DC-Solver: Improving Predictor-Corrector Diffusion Sampler via Dynamic Compensation *Supplementary Material*

A Detailed Background of Diffusion Models

A.1 Diffusion Models

In this section, we will provide a detailed background of diffusion probabilistic models (DPMs) [1, 11]. DPMs usually contain a forward diffusion process that gradually adds noise to the clean data and a backward denoising process that progressively removes the noise to obtain the cleaned data. The diffusion process can be defined either discretely [1] or continuously [11]. We will focus on the latter since continuous DPMs are usually used in the context of DPM samplers [6,7,14]. Let $\boldsymbol{x}_0$ be a random variable from the data distribution $q_0(\boldsymbol{x}_0)$, the forward (diffusion) process gradually adds noise via:

$$q_{t|0}(\boldsymbol{x}_t|\boldsymbol{x}_0) = \mathcal{N}(\boldsymbol{x}_t|\alpha_t\boldsymbol{x}_0, \sigma_t^2\boldsymbol{I}), \tag{1}$$

where α_t, σ_t control the noise schedule and the signal-to-noise-ratio α_t^2/σ_t^2 is decreasing w.r.t t. The noise schedule is designed such that the resulting distribution $q_T(\boldsymbol{x}_T)$ is approximately Gaussian. The forward process can be also formulated via an SDE [3]:

$$\mathrm{d}\boldsymbol{x}_t = f(t)\boldsymbol{x}_t\mathrm{d}t + g(t)\mathrm{d}\boldsymbol{w}_t, \quad \boldsymbol{x}_0 \sim q_0(\boldsymbol{x}_0) \tag{2}$$

where $f(t) = \frac{\mathrm{d}\log\alpha_t}{\mathrm{d}t}$, $g^2(t) = \frac{\mathrm{d}\sigma_t^2}{\mathrm{d}t} - 2\frac{\mathrm{d}\log\alpha_t}{\mathrm{d}t}\sigma_t^2$ and $\boldsymbol{w}_t$ is the standard Wiener process. The reverse process can be analytically computed under some conditons [11]:

$$\mathrm{d}\boldsymbol{x}_t = [f(t)\boldsymbol{x}_t - g^2(t)\nabla_{\boldsymbol{x}}\log q_t(\boldsymbol{x}_t)]\mathrm{d}t + g(t)\mathrm{d}\bar{\boldsymbol{w}}_t, \tag{3}$$

where $\bar{\boldsymbol{w}}_t$ is the standard Winer process in the reverse time. DPM is trained to estimate the scaled score function $-\sigma_t\nabla_{\boldsymbol{x}}\log q_t(\boldsymbol{x}_t)$ via a neural network $\boldsymbol{\epsilon}_\theta$, and the corresponding SDE during sampling is

$$\mathrm{d}\boldsymbol{x}_t = \left[f(t)\boldsymbol{x}_t + \frac{g^2(t)}{\sigma_t}\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t)\right]\mathrm{d}t + g(t)\mathrm{d}\bar{\boldsymbol{w}}_t. \tag{4}$$

A.2 ODE-based DPM samplers

Although one can numerally solve the diffusion SDE by discretizing (4), the stochasticity would harm the sampling quality especially when the step size is large. On the contrary, the probability flow ODE [11] is more practical:

$$\frac{\mathrm{d}\boldsymbol{x}_t}{\mathrm{d}t} = f(t)\boldsymbol{x}_t - \frac{g^2(t)}{2}\nabla_{\boldsymbol{x}} \log q_t(\boldsymbol{x}_t). \tag{5}$$

Modern fast samplers of DPMs [6, 7, 14] aim to efficiently solve the above ODE with small numbers of function evaluations (NFE) by introducing several useful techniques such as the exponential integrator [6, 13], the multi-step method [7, 13], data-prediction [7], and predictor-corrector paradigm [14]. For example, the deterministic version of DDIM [10] can be viewed as a 1-order discretization of the diffusion probability flow ODE. DPM-Solver [6] leverages an insightful parameterization (logSNR) and exponential integrator to achieve a high-order solver. DPM-Solver++ [7] further adopts the multi-step method to estimate high-order derivatives. Specifically, one can use a buffer to store the outputs of $\boldsymbol{\epsilon}_\theta$ on previous points and use them to increase the order of accuracy. PNDM [4] modified classical multi-step numerical methods to corresponding pseudo numerical methods for DPM sampling. UniPC [14] introduces a predictor-corrector framework that also uses the model output at the current point to improve the sampling quality, and bypasses the extra model evaluations by re-using the model outputs at the next sampling step. Generally speaking, the formulation of existing DPM samplers can be summarized as follows:

$$\tilde{\boldsymbol{x}}_{t_i} = A_{t_{i-1}}^{t_i} \tilde{\boldsymbol{x}}_{t_{i-1}}^{\mathrm{c}} + \sum_{m=1}^{p-1} B_{t_{i-m}}^{t_i} \boldsymbol{\beta}_\theta(\tilde{\boldsymbol{x}}_{t_{i-m}}, t_{i-m}), \tag{6}$$

$$\tilde{\boldsymbol{x}}_{t_i}^{\mathrm{c}} = C_{t_{i-1}}^{t_i} \tilde{\boldsymbol{x}}_{t_{i-1}}^{\mathrm{c}} + \sum_{m=0}^{p-1} D_{t_{i-m}}^{t_i} \boldsymbol{\beta}_\theta(\tilde{\boldsymbol{x}}_{t_{i-m}}, t_{i-m}), \tag{7}$$

where the corrector step (7) is optional and $\boldsymbol{x}_{t_i}^{\mathrm{c}} = \boldsymbol{x}_{t_i}$ if no corrector is used. We use $\boldsymbol{\beta}_\theta$ to represent different parameterizations during the sampling, such as the noise-prediction $\boldsymbol{\epsilon}_\theta$ [6, 13], data-prediction $\boldsymbol{x}_\theta$ [7, 14], v-prediction $\boldsymbol{v}_\theta$ [9], or the learned parameterization [15]. The coefficients (A, B, C, D) are determined by the specific sampler and differ across the sampling steps.

B Convergence of DC-Solver

In this section, we shall show that if the original sampler has the convergence order $p+1$ under mild conditions, then the same order of convergence is maintained when combined with our Dynamic Compensation. We will prove for both predictor-only samplers [7,10] and predictor-corrector samplers [14]. For the sake of simplicity, we use the $\ell-2$ norm by default to study the convergence.

B.1 Assumptions

We introduce some assumptions for the convenience of subsequent proofs. These assumptions are either common in ODE analysis or easy to satisfy.

Assumption 1 *The prediction model $\boldsymbol{\beta}_\theta(x,t)$ is Lipschitz continuous w.r.t. x.*

Assumption 2 $h = \max_{1\le i\le M} h_i = \mathcal{O}(1/M)$*, where h_i denotes the sampling step size, and M is the total number of sampling steps.*

Assumption 3 *The coefficients in* (7) *satisfy that* $0 < C_1 \le \|A_{t_{i-1}}^{t_i}\|_2 \le C_2$, $0 < C_3 h \le \|B_{t_{i-m}}^{t_i}\|_2 \le C_4 h$, $0 < C_5 \le \|C_{t_{i-1}}^{t_i}\|_2 \le C_6$ *and* $0 < C_7 h \le \|D_{t_{i-m}}^{t_i}\|_2 \le C_8 h$ *for sufficiently small h.*

Assumption 1 is common in the analysis of ODEs. Assumption 2 assures that the step size is basically uniform.

Assumption 3 can be easily verified by the formulation of the samplers. For example, in data-prediction mode of UniPC [14], we have $A_{t_{i-1}}^{t_i} = \alpha_{t_i}/\alpha_{t_{i-1}}$, which are constants independent of h_i. Note that $B_{t_{i-1}}^{t_i} = \sigma_{t_i}(e^{h_i}-1)\left[\sum_{m=1}^{p}\frac{a_m}{r_m}-1\right]$ and $B_{t_{i-m}}^{t_i} = -\sigma_{t_i}(e^{h_i}-1)\frac{a_m}{r_m}, m \neq 1$, where $a_m, r_m \in \mathcal{O}(1)$, we have $B_{t_i-m}^{t_i} = \mathcal{O}(h)$. For $C_{t_{i-1}}^{t_i}$ and $D_{t_{i-m}}^{t_i}$, we can analogically derive the bound for the two coefficients. By examining the analytical form of other existing solvers [4, 6, 7, 10, 13, 14], we can similarly find that Theorem 3 always holds.

B.2 Local Convergence

Theorem 4. *For any DPM sampler of $p+1$-th order of accuracy,* i.e., $\mathbb{E}\|\tilde{\boldsymbol{x}}_{t_{i+1}}^{\mathrm{c}} - \tilde{\boldsymbol{x}}_{t_{i+1}}\|_2 \le Ch_i^{p+2}$*, applying dynamic compensation with the ratio ρ_i^* will reduce the local truncation error and remain the $p+1$-th order of accuracy.*

Proof. Denote $\tilde{\boldsymbol{x}}_{t_{i+1}}^{\mathrm{c},\rho_i}$ as the intermediate result at the next sampling step by using dynamic compensation ratio ρ_i. Observe that $\rho_i = 1.0$ is equivalent to the original updating formula without the dynamic compensation, we have

$$\begin{aligned}\mathbb{E}\|\tilde{\boldsymbol{x}}_{t_{i+1}}^{\mathrm{c},\rho_i^*} - \tilde{\boldsymbol{x}}_{t_{i+1}}\|_2 &\le \mathbb{E}\|\tilde{\boldsymbol{x}}_{t_{i+1}}^{\mathrm{c},1.0} - \tilde{\boldsymbol{x}}_{t_{i+1}}\|_2 \\ &= \mathbb{E}\|\tilde{\boldsymbol{x}}_{t_{i+1}}^{\mathrm{c}} - \tilde{\boldsymbol{x}}_{t_{i+1}}\|_2 \le Ch_i^{p+2}.\end{aligned} \tag{8}$$

Therefore, the local truncation error is reduced and the order of accuracy after the DC is still $p+1$.

Note that the proof does not assume the detailed implementation of the sampler, indicating that the Theorem 4 holds for both predictor-only samplers and predictor-corrector samplers.

B.3 Global Convergence

We first investigate the global convergence of Dynamic Compensation with a p-th order predictor-only sampler.

Corollary 1. *Assume that we have* $\{\tilde{\boldsymbol{x}}_{t_{i-k}}\}_{k=1}^{p-1}$ *and* $\{\boldsymbol{\beta}_\theta^{\rho^*_{i-k}}(\tilde{\boldsymbol{x}}_{t_{i-k}}, t_{i-k})\}_{k=2}^{p-1}$ *(denoted as* $\{\boldsymbol{\beta}_\theta^{\rho^*_{i-k}}\}_{k=2}^{p-1}$*) satisfying* $\mathbb{E}\|\tilde{\boldsymbol{x}}_{t_{i-k}} - \boldsymbol{x}_{t_{i-k}}\|_2 = \mathcal{O}(h^p), 1 \leq k \leq p-1$*, and* $\mathbb{E}\|\boldsymbol{\beta}_\theta^{\rho^*_{i-k}} - \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_{i-k}}, t_{i-k})\|_2 = \mathcal{O}(h^{p-1}), 2 \leq k \leq p-1$*. If we use Predictor-p together with Dynamic Compensation to estimate* $\boldsymbol{x}_{t_i}$*, we shall get* $\boldsymbol{\beta}_\theta^{\rho^*_{i-1}}$ *and* $\tilde{\boldsymbol{x}}_{t_i}$ *that satisfy* $\mathbb{E}\|\boldsymbol{\beta}_\theta^{\rho^*_{i-1}} - \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_{i-1}}, t_{i-1})\|_2 = \mathcal{O}(h^{p-1})$ *and* $\mathbb{E}\|\tilde{\boldsymbol{x}}_{t_i} - \boldsymbol{x}_{t_i}\|_2 = \mathcal{O}(h^p)$*.*

Proof. It is obvious that for sufficiently large constants $C_{\boldsymbol{\beta}}, C_{\boldsymbol{x}}$, we have

$$\mathbb{E}\|\boldsymbol{\beta}_\theta^{\rho^*_{i-k}} - \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_{i-k}}, t_{i-k})\|_2 \leq C_{\boldsymbol{\beta}} h^{p-1}, 2 \leq k \leq p-1 \tag{9}$$

$$\mathbb{E}\|\tilde{\boldsymbol{x}}_{t_{i-k}} - \boldsymbol{x}_{t_{i-k}}\|_2 \leq C_x h^p, 1 \leq k \leq p-1 \tag{10}$$

When computer $\boldsymbol{x}_{t_i}$, we consider 3 different methods in this step. Firstly, if we continue to use Dynamic Compensation, we have

$$\tilde{\boldsymbol{x}}_{t_i} = A_{t_{i-1}}^{t_i} \tilde{\boldsymbol{x}}_{t_{i-1}} + \sum_{m=1}^{p-1} B_{t_{i-m}}^{t_i} \boldsymbol{\beta}_\theta^{\rho^*_{i-m}}. \tag{11}$$

Otherwise, if we use the standard Predictor-p at this step (which means to do not replace the $\boldsymbol{\beta}_\theta(\tilde{\boldsymbol{x}}_{t_{i-1}}, t_{i-1})$ with $\boldsymbol{\beta}_\theta^{\rho^*_{i-m}}$), we have the following result:

$$\tilde{\boldsymbol{x}}_{t_i}^{\mathrm{p}} = A_{t_{i-1}}^{t_i} \tilde{\boldsymbol{x}}_{t_{i-1}} + \sum_{m=2}^{p-1} B_{t_{i-m}}^{t_i} \boldsymbol{\beta}_\theta^{\rho^*_{i-m}} + B_{t_{i-1}}^{t_i} \boldsymbol{\beta}_\theta(\tilde{\boldsymbol{x}}_{t_{i-1}}, t_{i-1}). \tag{12}$$

In the third case, we adopt the Predictor-p to previous points on the ground truth trajectory:

$$\bar{\boldsymbol{x}}_{t_i} = A_{t_{i-1}}^{t_i} \boldsymbol{x}_{t_{i-1}} + \sum_{m=1}^{p-1} B_{t_{i-m}}^{t_i} \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_{i-m}}, t_{i-m}) \tag{13}$$

Due to the p-th order of accuarcy of Predictor-p, we have

$$\mathbb{E}\|\bar{\boldsymbol{x}}_{t_i} - \boldsymbol{x}_{t_i}\|_2 = \mathcal{O}(h^{p+1}) \tag{14}$$

Comparing (13) and (12), we obtain

$$\begin{aligned} \tilde{\boldsymbol{x}}_{t_i}^{\mathrm{p}} - \bar{\boldsymbol{x}}_{t_i} &= A_{t_{i-1}}^{t_i}(\tilde{\boldsymbol{x}}_{t_{i-1}} - \boldsymbol{x}_{t_{i-1}}) \\ &+ \sum_{m=2}^{p-1} B_{t_{i-m}}^{t_i} \left[\boldsymbol{\beta}_\theta^{\rho^*_{i-m}} - \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_{i-m}}, t_{i-m})\right] \\ &+ B_{t_{i-1}}^{t_i} \left[\boldsymbol{\beta}_\theta(\tilde{\boldsymbol{x}}_{t_{i-1}}, t_{i-1}) - \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_{i-1}}, t_{i-1})\right] \end{aligned} \tag{15}$$

Under Assumption 1, Assumption 3, (9) and (10), it follows that,

$$\begin{aligned}\mathbb{E}\|\tilde{\boldsymbol{x}}_{t_i}^{\mathrm{p}} - \bar{\boldsymbol{x}}_{t_i}\|_2 &\leq C_2 C_x h^p \\ &+ \sum_{m=2}^{p-1} C_4 C_{\boldsymbol{\beta}} h^p + C_4 L C_x h^{p+1} = \mathcal{O}(h^p)\end{aligned} \tag{16}$$

By (14) and (16), we have

$$\mathbb{E}\|\tilde{\boldsymbol{x}}_{t_i}^{\mathrm{p}} - \boldsymbol{x}_{t_i}\|_2 = \mathcal{O}(h^p) \tag{17}$$

Observing that DC-Solver-p is equivalent to Predictor-p when $\rho_{i-1} = 1.0$, we have

$$\mathbb{E}\|\tilde{\boldsymbol{x}}_{t_i} - \boldsymbol{x}_{t_i}\|_2 \leq \mathbb{E}\|\tilde{\boldsymbol{x}}_{t_i}^{\mathrm{p}} - \boldsymbol{x}_{t_i}\|_2 = \mathcal{O}(h^p). \tag{18}$$

Combining with (14), we get

$$\mathbb{E}\|\tilde{\boldsymbol{x}}_{t_i} - \bar{\boldsymbol{x}}_{t_i}\|_2 = \mathcal{O}(h^p) \leq C_9 h^p \tag{19}$$

Subtracting (13) from (11), we have

$$\begin{aligned}\tilde{\boldsymbol{x}}_{t_i} - \bar{\boldsymbol{x}}_{t_i} &= A_{t_{i-1}}^{t_i}(\tilde{\boldsymbol{x}}_{t_{i-1}} - \boldsymbol{x}_{t_{i-1}}) \\ &+ \sum_{m=2}^{p-1} B_{t_{i-m}}^{t_i}\left[\boldsymbol{\beta}_\theta^{\rho_{i-m}^*} - \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_{i-m}}, t_{i-m})\right] \\ &+ B_{t_{i-1}}^{t_i}\left[\boldsymbol{\beta}_\theta^{\rho_{i-1}^*} - \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_{i-1}}, t_{i-1})\right]\end{aligned} \tag{20}$$

Thus, given (19), (9), (10), we obtain

$$\begin{aligned}&\mathbb{E}\left\|B_{t_{i-1}}^{t_i}\left[\boldsymbol{\beta}_\theta^{\rho_{i-1}^*} - \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_{i-1}}, t_{i-1})\right]\right\|_2 \\ &= \left\|\tilde{\boldsymbol{x}}_{t_i} - \bar{\boldsymbol{x}}_{t_i} - A_{t_{i-1}}^{t_i}(\tilde{\boldsymbol{x}}_{t_{i-1}} - \boldsymbol{x}_{t_{i-1}}) \right.\\ &\left. - \sum_{m=2}^{p-1} B_{t_{i-m}}^{t_i}\left[\boldsymbol{\beta}_\theta^{\rho_{i-m}^*} - \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_{i-m}}, t_{i-m})\right]\right\|_2 \\ &\leq C_9 h^p + C_2 C_x h^p + \sum_{m=2}^{p-1} C_4 C_{\boldsymbol{\beta}} h^p \\ &= \mathcal{O}(h^p)\end{aligned} \tag{21}$$

Note that $\|B_{t_{i-1}}^{t_i}\|_2 \geq C_3 h$ according to Assumption 3, we have

$$\mathbb{E}\|\boldsymbol{\beta}_\theta^{\rho_{i-1}^*} - \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_{i-1}}, t_{i-1})\|_2 = \mathcal{O}(h^{p-1}). \tag{22}$$

Above all, (19) and (22) establish the correctness of the corollary.

Theorem 5. *For any predictor-only sampler of p-th order of convergence, applying Dynamic Compensation with ratio ρ_i^* will maintain the p-th order of convergence.*

Proof. We will use mathematical induction to prove it. Denote $\{\boldsymbol{\beta}_\theta^{\rho_k^*}\}_{k=0}^{i-1} = \{\boldsymbol{\beta}_\theta^{\rho_k^*}(\tilde{\boldsymbol{x}}_{t_k}, t_k)\}_{k=0}^{i-1}$, we define P_i as the proposition that $\mathbb{E}\|\boldsymbol{\beta}_\theta^{\rho_k^*} - \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_k}, t_k)\|_2 = \mathcal{O}(h^{p-1}), 0 \le k \le i-1$, and $\mathbb{E}\|\tilde{\boldsymbol{x}}_{t_k} - \boldsymbol{x}_{t_k}\|_2 = \mathcal{O}(h^p), 0 \le k \le i$.

In the first K steps (namely the warm-up steps), we only use the Predictor-p without the Dynamic Compensation. Since Predictor-p has p-th order of convergence, it's obvious that $\mathbb{E}\|\tilde{\boldsymbol{x}}_{t_k} - \boldsymbol{x}_{t_k}\|_2 = \mathcal{O}(h^p), 0 \le k \le K$. Under Assumption 1, we also have

$$\begin{aligned}
&\mathbb{E}\|\boldsymbol{\beta}_\theta^{\rho_k^*} - \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_k}, t_k)\|_2 = \mathbb{E}\|\boldsymbol{\beta}_\theta(\tilde{\boldsymbol{x}}_{t_k}, t_k) - \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_k}, t_k)\|_2 \\
&\le \mathbb{E}\|\tilde{\boldsymbol{x}}_{t_k} - \boldsymbol{x}_{t_k}\|_2 = \mathcal{O}(h^p) \le \mathcal{O}(h^{p-1}), \forall 0 \le k \le K-1
\end{aligned} \tag{23}$$

Thus, we show that P_K is true. Recall the result in Corollary 1, we can then use mathematical induction to prove that P_M is true, where M is the NFE. This indicates that $\mathbb{E}\|\tilde{\boldsymbol{x}}_{t_M} - \boldsymbol{x}_{t_M}\|_2 = \mathcal{O}(h^p)$, which concludes the proof that the convergence order is still p with the Dynamic Compensation

We then provide the proof of the convergence order when applying Dynamic Compensation to predictor-corrector solvers.

Corollary 2. *Assume that we have* $\{\tilde{\boldsymbol{x}}^{\mathrm{c}}_{t_{i-k}}\}_{k=1}^{p-1}$, $\{\tilde{\boldsymbol{x}}_{t_{i-k}}\}_{k=1}^{p-1}$, *and* $\{\boldsymbol{\beta}_\theta^{\rho_{i-k}^*}(\tilde{\boldsymbol{x}}^{\mathrm{c}}_{t_{i-k}}, t_{i-k})\}_{k=2}^{p-1}$ *(denoted as* $\{\boldsymbol{\beta}_\theta^{\rho_{i-k}^*}\}_{k=2}^{p-1}$*), which satisfy* $\mathbb{E}\|\boldsymbol{\beta}_\theta^{\rho_{i-k}^*} - \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_{i-k}}, t_{i-k})\|_2 = \mathcal{O}(h^p), 2 \le k \le p-1$, $\mathbb{E}\|\tilde{\boldsymbol{x}}^{\mathrm{c}}_{t_{i-k}} - \boldsymbol{x}_{t_{i-k}}\|_2 = \mathcal{O}(h^{p+1}), 1 \le k \le p-1$*, and* $\mathbb{E}\|\tilde{\boldsymbol{x}}_{t_{i-k}} - \boldsymbol{x}_{t_{i-k}}\|_2 = \mathcal{O}(h^p), 1 \le k \le p-1$*. Then using Predictor-Corrector-p combined with Dynamic Compensation to estimate* $\boldsymbol{x}_{t_i}$*, we can calculate* $\boldsymbol{\beta}_\theta^{\rho_{i-1}^*}, \tilde{\boldsymbol{x}}^{\mathrm{c}}_{t_i}, \tilde{\boldsymbol{x}}_{t_i}$*, that satisfy* $\mathbb{E}\|\boldsymbol{\beta}_\theta^{\rho_{i-1}^*} - \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_{i-1}}, t_{i-1})\|_2 = \mathcal{O}(h^p)$*,* $\mathbb{E}\|\tilde{\boldsymbol{x}}^{\mathrm{c}}_{t_i} - \boldsymbol{x}_{t_i}\|_2 = \mathcal{O}(h^{p+1})$ *and* $\mathbb{E}\|\tilde{\boldsymbol{x}}_{t_i} - \boldsymbol{x}_{t_i}\|_2 = \mathcal{O}(h^p)$

Proof. It is obvious that, there exists sufficiently large constants $C_{\boldsymbol{\beta}}, C_x, C_y$, such that

$$\mathbb{E}\|\boldsymbol{\beta}_\theta^{\rho_{i-k}^*} - \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_{i-k}}, t_{i-k})\|_2 \le C_{\boldsymbol{\beta}} h^p, 2 \le k \le p-1 \tag{24}$$

$$\mathbb{E}\|\tilde{\boldsymbol{x}}^{\mathrm{c}}_{t_{i-k}} - \boldsymbol{x}_{t_{i-k}}\|_2 \le C_x h^{p+1}, 1 \le k \le p-1 \tag{25}$$

$$\mathbb{E}\|\tilde{\boldsymbol{x}}_{t_{i-k}} - \boldsymbol{x}_{t_{i-k}}\|_2 \le C_y h^p, 1 \le k \le p-1 \tag{26}$$

When estimating $\boldsymbol{x}_{t_i}$, we consider three different methods in this step. First, if we use Dynamic Compensation, we have

$$\tilde{\boldsymbol{x}}_{t_i} = A_{t_{i-1}}^{t_i} \tilde{\boldsymbol{x}}^{\mathrm{c}}_{t_{i-1}} + \sum_{m=1}^{p-1} B_{t_{i-m}}^{t_i} \boldsymbol{\beta}_\theta^{\rho_{i-m}^*} \tag{27}$$

$$\tilde{\boldsymbol{x}}^{\mathrm{c}}_{t_i} = C_{t_{i-1}}^{t_i} \tilde{\boldsymbol{x}}^{\mathrm{c}}_{t_{i-1}} + \sum_{m=1}^{p-1} D_{t_{i-m}}^{t_i} \boldsymbol{\beta}_\theta^{\rho_{i-m}^*} + D_{t_i}^{t_i} \boldsymbol{\beta}_\theta(\tilde{\boldsymbol{x}}_{t_i}, t_i) \tag{28}$$

Otherwise, if we use the standard Predictor-Corrector-p without DC at this step, we get

$$\bar{\boldsymbol{x}}_{t_i} = A_{t_{i-1}}^{t_i} \tilde{\boldsymbol{x}}_{t_{i-1}}^{\mathrm{c}} + \sum_{m=2}^{p-1} B_{t_{i-m}}^{t_i} \boldsymbol{\beta}_\theta^{\rho_{i-m}^*} + B_{t_{i-1}}^{t_i} \boldsymbol{\beta}_\theta(\tilde{\boldsymbol{x}}_{t_{i-1}}, t_{i-1}) \tag{29}$$

$$\begin{aligned} \bar{\boldsymbol{x}}_{t_i}^{\mathrm{c}} &= C_{t_{i-1}}^{t_i} \tilde{\boldsymbol{x}}_{t_{i-1}}^{\mathrm{c}} + \sum_{m=2}^{p-1} D_{t_{i-m}}^{t_i} \boldsymbol{\beta}_\theta^{\rho_{i-m}^*} + D_{t_{i-1}}^{t_i} \boldsymbol{\beta}_\theta(\tilde{\boldsymbol{x}}_{t_{i-1}}, t_{i-1}) \\ &+ D_{t_i}^{t_i} \boldsymbol{\beta}_\theta(\bar{\boldsymbol{x}}_{t_i}, t_i) \end{aligned} \tag{30}$$

Finally, we use Predictor-Corrector-p to previous points on the ground truth trajectory, we have:

$$\hat{\boldsymbol{x}}_{t_i} = A_{t_{i-1}}^{t_i} \boldsymbol{x}_{t_{i-1}} + \sum_{m=1}^{p-1} B_{t_{i-m}}^{t_i} \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_{i-m}}, t_{i-m}) \tag{31}$$

$$\hat{\boldsymbol{x}}_{t_i}^{\mathrm{c}} = C_{t_{i-1}}^{t_i} \boldsymbol{x}_{t_{i-1}} + \sum_{m=1}^{p-1} D_{t_{i-m}}^{t_i} \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_{i-m}}, t_{i-m}) + D_{t_i}^{t_i} \boldsymbol{\beta}_\theta(\hat{\boldsymbol{x}}_{t_i}, t_i) \tag{32}$$

Due to Predictor-Corrector-p's $p+1$-th convergence order, we have

$$\mathbb{E}\|\hat{\boldsymbol{x}}_{t_i}^{\mathrm{c}} - \boldsymbol{x}_{t_i}\|_2 = \mathcal{O}(h^{p+2}) \tag{33}$$

Based on Assumption 1 and (26), we also know that

$$\begin{aligned} &\mathbb{E}\|\boldsymbol{\beta}_\theta(\tilde{\boldsymbol{x}}_{t_{i-1}}, t_{i-1}) - \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_{i-1}}, t_{i-1})\|_2 \\ &\leq L\mathbb{E}\|\tilde{\boldsymbol{x}}_{t_{i-1}} - \boldsymbol{x}_{t_{i-1}}\|_2 = \mathcal{O}(h^p) \end{aligned} \tag{34}$$

Subtracting (32) from (30), we obtain

$$\begin{aligned} \bar{\boldsymbol{x}}_{t_i}^{\mathrm{c}} - \hat{\boldsymbol{x}}_{t_i}^{\mathrm{c}} &= C_{t_{i-1}}^{t_i}(\tilde{\boldsymbol{x}}_{t_{i-1}}^{\mathrm{c}} - \boldsymbol{x}_{t_{i-1}}) \\ &+ \sum_{m=2}^{p-1} D_{t_{i-m}}^{t_i} \left[\boldsymbol{\beta}_\theta^{\rho_{i-m}^*} - \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_{i-m}}, t_{i-m})\right] \\ &+ D_{t_{i-1}}^{t_i} \left[\boldsymbol{\beta}_\theta(\tilde{\boldsymbol{x}}_{t_{i-1}}, t_{i-1}) - \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_{i-1}}, t_{i-1})\right] \\ &+ D_{t_i}^{t_i} \left[\boldsymbol{\beta}_\theta(\bar{\boldsymbol{x}}_{t_i}, t_i) - \boldsymbol{\beta}_\theta(\hat{\boldsymbol{x}}_{t_i}, t_i)\right] \end{aligned} \tag{35}$$

Under Assumption 1, Assumption 3, (34), (24), (25) and (26), it follows that,

$$\begin{aligned}
&\mathbb{E}\|\boldsymbol{\beta}_\theta(\bar{\boldsymbol{x}}_{t_i},t_i)-\boldsymbol{\beta}_\theta(\hat{\boldsymbol{x}}_{t_i},t_i)\|_2\le L\mathbb{E}\|\bar{\boldsymbol{x}}_{t_i}-\hat{\boldsymbol{x}}_{t_i}\|_2\\
&=L\mathbb{E}\|A_{t_{i-1}}^{t_i}(\tilde{\boldsymbol{x}}_{t_{i-1}}^{\mathrm{c}}-\boldsymbol{x}_{t_{i-1}})\\
&+\sum_{m=2}^{p-1}B_{t_{i-m}}^{t_i}\left[\boldsymbol{\beta}_\theta^{\rho_{i-m}^*}-\boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_{i-m}},t_{i-m})\right]\\
&+B_{t_{i-1}}^{t_i}\left[\boldsymbol{\beta}_\theta(\tilde{\boldsymbol{x}}_{t_{i-1}},t_{i-1})-\boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_{i-1}},t_{i-1})\right]\|_2\\
&\le L(C_2C_xh^{p+1}+\sum_{m=2}^{p-1}C_4C_{\boldsymbol{\beta}}h^{p+1}+C_4LC_yh^{p+1})\\
&=\mathcal{O}(h^{p+1})\le C_{10}h^{p+1}
\end{aligned}\tag{36}$$

Therefore, according to Assumption 3, (24), (25), (26), (35) and (36), we get

$$\begin{aligned}
\mathbb{E}\|\bar{\boldsymbol{x}}_{t_i}^{\mathrm{c}}-\hat{\boldsymbol{x}}_{t_i}^{\mathrm{c}}\|_2\le{}& C_6C_xh^{p+1}+\sum_{m=2}^{p-1}C_8C_{\boldsymbol{\beta}}h^{p+1}\\
&+C_8LC_yh^{p+1}+C_8C_{10}h^{p+2}\\
&=\mathcal{O}(h^{p+1})
\end{aligned}\tag{37}$$

Given (33), we have

$$\mathbb{E}\|\bar{\boldsymbol{x}}_{t_i}^{\mathrm{c}}-\boldsymbol{x}_{t_i}\|_2=\mathcal{O}(h^{p+1})\tag{38}$$

Observe that DC-Solver-p is equivalent to Predictor-Corrector-p when $\rho_{i-1}=1.0$, we have

$$\mathbb{E}\|\tilde{\boldsymbol{x}}_{t_i}^{\mathrm{c}}-\boldsymbol{x}_{t_i}\|_2\le\mathbb{E}\|\bar{\boldsymbol{x}}_{t_i}^{\mathrm{c}}-\boldsymbol{x}_{t_i}\|_2=\mathcal{O}(h^{p+1})\tag{39}$$

Combining with (38), we get

$$\mathbb{E}\|\tilde{\boldsymbol{x}}_{t_i}^{\mathrm{c}}-\bar{\boldsymbol{x}}_{t_i}^{\mathrm{c}}\|_2=\mathcal{O}(h^{p+1})\tag{40}$$

Comparing (28) and (30), we have

$$\begin{aligned}
\tilde{\boldsymbol{x}}_{t_i}^{\mathrm{c}}-\bar{\boldsymbol{x}}_{t_i}^{\mathrm{c}}={}&D_{t_{i-1}}^{t_i}\left[\boldsymbol{\beta}_\theta^{\rho_{i-1}^*}-\boldsymbol{\beta}_\theta(\tilde{\boldsymbol{x}}_{t_{i-1}},t_{i-1})\right]\\
&+D_{t_i}^{t_i}\left[\boldsymbol{\beta}_\theta(\tilde{\boldsymbol{x}}_{t_i},t_i)-\boldsymbol{\beta}_\theta(\bar{\boldsymbol{x}}_{t_i},t_i)\right]
\end{aligned}\tag{41}$$

Under Assumption 3 and 1, concerning about the order of the coefficients, we can know that

$$\begin{aligned}
&\mathbb{E}\|D_{t_i}^{t_i}\left[\boldsymbol{\beta}_\theta(\tilde{\boldsymbol{x}}_{t_i},t_i)-\boldsymbol{\beta}_\theta(\bar{\boldsymbol{x}}_{t_i},t_i)\right]\|_2\\
&\le L\|D_{t_i}^{t_i}\|_2\|B_{t_{i-1}}^{t_i}\|_2\mathbb{E}\|\boldsymbol{\beta}_\theta^{\rho_{i-1}^*}-\boldsymbol{\beta}_\theta(\tilde{\boldsymbol{x}}_{t_{i-1}},t_{i-1})\|_2\\
&\ll\mathbb{E}\|D_{t_{i-1}}^{t_i}\left[\boldsymbol{\beta}_\theta^{\rho_{i-1}^*}-\boldsymbol{\beta}_\theta(\tilde{\boldsymbol{x}}_{t_{i-1}},t_{i-1})\right]\|_2
\end{aligned}\tag{42}$$

Leveraging (40), (41) with (42), we have

$$\mathbb{E}\|D_{t_{i-1}}^{t_i}\left[\boldsymbol{\beta}_\theta^{\rho_{i-1}^*}-\boldsymbol{\beta}_\theta(\tilde{\boldsymbol{x}}_{t_{i-1}},t_{i-1})\right]\|_2=\mathcal{O}(h^{p+1})\tag{43}$$

Thus, considering that $\|D_{t_i}^{t_i}\|_2 \geq C_7 h$ in Assumption 3, we can get

$$\|\boldsymbol{\beta}_\theta^{\rho_{i-1}^*} - \boldsymbol{\beta}_\theta(\tilde{\boldsymbol{x}}_{t_{i-1}}, t_{i-1})\|_2 = \mathcal{O}(h^p) \tag{44}$$

Given (34) and (44), we further obtain

$$\|\boldsymbol{\beta}_\theta^{\rho_{i-1}^*} - \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_{i-1}}, t_{i-1})\|_2 = \mathcal{O}(h^p) \leq C_{11} h^p \tag{45}$$

Subtracting (31) from (27), we obtain

$$\begin{aligned}
&\mathbb{E}\|\tilde{\boldsymbol{x}}_{t_i} - \hat{\boldsymbol{x}}_{t_i}\|_2 = \mathbb{E}\|A_{t_{i-1}}^{t_i}(\tilde{\boldsymbol{x}}_{t_{i-1}}^{\mathrm{c}} - \boldsymbol{x}_{t_{i-1}}) \\
&+ B_{t_{i-1}}^{t_i}\left[\boldsymbol{\beta}_\theta^{\rho_{i-1}^*} - \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_{i-1}}, t_{i-1})\right] \\
&+ \sum_{m=2}^{p-1} B_{t_{i-m}}^{t_i}\left[\boldsymbol{\beta}_\theta^{\rho_{i-m}^*} - \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_{i-m}}, t_{i-m})\right]\|_2 \\
&\leq C_2 C_x h^{p+1} + C_4 C_{11} h^{p+1} + \sum_{m=2}^{p-1} C_4 C_{\boldsymbol{\beta}} h^{p+1} \\
&\leq \mathcal{O}(h^p)
\end{aligned} \tag{46}$$

Since $\mathbb{E}\|\hat{\boldsymbol{x}}_{t_i} - \boldsymbol{x}_{t_i}\|_2 = \mathcal{O}(h^{p+1})$, we have

$$\mathbb{E}\|\tilde{\boldsymbol{x}}_{t_i} - \boldsymbol{x}_{t_i}\|_2 \leq \mathcal{O}(h^p) \tag{47}$$

Above all, (39), (45) and (47) imply the validity of the corollary.

Theorem 6. *For any predictor-corrector sampler of $(p+1)$-th order of convergence, applying dynamic compensation with ratio ρ_i^* will remain the $(p+1)$-th order of convergence.*

Proof. We use mathematical induction to proof this. Suppose we have $\{\tilde{\boldsymbol{x}}_{t_k}^{\mathrm{c}}\}_{k=0}^{i}$, $\{\tilde{\boldsymbol{x}}_{t_k}\}_{k=0}^{i}$ and $\{\boldsymbol{\beta}_\theta^{\rho_k^*}(\tilde{\boldsymbol{x}}_{t_k}^{\mathrm{c}}, t_k)\}_{k=0}^{i-1}$ denoted as $\{\boldsymbol{\beta}_\theta^{\rho_k^*}\}_{k=0}^{i-1}$. First, we define P_i as the proposition that $\mathbb{E}\|\boldsymbol{\beta}_\theta^{\rho_k^*} - \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_k}, t_k)\|_2 = \mathcal{O}(h^p), 0 \leq k \leq i-1$, $\mathbb{E}\|\tilde{\boldsymbol{x}}_{t_k}^{\mathrm{c}} - \boldsymbol{x}_{t_k}\|_2 = \mathcal{O}(h^{p+1}), 0 \leq k \leq i$ and $\mathbb{E}\|\tilde{\boldsymbol{x}}_{t_k} - \boldsymbol{x}_{t_k}\|_2 = \mathcal{O}(h^p), 0 \leq k \leq i$.
In the first K steps, we only use Predictor-Corrector-p without the Dynamic Compensation. Since Predictor-Corrector-p has $(p+1)$-th order of convergence, it's obvious that $\mathbb{E}\|\tilde{\boldsymbol{x}}_{t_k}^{\mathrm{c}} - \boldsymbol{x}_{t_k}\|_2 = \mathcal{O}(h^{p+1}), 0 \leq k \leq K$, and $\mathbb{E}\|\tilde{\boldsymbol{x}}_{t_k} - \boldsymbol{x}_{t_k}\|_2 = \mathcal{O}(h^p), 0 \leq k \leq K$. Under Assumption 1, we also know, for $k \in [0, K-1]$,

$$\begin{aligned}
\mathbb{E}\|\boldsymbol{\beta}_\theta^{\rho_k^*} - \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_k}, t_k)\|_2 &= \mathbb{E}\|\boldsymbol{\beta}_\theta(\tilde{\boldsymbol{x}}_{t_k}, t_k) - \boldsymbol{\beta}_\theta(\boldsymbol{x}_{t_k}, t_k)\|_2 \\
&\leq L\mathbb{E}\|\tilde{\boldsymbol{x}}_{t_k} - \boldsymbol{x}_{t_k}\|_2 = \mathcal{O}(h^p)
\end{aligned} \tag{48}$$

Thus, we show that P_K is true. Similarly, using mathematical induction and the result in Corollary 2 we can know that P_M is true, which implies that $\mathbb{E}\|\tilde{\boldsymbol{x}}_{t_M}^{\mathrm{c}} - \boldsymbol{x}_{t_M}\|_2 = \mathcal{O}(h^{p+1})$ and ends the proof. Therefore, we reach the conclusion that for a predictor-corrector sampler, the Dynamic Compensation will preserve the $p+1$ convergence order.

Table 1: Detailed quantitative results on unconditional sampling. We provide the comparisons of the FID↓ of our DC-Solver and the previous method on FFHQ [2], LSUN-Church [12] and LSUN-Bedroom [12] with 5∼10 NFE. We observe that our DC-Solver achieves the lowest FID on all three datasets.

(a) FFHQ [2]

Method	NFE					
	5	6	7	8	9	10
DPM-Solver++ [7]	27.15	15.60	10.81	8.98	7.89	7.39
DEIS [13]	32.35	18.72	12.22	9.51	8.31	7.75
UniPC [14]	18.66	11.89	9.51	8.21	7.62	6.99
DC-Solver (Ours)	**10.38**	**8.39**	**7.66**	**7.14**	**6.92**	**6.82**

(b) LSUN-Church [12]

Method	NFE					
	5	6	7	8	9	10
DPM-Solver++ [7]	17.57	9.71	6.45	4.97	4.25	3.87
DEIS [13]	15.01	8.45	5.71	4.49	3.86	3.57
UniPC [14]	11.98	6.90	5.08	4.28	3.86	3.61
DC-Solver (Ours)	**7.47**	**4.70**	**3.91**	**3.46**	**3.23**	**3.06**

(c) LSUN-Bedroom [12]

Method	NFE					
	5	6	7	8	9	10
DPM-Solver++ [7]	18.13	8.33	5.15	4.14	3.77	3.61
DEIS [13]	16.68	8.75	6.13	5.11	4.66	4.41
UniPC [14]	12.14	6.13	4.53	4.05	3.81	3.64
DC-Solver (Ours)	**7.40**	**5.29**	**4.27**	**3.98**	**3.74**	**3.52**

C More Analyses

C.1 Quantitative Results

We now provide detailed quantitative results on both unconditional sampling and conditional sampling. For unconditional sampling, we list the numerical results on FFHQ [2], LSUN-Church [12] and LSUN-Bedroom [12] in Table 1. All the pre-trained DPMs are from Latent-Diffusion [8] and we use FID↓ as the evaluation metric. We demonstrate that our DC-Solver consistently attains the lowest FID on all three datasets. For conditional sampling, we summarize the

Table 2: Detailed quantitative results on conditional sampling. We provide the comparisons between our DC-Solver and the previous method on Stable-Diffusion-1.5 [8] with different classifier-free guidance scale (CFG) and NFE $\in [5, 10]$. The sampling quality is measured by the MSE↓ between the generated latents and the ground truth latents (obtained by a 999-step DDIM). We demonstrate that DC-Solver consistently achieves the best result for different sampling configurations.

(a) CFG = 1.0

Method	NFE					
	5	6	7	8	9	10
DPM-Solver++ [7]	0.277	0.232	0.204	0.188	0.177	0.169
DEIS [13]	0.299	0.252	0.223	0.203	0.191	0.181
UniPC [14]	0.245	0.206	0.184	0.172	0.166	0.161
DC-Solver (Ours)	**0.176**	**0.163**	**0.150**	**0.150**	**0.147**	**0.144**

(b) CFG = 1.5

Method	NFE					
	5	6	7	8	9	10
DPM-Solver++ [7]	0.288	0.242	0.213	0.195	0.182	0.173
DEIS [13]	0.307	0.260	0.229	0.209	0.194	0.184
UniPC [14]	0.260	0.219	0.194	0.180	0.170	0.163
DC-Solver (Ours)	**0.213**	**0.188**	**0.169**	**0.158**	**0.153**	**0.149**

(c) CFG = 2.5

Method	NFE					
	5	6	7	8	9	10
DPM-Solver++ [7]	0.339	0.293	0.262	0.239	0.221	0.208
DEIS [13]	0.354	0.307	0.274	0.250	0.231	0.217
UniPC [14]	0.321	0.277	0.247	0.226	0.208	0.195
DC-Solver (Ours)	**0.293**	**0.257**	**0.231**	**0.212**	**0.194**	**0.186**

(d) CFG = 3.5

Method	NFE					
	5	6	7	8	9	10
DPM-Solver++ [7]	0.409	0.360	0.323	0.295	0.272	0.255
DEIS [13]	0.419	0.369	0.332	0.303	0.280	0.262
UniPC [14]	0.397	0.349	0.312	0.285	0.262	0.245
DC-Solver (Ours)	**0.375**	**0.331**	**0.299**	**0.270**	**0.251**	**0.239**

(e) CFG = 4.5

Method	NFE					
	5	6	7	8	9	10
DPM-Solver++ [7]	0.490	0.437	0.392	0.358	0.330	0.308
DEIS [13]	0.496	0.441	0.397	0.364	0.336	0.314
UniPC [14]	0.483	0.430	0.386	0.352	0.324	0.302
DC-Solver (Ours)	**0.461**	**0.412**	**0.369**	**0.337**	**0.314**	**0.291**

(f) CFG = 5.5

Method	NFE					
	5	6	7	8	9	10
DPM-Solver++ [7]	0.580	0.517	0.468	0.427	0.395	0.368
DEIS [13]	0.581	0.519	0.469	0.430	0.398	0.372
UniPC [14]	0.577	0.516	0.468	0.428	0.395	0.367
DC-Solver (Ours)	**0.551**	**0.492**	**0.446**	**0.406**	**0.381**	**0.355**

(g) CFG = 6.5

Method	NFE					
	5	6	7	8	9	10
DPM-Solver++ [7]	0.687	0.612	0.556	0.512	0.474	0.441
DEIS [13]	0.684	0.610	0.554	0.511	0.474	0.442
UniPC [14]	0.691	0.618	0.563	0.517	0.479	0.445
DC-Solver (Ours)	**0.654**	**0.587**	**0.531**	**0.488**	**0.457**	**0.426**

(h) CFG = 7.5

Method	NFE					
	5	6	7	8	9	10
DPM-Solver++ [7]	0.812	0.719	0.648	0.597	0.554	0.518
DEIS [13]	0.802	0.712	0.643	0.592	0.552	0.517
UniPC [14]	0.825	0.733	0.666	0.612	0.570	0.530
DC-Solver (Ours)	**0.766**	**0.689**	**0.620**	**0.573**	**0.537**	**0.501**

results in Table 2, where we compare the sampling quality of different methods on various configurations of classifier-free guidance scale (CFG). Our results indicate that DC-Solver can outperform previous methods by large margins with different choices of CFG and NFE.

C.2 Qualitative Results

We present additional visualizations to showcase the superior qualitative performance of DC-Solver in both unconditional sampling and conditional sam-

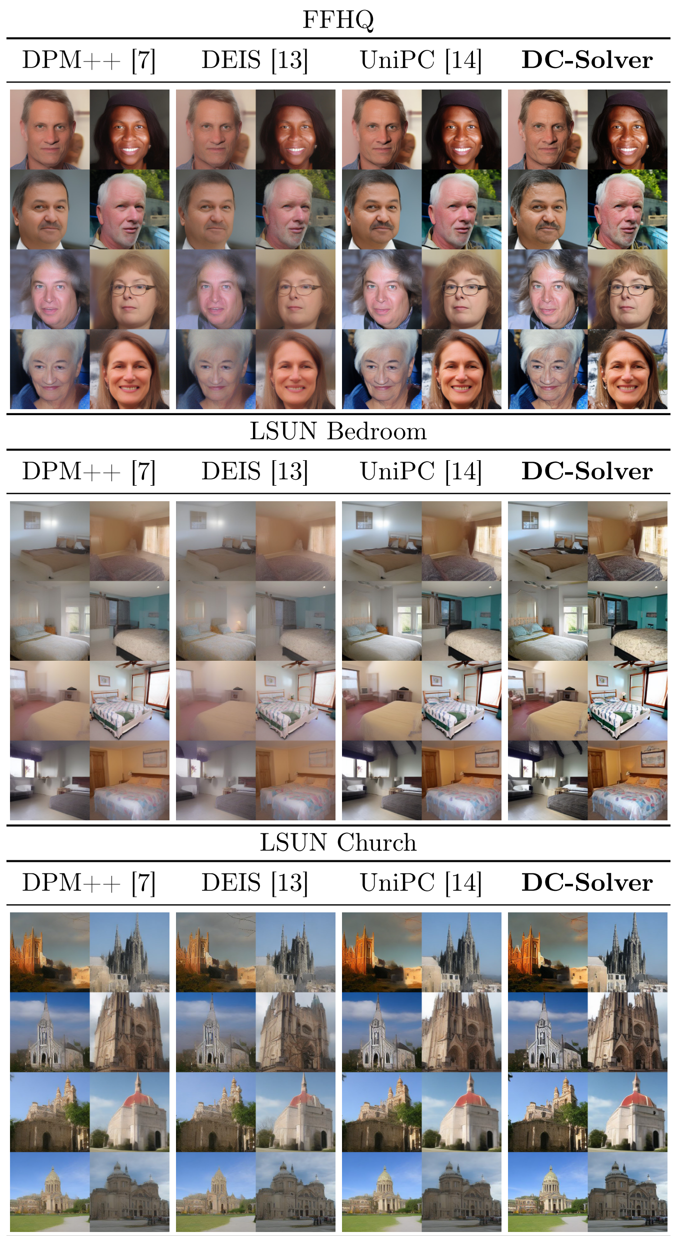
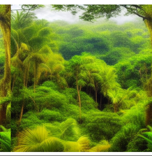
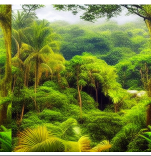
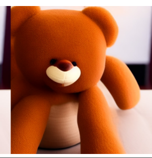
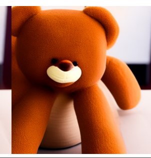
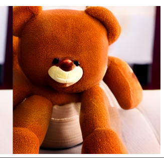


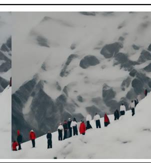
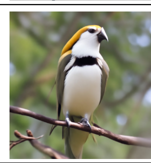
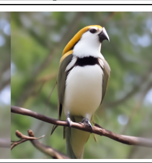
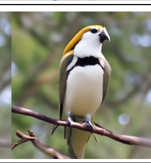
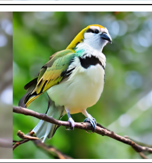
Fig. 1: Comparisons of unconditional sampling results across different datasets employing DC-Solver, UniPC [14], DPM-Solver++ [7] and DEIS [13]. Images are sampled using only 5 NFE.

pling. Initially, we compare the unconditional sampling quality of four different

Stable-Diffusion 1.5

Text Prompts	DPM++ [7]	DEIS [13]	UniPC [14]	**DC-Solver**
"A realistic photo of a tropical rainforest with diverse wildlife."				
"Close up of a teddy bear sitting on top of it."				

Stable-Diffusion 2.1

Text Prompts	DPM++ [7]	DEIS [13]	UniPC [14]	**DC-Solver**
"Group of people standing on top of a snow covered slope."				
"Close up of a bird perched on top of a tree."				

Stable-Diffusion XL

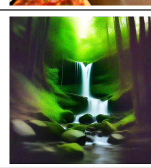
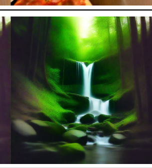
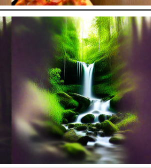
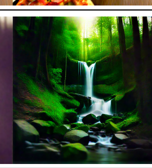
Text Prompts	DPM++ [7]	DEIS [13]	UniPC [14]	**DC-Solver**
"Pizza that is sitting on top of a plate."				
"A serene waterfall in a lush green forest."				

Fig. 2: Comparisons of text-to-image results on different pre-trained Stable-Diffusion models using DC-Solver, UniPC [14], DPM-Solver++ [7] and DEIS [13]. Images are sampled with a classifier-free guidance scale 7.5, using only 5 NFE.

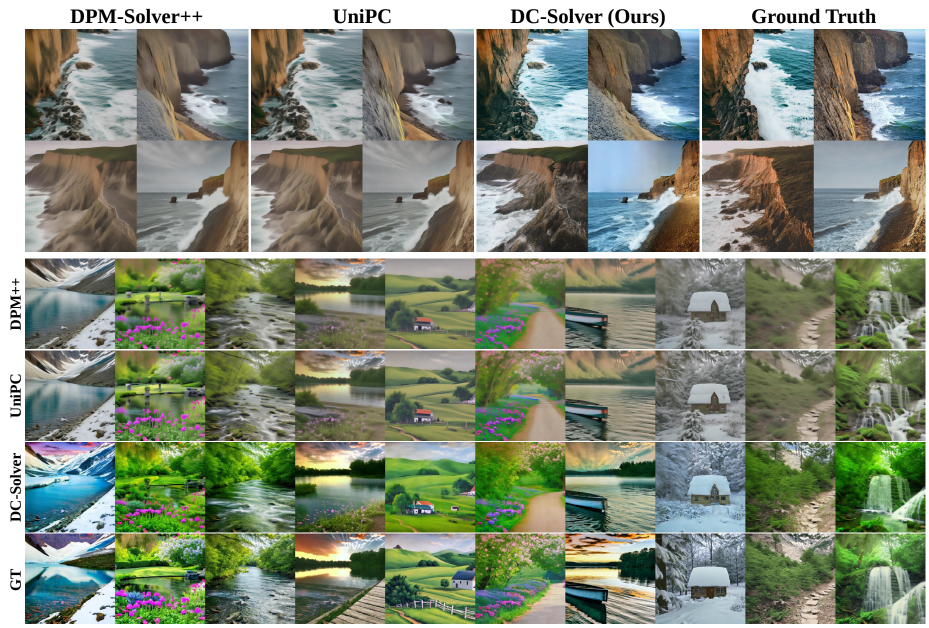


Fig. 3: Comparison with GT (upper part) and more uncurated results (lower part). For all the compared methods, we adopt NFE=5 and use the same initial noise. We can clearly find that DC-Solver outperforms other methods.

methods on FFHQ [2], LSUN-Church [12] and LSUN-Bedroom [12] in Figure 1, employing only 5 NFE. We show that DC-Solver can produce the clearest and most realistic images across all three datasets. Furthermore, we explore conditional sampling on different pre-trained Stable-Diffusion(SD) models, including SD1.5, SD2.1 and SDXL, with only 5 NFE. The reuslts in Figure 2 demonstrate that our DC-Solver is able to generate more realistic images with more details, consistently outperforming other methods on all three SD models.

D Implementation Details

Our DC-Solver is built on the predictor-corrector framework UniPC [14] by default. We set the order of the dynamic compensation $K = 2$ and skip the compensation when $i < K$, which is equivalent to $\rho_0 = \rho_1 = 1.0$. $K = 2$ also implies a parabola-like interpolation trajectory. During the searching stage, we set the number of datapoints $N = 10$. We use a 999-step DDIM [10] to generate the ground truth trajectory $\boldsymbol{x}_t^{\mathrm{GT}}$ in the conditional sampling while we found a 200-step DDIM is enough for unconditional sampling. We use AdamW [5] to optimize the compensation ratios for only $L = 40$ iterations and set the learning rate of learnable parameters as $\alpha = 0.1$. For the cascade polynomial regression, we use $p_1 = p_2 = 2$ and $p_3 = 3$. For the experiments on Latent-Diffusion [8], we adopt their original checkpoints and use the default latent size 64×64. For the experiments of conditional sampling using Stable-Diffusion [8], we use the default latent size of 64×64, 64×64, 96×96, 128×128 for SD1.4, SD1.5, SD2.1, SDXL, respectively. It is worth noting that our method can be scaled up to larger latent sizes and pre-trained DPMs mainly because of the effectiveness of the designed dynamic compensation, which can be controlled by several scalar parameters.

References

1. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. NeurIPS **33**, 6840–6851 (2020)
2. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: CVPR. pp. 4401–4410 (2019)
3. Kingma, D., Salimans, T., Poole, B., Ho, J.: Variational diffusion models. NeurIPS **34**, 21696–21707 (2021)
4. Liu, L., Ren, Y., Lin, Z., Zhao, Z.: Pseudo numerical methods for diffusion models on manifolds. ICLR (2022)
5. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
6. Lu, C., Zhou, Y., Bao, F., Chen, J., Li, C., Zhu, J.: Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps. NeurIPS (2022)
7. Lu, C., Zhou, Y., Bao, F., Chen, J., Li, C., Zhu, J.: Dpm-solver++: Fast solver for guided sampling of diffusion probabilistic models. arXiv preprint arXiv:2211.01095 (2022)
8. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: CVPR. pp. 10684–10695 (2022)
9. Salimans, T., Ho, J.: Progressive distillation for fast sampling of diffusion models. ICLR (2022)
10. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. ICLR (2021)
11. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. In: ICLR (2021)
12. Yu, F., Seff, A., Zhang, Y., Song, S., Funkhouser, T., Xiao, J.: Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop. arXiv preprint arXiv:1506.03365 (2015)
13. Zhang, Q., Chen, Y.: Fast sampling of diffusion models with exponential integrator. arXiv preprint arXiv:2204.13902 (2022)
14. Zhao, W., Bai, L., Rao, Y., Zhou, J., Lu, J.: Unipc: A unified predictor-corrector framework for fast sampling of diffusion models. NeurIPS (2023)
15. Zheng, K., Lu, C., Chen, J., Zhu, J.: Dpm-solver-v3: Improved diffusion ode solver with empirical model statistics. arXiv preprint arXiv:2310.13268 (2023)