

A generalization of the adaptive rejection sampling algorithm

Luca Martino · Joaquín Míguez

Abstract Rejection 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. The adaptive rejection sampling method is an efficient algorithm to sample from a log-concave target density, that attains high acceptance rates by improving the proposal density whenever a sample is rejected. In this paper we introduce a generalized adaptive rejection sampling procedure that can be applied with a broad class of target probability distributions, possibly non-log-concave and exhibiting multiple modes. The proposed technique yields a sequence of proposal densities that converge toward the target pdf, thus achieving very high acceptance rates. We provide a simple numerical example to illustrate the basic use of the proposed technique, together with a more elaborate positioning application using real data.

Keywords Rejection sampling · Adaptive rejection sampling · Gibbs sampling · Monte Carlo integration · Sensor networks · Target localization

L. Martino (✉) · J. 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

J. Míguez
e-mail: joaquin.miguez@uc3m.es

1 Introduction

Rejection sampling (see, e.g., Robert and Casella 2004, Chap. 2) is a standard Monte Carlo technique for universal sampling. It can be used to generate samples from a target probability density function (pdf) by drawing from a 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 (Robert and Casella 2004). Rejection sampling 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 (Gilks and Wild 1992; Gilks et al. 1994; Künsch 2005). An important limitation of rejection sampling methods, however, is the need to analytically establish a bound for the ratio of the target and proposal densities. There is a lack of general methods for the computation of exact bounds and so they have to be found for each specific problem at hand.

One exception to this rule is the so-called *adaptive rejection sampling* (ARS) method (Gilks and Wild 1992; Gilks 1992; Robert and Casella 2004) which, given a target density, yields both a suitable proposal pdf (easy to draw from) and an upper bound for the ratio of the target density over this proposal. The class of adaptive rejection sampling methods is particularly interesting because they ensure high acceptance rates. Indeed the standard ARS algorithm of Gilks and Wild (1992) yields a sequence of proposal functions that actually converge toward the target pdf when the procedure is iterated. As the proposal density becomes closer to the target pdf, the proportion of accepted samples grows. Unfortunately, this algorithm can only be used with log-concave target densities. An extension of the standard

ARS has been proposed in Hörmann (1995), where the same method of Gilks and Wild (1992) is applied to T -concave densities, with T being a monotonically increasing transformation, not necessarily the logarithm. The technique cannot be applied to multimodal densities either.

Indeed, when the target pdf presents several modes, an elaborate extension of the original ARS method is needed. The adaptive rejection Metropolis sampling (ARMS) procedure of Gilks et al. (1995) uses a Metropolis-Hastings algorithm to account for multiple modes. However, this approach is not fully satisfactory because the generated samples form a Markov chain. Hence, they are correlated and, for certain distributions, the chain can be easily trapped in a single mode.

The method in Evans and Swartz (1998) extends the technique of Hörmann (1995) to multimodal distributions. It involves the decomposition of the T -transformed density into pieces which are either convex and concave on disjoint intervals and can be handled separately. Unfortunately, this decomposition requires the ability to find all the inflection points of the T -transformed density, which can be something hard to do for practical problems.

More recently, it has been proposed to handle multimodal distributions by decomposing the log-density into a sum of concave and convex functions (Görür and Teh 2009). Then, every concave/convex element is handled using a method similar to the ARS procedure. A limitation for the applicability of this technique is the need to decompose the logarithm of the target pdf into concave and convex components, since no general procedure to carry out this decomposition is available. Also, the application of this technique to distributions with an infinite support requires that the tails of the log-densities be strictly concave.

In this paper, we introduce a generalization of the ARS method that can be applied to a large class of target pdf's, possibly not log-concave and possibly multimodal. We assume that the log-density $\log[p(x)]$ can be expressed as a sum of composed functions, $\log[p(x)] = -\sum_i^n (\bar{V}_i \circ g_i)(x) + cst$, where the \bar{V}_i 's are convex and the g_i 's are either convex or concave.¹ This is not a universal decomposition that can be applied to every density of interest, but the freedom in the choice of the \bar{V}_i 's and the g_i 's enables to describe a family of pdf's that includes, e.g., a posteriori distributions of random variables given a set of independent observations. Further remarks on the applicability of this approach are provided in Sect. 3.6.

The method is based on constructing piecewise-linear approximations of the nonlinearities g_i underlying the target density. The construction of these approximations requires a

¹These initial assumptions on the convexity of the \bar{V}_i 's and the g_i 's can be relaxed, as shown in Sect. 4.

sequence of calculations that can be relatively long depending on the target pdf. They are very systematic though, and the resulting piecewise-linear approximation yield an easy-to-sample proposal pdf. In a way also similar to the original ARS, the proposals are improved every time a candidate sample is rejected.

A major difference between the proposed method and the previous approaches, including both original methods by Gilks et al. (Gilks and Wild 1992; Gilks 1992) and posterior work on multimodal distributions (Evans and Swartz 1998; Görür and Teh 2009), is that we do not attempt to construct bounds for the log- or T -transformed density directly. Instead, we build approximations of the nonlinearities g_i . In the cases in which the latter functions can be identified, this approach is systematic because every non-linearity is handled using the same approximation procedure (as described in Sect. 3.2 and Appendix). One advantage of the proposed technique, when compared to the procedure of Görür and Teh (2009), is that it can be extended to handle target densities with an infinite support and tails which are possibly not log-concave, as shown in Sect. 4.3.

The new generalized ARS technique is conceived to be used within more elaborate Monte Carlo methods. It enables, for instance, a systematic implementation of the accept/reject particle filter of Künsch (2005). There is another potential application in the implementation of the Gibbs sampler for systems in which the conditional densities are complicated, as illustrated in the example of Sect. 5.2. The latter involves the Bayesian estimation of the position of an object using real signal-strength data collected using a network of wireless sensors. Besides the ARS-based implementation of the Gibbs sampler, this example is relevant because it shows how the proposed technique fits particularly well with Bayesian inference problems.

The rest of the paper is organized as follows. Background material is presented in Sect. 2. The basic form of the new algorithm is introduced in Sect. 3, with further extensions in Sect. 4. Section 5 is devoted to the examples. We conclude with a brief summary and conclusions in Sect. 6.

2 Background

2.1 Rejection sampling

Rejection sampling is a universal method for drawing independent samples from a target density $p_o(x) \geq 0$ known up to a proportionality constant (hence, we can evaluate $p(x) \propto p_o(x)$). Let $\exp\{-W(x)\}$ be an overbounding function for $p(x)$, i.e., $\exp\{-W(x)\} \geq p(x)$. We can generate N samples from $p_o(x)$ according to the standard rejection sampling algorithm:

1. Set $i = 1$.

2. Draw samples x' from $\pi(x) \propto \exp\{-W(x)\}$ and u' from $\mathcal{U}(0, 1)$, where $\mathcal{U}(0, 1)$ is the uniform pdf in $[0, 1]$.
3. If $\frac{p(x')}{\exp\{-W(x')\}} \geq u'$ then $x^{(i)} = x'$ and set $i = i + 1$, else discard x' and go back to step 2.
4. If $i > N$ then stop, else go back to step 2.

The fundamental figure of merit of a rejection sampler is the mean acceptance rate, i.e., the expected number of accepted samples over the total number of proposed candidates. In practice, finding a tight overbounding function is crucial for the performance of a rejection sampling algorithm.

2.2 Adaptive rejection sampling

The standard adaptive rejection sampling (ARS) (Gilks and Wild 1992) algorithm enables the construction of a sequence of proposal densities, $\{\pi_t(x)\}_{t \in \mathbb{N}}$, tailored to the target density $p_o(x) \propto p(x)$. 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, the 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, in subsequent sections, we proceed to introduce its extension for non-log-concave and multimodal target densities.

Let us assume that we want to draw from the pdf $p_o(x) \propto p(x) \geq 0$ with support in $\mathcal{D} \subseteq \mathbb{R}$. The ARS procedure can be applied when $\log[p(x)]$ is concave, i.e., when the potential function

$$V(x) \triangleq -\log[p(x)], \quad x \in \mathcal{D} \subseteq \mathbb{R}, \quad (1)$$

is strictly convex. The domain of $V(x)$ (and $p(x)$) is denoted as \mathcal{D} all through the paper.

Let $\mathcal{S}_t \triangleq \{s_1, s_2, \dots, s_{m_t}\} \subset \mathcal{D}$ be a set of support points, sorted in ascending order $s_1 < \dots < s_{m_t}$. The number of points m_t can grow with the iteration index t . From \mathcal{S}_t we build a piecewise-linear lower hull of $V(x)$, denoted $W_t(x)$, formed from segments of linear functions tangent to $V(x)$ at the support points s_k in \mathcal{S}_t . If we denote as $w_k(x)$ the linear function tangent to $V(x)$ at s_k , then the piecewise linear

function $W_t(x)$ can be defined as

$$W_t(x) \triangleq \max\{w_1(x), \dots, w_{m_t}(x)\}. \quad (2)$$

Figure 1 illustrates the construction of $W_t(x)$ with three support points for the convex potential function $V(x) = x^2$. It is apparent that $W_t(x) \leq V(x)$ by construction, therefore $\exp\{-W_t(x)\}$ is an overbounding function for $p(x)$, i.e.,

$$\exp\{-W_t(x)\} \geq p(x) = \exp\{-V(x)\}. \quad (3)$$

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

$$\pi_t(x) = c_t \exp[-W_t(x)], \quad (4)$$

where c_t is the proportionality constant. Therefore $\pi_t(x)$ is piecewise-exponential and very easy to sample from. Equation (3) can be rewritten as $\frac{1}{c_t} \pi_t(x) \geq p(x)$, hence we can apply the rejection sampling principle.

When a sample x' from $\pi_t(x)$ is rejected we can incorporate it into the set of support points, i.e., $\mathcal{S}_{t+1} = \mathcal{S}_t \cup \{x'\}$ and $m_{t+1} = m_t + 1$. Then, we compute a refined lower hull, $W_{t+1}(x)$, and a new proposal density $\pi_{t+1}(x) = c_{t+1} \exp[-W_{t+1}(x)]$. Table 1 summarizes the ARS algorithm.

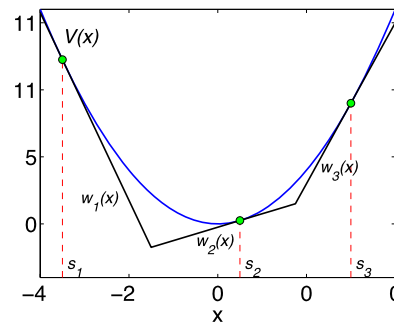


Fig. 1 Example of construction of the piecewise linear function $W_t(x)$ with three support points $\mathcal{S}_t = \{s_1, s_2, s_{m_t=3}\}$, as carried out by the original ARS technique. The potential is $V(x) = x^2$ and the function $W_t(x) \triangleq \max\{w_1(x), w_2(x), w_3(x)\}$ is formed from segments of linear functions tangent to the potential $V(x)$ at the support points in \mathcal{S}_t

Table 1 Adaptive rejection sampling algorithm

1. Start with $i = 1$, $t = 0$, $m_0 = 2$ $\mathcal{S}_0 = \{s_1, s_2\}$ where $s_1 < s_2$, and the derivatives of $V(x)$ in $s_1, s_2 \in \mathcal{D}$ have different signs. Let N be the number of desired samples from $p_o(x)$.
2. Build the piecewise-linear function $W_t(x)$ as shown in Fig. 1, using the tangent lines to $V(x)$ 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 $u' \leq \frac{p(x')}{\exp\{-W_t(x')\}}$ then accept $x^{(i)} = x'$ and set $\mathcal{S}_{t+1} = \mathcal{S}_t$, $m_{t+1} = m_t$, $i = i + 1$.
5. Otherwise, if $u' > \frac{p(x')}{\exp\{-W_t(x')\}}$, then reject x' , set $\mathcal{S}_{t+1} = \mathcal{S}_t \cup \{x'\}$ and update $m_{t+1} = m_t + 1$.
6. Sort \mathcal{S}_{t+1} in ascending order and increment $t = t + 1$. If $i > N$ then stop, else go back to step 2.

3 Generalized adaptive rejection sampling

3.1 Preliminary definitions and assumptions

Assume that the target pdf $p_o(x) \propto p(x)$, $x \in \mathcal{D}$, can be expressed as

$$\begin{aligned} p_o(x) \propto p(x) &= \exp \left\{ -c_n - \sum_{i=1}^n \bar{V}_i(g_i(x)) \right\} \\ &= \exp \left\{ -c_n - \sum_{i=1}^n (\bar{V}_i \circ g_i)(x) \right\}, \end{aligned} \quad (5)$$

where c_n is a constant value, \circ denotes the composition of functions and:

1. Functions $\bar{V}_i(\vartheta)$, for $i = 1, \dots, n$ (hereafter called marginal potentials), are convex with their unique minimum located at $\vartheta = \mu_i$ (an example can be seen in Fig. 4(b)).
2. Functions $g_i(x)$, $i = 1, \dots, n$, are either convex or concave (i.e., they have a second derivative with constant sign) and possibly nonlinear.

Note that, in general, $p(x)$ is non-log-concave, since each term $\bar{V}_i \circ g_i$ in the sum may have a second derivative with non-constant sign. Therefore, the standard ARS technique cannot be applied in this setup. The freedom in the choice of the \bar{V}_i 's and the g_i 's enables the representation of a broad class of exponential-type densities in this way (see Sect. 3.6 for further details).

We also introduce the set of *simple estimates* corresponding to the nonlinearity $g_i(x)$ as

$$\mathcal{X}_i \triangleq \{x_i \in \mathbb{R} : g_i(x_i) = \mu_i\}, \quad (6)$$

where μ_i is the position of the minimum of the marginal potential \bar{V}_i . Recall that each function $g_i(x)$ is assumed to have a second derivative with constant sign, hence the equation $\mu_i = g_i(x_i)$ can yield zero, one or two simple estimates.

3.2 Basic strategy

In this section we describe the basic procedure to build a proposal density $\pi(x)$ in a given interval $\mathcal{I} \subset \mathcal{D}$ of values of x . Later on, we generalize this procedure to yield an adaptive method. Let $\mathbf{g}(x) \triangleq [g_1(x), \dots, g_n(x)]$ denote a vector of nonlinearities. We now write the potential function $-\log[p(x)]$ as an explicit function of \mathbf{g} , i.e.,

$$V(x; \mathbf{g}) \triangleq -\log[p(x)] = c_n + \sum_{i=1}^n \bar{V}_i(g_i(x)). \quad (7)$$

Given an interval $\mathcal{I} \subset \mathcal{D}$, we proceed in two steps. First, we replace every nonlinearity $g_i(x)$ with a suitable linear

function $r_i(x)$. In this way we generate a modified potential $V(x, \mathbf{r})$, with $\mathbf{r}(x) = [r_1(x), r_2(x), \dots, r_n(x)]$, that lies below the original one, i.e., $V(x, \mathbf{r}) \leq V(x, \mathbf{g})$. Second, we construct a linear function $W(x)$ that is tangent at an (arbitrary) point $x^* \in \mathcal{I}$ to the modified potential $V(x, \mathbf{r})$. The two steps are described in detail below.

1. We build linear functions $r_i(x)$ such that

$$\bar{V}_i(r_i(x)) \leq \bar{V}_i(g_i(x)), \quad \forall x \in \mathcal{I}, \quad (8)$$

for every $i = 1, \dots, n$ (see Appendix for details). As a consequence, substituting \mathbf{g} by \mathbf{r} into the functional $V(x; \cdot)$, we obtain the inequality

$$\begin{aligned} V(x; \mathbf{r}) &\triangleq c_n + \sum_{i=1}^n \bar{V}_i(r_i(x)) \\ &\leq V(x; \mathbf{g}) = c_n + \sum_{i=1}^n \bar{V}_i(g_i(x)), \end{aligned} \quad (9)$$

$\forall x \in \mathcal{I}$. Note that $\exp\{-V(x; \mathbf{r})\}$ is already an overbounding function for $p(x)$, i.e.,

$$\exp\{-V(x; \mathbf{r})\} \geq \exp\{-V(x; \mathbf{g})\} = p(x). \quad (10)$$

However, it is not possible in general to draw from $\pi^*(x) \propto \exp\{-V(x; \mathbf{r})\}$ and we need to seek further simplifications.

2. Note that the modified potential $V(x; \mathbf{r})$ is convex in \mathcal{I} . Indeed,

$$\begin{aligned} \frac{d^2 \bar{V}_i(r_i(x))}{dx^2} &= \frac{d\bar{V}_i}{d\vartheta} \frac{d^2 r_i}{dx^2} + \left(\frac{dr_i}{dx} \right)^2 \frac{d^2 \bar{V}_i}{d\vartheta^2} \\ &= 0 + \left(\frac{dr_i}{dx} \right)^2 \frac{d^2 \bar{V}_i}{d\vartheta^2} \geq 0, \end{aligned} \quad (11)$$

where we have used that $\frac{d^2 r_i}{dx^2} = 0$ (since r_i is linear) and the convexity of the marginal potentials $\bar{V}_i(\vartheta)$, $i = 1, \dots, n$. Therefore, we can use a line tangent to $V(x; \mathbf{r})$ at an arbitrary point $x^* \in \mathcal{I}$ to build a linear function $W(x)$ such that $W(x) \leq V(x; \mathbf{r})$ for all $x \in \mathcal{I}$.

Thus,

$$\begin{aligned} \exp\{-W(x)\} &\geq \exp\{-V(x; \mathbf{r})\} \\ &\geq \exp\{-V(x; \mathbf{g})\} = p(x) \end{aligned} \quad (12)$$

is an overbounding function of $p(x)$. Since $W(x)$ is linear, it is straightforward to compute the proportionality constant

$$c^{-1} = \int_{x \in \mathcal{I}} \exp\{-W(x)\} dx \quad (13)$$

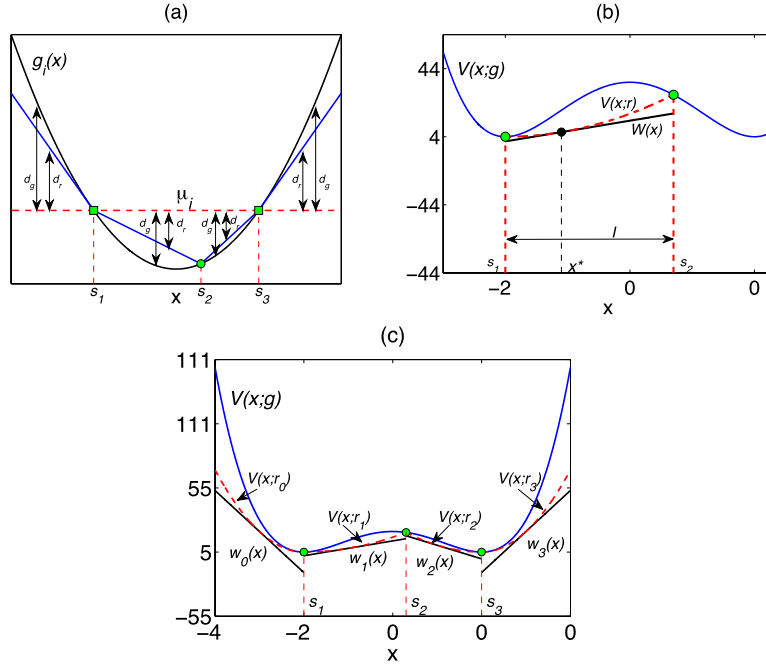


Fig. 2 (a) Example of construction of the linear function $r_i(x)$ in order to replace the nonlinearity $g_i(x) = x^2$ in different intervals \mathcal{I} . The absolute difference between the linear function $r_i(x)$ and the value μ_i , i.e., $d_r = |\mu_i - r_{i,k}(x)|$, is always less than the distance $d_g = |\mu_i - g_i(x)|$ in the interval \mathcal{I} , i.e., $d_r \leq d_g$ for all $x \in \mathcal{I}$. Hence, in $[-\infty, s_1]$ and $[s_3, +\infty]$ we use tangent straight lines while in $[s_1, s_2]$ and $[s_2, s_3]$ we use the linear functions passing through the two support points. (b) Example of construction of the linear function $W(x)$ inside a generic interval $\mathcal{I} = [s_1, s_2]$. The picture shows a non-convex potential $V(x; \mathbf{g})$ in solid line while the modified potential $V(x; \mathbf{r})$ is depicted in dashed line for $\forall x \in \mathcal{I}$. The linear function $W(x)$ is

tangent to $V(x; \mathbf{r})$ at an arbitrary point $x^* \in \mathcal{I}$. (c) Example of construction of the piecewise linear function $W_t(x)$ with three support points $\mathcal{S}_t = \{s_1, s_2, s_{m_t=3}\}$, as carried out by the generalized ARS technique. The potential is $V(x; \mathbf{g}) = 16 - 8x^2 + x^4 = (4 - x^2)^2$, therefore we can express it as $V(x; \mathbf{g}) = (\bar{V}_1 \circ g_1)(x)$ where $\bar{V}_1(\vartheta) = \vartheta^2$ and $g_1(x) = 4 - x^2$ (i.e., $n = 1$ and the vector of nonlinearities $\mathbf{g} = g_1$ is scalar). The modified potential $V(x; \mathbf{r}_k)$, for $x \in \mathcal{I}_k$, is depicted with a dashed line. The piecewise linear function $W_t(x)$ consists of segments of linear functions $w_k(x)$ tangent to the modified potential $V(x; \mathbf{r}_k)$ at arbitrary points $x_k^* \in \mathcal{I}_k$, with $k = 0, \dots, m_t = 3$, where $\mathcal{I}_0 = [-\infty, s_1]$, $\mathcal{I}_1 = [s_1, s_2]$, $\mathcal{I}_2 = [s_2, s_3]$ and $\mathcal{I}_3 = [s_3, +\infty]$

and to use the density $\pi(x) = c \exp\{-W(x)\}$ as a proposal function. Drawing from $\pi(x)$ is easy because it is a truncated exponential pdf, restricted to \mathcal{I} .

The details on the computation of the appropriate functions $r_i(x)$, with $i = 1, \dots, n$, are described in Appendix. However, Fig. 2(a) provides the basic idea of how to construct the linear function $r_i(x)$ that replaces the nonlinearity $g_i(x) = x^2$ in an interval \mathcal{I} . We seek a linear function $r_i(x)$ such that the absolute difference $d_r = |\mu_i - r_i(x)|$ is always less than the distance $d_g = |\mu_i - g_i(x)|$, i.e., $d_r \leq d_g$ in \mathcal{I} . Therefore, in the intervals $[-\infty, s_1]$ and $[s_3, +\infty]$ we use tangent straight lines while in $[s_1, s_2]$ and $[s_2, s_3]$ we use the linear functions passing through the two support points. Figure 2(b) shows an example of construction of the linear function $W(x)$ in a generic interval $\mathcal{I} \subset \mathcal{D}$. The picture represents a non-convex potential $V(x; \mathbf{g})$ (solid line) and the corresponding modified potential $V(x; \mathbf{r})$ in \mathcal{I} , depicted in dashed line. The linear function $W(x)$ is tan-

gent to the modified potential $V(x; \mathbf{r})$ in an arbitrary point $x^* \in \mathcal{I}$.

3.3 Adaptive algorithm

The basic method described above can be iterated to yield a sequence of proposal pdf's

$$\pi_1(x), \pi_2(x), \dots, \pi_t(x), \dots,$$

that converges to the target pdf $p_o(x)$. Let us consider the set of support points after the t th iteration

$$\mathcal{S}_t \triangleq \{s_1, s_2, \dots, s_{m_t}\} \subset \mathcal{D} \quad (14)$$

sorted in ascending order, $s_1 < \dots < s_{m_t}$, where m_t is the number of elements. From the points in \mathcal{S}_t we construct the closed intervals $\mathcal{I}_k \triangleq [s_k, s_{k+1}]$ for $k = 1, \dots, m_t - 1$, together with two semi-open intervals $\mathcal{I}_0 \triangleq (-\infty, s_1]$ and $\mathcal{I}_{m_t} \triangleq [s_{m_t}, +\infty)$. For each interval \mathcal{I}_k , $k = 0, \dots, m_t$, we

Table 2 Generalized adaptive rejection sampling algorithm

-
1. Start with $q = 1, t = 0$ and $\mathcal{S}_0 \triangleq \{s_j\}_{j=1}^{m_0}$. Let N be the number of desired samples from $p_o(x)$. At t th iteration perform the following steps.
 2. Build $r_{i,k}(x)$ for $i = 1, \dots, n, k = 0, \dots, m_t$.
 3. Build $W_t(x) \triangleq w_k(x) \forall x \in \mathcal{I}_k$, for every $k = 0, \dots, m_t$, where $w_k(x)$ is a tangent line to $V(x; \mathbf{r}_k)$ at an arbitrary point $x^* \in \mathcal{I}_k$.
 4. Draw a sample x' from $\pi_t(x) \propto \exp\{-W_t(x)\}$.
 5. Sample u' from $\mathcal{U}([0, 1])$.
 6. If $u' \leq \frac{p(x')}{\exp\{-W_t(x')\}}$ accept $x^{(q)} = x'$ and set $\mathcal{S}_{t+1} = \mathcal{S}_t, q = q + 1$.
 7. Otherwise, if $u' > \frac{p(x')}{\exp\{-W_t(x')\}}$ reject x' and update $\mathcal{S}_{t+1} = \mathcal{S}_t \cup \{x'\}$.
 8. Sort \mathcal{S}_{t+1} in ascending order, increment $t = t + 1$ and if $q > N$ then stop, else go back to step 2.
-

build suitable vectors of linear functions $\mathbf{r}_k(x) \triangleq [r_{1,k}(x), \dots, r_{n,k}(x)]$, using the technique in [Appendix](#), to comply with the inequality (8) for every $x \in \mathcal{I}_k$.

Since the modified potential $V(x; \mathbf{r}_k)$ is convex, we can build a piecewise-linear lower hull $W_t(x)$ such that $W_t(x) \leq V(x; \mathbf{r}_k) \leq V(x; \mathbf{g})$ for $\forall x \in \mathcal{I}_k$. Then, we build the tangent lines to the modified potential $V(x; \mathbf{r}_k)$ at arbitrary points $x_k^* \in \mathcal{I}_k = [s_k, s_{k+1}]$, denoted as $w_k(x)$. As a result, the piecewise linear function $W_t(x)$ at the t -iteration is

$$W_t(x) \triangleq \begin{cases} w_0(x), & \text{if } x \in \mathcal{I}_0, \\ \vdots \\ w_{m_t}(x), & \text{if } x \in \mathcal{I}_{m_t}, \end{cases} \quad (15)$$

and the t th proposal density is

$$\pi_t(x) \propto \exp\{-W_t(x)\}. \quad (16)$$

When a sample x' drawn from $\pi_t(x)$ is rejected, x' is incorporated as a support point in the new set $\mathcal{S}_{t+1} \triangleq \mathcal{S}_t \cup \{x'\}$ and, as a consequence, a refined lower hull $W_{t+1}(x)$ is constructed yielding a better approximation of the potential function $V(x; \mathbf{g})$. In this way, $\pi_{t+1}(x) \propto \exp\{-W_{t+1}(x)\}$ becomes closer to the target pdf $p_o(x)$ and it can be expected that the mean acceptance rate be higher.

Figure 2(c) illustrates the construction of the piecewise linear function $W_t(x)$ using the proposed technique for the non-convex potential $V(x; \mathbf{g}) = 16 - 8x^2 + x^4$ with three support points, $\mathcal{S}_t = \{s_1, s_2, s_{m_t=3}\}$. Indeed, this potential can be rewritten as $V(x; \mathbf{g}) = (4 - x^2)^2$, so that we can interpret it as composition of functions $(\bar{V}_1 \circ g_1)(x)$, where $\bar{V}_1(\vartheta) = \vartheta^2$ and $g_1(x) = 4 - x^2$ ($n = 1$). The dashed line shows the modified potentials $V(x; \mathbf{r}_k), k = 0, \dots, m_t = 3$. Function $W_t(x)$ consists of segments of linear functions $w_k(x)$ tangent to the modified potentials $V(x; \mathbf{r}_k)$ at arbitrary points $x_k^* \in \mathcal{I}_k$, with $k = 0, \dots, m_t = 3$.

3.4 Initialization and summary

Let us recall that the set of simple estimates corresponding to the nonlinearities $g_i(x)$ is defined as $\mathcal{X}_i = \{x_i \in \mathbb{R} :$

$g_i(x_i) = \mu_i\}$, where μ_i is the position of the minimum of the marginal potential \bar{V}_i . Since we have assumed nonlinearities $g_i(x)$ with second derivative with constant sign, each equation $\mu_i = g_i(x_i)$ can yield zero, one or two different simple estimates.

We initialize the algorithm with $\mathcal{S}_0 \triangleq \{s_j\}_{j=1}^{m_0}$ such that:

- (a) All simple estimates are contained in \mathcal{S}_0 , i.e., $X_i \subset \mathcal{S}_0, i = 1, \dots, n$.
- (b) For each non-monotonic function $g_i(x)$ that generates two simple estimates $x_{i,1} < x_{i,2}$, we incorporate in \mathcal{S}_0 an arbitrary support point to $s_j \in [x_{i,1}, x_{i,2}]$.
- (c) When the nonlinearity $g_i(x)$ is monotonic, we can have at most a single simple estimate, x_i . If $\frac{dg_i(x)}{dx} \times \frac{d^2g_i(x)}{dx^2} \geq 0$, we add an arbitrary point $s_j \in (-\infty, x_i]$ into \mathcal{S}_0 . Otherwise if $\frac{dg_i(x)}{dx} \times \frac{d^2g_i(x)}{dx^2} \leq 0$, we choose a support point in $s_j \in [x_i, +\infty)$.

The set \mathcal{S}_0 thus constructed enables us to build the linear functions $r_{i,k}(x)$ in the way described in [Appendix](#). If additional support points are included in \mathcal{S}_0 , the resulting proposal $\pi_0(x)$ becomes a tighter approximation of $p_o(x)$.

The proposed generalized adaptive rejection sampling (GARS) algorithm is summarized in [Table 2](#). [Figures 3](#) illustrates how the sequence of proposal pdf's converges toward the target density as the number of support points increases.

3.5 Improper proposals

The GARS algorithm summarized in [Table 2](#) breaks down when the potential function $V(x; \mathbf{g})$ has both an infinite support ($x \in \mathcal{D} = \mathbb{R}$) and concave tails. 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 ($\int_{-\infty}^{+\infty} \pi_t(x) dx \rightarrow +\infty$) and cannot be used for rejection sampling. We tackle this problem in [Sect. 4.3](#) below.

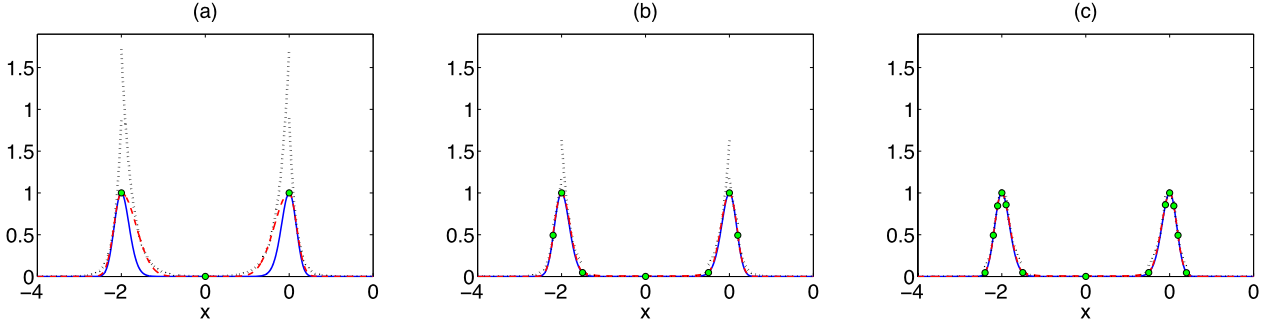


Fig. 3 Convergence of the overbounding functions $\exp\{-V(x; \mathbf{r}_k)\}$ (dashed line), and $\exp\{-W_i(x)\}$ (dotted line) toward the function $p(x) = \exp\{-(4-x^2)^2\}$ (solid line), with the GARS technique. The points, corresponding to the support points $\{s_j\}_j^{m_t}$, are depicted with

circles. (a) Construction of the overbounding functions with $m_t = 3$ support points. (b) Construction with $m_t = 7$. (c) Construction with $m_t = 13$

3.6 Applicability

In this subsection we briefly describe two general classes of target densities that appear often in practice and can be easily handled with the proposed method.²

Densities of the form of (5) appear naturally in Bayesian inference problems where it is desired to draw from the posterior pdf $p(x|\mathbf{y})$ with $\mathbf{y} = [y_1, y_2, \dots, y_n] \in \mathbb{R}^n$ of a random variable X given a collection of observations

$$Y_1 = \bar{g}_1(X) + \Theta_1, \quad \dots, \quad Y_n = \bar{g}_n(X) + \Theta_n, \quad (17)$$

where $\Theta_1, \dots, \Theta_n$ are independent “noise” variables. In fact, writing the noise pdf’s as $p(\theta_i) \propto \exp\{-\bar{V}_i(\theta_i)\}$, $i = 1, \dots, n$, the likelihood function can be expressed as

$$p(\mathbf{y}|x) \propto \exp\left\{-\sum_{i=1}^n \bar{V}_i(y_i - \bar{g}_i(x))\right\}. \quad (18)$$

Therefore, denoting $g_i(x) = y_i - \bar{g}_i(x)$ and writing the prior pdf as $p(x) \propto \exp\{-\bar{V}_{n+1}(g_{n+1}(x))\}$, the potential function is

$$\begin{aligned} V(x; \mathbf{g}) &= -\log[p(x|\mathbf{y})] \\ &= -\log[p(\mathbf{y}|x)p(x)] = \sum_{i=1}^{n+1} \bar{V}_i(g_i(x)). \end{aligned} \quad (19)$$

An example of this type is shown in Sect. 5.2.

Additionally, note that while the standard ARS can be applied for all pdfs $p_o(x) \propto \exp\{-h(x)\}$ where $h(x)$ is convex, our technique can be used, e.g., for distributions of the type

$$\begin{aligned} p_o(x) &\propto h(x) \exp\{-h(x)\} \\ &= \exp\{-h(x) + \log[h(x)]\}, \end{aligned} \quad (20)$$

where $h(x)$ can be either convex or concave. In fact, in this case we can interpret the $-\log[p_o(x)]$ as a combination of two functions, $\bar{V}_1 \circ g_1$, where $\bar{V}_1(\theta) \triangleq \theta - \log[\theta]$ (which is convex with a minimum at $\mu_1 = 1$) and $g_1(x) \triangleq h(x)$.

4 Extensions

Although the form allowed for the target pdf encompasses a large class of probability distributions, the aim of this paper is to provide a *general* adaptive rejection sampling scheme. Hence, in this section we investigate how to relax some constraints in the definition of $p_o(x)$, so that a broader class of densities can be handled. In particular, we first study the case in which the marginal potentials are non-convex functions, in Sect. 4.1. Then we consider the problem of dealing with arbitrary nonlinearities (possibly neither convex nor concave) in Sect. 4.2. Finally, in Sect. 4.3, we propose a solution to the problem of improper proposals pointed out in Sect. 3.5.

4.1 Non-convex marginal potentials

Consider a target pdf $p_o(x) \propto \exp\{-V(x; \mathbf{g})\}$ where the potential function $V(x; \mathbf{g}) = c_n + \sum_{i=1}^n \bar{V}_i(g_i(x))$ is built using non-convex marginal potentials. To be specific, we consider the case where each $\bar{V}_i(\vartheta)$, $i \in \{1, \dots, n\}$, is

- strictly increasing and concave for $\vartheta > \mu_i$, and
- strictly decreasing and concave for $\vartheta < \mu_i$.

Obviously $\bar{V}_i(\vartheta)$ has a unique minimum at $\vartheta^* = \mu_i$ and we also assume $\bar{V}_i(\mu_i) > -\infty$. Figure 4(a) shows an example

²We do not imply that only these two types of pdf’s can be sampled using our method. Other classes of densities can also fit the structure of (5).

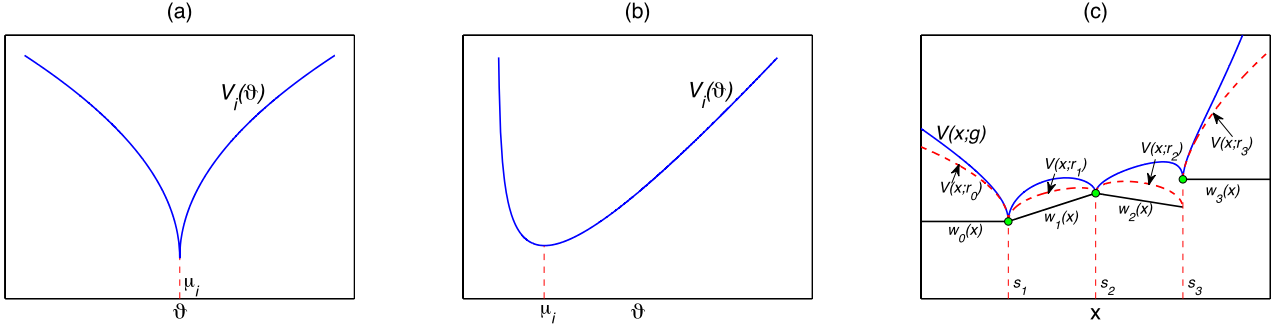


Fig. 4 (a) Example of generic marginal potential function $\bar{V}_i(\vartheta)$ strictly increasing and concave for $\vartheta > \mu_i$, strictly decreasing and concave for $\vartheta < \mu_i$. (b) Example of generic marginal potential function $\bar{V}_i(\vartheta)$ convex and with a minimum at $\vartheta^* = \mu_i$. (c) Example of construction of the modified potentials $V(x; \mathbf{r}_k)$ (dashed lines) and the piecewise linear function $W_t(x)$ (solid straight lines) when the potential function is $V(x; \mathbf{g}) = \sqrt{|x^2 - 4|} + \sqrt{|1 - \exp(x)|}$ (solid

line). We can express it as $V(x; \mathbf{g}) = \bar{V}_1(x^2 - 4) + \bar{V}_2(1 - \exp(x))$ where $\bar{V}_1(\vartheta) = \bar{V}_2(\vartheta) = \sqrt{|\vartheta|}$, i.e., the marginal potentials are concave strictly decreasing for $\vartheta < 0$ and strictly increasing for $\vartheta > 0$. The two nonlinearities $g_1(x) = x^2 - 4$ and $g_2(x) = 1 - \exp(x)$ generate the following sets of simple estimates $\mathcal{X}_1 = \{-2, 2\}$, $\mathcal{X}_2 = \{0\}$ that are contained in the set of support points $\mathcal{S}_t = \{s_1 = -2, s_2 = 0, s_3 = 2\}$

of this class of functions. These potentials describe super-Gaussian distributions, i.e., probability densities with positive kurtosis, which often appear in financial or biological applications (Reiss and Thomas 2007; Mayo et al. 2004).

With the assumptions above, the system potential $V(x; \mathbf{g})$ is differentiable almost everywhere, except for the (null measure) set of all simple estimates $\bigcup_{i=1}^n \mathcal{X}_i$ (we recall that $x \in \mathcal{X}_i$ if, and only if, $g_i(x) = \mu_i$). Moreover, since the set of support points \mathcal{S}_t includes all the simple estimates (see Sect. 3.4), i.e., $\bigcup_{i=1}^n \mathcal{X}_i \subset \mathcal{S}_t = \{s_k\}_{k=1}^{m_t}$, the simple estimates x_i can belong to the interval $\mathcal{I}_k = [s_k, s_{k+1}]$ only as border points. Therefore, replacing $g_i(x)$ with the linear function $r_{i,k}(x)$ in an interval $\mathcal{I}_k = [s_k, s_{k+1}]$ defined by two support points, we can write

$$\begin{aligned} \frac{d^2 \bar{V}_i(r_{i,k}(x))}{dx^2} &= \frac{d^2 r_{i,k}}{dx^2} \frac{d \bar{V}_i}{d \vartheta} + \left(\frac{dr_{i,k}}{dx} \right)^2 \frac{d^2 \bar{V}_i}{d \vartheta^2} \\ &= 0 + \left(\frac{dr_{i,k}}{dx} \right)^2 \frac{d^2 \bar{V}_i}{d \vartheta^2} \leq 0 \end{aligned} \quad (21)$$

for all $x \in \mathcal{I}_k$ except possibly the border points s_k or s_{k+1} if they are simple estimates. Hence, substituting the vector of nonlinearities \mathbf{g} with the vector of linear functions \mathbf{r}_k in \mathcal{I}_k , we obtain that the modified system potential $V(x; \mathbf{r}_k)$ is concave in \mathcal{I}_k . Thus we can build $W_t(x)$ for x in $\mathcal{I}_k = [s_k, s_{k+1}]$, for $k = 1, \dots, m_t - 1$, as the linear function passing through the points $(s_k, V(s_k; \mathbf{r}_k))$ and $(s_{k+1}, V(s_{k+1}; \mathbf{r}_{k+1}))$.

For $k = 0$ and $k = m_t$ we have, in general, semi-open intervals $\mathcal{I}_0 = [-\infty, s_1]$ and $\mathcal{I}_{m_t} = [s_{m_t}, +\infty]$, hence $W_t(x) = V(s_1; \mathbf{r}_1)$ for all $x \in \mathcal{I}_0$ and $W_t(x) = V(s_{m_t}; \mathbf{r}_{m_t})$ for all $x \in \mathcal{I}_{m_t}$, respectively. However, a constant value of $W_t(x)$ in a infinite interval yields an improper proposal $\pi_t(x) \propto \exp\{-W_t(x)\}$. Therefore this procedure can only be applied exactly either when the target pdf $p_o(x)$ has a

finite domain or using the alternative procedure explained in Sect. 4.3. See Fig. 4(c) for an example of this construction.

4.2 General nonlinearities

If some nonlinearity $g_i(x)$, defined in \mathcal{D} , has second derivative with non-constant sign, then it is not possible to apply directly the proposed GARS method.

However, this problem can be easily avoided in many cases. Indeed, if we can find a partition $\mathcal{D} = \bigcup_{j=1}^{q_i} \mathcal{B}_{i,j}$ (where $[\cdot]$ denotes the closure of an interval) such that $\mathcal{B}_{i,j} \cap \mathcal{B}_{i,k} = \emptyset$, $\forall j \neq k$, and such that $\frac{d^2 g_i}{dx^2}$ has a constant sign in every $\mathcal{B}_{i,j}$, $j = 1, \dots, q_i$, then we can incorporate this information into the initial set of support points \mathcal{S}_0 and apply the GARS algorithm.

Specifically, let $\mathcal{B}_{i,j} = (b_{i,j}^-, b_{i,j}^+)$. If we let $b_{i,j}^-, b_{i,j}^+ \in \mathcal{S}_0 = \{s_1, \dots, s_{m_0}\}$ (in addition to the rest of points indicated in Sect. 3.4), then $g_i(x)$ is either convex or concave in every $\mathcal{I}_k = [s_k, s_{k+1}]$ and we can apply the GARS algorithm as described in Sect. 3.

Note that the analytical study of the nonlinearities $g_i(x)$ is, in general, easier than the study of the entire log-density, as required, e.g., in Evans and Swartz (1998). For example, in the positioning application of Sect. 5.2 we can easily find the inflection points of each $g_i(x)$, $i = 1, \dots, n$, but the analysis over the whole log-density function is intractable.

4.3 Fixing improper proposals

We have seen that the application of the proposed method can yield improper proposal functions in some cases (see Sects. 3.5 and 4.1). The problem appears as follows: let $\mathcal{D} = \mathbb{R}$ and let $\mathcal{S}_t = \{s_0, \dots, s_{m_t}\}$ be the set of support points at the t th iteration and let $\mathcal{I}_0 = (-\infty, s_0)$, $\mathcal{I}_k = [s_k, s_{k+1})$,

$k = 1, \dots, m_t - 1$, and $\mathcal{I}_{m_t} = [s_{m_t}, +\infty)$ be the intervals in which we split the domain of $p(x)$. If $V(x; \mathbf{g})$ is concave for $x \in \mathcal{I}_0, x \in \mathcal{I}_{m_t}$ or both then $w_0(x), w_{m_t}(x)$, or both, are constant and $\pi_t(x)$ becomes improper.

We can circumvent this difficulty if, for some $j \in \{1, \dots, n\}$, the pdf defined as

$$p_j(x) \propto \exp\{-\bar{V}_j(g_j(x))\} \quad (22)$$

is such that: (a) we can integrate $p_j(x)$ over the intervals $\mathcal{I}_0, \mathcal{I}_1, \dots, \mathcal{I}_{m_t}$ and (b) we can sample from the density $p_j(x)$ restricted to every \mathcal{I}_k . To be specific, let us introduce the *reduced* potential

$$V_{-j}(x; \mathbf{g}) \triangleq c_n + \sum_{i=1, i \neq j}^n \bar{V}_i(g_i(x)), \quad (23)$$

attained by removing $\bar{V}_j(g_j(x))$ from $V(x; \mathbf{g})$.

It is straightforward to obtain lower bounds for $V_{-j}(x; \mathbf{g})$ at every interval \mathcal{I}_k . For instance, we can apply the standard technique of Sect. 3 to compute linear lower bounding functions $w_k(x) \leq V_{-j}(x; \mathbf{g}) \forall x \in \mathcal{I}_k$ and then take $\gamma_k = \min[w_k(s_k), w_k(s_{k+1})] \leq V_{-j}(x; \mathbf{g}) \forall x \in \mathcal{I}_k$. Once these bound are available, we build the proposal function by pieces as

$$\pi_t(x) \propto \begin{cases} \exp\{-\gamma_0 - \bar{V}_j(g_j(x))\}, & \forall x \in \mathcal{I}_0, \\ \vdots \\ \exp\{-\gamma_k - \bar{V}_j(g_j(x))\}, & \forall x \in \mathcal{I}_k, \\ \vdots \\ \exp\{-\gamma_{m_t} - \bar{V}_j(g_j(x))\}, & \forall x \in \mathcal{I}_{m_t}. \end{cases} \quad (24)$$

Notice that $\forall x \in \mathcal{I}_k, V(x; \mathbf{g}) - \gamma_k - \bar{V}_j(g_j(x)) = V_{-j}(x; \mathbf{g}) - \gamma_k \leq 0$, hence $\pi_t(x)$ is suitable for rejection sampling.

Finally, note that $\pi_t(x)$ is a mixture of truncated densities with non-overlapping supports. Indeed, let us define the mixture coefficients

$$\bar{\omega}_k \triangleq \exp\{-\gamma_k\} \int_{\mathcal{I}_k} p_j(x) dx \quad (25)$$

and normalize them as $\omega_k = \bar{\omega}_k / \sum_{k=0}^{m_t} \bar{\omega}_k$. Then,

$$\pi_t(x) \propto \sum_{k=1}^{m_t} \omega_k \chi_k(x) p_j(x) \quad (26)$$

where $\chi_k(x)$ is an indicator function ($\chi_k(x) = 1$ if $x \in \mathcal{I}_k$ and $\chi_k(x) = 0$ if $x \notin \mathcal{I}_k$). In order to draw from $\pi_t(x)$ we only need to be able to draw from truncated pieces of $p_j(x)$.

Let us remark that this construction is most natural in a Bayesian framework, when $p_0(x)$ is posterior density of the r.v. X given some data and $p_j(x)$ is the associated prior.

5 Examples

5.1 Toy example

We begin with a toy example in order to illustrate how to apply the GARS technique. Let us consider a target pdf $p_o(x) \propto p(x) = \exp\{-\sum_{i=0}^4 a_i x^i\}$ with $a_4 > 0$. Then, the potential function is a 4th order polynomial, $V(x; \mathbf{g}) = \sum_{i=0}^4 a_i x^i$. Every 4th order polynomial can be easily expressed as

$$V(x; \mathbf{g}) = \kappa + (\alpha + \beta x + \gamma x^2)^2 + (\delta + \eta x)^2, \quad (27)$$

where $\kappa, \alpha, \beta, \gamma, \eta$ and δ are real constants and, as a consequence, we can rewrite $V(x; \mathbf{g})$ using our notation as

$$\begin{aligned} V(x; \mathbf{g}) &= \kappa + \bar{V}_1(g_1(x)) + \bar{V}_2(g_2(x)) \\ &= \kappa + \sum_{i=1}^{n=2} \bar{V}_i(g_i(x)), \end{aligned} \quad (28)$$

where $\bar{V}_1(\vartheta) = \bar{V}_2(\vartheta) = \vartheta^2$, $g_1(x) \triangleq \alpha + \beta x + \gamma x^2$ is a 2nd-order polynomial and $g_2(x) \triangleq \delta + \eta x$ is linear. Since $\bar{V}_1(\vartheta) = \bar{V}_2(\vartheta)$ are convex, $\frac{d^2 g_1}{dx^2} = \gamma$ is constant and $g_2(x)$ is linear, it is straightforward to apply the basic GARS algorithm of Sect. 3 to this problem.

In particular, let $\mathcal{S}_t = \{s_1, s_2, \dots, s_{m_t}\}$ be the set of support points at the t th iteration of the algorithm. For each interval $\mathcal{I}_k = [s_k, s_{k+1}]$ we can build a suitable linear function $r_{i,k}(x) = a_{i,k}x + b_{i,k}$ (while $r_{i,k}(x) = g_2(x)$ for all x) using the method in Appendix in order to obtain a lower bound for the potential,

$$\begin{aligned} V(x; \mathbf{r}_k) &= \kappa + V(r_{1,k}(x)) + V(r_{2,k}(x)) \\ &\leq V(x; \mathbf{g}) = \kappa + V(g_1(x)) + V(g_2(x)), \end{aligned} \quad (29)$$

for all $x \in \mathcal{I}_k$. Note that, for this specific example, the proposal density

$$\pi_t^*(x) \propto \begin{cases} \exp\{-V(x; \mathbf{r}_0)\} & \text{if } x \in \mathcal{I}_0, \\ \vdots \\ \exp\{-V(x; \mathbf{r}_{m_t})\} & \text{if } x \in \mathcal{I}_{m_t} \end{cases} \quad (30)$$

is a piecewise pdf and could be used directly to draw candidate samples if an efficient method to sample from truncated Gaussian pdf's is at hand (Kotecha and Djurić 1999).

Alternatively, we can apply the complete GARS scheme. Specifically, since the modified potentials $V(x; \mathbf{r}_k)$ are convex in \mathcal{I}_k for all k , we can build a piecewise linear function $W_t(x)$ such that $W_t(x) \leq V(x; \mathbf{r}_k) \leq V(x; \mathbf{g}) \forall x \in \mathcal{I}_k$ and use the corresponding piecewise exponential density $\pi_t(x) \propto \exp\{-W_t(x)\}$ to draw candidates using the inversion method (Robert and Casella 2004, Chap. 2).

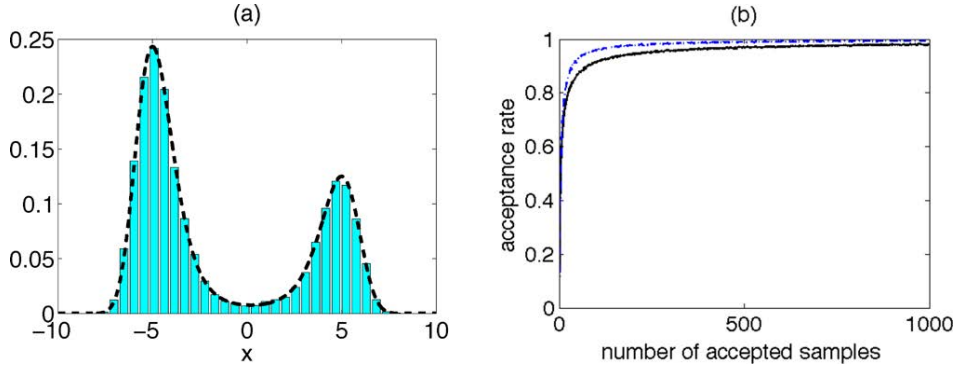


Fig. 5 (a) The target density $p_o(x) \propto \exp\{-\frac{1}{200}x^4 + \frac{1}{750}x^3 - \frac{1}{4}x^2 + \frac{1}{10}x\}$ (dashed line) and the normalized histogram of $N = 10,000$ samples obtained using the GARS algorithm. (b) The curve of acceptance rates (averaged over 10,000 simulations) as a function of the accepted

samples using the GARS algorithm with proposal function $\pi_t(x)$ (solid line) and using the method by Evans and Swartz (1998) with a log-transformation (dashed line)

Figure 5 illustrates the results obtained with this algorithm. The specific target density $p_o(x)$ results from the choice of parameters $\gamma = 0.0707$, $\beta = -0.0094$, $\alpha = -5.3033$, $\eta = 0.7071$, $\delta = 0$ and $\kappa = -28.1250$, and it is depicted in Fig. 5(a) with a dashed line. In the same plot, we observe the normalized histogram of 10,000 samples drawn with the GARS algorithm.

Let us note that, in this simple example, it is possible to analytically find the inflection points of the potential $V(x; \mathbf{g})$ and, as a consequence, we can apply the method in Evans and Swartz (1998) and compare it with the GARS technique.

Figure 5 shows the acceptance rates (averaged over 10,000 independent simulation runs) versus the number of accepted samples for the GARS algorithm using the proposal functions $\pi_t(x)$ in (16) (solid line) and the method by Evans and Swartz (1998) (dashed line), all of them with the same number of supports points. When the information about the inflection points of the potential $V(x; \mathbf{g})$ is available, the method by Evans and Swartz (1998) provides a tighter piecewise linear approximation of $V(x; \mathbf{g})$ and, as a result, attains a slightly higher acceptance rate. Unfortunately, for more complicated distributions the calculation of the inflection points of $V(x; \mathbf{g})$ becomes analytically intractable, as shown below in an experimental example.

5.2 Experimental example: target localization

In order to show how the proposed techniques can be used to draw samples from a multivariate distribution, we consider the problem of positioning a target in a 2-dimensional space using range measurements. This is a problem that appears frequently in localization applications using sensor networks (Ali et al. 2007; Patwari et al. 2003).

5.2.1 Experimental setup

We have carried out an experiment with a network consisting of four nodes. Three of them are placed at fixed positions and play the role of sensors that measure the strength of the radio signals transmitted by the target. The other node plays the role of the target to be localized. All nodes are bluetooth devices (Conceptronic CBT200U2A) with a nominal maximum range of 200 m.

The deployment of the network is sketched in Fig. 6(a). We consider a square monitored area of 4×4 m² and place the sensors at fixed positions $\mathbf{h}_1 = [h_{1,1} = 0.5, h_{1,2} = 1]$, $\mathbf{h}_2 = [h_{2,1} = 3.5, h_{2,2} = 1]$ and $\mathbf{h}_3 = [h_{3,1} = 2, h_{3,2} = 3]$, with all coordinates in meters. The target is located at $\mathbf{x} = [x_1 = 2.5, x_2 = 2]$.

The measurement provided by the i th sensor is denoted as a random variable Y_i . To describe the relationship between the observed radio signal strength, $Y_i = y_i$, and the target position, denoted by the random vector $\mathbf{X} = [X_1, X_2]$, we use the free space propagation model (Rappaport 2001)

$$Y_i = l - 10\gamma \log \left[\frac{D_i}{d_0} \right] + \Theta_i \quad (\text{dB}), \quad (31)$$

where the norm

$$D_i = \|\mathbf{X} - \mathbf{h}_i\| = \sqrt{[(X_1 - h_{i,1})^2 + (X_2 - h_{i,2})^2]}$$

is the distance between the i th sensor and the target, γ is a parameter that depends on the physical environment (for open space, $\gamma \approx 2$) and the constant l is the mean power received by each sensor when the target is located at a reference distance d_0 . The random variables Θ_i , $i = 1, 2, 3$, are i.i.d. Gaussian variates with density $N(\vartheta_i; 0, \sigma^2) \propto \exp\{-\frac{\vartheta_i^2}{2\sigma^2}\}$ that model the measurement noise.

For the experiment, the reference distance has been set to $d_0 = 0.3$ m. The parameters γ , l , and the noise variance σ^2

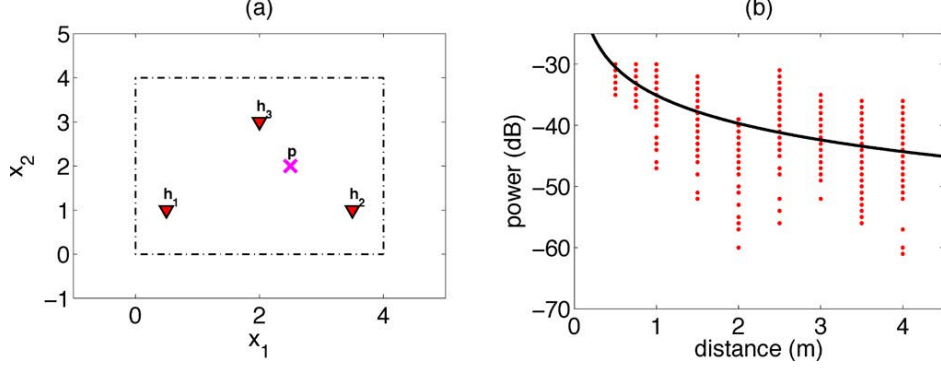


Fig. 6 (a) Deployment of the experimental sensor network over a rectangular surveillance area of $4 \times 4 \text{ m}^2$. The sensors are depicted with *triangles* while the target is depicted with a *cross*. (b) The least squares regression to adjust the parameters l and γ . The *points* indi-

cate the measurements collected by the sensors at different distances, and the *solid curve* denotes the function $\hat{l} - 10\hat{\gamma} \log[\frac{d}{d_0}]$ with $d_0 = 0.3$, $\hat{l} = -27.08 \text{ dB}$ and $\hat{\gamma} = 1.52$

5.2.2 Algorithm

Assume we collect M independent measurements from each sensor using the experimental setup just described. Let $\mathbf{Y} = [Y_{1,1}, \dots, Y_{1,M}, Y_{2,1}, \dots, Y_{2,M}, Y_{3,1}, \dots, Y_{3,M}]$ denote the random observation vector. For some fixed $\mathbf{Y} = \mathbf{y}$ the likelihood of the target position \mathbf{X} is Gaussian according to the model in (31), i.e.,

$$p(\mathbf{y}|\mathbf{x}) = \prod_{q=1}^3 \prod_{m=1}^M N(y_{q,m}; \hat{l} - 10\hat{\gamma} \log(\|\mathbf{X} - \mathbf{h}_q\|/d_0), \hat{\sigma}^2). \quad (32)$$

In order to perform inference on the position of the target, we aim at drawing from the posterior pdf

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}), \quad (33)$$

where $p(\mathbf{x})$ is the prior pdf of the target position \mathbf{X} . We assume $p(\mathbf{x}) = p(x_1, x_2) = p(x_1)p(x_2)$ where

$$p(x_k) = N(x_k; 1.5, 1/2), \quad k = 1, 2. \quad (34)$$

We apply the Gibbs sampler to draw N particles, denoted $\mathbf{x}^{(j)} = [x_1^{(j)}, x_2^{(j)}]$, from the posterior pdf $p(x_1, x_2|\mathbf{y}) \propto p(\mathbf{y}|x_1, x_2)p(x_1)p(x_2)$. The algorithm can be summarized as follows:

1. Set $j = 1$, and draw $x_2^{(1)}$ from the prior pdf $p(x_2)$.
2. Draw a sample $x_1^{(j)}$ from the conditional pdf $p(x_1|\mathbf{y}, x_2^{(j)})$, and set $\mathbf{x}^{(j)} = [x_1^{(j)}, x_2^{(j)}]$.

cate the measurements collected by the sensors at different distances, and the *solid curve* denotes the function $\hat{l} - 10\hat{\gamma} \log[\frac{d}{d_0}]$ with $d_0 = 0.3$, $\hat{l} = -27.08 \text{ dB}$ and $\hat{\gamma} = 1.52$

3. Draw a sample $x_2^{(j+1)}$ from the conditional pdf $p(x_2|\mathbf{y}, x_1^{(j)})$.
4. Increment $j = j + 1$. If $j > N$ 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 however, we have to be able to draw from the conditional densities $p(x_1|\mathbf{y}, x_2^{(j)})$ and $p(x_2|\mathbf{y}, x_1^{(j)})$. In general, these two conditional pdf's can be non-log-concave and can have several modes.

Next, we show how both $p(x_1|\mathbf{y}, x_2^{(j)})$ and $p(x_2|\mathbf{y}, x_1^{(j)})$ can be written using the potential-function notation in this paper, in order to sample from them using the proposed GARS method. Specifically, if we let $\mathbf{x}_1^{(j)} \triangleq [x_1, x_2^{(j)}]$ and $\mathbf{x}_2^{(j)} \triangleq [x_1^{(j)}, x_2]$, then we obtain that $p(x_1|\mathbf{y}, x_2^{(j)}) \propto \exp\{-V(x_1; \mathbf{g})\}$ and $p(x_2|\mathbf{y}, x_1^{(j)}) \propto \exp\{-V(x_2; \mathbf{g})\}$ where

$$V(x_k; \mathbf{g}) = \sum_{i=1}^{3M} \bar{V}_i(g_i(x_k)) + \bar{V}_{3M+1}(g_{3M+1}(x_k)), \quad (35)$$

and the functions $\bar{V}_i(g_i(x_k))$ have the form

$$\bar{V}_i(g_i(x_k)) = [y_{q,m} - \hat{l} + 10\hat{\gamma} \log(\|\mathbf{x}_k^{(j)} - \mathbf{h}_q\|/d_0)]^2, \quad (36)$$

with $k = 1, 2$, and the integers $q \in \{1, 2, 3\}$ and $m \in \{1, \dots, M\}$ are such that $i = (q - 1)M + m$ (in order to enumerate the elements in vector \mathbf{Y}). Finally

$$\bar{V}_{3M+1}(g_{3M+1}(x_k)) = [x_k^{(j)} - 1.5]^2. \quad (37)$$

Therefore, the vector \mathbf{g} ; consists of $3M$ nonlinearities

$$g_{(q-1)M+m}(x_k) = y_{q,m} - \hat{l} + 10\hat{\gamma} \log(\|\mathbf{x}_k^{(j)} - \mathbf{h}_q\|/d_0), \quad (38)$$

for $q = 1, 2, 3$ (sensors), $m = 1, \dots, M$ (measurements), $k = 1, 2$ and $j = 1, \dots, N$, plus one extra linear function $g_{3M+1}(x_k) = x_k - 1.5$.

Note that all the marginal potentials are purely quadratic functions, i.e., $\bar{V}_i(\vartheta) = \frac{1}{\sigma^2}\vartheta^2$ for $i = 1, \dots, 3M$, and $\bar{V}_{3M+1}(\vartheta) = \vartheta^2$.

The potential functions $V(x_k; \mathbf{g})$, $k = 1, 2$, are not convex in general. Their shape depends on the data set $\mathbf{Y} = \mathbf{y}$ and the fixed coordinates $x_1^{(j)}$ or $x_2^{(j)}$. Therefore, the ARS method cannot be applied to implement steps (2) and (3) of the Gibbs sampler. However, all the marginal potentials are convex and the support of the nonlinearities $g_i(x_k)$, $i = 1, \dots, 3M + 1$, can be partitioned as described in Sect. 4.2. As a consequence, we can use the proposed GARS technique to implement the Gibbs sampler. On the contrary, the form of (35), (36) and (37) makes the calculation of the inflection points (with respect x_k) intractable and, therefore, the methods of Evans and Swartz (1998) and Görür and Teh (2009) are not applicable in this example.

5.2.3 Results

We have run the Gibbs sampler (using the GARS algorithm to sample from the conditional pdf's) with three different data sets \mathbf{y} . In the first one we collected $M = 1$ observation per sensor, in the second one we recorded $M = 3$ observations per sensor and, finally, we obtained a data set with $M = 10$ measurements per sensors. The target was placed at $\mathbf{x} = [2.5, 2]$.

The average acceptance rate of the GARS algorithm was $\approx 30\%$ with $M = 1$, $\approx 37\%$ with $M = 3$ and $\approx 26\%$ with $M = 10$. Note that these rates are, indeed, averages, because the target pdf's are different at each step of the Markov chain (e.g., if $x_1^{(i)} \neq x_1^{(i-1)}$ then $p(x_2|\mathbf{y}, x_1^{(i)}) \neq p(x_2|\mathbf{y}, x_1^{(i-1)})$). The acceptance rates can be further improved by including additional support points in the initial set \mathcal{S}_0 .

Figures 7(a) and (d) show the shape of the true target density $p(x_1, x_2|\mathbf{y})$ with $M = 1$ and $M = 3$, respectively. Figures 7(b) and (e) depict the corresponding histograms using $N = 30,000$ samples. We observe that they approximate closely the shape of target pdf's. Finally, Figures 7(c) and (f) illustrate the normalized histograms corresponding to the number of proposed candidates which are needed to accept one sample. The acceptance rates can be further improved by including additional support points in the initial set \mathcal{S}_0 .

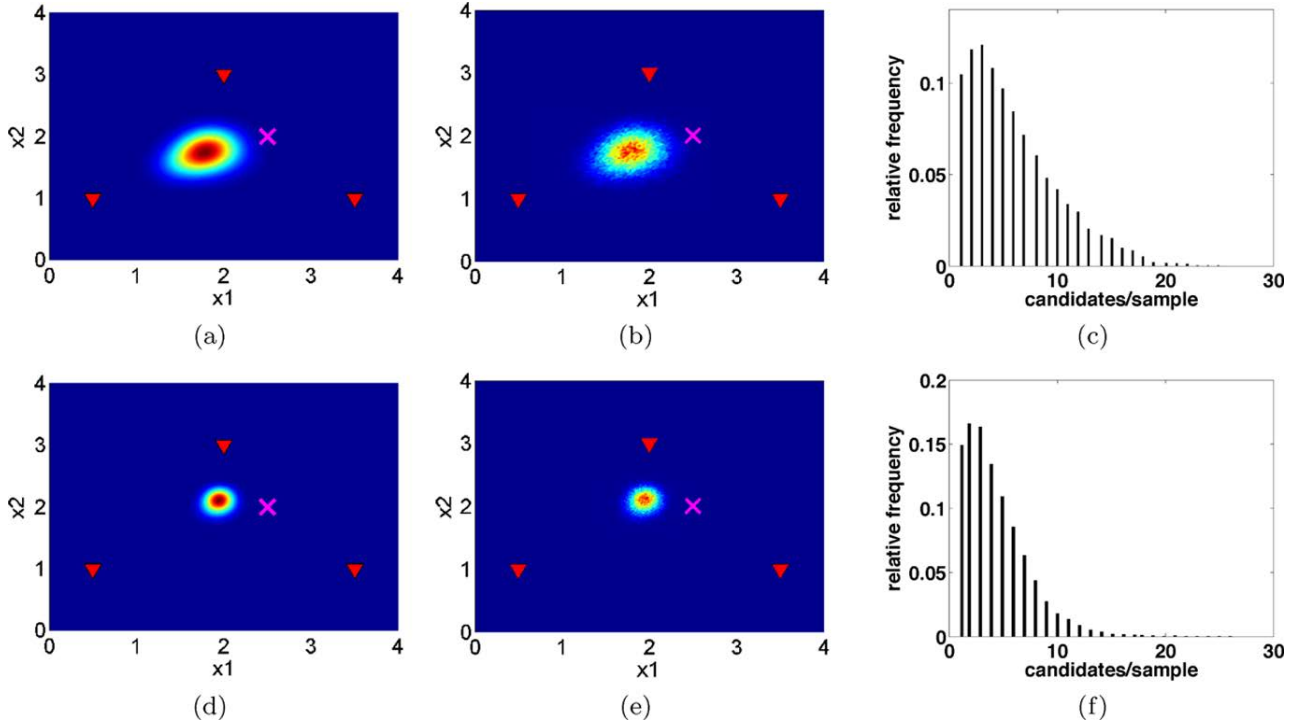


Fig. 7 (a) The shape of the true target density $p(x_1, x_2|\mathbf{y})$ with $M = 1$. (b) The normalized histogram, using $N = 20,000$ samples, corresponding to $p(x_1, x_2|\mathbf{y})$ with $M = 1$. (c) The normalized histogram corresponding to the number of proposed candidates which are needed to accept one sample, with $M = 1$. (d) The shape of the true target

density $p(x_1, x_2|\mathbf{y})$ with $M = 3$. (e) The normalized histogram, using $N = 20,000$ samples, corresponding to $p(x_1, x_2|\mathbf{y})$ with $M = 3$. (f) The normalized histogram corresponding to the number of proposed candidates which are needed to accept one sample, with $M = 3$

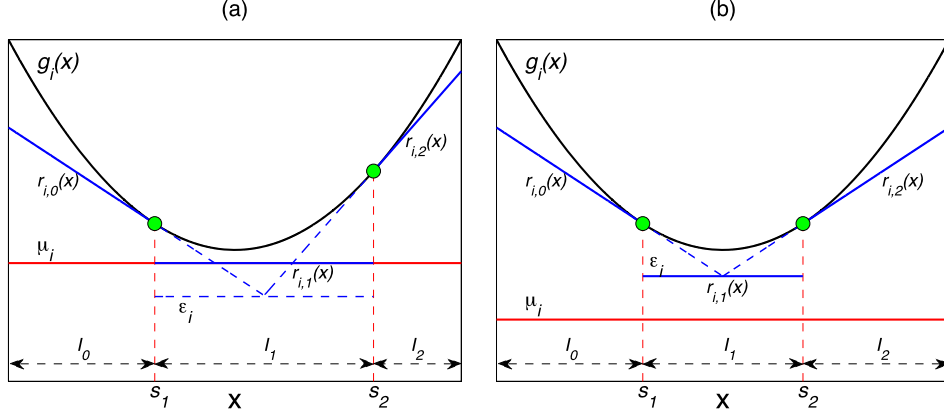


Fig. 9 Example of construction of the linear functions $r_{i,k}(x)$ when $|\mathcal{J}_i| = 0$ and $g_i(x)$ is non-monotonic and convex. We use two support points $\mathcal{S}_i = \{s_1, s_2\}$. Since $|\mathcal{J}_i \cap \mathcal{I}_0| = |\mathcal{J}_i \cap \mathcal{I}_2| = 0$, we construct $r_{i,0}(x)$ and $r_{i,2}(x)$ as target lines. In $\mathcal{I}_1 = [s_1, s_2]$, the nonlinear-

ity $g_i(x)$ is non-monotonic, since this interval contains the minimum value, and we set $r_{i,1}(x) = B_i$. (a) The value μ_i is greater than ϵ_i . So we choose $B_i = \mu_i$. (b) The value ϵ_i is greater than μ_i . Therefore, we set $B_i = \epsilon_i$

$g_i(x) = \mu_i$. If there is one solution, denoted x_i , then $\mathcal{J}_i = (-\infty, x_i]$ otherwise $\mathcal{J}_i = \emptyset$.

- If $g_i(x)$ is monotonic and $\frac{dg_i(x)}{dx} \times \frac{d^2g_i(x)}{dx^2} \leq 0$ then there may also exist at most one solution x_i . If x_i exists, then $\mathcal{J}_i = [x_i, +\infty)$, otherwise $\mathcal{J}_i = \emptyset$.

Take some \mathcal{I}_k , $k = 0, \dots, m_t$. With the above definition, \mathcal{J}_i is either disjoint of \mathcal{I}_k (except, maybe, for a single point, and $|\mathcal{J}_i \cap \mathcal{I}_k| = 0$ anyway) or a superset of the interval \mathcal{I}_k , i.e., $\mathcal{I}_k \cap \mathcal{J}_i = \mathcal{I}_k$. Any other possibility is excluded because the sets of support points $\mathcal{S}_0 \subseteq \dots \subseteq \mathcal{S}_t$ contain all simple estimates.

Now we provide a procedure for the construction of $r_{i,k}(x)$, $i \in \{1, \dots, n\}$, $k \in \{0, \dots, m_t\}$ with $x \in \mathcal{I}_0 = (-\infty, s_1]$ for $k = 0$, $x \in \mathcal{I}_k = [s_k, s_{k+1}]$ for $k = 1, \dots, m_t - 1$ and $x \in \mathcal{I}_{m_t} = [s_{m_t}, +\infty)$ for $k = m_t$.

1. If $\mathcal{I}_k \cap \mathcal{J}_i = \mathcal{I}_k$ then we choose the secant line $r_{i,k}(x)$ that connects the points $(s_k, g_i(s_k))$ and $(s_{k+1}, g_i(s_{k+1}))$.
2. If $|\mathcal{I}_k \cap \mathcal{J}_i| = 0$ and $g_i(x)$ is monotonic in \mathcal{I}_k (this is always true if $|\mathcal{J}_i| > 0$), then we set $r_{i,k}(x)$ as the tangent line to $g_i(x)$
 - (a) at s_k , if $\frac{dg_i(x)}{dx} \times \frac{d^2g_i(x)}{dx^2} \geq 0$ in \mathcal{I}_k , or
 - (b) at s_{k+1} , if $\frac{dg_i(x)}{dx} \times \frac{d^2g_i(x)}{dx^2} \leq 0$ in \mathcal{I}_k .
3. If $|\mathcal{J}_i| = 0$ and $g_i(x)$ is non-monotonic in \mathcal{I}_k then $r_{i,k}(x) = B_i$, where B_i is a bound of $g_i(x)$. Specially, we set

$$B_i \triangleq \begin{cases} \max[\mu_i, \epsilon_i], & \text{if } \frac{d^2g_i(x)}{dx^2} > 0, \\ \min[\mu_i, \epsilon_i], & \text{if } \frac{d^2g_i(x)}{dx^2} < 0, \end{cases} \quad (42)$$

where ϵ_i is the intersection point such that $r_{i,k-1}(x) = r_{i,k+1}(x) = \epsilon_i$ for $x \in \mathcal{I}_k$.

Figure 8 illustrates the construction of the linear functions $r_{i,k}(x)$ when $g_i(x)$ is non-monotonic and convex, and

$|\mathcal{J}_i| > 0$ (indeed, there are two simple estimates $x_{i,1}, x_{i,2}$). This depiction corresponds to the steps 1 and 2 of the proposed procedure.

Figure 9 depicts the construction of $r_{i,k}(x)$ when $|\mathcal{J}_i| = 0$ and corresponds to step 3 of the procedure. There are two support points and three intervals $\mathcal{I}_0, \mathcal{I}_1$ and \mathcal{I}_2 . The function $g_i(x)$ has a minimum in $\mathcal{I}_1 = [s_1, s_2]$. The linear functions $r_{i,0}(x)$ and $r_{i,2}(x)$ are tangent to $g_i(x)$ and they have an intersection at some $r_{i,0}(x') = r_{i,2}(x') = \epsilon_i$, $x' \in \mathcal{I}_1$.

Finally, note that, if $g_i(x)$ is monotonic, then either $\mathcal{J}_i = (-\infty, x_i]$ or $\mathcal{J}_i = [x_i, +\infty)$ and it occurs that $\mathcal{I}_0 = (-\infty, s_1] \subseteq \mathcal{J}_i$ or $\mathcal{I}_{m_t} = [s_{m_t}, +\infty) \subseteq \mathcal{J}_i$, respectively. In the first case, the construction algorithm yields $r_{i,0}(x) = g_i(s_1)$ while, in the second case, $r_{i,m_t}(x) = g_i(s_{m_t})$.

References

- Ali, A.M., Yao, K., Collier, T.C., Taylor, E., Blumstein, D., Girod, L.: An empirical study of collaborative acoustic source localization. In: Proceedings Information Processing in Sensor Networks, IPSN07, Boston (2007)
- Devroye, L.: Non-uniform Random Variate Generation. Springer, Berlin (1986)
- Evans, M., Swartz, T.: Random variate generation using concavity properties of transformed densities. J. Comput. Graph. Stat. **7**(4), 514–528 (1998)
- Gilks, W.R.: Derivative-free adaptive rejection sampling for Gibbs sampling. Bayesian Stat. **4**, 641–649 (1992)
- Gilks, W.R., Wild, P.: Adaptive rejection sampling for Gibbs sampling. J. R. Stat. Soc., Ser. C Appl. Stat. **41**(2), 337–348 (1992)
- Gilks, W.R., Robert, N.G.O., George, E.I.: Adaptive direction sampling. J. R. Stat. Soc., Ser. D Stat. **43**(1), 179–189 (1994)
- Gilks, W.R., Best, N.G., Tan, K.K.C.: Adaptive rejection metropolis sampling within Gibbs sampling. J. R. Stat. Soc., Ser. C Appl. Stat. **44**(4), 455–472 (1995)
- Görür, D., Teh, Y.W.: Concave convex adaptive rejection sampling. Technical Report, University College, London (2009)

- Hörmann, W.: A rejection technique for sampling from t-concave distributions. *ACM Trans. Math. Softw.* **21**(2), 182–193 (1995)
- Kotecha, J.H., Djurić, P.M.: Gibbs sampling approach for generation of truncated multivariate Gaussian random variables. In: *Proceedings of the IEEE ICASSP*, vol. 3, pp. 1757–1760 (1999)
- Künsch, H.R.: Recursive Monte Carlo filters: Algorithms and theoretical bounds. *Ann. Stat.* **33**(5), 1983–2021 (2005)
- Mayo, P., Rodenas, F., Verdú, G.: Comparing methods to denoise mammographic images. In: *Proceedings of the 26th IEEE EMBS*, vol. 1, pp. 247–250 (2004)
- Patwari, N., Hero, A.O., Perkins, M., Correal, N.S., O’Dea, R.J.: Relative location estimation in wireless sensor networks. *IEEE Trans. Signal. Process.* **51**(5), 2137–2148 (2003)
- Rappaport, T.S.: *Wireless Communications: Principles and Practice*, 2nd edn. Prentice-Hall, Upper Saddle River (2001)
- Reiss, R., Thomas, M.: *Statistical Analysis of Extreme Values: With Applications to Insurance, Finance, Hydrology and Other Fields*. Springer, Berlin (2007)
- Robert, C.P., Casella, G.: *Monte Carlo Statistical Methods*. Springer, Berlin (2004)