AN ADAPTIVE ACCEPT/REJECT SAMPLING ALGORITHM FOR 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

Accept/reject sampling is a well-known method to generate random samples from arbitrary target probability distributions. It demands the design of a suitable proposal probability density function (pdf) from which candidate samples can be drawn. These samples are either accepted or rejected depending on a test involving the ratio of the target and proposal densities. In this paper we introduce an adaptive method to build a sequence of proposal pdf's that approximate the target density and hence can ensure a high acceptance rate. In order to illustrate the application of the method we design an accept/reject particle filter and then assess its performance and sampling efficiency numerically, by means of computer simulations.

Index Terms— Rejection sampling; adaptive rejection sampling; particle filtering; Monte Carlo integration; sensor networks.

1. INTRODUCTION

The popularity of Bayesian methods in signal processing has brought a surge of interest in the Monte Carlo techniques that are often necessary for the implementation of optimal *a posteriori* estimators [1]. The accept/reject method, also known as rejection sampling (RS), is a fundamental technique that 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.

The class of adaptive RS (ARS) methods is particularly interesting because they ensure high acceptance rates. The standard ARS algorithm [2] yields a sequence of proposal functions that actually converge towards the target pdf as the procedure is iterated. As the proposal density becomes close to the target pdf, the proportion of accepted samples grows (and, in the limit, can also converge to 1). However, this algorithm can only be used with log-concave target densities. A variation of the standard procedure, termed adaptive rejection Metropolis sampling (ARMS) [3] can also be used with multimodal pdf's. However, the ARMS algorithm is based on the Metropolis-Hastings algorithm, so the resulting samples form a Markov chain. As a consequence, they are correlated and, for certain multimodal densities, the chain can be easily trapped in a single mode.

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 propose a novel ARS algorithm based on the construction of a piecewise-constant approximation of the log-density of interest. We derive the proposed method in the context of optimal Bayesian inference and, hence, we assume the target function is a posterior pdf. The resulting technique is very flexible, can be applied to a broad class of densities (possibly multimodal) and yields independent and identically distributed (i.i.d.) samples coming exactly from the desired probability distribution.

We formally state the problem to be investigated in Section 2. Background material, including some necessary definitions, assumptions and tools for the calculation of bounds, is presented in Section 3. The new ARS algorithm is introduced in Section 4 and in Section 5 we apply it to the design of an accept/reject particle filtering (ARPF) algorithm [4]. The performance and efficiency of the resulting method is numerically assessed, by means of computer simulations. Finally, Section 6 is devoted to the conclusions.

2. PROBLEM STATEMENT

Many signal processing problems involve the estimation of an unobserved signal of interest (SoI), denoted $\mathbf{x} \in \mathbb{R}^m$ (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 SoI, $p(\mathbf{x})$, 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

$$y_1 = g_1(\mathbf{x}) + \vartheta_1, \ldots, y_n = g_n(\mathbf{x}) + \vartheta_n \tag{1}$$

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 ϑ_i are independent noise variables, possibly with different distributions for each i. Let us assume noise pdf's of the form

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

where $k_i > 0$ is a real constant and $\bar{V}_i(\vartheta_i)$ is a real function, subsequently referred to as *marginal potential*. We assume that $\bar{V}_i$ is non-negative and convex, with a minimum at $\vartheta_i^* = 0$.

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 SoI $\mathbf{x}$, hence the *likelihood function*,

[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 the observation of y is written as $p(x|y)$.

978-1-4244-2710-9/09/$25.00 © 2009 IEEE

$\ell(\mathbf{x};\mathbf{y},\mathbf{g}) \triangleq p(\mathbf{y}|\mathbf{x})$, can be factorized as

$$\ell(\mathbf{x};\mathbf{y},\mathbf{g}) = \prod_{i=1}^{n} p(y_i|\mathbf{x}). \tag{3}$$

Moreover, since the noise terms are additive, we obtain

$$p(y_i|\mathbf{x}) = k_i \exp\left\{-\bar{V}_i(y_i - g_i(\mathbf{x}))\right\}, \tag{4}$$

and the likelihood in (3) induces a *system potential* $V(\mathbf{x};\mathbf{y},\mathbf{g}) : \mathbb{R}^m \to [0,+\infty)$, defined as

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

that is a function of $\mathbf{x}$ and depends on the observations $\mathbf{y}$ and the function $\mathbf{g}$. Using (4) and (5), we can rewrite the system potential in terms of the marginal potentials,

$$V(\mathbf{x};\mathbf{y},\mathbf{g}) = c_n + \sum_{i=1}^{n} \bar{V}_i(y_i - g_i(\mathbf{x})). \tag{6}$$

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 SoI 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.

Let $\eta(\mathbf{x})$ be a lower-bounding function for the system potential, i.e., $\eta(\mathbf{x}) \leq V(\mathbf{x};\mathbf{y},\mathbf{g})$, so that $L(\mathbf{x}) \triangleq \exp\{-\eta(\mathbf{x})\}$ is an upper-bounding function for the likelihood, i.e., $\ell(\mathbf{x};\mathbf{y},\mathbf{g}) \leq L(\mathbf{x})$. We aim to apply RS using a proposal function of the form $\pi(\mathbf{x}) \propto L(\mathbf{x})p(\mathbf{x})$. Specifically, we wish to draw N i.i.d. samples from $p(\mathbf{x}|\mathbf{y})$ using the algorithm below.

1. Set $i = 1$.
2. Draw $\mathbf{x}'$ from $\pi(\mathbf{x}) \propto L(\mathbf{x})p(\mathbf{x})$ and u' from $U(0,1)$, where $U(0,1)$ is the uniform pdf in $[0,1]$.
3. If $\frac{\ell(\mathbf{x}';\mathbf{y},\mathbf{g})}{L(\mathbf{x}')} > u'$ 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.

Then, we 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 obtaining a suitable over-bounding function $L(\mathbf{x})$, such that we are able to draw from $\pi(\mathbf{x}) \propto L(\mathbf{x})p(\mathbf{x})$. Note that, it is equivalent to maximize ℓ with respect to (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-bounding function $\eta(\mathbf{x}) \leq V(\mathbf{x};\mathbf{y},\mathbf{g})$, related to likelihood bound as $L(\mathbf{x}) = \exp\{-\eta(\mathbf{x})\}$. This problem is far from trivial, though. Even for very simple marginal potentials, $\bar{V}_i$, the system potential, V, can be highly multimodal w.r.t. $\mathbf{x}$ [5].

3. BACKGROUND: BOUNDS ON THE LIKELIHOOD

3.1. Notations and assumptions

Hereafter we restrict our attention to the case of a scalar SoI, $x \in \mathbb{R}$. 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 proposed algorithms can be extended to the multidimensional case, although this extension is not trivial.

We define the set of *simple estimates* as

$$\mathcal{X} \triangleq \{x_i \in \mathbb{R} : \ y_i = g_i(x_i) \ \text{ for } i = 1,\ldots,n\}, \tag{7}$$

where each equation $y_i = g_i(x_i)$, in general, can yield zero, one or several simple estimates.

Let us use $\mathcal{A} \subseteq \mathbb{R}$ to denote the support of the vector function $\mathbf{g}$, i.e., $\mathbf{g} : \mathcal{A} \subseteq \mathbb{R} \to \mathbb{R}^n$. We assume that there exists a partition $\{\mathcal{B}_j\}_{j=1}^{q}$ of $\mathcal{A}$ (i.e., $\mathcal{A} = \cup_{j=1}^{q}\mathcal{B}_j$ and $B_i \cap \mathcal{B}_j = \emptyset$, $\forall i \neq j$) such that we can define functions $g_{i,j} : \mathcal{B}_j \to \mathbb{R}, \quad j = 1,\ldots,q$ and $i = 1,\ldots,n$, as

$$g_{i,j}(x) \triangleq g_i(x), \ \ \forall x \in \mathcal{B}_j, \tag{8}$$

and: (a) every function $g_{i,j}$ is invertible in $\mathcal{B}_j$ and (b) every function $g_{i,j}$ is either convex in $\mathcal{B}_j$ or concave in $\mathcal{B}_j$. Assumptions (a) and (b) together mean that, for every i and all $x \in \mathcal{B}_j$, 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 simple estimates in each subset of the partition, denoted $x_{i,j} = g_{i,j}^{-1}(y_i)$, $i = 1,...,n$, and we write the set of simple estimates in $\mathcal{B}_j$ as $\mathcal{X}_j = \{x_{1,j},\ldots,x_{n,j}\}$. If $g_{i,j}$ is bounded and y_i is noisy, it is conceivable that $y_i > \max_{x\in[\mathcal{B}_j]} g_{i,j}(x)$ (or $y_i < \min_{x\in[\mathcal{B}_j]} g_{i,j}(x)$), where $[\mathcal{B}_j]$ denotes the closure of $\mathcal{B}_j$, hence $g_{i,j}^{-1}(y_i)$ may not exist. In such case, we define $x_{i,j} = \arg\max_{x\in[\mathcal{B}_j]} g_{i,j}(x)$ (or $x_{i,j} = \arg\min_{x\in[\mathcal{B}_j]} g_{i,j}(x)$, respectively), and admit $x_{i,j} = +\infty$ (respectively, $x_{i,j} = -\infty$) as valid solutions.

3.2. Basic algorithm for the calculation of bounds

We now briefly review a basic algorithm for the calculation of upper bounds for the likelihood function [5]. Our goal is to obtain an analytical method for the computation of a scalar $\gamma_{k,j} \in \mathbb{R}$ such that $\gamma_{k,j} \leq \inf_{x\in\mathcal{I}_{k,j}} V(x;\mathbf{y},\mathbf{g}_j)$ in an interval $x \in \mathcal{I}_{k,j} \triangleq (s_{k-1}, s_k)$ where $s_{k-1}, s_k \in \mathcal{B}_j$ (and $s_{k-1} < s_k$). The observations $\mathbf{y}$ are arbitrary but fixed and the nonlinearities $\mathbf{g}$ are assumed known.

The main difficulty to carry out this calculation is the nonlinearity $\mathbf{g}$ which, in many cases, renders the problem not directly tractable. In [5] it is described how to build, within each set $\mathcal{B}_j$ ($\mathcal{A} = \cup_{j=1}^{q}\mathcal{B}_j$), adequate linear functions $\left\{r_{i,j}^{(k)}\right\}_{i=1}^{n}$ in order to replace the nonlinearities $\{g_{i,j}\}_{i=1}^{n}$. We construct every $r_{i,j}^{(k)}$ in a way that ensures

$$\bar{V}_i(y_i - r_{i,j}^{(k)}(x)) \leq \bar{V}_i(y_i - g_{i,j}(x)), \quad \forall x \in \mathcal{I}_{k,j}. \tag{9}$$

Then, since $V(x;\mathbf{y},\mathbf{g}_j) = c_n + \sum_{i=1}^{n}\bar{V}_i(y_i - g_{i,j}(x))$ and $V(x;\mathbf{y},\mathbf{r}_j^{(k)}) = c_n + \sum_{i=1}^{n}\bar{V}_i(y_i - r_{i,j}^{(k)}(x))$ where $\mathbf{g}_j = [g_{1,j},\ldots,g_{n,j}]$ and $\mathbf{r}_j^{(k)} = [r_{1,j}^{(k)},\ldots,r_{n,j}^{(k)}]$, Eq. (9) together with the assumed properties of the marginal potentials imply that $V(x;\mathbf{y},\mathbf{r}_j^{(k)}) \leq V(x;\mathbf{y},\mathbf{g}_j)$, $\forall x \in \mathcal{I}_{k,j}$, and as a consequence,

$$\gamma_{k,j} = \inf_{x\in\mathcal{I}_{k,j}} V(x;\mathbf{y},\mathbf{r}_j^{(k)}) \leq \inf_{x\in\mathcal{I}_{k,j}} V(x;\mathbf{y},\mathbf{g}_j). \tag{10}$$

Therefore, it is possible to find a lower bound in $\mathcal{I}_{k,j}$ for the system potential $V(x;\mathbf{y},\mathbf{g}_j)$, denoted $\gamma_{k,j}$, by minimizing the modified potential $V(x;\mathbf{y},\mathbf{r}_j^{(k)})$ within $\mathcal{I}_{k,j}$.

The construction of each $r_{i,j}^{(k)}$ is straightforward dividing the problem into two cases. Case 1 corresponds to nonlinearities $g_{i,j}$ such

that $\frac{dg_{i,j}(x)}{dx} \times \frac{d^2 g_{i,j}(x)}{dx^2} \geq 0$, while case 2 corresponds to functions that comply with $\frac{dg_{i,j}(x)}{dx} \times \frac{d^2 g_{i,j}(x)}{dx^2} \leq 0$, when $x \in \mathcal{B}_j$.

In case 1, when the simple estimate $x_{i,j}$ belongs to the interval $\mathcal{I}_{k,j} = (s_{k-1}, s_k)$, we choose the linear function $r_{i,j}^{(k)}$ that passes through $(s_{k-1}, g_i(s_{k-1}))$ and the point corresponding to the simple estimate $(x_{i,j}, g_i(x_{i,j}))$. While in case 2, when $x_{i,j} \in \mathcal{I}_{k,j}$, we choose the linear function $r_{i,j}^{(k)}$ that passes through the points $(x_{i,j}, g_i(x_{i,j}))$ and $(s_k, g_i(s_k))$.

When $x_{i,j} \leq s_{k-1}$, the linear function $r_{i,j}^{(k)}$ is the tangent line to $g_i(x)$ at s_{k-1} in case 1, and the straight line that passes through the points $(s_{k-1}, g_i(s_{k-1}))$ and $(s_k, g_i(s_k))$ in case 2. When $x_{i,j} \geq s_k$, the linear function $r_{i,j}^{(k)}$ is the straight line that passes through the points $(s_{k-1}, g_i(s_{k-1}))$ and $(s_k, g_i(s_k))$ in case 1, and the tangent line to $g_i(x)$ at s_k in case 2. Finally, if $x_{i,j} = \pm\infty$ then $r_{i,j}^{(k)}(x)$ is a horizontal asymptote of $g_{i,j}(x)$.

It is often possible to find $\gamma_{k,j} = \inf_{x \in \mathcal{I}_{k,j}} V(x; \mathbf{y}, \mathbf{r}_j^{(k)}) \leq \inf_{x \in \mathcal{I}_{k,j}} V(x; \mathbf{y}, \mathbf{g}_j)$ in closed-form. However, if the minimization of $V(x; \mathbf{y}, \mathbf{r}_j^{(k)})$ remains intractable, we can use the method described below.

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

Assume that $\mathcal{A} = \{\mathcal{B}_j\}_{j=1}^q$ and that we have already found $r_{i,j}^{(k)}(x) = a_{i,j}^{(k)} x + b_{i,j}^{(k)}$, $i = 1, \ldots, n$ and $j = 1, \ldots, q$, using the procedure in Section 3.2. Since the marginal potentials $\bar{V}_i(\vartheta_i)$ are convex and the function $r_{i,j}^{(k)}(x)$ is linear, it turns out that $\bar{V}_i(y_i - r_{i,j}^{(k)}(x))$ is also convex. As a consequence

- the modified system potential, $V(x; \mathbf{y}, \mathbf{r}_j^{(k)}) = c_n + \sum_{i=1}^n \bar{V}_i(y_i - r_{i,j}^{(k)}(x))$, is also convex in $\mathcal{I}_{k,j}$ and
- we find a lower bound in $\mathcal{I}_{k,j}$ for the system potential $V(x; \mathbf{y}, \mathbf{g}_j)$ at the intersection of the tangents to $V(x; \mathbf{y}, \mathbf{r}_j^{(k)})$ at the limit points of $\mathcal{I}_{k,j}$, i.e, s_{k-1} and s_k.

4. ADAPTIVE REJECTION SAMPLING ALGORITHM

4.1. Algorithm

We introduce a new adaptive rejection sampling scheme. The proposed algorithm follows the steps specified in Section 2 but the function $L(x) = \exp\{-\eta(x)\} \geq \ell(x; \mathbf{y}, \mathbf{g})$ that overbounds the likelihood is computed adaptively, hence we write $L_t(x) = \exp\{-\eta_t(x)\}$, where $t \in \mathbb{N}$ is an iteration index. To be specific, for each partition $\mathcal{B}_j$, let us define the set of support points

$$\mathcal{S}_{t,j} = \{s_1, \ldots, s_{m_{t,j}}\}, \tag{11}$$

where $s_1 < s_2 < \ldots < s_{m_{t,j}} \in \mathcal{B}_j$ are sorted in ascending order. The piecewise-constant overbounding function $L_t(x) \triangleq L_{t,j}(x)$ for $x \in \mathcal{B}_j$, $j = 1, \ldots, q$, is constructed as $L_{t,j}(x) \triangleq \exp\{-\eta_{t,j}(x)\}$ where

$$\eta_{t,j}(x) \triangleq \begin{cases} \gamma_{1,j} & x \in \mathcal{I}_{1,j} = [-\infty, s_1] \\ \gamma_{2,j} & x \in \mathcal{I}_{2,j} = [s_1, s_2] \\ \vdots & \\ \gamma_{m_t,j} & x \in \mathcal{I}_{m_t,j} = [s_{m_{t-1,j}}, s_{m_{t,j}}] \\ \gamma_{m_t+1,j} & x \in \mathcal{I}_{m_t+1,j} = [s_{m_{t,j}}, +\infty], \end{cases} \tag{12}$$

and each $\gamma_{k,j}$, $k = 1, \ldots, m_{t,j} + 1$, is a constant lower bound for the system potential $V(x; \mathbf{y}, \mathbf{g}_j)$ in the corresponding interval $[s_{k-1}, s_k]$, calculated using the procedure described in Section 3.2. In order to draw from the posterior pdf $p(x|\mathbf{y})$, we sample x' from $\pi_t(x) \propto L_t(x)p(x)$ and accept it with probability $\frac{\exp\{-V(x'; \mathbf{y}, \mathbf{g})\}}{L_t(x')}$. If the sample x' is rejected, and $x' \in \mathcal{B}_j$, then we incorporate it into the set of support points $\mathcal{S}_{t+1,j} = \mathcal{S}_{t,j} \cup \{x'\}$.

Note that $L_t(x)$ becomes a tighter approximation of the likelihood function $p(\mathbf{y}|x) \propto \exp\{-V(x; \mathbf{y}, \mathbf{g})\}$ as we incorporate additional support poins. As a consequence, the proposal pdf, $\pi_t(x) \propto L_t(x)p(x)$, becomes closer to the target density, $p(x|\mathbf{y}) \propto p(\mathbf{y}|x)p(x)$, and the probability to accept candidate samples increases. Table 1 summarizes the algorithm. Details on the initialization procedure and sampling from $\pi_t(x)$ follow below.

4.2. Initialization

All stationary points of the system potential $V(x; \mathbf{y}, \mathbf{g}_j)$, for $x \in \mathcal{B}_j$, are contained in the interval $[\min(\mathcal{X}_j), \max(\mathcal{X}_j)]$, as proved in [5]. In particular, the system potential is strictly decreasing for all $x < \min(\mathcal{X}_j)$, and strictly increasing for all $x > \max(\mathcal{X}_j)$. As a consequence, the value $\gamma_{1,j} = V(\min(\mathcal{X}_j); \mathbf{y}, \mathbf{g}_j)$ is a lower bound of V in the interval $\mathcal{I}_{1,j} \triangleq [-\infty, \min(\mathcal{X}_j)]$, whereas $\gamma_{3,j} = V(\max(\mathcal{X}_j); \mathbf{y}, \mathbf{g}_j)$ is a lower bound of V in $\mathcal{I}_{3,j} \triangleq [\max(\mathcal{X}_j), +\infty]$. Furthermore, we can calculate a lower bound $\gamma_{2,j} \leq V(x; \mathbf{y}, \mathbf{g}_j)$ in the interval $[\min(\mathcal{X}_j), \max(\mathcal{X}_j)]$, using the technique in Section 3.2. As a result, the function

$$\eta_{0,j}(x) \triangleq \begin{cases} \gamma_{1,j} & x \in \mathcal{I}_{1,j} = [-\infty, \min(\mathcal{X}_j)] \\ \gamma_{2,j} & x \in \mathcal{I}_{2,j} = [\min(\mathcal{X}_j), \max(\mathcal{X}_j)] \\ \gamma_{3,j} & x \in \mathcal{I}_{3,j} = [\max(\mathcal{X}_j), +\infty] \end{cases} \tag{13}$$

bounds the system potential from below, i.e., $\eta_{0,j}(x) \leq V(x; \mathbf{y}, \mathbf{g}_j)$ for $x \in \mathcal{B}_j$, while the function $L_{0,j}(x) = \exp\{-\eta_{0,j}(x)\} \geq \exp\{-V(x'; \mathbf{y}, \mathbf{g}_j)\}$, bounds the likelihood from above, also for all $x \in \mathcal{B}_j$.

Table 1. Adaptive rejection sampling (ARS) algorithm.

1. Find an adequate partition $\{\mathcal{B}_j\}_{j=1}^q$ of the space of the SoI.
   Set $t = 0$, $m_0 = 2$, $\mathcal{S}_{0,j} \triangleq \{s_1 = \min(\mathcal{X}_j),\ s_2 = \max(\mathcal{X}_j)\}$.
2. For $j = 1, ..., q$, take the following steps:
   Compute $\gamma_{1,j} = V(s_1, \mathbf{y}, \mathbf{g}_j)$ and $\gamma_{m_t+1,j} = V(s_{m_{t,j}}, \mathbf{y}, \mathbf{g}_j)$.
   Compute the bounds $\gamma_{k,j}$ in $\mathcal{I}_{k,j} = (s_{k-1}, s_k)$,
   for $k = 2, \ldots, m_{t,j}$, using the technique in Section 3.2.
   Set $\eta_{t,j}(x) = \gamma_{k,j}$ for all $x \in \mathcal{I}_{k,j}$, with $k = 1, \ldots, m_t + 1$.
   Set $L_t(x) = L_{t,j}(x) = \exp\{-\eta_{t,j}(x)\}$ for all $x \in \mathcal{B}_j$.
3. Draw x' from $\pi_t(x) \propto L_t(x)p(x)$, and u' from $\mathcal{U}([0, 1])$.
4. If $u' \leq \frac{\exp\{-V(x'; \mathbf{y}, \mathbf{g})\}}{L_t(x')}$ then accept x'. If more samples are needed go to step 3.
5. Otherwise, if $u' > \frac{\exp\{-V(x'; \mathbf{y}, \mathbf{g})\}}{L_t(x')}$, then reject x'.
   If $x' \in \mathcal{B}_{j^*}$, then set $\mathcal{S}_{t+1,j^*} = \mathcal{S}_{t,j^*} \cup \{x'\}$ and update $m_{t+1,j^*} = m_{t,j^*} + 1$. Let $\mathcal{S}_{t+1,j} = \mathcal{S}_{t,j}$ and $m_{t+1,j} = m_{t,j}$ for all other $j \neq j^*$.
6. For $j = 1, ..., q$, sort $\mathcal{S}_{t+1,j}$ in ascending order, increment $t = t + 1$ and go back to the step 2.

4.3. Drawing from the proposal $\pi_t(x)$

The proposal function $\pi_t(x)$ is a mixture of truncated densities with non-overlapping supports. In order to draw from it, we need to be able to integrate the prior $p(x)$ in a finite interval. Specifically, let us denote

$$\bar{w}_{k,j} \triangleq \int_{\mathcal{I}_{k,j}} L_t(x)p(x)dx = \exp\{-\gamma_{k,j}\} \int_{\mathcal{I}_{k,j}} p(x)dx, \quad (14)$$

for $k = 1, ..., m_t+1$, $j = 1, ..., q$, and define the normalized weights $w_{k,j} = \bar{w}_{k,j} / \sum_{k,j} \bar{w}_{k,j}$. To sample from $\pi_t(x)$, we first draw a pair of random indices $(k, j) \in \{1, ..., m_t + 1\} \times \{1, ..., q\}$, from the probability mass function $p(k, j) = w_{k,j}$, and then draw x' from the prior $p(x)$ restricted to the interval $\mathcal{I}_{k,j}$.

5. EXAMPLE

In order to illustrate the use of the proposed ARS technique we investigate its application to implement an accept/reject particle filter (ARPF) [4]. Let us assume that the SoI is the state of the dynamical system

$$\begin{array}{l} x_k = |0.5x_{k-1} + \vartheta_{0,k}| \\ y_{1,k} = \log(x_k) + \vartheta_{1,k}, \;\; y_{2,k} = x_k^2 + \vartheta_{2,k} \end{array} \quad (15)$$

where $k \in \mathbb{N}$ denotes discrete time, $x_k \in \mathbb{R}^+$ is the (non-negative) system state at time k, $\mathbf{y}_k = [y_{1,k}, y_{2,k}]^\top$ is the k-th observation vector and $\mathbf{g}(x) = [g_1(x) = \log(x), g_2(x) = x^2]^\top$ is the vector-nonlinearity. The noise terms are i.i.d. random variables with Gaussian density $N(\vartheta_{i,k}; 0, 1/2) \propto \exp\{-\vartheta_{i,k}^2\}$.

An ARPF is a recursive algorithm that approximates the filtering pdf $p(x_k|\mathbf{y}_{1:k})$ by exact sampling (as opposed to the standard particle filters, that perform importance sampling). Specifically, let $\{x_{k-1}^{(i)}\}_{i=1}^N$ be a collection of samples from $p(x_{k-1}|\mathbf{y}_{1:k-1})$. If we approximate the predictive density as

$$\begin{aligned} p(x_k|\mathbf{y}_{1:k-1}) &= \int p(x_k|x_{k-1})p(x_{k-1}|\mathbf{y}_{1:k-1})dx_{k-1} \\ &\approx \frac{1}{N}\sum_{i=1}^{N} p(x_k|x_{k-1}^{(i)}) \end{aligned} \quad (16)$$

then the filtering pdf becomes

$$p(x_k|\mathbf{y}_{1:k}) \propto p(\mathbf{y}_k|x_k)\frac{1}{N}\sum_{i=1}^{N} p(x_k|x_{k-1}^{(i)}). \quad (17)$$

We can draw new samples $\{x_k^{(i)}\}_{i=1}^N$ from (17) using a RS scheme and, then, any integral $I(f)$ w.r.t. to the filtering pdf can be approximated as $I(f) \approx \hat{I}(f) = \frac{1}{N}\sum_{i=1}^N f(x_k^{(i)})$. In particular, we can readily apply the proposed ARS method to implement the ARPF. The likelihood at time k is $p(\mathbf{y}_k|x_k) \propto \exp\{-V(x_k; \mathbf{y}_k, \mathbf{g})\}$, where

$$V(x_k; \mathbf{y}_k, \mathbf{g}) = (y_{1,k} - \log(x_k))^2 + \left(y_{2,k} - x_k^2\right)^2 \quad (18)$$

is the system potential. It is important to remark that the system potential defined in (18) can have several minima depending on the observation vector $\mathbf{y}_k$, and, in general, it has a second derivate with non-constant sign. As a consequence the likelihood at time k is, in general, not a log-concave function.

Moreover, since both g_1 and g_2 are strictly monotonic functions, the partition of the support of x_k is trivial ($\mathcal{B} = \mathbb{R}$) and we can draw the partition index from the notation. Instead, we need to indicate the time k. Thus, we apply the ARS to build a sequence of lower-bounding functions $\eta_{t,k}(x_k) \leq V(x_k; \mathbf{y}_k, \mathbf{g})$, $t = 0, 1, 2, ...$, and overbounding functions $L_{t,k}(x_k) = \exp\{-\eta_{t,k}(x_k)\} \geq p(\mathbf{y}_k|x_k)$. We note that, for this system, the piecewise-constant functions $\eta_{t,k}(x_k)$ (and, hence, $L_{t,k}(x_k)$) can be found analytically from the modified potentials $V(x_k; \mathbf{y}_k, \mathbf{r}_t)$.

The ARS method is applied at each time step k, following the outline in Table 1. Sampling from $\pi_t(x_k) = \frac{L_{t,k}(x_k)}{N}\sum_{i=1}^N p(x_k|x_{k-1}^{(i)})$ is carried out as described in Section 4.3, with $p(x_k|\mathbf{y}_{1:k-1})$ playing the role of the prior.

Table 2 shows the mean acceptance rates obtained for the 1st, 2nd, 10th and 20th accepted samples. These rates have been approximated by averaging over 50 time steps in 10,000 independent simulation runs. Recall that each time a sample is rejected it is used to improve the bound $L_{t,k}(x_k)$. Therefore, the acceptance rate increases with the sample ordinal and, for the 20th sample, it is already as high as 88%.

For the sake of comparison, we have also applied the standard bootstrap filter with prior importance function [6] in the same set of simulations. For this algorithm, the mean square error (MSE) achieved in the estimation of x_k using $N = 30$ particles is MSE=0.19, while the ARPF can attain the same performance using only $N = 10$ particles. The ARPF is specially advantageous w.r.t. to the standard bootstrap algorithm when there is a significant discrepancy between the likelihood and prior functions.

Table 2. Acceptance rates.

Particle	1-th	2-th	10-th	20-th
Acceptance Rate	57%	68%	81%	88%

6. CONCLUSION

We have proposed a novel adaptive rejection sampling scheme that can be used to draw exactly from a large family of pdf's, not necessarily log-concave. The new method yields a sequence of proposal pdf's that converge towards the target density and, as a consequence, can attain high acceptance rates. We have successfully applied the proposed technique to the implementation of an accept/reject particle filter.

7. REFERENCES

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

[2] W. R. Gilks and P. Wild, "Adaptive Rejection Sampling for Gibbs Sampling," *Applied Statistics*, vol. 41, no. 2, pp. 337–348, 1992.

[3] N. G. Best W. R. Gilks and K. K. C. Tan, "Adaptive Rejection Metropolis Sampling within Gibbs Sampling," *Applied Statistics*, vol. 44, no. 4, pp. 455–472, 1995.

[4] H. R. Künsch, "Recursive Monte Carlo filters: Algorithms and theoretical bounds," *The Annals of Statistics*, vol. 33, no. 5, pp. 1983–2021, 2005.

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

[6] A. Doucet, N. de Freitas, and N. Gordon, "An introduction to sequential Monte Carlo methods," in *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. de Freitas, and N. Gordon, Eds., chapter 1, pp. 4–14. Springer, 2001.