Supplementary Materials for
Generative Adversarial Networks for Markovian Temporal Dynamics

Sung Woo Park [1] **Dong Wook Shu** [1] **Junseok Kwon** [1]

1. Mathematical Backgrounds

1.1. Notations

In this paper, a time-dependent system is defined on probability space $(\mathbb{R}^d, \mathcal{F}, \mathbb{P})$ with filtration $\mathcal{F}_t$. We assume that generated process X_t is $\mathcal{F}_t$-adapted for all t.

1.2. Assumptions

For rigorous technical results, we assume the following conditions for both generator and discriminator networks.

Assumptions for Generator Network.

- **G-H1**:
$$\left\|f^\theta(t,x) - f^\theta(t,y)\right\|^2 \vee \|\sigma(x,\theta) - \sigma(y,\theta)\|^2 \vee \mathbf{Tr}[\sigma(x,\theta)^T\sigma(x,\theta)] \leq K, \quad \forall x, y \in \mathbb{R}^d. \tag{1}$$
- **G-H2**: The infinitesimal generator of the parameterized Fokker-Planck equation induces the curvature-dimension condition: $CD(\kappa, \infty)$ (Villani, 2008; Bakry & Émery).

Assumptions for Discriminator Network.

- **D-H1**: The discriminator network is p-Lipschitz on $\mathcal{T}$, and q-Lipschitz on $\mathbb{R}^d$ in a global sense:
$$|D(\cdot, t_1) - D(\cdot, t_2)| \leq p\,|t_1 - t_2|\,, \quad |D(X_1, \cdot) - D(X_2, \cdot)| \leq q\,\|X_1 - X_2\| \tag{2}$$
for all $t_1 \neq t_2 \in \mathcal{T}$ and $X_1 \neq X_2 \in \mathbb{R}^d$.
- **D-H2**: The norm of second derivatives for the discriminator network is always bounded for some value $\hat{q}$: *i.e.*, $\sup_{i,j} \|\partial_i \partial_j D(x, \cdot)\| \leq \hat{q}$.

1.3. Stochastic Differential Equations

In the main paper, we use the integral formulation, but it is generally written as *Itô's diffusion*:

$$dX_t = f^\theta(X_t, t)dt + \sigma(X_t)dW_t, \tag{3}$$

where $X_t \in \mathbb{R}^d$ and $f : \mathbb{R}^d \times \mathbf{U} \to \mathbb{R}^d$ is a neural network parameterized by θ.

By the Lipschitz continuity (*i.e.*, **G-H1**) of both drift and diffusion functions, the solution to (3), X_t, is always a Markov process (Øksendal, 2003).

2. Proofs

Proposition 1. *(Controlled Stability of Discriminator) Let $G(X_0, s) = X_s$ be a generated sample obtained by the generator G. For simple analysis, let us consider $\sigma(X) := \sigma$ for some positive scalar $\sigma > 0$.*[1] *Then, the following probability*

[1]The result of this proposition can be easily extended to general measurable $\sigma(\cdot)$, if we clarify the explicit condition on σ.

inequality is satisfied:

$$\mathbb{P}\left[\sup_{0\le s\le t}\|D(X_s,s)-D(X_0,0)\|\ge\epsilon\right]\le\frac{2}{3\epsilon}\Big\{(p\vee q)\mathbb{E}\|X_t-X_0\|+tC\Big\},\tag{4}$$

where the numerical constant C is linearly dependent on σ. In other words, $C\propto\sigma$.

Proof. It is difficult to directly analyse the time-inhomogeneous Feller process, X_t, without appropriate and complicated assumptions on f^θ. Because using a time-variable is to generate high-dimensional and complex data, we transform the time-inhomogeneous Markov process, X_t, into a desirable form and analyze probabilistic properties of X_t. Let $\tilde{X}_t=(X_t,t)$ be a time-augmented stochastic process suggested in (Bossy & Champagnat, 2010; Böttcher, 2014), it is easily shown that the aforementioned time-augmented Markov semigroup can be defined. Let $\mathcal{F}_t=\sigma(X_s;s\le t)$ be a canonical filtration of X_t. Let $\mu\in\mathcal{P}(\mathbb{R}^d)$ be a probability measure and define $\mathbb{P}:=\mathbb{P}^\mu(X_0\in A)\otimes\mathcal{T}(t_0\in T)$ for all $A\in\Sigma(\mathbb{R}^d\times\mathcal{T})$.

Theorem 1. *(Böttcher, 2014) Let $\mathcal{T}=[0,C]$, assume that X_t is a time-inhomogeneous Feller process and has right-continuous infinitesimal generator A_s^+. Let $f\in C_\infty(\mathbb{R}^d\times\mathcal{T}),\pi_1\circ f\in C^1(\mathcal{T}),\pi_2\circ f\in\mathcal{D}(A_s^+)$. Then, $\tilde{X}_t$ is a time-homogeneous Feller process with generator $\tilde{L}$ defined as follows:*

$$Lf(\tilde{X})=\partial_s f(s,x)+A_s^+f(s,x),\tag{5}$$

where $A_s^+f(s,\cdot)=\frac{\sigma}{2}\sum_i^d\nabla_i^2 f(s,\cdot)+\nabla^T f(s,\cdot)$.

Based on the tools above, we reveal the impact to the probabilistic bound of perturbation according to varying magnitudes of σ in our SDE model. Let us first define the stochastic process $M_t^D:\mathbb{R}^d\times\mathcal{T}\to\mathbb{R}$ as follows:

$$M_t^D=D(\hat{X}_t)-D(\hat{X}_0)-\int_0^t(\partial_u+A_u^+)D(X_u,u)du.\tag{6}$$

This form is the time-inhomogeneous type of martingale formulation (Bossy & Champagnat, 2010) for itó's formula over discriminator D, *i.e.*, $\mathbb{E}[M_t^D|\mathcal{F}_s]=M_s^D$. In this form, the distortions induced by inhomogeneity are compensated by differential operator ∂_s. As M_t^D is martingale, one can induce the following probability inequality by applying Doob's maximal martingale inequality (Øksendal, 2003) to M_t^D:

$$\mathbb{P}\left[\sup_{0\le s\le t}\left\|M_s^D\right\|_2\ge\epsilon\right]\le\frac{1}{\epsilon^2}\mathbb{E}\left[\left\|M_t^D\right\|\right].\tag{7}$$

From (7), we can obtain the following inequality:

$$\begin{aligned}\epsilon\le\sup_{0\le s\le t}\left\|M_s^D\right\|_2&\le\sup_{0\le s\le t}\left[\|D(X_s,s)-D(X_{s=0},s=0)\|_2+\left\|\int_0^t-\partial_u D(X_u,u)du\right\|_2+\left\|\int_0^t-A_u^+D(X_u,u)du\right\|_2\right]\\&\le\sup_{0\le s\le t}\|D(X_s,s)-D(X_{s=0},s=0)\|_2+\sup_{0\le s\le t}\int_0^t\|-\partial_u D(X_u,u)\|_2\,du+\sup_{0\le s\le t}\int_0^t\left\|-A_u^+D(X_u,u)\right\|_2du.\end{aligned}\tag{8}$$

The second inequality is induced by applying Jensen's inequality to Lebesgue measure du with convex function $\|\cdot\|_2$, and the inequality $\sup_s[A(s)+B(s)+C(s)]\le\sup_s A(s)+\sup_s B(s)+\sup_s C(s)$.

$$\begin{aligned}\frac{1}{\epsilon^2}\mathbb{E}\left[\left\|M_t^D\right\|\right]\ge\mathbb{P}\left[\epsilon\le\sup_{0\le s\le t}\left\|M_s^D\right\|_2\right]\ge\mathbb{P}\left[\frac{\epsilon}{3}\le\sup_{0\le s\le t}\|D(X_s,s)-D(X_0,0)\|_2\right]\\+\mathbb{P}\left[\frac{\epsilon}{3}\le\sup_{0\le s\le t}B(t)\right]+\mathbb{P}\left[\frac{\epsilon}{3}\le\sup_{0\le s\le t}C(t)\right].\end{aligned}\tag{9}$$

By rewriting inequality above using $\mathbb{P}(A^c)=1-\mathbb{P}(A),\forall A\in\Sigma$, and rescaling dummy variable $\epsilon\to3\epsilon$, we get

$$\mathbb{P}\left[\epsilon\le\sup_{0\le s\le t}\left\|D(\tilde{X}_s)-D(\tilde{X}_0)\right\|_2\right]\le\frac{1}{3\epsilon}\mathbb{E}\left[\left\|M_t^D\right\|\right]+\mathbb{P}\left[\epsilon\ge\sup_{0\le s\le t}B(t)\right]+\mathbb{P}\left[\epsilon\ge\sup_{0\le s\le t}C(t)\right]-2.\tag{10}$$

Then, we use the assumptions to remove second and third term in (10). While we assume that D is global p-Lipschitz continuous on $\mathcal{T}$, the following inequality is induced by **D-H1**.

$$\mathbb{P}\big[\, |D(X,t_1) - D(X,t_2)| \le p \,\|(X,t_1) - (X,t_2)\| = p\,|t_1 - t_2| \,\big] = 1. \tag{11}$$

Let us assume $\epsilon \le pt$. As the inequality in (11) is equivalent to $\|\partial_u D\| \le p$ iff D is p-Lipschitz continuous, the second term in right-hand side of (10) is naturally bounded above with the following inequality because we assume that $\epsilon \le pt$. Subsequently, we get followings:

$$\begin{aligned}\mathbb{P}\left[\sup_{0\le s\le t} B(t) \le \epsilon\right] &\le \mathbb{P}\left[\sup_{0\le s\le t}\int_0^t \|\partial_u D(G(X_0,u),u)\|\,du \le \sup_{0\le s\le t}\int_0^t p du = pt\right]\\ &\le \mathbb{P}\left[\sup_{0\le s\le t}\int_0^t \|\partial_u D(X_u,u)\|\,du \le pt\right] = 1.\end{aligned} \tag{12}$$

As our discriminator is assumed to be q-Lipschitz on data dimension, the following inequality is naturally induced. The probability densities, P and Q induced by both $d\mathbb{P}_t = p_t(x)d\mathcal{L}(x)$, $d\mathbb{Q}_t = q_t(x)d\mathcal{L}(x)$ for Lebesgue measure $\mathcal{L}(x)$ with respect to $\mathbb{R}^d$, and $U_{P,Q} \coloneqq \mathbf{supp}(p_t) \cup \mathbf{supp}(q_t) \subset \mathbb{R}^d$. Based on the assumption that $D(\cdot,t)$ is global p-Lipschitz continuous for any $t \in \mathcal{T}$, we can induce the following set inclusion:

$$\{w : D\left[G(X_0,t)(w),t\right] \in \mathbf{Lip}_p(U_{P,Q})\} \subset \{w : D(Y(w),t) \in \mathbf{Lip}_p(\mathbb{R}^d), \forall\, Y \in \mathbb{R}^d\}. \tag{13}$$

Assume that $D(\cdot,t)$ vanishes outside of $\mathbf{supp}(p_t)$. Then, in probability, we can induce

$$\mathbb{P}\big[D\left[G(X(w),t),t\right] \in \mathbf{Lip}_p(\mathbf{supp}(P))\big] \le \mathbb{P}\left[D(X,t) \in \mathbf{Lip}_p(\mathbb{R}^d)\right] = \mathbb{P}\left[\|\nabla_i D(X,t)\| \le q\right] = 1. \tag{14}$$

$$\left\|A_s^+ D(X_s,s)\right\| \le \frac{\sigma}{2}\underbrace{\left\|\sum_i^d \nabla_i^2 D(X_s,s)\right\|}_{\text{bounded second derivative}} + \underbrace{\left\|\sum_i^d \nabla_i D(X_s,s)\right\|}_{\text{q-Lipschitz on data space}} \le 2^{-1}\sigma d \sup_{0\le i\le d}\hat{q}_i + q. \tag{15}$$

We denote $\hat{q} = \sup_{0\le i\le d}\hat{q}_i$ for simplicity. The following equality is naturally induced by the assumption **D-H2**.

$$\mathbb{P}\left[\left\|A_s^+ D(X_s,s)\right\| \le 2^{-1}\sigma d\hat{q} + q\right] = 1,\ \forall\, 0 \le s \le t. \tag{16}$$

If $\epsilon \le 2^{-1}\sigma d\hat{q} + q$, this naturally induces the following probability inequality:

$$\mathbb{P}\left[\sup_{0\le s\le t} C(t) \le \epsilon\right] \le \mathbb{P}\left[\sup_{0\le s\le t}\int_0^t \left\|A_u^+ D(X_u,u)\right\| du \le (2^{-1}\sigma d\hat{q} + q)t\right] = 1. \tag{17}$$

Lemma 1. *The discriminator network D is $2(p\vee q)$-Lipschitz on $\mathbb{R}^d \times \mathcal{T}$.*

Proof. The proof is trivial by the triangle inequality.

$$\begin{aligned}|D(X_1,t_1) - D(X_2,t_2)| &\le |D(X_1,t_1) - D(X_1,t_2)| + |D(X_1,t_2) - D(X_2,t_2)|\\ &\le p\,|t_1 - t_2| + q\,\|X_1 - X_2\|\\ &\le 2(p\vee q)\left[|t_1 - t_2| + \|X_1 - X_2\|\right] = 2(p\vee q)\,\|(X_1,t_1) - (X_2,t_2)\|\,.\end{aligned} \tag{18}$$

□

By integrating inequalities in (17) and (12) into (10), we can obtain the following inequality:

$$\begin{aligned}\mathbb{P}\left[\sup_{0\le s\le t}\|D(X_s,s) - D(X_0,0)\| \ge \epsilon\right] &\le \frac{1}{3\epsilon}\mathbb{E}\left[\left\|M_t^D\right\|\right] \le \frac{1}{3\epsilon}\Big\{\mathbb{E}\,\|D(X_t,t) - D(X_0,0)\| + t\left[p + q + 2^{-1}\sigma d\hat{q}\right]\Big\}\\ &\le \frac{2}{3\epsilon}\Big\{(p\vee q)\mathbb{E}\,\|X_t - X_0\| + t\left[2(p\vee q) + 4^{-1}\sigma d\hat{q}\right]\Big\},\end{aligned} \tag{19}$$

where $\epsilon = [p \wedge (2^{-1}\sigma d\hat{q} + q)]t$. The second inequality induced by the fact that D is global $2(p \vee q)$-Lipschitz continuous by Lemma 1, and the metric on $\mathbb{R}^d \times \mathcal{T}$ can be decomposed into metrics on $\mathbb{R}^d$ and $\mathcal{T}$.

$$\|\mathbb{E}\left[D(X_t,t) - D(X_0,0)\right]\| \leq \mathbb{E}\left\|D(X_t,t) - D(X_0,0)\right\| \leq 2(p\vee q)\mathbb{E}\left\|\hat{X}_t - \hat{X}_0\right\| \leq 2(p\vee q)\left(\mathbb{E}\left\|X_t - X_0\right\| + t\right). \tag{20}$$

The proof is completed by setting $C\left(p,q,\sigma,d,\left\|\nabla^2 D\right\|\right) = 2(p\vee q) + 4^{-1}\sigma d\hat{q}$. Please note that this numerical constant is linearly dependent on σ. □

Proposition 2. *(Controlled Stability of Wasserstein distance) Let us define the spatial-temporal gradient operator as $\tilde{\nabla}_{x,t} = \nabla_x + \partial_t$. Then, the expectation norm of the spatial-temporal gradient for the conditional distance is bounded as follows:*

$$\mathbb{E}\left[\left\|\tilde{\nabla}_{x,t}\mathcal{W}^{\varphi}(\mathbb{P}_t|x,\mathbb{Q}_t)\right\|\right] \leq C + (p\vee q)(1+e^{-\kappa t}) \tag{21}$$

for some numerical constants $\kappa, C > 0$.

Proof. The left-hand side of inequality in (21) can be divided into two terms as follows:

$$\mathbb{E}\left[\left\|\tilde{\nabla}_{x,t}\mathcal{W}^{\varphi}(\mathbb{P}_t|x,\mathbb{Q}_t)\right\|\right] \leq \mathbb{E}\left[\left\|\nabla\mathcal{W}^{\varphi}(\mathbb{P}_t|x,\mathbb{Q}_t)\right\|\right] + \left[\left\|\partial_t\mathcal{W}^{\varphi}(\mathbb{P}_t|x,\mathbb{Q}_t)\right\|\right]. \tag{22}$$

First, We investigate the first term of right-hand side in (22):

$$\begin{aligned}
\mathbb{E}_x\left\|\nabla\mathcal{W}^{\varphi}(\mathbb{P}_t|x,\mathbb{Q}_t)\right\| &= \int\left\|\nabla M_t D^{\varphi}(x,0) - \nabla\mathbb{E}_{Y_t\sim\mathbb{Q}_t}D^{\varphi}(Y_t,t)\right\|d\mathbb{P}_0(x) \leq \int e^{-\kappa t}\mathbb{E}_x M_t\left(\left\|\nabla D(x)\right\|\right)d\mathbb{P}_0(x)\\
&= e^{-\kappa t}\int\left\|\nabla D_1\right\|(x)p(t,y|0,x)p_0(x)d\mathcal{L}(x) = e^{-\kappa t}\int\left\|\nabla D_1\right\|d\mathbb{P}_t = e^{-\kappa t}\mathbb{E}\left[\left\|\nabla D(X_t)\right\|\right]\\
&\leq e^{-\kappa t}q.
\end{aligned} \tag{23}$$

The first inequality is induced by the assumption **G-H2** on curvature-dimension condition $CD(\kappa,\infty)$ of our parameterized Fokker-Planck equation. By the spatial constraints assumption **D-H2**, The last inequality is induced as $\|\nabla D\| \leq q$, $d\mathbb{P}_t(x)$-almost surely. Subsequently, we investigate the second term of right-hand side in (22):

$$\begin{aligned}
\mathbb{E}_x\left\|\partial_t\mathcal{W}^{\varphi}(\mathbb{P}_t|x,\mathbb{Q}_t)\right\| &= \int\left\|\partial_t M_t D^{\varphi}(x,0) - \partial_t\mathbb{E}_{Y_t\sim\mathbb{Q}_t}D^{\varphi}(Y_t,t)\right\|d\mathbb{P}_0(x,0)\\
&= \int\left\|M_t L D^{\varphi}(x,0) - \mathbb{E}\left[\partial_t D^{\varphi}(Y_t,t)\right]\right\|d\mathbb{P}_0(x)\\
&\leq \int\left\|M_t L D^{\varphi}(x,0)\right\|d\mathbb{P}_0(x) + \int\mathbb{E}\left[\left\|\partial_t D^{\varphi}(Y_t,t)\right\|\right]d\mathbb{P}_0(x)\\
&\leq \int M_t\left[\left\|\partial_t D(x,t)\right\| + \frac{\sigma}{2}d\sup_{0\leq i\leq d}\hat{q}_i + \left\|\nabla D(X)\right\|\right]d\mathbb{P}_0(x)\\
&= p + \frac{\sigma}{2}d\sup_{0\leq i\leq d}\hat{q}_i + \mathbb{E}\left\|\nabla D(X_t)\right\| \leq p + q + 2^{-1}\sigma d\hat{q}.
\end{aligned} \tag{24}$$

The first inequality is induced by the dual identity of Fokker-Planck equation: $\partial_t M_t f = M_t L f$ for Markovian generator L, and we use the fact that $\partial_t\mathbb{E}f(x,t) = \mathbb{E}\partial_t f(x,t)$ for bounded and second differentiable $f(x,t)$. The second inequality is induced by dividing L defined in Theorem 1 into two terms. The third equality holds as $\mathbb{E}_{x_0}M_t\left\|\nabla D(X_0)\right\| = \mathbb{E}\left\|\nabla D(X_t)\right\|$, which is bounded above q almost surely. Combining these results, it is easy to see that the following inequality is satisfied: $2^{-1}\sigma d\hat{q} + p + (1+e^{-\kappa t})q \leq 2^{-1}\sigma d\hat{q} + (p\vee q)(1+e^{-\kappa t})$ where $C = 2^{-1}\sigma d\hat{q}$. By the fact that $\int e^{-\kappa t}dt \leq \frac{1-e^{-\kappa T}}{\kappa}$, the proof is completed. □

Proposition 3. *Let V^{λ} be the function defined above, and $x, \hat{x}$ be two initial states such that $\hat{\mathbb{P}}_t = h_{\#}[\mathbb{P}_t]$. If the generator solves the regularization term in (25), such that*

$$\min_{\theta}\mathbb{E}_{X_t\sim\mathbb{P}_t, Z_t\sim\hat{\mathbb{P}}_t}V^{\lambda}(\theta, X_t, Z_t) = 0, \tag{25}$$

the following inequality holds:

$$\mathcal{W}_2(\mathbb{P}_t, \hat{\mathbb{P}}_t) \leq \sqrt{A + e^{-2\lambda t} \|h\|_{\mathbf{L}_2(\mathbb{P})}}, \tag{26}$$

where $\|\cdot\|_{\mathbf{L}_2(\mathbb{P})}$ *denotes* L_2*-norm over probability measure* $\mathbb{P}$*, for some* $A > 0$.

Proof. Assume that the function V^λ vanishes for some for some fixed x, y and parameter $\theta^\star$. That is, $V^\lambda(\theta^\star, x, y) := 0$. In this case, (25) indicates the following inequality:

$$(x-y)^T \left[\nabla f^{\theta^\star}_{\alpha x + (1-\alpha) y}\right](x-y) \leq -\lambda (x-y)^T I (x-y), \tag{27}$$

where we simply denote $\nabla f^{\theta^\star} = \nabla_x f(\theta^\star, x, t)$ for the fixed t. The drift function satisfying the inequality above is called *contraction function*. By the Theorem 2 (Pham et al., 2009), this property gives powerful stochastic contraction for processes X^x_t, X^y_t starting at different initial states $x \sim \mu, y \sim \nu$. In particular, any diffusion Markov process of which drift functions satisfy inequality in (27) can induce the following property:

$$\mathbb{E}\left\|X^x_t - X^y_t\right\|^2 \leq \frac{K}{\lambda} + e^{-2\lambda t} \int_{\mathcal{A}} \|x-y\|^2 \, d(\mu \otimes \nu), \tag{28}$$

where $\mathcal{A} = \mathbf{supp}(\mu_0) \cup \mathbf{supp}(\nu_0) \subset \mathbb{R}^d$. Let us consider $\hat{x} = x + h(x)$ for some measurable h. As the optimal transport between $\mathbb{P}^x_t, \mathbb{P}^y_t$ always exists, which is denoted as $\pi^{x,y}_t$, inducing the followings are straightforward.

We consider the system of SDEs consist of trained drift, diffusion functions $f(\theta^\star), and \sigma(\theta^\star)$ with different initial states.

$$\begin{cases} dX_t = f(\theta^\star, X_t, t) + \sigma(X_t) dW^1_t \\ d\hat{X}_t = f(\theta^\star, \hat{X}_t, t) + \sigma(\hat{X}_t) dW^2_t \end{cases} \tag{29}$$

with i.i.d Wiener processes W^1_t, W^2_t. In this case, it is easy to see that $Z_t = (X_t, \hat{X}_t)$ is also a Markov process on $\mathbb{R}^d \times \mathbb{R}^d$. We define $\iota(Z_t) = d^2(X_t, \hat{X}_t)$ for the Euclidean metric d on $\mathbb{R}^d$ and define Π as an optimal transport between initial state measures μ and ν. Expectation of Markov semi-group $M_t\iota$ over π yields followings:

$$\begin{aligned} \int_{\mathcal{A}^2} M_t\iota(z) d\Pi &= \int_{\mathcal{A}^2} \mathbb{E}[\iota(Z_t)|z = (x,y)] d\Pi(x,y) = \int_{\mathcal{A}^2} \mathbb{E}\left[d^2(X_t, \hat{X}_t)^2 \middle| (X_0, \hat{X}_0) = (x,y)\right] d\Pi(x,y) \\ &\leq \frac{K}{\lambda} + e^{-2\lambda t} \int_{\mathcal{A}^2} \int_{\mathcal{A}} \iota(Z_0) d(\mu \otimes \nu) d\Pi(x,y) = \frac{K}{\lambda} + e^{-2\lambda t} \mathcal{W}^2_2(\mu, \nu) = \frac{K}{\lambda} + e^{-2\lambda t} \|h\|^2_{\mathbf{L}_2(\mu)}. \end{aligned} \tag{30}$$

$\Gamma_t = \mathbb{E}_{z \sim \Pi} p(t, z, \cdot)$ denotes a push forward of Π through transition kernel. Then, for the any Z_t,

$$\mathcal{W}^2_2(\mathbb{P}^{x\sim\mu}_t, \mathbb{P}^{\hat{x}\sim\nu}_t) = \inf_{\Pi_t} \int \iota(Z_t) d\Pi_t(Z_t) \leq \int \iota d\Gamma_t. \tag{31}$$

By combining two inequalities above and the fact that $\mathbb{E}_\Pi[M_t\iota] = \mathbb{E}_{\Gamma_t}[\iota]$, we can conclude that $\mathcal{W}_2(\mathbb{P}^x_t, \mathbb{P}^{\hat{x}}_t) \leq \sqrt{\frac{K}{\lambda} + e^{-2\lambda t} \|h\|^2_{\mathbf{L}_2(\mu)}}$, where $A = K\lambda^{-1}$.

□

References

Bakry, D. and Émery, M. Diffusions hypercontractives. *Séminaire de probabilités de Strasbourg*, pp. 177–206.

Bossy, M. and Champagnat, N. Markov processes and parabolic partial differential equations. In *Encyclopedia of Quantitative Finance*, pp. 1142–1159. 2010.

Böttcher, B. Feller evolution systems: Generators and approximation. *Stochastics and Dynamics*, 14(03):1350025, 2014.

Øksendal, B. Stochastic differential equations. In *Stochastic differential equations*, pp. 65–84. Springer, 2003.

Pham, Q.-C., Tabareau, N., and Slotine, J.-J. A contraction theory approach to stochastic incremental stability. *IEEE Transactions on Automatic Control*, 54(4):816–820, 2009.

Villani, C. *Optimal Transport: Old and New*. Springer Berlin Heidelberg, 2008.