Generalized rejection sampling schemes and applications in signal processing

Luca Martino *, Joaquín Míguez

Department of Signal Theory and Communications, Universidad Carlos III de Madrid, Avenida de la Universidad 30, 28911 Leganés, Madrid, Spain

ABSTRACT

Bayesian methods and their implementations by means of sophisticated Monte Carlo techniques, such as Markov chain Monte Carlo (MCMC) and particle filters, have become very popular in signal processing over the last years. However, in many problems of practical interest these techniques demand procedures for sampling from probability distributions with non-standard forms, hence we are often brought back to the consideration of fundamental simulation algorithms, such as rejection sampling (RS). 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 introduce new schemes for (a) obtaining upper bounds for likelihood functions and (b) adaptively computing proposal densities that approximate the target pdf closely. The former class of methods provides the tools to easily sample from *a posteriori* probability distributions (that appear very often in signal processing problems) by drawing candidates from the prior distribution. However, they are even more useful when they are exploited to derive the generalized adaptive RS (GARS) algorithm introduced in the second part of the paper. The proposed GARS method yields a sequence of proposal densities that converge towards the target pdf and enable a very efficient sampling of a broad class of probability distributions, possibly with multiple modes and non-standard forms. We provide some simple numerical examples to illustrate the use of the proposed techniques, including an example of target localization using range measurements, often encountered in sensor network applications.

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 [5,7,16,17]. Indeed, Monte Carlo statistical methods are powerful tools for numerical inference and optimization [17]. Currently, there exist several classes of MC techniques, including the popular Markov chain Monte Carlo (MCMC) [3,7,15] and particle filtering [4,5] families of algorithms, which enjoy numerous applications. However, in many problems of practical interest these techniques demand procedures for sampling from probability distributions with non-standard forms, hence we are often brought back to the consideration of fundamental simulation algorithms, such as importance

---

* Corresponding author.
   E-mail addresses: luca@tsc.uc3m.es (L. Martino), joaquin.miguez@uc3m.es (J. Míguez).

sampling [2], inversion procedures [17] and the accept/reject method, also known as rejection sampling (RS).

The RS approach [17, Chapter 2] is a classical Monte Carlo technique 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 density. 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 [10,11,14]. An important limitation of RS methods is the need to analytically establish a bound for the ratio of the target and proposal densities, since there is a lack of general methods for the computation of exact bounds.

One exception is the so-called *adaptive rejection sampling* (ARS) method [8,11,17] which, given a target density, provides a procedure to obtain both a suitable proposal pdf (easy to draw from) and the upper bound for the ratio of the target density over this proposal. Unfortunately, this procedure is only valid when the target pdf is strictly log-concave, which is not the case in most practical cases. Although an extension has been proposed [6,12] that enables the application of the ARS algorithm with *T*-concave distributions (where *T* is a monotonically increasing function, not necessarily the logarithm), it does not address the main limitations of the original method (e.g., the impossibility to draw from multimodal distributions) and is hard to apply, due to the difficulty to find adequate *T* transformations other than the logarithm. Another algorithm, called *adaptive rejection metropolis sampling* (ARMS) [9], is an attempt to extend the ARS to multimodal densities by adding Metropolis–Hastings steps. However, the use of an MCMC procedure has two important consequences. First, the resulting samples are correlated (unlike in the original ARS method), and, second, for multimodal distributions the Markov Chain often tends to get trapped in a single mode.

In this paper we propose general procedures to apply RS when the target pdf is the posterior density of a signal of interest (SoI) given a collection of observations. Unlike the ARS technique, our methods can handle target pdf's with several modes (hence non-log-concave) and, unlike the ARMS algorithm, they do not involve MCMC steps. Hence, the resulting samples are independent and come exactly from the target pdf.

We first tackle the problem of computing an upper bound for the likelihood of the SoI given fixed observations. The proposed solutions, that include both closed-form bounds and iterative procedures, are useful when we draw the candidate samples from the prior pdf.

In the second part of the paper, we extend our approach to devise a generalization of the ARS method that can be applied to a broad class of pdf's, possibly multimodal. The generalized algorithm yields an efficient proposal density, tailored to the target density, that can attain a much better acceptance rate than the prior distribution. We remark that the accepted samples are independent and identically distributed (i.i.d.) drawings from the target pdf.

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 propose a general procedure to compute upper bounds for a large family of likelihood functions. The ARS method is briefly reviewed in Section 5, while the generalization of the ARS algorithm, is introduced in Section 6. The use of the proposed technique with multidimensional posterior densities is discussed in Section 7. Section 8 is devoted to simple numerical examples and we conclude with a brief summary in Section 9.

2. Model and problem statement

2.1. Notation

Scalar magnitudes are denoted using regular face letters, e.g., x, X, while vectors are displayed as bold-face letters, e.g., $\mathbf{x}$, $\mathbf{X}$. We indicate random variates with upper-case letters, e.g., X, $\mathbf{X}$, while we use lower-case letters to denote the corresponding realizations, e.g., x, $\mathbf{x}$. We use letter p to denote the true probability density function (pdf) of a random variable or vector. This is an argument-wise notation, common in Bayesian analysis. For two random variables X and Y, $p(x)$ is the true pdf of X and $p(y)$ is the true pdf of Y, possibly different. The conditional pdf of X given $Y=y$ is written $p(x|y)$. Sets are denoted with calligraphic upper-case letters, e.g., $\mathcal{R}$.

2.2. Signal model

Many problems in science and engineering involve the estimation of an unobserved SoI, $\mathbf{x} \in \mathbb{R}^m$, from a sequence of related observations. We assume an arbitrary prior probability density function (pdf) for the SoI, $\mathbf{X} \sim p(\mathbf{x})$, and consider n scalar random 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

$$Y_1 = g_1(\mathbf{X}) + \Theta_1, \ldots, Y_n = g_n(\mathbf{X}) + \Theta_n, \tag{1}$$

where $\mathbf{Y} = [Y_1, \ldots, Y_n]^\top \in \mathbb{R}^n$ is the random observation vector, $g_i : \mathbb{R}^m \to \mathbb{R}$, $i = 1, \ldots, n$, are nonlinearities and Θ_i are independent noise variables, possibly with different distributions for each i. We write $\mathbf{y} = [y_1, \ldots, y_n]^\top \in \mathbb{R}^n$ for the vector of available observations, i.e., a realization of $\mathbf{Y}$.

We assume exponential-type noise pdf's, of the form

$$\Theta_i \sim p(\vartheta_i) = k_i \exp\{-\overline{V}_i(\vartheta_i)\}, \tag{2}$$

where $k_i > 0$ is a real constant and $\overline{V}_i(\vartheta_i)$ is a function, subsequently referred to as marginal potential, with the following properties:

(P1) It is real and non-negative, i.e., $\overline{V}_i : \mathbb{R} \to [0, +\infty)$.
(P2) It is increasing ($d\overline{V}_i/d\vartheta_i > 0$) for $\vartheta_i > 0$ and decreasing ($d\overline{V}_i/d\vartheta_i < 0$) for $\vartheta_i < 0$.

These conditions imply that $\overline{V}_i(\vartheta_i)$ has a unique minimum at $\vartheta_i^* = 0$ and, as a consequence, $p(\vartheta_i)$ has only one maximum (mode) at $\vartheta_i^* = 0$. Since the noise variables are independent, the joint pdf $p(\vartheta_1, \vartheta_2, \ldots, \vartheta_n) = \prod_{i=1}^n p(\vartheta_n)$ is easy to construct and we can define a joint potential function $V^{(n)} : \mathbb{R}^n \to [0, +\infty)$ as

$$V^{(n)}(\vartheta_1, \ldots, \vartheta_n) \triangleq -\log[p(\vartheta_1, \ldots, \vartheta_n)] = -\sum_{i=1}^n \log[p(\vartheta_n)]. \quad (3)$$

Substituting (2) into (3) yields

$$V^{(n)}(\vartheta_1, \ldots, \vartheta_n) = c_n + \sum_{i=1}^n \overline{V}_i(\vartheta_i), \quad (4)$$

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

$$V_l^{(n)}(\vartheta_1, \ldots, \vartheta_n) = \sum_{i=1}^n |\vartheta_i|^l, \quad 0 < l < +\infty, \quad (5)$$

where the subscript l identifies the specific member of the class. In particular, the function obtained for $l=2$, $V_2^{(n)}(\vartheta_1, \ldots, \vartheta_n) = \sum_{i=1}^n |\vartheta_i|^2$ is termed quadratic potential.

Let $\mathbf{g} = [g_1, \ldots, g_n]^\top$ be the vector-valued nonlinearity defined as $\mathbf{g}(\mathbf{x}) \triangleq [g_1(\mathbf{x}), \ldots, g_n(\mathbf{x})]^\top$. The scalar observations are conditionally independent given a realization of the SoI, $\mathbf{X} = \mathbf{x}$, hence the *likelihood function* $\ell(\mathbf{x}; \mathbf{y}, \mathbf{g}) \triangleq p(\mathbf{y}|\mathbf{x})$, can be factorized as

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

where $p(y_i|\mathbf{x}) = k_i \exp\{-\overline{V}_i(y_i - g_i(\mathbf{x}))\}$. The likelihood in (6) induces a system potential function $V(\mathbf{x}; \mathbf{y}, \mathbf{g}) : \mathbb{R}^m \to [0, +\infty)$, defined as

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

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

$$V(\mathbf{x}; \mathbf{y}, \mathbf{g}) = c_n + \sum_{i=1}^n \overline{V}_i(y_i - g_i(\mathbf{x})) = V^{(n)}(y_1 - g_1(\mathbf{x}), \ldots, y_n - g_n(\mathbf{x})). \quad (8)$$

2.3. Rejection sampling

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 $p(\mathbf{x}|\mathbf{y})$, so we need to apply simulation techniques to generate adequate samples. One appealing possibility is to perform RS using the prior, $p(\mathbf{x})$, as a proposal function. In such case, let γ be a lower bound for the system potential, $\gamma \leq V(\mathbf{x}; \mathbf{y}, \mathbf{g})$, so that $L \triangleq \exp\{-\gamma\}$ is an upper bound for the likelihood, $\ell(\mathbf{x}; \mathbf{y}, \mathbf{g}) \leq L$. We can generate N samples according to the standard RS algorithm.

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

Then, $I(f)$ can be approximated as $I(f) \approx \hat{I}(f) = (1/N) \sum_{i=1}^N f(\mathbf{x}^{(i)})$. The fundamental figure of merit of a rejection sampler is the acceptance rate, i.e., the mean number of accepted samples over the total number of proposed candidates.

In Section 4, we address the problem of analytically calculating the bound $L = \exp\{-\gamma\}$. 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 γ for $V(\mathbf{x}; \mathbf{y}, \mathbf{g})$. Note that this problem is far from trivial. Even for very simple marginal potentials, $\overline{V}_i$, $i=1, \ldots, n$, the system potential can be highly multimodal w.r.t. $\mathbf{x}$. See the example in Section 8.1 for an illustration.

3. Definitions 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 techniques to be described in Sections 4–6 can be extended to the general case, although this extension is not trivial. We briefly discuss how to deal with multidimensional random variates in Section 7, and we provide a numerical example in Section 8.3.

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., $\mathcal{A} = \bigcup_{j=1}^q \mathcal{B}_j$ and $\mathcal{B}_i \cap \mathcal{B}_j = \emptyset, \forall i \neq j$) such that the subsets $\mathcal{B}_j$ are intervals in $\mathbb{R}$ and we can define functions $g_{i,j} : \mathcal{B}_j \to \mathbb{R}, j = 1, \ldots, q$ and $i=1, \ldots, n$, as

$$g_{i,j}(x) \triangleq g_i(x), \quad \forall x \in \mathcal{B}_j, \quad (9)$$

i.e., $g_{i,j}$ is the restriction of g_i to the interval $\mathcal{B}_j$. We further assume that (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 $dg_{i,j}/dx$ is either strictly positive or strictly negative and the second derivative $d^2 g_{i,j}/dx^2$ is either non-negative or non-positive.

For a given vector of observations $\mathbf{Y} = \mathbf{y}$ and an interval $\mathcal{B}_j, j \in \{1, \ldots, q\}$ in the partition of the support $\mathcal{A}$, we define the set of simple estimates of SoI as

$$\mathcal{X}_j \triangleq \left\{ x_{i,j} \in \mathbb{R} : x_{i,j} = \underset{x \in \mathcal{B}_j}{\arg\min} |g_i(x) - y_i|, i = 1, \ldots, n \right\}. \quad (10)$$

Let $[\mathcal{B}_j]$ denote the closure of $\mathcal{B}_j$ [13, Chapter 2]. For every observation

$$y_i \in \left(\underset{x \in [\mathcal{B}_j]}{\min} g_{i,j}(x), \underset{x \in [\mathcal{B}_j]}{\max} g_{i,j}(x) \right),$$

$i = 1,\ldots,n$, there exists a single estimate of the form $x_{i,j} = g_{i,j}^{-1}(y_i)$ (note that $g_{i,j}$ is constructed to be invertible). However, if the observation y_i is "out of range", i.e., $y_i < \min_{x \in [\mathcal{B}_j]} g_{i,j}(x)$ or $y_i > \max_{x \in [\mathcal{B}_j]} g_{i,j}(x)$, the inverse $g_{i,j}^{-1}(y_i)$ does not exist. In such cases, the simple estimates still exist but have the form

$$x_{i,j} = \operatorname*{argmin}_{x \in [\mathcal{B}_j]} g_{i,j}(x) \quad \left(\text{if } y_i < \min_{x \in [\mathcal{B}_j]} g_{i,j}(x) \right) \tag{11}$$

and

$$x_{i,j} = \operatorname*{argmax}_{x \in [\mathcal{B}_j]} g_{i,j}(x) \quad \left(\text{if } y_i > \max_{x \in [\mathcal{B}_j]} g_{i,j}(x) \right). \tag{12}$$

Specifically, if the nonlinearity has a horizontal asymptote we obtain that either $\operatorname{argmin}_{x \in [\mathcal{B}_j]} g_{i,j}(x) = \pm \infty$ or $\operatorname{argmax}_{x \in [\mathcal{B}_j]} g_{i,j}(x) = \pm \infty$. We admit these infinite values as valid simple estimates and the techniques proposed below still work if there exists some $x_{i,j} = \pm \infty$. As a consequence, every set of simple estimates $\mathcal{X}_j, j = 1,\ldots,q$, has exactly n elements, $x_{1,j},\ldots,x_{n,j}$ (one per observation).

4. Computation of upper bounds on the likelihood

4.1. Basic method

Let $\mathbf{y}$ be an arbitrary but fixed realization of the observation vector $\mathbf{Y}$. Our goal is to obtain an analytical method for the computation of a scalar $\gamma(\mathbf{y}) \in \mathbb{R}$ such that $\gamma(\mathbf{y}) \leq \inf_{x \in \mathbb{R}} V(x; \mathbf{y}, \mathbf{g})$. Hereafter, we omit the dependence on the observation vector and write simply γ. 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 $\mathcal{B}_j, j = 1,\ldots,q$, in the partition of $\mathcal{A}$. Within $\mathcal{B}_j$, we 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

$$|y_i - r_{i,j}(x)| \leq |y_i - g_{i,j}(x)|, \tag{13}$$

and

$$(y_i - r_{i,j}(x))(y_i - g_{i,j}(x)) \geq 0 \tag{14}$$

hold jointly for all $i = 1,\ldots,n$, and all $x \in \mathcal{I}_j \subset \mathcal{B}_j$, where $\mathcal{I}_j$ is any closed interval in $\mathcal{B}_j$ such that $\hat{x}_j \in \operatorname{argmin}_{x \in [\mathcal{B}_j]} V(x; \mathbf{y}, \mathbf{g})$ (i.e., any maximum likelihood estimator of the SoI X restricted to $\mathcal{B}_j$, possibly non-unique) is contained in $\mathcal{I}_j$. The latter requirement can be fulfilled if we choose $\mathcal{I}_j \triangleq [\min(\mathcal{X}_j), \max(\mathcal{X}_j)]$ (see the Appendix for a proof).

If (13) and (14) hold, we can write

$$\overline{V}_i(y_i - r_{i,j}(x)) \leq \overline{V}_i(y_i - g_{i,j}(x)), \quad \forall x \in \mathcal{I}_j, \tag{15}$$

which follows easily from the properties (P1) and (P2) of the marginal potential functions $\overline{V}_i$ as described in Section 2.2. Moreover, since $V(x; \mathbf{y}, \mathbf{g}_j) = c_n + \sum_{i=1}^{n} \overline{V}_i(y_i - g_{i,j}(x))$ and $V(x; \mathbf{y}, \mathbf{r}_j) = c_n + \sum_{i=1}^{n} \overline{V}_i(y_i - r_{i,j}(x))$ (this function will be subsequently referred as the modified system potential) where $\mathbf{g}_j(x) \triangleq [g_{1,j}(x),\ldots,g_{n,j}(x)]$ and $\mathbf{r}_j(x) \triangleq [r_{1,j}(x),\ldots,r_{n,j}(x)]$, Eq. (15) implies that $V(x; \mathbf{y}, \mathbf{r}_j) \leq V(x; \mathbf{y}, \mathbf{g}_j), \forall x \in \mathcal{I}_j$,

and, as a consequence,

$$\gamma_j = \inf_{x \in \mathcal{I}_j} V(x; \mathbf{y}, \mathbf{r}_j) \leq \inf_{x \in \mathcal{I}_j} V(x; \mathbf{y}, \mathbf{g}_j) = \inf_{x \in \mathcal{B}_j} V(x; \mathbf{y}, \mathbf{g}). \tag{16}$$

Therefore, it is possible to find a lower bound in $\mathcal{B}_j$ 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 $\mathcal{I}_j$.

All that remains is to actually build the linear functions $\{r_{i,j}\}_{i=1}^{n}$. This construction is straightforward and can be described graphically by splitting the problem into two cases. Case 1 corresponds to nonlinearities $g_{i,j}$ such that $dg_{i,j}(x)/dx \times d^2 g_{i,j}(x)/dx^2 \geq 0$ (i.e., $g_{i,j}$ is either increasing and convex or decreasing and concave), while case 2 corresponds to functions that comply with $dg_{i,j}(x)/dx \times d^2 g_{i,j}(x)/dx^2 \leq 0$ (i.e., $g_{i,j}$ is either increasing and concave or decreasing and convex), when $x \in \mathcal{B}_j$.

Fig. 1(a)–(b) depicts the construction of $r_{i,j}$ in case 1. We choose a linear function $r_{i,j}$ that connects the point $(\min(\mathcal{X}_j), g(\min(\mathcal{X}_j)))$ and the point corresponding to the simple estimate, $(x_{i,j}, g(x_{i,j}))$. 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 that $d_r \leq d_g$ for all $x \in \mathcal{I}_j$, hence inequality (13) is granted. Inequality (14) also holds for all $x \in \mathcal{I}_j$, since $r_{i,j}(x)$ and $g_{i,j}(x)$ are either simultaneously greater than (or equal to) y_i, or simultaneously lesser than (or equal to) y_i.

Fig. 1(c)–(d) depicts the construction of $r_{i,j}$ in case 2. We choose a linear function $r_{i,j}$ that connects the point $(\max(\mathcal{X}_j), g(\max(\mathcal{X}_j)))$ and the point corresponding to the simple estimate, $(x_{i,j}, g(x_{i,j}))$. Again, 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 (13) and (14) hold for all $x \in \mathcal{I}_j$.

A special subcase of 1 (respectively, of 2) occurs when $x_i = \min(\mathcal{X}_j)$ (respectively, $x_{i,j} = \max(\mathcal{X}_j)$). Then, $r_{i,j}(x)$ is the tangent to $g_{i,j}(x)$ at $x_{i,j}$. If $x_{i,j} = \pm \infty$ then $r_{i,j}(x)$ is a horizontal asymptote of $g_{i,j}(x)$.

It is often possible to find $\gamma_j = \inf_{x \in \mathcal{I}_j} V(x; \mathbf{y}, \mathbf{r}_j) \leq \inf_{x \in \mathcal{I}_j} V(x; \mathbf{y}, \mathbf{g}_j)$ 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, that will be subsequently referred to as bounding method 1 (BM1) for conciseness.

4.2. Iterative implementation

The quality of the bound γ_j depends, to a large extent, on the length of the interval $\mathcal{I}_j$, denoted $|\mathcal{I}_j|$. This is clear if we think of $r_{i,j}(x)$ as a linear approximation on $\mathcal{I}_j$ of the nonlinearity $g_{i,j}(x)$. Since we have assumed $g_{i,j}(x)$ is continuous and bounded in $\mathcal{I}_j$, the procedure to build $r_{i,j}(x)$ in BM1 implies that

$$\lim_{|\mathcal{I}_j| \to 0} |g_{i,j}(x) - r_{i,j}(x)| \leq \lim_{|\mathcal{I}_j| \to 0} |\sup_{x \in \mathcal{I}_j} g_{i,j}(x) - \inf_{x \in \mathcal{I}_j} g_{i,j}(x)| = 0, \tag{17}$$

for all $x \in \mathcal{I}_j$. Therefore, if we consider intervals $\mathcal{I}_j$ which are shorter and shorter, then the modified potential function $V(x; \mathbf{y}, \mathbf{r}_j)$ will be closer and closer to the true potential function $V(x; \mathbf{y}, \mathbf{g}_j)$, and hence the bound $\gamma_j \leq V(x; \mathbf{y}, \mathbf{r}_j) \leq V(x; \mathbf{y}, \mathbf{g}_j)$ will be tighter.
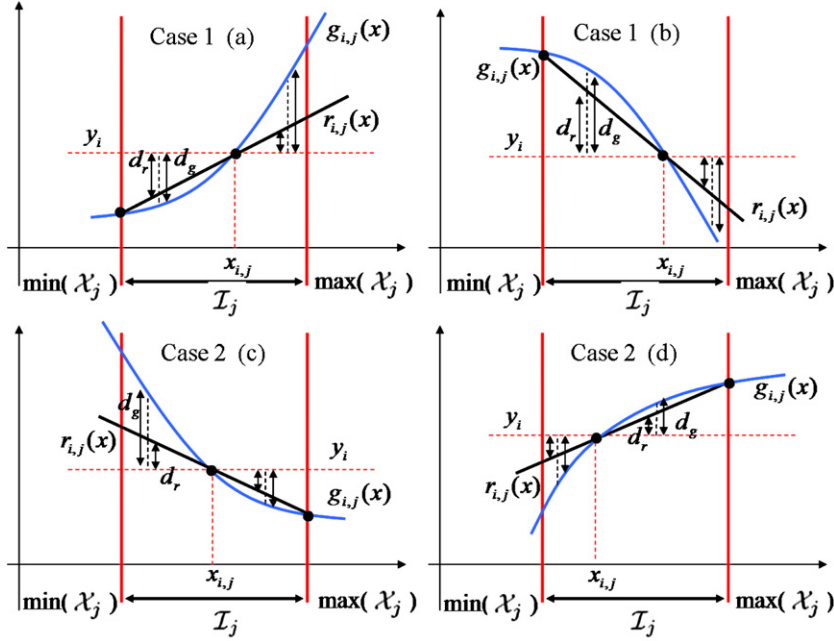
4

Fig. 1. Construction of the auxiliary linear functions $\{r_{i,j}\}_{i=1}^{n}$. We indicate $d_r = |y_i - r_{i,j}(x)|$ and $d_g = |y_i - g_{i,j}(x)|$, respectively. It is apparent that $d_r \leq d_g$ and $r_{i,j}(x)$ and $g_{i,j}(x)$ are either simultaneously greater than (or equal to) y_i, or simultaneously lesser than (or equal to) y_i, for all $x \in \mathcal{I}_j$. Hence, the inequalities (13) and (14) are satisfied $\forall x \in \mathcal{I}_j$. (a) Function $g_{i,j}$ is increasing and convex (case 1). (b) Function $g_{i,j}$ is decreasing and concave (case 1). (c) Function $g_{i,j}$ is decreasing and convex (case 2). (d) Function $g_{i,j}$ is increasing and concave (case 2).

Table 1
Bounding method 1.

---

1. Find a partition $\{\mathcal{B}_j\}_{j=1}^{q}$ of the space of the SoI.
2. Compute the simple estimates $\mathcal{X}_j = \{x_{1,j}, \ldots, x_{n,j}\}$ for each $\mathcal{B}_j$.
3. Calculate $\mathcal{I}_j \triangleq [\min(\mathcal{X}_j), \max(\mathcal{X}_j)]$ and build $r_{i,j}(x)$, for $x \in \mathcal{I}_j$ 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 $\gamma = \min_j \gamma_j$.

Table 2
Iterative algorithm to improve γ_j.

---

1. Start with $\mathcal{I}_{1,2} = \mathcal{I}_j$, and $\mathcal{S}_{j,1} = \{\min(\mathcal{X}_j), \max(\mathcal{X}_j)\}$. Let $\nu^* = 1$ and $k = 1$.
2. Choose an arbitrary interior point s^* in $\mathcal{I}_{\nu*, \nu*+1}$, and update the set of support points $\mathcal{S}_{j,k} = \mathcal{S}_{j,k-1} \cup \{s^*\}$.
3. Sort $\mathcal{S}_{j,k}$ in ascending order, so that $\mathcal{S}_{j,k} = \{s_1, \ldots, s_{k+1}\}$ where $s_1 = \min(\mathcal{X}_j)$, $s_{k+1} = (\max \mathcal{X}_j)$, and $k+1$ is the number of elements of $\mathcal{S}_{j,k}$.
4. Build $\mathbf{r}_j^{(\nu)}(x)$ for each interval $\mathcal{I}_{\nu, \nu+1} = [s_\nu, s_{\nu+1}]$ with $\nu = 1, \ldots, k$.
5. Find $\gamma_j^{(\nu)} = \min V(x; \mathbf{y}, \mathbf{r}_j^{(\nu)})$, for $\nu = 1, \ldots, k$.
6. Set the refined bound $\gamma_{j,k} = \min_{\nu \in \{1, \ldots, k\}} \gamma_j^{(\nu)}$, and set $\nu_* = \operatorname{argmin}_\nu \gamma_j^{(\nu)}$.
7. To iterate, go back to step 2.

The latter observation suggests a procedure to improve the bound γ_j for a given interval $\mathcal{I}_j$. Indeed, let us subdivide $\mathcal{I}_j$ into k subintervals denoted $\mathcal{I}_{\nu, \nu+1} \triangleq [s_\nu, s_{\nu+1}]$ where $\nu = 1, \ldots, k$ and $s_\nu, s_{\nu+1} \in \mathcal{I}_j$. We refer to the elements in the collection $\mathcal{S}_{j,k} = \{s_1, \ldots, s_{k+1}\}$, with $s_1 < s_2 < \cdots < s_{k+1}$, as support points in the interval $\mathcal{I}_j$. We can build linear functions $\mathbf{r}_j^{(\nu)} = [r_{1,j}^{(\nu)}, \ldots, r_{n,j}^{(\nu)}]$ for every subinterval $\mathcal{I}_{\nu, \nu+1}$, using the procedure described in Section 4.1. We recall that this procedure is graphically depicted in Fig. 1, where we simply need to

- substitute $\mathcal{I}_j$ by $\mathcal{I}_{\nu, \nu+1}$ and
- when the simple estimate $x_{i,j} \notin \mathcal{I}_{\nu, \nu+1}$, substitute $x_{i,j}$ by s_ν ($x_{i,j}$ by $s_{\nu+1}$) if $x_{i,j} < s_\nu$ (if $x_{i,j} > s_{\nu+1}$, respectively).

Using $\mathbf{r}_j^{(\nu)}$ we compute a bound $\gamma_j^{(\nu)}$, $\nu = 1, \ldots, k$, and then select $\gamma_{j,k} = \min_{\nu \in \{1, \ldots, k\}} \gamma_j^{(\nu)}$. Note that the subscript k in $\gamma_{j,k}$ indicates how many support points have been used to compute the bound in $\mathcal{I}_j$ (which becomes tighter as k increases). Moreover if we take a new (arbitrary) support point s^* from the subinterval $\mathcal{I}_{\nu*, \nu*+1}$ that contains $\gamma_{j,k}$,

and extend the set of support points with it, $\mathcal{S}_{j,k+1} = \{s_1, \ldots, s^*, \ldots, s_{k+2}\}$ with $s_1 < s_2 < \cdots < s^* < \cdots < s_{k+2}$, then we can iterate the proposed procedure and obtain a refined version of the bound, denoted $\gamma_{j,k+1}$.

The proposed iterative algorithm is described, with detail, in Table 2. Note that k is an iteration index that makes explicit the number of support points s_ν. If we plug this iterative procedure for the computation of γ_j into BM1 (specifically, replacing steps 3 and 4 of Table 1), we obtain a new technique that we will hereafter term bounding method 2 (BM2).

As an illustration, Fig. 2 shows four steps of the iterative algorithm. In Fig. 2(a) there are two support points $\mathcal{S}_{j,1} = \{\min(\mathcal{X}_j), \max(\mathcal{X}_j)\}$, which yield a single interval $\mathcal{I}_{1,2} = \mathcal{I}_j$. In Figs. 2(b)–(d), we successively add a point s^* chosen in the interval $\hat{\mathcal{I}}_{\nu*, \nu*+1}$ that contains the
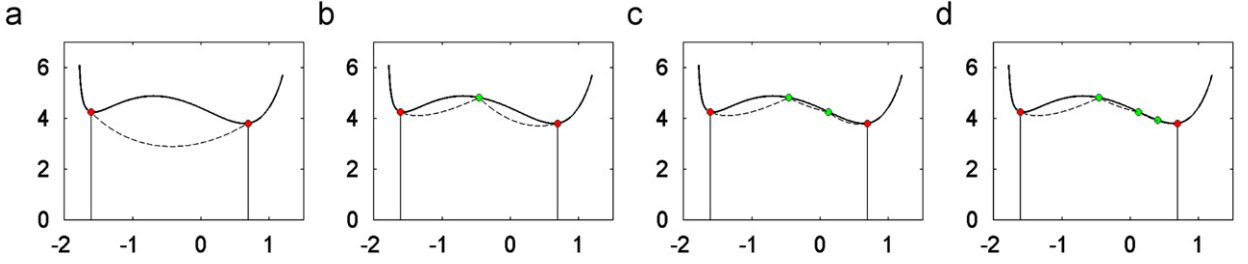
Fig. 2. Four steps of the iterative algorithm choosing s_* as the middle point of the subinterval $\mathcal{I}_{v*,v*+1}$. The solid line shows the system potential $V(x;\mathbf{y},\mathbf{g})=(y_1-\exp(x))^2-\log(y_2-\exp(-x)+1)+(y_2-\exp(-x))+1$ (see the example in Section 8.1), with y_1=5 and y_2=2, while the dashed line shows the modified potential $V(x;\mathbf{y},\mathbf{r}_j)$. We start in plot (a) with two points $\mathcal{S}_{j,1}=\{\min(\mathcal{X}_j),\max(\mathcal{X}_j)\}$. At each iteration, we add a new point chosen in the subinterval $\mathcal{I}_{v*,v*+1}$ that contains the latest bound. It is apparent that $V(x;\mathbf{y},\mathbf{r}_j)$ becomes a better approximation of $V(x;\mathbf{y},\mathbf{g}_j)$ each time we add a new support point.

latest bound. In this example, the point $s*$ is chosen deterministically as the mean of the extremes of the interval $\mathcal{I}_{v*,v*+1}$.

4.3. 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,…,q, and construct the set of linear functions $r_{i,j}(x)=a_{i,j}x+b_{i,j}$, for i=1,…,n and j=1,…,q. The modified system potential in $\mathcal{B}_j$ becomes

$$V_2(x;\mathbf{y},\mathbf{r}_j)=\sum_{i=1}^{n}(y_i-r_{i,j}(x))^2=\sum_{i=1}^{n}(y_i-a_{i,j}x-b_{i,j})^2, \quad (18)$$

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

$$\tilde{x}_j=\arg\min_{x\in\mathcal{B}_j}V(x;\mathbf{y},\mathbf{r}_j)=\frac{\mathbf{a}_j^\top\mathbf{w}_j}{\mathbf{a}_j^\top\mathbf{a}_j}, \quad (19)$$

and $\gamma_{2,j}=V(\tilde{x}_j;\mathbf{y},\mathbf{r}_j)$. It is apparent that $\gamma_2=\min_j\gamma_{2,j}\leq V(x;\mathbf{y},\mathbf{g})$. Furthermore, $\tilde{x}_j$ is an approximation of the maximum likelihood estimator restricted to $\mathcal{B}_j$.

4.4. Adaptation of γ_2 for generic system potentials

If the joint potential is not quadratic, in general it can still be difficult to minimize the modified function $V(x;\mathbf{y},\mathbf{r})$, despite the replacement of the nonlinearities $g_{i,j}(x)$ with the linear functions $r_{i,j}(x)$. In this section, we propose a method to transform the bound for a quadratic potential, γ_2, into a bound for some other, non-quadratic, potential function.

Consider an arbitrary joint potential $V^{(n)}$ and assume the availability of an invertible increasing 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

$$(R\circ V)(x;\mathbf{y},\mathbf{g})\geq V_2^{(n)}(y_1-g_1(x),\ldots,y_n-g_n(x))=\sum_{i=1}^{n}(y_i-g_i(x))^2\geq\gamma_2. \quad (20)$$

and, as consequence, $V(x;\mathbf{y},\mathbf{g})\geq R^{-1}(\gamma_2)=\gamma$, hence γ is a lower bound for the non-quadratic system potential $V(x;\mathbf{y},\mathbf{g})$ constructed from $V^{(n)}$.

For instance, consider the family of joint potentials $V_p^{(n)}$. Using the monotonicity of $\mathcal{L}^p$ norms, it is possible to prove [18] that

$$\left(\sum_{i=1}^{n}|\vartheta_i|^p\right)^{1/p}\geq\left(\sum_{i=1}^{n}\vartheta_i^2\right)^{1/2}, \quad 0\leq p\leq 2, \quad (21)$$

and

$$n^{((p-2)/2p)}\left(\sum_{i=1}^{n}|\vartheta_i|^p\right)^{1/p}\geq\left(\sum_{i=1}^{n}\vartheta_i^2\right)^{1/2}, \quad 2\leq p\leq+\infty. \quad (22)$$

Let $R_1(v)=v^{2/p}$. Since this function is, indeed, strictly increasing, we can transform the inequality (21) into

$$R_1\left(\sum_{i=1}^{n}|y_i-g_i(x)|^p\right)\geq\sum_{i=1}^{n}(y_i-g_i(x))^2, \quad (23)$$

which yields

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

$$=\left(\sum_{i=1}^{n}(y_i-g_i(x))^2\right)^{p/2}\geq\gamma_2^{p/2}, \quad (24)$$

hence the transformation $\gamma_2^{p/2}$ of the quadratic bound γ_2 is a lower bound for $V_p^{(n)}$ with $0<p\leq 2$. Similarly, if we let $R_2(v)=(n^{((p-2)/2p)}v^{1/p})^2$, the inequality (22) yields

$$\sum_{i=1}^{n}|y_i-g_i(x)|^p\geq R_2^{-1}\left(\sum_{i=1}^{n}(y_i-g_i(x))^2\right)$$

$$=\left[n^{(-(p-2)/2p)}\left(\sum_{i=1}^{n}(y_i-g_i(x))^2\right)^{1/2}\right]^p\geq n^{(-(p-2)/2)}\gamma_2^{p/2}, \quad (25)$$

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 $2\leq p<+\infty$.

It is possible to devise a systematic procedure to find a suitable function R given an arbitrary joint potential $V^{(n)}(\boldsymbol{\vartheta})$, where $\boldsymbol{\vartheta}\triangleq[\vartheta_1,\ldots,\vartheta_n]^T$. Let us define the manifold $\Gamma_v\triangleq\{\boldsymbol{\vartheta}\in\mathbb{R}^n:V^{(n)}(\boldsymbol{\vartheta})=v\}$. We can construct R by assigning $R(v)$ with the maximum of the quadratic potential $\sum_i^n\vartheta_i^2$
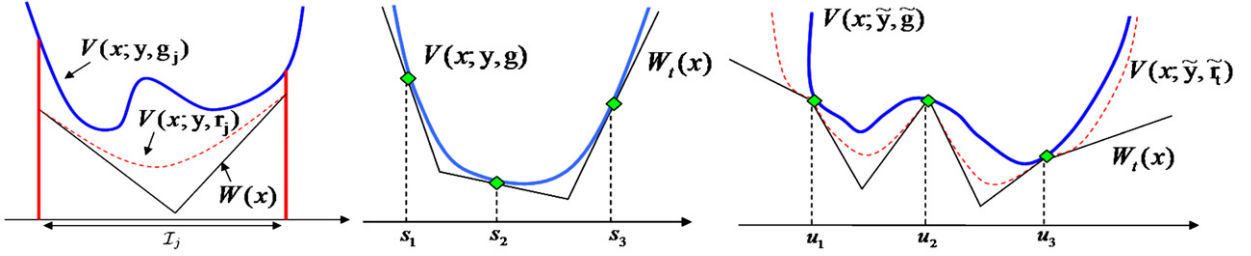
6

Fig. 3. Left: The intersection of the tangents to $V(x;\mathbf{y},\mathbf{r}_j)$ (dashed line) at $\min(\mathcal{X}_j)$ and $\max(\mathcal{X}_j)$ is a lower bound for $V(x;\mathbf{y},\mathbf{g}_j)$ (solid line). Moreover, note that the resulting piecewise-linear function $W(x)$ satisfies the inequality $V(x;\mathbf{y},\mathbf{g}_j) \geq V(x;\mathbf{y},\mathbf{r}_j) \geq W(x)$, for all $x \in \mathcal{I}_j$. Middle: Example of construction of the piecewise-linear function $W_t(x)$ with three support points $\mathcal{S}_t = \{s_1, s_2, s_3\}$, as carried out in the ARS technique. The function $W_t(x)$ is formed from segments of linear functions tangent to $V(x;\mathbf{y},\mathbf{g})$ at the support points in $\mathcal{S}_t$. Right: Construction of the piecewise linear function $W_t(x)$ as tangent lines to the modified potential $V(x;\tilde{\mathbf{y}},\tilde{\mathbf{r}}_t)$ at three intersection points u_1, u_2 and u_3, as carried out in the ARS technique.

when $\boldsymbol{\vartheta} \in \Gamma_v$, i.e., we define

$$R(v) \triangleq \max_{\boldsymbol{\vartheta} \in \Gamma_v} \sum_{i=1}^{n} \vartheta_i^2. \tag{26}$$

Note that (26) is a constrained optimization problem that can be solved using, e.g., Lagrangian multipliers.

From the definition in (26) we obtain that, $\forall \boldsymbol{\vartheta} \in \Gamma_v$, $R(v) \geq \sum_{i=1}^{n} \vartheta_i^2$. In particular, since $V^{(n)}(\boldsymbol{\vartheta}) = v$ from the definition of Γ_v, we obtain the desired relationship,

$$R(V^{(n)}(\vartheta_1, \ldots, \vartheta_n)) \geq \sum_{i=1}^{n} \vartheta_i^2. \tag{27}$$

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.

4.5. Convex marginal potentials $\overline{V}_i$

Assume that $\mathcal{A} = \{\mathcal{B}_j\}_{j=1}^{q}$ and that we have already found $r_{i,j}(x) = a_{i,j}x + b_{i,j}$, $i = 1, \ldots, n$ and $j = 1, \ldots, q$, using the technique in Section 4.1. If a marginal potential $\overline{V}_i(\vartheta_i)$ is convex, the function $\overline{V}_i(y_i - r_{i,j}(x))$ is also convex in $\mathcal{B}_j$. Indeed, for all $x \in \mathcal{B}_j$

$$\frac{d^2\overline{V}_i(y_i - r_{i,j}(x))}{dx^2} = \frac{d^2 r_{i,j}}{dx^2}\frac{d\overline{V}_i}{d\vartheta_i} + \left(\frac{dr_{i,j}}{dx}\right)^2 \frac{d^2\overline{V}_i}{d\vartheta_i^2} = 0 + a_i^2 \frac{d^2\overline{V}_i}{d\vartheta_i^2} \geq 0, \tag{28}$$

where we have used that $d^2 r_{i,j}/dx^2 = 0$ (since $r_{i,j}$ is linear).

As a consequence, if all marginal potentials $\overline{V}_i(\vartheta_i)$ are convex, then the modified system potential, $V(x;\mathbf{y},\mathbf{r}_j) = c_n + \sum_{i=1}^{n} \overline{V}_i(y_i - r_{i,j}(x))$, is also convex in $\mathcal{B}_j$. This is easily shown using (28), to obtain

$$\frac{d^2 V(x;\mathbf{y},\mathbf{r}_j)}{dx^2} = \sum_{i=1}^{n} a_i^2 \frac{d^2\overline{V}_i}{d\vartheta_i^2} \geq 0, \quad \forall x \in \mathcal{B}_j. \tag{29}$$

Therefore, we can use the tangents to $V(x;\mathbf{y},\mathbf{r}_j)$ at the limit points of $\mathcal{I}_j$ (i.e., $\min(\mathcal{X}_j)$ and $\max(\mathcal{X}_j)$) to find a lower bound for the system potential $V(x;\mathbf{y},\mathbf{g}_j)$. Fig. 3 (left) depicts a system potential $V(x;\mathbf{y},\mathbf{g}_j)$ (solid line), the corresponding modified potential $V(x;\mathbf{y},\mathbf{r}_j)$ (dotted line) and the two tangent lines at $\min(\mathcal{X}_j)$ and $\max(\mathcal{X}_j)$. It is apparent that the intersection of the two tangents yields a lower bound in $\mathcal{B}_j$. Specifically, if we let $W(x)$ be the

piecewise-linear function composed of the two tangents, then the inequality $V(x;\mathbf{y},\mathbf{g}_j) \geq V(x;\mathbf{y},\mathbf{r}_j) \geq W(x)$ is satisfied for all $x \in \mathcal{I}_j$.

5. Adaptive rejection sampling

The adaptive rejection sampling (ARS) [11] algorithm enables the construction of a sequence of proposal densities, $\{\pi_t(x)\}_{t\in\mathbb{N}}$, tailored to the target density. Its most appealing feature is that each time we draw a sample from a proposal π_t and it is rejected, we can use this sample to build an improved proposal, π_{t+1}, with a higher mean acceptance rate.

Unfortunately, this attractive ARS method can only be applied with target pdf's which are log-concave (hence, unimodal), which is a very stringent constraint for many practical applications. Next, we briefly review the ARS algorithm and then proceed to introduce its extension for non-log-concave and multimodal target densities.

Let $p(x|\mathbf{y})$ denote the target pdf.[1] The ARS procedure can be applied when $\log[p(x|\mathbf{y})]$ is concave, i.e., when the potential function $V(x;\mathbf{y},\mathbf{g}) \triangleq -\log[p(x|\mathbf{y})] + c_0$ is strictly convex. Let $\mathcal{S}_t = \{s_1, s_2, \ldots, s_{k_t}\}$ be a set of support points in the domain D of $V(x;\mathbf{y},\mathbf{g})$. From $\mathcal{S}_t$ we build a piecewise-linear lower hull of $V(x;\mathbf{y},\mathbf{g})$, denoted $W_t(x)$, formed from segments of linear functions tangent to $V(x;\mathbf{y},\mathbf{g})$ at the support points in $\mathcal{S}_t$. Fig. 3 (middle) illustrates the construction of $W_t(x)$ with three support points for a generic log-concave potential function $V(x;\mathbf{y},\mathbf{g})$.

Once $W_t(x)$ is built, we can use it to obtain an exponential-family proposal density

$$\pi_t(x) = c_t \exp[-W_t(x)], \tag{30}$$

where c_t is the proportionality constant. Therefore $\pi_t(x)$ is piecewise-exponential and very easy to sample from. Since $W_t(x) \leq V(x;\mathbf{y},\mathbf{g}) \forall x$, we trivially obtain that $(1/c_t)\pi_t(x) \geq \exp\{-V(x;\mathbf{y},\mathbf{g})\} \propto p(x|\mathbf{y})$ and we can apply the RS principle.

When a sample x' from $\pi_t(x)$ is rejected we can incorporate it into the set of support points, $\mathcal{S}_{t+1} = \mathcal{S}_t \cup \{x'\}$ (and $k_{t+1} = k_{t+1}$). Then we compute a refined

---

[1] The method does not require that the target density be a posterior pdf, but we prefer to keep the same notation as in the previous section for coherence.

Table 3
Adaptive rejection sampling algorithm.

---
1. Start with $t=0$, $\mathcal{S}_0 = \{s_1, s_2\}$ where $s_1 < s_2$ and the derivatives of $V(x,\mathbf{y},\mathbf{g})$ at s_1,s_2 have different signs.
2. Build the piecewise-linear function $W_t(x)$ as shown in Fig. 3 (middle), using the tangent lines to $V(x;\mathbf{y},\mathbf{g})$ at the support points in $\mathcal{S}_t$.
3. Sample x' from $\pi_t(x) \propto \exp\{-W_t(x)\}$, and u' from $\mathcal{U}([0,1])$.
4. If $u' \leq \dfrac{\exp\{-V(x';\mathbf{y},\mathbf{g})\}}{\exp\{-W_t(x')\}}$ then accept x' and set $\mathcal{S}_{t+1} = \mathcal{S}_t$, $k_{t+1} = k_t$.
5. Otherwise, if $u' > \dfrac{\exp\{-V(x';\mathbf{y},\mathbf{g})\}}{\exp\{-W_t(x')\}}$, then reject x', set $\mathcal{S}_{t+1} = \mathcal{S}_t \cup \{x'\}$ and update $k_{t+1} = k_t + 1$.
6. Sort $\mathcal{S}_{t+1}$ in ascending order, increment t and go back to step 2.

---

lower hull, $W_{t+1}(x)$, and a new proposal density $\pi_{t+1}(x) = c_{t+1}\exp\{-W_{t+1}(x)\}$. Table 3 summarizes the ARS algorithm.

6. Generalization of the ARS method

In this section we introduce a generalization of the standard ARS scheme that can cope with a broader class of target pdf's, including many multimodal distributions. The standard algorithm of [11], described in Table 3, is a special case of the method described below.

6.1. Generalized adaptive rejection sampling

We wish to draw samples from the posterior $p(x|\mathbf{y})$. For this purpose, we assume that

- the prior pdf has the form $p(x) \propto \exp\{-\overline{V}_{n+1}(\mu - g_{n+1}(x))\}$,
- all marginal potential functions, $\overline{V}_i(\vartheta_i)$, $i=1,\ldots,n+1$, are strictly convex,
- the nonlinearities $g_i(x)$, $i=1,\ldots,n+1$, are either convex or concave, not necessarily monotonic.

We incorporate the information of the prior by defining an extended observation vector, $\tilde{\mathbf{y}} \triangleq [y_1,\ldots,y_n,y_{n+1}=\mu]^\top$, and an extended vector of nonlinearities, $\tilde{\mathbf{g}}(x) \triangleq [g_1(x),\ldots,g_n(x),g_{n+1}(x)]^\top$. As a result, we introduce the extended system potential function

$$V(x;\tilde{\mathbf{y}},\tilde{\mathbf{g}}) \triangleq V(x;\mathbf{y},\mathbf{g}) + \overline{V}_{n+1}(\mu - g_{n+1}(x)) = -\log[p(x|\mathbf{y})] + c_0, \quad (31)$$

where c_0 accounts for the superposition of constant terms that do not depend on x. We remark that the function $V(x;\tilde{\mathbf{y}},\tilde{\mathbf{g}})$ constructed in this way is not necessarily convex. It can present several minima and, as a consequence, $p(x|\mathbf{y})$ can present several maxima.

Our technique is adaptive, i.e., it is aimed at the construction of a sequence of proposals, denoted $\pi_t(x)$, $t \in \mathbb{N}$, but relies on the same basic arguments already exploited to devise the BM1. To be specific, at the t-th iteration of the algorithm we seek to replace the nonlinearities $\{g_i\}_{i=1}^{n+1}$ by piecewise-linear functions $\{r_{i,t}\}_{i=1}^{n+1}$ in such a way that the inequalities

$$|y_i - r_{i,t}(x)| \leq |y_i - g_i(x)| \quad (32)$$

and

$$(y_i - r_{i,t}(x))(y_i - g_i(x)) \geq 0 \quad (33)$$

are satisfied $\forall x \in \mathbb{R}$. Therefore, we repeat the same conditions as in Eqs. (13) and (14) but the derivation of the generalized ARS (GARS) algorithm does not require the partition of the SoI space, as it was needed for the BM1.

We will show that it is possible to construct adequate piecewise-linear functions of the form

$$r_{i,t}(x) \triangleq \begin{cases} \max[\overline{r}_{i,1}(x),\ldots,\overline{r}_{i,K_t}(x)] & \text{if } g_i \text{ is convex,} \\ \min[\overline{r}_{i,1}(x),\ldots,\overline{r}_{i,K_t}(x)] & \text{if } g_i \text{ is concave,} \end{cases} \quad (34)$$

where $i=1,\ldots,n+1$ and each $\overline{r}_{i,j}(x)$, $j=1,\ldots,K_t$, is a purely linear function. The number of linear functions involved in the construction of $r_{i,t}(x)$ at the t-th iteration of the algorithm, denoted K_t, determines how tightly $\pi_t(x)$ approximates the true density $p(x|\mathbf{y})$ and, therefore, the higher K_t, the higher mean acceptance rate of the sampler. In Section 6.2 below, we explicitly describe how to choose the linear functions $\overline{r}_{i,j}(x)$, $j=1,\ldots,K_t$, in order to ensure that (32) and (33) hold. We will also show that, when a proposed sample x' is rejected, K_t can be increased ($K_{t+1} = K_t + 1$) to improve the acceptance rate.

Let $\tilde{\mathbf{r}}_t \triangleq [r_{1,t}(x),\ldots,r_{n,t}(x),r_{n+1,t}(x)]^\top$ be the extended vector of piecewise-linear functions, that yields the modified potential $V(x;\tilde{\mathbf{y}},\tilde{\mathbf{r}}_t)$. The same argument used in Section 4.1 to derive the BM1 shows that, if (32) and (33) hold, then $V(x;\tilde{\mathbf{y}},\tilde{\mathbf{r}}_t) \leq V(x;\tilde{\mathbf{y}},\tilde{\mathbf{g}})$, $\forall x \in \mathbb{R}$. Finally, we build a piecewise-linear lower hull $W_t(x)$ for the modified potential, as explained below, to obtain $W_t(x) \leq V(x;\tilde{\mathbf{y}},\tilde{\mathbf{r}}_t) \leq V(x;\tilde{\mathbf{y}},\tilde{\mathbf{g}})$.

The definition of the piecewise-linear function $r_{i,t}(x)$ in (34) can be rewritten in another form

$$r_{i,t}(x) \triangleq \overline{r}_{i,j}(x) \quad \text{for } x \in [a,b], \quad (35)$$

where a is the abscissa of the intersection between the linear functions $\overline{r}_{i,j-1}(x)$ and $\overline{r}_{i,j}(x)$, and b is the abscissa of the intersection between $\overline{r}_{i,j}(x)$ and $\overline{r}_{i,j+1}(x)$. Therefore, we can define the set of all abscissas of intersection points

$$\mathcal{E}_t = \{u \in \mathbb{R} : \overline{r}_{i,j}(u) = \overline{r}_{i,j+1}(u) \text{ for } i=1,\ldots,n+1, \\ j=1,\ldots,K_t-1\}, \quad (36)$$

and sort them in ascending order

$$u_1 < u_2 < \cdots < u_{Q_t}, \quad (37)$$

where Q_t is the total number of intersections. Then

(a) since we have assumed that the marginal potentials are convex, we can use Eq. (35) and the argument of Section 4.5 to show that the modified function $V(x;\tilde{\mathbf{y}},\tilde{\mathbf{r}}_t)$ is convex in each interval $[u_q,u_{q+1}]$, with $q=1,\ldots,Q_t$, and,
(b) as a consequence, we can to build $W_t(x)$ by taking the linear functions tangent to $V(x;\tilde{\mathbf{y}},\tilde{\mathbf{r}}_t)$ at every intersection point u_q, $q=1,\ldots,Q_t$.

Fig. 3 (right) depicts the relationship among $V(x;\tilde{\mathbf{y}},\tilde{\mathbf{g}})$, $V(x;\tilde{\mathbf{y}},\tilde{\mathbf{r}}_t)$ and $W_t(x)$. Since $W_t(x)$ is piecewise linear, the corresponding pdf $\pi_t(x) \propto \exp\{-W_t(x)\}$ is piecewise exponential and can be easily used in a rejection sampler

(we remark that $W_t(x) \leq V(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}})$, hence $\pi_t(x) \propto \exp\{-W_t(x)\} \geq \exp\{-V(x; \tilde{\mathbf{y}}, \tilde{\mathbf{g}})\}$).

Next subsection is devoted to the derivation of the linear functions needed to construct $\tilde{\mathbf{r}}_t$. Then, we describe how the algorithm is iterated to obtain a sequence of improved proposal densities and provide a pseudo-code. We also discuss the increase the mean of acceptance rate as a tentative densities $\pi_t(x)$ converge to the target pdf $p(x|\mathbf{y})$. Finally, we describe a limitation of the procedure that yields improper proposals in a specific scenario.

6.2. Construction of linear functions $\bar{r}_{i,j}(x)$

A basic element in the description of the GARS algorithm in the previous section is the construction of the linear functions $\bar{r}_{i,j}(x)$. This issue is addressed below. For clarity, we consider two cases corresponding to non-monotonic and monotonic nonlinearities, respectively. It is important to remark that the nonlinearities $g_i(x)$, $i=1,\dots,n+1$, can belong to different cases.

6.2.1. Non-monotonic nonlinearities

Assume $g_i(x)$ is a non-monotonic, either concave or convex, function. We have three possible scenarios depending on the number of simple estimates for $g_i(x)$: (a) there exist two simple estimates, $x_{i,1} < x_{i,2}$, (b) there exists a single estimate, $x_{i,1}=x_{i,2}$, or (c) there is no solution for the equation $y_i=g_i(x)$.

Let us assume that $x_{i,1} < x_{i,2}$ and denote $\mathcal{J}_i \triangleq [x_{i,1},x_{i,2}]$. Let us also introduce a set of support points $\mathcal{S}_t \triangleq \{s_1,\dots,s_{k_t}\}$ that contains at least the simple estimates and an arbitrary point $s \in \mathcal{J}_i$, i.e., $x_{i,1},x_{i,2} \in \mathcal{S}_t$. The number of support points, k_t, determines the accuracy of the approximation of the nonlinearity $g_i(x)$ that can be achieved with the piecewise-linear function $r_{i,t}(x)$. In Section 6.3 we show how this number increases as the GARS algorithm iterates. Now, we assume it is given and fixed.
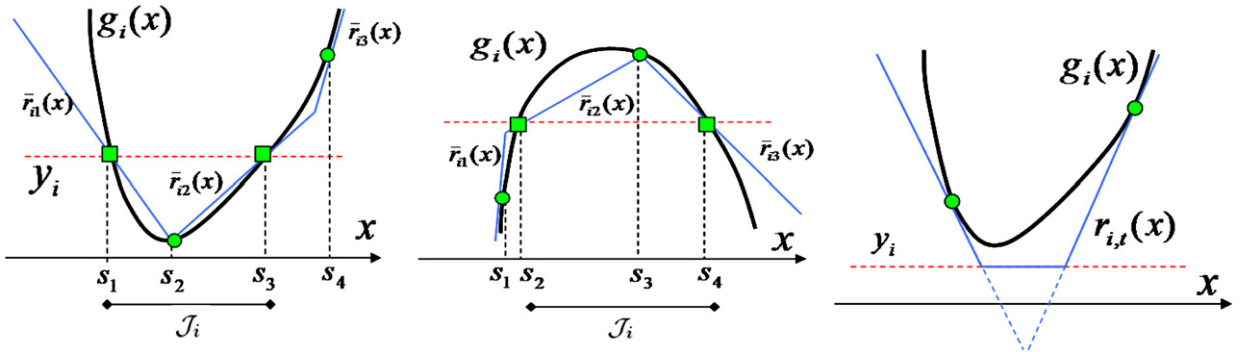
Fig. 4 illustrates the construction of $\bar{r}_{i,j}(x)$, $j=1,\dots,K_t$ where $K_t=k_t-1$, and $r_{i,t}(x)$ for a convex nonlinearity $g_i(x)$ (the procedure is completely analogous for concave $g_i(x)$). Assume that the two simple estimates $x_{i,1} < x_{i,2}$ exist, hence $|\mathcal{J}_i| > 0$. For each $j \in \{1,\dots,k_t\}$, the linear function $\bar{r}_{i,j}(x)$ is constructed in one out of two ways:

(a) if $[s_j,s_{j+1}] \subseteq \mathcal{J}_i$, then $\bar{r}_{i,j}(x)$ connects the points $(s_j,g_i(s_j))$ and $(s_{j+1},g_i(s_{j+1}))$, else
(b) if $s_j \notin \mathcal{J}_i$, then $\bar{r}_{i,j}(x)$ is tangent to $g_i(x)$ at $x=s_j$.

From Fig. 4 (left) it is apparent that $r_{i,t}(x) = \max[\bar{r}_{i,1}(x),\dots,\bar{r}_{i,K_t}(x)]^\top$ built in this way satisfies the inequalities (32) and (33), as required. For concave $g_i(x)$, see Fig. 4 (middle), the conditions (32) and (33) are satisfied if we choose $r_{i,t}(x) = \min[\bar{r}_{i,1}(x),\dots,\bar{r}_{i,K_t}(x)]^\top$.

When $|\mathcal{J}_i| = 0$ (i.e., $x_{i,1}=x_{i,2}$ or there is no solution for the equation $y_i=g_i(x)$), then each $\bar{r}_{i,j}(x)$ is tangent to $g_i(x)$ at $x=s_j$, $\forall s_j \in \mathcal{S}_t$, and in order to satisfy (32) and (33), we need to select

$$r_{i,t}(x) \triangleq \begin{cases} \max[\bar{r}_{i,1}(x),\dots,\bar{r}_{i,K_t}(x),y_i] & \text{if } g_i \text{ is convex,} \\ \min[\bar{r}_{i,1}(x),\dots,\bar{r}_{i,K_t}(x),y_i] & \text{if } g_i \text{ is concave} \end{cases} \quad (38)$$

as illustrated in Fig. 4 (right).

6.2.2. Monotonic nonlinearities

In this case $g_i(x)$ is invertible and there are two possibilities: there exists a single estimate, $x_i=g_i^{-1}(y_i)$ and there is no solution for the equation $y_i=g_i(x)$ (when y_i does not belong to the range of $g_i(x)$). Similarly to the construction in Section 4.1, we distinguish two cases:

(a) if $dg_i(x)/dx \times d^2g_i(x)/dx^2 \geq 0$, then we define $\mathcal{J}_i \triangleq (-\infty,x_i]$, and
(b) if $dg_i(x)/dx \times d^2g_i(x)/dx^2 \leq 0$, then we define $\mathcal{J}_i \triangleq [x_i,+\infty)$.

Fig. 4. Construction of the piecewise linear function $r_{i,t}(x)$ for non-monotonic functions. The straight lines $\bar{r}_{i,j}(x)$ form a piecewise linear function that is closer to the observation value y_i (dashed line) than the nonlinearity $g_i(x)$, i.e., $|y_i-r_{i,t}(x)| \leq |y_i-g_i(x)|$. Moreover, $r_{i,t}(x)$ and $g_i(x)$ are either simultaneously greater than (or equal to) y_i, or simultaneously lesser than (or equal to) y_i, i.e., $(y_i-r_{i,t}(x))(y_i-g_i(x)) \geq 0$. Therefore, the inequalities (32) and (33) are satisfied. The point $(s_j,g_i(s_j))$, corresponding to support point s_j, is represented either by a square or a circle, depending on whether it is a simple estimate or not, respectively. *Left*: construction of $r_{i,t}(x)$ with $k_t=4$ support points when the nonlinearity $g_i(x)$ is convex, therefore $r_{i,t}(x) = \max[\bar{r}_{i,1}(x),\dots,\bar{r}_{i,3}(x)]$ ($K_t=k_t-1=3$). We use the tangent to $g_i(x)$ at $x=s_4$ because $s_4 \notin \mathcal{J}_i = [s_1,s_3]$, where $s_1=x_{i,1}$ and $s_3=x_{i,2}$ are the simple estimates (represented with squares). *Middle*: since the nonlinearity $g_i(x)$ is concave, $r_{i,t}(x) = \min[\bar{r}_{i,1}(x),\dots,\bar{r}_{i,3}(x)]$. We use the tangent to $g_i(x)$ at s_4 because $s_1 \notin \mathcal{J}_i = [s_2,s_4]$, where $s_2=x_{i,1}$ and $s_4=x_{i,2}$ are the simple estimates (represented with squares). *Right*: construction of the $r_{i,t}(x)$, with two support points, when there are not simple estimates. We use the tangent lines, but we need a correction in the definition of $r_{i,t}(x)$ in order to satisfy the inequalities (32) and (33). Since $g_i(x)$ in the figure is convex, we take $r_{i,t}(x) = \max[\bar{r}_{i,1}(x),\dots,\bar{r}_{i,K_t}(x),y_i]$.

The set of support points is $\mathcal{S}_t \triangleq \{s_1, \ldots, s_{k_t}\}$, with $s_1 < s_2 < \cdots < s_{k_t}$, and includes at least the simple estimate x_i and an arbitrary point $s \in \mathcal{J}_i$, i.e., $x_i, s \in \mathcal{S}_t$.

The procedure to build $\bar{r}_{i,j}(x)$, for $j = 1, \ldots, K_t$, with $K_t = k_t$, is similar to Section 6.2.1. Consider case (a) first. For each $j \in \{2, \ldots, k_t\}$, if $[s_{j-1}, s_j] \subset \mathcal{J}_i = (-\infty, x_i]$, then $\bar{r}_{i,j}(x)$ is the linear function that connects the points $(s_{j-1}, g_i(s_{j-1}))$ and $(s_j, g_i(s_j))$. Otherwise, if $s_j \notin \mathcal{J}_i = (-\infty, x_i]$, $\bar{r}_{i,j}(x)$ is tangent to $g_i(x)$ at $x = s_j$. Finally, we set $\bar{r}_{i,1}(x) = g_i(s_1)$ for all $x \in \mathbb{R}$. The piecewise linear function $r_{i,t}$ is $r_{i,t}(x) = \max[\bar{r}_{i,1}(x), \ldots, \bar{r}_{i,K_t}(x)]$. This construction is depicted in Fig. 5 (left).

Case (b) is similar. For each $j \in \{1, \ldots, k_t\}$, if $[s_j, s_{j+1}] \subset \mathcal{J}_i = [x_i, +\infty)$, then $\bar{r}_{i,j}(x)$ is the linear function that connects the points $(s_j, g_i(s_j))$ and $(s_{j+1}, g_i(s_{j+1}))$. Otherwise, if $s_j \notin \hat{\mathcal{I}}_i = [x_i, +\infty)$, $\bar{r}_{i,j}(x)$ is tangent to $g_i(x)$ at $x = s_j$. Finally, we set $\bar{r}_{i,k_t}(x) = g_i(s_{k_t})$ (remember that, in this case, $K_t = k_t$), for all $x \in \mathbb{R}$. The piecewise linear function $r_{i,t}$ will be $r_{i,t}(x) = \min[\bar{r}_{i,1}(x), \ldots, \bar{r}_{i,K_t}(x)]$. This construction is depicted in Fig. 5 (right).

It is straightforward to check that the inequalities (32) and (33) are satisfied. Note that, in this case, the number of linear functions $\bar{r}_{i,j}(x)$ coincides with the number of support points. If there is no solution for the equation $y_i = g_i(x)$ (y_i does not belong to the range of $g_i(x)$), then (32) and (33) are satisfied if we use (38) to build $r_{i,t}(x)$.
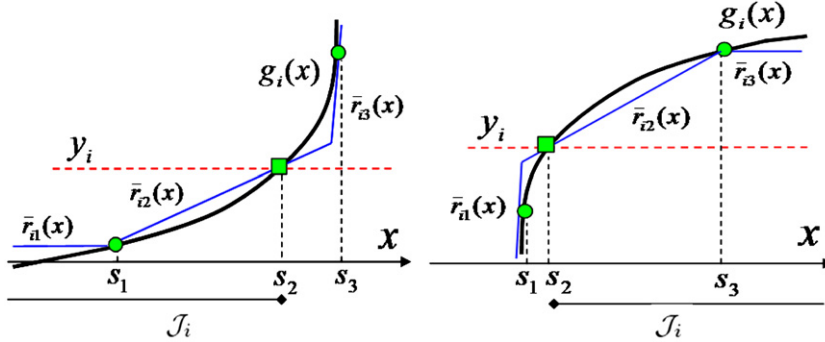
6.3. Outline of the algorithm

We can combine the elements described in Sections 6.2.1 and 6.2.2 into an adaptive algorithm that improves the proposal density $\pi_t(x) \propto \exp\{-W_t(x)\}$ each time a sample is rejected.

Let $\mathcal{S}_t$ denote the set of support points after the t-th iteration. We initialize the algorithm with $\mathcal{S}_0 \triangleq \{s_j\}_{j=1}^{k_0}$ such that

- all simple estimates are contained in $\mathcal{S}_0$, and
- for each interval $\mathcal{J}_i, i = 1, \ldots, n+1$, with non-zero length $(|\mathcal{J}_i| > 0)$, there is at least one (arbitrary) support point contained in $\mathcal{J}_i$.

The proposed GARS algorithm is described in Table 4.

6.4. Convergence

Note that every time a sample x' drawn from $\pi_t(x)$ is rejected, x' is incorporated as a support point in the new set $\mathcal{S}_{t+1} = \mathcal{S}_t \cup \{x'\}$ and, as a consequence, a refined lower hull $W_{t+1}(x)$ is constructed yielding a better approximation of the system potential function. In this way, $\pi_{t+1}(x) \propto \exp$

Fig. 5. Examples of construction of the piecewise-linear function $r_{i,t}(x)$ with $k_t = 3$ support points, for the two subcases. It is apparent that $|y_i - r_{i,t}(x)| \leq |y_i - g_i(x)|$ and that $r_{i,t}(x)$ and $g_i(x)$ are either simultaneously greater than (or equal to) y_i, or simultaneously lesser than (or equal to) y_i, i.e., $(y_i - r_{i,t}(x))(y_i - g_i(x)) \geq 0$. The simple estimates are represented by squares while all other support points are drawn as circles. Left: the figure corresponds to the subcase 1 where $r_{i,t}(x) = \max[\bar{r}_{i,1}(x), \ldots, \bar{r}_{i,3}(x)]$ ($K_t = k_t = 3$). Right: the figure corresponds to the subcase 2 where $r_{i,t}(x) = \min[\bar{r}_{i,1}(x), \ldots, \bar{r}_{i,3}(x)]$ ($K_t = k_t = 3$).

Table 4
Steps of generalized adaptive rejection sampling.

1. Start with $t=0$ set $\mathcal{S}_0 \triangleq \{s_j\}_{j=1}^{k_0}$.
2. Build $\bar{r}_{i,j}(x)$ for $i = 1, \ldots, n+1, j = 1, \ldots, K_t$, where $K_t = k_t - 1$ or $K_t = k_t$ depending on whether $g_i(x)$ is non-monotonic or monotonic, respectively.
3. Calculate the set of intersection points $\mathcal{E}_t \triangleq \{u \in \mathbb{R} : \bar{r}_{i,j}(u) = \bar{r}_{i,j+1}(u)$ for $i = 1, \ldots, n+1, j = 1, \ldots, K_t - 1\}$.
   Let $Q_t = |\mathcal{E}_t|$ be the number of elements in $\mathcal{E}_t$.
4. Build $W_t(x)$ using the tangent lines to $V(x; \tilde{\mathbf{y}}, \tilde{\mathbf{r}}_t)$ at the points $u_q \in \mathcal{E}_t$, $q = 1, \ldots, Q_t$.
5. Draw a sample x' from $\pi_t(x) \propto \exp[-W_t(x)]$.
6. Sample u' from $\mathcal{U}([0,1])$.
7. If $u' \leq \dfrac{\exp\{-V(x'; \mathbf{y}, \mathbf{g})\}}{\exp\{-W_t(x')\}}$ then accept x' and set $\mathcal{S}_{t+1} = \mathcal{S}_t$.
8. Otherwise, if $u' > \dfrac{\exp\{-V(x'; \mathbf{y}, \mathbf{g})\}}{\exp\{-W_t(x')\}}$ then reject x' and update $\mathcal{S}_{t+1} = \mathcal{S}_t \cup \{x'\}$.
9. Sort $\mathcal{S}_{t+1}$ in ascending order, set $t=t+1$ and go back to step 2.

$\{-W_{t+1}(x)\}$ becomes "closer" to $p(x|\mathbf{y})$ and it can be expected that the mean acceptance rate be higher.

To be precise the probability of accepting a sample $x \in \mathcal{A}$ drawn from $\pi_t(x)$ is

$$a_t(x) \triangleq \frac{\exp\{-V(x;\tilde{\mathbf{y}},\tilde{\mathbf{g}})\}}{\exp\{-W_t(x)\}} = \exp\{-[V(x;\tilde{\mathbf{y}},\tilde{\mathbf{g}})-W_t(x)]\}, \qquad (39)$$

and we define the acceptance rate at the t-th iteration of the GARS algorithm, denoted as $\hat{a}_t$, as the expected value of $a_t(x)$ with respect to the density $\pi_t(x)$, i.e.,

$$\hat{a}_t \triangleq E[a_t(x)] = \int_{\mathcal{A}} a_t(x)\pi_t(x)\, dx$$

$$= c_t \int_{\mathcal{A}} \exp\{-V(x;\tilde{\mathbf{y}},\tilde{\mathbf{g}})\}\, dx = \frac{c_t}{c_v} \leq 1, \qquad (40)$$

where c_t and c_v are the proportionality constants for $\pi_t(x)$ and $p(x|\mathbf{y})$, respectively, i.e.,

$$c_t = \left(\int_{\mathcal{A}} \exp\{-W_t(x)\}\, dx\right)^{-1}$$

and

$$c_v = \left(\int_{\mathcal{A}} \exp\{-V(x;\tilde{\mathbf{y}},\tilde{\mathbf{g}})\}\, dx\right)^{-1}.$$

Note that $c_t/c_v \leq 1$ because $W_t(x) \leq V(x;\tilde{\mathbf{y}},\tilde{\mathbf{g}})\ \forall x \in \mathcal{A}$.

From Eq. (40), we obtain that $\hat{a}_t = 1$ if, and only if, $c_t=c_v$ or, equivalently, $\hat{a}_t = 1$ if and only if the integral

$$e(t) \triangleq \int_{\mathcal{A}} [\exp\{-W_t(x)\}-\exp\{-V(x;\tilde{\mathbf{y}},\tilde{\mathbf{g}})\}]\, dx \qquad (41)$$

vanishes, i.e., $e(t)$=0.

The error signal $e(t)$ can be interpreted as a divergence between $\pi_t(x)$ and $p(x|\mathbf{y})$. In particular if $e(t)$ decreases, the acceptance rate $\hat{a}_t = c_t/c_v$ increases and since $\exp\{-W_t(x)\} \geq \exp\{-V(x;\tilde{\mathbf{y}},\tilde{\mathbf{g}})\}\ \forall x \in \mathcal{A}$, $e(t)$=0 if, and only if, $W_t(x) = V(x;\tilde{\mathbf{y}},\tilde{\mathbf{g}})$ almost everywhere. Equivalently, $\hat{a}_t = 1$ if, and only if, $\pi_t(x) = p(x|\mathbf{y})$ almost everywhere. The convergence of $\hat{a}_t$ toward 1 is illustrated numerically in Section 8.2.

6.5. Improper proposals

The GARS algorithm as described in Table 4 breaks down when every $g_i(x)$, i=1,…,n+1, is nonlinear and convex (or concave) monotonic. In this case, the proposed construction procedure yields a piecewise lower hull $W_t(x)$ which is positive and constant in an interval of infinite length. Thus, the resulting proposal, $\pi_t(x) \propto \exp\{-W_t(x)\}$ is improper ($\int_{-\infty}^{+\infty} \pi_t(x)\, dx \to +\infty$) and cannot be used for RS. One practical solution is to substitute the constant piece of $W_t(x)$ by a linear function with a small slope. In that case, $\pi_t(x)$ is proper but we cannot guarantee that the samples drawn using the GARS algorithm come exactly from the target pdf. If we assume, e.g., that $g_{n+1}(x)$=x implying that we choose *a priori* of the form $p(x) \propto \exp\{-\overline{V}_{n+1}(\mu-x)\}$, then we can guarantee that $\pi_t(x)$ be proper.

7. Multidimensional signal of interest

We have restricted our study to the case in which the SoI is scalar, $x \in \mathbb{R}$. A straightforward approach to draw samples from the posterior density $p(\mathbf{x}|\mathbf{y})$ of a random vector $\mathbf{X}$ taking values in $\mathbb{R}^m$ consists in combining the proposed GARS algorithm with a Gibbs sampler.

The Gibbs sampling algorithm to draw from the posterior pdf $p(\mathbf{x}|\mathbf{y}) = p(x_1,\ldots,x_m|\mathbf{y})$ can be summarized as follows:

(1) Set i=1, and draw $\mathbf{x}^{(1)} = [x_1^{(1)},\ldots,x_m^{(1)}]^\top$ from the prior pdf $p(\mathbf{x})$. Set an integer $N > 1$.
(2) Draw samples $x_k^{(i+1)}$, k=1,…,m from the conditional densities

$$x_1^{(i+1)} \sim p(x_1|\mathbf{y},x_2^{(i-1)},x_3^{(i-1)},\ldots,x_m^{(i-1)}),$$

$$x_2^{(i+1)} \sim p(x_2|\mathbf{y},x_1^{(i)},x_3^{(i-1)},x_4^{(i-1)}\ldots,x_m^{(i-1)}),$$

$$x_3^{(i+1)} \sim p(x_3|\mathbf{y},x_1^{(i)},x_2^{(i)},x_4^{(i-1)},\ldots,x_m^{(i-1)})$$

$$\vdots$$

$$x_m^{(i+1)} \sim p(x_m|\mathbf{y},x_1^{(i)},x_2^{(i)},x_3^{(i)},\ldots,x_{m-1}^{(i)}). \qquad (42)$$

(3) Set $\mathbf{x}^{(i+1)} = [x_1^{(i+1)},\ldots,x_m^{(i+1)}]^\top$.
(4) Increment i=i+1. If $i > N$ stop, else go back to step 2.

After a "burning period", the Markov chain $\mathbf{x}^{(1)}, \mathbf{x}^{(2)},\ldots,\mathbf{x}^{(k)},$ …converges to a stationary distribution with density $p(\mathbf{x}|\mathbf{y})$ [7]. However, to apply this technique we need to be able to draw from the posterior conditional pdf's $p(x_k|\mathbf{y},x_1^{(i)},\ldots,x_{k-1}^{(i)},x_{k+1}^{(i-1)},\ldots,x_m^{(i-1)})$, k=1,…,m. In many cases, this is feasible using the proposed GARS technique. An example is presented in Section 8.3.

Alternatively, it is also possible to extend the proposed approach to draw i.i.d. samples directly from $p(\mathbf{x}|\mathbf{y})$, $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^n$, $m \leq n$. Indeed, the general scheme of the algorithm remains the same and only the linear functions $\bar{r}_{i,j}(x)$ need to be extended. Specifically, they are generalized to m-dimensional linear manifolds embedded in an n-dimensional metric space ($\mathbb{R}^n,d$), where d is a suitably defined distance. Unfortunately, a complete description of this extension requires the introduction of a considerable amount of new notations and mathematical tools that fall out of the scope of the present paper.

8. Examples

8.1. Example 1: calculation of upper bounds for the likelihood function

Let X be a scalar SoI with prior density $X \sim p(x) = N(x;0,2)$ and the random observations

$$Y_1 = \exp(X)+\Theta_1, \quad Y_2 = \exp(-X)+\Theta_2, \qquad (43)$$

where Θ_1, Θ_2 are independent noise variables. Specifically, Θ_1 is Gaussian noise with $N(\vartheta_1;0,1/2) = k_1\exp\{-(\vartheta_1)^2\}$, and Θ_2 has a gamma pdf, $\Theta_2 \sim \Gamma(\vartheta_2;\theta,\lambda) = k_2\vartheta_2^{\theta-1}\exp\{-\lambda\vartheta_2\}$, with parameters θ=2, λ=1.

The marginal potentials are $\overline{V}_1(\vartheta_1) = \vartheta_1^2$ and $\overline{V}_2(\vartheta_2) = -\log(\vartheta_2) + \vartheta_2$. Since the minimum of $\overline{V}_2(\vartheta_2)$ occurs in $\vartheta_2 = 1$, we replace Y_2 with the shifted observation $Y_2^* = \exp(-X) + \Theta_2^*$, where $Y_2^* = Y_2 - 1$, $\Theta_2^* = \Theta_2 - 1$. Hence, the marginal potential becomes $\overline{V}_2(\vartheta_2^*) = -\log(\vartheta_2^* + 1) + \vartheta_2^* + 1$, with a minimum at $\vartheta_2^* = 0$, the vector of observations is $\mathbf{Y} = [Y_1, Y_2^*]^\top$ and the vector of nonlinearities is $\mathbf{g}(x) = [\exp(x), \exp(-x)]^\top$. Due to the monotonicity and convexity of g_1 and g_2, we can work with a partition of $\mathbb{R}$ consisting of just one set, $\mathcal{B}_1 \equiv \mathbb{R}$. The joint potential is $V^{(2)}(\vartheta_1, \vartheta_2^*) = \sum_{i=1}^2 \overline{V}_i(\vartheta_i) = \vartheta_1^2 - \ln(\vartheta_2^* + 1) + \vartheta_2^* + 1$ and the system potential is

$$V(x; \mathbf{y}, \mathbf{g}) = V^{(2)}(y_1 - \exp(x), y_2^* - \exp(-x))$$

$$= (y_1 - \exp(x))^2 - \log(y_2^* - \exp(-x) + 1) + (y_2^* - \exp(-x)) + 1. \qquad (44)$$

Assume that, $\mathbf{Y} = \mathbf{y} = [2,5]^\top$. The simple estimates are $\mathcal{X} = \{x_1 = \log(2), x_2 = -\log(5)\}$, and, therefore, we can restrict the search of the bound to the interval $\mathcal{I} = [\min(\mathcal{X}) = -\log(5), \max(\mathcal{X}) = \log(2)]$ (note that we omit the subscript because we have just one set, $\mathcal{B}_1 \equiv \mathbb{R}$). Using the BM1 technique in Section 4.1, we find the linear functions $r_1(x) = 0.78x + 1.45$ and $r_2(x) = -1.95x + 1.85$.

In this case, we can analytically minimize the modified system potential, to obtain $\tilde{x} = -0.4171 = \operatorname{argmin}_{x \in \mathcal{I}} V(x, \mathbf{y}, \mathbf{r})$. The associated lower bound is $\gamma = V(\tilde{x}, \mathbf{y}, \mathbf{r}) = 2.89$ (the true global minimum of the system potential is 3.78). We can also use the technique in Section 4.4 with $R^{-1}(v) = -\log(\sqrt{v} + 1) + \sqrt{v} + 1$. The lower bound for the quadratic potential is $\gamma_2 = 2.79$ and we can readily compute a lower bound $\gamma = R^{-1}(\gamma_2) = 1.68$ for $V(x; \mathbf{y}, \mathbf{g})$. Since the marginal potentials are both convex, we can also use the procedure described in Section 4.5, obtaining the lower bound $\gamma = 1.61$.

Fig. 6(a) depicts the system potential $V(x; \mathbf{y}, \mathbf{g})$, and the lower bounds obtained with the three methods. It is the standard BM1 algorithm that yields the best bound.
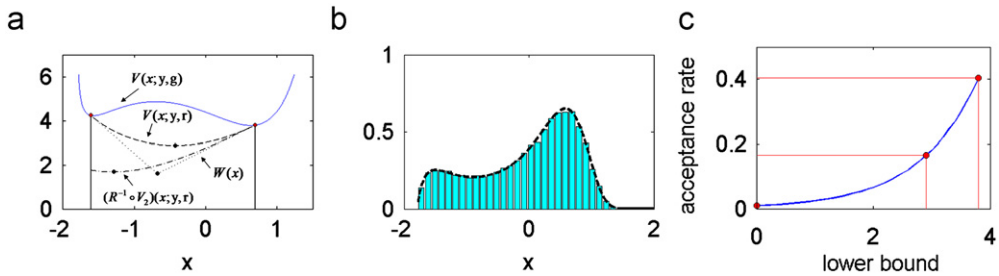
In order to improve the bound, we can use the iterative BM2 technique described in Section 4.2. With only three iterations of BM2, and minimizing analytically the modified potential $V(x, \mathbf{y}, \mathbf{r})$, we find a very tight lower bound $\gamma = \min_{x \in \mathcal{I}}(V(x, \mathbf{y}, \mathbf{r})) = 3.77$ (recall that the optimal bound is 3.78). Table 5 summarizes the bounds computed with the different techniques.

Next, we implement a rejection sampler, using the prior pdf $p(x) = N(x; 0, 2) \propto \exp\{-x^2/4\}$ as a proposal function and the upper bound for the likelihood $L = \exp\{-3.77\}$. The posterior density has the form

$$p(x|\mathbf{y}) \propto p(\mathbf{y}|x)p(x) = \exp\{-V(x; \mathbf{y}, \mathbf{g}) - x^2/4\}. \qquad (45)$$

Fig. 6(b) shows the normalized histogram of $N = 10,000$ samples generated by the RS algorithm, together with the true target pdf $p(x|\mathbf{y})$ depicted as a dashed line. The histogram follows closely the shape of the true posterior pdf. Fig. 6(c) shows the acceptance rates (averaged over 10,000 simulations) as a function of the bound γ. We start with the trivial lower bound $\gamma = 0$ and increase it progressively, up to the global minimum $\gamma = 3.78$. The resulting acceptance rates are 1.1% for the trivial bound $\gamma = 0$, 18% with $\gamma = 2.89$ (BM1) and approximately 40% with $\gamma = 3.77$ (BM2). Note that the acceptance rate is $\approx 41\%$ for the optimal bound and we cannot improve it any further. This is an intrinsic drawback of a rejection sampler with constant bound L and the principal argument that suggests the use of adaptive procedures.

8.2. Example 2: comparison of ARMS and GARS techniques

Consider the problem of sampling a scalar random variable X from a posterior bimodal density $p(x|y) \propto p(y|x)p(x)$, where the likelihood function is $p(y|x) \propto \exp\{-\cosh(y - x^2)\}$ (note that we have a single observation $Y = y_1$) and the prior pdf is $p(x) \propto \exp\{-\alpha(\eta - \exp(|x|))^2\}$, with constant parameters $\alpha > 0$ and η. Therefore, the posterior pdf is $p(x|y) \propto \exp\{-V(x; \tilde{y}, \tilde{\mathbf{g}})\}$,

Fig. 6. (a) The system potential $V(x, \mathbf{y}, \mathbf{g})$ (solid), the modified system potential $V(x, \mathbf{y}, \mathbf{r})$ (dashed), function $(R^{-1} \circ V_2)(x, \mathbf{y}, \mathbf{r})$ (dot-dashed) and the piecewise-linear function $W(x)$ formed by the two tangent lines to $V(x, \mathbf{y}, \mathbf{r})$ at $\min(\mathcal{X})$ and $\max(\mathcal{X})$ (dotted). The corresponding bounds are marked with dark circles. (b) The target density $p(x|\mathbf{y}) \propto p(\mathbf{y}|x)p(x)$ (dashed) and the normalized histogram of $N = 10,000$ samples using RS with the calculated bound L. (c) The curve of acceptance rates (averaged over 10,000 simulations) as a function of the lower bound γ. The acceptance rate is 1.1% for the trivial bound $\gamma = 0$, 18% with $\gamma = 2.89$, approximately 40% with $\gamma = 3.77$ and 41% with the optimal bound $\gamma = 3.78$.

Table 5
Lower bounds of the system potential function.

Method	BM1	BM1 + transformation R	BM1 + tangent lines	BM2	Optimal bound
Lower bound γ	2.89	1.68	1.61	3.77	3.78

where $\tilde{\mathbf{y}} = [y,\eta]^\top$, $\tilde{\mathbf{g}}(x) = [g_1(x),g_2(x)]^\top = [x^2,\exp(|x|)]^\top$ and the extended system potential function becomes

$$V(x;\tilde{\mathbf{y}},\tilde{\mathbf{g}}) = \cosh(y-x^2) + \alpha(\eta-\exp(|x|))^2. \qquad (46)$$

The marginal potentials are $\overline{V}_1(\vartheta_1) = \cosh(\vartheta_1)$ and $\overline{V}_2(\vartheta_2) = \alpha\vartheta_2^2$. Note that the density $p(x|y)$ is an even function, $p(x|y) = p(-x|y)$, hence it has a zero mean, $\mu = \int x p(x|y)\,dx = 0$. The constant α is a scale parameter that allows to control the variance of the random variable X, both *a priori* and *a posteriori*. The higher the value of α, the more skewed the modes of $p(x|y)$ become.

There are no standard methods to sample directly from $p(x|y)$. Moreover, since the posterior density $p(x|y)$ is bimodal, the system potential is non-log-concave and the ARS technique cannot be applied. However, we can easily use the GARS technique. If, e.g., $\tilde{\mathbf{Y}} = \tilde{\mathbf{y}} = [y = 5, \eta = 10]^\top$ then the simple estimates corresponding to $g_1(x)$ are $x_{1,1} = -\sqrt{5}$ and $x_{1,2} = \sqrt{5}$, so that $\mathcal{J}_1 = [-\sqrt{5},\sqrt{5}]$. In the same way, the simple estimates corresponding to $g_2(x)$ are $x_{2,1} = -\log(10)$ and $x_{2,2} = \log(10)$, therefore $\mathcal{J}_2 = [-\log(10),\log(10)]$.

An alternative possibility to draw from this density is to use the ARMS method [9]. Therefore, in this section we compare the two algorithms. Specifically, we look into the accuracy in the approximation of the posterior mean $\mu = 0$ by way of the sample mean estimate, $\hat{\mu} = (1/N)\sum_{i=1}^{N} x^{(i)}$, for different values of the scale parameter α.

In particular, we have considered ten equally spaced values of α in the interval $[0.2,5]$ and then performed 10,000 independent simulations for each value of α, each simulation consisting of drawing 5000 samples with the GARS method and the ARMS algorithm. Both techniques can be sensitive to their initialization. The ARMS technique starts with five points selected randomly in $[-3.5,3.5]$ (with uniform distribution). The GARS starts with the set of support points $\mathcal{S}_0 = \{x_{2,1},x_{1,1},s,x_{1,2},x_{2,2}\}$ sorted in ascending order, including all simple estimates and an arbitrary point s needed to enable the construction in Section 6.2. Point s is randomly chosen in each simulation, with uniform pdf in $\mathcal{J}_1 = [x_{1,1},x_{1,2}]$.

The simulation results show that the two techniques attain similar performance when $\alpha \in [0.2,1]$ (the modes of $p(x|y)$ are relatively flat). When $\alpha \in [1,4]$ the modes become more skewed and the Markov chain generated by the ARMS algorithm remains trapped at one of the two modes in $\approx 10\%$ of the simulations. When $\alpha \in [4,5]$ the same problem occurs in $\approx 25\%$ of the simulations. The performance of the GARS algorithm, on the other hand, is comparatively insensitive to the value of α.

Table 6 illustrates the estimated posterior mean $\hat{\mu}$ in five independent simulations, obtained with the two techniques when $\alpha = 5$. We can see that ARMS remains trapped at the negative mode in the first simulation ($\hat{\mu} = -2.2981$) and at the positive mode ($\hat{\mu} = 2.2994$) in the last simulation.

Fig. 7(a) shows the posterior density $p(x|y) \propto \exp\{-\cosh(y_1-x^2)-\alpha(\mu-\exp(|x|))^2\}$ with $\alpha = 0.2$ depicted as a dashed line, and the normalized histogram obtained samples drawn using the GARS technique.

Fig. 7(b) depicts the acceptance rate $\hat{a}_t$ (averaged over 10,000 simulations) as a function of the iteration index t,

Table 6
Estimated posterior mean, $\hat{\mu}$ (for $\alpha = 5$).

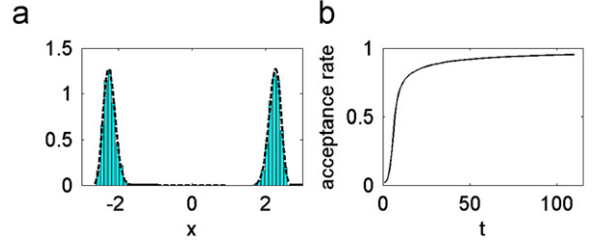| Simulation	1	2	3	4	5
ARMS	−2.2981	0.0267	0.0635	0.0531	2.2994
GARS	0.0772	−0.0143	0.0029	0.0319	0.0709

Fig. 7. (a) The bimodal density $p(x|\mathbf{y}) \propto \exp\{-V(x;\tilde{\mathbf{y}},\tilde{\mathbf{g}})\}$ (dashed line) and the normalized histogram of $N = 5000$ samples obtained using GARS algorithm. (b) The curve of acceptance rate $\hat{a}_t$ (averaged over 10,000 simulations) as a function of the iteration index t.

using the GARS algorithm. Every time a sample x' drawn from $\pi_t(x)$ is rejected, it is incorporated as a support point. Then, the proposal pdf $\pi_t(x)$ becomes closer to target pdf $p(x|y)$ and, as a consequence, the acceptance rate becomes higher. For instance, the acceptance rate for $t=1$ is approximately 1.8%, but for $t=10$, it is already $\approx 71\%$. The acceptance rate $\hat{a}_t$ for $t=100$ is $\approx 95\%$.

8.3. Example 3: target localization with a sensor network

In order to show how the proposed techniques can be used to draw samples from a multivariate (non-scalar) SoI, we consider the problem of positioning a target in a two-dimensional space using range measurements. This is a problem that appears frequently in localization applications using sensor networks [1].

We use a random vector $\mathbf{X} = [X_1,X_2]^\top$ to denote the target position in the plane $\mathbb{R}^2$. The prior density of X is $p(x_1,x_2) = p(x_1)p(x_2)$, where $p(x_i) = N(x_i; 0,1/2) = k \exp\{-(x_i)^2\}$, $i=1,2$, i.e., the coordinate X_1 and X_2 are i.i.d. Gaussian. The range measurements are obtained from two sensor located at $\mathbf{h}_1 = [0,0]^\top$ and $\mathbf{h}_2 = [2,2]^\top$, respectively. The effective observations are the (square) Euclidean distances from the target to the sensors, contaminated with Gaussian noise, i.e.,

$$Y_1 = X_1^2 + X_2^2 + \Theta_1,$$

$$Y_2 = (X_1-2)^2 + (X_2-2)^2 + \Theta_2, \qquad (47)$$

where Θ_i, $i=1,2$, are independent Gaussian variables with identical pdf's, $N(\vartheta_i; 0,1/2) = k_i\exp\{-\vartheta_i^2\}$. Therefore, the marginal potentials are quadratic, $\overline{V}_i(\vartheta_i) = \vartheta_i^2$, $i=1,2$. The random observation vector is denoted $Y = [Y_1,Y_2]^\top$. We note that one needs three range measurements to uniquely determine the position of a target in the plane, so the posterior pdf $p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$ is bimodal.

We apply the Gibbs sampler to draw N particles $\mathbf{x}^{(i)} = [x_1^{(i)}, x_2^{(i)}]^\top$, $i = 1, \ldots, N$, from the posterior density $p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|x_1, x_2) p(x_1) p(x_2)$. The algorithm consists of the following steps:

(1) Set $i = 1$, and draw $x_2^{(1)}$ from the prior pdf $p(x_2)$.
(2) Draw a sample $x_1^{(i)}$ from the conditional pdf $p(x_1|\mathbf{y}, x_2^{(i)})$, and set $\mathbf{x}^{(i)} = [x_1^{(i)}, x_2^{(i)}]^\top$.
(3) Draw a sample $x_2^{(i+1)}$ from the conditional pdf $p(x_2|\mathbf{y}, x_1^{(i)})$.
(4) Increment $i = i + 1$. If $i > N$ stop, else go back to step 2.

The Markov chain generated by the Gibbs sampler converges to a stationary distribution with pdf $p(x_1, x_2|\mathbf{y})$.

In order to use Gibbs sampling, we have to be able to draw from the conditional densities $p(x_1|\mathbf{y}, x_2^{(i)})$ and $p(x_2|\mathbf{y}, x_1^{(i)})$. In general, these two conditional pdf's can be non-log-concave and can have several modes. Specifically, the density $p(x_1|\mathbf{y}, x_2^{(i)}) \propto p(\mathbf{y}|x_1, x_2^{(i)}) p(x_1)$ can be expressed as $p(x_1|\mathbf{y}, x_2^{(i)}) \propto \exp\{-V(x_1; \tilde{\mathbf{y}}_1, \tilde{\mathbf{g}}_1)\}$ where $\tilde{\mathbf{y}}_1 = [y_1 - (x_2^{(i)})^2, y_2 - (x_2^{(i)} - 2)^2, 0]^\top$, $\tilde{\mathbf{g}}_1(x) = [x^2, (x-2)^2, x]^\top$ and

$$V(x_1; \tilde{\mathbf{y}}_1, \tilde{\mathbf{g}}_1) = [y_1 - (x_2^{(i)})^2 - x_1^2]^2 + [y_2 - (x_2^{(i)} - 2)^2 - (x_1 - 2)^2]^2 + x_1^2, \tag{48}$$

while the pdf $p(x_2|\mathbf{y}, x_1^{(i)}) \propto p(\mathbf{y}|x_2, x_1^{(i)}) p(x_2)$ can be expressed as $p(x_2|\mathbf{y}, x_1^{(i)}) \propto \exp\{-V(x_2; \tilde{\mathbf{y}}_2, \tilde{\mathbf{g}}_2)\}$ where $\tilde{\mathbf{y}}_1 = [y_1 - (x_1^{(i)})^2, y_2 - (x_1^{(i)} - 2)^2, 0]^\top$, $\tilde{\mathbf{g}}_2(x) = [x^2, (x-2)^2, x]^\top$ and

$$V(x_2; \tilde{\mathbf{y}}_2, \tilde{\mathbf{g}}_2) = [y_1 - (x_1^{(i)})^2 - x_2^2]^2 + [y_2 - (x_1^{(i)} - 2)^2 - (x_2 - 2)^2]^2 + x_2^2. \tag{49}$$

Since the marginal potentials and the nonlinearities are convex, we can use the GARS technique to sample the conditional pdf's.

We have generated $N = 10,000$ samples from the Markov chain, with fixed observations $y_1 = 5$ and $y_2 = 2$. The average acceptance rate of the GARS algorithm was $\approx 30\%$ both for $p(x_1|\mathbf{y}, x_2)$ and $p(x_2|\mathbf{y}, x_1)$. Note that this rate is indeed as an average because, at each step of the chain, the target pdf's are different (if, e.g., $x_1^{(i)} \neq x_1^{(i-1)}$ then $p(x_2|\mathbf{y}, x_1^{(i)}) \neq p(x_2|\mathbf{y}, x_1^{(i-1)})$).

Fig. 8(a) shows the shape of the true target density $p(x_1, x_2|\mathbf{y})$, while Fig. 8(b) depicts the normalized histogram with $N = 10,000$ samples.

Finally, it is illustrative to consider the computational savings attained by using the GARS method when compared with a rejection sampler with a fixed bound.
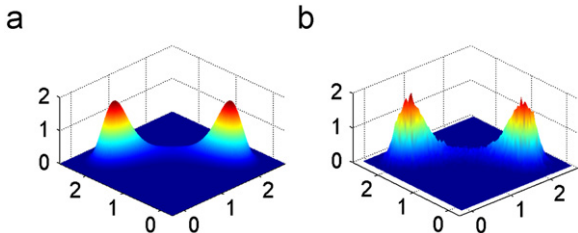
Specifically, we have run again the Gibbs sampler to generate a chain of 10,000 samples but, when drawing from $p(x_1|\mathbf{y}, x_2)$ and $p(x_2|\mathbf{y}, x_1)$, we have used RS with prior proposals ($p(x_1)$ and $p(x_2)$, respectively) and a fixed bound computed (analytically) with the method in Section 4.3 for quadratic potentials. The average acceptance rate for the rejection sampler was $\approx 4\%$ and the time needed to generate the chain was approximately 10 times the time needed in the simulation with the GARS algorithm.

9. Conclusions

We have proposed families of generalized rejection sampling schemes that are particularly, but not only, useful for efficiently drawing independent samples from *a posteriori* probability distributions. The problem of drawing from posterior distributions appears very often in signal processing, e.g., see the target localization example in this paper or virtually any application that involves the estimation of a physical magnitude given a set of observations collected by a sensor network. We have introduced two classes of schemes. The procedures in the first class are aimed at the computation of upper bounds for the likelihood function of the signal of interest given the set of available observations. They provide the means to (quickly and easily) design sampling schemes for posterior densities using the prior pdf as a proposal function. Then, we have elaborated on the bound-calculation procedures to devise a generalized adaptive rejection sampling (GARS) algorithm. The latter is a method to construct a sequence of proposal pdf's that converge towards the target density and, therefore, can attain very high acceptance rates. It should be noted that the method introduced in this paper includes the classical adaptive rejection sampling scheme of [11] as a particular case. We have provided some simple numerical examples to illustrate the use of the proposed techniques, including sampling from multimodal distributions (both with fixed and adaptive proposal functions) and an example of target localization using range measurements. The latter problem is often encountered in positioning applications of sensor networks.

Fig. 8. (a) The target density $p(\mathbf{x}|\mathbf{y}) = p(x_1, x_2|\mathbf{y}) \propto p(\mathbf{y}|x_1, x_2) p(x_1) p(x_2)$. (b) The normalized histogram with $N = 10,000$ samples, using the GARS algorithm within a Gibbs sampler.

Appendix

Proposition. *The state estimators* $\hat{x}_j \in \mathrm{argmax}_{x \in [\mathcal{B}_j]} \ell(x; \mathbf{y}, \mathbf{g})$ $= \mathrm{argmin}_{x \in [\mathcal{B}_j]} V(x; \mathbf{y}, \mathbf{g})$ *belong to the interval* $\mathcal{I}_j$, *i.e.*,

$$\hat{x}_j \in \mathcal{I}_j \triangleq [\min(\mathcal{X}_j), \max(\mathcal{X}_j)], \tag{50}$$

where $\mathcal{X}_j \triangleq \{x_{1,j}, \ldots, x_{n,j}\}$ *is the set of all simple estimates in* $\mathcal{B}_j$ *and* $\mathcal{I}_j \subseteq \mathcal{B}_j$.

Proof. We have to prove that the derivative of the system potential function is

$$\frac{dV}{dx} < 0 \quad \text{for all } x < \min(\mathcal{X}_j) \ (x \in [\mathcal{B}_j]), \tag{51}$$

and

$$\frac{dV}{dx} > 0 \quad \text{for all } x > \max(\mathcal{X}_j) \ (x \in [\mathcal{B}_j]), \tag{52}$$

so that all stationary points of V stay inside $\mathcal{I}_j = [\min(\mathcal{X}_j), \max(\mathcal{X}_j)]$. Routine calculations yield the derivative

$$\frac{dV}{dx} = -\sum_{i=1}^{n} \frac{dg_i}{dx} \left[\frac{d\overline{V}_i}{d\vartheta_i} \right]_{\vartheta_i = y_i - g_i(x)} \tag{53}$$

and we aim to evaluate it outside the interval $\mathcal{I}_j$. To do it, let us denote $x_{min} = \min(\mathcal{X}_j)$ and $x_{max} = \max(\mathcal{X}_j)$ and consider the cases $dg_i/dx > 0$ and $dg_i/dx < 0$ separately (recall that we have assumed the sign of dg_i/dx to remain constant in $\mathcal{B}_j$).

When $dg_i/dx > 0$ and since, for every simple estimate, $x_{i,j} \geq x_{min}$, we obtain that $y_i = g_i(x_{i,j}) \geq g_i(x_{min}) > g_i(x) \forall x < x_{min}$. Then $y_i - g_i(x) > 0$, for all $x < x_{min}$, and, due to properties (P1) and (P2) of marginal potential functions, $[d\overline{V}_i/d\vartheta_i]_{\vartheta_i = y_i - g_i(x) > 0} > 0$ for all i. As a consequence, $dV/dx < 0 \ \forall x < x_{min}, x \in [\mathcal{B}_j]$.

When $dg_i/dx < 0$ and $x_{i,j} \geq x_{min}$, we obtain that $y_i = g_i(x_{i,j}) \leq g_i(x_{min}) < g_i(x)$, $\forall x < x_{min}$. Then $y_i - g_i(x) < 0$ for all $x < x_{min}$ and $[d\overline{V}_i/d\vartheta_i]_{\vartheta_i = y_i - g_i(x) < 0} < 0$, again because of (P1) and (P2). As a consequence, $dV/dx < 0 \ \forall x < x_{min}, x \in [\mathcal{B}_j]$.

A similar argument for $x > x_{max}$ yields $dV/dx > 0$ for all $x > x_{max}$ and completes the proof. $\square$

References

[1] A.M. Ali, K. Yao, T.C. Collier, E. Taylor, D. Blumstein, L. Girod, An empirical study of collaborative acoustic source localization, in: Proceedings of the Information Processing in Sensor Networks (IPSN07), Boston, April 2007.

[2] M.H. DeGroot, M.J. Schervish, Probability and Statistics, third ed., Addison-Wesley, New York, 2002.

[3] P.M. Djuric, Y. Huang, T. Ghirmai, Perfect sampling: a review and applications to signal processing, IEEE Transaction on Signal Processing 50 (2) (February 2002) 345–356.

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

[5] A. Doucet, N. deFreitas, N. Gordon (Eds.), Sequential Monte Carlo Methods in Practice, Springer, New York, USA, 2001.

[6] M. Evans, T. Swartz, Random variate generation using concavity properties of transformed densities, Journal of Computational and Graphical Statistics 7 (4) (1998) 514–528.

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

[8] W.R. Gilks, Derivative-free adaptive rejection sampling for Gibbs sampling, Bayesian Statistics (4) (1992) 641–649.

[9] W.R. Gilks, N.G. Best, K.K.C. Tan, Adaptive rejection metropolis sampling within Gibbs sampling, Applied Statistics 44 (4) (1995) 455–472.

[10] N.G.O. Robert, E.I. George, W.R. Gilks, Adaptive direction sampling, The Statistician 43 (1) (1994) 179–189.

[11] W.R. Gilks, P. Wild, Adaptive rejection sampling for Gibbs sampling, Applied Statistics 41 (2) (1992) 337–348.

[12] W. Hoermann, A rejection technique for sampling from t-concave distributions, ACM Transactions on Mathematical Software 21 (2) (1995) 182–193.

[13] A.N. Kolmogorov, S.V. Fomin, Introductory Real Analysis, Dover Publications, New York, 1975.

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

[15] J.R. Larocque, P. Reilly, Reversible jump MCMC for joint detection and estimation of sources in colored noise, IEEE Transactions on Signal Processing 50 (2) (February 1998).

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

[17] C.P. Robert, G. Casella, Monte Carlo Statistical Methods, Springer, 2004.

[18] D. Williams, Probability with Martingales, Cambridge University Press, Cambridge, UK, 1991.