STOCHASTIC LOCALIZATION AND SAMPLING VIA TIME REVERSAL

ARIF KEREM DAYI

ABSTRACT. In this expository project, we establish formal connections between sampling, stochastic localization, and filtering. In particular, we consider a special case of the localization process used in [Eld13]. We first establish formal connections of this process to sampling via denoising diffusions, specifically the *variance exploding* formulation. Then, we derive the self-governing dynamics of the localization process using time-reversals of SDEs. Afterwards, we give an alternative characterization of the dynamics of localization via *filtering*, without referring to time reversals. Finally, we formally connect our results to the stochastic localization framework in [Eld13] as a measure valued martingale that converges to a random dirac distribution.

Our exposition clarifies the formal connections between sampling, localization, and filtering. Furthermore, we aim to present complete proofs and derivations as possible, commenting on connections to practical applications in diffusion modeling in generative AI.

1. INTRODUCTION

Let μ be a probability measure over $\mathbb{R}^d$, where d is potentially large. How can we draw samples from μ? This has been one of the central questions in the modern study of generative modeling, where diffusion modeling and SDE time reversals have driven significant progress in our ability to sample from high-dimensional multimodal distributions. These methods are able to go beyond the limitations of log-concave sampling[1] which significantly restrict the multimodality of μ. The core intuition is quite simple: Suppose $x^* \sim \mu$, and consider the following process

$$Z_t = x^* + B_t \tag{1}$$

where $(B_t)_{t\geqslant 0}$ is a standard brownian motion, started at 0. The process in Eq 1 turns samples from μ into "noise" and destroying information about x^*. [2] Then, eventually Z_t will 'forget' about x^*. Hence, if one could *reverse* this process, one can obtain a sample from μ. In particular, consider the following time/scaling change

$$Y_t \triangleq tZ_{1/t} = tx^* + tB_{1/t} \overset{(a)}{=} tx^* + \tilde{B}_t \tag{2}$$

where we first reverse time $t \mapsto 1/t$ and apply a scaling by t. In (a), we use the fact that $(\tilde{B}_t)_{t\geqslant 0} = (tB_{1/t})_{t\geqslant 0}$ is also a standard BM, when $(B_t)_{t\geqslant 0}$ is a standard BM. Thus, if we could simulate Eq 2, we could obtain a sample $x^* \sim \mu$, by considering the almost sure limit $x^* = \lim_{t\to\infty} Y_t/t$. In this aspect, we could say that $(Y_t)_{t\geqslant 0}$ *localizes* the sample x^*, as the posterior distribution of x^* given Y_t converges to δ_{x^*} in the limit.

As introduced, the dynamics of Y_t depend on the *unknown latent* x^*. As a result, one might think the process in Eq 2 is useless for sampling. Not all hope is lost though. One of the central results of Stochastic Localization is the following. An observer of $(Y_t)_{t\geqslant 0}$ can describe its dynamics autonomously:

---

[1] Or more generally, methods that work under certain functional inequalities, i.e. log-sobolev or poincare inequality.

[2] E.g. this could be made formal if μ has bounded support or finite second moment

Proposition 1. *Consider the process $Y_t = tx^* + B_t$, where $x^* \sim \mu$ and $(B_t)_{t\geqslant 0}$ is standard brownian motion. Then, Y_t is the unique strong solution of the following*

$$\mathrm{d}Y_t = m(Y_t, t)\mathrm{d}t + \mathrm{d}B_t, \qquad Y_0 = 0$$

where $m : \mathbb{R}^d \times [0, \infty)$ is defined as the following conditional expectation of x^* given Y_t

$$m(y, t) = \mathbb{E}[x^* | tx^* + \sqrt{t}G = y]$$

where $(x^*, G) \sim \mu \otimes \mathcal{N}(0, I)$ are independent.

This result essentially says the following: If we consider the information of the observer of Y_t (without the knowledge of x^*), Y_t follows a self-governing SDE. Furthermore, Y_t eventually reveals some x^* which is random, and distributed as μ. Hence, Y_t *localizes* a random x^*. As a result, if we can compute the conditional mean $m(y, t)$, then we can draw a sample from μ by simulating the SDE in Proposition 1 to some large T, and taking $\hat{x} = Y_T/T$. In fact, the SDE in Proposition 1 is very closely related to the *variance exploding* formulation of denoising diffusions in the context of generative AI.

In light of the above example, the goals of this exposition is as follows: (i) formally establish[3] the time-reversal SDE given in Proposition 1, (ii) establish the formal connections of the above $(Y_t)_{t\geqslant 0}$ to more general stochastic localization. While this exposition was inspired largely by [Mon23], we significantly deviate in our proofs, and aim to be a more self contained exposition with more proofs included.

2. Sampling and localization via time reversal

In this section, we first derive the SDE in Proposition 1 and illustrate the derivation of the SDE for the time reversal $(Y_t)_{t\geqslant 0}$ using *Kolmogorov equations*. Then, we will elaborate on further technical conditions on when there is *pathwise time reversal*. Motivated by the above discussion, consider the following SDE:

$$\mathrm{d}Z_t = \mathrm{d}B_t, \qquad Z_0 \sim \mu$$

Then, by the Kolmogorov equations, we have that if $p_t(\cdot)$ is the density of Z_t at time t, then

$$\frac{\mathrm{d}p_t(x)}{\mathrm{d}t} = \frac{1}{2}\Delta p_t(x)$$

Our goal is to understand the density of the time reversal $(Y_t)_{t\geqslant 0}$ which is given by $tZ_{1/t}$. Then, consider $q_t(x) = p_{1/t}(x)$. Hence, by the chain rule

$$\begin{aligned} \frac{\mathrm{d}q_t(x)}{\mathrm{d}t} &= -\frac{1}{2t^2} \cdot \Delta q_t(x) = -\frac{1}{t^2}\Delta q_t(x) + \frac{1}{2t^2}\Delta q_t(x) \\ &= -\mathrm{div}(\nabla q_t(x)) + \frac{1}{2} \cdot \frac{1}{t^2}\Delta q_t(x) \\ &= -\mathrm{div}(q_t(x)\nabla \log q_t(x)) + \frac{1}{2} \cdot \frac{1}{t^2}\Delta q_t(x) \end{aligned} \tag{3}$$

Now, the first term corresponds to a drift $\frac{1}{t^2}\nabla \log q_t(x)\mathrm{d}t$ and the second term corresponds to the diffusion $\frac{1}{t}\mathrm{d}B_t$. Pattern matching these terms, consider the following SDE:

$$\mathrm{d}\tilde{Y}_t = \frac{1}{t^2}\nabla \log q_t(\tilde{Y}_t)\mathrm{d}t + \frac{1}{t}\mathrm{d}B_t \tag{4}$$

[3]Modulo some technical details.

where we take the Brownian motion independent of the previous constructions.[4] One can easily verify that the SDE 4 with density $q_t(x)$ satisfies the PDE in Eq 3, which is by construction. Then, letting $Y_t = t\tilde{Y}_t$ and applying Itô, we have

$$\mathrm{d}Y_t = \underbrace{\frac{Y_t}{t}\mathrm{d}t}_{\text{time derivative term}} + \underbrace{\frac{1}{t}\nabla \log q_t(Y_t/t)\mathrm{d}t + \mathrm{d}B_t}_{\text{space derivative terms}}$$

At this point, we recognize that $q_t = p_{1/t}$ is given by a gaussian convolution. Then, to simplify further, we refer to the following:

Proposition 2 (Tweedie's formula)**.** *Let $X \sim \mu$ and $Y = \alpha X + \sigma Z$. Then, if $p : \mathbb{R}^d \to \mathbb{R}_{>0}$ is the density of Y, we have*

$$\nabla \log p(y) = \frac{\alpha \mathbb{E}[X|Y = y] - y}{\sigma^2}$$

A simple proof of this using exponential families could be found in Appendix A. Hence, noting that $p_{1/t}$ is given by the density of $Z_{1/t} = Z_0 + \frac{1}{\sqrt{t}}G$, we have

$$\frac{1}{t}\nabla \log q_t(Y_t/t) = \frac{1}{t} \cdot \frac{\mathbb{E}[Z_0|Z_0 + \frac{1}{\sqrt{t}}G = Y_t/t] - Y_t/t}{1/t} = \mathbb{E}[Z_0|tZ_0 + \sqrt{t}G = Y_t] - Y_t/t$$

Then, plugging back into the SDE, we obtain:

$$\mathrm{d}Y_t = \mathbb{E}[Z_0|tZ_0 + \sqrt{t}G = Y_t]\mathrm{d}t + \mathrm{d}B_t$$

Since $Z_0 \sim \mu$, we get that

$$\mathrm{d}Y_t = m(Y_t, t)\mathrm{d}t + \mathrm{d}B_t$$

However, we did not specify the initial condition. Recall $\tilde{Y}_t \sim \mu \circledast \mathcal{N}(0, t)$. Hence, $Y_t \sim (\mu \circledast \mathcal{N}(0, t)) \cdot t$, which, say for bounded support distributions μ, have $\lim_{t \to 0} Y_t = 0$. Hence, we conclude that the initial condition is $Y_0 = 0$. Some remarks:

(1): To implement the SDE for $(Y_t)_{t \geqslant 0}$, one needs only to compute the function $m(y, t)$ which is a conditional expectation. One way to do this is estimating m from data. In particular, consider i.i.d. samples $X^i \sim \mu$, and for some function class $\mathcal{F}$ over functions $f(y, t)$, consider the following objective:

$$\theta^* = \arg\min_\theta \frac{1}{N}\sum_{i=1}^{N}(X^i - f_\theta(tX^i + \sqrt{t}G^i, t))^2$$

Then, as $\mathcal{F}$ becomes more expressive and $N \to \infty$, we have that $f_\theta(y, t) \to \mathbb{E}[x^*|tx^* + \sqrt{t}G = y] = m(y, t)$. This is known as a denoising loss, and has been widely used in practice to train generative models for continuous data (such as images, proteins, etc).[5]

(2): While the above derivations shows how to match the marginals of the reversal $tZ_{1/t}$, it is indeed true that we have a pathwise reversal. Formally, this requires some work. In particular, one can take the results of [HP86], and first apply time reversal on $[\epsilon, T]$, and then monotone time rescaling with spatial rescaling. We defer the formal derivation to Appendix B. Note that

[4] With the appropriate initial condition.

[5] Albeit with a slightly different process. In practice, it is more common to reverse variance preserving processes like $\mathrm{d}X_t = -X_t + \sqrt{2}\mathrm{d}B_t$ where the variance is preserved.

the formal derivation is non-trivial since we could have matched the marginals with alternative processes, i.e. any SDE of the form

$$\mathrm{d}\tilde{Y}_t = \frac{1+\alpha}{2t^2}\nabla \log q_t(\tilde{Y}_t)\mathrm{d}t + \frac{\alpha}{t}\mathrm{d}B_t$$

In fact, SDE's of this form are widely used in practice for sampling (since we do not always care to match the entire path).[6] It turns out that the correct time-reversal is the one that matches the quadratic variation of the forward process (up to time scaling).

3. Sampling and localization via filtering

In this section, we will (i) provide an alternative derivation of Proposition 1 filtering and (ii) introduce stochastic localization as a generalization. Recall, the process $(Y_t)_{t\geqslant 0}$ was defined initially as $Y_t = tx^* + B_t$ for some brownian motion $(B_t)_{t\geqslant 0}$ and $x^* \sim \mu$. The main issue for sampling was that the latent x^* was unknown. Thus, a natural question is, what does $(Y_t)_{t\geqslant 0}$ look like given an observer, who sees the filtrations $\mathcal{F}_t^Y = \sigma(Y_s, s \leqslant t)$. We already know from earlier that it should be an SDE, but we have the following:

Proposition 3. *Let $\mathcal{F}_t^Y = \sigma(Y_s, s \leqslant t)$. Define $m_t = \mathbb{E}[x^*|\mathcal{F}_t^Y]$. Then,*

$$\tilde{B}_t = Y_t - \int_0^t m_s \mathrm{d}s$$

is an $\mathcal{F}_t^Y$-brownian motion.

Proof. We will use Levy's characterization of brownian motion. First, $\tilde{B}_t$ is $\mathcal{F}_t^Y$-adapted by definition. Then, note

$$\mathbb{E}[\tilde{B}_t - \tilde{B}_s|\mathcal{F}_s^Y] = \mathbb{E}[\mathbb{E}[\tilde{B}_t - \tilde{B}_s|x^*]|\mathcal{F}_s^Y]$$

However, conditioning on both x^* and $\mathcal{F}_s^Y$ we can equivalently condition on x^* and B_s, note

$$\mathbb{E}[Y_t - Y_s|x^*, \mathcal{F}_s^Y] = \mathbb{E}[B_t - B_s|B_s] + (t-s)x^* = (t-s)x^*$$

Hence, noting that $\mathbb{E}[\tilde{B}_t - \tilde{B}_s|\mathcal{F}_s^Y] = \mathbb{E}[Y_t - Y_s - \int_s^t m_r \mathrm{d}r|\mathcal{F}_s^Y] = (t-s)m_s - \mathbb{E}\left[\int_s^t m_r \mathrm{d}r|\mathcal{F}_s^Y\right]$, and that $\int_s^t \mathbb{E}[m_r|\mathcal{F}_s^Y]\mathrm{d}r = \int_s^t m_s \mathrm{d}r = (t-s)m_s$,[7] we have that $\mathbb{E}[\tilde{B}_t - \tilde{B}_s|\mathcal{F}_s^Y] = 0$. We conclude $\tilde{B}_t$ is an $\mathcal{F}_t^Y$-martingale. Finally, the quadratic variation of $\tilde{B}_t$ has $\langle \tilde{B}\rangle_t = \langle Y\rangle_t = \langle B\rangle_t = tI$. By Levy's characterization, $(\tilde{B}_t)_{t\geqslant 0}$ is a $\mathcal{F}_t^Y$ brownian motion. ∎

Applying this result, we get

$$Y_t = \int_0^t m_s \mathrm{d}s + \tilde{B}_t$$

where $\tilde{B}_t$ is $\mathcal{F}_t^Y$ adapted brownian motion. Equivalently, $(\Omega, \mathcal{F}, \mathcal{F}_t^Y, \mathbb{P}, \tilde{B}_t, Y_t)$ form a solution to the SDE

$$\mathrm{d}Y_t = m_t \mathrm{d}t + \mathrm{d}\tilde{B}_t$$

which recovers the SDE in Proposition 1. At this point, it should be surprising that one can derive the same SDE in two seemingly unrelated ways: (i) time reversal of SDEs (ii) a filtering based argument.

[6] For instance, $\alpha = 0$ recovers the probability flow ODE.

[7] This is because $m_t = \mathbb{E}[x^*|\mathcal{F}_t^Y]$ which makes it a martingale.

Now, we return to stochastic localization. Let $\mu_t(\cdot) = \mathbb{P}_x(\cdot|Y_t) = \mathbb{P}_x(\cdot|(Y_t)_{t\geqslant 0})$. Then, by a straightforward application of Bayes' rule,

$$\mu_t(\mathrm{d}x)\propto\mu(\mathrm{d}x)\exp\left\{-t\frac{\|x\|^2}{2}+\langle Y_t, x\rangle\right\}$$

Hence, the measures $\mu_t(\mathrm{d}x)$ form a measure valued process, for which $\mu_t \to \delta_{x^*}$. In fact, $\mu_t(\cdot)$ form a *martingale*, which can be shown as a direct application of Itô on the Bayes rule characterization:

Proposition 4 (μ_t follow an SDE)**.** *Let $Y_t, \mu_t, \mathcal{F}_t^Y$ be as previously defined. Then,*

$$\mathrm{d}\mu_t(x) = \mu_t(x)\langle x - m_t, \mathrm{d}\tilde{B}_t\rangle, \qquad \mu_0 = \mu$$

Proof. Let $Z_t = \int \mu(\mathrm{d}x)\exp\{-t\|x\|^2/2 + \langle Y_t, x\rangle\}$. Then, note by Ito's lemma

$$\mathrm{d}\log\mu_t(x) = -\frac{\|x\|^2}{2}\mathrm{d}t + x\mathrm{d}Y_t - \mathrm{d}\log Z_t$$

Furthermore, using Itô's lemma and properties of the cumulant generating function $\log Z_t$

$$\mathrm{d}\log Z_t = -\left(\int\frac{\|x\|^2}{2}\mu_t(\mathrm{d}x)\right)\mathrm{d}t + \left(\int x\mu_t(\mathrm{d}x)\right)\mathrm{d}Y_t + \left(\frac{1}{2}\int\langle I, xx^\top - m_t m_t^\top\rangle\mu_t(\mathrm{d}x)\right)\mathrm{d}\langle Y\rangle_t$$

However, $\mathrm{d}\langle Y\rangle_t = d\langle\tilde{B}\rangle_t = \mathrm{d}t$. Thus, $\mathrm{d}\log Z_t = m_t dY_t - \frac{1}{2}\|m_t\|^2\mathrm{d}t$. Hence,

$$\mathrm{d}\log\mu_t(x) = \frac{1}{2}\|m_t\|^2\mathrm{d}t - \frac{1}{2}\|x\|^2\mathrm{d}t + \langle x - m_t, dY_t\rangle$$

Noting $dY_t = m_t\mathrm{d}t + \mathrm{d}\tilde{B}_t$, we have

$$\mathrm{d}\log\mu_t = -\frac{1}{2}\|x - m_t\|^2\mathrm{d}t + \langle x - m_t, \mathrm{d}\tilde{B}_t\rangle$$

Then, applying Itô's lemma again with $f(x) = e^x$, we have

$$\mathrm{d}\mu_t(x) = \mu_t(x)\mathrm{d}\log\mu_t(x) + \frac{1}{2}\mu_t(x)\mathrm{d}\langle\log\mu\rangle_t$$

Then, note $d\langle\log\mu\rangle_t = \|x - m_t\|^2$. Thus, we get

$$\begin{aligned}\mathrm{d}\mu_t(x) &= -\frac{1}{2}\mu_t(x)\|x - m_t\|^2\mathrm{d}t + \mu_t(x)\langle x - m_t, \mathrm{d}\tilde{B}_t\rangle + \frac{1}{2}\mu_t(x)\|x - m_t\|^2\\ &= \mu_t(x)\langle x - m_t, \mathrm{d}\tilde{B}_t\rangle\end{aligned}$$

□

We remark that the above implies that $\mu_t(\cdot)$ itself is a measure valued martingale! This verifies that the μ_t matches the definition of the stochastic localization process used in the seminal work [Eld13]. In particular, μ_t is a measure valued martingale such that $\mu_0 = \mu$, $\lim_{t\to\infty}\mu_t = \delta_{x^*}$ for $x^* \sim \mu$.

4. Conclusions

In this expository work, we have established formal connections between sampling and stochastic localization. First, we showed that the stochastic localization process itself can be used for sampling, and gave a derivation of the SDE for $(Y_t)_{t\geqslant 0}$ using a time reversal argument. This is more closely aligned with works on sampling in diffusion processes. Second, we derived the equivalent localization SDE from the perspective of filtering, and thus showed an equivalence between filtering and sampling. Finally, we formally establish that the μ_t process is indeed a stochastic localization process by using the previous SDEs we have derived.

This project was quite interesting to me, as I learned this connection between sampling, localization, and filtering. This reveals an important connection: Denoising based sampling could be viewed as filtering, where we are aiming to iteratively 'clean up' a sample. This is quite satisfying, as it matches the intuition from practice about denoising diffusions. In the future, I am curious to see how the tools from localization can be applied to learn more about diffusion models. In particular, I would like to apply the different approaches (filtering, localization, reverse diffusions) to various problems I encounter in my own research of diffusion models.

References

[Eld13] Ronen Eldan. Thin shell implies spectral gap up to polylog via a stochastic localization scheme. *Geometric and Functional Analysis*, 23(2):532–569, 2013.

[HP86] Ulrich G Haussmann and Etienne Pardoux. Time reversal of diffusions. *The Annals of Probability*, pages 1188–1205, 1986.

[LG16] Jean-François Le Gall. *Brownian motion, martingales, and stochastic calculus*, volume 274. Springer, 2016.

[Mon23] Andrea Montanari. Sampling, diffusions, and stochastic localization. *arXiv preprint arXiv:2305.10690*, 2023.

Appendix A. Omitted proofs from earlier sections

Here we give a concise proof of Tweedie's utilizing properties of exponential families. This makes it easy to remember.

Proof of Tweedie's formula (Proposition 2). Recall $X \sim \mu$ and $Y = \alpha X + \sigma G$ where $G \sim \mathcal{N}(0, I)$ independent of X. Then, note

$$p(y) = \int_{\mathbb{R}^d} \frac{1}{Z} \exp\left\{-\frac{\|\alpha x - y\|^2}{2\sigma^2}\right\} \mathrm{d}\mu(x)$$

$$= \exp\left\{-\frac{\|y\|^2}{2\sigma^2}\right\} \int_{\mathbb{R}^d} \frac{1}{Z} \exp\left\{-\frac{\alpha^2 \|x\|^2}{2\sigma^2} + \frac{\alpha}{\sigma^2}\langle x, y\rangle\right\} \mathrm{d}\mu(x)$$

where $Z = (2\pi\sigma^2)^{-d/2}$ is the gaussian normalizing constant. Then, considering the integral above, define the following exponential family, parameterized by y:

$$\mathcal{E} = \left\{p(x; y) : \frac{\mathrm{d}p(x; y)}{\mathrm{d}\mu(x)} = \exp\left(-\alpha^2 \frac{\|x\|^2}{2\sigma^2} + \frac{\alpha\langle x, y\rangle}{\sigma^2} - \log Z(y)\right)\right\}$$

where $Z(y)$ is the partition function. Then, by the properties of exponential families, note $\nabla \log Z(y) = \frac{\alpha}{\sigma^2}\mathbb{E}_{X\sim p(\cdot;y)}[X] = \frac{\alpha}{\sigma^2}\mathbb{E}[X|Y = y]$. Furthermore, note

$$p(y) = Z(y) \exp\left\{-\frac{\|y\|^2}{2\sigma^2}\right\}$$

using the definition. Hence, we get that

$$\nabla \log p(y) = \nabla \log Z(y) - \nabla \left(\frac{\|y\|^2}{2\sigma^2}\right) = \frac{\alpha \mathbb{E}[X|Y=y] - y}{\sigma^2}$$

as claimed. □

Appendix B. Formal time reversal of SDEs

Here, we first state the formal time reversal result from [HP86] and connect it to our derivation of Proposition 1. In particular, we need to handle first the time reversal $t \mapsto T - t$ and then the monotone time change $T - \frac{1}{t}$.

Theorem 1 ([HP86] Time reversal). *Suppose we have a SDE*

$$\mathrm{d}X_t = b(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}B_t$$

in $\mathbb{R}^d$, for $t \in [0,1]$. Suppose the following hold

(1) b, σ are K-lipschitz in the space coordinate, and have at most K-linear growth (i.e. $\|b(t,x)\|_2 + \|\sigma(t,x)\|_2 \leqslant K(1 + \|x\|_2)$).

(2) The density of X_t, p_t satisfies the following for all $t_0 > 0$ and $j \in [d]$

$$\int_{t_0}^{1} \int_{\mathcal{O}} (|p(t,x)|^2 + \left\|\sigma(t,x)^\top \nabla p(t,x)\right\|^2)\mathrm{d}x\mathrm{d}t < \infty$$

Then, $\bar{X}_t = X_{1-t}$ is a weak solution to the SDE

$$\mathrm{d}\bar{X}_t = \bar{b}(\bar{X}_t, t)\mathrm{d}t + \bar{\sigma}(\bar{X}_t, t)\mathrm{d}\bar{B}_t$$

where

$$\begin{aligned}\bar{b}(x,t) &= -b(x, 1-t) + \nabla \cdot (\sigma(x,1-t)\sigma(x,1-t)^\top) \\ &\quad + \sigma(x,1-t)\sigma(x,1-t)^\top \nabla \log p_{1-t}(x) \\ \bar{\sigma}(x,t) &= \sigma(x, 1-t)\end{aligned}$$

Consider the SDE we have defined

$$\mathrm{d}Z_t = \mathrm{d}B_t, Z_0 \sim \mu$$

Here $b = 0$ and $\sigma = I$. Then, suppose for simplicity that μ has bounded support. Then, assumptions (1) and (2) in the statement of Theorem 1 hold. Furthermore, without harm, we can consider a time interval $[0,T]$ instead of $[0,1]$. Thus, defining $\bar{Z}_t = Z_{T-t}$ for $t \in [0,T]$, we have

$$\mathrm{d}\bar{Z}_t = \nabla \log p_{T-t}(\bar{Z}_t)\mathrm{d}t + \mathrm{d}\bar{B}_t$$

for some brownian motion $\bar{B}_t$. Now, consider the monotone increasing time change $t \mapsto \tau(t) = T - \frac{1}{t}$ mapping $[1/T, \infty)$ to $[0,T]$. And the time changed process $\hat{Z}_t = \bar{Z}_{\tau(t)}$. Then, the time changed process $\hat{Z}_t = \bar{Z}_{\tau(t)}$ has

$$\hat{Z}_t = \hat{Z}_0 + \int_0^{\tau(t)} \nabla \log p_{T-s}(\bar{Z}_s)\mathrm{d}s + \bar{B}_{\tau(t)}$$

Applying the change of variables $s = \tau(u)$ with $\frac{1}{u^2}\mathrm{d}u = \mathrm{d}s$ in the integral, we get $\int_0^{\tau(t)} \nabla \log p_{T-s}(\bar{Z}_s)\mathrm{d}s = \int_{1/T}^{t} \frac{1}{u^2}\nabla \log p_{1/u}(\hat{Z}_u)\mathrm{d}u$. Thus,

$$\hat{Z}_t - \int_{1/T}^{t} \frac{1}{s^2}\nabla \log p_{1/s}(\hat{Z}_s)\mathrm{d}s = \bar{B}_{\tau(t)}$$

In particular, the RHS is a martingale with the quadratic variation $\langle \bar{B}\rangle_{\tau(t)} - \langle \bar{B}\rangle_{\tau(1/T)} = T - \frac{1}{t}$. Equivalently, consider the martingale

$$\hat{M}_t = \int_{1/T}^{t} \frac{1}{s}\mathrm{d}\hat{B}_s$$

which has the same quadratic variation. Then, referring to martingale representation theorem [LG16], we have that for some brownian motion $\hat{B}_t$, for $t \in [1/T, \infty)$

$$\mathrm{d}\hat{Z}_t = \frac{1}{t^2}\nabla \log p_{1/t}(\hat{Z}_t)\mathrm{d}t + \frac{1}{t}\mathrm{d}\hat{B}_t$$

for $t \in [1/T, \infty)$ with $\hat{Z}_{1/T} = \bar{Z}_0$. Applying the rescaling $t\hat{Z}_t$ and sending $T \to \infty$ recovers the result of Proposition 1.