A EQUIVALENT DEFINITIONS OF DDIM AND DDPM

The DDPM and DDIM samplers are usually described in a different coordinate system z_t defined by parameters $\bar{\alpha}_t$ and the following relations , where the noise model is defined by a schedule $\bar{\alpha}_t$:

$$y \approx \sqrt{\bar{\alpha}_t} z + \sqrt{1-\bar{\alpha}_t}\epsilon, \tag{13}$$

with the estimate $\hat{z}_0^t := \hat{z}_0(z_t, t)$ given by

$$\hat{z}_0(y,t) := \frac{1}{\sqrt{\bar{\alpha}_t}}(y - \sqrt{1-\bar{\alpha}_t}\epsilon'_\theta(y,t)). \tag{14}$$

We have the following conversion identities between the x and z coordinates:

$$x_0 = z_0, \quad x_t = z_t/\sqrt{\bar{\alpha}_t}, \quad \sigma_t = \sqrt{\frac{1-\bar{\alpha}_t}{\bar{\alpha}_t}}, \quad \epsilon_\theta(y, \sigma_t) = \epsilon'_\theta(y/\sqrt{\bar{\alpha}_t}, t). \tag{15}$$

While this change-of-coordinates is used in Song et al. (2020a, Section 4.3) and in Karras et al. (2022)–and hence not new– we rigorously prove equivalence of the DDIM and DDPM samplers given in Section 2 with their original definitions.

DDPM Given initial z_N, the DDPM sampler constructs the sequence

$$z_{t-1} = \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t}\hat{z}_0^t + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} z_t + \sqrt{\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}(1-\alpha_t)}w_t, \tag{16}$$

where $\alpha_t := \bar{\alpha}_t/\bar{\alpha}_{t-1}$ and $w_t \sim \mathcal{N}(0, I)$. This is interpreted as sampling z_{t-1} from a Gaussian distribution conditioned on z_t and $\hat{z}_0^t$ (Ho et al., 2020).

Proposition A.1 (DDPM change of coordinates)**.** *The sampling update* (3) *is equivalent to the update* (16) *under the change of coordinates* (15)*.*

Proof. First we write (3) in terms of z_t, $\epsilon'_\theta(z_t, t)$ and w_t using (14):

$$\begin{aligned} z_{t-1} &= \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{\sqrt{\bar{\alpha}_t}(1-\bar{\alpha}_t)}\left(z_t - \sqrt{1-\bar{\alpha}_t}\epsilon'_\theta(z_t,t)\right) + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} z_t + \sqrt{\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}(1-\alpha_t)}w_t \\ &= \frac{z_t}{\sqrt{\alpha_t}} + \frac{\alpha_t - 1}{\sqrt{\alpha_t(1-\bar{\alpha}_t)}}\epsilon'_\theta(z_t,t) + \sqrt{\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}(1-\alpha_t)}w_t. \end{aligned}$$

Next we divide both sides by $\sqrt{\bar{\alpha}_{t-1}}$ and change z_t and z_{t-1} to x_t and x_{t-1}:

$$x_{t-1} = x_t + \frac{\alpha_t - 1}{\sqrt{\bar{\alpha}_t(1-\bar{\alpha}_t)}}\epsilon_\theta(x_t, \sigma_t) + \sqrt{\frac{1-\bar{\alpha}_{t-1}}{\bar{\alpha}_{t-1}}\frac{1-\alpha_t}{1-\bar{\alpha}_t}}w_t.$$

Now if we define

$$\begin{aligned} \eta &:= \sqrt{\frac{1-\bar{\alpha}_{t-1}}{\bar{\alpha}_{t-1}}\frac{1-\alpha_t}{1-\bar{\alpha}_t}} = \sigma_{t-1}\sqrt{\frac{1-\bar{\alpha}_t/\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}}, \\ \sigma_{t'} &:= \sqrt{\sigma_{t-1}^2 - \eta^2} = \sigma_{t-1}\sqrt{\frac{\bar{\alpha}_t(1/\bar{\alpha}_{t-1}-1)}{1-\bar{\alpha}_t}} = \frac{\sigma_{t-1}^2}{\sigma_t}, \end{aligned}$$

it remains to check that

$$\sigma_{t'} - \sigma_t = \frac{\sigma_{t-1}^2 - \sigma_t^2}{\sigma_t} = \frac{1/\bar{\alpha}_{t-1} - 1/\bar{\alpha}_t}{\sqrt{1-\bar{\alpha}_t}/\sqrt{\bar{\alpha}_t}} = \frac{\alpha_t - 1}{\sqrt{\bar{\alpha}_t(1-\bar{\alpha}_t)}}.$$

□

DDIM Given initial z_N, the DDIM sampler constructs the sequence

$$z_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\hat{z}_0^t + \sqrt{1-\bar{\alpha}_{t-1}}\epsilon'_\theta(z_t, t), \tag{17}$$

i.e., it estimates $\hat{z}_0^t$ from z_t and then constructs z_{t-1} by simply updating $\bar{\alpha}_t$ to $\bar{\alpha}_{t-1}$. This sequence can be equivalently expressed in terms of $\hat{z}_0^t$ as

$$z_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\hat{z}_0^t + \sqrt{\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}}(z_t - \sqrt{\bar{\alpha}_t}\hat{z}_0^t). \tag{18}$$

Proposition A.2 (DDIM change of coordinates)**.** *The sampling update* (4) *is equivalent to the update* (18) *under the change of coordinates* (15)*.*

Proof. First we write (17) in terms of z_t and $\epsilon'_\theta(z_t, t)$ using (14):

$$z_{t-1} = \sqrt{\frac{\bar{\alpha}_{t-1}}{\bar{\alpha}_t}}z_t + \left(\sqrt{1-\bar{\alpha}_{t-1}} - \sqrt{\frac{\bar{\alpha}_{t-1}}{\bar{\alpha}_t}}\sqrt{1-\bar{\alpha}_t}\right)\epsilon'_\theta(z_t, t).$$

Next we divide both sides by $\sqrt{\bar{\alpha}_{t-1}}$ and change z_t and z_{t-1} to x_t and x_{t-1}:

$$\begin{aligned} x_{t-1} &= x_t + \left(\sqrt{\frac{1-\bar{\alpha}_{t-1}}{\bar{\alpha}_{t-1}}} - \sqrt{\frac{\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}}\right)\epsilon_\theta(x_t, \sigma_t) \\ &= x_t + (\sigma_{t-1} - \sigma_t)\epsilon_\theta(x_t, \sigma_t). \end{aligned}$$

□

B Formal Comparison of Denoising and Projection

Our proof uses local Lipschitz continuity of the projection operator, stated formally as follows.

Proposition B.1 (Theorem 6.2(vi), Chapter 6 of Delfour & Zolésio (2011))**.** *Suppose* $0 < \operatorname{reach}(\mathcal{K}) < \infty$. *Consider* $h > 0$ *and* $x, y \in \mathbb{R}^n$ *satisfying* $0 < h < \operatorname{reach}(\mathcal{K})$ *and* $\operatorname{dist}_\mathcal{K}(x) \le h$ *and* $\operatorname{dist}_\mathcal{K}(y) \le h$*. Then the projection map satisfies* $\|\operatorname{proj}_\mathcal{K}(y) - \operatorname{proj}_\mathcal{K}(x)\| \le \frac{\operatorname{reach}(\mathcal{K})}{\operatorname{reach}(\mathcal{K})-h}\|y - x\|$.

Decomposing random noise $\sigma\epsilon$ as

$$\sigma\epsilon = w_N + w_T \tag{19}$$

for $w_N \in N_\mathcal{K}(x_0)$ and $w_T \in N_\mathcal{K}(x_0)^\perp$ and using Lemma 3.1 allows us to show that $\operatorname{proj}_\mathcal{K}(x_\sigma) \approx x_0$.

Theorem B.1 (Denoising vs Projection)**.** *Fix* $\sigma > 0$ *and suppose* $\mathcal{K}$ *and* $t > 0$ *satisfies* $\operatorname{reach}(\mathcal{K}) > \sigma(\sqrt{n} + t)$*. Given* $x_0 \in \mathcal{K}$ *and* $\epsilon \sim \mathcal{N}(0, I)$*, let* $x_\sigma = x_0 + \sigma\epsilon$ *and* $w := \sigma\epsilon = w_N + w_T$ *by the decomposition* (19)*. The following statements hold with probability at least* $1 - \exp(-\alpha t^2)$*, where* $\alpha > 0$ *is an absolute constant.*

- *(Backward error)* $x_0 = \operatorname{proj}_\mathcal{K}(x_\sigma - w_T)$.
- *(Forward error)* $\|\operatorname{proj}_\mathcal{K}(x_\sigma) - x_0\| \le C\sigma(\sqrt{d} + t)$*, where* $C = \frac{\operatorname{reach}(\mathcal{K})}{\operatorname{reach}(\mathcal{K}) - \sigma(\sqrt{n}+t)}$.

Proof. Let $B \in \mathbb{R}^{n\times d}$ denote an orthonormal basis for $N_\mathcal{K}(x_0)^\perp$, such that $w_T = BB^T w$, $\|w_T\| = \|B^T w\|$ and we have

$$\mathbf{E}[\|w_T\|^2] = \mathbf{E}[\|B^T w\|]^2 + \operatorname{Tr}\operatorname{cov}(B^T w) = \operatorname{Tr}\operatorname{cov}(B^T w) = \sigma^2 \operatorname{Tr} B^T B = \sigma^2 d. \tag{20}$$

Using a standard concentration inequality (Vershynin, 2018, page 44, Equation 3.3), we get that for a universal constant α, with probability at least $1 - \exp(-\alpha t^2)$, we have $\|\epsilon\| \le \sqrt{n} + t$ and $\|w_T\| \le \sigma(\sqrt{d} + t)$. Using Lemma 3.1 and the fact that $\|w_N\| \le \|\sigma\epsilon\| \le \sigma(\sqrt{n} + t) < \operatorname{reach}(\mathcal{K})$, we get

$$\operatorname{proj}(x_\sigma - w_T) = \operatorname{proj}(x_0 + w_N) = x_0,$$

proving the first statement. To prove the second statement, we observe that

$$\begin{aligned}\|\operatorname{proj}(x_\sigma) - x_0\| &= \|\operatorname{proj}(x_0 + w_N + w_T) - x_0\| \\ &= \|\operatorname{proj}(x_0 + w_N) - x_0 + \operatorname{proj}(x_0 + w_N + w_T) - \operatorname{proj}(x_0 + w_N)\| \\ &= \|\operatorname{proj}(x_0 + w_N) - \operatorname{proj}(x_0 + w_N + w_T)\| \\ &\leq C\|w_T\| \\ &\leq C\sigma(\sqrt{n} + t)\end{aligned}$$

where the second-to-last inequality comes from Proposition B.1, the assumption that $\operatorname{reach}(\mathcal{K}) > \sigma(\sqrt{n}+t)$, and the inequalities $\operatorname{dist}_\mathcal{K}(x_0+w_N) \leq \|w_N\| \leq \sigma(\sqrt{n}+t)$ and $\operatorname{dist}_\mathcal{K}(x_0+w_N+w_T) \leq \|w\| \leq \sigma(\sqrt{n}+t)$. □

C DDIM WITH PROJECTION ERROR ANALYSIS

C.1 PROOF OF THEOREM 4.1

Make the inductive hypothesis that $\operatorname{dist}(x_t) = \sqrt{n}\sigma_t$. From the definition of DDIM (4), we have

$$x_{t-1} = x_t + (\frac{\sigma_{t-1}}{\sigma_t} - 1)\sigma_t\epsilon_\theta(x_t, \sigma_t).$$

Under Assumption 1 and the inductive hypothesis, we conclude

$$\begin{aligned}x_{t-1} &= x_t + (\frac{\sigma_{t-1}}{\sigma_t} - 1)\nabla f(x_t) \\ &= x_t - \beta_t\nabla f(x_t)\end{aligned}$$

Using Lemma 4.1 we have that

$$\operatorname{dist}(x_{t-1}) = (1 - \beta_t)\operatorname{dist}(x_t) = \frac{\sigma_{t-1}}{\sigma_t}\operatorname{dist}(x_t) = \sqrt{n}\sigma_{t-1}$$

The base case holds by assumption, proving the claim.

C.2 PROOF OF LEMMA 4.1

Letting $x_0 = \operatorname{proj}_\mathcal{K}(x)$ and noting $\nabla f(x) = x - x_0$, we have

$$\begin{aligned}\operatorname{dist}_\mathcal{K}(x_+) &= \operatorname{dist}_\mathcal{K}(x + \beta(x_0 - x)) \\ &= \|x + \beta(x_0 - x) - x_0\| \\ &= \|(x - x_0)(1 - \beta)\| \\ &= (1 - \beta)\operatorname{dist}_\mathcal{K}(x)\end{aligned}$$

C.3 PROOF OF LEMMA 4.2

By (Delfour & Zolésio, 2011, Chapter 6, Theorem 2.1), $|\operatorname{dist}_\mathcal{K}(u) - \operatorname{dist}_\mathcal{K}(v)| \leq \|u - v\|$, which is equivalent to

$$\operatorname{dist}_\mathcal{K}(u) - \operatorname{dist}_\mathcal{K}(v) \leq \|u - v\|, \operatorname{dist}_\mathcal{K}(v) - \operatorname{dist}_\mathcal{K}(u) \leq \|u - v\|.$$

Rearranging proves the claim.

C.4 PROOF OF LEMMA 4.3

We first restate the full version of Lemma 4.3.

Lemma C.1. *For $\mathcal{K} \subseteq \mathbb{R}^n$, let $f(x) := \frac{1}{2}\operatorname{dist}_\mathcal{K}(x)^2$. The following statements hold.*

(a) If $x_+ = x - \beta(\nabla f(x) + e)$ for e satisfying $\|e\| \leq \eta\operatorname{dist}_\mathcal{K}(x)$ and $0 \leq \beta \leq 1$, then

$$(1 - \beta(\eta + 1))\operatorname{dist}_\mathcal{K}(x) \leq \operatorname{dist}_\mathcal{K}(x_+) \leq (1 + \beta(\eta - 1))\operatorname{dist}_\mathcal{K}(x).$$

(b) If $x_{t-1} = x_t - \beta_t(\nabla f(x_t) + e_t)$ *for* e_t *satisfying* $\|e_t\| \leq \eta \mathrm{dist}_{\mathcal{K}}(x_t)$ *and* $0 \leq \beta_t \leq 1$*, then*

$$\mathrm{dist}_{\mathcal{K}}(x_N)\prod_{i=t}^{N}(1-\beta_i(\eta+1)) \leq \mathrm{dist}_{\mathcal{K}}(x_{t-1}) \leq \mathrm{dist}_{\mathcal{K}}(x_N)\prod_{i=t}^{N}(1+\beta_i(\eta-1).)$$

For Item (a) we apply Lemma 4.2 at points $u = x_+$ and $v = x - \beta\nabla f(x)$. We also use $\mathrm{dist}(v) = (1-\beta)\mathrm{dist}_{\mathcal{K}}(x)$, since $0 \leq \beta \leq 1$, to conclude that

$$(1-\beta)\mathrm{dist}_{\mathcal{K}}(x) - \beta\|e\| \leq \mathrm{dist}_{\mathcal{K}}(x_+) \leq (1-\beta)\mathrm{dist}_{\mathcal{K}}(x) + \beta\|e\|.$$

Using the assumption that $\|e\| \leq \eta\mathrm{dist}_{\mathcal{K}}(x)$ gives

$$(1-\beta-\eta\beta)\mathrm{dist}_{\mathcal{K}}(x) \leq \mathrm{dist}_{\mathcal{K}}(x_+) \leq (1-\beta+\eta\beta)\mathrm{dist}_{\mathcal{K}}(x)$$

Simplifying completes the proof. Item (b) follows from Item (a) and induction.

C.5 Proof of Theorem 4.2

We first state and prove an auxillary theorem:

Theorem C.1. *Suppose Assumption 2 holds for* $\nu \geq 1$ *and* $\eta > 0$*. Given* x_N *and* $\{\beta_t, \sigma_t\}_{i=1}^N$*, recursively define* $x_{t-1} = x_t + \beta_t\sigma_t\epsilon_\theta(x_t, t)$ *and suppose that* $\mathrm{proj}_{\mathcal{K}}(x_t)$ *is a singleton for all* t*. Finally, suppose that* $\{\beta_t, \sigma_t\}_{i=1}^N$ *satisfies* $\frac{1}{\nu}\mathrm{dist}_{\mathcal{K}}(x_N) \leq \sqrt{n}\sigma_N \leq \nu\mathrm{dist}_{\mathcal{K}}(x_N)$ *and*

$$\frac{1}{\nu}\mathrm{dist}_{\mathcal{K}}(x_N)\prod_{i=t}^{N}(1+\beta_i(\eta-1)) \leq \sqrt{n}\sigma_{t-1} \leq \nu\mathrm{dist}_{\mathcal{K}}(x_N)\prod_{i=t}^{N}(1-\beta_i(\eta+1)). \tag{21}$$

The following statements hold.

- $\mathrm{dist}_{\mathcal{K}}(x_N)\prod_{i=t}^{N}(1-\beta_i(\eta+1)) \leq \mathrm{dist}_{\mathcal{K}}(x_{t-1}) \leq \mathrm{dist}_{\mathcal{K}}(x_N)\prod_{i=t}^{N}(1+\beta_i(\eta-1))$
- $\frac{1}{\nu}\mathrm{dist}_{\mathcal{K}}(x_{t-1}) \leq \sqrt{n}\sigma_{t-1} \leq \nu\mathrm{dist}_{\mathcal{K}}(x_{t-1})$

Proof. Since $\mathrm{proj}_{\mathcal{K}}(x_t)$ is a singleton, $\nabla f(x_t)$ exists. Hence, the result will follow from (7) in Lemma 4.3 if we can show that $\|\beta_t\sigma_t\epsilon_\theta(x_t,t) - \nabla f(x_t)\| \leq \eta\mathrm{dist}_{\mathcal{K}}(x_t)$. Under Assumption 2, it suffices to show that

$$\frac{1}{\nu}\mathrm{dist}_{\mathcal{K}}(x_t) \leq \sqrt{n}\sigma_t \leq \nu\mathrm{dist}_{\mathcal{K}}(x_t) \tag{22}$$

holds for all t. We use induction, noting that the base case ($t = N$) holds by assumption. Suppose then that (22) holds for all $t, t+1, \ldots, N$. By Lemma 4.3 and Assumption 2, we have

$$\mathrm{dist}_{\mathcal{K}}(x_N)\prod_{i=t}^{N}(1-\beta_i(\eta+1)) \leq \mathrm{dist}_{\mathcal{K}}(x_{t-1}) \leq \mathrm{dist}_{\mathcal{K}}(x_N)\prod_{i=t}^{N}(1+(\eta-1)\beta_i)$$

Combined with (21) shows

$$\frac{1}{\nu}\mathrm{dist}_{\mathcal{K}}(x_{t-1}) \leq \sqrt{n}\sigma_{t-1} \leq \nu\mathrm{dist}_{\mathcal{K}}(x_{t-1}),$$

proving the claim. □

The proof of Theorem 4.2 follows that of Theorem C.1 by additionally observing $\eta < 1$ implies that $\mathrm{dist}_{\mathcal{K}}(x_t) < \mathrm{reach}(\mathcal{K})$ for all t, which implies $\mathrm{proj}_{\mathcal{K}}(x_t)$ is a singleton.

C.6 Proof of Theorem 4.3

Assuming constant step-size $\beta_i = \beta$ and dividing (8) by $\prod_{i=1}^{N}(1-\beta)$ gives the conditions

$$\left(1+\eta\frac{\beta}{1-\beta}\right)^N \leq \nu, \qquad \left(1-\eta\frac{\beta}{1-\beta}\right)^N \geq \frac{1}{\nu}.$$

Rearranging and defining $a = \eta\frac{\beta}{1-\beta}$ and $b = \nu^{\frac{1}{N}}$ gives

$$a \leq b - 1, \qquad a \leq 1 - b^{-1}.$$

Since $b - 1 - (1 - b^{-1}) = b + b^{-1} - 2 \geq 0$ for all $b > 0$, we conclude $a \leq b - 1$ holds if $a \leq 1 - b^{-1}$ holds. We therefore consider the second inequality $\eta\frac{\beta}{1-\beta} \leq 1 - \nu^{-1/N}$, noting that it holds for all $0 \leq \beta < 1$ if and only if $0 \leq \beta \leq \frac{k}{1+k}$ for $k = \frac{1}{\eta}(1 - \nu^{-1/N})$, proving the claim.

C.7 PROOF OF THEOREM 4.4

The value of σ_0/σ_N follows from the definition of σ_t and and the upper bound for $\text{dist}_{\mathcal{K}}(x_0)/\text{dist}_{\mathcal{K}}(x_N)$ follows from Theorem 4.3. We introduce the parameter μ to get a general form of the expression inside the limit:

$$(1-\mu\beta_{*,N})^N = \left(1-\mu\frac{1-\nu^{-1/N}}{\eta+1-\nu^{-1/N}}\right)^N.$$

Next we take the limit using L'Hôpital's rule:

$$\begin{aligned}
\lim_{N\to\infty}\left(1-\mu\frac{1-\nu^{-1/N}}{\eta+1-\nu^{-1/N}}\right)^N &= \exp\left(\lim_{N\to\infty}\log\left(1-\mu\frac{1-\nu^{-1/N}}{\eta+1-\nu^{-1/N}}\right)/(1/N)\right)\\
&= \exp\left(\lim_{N\to\infty}\frac{\eta\mu\log(\nu)}{(\nu^{-1/N}-\eta-1)(\nu^{1/N}(\eta-\mu+1)+\mu-1)}\right)\\
&= \exp\left(-\frac{\mu\log(\nu)}{\eta}\right)\\
&= (1/\nu)^{\mu/\eta}.
\end{aligned}$$

For the first limit, we set $\mu = 1$ to get

$$\lim_{N\to\infty}(1-\beta_{*,N})^N = (1/\nu)^{1/\eta}.$$

For the second limit, we set $\mu = 1-\eta$ to get

$$\lim_{N\to\infty}(1+(\eta-1)\beta_{*,N})^N = (1/\nu)^{\frac{1-\eta}{\eta}}.$$

C.8 DENOISER ERROR

Assumption 2 places a condition directly on the approximation of $\nabla f(x)$, where $f(x) := \frac{1}{2}\text{dist}_{\mathcal{K}}(x)$, that is jointly obtained from σ_t and the denoiser ϵ_θ. We prove this assumption holds under a direct assumption on $\nabla\text{dist}_{\mathcal{K}}(x)$, which is easier to verify in practice.

Assumption 3. *There exists $\nu \geq 1$ and $\eta > 0$ such that if $\frac{1}{\nu}\text{dist}_{\mathcal{K}}(x) \leq \sqrt{n}\sigma_t \leq \nu\text{dist}_{\mathcal{K}}(x)$ then $\|\epsilon_\theta(x,t) - \sqrt{n}\nabla\text{dist}_{\mathcal{K}}(x)\| \leq \eta$*

Lemma C.2. *If Assumption 3 holds with (ν,η), then Assumption 2 holds with $(\hat{\nu},\hat{\eta})$, where $\hat{\eta} = \frac{1}{\sqrt{n}}\eta\nu + \max(\nu-1, 1-\frac{1}{\nu})$ and $\hat{\nu} = \nu$.*

Proof. Multiplying the error-bound on ϵ_θ by σ_t and using $\sqrt{n}\sigma_t \leq \nu\text{dist}_{\mathcal{K}}(x)$ gives

$$\|\sigma_t\epsilon_\theta(x,t) - \sqrt{n}\sigma_t\nabla\text{dist}_{\mathcal{K}}(x)\| \leq \eta\sigma_t \leq \eta\nu\frac{1}{\sqrt{n}}\text{dist}_{\mathcal{K}}(x)$$

Defining $C = \sqrt{n}\sigma_t - \text{dist}_{\mathcal{K}}(x)$ and simplifying gives

$$\begin{aligned}
\eta\nu\frac{1}{\sqrt{n}}\text{dist}_{\mathcal{K}}(x) &\geq \|\sigma_t\epsilon_\theta(x,t) - \sqrt{n}\sigma_t\nabla\text{dist}_{\mathcal{K}}(x)\|\\
&= \|\sigma_t\epsilon_\theta(x,t) - \nabla f(x) - C\nabla\text{dist}_{\mathcal{K}}(x)\|\\
&\geq \|\sigma_t\epsilon_\theta(x,t) - \nabla f(x)\| - \|C\nabla\text{dist}_{\mathcal{K}}(x)\|\\
&= \|\sigma_t\epsilon_\theta(x,t) - \nabla f(x)\| - |C|
\end{aligned}$$

Since $(\frac{1}{\nu}-1)\text{dist}_{\mathcal{K}}(x) \leq C \leq (\nu-1)\text{dist}_{\mathcal{K}}(x)$ and $\nu \geq 1$, the Assumption 2 error bound holds for the claimed $\hat{\eta}$. □

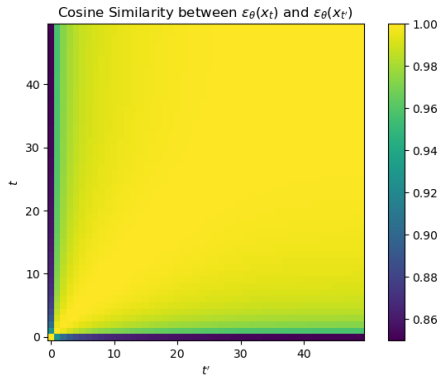

Figure 6: Plot of the cosine similarity between $\epsilon_\theta(x_t, t)$ and $\epsilon_\theta(x_{t'}, t')$ over $N = 50$ steps of DDIM denoising on the CIFAR-10 dataset. Each cell is the average result of 1000 runs.

D Derivation of Gradient Estimation Sampler

To choose W, we make two assumptions on the denoising error: the coordinates $e_t(\epsilon)_i$ and $e_t(\epsilon)_j$ are uncorrelated for all $i \neq j$, and $e_t(\epsilon)_i$ is only correlated with $e_{t+1}(\epsilon)_i$ for all i. In other words, we consider W of the form

$$W = \begin{bmatrix} aI & bI \\ bI & cI \end{bmatrix} \tag{23}$$

and next show that this choice leads to a simple rule for selecting $\bar{\epsilon}$. From the optimality conditions of the quadratic optimization problem (11), we get that

$$\bar{\epsilon}_t = \frac{a+b}{a+c+2b}\epsilon_\theta(x_t, \sigma_t) + \frac{c+b}{a+c+2b}\epsilon_\theta(x_{t+1}, \sigma_{t+1}).$$

Setting $\gamma = \frac{a+b}{a+c+2b}$, we get the update rule (12). When $b \geq 0$, the minimizer $\bar{\epsilon}_t$ is a simple convex combination of denoiser outputs. When $b < 0$, we can have $\gamma < 0$ or $\gamma > 1$, i.e., the weights in (12) can be negative (but still sum to 1). Negativity of the weights can be interpreted as cancelling positively correlated error ($b < 0$) in the denoiser outputs. Also note we can implicitly search over W by directly searching for γ.

E Further Experiments

E.1 Denoising Approximates Projection

We test our interpretation that denoising approximates projection on pretrained diffusion models on the CIFAR-10 dataset. In these experiments, we take a 50-step DDIM sampling trajectory, extract $\epsilon(x_t, \sigma_t)$ for each t and compute the cosine similarity for every pair of $t, t' \in [1, 50]$. The results are plotted in Figure 6. They show that the direction of $\epsilon(x_t, \sigma_t)$ over the entire sampling trajectory is close to the first step's output $\epsilon(x_N, \sigma_N)$. On average over 1000 trajectories, the minimum similarity (typically between the first step when $t = 50$ and last step when $t' = 1$) is 0.85, and for the vast majority (over 80%) of pairs the similarity is > 0.99, showing that the denoiser outputs approximately align in the same direction, validating our intuitive picture in Figure 2.

E.2 Distance Function Properties

We test Assumption 1 and Assumption 2 on pretrained networks. If Assumption 1 is true, then $\|\epsilon_\theta(x_t, \sigma_t)\| \sqrt{n} = \|\nabla \text{dist}_{\mathcal{K}}(x_t)\| = 1$ for every x_t along the DDIM trajectory. In Figure 7a, we plot the distribution of norm of the denoiser $\epsilon_\theta(x_t, \sigma_t)$ over the course of many runs of the DDIM sampler on the CIFAR-10 model for $N = 100$ steps ($t = 1000, 990, \ldots, 20, 10, 0$). This plot shows that $\|\epsilon_\theta(x_t, \sigma_t)\| / \sqrt{n}$ stays approximately constant and is close to 1 until the end of the sampling

(a) Plot of $\|\epsilon_\theta(x_t, \sigma_t)\| / \sqrt{n}$ against t. (b) Plot of $\|\epsilon_\theta(x_0 + \sigma_t\epsilon, \sigma_t) - \epsilon\| / \sqrt{n}$ against t.

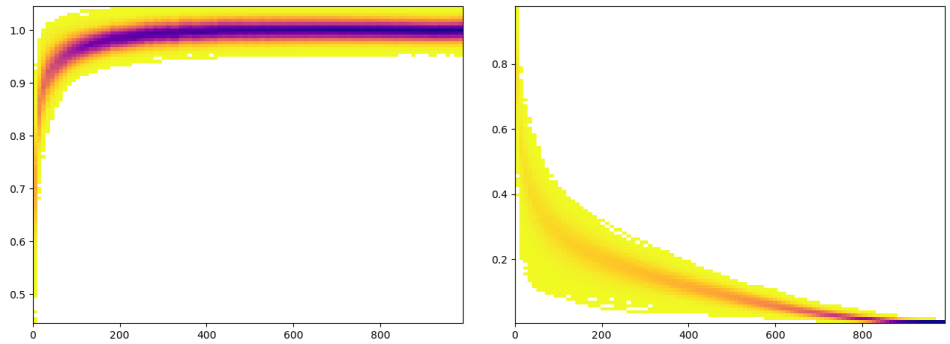

Figure 7: Plots of the norm of the denoiser at different stages of denoising, as well as the ability of the denoiser to accurately predict the added noise as a function of noise added.

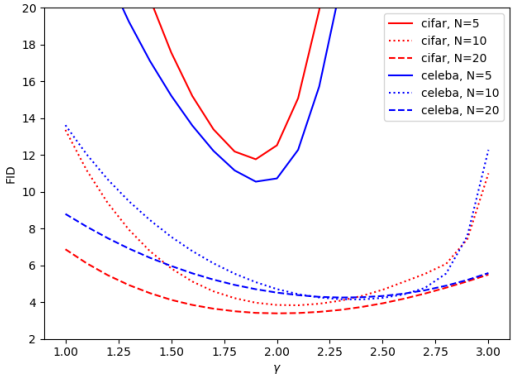

Figure 8: Plot of FID score against γ for our second-order sampling algorithm on the CIFAR-10 and CelebA datasets for $N = 5, 10, 20$ steps.

process. We next test Assumption 3, which implies Assumption 2 by Lemma C.2. We do this by first sampling a fixed noise vector ϵ, next adding different levels of noise σ_t, then using the denoiser to predict $\epsilon_\theta(x_0 + \sigma_t\epsilon, \sigma_t)$. In Figure 7b, we plot the distribution of $\|\epsilon_\theta(x_0 + \sigma_t\epsilon, \sigma_t) - \epsilon\| / \sqrt{n}$ over different levels of t, as a measure of how well the denoiser predicts the added noise.

E.3 Choice of γ

We motivate our choice of $\gamma = 2$ in Algorithm 2 with the following experiment. For varying γ, Figure 8 reports FID scores of our sampler on the CIFAR-10 and CelebA models for $N = 5, 10, 20$ timesteps using the σ_t schedule described in Appendix F.3. As shown, $\gamma \approx 2$ achieves the optimal FID score over different datasets and choices of N.

F Experiment Details

F.1 Pretrained Models

The CIFAR-10 model and architecture were based on that in Ho et al. (2020), and the CelebA model and architecture were based on that in Song et al. (2020a). The specific checkpoints we use are provided by Liu et al. (2022). We also use Stable Diffusion 2.1 provided in `https://huggingface.co/stabilityai/stable-diffusion-2-1`. For the comparison experiments in Figure 1, we implemented our gradient estimation sampler to interface with the Hug-

gingFace diffusers library and use the corresponding implementations of UniPC, DPM++, PNDM and DDIM samplers with default parameters.

F.2 FID SCORE CALCULATION

For the CIFAR-10 and CelebA experiments, we generate 50000 images using our sampler and calculate the FID score using the library in `https://github.com/mseitzer/pytorch-fid`. The statistics on the training dataset were obtained from the files provided by Liu et al. (2022). For the MS-COCO experiments, we generated images from 30k text captions drawn from the validation set, and computed FID with respect to the 30k corresponding images.

F.3 OUR SELECTION OF σ_t

Let $\sigma_1^{\text{DDIM}(N)}$ be the noise level at $t = 1$ for the DDIM sampler with N steps. For the CIFAR-10 and CelebA models, we choose $\sigma_1 = \sqrt{\sigma_1^{\text{DDIM}(N)}}$ and $\sigma_0 = 0.01$. For CIFAR-10 $N = 5, 10, 20, 50$ and CelebA $N = 5$ we choose $\sigma_N = 40$ and for CelebA $N = 10, 20, 50$ we choose $\sigma_N = 80$. For Stable Diffusion, we use the same sigma schedule as that in DDIM.

F.4 TEXT PROMPTS

For the text to image generation in Figure 1, the text prompts used are:

- "A digital Illustration of the Babel tower, 4k, detailed, trending in artstation, fantasy vivid colors"
- "London luxurious interior living-room, light walls"
- "Cluttered house in the woods, anime, oil painting, high resolution, cottagecore, ghibli inspired, 4k"