NEW ACCEPT/REJECT METHODS FOR INDEPENDENT SAMPLING FROM POSTERIOR PROBABILITY DISTRIBUTIONS

Luca Martino and Joaquín Míguez

Department of Signal Theory and Communications, Universidad Carlos III de Madrid.
Avenida de la Universidad 30, 28911 Leganés, Madrid, Spain.
E-mail: luca@tsc.uc3m.es, joaquin.miguez@uc3m.es

ABSTRACT

Rejection sampling (RS) is a well-known method to generate (pseudo-)random samples from arbitrary probability distributions that enjoys important applications, either by itself or as a tool in more sophisticated Monte Carlo techniques. Unfortunately, the use of RS techniques demands the calculation of tight upper bounds for the ratio of the target probability density function (pdf) over the proposal density from which candidate samples are drawn. Except for the class of log-concave target pdf's, for which an efficient algorithm exists, there are no general methods to analytically determine this bound, which has to be derived from scratch for each specific case. In this paper, we tackle the general problem of applying RS to draw from an arbitrary posterior pdf using the prior density as a proposal function. This is a scenario that appears frequently in Bayesian signal processing methods. We derive a general geometric procedure for the calculation of upper bounds that can be used with a broad class of target pdf's, including scenarios with correlated observations, multimodal and/or mixture measurement noises. We provide some simple numerical examples to illustrate the application of the proposed techniques.

1. INTRODUCTION

Bayesian methods have become very popular in signal processing during the past decades and, with them, there has been a surge of interest in the Monte Carlo techniques that are often necessary for the implementation of optimal *a posteriori* estimators [4, 3, 8, 6]. Indeed, the application of Markov Chain Monte Carlo (MCMC) [4] and particle filtering [3, 2] algorithms has become a commonplace in the current signal processing literature. However, in many problems of practical interest these techniques demand procedures for sampling from probability distributions with non-standard forms and the researcher is brought back to the consideration of fundamental simulation algorithms, such as importance sampling [1], inversion procedures [8] and the accept/reject method, also known as *rejection sampling* (RS).

The RS approach [8, Chapter 2] is a classical Monte Carlo technique for "universal sampling". It can be used to generate samples from any target probability density function (pdf), that we can evaluate up to a proportionality constant, by drawing from a possibly simpler proposal density. The sample is either accepted or rejected by an adequate test of the ratio of the two pdf's and it can be proved that accepted samples are actually distributed according to the target distribution. An important limitation of RS methods is the need to analytically establish an upper bound for the ratio of the target and proposal densities. With the exception of strictly log-concave pdf's, which can be efficiently dealt with using the adaptive rejection sampling (ARS) method of [5], there is a lack of systematic methods to obtain these upper bounds.

This work has been partially supported by *Ministerio de Ciencia e Innovación* of Spain (project MONIN, ref. TEC-2006-13514-C02-01/TCM), and Program Consolider-Ingenio 2010, ref. CSD2008-00010 COMONSENS) and *Comunidad Autónoma de Madrid* (project PROMULTIDIS-CM, ref. S-0505/TIC/0233).

In this paper, we aim at a general procedure to apply RS in scenarios where the target pdf is the posterior density of a signal of interest (SI) given a collection of observations (this statement can be readily connected to problems of distributed estimation in sensor networks) and the proposal density is the prior of the SI. In [7] we sketched a partial solution to this problem, restricted to cases where each observation is contaminated by independent additive noise with a exponential-family unimodal distribution. In the present work we extend these results to encompass a much broader class of target distributions (including those resulting from correlated noises, mixture or multi-modal marginal noise densities) and introduce a very general, yet simple, method for obtaining closed-form upper bounds derived from the solution for Gaussian target pdf's. Some extensions of the basic approach in [7] that can be used to obtain tighter bounds in specific scenarios are also proposed.

The remaining of the paper is organized as follows. The basic problem is stated in Section 2, where the RS algorithm is also briefly reviewed. Some basic definitions and assumptions are presented in Section 3. The restricted technique of [7] is concisely described in Section 4 for completeness and then we proceed to introduce more general bounding algorithms in Section 5 (for closed-form upper bounds) and Section 6 (for scenario-specific extensions of the restricted algorithm). Two illustrative examples are given in Section 7 and, finally, Section 8 is devoted to the conclusions.

2. PROBLEM STATEMENT

2.1 Signal model

Many signal processing problems involve the estimation of an unobserved SI, $\mathbf{x} \in \mathbb{R}^m$ (vectors are denoted as lower-case bold-face letters all through the paper), from a sequence of related observations. We assume an arbitrary prior probability density function[1] (pdf) for the SI, $p(\mathbf{x})$, and consider n scalar observations, $y_i \in \mathbb{R}$, $i = 1, \ldots, n$, which are obtained through nonlinear transformations of the signal $\mathbf{x}$ contaminated with additive noise. Formally, we write

$$y_1 = g_1(\mathbf{x}) + \xi_1, \ldots, y_n = g_n(\mathbf{x}) + \xi_n \tag{1}$$

where $\mathbf{y} = [y_1, \ldots, y_n]^T \in \mathbb{R}^n$ is the vector of available observations, $g_i : \mathbb{R}^m \to \mathbb{R}$, $i = 1, \ldots, n$, are nonlinearities and ξ_i are independent noise variables, possibly with different distributions for each i. Let us initially assume noise pdf's of the form

$$p(\xi_i) = k_i \exp\{-\bar{V}_i(\xi_i)\}, \qquad k_i > 0, \tag{2}$$

where k_i is a real constant and $\bar{V}(\xi_i)$ is a function, subsequently referred to as *marginal potential*. We assume that it is a real and non-negative function, $\bar{V}_i : \mathbb{R} \to [0, +\infty)$, and in general multimodal. If the noise variables are independent, the joint pdf $p(\xi_1, \xi_2, \ldots, \xi_n) =$

[1] We use $p(\cdot)$ to denote the probability density function (pdf) of a random variate, i.e., $p(x)$ denotes the pdf of x and $p(y)$ is the pdf of y, possibly different.

$\prod_{i=1}^{n} p(\xi_n)$ is easy to construct and we can define a *joint potential* $V^{(n)}: \mathbb{R}^n \to [0,+\infty)$ as

$$V^{(n)}(\xi_1,\ldots,\xi_n) \triangleq -\log\left[p(\xi_1,\ldots,\xi_n)\right] = -\sum_{i=1}^{n} \log p(\xi_n). \quad (3)$$

Substituting (2) into (3) yields

$$V^{(n)}(\xi_1,\ldots,\xi_n) = c_n + \sum_{i=1}^{n} \bar{V}_i(\xi_i) \quad (4)$$

where $c_n = -\sum_{i=1}^{n} \log k_i$ is a constant.
In subsequent sections we will be interested in a particular class of joint potential functions denoted as

$$V_l^{(n)}(\xi_1,\ldots,\xi_n) = \sum_{i=1}^{n} |\xi_i|^l, \quad 0 < l < +\infty, \quad (5)$$

where the subscript l identifies the specific member of the class. In particular, the function obtained for $l = 2$, $V_2^{(n)}(\xi_1,\ldots,\xi_n) = \sum_{i=1}^{n} |\xi_i|^2$ will be termed *quadratic potential* and it yields a Gaussian density when plugged into Eq. (2).
Let $\mathbf{g} = [g_1,\ldots,g_n]^T$ be the vector-valued nonlinearity defined as $\mathbf{g}(\mathbf{x}) \triangleq [g_1(\mathbf{x}),\ldots,g_n(\mathbf{x})]^T$. The scalar observations are conditionally independent given the SI $\mathbf{x}$, hence the *likelihood function*, $\ell(\mathbf{x};\mathbf{y},\mathbf{g}) \triangleq p(\mathbf{y}|\mathbf{x})$, can be factorized as

$$\ell(\mathbf{x};\mathbf{y},\mathbf{g}) = \prod_{i=1}^{n} p(y_i|\mathbf{x}). \quad (6)$$

Since we are assuming additive noises, $p(y_i|\mathbf{x}) = k_i \exp\{-\bar{V}_i(y_i - g_i(\mathbf{x}))\}$ and the likelihood in (6) induces a *system potential* $V(\mathbf{x};\mathbf{y},\mathbf{g}): \mathbb{R}^m \to [0,+\infty)$, defined as

$$V(\mathbf{x};\mathbf{y},\mathbf{g}) \triangleq -\ln[\ell(\mathbf{x};\mathbf{y},\mathbf{g})] = -\sum_{i=1}^{n} \log[p(y_i|\mathbf{x})], \quad (7)$$

that is a function of $\mathbf{x}$ and depends on the observations $\mathbf{y}$ and the function $\mathbf{g}$. Using (4) and (7), we write the system potential in terms of the joint potential,

$$V(\mathbf{x};\mathbf{y},\mathbf{g}) = V^{(n)}(y_1 - g_1(\mathbf{x}),\ldots,y_n - g_n(\mathbf{x})), \quad (8)$$

i.e., $V(\mathbf{x};\mathbf{y},\mathbf{g}) = c_n + \sum_{i=1}^{n} \bar{V}_i(y_i - g_i(\mathbf{x}))$.

2.2 Rejection sampling

Let us now assume that we wish to approximate, by sampling, some integral of the form $I(f) = \int_{\mathbb{R}} f(\mathbf{x})p(\mathbf{x}|\mathbf{y})d\mathbf{x}$, where f is some measurable function of $\mathbf{x}$ and $p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{x})\ell(\mathbf{x};\mathbf{y},\mathbf{g})$ is the posterior pdf of the SI given the observations. Unfortunately, it may not be possible in general to draw directly from $p(\mathbf{x}|\mathbf{y})$ and we must apply simulation techniques to generate adequate samples. One appealing possibility is to carry out rejection sampling using the prior, $p(\mathbf{x})$, as a proposal function. If we let L be an upper bound for the likelihood, $\ell(\mathbf{x};\mathbf{y},\mathbf{g}) \leq L$, then we can generate N independent samples from $p(\mathbf{x}|\mathbf{y})$ according to the following algorithm.

1. Set $i = 1$.
2. Draw $\mathbf{x}' \sim p(\mathbf{x})$ and $u' \sim U(0,1)$, where $U(0,1)$ is the uniform pdf in $[0,1]$.
3. If $\frac{\ell(\mathbf{x}';\mathbf{y},\mathbf{g})}{L} > u'$ then $\mathbf{x}_i = \mathbf{x}'$, else discard $\mathbf{x}'$ and go to step 2.
4. Set $i = i+1$. If $i > N$ stop, else go back to step 2.

We can approximate $I(f) \approx \hat{I}(f) = \frac{1}{N}\sum_{i=1}^{N} f(\mathbf{x}_i)$.
In the sequel, we address the problem of analytically calculating the bound L. Note that, since the log function is monotonous, it is equivalent to maximize ℓ w.r.t. $\mathbf{x}$ and to minimize the system potential V also w.r.t. $\mathbf{x}$. As a consequence, we may focus on the calculation of a lower bound for $V(\mathbf{x};\mathbf{y},\mathbf{g})$. Note that this problem is far from trivial. Even for very simple marginal potentials, $\bar{V}_i$, the system potential can be highly multimodal w.r.t. $\mathbf{x}$ [7].

3. DEFINITIONS AND ASSUMPTIONS

Hereafter we restrict our attention to the case of a scalar SI, $x \in \mathbb{R}$. This is done for the sake of clarity, since dealing with the general case $\mathbf{x} \in \mathbb{R}^m$ requires additional definitions and notations. The techniques to be described in Section 4.1 can be extended to the general case, although this extension is not trivial. We define the set of *state predictions* as

$$\mathscr{X} \triangleq \{x_i \in \mathbb{R}: \ y_i = g_i(x_i) \ \text{ for } i = 1,\ldots,n\} \quad (9)$$

Each equation $y_i = g_i(x_i)$, in general, can yield zero, one or several state predictions. We also introduce the maximum likelihood (ML) state estimator $\hat{x}$, as

$$\hat{x} \in \arg\max_{x\in\mathbb{R}} \ell(x|\mathbf{y},\mathbf{g}) = \arg\min_{x\in\mathbb{R}} V(x;\mathbf{y},\mathbf{g}), \quad (10)$$

not necessarily unique.
Let us use $\mathscr{A} \subseteq \mathbb{R}$ to denote the support of the vector function $\mathbf{g}$, i.e., $\mathbf{g}: \mathscr{A} \subseteq \mathbb{R} \to \mathbb{R}^n$. We assume that there exists a partition $\{\mathscr{B}_j\}_{j=1}^{q}$ of $\mathscr{A}$ (i.e., $\mathscr{A} = \cup_{j=1}^{q} \mathscr{B}_j$ and $B_i \cap \mathscr{B}_j = \emptyset$, $\forall i \neq j$) such that we can define functions $g_{i,j}: \mathscr{B}_j \to \mathbb{R}$, $j = 1,\ldots,q$ and $i = 1,\ldots,n$, as

$$g_{i,j}(x) \triangleq g_i(x), \ \ \forall x \in \mathscr{B}_j, \quad (11)$$

and: (a) every function $g_{i,j}$ is invertible in $\mathscr{B}_j$ and (b) every function $g_{i,j}$ is either convex in $\mathscr{B}_j$ or concave in $\mathscr{B}_j$. Assumptions (a) and (b) together mean that, for every i and all $x \in \mathscr{B}_j$, the first derivative $\frac{dg_{i,j}}{dx}$ is either strictly positive or strictly negative and the second derivative $\frac{d^2 g_{i,j}}{dx^2}$ is either non-negative or non-positive. As a consequence, there are exactly n state predictions in each subset of the partition, $x_{i,j} = g_{i,j}^{-1}(y_i)$. We write the set of predictions in $\mathscr{B}_j$ as $\mathscr{X}_j = \{x_{1,j},\ldots,x_{n,j}\}$. If $g_{i,j}$ is bounded and y_i is noisy, it is conceivable that $y_i > \max_{x\in[\mathscr{B}_j]} g_{i,j}(x)$ (or $y_i < \min_{x\in[\mathscr{B}_j]} g_{i,j}(x)$), where $[\mathscr{B}_j]$ denotes the closure of set $\mathscr{B}_j$, hence $g_{i,j}^{-1}(y_i)$ may not exist. In such case, we define $x_{i,j} = \arg\max_{x\in[\mathscr{B}_j]} g_{i,j}(x)$ (or $x_{i,j} = \arg\min_{x\in[\mathscr{B}_j]} g_{i,j}(x)$, respectively), and admit $x_{i,j} = +\infty$ (respectively, $x_{i,j} = -\infty$) as valid solutions.

4. BACKGROUND

We review the basic bounding algorithm of [7]. This technique is valid only if we impose an important additional constraints on the model of Section 2, namely that each marginal potential $\bar{V}_i(\xi_i)$ has a unique minimum at $\xi_i = 0$ and $\frac{d\bar{V}_i}{d\xi_i} \neq 0$ for all $\xi_i \neq 0$.

4.1 Basic algorithm for the calculation of bounds

Our goal is to obtain an analytical method for the computation of a scalar $\gamma \in \mathbb{R}$ such that $\gamma \leq \inf_{x\in\mathbb{R}} V(x;\mathbf{y},\mathbf{g})$ for arbitrary (but fixed) observations $\mathbf{y}$ and known nonlinearities $\mathbf{g}$. The main difficulty to carry out this calculation is the nonlinearity $\mathbf{g}$, which renders the problem not directly tractable. In [7] it is described how to build, within each set $\mathscr{B}_j$ ($\mathscr{A} = \cup_{j=1}^{q} \mathscr{B}_j$), adequate linear functions $\{r_{i,j}\}_{i=1}^{n}$ in order to replace the nonlinearities $\{g_{i,j}\}_{i=1}^{n}$. We construct every $r_{i,j}$ in a way that ensures

$$\bar{V}_i(y_i - r_{i,j}(x)) \leq \bar{V}_i(y_i - g_{i,j}(x)), \quad \forall x \in \mathscr{I}_j, \quad (12)$$

where $\mathscr{I}_j$ is any closed interval in $\mathscr{B}_j$ such that $\hat{x}_j \in \arg\min_{x\in[\mathscr{B}_j]} V(x;\mathbf{y},\mathbf{g})$ (i.e., a ML state estimator of x restricted to $\mathscr{B}_j$, possibly non-unique) is contained in $\mathscr{I}_j$. The latter requirement can be fulfilled if we choose $\mathscr{I}_j \triangleq [\min(\mathscr{X}_j), \max(\mathscr{X}_j)]$ [7].

Moreover, since $V(x;\mathbf{y},\mathbf{g}_j) = c_n + \sum_{i=1}^n \bar{V}_i(y_i - g_{i,j}(x))$ and $V(x;\mathbf{y},\mathbf{r}_j) = c_n + \sum_{i=1}^n \bar{V}_i(y_i - r_{i,j}(x))$ where $\mathbf{g}_j = [g_{1,j},\ldots,g_{n,j}]$ and $\mathbf{r}_j = [r_{1,j},\ldots,r_{n,j}]$, Eq. (12) implies that $V(x;\mathbf{y},\mathbf{r}_j) \le V(x;\mathbf{y},\mathbf{g}_j)$, $\forall x \in \mathscr{I}_j$, and as a consequence,

$$\gamma_j = \inf_{x\in\mathscr{I}_j} V(x;\mathbf{y},\mathbf{r}_j) \le \inf_{x\in\mathscr{I}_j} V(x;\mathbf{y},\mathbf{g}_j) = \inf_{x\in\mathscr{B}_j} V(x;\mathbf{y},\mathbf{g}). \quad (13)$$

Therefore, it is possible to find a lower bound in $\mathscr{B}_j$ for the system potential $V(x;\mathbf{y},\mathbf{g}_j)$, denoted γ_j, by minimizing the modified potential $V(x;\mathbf{y},\mathbf{r}_j)$ within $\mathscr{I}_j$. Repeating this procedure for every $\mathscr{B}_j$, and choosing $\gamma = \min_j \gamma_j$, we obtain that $\gamma \le \inf_{x\in\mathbb{R}} V(x,\mathbf{y},\mathbf{g})$ is a global lower bound of the system potential.

The construction of each $r_{i,j}$ is straightforward dividing the problem into two cases. Case I corresponds to nonlinearities $g_{i,j}$ such that $\frac{dg_{i,j}(x)}{dx} \times \frac{d^2 g_{i,j}(x)}{dx^2} \ge 0$, while case II corresponds to functions that comply with $\frac{dg_{i,j}(x)}{dx} \times \frac{d^2 g_{i,j}(x)}{dx^2} \le 0$, when $x \in \mathscr{B}_j$. In case I we choose a linear function $r_{i,j}$ that passes through $\min(\mathscr{X}_j)$ and the state prediction $x_{i,j} \in \mathscr{X}_j$, while in case II we choose a linear function $r_{i,j}$ that passes through $\max(\mathscr{X}_j)$ and the state prediction $x_{i,j} \in \mathscr{X}_j$.

Unfortunately, it is possible that the minimization of the modified system potential $V(x;\mathbf{y},\mathbf{r}_j)$ remains difficult and analytically intractable. We propose a general method to circumvent this limitation in Section 5. To develop this new technique, we start from the lower bound of the quadratic potential, $V_2^{(n)}$, presented next.

4.2 Lower Bound γ_2 for quadratic potentials

Assume that the joint potential is quadratic, i.e., $V_2^{(n)}(y_1 - g_{1,j}(x),\ldots,y_n - g_{n,j}(x)) = \sum_{i=1}^n (y_i - g_{i,j}(x))^2$ for each $j = 1,\ldots,q$, and construct the set of linearities $r_{i,j}(x) = a_{i,j}x + b_{i,j}$, $i = 1,\ldots,n$ and $j = 1,\ldots,q$. The modified system potential in $\mathscr{B}_j$ becomes

$$V(x;\mathbf{y},\mathbf{r}_j) = \sum_{i=1}^n (y_i - r_{i,j}(x))^2 = \sum_{i=1}^n (y_i - a_{i,j}x - b_{i,j})^2, \quad (14)$$

and it turns out straightforward to compute $\gamma_{2,j} = \min_{x\in\mathscr{B}_j} V(x;\mathbf{y},\mathbf{r}_j)$. Indeed, if we denote $\mathbf{a}_j = [a_{1,j},\ldots,a_{n,j}]^T$ and $\mathbf{w}_j = [y_1 - b_{1,j},\ldots,y_n - b_{n,j}]^T$, then

$$\hat{x}_{L,j} = \arg\min_{x\in\mathscr{B}_j} V(x;\mathbf{y},\mathbf{r}_j) = \frac{\mathbf{a}_j^T \mathbf{w}_j}{\mathbf{a}_j^T \mathbf{a}_j}, \quad (15)$$

and $\gamma_{2,j} = V(x_{L,j};\mathbf{y},\mathbf{r}_j)$. It is apparent that $\gamma_2 = \min_j \gamma_{2,j} \le V(x;\mathbf{y},\mathbf{g})$.

5. CLOSED-FORM UPPER BOUNDS

We assume the availability of an invertible increasing function $R(v) : \mathbb{R}^+ \to \mathbb{R}^+$ such that $R \circ V^{(n)} \ge V_2^{(n)}$, where $\circ$ denotes the composition of functions. Then, we can write

$$\begin{aligned}(R\circ V)(x;\mathbf{y},\mathbf{g}) &\ge V_2^{(n)}(y_1 - g_1(x),\ldots,y_n - g_n(x)) \\ &= \sum_{i=1}^n (y_i - g_i(x))^2 \ge \gamma_2.\end{aligned} \quad (16)$$

where γ_2 is the bound obtained in Section 4.2 and, as consequence, $V(x;\mathbf{y},\mathbf{g}) \ge R^{-1}(\gamma_2) = \gamma$.

For instance, consider the family of joint potentials $V_p^{(n)}$. Using the monotonicity of $\mathscr{L}^p$ norms, it is possible to prove [?] that

$$\left(\sum_{i=1}^n |\xi_i|^p\right)^{\frac{1}{p}} \ge \left(\sum_{i=1}^n \xi_i^2\right)^{\frac{1}{2}}, \text{ for } 0 \le p \le 2, \text{ and} \quad (17)$$

$$n^{\left(\frac{p-2}{2p}\right)}\left(\sum_{i=1}^n |\xi_i|^p\right)^{\frac{1}{p}} \ge \left(\sum_{i=1}^n \xi_i^2\right)^{\frac{1}{2}}, \text{ for } 2 \le p \le +\infty. \quad (18)$$

Let $R_1(v) = v^{2/p}$. Since this function is indeed strictly increasing, we can transform the inequality (17) into

$$R_1\left(\sum_{i=1}^n |y_i - g_i(x)|^p\right) \ge \sum_{i=1}^n (y_i - g_i(x))^2, \quad (19)$$

which yields

$$\begin{aligned}\sum_{i=1}^n |y_i - g_i(x)|^p &\ge R_1^{-1}\left(\sum_{i=1}^n (y_i - g_i(x))^2\right) \\ &= \left(\sum_{i=1}^n (y_i - g_i(x))^2\right)^{p/2} \ge \gamma_2^{p/2},\end{aligned} \quad (20)$$

hence the transformation $\gamma_2^{p/2}$ of the quadratic bound γ_2 is a lower bound for $V_p^{(n)}$ with $0 < p \le 2$. Similarly, if we let $R_2(v) = \left(n^{\left(\frac{p-2}{2p}\right)} v^{1/p}\right)^2$, the inequality (18) yields

$$\begin{aligned}&\sum_{i=1}^n |y_i - g_i(x)|^p \ge R_2^{-1}\left(\sum_{i=1}^n (y_i - g_i(x))^2\right) \\ &= \left[n^{\left(-\frac{p-2}{2p}\right)}\left(\sum_{i=1}^n (y_i - g_i(x))^2\right)^{1/2}\right]^p \ge n^{\left(-\frac{p-2}{2}\right)}\gamma_2^{p/2},\end{aligned} \quad (21)$$

hence the transformation $R_2^{-1}(\gamma_2) = n^{-(p-2)/2}\gamma_2^{p/2}$ is a lower bound for $V_p^{(n)}$ when $2 \le p < +\infty$.

It is possible to devise a systematic procedure to derive a suitable function R given an arbitrary joint potential $V^{(n)}(\mathbf{e})$, where $\mathbf{e} \triangleq [\xi_1,\ldots,\xi_n]^T$. Let us define the manifold $\Gamma_v \triangleq \left\{\mathbf{e} \in \mathbb{R}^n : V^{(n)}(\mathbf{e}) = v\right\}$. We can construct R by assigning $R(v)$ with the maximum of the quadratic potential $\sum_i^n \xi_i^2$ when $\mathbf{e} \in \Gamma_v$, i.e., we define

$$R(v) \triangleq \max_{\mathbf{e}\in\Gamma_v} \sum_{i=1}^n \xi_i^2. \quad (22)$$

Note that (22) is a constrained optimization problem that can be solved using, e.g., Lagrangian multipliers.

From the definition in (22) we obtain that, $\forall \mathbf{e} \in \Gamma_v$, $R(v) \ge \sum_{i=1}^n \xi_i^2$. In particular, since $V^{(n)}(\mathbf{e}) = v$ from the definition of Γ_v, we readily obtain the desired relationship,

$$R\left(V^{(n)}(\xi_1,\ldots,\xi_n)\right) \ge \sum_{i=1}^n \xi_i^2. \quad (23)$$

We additionally need to check whether R is a strictly increasing function of v. The two functions in the earlier examples of this Section, R_1 and R_2, can be readily found using this method.

Let us remark the generality of this approach. Indeed, some of the assumptions imposed on the model of Section 2 are not needed anymore. For example, the joint potential $V^{(n)}$ does not have to be defined as an addition of marginal potentials (see Eq. (4)) but may take more general forms. This technique also provides a suitable tool to compute bounds in scenarios where the noise random variables ξ_i, $i = 1,\ldots,n$, are correlated.

6. ENHANCEMENTS

The method based on the solution for quadratic potentials may admittedly lead to upper bounds which are not tight enough. In that case, the rejection rate (and, as a consequence, the average computational load) of the RS algorithm grows. In this Section we propose specific methods for the calculation of bounds that account for observation models where the observational noise components have possibly multimodal pdfs from either exponential or mixture families. The approach in Section 5 has a more general scope but the procedures in this Section can yield tighter bounds and, consequently, more efficient RS algorithms.

6.1 Multimodal exponential noise densities

Let us assume that the i-th noise variable, ξ_i, is distributed according to the pdf $p(\xi_i) = \prod_{t=1}^{T_i} p_t(\xi_i)$ where $p_t(\xi_i) \triangleq k_{i,t} \exp\left\{-\bar{V}_{i,t}(\xi_i - \mu_{i,t})\right\}$ is, itself, an exponential density with location parameter $\mu_{i,t}$ and proportionality constant $k_{i,t}$. If we assume that each $\bar{V}_{i,t}$ is a marginal potential function with a single minimum at $\mu_{i,t}$, then the noise pdf

$$p(\xi_i) = \prod_{t=1}^{T_i} p_t(\xi_i) = k_i \exp\left\{-\sum_{t=1}^{T_i} \bar{V}_{i,t}(\xi_i - \mu_{i,t})\right\}, \qquad (24)$$

is possibly multimodal and the basic algorithm in Section 4 cannot be applied directly to obtain an upper bound on the resulting likelihood function.

In order to tackle this problem, let us split the i-th observation equation, $y_i = g_i(x) + \xi_i$, in T_i equation which are (jointly) equivalent to the original one,

$$y_{i,1} - \mu_{i,1} = g_i(x) + \xi_{i,1}, \ldots, y_{i,T_i} - \mu_{i,T_i} = g_i(x) + \xi_{i,T_i}. \qquad (25)$$

It turns out possible to extend the bounding procedure of Section 4 to this scenario by introducing the extended observation vector

$$\mathbf{y}_e = [y_{1,1} - \mu_{1,1}, \ldots, y_{1,T_1} - \mu_{1,T_1}, \ldots, y_{n,T_n} - \mu_{n,T_n}] \qquad (26)$$

with dimension $T \times 1$, where $T = \sum_1^n T_i$. In a similar way, we can build up an extended vector-nonlinearity

$$\mathbf{g}_e(x) = [\underbrace{g_1(x), \ldots, g_1(x)}_{T_1}, \ldots, \underbrace{g_n(x), \ldots, g_n(x)}_{T_n}]^T, \qquad (27)$$

and an extended joint potential

$$V^{(n)}(\underbrace{\xi_1, \ldots, \xi_1}_{T_1}, \ldots, \underbrace{\xi_n, \ldots, \xi_n}_{T_n}). \qquad (28)$$

Since the system potential actually remains the same, i.e., $V(x; \mathbf{y}_e; \mathbf{g}_e) = V(x; \mathbf{y}; \mathbf{g})$, we can apply the procedure of Section 4 to compute a bound $\gamma \leq V(x; \mathbf{y}_e; \mathbf{g}_e)$ which is also a bound for $V(x; \mathbf{y}; \mathbf{g})$.

6.2 Mixture noise densities

In many application it is common to model noise random variables by means of a mixture of pdfs. In particular, let ξ_i have the density

$$p(\xi_i) = \sum_{s=1}^{S_i} \lambda_s p_s(\xi_i) \qquad (29)$$

and assume all other variables, $\xi_{j \neq i}$, have "simple" pdf's of the form (2). Again, we assume $p_s(\xi_i) = k_{i,s} \exp\left\{-\bar{V}_{i,s}(\xi_i - \mu_{i,s})\right\}$, is an exponential density with a proportionality constant $k_{i,s}$ and location parameter $\mu_{i,s}$. The mixture coefficients are normalized to yield $\sum_{s=1}^{S_i} \lambda_s = 1$ and each $\bar{V}_{i,s}$ has a unique minimum at $\mu_{i,s}$.

It is apparent that (due to the normalization of the λ_s's)

$$-\log[p(\xi_i)] = -\log\left(\sum_{s=1}^{S_i} \lambda_s p_s(\xi_i)\right) \geq \min_{s \in \{1,\ldots,S_i\}} \bar{V}_{i,1}(\xi_i - \mu_{i,s}). \qquad (30)$$

The latter equation implies that we can compute a lower bound for the system potential $V(x; \mathbf{y}, \mathbf{g})$ by exploring the S_i possible systems of the form

$$y_1 = g_1(x) + \xi_1, \ldots, y_i - \mu_{i,s} = g_i(x) + \tilde{\xi}_{i,s}, \ldots, y_n = g_n(x) + \xi_n \qquad (31)$$

where $s \in \{1, \ldots, S_i\}$ and $p_s(\tilde{\xi}_i) = k_{i,s} \exp\left\{-\bar{V}_{i,s}(\tilde{\xi}_i)\right\}$. Let γ_s be the lower bound computed for the s-th problem, then the global lower bound for the system potential $V(x; \mathbf{y}, \mathbf{g})$ is $\gamma = -\log\left(\sum_{s=1}^{S_i} \lambda_s \exp\{-\gamma_s\}\right)$

7. EXAMPLES

We show the application of the proposed bounding methods by way of two simple examples. They have been chosen for the sake of illustration only and are not intended to represent any specific practical system. In fact, we do not even claim that RS be the best way to draw from these distributions: they are simply convenient to put the proposed techniques at work.

7.1 Example 1: Independent noise variables

Given $x \in \mathbb{R}$ with prior density $p(x) = N(x; -2, 1)$ (Gaussian, with mean -2 and variance 1) and the system with $\mathbf{y} = [y_1, y_2]^\top \in \mathbb{R}^2$ given by

$$y_1 = \exp(x) + \xi_1, \; y_2 = \exp(-x) + \xi_2 \qquad (32)$$

where $p(\xi_1) \propto \exp\left\{-|\xi_1 - 1|^3 - |\xi_1 + 1|^3\right\}$ and ξ_2 is a mixture of two pdfs, $p(\xi_2) \propto 0.8 \exp\left\{-|\xi_2|^3\right\} + 0.2 \exp\left\{-|\xi_2 - 4|^3\right\}$. The resulting system potential is

$$\begin{aligned} V(x; \mathbf{y}, \mathbf{g}) =& |y_1 - 1 - \exp(x)|^3 + |y_1 + 1 - \exp(x)|^3 \\ & - \log[0.8 \exp(-|y_2 - \exp(-x)|^3) \\ & + 0.2 \exp(-|y_2 - 4 - \exp(-x)|^3)]. \end{aligned} \qquad (33)$$

Since the density of ξ_2 is a mixture, we need to apply the technique proposed in Section (6.2). Therefore, we consider two different bounding problems, one for each component in the mixture of $p(\xi_2)$. For the first problem, we use the first mixture term which yields the density $p(\xi_2) \propto \exp\left\{-|\xi_2|^3\right\}$. The resulting potential function that we wish to lower bound for this case is denoted

$$\begin{aligned} \tilde{V}(x; \mathbf{y}, \mathbf{g})_1 =& |y_1 - 1 - \exp(x)|^3 + |y_1 + 1 - \exp(x)|^3 \\ & + |y_2 - \exp(-x)|^3. \end{aligned} \qquad (34)$$

where the subscript in $\tilde{V}(\cdot)_1$, indicates that we are handling the first mixture component.

There is an additional difficulty that arises because of the form of $p(\xi_1) \propto \exp\left\{-|\xi_1 - 1|^3\right\} \exp\left\{-|\xi_1 + 1|^3\right\}$, which fits the pattern of Eq. (24). Thus, using the approach in Section 6 we build extended vectors for the observations

$$\mathbf{y}_e = [y_1 - 1, y_1 + 1, y_2]^\top, \qquad (35)$$

and the nonlinearities,

$$\mathbf{g}_e(x) = [\exp(x), \exp(x), \exp(-x)]^\top \qquad (36)$$

that yield the "extended version" of (34), denoted $\tilde{V}(x; \mathbf{y}_e, \mathbf{g}_e)_1$. The latter potential can be bounded using the method in Section 4.

To be specific, if we let, e.g., $\mathbf{y} = [1.5, 9]^\top$, the linear functions associated to the nonlinearities in $\mathbf{g}_e$ (calculated with the procedure in Section 4) are

$$\begin{aligned} r_1(x) &= 0.25x + 0.67, \\ r_2(x) &= 0.76x + 1.79, \\ r_3(x) &= -2.76x + 2.93 \end{aligned} \quad (37)$$

Note that the nonlinearities are monotonous and convex, so we have a trivial partition $\mathscr{B}_1 \equiv \mathbb{R}$ and we can skip the second subscript in the linear function $r_{i,j}(x)$. Substituting $\mathbf{g}_e$ by $\mathbf{r}_e = [r_1, r_2, r_3]^\top$ into a quadratic potential we obtain

$$\begin{aligned} \tilde{V}_2(x; \mathbf{y}_e, \mathbf{r}_e)_1 =& (y_1 - 1 - r_1(x))^2 + (y_1 + 1 - r_2(x))^2 \\ &+ (y_2 - r_3(x))^2. \end{aligned} \quad (38)$$

An analytical lower bound for (38) is easily calculated and it yields $\gamma_2^{(1)} = 5.40$. Next we can apply transformation $R_2(v)$ in Section 5 (see Eq. (21)), we arrive at a lower bound $\gamma^{(1)} = R_2^{-1}(\gamma_2^{(1)}) = 11.17 \leq \tilde{V}(x; \mathbf{y}, \mathbf{g})_1$, for the first sub-problem.

The second sub-problem arises from the second term in the mixture of $p(\xi_2)$. In this case, $p(\xi_2) \propto \exp\{-|\xi_2 - 4|^3\}$ and the potential to be bounded in this second case turns out to be

$$\begin{aligned} \tilde{V}(x; \mathbf{y}, \mathbf{g})_2 = & |y_1 - 1 - \exp(x)|^3 + |y_1 + 1 - \exp(x)|^3 \\ &+ |y_2 - 4 - \exp(-x)|^3. \end{aligned} \quad (39)$$

The problem is handled exactly in the same way as the first one. The extended observation vector, is $\mathbf{y}_e = [y_1 - 1, y_1 + 1, y_2 - 4]^\top$ and the nonlinearity vector remains equal to the sub-problem 1. We obtain a lower bond for (39) $\gamma^{(2)} = R_2^{-1}(\gamma_2^{(1)}) = 7.77 \leq \tilde{V}(x; \mathbf{y}, \mathbf{g})_2$.

The bounds $\gamma^{(1)}$ and $\gamma^{(2)}$ are combined according to the mixture coefficients in $p(\xi_2)$. to yield a global lower bound $\gamma = -\log[0.8\exp(-\gamma^{(1)}) + 0.2\exp(-\gamma^{(2)})] = 9.25$. The real minimum of the system potential is 13.57.

Now we use the prior pdf $p(x)$, and the upper bound $L = \exp\{-9.25\} = 9.5 \cdot 10^{-5}$ to implement a rejection sampler that draws from $p(x|\mathbf{y})$.

Figure 1 (Left plot) shows the target function $p(x|\mathbf{y})$ and the histogram of $N = 10,000$ samples generated by the RS algorithm. The histogram follows closely the shape of the true posterior pdf.

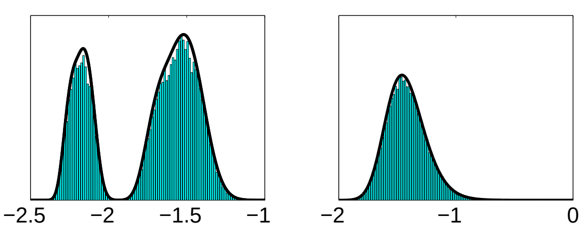

Figure 1: Left plot: The target density $p(x|\mathbf{y})$ of Example 1 and the histogram of $N = 10000$ samples using RS with the calculated bound. Right plot: The target density $p(x|\mathbf{y})$ of Example 2 and the histogram of $N = 10000$ samples using RS with the calculated bound.

7.2 Example 2: Correlated noise variables

Let $x \in \mathbb{R}$ with prior density $p(x) = N(x; 0, 1/2)$ and consider the system with $\mathbf{y} = [y_1, y_2]^\top$

$$y_1 = \exp(x) + \xi_1, \; y_2 = \exp(-x) + \xi_2, \quad (40)$$

where the vector of noise variables is jointly distributed as $p(\xi_1, \xi_2) \propto \exp\{-\xi_1^2 - \xi_2^2 - \rho\xi_1\xi_2\}$, so that the joint potential is $V^{(2)}(\xi_1, \xi_2) = +\xi_1^2 + \xi_2^2 + \rho\xi_1\xi_2$. Due to the monotonicity and convexity of of g_1 and g_2, we can work with a partition of $\mathbb{R}$ consisting of just one set, $\mathscr{B}_1 \equiv \mathbb{R}$. The system potential $V(x; \mathbf{y}, \mathbf{g}) = -\log[p(\mathbf{y}, x)]$ becomes

$$\begin{aligned} V(x; \mathbf{y}, \mathbf{g}) =& (y_1 - \exp(x))^2 + (y_2 - \exp(-x))^2 \\ &+ \rho(y_1 - \exp(x))(y_2 - \exp(-x)). \end{aligned} \quad (41)$$

If, e.g., $\mathbf{y} = [2, 5]^\top$, the state predictions are $x_1 = \log(2)$ and $x_2 = -\log(5)$, therefore $\mathscr{I}_1 = [-\log(5), \log(2)]$. Using the technique in Section 4.1, we find $r_1(x) = 0.78x + 1.45$ and $r_2(x) = -1.95x + 1.85$. We can easily minimize the quadratic potential associated to these linear functions, namely

$$V_2(x; \mathbf{y}, \mathbf{r}) = (y_1 - r_1(x))^2 + (y_2 - r_2(x))^2, \quad (42)$$

to find $\gamma_2 = 2.79 \leq V_2(x; \mathbf{y}, \mathbf{r})$. Using the technique in Section 5, we readily find an increasing function $R(v) = \frac{2}{(2+\rho)}v$ such that $R \circ V \geq V_2$, hence $\gamma = R^{-1}(\gamma_2) \leq V(x; \mathbf{y}, \mathbf{g})$. In particular if, e.g., $\rho = -0.7$, we obtain $\gamma = R^{-1}(\gamma_2) = 1.75$ (the true global minimum of the system potential being 2.72). Finally, the likelihood upper bound is $L = \exp(-1.75)$.

Figure 1 (lower plot) shows the target function $p(x|\mathbf{y})$ (solid line) and the histogram constructed from $N = 10,000$ independent samples generated by the RS algorithm.

8. CONCLUSIONS

We have proposed new methods for the systematic computation of the bounds needed for the implementation of rejection sampling schemes. In particular, we have considered the problem of drawing from a posterior probability distribution using the prior as a proposal function to generate candidate samples. In this setup, the problem consists in finding a tight upper bound for the likelihood function induced by the available data. We have briefly reviewed previous results and then extended them to: (a) provide a more general bounding algorithm that can be used with a very broad class of noise distributions (including correlated noise variables) and (b) propose scenario-specific algorithms that yield more accurate bounds in cases where the noise variables are described by exponential-multimodal and mixture densities. Two simple examples have been provided to illustrate the application of the proposed techniques.

References

[1] M. H. DeGroot and M. J. Schervish. *Probability and Statistics, 3rd ed.* Addison-Wesley, New York, 2002.

[2] P. M. Djurić, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Míguez. Particle filtering. *IEEE Signal Processing Magazine*, 20(5):19–38, September 2003.

[3] A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer, New York (USA), 2001.

[4] W. J. Fitzgerald. Markov chain Monte Carlo methods with applications to signal processing. *Signal Processing*, 81(1):3–18, January 2001.

[5] W. R. Gilks and P. Wild. Adaptive Rejection Sampling for Gibbs Sampling. *Applied Statistics*, 41(2):337–348, 1992.

[6] J. S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer, 2004.

[7] L. Martino and J. Míguez. A novel rejection sampling scheme for posterior probability distributions. *Proc. of the 34th IEEE ICASSP*, April 2009.

[8] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, 2004.