Stochastics and Stochastics Reports, Vol. 70, pp. 109–129
Reprints available directly from the publisher
Photocopying permitted by license only

SMALL NOISE ASYMPTOTICS OF THE GLR TEST FOR OFF-LINE CHANGE DETECTION IN MISSPECIFIED DIFFUSION PROCESSES

FABIEN CAMPILLO[a], YURY KUTOYANTS[b] and FRANÇOIS LE GLAND[c,]*

[a]*LATP/INRIA, 38 rue Joliot Curie, 13451 MARSEILLE Cédex 20, France;* [b]*Laboratoire de Statistique et Processus, Université du Maine, 72085 LE MANS Cédex 9, France;* [c]*IRISA/INRIA, Campus de Beaulieu, 35042 RENNES Cédex, France*

(Received 17 May 1999; In final form 18 February 2000)

We consider the problem of the non-sequential detection of a change in the drift coefficient of a stochastic differential equation, when a misspecified model is used. We formulate the generalized likelihood ratio (GLR) test for this problem, and we study the behaviour of the associated error probabilities (false alarm and nodetection) in the small noise asymptotics. We obtain the following robustness result: even though a wrong model is used, the error probabilities go to zero with exponential rate, and the maximum likelihood estimator (MLE) of the change time is consistent, provided the change to be detected is larger (in some sense) than the misspecification error. We give also computable bounds for selecting the threshold of the test so as to achieve these exponential rates.

Keywords: Change detection; diffusion process; generalized likelihood ratio test; small noise asymptotics; threshold selection; robustness

AMS Mathematics Subject Classifications (1991): 62M02, 60J60

1. INTRODUCTION

The problem of detecting abrupt changes at some unknown change time in the statistical characteristics of a dynamical system has numerous

*Corresponding author. Tel.: +33 2 99 84 73 62, Fax: +33 2 99 84 71 71, e-mail: legland@irisa.fr

applications, *e.g.*, in industrial quality control, navigation system monitoring, seismic data processing, segmentation of signals, edge detection in images, *etc.* A recent and very complete presentation of change detection problems in discrete-time stochastic systems, including both theory and applications, is given in Basseville and Nikiforov [1]. In general, three types of problems can be considered:

- off-line detection of change (non-sequential hypotheses testing),
- on-line quickest detection of change (sequential hypotheses testing),
- estimation of the change time (parameter estimation).

The quickest detection problem for a Wiener process with drift is considered in Shiryayev [9, Section 4.4]. Using a Bayesian framework, an optimal stopping time τ^* is found, which minimizes a risk function, and the decision is then to accept the hypothesis that a change has occurred in the time interval $[0, \tau^*]$. The problem of estimating the change time in a deterministic signal with additive Gaussian white noise is considered in Ibragimov and Khasminskii [3, Section VII.2]. The asymptotic properties of the maximum likelihood and Bayes estimators of the change time are described as the noise intensity goes to zero. The same problem is considered as the observation time goes to infinity in Vostrikova [12]. These results are generalized to diffusion processes with small noise in Kutoyants [4, Section 3.5] and [6, Chapter 5].

We consider here the problem of the non-sequential detection of a change in the drift coefficient of a stochastic differential equation, when a misspecified model is used. Our purpose is to formulate the generalized likelihood ratio (GLR) test for this problem, and to study the behaviour of the associated error probabilities (false alarm and nodetection) in the small noise asymptotics.

Here, misspecification is understood as follows: There always exists a gap between the idealized mathematical model which is used in the statistical algorithm, and the mechanism which generates the actual observations. Also, it can happen that the true model is very complicated, and has to be replaced by a simpler (and wrong) model. As a consequence, it is important to study the behaviour of the statistical algorithms under misspecification, *i.e.*, when the mathematical model does not coincide with the actual observations. One possible approach for taking robustness into account, is to consider local deviations on the parametric model. This approach has proved rather fruitful in general situations, see Huber [2]. Sometimes however, the assumption that the mathematical model and the actual observations are close is not realistic. In such situations, and if the underlying models are *regular* (in the statistical sense introduced by LeCam), then the estimators are not consistent, since the *true value* of the parameter does not exist. Nevertheless, it is possible to study the behaviour of some estimators under these circumstances, see White [13], McKeague [8], and Kutoyants [5] and [6, Section 2.6]. The change detection problem, which is the problem considered here, corresponds to a *non-regular* situation. For this reason, the *true value* of the parameter does exist, and the algorithm of change detection is consistent, *i.e.*, we obtain *a consistent algorithm based on the wrong model.*

Our motivation for using the small noise asymptotics is twofold: First, the models with small noise are simple to analyze, even simpler than models with a large number of i.i.d. observations. Secondly, they are nonlinear and sufficiently general, and the results obtained for these models could be later generalized to other models. The small noise asymptotics is also justified from the point of view of applications, since many dynamical systems can be considered as deterministic systems with small random perturbations.

The structure of the paper is the following: In Section 2 we define the statistical model and we formulate the GLR test for the problem of change detection. In Sections 3 and 4 we study the asymptotic behaviour of the probability of false alarm and the probability of nodetection, respectively, when the model is misspecified. We prove that the error probabilities go to zero with exponential rate, as the noise intensity goes to zero, provided some detectability assumption is satisfied by the limiting deterministic system, and provided that the modeling errors are smaller (in some sense) than the change to be detected. In Section 5 we discuss how the threshold should be selected so as to guarantee that both the probability of false alarm and the probability of nodetection go to zero with exponential rate. Under the hypothesis that a change has occurred, we prove also that the maximum likelihood estimator (MLE) of the change time is consistent. This means that the correct decision and estimation are made, even though a wrong model is used.

2. STATISTICAL MODEL

Let $\{X_t, 0 \leq t \leq T\}$ denote the available m-dimensional observation over a finite time interval $[0, T]$, and consider the problem of deciding between the following two hypotheses

- Under the null hypothesis $\mathbf{H}_0$

$$dX_t = b(X_t)dt + \varepsilon d W_t^\varepsilon, \quad X_0 = z, \;\; 0 \le t \le T, \tag{1}$$

where $\{W_t^\varepsilon, 0 \le t \le T\}$ is a m-dimensional Wiener process, with identity covariance matrix.

- Under the alternate hypothesis $\mathbf{H}_1$, there exists a change time $0 \le \tau \le T'$ with $T' < T$, such that

$$dX_t = [b(X_t) + a(X_t)\mathbf{1}_{(t\ge\tau)}]dt + \varepsilon d W_t^{\varepsilon,\tau}, \quad X_0 = z, \;\; 0 \le t \le T, \tag{2}$$

where $\{W_t^{\varepsilon,\tau}, 0 \le t \le T\}$ is a m-dimensional Wiener process, with identity covariance matrix.

Let $\mathbf{Q}^\varepsilon$ and $\{\mathbf{Q}_\tau^\varepsilon, 0 \le \tau \le T'\}$ denote the corresponding probability measures induced on the canonical space $C([0,T];\mathbb{R}^m)$.

Remark 2.1 The reason for requiring $T' < T$ is intuitively clear. It is impossible to detect a change occurring immediately before the end of the observation interval $[0,T]$, unless increasing dramatically the probability of false alarm. From the mathematical point of view, this will be reflected in the *detectability* assumption (18).

In the model above and throughout the paper, it is assumed that all the drift and change coefficients are Lipschitz continuous. Since any Lipschitz continuous function satisfies a linear growth condition, this assumption is sufficient to guarantee the existence and uniqueness of a solution to the stochastic differential equations (1) and (2), see Liptser–Shiryayev [7, Section 4.4]. In addition, the change coefficient is assumed to be *bounded*, so as to obtain exponential bounds for probability errors, based on the exponential bound (20) for continuous martingales. It is assumed for simplicity that the diffusion coefficient is constant, but all the results in this paper would generalize immediately to the case of a bounded non-degenerate state dependent diffusion coefficient.

Let $\ell_\varepsilon(\tau)$ denote the suitably normalized log-likelihood function for estimating the change time τ

$$\ell_\varepsilon(\tau) = \varepsilon^2 \log \frac{d\mathbf{Q}_\tau^\varepsilon}{d\mathbf{Q}^\varepsilon}.$$

The following expression holds

$$\ell_\varepsilon(\tau) = \int_\tau^T a^*(X_t)[dX_t - b(X_t)dt] - \frac{1}{2}\int_\tau^T |a(X_t)|^2 dt, \tag{3}$$

which depends only on the available observations $\{X_t,\ 0 \le t \le T\}$. Using Eqs. (1) and (2), the following equivalent expressions are obtained

$$\ell_\varepsilon(\tau) = \varepsilon \int_\tau^T a^*(X_t) d W_t^\varepsilon - \frac{1}{2} \int_\tau^T |a(X_t)|^2 dt, \tag{4}$$

and

$$\ell_\varepsilon(\tau) = \varepsilon \int_\tau^T a^*(X_t)\, d W_t^{\varepsilon,\tau} + \frac{1}{2} \int_\tau^T |a(X_t)|^2 dt. \tag{5}$$

Heuristically, if no change occurs at all, then it follows from expression (4) that $\ell_\varepsilon(\tau)$ will be nonpositive, asymptotically as $\varepsilon \downarrow 0$, for all $0 \le \tau \le T$. On the other hand, if a change actually occurs at time τ, then it follows from expression (5) that $\ell_\varepsilon(\tau)$ will be nonnegative, asymptotically as $\varepsilon \downarrow 0$, at least for this particular value τ. This motivates the introduction of the following statistics.

According to Van Trees [11, Section 2.5] the generalized likelihood ratio (GLR) test for deciding between the simple hypothesis $\mathbf{H}_0$ and the composite hypothesis $\mathbf{H}_1$ is defined by the following region

$$D_\varepsilon = \left\{ \sup_{0 \le \tau \le T'} \ell_\varepsilon(\tau) > c \right\},$$

for rejecting $\mathbf{H}_0$, where c is a given threshold. In other words, if the event D_ε holds, then the null hypothesis $\mathbf{H}_0$ is rejected, and the maximum likelihood estimator (MLE) for the change time τ is defined by

$$\hat{\tau}_\varepsilon \in \operatorname*{argmax}_{0 \le \tau \le T'} \ell_\varepsilon(\tau).$$

Under the alternate hypothesis $\mathbf{H}_1$, the behaviour (consistency, asymptotic probability distribution) of the estimator $\hat{\tau}_\varepsilon$ has been investigated for the one-dimensional case in Kutoyants [4, Theorem 3.5.2].

The purpose of this paper is to prove that both the probability of false alarm and the probability of nodetection associated with the above GLR test, go to zero with exponential rate when $\varepsilon \downarrow 0$. These error probabilities are evaluated under the *true* model, which may be different from the model (1), (2) used to build the test. The *true* model is defined as follows: If no change actually occurs, then

$$dX_t = b_0(X_t)dt + \varepsilon d V_t^\varepsilon, \quad X_0 = z, \quad 0 \le t \le T, \tag{6}$$

where $\{V_t^\varepsilon, 0 \leq t \leq T\}$ is a m-dimensional Wiener process with identity covariance matrix, and under the alternate hypothesis, if a change actually occurs at time τ, then

$$dX_t = [b_0(X_t) + a_0(X_t)\mathbf{1}_{(t \geq \tau)}]dt + \varepsilon d\, V_t^{\varepsilon,\tau}, \quad X_0 = z, \quad 0 \leq t \leq T, \tag{7}$$

where $\{V_t^{\varepsilon,\tau}, 0 \leq t \leq T\}$ is a m-dimensional Wiener process with identity covariance matrix. Let $\mathbf{P}^\varepsilon$ and $\{\mathbf{P}_\tau^\varepsilon, 0 \leq \tau \leq T'\}$ denote the corresponding probability measures induced on the canonical space $C([0, T]; \mathbb{R}^m)$. The probability of false alarm and the probability of nodetection are defined respectively as

$$F_\varepsilon = \mathbf{P}^\varepsilon(D_\varepsilon) \quad \text{and} \quad N_\varepsilon = \sup_{0 \leq \tau_0 \leq T'} \mathbf{P}_{\tau_0}^\varepsilon(D_\varepsilon^c).$$

Here and throughout the paper, τ_0 will denote the *true* but unknown change time.

Notice that

$$d\,X_t - b(X_t)dt = [b_0(X_t) - b(X_t)]dt + \varepsilon d\, V_t^\varepsilon, \quad 0 \leq t \leq T,$$

and for all $0 \leq \tau_0 \leq T'$

$$d\,X_t - b(X_t)dt = [[b_0(X_t) - b(X_t)] + a_0(X_t)\mathbf{1}_{(t \geq \tau_0)}]dt + \varepsilon d\, V_t^{\varepsilon,\tau_0}, \quad 0 \leq t \leq T,$$

which yield the following two equivalent expressions for the normalized log-likelihood function $\ell_\varepsilon(\tau)$, after substitution into (3)

$$\ell_\varepsilon(\tau) = \varepsilon \int_\tau^T a^*(X_t)d\, V_t^\varepsilon + \int_\tau^T a^*(X_t)[b_0(X_t) - b(X_t)]dt - \frac{1}{2}\int_\tau^T |a(X_t)|^2 dt, \tag{8}$$

and

$$\ell_\varepsilon(\tau) = \varepsilon \int_\tau^T a^*(X_t)d\, V_t^{\varepsilon,\tau_0} + \int_\tau^T a^*(X_t)[b_0(X_t) - b(X_t)]dt + \int_{\tau \vee \tau_0}^T a^*(X_t)a_0(X_t)dt - \frac{1}{2}\int_\tau^T |a(X_t)|^2 dt, \tag{9}$$

respectively.

We prove below that the error probabilities F_ε and N_ε go to zero with exponential rate, as the noise intensity ε goes to zero, under some additional conditions expressing that the modeling errors are smaller (in some sense) than the change to be detected. Under the hypothesis that a change has occurred, we prove also that the maximum likelihood estimator $\hat{\tau}_\varepsilon$ of the change time is consistent. The behaviour (consistency, asymptotic probability distribution) of the estimator $\hat{\tau}_\varepsilon$ is studied for the one-dimensional case in Kutoyants [6, Theorems 5.6 and 5.7]. This means that the correct decision and estimation are made, even though a wrong model is used.

3. PROBABILITY OF FALSE ALARM

Recall the expression (8) for the normalized log-likelihood function

$$\begin{aligned}\ell_\varepsilon(\tau) &= \varepsilon \int_\tau^T a^*(X_t)d\,V_t^\varepsilon + \int_\tau^T a^*(X_t)[b_0(X_t) - b(X_t)]dt - \frac{1}{2}\int_\tau^T |a(X_t)|^2 dt \\ &= \varepsilon \int_\tau^T a^*(X_t)d\,V_t^\varepsilon - \int_\tau^T \left[\frac{1}{2}|a(X_t)|^2 - a^*(X_t)\chi(X_t)\right]dt,\end{aligned}$$

where the misspecification coefficient $\chi \triangleq b_0 - b$ is the difference between the *true* and the *assumed* drift coefficients, and where $\{V_t^\varepsilon, 0 \le t \le T\}$ is a m-dimensional Wiener process under the probability measure $\mathbf{P}^\varepsilon$. Introduce the limiting expression obtained as $\varepsilon \downarrow 0$ under the probability measure $\mathbf{P}^\varepsilon$

$$\ell_0(\tau) = -\int_\tau^T \left[\frac{1}{2}|a(x_t)|^2 - a^*(x_t)\chi(x_t)\right]dt,$$

where $\{x_t, 0 \le t \le T\}$ is the solution of the limiting deterministic system

$$\dot{x}_t = b_0(x_t), \quad x_0 = z, \quad 0 \le t \le T. \tag{10}$$

Notice that the misspecification coefficient χ is bounded along the limiting trajectory, *i.e.*,

$$\sup_{0 \le t \le T} |\chi(x_t)| < \infty. \tag{11}$$

We introduce the following *consistency* assumption.

ASSUMPTION FA *For all* $0 \le t \le T$

$$\frac{1}{2}|a(x_t)|^2 - a^*(x_t)\chi(x_t) \ge 0.$$

Remark 3.1 Using the identity

$$\frac{1}{2}|a|^2 - a^*\chi = \frac{1}{2}|a - \chi|^2 - \frac{1}{2}|\chi|^2,$$

an equivalent form of Assumption FA is: For all $0 \leq t \leq T$

$$|b_0(x_t) - b(x_t)| \leq |b_0(x_t) - [b(x_t) + a(x_t)]|,$$

which means that along the limiting trajectory, the true drift b_0 should be closer to the assumed drift before change b, than it is to the assumed drift after change $b+a$. In general, it is difficult to check this assumption since the limiting trajectory $\{x_t,\ 0 \leq t \leq T\}$ depends also on the unknown coefficient b_0. However, a sufficient condition for Assumption FA to hold is the following stronger assumption.

ASSUMPTION FA′ *For all* $x \in \mathbb{R}^m$

$$\frac{1}{2}|a(x)|^2 - a^*(x)\chi(x) \geq 0.$$

Remark 3.2 If there is no misspecification, *i.e.*, if $b = b_0$, then Assumption FA′, and a fortiori Assumption FA, is always satisfied.

Under Assumption FA, it is easily checked that the mapping $\tau \mapsto \ell_0(\tau)$ achieves its maximum for $\tau = T'$, *i.e.*,

$$\begin{aligned} \ell_0^* \triangleq \sup_{0 \leq \tau \leq T'} \ell_0(\tau) &= \ell_0(T') \\ &= -\int_{T'}^{T} \left[\frac{1}{2}|a(x_t)|^2 - a^*(x_t)\chi(x_t)\right] dt \leq 0. \end{aligned} \tag{12}$$

Heuristically, as $\varepsilon \downarrow 0$

$$\mathbf{P}^\varepsilon\left[\sup_{0 \leq \tau \leq T'} \ell_\varepsilon(\tau) > c\right] \approx \mathbf{1}_{(\ell_0^* > c)},$$

hence $F_\varepsilon \approx 0$ if $\ell_0^* < c$. This is made rigorous in the following statement, where a nonasymptotic upper bound is given for the exponential rate of convergence of F_ε to zero.

THEOREM 3.3 *Assume that*

- *the drift and change coefficients* b_0, b *and* a *are Lipschitz continuous,*
- *the assumed change coefficient* a *is bounded.*

Then, the probability of false alarm F_ε satisfies

$$F_\varepsilon \le C \exp\left\{ -\frac{\theta^2 \delta_F^2}{8K^2 T \varepsilon^2} \right\},$$

provided Assumption FA holds, and provided the threshold c satisfies

$$\delta_F \triangleq c - \ell_0^* > 0.$$

The constant K depends only on the uniform bound of the assumed change coefficient a and on the bound of the misspecification coefficient χ along the limiting trajectory, and the constant $0 < \theta \le 1$ depends only on the Lipschitz constants and on the final time T.

Proof Assume first that the following large deviations estimate

$$\mathbf{P}^\varepsilon\left[\sup_{0\le\tau\le T'} |\ell_\varepsilon(\tau) - \ell_0(\tau)| > \delta \right] \le C \exp\left\{ -\frac{\theta^2 \delta^2}{8K^2 T \varepsilon^2} \right\}, \tag{13}$$

holds for any $\delta > 0$. Notice that

$$\ell_\varepsilon(\tau) \le \ell_0(\tau) + |\ell_\varepsilon(\tau) - \ell_0(\tau)|,$$

hence

$$\sup_{0\le\tau\le T'} \ell_\varepsilon(\tau) \le \ell_0^* + \sup_{0\le\tau\le T'} |\ell_\varepsilon(\tau) - \ell_0(\tau)|.$$

If the threshold c satisfies

$$\delta_F \triangleq c - \ell_0^* > 0,$$

then

$$F_\varepsilon = \mathbf{P}^\varepsilon\left[\sup_{0\le\tau\le T'} \ell_\varepsilon(\tau) > c \right] \le \mathbf{P}^\varepsilon\left[\sup_{0\le\tau\le T'} |\ell_\varepsilon(\tau) - \ell_0(\tau)| > \delta_F \right]$$
$$\le C \exp\left\{ -\frac{\theta^2 \delta_F^2}{8K^2 T \varepsilon^2} \right\}.$$

Turning now to the proof of the large deviations estimate (13), the following decomposition holds

$$\ell_\varepsilon(\tau) - \ell_0(\tau) = \varepsilon \int_\tau^T a^*(X_t) d\, V_t^\varepsilon - \frac{1}{2} \int_\tau^T [|a(X_t)|^2 - |a(x_t)|^2] dt + \int_\tau^T [a^*(X_t)\chi(X_t) - a^*(x_t)\chi(x_t)] dt = I'(\tau) + I''(\tau).$$

• **Study of $I'(\tau)$** By definition

$$I'(\tau) = \varepsilon \int_0^T a^*(X_t) d\, V_t^\varepsilon - \varepsilon \int_0^\tau a^*(X_t) d\, V_t^\varepsilon,$$

and therefore

$$\sup_{0 \le \tau \le T'} |I'(\tau)| \le \varepsilon \left| \int_0^T a^*(X_t) d\, V_t^\varepsilon \right| + \varepsilon \sup_{0 \le \tau \le T'} \left| \int_0^\tau a^*(X_t) d\, V_t^\varepsilon \right| \le 2\varepsilon \sup_{0 \le \tau \le T} \left| \int_0^\tau a^*(X_t) d\, V_t^\varepsilon \right|.$$

The uniform boundedness of the assumed change coefficient a, and the exponential bound (20) imply

$$\mathbf{P}^\varepsilon \left[\sup_{0 \le \tau \le T'} |I'(\tau)| > \theta\delta \right] \le 2 \exp\left\{ -\frac{\theta^2 \delta^2}{8K^2 T \varepsilon^2} \right\},$$

for any $\delta > 0$ and any $\theta \ge 0$.

• **Study of $I''(\tau)$** By definition

$$I''(\tau) = -\frac{1}{2} \int_\tau^T [a(X_t) - a(x_t)]^* [a(X_t) + a(x_t)] dt + \int_\tau^T [a(X_t) - a(x_t)]^* \chi(x_t) dt + \int_\tau^T a^*(X_t)[\chi(X_t) - \chi(x_t)] dt.$$

The uniform boundedness and Lipschitz continuity of the assumed change coefficient a, the Lipschitz continuity of the misspecification coefficient χ, and the estimates (11) and (21) imply

$$\sup_{0 \le \tau \le T'} |I''(\tau)| \le KL \int_0^T |X_t - x_t| dt \le KLC\, \varepsilon \sup_{0 \le t \le T} |V_t^\varepsilon|,$$

where the constant $C > 0$ depends only on the Lipschitz constant of the *true* drift coefficient b_0, and on the final time T. Finally, the exponential bound (20) implies

$$\mathbf{P}^{\varepsilon}\left[\sup_{0\le t\le T'} |I''(\tau)| > (1-\theta)\delta\right] \le 2m\exp\left\{-\frac{(1-\theta)^2\delta^2}{2K^2L^2C^2T\varepsilon^2}\right\},$$

for any $\delta > 0$ and any $\theta \le 1$.

Collecting the two above estimates, we obtain

$$\mathbf{P}^{\varepsilon}\left[\sup_{0\le \tau\le T'} |\ell_\varepsilon(\tau)-\ell_0(\tau)| > \delta\right] \le 2\exp\left\{-\frac{\theta^2\delta^2}{8K^2T\varepsilon^2}\right\} + 2m\exp\left\{-\frac{(1-\theta)^2\delta^2}{2K^2L^2C^2T\varepsilon^2}\right\},$$

for any $\delta > 0$ and any $0\le\theta\le 1$. For the particular choice $\theta = (1+LC/2)^{-1}$, the two exponential expressions have the same argument, and

$$\mathbf{P}^{\varepsilon}\left[\sup_{0\le \tau\le T'} |\ell_\varepsilon(\tau)-\ell_0(\tau)| > \delta\right] \le C\exp\left\{-\frac{\theta^2\delta^2}{8K^2T\varepsilon^2}\right\},$$

for any $\delta > 0$, where $C = 2(m+1)$. ■

4. PROBABILITY OF NODETECTION

Using the identity

$$a^*(b_0-b)+a^*a_0-\frac{1}{2}|a|^2 = a^*(b_0-b)+a^*(a_0-a)+\frac{1}{2}|a|^2,$$

the expression (9) for the normalized log-likelihood function reads

$$\begin{aligned}
\ell_\varepsilon(\tau) &= \varepsilon\int_\tau^T a^*(X_t)\,dV_t^{\varepsilon,\tau_0} + \int_\tau^T a^*(X_t)[b_0(X_t)-b(X_t)]dt \\
&\quad + \int_{\tau\vee\tau_0}^T a^*(X_t)a_0(X_t)dt - \frac{1}{2}\int_\tau^T |a(X_t)|^2 dt \\
&= \varepsilon\int_\tau^T a^*(X_t)\,dV_t^{\varepsilon,\tau_0} - \int_\tau^{\tau\vee\tau_0}\left[\frac{1}{2}|a(X_t)|^2 - a^*(X_t)\chi(X_t)\right]dt \\
&\quad + \int_{\tau\vee\tau_0}^T \left[\frac{1}{2}|a(X_t)|^2 + a^*(X_t)\eta(X_t)\right]dt,
\end{aligned}$$

where the misspecification coefficient $\chi \triangleq b_0 - b$ is the difference between the *true* and the *assumed* drift coefficients before the change time, the misspecification coefficient $\eta \triangleq (b_0 + a_0) - (b + a)$ is the difference between the *true* and the *assumed* drift coefficients after the change time, and where $\{V_t^{\varepsilon,\tau_0}, 0 \leq t \leq T\}$ is a m-dimensional Wiener process under the probability measure $\mathbf{P}_{\tau_0}^{\varepsilon}$. Introduce the limiting expression obtained as $\varepsilon \downarrow 0$ under the probability measure $\mathbf{P}_{\tau_0}^{\varepsilon}$

$$\ell(\tau_0, \tau) = -\int_{\tau}^{\tau \vee \tau_0} \left[\frac{1}{2}|a(x_t^{\tau_0})|^2 - a^*(d_t^{\tau_0})\chi(x_t^{\tau_0})\right] dt + \int_{\tau \vee \tau_0}^{T} \left[\frac{1}{2}|a(x_t^{\tau_0})|^2 + a^*(x_t^{\tau_0})\eta(x_t^{\tau_0})\right] dt,$$

where for all $0 \leq \tau_0 \leq T'$, $\{x_t^{\tau_0}, 0 \leq t \leq T\}$ is the solution of the limiting deterministic system

$$\dot{x}_t^{\tau_0} = b_0(x_t^{\tau_0}) + a_0(x_t^{\tau_0})\mathbf{1}_{(t \geq \tau_0)}, \quad x_0^{\tau_0} = z, \ 0 \leq t \leq T. \tag{14}$$

Notice that $x_t = x_t^{\tau_0}$ for all $0 \leq t \leq \tau_0$, *i.e.*, the solutions of the limiting deterministic systems (10) and (14) coincide before the change time τ_0, hence the equivalent expression

$$\ell(\tau_0, \tau) = -\int_{\tau}^{\tau \vee \tau_0} \left[\frac{1}{2}|a(x_t)|^2 - a^*(x_t)\chi(x_t)\right] dt + \int_{\tau \vee \tau_0}^{T} \left[\frac{1}{2}|a(x_t^{\tau_0})|^2 + a^*(x_t^{\tau_0})\eta(x_t^{\tau_0})\right] dt.$$

Notice also that uniformly for all $0 \leq \tau_0 \leq T'$, the misspecification coefficient η is bounded along the limiting trajectory after the change time τ_0, *i.e.*,

$$\sup_{0 \leq \tau_0 \leq T'} \sup_{\tau_0 \leq t \leq T} |\eta(x_t^{\tau_0}| < \infty. \tag{15}$$

We introduce the following *consistency* assumption.

ASSUMPTION ND(τ_0) *For all $0 \leq t \leq \tau_0$ (τ_0 fixed)*

$$\frac{1}{2}|a(x_t)|^2 - a^*(x_t)\chi(x_t) \geq 0,$$

and for all $\tau_0 \leq t \leq T$ (τ_0 fixed)

$$\frac{1}{2}|a(x_t^{\tau_0})|^2 + a^*(x_t^{\tau_0})\eta(x_t^{\tau_0}) \geq 0.$$

Remark 4.1 Using the identities

$$\frac{1}{2}|a|^2 - a^*\chi = \frac{1}{2}|a-\chi|^2 - \frac{1}{2}|\chi|^2 \quad \text{and} \quad \frac{1}{2}|a|^2 + a^*\eta = \frac{1}{2}|a+\eta|^2 - \frac{1}{2}|\eta|^2,$$

an equivalent form of Assumption ND(τ_0) is: For all $0 \le t \le \tau_0$ (τ_0 fixed)

$$|b_0(x_t) - b(x_t)| \le |b_0(x_t) - [b(x_t) + a(x_t)]|,$$

and for all $\tau_0 \le t \le T$ (τ_0 fixed)

$$|[b_0(x_t^{\tau_0}) + a_0(x_t^{\tau_0})] - [b(x_t^{\tau_0}) + a(x_t^{\tau_0})]| \le |[b_0(x_t^{\tau_0}) + a_0(x_t^{\tau_0})] - b(x_t^{\tau_0})|,$$

which means respectively that (i) along the limiting trajectory before the change time τ_0, the true drift b_0 should be closer to the assumed drift before change b, than it is to the assumed drift after change $b+a$, and that (ii) along the limiting trajectory after the change time τ_0, the true drift b_0+a_0 should be closer to the assumed drift after change $b+a$, than it is to the assumed drift before change b. In general, it is difficult to check this assumption since the limiting trajectory $\{x_t^{\tau_0}, 0 \le t \le T\}$ depends also on the unknown coefficients b_0 and a_0. However, a sufficient condition for Assumption ND(τ_0) to hold for all $0 \le \tau_0 \le T'$ is the following stronger assumption, which also includes Assumption FA$'$.

Assumption ND$'$ *For all* $x \in \mathbb{R}^m$

$$\frac{1}{2}|a(x)|^2 - a^*(x)\chi(x) \ge 0,$$

and

$$\frac{1}{2}|a(x)|^2 + a^*(x)\eta(x) \ge 0.$$

Remark 4.2 If there is no misspecification, *i.e.*, if $b = b_0$ and $a = a_0$, then Assumption ND$'$, and a fortiori Assumption ND(τ_0), is always satisfied.

Under Assumption ND(τ_0), it is easily checked that the mapping $\tau \mapsto \ell(\tau_0, \tau)$ achieves its maximum for $\tau = \tau_0$, *i.e.*,

$$\begin{aligned}\ell^*(\tau_0) &\triangleq \sup_{0 \le \tau \le T'} \ell(\tau_0, \tau) = \ell(\tau_0, \tau_0) \\ &= + \int_{\tau_0}^{T} \left[\frac{1}{2}|a(x_t^{\tau_0})|^2 + a^*(x_t^{\tau_0})\eta(x_t^{\tau_0})\right] dt \ge 0, \qquad (16)\end{aligned}$$

and

$$\tau_0 \in M(\tau_0) \triangleq \operatorname*{argmax}_{0 \le \tau \le T'} \ell(\tau_0, \tau).$$

Remark 4.3 In the simple case where $a(x) \equiv \alpha$ and $\eta(x) \equiv 0$ for all $x \in \mathbb{R}^m$, we have $\ell^*(\tau_0) = |\alpha|^2 (T - \tau_0)/2$, and obviously the minimum of this expression w.r.t. $0 \le \tau_0 \le T'$ is achieved for $\tau_0 = T'$. Unfortunately, the situation is not always as simple, *i.e.*, the minimum of $\ell^*(\tau_0)$ w.r.t. $0 \le \tau_0 \le T'$ is sometimes achieved for some $0 < \tau_0^* < T'$, as illustrated by the following example.

Given a constant $B > 0$, consider the following one-dimensional model without misspecification: the drift coefficient is defined by $b_0(x) = b(x) = Bx$ for all $x \in \mathbb{R}$, and the change coefficient is defined by $a_0(x) = a(x) = -Bx$ for all $x \in \mathbb{R}$. In this simple case, the limiting trajectory can be computed explicitely, *i.e.*, $x_t^{\tau_0} = e^{B(t \wedge \tau_0)} z$ for all $0 \le t \le T$. From these preliminary computations, we obtain

$$\ell^*(\tau_0) = \frac{1}{2} \int_{\tau_0}^{T} |a(x_t^{\tau_0})|^2 dt = \frac{1}{2} B^2 |z|^2 (T - \tau_0) e^{2B\tau_0}.$$

Provided B satisfies $1 < 2BT < T/(T - T')$, the unique minimum of $\ell^*(\tau_0)$ w.r.t. to $0 \le \tau_0 \le T'$ is achieved for $0 < \tau_0^* = T - 1/2B < T'$.

Heuristically, as $\varepsilon \downarrow 0$

$$\mathbf{P}_{\tau_0}^{\varepsilon} \left[\sup_{0 \le \tau \le T'} \ell_\varepsilon(\tau) \le c \right] \approx \mathbf{1}_{(\ell^*(\tau_0) \le c)},$$

hence $N_\varepsilon \approx 0$ if $\ell^*(\tau_0) > c$ for all $0 \le \tau_0 \le T'$. This is made rigorous in the following statement, where a nonasymptotic upper bound is given for the exponential rate of convergence of N_ε to zero.

THEOREM 4.4 *Assume that*

- *the drift and change coefficients b_0, b, a_0 and a are Lipschitz continuous,*
- *the assumed change coefficient a is bounded.*

Then, the probability of nodetection N_ε satisfies

$$N_\varepsilon \le C \exp\left\{ -\frac{\theta^2 \delta_N^2}{8K^2 T \varepsilon^2} \right\},$$

provided Assumption $ND(\tau_0)$ *holds for all* $0 \le \tau_0 \le T'$, *and provided the threshold* c *satisfies*

$$\delta_N \triangleq \inf_{0 \le \tau_0 \le T'} \ell^*(\tau_0) - c > 0.$$

The constant K *depends only on the uniform bound of the assumed change coefficient* a *and on the bounds of the misspecification coefficients* χ *and* η *along the limiting trajectories, and the constant* $0 < \theta \le 1$ *depends only on the Lipschitz constants and on the final time* T.

Proof Assume first that the following large deviations estimate

$$\mathbf{P}^{\varepsilon}_{\tau_0}\left[\sup_{0 \le \tau \le T'} |\ell_\varepsilon(\tau) - \ell(\tau_0, \tau)| > \delta\right] \le C \exp\left\{-\frac{\theta^2 \delta^2}{8K^2 T \varepsilon^2}\right\}, \tag{17}$$

holds for any $\delta > 0$. Notice that

$$\ell(\tau_0, \tau) \le \ell_\varepsilon(\tau) + |\ell_\varepsilon(\tau) - \ell(\tau_0, \tau)|,$$

hence

$$\ell^*(\tau_0) \le \sup_{0 \le \tau \le T'} \ell_\varepsilon(\tau) + \sup_{0 \le \tau \le T'} |\ell_\varepsilon(\tau) - \ell(\tau_0, \tau)|.$$

If the threshold c satisfies

$$\delta(\tau_0) \triangleq \ell^*(\tau_0) - c > 0,$$

then

$$\begin{aligned}\mathbf{P}^{\varepsilon}_{\tau_0}\left[\sup_{0 \le \tau \le T'} \ell_\varepsilon(\tau) < c\right] &\le \mathbf{P}^{\varepsilon}_{\tau_0}\left[\sup_{0 \le \tau \le T'} |\ell_\varepsilon(\tau) - \ell(\tau_0, \tau)| > \delta(\tau_0)\right] \\ &\le C \exp\left\{-\frac{\theta^2 \delta^2(\tau_0)}{8K^2 T \varepsilon^2}\right\}.\end{aligned}$$

If in addition the threshold c satisfies

$$\delta_N \triangleq \inf_{0 \le \tau_0 \le T'} \delta(\tau_0) = \inf_{0 \le \tau_0 \le T'} \ell^*(\tau_0) - c > 0$$

then

$$N_\varepsilon = \sup_{0 \le \tau_0 \le T'} \mathbf{P}^{\varepsilon}_{\tau_0}\left[\sup_{0 \le \tau \le T'} \ell_\varepsilon(\tau) < c\right] \le C \exp\left\{-\frac{\theta^2 \delta_N^2}{8K^2 T \varepsilon^2}\right\}.$$

Turning now to the proof of the large deviations estimate (17), the following decomposition holds

$$\ell_\varepsilon(\tau) - \ell(\tau_0, \tau) = \varepsilon \int_\tau^T a^*(X_t) d V_t^{\varepsilon,\tau_0} - \frac{1}{2}\int_\tau^{\tau\vee\tau_0} [|a(X_t)|^2 - |a(x_t)|^2] dt$$
$$+ \int_\tau^{\tau\vee\tau_0} [a^*(X_t)\chi(X_t) - a^*(x_t)\chi(x_t)] dt$$
$$+ \frac{1}{2}\int_{\tau\vee\tau_0}^T [|a(X_t)|^2 - |a(x_t^{\tau_0})|^2] dt$$
$$+ \int_{\tau\vee\tau_0}^T [a^*(X_t)\eta(X_t) - a^*(x_t^{\tau_0})\eta(x_t^{\tau_0})] dt = I'(\tau) + I''(\tau).$$

- **Study of $I'(\tau)$** By definition

$$I'(\tau) = \varepsilon \int_0^T a^*(X_t) d V_t^{\varepsilon,\tau_0} - \varepsilon \int_0^\tau a^*(X_t) d V_t^{\varepsilon,\tau_0},$$

and therefore

$$\sup_{0\le r\le T'} |I'(\tau)| \le \varepsilon \left| \int_0^T a^*(X_t) d V_t^{\varepsilon,\tau_0} \right| + \varepsilon \sup_{0\le r\le T'} \left| \int_0^\tau a^*(X_t) d V_t^{\varepsilon,\tau_0} \right|$$
$$\le 2\varepsilon \sup_{0\le r\le T} \left| \int_0^\tau a^*(X_t) d V_t^{\varepsilon,\tau_0} \right|.$$

The uniform boundedness of the assumed change coefficient a, and the exponential bound (20) imply

$$\mathbf{P}_{\tau_0}^{\varepsilon}\left[\sup_{0\le\tau\le T'} |I'(\tau)| > \theta\delta \right] \le 2\exp\left\{ -\frac{\theta^2\delta^2}{8K^2T\varepsilon^2} \right\},$$

for any $\delta > 0$ and any $\theta \ge 0$.

- **Study of $I''(\tau)$** By definition

$$I''(\tau) = -\frac{1}{2}\int_\tau^{\tau\vee\tau_0} [a(X_t) - a(x_t)]^* [a(X_t) + a(x_t)] dt$$
$$+ \int_\tau^{\tau\vee\tau_0} [a(X_t) - a(x_t)]^* \chi(x_t) dt + \int_\tau^{\tau\vee\tau_0} a^*(X_t)[\chi(X_t) - \chi(x_t)] dt$$
$$+ \frac{1}{2}\int_{\tau\vee\tau_0}^T \left[a(X_t) - a(x_t^{\tau_0})\right]^* \left[a(X_t) + a(x_t^{\tau_0})\right] dt$$
$$+ \int_{\tau\vee\tau_0}^T \left[a(X_t) - a(x_t^{\tau_0})\right]^* \eta(x_t^{\tau_0}) dt + \int_{\tau\vee\tau_0}^T a^*(X_t)\left[\eta(X_t) - \eta(x_t^{\tau_0})\right] dt.$$

The uniform boundedness and Lipschitz continuity of the assumed change coefficient a, the Lipschitz continuity of the two misspecification coefficients χ and η, and the estimates (11), (15) and (21) imply

$$\sup_{0\leq\tau\leq T'} |I''(\tau)| \leq KL \int_0^T |X_t - x_t^{\tau_0}| dt \leq KLC\,\varepsilon \sup_{0\leq t\leq T} |V_t^{\varepsilon,\tau_0}|,$$

where the constant $C > 0$ depends only on the Lipschitz constants of the *true* drift and change coefficients b_0 and a_0, and on the final time T. Finally, the exponential bound (20) implies

$$\mathbf{P}_{\tau_0}^{\varepsilon}\left[\sup_{0\leq t\leq T'} |I''(\tau)| > (1-\theta)\delta\right] \leq 2m\exp\left\{-\frac{(1-\theta)^2\delta^2}{2K^2L^2C^2T\varepsilon^2}\right\},$$

for any $\delta > 0$ and any $\theta \leq 1$.

The end of the proof is exactly the same as the end of the proof of Theorem 3.3 above, and is therefore omitted. ■

5. DETECTABILITY AND THRESHOLD SELECTION

It has been proved in Theorems 3.3 and 4.4 above, that both the probability of false alarm and the probability of nodetection go to zero with exponential rate when $\varepsilon \downarrow 0$, provided Assumption FA holds, provided Assumption ND(τ_0) holds for all $0 \leq \tau_0 \leq T'$, and provided the threshold c satisfies simultaneously

$$\ell_0^* < c < \inf_{0\leq\tau_0\leq T'} \ell^*(\tau_0),$$

where ℓ_0^* and $\ell^*(\tau_0)$ are defined in (12) and (16) respectively. This is possible under the following additional *detectability* assumption

$$\begin{aligned} -\int_{T'}^{T}&\left[\frac{1}{2}|a(x_t)|^2 - a^*(x_t)\chi(x_t)\right]dt \\ &< \inf_{0\leq\tau_0\leq T'} \int_{\tau_0}^{T}\left[\frac{1}{2}\left|a(x_t^{\tau_0})\right|^2 + a(x_t^{\tau_0})\eta(x_t^{\tau_0})\right]dt. \end{aligned} \tag{18}$$

Note that for this assumption to hold, it is necessary that $T' < T$.

Remark 5.1 It follows from the estimates given in Theorems 3.3 and 4.4 above that the larger is the upper bound K on the assumed change coefficient a, the slower is the exponential decay to zero of the two error probabilities. This can look surprising at first sight, since one would expect large changes to be better detected. However, what really matters for the performance of the test, is rather the lower bound on the change coefficient. Indeed, assume that for all $x \in \mathbb{R}^m$

$$0 \leq \frac{1}{2} K_F^2 \leq \frac{1}{2} |a(x)|^2 - a^*(x)\chi(x)$$
$$0 \leq \frac{1}{2} K_N^2 \leq \frac{1}{2} |a(x)|^2 + a^*(x)\eta(x),$$

which is always possible under the stronger Assumption ND$'$. Then

$$-\int_{T'}^{T} \left[\frac{1}{2}|a(x_t)|^2 - a^*(x_t)\chi(x_t)\right] dt \leq -\frac{1}{2} K_F^2 (T - T')$$
$$\leq \frac{1}{2} K_N^2 (T - T') \leq \inf_{0 \leq \tau_0 \leq T'} \int_{\tau_0}^{T} \left[\frac{1}{2}\left|a(x_t^{\tau_0})\right|^2 + a^*(x_t^{\tau_0})\eta(x_t^{\tau_0})\right] dt,$$

which results in the following lower bound

$$\Delta \geq \frac{1}{2}\left(K_F^2 + K_N^2\right)(T - T'),$$

for the *threshold margin* Δ, *i.e.*, for the length of the interval in which the threshold can be selected. Therefore, a *sufficient* condition for the *detectability* assumption (18) to hold is simply $(K_F^2 + K_N^2) > 0$ and $T' < T$. In addition, the particular choice $c = (K_N^2 - K_F^2)/4$ for the threshold results in the following lower bounds

$$\delta_F \geq \frac{1}{4}\left(K_F^2 + K_N^2\right)(T - T') \quad \text{and} \quad \delta_N \geq \frac{1}{4}\left(K_F^2 + K_N^2\right)(T - T')$$

for the constants governing the exponential decay of the error probabilities in the small noise asymptotics.

Considering the change time estimation, it follows from a large deviations estimate similar to (17), that the following *consistency* result holds

$$\mathbf{P}_{\tau_0}^{\varepsilon}[d(\hat{\tau}_\varepsilon, M(\tau_0)) > \delta] \xrightarrow{\varepsilon \downarrow 0} 0,$$

where $\{\hat{\tau}_\varepsilon, \varepsilon > 0\}$ is any MLE sequence for the change time, and

$$M(\tau_0) \stackrel{\Delta}{=} \operatorname*{argmax}_{0 \leq \tau \leq T'} \ell(\tau_0, \tau),$$

is the set-valued deterministic change time estimator. If in addition to Assumption ND(τ_0), the following *identifiability* assumption

$$\begin{aligned} &\frac{1}{2}\left|a(x_{\tau_0})\right|^2 - a^*(x_{\tau_0})\chi(x_{\tau_0}) > 0 \\ &\frac{1}{2}\left|a(x_{\tau_0})\right|^2 + a^*(x_{\tau_0})\eta(x_{\tau_0}) > 0, \end{aligned} \tag{19}$$

holds, then there is no other point than τ_0 in the set $M(\tau_0)$. This consistency result is proved for the one-dimensional case in Kutoyants [6, Theorem 5.6], and the asymptotic probability distribution of the estimator $\hat{\tau}_\varepsilon$ is given in [6, Theorem 5.7]. If there is no misspecification, *i.e.*, if $b = b_0$ and $a = a_0$, then condition (19) reduces to

$$a(x_{\tau_0}) \neq 0,$$

and the corresponding results are proved for the one-dimensional case in Kutoyants [4, Theorem 3.5.2].

In conclusion, it is possible both to detect a change and to estimate the change time, using a *wrong* model, provided the change to be estimated is larger then the misspecification error.

Acknowledgements

This work was partially supported by the Commission of the European Communities, under the Science project *System Identification*, project number SC1*-CT92-0779, and under the HCM project *Statistical Inference for Stochastic Processes*, project number CHRX-CT92-0078, and by the Army Research Office, under grant DAAH04-95-1-0164.

References

[1] Basseville, M. and Nikiforov, I. V., *Detection of Abrupt Changes: Theory and Application*. Information and System Sciences Series. Prentice-Hall, Englewood Cliffs, 1993.
[2] Huber, P. J., *Robust Statistics*. John Wiley & Sons, New York, 1981.
[3] Ibragimov, I. A. and Khasminskii, R. Z., *Statistical Estimation. Asymptotic Theory, Volume 16 of Applications of Mathematics*. Springer-Verlag, New York, 1981.
[4] Kutoyants, Y. A., *Parameter Estimation for Stochastic Processes, Volume 6 of Research and Exposition in Mathematics*. Heldermann Verlag, Berlin, 1984.
[5] Kutoyants, Y. A. (1988). On an identification problem of dynamical systems with a small noise. *Soviet Journal of Contemporary Mathematical Analysis*, **23**(3), 79–95.
[6] Kutoyants, Y. A., *Identification of Dynamical Systems with Small Noise, Volume 300 of Mathematics and its Applications*. Kluwer Academic Publisher, Dordrecht, 1994.
[7] Liptser, R. S. and Shiryayev, A. N., *Statistics of Random Processes I. General Theory, Volume 5 of Applications of Mathematics*. Springer-Verlag, New York, 1977.

[8] McKeague, I. W., Estimation for diffusion processes under misspecified models. *Journal of Applied Probability*, **21**(3), 511–520, Sept., 1984.
[9] Shiryayev, A. N., *Optimal Stopping Rules, Volume 8 of Applications of Mathematics*. Springer-Verlag, New York, 1978.
[10] Stroock, D. W. and Varadhan, S. R. S., *Multidimensional Diffusion Processes, Volume 233 of Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin, 1979.
[11] van Trees, H. L., *Detection, Estimation and Modulation Theory. Part I: Detection, Estimation and Linear Modulation Theory*. John Wiley & Sons, New York, 1968.
[12] Vostrikova, L. Y. (1981). Detection of a "disorder" in a Wiener process. *Theory of Probability and its Applications*, **XXVI**(2), 356–362.
[13] White, H., Maximum likelihood estimation of misspecified models. *Econometrica*, **50**(1), 1–25, Jan., 1982.

A: EXPONENTIAL BOUNDS

Let $\{M_t, 0 \le t \le T\}$ be a m-dimensional continuous martingale on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with $M_0 = 0$, and let $\{\langle M\rangle_t, 0 \le t \le T\}$ denote the corresponding increasing (quadratic variation) process, with values in the space of symmetric nonnegative $m \times m$ matrices. If

$$\langle M\rangle_T \le \alpha I,$$

then

$$\mathbf{P}\left[\sup_{0\le t\le T} |M_t| > \delta\right] \le 2m \exp\left\{-\frac{\delta^2}{2\alpha}\right\}, \tag{20}$$

for all $\delta > 0$, see Stroock and Varadhan [10, Theorem 4.2.1].

This bound can be used to estimate the difference between the solution $\{X_t, 0 \le t \le T\}$ of the stochastic differential equation

$$dX_t = \phi_t(X_t)dt + dB_t, \quad X_0 = z, \;\; 0 \le t \le T,$$

where $\{B_t, 0 \le t \le T\}$ is a m-dimensional Wiener process, and the solution $\{x_t, 0 \le t \le T\}$ of the corresponding ordinary differential equation

$$\dot{x}_t = \phi_t(x_t), \quad x_0 = z, \;\; 0 \le t \le T.$$

If the drift coefficient is Lipschitz continuous, *i.e.*,

$$|\phi_t(x) - \phi_t(x')| \le L|x - x'|,$$

for all $x, x' \in \mathbb{R}^m$, and all $0 \le t \le T$, the Gronwall lemma gives

$$\int_0^T |X_t - x_t| dt \le C_L \sup_{0\le t\le T} |B_t|, \tag{21}$$

and

$$\sup_{0 \le t \le T} |X_t - x_t| \le D_L \sup_{0 \le t \le T} |B_t|,$$

where the constants $C_L = (e^{LT} - 1)/L$ and $D_L = e^{LT}$ depend only on the Lipschitz constant L, and on the final time T. Using the exponential bound (20), the following exponential estimates hold for probabilities of large deviations

$$\mathbf{P}\left[\int_0^T |X_t - x_t| dt > \delta\right] \le 2m \exp\left\{-\frac{\delta^2}{2C_L^2 T}\right\},$$

and

$$\mathbf{P}\left[\sup_{0 \le t \le T} |X_t - x_t| > \delta\right] \le 2m \exp\left\{-\frac{\delta^2}{2D_L^2 T}\right\},$$

for all $\delta > 0$.