

Information-Theoretic Extensions of the Shannon-Nyquist Sampling Theorem

Xianming Liu
Huazhong University of Science and Technology
email: xmliu@hust.edu.cn

Guangyue Han
The University of Hong Kong
email: ghan@hku.hk

December 5, 2018

Abstract

A continuous-time white Gaussian channel can be formulated using a white Gaussian noise, and a conventional way for examining such a channel is the sampling approach based on the classical Shannon-Nyquist sampling theorem, where the original continuous-time channel is converted to an equivalent discrete-time channel, to which a great variety of established tools and methodology can be applied. However, one of the key issues of this scheme is that continuous-time feedback cannot be incorporated into the channel model. It turns out that this issue can be circumvented by considering the Brownian motion formulation of a continuous-time white Gaussian channel. Nevertheless, as opposed to the white Gaussian noise formulation, a link that establishes the information-theoretic connection between a continuous-time white Gaussian channel under the Brownian motion formulation and its discrete-time counterparts has long been missing. This paper is to fill this gap by establishing information-theoretic extensions of the Shannon-Nyquist theorem, which naturally yield causality-preserving connections between continuous-time Gaussian feedback channels and their associated discrete-time versions in the forms of sampling and approximation theorems. As an example of the possible applications of the extensions, we use the above-mentioned connections to analyze the capacity of a continuous-time white Gaussian feedback channel.

1 Introduction

Continuous-time Gaussian channels were considered at the very inception of information theory. In his celebrated paper [26] birthing information theory, Shannon studied the following continuous-time white Gaussian channels:

$$Y(t) = X(t) + Z(t), \quad t \in \mathbb{R}, \quad (1)$$

where the stochastic process $\{X(t)\}$ is the *channel input* with average power limit P , $\{Z(t)\}$ is the white Gaussian noise with flat power spectral density 1 and $\{Y(t)\}$ is the *channel output*. Shannon actually only considered the case that the channel has bandwidth limit ω , namely, the support of the Fourier transform of $\{X(t)\}$ is contained in $[-\omega, \omega]$. Using the

celebrated Shannon-Nyquist sampling theorem [22, 27], the continuous-time channel (1) can be equivalently represented by a parallel Gaussian channel:

$$Y_n^{(\omega)} = X_n^{(\omega)} + Z_n^{(\omega)}, \quad n \in \mathbb{Z}, \quad (2)$$

where the noise process $\{Z_n^{(\omega)}\}$ is i.i.d. with variance 1 [5]. Regarding the “space” index n as time, the above parallel channel can be interpreted as a discrete-time Gaussian channel associated with the continuous-time channel (1). It is well known from the theory of discrete-time Gaussian channels that the capacity of the channel (2) can be computed as

$$C^{(\omega)} = \omega \log \left(1 + \frac{P}{2\omega} \right). \quad (3)$$

Then, the capacity C of the channel (1) can be computed by taking the limit of the above expression as ω tends to infinity:

$$C = \lim_{\omega \rightarrow \infty} C^{(\omega)} = P/2. \quad (4)$$

The *sampling approach* consisting of (1)-(4) as above, which serves as a link between the continuous-time channel (1) and the discrete-time channel (2), typifies a conventional way to examine continuous-time Gaussian channels: convert them into associated discrete-time Gaussian channels, for which we have ample ammunition at hands. Moments of reflection, however, reveals that the sampling approach for the channel capacity (with bandwidth limit or not) is heuristic in nature: For one thing, a bandwidth-limited signal cannot be time-limited, which renders it infeasible to define the data transmission rate if assuming a channel has bandwidth limit. In this regard, rigorous treatments coping with this issue and other technicalities can be found in [29, 9]; see also [28] for a relevant in-depth discussion. Another issue is that, even disregarding the above technical nuisance arising from the bandwidth limit assumption, the sampling approach only gives a lower bound on the capacity of (1): it shows that $P/2$ is achievable via a class of special coding schemes, but it is not clear that why transmission rate higher than $P/2$ cannot be achieved by other coding schemes. The capacity of (1) was rigorously studied in [8, 4], and a complete proof establishing $P/2$ as its capacity can be found in [1, 2].

Alternatively, the continuous-time white Gaussian channel (1) can be examined [14] under the Brownian motion formulation:

$$Y(t) = \int_0^t X(s)ds + B(t), \quad t \in \mathbb{R}_+ \quad (5)$$

where slightly abusing the notation, we still use $\{Y(t)\}$ to denote the output corresponding to the input $\{X(s)\}$, and $\{B(t)\}$ denotes the standard Brownian motion. Here we remark that, the formulation in (5) is often regarded as the integral version of that in (1) due to the long-held heuristic interpretation of a white Gaussian noise as the “derivative” of a Brownian motion, and via a routine orthonormal decomposition argument, both of the two channels are equivalent to a parallel channel consisting of infinitely many Gaussian sub-channels [3].

An immediate and convenient consequence of such a formulation is that many notions in discrete time, including mutual information and typical sets, carry over to the continuous-time setting, which will rid us of the nuisances arising from the bandwidth limit assumption.

Indeed, such a framework yields a fundamental formula for the mutual information of the channel (5) [6, 16] and a clean and direct proof [16] that the capacity of (5) is $P/2$; moreover, as evidenced by numerous results collected in [14] on continuous-time Gaussian channels, the use of Brownian motions makes our treatment more rigorous and equips us with a wide range of established techniques and tools from stochastic calculus. Here we remark that Girsanov's theorem, one of the fundamental theorems in stochastic calculus, lays the foundation of our treatment. We refer to [20, 14], where Girsanov's theorem (and its numerous variants) and its wide range of applications in information theory are discussed in great details.

Furthermore, the Brownian motion formulation is also versatile enough to accommodate feedback. Here we note that a continuous-time Gaussian channel as in (5) can be alternatively written as

$$Y(t) = \int_0^t g(s, M) ds + B(t), \quad t \in \mathbb{R}_+, \quad (6)$$

where M is a random variable taking values in a finite alphabet \mathcal{M} , interpreted as the message to be transmitted through the channel, and g is a real-valued deterministic function depending on $s \in \mathbb{R}_+$, $M \in \mathcal{M}$, interpreted as the channel input. The formulation in (6) can be readily extended to model a continuous-time white Gaussian channel with feedback, or simply, Gaussian feedback channel, which is characterized by the following stochastic differential equation (SDE) [14]:

$$Y(t) = \int_0^t g(s, M, Y_0^s) ds + B(t), \quad t \in \mathbb{R}_+, \quad (7)$$

where the channel input g also depends on $Y_0^s \triangleq \{Y(r) : 0 \leq r \leq s\}$, the channel output up to time s that is fed back to the sender, which will be referred to as the *channel feedback*. For obvious reasons, with a same set of constraints, the capacity of the channel (7) is greater than or equal to that of (6), that is to say, feedback increases the capacity in general. On the other hand though, for much subtler reasons, with the average power limit P , the capacity of (7) is still $P/2$, namely, feedback does not help with the average power constraint [16].

As opposed to the white Gaussian noise formulation, a feedback channel under the Brownian motion formulation can be naturally translated to the discrete-time setting: the pathwise continuity of a Brownian motion allows the inheritance of temporal causality when the channel is sampled (see Section 3.1) or approximated (see Section 3.2). On the other hand, the white Gaussian noise formulation is facing inherent difficulty as far as inheriting temporal causality is concerned: in converting (1) to (2), while $X_n^{(w)}$ are obtained as “time” samples of $X(t)$, $Z_n^{(w)}$ are in fact “space” samples of $Z(t)$, as they are merely the coefficients of the (extended) Karhunen-Loeve decomposition of $Z(t)$ [10, 11, 12].

On the other hand though, as opposed to the white Gaussian noise formulation, a link that establishes the information-theoretic connection between the continuous-time channel (7) and its discrete-time counterparts has long been missing, which may explain why discrete-time and continuous-time information theory (under the Brownian motion formulation) have largely gone separate ways with little interaction for the past several decades. In this paper, we will fill this gap by establishing information-theoretic extensions of the Shannon-Nyquist theorem, which naturally give causality-preserving connections between continuous-time Gaussian feedback channels and their associated discrete-time versions. We believe the extensions will serve as the above-mentioned missing links and play important

roles in the long run for further developing continuous-time information theory, particularly for the communication scenarios when feedback is present; see some preliminary results collected in the recent work [18, 19].

The remainder of the paper is organized as follows. In Section 2, we introduce our notations and recall some basic notions and facts that will be used in our proofs. In Section 3.1, we prove Theorems 3.2, a sampling theorem for a continuous-time Gaussian feedback channel, which naturally connect such a channel with their sampled discrete-time versions. And in Section 3.2, we prove Theorem 3.6, the so-called approximation theorem, which connects a continuous-time Gaussian feedback channel with its approximated discrete-time versions (in the sense of the Euler-Maruyama approximation [13]). Roughly speaking, a sampling theorem says that a time-sampled channel is “close” to the original channel if the sampling is fine enough, and an approximation theorem says that an approximated channel is “close” to the original channel if the approximation is fine enough, both in an information-theoretic sense. Note that, as elaborated in Remark 3.7, the approximation theorem boils down to the sampling theorem when there is no feedback in the channel. In Section 4, for the purpose of illustrating how our results can be used, we use Theorem 3.6 to give alternative derivation of the capacity of the channel (7).

2 Notations and Preliminaries

We use $(\Omega, \mathcal{F}, \mathbb{P})$ to denote the underlying probability space, and \mathbb{E} to denote the expectation with respect to the probability measure \mathbb{P} . As is typical in the theory of SDEs, we assume the probability space is equipped with a filtration $\{\mathcal{F}_t : 0 \leq t < \infty\}$, which satisfies the *usual conditions* [17] and is rich enough to accommodate the standard Brownian motion $\{B(t) : 0 \leq t < \infty\}$. Throughout the paper, we will use uppercase letters (e.g., $X, Y, Y^{(n)}$) to denote random variables, and their lowercase counterparts (e.g., $x, y, y^{(n)}$) to denote their realizations.

Let $C[0, \infty)$ denote the space of all continuous functions over $[0, \infty)$, and for any $t > 0$, let $C[0, t]$ denote the space of all continuous functions over $[0, t]$. As usual, we will equip the space $C[0, \infty)$ with the filtration $\{\mathcal{B}_t\}_{0 \leq t < \infty}$, where \mathcal{B}_∞ denotes the standard Borel σ -algebra on the space $C[0, \infty)$ and $\mathcal{B}_t = \pi_t^{-1}(\mathcal{B}_\infty)$, where $\pi_t : C[0, \infty) \rightarrow C[0, t]$ is given by the map $(\pi_t x)(s) = x(t \wedge s)$.

For any $\varphi \in C[0, \infty)$, we use $\varphi(\{t_1, t_2, \dots, t_m\})$ to denote $\{\varphi(t_1), \varphi(t_2), \dots, \varphi(t_m)\}$ and φ_0^t to denote $\{\varphi(s) : 0 \leq s \leq t\}$. The sup-norm of φ_0^t , denoted by $\|\varphi_0^t\|$, is defined as $\|\varphi_0^t\| \triangleq \sup_{0 \leq s \leq t} |\varphi(s)|$; and similarly, we define $\|\varphi_0^t - \psi_0^t\| \triangleq \sup_{0 \leq s \leq t} |\varphi(s) - \psi(s)|$. For any $\varphi, \psi \in C[0, \infty)$, slightly abusing the notation, we define $\|\varphi_0^s - \psi_0^t\| \triangleq \|\hat{\varphi}_0^\infty - \hat{\psi}_0^\infty\|$, where $\hat{\varphi}, \hat{\psi} \in C[0, \infty)$ are “stopped” versions of φ, ψ at time s, t , respectively, with $\hat{\varphi}(r) = \varphi(r \wedge s)$ and $\hat{\psi}(r) = \psi(r \wedge t)$ for any $r \in \mathbb{R}_+$.

For any two probability measures μ and ν , we write $\mu \sim \nu$ to mean they are equivalent, namely, μ is absolutely continuous with respect to ν and vice versa. For any two processes $X_0^t = \{X(s); 0 \leq s \leq t\}$ and $Y_0^t = \{Y(s); 0 \leq s \leq t\}$, we use $\mu_{X_0^t}$ and $\mu_{Y_0^t}$ to denote the probability distributions on \mathcal{B}_t induced by X_0^t and Y_0^t , respectively; and if $\mu_{Y_0^t}$ is absolutely continuous with respect to $\mu_{X_0^t}$, we write the Radon-Nikodym derivative of $\mu_{Y_0^t}$ with respect to $\mu_{X_0^t}$ as $d\mu_{Y_0^t}/d\mu_{X_0^t}$. We use $\mu_{Y_0^t|Z=z}$ denote the probability distribution on \mathcal{B}_t induced by Y_0^t given $Z = z$, and $d\mu_{Y_0^t|Z=z}/d\mu_{X_0^t|Z=z}$ to denote the Radon-Nikodym derivative of Y_0^t with

respect to X_0^t given $Z = z$. Obviously, when Z is independent of X , $d\mu_{Y_0^t|Z=z}/d\mu_{X_0^t|Z=z} = d\mu_{Y_0^t|Z=z}/d\mu_{X_0^t}$.

We next present some basic notions and facts from information theory and introduce the corresponding notations. For more comprehensive expositions, we refer to [5, 14].

Let X, Y, Z be random variables defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which will be used to illustrate most of the notions and facts in this section (note that the same notations may have different connotations in other sections). Particularly in this paper, random variables can be discrete-valued with a probability mass function, real-valued with a probability density function or path-valued (more precisely, $C[0, \infty)$ -valued or $C[0, t]$ -valued).

By definition, for $\mathbb{E}[X|\sigma(Y, Z)]$, the conditional expectation of X with respect to the σ -algebra generated by Y and Z , there exists a $\sigma(Y) \otimes \sigma(Z)$ -measurable function $\Psi(\cdot, \cdot)$ such that $\Psi(Y, Z) = \mathbb{E}[X|\sigma(Y, Z)]$. For notational convenience, we will in this paper simply write $\mathbb{E}[X|\sigma(Y, Z)]$ as $\mathbb{E}[X|Y, Z]$, and $\Psi(y, z)$ as $\mathbb{E}[X|y, z]$ and furthermore, $\Psi(Y, z)$ as $\mathbb{E}[X|Y, z]$.

A *partition* of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a disjoint collection of elements of \mathcal{F} whose union is Ω . It is well known there is a one-to-one correspondence between finite partitions and finite sub- σ -algebras of \mathcal{F} . For a finite sub- σ -algebra $\mathcal{H} \subset \mathcal{F}$, let $\eta(\mathcal{H})$ denote the corresponding finite partition. The entropy of a finite partition $\xi = \{A_1, A_2, \dots, A_m\}$, denoted by $H(\xi)$, is defined as $H(\xi) \triangleq \sum_{i=1}^m -\mathbb{P}(A_i) \log \mathbb{P}(A_i)$, whereas the conditional entropy of ξ given another finite partition $\zeta = \{B_1, B_2, \dots, B_n\}$, denoted by $H(\xi|\zeta)$, is defined as $H(\xi|\zeta) \triangleq \sum_{j=1}^n \sum_{i=1}^m -\mathbb{P}(A_i \cap B_j) \log \mathbb{P}(A_i|B_j)$. The mutual information between the above-mentioned two partitions ξ and ζ , denoted by $I(\xi; \zeta)$, is defined as $I(\xi; \zeta) \triangleq \sum_{j=1}^n \sum_{i=1}^m -\mathbb{P}(A_i \cap B_j) \log \mathbb{P}(A_i \cap B_j) / \mathbb{P}(A_i) \mathbb{P}(B_j)$.

For the random variable X , we define

$$\eta(X) \triangleq \{\eta(\mathcal{H}) : \mathcal{H} \text{ is a finite sub-}\sigma\text{-algebra of } \sigma(X)\}.$$

The *entropy* of the random variable X , denoted by $H(X)$, is defined as

$$H(X) \triangleq \sup_{\xi \in \eta(X)} H(\xi).$$

The *conditional entropy* of Y given X , denoted by $H(Y|X)$, is defined as

$$H(Y|X) \triangleq \inf_{\xi \in \eta(X)} \sup_{\zeta \in \eta(Y)} H(\zeta|\xi).$$

Here, we note that if X and Y are independent, then obviously it holds that

$$H(Y|X) = H(Y). \tag{8}$$

The *mutual information* between X and Y , denoted by $I(X; Y)$, is defined as

$$I(X; Y) \triangleq \sup_{\xi \in \eta(X), \zeta \in \eta(Y)} I(\xi; \zeta).$$

A couple of properties of mutual information are in order. First, it can be shown, via a concavity argument, that the mutual information is always non-negative. Second, the mutual information is determined by the σ -algebras generated by the corresponding random variables; more specifically, for any random variables X', Y', X'', Y'' ,

$$I(X'; Y') = I(X'; Y'') \text{ if } \sigma(X') = \sigma(X'') \text{ and } \sigma(Y') = \sigma(Y'') \tag{9}$$

and

$$I(X'; Y') \leq I(X'; Y'') \text{ if } \sigma(X') \subset \sigma(X'') \text{ and } \sigma(Y') \subset \sigma(Y''). \quad (10)$$

For a quick example, we have $I(X; Y) = I(X, X; Y, Y + X)$ and $I(X; Y) \leq I(X; Y, Z)$.

It turns out that for the case that X, Y, Z are all discrete random variables, all the above-mentioned notions are well-defined and can be computed rather explicitly: $H(X)$ can be computed as $H(X) = \mathbb{E}[-\log p_X(X)]$, where $p_X(\cdot)$ denotes the probability mass function of X ; $H(Y|X)$ can be computed as $H(Y|X) = \mathbb{E}[-\log p_{Y|X}(Y|X)]$, where $p_{Y|X}(\cdot|\cdot)$ denotes the conditional probability mass function of Y given X ; $I(X; Y)$ can be computed as

$$I(X; Y) = \mathbb{E} \left[\log \frac{p_{Y|X}(Y|X)}{p_Y(Y)} \right]. \quad (11)$$

The mutual information is intimately related to entropy. As an example, one verifies that

$$I(X; Y) = H(Y) - H(Y|X). \quad (12)$$

Note that the quality (12) may fail if non-discrete random variables are involved, since the corresponding entropies $H(Y)$ and $H(Y|X)$ can be infinity. For the case of real-valued random variables with density, this issue can be circumvented using the notion of differential entropy, as elaborated below.

Now, let Y be a real-valued random variable with probability density function $f_Y(\cdot)$. The *differential entropy* of Y , denoted by $h(Y)$, is defined as $h(Y) \triangleq \mathbb{E}[-\log f_Y(Y)]$. And the *differential conditional entropy* of Y given a finite partition ξ , denoted by $h(Y|\xi)$, is defined as $h(Y|\xi) \triangleq \sum_{j=1}^n \mathbb{P}(A_j) \int f_{Y|A_j}(x) \log f_{Y|A_j}(x) dx$. The *differential conditional entropy* of Y given X (which may not be real-valued), denoted by $h(Y|X)$, is defined as $h(Y|X) \triangleq \inf_{\xi \in \eta(X)} h(Y|\xi)$ (which, similarly as in (8), reduces to $h(Y)$ if Y is independent of X); in particular, if the conditional probability density function $f_{Y|X}(\cdot|\cdot)$ exists, then $h(Y|X)$ can be explicitly computed as $\mathbb{E}[-\log f_{Y|X}(Y|X)]$. As mentioned before, the aforementioned failure of (12) can be salvaged with the notion of differential entropy:

$$I(X; Y) = h(Y) - h(Y|X). \quad (13)$$

Note that this equality, together with the fact that the mutual information is non-negative, immediately implies that

$$h(Y|X) \leq h(Y); \quad (14)$$

in other words, conditioning reduces the differential entropy (the same statement holds for entropy, which however is not needed in this paper). We will use the fact that for a given variance, a Gaussian distribution maximizes the differential entropy. More precisely, for a real-valued random variable Y with variance less than or equal to σ^2 , we have

$$h(X) \leq h(N) = \frac{1}{2} \log(2\pi e \sigma^2), \quad (15)$$

where N is a Gaussian random variable with mean 0 and variance σ^2 .

Here we emphasize that all the above-mentioned definitions naturally carry over to the setting where some/all of random variables are vector-valued. For a quick example, let

$Y = \{Y_1, Y_2, \dots, Y_n\}$, where each Y_i is a real-valued random variable with density. Then, the differential entropy $h(Y)$ of Y is defined as

$$h(Y) = h(Y_1, Y_2, \dots, Y_n) \triangleq \mathbb{E}[-\log f_{Y_1, Y_2, \dots, Y_n}(Y_1, Y_2, \dots, Y_n)],$$

where f_{Y_1, Y_2, \dots, Y_n} is the joint probability density function of Y_1, Y_2, \dots, Y_n . The *chain rule* for differential entropy states that

$$h(Y_1, Y_2, \dots, Y_n) = h(Y_1) + h(Y_2|Y_1) + \dots + h(Y_n|Y_1, Y_2, \dots, Y_{n-1}), \quad (16)$$

whose conditional version given a random variable X reads

$$h(Y_1, Y_2, \dots, Y_n|X) = h(Y_1|X) + h(Y_2|X, Y_1) + \dots + h(Y_n|X, Y_1, Y_2, \dots, Y_{n-1}). \quad (17)$$

Consider the continuous-time white Gaussian feedback channel in (7), the mutual information of which over the time interval $[0, T]$ can be computed as below (see, e.g., [24, 14]):

$$I(M; Y_0^T) = \begin{cases} \mathbb{E} \left[\log \frac{d\mu_{M, Y_0^T}}{d\mu_M \times \mu_{Y_0^T}}(M, Y_0^T) \right], & \text{if } \frac{d\mu_{M, Y_0^T}}{d\mu_M \times \mu_{Y_0^T}} \text{ exists,} \\ \infty, & \text{otherwise,} \end{cases} \quad (18)$$

where $d\mu_{M, Y_0^T}/d\mu_M \times \mu_{Y_0^T}$ denotes the Radon-Nikodym derivative of μ_{M, Y_0^T} with respect to $d\mu_M \times \mu_{Y_0^T}$. It turns out that the mutual information is intimately connected to the channel capacity, detailed below.

For $T, R, P > 0$, a (T, e^{TR}, P) -code for the above-mentioned continuous-time Gaussian channel consists of a set of integers $\mathcal{M} = \{1, 2, \dots, e^{TR}\}$, the *message set* for receiver, and an *encoding function*, $g : \mathcal{M} \rightarrow C[0, T]$, which satisfies the following average power constraint: with probability 1,

$$\frac{1}{T} \int_0^T \mathbb{E}[g^2(s, M, Y_0^s)] ds \leq P, \quad (19)$$

and a *decoding functions*, $h : C[0, T] \rightarrow \mathcal{M}$. The *average probability of error* for the (T, e^{TR}, P) -code is defined as

$$P_e^{(T)} = \frac{1}{e^{TR}} \sum_{M \in \mathcal{M}} \mathbb{P}(h(Y_0^T) \neq M \mid M \text{ sent}).$$

A rate R is said to be *achievable* for the channel if there exists a sequence of (T, e^{TR}, P) -codes with $P_e^{(T)} \rightarrow 0$ as $T \rightarrow \infty$. The *capacity* \mathcal{C} of the channel is the supremum of all such achievable rates.

The celebrated Shannon's channel coding theorem [26] states that

$$\mathcal{C} = \lim_{T \rightarrow \infty} \frac{1}{T} \sup_{M, g} I(M; Y_0^T),$$

where the supremum is over all choices of the message M (including its alphabet and distribution), and the encoding function g . As mentioned in Section 1, the capacity of the channel (7) is $P/2$, the same as that of the channel (6), where the feedback is absent.

3 Information-Theoretic Extensions

In this section, we will establish a sampling and approximation theorem for the channel (7) restricted to the time interval $[0, T]$:

$$Y(t) = \int_0^t g(s, M, Y_0^s) ds + B(t), \quad t \in [0, T], \quad (20)$$

which naturally connect such a channel with its discrete-time versions obtained via sampling and approximation.

The following regularity conditions may be imposed to our channel (20):

(a) The solution $\{Y(t)\}$ to the stochastic differential equation (20) uniquely exists.

(b)

$$\mathbb{P} \left(\int_0^T g^2(t, M, Y_0^t) dt < \infty \right) = \mathbb{P} \left(\int_0^T g^2(t, M, B_0^t) dt < \infty \right) = 1.$$

(c)

$$\int_0^T \mathbb{E}[|g(t, M, Y_0^t)|] dt < \infty.$$

(d) **The uniform Lipschitz condition:** There exists a constant $L > 0$ such that for any $0 \leq s_1, s_2, t_1, t_2 \leq T$, any Y_0^T, Z_0^T ,

$$|g(s_1, M, Y_0^{s_2}) - g(t_1, M, Z_0^{t_2})| \leq L(|s_1 - t_1| + \|Y_0^{s_2} - Z_0^{t_2}\|).$$

(e) **The uniform linear growth condition:** There exists a constant $L > 0$ such that for any M and any Y_0^T ,

$$|g(t, M, Y_0^t)| \leq L(1 + \|Y_0^t\|).$$

The following lemma says that Conditions (d)-(e) are stronger than Conditions (a)-(c). We remark that Conditions (d)-(e) are still rather mild assumptions: conditions of similar nature are typically imposed to guarantee the existence and uniqueness of the solution to a given stochastic differential equation, and it is possible that the solution may not uniquely exist if these two conditions are violated (see, e.g., [21]).

Lemma 3.1. *Assume Conditions (d)-(e). Then, there exists a unique strong solution of (7) with initial value $Y(0) = 0$. Moreover, there exists $\varepsilon > 0$ such that*

$$\mathbb{E}[e^{\varepsilon \|Y_0^T\|^2}] < \infty, \quad (21)$$

which immediately implies Conditions (b) and (c).

Proof. With Conditions (d)-(e), the proof of the existence and uniqueness of the solution to (7) is somewhat standard; see, e.g., Section 5.4 in [21]. So, in the following, we will only prove (21).

For the stochastic differential equation (7), applying Condition (e), we deduce that there exists $L_1 > 0$ such that

$$\|Y_0^T\| \leq \int_0^T L_1(1 + \|Y_0^t\|)dt + \|B_0^T\| \leq L_1T + \|B_0^T\| + \int_0^T L_1\|Y_0^t\|dt.$$

Then, applying the Gronwall inequality followed by a straightforward bounding analysis, we deduce that there exists $L_2 > 0$ such that

$$\|Y_0^T\| \leq (L_1T + \|B_0^T\|)e^{\int_0^T L_1dt} = e^{L_1T}(L_1T + \|B_0^T\|) \leq L_2 + L_2\|B_0^T\|.$$

Now, for any $\varepsilon > 0$, applying Doob's submartingale inequality, we have

$$\begin{aligned} \mathbb{E}[e^{\varepsilon\|Y_0^T\|^2}] &\leq \mathbb{E}[e^{\varepsilon(L_2+L_2\|B_0^T\|)^2}] \\ &\leq \mathbb{E}[e^{2\varepsilon(L_2^2+L_2^2\|B_0^T\|^2)}] \\ &= e^{2\varepsilon L_2^2}\mathbb{E}[e^{2\varepsilon L_2^2\|B_0^T\|^2}] \\ &= e^{2\varepsilon L_2^2}\mathbb{E}[\sup_{0 \leq t \leq T} e^{2\varepsilon L_2^2 B(t)^2}] \\ &\leq 4e^{2\varepsilon L_2^2}\mathbb{E}[e^{2\varepsilon L_2^2 B(T)^2}], \end{aligned}$$

which is finite provided that ε is small enough. □

For any $n \in \mathbb{N}$, choose time points $t_0^{(n)}, t_1^{(n)}, \dots, t_n^{(n)} \in \mathbb{R}$ such that

$$0 = t_0^{(n)} < t_1^{(n)} < \dots < t_n^{(n)} = T,$$

and we define

$$\Delta_n \triangleq \{t_0^{(n)}, t_1^{(n)}, \dots, t_n^{(n)}\}.$$

For any time point sequence Δ_n , we will use δ_{Δ_n} to denote its maximal stepsize, namely,

$$\delta_{\Delta_n} \triangleq \max_{i=1,2,\dots,n} (t_i^{(n)} - t_{i-1}^{(n)}).$$

Δ_n is said to be *evenly spaced* if $t_i^{(n)} - t_{i-1}^{(n)} = T/n$ for all feasible i , and we will use the shorthand notation δ_n to denote its stepsize, i.e., $\delta_n \triangleq t_1^{(n)} - t_0^{(n)} = T/n$. Apparently, evenly spaced time point sequences are natural choices for a conversion from a continuous-time setting to a discrete-time setting; particularly, such a time point sequence is used in the classical Shannon-Nyquist sampling theorem.

In the following, we will present our extensions of the Shannon-Nyquist sampling theorem via sampling or approximation with respect to Δ_n .

3.1 Extension via Sampling

In this section, we will establish a sampling theorem for the channel (7), which naturally connect such a channel with their discrete-time versions obtained via sampling.

Sampling the channel (7) over the time interval $[0, T]$ with respect to Δ_n , we obtain its sampled discrete-time version as follows:

$$Y(t_i^{(n)}) = \int_0^{t_i^{(n)}} g(s, M, Y_0^s) ds + B(t_i^{(n)}), \quad i = 0, 1, \dots, n. \quad (22)$$

Roughly speaking, the following sampling theorem states that for any sequence of “increasingly finer” samplings, the mutual information of the sampled discrete-time channel (22) will converge to that of the original channel (20).

Theorem 3.2. 1) *Assume Conditions (a)-(c). Suppose that $\Delta_n \subset \Delta_{n+1}$ for all n and that $\delta_{\Delta_n} \rightarrow 0$ as n tends to infinity. Then, we have*

$$\lim_{n \rightarrow \infty} I(M; Y(\Delta_n)) = I(M; Y_0^T),$$

where we recall from Section 2 that $Y(\Delta_n) = \{Y(t_0^{(n)}), Y(t_1^{(n)}), \dots, Y(t_n^{(n)})\}$.

2) *Assume Conditions (d)-(e). Suppose that $\{\Delta_n\}$ with $\delta_{\Delta_n} \rightarrow 0$ as n tends to infinity. Then, we have*

$$\lim_{n \rightarrow \infty} I(M; Y(\Delta_n)) = I(M; Y_0^T).$$

Proof. The proof is rather technical and lengthy, and thereby postponed to Section 5. \square

Remark 3.3. Regarding the assumptions of Theorem 3.2, Conditions (a)-(c) in 1) are rather weak, but the condition that “ $\Delta_n \subset \Delta_{n+1}$ for all n ” is somewhat restrictive, which, in particular, is not satisfied by the set $\{\Delta_n\}$ of all evenly spaced time point sequences. As shown in 2), this condition can be removed at the expenses of the extra regularity conditions (d)-(e): The same theorem holds as long as the stepsize of the sampling tends to 0, which, in particular, is satisfied by the set of all evenly spaced time point sequences.

3.2 Extension via Approximation

In this section, we will establish an approximation theorem for the channel (20), which naturally connect such a channel with their discrete-time versions obtained via approximation.

An application of the Euler-Maruyama approximation [13] with respect to Δ_n to (20) will yield a discrete-time sequence $\{Y^{(n)}(t_i^{(n)}) : i = 0, 1, \dots, n\}$ and a continuous-time process $\{Y^{(n)}(t) : t \in [0, T]\}$, a linear interpolation of $\{Y(t_i^{(n)})\}$, as follows: Initializing with $Y^{(n)}(0) = 0$, we recursively compute, for each $i = 0, 1, \dots, n - 1$,

$$Y^{(n)}(t_{i+1}^{(n)}) = Y^{(n)}(t_i^{(n)}) + \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} g(s, M, Y_0^{(n), t_i^{(n)}}) ds + B(t_{i+1}^{(n)}) - B(t_i^{(n)}), \quad (23)$$

$$Y^{(n)}(t) = Y^{(n)}(t_i^{(n)}) + \frac{t - t_i^{(n)}}{t_{i+1}^{(n)} - t_i^{(n)}} (Y^{(n)}(t_{i+1}^{(n)}) - Y^{(n)}(t_i^{(n)})), \quad t_i^{(n)} \leq t \leq t_{i+1}^{(n)}. \quad (24)$$

We will first show the following lemma, which is parallel to Lemma 3.1.

Lemma 3.4. *Assume Conditions (d)-(e). Then, there exists $\varepsilon > 0$ and a constant $C > 0$ such that for all n ,*

$$\mathbb{E}[e^{\varepsilon \|Y_0^{(n),T}\|^2}] < C. \quad (25)$$

Proof. A discrete-time version of the proof of Lemma 3.1 implies that there exists $\varepsilon > 0$ and a constant $C > 0$ such that for all n

$$\mathbb{E}[e^{\varepsilon \sup_{i \in \{0,1,\dots,n\}} (Y^{(n)}(t_i^{(n)}))^2}] < C,$$

which, together with (24), immediately implies (25). \square

We also need the following lemma, which is parallel to Theorem 10.2.2 in [13].

Lemma 3.5. *Assume Conditions (d)-(e). Then, there exists a constant $C > 0$ such that for all n ,*

$$\mathbb{E}[\|Y_0^{(n),T} - Y_0^T\|^2] \leq C\delta_{\Delta_n}.$$

Proof. Note that for any n , we have

$$Y(t_{i+1}^{(n)}) = Y(t_i^{(n)}) + \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} g(s, M, Y_0^s) ds + B(t_{i+1}^{(n)}) - B(t_i^{(n)}),$$

and

$$Y^{(n)}(t_{i+1}^{(n)}) = Y^{(n)}(t_i^{(n)}) + \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} g(s, M, Y_0^{(n),t_i^{(n)}}) ds + B(t_{i+1}^{(n)}) - B(t_i^{(n)}).$$

It then follows that

$$Y(t_{i+1}^{(n)}) - Y^{(n)}(t_{i+1}^{(n)}) = Y(t_i^{(n)}) - Y^{(n)}(t_i^{(n)}) + \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (g(s, M, Y_0^s) - g(s, M, Y_0^{(n),t_i^{(n)}})) ds. \quad (26)$$

Now, for any t , choose n_0 such that $t_{n_0}^{(n)} \leq t < t_{n_0+1}^{(n)}$. Now, a recursive application of (26), coupled with Conditions (d) and (e), yields that for some $L > 0$,

$$\begin{aligned} Y(t) - Y^{(n)}(t) &= \sum_{i=0}^{n_0} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (g(s, M, Y_0^s) - g(t_i^{(n)}, M, Y_0^{(n),t_i^{(n)}})) ds + \int_{t_{n_0+1}^{(n)}}^t (g(s, M, Y_0^s) - g(t_i^{(n)}, M, Y_0^{(n),t_{n_0+1}^{(n)}})) ds \\ &\leq \sum_{i=0}^{n_0} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} L|s - t_i^{(n)}| + L\|Y_0^s - Y_0^{(n),s}\| + L\|Y_0^{(n),s} - Y_0^{(n),t_i^{(n)}}\| ds \\ &\quad + \int_{t_{n_0+1}^{(n)}}^t L|s - t_{n_0+1}^{(n)}| + L\|Y_0^s - Y_0^{(n),s}\| + L\|Y_0^{(n),s} - Y_0^{(n),t_{n_0+1}^{(n)}}\| ds. \end{aligned}$$

Noticing that for any s with $t_i^{(n)} \leq s < t_{i+1}^{(n)}$, we have

$$\|Y_0^{(n),s} - Y_0^{(n),t_i^{(n)}}\|^2 \leq |Y^{(n)}(t_{i+1}^{(n)}) - Y^{(n)}(t_i^{(n)})|^2 \leq 2 \left| \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} g(s, M, Y_0^{(n),t_i^{(n)}}) ds \right|^2 + 2|B(t_{i+1}^{(n)}) - B(t_i^{(n)})|^2,$$

which, together with Condition (e) and the fact that for all n and i ,

$$\mathbb{E}[|B(t_{i+1}^{(n)}) - B(t_i^{(n)})|^2] = O(\delta_{\Delta_n}), \quad (27)$$

implies that

$$\mathbb{E}[\|Y_0^{(n),s} - Y_0^{(n),t_i^{(n)}}\|^2] = O(\delta_{\Delta_n}). \quad (28)$$

Noting that the constants in the two terms $O(\delta_{\Delta_n})$ in (27) and (28) can be chosen uniform over all n , a usual argument with the Gronwall inequality applied to $\mathbb{E}[\|Y_0^t - Y_0^{(n),t}\|^2]$ completes the proof of the lemma. \square

We are now ready to state and prove the following theorem:

Theorem 3.6. *Assume Conditions (d)-(e). Then, we have*

$$\lim_{n \rightarrow \infty} I(M; Y^{(n)}(\Delta_n)) = I(M; Y_0^T).$$

Proof. The proof is rather technical and lengthy, and thereby postponed to Section 6. \square

Remark 3.7. When there is no feedback, Theorem 3.6 boils down to Theorem 3.2: obviously it holds that for any feasible i ,

$$Y^{(n)}(t_i^{(n)}) = Y(t_i^{(n)}),$$

which means that Theorem 3.6 actually states

$$\lim_{n \rightarrow \infty} I(M; Y(\Delta_n)) = I(M; Y_0^T),$$

which is precisely the conclusion of Theorem 3.2.

4 Applications of Our Results

In this section, we discuss possible applications of our extensions. Clearly, establishing causality-preserving connections between continuous-time and discrete-time Gaussian channels with possible feedback, our results may help channel the ideas and techniques in the discrete-time regime to the continuous-time one, which is evidenced by some preliminary results collected in the recent work [18, 19]. More specifically, Theorem 3.2 has been used in [18] to conduct a spectral analysis of a continuous-time Gaussian channel to recover a classical information-theoretic formula with an elementary proof; moreover, as opposed to the sampling approach based on the Shannon-Nyquist sampling theorem, the so-called approximation approach, a new approach based on Theorem 3.6, has been employed in [19] to analyze the capacity regions of continuous-time Gaussian channels with multiple users. Below, we use an example to illustrate how our the causality-preserving connections prescribed by the extensions can be used.

Consider the continuous-time white Gaussian feedback channel as in (7) and assume that Conditions (d) and (e) are satisfied, and moreover the following average power constraint is satisfied: there exists $P > 0$ such that for any T ,

$$\frac{1}{T} \int_0^T \mathbb{E}[g^2(s, M, Y_0^s)] ds \leq P. \quad (29)$$

It has been established in [16] that

$$I(M; Y_0^T) = \frac{1}{2} \int_0^T \mathbb{E}[g^2(s, M, Y_0^s)] - \mathbb{E}[\mathbb{E}^2[g(s, M, Y_0^s)|Y_0^s]] ds,$$

which, together with (29), immediately implies that

$$I(M; Y_0^T) \leq \frac{PT}{2}. \quad (30)$$

Below, we will use Theorem 3.6 and some basic facts for discrete-time Gaussian channels to derive (30), which, in combination with the proven fact that one can choose g , M and sufficiently large T so that $I(M; Y_0^T)/T$ is arbitrarily close to $P/2$ (see Theorem 6.4.1 in [14]), implies that the capacity \mathcal{C} of the channel (7) is indeed $P/2$.

First of all, for fixed $T > 0$, consider the evenly spaced Δ_n with stepsize $\delta_n = T/n$. Applying the Euler-Maruyama approximation (23) to the channel (20) over the time window $[0, T]$, we obtain

$$Y^{(n)}(t_{i+1}^{(n)}) = Y^{(n)}(t_i^{(n)}) + \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} g(s, M, Y_0^{(n), t_i^{(n)}}) ds + B(t_{i+1}^{(n)}) - B(t_i^{(n)}). \quad (31)$$

By Theorem 3.6, we have

$$I(M; Y_0^T) = \lim_{n \rightarrow \infty} I(M; Y^{(n)}(\Delta_n)). \quad (32)$$

Note that, it can be easily verified that

$$\begin{aligned} h(\{Y^{(n)}(t_i^{(n)}) - Y^{(n)}(t_{i-1}^{(n)})\}_{i=1}^n | M) &\stackrel{(a)}{=} \sum_{i=1}^n h(Y^{(n)}(t_i^{(n)}) - Y^{(n)}(t_{i-1}^{(n)}) | \{Y^{(n)}(t_j^{(n)}) - Y^{(n)}(t_{j-1}^{(n)})\}_{j=1}^{i-1}, M) \\ &\stackrel{(b)}{=} h(\{B(t_i^{(n)}) - B(t_{i-1}^{(n)})\}_{i=1}^n | M) \\ &\stackrel{(c)}{=} h(\{B(t_i^{(n)}) - B(t_{i-1}^{(n)})\}_{i=1}^n) \\ &\stackrel{(d)}{=} \sum_{i=1}^n h(B(t_i^{(n)}) - B(t_{i-1}^{(n)})), \end{aligned}$$

where we have used the conditional chain rule for differential entropy for (a); and (31) for (b); and the independence between $\{B(t)\}$ and M for (c); and (16) and the fact that a Brownian motion has independent increments for (d). Then, using (9), (13), (16) and (14), we find that

$$\begin{aligned} I(M; Y^{(n)}(\Delta_n)) &= I(M; \{Y^{(n)}(t_i^{(n)}) - Y^{(n)}(t_{i-1}^{(n)})\}_{i=1}^n) \\ &= h(\{Y^{(n)}(t_i^{(n)}) - Y^{(n)}(t_{i-1}^{(n)})\}_{i=1}^n) - h(\{Y^{(n)}(t_i^{(n)}) - Y^{(n)}(t_{i-1}^{(n)})\}_{i=1}^n | M) \\ &= \sum_{i=1}^n h(Y^{(n)}(t_i^{(n)}) - Y^{(n)}(t_{i-1}^{(n)}) | \{Y^{(n)}(t_j^{(n)}) - Y^{(n)}(t_{j-1}^{(n)})\}_{j=1}^{i-1}) - \sum_{i=1}^n h(B(t_i^{(n)}) - B(t_{i-1}^{(n)})) \\ &\leq \sum_{i=1}^n h(Y^{(n)}(t_i^{(n)}) - Y^{(n)}(t_{i-1}^{(n)})) - \sum_{i=1}^n h(B(t_i^{(n)}) - B(t_{i-1}^{(n)})). \end{aligned}$$

Next, using the Cauchy-Schwarz inequality, we reach

$$\begin{aligned}
\text{Var}(Y^{(n)}(t_i^{(n)}) - Y^{(n)}(t_{i-1}^{(n)})) &= \mathbb{E}[(Y^{(n)}(t_i^{(n)}) - Y^{(n)}(t_{i-1}^{(n)}))^2] \\
&\leq \delta_n \mathbb{E} \left[\int_{t_{i-1}^{(n)}}^{t_i^{(n)}} g^2(s, M, Y_0^{(n), t_{i-1}^{(n)}}) ds \right] + \mathbb{E}[(B(t_i^{(n)}) - B(t_{i-1}^{(n)}))^2] \\
&= \delta_n \int_{t_{i-1}^{(n)}}^{t_i^{(n)}} \mathbb{E} \left[g^2(s, M, Y_0^{(n), t_{i-1}^{(n)}}) \right] ds + \delta_n,
\end{aligned}$$

which, together with (15), further implies that

$$I(M; Y^{(n)}(\Delta_n)) \leq \frac{1}{2} \sum_{i=0}^n \log \left(1 + \int_{t_{i-1}^{(n)}}^{t_i^{(n)}} \mathbb{E} \left[g^2(s, M, Y_0^{(n), t_{i-1}^{(n)}}) \right] ds \right) \quad (33)$$

$$\leq \frac{1}{2} \sum_{i=0}^n \int_{t_{i-1}^{(n)}}^{t_i^{(n)}} \mathbb{E} \left[g^2(s, M, Y_0^{(n), t_{i-1}^{(n)}}) \right] ds. \quad (34)$$

It then follows from (32), Condition (d), Lemma 3.5 and (29) that

$$I(M; Y_0^T) \leq \frac{1}{2} \int_0^T \mathbb{E}[g^2(s, M, Y_0^s)] ds \leq \frac{PT}{2}, \quad (35)$$

as desired.

5 Proof of Theorem 3.2

First of all, recall from Section 2 that for a stochastic process $\{X(s)\}$ and any $t \in \mathbb{R}_+$, we use $\mu_{X_0^t}$ to denote the distribution on $C[0, t]$ induced by X_0^t . Throughout the proof, we only have to deal with the case $t = T$, and so we will simply write $\mu_{X_0^T}$ as μ_X for notational simplicity. And for cosmetic reasons, we will write the summation $\sum_m(\cdot)p_M(m)$ as the integral $\int(\cdot)d\mu_M(m)$.

5.1 Proof of 1)

Note that an application of Theorem 7.14 of [20] with Conditions (b) and (c) yields that

$$\mathbb{P} \left(\int_0^T \mathbb{E}^2[g(t, M, Y_0^t) | Y_0^t] dt < \infty \right) = 1. \quad (36)$$

Then one verifies that the assumptions of Lemma 7.7 of [20] are all satisfied (this lemma is stated under very general assumptions, which are exactly Conditions (b), (c) and (36) when restricted to our settings), which implies that for any m ,

$$\mu_Y \sim \mu_{Y|M=m} \sim \mu_B, \quad (37)$$

and moreover, with probability 1,

$$\frac{d\mu_{Y|M}}{d\mu_B}(Y_0^T|M) = \frac{1}{\mathbb{E}[e^{\rho_1(M, Y_0^T)}|Y_0^T, M]}, \quad \frac{d\mu_Y}{d\mu_B}(Y_0^T) = \frac{1}{\mathbb{E}[e^{\rho_1(M, Y_0^T)}|Y_0^T]}, \quad (38)$$

where

$$\rho_1(m, Y_0^T) \triangleq - \int_0^T g(s, m, Y_0^s) dY(s) + \frac{1}{2} \int_0^T g(s, m, Y_0^s)^2 ds.$$

Here we note that $\mathbb{E}[e^{\rho_1(M, Y_0^T)}|Y_0^T, M]$ is in fact equal to $e^{\rho_1(M, Y_0^T)}$, but we keep it the way it is as above for an easy comparison.

Note that it follows from $\mathbb{E}[d\mu_B/d\mu_Y(Y_0^T)] = 1$ that $\mathbb{E}[e^{\rho_1(M, Y_0^T)}] = 1$, which is equivalent to

$$\mathbb{E}[e^{-\int_0^T g(s, M, Y_0^s) dB(s) - \frac{1}{2} \int_0^T g(s, M, Y_0^s)^2 ds}] = 1. \quad (39)$$

Then, a parallel argument as in the proof of Theorem 7.1 of [20] (which requires the condition (39)) further implies that, for any n ,

$$\frac{d\mu_{Y(\Delta_n)|M}}{d\mu_{B(\Delta_n)}}(Y(\Delta_n)|M) = \frac{1}{\mathbb{E}[e^{\rho_1(M, Y_0^T)}|Y(\Delta_n), M]}, \quad \frac{d\mu_{Y(\Delta_n)}}{d\mu_{B(\Delta_n)}}(Y(\Delta_n)) = \frac{1}{\mathbb{E}[e^{\rho_1(M, Y_0^T)}|Y(\Delta_n)]}, \quad \text{a.s.} \quad (40)$$

Notice that it can be easily checked that $e^{\rho_1(M, Y_0^T)}$ is integrable, which, together with the assumption that $\Delta_n \subset \Delta_{n+1}$ for all n , further implies that

$$\{\mathbb{E}[e^{\rho_1(M, Y_0^T)}|Y(\Delta_n), M]\}, \quad \{\mathbb{E}[e^{\rho_1(M, Y_0^T)}|Y(\Delta_n)]\}$$

are both martingales, and therefore, by Doob's martingale convergence theorem [7],

$$\frac{d\mu_{Y(\Delta_n)|M}}{d\mu_{B(\Delta_n)}}(Y(\Delta_n)|M) \rightarrow \frac{d\mu_{Y|M}}{d\mu_B}(Y_0^T|M), \quad \frac{d\mu_{Y(\Delta_n)}}{d\mu_{B(\Delta_n)}}(Y(\Delta_n)) \rightarrow \frac{d\mu_Y}{d\mu_B}(Y_0^T), \quad \text{a.s.}$$

Now, by Jensen's inequality, we have

$$\mathbb{E}[\rho_1(M, Y_0^T)|Y(\Delta_n), M] \leq \log \mathbb{E}[e^{\rho_1(M, Y_0^T)}|Y(\Delta_n), M], \quad (41)$$

and, by the fact that $\log x \leq x$ for any $x > 0$, we have

$$\log \mathbb{E}[e^{\rho_1(M, Y_0^T)}|Y(\Delta_n), M] \leq \mathbb{E}[e^{\rho_1(M, Y_0^T)}|Y(\Delta_n), M]. \quad (42)$$

It then follows from (41) and (42) that

$$\left| \log \mathbb{E}[e^{\rho_1(M, Y_0^T)}|Y(\Delta_n), M] \right| \leq \left| \mathbb{E}[\rho_1(M, Y_0^T)|Y(\Delta_n), M] \right| + \mathbb{E}[e^{\rho_1(M, Y_0^T)}|Y(\Delta_n), M].$$

Applying the general Lebesgue dominated convergence theorem (see, e.g., Theorem 19 on Page 89 of [25]), we then have

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\log \frac{d\mu_{Y|M}}{d\mu_B}(Y(\Delta_n)|M) \right] = \mathbb{E}[\log \mathbb{E}[e^{\rho_1(M, Y_0^T)}|Y_0^T, M]] = \mathbb{E} \left[\log \frac{d\mu_{Y|M}}{d\mu_B}(Y_0^T|M) \right]. \quad (43)$$

A completely parallel argument yields that

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\log \frac{d\mu_{Y(\Delta_n)}}{d\mu_{B(\Delta_n)}}(Y(\Delta_n)) \right] = \mathbb{E}[\log \mathbb{E}[\exp(\rho_1(M, Y_0^T)) | Y_0^T]] = \mathbb{E} \left[\log \frac{d\mu_Y}{d\mu_B}(Y_0^T) \right]. \quad (44)$$

Next, by the definition of mutual information, we have

$$\begin{aligned} I(M; Y(\Delta_n)) &= \mathbb{E} \left[\log f_{Y(\Delta_n)|M}(Y(\Delta_n)|M) \right] - \mathbb{E} \left[\log f_{Y(\Delta_n)}(Y(\Delta_n)) \right] \\ &= \mathbb{E} \left[\log \frac{d\mu_{Y(\Delta_n)|M}}{d\mu_{B(\Delta_n)}}(Y(\Delta_n)|M) \right] - \mathbb{E} \left[\log \frac{d\mu_{Y(\Delta_n)}}{d\mu_{B(\Delta_n)}}(Y(\Delta_n)) \right], \end{aligned} \quad (45)$$

and

$$\begin{aligned} I(M; Y_0^T) &= \mathbb{E} \left[\log \frac{d\mu_{M, Y_0^T}}{d\mu_M \times \mu_{Y_0^T}}(M, Y_0^T) \right] \\ &= \mathbb{E} \left[\log \frac{d\mu_{Y|M}}{d\mu_B}(Y_0^T|M) \right] - \mathbb{E} \left[\log \frac{d\mu_Y}{d\mu_B}(Y_0^T) \right], \end{aligned} \quad (46)$$

where the well-definedness of the Radon-Nikodym derivatives are guaranteed by (37).

Finally, with (43), (44), (45) and (46), we conclude that

$$\lim_{n \rightarrow \infty} I(M; Y(\Delta_n)) = \mathbb{E} \left[\log \frac{d\mu_{Y|M}}{d\mu_B}(Y_0^T|M) \right] - \mathbb{E} \left[\log \frac{d\mu_Y}{d\mu_B}(Y_0^T) \right] = I(M; Y_0^T),$$

as desired.

5.2 Proof of 2)

We proceed in the following steps.

Step 1. In this step, we establish the theorem assuming that there exists $C > 0$ such that for all $m \in \mathcal{M}$ and all $y_0^T \in C[0, T]$,

$$\int_0^T g^2(s, m, y_0^s) ds < C. \quad (47)$$

By (45), (38) and (40), we have

$$\begin{aligned} I(M; Y(\Delta_n)) &= \mathbb{E} \left[\log \frac{d\mu_{Y(\Delta_n)|M}}{d\mu_{B(\Delta_n)}}(Y(\Delta_n)|M) \right] - \mathbb{E} \left[\log \frac{d\mu_{Y(\Delta_n)}}{d\mu_{B(\Delta_n)}}(Y(\Delta_n)) \right] \\ &= -\mathbb{E}[\log \mathbb{E}[e^{\rho_1(M, Y_0^T)} | Y(\Delta_n), M]] + \mathbb{E}[\log \mathbb{E}[e^{\rho_1(M, Y_0^T)} | Y(\Delta_n)]]. \end{aligned}$$

Step 1.1. In this step, we prove that as n tends to infinity,

$$\log \mathbb{E}[e^{\rho_1(M, Y_0^T)} | Y(\Delta_n), M] \rightarrow \rho_1(M, Y_0^T), \quad (48)$$

in probability.

Let $\bar{Y}_{\Delta_n, 0}^T$ denote the piecewise linear version of Y_0^T with respect to Δ_n ; more precisely, for any $i = 0, 1, \dots, n$, let $\bar{Y}_{\Delta_n}(t_i^{(n)}) = Y(t_i^{(n)})$, and for any $t_{i-1}^{(n)} < s < t_i^{(n)}$ with $s = \lambda t_{i-1}^{(n)} + (1 - \lambda)t_i^{(n)}$

for some $0 < \lambda < 1$, let $\bar{Y}_{\Delta_n}(s) = \lambda Y(t_{i-1}^{(n)}) + (1 - \lambda)Y(t_i^{(n)})$. Let $\bar{g}_{\Delta_n}(s, M, \bar{Y}_{\Delta_n,0}^s)$ denote the piecewise “flat” version of $g(s, M, \bar{Y}_{\Delta_n,0}^s)$ with respect to Δ_n ; more precisely, for any $t_{i-1}^{(n)} \leq s < t_i^{(n)}$, $\bar{g}_{\Delta_n}(s, M, \bar{Y}_{\Delta_n,0}^s) = g(t_{i-1}^{(n)}, M, \bar{Y}_{\Delta_n,0}^{t_{i-1}^{(n)}})$.

Letting

$$\rho_2(\Delta_n, m, Y_0^T) \triangleq - \int_0^T \bar{g}_{\Delta_n}(s, m, \bar{Y}_{\Delta_n,0}^s) dY(s) + \frac{1}{2} \int_0^T \bar{g}_{\Delta_n}^2(s, m, \bar{Y}_{\Delta_n,0}^s) ds,$$

we have

$$\begin{aligned} \log \mathbb{E}[e^{\rho_1(M, Y_0^T)} | Y(\Delta_n), M] &= \log \mathbb{E}[e^{\rho_2(\Delta_n, M, Y_0^T) + \rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)} | Y(\Delta_n), M] \\ &= \log e^{\rho_2(\Delta_n, M, Y_0^T)} \mathbb{E}[e^{\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)} | Y(\Delta_n), M] \\ &= \rho_2(\Delta_n, M, Y_0^T) + \log \mathbb{E}[e^{\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)} | Y(\Delta_n), M], \end{aligned}$$

where we have used the fact that

$$\mathbb{E}[e^{\rho_2(\Delta_n, M, Y_0^T)} | Y(\Delta_n), M] = e^{\rho_2(\Delta_n, M, Y_0^T)},$$

since $\rho_2(\Delta_n, M, Y_0^T)$ only depends on M and $Y(\Delta_n)$.

We now prove the following convergence:

$$\lim_{n \rightarrow \infty} \mathbb{E} [(\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T))^2] = 0, \quad (49)$$

which will imply that

$$\rho_2(\Delta_n, M, Y_0^T) \rightarrow \rho_1(M, Y_0^T)$$

in probability. To prove (49), we note that

$$\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T) = - \int_0^T (g(s) - \bar{g}_{\Delta_n}(s)) dB(s) - \frac{1}{2} \int_0^T (g(s) - \bar{g}_{\Delta_n}(s))^2 ds,$$

where we have rewritten $g(s, M, Y_0^s)$ as $g(s)$, and $\bar{g}_{\Delta_n}(s, M, \bar{Y}_{\Delta_n,0}^s)$ as $\bar{g}_{\Delta_n}(s)$. It then follows that (49) boils down to

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\left(- \int_0^T (g(s) - \bar{g}_{\Delta_n}(s)) dB(s) - \frac{1}{2} \int_0^T (g(s) - \bar{g}_{\Delta_n}(s))^2 ds \right)^2 \right] = 0. \quad (50)$$

To establish (50), notice that, by the Itô isometry [23], we have

$$\mathbb{E} \left[\left(\int_0^T (g(s) - \bar{g}_{\Delta_n}(s)) dB(s) \right)^2 \right] = \mathbb{E} \left[\int_0^T (g(s) - \bar{g}_{\Delta_n}(s))^2 ds \right],$$

which means we only need to prove that

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\left(\int_0^T (g(s) - \bar{g}_{\Delta_n}(s))^2 ds \right)^2 \right] = 0. \quad (51)$$

To see this, we note that, by Conditions (d) and (e), there exists $L_1 > 0$ such that for any $s \in [0, T]$ with $t_{i-1}^{(n)} \leq s < t_i^{(n)}$,

$$\begin{aligned}
|g(s, M, \bar{Y}_{\Delta_n, 0}^s) - \bar{g}_{\Delta_n}(s, M, \bar{Y}_{\Delta_n, 0}^s)| &= |g(s, M, \bar{Y}_{\Delta_n, 0}^s) - g(t_{i-1}^{(n)}, M, \bar{Y}_{\Delta_n, 0}^{t_{i-1}^{(n)}})| \\
&\leq L_1(|s - t_{i-1}^{(n)}| + \|\bar{Y}_{\Delta_n, 0}^s - \bar{Y}_{\Delta_n, 0}^{t_{i-1}^{(n)}}\|) \\
&\leq L_1(|s - t_{i-1}^{(n)}| + |Y(t_i^{(n)}) - Y(t_{i-1}^{(n)})|) \\
&\leq L_1\delta_{\Delta_n} + L_1\delta_{\Delta_n} + L_1\delta_{\Delta_n}\|Y_0^T\| + |B(t_i^{(n)}) - B(t_{i-1}^{(n)})|. \tag{52}
\end{aligned}$$

Moreover, by Lemma 3.1, $\|Y_0^T\|^4$ is integrable. And furthermore, one easily verifies that that

$$\mathbb{E}[(B(t_i^{(n)}) - B(t_{i-1}^{(n)}))^4] = 3(t_i^{(n)} - t_{i-1}^{(n)})^2 \leq 3\delta_{\Delta_n}^2. \tag{53}$$

It can be readily checked that (52) and (53) imply (51), which in turn implies (49), as desired.

We now prove the following convergence:

$$\lim_{n \rightarrow \infty} \mathbb{E}[|\mathbb{E}[e^{\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)} | Y(\Delta_n), M] - 1|] = 0, \tag{54}$$

which will imply that

$$\log \mathbb{E}[e^{\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)} | Y(\Delta_n), M] \rightarrow 0$$

in probability and furthermore (48). To establish (54), we first note that

$$\begin{aligned}
&\mathbb{E}[|\mathbb{E}[e^{\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)} | Y(\Delta_n), M] - 1|] \\
&\leq \mathbb{E}[|\mathbb{E}[e^{\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)} - 1] | Y(\Delta_n), M|] \\
&= \mathbb{E}[|e^{\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)} - 1|] \\
&\leq \mathbb{E}\left[|\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)| e^{|\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)|}\right] \\
&\leq \mathbb{E}\left[|\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)|^2\right] \mathbb{E}\left[e^{2|\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)|}\right].
\end{aligned}$$

It then follows from (49) that, to prove (54), we only need to prove that if δ_{Δ_n} is small enough,

$$\mathbb{E}\left[e^{2|\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)|}\right] < \infty. \tag{55}$$

Since

$$\mathbb{E}\left[e^{2|\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)|}\right] \leq \mathbb{E}\left[e^{2(\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T))}\right] + \mathbb{E}\left[e^{-2(\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T))}\right], \tag{56}$$

we only have to prove that the two terms in the above upper bound are both finite provided that δ_{Δ_n} is small enough. Note that for the first term, applying the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\mathbb{E}[e^{2(\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T))}] &= \mathbb{E}[e^{\int_0^T 2(g(s) - \bar{g}_{\Delta_n}(s))dB(s) - \int_0^T 4(g(s) - \bar{g}_{\Delta_n}(s))^2 ds + 3 \int_0^T (g(s) - \bar{g}_{\Delta_n}(s))^2 ds}] \\
&\leq \mathbb{E}[e^{\int_0^T 4(g(s) - \bar{g}_{\Delta_n}(s))dB(s) - \int_0^T 8(g(s) - \bar{g}_{\Delta_n}(s))^2 ds}] \mathbb{E}[e^{6 \int_0^T (g(s) - \bar{g}_{\Delta_n}(s))^2 ds}].
\end{aligned}$$

Then, an application of Fatou's lemma yields that

$$\mathbb{E}[e^{\int_0^T 4(g(s) - \bar{g}_{\Delta_n}(s))dB(s) - \int_0^T 8(g(s) - \bar{g}_{\Delta_n}(s))^2 ds}] \leq 1. \tag{57}$$

and by (52), we deduce that there exists $L_3 > 0$ such that

$$\mathbb{E}[e^{6 \int_0^T (g(s) - \bar{g}_{\Delta_n}(s))^2 ds}] \leq e^{L_3 \delta_{\Delta_n}^2} \mathbb{E}[e^{L_3 \|B_0^{\delta_{\Delta_n}}\|^2}] \mathbb{E}[e^{L_3 \delta_{\Delta_n}^2 \|Y_0^T\|^2}].$$

Note that it follows from Doob's submartingale inequality that if δ_{Δ_n} is small enough,

$$\mathbb{E}[e^{L_3 \|B_0^{\delta_{\Delta_n}}\|^2}] < \infty.$$

Furthermore, via a usual argument with the Gronwall inequality, we infer that there exists $L_2 > 0$ such that

$$\|Y_0^T\| \leq L_2(1 + \|B_0^T\|),$$

which, together with Doob's submartingale inequality, implies that if δ_{Δ_n} is small enough,

$$\mathbb{E}[e^{L_3 \delta_{\Delta_n}^2 \|Y_0^T\|^2}] < \infty.$$

It then follows that for the first term in (56)

$$\mathbb{E}[e^{2(\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T))}] < \infty. \quad (58)$$

A completely parallel argument will yield that for the second term in (56)

$$\mathbb{E}\left[e^{-2(\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T))}\right] < \infty,$$

which, together with (58), immediately implies (55), which in turn implies (54), as desired.

Step 1.2. In this step, we prove that as n tends to infinity,

$$\log \mathbb{E}[e^{\rho_1(M, Y_0^T)} | Y(\Delta_n)] \rightarrow \log \mathbb{E}[e^{\rho_1(M, Y_0^T)} | Y_0^T], \quad (59)$$

in probability.

First, note that by Theorem 7.23 of [20], we have,

$$\frac{d\mu_Y}{d\mu_B}(Y_0^T) = \int \frac{d\mu_{Y|M}}{d\mu_B}(Y_0^T | m) d\mu_M(m),$$

where

$$\frac{d\mu_{Y|M}}{d\mu_B}(Y_0^T | m) = e^{\int_0^T g(s, m, Y_0^s) dY(s) - \frac{1}{2} \int_0^T g^2(s, m, Y_0^s) ds}.$$

It then follows from (38) that

$$\begin{aligned} \log \mathbb{E}[e^{\rho_1(M, Y_0^T)} | Y_0^T] &= -\log \int \frac{d\mu_{Y|M}}{d\mu_B}(Y_0^T | m) d\mu_M(m) \\ &= -\log \int e^{-\rho_1(m, Y_0^T)} d\mu_M(m). \end{aligned}$$

Similarly, we have

$$\begin{aligned} \frac{d\mu_{Y(\Delta_n)}}{d\mu_{B(\Delta_n)}}(Y(\Delta_n)) &= \int \frac{d\mu_{Y(\Delta_n)|M}}{d\mu_{B(\Delta_n)}}(Y(\Delta_n) | m) d\mu_M(m) \\ &= \int \frac{1}{\mathbb{E}[e^{\rho_1(M, Y_0^T)} | Y(\Delta_n), m]} d\mu_M(m). \end{aligned}$$

It then follows from (40) that

$$\log \mathbb{E}[e^{\rho_1(M, Y_0^T)} | Y(\Delta_n)] = -\log \int \frac{1}{\mathbb{E}[e^{\rho_1(M, Y_0^T)} | Y(\Delta_n), m]} d\mu_M(m).$$

Now, we consider the following difference:

$$\begin{aligned} & \int e^{-\rho_1(m, Y_0^T)} d\mu_M(m) - \int \frac{1}{\mathbb{E}[e^{\rho_1(M, Y_0^T)} | Y(\Delta_n), m]} d\mu_M(m) \\ &= \int e^{-\rho_1(m, Y_0^T)} - e^{-\rho_2(\Delta_n, m, Y_0^T)} d\mu_M(m) \\ & \quad + \int e^{-\rho_2(\Delta_n, m, Y_0^T)} \times \frac{\mathbb{E}[e^{\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^t)} | Y(\Delta_n), m] - 1}{\mathbb{E}[e^{\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^t)} | Y(\Delta_n), m]} d\mu_M(m). \end{aligned}$$

Applying the inequality that for any $x, y \in \mathbb{R}$,

$$|e^x - e^y| \leq |x - y|(e^x \vee e^y) \leq |x - y|(e^x + e^y), \quad (60)$$

we have, for some positive constant L ,

$$\begin{aligned} & \mathbb{E} \left[\left| \int e^{-\rho_1(m, Y_0^T)} - e^{-\rho_2(\Delta_n, m, Y_0^T)} d\mu_M(m) \right| \right] \\ & \leq \int \mathbb{E} \left[\left| e^{-\rho_1(m, Y_0^T)} - e^{-\rho_2(\Delta_n, m, Y_0^T)} \right| \right] d\mu_M(m) \\ & \leq \int \mathbb{E} \left[\left| \rho_1(m, Y_0^T) - \rho_2(m, Y_0^T) \right| \left(e^{-\rho_1(m, Y_0^T)} + e^{-\rho_2(m, Y_0^T)} \right) \right] d\mu_M(m) \\ & \leq \int \mathbb{E} \left[\left(\left| \int_0^T (g(m) - \bar{g}_{\Delta_n}(m)) dB(s) \right| + \left| \int_0^T (g(m) - \bar{g}_{\Delta_n}(m))(g(s) - \frac{1}{2}g(m) - \frac{1}{2}\bar{g}_{\Delta_n}(m)) ds \right| \right) \right. \\ & \quad \times \left. \left(e^{-\rho_1(m, Y_0^T)} + e^{-\rho_2(m, Y_0^T)} \right) \right] d\mu_M(m) \\ & \leq \int \mathbb{E} \left[\left(\left| \int_0^T (g(m) - \bar{g}_{\Delta_n}(m)) dB(s) \right| \right. \right. \\ & \quad \left. \left. + (L\delta_{\Delta_n} + L\delta_{\Delta_n} + L\delta_{\Delta_n} \|Y_0^T\| + \sup_{|s-t| \leq \delta_{\Delta_n}} |B(s) - B(t)| \int_0^T \left| g(s) - \frac{1}{2}g(m) - \frac{1}{2}\bar{g}_{\Delta_n}(m) \right| ds \right) \right. \\ & \quad \left. \times \left(e^{-\rho_1(m, Y_0^T)} + e^{-\rho_2(m, Y_0^T)} \right) \right] d\mu_M(m), \end{aligned}$$

where we have rewritten $g(s, m, Y_0^s)$ as $g(m)$, $\bar{g}_{\Delta_n}(s, m, \bar{Y}_{\Delta_n, 0}^s)$ as $\bar{g}_{\Delta_n}(m)$ for notational simplicity.

Now, using (52) and the Itô isometry, we deduce that

$$\lim_{n \rightarrow \infty} \int \mathbb{E} \left[\left| \int_0^T g(m) - \bar{g}_{\Delta_n}(m) dB(s) \right|^2 \right] d\mu_M(m) = 0, \quad (61)$$

and

$$\lim_{n \rightarrow \infty} \int \mathbb{E} \left[(L\delta_{\Delta_n} + L\delta_{\Delta_n} + L\delta_{\Delta_n} \|Y_0^T\| + \sup_{|s-t| \leq \delta_{\Delta_n}} |B(s) - B(t)|)^2 \right] d\mu_M(m) = 0. \quad (62)$$

Now, using a similar argument as above with (47) and Lemma 3.1, we can show that for any constant K ,

$$\mathbb{E}[e^{\int_0^T K \bar{g}_{\Delta_n}^2(s) ds}] = \mathbb{E}[e^{\int_0^T K (\bar{g}_{\Delta_n}(s) - g(s) + g(s))^2 ds}] = \mathbb{E}[e^{\int_0^T K (2(\bar{g}_{\Delta_n}(s) - g(s))^2 + 2g^2(s)) ds}] < \infty, \quad (63)$$

provided that n is large enough, which, coupled with a similar argument as in the derivation of (58), proves that for n large enough,

$$\int \mathbb{E} \left[\left(e^{-\rho_1(m, Y_0^T)} + e^{-\rho_2(\Delta_n, m, Y_0^T)} \right)^2 \right] d\mu_M(m) < \infty, \quad (64)$$

and furthermore

$$\int \mathbb{E} \left[\left(\int_0^T \left| g(s) - \frac{1}{2}g(m) - \frac{1}{2}\bar{g}_{\Delta_n}(m) \right| ds \right)^2 \times \left(e^{-\rho_1(m, Y_0^T)} + e^{-\rho_2(\Delta_n, m, Y_0^T)} \right)^2 \right] d\mu_M(m) < \infty. \quad (65)$$

It then immediately follows that

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\left| \int e^{-\rho_1(m, Y_0^T)} - e^{-\rho_2(\Delta_n, m, Y_0^T)} d\mu_M(m) \right| \right] = 0. \quad (66)$$

Now, using the shorthand notations as before, we have

$$\begin{aligned} & \mathbb{E} \left[\left| \int e^{-\rho_2(\Delta_n, m, Y_0^T)} \frac{\mathbb{E}[e^{\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)} | Y(\Delta_n), m] - 1}{\mathbb{E}[e^{\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)} | Y(\Delta_n), m]} d\mu_M(m) \right| \right] \\ &= \mathbb{E} \left[\left| \int \frac{\mathbb{E}[e^{\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)} - 1 | Y(\Delta_n), m]}{\mathbb{E}[e^{\rho_1(M, Y_0^T)} | Y(\Delta_n), m]} d\mu_M(m) \right| \right] \\ &\leq \mathbb{E} \left[\int \frac{\mathbb{E}[|e^{\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)} - 1| | Y(\Delta_n), m]}{\mathbb{E}[e^{\rho_1(M, Y_0^T)} | Y(\Delta_n), m]} d\mu_M(m) \right] \\ &\stackrel{(a)}{\leq} \mathbb{E} \left[\int \mathbb{E}[|-\rho_1(M, Y_0^T) + \rho_2(\Delta_n, M, Y_0^T)| e^{|\rho_1(M, Y_0^T) + \rho_2(\Delta_n, M, Y_0^T)|} | Y(\Delta_n), m] \right. \\ &\quad \times \mathbb{E}[e^{-\rho_1(M, Y_0^T)} | Y(\Delta_n), m] d\mu_M(m) \left. \right] \\ &= \mathbb{E} \left[\mathbb{E}[|-\rho_1(M, Y_0^T) + \rho_2(\Delta_n, M, Y_0^T)| e^{|\rho_1(M, Y_0^T) + \rho_2(\Delta_n, M, Y_0^T)|} | Y(\Delta_n), M] \right. \\ &\quad \times \mathbb{E}[e^{-\rho_1(M, Y_0^T)} | Y(\Delta_n), M] \left. \left(\frac{d\mu_Y}{d\mu_B}(Y_0^T) \right) / \left(\frac{d\mu_{Y|M}}{d\mu_B}(Y_0^T | M) \right) \right] \\ &= \mathbb{E} \left[|-\rho_1(M, Y_0^T) + \rho_2(\Delta_n, M, Y_0^T)| e^{|\rho_1(M, Y_0^T) + \rho_2(\Delta_n, M, Y_0^T)|} \mathbb{E}[e^{-\rho_1(M, Y_0^T)} | Y(\Delta_n), M] \right. \\ &\quad \times \mathbb{E} \left[\left(\frac{d\mu_Y}{d\mu_B}(Y_0^T) \right) / \left(\frac{d\mu_{Y|M}}{d\mu_B}(Y_0^T | M) \right) | Y(\Delta_n), M \right] \left. \right], \end{aligned}$$

where we have used Jensen's inequality and the convexity of the inverse function $1/x : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ in (a). Now, a similar argument as in (61)-(65), together with the well-known fact (see, e.g., Theorem 6.2.2 in [14]) that

$$\frac{d\mu_Y}{d\mu_B}(Y_0^T) = e^{\int_0^T \mathbb{E}[g(s) | Y_0^s] dY(s) - \frac{1}{2} \int_0^T \mathbb{E}^2[g(s) | Y_0^s] ds}, \quad \frac{d\mu_{Y|M}}{d\mu_B}(Y_0^T | M) = e^{\int_0^T g(s) dY(s) - \frac{1}{2} \int_0^T g^2(s) ds},$$

yields that

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\left| \int e^{-\rho_2(\Delta_n, m, Y_0^T)} \frac{\mathbb{E}[e^{\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)} | Y(\Delta_n), m] - 1}{\mathbb{E}[e^{\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)} | Y(\Delta_n), m]} d\mu_M(m) \right| \right] = 0.$$

Now, we are ready to conclude that as n tends to infinity,

$$\mathbb{E}[e^{-\int_0^T g(s)dY(s) + \frac{1}{2} \int_0^T g^2(s)ds} | Y(\Delta_n)] \rightarrow \mathbb{E}[e^{-\int_0^T g(s)dY + \frac{1}{2} \int_0^T g^2(s)ds} | Y_0^T]$$

in probability and furthermore (59), as desired.

Step 1.3. In this step, we will show the convergence of $\{\mathbb{E}[\log \mathbb{E}[e^{\rho_1(M, Y_0^T)} | Y(\Delta_n), M]]\}$ and $\{\mathbb{E}[\log \mathbb{E}[e^{\rho_1(M, Y_0^T)} | Y(\Delta_n)]]\}$ and further establish the theorem under the condition (47).

First of all, using the concavity of the log function and the fact that $\log x \leq x$, we can obtain the following bounds:

$$|\log \mathbb{E}[e^{\rho_1(M, Y_0^T)} | Y(\Delta_n), M]| \leq |\mathbb{E}[\rho_1(M, Y_0^T) | Y(\Delta_n), M]| + \mathbb{E}[e^{\rho_1(M, Y_0^T)} | Y(\Delta_n), M],$$

and

$$|\log \mathbb{E}[e^{\rho_1(M, Y_0^T)} | Y(\Delta_n)]| \leq |\mathbb{E}[\rho_1(M, Y_0^T) | Y(\Delta_n)]| + \mathbb{E}[e^{\rho_1(M, Y_0^T)} | Y(\Delta_n)].$$

And furthermore, using a similar argument as in **Step 1.1**, we can show that,

$$\begin{aligned} \mathbb{E}[\rho_1(M, Y_0^T) | Y(\Delta_n), M] &= \rho_2(\Delta_n, M, Y_0^T) \times \mathbb{E}[\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T) | Y(\Delta_n), M] \\ &\xrightarrow{n \rightarrow \infty} \rho_1(M, Y_0^T) \end{aligned}$$

in probability, and

$$\begin{aligned} \mathbb{E}[e^{\rho_1(M, Y_0^T)} | Y(\Delta_n), M] &= e^{\rho_2(\Delta_n, M, Y_0^T)} \mathbb{E}[e^{\rho_1(M, Y_0^T) - \rho_2(\Delta_n, M, Y_0^T)} | Y(\Delta_n), M] \\ &\xrightarrow{n \rightarrow \infty} e^{\rho_1(M, Y_0^T)} \end{aligned}$$

in probability. It then follows from the general Lebesgue dominated convergence theorem that

$$\lim_{n \rightarrow \infty} \mathbb{E}[\log \mathbb{E}[e^{\rho_1(M, Y_0^T)} | Y(\Delta_n), M]] = \mathbb{E}[\rho_1(M, Y_0^T)].$$

A parallel argument can be used to show that

$$\lim_{n \rightarrow \infty} \mathbb{E}[\log \mathbb{E}[e^{\rho_1(M, Y_0^T)} | Y(\Delta_n)]] = \mathbb{E}[\log \mathbb{E}[e^{\rho_1(M, Y_0^T)} | Y_0^T]].$$

So, under the condition (47), we have shown that

$$\lim_{n \rightarrow \infty} I(M; Y(\Delta_n)) = I(M; Y_0^T).$$

Step 2. In this step, we will use the convergence in **Step 1** and establish the theorem without the condition (47).

Following Page 264 of [20], we define, for any k ,

$$\tau_k = \begin{cases} \inf\{t \leq T : \int_0^t g^2(s, M, Y_0^s) ds \geq k\}, & \text{if } \int_0^T g^2(s, M, Y_0^s) ds \geq k, \\ T, & \text{if } \int_0^T g^2(s, M, Y_0^s) ds < k. \end{cases} \quad (67)$$

Then, we again follow [20] and define a truncated version of g as follows:

$$g_{(k)}(t, \gamma_0^t, \phi_0^t) = g(t, \gamma_0^t, \phi_0^t) \mathbb{1}_{\int_0^t g^2(s, \gamma_0^s, \phi_0^s) ds < k}.$$

Now, define a truncated version of Y as follows:

$$Y_{(k)}(t) = \int_0^t g_{(k)}(s, M, Y_0^s) ds + B(t), \quad t \in [0, T],$$

which, as elaborated on Page 265 in [20], can be rewritten as

$$Y_{(k)}(t) = \int_0^t g_{(k)}(s, M, Y_{(k),0}^s) ds + B(t), \quad t \in [0, T]. \quad (68)$$

Note that for any fixed k , the system in (68) satisfies the condition (47), and so the theorem holds true. To be more precise, we have

$$I(M; Y_{(k),0}^T) = \mathbb{E} \left[\log \frac{d\mu_{Y_{(k)}|M}}{d\mu_B}(Y_{(k),0}^T | M) \right] - \mathbb{E} \left[\log \frac{d\mu_{Y_{(k)}}}{d\mu_B}(Y_{(k),0}^T) \right],$$

where

$$\frac{d\mu_{Y_{(k)}|M}}{d\mu_B}(Y_{(k),0}^T | M) = e^{\int_0^T g_{(k)}(s) dY_{(k)}(s) - \frac{1}{2} \int_0^T g_{(k)}^2(s) ds} = e^{\int_0^{\tau_k} g(s) dY(s) - \frac{1}{2} \int_0^{\tau_k} g^2(s) ds},$$

and

$$\frac{d\mu_{Y_{(k)}}}{d\mu_B}(Y_{(k),0}^T) = e^{\int_0^T \hat{g}_{(k)}(s) dY_{(k)}(s) - \frac{1}{2} \int_0^T \hat{g}_{(k)}^2(s) ds} = e^{\int_0^{\tau_k} \hat{g}(s) dY(s) - \frac{1}{2} \int_0^{\tau_k} \hat{g}^2(s) ds},$$

where

$$\hat{g}_{(k)}(s) = \mathbb{E}[g_{(k)}(s, M, Y_0^s) | Y_{(k),0}^s], \quad \hat{g}(s) = \mathbb{E}[g(s, M, Y_0^s) | Y_0^s].$$

It then follows from straightforward computations that

$$I(M; Y_{(k),0}^T) = \frac{1}{2} \mathbb{E} \left[\int_0^{\tau_k} (g(s) - \hat{g}(s))^2 ds \right].$$

Notice that it can be easily verified that $\tau_k \rightarrow T$ as k tends to infinity, which, together with the monotone convergence theorem, further yields that monotone increasingly,

$$I(M; Y_{(k),0}^T) = \frac{1}{2} \mathbb{E} \left[\int_0^{\tau_k} (g(s) - \hat{g}(s))^2 ds \right] \rightarrow I(M; Y_0^T) = \frac{1}{2} \mathbb{E} \left[\int_0^T (g(s) - \hat{g}(s))^2 ds \right],$$

as k tends to infinity. By **Step 1**, for any fixed k_i ,

$$\lim_{n \rightarrow \infty} I(M; Y_{(k_i)}(\Delta_n)) = \lim_{n \rightarrow \infty} I(M; Y_{(k_i)}(\Delta_n \cap [0, \tau_{k_i}])) = I(M; Y_{(k_i),0}^T),$$

which means that there exists a sequence $\{n_i\}$ such that, as i tends to infinity, we have, monotone increasingly,

$$I(M; Y_{(k_i)}(\Delta_{n_i} \cap [0, \tau_{k_i}])) \rightarrow I(M; Y_0^T).$$

Since, by the fact that $Y_{(k_i),0}^T$ coincides with Y_0^T on $t \in [0, \tau_{k_i} \wedge T]$ and then (10), we have

$$I(M; Y(\Delta_{n_i})) \geq I(M; Y_{(k_i)}(\Delta_{n_i} \cap [0, \tau_{k_i}])).$$

Now, using the fact that

$$I(M; Y(\Delta_{n_i})) \leq I(M; Y_0^T),$$

we conclude that as i tends to infinity,

$$\lim_{i \rightarrow \infty} I(M; Y(\Delta_{n_i})) = I(M; Y_0^T).$$

A similar argument can be readily applied to any subsequence of $\{I(M; Y(\Delta_n))\}$, which will establish the existence of its further subsubsequence that converges to $I(M; Y_0^T)$, which implies that

$$\lim_{n \rightarrow \infty} I(M; Y(\Delta_n)) = I(M; Y_0^T),$$

completing the proof of the theorem.

6 Proof of Theorem 3.6

Throughout the proof, we will rewrite $t_i^{(n)}$ as t_i for notational simplicity. As in the proof of Theorem 3.2, we will again write the summation $\sum_m(\cdot)p_M(m)$ as the integral $\int(\cdot)d\mu_M(m)$. We proceed in the following two steps.

Step 1. In this step, we establish the theorem assuming that there exists a constant $C > 0$ such that for all $m \in \mathcal{M}$ and all $y_0^T \in C[0, T]$,

$$\int_0^T g^2(s, m, y_0^s) ds < C. \quad (69)$$

We first note that straightforward computations yield

$$f_{Y^{(n)}(\Delta_n)|M}(Y^{(n)}(\Delta_n)|M) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} \exp\left(-\frac{(Y^{(n)}(t_i) - Y^{(n)}(t_{i-1}) - \int_{t_{i-1}}^{t_i} g(s, M, Y_0^{(n), t_{i-1}}) ds)^2}{2(t_i - t_{i-1})}\right), \quad (70)$$

and

$$f_{Y^{(n)}(\Delta_n)}(Y^{(n)}(\Delta_n)) = \int \prod_{i=1}^n \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} \exp\left(-\frac{(Y^{(n)}(t_i) - Y^{(n)}(t_{i-1}) - \int_{t_{i-1}}^{t_i} g(s, m, Y_0^{(n), t_{i-1}}) ds)^2}{2(t_i - t_{i-1})}\right) d\mu_M(m). \quad (71)$$

With (70) and (71), we have

$$\begin{aligned} I(M; Y^{(n)}(\Delta_n)) &= \mathbb{E}[\log f_{Y^{(n)}(\Delta_n)|M}(Y^{(n)}(\Delta_n)|M)] - \mathbb{E}[\log f_{Y^{(n)}(\Delta_n)}(Y^{(n)}(\Delta_n))] \\ &= \mathbb{E}[-\rho_3(M, Y_0^{(n), T})] - \mathbb{E}\left[\log \int e^{-\rho_3(m, Y_0^{(n), T})} d\mu_M(m)\right], \end{aligned} \quad (72)$$

where

$$\rho_3(m, Y_0^{(n), T}) \triangleq \sum_{i=1}^n \left(\frac{-2 \int_{t_{i-1}}^{t_i} g(s, m, Y_0^{(n), t_{i-1}}) ds (Y^{(n)}(t_i) - Y^{(n)}(t_{i-1})) + (\int_{t_{i-1}}^{t_i} g(s, m, Y_0^{(n), t_{i-1}}) ds)^2}{2(t_i - t_{i-1})} \right).$$

On the other hand, by (46), we have

$$\begin{aligned} I(M; Y_0^T) &= \mathbb{E}\left[\log \frac{d\mu_{Y|M}}{d\mu_B}(Y_0^T|M)\right] - \mathbb{E}\left[\log \frac{d\mu_Y}{d\mu_B}(Y_0^T)\right] \\ &= \mathbb{E}\left[\log e^{-\rho_1(M, Y_0^T)}\right] - \mathbb{E}\left[\log \int e^{-\rho_1(m, Y_0^T)} d\mu_M(m)\right] \\ &= \mathbb{E}[-\rho_1(M, Y_0^T)] - \mathbb{E}\left[\log \int e^{-\rho_1(m, Y_0^T)} d\mu_M(m)\right]. \end{aligned} \quad (73)$$

Now, we compute

$$\begin{aligned} &\int_0^T g(s, M, Y_0^s) dY(s) - \sum_{i=1}^n \frac{\int_{t_{i-1}}^{t_i} g(s, M, Y_0^{(n), t_{i-1}}) ds (Y^{(n)}(t_i) - Y^{(n)}(t_{i-1}))}{t_i - t_{i-1}} \\ &= \int_0^T g(s, M, Y_0^s) dY(s) - \sum_{i=1}^n g(t_{i-1}, M, Y_0^{(n), t_{i-1}}) (Y^{(n)}(t_i) - Y^{(n)}(t_{i-1})) \\ &\quad - \sum_{i=1}^n \frac{\int_{t_{i-1}}^{t_i} (g(s, M, Y_0^{(n), t_{i-1}}) - g(t_{i-1}, M, Y_0^{(n), t_{i-1}})) ds (Y^{(n)}(t_i) - Y^{(n)}(t_{i-1}))}{t_i - t_{i-1}}. \end{aligned}$$

It can be easily checked that the second term of the right hand side of the above equality converges to 0 in mean. And for the first term, we have

$$\begin{aligned}
& \int_0^T g(s, M, Y_0^s) dY(s) - \sum_{i=1}^n g(t_{i-1}, M, Y_0^{(n), t_{i-1}}) (Y^{(n)}(t_i) - Y^{(n)}(t_{i-1})) \\
&= \sum_{i=1}^n \int_{t_{i-1}}^{t_i} g(s, M, Y_0^s) dY(s) - \sum_{i=1}^n g(t_{i-1}, M, Y_0^{(n), t_{i-1}}) (Y(t_i) - Y(t_{i-1})) \\
&\quad + \sum_{i=1}^n g(t_{i-1}, M, Y_0^{(n), t_{i-1}}) ((Y(t_i) - Y(t_{i-1})) - (Y^{(n)}(t_i) - Y^{(n)}(t_{i-1}))) \\
&= \sum_{i=1}^n \int_{t_{i-1}}^{t_i} g(s, M, Y_0^s) dY(s) - \sum_{i=1}^n \int_{t_{i-1}}^{t_i} g(t_{i-1}, M, Y_0^{(n), t_{i-1}}) dY(s) \\
&\quad + \sum_{i=1}^n g(t_{i-1}, M, Y_0^{(n), t_{i-1}}) ((Y(t_i) - Y(t_{i-1})) - (Y^{(n)}(t_i) - Y^{(n)}(t_{i-1}))) \\
&= \sum_{i=1}^n \int_{t_{i-1}}^{t_i} (g(s, M, Y_0^s) - g(t_{i-1}, M, Y_0^{(n), t_{i-1}})) dY(s) \\
&\quad + \sum_{i=1}^n g(t_{i-1}, M, Y_0^{(n), t_{i-1}}) ((Y(t_i) - Y(t_{i-1})) - (Y^{(n)}(t_i) - Y^{(n)}(t_{i-1}))).
\end{aligned}$$

It then follows from Conditions (d) and (e), Lemmas 3.1, 3.4 and 3.5 that

$$\mathbb{E} \left[\left| \sum_{i=1}^n \frac{\int_{t_{i-1}}^{t_i} g(s, M, Y_0^{(n), t_{i-1}}) ds (Y^{(n)}(t_i) - Y^{(n)}(t_{i-1}))}{t_i - t_{i-1}} - \int_0^T g(s, M, Y_0^s) dY(s) \right| \right] = O(\delta_{\Delta_n}^{\frac{1}{2}}). \quad (74)$$

And using a similar argument as above, we deduce that

$$\mathbb{E} \left[\left| \frac{1}{2} \sum_{i=1}^n \frac{(\int_{t_{i-1}}^{t_i} g(s, M, Y_0^{(n), t_{i-1}}) ds)^2}{t_i - t_{i-1}} - \frac{1}{2} \int_0^T g(s, M, Y_0^s)^2 ds \right| \right] = O(\delta_{\Delta_n}^{\frac{1}{2}}). \quad (75)$$

It then follows from (74) and (75) that

$$\mathbb{E} \left[\left| \rho_3(M, Y_0^{(n), T}) - \rho_1(M, Y_0^T) \right| \right] = O(\delta_{\Delta_n}^{\frac{1}{2}}). \quad (76)$$

We now establish the following convergence:

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\log \int e^{-\rho_3(m, Y_0^{(n), T})} d\mu_M(m) \right] = \mathbb{E} \left[\log \int e^{-\rho_1(m, Y_0^T)} d\mu_M(m) \right]. \quad (77)$$

Note that using a parallel argument as in the derivation of (76), we can establish

$$\lim_{n \rightarrow \infty} \mathbb{E} \int \left| -\rho_3(m, Y_0^{(n), T}) - \rho_1(m, Y_0^T) \right| d\mu_M(m) = 0; \quad (78)$$

and similarly as in the derivation of (66), from Conditions (d) and (e), Lemmas 3.1, 3.4 and 3.5, we deduce that

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\int \left| e^{-\rho_3(m, Y_0^{(n), T})} - e^{-\rho_1(m, Y_0^T)} \right| d\mu_M(m) \right] = 0. \quad (79)$$

Moreover, we note that it always holds that

$$\left| \log \int e^{-\rho_3(m, Y_0^{(n), T})} d\mu_M(m) \right| \leq \int e^{-\rho_3(m, Y_0^{(n), T})} d\mu_M(m) + \left| \int \rho_3(m, Y_0^{(n), T}) d\mu_M(m) \right|. \quad (80)$$

Then, the desired (77) follows from an application of the general Lebesgue dominated convergence theorem with (78), (79) and (80).

Finally, with (72), (73), (76) and (77), we conclude that

$$\lim_{n \rightarrow \infty} I(M; Y^{(n)}(\Delta_n)) = I(M; Y_0^T),$$

establishing the theorem with the extra condition (69).

Step 2. In this step, we will use the convergence in **Step 1** and establish the theorem without the condition (69).

Defining the stopping τ_k , $g_{(k)}$ and $Y_{(k)}$ as in the proof of 2) of Theorem 3.2, we again have

$$Y_{(k)}(t) = \int_0^t g_{(k)}(s, M, Y_{(k),0}^s) ds + B(t), \quad t \in [0, T].$$

For any fixed k , applying the Euler-Maruyama approximation as in (23) and (24) to the above channel with respect to Δ_n , we obtain the process $\{Y_{(k)}^{(n)}(t)\}$.

Now, recall that $I(M; Y^{(n)}(\Delta_n))$ can be computed as in (72), and moreover, by some straightforward computations,

$$\mathbb{E}[-\rho_3(M, Y_0^{(n), T})] = \mathbb{E} \left[\sum_{i=1}^n \frac{(\int_{t_{i-1}}^{t_i} g(s, M, Y_0^{(n), t_{i-1}}) ds)^2}{2(t_i - t_{i-1})} \right].$$

Similarly we have

$$\begin{aligned} I(M; Y_{(k)}^{(n)}(\Delta_n)) &= \mathbb{E}[\log f_{Y_{(k)}^{(n)}(\Delta_n)|M}(Y_{(k)}^{(n)}(\Delta_n)|M)] - \mathbb{E}[\log f_{Y_{(k)}^{(n)}(\Delta_n)}(Y_{(k)}^{(n)}(\Delta_n))] \\ &= \mathbb{E}[-\rho_4(M, Y_{(k),0}^{(n), T})] - \mathbb{E} \left[\log \int e^{-\rho_4(m, Y_{(k),0}^{(n), T})} d\mu_M(m) \right], \end{aligned}$$

where

$$\rho_4(m, Y_{(k),0}^{(n), T}) = \sum_{i=1}^n \left(\frac{-2 \int_{t_{i-1}}^{t_i} g_{(k)}(s, m, Y_{(k),0}^{(n), t_{i-1}}) ds (Y_{(k)}^{(n)}(t_i) - Y_{(k)}^{(n)}(t_{i-1})) + (\int_{t_{i-1}}^{t_i} g_{(k)}(s, m, Y_{(k),0}^{(n), t_{i-1}}) ds)^2}{2(t_i - t_{i-1})} \right),$$

and moreover, by some straightforward computations,

$$\mathbb{E}[-\rho_4(M, Y_{(k),0}^{(n), T})] = \mathbb{E} \left[\sum_{i=1}^n \frac{(\int_{t_{i-1}}^{t_i} g_{(k)}(s, M, Y_{(k),0}^{(n), t_{i-1}}) ds)^2}{2(t_i - t_{i-1})} \right].$$

Note that it can be easily verified that

$$\frac{1}{f(Y^{(n)}(\Delta_n)|Y_{(k)}^{(n)}(\Delta_n), \tau_k)} = \mathbb{E} \left[\frac{1}{f(Y^{(n)}(\Delta_n)|Y_{(k)}^{(n)}(\Delta_n), \tau_k, M)} \middle| Y^{(n)}(\Delta_n), \tau_k \right],$$

which boils down to

$$\frac{\int e^{-\rho_4(m, Y_{(k),0}^{(n),T})} d\mu_M(m)}{\int e^{-\rho_3(m, Y_0^{(n),T})} d\mu_M(m)} = \mathbb{E} \left[e^{\rho_3(M, Y_0^{(n),T}) - \rho_4(M, Y_{(k),0}^{(n),T})} \middle| Y^{(n)}(\Delta_n), \tau_k \right].$$

Using this and Jensen's inequality, we deduce that

$$\begin{aligned} \mathbb{E} \left[\log \frac{\int e^{-\rho_4(m, Y_{(k),0}^{(n),T})} d\mu_M(m)}{\int e^{-\rho_3(m, Y_0^{(n),T})} d\mu_M(m)} \right] &= \mathbb{E} \left[\log \mathbb{E} \left[e^{\rho_3(M, Y_0^{(n),T}) - \rho_4(M, Y_{(k),0}^{(n),T})} \middle| Y^{(n)}(\Delta_n), \tau_k \right] \right] \\ &\leq \log \mathbb{E} \left[e^{\rho_3(M, Y_0^{(n),T}) - \rho_4(M, Y_{(k),0}^{(n),T})} \right] \\ &\leq 0, \end{aligned} \tag{81}$$

where for the last inequality, we have applied Fatou's lemma as in deriving (57).

Now, using (10) and the fact that $Y^{(n)}$ and $Y_{(k)}^{(n)}$ coincide over $[0, \tau_k \wedge T]$, we infer that

$$I(M; Y^{(n)}(\Delta_n)) - I(M; Y_{(k)}^{(n)}(\Delta_n)) \geq 0, \tag{82}$$

and furthermore, one verifies that for any $\varepsilon > 0$,

$$\begin{aligned} &I(M; Y^{(n)}(\Delta_n)) - I(M; Y_{(k)}^{(n)}(\Delta_n)) \\ &= (\mathbb{E}[-\rho_3(M, Y_0^{(n),T})] - \mathbb{E}[-\rho_4(M, Y_{(k),0}^{(n),T})]) \\ &\quad - \left(\mathbb{E} \left[\log \int e^{-\rho_3(m, Y_0^{(n),T})} d\mu_M(m) \right] - \mathbb{E} \left[\log \int e^{-\rho_4(m, Y_{(k),0}^{(n),T})} d\mu_M(m) \right] \right) \\ &\stackrel{(a)}{\leq} \mathbb{E} \left[\sum_{i=1}^n \frac{(\int_{t_{i-1}}^{t_i} g(s, M, Y_0^{(n),t_{i-1}}) ds)^2}{2(t_i - t_{i-1})} \right] - \mathbb{E} \left[\sum_{i=1}^n \frac{(\int_{t_{i-1}}^{t_i} g_{(k)}(s, M, Y_{(k),0}^{(n),t_{i-1}}) ds)^2}{2(t_i - t_{i-1})} \right] \\ &= \mathbb{E} \left[\sum_{i=1}^n \frac{(\int_{t_{i-1}}^{t_i} g(s, M, Y_0^{(n),t_{i-1}}) ds)^2}{2(t_i - t_{i-1})} \right] - \mathbb{E} \left[\sum_{i=1}^n \frac{(\int_{t_{i-1} \wedge \tau_k}^{t_i \wedge \tau_k} g(s, M, Y_0^{(n),t_{i-1}}) ds)^2}{2(t_i - t_{i-1})} \right] \\ &= \mathbb{E} \left[\sum_{i=1}^n \frac{(\int_{t_{i-1} \vee \tau_k}^{t_i \vee \tau_k} g(s, M, Y_0^{(n),t_{i-1}}) ds)^2}{2(t_i - t_{i-1})} \right] \\ &\leq \mathbb{E} \left[\int_{\tau_k}^T g(s, M, Y_0^{(n), \lfloor s \rfloor_{\Delta_n}})^2 ds \right] \\ &\leq \mathbb{E} \left[\int_{\tau_k}^T g(s, M, Y_0^{(n), \lfloor s \rfloor_{\Delta_n}})^2 ds; T - \tau_k \leq \varepsilon \right] + \mathbb{E} \left[\int_{\tau_k}^T g(s, M, Y_0^{(n), \lfloor s \rfloor_{\Delta_n}})^2 ds; T - \tau_k > \varepsilon \right] \\ &\leq \int_{T-\varepsilon}^T \mathbb{E} \left[g(s, M, Y_0^{(n), \lfloor s \rfloor_{\Delta_n}})^2 \right] ds + \mathbb{E} \left[\int_{\tau_k}^T g(s, M, Y_0^{(n), \lfloor s \rfloor_{\Delta_n}})^2 ds; T - \tau_k > \varepsilon \right]. \end{aligned}$$

where we have used (81) for (a) and $\lfloor s \rfloor_{\Delta_n}$ denotes the unique number n_0 such that $t_{n_0} \leq s < t_{n_0+1}$. Using the easily verifiable fact that $\tau_k \rightarrow T$ almost surely as k tends to infinity, (82) and the fact that ε can be arbitrarily small, we conclude that as k tends to infinity, uniformly over all n ,

$$I(M; Y_{(k)}^{(n)}(\Delta_n)) \rightarrow I(M; Y^{(n)}(\Delta_n)). \tag{83}$$

Next, an application of the monotone convergence theorem with the fact that $\tau_k \rightarrow T$ as k tends to infinity yields that monotone increasingly

$$I(M; Y_{(k),0}^T) = \frac{1}{2} \mathbb{E} \left[\int_0^{\tau_k} (g(s) - \hat{g}(s))^2 ds \right] \rightarrow I(M; Y_0^T) = \frac{1}{2} \mathbb{E} \left[\int_0^T (g(s) - \hat{g}(s))^2 ds \right]$$

as n tends to infinity. By **Step 1**, for any fixed k_i ,

$$\lim_{n \rightarrow \infty} I(M; Y_{(k_i)}^{(n)}(\Delta_n)) = I(M; Y_{(k_i),0}^T),$$

which means that there exists a sequence $\{n_i\}$ such that, as i tends to infinity,

$$I(M; Y_{(k_i)}^{(n_i)}(\Delta_{n_i})) \rightarrow I(M; Y_0^T).$$

Moreover, by (83),

$$\lim_{i \rightarrow \infty} I(M; Y_{(k_i)}^{(n_i)}(\Delta_{n_i})) = \lim_{i \rightarrow \infty} I(M; Y^{(n_i)}(\Delta_{n_i})),$$

which further implies that

$$\lim_{i \rightarrow \infty} I(M; Y^{(n_i)}(\Delta_{n_i})) = I(M; Y_0^T).$$

The theorem then follows from a usual subsequence argument as in the proof of 2) of Theorem 3.2.

References

- [1] R. Ash. Capacity and error bounds for a time-continuous Gaussian channel. *Information and Control*, vol. 6, pp. 14-27, 1963.
- [2] R. Ash. Further discussion of a time-continuous Gaussian channel. *Information and Control*, vol. 7, pp. 78-83, 1964.
- [3] R. Ash. *Information Theory*, Wiley Interscience, New York, 1965.
- [4] P. Bethoux. Test et estimations concernant certaines fonctions aleatoires en particulier Laplaciennes. *Ann. Inst. Henri Poincare*, vol. 27, pp. 255-322, 1962.
- [5] T. Cover and J. Thomas. *Elements of Information Theory*, 2nd ed., Wiley Interscience, New York, 2006.
- [6] T. Duncan. On the calculation of mutual information. *SIAM Journal on Applied Mathematics*, vol. 19, no. 1, pp. 215-220, 1970.
- [7] R. Durrett. *Probability: Theory and Examples*, 4th ed., Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 2010.
- [8] R. Fortet. Hypothesis testing and Estimation for Laplacian Functions. *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 289-305, 1961.
- [9] R. Gallager. *Information Theory and Reliable Communication*, Wiley, New York, 1968.
- [10] A. Gelfand and I. Yaglom. Calculation of the amount of information about a random function contained in another such function. *Uspekhi Mat. Nauk*, vol. 12, pp. 3-52, 1957. Trans. in *Am. Math. Soc. Trans.*, Ser. 2, vol. 12, pp. 199-247, 1959.

- [11] R. Huang and R. Johnson. Information capacity of time-continuous channels. *IEEE Trans. Info. Theory*, IT-8, pp. 191-198, 1962.
- [12] R. Huang and R. Johnson. Information transmission with time-continuous random processes. *IEEE Trans. Info. Theory*, vol. 9, no. 2, pp. 84-94, 1963.
- [13] P. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*, Stochastic Modelling and Applied Probability, vol. 23, Springer-Verlag, Berlin Heidelberg, 1992.
- [14] S. Ihara. *Information Theory for Continuous Systems*, World Scientific, Singapore, 1993.
- [15] B. Jacob, M. Zakai and J. Ziv. On the ε -entropy and the rate-distortion function of certain non-Gaussian processes. *IEEE Trans. Info. Theory*, vol. 20, no. 4, pp. 517-524, 1974.
- [16] T. Kadota, M. Zakai and J. Ziv. Mutual information of the white Gaussian channel with and without feedback. *IEEE Trans. Info. Theory*, vol. 17, pp. 368-371, 1971.
- [17] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus*, New York: Springer-Verlag, 1991.
- [18] X. Liu, R. Bustin, G. Han and S. Shamai. An elementary proof of a classical information-theoretic formula. Preprint, 2018, Arxiv:1809.01422.
- [19] X. Liu and G. Han. On continuous-time Gaussian channels. Preprint, 2017, available at Arxiv:1704.02569.
- [20] R. Liptser and A. Shiryaev. *Statistics of random processes (I): General theory*, 2nd edition, Springer-Verlag, Berlin, 2001.
- [21] X. Mao. *Stochastic Differential Equations and Applications*, Horwood, 1997.
- [22] H. Nyquist. Certain factors affecting telegraph speed. *The Bell System Technical Journal*, pp. 324, 1924.
- [23] B. Oksendal. *Stochastic Differential Equations: An Introduction with Applications*, Springer, Berlin, 1995.
- [24] M. Pinsker. *Information and Information Stability of Random Variables and Processes*, Holden-Day, San Francisco, 1964.
- [25] H. Royden. *Real analysis*, 4th edition, Prentice Hall, Boston, 2010.
- [26] C. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, vol. 27, pp. 379-423, 1948.
- [27] C. Shannon. Communication in the presence of noise. *Proc. IRE*, vol. 37, pp. 10-21, 1949.
- [28] D. Slepian. On Bandwidth. *Proc. IEEE*, vol. 64, no. 3, 1976.
- [29] A. Wyner. The capacity of the band-limited Gaussian channel, *The Bell System Technical Journal*, pp. 359-395, 1966. Also reprinted in *Key Papers in the Development of Information Theory*, edited by D. Slepian, IEEE Press, New York, pp. 190-193, 1974.