

# Two adaptive rejection sampling schemes for probability density functions log-convex tails

Luca Martino and Joaquín Míguez

Department of Signal Theory and Communications, Universidad Carlos III de Madrid.

Avenida de la Universidad 30, 28911 Leganés, Madrid, Spain.

E-mail: [luca@tsc.uc3m.es](mailto:luca@tsc.uc3m.es), [joaquin.miguez@uc3m.es](mailto:joaquin.miguez@uc3m.es)

## Abstract

Monte Carlo methods are often necessary for the implementation of optimal Bayesian estimators. A fundamental technique that can be used to generate samples from virtually any target probability distribution is the so-called rejection sampling method, which generates candidate samples from a proposal distribution and then accepts them or not by testing the ratio of the target and proposal densities. The class of adaptive rejection sampling (ARS) algorithms is particularly interesting because they can achieve high acceptance rates. However, the standard ARS method can only be used with log-concave target densities. For this reason, many generalizations have been proposed.

In this work, we investigate two different adaptive schemes that can be used to draw exactly from a large family of univariate probability density functions (pdf's), not necessarily log-concave, possibly multimodal and with tails of arbitrary concavity. These techniques are adaptive in the sense that every time a candidate sample is rejected, the acceptance rate is improved. The two proposed algorithms can work properly when the target pdf is multimodal, with first and second derivatives analytically intractable, and when the tails are log-convex in a infinite domain. Therefore, they can be applied in a number of scenarios in which the other generalizations of the standard ARS fail. Two illustrative numerical examples are shown.

## Index Terms

Rejection sampling; adaptive rejection sampling; ratio of uniforms method; particle filtering; Monte Carlo integration; volatility model.

## I. INTRODUCTION

Monte Carlo methods are often necessary for the implementation of optimal Bayesian estimators and, several families of techniques have been proposed [5, 12, 18, 1] that enjoy numerous applications. A fundamental technique that can be used to generate independent and identically distributed (i.i.d.) samples from virtually any target probability distribution is the so-called rejection sampling method, which generates candidate samples from a proposal distribution and then accepts them or not by testing the ratio of the target and proposal densities.

Several computationally efficient methods have been designed in which samples from a scalar random variable (r.v.) are accepted with high probability. The class of adaptive rejection sampling (ARS) algorithms [11] is particularly interesting because high acceptance rates can be achieved. The standard ARS algorithm yields a sequence of proposal functions that actually converge towards the target probability density distribution (pdf) when the procedure is iterated. As the proposal density becomes closer to the target pdf, the proportion of accepted samples grows (and, in the limit, can also converge to 1). However, this algorithm can only be used with log-concave target densities. A variation of the standard procedure, termed adaptive rejection Metropolis sampling (ARMS) [10] can also be used with multimodal pdf's. However, the ARMS algorithm is based on the Metropolis-Hastings algorithm, so the resulting samples form a Markov chain. As a consequence, they are correlated and, for certain multimodal densities, the chain can be easily trapped in a single mode (see, e.g., [23]).

Another extension of the standard ARS technique has been proposed in [14], where the same method of [11] is applied to  $T$ -concave densities, with  $T$  being a monotonically increasing transformation, not necessarily the logarithm. However, in practice it is hard to find useful transformations other than the logarithm and the technique cannot be applied to multimodal densities either. The method in [6] generalizes the technique of [14] 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 even for relatively simple 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 [13]. Then, every concave/convex element is handled using a method similar to the ARS procedure. A drawback of this technique is the need to decompose the target log-pdf into concave and convex components. Although this can be natural for some examples, it can

also be very tough for some practical problems (in general, identifying the convex and concave parts of the log-density may require the computation of all its inflection points). Moreover, the application of this technique to distributions with an infinite support requires that the tails of the log-densities be strictly concave. Other adaptive schemes that extend the standard ARS method for multimodal densities have been very recently introduced in [22, 23]. These techniques include the original ARS algorithm as a particular case and they can be relatively simpler than the approaches in [6] or [13] because they do not require the computation of the inflection points of the entire (transformed) target density. However, the method in [23] and the basic approach in [22] can also break down when the tails of the target pdf are log-convex, the same as the techniques in [11, 13, 14].

In this paper, we introduce two different ARS schemes that can be used to draw exactly from a large family of univariate pdf's, not necessarily log-concave and including cases in which the pdf has log-convex tails in an infinite domain. Therefore, the new methods can be applied to problems where the algorithms of, e.g., [11, 14, 13, 22] are invalid.

The first adaptive scheme described in this work was briefly suggested in [22] as alternative strategy. Here, we take this suggestion and elaborate it to provide a full description of the resulting algorithm. This procedure can be easy to implement and provide good performance as shown in a numerical example. However, it presents some technical requirements that can prevent its use with certain densities, as we also illustrate in a second example.

The second approach introduced here is more general and it is based on the ratio of uniforms (RoU) technique [3, 17, 25]. The RoU method enables us to obtain a two dimensional region  $\mathcal{A}$  such that drawing from the univariate target density is equivalent to drawing uniformly from  $\mathcal{A}$ . When the tails of the target density decay as  $1/x^2$  (or faster), the region  $\mathcal{A}$  is bounded and, in such case, we introduce an adaptive technique that generates a collection of non-overlapping triangular regions that cover  $\mathcal{A}$  completely. Then, we can efficiently use rejection sampling to draw uniformly from  $\mathcal{A}$  (by first drawing uniformly from the triangles). Let us remark that a basic adaptive rejection sampling scheme based on the RoU technique was already introduced in [19, 20] but it only works when the region  $\mathcal{A}$  is strictly convex<sup>1</sup>. The adaptive scheme that we introduce can also be used with non-convex sets.

The rest of the paper is organized as follows. The necessary background material is presented in Section II. The first adaptive procedure is described in Section III while the adaptive RoU scheme is introduced in Section IV. In Section V we present two illustrative examples and we conclude with a brief summary

<sup>1</sup>It can be seen as a RoU-based counterpart of the original ARS algorithm in [11], that requires the log-density to be concave.

and conclusions in Section VI.

## II. BACKGROUND

In this Section we recall some background material needed for the remaining of the paper. First, in Sections II-A and II-B we briefly review the rejection sampling method and its adaptive implementation, respectively. The difficulty of handling target densities with log-convex tails is discussed in Section II-C. Finally, we present the ratio of uniforms (RoU) method in Section II-D.

### A. Rejection sampling

Rejection sampling [3, Chapter 2] 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  $N$  i.i.d. 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. In order to improve the mean acceptance rate many adaptive variants have been proposed [6, 11, 13, 14, 22].

### B. An adaptive rejection sampling scheme

We describe the method in [22], that contains the the standard ARS algorithm of [11] as a particular case. Let us write the target pdf as  $p_o(x) \propto p(x) = \exp\{-V(x; \mathbf{g})\}$ ,  $\mathcal{D} \subset \mathbb{R}$ , where  $\mathcal{D} \subset \mathbb{R}$  is the support of  $p_o(x)$  and  $V(x; \mathbf{g})$  is termed the potential function of  $p_o(x)$ . In order to apply the technique of [22], we assume that the potential function can be expressed as the addition of  $n$  terms,

$$V(x; \mathbf{g}) \triangleq \sum_{i=1}^n \bar{V}_i(g_i(x)), \quad (1)$$

where the functions  $\bar{V}_i$ ,  $i = 1, \dots, n$ , referred to as marginal potentials, are convex while every  $g_i(x)$ ,  $i = 1, \dots, n$ , is either convex or concave. As a consequence, the potential  $V(x; \mathbf{g})$  is possibly non-convex (hence, the standard ARS algorithm cannot be applied) and can have rather complicated forms.

The sampling method is iterative. Assume that, after the  $(t-1)$ -th iteration, there is available a set of  $m_t$  distinct support points,  $\mathcal{S}_t = \{s_1, s_2, \dots, s_{m_t}\} \subset \mathcal{D}$ , sorted in ascending order, i.e.,  $s_1 < s_2 < \dots < s_{m_t}$ . From this set, we define  $m_t + 1$  intervals of the form  $\mathcal{I}_0 = (-\infty, s_1]$ ,  $\mathcal{I}_k = [s_k, s_{k+1}]$ ,  $k = 1, \dots, m_t - 1$ , and  $\mathcal{I}_{m_t} = [s_{m_t}, +\infty)$ . For each interval  $\mathcal{I}_k$ , it is possible to construct a vector of linear functions  $\mathbf{r}_k(x) = [r_{1,k}(x), \dots, r_{n,k}(x)]$  such that  $\bar{V}_i(r_{i,k}(x)) \leq \bar{V}_i(g_i(x))$  (see [22]) and, as a consequence,

$$V(x; \mathbf{r}_k) \leq V(x; \mathbf{g}), \quad \text{for all } x \in \mathcal{I}_k \quad (2)$$

and  $k = 0, \dots, m_t$ . Hence, it is possible to obtain  $\exp\{-V(x; \mathbf{r}_k)\} \geq p(x)$ ,  $\forall x \in \mathcal{I}_k$ . Moreover, the modified potential  $V(x; \mathbf{r}_k)$  is strictly convex in the interval  $\mathcal{I}_k$  and, therefore, it is straightforward to build a linear function  $w_k(x)$ , tangent to  $V(x; \mathbf{r}_k)$ , such that  $w_k(x) \leq V(x; \mathbf{r}_k)$  and  $\exp\{-w_k(x)\} \geq p(x)$  for all  $x \in \mathcal{I}_k$ .

With these ingredients, we can outline the generalized ARS algorithm as follows.

- 1) **Initialization.** Set  $i = 1$ ,  $t = 0$  and choose  $m_1$  support points,  $\mathcal{S}_1 = \{s_1, \dots, s_{m_1}\}$ .
- 2) **Iteration.** For  $t \geq 1$ , take the following steps.
  - From  $\mathcal{S}_t$ , determine the intervals  $\mathcal{I}_0, \dots, \mathcal{I}_{m_t}$ .
  - For  $k = 0, \dots, m_t$ , construct  $\mathbf{r}_k$ ,  $V(x; \mathbf{r}_k)$  and  $w_k(x)$ .
  - Let  $W_t(x) = w_k(x)$ , if  $x \in \mathcal{I}_k$ ,  $k \in \{0, \dots, m_t\}$ .
  - Build the proposal pdf  $\pi_t(x) \propto \exp\{-W_t(x)\}$ .
  - Draw  $x'$  from the proposal  $\pi_t(x)$  and  $u'$  from  $\mathcal{U}(0, 1)$ .
  - If  $u' \leq \frac{p(x')}{\exp\{-W_t(x')\}}$ , then accept  $x^{(i)} = x'$  and set  $\mathcal{S}_{t+1} = \mathcal{S}_t$ ,  $m_{t+1} = m_t$ ,  $i = i + 1$ .
  - 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  $m_{t+1} = m_t + 1$ .
  - Sort  $\mathcal{S}_{t+1}$  in ascending order and increment  $t = t + 1$ . If  $i > N$ , then stop the iteration.

Note that it is very easy to draw from  $\pi_t(x)$  because it consists of pieces of exponential densities. Moreover, every time we reject a sample  $x'$ , it is incorporated into the set of support points and, as a consequence, the shape of the proposal  $\pi_t(x)$  becomes closer to the shape of the target  $p_o(x)$  and the acceptance rate of the candidate samples is improved [22].

### C. Log-convex tails

The algorithm of Section II-B breaks down when the potential function  $V(x; \mathbf{g})$  has both an infinite support ( $x \in \mathcal{D} = \mathbb{R}$ ) and concave tails (i.e., the target pdf  $p_o(x)$  has log-convex tails). In this case, the function  $W_t(x)$  becomes constant over an interval of infinite length and we cannot obtain a proper

proposal pdf  $\pi_t(x)$ . To be specific, if  $V(x; \mathbf{g})$  is concave in the intervals  $(-\infty, s_1]$ ,  $[s_{m_t}, +\infty)$  or both, then  $w_0(x)$ ,  $w_{m_t}(x)$ , or both, are constant and, as a consequence,  $\int_{-\infty}^{+\infty} \exp\{-W_t(x)\}dx = +\infty$ .

This difficulty with the tails is actually shared by other adaptive rejection sampling techniques in the literature. A theoretical solution to the problem is to find an invertible transformation  $G : \mathcal{D} \rightarrow \mathcal{D}^*$ , where  $\mathcal{D}^* \subset \mathbb{R}$  is a bounded set [3, 15, 21, 26]. Then, we can define a random variable  $Y = G(x)$  with density  $q(y)$ , draw samples  $y^{(1)}, \dots, y^{(N)}$  and then convert them into samples  $x^{(1)} = G^{-1}(y^{(1)}), \dots, x^{(N)} = G^{-1}(y^{(N)})$  from the target pdf  $p_o(x)$  of the r.v.  $X$ . However, in practice, it is difficult to find a suitable transformation  $G$ , since the resulting density  $q(y)$  may not have a structure that makes sampling any easier than in the original setting.

A similar, albeit more sophisticated, approach is to use the method of [6]. In this case, we need to build a partition of the domain  $\mathcal{D}$  with disjoint intervals  $\mathcal{D} = \cup_{i=1}^m \mathcal{D}_i$  and then apply invertible transformations  $T_i : \mathcal{D}_i \rightarrow \mathbb{R}$ ,  $i = 1, \dots, m$ , to the target function  $p(x)$ . In particular, the intervals  $\mathcal{D}_1$  and  $\mathcal{D}_m$  contain the tails of  $p(x)$  and the method works correctly if the composed functions  $(T_1 \circ p)(x)$  and  $(T_m \circ p)(x)$  are concave. However, finding adequate  $T_1$  and  $T_m$  is not necessarily a simple task and, even if they are obtained, applying the algorithm of [6] requires the ability to compute all the inflection points of the target function  $p(x)$ .

#### D. Ratio of uniforms method

The RoU method [2, 17, 25] is a sampling technique that relies on the following result.

**Theorem 1:** Let  $q(x) \geq 0$  be a pdf known only up to a proportionality constant. If  $(u, v)$  is a sample drawn from the uniform distribution on the set

$$\mathcal{A} = \left\{ (v, u) \in \mathbb{R}^2 : 0 \leq u \leq \sqrt{q(v/u)} \right\}, \quad (3)$$

then  $x = \frac{v}{u}$  is a sample from  $q(x)$ .

**Proof:** See [3, Theorem 7.1].

Therefore, if we are able to draw uniformly from  $\mathcal{A}$ , then we can also draw from the pdf  $q_o(x) \propto q(x)$ . The cases of practical interest are those in which the region  $\mathcal{A}$  is bounded, and  $\mathcal{A}$  is bounded if, and only if, both  $\sqrt{q(x)}$  and  $x\sqrt{q(x)}$  are bounded. Moreover, the function  $\sqrt{q(x)}$  is bounded if, and only if, the target density  $q_o(x) \propto q(x)$  is bounded and, assuming we have a bounded target function  $q(x)$ , the function  $x\sqrt{q(x)}$  is bounded if, and only if, the tails of  $q(x)$  decay as  $1/x^2$  or faster. Chung97, Wakefield91 defining different region  $\mathcal{A}$  that could be used for fatter tails.

Figure 1 (a) depicts a bounded set  $\mathcal{A}$ . Note that, for every angle  $\alpha \in (-\pi/2, +\pi/2)$  rad, we can draw a straight line that passes through the origin  $(0,0)$  and contains points  $(v_i, u_i) \in \mathcal{A}$  such that  $x = \frac{v_i}{u_i} = \tan(\alpha)$ , i.e., every point  $(v_i, u_i)$  in the straight line with angle  $\alpha$  yields the same value of  $x$ .

From the definition of  $\mathcal{A}$ , we obtain  $u_i \leq q(x)$  and  $v_i = xu_i \leq x\sqrt{q(x)}$ , hence, if we choose the point  $(v_2, u_2)$  that lies on the boundary of  $\mathcal{A}$ ,  $u_2 = \sqrt{q(x)}$  and  $v_2 = x\sqrt{q(x)}$ . Moreover, we can embed the set  $\mathcal{A}$  in the rectangular region

$$\mathcal{R} = \left\{ (v', u') : 0 \leq u' \leq \sup_x \sqrt{q(x)}, \right. \\ \left. \inf_x x\sqrt{q(x)} \leq v' \leq \sup_x x\sqrt{q(x)} \right\}, \quad (4)$$

as depicted in Fig. 1 (b).

Once  $\mathcal{R}$  is constructed, it is straightforward to draw uniformly from  $\mathcal{A}$  by rejection sampling: simply draw uniformly from  $\mathcal{R}$  and then check whether the candidate point belongs to  $\mathcal{A}$ .

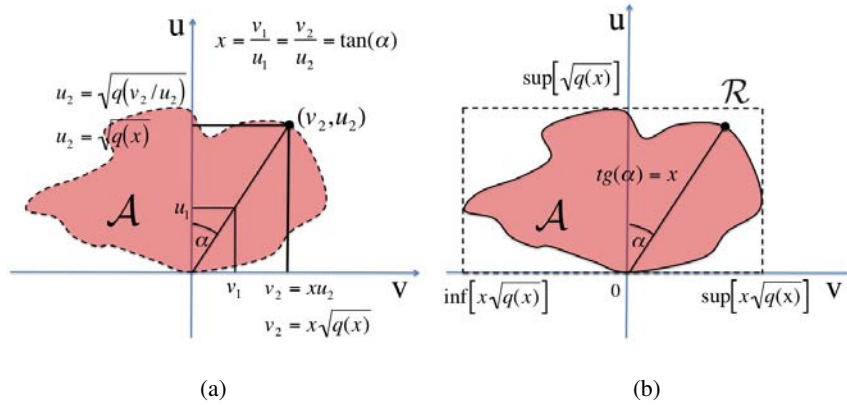


Fig. 1. **(a)** A bounded region  $\mathcal{A}$  and the straight line  $v = xu$  corresponding to the sample  $x = \tan(\alpha)$ . Every point in the intersection of the line  $v = xu$  and the set  $\mathcal{A}$  yields the same sample  $x$ . The point on the boundary,  $(v_2, u_2)$ , has coordinates  $v_2 = x\sqrt{q(x)}$  and  $u_2 = \sqrt{q(x)}$ . **(b)** If the two functions  $\sqrt{q(x)}$  and  $x\sqrt{q(x)}$  are bounded, the set  $\mathcal{A}$  is bounded and embedded in the rectangle  $\mathcal{R}$ .

### III. ADAPTIVE REJECTION SAMPLING WITH LOG-CONVEX TAILS

In this section, we investigate a strategy proposed in [22] to obtain an adaptive rejection sampling algorithm that remains valid the tails of the potential function  $V(x; \mathbf{g})$  are concave (i.e., the target pdf  $p_o(x)$  has log-convex tails).

Let us assume that for some  $j \in \{1, \dots, n\}$ , the pdf defined as

$$q(x) \propto \exp\{-\bar{V}_j(g_j(x))\} \quad (5)$$

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

$$V_{-j}(x; \mathbf{g}) \triangleq \sum_{i=1, i \neq j}^n \bar{V}_i(g_i(x)), \quad (6)$$

attained by removing  $\bar{V}_j(g_j(x))$  from  $V(x; \mathbf{g})$ . It is straightforward to obtain lower bounds  $\gamma_k \leq V_{-j}(x; \mathbf{g})$  by applying the procedure explained in the Appendix to the reduced potential  $V_{-j}(x; \mathbf{g})$  in every interval  $\mathcal{I}_k$ ,  $k = 0, \dots, m_t$ . Once these bounds are available, we set  $L_k \triangleq \exp\{-\gamma_k\}$  and build the piecewise proposal function

$$\pi_t(x) \propto \begin{cases} L_0 \exp\{-\bar{V}_j(g_j(x))\}, & \forall x \in \mathcal{I}_0, \\ \vdots \\ L_k \exp\{-\bar{V}_j(g_j(x))\}, & \forall x \in \mathcal{I}_k, \\ \vdots \\ L_{m_t} \exp\{-\bar{V}_j(g_j(x))\}, & \forall x \in \mathcal{I}_{m_t}. \end{cases} \quad (7)$$

Notice that, for all  $x \in \mathcal{I}_k$ , we have  $L_k \geq \exp\{-V_{-j}(x; \mathbf{g})\}$  and multiplying both sides of this inequality by the positive factor  $\exp\{-\bar{V}_j(g_j(x))\} \geq 0$ , we obtain

$$L_k \exp\{-\bar{V}_j(g_j(x))\} \geq \exp\{-V(x; \mathbf{g})\}, \quad \forall x \in \mathcal{I}_k,$$

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

$$\bar{\alpha}_k \triangleq L_k \int_{\mathcal{I}_k} q(x) dx \quad (8)$$

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

$$\pi_t(x) = \sum_{k=0}^{m_t} \alpha_k q(x) \chi_k(x) \quad (9)$$

where  $\chi_k(x)$  is an indicator function ( $\chi_k(x) = 1$  if  $x \in \mathcal{I}_k$  and  $\chi_k(x) = 0$  if  $x \notin \mathcal{I}_k$ ). The complete algorithm is summarized below.

- 1) **Initialization.** Set  $i = 1$ ,  $t = 0$  and choose  $m_1$  support points,  $\mathcal{S}_1 = \{s_1, \dots, s_{m_1}\}$ .
- 2) **Iteration.** For  $t \geq 1$ , take the following steps.
  - From  $\mathcal{S}_t$ , determine the intervals  $\mathcal{I}_0, \dots, \mathcal{I}_{m_t}$ .
  - Choose a suitable pdf  $q(x) \propto \exp\{-\bar{V}_j(g_j(x))\}$  for some  $j \in \{1, \dots, n\}$ .
  - For  $k = 0, \dots, m_t$ , compute the lower bounds  $\gamma_k \leq V_{-j}(x, \mathbf{g})$ ,  $\forall x \in \mathcal{I}_k$ , and set  $L_k = \exp\{-\gamma_k\}$  (see the Appendix).



- Build the proposal pdf  $\pi_t(x) \propto L_k q(x)$  for  $x \in \mathcal{I}_k$ ,  $k \in \{0, \dots, m_t\}$ .
- Draw  $x'$  from the proposal  $\pi_t(x)$  in Eq. (9) and  $u'$  from  $\mathcal{U}(0, 1)$ .
- If  $u' \leq \frac{\exp\{-V_{-j}(x'; \mathbf{g})\}}{L_k}$ , then accept  $x^{(i)} = x'$  and set  $\mathcal{S}_{t+1} = \mathcal{S}_t$ ,  $m_{t+1} = m_t$ ,  $i = i + 1$ .
- Otherwise, if  $u' > \frac{\exp\{-V_{-j}(x'; \mathbf{g})\}}{L_k}$ , then reject  $x'$ , set  $\mathcal{S}_{t+1} = \mathcal{S}_t \cup \{x'\}$  and update  $m_{t+1} = m_t + 1$ .
- Sort  $\mathcal{S}_{t+1}$  in ascending order and increment  $t = t + 1$ . If  $i > N$ , then stop the iteration.

When a sample  $x'$  drawn from  $\pi_t(x)$  is rejected,  $x'$  is added to the set of support points  $\mathcal{S}_{t+1} \triangleq \mathcal{S}_t \cup \{x'\}$ . Hence, we improve the piecewise constant approximation of  $V_{-j}(x; \mathbf{g})$  (formed by the upper bounds  $L_k$ ) and proposal pdf  $\pi_{t+1}(x) \propto L_k \exp\{-\bar{V}_j(g_j(x))\}$ ,  $\forall x \in \mathcal{I}_k$ , becomes closer to the target pdf  $p_o(x)$ .

Figure 2 (a) illustrates the reduced potential  $\exp\{-V_{-j}(x; \mathbf{g})\}$  and its stepwise approximation  $L_k = \exp\{-\gamma_k\}$  built using  $m_t = 4$  support points. Figure 2 (b) depicts the target pdf  $p_o(x)$  jointly with the proposal pdf  $\pi_t(x)$ , composed by weighted pieces of the function  $\exp\{-\bar{V}_j(g_j(x))\}$  (shown, in dashed line).

This procedure is feasible only if we can find a pair  $\bar{V}_j, g_j$ , for some  $j \in \{1, \dots, n\}$ , such that the pdf  $q(x) \propto \exp\{-\bar{V}_j(g_j(x))\}$  is

- is analytically integrable in every interval  $\mathcal{I} \subset \mathcal{D}$ , given otherwise the weights  $\alpha_1, \dots, \alpha_{m_t}$  in Eq. (9) cannot be computed in general, and
- is easy to draw from when truncated into a finite or an infinite interval, since otherwise we cannot generate samples from it.

Note that in order to draw from  $\pi_t(x)$  we need to be able to draw from and to integrate analytically truncated pieces of  $q(x)$ . This procedure is possible only if we have on hand a suitable pdf  $q(x) \propto \exp\{-\bar{V}_j(g_j(x))\}$ .

The technique that we introduce in the next section, based on ratio of uniforms method, overcomes these constraints.

#### IV. ADAPTIVE ROU SCHEME

The RoU method can be a useful tool to deal with target pdf's  $p_o(x) \propto p(x) = \exp\{-V(x; \mathbf{g})\}$  where  $V(x; \mathbf{g})$  of the form in Eq. (1) in a infinite domain and with tails of arbitrary concavity. Indeed, as long as the tails of  $p(x)$  decay as  $1/x^2$  (or faster), the region  $\mathcal{A}$  defined by the RoU transformation is bounded and, therefore, the problem of drawing samples from the target density becomes equivalent to that of drawing uniformly from a finite region.

In order to draw uniformly from  $\mathcal{A}$  by rejection sampling, we need to build a suitable set  $\mathcal{P} \supseteq \mathcal{A}$  for which we are able to generate uniform samples easily. This, in turn, requires knowledge of upper bounds

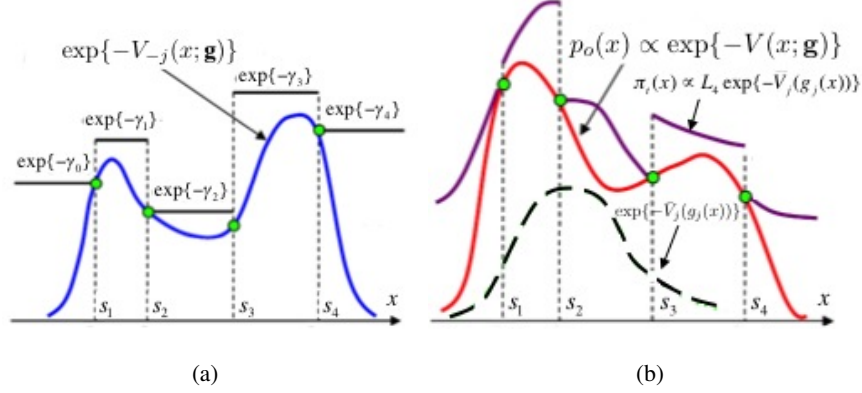


Fig. 2. **(a)** Example of the function  $\exp\{-V_{-j}(x; \mathbf{g})\}$  and its stepwise approximation  $L_k = \exp\{-\gamma_k\}$ ,  $k = 0, \dots, m_t = 4$ , constructed with the proposed technique using four support points  $\mathcal{S}_t = \{s_1, s_2, s_3, s_4\}$ . **(b)** Our target pdf  $p_o(x) \propto \exp\{-V(x; \mathbf{g})\} = \exp\{-V_{-j}(x; \mathbf{g}) - \bar{V}_j(g_j(x))\}$  obtained by multiplying the previous function  $\exp\{-V_{-j}(x; \mathbf{g})\}$  times  $\exp\{-\bar{V}_j(g_j(x))\}$  (shown with dashed line). The picture also shows the shape of the proposal density  $\pi_t(x)$  consisting of pieces of the function  $\exp\{-\bar{V}_j(g_j(x))\}$  scaled by the constant values  $L_k = \exp\{-\gamma_k\}$ .

for the functions  $\sqrt{p(x)}$  and  $x\sqrt{p(x)}$ , as described in Section II-D. Indeed, note that to apply the RoU transformation directly to an improper proposal  $\pi_t(x) \propto \exp\{W_t(x)\}$ , with  $\exp\{W_t(x)\} \geq p(x)$  and  $\int_{-\infty}^{+\infty} \pi_t(x) dx \rightarrow +\infty$ , provides us an unbounded region  $\mathcal{A}$  so that this strategy is useless.

In the sequel, we describe how to adaptively build a polygonal sets  $\mathcal{P}_t$  ( $t = 1, 2, \dots$ ) composed by non-overlapping triangular subsets that embed the region  $\mathcal{A}$ . To draw uniformly from  $\mathcal{P}_t$ , we randomly select one triangle (with probability proportional to its area) and then generate uniformly a sample point from it using the algorithm in the Appendix. If the sample belongs to  $\mathcal{A}$ , it is accepted, and otherwise is rejected and the region  $\mathcal{P}_t$  is improved by adding another triangular piece.

In Section IV-A below, we provide a detailed description of the proposed adaptive technique. Details of the computation of some bounds that are necessary for the algorithm are given in Section IV-B and in the Appendix.

#### A. Adaptive Algorithm

Given a set of support points

$$\mathcal{S}_t = \{s_1, \dots, s_{m_t}\},$$

we assume that there always exists  $k' \in \{0, \dots, m_t\}$  such that  $s_{k'} = 0$ , i.e., the point zero is included in the set of support points  $\mathcal{S}_t$ . Hence, in an interval  $\mathcal{I}_k = [s_k, s_{k+1}]$  the points  $s_k$  and  $s_{k+1}$  are both non-positive or both non-negative.

We also assume that we are able to compute upper bounds  $L_k^{(1)} \geq \sqrt{p(x)}$ ,  $L_k^{(2)} \geq x\sqrt{p(x)}$  (for  $x > 0$ ) and  $L_k^{(3)} \geq -x\sqrt{p(x)}$  (for  $x < 0$ ) within every interval  $x \in \mathcal{I}_k = [s_k, s_{k+1}]$ ,  $k = 0, \dots, m_t$ , where  $\mathcal{I}_0 = (-\infty, s_1]$  and  $\mathcal{I}_{m_t} = [s_{m_t}, +\infty)$ .

1) *Construction of  $\mathcal{P}_t \supseteq \mathcal{A}$ :* Consider the construction in Figure 3 (a). For a pair of angles  $\alpha_k \triangleq \arctan(s_k)$  and  $\alpha_{k+1} \triangleq \arctan(s_{k+1})$ , we define the subset  $\mathcal{A}_k \triangleq \mathcal{A} \cap \mathcal{J}_k$ , where  $\mathcal{J}_k$  is the cone with vertex at the origin  $(0, 0)$  and delimited by the two straight lines that form angles  $\alpha_k$  and  $\alpha_{k+1}$  w.r.t. the  $u$  axis. Note that, clearly,  $\mathcal{A} = \cup_{k=0}^{m_t} \mathcal{A}_k$ .

For each  $k = 0, \dots, m_t$  the subset  $\mathcal{A}_k$  is contained in a piece of circle  $\mathcal{C}_k$  ( $\mathcal{A}_k \subseteq \mathcal{C}_k$ ) delimited by the angles  $\alpha_k$ ,  $\alpha_{k+1}$  and radius

$$r_k = \begin{cases} \sqrt{\left(L_k^{(1)}\right)^2 + \left(L_k^{(2)}\right)^2}, & \text{if } s_k, s_{k+1} \geq 0 \\ \sqrt{\left(L_k^{(1)}\right)^2 + \left(L_k^{(3)}\right)^2}, & \text{if } s_k, s_{k+1} \leq 0 \end{cases} \quad (10)$$

also shown (with a dashed line) in Fig. 3 (a).

Unfortunately, it is not straightforward to generate samples uniformly from  $\mathcal{C}_k$ , but we can easily draw samples uniformly from a triangle in the plane  $\mathbb{R}^2$ , as explained in the Appendix. Hence, we can choose an arbitrarily a point in the arc of circumference that delimits  $\mathcal{C}_k$  (e.g., the point  $(L_k^{(2)}, L_k^{(1)})$  in Fig. 3 (a)) and calculate the tangent line to the arc at this point. In this way, we build a triangular region  $\mathcal{T}_k$  such that  $\mathcal{T}_k \supseteq \mathcal{C}_k \supseteq \mathcal{A}_k$ , with a vertex at  $(0, 0)$ .

We can repeat the procedure for every  $k = 0, \dots, m_t$ , with different angles  $\alpha_k = \arctan(s_k)$  and  $\alpha_{k+1} = \arctan(s_{k+1})$ , and define the polygonal region  $\mathcal{P}_t \triangleq \cup_{k=0}^{m_t} \mathcal{T}_k$  composed by non-overlapping triangular subsets. Note that, by construction,  $\mathcal{P}_t$  embeds the entire region  $\mathcal{A}$ , i.e.,  $\mathcal{A} \subseteq \mathcal{P}_t$ .

Figure 3 summarizes the procedure to build the set  $\mathcal{P}_t$ . In Figure 3 (a) we show the construction of a triangle  $\mathcal{T}_k$  within the angles  $\alpha_k = \arctan(s_k)$ ,  $\alpha_{k+1} = \arctan(s_{k+1})$ , using the upper bounds  $L_k^{(1)}$  and  $L_k^{(2)}$  for a single interval  $x \in \mathcal{I}_k = [s_k, s_{k+1}]$ . Figure 3 (b) illustrates the entire region  $\mathcal{P}_t \triangleq \cup_{k=0}^{m_t} \mathcal{T}_k$  formed by  $m_t + 1 = 6$  triangular subsets that covers completely the region  $\mathcal{A}$ , i.e.,  $\mathcal{A} \subseteq \mathcal{P}_t$ .

2) *Adaptive sampling:* To generate samples uniformly in  $\mathcal{P}_t$ , we first have to select a triangle proportionally to the areas  $|\mathcal{T}_k|$ ,  $k = 0, \dots, m_t$ . Therefore, we define the normalized weights

$$w_k \triangleq \frac{|\mathcal{T}_k|}{\sum_{i=0}^{m_t} |\mathcal{T}_i|}, \quad (11)$$

and then we choose a triangular piece by drawing an index  $k' \in \{0, \dots, m_t\}$  from the probability distribution  $P(k) = w_k$ . Using the method in the Appendix, we can easily generate a point  $(v', u')$  uniformly in the selected triangular region  $\mathcal{T}_{k'}$ . If this point  $(v', u')$  belongs to  $\mathcal{A}$ , we accept the sample

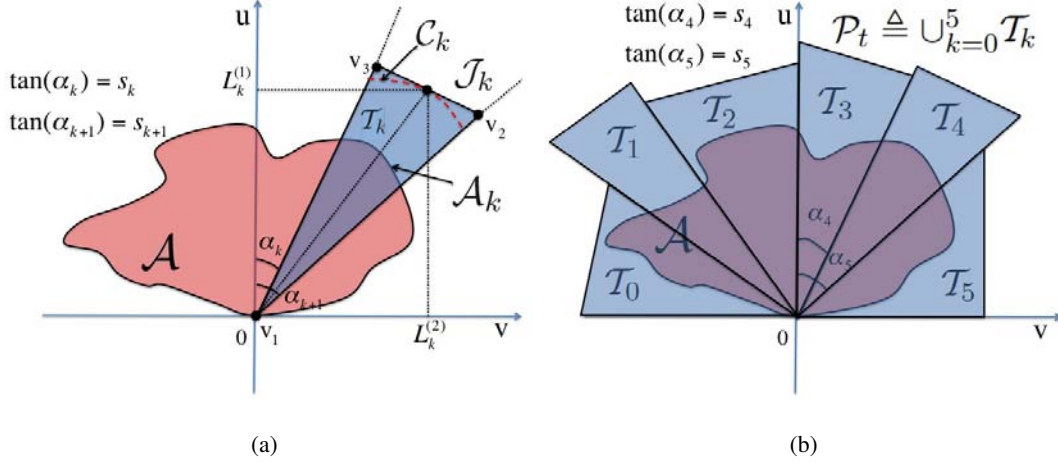


Fig. 3. (a) A region  $\mathcal{A}$  constructed by the RoU method and a triangular region  $\mathcal{T}_k$  defining by the vertices  $\mathbf{v}_1$ ,  $\mathbf{v}_2$  and  $\mathbf{v}_3$  built using the upper bounds  $L_k^{(1)}$ ,  $L_k^{(2)}$  for the functions  $\sqrt{p(x)}$  and  $x\sqrt{p(x)}$  for all  $x \in \mathcal{I}_k = [s_k, s_{k+1}]$ . The dashed line depicts the piece of circumference  $\mathcal{C}_k$  with radius  $r_k = \sqrt{(L_k^{(1)})^2 + (L_k^{(2)})^2}$ . The set  $\mathcal{T}_k$  embeds the subset  $\mathcal{A}_k = \mathcal{A} \cap \mathcal{J}_k$  where  $\mathcal{J}_k$  is the cone defined as  $\mathbf{v} \in \mathcal{J}_k$  if and only if  $\mathbf{v} = \theta_1 \mathbf{v}_2 + \theta_2 \mathbf{v}_3$  and  $\theta_1, \theta_2 \geq 0$ . (b) Construction of the polygonal region  $\mathcal{P}_t = \cup_{k=0}^5 \mathcal{T}_k$  using  $m_t = 5$  support points, i.e.,  $\mathcal{S}_t = \{s_1, s_2, s_3 = 0, s_4, s_5\}$ . Observe that each triangle  $\mathcal{T}_k$  has a vertex at  $(0, 0)$ . The set  $\mathcal{P}_t$  covers completely the region  $\mathcal{A}$  obtained by the RoU method, i.e.,  $\mathcal{A} \subset \mathcal{P}_t$ .

$x' = v'/u'$  and set  $m_{t+1} = m_t$ ,  $\mathcal{S}_{t+1} = \mathcal{S}_t$  and  $\mathcal{P}_t = \mathcal{P}_t$ . Otherwise, we discard the sample  $x' = v'/u'$  and incorporate it to the set of support points,  $\mathcal{S}_{t+1} = \mathcal{S}_t \cup \{x'\}$ , so that  $m_{t+1} = m_t + 1$  and the region  $\mathcal{P}_t$  is improved by adding another triangle.

3) *Summary of the algorithm:* The adaptive RS algorithm to generate  $N$  samples from  $p(x)$  using RoU is summarized as follows.

- 1) **Initialization.** Start with  $i = 1$ ,  $t = 0$  and choose  $m_1$  support points,  $\mathcal{S}_1 = \{s_1, \dots, s_{m_1}\}$ .
- 2) **Iteration.** For  $t \geq 1$ , take the following steps.
  - From  $\mathcal{S}_t$ , determine the intervals  $\mathcal{I}_0, \dots, \mathcal{I}_{m_t}$ .
  - Compute the upper bounds  $L_k^{(j)}$ ,  $j = 1, 2, 3$ , for each  $k = 0, \dots, m_t$  (see Section IV-B and the Appendix).
  - Construct the triangular regions  $\mathcal{T}_k$  as described in Section IV-A1,  $k = 0, \dots, m_t$ .
  - Calculate the area  $|\mathcal{T}_k|$  of every triangle, and compute the normalized weights

$$w_k \triangleq \frac{|\mathcal{T}_k|}{\sum_{i=0}^{m_t} |\mathcal{T}_i|} \quad (12)$$

with  $k = 0, \dots, m_t$ .

- Draw an index  $k' \in \{0, \dots, m_t\}$  from the probability distribution  $P(k) = w_k$ .

- Generate a point  $(v', u')$  uniformly from the region  $\mathcal{T}_{k'}$  (see the Appendix).
- If  $u' \leq \sqrt{p\left(\frac{v'}{u'}\right)}$ , then accept the sample  $x^{(i)} = x' = \frac{v'}{u'}$ , set  $i = i + 1$ ,  $\mathcal{S}_{t+1} = \mathcal{S}_t$  and  $m_{t+1} = m_t$ .
- Otherwise, if  $u' > \sqrt{p\left(\frac{v'}{u'}\right)}$ , then reject the sample  $x' = \frac{v'}{u'}$ , set  $\mathcal{S}_{t+1} = \mathcal{S}_t \cup \{x'\}$  and  $m_{t+1} = m_t + 1$ .
- Sort  $\mathcal{S}_{t+1}$  in ascending order and increment  $t = t + 1$ . If  $i > N$ , then stop the iteration.

It is interesting to note that the region  $\mathcal{P}_t$  is equivalent, in the domain of  $x$ , to a proposal function  $\pi_t(x)$  formed by pieces of reciprocal uniform distributions scaled and translated (as seen in Section II-D), i.e.,  $\pi_t(x) \propto 1/(\lambda_k x + \beta_k)^2$  in every interval  $\mathcal{I}_k$ , for some  $\lambda_k$  and  $\beta_k$ .

### B. Bounds for other potentials

In this section we provide the details on the computation of the bounds  $L_k^{(j)}$ ,  $j = 1, 2, 3$ , needed for the implementation of the algorithm above.

We associate a potential  $V^{(j)}$  to each function of interest. Specifically, since  $p(x) \propto \exp\{-V(x; \mathbf{g})\}$  we readily obtain that

$$\begin{aligned} \sqrt{p(x)} &\propto \exp\{-V^{(1)}(x; \mathbf{g})\}, \quad \text{with } V^{(1)}(x; \mathbf{g}) \triangleq \frac{1}{2}V(x; \mathbf{g}), \\ x\sqrt{p(x)} &\propto \exp\{-V^{(2)}(x; \mathbf{g})\}, \quad \text{with } V^{(2)}(x; \mathbf{g}) \triangleq \frac{1}{2}V(x; \mathbf{g}) - \log(x), (x > 0), \\ -x\sqrt{p(x)} &\propto \exp\{-V^{(3)}(x; \mathbf{g})\}, \quad \text{with } V^{(3)}(x; \mathbf{g}) \triangleq \frac{1}{2}V(x; \mathbf{g}) - \log(-x), (x < 0), \end{aligned} \quad (13)$$

respectively.

Note that it is equivalent to maximize the functions  $\sqrt{p(x)}$ ,  $x\sqrt{p(x)}$ ,  $-x\sqrt{p(x)}$  w.r.t.  $x$  and to minimize the corresponding potentials  $V^{(j)}$ ,  $j = 1, 2, 3$ , also w.r.t.  $x$ . As a consequence, we may focus on the calculation of lower bounds  $\gamma_k^{(j)} \leq V^{(j)}(x; \mathbf{g})$  in an interval  $x \in \mathcal{I}_k$ , related to the upper bounds as  $L_k^{(j)} = \exp\{-\gamma_k^{(j)}\}$ ,  $j = 1, 2, 3$  and  $k = 0, \dots, m_t$ . This problem is far from trivial, though. Even for very simple marginal potentials,  $\bar{V}_i$ ,  $i = 1, \dots, n$ , the potential functions,  $V^{(j)}$ ,  $j = 1, 2, 3$ , can be highly multimodal w.r.t.  $x$  [23].

In the Appendix we describe a procedure to find a lower bound  $\gamma_k$  for the potential  $V(x; \mathbf{g})$ . We can apply the same technique to the function  $V^{(1)}(x; \mathbf{g}) = \frac{1}{2}V(x; \mathbf{g})$ , associated to the function  $\sqrt{p(x)}$ , since  $V^{(1)}$  is a scaled version of the generalized system potential  $V(x; \mathbf{g})$ . Therefore, we can easily compute a lower bound  $\gamma_k^{(1)} \leq V^{(1)}(x; \mathbf{g})$  in the interval  $\mathcal{I}_k$ .

The procedure in the Appendix can also be applied to find upper bounds for  $x\sqrt{p(x)}$ , with  $x > 0$ , and  $-x\sqrt{p(x)}$  with  $x < 0$ . Indeed, recalling that the associated potentials are  $V^{(2)}(x; \mathbf{g}) = \frac{1}{2}V(x; \mathbf{g}) - \log(x)$ ,

$x > 0$ , and  $V^{(3)}(x; \mathbf{g}) = \frac{1}{2}V(x; \mathbf{g}) - \log(-x)$ ,  $x < 0$ , it is straightforward to realize that the corresponding modified potentials

$$\begin{aligned} V^{(2)}(x; \mathbf{r}_k) &\triangleq \frac{1}{2}V(x; \mathbf{r}_k) - \log(x), \\ V^{(3)}(x; \mathbf{r}_k) &\triangleq \frac{1}{2}V(x; \mathbf{r}_k) - \log(-x), \end{aligned} \quad (14)$$

are convex in  $\mathcal{I}_k$ , since the functions  $-\log(x)$  ( $x > 0$ ) and  $-\log(-x)$  ( $x < 0$ ), are also convex. Therefore, it is always possible to compute lower bounds  $\gamma_k^{(2)} \leq V^{(2)}(x; \mathbf{g})$  and  $\gamma_k^{(3)} \leq V^{(3)}(x; \mathbf{g})$ .

The corresponding upper bounds are  $L_k^{(j)} = \exp\{-\gamma_k^{(j)}\}$ ,  $j = 1, 2, 3$  for all  $x \in \mathcal{I}_k$ .

### C. Heavier tails

The standard RoU method can be easily generalized in this way: given the area

$$\mathcal{A}_\rho = \left\{ (v, u) : 0 \leq u \leq [p(v/u^\rho)]^{1/(\rho+1)} \right\}, \quad (15)$$

if we are able to draw uniformly from it the pair  $(v, u)$  then  $x = \frac{v}{u^\rho}$  is a sample from  $p_o(x) \propto p(x)$ .

The region  $\mathcal{A}_\rho$  is bounded if  $[p(x)]^{1/(\rho+1)}$  and  $x[p(x)]^{\rho/(\rho+1)}$  are bounded. It occurs when  $p(x)$  is bounded and the its tails decay as  $1/x^{(\rho+1)/\rho}$  or faster. Hence, for  $\rho > 1$  we can handle pdf's with fatter tails than with the standard RoU method.

In this case the potentials associated to our target pdf  $p_o(x) \propto \exp\{-V(x; \mathbf{g})\}$  are

$$\begin{aligned} V^{(1)}(x; \mathbf{g}) &\triangleq \frac{1}{\rho+1}V(x; \mathbf{g}), \\ V^{(2)}(x; \mathbf{g}) &\triangleq \frac{\rho}{\rho+1}V(x; \mathbf{g}) - \log(x), \text{ for } x > 0 \\ V^{(3)}(x; \mathbf{g}) &\triangleq \frac{\rho}{\rho+1}V(x; \mathbf{g}) - \log(-x), \text{ for } x < 0 \end{aligned} \quad (16)$$

Obviouly, we can use the technique in the Appendix to obtain lower bounds and to build an extended adaptive RoU scheme to tackle target pdf's with fatter tails. Moreover, the constant parameter  $\rho$  affects to the shape of  $\mathcal{A}_\rho$  and, as a consequence, also the acceptance rate of a rejection sampler. Hence, in some case it is interesting to find the optimal value of  $\rho$  that maximizes the acceptance rate.

## V. EXAMPLES

In this section we illustrate the application of the proposed techniques. The first example is devoted to compare the performance of the first adaptive technique in Section III and the adaptive RoU algorithm in Section IV using an artificial model. In the second example, we apply adaptive RoU scheme as a building block of an accept/reject particle filter [18] for inference in a financial volatility model. In this second example the first adaptive technique cannot implemented. Moreover, note that the other generalizations of the standard ARS method proposed in [6, 11, 13, 14] cannot be applied in both examples.

### A. Artificial example

Let  $x$  be a positive scalar signal of interest,  $x \in \mathbb{R}^+$ , with exponential prior,  $q(x) \propto \exp\{-\lambda x\}$ ,  $\lambda > 0$ , and consider the system with three observations,  $\mathbf{y} = [y_1, y_2, y_3] \in \mathbb{R}^{n=3}$ ,

$$y_1 = a \exp(-bx) + \vartheta_1, \quad y_2 = c \log(dx + 1) + \vartheta_2, \quad \text{and} \quad y_3 = (x - e)^2 + \vartheta_3, \quad (17)$$

where  $\vartheta_1, \vartheta_2, \vartheta_3$  are independent noise variables and  $a, b, c, d, e$  are constant parameters. Specifically,  $\vartheta_1$  and  $\vartheta_2$  are independent with generalized gamma pdf  $\Gamma_g(\vartheta_i; \alpha_i, \beta_i) \propto \vartheta_i^{\alpha_i} \exp\{-\vartheta_i^{\beta_i}\}$ ,  $i = 1, 2$ , with parameters  $\alpha_1 = 4$ ,  $\beta_1 = 2$  and  $\alpha_2 = 2$ ,  $\beta_2 = 2$ , respectively. The variable  $\vartheta_3$  has a Gaussian density  $N(\vartheta_3; 0, 1/2) \propto \exp\{-\vartheta_3^2\}$ .

Our goal is to generate samples from the posterior pdf  $p(x|\mathbf{y}) \propto p(\mathbf{y}|x)q(x)$ . Given the system (17) our target density can be written as

$$p_o(x) = p(x|\mathbf{y}) \propto \exp\{-V(x; \mathbf{g})\}, \quad (18)$$

where the potential is

$$\begin{aligned} V(x; \mathbf{g}) = & (y_1 - a \exp(-bx))^2 - \log[y_1 - a \exp(-bx)]^4 + \\ & + (y_2 - c \log(dx + 1))^2 - \log[(y_2 - c \log(dx + 1))^2] + \\ & + (y_3 - (x - e)^2)^2 + \lambda x. \end{aligned} \quad (19)$$

We can interpreted it as

$$V(x; \mathbf{g}) = \bar{V}_1(g_1(x)) + \bar{V}_2(g_2(x)) + \bar{V}_3(g_3(x)) + \bar{V}_4(g_4(x)) \quad (20)$$

where the marginal potentials are  $\tilde{V}_1(\vartheta) = \vartheta^2 - \log[\vartheta^4]$  with minimum at  $\mu_1 = \sqrt{2}$ ,  $\tilde{V}_2(\vartheta) = \vartheta^2 - \log[\vartheta^2]$  with minimum at  $\mu_2 = 1$ ,  $\tilde{V}_3(\vartheta) = \vartheta^2$  with minimum at  $\mu_3 = 0$  and  $\tilde{V}_4(\vartheta) = \lambda|\vartheta|$  with minimum at  $\mu_4 = 0$  (we recall that  $\lambda > 0$  and  $x \in \mathbb{R}^+$ ). Moreover, the vector of nonlinearities is

$$\mathbf{g}(x) = [g_1(x) = a \exp(-bx), g_2(x) = c \log(dx + 1), g_3(x) = (x - e)^2, g_4(x) = x]. \quad (21)$$

Since all marginal potentials are convex and all nonlinearities are convex or concave, we can apply the proposed adaptive RoU technique. Moreover, since  $\exp\{-\bar{V}_4(g_4(x))\} = \exp\{-\lambda x\}$  is easy to draw from, we can also applied the first technique described in Section III.

Note, however, that

- 1) the potential  $V(x; \mathbf{g})$  in Eq. (19) is not convex,
- 2) the target function  $p(x|\mathbf{y}) \propto \exp\{-V(x; \mathbf{g})\}$  can be multimodal,
- 3) the tails of the potential  $V(x; \mathbf{g})$  are not convex and

4) it is not possible to study analytically the first and the second derivatives of the potential  $V(x; \mathbf{g})$ . Therefore, the techniques in [11, 6, 13, 14] and the basic strategy in [22] cannot be used.

The procedure in Section III can be used because of the pdf  $q_4(x) \propto \exp\{-V_4(g_4(x))\} = \exp\{-x\}$  can be easily integrated and sampled, even if it is restricted in a interval  $\mathcal{I}_k$ . In this case, the reduced potential is  $V_{-4}(x; \mathbf{g}) = \bar{V}_1(g_1(x)) + \bar{V}_2(g_2(x)) + \bar{V}_3(g_3(x))$  and the corresponding function  $\exp\{-V_{-4}(x; \mathbf{g})\}$  is the likelihood, i.e.,  $p(\mathbf{y}|x) = \exp\{-V_{-4}(x; \mathbf{g})\}$ , since  $q_4(x)$  coincides with the prior pdf. Moreover, in order to use the RoU method we additionally need to study the potential functions  $V^{(1)}(x; \mathbf{g}) = \frac{1}{2}V(x; \mathbf{g})$  and  $V^{(2)}(x; \mathbf{g}) = \frac{1}{2}V(x; \mathbf{g}) - \log(x)$ . Since we assume  $x \geq 0$ , it is not necessary to study  $V^{(3)}$  as described in Section IV-B.

We set, e.g.,  $a = -2$ ,  $b = 1.1$ ,  $c = -0.8$ ,  $d = 1.5$ ,  $e = 2$ ,  $\lambda = 0.2$  and  $\mathbf{y} = [2.314, 1.6, 2]$ , and we start with the set of support points  $\mathcal{S}_0 = \{0, 2 - \sqrt{2}, 2, 2 + \sqrt{2}\}$  (details of the initialization step are explained in [22]). Figure 4 (a) shows the posterior density  $p(x|\mathbf{y})$  and the corresponding normalized histogram obtained by the first adaptive rejection sampling scheme. Figure 4 (b) depicts, jointly, the likelihood function  $p(\mathbf{y}|x) = \exp\{-V_{-4}(x; \mathbf{g})\}$  and its stepwise approximation  $\exp\{-\gamma_k\} \forall x \in \mathcal{I}_k, k = 0, \dots, m_t$ . The computation of the lower bounds  $\gamma_k \leq V_{-4}(x; \mathbf{g})$  has been explained in Appendix and Section III.

Figure 4 (c) depicts the set  $\mathcal{A}$  (solid) obtained by way of the RoU method, corresponding to the posterior density  $p(x|\mathbf{y})$  and the region  $\cup_{k=0}^{m_t} \mathcal{T}_k$  formed by triangular pieces, constructed as described in Section IV-A with  $m_t = 9$  support points.

The simulations shows that both method attain very similar acceptance rates<sup>2</sup>. In Figure 5 (a)-(b) are shown the curves of acceptance rates (averaged over 10,000 independent simulation runs) versus the first 1000 accepted samples using the technique described in Section III and the adaptive RoU algorithm explained in Section IV, respectively. More specifically, every point in these curves represent the probability of accepting the  $i$ -th sample. We can see that the methods are equal efficient, and the rates converge quickly close to 1.

### B. Stochastic volatility model

In this example, we study a stochastic volatility model where only the adaptive RoU scheme can be applied. Let be  $x_k \in \mathbb{R}^+$  the volatility of a financial time series at time  $k$ , and consider the following

<sup>2</sup>We define the acceptance rate  $R_i$  as the mean probability of accepting the  $i$ -th sample from an adaptive RS algorithm. In order to estimate it, we have run  $M = 10,000$  independent simulations and recorded the numbers  $\kappa_{i,j}$ ,  $j = 1, \dots, M$ , of candidate samples that were needed in order to accept the  $i$ -th sample. Then, the resulting empirical acceptance rate is  $\hat{R}_i = \sum_{j=1}^M \kappa_{i,j}^{-1} / M$ .



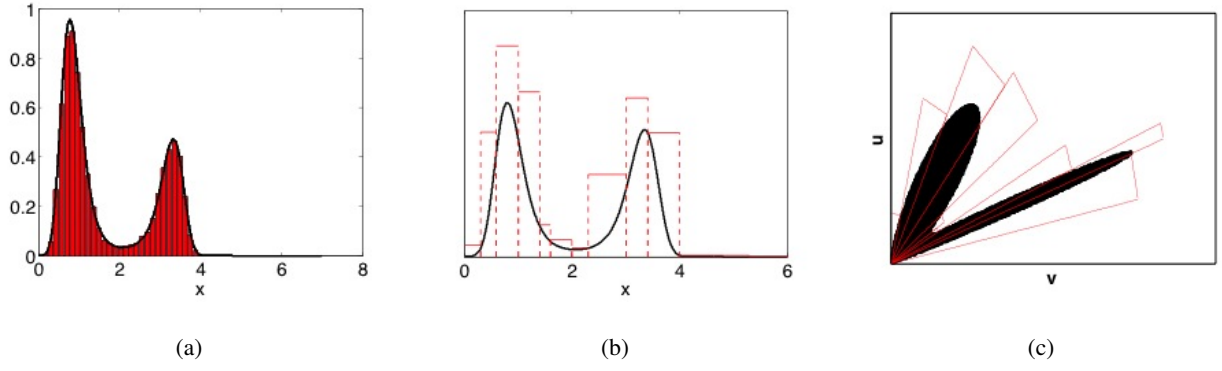


Fig. 4. **(a)** The target function  $p_o(x) = p(x|y)$  and the normalized histogram obtained by the adaptive RoU scheme. **(b)** The function  $\exp\{-V_{-4}(x; \mathbf{g})\}$  obtained using the reduced potential  $V_{-4}(x; \mathbf{g}) = V(x; \mathbf{g}) - \bar{V}_4(g_4(x))$  and its constant upper bounds  $\exp\{-\gamma_k\}$ ,  $k = 0, \dots, m_t$ . In this example, the function  $p(y|x) = \exp\{-V_{-4}(x; \mathbf{g})\}$  coincides with the likelihood. **(c)** The set  $\mathcal{A}$  corresponding to the target pdf  $p(x|y)$  using the RoU method and the region  $\cup_{k=0}^{m_t=9} \mathcal{T}_k$ , formed by triangles, constructed using the second adaptive rejection sampling scheme.

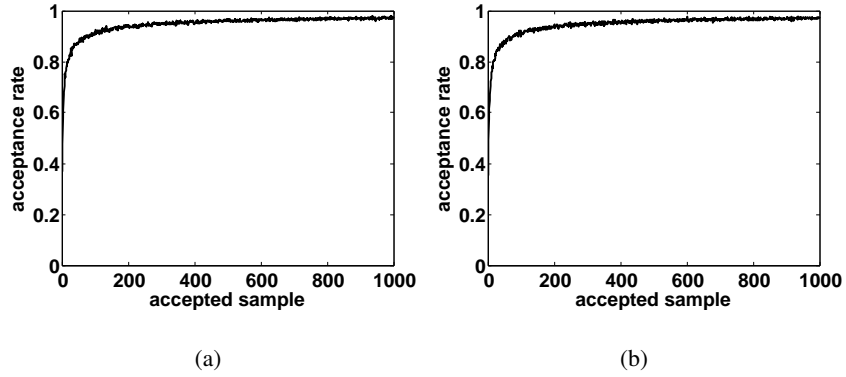


Fig. 5. **(a)** The curve of acceptance rates (averaged over 10,000 simulations) as a function of the first 1000 accepted samples using the first adaptive scheme described in Section III. **(b)** The curve of acceptance rates (averaged over 10,000 simulations) as a function of the first 1000 accepted samples using the proposed adaptive RoU technique.

state space system [5, Chapter 9]-[16] with scalar observation,  $y_k \in \mathbb{R}$ ,

$$\begin{cases} \log(x_k^2) = \beta \log(x_{k-1}^2) + \vartheta_{2,k}, \\ y_k = \log(x_k^2) + \vartheta_{1,k}, \end{cases} \quad (22)$$

where  $k \in \mathbb{N}$  denotes discrete time,  $\beta$  is a constant,  $\vartheta_{2,k} \sim N(\vartheta_{2,k}; 0, \sigma^2)$  is a Gaussian noise, i.e.,  $p(\vartheta_{2,k}) \propto \exp\{-\vartheta_{2,k}^2/2\sigma^2\}$ , while  $\vartheta_{1,k}$  has a density  $p(\vartheta_{1,k}) \propto \exp\{\vartheta_{1,k}/2 - \exp(\vartheta_{1,k})/2\}$  obtained from the transformation  $\vartheta_{1,k} = \log[\vartheta_{0,k}^2]$  of a standard Gaussian variable  $\vartheta_{0,k} \sim N(\vartheta_{0,k}; 0, 1)$ .

Given the system in Eq. (22), the transition (prior) pdf

$$p(x_k|x_{k-1}) \propto \exp \left\{ -\frac{\log(x_k^2) - \beta \log(x_{k-1}^2)}{2\sigma^2} \right\}, \quad (23)$$

and the likelihood function

$$\begin{aligned} p(y_k|x_k) &\propto \exp \left\{ \frac{y_k - \log(x_k^2)}{2} - \frac{\exp(y_k - \log(x_k^2))}{2} \right\} = \\ &= \exp \left\{ \frac{y_k - \log(x_k^2)}{2} - \frac{\exp(y_k) - 1/x_k^2}{2} \right\}, \end{aligned} \quad (24)$$

We can apply the proposed adaptive scheme to implement a particle filter. Specifically, let  $\{x_{k-1}^{(i)}\}_{i=1}^N$  be a collection of samples from  $p(x_{k-1}|y_{1:k-1})$ . We can approximate the predictive density as

$$\begin{aligned} p(x_k|y_{1:k-1}) &= \int p(x_k|x_{k-1})p(x_{k-1}|y_{1:k-1})dx_{k-1} \\ &\approx \frac{1}{N} \sum_{i=1}^N p(x_k|x_{k-1}^{(i)}) \end{aligned} \quad (25)$$

and then the filtering pdf as [4]

$$p(x_k|y_{1:k}) \propto p(y_k|x_k) \frac{1}{N} \sum_{i=1}^N p(x_k|x_{k-1}^{(i)}). \quad (26)$$

If we can draw exact samples  $\{x_k^{(i)}\}_{i=1}^N$  from (26) using a RS scheme, then the integrals of measurable functions  $I(f) = \int f(x_k)p(x_k|y_{1:k})dx_k$  w.r.t. to the filtering pdf can be approximated as  $I(f) \approx I_N(f) = \frac{1}{N} \sum_{i=1}^N f(x_k^{(i)})$ .

To show it, we first recall that the problem of drawing from the pdf in Eq. (26) can be reduced to generate an index  $j \in \{1, \dots, N\}$  with uniform probabilities and then draw a sample from the pdf

$$x_k^{(i)} \sim p_{o,j}(x_k) \propto p(y_k|x_k)p(x_k|x_{k-1}^{(j)}). \quad (27)$$

Note that, in order to take advantage of the adaptive nature of the proposed technique, one can draw first  $N$  indices  $j_1, \dots, j_r, \dots, j_N$ , with  $j_r \in \{1, \dots, N\}$ , and then we sample  $N_r$  particles from the same proposal,  $x_k^{(m)} \sim p_{o,j_r}(x_k)$ ,  $m = 1, \dots, N_r$ , where  $N_r$  is the number of times the index  $r \in \{1, \dots, N\}$  has been drawn. Obviously,  $N_1 + N_2 + \dots + N_N = N$ .

The potential function associated with the pdf in Eq. (27) is

$$V(x_k; \mathbf{g}) = -\frac{y_k - \log(x_k^2)}{2} + \frac{\exp\{y_k\}}{2} + \frac{1}{2x_k^2} + \frac{(\log(x_k^2) - \alpha_k)^2}{2\sigma^2} \quad (28)$$

where  $\alpha_k = \beta \log[(x_{k-1}^{(j)})^2]$  is a constant and  $\mathbf{g}(x_k) = [g_1(x_k) \triangleq \log(x_k^2), g_2(x_k) \triangleq -\log(x_k^2) + \alpha_k]$ .

The potential function in Eq. (28) can be expressed as

$$V(x_k; \mathbf{g}) = \bar{V}_1(g_1(x_k)) + \bar{V}_2(g_2(x_k)), \quad (29)$$

where the marginal potentials are  $\bar{V}_1(\vartheta_1) = \frac{1}{2}(-\vartheta_1 + \exp\{\vartheta_1\})$  and  $\bar{V}_2(\vartheta_2) = \frac{1}{2\sigma^2}\vartheta_2^2$ . Note that in this example:

- Since the potential  $V(x_k; \mathbf{g})$  is in general a non-convex function and the study of the first and the second derivatives is not analytically tractable, the methods in [6, 11] cannot be applied.
- Moreover, since the potential function  $V(x_k; \mathbf{g})$  has concave tails, as shown in Figure 6 (a), the method in [13] and the basic adaptive technique described in [22] cannot be used either.
- In addition, we also cannot apply the first adaptive procedure in Section III because there are no simple (direct) techniques to draw from (and to integrate) the function  $\exp\{-\bar{V}_1(g_1(x_k))\}$  or  $\exp\{-\bar{V}_2(g_2(x_k))\}$ .

But, since the marginal potentials  $\bar{V}_1$  and  $\bar{V}_2$  are convex and we also know the concavity of the nonlinearities  $g_1(x_k)$  and  $g_2(x_k)$ , we can apply the proposed adaptive RoU scheme. Therefore, the used procedure can be summarized in this way: at each time step  $k$ , we draw  $N$  time from the set of index  $r \in \{1, \dots, N\}$  with uniform probabilities and denote as  $N_r$  the number of repetition of index  $r$  ( $N_1 + \dots + N_N = N$ ). Then we apply the proposed adaptive RoU method, described in Section IV-A, to draw  $N_r$  particles from each pdf of the form of Eq. (27) with  $j = r$  and  $r = 1, \dots, N$ .

Setting the constant parameters as  $\beta = 0.8$  and  $\sigma = 0.9$ , we obtain an acceptance rate  $\approx 42\%$  (averaged over 40 time steps in 10,000 independent simulation runs). It is important to remark that it is not easy to implement a standard particle filter to make inference directly about  $x_k$  (not about  $\log(x_k)$ ), because it is not straightforward to draw from the prior density in Eq. (23). Indeed, there are no direct methods to sample from this prior pdf and, in general, we need to use a rejection sampling or a MCMC approach.

Figure 6 (b) depicts 40 time steps of a real trajectory (solid line) of the signal of interest  $x_k$  generated by the system in Eq. (22), with  $\beta = 0.8$  and  $\sigma = 0.9$ . In dashed line, we see the estimated trajectory obtained by the particle filter using the adaptive RoU scheme with  $N = 1000$  particles. The shadowed area illustrates the standard deviation of the estimation obtained by our filter. The mean square error (MSE) achieved in the estimation of  $x_k$  using  $N = 1000$  particles is 1.48.

The designed particle filter is specially advantageous w.r.t. to the bootstrap filter when there is a significant discrepancy between the likelihood and prior functions.

## VI. CONCLUSIONS

We have proposed two adaptive rejection sampling schemes that can be used to draw exactly from a large family of pdfs, not necessarily log-concave. Probability distributions of this family appear in many

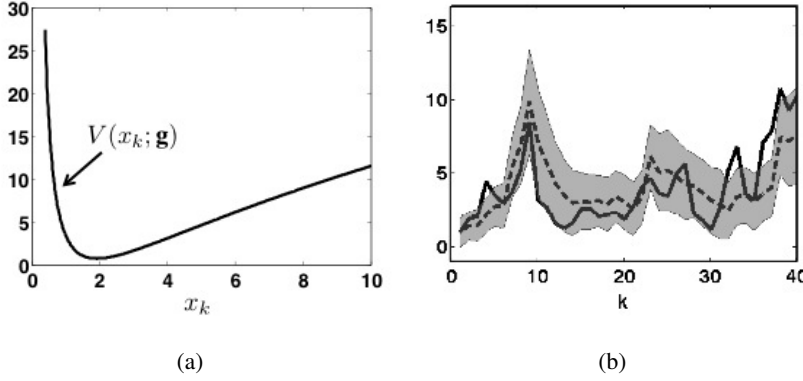


Fig. 6. **(a)** An example of potential function  $V(x_k; \mathbf{g})$  defined in Eq. (28), when  $\alpha_k = 1$ ,  $y_k = 2$  and  $\sigma = 0.8$ . The right tail is concave. **(b)** An example of trajectory with 40 time steps of the signal of interest  $x_k$ . The solid line illustrates the real trajectory generated by the system in Eq. (22) with  $\beta = 0.8$  and  $\sigma = 0.9$  while the dashed line shows the estimated trajectory attained by the particle filter with  $N = 1000$  particles. The shadowed area shows the standard deviation of the estimation.

inference problems as, for example, localization in sensor networks [18, 22, 23], stochastic volatility [5, Chapter 9], [16] or hierarchical models [12, Chapter 9], [7, 8, 9].

The new method yields a sequence of proposal pdfs that converge towards the target density and, as a consequence, can attain high acceptance rates. Moreover, it can also be applied when the tails of the target density are log-convex.

The new techniques are conceived to be used within more elaborate Monte Carlo methods. It enables, for instance, a systematic implementation of the accept/reject particle filter of [18]. There is another generic application in the implementation of the Gibbs sampler for systems in which the conditional densities are complicated. Since they can be applied to multimodal densities also when the tails are log-convex, these new techniques has an extended applicability compared to the related methods in literature [6, 11, 13, 20, 22], as shown in the two numerical examples.

The first adaptive approach in Section III is easier to implement. However, the proposed adaptive RoU technique is more general than the first scheme, as show in our application to a stochastic volatility model. Moreover using the adaptive RoU scheme to generate a candidate sample we only need to draw two uniform random variates but, in exchange, we have to find bound for three (similar) potential functions.

## VII. ACKNOWLEDGEMENTS

This work has been partially supported by the Ministry of Science and Innovation of Spain (project DEIPRO, ref. TEC-2009-14504-C02-01 and program Consolider-Ingenio 2010 CSD2008-00010

COMONSENS). This work was partially presented at the Valencia 9 meeting.

## VIII. APPENDIX

### A. Calculation of lower bounds

In some cases, it can be useful to find a lower bound  $\gamma_k \in \mathbb{R}$  such that

$$\gamma_k \leq \min_{x \in \mathcal{I}_k} V(x; \mathbf{g}), \quad (30)$$

in some interval  $\mathcal{I}_k = [s_k, s_{k+1}]$ .

If we are able to minimize analytically the modified potential  $V(x; \mathbf{r}_k)$ , we obtain

$$\gamma_k = \min_{x \in \mathcal{I}_k} V(x; \mathbf{r}_k) \leq \min_{x \in \mathcal{I}_k} V(x; \mathbf{g}). \quad (31)$$

If the analytical minimization of the modified potential  $V(x; \mathbf{r}_k)$  remains intractable, since the modified potential  $V(x; \mathbf{r}_k)$  is convex  $\forall x \in \mathcal{I}_k$ , we can use the tangent straight line  $w_k(x)$  to  $V(x; \mathbf{r}_k)$  at an arbitrary point  $x^* \in \mathcal{I}_k$  to attain a bound. Indeed, a lower bound  $\forall x \in \mathcal{I}_k = [s_k, s_{k+1}]$  can be defined as

$$\gamma_k \triangleq \min[w_k(s_k), w_k(s_{k+1})] \leq \min_{x \in \mathcal{I}_k} V(x; \mathbf{r}_k) \leq \min_{x \in \mathcal{I}_k} V(x; \mathbf{g}). \quad (32)$$

The two steps of the algorithm in Section II-B are described in the Figure 7 (a) shows an example of construction of the linear function  $w_k(x)$  in a generic interval  $\mathcal{I}_k \subset \mathcal{D}$ . Figure 7 (b) illustrates the construction of the piecewise linear function  $W_t(x)$  using three support points,  $m_t = 3$ . Function  $W_t(x)$  consists of segments of linear functions  $w_k(x)$ . Figure 8 (b) depicts this procedure to obtain a lower bound in an interval  $\mathcal{I}_k$  for a system potential  $V(x; \mathbf{g})$  (solid line) using a tangent line (dotted line) to the modified potential  $V(x; \mathbf{r}_k)$  (dashed line) at an arbitrary point  $x^* \in \mathcal{I}_k$ .

### B. Sampling uniformly in a triangular region

Consider a triangular set  $\mathcal{T}$  in the plane  $\mathbb{R}^2$  defined by the vertices  $\mathbf{v}_1$ ,  $\mathbf{v}_2$  and  $\mathbf{v}_3$ . We can draw uniformly from a triangular region [24], [3, p. 570] with the following steps:

- 1) Sample  $u_1 \sim \mathcal{U}([0, 1])$  and  $u_2 \sim \mathcal{U}([0, 1])$ .
- 2) The resulting sample is generated by

$$\mathbf{x}' = \mathbf{v}_1 \min[u_1, u_2] + \mathbf{v}_2(1 - \max[u_1, u_2]) + \mathbf{v}_3(\max[u_1, u_2] - \min[u_1, u_2]). \quad (33)$$

The samples  $\mathbf{x}'$  drawn with this convex linear combination are uniformly distributed within the triangle  $\mathcal{T}$  with vertices  $\mathbf{v}_1$ ,  $\mathbf{v}_2$  and  $\mathbf{v}_3$ .

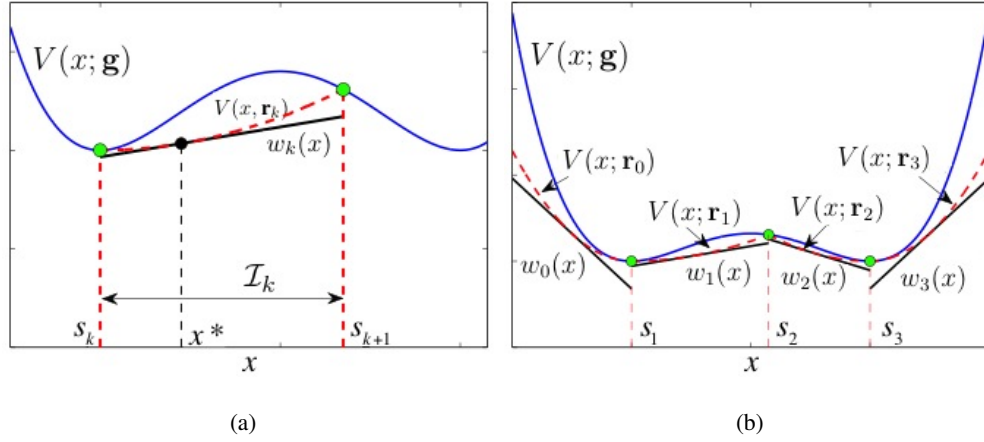


Fig. 7. **(a)** Example of construction of the linear function  $w_k(x)$  inside a generic interval  $\mathcal{I} = [s_k, s_{k+1}]$ . The picture shows a non-convex potential  $V(x; \mathbf{g})$  in solid line while the modified potential  $V(x; \mathbf{r}_k)$  is depicted in dashed line for  $\forall x \in \mathcal{I}_k$ . The linear function  $w_k(x)$  is tangent to  $V(x; \mathbf{r}_k)$  at an arbitrary point  $x^* \in \mathcal{I}_k$ . **(b)** Example of construction of the piecewise linear function  $W_t(x)$  with three support points  $\mathcal{S}_t = \{s_1, s_2, s_{m_t=3}\}$ . 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  $V(x; \mathbf{r}_k)$ .

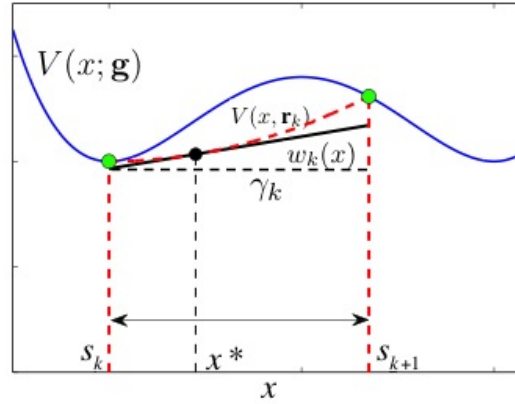


Fig. 8. The picture shows the potential  $V(x; \mathbf{g})$ , the modified potential  $V(x; \mathbf{r}_k)$  and the tangent line  $w_k(x)$  to the modified potential at an arbitrary point  $x^*$  in  $\mathcal{I}_k$ . The lower bound  $\gamma_k$  is obtained as  $\gamma_k = \min[w_k(s_k), w_k(s_{k+1})]$ . In this picture, we have  $\gamma_k = w_k(s_k)$ .

## REFERENCES

- [1] J. H. Ahrens. A one-table method for sampling from continuous and discrete distributions. *Computing*, 54(2):127–146, June 1995.
- [2] Y. Chung and S. Lee. The generalized ratio-of-uniform method. *Journal of Applied Mathematics*

- and Computing*, 4(2):409–415, August 1997.
- [3] L. Devroye. *Non-Uniform Random Variate Generation*. Springer, 1986.
  - [4] A. Doucet, N. de Freitas, and N. Gordon. An introduction to sequential Monte Carlo methods. In A. Doucet, N. de Freitas, and N. Gordon, editors, *Sequential Monte Carlo Methods in Practice*, chapter 1, pages 4–14. Springer, 2001.
  - [5] A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer, New York (USA), 2001.
  - [6] M. Evans and T. Swartz. Random variate generation using concavity properties of transformed densities. *Journal of Computational and Graphical Statistics*, 7(4):514–528, 1998.
  - [7] P. H. Everson. Exact bayesian inference for normal hierarchical models. *Journal of Statistical Computation and Simulation*, 68(3):223 – 241, February 2001.
  - [8] S. Frühwirth-Schnatter, R. Frühwirth, L. Held, and H. Rue. Improved auxiliary mixture sampling for hierarchical models of non-gaussian data. *Statistics and Computing*, 19(4):479–492, December 2009.
  - [9] D. Gamerman. Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing*, 7(1):57–68, 1997.
  - [10] W. R. Gilks, N. G. Best, and K. K. C. Tan. Adaptive Rejection Metropolis Sampling within Gibbs Sampling. *Applied Statistics*, 44(4):455–472, 1995.
  - [11] W. R. Gilks and P. Wild. Adaptive Rejection Sampling for Gibbs Sampling. *Applied Statistics*, 41(2):337–348, 1992.
  - [12] W.R. Gilks, S. Richardson, and D. Spiegelhalter. *Markov Chain Monte Carlo in Practice: Interdisciplinary Statistics*. Taylor & Francis, Inc., UK, 1995.
  - [13] Dilan Görür and Yee Whye Teh. Concave convex adaptive rejection sampling. *Journal of Computational and Graphical Statistics (to appear)*, 2010.
  - [14] W. Hörmann. A rejection technique for sampling from t-concave distributions. *ACM Transactions on Mathematical Software*, 21(2):182–193, 1995.
  - [15] W. Hörmann and G. Derflinger. The transformed rejection method for generating random variables, an alternative to the ratio of uniforms method. *Manuskript, Institut f. Statistik, Wirtschaftsuniversität*, 1994.
  - [16] E. Jacquier, N. G. Polson, and P. E. Rossi. Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics*, 12(4):371–389, October 1994.
  - [17] A. J. Kinderman and J. F. Monahan. Computer generation of random variables using the ratio of

- uniform deviates. *ACM Transactions on Mathematical Software*, 3(3):257–260, September 1977.
- [18] H. R. Künsch. Recursive Monte Carlo filters: Algorithms and theoretical bounds. *The Annals of Statistics*, 33(5):1983–2021, 2005.
  - [19] Josef Leydold. Automatic sampling with the ratio-of-uniforms method. *ACM Transactions on Mathematical Software*, 26(1):78–98, 2000.
  - [20] Josef Leydold. Short universal generators via generalized ratio-of-uniforms method. *Mathematics of Computation*, 72:1453–1471, 2003.
  - [21] G. Marsaglia. The exact-approximation method for generating random variables in a computer. *American Statistical Association*, 79(385):218–221, March 1984.
  - [22] L. Martino and J. Míguez. A generalization of the adaptive rejection sampling algorithm. *Statistics and Computing (to appear)*, July 2010.
  - [23] L. Martino and J. Míguez. Generalized rejection sampling schemes and applications in signal processing. *Signal Processing*, 90(11):2981–2995, November 2010.
  - [24] M. F. Keblis W. E. Stein. A new method to simulate the triangular distribution. *Mathematical and Computer Modelling*, 49(5):1143–1147, 2009.
  - [25] J. C. Wakefield, A. E. Gelfand, and A. F. M. Smith. Efficient generation of random variates via the ratio-of-uniforms method. *Statistics and Computing*, 1(2):129–133, August 1991.
  - [26] C.S. Wallace. Transformed rejection generators for gamma and normal pseudo-random variables. *Australian Computer Journal*, 8:103–105, 1976.