MULTIDIMENSIONAL LAPLACE FORMULAS FOR NONLINEAR BAYESIAN ESTIMATION

Paul BUI QUANG[1,2], Christian MUSSO[1], François LE GLAND[3]

[1]Onera – The French Aerospace Lab, Palaiseau, France
[2]Université de Rennes 1, Rennes, France
[3]Inria, Rennes Bretagne–Atlantique research centre, Rennes, France
{paul.bui_quang,christian.musso}@onera.fr, francois.le_gland@inria.fr

ABSTRACT

The Laplace method and Monte Carlo methods are techniques to approximate integrals which are useful in nonlinear Bayesian computation. When the model is one–dimensional, Laplace formulas to compute posterior expectations and variances have been proposed by Tierney, Kass and Kadane (1989). We provide in this article formulas for the multidimensional case. We demonstrate the accuracy of these formulas and show how to use them in importance sampling to design an importance density function which reduces the Monte Carlo error.

Index Terms— Nonlinear Bayesian estimation, Laplace method, Monte Carlo methods, importance sampling

1. INTRODUCTION

The Laplace method and importance sampling (IS) are useful approximation techniques to estimate the hidden state of a system given nonlinear observations. The Laplace method is a deterministic method which is consistent as the data sample size goes to infinity or as the observation noise intensity goes to zero. IS is a Monte Carlo (MC) method that is consistent as the number of simulations goes to infinity.

Previous work dealing with MC and the Laplace method includes, e.g., comparison [1] and combination [2] of the Laplace method with Monte Carlo Markov chain (MCMC) algorithms.

In this article, we compare the Laplace method and IS for nonlinear Bayesian estimation. After defining the Bayesian estimation framework we consider in section 2, we provide multidimensional formulas to compute posterior moments in section 3. These formulas have been established by Tierney et al. in [3] when the dimension of the model is one. The IS algorithm is presented in section 4. In this section, we propose to choose the importance density function by shifting and scaling the sampling density using the posterior expectation and covariance computed with the Laplace method. IS and the Laplace method are numerically compared in section 5 on a simple two-dimensional triangulation problem. Numerical experiments illustrate the accuracy of the Laplace method.

2. BAYESIAN MODEL

Consider a system with $\mathbb{R}^d$–valued hidden state X and delivering an observation Y according to the model

$$Y = H(X) + W$$

where W is a random variable independent from X. The likelihood function of the observation Y is denoted by $g(x) = \mathbb{P}[Y|X = x]$. X admits the prior probability density function (pdf) q. Bayesian estimation consists in computing the posterior pdf p of X conditionally to the observation Y. p is given by the Bayes formula

$$p(x) = \frac{g(x)q(x)}{\int_{\mathbb{R}^d} g(x)q(x)dx}. \tag{1}$$

In the sequel we assume that the posterior pdf p admits an unique global maximum $\hat{x}$, called the maximum a posteriori (MAP).

Throughout the article, the posterior expectation and covariance will respectively be denoted by

$$\bar{x} = \int_{\mathbb{R}^d} xp(x)dx \quad \text{and} \quad \bar{\Sigma} = \int_{\mathbb{R}^d} xx^T p(x)dx - \bar{x}\bar{x}^T.$$

3. LAPLACE METHOD

3.1. Laplace approximation

Let $\{h_\lambda\}_{\lambda>0}$ be a collection of functions in $C^4(\mathbb{R}^d)$. Consider the integral

$$\mathcal{I}_\lambda = \int_{\mathbb{R}^d} e^{-\lambda h_\lambda(x)}dx.$$

Theorem 3.1 (Laplace approximation)**.** *Assume that there exists λ_0 such that, for all $\lambda > \lambda_0$,*

(i) $\int_{\mathbb{R}^d} e^{-\lambda h_\lambda(x)}dx < \infty$,

(ii) h_λ has a unique global minimum $\hat{x}_\lambda$,

(iii) all the k–th order partial derivatives of h_λ, for $k \in \{0, \dots, 4\}$, are bounded independently of λ.

Then, when $\lambda \to +\infty$,

$$\mathcal{I}_\lambda = (2\pi)^{d/2} e^{-\lambda h(\hat{x}_\lambda)} \det[\lambda h_\lambda''(\hat{x}_\lambda)]^{-1/2} \{1 + O(\lambda^{-1})\}.$$

An important property is that the error made by this approximation is a relative error. The proof of theorem 3.1 can be found in [4].

3.2. Bayesian application

Consider the Bayesian model defined in section 2. Let $J(x)$ be the observed information matrix ($d \times d$), $J(x) = -(\log g)''(x) - (\log q)''(x)$, and $\hat{x} = \arg\max_{x \in \mathbb{R}^d}\{g(x)q(x)\}$ be the MAP.

Let $\phi \in C^4(\mathbb{R}^d)$ be a strictly positive function. Suppose that $H \in C^4(\mathbb{R}^d)$ and $\log q \in C^4(\mathbb{R}^d)$. Suppose the Bayesian model defined by g and q is identifiable. Then, theorem 3.1 can be used to compute the conditional expectation

$$\mathbb{E}[\phi(X)|Y] = \frac{\int_{\mathbb{R}^d} \phi(x)g(x)q(x)dx}{\int_{\mathbb{R}^d} g(x)q(x)dx}.$$

The Laplace method is applied on the numerator and the denominator. It yields the approximation

$$\frac{\phi(x^*)g(x^*)q(x^*)}{g(\hat{x})q(\hat{x})}$$
$$\times \left(\frac{\det[-(\log g)''(\hat{x})-(\log q)''(\hat{x})]}{\det[-(\log g)''(x^*)-(\log q)''(x^*)-(\log \phi)''(x^*)]} \right)^{1/2}$$

where $x^* = \arg\max_{x \in \mathbb{R}^d}\{\phi(x)g(x)q(x)\}$, which is called fully exponential Laplace approximation [3].

Let M be the moment–generating function (mgf) associated with the posterior distribution. It is defined by

$$M(a) = \mathbb{E}[e^{a^T X}|Y], \quad a \in \mathbb{R}^d,$$

and it verifies $\mathbb{E}[X|Y] = M'(0)^T$ and $\mathbb{V}[X|Y] = M''(0) - M'(0)^T M'(0)$. The fully exponential Laplace approximation of $M(a)$ is

$$M^L(a) = \frac{e^{a^T \tilde{x}}g(\tilde{x})q(\tilde{x})}{g(\hat{x})q(\hat{x})}$$
$$\times \left(\frac{\det[-(\log g)''(\hat{x})-(\log q)''(\hat{x})]}{\det[-(\log g)''(\tilde{x})-(\log q)''(\tilde{x})]} \right)^{1/2} \quad (2)$$

where

$$\tilde{x} = \tilde{x}(a) = \arg\max_{x \in \mathbb{R}^d}\{e^{a^T x}g(x)q(x)\}. \quad (3)$$

In [3], Tierney et al. propose to approximate the posterior expectation and variance by differentiating the Laplace approximation of the mgf. They prove the consistency of the approximations obtained after differentiation as the noise standard–deviation σ of $Y \,|\, X = x$ goes to 0 or, equivalently, when the number of observations (i.e., the dimension of Y) goes to infinity.

Theorem 3.2 (Tierney, Kass, Kadane (1989)).

$$\mathbb{E}[X|Y] = (M^L)'(0)^T + O(\sigma^4)$$

and

$$\mathbb{V}[X|Y] = (M^L)''(0) - (M^L)'(0)^T(M^L)'(0) + O(\sigma^6).$$

Thereafter we derive approximate multidimensional formulas for the posterior moments: $\bar{x}^L \approx \mathbb{E}[X|Y]$ and $\bar{\Sigma}^L \approx \mathbb{V}[X|Y]$.

Theorem 3.3. *Let $\bar{x}^L = (M^L)'(0)^T$ and $\bar{\Sigma}^L = (M^L)''(0) - (M^L)'(0)^T(M^L)'(0)$. Then,*

$$\bar{x}^L = \hat{x} - \frac{1}{2}\hat{J}^{-1}\hat{J}'^T \text{vec}[\hat{J}^{-1}] \quad (4)$$

and

$$\bar{\Sigma}^L = \hat{J}^{-1} + \frac{1}{2}\hat{J}^{-1}\hat{J}'^T(\hat{J}^{-1} \otimes \hat{J}^{-1})\hat{J}'\hat{J}^{-1}$$
$$+ \frac{1}{2}(I_d \otimes \text{vec}(\hat{J}^{-1})^T\hat{J}')(\hat{J}^{-1} \otimes \hat{J}^{-1})\hat{J}'\hat{J}^{-1}$$
$$- \frac{1}{2}\hat{J}^{-1}(I_d \otimes \text{vec}(\hat{J}^{-1})^T)\hat{J}''\hat{J}^{-1}. \quad (5)$$

where $\hat{J} = J(\hat{x})$ and $\hat{x}$ is the MAP.

The vec operator vectorizes a matrix by stacking its columns and $\otimes$ denotes the Kronecker product. J' and J'' denote the first derivative and the second derivative (Hessian matrix) of J, respectively. See [5] for details on matrix calculus.

Sketch of the proof. We give here a sketch of the proof in the multidimensional case for the posterior expectation (4). Details are provided in the Appendix. The matrix calculus rules used are defined in [5]. Starting from (2), we decompose the derivative of $M^L(a)$ as,

$$(M^L)'(a) = \frac{\partial M^L}{\partial \tilde{x}}\frac{d\tilde{x}}{da} + \frac{\partial M^L}{\partial a}.$$

- For all $a \in \mathbb{R}^d$, $\tilde{x}(a)$ cancels the gradient of (3): $a + (\log g)'(\tilde{x}(a)) + (\log g)'(\tilde{x}(a)) = 0$, which gives $\frac{d\tilde{x}}{da} = J(\tilde{x}(a))^{-1}$ after differentiation w.r.t. a.

- $\frac{\partial M^L}{\partial \tilde{x}}$ is computed using the matrix derivative formula $\frac{d \det J(x)}{dx_1} = (\det J(x)) \text{tr}\left[J(x)^{-1}\frac{\partial J(x)}{\partial x_1}\right]$.

- $\frac{\partial M^L}{\partial a}(a = 0) = \hat{x}$, since $\tilde{x}(0) = \hat{x}$ according to (3).

Finally we obtain the following approximation,

$$\mathbb{E}[X|Y] \approx \hat{x} - \frac{1}{2}J(\hat{x})^{-1}\begin{pmatrix} \text{tr}\left[J(\hat{x})^{-1}\frac{\partial J}{\partial x_1}(\hat{x})\right] \\ \vdots \\ \text{tr}\left[J(\hat{x})^{-1}\frac{\partial J}{\partial x_d}(\hat{x})\right] \end{pmatrix} \quad (6)$$

which has the compact expression (4). $\qquad \square$

Expression (6) gives the posterior expectation as a function of the MAP. The second term in (6), which involves the derivative of the information matrix J, is related to the skewness of the posterior pdf p. The formulas in Theorem 3.3 are often very accurate. For example, in a one–dimensional case, consider the gamma distribution $\gamma(k, \theta)$, with pdf $p(x) \propto x^{k-1}e^{-\frac{x}{\theta}}$. p has various shapes when (k, θ) varies. Applying (4) and (5) with $J(x) = \frac{k-1}{x^2}$ and $\hat{x} = (k-1)\theta$ yields: $\bar{x}^L = \hat{x} - \frac{1}{2}\hat{J}^{-2}\hat{J}' = k\theta$ and $\bar{\Sigma}^L = \frac{1}{2}\hat{J}^{-4}[2\hat{J}^3 + 2\hat{J}'^2 - \hat{J}\hat{J}''] = k\theta^2$. These are the exact values of the expectation and the variance of the gamma distribution.

4. IMPORTANCE SAMPLING

4.1. Basics on importance sampling

A classical MC method to approximate the posterior distribution and its moments is importance sampling. Let $\tilde{q}$ be a pdf, called importance pdf, such that $\mathrm{supp}\{p\} \subset \mathrm{supp}\{\tilde{q}\}$. IS consists in generating N random variables $X^1, \ldots, X^N$ from $\tilde{q}$, and weighting each X^i for $i = 1, \ldots, N$ by the weights,

$$w^i = \frac{g(X^i)q(X^i)}{\tilde{q}(X^i)}. \tag{7}$$

The approximation p^N of the posterior pdf p is then

$$p^N = \sum_{i=1}^{N} \bar{w}^i \delta_{X^i} \tag{8}$$

where $\bar{w}^i = \frac{w^i}{\sum_{j=1}^{N} w^j}$ is the normalized weight and δ_{X^i} is the Dirac measure centered at X^i for $i = 1, \ldots, N$. The approximations of the posterior expectation and covariance are thus respectively

$$\bar{x}^N = \sum_{i=1}^{N} \bar{w}^i X^i \quad \text{and} \quad \bar{\Sigma}^N = \sum_{i=1}^{N} \bar{w}^i (X^i - \bar{x}^N)(X^i - \bar{x}^N)^T. \tag{9}$$

The prior pdf is often chosen as an importance pdf, i.e. $q = \tilde{q}$. This is the choice we make in section 5.2. In the sequel we assume that we can sample from q.

4.2. Improved importance pdf

The choice of the importance pdf $\tilde{q}$ is important to control the MC error, namely the distance between p^N (8) and p. Precisely, the MC error is related to the asymptotic variance of the weights $\bar{w}^i$

$$V(\tilde{q}) = \frac{1}{N} \left(\frac{\int_{\mathbb{R}^d} g(x)^2 \frac{q(x)^2}{\tilde{q}(x)} dx}{(\int_{\mathbb{R}^d} g(x)q(x)dx)^2} - 1 \right),$$

see [6].

The variance $V(\tilde{q})$ is large when $\tilde{q}$ does not overlap g. It is minimal when the $\tilde{q}$ is equal to the posterior pdf p, i.e. $\tilde{q}_{\mathrm{opt}}(x) = p(x) \propto g(x)q(x)$, in which case it is 0. But of course, it is not easy to sample directly from p because its normalizing constant is unknown.

There are several strategies to design a good $\tilde{q}$, such as the cross–entropy method [7]. We propose to approach $\tilde{q}_{\mathrm{opt}}$ by shifting and rescaling the prior pdf q as

$$\tilde{q}(x) = \sqrt{\frac{\det \Sigma_q}{\det \bar{\Sigma}^L}} \, q\left(\Sigma_q^{1/2}(\bar{\Sigma}^L)^{-1/2}(x - \bar{x}^L) + \mu_q \right) \tag{10}$$

where μ_q and Σ_q are respectively the expectation and the covariance matrix associated to q. $\bar{x}^L$ and $\bar{\Sigma}^L$ are computed thanks to Theorem 3.3. Thus, the sampling pdf q is shifted and rescaled in order to get the same expectation and covariance as the Laplace approximation of the posterior expectation and covariance.

Note that it is easy to sample according to $\tilde{q}$ defined by (10), by shifting and scaling the sample $X^1, \ldots, X^N$ generated from q. The new weights are computed according to (7). With this choice of importance pdf, the MC error can be drastically reduced as illustrated in section 5.

As a perspective, we can use such a shifted–rescaled importance pdf in sequential Monte Carlo algorithms, also called particle filters (PF) (see [8] for a tutorial on PF). At each correction step, the particles are sampled from (10), with the empirical mean and covariance of the predictive distribution in place of μ_q and Σ_q. However, in PF, we cannot evaluate the function q (because the pdf of the predictive distribution is unknown) to compute the weights (7), we can only sample from it. The simplest solution consists in assuming, for the computation of the weights only, that q is a Gaussian pdf. Some preliminary results confirm the performance of this solution.

5. ILLUSTRATION EXAMPLES

5.1. Optimal IS on a simple example

Consider the one-dimensional posterior pdf $p(x) \propto x^2 e^{-\frac{x^2}{2\sigma_q^2}}$. We set $q(x) = \frac{1}{\sqrt{2\pi}\sigma_q} e^{-\frac{x^2}{2\sigma_q^2}}$ and $g(x) = x^2$ in (1). Consider the shifted and rescaled importance pdf $\tilde{q}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. The optimal shifting and rescaling parameters in this case are $\mu_{\mathrm{opt}} = 0$ and $\sigma_{\mathrm{opt}} = \sqrt{3}\sigma_q$ (see figure 1), in the sense that they minimize $V(\tilde{q})$ (which is then equal to $0.5/N$). These values coincide with the posterior expectation and standard deviation.

5.2. Triangulation

Let $X = (X_1, X_2)^T$ be the unknown position of a target in the Cartesian plane. The azimuths Y_0 and Y_1 of the target
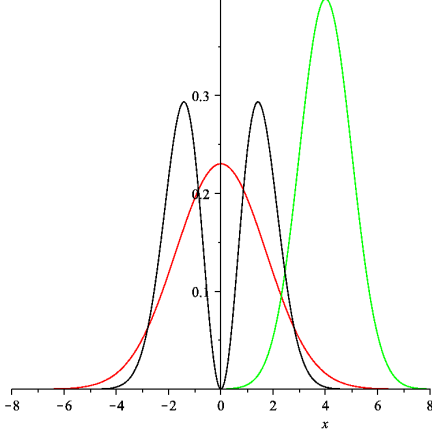
Fig. 1. Shifted and rescaled IF. p (black), $\tilde{q}$ (green), $\tilde{q}_{\mathrm{opt}}$ with μ_{opt} and σ_{opt} (red).

are measured by two passive sensors, with positions $s_0 = (s_{0,1}, s_{0,2})^T$ and $s_1 = (s_{1,1}, s_{1,2})^T$. The observation model is

$$Y = \begin{pmatrix} Y_0 \\ Y_1 \end{pmatrix} = \begin{pmatrix} \arctan \frac{X_2 - s_{0,2}}{X_1 - s_{0,1}} \\ \arctan \frac{X_2 - s_{1,2}}{X_1 - s_{1,1}} \end{pmatrix} + W,$$

with $W \sim \mathcal{N}(0, \sigma^2 I_2)$. X admits a Gaussian prior pdf with mean m_X and covariance matrix Q_X.

We aim at estimating the posterior expectation $\bar{x}$ and covariance $\bar{\Sigma}$ with IS (9) and with the Laplace method (4)–(5), and comparing their performance. The root mean squared errors (RMSE) $\mathbb{E}[\|\bar{x} - \tilde{x}\|^2]^{1/2}$ and $\mathbb{E}[\|\bar{\Sigma} - \tilde{\Sigma}\|^2]^{1/2}$ are empirically computed over 100 MC runs, for $\tilde{x} = \bar{x}^N$ and $\tilde{\Sigma} = \bar{\Sigma}^N$ (IS), and for $\tilde{x} = \bar{x}^L$ and $\tilde{\Sigma} = \bar{\Sigma}^L$ (Laplace method). We also compute the RMSE of the MAP $\hat{x}$, which is needed to compute the Laplace approximations.

The simulation parameters are $\sigma = 1°$, $s_0 = (0,0)$ (m), $s_1 = (0, 50)$ (m), $m_X = (2000, 3000)$ (m) and $Q_X = 1000^2 I_2$ (m^2). The true values $\bar{x}$ and $\bar{\Sigma}$ are computed by numerical integration. Results are shown on figure 2. For both the posterior expectation and covariance, it can be seen that a very large number of particles ($> 10^5$) is needed for the IS approximation to outperform the Laplace approximation. Besides, the "offset" observed on the RMSE of the MAP $\hat{x}$ shows that the posterior pdf p is asymmetric in this model. This asymmetry is quantified in formula (4) by the second term.

6. CONCLUSION

We provide multidimensional Laplace formulas to compute the posterior expectation and covariance matrix in a nonlinear Bayesian context. Based on this formula, a method to design an importance pdf is proposed. This density is obtained by shifting and rescaling the sampling pdf in order to reduce the Monte Carlo error. Numerical experiments on simple examples illustrate the accuracy of this method.
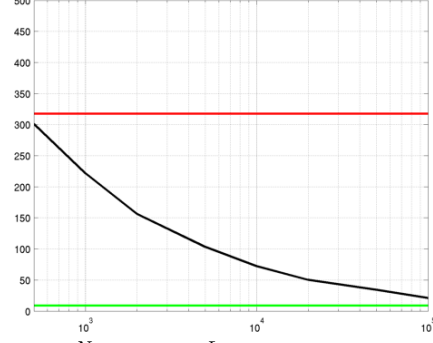
(a) RMSE of $\bar{x}^N$ (black), $\bar{x}^L$ (green) and of $\hat{x}$ (red) vs. N.
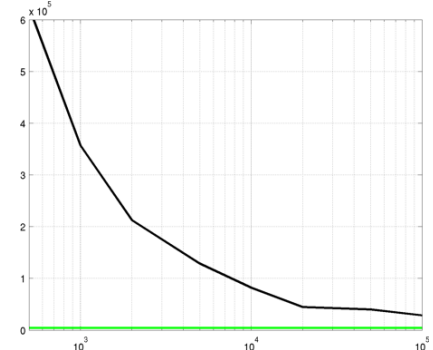
(b) RMSE of $\bar{\Sigma}^N$ (black) and of $\bar{\Sigma}^L$ (green) vs. N.

Fig. 2. The Laplace method compared to IS for an increasing number of particles.

Appendix

Proof of theorem 3.3. Let us first derive the Laplace approximation $\bar{x}^L$ of the posterior expectation. We have $(M^L)'(a) = \frac{\partial M^L}{\partial \tilde{x}} \frac{d\tilde{x}}{da} + \frac{\partial M^L}{\partial a}$ and therefore

$$\bar{x}^L = \left[\frac{\partial M^L}{\partial \tilde{x}} (\tilde{x} = \hat{x}) \frac{d\tilde{x}}{da}(a=0) + \frac{\partial M^L}{\partial a}(a=0) \right]^T.$$

First of all, $\frac{\partial M^L}{\partial a} = M^L(a)\tilde{x}(a)^T$, so that $\frac{\partial M^L}{\partial a}(a=0) = \hat{x}^T$, since $\tilde{x}(0) = \hat{x}$ according to (3). Besides,

$$\frac{\partial M^L}{\partial \tilde{x}} = \frac{\det[\hat{J}]^{1/2}}{p(\hat{x})} \frac{\partial}{\partial \tilde{x}} \left(\frac{e^{a^T \tilde{x}} p(\tilde{x})}{\det[J(\tilde{x})]^{1/2}} \right),$$

with

$$\frac{\partial}{\partial \tilde{x}} \left(\frac{e^{a^T \tilde{x}} p(\tilde{x})}{\det[J(\tilde{x})]^{1/2}} \right)$$
$$= \frac{a^T e^{a^T \tilde{x}} p(\tilde{x}) + e^{a^T \tilde{x}} p'(\tilde{x})}{\det[J(\tilde{x})]^{1/2}} + e^{a^T \tilde{x}} p(\tilde{x}) \frac{d}{d\tilde{x}} \left(\frac{1}{\det[J(\tilde{x})]^{1/2}} \right)$$
$$= \frac{a^T e^{a^T \tilde{x}} p(\tilde{x}) + e^{a^T \tilde{x}} p'(\tilde{x})}{\det[J(\tilde{x})]^{1/2}} - \frac{1}{2} \frac{e^{a^T \tilde{x}} p(\tilde{x})}{\det[J(\tilde{x})]^{1/2}} \mathrm{vec}(J(\tilde{x})^{-1})^T J'(\tilde{x})$$

so that, $\frac{\partial M^L}{\partial \tilde{x}}(\tilde{x} = \hat{x}) = -\frac{1}{2}\text{vec}(\hat{J}^{-1})^T \hat{J}'$. Moreover, (3) implies that $\varphi(a) := a + \frac{\partial \log p}{\partial x}(\tilde{x}(a)) = 0$ for all a, so that

$$\frac{d\varphi}{da} = \frac{\partial\varphi}{\partial x}(\tilde{x}(a))\frac{d\tilde{x}}{da} + \frac{\partial\varphi}{\partial a} = -J(\tilde{x}(a))\frac{d\tilde{x}}{da} + I_d = 0,$$

which gives $\frac{d\tilde{x}}{da} = J(\tilde{x}(a))^{-1}$. Finally,

$$\bar{x}^L = (M^L)'(0)^T = \hat{x} - \frac{1}{2}\hat{J}^{-1}\hat{J}'^T\text{vec}(\hat{J}^{-1}).$$

The sketch of the proof to derive the Laplace approximation $\bar{\Sigma}^L$ of the posterior covariance is given here. The detailed proof is available from the first author. We have

$$
\begin{aligned}
(M^L)''(a) &= \frac{d}{da}\left(\frac{\partial M^L}{\partial \tilde{x}}\frac{d\tilde{x}}{da}\right) + \frac{d}{da}\left(\frac{\partial M^L}{\partial a}\right) \\
&= \left[\frac{d\tilde{x}}{da}\right]^T \frac{d}{da}\left(\frac{\partial M^L}{\partial \tilde{x}}\right) + \left(I_d \otimes \frac{\partial M^L}{\partial \tilde{x}}\right)\frac{d^2\tilde{x}}{da^2} \\
&\quad + \frac{d}{da}\left(\frac{\partial M^L}{\partial a}\right),
\end{aligned}
$$

where $\frac{d}{da}\left(\frac{\partial M^L}{\partial \tilde{x}}\right) = \frac{\partial^2 M^L}{\partial \tilde{x}^2}\frac{d\tilde{x}}{da} + \frac{\partial^2 M^L}{\partial a \partial \tilde{x}}$ and $\frac{d}{da}\left(\frac{\partial M^L}{\partial a}\right) = \frac{\partial^2 M^L}{\partial \tilde{x}\partial a}\frac{d\tilde{x}}{da} + \frac{\partial^2 M^L}{\partial a^2}$, so that

$$
\begin{aligned}
(M^L)''(a) &= \left[\frac{d\tilde{x}}{da}\right]^T \frac{\partial^2 M^L}{\partial \tilde{x}^2}\frac{d\tilde{x}}{da} + \left(I_d \otimes \frac{\partial M^L}{\partial \tilde{x}}\right)\frac{d^2\tilde{x}}{da^2} \\
&\quad + \left[\frac{d\tilde{x}}{da}\right]^T \frac{\partial^2 M^L}{\partial a \partial \tilde{x}} + \frac{\partial^2 M^L}{\partial \tilde{x}\partial a}\frac{d\tilde{x}}{da} + \frac{\partial^2 M^L}{\partial a^2}.
\end{aligned}
$$

The calculation of the four terms of the sum yields:

- $\frac{d\tilde{x}}{da}(a=0)\frac{\partial^2 M^L}{\partial \tilde{x}^2}(a=0, \tilde{x}=\hat{x})\frac{d\tilde{x}}{da}(a=0)$

  $= -\hat{J}^{-1} + \frac{1}{4}\hat{J}^{-1}\hat{J}'^T\text{vec}(\hat{J}^{-1})\text{vec}(\hat{J}^{-1})^T\hat{J}'\hat{J}^{-1}$

  $\quad + \frac{1}{2}\hat{J}^{-1}\hat{J}'^T(\hat{J}^{-1} \otimes \hat{J}^{-1})\hat{J}'\hat{J}^{-1}$

  $\quad - \frac{1}{2}\hat{J}^{-1}\left(I_d \otimes \text{vec}(\hat{J}^{-1})^T\right)\hat{J}''\hat{J}^{-1},$

- $\left(I_d \otimes \frac{\partial M^L}{\partial \tilde{x}}(\tilde{x}=\hat{x})\right)\frac{d^2\tilde{x}}{da^2}(a=0)$

  $= \frac{1}{2}\left(I_d \otimes \text{vec}(\hat{J}^{-1})^T\hat{J}'\right)(\hat{J}^{-1} \otimes \hat{J}^{-1})\hat{J}'\hat{J}^{-1},$

- $\left[\frac{d\tilde{x}}{da}(a=0)\right]^T \frac{\partial^2 M^L}{\partial a \partial \tilde{x}}(a=0, \tilde{x}=\hat{x})$

  $\quad + \frac{\partial^2 M^L}{\partial \tilde{x}\partial a}(a=0, \tilde{x}=\hat{x})\frac{d\tilde{x}}{da}(a=0)$

  $= 2\hat{J}^{-1} - \frac{1}{2}(\hat{J}^{-1}\hat{J}'^T\text{vec}(\hat{J}^{-1})\hat{x}^T$

  $\quad + \hat{x}\,\text{vec}(\hat{J}^{-1})^T\hat{J}'\hat{J}^{-1}),$

- $\frac{\partial^2 M^L}{\partial a^2}(a=0) = \hat{x}\hat{x}^T.$

After summing these four terms, we obtain $(M^L)''(0)$, and $\bar{\Sigma}^L$ is given by $\bar{\Sigma}^L = (M^L)''(0) - \bar{x}^L(\bar{x}^L)^T$. $\qquad\square$

7. REFERENCES

[1] H. Rue, S. Martino, and N. Chopin, "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations," *J. of the Roy. Stat. Soc. Series B*, vol. 71, pp. 319–392, 2009.

[2] C. Guihenneuc-Jouyaux and J. Rousseau, "Laplace expansions in Markov chain Monte Carlo algorithms," *J. of Comp. and Graph. Statistics*, vol. 14, no. 1, pp. 75–94, 2005.

[3] L. Tierney, R.E. Kass, and J.B. Kadane, "Fully Exponential Laplace Approximations to Expectations and Variances of Nonpositive Functions," *J. of the Am. Stat. Assoc.*, vol. 84, no. 407, pp. 710–716, 1989.

[4] R.E. Kass, L. Tierney, and J.B. Kadane, "The validity of posterior expansions based on Laplace's method," in *Bayesian and Likelihood Methods in Statistics and Econometrics*, S. Geisser, J. S. Hodges, S. J. Press, and A. Zellner, Eds., pp. 473–488. Elsevier Science, 1990.

[5] P. L. Fackler, "Notes on Matrix Calculus," Tech. Rep., North Carolina State Univ., 2005.

[6] A. Kong, "A Note on Importance Sampling using Standardized Weights," Tech. Rep. 348, Univ. of Chicago, Dept. of Statistics, 1992.

[7] R.Y. Rubinstein and D.P. Kroese, *The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation, and machine learning*, Springer, 2004.

[8] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.