Appendix

Table of Contents

A Acknowledgment of LLM Usage

We used a large language model (ChatGPT) to polish this paper. Its use was limited to grammar checking, fixing typos, rephrasing sentences for clarity, and improving word choice. All conceptual contributions, methodological designs, experiments, and analyses were carried out entirely by the authors. The use of an LLM does not affect the reproducibility or scientific validity of our work.

B Proof

Proof of Equation (11) . For a Gaussian distribution $q(\boldsymbol{x}) \approx \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\boldsymbol{\Sigma}_t \in \mathbb{R}^{n\times n}$.

$$\log q(\boldsymbol{x}) = \log\left((2\pi)^{\frac{-n}{2}}\right) + \log\left(\det(\boldsymbol{\Sigma}_t)^{-\frac{1}{2}}\right) - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{x}-\boldsymbol{\mu}). \tag{20}$$

The Jacobian matrix and Hessian matrix are

$$\nabla \log q(\boldsymbol{x}) = -\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{x}-\boldsymbol{\mu}), \nabla^2 \log q(\boldsymbol{x}) = \boldsymbol{\Sigma}_t^{-1}. \tag{21}$$

Proof of Equation (15) As $\mu_{t-1|t}(\boldsymbol{x}) = \mathbb{E}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t) = \int \boldsymbol{x}_{t-1} q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t) d\boldsymbol{x}_{t-1}$, the derivative of $\mu_{t-1|t}$ is

$$\nabla_{\boldsymbol{x}_t}\mu_{t-1|t}(\boldsymbol{x}_t) = \nabla_{\boldsymbol{x}_t}\int \boldsymbol{x}_{t-1} q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t) d\boldsymbol{x}_{t-1} = \int \boldsymbol{x}_{t-1}\nabla_{\boldsymbol{x}_t} q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t) d\boldsymbol{x}_{t-1} \tag{22}$$

For $\nabla_{\boldsymbol{x}_t} q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)$ part, using bayes rule to get

$$\begin{aligned}
\nabla_{\boldsymbol{x}_t} q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t) &= \nabla_{\boldsymbol{x}_t}\left(\frac{q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})q(\boldsymbol{x}_{t-1})}{q(\boldsymbol{x}_t)}\right) \\
&= q(\boldsymbol{x}_{t-1})\nabla_{\boldsymbol{x}_t}\left(\frac{q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})}{q(\boldsymbol{x}_t)}\right) \\
&= q(\boldsymbol{x}_{t-1})\left(\frac{\nabla_{\boldsymbol{x}_t} q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})}{q(\boldsymbol{x}_t)} - \frac{q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})\nabla_{\boldsymbol{x}_t} q(\boldsymbol{x}_t)}{q(\boldsymbol{x}_t)^2}\right) \\
&= \frac{q(\boldsymbol{x}_{t-1})q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})}{q(\boldsymbol{x}_t)}\left(\frac{\nabla_{\boldsymbol{x}_t} q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})}{q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})} - \frac{\nabla_{\boldsymbol{x}_t} q(\boldsymbol{x}_t)}{q(\boldsymbol{x}_t)}\right) \\
&= q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)(\nabla_{\boldsymbol{x}_t}\log q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) - \nabla_{\boldsymbol{x}_t}\log q(\boldsymbol{x}_t))
\end{aligned} \tag{23}$$

Substitude $\nabla_{\boldsymbol{x}_t} q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)$ in Equation (22)

$$\begin{aligned}\nabla_{\boldsymbol{x}_t}\mu_{t-1|t}(\boldsymbol{x}_t) = &\int \boldsymbol{x}_{t-1} q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)\nabla_{\boldsymbol{x}_t}\log q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})d\boldsymbol{x}_{t-1} \\ &- \int \boldsymbol{x}_{t-1} q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)\nabla_{\boldsymbol{x}_t}\log q(\boldsymbol{x}_t)d\boldsymbol{x}_{t-1}\end{aligned} \tag{24}$$

As forward diffusion process is Gaussian process $q(\boldsymbol{x}_t;\boldsymbol{x}_{t-1}) = \mathcal{N}(\boldsymbol{x}_t|\hat{\alpha}_t\boldsymbol{x}_{t-1}, \hat{\sigma}_t^2 I)$, the partial derivative of $\log q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ with respect to latent variable $\boldsymbol{x}_t$ is

$$\nabla_{\boldsymbol{x}_t}\log q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) = -\frac{\boldsymbol{x}_t - \hat{\alpha}_t\boldsymbol{x}_{t-1}}{\hat{\sigma}_t^2}. \tag{25}$$

The first term of Equation (24) can be calculated by

$$\begin{aligned}\int \boldsymbol{x}_{t-1} q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)\nabla_{\boldsymbol{x}}\log q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})d\boldsymbol{x}_{t-1} &= \int \boldsymbol{x}_{t-1}(\frac{\hat{\alpha}_t\boldsymbol{x}_{t-1} - \boldsymbol{x}_t}{\hat{\sigma}_t^2})q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)d\boldsymbol{x}_{t-1} \\ &= \frac{\hat{\alpha}_t}{\hat{\sigma}_t^2}\int \boldsymbol{x}_{t-1}(\boldsymbol{x}_{t-1} - \frac{\boldsymbol{x}_t}{\hat{\alpha}_t})q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)d\boldsymbol{x}_{t-1}\end{aligned} \tag{26}$$

Decompose the term $\boldsymbol{x}_{t-1}(\hat{\alpha}_t\boldsymbol{x}_{t-1} - \boldsymbol{x}_t)$ in Equation (26) by

$$\begin{aligned}&\boldsymbol{x}_{t-1}(\boldsymbol{x}_{t-1} - \frac{\boldsymbol{x}_t}{\hat{\alpha}_t}) \\ &= (\boldsymbol{x}_{t-1} - \overline{\boldsymbol{x}}_{t-1} + \overline{\boldsymbol{x}}_{t-1})(\boldsymbol{x}_{t-1} - \overline{\boldsymbol{x}}_{t-1} + \overline{\boldsymbol{x}}_{t-1} - \frac{\boldsymbol{x}_t}{\hat{\alpha}_t}) \\ &= (\boldsymbol{x}_{t-1} - \overline{\boldsymbol{x}}_{t-1})(\boldsymbol{x}_{t-1} - \overline{\boldsymbol{x}}_{t-1}) + \overline{\boldsymbol{x}}_{t-1}(\boldsymbol{x}_{t-1} - \overline{\boldsymbol{x}}_{t-1}) \\ &\quad + (\boldsymbol{x}_{t-1} - \overline{\boldsymbol{x}}_{t-1})(\overline{\boldsymbol{x}}_{t-1} - \frac{\boldsymbol{x}_t}{\hat{\alpha}_t}) + \overline{\boldsymbol{x}}_{t-1}(\overline{\boldsymbol{x}}_{t-1} - \frac{\boldsymbol{x}_t}{\hat{\alpha}_t}).\end{aligned} \tag{27}$$

Integrating term by term gives

$$\int (\boldsymbol{x}_{t-1} - \overline{\boldsymbol{x}}_{t-1})(\boldsymbol{x}_{t-1} - \overline{\boldsymbol{x}}_{t-1})q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)d\boldsymbol{x}_{t-1} = \Sigma_{t-1|t}(\boldsymbol{x}), \tag{28}$$

$$\int \overline{\boldsymbol{x}}_{t-1}(\boldsymbol{x}_{t-1} - \overline{\boldsymbol{x}}_{t-1})q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)d\boldsymbol{x}_{t-1} = 0, \tag{29}$$

$$\int (\boldsymbol{x}_{t-1} - \overline{\boldsymbol{x}}_{t-1})(\overline{\boldsymbol{x}}_{t-1} - \boldsymbol{x}_t)q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)d\boldsymbol{x}_{t-1} = 0, \tag{30}$$

$$\int \overline{\boldsymbol{x}}_{t-1}(\overline{\boldsymbol{x}}_{t-1} - \frac{\boldsymbol{x}_t}{\hat{\alpha}_t})q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)d\boldsymbol{x}_{t-1} = \overline{\boldsymbol{x}}_{t-1}(\overline{\boldsymbol{x}}_{t-1} - \frac{\boldsymbol{x}_t}{\hat{\alpha}_t}). \tag{31}$$

Therefore, the Equation (26), which is the first term of Equation (24):

$$\int \boldsymbol{x}_{t-1} q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)\nabla_{\boldsymbol{x}}\log q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})d\boldsymbol{x}_{t-1} = \frac{\hat{\alpha}_t}{\hat{\sigma}_t^2}\big(\Sigma_{t-1|t}(\boldsymbol{x}) + \overline{\boldsymbol{x}}_{t-1}(\overline{\boldsymbol{x}}_{t-1} - \frac{\boldsymbol{x}_t}{\hat{\alpha}_t})\big). \tag{32}$$

The second term of Equation (24) is calculated by

$$\int \boldsymbol{x}_{t-1} q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)\nabla_{\boldsymbol{x}_t}\log q(\boldsymbol{x}_t)d\boldsymbol{x}_{t-1} = \overline{\boldsymbol{x}}_{t-1}\nabla_{\boldsymbol{x}_t}\log q(\boldsymbol{x}_t). \tag{33}$$

Decompose the term $\nabla_{\boldsymbol{x}_t} \log q(\boldsymbol{x}_t)$ by

$$\begin{aligned}
\nabla_{\boldsymbol{x}} \log q(\boldsymbol{x}_t) &= \frac{\nabla_{\boldsymbol{x}_t} q(\boldsymbol{x}_t)}{q(\boldsymbol{x}_t)} \\
&= \frac{\int \nabla_{\boldsymbol{x}_t} q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) q(\boldsymbol{x}_{t-1}) d\boldsymbol{x}_{t-1}}{q(\boldsymbol{x}_t)} \\
&= \frac{\int q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) \nabla_{\boldsymbol{x}_t} \log q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) q(\boldsymbol{x}_{t-1}) d\boldsymbol{x}_{t-1}}{q(\boldsymbol{x}_t)} \\
&= \frac{\int q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) \frac{\hat{\alpha}_t}{\hat{\sigma}_t^2} (\boldsymbol{x}_{t-1} - \frac{\boldsymbol{x}_t}{\hat{\alpha}_t}) q(\boldsymbol{x}_{t-1}) d\boldsymbol{x}_{t-1}}{q(\boldsymbol{x}_t)} \\
&= \int \frac{q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) q(\boldsymbol{x}_{t-1})}{q(\boldsymbol{x}_t)} \frac{\hat{\alpha}_t}{\hat{\sigma}_t^2} (\boldsymbol{x}_{t-1} - \frac{\boldsymbol{x}_t}{\hat{\alpha}_t}) d\boldsymbol{x}_{t-1} \\
&= \frac{\hat{\alpha}_t}{\hat{\sigma}_t^2} \int q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t) (\boldsymbol{x}_{t-1} - \frac{\boldsymbol{x}_t}{\hat{\alpha}_t}) d\boldsymbol{x}_{t-1} \\
&= \frac{\hat{\alpha}_t}{\hat{\sigma}_t^2} (\overline{\boldsymbol{x}}_{t-1} - \frac{\boldsymbol{x}_t}{\hat{\alpha}_t})
\end{aligned} \tag{34}$$

Therefore, the second term of Equation (24)

$$\int \boldsymbol{x}_{t-1} q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t) \nabla_{\boldsymbol{x}} \log q(\boldsymbol{x}_t) d\boldsymbol{x}_{t-1} = \frac{\hat{\alpha}_t}{\hat{\sigma}_t^2} \overline{\boldsymbol{x}}_{t-1} (\overline{\boldsymbol{x}}_{t-1} - \frac{\boldsymbol{x}_t}{\hat{\alpha}_t}) \tag{35}$$

Substitute the first term and second term in Equation (24) by Equation (32) and Equation (35) to achieve

$$\nabla_{\boldsymbol{x}} \mu_{t-1|t}(\boldsymbol{x}) = \frac{\hat{\alpha}_t}{\hat{\sigma}_t^2} \Sigma_{t-1|t}(\boldsymbol{x}) \tag{36}$$

Therefore, the Hessian matrix in Equation (15) is

$$\nabla_{\boldsymbol{x}}^2 \log q_t(\boldsymbol{x}) = \frac{\hat{\alpha}_t^2}{\hat{\sigma}_t^4} \Sigma_{t-1|t}(\boldsymbol{x}) - \frac{1}{\hat{\sigma}_t^2} \mathbf{I} \tag{37}$$

C MORE EXPERIMENTS

Table 4 and Table 5 report results on LSUN bedroom 256×256 and ImageNet 256×256 across NEF from 3 to 20. Our method yields consistent gains at every NEF, improving FID over all baselines without extra calculation.

Table 4: Experiments on LSUN bedroom 256×256 guided with various NFE. The quality of images is measured by FID. We use the implementation from the DPM-Solver-v3 repository. * We borrow results reported in AMED-Plugin, "(NEF)" denotes the actual NEF corresponding to the reported FID.

NFE	3	4	5	6	8	10	12	15	20
AMED-Plugin*	101.5	-	25.68	8.63 (7)	7.82 (9)	-	-	-	-
UniPC	109.31	39.89	13.99	6.55	4.00	3.57	3.35	3.18	3.07
+ Ours	59.27	20.08	8.96	5.48	3.79	3.38	3.19	3.08	3.02
DPM-Solver-v3	64.43	19.17	8.96	5.13	3.56	3.20	3.12	3.12	3.10
+ Ours	54.41	15.49	6.77	4.55	3.45	3.14	3.05	3.04	3.04

D MULTI UPDATE

In Table 6, we examine the effect of applying multiple updates within ParaDiGMS. In particular, we compare the latent variable before each Picard iteration with the updated latent variable after the iteration when our refinement method is applied.

Table 5: Experiments on ImageNet 256×256 guided with various NFE. The quality of images is measured by FID. We use the implementation from the DPM-Solver-v3 repository. * We borrow the results (NFE∈[5,20]) reported in DPM-Solver-v3.

NFE	3	4	5	6	8	10	12	15	20
UniPC*	52.21	24.53	15.62	11.91	9.29	8.35	7.95	7.64	7.44
+ Ours	50.79	22.60	14.38	11.07	9.06	8.26	7.97	7.74	7.46
DPM-Solver-v3*	65.38	26.37	15.10	11.39	8.96	8.27	7.94	7.62	7.39
+ Ours	60.80	24.34	14.53	10.84	8.86	8.08	7.85	7.55	7.46

Table 6: Experiments on multi-update

Number of Iters	1	2	3	4	5
Difference	51.7352	1.2923	1.2919	1.2919	1.2919

E Ablation study

Table 7 presents an ablation study on the choice of the parameter λ. We compare different strategies, including the default setting $-Hm(\boldsymbol{x}_{t-1|t}) = \frac{1}{\sigma_{t-1}^2}$, fixed constant values, and empirical values collected from experiments.

Table 7: Experiments on LSUN bedroom 256×256 guided with various NFE. The quality of images is measured by FID. We use the implementation from the DPM-Solver-v3 repository.

NFE	3	4	5	6	8	10	12	15	20
DPM-Solver-v3	64.43	19.17	8.96	5.13	3.56	3.20	3.12	3.12	3.10
+ Ours	54.41	15.49	6.77	4.55	3.45	3.14	3.05	3.04	3.04
+ Ours $\lambda = 1$	63.42	18.71	7.42	4.76	3.52	3.18	3.09	3.08	3.06
+ Ours $\lambda = 0.5$	49.44	14.68	6.74	4.37	3.42	3.14	3.05	3.05	3.05
+ Ours empirical	53.17	15.61	6.89	4.61	3.48	3.16	3.07	3.06	3.06

F Checkpoints

For all experiments, we employ publicly released model checkpoints to ensure reproducibility and fair comparison. In particular, we reference the official implementations of DPM-Solver-v3 and ParaDiGMS, and list the specific checkpoints used in Table 8.

G More Visualization

Figure 4 presents additional visualizations of generated images on the LSUN Bedroom 256×256. Integrating our plug-in refinement yields images with richer details and improved semantic coherence compared to the baselines.

Table 8: Checkpoints of models in experiments

Datasets	URL
ImageNet 256×256 guided	`https://openaipublic.blob.core.windows.net/diffusion/jul-2021/256x256_diffusion.pt`
LSUN bedroom 256×256	`https://openaipublic.blob.core.windows.net/diffusion/jul-2021/lsun_bedroom.pt`
Stable Diffusion	`https://huggingface.co/CompVis/stable-diffusion-v-1-4-original/resolve/main/sd-v1-4.ckpt`
Stable Diffusion v2	`https://huggingface.co/stabilityai/stable-diffusion-2`

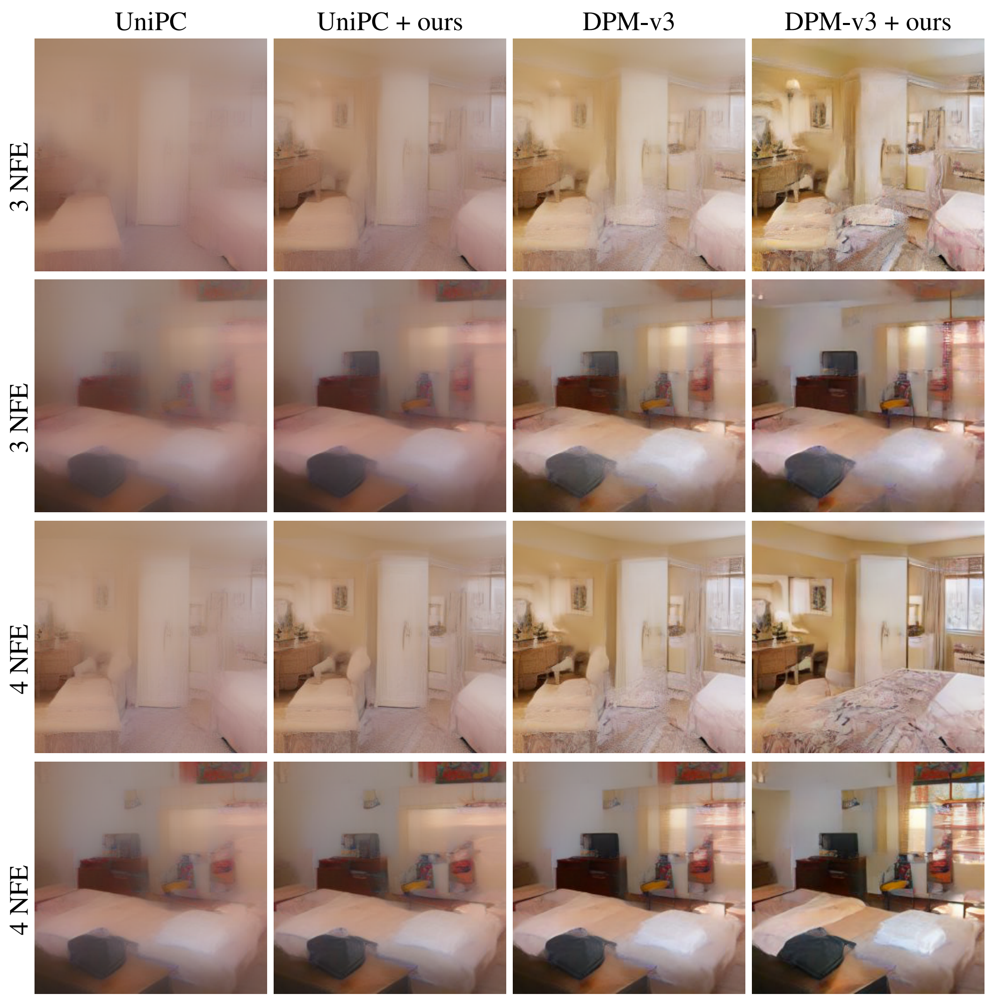

Figure 4: Visualization of generated images using LUSN bedroom 256×256of UniPC, DPM-Solver-v3, and our method integrated with 3 and 4 NFEs. We refer to DPM-Solver-v3 as DPM-v3.