PROVABLE DIFFUSION-BASED POSTERIOR SAMPLING FOR LINEAR INVERSE PROBLEMS VIA DDIM

Anonymous authors
Paper under double-blind review

ABSTRACT

Diffusion-based methods have achieved remarkable empirical success in solving inverse problems. However, many existing posterior samplers either lack rigorous theoretical guarantees or incur substantial computational overhead. We propose a simple and efficient algorithm, called `Posterior-DDIM`, for solving linear inverse problems with diffusion priors via a DDIM-type sampler. Our method requires only lightweight, coordinate-wise modifications to the standard DDIM update, while explicitly incorporating the measurement model. The key idea is to perform posterior sampling separately along each singular direction of the measurement operator: for each direction, the sampler follows the learned diffusion prior when the observation signal-to-noise ratio (SNR) is below the corresponding diffusion SNR, and switches to a calibrated measurement-based predictor otherwise. We prove that the proposed sampler converges to the Bayesian posterior conditioned on the measurements. Empirical results show that the proposed sampler performs favorably against existing diffusion-based posterior samplers across a range of image restoration tasks, achieving the best performance on the majority of evaluation metrics considered. Overall, our results convert posterior sampling for noisy linear inverse problems to simple coordinate-wise DDIM updates, yielding an efficient, easy-to-implement algorithm with provable posterior consistency. Our code is available at `https://anonymous.4open.science/r/DDIM_inv-E3C1`.

1 INTRODUCTION

Diffusion models (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song et al., 2020; Lai et al., 2025) have emerged as one of the most successful paradigms for modern generative modeling, achieving remarkable performance across a wide range of domains, from image synthesis, video generation, to scientific data generation. Beyond sample generation, their ability to capture complex, high-dimensional data distributions makes them particularly well-suited as flexible, expressive, and fully data-driven priors over the signals of interest, offering a versatile framework for Bayesian inference in inverse problems (Song et al., 2022; Chung et al., 2022; 2023a; Kawar et al., 2022; Song et al., 2023a; Rout et al., 2023; Daras et al., 2024).

Motivation: inverse problems with diffusion priors. In an inverse problem, one is asked to infer an unknown random signal $X_0 \in \mathbb{R}^d$ from noisy observations

$$Y = \mathcal{A}(X_0) + \varepsilon, \tag{1}$$

where $\mathcal{A} : \mathbb{R}^d \to \mathbb{R}^n$ is a known measurement operator, and $\varepsilon \in \mathbb{R}^n$ represents the measurement noise. Such problems are often ill-posed: even in the absence of noise, the observations y alone may be insufficient to uniquely determine the underlying signal X_0, particularly when $n < d$. A principled approach to resolving this ambiguity and improving statistical performance is provided by Bayesian inference, which combines the measurement model with a prior distribution over the unknown signal X_0. Specifically, given a prior p_{X_0}, the central objective of Bayesian inference is to characterize, or sample from, the posterior distribution $p_{X_0|y}$.

The effectiveness of Bayesian inference therefore depends largely on the choice of prior. Classical approaches typically impose hand-crafted priors, such as sparsity or low-rank structure (Daubechies

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

et al., 2004; Fan & Li, 2001; Candès et al., 2006; Donoho, 2006; Park & Casella, 2008; Ji et al., 2008; Babacan et al., 2009; Donoho et al., 2009; Candes & Recht, 2012; Candès, 2014; Rockova & George, 2018; Chi et al., 2019). While these priors have led to many important advances, they often oversimplify the structure of natural signals. Diffusion models offer a fundamentally different alternative: rather than manually specifying a prior or regularizer, they use a pre-trained generative model as an implicit prior over the signal distribution. This perspective has inspired a rapidly growing line of work on posterior sampling for inverse problems with diffusion priors (Song et al., 2022; Chung et al., 2023a; Kawar et al., 2022; Song et al., 2023a; Xu & Chi, 2024; Wang et al., 2024; Zhang et al., 2025). Moreover, because pre-trained diffusion models can be used as "plug-and-play" priors (Venkatakrishnan et al., 2013; Sreehari et al., 2016; Chan et al., 2017; Xu & Chi, 2024), they often require no task-specific retraining and readily transfer across sensing modalities and application domains.

Using diffusion models as priors, however, presents a new suite of algorithmic challenges. Rather than assuming access to an explicit, closed-form expression for the prior distribution, the diffusion-prior framework models the prior implicitly through a pre-trained diffusion model that provides score estimates (i.e., estimates of the Stein score functions) for progressively diffused versions of the prior. A central algorithmic challenge is therefore to combine the learned diffusion prior with the measurement model (2) in order to generate samples from the posterior distribution. Existing diffusion-based methods typically address this through heuristic correction steps, projection operators, likelihood-gradient guidance, approximate posterior updates, and so on (Chung et al., 2022; Chung & Ye, 2022; Choi et al., 2021; Kawar et al., 2022; Yuan et al., 2026). Although these methods have demonstrated strong empirical performance, many lack rigorous guarantees that the generated samples converge to the true Bayesian posterior. Furthermore, some require computationally intensive modifications to the underlying sampling procedure. For example, Chung et al. (2023a); Song et al. (2023a) required computing the gradient of the score functions, while the sequential Monte Carlo approach of Cardoso et al. (2023) may require a prohibitively large number of particles. These considerations can limit their practical applicability in settings where both computational efficiency and statistical reliability are important.

These issues highlight the need for diffusion-based inverse problem solvers that achieve rigorous theoretical guarantees and computational efficiency at once. Such properties are particularly important in high-stakes applications such as medical imaging and scientific discovery. At the same time, practical deployment often favors algorithms that preserve the simplicity and efficiency of standard diffusion samplers.

An overview of our contributions. In this paper, we make progress towards developing a theoretically principled, and computationally efficient framework for posterior sampling in *linear* inverse problems with diffusion priors. Specifically, we consider the noisy linear measurement model:

$$Y = AX_0 + \varepsilon, \tag{2}$$

where $X_0 \in \mathbb{R}^d$ is the unknown signal (generated randomly from a data distribution p_0), $A : \mathbb{R}^d \to \mathbb{R}^n$ represents a linear measurement operator known *a priori*, and $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$ stands for Gaussian measurement noise with $\sigma > 0$ the noise standard deviation. Throughout, we assume that the prior distribution of X_0 can be represented implicitly by a pre-trained diffusion model. The objective is to generate samples from the posterior distribution of X_0 conditioned on the observation y.

Our starting point is the observation that, under a linear measurement model, different singular directions of the measurement operator carry different amounts of statistical information. Rather than treating all directions uniformly, we design a DDIM-type posterior sampler, called `Posterior-DDIM`, that adaptively determines, for each singular direction, whether the update should be based primarily on the learned diffusion prior or the observed data. This results in a simple, lightweight, coordinate-wise modification of the standard DDIM sampler that explicitly exploits the structure of the measurement operator, while preserving the efficiency of DDIM-type sampling. Along directions where the measurements have sufficient signal-to-noise ratio (SNR), the algorithm injects a carefully calibrated noisy version of the observations so that the resulting coordinates have the correct forward-diffusion marginal distribution. In contrast, directions that are either observed with low SNR or entirely unobserved continue to evolve according to the standard DDIM dynamics induced by the prior. Consequently, posterior sampling reduces to a sequence of simple, direction-wise DDIM updates that are both computationally efficient and easy to implement.

Theoretically, our main contribution is to establish posterior consistency of the proposed `Posterior-DDIM` sampler. Under an exact diffusion prior and a sufficiently fine time discretization, we prove that the sampler converges to the posterior distribution of the unknown signal conditioned on the measurements. This result provides rigorous theoretical support for DDIM-type samplers in linear inverse problems, and suggests a principled approach to integrating measurement information with diffusion priors in noisy, potentially ill-conditioned or rank-deficient settings.

Empirically, we evaluate the proposed sampler on a range of image restoration tasks, including inpainting, super-resolution, Gaussian deblurring, and compressed sensing, using PSNR, SSIM, LPIPS, and FID as evaluation metrics. Across these tasks, our `Posterior-DDIM` algorithm performs favorably against existing diffusion-based posterior samplers: it achieves the best performance on at least three of the four metrics in most tasks, and on at least two metrics in every task considered.

Related work. Despite their significant empirical success, existing diffusion-based inverse solvers still leave room for improvement. One prominent class of methods (Song et al., 2020; Chung et al., 2022; Chung & Ye, 2022; Choi et al., 2021; Kawar et al., 2022; Chung et al., 2023b) enforces measurement consistency by incorporating projection steps into unconditional diffusion sampling procedures. While these approaches have demonstrated strong empirical performance, they do not provide theoretical guarantees on the convergence or posterior correctness of the resulting sampling procedures. A second line of work (Chung et al., 2023a; Song et al., 2023a; Kawar et al., 2021; Cardoso et al., 2023; Song et al., 2023b) attempted to approximate the (*intractable*) posterior score and subsequently apply standard diffusion samplers for posterior sampling. Although empirically effective, it remains unclear whether the resulting sampler faithfully yields the true posterior distribution. In addition, Song et al. (2022) proposed `ReSample`, which promotes measurement consistency by repeatedly optimizing the clean image estimate at intermediate time steps and then re-noising it to continue the reverse diffusion process. Similarly, Graikos et al. (2022) proposed to iteratively optimize the clean image via gradient descent in the reverse diffusion process. A different perspective was introduced by Mardani et al. (2023), who formulated inverse problems with diffusion priors as a variational inference problem via direct KL minimization rather than reverse diffusion sampling. Like the aforementioned approaches, these methods currently lack rigorous theoretical guarantees. Another strand of work Trippe et al. (2022); Wu et al. (2023); Dou & Song (2024); Cardoso et al. (2023) combined Sequential Monte Carlo (MC) methods (Doucet et al., 2001) with (unconditional) scores to sample from the correct posterior. In particular, Cardoso et al. (2023); Dou & Song (2024) established convergence guarantees as the number of particles tends to infinity. However, these methods incur a prohibitive computational cost. In particular, (Gupta et al., 2024) showed that polynomially many particles are in general insufficient for the above Sequential MC approaches to converge. More recently, Xu & Chi (2024) proposed a theoretically justified alternating framework for nonlinear inverse problems that interleaves likelihood-based proximal consistency sampling with score-based diffusion sampling.

2 PRELIMINARIES

We briefly review the background on diffusion models and their connections to inverse problems.

Diffusion models. We first review the basics of diffusion models. Let $X_0 \sim p_0$ denote a random sample in $\mathbb{R}^d$ drawn from an underlying data distribution p_0. A diffusion model defines a forward stochastic process $(X_t)_{t\in[0,T]}$ that corrupts X_0 by injecting Gaussian noise according to:

$$X_t \overset{\mathrm{d}}{=} \alpha_t X_0 + \sigma_t Z_t, \qquad Z_t \sim \mathcal{N}(0, I_d) \tag{3}$$

for all $0 \le t \le T$, where the coefficients $\alpha_t \ge 0$ and $\sigma_t \ge 0$ specify the noise schedule. Throughout this paper, we assume that the SNR α_t/σ_t decreases monotonically and continuously as t grows, and satisfies $\alpha_0/\sigma_0 = \infty$ and $\alpha_\infty/\sigma_\infty = 0$. As t increases, the signal component gradually diminishes while the noise level increases, so that, under an appropriate noise schedule, the distribution of X_t transitions smoothly from the data distribution toward an approximately Gaussian distribution. A central goal of a diffusion model is to learn its reverse process, which starts from a Gaussian sample and progressively removes the injected noise to yield a sample from the data distribution of interest.

Score functions. A fundamental object governing the reverse process is the (Stein) score function

$$s_t(x) \coloneqq \nabla \log p_{X_t}(x), \qquad 0 \le t \le T, \tag{4}$$

where p_{X_t} denotes the marginal distribution of X_t. The score function specifies the direction in which the log-density of X_t increases most rapidly and plays a central role in reversing the forward process (Anderson, 1982; Haussmann & Pardoux, 1986). In practice, the score functions are often parameterized through the *noise predictor* $\epsilon_t(x)$ or the *data predictor* $\mu_t(x)$:

$$\epsilon_t(x) \coloneqq \mathbb{E}[Z_t \mid X_t = x], \qquad \mu_t(x) \coloneqq \mathbb{E}[X_0 \mid X_t = x]. \tag{5}$$

These objects are linked through Tweedie's formula (Efron, 2011), which yields the identities

$$\epsilon_t(x) = -\sigma_t s_t(x) \qquad \text{and} \qquad \mu_t(x) = \alpha_t^{-1}[x - \sigma_t \epsilon_t(x)]. \tag{6}$$

Reparameterzed reverse process and DDIM-type sampler. The forward diffusion process (3) admits a reverse-time stochastic differential equation (SDE), originally characterized by Anderson (1982). It is sometimes convenient to reparameterize time t using the following log SNR parameter:

$$\lambda_t \coloneqq \log\big(\alpha_t/\sigma_t\big). \tag{7}$$

Under this parameterization, the reverse SDE admits the following simple form using the λ-variable (i.e., the log SNR variable) as the "time variable" in place of t:

$$\mathrm{d}\big(\mathrm{e}^{\lambda} X/\sigma\big) = 2e^{2\lambda}\mu_\lambda(X)\,\mathrm{d}\lambda + \sqrt{2}e^{\lambda}\,\mathrm{d}B_\lambda, \tag{8}$$

where B_λ denotes a standard Brownian motion indexed by λ, and we abuse the notation by letting μ_λ be a shorthand for μ_t with $\lambda = \lambda_t$. Practical sampling procedures can be obtained by approximating this reverse SDE (8) through suitable time discretization. More precisely, consider time points $t_0 > t_1 > \cdots > t_M$ with corresponding log SNR values $\lambda_{t_0}, \cdots, \lambda_{t_M}$ (cf. (7)). Let

$$\delta_{i+1} \coloneqq \lambda_{t_i} - \lambda_{t_{i+1}}. \tag{9}$$

A first-order discretization yields the first-order SDE-DPMSolver-1 update (Lu et al., 2022)

$$\frac{X_{t_{i+1}}}{\sigma_{t_{i+1}}} = e^{-\delta_{i+1}}\frac{X_{t_i}}{\sigma_{t_i}} + e^{\lambda_{t_{i+1}}}\left(1 - e^{-2\delta_{i+1}}\right)\mu_{t_i}(X_{t_i}) + \sqrt{1 - e^{-2\delta_{i+1}}}\,Z_{t_i}. \tag{10}$$

where $Z_{t_i} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I)$. More generally, DDIM-type samplers (Song et al., 2021) interpolate between deterministic and stochastic reverse updates, and for any $0 \le \eta \le 1$ can be written as

$$\begin{aligned}\frac{X_{t_{i+1}}}{\sigma_{t_{i+1}}} = \frac{X_{t_i}}{\sigma_{t_i}}\sqrt{1 - \eta\big(1 - e^{-2\delta_{i+1}}\big)} + e^{\lambda_{t_{i+1}}}\Big(1 - e^{-\delta_{i+1}}\sqrt{1 - \eta\big(1 - e^{-2\delta_{i+1}}\big)}\Big)\mu_{t_i}(X_{t_i}) \\ + \sqrt{\eta\big(1 - e^{-2\delta_{i+1}}\big)}\,Z_{t_i}.\end{aligned} \tag{11}$$

Notably, η controls the stochasticity of the update: the choice $\eta = 0$ yields the deterministic DDIM-type update, while $\eta = 1$ recovers the stochastic SDE-DPMSolver-1 update in (10).

Diffusion priors for inverse problems. Next, we briefly describe how diffusion models are commonly used as priors for linear inverse problems. Given a pre-trained diffusion model representing the prior distribution of X_0, the aim is to generate samples from the posterior distribution $p_{X_0|Y=y}$. In analogy with the unconditional case, consider the conditional data predictor

$$\mu_{t,y}(x) \coloneqq \mathbb{E}[X_0 \mid X_t = x,\, AX_0 + \varepsilon = y]. \tag{12}$$

If this conditional predictor were available, then one could simply replace the unconditional predictor μ_t by $\mu_{t,y}$ in the DDIM-type sampler (11), thereby directly obtaining samples from the posterior distribution $p_{X_0|y}$. Consequently, a central challenge in diffusion-based inverse problems is to approximate the conditional predictor $\mu_{t,y}$ using only the pre-trained unconditional predictor μ_t together with the measurement y.

Many existing methods are built upon this idea. Starting from the unconditional prediction $\mu_t(x)$, they seek to compute a measurement-compatible data predictor μ by balancing two competing objectives: remaining faithful to to the diffusion prior that yields the unconditional data predictor $\mu_t(\cdot)$, while simultaneously enforcing consistency with the observed measurements. To approximate the conditional data predictor, existing approaches often rely on heuristic approximations. A common strategy is to combine the unconditional predictor $\mu_t(\cdot)$ and the measurement Y in a simple — often linear — manner (Wang et al., 2022b; Kawar et al., 2022; Song et al., 2022), while another is to use $\mu_t(\cdot)$ directly as a proxy for the unknown signal X_0 (Chung et al., 2023a).

3 MAIN RESULTS

In this section, we develop a posterior-consistent DDIM-type sampler for noisy linear inverse problems, and establish its theoretical guarantees.

Before proceeding, we find it convenient to introduce notation associated with the singular value decomposition (SVD) of the measurement matrix $A \in \mathbb{R}^{n\times d}$. Suppose that A has rank r, and let

$$A = U\Sigma V^\top = U_{\mathcal{S}}\Sigma_{\mathcal{S}}V_{\mathcal{S}}^\top \tag{13}$$

represent the full and compact singular value decompositions (SVDs) of A, where $U \in \mathbb{R}^{n\times n}$ and $V \in \mathbb{R}^{d\times d}$ are orthonormal matrices, $U_{\mathcal{S}} \in \mathbb{R}^{n\times r}$ and $V_{\mathcal{S}} \in \mathbb{R}^{d\times r}$ consist of the left and right singular vectors associated with the nonzero singular values, and $\Sigma \in \mathbb{R}^{n\times d}$ and $\Sigma_{\mathcal{S}} \in \mathbb{R}^{r\times r}$ are the rectangular diagonal and diagonal matrices of singular values, respectively. We denote by

$$\mathcal{S} \coloneqq \{s \in [\min\{n,d\}] : \Sigma_{ss} > 0\} \tag{14}$$

the indices of the nonzero singular values, and hence the notation $U_{\mathcal{S}}, V_{\mathcal{S}}$ and $\Sigma_{\mathcal{S}}$.We also write

$$U_{\mathcal{S}} \coloneqq [u_s]_{s\in\mathcal{S}} \in \mathbb{R}^{n\times r}, \quad V_{\mathcal{S}} \coloneqq [v_s]_{s\in\mathcal{S}} \in \mathbb{R}^{d\times r}, \quad \text{and} \quad \Sigma_{\mathcal{S}} \coloneqq \mathsf{diag}([\Sigma_{ss}]_{s\in\mathcal{S}}) \in \mathbb{R}^{r\times r}. \tag{15}$$

3.1 MOTIVATION: SNR-GUIDED PARTITION OF SINGULAR DIRECTIONS

The SVD (13) of the measurement matrix A naturally decouples the inverse problem into a sequence of scalar inference problems along its singular directions. Indeed, expressing the measurement model $Y = AX_0 + \varepsilon$ (i.e., (2)) in the SVD basis yields

$$\Sigma_{\mathcal{S}}^{-1}U_{\mathcal{S}}^\top Y = V_{\mathcal{S}}^\top X_0 + \varepsilon' \qquad \text{with } \varepsilon' \coloneqq \Sigma_{\mathcal{S}}^{-1}U_{\mathcal{S}}^\top\varepsilon \sim \mathcal{N}\big(0, \sigma^2\Sigma_{\mathcal{S}}^{-2}\big); \tag{16}$$

or equivalently, for every $s \in \mathcal{S}$,

$$\Sigma_{ss}^{-1}u_s^\top Y = v_s^\top X_0 + \varepsilon_s' \qquad \text{with } \varepsilon_s' \sim \mathcal{N}\left(0, \frac{\sigma^2}{\Sigma_{ss}^2}\right). \tag{17}$$

Thus, after transforming into the singular-vector basis, the effective *observation SNR* along the s-th singular direction is naturally measured by Σ_{ss}/σ. Meanwhile, projecting the forward diffusion process (3) along the same singular direction gives

$$v_s^\top X_t = \alpha_t v_s^\top X_0 + \sigma_t Z_s \qquad \text{with } Z_s \sim \mathcal{N}(0,1), \tag{18}$$

whose SNR — which we often refer to as *diffusion SNR* — is given by α_t/σ_t. Consequently, the relative magnitudes of these two SNRs dictate whether the posterior update should rely primarily on the measurements or on the diffusion prior.

This comparison naturally leads to a partition of the singular directions. For each diffusion time t, we partition the singular directions according to

$$\mathcal{S}_t^{\mathsf{meas}} \coloneqq \{s \in \mathcal{S} : \alpha_t\sigma < \Sigma_{ss}\sigma_t\} \qquad \text{and} \qquad \mathcal{S}_t^{\mathsf{prior}} \coloneqq \mathcal{S} \setminus \mathcal{S}_t^{\mathsf{meas}}. \tag{19}$$

For each $s \in \mathcal{S}$, we define the threshold time $\tau_s \in \mathbb{R}_{\geq 0} \cup \{\infty\}$ such that $\frac{\alpha_{\tau_s}}{\sigma_{\tau_s}} = \frac{\Sigma_{ss}}{\sigma}$, namely, the diffusion time at which the diffusion SNR matches the observation SNR along the s-th singular direction. When α_t/σ_t is continuous and strictly decreasing in t, with $\alpha_0/\sigma_0 = \infty$ and $\alpha_\infty/\sigma_\infty = 0$, the threshold τ_s is well defined. It can be easily verified that

$$s \in \mathcal{S}_t^{\mathsf{meas}} \qquad \Longleftrightarrow \qquad t \geq \tau_s. \tag{20}$$

Therefore, at each diffusion time t, it is natural to partition the singular directions into three groups:

- *Measurement-dominated directions*: $s \in \mathcal{S}_t^{\mathsf{meas}}$, that is, directions obeying $t \geq \tau_s$;
- *Prior-dominated observed directions*: $s \in \mathcal{S}_t^{\mathsf{prior}}$, that is, directions in $\mathcal{S}$ obeying $t < \tau_s$;
- *Unobserved directions*: $s \in \mathcal{S}^{\mathrm{c}} \coloneqq [d] \setminus \mathcal{S}$.

Our sampler treats these three groups differently. As we shall see, each case admits a simple DDIM-type update, resulting in an efficient posterior sampler that adapts to the SNR along each direction.

Algorithm 1: `Posterior-DDIM` sampler for the linear inverse problem (2)

Input: time points $t_0 > \cdots > t_M$, data predictor $\{\mu_{t_i}(\cdot)\}_{i=0}^{M}$, schedule $\{(\alpha_{t_i}, \sigma_{t_i})\}_{i=0}^{M}$, parameters η_0, η_1.
Output: the generated sample $\widehat{X}_{t_M} = \sum_s v_s v_s^\top \widehat{X}_{t_M}$.

initialize $\widehat{X}_{t_0}, Z = [Z_s]_{1 \le s \le d} \sim \mathcal{N}(0, I_d)$; compute SVD of A (cf. (13)); set $\mathcal{S} \leftarrow \{i : \Sigma_{ii} > 0\}$.
for $s = 1, \cdots, d$ **do**
  if $s \le \min\{n, d\}$ *and* $\alpha_{t_0}\sigma < \Sigma_{ss}\sigma_{t_0}$ **then** $v_s^\top \widehat{X}_{t_0} \leftarrow \alpha_{t_0}\Sigma_{ss}^{-1} u_s^\top Y + \sqrt{\sigma_{t_0}^2 - \alpha_{t_0}^2\sigma^2\Sigma_{ss}^{-2}}\, Z_s$
  else $v_s^\top \widehat{X}_{t_0} \leftarrow Z_s$
end
for $i = 0, \cdots, M-1$ **do**
  $\mathcal{S}_{t_{i+1}}^{\mathsf{meas}} \leftarrow \{s : \alpha_{t_{i+1}}\sigma < \Sigma_{ss}\sigma_{t_{i+1}}\}$; $\mathcal{S}_{t_{i+1}}^{\mathsf{prior}} \leftarrow \mathcal{S} \setminus \mathcal{S}_{t_{i+1}}^{\mathsf{meas}}$; draw $Z_{t_{i+1}} \sim \mathcal{N}(0, I)$.
  for $s \in \mathcal{S}$ **do**
    if $t_{i+1} < \tau_s \le t_i$ **then** $v_s^\top \widehat{X}_{\tau_s} \leftarrow v_s^\top \widehat{X}_{t_i}$
  end
  for $s \in \mathcal{S}_{t_{i+1}}^{\mathsf{meas}}$ **do**
    $v_s^\top \widehat{X}_{t_{i+1}} \leftarrow \alpha_{t_{i+1}}\Sigma_{ss}^{-1} u_s^\top Y + \sqrt{\sigma_{t_{i+1}}^2 - \alpha_{t_{i+1}}^2\sigma^2\Sigma_{ss}^{-2}}\, Z_s$ `// measurement-dominated`
  end
  for $s \in \mathcal{S}_{t_{i+1}}^{\mathsf{prior}}$ **do**
    `/* prior-dominated */`
    if *all nonzero singular values of* A *are equal* **then** compute $v_s^\top \widehat{X}_{t_{i+1}}$ via (23) with $\eta = \eta_0$
    else compute $v_s^\top \widehat{X}_{t_{i+1}}$ via (24) with $\eta = \eta_0$ `// add correction`
  end
  for $s \in \mathcal{S}^{\mathsf{c}}$ **do** compute $v_s^\top \widehat{X}_{t_{i+1}}$ via (23) with $\eta = \eta_1$ `// unobserved`
end

3.2 Algorithm: `Posterior-DDIM`

We now describe the proposed sampler, based on the partition of singular directions developed in Section 3.1. The central idea is to exploit the measurements whenever the corresponding observation SNR exceeds the diffusion SNR, and otherwise rely on the learned diffusion prior. This idea gives rise to distinct update rules for the three groups of singular directions, which we describe below. The complete algorithm of the proposed DDIM-based posterior sampler, hereafter referred to as `Posterior-DDIM`, is summarized in Algorithm 1.

Measurement-dominated directions. For directions $s \in \mathcal{S}_t^{\mathsf{meas}}$, the measurements are sufficiently informative to be used directly. To this end, consider the auxiliary forward process

$$\xi_{t,s} := \alpha_t \xi_{0,s} + \sqrt{\sigma_t^2 - \alpha_t^2\sigma^2\Sigma_{ss}^{-2}}\, Z_s, \quad Z_s \sim \mathcal{N}(0,1), \tag{21}$$

where $\xi_{0,s}$ is taken to be $\xi_{0,s} := \Sigma_{ss}^{-1} u_s^\top Y$. By construction, this auxiliary forward process $\xi_{t,s}$ has the same marginal distribution as the forward diffusion process X_t (cf. (3)) projected onto the direction the s-th singular direction v_s, namely,

$$\xi_{t,s} \overset{\mathrm{d}}{=} v_s^\top X_t = v_s^\top(\alpha_t X_0 + \sigma_t Z).$$

Consequently, sampling $v_s^\top \widehat{X}_{t_i}$ at the time point t_i can be accomplished by generating

$$v_s^\top \widehat{X}_{t_i} = \alpha_{t_i}\Sigma_{ss}^{-1} u_s^\top Y + \sqrt{\sigma_{t_i}^2 - \alpha_{t_i}^2\sigma^2\Sigma_{ss}^{-2}}\, Z_s \quad \text{with } Z_s \sim \mathcal{N}(0,1), \tag{22}$$

provided that $s \in \mathcal{S}_{t_i}^{\mathsf{meas}}$. A formal proof is given in Appendix C.1.

Prior-dominated observed directions. For directions $s \in \mathcal{S}_t^{\mathsf{prior}}$, the measurements are less informative than the diffusion prior. Consequently, rather than directly incorporating the observations, we update these directions using the standard DDIM sampler driven primarily by the learned data predictor.

We begin with the special case in which all nonzero singular values of A are identical. In this setting, for $s \in \mathcal{S}_{t_{i+1}}^{\mathsf{prior}}$, we update $v_s^\top \widehat{X}_t$ according to the standard DDIM-type update rule (cf. (23)) as follows

$$\frac{v_s^\top \widehat{X}_{t_{i+1}}}{\sigma_{t_{i+1}}} = \frac{v_s^\top \widehat{X}_{t_i}}{\sigma_{t_i}} \sqrt{1-\eta(1-e^{-2\delta_{i+1}})} + e^{\lambda_{t_{i+1}}} \left(1 - e^{-\delta_{i+1}} \sqrt{1-\eta(1-e^{-2\delta_{i+1}})}\right) v_s^\top \mu_{t_i}(\widehat{X}_{t_i}) + \sqrt{\eta(1-e^{-2\delta_{i+1}})}\, v_s^\top Z_{t_{i+1}}, \qquad Z_{t_{i+1}} \sim \mathcal{N}(0, I_d), \tag{23}$$

where the stochasticity parameter is set to $\eta = \eta_0$. Note that the noise $Z_{t_{i+1}}$ is independent across time steps and is also independent of Z_s in (21).

The above update extends naturally to the general setting where A has arbitrary nonzero singular values. The only modification is the addition of a correction term, which arises from the conditional distribution given $v_s^\top X_{\tau_s}$. Specifically, for $s \in \mathcal{S}_{t_i}^{\mathsf{prior}}$, we replace (23) with

$$\frac{v_s^\top \widehat{X}_{t_{i+1}}}{\sigma_{t_{i+1}}} = \frac{v_s^\top \widehat{X}_{t_i}}{\sigma_{t_i}} \sqrt{1-\eta(1-e^{-2\delta_{i+1}})} + \sqrt{\eta(1-e^{-2\delta_{i+1}})}\, v_s^\top Z_{t_{i+1}} + e^{\lambda_{t_{i+1}}} \left(1 - e^{-\delta_{i+1}} \sqrt{1-\eta(1-e^{-2\delta_{i+1}})}\right) v_s^\top \left(\mu_{t_i}(\widehat{X}_{t_i}) + \underbrace{\frac{\alpha_{t_i}\widehat{X}_{\tau_s} - \alpha_{\tau_s}\widehat{X}_{t_i}}{\alpha_{\tau_s}\alpha_{t_i}(\mathrm{e}^{2(\lambda_{t_i}-\lambda_{\tau_s})}-1)}}_{\text{correction term}}\right), \tag{24}$$

where $Z_{t_{i+1}} \sim \mathcal{N}(0, I_d)$. The derivation of this modified update rule is deferred to Appendix D.

Unobserved directions. For directions $s \in \mathcal{S}^{\mathsf{c}}$, the measurements contain no information about $v_s^\top X_0$. Accordingly, these coordinates should be sampled entirely from the learned diffusion prior, given the coordinates that are observed. As in the prior-dominated observed case, we employ the DDIM update (23). However, we choose a larger stochasticity parameter $\eta = \eta_1$ to encourage sufficient exploration of the unobserved subspace. This is important because the posterior uncertainty along $\mathcal{S}^{\mathsf{c}}$ is determined entirely by the learned diffusion prior rather than the measurement model.

Choices of the stochasticity parameters η_0 and η_1. We take a moment to discuss the roles of the two DDIM stochasticity parameters η_0 and η_1. Although both parameters control the amount of randomness injected into the DDIM update, they serve fundamentally different purposes because the target conditional distributions differ across the two subspaces. For directions $s \in \mathcal{S}_t^{\mathsf{prior}}$, the corresponding coordinates are observed, although the measurements are not yet sufficiently informative to be imposed directly. The update should therefore approximate the conditional reverse transition given the current diffusion state and the measurements. This motivates using a relatively small value of η_0, so as to keep the update closer to the deterministic DDIM sampler. In contrast, for directions $s \in \mathcal{S}^{\mathsf{c}}$, no measurements are available. These coordinates must be sampled entirely from the learned conditional prior given the observed coordinates. A larger value of η_1 makes the DDIM update more Langevin-like, thereby promoting exploration of the unobserved subspace.

3.3 Theoretical guarantees: asymptotic posterior correctness

In this subsection, we establish the posterior correctness of the proposed `Posterior-DDIM` sampler. We begin with the special case in which all nonzero singular values of the measurement matrix A are identical, as this setting captures the main ideas while allowing for cleaner algorithm design analysis. Here and throughout, $\mathsf{TV}(p, q)$ denotes the total-variation (TV) distance between two probability measures p and q.

Theorem 1. *Consider a sequence of time points $t_0 > \cdots > t_M$. Recall the definition of δ_i in (9), and set $\delta := \max\{\max_{1\le i\le M} \delta_i, \alpha_{t_0}, |1-\sigma_{t_0}|\}$. Assume that the support of X_0 is bounded, and the data predictor is exact, i.e., $\mu_t(y) = \mathbb{E}[X_0 \,|\, X_t = y]$. We further assume that $\eta_1 = \delta^{-p}$ with $1/2 < p < 2/3$, and that $\eta_0 = 1$ when all nonzero singular values of A are equal, while $\eta_0 = \eta_1$ otherwise. The `Posterior-DDIM` sampler (see Algorithm 1) achieves*

$$\lim_{t_M \to 0} \lim_{\delta \to 0} \mathsf{TV}\big(p_{\widehat{X}_{t_M}|Y}, p_{X_0|Y}\big) = 0. \tag{25}$$

In words, Theorem 1 establishes that, the distribution of the output generated by Algorithm 1 converges to the posterior distribution of X_0 conditioned on the measurements. This guarantee is asymptotic and holds in the regime where (i) the terminal time t_M tends to 0, approaching the initial time of the forward diffusion process; (ii) the initial sampling time t_0 is chosen such that $\alpha_{t_0} \to 0$ and $\sigma_{t_0} \to 1$, ensuring that the SNR at the initialization stage vanishes; and (iii) the maximum gap between the log-SNRs of consecutive time points vanishes, which in particular requires the number of sampling steps M to tend to infinity. The proof is deferred to Appendix B.

For an inaccurate data predictor $\mu_t(y)$, the influence of the data predictor error can be analyzed by defining an auxiliary sequence that employs the exact data predictor, bounding the KL divergence between this sequence and our sampler, and then converting to a TV distance via Pinsker's inequality. Compared with the bound in Theorem 1, the total variation distance then increases by an additional term induced by the data predictor estimation error, which is proportional to the following error term: $\varepsilon_{\mathsf{sc}} = \sqrt{\sum_{i=0}^{M-1} \sigma_{t_i}^2 \delta_{i+1} \mathbb{E}[\|s_{t_i}(X_{t_i}) - \nabla \log p_{X_{t_i}}(X_{t_i})\|_2^2]}$.

4 Experiments

We begin by describing the common experimental setup. Unless otherwise stated, the experiments are run with the NFE (number of function evaluations) set to 100, except for `DPS`, for which we follow Zhang et al. (2024) and set the NFE to be 1000. To justify our default choice, we also investigate the effect of NFE and find that the reconstruction quality largely saturates once the NFE exceeds 80 (see Figure 2 in Appendix). We set η_0 and η_1 in `Posterior-DDIM` to 0 and 16 respectively, as these values yield the best overall performance across most evaluation metrics in our hyperparameter sensitivity experiments (cf. Appendix A.1). We adopt $\sqrt{\max\{1-\eta(1-e^{-2\delta_{i+1}}), 0\}}$ in implementation to avoid negativity for large η. All experiments are conducted on a server equipped with NVIDIA A100-PCIE-40GB GPUs and Intel Xeon (Cascade Lake) CPUs.

We compare `Posterior-DDIM` against several representative diffusion-based image restoration algorithms, including `DDRM` (Kawar et al., 2022), `DDNM+` (Wang et al., 2022a), `DPS` (Chung et al., 2023a), `RED-diff` (Mardani et al., 2023), and `ProjDiff` (Zhang et al., 2024).[1] As a classical baseline, we also include the least-squares estimator, $\widehat{X}^{\mathsf{ls}} = A^\dagger Y$. Experiments are conducted on four image restoration tasks: 50% random inpainting and $4\times$ super-resolution with average pooling, Gaussian deblurring and compressed sensing, using the CelebA (Liu et al., 2015) and ImageNet (Deng et al., 2009) datasets. For Gaussian deblurring, we use a one-dimensional Gaussian blur kernel of size 5 with standard deviation 10. For compressed sensing, we employ a Walsh–Hadamard sampling matrix with a sampling ratio of 0.5. We set the noise standard deviation to $\sigma = 0.05$ (doubled when pixels are rescaled to $[-1, 1]$). For ImageNet, we use the pretrained 256×256 model without classifier guidance from Dhariwal & Nichol (2021). For CelebA, we use the pretrained model released with SDEdit (Meng et al., 2021). Following Wang et al. (2022a); Zhang et al. (2024), we evaluate on the 1K ImageNet sub-test set and the CelebA test set provided by Wang et al. (2022a). Performance is measured using PSNR (peak SNR), SSIM (Wang et al., 2004), LPIPS (Zhang et al., 2018), and FID (Heusel et al., 2017). For readability, all reported LPIPS values are multiplied by 100.

In addition, note that the posterior mean can be approximately computed via Monte Carlo sampling, i.e, $\mathbb{E}[X_0 \,|\, Y] \approx \frac{1}{n}\sum_{i=1}^N \widehat{X}^{(i)}$, where $\widehat{X}^{(i)} \overset{\text{i.i.d.}}{\sim} p_{X_0|Y}$. As is well-known, the posterior mean achieves the optimal reconstruction accuracy under the mean squared estimation error metric. Consequently, reconstruction fidelity can be improved by averaging samples generated from independent noise realizations. We observe, however, that this approach may sometimes degrade quality as measured by FID. To illustrate this effect, we report both single-sample results and posterior-mean results, where the latter are obtained by averaging 4 independent trials.

The results are reported in Tables 1–2, and Tables 5–6 in Appendix. Boldface indicates the best overall result, while underlining indicates the best sampling-based result. Across all tasks and datasets, the proposed `Posterior-DDIM` method consistently delivers strong performance. For super-resolution, it achieves the best performance on three and two of the four evaluation metrics

[1] In addition to our algorithm, we also implement `ProjDiff` and `DDNM+`. The remaining results reported in this section are taken from Zhang et al. (2024).

Table 1: Empirical results for super-resolution on CelebA and ImageNet.

	Method	CelebA				ImageNet			
		PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓
Single-sample	$A^\dagger y$	23.64	0.49	68.72	147.89	21.85	0.45	65.34	183.32
	DPS	27.98	0.78	23.10	39.91	24.44	0.67	**31.81**	**36.17**
	DDRM	29.20	0.82	21.92	40.14	25.66	0.72	34.88	55.71
	RED-diff	24.98	0.55	50.59	73.89	22.74	0.49	53.24	96.26
	DDNM+	29.20	0.82	21.91	39.96	25.62	0.72	34.39	53.78
	ProjDiff	29.49	0.83	**20.89**	36.61	25.73	0.72	33.03	49.70
	Posterior-DDIM (Ours)	29.84	0.84	22.02	**34.16**	25.90	0.73	32.84	49.99
Posterior mean	DDNM+	30.22	**0.85**	21.97	43.54	26.11	0.73	34.38	54.27
	ProjDiff	30.31	**0.85**	22.55	38.34	26.10	**0.74**	33.13	50.27
	Posterior-DDIM (Ours)	**30.35**	**0.85**	22.55	40.62	**26.13**	**0.74**	33.35	52.91

Table 2: Empirical results for inpainting on CelebA and ImageNet.

	Method	CelebA				ImageNet			
		PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓
Single-sample	$A^\dagger y$	13.70	0.19	76.07	226.28	14.21	0.24	67.42	176.52
	DPS	32.80	0.90	16.32	32.80	30.15	0.86	17.76	22.03
	DDRM	32.81	0.90	16.78	35.28	29.99	0.87	17.11	19.88
	RED-diff	7.96	0.18	78.27	192.32	9.85	0.17	83.92	281.65
	DDNM+	32.81	0.90	16.78	35.33	30.00	0.87	17.07	19.92
	ProjDiff	33.44	0.91	15.32	31.13	31.32	0.89	13.87	**14.97**
	Posterior-DDIM (Ours)	33.23	0.91	15.06	**25.33**	31.06	0.89	14.31	15.85
Posterior mean	DDNM+	33.98	**0.92**	16.06	37.64	31.22	0.89	16.44	22.07
	ProjDiff	34.22	**0.92**	14.41	32.40	**32.07**	**0.90**	13.78	16.57
	Posterior-DDIM (Ours)	**34.24**	**0.92**	**13.81**	27.15	32.04	**0.90**	**13.67**	16.75

on CelebA and ImageNet datasets respectively. For inpainting, it outperforms the reference algorithms on all four metrics for CelebA and on two metrics for ImageNet. For Gaussian deblurring, Posterior-DDIM achieves the best performance on three metrics for CelebA and two metrics for ImageNet. For compressed sensing, Posterior-DDIM outperforms DDNM+ across three evaluation metrics by posterior mean. Overall, whereas existing methods typically achieve the best result on only one or two metrics in most settings, our method attains the best performance on **at least two** evaluation metrics across all tasks and datasets, demonstrating its efficacy and robustness across a diverse range of inverse problems. Results for Posterior-DDIM, DDNM+, and ProjDiff under different numbers of Monte Carlo samples used to estimate the posterior mean are in Appendix.

5 Discussion

In this work, we have proposed a novel DDIM-type posterior sampler for solving linear inverse problems with diffusion priors. In contrast to many existing methods that rely on heuristic projection steps or approximate posterior scores, our approach follows a different route by treating different singular directions of the measurement operator differently according to the relative strengths of the observation noise and the diffusion noise. This leads to a simple, easy-to-implement, DDIM-type sampling procedure that is both practically effective and theoretically principled. In particular, we have established that, under mild conditions, the proposed Posterior-DDIM sampler converges to the true posterior distribution of the signal given the measurement. Empirically, our experiments on a range of real-world image restoration tasks have shown that the proposed method consistently performs favorably against existing diffusion-based posterior samplers, achieving the best performance on the majority of evaluation metrics across all tasks.

Several directions merit further investigation. For instance, the present work has focused on linear inverse problems, and a fundamentally important next step is to extend the proposed framework to broader nonlinear inverse problems, which naturally arise in wide-ranging real-world applications. In addition, while our theoretical analysis has established asymptotic convergence to the true posterior distribution, it falls short of characterizing the rate of convergence, which would be of interest to quantify. It would also be valuable to understand how the convergence behavior and sampling quality depend on additional factors such as the noise level, the properties of the measurement operator, and the choice of time discretization of the diffusion process.

AI USE STATEMENT

In preparing this paper, large language models (LLMs) were employed solely as assistive tools for minor language refinement. All technical contributions, results, and conclusions are entirely the work of the authors.

REFERENCES

Brian DO Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982.

S Derin Babacan, Rafael Molina, and Aggelos K Katsaggelos. Bayesian compressive sensing using laplace priors. *IEEE Transactions on image processing*, 19(1):53–63, 2009.

Emmanuel Candes and Benjamin Recht. Exact matrix completion via convex optimization. *Communications of the ACM*, 55(6):111–119, 2012.

Emmanuel J Candès. Mathematics of sparsity (and a few other things). In *Proceedings of the International Congress of Mathematicians, Seoul, South Korea*, volume 123, pp. 235–258. Citesee, 2014.

Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, 2006.

Gabriel Cardoso, Yazid Janati El Idrissi, Sylvain Le Corff, and Eric Moulines. Monte carlo guided diffusion for bayesian linear inverse problems. *arXiv preprint arXiv:2308.07983*, 2023.

Stanley H. Chan, Xiran Wang, and Omar A. Elgendy. Plug-and-play admm for image restoration: Fixed-point convergence and applications. *IEEE Transactions on Computational Imaging*, 3(1): 84–98, 2017.

Yuejie Chi, Yue M Lu, and Yuxin Chen. Nonconvex optimization meets low-rank matrix factorization: An overview. *IEEE Transactions on Signal Processing*, 67(20):5239–5269, 2019.

Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. Ilvr: Conditioning method for denoising diffusion probabilistic models. *arXiv preprint arXiv:2108.02938*, 2021.

Hyungjin Chung and Jong Chul Ye. Score-based diffusion models for accelerated mri. *Medical image analysis*, 80:102479, 2022.

Hyungjin Chung, Byeongsu Sim, and Jong Chul Ye. Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12413–12422, 2022.

Hyungjin Chung, Jeongsol Kim, Michael Thompson McCann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations*, 2023a.

Hyungjin Chung, Suhyeon Lee, and Jong Chul Ye. Fast diffusion sampler for inverse problems by geometric decomposition. *arXiv preprint arXiv:2303.05754*, 2023b.

Giannis Daras, Hyungjin Chung, Chieh-Hsin Lai, Yuki Mitsufuji, Jong Chul Ye, Peyman Milanfar, Alexandros G Dimakis, and Mauricio Delbracio. A survey on diffusion models for inverse problems. *arXiv preprint arXiv:2410.00083*, 2024.

Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 57(11):1413–1457, 2004.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

David L Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.

Zehao Dou and Yang Song. Diffusion posterior sampling for linear inverse problem solving: A filtering perspective. In *The Twelfth International Conference on Learning Representations*, 2024.

Arnaud Doucet, Nando De Freitas, and Neil Gordon. An introduction to sequential monte carlo methods. In *Sequential Monte Carlo methods in practice*, pp. 3–14. Springer, 2001.

Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011.

Jianqing Fan and Runze Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456):1348–1360, 2001.

Alexandros Graikos, Nikolay Malkin, Nebojsa Jojic, and Dimitris Samaras. Diffusion models as plug-and-play priors. *arXiv preprint arXiv:2206.09012*, 2022.

Shivam Gupta, Ajil Jalal, Aditya Parulekar, Eric Price, and Zhiyang Xun. Diffusion posterior sampling is computationally intractable. *arXiv preprint arXiv:2402.12727*, 2024.

Ulrich G Haussmann and Etienne Pardoux. Time reversal of diffusions. *The Annals of Probability*, pp. 1188–1205, 1986.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Shihao Ji, Ya Xue, and Lawrence Carin. Bayesian compressive sensing. *IEEE Transactions on signal processing*, 56(6):2346–2356, 2008.

Bahjat Kawar, Gregory Vaksman, and Michael Elad. Snips: Solving noisy inverse problems stochastically. *Advances in Neural Information Processing Systems*, 34:21757–21769, 2021.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Chieh-Hsin Lai, Yang Song, Dongjun Kim, Yuki Mitsufuji, and Stefano Ermon. The principles of diffusion models. *arXiv preprint arXiv:2510.21890*, 2025.

Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision*, pp. 3730–3738, 2015.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver++: Fast solver for guided sampling of diffusion probabilistic models. *arXiv preprint arXiv:2211.01095*, 2022.

Morteza Mardani, Jiaming Song, Jan Kautz, and Arash Vahdat. A variational perspective on solving inverse problems with diffusion models. *arXiv preprint arXiv:2305.04391*, 2023.

Chenlin Meng, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. Sdedit: Image synthesis and editing with stochastic differential equations. *arXiv preprint arXiv:2108.01073*, 2021.

Trevor Park and George Casella. The Bayesian lasso. *Journal of the american statistical association*, 103(482):681–686, 2008.

Veronika Rockova and Edward I George. The spike-and-slab lasso. *Journal of the American Statistical Association*, 113(521):431–444, 2018.

Litu Rout, Negin Raoof, Giannis Daras, Constantine Caramanis, Alex Dimakis, and Sanjay Shakkottai. Solving linear inverse problems provably via posterior sampling with latent diffusion models. *Advances in Neural Information Processing Systems*, 36:49960–49990, 2023.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pp. 2256–2265. pmlr, 2015.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021.

Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *International conference on learning representations*, 2023a.

Jiaming Song, Qinsheng Zhang, Hongxu Yin, Morteza Mardani, Ming-Yu Liu, Jan Kautz, Yongxin Chen, and Arash Vahdat. Loss-guided diffusion models for plug-and-play controllable generation. In *International Conference on Machine Learning*, pp. 32483–32498. PMLR, 2023b.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. *International Conference on Learning Representations*, 2022.

Suhas Sreehari, S Venkat Venkatakrishnan, Brendt Wohlberg, Gregery T Buzzard, Lawrence F Drummy, Jeffrey P Simmons, and Charles A Bouman. Plug-and-play priors for bright field electron tomography and sparse interpolation. *IEEE Transactions on Computational Imaging*, 2(4): 408–423, 2016.

Brian L Trippe, Jason Yim, Doug Tischer, David Baker, Tamara Broderick, Regina Barzilay, and Tommi Jaakkola. Diffusion probabilistic modeling of protein backbones in 3d for the motif-scaffolding problem. *arXiv preprint arXiv:2206.04119*, 2022.

Singanallur V Venkatakrishnan, Charles A Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *2013 IEEE global conference on signal and information processing*, pp. 945–948. IEEE, 2013.

Hengkang Wang, Xu Zhang, Taihui Li, Yuxiang Wan, Tiancong Chen, and Ju Sun. Dmplug: A plug-in method for solving inverse problems with diffusion models. *Advances in Neural Information Processing Systems*, 37:117881–117916, 2024.

Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. *arXiv preprint arXiv:2212.00490*, 2022a.

Zhendong Wang, Jonathan J Hunt, and Mingyuan Zhou. Diffusion policies as an expressive policy class for offline reinforcement learning. *arXiv preprint arXiv:2208.06193*, 2022b.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

Table 3: Empirical results for inpainting on CelebA and ImageNet under different η.

η_0	η_1	CelebA				ImageNet			
		PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓
0	0	27.22	0.79	27.11	61.46	21.24	0.54	46.72	89.56
	1	31.32	0.88	16.97	30.24	26.13	0.78	26.17	37.76
	2	32.55	0.90	15.44	26.62	28.76	0.86	16.74	18.60
	4	33.16	**0.91**	15.07	25.42	30.72	**0.89**	14.45	**15.74**
	8	**33.23**	**0.91**	**15.06**	**25.33**	**31.06**	**0.89**	**14.31**	15.86
	16	**33.23**	**0.91**	**15.06**	**25.33**	**31.06**	**0.89**	**14.31**	15.85
	32	**33.23**	**0.91**	**15.06**	**25.33**	**31.06**	**0.89**	**14.31**	15.85
0	16	**33.23**	**0.91**	**15.06**	**25.33**	**31.06**	**0.89**	**14.31**	15.85
0.5		33.03	0.90	15.66	26.60	30.86	0.88	15.33	17.89
1		32.84	0.90	16.36	28.42	30.67	0.88	16.61	20.85
2		32.51	0.89	17.89	33.21	30.26	0.86	19.52	28.71

Table 4: Empirical results for super-resolution on CelebA and ImageNet under different η.

η_0	η_1	CelebA				ImageNet			
		PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓
0	0	27.77	0.78	22.49	32.97	22.48	0.56	38.50	53.08
	1	28.39	0.80	**21.09**	**31.24**	24.81	0.70	**31.22**	**39.94**
	2	28.92	0.82	21.31	32.56	25.67	0.72	32.03	45.74
	4	29.59	0.83	21.85	33.49	**25.93**	**0.73**	32.73	49.06
	8	29.83	**0.84**	22.02	34.13	25.90	**0.73**	32.84	49.99
	16	**29.84**	**0.84**	22.02	34.16	25.90	**0.73**	32.84	49.99
	32	**29.84**	**0.84**	22.02	34.16	25.90	**0.73**	32.84	49.99
0	16	**29.84**	**0.84**	22.02	34.16	25.90	**0.73**	32.84	49.99
0.5		29.48	0.83	22.86	35.35	25.64	0.72	34.11	52.27
1		29.21	0.82	23.57	37.15	25.44	0.70	35.23	55.03
2		28.82	0.81	24.89	41.79	25.13	0.69	37.24	61.86

Luhuan Wu, Brian Trippe, Christian Naesseth, David Blei, and John P Cunningham. Practical and asymptotically exact conditional sampling in diffusion models. *Advances in Neural Information Processing Systems*, 36:31372–31403, 2023.

Xingyu Xu and Yuejie Chi. Provably robust score-based diffusion posterior sampling for plug-and-play image reconstruction. *arXiv preprint arXiv:2403.17042*, 2024.

Zijie Yuan, Ji Li, and Zhaoqiang Liu. Improving diffusion inverse problem solving with structure consistency regularization. In *ICASSP 2026 - 2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 11497–11501, 2026. doi: 10.1109/ICASSP55912.2026.11464920.

Jiawei Zhang, Jiaxin Zhuang, Cheng Jin, Gen Li, and Yuantao Gu. Unleashing the denoising capability of diffusion prior for solving inverse problems. *arXiv preprint arXiv:2406.06959*, 2024.

Jiawei Zhang, Ziyuan Liu, Leon Yan, Gen Li, and Yuantao Gu. Improving diffusion-based inverse algorithms under few-step constraint via learnable linear extrapolation. *arXiv preprint arXiv:2503.10103*, 2025.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 586–595, 2018.

Table 5: Empirical results for Gaussian deblurring on CelebA and ImageNet.[2]

	Method	CelebA				ImageNet			
		PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓
Single-sample	$A^\dagger y$	18.06	0.33	61.51	93.90	17.78	0.29	60.25	100.05
	DPS	29.46	0.81	21.19	38.54	24.26	0.64	37.03	50.17
	DDRM	30.51	0.85	19.89	38.24	27.82	0.79	28.62	45.96
	DDNM+	30.60	0.85	19.79	38.23	27.90	0.79	28.63	46.54
	RED-diff	26.32	0.56	41.23	56.59	23.74	0.50	48.12	68.23
	ProjDiff	<u>31.41</u>	**<u>0.87</u>**	**<u>18.12</u>**	34.89	<u>27.91</u>	**<u>0.79</u>**	23.85	29.12
	Posterior-DDIM (Ours)	30.18	0.84	18.90	<u>27.70</u>	26.99	0.76	<u>23.71</u>	<u>24.54</u>
Posterior mean	ProjDiff	**31.44**	**0.87**	18.22	35.96	**27.96**	**0.79**	23.93	29.41
	Posterior-DDIM (Ours)	**31.44**	**0.87**	18.18	**27.15**	27.28	0.77	**23.05**	**23.69**

Table 6: Empirical results for compressed sensing on CelebA and ImageNet.[3]

	Method	CelebA				ImageNet			
		PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓
Single-sample	$A^\dagger y$	14.96	0.34	66.48	397.50	15.08	0.37	60.45	257.70
	DDNM+	29.05	0.88	18.46	36.43	23.64	0.81	21.10	25.38
	Posterior-DDIM (Ours)	28.94	0.88	16.91	**26.69**	23.39	0.82	18.67	**19.85**
Posterior mean	DDNM+	30.10	**0.90**	17.72	39.07	24.40	0.83	20.02	25.54
	Posterior-DDIM (Ours)	**30.35**	**0.90**	**15.20**	28.57	**24.70**	**0.85**	**16.68**	18.73

A Additional experimental results

A.1 Choices of η_0 and η_1

We evaluate the influence of the hyperparameters η_0 and η_1 in Posterior-DDIM, which govern the DDIM updates on $\mathcal{S}^{\mathrm{c}}$ and $\mathcal{S}_t^{\mathsf{prior}}$, respectively. Experiments are conducted on two image restoration tasks: 50% random inpainting and $4\times$ super-resolution with average pooling, using the CelebA (Liu et al., 2015) and ImageNet (Deng et al., 2009) datasets. We set the noise standard deviation to $\sigma = 0.05$ (doubled when pixels are rescaled to $[-1, 1]$).

We vary η_1 from 0 to 32 while fixing $\eta_0 = 0$, and vary η_0 from 0 to 2 while fixing $\eta_1 = 16$. In implementation, we use $\sqrt{\max\{1 - \eta(1 - e^{-2\delta_{i+1}})\}}$ in case $1-\eta(1-e^{-2\delta_{i+1}})$ is negative for large η. The results for inpainting and super-resolution are reported in Table 3 and 4, respectively. For fixed η_0, increasing η_1 generally improves performance across most evaluation metrics, although the improvement is not strictly monotone and small fluctuations are observed. This overall trend is consistent with our theoretical analysis that requires η_1 to be sufficiently large. Turning to the second set of experiments, where η_1 is fixed and η_0 is varied, we find that fine-tuning η_0 can yield additional, albeit modest, performance gains. Empirically, the best empirical performance is attained at $\eta_0 = 0$, which corresponds to the deterministic DDIM update.

A.2 Additional results for empirical comparisons with reference algorithms

Table 5 presents the results on Gaussian deblurring task. Our algorithm achieves the best performance on three metrics on CelebA and two metrics on ImageNet. Table 6 compares our algorithm with DDNM+ on a compressed sensing task employing a Walsh-Hadamard sampling matrix at a compression ratio of 0.5. Our algorithm outperforms DDNM+ on three of the four metrics.

[2]When reproducing DDNM+ on the Gaussian deblurring task, we observe that the obtained results are substantially inferior to those reported in Zhang et al. (2024). Consequently, we adopt the results from Zhang et al. (2024); since only the published single-sample results are available, posterior-mean results for DDNM+ is lacking.

[3]Results for the compressed sensing task are not implemented in Zhang et al. (2024); therefore, we compare only with DDNM+. We note that Wang et al. (2022a) reported that DDNM+ outperforms DDRM with PSNR, SSIM, and FID metrics, and our algorithm outperforms DDNM+.

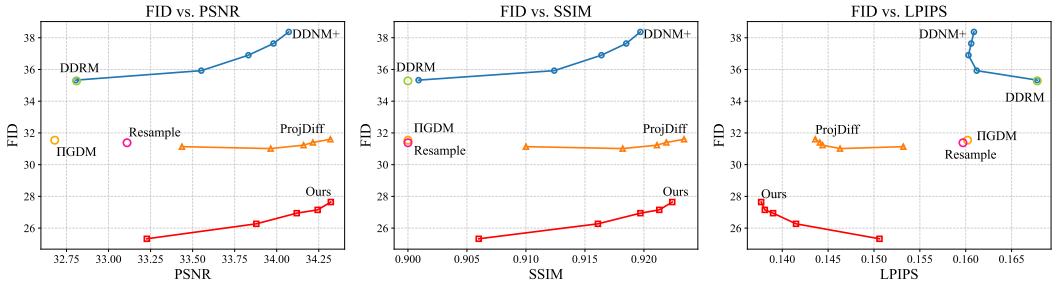

Figure 1: FID versus PSNR, SSIM, and LPIPS for different inpainting algorithms on the CelebA dataset.

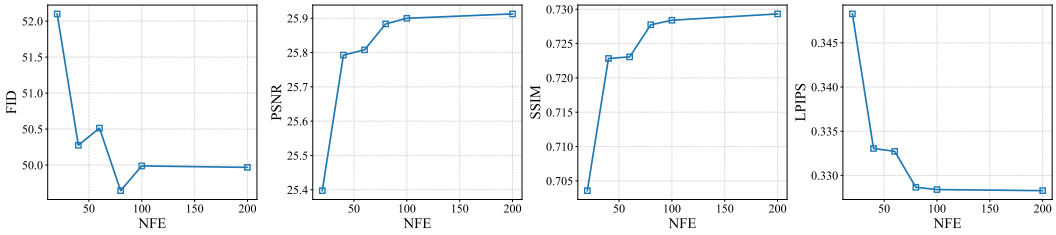

Figure 2: Effect of the NFE on the ImageNet super-resolution task.

A.3 Posterior mean

Figure 1 shows the performance of `Posterior-DDIM`, `DDNM+`, and `ProjDiff` under different numbers of Monte Carlo samples used to estimate the posterior mean. As discussed previously, increasing the number of samples improves reconstruction fidelity, as reflected by PSNR, SSIM, and LPIPS, but comes at the expense of reduced sample diversity, leading to worse FID scores. Furthermore, among all reference algorithms considered, our method achieves the best trade-off between fidelity and diversity on this task.

A.4 Impact of the NFE

Figure 2 illustrates the effect of the NFE on the super-resolution task using the ImageNet dataset. As the NFE increases, the performance improves, and becomes stable once NFE exceeds 80, suggesting that our default choice of NFE=100 provides a favorable balance between computational cost and reconstruction quality.

B Proof of Theorem 1 for identical singular values

For clarity, this section focuses on the proof under the assumption that all non-zero singular values of matrix A are equal. The proof without this assumption is similar and is deferred to Appendix C.8. For convenience, we introduce the following notations: for any vector $X_{t_i} \in \mathbb{R}^d$, denote

$$X_{t_i}^{\mathsf{st}} \coloneqq V_{\mathcal{S}_{t_i}^{\mathsf{meas}}}^\top X_{t_i}, \qquad X_{t_i}^{\mathsf{sc}} \coloneqq V_{\mathcal{S}_{t_i}^{\mathsf{prior}}}^\top X_{t_i}, \qquad X_{t_i}^{\mathsf{c}} \coloneqq V_{\mathcal{S}^{\mathsf{c}}}^\top X_{t_i}, \qquad X_{t_i}^{\mathsf{s}} \coloneqq V_{\mathcal{S}}^\top X_{t_i},$$

the projection of observation $\overline{Y}_{t_i} \coloneqq \Sigma_{\mathcal{S}_{t_i}^{\mathsf{meas}}}^{-1} U_{\mathcal{S}_{t_i}^{\mathsf{meas}}}^\top Y$, and the noise $\overline{Z}_{t_i} \coloneqq [Z_s]_{s \in \mathcal{S}_{t_i}^{\mathsf{meas}}}$ used to update $V_{\mathcal{S}_{t_i}^{\mathsf{meas}}}^\top \widehat{X}_{t_i}$. Additionally, throughout this paper, we denote by $\mathsf{KL}(p \,\|\, q)$ the Kullback-Leibler (KL) divergence between two probability measures p and q.

We intend to establish the following stronger conclusion than Theorem 1: for any t_i, conditioned on $V_{\mathcal{S}_{t_i}^{\mathsf{meas}}}^\top \widehat{X}_{t_i}$ and $v_s^\top \widehat{X}_{\tau_s}$ for all $\tau_s > t_i$, we have

$$\mathsf{TV}\big(p_{\widehat{X}_{t_i}|V_{\mathcal{S}_{t_i}^{\mathsf{meas}}}^\top \widehat{X}_{t_i}=V_{\mathcal{S}_{t_i}^{\mathsf{meas}}}^\top x_{t_i}, v_s^\top \widehat{X}_{\tau_s}=v_s^\top x_{\tau_s}}, p_{X_{t_i}|V_{\mathcal{S}_{t_i}^{\mathsf{meas}}}^\top X_{t_i}=V_{\mathcal{S}_{t_i}^{\mathsf{meas}}}^\top x_{t_i}, v_s^\top X_{\tau_s}=v_s^\top x_{\tau_s}}\big) \to 0.$$

When t_M is sufficiently close to 0, the set $\mathcal{S}_{t_M}^{\mathsf{meas}}$ becomes empty, and we obtain

$$\mathsf{TV}\big(p_{\widehat{X}_{t_M}|v_s^\top \widehat{X}_{\tau_s}=v_s^\top x_{\tau_s}}, p_{X_{t_M}|v_s^\top X_{\tau_s}=v_s^\top x_{\tau_s}}\big) \to 0.$$

Moreover, noticing that $v_s^\top \widehat{X}_{\tau_s} = \Sigma_{ss}^{-1} u_s^\top Y$, and $(X_{t_M}, v_s^\top X_{\tau_s})$ have the same joint distribution as $(X_{t_M}, \Sigma_{ss}^{-1} u_s^\top Y)$, we can complete the proof of Theorem 1.

Step 1. construct auxiliary sequences. We begin with introducing some auxiliary sequences. Note that $\sum_{i=1}^{M-1} \delta_i = \Theta(1)$. Denote

$$\gamma_i \coloneqq \frac{\eta_1(1-\mathrm{e}^{-2\delta_{i+1}})}{2}\sigma_{t_{i+1}}^2, \qquad p_{\lambda_i}(\cdot,\cdot) \coloneqq p_{X_{t_i}^{\mathsf{c}}, X_{t_i}^{\mathsf{sc}}|X_{t_i}^{\mathsf{st}}}(\cdot,\cdot|\alpha_{t_i}\overline{Y}_{t_i} + \widehat{\sigma}_{t_i}\overline{Z}_{t_i}), \tag{26}$$

where $\widehat{\sigma}_{t_i}$ is defined in (21). We have for $\delta_{i+1} \le 1/2$,

$$\frac{1}{2}\eta_1\delta_{i+1}\sigma_{t_{i+1}}^2 = \frac{\eta_1\delta_{i+1}}{2}\sigma_{t_{i+1}}^2 \le \gamma_i \le \frac{\eta_1\cdot 2\delta_{i+1}}{2}\sigma_{t_{i+1}}^2 = \eta_1\delta_{i+1}\sigma_{t_{i+1}}^2. \tag{27}$$

An auxiliary sequence $\widetilde{X}_k$ with initialization $\widetilde{X}_0 = \widehat{X}_{t_0}$ is iteratively defined as

$$\begin{aligned}
\widetilde{X}_{k+1}^{\mathsf{st}} &= \alpha_{t_{k+1}}\overline{Y}_{t_{k+1}} + \widehat{\sigma}_{t_{k+1}}\overline{Z}_{t_{k+1}},\\
\widetilde{X}_{k+1}^{\mathsf{sc}} &= \frac{\alpha_{t_{k+1}}}{\alpha_{t_k}}\widetilde{X}_{t_k}^{\mathsf{sc}} + \frac{\alpha_{t_{k+1}}\sigma_{t_k}^2}{\alpha_{t_k}}\left(1-\mathrm{e}^{-2\delta_{k+1}}\right)\nabla_{\widetilde{X}_k^{\mathsf{sc}}}\log p_{\lambda_k}(\widetilde{X}_k^{\mathsf{c}}, \widetilde{X}_k^{\mathsf{sc}}) + \sigma_{t_{k+1}}\sqrt{1-e^{-2\delta_{k+1}}}Z_k^{\mathsf{sc}},\\
\widetilde{X}_{k+1}^{\mathsf{c}} &= \widetilde{X}_k^{\mathsf{c}} + \gamma_k\nabla_{\widetilde{X}_k^{\mathsf{c}}}\log p_{\lambda_{k+1}}(\widetilde{X}_k^{\mathsf{c}}, \widetilde{X}_{k+1}^{\mathsf{sc}}) + \sqrt{2\gamma_k}Z_k^{\mathsf{c}},
\end{aligned} \tag{28a}$$

where Z_k^{sc} and Z_k^{c} are independent Gaussian random processes. Another auxiliary sequence is defined as

$$\overline{X}_{k+1}^{\mathsf{st}} = \alpha_{t_{k+1}}\overline{Y}_{t_{k+1}} + \widehat{\sigma}_{t_{k+1}}\overline{Z}_{t_{k+1}}, \quad \overline{X}_{k+1}^{\mathsf{sc}} \sim p_{X_{k+1}^{\mathsf{sc}}|X_k}(\cdot \mid \overline{X}_k), \quad \overline{X}_{k+1}^{\mathsf{c}} \sim p_{X_{k+1}^{\mathsf{c}}|X_{k+1}^{\mathsf{s}}}(\cdot \mid \overline{X}_{k+1}^{\mathsf{s}}). \tag{29}$$

Based on these sequences, we have the following lemma, whose proof is postponed to Section C.3.

Lemma 1. *The auxiliary sequence defined in (28) satisfies*

$$\mathsf{KL}(p_{\widetilde{X}_{0:M}\,|\,Y,Z} \,\|\, p_{\widehat{X}_{t_{0:M}}\,|\,Y,Z}) \to 0, \qquad \text{as} \quad \eta_1^4\delta^3 \to 0, \quad \text{and} \quad \eta_1 \to \infty. \tag{30}$$

This implies that conditioned on observation Y and Gaussian noise Z, random variables $\widetilde{X}_{0:M}$ converges to the same distribution with $\widehat{X}_{t_{0:M}}$ if this limit exists. Moreover, the sequence defined in (29) satisfies $p_{\overline{X}_k\,|\,V_{\mathcal{S}_{t_k}^{\mathsf{meas}}}^\top \overline{X}_k, v_s^\top \overline{X}_{k_{\tau_s}}} = p_{X_{t_k}\,|\,V_{\mathcal{S}_{t_k}^{\mathsf{meas}}}^\top X_{t_k}, v_s^\top X_{\tau_s}}$, where k_{τ_s} satisfies $t_{k_{\tau_s}} = \tau_s$.

This lemma tells us that the marginal distribution $\mathsf{TV}\big(p_{\widetilde{X}_k\,|\,Y,Z}, p_{\widehat{X}_k\,|\,Y,Z}\big) \to 0$. Denote $\overline{Z}_{t_i}^{\mathsf{c}} = [Z_s]_{s\in\mathcal{S}_{t_i}^{\mathsf{prior}}}$. The convergence of marginal distribution leads to

$$\begin{aligned}
&\mathsf{TV}\big(p_{\widetilde{X}_k\,|\,V_{\mathcal{S}_{t_k}^{\mathsf{meas}}}^\top \widetilde{X}_k, v_s^\top \widetilde{X}_{\tau_s}}(\cdot, x^{\mathsf{st}}, v_s^\top x_{\tau_s}), p_{\widehat{X}_{t_k}\,|\,V_{\mathcal{S}_{t_k}^{\mathsf{meas}}}^\top \widehat{X}_{t_k}, v_s^\top \widehat{X}_{\tau_s}}(\cdot, x^{\mathsf{st}}, v_s^\top x_{\tau_s})\big)\\
&\quad\le \mathbb{E}_{\overline{Y}_{t_k}, \overline{Z}_{t_k}^{\mathsf{c}}|V_{\mathcal{S}_{t_k}^{\mathsf{meas}}}^\top \widetilde{X}_k}\big[\mathsf{TV}\big(p_{\widetilde{X}_k\,|\,V_{\mathcal{S}_{t_k}^{\mathsf{meas}}}^\top \widetilde{X}_k, v_s^\top \widetilde{X}_{\tau_s}, \overline{Y}_{t_k}, \overline{Z}_{t_k}^{\mathsf{c}}}(\cdot\,|\,x^{\mathsf{st}}, v_s^\top x_{\tau_s}, \overline{y}_{t_k}, \widehat{z}),\\
&\qquad\qquad p_{\widehat{X}_{t_k}\,|\,V_{\mathcal{S}_{t_k}^{\mathsf{meas}}}^\top \widehat{X}_{t_k}, v_s^\top \widehat{X}_{\tau_s}, \overline{Y}_{t_k}, \overline{Z}_{t_k}^{\mathsf{c}}}(\cdot\,|\,x^{\mathsf{st}}, v_s^\top x_{\tau_s}, \overline{y}_{t_k}, \widehat{z})\big)\big]\\
&\quad= \mathbb{E}_{\overline{Y}_{t_k}, \overline{Z}_{t_k}^{\mathsf{c}}|V_{\mathcal{S}_{t_k}^{\mathsf{meas}}}^\top \widetilde{X}_k}\big[\mathsf{TV}\big(p_{\widetilde{X}_k\,|\,Y,Z}(\cdot\,|\,y,z), p_{\widehat{X}_{t_k}\,|\,Y,Z}(\cdot\,|\,y,z)\big)\big] \to 0,
\end{aligned} \tag{31}$$

where $v_s^\top x_{\tau_s} = \Sigma_{ss}^{-1} u_s^\top y$, $\overline{z}_{t_k} = \widehat{\sigma}_{t_k}^{-1}(x^{\mathsf{st}} - \alpha_{t_k}\overline{y}_{t_k})$, and the last equation holds since $p_{\widetilde{X}_k \mid V_{\mathcal{S}^{\mathsf{meas}}_{t_k}}^\top \widetilde{X}_k, v_s^\top \widetilde{X}_{\tau_s}, \overline{Y}_{t_k}, \overline{Z}^{\mathsf{c}}_{t_k}} = p_{\widetilde{X}_k \mid Y,Z}$ and $p_{\widehat{X}_k \mid V_{\mathcal{S}^{\mathsf{meas}}_{t_k}}^\top \widehat{X}_{t_k}, v_s^\top \widehat{X}_{\tau_s}, \overline{Y}_{t_k}, \overline{Z}^{\mathsf{c}}_{t_k}} = p_{\widehat{X}_{t_k} \mid Y,Z}$. Based on Lemma 1 and (31), it suffices to prove that for any x^{st} and x_{τ_s},

$$\mathsf{TV}\big(p_{\widetilde{X}_k \mid \widetilde{X}_k^{\mathsf{st}}, v_s^\top \widetilde{X}_{\tau_s}}(\cdot \mid x^{\mathsf{st}}, v_s^\top x_{\tau_s}), p_{\overline{X}_{t_k} \mid \overline{X}_{t_k}^{\mathsf{st}}, v_s^\top \overline{X}_{\tau_s}}(\cdot \mid x^{\mathsf{st}}, v_s^\top x_{\tau_s})\big) \to 0.$$

By a similar argument as in (31), it suffices to prove that for any y, z,

$$\mathsf{TV}\big(p_{\widetilde{X}_k \mid Y,Z}(\cdot \mid y, z), p_{\overline{X}_{t_k} \mid Y,Z}(\cdot \mid y, z)\big) \to 0.$$

Step 2. transform to the analysis of auxiliary sequences. We begin with the decomposition that

$$p_{\widetilde{X}_k \mid Y,Z}(x|y,z) = p_{\widetilde{X}_k^{\mathsf{c}} \mid \widetilde{X}_k^{\mathsf{s}}, Y,Z}(x^{\mathsf{c}}|x^{\mathsf{s}}, y, z) p_{\widetilde{X}_k^{\mathsf{sc}} \mid Y,Z}(x^{\mathsf{sc}}|y,z),$$
$$p_{\overline{X}_k \mid Y,Z}(x \mid y,z) = p_{\overline{X}_k^{\mathsf{c}} \mid \overline{X}_k^{\mathsf{s}}, Y,Z}(x^{\mathsf{c}} \mid x^{\mathsf{s}}, y, z) p_{\overline{X}_k^{\mathsf{sc}} \mid Y,Z}(x^{\mathsf{sc}} \mid y, z)$$

Then the total variation distance can be decomposed by

$$\begin{aligned}\mathsf{TV}\Big(p_{\widetilde{X}_k \mid Y,Z}, p_{\overline{X}_k \mid Y,Z}\Big) &\le \mathsf{TV}\Big(p_{\widetilde{X}_k^{\mathsf{sc}} \mid Y,Z}, p_{\overline{X}_k^{\mathsf{sc}} \mid Y,Z}\Big) \\ &\quad + \mathbb{E}_{x^{\mathsf{sc}} \sim p_{\widetilde{X}_k^{\mathsf{sc}} \mid Y,Z}}\left[\mathsf{TV}\left(p_{\widetilde{X}_k^{\mathsf{c}} \mid \widetilde{X}_k^{\mathsf{s}}, Y,Z}(\cdot|x^{\mathsf{s}}, y, z), p_{\overline{X}_k^{\mathsf{c}} \mid \overline{X}_k^{\mathsf{s}}, Y,Z}(\cdot|x^{\mathsf{s}}, y, z)\right)\right]. \end{aligned} \tag{32}$$

The convergence of the second terms comes from the following lemma:

Lemma 2. *Suppose that η_1 is large enough and δ small enough. As $\eta_1 \delta \log \frac{1}{\delta} \to 0$ and $\eta_1^2 \delta / \log \frac{1}{\delta} \to \infty$, for any x,*

$$\mathbb{E}_{x \sim p_{\widehat{X}_t}}\left[\mathsf{TV}\left(p_{V_{\mathcal{S}^{\mathsf{c}}}^\top \widehat{X}_t | V_{\mathcal{S}}^\top \widehat{X}_t}(\cdot \mid V_{\mathcal{S}}^\top x), p_{V_{\mathcal{S}^{\mathsf{c}}}^\top X_t | V_{\mathcal{S}}^\top X_t}(\cdot \mid V_{\mathcal{S}}^\top x)\right)\right] \to 0.$$

The proof is postponed to Appendix C.4. Here we use a more strong conclusion than Lemma 2, i.e., for any y, z,

$$\mathbb{E}_{x^{\mathsf{sc}} \sim \widetilde{X}_k^{\mathsf{sc}} \mid Y,Z}\left[\mathsf{TV}\left(p_{\widetilde{X}_k^{\mathsf{c}} | \widetilde{X}_k^{\mathsf{s}}, Y,Z}(\cdot \mid x^{\mathsf{s}}, y, z), p_{X_{t_k}^{\mathsf{c}} | X_{t_k}^{\mathsf{s}}}(\cdot \mid x^{\mathsf{s}})\right)\right] \to 0. \tag{33}$$

The proof is postponed to Section C.4 (cf. (59)). Then the second term goes to zero:

$$\begin{aligned}&\mathbb{E}_{x^{\mathsf{sc}} \sim p_{\widetilde{X}_k^{\mathsf{sc}} \mid Y,Z}}\left[\mathsf{TV}\left(p_{\widetilde{X}_k^{\mathsf{c}} \mid \widetilde{X}_k^{\mathsf{s}}, Y,Z}(\cdot|x^{\mathsf{s}}, y, z), p_{\overline{X}_k^{\mathsf{c}} \mid \overline{X}_k^{\mathsf{s}}, Y,Z}(\cdot|x^{\mathsf{s}}, y, z)\right)\right] \\ &\overset{\text{(i)}}{=} \mathbb{E}_{x^{\mathsf{sc}} \sim p_{\widetilde{X}_k^{\mathsf{sc}} \mid Y,Z}}\left[\mathsf{TV}\left(p_{\widetilde{X}_k^{\mathsf{c}} \mid \widetilde{X}_k^{\mathsf{s}}, Y,Z}(\cdot|x^{\mathsf{s}}, y, z), p_{X_{t_k}^{\mathsf{c}} \mid X_{t_k}^{\mathsf{s}}}(\cdot|x^{\mathsf{s}})\right)\right] \to 0, \end{aligned} \tag{34}$$

where (i) arises from the fact that $p_{\overline{X}_k^{\mathsf{c}} \mid \overline{X}_k^{\mathsf{s}}, Y,Z} = p_{X_k^{\mathsf{c}} \mid X_k^{\mathsf{s}}}$. It suffices to prove the first term goes to zero, i.e.,

$$\mathsf{TV}\big(p_{\widetilde{X}_k^{\mathsf{sc}} \mid Y,Z}, p_{\overline{X}_k^{\mathsf{sc}} \mid Y,Z}\big) \to 0. \tag{35}$$

Step 3. decompose the error terms. By virtue of Pinsker's inequality, we further decompose the total variation distance as

$$\begin{aligned}&\mathsf{TV}\big(p_{\widetilde{X}_k^{\mathsf{sc}} \mid Y,Z}, p_{\overline{X}_k^{\mathsf{sc}} \mid Y,Z}\big)^2 \\ &\le \mathsf{KL}\left(p_{\widetilde{X}_k^{\mathsf{sc}} \mid Y,Z} \,\|\, p_{\overline{X}_k^{\mathsf{sc}} \mid Y,Z}\right) \\ &\le \sum_{i=1}^k \mathbb{E}_{x_{i-1}^{\mathsf{sc}} \sim p_{\widetilde{X}_{i-1}^{\mathsf{sc}} \mid Y,Z}}\left[\mathsf{KL}\left(p_{\widetilde{X}_i^{\mathsf{sc}} \mid \widetilde{X}_{i-1}^{\mathsf{s}}, Y,Z}(\cdot \mid x_{i-1}^{\mathsf{s}}, y, z) \,\|\, p_{\overline{X}_i^{\mathsf{sc}} \mid \overline{X}_{i-1}^{\mathsf{s}}, Y,Z}(\cdot \mid x_{i-1}^{\mathsf{s}}, y, z)\right)\right] + \mathsf{KL}(p_{\widetilde{X}_0^{\mathsf{sc}}} \,\|\, p_{\overline{X}_0^{\mathsf{sc}}}) \\ &= \sum_{i=1}^k \mathbb{E}_{x_{i-1}^{\mathsf{sc}} \sim p_{\widetilde{X}_{i-1}^{\mathsf{sc}} \mid Y,Z}}\left[\mathsf{KL}\left(p_{\widetilde{X}_i^{\mathsf{sc}} \mid \widetilde{X}_{i-1}^{\mathsf{s}}, Y,Z}(\cdot|x_{i-1}^{\mathsf{s}}, y, z) \| p_{\overline{X}_i^{\mathsf{sc}} \mid \overline{X}_{i-1}^{\mathsf{s}}}(\cdot|x_{i-1}^{\mathsf{s}})\right)\right] + \mathsf{KL}(p_{\widetilde{X}_0^{\mathsf{sc}}} \,\|\, p_{\overline{X}_0^{\mathsf{sc}}}), \end{aligned} \tag{36}$$

where x_{i-1}^{s} satisfies $x_{i-1}^{\mathsf{st}} = \alpha_{t_{i-1}} \Sigma_{\mathcal{S}_{t_{i-1}}}^{-1} U_{\mathcal{S}_{t_{i-1}}}^{\top} y + \widehat{\sigma}_{t_{i-1}} [Z_s]_{s \in \mathcal{S}_{t_{i-1}}}$, and the last line holds since $p_{\overline{X}_i^{\mathsf{sc}} \mid \overline{X}_{i-1}^{\mathsf{s}}, Y, Z} = p_{\overline{X}_i^{\mathsf{sc}} \mid \overline{X}_{i-1}^{\mathsf{s}}}$. According to the definition of $\overline{X}_i$, we further have

$$
\begin{aligned}
&\mathsf{KL}\left(p_{\widetilde{X}_i^{\mathsf{sc}} \mid \widetilde{X}_{i-1}^{\mathsf{s}}, Y, Z}(\cdot | x_{i-1}^{\mathsf{s}}, y, z) \,\|\, p_{\overline{X}_i^{\mathsf{sc}} \mid \overline{X}_{i-1}^{\mathsf{s}}}(\cdot | x_{i-1}^{\mathsf{s}}) \right) \\
&\quad \overset{\text{(i)}}{=} \mathsf{KL}\left(p_{\widetilde{X}_i^{\mathsf{sc}} \mid \widetilde{X}_{i-1}^{\mathsf{s}}, Y, Z}(\cdot | x_{i-1}^{\mathsf{s}}, y, z) \,\|\, p_{X_{t_i}^{\mathsf{sc}} \mid X_{t_{i-1}}^{\mathsf{s}}}(\cdot | x_{i-1}^{\mathsf{s}}) \right) \\
&\quad = \int \log \frac{p_{\widetilde{X}_i^{\mathsf{sc}} \mid \widetilde{X}_{i-1}^{\mathsf{s}}, Y, Z}(x_i^{\mathsf{sc}} \mid x_{i-1}^{\mathsf{s}}, y, z)}{p_{X_{t_i}^{\mathsf{sc}} \mid X_{t_{i-1}}^{\mathsf{s}}}(x_i^{\mathsf{sc}} \mid x_{i-1}^{\mathsf{s}})} p_{\widetilde{X}_i^{\mathsf{sc}} \mid \widetilde{X}_{i-1}^{\mathsf{s}}, Y, Z}(x_i^{\mathsf{sc}} \mid x_{i-1}^{\mathsf{s}}, y, z) \mathrm{d} x_i^{\mathsf{sc}}, \qquad (37)
\end{aligned}
$$

where (i) holds since

$$
\begin{aligned}
p_{\overline{X}_i^{\mathsf{sc}} \mid \overline{X}_{i-1}^{\mathsf{s}}}(x_i^{\mathsf{sc}} \mid x_{i-1}^{\mathsf{s}}) &= \int p_{\overline{X}_i^{\mathsf{sc}} \mid \overline{X}_{i-1}}(x_i^{\mathsf{sc}} \mid x_{i-1}) p_{\overline{X}_{i-1}^{\mathsf{c}} \mid \overline{X}_{i-1}^{\mathsf{s}}}(x_{i-1}^{\mathsf{c}} \mid x_{i-1}^{\mathsf{s}}) \mathrm{d} x_{i-1}^{\mathsf{c}} \\
&= \int p_{X_{t_i}^{\mathsf{sc}} \mid X_{t_{i-1}}}(x_i^{\mathsf{sc}} \mid x_{i-1}) p_{X_{t_{i-1}}^{\mathsf{c}} \mid X_{t_{i-1}}^{\mathsf{s}}}(x_{i-1}^{\mathsf{c}} \mid x_{i-1}^{\mathsf{s}}) \mathrm{d} x_{i-1}^{\mathsf{c}} = p_{X_{t_i}^{\mathsf{sc}} \mid X_{t_{i-1}}^{\mathsf{s}}}(x_i^{\mathsf{sc}} \mid x_{i-1}^{\mathsf{s}}).
\end{aligned}
$$

We further decompose the KL divergence in (37) as

$$
\begin{aligned}
&\mathsf{KL}\left(p_{\widetilde{X}_i^{\mathsf{sc}} \mid \widetilde{X}_{i-1}^{\mathsf{s}}, Y, Z}(\cdot | x_{i-1}^{\mathsf{s}}, y, z) \,\|\, p_{\overline{X}_i^{\mathsf{sc}} \mid \overline{X}_{i-1}^{\mathsf{s}}}(\cdot | x_{i-1}^{\mathsf{s}}) \right) \\
&\quad = \int \log \frac{\int p_{\widetilde{X}_i^{\mathsf{sc}} \mid \widetilde{X}_{i-1}, Y, Z}(x_i^{\mathsf{sc}} \mid x_{i-1}, y, z) p_{\widetilde{X}_{i-1}^{\mathsf{c}} \mid \widetilde{X}_{i-1}^{\mathsf{s}}, Y, Z}(x_{i-1}^{\mathsf{c}} \mid x_{i-1}^{\mathsf{s}}, y, z) \mathrm{d} x_{i-1}^{\mathsf{c}}}{\int p_{X_{t_i}^{\mathsf{sc}} \mid X_{t_{i-1}}}(x_i^{\mathsf{sc}} \mid x_{i-1}) p_{X_{t_{i-1}}^{\mathsf{c}} \mid X_{t_{i-1}}^{\mathsf{s}}}(x_{i-1}^{\mathsf{c}} \mid x_{i-1}^{\mathsf{s}}) \mathrm{d} x_{i-1}^{\mathsf{c}}} \\
&\qquad \times p_{\widetilde{X}_i^{\mathsf{sc}} \mid \widetilde{X}_{i-1}^{\mathsf{s}}, Y, Z}(x_i^{\mathsf{sc}} \mid x_{i-1}^{\mathsf{s}}, y, z) \mathrm{d} x_i^{\mathsf{sc}} \\
&\quad = \int \log \frac{\int p_{\widetilde{X}_i^{\mathsf{sc}} \mid \widetilde{X}_{i-1}, Y, Z}(x_i^{\mathsf{sc}} \mid x_{i-1}, y, z) p_{X_{t_{i-1}}^{\mathsf{c}} \mid X_{t_{i-1}}^{\mathsf{s}}}(x_{i-1}^{\mathsf{c}} \mid x_{i-1}^{\mathsf{s}}) \mathrm{d} x_{i-1}^{\mathsf{c}}}{\int p_{X_{t_i}^{\mathsf{sc}} \mid X_{t_{i-1}}}(x_i^{\mathsf{sc}} \mid x_{i-1}) p_{X_{t_{i-1}}^{\mathsf{c}} \mid X_{t_{i-1}}^{\mathsf{s}}}(x_{i-1}^{\mathsf{c}} \mid x_{i-1}^{\mathsf{s}}) \mathrm{d} x_{i-1}^{\mathsf{c}}} \\
&\qquad \times p_{\widetilde{X}_i^{\mathsf{sc}} \mid \widetilde{X}_{i-1}^{\mathsf{s}}, Y, Z}(x_i^{\mathsf{sc}} \mid x_{i-1}^{\mathsf{s}}, y, z) \mathrm{d} x_i^{\mathsf{sc}} \\
&\qquad + \int \log \frac{\int p_{\widetilde{X}_i^{\mathsf{sc}} \mid \widetilde{X}_{i-1}, Y, Z}(x_i^{\mathsf{sc}} \mid x_{i-1}, y, z) p_{\widetilde{X}_{i-1}^{\mathsf{c}} \mid \widetilde{X}_{i-1}^{\mathsf{s}}, Y, Z}(x_{i-1}^{\mathsf{c}} \mid x_{i-1}^{\mathsf{s}}, y, z) \mathrm{d} x_{i-1}^{\mathsf{c}}}{\int p_{\widetilde{X}_i^{\mathsf{sc}} \mid \widetilde{X}_{i-1}, Y, Z}(x_i^{\mathsf{sc}} \mid x_{i-1}, y, z) p_{X_{t_{i-1}}^{\mathsf{c}} \mid X_{t_{i-1}}^{\mathsf{s}}}(x_{i-1}^{\mathsf{c}} \mid x_{i-1}^{\mathsf{s}}) \mathrm{d} x_{i-1}^{\mathsf{c}}} \\
&\qquad \times p_{\widetilde{X}_i^{\mathsf{sc}} \mid \widetilde{X}_{i-1}^{\mathsf{s}}, Y, Z}(x_i^{\mathsf{sc}} \mid x_{i-1}^{\mathsf{s}}, y, z) \mathrm{d} x_i^{\mathsf{sc}}. \qquad (38)
\end{aligned}
$$

The first term is further decomposed by

$$
\begin{aligned}
&\int \log \frac{\int p_{\widetilde{X}_i^{\mathsf{sc}} \mid \widetilde{X}_{i-1}, Y, Z}(x_i^{\mathsf{sc}} \mid x_{i-1}, y, z) p_{X_{t_{i-1}}^{\mathsf{c}} \mid X_{t_{i-1}}^{\mathsf{s}}}(x_{i-1}^{\mathsf{c}} \mid x_{i-1}^{\mathsf{s}}) \mathrm{d} x_{i-1}^{\mathsf{c}}}{\int p_{X_{t_i}^{\mathsf{sc}} \mid X_{t_{i-1}}}(x_i^{\mathsf{sc}} \mid x_{i-1}) p_{X_{t_{i-1}}^{\mathsf{c}} \mid X_{t_{i-1}}^{\mathsf{s}}}(x_{i-1}^{\mathsf{c}} \mid x_{i-1}^{\mathsf{s}}) \mathrm{d} x_{i-1}^{\mathsf{c}}} p_{\widetilde{X}_i^{\mathsf{sc}} \mid \widetilde{X}_{i-1}^{\mathsf{s}}, Y, Z}(x_i^{\mathsf{sc}} \mid x_{i-1}^{\mathsf{s}}, y, z) \mathrm{d} x_i^{\mathsf{sc}} \\
&\quad = \int f(x_i^{\mathsf{sc}}, x_{i-1}^{\mathsf{s}}) \int p_{\widetilde{X}_i^{\mathsf{sc}} \mid \widetilde{X}_{i-1}, Y, Z}(x_i^{\mathsf{sc}} \mid x_{i-1}, y, z) p_{X_{t_{i-1}}^{\mathsf{c}} \mid X_{t_{i-1}}^{\mathsf{s}}}(x_{i-1}^{\mathsf{c}} \mid x_{i-1}^{\mathsf{s}}) \mathrm{d} x_{i-1}^{\mathsf{c}} \mathrm{d} x_i^{\mathsf{sc}} \\
&\qquad + \int f(x_i^{\mathsf{sc}}, x_{i-1}^{\mathsf{s}}) \int p_{\widetilde{X}_i^{\mathsf{sc}} \mid \widetilde{X}_{i-1}, Y, Z}(x_i^{\mathsf{sc}} \mid x_{i-1}, y, z) \\
&\qquad \cdot \left(p_{\widetilde{X}_{i-1}^{\mathsf{c}} \mid \widetilde{X}_{i-1}^{\mathsf{s}}, Y, Z}(x_{i-1}^{\mathsf{c}} \mid x_{i-1}^{\mathsf{s}}, y, z) - p_{X_{t_{i-1}}^{\mathsf{c}} \mid X_{t_{i-1}}^{\mathsf{s}}}(x_{i-1}^{\mathsf{c}} \mid x_{i-1}^{\mathsf{s}}) \right) \mathrm{d} x_{i-1}^{\mathsf{c}} \mathrm{d} x_i^{\mathsf{sc}}, \qquad (39)
\end{aligned}
$$

where

$$
f(x_i^{\mathsf{sc}}, x_{i-1}^{\mathsf{s}}) \coloneqq \log \frac{\int p_{\widetilde{X}_i^{\mathsf{sc}} \mid \widetilde{X}_{i-1}, Y, Z}(x_i^{\mathsf{sc}} \mid x_{i-1}, y, z) p_{X_{t_{i-1}}^{\mathsf{c}} \mid X_{t_{i-1}}^{\mathsf{s}}}(x_{i-1}^{\mathsf{c}} \mid x_{i-1}^{\mathsf{s}}) \mathrm{d} x_{i-1}^{\mathsf{c}}}{\int p_{X_{t_i}^{\mathsf{sc}} \mid X_{t_{i-1}}}(x_i^{\mathsf{sc}} \mid x_{i-1}) p_{X_{t_{i-1}}^{\mathsf{c}} \mid X_{t_{i-1}}^{\mathsf{s}}}(x_{i-1}^{\mathsf{c}} \mid x_{i-1}^{\mathsf{s}}) \mathrm{d} x_{i-1}^{\mathsf{c}}}.
$$

By using (33), the second term in the right-hand-side of (39) satisfies

$$
\begin{aligned}
&\mathbb{E}_{x_{i-1}^{\mathsf{sc}} \sim p_{\overline{X}_{i-1}^{\mathsf{sc}} \mid Y, Z}} \Big[\int f(x_i^{\mathsf{sc}}, x_{i-1}^{\mathsf{s}}) \int p_{\widetilde{X}_i^{\mathsf{sc}} \mid \widetilde{X}_{i-1}, Y, Z}(x_i^{\mathsf{sc}} \mid x_{i-1}, y, z) \\
&\qquad \cdot \left(p_{\widetilde{X}_{i-1}^{\mathsf{c}} \mid \widetilde{X}_{i-1}^{\mathsf{s}}, Y, Z}(x_{i-1}^{\mathsf{c}} \mid x_{i-1}^{\mathsf{s}}, y, z) - p_{X_{t_{i-1}}^{\mathsf{c}} \mid X_{t_{i-1}}^{\mathsf{s}}}(x_{i-1}^{\mathsf{c}} \mid x_{i-1}^{\mathsf{s}}) \right) \mathrm{d} x_{i-1}^{\mathsf{c}} \mathrm{d} x_i^{\mathsf{sc}} \Big]
\end{aligned}
$$

$$
\begin{aligned}
&= \mathbb{E}_{x^{\mathsf{sc}}_{i-1}\sim p_{\widetilde{X}^{\mathsf{sc}}_{i-1}\,|\,Y,Z}}\Big[\int\Big(\int f(x^{\mathsf{sc}}_i, x^{\mathsf{s}}_{i-1}) p_{\widetilde{X}^{\mathsf{sc}}_i\,|\,\widetilde{X}_{i-1},Y,Z}(x^{\mathsf{sc}}_i\,|\,x_{i-1},y,z)\mathrm{d}x^{\mathsf{sc}}_i\Big)\\
&\qquad\cdot\Big(p_{\widetilde{X}^{\mathsf{c}}_{i-1}\,|\,\widetilde{X}^{\mathsf{s}}_{i-1},Y,Z}(x^{\mathsf{c}}_{i-1}\,|\,x^{\mathsf{s}}_{i-1},y,z) - p_{X^{\mathsf{c}}_{t_{i-1}}\,|\,X^{\mathsf{s}}_{t_{i-1}}}(x^{\mathsf{c}}_{i-1}\,|\,x^{\mathsf{s}}_{i-1})\Big)\,\mathrm{d}x^{\mathsf{c}}_{i-1}\Big]\\
&\le \mathbb{E}_{x^{\mathsf{sc}}_{i-1}\sim p_{\widetilde{X}^{\mathsf{sc}}_{i-1}\,|\,Y,Z}}\Big[\int\Big(\int f(x^{\mathsf{sc}}_i, x^{\mathsf{s}}_{i-1}) p_{\widetilde{X}^{\mathsf{sc}}_i\,|\,\widetilde{X}_{i-1},Y,Z}(x^{\mathsf{sc}}_i\,|\,x_{i-1},y,z)\mathrm{d}x^{\mathsf{sc}}_i\Big)^2\\
&\qquad\cdot\Big|p_{\widetilde{X}^{\mathsf{c}}_{i-1}\,|\,\widetilde{X}^{\mathsf{s}}_{i-1},Y,Z}(x^{\mathsf{c}}_{i-1}\,|\,x^{\mathsf{s}}_{i-1},y,z) - p_{X^{\mathsf{c}}_{t_{i-1}}\,|\,X^{\mathsf{s}}_{t_{i-1}}}(x^{\mathsf{c}}_{i-1}\,|\,x^{\mathsf{s}}_{i-1})\Big|\,\mathrm{d}x^{\mathsf{c}}_{i-1}\Big]^{1/2}\\
&\qquad\cdot\mathbb{E}_{x^{\mathsf{sc}}_{i-1}\sim p_{\widetilde{X}^{\mathsf{sc}}_{i-1}\,|\,Y,Z}}\Big[\int\Big|p_{\widetilde{X}^{\mathsf{c}}_{i-1}\,|\,\widetilde{X}^{\mathsf{s}}_{i-1},Y,Z}(x^{\mathsf{c}}_{i-1}\,|\,x^{\mathsf{s}}_{i-1},y,z) - p_{X^{\mathsf{c}}_{t_{i-1}}\,|\,X^{\mathsf{s}}_{t_{i-1}}}(x^{\mathsf{c}}_{i-1}\,|\,x^{\mathsf{s}}_{i-1})\Big|\,\mathrm{d}x^{\mathsf{c}}_{i-1}\Big]^{1/2}\\
&\to 0. \qquad (40)
\end{aligned}
$$

For the first term in the right-hand-side of (39), note that it is the KL divergence between the marginal distribution of x^{sc}_i, and it is smaller than the KL divergence of the joint distribution of x^{sc}_i and x^{c}_{i-1}. Therefore we have

$$
\begin{aligned}
&\int\log\frac{\int p_{\widetilde{X}^{\mathsf{sc}}_i\,|\,\widetilde{X}_{i-1},Y,Z}(x^{\mathsf{sc}}_i\,|\,x_{i-1},y,z)p_{X^{\mathsf{c}}_{t_{i-1}}\,|\,X^{\mathsf{s}}_{t_{i-1}}}(x^{\mathsf{c}}_{i-1}\,|\,x^{\mathsf{s}}_{i-1})\mathrm{d}x^{\mathsf{c}}_{i-1}}{\int p_{X^{\mathsf{sc}}_{t_i}\,|\,X_{t_{i-1}}}(x^{\mathsf{sc}}_i\,|\,x_{i-1})p_{X^{\mathsf{c}}_{t_{i-1}}\,|\,X^{\mathsf{s}}_{t_{i-1}}}(x^{\mathsf{c}}_{i-1}\,|\,x^{\mathsf{s}}_{i-1})\mathrm{d}x^{\mathsf{c}}_{i-1}}\\
&\qquad\cdot\int p_{\widetilde{X}^{\mathsf{sc}}_i\,|\,\widetilde{X}_{i-1},Y,Z}(x^{\mathsf{sc}}_i\,|\,x_{i-1},y,z)p_{X^{\mathsf{c}}_{t_{i-1}}\,|\,X^{\mathsf{s}}_{t_{i-1}}}(x^{\mathsf{c}}_{i-1}\,|\,x^{\mathsf{s}}_{i-1})\mathrm{d}x^{\mathsf{c}}_{i-1}\mathrm{d}x^{\mathsf{sc}}_i\\
&\le\iint\log\frac{p_{\widetilde{X}^{\mathsf{sc}}_i\,|\,\widetilde{X}_{i-1},Y,Z}(x^{\mathsf{sc}}_i\,|\,x_{i-1},y,z)p_{X^{\mathsf{c}}_{t_{i-1}}\,|\,X^{\mathsf{s}}_{t_{i-1}}}(x^{\mathsf{c}}_{i-1}\,|\,x^{\mathsf{s}}_{i-1})}{p_{X^{\mathsf{sc}}_{t_i}\,|\,X_{t_{i-1}}}(x^{\mathsf{sc}}_i\,|\,x_{i-1})p_{X^{\mathsf{c}}_{t_{i-1}}\,|\,X^{\mathsf{s}}_{t_{i-1}}}(x^{\mathsf{c}}_{i-1}\,|\,x^{\mathsf{s}}_{i-1})}\\
&\qquad\times p_{\widetilde{X}^{\mathsf{sc}}_i\,|\,\widetilde{X}_{i-1},Y,Z}(x^{\mathsf{sc}}_i\,|\,x_{i-1},y,z)p_{X^{\mathsf{c}}_{t_{i-1}}\,|\,X^{\mathsf{s}}_{t_{i-1}}}(x^{\mathsf{c}}_{i-1}\,|\,x^{\mathsf{s}}_{i-1})\mathrm{d}x^{\mathsf{c}}_{i-1}\mathrm{d}x^{\mathsf{sc}}_i\\
&= \mathbb{E}_{x^{\mathsf{c}}_{i-1}\sim X^{\mathsf{c}}_{t_{i-1}}\,|\,X^{\mathsf{s}}_{t_{i-1}}}\big[\mathsf{KL}(p_{\widetilde{X}^{\mathsf{sc}}_i\,|\,\widetilde{X}_{i-1},Y,Z}(\cdot\,|\,x_{i-1},y,z)\,\|\,p_{X^{\mathsf{sc}}_{t_i}\,|\,X_{t_{i-1}}}(\cdot\,|\,x_{i-1}))\big]. \qquad (41)
\end{aligned}
$$

Substituting (41) and (40) into (39), we have

$$
\begin{aligned}
&\int\log\frac{\int p_{\widetilde{X}^{\mathsf{sc}}_i\,|\,\widetilde{X}_{i-1},Y,Z}(x^{\mathsf{sc}}_i\,|\,x_{i-1},y,z)p_{X^{\mathsf{c}}_{t_{i-1}}\,|\,X^{\mathsf{s}}_{t_{i-1}}}(x^{\mathsf{c}}_{i-1}\,|\,x^{\mathsf{s}}_{i-1})\mathrm{d}x^{\mathsf{c}}_{i-1}}{\int p_{X^{\mathsf{sc}}_{t_i}\,|\,X_{t_{i-1}},Y,Z}(x^{\mathsf{sc}}_i\,|\,x_{i-1})p_{X^{\mathsf{c}}_{t_{i-1}}\,|\,X^{\mathsf{s}}_{t_{i-1}}}(x^{\mathsf{c}}_{i-1}\,|\,x^{\mathsf{s}}_{i-1})\mathrm{d}x^{\mathsf{c}}_{i-1}}p_{\widetilde{X}^{\mathsf{sc}}_i\,|\,\widetilde{X}^{\mathsf{s}}_{i-1},Y,Z}(x^{\mathsf{sc}}_i\,|\,x^{\mathsf{s}}_{i-1},y,z)\mathrm{d}x^{\mathsf{sc}}_i\\
&\quad\to\mathbb{E}_{x^{\mathsf{c}}_{i-1}\sim X^{\mathsf{c}}_{t_{i-1}}\,|\,X^{\mathsf{s}}_{t_{i-1}}}\big[\mathsf{KL}(p_{\widetilde{X}^{\mathsf{sc}}_i\,|\,\widetilde{X}_{i-1},Y,Z}(\cdot\,|\,x_{i-1},y,z)\,\|\,p_{X^{\mathsf{sc}}_{t_i}\,|\,X_{t_{i-1}}}(\cdot\,|\,x_{i-1}))\big]. \qquad (42)
\end{aligned}
$$

Combining (36), (38) and (42), we have

$$
\begin{aligned}
&\lim\mathsf{TV}\Big(p_{\widetilde{X}^{\mathsf{sc}}_k\,|\,Y,Z},p_{\overline{X^{\mathsf{sc}}_k}\,|\,Y,Z}\Big)^2\\
&\le\lim\sum_{i=1}^k\mathbb{E}_{x^{\mathsf{s}}_{i-1}\sim\widetilde{X}^{\mathsf{s}}_{i-1}\,|\,Y,Z}\Big[\mathbb{E}_{x^{\mathsf{c}}_{i-1}\sim X^{\mathsf{c}}_{t_{i-1}}\,|\,X^{\mathsf{s}}_{t_{i-1}}}\big[\mathsf{KL}(p_{\widetilde{X}^{\mathsf{sc}}_i\,|\,\widetilde{X}_{i-1},Y,Z}(\cdot\,|\,x_{i-1},y,z)\,\|\,p_{X^{\mathsf{sc}}_{t_i}\,|\,X_{t_{i-1}}}(\cdot\,|\,x_{i-1}))\big]\Big]\\
&\quad+\lim\sum_{i=1}^k\mathbb{E}_{x^{\mathsf{s}}_{i-1}\sim\widetilde{X}^{\mathsf{s}}_{i-1}\,|\,Y,Z}\Big[\int\log\frac{\int p_{\widetilde{X}^{\mathsf{sc}}_i\,|\,\widetilde{X}_{i-1},Y,Z}(x^{\mathsf{sc}}_i\,|\,x_{i-1},y,z)p_{\widetilde{X}^{\mathsf{c}}_{i-1}\,|\,\widetilde{X}^{\mathsf{s}}_{i-1},Y,Z}(x^{\mathsf{c}}_{i-1}\,|\,x^{\mathsf{s}}_{i-1},y,z)\mathrm{d}x^{\mathsf{c}}_{i-1}}{\int p_{\widetilde{X}^{\mathsf{sc}}_i\,|\,\widetilde{X}_{i-1},Y,Z}(x^{\mathsf{sc}}_i\,|\,x_{i-1},y,z)p_{X^{\mathsf{c}}_{t_{i-1}}\,|\,X^{\mathsf{s}}_{t_{i-1}},Y,Z}(x^{\mathsf{c}}_{i-1}\,|\,x^{\mathsf{s}}_{i-1},y,z)\mathrm{d}x^{\mathsf{c}}_{i-1}}\\
&\qquad\qquad\times p_{\widetilde{X}^{\mathsf{sc}}_i\,|\,\widetilde{X}^{\mathsf{s}}_{i-1},Y,Z}(x^{\mathsf{sc}}_i\,|\,x^{\mathsf{s}}_{i-1},y,z)\mathrm{d}x^{\mathsf{sc}}_i\Big]. \qquad (43)
\end{aligned}
$$

Step 4. control error terms. Particularly, with $\eta_0 = 1$ which corresponds to DDPM, we have the following result, whose proof is postponed to Appendix C.5.

Lemma 3. *Suppose that $\eta_0 = 1$ and δ small enough. For any x,*

$$
\mathsf{KL}\Big(p_{V^\top_{S^{\mathsf{prior}}_{t_i}}\widehat{X}_{t_i}\,|\,\widehat{X}_{t_{i-1}}=x}\,\|\,p_{V^\top_{S^{\mathsf{prior}}_{t_i}}X_{t_i}\,|\,X_{t_{i-1}}=x}\Big) = o(\delta_i). \qquad (44)
$$

By using Lemma 3, and notice that $\widetilde{X}^{\mathsf{sc}}_i\,|\,\widetilde{X}_{i-1},Y,Z$ shares the same transition probability as $\widehat{X}^{\mathsf{sc}}_{t_i}\,|\,\widehat{X}_{t_{i-1}}$, we have

$$
\sum_{i=1}^k\mathbb{E}_{x^{\mathsf{s}}_{i-1}\sim\widetilde{X}^{\mathsf{s}}_{i-1}\,|\,Y,Z}\Big[\mathbb{E}_{x^{\mathsf{c}}_{i-1}\sim X^{\mathsf{c}}_{t_{i-1}}\,|\,X^{\mathsf{s}}_{t_{i-1}}}\big[\mathsf{KL}(p_{\widetilde{X}^{\mathsf{sc}}_i\,|\,\widetilde{X}_{i-1},Y,Z}(\cdot\,|\,x_{i-1},y,z)\,\|\,p_{X^{\mathsf{sc}}_{t_i}\,|\,X_{t_{i-1}}}(\cdot\,|\,x_{i-1}))\big]\Big]
$$

$$= \sum_{i=1}^{k} \mathbb{E}_{x_{i-1}^{\mathsf{s}} \sim \widetilde{X}_{i-1}^{\mathsf{s}} \mid Y,Z;\, x_{i-1}^{\mathsf{c}} \sim X_{t_{i-1}}^{\mathsf{c}} \mid X_{t_{i-1}}^{\mathsf{s}} = x_{i-1}^{\mathsf{s}}} \left[\mathsf{KL}(p_{\widehat{X}_{t_i}^{\mathsf{sc}} \mid \widehat{X}_{t_{i-1}}}(\cdot \mid x_{i-1}) \,\|\, p_{X_{t_i}^{\mathsf{sc}} \mid X_{t_{i-1}}}(\cdot \mid x_{i-1})) \right]$$

$$= \sum_{i=1}^{k} o(\delta_i) \to 0, \text{ as } \delta \to 0. \tag{45}$$

The second term describes the influence of the sampling error of x_{i-1}^{c} on the distribution of x_i^{sc}. It is given by the following lemma, whose proof is postponed to Section C.6.

Lemma 4. *If $p_{\widetilde{X}_{i-1}^{\mathsf{c}} \mid \widetilde{X}_{i-1}^{\mathsf{s}}}(x_{i-1}^{\mathsf{c}} \mid x_{i-1}^{\mathsf{s}}) \to p_{X_{t_{i-1}}^{\mathsf{c}} \mid X_{t_{i-1}}^{\mathsf{s}}}(x_{i-1}^{\mathsf{c}} \mid x_{i-1}^{\mathsf{s}})$ in distribution, then we have*

$$\sum_{i=1}^{k} \mathbb{E}_{x_{i-1}^{\mathsf{s}} \sim \widetilde{X}_{i-1}^{\mathsf{s}} \mid Y,Z} \Bigg[\int \log \frac{\int p_{\widetilde{X}_i^{\mathsf{sc}} \mid \widetilde{X}_{i-1}}(x_i^{\mathsf{sc}} \mid x_{i-1}) p_{\widetilde{X}_{i-1}^{\mathsf{c}} \mid \widetilde{X}_{i-1}^{\mathsf{s}}, Y, Z}(x_{i-1}^{\mathsf{c}} \mid x_{i-1}^{\mathsf{s}}, y, z) \mathrm{d}x_{i-1}^{\mathsf{c}}}{\int p_{\widetilde{X}_i^{\mathsf{sc}} \mid \widetilde{X}_{i-1}}(x_i^{\mathsf{sc}} \mid x_{i-1}) p_{X_{t_{i-1}}^{\mathsf{c}} \mid X_{t_{i-1}}^{\mathsf{s}}}(x_{i-1}^{\mathsf{c}} \mid x_{i-1}^{\mathsf{s}}) \mathrm{d}x_{i-1}^{\mathsf{c}}}$$

$$\times\, p_{\widetilde{X}_i^{\mathsf{sc}} \mid \widetilde{X}_{i-1}^{\mathsf{s}}, Y, Z}(x_i^{\mathsf{sc}} \mid x_{i-1}^{\mathsf{s}}, y, z) \mathrm{d}x_i^{\mathsf{sc}} \Bigg] \to 0. \tag{46}$$

Combining (45) and Lemma 4 together with (43), we have

$$\mathsf{TV}\big(p_{\widetilde{X}_k^{\mathsf{sc}} \mid Y,Z}, p_{\overline{X_k^{\mathsf{sc}}} \mid Y,Z}\big) \to 0. \tag{47}$$

Thus we complete the proof.

C Proof of auxiliary lemmas and facts

C.1 Proof of (22)

We first present the conclusion as the following lemma, and then prove it.

Lemma 5. *The update rule (22) is equivalent to applying DDIM with $\eta = 0$ by taking (21) as forward process. That is,*

$$\frac{v_s^\top \widehat{X}_{t_{i+1}}}{\widehat{\sigma}_{t_{i+1}}} = \frac{v_s^\top \widehat{X}_{t_i}}{\widehat{\sigma}_{t_i}} + \mathrm{e}^{\widehat{\lambda}_{t_i}} (1 - \mathrm{e}^{-\widehat{\delta}_{i+1}}) v_s^\top \mu_{t_i}(\widehat{X}_{t_i}), \tag{48}$$

where $\widehat{\alpha}_{t_i} = \alpha_{t_i}$, $\widehat{\sigma}_{t_i}^2 = \sigma_{t_i}^2 - \alpha_{t_i}^2 \sigma^2 \Sigma_{ss}^{-1}$, $\widehat{\lambda}_{t_i} = \log(\widehat{\alpha}_{t_i} / \widehat{\sigma}_{t_i})$ and $\widehat{\delta}_{i+1} = \widehat{\lambda}_{t_{i+1}} - \widehat{\lambda}_{t_i}$, and $v_s^\top \mu_{t_i}(\widehat{X}_{t_i}) = \Sigma_s^{-1} u_s^\top Y$.

Proof. By simplification, the update rule (48) is equivalent to

$$v_s^\top \widehat{X}_{t_{i+1}} = \frac{\widehat{\sigma}_{t_{i+1}}}{\widehat{\sigma}_{t_i}} v_s^\top \widehat{X}_{t_i} + \left(\widehat{\alpha}_{t_{i+1}} - \frac{\widehat{\sigma}_{t_{i+1}} \widehat{\alpha}_{t_i}}{\widehat{\sigma}_{t_i}} \right) v_s^\top \mu_{t_i}(\widehat{X}_{t_i}). \tag{49}$$

Assume that $v_s^\top \widehat{X}_{t_i} = \alpha_{t_i} \Sigma_s^{-1} U^\top Y + (\sigma_{t_i}^2 - \alpha_{t_i}^2 \sigma^2 \Sigma_{\mathcal{S}_{t_i}}^{-2})^{1/2} Z_s$. Instituting into (49) leads to

$$\begin{aligned} v_s^\top \widehat{X}_{t_{i+1}} &= \frac{\widehat{\sigma}_{t_{i+1}}}{\widehat{\sigma}_{t_i}} \left(\alpha_{t_i} \Sigma_s^{-1} U^\top Y + (\sigma_{t_i}^2 - \alpha_{t_i}^2 \sigma^2 \Sigma_{\mathcal{S}_{t_i}}^{-2})^{1/2} Z_s \right) + \left(\widehat{\alpha}_{t_{i+1}} - \frac{\widehat{\sigma}_{t_{i+1}} \widehat{\alpha}_{t_i}}{\widehat{\sigma}_{t_i}} \right) \Sigma_s^{-1} u_s^\top Y \\ &= \widehat{\alpha}_{t_{i+1}} \Sigma_s^{-1} u_s^\top Y + \frac{\widehat{\sigma}_{t_{i+1}}}{\widehat{\sigma}_{t_i}} (\sigma_{t_i}^2 - \alpha_{t_i}^2 \sigma^2 \Sigma_s^{-2})^{\frac{1}{2}} Z_s \\ &= \alpha_{t_{i+1}} \Sigma_s^{-1} u_s^\top Y + (\sigma_{t_{i+1}}^2 - \alpha_{t_{i+1}}^2 \sigma^2 \Sigma_s^{-2})^{\frac{1}{2}} Z_s, \end{aligned}$$

where the last equation arises from the definition of $\widehat{\sigma}_{t_i}$ and $\widehat{\alpha}_{t_i}$. □

C.2 Preliminaries

Throughout the proof, we redefine η_1 as η for ease of notations. We begin with approximating the update rule (23) for $\widehat{X}_{t_i}^{\mathsf{c}}$. We have

$$\widehat{X}_{t_{i+1}}^{\mathsf{c}} = \frac{\sigma_{t_{i+1}}}{\sigma_{t_i}} \sqrt{1 - \eta(1 - e^{-2\delta_{i+1}})} \widehat{X}_{t_i}^{\mathsf{c}} + \alpha_{t_{i+1}} \big(1 - e^{-\delta_{i+1}} \sqrt{1 - \eta(1 - e^{-2\delta_{i+1}})}\big) \mu_{t_i}^{\mathsf{c}}(\widehat{X}_{t_i})$$

$$+ \sigma_{t_{i+1}} \sqrt{\eta(1-e^{-2\delta_{i+1}})} Z^{\mathsf{c}}_{t_i}. \tag{50}$$

For the first term, we have

$$\begin{aligned}\frac{\sigma_{t_{i+1}}}{\sigma_{t_i}}\sqrt{1-\eta(1-e^{-2\delta_{i+1}})}\widehat{X}^{\mathsf{c}}_{t_i} &= \sqrt{1-\eta(1-e^{-2\delta_{i+1}})}\widehat{X}^{\mathsf{c}}_{t_i} + \left(\frac{\sigma_{t_{i+1}}}{\sigma_{t_i}} - 1\right)\sqrt{1-\eta(1-e^{-2\delta_{i+1}})}\widehat{X}^{\mathsf{c}}_{t_i} \\ &=: \sqrt{1-\eta(1-e^{-2\delta_{i+1}})}\widehat{X}^{\mathsf{c}}_{t_i} + e_{t_i,1}(\widehat{X}^{\mathsf{c}}_{t_i}).\end{aligned} \tag{51}$$

For the second term, we have

$$\begin{aligned}&\alpha_{t_{i+1}}\big(1 - e^{-\delta_{i+1}}\sqrt{1-\eta(1-e^{-2\delta_{i+1}})}\big)\mu^{\mathsf{c}}_{t_i}(\widehat{X}_{t_i}) \\ &= \alpha_{t_i}\big(1 - \sqrt{1-\eta(1-e^{-2\delta_{i+1}})}\big)\mu^{\mathsf{c}}_{t_i}(\widehat{X}_{t_i}) \\ &\quad + \underbrace{\Big(\big(\alpha_{t_{i+1}} - \alpha_{t_i}\big) + \big(\alpha_{t_{i+1}}e^{-\delta_{i+1}} - \alpha_{t_i}\big)\sqrt{1-\eta(1-e^{-2\delta_{i+1}})}\Big)\mu^{\mathsf{c}}_{t_i}(\widehat{X}_{t_i})}_{e_{t_i,2}(\widehat{X}_{t_i})} \\ &= \big(1 - e^{-\delta_{i+1}}\sqrt{1-\eta(1-e^{-2\delta_{i+1}})}\big)\Big(\widehat{X}^{\mathsf{c}}_{t_i} - \sigma_{t_i}\epsilon^{\mathsf{c}}_{t_i}(\widehat{X}_{t_i})\Big) + e_{t_i,2}(\widehat{X}_{t_i}) \\ &= \big(1 - \sqrt{1-\eta(1-e^{-2\delta_{i+1}})}\big)\Big(\widehat{X}^{\mathsf{c}}_{t_i} - \sigma_{t_{i+1}}\epsilon^{\mathsf{c}}_{t_{i+1}}(\widehat{X}^{\mathsf{est}}_{t_{i+1}})\Big) + e_{t_i,2}(\widehat{X}_{t_i}) \\ &\quad + \underbrace{\big(1 - \sqrt{1-\eta(1-e^{-2\delta_{i+1}})}\big)\big(\sigma_{t_i}\epsilon^{\mathsf{c}}_{t_i}(\widehat{X}_{t_i}) - \sigma_{t_{i+1}}\epsilon^{\mathsf{c}}_{t_{i+1}}(\widehat{X}^{\mathsf{est}}_{t_{i+1}})\big)}_{e_{t_i,3}(\widehat{X}_{t_i},\widehat{X}^{\mathsf{s}}_{t_{i+1}})},\end{aligned} \tag{52}$$

where $\widehat{X}^{\mathsf{est}}_{t_{i+1}}$ is defined by $V_{\mathcal{S}^{\mathsf{c}}}^\top \widehat{X}^{\mathsf{est}}_{t_{i+1}} = \widehat{X}^{\mathsf{c}}_{t_i}$ and $V_{\mathcal{S}}^\top \widehat{X}^{\mathsf{est}}_{t_{i+1}} = \widehat{X}^{\mathsf{s}}_{t_{i+1}}$. Note that

$$\epsilon^{\mathsf{c}}_{t_i}(x) \coloneqq \frac{1}{\sigma_{t_i}}\mathbb{E}[x^{\mathsf{c}} - \alpha_{t_i}X^{\mathsf{c}}_0 \,|\, X_{t_i} = x] = -\sigma_{t_i}\nabla_{x^{\mathsf{c}}}\log p_{X^{\mathsf{c}}_{t_i}|X^{\mathsf{s}}_{t_i}}(x^{\mathsf{c}}|x^{\mathsf{s}}), \tag{53}$$

where $X_{t_i} \overset{\mathrm{d}}{=} \alpha_{t_i}X_0 + \sigma_{t_i}\mathcal{N}(0, I_d)$, and the second equation holds because

$$\nabla_{x^{\mathsf{c}}}\log p_{X^{\mathsf{c}}_{t_i}|X^{\mathsf{s}}_{t_i}}(x^{\mathsf{c}}|x^{\mathsf{s}}) = \nabla_{x^{\mathsf{c}}}\log p_{X_{t_i}}(x) = -\frac{1}{\sigma^2_{t_i}}\mathbb{E}[x^{\mathsf{c}} - \alpha_{t_i}X^{\mathsf{c}}_0|X_{t_i} = x]. \tag{54}$$

Combining (51) and (52) together with (50), we have

$$\begin{aligned}\widehat{X}^{\mathsf{c}}_{t_{i+1}} &= \widehat{X}^{\mathsf{c}}_{t_i} - \sigma_{t_{i+1}}\big(1 - \sqrt{1-\eta(1-e^{-2\delta_{i+1}})}\big)\epsilon^{\mathsf{c}}_{t_{i+1}}(\widehat{X}^{\mathsf{est}}_{t_{i+1}}) + \sigma_{t_{i+1}}\sqrt{\eta(1-e^{-2\delta_{i+1}})}Z^{\mathsf{c}}_{t_i} \\ &\quad + e_{t_i,1}(\widehat{X}^{\mathsf{c}}_{t_i}) + e_{t_i,2}(\widehat{X}_{t_i}) + e_{t_i,3}(\widehat{X}_{t_i}, \widehat{X}^{\mathsf{s}}_{t_{i+1}}) \\ &=: \widehat{X}^{\mathsf{c}}_{t_i} - \frac{\eta(1-e^{-2\delta_{i+1}})}{2}\sigma_{t_{i+1}}\epsilon^{\mathsf{c}}_{t_{i+1}}(\widehat{X}^{\mathsf{est}}_{t_{i+1}}) + \sigma_{t_{i+1}}\sqrt{\eta(1-e^{-2\delta_{i+1}})}Z^{\mathsf{c}}_{t_i} + e_{t_i}(\widehat{X}_{t_i}, \widehat{X}^{\mathsf{s}}_{t_{i+1}})\end{aligned} \tag{55}$$

where $e_{t_i}(\widehat{X}_{t_i}, \widehat{X}^{\mathsf{s}}_{t_{i+1}})$ is defined as

$$\begin{aligned}e_{t_i}(\widehat{X}_{t_i}, \widehat{X}^{\mathsf{s}}_{t_{i+1}}) &\coloneqq e_{t_i,1}(\widehat{X}^{\mathsf{c}}_{t_i}) + e_{t_i,2}(\widehat{X}_{t_i}) + e_{t_i,3}(\widehat{X}_{t_i}, \widehat{X}^{\mathsf{s}}_{t_{i+1}}) \\ &\quad + \left(1 - \sqrt{1-\eta(1-e^{-2\delta_{i+1}})} - \frac{\eta(1-e^{-2\delta_{i+1}})}{2}\right)\sigma_{t_{i+1}}\epsilon^{\mathsf{c}}_{t_{i+1}}(\widehat{X}^{\mathsf{est}}_{t_{i+1}}).\end{aligned} \tag{56}$$

We have the following lemma, whose proof is postponed to Appendix C.7.

Lemma 6. *In (55), we have*

$$\mathbb{E}_{\widehat{X}^{\mathsf{sc}}_{t_{i+1}}|\widehat{X}_{t_i},Y,Z}[\|e_{t_i}(\widehat{X}_{t_i}, \widehat{X}^{\mathsf{sc}}_{t_{i+1}})\|_2^2] \lesssim \left(\|\widehat{X}^{\mathsf{c}}_{t_i}\|_2 + \|\widehat{X}^{\mathsf{sc}}_{t_i}\|_2 + \|Z\|_2 + 1\right)^2(\delta^2_{i+1} + \gamma^4_i) + \delta_{i+1}\gamma^2_i.$$

Next, we rewrite the update rule (23) for $\widehat{X}^{\mathsf{sc}}_{t_i}$. When $\eta_0 = 1$, it is equivalent to

$$\begin{aligned}
\widehat{X}^{\mathsf{sc}}_{t_{i+1}} &= \frac{\sigma_{t_{i+1}}}{\sigma_{t_i}} e^{-\delta_{i+1}} \widehat{X}^{\mathsf{sc}}_{t_i} + \sigma_{t_{i+1}} e^{\lambda_{t_{i+1}}} \big(1 - e^{-2\delta_{i+1}}\big) \mu^{\mathsf{sc}}_{t_i}(\widehat{X}_{t_i}) + \sigma_{t_{i+1}} \sqrt{1 - e^{-2\delta_{i+1}}}\, Z^{\mathsf{sc}}_{t_i} \\
&= \frac{\sigma_{t_{i+1}}}{\sigma_{t_i}} e^{-\delta_{i+1}} \widehat{X}^{\mathsf{sc}}_{t_i} + \frac{\alpha_{t_{i+1}}}{\alpha_{t_i}} \big(1 - e^{-2\delta_{i+1}}\big) (\widehat{X}^{\mathsf{sc}}_{t_i} - \sigma_{t_i} \epsilon^{\mathsf{sc}}_{t_i}) + \sigma_{t_{i+1}} \sqrt{1 - e^{-2\delta_{i+1}}}\, Z^{\mathsf{sc}}_{t_i} \\
&= \left(\frac{\sigma_{t_{i+1}}}{\sigma_{t_i}} e^{-\delta_{i+1}} + \frac{\alpha_{t_{i+1}}}{\alpha_{t_i}} - \frac{\alpha_{t_{i+1}}}{\alpha_{t_i}} e^{-2\delta_{i+1}} \right) \widehat{X}^{\mathsf{sc}}_{t_i} - \frac{\alpha_{t_{i+1}}}{\alpha_{t_i}} \big(1 - e^{-2\delta_{i+1}}\big) \sigma_{t_i} \epsilon^{\mathsf{sc}}_{t_i} + \sigma_{t_{i+1}} \sqrt{1 - e^{-2\delta_{i+1}}}\, Z^{\mathsf{sc}}_{t_i} \\
&= \frac{\alpha_{t_{i+1}}}{\alpha_{t_i}} \widehat{X}^{\mathsf{sc}}_{t_i} - \alpha_{t_{i+1}} \left(\mathrm{e}^{-\lambda_{t_i}} - \mathrm{e}^{-2\lambda_{t_{i+1}} + \lambda_{t_i}} \right) \epsilon^{\mathsf{sc}}_{t_i} + \sigma_{t_{i+1}} \sqrt{1 - e^{-2\delta_{i+1}}}\, Z^{\mathsf{sc}}_{t_i} \\
&= \frac{\alpha_{t_{i+1}}}{\alpha_{t_i}} \widehat{X}^{\mathsf{sc}}_{t_i} - \frac{\alpha_{t_{i+1}} \sigma_{t_i}}{\alpha_{t_i}} \left(1 - \mathrm{e}^{-2\delta_{i+1}} \right) \epsilon^{\mathsf{sc}}_{t_i} + \sigma_{t_{i+1}} \sqrt{1 - e^{-2\delta_{i+1}}}\, Z^{\mathsf{sc}}_{t_i}. \qquad (57)
\end{aligned}$$

C.3 Proof of Lemma 1

Combining (22), (55) and (57), the update rule of $\widehat{X}_{t_k}$ can be rewritten as

$$\begin{aligned}
\widehat{X}^{\mathsf{st}}_{t_{k+1}} &= \alpha_{t_{k+1}} \overline{Y}_{t_{k+1}} + \widehat{\sigma}_{t_{k+1}} \overline{Z}_{t_{k+1}}, \\
\widehat{X}^{\mathsf{sc}}_{t_{k+1}} &= \frac{\alpha_{t_{k+1}}}{\alpha_{t_k}} \widehat{X}^{\mathsf{sc}}_{t_k} + \alpha_{t_{k+1}} \left(\mathrm{e}^{-\lambda_{t_k}} - \mathrm{e}^{-2\lambda_{t_{k+1}} + \lambda_{t_k}} \right) \frac{\nabla_{\widehat{X}^{\mathsf{sc}}_{t_k}} \log p_{\lambda_k}(\widehat{X}^{\mathsf{c}}_{t_k}, \widehat{X}^{\mathsf{sc}}_{t_k})}{\sigma_{t_k}} + \sigma_{t_{k+1}} \sqrt{1 - e^{-2\delta_{k+1}}}\, Z^{\mathsf{sc}}_{t_k}, \\
\widehat{X}^{\mathsf{c}}_{t_{k+1}} &= \widehat{X}^{\mathsf{c}}_{t_k} + \gamma_k \nabla \log p_{\lambda_{k+1}}(\widehat{X}^{\mathsf{c}}_{t_k}, \widehat{X}^{\mathsf{sc}}_{t_{k+1}}) + \sigma_{t_{k+1}} \sqrt{\eta(1 - e^{-2\delta_{k+1}})}\, Z^{\mathsf{c}}_{t_k} + e_{t_k}(\widehat{X}_{t_k}, \widehat{X}^{\mathsf{sc}}_{t_{k+1}}).
\end{aligned} \qquad (58)$$

The KL divergence between $\widehat{X}_{t_{0:M}}$ and $\widetilde{X}_{0:M}$ conditoned on Y and Z can be decomposed as

$$\begin{aligned}
&\mathsf{KL}(p_{\widetilde{X}_{0:M} \mid Y,Z} \| p_{\widehat{X}_{t_{0:M}} \mid Y,Z}) \\
&\quad = \mathsf{KL}(p_{\widetilde{X}_0 \mid Y,Z} \| p_{\widehat{X}_{t_0} \mid Y,Z}) + \sum_{i=0}^{M-1} \mathbb{E}_{x \sim p_{\widetilde{X}_{t_i} \mid Y,Z}} \mathsf{KL}\big(p_{\widetilde{X}^{\mathsf{sc}}_{i+1} | \widetilde{X}_i}(\cdot|x) \| p_{\widehat{X}^{\mathsf{sc}}_{t_{i+1}} | \widehat{X}_{t_i}}(\cdot|x)\big) \\
&\qquad + \sum_{i=0}^{M-1} \mathbb{E}_{x,x' \sim p_{\widetilde{X}^{\mathsf{sc}}_{i+1}, \widetilde{X}_i \mid Y,Z}} \mathsf{KL}\big(p_{\widetilde{X}^{\mathsf{c}}_{i+1} | \widetilde{X}^{\mathsf{sc}}_{i+1}, \widetilde{X}_i}(\cdot|x,x') \| p_{\widehat{X}^{\mathsf{c}}_{t_{i+1}} | \widehat{X}^{\mathsf{sc}}_{t_{i+1}}, \widehat{X}_{t_i}}(\cdot|x,x')\big) \\
&\quad = \sum_{i=0}^{M-1} \mathbb{E}_{x,x' \sim p_{\widetilde{X}^{\mathsf{sc}}_{i+1}, \widetilde{X}_i \mid Y,Z}} \mathsf{KL}\big(p_{\widetilde{X}^{\mathsf{c}}_{i+1} | \widetilde{X}^{\mathsf{sc}}_{i+1}, \widetilde{X}_i}(\cdot|x,x') \| p_{\widehat{X}^{\mathsf{c}}_{t_{i+1}} | \widehat{X}^{\mathsf{sc}}_{t_{i+1}}, \widehat{X}_{t_i}}(\cdot|x,x')\big).
\end{aligned}$$

Note that the both of the above conditional distributions are Gaussian ditributions with variance $2\gamma_i$ and mean vector $(x')^{\mathsf{c}} + \gamma_i \nabla \log p_{\lambda_{i+1}}((x')^{\mathsf{c}}, x^{\mathsf{sc}}) + e_{t_i}(x', x^{\mathsf{sc}})$ and $(x')^{\mathsf{c}} + \gamma_i \nabla \log p_{\lambda_{i+1}}((x')^{\mathsf{c}}, x^{\mathsf{sc}})$, respectively. Combining with Lemma 6, we finally get

$$\begin{aligned}
&\mathbb{E}_{x,x' \sim p_{\widetilde{X}^{\mathsf{sc}}_{i+1}, \widetilde{X}_i \mid Y,Z}} \big[\mathsf{KL}\big(p_{\widetilde{X}^{\mathsf{c}}_{i+1} | \widetilde{X}^{\mathsf{sc}}_{i+1}, \widetilde{X}_i}(\cdot|x,x') \| p_{\widehat{X}^{\mathsf{c}}_{t_{i+1}} | \widehat{X}^{\mathsf{sc}}_{t_{i+1}}, \widehat{X}_{t_i}}(\cdot|x,x')\big) \big] \\
&\quad = \frac{\mathbb{E}_{x,x' \sim p_{\widetilde{X}^{\mathsf{sc}}_{i+1}, \widetilde{X}_i \mid Y,Z}} \big[\| e_{t_i}(x', x) \|_2^2 \big]}{4\gamma_i} \\
&\quad \lesssim \frac{1}{\gamma_i} \left(\mathbb{E}\left[\left(\|\widetilde{X}^{\mathsf{c}}_i\|_2 + \|\widetilde{X}^{\mathsf{sc}}_i\|_2 + \|Z\|_2 + 1 \right)^2 \right] \right) (\delta_{i+1}^2 + \gamma_i^4) + \delta_{i+1} \gamma_i.
\end{aligned}$$

Summarizing from $i = 0$ to $M - 1$ yields

$$\begin{aligned}
&\sum_{i=0}^{M-1} \mathbb{E}_{x,x' \sim p_{\widetilde{X}^{\mathsf{sc}}_{i+1}, \widetilde{X}_i \mid Y,Z}} \big[\mathsf{KL}\big(p_{\widetilde{X}^{\mathsf{c}}_{i+1} | \widetilde{X}^{\mathsf{sc}}_{i+1}, \widetilde{X}_i}(\cdot|x,x') \| p_{\widehat{X}^{\mathsf{c}}_{t_{i+1}} | \widehat{X}^{\mathsf{sc}}_{t_{i+1}}, \widehat{X}_{t_i}}(\cdot|x,x')\big) \big] \\
&\quad \overset{\text{(i)}}{\lesssim} \max \frac{\delta_{i+1}}{\gamma_i} + \eta \max \gamma_i^2 + \max \gamma_i \overset{\text{(ii)}}{\lesssim} \frac{1}{\eta} + \eta \max \gamma_i^2 + \max \gamma_i \\
&\quad = O\big(\eta^{-1}\big) + O(\eta^3 \delta^2) + O(\eta\delta),
\end{aligned}$$

where (i) uses $\sum_{i=0}^{M-1}\delta_{i+1}=O(1)$, $\sum_{i=0}^{M-1}\gamma_i=O(\eta)$, and $\mathbb{E}[\|\widetilde{X}_i^{\mathsf{c}}\|_2^2+\|\widetilde{X}_i^{\mathsf{sc}}\|_2^2+\|Z\|_2^2]=O(1)$, and (ii) uses (27). As $\eta^3\delta^2\to 0$ and $\eta\to\infty$, which is satisfied by $\eta=\Theta(\delta^{-7/12})$, we have $\mathsf{KL}\big(p_{\widetilde{X}_{0:M}\,|\,Y,Z}\|p_{\widehat{X}_{t_{0:M}}\,|\,Y,Z}\big)\to 0$ and complete the proof.

Next, we intend to prove that $p_{\overline{X}_k\,|\,V_{\mathcal{S}_{t_k}^{\mathsf{meas}}}^\top\overline{X}_k,v_s^\top\overline{X}_{k_{\tau_s}}}=p_{X_{t_k}\,|\,V_{\mathcal{S}_{t_k}^{\mathsf{meas}}}^\top X_{t_k},v_s^\top X_{\tau_s}}$. When $\sigma_{t_k}>\alpha_{t_k}\Sigma_{ss}^{-1}\sigma$, we have $\mathcal{S}_{t_k}^{\mathsf{meas}}=\mathcal{S}$ and $\mathcal{S}_{t_k}^{\mathsf{prior}}$ is empty. Thus we have

$$p_{\overline{X}_k\,|\,V_{\mathcal{S}_{t_k}^{\mathsf{meas}}}^\top\overline{X}_k,v_s^\top\overline{X}_{k\tau_s}}=p_{\overline{X}_k\,|\,V_{\mathcal{S}_{t_k}^{\mathsf{meas}}}^\top\overline{X}_k}=p_{X_{t_k}\,|\,V_{\mathcal{S}_{t_k}^{\mathsf{meas}}}^\top X_{t_k}}=p_{X_{t_k}\,|\,V_{\mathcal{S}_{t_k}^{\mathsf{meas}}}^\top X_{t_k},v_s^\top X_{\tau_s}},$$

which makes Lemma 1 hold obviously. In particular, this leads to the following equation when $t_i=\tau_s$:

$$p_{\overline{X}_{k_{\tau_s}}\,|\,V_{\mathcal{S}}^\top\overline{X}_{k_{\tau_s}}}=p_{X_{\tau_s}\,|\,V_{\mathcal{S}}^\top X_{\tau_s}}.$$

When $\sigma_{t_k}\le\alpha_{t_k}\Sigma_{ss}^{-1}\sigma$, we have $\mathcal{S}_{t_k}^{\mathsf{prior}}=\mathcal{S}$ and $\mathcal{S}_{t_k}^{\mathsf{meas}}$ is empty. Thus we have

$$p_{\overline{X}_k\,|\,\overline{X}_{k-1}}=p_{\overline{X}_k^{\mathsf{c}}\,|\,\overline{X}_k^{\mathsf{s}}}p_{\overline{X}_k^{\mathsf{s}}\,|\,\overline{X}_{k-1}}=p_{X_{t_k}^{\mathsf{c}}\,|\,X_{t_k}^{\mathsf{s}}}p_{X_{t_k}^{\mathsf{s}}\,|\,X_{t_{k-1}}}=p_{X_{t_k}\,|\,X_{t_{k-1}}}.$$

Therefore we have

$$\begin{aligned}
&p_{\overline{X}_k\,|\,V_{\mathcal{S}_{t_k}^{\mathsf{meas}}}^\top\overline{X}_k,v_s^\top\overline{X}_{k_{\tau_s}}}(x_k\,|\,\cdot,x_{\tau_s}^{\mathsf{s}})\\
&\quad=p_{\overline{X}_k\,|\,V_{\mathcal{S}}^\top\overline{X}_{k_{\tau_s}}}(x_k\,|\,x_{\tau_s}^{\mathsf{s}})\\
&\quad=\int\prod_{j=k_{\tau_s}+1}^{k}p_{\overline{X}_j\,|\,\overline{X}_{j-1}}(x_j\,|\,x_{j-1})p_{\overline{X}_{k_{\tau_s}}\,|\,V_{\mathcal{S}}^\top\overline{X}_{k_{\tau_s}}}(x_{\tau_s}\,|\,x_{\tau_s}^{\mathsf{s}})\prod_{j=k_{\tau_s}+1}^{k-1}\mathrm{d}x_j\mathrm{d}x_{\tau_s}^{\mathsf{c}}\\
&\quad=\int\prod_{j=k_{\tau_s}+1}^{k}p_{X_{t_j}\,|\,X_{t_{j-1}}}(x_j\,|\,x_{j-1})p_{X_{\tau_s}\,|\,V_{\mathcal{S}}^\top X_{\tau_s}}(x_{\tau_s}\,|\,x_{\tau_s}^{\mathsf{s}})\prod_{j=k_{\tau_s}+1}^{k-1}\mathrm{d}x_j\mathrm{d}x_{\tau_s}^{\mathsf{c}}\\
&\quad=p_{X_{t_k}\,|\,V_{\mathcal{S}}^\top X_{\tau_s}}(x_k\,|\,x_{\tau_s}^{\mathsf{s}})=p_{X_{t_k}\,|\,V_{\mathcal{S}_{t_k}^{\mathsf{meas}}}^\top X_{t_k},v_s^\top X_{\tau_s}}(x_k\,|\,\cdot,x_{\tau_s}^{\mathsf{s}}).
\end{aligned}$$

C.4 Proof of Lemma 2

According to Lemma 1, it sufficies to prove that

$$\mathbb{E}_{V_{\mathcal{S}}^\top\widetilde{X}_k\,|\,Y,Z}\big[\mathsf{TV}(p_{V_{\mathcal{S}^{\mathsf{c}}}^\top\widetilde{X}_k|V_{\mathcal{S}}^\top\widetilde{X}_k,Y,Z},\|p_{V_{\mathcal{S}^{\mathsf{c}}}^\top X_{t_k}|V_{\mathcal{S}}^\top X_{t_k}})\big]\to 0\qquad\mathsf{as}\qquad\delta\eta\log\frac{1}{\delta}\to 0\quad\text{and}\quad\frac{\delta\eta^2}{\log\frac{1}{\delta}}\to\infty.\tag{59}$$

Then by using Lemma 1, we have

$$\begin{aligned}
&\mathbb{E}_{x_k^{\mathsf{s}}\sim\widehat{X}_{t_k}^{\mathsf{s}}}\big[\mathsf{TV}(p_{V_{\mathcal{S}^{\mathsf{c}}}^\top\widehat{X}_{t_k}|V_{\mathcal{S}}^\top\widehat{X}_{t_k}}(\cdot\,|\,x_k^{\mathsf{s}}),p_{V_{\mathcal{S}^{\mathsf{c}}}^\top X_{t_k}|V_{\mathcal{S}}^\top X_{t_k}}(\cdot\,|\,x_k^{\mathsf{s}}))\big]\\
&\quad=\mathbb{E}_{x_k^{\mathsf{s}}\sim\widehat{X}_{t_k}^{\mathsf{s}}}\big[\mathsf{TV}(p_{\widehat{X}_{t_k}^{\mathsf{c}}|\widehat{X}_{t_k}^{\mathsf{s}}}(\cdot\,|\,x_k^{\mathsf{s}}),p_{X_{t_k}^{\mathsf{c}}|X_{t_k}^{\mathsf{s}}}(\cdot\,|\,x_k^{\mathsf{s}}))\big]\\
&\quad=\mathbb{E}_{Y,Z}\Big[\mathbb{E}_{x_k^{\mathsf{s}}\sim\widehat{X}_{t_k}^{\mathsf{s}}\,|\,Y,Z}\big[\mathsf{TV}(p_{\widehat{X}_{t_k}^{\mathsf{c}}|\widehat{X}_{t_k}^{\mathsf{s}},Y,Z}(\cdot\,|\,x_k^{\mathsf{s}},y,z),p_{X_{t_k}^{\mathsf{c}}|X_{t_k}^{\mathsf{s}}}(\cdot\,|\,x_k^{\mathsf{s}}))\big]\Big]\\
&\quad\le\mathbb{E}_{Y,Z}\Big[\mathbb{E}_{x_k^{\mathsf{s}}\sim\widehat{X}_{t_k}^{\mathsf{s}}\,|\,Y,Z}\big[\mathsf{TV}(p_{\widehat{X}_{t_k}^{\mathsf{c}}|\widehat{X}_{t_k}^{\mathsf{s}},Y,Z}(\cdot\,|\,x_k^{\mathsf{s}},y,z),p_{\widetilde{X}_k^{\mathsf{c}}|\widetilde{X}_k^{\mathsf{s}},Y,Z}(\cdot\,|\,x_k^{\mathsf{s}},y,z))\big]\Big]\\
&\qquad+\mathbb{E}_{Y,Z}\Big[\mathbb{E}_{x_k^{\mathsf{s}}\sim\widehat{X}_{t_k}^{\mathsf{s}}\,|\,Y,Z}\big[\mathsf{TV}(p_{\widetilde{X}_k^{\mathsf{c}}|\widetilde{X}_k^{\mathsf{s}},Y,Z}(\cdot\,|\,x_k^{\mathsf{s}},y,z),p_{X_{t_k}^{\mathsf{c}}|X_{t_k}^{\mathsf{s}}}(\cdot\,|\,x_k^{\mathsf{s}}))\big]\Big]\\
&\quad\le\mathbb{E}_{Y,Z}\Big[\mathsf{TV}(p_{\widehat{X}_{t_k}\,|\,Y,Z},p_{\widetilde{X}_k\,|\,Y,Z})+\mathsf{TV}(p_{\widehat{X}_{t_k}^{\mathsf{s}}\,|\,Y,Z},p_{\widetilde{X}_k^{\mathsf{s}}\,|\,Y,Z})\Big]\\
&\qquad+\mathbb{E}_{Y,Z}\Big[\mathbb{E}_{x_k^{\mathsf{s}}\sim\widehat{X}_{t_k}^{\mathsf{s}}\,|\,Y,Z}\big[\mathsf{TV}(p_{\widetilde{X}_k^{\mathsf{c}}|\widetilde{X}_k^{\mathsf{s}},Y,Z}(\cdot\,|\,x_k^{\mathsf{s}},y,z),p_{X_{t_k}^{\mathsf{c}}|X_{t_k}^{\mathsf{s}}}(\cdot\,|\,x_k^{\mathsf{s}}))\big]\Big]\\
&\quad\le\mathbb{E}_{Y,Z}\Big[\mathsf{TV}(p_{\widehat{X}_{t_k}\,|\,Y,Z},p_{\widetilde{X}_k\,|\,Y,Z})+2\mathsf{TV}(p_{\widehat{X}_{t_k}^{\mathsf{s}}\,|\,Y,Z},p_{\widetilde{X}_k^{\mathsf{s}}\,|\,Y,Z})\Big]\\
&\qquad+\mathbb{E}_{Y,Z}\Big[\mathbb{E}_{x_k^{\mathsf{s}}\sim\widetilde{X}_k^{\mathsf{s}}\,|\,Y,Z}\big[\mathsf{TV}(p_{\widetilde{X}_k^{\mathsf{c}}|\widetilde{X}_k^{\mathsf{s}},Y,Z}(\cdot\,|\,x_k^{\mathsf{s}},y,z),p_{X_{t_k}^{\mathsf{c}}|X_{t_k}^{\mathsf{s}}}(\cdot\,|\,x_k^{\mathsf{s}}))\big]\Big]\to 0.
\end{aligned}\tag{60}$$

We introduce an auxiliary sequence as below: for $k-N \le i < k$, with initilization $\breve{X}_{k-N} \sim p_{\widetilde{X}_{k-N}}$, and iterates as

$$\breve{X}^{\mathsf{st}}_{i+1} = \widetilde{X}^{\mathsf{st}}_{i+1}, \tag{61a}$$

$$\breve{X}^{\mathsf{sc}}_{i+1} = \frac{\alpha_{t_{i+1}}}{\alpha_{t_i}} \breve{X}^{\mathsf{sc}}_i + \frac{\alpha_{t_{i+1}} \sigma^2_{t_k}}{\alpha_{t_i}} \left(1 - \mathrm{e}^{-2\delta_{i+1}}\right) \nabla_{\breve{X}^{\mathsf{sc}}_i} \log p_{\lambda_i}(\breve{X}^{\mathsf{c}}_{i-N}, \breve{X}^{\mathsf{sc}}_i) + \sigma_{t_{i+1}} \sqrt{1 - e^{-2\delta_{i+1}}} Z^{\mathsf{sc}}_i, \tag{61b}$$

$$\breve{X}^{\mathsf{c}}_{i+1} = \breve{X}^{\mathsf{c}}_i + \gamma_i \nabla_{\breve{X}^{\mathsf{c}}_i} \log p_{\lambda_{i+1}}(\breve{X}^{\mathsf{c}}_i, \breve{X}^{\mathsf{sc}}_{i+1}) + \sqrt{2\gamma_i} Z^{\mathsf{c}}_i. \tag{61c}$$

We choose N satisfying $\sum_{i=k-N}^{k} \delta_{i+1} = \Theta(\delta^{\frac{1}{2}})$. For any y, z, we have

$$\mathsf{TV}(p_{\breve{X}_k \,|\, Y,Z}, p_{\widetilde{X}_k \,|\, Y,Z}) \to 0. \tag{62}$$

The proof is postponed to the end of this subsection. Armed with this auxiliary sequence, we prove (59) via establishing

$$\mathbb{E}_{\widetilde{X}^{\mathsf{s}}_k \,|\, Y,Z}\left[\mathsf{TV}(p_{\breve{X}^{\mathsf{c}}_k | \breve{X}^{\mathsf{s}}_k, Y, Z}, \|p_{X^{\mathsf{c}}_{t_k} | X^{\mathsf{s}}_{t_k}})\right] \to 0 \qquad \mathsf{as} \quad \delta\eta \log \frac{1}{\delta} \to 0 \quad \text{and} \quad \frac{\delta\eta^2}{\log \frac{1}{\delta}} \to \infty. \tag{63}$$

which implies that

$$\begin{aligned}
&\mathbb{E}_{x^{\mathsf{sc}}_k \sim \widetilde{X}^{\mathsf{sc}}_k \,|\, Y,Z}[\mathsf{TV}(p_{\widetilde{X}^{\mathsf{c}}_k \,|\, \widetilde{X}^{\mathsf{sc}}_k, Y, Z}(\cdot \,|\, x^{\mathsf{sc}}_k, y, z), p_{X^{\mathsf{c}}_{t_k} \,|\, X^{\mathsf{sc}}_{t_k}}(\cdot \,|\, x^{\mathsf{sc}}_k))] \\
&\quad \le \mathbb{E}_{x^{\mathsf{sc}}_k \sim \widetilde{X}^{\mathsf{sc}}_k \,|\, Y,Z}[\mathsf{TV}(p_{\breve{X}^{\mathsf{c}}_k \,|\, \breve{X}^{\mathsf{sc}}_k, Y, Z}(\cdot \,|\, x^{\mathsf{sc}}_k, y, z), p_{X^{\mathsf{c}}_{t_k} \,|\, X^{\mathsf{sc}}_{t_k}}(\cdot \,|\, x^{\mathsf{sc}}_k))] \\
&\qquad + \mathbb{E}_{x^{\mathsf{sc}}_k \sim \widetilde{X}^{\mathsf{sc}}_k \,|\, Y,Z}[\mathsf{TV}(p_{\breve{X}^{\mathsf{c}}_k \,|\, \breve{X}^{\mathsf{sc}}_k, Y, Z}(\cdot \,|\, x^{\mathsf{sc}}_k, y, z), p_{\widetilde{X}^{\mathsf{c}}_k \,|\, \widetilde{X}^{\mathsf{sc}}_k, Y, Z}(\cdot \,|\, x^{\mathsf{sc}}_k, y, z))] \\
&\quad \le \mathbb{E}_{x^{\mathsf{sc}}_k \sim \widetilde{X}^{\mathsf{sc}}_k \,|\, Y,Z}[\mathsf{TV}(p_{\breve{X}^{\mathsf{c}}_k \,|\, \breve{X}^{\mathsf{sc}}_k, Y, Z}(\cdot \,|\, x^{\mathsf{sc}}_k, y, z), p_{X^{\mathsf{c}}_{t_k} \,|\, X^{\mathsf{sc}}_{t_k}}(\cdot \,|\, x^{\mathsf{sc}}_k))] \\
&\qquad + \mathsf{TV}(p_{\breve{X}_k \,|\, Y,Z}(\cdot \,|\, y, z), p_{\widetilde{X}_k \,|\, Y,Z}(\cdot \,|\, y, z)) + \mathsf{TV}(p_{\breve{X}^{\mathsf{sc}}_k \,|\, Y,Z}(\cdot \,|\, y, z), p_{\widetilde{X}^{\mathsf{sc}}_k \,|\, Y,Z}(\cdot \,|\, y, z)) \to 0.
\end{aligned}$$

To establish (63), we begin with considering given $\breve{X}^{\mathsf{sc}}_{1:k} = \breve{x}^{\mathsf{sc}}_{1:k}$, which satisfies

$$\mathcal{E}_k \coloneqq \left\{\|\breve{x}^{\mathsf{sc}}_{i+1} - \breve{x}^{\mathsf{sc}}_i\|_2 \lesssim \delta_{i+1} \log \frac{1}{\delta}, \forall i \le k-1\right\}. \tag{64}$$

For ease of notation, denote the distribution (pdf) of $\breve{X}^{\mathsf{c}}_k$ under given $\breve{X}^{\mathsf{s}}_{1:k} = \breve{x}^{\mathsf{sc}}_{1:k}$ as ψ_k. Denote the target conditional distribution

$$p^{\mathsf{c}}_k(x) \coloneqq p_{X^{\mathsf{c}}_{t_k} | X^{\mathsf{s}}_{t_k}}(x | \breve{x}^{\mathsf{s}}_k).$$

By deifnition of p_{λ_k}, we have $\nabla_x \log p_{\lambda_k}(x, \breve{x}^{\mathsf{sc}}_k) = \nabla_x \log p^{\mathsf{c}}_k(x)$. Moreover, denote P_{γ_k} as the transition probability of the following SDE from t to $t + \gamma_k$:

$$\mathrm{d}X(t) = \nabla \log p^{\mathsf{c}}_k(X(t))\mathrm{d}t + \sqrt{2}\mathrm{d}W_t, \tag{65}$$

where W_t denotes the Brownian motion. According to the analysis of Langevin, we have

$$\mathsf{KL}(P_{\gamma_k}\psi_k \| p^{\mathsf{c}}_{k+1}) \le \mathrm{e}^{-2\kappa\gamma_k} \mathsf{KL}(\psi_k \| p^{\mathsf{c}}_{k+1}),$$

where $\kappa < \infty$ is a constant. Therefore, we have

$$\begin{aligned}
\mathsf{KL}(\psi_{k+1} \| p^{\mathsf{c}}_{k+1}) &= \mathsf{KL}(P_{\gamma_k}\psi_k \| p^{\mathsf{c}}_{k+1}) + \mathsf{KL}(\psi_{k+1} \| p^{\mathsf{c}}_{k+1}) - \mathsf{KL}(P_{\gamma_k}\psi_k \| p^{\mathsf{c}}_{k+1}) \\
&\le \mathrm{e}^{-2\kappa\gamma_k} \mathsf{KL}(\psi_k \| p^{\mathsf{c}}_{k+1}) + \mathsf{KL}(\psi_{k+1} \| p^{\mathsf{c}}_{k+1}) - \mathsf{KL}(P_{\gamma_k}\psi_k \| p^{\mathsf{c}}_{k+1}) \\
&= \mathrm{e}^{-2\kappa\gamma_k} \mathsf{KL}(\psi_k \| p^{\mathsf{c}}_k) + \mathrm{e}^{-2\kappa\gamma_k} \left(\mathsf{KL}(\psi_k \| p^{\mathsf{c}}_{k+1}) - \mathsf{KL}(\psi_k \| p^{\mathsf{c}}_k)\right) \\
&\quad + \mathsf{KL}(\psi_{k+1} \| p^{\mathsf{c}}_{k+1}) - \mathsf{KL}(P_{\gamma_k}\psi_k \| p^{\mathsf{c}}_{k+1}).
\end{aligned} \tag{66}$$

Next, we control the two error terms in the right-hand-side of (66) separately.

- *Control the term* $\mathsf{KL}(\psi_k \| p^{\mathsf{c}}_{k+1}) - \mathsf{KL}(\psi_k \| p^{\mathsf{c}}_k)$. According to the definition of KL divergence, we have

$$\mathsf{KL}(\psi_k \| p^{\mathsf{c}}_{k+1}) - \mathsf{KL}(\psi_k \| p^{\mathsf{c}}_k) = \int \log \frac{p^{\mathsf{c}}_k(x)}{p^{\mathsf{c}}_{k+1}(x)} \psi_k(x) \mathrm{d}x. \tag{67}$$

According to the definition of $p^{\mathsf{c}}_{k+1}(x)$, we have

$$p^{\mathsf{c}}_k(x) = \frac{p_{X_{t_k}}(x)}{p_{X^{\mathsf{s}}_{t_k}}(\breve{x}^{\mathsf{s}}_k)}$$

Then we have

$$\begin{aligned}\frac{\partial \log p^{\mathsf{c}}_k(x)}{\partial \alpha_{t_i}} &= \frac{\partial \log p_{X_{t_k}}(x)}{\partial \alpha_{t_k}} - \frac{\partial \log p_{X^{\mathsf{s}}_{t_k}}(\breve{x}^{\mathsf{s}}_k)}{\partial \alpha_{t_k}} \\ &= \frac{1}{\sigma^2_{t_k}}\mathbb{E}[(x - \alpha_{t_k}X^{\mathsf{c}}_0)^\top X^{\mathsf{c}}_0 + (\breve{x}_k - \alpha_{t_k}X^{\mathsf{s}}_0)^\top X^{\mathsf{s}}_0 | X^{\mathsf{c}}_{t_k} = x, X^{\mathsf{s}}_{t_k} = \breve{x}^{\mathsf{s}}_k] \\ &\quad - \frac{1}{\sigma^2_{t_k}}\mathbb{E}[(\breve{x}^{\mathsf{s}}_k - \alpha_{t_k}X^{\mathsf{s}}_0)^\top X^{\mathsf{s}}_0 | X^{\mathsf{s}}_{t_k} = \breve{x}^{\mathsf{s}}_k].\end{aligned}$$

Recalling that X_0 is bounded, we further have

$$\left|\frac{\partial \log p^{\mathsf{c}}_k(x)}{\partial \alpha_{t_k}}\right| \lesssim (\|x\|_2 + \|\breve{x}^{\mathsf{s}}_k\|_2 + 1).$$

By similar calculation, we have

$$\left|\frac{\partial \log p^{\mathsf{c}}_k(x)}{\partial \sigma_{t_i}}\right| \lesssim (\|x\|^2_2 + \|\breve{x}^{\mathsf{s}}_k\|_2 + 1), \qquad \|\nabla_{x^{\mathsf{s}}} \log p^{\mathsf{c}}_k(x)\|_2 \lesssim (\|x\|_2 + \|\breve{x}^{\mathsf{s}}_k\|_2 + 1).$$

Thus we have

$$\begin{aligned}\log \frac{p^{\mathsf{c}}_{k+1}(x)}{p^{\mathsf{c}}_k} &= O(|\alpha_{t_{k+1}} - \alpha_{t_k}| \\ &\qquad + |\sigma_{t_{k+1}} - \sigma_{t_k}|)(\|x\|^2_2 + \|\breve{x}^{\mathsf{s}}_k\|_2 + 1) + O(\|\breve{x}^{\mathsf{s}}_{k+} - \breve{x}^{\mathsf{s}}_k\|_2)(\|x\|_2 + \|\breve{x}^{\mathsf{s}}_k\|_2 + 1) \\ &= (\|x\|^2_2 + \|\breve{x}^{\mathsf{s}}_k\|_2 + 1)O\left(\delta_{k+1} + \sqrt{\delta_{k+1}\log\frac{1}{\delta}}\right),\end{aligned}$$

where the last line arises from (92), (89), and (64). Based on this observation, we have

$$\begin{aligned}\int \log \frac{p^{\mathsf{c}}_k(x)}{p^{\mathsf{c}}_{k+1}(x)}\psi^{\mathsf{c}}_k(x)\mathrm{d}x &= -\int \log \frac{p^{\mathsf{c}}_{k+1}(x)}{p^{\mathsf{c}}_k(x)}\psi^{\mathsf{c}}_k(x)\mathrm{d}x \\ &= O(\sqrt{\delta_{k+1}\log\delta^{-1}})\big(\mathbb{E}[\|\breve{X}^{\mathsf{c}}_k\|^2_2 \,|\, \breve{X}^{\mathsf{s}}_k = \breve{x}^{\mathsf{sc}}_{1:k}] + \|\breve{x}^{\mathsf{sc}}_k\|_2 + 1\big) \\ &= O(\sqrt{\delta_{k+1}\log\delta^{-1}}),\end{aligned}$$

where the last line uses $\mathbb{E}[\|\breve{X}^{\mathsf{c}}_k\|^2_2 \,|\, \breve{X}^{\mathsf{s}}_k = \breve{x}^{\mathsf{sc}}_{1:k}] < \infty$. Combining these with (67) yields

$$\big|\mathsf{KL}(\psi_k\|p^{\mathsf{c}}_{k+1}) - \mathsf{KL}(\psi_k\|p^{\mathsf{c}}_{k+1})\big| \lesssim \sqrt{\delta_{k+1}\log\delta^{-1}}. \tag{68}$$

- *Control the term* $\mathsf{KL}(\psi_{k+1}\|p^{\mathsf{c}}_{k+1}) - \mathsf{KL}(P_{\gamma_k}\psi_k\|p^{\mathsf{c}}_{k+1})$. For any smooth function f, we have

$$\begin{aligned}\mathbb{E}_{\breve{X}^{\mathsf{c}}_{k+1}}[f(\breve{X}^{\mathsf{c}}_{k+1})] &= \mathbb{E}_{\breve{X}^{\mathsf{c}}_k, Z^{\mathsf{c}}_k}\big[f\big(\breve{X}^{\mathsf{c}}_k + \gamma_k\nabla\log p^{\mathsf{c}}_{k+1}(\breve{X}^{\mathsf{c}}_k) + \sqrt{2\gamma_k}Z^{\mathsf{c}}_k\big)\big] \\ &= \mathbb{E}\big[f\big(\breve{X}^{\mathsf{c}}_k\big) + \gamma_k\nabla f(\breve{X}^{\mathsf{c}}_k)^\top\nabla\log p^{\mathsf{c}}_{k+1}(\breve{X}^{\mathsf{c}}_k) + \gamma_k\Delta f(\breve{X}^{\mathsf{c}}_k) + O(\gamma^2_k)\big],\end{aligned} \tag{69}$$

where $\Delta f(x) \coloneqq \mathsf{Tr}(\nabla^2_x f(x))$. In contrast, by using the geneator operator of Langevin proccess $\mathcal{L} = \nabla\log p^\top_{k+1}\nabla + \Delta$, we have

$$\begin{aligned}\mathbb{E}_{X\sim P_{\gamma_k}\psi_k}[f(X)] &= \mathbb{E}_{X\sim\psi_k}[f(X) + \gamma_k(\mathcal{L}f)(X) + O(\gamma^2_k)] \\ &= \mathbb{E}_{X\sim\psi_k}[f(X) + \gamma_k\nabla\log p^\top_{k+1}(X)\nabla f(X) + \gamma_k\Delta f(X) + O(\gamma^2_k)]\end{aligned} \tag{70}$$

Combining (69) and (70), we have

$$\mathbb{E}_{X\sim\psi_{k+1}}[f(X)] - \mathbb{E}_{X\sim P_{\gamma_k}\psi_k}[f(X)] = O(\gamma^2_k). \tag{71}$$

Armed with this, we have

$$\begin{aligned}
&\mathsf{KL}(\psi_{k+1}\|p^{\mathsf{c}}_{k+1}) - \mathsf{KL}(P_{\gamma_k}\psi_k\|p^{\mathsf{c}}_{k+1}) \\
&\quad= \mathbb{E}_{X\sim\psi_{k+1}}\Big[\log\frac{\psi_{k+1}(X)}{p^{\mathsf{c}}_{k+1}(X)}\Big] - \mathbb{E}_{X\sim P_{\gamma_k}\psi_k}\Big[\log\frac{P_{\gamma_k}\psi_k(X)}{p^{\mathsf{c}}_{k+1}(X)}\Big] \\
&\quad= \mathbb{E}_{X\sim\psi_{k+1}}\Big[\log\frac{\psi_{k+1}(X)}{p^{\mathsf{c}}_{k+1}(X)}\Big] - \mathbb{E}_{X\sim P_{\gamma_k}\psi_k}\Big[\log\frac{\psi_{k+1}(X)}{p^{\mathsf{c}}_{k+1}(X)}\Big] - \mathbb{E}_{X\sim P_{\gamma_k}\psi_k}\Big[\log\frac{P_{\gamma_k}\psi_k(X)}{\psi_{k+1}(X)}\Big].
\end{aligned}$$

Notice that as the KL divergence between $P_{\gamma_k}\psi_k$ and ψ_{k+1} is non-negative, i.e.,

$$\mathbb{E}_{X\sim P_{\gamma_k}\psi_k}\Big[\log\frac{P_{\gamma_k}\psi_k(X)}{\psi_{k+1}(X)}\Big] \geq 0.$$

By using (71), we have

$$\begin{aligned}
&\mathsf{KL}(\psi_{k+1}\|p^{\mathsf{c}}_{k+1}) - \mathsf{KL}(P_{\gamma_k}\psi_k\|p^{\mathsf{c}}_{k+1}) \\
&\quad\leq \mathbb{E}_{X\sim\psi_{k+1}}\Big[\log\frac{\psi_{k+1}(X)}{p^{\mathsf{c}}_{k+1}(X)}\Big] - \mathbb{E}_{X\sim P_{\gamma_k}\psi_k}\Big[\log\frac{\psi_{k+1}(X)}{p^{\mathsf{c}}_{k+1}(X)}\Big] = O(\gamma_k^2).
\end{aligned} \tag{72}$$

Substituting (68) and (72) into (66), we have

$$\begin{aligned}
\mathsf{KL}(\psi_{k+1}\|p^{\mathsf{c}}_{k+1}) &\leq \mathrm{e}^{-2\kappa\gamma_k}\mathsf{KL}(\psi_k\|p^{\mathsf{c}}_k) + O\big(\sqrt{\delta_{k+1}\log\delta^{-1}}\big) + O(\gamma_k^2) \\
&\leq \mathrm{e}^{-2\kappa\sum_{i=k-N}^{k}\gamma_i}\mathsf{KL}(\psi_{k-N}\|p^{\mathsf{c}}_{k-N}) + O\Big(\frac{\sqrt{\delta}\log\delta^{-1}}{\delta\eta}\Big) + O(\max_k\gamma_k) \\
&\lesssim \mathrm{e}^{-2\Theta(\sqrt{\delta})\eta} + \frac{\log\delta^{-1}}{\sqrt{\delta}\eta} + \max_k\gamma_k \to 0, \qquad \text{as} \quad \frac{\delta\eta^2}{\log\delta^{-1}}\to\infty, \quad \delta\eta\to 0,
\end{aligned} \tag{73}$$

where we have used the assumption that $\sum_{i=k-N}^{k}\gamma_i \asymp \sqrt{\delta}\eta \to \infty$. Futhermore, notice that

$$\mathbb{P}(\mathcal{E}_k^{\mathsf{c}}) \leq \sum_{i=1}^{k+1}\mathbb{P}\left(O(\delta_{i+1}) + \sqrt{\delta_{i+1}}\|Z^{\mathsf{sc}}_{t_i}\|_2 \leq O(\sqrt{\delta_{i+1}}\log\delta^{-1})\right) \leq \delta^{10} \to 0, \qquad \text{as} \quad \delta\to 0. \tag{74}$$

We have

$$\begin{aligned}
&\mathsf{TV}(p_{\breve{X}^{\mathsf{c}}_k|\breve{X}^{\mathsf{s}}_k}(\cdot\,|\,\breve{x}^{\mathsf{s}}_k), p_{X^{\mathsf{c}}_{t_k}|X^{\mathsf{s}}_{t_k}}(\cdot\,|\,\breve{x}^{\mathsf{s}}_k)) \\
&\quad\leq \mathbb{E}_{\breve{x}^{\mathsf{s}}_{1:k-1}\sim\breve{X}^{\mathsf{s}}_{1:k-1}\,|\,\breve{X}^{\mathsf{s}}_k}\big[\mathsf{TV}(p_{\breve{X}^{\mathsf{c}}_k|\breve{X}^{\mathsf{s}}_{1:k}}(\cdot\,|\,\breve{x}^{\mathsf{s}}_{1:k}), p_{X^{\mathsf{c}}_{t_k}|X^{\mathsf{s}}_{t_k}}(\cdot\,|\,x^{\mathsf{s}}_k))\big] \\
&\quad\leq \mathbb{E}_{\breve{x}^{\mathsf{s}}_{1:k-1}\sim\breve{X}^{\mathsf{s}}_{1:k-1}\,|\,\breve{X}^{\mathsf{s}}_k,\breve{x}^{\mathsf{s}}_{1:k-1}\in\mathcal{E}_k}\big[\mathsf{TV}(p_{\breve{X}^{\mathsf{c}}_k|\breve{X}^{\mathsf{s}}_{1:k}}(\cdot\,|\,\breve{x}^{\mathsf{s}}_{1:k}), p_{X^{\mathsf{c}}_{t_k}|X^{\mathsf{s}}_{t_k}}(\cdot\,|\,x^{\mathsf{s}}_k))\big] + \mathbb{P}(\mathcal{E}_k^{\mathsf{c}}) \to 0.
\end{aligned}$$

We complete the proof.

Proof of (62). By Pinsker's inequality, we have

$$\begin{aligned}
&\mathsf{TV}(p_{\breve{X}_k\,|\,Y,Z}, p_{\widetilde{X}_k\,|\,Y,Z})^2 \leq \mathsf{KL}(p_{\widetilde{X}_k\,|\,Y,Z}\|p_{\breve{X}_k\,|\,Y,Z}) \\
&\leq \sum_{i=k-N}^{k-1}\mathbb{E}_{x^{\mathsf{sc}}_i\sim\widetilde{X}^{\mathsf{sc}}_i, x^{\mathsf{c}}_i\sim\widetilde{X}^{\mathsf{c}}_i, x^{\mathsf{c}}_{i-N}\sim\widetilde{X}^{\mathsf{c}}_{i-N}}\Big[\mathsf{KL}(p_{\widetilde{X}^{\mathsf{sc}}_{i+1}\,|\,\widetilde{X}^{\mathsf{sc}}_i,\widetilde{X}^{\mathsf{c}}_i}(\cdot\,|\,x^{\mathsf{sc}}_i, x^{\mathsf{c}}_i)\|p_{\breve{X}^{\mathsf{sc}}_{i+1}\,|\,\breve{X}^{\mathsf{sc}}_i,\breve{X}^{\mathsf{c}}_{i-N}}(\cdot\,|\,x^{\mathsf{sc}}_i, x^{\mathsf{c}}_{i-N}))\Big] \\
&\quad+ \sum_{i=k-N}^{k-1}\mathbb{E}_{x^{\mathsf{c}}_i\sim\widetilde{X}^{\mathsf{c}}_i, x^{\mathsf{sc}}_{i+1}\sim\widetilde{X}^{\mathsf{sc}}_{i+1}}\Big[\mathsf{KL}(p_{\widetilde{X}^{\mathsf{c}}_{i+1}\,|\,\widetilde{X}^{\mathsf{c}}_i,\widetilde{X}^{\mathsf{sc}}_{i+1}}(\cdot\,|\,x^{\mathsf{c}}_i, x^{\mathsf{sc}}_{i+1})\|p_{\breve{X}^{\mathsf{c}}_{i+1}\,|\,\breve{X}^{\mathsf{c}}_i,\breve{X}^{\mathsf{sc}}_{i+1}}(\cdot\,|\,x^{\mathsf{c}}_i, x^{\mathsf{sc}}_{i+1}))\Big] \\
&= \sum_{i=k-N}^{k-1}\mathbb{E}_{x^{\mathsf{sc}}_i\sim\widetilde{X}^{\mathsf{sc}}_i, x^{\mathsf{c}}_i\sim\widetilde{X}^{\mathsf{c}}_i, x^{\mathsf{c}}_{i-N}\sim\widetilde{X}^{\mathsf{c}}_{i-N}}\Big[\mathsf{KL}(p_{\widetilde{X}^{\mathsf{sc}}_{i+1}\,|\,\widetilde{X}^{\mathsf{sc}}_i,\widetilde{X}^{\mathsf{c}}_i}(\cdot\,|\,x^{\mathsf{sc}}_i, x^{\mathsf{c}}_i)\|p_{\breve{X}^{\mathsf{sc}}_{i+1}\,|\,\breve{X}^{\mathsf{sc}}_i,\breve{X}^{\mathsf{c}}_{i-N}}(\cdot\,|\,x^{\mathsf{sc}}_i, x^{\mathsf{c}}_{i-N}))\Big],
\end{aligned}$$

where the last equation holds since $p_{\widetilde{X}^{\mathsf{c}}_{i+1}\,|\,\widetilde{X}^{\mathsf{c}}_i,\widetilde{X}^{\mathsf{sc}}_{i+1}}(\cdot\,|\,x^{\mathsf{c}}_i,x^{\mathsf{sc}}_{i+1}) = p_{\breve{X}^{\mathsf{c}}_{i+1}\,|\,\breve{X}^{\mathsf{c}}_i,\breve{X}^{\mathsf{sc}}_{i+1}}(\cdot\,|\,x^{\mathsf{c}}_i,x^{\mathsf{sc}}_{i+1})$ by (61c). According to (61b), both $p_{\widetilde{X}^{\mathsf{s}}_{i+1}\,|\,\widetilde{X}^{\mathsf{s}}_i,\widetilde{X}^{\mathsf{c}}_i}(\cdot\,|\,x^{\mathsf{s}}_i,x^{\mathsf{c}}_i)$ and $p_{\breve{X}^{\mathsf{sc}}_{i+1}\,|\,\breve{X}^{\mathsf{sc}}_i,\breve{X}^{\mathsf{c}}_{i-N}}(\cdot\,|\,x^{\mathsf{s}}_i,x^{\mathsf{c}}_{i-N})$ are Gaussian distributions with the same variance $\sigma^2_{t_{i+1}}(1-\mathrm{e}^{-2\delta_{i+1}})I_d$ and different mean vectors. We have

$$\begin{aligned}
&\mathsf{KL}\left(p_{\widetilde{X}^{\mathsf{s}}_{i+1}\,|\,\widetilde{X}^{\mathsf{s}}_i,\widetilde{X}^{\mathsf{c}}_i}(\cdot\,|\,x^{\mathsf{s}}_i,x^{\mathsf{c}}_i)\|p_{\breve{X}^{\mathsf{sc}}_{i+1}\,|\,\breve{X}^{\mathsf{sc}}_i,\breve{X}^{\mathsf{c}}_{i-N}}(\cdot\,|\,x^{\mathsf{s}}_i,x^{\mathsf{c}}_{i-N})\right)\\
&\quad=\frac{\alpha^2_{t_{i+1}}\sigma^4_{t_i}(1-\mathrm{e}^{-2\delta_{i+1}})^2}{2\alpha^2_{t_i}\sigma^2_{t_{i+1}}(1-\mathrm{e}^{-2\delta_{i+1}})}\left\|\nabla_{x^{\mathsf{sc}}_i}\log p_{\lambda_i}(x^{\mathsf{c}}_{i-N},x^{\mathsf{sc}}_i)-\nabla_{x^{\mathsf{sc}}_i}\log p_{\lambda_i}(x^{\mathsf{c}}_i,x^{\mathsf{sc}}_i)\right\|_2^2\\
&\quad\lesssim\delta_{i+1}\left\|\nabla_{x^{\mathsf{sc}}_i}\log p_{\lambda_i}(x^{\mathsf{c}}_{i-N},x^{\mathsf{sc}}_i)-\nabla_{x^{\mathsf{sc}}_i}\log p_{\lambda_i}(x^{\mathsf{c}}_i,x^{\mathsf{sc}}_i)\right\|_2^2.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&\mathsf{TV}(p_{\breve{X}_k\,|\,Y,Z},p_{\widetilde{X}_k\,|\,Y,Z})^2\\
&\lesssim\sum_{i=k-N}^{k-1}\delta_{i+1}\mathbb{E}_{x^{\mathsf{sc}}_i\sim\widetilde{X}^{\mathsf{sc}}_i,x^{\mathsf{c}}_i\sim\widetilde{X}^{\mathsf{c}}_i,x^{\mathsf{c}}_{i-N}\sim\widetilde{X}^{\mathsf{c}}_{i-N}}\left[\left\|\nabla_{x^{\mathsf{sc}}_i}\log p_{\lambda_i}(x^{\mathsf{c}}_{i-N},x^{\mathsf{sc}}_i)-\nabla_{x^{\mathsf{sc}}_i}\log p_{\lambda_i}(x^{\mathsf{c}}_i,x^{\mathsf{sc}}_i)\right\|_2^2\right]\\
&\lesssim\sum_{i=k-N}^{k-1}\delta_{i+1}\asymp\delta^{1/2}\to0,\qquad\text{as}\quad\delta\to0.
\end{aligned}$$

Thus we complete the proof.

C.5 Proof of Lemma 3

It suffices to prove that

$$\mathsf{KL}(p_{\widehat{X}^{\mathsf{sc}}_{t_i}\,|\,\widehat{X}_{t_{i-1}}}\|p_{X^{\mathsf{sc}}_{t_i}\,|\,X_{t_{i-1}}})=o(\delta_i).\tag{75}$$

According to the update rule, $p_{\widehat{X}^{\mathsf{sc}}_{t_i}|\widehat{X}_{t_{i-1}}}$ is a Gaussian distribution. We consider the following density function:

$$\widehat{p}_i(x_i|x_{i-1})\coloneqq(2\pi\varrho^2)^{-d/2}\exp\left(-\frac{1}{2\varrho^2}\|x_i-\mu_0-\Delta\|_2^2\right),\tag{76}$$

where variance $\varrho^2=\sigma^2_{t_i}(1-\mathrm{e}^{-2\delta_i})$ and the mean vecotr $\mu_0+\Delta$ is given by

$$\mu_0=\frac{\alpha_{t_i}}{\alpha_{t_{i-1}}}\widehat{X}_{t_{i-1}},\qquad\Delta=\alpha_{t_i}\sigma_{t_{i-1}}\left(\mathrm{e}^{-\lambda_{t_{i-1}}}-\mathrm{e}^{-2\lambda_{t_i}+\lambda_{t_{i-1}}}\right)\nabla_{x_{i-1}}\log p_{X_{t_{i-1}}}(x_{i-1}).$$

By noticing that $V^\top_{\mathcal{S}^{\mathsf{prior}}_{t_i}}\nabla_{x_{i-1}}\log p_{X_{t_{i-1}}}(x_{i-1})=\nabla_{x^{\mathsf{sc}}_{i-1}}\log p_{\lambda_{i-1}}(x^{\mathsf{sc}}_{i-1},x^{\mathsf{c}}_{i-1})$, we conclude that $\widehat{p}_i(x_i|x_{i-1})$ has the same marginal distribution on directions $\mathcal{S}^{\mathsf{prior}}_{t_i}$ as $p_{\widehat{X}^{\mathsf{sc}}_{t_i}|\widehat{X}_{t_{i-1}}}$. It suffices to prove that

$$\mathsf{KL}(p_{\widehat{X}^{\mathsf{sc}}_{t_i}\,|\,\widehat{X}_{t_{i-1}}}\|p_{X^{\mathsf{sc}}_{t_i}\,|\,X_{t_{i-1}}})\le\mathsf{KL}(\widehat{p}_i\|p_{X_{t_i}\,|\,X_{t_{i-1}}})=o(\delta_i).\tag{77}$$

Towards this, we decompose $\widehat{p}_i$ and $p_{X_{t_i}\,|\,X_{t_{i-1}}}$ seperately.

- *Decomposition of $\widehat{p}_i(x_i|x_{i-1})$.* Expanding the exponent expression in (76) yields

$$\widehat{p}_i(x_i|x_{i-1})=(2\pi\varrho^2)^{-d/2}\exp\left(-\frac{1}{2\varrho^2}\|x_i-\mu_0\|_2^2+\frac{1}{\varrho^2}\Delta^\top(x_i-\mu_0)+\frac{1}{2\varrho^2}\|\Delta\|_2^2\right).\tag{78}$$

  Note that

$$\begin{aligned}
\frac{1}{\varrho^2}\Delta^\top(x_i-\mu_0)&=\frac{1}{\varrho^2}\alpha_{t_i}\sigma_{t_{i-1}}\left(\mathrm{e}^{-\lambda_{t_{i-1}}}-\mathrm{e}^{-2\lambda_{t_i}+\lambda_{t_{i-1}}}\right)\nabla_{x_{i-1}}\log p_{X_{t_{i-1}}}(x_{i-1})^\top(x_i-\mu_0)\\
&=\frac{1}{\varrho^2}\alpha_{t_i}\sigma_{t_{i-1}}\mathrm{e}^{-\lambda_{t_{i-1}}}\left(1-\mathrm{e}^{-2\delta_i}\right)\nabla_{x_{i-1}}\log p_{X_{t_{i-1}}}(x_{i-1})^\top(x_i-\mu_0)
\end{aligned}$$

$$= \frac{1}{\varrho^2}\frac{\alpha_{t_i}^2\sigma_{t_{i-1}}^2}{\alpha_{t_{i-1}}^2}\left(1-\mathrm{e}^{-2\delta_i}\right)\nabla_{x_{i-1}}\log p_{X_{t_{i-1}}}(x_{i-1})^\top\big(\frac{\alpha_{t_{i-1}}}{\alpha_{t_i}}x_i-\widehat{X}_{t_{i-1}}\big)$$
$$= \mathrm{e}^{2\delta_i}\nabla_{x_{i-1}}\log p_{X_{t_{i-1}}}(x_{i-1})^\top\big(\frac{\alpha_{t_{i-1}}}{\alpha_{t_i}}x_i-\widehat{X}_{t_{i-1}}\big).$$

By denoting $\widehat{x}_i=\alpha_{t_{i-1}}x_i/\alpha_{t_i}$, and noticing that $\|\Delta\|_2=O(\delta_i)$, we have

$$\widehat{p}_i(x_i|x_{i-1}) = (2\pi\varrho^2)^{-d/2}\exp\Big(-\frac{1}{2\varrho^2}\|x_i-\mu_0\|_2^2 + \mathrm{e}^{2\delta_i}\nabla_{x_{i-1}}\log p_{X_{t_{i-1}}}(x_{i-1})^\top\big(\widehat{x}_i-\widehat{X}_{t_{i-1}}\big)+O(\delta_i)\Big) \tag{79}$$

- *Decomposition of* $p_{X_{t_i}\,|\,X_{t_{i-1}}}(x_i|x_{i-1})$.

$$p_{X_{t_i}\,|\,X_{t_{i-1}}}(x_i|x_{i-1}) = p_{X_{t_{i-1}}\,|\,X_{t_i}}(x_{i-1}|x_i)\frac{p_{X_{t_i}}(x_i)}{p_{X_{t_{i-1}}}(x_{i-1})}. \tag{80}$$

Furthermore, according to the forward process

$$\begin{aligned}X_{t_{i-1}} &= \frac{\alpha_{t_{i-1}}}{\alpha_{t_i}}X_{t_i}+\left(\sigma_{t_{i-1}}^2-\frac{\alpha_{t_{i-1}}}{\alpha_{t_i}}\sigma_{t_i}^2\right)\mathcal{N}(0,I_d)\\ &= \frac{\alpha_{t_{i-1}}}{\alpha_{t_i}}X_{t_i}+\sigma_{t_{i-1}}\left(1-\mathrm{e}^{-2\delta_i}\right)^{1/2}\mathcal{N}(0,I_d)\\ &= \frac{\alpha_{t_{i-1}}}{\alpha_{t_i}}X_{t_i}+\frac{\sigma_{t_{i-1}}}{\sigma_{t_i}}\varrho\mathcal{N}(0,I_d),\end{aligned}$$

we have

$$\begin{aligned}p_{X_{t_{i-1}}\,|\,X_{t_i}}(x_{i-1}|x_i) &= (2\pi\varrho^2)^{-d/2}\left(\frac{\sigma_{t_i}}{\sigma_{t_{i-1}}}\right)^d\exp\left(-\frac{\sigma_{t_i}^2}{2\varrho^2\sigma_{t_{i-1}}^2}\big\|\widehat{X}_{t_{i-1}}-\frac{\alpha_{t_{i-1}}}{\alpha_{t_i}}\widehat{X}_{t_i}\big\|_2^2\right)\\ &= (2\pi\varrho^2)^{-d/2}\left(\frac{\sigma_{t_i}}{\sigma_{t_{i-1}}}\right)^d\exp\left(-\frac{\sigma_{t_i}^2\alpha_{t_{i-1}}^2}{2\varrho^2\sigma_{t_{i-1}}^2\alpha_{t_i}^2}\big\|\frac{\alpha_{t_i}}{\alpha_{t_{i-1}}}\widehat{X}_{t_{i-1}}-\widehat{X}_{t_i}\big\|_2^2\right)\\ &= (2\pi\varrho^2)^{-d/2}\left(\frac{\sigma_{t_i}}{\sigma_{t_{i-1}}}\right)^d\exp\left(-\frac{\mathrm{e}^{-2\delta_i}}{2\varrho^2}\big\|\widehat{X}_{t_i}-\mu_0\big\|_2^2\right).\end{aligned} \tag{81}$$

For the ratio $\frac{p_{X_{t_i}}(x_i)}{p_{X_{t_{i-1}}}(x_{i-1})}$, we have

$$\begin{aligned}\frac{p_{X_{t_i}}(x_i)}{p_{X_{t_{i-1}}}(x_{i-1})} &= \frac{p_{X_{t_i}}(x_i)}{p_{X_{t_{i-1}}}(\widehat{x}_i)}\frac{p_{X_{t_{i-1}}}(\widehat{x}_i)}{p_{X_{t_{i-1}}}(x_{i-1})} = \frac{p_{X_{t_i}}(x_i)}{p_{X_{t_{i-1}}}(\widehat{x}_i)}\exp\Big(\log p_{X_{t_{i-1}}}(\widehat{x}_i)-\log p_{X_{t_{i-1}}}(x_{i-1})\Big)\\ &= \frac{p_{X_{t_i}}(x_i)}{p_{X_{t_{i-1}}}(\widehat{x}_i)}\exp\Big(\nabla\log p_{X_{t_{i-1}}}(x_{i-1})^\top(\widehat{x}_i-x_{i-1})\Big).\end{aligned} \tag{82}$$

Furthermore, we have

$$\begin{aligned}&\frac{p_{X_{t_i}}(x_i)}{p_{X_{t_{i-1}}}(\widehat{x}_i)}\\ &= \left(\frac{\sigma_{t_{i-1}}}{\sigma_{t_i}}\right)^d\frac{\int p_{X_0}(x_0)\exp\left(-\frac{1}{2\sigma_{t_i}^2}\|x_i-\alpha_{t_i}x_0\|_2^2\right)\mathrm{d}x_0}{\int p_{X_0}(x_0)\exp\left(-\frac{\mathrm{e}^{-2\delta_i}}{2\sigma_{t_i}^2}\|x_i-\alpha_{t_i}x_0\|_2^2\right)\mathrm{d}x_0}\\ &= \left(\frac{\sigma_{t_{i-1}}}{\sigma_{t_i}}\right)^d\frac{\int p_{X_0}(x_0)\exp\left(-\frac{\mathrm{e}^{-2\delta_i}}{2\sigma_{t_i}^2}\|x_i-\alpha_{t_i}x_0\|_2^2\right)\left(1-\frac{(1-\mathrm{e}^{-2\delta_i})}{2\sigma_{t_i}^2}\|x_i-\alpha_{t_i}x_0\|_2^2+O(\delta_i^2)\right)\mathrm{d}x_0}{\int p_{X_0}(x_0)\exp\left(-\frac{\mathrm{e}^{-2\delta_i}}{2\sigma_{t_i}^2}\|x_i-\alpha_{t_i}x_0\|_2^2\right)\mathrm{d}x_0},\end{aligned}$$

where the first equation comes from

$$p_{X_{t_i}}(x_i) = (2\pi\sigma_{t_i}^2)^{-d/2}\int p_{X_0}(x_0)\exp\left(-\frac{1}{2\sigma_{t_i}^2}\|x_i-\alpha_{t_i}x_0\|_2^2\right)\mathrm{d}x_0$$

$$
\begin{aligned}
p_{X_{t_{i-1}}}(\widehat{x}_i) &= (2\pi\sigma_{t_{i-1}}^2)^{-d/2} \int p_{X_0}(x_0) \exp\left(-\frac{1}{2\sigma_{t_{i-1}}^2}\Big\|\frac{\alpha_{t_{i-1}}}{\alpha_{t_i}}x_i - \alpha_{t_{i-1}}x_0\Big\|_2^2\right) \mathrm{d}x_0 \\
&= (2\pi\sigma_{t_{i-1}}^2)^{-d/2} \int p_{X_0}(x_0) \exp\left(-\frac{\alpha_{t_{i-1}}^2}{2\sigma_{t_{i-1}}^2\alpha_{t_i}^2}\big\|x_{i-1} - \alpha_{t_i}x_0\big\|_2^2\right) \mathrm{d}x_0 \\
&= (2\pi\sigma_{t_{i-1}}^2)^{-d/2} \int p_{X_0}(x_0) \exp\left(-\frac{\mathrm{e}^{-2\delta_i}}{2\sigma_{t_i}^2}\big\|x_i - \alpha_{t_i}x_0\big\|_2^2\right) \mathrm{d}x_0.
\end{aligned}
$$

Thus we have

$$
\begin{aligned}
&\left|\left(\frac{\sigma_{t_{i-1}}}{\sigma_{t_i}}\right)^{-d} \frac{p_{X_{t_i}}(x_i)}{p_{X_{t_{i-1}}}(\widehat{x}_i)} - 1\right| \\
&\le \frac{\int p_{X_0}(x_0) \exp\left(-\frac{\mathrm{e}^{-2\delta_i}}{2\sigma_{t_i}^2}\|x_i - \alpha_{t_i}x_0\|_2^2\right)\left(\frac{2\delta_i}{2\sigma_{t_i}^2}\|x_i - \alpha_{t_i}x_0\|_2^2 + O(\delta_i^2)\right)\mathrm{d}x_0}{\int p_{X_0}(x_0) \exp\left(-\frac{\mathrm{e}^{-2\delta_i}}{2\sigma_{t_i}^2}\|x_i - \alpha_{t_i}x_0\|_2^2\right)\mathrm{d}x_0} \\
&= O(\delta_i)(\|x_i\|_2^2 + 1).
\end{aligned}
$$

Instituting into (82), we have

$$
\frac{p_{X_{t_i}}(x_i)}{p_{X_{t_{i-1}}}(x_{i-1})} = \left(\frac{\sigma_{t_{i-1}}}{\sigma_{t_i}}\right)^{-d} \exp\Big(\nabla \log p_{X_{t_{i-1}}}(x_{i-1})^\top(\widehat{x}_i - x_{i-1})\Big)\big(1 + O(\delta_i)(1 + \|x_i\|_2^2)\big) \tag{83}
$$

Combining this with (81) and (80) yields

$$
\begin{aligned}
p_{X_{t_i}\,|\,X_{t_{i-1}}}(x_i|x_{i-1}) &= (2\pi\varrho^2)^{-d/2} \exp\left(-\frac{\mathrm{e}^{-2\delta_i}}{2\varrho^2}\big\|\widehat{X}_{t_i} - \mu_0\big\|_2^2\right) \\
&\quad \cdot \exp\Big(\nabla \log p_{X_{t_{i-1}}}(x_{i-1})^\top(\widehat{x}_i - x_{i-1})\Big)\big(1 + O(\delta_i)(1 + \|x_i\|_2^2)\big) \\
&= (2\pi\varrho^2)^{-d/2} \exp\left(-\frac{\mathrm{e}^{-2\delta_i}}{2\varrho^2}\big\|\widehat{X}_{t_i} - \mu_0\big\|_2^2 + \nabla \log p_{X_{t_{i-1}}}(x_{i-1})^\top(\widehat{x}_i - x_{i-1})\right) \\
&\quad \cdot \big(1 + O(\delta_i)(1 + \|x_i\|_2^2)\big)
\end{aligned} \tag{84}
$$

Combining (79) and (84) leads to

$$
\begin{aligned}
\frac{p_{X_{t_i}\,|\,X_{t_{i-1}}}(x_i|x_{i-1})}{\widehat{p}_i(x_i|x_{i-1})} &= \exp\left(\frac{(1 - \mathrm{e}^{-2\delta_i})}{2\varrho^2}\big\|x_i - \mu_0\big\|_2^2 - (\mathrm{e}^{2\delta_i} - 1)\nabla \log p_{X_{t_{i-1}}}(x_{i-1})^\top(\widehat{x}_i - x_{i-1}) + O(\delta_i)\right) \\
&\quad \cdot (1 + O(\delta_i)(1 + \|x_i\|_2^2))
\end{aligned}
$$

and

$$
\frac{p_{X_{t_i}\,|\,X_{t_{i-1}}}(x_i|x_{i-1})}{\widehat{p}_i(x_i|x_{i-1})} - 1 = O(\delta_i)\frac{\|x_i - \mu_0\|_2^2}{\varrho^2} + O(\delta_i)\|\widehat{x}_i - x_{i-1}\|_2 + O(\delta_i)\|x_i\|^2 + O(\delta_i).
$$

By using the definition of KL divergence, we have

$$
\begin{aligned}
\mathsf{KL}(\widehat{p}_i \| p_{X_{t_i}\,|\,X_{t_{i-1}}}) &= \int \log \frac{\widehat{p}_i(x_i|x_{i-1})}{p_{X_{t_i}\,|\,X_{t_{i-1}}}(x_i|x_{i-1})} \widehat{p}_i(x_i|x_{i-1})\mathrm{d}x_i \\
&= -\int \log \frac{p_{X_{t_i}\,|\,X_{t_{i-1}}}(x_i|x_{i-1})}{\widehat{p}_i(x_i|x_{i-1})} \widehat{p}_i(x_i|x_{i-1})\mathrm{d}x_i \\
&= -\int \left(\frac{p_{X_{t_i}\,|\,X_{t_{i-1}}}(x_i|x_{i-1})}{\widehat{p}_i(x_i|x_{i-1})} - 1\right) \widehat{p}_i(x_i|x_{i-1})\mathrm{d}x_i \\
&\quad + \frac{1}{2}\int \left(\frac{p_{X_{t_i}\,|\,X_{t_{i-1}}}(x_i|x_{i-1})}{\widehat{p}_i(x_i|x_{i-1})} - 1\right)^2 \widehat{p}_i(x_i|x_{i-1})\mathrm{d}x_i
\end{aligned}
$$

$$= \frac{1}{2}\int \left(\frac{p_{X_{t_i}\,|\,X_{t_{i-1}}}(x_i|x_{i-1})}{\widehat{p}_i(x_i|x_{i-1})} - 1\right)^2 \widehat{p}_i(x_i|x_{i-1})\mathrm{d}x_i$$
$$= O(\delta_i^2)\mathbb{E}_{x_i\sim\widehat{p}_i(\cdot\,|\,x_{i-1})}\left[\frac{\|x_i-\mu_0\|_2^2}{\varrho^2} + \|\widehat{x}_i - x_{i-1}\|_2 + \|x_i\|^2 + 1\right] = O(\delta_i^2),$$

where the last equation holds since by (76),

$$\frac{\|x_i-\mu_0\|_2^2}{\varrho^2} = \frac{\|\Delta\|_2^2+\varrho^2}{\varrho^2} = O(1).$$

Combining with (77), we complete the proof.

C.6 Proof of Lemma 4

We start from the observation that $p_{\widetilde{X}_i^{\mathsf{sc}}\,|\,\widetilde{X}_{i-1}}(x_i^{\mathsf{sc}}\,|\,x_{i-1})$ is a Gaussian distribution and x_{i-1}^{c} influences the mean vector:

$$p_{\widetilde{X}_i^{\mathsf{sc}}\,|\,\widetilde{X}_{i-1}}(x_i^{\mathsf{sc}}\,|\,x_{i-1}) = \phi(x_i^{\mathsf{sc}};\mu_0+\Delta(x_{i-1}^{\mathsf{c}}),\sigma_{t_i}^2(1-\mathrm{e}^{-2\delta_i})),$$

where $\phi(\cdot;\mu,\sigma^2)$ denotes the density function of Gaussian distribution with mean vector μ and variance σ^2, and

$$\mu_0 \coloneqq \frac{\alpha_{t_i}}{\alpha_{t_{i-1}}}x_{i-1}^{\mathsf{sc}},$$
$$\Delta(x_{i-1}^{\mathsf{c}}) \coloneqq \alpha_{t_i}\sigma_{t_{i-1}}\left(\mathrm{e}^{-\lambda_{t_{i-1}}} - \mathrm{e}^{-2\lambda_{t_i}+\lambda_{t_{i-1}}}\right)\nabla_{x_{i-1}^{\mathsf{sc}}}\log p_{\lambda_{i-1}}(x_{i-1}^{\mathsf{c}},x_{i-1}^{\mathsf{sc}}). \tag{85}$$

It can be immediately check that

$$\|\Delta(x_{i-1}^{\mathsf{c}})\|_2 \lesssim O(\delta). \tag{86}$$

By Taylor expansion and denoting the variance as $\varrho^2 = \sigma_{t_i}^2(1-\mathrm{e}^{-2\delta_i})$, we have

$$p_{\widetilde{X}_i^{\mathsf{sc}}\,|\,\widetilde{X}_{i-1}}(x_i^{\mathsf{sc}}\,|\,x_{i-1}) = \phi(x_i^{\mathsf{sc}};\mu_0,\varrho^2)\left(1+\frac{(x_i^{\mathsf{sc}}-\mu_0)^\top\Delta(x_{i-1}^{\mathsf{c}})}{\varrho^2} - \frac{\|\Delta(x_{i-1}^{\mathsf{c}})\|_2^2}{\varrho^2} + \frac{((x_i^{\mathsf{sc}}-\mu_0)^\top\Delta(x_{i-1}^{\mathsf{c}}))^2}{\varrho^2}\right)$$
$$+ O(\|\Delta(x_{i-1}^{\mathsf{c}})\|_2^3).$$

Thus we have

$$\int p_{\widetilde{X}_i^{\mathsf{sc}}\,|\,\widetilde{X}_{i-1}}(x_i^{\mathsf{sc}}\,|\,x_{i-1})p_{X_{t_i}^{\mathsf{sc}}\,|\,X_{t_{i-1}}}(x_{i-1}^{\mathsf{c}}\,|\,x_{i-1}^{\mathsf{s}})\mathrm{d}x_{i-1}^{\mathsf{c}}$$
$$= \int \phi(x_i^{\mathsf{sc}};\mu_0,\varrho^2)\left(1+\frac{(x_i^{\mathsf{sc}}-\mu_0)^\top\Delta(x_{i-1}^{\mathsf{c}})}{\varrho^2} - \frac{\|\Delta(x_{i-1}^{\mathsf{c}})\|_2^2}{\varrho^2} + \frac{((x_i^{\mathsf{sc}}-\mu_0)^\top\Delta(x_{i-1}^{\mathsf{c}}))^2}{\varrho^2}\right)$$
$$\cdot\, p_{X_{t_{i-1}}^{\mathsf{c}}\,|\,X_{t_{i-1}}^{\mathsf{s}}}(x_{i-1}^{\mathsf{c}}\,|\,x_{i-1}^{\mathsf{s}})\mathrm{d}x_{i-1}^{\mathsf{c}}$$
$$+ O(\mathbb{E}_{X_{t_{i-1}}^{\mathsf{c}}\,|\,X_{t_{i-1}}^{\mathsf{s}}}[\|\Delta(X_{t_{i-1}}^{\mathsf{c}})\|_2^3])$$
$$= \phi(x_i^{\mathsf{sc}};\mu_0,\varrho^2)\Bigg(1+\varrho^{-2}(x_i^{\mathsf{sc}}-\mu_0)^\top\mathbb{E}_{X_{t_{i-1}}^{\mathsf{c}}\,|\,X_{t_{i-1}}^{\mathsf{s}}}[\Delta(X_{t_{i-1}}^{\mathsf{c}})] - \varrho^{-2}\mathbb{E}_{X_{t_{i-1}}^{\mathsf{c}}\,|\,X_{t_{i-1}}^{\mathsf{s}}}[\|\Delta(X_{t_{i-1}}^{\mathsf{c}})\|_2^2]$$
$$+ \varrho^{-2}\mathbb{E}_{X_{t_{i-1}}^{\mathsf{c}}\,|\,X_{t_{i-1}}^{\mathsf{s}}}[((x_i^{\mathsf{sc}}-\mu_0)^\top\Delta(X_{t_{i-1}}^{\mathsf{c}}))^2]\Bigg) + O(\mathbb{E}_{X_{t_{i-1}}^{\mathsf{c}}\,|\,X_{t_{i-1}}^{\mathsf{s}}}[\|\Delta(X_{t_{i-1}}^{\mathsf{c}})\|_2^3]).$$

Similarly, we have

$$\int p_{\widetilde{X}_i^{\mathsf{sc}}\,|\,\widetilde{X}_{i-1}}(x_i^{\mathsf{sc}}\,|\,x_{i-1})p_{\widetilde{X}_{i-1}^{\mathsf{c}}\,|\,\widetilde{X}_{i-1}^{\mathsf{s}}}(x_{i-1}^{\mathsf{c}}\,|\,x_{i-1}^{\mathsf{s}})\mathrm{d}x_{i-1}^{\mathsf{c}}$$
$$= \phi(x_i^{\mathsf{sc}};\mu_0,\varrho^2)\Bigg(1+\varrho^{-2}(x_i^{\mathsf{sc}}-\mu_0)^\top\mathbb{E}_{\widetilde{X}_{i-1}^{\mathsf{c}}\,|\,\widetilde{X}_{i-1}^{\mathsf{s}}}[\Delta(\widetilde{X}_{i-1}^{\mathsf{c}})] - \varrho^{-2}\mathbb{E}_{\widetilde{X}_{i-1}^{\mathsf{c}}\,|\,\widetilde{X}_{i-1}^{\mathsf{s}}}[\|\Delta(\widetilde{X}_{i-1}^{\mathsf{c}})\|_2^2]$$

$$+ \varrho^{-2}\mathbb{E}_{\widetilde{X}^{\mathsf{c}}_{i-1}\,|\,\widetilde{X}^{\mathsf{s}}_{i-1}}[((x^{\mathsf{sc}}_i-\mu_0)^\top\Delta(\widetilde{X}^{\mathsf{c}}_{i-1}))^2]\Big) + O(\mathbb{E}_{\widetilde{X}^{\mathsf{c}}_{i-1}\,|\,\widetilde{X}^{\mathsf{s}}_{i-1}}[\|\Delta(\widetilde{X}^{\mathsf{c}}_{i-1})\|_2^3]).$$

These lead to

$$\begin{aligned}
&\log\frac{\int p_{\widetilde{X}^{\mathsf{sc}}_i\,|\,\widetilde{X}_{i-1}}(x^{\mathsf{sc}}_i\,|\,x_{i-1})p_{\widetilde{X}^{\mathsf{c}}_{i-1}\,|\,\widetilde{X}^{\mathsf{s}}_{i-1},Y,Z}(x^{\mathsf{c}}_{i-1}\,|\,x^{\mathsf{s}}_{i-1},y,z)\mathrm{d}x^{\mathsf{c}}_{i-1}}{\int p_{\widetilde{X}^{\mathsf{sc}}_i\,|\,\widetilde{X}_{i-1}}(x^{\mathsf{sc}}_i\,|\,x_{i-1})p_{X^{\mathsf{c}}_{t_{i-1}}\,|\,X^{\mathsf{s}}_{t_{i-1}}}(x^{\mathsf{c}}_{i-1}\,|\,x^{\mathsf{s}}_{i-1})\mathrm{d}x^{\mathsf{c}}_{i-1}}\\
&\quad=\log\frac{1+\varrho^{-2}(x^{\mathsf{sc}}_i-\mu_0)^\top\mathbb{E}[\Delta(\widetilde{X}^{\mathsf{c}}_{i-1})]-\varrho^{-2}\mathbb{E}[\|\Delta(\widetilde{X}^{\mathsf{c}}_{i-1})\|_2^2]+\varrho^{-2}\mathbb{E}[((x^{\mathsf{sc}}_i-\mu_0)^\top\Delta(\widetilde{X}^{\mathsf{c}}_{i-1}))^2]}{1+\varrho^{-2}(x^{\mathsf{sc}}_i-\mu_0)^\top\mathbb{E}[\Delta(X^{\mathsf{c}}_{t_{i-1}})]-\varrho^{-2}\mathbb{E}[\|\Delta(X^{\mathsf{c}}_{t_{i-1}})\|_2^2]+\varrho^{-2}\mathbb{E}[((x^{\mathsf{sc}}_i-\mu_0)^\top\Delta(X^{\mathsf{c}}_{t_{i-1}}))^2]}+O(\delta^3)\\
&\quad\le\frac{\varrho^{-2}(x^{\mathsf{sc}}_i-\mu_0)^\top\big(\mathbb{E}[\Delta(\widetilde{X}^{\mathsf{c}}_{i-1})-\Delta(X^{\mathsf{c}}_{t_{i-1}})]\big)-\varrho^{-2}\big(\mathbb{E}[\|\Delta(\widetilde{X}^{\mathsf{c}}_{i-1})\|_2^2-\|\Delta(X^{\mathsf{c}}_{t_{i-1}})\|_2^2]\big)}{1+\varrho^{-2}(x^{\mathsf{sc}}_i-\mu_0)^\top\mathbb{E}[\Delta(X^{\mathsf{c}}_{t_{i-1}})]-\varrho^{-2}\mathbb{E}[\|\Delta(X^{\mathsf{c}}_{t_{i-1}})\|_2^2]+\varrho^{-2}\mathbb{E}[((x^{\mathsf{sc}}_i-\mu_0)^\top\Delta(X^{\mathsf{c}}_{t_{i-1}}))^2]}\\
&\qquad+\frac{\varrho^{-2}\mathbb{E}[((x^{\mathsf{sc}}_i-\mu_0)^\top\Delta(\widetilde{X}^{\mathsf{c}}_{i-1}))^2-((x^{\mathsf{sc}}_i-\mu_0)^\top\Delta(X^{\mathsf{c}}_{t_{i-1}}))^2]}{1+\varrho^{-2}(x^{\mathsf{sc}}_i-\mu_0)^\top\mathbb{E}[\Delta(X^{\mathsf{c}}_{t_{i-1}})]-\varrho^{-2}\mathbb{E}[\|\Delta(X^{\mathsf{c}}_{t_{i-1}})\|_2^2]+\varrho^{-2}\mathbb{E}[((x^{\mathsf{sc}}_i-\mu_0)^\top\Delta(X^{\mathsf{c}}_{t_{i-1}}))^2]}.
\end{aligned}$$

Therefore, the term considered in (46) is bounded by

$$\begin{aligned}
&\int\log\frac{\int p_{\widetilde{X}^{\mathsf{sc}}_i\,|\,\widetilde{X}_{i-1}}(x^{\mathsf{sc}}_i\,|\,x_{i-1})p_{\widetilde{X}^{\mathsf{c}}_{i-1}\,|\,\widetilde{X}^{\mathsf{s}}_{i-1},Y,Z}(x^{\mathsf{c}}_{i-1}\,|\,x^{\mathsf{s}}_{i-1},y,z)\mathrm{d}x^{\mathsf{c}}_{i-1}}{\int p_{\widetilde{X}^{\mathsf{sc}}_i\,|\,\widetilde{X}_{i-1}}(x^{\mathsf{sc}}_i\,|\,x_{i-1})p_{X^{\mathsf{c}}_{t_{i-1}}\,|\,X^{\mathsf{s}}_{t_{i-1}}}(x^{\mathsf{c}}_{i-1}\,|\,x^{\mathsf{s}}_{i-1})\mathrm{d}x^{\mathsf{c}}_{i-1}}p_{\widetilde{X}^{\mathsf{sc}}_i\,|\,\widetilde{X}^{\mathsf{s}}_{i-1},Y,Z}(x^{\mathsf{sc}}_i\,|\,x^{\mathsf{s}}_{i-1},y,z)\mathrm{d}x^{\mathsf{sc}}_i\\
&\lesssim\int\Big(\varrho^{-2}(x^{\mathsf{sc}}_i-\mu_0)^\top\big(\mathbb{E}[\Delta(\widetilde{X}^{\mathsf{c}}_{i-1})-\Delta(X^{\mathsf{c}}_{t_{i-1}})]\big)-\varrho^{-2}\big(\mathbb{E}[\|\Delta(\widetilde{X}^{\mathsf{c}}_{i-1})\|_2^2-\|\Delta(X^{\mathsf{c}}_{t_{i-1}})\|_2^2]\big)\\
&\qquad+\varrho^{-2}\mathbb{E}[((x^{\mathsf{sc}}_i-\mu_0)^\top\Delta(\widetilde{X}^{\mathsf{c}}_{i-1}))^2-((x^{\mathsf{sc}}_i-\mu_0)^\top\Delta(X^{\mathsf{c}}_{t_{i-1}}))^2]\Big)p_{\widetilde{X}^{\mathsf{sc}}_i\,|\,\widetilde{X}^{\mathsf{s}}_{i-1}}(x^{\mathsf{sc}}_i\,|\,x^{\mathsf{s}}_{i-1})\mathrm{d}x^{\mathsf{sc}}_i\\
&\le\varrho^{-2}\|\mathbb{E}[x^{\mathsf{sc}}_i-\mu_0]\|_2^\top\|\mathbb{E}[\Delta(X^{\mathsf{c}}_{t_{i-1}})-\Delta(\widetilde{X}^{\mathsf{c}}_{i-1})]\|_2+\varrho^{-2}\big|\mathbb{E}[\|\Delta(X^{\mathsf{c}}_{t_{i-1}})\|_2^2-\|\Delta(\widetilde{X}^{\mathsf{c}}_{i-1})\|_2^2]\big|\\
&\quad+\varrho^{-2}\mathbb{E}[\|x^{\mathsf{sc}}_i-\mu_0\|_2^2]\big\|\mathbb{E}[\Delta(X^{\mathsf{c}}_{t_{i-1}})\Delta(X^{\mathsf{c}}_{t_{i-1}})^\top-\Delta(\widetilde{X}^{\mathsf{c}}_{i-1})\Delta(\widetilde{X}^{\mathsf{c}}_{i-1})^\top]\big\|_{\mathsf{op}}.
\end{aligned}\tag{87}$$

According to the update rule of $\widetilde{X}^{\mathsf{sc}}_i$, the expectation of $\widetilde{X}^{\mathsf{sc}}_i-\mu_0$ conditioned on $\widetilde{X}^{\mathsf{s}}_{i-1}=x^{\mathsf{s}}_{i-1}$ is bounded by

$$\begin{aligned}
\varrho^{-2}\|\mathbb{E}[\widetilde{x}^{\mathsf{sc}}_i-\mu_0]\|_2&=\varrho^{-2}\|\mathbb{E}[\Delta(\widetilde{X}^{\mathsf{c}}_{i-1})\,|\,\widetilde{X}^{\mathsf{s}}_{i-1}=x^{\mathsf{s}}_{i-1}]\|_2\\
&=\frac{\alpha_{t_i}\sigma_{t_{i-1}}\mathrm{e}^{-\lambda_{t_{i-1}}}(1-\mathrm{e}^{-2\delta_i})}{\sigma_{t_i}^2(1-\mathrm{e}^{-2\delta_i})}\|\mathbb{E}[\nabla_{x^{\mathsf{sc}}_{i-1}}\log p_{\lambda_{i-1}}(\widetilde{X}^{\mathsf{c}}_{i-1},x^{\mathsf{sc}}_{i-1})\,|\,\widetilde{X}^{\mathsf{s}}_{i-1}=x^{\mathsf{s}}_{i-1}]\|_2\\
&=O(1).
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\varrho^{-2}\mathbb{E}[\|x^{\mathsf{sc}}_i-\mu_0\|_2^2]&=\frac{\alpha_{t_i}^2\sigma_{t_{i-1}}^2\mathrm{e}^{-2\lambda_{t_{i-1}}}(1-\mathrm{e}^{-2\delta_i})^2}{\sigma_{t_i}^2(1-\mathrm{e}^{-2\delta_i})}\mathbb{E}[\|\nabla_{x^{\mathsf{sc}}_{i-1}}\log p_{\lambda_{i-1}}(\widetilde{X}^{\mathsf{c}}_{i-1},x^{\mathsf{sc}}_{i-1})\|_2^2\,|\,\widetilde{X}^{\mathsf{s}}_{i-1}=x^{\mathsf{s}}_{i-1}]+d\\
&=O(1).
\end{aligned}$$

Morever, define

$$\overline{\Delta}(x^{\mathsf{c}}_{i-1}):=\delta_i^{-1}\Delta(x^{\mathsf{c}}_{i-1}).$$

We have $\mathbb{E}[\|\overline{\Delta}(X^{\mathsf{c}}_{t_{i-1}})\|_2]<\infty$ and $\mathbb{E}[\|\overline{\Delta}(\widetilde{X}^{\mathsf{c}}_{i-1})\|_2]<\infty$. Instituting into (87), we have

$$\begin{aligned}
&\int\log\frac{\int p_{\widetilde{X}^{\mathsf{sc}}_i\,|\,\widetilde{X}_{i-1}}(x^{\mathsf{sc}}_i\,|\,x_{i-1})p_{\widetilde{X}^{\mathsf{c}}_{i-1}\,|\,\widetilde{X}^{\mathsf{s}}_{i-1},Y,Z}(x^{\mathsf{c}}_{i-1}\,|\,x^{\mathsf{s}}_{i-1},y,z)\mathrm{d}x^{\mathsf{c}}_{i-1}}{\int p_{\widetilde{X}^{\mathsf{sc}}_i\,|\,\widetilde{X}_{i-1}}(x^{\mathsf{sc}}_i\,|\,x_{i-1})p_{X^{\mathsf{c}}_{t_{i-1}}\,|\,X^{\mathsf{s}}_{t_{i-1}}}(x^{\mathsf{c}}_{i-1}\,|\,x^{\mathsf{s}}_{i-1})\mathrm{d}x^{\mathsf{c}}_{i-1}}p_{\widetilde{X}^{\mathsf{sc}}_i\,|\,\widetilde{X}^{\mathsf{s}}_{i-1},Y,Z}(x^{\mathsf{sc}}_i\,|\,x^{\mathsf{s}}_{i-1},y,z)\mathrm{d}x^{\mathsf{sc}}_i\\
&\lesssim\delta_i\|\mathbb{E}[\overline{\Delta}(X^{\mathsf{c}}_{t_{i-1}})-\overline{\Delta}(\widetilde{X}^{\mathsf{c}}_{i-1})]\|_2+\delta_i|\mathbb{E}[\|\overline{\Delta}(X^{\mathsf{c}}_{t_{i-1}})\|_2^2-\|\overline{\Delta}(\widetilde{X}^{\mathsf{c}}_{i-1})\|_2^2]|\\
&\quad+\delta_i^2\|\mathbb{E}[\overline{\Delta}(X^{\mathsf{c}}_{t_{i-1}})\overline{\Delta}(X^{\mathsf{c}}_{t_{i-1}})^\top-\overline{\Delta}(\widetilde{X}^{\mathsf{c}}_{i-1})\overline{\Delta}(\widetilde{X}^{\mathsf{c}}_{i-1})^\top]\|_{\mathsf{op}}.
\end{aligned}\tag{88}$$

Thus we have

$$\sum_{i=1}^k\mathbb{E}_{x^{\mathsf{s}}_{i-1}\sim\widetilde{X}^{\mathsf{s}}_{i-1}\,|\,Y,Z}\Big[\int\log\frac{\int p_{\widetilde{X}^{\mathsf{sc}}_i\,|\,\widetilde{X}_{i-1}}(x^{\mathsf{sc}}_i\,|\,x_{i-1})p_{\widetilde{X}^{\mathsf{c}}_{i-1}\,|\,\widetilde{X}^{\mathsf{s}}_{i-1},Y,Z}(x^{\mathsf{c}}_{i-1}\,|\,x^{\mathsf{s}}_{i-1},y,z)\mathrm{d}x^{\mathsf{c}}_{i-1}}{\int p_{\widetilde{X}^{\mathsf{sc}}_i\,|\,\widetilde{X}_{i-1}}(x^{\mathsf{sc}}_i\,|\,x_{i-1})p_{X^{\mathsf{c}}_{t_{i-1}}\,|\,X^{\mathsf{s}}_{t_{i-1}}}(x^{\mathsf{c}}_{i-1}\,|\,x^{\mathsf{s}}_{i-1})\mathrm{d}x^{\mathsf{c}}_{i-1}}$$

$$\cdot\, p_{\widetilde{X}_i^{\mathsf{sc}} \mid \widetilde{X}_{i-1}^{\mathsf{s}}}(x_i^{\mathsf{sc}} \mid x_{i-1}^{\mathsf{s}}, y, z)\mathrm{d}x_i^{\mathsf{sc}}\Big]$$

$$\lesssim \max_i \mathbb{E}_{x_{i-1}^{\mathsf{sc}} \sim p_{\widetilde{X}_{i-1}^{\mathsf{sc}} \mid Y,Z}}\big[\|\mathbb{E}[\overline{\Delta}(X_{t_{i-1}}^{\mathsf{c}}) - \overline{\Delta}(\widetilde{X}_{i-1}^{\mathsf{c}}) \mid \widetilde{X}_{i-1}^{\mathsf{sc}} = X_{t_{i-1}}^{\mathsf{sc}} = x_{i-1}^{\mathsf{sc}}]\|_2\big]$$
$$+ \max_i \mathbb{E}_{x_{i-1}^{\mathsf{sc}} \sim p_{\widetilde{X}_{i-1}^{\mathsf{sc}} \mid Y,Z}}\big[|\mathbb{E}[\|\overline{\Delta}(X_{t_{i-1}}^{\mathsf{c}})\|_2^2 - \|\overline{\Delta}(\widetilde{X}_{i-1}^{\mathsf{c}})\|_2^2 \mid \widetilde{X}_{i-1}^{\mathsf{sc}} = X_{t_{i-1}}^{\mathsf{sc}} = x_{i-1}^{\mathsf{sc}}]|\big]$$
$$+ \delta \max_i \mathbb{E}_{x_{i-1}^{\mathsf{sc}} \sim p_{\widetilde{X}_{i-1}^{\mathsf{sc}} \mid Y,Z}}\big[\|\mathbb{E}[\overline{\Delta}(X_{t_{i-1}}^{\mathsf{c}})\overline{\Delta}(X_{t_{i-1}}^{\mathsf{c}})^\top - \overline{\Delta}(\widetilde{X}_{i-1}^{\mathsf{c}})\overline{\Delta}(\widetilde{X}_{i-1}^{\mathsf{c}})^\top \mid \widetilde{X}_{i-1}^{\mathsf{sc}} = X_{t_{i-1}}^{\mathsf{sc}} = x_{i-1}^{\mathsf{sc}}]\|_{\mathsf{op}}\big]$$
$$\to 0,$$

as Lemma 2 implies that the following limit holds for any function f with finite expectation:

$$\mathbb{E}_{x_{i-1}^{\mathsf{sc}} \sim p_{\widetilde{X}_{i-1}^{\mathsf{sc}} \mid Y,Z}}\big[\mathbb{E}[f(X_{t_{i-1}}^{\mathsf{c}}) - f(\widetilde{X}_{i-1}^{\mathsf{c}}) \mid \widetilde{X}_{i-1}^{\mathsf{sc}} = X_{t_{i-1}}^{\mathsf{sc}} = x_{i-1}^{\mathsf{sc}}]\big] \to 0.$$

C.7 Proof of Lemma 6

We will bound the norm of $\|e_{t_i,j}\|_2$ for $j = 1, 2, 3$ separately. We start from $\|e_{t_i,1}\|_2$, which is controlled by $|\frac{\sigma_{t_{i+1}}}{\sigma_{t_i}} - 1|\|\widehat{X}_{t_i}^{\mathsf{c}}\|_2$ according to its definition. Note that $\sigma_{t_{i+1}} \le \sigma_{t_i}$ and $\alpha_{t_{i+1}} \ge \alpha_{t_i}$. Thus we have

$$\big|\frac{\sigma_{t_{i+1}}}{\sigma_{t_i}} - 1\big| = 1 - \frac{\sigma_{t_{i+1}}}{\sigma_{t_i}} \le 1 - \frac{\alpha_{t_i}}{\alpha_{t_{i+1}}}\frac{\sigma_{t_{i+1}}}{\sigma_{t_i}} = 1 - \frac{\mathrm{e}^{-\lambda_{t_{i+1}}}}{\mathrm{e}^{-\lambda_{t_i}}} = 1 - \mathrm{e}^{-\delta_{i+1}} \le \delta_{i+1}. \tag{89}$$

Thus we have

$$\|e_{t_i,1}(x_{t_i}^{\mathsf{c}})\|_2^2 = \Big\| \left(\frac{\sigma_{t_{i+1}}}{\sigma_{t_i}} - 1\right)\sqrt{1 - \eta(1 - e^{-2\delta_{i+1}})}x_{t_i}^{\mathsf{c}} \Big\|_2^2 \le \delta_{i+1}^2 \|x_{t_i}^{\mathsf{c}}\|_2^2. \tag{90}$$

For the error term $\|e_{t_i,2}\|_2$, recalling that x_0 is bounded, which implies that $\mu_{t_i}^{\mathsf{c}}$ is also bounded by some constant B as it is a conditional expectation of x_0. Thus we have

$$\begin{aligned}\|e_{t_i,2}(x_{t_i}^{\mathsf{c}})\|_2 &= \Big\| \big(\left(\alpha_{t_{i+1}} - \alpha_{t_i}\right) + \left(\alpha_{t_{i+1}} e^{-\delta_{i+1}} - \alpha_{t_i}\right)\sqrt{1 - \eta(1 - e^{-2\delta_{i+1}})}\big)\mu_{t_i}^{\mathsf{c}}(x_{t_i}) \Big\|_2 \\ &\le B\left(\alpha_{t_{i+1}} - \alpha_{t_i} + |\alpha_{t_{i+1}}\mathrm{e}^{-\delta_{i+1}} - \alpha_{t_i}|\right) \le B\left(2\alpha_{t_{i+1}} - 2\alpha_{t_i} + \alpha_{t_{i+1}}(1 - \mathrm{e}^{-\delta_{i+1}})\right).\end{aligned} \tag{91}$$

Note that

$$\alpha_{t_{i+1}} - \alpha_{t_i} = \alpha_{t_{i+1}}\left(1 - \frac{\alpha_{t_i}}{\alpha_{t_{i+1}}}\right) \le \alpha_{t_{i+1}}\left(1 - \frac{\sigma_{t_{i+1}}}{\sigma_{t_i}}\frac{\alpha_{t_i}}{\alpha_{t_{i+1}}}\right) = \alpha_{t_{i+1}}\left(1 - \mathrm{e}^{-\delta_{i+1}}\right) \le \alpha_{t_{i+1}}\delta_{i+1}. \tag{92}$$

Inserting into (91), we have

$$\|e_{t_i,2}(x_{t_i}^{\mathsf{c}})\|_2^2 \le B^2\alpha_{t_{i+1}}^2\left(2\delta_{i+1} + \delta_{i+1}\right)^2 = O(\delta_{i+1}^2). \tag{93}$$

For the error term $\|e_{t_i,3}\|_2$, we have

$$\begin{aligned}\|e_{t_i,3}(x_{t_i}, x_{t_{i+1}}^{\mathsf{s}})\|_2 &= \Big\| \big(1 - \sqrt{1 - \eta(1 - e^{-2\delta_{i+1}})}\big)\left(\sigma_{t_i}\epsilon_{t_i}^{\mathsf{c}}(X_{t_i}) - \sigma_{t_{i+1}}\epsilon_{t_{i+1}}^{\mathsf{c}}(x_{t_{i+1}}^{\mathsf{est}})\right) \Big\|_2 \\ &\overset{\text{(i)}}{\le} \sigma_{t_{i+1}}^{-2}\gamma_i \Big\| \sigma_{t_i}\epsilon_{t_i}^{\mathsf{c}} - \sigma_{t_{i+1}}\epsilon_{t_{i+1}}^{\mathsf{c}} \Big\|_2 \\ &= \sigma_{t_{i+1}}^{-2}\gamma_i \Big\| \sigma_{t_i}^2 \nabla \log p_{X_{t_i}^{\mathsf{c}}, X_{t_i}^{\mathsf{sc}} \mid X_{t_i}^{\mathsf{s}}}(x_{t_i}^{\mathsf{c}}, x_{t_i}^{\mathsf{sc}} \mid \alpha_{t_i}\overline{Y}_{t_i} + \hat{\sigma}_{t_i}\overline{Z}_{t_i}) \\ &\qquad - \sigma_{t_{i+1}}^2 \nabla \log p_{X_{t_{i+1}}^{\mathsf{c}}, X_{t_{i+1}}^{\mathsf{sc}} \mid x_{t_{i+1}}^{\mathsf{s}}}(x_{t_i}^{\mathsf{c}}, x_{t_{i+1}}^{\mathsf{sc}} \mid \alpha_{t_{i+1}}\overline{Y}_{t_i} + \hat{\sigma}_{t_{i+1}}\overline{Z}_{t_i}) \Big\|,\end{aligned}$$

where $x_{t_{i+1}}^{\mathsf{est}}$ is defined by $V_{\mathcal{S}^{\mathsf{c}}}^\top x_{t_{i+1}}^{\mathsf{est}} = x_{t_i}^{\mathsf{c}}$ and $V_{\mathcal{S}}^\top x_{t_{i+1}}^{\mathsf{est}} = x_{t_{i+1}}^{\mathsf{s}}$, (i) arises from

$$1 - \sqrt{1 - \eta(1 - \mathrm{e}^{-2\delta_{i+1}})} \le \eta(1 - \mathrm{e}^{-2\delta_{i+1}}) = \sigma_{t_{i+1}}^{-2}\gamma_i.$$

and the last equation comes from (53). Combining with (54), we have

$$
\begin{aligned}
&\|e_{t_i,3}(x_{t_i}, x^{\mathsf{s}}_{t_{i+1}})\|_2 \\
&\le \sigma^{-2}_{t_{i+1}}\gamma_i \Big\| \mathbb{E}_{X_0^{\mathsf{c}}}[x^{\mathsf{c}}_{t_i} - \alpha_{t_i} X^{\mathsf{c}}_0 | X^{\mathsf{c}}_{t_i} = x^{\mathsf{c}}_{t_i}, X^{\mathsf{sc}}_{t_i} = x^{\mathsf{sc}}_{t_i}, X^{\mathsf{s}}_{t_i} = \alpha_{t_i}\overline{Y}_{t_i} + \hat{\sigma}_{t_i}\overline{Z}_{t_i}] \\
&\qquad - \mathbb{E}_{X_0^{\mathsf{c}}}[x^{\mathsf{c}}_{t_i} - \alpha_{t_{i+1}} X^{\mathsf{c}}_0 | X^{\mathsf{c}}_{t_{i+1}} = x^{\mathsf{c}}_{t_i}, X^{\mathsf{sc}}_{t_{i+1}} = x^{\mathsf{sc}}_{t_{i+1}}, X^{\mathsf{s}}_{t_{i+1}} = \alpha_{t_{i+1}}\overline{Y}_{t_{i+1}} + \hat{\sigma}_{t_{i+1}}\overline{Z}_{t_{i+1}}] \Big\|_2 \\
&= \sigma^{-2}_{t_{i+1}}\gamma_i \Big\| \mathbb{E}_{X_0^{\mathsf{c}}}[\alpha_{t_i} X^{\mathsf{c}}_0 | X^{\mathsf{c}}_{t_i} = x^{\mathsf{c}}_{t_i}, X^{\mathsf{sc}}_{t_i} = x^{\mathsf{sc}}_{t_i}, X^{\mathsf{s}}_{t_i} = \alpha_{t_i}\overline{Y}_{t_i} + \hat{\sigma}_{t_i}\overline{Z}_{t_i}] \\
&\qquad - \mathbb{E}_{X_0^{\mathsf{c}}}[\alpha_{t_{i+1}} X^{\mathsf{c}}_0 | X^{\mathsf{c}}_{t_{i+1}} = x^{\mathsf{c}}_{t_i}, X^{\mathsf{sc}}_{t_{i+1}} = x^{\mathsf{sc}}_{t_{i+1}}, X^{\mathsf{s}}_{t_{i+1}} = \alpha_{t_{i+1}}\overline{Y}_{t_{i+1}} + \hat{\sigma}_{t_{i+1}}\overline{Z}_{t_{i+1}}] \Big\|_2 \\
&\le \sigma^{-2}_{t_{i+1}}\gamma_i(\alpha_{t_{i+1}} - \alpha_{t_i}) \Big\| \mathbb{E}_{X_0^{\mathsf{c}}}[X^{\mathsf{c}}_0 | X^{\mathsf{c}}_{t_i} = x^{\mathsf{c}}_{t_i}, X^{\mathsf{sc}}_{t_i} = x^{\mathsf{sc}}_{t_i}, X^{\mathsf{s}}_{t_i} = \alpha_{t_i}\overline{Y}_{t_i} + \hat{\sigma}_{t_i}\overline{Z}_{t_i}] \Big\|_2 \\
&\qquad + \sigma^{-2}_{t_{i+1}}\gamma_i\alpha_{t_{i+1}} \Big\| \mathbb{E}_{X_0^{\mathsf{c}}}[X^{\mathsf{c}}_0 | X^{\mathsf{c}}_{t_i} = x^{\mathsf{c}}_{t_i}, X^{\mathsf{sc}}_{t_i} = x^{\mathsf{sc}}_{t_i}, X^{\mathsf{s}}_{t_i} = \alpha_{t_i}\overline{Y}_{t_i} + \hat{\sigma}_{t_i}\overline{Z}_{t_i}] \\
&\qquad - \mathbb{E}_{X_0^{\mathsf{c}}}[X^{\mathsf{c}}_0 | X^{\mathsf{c}}_{t_{i+1}} = x^{\mathsf{c}}_{t_i}, X^{\mathsf{sc}}_{t_{i+1}} = x^{\mathsf{sc}}_{t_{i+1}}, X^{\mathsf{s}}_{t_{i+1}} = \alpha_{t_{i+1}}\overline{Y}_{t_{i+1}} + \hat{\sigma}_{t_{i+1}}\overline{Z}_{t_{i+1}}] \Big\|_2 . \qquad (94)
\end{aligned}
$$

Recall that X_0 is bounded. By using (92), the first term is bounded by

$$
(\alpha_{t_{i+1}} - \alpha_{t_i}) \Big\| \mathbb{E}_{X_0^{\mathsf{c}}}[X^{\mathsf{c}}_0 | X^{\mathsf{c}}_{t_i} = x^{\mathsf{c}}_{t_i}, X^{\mathsf{sc}}_{t_i} = x^{\mathsf{sc}}_{t_i}, X^{\mathsf{s}}_{t_i} = \alpha_{t_i}\overline{Y}_{t_i} + \hat{\sigma}_{t_i}\overline{Z}_{t_i}] \Big\|_2 = O(\delta_{i+1}). \qquad (95)
$$

We claim that the second term is bounded by the following inequality, whose proof is postponed to the end of this section.

$$
\begin{aligned}
&\Big\| \mathbb{E}_{X_0^{\mathsf{c}}}[X^{\mathsf{c}}_0 | X^{\mathsf{c}}_{t_i} = x^{\mathsf{c}}_{t_i}, X^{\mathsf{sc}}_{t_i} = x^{\mathsf{sc}}_{t_i}, X^{\mathsf{s}}_{t_i} = \alpha_{t_i}\overline{Y}_{t_i} + \hat{\sigma}_{t_i}\overline{Z}_{t_i}] \\
&\qquad - \mathbb{E}_{X_0^{\mathsf{c}}}[X^{\mathsf{c}}_0 | X^{\mathsf{c}}_{t_{i+1}} = x^{\mathsf{c}}_{t_i}, X^{\mathsf{sc}}_{t_{i+1}} = x^{\mathsf{sc}}_{t_{i+1}}, X^{\mathsf{st}}_{t_{i+1}} = \alpha_{t_{i+1}}\overline{Y}_{t_{i+1}} + \hat{\sigma}_{t_{i+1}}\overline{Z}_{t_{i+1}}] \Big\|_2 \\
&\qquad \lesssim \Big(\|x^{\mathsf{c}}_{t_i}\|_2 + \|x^{\mathsf{sc}}_{t_i}\|_2 + \|Z\|_2 + 1 \Big)\delta_{i+1} + \|x^{\mathsf{sc}}_{t_{i+1}} - x^{\mathsf{sc}}_{t_i}\|_2 . \qquad (96)
\end{aligned}
$$

Combining (96) and (95) together with (94), we have

$$
\|e_{t_i,3}(x_{t_i}, x^{\mathsf{s}}_{t_{i+1}})\|_2 \lesssim \Big(\|x^{\mathsf{c}}_{t_i}\|_2 + \|x^{\mathsf{sc}}_{t_i}\|_2 + \|Z\|_2 + 1 \Big)\delta_{i+1}\gamma_i + \gamma_i \|x^{\mathsf{sc}}_{t_{i+1}} - x^{\mathsf{sc}}_{t_i}\|_2 . \qquad (97)
$$

For the last term in the right-hand-side of (56), we have

$$
\begin{aligned}
&\Big\| \Big(1 - \sqrt{1 - \eta(1 - e^{-2\delta_{i+1}})} - \frac{\eta(1 - e^{-2\delta_{i+1}})}{2} \Big) \sigma_{t_{i+1}} \epsilon^{\mathsf{c}}_{t_{i+1}} \Big\|_2 \\
&\le \Big| 1 - \sqrt{1 - \eta(1 - e^{-2\delta_{i+1}})} - \frac{\eta(1 - e^{-2\delta_{i+1}})}{2} \Big| \sigma_{t_{i+1}} \|\epsilon^{\mathsf{c}}_{t_{i+1}}\|_2 \\
&\lesssim \gamma_i^2 \frac{\|\epsilon^{\mathsf{c}}_{t_{i+1}}\|_2}{\sigma^3_{t_{i+1}}} \le \frac{\gamma_i^2}{\sigma^4_{t_{i+1}}} (\|x^{\mathsf{c}}_{t_i}\|_2 + \alpha_{t_i} B). \qquad (98)
\end{aligned}
$$

Combining (90), (93), (97), and (98), we have

$$
\|e_{t_i}(x_{t_i}, x^{\mathsf{s}}_{t_{i+1}})\|_2^2 \lesssim \Big(\|x^{\mathsf{c}}_{t_i}\|_2 + \|x^{\mathsf{sc}}_{t_i}\|_2 + \|Z\|_2 + 1 \Big)^2 \delta^2_{i+1} + (\|x^{\mathsf{c}}_{t_i}\|_2^2 + 1)\gamma_i^4 + \gamma_i^2 \|x^{\mathsf{sc}}_{t_{i+1}} - x^{\mathsf{sc}}_{t_i}\|_2^2 .
$$

It suffices to prove that

$$
\mathbb{E}_{\widehat{X}^{\mathsf{sc}}_{t_{i+1}} | \widehat{X}_{t_i} = x_{t_i}} [\|\widehat{X}^{\mathsf{sc}}_{t_{i+1}} - \widehat{X}^{\mathsf{sc}}_{t_i}\|_2^2] = O(\delta^2_{i+1})(\|x_{t_i}\|_2^2 + 1) + O(\delta_i).
$$

According to the update rule of $\widehat{X}^{\mathsf{sc}}_{t_{i+1}}$, we have

$$
\begin{aligned}
\mathbb{E}_{\widehat{X}^{\mathsf{sc}}_{t_{i+1}} | \widehat{X}_{t_i} = x_{t_i}} [\|\widehat{X}^{\mathsf{sc}}_{t_{i+1}} - \widehat{X}^{\mathsf{sc}}_{t_i}\|_2^2] &\lesssim \Big(\frac{\sigma_{t_{i+1}}}{\sigma_{t_i}} \mathrm{e}^{-\delta_{i+1}} - 1 \Big)^2 \|x^{\mathsf{sc}}_{t_i}\|_2^2 + \delta^2_{i+1} + \delta_{i+1} \mathbb{E}[\|z^{\mathsf{sc}}_{t_i}\|_2^2] \\
&= \delta^2_{i+1}(\|x^{\mathsf{sc}}_{t_i}\|_2^2 + 1).
\end{aligned}
$$

and complete the proof.

Proof of (96). For the second term, we introduce the following notation $q_{\lambda_i}(x_0|x')$ to simplify the expression:

$$q_{\lambda_i}(x_0|x') \coloneqq p_{X_0|X_{t_i}}(x_0|x') = \frac{\phi(x'|\alpha_{t_i}x_0, \sigma_{t_i})p_{X_0}(x_0)}{p_{X_{t_i}}(x')},$$

where $\phi(\cdot|\mu, \sigma)$ denotes the Gaussian density functon with mean vector μ and covariance matrix $\sigma^2 I$. Armed with this notation, we can control the second term as

$$\begin{aligned}
&\Big\|\mathbb{E}_{X_0^{\mathsf{c}}}[X_0^{\mathsf{c}}|X_{t_i}^{\mathsf{c}} = x_{t_i}^{\mathsf{c}}, X_{t_i}^{\mathsf{sc}} = x_{t_i}^{\mathsf{sc}}, X_{t_i}^{\mathsf{s}} = \alpha_{t_i}\overline{Y}_{t_i} + \hat{\sigma}_{t_i}\overline{Z}_{t_i}] \\
&\quad - \mathbb{E}_{X_0^{\mathsf{c}}}[X_0^{\mathsf{c}}|X_{t_{i+1}}^{\mathsf{c}} = x_{t_i}^{\mathsf{c}}, X_{t_{i+1}}^{\mathsf{sc}} = x_{t_{i+1}}^{\mathsf{sc}}, X_{t_{i+1}}^{\mathsf{s}} = \alpha_{t_{i+1}}\overline{Y}_{t_{i+1}} + \hat{\sigma}_{t_{i+1}}\overline{Z}_{t_{i+1}}]\Big\|_2 \\
&= \left\|\int x_0 \left(q_{\lambda_i}(x_0|\tilde{x}_{t_i}) - q_{\lambda_{i+1}}(x_0|\tilde{x}_{t_{i+1}})\right) \mathrm{d}x_0\right\|_2 \leq B\int |q_{\lambda_i}(x_0|\tilde{x}_{t_i}) - q_{\lambda_{i+1}}(x_0|\tilde{x}_{t_{i+1}})|\mathrm{d}x_0,
\end{aligned} \tag{99}$$

where $\tilde{x}_{t_i} \coloneqq [x_{t_i}^{\mathsf{c}}, x_{t_i}^{\mathsf{sc}}, \alpha_{t_i}\overline{Y}_{t_i} + \hat{\sigma}_{t_i}\overline{Z}_{t_i}]$ and $\tilde{x}_{t_{i+1}} \coloneqq [x_{t_i}^{\mathsf{c}}, x_{t_{i+1}}^{\mathsf{sc}}, \alpha_{t_{i+1}}\overline{Y}_{t_{i+1}} + \hat{\sigma}_{t_{i+1}}\overline{Z}_{t_{i+1}}]$. According to the definition of q_{λ_i}, we have

$$\begin{aligned}
\frac{1}{q_{\lambda_i}(x|x')}\left|\frac{\partial q_{\lambda_i}(x|x')}{\partial \alpha_{t_i}}\right| &= \frac{\partial \log q_{\lambda_i}(x|x')}{\partial \alpha_{t_i}} = \frac{\partial \log \phi(x'|\alpha_{t_i}x, \sigma_{t_i})}{\partial \alpha_{t_i}} - \frac{\partial \log p_{X_{t_i}}(x')}{\partial \alpha_{t_i}} \\
&= -\frac{1}{2}\frac{\partial}{\partial \alpha_{t_i}}\frac{1}{\sigma_{t_i}^2}\|x' - \alpha_{t_i}x\|_2^2 - \frac{\partial \log p_{X_{t_i}}(x')}{\partial \alpha_{t_i}}.
\end{aligned} \tag{100}$$

Moreover, we have

$$\begin{aligned}
\left|\frac{\partial \log p_{X_{t_i}}(x')}{\partial \alpha_{t_i}}\right| &= \frac{1}{2\sigma_{t_i}^2}\left|\int \frac{\partial}{\partial \alpha_{t_i}}\|x' - \alpha_{t_i}x\|_2^2 \frac{\phi(x'|\alpha_{t_i}x, \sigma_{t_i})p_{X_0}(x)}{p_{X_{t_i}}(x')}\mathrm{d}x\right| \\
&= \frac{1}{\sigma_{t_i}^2}\left|\int (x' - \alpha_{t_i}x)^\top x \frac{\phi(x'|\alpha_{t_i}x, \sigma_{t_i})p_{X_0}(x)}{p_{X_{t_i}}(x')}\mathrm{d}x\right| \\
&= \frac{1}{\sigma_{t_i}^2}\left|\int (x' - \alpha_{t_i}x)^\top x q_{\lambda_i}(x|x')\mathrm{d}x\right| \\
&\lesssim O(\|x'\|_2 + 1).
\end{aligned}$$

Substituting into (100), we have

$$\frac{1}{q_{\lambda_i}(x|x')}\left|\frac{\partial q_{\lambda_i}(x|x')}{\partial \alpha_{t_i}}\right| \lesssim O(\|x'\|_2 + 1).$$

Furthermore, we have the following inequality in a similar way

$$\begin{aligned}
\frac{1}{q_{\lambda_i}(x|x')}\frac{\partial q_{\lambda_i}(x|x')}{\partial \sigma_{t_i}} &= -\frac{1}{2}\frac{\partial}{\partial \sigma_{t_i}}\frac{1}{\sigma_{t_i}^2}\|x' - \alpha_{t_i}x\|_2^2 - \frac{d}{\sigma_{t_i}} - \frac{\partial \log p_{X_{t_i}}(x')}{\partial \sigma_{t_i}} \\
&\overset{\text{(i)}}{=} \sigma_{t_i}^{-3}\|x' - \alpha_{t_i}x\|_2^2 - \int \sigma_{t_i}^{-3}\|x' - \alpha_{t_i}x\|_2^2 q_{\lambda_i}(x|x')\mathrm{d}x,
\end{aligned} \tag{101}$$

where (i) comes from

$$\frac{\partial \log p_{X_{t_i}}(x')}{\partial \sigma_{t_i}} = \int \sigma_{t_i}^{-3}\|x' - \alpha_{t_i}x\|_2^2 q_{\lambda_i}(x|x')\mathrm{d}x - \frac{d}{\sigma_{t_i}}.$$

By using the bounded property of X_0 again, we have

$$\left|\frac{1}{q_{\lambda_i}(x|x')}\frac{\partial q_{\lambda_i}(x|x')}{\partial \sigma_{t_i}}\right| = O(\|x'\|_2 + 1).$$

In addition, we have

$$\nabla_{x'}\log q_{\lambda_i}(x|x') = -\nabla_{x'}\frac{1}{2\sigma_{t_i}^2}\|x' - \alpha_{t_i}x\|_2^2 - \nabla_{x'}\log p_{X_{t_i}}(x')$$

$$= -\frac{1}{\sigma_{t_i}^2}(x' - \alpha_{t_i}x) + \frac{1}{\sigma_{t_i}^2}\mathbb{E}[x' - \alpha_{t_i}X_0 | X_{t_i} = x'],$$
$$\|\nabla_{x'} \log q_{\lambda_i}(x|x')\|_2 = O(\|x'\|_2).$$

Moreover, we have

$$\|\tilde{x}_{t_{i+1}} - \tilde{x}_{t_i}\|_2 \le (\alpha_{t_{i+1}} - \alpha_{t_i})\|\overline{y}_{t_i}\|_2 + (\sigma_{t_{i+1}} - \sigma_{t_i})\|\overline{Z}_{t_i}\|_2 + \|x^{\mathsf{sc}}_{t_{i+1}} - x^{\mathsf{sc}}_{t_i}\|_2.$$

Combining these together, we have

$$\begin{aligned}
&|q_{\lambda_i}(x|\tilde{x}_{t_i}) - q_{\lambda_{i+1}}(x|\tilde{x}_{t_{i+1}})| \\
&\lesssim q_{\lambda_i}(x|\tilde{x}_{t_i})(\|x^{\mathsf{c}}_{t_i}\|_2 + \|x^{\mathsf{sc}}_{t_i}\|_2 + \|Z\|_2 + 1)(|\alpha_{t_i} - \alpha_{t_{i+1}}| + |\sigma_{t_i} - \sigma_{t_{i+1}}|) \\
&\quad + q_{\lambda_i}(x|\tilde{x}_{t_i})\|x^{\mathsf{sc}}_{t_{i+1}} - x^{\mathsf{sc}}_{t_i}\|_2 + O((|\alpha_{t_i} - \alpha_{t_{i+1}}| + |\sigma_{t_i} - \sigma_{t_{i+1}}|)^2) + O(\|x^{\mathsf{sc}}_{t_{i+1}} - x^{\mathsf{sc}}_{t_i}\|_2^2)
\end{aligned}$$

According to (92) and (89), we have $|\alpha_{t_i} - \alpha_{t_{i+1}}| = O(\delta_{i+1})$ and $|\sigma_{t_i} - \sigma_{t_{i+1}}| = O(\delta_{i+1})$, and thus

$$\begin{aligned}
|q_{\lambda_i}(x|\tilde{x}_{t_i}) - q_{\lambda_{i+1}}(x|\tilde{x}_{t_{i+1}})| \lesssim q_{\lambda_i}(x|\tilde{x}_{t_i})\bigg(\|x^{\mathsf{c}}_{t_i}\|_2 + \|x^{\mathsf{sc}}_{t_i}\|_2 + \|Z\|_2 + 1\bigg)\delta_{i+1} \\
+ q_{\lambda_i}(x|\tilde{x}_{t_i})\|x^{\mathsf{sc}}_{t_{i+1}} - x^{\mathsf{sc}}_{t_i}\|_2 + O(\delta_{i+1}^2) + O(\|x^{\mathsf{sc}}_{t_{i+1}} - x^{\mathsf{sc}}_{t_i}\|_2^2).
\end{aligned} \tag{102}$$

Substituting (102) into (99), we complete the proof.

C.8 Proof of Theorem 1 for general singular values

Akin to the proof in Section B, we first construct the following auxiliary sequence:

$$\overline{X}^{\mathsf{st}}_{k+1} = \alpha_{t_{k+1}}\overline{Y}_{t_{k+1}} + \widehat{\sigma}_{t_{k+1}}\overline{Z}_{t_{k+1}}, \quad \overline{X}^{\mathsf{sc}}_{k+1}, \overline{X}^{\mathsf{c}}_{k+1} \sim p_{X^{\mathsf{sc}}_{k+1}, X^{\mathsf{c}}_{k+1} | X^{\mathsf{st}}_{k+1}, X^{\mathsf{sc},t_{k+1}}_{\tau_s}}(\cdot, \cdot \mid \overline{X}^{\mathsf{st}}_{k+1}, \overline{X}^{\mathsf{sc},t_{k+1}}_{k_{\tau_s}}), \tag{103a}$$

$$\widetilde{X}^{\mathsf{st}}_{k+1} = \alpha_{t_{k+1}}\overline{Y}_{t_{k+1}} + \widehat{\sigma}_{t_{k+1}}\overline{Z}_{t_{k+1}}, \tag{103b}$$

$$\widetilde{X}^{\mathsf{sc}}_{k+1} = \widetilde{X}^{\mathsf{sc}}_{k} + \gamma_k \nabla_{\widetilde{X}^{\mathsf{sc}}_k} \log q_{\lambda_{k+1}}(\widetilde{X}^{\mathsf{c}}_k, \widetilde{X}^{\mathsf{sc}}_k, \widetilde{X}^{\mathsf{st}}_{k+1}) + \sqrt{2\gamma_k} Z^{\mathsf{sc}}_k, \tag{103c}$$

$$\widetilde{X}^{\mathsf{c}}_{k+1} = \widetilde{X}^{\mathsf{c}}_{k} + \gamma_k \nabla_{\widetilde{X}^{\mathsf{c}}_k} \log q_{\lambda_{k+1}}(\widetilde{X}^{\mathsf{c}}_k, \widetilde{X}^{\mathsf{sc}}_k, \widetilde{X}^{\mathsf{st}}_{k+1}) + \sqrt{2\gamma_k} Z^{\mathsf{c}}_k, \tag{103d}$$

where $q_{\lambda_k}(x^{\mathsf{c}}_k, x^{\mathsf{sc}}_k, x^{\mathsf{st}}_{k+1})$ is defined as

$$q_{\lambda_{k+1}}(x^{\mathsf{c}}_k, x^{\mathsf{sc}}_k, x^{\mathsf{st}}_k) \coloneqq p_{X^{\mathsf{sc}}_{t_{k+1}}, X^{\mathsf{c}}_{t_{k+1}} | X^{\mathsf{st}}_{t_{k+1}}, X^{\mathsf{sc},t_{k+1}}_{\tau_s}}(x^{\mathsf{c}}_k, x^{\mathsf{sc}}_k \mid x^{\mathsf{st}}_{k+1}, x^{\mathsf{sc},t_{k+1}}_{k_{\tau_s}}),$$

and $X^{\mathsf{sc},t_{k+1}}_{\tau_s} = \{v_s^\top X_{\tau_s}\}_{s \in \mathcal{S}^{\mathsf{c}}_{t_{k+1}}}$, $\overline{X}^{\mathsf{sc},t_{k+1}}_{k_{\tau_s}} = \{v_s^\top \overline{X}_{k_{\tau_s}}\}_{s \in \mathcal{S}^{\mathsf{c}}_{t_{k+1}}}$ By a similar argument with Lemma 1, we have

$$\mathsf{KL}(p_{\widetilde{X}_{0:M} | Y,Z} \| p_{\widehat{X}_{t_{0:M}} | Y,Z}) \to 0, \qquad \text{as} \quad \eta^3\delta^2 \to 0, \quad \text{and} \quad \eta \to \infty. \tag{104}$$

Moreover, the sequence defined in (29) satisfies $p_{\overline{X}_k | V^\top_{\mathcal{S}_{t_k}} \overline{X}_k, v_s^\top \overline{X}_{k_{\tau_s}}} = p_{X_{t_k} | V^\top_{\mathcal{S}_{t_k}} X_{t_k}, v_s^\top X_{\tau_s}}$. By a similar argument as in (31), it suffices to prove that for any y, z,

$$\mathsf{TV}\big(p_{\widetilde{X}_k | Y,Z}(\cdot \mid y, z), p_{\overline{X}_k | Y,Z}(\cdot \mid y, z)\big) \to 0.$$

By decomposition, we have

$$\begin{aligned}
\mathsf{TV}\big(p_{\widetilde{X}_k | Y,Z}(\cdot \mid y, z), p_{\overline{X}_k | Y,Z}(\cdot \mid y, z)\big) &\le \mathsf{TV}\big(p_{\widetilde{X}^{\mathsf{st}}_k, \widetilde{X}^{\mathsf{sc},t_k}_{k_{\tau_s}} | Y,Z}(\cdot \mid y, z), p_{\overline{X}^{\mathsf{st}}_k, \overline{X}^{\mathsf{sc},t_k}_{k_{\tau_s}} | Y,Z}(\cdot \mid y, z)\big) \\
&\quad + \mathsf{TV}\big(p_{\widetilde{X}_k | \widetilde{X}^{\mathsf{st}}_k, \widetilde{X}^{\mathsf{sc},t_k}_{k_{\tau_s}}, Y,Z}(\cdot \mid x^{\mathsf{st}}_k, x_\tau, y, z), p_{\overline{X}_k | \overline{X}^{\mathsf{st}}_k, \overline{X}^{\mathsf{sc},t_k}_{k_{\tau_s}}, Y,Z}(\cdot \mid x^{\mathsf{st}}_k, x_\tau, y, z)\big) \\
&= \mathsf{TV}\big(p_{\widetilde{X}_k | \widetilde{X}^{\mathsf{st}}_k, \widetilde{X}^{\mathsf{sc},t_k}_{k_{\tau_s}}, Y,Z}(\cdot \mid x^{\mathsf{st}}_k, x_\tau, y, z), p_{X_{t_k} | X^{\mathsf{st}}_{t_k}, X^{\mathsf{sc},t_k}_{\tau_s}}(\cdot \mid x^{\mathsf{st}}_k, x_\tau)\big),
\end{aligned}$$

where $x^{\mathsf{st}}_k = \alpha_{t_k} y + \widehat{\sigma}_{t_k} z$, and the last equation uses $p_{\overline{X}_k | \overline{X}^{\mathsf{st}}_k, \overline{X}^{\mathsf{sc},t_k}_{k_{\tau_s}}, Y,Z}(\cdot \mid x^{\mathsf{st}}_k, x_\tau, y, z) = p_{X_{t_k} | X^{\mathsf{st}}_{t_k}, X^{\mathsf{sc},t_k}_{\tau_s}}(\cdot \mid x^{\mathsf{st}}_k, x_\tau)$. The remaining proof is similar to the proof of Lemma 2. Specificially,

we define a new random variable $\widetilde{X}_k^{\mathsf{stc}} \coloneqq [\widetilde{X}_k^{\mathsf{sc}}, \widetilde{X}_k^{\mathsf{c}}]$ and $X_{t_k}^{\mathsf{stc}} \coloneqq [X_{t_k}^{\mathsf{sc}}, X_{t_k}^{\mathsf{c}}]$. Then the update rule in (103c) - (103d) can be integrated as

$$\widetilde{X}_{k+1}^{\mathsf{stc}} = \widetilde{X}_k^{\mathsf{stc}} + \gamma_k \nabla_{\widetilde{X}_k^{\mathsf{stc}}} \log q_{\lambda_{k+1}}(\widetilde{X}_k^{\mathsf{c}}, \widetilde{X}_k^{\mathsf{sc}}, \widetilde{X}_{k+1}^{\mathsf{st}}) + \sqrt{2\gamma_k}[Z_k^{\mathsf{sc}}, Z_k^{\mathsf{c}}].$$

For convenience, denote distributions

$$p_k^{\mathsf{stc}}(x_{t_k}^{\mathsf{stc}}) \coloneqq p_{X_{t_k}^{\mathsf{stc}} \mid X_{t_k}^{\mathsf{st}}}(x_{t_k}^{\mathsf{stc}} \mid x_{t_k}^{\mathsf{st}}), \qquad \psi_k(x_{t_k}^{\mathsf{stc}}) \coloneqq p_{\widetilde{X}_k^{\mathsf{stc}} \mid \widetilde{X}_k^{\mathsf{st}}, \widetilde{X}_{k_{\tau_s}}^{\mathsf{sc},t_k}, Y, Z}(x_{t_k}^{\mathsf{stc}} \mid x_k^{\mathsf{st}}, x_\tau, y, z)$$

and denote P_{γ_k} as the transition probability of the following SDE from t to $t+\gamma_k$:

$$\mathrm{d}X(t) = \nabla \log p_{k+1}^{\mathsf{stc}}(X(t))\mathrm{d}t + \sqrt{2}\mathrm{d}W_t, \tag{105}$$

where W_t denotes the Brownian motion. According to the analysis of Langevin, we have

$$\begin{aligned}
\mathsf{KL}(\psi_{k+1} \| p_{k+1}^{\mathsf{stc}}) &\le \mathsf{KL}(P_{\gamma_k}\psi_k \| p_{k+1}^{\mathsf{stc}}) + \mathsf{KL}(\psi_{k+1} \| p_{k+1}^{\mathsf{stc}}) - \mathsf{KL}(P_{\gamma_k}\psi_k \| p_{k+1}^{\mathsf{stc}}) \\
&\le \mathrm{e}^{-2\kappa\gamma_k}\mathsf{KL}(\psi_k \| p_k^{\mathsf{stc}}) + \mathrm{e}^{-2\kappa\gamma_k}\left(\mathsf{KL}(\psi_k \| p_{k+1}^{\mathsf{stc}}) - \mathsf{KL}(\psi_k \| p_k^{\mathsf{stc}})\right) \\
&\quad + \mathsf{KL}(\psi_{k+1} \| p_{k+1}^{\mathsf{stc}}) - \mathsf{KL}(P_{\gamma_k}\psi_k \| p_{k+1}^{\mathsf{stc}})
\end{aligned}$$

where $\kappa < \infty$ is a constant. We claim that

$$\mathsf{KL}(\psi_k \| p_{k+1}^{\mathsf{stc}}) - \mathsf{KL}(\psi_k \| p_k^{\mathsf{stc}}) = O(\delta_{k+1}). \tag{106}$$

By similar argument as in (69) - (72), we have

$$\mathsf{KL}(\psi_{k+1} \| p_{k+1}^{\mathsf{stc}}) - \mathsf{KL}(P_{\gamma_k}\psi_k \| p_{k+1}^{\mathsf{stc}}) = O(\gamma_k^2).$$

Therefore, we have

$$\mathsf{KL}(\psi_{k+1} \| p_{k+1}^{\mathsf{stc}}) \le \mathrm{e}^{-2\kappa\gamma_k}\mathsf{KL}(\psi_k \| p_k^{\mathsf{stc}}) + O(\delta_{k+1} + \gamma_k^2) \quad \lesssim \frac{1}{\eta} + \max_k \gamma_k \to 0, \qquad \text{as} \quad \eta \to \infty, \delta\eta \to 0.$$

Finally applying Pinsker's inequality, we complete the proof.

Proof of (106). According to the definition of KL divergence, we have

$$\mathsf{KL}(\psi_k \| p_{k+1}^{\mathsf{stc}}) - \mathsf{KL}(\psi_k \| p_k^{\mathsf{stc}}) = \int \log \frac{p_k^{\mathsf{stc}}(x)}{p_{k+1}^{\mathsf{stc}}(x)} \psi_{k+1}(x)\mathrm{d}x.$$

It suffices to prove that

$$\log p_k^{\mathsf{stc}}(x) - \log p_{k+1}^{\mathsf{stc}}(x) = O(\delta_{k+1}),$$

which is implied by

$$\left|\frac{\partial}{\partial \alpha_k} \log p_k^{\mathsf{stc}}(x)\right| = O(\|x\|_2 + 1), \qquad \left|\frac{\partial}{\partial \sigma_k} \log p_k^{\mathsf{stc}}(x)\right| = O(\|x\|_2 + 1),$$

$$\|\nabla_{x_{t_k}^{\mathsf{st}}} \log p_{X_{t_k}^{\mathsf{stc}} \mid X_{t_k}^{\mathsf{st}}}(x_{t_k}^{\mathsf{stc}} \mid x_{t_k}^{\mathsf{st}})\|_2 = O(\|x_{t_k}^{\mathsf{st}}\|_2 + 1).$$

To this end, by the definition of $p_k^{\mathsf{stc}}(x)$, we have

$$\begin{aligned}
\left|\frac{\partial}{\partial \alpha_{t_k}} \log p_k^{\mathsf{stc}}(x)\right| &\le \left|\frac{\partial}{\partial \alpha_k} \log p_{X_{t_k}^{\mathsf{stc}}, X_{t_k}^{\mathsf{st}}}(x^{\mathsf{stc}}, x^{\mathsf{st}})\right| + \left|\frac{\partial}{\partial \alpha_k} \log p_{X_{t_k}^{\mathsf{st}}}(x^{\mathsf{st}})\right| \\
&= \frac{\left|\mathbb{E}[(x^{\mathsf{stc}} - \alpha_{t_k} X_0^{\mathsf{stc}})^\top X_0^{\mathsf{stc}} \mid X_{t_k} = x]\right|}{\sigma_{t_k}^2} + \frac{\left|\mathbb{E}[(x^{\mathsf{st}} - \alpha_{t_k} X_0^{\mathsf{st}})^\top X_0^{\mathsf{st}} \mid X_{t_k} = x]\right|}{\sigma_{t_k}^2} \\
&\quad + \frac{\left|\mathbb{E}[(x^{\mathsf{st}} - \alpha_{t_k} X_0^{\mathsf{st}})^\top X_0^{\mathsf{st}} \mid X_{t_k}^{\mathsf{st}} = x^{\mathsf{st}}]\right|}{\sigma_{t_k}^2} \\
\left|\frac{\partial}{\partial \sigma_{t_k}} \log p_k^{\mathsf{stc}}(x)\right| &\le \frac{\mathbb{E}[\|x^{\mathsf{stc}} - \alpha_{t_k} X_0^{\mathsf{stc}}\|_2^2 \mid X_{t_k} = x]}{\sigma_{t_k}^3} + \frac{\mathbb{E}[\|x^{\mathsf{st}} - \alpha_{t_k} X_0^{\mathsf{st}}\|_2^2 \mid X_{t_k} = x]}{\sigma_{t_k}^3} \\
&\quad + \frac{\mathbb{E}[\|x^{\mathsf{st}} - \alpha_{t_k} X_0^{\mathsf{st}}\|_2^2 \mid X_{t_k}^{\mathsf{st}} = x^{\mathsf{st}}]}{\sigma_{t_k}^3} - \frac{|\mathcal{S}_t^{\mathsf{c}} \cup \mathcal{S}^{\mathsf{c}}|}{\sigma_{t_k}}, \\
\|\nabla_{x_{t_k}^{\mathsf{st}}} \log p_{X_{t_k}^{\mathsf{stc}} \mid X_{t_k}^{\mathsf{st}}}(x_{t_k}^{\mathsf{stc}} \mid x_{t_k}^{\mathsf{st}})\|_2 &\le \|\nabla_{x_{t_k}^{\mathsf{st}}} \log p_{X_{t_k}^{\mathsf{stc}}, X_{t_k}^{\mathsf{st}}}(x_{t_k}^{\mathsf{stc}}, x_{t_k}^{\mathsf{st}})\|_2 + \|\nabla_{x_{t_k}^{\mathsf{st}}} \log p_{X_{t_k}^{\mathsf{st}}}(x_{t_k}^{\mathsf{st}})\|_2 \\
&\le \frac{\mathbb{E}[\|x^{\mathsf{st}} - \alpha_{t_k} X_0^{\mathsf{st}}\|_2 \mid X_{t_k} = x]}{\sigma_{t_k}^2} + \frac{\mathbb{E}[\|x^{\mathsf{st}} - \alpha_{t_k} X_0^{\mathsf{st}}\|_2 \mid X_{t_k}^{\mathsf{st}} = x^{\mathsf{st}}]}{\sigma_{t_k}^2}.
\end{aligned}$$

Therefore, we complete the proof.

D EXTENSION TO GENERAL A

For $V_{\mathcal{S}^c}^\top x_{t_i}, V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top x_{t_i}$, we aim to approximately generate them according to the following conditional distribution:

$$V_{\mathcal{S}^c}^\top \widehat{X}_{t_i}, V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top \widehat{X}_{t_i} \mid V_{\mathcal{S}_{t_i}^{\text{meas}}}^\top \widehat{X}_{t_i}, V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top \widehat{X}_\tau \sim p_{V_{\mathcal{S}^c}^\top X_{t_i}, V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top X_{t_i} \mid V_{\mathcal{S}_{t_i}^{\text{meas}}}^\top X_{t_i}, V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top X_\tau}$$

Towards this, we need to calculate the score function, i.e., the log gradient of the above conditional density with respect to $V_{\mathcal{S}^c}^\top x_{t_i}$ and $V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top x_{t_i}$. In particular, the gradient with respect to $V_{\mathcal{S}^c}^\top x_{t_i}$ is given by

$$\begin{aligned}
&\nabla_{V_{\mathcal{S}^c}^\top x_{t_i}} \log p_{V_{\mathcal{S}^c}^\top X_{t_i}, V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top X_{t_i} \mid V_{\mathcal{S}_{t_i}^{\text{meas}}}^\top X_{t_i}, V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top X_\tau}(V_{\mathcal{S}^c}^\top x_{t_i}, V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top x_{t_i} \mid V_{\mathcal{S}_{t_i}^{\text{meas}}}^\top x_{t_i}, V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top x_\tau) \\
&\quad = \nabla_{V_{\mathcal{S}^c}^\top x_{t_i}} \log p_{V_{\mathcal{S}^c}^\top X_{t_i}, V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top X_{t_i}, V_{\mathcal{S}_{t_i}^{\text{meas}}}^\top X_{t_i}}(V_{\mathcal{S}^c}^\top x_{t_i}, V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top x_{t_i}, V_{\mathcal{S}_{t_i}^{\text{meas}}}^\top x_{t_i}) = -\frac{1}{\sigma_{t_i}} V_{\mathcal{S}^c}^\top \epsilon_{t_i}(x_{t_i}),
\end{aligned}$$

which is the same as previously used score function defined in (12). the gradient with respect to $V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top x_{t_i}$ is given by

$$\begin{aligned}
&\nabla_{V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top x_{t_i}} \log p_{V_{\mathcal{S}^c}^\top X_{t_i}, V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top X_{t_i} \mid V_{\mathcal{S}_{t_i}^{\text{meas}}}^\top X_{t_i}, V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top X_\tau}(V_{\mathcal{S}^c}^\top x_{t_i}, V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top x_{t_i} \mid V_{\mathcal{S}_{t_i}^{\text{meas}}}^\top x_{t_i}, V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top x_\tau) \\
&\quad = \nabla_{V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top x_{t_i}} \log p_{V_{\mathcal{S}^c}^\top X_{t_i}, V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top X_{t_i}, V_{\mathcal{S}_{t_i}^{\text{meas}}}^\top X_{t_i}}(V_{\mathcal{S}^c}^\top x_{t_i}, V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top x_{t_i}, V_{\mathcal{S}_{t_i}^{\text{meas}}}^\top x_{t_i}) \\
&\qquad + \nabla_{V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top x_{t_i}} \log p_{V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top X_\tau \mid V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top X_{t_i}}(V_{\mathcal{S}_{t_i}^{\text{meas}}}^\top x_\tau, V_{\mathcal{S}_{t_i}^{\text{meas}}}^\top x_{t_i}) \\
&\quad = -\frac{1}{\sigma_{t_i}} V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top \epsilon_{t_i}(x_{t_i}) + \frac{1}{\sigma_{t_i}^2(\mathrm{e}^{2(\lambda_{t_i}-\lambda_\tau)}-1)} \left(\frac{\alpha_{t_i}}{\alpha_\tau} V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top x_\tau - V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top x_{t_i} \right).
\end{aligned}$$

Compared to previously used score function, an additional term is introduced to capture the condition information between $V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top \widehat{X}_{t_i}$ and $V_{\mathcal{S}_{t_i}^{\text{prior}}}^\top \widehat{X}_\tau$. Inserting this score function into the DDIM update rule yields the update formula in (24).