S1 Organization of the Supplementary Material

We provide the pseudocode for the WSGM algorithm in Appendix S2. Appendix S3 contains an introduction to wavelet transforms, and their whitening properties are presented in Appendix S4. The proofs of Section 2 and Section 3 are gathered Appendix S5 and Appendix S6 respectively. Details about the Gaussian model and the φ^4 model are given in Appendix S7 and Appendix S8 respectively. Finally, experimental details and additional experiments are described in Appendix S9.

S2 WSGM Algorithm

In Algorithm 1, we provide the pseudocode for WSGM. Notice that the training of score models at each scale can be done in parallel, while the sampling is done sequentially one scale after the next.

Algorithm 1 Wavelet Score-based Generative Model

Require: J, N_{iter}, N, T,$\{\bar{\theta}_{j,0}, \theta_{J,0}\}_{j=0}^{J}$, $\{x_0^m\}_{m=1}^{M}$

/// WAVELET TRANSFORM ///
for $j \in \{1, \dots, J\}$ **do**
  for $m \in \{1, \dots, M\}$ **do**
    $x_j^m = \gamma_j^{-1} G x_{j-1}^m$, $\bar{x}_j^m = \gamma_j^{-1} \bar{G} x_{j-1}^m$ ▷ Wavelet transform of the dataset
  end for
end for
/// TRAINING ///
Train score network $s_{\theta_J^\star}$ at scale J with dataset $\{x_J^m\}_{m=0}^{M}$ ▷ Unconditional SGM training
for $j \in \{J, \dots, 1\}$ **do** ▷ Can be run in parallel
  for $n \in \{0, \dots, N_{\text{iter}} - 1\}$ **do**
    Sample $(\bar{x}_{j,0}, x_j)$ from $\{\bar{x}_j^m, x_j^m\}_{m=1}^{M}$
    Sample t in $[0, T]$ and $\bar{Z} \sim \mathrm{N}(0, \mathrm{Id})$
    $\bar{x}_{j,t} = \mathrm{e}^{-t} \bar{x}_{j,0} + (1 - \mathrm{e}^{-2t})^{1/2} \bar{Z}$
    $\ell(\bar{\theta}_{j,n}) = \|(\mathrm{e}^{-t} \bar{x}_{j,0} - \bar{x}_{j,t}) - (1 - \mathrm{e}^{-2t})^{1/2} \bar{s}_{\bar{\theta}_{j,n}}(t, \bar{x}_{j,t} | x_j)\|^2$
    $\bar{\theta}_{j,n+1} = \texttt{optimizer\_update}(\bar{\theta}_{j,n}, \ell(\bar{\theta}_{j,n}))$ ▷ ADAM optimizer step
  end for
  $\bar{\theta}_j^\star = \bar{\theta}_{j,N_{\text{iter}}}$
end for
/// SAMPLING ///
$x_J = \texttt{EulerMaruyama}(T, N, s_{\theta_J^\star})$ ▷ Euler-Maruyama recursion following (16)
for $j \in \{J, \dots, 1\}$ **do**
  $\bar{x}_j = \texttt{EulerMaruyama}(T, N, \bar{s}_{\bar{\theta}_j^\star}(\cdot, \cdot | x_j))$ ▷ Euler-Maruyama recursion following (17)
  $x_{j-1} = \gamma_j G^\top x_j + \gamma_j \bar{G}^\top \bar{x}_j$ ▷ Wavelet reconstruction
end for
return $\{\bar{\theta}_j^\star, \theta_J^\star\}_{j=1}^{J}, x_0$ ▷ Returns learned parameters and generated samples

S3 Introduction to the Fast Orthogonal Wavelet Transform

This section introduces the fast orthogonal wavelet transform introduced in [30]. It is computed with convolutional operators G and $\bar{G}$. In this section, we deal with the non-normalized wavelet transform, which is obtained by setting $\gamma_j = 1$. To avoid confusion with normalized wavelet coefficients $(x_j, \bar{x}_j)$, we denote the non-normalized wavelet coefficients with a w exponent: $(x_j^w, \bar{x}_j^w)$.

Let x_0^w be a signal. The index u in $x_0^w(u)$ belongs to an n-dimensional grid of linear size L and hence with L^n sites, with $n = 2$ for images. Let us denote x_j^w the coarse-grained version of x_0^w at a scale 2^j defined over a coarser grid with intervals 2^j and hence $(2^{-j}L)^n$ sites. The coarser signal x_j^w is iteratively computed from x_{j-1}^w by applying a coarse-graining operator, which acts as a scaling filter G which eliminates high frequencies and subsamples the grid:

$$(G x_{j-1}^w)(u) = \sum_{u'} x_{j-1}^w(u') \, G(2u - u') \; . \tag{S1}$$

The index u on the left-hand side runs on the coarser grid, whereas u' runs on the finer one.

The degrees of freedom of x_{j-1}^w that are not in x_j^w are encoded in orthogonal wavelet coefficients $\bar{x}_j^w$. The representation $(x_j^w, \bar{x}_j^w)$ is an orthogonal change of basis calculated from x_{j-1}^w. The coarse signal x_j^w is calculated in (S1) with a low-pass scaling filter G and a subsampling. In dimension n, the wavelet coefficients $\bar{x}_j^w$ have $2^n - 1$ channels computed with a convolution and subsampling operator $\bar{G}$. We thus have:

$$x_j^w = G\, x_{j-1}^w \text{ and } \bar{x}_j^w = \bar{G}\, x_{j-1}^w. \tag{S2}$$

The wavelet filter $\bar{G}$ computes $2^n - 1$ wavelet coefficients $\bar{x}_j^w(u,k)$ indexed by $1 \leq k \leq 2^n - 1$, with separable high-pass filters $\bar{G}_k(u)$:

$$\bar{x}_j^w(u,k) = \sum_{u'} x_{j-1}^w(u')\, \bar{G}_k(2u - u').$$

As an example, the Haar wavelet leads to a block averaging filter G. In dimension $n = 1$

$$x_j^w(u) = \frac{x_{j-1}^w(2u) + x_{j-1}^w(2u+1)}{\sqrt{2}},$$

and there is a single wavelet channel in $\bar{x}_j^w$. The corresponding wavelet filter $\bar{G}$ computes the wavelet coefficients with increments divided by $\sqrt{2}$:

$$\bar{x}_j^w(u) = \frac{x_{j-1}^w(2u) - x_{j-1}^w(2u+1)}{\sqrt{2}}.$$

If $n = 2$, then there are $2^n - 1 = 3$ wavelet channels as shown in Figure 1.

The fast wavelet transform cascades (S2) for $1 \leq j \leq J$ to compute the decomposition of the high-resolution signal x_0^w into its orthogonal wavelet representation over J scales:

$$\{x_J^w\, , \bar{x}_j^w\}_{1 \leq j \leq J}. \tag{S3}$$

The wavelet orthonormal filters G and $\bar{G}$ define a unitary transformation, which satisfies:

$$\bar{G}G^\top = G\bar{G}^\top = 0 \text{ and } G^\top G + \bar{G}^T\bar{G} = \mathrm{Id}\,,$$

where Id is the identity. Conjugate mirror conditions are given in [30] on the Fourier transforms of G and $\bar{G}$ to build such unitary filters. The filtering equations (S2) can then be inverted with the adjoint operators:

$$x_{j-1}^w = G^\top x_j^w + \bar{G}^\top \bar{x}_j^w\ . \tag{S4}$$

The adjoint $G^\top$ enlarge the grid size of x_j^w by inserting a zero between each coefficients, and then filters the output:

$$(G^\top x_j^w)(u) = \sum_{u'} x_j^w(u')\, G(2u' - u).$$

The adjoint of $\bar{G}$ performs the same operations over the $2^n - 1$ channels and adds them:

$$(\bar{G}^\top \bar{x}_j^w)(u) = \sum_{k=1}^{2^n-1} \sum_{u'} \bar{x}_j^w(u',k)\, \bar{G}_k(2u' - u).$$

The fast inverse wavelet transform [30] recovers x_0^w from its wavelet representation (S3) by progressively recovering x_{j-1}^w from x_j^w and $\bar{x}_j^w$ with (S4), for j going from J to 1.

S4 Orthogonal Wavelet Bases and Preconditioning of Operators

This appendix relates the fast discrete wavelet transform to decomposition of finite energy functions in orthonormal bases of $\mathbf{L^2}([0,1]^n)$. Although the covariance of normalized wavelet coefficients of multiscale processes are badly conditioned, after normalisation these covariance matrices become well conditioned because the normalisation acts as a preconditioning operator [15]. This is a central result to prove Theorem 3. The results of this appendix are based on the multiresolution theory [30, 31] and the representation of elliptic singular operators in wavelet orthonormal bases [34].

Orthonormal wavelet bases From an input discrete signal $x_0(u) = x(u)$ defined over an n-dimensional grid of width L, we introduced in (14) a normalized wavelet transform which computes wavelet coefficients $\bar{x}_j(u,k)$ having $2^n - 1$ channels $1 \leq k < 2^n$. The orthonormal wavelet transform without renormalization is obtained by setting $\gamma_j = 1$ and has been introduced in appendix S3. We write $\bar{x}^w = (\bar{x}^w_j, x^w_J)_{j \leq J}$ the vector of non-normalized wavelet coefficients.

The multiresolution wavelet theory [31, 34] proves that the coefficients of $\bar{x}^w$ can also be written as the decomposition coefficients of a finite energy function, in a wavelet orthonormal basis of the space $\mathbf{L^2}(\mathbb{R}^n)$ of finite energy functions. These wavelets arise from the cascade of the convolutional filters G and $\bar{G}$ in (**??**) when we iterate on j [31]. This wavelet orthonormal basis is thus entirely specified by the choice of the filters G and $\bar{G}$. A wavelet orthonormal basis is defined by a *scaling function* $\psi^0(v)$ for $v \in \mathbb{R}^n$ which has a unit integral $\int \psi^0(v)\,\mathrm{d}v = 1$, and $2^n - 1$ *wavelets* which have a zero integral $\int \psi^k(v)\,\mathrm{d}v = 0$ for $1 \leq k < 2^n$. Each of these functions are dilated and translated by $u \in \mathbb{Z}^n$, for $1 \leq k < 2^n$ and $j \in \mathbb{Z}$:

$$\psi^k_{j,u}(v) = 2^{-nj/2}\,\psi^k(2^{-j}v - u).$$

The main result proved in [31, 34], is that for appropriate filters G and $\bar{G}$ such that $(G, \bar{G})$ is unitary, the family of translated and dilated wavelets up to the scale 2^J:

$$\{\psi^0_{J,u}\,,\;\psi^k_{j,u}\}_{1 \leq k < 2^n\,,\,j \leq J\,,\,u \in \mathbb{Z}^n}$$

is an orthonormal basis of $\mathbf{L^2}(\mathbb{R}^n)$. A periodic wavelet basis of $\mathbf{L^2}([0,1]^n)$ is defined by replacing each wavelet $\psi^k_{j,u}$ by the periodic function $\sum_{r \in \mathbb{Z}^n} \psi^k_{j,u}(v - r)$ which we shall still write $\psi^k_{j,u}$.

The properties of the wavelets $\psi^k_{j,u}$ depend upon the choice of the filters G and $\bar{G}$. If these filters have a compact support then one can verify [31] that all wavelets $\psi^k_{j,u}$ have a compact support of size proportional to 2^j. With an appropriate choice of filters, one can also define wavelets having q vanishing moments, which means that they are orthogonal to any polynomial $Q(v)$ of degree strictly smaller than q:

$$\textstyle\int_{[0,1]^n} Q(v)\,\psi^k_{j,u}(v)\,\mathrm{d}v = 0.$$

One can also ensure that wavelet are q times continuously differentiable. Daubechies wavelets [31] are examples of orthonormal wavelets which can have q vanishing moments and be $\mathbf{C}^q$ for any q.

The relation between the fast wavelet transform and these wavelet orthonormal bases proves [31] that any discrete signal $x_0(u)$ of width L can be written as a discrete approximation at a scale $2^\ell = L^{-1}$ ($\ell < 0$) of a (non-unique) function $f \in \mathbf{L^2}([0,1]^n)$. The support of f is normalized whereas the approximation scale 2^ℓ decreases as the number of samples L increases. The coefficients $x_0(u)$ are inner products of f with the orthogonal family of scaling functions at the scale 2^ℓ for all $u \in \mathbb{Z}^n$ and $2^\ell u \in [0,1]^n$:

$$x_0(u) = \textstyle\int_{[0,1]^n} f(v)\,\psi^0_{\ell,u}(v)\,\mathrm{d}v = \langle f, \psi^0_{\ell,u}\rangle.$$

Let V_ℓ be the space generated by the orthonormal family of scaling functions $\{\psi^0_{\ell,u}\}_{2^\ell u \in [0,1]^n}$, and $P_{V_\ell} f$ be the orthogonal projection of f in V_ℓ. The signal x_0 gives the orthogonal decomposition coefficients of $P_{V_\ell} f$ in this family of scaling functions. One can prove [31] that the non-normalized wavelet coefficients $\bar{x}^w_j$ of x_0 computed with a fast wavelet transform are equal to the orthogonal wavelet coefficients of f at the scale $2^{j+\ell}$, for all $u \in \mathbb{Z}^n$ and $2^{j+\ell} u \in [0,1]^n$:

$$\bar{x}^w_j(u,k) = \textstyle\int_{[0,1]^n} f(v)\,\psi^k_{j+\ell,u}(v)\,\mathrm{d}v = \langle f, \psi^k_{j+\ell,u}\rangle.$$

and at the largest scale 2^J

$$x^w_J(u,k) = \textstyle\int_{[0,1]^n} f(v)\,\psi^0_{J+\ell,u}(v)\,\mathrm{d}v = \langle f, \psi^k_{j+J,u}\rangle.$$

Normalized covariances We now consider a periodic stationary multiscale random process $x(u)$ of width L. It covariance is diagonalised in a Fourier basis and its power spectrum (eigenvalues) has a power-law decay $P(\omega) = c(\xi^\eta + |\omega|^\eta)^{-1}$, for frequencies $\omega = 2\pi m/L$ with $m \in \{0, \ldots, L-1\}^n$. The following lemma proves that the covariance matrix $\bar{\Sigma}$ of the normalized wavelet coefficients $\bar{x}$ of x is well conditioned, with a condition number which does not depend upon L. It relies on an equivalence between Sobolev norms and weighted norms in a wavelet orthonormal basis.

Lemma S4. *For a wavelet transform corresponding to wavelets having $q > \eta$ vanishing moments, which have a compact support and are q times continuously differentiable, there exists $C_2 \geq C_1 > 0$ such that for any L the covariance $\bar{\Sigma}$ of $\bar{x} = (\bar{x}_j, x_J)_{j \leq J}$ satisfies:*

$$C_1 \; \mathrm{Id} \leq \bar{\Sigma} \leq C_2 \; \mathrm{Id}\,. \tag{S5}$$

The remaining of the appendix is a proof of this lemma. Without loss of generality, we shall suppose that $\mathbb{E}[x] = 0$. Let $\sigma_{j,k}^2$ be the variance of $\bar{x}_j^w(u,k)$, and D be the diagonal matrix whose diagonal values are $\sigma_{j,k}^{-1}$. The vector of normalized wavelet coefficients $\bar{x} = (\bar{x}_j, x_J)_{j \leq J}$ are related to the non-normalized wavelet coefficients $\bar{x}^w$ by a multiplication by D:

$$\bar{x} = D\,\bar{x}^w.$$

Let $\bar{\Sigma}_w$ be the covariance of $\bar{x}^w$. It results from this equation that the covariance $\bar{\Sigma}$ of $\bar{x}$ and the covariance $\bar{\Sigma}_w$ of $\bar{x}^w$ satisfy:

$$\bar{\Sigma} = D\bar{\Sigma}_w D.$$

The diagonal normalization D is adjusted so that the variance of each coefficient of $\bar{x}$ is equal to 1, which implies that the diagonal of $\bar{\Sigma}$ is the identity. We must now prove that $\bar{\Sigma}$ satisfies (S5), which is equivalent to prove that there exists C_1 and C_2 such that:

$$C_1 \; \mathrm{Id} \leq D\bar{\Sigma}_w D \leq C_2 \; \mathrm{Id}\,. \tag{S6}$$

To prove (S6), we relate it to Sobolev norm equivalences that have been proved in harmonic analysis. We begin by stating the result on Sobolev inequalities and then prove that it implies (S6) for appropriate constants C_1 and C_2.

Let Σ_∞ be the singular self-adjoint convolutional operator over $\mathbf{L^2}(\mathbb{R}^n)$ defined in the Fourier domain for all $\omega \in \mathbb{R}^n$:

$$\widehat{\Sigma_\infty f}(\omega) = \hat{f}(\omega)\,(\xi^\eta + |\omega|^\eta).$$

Observe that:

$$\langle \Sigma_\infty f, f\rangle = \tfrac{1}{(2\pi)^n} \int_{\mathbb{R}^n} |\hat{f}(\omega)|^2\,(\xi^\eta + |\omega|^\eta)\,\mathrm{d}\omega$$

is a Sobolev norm of exponent η. Such Sobolev norms are equivalent to weighted norms in wavelet bases, as proved in Theorem 4, Chapter 3 in [34]. To take into account the constant ξ, we introduce a maximum scale $2^{J'} = \xi^{-1}$. For all $f \in \mathbf{L^2}(\mathbb{R}^n)$, there exists $B \geq A > 0$ such that:

$$\begin{aligned} A\,\langle \Sigma_\infty f, f\rangle \leq \textstyle\sum_{u\in\mathbb{Z}^n} 2^{-J\eta} |\langle f, \psi^0_{J',u}\rangle|^2 \\ + \textstyle\sum_{j=-\infty}^{J'} \sum_{u\in\mathbb{Z}^n} \sum_{k=1}^{2^n-1} 2^{-j\eta}\,|\langle f, \psi^k_{j,u}\rangle|^2 \leq B\,\langle \Sigma_\infty f, f\rangle. \end{aligned} \tag{S7}$$

The remaining of the proof shows that these inequalities imply similar inequalities over the covariance of discrete wavelet coefficients. This is done by first restricting it to a finite support and then using the correspondence between the orthonormal wavelet coefficients of f and the discrete wavelet coefficients $\bar{x}^w$ of x_0.

One can verify that the equivalence (S7) remains valid for functions $f \in \mathbf{L^2}([0,1]^n)$ decomposed over periodic wavelet bases, because functions in $\mathbf{L^2}([0,1]^n)$ can be written $f(v) = \sum_{r\in\mathbb{Z}^n} \tilde{f}(v-r)$ with $\tilde{f} \in \mathbf{L^2}(\mathbb{R}^n)$:

$$\begin{aligned} A\,\langle \Sigma_\infty f, f\rangle \leq \textstyle\sum_{2^j u\in[0,1]^n} 2^{-J'\eta} |\langle f, \psi^0_{J',u}\rangle|^2 \\ + \textstyle\sum_{j=-\infty}^{J'} \sum_{2^j u\in[0,1]^n} \sum_{k=1}^{2^n-1} 2^{-j\eta}\,|\langle f, \psi^k_{j,u}\rangle|^2 \leq B\,\langle \Sigma_\infty f, f\rangle. \end{aligned}$$

Applying this result to $f \in V_\ell$ is equivalent to restricting Σ_∞ to V_ℓ, which proves that $\Sigma_\ell = P_{V_\ell}\Sigma_\infty P_{V_\ell}$ satisfies:

$$\begin{aligned} A\,\langle \Sigma_\ell f, f\rangle \leq \textstyle\sum_{2^{J'} u\in[0,1]^n} 2^{-J'\eta} |\langle f, \psi^0_{J',u}\rangle|^2 \\ + \textstyle\sum_{j=\ell+1}^{J'} \sum_{2^j u\in[0,1]^n} \sum_{k=1}^{2^n-1} 2^{-j\eta}\,|\langle f, \psi^k_{j,u}\rangle|^2 \leq B\,\langle \Sigma_\ell f, f\rangle. \end{aligned} \tag{S8}$$

The operator $\Sigma_\ell = P_{V_\ell}\,\Sigma_\infty\,P_{V_\ell}$ is covariant with respect to shifts by any $m2^\ell$ for $m \in \mathbb{Z}^n$ because P_{V_ℓ} and Σ_∞ are covariant to such shifts. Its representation in the basis of scaling functions

$\{\psi^0_{\ell,u}\}_{2^\ell u\in[0,1]^n}$ is thus a Toeplitz matrix which is diagonalized by a Fourier transform. There exists $0 < A_1 \le B_1$ such that for all $\ell < 0$ and all $\omega \in [-2^{-\ell}\pi, 2^{-\ell}\pi]^n$,

$$A_1\,(\xi^\eta + |\omega|^\eta) \le P_\ell(\omega) \le B_1\,(\xi^\eta + |\omega|^\eta). \tag{S9}$$

Indeed, the spectrum of Σ_∞ is $c\,(\xi^\eta + |\omega|^\eta)$ for $\omega \in \mathbb{R}^n$ and P_{V_ℓ} performs a filtering with the scaling function ψ^0_ℓ whose support is essentially restricted to the frequency interval $[-\pi 2^{-\ell}, \pi 2^{-\ell}]$ so that the spectrum of $P_{V_\ell}\,\Sigma_\infty\,P_{V_\ell}$ is equivalent to the spectrum of Σ_∞ restricted to this interval.

The lemma hypothesis supposes that the covariance $\widetilde{\Sigma}$ of x_0 has a spectrum equal to $c\,(\xi^\eta + |\omega|^\eta)^{-1}$ and hence that the spectrum of $\widetilde{\Sigma}^{-1}$ is $c^{-1}\,(\xi^\eta + |\omega|^\eta)$. Since x_0 are decomposition coefficients of $f \in V_\ell$ in the basis of scaling functions, equation (S9) can be rewritten for any $f \in V_\ell$:

$$A_1\,c\,\langle \widetilde{\Sigma}^{-1} x_0, x_0\rangle \le \langle \Sigma_\ell f, f\rangle \le B_1\,c\,\langle \widetilde{\Sigma}^{-1} x_0, x_0\rangle. \tag{S10}$$

Since the orthogonal wavelet coefficients $\bar{x}^w$ defines an orthonormal representation of x_0, the covariance $\bar{\Sigma}_w$ of $\bar{x}^w$ satisfies $\langle \bar{\Sigma}_w^{-1}\bar{x}^w, \bar{x}^w\rangle = \langle \widetilde{\Sigma}^{-1} x_0, x_0\rangle$. Moreover, we saw that that the wavelet coefficients $\bar{x}^w$ of x_0 satisfy $\bar{x}^w_j(u,k) = \langle f, \psi^k_{j+\ell,u}\rangle$ and at the largest scale $\bar{x}^w_J(u,k) = \langle f, \psi^0_{J+\ell,u}\rangle$. Hence for $J + \ell = J'$, we derive from (S8) and (S10) that:

$$\begin{aligned} A\,A_1\,c\,\langle \bar{\Sigma}_w^{-1}\bar{x}^w, \bar{x}^w\rangle &\le \textstyle\sum_{2^J u\in[0,1]^n} 2^{-(J+\ell)\eta}\,|x^w_J(u)|^2 \\ &\quad + \textstyle\sum_{j=1}^{J}\sum_{2^j u\in[0,1]^n}\sum_{k=1}^{2^n-1} 2^{-(j+\ell)\eta}\,|\bar{x}^w_j(u,k)|^2 \le B\,B_1\,c\,\langle \bar{\Sigma}_w^{-1}\bar{x}^w, \bar{x}^w\rangle. \end{aligned}$$

It results that for $A_2 = A\,A_1\,c$ and $B_2 = B\,B_1\,c$ we have:

$$A_2\,\langle \bar{\Sigma}_w^{-1}\bar{x}^w, \bar{x}^w\rangle \le 2^{-(J+\ell)\eta}\,\|x^w_J\|^2 + \sum_{j=1}^{J} 2^{-(j+\ell)\eta}\,\|\bar{x}^w_j\|^2 \le B_2\,\langle \bar{\Sigma}_w^{-1}\bar{x}^w, \bar{x}^w\rangle.$$

Let $\tilde{D}$ be the diagonal operator over the wavelet coefficients $\bar{x}^w$, whose diagonal values are $2^{-\eta(j+\ell)/2}$ at all scales 2^j. These inequalities can be rewritten as operator inequalities:

$$A_2\,\bar{\Sigma}_w^{-1} \le \tilde{D}^2 \le B_2\,\bar{\Sigma}_w^{-1},$$

and hence:

$$A_2\,\mathrm{Id} \le \tilde{D}\bar{\Sigma}_w\tilde{D} \le B_2\,\mathrm{Id}\,. \tag{S11}$$

Since D^{-2} is the diagonal of $\bar{\Sigma}_w$, we derive from (S11) that:

$$A_2\,\tilde{D}^{-2} \le D^{-2} \le B_2\,\tilde{D}^{-2}.$$

Inserting this equation in (S11) proves that:

$$A_2\,B_2^{-1}\;\mathrm{Id} \le D\,\bar{\Sigma}_w\,D \le B_2\,A_2^{-1}\;\mathrm{Id},$$

and since $\bar{\Sigma} = D\,\bar{\Sigma}_w\,D$ it proves the lemma result (S6), with $C_1 = A_2 B_2^{-1}$ and $C_2 = B_2 A_2^{-1}$.

S5 Proof of Theorems 1 and 2

In this section, we first present the continuous-time framework in a Gaussian setting in Appendix S5.1. The general outline of the proof of Theorem 1 is presented in Appendix S5.2. Technical lemmas are gathered in Appendix S5.3. The proof of Theorem 2 is presented in Appendix S5.4.

S5.1 Gaussian setting

In what follows we present the Gaussian setting used in the proof of Theorem 1. We assume that $p_0 = \mathrm{N}(0, \Sigma)$ with $\Sigma \in \mathbb{S}_d(\mathbb{R})_+$. Let $\mathrm{D} \in \mathcal{M}_d(\mathbb{R})_+$ a diagonal positive matrix such that $\Sigma = \mathrm{P}^\top \mathrm{D}\mathrm{P}$ with P an orthonormal matrix. We consider the following forward dynamics

$$\mathrm{d}x_t = -x_t\mathrm{d}t + \sqrt{2}\mathrm{d}w_t,$$

with $\mathcal{L}(x_0) = p_0$. We also consider the backward dynamics given by

$$\mathrm{d}y_t = \{y_t + 2\nabla \log p_{T-t}(y_t)\}\mathrm{d}t + \sqrt{2}\mathrm{d}w_t,$$

with $\mathcal{L}(y_0) = p_\infty = \mathrm{N}(0, \mathrm{Id})$. Note that since for any $t \in [0, T]$ and $x \in \mathbb{R}^d$, $\nabla \log p_t(x) = -\Sigma_t^{-1} x$ with $\Sigma_t = \exp[-2t]\Sigma + (1 - \exp[-2t])\,\mathrm{Id}$, we have that $(y_t)_{t\in[0,T]}$ is a Gaussian process. In particular, we can compute the mean and the covariance matrix of y_t in a closed form for any $t \in [0, T]$. The results of Proposition S5 will not be used to prove Theorem 1. However, they provide some insights regarding the evolution of the mean and covariance of the backward process.

Proposition S5. *For any $t \in [0, T]$, we have that $\mathcal{L}(y_t) = \mathrm{N}(0, \bar{\Sigma}_t)$ with*

$$\bar{\Sigma}_t = \mathrm{P}^\top((1 - \exp[-2t])\bar{\mathrm{D}}_t + \exp[-2t]\bar{\mathrm{D}}_t^2)\mathrm{P},$$

and

$$\bar{\mathrm{D}}_t = (\mathrm{Id} + (\mathrm{D} - \mathrm{Id})\exp[-2(T-t)]) \oslash (\mathrm{Id} + (\mathrm{D} - \mathrm{Id})\exp[-2T])\ .$$

Note that $\bar{\mathrm{D}}_0 = \mathrm{Id}$ and $\bar{\mathrm{D}}_T = \mathrm{D} \oslash (\mathrm{Id} + (\mathrm{D} - \mathrm{Id})\exp[-2T]) \approx \mathrm{D}$. Hence, we have $\bar{\Sigma}_T \approx \Sigma$ and therefore $\mathcal{L}(y_T) \approx p_0$.

Proof. First, note that for any $t \in [0, T]$ we have that

$$y_t = y_0 + \textstyle\int_0^t (\mathrm{Id} - 2\Sigma_{T-t}^{-1})y_t + \sqrt{2}w_t = y_0 + \int_0^t (\mathrm{Id} - 2\mathrm{P}^\top D_{T-t}^{-1}\mathrm{P})y_t + \sqrt{2}w_t,$$

with $\mathrm{D}_{T-t} = \exp[-2(T-t)]\mathrm{D} + (1 - \exp[-2(T-t)])\,\mathrm{Id}$. Denote $\{y_t^{\mathrm{P}}\}_{t\in[0,T]} = \{\mathrm{P}y_t\}_{t\in[0,T]}$. Using that $\mathrm{P}^\top\mathrm{P} = \mathrm{Id}$, we have that for any $t \in [0, T]$

$$y_t^{\mathrm{P}} = y_0^{\mathrm{P}} + \textstyle\int_0^t (\mathrm{Id} - 2D_{T-t}^{-1})y_t^{\mathrm{P}} + \sqrt{2}w_t^{\mathrm{P}},$$

where $\{w_t^{\mathrm{P}}\}_{t\in[0,T]} = \{\mathrm{P}w_t\}_{t\in[0,T]}$. Note that since P is orthonormal, $\{w_t^{\mathrm{P}}\}_{t\in[0,T]}$ is also a d-dimensional Brownian motion. We also have that $\mathcal{L}(y_0^{\mathrm{P}}) = \mathrm{N}(0, \mathrm{Id})$. Hence for any $\{\{y_t^{\mathrm{P},i}\}_{t\in[0,T]}\}_{i=1}^d$ is a collection of d independent Gaussian processes, where for any $i \in \{1, \dots, d\}$ and $t \in [0, T]$, $y_t^{\mathrm{P},i} = \langle y_t^{\mathrm{P}}, e_i\rangle$ and $\{e_i\}_{i=1}^d$ is the canonical basis of $\mathbb{R}^d$. Let $i \in \{1, \dots, d\}$ and for any $t \in [0, T]$ denote $u_t^i = \mathbb{E}[y_t^{\mathrm{P},i}]$ and $v_t^i = \mathbb{E}[(y_t^{\mathrm{P},i})^2]$. We have that for any $t \in [0, T]$, $\partial_t u_t^i = (1 - 1/\mathrm{D}_t^i)u_t^i$ with $u_0 = 0$ and $\mathrm{D}_t^i = \exp[-2t]\mathrm{D}_i + 1 - \exp[-2t]$. Hence, we get that for any $t \in [0, T]$, $u_t^i = 0$. Using Itô's lemma we have that

$$\partial v_t^i = \{2 - 4/\mathrm{D}_{T-t}^i\}v_t^i + 2, \tag{S12}$$

with $v_0^i = 1$. Denote $\alpha_T^i = (\mathrm{D}^i - 1)\exp[-2T]$, we have that for any $t \in [0, T]$, $\mathrm{D}_{T-t}^i = 1 + \alpha_T^i \exp[2t]$. Therefore, we get that for any $t \in [0, T]$

$$2 - 4/\mathrm{D}_{T-t}^i = -2 + 2 \times (2\alpha_T^i)\exp[2t]/(1 + \alpha_T^i\exp[2t]) = -2 + 2\partial_t \log(1 + \alpha_T^i \exp[2t]).$$

Hence, we have that for any $t \in [0, T]$

$$\textstyle\int_0^t 2 - 4/\mathrm{D}_{T-s}^i \mathrm{d}s = -2t + \log((1 + \alpha_T^i\exp[2t])^2/(1 + \alpha_T^i)^2).$$

Hence, there exists $C_t^i \in \mathrm{C}^1([0, T], \mathbb{R})$ such that for any $t \in [0, T]$, $v_t^i = C_t^i \exp[-2t](1 + \alpha_T^i\exp[2t])^2/(1 + \alpha_T^i)^2$. Using (S12), we have that for any $t \in [0, T]$

$$\partial_t C_t^i = 2\exp[2t]((1 + \alpha_T^i\exp[2t])/(1 + \alpha_T^i))^{-2} = -(1/\alpha_T^i)(1 + \alpha_T^i)^2 \partial_t (1 + \alpha_T^i\exp[2t])^{-1}.$$

Hence, we have that for any $t \in [0, T]$

$$C_t^i = (1/\alpha_T^i)(1 + \alpha_T^i)^2[(1 + \alpha_T^i)^{-1} - (1 + \alpha_T^i\exp[2t])^{-1}] + A,$$

with $A \geq 0$. Hence, we get that for any $t \in [0, T]$

$$\begin{aligned} v_t^i = {} & (1/\alpha_T^i)\exp[-2t](1 + \alpha_T^i\exp[2t])[(1 + \alpha_T^i\exp[2t])/(1 + \alpha_T^i) - 1] \\ & + A\exp[-2t](1 + \alpha_T^i\exp[2t])^2/(1 + \alpha_T^i)^2. \end{aligned}$$

In addition, we have that $v_0^i = 1$ and therefore $A = 1$. Therefore, for any $t \in [0, T]$ we have

$$\begin{aligned} v_t^i &= (1/\alpha_T^i)\exp[-2t](1 + \alpha_T^i\exp[2t])[(1 + \alpha_T^i\exp[2t])/(1 + \alpha_T^i) - 1] \\ &\qquad + \exp[-2t](1 + \alpha_T^i\exp[2t])^2/(1 + \alpha_T^i)^2 \\ &= (1 - \exp[-2t])(1 + \alpha_T^i\exp[2t])/(1 + \alpha_T^i) + \exp[-2t](1 + \alpha_T^i\exp[2t])^2/(1 + \alpha_T^i)^2, \end{aligned}$$

which concludes the proof. □

S5.2 Convergence results for the discretization

In what follows, we denote $(Y_k)_{k\in\{0,\dots,N-1\}} = (\bar{x}_{t_k})_{k\in\{0,\dots,N-1\}}$, the sequence given by (6). The following result gives an expansion of the covariance matrix and the mean of Y_N, i.e. the output of SGM, in the case where $p = \mathrm{N}(\mu, \Sigma)$.

Theorem S6. *Let $N \in \mathbb{N}$, $\delta > 0$ and $T = N\delta$. Then, we have that $\bar{x}_{t_N} \sim \mathrm{N}(\hat{\mu}_N, \hat{\Sigma}_N)$ with*

$$\hat{\Sigma}_N = \Sigma + \exp[-4T]\hat{\Sigma}_T + \delta\hat{\mathrm{E}}_T + \delta^2\mathrm{R}_{T,\delta} \ , \qquad \hat{\mu}_N = \mu + \exp[-2T]\hat{\mu}_T + \delta\hat{e}_T + \delta^2 r_{T,\delta} \ ,$$

where $\hat{\Sigma}_T, \hat{\mathrm{E}}_T, \mathrm{R}_{T,\delta} \in \mathbb{R}^{d\times d}$, $\hat{\mu}_T, \hat{e}_T, r_{T,\delta} \in \mathbb{R}^d$ and $\|\mathrm{R}_{T,\delta}\| + \|r_{T,\delta}\| \leq R$ not dependent on $T \geq 0$ and $\delta > 0$. We have that

$$\begin{aligned}\hat{\Sigma}_T &= -(\Sigma - \mathrm{Id})(\Sigma\Sigma_T^{-1})^2 \ , \\ \hat{\mathrm{E}}_T &= \mathrm{Id} - (1/2)\Sigma^2(\Sigma - \mathrm{Id})^{-1}\log(\Sigma) + \exp[-2T]\tilde{\mathrm{E}}_T \ .\end{aligned} \tag{S13}$$

In addition, we have

$$\begin{aligned}\hat{\mu}_T &= -\Sigma_T^{-1}\Sigma\mu \ , \\ \hat{e}_T &= \{-2\Sigma^{-1} - (1/4)\Sigma(\Sigma - \mathrm{Id})^{-1}\log(\Sigma)\}\mu + \exp[-2T]\tilde{\mu}_T \ ,\end{aligned}$$

with $\tilde{\mathrm{E}}_T, \tilde{\mu}_T$ bounded and not dependent on T.

Before turning to the proof of Theorem S6, we state a few consequences of this result.

Corollary S7. *Let $\{\bar{x}_{t_k}\}_{k=0}^N$ the sequence defined by (6). We have that $\bar{x}_{t_N} \sim \mathrm{N}(\mu_N, \Sigma_N)$ with*

$$\begin{aligned}\Sigma_N &= \Sigma + \delta\Sigma_\delta + \exp[-4T]\Sigma_T + \Sigma_{\delta,T} \ , \\ \mu_N &= \mu + \delta\mu_\delta + \exp[-2T]\mu_T + \mu_{\delta,T} \ ,\end{aligned}$$

with

$$\begin{aligned}\Sigma_T &= -(\Sigma - \mathrm{Id})\Sigma^2 \ , \\ \Sigma_\delta &= \mathrm{Id} - (1/2)\Sigma^2(\Sigma - \mathrm{Id})^{-1}\log(\Sigma) \ , \\ \mu_T &= \Sigma\mu, \\ \mu_\delta &= \{-2\Sigma^{-1} - (1/4)\Sigma(\Sigma - \mathrm{Id})^{-1}\log(\Sigma)\}\mu \ .\end{aligned}$$

In addition, we have $\lim_{\delta\to 0, T\to+\infty} \|\Sigma_{\delta,T}\|/(\delta + \exp[-4T]) = 0$ and $\lim_{\delta\to 0, T\to+\infty} \|\mu_{\delta,T}\|/(\delta + \exp[-2T]) = 0$.

At first sight, it might appear surprising that Σ^{-1} does not appear in Σ_T and μ_T. Note that in the extreme case where $\Sigma = 0$ and $\delta \to 0$, i.e. we only consider the error associated with the fact that $T \neq +\infty$, then we have no error. This is because in this case the associated continuous-time process is an Ornstein-Uhlenbeck bridge which has distribution $\mathrm{N}(\mu, 0)$ at time T.

We will use the following result.

Lemma S8. *Let $\pi_i = \mathrm{N}(\mu_i, \Sigma_i)$ for $i \in \{0, 1\}$, with $\mu_0, \mu_1 \in \mathbb{R}^d$ and $\Sigma_0, \Sigma_1 \in \mathbb{S}_d(\mathbb{R})_+$. Then, we have that*

$$\mathrm{KL}(\pi_0\|\pi_1) = (1/2)\{\log(\det(\Sigma_1)/\det(\Sigma_0)) - d + \mathrm{Tr}(\Sigma_1^{-1}\Sigma_0) + (\mu_1 - \mu_0)^\top\Sigma_1^{-1}(\mu_1 - \mu_0)\}.$$

In particular, applying Lemma S8 we have that for any $\Sigma \in \mathbb{S}_d(\mathbb{R})_+$

$$\mathrm{KL}(\mathrm{N}(0, \Sigma)\|\mathrm{N}(0, \mathrm{Id})) = (1/2)\{-\log(\det(\Sigma)) + \mathrm{Tr}(\Sigma) - d\}. \tag{S14}$$

Proposition S9. *Let $\{\bar{x}_{t_k}\}_{k=0}^N$ the sequence defined by (6). We have that $\bar{x}_{t_N} \sim \mathrm{N}(\mu_N, \Sigma_N)$, with μ_N, Σ_N given by Corollary S7. We have that*

$$\mathrm{KL}(\mathrm{N}(\mu, \Sigma)\|\mathrm{N}(\mu_N, \Sigma_N)) \leq \delta|\mathrm{Tr}(\Sigma^{-1}\Sigma_\delta)| + \exp[-4T]|\mathrm{Tr}(\Sigma^{-1}\Sigma_T)| + \exp[-4T]\mu^\top\Sigma\mu + E_{T,\delta} \ ,$$

with $E_{T,\delta}$ a higher order term such that $\lim_{T\to+\infty,\delta\to 0} E_{T,\delta}/(\delta + \exp[-4T]) = 0$.

We now prove Theorem S6.

Proof. For any k, denote $Y_k = \bar{x}_{t_{N-k}}$. First, we recall that for any $k \in \{0, \dots, N-1\}$ and $x \in \mathbb{R}^d$, $\nabla \log p_{T-k\gamma}(x) = -\Sigma_{T-k\gamma}^{-1} x$ where for any $t \in [0,T]$

$$\Sigma_t = (1 - \exp[-2t]) \operatorname{Id} + \exp[-2t]\Sigma \ .$$

Hence, we get that for any $k \in \{0, \dots, N-1\}$

$$Y_{k+1} = ((1+\gamma)\operatorname{Id} - 2\gamma\Sigma_{T-k\gamma}^{-1})Y_k + 2\gamma\Sigma_{T-k\gamma}^{-1} M_{T-k\gamma} + \sqrt{2\gamma} Z_{k+1} \ , \tag{S15}$$

where for any $t \in [0,T]$, $M_t = \exp[-t]\mu$. Therefore, we get that for any $k \in \{0, \dots, N\}$, Y_k is a Gaussian random variable. Using (S23), we have that for any $k \in \{0, \dots, N-1\}$

$$\mathbb{E}[\hat{Y}_{k+1}\hat{Y}_{k+1}^\top] = ((1+\gamma)\operatorname{Id} - 2\gamma\Sigma_{T-k\gamma}^{-1})\mathbb{E}[\hat{Y}_k\hat{Y}_k^\top]((1+\gamma)\operatorname{Id} - 2\gamma\Sigma_{T-k\gamma}^{-1}) + 2\gamma\operatorname{Id} \ , \tag{S16}$$

where for any $k \in \{0, \dots, N\}$, $\hat{Y}_k = Y_k - \mathbb{E}[Y_k]$. There exists $\mathrm{P} \in \mathbb{R}^{d\times d}$ orthogonal such that $\mathrm{D} = \mathrm{P}\Sigma\mathrm{P}^\top$ is diagonal. Note that for any $k \in \{0, \dots, N-1\}$, we have that $\Lambda_k = \mathrm{P}((1+\gamma)\operatorname{Id} - 2\gamma\Sigma_{T-k\gamma}^{-1})\mathrm{P}^\top$ is diagonal. For any $k \in \{0, \dots, N\}$, define $\mathrm{H}_k = \mathrm{P}\mathbb{E}[\hat{Y}_k\hat{Y}_k^\top]\mathrm{P}^\top$. Note that $\mathrm{H}_0 = \operatorname{Id}$. Using (S16), we have that for any $k \in \{0, \dots, N-1\}$

$$\mathrm{H}_{k+1} = \Lambda_k^2 \mathrm{H}_k + 2\gamma\operatorname{Id} \ . \tag{S17}$$

Hence, for any $k \in \{0, \dots, N\}$, H_k is diagonal. For any diagonal matrix $C \in \mathbb{R}^{d\times d}$ denote $\{c^1, \dots, c^d\}$ its diagonal elements. Let $i \in \{1, \dots, d\}$. Using (S17), we have that for any $k \in \{0, \dots, N-1\}$

$$h_{k+1}^i = (\lambda_k^i)^2 h_k^i + 2\gamma \ .$$

Using this result we have that for any $k \in \{0, \dots, N\}$

$$h_k^i = (\textstyle\prod_{\ell=0}^{k-1}\lambda_\ell^i)^2 + 2\gamma\sum_{\ell=0}^{k-1}(\prod_{j=0}^{\ell-1}\lambda_{k-1-j}^i)^2 = (\prod_{\ell=0}^{k-1}\lambda_\ell^i)^2 + 2\gamma\sum_{\ell=0}^{k-1}(\prod_{j=k-\ell}^{k-1}\lambda_j^i)^2 \ .$$

Let $k_1, k_2 \in \{0, \dots, N\}$ with $k_1 < k_2$. In what follows, we derive an expansion of $I_{k_1,k_2} = \prod_{k=k_1}^{k_2}\lambda_k^i$ w.r.t. $\gamma > 0$. We have that

$$I_{k_1,k_2} = \textstyle\prod_{k=k_1}^{k_2}\lambda_k^i = \exp[\sum_{k=k_1}^{k_2}\log(\lambda_k^i)] = \exp[\sum_{k=k_1}^{k_2}\log(1+\gamma a_k^i)] \ , \tag{S18}$$

where for any $k \in \{0, \dots, N\}$, $a_k^i = 1 - 2/d_{(N-k)\gamma}^i$, with $d_{(N-k)\gamma}^i = 1 + \exp[-2(N-k)\gamma](d^i - 1)$. Hence, there exist $(b_{k,\gamma}^i)_{k\in\{0,\dots,N\}}$ bounded such that for any $k \in \{0, \dots, N\}$ we have

$$\log(1+\gamma a_k^i) = \gamma a_k^i - (\gamma^2/2)(a_k^i)^2 + \gamma^3 b_{k,\gamma}^i \ .$$

In addition, using Proposition S10, there exists $C_{k_1,k_2}^\gamma \geq 0$ such that $\gamma C_{k_1,k_2}^\gamma \leq C$ with $C \geq 0$ not dependent on $k_2, k_2 \in \{0, \dots, N\}$, $\gamma > 0$ and

$$\textstyle\sum_{k=k_1}^{k_2}\log(1+\gamma a_k^i) = \int_{t_1}^{t_2^+} a^i(t)\mathrm{d}t - (\gamma/2)[\int_{t_1}^{t_2^+} a^i(t)^2\mathrm{d}t + a^i(t_2^+) - a^i(t_1)] + C_{k_1,k_2}^\gamma\gamma^3 \ ,$$

with $t_1 = k_1\gamma$, $t_2^+ = (k_2+1)\gamma$ and for any $t \in [0,T]$, $a_t = 1 - 2/d_{T-t}^i$ with $d_{T-t}^i = 1 + \exp[-2(T-t)](d^i - 1)$. Hence, using this result and (S18), we get that there exists $D_{k_1,k_2}^\gamma \geq 0$ such that $\gamma D_{k_1,k_2}^\gamma \leq D$ with $D \geq 0$ not dependent on $k_2, k_2 \in \{0, \dots, N\}$, $\gamma > 0$ and

$$I_{k_1,k_2} = \exp[\textstyle\int_{t_1}^{t_2^+} a^i(t)\mathrm{d}t] - \exp[\int_{t_1}^{t_2^+} a^i(t)](\gamma/2)[\int_{t_1}^{t_2^+} a^i(t)^2\mathrm{d}t + a^i(t_2^+) - a^i(t_1)] + \gamma^3 D_{k_1,k_2}^\gamma \ . \tag{S19}$$

Using this result, we get that there exists $E_1^\gamma \geq 0$ such that $\gamma E_1^\gamma \leq E$ with $E \geq 0$ not dependent on γ such that

$$(\textstyle\prod_{\ell=0}^{N-1}\lambda_\ell^i)^2 = \exp[2\int_0^T a^i(t)\mathrm{d}t] - \gamma\exp[2\int_0^T a^i(t)\mathrm{d}t][\int_0^T a^i(t)^2\mathrm{d}t + a^i(T) - a^i(0)] + \gamma^3 E_1^\gamma \ . \tag{S20}$$

Similarly, using (S19), there exist $E \geq 0$ and $(E_{2,\ell}^\gamma)_{\ell\in\{0,\dots,N\}}$ such that for any $\ell \in \{0, \dots, N\}$, $E_{2,\ell}^\gamma \geq 0$ and $\gamma E_{2,\ell}^\gamma \leq E$ with $E \geq 0$ not dependent on γ and ℓ such that

$$\begin{aligned} 2\gamma\textstyle\sum_{\ell=0}^{N-1}(\prod_{j=N-\ell}^{N-1}\lambda_j^i)^2 = (2\gamma)\sum_{\ell=0}^{N-1}\exp[2\int_{T-\ell\gamma}^T a^i(t)\mathrm{d}t]\} \\ -2\gamma^2\textstyle\sum_{\ell=0}^{N-1}\{\exp[2\int_{T-\ell\gamma}^T a^i(t)\mathrm{d}t][\int_{T-\ell\gamma}^T a^i(t)^2\mathrm{d}t + a^i(T) - a^i(T-\ell\gamma)]\} \\ +\gamma^4\textstyle\sum_{\ell=0}^{N-1}E_{2,\ell}^\gamma \ . \end{aligned}$$

Therefore, using Proposition S10, there exists E_3^γ such that $\gamma E_3^\gamma \leq E$ with $E \geq 0$ not dependent on γ and

$$\begin{aligned}2\gamma \sum_{\ell=0}^{N-1} (\prod_{j=N-\ell}^{N-1} \lambda_j^i)^2 = 2\int_0^T \exp[2\int_{T-t}^T a^i(s)\mathrm{d}s]\mathrm{d}t + \gamma(1-\exp[2\int_0^T a^i(t)\mathrm{d}t]) \\ -2\gamma \int_0^T \{\exp[2\int_{T-t}^T a^i(s)\mathrm{d}s][\int_{T-t}^T a^i(s)^2\mathrm{d}t + a^i(T) - a^i(T-t)]\}\mathrm{d}t \\ +\gamma^3 E_3^\gamma \ . \end{aligned} \tag{S21}$$

Hence, combining (S20) and (S21) we get that

$$h_N^i = c_T^i - \gamma e_T^i + \gamma^3 E^\gamma \ ,$$

with

$$c_T^i = \exp[2\int_0^T a^i(t)\mathrm{d}t] + 2\int_0^T \exp[2\int_{T-t}^T a^i(s)\mathrm{d}s]\mathrm{d}t \ . \tag{S22}$$

and

$$\begin{aligned}e_T^i = -\exp[2\int_0^T a^i(t)\mathrm{d}t][\int_0^T a^i(t)^2\mathrm{d}t + a^i(T) - a^i(0)] + 1 - \exp[2\int_0^T a^i(t)\mathrm{d}t] \\ -2\int_0^T \exp[2\int_{T-t}^T a^i(s)\mathrm{d}s][\int_{T-t}^T a^i(s)^2\mathrm{d}s + a^i(T) - a^i(T-t)]\mathrm{d}t \ .\end{aligned}$$

In what follows, we compute c_T^i and e_T^i.

(i) Using Lemma S11 we have

$$\exp[2\int_0^T a^i(t)\mathrm{d}t] = d^2\exp[-2T]/(1+\exp[-2T](d-1))^2 \ .$$

In addition, using Lemma S12 we have

$$\int_0^T \exp[2\int_{T-t}^T a^i(s)\mathrm{d}s] = (d/2)(1-\exp[-2T])/(1+\exp[-2T](d-1)) \ .$$

Combining these results and (S22), we get that

$$\begin{aligned}c_T^i &= d + d^2\exp[-2T](1-(1+\exp[-2T](d-1))^{-1})/(1+\exp[-2T](d-1)) \\ &= d + d^2(d-1)\exp[-4T]/(1+\exp[-2T](d-1))^2 \ .\end{aligned}$$

(ii) We conclude for e_T^i using Proposition S20 with $\lambda = d^i - 1$.

This concludes the proof of (S13). Next, we compute the evolution of the mean. Using (S23), we have

$$\mathbb{E}[Y_{k+1}] = ((1+\gamma)\,\mathrm{Id} - 2\gamma\Sigma_{T-k\gamma}^{-1})\mathbb{E}[Y_k] + 2\gamma\mathbb{E}[\Sigma_{T-k\gamma}^{-1} M_{T-k\gamma}] \ , \tag{S23}$$

Note that for any $k \in \{0, \dots, N-1\}$, we have that $\Lambda_k = \mathrm{P}((1+\gamma)\,\mathrm{Id} - 2\gamma\Sigma_{T-k\gamma}^{-1})\mathrm{P}^\top$ is diagonal. For any $k \in \{0, \dots, N\}$, define $\mathrm{H}_k = \mathrm{P}\mathbb{E}[Y_k]\mathrm{P}^\top$. Note that $\mathrm{H}_0 = 0$. For any $k \in \{0, \dots, N-1\}$ we have that

$$\mathrm{H}_{k+1} = \Lambda_k \mathrm{H}_k + 2\gamma \mathrm{D}_{T-k\gamma}^{-1}\mathrm{V}_{T-k\gamma} \ , \tag{S24}$$

where for any $t \in [0,T]$, $\mathrm{D}_t = \mathrm{P}\Sigma_t\mathrm{P}^\top$ and $\mathrm{V}_t = \mathrm{P}M_t$. Let $i \in \{1, \dots, d\}$. Using (S24), we have for any $k \in \{0, \dots, N-1\}$

$$h_{k+1}^i = \lambda_k^i h_k^i + 2\gamma v_{T-k\gamma}^i / d_{T-k\gamma}^i \ . \tag{S25}$$

In what follows, we define for any $t \in [0,T]$, $r(t)^i = v_{T-t}^i/d_{T-t}^i$ and note that for any $t \in [0,T]$

$$r(t)^i = \exp[-(T-t)]/(1+\exp[-2(T-t)](d^i-1))(\mathrm{P}\mu)^i \ . \tag{S26}$$

Using (S25) and that $h_0^i = 0$, we have that for any $k \in \{0, \dots, N\}$

$$h_k^i = 2\gamma \sum_{\ell=0}^{k-1} r((k-\ell-1)\gamma) \prod_{j=0}^{\ell-1} \lambda_{k-1-j}^i = 2\gamma \sum_{\ell=0}^{k-1} r((k-\ell-1)\gamma) \prod_{j=k-\ell}^{k-1} \lambda_j^i \ .$$

Using (S19), we get that there exists $D^\gamma \geq 0$ such that $\gamma D^\gamma \leq D$ not dependent on γ and

$$\begin{aligned}
h_N^i &= 2\gamma \textstyle\sum_{k=0}^{N-1} r(T-(k+1)\gamma)\exp[\int_{T-k\gamma}^T a^i(t)\mathrm{d}t] \\
&\quad -\gamma^2 \textstyle\sum_{k=0}^{N-1} r(T-(k+1)\gamma)\exp[\int_{T-k\gamma}^T a^i(t)\mathrm{d}t][\int_{T-k\gamma}^T a^i(t)^2\mathrm{d}t + a^i(T) - a^i(T-k\gamma)] \\
&\quad + \textstyle\sum_{k=0}^{N-1} \gamma^4 D_{k,N}^\gamma \\
&= 2\gamma \textstyle\sum_{k=0}^{N-1} r(T-(k+1)\gamma)\exp[\int_{T-k\gamma}^T a^i(t)\mathrm{d}t] \\
&\quad -\gamma^2 \textstyle\sum_{k=0}^{N-1} r(T-(k+1)\gamma)\exp[\int_{T-k\gamma}^T a^i(t)\mathrm{d}t][\int_{T-k\gamma}^T a^i(t)^2\mathrm{d}t + a^i(T) - a^i(T-k\gamma)] \\
&\quad +\gamma^3 D^\gamma \ .
\end{aligned}$$

Using Proposition S10, we get that there exists $E^\gamma \geq 0$ such that $\gamma E^\gamma \leq E$ not dependent on γ and

$$\begin{aligned}
h_N^i &= 2\gamma \textstyle\sum_{k=0}^{N-1} r(T-(k+1)\gamma)\exp[\int_{T-k\gamma}^T a^i(t)\mathrm{d}t] \\
&\quad -\gamma \textstyle\int_0^T r(T-t)\exp[\int_{T-t}^T a^i(s)\mathrm{d}s][\int_{T-t}^T a^i(t)^2\mathrm{d}t + a^i(T) - a^i(T-t)]\mathrm{d}t \\
&\quad +\gamma^3 E^\gamma \ .
\end{aligned}$$

In addition, for any $k \in \{0, \dots, N\}$, there exists $u_k \geq 0$ with $u_k \leq u$ and $u \geq 0$ not dependent on k and

$$r(T-(k+1)\gamma) = r(T-k\gamma) - r'(T-k\gamma)\gamma + u_k\gamma^2 \ .$$

Using this result, we get that exists $F^\gamma \geq 0$ such that $\gamma F^\gamma \leq F$ not dependent on γ and

$$\begin{aligned}
h_N^i &= 2\gamma \textstyle\sum_{k=0}^{N-1} r(T-k\gamma)\exp[\int_{T-k\gamma}^T a^i(t)\mathrm{d}t] \\
&\quad -2\gamma^2 \textstyle\sum_{k=0}^{N-1} r'(T-k\gamma)\exp[\int_{T-k\gamma}^T a^i(t)\mathrm{d}t] \\
&\quad -\gamma \textstyle\int_0^T r(T-t)\exp[\int_{T-t}^T a^i(s)\mathrm{d}s][\int_{T-t}^T a^i(t)^2\mathrm{d}t + a^i(T) - a^i(T-t)]\mathrm{d}t \\
&\quad +\gamma^3 F^\gamma \ .
\end{aligned}$$

Using Proposition S10, we get that there exists $G^\gamma \geq 0$ such that $\gamma G^\gamma \leq G$ not dependent on γ and

$$\begin{aligned}
h_N^i &= 2\gamma \textstyle\sum_{k=0}^{N-1} r(T-k\gamma)\exp[\int_{T-k\gamma}^T a^i(t)\mathrm{d}t] \\
&\quad -2\gamma \textstyle\int_0^T r'(T-t)\exp[\int_{T-t}^T a^i(t)\mathrm{d}s]\mathrm{d}t \\
&\quad -\gamma \textstyle\int_0^T r(T-t)\exp[\int_{T-t}^T a^i(s)\mathrm{d}s][\int_{T-t}^T a^i(t)^2\mathrm{d}t + a^i(T) - a^i(T-t)]\mathrm{d}t \\
&\quad +\gamma^3 G^\gamma \ .
\end{aligned}$$

In addition, using Proposition S10, we get that there exists $H^\gamma \geq 0$ such that $\gamma H^\gamma \leq H$ not dependent on γ and

$$\begin{aligned}
h_N^i &= 2 \textstyle\int_0^T r(T-t)\exp[\int_{T-t}^T a^i(s)\mathrm{d}s]\mathrm{d}t \\
&\quad -\gamma\{r(0)\exp[\int_0^T a^i(t)\mathrm{d}t] - r(T)\} - 2\gamma \textstyle\int_0^T r'(T-t)\exp[\int_{T-t}^T a^i(s)\mathrm{d}s]\mathrm{d}t \\
&\quad -\gamma \textstyle\int_0^T r(T-t)\exp[\int_{T-t}^T a^i(s)\mathrm{d}s][\int_{T-t}^T a^i(t)^2\mathrm{d}t + a^i(T) - a^i(T-t)]\mathrm{d}t \\
&\quad +\gamma^3 H^\gamma \ .
\end{aligned} \tag{S27}$$

In addition, we have by integration by part

$$\begin{aligned}
&\textstyle\int_0^T r'(T-t)\exp[\int_{T-t}^T a^i(s)\mathrm{d}s]\mathrm{d}t \\
&\qquad = -\{r(0)\exp[\int_0^T a^i(t)\mathrm{d}t] - r(T)\} - \textstyle\int_0^T r(T-t)a^i(T-t)\exp[\int_{T-t}^T a^i(s)\mathrm{d}s]\mathrm{d}t \ .
\end{aligned}$$

Combining this result and (S27) we get that

$$\begin{aligned}
h_N^i &= 2 \textstyle\int_0^T r(T-t)\exp[\int_{T-t}^T a^i(s)\mathrm{d}s]\mathrm{d}t \\
&\quad +\gamma\{r(0)\exp[\int_0^T a^i(t)\mathrm{d}t] - r(T)\} \\
&\quad -\gamma \textstyle\int_0^T r(T-t)\exp[\int_{T-t}^T a^i(s)\mathrm{d}s][\int_{T-t}^T a^i(t)^2\mathrm{d}t + a^i(T) - 3a^i(T-t)]\mathrm{d}t \\
&\quad +\gamma^3 H^\gamma \ .
\end{aligned} \tag{S28}$$

In what follows, we assume that $d^i \neq 0$. The case where $d^i = 0$ is left to the reader. Finally using (S26) and Lemma S11 we have that for any $t \in [0, T]$

$$\begin{aligned}\exp[\textstyle\int_{T-t}^{T} a^i(s)\mathrm{d}s] r(T-t)^i &= \exp[-2t]/(1+\exp[-2t](d^i-1))^2 (\mathrm{P}\mu)^i d^i \\ &= \exp[2\textstyle\int_{T-t}^{T} a^i(s)\mathrm{d}s](\mathrm{P}\mu)^i / d^i \ .\end{aligned}$$

Therefore, combining this result and (S28), we get that

$$\begin{aligned}h_N^i = (\mathrm{P}\mu)^i/d^i[&2\textstyle\int_0^T \exp[2\int_{T-t}^{T} a^i(s)\mathrm{d}s]\mathrm{d}t \\ &+\gamma\{\exp[2\textstyle\int_0^T a^i(t)\mathrm{d}t]-1\} \\ &-\gamma\textstyle\int_0^T \exp[2\int_{T-t}^{T} a^i(s)\mathrm{d}s][\int_{T-t}^{T} a^i(t)^2\mathrm{d}t + a^i(T) - 3a^i(T-t)]\mathrm{d}t] \\ &+\gamma^3 H^\gamma \ ,\end{aligned}$$

which concludes the proof upon using Lemma S12 and Proposition S21. ☐

S5.3 Technical lemmas

We are going to make use of the following lemma which is a direct consequence of the Euler-MacLaurin formula.

Proposition S10. *Let $f \in \mathrm{C}^\infty([0,T])$, and $(u_k^\gamma)_{k\in\{0,\dots,N-1\}}$ with $N \in \mathbb{N}$ and $\gamma = T/N > 0$ such that for any $k \in \{0,\dots,N-1\}$, $u_k^\gamma = f(k\gamma)$. Then, there exists $C \geq 0$ such that*

$$\textstyle\int_0^T f(t)\mathrm{d}t - \gamma\sum_{k=0}^{N-1} u_k^\gamma - (\gamma/2)\{f(T)-f(0)\} = C\gamma^2 \ .$$

Proof. Apply the classical Euler-MacLaurin formula to $t \mapsto f(t\gamma)$. ☐

We will also use the following lemmas.

Lemma S11. *Let $\lambda \in (-1,+\infty)$ and $a : [0,T] \to \mathbb{R}$ such that for any $t \in [0,T]$,*

$$a(t) = 1 - 2/(1+\exp[-2(T-t)]\lambda) \ .$$

Then, we have that for any $t \in [0,T]$,

$$\textstyle\int_{T-t}^{T} a(s)\mathrm{d}s = t + \log((1+\lambda)/(\exp[2t]+\lambda)) \ .$$

In particular, we have that for any $t \in [0,T]$

$$\exp[2\textstyle\int_{T-t}^{T} a(s)\mathrm{d}s] = \exp[-2t](1+\lambda)^2/(1+\lambda\exp[-2t])^2 \ .$$

Proof. Let $t \in [0,T]$. We have that $\int_{T-t}^{T} a(s)\mathrm{d}s = \int_0^t a(T-s)\mathrm{d}s$. Define b such that for any $t \in [0,T]$, $b(t) = a(T-t)$. In particular, we have that for any $t \in [0,T]$

$$b(t) = 1 - 2/(1+\lambda\exp[-2t]) \ .$$

Hence, we have

$$\begin{aligned}\textstyle\int_0^t b(s)\mathrm{d}s &= t - 2\textstyle\int_0^t (1+\lambda\exp[-2s])^{-1}\mathrm{d}s \\ &= t - \textstyle\int_0^t 2\exp[2s]/(\exp[2s]+\lambda)\mathrm{d}s \\ &= t + \log((1+\lambda)/(\exp[2t]+\lambda)) \ ,\end{aligned}$$

which concludes the proof. ☐

Lemma S12. *Let $\lambda \in (-1,+\infty)$ and $a : [0,T] \to \mathbb{R}$ such that for any $t \in [0,T]$,*

$$a(t) = 1 - 2/(1+\exp[-2(T-t)]\lambda) \ .$$

Then, we have that for any $t \in [0,T]$,

$$\begin{aligned}\textstyle\int_0^t \exp[2\int_{T-s}^{T} a(u)\mathrm{d}u]\mathrm{d}s &= (1/2)(1+\lambda)^2[(1+\lambda\exp[-2t])^{-1} - 1/(1+\lambda)]/\lambda \\ &= (1/2)(1+\lambda)(1-\exp[-2t])/(1+\lambda\exp[-2t]) \ .\end{aligned}$$

Proof. Let $t \in [0,T]$. Using Lemma S11 we have that for any $s \in [0,T]$

$$\exp[2\int_{T-s}^{T} a(u)\mathrm{d}u] = (1+\lambda)^2 \exp[2s]/(\lambda+\exp[2s])^2 = (1+\lambda)^2\exp[-2s]/(1+\lambda\exp[-2s])^2 \ .$$

Assume that $\lambda \neq 0$. Then, we have that

$$\begin{aligned}\int_0^t \exp[2\int_{T-s}^{T} a(u)\mathrm{d}u]\mathrm{d}s &= (1/2)(1+\lambda)^2/\lambda \int_0^t 2\lambda\exp[-2t]/(1+\lambda\exp[-2t])^2\mathrm{d}s \\ &= (1/2)(1+\lambda)^2[(1+\lambda\exp[-2t])^{-1} - 1/(1+\lambda)]/\lambda \\ &= (1/2)(1+\lambda)(1-\exp[-2t])/(1+\lambda\exp[-2t]) \ .\end{aligned}$$

We conclude the proof upon remarking that his result still holds in the case where $\lambda = 0$. □

Lemma S13. *Let $\lambda \in (-1,+\infty)$ and $a:\ [0,T] \to \mathbb{R}$ such that for any $t \in [0,T]$,*

$$a(t) = 1 - 2/(1+\exp[-2(T-t)]\lambda) \ .$$

Then, if $\lambda \neq 0$, we have that for any $t \in [0,T]$,

$$\begin{aligned}&\int_0^t \exp[2\int_{T-s}^{T} a(u)\mathrm{d}u]/(1+\lambda\exp[-2s])\mathrm{d}s \\ &\qquad = (1/4)(1+\lambda)^2[(1+\lambda\exp[-2t])^{-2} - 1/(1+\lambda)^2]/\lambda \\ &\qquad = (1/4)(1-\exp[-2t])(2+\lambda(1+\exp[-2t]))/(1+\lambda\exp[-2t])^2 \ .\end{aligned}$$

If $\lambda = 0$ we have

$$\int_0^t \exp[2\int_{T-s}^{T} a(u)\mathrm{d}u]/(1+\lambda\exp[-2s])\mathrm{d}s = (1/2)(1-\exp[-2t]) \ .$$

Proof. Let $t \in [0,T]$. Using Lemma S11 we have that for any $s \in [0,T]$

$$\exp[\int_{T-s}^{T} a(u)\mathrm{d}u]/(1+\lambda\exp[-2s]) = (1+\lambda)^2\exp[-2s]/(1+\lambda\exp[-2s])^3 \ .$$

Assume that $\lambda \neq 0$. Then, we have that

$$\begin{aligned}\int_0^t \exp[\int_{T-s}^{T} a(u)\mathrm{d}u]/(1+\lambda\exp[-2s])\mathrm{d}s &= (1/2)(1+\lambda)^2/\lambda \int_0^t 2\lambda\exp[-2t]/(1+\lambda\exp[-2t])^3\mathrm{d}s \\ &= (1/4)(1+\lambda)^2[(1+\lambda\exp[-2t])^{-2} - 1/(1+\lambda)^2]/\lambda \\ &= (1/4)(1-\exp[-2t])(2+\lambda(1+\exp[-2t]))/(1+\lambda\exp[-2t])^2 \ .\end{aligned}$$

We conclude the proof upon remarking that his result still holds in the case where $\lambda = 0$. □

Lemma S14. *Let $\lambda \in (-1,+\infty)$ and $a:\ [0,T] \to \mathbb{R}$ such that for any $t \in [0,T]$,*

$$a(t) = 1 - 2/(1+\exp[-2(T-t)]\lambda) \ .$$

Then, we have that for any $t \in [0,T]$

$$\begin{aligned}&\int_0^t \exp[2\int_{T-s}^{T} a(u)\mathrm{d}u]a(T-s)\mathrm{d}s \\ &\qquad = -(1/2)(1-\exp[-2t])(1-\lambda^2\exp[-2t])/(1+\lambda\exp[-2t])^2 \ .\end{aligned}$$

Proof. Let $t \in [0,T]$. We have that

$$\begin{aligned}&\int_0^t \exp[2\int_{T-s}^{T} a(u)\mathrm{d}u]a(T-s)\mathrm{d}s \\ &\qquad = \int_0^t \exp[2\int_{T-s}^{T} a(u)\mathrm{d}u]\mathrm{d}s - 2\int_0^t \exp[2\int_{T-s}^{T} a(u)\mathrm{d}u]/(1+\lambda\exp[-2s])\mathrm{d}s \ .\end{aligned} \tag{S29}$$

Using Lemma S12, we have that

$$\int_0^t \exp[2\int_{T-s}^{T} a(u)\mathrm{d}u]\mathrm{d}t = (1/2)(1+\lambda)(1-\exp[-2t])/(1+\lambda\exp[-2t]) \ . \tag{S30}$$

In addition, using Lemma S13, we have

$$\begin{aligned}&\int_0^t \exp[2\int_{T-s}^{T} a(u)\mathrm{d}u]/(1+\lambda\exp[-2s])\mathrm{d}s \\ &\qquad = (1/4)(1-\exp[-2t])(2+\lambda(1+\exp[-2t]))/(1+\lambda\exp[-2t])^2 \ .\end{aligned} \tag{S31}$$

Combining (S30) and (S31) in (S29) we have that

$$\begin{aligned}
&\textstyle\int_0^t \exp[2\int_{T-s}^T a(u)\mathrm{d}u]a(T-s)\mathrm{d}s \\
&\quad = (1/2)(1-\exp[-2t])[(1+\lambda)(1+\lambda\exp[-2t])]/(1+\lambda\exp[-2t])^2 \\
&\qquad - (1/2)(1-\exp[-2t])(2+\lambda(1+\exp[-2t]))/(1+\lambda\exp[-2t])^2 \\
&\quad = -(1/2)(1-\exp[-2t])(1-\lambda^2\exp[-2t])/(1+\lambda\exp[-2t])^2 \ ,
\end{aligned}$$

which concludes the proof. □

Lemma S15. *Let $\lambda \in (-1,+\infty)$ we have that for any $t \in [0,T]$*

$$\textstyle\int_0^t (1+\lambda\exp[-2s])^{-1}\mathrm{d}s = (1/2)\log((\lambda+\exp[2t])/(\lambda+1)) \ .$$

In addition, we have for any $t \in [0,T]$

$$\textstyle\int_0^t (1+\lambda\exp[-2s])^{-2}\mathrm{d}s = (1/2)\log((\lambda+\exp[2t])/(\lambda+1)) + (\lambda/2)[(\exp[2t]+\lambda)^{-1}-(\lambda+1)^{-1}] \ .$$

Finally, we have that for any $t \in [0,T]$

$$\begin{aligned}
\textstyle\int_0^t (1+\lambda\exp[-2s])^{-3}\mathrm{d}s = (1/2)\log((\lambda+\exp[2t])/(\lambda+1)) + \lambda[(\exp[2t]+\lambda)^{-1}-(\lambda+1)^{-1}] \\
- (\lambda^2/4)[(\exp[2t]+\lambda)^{-2}-(\lambda+1)^{-2}] \ .
\end{aligned}$$

Proof. Let $k \in \{1,2,3\}$. Using the change of variable $u \mapsto \exp[2u]$ we have that

$$\textstyle\int_0^t (1+\lambda\exp[-2s])^{-k}\mathrm{d}s = (1/2)\int_1^{\exp[2t]} u^{k-1}/(u+\lambda)\mathrm{d}u \ .$$

Therefore, we have that

$$\textstyle\int_0^t (1+\lambda\exp[-2s])^{-1}\mathrm{d}s = (1/2)\int_1^{\exp[2t]}(u+\lambda)^{-1}\mathrm{d}u = (1/2)\log((\lambda+\exp[2t])/(\lambda+1)) \ .$$

In addition, using that for any $u \in [0,T]$, $u = (u+\lambda)-\lambda$ we have that

$$\begin{aligned}
&\textstyle\int_0^t (1+\lambda\exp[-2s])^{-2}\mathrm{d}s = (1/2)\int_1^{\exp[2t]} u(u+\lambda)^{-2}\mathrm{d}u \\
&\quad = \textstyle(1/2)\int_1^{\exp[2t]}(u+\lambda)^{-1}\mathrm{d}u - (\lambda/2)\int_1^{\exp[2t]}(u+\lambda)^{-2}\mathrm{d}u \\
&\quad = (1/2)\log((\lambda+\exp[2t])/(\lambda+1)) + (\lambda/2)[(\exp[2t]+\lambda)^{-1}-(\lambda+1)^{-1}] \ .
\end{aligned}$$

Finally, using that for any $u \in [0,T]$, $u \in [0,T]$, $u^2 = (u+\lambda)^2 - 2\lambda(u+\lambda)+\lambda^2$ we have that

$$\begin{aligned}
&\textstyle\int_0^t (1+\lambda\exp[-2s])^{-3}\mathrm{d}s = (1/2)\int_1^{\exp[2t]} u^2(u+\lambda)^{-2}\mathrm{d}u \\
&\quad = \textstyle(1/2)\int_1^{\exp[2t]}(u+\lambda)^{-1}\mathrm{d}u - \lambda\int_1^{\exp[2t]}(u+\lambda)^{-2}\mathrm{d}u + (\lambda^2/2)\int_1^{\exp[2t]}(u+\lambda)^{-3}\mathrm{d}u \\
&\quad = (1/2)\log((\lambda+\exp[2t])/(\lambda+1)) + \lambda[(\exp[2t]+\lambda)^{-1}-(\lambda+1)^{-1}] \\
&\qquad - (\lambda^2/4)[(\exp[2t]+\lambda)^{-2}-(\lambda+1)^{-2}] \ ,
\end{aligned}$$

which concludes the proof. □

Lemma S16. *Let $\lambda \in (-1,+\infty)$ and $a: [0,T] \to \mathbb{R}$ such that for any $t \in [0,T]$,*

$$a(t) = 1 - 2/(1+\exp[-2(T-t)]\lambda) \ .$$

Then, we have that for any $t \in [0,T]$,

$$\textstyle\int_{T-t}^T a(s)^2\mathrm{d}s = t - 2\lambda(1-\exp[-2t])/[(1+\lambda)(1+\lambda\exp[-2t])] \ .$$

Proof. Let $t \in [0,T]$. Similarly to the proof of Lemma S11, we have that $\int_{T-t}^T a(s)\mathrm{d}s = \int_0^t a(T-s)\mathrm{d}s$. Define b such that for any $t \in [0,T]$, $b(t) = a(T-t)$. In particular, we have that for any $t \in [0,T]$

$$b(t) = 1 - 2/(1+\lambda\exp[-2t]) \ .$$

We have that

$$\int_{T-t}^{T} a(s)^2 \mathrm{d}s = \int_0^t b(s)^2 \mathrm{d}s = \int_0^t (1 - 4/(1+\lambda\exp[-2s]) + 4/(1+\lambda\exp[-2s])^2) \mathrm{d}s \ .$$

Combining this result and Lemma S15, we have

$$\begin{aligned}\int_{T-t}^{T} a(s)^2 \mathrm{d}s &= t + 2\lambda[(\lambda + \exp[2t])^{-1} - (\lambda+1)^{-1}] \\ &= t - 2\lambda(1-\exp[-2t])/[(1+\lambda)(1+\lambda\exp[-2t])] \ .\end{aligned}$$

□

Lemma S17. *Let $\lambda \in (-1, +\infty)$ and $a : \ [0,T] \to \mathbb{R}$ such that for any $t \in [0,T]$,*

$$a(t) = 1 - 2/(1 + \exp[-2(T-t)]\lambda) \ .$$

Then, if $\lambda \neq 0$, we have that

$$\begin{aligned}\int_0^T \exp[2\textstyle\int_{T-t}^T a(s)\mathrm{d}s](\int_{T-t}^T a(s)^2\mathrm{d}s)\mathrm{d}t &= -(T/2)(1+\lambda)^2\exp[-2T]/(1+\lambda\exp[-2T]) \\ &\quad + (1+\lambda)^2/(4\lambda)\log((1+\lambda)/(1+\lambda\exp[-2T])) \\ &\quad - (\lambda/2)(1-\exp[-2T])^2/(1+\lambda\exp[-2T])^2 \ . \end{aligned} \tag{S32}$$

If $\lambda = 0$, we have that

$$\int_0^T \exp[2\textstyle\int_{T-t}^T a(s)\mathrm{d}s](\int_{T-t}^T a(s)^2\mathrm{d}s)\mathrm{d}t = -(T/2)\exp[-2T] + (1/4)(1-\exp[-2T]) \ . \tag{S33}$$

Note that taking $\lambda \to 0$ in (S32) we recover (S33), using that for any $u > 0$, $\lim_{\lambda\to 0}\log(1+\lambda u)/\lambda = u$.

Proof. We first start with the case $\lambda \neq 0$. Similarly to the proof of Lemma S11, we have that $\int_{T-t}^T a(s)\mathrm{d}s = \int_0^t a(T-s)\mathrm{d}s$. Define b such that for any $t \in [0,T]$, $b(t) = a(T-t)$. We have that

$$\int_0^T \exp[2\textstyle\int_{T-t}^T a(s)\mathrm{d}s](\int_{T-t}^T a(s)^2\mathrm{d}s)\mathrm{d}t = \int_0^T \exp[2\int_{T-t}^T a(s)\mathrm{d}s](\int_0^t b(s)^2\mathrm{d}s)\mathrm{d}t \ .$$

Let $A : \ [0,T] \to \mathbb{R}$ such that for any $t \in [0,T]$,

$$A(t) = \int_0^t \exp[2\textstyle\int_{T-s}^T a(u)\mathrm{d}u]\mathrm{d}s \ .$$

Note that $A(0) = 0$. Hence, by integration by parts, we have

$$\int_0^T \exp[2\textstyle\int_{T-t}^T a(s)\mathrm{d}s](\int_0^t b(s)^2\mathrm{d}s)\mathrm{d}t = A(T)\int_0^T b(t)^2\mathrm{d}t - \int_0^T A(t)b(t)^2\mathrm{d}t \ .$$

In what follows, we compute $\int_0^T A(t)b(t)^2\mathrm{d}t$. First, we recall that for any $t \in [0,T]$

$$b(t)^2 = (1 - 2/(1+\lambda\exp[-2t]))^2 = 1 - 4/(1+\lambda\exp[-2t]) + 4/(1+\lambda\exp[-2t])^2 \ . \tag{S34}$$

In addition, using Lemma S12, we have that for any $t \in [0,T]$

$$A(t) = (1/2)\{(1+\lambda)^2/(\lambda(1+\lambda\exp[-2t])) - (1+\lambda)/\lambda\} \ . \tag{S35}$$

Using (S34) and (S35) we have that for any $t \in [0,T]$

$$\begin{aligned}2A(t)b(t)^2 &= -(1+\lambda)/\lambda + [4(1+\lambda)/\lambda + (1+\lambda)^2/\lambda]u_1(t) \\ &\quad - [4(1+\lambda)^2/\lambda + 4(1+\lambda)/\lambda]u_2(t) + [4(1+\lambda)^2/\lambda]u_3(t) \\ &= -(1+\lambda)/\lambda + [(1+\lambda)(5+\lambda)/\lambda]u_1(t) \\ &\quad - [4(1+\lambda)(2+\lambda)/\lambda]u_2(t) + [4(1+\lambda)^2/\lambda]u_3(t) \ ,\end{aligned} \tag{S36}$$

where for any $k \in \{1,2,3\}$ and $t \in [0,T]$ we have

$$u_k(t) = (1+\lambda\exp[-2t])^{-k} \ .$$

For any $k \in \{0,1,2\}$ denote $v_k : \ [0,T] \to \mathbb{R}$ such that for any $t \in [0,T]$ and $k \in \{1,2\}$

$$v_0(t) = \log((\lambda+\exp[2t])/(\lambda+1)) \ , \qquad v_k(t) = (\exp[2t]+\lambda)^{-k} - (1+\lambda)^{-k} \ .$$

Combining (S36) and Lemma S15, we get that for any $t \in [0, T]$

$$
\begin{aligned}
2 \int_0^t A(s) b(s)^2 \mathrm{d}s &= -[(1+\lambda)/\lambda] t \\
&\quad + (1/2)\{[(1+\lambda)(5+\lambda)/\lambda] - [4(1+\lambda)(2+\lambda)/\lambda] + [4(1+\lambda)^2/\lambda]\} v_0(t) \\
&\quad + \{-(\lambda/2)[4(1+\lambda)(2+\lambda)/\lambda] + \lambda[4(1+\lambda)^2/\lambda]\} v_1(t) \\
&\quad - (\lambda^2/4)[4(1+\lambda)^2/\lambda] v_2(t) \\
&= -[(1+\lambda)/\lambda] t + (1+\lambda)^2/(2\lambda) v_0(t) + 2(1+\lambda)\lambda v_1(t) - \lambda(1+\lambda)^2 v_2(t) \ .
\end{aligned}
$$

In addition, we have that

$$
\begin{aligned}
-[(1+\lambda)/\lambda] t + (1+\lambda)^2/(2\lambda) v_0(t) &= -[(1+\lambda)/\lambda] t + [(1+\lambda)^2/\lambda] t \\
&\quad + (1+\lambda)^2/(2\lambda) \log((1+\lambda\exp[-2t])/(1+\lambda)) \\
&= (1+\lambda) t + (1+\lambda)^2/(2\lambda) \log((1+\lambda\exp[-2t])/(1+\lambda)) \ .
\end{aligned}
$$

Therefore, we get that

$$
\begin{aligned}
2 \int_0^t A(s) b(s)^2 \mathrm{d}s &= (1+\lambda) t + (1+\lambda)^2/(2\lambda) \log((1+\lambda\exp[-2t])/(1+\lambda)) \\
&\quad + 2(1+\lambda)\lambda v_1(t) + \lambda(1+\lambda)^2 v_2(t) \ .
\end{aligned} \tag{S37}
$$

In addition, we have that

$$
(1+\lambda)\lambda v_1(t) = -\lambda(1-\exp[-2T])/(1+\lambda\exp[-2T]) \ . \tag{S38}
$$

We also have that

$$
\begin{aligned}
\lambda(1+\lambda)^2 v_2(t) &= \lambda(2\lambda + 1 - 2\lambda\exp[2T] - \exp[4T])/(\exp[2T]+\lambda)^2 \\
&= \lambda(1-\exp[2T])(1+2\lambda+\exp[2T])/(\exp[2T]+\lambda)^2 \\
&= -\lambda(1-\exp[-2T])(1+(1+2\lambda)\exp[-2T])/(1+\lambda\exp[-2T])^2 \ .
\end{aligned} \tag{S39}
$$

Finally, using Lemma S12 and Lemma S16 we have

$$
\begin{aligned}
A(T) \int_0^T b(t)^2 \mathrm{d}t &= (1/2)(1+\lambda)(1-\exp[-2T])/(1+\lambda\exp[-2T]) \\
&\quad \times (T - 2\lambda(1-\exp[-2T])/[(1+\lambda)(1+\lambda\exp[-2T])]) \\
&= (T/2)(1+\lambda)(1-\exp[-2T])/(1+\lambda\exp[-2T]) \\
&\quad - \lambda(1-\exp[-2T])^2/(1+\lambda\exp[-2T])^2 \ .
\end{aligned} \tag{S40}
$$

Combining (S37), (S38), (S39) and (S40) we get

$$
\begin{aligned}
\int_0^T \exp[2\int_{T-t}^T a(s)\mathrm{d}s](\int_{T-t}^T a(s)^2 \mathrm{d}s)\mathrm{d}t &= (T/2)(1+\lambda)(1-\exp[-2T])/(1+\lambda\exp[-2T]) \\
&\quad - \lambda(1-\exp[-2T])^2/(1+\lambda\exp[-2T])^2 \\
&\quad - (1+\lambda)(T/2) + (1+\lambda)^2/(4\lambda)\log((1+\lambda)/(1+\lambda\exp[-2T])) \\
&\quad + \lambda(1-\exp[-2T])/(1+\lambda\exp[-2T]) \\
&\quad - (\lambda/2)(1-\exp[-2T])((1+2\lambda)\exp[-2T]+1)/(1+\lambda\exp[-2T])^2 \ .
\end{aligned} \tag{S41}
$$

In addition, we have that

$$
\begin{aligned}
&-(\lambda/2)(1-\exp[-2T])^2/(1+\lambda\exp[-2T])^2 \\
&\qquad = -\lambda(1-\exp[-2T])^2/(1+\lambda\exp[-2T])^2 \\
&\qquad\quad + \lambda(1-\exp[-2T])/(1+\lambda\exp[-2T]) \\
&\qquad\quad - (\lambda/2)(1-\exp[-2T])((1+2\lambda)\exp[-2T]+1)/(1+\lambda\exp[-2T])^2 \ .
\end{aligned}
$$

Combining this result and (S41), we get

$$
\begin{aligned}
\int_0^T \exp[2\int_{T-t}^T a(s)\mathrm{d}s](\int_{T-t}^T a(s)^2 \mathrm{d}s)\mathrm{d}t &= (T/2)(1+\lambda)(1-\exp[-2T])/(1+\lambda\exp[-2T]) \\
&\quad - (1+\lambda)(T/2) + (1+\lambda)^2/(4\lambda)\log((1+\lambda)/(1+\lambda\exp[-2T])) \\
&\quad - (1/2)(1-\exp[-2T])^2/(1+\lambda\exp[-2T])^2 \ .
\end{aligned} \tag{S42}
$$

Finally, we have

$$
\begin{aligned}
&(T/2)(1+\lambda)(1-\exp[-2T])/(1+\lambda\exp[-2T]) - (T/2)(1+\lambda) \\
&\quad = -(T/2)(1+\lambda)^2\exp[-2T]/(1+\lambda\exp[-2T]) \ ,
\end{aligned}
$$

which concludes the proof in the case $\lambda \neq 0$ upon combining this result and (S42). In the case $\lambda = 0$, we have that for any $t \in [0,T]$, $a(t) = -1$ and therefore by integration by part we have

$$
\int_0^T \exp[2\int_{T-t}^T a(s)\mathrm{d}s](\int_{T-t}^T a(s)^2\mathrm{d}s)\mathrm{d}t = -(T/2)\exp[-2T] + (1/4)(1-\exp[-2T]) \ ,
$$

which concludes the proof. □

We are now ready to prove the following results.

Proposition S18. *Let $\lambda \in (-1,+\infty)$ and $a : [0,T] \to \mathbb{R}$ such that for any $t \in [0,T]$,*

$$
a(t) = 1 - 2/(1+\exp[-2(T-t)]\lambda) \ .
$$

Then, we have that for any $t \in [0,T]$,

$$
\begin{aligned}
&\exp[2\int_0^T a(t)\mathrm{d}t]\{\int_0^T a(t)^2\mathrm{d}t + a(T) - a(0) + 1\} \\
&\qquad = (T+1)\exp[-2T](\lambda+1)^2/(1+\lambda\exp[-2T])^2 \ .
\end{aligned}
$$

Proof. First, we have that

$$
\begin{aligned}
a(T) - a(0) &= 1 - 2/(1+\lambda) - 1 + 2/(1+\lambda\exp[-2T]) \\
&= 2\lambda(1-\exp[-2T])/[(1+\lambda)(1+\lambda\exp[-2T])] \ .
\end{aligned} \tag{S43}
$$

In addition, using Lemma S16 we have

$$
\int_0^T a(s)^2\mathrm{d}s = T - 2\lambda(1-\exp[-2T])/[(1+\lambda)(1+\lambda\exp[-2T])] \ . \tag{S44}
$$

Finally, using Lemma S11 we have that

$$
\exp[2\int_0^T a(s)\mathrm{d}s] = \exp[-2T](\lambda+1)^2/(1+\lambda\exp[-2T])^2 \ . \tag{S45}
$$

We conclude the proof upon combining (S43), (S44) and (S45). □

Finally, we have the following proposition.

Proposition S19. *Let $\lambda \in (-1,+\infty)$ and $a : [0,T] \to \mathbb{R}$ such that for any $t \in [0,T]$,*

$$
a(t) = 1 - 2/(1+\exp[-2(T-t)]\lambda) \ .
$$

Then, if $\lambda \neq 0$, we have that for any $t \in [0,T]$,

$$
\begin{aligned}
&\int_0^T \exp[2\int_{T-t}^T a(s)\mathrm{d}s][\int_{T-t}^T a(s)^2\mathrm{d}s + a(T) - a(T-t)]\mathrm{d}t \\
&\quad = -(T/2)(1+\lambda)^2\exp[-2T]/(1+\lambda\exp[-2T]) \\
&\qquad + (1+\lambda)^2/(4\lambda)\log((1+\lambda)/(1+\lambda\exp[-2T])) \\
&\qquad + (\lambda/2)\exp[-2T]/(1+\lambda\exp[-2T])^2 \ .
\end{aligned}
$$

If $\lambda = 0$, we have that

$$
\begin{aligned}
&\int_0^T \exp[2\int_{T-t}^T a(s)\mathrm{d}s][\int_{T-t}^T a(s)^2\mathrm{d}s + a(T) - a(T-t)]\mathrm{d}t \\
&\qquad = -(T/2)\exp[-2T] + (1/4)(1-\exp[-2T]) \ .
\end{aligned}
$$

Proof. We assume that $\lambda \neq 0$. The case where $\lambda = 0$ is left to the reader. First, using Lemma S17, we have that

$$
\begin{aligned}
\int_0^T \exp[2\int_{T-t}^T a(s)\mathrm{d}s](\int_{T-t}^T a(s)^2\mathrm{d}s)\mathrm{d}t &= -(T/2)(1+\lambda)^2\exp[-2T]/(1+\lambda\exp[-2T]) \\
&+ (1+\lambda)^2/(4\lambda)\log((1+\lambda)/(1+\lambda\exp[-2T])) \\
&+ (3\lambda/2)\exp[-2T](1-\exp[-2T])(1+\lambda)/(1+\lambda\exp[-2T])^2 \ .
\end{aligned} \tag{S46}
$$

Second, using Lemma S14, we have that

$$\int_0^T \exp[2\int_{T-t}^T a(u)\mathrm{d}u]a(T-t)\mathrm{d}t$$
$$= -(1/2)(1-\exp[-2T])(1-\lambda^2\exp[-2T])/(1+\lambda\exp[-2T])^2 \ . \quad \text{(S47)}$$

Third, using Lemma S12 and that $a(T) = 1 - 2/(1+\lambda)$, we have that

$$\begin{aligned} a(T)\int_0^T \exp[2\int_{T-t}^T a(s)\mathrm{d}s]\mathrm{d}t &= a(T)(1/2)(1+\lambda)(1-\exp[-2T])/(1+\lambda\exp[-2T]) \\ &= (1/2)(1+\lambda)(1-\exp[-2T])/(1+\lambda\exp[-2T]) \\ &\quad - (1-\exp[-2T])/(1+\lambda\exp[-2T]) \ . \end{aligned} \quad \text{(S48)}$$

Combining (S46), (S47) and (S48) we get

$$\begin{aligned} &\int_0^T \exp[2\int_{T-t}^T a(s)\mathrm{d}s][\int_{T-t}^T a(s)^2\mathrm{d}s + a(T) - a(T-t)]\mathrm{d}t \\ &\quad = -(T/2)(1+\lambda)^2\exp[-2T]/(1+\lambda\exp[-2T]) \\ &\quad\quad + (1+\lambda)^2/(4\lambda)\log((1+\lambda)/(1+\lambda\exp[-2T])) \\ &\quad\quad + (3\lambda/2)\exp[-2T](1-\exp[-2T])(1+\lambda)/(1+\lambda\exp[-2T])^2 \\ &\quad\quad + (1/2)(1-\exp[-2T])(1-\lambda^2\exp[-2T])/(1+\lambda\exp[-2T])^2 \\ &\quad\quad + (1/2)(1+\lambda)(1-\exp[-2T])/(1+\lambda\exp[-2T]) \\ &\quad\quad - (1-\exp[-2T])/(1+\lambda\exp[-2T]) \end{aligned}$$

In addition, we have

$$\begin{aligned} &(\lambda/2)(1-\exp[-2T])/(1+\lambda\exp[-2T])^2 \\ &\quad = (1/2)(1-\exp[-2T])(1-\lambda^2\exp[-2T])/(1+\lambda\exp[-2T])^2 \\ &\quad\quad + (1/2)(1+\lambda)(1-\exp[-2T])/(1+\lambda\exp[-2T]) \\ &\quad\quad - (1-\exp[-2T])/(1+\lambda\exp[-2T]) \ , \end{aligned}$$

Finally, we have

$$\begin{aligned} &(\lambda/2)\exp[-2T]/(1+\lambda\exp[-2T])^2 \\ &\quad = -(\lambda/2)(1-\exp[-2T])^2/(1+\lambda\exp[-2T])^2 \\ &\quad\quad + (\lambda/2)(1-\exp[-2T])/(1+\lambda\exp[-2T])^2 \ . \end{aligned}$$

which concludes the proof. □

Finally, we have the following result.

Proposition S20. *Let $\lambda \in (-1, +\infty)$ and $a : [0,T] \to \mathbb{R}$ such that for any $t \in [0,T]$,*

$$a(t) = 1 - 2/(1+\exp[-2(T-t)]\lambda) \ .$$

Then, if $\lambda \neq 0$, we have that for any $t \in [0,T]$,

$$\begin{aligned} &-\exp[2\int_0^T a(t)\mathrm{d}t]\{\int_0^T a(t)^2\mathrm{d}t + a(T) - a(0)\} + 1 - \exp[2\int_0^T a(t)\mathrm{d}t] \\ &\quad - 2\int_0^T \exp[2\int_{T-t}^T a(s)\mathrm{d}s][\int_{T-t}^T a(s)^2\mathrm{d}s + a(T) - a(T-t)]\mathrm{d}t \\ &\quad\quad = 1 - \exp[-2T](1-\lambda T\exp[-2T])(\lambda+1)^2/(1+\lambda\exp[-2T])^2 \\ &\quad\quad\quad - (1+\lambda)^2/(2\lambda)\log((1+\lambda)/(1+\lambda\exp[-2T])) \\ &\quad\quad\quad - \lambda\exp[-2T]/(1+\lambda\exp[-2T])^2 \ . \end{aligned}$$

In particular, we have that

$$\begin{aligned} &-\exp[2\int_0^T a(t)\mathrm{d}t]\{\int_0^T a(t)^2\mathrm{d}t + a(T) - a(0)\} + 1 - \exp[2\int_0^T a(t)\mathrm{d}t] \\ &\quad - 2\int_0^T \exp[2\int_{T-t}^T a(s)\mathrm{d}s][\int_{T-t}^T a(s)^2\mathrm{d}s + a(T) - a(T-t)]\mathrm{d}t \\ &\quad\quad = 1 - (1/2)(1+\lambda)^2\log(1+\lambda)/\lambda + O(\exp[-2T]) \ . \end{aligned}$$

If $\lambda = 0$, we have that for any $t \in [0, T]$

$$
\begin{aligned}
- \exp[2 \textstyle\int_0^T a(t)\mathrm{d}t] \{ \textstyle\int_0^T a(t)^2 \mathrm{d}t + a(T) - a(0) \} + 1 - \exp[2 \textstyle\int_0^T a(t)\mathrm{d}t] \\
- 2 \textstyle\int_0^T \exp[2 \textstyle\int_{T-t}^T a(s)\mathrm{d}s] [\textstyle\int_{T-t}^T a(s)^2 \mathrm{d}s + a(T) - a(T-t)] \mathrm{d}t \\
= (1/2)(1 - \exp[-2T]) \ .
\end{aligned}
$$

Proof. The proof is a direct consequence of Proposition S18, Proposition S19 and the fact that

$$
\begin{aligned}
&- \exp[-2T](1 - \lambda T \exp[-2T])(\lambda + 1)^2 / (1 + \lambda \exp[-2T])^2 \\
&\quad = -(T+1) \exp[-2T](1 + \lambda^2)/(1 + \lambda \exp[-2T])^2 \\
&\qquad + T(1+\lambda)^2 \exp[-2T]/(1 + \lambda \exp[-2T]) \ .
\end{aligned}
$$

□

Proposition S21. *Let $\lambda \in (-1, +\infty)$ and $a : [0, T] \to \mathbb{R}$ such that for any $t \in [0, T]$,*

$$
a(t) = 1 - 2/(1 + \exp[-2(T-t)]\lambda) \ .
$$

Then, we have that

$$
\begin{aligned}
&\exp[2 \textstyle\int_0^T a(t)\mathrm{d}t] - 1 - \textstyle\int_0^T \exp[2 \textstyle\int_{T-t}^T a(s)\mathrm{d}s] \{ \textstyle\int_{T-t}^T a(s)^2 \mathrm{d}s + a(T) - 3a(T-t) \} \mathrm{d}t \\
&\quad = \exp[-2T](1+\lambda)^2/(1 + \lambda \exp[-2T])^2 - 1 \\
&\qquad - (1/2)(1+\lambda)(1 - \exp[-2T])/(1 + \lambda \exp[-2T])(1 - 2/(1+\lambda)) \\
&\qquad - (3/2)(1 - \exp[-2T])(1 - \lambda^2 \exp[-2T])/(1 + \lambda \exp[-2T])^2 \\
&\qquad + (T/2)(1+\lambda)^2 \exp[-2T]/(1 + \lambda \exp[-2T]) \\
&\qquad - (1+\lambda)^2/(4\lambda) \log((1+\lambda)/(1 + \lambda \exp[-2T])) \\
&\qquad + (\lambda/2)(1 - \exp[-2T])^2/(1 + \lambda \exp[-2T])^2 \ .
\end{aligned}
$$

In particular, we have

$$
\begin{aligned}
&\exp[2 \textstyle\int_0^T a(t)\mathrm{d}t] - 1 - \textstyle\int_0^T \exp[2 \textstyle\int_{T-t}^T a(s)\mathrm{d}s] \{ \textstyle\int_{T-t}^T a(s)^2 \mathrm{d}s + a(T) - 3a(T-t) \} \mathrm{d}t \\
&\qquad = -2 - (1+\lambda)^2/(4\lambda) \log(1+\lambda) + O(\exp[-2T]) \ .
\end{aligned}
$$

Proof. Using Lemma S11, we have that

$$
\exp[2 \textstyle\int_0^T a(t)\mathrm{d}t] = \exp[-2T](1+\lambda)^2/(1 + \lambda \exp[-2T])^2 \ . \tag{S49}
$$

Using Lemma S12, we have

$$
\textstyle\int_0^T \exp[2 \textstyle\int_{T-t}^T a(s)\mathrm{d}s]\mathrm{d}t = (1/2)(1+\lambda)(1 - \exp[-2t])/(1 + \lambda \exp[-2t]) \ . \tag{S50}
$$

Using Lemma S14, we have

$$
\begin{aligned}
&\textstyle\int_0^T \exp[2 \textstyle\int_{T-t}^T a(s)\mathrm{d}s] a(T-t)\mathrm{d}t \\
&\qquad = -(1/2)(1 - \exp[-2T])(1 - \lambda^2 \exp[-2T])/(1 + \lambda \exp[-2T])^2 \ .
\end{aligned} \tag{S51}
$$

Finally, using Lemma S17, we have

$$
\begin{aligned}
\textstyle\int_0^T \exp[2 \textstyle\int_{T-t}^T a(s)\mathrm{d}s](\textstyle\int_{T-t}^T a(s)^2 \mathrm{d}s)\mathrm{d}t &= -(T/2)(1+\lambda)^2 \exp[-2T]/(1 + \lambda \exp[-2T]) \\
&\quad + (1+\lambda)^2/(4\lambda) \log((1+\lambda)/(1 + \lambda \exp[-2T])) \\
&\quad - (\lambda/2)(1 - \exp[-2T])^2/(1 + \lambda \exp[-2T])^2 \ .
\end{aligned} \tag{S52}
$$

We conclude upon combining (S49), (S50), (S51), (S52) and that $a(T) = 1 - 2/(1+\lambda)$. □

S5.4 General setting

In this section, we prove Theorem 2. In order to compare our results with [5, Theorem 1], we redefine a few processes. Let $p \in \mathscr{P}(\mathbb{R}^d)$ be the target distribution. Consider the Ornstein-Ulhenbeck forward dynamics $(x_t)_{t \in [0,T]}$ such that $\mathrm{d}x_t = -x_t \mathrm{d}t + \sqrt{2}\mathrm{d}w_t$ and x_0 has distribution p_0. We consider the backward chain $(X_k)_{k \in \{0,\dots,N\}}$ such that for any $k \in \{0, \dots, N-1\}$,

$$X_k = X_{k+1} + \gamma_{k+1}\{X_{k+1} + 2\nabla \log p_{t_{k+1}}(X_{k+1})\} + \sqrt{2\gamma_{k+1}} Z_{k+1}, \tag{S53}$$

with $\{Z_k\}_{k \in \mathbb{N}}$ a family of i.i.d. Gaussian random variables with zero mean and identity covariance matrix, $t_k = \sum_{\ell=1}^k \gamma_\ell$, $\sum_{\ell=1}^N \gamma_\ell = T$ and X_N has distribution $p_0 = \mathrm{N}(0, \mathrm{Id})$ (independent from $\{Z_k\}_{k \in \mathbb{N}}$). Notice that here we do not consider a score approximation in the recursion in order to clarify our approximation results. We recall the following result from [5, Theorem 1].

Theorem S22. *Assume that p_0 admits a bounded density (w.r.t. the Lebesgue measure) $p_0 \in \mathrm{C}^3(\mathbb{R}^d, (0, +\infty))$ and that there exist $d_1, A_1, A_2, A_3 \geq 0$, $\beta_1, \beta_2, \beta_3 \in \mathbb{N}$ and $\mathtt{m}_1 > 0$ such that for any $x \in \mathbb{R}^d$ and $i \in \{1, 2, 3\}$*

$$\|\nabla^i \log p_0(x)\| \leq A_i(1 + \|x\|^{\beta_i}), \quad \langle \nabla \log p_0(x), x\rangle \leq -\mathtt{m}_1\|x\|^2 + d_1\|x\|,$$

with $\beta_1 = 1$. Then there exist $B, C, D \geq 0$ such that for any $N \in \mathbb{N}$ and $\{\gamma_k\}_{k=1}^N$ with $\gamma_k > 0$ for any $k \in \{1, \dots, N\}$ we have

$$\|\mathcal{L}(X_0) - p_0\|_{\mathrm{TV}} \leq C \exp[DT]\sqrt{\gamma^\star} + B\exp[-T]. \tag{S54}$$

where $\gamma^\star = \sup_{k \in \{1,\dots,N\}} \gamma_k$ and $\mathcal{L}(X_0)$ is the distribution of X_0 given in (S53).

In the rest of this note we improve the theorem in the following way:

(a) We remove the exponential dependency w.r.t. the time in the first term of the RHS of (S54).

(b) We provide explicit bounds $B, C, D \geq 0$ depending on the parameters of p_0.

Lemma S23. *Assume*

$$\sup_{x,t}\|\nabla^2 \log p_t(x)\| \leq K \;\text{ and }\; \|\partial_t \nabla \log p_t(x)\| \leq M\,\mathrm{e}^{-\alpha t}\,\|x\|.$$

Then there exists $D \geq 0$ such that for any $x \in \mathbb{R}^d$ and $t \in [0, T]$, $\|\nabla \log p_t(x)\| \leq D(1 + \|x\|)$ with $D = \|\nabla \log p_0(0)\| + K + CT$.

Proof. Let $x \in \mathbb{R}^d$ and $t \in [0, T]$. Since $(t, x) \mapsto \log p_t(x) \in \mathrm{C}^2([0, T] \times \mathbb{R}^d, (0, +\infty))$, we have that

$$\begin{aligned}\nabla \log p_t(x) &= \nabla \log p_0(x) + \textstyle\int_0^t \partial_s \nabla \log p_s(x)\mathrm{d}s \\ &= \nabla \log p_0(0) + \textstyle\int_0^1 \nabla^2 \log p_0(ux)(x)\mathrm{d}u + \int_0^t \partial_s \nabla \log p_s(x)\mathrm{d}s.\end{aligned}$$

Therefore, we have that

$$\begin{aligned}\|\nabla \log p_t(x)\| &\leq \|\nabla \log p_0(0)\| + K\|x\| + \textstyle\int_0^t \|\partial_s \nabla \log p_s(x)\|\mathrm{d}s \\ &\leq \|\nabla \log p_0(0)\| + K\|x\| + M\textstyle\sum_{k=0}^{N-1}(t_k - t_{k-1})\exp[-\alpha t_k]\|x\| \\ &\leq \|\nabla \log p_0(0)\| + K\|x\| + MT\|x\|,\end{aligned}$$

which concludes the proof. □

Note that in the previous proposition we can derive a tighter bound for D which does not depend on the limiting time $T > 0$. However, we do not use the bound $D > 0$ in our quantitative result and therefore our simple bound suffices.

We also have the following useful lemma.

Lemma S24. *Let $T \geq \log(2\mathbb{E}[\|X_0\|^2]) + \log(2)/2$ and assume that there exists $\eta > 0$ such that aaa Then, we have*

$$\textstyle\int_{\mathbb{R}^d} p_\infty(x_T)^2/p_T(x_T)\mathrm{d}x_T \leq \exp[4] + E_T\ ,$$

with $E_T \sim C\exp[-T]$ when $T \to +\infty$ and $C \geq 0$.

If p_∞ satisfies the following Φ-entropy inequality for any $f : \mathbb{R}_d \to (0, \infty)$ measurable

$$\int_{\mathbb{R}^d} \|\nabla f(x)\|^2 / f(x)^3 p_\infty(x) \mathrm{d}x \leq C[\int_{\mathbb{R}^d} (1/f(x)) p_\infty(x) \mathrm{d}x - 1/(\int_{\mathbb{R}^d} f(x) p_\infty(x) \mathrm{d}x)] \ , \quad \text{(S55)}$$

with $C \geq 0$. Then, we have as in [1, Proposition 7.6.1]

$$\chi^2(p_\infty || p_t) = \int_{\mathbb{R}^d} p_\infty^2(x)/p_T(x) \mathrm{d}x - 1 \leq \mathrm{e}^{-Ct} \ ,$$

which immediately concludes the proof of Lemma S24. However, to the best of our knowledge, establishing (S55) remains an open problem. Note that controlling $\chi^2(p_t || p_\infty)$ is much easier as the exponential decay of this divergence is linked with the Poincaré inequality which is satisfied in our Gaussian setting. In what follows, we consider another approach which relies on the structure of the Ornstein-Ulhenbeck transition kernel and provide non-tight upper bounds.

Proof. Let $T \geq 0$, $\varepsilon > 0$ and $x_T \in \mathbb{R}^d$

$$\|x_T - \mathrm{e}^{-T} x_0\|^2 \leq (1 + \varepsilon)\|x_T\|^2 + (1 + 1/\varepsilon)\mathrm{e}^{-2T}\|x_0\|^2 \ .$$

Let $\varepsilon > 0$ and $x_T \in \mathbb{R}^d$, we have

$$p_T(x_T)^{-2} \leq \exp[(1 + \varepsilon)/\sigma_T^2 \|x_T\|^2] \\ \times (\int_{\mathbb{R}^d} p(x_0) \exp[-\mathrm{e}^{-2T}(1 + 1/\varepsilon)/(2\sigma_T^2)\|x_0\|^2] \mathrm{d}x_0)^{-2} (2\pi\sigma_T^2)^d \ .$$

For any $x_T \in \mathbb{R}^d$, we have

$$p_\infty(x_T)^2 / p_T(x_T) \leq \exp[\{-1 + (1 + \varepsilon)/(2\sigma_T^2)\}\|x_T\|^2](2\pi/\sigma_T^2)^{-d/2} \\ (\int_{\mathbb{R}^d} p(x_0) \exp[-\mathrm{e}^{-2T}(1 + 1/\varepsilon)/(2\sigma_T^2)\|x_0\|^2] \mathrm{d}x_0)^{-1} \ .$$

In what follows, we set $\varepsilon = \mathrm{e}^{-T}$. We have that

$$-1 + (1 + \varepsilon)/(2\sigma_T^2) = (2\sigma_T^2)^{-1}(-2\sigma_T^2 + 1 + \varepsilon) = -(1 - 2\mathrm{e}^{-T} + \varepsilon)/(2\sigma_T^2) = -(1 - \mathrm{e}^{-T})/(2\sigma_T^2) \ .$$

Therefore, we get that

$$\int_{\mathbb{R}^d} \exp[\{-1 + (1 + \varepsilon)/(2\sigma_T^2)\}\|x_T\|^2](2\pi/\sigma_T^2)^{-d/2} \mathrm{d}x_T = (1 - \mathrm{e}^{-T})^{-d/2} \ . \quad \text{(S56)}$$

In addition, we have that for any $R \geq 0$ using that $\sigma_T^2 \geq 1/2$ since $T \geq \log(2)/2$

$$\begin{aligned} &\int_{\mathbb{R}^d} p(x_0) \exp[-\mathrm{e}^{-2T}(1 + 1/\varepsilon)/\sigma_T^2 \|x_0\|^2] \mathrm{d}x_0 \\ &\qquad \geq \mathbb{P}(X_0 \in \bar{\mathrm{B}}(0, R)) \exp[-\mathrm{e}^{-2T}(1 + 1/\varepsilon)/\sigma_T^2 R^2] \\ &\qquad \geq \mathbb{P}(X_0 \in \bar{\mathrm{B}}(0, R)) \exp[-4\mathrm{e}^{-T} R^2] \end{aligned} \quad \text{(S57)}$$

Now let $R^2 = \mathrm{e}^T$. We obtain

$$\int_{\mathbb{R}^d} p(x_0) \exp[-\mathrm{e}^{-2T}(1 + 1/\varepsilon)/\sigma_T^2 \|x_0\|^2] \mathrm{d}x_0 \geq \mathbb{P}(X_0 \in \bar{\mathrm{B}}(0, \mathrm{e}^{T/2})) \exp[-4] \ .$$

In addition, using Markov inequality, we have

$$\mathbb{P}(X_0 \in \bar{\mathrm{B}}(0, \mathrm{e}^{T/2})) = 1 - \mathbb{P}(\|X_0\|^2 \geq \mathrm{e}^T) \geq 1 - \mathbb{E}[\|X_0\|^2]\mathrm{e}^{-T} \geq 0 \ .$$

Therefore, combining this result and (S57), we have

$$\int_{\mathbb{R}^d} p(x_0) \exp[-\mathrm{e}^{-2T}(1 + 1/\varepsilon)/\sigma_T^2 \|x_0\|^2] \mathrm{d}x_0 \geq \exp[-4](1 - \mathbb{E}[\|X_0\|^2]\mathrm{e}^{-T}) > 0 \ . \quad \text{(S58)}$$

We conclude upon combining (S56) and (S58). □

We are now ready to state the following lemma.

Lemma S25. *There exists a unique strong solution to the SDE* $\mathrm{d}y_t = \{y_t + 2\nabla \log p_{T-t}(y_t)\}\mathrm{d}t + \sqrt{2}\mathrm{d}w_t$ *with initial condition* $\mathcal{L}(y_0) = p_\infty$*. In addition, we have that* $\mathbb{E}[\sup_{t \in [0,T]} \|y_t\|^\alpha] < +\infty$ *for any* $\alpha > 0$.

Proof. Let $b: [0,T] \times \mathbb{R}^d$ given for any $t \in [0,T]$ and $x \in \mathbb{R}^d$ by $b(t,x) = x + 2\nabla \log p_t(x)$. We have that $b \in \mathrm{C}^1([0,T] \times \mathbb{R}^d, \mathbb{R}^d)$ and in particular is locally Lipschitz. In addition, using Lemma S23 we have that for any $t \in [0,T]$ and $x \in \mathbb{R}^d$, $\|b(t,x)\| \leq (1+D)\|x\|$. Hence using [14, Theorem 2.3, Theorem 3.1] and [35, Theorem 2.1] (with $V(x) = (1/2)\|x\|^2$) there exists a unique strong solution to the SDE $\mathrm{d}y_t = \{y_t + 2\nabla \log p_{T-t}(y_t)\}\mathrm{d}t + \sqrt{2}\mathrm{d}w_t$ with initial condition $\mathcal{L}(y_0) = p_\infty$. Let $\alpha > 1$, then we have for any $t \in [0,T]$

$\sup_{s\in[0,t]} \|y_t\|^\alpha \leq 3^{\alpha-1}[\|y_0\|^\alpha + t^{\alpha-1}(1+D)^\alpha \int_0^t \sup_{u\in[0,s]} \|y_u\|^\alpha \mathrm{d}u + 2^{\alpha/2} \sup_{s\in[0,t]} \|w_u\|^\alpha]$.

Using that $\mathbb{E}[\sup_{s\in[0,T]} \|w_u\|^\alpha]$ and Grönwall's lemma, we get that $\mathbb{E}[\sup_{t\in[0,T]} \|y_t\|^\alpha] < +\infty$ for any $\alpha > 1$. The result is extended to any $\alpha > 0$ since for any $\alpha \in (0,1]$ we have that

$$\mathbb{E}[\sup_{t\in[0,T]} \|y_t\|^\alpha] \leq \mathbb{E}[\sup_{t\in[0,T]} \|y_t\|]^\alpha < +\infty.$$

□

We are now ready to prove Theorem 2.

Proof. The beginning of the proof is similar to the one of [5, Theorem 1]. For any $k \in \{1,\dots,N\}$, denote R_k the Markov kernel such that for any $x \in \mathbb{R}^d$, $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d)$ and $k \in \{0,\dots,N-1\}$ we have

$$\mathrm{R}_{k+1}(x,\mathsf{A}) = (4\pi\gamma_{k+1})^{-d/2} \int_\mathsf{A} \exp[-\|\tilde{x} - \mathcal{T}_{k+1}(x)\|^2/(4\gamma_{k+1})]\mathrm{d}\tilde{x},$$

where for any $x \in \mathbb{R}^d$, $\mathcal{T}_{k+1}(x) = x + \gamma_{k+1}\{x + 2\nabla \log p_{t_{k+1}}(x)\}$. Define for any $k_0, k_1 \in \{1,\dots,N\}$ with $k_1 \geq k_0$ $\mathrm{Q}_{k_0,k_1} = \prod_{\ell=k_0}^{k_1} \mathrm{R}_{k_1+k_0-\ell}$. Finally, for ease of notation, we also define for any $k \in \{1,\dots,N\}$, $\mathrm{Q}_k = \mathrm{Q}_{k+1,N}$. Note that for any $k \in \{1,\dots,N\}$, X_k has distribution $p_\infty \mathrm{Q}_k$, where $p_\infty \in \mathscr{P}(\mathbb{R}^d)$ with density w.r.t. the Lebesgue measure p_∞. Let $\mathbb{P} \in \mathscr{P}(\mathcal{C})$ be the probability measure associated with the diffusion

$$\mathrm{d}x_t = -x_t\mathrm{d}t + \sqrt{2}\mathrm{d}w_t, \quad x_0 \sim p_0,$$

First, we have for any $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d)$

$$p_0\mathbb{P}_{T|0}(\mathbb{P}^R)_{T|0}(\mathsf{A}) = \mathbb{P}_T(\mathbb{P}^R)_{T|0}(\mathsf{A}) = (\mathbb{P}^R)_0(\mathbb{P}^R)_{T|0}(\mathsf{A}) = (\mathbb{P}^R)_T(\mathsf{A}) = p_0(\mathsf{A}).$$

Hence $p_0 = p_0\mathbb{P}_{T|0}(\mathbb{P}^R)_{T|0}$. Using this result we have

$$\begin{aligned}
\|p_0 - p_\infty \mathrm{Q}_0\|_{\mathrm{TV}} &= \|p_0\mathbb{P}_{T|0}(\mathbb{P}^R)_{T|0} - p_\infty \mathrm{Q}_0\|_{\mathrm{TV}} \\
&\leq \|p_0\mathbb{P}_{T|0}(\mathbb{P}^R)_{T|0} - p_\infty(\mathbb{P}^R)_{T|0}\|_{\mathrm{TV}} + \|p_\infty(\mathbb{P}^R)_{T|0} - p_\infty \mathrm{Q}_0\|_{\mathrm{TV}} \\
&\leq \|p_0\mathbb{P}_{T|0} - p_\infty\|_{\mathrm{TV}} + \|p_\infty(\mathbb{P}^R)_{T|0} - p_\infty \mathrm{Q}_0\|_{\mathrm{TV}}.
\end{aligned}$$

Note that $\mathcal{L}(X_0) = p_\infty \mathrm{Q}_0$ and therefore

$$\|\mathcal{L}(X_0) - p_0\|_{\mathrm{TV}} \leq \|p_0\mathbb{P}_{T|0} - p_\infty\|_{\mathrm{TV}} + \|p_\infty(\mathbb{P}^R)_{T|0} - p_\infty \mathrm{Q}_0\|_{\mathrm{TV}}.$$

We now bound each one of these terms.

(a) First, we bound $\|p_0\mathbb{P}_{T|0} - p_\infty\|_{\mathrm{TV}}$. Using the Pinsker inequality [1, Equation 5.2.2] we have that

$$\|p_0\mathbb{P}_{T|0} - p_\infty\|_{\mathrm{TV}} \leq \sqrt{2}\,\mathrm{KL}(p_0\mathbb{P}_{T|0}\|p_\infty)^{1/2}. \tag{S59}$$

In addition, p_∞ satisfies the log-Sobolev inequality with constant $C = 1$, [8]. Namely, for any $f \in \mathrm{C}^1(\mathbb{R}^d, (0,+\infty))$ such that $f \in \mathrm{L}^1(p_\infty)$ and $\int_{\mathbb{R}^d} \|\nabla \log f(x)\|^2 f(x)\mathrm{d}p_\infty(x) < +\infty$ we have

$$\begin{aligned}
&\int_{\mathbb{R}^d} f(x)\log f(x)\mathrm{d}p_\infty(x) - (\int_{\mathbb{R}^d} f(x)\mathrm{d}p_\infty(x))(\log \int_{\mathbb{R}^d} f(x)\mathrm{d}p_\infty(x)) \\
&\qquad \leq (C/2)\int_{\mathbb{R}^d} \|\nabla \log f(x)\|^2 f(x)\mathrm{d}p_\infty(x),
\end{aligned}$$

with $C = 1$. Therefore, using [1, Theorem 5.2.1] we have that for any $f \in \mathrm{L}^1(p_\infty)$ with $\int_{\mathbb{R}^d} |f(x)||\log f(x)|\mathrm{d}p_\infty(x) < +\infty$

$$\mathrm{Ent}_{p_\infty}(\mathbb{P}_{T|0}[f]) \leq \exp[-2T]\mathrm{Ent}_{p_\infty}(f), \tag{S60}$$

where for any $g \in \mathrm{L}^1(p_\infty)$ with $\int_{\mathbb{R}^d} |g(x)||\log g(x)| \mathrm{d}p_\infty(x) < +\infty$ we define

$$\mathrm{Ent}_{p_\infty}(g) = \int_{\mathbb{R}^d} g(x) \log g(x) \mathrm{d}p_\infty(x) - (\int_{\mathbb{R}^d} g(x) \mathrm{d}p_\infty(x))(\log \int_{\mathbb{R}^d} g(x) \mathrm{d}p_\infty(x)).$$

Note that $(\mathrm{d}p_T/\mathrm{d}p_\infty) = \mathbb{P}_{T|0}[\mathrm{d}p_0/\mathrm{d}p_\infty]$ and that for any $\mu \in \mathscr{P}(\mathbb{R}^d)$ with $\mathrm{KL}(\mu \| p_\infty) < +\infty$ we have $\mathrm{Ent}_{p_\infty}(\mathrm{d}\mu/\mathrm{d}p_\infty) = \mathrm{KL}(\mu \| p_\infty)$. Using these results, (S59) and (S60) we get that

$$\|p_0 \mathbb{P}_{T|0} - p_\infty\|_{\mathrm{TV}} \leq \sqrt{2} \exp[-T] \, \mathrm{KL}(p_0 \| p_\infty)^{1/2}. \tag{S61}$$

In addition, we have that

$$\mathrm{KL}(p_0 \| p_\infty) = (d/2) \log(2\pi) + \int_{\mathbb{R}^d} \|x\|^2 \mathrm{d}p_0(x) - \mathrm{H}(p_0),$$

where $\mathrm{H}(p_0) = -\int_{\mathbb{R}^d} \log(p_0(x)) p_0(x) \mathrm{d}x$. Combining this result and (S61) we get that

$$\|p_0 \mathbb{P}_{T|0} - p_\infty\|_{\mathrm{TV}} \leq \sqrt{2} \exp[-T]((d/2) \log(2\pi) + \int_{\mathbb{R}^d} \|x\|^2 \mathrm{d}p_0(x) - \mathrm{H}(p_0))^{1/2},$$

which concludes the first part of the proof.

(b) First, let $\mathbb{Q} \in \mathscr{P}(\mathcal{C})$ such that $\mathbb{Q} = p_\infty \mathbb{P}^R_{|0}$, where $\mathbb{P}^R_{|0}$ is the disintegration of $\mathbb{P}^R$ w.r.t. $\phi : \mathcal{C} \to \mathbb{R}^d$ given for any $\omega \in \mathcal{C}$ by $\phi(\omega) = \omega_T$, see [25] for instance. Note that for any $f \in \mathrm{C}(\mathcal{C})$ with f bounded we have

$$\begin{aligned}\mathbb{Q}[f] = \int_{\mathbb{R}^d} \int_{\mathcal{C}} f(\omega) \mathbb{P}^R_{|0}(\omega_0, \mathrm{d}\omega) \mathrm{d}p_\infty(\omega_0) &= \int_{\mathbb{R}^d} \int_{\mathcal{C}} f(\omega) \mathbb{P}^R_{|0}(\omega_0, \mathrm{d}\omega)(\mathrm{d}p_\infty/\mathrm{d}p_T)(\omega_0) \mathrm{d}p_T(\omega_0) \\ &= \int_{\mathcal{C}} f(\omega)(\mathrm{d}p_\infty/\mathrm{d}p_T)(\omega_0) \mathrm{d}\mathbb{P}^R(\omega).\end{aligned}$$

Therefore, we get that for any $\omega \in \mathcal{C}$, $(\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}^R)(\omega) = (\mathrm{d}p_\infty/\mathrm{d}p_T)(\omega_0)$. Let $\mathbb{R} = p_\infty \mathbb{P}_{|0}$. Note that for any $t \in [0, T]$, $\mathbb{R}_t = p_\infty$ and that $\mathbb{R}$ is associated with the process $\mathrm{d}x_t = -x_t \mathrm{d}t + \sqrt{2} \mathrm{d}w_t$ with $\mathcal{L}(x_0) = p_\infty$. In particular, $\mathbb{R}$ satisfies [2, Hypothesis 1.8]. Using [25, Theorem 2.4] we have that

$$\mathrm{KL}(\mathbb{P} \| \mathbb{R}) = \mathrm{KL}(p_0 \| p_\infty) + \int_{\mathbb{R}^d} \mathrm{KL}(\mathbb{P}_{|0}(x_0) \| \mathbb{P}_{|0}(x_0)) \mathrm{d}p_0(x_0) = \mathrm{KL}(p_0 \| p_\infty) < +\infty.$$

Therefore, we can apply [2, Theorem 4.9]. Let $u \in \mathrm{C}^\infty_c(\mathbb{R}^d, \mathbb{R})$, we have that $(\mathbf{M}^u_t(y))_{t \in [0,T]}$ is a local martingale, where we have for any $t \in [0, T]$

$$\mathbf{M}^u_t(y) = u(y_t) - u(y_0) - \int_0^t \{\langle \nabla u(y_s), y_s + 2\nabla \log p_{T-s}(y_s) \rangle + \Delta u(y_s)\} \mathrm{d}s,$$

where $\mathcal{L}(y) = \mathbb{P}^R$. Since u is compactly supported we have that $\sup_{\omega \in \mathcal{C}} \sup_{t \in [0,T]} |\mathbf{M}^u_t(\omega)| < +\infty$ and therefore $(\mathbf{M}^u_t(y))_{t \in [0,T]}$ is a martingale. We now show that $(\mathbf{M}^u_t(y))_{t \in [0,T]}$ is a martingale, with $\mathcal{L}(y) = \mathbb{Q}$. Since $\sup_{\omega \in \mathcal{C}} \sup_{t \in [0,T]} |\mathbf{M}^u_t(\omega)| < +\infty$, we have that for any $t \in [0, T]$, $\mathbb{E}[|\mathbf{M}^u_t|] < +\infty$. Let $t, s \in [0, T]$ with $t > s$ and $g : \mathcal{C} \to \mathbb{R}^d$ bounded. We have that $\mathbb{E}[|g(\{x_{T-s}\}_{s \in [0,t]})|^2 (\mathrm{d}p_\infty/\mathrm{d}p_T)(x_T)^2] < +\infty$. Hence, we have that

$$\mathbb{E}[(\mathbf{M}^u_t(x_{T-\cdot}) - \mathbf{M}^u_s(x_{T-\cdot}))g(\{x_{T-s}\}_{s \in [0,t]})(\mathrm{d}p_\infty/\mathrm{d}p_T)(x_T)] = 0.$$

Using this result and that for any $\omega \in \mathcal{C}$, $(\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}^R)(\omega) = (\mathrm{d}p_\infty/\mathrm{d}p_T)(\omega_0)$ we get

$$\mathbb{E}[(\mathbf{M}^u_t(y) - \mathbf{M}^u_s(y))g(\{y_s\}_{s \in [0,t]})] = 0.$$

Hence, for any $u \in \mathrm{C}^2_c(\mathbb{R}^d, \mathbb{R})$, $(\mathbf{M}^u_t(y))_{t \in [0,T]}$ is a martingale. In addition, $(\mathbf{M}^u_t(\mathbf{Z}))_{t \in [0,T]}$ is a martingale using Lemma S25 and Itô's lemma, where $\mathbf{Z}$ is the solution to the SDE in Lemma S25. In addition, we have that $\mathcal{L}(\mathbf{Z}_0) = \mathcal{L}(y_0) = p_\infty$. Using Lemma S23 and the remark following [2, Hypothesis 1.8], we get that $\mathcal{L}(\mathbf{Z}) = \mathcal{L}(y) = \mathbb{Q}$. We have just shown that the time-reversed process with initialisation p_∞ can be obtained as a strong solution of an SDE. Using Lemma S23 and Lemma S25, we have that for any $t \in [0, T]$

$$\mathbb{E}[\int_0^t \|x_s + 2\nabla \log p_s(x_s)\|^2 \mathrm{d}s + \int_0^t \|w_s + 2\nabla \log p_s(w_s)\|^2 \mathrm{d}s] < +\infty.$$

Combining this result and [5, Lemma S13] we have that

$$\|p_\infty \mathbb{P}^R_{T|0} - p_\infty \mathrm{Q}_0\|^2_{\mathrm{TV}} \leq (1/2) \int_0^T \mathbb{E}[\|b_1(t, (y_s)_{s \in [0,T]}) - b_2(t, (y_s)_{s \in [0,T]})\|^2] \mathrm{d}t, \tag{S62}$$

where for any $t \in [0,T]$ and $\omega \in \mathcal{C}$ we have that

$$b_1(t,\omega) = \omega_t + 2\nabla \log p_{T-t}(\omega_t), \qquad b_2(t,\omega) = \omega_{t_\gamma} + 2\nabla \log p_{T-t_\gamma}(\omega_{t_\gamma}),$$

where $t_\gamma = \sum_{k=0}^{N-1} \mathbb{1}_{[T-t_{k+1},T-t_k)}(t)(T-t_{k+1})$. Noting that $(y_t)_{t\in[0,T]}$ is distributed according to $\mathbb{Q}$ and using that $(\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}^R)(\omega) = (\mathrm{d}p_\infty/\mathrm{d}p_T)(\omega_0)$, (S62) and the Cauchy-Schwarz inequality we have

$$\begin{aligned}
&\|p_\infty \mathbb{P}^R_{T|0} - p_\infty \mathrm{Q}_0\|^2_{\mathrm{TV}} \qquad\qquad (\mathrm{S}63)\\
&\leq (1/2)\mathbb{E}[(\mathrm{d}p_\infty/\mathrm{d}p_T)(x_T)^2]^{1/2} \textstyle\int_0^T \mathbb{E}^{1/2}[\|b_1(t,(x_{T-s})_{s\in[0,T]}) - b_2(t,(x_{T-s})_{s\in[0,T]})\|^4]\mathrm{d}t\\
&\leq (1/2)\mathbb{E}[(\mathrm{d}p_\infty/\mathrm{d}p_T)(x_T)^2]^{1/2}\\
&\qquad \times \textstyle\int_0^T \mathbb{E}^{1/2}[\|b_1(T-t,(x_{T-s})_{s\in[0,T]}) - b_2(T-t,(x_{T-s})_{s\in[0,T]})\|^4]\mathrm{d}t.
\end{aligned}$$

In addition, we have that for any $t \in [0,T]$ and $\omega \in \mathcal{C}$ we have

$$\begin{aligned}
&\|b_1(t,\omega) - b_2(t,\omega)\|\\
&\leq \|\omega_t - \omega_{t_\gamma}\| + 2\|\nabla \log p_{T-t}(\omega_t) - \nabla \log p_{T-t_\gamma}(\omega_t)\|\\
&\qquad\qquad + 2\|\nabla \log p_{T-t_\gamma}(\omega_t) - \nabla \log p_{T-t_\gamma}(\omega_{t_\gamma})\|\\
&\leq (1 + 2\sup_{s\in[0,T]}\sup_{x\in\mathbb{R}^d} \|\nabla^2 \log p_s(x)\|)\|\omega_t - \omega_{t_\gamma}\|\\
&\qquad\qquad +2\sup_{s\in[T-t,T-t_\gamma]} \|\partial_t \nabla \log p_t(\omega_t)\|(t-t_\gamma)\\
&\leq (1+2K)\|\omega_t - \omega_{t_\gamma}\| + 2\sup_{s\in[T-t,T-t_\gamma]} \|\partial_s \nabla \log p_s(\omega_t)\|(t-t_\gamma).
\end{aligned}$$

Note that

$$T - (T-t)_\gamma = T - \textstyle\sum_{k=0}^{N-1} \mathbb{1}_{[T-t_{k+1},T-t_k)}(T-t)(T-t_{k+1}) = \sum_{k=0}^{N-1} \mathbb{1}_{(t_k,t_{k+1}]}(t)t_{k+1}.$$

For any $t \in [0,T]$, denote $t^\gamma = T - (T-t)_\gamma = \sum_{k=0}^{N-1} \mathbb{1}_{(t_k,t_{k+1}]}(t)t_{k+1}$. Therefore, we get that for any $t \in (t_k, t_{k+1}]$

$$\begin{aligned}
&\|b_1(T-t,\omega) - b_2(T-t,\omega)\|\\
&\leq (1+2K)\|\omega_{T-t} - \omega_{(T-t)_\gamma}\| + 2\sup_{s\in[t,t^\gamma]} \|\partial_s \nabla \log p_s(\omega_{T-t})\|(t^\gamma - t)\\
&\leq (1+2K)\|\omega_{T-t} - \omega_{(T-t)_\gamma}\| + 2\sup_{s\in[t_k,t_{k+1}]} \|\partial_s \nabla \log p_s(\omega_{T-t})\|\gamma_{k+1}\\
&\leq (1+2K)\|\omega_{T-t} - \omega_{(T-t)_\gamma}\| + 2\mathrm{S}_{t_k}(\omega_{T-t})\gamma_{k+1}.
\end{aligned}$$

Combining this result and that for any $a,b \geq 0$, $(a+b)^4 \leq 8a^4 + 8b^4$ we get that for any $t \in (t_k, t_{k+1}]$

$$\begin{aligned}
&\mathbb{E}[\|b_1(T-t,(x_{T-s})_{s\in[0,T]}) - b_2(T-t,(x_{T-s})_{s\in[0,T]})\|^4] \qquad (\mathrm{S}64)\\
&\qquad \leq 8(1+2K)^4\mathbb{E}[\|x_t - x_{t_k}\|^4] + 16\mathbb{E}[\mathrm{S}_{t_k}(x_t)^4]\gamma^4_{k+1}.
\end{aligned}$$

In addition, we have that for any $t \in [0,T]$, $x_t = \exp[-t]x_0 + w_{(1-\exp[-2t])^{1/2}}$. Hence, for any $s,t \in [0,T]$ with $t > s$ we have

$$\|x_t - x_s\| \leq \exp[-s](\exp[t-s]-1)\|x_0\| + \|w_{(1-\exp[-2t])} - w_{(1-\exp[-2s])}\|.$$

Therefore, we have that for any $s,t \in [0,T]$ with $t > s$

$$\begin{aligned}
\mathbb{E}[\|x_t - x_s\|^4] &\leq 8\exp[-4s](1-\exp[-t+s])^4\mathbb{E}[\|x_0\|^4] + 8\mathbb{E}[\|w_{(1-\exp[-2t])} - w_{(1-\exp[-2s])}\|^4]\\
&\leq 8\exp[-4s](1-\exp[-t+s])^4\mathbb{E}[\|x_0\|^4] + 24(\exp[-t]-\exp[-s])^2\\
&\leq 8\exp[-4s](1-\exp[-t+s])^4\mathbb{E}[\|x_0\|^4] + 24\exp[-2s](1-\exp[-t+s])^2\\
&\leq 8\mathbb{E}[\|x_0\|^4]\exp[-4s](t-s)^4 + 24\exp[-2s](t-s)^2. \qquad (\mathrm{S}65)
\end{aligned}$$

In addition, using that that for any $k \in \{0,\dots,N-1\}$ and $x \in \mathbb{R}^d$, $\mathtt{S}_{t_k}(x) \leq M\exp[-\alpha t_k]\|x\|$ we get that

$$\mathbb{E}[\mathtt{S}_{t_k}(x_t)^4] \leq 24M^4\exp[-4\alpha t_k]\{1 + \mathbb{E}[\|x_0\|^4]\}.$$

Combining this result, (S64) and (S65) we get that for any $t \in (t_k, t_{k+1}]$

$$\begin{aligned}
&\mathbb{E}[\|b_1(T-t,(x_{T-s})_{s\in[0,T]}) - b_2(T-t,(x_{T-s})_{s\in[0,T]})\|^4] \\
&\qquad \leq 64(1+2K)^4 \mathbb{E}[\|x_0\|^4]\exp[-4t_k]\gamma_{k+1}^4 \\
&\qquad\qquad + 192(1+2K)^4\exp[-2t_k]\gamma_{k+1}^2 + 384M^4\exp[-4\alpha t_k]\{1+\mathbb{E}[\|x_0\|^4]\}\gamma_{k+1}^4 .
\end{aligned}$$

Using this result and that for any $a, b \geq 0$, $(a+b)^{1/2} \leq a^{1/2} + b^{1/2}$, we have for any $t \in (t_k, t_{k+1}]$

$$\begin{aligned}
&\mathbb{E}^{1/2}[\|b_1(T-t,(x_{T-s})_{s\in[0,T]}) - b_2(T-t,(x_{T-s})_{s\in[0,T]})\|^4] \\
&\qquad \leq 8(1+2K)^2\mathbb{E}^{1/2}[\|x_0\|^4]\exp[-2t_k]\gamma_{k+1}^2 \\
&\qquad\qquad + 14(1+2K)^2\exp[-t_k]\gamma_{k+1} + 20M^2\exp[-2\alpha t_k]\{1+\mathbb{E}^{1/2}[\|x_0\|^4]\}\gamma_{k+1}^2 .
\end{aligned} \tag{S66}$$

We have that for any $\beta > 0$,

$$\textstyle\sum_{k=0}^{N-1}\exp[-\beta t_k] \leq \sum_{k\in\mathbb{N}}\exp[-\beta\gamma_\star k] \leq (1-\exp[-\beta\gamma_\star])^{-1} \leq 1+\beta/\gamma_\star .$$

Then using this result, (S66) and (S63) we get that

$$\begin{aligned}
\|p_\infty \mathbb{P}^R_{T|0} - p_\infty \mathrm{Q}_0\|^2_{\mathrm{TV}} &\leq \mathbb{E}[(\mathrm{d}p_\infty/\mathrm{d}p_T)(x_T)^2]^{1/2}[4(1+2K)^2\mathbb{E}^{1/2}[\|x_0\|^4](1+/(2\gamma_\star))(\gamma^\star)^3 \\
&\qquad + 7(1+2K)^2(1+1/\gamma_\star)(\gamma^\star)^2 + 10M^2\{1+\mathbb{E}^{1/2}[\|x_0\|^4]\}(1+1/(2\alpha\gamma_\star))(\gamma^\star)^3].
\end{aligned}$$

Therefore, we get that

$$\begin{aligned}
\|p_\infty \mathbb{P}^R_{T|0} - p_\infty \mathrm{Q}_0\|_{\mathrm{TV}} &\leq \mathbb{E}[(\mathrm{d}p_\infty/\mathrm{d}p_T)(x_T)^2]^{1/4}[2(1+2K)\mathbb{E}^{1/4}[\|x_0\|^4](1+/(2\gamma_\star)^{1/2})(\gamma^\star)^{3/2} \\
&\qquad + 3(1+2K)(1+1/\gamma_\star^{1/2})\gamma^\star + 4M\{1+\mathbb{E}^{1/4}[\|x_0\|^4]\}(1+1/(2\alpha\gamma_\star)^{1/2})(\gamma^\star)^{3/2}] \\
&\leq \mathbb{E}[(\mathrm{d}p_\infty/\mathrm{d}p_T)(x_T)^2]^{1/4}[6(1+2K)(1+\mathbb{E}^{1/4}[\|x_0\|^4]) \\
&\qquad + 4M\{1+\mathbb{E}^{1/4}[\|x_0\|^4]\}(1+1/(2\alpha)^{1/2})]((\gamma^\star)^2/\gamma_\star)^{1/2} \\
&\leq 6(1+\mathbb{E}^{1/4}[\|x_0\|^4])\mathbb{E}[(\mathrm{d}p_\infty/\mathrm{d}p_T)(x_T)^2]^{1/4}[1+K+M(1+1/(2\alpha)^{1/2})]((\gamma^\star)^2/\gamma_\star)^{1/2},
\end{aligned}$$

which concludes the proof upon using Lemma S24.

□

We now check that the assumption of Theorem 2 are satisfied in a Gaussian setting.

Proposition S26. *Assume that $p_0 = \mathrm{N}(0,\Sigma)$ and that $T \geq 1 + (1/2)[\log^+(\|\Sigma\|) + \log(d+1)]$ then we have that for any $t \in [0,T]$ and $x \in \mathbb{R}^d$*

$$\|\nabla^2 \log p_t(x)\| \leq \max(1, \|\Sigma^{-1}\|), \qquad \|\partial_t \nabla \log p_t(x)\| \leq 2\exp[-2t]\max(1,\|\Sigma^{-1}\|)^2\|\Sigma - \mathrm{Id}\,\|\|x\|.$$

In addition, we have that $\int_{\mathbb{R}^d} p_\infty(x)^2/p_T(x)\mathrm{d}x \leq \sqrt{2}$.

Proof. Recall that for any $t \in [0,T]$, $x_t = \exp[-t]x_0 + w_{1-\exp[-2t]}$. Therefore, we have that for any $t \in [0,T]$, $p_t = \mathrm{N}(0,\Sigma_t)$ with $\Sigma_t = \exp[-2t]\Sigma + (1-\exp[-2t])\,\mathrm{Id}$. Hence, we get that for any $t \in [0,T]$ and $x \in \mathbb{R}^d$, $\nabla^2 \log p_t(x) = (\exp[-2t]\Sigma + (1-\exp[-2t]\,\mathrm{Id})^{-1}$. Using this result, we have that for any $t \in [0,T]$ and $x \in \mathbb{R}^d$, $\|\nabla^2 \log p_t(x)\| \leq \max(1, \|\Sigma^{-1}\|)$. Similarly, for any $t \in [0,T]$ and $x \in \mathbb{R}^d$ we have

$$\partial_t \nabla \log p_t(x) = \partial_t \Sigma_t^{-1} x = -\Sigma_t^{-1}(\partial_t \Sigma_t)\Sigma_t^{-1} x .$$

Hence, for any $t \in [0,T]$ and $x \in \mathbb{R}^d$ we have $\|\partial_t \nabla \log p_t(x)\| \leq 2\exp[-2t]\max(1,\|\Sigma^{-1}\|)^2\|\Sigma - \mathrm{Id}\,\|\|x\|$. Finally, we have that for any $t \in [0,T]$ and $x \in \mathbb{R}^d$

$$\langle x, [2\,\mathrm{Id} - (\exp[-2t]\Sigma + (1-\exp[-2t])\,\mathrm{Id})^{-1}]x\rangle \geq (2 - (\exp[-2t]\|\Sigma^{-1}\|^{-1} + (1-\exp[-2t]))^{-1})\|x\|^2 .$$

Let $\varepsilon \in (0,1/2]$. For any $t \in [0,T]$, we have that $2-(\exp[-2t]\|\Sigma^{-1}\|^{-1}+(1-\exp[-2t]))^{-1} \geq 1-\varepsilon$ if and only if $\exp[-2t](1-\|\Sigma^{-1}\|^{-1}) \leq 1-(1+\varepsilon)^{-1}$. Using that $-\log(1-(1+\varepsilon)^{-1}) = \log(1+\varepsilon^{-1})$ we have that for any $t \geq (1/2)\log(1+\varepsilon^{-1})$ and $x \in \mathbb{R}^d$

$$p_\infty(x)^2/p_t(x) \leq \exp[-\|x\|^2/4](2\pi)^{-d/2}\det(\Sigma_t).$$

Combining this result and the fact that $\int_{\mathbb{R}^d}\exp[-\|x\|^2/2(1-\varepsilon)]\mathrm{d}x = (2(1-\varepsilon)\pi)^{d/2}$, we get that for any $t \geq (1/2)\log(1+\varepsilon^{-1})$

$$\int_{\mathbb{R}^d} p_\infty(x)^2/p_t(x)\mathrm{d}x \leq \int_{\mathbb{R}^d}\exp[-\|x\|^2/2(1-\varepsilon)](2\pi)^{-d/2}\det(\Sigma_t)^{1/2}\mathrm{d}x \leq (1-\varepsilon)^{d/2}\det(\Sigma_t)^{1/2}.$$

Let $\varepsilon = 1/(2d) \leq 1/2$. Note that $T \geq (1/2)\{-\log(|\|\Sigma^{-1}\|^{-1} - 1|) + \log(1+2d)\}$. Hence, we have that

$$\int_{\mathbb{R}^d} p_\infty(x)^2/p_T(x)\mathrm{d}x \leq \exp[-\log(1-1/(2d))(d/2)]\det(\Sigma_T)^{1/2}.$$

Since for any $t \in [0,1/2)$, $-\log(1-t) \leq \log(2)t$ we get that

$$\int_{\mathbb{R}^d} p_\infty(x)^2/p_T(x)\mathrm{d}x \leq 2^{1/4}\det(\Sigma_T)^{1/2}. \tag{S67}$$

Finally, using that $\Sigma_T = \exp[-2T]\Sigma + (1-\exp[-2T])\,\mathrm{Id}$ we have that

$$\det(\Sigma_T)^{1/2} \leq (\exp[-2T]\|\Sigma\| + 1 - \exp[-2T])^{d/2} \leq (1+\exp[-2T]\|\Sigma\|)^{d/2}.$$

Hence, using that result and that for any $t \geq 0$, $\log(1+t) \leq t$ we have

$$\det(\Sigma_T)^{1/2} \leq \exp[\exp[-2T]\|\Sigma\|(d/2)]\ .$$

Since, $T \geq (1/2)\{\log(\|\Sigma\|) + \log(d) + \log(2) - \log(\log(2^{1/4}))\}$, we get that $\det(\Sigma_T)^{1/2} \leq 2$, which concludes the proof upon combining this result and (S67). □

Therefore, we get the following simplified result in the Gaussian setting.

Corollary S27. *Assume that $p = \mathrm{N}(0,\Sigma)$, with $\|\Sigma^{-1}\| \geq 1$, $\gamma^\star = \gamma_\star = \gamma > 0$ and $T \geq 1 + (1/2)[\log^+(\|\Sigma\|) + \log(d+1)]$, then we have*

$$\begin{aligned}\|\mathcal{L}(X_0) - p_0\|_{\mathrm{TV}} \leq \exp[-T/2](\log^+(\|\Sigma^{-1}\|) + \|\Sigma - \mathrm{Id}\,\|)^{1/2}\\ +48(1+\|\Sigma\|^{1/2}d^{1/2})\|\Sigma^{-1}\|^2[1+\|\Sigma - \mathrm{Id}\,\|]\sqrt{\gamma}.\end{aligned}$$

Proof. Using (S14) and Proposition S26 we have

$$\begin{aligned}\|\mathcal{L}(X_0) - p_0\|_{\mathrm{TV}} &\leq \exp[-T/2](-\log(\det(\Sigma)) + \mathrm{Tr}(\Sigma) - d)^{1/2}\\
&\quad +12(1+(\textstyle\int_{\mathbb{R}^d}\|x\|^4\mathrm{d}p_0(x))^{1/4})[1+K+2C]\sqrt{(\gamma^\star)^2/\gamma_\star}\\
&\leq \exp[-T/2](-\log(\det(\Sigma)) + \mathrm{Tr}(\Sigma) - d)^{1/2}\\
&\quad +12(1+(\textstyle\int_{\mathbb{R}^d}\|x\|^4\mathrm{d}p_0(x))^{1/4})[1+\|\Sigma^{-1}\| + 2\|\Sigma^{-1}\|^2\|\Sigma - \mathrm{Id}\,\|]\sqrt{(\gamma^\star)^2/\gamma_\star}\\
&\leq \exp[-T/2](-\log(\det(\Sigma)) + \mathrm{Tr}(\Sigma) - d)^{1/2}\\
&\quad +12(1+3^{1/4}\|\Sigma\|^{1/2}d^{1/2})[1+\|\Sigma^{-1}\| + 2\|\Sigma^{-1}\|^2\|\Sigma - \mathrm{Id}\,\|]\sqrt{(\gamma^\star)^2/\gamma_\star}\\
&\leq \exp[-T/2](-\log(\det(\Sigma)) + \mathrm{Tr}(\Sigma) - d)^{1/2}\\
&\quad +48(1+\|\Sigma\|^{1/2}d^{1/2})\|\Sigma^{-1}\|^2[1+\|\Sigma - \mathrm{Id}\,\|]\sqrt{(\gamma^\star)^2/\gamma_\star}\\
&\leq \exp[-T/2](\log^+(\|\Sigma^{-1}\|) + \|\Sigma - \mathrm{Id}\,\|)^{1/2}\\
&\quad +48(1+\|\Sigma\|^{1/2}d^{1/2})\|\Sigma^{-1}\|^2[1+\|\Sigma - \mathrm{Id}\,\|]\sqrt{(\gamma^\star)^2/\gamma_\star}.\end{aligned}$$

□

S6 Proof of Theorem 3

Proof. For any x and j, denote $\bar{p}_{j,0}(\cdot|x)$ the distribution of $\bar{x}_{j,0}$ given $x_j = x$ and $p_{j,0}$ the distribution of $\tilde{x}_{j,0}$. For any j we have

$$\mathrm{KL}(p_j \| p_{j,0}) = \mathrm{KL}(p_{j+1} \| p_{j+1,0}) + \mathbb{E}[\mathrm{KL}(\bar{p}_{j+1}(\cdot|x_{j+1}) \| \bar{p}_{j+1,0}(\cdot|x_{j+1}))].$$

By recursion, we have that

$$\mathrm{KL}(p \| p_0) = \mathrm{KL}(p_J \| p_{j,0}) + \textstyle\sum_{j=1}^{J} \mathbb{E}[\mathrm{KL}(\bar{p}_j(\cdot|x_j) \| \bar{p}_{j,0}(\cdot|x_j))].$$

Combining Proposition S9 and Lemma S4, we get that

$$\mathrm{KL}(p \| p_0) \leq (\delta + \exp[-4T])(2^{-J}L)^n + \textstyle\sum_{j=1}^{J} (\delta + \exp[-4T])(2^{-j}L)^n(2^n - 1) + E_{T,\delta}.$$

Therefore, $\mathrm{KL}(p \| p_0) \leq (\delta + \exp[-4T])L^n + E_{T,\delta}$, which concludes the proof. □

S7 Experimental Details on Gaussian Experiments

We now give some details on the experiments in Section 3.2 (Figure 2). We use the exact formulas for the Stein score of p_t in this case: if $x_0 \sim \mathcal{N}(M, \Sigma)$, then $x_t \sim N(M_t, \Sigma_t)$ with $M_t = e^{-t}M$ and

$$\Sigma_t = \mathrm{e}^{-2t}\Sigma + (1 - \mathrm{e}^{-2t})\,\mathrm{Id}\,.$$

Under an ideal situation where there is no score error, the discretization of the (backward) generative process is given by equation (S23):

$$x_{k+1} = ((1+\delta)\,\mathrm{Id} - 2\delta\Sigma_{T-k\delta}^{-1})x_k + 2\delta\Sigma_{T-k\delta}^{-1}M_{T-k\delta} + \sqrt{2\delta}z_{k+1}\,,$$

where δ is the uniform step size and z_k are iid white Gaussian random variables. For the SGM case, $M = 0$. The starting step of this discretization is itself $x_0 \sim \mathcal{N}(0, \mathrm{Id})$. From this formula, the covariance matrix $\hat{\Sigma}_k$ of x_k satisfies the recursion (S16):

$$\hat{\Sigma}_{k+1} = ((1+\delta)\,\mathrm{Id} - 2\delta\Sigma_{T-k\delta}^{-1})\hat{\Sigma}_k((1+\delta)\,\mathrm{Id} - 2\delta\Sigma_{T-k\delta}^{-1}) + 2\delta\,\mathrm{Id}\,,$$

from which we can exactly compute $\hat{\Sigma}_k$ for very k, and especially for $k = N = T/\delta$, as a function of Σ, the final time T, and the step size δ. In all our experiments, we choose stationary processes: their covariance Σ is diagonal in a Fourier basis, with eigenvalues (*power spectrum*) noted $\hat{P}_k$. All the x_k remain stationary so $\hat{\Sigma}_k$ is still diagonal in a Fourier basis, with power spectrum noted $\hat{P}_k$. The error displayed in the left panel of figure 2 is:

$$\|\hat{P}_N - P\| = \max_{\omega} |\hat{P}_N(\omega) - P(\omega)| / \max_{\omega} |P(\omega)|,$$

normalized by the operator norm of Σ.

The illustration in the middle panel of Figure 2, for WSGM, is done for simplicity only at one scale (ie, at $j = 1$ in Algorithm 1): instead of stacking the full cascade of conditional distributions for all $j = J, \ldots, 1$, we use the true low-frequencies $x_{j,0} = x_1$. Here, we use Daubechies-4 wavelets. We sample $\bar{x}_{j,0}$ using the Euler-Maruyama recursion (S23)-(S16) for the conditional distribution. We recall that in the Gaussian case, $\bar{x}_1$ and x_1 are jointly Gaussian. The conditional distribution of $\bar{x}_1$ given x_1 is known to be $\mathcal{N}(Ax_1, \Gamma)$, where:

$$A = -\mathrm{Cov}(\bar{x}_1, x_1)\mathrm{Var}(x_1)^{-1}, \quad \Gamma = \mathrm{Var}(\bar{x}_1) - \mathrm{Cov}(\bar{x}_1, x_1)\mathrm{Var}(x_1)^{-1}\mathrm{Cov}(\bar{x}_1, x_1)^\top.$$

We solve the recursion (S16) with a step size δ and $N = T/\delta$ steps; the sampled conditional wavelet coefficients $\bar{x}_{j,0}$ have conditional distribution noted $\mathcal{N}(\hat{A}_N x, \hat{\Gamma}_N)$. The full covariance of $(\tilde{x}_{j,0}, \bar{x}_{j,0})$, written in the basis given by the high/low frequencies, is now given by

$$\hat{\Sigma}_N = \begin{bmatrix} \hat{\Gamma}_N & \mathrm{Cov}(x_1, \bar{x}_1)\hat{A}_N^\top \\ \hat{A}_N \mathrm{Cov}(x_1, \bar{x}_1)^\top & \mathrm{Cov}(x_1, x_1) \end{bmatrix}.$$

Figure 2, middle panel compares the eigenvalues (power spectrum) of these covariances, as a function of δ, with the ones of Σ.

The right panel of 2 gives the smallest N needed to reach $\|\hat{P}_N - P\| = 0.1$ in both cases (SGM and WSGM), based on a power law extrapolation of the curves $N \mapsto \hat{P}_N$.

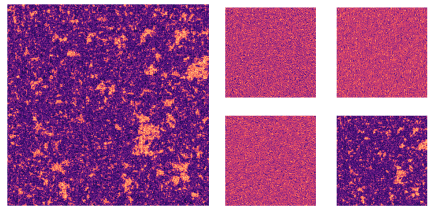

Figure S1: Example of a realization of a φ^4 critical field ($L = 256$) with its wavelet decomposition on the left (lower-frequencies are on bottom right panel).

S8 Experimental Details on the φ^4 Model

In this section, we develop and and make more precise the results in Section 4.1.

S8.1 The Critical φ^4 Process and its Stein Score Regularity

The macroscopic energy of non-Gaussian distributions can be specified as shown in (20), where K is a coupling matrix and V is non-quadratic potential. The φ^4-model over the $L \times L$ periodic grid is the special case defined by $C = -\Delta$ (the negative two-dimensional discrete Laplacian) and V is a quartic potential:

$$E(x) = \tfrac{\beta}{2} \textstyle\sum_{|u-v|=1} (x(u) - x(v))^2 + \sum_u (x(u)^2 - 1)^2.$$

Here, β is a parameter proportional to an inverse temperature.

In physics, the φ^4 model is a typical example of second-order phase transitions: the quadratic part reduces spatial fluctuations, and V favors configurations whose entries remain close to ± 1 (in physics, this is often called a *double-well potential*). In the thermodynamic limit $L \to \infty$, both term compete according to the value of β.

- For $\beta \ll 1$, the quadratic term becomes negligible and the marginals of the field become independent; this is the disordered state.
- For $\beta \gg 1$, the quadratic term favors configuration which are spatially smooth and the potential term drives the values towards ± 1, resulting in an ordered state, where all values of the field are simultaneously close to either $+1$ or to -1.

A phase transition occurs between these two regimes at a critical temperature $\beta_c \sim 0.68$ [36, 20]. At this point, the φ^4 field display very long-range correlations and an ill-conditioned Hessian $\nabla^2 \log p$. The sampling of φ^4 at this critical point becomes very difficult. This "critical slowing down" phenomenon is why, from a machine learning point of view, the critical φ^4 field is an excellent example of hard-to-learn and hard-to-sample distribution, yet still accessible for mathematical analysis.

Our wavelet diffusion considers the sampling of the conditional probability $p(\bar{x}_1|x_1)$ instead of $p(x_0)$, by inverting the noise diffusion projected on the wavelet coefficients. Theorem 2 indicates that the loss obtained with any SGM-type method depends on the regularity parameters of $\nabla \log p_t$ in (10).

Strictly speaking, to get a bound on K we should control the norm of $\nabla^2 \log p_t$ over all x and t. However, a look at the proof of the theorem indicates that this control does not have to be uniform in x; for instance, there is no need to control this Hessian in domains which have an extremely small probability under p_t. Moreover, since p_t is a convolution between p_0 and a Gaussian, we expect that a control over $\nabla^2 \log p_0(x)$ will actually be sufficient to control $\nabla^2 \log p_t(x)$ for all $t > 0$; these facts are non-rigorous for the moment. The distribution of some spectral statistics of $\nabla^2 \log p_0$ over samples drawn from the φ^4-model are shown in Figure S2 (blue).

Considering conditional probabilities $\bar{p}$ instead of p acts on the Hessian of the φ^4-energy as a projection over the wavelet field: in the general context of (20),

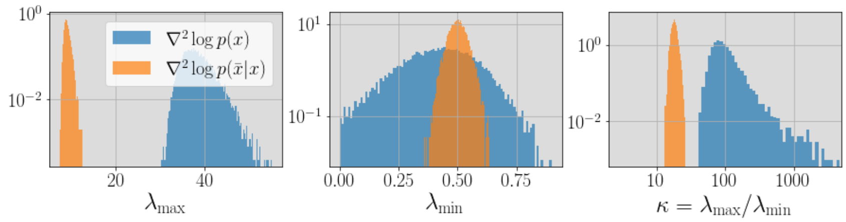

Figure S2: Histograms of 10^5 realizations of $\lambda_{\min}, \lambda_{\max}$ and $\kappa = \lambda_{\max}/\lambda_{\min}$ of the Hessian matrices in (S68) for the critical φ^4 model in dimension $L = 32$. The mean values of κ are respectively $\mu = 18.32$ and $\bar{\mu} = 210.53$; standard deviations are $\sigma = 1.78$ and $\bar{\sigma} = 9451.37$.

$$-\nabla_x^2 \log p(x_0) = K + \nabla^2 V(x_0), \qquad -\nabla_{\bar{x}_1}^2 \log p(\bar{x}_1|x_1) = \gamma^2 \bar{G}(K + \nabla^2 V(x_0))\bar{G}^\top. \quad \text{(S68)}$$

The proof is in Appendix S8. The distribution of the conditioning number of $\nabla_x^2 \log p$ and $\nabla_{\bar{x}}^2 \log p$ over samples drawn from the φ^4 model is shown at Figure S2: the Hessian of the wavelet log-probability is orders-of-magnitude better conditioned than its single-scale counterpart, with a very concentrated distribution. The same phenomenon occurs at each scale j, and the same is true for $\lambda_{\min}, \lambda_{\max}$. It turns out that considering wavelet coefficient not only concentrates these eigenvalues, but also drives $\lambda_{\min}$ away from 0. In the context of multiscale Gaussian processes, Theorem S4 gives a rigorous proof of this phenomenon. In the general case, $\nabla^2 \log p_t$ is not reduced to the inverse of a covariance matrix, but we expect the same phenomenon to be true.

S8.2 Score Models and Details on our Numerical Experiments of φ^4

In this section, we give some details on our numerical experiments from Section 4.1.

Training Data and Wavelets

We used samples from the φ^4 model generated using a classical MCMC algorithm — the sampling script will be publicly available in our repository.

The wavelet decompositions of our fields were performed using Python's `pywavelets` package and `Pytorch Wavelets` package. For synthetic experiments, we used the Daubechies wavelets with $p = 4$ vanishing moments (see [31, Section 7.2.3]).

Score Model

At the first scale $j = 0$, the distribution of the φ^4 model falls into the general form given in (20), and it is assumed that at each scale j, the distribution of the field at scale j still assumes this shape — with modified constants and coupling parameters. The score model we use at each scale is given by:

$$s_{K,\theta}(x) = \tfrac{1}{2} x^\top K x + \textstyle\sum_u (\theta_1 v_1(x(u)) + \cdots + \theta_m v_m(x(u))),$$

where the parameters are $K, \theta_1, \ldots, \theta_m$ and v_i are a family of smooth functions. One can also represent this score as $s_{K,\theta} = K \cdot xx^\top + \theta^\top U(x)$ where $U_i(x) = \sum_u v_i(x(u))$.

Learning

We trained our various algorithms using SGM or WSGM up to a time $T = 5$ with $n_{\text{train}} = 2000$ steps of forward diffusion. At each step t, the parameters were learned by minimizing the score loss:

$$\ell(K, \theta) = \mathbb{E}[|\nabla s_{K,\theta}(x_t)|^2 + 2\Delta_x s_{K,\theta}(x_t)]$$

using the Adam optimiser with learning rate $\texttt{lr} = 0.01$ and default parameters α, β. At the start of the diffusion ($t = 0$) we use 10000 steps of gradient descent. For $t > 1$, we use only 100 steps of gradient descent, but initialized at (K_{t-1}, θ_{t-1}).

Sampling

For the sampling, we used uniform steps of discretization.

For the error metric, we first measure the L^2-norm between the power spectra $P, \hat{P}$ of the true φ^4 samples and our synthesized examples; more precisely, we set:

$$D_1 = \|P - \hat{P}\|^2.$$

This error on second-order statistics is perfectly suitable for Gaussian processes, but must be refined for non-Gaussian processes. We also consider the total variation distance between the histograms of the marginal distributions (in the case of two-dimensions, pixel-wise histograms). We note this error D_2; our final error measure is $D_1 + D_2$. This is the error used in Figure S2.

S8.3 Proofs of (S68)

In the sequel, ∇f is the gradient of a function f, and ∇^2 is the Hessian matrix of f. The *Laplacian* of f is the trace of the Hessian.

Lemma S28. *Let $U : \mathbb{R}^n \to \mathbb{R}$ be smooth and M be a $n \times m$ matrix. We set $F(x) = U(Mx)$ where $x \in \mathbb{R}^m$. Then, $\nabla^2 F(x) = M^\top \nabla^2 U(x) M$.*

Proof. Let $U : \mathbb{R}^n \to \mathbb{R}$ be smooth and M be a $n \times m$ matrix. Then,

$$\partial_k F(x) = \sum_{i=1}^{n} M_{i,k} (\partial_i U)(Mx).$$

Similarly,

$$\partial_{k,\ell} F(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} M_{i,k} M_{j,\ell} \partial_j (\partial_i U)(Mx). \tag{S69}$$

This is equal to $(M^\top \nabla^2 U M)_{k,\ell}$. □

Lemma S29. *Under the setting of the preceding lemma, if $U(x) = \sum_{i=1}^{n} f(x_i)$, then (i) $\nabla^2 U(x) = \mathrm{diag}(u''(x_1), \dots, u''(x_n))$ and (ii) the Laplacian of $F(x) = U(Mx)$ is given by*

$$\Delta F(x) = \sum_{i=1}^{n} (M^\top M)_{i,i} u''(x_i).$$

Proof. The proof of (i) comes from the fact that $\partial_i U(x) = u'(x_i)$, hence $\partial_j \partial_i U(x) = u''(x_i)$ if $i = j$, zero otherwise. The proof of (ii) consists in summing the $k = \ell$ terms in (S69) and using (i). □

For simplicity, let us note $p(x) = \mathrm{e}^{-H(x)}/Z$ where Z_0 is a normalization constant and $H(x) = x^\top K x/2 + V(x)$. Then,

$$\nabla_x p(x) = -\nabla_x H(x), \qquad \nabla_x^2 p(x) = -\nabla_x^2 H(x),$$

and the formula in the left of (S68) comes from the fact that the Hessian of $x^\top K x$ is $2K$.

For the second term, let us first recall that if $\bar{x}_1$ and x_1 are the wavelet coefficients and low-frequencies of x, they are linked by (18). Consequently, the joint density of $(\bar{x}_1, x_1)$ is:

$$q(\bar{x}_1, x_1) = \mathrm{e}^{-H(\gamma G^\top x_1 + \gamma \bar{G}^\top \bar{x}_1)}/Z_1$$

where Z_1 is another normalization constant. The conditional distribution of $\bar{x}_1$ given x_1 is:

$$q(\bar{x}_1 | x_1) = \frac{q(\bar{x}_1, x_1)}{Z_1(x_1)}$$

where $Z_1(x) = \int q(\bar{x}_1, u)\mathrm{d}u$. Consequently,

$$\begin{aligned}\nabla_{\bar{x}_1} \log q(\bar{x}_1|x_1) &= \nabla_{\bar{x}_1}(-H(\gamma G^\top x_1 + \gamma \bar{G}^\top \bar{x}_1) - \log Z_1) - \nabla_{\bar{x}_1} Z_1(x_1) \\ &= -\nabla_{\bar{x}_1} H(\gamma G^\top x_1 + \gamma \bar{G}^\top \bar{x}_1)\end{aligned}$$

and additionally:

$$\nabla^2_{\bar{x}_1} q(\bar{x}_1|x_1) = -\nabla^2_{\bar{x}_1} H(\gamma G^\top x_1 + \gamma \bar{G}^\top \bar{x}_1).$$

The RHS of (S68) then follows from the lemmas in the preceding section.

S9 Experimental Details on CelebA-HQ

Data We use Haar wavelets. The 128×128 original images are thus successively brought to the 64×64 and 32×32 resolutions, separately for each color channel. Each of the 3 channels of x_j and 9 channels of $\bar{x}_j$ are normalized to have zero mean and unit variance.

Architecture Following [38], both the conditional and unconditional scores are parameterized by a neural network with a U-Net architecture. It has 3 residual blocks at each scale, with a base number of channels of $C = 128$. The number of channels at the k-th scale is $a_k C$, where the multipliers $(a_k)_k$ depend on the resolution of the generated images. These multipliers are $(1, 2, 2, 4, 4)$ for models at the 128×128 resolution, $(2, 2, 4, 4)$ for models at the 64×64 resolution, $(4, 4)$ for the conditional model at the 32×32 resolution, and $(1, 2, 2, 2)$ for the unconditional model at the 32×32 resolution. All models include multi-head attention layers in blocks operating on images at resolutions 16×16 and 8×8. The conditioning on the low frequencies x_j is done with a simple input concatenation along channels, while conditioning on time is done through affine rescalings with learned time embeddings at each GroupNorm layer [38, 40].

Training The networks are trained with the (conditional) denoising score matching losses:

$$\begin{aligned}\ell(\theta_J) &= \mathbb{E}_{x_J,t,z}\left[||s_{\theta_J}(t, \mathrm{e}^{-t}x_J + \sqrt{1-\mathrm{e}^{-2t}}z) - \frac{z}{\sqrt{1-\mathrm{e}^{-2t}}}||^2\right] \\ \ell(\bar{\theta}_j) &= \mathbb{E}_{\bar{x}_j,x_j,t,z}\left[||\bar{s}_{\bar{\theta}_j}(t, \mathrm{e}^{-t}\bar{x}_j + \sqrt{1-\mathrm{e}^{-2t}}z \,|\, x_j) - \frac{z}{\sqrt{1-\mathrm{e}^{-2t}}}||^2\right]\end{aligned}$$

where $z \sim \mathcal{N}(0, \mathrm{Id})$ and the time t is distributed as Tu^2 with $u \sim \mathcal{U}([0, 1])$. We fix the maximum time $T = 5$ for all scales. Networks are trained for 5×10^5 gradient steps with a batch size of 128 at the 32×32 resolution and 64 otherwise. We use the Adam [21] optimizer with a learning rate of 10^{-4} and no weight decay.

Sampling For sampling, we use model parameters from an exponential moving average with a rate of 0.9999. For each number of discretization steps N, we use the Euler-Maruyama discretization with a uniform step size $\delta_k = T/N$ starting from $T = 5$. This discretization scheme is used at all scales. For FID computations, we generate $30,000$ samples in each setting.