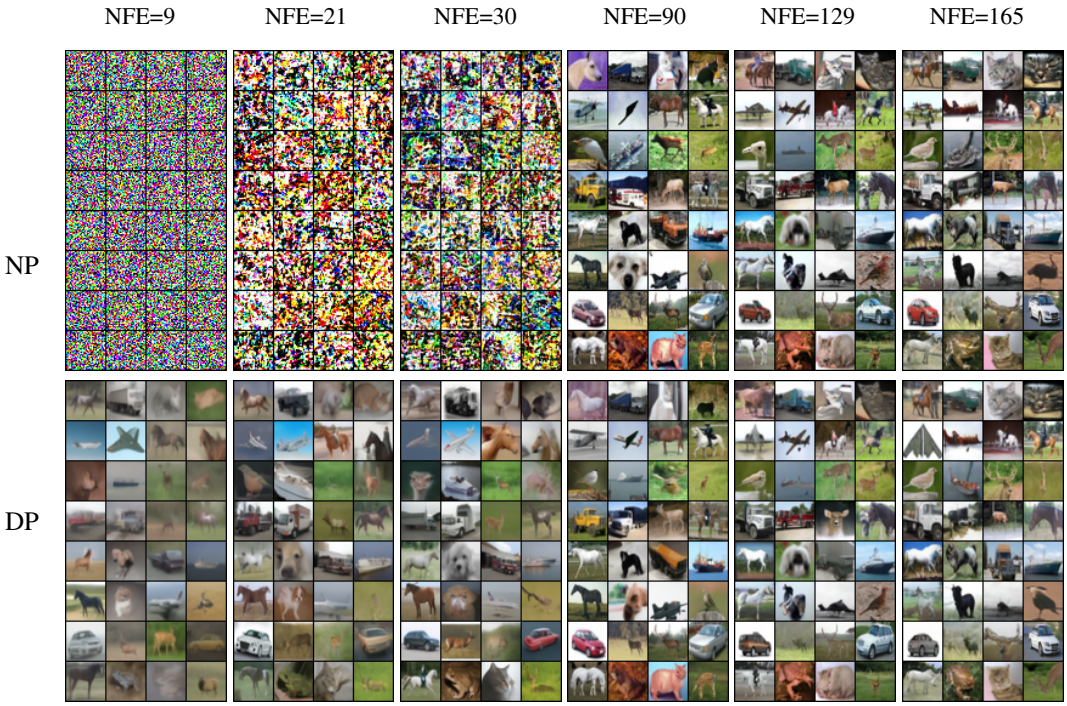


Figure 3: Samples on CIFAR-10 from low to high NFEs by SEEDS-3 in Noise Prediction (NP) and Data Prediction (DP) modes, using conditional VP continuous baseline model [12].

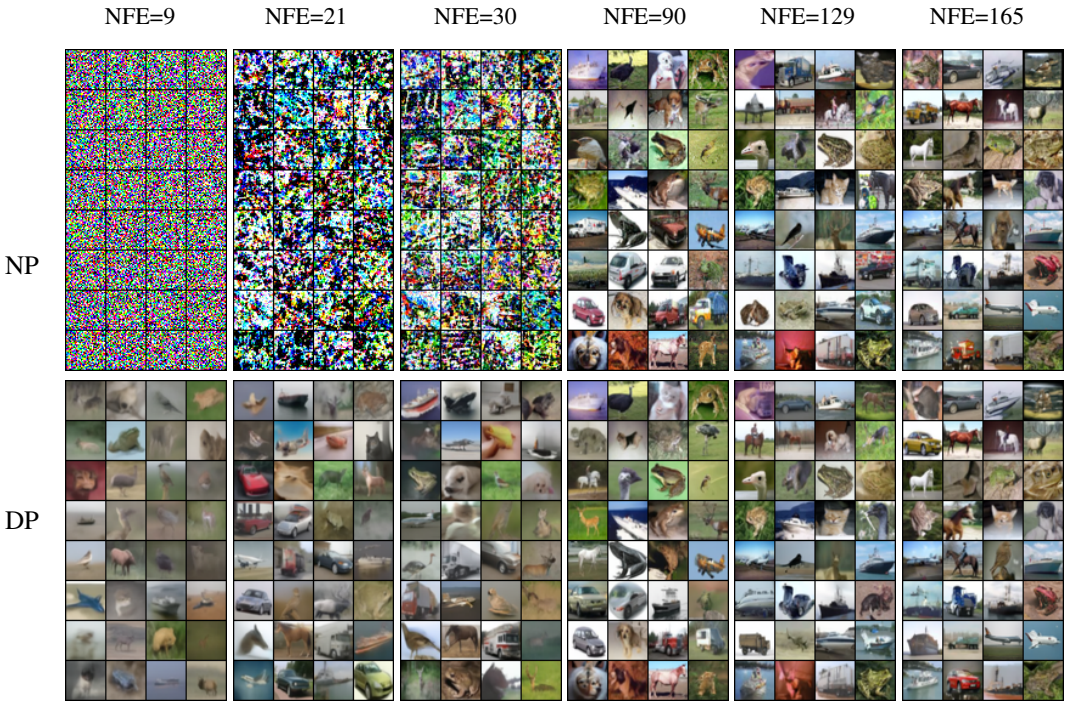


Figure 4: Samples on CIFAR-10 from low to high NFEs by SEEDS-3 in Noise Prediction (NP) and Data Prediction (DP) modes, using unconditional VP continuous baseline model [12].

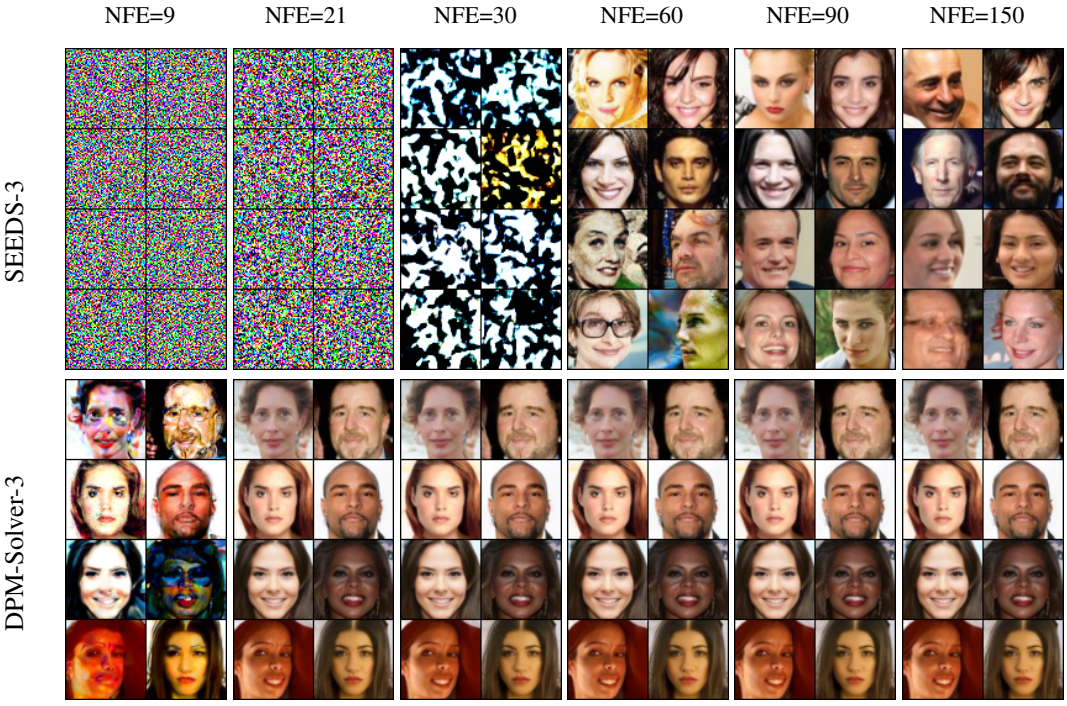


Figure 5: Samples on CelebA-64 from low to high NFEs by SEEDS-3 and DPM-Solver-3, using pre-trained model from [22].

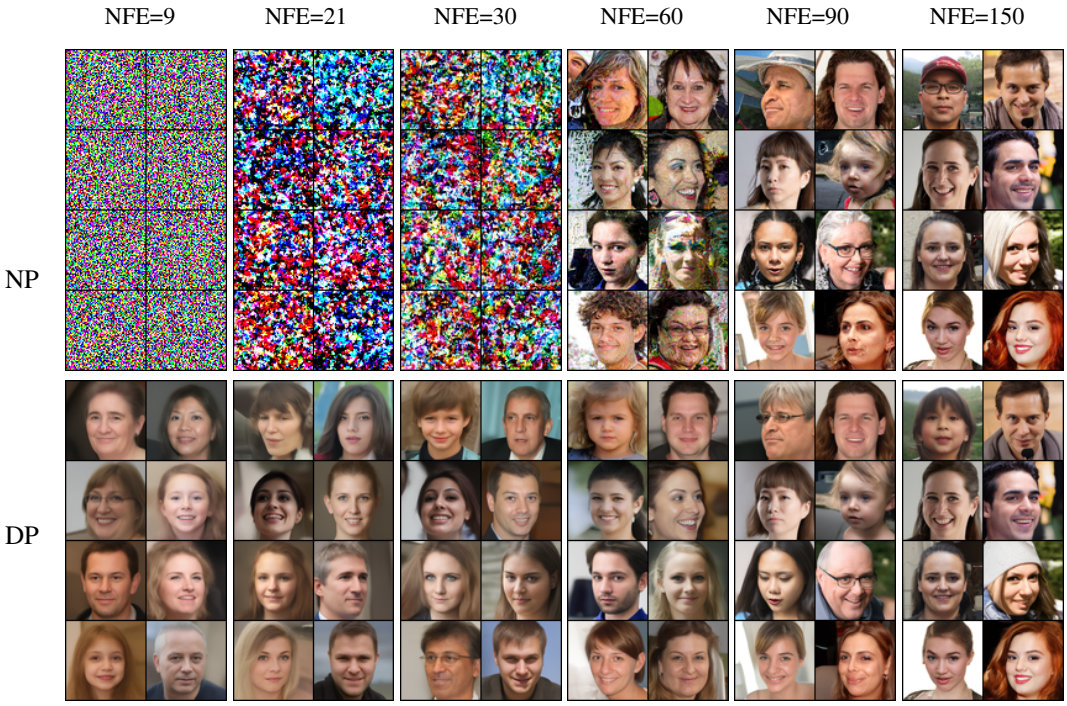


Figure 6: Samples on FFHQ-64 from low to high NFEs by SEEDS-3 in Noise Prediction (NP) and Data Prediction (DP) modes, using unconditional VP continuous baseline model [12].

Figure 7: Samples on ImageNet-64 from low to high NFEs in Noise Prediction (NP) and Data Prediction (DP) modes, using conditional EDM optimized model [12].

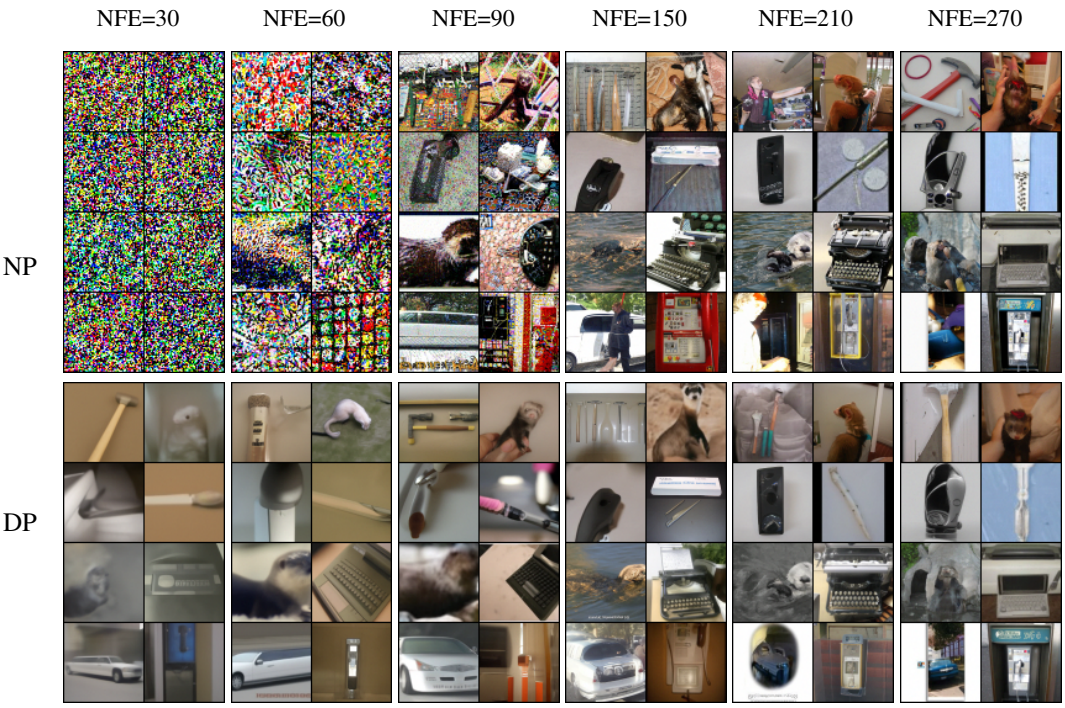


Figure 8: Samples on ImageNet-64 from low to high NFEs in Noise Prediction (NP) and Data Prediction (DP) modes, using conditional EDM baseline model [12].

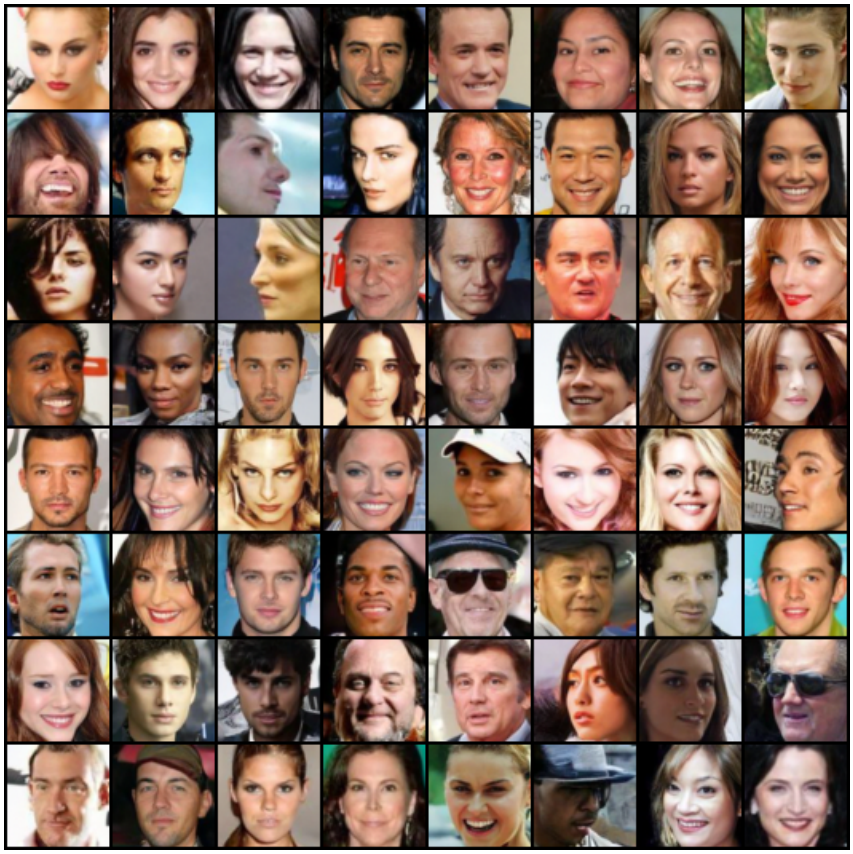
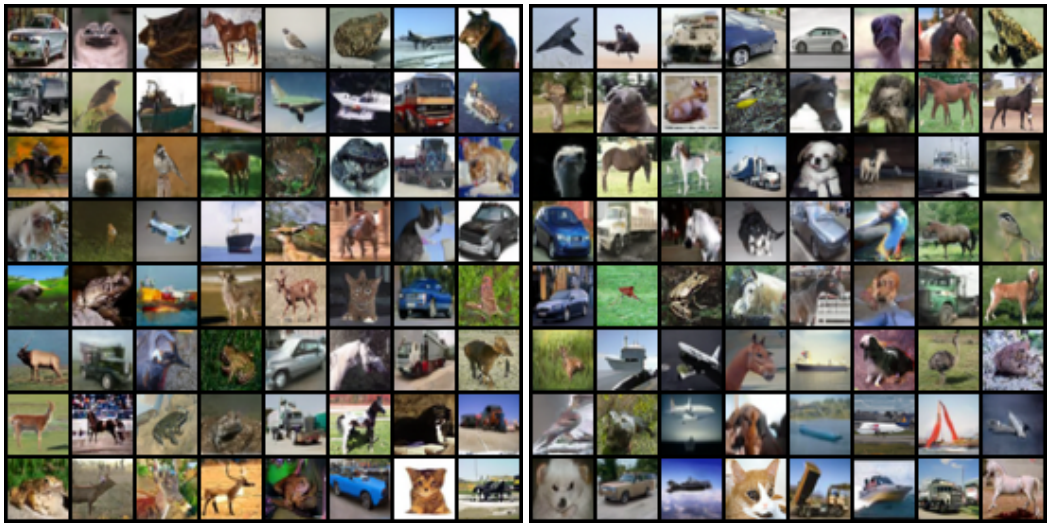


Figure 9: Example of samples on CelebA-64 generated by SEEDS-3 in 90 NFEs, using pre-trained model from [22].

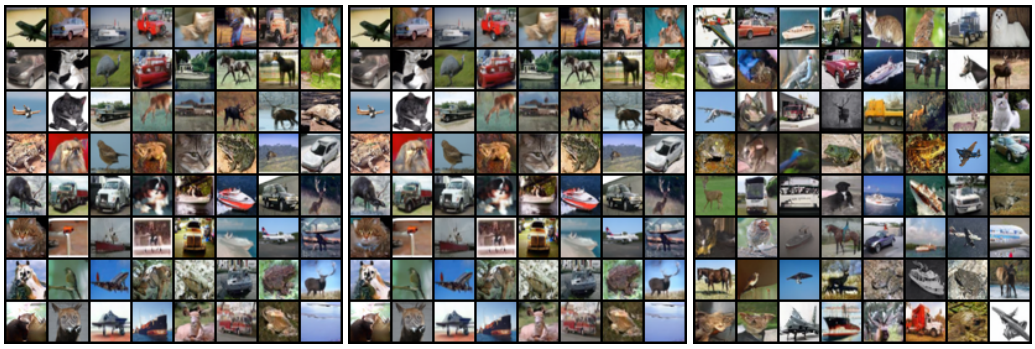


(a) DPM-Solver-3 at 44 NFE. (b) SEEDS-3 at 201 NFE.

Figure 10: Visual sample quality comparison between DPM-Solver-3 and SEEDS-3 using their optimal settings and unconditional CIFAR-10 discrete model [23].

(a) DPM-Solver-3 at 80 NFE. (b) EDM at 80 NFE. (c) SEEDS-3 at 150 NFE.

Figure 11: Visual sample quality comparison between DPM-Solver-3, EDM and SEEDS-3 using their optimal settings and unconditional VP FFHQ-64 continuous model [12].



(a) DPM-Solver-3 at 30 NFE. (b) EDM at 90 NFE. (c) SEEDS-3 at 150 NFE.

Figure 12: Visual sample quality comparison between DPM-Solver-3, EDM and SEEDS-3 using their optimal settings and conditional VP CIFAR-10 baseline continuous model [12].

Figure 13: Example of samples on LSUN-Bedroom-256 by SEEDS-3 in 201 NFEs, using pre-trained model from [5].