

ENTROPY FLOW AND DE BRUIJN'S IDENTITY FOR A CLASS OF STOCHASTIC DIFFERENTIAL EQUATIONS DRIVEN BY FRACTIONAL BROWNIAN MOTION

MICHAEL C.H. CHOI 

Institute for Data and Decision Analytics, The Chinese University of Hong Kong, Shenzhen, Guangdong, 518172, P.R. China and Shenzhen Institute of Artificial Intelligence and Robotics for Society
E-mail: michaelchoi@cuhk.edu.cn

CHIHOOON LEE

School of Business, Stevens Institute of Technology, Hoboken, NJ 07030, USA and Institute for Data and Decision Analytics, The Chinese University of Hong Kong, Shenzhen, Guangdong 518172, P.R. China
E-mail: clee4@stevens.edu

JIAN SONG 

Research Center for Mathematics and Interdisciplinary Sciences, Shandong University, Qingdao, Shandong, 266237, China and School of Mathematics, Shandong University, Jinan, Shandong, 250100, China
E-mail: txjsong@hotmail.com

Motivated by the classical De Bruijn's identity for the additive Gaussian noise channel, in this paper we consider a generalized setting where the channel is modelled via stochastic differential equations driven by fractional Brownian motion with Hurst parameter $H \in (0, 1)$. We derive generalized De Bruijn's identity for Shannon entropy and Kullback–Leibler divergence by means of Itô's formula, and present two applications. In the first application we demonstrate its equivalence with Stein's identity for Gaussian distributions, while in the second application, we show that for $H \in (0, 1/2]$, the entropy power is concave in time while for $H \in (1/2, 1)$ it is convex in time when the initial distribution is Gaussian. Compared with the classical case of $H = 1/2$, the time parameter plays an interesting and significant role in the analysis of these quantities.

Keywords: de bruijn's identity, entropy power, fokker–planck equation, fractional brownian motion

Jel Classification: 60G22, 60G15

1. INTRODUCTION

Consider an additive Gaussian noise channel modelled by

$$X_t = X_0 + \sqrt{t}Z,$$

where $t \geq 0$, Z is a standard normal random variable and the initial value X_0 is independent of Z . In information theory, the classical De Bruijn's identity, first studied by Stam [19], establishes a relationship between the time derivative of the entropy of X_t to the Fisher information of X_t . While such a Gaussian channel is very popular in the literature (see e.g., [12,16]), in recent years researchers have been investigating into various generalizations of the noise channel. This includes the Fokker–Planck channel [22] in which it is modelled via the stochastic differential equation driven by Brownian motion with general drift and diffusion, and also the dependent case [13] where X_0 and Z are jointly distributed as Archimedean or Gaussian copulas.

In reality, however, the channel may exhibit features that are not adequately modelled by the classical model. For example, in the area of Ethernet traffic [23], it has been reported that the traffic exhibits self-similarity and long-range dependence. A similar phenomenon is also observed in analyzing video conference traffic [3]. As these non-standard characteristics, in particular long-range dependence, cannot be effectively captured in the traditional additive Gaussian noise model, it motivates us to consider channel driven by fractional Brownian motion (fBm) naturally as a possible generalization. In particular, long-range dependence is a significant feature possessed by with Hurst parameter greater than $1/2$. In this paper we derive generalized De Bruijn's identity for such channel and discuss its relationship with Stein's identity as well as entropy power. Interestingly, the time parameter t and the Hurst parameter H of the fBm both play an important role in these results.

Our mathematical contributions in this paper are as follows.

- We derive a generalized version of the celebrated De Bruijn's identity for channel driven by fBm. It involves a combination of techniques such as Itô's formula and the Fokker–Planck equation.
- We build the connection and prove the equivalence between Stein's identity for Gaussian distributions and the generalized De Bruijn's identity, when the initial noise is a normal distribution.
- As another application of the generalized De Bruijn's identity, we demonstrate that the convexity/concavity of the entropy power depends on the Hurst parameter. This phenomenon is not observed in the classical Brownian motion case.

The rest of the paper is organized as follows. In Section 2, we first introduce the channel driven by fBm, followed by stating the results for generalized De Bruijn's identity as well as their proofs. In Section 3, we present two applications. In Section 3.1, we prove the equivalence between the generalized De Bruijn's identity and the Stein's identity for Gaussian distribution, while in Section 3.2, we prove that the entropy power is convex or concave depending on the value of H . Finally in Section 4, we conclude our paper along with future research directions.

Before we proceed to the main results of the paper, we first review a few important concepts that will be frequently used in subsequent sections. A fBm $B^H = (B_t^H)_{t \geq 0}$ with Hurst parameter $H \in (0, 1)$ is a centered Gaussian process with stationary increments and

covariance function given by

$$\mathbb{E}B_s^H B_t^H = \frac{1}{2}(t^{2H} + s^{2H} - |t - s|^{2H}).$$

For further references of fBm, we refer readers to Mandelbrot and Van Ness [14]. The Shannon entropy of a random variable X with probability density function f_X , denoted by $h(X)$, is given by

$$h(X) := -\mathbb{E}(\ln f_X(X)) = -\int_{\mathbb{R}} f_X(x) \ln f_X(x) dx. \tag{1.1}$$

Let $b : \mathbb{R} \rightarrow (0, \infty)$ be a positive function. The generalized Fisher information with respect to b , first introduced by Wibisono et al. [22], is given by

$$J_b(X) := \mathbb{E} \left[b(X) \left(\frac{\partial}{\partial x} \ln f_X(X) \right)^2 \right]. \tag{1.2}$$

Note that when $b=1$, $J_1(X)$ is simply the classical Fisher information. When X follows a parametric distribution, say with location parameter θ , then the Cramér-Rao lower bound states that the variance of any unbiased estimator of θ is lower bounded by the reciprocal of $J_1(X)$. The Kullback–Leibler (KL) divergence, or the relative entropy, between X and random variable Y with density f_Y , written as $K(X||Y)$, is given by

$$K(X||Y) := \mathbb{E} \left[\ln \frac{f_X(X)}{f_Y(X)} \right] = \int_{\mathbb{R}} f_X(x) \ln \frac{f_X(x)}{f_Y(x)} dx. \tag{1.3}$$

For two random variables X and Y , the relative Fisher information with respect to b is

$$J_b(X||Y) := \mathbb{E} \left[b(X) \left(\frac{\partial}{\partial x} \ln \frac{f_X(X)}{f_Y(X)} \right)^2 \right]. \tag{1.4}$$

2. GENERALIZED DE BRUIJN'S IDENTITY

In this section, we derive the generalized De Bruijn's identity for channel modelled via stochastic differential equation driven by fBm. More precisely, consider a channel governed by

$$dX_t = \sigma(X_t) \circ dB_t^H, \tag{2.1}$$

with initial value $X_0 = x_0$, where $(B_t^H)_{t \geq 0}$ is a fBm with Hurst parameter $H \in (0, 1)$. The stochastic integral $\sigma(X_t) \circ dB_t^H$ is in the “pathwise” sense, i.e., if $H \in (1/2, 1)$, the integral is understood as Young's integration [24]; if $H \in (1/4, 1)$ it is understood in the rough paths sense of Lyons [8]; if $H \in (1/6, 1)$ it is understood in the sense of symmetric integral [18]; and if $H \in (0, 1)$, and it is understood as m -order Newton-Cotes functional [11].

By Theorem 2.10 in Nourdin [15], assuming that the diffusion coefficient $\sigma(x)$ is sufficiently regular (say, infinitely differentiable with bounded derivatives of all orders), the one-dimensional SDE (2.1) with $H \in (0, 1)$ has a unique solution $X_t = \varphi(B_t^H)$, where $\varphi'(x) = \sigma(\varphi(x))$ with $\varphi(0) = x_0$, which can be obtained by using the Doss–Sussman transformation as in Sussmann [20]. Note that the solution X_t is a function of B_t^H , rather than a functional of $(B_s^H)_{0 \leq s \leq t}$. This particular form allows functions of X_t to have a simple Itô's formula (Lemma 2.1) without involving Malliavian derivatives. As a consequence, the

Fokker–Planck equation (Lemma 2.2) can be derived, and furthermore, a Feynman–Kac type formula can also be obtained for a class of partial differential equations (see Corollary 26 and Example 28 in Baudoin and Coutin [2]). Note that by Remark 27 in Baudoin and Coutin [2], this type of formulas only hold for the SDEs driven by fBm in the commutative case which is in the form of (2.1) if the dimension is one.

THEOREM 2.1 (Generalized De Bruijn’s identity for Shannon entropy of fBm): *Consider the channel $X = (X_t)_{t \geq 0}$ modeled by Eq. (2.1) with Hurst parameter $H \in (0, 1)$ and initial value $X_0 = x_0$. Assume that the diffusion coefficient $\sigma(x) \in C^\infty(\mathbb{R})$ has bounded derivatives of all orders. The entropy flow of X is given by*

$$\frac{d}{dt}h(X_t) = Ht^{2H-1} \left\{ J_{\sigma^2}(X_t) - \mathbb{E} \left[\frac{\partial^2}{\partial x^2} \sigma^2(X_t) \right] + \mathbb{E} [\sigma''(X_t)\sigma(X_t) + (\sigma'(X_t))^2] \right\}, \tag{2.2}$$

where we recall that the generalized Fisher information $J_{\sigma^2}(X_t)$ is defined in (1.2).

Remark 2.1: Note that when $H = 1/2$, the fBm $W = B^H$ is a Brownian motion, and the Stratonovich Eq. (2.1) becomes

$$dX_t = \frac{\sigma(X_t)\sigma'(X_t)}{2}dt + \sigma(X_t) \diamond dW_t$$

where the stochastic integral is in the Itô sense. Then formula (2.2) coincides with the classical De Bruijn’s identity [13,22]. That is, when $H = 1/2$, (2.2) becomes

$$\frac{d}{dt}h(X_t) = \frac{1}{2} \left\{ J_{\sigma^2}(X_t) - \mathbb{E} \left[\frac{\partial^2}{\partial x^2} \sigma^2(X_t) \right] + \mathbb{E} [\sigma''(X_t)\sigma(X_t) + (\sigma'(X_t))^2] \right\},$$

which is the result in Wibisono et al. [22, Thm. 5] with drift $\sigma(x)\sigma'(x)/2$ and diffusion coefficient $\sigma(x)$. In particular, when $\sigma = 1$, we have

$$\frac{d}{dt}h(X_t) = \frac{1}{2}J_1(X_t).$$

THEOREM 2.2 (Generalized De Bruijn’s identity for KL divergence of fBm): *Consider the channel $X = (X_t)_{t \geq 0}$ (resp. $Y = (Y_t)_{t \geq 0}$) modeled by Eq. (2.1) with Hurst parameter $H \in (0, 1)$ and initial value $X_0 = x_0$ (resp. $Y_0 = y_0$). Assume that the diffusion coefficient $\sigma(x) \in C^\infty(\mathbb{R})$ has bounded derivatives of all orders. The time derivative of the KL divergence between X_t and Y_t is given by*

$$\frac{d}{dt}K(X_t||Y_t) = -Ht^{2H-1}J_{\sigma^2}(X_t||Y_t), \tag{2.3}$$

where we recall that the relative Fisher information $J_{\sigma^2}(X_t||Y_t)$ is defined in (1.4). In particular, $K(X_t||Y_t)$ is non-increasing in t .

Remark 2.2: Note that again when $H = 1/2$, (2.3) becomes

$$\frac{d}{dt}K(X_t||Y_t) = -\frac{1}{2}J_{\sigma^2}(X_t||Y_t),$$

which is Wibisono et al. [22, Thm. 6].

In the first two main results above, we assume that the initial value is $X_0 = x_0$. In the following result, we assume that the channel is of the form

$$X_t = X_0 + B_t^H, \tag{2.4}$$

where the initial value X_0 is independent of the fBm. We shall derive the generalized De Bruijn's identity via the classical version:

THEOREM 2.3 (Deriving the generalized De Bruijn's identity via the classical De Bruijn's identity): *Consider the channel $X = (X_t)_{t \geq 0}$ modeled by Eq. (2.4) with Hurst parameter $H \in (0, 1)$, initial value X_0 independent of the fBm and has a finite second moment. The entropy flow of X is given by*

$$\frac{d}{dt}h(X_t) = Ht^{2H-1}J_1(X_t). \tag{2.5}$$

In particular, when X_0 is a Gaussian distribution with mean 0 and variance σ_0^2 , we then have

$$\frac{d}{dt}h(X_t) = \frac{Ht^{2H-1}}{\sigma_0^2 + t^{2H}}.$$

2.1. Proof of Theorem 2.1, Theorem 2.2 and Theorem 2.3

We first present two lemmas that will be used in our proofs of Theorems 2.1 and 2.2.

LEMMA 2.1 (Itô's formula): *Consider the channel $X = (X_t)_{t \geq 0}$ modeled by Eq. (2.1) with Hurst parameter $H \in (0, 1)$, initial value $X_0 = x_0$ and twice differentiable diffusion coefficient $\sigma(x)$. Suppose that $f(t, x)$ is any twice differentiable function of two variables. Assume that the functions $\sigma(x)$ and $f(t, x)$ and their (partial) derivatives are at polynomial growth. Then*

$$\begin{aligned} f(t, X_t) &= f(0, x_0) + \int_0^t f_s(s, X_s)ds + \int_0^t f_x(s, X_s)\sigma(X_s) \diamond dB_s^H \\ &\quad + H \int_0^t s^{2H-1} \left(f_{xx}(s, X_s)\sigma(X_s) + f_x(s, X_s)\sigma'(X_s) \right) \sigma(X_s)ds. \end{aligned}$$

PROOF: Note that $X_t = \varphi(B_t^H)$ with $\varphi'(x) = \sigma(\varphi(x))$ and $\varphi(0) = x_0$. By Itô's formula in Section 8 of Alòs et al. [1] for $H \in (1/4, 1)$ and in Corollary 4.8 of Cheridito and Nualart [5] for $H \in (0, 1/2)$, noting that $\varphi''(x) = \sigma(\varphi(x))\varphi'(x)$, we have

$$\begin{aligned} f(t, X_t) &= f(t, \varphi(B_t^H)) \\ &= f(0, x_0) + \int_0^t f_s(s, \varphi(B_s^H))ds + \int_0^t f_x(s, \varphi(B_s^H))\varphi'(B_s^H) \diamond dB_s^H \\ &\quad + H \int_0^t s^{2H-1} \left(f_{xx}(s, \varphi(B_s^H))(\varphi'(B_s^H))^2 + f_x(s, \varphi(B_s^H))\varphi''(B_s^H) \right) ds \\ &= f(0, x_0) + \int_0^t f_s(s, X_s)ds + \int_0^t f_x(s, X_s)\sigma(X_s) \diamond dB_s^H \\ &\quad + H \int_0^t s^{2H-1} \left(f_{xx}(s, X_s)\sigma(X_s) + f_x(s, X_s)\sigma'(X_s) \right) \sigma(X_s)ds. \end{aligned} \quad \blacksquare$$

Remark 2.3 (The condition imposed on $\sigma(x)$ in [Theorems 2.1](#) and [2.2](#)): The proofs of [Theorems 2.1](#) and [2.2](#) rely on the Itô's formula given in [Lemma 2.1](#) and the existence and uniqueness of the solution to (2.1). On the one hand, in order to apply [Lemma 2.1](#), it is natural to assume that $\sigma(x)$ is twice differentiable with the derivatives growing at most polynomially fast. On the other hand, more regularity condition on $\sigma(x)$ was imposed to establish the existence and uniqueness of the solution to (2.1) in Theorem 2.10 of Nourdin [15] for small Hurst parameter H . More precisely, for $H \in (1/4m + 2, 1)$ with $m \in \mathbb{N}$, it is assumed that, $\sigma(x)$ belongs to $C^{4m+1}(\mathbb{R})$ and is Lipschitz.

In [Theorems 2.1](#) and [2.2](#) for all $H \in (0, 1)$, we simply assume that $\sigma(x)$ is smooth and all derivatives are bounded, which clearly satisfies the conditions in [Lemma 2.1](#) above and Theorem 2.10 of Nourdin [15].

LEMMA 2.2 (Fokker–Planck equation): *Consider the channel $X = (X_t)_{t \geq 0}$ modeled by Eq. (2.1) with Hurst parameter $H \in (0, 1)$ and initial value $X_0 = x_0$. Assume that the diffusion coefficient $\sigma(x) \in C^\infty(\mathbb{R})$ has bounded derivatives of all orders. Let $P_t(x)$ be the probability density function of X_t , then*

$$\frac{\partial}{\partial t} P_t(x) = Ht^{2H-1} \left(-\frac{\partial}{\partial x} \sigma'(x)\sigma(x)P_t(x) + \frac{\partial^2}{\partial x^2} \sigma^2(x)P_t(x) \right).$$

PROOF: Let $g(x)$ be a twice differentiable function, and we substitute $f(t, x) = g(x)$ in [Lemma 2.1](#). We arrive at

$$\frac{d}{dt} \mathbb{E}[g(X_t)] = Ht^{2H-1} \left(\mathbb{E} [g_{xx}(X_t)\sigma^2(X_t)] + \mathbb{E} [g_x(X_t)\sigma'(X_t)\sigma(X_t)] \right). \tag{2.6}$$

Note that the left hand side of (2.6) is

$$\frac{d}{dt} \mathbb{E}[g(X_t)] = \int_{\mathbb{R}} g(x) \frac{\partial}{\partial t} P_t(x) dx.$$

Using integration by part, the right hand side of (2.6) can be written as

$$Ht^{2H-1} \left(\int_{\mathbb{R}} g(x) \left(\frac{\partial^2}{\partial x^2} \sigma^2(x)P_t(x) \right) dx - \int_{\mathbb{R}} g(x) \left(\frac{\partial}{\partial x} \sigma'(x)\sigma(x)P_t(x) \right) dx \right).$$

The desired result follows since g is arbitrary. ■

2.1.1. Proof of [Theorem 2.1](#) Denote by $P_t(x)$ the probability density of X_t , and let $f(s, x) = -\ln P_s(x)$ in [Lemma 2.1](#). Then we have

$$f_x(s, x) = -(P_s(x))^{-1} \frac{\partial}{\partial x} P_s(x),$$

and

$$f_{xx}(s, x) = (P_s(x))^{-2} \left(\frac{\partial}{\partial x} P_s(x) \right)^2 - (P_s(x))^{-1} \frac{\partial^2}{\partial x^2} P_s(x).$$

Thus,

$$\begin{aligned} \mathbb{E}[f_x(s, X_s)\sigma'(X_s)\sigma(X_s)] &= - \int_{\mathbb{R}} \frac{\partial}{\partial x} P_s(x)\sigma'(x)\sigma(x)dx \\ &= \int_{\mathbb{R}} P_s(x)(\sigma'(x)\sigma(x))'dx = \mathbb{E} [\sigma''(X_s)\sigma(X_s) + (\sigma'(X_s))^2], \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}[f_{xx}(s, X_s)\sigma^2(X_s)] &= \int_{\mathbb{R}} \left[(P_s(x))^{-1} \left(\frac{\partial}{\partial x} P_s(x) \right)^2 - \frac{\partial^2}{\partial x^2} P_s(x) \right] \sigma^2(x) dx \\ &= \mathbb{E} \left[\sigma^2(X_s) \left(\frac{\partial}{\partial x} \ln P_s(X_s) \right)^2 \right] - \mathbb{E} \left[\frac{\partial^2}{\partial x^2} \sigma^2(X_s) \right]. \end{aligned}$$

Therefore, we have the following formula.

$$\begin{aligned} -\frac{d}{dt} \mathbb{E}[\ln P_t(X_t)] &= Ht^{2H-1} \left\{ \mathbb{E} \left[\sigma^2(X_t) \left(\frac{\partial}{\partial x} \ln P_t(X_t) \right)^2 \right] - \mathbb{E} \left[\frac{\partial^2}{\partial x^2} \sigma^2(X_t) \right] \right. \\ &\quad \left. + \mathbb{E} [\sigma''(X_t)\sigma(X_t) + (\sigma'(X_t))^2] \right\}. \end{aligned} \tag{2.7}$$

2.1.2. *Proof of Theorem 2.2* In this proof, we write $P_t(x)$ to be the probability density of X_t , $Q_t(y)$ to be the probability density of Y_t and let $f(s, x) = \ln(P_s(x)/Q_s(x))$ in Lemma 2.1. Then we have

$$\begin{aligned} f_x(s, x) &= \left(\frac{P_s(x)}{Q_s(x)} \right)^{-1} \frac{\partial}{\partial x} \frac{P_s(x)}{Q_s(x)}, \\ f_{xx}(s, x) &= \left(\frac{P_s(x)}{Q_s(x)} \right)^{-1} \left(\frac{\partial^2}{\partial x^2} \frac{P_s(x)}{Q_s(x)} \right) - \left(\frac{P_s(x)}{Q_s(x)} \right)^{-2} \left(\frac{\partial}{\partial x} \frac{P_s(x)}{Q_s(x)} \right)^2 \\ &= \left(\frac{P_s(x)}{Q_s(x)} \right)^{-1} \left(\frac{\partial^2}{\partial x^2} \frac{P_s(x)}{Q_s(x)} \right) - \left(\frac{\partial}{\partial x} \ln \frac{P_s(x)}{Q_s(x)} \right)^2, \\ f_s(s, x) &= \frac{1}{P_s(x)} \frac{\partial}{\partial s} P_s(x) - \frac{1}{Q_s(x)} \frac{\partial}{\partial s} Q_s(x). \end{aligned}$$

As a result, using integration by part we arrive at

$$\begin{aligned} \mathbb{E}[f_s(s, X_s)] &= - \int \left(\frac{\partial}{\partial s} Q_s(x) \right) \frac{P_s(x)}{Q_s(x)} dx, \\ \mathbb{E}[f_x(s, X_s)\sigma'(X_s)\sigma(X_s)] &= - \int \left(\frac{\partial}{\partial x} \sigma'(x)\sigma(x)Q_s(x) \right) \frac{P_s(x)}{Q_s(x)} dx, \\ \mathbb{E}[f_{xx}(s, X_s)\sigma^2(X_s)] &= \int \left(\frac{\partial^2}{\partial x^2} \sigma^2(x)Q_s(x) \right) \frac{P_s(x)}{Q_s(x)} dx - J_{\sigma^2}(X_s||Y_s). \end{aligned}$$

Now, by Lemma 2.1 we note that

$$\begin{aligned} \frac{d}{dt} K(X_t||Y_t) &= \frac{d}{dt} \mathbb{E} \left[\ln \frac{P_t(X_t)}{Q_t(X_t)} \right] \\ &= \mathbb{E}[f_t(t, X_t)] + Ht^{2H-1} (\mathbb{E}[f_x(t, X_t)\sigma'(X_t)\sigma(X_t)] + \mathbb{E}[f_{xx}(t, X_t)\sigma^2(X_t)]) \\ &= -Ht^{2H-1} J_{\sigma^2}(X_t||Y_t), \end{aligned}$$

where the last equality follows from Lemma 2.2.

2.1.3. *Proof of Theorem 2.3* Let $t \geq 0$ be such that $s^H = \sqrt{t}$ and denote Z to follow the standard normal distribution. Using chain rule we have

$$\begin{aligned} \frac{d}{ds}h(X_s) &= \frac{d}{ds}h(X_0 + s^H Z) \\ &= \frac{d}{ds}h(X_0 + \sqrt{t}Z) \\ &= \frac{d}{dt}h(X_0 + \sqrt{t}Z) \frac{dt}{ds} \\ &= \frac{1}{2}J_1(X_0 + \sqrt{t}Z)2Hs^{2H-1} \\ &= Hs^{2H-1}J_1(X_s), \end{aligned}$$

where the fourth equality follows from the classical De Bruijn’s identity (see e.g., [9]). In particular when X_0 is Gaussian, then X_s is also Gaussian with mean 0 and variance $\sigma_0^2 + s^{2H}$. Since for normal distribution the Fisher information is the reciprocal of the variance, we have

$$\frac{d}{ds}h(X_s) = Hs^{2H-1}J_1(X_s) = \frac{Hs^{2H-1}}{\text{Var}(X_s)} = \frac{Hs^{2H-1}}{\sigma_0^2 + s^{2H}}.$$

3. APPLICATIONS

In this section, we present two applications of the generalized De Bruijn’s identity. In the first application in Section 3.1, we demonstrate its equivalence with the Stein’s identity for Gaussian distribution, while in Section 3.2, we prove the convexity or the concavity of entropy power, which depends on the Hurst parameter H . Throughout this section, we assume that the channel is of the form

$$X_t = X_0 + B_t^H,$$

where the initial value X_0 is independent of the fBm and the Hurst parameter $H \in (0, 1)$.

3.1. Equivalence of the generalized De Bruijn’s identity and Stein’s identity for normal distribution

It is known that the classical De Bruijn’s identity is equivalent to the Stein’s identity for normal distribution as well as the heat equation identity, provided that the initial noise X_0 is Gaussian, see e.g., [4,17]. These identities are equivalent in the sense that one can derive the others using any one of them. It is therefore natural for us to guess that the same equivalence also holds for the proposed generalized De Bruijn’s identity. To this end, let us recall the classical Stein’s identity for normal distribution. Writing Y to be the normal distribution with mean μ and variance σ^2 , the Stein’s identity is given by

$$\mathbb{E}[r(Y)(Y - \mu)] = \sigma^2\mathbb{E}\left[\frac{d}{dy}r(Y)\right], \tag{3.1}$$

where r is a differentiable function such that the above expectations exist. In the following result, we prove that the generalized De Bruijn’s identity presented in Theorem 2.3 is equivalent to the Stein’s identity,

THEOREM 3.1 (Equivalence of the generalized De Bruijn's identity (2.5) and Stein's identity): *Consider the channel $X = (X_t)_{t \geq 0}$ modeled by Eq. (2.4) with Hurst parameter $H \in (0, 1)$ and initial Gaussian X_0 independent of the fBm. Then the generalized De Bruijn's identity (2.5) is equivalent to the Stein's identity (3.1).*

Remark 3.1 (Equivalence of the generalized De Bruijn's identity (2.2) and Stein's identity): While Theorem 3.1 is established for channel (2.4), it is natural to consider if the same equivalence holds between (2.2) and Stein's identity for channel (2.1). One direction is straightforward: with (2.2) we have the classical De Bruijn's identity by taking the diffusion coefficient to be $\sigma(x) = 1$, and so we can derive the Stein's identity. However, for the opposite direction, we did not manage to prove (2.2) via the classical De Bruijn's identity in the presence of general diffusion coefficient $\sigma(x)$. The trick employed in proving Theorem 2.3, where the diffusion coefficient is simply 1, does not seem to carry over to this setting. If this can be proved by other means, then the equivalence could be established.

PROOF: If we have the Stein's identity, then we can derive the classical De Bruijn's identity [17], and so we have the generalized De Bruijn's identity by Theorem 2.3. For the other direction, if we have the generalized De Bruijn's identity, then we can derive the classical De Bruijn's identity by taking $H = 1/2$, and from it we can derive the Stein's identity by Park et al. [17]. ■

3.2. Convexity/concavity of the entropy power

Recall that the entropy power of a random variable X is defined to be

$$N(X) := \frac{1}{2\pi e} e^{2h(X)}. \tag{3.2}$$

In the classical setting when the channel X_t is of the form (2.4) with X_0 being an arbitrary initial noise, Costa [6], Dembo [10] prove that the entropy power of $N(X_t)$ is concave in time t . Recently in Khoolenjani and Alamatsaz [13] the authors extend the concavity of entropy power to the dependent case where the dependency structure between the initial value X_0 and the channel X_t is specified by Archimedean and Gaussian copulas. In our case, interestingly convexity/concavity of the entropy power depends on the Hurst parameter H :

THEOREM 3.2 (Convexity/concavity of the entropy power): *Consider the channel $X = (X_t)_{t \geq 0}$ modeled by Eq. (2.4) with Hurst parameter $H \in (0, 1)$, initial value X_0 independent of the fBm and has a finite second moment. We have*

$$\begin{aligned} \frac{d^2}{dt^2} N(X_t) &= 2N(X_t) \left(2H^2 t^{4H-2} J_1(X_t)^2 + H(2H - 1)t^{2H-2} J_1(X_t) + Ht^{2H-1} \frac{d}{dt} J_1(X_t) \right) \\ &= 2N(X_t)g(t, H, X_t), \end{aligned}$$

where $g(t, H, X_t) = 2H^2 t^{4H-2} J_1(X_t)^2 + H(2H - 1)t^{2H-2} J_1(X_t) + Ht^{2H-1}(d/dt)J_1(X_t)$. Consequently,

$$N(X_t) = \begin{cases} \text{convex in } t & \text{if } g(t, H, X_t) > 0, \\ \text{concave in } t & \text{if } g(t, H, X_t) \leq 0. \end{cases}$$

In particular, when X_0 is a Gaussian distribution with mean 0 and variance σ_0^2 , we then have $g(t, H, X_t) = H(2H - 1)t^{2H-2}J_1(X_t)$ and

$$N(X_t) = \begin{cases} \text{convex in } t & \text{if } H \in (1/2, 1), \\ \text{concave in } t & \text{if } H \in (0, 1/2]. \end{cases}$$

Remark 3.2: In the special case when $H = 1/2$ and X_0 is Gaussian, we retrieve the classical result that $N(X_t)$ is linear and hence concave (or convex) in t .

Remark 3.3 (The role of time parameter t): In [Theorem 3.2](#), we see that there are factors such as t^{2H-1} or t^{4H-2} appearing in the function $g(t, H, X_t)$. While these terms equal to 1 in the classical $H = 1/2$ case, as we shall see in the proof these terms play an important and interesting role in determining the second-order behavior of the entropy power in the general fBm case. Note that these terms all come from the factor t^{2H-1} in front of the Fisher information J_1 in [\(2.5\)](#).

Remark 3.4 (On establishing the convexity/concavity of the entropy power in model [\(2.1\)](#)): While [Theorem 3.2](#) is stated for model [\(2.4\)](#), we can in fact state a similar result for model [\(2.1\)](#) using [Theorem 2.1](#). However, there will not be a clear cut distinction between the two cases $H \leq 1/2$ and $H > 1/2$ as in [Theorem 3.2](#); it will also depend on the derivatives of the diffusion coefficient $\sigma(x)$, which may not be tractable in general.

PROOF: Using the definition of the entropy power [\(3.2\)](#), we have

$$\begin{aligned} \frac{d^2}{dt^2}N(X_t) &= \frac{d}{dt} \left(2N(X_t) \frac{d}{dt}h(X_t) \right) \\ &= 2N(X_t) \left(2 \left(\frac{d}{dt}h(X_t) \right)^2 + \frac{d^2}{dt^2}h(X_t) \right) \\ &= 2N(X_t) \left(2H^2t^{4H-2}J_1(X_t)^2 + H(2H - 1)t^{2H-2}J_1(X_t) + Ht^{2H-1} \frac{d}{dt}J_1(X_t) \right) \\ &= 2N(X_t)g(t, H, X_t), \end{aligned}$$

where we make use of the generalized De Bruijn’s identity [\(2.5\)](#) in the third equality. Since $N(X_t) \geq 0$, convexity/concavity of $N(X_t)$ thus depends on the sign of the function $g(t, H, X_t)$. In particular, when X_0 is Gaussian with mean 0 and variance σ_0^2 , we have

$$\begin{aligned} J_1(X_t) &= \frac{1}{\sigma_0^2 + t^{2H}}, \\ \frac{d}{dt}J_1(X_t) &= \frac{-2Ht^{2H-1}}{(\sigma_0^2 + t^{2H})^2} = -2Ht^{2H-1}J_1(X_t)^2, \\ g(t, H, X_t) &= H(2H - 1)t^{2H-2}J_1(X_t) \\ &= \begin{cases} > 0 & \text{if } H \in (1/2, 1), \\ \leq 0 & \text{if } H \in (0, 1/2]. \end{cases} \quad \blacksquare \end{aligned}$$

Theorem 3.2 implies that, for channel of the form (2.4) and Gaussian distributed X_0 , for $t \in [0, 1]$, we have

$$N(X_t) \begin{cases} \leq tN(X_0) + (1-t)N(X_1) & \text{if } H \in (1/2, 1), \\ \geq tN(X_0) + (1-t)N(X_1) & \text{if } H \in (0, 1/2]. \end{cases}$$

One application of the above equation lies in determining the so-called capacity region in a Gaussian interference channel; see the paper [7]. It turns out that the concavity of entropy power is a crucial step in the proof of Theorem 2 in Costa [7]. With our **Theorem 3.2**, it seems possible to study the capacity region in an interference channel with fBm noise and generalize the result to $H \in (0, 1/2]$. We leave this as one of our future research directions.

4. CONCLUSION

In this paper, we present the generalized De Bruijn's identity for the channel driven by fBm with Hurst parameter $H \in (0, 1)$. Compared with the classical Brownian motion, i.e., $H = \frac{1}{2}$, in our setting, the term t^{2H-1} in general does not degenerate unless $H = \frac{1}{2}$, and thus plays an essential role in the derivation of the identity. Consequently, we also investigate its equivalence with the Stein's identity and study the second-order behavior of the entropy power. We hope that this paper can open new doors in analyzing the De Bruijn's identity in a more general context to model phenomenon such as self-similarity and long-range dependency.

There are at least two future research directions. First, as mentioned in Section 3.2, we can study the capacity region in an interference channel with fBm noise, where our convexity/concavity result should be an important step in the analysis. Second, we can consider more general channel driven by stochastic processes such as the Gaussian Volterra process [21]. Members of this broad family include fBm as well as the Riemann–Liouville process. This shall provide a unified framework in studying the De Bruijn's identity and its applications for channels driven by the more general Gaussian processes.

Acknowledgments

The authors would like to thank the associate editor and the referee for constructive comments that improve the quality of the manuscript. Michael Choi acknowledges the support from the Chinese University of Hong Kong, Shenzhen grant PF01001143 and AIRS - Shenzhen Institute of Artificial Intelligence and Robotics for Society.

References

1. Alòs, E., Mazet, O., & Nualart, D. (2001). Stochastic calculus with respect to Gaussian processes. *Annals of Probability* 29(2): 766–801.
2. Baudoin, F. & Coutin, L. (2007). Operators associated with a stochastic differential equation driven by fractional Brownian motions. *Stochastic Processes and Their Applications* 117(5): 550–574.
3. Beran, J., Sherman, R., Taqqu, M.S., & Willinger, W. (1995). Long-range dependence in variable-bit-rate video traffic. *IEEE Transactions on Communications* 43(2/3/4): 1566–1579.
4. Brown, L., DasGupta, A., Haff, L.R., & Strawderman, W.E. (2006). The heat equation and Stein's identity: connections, applications. *Journal of Statistical Planning and Inference* 136(7): 2254–2278.
5. Cheridito, P. & Nualart, D. (2005). Stochastic integral of divergence type with respect to fractional Brownian motion with Hurst parameter $H \in (0, \frac{1}{2})$. *Annales de l'Institut Henri Poincaré B: Probability and Statistics* 41(6): 1049–1081, ISSN 0246–0203.
6. Costa, M.H.M. (1985). A new entropy power inequality. *IEEE Transactions on Information Theory* 31(6): 751–760.

7. Costa, M.H.M. (1985). On the Gaussian interference channel. *IEEE Transactions on Information Theory* 31(5): 607–615.
8. Coutin, L. & Qian, Z. (2002). Stochastic analysis, rough path analysis and fractional Brownian motions. *Probability Theory and Related Fields* 122(1): 108–140.
9. Cover, T.M. & Thomas, J.A. (2006). *Elements of information theory* (2nd ed.). Hoboken, NJ: Wiley-Interscience [John Wiley & Sons].
10. Dembo, A. (1989). Simple proof of the concavity of the entropy power with respect to added Gaussian noise. *IEEE Transactions on Information Theory* 35(4): 887–888.
11. Gradinaru, M., Nourdin, I., Russo, F., & Vallois, P. (2005). m -order integrals and generalized Itô's formula: the case of a fractional Brownian motion with any Hurst index. *Annales de l'Institut Henri Poincaré B: Probability and Statistics* 41(4): 781–806, ISSN 0246–0203.
12. Guo, D., Shamai, S., & Verdú, S. (2005). Mutual information and minimum mean-square error in Gaussian channels. *IEEE Transactions on Information Theory* 51(4): 1261–1282.
13. Khoolejani, N.B. & Alamatsaz, M.H. (2016). A De Bruijn's identity for dependent random variables based on copula theory. *Probability in the Engineering and Informational Sciences* 30(1): 125–140.
14. Mandelbrot, B.B. & Van Ness, J.W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review* 10: 422–437.
15. Nourdin, I. (2008). A simple theory for the study of SDEs driven by a fractional Brownian motion, in dimension one. In *Séminaire de probabilités XLI*, vol. 1934, Lecture Notes in Math, Springer, Berlin, pp. 181–197.
16. Palomar, D.P. & Verdú, S. (2006). Gradient of mutual information in linear vector Gaussian channels. *IEEE Transactions on Information Theory* 52(1): 141–154.
17. Park, S., Serpedin, E., & Qaraqe, K. (2012). On the equivalence between Stein and De Bruijn identities. *IEEE Transactions on Information Theory* 58(12): 7045–7067.
18. Russo, F. & Vallois, P. (1993). Forward, backward and symmetric stochastic integration. *Probability Theory and Related Fields* 97(3): 403–421, ISSN 0178–8051.
19. Stam, A.J. (1959). Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control* 2: 101–112.
20. Sussmann, H.J. (1978). On the gap between deterministic and stochastic ordinary differential equations. *Annals of Probability* 6(1): 19–41.
21. Valdivia, A. (2017). Information loss on Gaussian Volterra process. *Electronic Communications in Probability* 22: 5, Paper No. 60.
22. Wibisono, A., Jog, V., & Loh, P. (2017). Information and estimation in Fokker-Planck channels. In *2017 IEEE International Symposium on Information Theory, ISIT 2017, IEEE, Aachen, Germany, June 25–30, 2017*, pp. 2673–2677.
23. Willinger, W., Taqqu, M.S., Leland, W.E., & Wilson, D.V. (1995). Self-similarity in high-speed packet traffic: analysis and modeling of Ethernet traffic measurements. *Statistical Science* 10(1): 67–85.
24. Young, L.C. (1936). An inequality of the Hölder type, connected with Stieltjes integration. *Acta Mathematica* 67(1): 251–282, ISSN 0001–5962.