

Non-asymptotic bounds for forward processes in denoising diffusions: Ornstein-Uhlenbeck is hard to beat

Miha Brešar and Aleksandar Mijatović

Department of Statistics



17 August 2024

Part II: General result and its proof

EPSRC

Engineering and Physical Sciences
Research Council

EP/W006227/1 & EP/V009478/1

Problem

$dX_t = -X_t dt + \sqrt{2} dB_t$, with std. normal invariant measure π_X and $X_0 = \rho_0$.

$dY_t = b(Y_t)dt + \sigma(Y_t)dB_t$, with invariant measure π_Y and $Y_0 = \rho_0$.

Problem

$dX_t = -X_t dt + \sqrt{2} dB_t$, with std. normal invariant measure π_X and $X_0 = \rho_0$.

$dY_t = b(Y_t)dt + \sigma(Y_t)dB_t$, with invariant measure π_Y and $Y_0 = \rho_0$.

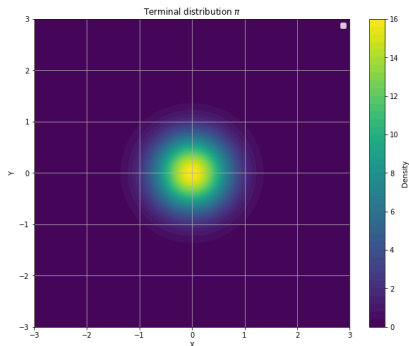
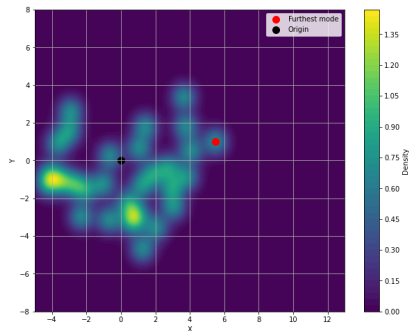
Compare $\|\mathbb{P}_{\rho_0}(X_t \in \cdot) - \pi_X\|_{TV}$ and $\|\mathbb{P}_{\rho_0}(Y_t \in \cdot) - \pi_Y\|_{TV}$.

Problem

$d\mathbf{X}_t = -\mathbf{X}_t dt + \sqrt{2} dB_t$, with std. normal invariant measure π_X and $\mathbf{X}_0 = \rho_0$.

$d\mathbf{Y}_t = b(\mathbf{Y}_t)dt + \sigma(\mathbf{Y}_t)dB_t$, with invariant measure π_Y and $\mathbf{Y}_0 = \rho_0$.

Compare $\|\mathbb{P}_{\rho_0}(X_t \in \cdot) - \pi_X\|_{TV}$ and $\|\mathbb{P}_{\rho_0}(Y_t \in \cdot) - \pi_Y\|_{TV}$.

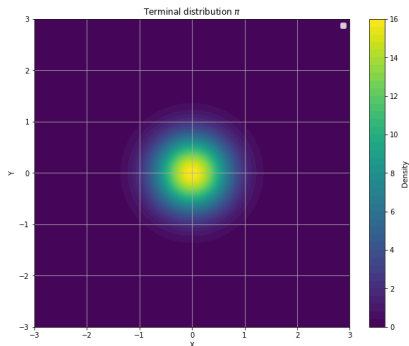
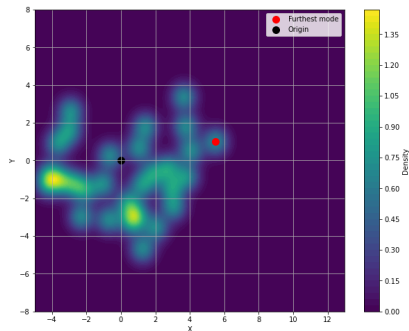


Problem

$d\mathbf{X}_t = -\mathbf{X}_t dt + \sqrt{2} dB_t$, with std. normal invariant measure π_X and $\mathbf{X}_0 = \rho_0$.

$d\mathbf{Y}_t = b(\mathbf{Y}_t)dt + \sigma(\mathbf{Y}_t)dB_t$, with invariant measure π_Y and $\mathbf{Y}_0 = \rho_0$.

Compare $\|\mathbb{P}_{\rho_0}(\mathbf{X}_t \in \cdot) - \pi_X\|_{TV}$ and $\|\mathbb{P}_{\rho_0}(\mathbf{Y}_t \in \cdot) - \pi_Y\|_{TV}$.



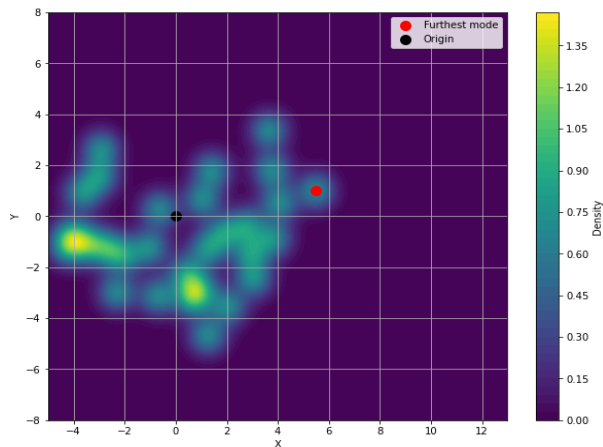
See part I for the context and connection to DDPMs.

Assumption (DATA): multi-modality of initial data law

For $\rho_0 \in \mathcal{M}_{1,R,\varepsilon}(\mathbb{R}^d)$, there exists $x_0 \in \mathbb{R}^d$ with $|x_0| = R(1 + \delta)$, such that $b_\rho = \rho_0(B(x_0, R\delta))$ and $\rho_0(B(0, R(1 + 2\delta))^c) < \varepsilon/2$.

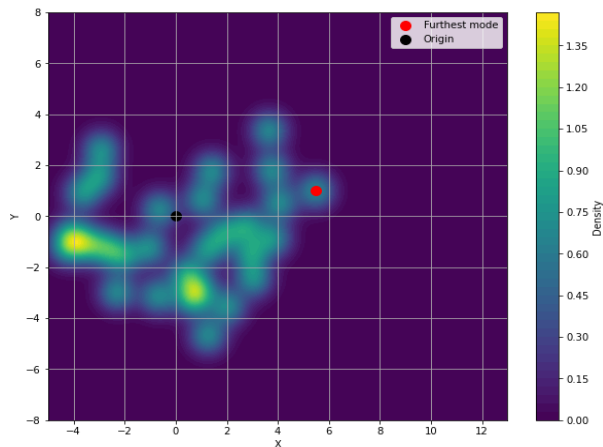
Assumption (DATA): multi-modality of initial data law

For $\rho_0 \in \mathcal{M}_{1,R,\varepsilon}(\mathbb{R}^d)$, there exists $x_0 \in \mathbb{R}^d$ with $|x_0| = R(1 + \delta)$, such that $b_\rho = \rho_0(B(x_0, R\delta))$ and $\rho_0(B(0, R(1 + 2\delta))^c) < \varepsilon/2$.



Assumption (DATA): multi-modality of initial data law

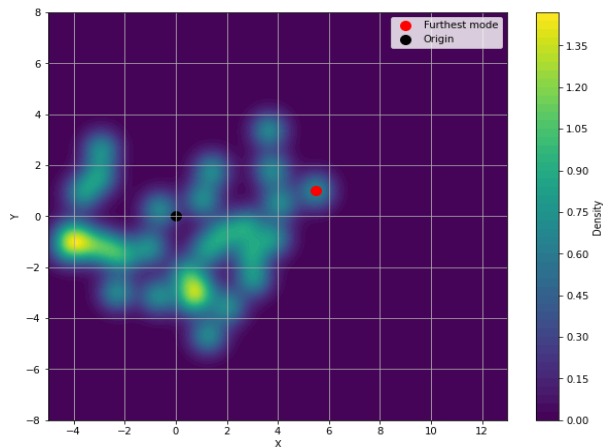
For $\rho_0 \in \mathcal{M}_{1,R,\varepsilon}(\mathbb{R}^d)$, there exists $x_0 \in \mathbb{R}^d$ with $|x_0| = R(1 + \delta)$, such that $b_\rho = \rho_0(B(x_0, R\delta))$ and $\rho_0(B(0, R(1 + 2\delta))^c) < \varepsilon/2$.



Remark. In (DATA), **no** assumptions on **moments**, **log-concavity**, **existence of density** of ρ_0 or **KL-divergence** $\text{KL}(\rho_0 \parallel \text{Gauss}) < \infty$.

Assumption (DATA): multi-modality of initial data law

For $\rho_0 \in \mathcal{M}_{1,R,\varepsilon}(\mathbb{R}^d)$, there exists $x_0 \in \mathbb{R}^d$ with $|x_0| = R(1 + \delta)$, such that $b_\rho = \rho_0(B(x_0, R\delta))$ and $\rho_0(B(0, R(1 + 2\delta))^c) < \varepsilon/2$.



Remark. In (DATA), **no** assumptions on **moments**, **log-concavity**, **existence of density** of ρ_0 or **KL-divergence** $\text{KL}(\rho_0 \parallel \text{Gauss}) < \infty$. In particular, (DATA) allows for the **Manifold Hypothesis** [FMN16].

Assumption (ForProc): Forward process dynamics

Consider a $d \geq 3$ dimensional ergodic SDE with invariant measure π

$$dY_t = b(Y_t)dt + \sigma(Y_t)dB_t.$$

Assumption (ForProc): Forward process dynamics

Consider a $d \geq 3$ dimensional ergodic SDE with invariant measure π

$$dY_t = b(Y_t)dt + \sigma(Y_t)dB_t.$$

(I) the drift b exhibits at most linear growth in each direction:

$$|\langle b(x), u/|u| \rangle| \leq |\langle x, u/|u| \rangle| \quad \text{for all } x \in \mathbb{R}^d \text{ and } u \in \mathbb{R}^d \setminus \{0\}.$$

Assumption (ForProc): Forward process dynamics

Consider a $d \geq 3$ dimensional ergodic SDE with invariant measure π

$$dY_t = b(Y_t)dt + \sigma(Y_t)dB_t.$$

(I) the drift b exhibits at most linear growth in each direction:

$$|\langle b(x), u/|u| \rangle| \leq |\langle x, u/|u| \rangle| \quad \text{for all } x \in \mathbb{R}^d \text{ and } u \in \mathbb{R}^d \setminus \{0\}.$$

(II) for some $k \in \{3, \dots, d\}$ and any $y_1 \in \mathbb{R}^d$ with $|y_1| = 1$, there exist ON base $\{y_1, \dots, y_k\}$ of $\mathcal{Y}$ s.t. ON projection $G_{\mathcal{Y}} : \mathbb{R}^d \rightarrow \mathcal{Y}$ and dispersion $a = \sigma\sigma^T$ satisfy

$$\sum_{j=1}^k \langle ay_j, y_j \rangle \geq 3 \langle a\hat{G}_{\mathcal{Y}}, \hat{G}_{\mathcal{Y}} \rangle \quad \text{on } \mathbb{R}^d, \quad \hat{G}_{\mathcal{Y}} = (1 + |G_{\mathcal{Y}}|^2)^{-1/2} G_{\mathcal{Y}}. \quad (1)$$

Assumption (ForProc): Forward process dynamics

Consider a $d \geq 3$ dimensional ergodic SDE with invariant measure π

$$dY_t = b(Y_t)dt + \sigma(Y_t)dB_t.$$

(I) the drift b exhibits at most linear growth in each direction:

$$|\langle b(x), u/|u| \rangle| \leq |\langle x, u/|u| \rangle| \quad \text{for all } x \in \mathbb{R}^d \text{ and } u \in \mathbb{R}^d \setminus \{0\}.$$

(II) for some $k \in \{3, \dots, d\}$ and any $y_1 \in \mathbb{R}^d$ with $|y_1| = 1$, there exist ON base $\{y_1, \dots, y_k\}$ of $\mathcal{Y}$ s.t. ON projection $G_{\mathcal{Y}} : \mathbb{R}^d \rightarrow \mathcal{Y}$ and dispersion $a = \sigma\sigma^T$ satisfy

$$\sum_{j=1}^k \langle ay_j, y_j \rangle \geq 3 \langle a\hat{G}_{\mathcal{Y}}, \hat{G}_{\mathcal{Y}} \rangle \quad \text{on } \mathbb{R}^d, \quad \hat{G}_{\mathcal{Y}} = (1 + |G_{\mathcal{Y}}|^2)^{-1/2} G_{\mathcal{Y}}. \quad (1)$$

Note that for $a(x) = \sigma^2(x)\text{Id}$ (1) holds iff $k \geq 3$.

Main result

Let π be the invariant measure of Y , $\varepsilon > 0$ and pick $r_k \in (1, \infty)$ s. t.

$$\pi(\{x \in \mathbb{R}^d : |G_Y(x)|^2 + 1 \leq r_k^2\}) > 1 - \varepsilon/2.$$

Main result

Let π be the invariant measure of Y , $\varepsilon > 0$ and pick $r_k \in (1, \infty)$ s. t.

$$\pi(\{x \in \mathbb{R}^d : |G_Y(x)|^2 + 1 \leq r_k^2\}) > 1 - \varepsilon/2.$$

Theorem

Let the data distribution $\rho_0 \in \mathcal{M}_{1,R,\varepsilon}(\mathbb{R}^d)$ satisfy **(DATA)** and the diffusion Y satisfy **(ForProc)**, respectively, for some $0 < \varepsilon, \delta, \beta \ll 1$ and $R, r_k \in (0, \infty)$ with $2r_k \leq R$. Then we have

$$\|\mathbb{P}_{\rho_0}(Y_{T_c} \in \cdot) - \pi\|_{\text{TV}} > (b_\rho - \varepsilon)/2 \gg \varepsilon, \quad \text{where } T_c = \log\left(\frac{R}{2r_k}\right).$$

Main result

Let π be the invariant measure of Y , $\varepsilon > 0$ and pick $r_k \in (1, \infty)$ s. t.

$$\pi(\{x \in \mathbb{R}^d : |G_Y(x)|^2 + 1 \leq r_k^2\}) > 1 - \varepsilon/2.$$

Theorem

Let the data distribution $\rho_0 \in \mathcal{M}_{1,R,\varepsilon}(\mathbb{R}^d)$ satisfy **(DATA)** and the diffusion Y satisfy **(ForProc)**, respectively, for some $0 < \varepsilon, \delta, \beta \ll 1$ and $R, r_k \in (0, \infty)$ with $2r_k \leq R$. Then we have

$$\|\mathbb{P}_{\rho_0}(Y_{T_c} \in \cdot) - \pi\|_{\text{TV}} > (b_\rho - \varepsilon)/2 \gg \varepsilon, \quad \text{where } T_c = \log\left(\frac{R}{2r_k}\right).$$

The OU process X satisfies:

$$\|\mathbb{P}_{\rho_0}(X_{T_{\text{OU}}} \in \cdot) - \pi_X\|_{\text{TV}} < \varepsilon,$$

where $T_{\text{OU}} = \max\{\log(2d^{1/4}/\varepsilon^{1/2}), \log(2R(1+2\delta)/\varepsilon)\}$.

Assumption $(\text{DATA}) \leftrightarrow (\text{ForProc})$

Let (DATA) and (ForProc) hold with dimension $\mathbf{d}$, distance to the furthest mode $\mathbf{R}$, its radius δ and error tolerance ϵ .

Assumption $(\text{DATA}) \leftrightarrow (\text{ForProc})$

Let (DATA) and (ForProc) hold with dimension $\mathbf{d}$, distance to the furthest mode $\mathbf{R}$, its radius δ and error tolerance ε .

(i) $R \geq \varepsilon^{1/2} d^{1/4}$.

Assumption $(\text{DATA}) \leftrightarrow (\text{ForProc})$

Let (DATA) and (ForProc) hold with dimension $\mathbf{d}$, distance to the furthest mode $\mathbf{R}$, its radius δ and error tolerance ε .

$$(i) \ R \geq \varepsilon^{1/2} d^{1/4}.$$

Assume the following holds for $0 < \beta \ll 1$ small.

$$(ii) \ R^\beta \geq 2(1 + 2\delta)/\varepsilon.$$

Assumption (DATA) $\leftrightarrow$ (ForProc)

Let (DATA) and (ForProc) hold with dimension $\mathbf{d}$, distance to the furthest mode $\mathbf{R}$, its radius δ and error tolerance ε .

(i) $R \geq \varepsilon^{1/2} d^{1/4}$.

Assume the following holds for $0 < \beta \ll 1$ small.

(ii) $R^\beta \geq 2(1 + 2\delta)/\varepsilon$.

(iii) the $\varepsilon/2$ -quantile r_k of the k -dimensional projection onto π :
 $\pi(\{x \in \mathbb{R}^d : |G_Y(x)|^2 + 1 \leq r_k^2\}) > 1 - \varepsilon/2$ satisfies $2r_k \leq R^\beta$.

Main result in the context of DDPMs

Corollary

Let the data distribution $\rho_0 \in \mathcal{M}_{1,R,\varepsilon}(\mathbb{R}^d)$ and the diffusion Y satisfy $(DATA) \leftrightarrow (ForProc)$ with parameters $R, \mu \in (0, \infty)$ and $0 < \varepsilon, \delta, \beta \ll 1$. Then we have

$$\|\mathbb{P}_{\rho_0}(Y_{T_c} \in \cdot) - \pi\|_{TV} > (b_\rho - \varepsilon)/2 > \varepsilon, \quad \text{where } T_c = (1 - \beta) \log R.$$

Main result in the context of DDPMs

Corollary

Let the data distribution $\rho_0 \in \mathcal{M}_{1,R,\varepsilon}(\mathbb{R}^d)$ and the diffusion Y satisfy $(DATA) \leftrightarrow (ForProc)$ with parameters $R, \mu \in (0, \infty)$ and $0 < \varepsilon, \delta, \beta \ll 1$. Then we have

$$\|\mathbb{P}_{\rho_0}(Y_{T_c} \in \cdot) - \pi\|_{TV} > (b_\rho - \varepsilon)/2 > \varepsilon, \quad \text{where } T_c = (1 - \beta) \log R.$$

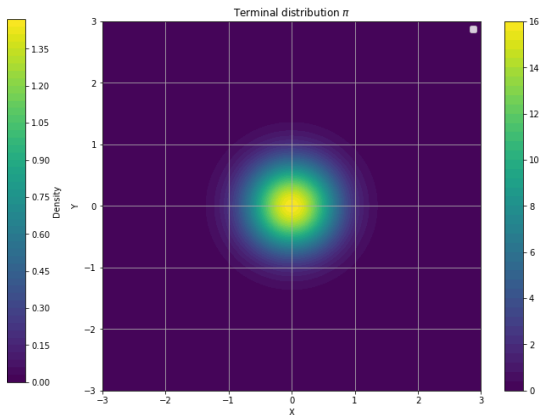
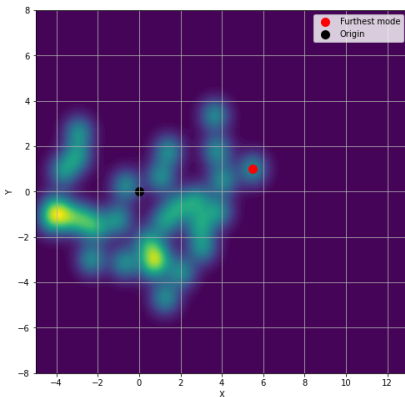
The OU process X satisfies

$$\|\mathbb{P}_{\rho_0}(X_{T_{OU}} \in \cdot) - \pi_X\|_{TV} < \varepsilon, \quad \text{where } T_{OU} = (1 + \beta) \log R.$$

Problem

$d\mathbf{X}_t = -\mathbf{X}_t dt + \sqrt{2} dB_t$, with std. normal invariant measure π_X and $\mathbf{X}_0 = \rho_0$

$d\mathbf{Y}_t = b(\mathbf{Y}_t)dt + \sigma(\mathbf{Y}_t)dB_t$, with invariant measure π_Y and $\mathbf{Y}_0 = \rho_0$



Problem

Minimal assumptions on ρ_0 and Y .

Problem

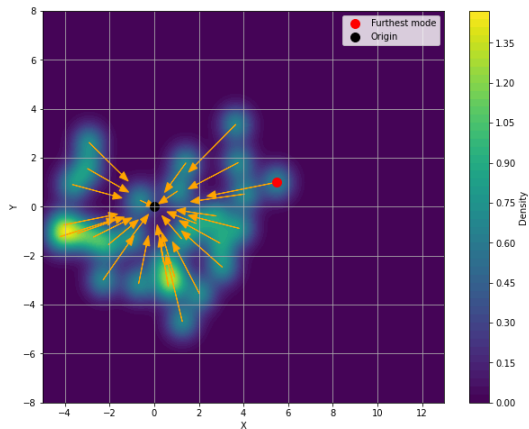
Minimal assumptions on ρ_0 and Y .

Non-asymptotic bounds on the convergence rate.

Problem

Minimal assumptions on ρ_0 and Y .

Non-asymptotic bounds on the convergence rate.



Lower bounds: key concepts

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \quad X_0 = x, \pi\text{-invariant, heavy tailed}$$

Hairer's idea [Hai09], a good **lower** bound

$$\|\mathbb{P}_x(X_T \in \cdot) - \pi\|_{\text{TV}} \geq \pi(\{|\cdot| \geq r\}) - \mathbb{P}_x(|X_T| \geq r).$$

Lower bounds: key concepts

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \quad X_0 = x, \pi\text{-invariant, heavy tailed}$$

Hairer's idea [Hai09], a good **lower** bound

$$\|\mathbb{P}_x(X_T \in \cdot) - \pi\|_{\text{TV}} \geq \pi(\{|\cdot| \geq r\}) - \mathbb{P}_x(|X_T| \geq r).$$

Reverse situation: **light tailed** π with **multi-modal** ρ_0

Lower bounds: key concepts

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \quad X_0 = x, \pi\text{-invariant, heavy tailed}$$

Hairer's idea [Hai09], a good **lower** bound

$$\|\mathbb{P}_x(X_T \in \cdot) - \pi\|_{\text{TV}} \geq \pi(\{|\cdot| \geq r\}) - \mathbb{P}_x(|X_T| \geq r).$$

Reverse situation: **light tailed** π with **multi-modal** ρ_0

$$\|\mathbb{P}_{\rho_0}(X_T \in \cdot) - \pi\|_{\text{TV}} \geq \pi(\{|\cdot| \leq r\}) - \mathbb{P}_{\rho_0}(|X_T| \leq r).$$

Lower bounds: Key lemma

Lemma

Let X be a Markov process with a generator $\mathcal{A}$ and an invariant measure π . Assume that $H : \mathbb{R}^d \rightarrow [0, 1]$ is such that $H \in \mathcal{D}(\mathcal{A})$. Assume $\mathcal{A}H \leq \xi \circ H$ on $\mathbb{R}^d$ with $\xi : [0, 1] \rightarrow (0, 1)$ increasing concave. Fix $\rho_0 \in \mathcal{M}_1(\mathbb{R}^n)$ and $T \geq 1$.

Lower bounds: Key lemma

Lemma

Let X be a Markov process with a generator $\mathcal{A}$ and an invariant measure π . Assume that $H : \mathbb{R}^d \rightarrow [0, 1]$ is such that $H \in \mathcal{D}(\mathcal{A})$.

Assume $\mathcal{A}H \leq \xi \circ H$ on $\mathbb{R}^d$ with $\xi : [0, 1] \rightarrow (0, 1)$ increasing concave. Fix $\rho_0 \in \mathcal{M}_1(\mathbb{R}^n)$ and $T \geq 1$.

Then, for every $r \geq 1$, it holds that

$$\begin{aligned} & \|\mathbb{P}_{\rho_0}(X_T \in \cdot) - \pi\|_{\text{TV}} \\ & \geq \pi(\{H \geq 1/r\}) - \rho_0(\{H \geq C_{r,T}\}) - \int_{\{H < C_{r,T}\}} r \Xi_{H(x)}(T)^{-1} \rho_0(dx), \end{aligned}$$

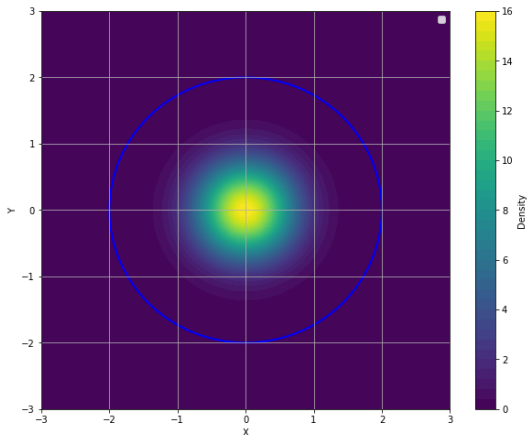
$$\text{where } \Xi_{H(x)}(T) = \int_{H(x)}^T \frac{1}{\xi(s)} ds$$

$$\begin{aligned}
& \|\mathbb{P}_{\rho_0}(X_T \in \cdot) - \pi\|_{\text{TV}} \\
& \geq \pi(\{H \geq 1/r\}) - \rho_0(\{H \geq C_{r,T}\}) - \int_{\{H < C_{r,T}\}} r \Xi_{H(x)}^{-1}(T) \rho_0(dx),
\end{aligned}$$

$$\|\mathbb{P}_{\rho_0}(X_T \in \cdot) - \pi\|_{\text{TV}}$$

$$\geq \pi(\{H \geq 1/r\}) - \rho_0(\{H \geq C_{r,T}\}) - \int_{\{H < C_{r,T}\}} r \Xi_{H(x)}^{-1}(T) \rho_0(dx),$$

Natural choice $H(x) = (1 + |x|)^{-1}$.



$$\begin{aligned} & \|\mathbb{P}_{\rho_0}(X_T \in \cdot) - \pi\|_{\text{TV}} \\ & \geq \pi(\{H \geq 1/r\}) - \rho_0(\{H \geq C_{r,T}\}) - \int_{\{H < C_{r,T}\}} r \Xi_{H(x)}^{-1}(T) \rho_0(dx), \end{aligned}$$

Natural choice $H(x) = (1 + |x|)^{-1}$.

Problem with a **high dimension** d : $\pi(\{H \geq 1/r\}) > 1 - \varepsilon$, if $r \approx \sqrt{d}$.

$$\begin{aligned} & \|\mathbb{P}_{\rho_0}(X_T \in \cdot) - \pi\|_{\text{TV}} \\ & \geq \pi(\{H \geq 1/r\}) - \rho_0(\{H \geq C_{r,T}\}) - \int_{\{H < C_{r,T}\}} r \Xi_{H(x)}^{-1}(T) \rho_0(dx), \end{aligned}$$

Natural choice $H(x) = (1 + |x|)^{-1}$.

Problem with a **high dimension** d : $\pi(\{H \geq 1/r\}) > 1 - \varepsilon$, if $r \approx \sqrt{d}$.

Solution: projection to **critical submanifold** $\mathcal{Y}$, related to **Manifold hypothesis** [De 22].

$$\begin{aligned} & \|\mathbb{P}_{\rho_0}(X_T \in \cdot) - \pi\|_{\text{TV}} \\ & \geq \pi(\{H \geq 1/r\}) - \rho_0(\{H \geq C_{r,T}\}) - \int_{\{H < C_{r,T}\}} r \Xi_{H(x)}^{-1}(T) \rho_0(dx), \end{aligned}$$

(DATA) For $\rho_0 \in \mathcal{M}_{1,R,\varepsilon}(\mathbb{R}^d)$, there exists $x_0 \in \mathbb{R}^d$ with $|x_0| = R(1 + \delta)$, such that $b_\rho = \rho_0(B(x_0, R\delta))$.

$$\begin{aligned} & \|\mathbb{P}_{\rho_0}(X_T \in \cdot) - \pi\|_{\text{TV}} \\ & \geq \pi(\{H \geq 1/r\}) - \rho_0(\{H \geq C_{r,T}\}) - \int_{\{H < C_{r,T}\}} r \Xi_{H(x)}^{-1}(T) \rho_0(dx), \end{aligned}$$

(DATA) For $\rho_0 \in \mathcal{M}_{1,R,\varepsilon}(\mathbb{R}^d)$, there exists $x_0 \in \mathbb{R}^d$ with $|x_0| = R(1 + \delta)$, such that $b_\rho = \rho_0(B(x_0, R\delta))$.

(ForProc) Let $y_1 = x_0/|x_0|$, $k \in \{3, \dots, d\}$, $y_1, \dots, y_k$ ON base of $\mathcal{Y}$ with ON base $y_1, \dots, y_k$ and $G_{\mathcal{Y}}$ projection of x onto $\mathcal{Y}$ such that

$$\sum_{j=1}^k \langle ay_j, y_j \rangle \geq 3 \langle aG_{\mathcal{Y}}/(1 + |G_{\mathcal{Y}}|), G_{\mathcal{Y}}/(1 + |G_{\mathcal{Y}}|) \rangle \quad \text{on } \mathbb{R}^d \quad (2)$$

$$\begin{aligned} & \|\mathbb{P}_{\rho_0}(X_T \in \cdot) - \pi\|_{\text{TV}} \\ & \geq \pi(\{H \geq 1/r\}) - \rho_0(\{H \geq C_{r,T}\}) - \int_{\{H < C_{r,T}\}} r \Xi_{H(x)}^{-1}(T) \rho_0(dx), \end{aligned}$$

(DATA) For $\rho_0 \in \mathcal{M}_{1,R,\varepsilon}(\mathbb{R}^d)$, there exists $x_0 \in \mathbb{R}^d$ with $|x_0| = R(1 + \delta)$, such that $b_\rho = \rho_0(B(x_0, R\delta))$.

(ForProc) Let $y_1 = x_0/|x_0|$, $k \in \{3, \dots, d\}$, $y_1, \dots, y_k$ ON base of $\mathcal{Y}$ with ON base $y_1, \dots, y_k$ and $G_{\mathcal{Y}}$ projection of x onto $\mathcal{Y}$ such that

$$\sum_{j=1}^k \langle ay_j, y_j \rangle \geq 3 \langle aG_{\mathcal{Y}}/(1 + |G_{\mathcal{Y}}|), G_{\mathcal{Y}}/(1 + |G_{\mathcal{Y}}|) \rangle \quad \text{on } \mathbb{R}^d \quad (2)$$

Our function

$$H = H_{\mathcal{Y}}(x) = (1 + |G_{\mathcal{Y}}(x)|^2)^{-1/2} \quad \text{satisfies } \mathcal{A}H_{\mathcal{Y}} \leq H_{\mathcal{Y}} \quad \text{on } R^d$$

Lower bound

$$\begin{aligned} & \|\mathbb{P}_{\rho_0}(X_T \in \cdot) - \pi\|_{\text{TV}} \\ & \geq \pi(\{H_{\mathcal{Y}} \geq 1/r\}) - \rho_0(\{H_{\mathcal{Y}} \geq r \exp(T)\}) - \int_{\{H_{\mathcal{Y}} < r \exp(T)\}} r \frac{\exp(T)}{H_{\mathcal{Y}}(x)} \rho_0(\mathrm{d}x), \end{aligned}$$

Lower bound

$$\begin{aligned} & \|\mathbb{P}_{\rho_0}(X_T \in \cdot) - \pi\|_{\text{TV}} \\ & \geq \pi(\{H_{\mathcal{Y}} \geq 1/r\}) - \rho_0(\{H_{\mathcal{Y}} \geq r \exp(T)\}) - \int_{\{H_{\mathcal{Y}} < r \exp(T)\}} r \frac{\exp(T)}{H_{\mathcal{Y}}(x)} \rho_0(dx), \end{aligned}$$

Note that for some r_k, ε , we have $\pi(\{H_{\mathcal{Y}} \geq 1/r_k\}) > 1 - \varepsilon/2$.

Lower bound

$$\begin{aligned} & \|\mathbb{P}_{\rho_0}(X_T \in \cdot) - \pi\|_{\text{TV}} \\ & \geq \pi(\{H_Y \geq 1/r\}) - \rho_0(\{H_Y \geq r \exp(T)\}) - \int_{\{H_Y < r \exp(T)\}} r \frac{\exp(T)}{H_Y(x)} \rho_0(dx), \end{aligned}$$

Note that for some r_k, ε , we have $\pi(\{H_Y \geq 1/r_k\}) > 1 - \varepsilon/2$.

(DATA) The furthest mode $|x_0| = R(1 + \delta)$ satisfies $b_\rho = \rho_0(B(x_0, \delta R))$.

$$\|\mathbb{P}_{\rho_0}(Y_{T_c} \in \cdot) - \pi\|_{\text{TV}} > (b_\rho - \varepsilon)/2 > \varepsilon, \quad \text{where } T_c = \log \left(\frac{R}{2r_k} \right).$$

Lower bound

$$\begin{aligned} & \|\mathbb{P}_{\rho_0}(X_T \in \cdot) - \pi\|_{\text{TV}} \\ & \geq \pi(\{H_Y \geq 1/r\}) - \rho_0(\{H_Y \geq r \exp(T)\}) - \int_{\{H_Y < r \exp(T)\}} r \frac{\exp(T)}{H_Y(x)} \rho_0(dx), \end{aligned}$$

Note that for some r_k, ε , we have $\pi(\{H_Y \geq 1/r_k\}) > 1 - \varepsilon/2$.

(DATA) The furthest mode $|x_0| = R(1 + \delta)$ satisfies $b_\rho = \rho_0(B(x_0, \delta R))$.

$$\|\mathbb{P}_{\rho_0}(Y_{T_c} \in \cdot) - \pi\|_{\text{TV}} > (b_\rho - \varepsilon)/2 > \varepsilon, \quad \text{where } T_c = \log \left(\frac{R}{2r_k} \right).$$

Upper Bounds: Explicit density of OU + Pinsker's inequality.

Concluding remarks

Our robust approach is **self-contained** and applicable in a more general setting.

Concluding remarks

Our robust approach is **self-contained** and applicable in a more general setting.

We provide **lower bounds** for the convergence of the **forward** process. The bounds for the **reverse** process and the full algorithm remain open.

Concluding remarks

Our robust approach is **self-contained** and applicable in a more general setting.

We provide **lower bounds** for the convergence of the **forward** process. The bounds for the **reverse** process and the full algorithm remain open.

Thank you!

- [De 22] Valentin De Bortoli. “Convergence of denoising diffusion models under the manifold hypothesis”. *Transactions on Machine Learning Research* (2022).
- [FMN16] Charles Fefferman, Sanjoy Mitter, and Hariharan Narayanan. “Testing the manifold hypothesis”. *J. Amer. Math. Soc.* 29.4 (2016), pp. 983–1049.
- [Hai09] Martin Hairer. “How hot can a heat bath get?” *Comm. Math. Phys.* 292.1 (2009), pp. 131–177.