

**MAXIMUM LIKELIHOOD ESTIMATOR FOR THE DRIFT OF A
BROWNIAN FLOW**

Mine ÇAĞLAR

Princeton University, Princeton, NJ, USA¹

¹Current address: Guttenberg, NJ USA

SUMMARY

The maximum likelihood estimator for the drift of a Brownian flow on \mathbb{R}^d , $d \geq 2$, is found with the assumption that the covariance is known. By approximation of the drift with known functions, the statistical model is reduced to a parametric one that is a curved exponential family. The data is the n -point motion of the Brownian flow throughout the time interval $[0, T]$. The asymptotic properties of the MLE are also investigated.

KEY WORDS: Brownian flows, maximum likelihood estimation, function approximation, oceanography.

1 Introduction

The theory of stochastic flows have been built on that of stochastic differential equations which was initiated by K. Itô in 1942. Since then the theory has been developed in various directions.¹ In particular, a Brownian flow is generated by a stochastic differential equation involving standard Brownian motions, or Wiener processes, as the so called driving processes. The property that a Brownian flow is characterized by two functions only, namely drift and covariance, establishes its importance also from statistical point of view. In this paper, we study the estimation of the drift assuming that the covariance function is known.

The motivation for the sampling scheme comes from the drifter experiments in oceanography. Certain number of drifters are placed in a region of the ocean, and then their trajectories are fixed by satellites at frequent intervals. Assuming that they completely follow the flow, hence behave as proxies for fluid particles, and the trajectories are observed continuously in time, we define our sample data as the n -point motion of a Brownian flow in interval $[0, T]$. We define and briefly review Brownian flows in Section 2. The flow is defined as a process taking values in the space of homeomorphisms of \mathbb{R}^d , $d \geq 2$, but the n -point motion is essentially a diffusion on

\mathbb{R}^{nd} . The tools of stochastic calculus are used to get the maximum likelihood estimator (MLE) of the drift based on this data, in Section 3.

The estimation of the drift of a diffusion has been extensively studied in the literature where the basic approach is to parametrize it. Brown and Hewitt² employ the approximation of the drift through known functions for one-dimensional diffusions. Our problem is the estimation of a d -dimensional drift which we transform to the case of nd -dimensional diffusions properly. The usual approach is to assume that the drift is a linear combination of given functions. Hence estimation of a function reduces to estimation of real parameters.

The statistical model for Brownian flows with parametrization belongs to the exponential family.³ Although the multidimensionality of the parameter space necessitates stronger assumptions for asymptotic results, the MLE exists and can be displayed explicitly. The asymptotic properties are given in Section 4.

2 Preliminaries

On the probability space $(\Omega, \mathcal{H}, \mathbb{P})$, we define a stochastic flow of homeomorphisms as a collection $\{F_{st} : 0 \leq s \leq t\}$ of continuous random maps from \mathbb{R}^d to \mathbb{R}^d which satisfy the following properties almost surely:

- (a) $F_{ss} = \text{identity map}$ for all $s \in \mathbb{R}_+$,
- (b) $F_{st} \circ F_{rs} = F_{rt}$ for all $0 \leq r \leq s \leq t$,
- (c) $F_{st} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a homeomorphism for all $0 \leq s \leq t$.

In particular, a Brownian flow is a stochastic flow which further satisfies the following property:

- (d) for each integer $n \geq 1$ and all $0 \leq t_1 \leq t_2 \leq \dots \leq t_n$, the random maps $F_{t_1 t_2}, \dots, F_{t_{n-1} t_n}$ are independent.

We are interested in temporally homogeneous Brownian flows which means that the distribution of $F_{s+h,t+h}$ is free of $h \in \mathbb{R}_+$, for all $0 \leq s \leq t$. The *infinitesimal mean* u_0 and the *infinitesimal covariance* a of F are found by

$$u_0(x) = \lim_{h \downarrow 0} \frac{1}{h} \mathbb{E}[F_{t,t+h}(x) - x]$$

$$a(x, y) = \lim_{h \downarrow 0} \frac{1}{h} \mathbb{E}[(F_{t,t+h}(x) - x)(F_{t,t+h}(y) - y)^T]$$

A Brownian motion U taking values in $C(\mathbb{R}^d \rightarrow \mathbb{R}^d)$, the continuous functions from \mathbb{R}^d to \mathbb{R}^d , is defined as a stochastic process whose increments are Gaussian random fields on \mathbb{R}^d and are independent. That is, for every $t_1 \leq \dots \leq t_n$, the collection

$$\{U(\cdot, t_{i+1}) - U(\cdot, t_i) : i = 1, \dots, n\}$$

is an independent collection of continuous random fields on \mathbb{R}^d .

We have the following decomposition of U into independent, one-dimensional Wiener Processes W_1, W_2, \dots . There exist deterministic, continuous vector fields u_0, u_1, \dots on \mathbb{R}^d such that

$$U(x, t) = u_0(x)t + \sum_{k=1}^{\infty} u_k(x)W_k(t) . \quad (2.1)$$

The vector field u_0 is called the *drift* and the matrix function

$$a(x, y) = \sum_{k=1}^{\infty} u_k(x)u_k(y)^T \quad (2.2)$$

is called the *covariance* of U . The law of a Brownian motion is characterized by its drift and covariance functions. We have

$$\mathbb{E}U(x, t) = u_0(x)t \quad x \in \mathbb{R}^d, t \in \mathbb{R}_+ ,$$

and

$$\text{Cov}(U^i(x, s), U^j(y, t)) = a^{ij}(x, y)(s \wedge t) \quad i, j = 1, \dots, d, \quad x, y \in \mathbb{R}^d, \quad s, t \in \mathbb{R}_+ .$$

Given a Brownian flow $\{F_{st}, 0 \leq s \leq t\}$ satisfying global Lipschitz and linear growth conditions

$$|u_0(x)| \leq K(1 + |x|), \quad |u_0(x) - u_0(y)| \leq K|x - y|, \quad x, y \in \mathbb{R}^d$$

$$\|a(x, y)\| \leq (1 + |x|)(1 + |y|),$$

$$\|a(x, y) - a(x', y) - a(x, y') + a(x', y')\| \leq K|x - x'| |y - y'|, \quad x, x', y, y' \in \mathbb{R}^d$$

and the condition

$$|\mathbb{E}[F_{st}(x) - x]| \leq K(1 + |x|)|t - s|$$

for some $K > 0$, one can find a unique Brownian motion U with drift u_0 and covariance a such that for every fixed s and x , we have

$$F_{st}(x) = x + \int_s^t U(F_{sr}(x), dr), \quad (2.3)$$

which is a generalization of Itô's stochastic differential equation based on U .¹

Conversely, suppose the fields u_0, u_1, \dots that form a Brownian motion U as in (2.1) satisfy

$$\sum_{k=0}^{\infty} |u_k(x)|^2 \leq K(1 + |x|)^2 \quad (2.4)$$

and

$$\sum_{k=0}^{\infty} |u_k(x) - u_k(y)|^2 \leq K|x - y|^2. \quad (2.5)$$

Then, Equation (2.3) has a unique solution F which is a Brownian flow with infinitesimal mean u_0 and covariance a as given in (2.2) (Ref.1, pg.106). Hence, F is also characterized by u_0 and a .

The n -point motion of a Brownian flow can be interpreted as the finite dimensional projection of it. It is the process $(F_{0t}(x_1), \dots, F_{0t}(x_n))$ for fixed x_1, \dots, x_n in \mathbb{R}^d and is a diffusion on \mathbb{R}^{nd} with infinitesimal generator $A^{(n)}$ which satisfies

$$\begin{aligned} & (A^{(n)}f)(x_1, \dots, x_n) \\ &= \frac{1}{2} \sum_{p,q=1}^n \sum_{i,j=1}^d a_{ij}(x_p, x_q) \frac{\partial^2 f}{\partial x_p^i \partial x_q^j}(x_1, \dots, x_n) + \sum_{p=1}^n \sum_{i=1}^d u_0^i(x_p) \frac{\partial f}{\partial x_p^i}(x_1, \dots, x_n) \end{aligned}$$

for all $f \in C^2(\mathbb{R}^{nd} \rightarrow \mathbb{R})$ and $x_1, \dots, x_n \in \mathbb{R}^d$. The operator $A^{(2)}$ involves the covariance $a(x, y)$ for all $x, y \in \mathbb{R}^d$, consequently the law of two-point motion determines the law of a Brownian flow.

Using the relation (2.1), we can write Equation (2.3) in differential form as

$$dF_{st}(x) = u_0(F_{st}(x))dt + \sum_{k=1}^{\infty} u_k(F_{st}(x))W_k(dt) \quad (2.6)$$

for each $x \in \mathbb{R}^d$ and $0 \leq s \leq t$. In the context of the flow, $F_{st}(x)$ denotes the position of the particle at time t which started from position x at time s . Note that the time homogeneity of F is reflected in the fact that the drift and covariance of U do not depend on time.

3 The Maximum Likelihood Estimator

3.1 The Likelihood

Let (\mathcal{F}_t) be a filtration on the probability space $(\Omega, \mathcal{H}, \mathbb{P})$. The likelihood in estimation for stochastic processes is analogous to the classical likelihood definition in parameter estimation for random variables. Two measures μ and ν on the same measure space E are said to be absolutely continuous with respect to each other, when their null sets are the same. Then, according to the Radon-Nikodym Theorem there exists a unique density p such that

$$\mu(dx) = p(x)\nu(dx) .$$

Since the distribution of a process is a measure on the space it takes values, we define the *likelihood* as the density of the distribution of the process of interest, with respect to that of another process. The latter is to be chosen according to the parameters to be estimated. Below, we write $d\mu/d\nu$ instead of p .

The following theorem will be used to derive the likelihood function for the estimation of the drift in a Brownian flow. Here, two diffusion processes⁴ X and Y are considered. They have the

same covariance, but different drifts. The conditions of the theorem are regularity conditions which establish the existence and uniqueness of the processes X and Y , as well.

Theorem 3.1 *Let X and Y be \mathbb{R}^d -valued processes having the following differentials*

$$\begin{aligned} dX_t &= v_0(X_t, t)dt + \sum_{k=1}^{\infty} v_k(X_t, t)W_k(dt), & X_0 = Y_0 \\ dY_t &= w_0(Y_t, t)dt + \sum_{k=1}^{\infty} v_k(Y_t, t)W_k(dt) \end{aligned}$$

where W_k , $k = 1, 2, \dots$ are independent Wiener processes adapted to (\mathcal{F}_t) , and w_0, v_k , $k = 0, 1, \dots$ are \mathbb{R}^d -valued continuous functions defined on $\mathbb{R}^d \times \mathbb{R}_+$. Suppose that the following hold:

I. *There exists $K > 0$ such that*

$$\begin{aligned} |w_0(x, t)|^2 + \sum_{k=0}^{\infty} |v_k(x, t)|^2 &\leq K(1 + |x|^2) \\ |w_0(x, t) - w_0(y, t)|^2 + \sum_{k=0}^{\infty} |v_k(x, t) - v_k(y, t)|^2 &\leq K|x - y|^2 \end{aligned}$$

for all $x, y \in \mathbb{R}^d$, $t \in [0, T]$.

II. *The equation*

$$\sum_{k=1}^{\infty} v_k(X_t, t)\alpha_k(X_t, t) = v_0(X_t, t) - w_0(X_t, t)$$

has a bounded solution (α_k) for each $t \in [0, T]$.

III. *Almost surely*

$$\begin{aligned} \int_0^T v_0(Z_t, t)^T \left[\sum_{k=1}^{\infty} v_k(Z_t, t)v_k^T(Z_t, t) \right]^+ v_0(Z_t, t) dt \\ + \int_0^T w_0(Z_t, t)^T \left[\sum_{k=1}^{\infty} v_k(Z_t, t)v_k^T(Z_t, t) \right]^+ w_0(Z_t, t) dt < \infty \end{aligned}$$

for $Z = X$ and $Z = Y$, where $[N]^+$ denotes the pseudo inverse of a matrix N .

Then μ_X and μ_Y , the distributions of $\{X_t : t \in [0, T]\}$ and $\{Y_t : t \in [0, T]\}$ respectively, are absolutely continuous with respect to each other and

$$\begin{aligned} \frac{d\mu_X}{d\mu_Y}(X) &= \exp\left\{\int_0^T [v_0(X_t, t) - w_0(X_t, t)]^T \left[\sum_{k=1}^{\infty} v_k(X_t, t)v_k^T(X_t, t)\right]^+ dX_t \right. \\ &\quad \left. - \frac{1}{2} \int_0^T [v_0(X_t, t) - w_0(X_t, t)]^T \left[\sum_{k=1}^{\infty} v_k(X_t, t)v_k^T(X_t, t)\right]^+ [v_0(X_t, t) - w_0(X_t, t)] dt\right\} \end{aligned}$$

Proof. This theorem is the natural generalization of (the corollary to) Theorem 7.18 of Ref. 5, where X and Y satisfy the same differentials but with $k = 1, \dots, M$, $M < \infty$. \square

Let F be the flow based on the Brownian motion U of (2.1), as defined in Section 2. Since it is time homogeneous, we will concentrate on the processes $\{F_{0t}(x) : t \in [0, T]\}$ while x varies on \mathbb{R}^d , and $T > 0$. From (2.6), for each $x \in \mathbb{R}^d$, $F_t x \equiv F_{0t}(x)$ satisfies

$$dF_t x = u_0(F_t x)dt + \sum_{k=1}^{\infty} u_k(F_t x)W_k(dt) \quad . \quad (3.1)$$

Our aim is to estimate u_0 which is deterministic, yet unknown. Although the covariance function a has to be estimated also, we will assume that it is known in order to find the MLE of u_0 below, and then comment on how an estimate of it can be obtained for our purposes.

The estimation procedure will be based on the n -point motion data through $[0, T]$, that is, we observe the process $\{F_t x_i : x_i \in \mathbb{R}^d, i = 1, \dots, n, 0 \leq t \leq T\}$ where x_1, \dots, x_n are assumed to be pairwise different. This is a diffusion on \mathbb{R}^{nd} and satisfies the following differential equation

$$d\tilde{F}_t \tilde{x} = \tilde{u}_0(\tilde{F}_t \tilde{x})dt + \sum_k \tilde{u}_k(\tilde{F}_t \tilde{x})W_k(dt)$$

where

$$\tilde{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \tilde{F}_t \tilde{x} = \begin{bmatrix} F_t x_1 \\ F_t x_2 \\ \vdots \\ F_t x_n \end{bmatrix} \quad \tilde{u}_k(\tilde{x}) = \begin{bmatrix} u_k(x_1) \\ u_k(x_2) \\ \vdots \\ u_k(x_n) \end{bmatrix}$$

$x_i \in \mathbb{R}^d$, $i = 1, \dots, n$, $k = 0, 1, 2, \dots$, and W_k , $k = 1, 2, \dots$ are independent Wiener processes as before.

We will write the density of the law, that is the distribution, of the n -point motion $\tilde{F}\tilde{x}$, denoted by $\mu_{\tilde{F}\tilde{x}}$, with respect to the law of the n -point motion of a 0-drift Brownian flow G , denoted by $\mu_{\tilde{G}\tilde{x}}$. So, we choose G identical to F , except that its drift is zero. Analogous to F , the n -point motion of G satisfies

$$d\tilde{G}_t\tilde{x} = \sum_{k=1}^{\infty} \tilde{u}_k(\tilde{G}_t\tilde{x})W_k(dt) \quad .$$

Assuming for a moment that the assertions of Theorem 3.1 hold, we write the density as

$$\begin{aligned} \frac{d\mu_{\tilde{F}\tilde{x}}(\{\tilde{F}_t\tilde{x} : 0 \leq t \leq T\})}{d\mu_{\tilde{G}\tilde{x}}} = & \quad (3.2) \\ \exp\left\{\int_0^T [\tilde{u}_0(\tilde{F}_t\tilde{x})]^T [U(\tilde{F}_t\tilde{x})]^{-1} d\tilde{F}_t\tilde{x} - \frac{1}{2} \int_0^T [\tilde{u}_0(\tilde{F}_t\tilde{x})]^T [U(\tilde{F}_t\tilde{x})]^{-1} \tilde{u}_0(\tilde{F}_t\tilde{x}) dt\right\} \end{aligned}$$

where $U(\tilde{F}_t\tilde{x})$ is the matrix

$$\begin{bmatrix} K^{11} & K^{12} & \dots & K^{1n} \\ K^{21} & K^{22} & \dots & K^{2n} \\ \vdots & \vdots & & \vdots \\ K^{n1} & K^{n2} & \dots & K^{nn} \end{bmatrix} \quad (3.3)$$

formed by the blocks of $d \times d$ matrices

$$K^{ij} = a(F_t x_i, F_t x_j) \quad i, j = 1, \dots, n.$$

Theorem 3.1 states the sufficient conditions for absolute continuity of the laws of F and G when $n = 1$. However, for n -point motions with $n \geq 2$, we need stronger conditions. We refer to Theorems III.3.24 and III.5.19 in Ref.6 for this purpose. These theorems imply that if U is invertible, then the processes are well defined in (3.2) and it is indeed the desired density for all $n \in \mathbb{N}$.

It can easily be verified that U is symmetric and positive definite by the fact that a is symmetric and positive definite. For invertibility, we assume that it is strictly positive definite.

All possible matrices U given in (3.3) are strictly positive definite if and only if for all $n \in \mathbb{N}$, and all x_1, \dots, x_n in \mathbb{R}^d pairwise different, the matrices $[a(x_i, x_j)]_{i,j=1,\dots,n}$ are invertible. Considering all x_1, \dots, x_n in \mathbb{R}^d pairwise different is sufficient as F is a flow of homeomorphisms, since as a result, $F_t x_1, \dots, F_t x_n$ in \mathbb{R}^d remain pairwise different for all $t \geq 0$. In terms of the velocity fields u_k , $k = 1, \dots, n$, this is equivalent to the following condition.

Condition 3.1 For all $n \in \mathbb{N}$ and all $x_1, \dots, x_n \in \mathbb{R}^d$ pairwise different and for all $h_1, \dots, h_n \in \mathbb{R}^d$

$$\sum_{k=1}^{\infty} \left(\sum_{i=1}^n h_i^T u_k(x_i) \right)^2 > 0.$$

Equivalently,

$$\text{span} \left\{ \begin{bmatrix} u_k(x_1) \\ \vdots \\ u_k(x_n) \end{bmatrix} : k = 1, 2, \dots \right\} = \mathbb{R}^{\text{nd}}.$$

As a result, we assume Condition 3.1 is satisfied in addition to the conditions I,II,III of Theorem 3.1. Note that condition I is sufficient for the existence and uniqueness of the flow in view of equations (2.4) and (2.5). For the estimation purpose of u_0 , the density $d\mu_{\tilde{F}\tilde{x}}/d\mu_{\tilde{G}\tilde{x}}$ will now be interpreted as the likelihood of \tilde{u}_0 (or u_0) given the realization $\{(F_t x_i)(\omega) : 0 \leq t \leq T, x_i \in \mathbb{R}^d, i = 1, \dots, n\}$, $\omega \in \Omega$, and will be denoted by $\mathcal{L}(\tilde{u}_0; \{\tilde{F}_t \tilde{x} : 0 \leq t \leq T\})$ below.

Remark In general, absolute continuity of the laws of n -point motions requires much stronger conditions than those for the absolute continuity of one point motions in a flow. A striking case⁷ of an Ornstein-Uhlenbeck flow is iterated in Ref.8 as an example. Also, in some degenerate cases where $u_k \equiv 0$ for $k > m$ for some $m \in \mathbb{N}$, certain structural assumptions on u_0 are needed.⁸ In our case, note that Condition 3.1 excludes such a degeneracy.

3.2 The Estimator

In this subsection, our aim is to find an estimator for \tilde{u}_0 which maximizes the function $\mathcal{L}(\tilde{u}_0; \{\tilde{F}_t \tilde{x} : 0 \leq t \leq T\})$. Suppose that each component of the vector field u_0 is in $L^2(\mathbb{R}^d)$, in (3.1). Our approach will be to approximate u_0 by an orthonormal basis for $L^2(\mathbb{R}^d)$.

We assume that each component u_0^i , $i = 1, \dots, d$, is approximated by a finite number of functions ψ_m , $m = 1, \dots, M$, from an orthonormal basis as follows:

$$u_0^i \cong \sum_{m=1}^M \theta_{im} \psi_m \quad i = 1, \dots, d$$

where $\theta_{im} \in \mathbb{R}$, and M depends on the precision we want to achieve. In the n-point motion model of the previous subsection, we must then parametrize the velocity field \tilde{u}_0 in the following way

$$\tilde{u}_0(\tilde{x}) = \begin{bmatrix} u_0(x_1) \\ u_0(x_2) \\ \vdots \\ u_0(x_n) \end{bmatrix} = \begin{bmatrix} \theta & & & \\ & \theta & & \\ & & \ddots & \\ & & & \theta \end{bmatrix} \begin{bmatrix} \underline{\psi}(x_1) \\ \underline{\psi}(x_2) \\ \vdots \\ \underline{\psi}(x_n) \end{bmatrix} = \Theta \Psi(\tilde{x}) \quad (3.4)$$

where the matrix $\theta = [\theta_{ij}]$, $i = 1, \dots, d$, $j = 1, \dots, M$, and $\underline{\psi}(x)$ is the vector $[\psi_1(x) \dots \psi_M(x)]^T$ for each $x \in \mathbb{R}^d$. Hence, we will estimate u_0 using the n-point motion data. The log of the likelihood in Equation (3.2) is now given by

$$\begin{aligned} \log \mathcal{L}(\theta) &= \log \mathcal{L}(\tilde{u}_0; \{\tilde{F}_t \tilde{x} : 0 \leq t \leq T\}) \\ &= \int_0^T [\Theta \Psi(\tilde{F}_t \tilde{x})]^T [U(\tilde{F}_t \tilde{x})]^{-1} d\tilde{F}_t \tilde{x} \\ &\quad - \frac{1}{2} \int_0^T [\Theta \Psi(\tilde{F}_t \tilde{x})]^T [U(\tilde{F}_t \tilde{x})]^{-1} [\Theta \Psi(\tilde{F}_t \tilde{x})] dt \end{aligned} \quad (3.5)$$

since we have parametrized the estimation problem. Let B^{ij} , $i, j = 1, \dots, n$ be $d \times d$ matrix blocks in U^{-1} , as K^{ij} are in U as given in Equation (3.3). Then, we can display (3.5) more explicitly as

$$\log \mathcal{L}(\theta) = \int_0^T \sum_{i,j=1}^n \underline{\psi}(F_t x_i)^T \theta^T B^{ij}(\tilde{F}_t \tilde{x}) dF_t x_j \quad (3.6)$$

$$\begin{aligned}
& -\frac{1}{2} \int_0^T \sum_{i,j=1}^n \underline{\psi}(F_t x_i)^\top \theta^\top B^{ij}(\tilde{F}_t \tilde{x}) \theta \underline{\psi}(F_t x_j) dt \\
= & \sum_{i,j=1}^n \int_0^T \sum_{k=1}^d \sum_{m=1}^M \psi_m(F_t x_i) [\theta^\top]_{mk} [B^{ij}(\tilde{F}_t \tilde{x}) dF_t x_j]_k \\
& -\frac{1}{2} \sum_{i,j=1}^n \sum_{k,s=1}^d \sum_{r,m=1}^M \theta_{sr} [\theta^\top]_{mk} \int_0^T \psi_m(F_t x_i) [B^{ij}(\tilde{F}_t \tilde{x})]_{ks} \psi_r(F_t x_j) dt
\end{aligned}$$

where $[N]_{ij}$ denotes the (i, j) -entry of a matrix N . This notation is more useful in these intermediate steps; of course, $[\theta^\top]_{mk} = \theta_{km}$.

Let $\tilde{\theta} \in \mathbb{R}^{Md}$ denote a vector whose entries are obtained from the matrix θ in one-to-one correspondence. Then, (3.6) is equivalent to

$$\log \mathcal{L}(\theta) = b^\top \tilde{\theta} - \frac{1}{2} \tilde{\theta}^\top A \tilde{\theta}$$

where

$$\begin{aligned}
A_{pq} &= \sum_{i,j=1}^n \int_0^T \psi_m(F_t x_i) [B^{ij}(\tilde{F}_t \tilde{x})]_{ks} \psi_r(F_t x_j) dt \\
b_p &= \sum_{i,j=1}^n \int_0^T \psi_m(F_t x_i) [B^{ij}(\tilde{F}_t \tilde{x}) dF_t x_j]_k \\
&\quad \text{if } \tilde{\theta}_p = \theta_{km} \quad \text{and} \quad \tilde{\theta}_q = \theta_{sr}
\end{aligned} \tag{3.7}$$

$p, q \in \{1, 2, \dots, Md\}$, $k, s \in \{1, \dots, d\}$, $m, r \in \{1, \dots, M\}$. The following computation shows that A is strictly positive definite. Let $\tilde{y} \in \mathbb{R}^{Md}$ be arbitrary. Let y denote a $d \times M$ matrix formed by the elements of \tilde{y} with the following rule

$$y_{km} = \tilde{y}_p \quad \text{if} \quad \tilde{\theta}_p = \theta_{km}$$

$p = 1, \dots, Md$, $k = 1, \dots, d$, $m = 1, \dots, M$. Then, we have

$$\begin{aligned}
\tilde{y}^\top A \tilde{y} &= \sum_{k,s=1}^d \sum_{r,m=1}^M y_{sr} [y^\top]_{mk} \sum_{i,j=1}^n \int_0^T \psi_m(F_t x_i) [B^{ij}(\tilde{F}_t \tilde{x})]_{ks} \psi_r(F_t x_j) dt \\
&= \int_0^T \sum_{i,j=1}^n \underline{\psi}(F_t x_i)^\top y^\top B^{ij}(\tilde{F}_t \tilde{x}) y \underline{\psi}(F_t x_j) dt.
\end{aligned} \tag{3.8}$$

Recall that U is positive definite, moreover we have assumed that it is strictly positive definite.

It follows that U^{-1} is also strictly positive definite, that is, for any $z_i \in \mathbb{R}^d$, $i = 1, \dots, n$

$$\sum_{i,j=1}^n z_i^T B^{ij}(\cdot) z_j > 0 .$$

So, with $z_i = y \psi(F_t x_i)$, $i = 1, \dots, n$, the integrand in the second equality in (3.8) is strictly positive, for each t . This proves that A is strictly positive definite, which in turn implies that A is invertible. So, $\log \mathcal{L}$ is a concave quadratic function of $\tilde{\theta}$ and it has a maximum. Now, taking the derivative with respect to $\tilde{\theta}$, we get

$$\nabla \log \mathcal{L}(\theta) = b - A\tilde{\theta} .$$

Setting this equal to 0, we get the unique maximum likelihood estimator $\hat{\tilde{\theta}}$ of $\tilde{\theta}$ as

$$\hat{\tilde{\theta}} = A^{-1}b . \tag{3.9}$$

With the original indexing scheme, we denote the matrix obtained from $\hat{\tilde{\theta}}$ as $\hat{\theta}$. Then since $\hat{\theta}$ is the MLE for θ , the MLE \hat{u}_0 for u_0 is

$$\hat{u}_0 = \hat{\theta} \underline{\psi} .$$

Remark 1 In the previous subsection, we considered a time homogeneous Brownian motion, and the flow generated by this, that is, the drift and the covariance are not functions of time. However, Theorem 3.1 allowed them to be so and the discussion above would be still all right if we replaced $u_0(x)$ and $a(x, y)$ by $u_0(x, t)$ and $a(x, y, t)$, respectively, provided that we used the correct basis to decompose u_0 which would then be defined on $\mathbb{R}^d \times \mathbb{R}_+$.

Remark 2 Above, we have assumed a is known. If it is not, in order to get \hat{u}_0 , we must use an approximation for $B^{ij}(\tilde{F}_t \tilde{x})$ in (3.7). Usually, the data $\{(\tilde{F}_t \tilde{x})(\omega), 0 \leq t \leq T\}$ is not available for all $t \in [0, T]$, but only at $t_1, t_2, \dots, t_m \in [0, T]$, and we have to approximate the integrals in (3.7) by discrete sums. So, we suggest the following approximation \hat{a} to a :

$$\hat{a}(F_{t_k} x_i, F_{t_k} x_j) = \frac{[F_{t_{k+1}} x_i - F_{t_k} x_i][F_{t_{k+1}} x_j - F_{t_k} x_j]^T}{t_{k+1} - t_k} .$$

4 The Properties

The likelihood function can be written in a simplified form with the notation introduced in Section 3, as

$$\mathcal{L}(\theta) = \exp\{\tilde{\theta}^T b - \frac{1}{2}\tilde{\theta}^T A \tilde{\theta}\}$$

Because of this structure, the model is a curved exponential family with minimal sufficient statistic (b, A) .³ Moreover, the conditional Fisher information coincides with the observed one in this model and is given by A . Note that A depends on T and for asymptotic results below it is more convenient to regard it as a process $\{A(t) : t \geq 0\}$. Then each component A_{pp} can be interpreted as the Fisher information process associated with $\tilde{\theta}_p$, $p \in \{1, \dots, Md\}$.

From Equation (3.1) and Equation (3.4), we can write

$$dF_t x_i = \theta \underline{\psi}(x_i) dt + \sum_{k=1}^{\infty} u_k(F_t x_i) W_k(dt) \quad i = 1, \dots, n \quad (4.1)$$

where θ is the true parameter in matrix form. Let θ^0 be its vector form with the indexing scheme introduced in Section 3. Taking Equation (4.1) into account, from the expression for b in (3.7), we get

$$b = b' + A\theta^0$$

where

$$b'_p = \sum_{i,j=1}^n \sum_{m=1}^d \sum_{k=1}^{\infty} \int_0^T \psi_r(F_t x_i) [B^{ij}(\tilde{F}_t \tilde{x})]_{sm} [u_k(F_t x_j)]_m W_k(dt) \quad \text{if } \tilde{\theta}_p = \theta_{sr} .$$

Then since $A\hat{\tilde{\theta}} = b$ from (3.9), we get

$$b' = A(\hat{\tilde{\theta}} - \theta^0) . \quad (4.2)$$

Note that b' is a local martingale and the following computation shows that A is the quadratic variational process of b' .⁴ Let $\theta_{\tilde{k}\tilde{m}} = \tilde{\theta}_p$ and $\theta_{sr} = \tilde{\theta}_q$. From the stochastic calculus of Wiener processes, we have $W_k(dt)W_{k'}(dt') = dt$ if $k = k'$ and $t = t'$, and $W_k(dt)W_{k'}(dt') = 0$, otherwise.

Then, the cross variational process of b'_p and b'_q is

$$\sum_{i,j,i',j'=1}^n \sum_{m,m'=1}^d \int_0^T \psi_{\tilde{m}}(F_t x_{i'}) [B^{i'j'}(\tilde{F}_t \tilde{x})]_{\tilde{k}m'} \psi_r(F_t x_i) [B^{ij}(\tilde{F}_t \tilde{x})]_{sm} \sum_{k=1}^{\infty} [u_k(F_t x_j)]_m [u_k(F_t x_{j'})]_{m'} dt,$$

and in view of (2.2) and (3.3), we get

$$\sum_{i,j,i',j'=1}^n \sum_{m,m'=1}^d \int_0^T \psi_{\tilde{m}}(F_t x_{i'}) [B^{i'j'}(\tilde{F}_t \tilde{x})]_{\tilde{k}m'} [B^{ij}(\tilde{F}_t \tilde{x})]_{sm} [K^{jj'}]_{mm'} \psi_r(F_t x_i) dt . \quad (4.3)$$

But, $\sum_{j,m} [B^{ij}]_{sm} [K^{jj'}]_{mm'}$ is equal to 1 if $s = m'$ and $i = j'$, and is 0 otherwise, since K and B are blocks in U and U^{-1} , respectively. So, (4.3) is simplified as

$$\sum_{i,i'=1}^n \int_0^T \psi_{\tilde{m}}(F_t x_{i'}) [B^{i'i}(\tilde{F}_t \tilde{x})]_{\tilde{k}s} \psi_r(F_t x_i) dt$$

which is A_{pq} .

When $\mathbb{E}A(t) < \infty$ holds for all $t \geq 0$, the process $\{b'(t) : t \geq 0\}$ becomes a martingale. This allows us to make use of a central limit theorem (CLT) for martingales. Hence, we have the following theorem which establishes that the MLE for θ^0 is asymptotically Gaussian and is consistent, as $t \rightarrow \infty$. The conditions of the theorem are on $A(t)$, but implicitly are on the covariance function a of the Brownian motion U of (2.1). The notations $X \xrightarrow{\mathbb{P}} Y$ and $X \xrightarrow{d} Y$ denote convergence of X to Y in probability and distribution, respectively.

Theorem 4.1 *Suppose that $L_t = \mathbb{E}A(t) < \infty$ for all $t \geq 0$ and the diagonal elements of L_t tend to infinity as $t \rightarrow \infty$. Let $K_t = \text{diag}(k_t^1, \dots, k_t^{Md})$ where $(k_t^i)^2 = L_t(i, i) = \mathbb{E}A_{ii}(t)$, $i = 1, \dots, Md$. If as $t \rightarrow \infty$*

$$K_t^{-1}A(t)K_t^{-1} \xrightarrow{\mathbb{P}} \eta$$

and

$$K_t^{-1}L_tK_t^{-1} \longrightarrow \Sigma$$

where η is a random positive definite matrix and Σ is a deterministic strictly positive definite matrix, then conditionally on $\{\det(\eta) > 0\}$, we have

$$(K_t^{-1}A(t)K_t^{-1})^{1/2}K_t(\hat{\theta}(t) - \theta^0) \xrightarrow{d} Z \quad (4.4)$$

$$(\hat{\theta}(t) - \theta^0)^T A(t) (\hat{\theta}(t) - \theta^0) \xrightarrow{d} Y \quad (4.5)$$

and

$$\hat{\theta}(t) \xrightarrow{\mathbb{P}} \theta^0 \quad (4.6)$$

as $t \rightarrow \infty$, where Z is a random vector with Gaussian distribution of mean 0 and covariance matrix I , identity matrix, and Y is a random vector with χ^2 distribution with Md degrees of freedom. In case $k_t^i, i = 1, \dots, Md$ tend to infinity at the same rate, we have

$$A^{1/2}(t) (\hat{\theta}(t) - \theta^0) \xrightarrow{d} Z \quad . \quad (4.7)$$

Proof. Because of the assumption $L_t = \mathbb{E}A(t) < \infty$ for all $t \geq 0$, the process b' becomes a martingale. Then, applying the martingale central limit theorem⁹ to b' , we get

$$(K_t A^{-1}(t) K_t)^{1/2} K_t^{-1} b'(t) \xrightarrow{d} Z$$

But from Equation (4.2) $b' = A(\hat{\theta} - \theta^0)$, so by the identity

$$(K_t A^{-1}(t) K_t)^{1/2} K_t^{-1} A = (K_t^{-1} A(t) K_t^{-1})^{1/2} K_t$$

we get (4.4). The statement (4.5) follows directly from the martingale central limit theorem, and (4.6) is a consequence of (4.4). Finally, (4.7) is a particular case of (4.4). \square

Acknowledgement I would like to thank an anonymous referee for helpful and detailed comments, in particular, concerning the sufficiency conditions for absolute continuity.

REFERENCES

1. H. Kunita, *Stochastic Flows and Stochastic Differential Equations*, Cambridge University Press, Cambridge, 1990.
2. B.M. Brown and J.I. Hewitt, 'Asymptotic Likelihood Theory for Diffusion Processes', *J. Appl. Prob.* **12**, 228-238, 1975.

3. M. Sorensen, 'On Maximum Likelihood Estimation in Randomly Stopped Diffusion-Type Processes', *International Statistical Review*, **51**, 93-110, 1983.
4. N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, second edn., North Holland and Kodansha Ltd., Tokyo, 1989.
5. R.S Liptser and A. Shiryaev, *Statistics of Random Processes*, vol. I,II. Springer-Verlag, New York, 1977-1978.
6. J. Jacod and A. Shiryaev, *Limit Theorems for Stochastic Processes*, Springer-Verlag, Berlin, 1987.
7. E. Çinlar and J.S. Kao, 'Birth and Death on a Flow' in *Diffusion Processes and Related Problems in Analysis, II. Stochastic Flows*, Pinsky, Wihstutz (Eds.), Birkhauser, Boston, 1992.
8. R. Höpfner and E. Löcherbach, 'On Statistical Models for Birth and Death on a Flow: Local Absolute Continuity and Likelihood Ratio Processes', *Scand. J. Stat.* **26**, 107-128, 1999.
9. M. Sorensen, 'Likelihood methods for diffusions with jumps', in *Statistical Inference in Stochastic Processes*, M. Prabhu and I.V. Basawa, Eds, 1991, Appendix.