Small Noise Asymptotics of Nonlinear Filters with Nonobservable Limiting Deterministic System*

Marc Joannides and François LeGland
IRISA / INRIA
Campus de Beaulieu
35042 Rennes Cédex, France
e-mail : {mjoannid,legland}@irisa.fr

Abstract

We study the asymptotic behaviour of the Bayesian estimator for a deterministic signal in additive Gaussian white noise, in the case where the set of minima of the Kullback–Leibler information is a submanifold of the parameter space. This problem includes as a special case the study of the asymptotic behaviour of the nonlinear filter, when the state equation is noise–free, and when the limiting deterministic system is nonobservable. We present a practical example where this situation occurs. We give an explicit expression of the limit, as the noise intensity goes to zero, of the posterior probability distribution of the parameter, and we study the rate of convergence.

1 Introduction

Consider a nonlinear filtering problem with noise–free dynamics, where the unobserved process $\{X_t\,,\,0 \leq t \leq T\}$ evolves according to the ODE

$$\dot{X}_t = b(X_t)\ ,$$

with unknown initial condition X_0, and the observations are corrupted by some small additive white noise

$$dY_t = h(X_t)\,dt + \varepsilon\,dV_t\ .$$

The general nonlinear filtering problem consists in computing the conditional distribution μ_t^ε of the state X_t given the past observations $\mathcal{Y}_t = \sigma(Y_s\,,\,0 \leq s \leq t)$. Let $\{\phi_t(\cdot)\,,\,t \geq 0\}$ denote the flow of diffeomorphisms associated with the ODE, and consider the following deterministic system

$$(\Sigma) \qquad \left\{ \begin{aligned} \dot{x}_t &= b(x_t) \\ z_t &= h(x_t) \end{aligned} \right.$$

*This work was partially supported by the Commission of the European Communities, 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.

obtained in the limit as $\varepsilon \downarrow 0$. If the system (Σ) is *observable* on the time interval $[0,t]$, in the sense that the mapping

$$x \mapsto (h \circ \phi_s(x)\,,\,0 \leq s \leq t)\ ,$$

is injective, then

$$\mu_t^\varepsilon \Longrightarrow \delta_{\phi_t(x_0)}$$

as $\varepsilon \downarrow 0$, where x_0 is the *true* initial condition. The purpose of this paper is to describe the asymptotic behaviour of the conditional distribution μ_t^ε as $\varepsilon \downarrow 0$, when the system (Σ) is *nonobservable.*

This problem can be considered as a particular case of the problem of studying the small noise asymptotics of the Bayesian estimator in nonidentifiable nonlinear regressions. To be more specific, assume that the m–dimensional observation $\{X_t\,,\,0 \leq t \leq T\}$ has the differential

$$dX_t = m_t(\theta)\,dt + \varepsilon\,dW_t^\theta\ ,$$

where $\theta \in \Theta \subset \mathbf{R}^p$ is an unknown parameter and $\{W_t^\theta\,,\,0 \leq t \leq T\}$ is a standard Wiener process. The process $\{X_t\,,\,0 \leq t \leq T\}$ could be thought of as the sum of a parameter dependent signal corrupted by an additive white noise, see Ibragimov and Khasminskii [7]. The problem is to estimate the unknown parameter θ, given the observations $\{X_t\,,\,0 \leq t \leq T\}$. We assume that for any $\theta \in \Theta$, the mapping $t \longmapsto m_t(\theta)$ is measurable and satisfy the *finite energy* condition

$$\int_0^T |m_t(\theta)|^2\,dt < \infty\ .$$

In addition, we assume that for a.e. $0 \leq t \leq T$, the mapping $\theta \longmapsto m_t(\theta)$ is continuously differentiable, and for any $\theta \in \Theta$, the $p \times p$ symmetric nonnegative matrix (Fisher information matrix)

$$I(\theta) = \int_0^T [\dot{m}_t(\theta)]^*\,\dot{m}_t(\theta)\,dt$$

can be defined. For any $\varepsilon > 0$, let $\{\mathbf{P}_\theta^\varepsilon\,,\,\theta \in \Theta\}$ be the family of the probability measures generated on the

0-7803-4187-2 1663-1668

canonical space $C([0,T]\,;\mathbf{R}^m)$ by the process $\{X_t\,,\,0 \le t \le T\}$ for different values of the parameter θ. The likelihood function for the estimation of θ based on $\{X_t\,,\,0 \le t \le T\}$ is given by

$$L^\varepsilon(\theta) \;=\; \exp\{\frac{1}{\varepsilon^2}\int_0^T [m_t(\theta)]^*\, dX_t - \frac{1}{2\,\varepsilon^2}\int_0^T |m_t(\theta)|^2\, dt\}\;.$$

Using the Bayesian approach, we model the *a priori* information on the unknown parameter θ by the prior probability distribution $p(\theta)\,d\theta$. The *posterior* probability distribution μ^ε is then defined by the Bayes rule

$$\langle \mu^\varepsilon, \phi\rangle = \frac{\displaystyle\int_\Theta \phi(\theta)\, L^\varepsilon(\theta)\, p(\theta)\, d\theta}{\displaystyle\int_\Theta L^\varepsilon(\theta)\, p(\theta)\, d\theta}\;.$$

The Bayesian point estimator $\widetilde{\theta}^\varepsilon$ associated with a quadratic loss function coincides with the conditional mean, i.e.

$$\widetilde{\theta}^\varepsilon = \int_\Theta \theta\, \mu^\varepsilon(d\theta)\;.$$

We introduce the *contrast process* $\ell^\varepsilon(\theta) = -\varepsilon^2 \log L^\varepsilon(\theta)$ and we denote by α the *true* value of the parameter. Discarding additional terms that are independent of θ, the contrast process could be written

$$\ell^\varepsilon(\theta) \;=\; -\varepsilon \int_0^T [m_t(\theta) - m_t(\alpha)]^*\, dW_t^\alpha + \tfrac{1}{2}\int_0^T |m(\theta) - m_t(\alpha)|^2\, dt\;,$$

and converges to the Kullback–Leibler information

$$K_\alpha(\theta) = \tfrac{1}{2}\int_0^T |m_t(\theta) - m_t(\alpha)|^2\, dt\;,$$

in $\mathbf{P}^\varepsilon_\alpha$–probability, as $\varepsilon \downarrow 0$.

Under the usual *identifiability* assumption that the true value of the parameter is the only minimum point of $K_\alpha(\theta)$, the Bayesian estimator is consistant, i.e.

$$\widetilde{\theta}^\varepsilon \longrightarrow \alpha \qquad \text{or equivalently} \qquad \mu^\varepsilon \Longrightarrow \delta_\alpha\;,$$

in $\mathbf{P}^\varepsilon_\alpha$–probability, as $\varepsilon \downarrow 0$, and is asymptotically normal, i.e.

$$\frac{1}{\varepsilon}\,[\widetilde{\theta}^\varepsilon - \alpha] \Longrightarrow \mathcal{N}(0, I^{-1}(\alpha))\;,$$

as $\varepsilon \downarrow 0$, provided the Fisher information matrix $I(\alpha)$ is invertible, see Kutoyants [11].

However, there are some practical situations, see Example 2.1 below, where *nonidentifiablilty* occurs, i.e. $M_\alpha \neq \{\alpha\}$ where

$$M_\alpha \triangleq \operatorname*{argmin}_{\theta\in\Theta} K_\alpha(\theta)\;,$$

by definition. In this case, the point estimator $\widetilde{\theta}^\varepsilon$ is not relevant, and we are rather interested in the asymptotic behavior of the *posterior* probability distribution μ^ε as $\varepsilon \downarrow 0$.

In this paper, we address the case where the set of points with minimum contrast (Kullback–Leibler information) is a submanifold of $\mathbf{R}^p$. It is easy to show that asymptotically as $\varepsilon \downarrow 0$ the probability distribution μ^ε is supported by M_α. We show that, using first order terms, such as the Fisher information matrix, it is possible to characterize the limit as $\varepsilon \downarrow 0$ of the probability distribution μ^ε, as a *random* probability distribution μ_α on M_α, absolutely continuous w.r.t. the canonical measure on M_α, and to provide an explicit expression for the density. We also study the rate of convergence. These results have been announced in Joannides and LeGland [10].

2 Prototypical situation, and example

As a prototypical situation where nonidentifiability occurs, consider the nonlinear filtering problem introduced at the beginning, where the unobserved process $\{X_t\,,\,0 \le t \le T\}$ evolves according to the ODE

$$\dot{X}_t = b(X_t)\;,$$

with unknown initial condition X_0, and the observations are corrupted by some small additive white noise

$$dY_t = h(X_t)\, dt + \varepsilon\, dV_t = h\circ\phi_t(X_0)\, dt + \varepsilon\, dV_t\;.$$

In the Bayesian approach, the unknown initial condition X_0 is given the probability distribution $p_0(x)\,dx$, and the general nonlinear filtering problem consists in computing the conditional distribution of the state X_t given the past observations $\mathcal{Y}_t = \sigma(Y_s\,,\,0 \le s \le t\}$.

The limiting deterministic system (Σ) is *nonobservable* if

$$M_0 = \operatorname*{argmin}_{x\in\mathbf{R}^m}\int_0^T |h\circ\phi_t(x) - h\circ\phi_t(x_0)|^2\, dt \neq \{x_0\}\;,$$

where x_0 is the true initial condition. In this example, the notion of nonobservability of the limiting deterministic system is equivalent to the notion of nonidentifiability of the corresponding statistical problem. In this context, the Fisher information matrix reads as follows :

$$I(x) = \int_0^T [\,h'[\phi_t(x)]\,\phi_t'(x)\,]^*\; h'[\phi_t(x)]\,\phi_t'(x)\, dt$$

for any $x \in \mathbf{R}^m$, and has been introduced in James [8] as the *observability Grammian* for the limiting deterministic system (Σ).

Example 2.1 Let us mention the target motion analysis (TMA), or tracking with bearings only measurements, as a typical application where nonobservability occurs, see Lévine and Marino [12].

In this problem, a ship (the platform) is trying to estimate the position and velocity of another ship (the target), using angle only measurements provided by a passive sonar. We denote by $r = (r^x, r^y)$ and $v = (v^x, v^y)$ the relative position and velocity of the target w.r.t. the platform. The measurements available at time $t \geq 0$ are of the form

$$dY_t = h(r_t, v_t)\, dt + \varepsilon\, dW_t$$

where the observation function is

$$h(r, v) = \arctan \frac{r^x}{r^y} \ .$$

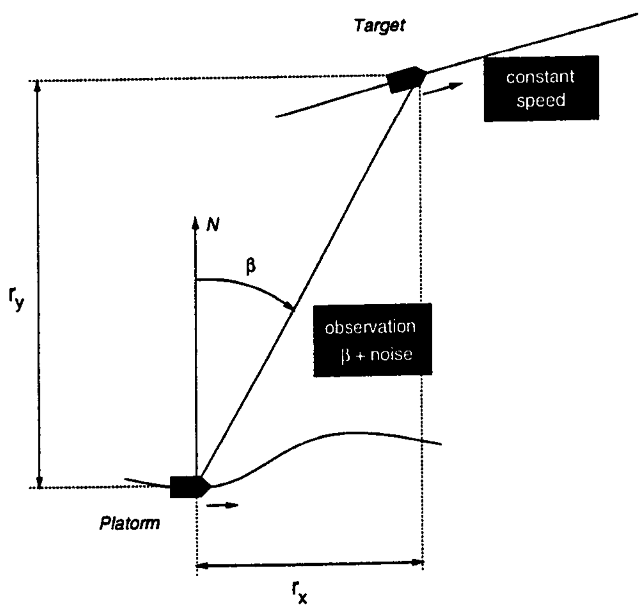

Figure 1: Target motion analysis (TMA)

In general, both ships move at constant velocity along straight lines, so that the motion of the target could be described by the equation

$$r_t^x = r_0^x + v_0^x\, t \ , \qquad r_t^y = r_0^y + v_0^y\, t \ ,$$

$$v_t^x = v_0^x \ , \qquad v_t^y = v_0^y \ .$$

Denoting by $\theta \triangleq (r_0^x, r_0^y, v_0^x, v_0^y)$ the initial condition of the above set of equations, the problem reduces to estimating θ based on the observations

$$dY_t = m_t(\theta)\, dt + \varepsilon\, dW_t^\theta \ ,$$

where

$$m_t(\theta) = \arctan \frac{\theta_1 + t\,\theta_3}{\theta_2 + t\,\theta_4} \ .$$

Denote by $M_\alpha = \{\theta \in \mathbf{R}^4 \, : \, m_t(\theta) = m_t(\alpha)\, , \, 0 \leq t \leq T\}$ the set of points that cannot be distinguished from α. Clearly, $\theta \in M_\alpha$ if and only if for any $0 \leq t \leq T$

$$\frac{\theta_1 + t\,\theta_3}{\theta_2 + t\,\theta_4} = \frac{\alpha_1 + t\,\alpha_3}{\alpha_2 + t\,\alpha_4} \ ,$$

i.e. if and only if

$$\begin{pmatrix} \alpha_2 & -\alpha_1 & 0 & 0 \\ \alpha_4 & -\alpha_3 & \alpha_2 & -\alpha_1 \\ 0 & 0 & \alpha_4 & -\alpha_3 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{pmatrix} = 0 \ .$$

Notice that the rows of this matrix are linearly independent vectors of $\mathbf{R}^4$, provided that $\alpha_4\, \alpha_1 - \alpha_2\, \alpha_3 \neq 0$. If this is the case, then M_α is the one–dimensional kernel of this matrix. We can easily check that α is in the kernel, hence

$$M_\alpha = \{y \in \mathbf{R}^4 \, : \, y = \rho\, \alpha \quad \text{for some } \rho \in \mathbf{R}\} \ .$$

Indeed there exist a one–dimensional manifold, in the four–dimensional parameter space (initial relative position and velocity of the target w.r.t. the platform), of points that cannot be distinguished from the *true* value. One would need additional measurements, e.g. range measurements provided by an active sonar, to overcome this problem.

The case where $\alpha_4\, \alpha_1 - \alpha_2\, \alpha_3 = 0$ corresponds to some values of α for which the platform and the target are moving along the same line. We do not consider that type of problem here and therefore, we discard these values of α from the set of parameter Θ.

3 Convergence result

For $\theta \in \Theta$, recall the Fisher information matrix

$$I(\theta) = \int_0^T [\dot{m}_t(\theta)]^*\, \dot{m}_t(\theta)\, dt \ ,$$

and notice that for any $y \in M_\alpha$, $I(y)$ is the Hessian matrix at point $\theta = y$ of the Kullback–Leibler information $K_\alpha(\theta)$. If $I(y)$ has full rank p, then by the local inversion theorem, y is an isolated minimum point. This case has been considered in Kutoyants [11] for a slightly different model however. Since our goal is to address the case where the set of points of minimum contrast is a submanifold, we will make hereafter the following

Assumption A:

> The Fisher information matrix $I(y)$ has constant rank $d \leq p$ for any $y \in M_\alpha$.

By the rank theorem, this assumption ensures that M_α is a $(p-d)$–dimensional submanifold of $\mathbf{R}^p$. For $y \in M_\alpha$, we denote respectively by $T_y M_\alpha$ and $N_y M_\alpha$ the tangent and normal spaces to M_α at point y. We easily check that $T_y M_\alpha = \ker I(y)$, and $N_y M_\alpha = \operatorname{Im} I(y)$. Indeed, for any tangent vector v to M_α, let $\{\gamma(s)\,,\, s \in I\}$ be a smooth curve in M_α such that $\gamma(0) = y$ and $\dot\gamma(0) = v$. For any $s \in I$ and for a.e. $0 \leq t \leq T$, $m_t[\gamma(s)] = m_t(\alpha)$, hence $\dot m_t[\gamma(s)]\,\dot\gamma(s) = 0$ and in particular for $s = 0$, $\dot m_t(y)\,v = 0$. Since the latter holds for a.e. $0 \leq t \leq T$, we have $I(y)\,v = 0$ which implies $\operatorname{Im} I(y) \subset N_y M_\alpha$ since for any $u \in \mathbf{R}^p$, $[I(y)\,u]^* v = 0$. By Assumption A, $\operatorname{Im} I(y)$ is a d–dimensional subspace of $\mathbf{R}^d$, hence $N_y M_\alpha = \operatorname{Im} I(y)$, which in turn implies $T_y M_\alpha = \ker I(y)$.

The above discussion also shows that the Gaussian process

$$\zeta_\alpha(y) = \int_0^T [\dot m_t(y)]^*\, dW_t^\alpha \ ,$$

indexed by $y \in M_\alpha$ belongs to $N_y M_\alpha$, since $v^*\,\zeta_\alpha(y) = 0$ for any $v \in T_y M_\alpha$. In addition, for any $y \in M_\alpha$, the restriction $I_\perp(y)$ of the linear mapping $I(y)$ to the normal space $N_y M_\alpha$ is invertible, since it has full rank d. Thanks to these remarks, the Gaussian random field

$$\xi_\alpha(y) = [I_\perp(y)]^{-1/2}\,\zeta_\alpha(y)$$

indexed by $y \in M_\alpha$ is well defined. We also introduce the *random* probability measure on M_α

$$\mu_\alpha(dy) = c_\alpha \exp\{\tfrac{1}{2}\,|\xi_\alpha(y)|^2\}\, \frac{p(y)}{\sqrt{\det I_\perp(y)}}\, \lambda_\alpha(dy) \ ,$$

where c_α is a normalizing constant. In this definition, λ_α denotes the *canonical* (or Lebesgue) measure on M_α, see Berger and Gostiaux [1, Chapter 3]. We recall at this point that the normal space $N_y M_\alpha$ is also equipped with a canonical Lebesgue measure $\lambda_{\alpha,y}$, as a d–dimensional vector subspace of $\mathbf{R}^p$.

Notice that both μ^ε and μ_α are random measures, i.e. random variables with value in $\mathcal{M}$, the set of probability measure on $\mathbf{R}^p$. Since μ^ε is absolutely continuous w.r.t. the Lebesgue measure on $\mathbf{R}^p$, and μ_α is supported by a set of zero Lebesgue measure, the topology associated with the total variation distance is obviously too strong, and we use instead the topology associated with the following norm on $\mathcal{M}$, which is equivalent to the Prokhorov distance

$$\|\mu - \mu'\|_{\mathrm{BL}*} = \sup_{\|f\|_{\mathrm{BL}}=1} |\,\langle \mu, f\rangle - \langle \mu', f\rangle\,| \ ,$$

where $\|\cdot\|_{\mathrm{BL}}$ is the norm on the space of bounded and Lipschitz continuous functions, i.e.

$$\|f\|_{\mathrm{BL}} = \max\{\|f\|, \|f\|_{\mathrm{L}}\} \ ,$$

where $\|\cdot\|$ and $\|\cdot\|_{\mathrm{L}}$ denotes respectively the supremum and the Lipschitz norm. The equivalence between the Prokhorov distance and the norm $\|\cdot\|_{\mathrm{BL}*}$ is proved in Dudley [2, 3].

We now state the main result of this paper.

Theorem 3.1 *We assume that Θ is a* compact *subset of $\mathbf{R}^p$, and that the prior density p is continuous and strictly positive on M_α. Under Assumption A, and under additional regularity assumptions*

$$\|\mu^\varepsilon - \mu_\alpha\|_{\mathrm{BL}*} \longrightarrow 0 \ ,$$

in $\mathbf{P}_\alpha^\varepsilon$-probability as $\varepsilon \downarrow 0$.

SKETCH OF THE PROOF. From the definition of the norm $\|\cdot\|_{\mathrm{BL}*}$, it is enough to look at the difference

$$\langle \widetilde\mu^\varepsilon, \phi\rangle - \langle \widetilde\mu_\alpha, \phi\rangle \ ,$$

between the unnormalized probability distributions $\widetilde\mu^\varepsilon$ and $\widetilde\mu_\alpha$ defined by

$$\langle \widetilde\mu^\varepsilon, \phi\rangle = \varepsilon^{-d} \int_\Theta \phi(\theta)\, p(\theta)\, \exp\{-\frac{1}{\varepsilon}\,\ell^\varepsilon(\theta)\}\, d\theta \ , \qquad (1)$$

and

$$\langle \widetilde\mu_\alpha, \phi\rangle = (2\,\pi)^{d/2} \int_{M_\alpha} \phi(y)\, \exp\{\tfrac{1}{2}\,|\xi_\alpha(y)|^2\} \frac{p(y)}{\sqrt{\det I_\perp(y)}}\, \lambda_\alpha(dy) \ .$$

respectively. The approach is to obtain suitable estimates for this difference, holding at least on some *good* sets, and to control the probability of the complementary *bad* sets.

The integral in (1) is viewed as a Laplace integral, hence for small values of ε, we expect its behaviour to be determined only by the set of minimum points of ℓ^ε. As $\varepsilon \downarrow 0$, this set shrinks to M_α, so we are naturally lead to evaluating this integral over a small neighborhood of M_α. Following the same approach as in Hwang [6], the first step consists in replacing the integral over the whole space Θ by the integral over a small tubular neighbourhood of M_α, and to estimate the resulting approximation error. For $r > 0$, the r–neighborhood of M_α is the set

$$M_\alpha^r \triangleq \{\theta \in \Theta \, : \, d(\theta, M_\alpha) < r\} \; ,$$

and we show that the contribution of $\Theta \backslash M_\alpha^r$ is exponentially small as $\varepsilon \downarrow 0$. The next step is to evaluate the integral over M_α^r using a Fubini–like theorem. Since M_α is compact, then by the tubular neighbourhood theorem, see Berger and Gostiaux [1, Theorem 2.7.12], for $r > 0$ small enough

$$M_\alpha^r \simeq \{(y, u) \, : \, y \in M_\alpha, u \in N_y M_\alpha, |u| \leq r\} \triangleq N^r M_\alpha \; ,$$

under the diffeomorphism

$$j_\alpha \; : \; (y, u) \in N^r M_\alpha \longmapsto \theta = y + u \in M_\alpha^r \; .$$

It is proved in Weyl [13] that the Jacobian determinant $J_\alpha(y, u)$ of the transformation j_α is a positive continuous function, identically equal to 1 on the manifold M_α, i.e. for $u = 0$. The change of variable $\theta = y + u$ in the integral over M_α^r yields

$$\begin{aligned} &\varepsilon^{-d} \int_{M_\alpha} \{ \int_{N_y M_\alpha^r} \phi(y+u) \, \exp\{-\frac{1}{\varepsilon^2} \, \ell^\varepsilon(y+u)\} \\ &\qquad p(y+u) \, J_\alpha(y,u) \, \lambda_{\alpha,y}(du) \; \} \, \lambda_\alpha(dy) \\ &= \int_{M_\alpha} \{ \int_{N_y M_\alpha^{r/\varepsilon}} \phi(y+\varepsilon\, u) \, \exp\{-\frac{1}{\varepsilon^2} \, \ell^\varepsilon(y+\varepsilon\, u)\} \\ &\qquad p(y+\varepsilon\, u) \, J_\alpha(y,\varepsilon\, u) \, \lambda_{\alpha,y}(du) \; \} \, \lambda_\alpha(dy) \; , \end{aligned}$$

after rescaling the innermost integral. Notice that the factor ε^{-d} was chosen in order to balance this rescaling. The only indetermination left in the rescaled integral is with the exponential. We therefore focus on the expression

$$\begin{aligned} -\frac{1}{\varepsilon^2} \, \ell^\varepsilon(y+\varepsilon\, u) &= \int_0^T [\frac{m_t(y+\varepsilon\, u) - m_t(y)}{\varepsilon}]^* \, dW_t^\alpha \\ &\quad - \tfrac{1}{2} \int_0^T |\frac{m_t(y+\varepsilon\, u) - m_t(y)}{\varepsilon}|^2 \, dt \; . \end{aligned}$$

Taylor expansion to the first order yields

$$-\frac{1}{\varepsilon^2} \, \ell^\varepsilon(y+\varepsilon\, u) \longrightarrow Q_{\alpha,y}(u) = u^* \, \zeta_\alpha(y) - \tfrac{1}{2} \, u^* I_\perp(y) \, u \; ,$$

as $\varepsilon \downarrow 0$, where $I_\perp(y)$ and $\zeta_\alpha(y)$ have been defined above in Section 3. Notice that

$$\begin{aligned} &\exp\{Q_\alpha(y,u)\} \, \lambda_{\alpha,y}(du) = \\ &= (2\,\pi)^{d/2} \, \exp\{\tfrac{1}{2}|\xi_\alpha(y)|^2\} \, \frac{1}{\sqrt{\det I_\perp(y)}} \, \Gamma_\alpha(y, du) \; , \end{aligned}$$

where $\Gamma_\alpha(y, du)$ is a Gaussian probability distribution on $N_y M_\alpha$, with *random* mean vector (Gaussian r.v. with values in $N_y M_\alpha$)

$$\chi_\alpha(y) \triangleq [I_\perp(y)]^{-1} \int_0^T [\dot{m}_t(y)]^* \, dW_t^\alpha$$

and covariance matrix $[I_\perp(y)]^{-1}$. As a result, the following convergence holds

$$\langle \widetilde{\mu}^\varepsilon, \phi \rangle \longrightarrow \langle \widetilde{\mu}_\alpha, \phi \rangle \; ,$$

as $\varepsilon \downarrow 0$. Further details are given in Joannides [9]. □

4 Rate of convergence

To study the rate of convergence, we define as follows the projection π on M_α

$$\pi(\theta) = \begin{cases} y \; , & \text{if } \theta = y + u \in M_\alpha^r, \\ 0 \; , & \text{otherwise.} \end{cases}$$

and we consider the small noise asymptotics of

$$\langle \nu^\varepsilon, f \rangle = \mathbf{E}^\varepsilon [f(\frac{\theta - \pi(\theta)}{\varepsilon}) \mid \mathcal{X}] \; ,$$

which is the conditional probability distribution of the r.v. $\frac{1}{\varepsilon}[\theta - \pi(\theta)]$. Following the same approach as in Ellis–Rosen [4], we define the following mixture of *random* Gaussian probability distributions

$$\langle \nu_\alpha, f \rangle = \int_{M_\alpha} \{ \int_{N_y M_\alpha} f(u) \, \Gamma_\alpha(y, du) \; \} \, \mu_\alpha(dy) \; ,$$

and we obtain our second theorem.

Theorem 4.1 *Under the same assumptions as in Theorem 3.1*

$$\|\nu^\varepsilon - \nu_\alpha\|_{\mathrm{TV}} \longrightarrow 0 \; ,$$

in $\mathbf{P}_\alpha^\varepsilon$-probability as $\varepsilon \downarrow 0$.

In particular, the mean value of the probability distribution ν_α satisfies

$$\int u \, \nu_\alpha(du) = \int_{M_\alpha} \chi_\alpha(y) \, \mu_\alpha(dy) \; .$$

5 Application to TMA

The results obtained in the previous two sections can be applied to the target motion analysis example, presented in Example 2.1 above. Simple calculations yield

$$[\dot{m}_t(\theta)]^* = \frac{1}{[\theta_1 + t\,\theta_3]^2 + [\theta_2 + t\,\theta_4]^2} \begin{pmatrix} \theta_2 + t\,\theta_4 \\ -[\theta_1 + t\,\theta_3] \\ t\,[\theta_2 + t\,\theta_4] \\ -t\,[\theta_1 + t\,\theta_3] \end{pmatrix} \; ,$$

for any $\theta \in \Theta$, hence for any $y \in M_\alpha$ of the form $y = \rho\,\alpha$ for some $\rho \in \mathbf{R}$

$$\dot{m}_t(y) = \frac{1}{\rho} \, \dot{m}_t(\alpha) \; .$$

A

It follows that for any $y \in M_\alpha$ of the form $y = \rho\,\alpha$ for some $\rho \in \mathbf{R}$

$$I(y) = \int_0^T [\dot{m}_t(y)]^* \, \dot{m}_t(y)\, dt = \frac{1}{\rho^2}\, I(\alpha) \ ,$$

and similarly for the 3×3 matrix

$$I_\perp(y) = \frac{1}{\rho^2}\, I_\perp(\alpha) \ ,$$

hence

$$\sqrt{\det I_\perp(y)} = \frac{1}{\rho^3}\, \sqrt{\det I_\perp(\alpha)} \ .$$

It follows also that for any $y \in M_\alpha$ of the form $y = \rho\,\alpha$ for some $\rho \in \mathbf{R}$

$$\xi_\alpha(y) = [I_\perp(y)]^{-1/2} \int_0^T [\dot{m}_t(y)]^* \, dW_t^\alpha = \xi_\alpha(\alpha) \ ,$$

does not depend on ρ, i.e. is constant over M_α, and

$$\chi_\alpha(y) = [I_\perp(y)]^{-1} \int_0^T [\dot{m}_t(y)]^* \, dW_t^\alpha = \rho\, \chi_\alpha(\alpha) \ .$$

Therefore, the limiting probability distribution μ_α on M_α is *nonrandom*, and has the form

$$\mu_\alpha(d\rho) = c_\alpha\, \rho^3\, p(\rho\,\alpha)\, d\rho \ ,$$

where c_α is a normalizing constant.

Since M_α is a one–dimensional *linear* submanifold, the normal spaces $N_y M_\alpha$ for $y \in M_\alpha$ are three–dimensional linear submanifolds, all parallel to the single vector space

$$M_\alpha^\perp = \{ v \in \mathbf{R}^4 \ : \ v^*\, \alpha = 0 \} \ ,$$

and the probability distribution governing the rate of convergence has the form

$$\langle \nu_\alpha, f \rangle = \int_{M_\alpha} (2\,\pi)^{-d/2} \sqrt{\det I_\perp(\alpha)} \int_{M_\alpha^\perp} f(\rho\, v)$$

$$\exp\{ -\tfrac{1}{2}\, [v - \chi_\alpha(\alpha)]^*\, I_\perp(\alpha)\, [v - \chi_\alpha(\alpha)] \}\, dv\ \mu_\alpha(d\rho) \ .$$

In particular, the mean value has the form

$$\int u\, \nu_\alpha(du) = \chi_\alpha(\alpha)\, \frac{\displaystyle\int_{M_\alpha} \rho^4\, p(\rho\,\alpha)\, d\rho}{\displaystyle\int_{M_\alpha} \rho^3\, p(\rho\,\alpha)\, d\rho} \ .$$

Acknowledgment

The authors gratefully acknowledge Matthew James for bringing the papers [4] and [5] to their attention, and Fabien Campillo for suggesting to look at the problem.

References

[1] M. BERGER and B. GOSTIAUX. *Differential Geometry : Manifolds, Curves and Surfaces*, volume 115 of *Graduate Texts in Mathematics*. Springer–Verlag, New York, 1988.

[2] R.M. DUDLEY. Convergence of Baire measures. *Studia Mathematica*, XXVII:251–268, 1966.

[3] R.M. DUDLEY. Distance of probability measures and random variables. *The Annals of Mathematical Statistics*, 39(5):1563–1572, 1968.

[4] R.S. ELLIS and J.S. ROSEN. Asymptotic analysis of Gaussian integrals, II : Manifold of minimum points. *Communications in Mathematical Physics*, 82:153–181, 1981.

[5] R.S. ELLIS and J.S. ROSEN. Asymptotic analysis of Gaussian integrals. I : Isolated minimum points. *Transactions of the American Mathematical Society*, 273(2):447–481, October 1982.

[6] HWANG Chii–Ruey. Laplace's method revisited : weak convergence of probability measures. *The Annals of Probability*, 8(6):1177–1182, 1980.

[7] I.A. IBRAGIMOV and R.Z. KHASMINSKII. *Statistical Estimation. Asymptotic Theory*, volume 16 of *Applications of Mathematics*. Springer Verlag, New York, 1981.

[8] M.R. JAMES. Finite time observer design by probabilistic–variational methods. *SIAM Journal on Control and Optimization*, 29(4):954–967, 1991.

[9] M. JOANNIDES. *Navigation Intégrée d'un Engin Sous–Marin Remorqué. Filtrage Non–Linéaire des Systèmes avec Contraintes et/ou Mesures Parfaites*. Thèse de Doctorat, Université de Provence, Marseille, March 1997. `http://www.irisa.fr/prive/legland/pub/joannides-phd.ps.gz`.

[10] M. JOANNIDES and F. LE GLAND. Small noise asymptotics of the Bayesian estimator in nonidentifiable nonlinear regressions. In *Proceedings of the 2nd Portuguese Conference on Automatic Control, Porto 1996*, pages 257–261. APCA, September 1996.

[11] Yu.A. KUTOYANTS. *Identification of Dynamical Systems with Small Noise*, volume 300 of *Mathematics and its Applications*. Kluwer Academic Publisher, Dordrecht, 1994.

[12] J. LÉVINE and R. MARINO. Constant–speed target tracking via bearings–only measurements. *IEEE Transactions on Aerospace and Electronic Systems*, AES–28(1):174–182, January 1992.

[13] H. WEYL. On the volume of tubes. *American Journal of Mathematics*, 61:461–472, 1939.