

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

Miha Brešar and Aleksandar Mijatović

Department of Statistics



19 August 2024

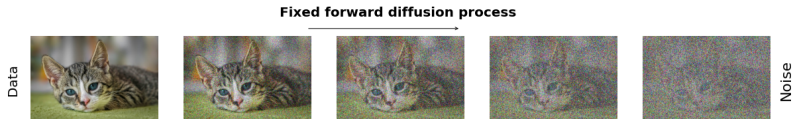
Part I: setting and key results

EPSRC

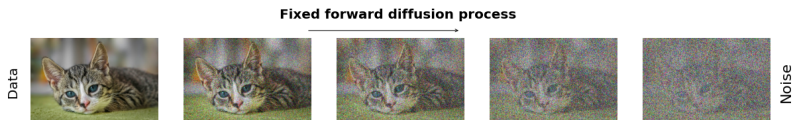
Engineering and Physical Sciences
Research Council

EP/W006227/1 & EP/V009478/1

Denoising Diffusion Probabilistic Models (DDPMs)



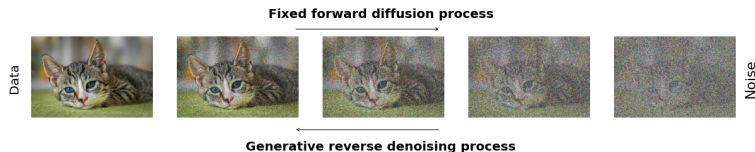
Denoising Diffusion Probabilistic Models (DDPMs)



Data distribution ρ_0 and forward SDE

$$dZ_t = b(Z_t)dt + \sigma(Z_t)dB_t, \quad \text{with } Z_0 \sim \rho_0.$$

Denoising Diffusion Probabilistic Models (DDPMs)



Data distribution ρ_0 and forward SDE (with stationary law π_Z)

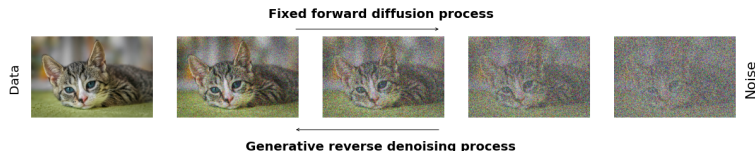
$$dZ_t = b(Z_t)dt + \sigma(Z_t)dB_t, \quad \text{with } Z_0 \sim \rho_0.$$

For $T > 0$, time-reversed process $(Z_{T-t})_{t \in [0, T]} \stackrel{d}{=} (\tilde{Z}_t)_{t \in [0, T]}$ satisfies SDE, started at $\tilde{Z}_0 \stackrel{d}{=} Z_T$,

$$d\tilde{Z}_t = -(b(\tilde{Z}_t) - \frac{1}{2}\nabla \cdot a(\tilde{Z}_t) - \frac{1}{2}a(\tilde{Z}_t)\nabla_x \log q_{T-t}(\tilde{Z}_t))dt + \sigma(\tilde{Z}_t)dB'_t,$$

where $a = \sigma\sigma^\top$ and q_t is the marginal density of Z_t [And82].

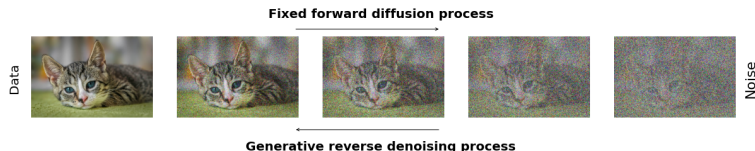
Denoising Diffusion Probabilistic Models (DDPMs)



SDE for $\overleftarrow{Z}$ yields an **approximate** simulation algorithm for the data distribution ρ_0 [Che+22; CDS23]:

$$dZ_t = b(Z_t)dt + \sigma(Z_t)dB_t, \quad \text{with } Z_0 \sim \rho_0.$$

Denoising Diffusion Probabilistic Models (DDPMs)



SDE for $\overleftarrow{Z}$ yields an **approximate** simulation algorithm for the data distribution ρ_0 [Che+22; CDS23]:

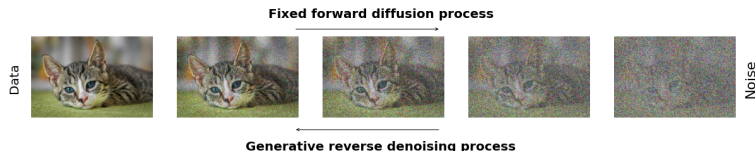
$$dZ_t = b(Z_t)dt + \sigma(Z_t)dB_t, \quad \text{with } Z_0 \sim \rho_0.$$

Since $\mathcal{L}(Z_T) \neq \pi_Z$, starting $\overleftarrow{Z}_0 \sim \pi_Z$ and evolving $\overleftarrow{Z}$ according to SDE

$$d\overleftarrow{Z}_t = -(b(\overleftarrow{Z}_t) - \frac{1}{2}\nabla \cdot a(\overleftarrow{Z}_t) - \frac{1}{2}a(\overleftarrow{Z}_t)\nabla_x \log q_{T-t}(\overleftarrow{Z}_t))dt + \sigma(\overleftarrow{Z}_t)dB'_t,$$

the total variation (TV) distance satisfies

Denoising Diffusion Probabilistic Models (DDPMs)



SDE for $\overleftarrow{Z}$ yields an **approximate** simulation algorithm for the data distribution ρ_0 [Che+22; CDS23]:

$$dZ_t = b(Z_t)dt + \sigma(Z_t)dB_t, \quad \text{with } Z_0 \sim \rho_0.$$

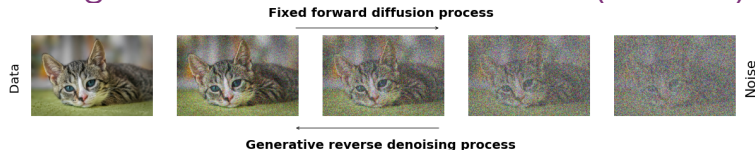
Since $\mathcal{L}(Z_T) \neq \pi_Z$, starting $\overleftarrow{Z}_0 \sim \pi_Z$ and evolving $\overleftarrow{Z}$ according to SDE

$$d\overleftarrow{Z}_t = -(b(\overleftarrow{Z}_t) - \frac{1}{2}\nabla \cdot a(\overleftarrow{Z}_t) - \frac{1}{2}a(\overleftarrow{Z}_t)\nabla_x \log q_{T-t}(\overleftarrow{Z}_t))dt + \sigma(\overleftarrow{Z}_t)dB'_t,$$

the total variation (TV) distance satisfies

$$\|\rho_0 - \mathcal{L}(\overleftarrow{Z}_T)\|_{\text{TV}} \leq \|\mathbb{P}_{\rho_0}(Z_T \in \cdot) - \pi_Z\|_{\text{TV}} + \dots$$

Denoising Diffusion Probabilistic Models (DDPMs)



SDE for $\overleftarrow{Z}$ yields an **approximate** simulation algorithm for the data distribution ρ_0 [Che+22; CDS23]:

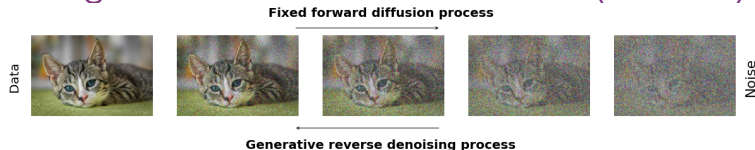
$$dZ_t = b(Z_t)dt + \sigma(Z_t)dB_t, \quad \text{with } Z_0 \sim \rho_0$$

For $\overleftarrow{Z}_0 \sim \pi_Z$ and evolution $\overleftarrow{Z}$ given by

$$d\overleftarrow{Z}_t = -(b(\overleftarrow{Z}_t) - \frac{1}{2}\nabla \cdot a(\overleftarrow{Z}_t) - \frac{1}{2}a(\overleftarrow{Z}_t)s_{T-t}(\overleftarrow{Z}_t))dt + \sigma(\overleftarrow{Z}_t)dB'_t,$$

$\nabla_x \log q_{T-t}$ is approximated with s_{T-t} using **score matching**.

Denoising Diffusion Probabilistic Models (DDPMs)



SDE for $\overleftarrow{Z}$ yields an **approximate** simulation algorithm for the data distribution ρ_0 [Che+22; CDS23]:

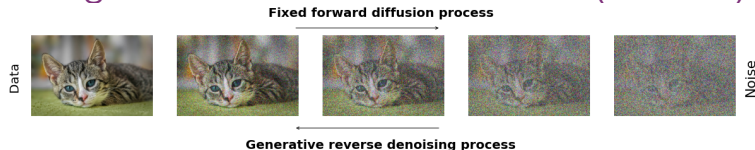
$$dZ_t = b(Z_t)dt + \sigma(Z_t)dB_t, \quad \text{with } Z_0 \sim \rho_0$$

For $\overleftarrow{Z}_0 \sim \pi_Z$ and evolution $\overleftarrow{Z}$ given by

$$d\overleftarrow{Z}_t = -(b(\overleftarrow{Z}_t) - \frac{1}{2}\nabla \cdot a(\overleftarrow{Z}_t) - \frac{1}{2}a(\overleftarrow{Z}_t)s_{T-t}(\overleftarrow{Z}_t))dt + \sigma(\overleftarrow{Z}_t)dB'_t,$$

$\nabla_x \log q_{T-t}$ is approximated with s_{T-t} using **score matching**. Assuming $\mathbb{E}_{q_t}[|s_t - \nabla_x \log q_t|^2] \leq \varepsilon_{\text{score}}^2$ leads to

Denoising Diffusion Probabilistic Models (DDPMs)



SDE for $\overleftarrow{Z}$ yields an **approximate** simulation algorithm for the data distribution ρ_0 [Che+22; CDS23]:

$$dZ_t = b(Z_t)dt + \sigma(Z_t)dB_t, \quad \text{with } Z_0 \sim \rho_0$$

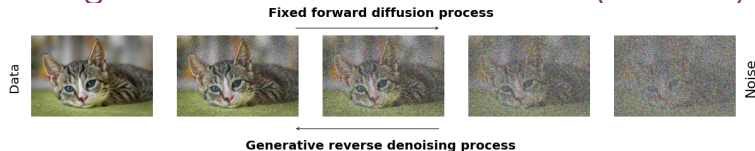
For $\overleftarrow{Z}_0 \sim \pi_Z$ and evolution $\overleftarrow{Z}$ given by

$$d\overleftarrow{Z}_t = -(b(\overleftarrow{Z}_t) - \frac{1}{2}\nabla \cdot a(\overleftarrow{Z}_t) - \frac{1}{2}a(\overleftarrow{Z}_t)s_{T-t}(\overleftarrow{Z}_t))dt + \sigma(\overleftarrow{Z}_t)dB'_t,$$

$\nabla_x \log q_{T-t}$ is approximated with s_{T-t} using **score matching**. Assuming $\mathbb{E}_{q_t}[|s_t - \nabla_x \log q_t|^2] \leq \varepsilon_{\text{score}}^2$ leads to

$$\|\rho_0 - \mathcal{L}(\overleftarrow{Z}_T)\|_{\text{TV}} \leq \|\mathbb{P}_{\rho_0}(Z_T \in \cdot) - \pi\|_{\text{TV}} + \varepsilon_{\text{score}}\sqrt{T} + \dots$$

Denoising Diffusion Probabilistic Models (DDPMs)



SDE for $\overleftarrow{Z}$ yields an **approximate** simulation algorithm for the data distribution ρ_0 [Che+22; CDS23]:

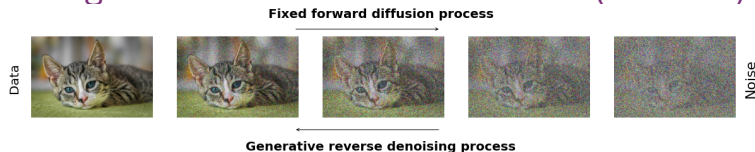
$$dZ_t = b(Z_t)dt + \sigma(Z_t)dB_t, \quad \text{with } Z_0 \sim \rho_0$$

For $\overleftarrow{Z}_0 \sim \pi_Z$ and evolution $\overleftarrow{Z}$ given by

$$d\overleftarrow{Z}_t = -(b(\overleftarrow{Z}_t) - \frac{1}{2}\nabla \cdot a(\overleftarrow{Z}_t) - \frac{1}{2}a(\overleftarrow{Z}_t)s_{T-t}(\overleftarrow{Z}_t))dt + \sigma(\overleftarrow{Z}_t)dB'_t,$$

Euler “type” schemes for Z and $\overleftarrow{Z}$ lead to the discretisation error

Denoising Diffusion Probabilistic Models (DDPMs)



SDE for $\overleftarrow{Z}$ yields an **approximate** simulation algorithm for the data distribution ρ_0 [Che+22; CDS23]:

$$dZ_t = b(Z_t)dt + \sigma(Z_t)dB_t, \quad \text{with } Z_0 \sim \rho_0$$

For $\overleftarrow{Z}_0 \sim \pi_Z$ and evolution $\overleftarrow{Z}$ given by

$$d\overleftarrow{Z}_t = -(b(\overleftarrow{Z}_t) - \frac{1}{2}\nabla \cdot a(\overleftarrow{Z}_t) - \frac{1}{2}a(\overleftarrow{Z}_t)s_{T-t}(\overleftarrow{Z}_t))dt + \sigma(\overleftarrow{Z}_t)dB'_t,$$

Euler “type” schemes for Z and $\overleftarrow{Z}$ lead to the discretisation error

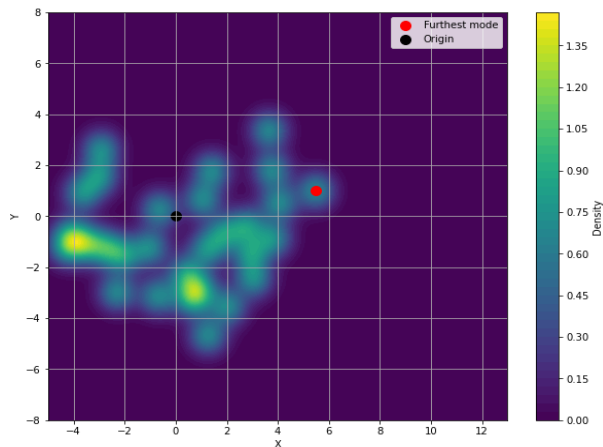
$$\|\rho_0 - \mathcal{L}(\overleftarrow{Z}_T)\|_{\text{TV}} \leq \|\mathbb{P}_{\rho_0}(Z_T \in \cdot) - \pi_Z\|_{\text{TV}} + \varepsilon_{\text{score}}\sqrt{T} + \varepsilon_{\text{disc}}\sqrt{T}$$

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 $b_\rho = \rho_0(B(x_0, R\delta)) > 3\varepsilon$ and $\rho_0(B(0, R(1 + 2\delta))^c) < \varepsilon/2$.

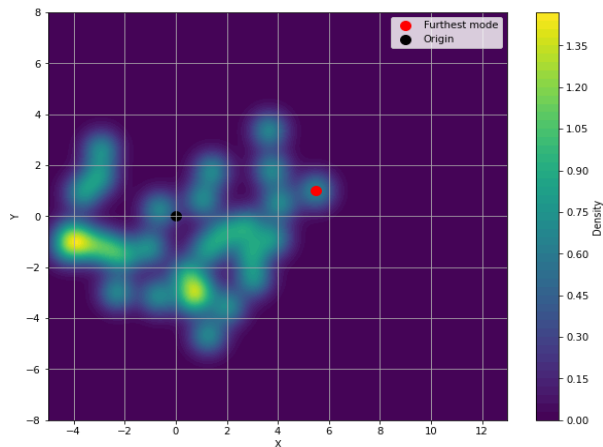
Assumption (DATA): multi-modality of initial data law

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 $b_\rho = \rho_0(B(x_0, R\delta)) > 3\varepsilon$ and $\rho_0(B(0, R(1 + 2\delta)))^c < \varepsilon/2$.



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

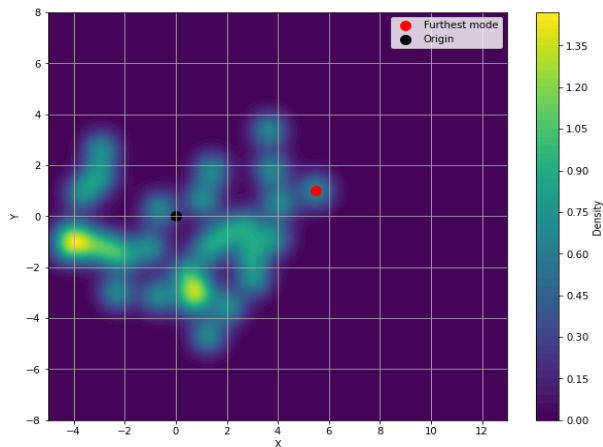
For $\rho_0 \in \mathcal{M}_{1,R,\varepsilon}(\mathbb{R}^d)$, there exists $x_0 \in \mathbb{R}^d$ with $|x_0| = R(1 + \delta)$, such that $b_\rho = \rho_0(B(x_0, R\delta)) > 3\varepsilon$ and $\rho_0(B(0, R(1 + 2\delta)))^c < \varepsilon/2$.



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 $b_\rho = \rho_0(B(x_0, R\delta)) > 3\varepsilon$ 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].

Convergence of OU process and cut-off

Ornstein-Uhlenbeck (OU) process: $dX_t = -X_t dt + \sqrt{2} dB_t$

Convergence of OU process and cut-off

Ornstein-Uhlenbeck (OU) process: $dX_t = -X_t dt + \sqrt{2} dB_t$

Proposition

Assume $\rho_0 \in \mathcal{M}_{1,R,\varepsilon}$ satisfying (DATA) with R large. Consider the OU process X and its Gaussian invariant measure π_X and recall $b_\rho = \rho_0(B(x_0, R\delta))$. Then

$$\|\mathbb{P}_{\rho_0}(X_{T_c} \in \cdot) - \pi_X\|_{TV} \geq b_\rho/2 \gg \varepsilon$$

for $T_c = \log R - \max\{\sqrt{2 \log(1/\varepsilon)}, 1\}$ and

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

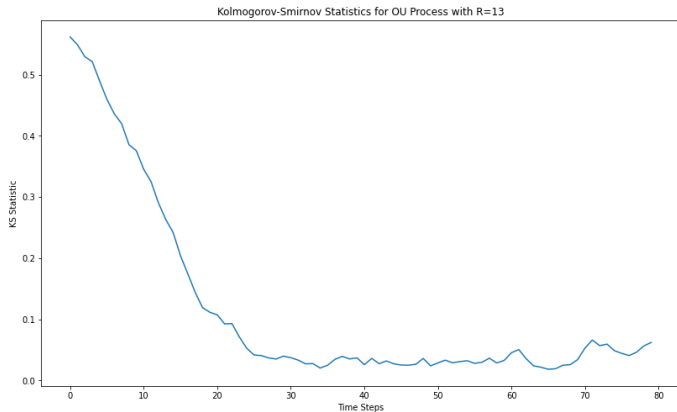
for $T_\varepsilon = \log R + \log(1 + 2\delta) + \log(1/\varepsilon)$.

Remark: Fixing the values of δ, ε leads to a cut-off like phenomenon i.e.,

$$\lim_{R \rightarrow \infty} \frac{T_\varepsilon}{T_b} = 1.$$

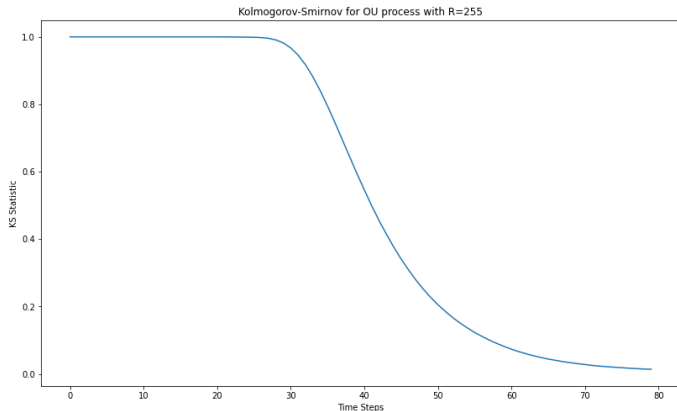
Simulation of OU: $d = 32 \times 32$ with $R = 13$

We simulate OU process X started at an image of dimension $d = 32 \times 32$ and measure distance between marginal X_t and the Gaussian invariant measure π_X .



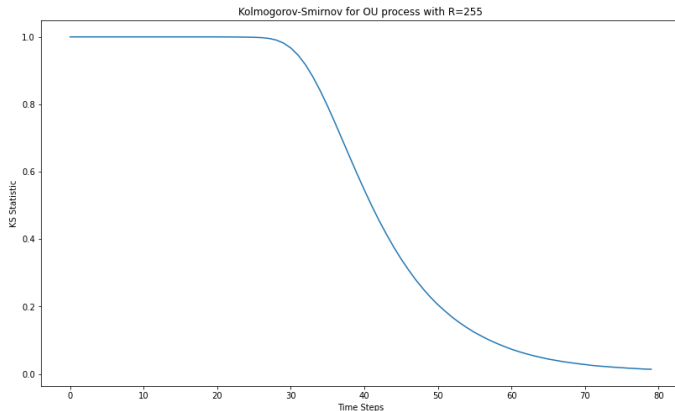
Simulation of OU: $d = 1024 \times 1024$ with $R = 255$

We simulate OU process X started at an image of dimension 1024×1024 and measure distance between the OU marginal X_t and the Gaussian invariant measure π_X .



Simulation of OU: $d = 1024 \times 1024$ with $R = 255$

We simulate OU process X started at an image of dimension 1024×1024 and measure distance between the OU marginal X_t and the Gaussian invariant measure π_X .



Can replacing the OU process X with an ergodic diffusion Y speed-up the convergence in the presence of **large R**?

Motivation: ergodicity of Langevin diffusions in $\mathbb{R}^d$

Let π have density proportional to $x \mapsto \exp(-h_\pi(x))$ and

$$dY_t = \nabla h_\pi(Y_t)dt + \sqrt{2}dB_t.$$

Motivation: ergodicity of Langevin diffusions in $\mathbb{R}^d$

Let π have density proportional to $x \mapsto \exp(-h_\pi(x))$ and

$$dY_t = \nabla h_\pi(Y_t)dt + \sqrt{2}dB_t.$$

Assume $h_\pi(x) \sim |x|^p$ as $|x| \rightarrow \infty$ for $p \in (0, 1)$.

Motivation: ergodicity of Langevin diffusions in $\mathbb{R}^d$

Let π have density proportional to $x \mapsto \exp(-h_\pi(x))$ and

$$dY_t = \nabla h_\pi(Y_t)dt + \sqrt{2}dB_t.$$

Assume $h_\pi(x) \sim |x|^p$ as $|x| \rightarrow \infty$ for $p \in (0, 1)$. Then, by [DFG09; BM24], for any $\varepsilon > 0$ and all $t \in [1, \infty)$ we set $\eta_0 = p/(2 - p)$ and get

$$C_-(y) \exp(-c_- t^{\eta_0}) \leq \|\mathbb{P}_y(Y_t \in \cdot) - \pi\|_{\text{TV}} \leq C_+(y) \exp(-c_+ t^{\eta_0}). \quad (1)$$

Motivation: ergodicity of Langevin diffusions in $\mathbb{R}^d$

Let π have density proportional to $x \mapsto \exp(-h_\pi(x))$ and

$$dY_t = \nabla h_\pi(Y_t)dt + \sqrt{2}dB_t.$$

Assume $h_\pi(x) \sim |x|^p$ as $|x| \rightarrow \infty$ for $p \in (0, 1)$. Then, by [DFG09; BM24], for any $\varepsilon > 0$ and all $t \in [1, \infty)$ we set $\eta_0 = p/(2 - p)$ and get

$$C_-(y) \exp(-c_- t^{\eta_0}) \leq \|\mathbb{P}_y(Y_t \in \cdot) - \pi\|_{\text{TV}} \leq C_+(y) \exp(-c_+ t^{\eta_0}). \quad (1)$$

Compare to *tempered Langevin diffusion* for $\ell \geq 0$:

$$dY_t = -h_\pi(x)^{2\ell-1}(h_\pi(x) - 2\ell)\nabla h_\pi(x)dt + \sqrt{2}h_\pi(Y_t)^\ell dB_t.$$

Motivation: ergodicity of Langevin diffusions in $\mathbb{R}^d$

Let π have density proportional to $x \mapsto \exp(-h_\pi(x))$ and

$$dY_t = \nabla h_\pi(Y_t)dt + \sqrt{2}dB_t.$$

Assume $h_\pi(x) \sim |x|^p$ as $|x| \rightarrow \infty$ for $p \in (0, 1)$. Then, by [DFG09; BM24], for any $\varepsilon > 0$ and all $t \in [1, \infty)$ we set $\eta_0 = p/(2 - p)$ and get

$$C_-(y) \exp(-c_- t^{\eta_0}) \leq \|\mathbb{P}_y(Y_t \in \cdot) - \pi\|_{\text{TV}} \leq C_+(y) \exp(-c_+ t^{\eta_0}). \quad (1)$$

Compare to *tempered Langevin diffusion* for $\ell \geq 0$:

$$dY_t = -h_\pi(x)^{2\ell-1}(h_\pi(x) - 2\ell)\nabla h_\pi(x)dt + \sqrt{2}h_\pi(Y_t)^\ell dB_t.$$

For $\ell := 1/p - 1 > 0$, we get:

- **multiplicative** noise $h_\pi(x)^\ell \sim |x|^{1-p}$ and drift, both sublinear, and

Motivation: ergodicity of Langevin diffusions in $\mathbb{R}^d$

Let π have density proportional to $x \mapsto \exp(-h_\pi(x))$ and

$$dY_t = \nabla h_\pi(Y_t)dt + \sqrt{2}dB_t.$$

Assume $h_\pi(x) \sim |x|^p$ as $|x| \rightarrow \infty$ for $p \in (0, 1)$. Then, by [DFG09; BM24], for any $\varepsilon > 0$ and all $t \in [1, \infty)$ we set $\eta_0 = p/(2 - p)$ and get

$$C_-(y) \exp(-c_- t^{\eta_0}) \leq \|\mathbb{P}_y(Y_t \in \cdot) - \pi\|_{\text{TV}} \leq C_+(y) \exp(-c_+ t^{\eta_0}). \quad (1)$$

Compare to *tempered Langevin diffusion* for $\ell \geq 0$:

$$dY_t = -h_\pi(x)^{2\ell-1}(h_\pi(x) - 2\ell)\nabla h_\pi(x)dt + \sqrt{2}h_\pi(Y_t)^\ell dB_t.$$

For $\ell := 1/p - 1 > 0$, we get:

- **multiplicative** noise $h_\pi(x)^\ell \sim |x|^{1-p}$ and drift, both sublinear, and
- the rate in (1) is **exponential!!!**

Main result

Class of forward processes:

- tempered Langevin diffusion Y with drift b satisfying $|b(x)| \leq \mu|x|$ for a fixed constant $\mu > 0$;

Main result

Class of forward processes:

- tempered Langevin diffusion Y with drift b satisfying $|b(x)| \leq \mu|x|$ for a fixed constant $\mu > 0$;
- invariant measure π of Y is spherically symmetric,
 $\pi(B_3(r_3) \times \mathbb{R}^{d-3}) > 1-\varepsilon/2$ and $2r_3 \leq R^\beta$ for small $0 < \beta \ll 1$,
where $B_3(r_3)$ is the closed ball in $\mathbb{R}^3$ with radius $(r_3^2 - 1)^{1/2}$.

Main result

Class of forward processes:

- tempered Langevin diffusion Y with drift b satisfying $|b(x)| \leq \mu|x|$ for a fixed constant $\mu > 0$;
- invariant measure π of Y is spherically symmetric,

$\pi(B_3(r_3) \times \mathbb{R}^{d-3}) > 1 - \varepsilon/2$ and $2r_3 \leq R^\beta$ for small $0 < \beta \ll 1$,
where $B_3(r_3)$ is the closed ball in $\mathbb{R}^3$ with radius $(r_3^2 - 1)^{1/2}$.

Theorem

Let **(DATA)** hold and pick $\rho_0 \in \mathcal{M}_{1,R,\varepsilon}(\mathbb{R}^d)$ with distance to furthest mode R sufficiently large. For Y in the class above,

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

Main result

Class of forward processes:

- tempered Langevin diffusion Y with drift b satisfying $|b(x)| \leq \mu|x|$ for a fixed constant $\mu > 0$;
- invariant measure π of Y is spherically symmetric,
 $\pi(B_3(r_3) \times \mathbb{R}^{d-3}) > 1 - \varepsilon/2$ and $2r_3 \leq R^\beta$ for small $0 < \beta \ll 1$,
where $B_3(r_3)$ is the closed ball in $\mathbb{R}^3$ with radius $(r_3^2 - 1)^{1/2}$.

Theorem

Let **(DATA)** hold and pick $\rho_0 \in \mathcal{M}_{1,R,\varepsilon}(\mathbb{R}^d)$ with distance to furthest mode R sufficiently large. For Y in the class above,

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

Moreover, the OU process X satisfies

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

Concluding remarks

OU process cannot be substantially improved within the large class of diffusions with (possibly) multiplicative noise under the Manifold Hypothesis.

Concluding remarks

OU process cannot be substantially improved within the large class of diffusions with (possibly) multiplicative noise under the Manifold Hypothesis.

Our theorem suggests that the terminal time horizon T for the forward diffusion should be taken as $\log R + \log(1/\varepsilon)$.

Concluding remarks

OU process cannot be substantially improved within the large class of diffusions with (possibly) multiplicative noise under the Manifold Hypothesis.

Our theorem suggests that the terminal time horizon T for the forward diffusion should be taken as $\log R + \log(1/\varepsilon)$.

The growth $T_c \approx \log R$ motivates the use of Schrödinger bridges [De+21] and development of efficient sampling algorithms for uniformly ergodic diffusions [Liv+19]

Concluding remarks

OU process cannot be substantially improved within the large class of diffusions with (possibly) multiplicative noise under the Manifold Hypothesis.

Our theorem suggests that the terminal time horizon T for the forward diffusion should be taken as $\log R + \log(1/\varepsilon)$.

The growth $T_c \approx \log R$ motivates the use of Schrödinger bridges [De+21] and development of efficient sampling algorithms for uniformly ergodic diffusions [Liv+19]

Can cut-off for the reverse process be proved rigorously?

Bibliography

- [And82] Brian D. O. Anderson. “Reverse-time diffusion equation models”. *Stochastic Process. Appl.* 12 (1982).
- [BM24] Miha Brešar and Aleksandar Mijatović. “Subexponential lower bounds for f -ergodic Markov processes”. *arXiv:2403.14826 [math.PR]*, to appear in *PTRF* (2024), 47 pages.
- [CDS23] Giovanni Conforti, Alain Durmus, and Marta Gentiloni Silveri. “Score diffusion models without early stopping: finite Fisher information is all you need”. *arXiv:2308.12240* (2023).
- [Che+22] Sitan Chen, Sinho Chewi, Jerry Li, Yuanzhi Li, Adil Salim, and Anru Zhang. “Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions”. *ICLR*. 2022.
- [De +21] Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. “Diffusion schrödinger bridge with applications to score-based generative modeling”. *NeurIPS* 34 (2021).
- [DFG09] Randal Douc, Gersende Fort, and Arnaud Guillin. “Subgeometric rates of convergence of f -ergodic strong Markov processes”. *Stochastic Process. Appl.* 119.3 (2009).
- [FMN16] Charles Fefferman, Sanjoy Mitter, and Hariharan Narayanan. “Testing the manifold hypothesis”. *JAMS* 29.4 (2016), pp. 983–1049.
- [Liv+19] Samuel Livingstone, Michael Betancourt, Simon Byrne, and Mark Girolami. “On the geometric ergodicity of Hamiltonian Monte Carlo”. *Bernoulli* 25.4A (2019), pp. 3109–3138.