Supplementary Material: Continuous-Time Functional Diffusion Processes

A Reverse Functional Diffusion Processes

In this Section, we review the mathematical details to obtain the backward FDP discussed in Theorem 1. Depending on the considered class of noise, different approaches are needed. First, we present in Appendix A.1 the conditions to ensure existence of the backward process , which we use if the C operator is an identity matrix, $C = I$. Then we move to a different approach in Appendix A.2 for the case $C \neq I$.

A.1 Follmer Formulation

The work in Föllmer (1986) is based on a finite entropy condition, which we report here as Condition 1. One simple way to ensure that the condition is satisfied is to assume:

Condition 1. *For a given k, define $\mathbb{Q}_{(k)}$ to be the path measure corresponding to the (infinite) system*

$$\begin{cases} \mathrm{d}X_t^i = b^i(X_t, t)\mathrm{d}t + \mathrm{d}W_t^i, & i \neq k \\ \mathrm{d}X_t^i = \mathrm{d}W_t^k, & i = k. \end{cases} \tag{22}$$

We say that $\mathbb{Q}$ satisfies the finite local entropy condition if $\mathrm{KL}\left[\mathbb{Q} \parallel \mathbb{Q}_{(k)}\right] < \infty, \forall k$.

Define $\mathcal{F}_t^{(i)} = \sigma(X_0^i, X_s^j, 0 \leq s \leq t, j \neq i)$.

Assumption 1.

$$\int_0^T b^i(X_t, t)^2 \mathrm{d}t + \sum_{j \neq i} \mathbb{E}[\int_0^T \left(b^j(X_t, t) - \mathbb{E}\left[b^j(X_t, t) \,|\, \mathcal{F}_t^{(i)}\right]\right)^2 \mathrm{d}t] < \infty, \mathbb{Q}_{(i)} a.s. \tag{23}$$

Notice that if Assumption 1 is true, then Condition 1 holds (Föllmer (1986), Thm. 2.23)

Theorem 3. *If* $\mathrm{KL}\left[\mathbb{Q} \parallel \mathbb{Q}_{(k)}\right] < \infty$*, then* $\mathrm{KL}\left[\hat{\mathbb{Q}} \parallel \hat{\mathbb{Q}}_{(k)}\right] < \infty$.

Proof. The proof can be obtained by adapting the result of Lemma 3.6 of Föllmer & Wakolbinger (1986). ☐

This Theorem states that if the forward FDP path measure $\mathbb{Q}$ satisfies the finite local entropy condition, then also the reverse FDP path measure $\hat{\mathbb{Q}}$ satisfies the finite local entropy condition.

Theorem 4. *Let $\mathbb{Q}$ be a finite entropy measure. Then:*

$$\begin{cases} \mathrm{d}X_t^k = b^k(X_t, t)\mathrm{d}t + \mathrm{d}W_t^k, & \textit{under} \quad \mathbb{Q} \\ \mathrm{d}\hat{X}_t^k = \hat{b}^k(\hat{X}_t, t)\mathrm{d}t + \mathrm{d}\hat{W}_t^k, & \textit{under} \quad \hat{\mathbb{Q}} \end{cases} \tag{24}$$

where:

$$\frac{\partial \log\left(\rho_t^{(d)}(x^k \,|\, x^j, j \neq k)\right)}{\partial x^k} = \hat{b}^k(x, T-t) + b^k(x, t) \tag{25}$$

Proof. For the proof, we refer to Theorem 3.14 of Föllmer & Wakolbinger (1986). ☐

A.2 Millet Formulation

Let $L^2(R) = \{x \in H : \sum r^i (x^i)^2 < \infty\}$. For simplicity, we overload the notation of the letter K, and use it for generic constants, that might be different on a case by case basis.

Assumption 2.

$$\forall x \in L^2(R), \sup_t \{\sum r^i (b^i(x,t))^2\} + \sum (r^i)^2 \leq K(1 + \sum r^i (x^i)^2)$$

$$\forall x, y \in L^2(R), \sup_t \{\sum r^i (b^i(x,t) - b^i(y,t))^2\} \leq K \sum r^i (x^i - y^i)^2$$

This assumption is simply the translation of H1 from Millet et al. (1989) to our notation.

Assumption 3. *There exists an increasing sequence of finite subsets* $J(n), n \in \mathrm{N}, \cup_n J(n) = \mathrm{N}$ *such that* $\forall n \in \mathrm{N}, M > 0$ *there exists a constant* $K(M,n)$ *such that the following holds:*

$$\sup_t \left(\sup_{i \in J(n)} \left(\left(\sup_x |b^i(x,t)| : \sup_{j \in J(n)} |x^j| \leq M \right) + \sum_j r^j \right) \right) \leq K(M,n).$$

Again, this assumption is simply the translation of H5 from Millet et al. (1989) to our notation.

Assumption 4. *Either i):*

$$\forall x, y \in L^2(R), \sup_t \{ \sum r^i (b^i(x,t) - b^i(y,t))^2 \} \leq K \sum (r^i)^2 (x^i - y^i)^2,$$

or ii): $\forall i, b^i(x)$ *is a function of* x *for at most* M *coordinates and*

$$\forall x, y \in L^2(R), \sup_t \{ \sum (r^i)^2 (b^i(x,t) - b^i(y,t))^2 \} \leq K \sum (r^i)^2 (x^i - y^i)^2.$$

This corresponds to satisfying either H3 or jointly H2 and H4 of Millet et al. (1989). For simplicity, we can combine together the different assumptions into

Assumption 5. *Let Assumption 2, Assumption 3, and Assumption 4 hold.*

Finally, we state required assumptions about the density:

Assumption 6. *Suppose that the initial condition is* $X_0 \in L^2(R)$.

- *Assume that the conditional law of* x^i *given* $x^j, j \neq i$ *has density* $\rho_t^{(d)}(x^i \mid x^j, j \neq i)$ *w.r.t Lebesgue measure on* $\mathbb{R}$.
- *Assume that* $\int_{t_0}^1 \int_{D_J} |r^i \frac{\mathrm{d}}{\mathrm{d}x^i}(\rho_t^{(d)}(x^i \mid x^j, j \neq i))| \mathrm{d}x^i \rho_t(\mathrm{d}x^{j \neq i}) \mathrm{d}t < \infty$, *for fixed subset* $J \subset \mathrm{N}, t_0 > 0$ *and* $D_J = \{ (\prod_{j \in J} K_j) \times (\prod_{j \notin J} \mathbb{R}), K_j$ *compact in* $\mathbb{R} \} \cap L^2(R)$.

We reported in our notation the content of Theorem 4.3 of Millet et al. (1989). This can be used to prove the existence of the backward process.

A.3 Proof of Theorem 1

If $R = I$, then we assume Assumption 1. Consequently, $\mathbb{Q}$ is a finite entropy measure. Then Theorem 4 holds, from which the desired result. If, instead $R \neq I$, then we require Assumption 5,Assumption 6. Application of Thm 4.3 of Millet et al. (1989) allows to prove the validity of Theorem 1 also in this case.

A.3.1 Proof of Corollary 1

Assumption 5 is required directly. We need to show that with the considered restrictions Assumption 6 is valid.

Since $\sum_i r^i < \infty$, then $\sum_i (r^i)^2 = K_a < \infty$. Moreover, $(b^i(x^i,t))^2 < K_b^2 (x^i)^2$. Then, $\forall x \in L^2(R)$, the following holds $\sup_t \{ \sum r^i (b^i(x,t))^2 \} + \sum (r^i)^2 \leq \sum r^i K_b^2 (x^i)^2 + K_a \leq \max(K_a, K_b^2) \left(1 + \sum r^i (x^i)^2 \right)$. Similarly, $\forall x, y \in L^2(R)$ we have $\sup_t \{ \sum r^i (b^i(x,t) - b^i(y,t))^2 \} \leq \sum r^i K_b^2 (x^i - y^i)^2$. Thus Assumption 2 is satisfied.

Since $b^i(x,t)$ is bounded and independent on t, Assumption 3 is satisfied, as explicitly discussed in Millet et al. (1989).

Finally, since $b^i(x)$ is a function of x for $M = 1$ coordinate, and $\sup_t \{ \sum (r^i)^2 (b^i(x,t) - b^i(y,t))^2 \} \leq \sum (r^i)^2 K_b^2 (x^i - y^i)^2$, Assumption 4 is satisfied.

Then, combined toghether Assumption 5 holds.

A.4 Girsanov Regularity

Condition 2. *Assume that $\gamma_{\boldsymbol{\theta}}(x,t)$ is an $\hat{\mathcal{F}}$ measurable process and that either:*

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(\frac{1}{2}\int_0^T \left\|\gamma_{\boldsymbol{\theta}}(\hat{X}_t,t)\right\|^2_{R^{\frac{1}{2}}H}\mathrm{d}t\right)\right] = \mathbb{E}_{\mathbb{Q}}\left[\exp\left(\frac{1}{2}\int_0^T \left\|\gamma_{\boldsymbol{\theta}}(X_t,t)\right\|^2_{R^{\frac{1}{2}}H}\mathrm{d}t\right)\right] < \infty, \quad (26)$$

or

$$\exists \delta > 0 : \mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(\frac{1}{2}\left\|\gamma_{\boldsymbol{\theta}}(\hat{X}_\delta,\delta)\right\|_{R^{\frac{1}{2}}H}\mathrm{d}t\right)\right] < \infty. \quad (27)$$

This is equivalent to the regularity condition in eq. 10.23 of Da Prato & Zabczyk (2014) or Proposition 10.17 in Da Prato & Zabczyk (2014).

A.5 Proof of KL divergence expression

We leverage Equation (7) to express the Kullback-Leibler divergence as:

$$\mathrm{KL}\left[\hat{\mathbb{Q}} \parallel \hat{\mathbb{P}}^{\chi_T}\right] = \mathbb{E}_{\hat{\mathbb{Q}}}\left[\log\frac{\mathrm{d}\hat{\mathbb{Q}}_0}{\mathrm{d}\hat{\mathbb{P}}_0} + \log\frac{\mathrm{d}\rho_T}{\mathrm{d}\chi_T}\right] = \mathbb{E}_{\hat{\mathbb{Q}}}\left[\log\frac{\mathrm{d}\hat{\mathbb{Q}}_0}{\mathrm{d}\hat{\mathbb{P}}_0}\right] + \mathrm{KL}\left[\rho_T \parallel \chi_T\right] =$$

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[-\int_0^T \langle\gamma_{\boldsymbol{\theta}}(\hat{X}_t,t),\mathrm{d}\hat{W}_t\rangle_{R^{\frac{1}{2}}H} + \frac{1}{2}\int_0^T \left\|\gamma_{\boldsymbol{\theta}}(\hat{X}_t,t)\right\|^2_{R^{\frac{1}{2}}H}\mathrm{d}t\right] + \mathrm{KL}\left[\rho_T \parallel \chi_T\right] =$$

$$\frac{1}{2}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T \left\|\gamma_{\boldsymbol{\theta}}(\hat{X}_t,t)\right\|^2_{R^{\frac{1}{2}}H}\mathrm{d}t\right] + \mathrm{KL}\left[\rho_T \parallel \chi_T\right] = \frac{1}{2}\mathbb{E}_{\mathbb{Q}}\left[\int_0^T \left\|\gamma_{\boldsymbol{\theta}}(X_t,t)\right\|^2_{R^{\frac{1}{2}}H}\mathrm{d}t\right] + \mathrm{KL}\left[\rho_T \parallel \chi_T\right].$$

Moreover, since

$$\mathrm{KL}\left[\hat{\mathbb{Q}} \parallel \hat{\mathbb{P}}^{\chi_T}\right] = \mathbb{E}_{\mathbb{Q}}\left[\log\frac{\mathrm{d}\hat{\mathbb{Q}}_T}{\mathrm{d}\hat{\mathbb{P}}^{\chi_T}_T} + \log\frac{\mathrm{d}\rho_0}{\mathrm{d}\chi_0}\right] \geq \mathrm{KL}\left[\rho_0 \parallel \chi_0\right],$$

we can combine the two results and obtain Equation (8)

A.6 Conditional score matching

In this subsection we prove the equality in Equation (13):

$$\mathbb{E}_{\mathbb{Q}}\left[\int_0^T \left\|\gamma_{\boldsymbol{\theta}}(X_t,t)\right\|^2_{R^{\frac{1}{2}}H}\mathrm{d}t\right] = \int_0^T\int_H \left\|\gamma_{\boldsymbol{\theta}}(x,t)\right\|^2_{R^{\frac{1}{2}}H}\mathrm{d}t\mathrm{d}\rho_t(x) =$$

$$\int_0^T\int_H \left\|D_x\log\rho^{(d)}_{T-t}(x) - s_{\boldsymbol{\theta}}(x,T-t)\right\|^2_{R^{\frac{1}{2}}H}\mathrm{d}t\mathrm{d}\rho_t(x) =$$

$$\int_0^T\int_{H\times H} \left\|D_x\log\rho^{(d)}_{t}(x) - s_{\boldsymbol{\theta}}(x,t)\right\|^2_{R^{\frac{1}{2}}H}\mathrm{d}t\mathrm{d}\rho_t(x,x_0) =$$

$$\int_0^T\int_{H\times H} \left\|D_x\log\rho^{(d)}_{t}(x) - D_x\log\rho^{(d)}_{t}(x\,|\,x_0) + D_x\log\rho^{(d)}_{t}(x\,|\,x_0) - s_{\boldsymbol{\theta}}(x,t)\right\|^2_{R^{\frac{1}{2}}H}\mathrm{d}t\mathrm{d}\rho_t(x,x_0) =$$

$$\int_0^T\int_{H\times H} \left\|D_x\log\rho^{(d)}_{t}(x) - D_x\log\rho^{(d)}_{t}(x\,|\,x_0)\right\|^2_{R^{\frac{1}{2}}H} + \left\|D_x\log\rho^{(d)}_{t}(x\,|\,x_0) - s_{\boldsymbol{\theta}}(x,t)\right\|^2_{R^{\frac{1}{2}}H} +$$

$$2\left\langle D_x\log\rho^{(d)}_{t}(x) - D_x\log\rho^{(d)}_{t}(x\,|\,x_0), D_x\log\rho^{(d)}_{t}(x\,|\,x_0) - s_{\boldsymbol{\theta}}(x,t)\right\rangle \mathrm{d}t\mathrm{d}\rho_t(x,x_0).$$

To simplify the equality, we need to notice that:

$$\rho_t^{(d)}(x^i|x^{j\neq i})\mathrm{d}x^i = \mathrm{d}\rho_t(x^i|x^{j\neq i}) = \int_{x_0} \mathrm{d}\rho_t(x_0|x)\mathrm{d}\rho_t(x^i|x^{j\neq i}) = \int_{x_0} \mathrm{d}\rho_t(x^i, x_0|x^{j\neq i}) =$$
$$\int_{x_0} \mathrm{d}\rho_t(x^i|x_0, x^{j\neq i})\mathrm{d}\rho_t(x_0|x^{j\neq i}) = \mathrm{d}x^i \int_{x_0} \rho_t^{(d)}(x^i|x_0, x^{j\neq i})\mathrm{d}\rho_t(x_0|x^{j\neq i}).$$

Then, computing

$$\int_{x_0} \frac{\mathrm{d}}{\mathrm{d}x^i} \log \rho^{(d)}(x^i|x^{j\neq i}, x_0)\mathrm{d}\rho_t(x, x_0) = \int_{x_0} \frac{\frac{\mathrm{d}}{\mathrm{d}x^i}\rho^{(d)}(x^i|x^{j\neq i}, x_0)}{\rho^{(d)}(x^i|x^{j\neq i}, x_0)}\mathrm{d}\rho_t(x, x_0) =$$
$$\int_{x_0} \frac{\frac{\mathrm{d}}{\mathrm{d}x^i}\rho^{(d)}(x^i|x^{j\neq i}, x_0)}{\rho^{(d)}(x^i|x^{j\neq i}, x_0)}\mathrm{d}\rho_t(x^i|x^{j\neq i}, x_0)\mathrm{d}\rho_t(x_0, x^{j\neq i}) = \int_{x_0} \frac{\mathrm{d}}{\mathrm{d}x^i}\rho^{(d)}(x^i|x^{j\neq i}, x_0)\mathrm{d}x^i\mathrm{d}\rho_t(x_0, x^{j\neq i}) =$$
$$\int_{x_0} \frac{\mathrm{d}}{\mathrm{d}x^i}\rho^{(d)}(x^i|x^{j\neq i}, x_0)\mathrm{d}x^i\mathrm{d}\rho_t(x_0|x^{j\neq i})\mathrm{d}\rho_t(x^{j\neq i}) = \frac{\mathrm{d}}{\mathrm{d}x^i}\left(\int_{x_0} \rho^{(d)}(x^i|x^{j\neq i}, x_0)\mathrm{d}\rho_t(x_0|x^{j\neq i})\right)\mathrm{d}x^i\mathrm{d}\rho_t(x^{j\neq i}) =$$
$$\frac{\mathrm{d}}{\mathrm{d}x^i}\rho_t^{(d)}(x^i|x^{j\neq i})\mathrm{d}x^i\mathrm{d}\rho_t(x^{j\neq i}) = \frac{\mathrm{d}\log \rho_t^{(d)}(x^i|x^{j\neq i})}{\mathrm{d}x^i}\rho_t^{(d)}(x^i|x^{j\neq i})\mathrm{d}x^i\mathrm{d}\rho_t(x^{j\neq i}) = \frac{\mathrm{d}\log \rho_t^{(d)}(x^i|x^{j\neq i})}{\mathrm{d}x^i}\mathrm{d}\rho_t(x)$$

Consequently:

$$\int_{H\times H} \left\langle D_x \log \rho_t^{(d)}(x) - D_x \log \rho_t^{(d)}(x \,|\, x_0), s_{\boldsymbol{\theta}}(x, t)\right\rangle \mathrm{d}\rho_t(x, x_0) = 0.$$

Combining together and rearranging the terms, we get the desired Equation (13).

A.7 Explicit expression of score function

As mentioned in the text, we consider the case $f = 0$. In this case, there exists a weak solution to Equation (1) as:

$$X_t = \exp(t\mathcal{A})X_0 + \int_0^t \exp((t-s)\mathcal{A})\mathrm{d}W_s. \tag{28}$$

Consequently, the true score function has expression:

$$\frac{\mathrm{d}}{\mathrm{d}x^i}\log\rho_t^{(d)}(x^i|x^{j\neq i}) = \frac{\frac{\mathrm{d}}{\mathrm{d}x^i}\rho_t^{(d)}(x^i|x^{j\neq i})}{\rho_t^{(d)}(x^i|x^{j\neq i})} = \frac{\frac{\mathrm{d}}{\mathrm{d}x^i}\int_{x_0}\rho_t^{(d)}(x^i|x_0,x^{j\neq i})\mathrm{d}\rho_t(x_0|x^{j\neq i})}{\rho_t^{(d)}(x^i|x^{j\neq i})} =$$

$$\frac{-\int_{x_0}(s^i)^{-1}\left(x^i-\exp\left(tb^i\right)x_0^i\right)\rho_t^{(d)}(x^i|x_0,x^{j\neq i})\mathrm{d}\rho_t(x_0|x^{j\neq i})}{\rho_t^{(d)}(x^i|x^{j\neq i})} =$$

$$\frac{-(s^i)^{-1}\left(x^i\rho_t^{(d)}(x^i|x^{j\neq i})-\int_{x_0}\exp\left(tb^i\right)x_0^i\rho_t^{(d)}(x^i|x_0,x^{j\neq i})\mathrm{d}\rho_t(x_0|x^{j\neq i})\right)}{\rho_t^{(d)}(x^i|x^{j\neq i})} =$$

$$\frac{-(s^i)^{-1}\left(x^i\rho_t^{(d)}(x^i|x^{j\neq i})-\int_{x_0}\exp\left(tb^i\right)x_0^i\rho_t^{(d)}(x^i|x_0,x^{j\neq i})\mathrm{d}\rho_t(x_0|x^{j\neq i})\right)}{\rho_t^{(d)}(x^i|x^{j\neq i})} =$$

$$\frac{-(s^i)^{-1}\left(x^i\rho_t^{(d)}(x^i|x^{j\neq i})-\int_{x_0^i}\exp\left(tb^i\right)x_0^i\rho_t^{(d)}(x^i|x_0^i,x^{j\neq i})\mathrm{d}\rho_t(x_0^i|x^{j\neq i})\right)}{\rho_t^{(d)}(x^i|x^{j\neq i})} =$$

$$\frac{-(s^i)^{-1}\left(x^i\rho_t^{(d)}(x^i|x^{j\neq i})-\int_{x_0^i}\exp\left(tb^i\right)x_0^i\rho_t^{(d)}(x^i|x_0^i,x^{j\neq i})\rho^{(d)}(x_0^i|x^{j\neq i})\mathrm{d}x_0^i\right)}{\rho_t^{(d)}(x^i|x^{j\neq i})} =$$

$$\frac{-(s^i)^{-1}\left(x^i\rho_t^{(d)}(x^i|x^{j\neq i})-\int_{x_0^i}\exp\left(tb^i\right)x_0^i\rho_t^{(d)}(x^i,x_0^i|x^{j\neq i})\mathrm{d}x_0^i\right)}{\rho_t^{(d)}(x^i|x^{j\neq i})} =$$

$$\frac{-(s^i)^{-1}\left(x^i\rho_t^{(d)}(x^i|x^{j\neq i})-\int_{x_0^i}\exp\left(tb^i\right)x_0^i\rho_t^{(d)}(x_0^i|x)\mathrm{d}x_0^i\right)\rho_t^{(d)}(x^i|x^{j\neq i}))}{\rho_t^{(d)}(x^i|x^{j\neq i})} =$$

$$-(s^i)^{-1}\left(x^i-\int_{x_0^i}\exp(tb^i)x_0^i\rho_t^{(d)}(x_0^i|x)\mathrm{d}x_0^i\right)$$

where $s^i = r^i\frac{\exp\left(2b^it\right)-1}{2b^i}$. This is exactly the desired Equation (11). Similar calculations allow to prove $D_x\log\rho_t^{(d)}(x\,|\,x_0) = -\mathcal{S}(t)^{-1}\left(x-\exp(t\mathcal{A})x_0\right)$.

B Fokker Planck equation

In this Section we discuss the infinite dimensional generalization of the classical Fokker Planck equation. We can associate to Eq. (1) the differential operator:

$$\mathcal{L}_0u(x,t) = D_tu(x,t) + \underbrace{\frac{1}{2}\mathrm{Tr}\left\{RD_x^2u(x,t)\right\} + \langle\mathcal{A}x+f(x,t),D_xu(x,t)\rangle}_{\mathcal{L}u(x,t)},\quad x\in H, t\in[0,T], \tag{29}$$

where D_t is the time derivative, D_x, D_x^2 are first and second order Fréchet derivatives in space. The domain of the operator $\mathcal{L}_0$ is $D(\mathcal{L}_0)$, the linear span of real parts of functions $u_{\phi,h} = \phi(t)\exp(i\langle x,h(t)\rangle), x\in H, t\in[0,T]$ where $\phi\in C^1([0,T]), \phi(T)=0$, $h\in C^1([0,T];D(\mathcal{A}^\dagger))$, where $\dagger$ indicates adjoint. Provided appropriate conditions are satisfied, see for example Bogachev et al. (2009, 2011), the time varying measure $\rho_t(\mathrm{d}x)\mathrm{d}t$ exists, is unique, and solves the Fokker-Planck equation $\mathcal{L}_0^\dagger\rho_t = 0$.

C Uncertainty principle

We here clarify that Hilbert spaces of square integrable functions that are not, in general, simultaneously homogeneous and separable. For example, while $\mathbb{R}$ is homogeneous, the set of square integrable functions over $\mathbb{R}$ is not separable, since the basis is the *uncountable* set $\cos(2\pi\nu p), \sin(2\pi\nu p), \nu\in\mathbb{R}$. Then, FDP requirements are not met, as we need a countable basis. Moreover, we would need in general an infinite number of samples (grid over the whole $\mathbb{R}$) to reconstruct the functions. Conversely, a set like the interval $I = [0, 1] \subset \mathbb{R}$ has *countable* basis $\cos(2\pi tp), \sin(2\pi tp), t \in \mathbb{Z}$ (and thus is separable) and, considering x to be band-limited, a sampling grid with finite cardinality would allow to reconstruct of the function. However, I is not homogeneous as no isometry group exists. Consequently, Theorem 2 is not applicable. To fix the issue, one could naively think of extending any function defined over I to the whole $\mathbb{R}$ by considering $\bar{x}[p] = x[p], p \in I$ and $\bar{x}[p] = 0, p \notin I$. Obviously, if $x \in L_2(I)$ then $\bar{x} \in L_2(\mathbb{R})$. However, since $\bar{x}$ has finite support, it cannot be bandlimited, making such an approach not a viable solution. In classical signal processing literature, the problem is usually referred to as the *uncertainty principle* (Slepian, 1983).

D A complete example

We present an example in which we cast Equation (20) for square integrable functions over the interval $I = [0, 1]$, $L^2(I)$. In this case, one natural selection for the basis is the Fourier basis[2] $e^k = \{\dots, \exp(-j2\pi 2p), \exp(-j2\pi p), 1, \exp(j2\pi p), \exp(j2\pi 2p), \dots\}$. Assume the operator $\mathcal{A}$ to be a pseudo-differantial operator, such that $\langle \mathcal{A}x, e^k \rangle = b^k x^k$. Also, assume that b^k, r^k are selected such that conditions of Corollary 1 are met, and consequently the backward process exists. Since we are working with samples collected on the grid $x\left[{}^{i}/_{N}\right]$ we first map the samples to the frequency domain, and then build a Fourier-like representation with a finite set of sinusoids. We then define the mapping $\mathfrak{F}(z^i)^k \stackrel{\text{def}}{=} \sum_{i=0}^{N-1} z^i \exp\left(-j2\pi k \frac{i}{N}\right)$ and its inverse $\mathfrak{I}(z^i)^k \stackrel{\text{def}}{=} N^{-1} \sum_{i=0}^{N-1} z^i \exp\left(j2\pi k \frac{i}{N}\right)$. This suggests to consider th following expression for the interpolating functions:

$$\xi^i = \frac{1}{N} \sum_{k=0}^{N-1} e^k \exp\left(-j2\pi k \frac{i}{N}\right) = \frac{1}{N} \sum_{k=0}^{N-1} \exp\left(j2\pi k (p - \frac{i}{N})\right).$$

Those functions are indeed nothing but a frequency truncated version of the sinc function, which is the classical reconstruction function of the sampling theorem on 1-D signals. Moreover $\langle \xi^i, \zeta^k \rangle = \delta(i - k)$. We are now ready to show *i)* the expression of the forward process, *ii)* the expression of the parametric score function $s_{\boldsymbol{\theta}}$ and $\gamma_{\boldsymbol{\theta}}$, *iii)* the computation of the ELBO and finally *iv)* the expression for the backward process. We defer all detailed derivations to the Appendix.

The forward process defined in Equation (20) has expression:

$$\mathrm{d}X_t\left[{}^{k}/_{N}\right] = \mathfrak{I}\left(b^l \mathfrak{F}(X_t[{}^{i}/_{N}])^l\right)^k \mathrm{d}t + \mathrm{d}W_t\left[{}^{k}/_{N}\right], \quad k = 1, \dots, |Z|, \tag{30}$$

where $\mathrm{d}W_t\left[{}^{k}/_{N}\right] \simeq \mathfrak{F}(\mathrm{d}W_t^i)^k$. Simple calculations show that $X_t\left[{}^{k}/_{N}\right]$ is equivalent in distribution to

$$X_t\left[{}^{k}/_{N}\right] = \mathfrak{I}\left(\exp\left(b^l t\right) \mathfrak{F}(X_0[{}^{i}/_{N}])^l + \sqrt{c^l} \epsilon^l\right)^k, \tag{31}$$

where $s^l = \left\langle \mathcal{S}(t), e^l \right\rangle = r^l \frac{\exp\left(2b^l t\right) - 1}{2b^l}$ and $\epsilon^l \sim \mathcal{N}(0, 1)$, allowing simulation of the forward process in a single step.

The parametric score function can be approximated as:

$$\begin{aligned} & s_{\boldsymbol{\theta}}\left(\sum_i X_t\left[{}^{i}/_{N}\right] \xi^i, t\right)\left[{}^{i}/_{N}\right] = \\ & -\mathfrak{I}\left(\frac{\mathfrak{F}\left(X_t\left[{}^{i}/_{N}\right]\right)^k - \exp\left(b^k t\right) \mathfrak{F}\left(n(g(X_t\left[{}^{l}/_{N}\right]), t, \boldsymbol{\theta})\left[{}^{l}/_{N}\right]\right)}{s^k}\right)^i. \end{aligned} \tag{32}$$

Similarly:

$$\begin{aligned} & \tilde{\gamma}_{\boldsymbol{\theta}}\left(\sum_i X_t\left[{}^{i}/_{N}\right] \xi^i, \sum_i X_0\left[{}^{i}/_{N}\right] \xi^i, t\right)\left[{}^{i}/_{N}\right] = \\ & -\mathfrak{I}\left(\frac{\exp\left(b^k t\right)}{s^k}\left(\mathfrak{F}\left(n(g(X_t\left[{}^{l}/_{N}\right]), t, \boldsymbol{\theta}\right)\left[{}^{l}/_{N}\right] - X_0[{}^{l}/_{N}])^k\right)\right)^i. \end{aligned} \tag{33}$$

[2] We stress that although we should consider a real Hilbert space, we select the complex exponential to avoid cluttering the notation. It is possible to select $\{\cos(2\pi p), \sin(2\pi p), \cos(2\pi 2p), \sin(2\pi 2p), \dots\}$ as a basis, and redoing the calculations in this Section we can obtain a functionally equivalent scheme as the one with the real basis.

Combining Equation (31) and Equation (33) we can fully characterize the training objective defined in Equation (19). Then, it is possible to optimize the value of the parameters $\boldsymbol{\theta}$ with any gradient-based optimizer.

Finally, the backward process approximation is expressed as:

$$\mathrm{d}\hat{X}_t\,[k/N] = -\mathfrak{I}\left(b^l \mathfrak{F}(\hat{X}_t[i/N])^l\right)^k + \mathfrak{I}\left(r^l \mathfrak{F}\left(s_{\boldsymbol{\theta}}(\sum_i \hat{X}_t\,[i/N]\,\xi^i, T-t)\,[i/N]\right)^l\right)\mathrm{d}t + \mathrm{d}W_t\,[k/N] \tag{34}$$

$$k = 1, \dots, |Z|,$$

from which new samples can be generated.

D.1 Proofs

We start by proving Equation (30). Starting from the drift term of Equation (20), we have the following chain of equalities:

$$\left\langle \mathcal{A}\sum_{i=0}^{N-1} X_t[i/N]\xi^i, \zeta^k \right\rangle = \left\langle \sum_{i=0}^{N-1} X_t[i/N]\mathcal{A}\frac{1}{N}\sum_{l=0}^{N-1} e^l \exp\left(-j2\pi l\frac{i}{N}\right), \zeta^k \right\rangle =$$
$$\left\langle \sum_{i=0}^{N-1} X_t[i/N]\frac{1}{N}\sum_{l=0}^{N-1} b^l e^l \exp\left(-j2\pi l\frac{i}{N}\right), \zeta^k \right\rangle =$$
$$\sum_{i=0}^{N-1} X_t[i/N]\frac{1}{N}\sum_{l=0}^{N-1} b^l \exp(j2\pi l k/N)\exp(-j2\pi l i/N) =$$
$$\sum_{l=0}^{N-1} b^l \exp(j2\pi l k/N)\mathfrak{F}(X_t[i/N])^l =$$
$$\mathfrak{I}\left(b^l \mathfrak{F}(X_t[i/N])^l\right)^i.$$

The noise term $\mathrm{d}W_t\,[k/N]$ is approximated as:

$$\mathrm{d}W_t\,[k/N] = \langle \mathrm{d}W_t, \zeta^k\rangle = \langle \sum_{i=0}^{\infty} \mathrm{d}W_t^i e^i, \zeta^k\rangle = \sum_{i=0}^{\infty} \mathrm{d}W_t^i \exp\left(j2\pi i\frac{k}{N}\right) \simeq \mathfrak{F}(\mathrm{d}W_t^i)^k,$$

where we are truncating the sum. The score term has expression:

$$s_{\boldsymbol{\theta}}(\sum_i X_t\,[i/N]\,\xi^i, t) = -(\mathcal{S}(t))^{-1}\left(\sum_i X_t\,[i/N]\,\xi^i - \exp(t\mathcal{A})n(g(X_t\,[i/N]), t, \boldsymbol{\theta})\right) =$$
$$-\sum_k \frac{\overbrace{\sum_i X_t\,[i/N]\left\langle \xi^i, (e^k)^\dagger\right\rangle}^{=\mathfrak{F}(X_t[i/N]) \overset{\text{def}}{=} C_t^k} - \exp\left(b^k t\right)\left\langle n(g(X_t\,[i/N]), t, \boldsymbol{\theta}), (e^k)^\dagger\right\rangle}{s^k} e^k =$$
$$-\sum_k \frac{C_t^k - \exp\left(b^k t\right)\langle n(g(X_t\,[i/N]), t, \boldsymbol{\theta}), \exp(-j2\pi k p)\rangle}{s^k} e^k \simeq$$
$$-\sum_k \frac{C_t^k - \exp\left(b^k t\right)\left(N^{-1}\sum_r n(g(X_t\,[i/N]), t, \boldsymbol{\theta})\left[\frac{r}{N}\right], \exp\left(-j2\pi k\frac{r}{N}\right)\right)}{s^k} e^k,$$

where the approximation is due to the substitution of explicit scalar product with the discretized version trough $\mathfrak{F}$. When evaluated on the grid of interest:

$$
\begin{aligned}
&s_{\boldsymbol{\theta}}(\sum_i X_t\left[{}^i/_N\right]\xi^i, t)\left[{}^i/_N\right] = \\
&-\sum_k \frac{\left(C_t^k - \exp\left(b^k t\right)\left(N^{-1}\sum_r n(g(X_t\left[{}^i/_N\right]), t, \boldsymbol{\theta})\left[\frac{r}{N}\right], \exp\left(-j2\pi k\frac{r}{N}\right)\right)\right)}{s^k}\exp(j2\pi k{}^i/_N) = \\
&-\mathfrak{I}\left(\frac{\mathfrak{F}\left(X_t\left[{}^i/_N\right]\right) - \exp\left(b^k t\right)\mathfrak{F}\left(n(g(X_t\left[{}^i/_N\right]), t, \boldsymbol{\theta})\left[{}^i/_N\right]\right)}{s^k}\right).
\end{aligned}
$$

The value of $\tilde{\gamma}_{\boldsymbol{\theta}}$, Equation (33) and the expression of the backward process, Equation (34), are obtained similarly, considering the above results.

E Implementation Details and Additional Experiments

In all experiments we use the the complex Fourier basis for the Hilbert spaces, indexed by k. This extends to the 2-dimensional case what we described in Appendix D.1. AS stated in the main paper, our practical implementation sets $f = 0$: then, we only need to specify the value for the parameters b^k, r^k. In our implementation we consider an extended class of SDEs that include time-varying multiplying coefficients in front of the drift and diffusion terms, as done for example in the Variance Preserving SDE originally described by Song & Ermon (2020). This can be simply interpreted as the time-rescaled version of autonomous SDEs.

E.1 Architectural details

In our implementation, we use the original INR architecture (Sitzmann et al., 2020). For the specific denoising task we consider in our model, we extend the input of the network architecture to include the corrupted version of the input sample and the diffusion time t, in addition to the spatial coordinates. We emphasize that our architectural is simple, and does not require self-attention mechanisms (Song & Ermon, 2020). The non-linearity we use in our network is a Gabor wavelet activation function (Saragadam et al., 2023). Furthermore, we found beneficial the inclusion of skip connections.

As stated in the main paper, we consider the *modulation* approach to INRs. In particular, we implement the meta-learning scheme described by Dupont et al. (2022b); Finn et al. (2017). The outer loop is dedicated to learning the base parameters of the model, while the inner loop focuses on refining the base parameters for each input sample. In the outer loop, the optimization algorithm is AdaBelief (Zhuang et al., 2020), sweeping the learning rate over 1e-4, 1e-5, 1e-6. We found the use of a cosine warm-up schedule to be beneficial for avoiding training instabilities and convergence to sub-optimal solutions. The inner loop is implemented by using three steps of stochastic gradient descent (SGD), where the per-parameter learning rate are found using the Meta-SGD scheme described by Dupont et al. (2022b).

E.2 Additional results

E.2.1 A Toy example.

We here present some qualitative examples on a synthetic data-set of functions $\in L([-1, 1])$, and therefore consider the settings described in Appendix D. The *Quadratic* data is generated as in (Phillips et al., 2022), i.e. $X_0[p] = qp^2 + \epsilon$, where $\epsilon \sim \mathcal{N}(0, 0.1)$ and q is a binary random variable that take values $\{-1, 1\}$ with equal probability. Concerning the design of the forward SDE, we select $b^k = \min(\sqrt{k}, 10)$ and $r^k = k^{-2}$ (thus satisfying Corollary 1). The real data is generated considering a grid of 100 equally spaced points. We can see in Figure 2 some qualitative results. On the left real (red) and generated through FDP (blue) samples show good agreement. Center and right plots depict some example of diffused samples for times 0.2 and 1.0 respectively.

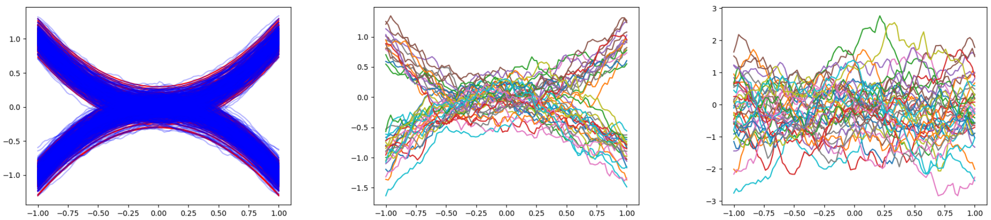

Figure 2: Left: real (red) and generated samples (blue). Center and Right: Samples diffused for times 0.2 and 1.0 respectively.

E.2.2 MNIST data-set

We evaluate our approach on a simple data-set, using MNIST 32×32 (LeCun et al., 2010). In this experiment, we compare our method against the baseline score-based diffusion models from Song et al. (2021), which we take from the official code repository `https://github.com/yang-song/score_sde`. The baseline implements the score network using a U-NET with self-attention and skip connections, as indicated by current best practices, which amounts to $O(10^8)$ parameters.

Instead, our method uses a score-network/INR implemented as a simple MLP with 8 layers and 128 neurons in each layer. The activation function is a sinusoidal non-linearity (Sitzmann et al., 2020). Our model counts $O(10^5)$ parameters. We consider an SDE with parameters $r^{k,m} = \frac{176}{k^2+m^2+2}$,[3] and $b^{k,m} = \min((k^2+m^2+0.3)^{-1} + \left(\frac{r^{k,m}}{33}\right)^{\frac{1}{4}}, 3.6)$. These values have been determined empirically by observing the power spectral density of the data-set, to ensure a well-behaved Signal to Noise ratio evolution throughout the diffusion process for all frequency components.

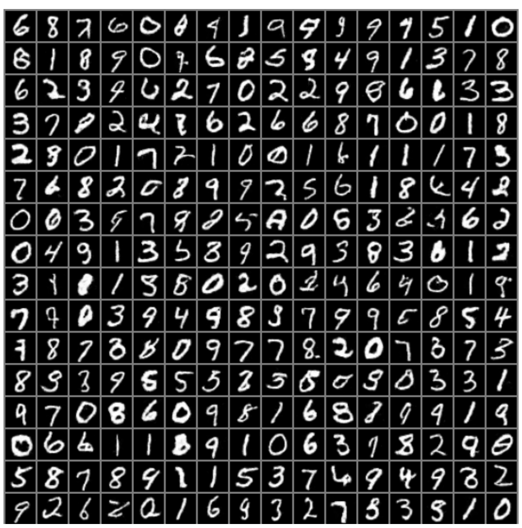

Figure 3: MNIST samples generated according to our proposed FDPs.

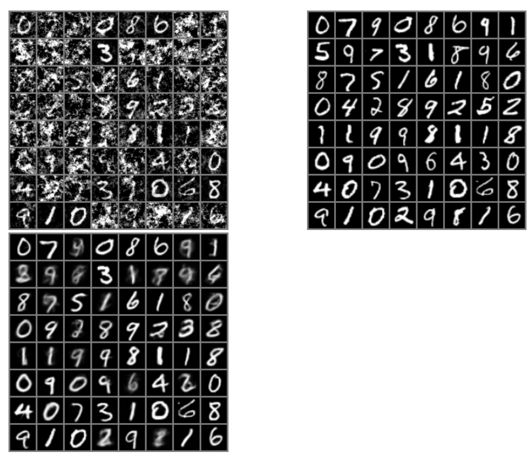

Figure 4: Top right: MNIST real samples. Top Left: Each sample is diffused for a given random time. Bottom: output of INR for corresponding input noisy image.

In Figure 3 we report un-curated samples generated according to our FDP. In Figure 4 we present instead various "intermediate" noisy versions of the training data, to illustrate the kind of noise we use to train the score network, and the output of the denoising INR. We also report the Fréchet Inception Distance (FID) score computed using 16k samples (lower is better). For the baseline we obtain FID=0.05, whereas for the proposed method we obtain FID=0.43. Although the FID score is in

[3] Strictly speaking, the sum of the series $r^{k,m}$ is not convergent. We experimented changing the decay to ensure convergence, but we observed no numerical difference with the settings we the setting we used. It is an interesting avenue for future work to study if this approximation has an impact for higher-resolution data-sets.

Figure 5: Uncurated CELEBA samples

favor of the baseline, we believe that our results – obtained with a simple MLP – are very promising, as further corroborated by experiments on a more complex dataset, which we show next.

E.2.3 CELEBA data-set

For the CELEBA data-set we considered the same SDE as for the MNIST experiment. Results reported in the main paper have been obtained using a numerical integration scheme with 300 steps of a variant of the predictor-corrector scheme of (Song & Ermon, 2020), which we adapted to the SDEs we consider in our work. In Figure 5 we report additional un-curated samples obtained with the configuration described above. We proceed to describe further experiments in the following section.

Conditional generation. In the following, we consider three use-cases for conditional generation: in-painting, de-blurring, and colorization, which we describe next. All these additional experiments were completed using the same architecture and configuration of the unconditional generation described above.

In-painting. We perform in-painting experiments by adopting the same approach described by Song & Ermon (2020), and report results in Figure 6. Original images (left-column of Figure 6) are masked (center-column of Figure 6), where we set the value corresponding to the missing pixels to 0. The right column of Figure 6 shows the results of the in-painting scheme where, qualitatively, it is possible to observe that the conditional generation is able to fill the missing portion of the images while maintaining good semantic coherence.

De-blurring. Our FDPs are naturally suited for the de-blurring use-case, as shown in Figure 8. In this experiment, we take the original images (left column of Figure 8) and filter them with a low pass filter (center column of Figure 8). The de-blurring scheme is implemented as the in-painting

Figure 6: In-painting experiment. Left: real samples, Center: Masked samples, Right: Reconstructed samples

approach described by Song & Ermon (2020), where the only difference is that the masking at each update is applied in the frequency domain. The right column of Figure 8 shows that our technique gracefully recovers missing details and is capable of producing high quality images conditioned on the distorted inputs.

Colorization. In this use-case, we adapt the approach from (Song & Ermon, 2020) to our setting. Figure 7 depicts qualitative results of the colorization experiment, confirming the flexibility of the proposed scheme.

Figure 7: Colorization experiment. Left: real samples, Center: Gray-scale samples, Right: Reconstructed samples

Figure 8: De-blurring experiment. Left: real samples, Center: blurred samples, Right: Reconstructed samples