A NOVEL REJECTION SAMPLING SCHEME FOR POSTERIOR PROBABILITY DISTRIBUTIONS

Luca Martino and Joaquín Míguez

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

ABSTRACT

Rejection sampling (RS) is a well-known method to draw from arbitrary target probability distributions, which has important applications by itself or as a building block for more sophisticated Monte Carlo techniques. The main limitation to the use of RS is the need to find an adequate upper bound for the ratio of the target probability density function (pdf) over the proposal pdf from which the samples are generated. There are no general methods to analytically find this bound, except in the particular case in which the target pdf is log-concave. In this paper we adopt a Bayesian view of the problem and propose a general RS scheme to draw from the posterior pdf of a signal of interest using its prior density as a proposal function. The method enables the analytical calculation of the bound and can be applied to a large class of target densities. We illustrate its use with a simple numerical example.

Index Terms— Rejection sampling; Monte Carlo methods; Monte Carlo integration; Overbounding; Sampling methods.

1. INTRODUCTION

Monte Carlo statistical methods are powerful tools for numerical inference and optimization [5]. Currently, there exist several classes of MC techniques, including the popular Markov Chain Monte Carlo (MCMC) [2] and particle filtering [1] algorithms, which enjoy numerous applications in signal processing.

Among these, rejection sampling (RS) [5, Chapter 2] is a classical Monte Carlo method for "universal sampling". It can be used to generate samples from a target probability density function (pdf) 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 probability distribution. RS can be applied as a tool by itself, in problems where the goal is to approximate integrals with respect to (w.r.t.) the pdf of interest, but more often it is a useful building block for more sophisticated Monte Carlo procedures [3, 4].

An important limitation of RS methods is the need to analytically establish a bound for the ratio of the target and proposal densities. There is a lack of general methods for bound computation, however. One exception is the adaptive rejection sampling (ARS) method [3, 5] which, given a target density, provides a procedure to obtain a suitable proposal pdf (for which the bound is easy to compute). This procedure is only valid when the target pdf is strictly log-concave, which is not the case in most practical cases. In this paper we adopt a Bayesian view of the problem and propose a general procedure to apply RS when the target pdf is the posterior pdf of a signal of interest (SI) given a collection of observations and the proposal density is the prior of the SI. Unlike the ARS technique, our method can handle target pdf's with several modes (hence not log-concave). The number of available observations can be arbitrary and each observation can be related to the SI through a different nonlinearity. We assume that the observations are contaminated with additive noise, but these random variables need not be identically distributed.

The remaining of the paper is organized as follows. We formally describe the signal model in Section 2. Some useful definitions and basic assumptions are introduced in Section 3. In Section 4 we derive the proposed method to find an upper bound of the target pdf in its general form, while Section 5 is devoted to specific cases of practical interest. The proposed technique is exemplified in Section 6 and Section 7 is devoted to the conclusions.

2. MODEL AND PROBLEM STATEMENT

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 $y_1 = g_1(\mathbf{x}) + \xi_1, \ldots, y_n = g_n(\mathbf{x}) + \xi_n$ 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. We will assume noise pdf's of the form

$$p(\xi_i) = k_i \exp\left\{-\alpha_i \bar{V}_i(\xi_i)\right\}, \qquad k_i, \alpha_i > 0, \tag{1}$$

where k_i, α_i are real constants and $\bar{V}(\xi_i)$ is a function, subsequently referred to as *marginal potential*, with the following properties: (I) it is real and non negative, $\bar{V}_i : \mathbb{R} \to [0, +\infty)$; and (II) it is increasing ($\frac{d\bar{V}_i}{d\xi_i} > 0$) for $\xi_i > 0$ and decreasing ($\frac{d\bar{V}_i}{d\xi_i} < 0$) for $\xi_i < 0$.

These conditions imply that $\bar{V}_i(\xi_i)$ has a unique minimum at $\xi_i^* = 0$ and, as a consequence, $p(\xi_i)$ has only one maximum (mode) at $\xi_i^* = 0$. Since the noise variables are independent, the joint pdf $p(\xi_1, \xi_2, \ldots, \xi_n) = \prod_{i=1}^n p(\xi_n)$ is easy to construct and we can define a *joint potential* $V^{(n)} : \mathbb{R}^n \to [0, +\infty)$ as

$$V^{(n)}(\xi_1, \ldots, \xi_n) \triangleq -\log\left[p(\xi_1, \ldots, \xi_n)\right] = -\sum_{i=1}^n \log p(\xi_n). \tag{2}$$

Substituting (1) into (2) yields

$$V^{(n)}(\xi_1, \ldots, \xi_n) = c_n + \sum_{i=1}^n \alpha_i \bar{V}_i(\xi_i) \tag{3}$$

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

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 $V_l^{(n)}(\xi_1, \ldots, \xi_n) = \sum_{i=1}^{n} |\xi_i|^l$ for $0 < l < +\infty$, 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 *Euclidean potential*. 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*, $\ell(\mathbf{x}; \mathbf{y}, \mathbf{g}) \triangleq p(\mathbf{y}|\mathbf{x})$, factorizes as

$$\ell(\mathbf{x}; \mathbf{y}, \mathbf{g}) = \prod_{i=1}^{n} p(y_i|\mathbf{x}), \tag{4}$$

where $p(y_i|\mathbf{x}) = k_i \exp\left\{-\alpha_i \bar{V}_i(y_i - g_i(\mathbf{x}))\right\}$. The likelihood in (4) induces a *system potential* $V(\mathbf{x}; \mathbf{y}, \mathbf{g}) : \mathbb{R}^m \to [0, +\infty)$,

$$V(\mathbf{x}; \mathbf{y}, \mathbf{g}) \triangleq -\ln[\ell(\mathbf{x}; \mathbf{y}, \mathbf{g})] = -\sum_{i=1}^{n} \log[p(y_i|\mathbf{x})], \tag{5}$$

that is a function of $\mathbf{x}$ and depends on the observations $\mathbf{y}$ and the function $\mathbf{g}$. Using (3) and (5), we write the system potential in terms of the joint potential,

$$V(\mathbf{x}; \mathbf{y}, \mathbf{g}) = V^{(n)}(y_1 - g_1(\mathbf{x}), \ldots, y_n - g_n(\mathbf{x})). \tag{6}$$

Assume 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. In such case, let L be an upper bound for the likelihood, $\ell(\mathbf{x}; \mathbf{y}, \mathbf{g}) \leq L$, and generate N samples according to the 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 $\frac{p(\mathbf{x}'|\mathbf{y})}{Lp(\mathbf{x}')} \propto \frac{\ell(\mathbf{x}'; \mathbf{y}, \mathbf{g})}{L} > u'$ then $\mathbf{x}_i = \mathbf{x}'$, else discard $\mathbf{x}'$ and go back to step 2. **4.-** Set $i = i + 1$. If $i > N$ stop, else go back to step 2. Then, $I(f)$ can be approximated as $I(f) \approx \hat{I}(f) = \frac{1}{N} \sum_{i=1}^{N} f(\mathbf{x}_i)$.

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, $\bar{V}_i$, $i = 1, ..., n$, the system potential can be highly multimodal w.r.t. $\mathbf{x}$. See the example in the Section 6 for an illustration.

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 can be extended to the general case, although this extension is not trivial. We define the set of *state predictions* as $\tilde{X} \triangleq \left\{\tilde{x}^{(i)} \in \mathbb{R} : \ y_i = g_i(\tilde{x}^{(i)}) \ \text{ for } i = 1, \ldots, n\right\}$. Each equation $y_i = g_i(\tilde{x}^{(i)})$, in general, can yield zero, one or more state predictions. We also introduce the maximum likelihood (ML) *state estimator* $\hat{x}$, as $\hat{x} \triangleq \arg\max_{x \in \mathbb{R}} \ell(x|\mathbf{y}, \mathbf{g}) = \arg\min_{x \in \mathbb{R}} V(x; \mathbf{y}, \mathbf{g})$, not necessarily unique.

Let us use $A \subseteq \mathbb{R}$ to denote the support of the vector function $\mathbf{g}$, i.e., $\mathbf{g} : A \subseteq \mathbb{R} \to \mathbb{R}^n$. We assume that there exists a partition $\{B_j\}_{j=1}^{q}$ of A (i.e., $A = \cup_{j=1}^{q} B_j$ and $B_i \cap B_j = \emptyset$, $\forall i \neq j$) such that we can define functions $g_{i,j} : B_j \to \mathbb{R}$, $\ j = 1, \ldots, q$ and $i = 1, \ldots, n$, as $g_{i,j}(x) \triangleq g_i(x)$, $\ \forall x \in B_j$, and: (a) every function $g_{i,j}$ is invertible in B_j and (b) every function $g_{i,j}$ is either convex in B_j or concave in B_j. Assumptions (a) and (b) together mean that, for every i and all $x \in B_j$, 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, $\tilde{x}^{(i)} = g_{i,j}^{-1}(y_i)$. We write the set of predictions in B_j as $\tilde{X}_j = \{\tilde{x}^{(1)}, \ldots, \tilde{x}^{(n)}\}$. If $g_{i,j}$ is bounded and y_i is noisy, it is conceivable that $y_i > \max_{x \in [B_j]} g_{i,j}(x)$ (or $y_i < \min_{x \in [B_j]} g_{i,j}(x)$), where $[B_j]$ denotes the closure of the set B_j, hence $g_{i,j}^{-1}(y_i)$ may not exist. In such case, we define $\tilde{x}^{(i)} = \arg\max_{x \in [B_j]} g_{i,j}(x)$ (or $\tilde{x}^{(i)} = \arg\min_{x \in [B_j]} g_{i,j}(x)$, respectively), and admit $\tilde{x}^{(i)} = +\infty$ (respectively, $\tilde{x}^{(i)} = -\infty$) as valid solutions.

4. GENERAL COMPUTATION 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. To circumvent this obstacle, we split the problem into q subproblems and address the computation of bounds for each set B_j, $j = 1, \ldots, q$, in the partition of A. Within B_j we will build adequate linear functions $\{r_{i,j}\}_{i=1}^{n}$ in order to replace the nonlinearities $\{g_{i,j}\}_{i=1}^{n}$. We require that, for every $r_{i,j}$, the inequalities

$$|y_i - r_{i,j}(x)| \leq |y_i - g_{i,j}(x)|, \text{ and} \tag{7}$$

$$(y_i - r_{i,j}(x))(y_i - g_{i,j}(x)) \geq 0 \tag{8}$$

hold jointly for all $i = 1, \ldots, n$, and all $x \in I_j \subset B_j$, where I_j is any closed interval in B_j such that $\hat{x}_j = \arg\min_{x \in [B_j]} V(x; \mathbf{y}, \mathbf{g})$ (i.e., the state estimator of x restricted to B_j, possible non unique) is contained in I_j. The latter requirement can be fulfilled if we choose $I_j \triangleq [\min(\tilde{X}_j), \max(\tilde{X}_j)]$ (see the Appendix for a proof).

If (7) and (8) hold, we can write

$$\bar{V}_i(y_i - r_{i,j}(x)) \leq \bar{V}_i(y_i - g_{i,j}(x)), \quad \forall x \in I_j, \tag{9}$$

which follows easily from the properties of the marginal potential functions $\bar{V}_i$. Moreover, since $V(x; \mathbf{y}, \mathbf{g}_j) = c_n + \sum_{i=1}^{n} \alpha_i \bar{V}_i(y_i - g_{i,j}(x))$ and $V(x; \mathbf{y}, \mathbf{r}_j) = c_n + \sum_{i=1}^{n} \alpha_i \bar{V}_i(y_i - r_{i,j}(x))$ where $\mathbf{g}_j = [g_{1,j}, \ldots, g_{n,j}]$ and $\mathbf{r}_j = [r_{1,j}, \ldots, r_{n,j}]$, Eq. (9) implies that $V(x; \mathbf{y}, \mathbf{r}_j) \leq V(x; \mathbf{y}, \mathbf{g}_j)$, $\forall x \in I_j$, and, as a consequence,

$$\gamma_j = \inf_{x \in I_j} V(x; \mathbf{y}, \mathbf{r}_j) \leq \inf_{x \in I_j} V(x; \mathbf{y}, \mathbf{g}_j) = \inf_{x \in B_j} V(x; \mathbf{y}, \mathbf{g}). \tag{10}$$

Therefore, it is possible to find a lower bound in B_j for the system potential $V(x; \mathbf{y}, \mathbf{g}_j)$, denoted γ_j, by minimizing the modified potential $V(x; \mathbf{y}, \mathbf{r}_j)$ in I_j.

All that remains is to actually build $\{r_{i,j}\}_{i=1}^{n}$. This construction is straightforward and can be described graphically by splitting the problem into two cases. Case I corresponds to nonlinearities $g_{i,j}$ such that $\frac{dg_{i,j}(x)}{dx} \times \frac{d^2 g_{i,j}(x)}{dx^2} \geq 0$ (i.e., $g_{i,j}$ is either increasing and convex

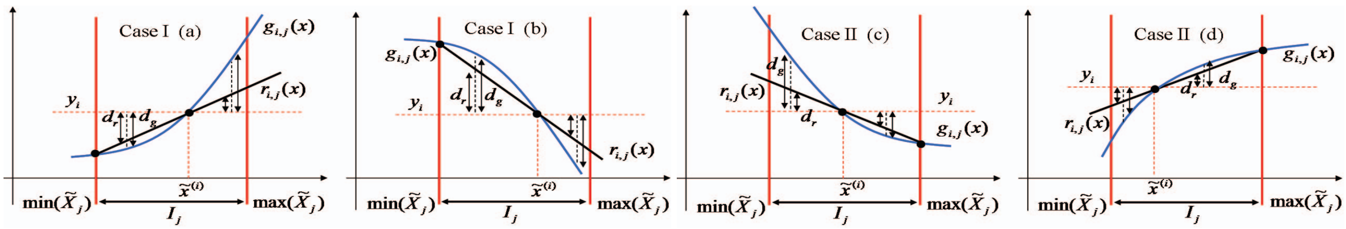

Fig. 1. Construction of the auxiliary linearities $\{r_{i,j}\}_{i=1}^{n}$. (a) Function $g_{i,j}$ is increasing and convex (case I). (b) Function $g_{i,j}$ is decreasing and concave (case I). (c) Function $g_{i,j}$ is decreasing and convex (case II). (d) Function $g_{i,j}$ is increasing and concave (case II).

or decreasing and concave), for case II $\frac{dg_{i,j}(x)}{dx} \times \frac{d^2 g_{i,j}(x)}{dx^2} \leq 0$ (i.e., $g_{i,j}$ is either increasing and concave or decreasing and convex), when $x \in B_j$.

Figure 1 (a)-(b) depicts the construction of $r_{i,j}$ in case I. We choose a linear function $r_{i,j}$ that passes through $\min(\tilde{X}_j)$ and the state prediction $\tilde{x}^{(i)}$. It is apparent that $d_r < d_g$ for $x \in I_j$, hence inequality (7) is granted. Inequality (8) also holds for all $x \in I_j$, since $r_{i,j}(x)$ and $g_{i,j}(x)$ are either simultaneously greater than, or simultaneously lesser than, y_i.

Figure 1 (c)-(d) depicts the construction of $r_{i,j}$ in case II. We choose a linear function $r_{i,j}$ that passes through $\max(\tilde{X}_j)$ and the state prediction $\tilde{x}^{(i)}$. In the figure d_r and d_g denote the distances $|y_i - r_{i,j}(x)|$ and $|y_i - g_{i,j}(x)|$, respectively. It is apparent from the two plots that inequalities (7) and (8) hold for all $x \in I_j$.

A special subcase of I (respectively, of II) occurs when $\tilde{x}^{(i)} = \min(\tilde{X}_{B_j})$ (respectively, $\tilde{x}^{(i)} = \max(\tilde{X}_{B_j})$). Then, $r_{i,j}(x)$ is the tangent to $g_{i,j}(x)$ in $\tilde{x}^{(i)}$. If $\tilde{x}^{(i)} = \pm\infty$ then $r_{i,j}(x)$ is a horizontal asymptote of $g_{i,j}(x)$.

It is often possible to find $\gamma_j = \inf_{x \in I_j} V(x; \mathbf{y}, \mathbf{r}_j) \leq \inf_{x \in I_j} V(x; \mathbf{y}, \mathbf{g}_j)$ in closed-form. If we choose $\gamma = \min_j \gamma_j$, then $\gamma \leq \inf_{x \in \mathbb{R}} V(x, \mathbf{y}, \mathbf{g})$ is a global lower bound of the system potential. Table 1 shows an outline of the proposed method.

Table 1. Algorithm to find a lower bound.

1. Find a partition $\{B_j\}_{j=1}^{q}$.
2. Compute the state prediction set $\tilde{X}_j$ for each B_j.
3. Calculate $\min(\tilde{X}_j)$ and $\max(\tilde{X}_j)$ and build $r_{i,j}(x)$ for $x \in I_j$ and $i = 1, \ldots, n$.
4. Replace $\mathbf{g}_j(x)$ with $\mathbf{r}_j(x)$, and minimize $V(x; \mathbf{y}, \mathbf{r}_j)$ to find the lower bound γ_j.
5. Find $\gamma = \min_j \gamma_j$.

5. SPECIAL CASES

5.1. Lower Bound γ_2 for Euclidean potentials

Assume a Euclidean potential, $V_2^{(n)}(y_1 - g_{1,j}(x), \ldots, y_n - g_{n,j}(x)) = \sum_{i=1}^{n}(y_i - g_{i,j}(x))^2$ 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 "linearized" system potential in B_j becomes $V_2(x; \mathbf{y}, \mathbf{r}_j) = \sum_{i=1}^{n}(y_i - r_{i,j}(x))^2 = \sum_{i=1}^{n}(y_i - a_{i,j}x - b_{i,j})^2$, and it turns out straightforward to compute $\gamma_{2,j} = \min_{x \in B_j} V(x; \mathbf{y}, \mathbf{r}_j)$. 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 $\hat{x}_{L,j} = \arg\min_{x \in B_j} V(x; \mathbf{y}, \mathbf{r}_j) = \frac{\mathbf{a}_j^T \mathbf{w}_j}{(\mathbf{a}_j^T \mathbf{a}_j)}$, and $\gamma_{2,j} = V(x_{L,j}; \mathbf{y}, \mathbf{r}_j)$. Apparently $\gamma_2 = \min_j \gamma_{2,j} \leq V(x; \mathbf{y}, \mathbf{g})$.

5.2. Adaptation of γ_2 for generic system potentials

Consider an arbitrary joint potential $V^{(n)}$ and assume the availability of an invertible function R such that $R \circ V^{(n)} \geq V_2^{(n)}$, where $\circ$ denotes the composition of functions. Then, for the system potential we can write $(R \circ V)(x; \mathbf{y}, \mathbf{g}) \geq \sum_{i=1}^{n}(y_i - g_i(x))^2 \geq \gamma_2$, and, as consequence, $V(x; \mathbf{y}, \mathbf{g}) \geq R^{-1}(\gamma_2) = \gamma$. As an example, consider joint potentials $V_p^{(n)}$, where

$$\left(\sum_{i=1}^{n} |\xi_i|^p\right)^{\frac{1}{p}} \geq \left(\sum_{i=1}^{n} \xi_i^2\right)^{\frac{1}{2}}, \text{ for } 0 \leq p \leq 2, \text{ and} \quad (11)$$

$$n^{\left(\frac{p-2}{2p}\right)} \left(\sum_{i=1}^{n} |\xi_i|^p\right)^{\frac{1}{p}} \geq \left(\sum_{i=1}^{n} \xi_i^2\right)^{\frac{1}{2}}, \text{ for } 2 \leq p \leq +\infty. \quad (12)$$

Both inequalities can be proved from the monotonicity of $\mathcal{L}^p$ norms [6]. We find invertible functions $R(v) = v^{2/p}$, for Eq. (11), and $R(v) = \left(n^{\left(\frac{p-2}{2p}\right)} v^{1/p}\right)^2 = n^{\left(\frac{p-2}{p}\right)} v^{2/p}$, for Eq. (12), that can be applied to yield $\gamma = \gamma_2^{p/2} \leq \sum_{i=1}^{n} |y_i - g_i(x)|^p$, for $0 < p \leq 2$, and $\gamma = \frac{\gamma_2^{p/2}}{n^{(p-2)/2p}} \leq \sum_{i=1}^{n} |y_i - g_i(x)|^p$, for $2 \leq p < +\infty$.

5.3. Convex marginal potentials $\bar{V}_i$

Assume that $A = \{B_j\}_{j=1}^{q}$ and we have already found $r_{i,j}(x) = a_{i,j}x + b_i$, $i = 1, \ldots, n$ and $j = 1, \ldots, q$, using the technique in Section 4. If all marginal potentials $\bar{V}_i(\xi_i)$ are convex, then the "linearized" system potential in B_j, $V(x; \mathbf{y}, \mathbf{r}_j) = c_n + \sum_{i=1}^{n} \alpha_i \bar{V}_i(y_i - r_{i,j}(x))$, is also convex and we can use the tangents to $V(x; \mathbf{y}, \mathbf{r}_j)$ at the limit points of I_j (i.e, $\min(\tilde{X}_j)$ and $\max(\tilde{X}_j)$) to find a bound. This is depicted in Fig. 2 (a), where it is seen that the intersection of the two tangents yields a lower bound in B_j.

6. EXAMPLE

Given $x \in \mathbb{R}$ with prior density $p(x) \sim N(x; 0, 2)$ and the system with $\mathbf{y} \in \mathbb{R}^{n=2}$

$$y_1 = \exp(x) + \xi_1, \quad y_2 = \exp(-x) + \xi_2, \quad (13)$$

where ξ_1 is Gaussian noise $N(\xi_1; 0, 1/2) = k_1 \exp\{-(\xi_1)^2\}$, and ξ_2 has a gamma pdf, $\Gamma(\xi_2; \theta, \lambda) = k_2 \xi_2^{\theta-1} \exp\{-\lambda \xi_2\}$, with parameters $\theta = 2, \lambda = 1$, the corresponding marginal potentials are $V_1(\xi_1) = \xi_1^2$ and $\bar{V}_2(\xi_2) = -\log(\xi_2) + \xi_2$. Since the minimum of $\bar{V}_2(\xi_2)$ occurs in $\xi_2 = 1$, we replace y_2 with the shifted observation

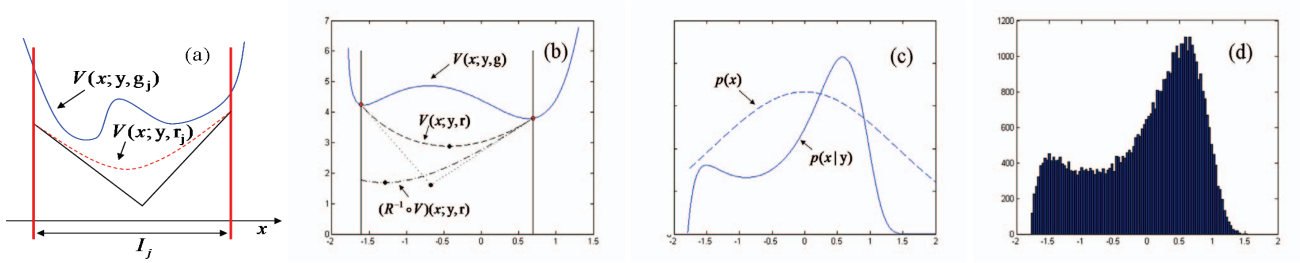

Fig. 2. (a) The intersection of the tangents to $V(x;\mathbf{y},\mathbf{r}_j)$ at $\min(\tilde{X}_j)$ and $\max(\tilde{X}_j)$ is a lower bound for $V(x;\mathbf{y},\mathbf{g}_j)$. (b) The system potential $V(x,\mathbf{y},\mathbf{g})$ (solid), the modified system potential $V(x,\mathbf{y},\mathbf{r})$ (dashed), function $(R^{-1}\circ V)(x,\mathbf{y},\mathbf{r})$ (dot-dashed) and two tangents lines (dotted)). (c) The target density $p(x|\mathbf{y})$ (solid), and the prior $p(x)$ (dashed). (d) The histogram of $N = 10000$ samples using RS.

$y_2^* = \exp(-x) + \xi_2^*$, where $y_2^* = y_2 - 1$, $\xi_2^* = \xi_2 - 1$. The marginal potential becomes $\bar{V}_2(\xi_2^*) = -\log(\xi_2^* + 1) + \xi_2^* + 1$ with a minimum at $\xi_2^* = 0$. The observation vector is $\mathbf{y} = [y_1, y_2^*]^T$ and the vector of nonlinearities is $\mathbf{g} = [\exp(x), \exp(-x)]^T$. 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, $B_1 \equiv \mathbb{R}$. The joint potential is $V^{(2)}(\xi_1, \xi_2^*) = \xi_1^2 - \ln(\xi_2^* + 1) + \xi_2^* + 1$ and the system potential

$$V(x;\mathbf{y},\mathbf{g}) = (y_1 - \exp(x))^2 - \log(y_2^* - \exp(-x) + 1) + (y_2^* - \exp(-x)) + 1. \quad (14)$$

If, e.g., $\mathbf{y} = [2, 5]$, the state predictions are $\tilde{x}^{(1)} = \log(2)$ and $\tilde{x}^{(2)} = -\log(5)$, therefore $I_{B_1} = [-\log(5), \log(2)]$. Using the technique in Section 4, we find $r_1(x) = -1.95x + 1.85$ and $r_2(x) = 0.78x + 1.45$. In this case, we can analytically minimize the modified system potential, $V(x;\mathbf{y};\mathbf{r})$, finding $\arg\min_{x\in I_{B_1}}(V(x,\mathbf{y},\mathbf{r})) = -0.4171$. The associated lower bound is $\gamma = \min_{x\in I_{B_1}}(V(x,\mathbf{y},\mathbf{r})) = 2.89$ (the true global minimum of the system potential is 3.78). Since the marginal potentials are both convex, we could have used the procedure described in Section 5.3, obtaining as lower bound the value $\gamma = 1.61$. We also could have used the technique in Section 5.2 with $R^{-1}(v) = -\log(\sqrt{v} + 1) + \sqrt{v} + 1$. Solving the problem for an Euclidean potential, we find $\gamma_2 = 2.79$. Therefore, in this case our lower bound is $\gamma = R^{-1}(\gamma_2) = 1.68$.
Now we use the prior pdf $p(x) = N(x; 0, 2)$ and the upper bound $L = \exp\{-2.89\}$ to implement a rejection sampler that draws from $p(x|\mathbf{y})$. Figure 2 (b) depicts the system potential $V(x;\mathbf{y},\mathbf{g})$ (solid line), the modified system potential $V(x;\mathbf{y},\mathbf{r})$ (dashed line), the function $(R^{-1}\circ V)(x;\mathbf{y},\mathbf{r})$ (dotted-dashed line) and the two tangents lines (dotted lines) for our example. Figure 2 (c) shows the target function $p(x|\mathbf{y})$ and the proposal $p(x)$. Finally, Figure 2 (d) depicts the histogram of $N = 10,000$ samples generated by the RS algorithm. The histogram follows closely the shape of the true posterior pdf. The acceptance rate for the sampler is $\approx 20\%$.

7. CONCLUSIONS

We have proposed a rejection sampling method to draw from the posterior pdf, $p(x|\mathbf{y})$, of a signal of interest x given a collection of nonlinear observations $\mathbf{y}$ in additive noise. The new technique uses the prior, $p(x)$, as a proposal density and yields the required bound of $p(\mathbf{y}|x)$ analytically. Our method can be used in different Monte Carlo schemes, including accept-reject particle filters [4] and MCMC methods [5]. It also yields a natural generalization of the adaptive rejection sampling scheme of [3] that can be applied without requiring the target pdf to be log-concave.

Appendix

Proposition: The state estimator $\hat{x}_j$ belongs to interval I_j, i.e., $\hat{x}_j \in I_j \triangleq [\min(\tilde{X}_j), \max(\tilde{X}_j)]$, where $I_j \subseteq B_j$.
Proof: We have to prove that $dV/dx < 0$, for all $x < \min(\tilde{X}_j)$, and $dV/dx > 0$, for all $x > \max(\tilde{X}_j)$, so that all stationary points of V stay inside $I_j = [\min(\tilde{X}_j), \max(\tilde{X}_j)]$. Routine calculations yield the derivative $dV/dx = -\sum_{i=1}^{n} \alpha_i dg_i/dx \left[d\bar{V}_i/d\xi_i\right]_{\xi_i = y_i - g_i(x)}$ and we aim to evaluate it outside the interval I_j. To do it, let us denote $\tilde{x}_{min} = \min(\tilde{X}_j)$ and $\tilde{x}_{max} = \max(\tilde{X}_j)$ and consider the cases $dg_i/dx > 0$ and $dg_i/dx < 0$ separately (the sign of dg_i/dx does not change in B_j). In the first case we can readily find that $dg_i/dx > 0$, $\tilde{x}^{(i)} \geq \tilde{x}_{min}$ together imply that $y_i = g_i(\tilde{x}^{(i)}) \geq g_i(\tilde{x}_{min}) > g_i(x)$ $\forall x < \tilde{x}_{min}$. Therefore to the assumed properties of potential functions, $\left[d\bar{V}_i/d\xi_i\right]_{\xi_i = y_i - g_i(x) > 0} > 0$. As a consequence, $\frac{dV}{dx} < 0$ $\forall x < \tilde{x}_{min}$. When $dg_i/dx < 0$ and $\tilde{x}^{(i)} \geq \tilde{x}_{min}$ we obtain that $y_i = g_i(\tilde{x}^{(i)}) \leq g_i(\tilde{x}_{min}) < g_i(x)$, $\forall x < \tilde{x}_{min}$. Then $y_i - g_i(x) < 0$ for all $x < \tilde{x}_{min}$ and $\left[d\bar{V}_i/d\xi_i\right]_{\xi_i = y_i - g_i(x) < 0} < 0$. As a consequence, $\frac{dV}{dx} < 0 \; \forall x < \tilde{x}_{min}$.
A similar argument for $x > \tilde{x}_{max}$ yields $\frac{dV}{dx} > 0$ for all $x > \tilde{x}_{max}$ and completes the proof. □

References

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

[2] W. R. Gilks and D. Spiegelhalter S. Richardson. *Markov Chain Monte Carlo in Pratice*. CRC Press, Interdisciplinary Statistics Series, 1995.

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

[4] H. R. Künsch. Recursive Monte Carlo filters: Algorithms and theoretical bounds. *The Annals of Statistics*, 33(5):1983–2021, 2005.

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

[6] D. Williams. *Probability with Martingales*. Cambridge University Press, Cambridge (UK), 1991.