A. Asymptotic Accuracy of LD-SMC

Here we provide a proof that $\mathbf{z}_0$ samples can be made arbitrarily accurate according to our model design. Our proof is composed of three parts, all of which have been previously established in the literature. We will restate them here and accommodate them to our setting. The three parts are, (1) augmenting the model with auxiliary random variables, ergo performing a completion of the desired marginal density $p_\theta(\mathbf{z}_0|\mathbf{y}_0)$ (Definition 10.3 in (Robert & Casella, 1999)), (2) Proving asymptotic accuracy of the SMC procedure following a similar lines of Theorem 2 in (Wu et al., 2024), (3) Showing sufficient conditions that the Markov chain generated by the Gibbs sampling procedure is ergodic and hence $p_\theta(\mathbf{z}_0|\mathbf{y}_0)$ is the limiting distribution from which $\mathbf{z}_0$'s are sampled (Theorem 10.6 in (Robert & Casella, 1999)).

First of all, here we provide a concise version of our Gibbs sampler presented in Algorithm 1. We denote by j the index of the Gibbs iterations:

1. Initialize $\mathbf{z}_{0:T}^0$
2. For $j = 0, ..., J-1$:
   (a) Sample, $\mathbf{y}_{1:T}^{j+1} \sim p_\theta(\mathbf{y}_{1:T}|\mathbf{z}_{0:T}^j, \mathbf{y}_0)$.
   (b) Sample, $\mathbf{z}_{0:T}^{j+1} \sim p_\theta(\mathbf{z}_{0:T}|\mathbf{y}_{1:T}^{j+1}, \mathbf{y}_0)$ using SMC.

Definition A.1. (Robert & Casella, 1999). Given a probability density $p_\theta(\mathbf{z}_0|\mathbf{y}_0)$, a density g that satisfies $\int g(\mathbf{z}_0, \mathbf{z}_{1:T}, \mathbf{y}_{1:T}|\mathbf{y}_0) d\mathbf{z}_{1:T} d\mathbf{y}_{1:T}$ is called a completion of $p_\theta(\mathbf{z}_0|\mathbf{y}_0)$.

We denote by $(\mathbf{z}_{0:T}^j, \mathbf{y}_{1:T}^j)$ the Markov chain generated by our proposed Gibbs sampler. Similarly, denote by $(\mathbf{z}_0^j)$ the corresponding subchain.

Theorem A.2. *(Robert & Casella, 1999). For the Gibbs sampler described in Section 4.2, if $(\mathbf{z}_{0:T}^j, \mathbf{y}_{1:T}^j)$ is ergodic, then the distribution g is a stationary distribution for it and $p_\theta(\mathbf{z}_0|\mathbf{y}_0)$ is the limiting distribution of the subchain $(\mathbf{z}_0^j)$.*

Proof. The support of g is $\mathbb{R}^d$, with $d = dim(\mathbf{z}_{0:T}, \mathbf{y}_{1:T})$, and hence is connected. Also, since each conditional distribution in the Gibbs sampling process is a Gaussian or a multiplication of Gaussian densities, they are all strictly positive. Following Lemma 10.11 in (Robert & Casella, 1999), $(\mathbf{z}_{0:T}^j, \mathbf{y}_{1:T}^j)$ is irreducible and aperiodic, i.e., it is ergodic. □

Specifically, Theorem A.2 shows that in order to obtain samples from the correct marginal distribution $p_\theta(\mathbf{z}_0|\mathbf{y}_0)$, we need to sample $\mathbf{y}_{1:T}$ and $\mathbf{z}_{0:T}$ from the correct conditional distribution according to the model at each iteration. Sampling $\mathbf{y}_{1:T}$ variables is straightforward as it requires sampling from independent Gaussian distributions, that is, $p_\theta(\mathbf{y}_t|\mathbf{z}_t) = \mathcal{N}(\mathbf{y}_t|\mathcal{A}(\mathcal{D}(\mathbf{z}_t)), \tau^2\mathbf{I})$ for all $t > 0$. On the other hand, sampling $\mathbf{z}_{0:T}$ requires sampling using an SMC procedure. As we will show next in the large particle limit samples from $p_\theta(\mathbf{z}_{0:T}|\mathbf{y}_{0:T})$ are accurate.

To prove the asymptotic accuracy of the SMC procedure in LD-SMC, we adopt the formulation of (Wu et al., 2024). In what follows, to prevent cluttered notation, we drop the index of the Gibbs iteration and note that all quantities are those of a specific iteration. Specifically, here we first reiterate the following 3 important quantities, the prior distribution $p_\theta(\mathbf{z}_{0:T})$, the proposal distributions $\pi_t(\mathbf{z}_t|\mathbf{z}_{t+1})$, and the weighting functions $w_t(\mathbf{z}_t, \mathbf{z}_{t+1})$ and then proceed to the proof.

Prior distribution. Let $p_\theta(\mathbf{z}_{0:T})$ denote the diffusion generative model defined according to Eq. 1. Then, the Markovian structure of the prior diffusion model goes as follows,

$$\begin{cases} p_\theta(\mathbf{z}_t|\mathbf{z}_{t+1}) = \mathcal{N}(\boldsymbol{\mu}_\theta(\mathbf{z}_{t+1}, t+1), \sigma_{t+1}^2\mathbf{I}) & \\ \qquad = \mathcal{N}(\mathbf{z}_t|\sqrt{\bar{\alpha}_t}\left(\frac{\mathbf{z}_{t+1} - \sqrt{1-\bar{\alpha}_{t+1}}\cdot\boldsymbol{\epsilon}_\theta(\mathbf{z}_{t+1}, t+1)}{\sqrt{\bar{\alpha}_{t+1}}}\right) + \sqrt{1-\bar{\alpha}_t - \sigma_{t+1}^2}\cdot\boldsymbol{\epsilon}_\theta(\mathbf{z}_{t+1}, t+1), \sigma_{t+1}^2\mathbf{I}), & 1 < t < T, \\ p_\theta(\mathbf{z}_T) = \mathcal{N}(0, \mathbf{I}) & t = T. \end{cases}$$

Proposal distributions. Denote the proposal distribution for timestep $t < T$ as $\pi_t(\mathbf{z}_t|\mathbf{z}_{t+1}) = \mathcal{N}(\mathbf{m}_t, \mathbf{S}_t)$, with parameters

$$\begin{aligned} \mathbf{S}_t &= \tilde{\sigma}_{t+1}^2\mathbf{I} \\ \mathbf{m}_t &= \boldsymbol{\mu}_\theta(\mathbf{z}_{t+1}, t+1) - (\gamma_t\nabla_{\mathbf{z}_{t+1}}\log\ \bar{p}_\theta(\mathbf{y}_0|\mathbf{z}_{t+1}) + \lambda_t\nabla_{\boldsymbol{\mu}_\theta(\mathbf{z}_{t+1}, t+1)}\log\ q_\theta(\mathbf{y}_t|\mathbf{z}_{t+1})). \end{aligned} \tag{7}$$

Where $\bar{p}_\theta(\mathbf{y}_0|\mathbf{z}_{t+1}) = \mathcal{N}(\mathbf{y}_0|\mathcal{A}(\mathcal{D}(\bar{\mathbf{z}}_0(\mathbf{z}_{t+1}))), (1-\bar{\alpha}_{t+1})\mathbf{I})$, and $q_\theta(\mathbf{y}_t|\mathbf{z}_{t+1}) = \mathcal{N}(\mathbf{y}_t|\mathcal{A}(\mathcal{D}(\boldsymbol{\mu}_\theta(\mathbf{z}_{t+1}, t+1))), \tau^2\mathbf{I})$, and γ_t, λ_t are finite scaling coefficients. The distribution of the proposal at time T, $\pi_T(\mathbf{z}_T)$, from which the initial sample is taken, is set to the diffusion prior distribution $p_\theta(\mathbf{z}_T) = \mathcal{N}(0, \mathbf{I})$.

Weighting functions. The unnormalized weighting functions for all timesteps t are summarized as follows,

$$\begin{cases} \tilde{w}_T(\mathbf{z}_T) = p_\theta(\mathbf{y}_T|\mathbf{z}_T)\bar{p}_\theta(\mathbf{y}_0|\mathbf{z}_T), \\ \tilde{w}_t(\mathbf{z}_t, \mathbf{z}_{t+1}) = p_\theta(\mathbf{y}_t|\mathbf{z}_t)\bar{p}_\theta(\mathbf{y}_0|\mathbf{z}_t)p_\theta(\mathbf{z}_t|\mathbf{z}_{t+1})/(\bar{p}_\theta(\mathbf{y}_0|\mathbf{z}_{t+1})\pi_t(\mathbf{z}_t|\mathbf{z}_{t+1})), \\ \tilde{w}_0(\mathbf{z}_0, \mathbf{z}_1) = p_\theta(\mathbf{y}_0|\mathbf{z}_0)p_\theta(\mathbf{z}_0|\mathbf{z}_1)/\bar{p}_\theta(\mathbf{y}_0|\mathbf{z}_1)\pi_0(\mathbf{z}_0|\mathbf{z}_1), \end{cases}$$

where $p_\theta(\mathbf{y}_t|\mathbf{z}_t) = \mathcal{N}(\mathbf{y}_t|\mathcal{A}(\mathcal{D}(\mathbf{z}_t)), \tau^2\mathbf{I})$.

In SMC the proposal distributions and weighting functions define a sequence of intermediate target distributions $\{\nu_t\}_{t=0}^T$ defined as follows,

$$\nu_t(\mathbf{z}_{t:T}) = \frac{1}{\mathcal{L}_t}\left(\pi_T(\mathbf{z}_T)\prod_{t'=t}^{T}\pi_{t'}(\mathbf{z}_{t'}|\mathbf{z}_{t'+1})\right)\left(\tilde{w}_T(\mathbf{z}_T)\prod_{t'=t}^{T}\tilde{w}_{t'}(\mathbf{z}_{t'}, \mathbf{z}_{t'+1})\right). \tag{8}$$

Importantly, if for given weighting functions and proposal distributions the final target distribution ν_0 coincides with the desired posterior distribution $p(\mathbf{z}_{0:T}|\mathbf{y}_{0:T})$, then samples that are approximately distributed according to $p(\mathbf{z}_{0:T}|\mathbf{y}_{0:T})$ can be obtained (Doucet et al., 2001a; Naesseth et al., 2019). The approximation becomes accurate in the limit of large number of particles. For $t = 0$, plugging in Eq. 8 the proposal distributions and weighting functions and observing that all appearances of the proposal distributions besides that of time T cancel out, we obtain:

$$\begin{aligned} \nu_0(\mathbf{z}_{0:T}) &= \frac{1}{\mathcal{L}_0}p_\theta(\mathbf{z}_T)p_\theta(\mathbf{y}_T|\mathbf{z}_T)\bar{p}_\theta(\mathbf{y}_0|\mathbf{z}_T)\left[\prod_{t=1}^{T-1}\frac{p_\theta(\mathbf{y}_t|\mathbf{z}_t)\bar{p}_\theta(\mathbf{y}_0|\mathbf{z}_t)p_\theta(\mathbf{z}_t|\mathbf{z}_{t+1})}{\bar{p}_\theta(\mathbf{y}_0|\mathbf{z}_{t+1})}\right]\frac{p_\theta(\mathbf{y}_0|\mathbf{z}_0)}{\bar{p}_\theta(\mathbf{y}_0|\mathbf{z}_1)}p_\theta(\mathbf{z}_0|\mathbf{z}_1) \\ &= \frac{1}{\mathcal{L}_0}\prod_{t=0}^{T}p_\theta(\mathbf{y}_t|\mathbf{z}_t)\prod_{t=0}^{T-1}p_\theta(\mathbf{z}_t|\mathbf{z}_{t+1})p_\theta(\mathbf{z}_T) = \frac{1}{\mathcal{L}_0}p_\theta(\mathbf{y}_{0:T}|\mathbf{z}_{0:T})p_\theta(\mathbf{z}_{0:T}) = p_\theta(\mathbf{z}_{0:T}|\mathbf{y}_{0:T}). \end{aligned} \tag{9}$$

Where in the second equality $\bar{p}_\theta$ terms cancel out and in the last equality we applied Bayes rule.

Now that we established necessary conditions to sample from $p_\theta(\mathbf{z}_{0:T}|\mathbf{y}_{0:T})$, the following theorem from (Chopin et al., 2020) characterizes the conditions under which SMC algorithms converge. We adopt the formulation of (Wu et al., 2024) and present it here, adapted to our case.

Theorem A.3. *((Chopin et al., 2020) – Proposition 11.4). Let $\{\mathbf{z}_{0:T}^{(i)}, w_0^{(i)}\}$ be the sequence of particles and final weights returned by the last iteration of the SMC algorithm with N particles and using multinomial resampling. If the weight functions of all time steps $w_t^{(i)}$ are positive and bounded, then for every ν_0-measurable function ϕ of $\mathbf{z}_{0:T}^{(i)}$,*

$$\lim_{N\to\infty}\sum_{i=1}^{N}w_0^{(i)}\phi(\mathbf{z}_{0:T}^{(i)}) = \int\phi(\mathbf{z}_{0:T})\nu_0(\mathbf{z}_{0:T})d\mathbf{z}_{0:T}$$

with probability one.

Specifically, taking $\phi(\mathbf{z}_{0:T}) = \mathbb{I}[\mathbf{z}_{0:T} \in E]$ for any ν_0-measurable set E implies the convergence of $\mathbb{P}_N(E) = \sum_{i=1}^N w_0^{(i)}\delta_{\mathbf{z}_{0:T}^{(i)}}(E)$. Here, $\delta_{\mathbf{z}}$ is the Dirac measure defined for a given point $\mathbf{z}$ and a ν_0-measurable set E. The following proposition characterizes the conditions under which Theorem A.3 applies in our case.

Proposition A.4. *Let $\mathbb{P}_N(E) = \sum_{i=1}^N w_0^{(i)}\delta_{\mathbf{z}_{0:T}^{(i)}}(E)$ be a discrete distribution over particles with $\{(\mathbf{z}_{0:T}^{(i)}, w_0^{(i)})\}_{i=1}^N$ returned by the SMC procedure in Section 4.2.2 with N particles. Assume for all t:*

(a) The likelihood functions $p_\theta(\mathbf{y}_t|\mathbf{z}_t) = \mathcal{N}(\mathbf{y}_t|\mathcal{A}(\mathcal{D}(\mathbf{z}_t)), \tau^2\mathbf{I})$, the ratios $\bar{p}_\theta(\mathbf{y}_0|\mathbf{z}_t)/\bar{p}_\theta(\mathbf{y}_0|\mathbf{z}_{t+1})$, and $\bar{p}_\theta(\mathbf{y}_0|\mathbf{z}_T)$, are all positive and bounded.

(b) $\log\bar{p}_\theta(\mathbf{y}_0|\mathbf{z}_t)$ is continuous and has bounded gradients in $\mathbf{z}_t$.

(c) $\log q_\theta(\mathbf{y}_t|\mathbf{z}_{t+1})$ is continuous and has bounded gradients in $\boldsymbol{\mu}_\theta(\mathbf{z}_{t+1}, t+1)$.

(d) The proposal variance is larger than the prior diffusion model variance, namely $\tilde{\sigma}_t^2 > \sigma_t^2$.

Then $\mathbb{P}_N$ converges setwise to $p_\theta(\mathbf{z}_{0:T}|\mathbf{y}_{0:T})$ with probability one, that is for every set E, $\lim_{N\to\infty}\mathbb{P}_N(E) = \int_E p_\theta(\mathbf{z}_{0:T}|\mathbf{y}_{0:T})d\mathbf{z}_{0:T}$.

Note that here, unlike (Wu et al., 2024), we do not need to assume that $\bar{p}_\theta(\mathbf{y}_0|\mathbf{z}_0) = p_\theta(\mathbf{y}_0|\mathbf{z}_0)$ due to our revised posterior distribution. Here, however, we add assumption (c). Assumptions (b) and (c) are the strongest assumptions as they may not apply even for linear transformations. But, for sufficiently smooth decoder and diffusion model in the input ($\mathbf{z}_t$ or $\boldsymbol{\mu}_\theta(\mathbf{z}_{t+1}, t+1)$) with uniformly bounded gradients, the assumptions will hold.

Proof. To prove the statement we need to show that (1) the target distribution of the SMC procedure at time $t = 0$ is $p_\theta(\mathbf{z}_{0:T}|\mathbf{y}_{0:T})$, and (2) the weighting functions $w_t := w_t(\mathbf{z}_t, \mathbf{z}_{t+1})$ of all timesteps are all positive and bounded. The first condition was already established in Eq. 9, hence we proceed to the second condition.

Since all the weights are defined through multiplications of density functions they are strictly positive. To show that they are bounded, we first note that by Assumption (a) $w_T(\mathbf{z}_T)$ is bounded. To show that the intermediate weights are bounded it is enough to consider the log transformation of the unnormalized weights,

$$\log\ \tilde{w}_t(\mathbf{z}_t, \mathbf{z}_{t+1}) = \begin{cases} \log\ p_\theta(\mathbf{y}_t|\mathbf{z}_t) + \log\ \frac{\bar{p}_\theta(\mathbf{y}_0|\mathbf{z}_t)}{\bar{p}_\theta(\mathbf{y}_0|\mathbf{z}_{t+1})} + \log\ \frac{p_\theta(\mathbf{z}_t|\mathbf{z}_{t+1})}{\pi_t(\mathbf{z}_t|\mathbf{z}_{t+1})}, & 0 < t < T, \\ \log\ p_\theta(\mathbf{y}_0|\mathbf{z}_0) + \log\ \frac{p_\theta(\mathbf{z}_0|\mathbf{z}_1)}{\pi_0(\mathbf{z}_0|\mathbf{z}_1)}, & t = 0. \end{cases}$$

By assumption (a), $\log\ p_\theta(\mathbf{y}_t|\mathbf{z}_t)$ and $\log\ \bar{p}_\theta(\mathbf{y}_0|\mathbf{z}_t)/\bar{p}_\theta(\mathbf{y}_0|\mathbf{z}_{t+1})$ are bounded for all t. We will show next that the $\log\ p_\theta(\mathbf{z}_t|\mathbf{z}_{t+1})/\pi_t(\mathbf{z}_t|\mathbf{z}_{t+1})$ terms are all bounded based on assumptions (b) - (d). We denote by $\boldsymbol{\mu}_t := \boldsymbol{\mu}_\theta(\mathbf{z}_{t+1}, t+1)$,

$$\begin{aligned}
\log\ \frac{p_\theta(\mathbf{z}_t|\mathbf{z}_{t+1})}{\pi_t(\mathbf{z}_t|\mathbf{z}_{t+1})} &= \log\ \frac{|2\pi\sigma_{t+1}^2\mathbf{I}|^{-0.5}\exp\{-(2\sigma_{t+1}^2)^{-1}||\mathbf{z}_t - \boldsymbol{\mu}_t||_2^2\}}{|2\pi\tilde{\sigma}_{t+1}^2\mathbf{I}|^{-0.5}\exp\{-(2\tilde{\sigma}_{t+1}^2)^{-1}||\mathbf{z}_t - \mathbf{m}_t||_2^2\}} \\
&\overset{c}{=} -\frac{1}{2}[\sigma_{t+1}^{-2}||\mathbf{z}_t - \boldsymbol{\mu}_t||_2^2 - \tilde{\sigma}_{t+1}^{-2}||\mathbf{z}_t - \mathbf{m}_t||_2^2] \\
&\text{Expanding, } ||\mathbf{z}_t - \mathbf{m}_t \pm \boldsymbol{\mu}_t||_2^2 = ||\mathbf{z}_t - \boldsymbol{\mu}_t||_2^2 + 2\langle\mathbf{z}_t - \boldsymbol{\mu}_t, \boldsymbol{\mu}_t - \mathbf{m}_t\rangle + ||\boldsymbol{\mu}_t - \mathbf{m}_t||_2^2 \\
&= -\frac{1}{2}[(\sigma_{t+1}^{-2} - \tilde{\sigma}_{t+1}^{-2})||\mathbf{z}_t - \boldsymbol{\mu}_t||_2^2 - 2\tilde{\sigma}_{t+1}^{-2}\langle\mathbf{z}_t - \boldsymbol{\mu}_t, \boldsymbol{\mu}_t - \mathbf{m}_t\rangle - \tilde{\sigma}_{t+1}^{-2}||\boldsymbol{\mu}_t - \mathbf{m}_t||_2^2] \\
&\text{Notice, } ||\boldsymbol{\mu}_t - \mathbf{m}_t||_2^2 = ||\gamma_t\nabla_{\mathbf{z}_{t+1}}\log\ \bar{p}_\theta(\mathbf{y}_0|\mathbf{z}_{t+1}) + \lambda_t\nabla_{\boldsymbol{\mu}_t}\log\ q_\theta(\mathbf{y}_t|\mathbf{z}_{t+1})||_2^2 < \infty \\
&\ \text{by assumptions (b) and (c)} \\
&\overset{c}{=} -\frac{1}{2}[(\sigma_{t+1}^{-2} - \tilde{\sigma}_{t+1}^{-2})||\mathbf{z}_t - \boldsymbol{\mu}_t||_2^2 - 2\tilde{\sigma}_{t+1}^{-2}\langle\mathbf{z}_t - \boldsymbol{\mu}_t, \boldsymbol{\mu}_t - \mathbf{m}_t\rangle] \\
&\text{Apply Cauchy-Schwartz inequality,} \\
&\leq -\frac{1}{2}(\sigma_{t+1}^{-2} - \tilde{\sigma}_{t+1}^{-2})||\mathbf{z}_t - \boldsymbol{\mu}_t||_2^2 + \tilde{\sigma}_{t+1}^{-2}||\boldsymbol{\mu}_t - \mathbf{m}_t||_2||\mathbf{z}_t - \boldsymbol{\mu}_t||_2] \\
&\text{Apply the inequality } -\frac{a}{2}x + bx \leq \frac{b^2}{2a}\text{for } a = \sigma_{t+1}^{-2} - \tilde{\sigma}_{t+1}^{-2} > 0 \text{ by assumption (d)} \\
&\leq \frac{\tilde{\sigma}_{t+1}^{-4}||\boldsymbol{\mu}_t - \mathbf{m}_t||_2^2}{2(\sigma_{t+1}^{-2} - \tilde{\sigma}_{t+1}^{-2})} < \infty.
\end{aligned}$$

Where the last inequality is again due to assumptions (b) and (c). Here, $\overset{c}{=}$ denotes an equality up to a constant. □