Weierstraß–Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint ISSN 0946 – 8633

On estimation of the linearized drift for nonlinear stochastic differential equations

R. Khasminskii[1] G.N. Milstein[2]

submitted: 25th January 2001

[1] Wayne State University
Detroit
USA
E-mail: Rafail@math.wayne.edu

[2] Weierstrass Institute for Applied
Analysis and Stochastics
Mohrenstr. 39
D-10117 Berlin, Germany
E-Mail: milstein@wias-berlin.de

and

Ural State University
Lenin Street 51
620083 Ekaterinburg
Russia

Preprint No. 635
Berlin 2001

2000 *Mathematics Subject Classification.* 60H10, 62F, 93E15.

Key words and phrases. asymptotic efficiency; maximum likelihood estimator; local asymptotic normality; stochastic stability .

Edited by
Weierstraß–Institut für Angewandte Analysis und Stochastik (WIAS)
Mohrenstraße 39
D — 10117 Berlin
Germany

Fax: + 49 30 2044975
E-Mail (X.400): c=de;a=d400-gw;p=WIAS-BERLIN;s=preprint
E-Mail (Internet): preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

ABSTRACT. The estimation of linearized drift for stochastic differential equations with equilibrium points is considered. It is proved that the linearized drift matrix can be estimated efficiently if the initial condition for the system is chosen close enough to the equilibrium point. Some bounds for initial conditions providing the asymptotical efficiency of estimators are found.

1. INTRODUCTION

It is known that solutions of nonlinear stochastic differential equations (SDEs) at the vicinity of equilibrium points can be approximated sufficiently well by solutions of the linearized equations. In particular (see, e.g., [4]), stability of equilibrium points for a nonlinear SDE can often be deduced from the stability of the linear approximation. Note also that the knowledge of parameters of the linearized equation allows to find the stability index (see [1]) and other useful characteristics for nonlinear SDEs. Therefore estimation of parameters for the linearized SDE is interesting for many applications in mechanics, biology, etc.

It is well known (see, for instance, discussion in [3]) that a value of the diffusion matrix at any point x can be evaluated precisely on the basis of observing the solution on an arbitrary small time interval $(t_0, t_0+\delta)$ with $X(t_0)=x$ (here and below we denote by $X(t)$ or X_t the solution of a SDE). So, in the paper we consider the estimation problem for the matrix $f'(0)$ only. Here $f(x)$ is a drift vector for the SDE.

The asymptotically efficient (a.e.) procedures for the drift estimation of linear homogeneous SDEs were proposed in [3], [5]. It was shown there that the estimation performance for linear homogeneous SDEs does not depend on the type of equation and on the choice of initial conditions: the drift coefficients can be estimated by the same procedures with the same rate of convergence of risks for stable, unstable, and neutral equations and with arbitrary nonzero initial conditions.

In general, there is no consistent estimator for $f'(0)$, because a trajectory $X(t)$ with an arbitrary initial condition of even a stable in probability nonlinear SDE may not visit a sufficiently small neighbourhood of the origin with positive probability (without the loss of generality we can identify the equilibrium point with the origin).

The aim of this paper is to propose and justify a.e. procedures for the estimation of $f'(0)$ in nonlinear SDEs. It is clear from the discussion above that for this type of SDEs a.e. estimators, as a rule, do not exist if a statistician can not choose the initial conditions sufficiently close to the origin. The main problems are: (i) to indicate how the initial conditions must be close to the origin to ensure existence of an a.e. estimator; (ii) to construct a.e. estimators.

2. ONE-DIMENSIONAL EQUATION

Let $X_t^x = X^x(t) \in \mathbf{R}^1$ be a Markov process described by the SDE

$$dX_t = f(X_t)dt + b(X_t)dw_t, \quad X^x(0) = x. \tag{2.1}$$

Let $f(0)=b(0)=0$, so that $x=0$ is an equilibrium point for (2.1). Moreover, we assume that

$$f(x)=f'(0)x+O(|x|^{1+\alpha}),\ b(x)=b'(0)x+O(|x|^{1+\alpha}),\ \alpha>0, \tag{2.2}$$

as $x\to 0$.

We consider the estimation problem for $\theta=f'(0)$. It is proved in [3] that for the linear equation

$$d\bar{X}_t=\theta\bar{X}_t dt+\sigma\bar{X}_t dw_t,\ \bar{X}_0=x\neq 0, \tag{2.3}$$

the estimator

$$\hat{\theta}_T=\frac{1}{T}\int_0^T\frac{d\bar{X}_t^x}{\bar{X}_t^x}\ \ (\text{i.e., } \hat{\theta}_T=\frac{1}{T}\ln\frac{\bar{X}_T^x}{x}+\frac{1}{2}\sigma^2) \tag{2.4}$$

is efficient in the sense (here and below $\mathcal{L}(.)$ is the distribution law of (.)):

$$\mathcal{L}\big(\frac{\sqrt{T}}{\sigma}(\hat{\theta}_T-\theta)\big)=\mathcal{N}(0,1)$$

and there is no estimator with uniformly in θ smaller risks. Denote $b'(0)$ by σ and rewrite equation (2.1) in the form

$$dX_t=(\theta X_t+\varphi(X_t))dt+(\sigma X_t+\psi(X_t))dw_t,\ X_0=x\neq 0. \tag{2.5}$$

Consider now some properties of the estimator (2.4) for the process (2.1). We have (along with notation w_t we use $w(t)$)

$$\hat{\theta}_T=\frac{1}{T}\int_0^T\frac{dX_t^x}{X_t^x}=\theta+\frac{1}{T}\int_0^T\frac{\varphi(X_t^x)}{X_t^x}dt+\sigma\frac{w(T)}{T}+\frac{1}{T}\int_0^T\frac{\psi(X_t^x)}{X_t^x}dw(t).$$

So

$$\sqrt{T}(\hat{\theta}_T-\theta)=\sigma\frac{w(T)}{\sqrt{T}}+\frac{1}{\sqrt{T}}\int_0^T\frac{\varphi(X_t^x)}{X_t^x}dt+\frac{1}{\sqrt{T}}\int_0^T\frac{\psi(X_t^x)}{X_t^x}dw(t) \tag{2.6}$$

$$:=\sigma\frac{w(T)}{\sqrt{T}}+\xi_T+\eta_T.$$

Below we use the notation x_0 for the initial condition $X(0)$ instead of x and consider x_0 depending on T: $X(0)=x_0=x_0(T)$.

Theorem 2.1. *Suppose the coefficients of (2.1) satisfy the conditions (2.2) for some $\alpha>0$ and*

$$f'(0)\le B \tag{2.7}$$

for a known constant B. Let $M:=B-\dfrac{\sigma^2}{2}\ge 0$. Then the estimator

$$\hat{\theta}_T=\frac{1}{T}\int_0^T\frac{dX_t^{x_0}}{X_t^{x_0}} \tag{2.8}$$

with $x_0 = x_0(T)$ *satisfying the condition*

$$0 < |x_0| < e^{-(M+\varepsilon)T}, \tag{2.9}$$

$\varepsilon > 0$ *is a constant, has the property*

$$\mathcal{L}(\frac{\sqrt{T}}{\sigma}(\hat{\theta}_T - \theta)) \to \mathcal{N}(0,1) \tag{2.10}$$

as $T \to \infty$.

Remark 2.1. It is known (see [4]), that the condition $\theta - \sigma^2/2 > 0$ provides instability of the origin for the solution of (2.5). So the solution of (2.1) can be unstable in probability under the conditions of Theorem 2.1. This is the reason why the asymptotic efficiency of the estimator (2.8) can be guaranteed under the very strong restriction (2.9) only. We will see below that for asymptotically stable SDEs this restriction can be essentially weakened. □

Proof. It is clear from (2.6) that it is enough to prove that $\xi_T \to 0$ and $\eta_T \to 0$ as $T \to \infty$ in probability for $x = x_0$ satisfying (2.9). Due to (2.2), we have for $x \to 0$

$$\frac{\varphi(x)}{x} = O(|x|^\alpha), \ \frac{\psi(x)}{x} = O(|x|^\alpha). \tag{2.11}$$

So $\xi_T \to 0$ in probability for $T \to \infty$ if

$$\lim_{T\to\infty} P\{\sup_{0\le t\le T} |X_t^{x_0}|^\alpha > \frac{1}{T^{1/2+\delta}}\} = 0, \tag{2.12}$$

for some $\delta > 0$.

Introduce

$$\tau = T \wedge \big(\inf\{t > 0 : |X_t^{x_0}|^\alpha > T^{-(1/2+\delta)}\}\big).$$

Since

$$|X_\tau^{x_0}|^\alpha = \frac{1}{T^{1/2+\delta}}$$

for $\tau \le T$ only, we obtain for any $\lambda > 0$ and $\gamma > 0$

$$\begin{aligned} P\{\sup_{0\le t\le T} |X_t^{x_0}|^\alpha > \frac{1}{T^{1/2+\delta}}\} &= P\{\tau < T\} \\ &= P\{e^{-\lambda\gamma\tau}|X_\tau^{x_0}|^\gamma > e^{-\lambda\gamma T}(\frac{1}{T^{1/2+\delta}})^{\gamma/\alpha}\} \\ &\le \mathbf{E}(e^{-\lambda\gamma\tau}|X_\tau^{x_0}|^\gamma)\cdot e^{\lambda\gamma T}(T^{1/2+\delta})^{\gamma/\alpha}. \end{aligned} \tag{2.13}$$

Consider now the auxiliary function $V(t,x) = e^{-\lambda\gamma t}|x|^\gamma$. Due to the Ito formula, we get

$$\mathbf{E}e^{-\lambda\gamma\tau}|X_\tau^{x_0}|^\gamma - |x_0|^\gamma$$

$$= \mathbf{E}\int_0^\tau e^{-\lambda\gamma t}(-\lambda\gamma|X_t^{x_0}|^\gamma + f(X_t^{x_0})\gamma|X_t^{x_0}|^{\gamma-1}\text{sign}X_t^{x_0} + \frac{1}{2}b^2(X_t^{x_0})\gamma(\gamma-1)|X_t^{x_0}|^{\gamma-2})dt$$

$$= \mathbf{E}\int_0^\tau e^{-\lambda\gamma t}|X_t^{x_0}|^\gamma(-\lambda\gamma + \gamma\frac{f(X_t^{x_0})}{X_t^{x_0}} + \frac{1}{2}\gamma(\gamma-1)\frac{b^2(X_t^{x_0})}{(X_t^{x_0})^2})dt.$$

Due to the fact that $|X_\tau^{x_0}|^\alpha \le 1/T^{1/2+\delta}$ for $0 \le t \le \tau$ and due to condition (2.2), the expression in the brackets is negative for T large enough if $\lambda > 0$ and $\gamma > 0$ are chosen so that

$$f'(0) + \frac{1}{2}(\gamma - 1)\sigma^2 \le B + \frac{1}{2}(\gamma - 1)\sigma^2 \le \lambda + \varepsilon, \tag{2.14}$$

where ε is a positive constant.

Therefore, for T large enough we have

$$\mathbf{E}\{e^{-\lambda\gamma\tau}|X_\tau^{x_0}|^\gamma\} \le |x_0|^\gamma.$$

From this inequality and (2.13) we get

$$P\{\sup_{0\le t\le T}|X_t^{x_0}|^\alpha > \frac{1}{T^{1/2+\delta}}\} \le e^{\lambda\gamma T}|x_0|^\gamma T^{(1/2+\delta)\gamma/\alpha}.$$

So (2.12) holds if x_0 satisfies the inequality

$$0 < |x_0| < e^{-(\lambda+\varepsilon)T}.$$

Set $\lambda = M,\ 0 < \gamma \le 2\varepsilon/\sigma^2$. Then (2.14) is valid and $\xi_T \to 0$ in probability under (2.9).

Thus we have

$$\eta_T = \frac{1}{\sqrt{T}}\int_0^\tau \frac{\psi(X_t^{x_0})}{X_t^{x_0}}dw(t) + \frac{1}{\sqrt{T}}\int_\tau^T \frac{\psi(X_t^{x_0})}{X_t^{x_0}}dw(t) := \eta_T^{(1)} + \eta_T^{(2)}.$$

It is clear that $P\{\eta_T^{(2)} \ne 0\} \to 0$ due to (2.12). Using (2.11), we get

$$\mathbf{E}(\eta_T^{(1)})^2 \le \frac{1}{T}\int_0^T \frac{dt}{T^{1+2\delta}} \to 0,\ T \to \infty. \ \square$$

Now we consider the systems with stable in probability equilibrium points assuming that $M < 0$.

Theorem 2.2. *Let the conditions* (2.2) *and* (2.7) *be valid. Let* $M = B - \dfrac{\sigma^2}{2} < 0$. *Then the estimator* (2.8) *has the property* (2.10) *for any* x_0 *satisfying the condition*

$$0 < |x_0| < T^{-(1/2+\beta)/\alpha}, \tag{2.15}$$

where β *is an arbitrary positive constant.*

Proof. Denote by $L = f(x)\dfrac{d}{dx} + \dfrac{\sigma^2(x)}{2}\dfrac{d^2}{dx^2}$ the generator of the process (2.1). Since $M < 0$, the function $|x|^\gamma$ with γ satisfying the bounds

$$0 < \gamma < \frac{\sigma^2 - 2B}{\sigma^2}$$

has the property

$$L(|x|^\gamma) < 0 \tag{2.16}$$

in a sufficiently small neighbourhood of the origin, so $|X^x(t)|^\gamma$ is a local supermartingale if x is small enough.

As in the proof of Theorem 2.1, it is sufficient to check that

$$\lim_{T\to\infty} P\{\sup_{0\le t\le T} |X_t^{x_0}|^\alpha > \frac{1}{T^{1/2+\delta}}\} = \lim_{T\to\infty} P\{\tau < T\} = 0 \tag{2.17}$$

for x_0 satisfying (2.15).

Making use of the supermartingale property of $|X^x(t)|^\gamma$, we obtain

$$|x_0|^\gamma \ge E|X_\tau^{x_0}|^\gamma \ge T^{-(1/2+\delta)\gamma/\alpha} P\{\tau < T\},$$

i.e. (see (2.15))

$$P\{\tau < T\} \le T^{(1/2+\delta)\gamma/\alpha} |x_0|^\gamma \le T^{(\delta-\beta)\gamma/\alpha}.$$

This bound implies (2.17) for $\beta > \delta$. As $\delta > 0$ can be arbitrary small, Theorem 2.2 is proved. □

Remark 2.2. The estimators from Theorems 2.1 and 2.2 are asymptotically efficient for the bounded loss functions in the following sense. For any loss function $l(x)$ with the properties

(i) $l(-x) = l(x)$ and $l(0) = 0$;

(ii) $l(x_2) \ge l(x_1)$ for $x_2 > x_1 > 0$;

(iii) $l(x) < K < \infty$,

we have

$$\lim_{T\to\infty} \mathbf{E} l(\frac{\sqrt{T}}{\sigma}(\hat\theta_T - \theta)) = \mathbf{E} l(\xi),\ \mathcal{L}(\xi) = \mathcal{N}(0,1),$$

and there is no estimator with uniformly in θ less risk. The last assertion follows from the fact that even for linear systems there is no uniformly better estimator (see [3], [5]).

The event $\{X(t) \nrightarrow 0 \text{ for } t \to \infty\}$ can have positive probability for a nonlinear (even stable) SDE with any initial condition $X(0) \ne 0$. Due to this fact, it is impossible to propose any estimator which is asymptotically efficient for unbounded loss functions. □

Remark 2.3. Comparing Theorems 2.1 and 2.2, we see that the choices of initial conditions for a.e. and even consistent estimation of $f'(0)$ are essentially different for stable and unstable SDEs. For unstable SDEs, we have to choose the initial condition exponentially close to the origin. Clearly such a choice is also necessary for multidimensional SDEs. This fact implies the essential difficulties for applications. So below (see Section 4) we restrict ourselves to consideration of asymptotically stable multidimensional SDEs. □

3. On drift estimation for linear equations

Consider the system of linear SDEs

$$d\bar X_t = A(\theta)\bar X_t dt + \sum_{r=1}^{q} \sigma_r \bar X_t dw_r(t). \tag{3.1}$$

Here $\bar{X}_t \in \mathbf{R}^d$, $w_r(t)$, $r = 1, ..., q$, are independent standard scalar Wiener processes, $\sigma_1, ..., \sigma_q$ are real $d \times d$ matrices,

$$A(\theta) = A_0 + \sum_{j=1}^{k} A_j \theta_j, \tag{3.2}$$

where the $d \times d$ matrices A_0, A_j are known and the scalars θ_j, $j = 1, ..., k$, are unknown parameters. The estimation of parameters for a linear SDE of the form (3.1) was considered in [3], [5]. Recall some facts from [3], [5] which we shall use below. We suppose that the conditions (C_1) - (C_4) are fulfilled:

(C_1) *The matrices $A_1, ..., A_k$ are linearly independent.*

(C_2) $A_j \in \text{span}(\sigma_1, ..., \sigma_q)$, $j = 1, ..., k$.

(C_3) *The weak Hörmander condition for the Markov diffusion process $\bar{\Lambda}(t) = \bar{X}(t)/|\bar{X}(t)|$ with values on the unit sphere $\mathbf{S}^{d-1} \subset \mathbf{R}^d$ is fulfilled:*

$$\dim LA\{h_0, h_1, ..., h_q\} = d - 1 \text{ for all } \lambda = x/|x| \in \mathbf{S}^{d-1},$$

where

$$h_0(\lambda) = \tilde{A}\lambda - (\tilde{A}\lambda, \lambda)\lambda, \ \tilde{A} = A - \frac{1}{2}\sum_{r=1}^{q} \sigma_r^2,$$

$$h_r(\lambda) = \sigma_r \lambda - (\sigma_r \lambda, \lambda)\lambda, \ r = 1, ..., q,$$

$LA\{\}$ denotes the Lie algebra generated by the vector fields which occur in the brackets (see [2]).

Introduce the diffusion matrix

$$B(x) = \sum_{r=1}^{q} \sigma_r x x^* \sigma_r^*$$

and denote by $B^+(x)$ the pseudoinverse of B.

(C_4) *The matrix $B^+(\lambda)$, $|\lambda| = 1$, is continuous on $\mathbf{S}^{d-1}$.*

Let us also consider the condition

(C_5) *The diffusion matrix $B(x)$ is non-singular for $x \neq 0$.*

Both the conditions (C_3) and (C_4) follow from (C_5). In the non-singular case the matrix $B(x)$ is invertible, i.e., $B^+(x) = B^{-1}(x)$, and

$$(B^{-1}(\lambda)z, z) \leq K|z|^2, \ \lambda \in \mathbf{S}^{d-1}, \ z \in \mathbf{R}^d,$$

where K is a positive constant.

The non-singularity condition (C_5) has been assumed in [3]. It sometimes is too restrictive (see the corresponding discussion in [5]). The authors of [5] eliminate condition (C_5) and consider the estimation problem for linear systems under conditions (C_1)-(C_3) only. For nonlinear systems, we need the condition (C_4) which is essentially less restrictive than (C_5).

The conditions (C_2), (C_4) imply that the measures $P_\theta^{(T)}$ and $P_0^{(T)}$ are mutually absolutely continuous ([6], Section 7.6). Here $P_\theta^{(T)}$ is the probability measure corresponding to the process X_t generated by the system (3.1) with parameter θ. The measure is defined on the space $\mathbf{C}([0,T],\mathbf{R}^d)$ of continuous functions of $[0,T]$ into $\mathbf{R}^d$.

The log-likelihood ratio has the form (we denote by $\bar{X}^T$ the trajectory of the observation process $\bar{X}_t$ for $0 \le t \le T$)

$$\ln \frac{dP_\theta^{(T)}}{dP_0^{(T)}}(\bar{X}^T) = \int_0^T (B^+(\bar{X}_t)\sum_{j=1}^k A_j\theta_j\bar{X}_t, d\bar{X}_t) \tag{3.3}$$

$$-\frac{1}{2}\int_0^T (B^+(\bar{X}_t)\sum_{j=1}^k A_j\theta_j\bar{X}_t, \sum_{j=1}^k A_j\theta_j\bar{X}_t + 2A_0\bar{X}_t)dt.$$

It can be seen from (3.3) that the likelihood ratio depends on the process $\bar{\Lambda}(t) = \bar{X}(t)/|\bar{X}(t)|$ only. It is known (see [2]) that (C_3) implies the existence of a unique invariant distribution for the process $\bar{\Lambda}$ on $\mathbf{S}^{d-1}$ having smooth density $\mu_\theta(\lambda) > 0$ with respect to the surface measure $S(\cdot)$ on $\mathbf{S}^{d-1}$.

It is not difficult to check that the maximum likelihood estimator (MLE) $\hat{\theta}$ is defined by the following system of linear algebraic equations

$$\frac{1}{2}\sum_{j=1}^k \int_0^T ((A_i^* B^+(\bar{X}_t)A_j + A_j^* B^+(\bar{X}_t)A_i)\bar{X}_t, \bar{X}_t)dt \cdot \theta_j \tag{3.4}$$

$$= \int_0^T (B^+(\bar{X}_t)A_i\bar{X}_t, d\bar{X}_t - A_0\bar{X}_t dt),\ i = 1, ..., k.$$

Denote by $H(\bar{X}^T)$ the $k \times k$ matrix of the system (3.4) divided by T. The elements $H_{ij}(\bar{X}^T)$ of this self-adjoint matrix are equal to

$$H_{ij}(\bar{X}^T) = \frac{1}{2T}\int_0^T ((A_i^* B^+(\bar{X}_t)A_j + A_j^* B^+(\bar{X}_t)A_i)\bar{X}_t, \bar{X}_t)dt.$$

Clearly, $H_{ij}(\bar{X}^T) = H_{ij}(\bar{\Lambda}^T)$. Denote also the right-hand side of (3.4) by $V(\bar{X}^T)$:

$$V_i(\bar{X}^T) = \int_0^T (B^+(\bar{X}_t)A_i\bar{X}_t, d\bar{X}_t - A_0\bar{X}_t dt) = V_i(\bar{\Lambda}^T).$$

The matrix $H(\bar{X}^T)$ and the vector $V(\bar{X}^T)$ were introduced in [5] (they have another form there). It is known, that under (C_1) - (C_4) the matrix $H(\bar{X}^T)$ is positively definite a.s.. Therefore

$$\hat{\theta} = \frac{1}{T}H^{-1}(\bar{X}^T)V(\bar{X}^T) = \frac{1}{T}H^{-1}(\bar{\Lambda}^T)V(\bar{\Lambda}^T). \tag{3.5}$$

Due to (3.5) and (3.1), we have

$$\sum_{j=1}^{k} H_{ij}(\bar{X}^T)\hat{\theta}_j = \frac{1}{T}V_i(\bar{X}^T) = \frac{1}{T}\int_0^T (B^+(\bar{X}_t)A_i\bar{X}_t, \sum_{j=1}^{k} A_j\bar{X}_t\theta_j)dt$$

$$+\frac{1}{T}\int_0^T (B^+(\bar{X}_t)A_i\bar{X}_t, \sum_{r=1}^{q} \sigma_r\bar{X}_t dw_r(t))$$

$$= \sum_{j=1}^{k} H_{ij}(\bar{X}^T)\theta_j + \frac{1}{T}\int_0^T (B^+(\bar{X}_t)A_i\bar{X}_t, \sum_{r=1}^{q} \sigma_r\bar{X}_t dw_r(t)).$$

So we have

$$H(\bar{X}^T)\sqrt{T}(\hat{\theta}-\theta) = \bar{\zeta}(T) = (\bar{\zeta}_1(T), ..., \bar{\zeta}_k(T))^\top,$$

where

$$\bar{\zeta}_i(T) = \frac{1}{\sqrt{T}}\int_0^T (B^+(\bar{X}_t)A_i\bar{X}_t, \sum_{r=1}^{q} \sigma_r\bar{X}_t dw_r(t)).$$

We get

$$\mathbf{E}\bar{\zeta}_i(T)\bar{\zeta}_j(T) = \frac{1}{T}\int_0^T E((B^+(\bar{X}_t)A_i\bar{X}_t)^* \sum_{r=1}^{q} \sigma_r\bar{X}_t(\sigma_r\bar{X}_t)^* B^+(\bar{X}_t)A_j\bar{X}_t)dt$$

$$= \frac{1}{T}\int_0^T \mathbf{E}(\bar{X}_t^* A_i^* B^+ B B^+ A_j\bar{X}_t)dt = \frac{1}{T}\int_0^T \mathbf{E}(\bar{X}_t^* A_i^* B^+(\bar{X}_t)A_j\bar{X}_t)dt = \mathbf{E}H_{ij}(\bar{X}^T).$$

It is known from [5] that under (C_1) - (C_4) there exists (a.s.) the limit

$$\lim_{T\to\infty} H_{ij}(\bar{X}^T) = \lim_{T\to\infty} H_{ij}(\bar{\Lambda}^T) = \lim_{T\to\infty} \mathbf{E}H_{ij}(\bar{X}^T) \tag{3.6}$$

$$= \frac{1}{2}\int_{\mathbf{S}^{d-1}} ((A_i^* B^+(\lambda)A_j + A_j^* B^+(\lambda)A_i)\lambda, \lambda)\mu_\theta(\lambda)S(d\lambda) := I_{ij}(\theta)$$

with $k \times k$ matrix $I(\theta) = \{I_{ij}(\theta)\}$ being deterministic and positively definite. The matrix $TI(\theta)$ is the Fisher information matrix.

The estimator (3.5) is asymptotically normal and asymptotically efficient for a wide class of the loss functions (see details in [5]).

4. Stable nonlinear systems

Consider the nonlinear SDE

$$dX_t = f(X_t)dt + \sum_{r=1}^{q} b_r(X_t)dw_r(t), \tag{4.1}$$

where $X_t \in \mathbf{R}^d$, f and b_r, $r = 1, ..., q$, are d-dimensional vectors with $f(0) = b_r(0) = 0$, and $w_r(t)$ are standard Wiener processes.

Assume that the coefficients of (4.1) are sufficiently smooth at 0 so that for $x \to 0$

$$f(x) = f'(0)x + \varphi(x), \ \frac{|\varphi(x)|}{|x|} = O(|x|^\alpha), \ \alpha > 0, \tag{4.2}$$

$$b_r(x) = \sigma_r x + \psi_r(x), \ \frac{|\psi_r(x)|}{|x|} = O(|x|^\alpha), \ r = 1, ..., q, \tag{4.3}$$

where $f'(0)$ and $\sigma_r := b_r'(0)$ are $d \times d$ matrices. We suppose that all $b_r(x)$ are known vector-functions (see comments in [3]). The problem is to estimate $f'(0)$. However some coefficients of this matrix can be known (see, e.g. [5] and Section 5). We assume that $f'(0)$ is a linear function of the finite-dimensional parameter $\theta = (\theta_1, ..., \theta_k)$

$$f'(0) = A(\theta) = A_0 + \sum_{j=1}^{k} A_j \theta_j. \tag{4.4}$$

Rewrite the equation (4.1) in the form

$$dX_t = A(\theta)X_t dt + \sum_{r=1}^{q} \sigma_r X_t dw_r(t) + \varphi(X_t)dt + \sum_{r=1}^{q} \psi_r(X_t)dw_r(t). \tag{4.5}$$

and consider the first order variation of (4.1)

$$d\bar{X}_t = A(\theta)\bar{X}_t dt + \sum_{r=1}^{q} \sigma_r \bar{X}_t dw_r(t). \tag{4.6}$$

Throughout this section we suppose that the conditions (C$_1$) - (C$_4$) are fulfilled.

Lemma 4.1. *Let the trivial solution of* (4.6) *be stable in probability and*

$$0 < |x_0(T)| \le T^{-(1/2+\beta)/\alpha}, \tag{4.7}$$

where β is an arbitrary positive constant. Then

$$\lim_{T\to\infty} P\{\sup_{0\le t\le T} |X_t^{x_0(T)}|^\alpha > \frac{1}{T^{1/2+\delta}}\} = 0. \tag{4.8}$$

Proof. Due to the condition of stability, there exists a positively definite homogeneous of some order $\gamma > 0$ function $v(x)$ such that $Lv < 0$ in a sufficiently small neighbourhood of the origin. So $v(X_t^x)$ is a local supermartingale if x is small enough. Since $v(x) = v(\lambda)|x|^\gamma$, $0 < k_1 \le v(\lambda) \le k_2$, for some positive constants k_1 and k_2, we have

$$k_1 E|X_\tau^{x_0}|^\gamma \le \mathbf{E}v(X_\tau^{x_0}) \le v(x_0) \le k_2|x_0|^\gamma.$$

Now the statement of the lemma can be easily proved by arguments similar to the ones which have been used in the proof of Theorem 2.2. □

Recall that $\bar{\Lambda}(t) = \bar{X}(t)/|\bar{X}(t)|$. If $\bar{X}(0) = x_0$, we use the notation $\bar{\Lambda}_t^{(x_0)} = \bar{X}_t^{x_0}/|\bar{X}_t^{x_0}|$. Clearly the initial value for $\bar{\Lambda}_t^{(x_0)}$ is equal to $\bar{\Lambda}^{(x_0)}(0) = x_0/|x_0| := \lambda_0$, i.e., $\bar{\Lambda}_t^{(x_0)} = \bar{\Lambda}_t^{\lambda_0}$. Introduce $\Lambda_t^{(x_0)} = X_t^{x_0}/|X_t^{x_0}|$, $\Delta_t^{(x_0)} = \Lambda_t^{(x_0)} - \bar{\Lambda}_t^{(x_0)}$. Clearly $\Lambda^{(x_0)}(0) = \lambda_0$, $\Delta^{(x_0)}(0) = 0$.

Now our goal is to prove that for a suitable choice of $x_0 = x_0(T) \to 0$ and $\delta(T) \to 0$ as $T \to \infty$

$$\lim_{T\to\infty} P\{\sup_{0\le t\le T} |\Lambda_t^{(x_0)} - \bar{\Lambda}_t^{(x_0)}| \le \delta(T)\} = 1. \tag{4.9}$$

We show in Lemmas 4.2-4.4 that in the case of stable systems the equation (4.9) holds for a choice of $x_0(T) \to 0$ exponentially with an arbitrary small exponent as $T \to \infty$.

We have

$$|\Lambda - \bar{\Lambda}| = \left| \frac{X}{|X|} \cdot \frac{|\bar{X}| - |X|}{|\bar{X}|} + \frac{Z}{|\bar{X}|} \right| \le \frac{2|Z|}{|\bar{X}|},$$

where $Z = Z_t^{(x_0)} := X_t^{x_0} - \bar{X}_t^{x_0}$.

So (4.9) follows from

$$\lim_{T\to\infty} P\{\sup_{0\le t\le T} \frac{|Z_t^{(x_0)}|}{|\bar{X}_t^{x_0}|} \le \frac{1}{2}\delta(T)\} = 1.$$

Lemma 4.2. *Let the Lyapunov exponent for the first order variation equation* (4.6) *be equal to* ρ. *Then for any* $\varepsilon > 0$, $\mu > 1$ *there exist* $\beta > 0$ *and* $K > 0$ *such that*

$$P\{\sup_{0\le t\le T} e^{(\rho-\varepsilon)t} |\bar{X}_t^{x_0(T)}|^{-1} > |x_0(T)|^{-\mu}\} \le K|x_0(T)|^{\beta} \tag{4.10}$$

for any $x_0(T)$ *with* $|x_0(T)| < 1$.

Proof. Introduce the variable $\tilde{X} = e^{-(\rho-\varepsilon)t}\bar{X}$. Then

$$d\tilde{X}_t = (-\rho + \varepsilon)\tilde{X}_t dt + A(\theta)\tilde{X}_t dt + \sum_{r=1}^{q} \sigma_r \tilde{X}_t dw_r(t). \tag{4.11}$$

The Lyapunov exponent for the system is equal to $\varepsilon > 0$. Therefore (see [4]) this system is exponentially q-unstable for all sufficiently small $q > 0$ and there exists a positively definite homogeneous of order $-q$ function $v(x)$ such that

$$k_1|x|^{-q} \le v(x) \le k_2|x|^{-q}, \ \tilde{L}v(x) \le -k_3|x|^{-q}, \tag{4.12}$$

where k_1, k_2, k_3 are positive constants and $\tilde{L}$ is the generator of system (4.11).

Let $|x_0(T)| < 1$, $\mu > 1$,

$$\tau = T \wedge \inf\{0 \le t \le T : |\tilde{X}_t^{x_0(T)}| < |x_0(T)|^{\mu}\}.$$

We have

$$\begin{aligned} P\{\sup_{0\le t\le T} |\tilde{X}_t^{x_0(T)}|^{-1} > |x_0(T)|^{-\mu}\} &= P\{\tau < T\} \\ &= P\{(|\tilde{X}_\tau^{x_0(T)}|^{-q} > |x_0(T)|^{-\mu q})\} \\ &\le (\mathbf{E}|\tilde{X}_\tau^{x_0(T)}|^{-q}) \cdot |x_0(T)|^{\mu q}. \end{aligned} \tag{4.13}$$

Due to (4.12), we get

$$\mathbf{E}|\tilde{X}_\tau^{x_0(T)}|^{-q} \le \frac{1}{k_1}\mathbf{E}v(\tilde{X}_\tau^{x_0(T)}) \le \frac{1}{k_1}\mathbf{E}v(x_0(T)) \le \frac{k_2}{k_1}|x_0(T)|^{-q}.$$

This inequality, (4.13), and the change of variables imply inequality (4.10) with

$$\beta = (\mu - 1)q.$$

□

Lemma 4.3. *Let the Lyapunov exponent for the equation (4.6) be equal to* $\rho < 0$. *Then for any* $0 < \varepsilon < |\rho|$ *there exist* $K > 0,\ \nu > 1,\ 0 < q < 1,$ *such that*

$$P\{\sup_{0\le t\le T} e^{-(\rho+\varepsilon)t}|Z_t^{(x_0(T))}| > |x_0(T)|^\nu\} \le K|x_0(T)|^q$$

for any $x_0(T) \to 0$ *as* $T \to \infty$.

Proof. Obviously, $Z_t^{(x_0)}$ is the solution of the problem

$$dZ_t = A(\theta)Z_t dt + \sum_{r=1}^q \sigma_r Z_t dw_r(t) + \varphi(X_t)dt + \sum_{r=1}^q \psi_r(X_t)dw_r(t),\ Z_0 = 0. \tag{4.14}$$

Let us consider the $2d$-dimensional system of equations (4.5), (4.14). We denote by $X_t^{x_0},\ Z_t^{x_0,z_0}$ the solution of this system with the initial conditions $X(0) = x_0,\ Z(0) = z_0$, so that $Z_t^{(x_0)} = Z_t^{x_0,0}$. Let us change the variables:

$$\tilde{X} = e^{-(\rho+\varepsilon)t}X,\ \tilde{Z} = \frac{e^{-(\rho+\varepsilon)t}}{|x_0(T)|^{\alpha_0}}Z,\ 0 < \varepsilon < -\rho,\ 0 < 2\alpha_0 < \alpha. \tag{4.15}$$

Then we have

$$d\tilde{X}_t = -(\rho+\varepsilon)\tilde{X}_t dt + A(\theta)\tilde{X}_t dt + \sum_{r=1}^q \sigma_r \tilde{X}_t dw_r(t) + \tilde{\varphi}(t,\tilde{X}_t)dt + \sum_{r=1}^q \tilde{\psi}_r(t,\tilde{X}_t)dw_r(t), \tag{4.16}$$

$$d\tilde{Z}_t = -(\rho+\varepsilon)\tilde{Z}_t dt + A(\theta)\tilde{Z}_t dt + \sum_{r=1}^q \sigma_r \tilde{Z}_t dw_r(t) \tag{4.17}$$

$$+\frac{1}{|x_0(T)|^{\alpha_0}}\tilde{\varphi}(t,\tilde{X}_t)dt + \sum_{r=1}^q \frac{1}{|x_0(T)|^{\alpha_0}}\tilde{\psi}_r(t,\tilde{X}_t)dw_r(t),$$

$$\tilde{X}(0) = X(0) = x_0,\ \tilde{Z}(0) = Z(0) = 0. \tag{4.18}$$

It follows from (4.2)-(4.3) that for x small enough (recall that $\rho + \varepsilon < 0$)

$$|\tilde{\varphi}(t,x)| = e^{-(\rho+\varepsilon)t}|\varphi(e^{(\rho+\varepsilon)t}x)| \le C|x|^{1+\alpha},$$

$$|\tilde{\psi}_r(t,x)| = e^{-(\rho+\varepsilon)t}|\psi_r(e^{(\rho+\varepsilon)t}x)| \le C|x|^{1+\alpha},\ r = 1,...,q,$$

where C is a constant.

For

$$|x| \leq |x_0(T)|^{1/2}$$

we have

$$\frac{1}{|x_0(T)|^{\alpha_0}}|\tilde{\varphi}(t,x)| \leq C|x|^{1+\alpha-2\alpha_0}, \ \frac{1}{|x_0(T)|^{\alpha_0}}|\tilde{\psi}_r(t,x)| \leq C|x|^{1+\alpha-2\alpha_0}. \tag{4.19}$$

For $\nu_0 < 1$, let us introduce

$$\begin{aligned} \tau_1 &= T \wedge \inf\{0 \leq t \leq T : |\tilde{X}_t^{x_0}| > |x_0(T)|^{1/2}\}, \\ \tau_2 &= T \wedge \inf\{0 \leq t \leq T : |\tilde{Z}_t^{x_0,0}| > |x_0(T)|^{\nu_0}\}, \\ \tau &= \tau_1 \wedge \tau_2. \end{aligned} \tag{4.20}$$

We get for any $\gamma > 0$

$$\begin{aligned} &P\{\sup_{0\leq t\leq T} |\tilde{Z}_t^{x_0,0}| > |x_0(T)|^{\nu_0}\} = P\{\tau_2 < T\} \leq P\{\tau < T\} \\ &= P\{(\tau_1 < T) \cup (\tau_2 < T)\} = P\{(\tau_1 < T)\} + P\{(\tau_2 < T)\backslash(\tau_1 < T)\} \\ &\leq P\{\tau_1 < T\} + P\{|\tilde{Z}_\tau^{x_0,0}|^\gamma > |x_0(T)|^{\gamma\nu_0}\} \\ &\leq P\{\tau_1 < T\} + \mathbf{E}|\tilde{Z}_\tau^{x_0,0}|^\gamma \cdot \frac{1}{|x_0(T)|^{\gamma\nu_0}}. \end{aligned} \tag{4.21}$$

The equations (4.16) and (4.17) have the same linear parts and their Lyapunov exponents are equal to $-\varepsilon < 0$. Therefore, for any $\gamma < \varepsilon$ there exists a positively definite homogenous of order $\gamma > 0$ function $V(x,z)$ such that $LV(x,z) < 0$. Here L denotes the generator corresponding to the linear part of system (4.16)-(4.17). It is known [4] that there exist positive constants $k_1,\ k_2,\ k_3,\ k_4$ such that

$$\begin{gathered} k_1|y|^\gamma \leq V(x,z) \leq k_2|y|^\gamma, \ LV(x,z) \leq -k_3|y|^\gamma, \\ \left|\frac{\partial V}{\partial y_i}\right| \leq k_4|y|^{\gamma-1}, \ \left|\frac{\partial^2 V}{\partial y_i \partial y_j}\right| \leq k_4|y|^{\gamma-2}, \ i,j = 1,...,2d, \end{gathered} \tag{4.22}$$

where y is the $2d$-dimensional vector consisting of the components of the vectors x and z.

Let $|x_0(T)| \to 0$ as $T \to \infty$. Due to (4.19) and (4.22), we obtain

$$\tilde{L}_t V(x,z) < 0 \tag{4.23}$$

for a sufficiently large T if (x,z) is such that

$$|x| \leq |x_0(T)|^{1/2}, \ |z| \leq |x_0(T)|^{\nu_0}. \tag{4.24}$$

In (4.23) $\tilde{L}_t$ is the generator of the diffusion process defined by the system (4.16)-(4.17).

Because of (4.23)-(4.24) and the definition of τ,

$$\tilde{L}_t V(\tilde{X}_t^{x_0}, \tilde{Z}_t^{x_0,0}) < 0, \ 0 \leq t \leq \tau,$$

and, consequently, $V(\tilde{X}_t, \tilde{Z}_t)$ is a local supermartingale.

Therefore we get

$$\mathbf{E}|\tilde{Z}_\tau^{x_0,0}|^\gamma \leq \mathbf{E}(|\tilde{X}_\tau^{x_0}|^2 + |\tilde{Z}_\tau^{x_0,0}|^2)^{\gamma/2} \leq \frac{1}{k_1}\mathbf{E}V(\tilde{X}_\tau^{x_0}, \tilde{Z}_\tau^{x_0,0}) \leq \frac{1}{k_1}V(\tilde{X}(0), \tilde{Z}(0)) \tag{4.25}$$

$$= \frac{1}{k_1} V(x_0(T), 0) \leq \frac{k_2}{k_1} |x_0(T)|^{\gamma}.$$

Hence

$$\mathbf{E}|\tilde{Z}_{\tau}^{x_0,0}|^{\gamma} \cdot \frac{1}{|x_0(T)|^{\gamma\nu_0}} \leq \frac{k_2}{k_1} |x_0(T)|^{\gamma(1-\nu_0)}.$$

Besides, from (4.20) and (4.25) we have

$$P\{\tau_1 < T\} \leq \mathbf{E}|\tilde{X}_{\tau}^{x_0}|^{\gamma} \cdot \frac{1}{|x_0(T)|^{\gamma/2}} \leq \frac{k_2}{k_1} |x_0(T)|^{\gamma/2}.$$

Now from (4.15) and (4.21) we obtain

$$P\{\sup_{0 \leq t \leq T} e^{-(\rho+\varepsilon)t} |Z_t^{(x_0)}| > |x_0(T)|^{\nu_0+\alpha_0}\} \leq \frac{k_2}{k_1} |x_0(T)|^{\gamma/2} + \frac{k_2}{k_1} |x_0(T)|^{\gamma(1-\nu_0)}. \tag{4.26}$$

As α_0 and ν_0 can be chosen so that $\nu := \alpha_0 + \nu_0 > 1$, the assertion of the lemma follows from (4.26). □

The following lemma follows from two previous ones.

Lemma 4.4. *Let the SDE (4.6) be stable and let δ be an arbitrary small positive number. Then for x_0 with $|x_0| = |x_0(T)| \leq \exp(-\delta T)$*

$$P\{\lim_{T\to\infty} \sup_{0 \leq t \leq T} |\Lambda_t^{(x_0)} - \bar{\Lambda}_t^{(x_0)}| \to 0\} = P\{\lim_{T\to\infty} \sup_{0 \leq t \leq T} \frac{|Z_t^{(x_0)}|}{|\bar{X}_t^{x_0}|} \to 0\} = 1. \tag{4.27}$$

Proof. Let us choose $\nu > \mu > 0$. Introduce the events

$$\mathcal{A} = \left\{\omega : \sup_{0 \leq t \leq T} e^{(\rho-\varepsilon)t} |\bar{X}_t^{x_0}|^{-1} \leq |x_0(T)|^{-\mu}\right\}, \ \mathcal{B} = \left\{\omega : \sup_{0 \leq t \leq T} e^{-(\rho+\varepsilon)t} |Z_t^{(x_0)}| \leq |x_0(T)|^{\nu}\right\}.$$

Due to Lemmas 4.2 and 4.3, we get

$$P(\mathcal{A}) \geq 1 - K|x_0(T)|^{\beta}, \ P(\mathcal{B}) \geq 1 - K|x_0(T)|^{q}.$$

Thus

$$P(\mathcal{A} \cap \mathcal{B}) \geq 1 - K|x_0(T)|^{\beta} - K|x_0(T)|^{q},$$

$$P\{\sup_{0 \leq t \leq T} e^{-2\varepsilon t} |\bar{X}_t^{x_0}|^{-1} |Z_t^{(x_0)}| \leq |x_0(T)|^{\nu-\mu}\} \geq 1 - K|x_0(T)|^{\beta} - K|x_0(T)|^{q},$$

and

$$P\{e^{-2\varepsilon T} \sup_{0 \leq t \leq T} |\bar{X}_t^{x_0}|^{-1} |Z_t^{(x_0)}| \leq |x_0(T)|^{\nu-\mu}\} \geq 1 - K|x_0(T)|^{\beta} - K|x_0(T)|^{q}.$$

The assertion of the lemma follows from the choice of ε satisfying the inequality $2\varepsilon < \delta(\nu - \mu)$. □

Remark 4.1. The analysis of a one-dimensional equation is much easier because $\Lambda_t^{(x_0)} = \bar{\Lambda}_t^{(x_0)}$ for $d = 1$, and we do not need Lemmas 4.2-4.4. Let us also note that the condition (C_3) is essential for the proof of Lemma 4.4 (both the uniqueness of the Lyapunov exponent and the ergodicity of $\bar{\Lambda}(t)$ follow from (C_3)). As a clarifying example, let us consider the following deterministic system with two Lyapunov exponents

$$\dot{X}_1 = aX_1,\ \dot{X}_2 = bX_2 + X_1^2,\ a < b < 0.$$

Setting $x_2^0 = 0$, we have

$$\frac{|X - \bar{X}|}{|\bar{X}|} = \frac{|x_1^0|}{|2a - b|}|e^{at} - e^{(b-a)t}|.$$

It follows from this equation that (4.27) is not valid for $x_2^0 = 0$. □

Now we will study the properties of the estimator (3.5) for the *nonlinear* equation (4.1). Because we are not confident in the existence of the inverse matrix $H^{-1}(X^T)$ a.s., we will use the pseudoinverse matrix H^+ in (3.5). Thus we consider the estimator

$$\hat{\theta} = \frac{1}{T}H^+(X^T)V(X^T) = \frac{1}{T}H^+(\Lambda^T)V(\Lambda^T). \tag{4.28}$$

Introduce the notation

$$\delta H_{ij} = H_{ij}(X^T) - H_{ij}(\bar{X}^T) = H_{ij}(\Lambda^T) - H_{ij}(\bar{\Lambda}^T),$$
$$\delta H = H(X^T) - H(\bar{X}^T) = H(\Lambda^T) - H(\bar{\Lambda}^T).$$

Then (below we use the more detailed notation $H(X^{x_0,T})$ for $H(X^T)$ et al. if $X(0) = x_0 = x_0(T)$)

$$H(X^{x_0,T}) = H(\bar{X}^{x_0,T}) + \delta H = H(\bar{X}^{x_0,T})[E + H^{-1}(\bar{X}^{x_0,T})\delta H]. \tag{4.29}$$

A sufficient condition for existence of the inverse matrix $[E + H^{-1}(\bar{X}^{x_0,T})\delta H]^{-1}$ consists in the inequality

$$||H^{-1}(\bar{X}^{x_0,T})\delta H|| \leq q < 1 \tag{4.30}$$

which is fulfilled with some probability depending on T and x_0. Under (4.30) we get from (4.29):

$$H^{-1}(X^{x_0,T}) = H^{-1}(\bar{X}^{x_0,T}) + \sum_{k=1}^{\infty}[H^{-1}(\bar{X}^{x_0,T})\delta H]^k H^{-1}(\bar{X}^{x_0,T}) \tag{4.31}$$

and therefore the norm of

$$\delta(H^{-1}) := \sum_{k=1}^{\infty}[H^{-1}(\bar{X}^{x_0,T})\delta H]^k H^{-1}(\bar{X}^{x_0,T}) \tag{4.32}$$

is small enough if $\|H^{-1}(\bar{X}^{x_0,T})\|$ is bounded and $\|\delta H\|$ is small enough. We emphasize that (4.31) and (4.32) take place if (4.30) is valid.

Let $||I^{-1}(\theta)|| = K$ (we recall that the matrix $I(\theta)$ is deterministic and positively definite). If we take $x_0(T) = |x_0(T)|\lambda_0$ with a fixed vector λ_0, then due to (3.6), we get

$$\lim_{T\to\infty} P\{||H^{-1}(\bar{X}^{x_0,T})|| \leq 2K\} = \lim_{T\to\infty} P\{||H^{-1}(\bar{\Lambda}^{\lambda_0,T})|| \leq 2K\} = 1. \tag{4.33}$$

Introduce the notation

$$\delta\Lambda_t = \Lambda_t^{x_0} - \bar\Lambda_t^{\lambda_0},\ \delta B_t^+ = B^+(\Lambda_t^{x_0}) - B^+(\bar\Lambda_t^{\lambda_0}).$$

According to Lemma 4.4 and condition (C_4), the following equations hold for $|x_0(T)| \leq \exp(-\delta T)$:

$$P\{\lim_{T\to\infty} \sup_{0\leq t\leq T} |\delta\Lambda_t| \to 0\} = 1,\ P\{\lim_{T\to\infty} \sup_{0\leq t\leq T} |\delta B_t^+| \to 0\} = 1. \tag{4.34}$$

Thus

$$\begin{aligned} \delta H_{ij} &= H_{ij}(\Lambda^{x_0,T}) - H_{ij}(\bar\Lambda^{\lambda_0,T}) \\ &= \frac{1}{2T}\int_0^T ((A_i^* B^+(\Lambda_t^{x_0})A_j + A_j^* B^+(\Lambda_t^{x_0})A_i)\Lambda_t^{x_0}, \Lambda_t^{x_0})dt \\ &- \frac{1}{2T}\int_0^T ((A_i^* B^+(\bar\Lambda_t^{\lambda_0})A_j + A_j^* B^+(\bar\Lambda_t^{\lambda_0})A_i)\bar\Lambda_t^{\lambda_0}, \bar\Lambda_t^{\lambda_0})dt \\ &= \frac{1}{2T}\int_0^T ((A_i^* \delta B_t^+ A_j + A_j^* \delta B_t^+ A_i)\Lambda_t^{x_0}, \Lambda_t^{x_0})dt \\ &+ \frac{1}{2T}\int_0^T ((A_i^* B^+(\bar\Lambda_t^{\lambda_0})A_j + A_j^* B^+(\bar\Lambda_t^{\lambda_0})A_i)(\Lambda_t^{x_0} + \bar\Lambda_t^{\lambda_0}), \delta\Lambda_t)dt. \end{aligned} \tag{4.35}$$

Making use of (4.32)-(4.35), we obtain the following result.

Lemma 4.5. *Let the SDE* (4.6) *be stable, δ be an arbitrary small positive number, and*

$$|x_0(T)| \leq e^{-\delta T}.$$

Then

(4.36)

$$P\{\lim_{T\to\infty} ||\delta H|| = 0\} = 1,\ P\{\lim_{T\to\infty} H(X^{x_0,T}) = I(\theta)\} = 1,\ P\{\lim_{T\to\infty} ||\delta(H^{-1})|| = 0\} = 1.$$

In particular

$$\lim_{T\to\infty} P\{H^+(X^{x_0,T}) = H^{-1}(X^{x_0,T})\} = 1.$$

Let us return to the formula (4.28). We have

$$\sum_{j=1}^k H_{ij}(X^{x_0,T})\hat\theta_j = \frac{1}{T}\int_0^T (B^+(X_t)A_iX_t, dX_t - A_0X_tdt)$$

$$= \frac{1}{T}\int_0^T (B^+(X_t)A_iX_t, \sum_{j=1}^k A_j\theta_jX_tdt + \varphi(X_t)dt + \sum_{r=1}^q \sigma_rX_tdw_r(t) + \sum_{r=1}^q \psi_r(X_t)dw_r(t))$$

$$= \sum_{j=1}^k H_{ij}(X^{x_0,T})\theta_j + \frac{1}{T}\int_0^T (B^+(X_t)A_iX_t, \sum_{r=1}^q \sigma_rX_tdw_r(t))$$

$$+\frac{1}{T}\int_0^T (B^+(X_t)A_iX_t, \varphi(X_t)dt + \sum_{r=1}^q \psi_r(X_t)dw_r(t)).$$

Thus

$$\sqrt{T}H(X^{x_0,T})(\hat{\theta}-\theta) = \zeta(T) + \xi(T),$$

where

$$\zeta_i(T) = \frac{1}{\sqrt{T}}\int_0^T (B^+(X_t)A_iX_t, \sum_{r=1}^q \sigma_r X_t dw_r(t)),$$

$$\xi_i(T) = \frac{1}{\sqrt{T}}\int_0^T (B^+(X_t)A_iX_t, \varphi(X_t)dt + \sum_{r=1}^q \psi_r(X_t)dw_r(t)),\ i = 1, ..., k.$$

Due to Lemma 4.1, $\xi(T) \to 0$ in probability as $T \to \infty$ (the proof is the same as in the one-dimensional case).

Further we have

$$E\zeta(T)\zeta^\top(T) = EH(X^{x_0,T}).$$

According to Lemma 4.5, $H(X^{x_0,T}) \to I(\theta)$ as $T \to \infty$. These facts and results from [5] imply the following theorem.

Theorem 4.1. *Let the conditions* (C_1)-(C_4) *hold. Let the equation of first order variation* (4.6) *be stable and δ be an arbitrary small positive number. Then the estimator*

$$\hat{\theta} = \frac{1}{T}H^+(X^{x_0(T),T})V(X^{x_0(T),T})$$

with $X(0) = x_0(T)$ satisfying the condition

$$|x_0(T)| \le e^{-\delta T} \tag{4.37}$$

has the property

$$\mathcal{L}(\sqrt{T}I^{1/2}(\theta)(\hat{\theta}-\theta)) \to \mathcal{N}(0, 1_k) \text{ as } T \to \infty,$$

where 1_k is the unit $k \times k$ matrix and $I(\theta)$ is defined by (3.6).

Remark 4.2. Theorem 4.1 allows to propose an "almost" asymptotically efficient estimator if the initial condition $x_0 = x_0(T)$ satisfies the condition

$$x_0(T) \to 0 \text{ as } T \to \infty, \tag{4.38}$$

which is essentially weaker than (4.37).

Let the system (4.6) be stable. Then there exists a homogeneous of order $p > 0$, positive for $x \ne 0$ function $V(x)$ such that for some $c > 0$, $\rho > 0$ the inequality $LV \le -cV$ is valid for $|x| \le \rho$, where L is the generator of the nonlinear SDE (4.1). Let $|x_0(T)| < \rho$ and $\tau = \tau^{x_0(T)}$ be an exit time of $X^{x_0(T)}$ from the ball $|x| \le \rho$. Clearly, $\tilde{V}(t, X^{x_0(T)}(t)) := e^{ct}V(X^{x_0(T)}(t))$ is a local supermartingale.

We have for some $K > 0$ and any $t \ge 0$

$$P\{|X^{x_0}(t)|^p e^{ct} \ge K\} = P\{(|X^{x_0}(t)|^p e^{ct} \ge K) \cap (\tau < t)\} \tag{4.39}$$
$$+P\{(|X^{x_0}(t)|^p e^{ct} \ge K) \cap (\tau \ge t)\} = P(A) + P(B).$$

Let $k_1|x|^p \le V(x) \le k_2|x|^p$. We get

$$(4.40) \qquad P(A) \le P(\tau^{x_0(T)} < t) = P\{\sup_{0\le s\le t} |X^{x_0(T)}(s)| \ge \rho\}$$
$$\le \frac{1}{k_1\rho^p} EV(X^{x_0(T)}(\tau\wedge t)) \le \frac{1}{k_1\rho^p} V(x_0(T)).$$

Further

$$(4.41) \qquad P(B) \le P\{|X^{x_0(T)}(\tau\wedge t)|^p e^{c(\tau\wedge t)} \ge K\} \le P\{V(X^{x_0(T)}(\tau\wedge t))e^{c(\tau\wedge t)} \ge k_2 K\}$$
$$\le \frac{1}{k_2K} EV(X^{x_0(T)}(\tau\wedge t))e^{c(\tau\wedge t)} \le \frac{1}{k_2K} V(x_0(T)).$$

From (4.39)-(4.41) we obtain that for $x_0(T)$ satisfying (4.38), $0 < \gamma < c$, and $N > 1$

$$\lim_{T\to\infty} P\{|X^{x_0(T)}(T/N)| < e^{-\gamma T}\} = 1.$$

Now we can use the estimator of Theorem 4.1 on the time interval $[T/N, T]$. Thus, the following result is valid.

Proposition 4.1. *Let the initial condition for the nonlinear SDE* (4.1) *satisfy the condition* (4.38) *and* T *be an observation time. For arbitrary* $N > 1$, *denote by* $\hat{\theta}_{[T/N,T]}$ *the estimator based on the solution of* (4.1) *on the interval* $[T/N, T]$ *with the initial condition* $X(T/N) = X^{x_0(T)}(T/N)$ *and defined by*

$$\hat{\theta}_{[T/N,T]} = \frac{1}{T(1-1/N)} H^+(X^{X(T/N),[T/N,T]})V(X^{X(T/N),[T/N,T]}).$$

Then

$$\mathcal{L}(\sqrt{T(1-1/N)} I^{1/2}(\theta)(\hat{\theta}-\theta)) \to \mathcal{N}(0, 1_k) \text{ as } T\to\infty.$$

It means that for $T\to\infty$ the asymptotic efficiency of estimator $\hat{\theta}_{[T/N,T]}$ with respect to the quadratic loss function is $(N-1)/N$.

5. Example. Stochastic oscillator

Consider the estimation problem of parameter θ in the system

$$(5.1) \qquad dX_1 = X_2 dt$$
$$dX_2 = -(a^2 \sin X_1 + \theta X_2)dt + \sigma_1 X_1 dw_1(t) + \sigma_2 X_2 dw_2(t).$$

The equation of the first order variation for (5.1) has the form

$$(5.2) \qquad d\bar{X}_1 = \bar{X}_2 dt$$
$$d\bar{X}_2 = -(a^2\bar{X}_1 + \theta\bar{X}_2)dt + \sigma_1\bar{X}_1 dw_1(t) + \sigma_2\bar{X}_2 dw_2(t).$$

We assume that $\sigma_1 \ne 0,\ \sigma_2 \ne 0$. The matrix $B^+(\lambda)$ is equal to

$$B^+(\lambda) = \begin{bmatrix} 0 & 0 \\ 0 & (\sigma_1^2\lambda_1^2 + \sigma_2^2\lambda_2^2)^{-1} \end{bmatrix}$$

and all the conditions (C$_1$)-(C$_4$) are fulfilled (while the condition (C$_5$) is not valid).

We have

$$\hat{\theta} = -\left(\int_0^T \frac{X_2^2}{\sigma_1^2 X_1^2 + \sigma_2^2 X_2^2} dt\right)^{-1} \cdot \left(\int_0^T \frac{a^2 X_1 X_2}{\sigma_1^2 X_1^2 + \sigma_2^2 X_2^2} dt + \int_0^T \frac{X_2 dX_2}{\sigma_1^2 X_1^2 + \sigma_2^2 X_2^2}\right). \tag{5.3}$$

If it is known that $\theta \geq \theta_0 > 0$ and σ_1, σ_2 are comparatively small, then the trivial solution of the system (5.2) is stable in probability. Due to Lemma 4.1 and Theorem 4.1, the initial data $X(0) = x_0(T)$ can be chosen as (here $\alpha = 1$)

$$0 < |x_0(T)| \leq e^{-\delta t},$$

where $\delta > 0$ is an arbitrary constant.

Remark 5.1. Consider a model of physical pendulum with unknown damping θ subject to a random perturbation by the white Gaussian noise of intensity σ. The equation of motion for this model has the form

$$\ddot{X} + a^2 \sin X + (\theta + \sigma \dot{w})\dot{X} = 0 \tag{5.4}$$

It is natural to treat this equation as an Ito SDE in the form (5.1) with $X = x_1$, $x_2 = \dot{X}$ and $\sigma_1 = 0$. The only nonzero element of the matrix $B^+(\lambda)$ is equal to $\sigma_2^{-2}\lambda_2^{-2}$ if $\lambda_2 \neq 0$, and it is equal to 0 if $\lambda_2 = 0$. Thus the condition (C$_4$) is not valid. The expression for $\hat{\theta}$ has the form

$$\hat{\theta} = -\frac{1}{T}\int_0^T \frac{a^2 X_1 dt + dX_2}{X_2}. \tag{5.5}$$

It is easy to see that the integral in (5.5) does not converge and therefore the estimator (5.5) has no sense. But the nonlinear estimator

$$\hat{\theta} = -\frac{1}{T}\int_0^T \frac{a^2 \sin X_1 dt + dX_2}{X_2}.$$

has sense because $a^2 \sin X_1 dt + dX_2 = -\theta X_2 dt + \sigma_2 X_2 dw_2(t)$.

Moreover, as

$$\hat{\theta} = \theta - \frac{\sigma_2}{T} w(T),$$

this estimator is asymptotically efficient. So we see that sometimes it is possible to create the efficient estimator even in a situation when the condition (C$_4$) is not valid.

6. Acknowledgment

The first author was partially supported by NSF Grant DMS 9971608. The second author was partially supported by NSF Grant DMS 9971608 and by the Russian Foundation for Basic Research (project 99-01-00134).

References

[1] L. Arnold and R. Khasminskii. Stability index for nonlinear stochastic differential equations. Proc. of Symposia in Pure Math. 57(1995), pp. 543-551.

[2] L. Arnold, E. Oeljeklaus, and E. Pardoux. Almost sure and moment stability for linear Ito equations. Lyapunov exponents (L. Arnold and V. Wihstutz, eds.), Lecture Notes in Math., vol. 1186, Springer, Berlin, 1986, pp.129-159.

[3] A. Jankunas and R. Khasminskii. Estimation of parameters of linear stochastic differential equations. Stoch. Proc. Applic. 72(1997), pp.205-219.

[4] R.Z. Khasminskii. Stochastic stability of differential equations. Nauka, Moscow, 1969; Engl. transl., Sijthoff and Noordhoff, Alphen, 1980.

[5] R. Khasminskii, N. Krylov, N. Moshchuk. On the estimation of parameters for linear stochastic differential equations. Probab. Theory Relat. Fields 113(1999), pp. 443-472.

[6] R. Liptser, A. Shyryaev. Statistics of random processes. Springer Verlag, New York, 1977.