FP08 4:30

Proceedings of the 34th
Conference on Decision & Control
New Orleans, LA - December 1995

Nonlinear Filtering with Perfect Discrete Time Observations*

Marc Joannides
Department of Mathematics
Imperial College
London SW7 2BZ, United Kingdom
mjo@ma.ic.ac.uk

François LeGland
IRISA / INRIA
Campus de Beaulieu
35042 Rennes Cédex, France
legland@irisa.fr

Abstract : We consider the problem of estimating the state of a diffusion process, based on discrete time observations in singular noise. We reduce the problem to a static problem, and we show that the solution is provided by the area or co-area formula of geometric measure theory, provided the observed value is a regular value of the observation function. In order to address the case of singular values, we propose another approach, based on small-noise perturbation and asymptotics of Laplace integrals.

1 INTRODUCTION

One major limiting assumption in the nonlinear filtering literature is the non degeneracy of the observation noise covariance matrix. However, there are numerous situations of practical interest, where some perfect noise-free information is available about the unknown state, and yet the problem of state estimation with degenerate observation noise has received little attention, except in the linear case.

An additional motivation for studying this problem, is the existing connection with various problems of state estimation for non classical dynamical systems, including : hybrid systems, i.e. systems with state constraints, stochastic differential-algebraic systems, systems with colored noise, see Korezlioglu and Runggaldier [8], systems with state-dependent observation noise, see Takeuchi and Akashi [11], etc.

Finally, a better understanding of the nonlinear filtering problem with noise-free observations should help designing robust and efficient numerical approximation schemes in the important case where the observation noise is *small.*

To be more specific, we consider the following state equation in $\mathbf{R}^m$

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dW_t \ , \qquad (1)$$

where $\{W_t \, , \, t \geq 0\}$ is a Wiener process of appropriate dimension, and the noise-free d-dimensional discrete time observations

$$z_k = h(X_{t_k}) \ .$$

*This work was partially supported by the Army Research Office, under grant DAAH04-95-1-0164.

The objective of nonlinear filtering is to compute the conditional probability distribution

$$\mu_k(dx) = \mathbf{P}[X_{t_k} \in dx \mid z_1, \cdots, z_k] \ .$$

What makes this problem singular is that the state X_{t_k} is known *exactly* to belong to the level set M_{z_k}, where for all $z \in \mathbf{R}^d$

$$M_z = h^{-1}(z) = \{x \in \mathbf{R}^m \ : \ h(x) = z\} \ .$$

Therefore, the conditional probability distribution μ_k is supported by the set M_{z_k}, which in general has null Lebesgue measure, and μ_k does not have a density w.r.t. the Lebesgue measure on $\mathbf{R}^m$. The question naturally arises whether μ_k has a density w.r.t. some *canonical* measure on the level set M_{z_k}. The objective of this paper is to give conditions under which this holds, and to provide an explicit expression for the density.

As usual, the transition from μ_k to μ_{k+1} consists in two steps.

- In the prediction step, i.e. between times t_k and t_{k+1}, we solve the Fokker-Planck equation

$$\frac{\partial \mu_t^k}{\partial t} = L^* \, \mu_t^k \ , \qquad \mu_{t_k}^k = \mu_k \ ,$$

  where L^* is the adjoint of the infinitesimal generator L associated with equation (1). For all $t_k \leq t \leq t_{k+1}$ it holds

$$\mu_t^k(dx) = \mathbf{P}[X_t \in dx \mid z_1, \cdots, z_k] \ .$$

  and in particular at final time t_{k+1} we obtain the *prior* probability distribution $\mu_{k+1}^- = \mu_{t_{k+1}}^k$ of the state $X_{t_{k+1}}$.

- In the correction step, i.e. at time t_{k+1}, we have to combine the prior probability distribution μ_{k+1}^- and the new observation z_{k+1} which defines the level set $M_{z_{k+1}}$, so as to obtain the *posterior* probability distribution μ_{k+1} of the state $X_{t_{k+1}}$.

We make now the following two assumptions.

Assumption A :

For all $0 \leq s < t$, and all $x \in \mathbf{R}^m$, the conditional probability distribution

$$\mu_t^{s,x}(dy) = \mathbf{P}[X_t \in dy \mid X_s = x] ,$$

which is the solution of the Fokker–Planck equation

$$\frac{\partial \mu_t^{s,x}}{\partial t} = L^* \mu_t^{s,x} , \qquad \mu_s^{s,x} = \delta_x ,$$

has a continuous density w.r.t. the Lebesgue measure on $\mathbf{R}^m$.

Assumption B :

The mapping h is continuously differentiable from $\mathbf{R}^m$ to $\mathbf{R}^d$.

The purpose of Assumption A is to make sure that, even though μ_k is supported by a set of null Lebesgue measure, the prior probability distribution μ_{k+1}^- has a continuous density w.r.t. the Lebesgue measure on $\mathbf{R}^m$. Assumption B is used to describe the level sets M_z. A stronger assumption (Assumption C) will be introduced later.

Under Assumptions A and B, the original problem reduces as follows :

Let X be a r.v. in $\mathbf{R}^m$, with absolutely continuous probability distribution μ, and continuous density p w.r.t. the Lebesgue measure on $\mathbf{R}^m$. Let the mapping h be continuously differentiable from $\mathbf{R}^m$ to $\mathbf{R}^d$. Compute the conditional probability distribution

$$\mu_z(dx) = \mathbf{P}[X \in dx \mid h(X) = z] ,$$

for all $z \in \mathbf{R}^d$.

Our objective is to provide an explicit expression for the conditional probability distribution μ_z. In particular, assuming that a canonical measure λ_z exists on the level set M_z, we want to give conditions under which μ_z is absolutely continuous w.r.t. λ_z, and to provide an explicit expression for the density.

2 DIRECT APPROACH FOR REGULAR VALUES

We begin this section with some definitions, and properties of the level sets associated with the continuously differentiable mapping h from $\mathbf{R}^m$ to $\mathbf{R}^d$. For all $z \in \mathbf{R}^d$, the level set M_z is defined by

$$M_z = h^{-1}(z) = \{x \in \mathbf{R}^m \, : \, h(x) = z\} .$$

Notice that for all $x \in \mathbf{R}^m$, $h'(x)$ is a linear mapping from $\mathbf{R}^m$ to $\mathbf{R}^d$, i.e. a $d \times m$ matrix. By definition, $x \in \mathbf{R}^m$ is a *regular point* iff the matrix $h'(x)$ has full rank $\min(m, d)$. We distinguish two cases :

- In the case $m \leq d$, $x \in \mathbf{R}^m$ is a *regular point* iff $h'(x)$ has rank m, i.e.

$$J_h(x) \triangleq \sqrt{\det \left([h'(x)]^* \, h'(x) \right)} > 0 .$$

- In the case $m \geq d$, $x \in \mathbf{R}^m$ is a *regular point* iff $h'(x)$ has rank d, i.e.

$$J_h(x) \triangleq \sqrt{\det \left(h'(x) \, [h'(x)]^* \right)} > 0 .$$

Remark 2.1 If the mapping h is only Lipschitz continuous, it follows from the theorem of Rademacher, see Federer [6] or Evans and Gariepy [5, Chapter 3], that h is differentiable a.e., and the Jacobian $J_h(x)$ is defined as above for a.e. $x \in \mathbf{R}^m$.

By definition, $z \in \mathbf{R}^d$ is a *regular value* iff all $x \in M_z$ are regular points. We distinguish again the same two cases :

- In the case $m \leq d$, if $z \in \mathbf{R}^d$ is a *regular value*, then M_z is a collection of isolated points, i.e. a 0–dimensional submanifold of $\mathbf{R}^m$, and the canonical measure λ_z on M_z is the counting measure.
- In the case $m \geq d$, if $z \in \mathbf{R}^d$ is a *regular value*, then M_z is a $(m-d)$–dimensional submanifold of $\mathbf{R}^m$, and the canonical measure λ_z on M_z is the *Lebesgue* measure on M_z, see Berger and Gostiaux [1, Chapter 6].

Under additional *smoothness* assumption on the mapping h, it follows from the theorem of Sard that the set of regular values has full Lebesgue measure on $\mathbf{R}^d$, see Berger and Gostiaux [1, Chapter 4] or Sternberg [10]. We are therefore focusing our attention below on regular values $z \in \mathbf{R}^d$. We first recall the area and co–area formula, see Federer [6] or Evans and Gariepy [5, Chapter 3].

Theorem 2.2 (Area formula) *Let the mapping h be Lipschitz continuous from $\mathbf{R}^m$ to $\mathbf{R}^d$, and assume $m \leq d$. Then for every $g \in L^1(\mathbf{R}^m)$*

$$\int_{\mathbf{R}^m} g(x) \, J_h(x) \, dx = \int_{\mathbf{R}^d} \{ \sum_{x \in M_z} g(x) \} \, \mathcal{H}^m(dz) ,$$

where $\mathcal{H}^m$ denotes the m–dimensional Hausdorff measure on $\mathbf{R}^d$.

Theorem 2.3 (Co–area formula) *Let the mapping h be Lipschitz continuous from $\mathbf{R}^m$ to $\mathbf{R}^d$, and assume $m \geq d$. Then for every $g \in L^1(\mathbf{R}^m)$, the restriction of g to the level set M_z is $\mathcal{H}^{m-d}$–integrable for almost all $z \in \mathbf{R}^d$, and*

$$\int_{\mathbf{R}^m} g(x) \, J_h(x) \, dx = \int_{\mathbf{R}^d} \{ \int_{M_z} g(x) \, \mathcal{H}^{m-d}(dx) \} \, dz ,$$

where $\mathcal{H}^{m-d}$ denotes the $(m-d)$–dimensional Hausdorff measure on $\mathbf{R}^m$.

Remark 2.4 For a regular value $z \in \mathbf{R}^d$, the restriction to M_z of the $(m-d)$–dimensional Hausdorff measure agrees with the canonical measure λ_z on M_z.

Remark 2.5 If the mapping h is only locally Lipschitz continuous from $\mathbf{R}^m$ to $\mathbf{R}^d$, the formulas still hold for every $g \in L^1(\mathbf{R}^m)$ vanishing outside a compact subset of $\mathbf{R}^m$.

Remark 2.6 If the mapping h is *proper* (i.e. $h^{-1}(K) \subset \mathbf{R}^m$ is compact, whenever $K \subset \mathbf{R}^d$ is compact), then all the level sets M_z are compact, and in addition the set of regular values is open in $\mathbf{R}^d$. Indeed, assume the contrary, i.e. assume that there exists a sequence $\{z_n\,,\, n \geq 1\}$ in $\mathbf{R}^d$, converging to the regular value z. By definition, there exists a sequence $\{x_n\,,\, n \geq 1\}$ in $\mathbf{R}^m$ such that $h(x_n) = z_n$ and $J_h(x_n) = 0$ for all $n \geq 1$. The converging sequence $\{z_n\,,\, n \geq 1\}$ stays in a compact $K \subset \mathbf{R}^d$, hence the sequence $\{x_n\,,\, n \geq 1\}$ stays in $h^{-1}(K) \subset \mathbf{R}^m$, which is compact since h is proper. Any limit point x^* of the sequence $\{x_n\,,\, n \geq 1\}$ satisfies $h(x^*) = z$ and $J_h(x^*) = 0$, which contradicts the assumption that z is a regular value. This motivates the introduction of the following stronger assumption.

Assumption C :

> The mapping h is continuously differentiable and proper from $\mathbf{R}^m$ to $\mathbf{R}^d$.

Under Assumption C, we deduce from the area and co–area formula, the following characterization of the conditional probability distribution μ_z for a regular value $z \in \mathbf{R}^d$.

Theorem 2.7 *Let Assumption C hold, and assume $m \leq d$. Let $z \in \mathbf{R}^d$ be a regular value, with* finite *level set $M_z = \{x_i\,,\, i \in I\}$. The following* summability *condition holds*

$$\sum_{i \in I} \frac{p(x_i)}{J_h(x_i)} < \infty\ .$$

If in addition the density p does not identically vanish on M_z, then

$$\mu_z = \sum_{i \in I} q_i\, \delta_{x_i}\ ,$$

with density

$$q_i = \frac{\dfrac{p(x_i)}{J_h(x_i)}}{\displaystyle\sum_{j \in I} \frac{p(x_j)}{J_h(x_j)}}\ .$$

Theorem 2.8 *Let Assumption C hold, and assume $m \geq d$. Let $z \in \mathbf{R}^d$ be a regular value, and let λ_z denote the canonical measure on the* compact *level set M_z. The following* integrability *condition holds*

$$\int_{M_z} \frac{p(x)}{J_h(x)}\, \lambda_z(dx) < \infty\ .$$

If in addition the density p does not identically vanish on M_z, then

$$\mu_z(dx) = q(x)\, \lambda_z(dx)\ ,$$

with density

$$q(x) = \frac{\dfrac{p(x)}{J_h(x)}}{\displaystyle\int_{M_z} \frac{p(y)}{J_h(y)}\, \lambda_z(dy)}\ .$$

Proof. Under Assumption C, there exists a compact neighbourhood V_z of the regular value $z \in \mathbf{R}^d$, such that all values in V_z are regular. For any test function ϕ defined in $\mathbf{R}^m$, and any Borel set $B \subset V_z$ in $\mathbf{R}^d$, define

$$g(x) = \phi(x)\, \mathbf{1}_{(h(x) \in B)}\, \frac{p(x)}{J_h(x)}\ .$$

This mapping is Lebesgue–integrable and vanishes outside $h^{-1}(V_z)$, which is a compact subset of $\mathbf{R}^m$. Therefore, it follows from Theorem 2.3 and Remark 2.5 that

$$\begin{aligned} &\mathbf{E}[\phi(X)\, \mathbf{1}_{(h(X) \in B)}] = \\ &= \int_{\mathbf{R}^m} g(x)\, J_h(x)\, dx \\ &= \int_{\mathbf{R}^d} \Big\{ \int_{M_{z'}} g(x)\, \mathcal{H}^{m-d}(dx) \Big\}\, dz' \\ &= \int_B \Big\{ \int_{M_{z'}} \phi(x)\, \frac{p(x)}{J_h(x)}\, \mathcal{H}^{m-d}(dx) \Big\}\, dz' \\ &= \int_B \Big\{ \int_{M_{z'}} \phi(x)\, \frac{p(x)}{J_h(x)}\, \lambda_{z'}(dx) \Big\}\, dz'\ . \end{aligned}$$

By taking $\phi(x) \equiv 1$, we get

$$\mathbf{P}[h(X) \in B] = \int_B \Big\{ \int_{M_{z'}} \frac{p(x)}{J_h(x)}\, \lambda_{z'}(dx) \Big\}\, dz'\ .$$

The result follows from the definition of $\mathbf{E}[\phi(X) \mid h(X) = z]$, see e.g. Breiman [2]. □

3 ASYMPTOTICS OF LAPLACE INTEGRALS

To obtain an expression of the conditional probability distribution μ_z for any (regular or singular) value $z \in \mathbf{R}^d$, and motivated by the design of robust and efficient numerical approximation schemes in nonlinear filtering with small observation noise, we introduce the following perturbation procedure : For $\varepsilon > 0$, define

$$Z^\varepsilon = Z + V^\varepsilon = h(X) + V^\varepsilon\ ,$$

where V^ε is a d–dimensional Gaussian r.v. with covariance matrix $\varepsilon\, \mathrm{I}_d$, independent of the r.v. X. Using the Bayes formula, it is easy to define a regular conditional probability distribution for the r.v. X given Z^ε : For any test function ϕ defined in $\mathbf{R}^m$

$$\mathbf{E}[\phi(X) \mid Z^\varepsilon] = $$

$$= \frac{\displaystyle\int_{\mathbf{R}^m} \phi(x)\, \exp\Big\{ -\frac{1}{2\varepsilon^2} |Z^\varepsilon - h(x)|^2 \Big\}\, p(x)\, dx}{\displaystyle\int_{\mathbf{R}^m} \exp\Big\{ -\frac{1}{2\varepsilon^2} |Z^\varepsilon - h(x)|^2 \Big\}\, p(x)\, dx}\ .$$

Two questions arise at this point :

- Does the left–hand side have a limit as $\varepsilon \downarrow 0$, and does this limit provide a version of the conditional probability distribution of the r.v. X given Z, i.e. given $h(X)$?
- How can we compute the limit of the right–hand side as $\varepsilon \downarrow 0$?

The answer to the first question is provided by the following result.

Proposition 3.1 *For any test function ϕ defined in $\mathbf{R}^m$*

$$\mathbf{E}[\,\mathbf{E}[\phi(X) \mid Z^\varepsilon] \mid Z] \longrightarrow \mathbf{E}[\phi(X) \mid Z] \ ,$$

in L^1, as $\varepsilon \downarrow 0$.

Proof. Given a d–dimensional Gaussian white noise sequence $\{V_n \, , \, n \geq 0\}$ with unit covariance matrix, we define

$$\widetilde{Z}_n = Z + \sum_{k=n}^{\infty} \delta_k \, V_k = Z + \widetilde{V}_n \ ,$$

with $\delta_k \triangleq \varepsilon_k - \varepsilon_{k+1}$, and $\{\varepsilon_n \, , \, n \geq 0\}$ is a sequence of positive numbers decreasing to zero. It follows from Takeuchi and Akashi [11, Lemma A.1] that

$$\mathbf{E}[\phi(X) \mid \widetilde{Z}_n] \longrightarrow \mathbf{E}[\phi(X) \mid Z] \ ,$$

a.s. and in L^1, as $n \to 0$. As a result

$$\mathbf{E}[\,\mathbf{E}[\phi(X) \mid \widetilde{Z}_n] \mid Z] \longrightarrow \mathbf{E}[\phi(X) \mid Z] \ ,$$

in L^1, as $n \to 0$. Notice that $(X, \widetilde{Z}_n)$ and (X, Z^{ε_n}) are identically distributed, hence

$$\mathbf{E}[\,\mathbf{E}[\phi(X) \mid \widetilde{Z}_n] \mid Z] = \mathbf{E}[\,\mathbf{E}[\phi(X) \mid Z^{\varepsilon_n}] \mid Z]$$

a.s., and and therefore

$$\mathbf{E}[\,\mathbf{E}[\phi(X) \mid Z^{\varepsilon_n}] \mid Z] \longrightarrow \mathbf{E}[\phi(X) \mid Z] \ ,$$

in L^1, as $n \to 0$. Since the convergence holds for an arbitrary sequence $\{\varepsilon_n \, , \, n \geq 0\}$ of positive numbers decreasing to zero, the result is proved. □

Corollary 3.2 *For any test function ϕ defined in $\mathbf{R}^m$*

$$\int_{\mathbf{R}^d} |\, \Phi_z^\varepsilon - \mathbf{E}[\phi(X) \mid h(X) = z] \,| \, \kappa(dz) \longrightarrow 0 \ ,$$

as $\varepsilon \downarrow 0$, where

$$\Phi_z^\varepsilon \triangleq \mathbf{E}[\,\mathbf{E}[\phi(X) \mid Z^\varepsilon] \mid h(X) = z] \ ,$$

and where κ denotes the probability distribution of the r.v. $h(X)$.

In the remaining of this section, we are going to answer the second question, i.e. to study the asymptotic behaviour of Φ_z^ε as $\varepsilon \downarrow 0$, for a given $z \in \mathbf{R}^d$. For every $x \in \mathbf{R}^m$, $v \in \mathbf{R}^d$, we define

$$\Gamma_z^\varepsilon(x,v) = \exp\{\, -\frac{1}{\varepsilon^2} H_z^\varepsilon(x,v) \,\} \; p(x) \ ,$$

with

$$H_z^\varepsilon(x,v) = \tfrac{1}{2} \, |h(x) - (z + \varepsilon \, v)|^2 \ ,$$

and we notice that

$$\Phi_z^\varepsilon = \int_{\mathbf{R}^d} \frac{\displaystyle\int_{\mathbf{R}^m} \phi(x) \; \Gamma_z^\varepsilon(x,v) \, dx}{\displaystyle\int_{\mathbf{R}^m} \Gamma_z^\varepsilon(x,v) \, dx} \; \frac{e^{-\frac{1}{2}|v|^2}}{(2\pi)^{d/2}} \, dv \ .$$

The first step is to show that the behaviour of Φ_z^ε is determined only by the points $x \in \mathbf{R}^m$ in a neighbourhood of the level set M_z. For this purpose, we introduce the following *identifiability* condition.

Assumption D :

For all $z \in \mathbf{R}^d$, $r > 0$

$$g_z(r) \triangleq \inf_{x \notin M_z^r} \; |h(x) - z| > 0 \ ,$$

where

$$M_z^r \triangleq \{x \in \mathbf{R}^m \; : \; d(x, M_z) < r\} \ .$$

Remark 3.3 Under Assumption C, this identifiability condition is always satisfied. Indeed, assume the contrary, i.e. assume that $g_z(r) = 0$ for some $z \in \mathbf{R}^d$, $r > 0$. By definition, there exists a sequence $\{x_n \, , \, n \geq 1\}$ in $\mathbf{R}^m \backslash M_z^r$ such that $h(x_n)$ converges to z as $n \to \infty$. The converging sequence $\{h(x_n) \, , \, n \geq 1\}$ stays in a compact $K \subset \mathbf{R}^d$, hence the sequence $\{x_n \, , \, n \geq 1\}$ stays in $h^{-1}(K) \subset \mathbf{R}^m$, which is compact since h is proper. Any limit point x^* of the sequence $\{x_n \, , \, n \geq 1\}$ satisfies $h(x^*) = z$ and $d(x^*, M_z) \geq r > 0$, which is a contradiction.

We define

$$D_z^{\varepsilon,r} \triangleq \{v \in \mathbf{R}^d \; : \; |\varepsilon \, v| < \tfrac{1}{2} \, g_z(r)\} \ ,$$

and we notice that

$$\Phi_z^\varepsilon = \int_{D_z^{\varepsilon,r}} \frac{\displaystyle\int_{\mathbf{R}^m} \phi(x) \; \Gamma_z^\varepsilon(x,v) \, dx}{\displaystyle\int_{\mathbf{R}^m} \Gamma_z^\varepsilon(x,v) \, dx} \; \frac{e^{-\frac{1}{2}|v|^2}}{(2\pi)^{d/2}} \, dv + R_z^{\varepsilon,r} \ ,$$

with

$$|R_z^{\varepsilon,r}| \leq \|\phi\| \; \mathbf{P}[\,|V^\varepsilon| \geq \tfrac{1}{2} \, g_z(r)] \ . \tag{2}$$

In addition

$$\int_{\mathbf{R}^m} \phi(x) \; \Gamma_z^\varepsilon(x,v) \, dx = \int_{M_z^r} \phi(x) \; \Gamma_z^\varepsilon(x,v) \, dx + \int_{\mathbf{R}^m \backslash M_z^r} \phi(x) \; \Gamma_z^\varepsilon(x,v) \, dx \ ,$$

and for any $x \notin M_z^r$, $v \in D_z^{\varepsilon,r}$

$$|h(x) - (z + \varepsilon \, v)| \geq |h(x) - z| - |\varepsilon \, v| \geq \tfrac{1}{2} \, g_z(r) \ ,$$

hence

$$\Gamma_z^\varepsilon(x,v) \leq \exp\{\, -\frac{1}{8\varepsilon^2} \, g_z^2(r) \,\} \; p(x) \ ,$$

and

$$| \int_{\mathbf{R}^m \setminus M_z^r} \phi(x)\ \Gamma_z^\varepsilon(x,v)\, dx | \leq$$
$$\leq \|\phi\|\ \exp\{ -\frac{1}{8\varepsilon^2}\, g_z^2(r) \}\ , \tag{3}$$

for any $v \in D_z^{\varepsilon,r}$. As a consequence, we concentrate below on studying the asymptotic behaviour of

$$\Phi_z^{\varepsilon,r}(v) \triangleq \frac{\varepsilon^{-d}}{(2\pi)^{d/2}} \int_{M_z^r} \phi(x)\ \Gamma_z^\varepsilon(x,v)\, dx\ ,$$

as $\varepsilon \downarrow 0$, for a given $v \in \mathbf{R}^d$. For the sake of brevity, we will limit ourselves to recover the expression given in the Theorem 2.8 above for a regular value $z \in \mathbf{R}^d$, in the case $m \geq d$, following the method used in Ellis and Rosen [3], [4] and Hwang [7].

Proposition 3.4 *Let Assumption C hold, and assume $m \geq d$. Let $z \in \mathbf{R}^d$ be a regular value, and let λ_z denote the canonical measure on the compact $(m-d)$–dimensional submanifold M_z. Then, for any fixed $v \in \mathbf{R}^d$, and any $r > 0$ small enough*

$$\Phi_z^{\varepsilon,r}(v) \longrightarrow \int_{M_z} \phi(x)\, \frac{p(x)}{J_h(x)}\, \lambda_z(dx)\ ,$$

as $\varepsilon \downarrow 0$.

PROOF. We prove the result under the additional assumption that the mapping h has a continuous second order derivative. For any $x \in M_z$, let $N_x M_z$ denote the normal space to the submanifold M_z at point x. This is a d–dimensional affine subspace of $\mathbf{R}^m$, isomorphic to $\mathbf{R}^d$. Define also

$$N_x^r M_z \triangleq \{\nu \in N_x M_z\ :\ |\nu| < r\}\ ,$$

$$N^r M_z \triangleq \{(x,\nu) \in M_z \times \mathbf{R}^m\ :\ \nu \in N_x^r M_z\}\ .$$

Since M_z is compact, we know that for $r > 0$ small enough, the canonical mapping

$$(x,\nu) \in N^r M_z \mapsto x + \nu \in M_z^r\ ,$$

is a diffeomorphism, see Berger and Gostiaux [1, Chapter 2]. Moreover, the Jacobian of this diffeomorphism is a continuous mapping $G_z(x,\nu)$, with $G_z(x,0) = 1$, for all $x \in M_z$. The change of variable formula gives

$$\Phi_z^{\varepsilon,r}(v) = \int_{M_z} \{ \frac{\varepsilon^{-d}}{(2\pi)^{d/2}} \int_{N_x^r M_z} \phi(x+\nu)$$
$$\Gamma_z^\varepsilon(x+\nu,v)\ G_z(x,\nu)\ \rho_{z,x}(d\nu)\ \}\ \lambda_z(dx)\ ,$$

where $\rho_{z,x}$ is the canonical measure on the d–dimensional linear space $N_x M_z$. To evaluate the inner integral

$$\Psi_z^{\varepsilon,r}(x,v) = \frac{1}{(2\pi)^{d/2}} \int_{N_x^{r/\varepsilon} M_z} \phi(x+\varepsilon\,\nu)$$
$$\Gamma_z^\varepsilon(x+\varepsilon\,\nu,v)\ G_z(x,\varepsilon\,\nu)\ \rho_{z,x}(d\nu)\ ,$$

we use a first order Taylor expansion for the mapping $\nu \mapsto H_z^\varepsilon(x+\varepsilon\,\nu,v)$, where $x \in M_z$ and $v \in \mathbf{R}^d$ are fixed, i.e.

$$H_z^\varepsilon(x+\varepsilon\,\nu,v) = H_z^\varepsilon(x,v) + \varepsilon\ [H_z^\varepsilon]'(x,v)\ \nu$$
$$+\,\varepsilon^2\ \nu^*\ \{ \int_0^1 (1-\theta)\ [H_z^\varepsilon]''(x+\theta\,\varepsilon\,\nu,v)\ d\theta\ \}\ \nu\ .$$

Notice that if $x \in M_z$, i.e. if x satisfies $h(x) = z$, then for every $v \in \mathbf{R}^d$

$$H_z^\varepsilon(x,v)\ =\ \tfrac{1}{2}\,\varepsilon^2\ |v|^2\ ,$$

$$[H_z^\varepsilon]'(x,v)\ =\ -\varepsilon\, v^*\ h'(x)\ ,$$

$$[H_z^\varepsilon]''(x,v)\ =\ -\varepsilon\, v^*\ h''(x) + [h'(x)]^*\ h'(x)\ .$$

It follows from these observations that

$$\frac{1}{\varepsilon^2} H_z^\varepsilon(x+\varepsilon\,\nu,v) \longrightarrow \tfrac{1}{2}\,|v - h'(x)\,\nu|^2\ ,$$

as $\varepsilon \downarrow 0$. Therefore, by the Lebesgue dominated convergence theorem

$$\Psi_z^{\varepsilon,r}(x,v) \longrightarrow \phi(x)\ p(x)\ I_z(x,v)\ ,$$

as $\varepsilon \downarrow 0$, where the integral $I_z(x,v)$ is defined as

$$\frac{1}{(2\pi)^{d/2}} \int_{N_x M_z} \exp\{ -\tfrac{1}{2}\,|v - h'(x)\,\nu|^2 \}\ \rho_{z,x}(d\nu)\ .$$

To evaluate this integral we select an orthonormal basis $\{u_1,\cdots,u_d\}$ of $N_x M_z$, and we define the linear mapping

$$U\ :\ \alpha \in \mathbf{R}^d \mapsto \sum_{k=1}^d \alpha_k\, u_k \in N_x M_z \simeq \mathbf{R}^d\ .$$

This mapping is *orthogonal*, i.e.

$$U^*\,U = \mathrm{I}_{\mathbf{R}^d} \qquad \text{and} \qquad U\,U^* = \mathrm{I}_{\mathcal{R}(U)}\ ,$$

and the change of variable formula gives

$$I_z(x,v) = \frac{1}{\sqrt{\det \Sigma(x)}}\ ,$$

where

$$\Sigma(x) = [h'(x)\,U]^*\ h'(x)\,U\ .$$

Since $h'(x)\,U$ is a square $d \times d$ matrix, we obtain

$$\det \Sigma(x) = \det\,(\ h'(x)\,U\ [h'(x)\,U]^*\ )\ .$$

Notice that the rows of $h'(x)$, i.e. the columns of $[h'(x)]^*$, are d linearly independent m–dimensional vectors in $N_x M_z = \mathcal{R}(U)$. It follows that, for any $\alpha, \beta \in \mathbf{R}^d$

$$\beta^*\ h'(x)\ U\,U^*\ [h'(x)]^*\ \alpha = \beta^*\ h'(x)\ [h'(x)]^*\ \alpha\ ,$$

since $[h'(x)]^*\ \alpha \in \mathcal{R}(U)$. Therefore

$$\sqrt{\det \Sigma(x)} = J_h(x)\ ,$$

which finishes the proof. □

Collecting this result with the estimates (2) and (3), we obtain the following limit, i.e. we recover the expression given in the Theorem 2.8 above.

Theorem 3.5 *Let Assumption C hold, and assume $m \geq d$. Let $z \in \mathbf{R}^d$ be a regular value, and let λ_z denote the canonical measure on the compact $(m-d)$-dimensional submanifold M_z. Then*

$$\Phi_z^\varepsilon \longrightarrow \frac{\displaystyle\int_{M_z} \phi(x)\, \frac{p(x)}{J_h(x)}\, \lambda_z(dx)}{\displaystyle\int_{M_z} \frac{p(x)}{J_h(x)}\, \lambda_z(dx)} ,$$

as $\varepsilon \downarrow 0$.

4 APPLICATION TO HYBRID SYSTEMS

With the approach presented in this paper, it is possible to address the following estimation problems, which for the sake of simplicity are stated in the static case, i.e. for finite–dimensional r.v.'s rather than for random sequences.

Let X be a r.v. in $\mathbf{R}^m$, with absolutely continuous probability distribution μ, and continuous density p w.r.t. the Lebesgue measure on $\mathbf{R}^m$.

- Assume that we learn that X satisfies the constraint
  $$0 = h(X) ,$$
  where $x \longmapsto h(x)$ is a proper continuously differential mapping from $\mathbf{R}^m$ to $\mathbf{R}^d$.

  If 0 is a regular value of the mapping $x \longmapsto h(x)$, then the conditional probability distribution $P[X \in dx \mid h(X) = 0]$ can be computed explicitly, and is supported by $M = \{x \in \mathbf{R}^m \, : \, h(x) = 0\}$.

- Assume that we learn that X satisfies instead the constraint
  $$0 = g(X, Y) ,$$
  where Y is some *observed deterministic* variable in $\mathbf{R}^p$, e.g. input, parameter, etc., and where $x \longmapsto g(x, y)$ is a proper and continuously differential mapping from $\mathbf{R}^m$ to $\mathbf{R}^d$, for any $y \in \mathbf{R}^p$.

  If we observe $Y = y$, and if 0 is a regular value of the mapping $x \longmapsto g_y(x) = g(x, y)$, then the conditional probability distribution $P[X \in dx \mid g_y(X) = 0]$ can be computed explicitly, and is supported by $M_y = \{x \in \mathbf{R}^m \, : \, g_y(x) = 0\}$.

5 EXTENSIONS

In the case of a singular value $z \in \mathbf{R}^d$, some partial results could be obtained in the case $m \leq d$, using ideas contained in Ellis and Rosen [3]. Only the most singular points will contribute to the conditional probability distribution μ_z.

The case of observations in continuous time should also be considered. A possible approach would consist to sample the observations and take the limit as the sampling rate goes to infinity, but this has not been addressed yet.

Let us finally mention another related problem, arising in parameter estimation for a diffusion process with small noise. In the case where the standard identifiability condition does not hold, i.e. where the corresponding Kullback–Leibler information does not have a *unique* minimum, the consistency of the Bayesian estimate could be obtained as an application of the asymptotics of Laplace integrals. Notice that the case of a finite number of minima of the contrast function has already been considered in Kutoyants [9].

ACKNOWLEDGEMENT

The authors gratefully acknowledge Matthew James for bringing the papers [4] and [3] to their attention.

6 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] L. BREIMAN. *Probability.* Addison–Wesley, Reading, MA, 1968.

[3] 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.

[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] L.C. EVANS and R.F. GARIEPY. *Measure Theory and Fine Properties of Functions.* CRC Press, Boca Raton, FL, 1992.

[6] H. FEDERER. *Geometric Measure Theory.* Volume 153 of *Grundlehren der mathematischen Wissenschaften*, Springer Verlag, Berlin, 1969.

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

[8] H. KOREZLIOGLU and W.J. RUNGGALDIER. Filtering for nonlinear systems driven by nonwhite noises : an approximation scheme. *Stochastics and Stochastics Reports*, 44:65–102, 1993.

[9] Yu.A. KUTOYANTS. Nonconsistent estimation by diffusion type observations. *Statistics & Probability Letters*, 20(1):1–7, May 1994.

[10] Sh. STERNBERG. *Lectures on Differential Geometry.* Prentice Hall, Englewood Cliffs, NJ, 1964.

[11] Y. TAKEUCHI and H. AKASHI. Least–squares state estimation of systems with state–dependent observation noise. *Automatica*, 21(3):303–313, 1985.