

Relative Entropy Convergence under Picard's Iteration for Stochastic Differential Equations

Tsz Hin, Ng

The University of Hong Kong
kennthin@hku.hk

Guangyue, Han

The University of Hong Kong
ghan@hku.hk

October 9, 2018

Abstract

For a family of stochastic differential equations, we investigate the asymptotic behaviors of its corresponding Picard's iteration, establishing convergence results in terms of relative entropy. Our convergence results complement the conventional ones in the L^2 and almost sure sense, revealing some previously unexplored aspects of the stochastic differential equations under consideration. For example, in combination with Pinsker's inequality, one of our results readily yields the convergence under Picard's iteration in the total variation sense, which does not seem to directly follow from any other known results. Moreover, our results promise possible further applications of SDEs in related disciplines. As an example of such applications, we establish the convergence of the corresponding mutual information sequence under Picard's iteration for a continuous-time Gaussian channel with feedback, which may pave way for effective computation of the mutual information of such a channel, a long open problem in information theory.

Keywords: relative entropy, mutual information, Girsanov's theorem, stochastic differential equations, Picard's iteration.

1 Introduction

Let $T > 0$, and let $C[0, T]$ denote the space of all real-valued continuous functions over the interval $[0, T]$. Let f be a *non-anticipative* functional from $[0, T] \times C[0, T] \times C[0, T]$ to \mathbb{R} in the sense that, roughly speaking, with t interpreted as the present time, the value of $f(t, \varphi, \psi)$ only causally depends on the "history" of φ and ψ (see the rigorous definition in Section 2.1). And let g be a non-anticipative functional from $[0, T] \times C[0, T]$ to \mathbb{R} .

In this paper, we are concerned with a stochastic differential equation (SDE) taking the following form:

$$dY(t) = f(t, \xi, Y)dt + g(t, Y)dW(t), \quad 0 \leq t \leq T, \quad (1)$$

with the initial condition $Y(0) = 0$, where $W = \{W(t); 0 \leq t \leq T\}$ is a standard Brownian motion and $\xi = \{\xi(t); 0 \leq t \leq T\}$ is a stochastic process independent of W with $\mathbb{E}[\sup_{0 \leq t \leq T} |\xi(t)|^2] < \infty$.

For the special case that there is no actual dependence of f on the second parameter φ , it is well known (see, e.g., [11, 15]) that a convergence analysis of the corresponding Picard's iteration can establish the existence and uniqueness of the strong solution to the equation (1), provided that f and g satisfy the conventional uniform Lipschitz and linear growth conditions. It turns out that, for the general case (that f may actually depends on φ), the aforementioned conditions and convergence analysis can be adapted to achieve the same objective. More precisely, assume that f and g satisfy the following conditions:

(L) (Uniform Lipschitz) There exists $K > 0$ such that for any $t \in [0, T]$ and any $\varphi, \psi, \tilde{\psi} \in C[0, T]$,

$$|f(t, \varphi, \psi) - f(t, \varphi, \tilde{\psi})|^2 + |g(t, \psi) - g(t, \tilde{\psi})|^2 \leq K \sup_{0 \leq s \leq t} |\psi(s) - \tilde{\psi}(s)|^2. \quad (2)$$

(G) (Linear Growth) There exists $L > 0$ such that for any $t \in [0, T]$ and $\varphi, \psi \in \mathbb{R}$,

$$|f(t, \varphi, \psi)|^2 + |g(t, \psi)|^2 \leq L \left(1 + \sup_{0 \leq s \leq t} |\varphi(s)|^2 + \sup_{0 \leq s \leq t} |\psi(s)|^2 \right). \quad (3)$$

And consider the sequence $\{Y^{(n)}\} = \{Y^{(n)}(t); 0 \leq t \leq T\}_{n=0}^{\infty}$ recursively defined by

$$Y^{(n)}(t) = \int_0^t f(s, \xi, Y^{(n-1)}) ds + \int_0^t g(s, Y^{(n-1)}) dW(s), \quad \text{for } n \in \mathbb{N}, \quad (4)$$

starting with $Y^{(0)}(t) \equiv 0$. Then, as elaborated in Theorem 2.1, the sequence $\{Y^{(n)}\}$ converges in the L^2 and almost-sure sense, which yields the unique strong solution $Y = \{Y(t); 0 \leq t \leq T\}$ to the equation (1).

In this paper, we will further characterize the convergence behaviors of $\{Y^{(n)}\}$ in terms of the so-called *relative entropy*, a.k.a. *Kullback-Leibler divergence*. Roughly speaking, the relative entropy between two probability measures μ and ν , denoted by $\mathbb{D}(\mu||\nu)$, is a “distance-like” function measuring how “close” μ and ν are; and our main results show that as n tends to infinity, the distribution of $\{Y^{(n)}\}$ converges to that of Y with respect to certain measures in terms of relative entropy. More precisely, let $B = \{B(t); 0 \leq t \leq T\}$

be the strong solution to the following SDE:

$$dB(t) = g(t, B)dW(t), \quad 0 \leq t \leq T, \quad (5)$$

with the initial condition $B(0) = 0$. And let $\mu_{Y^{(n)}}$, μ_Y and μ_B denote the probability measures on the space of $C[0, T]$ induced by $Y^{(n)}$, Y and B , respectively. Then, our main result can be stated as follows:

Theorem 1.1. *Suppose that f and g satisfy **(L)** and **(G)** and that g satisfies the following regularity condition:*

$$\mathbf{(R)} \quad \sup_{(t, \varphi) \in [0, T] \times C[0, T]} |g^{-1}(t, \psi)| < \infty.$$

Then, we have

- 1) $\lim_{n \rightarrow \infty} \mathbb{D}(\mu_{Y^{(n)}} \| \mu_Y) = 0$;
- 2) $\lim_{n \rightarrow \infty} \mathbb{D}(\mu_{Y^{(n)}} \| \mu_B) = \mathbb{D}(\mu_Y \| \mu_B)$.

Though often intuited as a way of measuring the distance between two probability distributions, relative entropy does not qualify as a metric due to its asymmetry (with respect to its two parameters) and failure to obey the triangle inequality. Nevertheless, relative entropy is intimately related to many other metrics such as total variation distance, Fisher information divergence, Wasserstein distance and so on; see, e.g., [23, 24] and references therein. This, together with many other desirable properties, explains its widespread use in diverse disciplines including probability theory [9], statistics [12, 22], statistical physics [21, 7], machine learning and neural science [2, 1, 17, 14], and information theory [3, 8]. Complementing the convergence of Picard's iteration in the conventional L^2 and almost sure sense, relative entropy convergence as characterized in Theorem 1.1 may reveal aspects of SDEs unexplored as yet, which hold out the promise of new applications of SDEs in some uncharted research territories.

Roughly speaking, the well-known Pinsker's inequality [4] states that the relative entropy between two probability measures dominates their total variation distance. As an example of the possible applications of our results, we note that the statement 1) of Theorem 1.1, in combination with Pinsker's inequality, immediately leads to the following corollary, which does not seem to directly follow from any other known results.

Corollary 1.1. *Suppose f and g satisfy **(L)**, **(G)** and **(R)**. Then, $\mu_{Y^{(n)}}$ converges to μ_Y in total variation distance.*

For another example, the equation (1) is often used to model a continuous-time Gaussian channel with feedback in information theory. As a corollary of the statement 2) of

Theorem 1.1, we establish the convergence of the mutual information of such a channel under Picard's iteration in Theorem 4.1. Here, we note that this convergence result requires rather mild assumptions and more importantly feature a finer characterization of the convergence behavior, and thereby may pave way for effective computation of the mutual information of a continuous-time Gaussian channel with feedback, which is a long open problem in information theory.

The remainder of the paper is organized as follows. We start with Section 2, which will introduce some basic notions and results that are instrumental in our proofs. The proof of our main result will then be presented in Section 3. In Section 4, we state and prove Theorem 4.1, which exemplifies the possible applications of Theorem 1.1 in information theory.

2 Preliminaries

2.1 Notations

We denote by \mathbb{N} and \mathbb{N}_0 the sets of positive and non-negative integers respectively. For any $\varphi \in C[0, T]$, let $\|\varphi\|_T$ denote its sup-norm, that is, $\|\varphi\|_T := \sup_{0 \leq t \leq T} |\varphi(t)|$. We will equip the space $C[0, T]$ with the filtration $\{\mathcal{B}_t\}_{0 \leq t \leq T}$, where \mathcal{B}_T denotes the Borel σ -algebra on the space $C[0, T]$ and $\mathcal{B}_t = \pi_t^{-1}(\mathcal{B}_T)$, where $\pi_t : C[0, T] \rightarrow C[0, T]$ is given by the map $(\pi_t x)(s) = x(t \wedge s)$. We say that a functional f from $[0, T] \times C[0, T] \times C[0, T]$ to \mathbb{R} is *non-anticipative* if f is $\mathcal{B}([0, T]) \otimes \mathcal{B}_T \otimes \mathcal{B}_T$ -measurable and for each $t \in [0, T]$, $f(t, \cdot, \cdot)$ is $\mathcal{B}_t \otimes \mathcal{B}_t$ -measurable, or equivalently, for any $\varphi, \psi \in C[0, T]$,

$$f(t, \varphi, \psi) = f(t, \{\varphi(s); 0 \leq s \leq t\}, \{\psi(s); 0 \leq s \leq t\}).$$

We use $(\Omega, \mathcal{F}, \mathbb{P})$ to denote the underlying probability space equipped with a filtration $\mathcal{F} = \{\mathcal{F}_t\}_{0 \leq t \leq T}$, which, as is typical in the theory of SDEs, satisfies the *usual conditions* [11] and is rich enough to accommodate ξ and W . For any \mathcal{F} -adapted process $X = \{X(t), \mathcal{F}_t; 0 \leq t \leq T\}$, we use $\mathcal{F}^X = \{\mathcal{F}_t^X; 0 \leq t \leq T\}$ to denote the *augmented filtration* generated by X .

For any two probability measures μ and ν , we write $\mu \ll \nu$ to mean μ is absolutely continuous with respect to ν and $\mu \sim \nu$ if they are equivalent. For any two processes $X = \{X(t); 0 \leq t \leq T\}$ and $Y = \{Y(t); 0 \leq t \leq T\}$, as mentioned before, we use μ_X, μ_Y to denote the probability distributions on \mathcal{B}_T induced by X, Y , respectively; and we write the Radon-Nikodym derivative of μ_Y with respect to μ_X as $d\mu_Y/d\mu_X$.

Given two expressions A_1, A_2 , we use $A_1 \lesssim_M A_2$ as a shorthand for the inequality $A_1 \leq C_M A_2$ for some constant C_M depending only on M .

2.2 The L^2 and Almost Sure Convergence of $\{Y^{(n)}\}$

The following theorem establishes the convergence of $\{Y^{(n)}\}$ in the L^2 and almost sure sense, and thereby the existence and uniqueness of the strong solution to (1). The proof is a slight modification of the conventional argument and is included in Appendix A for completeness. Here, we emphasize that the constants specified in the theorem will be used to characterize the rate of convergence in Theorem 4.1.

Theorem 2.1. *Consider the equation (1) satisfying the conditions **(L)** and **(G)**. Then, we have*

1) for all $n \in \mathbb{N}_0$,

$$\mathbb{E} \left[\|Y^{(n)}\|_T^2 \right] \leq k_1 e^{k_2 T} \quad \text{and} \quad \mathbb{E} \left[\|Y^{(n+1)} - Y^{(n)}\|_T^2 \right] \leq c_1 \frac{(c_2 T)^n}{n!}, \quad (6)$$

where $k_1 = 2LT(T+4)(1 + \mathbb{E}[\|\xi\|_T^2])$, $k_2 = 2L(T+4)$, $c_1 = 2TL(T+4)(1 + \mathbb{E}[\|\xi\|_T^2])$ and $c_2 = 2K(T+4)$;

2) there exists a unique strong solution $Y = \{Y(t); 0 \leq t \leq T\}$ to (1) with

$$\lim_{n \rightarrow \infty} \|Y^{(n)} - Y\|_T^2 = 0 \quad \text{and} \quad \mathbb{E} \left[\|Y^{(n)} - Y\|_T^2 \right] \leq c_3 \frac{(c_2 T)^n}{n!} \xrightarrow{n \rightarrow \infty} 0, \quad (7)$$

where $c_3 = k_1 e^{k_2 T}$. So, the solution Y is the pointwise and L^2 -limit of the sequence $\{Y^{(n)}\}_{n=0}^\infty$.

2.3 Absolute Continuity between Relevant Measures

In this section, assuming $g \equiv 1$, we will discuss the absolute continuity between various measures and derive some formulas for the corresponding Radon-Nikodym derivatives which will be used in our proofs. For notational convenience, we may simply write $f(t, \xi, Y^{(n)})$ as $f^{(n)}(t)$, and $f(t, \xi, Y)$ as $f(t)$.

First of all, it is easy to see that the linear growth condition **(G)** guarantees that

$$\int_0^T |f(t, \varphi, \psi)| dt < \infty \quad (8)$$

for any $\varphi, \psi \in C[0, T]$, and

$$\mathbb{P} \left(\int_0^T f^2(t, \xi, Y) dt < \infty \right) = 1 \quad \text{and} \quad \mathbb{P} \left(\int_0^T \mathbb{E}^2 [f(t, \xi, Y) | \mathcal{F}_t^Y] dt < \infty \right) = 1. \quad (9)$$

Hence, by [13, Theorem 7.14 and Lemma 7.7], we have

$$\mu_Y \sim \mu_W \text{ and } \mu_{\xi, Y} \sim \mu_\xi \times \mu_W, \quad (10)$$

and moreover,

$$\frac{d\mu_Y}{d\mu_W}(Y) = \exp \left(\int_0^T \mathbb{E} [f(t, \xi, Y) | \mathcal{F}_t^Y] dY(t) - \frac{1}{2} \int_0^T \mathbb{E}^2 [f(t, \xi, Y) | \mathcal{F}_t^Y] dt \right) \quad (11)$$

and

$$\frac{d\mu_{\xi, Y}}{d(\mu_\xi \times \mu_W)}(\xi, Y) = \exp \left(\int_0^T f(t, \xi, Y) dY(t) - \frac{1}{2} \int_0^T f^2(t, \xi, Y) dt \right). \quad (12)$$

Similarly, it follows from **(G)** and (7) that for any $n \in \mathbb{N}_0$,

$$\mathbb{P} \left(\int_0^T f^2(t, \xi, Y^{(n)}) dt < \infty \right) = 1 \quad \text{and} \quad \mathbb{P} \left(\int_0^T \mathbb{E}^2 [f(t, \xi, Y^{(n)}) | \mathcal{F}_t^{Y^{(n+1)}}] dt < \infty \right) = 1. \quad (13)$$

Hence, by [13, Theorem 7.14], we have $\mu_{Y^{(n)}} \ll \mu_W$ for all $n \in \mathbb{N}$, and furthermore,

$$\begin{aligned} \frac{d\mu_{Y^{(n)}}}{d\mu_W}(Y^{(n)}) &= \exp \left(\int_0^T \alpha^{(n)}(t) dY^{(n)}(t) - \frac{1}{2} \int_0^T (\alpha^{(n)}(t))^2 dt \right) \\ &= \exp \left(\int_0^T \alpha^{(n)}(t) f^{(n-1)}(t) dW(t) + \int_0^T f^{(n-1)}(t) \alpha^{(n)}(t) dt - \frac{1}{2} \int_0^T (\alpha^{(n)}(t))^2 dt \right), \end{aligned} \quad (14)$$

where $\alpha^{(n)}(t) := \mathbb{E} [f(t, \xi, Y^{(n-1)}) | \mathcal{F}_t^{Y^{(n)}}]$, and, as mentioned before, $f^{(n)}(t)$ is a shorthand notation for $f(t, \xi, Y^{(n)})$. The following proposition, whose proof is postponed to Appendix B, says that an extra condition will guarantee that $\mu_{Y^{(n)}}$ and μ_W are equivalent.

Proposition 2.1. If f satisfies **(L)** and there exists an $\varepsilon > 0$ such that

$$\mathbb{E} \left[e^{\varepsilon \|\xi\|_T^2} \right] < \infty, \quad (15)$$

then $\mu_{Y^{(n)}} \sim \mu_W$.

We next show the absolute continuity of $\mu_{\xi, Y^{(n)}}$ with respect to $\mu_\xi \times \mu_W$ and derive the formula for $\frac{d\mu_{\xi, Y^{(n)}}}{d(\mu_\xi \times \mu_W)}(\xi, Y^{(n)})$, for which we will extend the results in [13, Lemmas 7.6 and 7.7] below. Consider the process $Z = \{Z(t); 0 \leq t \leq T\}$ defined by

$$Z(t) = \int_0^t h(s, \xi, W) ds + W(t), \quad 0 \leq t \leq T, \quad (16)$$

where $h : [0, T] \times C[0, T] \times C[0, T] \rightarrow \mathbb{R}$ is a *non-anticipative* functional, and ξ and W are as in (1). Then, the extension (see Appendix C for the proof) says

Proposition 2.2. Suppose that

$$\mathbb{E} \left[\int_0^T h^2(t, \xi, W) dt \right] < \infty, \text{ and for any } x, w \in C[0, T], \int_0^T |h(t, x, w)| dt < \infty. \quad (17)$$

Then, $\mu_{\xi, Z} \ll \mu_{\xi} \times \mu_Z$ and moreover,

$$\frac{d\mu_{\xi, Z}}{d(\mu_{\xi} \times \mu_W)}(\xi, Z) = \exp \left(\int_0^T \mathbb{E} \left[h(t, \xi, W) | \mathcal{F}_t^{Z, \xi} \right] dZ(t) - \frac{1}{2} \int_0^T \mathbb{E}^2 \left[h(t, \xi, W) | \mathcal{F}_t^{Z, \xi} \right] dt \right). \quad (18)$$

Note that an application of Proposition 2.2 with $h(t, \xi, W) = f(t, \xi, Y^{(n)})$ immediately yields that $\mu_{\xi, Y^{(n)}}$ is absolutely continuous with respect to $\mu_{\xi} \times \mu_W$, and moreover,

$$\frac{d\mu_{\xi, Y^{(n)}}}{d(\mu_{\xi} \times \mu_W)}(\xi, Y^{(n)}) = \exp \left(\int_0^T \beta^{(n)}(t) dY^{(n)}(t) - \frac{1}{2} \int_0^T (\beta^{(n)}(t))^2 dt \right), \quad (19)$$

where $\beta^{(n)}(t) := \mathbb{E} \left[f(t, \xi, Y^{(n-1)}) | \mathcal{F}_t^{\xi, Y^{(n)}} \right]$.

2.4 Relative Entropy

Definition 2.1 (Relative entropy). For two probability measures μ and ν on a same probability space, the *relative entropy* of μ with respect to ν is defined as

$$\mathbb{D}(\mu \| \nu) = \begin{cases} \int \log \frac{d\mu}{d\nu}(x) \mu(dx), & \text{if } \mu \ll \nu; \\ \infty & \text{otherwise,} \end{cases}$$

where $d\mu/d\nu(x)$ is the Radon-Nikodym derivative of μ with respect to ν .

It is well known that relative entropy $\mathbb{D}(\mu \| \nu)$ is convex in (μ, ν) . To be exact, for any $\alpha \in [0, 1]$ and measures μ_i, ν_i , $i = 1, 2$, we have

$$\mathbb{D}(\mu \| \nu) \leq \alpha \mathbb{D}(\mu_1 \| \nu_1) + (1 - \alpha) \mathbb{D}(\mu_2 \| \nu_2),$$

where $\mu = \alpha \mu_1 + (1 - \alpha) \mu_2$ and $\nu = \alpha \nu_1 + (1 - \alpha) \nu_2$. It turns out, as elaborated in the following lemma (see Appendix D for the proof), the convexity can be generalized to our setting, which is crucial in our proof of Theorem 1.1.

Proposition 2.3. Let U, V and Z be $C[0, T]$ -valued random variables defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose $\mu_U \ll \mu_V$. Then,

$$\mathbb{D}(\mu_U \| \mu_V) \leq \int_{C[0, T]} \mathbb{D}(\mu_{U, Z}(\cdot, z) \| \mu_{V, Z}(\cdot, z)) \mu_Z(dz). \quad (20)$$

3 Proof of Theorem 1.1

First of all, we show that the proof boils down to that for the special case $g \equiv 1$.

Consider the process $\tilde{Y} = \{\tilde{Y}(t); 0 \leq t \leq T\}$ satisfying the following SDE:

$$d\tilde{Y}(t) = \tilde{f}(t, \xi, \tilde{Y})dt + dW(t), \quad 0 \leq t \leq T, \quad (21)$$

with the initial condition $Y(0) = 0$, where $\tilde{f} := fg^{-1}$. It follows from **(L)**, **(G)** and **(R)** that \tilde{f} satisfies **(L)** and **(G)**, and hence the solution to (21) uniquely exists. Then, by [13, Theorem 7.19], we have

$$\frac{d\mu_Y}{d\mu_B}(y) = \frac{d\mu_{\tilde{Y}}}{d\mu_W}(y) \quad \text{and} \quad \frac{d\mu_{\xi, Y}}{d(\mu_\xi \times \mu_B)}(x, y) = \frac{d\mu_{\xi, \tilde{Y}}}{d(\mu_\xi \times \mu_W)}(x, y), \quad (22)$$

where $B = \{B(t); 0 \leq t \leq T\}$ is the process defined in (5). Note that the Picard's iteration corresponding to (21) recursively computes: for any $n \in \mathbb{N}$,

$$\tilde{Y}^{(n)}(t) = \int_0^t \tilde{f}(s, \xi, \tilde{Y}^{(n-1)}) ds + W(t), \quad (23)$$

starting with $\tilde{Y}^{(0)} \equiv 0$. Therefore, following the proofs of [13, Theorems 7.14, 7.18 and 7.19], we have

$$\frac{d\mu_{Y^{(n)}}}{d\mu_B}(y) = \frac{d\mu_{\tilde{Y}^{(n)}}}{d\mu_W}(y) \quad \text{and} \quad \frac{d\mu_{\xi, Y^{(n)}}}{d(\mu_\xi \times \mu_B)}(x, y) = \frac{d\mu_{\xi, \tilde{Y}^{(n)}}}{d(\mu_\xi \times \mu_W)}(x, y). \quad (24)$$

It then immediately follows from (22) and (24) that Theorem 1.1 established for the special case $g \equiv 1$ implies that for the general case. So, throughout the remainder of the proof, we will assume that $g \equiv 1$.

3.1 Some Useful Relations

For each $x \in C[0, T]$, consider the process $Y^x = \{Y^x(t); 0 \leq t \leq T\}$ satisfying the following SDE:

$$dY^x(t) = f(t, x, Y^x)dt + dW(t), \quad 0 \leq t \leq T, \quad (25)$$

with the initial condition $Y^x(0) = 0$. The corresponding Picard's iteration recursively computes: for any $n \in \mathbb{N}$,

$$Y^{x, (n)}(t) = \int_0^t f(s, x, Y^{x, (n-1)}) ds + W(t), \quad (26)$$

starting with $Y^{(0)}(t) \equiv 0$. As discussed in Section 2.3, it holds that $\mu_{Y^x} \sim \mu_W$ and furthermore

$$\frac{d\mu_{Y^x}}{d\mu_W}(Y^x) = \exp\left(\int_0^T f(t, x, Y^x) dY^x(t) - \frac{1}{2} \int_0^T f^2(t, x, Y^x) dt\right). \quad (27)$$

We now claim that, for any $n \in \mathbb{N}_0$, there exists a non-anticipative functional $Q^{(n)} : [0, T] \times C[0, T] \times C[0, T] \rightarrow \mathbb{R}$ such that for any $x \in C[0, T]$,

$$\mathbb{P}(Y^{(n)}(t) = Q^{(n)}(t, \xi, W); 0 \leq t \leq T) = \mathbb{P}(Y^{x, (n)}(t) = Q^{(n)}(t, x, W); 0 \leq t \leq T) = 1. \quad (28)$$

Indeed, for $n = 0$, one can simply set $Q^{(0)} \equiv 0$. By way of induction, suppose that (28) is satisfied by $Q^{(n)}$ for a particular n . Then, we have

$$\begin{aligned} Y^{x, (n+1)}(t) &= \int_0^t f(s, x, Y^{x, (n)}) ds + W(t) \\ &= \int_0^t f(s, x, Q^{(n)}(\cdot, x, W)) ds + W(t). \end{aligned} \quad (29)$$

Setting $Q^{(n+1)}(t, x, y) := \int_0^t f(s, x, Q^{(n)}(\cdot, x, y)) ds + y(t)$, one readily verifies that for any $t \in [0, T]$,

$$Y^{(n+1)}(t) = Q^{(n+1)}(t, \xi, W) \text{ and } Y^{x, (n+1)}(t) = Q^{(n+1)}(t, x, W) \text{ } \mathbb{P}\text{-a.s.}$$

Since the above processes admit continuous sample paths, (28) holds true for $Q^{(n+1)}$, which in turn yields the claim.

Note that when ξ is fixed as x , (15) is trivially satisfied, and therefore, by Proposition 2.1, it holds that $\mu_{Y^{x, (n)}} \sim \mu_W$, and, by Proposition 2.2,

$$\frac{d\mu_{Y^{x, (n)}}}{d\mu_W}(Y^{x, (n)}) = \exp\left(\int_0^T \Phi^{(n)}(t, Y^{x, (n)}) dY^{x, (n)}(t) - \frac{1}{2} \int_0^T (\Phi^{(n)}(t, Y^{x, (n)}))^2 dt\right), \quad (30)$$

where $\Phi^{(n)} : [0, T] \times C[0, T] \rightarrow \mathbb{R}$ is a non-anticipative functional such that for any $t \in [0, T]$,

$$\Phi^{(n)}(t, Y^{x, (n)}) = \mathbb{E}\left[f(t, x, Y^{x, (n-1)}) \mid \mathcal{F}_t^{Y^{x, (n)}}\right] \text{ } \mathbb{P}\text{-a.s.} \quad (31)$$

Finally, mimicking the proof of [13, Lemma 7.6] with the representation as in (28), we can show that for all $x \in C[0, T]$ and $n \in \mathbb{N}_0$,

$$\frac{d\mu_{\xi, Y^{(n)}}}{d(\mu_\xi \times \mu_W)}(x, Y^{(n)}) = \frac{d\mu_{Y^{x, (n)}}}{d\mu_W}(Y^{(n)}) \text{ } \mathbb{P}\text{-a.s.}, \quad (32)$$

and

$$\frac{d\mu_{\xi,Y}}{d(\mu_{\xi} \times \mu_W)}(x, Y) = \frac{d\mu_{Y^x}}{d\mu_W}(Y) \quad \mathbb{P}\text{-a.s.} \quad (33)$$

And moreover, by [13, Lemma 7.7], we have the following relations

$$\frac{d\mu_{Y^{(n)}}}{d\mu_W}(Y^{(n)}) = \int_{C[0,T]} \frac{d\mu_{\xi,Y^{(n)}}}{d(\mu_{\xi} \times \mu_W)}(x, Y^{(n)}) \mu_{\xi}(dx) \quad \mathbb{P}\text{-a.s.}, \quad (34)$$

and

$$\frac{d\mu_Y}{d\mu_W}(Y) = \int_{C[0,T]} \frac{d\mu_{\xi,Y}}{d(\mu_{\xi} \times \mu_W)}(x, Y) \mu_{\xi}(dx) \quad \mathbb{P}\text{-a.s.} \quad (35)$$

3.2 Proof of 1)

Applying Lemma 2.3 with $U = Y^{(n)}$, $V = Y$, $Z = \xi$ and $\tilde{V} = W$, we arrive at

$$\begin{aligned} \mathbb{D}(\mu_{Y^{(n)}} \| \mu_Y) &\leq \int_{C[0,T]} \mathbb{D}(\mu_{Y^{(n)},\xi}(\cdot, x) \| \mu_{Y,\xi}(\cdot, x)) \mu_{\xi}(dx) \\ &= \int_{C[0,T]} \left(\mathbb{E} \left[\log \frac{d\mu_{\xi,Y^{(n)}}}{d(\mu_{\xi} \times \mu_W)}(x, Y^{(n)}) \right] - \mathbb{E} \left[\log \frac{d\mu_{\xi,Y}}{d(\mu_{\xi} \times \mu_W)}(x, Y) \right] \right) \mu_{\xi}(dx) \\ &= \int_{C[0,T]} \left(\mathbb{E} \left[\log \frac{d\mu_{Y^x,(n)}}{d\mu_W}(Y^{(n)}) \right] - \mathbb{E} \left[\log \frac{d\mu_{Y^x}}{d\mu_W}(Y^{(n)}) \right] \right) \mu_{\xi}(dx), \end{aligned}$$

where for the last equality we have used the relations (32) and (33).

As discussed in Sections 2.3 and 3.1, $\mu_{Y^x,(n)} \sim \mu_W$ for any fixed $x \in C[0, T]$ and $n \in \mathbb{N}_0$, and $\mu_{Y^{(n)}} \ll \mu_W$ for every $n \in \mathbb{N}_0$. It then immediately follows that $\mu_{Y^{(n)}} \ll \mu_{Y^x,(n)}$, which, together with [13, Lemma 4.10], implies that

$$\frac{d\mu_{Y^x,(n)}}{d\mu_W}(Y^{(n)}) = \exp \left(\int_0^T \Phi_t^{(n)}(Y^{(n)}) dY^{(n)}(t) - \frac{1}{2} \int_0^T \left(\Phi_t^{(n)}(Y^{(n)}) \right)^2 dt \right).$$

In a parallel fashion, we deduce that $\mu_{Y^{(n)}} \ll \mu_{Y^x}$, and moreover,

$$\frac{d\mu_{Y^x}}{d\mu_W}(Y^{(n)}) = \exp \left(\int_0^T f(t, x, Y^{(n)}) dY^{(n)}(t) - \frac{1}{2} \int_0^T f^2(t, x, Y^{(n)}) dt \right).$$

With $\Phi^{(n)}$ as defined in (31), we now show that, for each $t \in [0, T]$,

$$\Phi^{(n)}(t, Y^{(n)}) = \mathbb{E} \left[f(t, x, Y^{(n-1)}) | \mathcal{F}_t^{Y^{(n)}} \right] \quad \mathbb{P}\text{-a.s.} \quad (36)$$

To see this, recall that from Section 3.1 that for each $n \in \mathbb{N}_0$, there exists a non-anticipative functional $Q(n)$ satisfying (28). Then, for any bounded $\mathcal{F}_t^{Y^{(n)}}$ -measurable random variable

$\lambda = \lambda(Y^{(n)}(t))$, we have

$$\begin{aligned}\mathbb{E} [\Phi^{(n)}(t, Y^{(n)}) \lambda(Y^{(n)}(t))] &= \int_{C[0,T]} \left(\int_{C[0,T]} \Phi^{(n)}(t, Q^{(n)}(\cdot, x, w)) \lambda(Q^{(n)}(\cdot, x, w)) \mu_W(dw) \right) \mu_\xi(dx) \\ &= \int_{C[0,T]} \int_{C[0,T]} f(t, x, Q^{(n)}(\cdot, x, w)) \lambda(Q^{(n)}(\cdot, x, w)) \mu_W(dw) \times \mu_\xi(dx) \\ &= \mathbb{E} [f(t, x, Y^{(n)}) \lambda(Y^{(n)}(t))],\end{aligned}$$

where the second equality follows from (31). The claimed (36) then immediately follows.

Then, using (36) and rewriting $f(t, x, Y^{(n)})$ as $f_x^{(n)}(t)$, we have

$$\begin{aligned}\mathbb{D}(\mu_{Y^{(n)}} \|\mu_Y) &\leq \int_{C[0,T]} \mathbb{E} \left[\int_0^T \mathbb{E} [f_x^{(n-1)}(t) - f_x^{(n)}(t) | \mathcal{F}_t^{Y^{(n)}}] dY^{(n)}(t) \right. \\ &\quad \left. - \frac{1}{2} \int_0^T \left(\mathbb{E}^2 [f_x^{(n-1)}(t) | \mathcal{F}_t^{Y^{(n)}}] - (f_x^{(n)}(t))^2 \right) dt \right] \mu_\xi(dx) \\ &= \int_{C[0,T]} \left(\mathbb{E} \left[\int_0^T \beta_x^{(n)}(t) dW(t) \right] + \frac{1}{2} \mathbb{E} \left[\int_0^T (\beta_x^{(n)}(t))^2 dt \right] \right) \mu_\xi(dx),\end{aligned}\quad (37)$$

where $\beta_x^{(n)}(t) = \mathbb{E} [f_x^{(n-1)}(t) - f_x^{(n)}(t) | \mathcal{F}_t^{Y^{(n)}}]$. Now, by Jensen's inequality and the Itô isometry, we infer that, for the first term in (37),

$$\begin{aligned}\mathbb{E} \left[\int_0^T \beta_x^{(n)}(t) dW(t) \right] &\leq \mathbb{E}^{\frac{1}{2}} \left[\int_0^T (\beta_x^{(n)}(t))^2 dt \right] \\ &\leq \sqrt{\int_0^T \mathbb{E} \left[(f_x^{(n-1)}(t) - f_x^{(n)}(t))^2 \right] dt} \\ &\leq \sqrt{K \int_0^T \mathbb{E} [\|Y^{(n-1)} - Y^{(n)}\|_t^2] dt} \\ &\leq (c_1 KT)^{\frac{1}{2}} \sqrt{\frac{(c_2 T)^{n-1}}{(n-1)!}},\end{aligned}\quad (38)$$

where c_1 and c_2 are the constants as in Theorem 2.1. Similarly, for the second term in (37), we have

$$\mathbb{E} \left[\int_0^T (\beta_x^{(n)}(t))^2 dt \right] \leq c_1 KT \frac{(c_2 T)^{n-1}}{(n-1)!}.\quad (39)$$

With the two bounds in (38) and (39), we eventually reach

$$\mathbb{D}(\mu_{Y^{(n)}} \|\mu_Y) \leq (c_1 KT)^{\frac{1}{2}} \sqrt{\frac{(c_2 T)^{n-1}}{(n-1)!}} + c_1 KT \frac{(c_2 T)^{n-1}}{2(n-1)!},\quad (40)$$

which in turn implies 1) of Theorem 1.1.

3.3 Proof of 2)

We will need the following lemma, whose proof is postponed to Appendix E.

Lemma 3.1. Assume **(L)**, **(G)** and the following condition:

(B) (Uniform Boundedness) There exists $M > 0$ such that for all $\varphi, \psi \in C[0, T]$,

$$\int_0^T f^2(t, \phi, \varphi) dt \leq M. \quad (41)$$

Then, for any $n \in \mathbb{N}_0$ and p with $|p| \geq 1$,

$$\mathbb{E} \left[\left(\frac{d\mu_{Y^{(n)}}}{d\mu_W} (Y^{(n)}) \right)^p \right] \leq e^{\frac{p(p+1)M}{2}}. \quad (42)$$

Moreover, for any $x \in C[0, T]$, we have

$$\mathbb{E} \left[\left(\frac{d\mu_{Y^{x,(n)}}}{d\mu_W} (Y^{x,(n)}) \right)^p \right] \leq e^{\frac{p(p+1)M}{2}}. \quad (43)$$

Now, we are ready for the proof of 2), which consists of the following two steps:

Step 1. In this step, we establish 2) with an extra condition **(B)**.

First of all, we note that for each $n \in \mathbb{N}_0$, it can be readily verified that

$$\mathbb{D}(\mu_{Y^{(n)}} \| \mu_W) - \mathbb{D}(\mu_Y \| \mu_W) = \mathbb{D}(\mu_{Y^{(n)}} \| \mu_Y) + \mathbb{E} \left[\log \frac{d\mu_Y}{d\mu_W} (Y^{(n)}) \right] - \mathbb{E} \left[\log \frac{d\mu_Y}{d\mu_W} (Y) \right]. \quad (44)$$

Since it has been established in 1) that $\lim_{n \rightarrow \infty} \mathbb{D}(\mu_{Y^{(n)}} \| \mu_Y) = 0$, to prove (2), it suffices to show that

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\log \frac{d\mu_Y}{d\mu_W} (Y^{(n)}) \right] = \mathbb{E} \left[\log \frac{d\mu_Y}{d\mu_W} (Y) \right]. \quad (45)$$

Now, applying Jensen's inequality and Fubini's theorem, we have, for any $p \geq 1$,

$$\mathbb{E} \left[\left| \frac{d\mu_Y}{d\mu_W} (Y^{(n)}) - \frac{d\mu_Y}{d\mu_W} (Y) \right|^p \right] \leq \int_{C[0, T]} \mathbb{E} \left[\left| \frac{d\mu_{Y^x}}{d\mu_W} (Y^{(n)}) - \frac{d\mu_{Y^x}}{d\mu_W} (Y) \right|^p \right] \mu_\xi(dx). \quad (46)$$

Notice that as discussed before, $\mu_{Y^{(n)}} \ll \mu_W$ for all $n \in \mathbb{N}_0$, $\mu_Y \sim \mu_W$ and $\mu_{Y^x} \sim \mu_W$, we have $\mu_{Y^{(n)}} \ll \mu_{Y^x} \sim \mu_Y$, which, together with [13, Lemma 4.10], implies that

$$\frac{d\mu_{Y^x}}{d\mu_W} (Y^{(n)}) = \exp \left(\int_0^T f(t, x, Y^{(n)}) dY^{(n)}(t) - \frac{1}{2} \int_0^T f^2(t, x, Y^{(n)}) dt \right), \quad \mathbb{P}\text{-a.s.}, \quad (47)$$

and moreover,

$$\frac{d\mu_{Y^x}}{d\mu_W}(Y) = \exp\left(\int_0^T f(t, x, Y) dY(t) - \frac{1}{2} \int_0^T f^2(t, x, Y) dt\right), \quad \mathbb{P}\text{-a.s.} \quad (48)$$

It then follows from (47), (48), the condition **(B)** and the Cauchy-Schwarz inequality that

$$\begin{aligned} & \mathbb{E} \left[\left| \log\left(\frac{d\mu_{Y^x}}{d\mu_W}(Y^{(n)})\right) - \log\left(\frac{d\mu_{Y^x}}{d\mu_W}(Y)\right) \right|^2 \right] \\ & \leq 2 \mathbb{E} \left[\left| \int_0^T \left(f_x^{(n)}(t) f_x^{(n-1)}(t) - \frac{1}{2} \left(f_x^{(n)}(t) \right)^2 - \frac{1}{2} f_x^2(t) \right) dt \right|^2 \right] \\ & \quad + 2 \mathbb{E} \left[\left| \int_0^T \left(f_x^{(n)}(t) - f_x(t) \right) dW(t) \right|^2 \right] \\ & \leq 18 \mathbb{E} \left[\left| \int_0^T f_x^{(n)}(t) \left(f_x^{(n-1)}(t) - f_x(t) \right) dt \right|^2 \right] + 9 \mathbb{E} \left[\left| \int_0^T f_x^{(n)}(t) \left(f_x(t) - f_x^{(n)}(t) \right) dt \right|^2 \right] \\ & \quad + 9 \mathbb{E} \left[\int_0^T \left| f_x(t) \left(f_x^{(n)}(t) - f_x(t) \right) dt \right|^2 \right] + 2 \mathbb{E} \left[\int_0^T \left(f_x^{(n)}(t) - f_x(t) \right)^2 dt \right] \\ & \leq 18M \left(\mathbb{E} \left[\int_0^T \left(f_x^{(n-1)}(t) - f_x(t) \right)^2 dt \right] + \mathbb{E} \left[\int_0^T \left(f_x(t) - f_x^{(n)}(t) \right)^2 dt \right] \right) \\ & \quad + 2 \mathbb{E} \left[\int_0^T \left(f_x^{(n)}(t) - f_x(t) \right)^2 dt \right] \\ & \lesssim_{M,K,L} \frac{(c_2 T)^{n-1}}{(n-1)!} + \frac{(c_2 T)^n}{n!}, \end{aligned} \quad (49)$$

where $c_2 = 2KT$, the constants M, K, L in the last inequality does not depend on x (cf. Theorem 2.1 and Remark A.1), and we have rewritten $f(t, x, Y^{(n)})$, $f(t, x, Y)$ as $f_x^{(n)}(t)$, $f_x(t)$, respectively, for notational simplicity.

Note that it can be verified that for any $p \geq 1$ and any $x, y \in \mathbb{R}$,

$$|e^x - e^y|^p \leq |e^x - e^y|^{p-1} (e^x \vee e^y) |x - y|. \quad (50)$$

Using this, the Cauchy-Schwarz inequality and Proposition 3.1, we have the following estimate

$$\begin{aligned} \mathbb{E} \left[\left| \frac{d\mu_{Y^x}}{d\mu_W}(Y^{(n)}) - \frac{d\mu_{Y^x}}{d\mu_W}(Y) \right|^p \right] & \lesssim_M \sqrt{\mathbb{E} \left[\left| \log\left(\frac{d\mu_{Y^x}}{d\mu_W}(Y^{(n)})\right) - \log\left(\frac{d\mu_{Y^x}}{d\mu_W}(Y)\right) \right|^2 \right]} \\ & \lesssim_{M,K,L,p} \sqrt{\frac{(c_2 T)^{n-1}}{(n-1)!} + \frac{(c_2 T)^n}{n!}}. \end{aligned} \quad (51)$$

Since the bound in (51) does not depend on x , we infer from (46) that

$$\mathbb{E} \left[\left| \frac{d\mu_Y}{d\mu_W}(Y^{(n)}) - \frac{d\mu_Y}{d\mu_W}(Y) \right|^p \right] \lesssim_{M,K,L,p} \sqrt{\frac{(c_2 T)^{n-1}}{(n-1)!} + \frac{(c_2 T)^n}{n!}}. \quad (52)$$

Now, using the well-known inequality $|\log x| \leq |x-1| + |1/x - 1|$, Proposition 3.1, Hölder's inequality and (52), we arrive at

$$\begin{aligned} & \mathbb{E} \left[\left| \log \left(\frac{d\mu_Y}{d\mu_W}(Y^{(n)}) \right) - \log \left(\frac{d\mu_Y}{d\mu_W}(Y) \right) \right| \right] \\ & \leq \mathbb{E} \left[\left(\left(\frac{d\mu_Y}{d\mu_W}(Y) \right)^{-1} + \left(\frac{d\mu_Y}{d\mu_W}(Y^{(n)}) \right)^{-1} \right) \left| \frac{d\mu_Y}{d\mu_W}(Y^{(n)}) - \frac{d\mu_Y}{d\mu_W}(Y) \right| \right] \\ & \lesssim_{M,p} \mathbb{E}^{\frac{1}{p}} \left[\left| \frac{d\mu_Y}{d\mu_W}(Y^{(n)}) - \frac{d\mu_Y}{d\mu_W}(Y) \right|^p \right] \\ & \lesssim_{M,K,L,p} \left(\frac{(c_2 T)^n}{n!} + \frac{(c_2 T)^{n-1}}{(n-1)!} \right)^{\frac{1}{2p}}. \end{aligned} \quad (53)$$

Therefore, by (44), (40) and (53), we reach

$$|\mathbb{D}(\mu_{Y^{(n)}} \| W) - \mathbb{D}(\mu_Y \| \mu_W)| \lesssim_{M,K,L,p} \left(\frac{(c_2 T)^{n-1}}{(n-1)!} + \frac{(c_2 T)^n}{n!} \right)^{\frac{1}{2p}} + \sqrt{\frac{(c_2 T)^{n-1}}{(n-1)!} + \frac{(c_2 T)^n}{n!}}, \quad (54)$$

which converges to 0 as n tends to infinity, and thereby completing **Step 1**.

Remark 3.1. At the expense of a possible increase of the corresponding constant, the rate of convergence as in (54) can be made arbitrarily close to $\mathcal{O} \left(\sqrt{\frac{(c_2 T)^{n-1}}{(n-1)!}} \right)$ by choosing p close enough to 1.

Step 2. In this step, we finish the proof of 2) without the condition **(B)**.

First of all, for each $n, m \in \mathbb{N}$, we define the stopping times $\tau_m^{(n)}$ and τ_m by

$$\tau_m^{(n)} := \begin{cases} \inf\{t \leq T : \int_0^t f^2(s, \xi, Y^{(n)}) ds \geq m\}, & \text{for } \int_0^T f^2(t, \xi, Y^{(n)}) dt \geq m; \\ T, & \text{otherwise,} \end{cases} \quad (55)$$

and

$$\tau_m := \begin{cases} \inf\{t \leq T : \int_0^t f^2(s, \xi, Y) ds \geq m\}, & \text{for } \int_0^T f^2(t, \xi, Y) dt \geq m; \\ T, & \text{otherwise.} \end{cases} \quad (56)$$

Furthermore, we define the stopping time σ_m by

$$\sigma_m := \inf_{n \geq 0} \{\tau_m^{(n)}\}. \quad (57)$$

For each fixed $n \in \mathbb{N}_0$, it is easy to see that $\tau_m^{(n)} \uparrow T$ \mathbb{P} -a.s. as $m \rightarrow \infty$; and moreover, by the easily verifiable fact that

$$\lim_{n \rightarrow \infty} \int_0^T f^2(t, \xi, Y^{(n)}) dt = \int_0^T f^2(t, \xi, Y) dt,$$

we infer that $\sigma_m \uparrow T$ \mathbb{P} -a.s. as $m \rightarrow \infty$. Now, we are ready to define truncated versions of f and Y as below:

$$f_{(m)}(t, \phi, \varphi) := f(t, \phi, \varphi) \mathbb{1}_{\{\int_0^t f^2(s, \phi, \varphi) ds \leq m\}}, \quad (58)$$

and

$$Y_{(m)}(t) := \int_0^t f_{(m)}(s, \xi, Y) ds + W(t), \quad 0 \leq t \leq T. \quad (59)$$

Since for any $t \in [0, T]$, $Y_{(m)}(t) \mathbb{1}_{\{t \leq \tau_m\}} = Y(t) \mathbb{1}_{\{t \leq \tau_m\}}$ \mathbb{P} -a.s., and both the processes $\{Y_{(m)}(t) \mathbb{1}_{\{t \leq \tau_m\}}; 0 \leq t \leq T\}$ and $\{Y(t) \mathbb{1}_{\{t \leq \tau_m\}}\}$ admit left-continuous sample paths, they are indistinguishable to each other. Consequently, we have

$$\mathbb{P} \left(\int_0^t f_{(m)}(s, \xi, Y) ds = \int_0^t f_{(m)}(s, \xi, Y_{(m)}) ds; \quad 0 \leq t \leq T \right) = 1, \quad (60)$$

and moreover,

$$Y_{(m)}(t) = \int_0^t f_{(m)}(s, \xi, Y_{(m)}) ds + W(t), \quad 0 \leq t \leq T. \quad (61)$$

Note that, for any fixed $m \in \mathbb{N}$, the corresponding Picard's iteration $\{Y_{(m)}^{(n)}\}_{n=0}^{\infty}$ recursively computes: for any $n \in \mathbb{N}$,

$$Y_{(m)}^{(n)}(t) = \int_0^t f_{(m)}(s, \xi, Y_{(m)}^{(n-1)}) ds + W(t), \quad (62)$$

starting with $Y_{(m)}^{(0)}(t) \equiv 0$. By way of induction, we can show that, for all $n, m \in \mathbb{N}_0$ and $t \in [0, T]$,

$$Y_{(m)}^{(n)}(t) \mathbb{1}_{\{t \leq \sigma_m\}} = Y^{(n)}(t) \mathbb{1}_{\{t \leq \sigma_m\}} \quad \mathbb{P}\text{-a.s.} \quad \text{and} \quad Y_{(m)}(t) \mathbb{1}_{\{t \leq \sigma_m\}} = Y(t) \mathbb{1}_{\{t \leq \sigma_m\}} \quad \mathbb{P}\text{-a.s.}, \quad (63)$$

i.e., the process $\{Y_{(m)}^{(n)}(t) \mathbb{1}_{\{t \leq \sigma_m\}}; 0 \leq t \leq T\}$ is a modification of $\{Y^{(n)}(t) \mathbb{1}_{\{t \leq \sigma_m\}}; 0 \leq t \leq T\}$. Again, since both of them admit left-continuous sample paths, they are indistinguishable

to each other. And a similar argument can be applied to establish the indistinguishability between the two processes $\{Y_{(m)}(t) \mathbb{1}_{\{t \leq \sigma_m\}}; 0 \leq t \leq T\}$ and $\{Y(t) \mathbb{1}_{\{t \leq \sigma_m\}}; 0 \leq t \leq T\}$. Hence, when restricted to the event $\{\sigma_m = T\}$, the processes $Y_{(m)}^{(n)}$ and $Y_{(m)}$ share the same laws as $Y^{(n)}$ and Y , respectively.

With the above preparations, we are now ready to establish the desired convergence in 2). Splitting $\mathbb{E} \left[\log \frac{d\mu_Y}{d\mu_W} (Y) \right]$ with respect to the two non-overlapping events $\{\sigma_m = T\}$ and $\{\sigma_m < T\}$, we have

$$\begin{aligned} \mathbb{E} \left[\log \frac{d\mu_Y}{d\mu_W} (Y) \right] &= \mathbb{E} \left[\log \frac{d\mu_Y}{d\mu_W} (Y) (\mathbb{1}_{\{\sigma_m = T\}} + \mathbb{1}_{\{\sigma_m < T\}}) \right] \\ &= \mathbb{E} \left[\log \frac{d\mu_{Y_{(m)}}}{d\mu_W} (Y_{(m)}) \right] + \mathbb{E} \left[\log \frac{d\mu_Y}{d\mu_W} (Y) \mathbb{1}_{\{\sigma_m < T\}} \right]. \end{aligned} \quad (64)$$

Similarly, for each $n, m \in \mathbb{N}$, by (63), we have

$$\mathbb{E} \left[\log \left(\frac{d\mu_{Y^{(n)}}}{d\mu_W} (Y^{(n)}) \right) \right] = \mathbb{E} \left[\log \left(\frac{d\mu_{Y_{(m)}^{(n)}}}{d\mu_W} (Y_{(m)}^{(n)}) \right) \right] + \mathbb{E} \left[\log \left(\frac{d\mu_{Y^{(n)}}}{d\mu_W} (Y^{(n)}) \right) \mathbb{1}_{\{\sigma_m < T\}} \right]. \quad (65)$$

Note that by **Step 1**, we have

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\log \left(\frac{d\mu_{Y_{(m)}^{(n)}}}{d\mu_W} (Y_{(m)}^{(n+1)}) \right) \right] = \mathbb{E} \left[\log \frac{d\mu_{Y_{(m)}}}{d\mu_W} (Y_{(m)}) \right]. \quad (66)$$

And, by (64), (65) and (66), we have

$$\begin{aligned} &\mathbb{E} \left[\log \left(\frac{d\mu_Y}{d\mu_W} (Y) \right) \right] - \lim_{n \rightarrow \infty} \mathbb{E} \left[\log \left(\frac{d\mu_{Y^{(n)}}}{d\mu_W} (Y^{(n)}) \right) \right] \\ &= \mathbb{E} \left[\log \left(\frac{d\mu_Y}{d\mu_W} (Y) \right) \mathbb{1}_{\{\sigma_m < T\}} \right] - \lim_{n \rightarrow \infty} \mathbb{E} \left[\log \left(\frac{d\mu_{Y^{(n)}}}{d\mu_W} (Y^{(n)}) \right) \mathbb{1}_{\{\sigma_m < T\}} \right]. \end{aligned} \quad (67)$$

Noticing that (67) holds true for all $m \geq 0$ and using the fact that $\mathbb{1}_{\{\sigma_m < T\}} \downarrow 0$ as $m \rightarrow \infty$, we deduce that

$$\lim_{m \rightarrow \infty} \mathbb{E} \left[\log \left(\frac{d\mu_Y}{d\mu_W} (Y) \right) \mathbb{1}_{\{\sigma_m < T\}} \right] = 0, \quad (68)$$

where we have applied the dominated convergence theorem. The proof is then complete if we can show

$$\lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{E} \left[\log \left(\frac{d\mu_{Y^{(n)}}}{d\mu_W} (Y^{(n)}) \right) \mathbb{1}_{\{\sigma_m < T\}} \right] = 0. \quad (69)$$

To this end, we note that by (14),

$$\begin{aligned} & \mathbb{E} \left[\log \left(\frac{d\mu_{Y^{(n)}}}{d\mu_W} (Y^{(n)}) \right) \mathbb{1}_{\{\sigma_m < T\}} \right] = \\ & \mathbb{E} \left[\left(\int_0^T \alpha^{(n)}(t) dW(t) - \frac{1}{2} \int_0^T (\alpha^{(n)}(t))^2 dt + \int_0^T f^{(n-1)}(t) \alpha^{(n)}(t) dt \right) \mathbb{1}_{\{\sigma_m < T\}} \right], \end{aligned} \quad (70)$$

where $\alpha^{(n)}(t) := \mathbb{E} \left[f(t, \xi, Y^{(n-1)}) \mid \mathcal{F}_t^{Y^{(n)}} \right]$ and $f^{(n)}(t)$ represents $f(t, \xi, Y^{(n)})$. By the Itô isometry, the Cauchy-Schwarz inequality and Jensen's inequality, we have

$$\begin{aligned} \mathbb{E} \left[\left| \int_0^T \alpha^{(n)}(t) dW(t) \mathbb{1}_{\{\sigma_m < T\}} \right| \right] & \leq \sqrt{\mathbb{E} \left[\int_0^T (\alpha^{(n)}(t))^2 dt \right] \mathbb{P}(\sigma_m < T)} \\ & \leq \sqrt{\mathbb{E} \left[\int_0^T (f^{(n-1)}(t))^2 dt \right] \mathbb{P}(\sigma_m < T)}. \end{aligned} \quad (71)$$

Since $\int_0^T (f^{(n)}(t))^2 dt \xrightarrow{L^1} \int_0^T f^2(t) dt$ as $n \rightarrow \infty$, where $f(t)$ represents $f(t, \xi, Y)$, we have

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\left| \int_0^T \alpha^{(n)}(t) dW(t) \mathbb{1}_{\{\sigma_m < T\}} \right| \right] \leq \sqrt{\mathbb{E} \left[\int_0^T f^2(t) dt \right] \mathbb{P}(\sigma_m < T)}, \quad (72)$$

and therefore,

$$\lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{E} \left[\int_0^T \alpha^{(n)}(t) dW(t) \mathbb{1}_{\{\sigma_m < T\}} \right] = 0. \quad (73)$$

Using a similar argument, we deduce that

$$\begin{aligned} \mathbb{E} \left[\left| \int_0^T f^{(n-1)}(t) \alpha^{(n)}(t) dt \mathbb{1}_{\{\sigma_m < T\}} \right| \right] & \leq \mathbb{E} \left[\left(\int_0^T (f^{(n-1)}(t))^2 dt \right)^{1/2} \left(\int_0^T (\alpha^{(n)}(t))^2 dt \right)^{1/2} \mathbb{1}_{\{\sigma_m < T\}} \right] \\ & \leq \sqrt{\mathbb{E} \left[\int_0^T (f^{(n-1)}(t))^2 dt \mathbb{1}_{\{\sigma_m < T\}} \right] \mathbb{E} \left[\int_0^T (\alpha^{(n)}(t))^2 dt \right]} \\ & \leq \sqrt{\mathbb{E} \left[\int_0^T (f^{(n-1)}(t))^2 dt \mathbb{1}_{\{\sigma_m < T\}} \right] \mathbb{E} \left[\int_0^T (f^{(n-1)}(t))^2 dt \right]}, \end{aligned} \quad (74)$$

which, upon letting $n \rightarrow \infty$, yields

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\left| \int_0^T f^{(n-1)}(t) \alpha^{(n)}(t) dt \mathbb{1}_{\{\sigma_m < T\}} \right| \right] \leq \sqrt{\mathbb{E} \left[\int_0^T f^2(t) dt \mathbb{1}_{\{\sigma_m < T\}} \right] \mathbb{E} \left[\int_0^T f^2(t) dt \right]}. \quad (75)$$

A further application of the dominated convergence theorem then gives

$$\lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{E} \left[\int_0^T f^{(n-1)}(t) \alpha^{(n)}(t) dt \mathbb{1}_{\{\sigma_m < T\}} \right] = 0. \quad (76)$$

Next, by Jensen's inequality and **(L)**, we have

$$\begin{aligned} & \mathbb{E} \left[\int_0^T \left(\alpha^{(n)}(t) \right)^2 dt \mathbb{1}_{\{\sigma_m < T\}} \right] \\ & \leq \mathbb{E} \left[\int_0^T \mathbb{E} \left[\left(f^{(n-1)}(t) \right)^2 \middle| \mathcal{F}_t^{Y^{(n)}} \right] dt \mathbb{1}_{\{\sigma_m < T\}} \right] \\ & \leq \mathbb{E} \left[\int_0^T \mathbb{E} \left[L \left(1 + \sup_{0 \leq s \leq t} |\xi(s)|^2 + \sup_{0 \leq s \leq t} |Y^{(n-1)}(s)|^2 \right) \middle| \mathcal{F}_t^{Y^{(n)}} \right] dt \mathbb{1}_{\{\sigma_m < T\}} \right] \\ & \leq \mathbb{E} \left[\int_0^T \mathbb{E} \left[L \left(1 + \sup_{0 \leq s \leq t} |\xi(s)|^2 + 2 \sup_{0 \leq s \leq t} |Y^{(n)}(s) - Y^{(n-1)}(s)|^2 \right) \middle| \mathcal{F}_t^{Y^{(n)}} \right] dt \mathbb{1}_{\{\sigma_m < T\}} \right] \\ & \quad + 2 \mathbb{E} \left[\int_0^T \sup_{0 \leq s \leq t} |Y^{(n)}(s)|^2 dt \mathbb{1}_{\{\sigma_m < T\}} \right] \end{aligned} \quad (77)$$

$$\begin{aligned} & \leq L \left(T \mathbb{P}(\sigma_m < T) + \mathbb{E} \left[\int_0^T \mathbb{E} \left[\|\xi\|_T^2 + 2 \|Y^{(n)} - Y^{(n-1)}\|_T^2 \middle| \mathcal{F}_t^{Y^{(n)}} \right] dt \mathbb{1}_{\{\sigma_m < T\}} \right] \right. \\ & \quad \left. + 2T \mathbb{E} \left[\|Y^{(n)}\|_T^2 \mathbb{1}_{\{\sigma_m < T\}} \right] \right). \end{aligned} \quad (78)$$

It is easy to see that

$$\lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{E} \left[\|Y^{(n)}\|_T^2 \mathbb{1}_{\{\sigma_m < T\}} \right] = \lim_{m \rightarrow \infty} \mathbb{E} \left[\|Y\|_T^2 \mathbb{1}_{\{\sigma_m < T\}} \right] = 0, \quad (79)$$

and

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\int_0^T \mathbb{E} \left[\|Y^{(n)} - Y^{(n-1)}\|_T^2 \middle| \mathcal{F}_t^{Y^{(n)}} \right] dt \mathbb{1}_{\{\sigma_m < T\}} \right] \leq T \lim_{n \rightarrow \infty} \mathbb{E} \left[\|Y^{(n)} - Y^{(n-1)}\|_T^2 \right] = 0. \quad (80)$$

Finally, since $\mathbb{E} [\|\xi\|_T^2] < \infty$, by [6, Theorem 5.5.1], the collection

$$\left\{ \mathbb{E} \left[\|\xi\|_T^2 \middle| \mathcal{F}_t^{Y^{(n)}} \right]; n \in \mathbb{N}_0, 0 \leq t \leq T \right\}$$

is uniformly integrable. It then follows that

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{E} \left[\int_0^T \mathbb{E} \left[\|\xi\|_T^2 \middle| \mathcal{F}_t^{Y^{(n)}} \right] \mathbb{1}_{\{\sigma_m < T\}} dt \right] = 0. \quad (81)$$

Combining (77), (79), (80) and (81), we infer that

$$\lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{E} \left[\int_0^T \left(\alpha^{(n)}(t) \right)^2 dt \mathbb{1}_{\{\sigma_m < T\}} \right] = 0. \quad (82)$$

Hence, the desired (69) follows from (73), (76) and (82), which in turn completes the proof of the theorem.

4 Application in Information Theory

In the context of information theory, an SDE taking the form

$$Y(t) = \int_0^t f(s, \xi, Y) ds + W(t), \quad 0 \leq t \leq T, \quad (83)$$

is often used to model a continuous-time Gaussian channel with feedback, where ξ is interpreted as the message to be transmitted, f as the channel input and Y as the channel output. The *mutual information* $I_T(\xi; Y)$, a fundamental notion measuring the information transmission rate of the channel (83), is defined as

$$I_T(\xi; Y) := \mathbb{E} \left[\log \frac{d\mu_{\xi, Y}}{d(\mu_\xi \times \mu_Y)}(\xi, Y) \right] = \mathbb{D}(\mu_{\xi, Y} \| \mu_\xi \times \mu_Y). \quad (84)$$

Exemplifying the possible applications of Theorem 1.1, the following theorem show that $I_T(\xi; Y)$ can be approximated via the means of Picard's iteration.

Theorem 4.1. *Let $\{Y^{(n)}\}_{n=0}^\infty$ be as in (4) with $g \equiv 1$ and $\xi = \{\xi(t), \mathcal{F}_t; 0 \leq t \leq T\}$ be such that $\mathbb{E}[\|\xi\|_T^2] < \infty$. Suppose f and g satisfy **(L)** and **(G)**. Then,*

$$\lim_{n \rightarrow \infty} I_T(\xi; Y^{(n)}) = I_T(\xi; Y). \quad (85)$$

*In addition, if f further satisfies **(B)**, then, for any $p \geq 1$, we have*

$$|I_T(\xi; Y^{(n)}) - I_T(\xi; Y)| \lesssim_{M, K, L, p} \left(\frac{(c_1 T)^{n-1}}{(n-1)!} + \frac{(c_1 T)^n}{n!} \right)^{\frac{1}{2p}} + \sqrt{\frac{(c_2 T)^{n-1}}{(n-1)!} + \frac{(c_2 T)^n}{(n-1)!}}, \quad (86)$$

where c_1, c_2 are the constants as in Theorem 2.1.

Proof. First of all, we note that the mutual information $I_T(\xi; Y)$ can be split into

$$\begin{aligned} I_T(\xi; Y) &= \mathbb{E} \left[\log \left(\frac{d\mu_{\xi, Y}}{d(\mu_\xi \times \mu_Y)}(\xi, Y) \right) \right] \\ &= \mathbb{E} \left[\log \left(\frac{d\mu_{\xi, Y}}{d(\mu_\xi \times \mu_W)}(\xi, Y) \right) \right] - \mathbb{E} \left[\log \left(\frac{d\mu_Y}{d\mu_W}(Y) \right) \right] \\ &= \mathbb{D}(\mu_{\xi, Y} \| \mu_\xi \times \mu_W) - \mathbb{D}(\mu_Y \| \mu_W). \end{aligned} \quad (87)$$

Similarly, for each $n \in \mathbb{N}$, we have

$$I_T(\xi; Y^{(n)}) = \mathbb{D}(\mu_{\xi, Y^{(n)}} \| \mu_{\xi} \times \mu_W) - \mathbb{D}(\mu_{Y^{(n)}} \| \mu_W). \quad (88)$$

Then, by 2) of Theorem 1.1, we have

$$\lim_{n \rightarrow \infty} \mathbb{D}(\mu_{Y^{(n)}} \| \mu_W) = \mathbb{D}(\mu_Y \| \mu_W). \quad (89)$$

Hence, to prove (85), it remains to show

$$\lim_{n \rightarrow \infty} \mathbb{D}(\mu_{\xi, Y^{(n)}} \| \mu_W) = \mathbb{D}(\mu_{\xi, Y} \| \mu_W). \quad (90)$$

Towards this goal, applying Proposition 2.2, we have

$$\begin{aligned} \frac{d\mu_{\xi, Y^{(n)}}}{d(\mu_{\xi} \times \mu_W)}(\xi, Y^{(n)}) &= \exp\left(\int_0^T \beta^{(n)}(t) dY^{(n)}(t) - \frac{1}{2} \int_0^T (\beta^{(n)}(t))^2 dt\right) \\ &= \exp\left(\int_0^T \beta^{(n)}(t) dW(t) + \int_0^T \beta^{(n)}(t) f^{(n-1)}(t) dt - \frac{1}{2} \int_0^T (\beta^{(n)}(t))^2 dt\right), \end{aligned} \quad (91)$$

and furthermore,

$$\frac{d\mu_{\xi, Y}}{d(\mu_{\xi} \times \mu_W)}(\xi, Y) = \exp\left(\int_0^T f(t, \xi, Y) dY(t) - \frac{1}{2} \int_0^T f^2(t, \xi, Y) dt\right). \quad (92)$$

Here, $\beta^{(n)}(t) = \mathbb{E}\left[f^{(n-1)}(t) | \mathcal{F}_t^{\xi, Y^{(n)}}\right]$. Now, with a parallel argument as in the proof of Theorem 1.1, we reach

$$\begin{aligned} \mathbb{E} \left[\left| \log \left(\frac{d\mu_{\xi, Y^{(n)}}}{d(\mu_{\xi} \times \mu_W)}(\xi, Y^{(n)}) \right) - \log \left(\frac{d\mu_{\xi, Y}}{d(\mu_{\xi} \times \mu_W)}(\xi, Y) \right) \right| \right] \\ \leq 4(c_1 KT)^{1/2} \sqrt{\frac{(c_2 T)^n}{n!}} + 4(c_1 KT)^{1/2} \sqrt{\frac{(c_2 T)^{n-1}}{(n-1)!}} + 2c_1 KT \frac{(c_2 T)^n}{n!}, \end{aligned} \quad (93)$$

which implies (90) and hence (85). In addition, if f satisfies the condition **(B)**, we can combine (93) and (40) to establish (86). \square

Remark 4.1. Effective computation of the mutual information of a continuous-time Gaussian feedback channel under various input constraints has been a long open problem in information theory (see, e.g., [19, 5, 10, 20, 16]). Evidently, Theorem 4.1 suggests a solution via Picard's iteration, which, thanks to the rate of convergence in (84), can be particularly promising for the channel under the condition **(B)**, which is often termed as "peak power constraint" (see, e.g., [18]) in information theory.

5 Appendices

A Proof of Theorem 2.1

The proof closely follows those showing the existence and uniqueness of the solution to some “conventional” SDEs where ξ is absent (see, e.g., [11, 15]). In the following we only sketch the proof and emphasize the technical issues only present in our setting.

The uniqueness of the solution can be proven using Gronwall’s inequality in a parallel fashion as in [11, 15], so here we only prove the existence. More precisely, we achieve this by establishing (7), which in turn implies the well-definedness of the limit of the sequence $\{Y^{(n)}\}$.

To this end, we first note that for any $n \in \mathbb{N}_0$ and any $t \in [0, T]$,

$$\begin{aligned} \sup_{0 \leq s \leq t} |Y^{(n)}(s)|^2 &\leq 2 \sup_{0 \leq s \leq t} \left(\int_0^s f(u, \xi, Y^{(n-1)}) du \right)^2 + 2 \sup_{0 \leq s \leq t} \left(\int_0^s g(u, Y^{(n-1)}) dW(u) \right)^2 \\ &\leq 2t \int_0^t f^2(u, \xi, Y^{(n-1)}) du + 2 \sup_{0 \leq s \leq t} \left(\int_0^s g(u, Y^{(n-1)}) dW(u) \right)^2, \end{aligned} \quad (94)$$

where the last inequality follows from the Cauchy-Schwarz inequality. Taking expectation on both sides of (94) then yields

$$\begin{aligned} \mathbb{E} \left[\sup_{0 \leq s \leq t} |Y^{(n)}(s)|^2 \right] &\leq 2LT \left(T + T \mathbb{E} [\|\xi\|_T^2] + \int_0^t \mathbb{E} \left[\sup_{0 \leq v \leq u} |Y^{(n-1)}(v)|^2 \right] du \right) \\ &\quad + 2 \mathbb{E} \left[\sup_{0 \leq s \leq t} \left(\int_0^s g(u, Y^{(n-1)}) dW(u) \right)^2 \right]. \end{aligned} \quad (95)$$

For the second term in (95), by Doob’s submartingale inequality and the Itô isometry, we have

$$\begin{aligned} \mathbb{E} \left[\sup_{0 \leq s \leq t} \left| \int_0^s g(u, \xi, Y^{(n-1)}) dW(u) \right|^2 \right] &\leq 2^2 \mathbb{E} \left[\int_0^t g^2(u, \xi, Y^{(n-1)}) du \right] \\ &\leq 4L \left(T + T \mathbb{E} [\|\xi\|_T^2] + \int_0^t \mathbb{E} \left[\sup_{0 \leq v \leq u} |Y^{(n-1)}(v)|^2 \right] du \right). \end{aligned} \quad (96)$$

Now, letting $k_1 := 2LT(T + 4)(1 + \mathbb{E}[\|\xi\|_T^2])$ and $k_2 := 2L(T + 4)$, we have

$$\mathbb{E} \left[\sup_{0 \leq s \leq t} |Y^{(n)}(s)|^2 \right] \leq k_1 + k_2 \int_0^t \mathbb{E} \left[\sup_{0 \leq v \leq u} |Y^{(n-1)}(v)|^2 \right] du,$$

a recursive application of which yields

$$\mathbb{E} \left[\sup_{0 \leq s \leq t} |Y^{(n)}(s)|^2 \right] \leq k_1 e^{k_2 t}. \quad (97)$$

Next, using a similar approach as above and taking into account the fact that f and g satisfy **(L)**, we reach, for any $t \in [0, T]$,

$$\begin{aligned} \mathbb{E} \left[\sup_{0 \leq s \leq t} |Y^{(1)}(s) - Y^{(0)}(s)|^2 \right] &\leq 2t \int_0^t \mathbb{E} [f^2(u, \xi, Y^{(0)})] du + 2 \mathbb{E} \left[\sup_{0 \leq s \leq t} \left| \int_0^s g(u, Y^{(0)}) dW(u) \right|^2 \right] \\ &\leq 2T^2 L (1 + \mathbb{E} [\|\xi\|_T^2]) + 2 \mathbb{E} \left[\sup_{0 \leq s \leq t} \left| \int_0^s g(u, Y^{(0)}) dW(u) \right|^2 \right]. \end{aligned} \quad (98)$$

Again, by Doob's submartingale inequality and the Itô isometry,

$$\begin{aligned} \mathbb{E} \left[\sup_{0 \leq s \leq t} \left| \int_0^s g(u, Y^{(0)}) dW(u) \right|^2 \right] &\leq 2^2 \mathbb{E} \left[\int_0^t g^2(u, Y^{(0)}) du \right] \\ &\leq 4LT (1 + \mathbb{E} [\|\xi\|_T^2]). \end{aligned} \quad (99)$$

Hence, from (98) and (99), we have

$$\mathbb{E} \left[\sup_{0 \leq s \leq t} |Y^{(1)}(s) - Y^{(0)}(s)|^2 \right] \leq 2TL(T+4) (1 + \mathbb{E} [\|\xi\|_T^2]) =: c_1, \quad (100)$$

for any $t \in [0, T]$. It then follows from the fact that f and g satisfy **(L)** that

$$\begin{aligned} \mathbb{E} \left[\sup_{0 \leq s \leq t} |Y^{(2)}(s) - Y^{(1)}(s)|^2 \right] &\leq 2 \mathbb{E} \left[\left| \int_0^t (f(u, \xi, Y^{(1)}) - f(u, \xi, Y^{(0)})) du \right|^2 \right] \\ &\quad + 8 \mathbb{E} \left[\left| \int_0^t (g(u, \xi, Y^{(1)}) - g(u, \xi, Y^{(0)})) dW(u) \right|^2 \right] \\ &\leq 2K(T+4) \int_0^t \mathbb{E} \left[\sup_{0 \leq v \leq u} |Y^{(1)}(v) - Y^{(0)}(v)|^2 \right] du \\ &\leq c_1 c_2 t, \end{aligned}$$

where $c_2 := 2K(T+4)$. Now, by way of induction, one can readily show that

$$\mathbb{E} \left[\sup_{0 \leq s \leq t} |Y^{(n+1)}(t) - Y^{(n)}(t)|^2 \right] \leq c_1 \frac{(c_2 t)^n}{n!}, \quad (101)$$

which can be used to establish

$$\mathbb{E} \left[\|Y^{(n)}\|_T^2 \right] \leq k_1 e^{k_2 T},$$

and thereby completing the proof of (6). Consequently, the sequence $\{Y^{(n)}\}_{n=0}^\infty$ is Cauchy in $L^2(\Omega, \mathcal{F}, \mathbb{P})$ with the limit $Y = \{Y(t), \mathcal{F}; 0 \leq t \leq T\}$. Furthermore, by Fatou's lemma, the limit Y satisfies

$$\mathbb{E} [\|Y\|_T^2] \leq \liminf_{n \rightarrow \infty} \mathbb{E} [\|Y^{(n)}\|_T^2] \leq k_1 e^{k_2 T}. \quad (102)$$

We are now ready to establish (7). To this end, note that for any $t \in [0, T]$, we have

$$\mathbb{E} \left[\sup_{0 \leq s \leq t} |Y^{(n)}(s) - Y(s)|^2 \right] \leq c_2 \int_0^t \mathbb{E} \left[\sup_{0 \leq v \leq u} |Y^{(n-1)}(v) - Y(v)|^2 \right] du,$$

which, upon a sequence of recursive substitutions, yields that

$$\mathbb{E} \left[\sup_{0 \leq s \leq t} |Y^{(n)}(s) - Y(s)|^2 \right] \leq \mathbb{E} [\|Y\|_T^2] \frac{(c_2 t)^n}{n!}, \quad (103)$$

establishing the L^2 convergence of $\{Y^{(n)}\}_{n=0}^\infty$. To establish that Y is the uniform pointwise limit of the sequence $\{Y^{(n)}\}_{n=0}^\infty$ to complete the proof of (7), one can use the estimate in (101) and the Borel-Cantelli lemma; moreover, together with the estimate in (103), one can show that Y indeed satisfies (1). These assertions can be shown in the same way as in [11, 15], and thereby we omit the details thereof.

Remark A.1. For the special case that $g \equiv 1$ and f satisfies **(B)**, c_1 can be replaced by $\tilde{c}_1 := 2T(M+4)$, a constant that is independent of ξ , which can be justified by noting

$$\mathbb{E} \left[\sup_{0 \leq s \leq t} |Y^{(1)}(s) - Y^{(0)}(s)|^2 \right] \leq 2t \mathbb{E} \left[\int_0^t f^2(u, \xi, Y^{(0)}) du \right] + 8T \leq 2T(M+4) = \tilde{c}_1. \quad (104)$$

It can be verified that, via a similar argument, the constants k_1 and c_3 can also be respectively replaced by some constants that are independent of ξ .

B Proof of Proposition 2.1

By [13, Theorem 7.1], it suffices to establish

$$\mathbb{E} \left[\exp \left(- \int_0^T f^{(n)}(t) dW(t) - \frac{1}{2} \int_0^T (f^{(n)}(t))^2 dt \right) \right] = 1, \text{ for all } n \in \mathbb{N}_0. \quad (105)$$

In the meantime, to prove (105), by the well-known Novikov's condition [11], it suffices to construct, for each $n \in \mathbb{N}$, a sequence $\{t_m^{(n)}\}_{m=0}^{M_n}$ with $0 = t_0^{(n)} < t_1^{(n)} < \dots < t_{M_n}^{(n)} = T$, such that for any $m = 1, 2, \dots, M_n$,

$$\mathbb{E} \left[\exp \left(\int_{t_{m-1}^{(n)}}^{t_m^{(n)}} (f^{(n)}(t))^2 dt \right) \right] < \infty. \quad (106)$$

To this end, we first note that it follows from **(L)** that, for each $n \in \mathbb{N}_0$,

$$\begin{aligned} \|Y^{(n+1)}\|_T^2 &\leq 2\|W\|_T^2 + 2 \int_0^T L \left(1 + \sup_{0 \leq s \leq t} |\xi(s)|^2 + \sup_{0 \leq s \leq t} |Y^{(n)}(s)|^2 \right) dt \\ &\leq 2\|W\|_T^2 + 2LT(1 + \|\xi\|_T^2) + 2L \int_0^T \sup_{0 \leq s \leq t} |Y^{(n)}(s)|^2 dt, \end{aligned}$$

a recursive application of which yields

$$\|Y^{(n+1)}\|_T^2 \leq 2C_{n+1}\|W\|_T^2 + 2LTC_{n+1}(1 + \|\xi\|_T^2),$$

where $C_{n+1} = \frac{(2L)^{n+1}-1}{2L-1}$. Then, with the linear growth assumption **(G)**, we derive

$$\begin{aligned} \mathbb{E} \left[\exp \left(\frac{1}{2} \int_{t_{m-1}^{(n)}}^{t_m^{(n)}} (f^{(n)}(t))^2 dt \right) \right] &\leq \mathbb{E} \left[\exp \left(\frac{1}{2} L (t_m^{(n)} - t_{m-1}^{(n)}) (1 + \|\xi\|_T^2 + \|Y^{(n)}\|_T^2) \right) \right] \\ &\leq e^{LTD_{n,m}} \mathbb{E} \left[e^{LTD_{n,m}\|\xi\|_T^2} \right] \mathbb{E} \left[e^{D_{n,m}\|W\|_T^2} \right], \end{aligned}$$

where $D_{n,m} = LC_n (t_m^{(n)} - t_{m-1}^{(n)})$. Now, applying Doob's submartingale inequality, we conclude that, as long as $\max_{m \in \{1, 2, \dots, M_n\}} \{t_m^{(n)} - t_{m-1}^{(n)}\} \leq \frac{\varepsilon}{L^2 C_n T}$,

$$\mathbb{E} \left[\exp \left(\frac{1}{2} \int_{t_{m-1}^{(n)}}^{t_m^{(n)}} (f^{(n)}(t))^2 dt \right) \right] \leq 4e^{LTD_{n,m}} \mathbb{E} \left[e^{LTD_{n,m}\|\xi\|_T^2} \right] \mathbb{E} \left[e^{D_{n,m}W(T)^2} \right] < \infty,$$

which immediately implies the proposition.

C Proof of Proposition 2.2

Fix $x \in C[0, T]$ and consider the process Z^x satisfying

$$Z^x(t) = \int_0^t h(s, x, W) ds + W(t), \quad 0 \leq t \leq T.$$

Using similar arguments as in Section 3.1, we infer that there exists a non-anticipative functional $Q : [0, T] \times C[0, T] \rightarrow \mathbb{R}$, such that for any $t \in [0, T]$,

$$Z(t) = Q(t, \xi, W) \text{ and } Z^x(t) = Q(t, x, W).$$

Note that [13, Theorem 7.14], together with the condition (17), implies that $\mu_Z \ll \mu_W$ and $\mu_{Z^x} \ll \mu_W$, and furthermore,

$$\frac{d\mu_{\xi, Z}}{d(\mu_\xi \times \mu_W)}(x, w) = \frac{d\mu_{Z^x}}{d\mu_W}(w) \quad \mu_\xi \times \mu_W \text{-a.s.},$$

and

$$\frac{d\mu_{\xi, Z}}{d(\mu_\xi \times \mu_W)}(x, W) = \frac{d\mu_{Z^x}}{d\mu_W}(W) = \exp\left(\int_0^T \gamma(t, x, W)dW(t) - \frac{1}{2} \int_0^T \gamma^2(t, x, W)dt\right),$$

where $\gamma : [0, T] \times C[0, T] \rightarrow \mathbb{R}$ is a non-anticipative functional such that

$$\gamma(t, x, Z^x) = \gamma(t, x, Q(\cdot, x, W)) = \mathbb{E}[h(t, x, W) | \mathcal{F}_t^{Z^x}].$$

Since $\mu_Z \ll \mu_W$, by [13, Lemma 4.10], we have

$$\frac{d\mu_{\xi, Z}}{d(\mu_\xi \times \mu_W)}(\xi, Z) = \exp\left(\int_0^T \gamma(t, \xi, Z)dZ(t) - \frac{1}{2} \int_0^T \gamma^2(t, \xi, Z)dt\right).$$

We now claim that for each $t \in [0, T]$, $\gamma(t, \xi, Z) = \mathbb{E}[h(t, \xi, W) | \mathcal{F}_t^{\xi, Z}]$ \mathbb{P} -a.s., which will imply the proposition. Indeed, for any bounded, $\mathcal{F}_t^{\xi, Z}$ -measurable random variable $\lambda(\xi, Z)$, we have

$$\begin{aligned} \mathbb{E}[\lambda(\xi, Z)\gamma(t, \xi, Z)] &= \mathbb{E}[\lambda(\xi, Q(t, \xi, W))\gamma(t, \xi, Q(t, \xi, W))] \\ &= \int_{C[0, T]} \int_{C[0, T]} \lambda(x, Q(t, x, w))\gamma(t, x, Q(x, w))\mu_W(dw) \times \mu_\xi(dx) \\ &= \int_{C[0, T]} \int_{C[0, T]} \lambda(x, Q(t, x, w))h(t, x, w)\mu_W(dw) \times \mu_\xi(dx) \\ &= \mathbb{E}[\lambda(\xi, Z)h(t, \xi, W)], \end{aligned}$$

where the second equality follows from the independence of ξ and W , and the third one follows from the fact that $\gamma(t, x, Q(x, W)) = \mathbb{E}[h(t, x, W) | \mathcal{F}_t^{Z^x}]$. The proof of the claim is then completed by noting that $\gamma(t, \xi, Z)$ is $\mathcal{F}_t^{\xi, Z}$ -measurable.

D Proof of Proposition 2.3

We will need the following lemma, which generalizes the well-known log-sum inequality for discrete probability distributions (see, e.g., [3]).

Lemma D.1. Let f, g be positive measurable functions defined on a probability space (E, \mathcal{E}, μ) such that

$$\mu(f), \mu(g), \mu(f \log f), \mu(f \log g) < \infty,$$

where $\mu(\cdot)$ means the integral with respect to μ . Then, we have

$$\int_E f(x) \log \frac{f(x)}{g(x)} \mu(dx) \geq \mu(f) \log \frac{\mu(f)}{\mu(g)}. \quad (107)$$

Proof. We define two probability measures ν, κ as follows: for any $A \in \mathcal{E}$,

$$\nu(A) := \int_A \frac{f(x)}{\mu(f)} \mu(dx) \quad \text{and} \quad \kappa(A) := \int_A \frac{g(x)}{\mu(g)} \mu(dx).$$

Since $f/g > 0$, we have $\nu \ll \kappa$ with $d\nu/d\kappa = (\mu(f)/\mu(g))^{-1} f/g$. Therefore, by the well-known fact that relative entropy is non-negative [8], we have

$$\begin{aligned} 0 \leq \mathbb{D}(\nu \parallel \kappa) &= \int_E \log \frac{d\nu}{d\kappa}(x) \nu(dx) = \int_E \frac{d\nu}{d\mu}(x) \log \frac{d\nu}{d\kappa}(x) \mu(dx) \\ &= \mu(f)^{-1} \int_E f(x) \log \frac{f(x)}{g(x)} \mu(dx) - \log \frac{\mu(f)}{\mu(g)}, \end{aligned}$$

establishing the lemma. □

Proof of Lemma 2.3. First of all, we define

$$f(w, z) := \frac{d\mu_{U,Z}}{d(\mu_V \times \mu_Z)}(w, z), \quad \text{and} \quad g(w, z) := \frac{d\mu_{V,Z}}{d(\mu_V \times \mu_Z)}(w, z).$$

Then, by Lemma D.1, we have, for each $w \in E$,

$$\int_E f(w, z) \log \frac{f(w, z)}{g(w, z)} \mu_Z(dz) \geq \mu_Z(f(w, \cdot)) \log \frac{\mu_Z(f(w, \cdot))}{\mu_Z(g(w, \cdot))}, \quad (108)$$

where

$$\mu_Z(f(w, \cdot)) = \int_E \frac{d\mu_{U,Z}}{d(\mu_V \times \mu_Z)}(w, z) \mu_Z(dz) = \frac{d\mu_U}{d\mu_V}(w)$$

and

$$\mu_Z(g(w, \cdot)) = \int_E \frac{d\mu_{V,Z}}{d(\mu_V \times \mu_Z)}(w, z) \mu_Z(dz) = 1.$$

Finally, integrating both sides of (108) with respect to the measure μ_V yields the desired result. \square

E Proof of Lemma 3.1

By [13, Theorem 7.1], we have

$$\frac{d\mu_{Y^{(n+1)}}(Y^{(n+1)})}{d\mu_W} = \frac{1}{\mathbb{E} \left[\exp \left(- \int_0^T f^{(n)}(t) dY^{(n+1)}(t) + \frac{1}{2} \int_0^T (f^{(n)}(t))^2 dt \right) \middle| \mathcal{F}_T^{Y^{(n+1)}} \right]}. \quad (109)$$

Then, applying Jensen's inequality, we establish (42) as follows:

$$\begin{aligned} \mathbb{E} \left[\left(\frac{d\mu_{Y^{(n+1)}}(Y^{(n+1)})}{d\mu_W} \right)^p \right] &= \mathbb{E} \left[\frac{1}{\mathbb{E}^p \left[\exp \left(- \int_0^T f^{(n)}(t) dY^{(n+1)}(t) + \frac{1}{2} \int_0^T (f^{(n)}(t))^2 dt \right) \middle| \mathcal{F}_T^{Y^{(n+1)}} \right]} \right] \\ &\leq \mathbb{E} \left[\exp \left(p \int_0^T f^{(n)}(t) dY^{(n+1)}(t) - \frac{p}{2} \int_0^T (f^{(n)}(t))^2 dt \right) \right] \\ &= \mathbb{E} \left[\exp \left(p \int_0^T f^{(n)}(t) dW(t) - \frac{p^2}{2} \int_0^T (f^{(n)}(t))^2 dt + \frac{p(p+1)}{2} \int_0^T (f^{(n)}(t))^2 dt \right) \right] \\ &\leq e^{\frac{p(p+1)M}{2}} \mathbb{E} \left[\exp \left(p \int_0^T f^{(n)}(t) dW(t) - \frac{p^2}{2} \int_0^T (f^{(n)}(t))^2 dt \right) \right] \\ &= e^{\frac{p(p+1)M}{2}}, \end{aligned}$$

where the last equality follows from the martingale property of the process

$$\left\{ \exp \left(p \int_0^t f^{(n)}(s) dW(s) - \frac{p^2}{2} \int_0^t (f^{(n)}(s))^2 ds \right), 0 \leq t \leq T \right\}.$$

A parallel argument can be used to establish (43).

References

- [1] Chen, Y., Jia, L. and Wang, J. (2004). *Machine Learning and Statistical Modeling Approaches to Image Retrieval*. Springer, Boston, MA.
- [2] Bishop, C. (2006). *Pattern Recognition and Machine Learning*. Springer, New York.
- [3] Cover, T and Thomas, J. (2006). *Elements of Information Theory*, 2nd ed., Wiley Interscience, New York.
- [4] Csiszar, I. and Korner, J. (2011). *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, Cambridge.

- [5] Duncan, T. E. (1970). On the calculation of the mutual information. *SIAM Journal on Applied Mathematics*, vol. 19(1), 215-220.
- [6] Durrett, R. (2013). *Probability: Theory and Examples*, 4th edition. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge.
- [7] Hobson, A. (1971). *Concepts in statistical mechanics*. Gordon and Breach, New York.
- [8] Ihara, S. (1993). *Information Theory for Continuous Systems*. World Scientific, Singapore.
- [9] Johnson, O. (2004). *Information theory and the central limit theorem*. Imperial College Press, London, 2004.
- [10] Kadota, T., Zakai, M. and Ziv, J. (1971). Mutual information of the white Gaussian channel with and without feedback. *IEEE Trans. Info. Theory*, vol. 17, pp. 368-371, 1971.
- [11] Karatzas, I. and Shreve, S. E. (1998). *Brownian Motion and Stochastic Calculus*. Springer Verlag, New York.
- [12] Kullback, S. (1959). *Information Theory and Statistics*. Wiley, New York.
- [13] Liptser, R. S. and Shiriyayev, A. N. (1977). *Statistics of Random Process I: General Theory*. Springer Verlag, New York.
- [14] MacKay, D. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, Cambridge, UK.
- [15] Mao, X. (2007). *Stochastic Differential Equations and Applications*. Woodhead Publishing, Cambridge.
- [16] Mayer-Wolf, E. and Zakai, M. (2007). Some relations between mutual information and estimation error in Wiener space. *The Annals of Applied Probability*, vol. 17, no. 3, 1102-1116.
- [17] Michie, D., Spiegelhalter, J. and Taylor, C. (1994). *Machine learning, neural and statistical classification*. Prentice Hall, Englewood Cliffs, NJ.
- [18] Ozarow, L. H., Wyner, A.D. and Ziv, J. (1988). Achievable rates for a constrained Gaussian channel. *IEEE Transactions on Information Theory*, vol. 34, no.3, 365-370.
- [19] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, vol. 27, 379-423, 623-656.

- [20] Snyders, J. (1972). Error formulae for optimal linear filtering, prediction and interpolation of stationary time series. *Ann. Math. Statist.*, vol. 43, pp. 1935-1943.
- [21] Tribus, M. (1961). *Thermostatistics and Thermodynamics: An Introduction to Energy, Information and States of Matter, with Engineering Applications*, Van Nostrand, Princeton, NJ.
- [22] Tsybakov, A. (2009). *Introduction to Nonparametric Estimation*. Springer New York, New York, NY.
- [23] Villani, C. (2003). *Topics in optimal transportation*. American Mathematical Society, Providence, RI.
- [24] Villani, C. (2009). *Optimal Transport: Old and New*. Springer, Berlin Heidelberg.