UNIVERSIDAD CARLOS III DE MADRID
Departamento de Teoría de la Señal y Comunicaciones

DOCTORAL THESIS

NOVEL SCHEMES FOR
ADAPTIVE REJECTION SAMPLING

Author: LUCA MARTINO
Supervised by: JOAQUÍN MÍGUEZ ARENAS
July 2011

i

Tesis Doctoral:	NOVEL SCHEMES FOR ADAPTIVE REJECTION SAMPLING
Autor:	Luca Martino
Director:	D. Joaquín Míguez Arenas
Fecha:	

Tribunal

Presidente:

Vocal 1:

Vocal 2:

Vocal 3:

Secretario:

Agradecimientos y reflexiones

" ...non al denaro, non all'amore né al cielo..."
"...in direzione ostinata e contraria..."
Fabrizio De André

A pesar de todo, y aunque tarde, he llegado. Un paso insignificante en sentido absoluto, pero muy importante para mí. Repito insignificante en sentido absoluto, basta tener un gran director de tesis para lograrlo. De hecho, me siento bastante *peor* persona (haciendo una integral sobre todos los aspectos de la vida, $x \in$ Vida, $\int \text{Luca}(x)dx$) respecto a hace cuatros años y medio.

El 80% de este trabajo hay que atribuirlo (en order transcendental) a Dios, a mi familia, a Vera y a Joaquín.

Sobre Dios me gustaría hablar mucho pero necesitaría otra tesis. De todas formas le agradeceré yo directamente, dado que, asumiendo su existencia, unicidad y omnipotencia en todo el conjunto continuo y compacto como es el universo, seguramente está dotado de tecnología inalámbrica (no creo, por ejemplo, que se necesite el trámite de un hombre disfrazado de payaso para lograr una comunicación fiable. Cualquier referencia a curas y Papas es pura coincidencia).

A mi familia (padres, hermano y hermana, tíos, tías, abuelos y abuelitas, etc.) debo un apoyo y una paciencia infinita. Por haberme permitido los estudios y haberme acompañado hasta aquí a pesar de mis numerosas "aventuras".

A Vera sencillamente le debo el estar aquí. Ha sido como un ángel. Ha tenido la tarea de un ángel. Además, creo que ha sufrido esta tesis más que yo.

Joaquín. Por dónde empezar... Bueno, con algo que lo resume todo: creo, y pienso de verdad, que cualquier ser viviente lograría escribir una tesis con Joaquín como director. Sí, sí, también un organismo unicelular. Esta tesis es fruto de su paciencia, de su persistencia, de su capacidad de saber escuchar y de su manera moderada de expresarse y pedir las cosas. Además ha sabido encontrar y dirigir unas ideas básicas hasta la formulación completa de esta tesis. Por ultimo, pero no menos importante, *"por terse comportado como amigo ademaís de como xefe"*. He disfrutado muchísimo trabajando, escribiendo los artículos y la tesis, y esto puede considerarse otro logro de Joaquín.

Además, en orden aleatorio coherentemente con este trabajo, tengo que agradecer a: Blanca, Fran, Sandra, Raul, Giorgio.es, Camilo, Isabel, Rafa, Pablo, Aurora, la hermana de Aurora, Wilton, Casquero, David Luengo, Angel Bravo, Jesse, Niamh, Cristina, Concha, Katrin, Jair, Eduardo-Novio-Kat, Edu-CaraFiga, Paloma, Miguel, Eugenia, Mais (o algo así)-*from* Nueva York, David-Marido-Eugenia, Hanna, Giorgio.mex, Jesus El Infiltrado, Bernd, Ricardo el Gallego, Fernandinho, Ramón, David-Novio-Cris, William, David Delgado, toda la gente del fútbol, los chicos de Addiopizzo (una asociación anti-mafia de Palermo), los chicos de residencia "La Cordata" de Milan (Beniamino, Michele Giove, Emilio, Emanuele, Antonio y locos que eran más normales que nosotros), mis amgios de Palermo (Piero, Frenquy, Manuel, Giuvá, Simulune, Zizzo etc.) todos mis amigos brasileños y el padre/exorcista La Grua *from* Palermo (querido padre, no ha funcionado, lo siento mucho).

Tengo que agradecer de forma especial a las siguientes personas:

1. Sandra, por todo que ha tenido que aguantar: bailes, ruidos, abre la ventana, cierra la ventana, gritos, y abre otra vez la ventana, y cierra la ventana, otros bailes, otras palmas y otras bromas, etc.

2. Fran, sobre todo porque es el mejor anti-depresivo que exista. Deberían venderlo en las farmacias.

3. Giorgio.es, porque ha intentado sinceramente cuidarme en custiones personales muy complejas.

4. Blanca, porque creo que me tiene bastante aprecio y no se el porqué.

5. Isabel, Jesse y Wilton por aguantar mis llamadas de todo tipo a todas horas.

6. Ramón, por haber compartido muchas horas con comida y cervezas por delante.

7. Todas las chicas por las "guarradas" que he dicho (y diré).

De Berlusconi no voy a hablar para no arruinar la tesis. Pero sí me gustaría terminar con unas frases que resumen perfectamente lo que pienso sobre las democracias representativas:
"La libertá non é star sopra un albero,

non é neanche avere una opinione,
la libertá non é uno spazio libero,
libertá é partecipazione.
La libertá non é star sopra un albero,
non é neanche il volo di un moscone,
la libertá non é uno spazio libero,
libertá é partecipazione." Giorgio Gaber (un gran italiano).

La libertad no es ir a votar un día y olvidarte al día siguiente, delegando tu futuro a un grupo de delincuentes, *la libertad es participación* (en la vida pública).

Abstract

We address the problem of generating random samples from a target probability distribution, with density function p_o, using accept/reject methods. An "accept/reject sampler" (or, simply, a "rejection sampler") is an algorithm that first draws a random variate from a proposal distribution with density π (where $\pi \neq p_o$, in general) and then performs a test to determine whether the variate can be accepted as a sample from the target distribution or not. If we apply the algorithm repeatedly until we accept N times, then we obtain a collection of N independent and identically distributed (i.i.d.) samples from the distribution with density p_o.

The goal of the present work is to devise and analyze *adaptive* rejection samplers that can be applied to generate i.i.d. random variates from the broadest possible class of probability distributions. Adaptive rejection sampling algorithms typically construct a sequence of proposal functions π_0, $\pi_1,...,\pi_t,...$ such that

(a) it is easy to draw i.i.d. samples from them and

(b) they converge, in some way, to the density p_o of the target probability distribution.

When surveying the literature, it is simple to identify several such methods but virtually all of them present severe limitations in the class of target densities, p_o, for which they can be applied. The "standard" adaptive rejection sampler by Gilks and Wild, for instance, only works when p_o is strictly log-concave.

Through Chapters 3, 4 and 5 we introduce a new methodology for adaptive rejection sampling that can be used with a broad family of target probability densities (including, e.g., multimodal functions) and subsumes Gilks and Wild's method as a particular case. We discuss several variations of the main algorithm that enable, e.g., sampling from some particularly "difficult" distributions (for instance, cases where p_o has log-convex tails and infinite support) or yield "automatic" software implementations using little analytical information about the target density p_o. Several numerical examples, including comparisons with some of the most relevant techniques in the literature, are also shown in Chapter 6.

x

Resumen

En este trabajo abordamos el problema de generar muestras aleatorias de una distribución de probabilidad objetivo, con densidad p_o, empleando métodos de aceptación/rechazo. En un algoritmo de aceptación/rechazo, primero se genera una realización aleatoria de una distribución tentativa con densidad π (donde $\pi \neq p_o$, en general) y a continuación se realiza un test para determinar si la muestra se puede aceptar como proveniente de la distribución objetivo o no. Si se aplica el algoritmo repetidamente hasta aceptar N veces, obtenemos una colección de N muestras independientes y identicamente distribuidas (i.i.d.) de la distribución con densidad p_o.

El objetivo del trabajo es proponer y analizar nuevos métodos de aceptación/rechazo *adaptativos* que pueden ser aplicados para generar muestras i.i.d. de la clase más amplia posible de distribuciones de probabilidad. Los algoritmos de aceptación/rechazo adaptativos suelen construir una secuencia de funciones tentativas π_0, π_1,...,π_t,... tales que

(a) sea fácil generar muestras i.i.d. a partir de ellas y

(b) converjan, de manera adecuada, hacia la densidad p_o de la distribución objetivo.

Al revisar la literatura, es sencillo identificar varios métodos de este tipo pero todos ellos presentan limitaciones importantes en cuanto a las clases de densidades objetivo a las que se pueden aplicar.

El método original de Gilks y Wild, por ejemplo, sólo funciona si p_o es estrictamente log-concava. En los Capítulos 3, 4 y 5 presentamos una nueva metodología para muestreo adaptativo por aceptación/rechazo que se puede utilizar con una amplia clase de densidades de probabilidad objetivo (incluyendo, por ejemplo, funciones multimodales) y comprende al método de Gilks y Wild como un caso particular. Se discuten diversas variaciones del algoritmo principal que facilitan, por ejemplo, el muestreo de algunas distribuciones particularmente "díficiles" (e.g., casos en los que p_o tiene colas log-convexas y con soporte infinito) o una implementación *software* prácticamente "automática", en el sentido de que necesitamos poca información analítica acerca de la función p_o.

En el Capítulo 6 mostramos varios ejemplos numéricos, incluyendo comparaciones con algunas de las técnicas más relevantes que se pueden encontrar en la literatura.

Contents

List of Tables

Chapter 1

Introduction

"Anyone who considers arithmetic methods of producing random digits is, of course, in a state of sin." J. von Neumann (1940). This sentence means that there are no true random "numbers", just means to produce them, and "a strict arithmetic procedure is not such a method".

1.1   Monte Carlo: a brief history

Monte Carlo methods have been used since, at least, the eighteenth century but only in the past few decades has the technique gained the status of a full-fledged numerical methodology capable of addressing the most complex applications. The name "Monte Carlo" was used as a code name of a secret project of the U. S. Defense Department. This project was headed by John von Neumann (1903-1957, Hungarian-American mathematician) and Stanislas Ulam (1909-1984, Polish mathematician) at the Los Alamos Scientific Laboratory. The project used random numbers to simulate complicated sequences of connected events. The name comes from the city in the Monaco principality, because of the roulette (the capital of Monaco was a center for gambling), a simple random number generator. A gossip about the name reports that Ulam's uncle would borrow money from the family by saying that "I just have to go to Monte Carlo". The name seems to have held on.

The systematic development of Monte Carlo methods can be dated from about 1940. Indeed, in the 1940's, a formal foundation for the methodology was developed by von Neumann, who established the mathematical basis for

pseudo-random number generators. The work was done in collaboration with Stanislaw Ulam, who realized the importance of the digital computer in the implementation of the approach.

However, there is a number of isolated and undeveloped instances on much earlier occasions. For example, in the second half of the nineteenth century a number of people performed experiments [57] in which they threw a needle in a haphazard manner (i.e., like "random") onto a board ruled with parallel straight lines and inferred the value of $\pi \approx 3.14$ from observations of the number of intersections between the needle and the lines. This problem was first stated in 1777 by Georges-Louis Leclerc, comte de Buffon.

In 1899, Lord Rayleigh showed that a one-dimensional random walk without absorbing barriers could provide an approximate solution to a parabolic differential equation. In 1930, Fermi used sampling methods to estimate quantities involved in controlled fission. In 1931, Kolmogorov showed the relationship between Markov stochastic processes and certain integro-differential equations.

But the real use of Monte Carlo methods as a research tool originates from work on the atomic bomb during the World War II. This work involved a direct simulation of the probabilistic problems concerned with random neutron diffusion in fissile material. Later on, about 1948, Fermi, Metropolis and Ulam obtained Monte Carlo estimates for the eigenvalues of the Schrodinger equation.

Although the method had already been used for secret projects, the "official history" of Monte Carlo methods began in 1949 with the publication of a paper by Metropolis and Ulam [119]. A team headed by Nicolas Metropolis (1915-1999) carried out the first actual Monte Carlo calculations on the ENIAC computer (the world's first electronic digital computer, built at the University of Pennsylvania) in 1948. The Metropolis algorithm, first described in a 1953 paper by N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller and E. Teller [118], was cited in Computing in Science and Engineering as being among the top 10 algorithms having the "greatest influence on the development and practice of science and engineering in the 20th century".

About 1970, the newly developing theory of computational complexity began to provide a more precise and persuasive rationale for employing the Monte Carlo method.

The main requirement to use the Monte Carlo method for simulation of a physical system is that it must be possible to describe the system in terms of

a probability density function (or a cumulative distribution function). Once the density function of a system is known, then the simulation begins to generate random numbers from this density. There must be a rule available, based on some reasonable mathematical and/or physical theory, to decide the outcome of such a trial. Many trials are conducted and outcomes of all of these trials are recorded. The final step in the Monte Carlo method is that the behavior of the overall system is obtained by computing the average of the outcomes of the trials.

The spirit of Monte Carlo can be summarized by the following example (by von Neumann): consider a spherical core of fissile material surrounded by a shell of tamper material. Assume some initial distribution of neutrons in space and in velocity. The underlying idea is to follow the development of a large number of individual neutron chains as a consequence of scattering, absorption, fission and escape. At each stage, a sequence of "decisions" has to be made based on statistical probabilities appropriate to the physical and geometric factors. The various decisions are made using pseudo-random numbers with desired properties.

Table 1.1 summarizes the most important steps in the development of Monte Carlo approach.

Because of the great potential of this methodology, various techniques are still being developed by researchers. Recent advances of Monte Carlo algorithms include the cluster method, data augmentation, rejection control [97, 99], umbrella sampling [158], density-scaling Monte Carlo [151], multigrid Monte Carlo [52], hybrid Monte Carlo [35], simulated annealing [81, 102], simulated tempering [103], parallel tempering [46, 143], multiple try Metropolis [100, 97, 95], adaptive and sequential MCMC [5, 49], adaptive importance sampling [132], sequential Monte Carlo [34, 32, 98], particle filtering [7, 13, 39, 53, 86, 133] etc...

There is also a trend in studying and analyzing *population-based* methods [19, 49, 71, 97]. The underlying idea is to generate a collection of random variables in parallel and then incorporate an additional step of information exchange. For instance, population-based Markov chain Monte Carlo operates by embedding the target into a sequence of related probability measures and simulating N parallel chains (the population). In addition, the chains are allowed to interact via various crossover moves.

In some cases, many algorithms are related, and sometimes they are even identical. For instance, the configuration bias Monte Carlo [137, 138] is equivalent to a sequential importance sampling combined with

Table 1.1: Most important stages in Monte Carlo research.

1949: paper of N. Metropolis and S. Ulam [119].

1951: J. von Neumann: Rejection sampling [152].

1953: N. Metropolis: Metropolis algorithm [118].

1954: J. Harmersly, K. Morton: sequential sampling [58].

1955: A. Rosenbluth, M. Rosenbluth: sequential sampling [135].

1956: A. Marshall: Importance sampling [108].

1970: Hasting: Metropolis-Hasting algorithm [59].

1987: Rubin: Sampling importance resampling [136].

the Metropolis-Hasting algorithm with independent proposal (transition) density. The multiple try Metropolis [100, 97, 95] is an extension of a technique described in [41]. The exchange Monte Carlo [67] recalls the parallel tempering [46, 143]. Sequential Monte Carlo and particle filtering are often used a synonimus.

1.2 The need of Monte Carlo

The range of application of Monte Carlo algorithms is enormous from statistical physics problems [135, 137, 138] (e.g. the simulation of galactic formation [128]) to nuclear medicine applications [101] (to predict the exact path of photons, electrons or α-particles that traverse different regions of the body and to identify the correct activity [139]). Other examples of applications are in finance [70], in genetics [56], state space models (epidemiology [156] and meteorology [160]), time series analysis [9], mixture models for inference [40], operations research [82] (traffic control, quality

control, production optimization). Monte Carlo methods are employed in many other areas of engineering too. For example, they have been used to simulate the turbulent combustion of a diesel engine spray injection [129] or to track moving targets and estimate their positions [7].

Monte Carlo techniques have been applied to all of these problems in order to calculate complicated integrals, to simulate a complex phenomenon or to reduce the amount of computation. The resulting algorithms are often concurrent and well suited to implementation on parallel computers.

1.2.1   Numerical integration

The best known (and maybe the most important) applications of Monte Carlo techniques involve the approximation of complicated integrals. Given an m-dimensional variable $\mathbf{x} \in \mathbb{R}^m$, a crucial part of many scientific problems is the computation of integrals of the form

$$I = \int_{\mathcal{D}} f(\mathbf{x})d\mathbf{x}, \tag{1.1}$$

where $\mathcal{D} \subseteq \mathbb{R}^m$. Let us assume that we are able to generate N random samples $\mathbf{x}^{(1)}, ..., \mathbf{x}^{(N)}$ uniformly in $\mathcal{D}$. Then we can approximate the integral I as

$$\hat{I}_N = \frac{1}{N} \left(f(\mathbf{x}^{(1)}) + ... + f(\mathbf{x}^{(N)}) \right). \tag{1.2}$$

Furthermore, given a collection of independent random variables $X_1, ..., X_N$ with common mean μ and finite variances, the *strong law of large numbers* [28, 76, 157] states that

$$\frac{X_1 + ... + X_N}{N} \to \mu,$$

when $N \to +\infty$, and based on this result we can claim that

$$\lim_{N \to +\infty} \hat{I}_N \to I, \tag{1.3}$$

with probability 1. This is the basic formulation of a Monte Carlo technique. We can provide further theoretical ground for the methodology by invoking the *central limit theorem* [28, 76, 157] that yields

$$\sqrt{N}(\hat{I}_N - I) \to \mathcal{N}(0, \sigma^2)^1 \tag{1.4}$$

---

[1] $\mathcal{N}(0, \sigma^2)$ denotes a Gaussian distribution with 0 mean and σ^2 variance.

in distribution, as $N \to +\infty$, where

$$\sigma^2 = \text{var}[f(\mathbf{x})].$$

It is important to remark that this variance σ^2 measures how the function $f(\mathbf{x})$ is distributed over $\mathcal{D}$ (if $f(\mathbf{x})$ is constant in $\mathcal{D}$ the variance is zero, i.e., $\sigma^2 = 0$).

Moreover, given Eq. (1.4), we can state that the Monte Carlo approximation error $(\hat{I}_N - I) \to 0$ decays to zero with $1/\sqrt{N}$, in spite of the dimensionality of $\mathbf{x} \in \mathbb{R}^m$ (i.e., *independently* of m). This is probably the main advantage of the Monte Carlo techniques when we compare with their deterministic counterparts (see, e.g., [18, 91]).

To see how the Monte Carlo methodology works with multidimensional problems, let first consider the simplest case $m = 1$. In this situation, we can carry out deterministic approximation of I, such as the *Riemann approximation* [91], obtaining an error rate that decays as $1/N$, better than the Monte Carlo method. Moreover, using more sophisticated deterministic algorithms, as the *Simpson's rule* or the *Newton-Cote's rule* [91], we can improve the approximation. However, these deterministic methods are computationally expensive when the dimension m increases. For instance, for $m = 20$ we need to evaluate N^{20} grid points in the Riemann approximation to obtain an accuracy $O(1/N)$. On the contrary, theoretically, drawing N points $\mathbf{x}^{(1)}, ..., \mathbf{x}^{(N)}$ uniformly in $\mathcal{D}$, the Monte Carlo scheme has an accuracy of $O(1/\sqrt{N})$ regardless of the dimension of $\mathcal{D}$. From this point of view, many researches have argued that the Monte Carlo approach breaks the "curse of dimensionality" [97, 134].

In practice, however, there are also some relevant drawbacks. For example, it is important to remark that the errors of deterministic algorithms (trapezoidal rule, Simpson's rule, Newton-Cote's rule, etc...) are themselves deterministic whereas the Monte Carlo error is a variance bound. Moreover, when the dimensionality increases, the "crude" Monte Carlo technique has further problems, e.g.,

1. the variance σ^2 can be very large,

2. the accuracy $O\left(1/\sqrt{N}\right)$ is only a probabilistic bound,

3. the regularity of the integrand $f(\mathbf{x})$ is not exploited,

4. and we may not be able to draw uniformly from $\mathcal{D}$.

The second point means there is no guarantee that the expected accuracy is achieved in a specific calculation. Moreover, the third point remarks that the probabilistic bound $O\left(1/\sqrt{N}\right)$ is obtained under very weak regularity conditions (i.e., for any square-integrable integrand $f(\mathbf{x})$), but we do not make any improvements from any additional regularity properties of $f(\mathbf{x})$.

In order to overcome some of these difficulties, the concept of *importance sampling* [108, 34, 32, 134] has been introduced in the literature. This approach consists in drawing N samples $\mathbf{x}^{(1)}, ..., \mathbf{x}^{(N)}$ from a nonuniform density function $\pi(\mathbf{x})$ that puts more probability mass on the "important" parts of the region $\mathcal{D}$ (in oder to save the computational resources). In this case, the approximation of I is

$$\hat{I}_N = \frac{1}{N}\left(\frac{f(\mathbf{x}^{(1)})}{\pi(\mathbf{x}^{(1)})} + ... + \frac{f(\mathbf{x}^{(N)})}{\pi(\mathbf{x}^{(N)})}\right). \tag{1.5}$$

This estimator is unbiased and it has variance $\sigma^2 = \text{var}_\pi[f(\mathbf{x})/\pi(\mathbf{x})]$. In a fortunate case, if f is non-negative and I is finite, we may select $\pi(\mathbf{x}) \propto f(\mathbf{x})$ so that $\hat{I}_N = I$. However, in most practical problems we can only try to find a "good" proposal density $\pi(\mathbf{x})$ that is reasonably "close" in shape to $g(\mathbf{x})$. Therefore, the challenge is to draw random samples from a good proposal density π.

If we know the analytic form of the proposal density $\pi(\mathbf{x})$ only up to a multiplicative constant (i.e., we know an unnormalized function $\pi_u(\mathbf{x}) \propto \pi(\mathbf{x})$), we can construct a *biased* estimator

$$\hat{I}_N = \frac{1}{w^{(1)} + ... + w^{(N)}}\left(w^{(1)}f(\mathbf{x}^{(1)}) + ... + w^{(N)}f(\mathbf{x}^{(N)})\right), \tag{1.6}$$

where $w^{(i)} = 1/\pi_u(\mathbf{x}^{(i)})$, $i = 1, ..., N$. With this approach, we can effectively use an unnormalized function $\pi_u(\mathbf{x})$ and, moreover, $\hat{I}_N$ often has a smaller mean square error than the unbiased estimator of Eq. (1.5). However, we remark that the choice of π (or π_u) is crucial to the performance of both estimators, i.e., a good choice of the proposal can reduce drastically the variance of the estimate.

Another class of algorithms introduced to overcome the drawbacks of the Monte Carlo methodology are the so-called *quasi-Monte Carlo methods* [124, 155]. The basic idea of a quasi-Monte Carlo technique is to replace the random samples in a Monte Carlo method by *well-chosen* deterministic points, often termed *nodes*. These nodes are chosen judiciously in order to

guarantee a small error in the numerical approximation of the integral I. The selection criterion is based on the concepts of *uniformly distributed sequence* and *discrepancy* [124]. The discrepancy is a measure of the deviation from the uniform distribution. For a suitable choice of N nodes, quasi-Monte Carlo methods can obtain a deterministic error bound of $O(N^{-1}\log(N)^{m-1})$, where m is the dimension of the space.

1.3   Random number generation

Random number generation is the core issue of Monte Carlo simulations. A reliable random number generator is critical for the success of a Monte Carlo method. Indeed, the success of Monte Carlo calculations depends on the pertinence and suitability of the underlying stochastic model, but also on how well the random numbers simulate the random variables in the model.

Random numbers are needed in a variety of areas. We have emphasized their applications in Monte Carlo methods, but they also play a crucial role in simulation problems, computational statistics, VLSI testing, cryptography and computer games.

Also many problems in bioinformatics, computational chemistry, physics, biology, statistics, etc... require sampling from a nonuniform distribution $p_o(\mathbf{x})$. In these scientific applications, $p_o(\mathbf{x})$ represents the probability density of a complex system, where $\mathbf{x}$ denotes the (random) configuration of system. For instance, in an analysis of a macromolecule, $\mathbf{x}$ represents the three-dimensional coordinates of all atoms in a molecule. The target density is defined as

$$p_o(\mathbf{x}) = z(T)\exp\{-V(\mathbf{x})/kT\},$$

and corresponds to the so-called Boltzmann distribution [51, 95, 97], where k is the Boltzmann constant, T is the system temperature, $V(\mathbf{x})$ is the energy function and $z(T)$ is the *partition function*, which is difficult to calculate in general. In Bayesian statistical inference, $\mathbf{x}$ usually represents the missing data jointly with the unknown parameter values and $p_o(\mathbf{x})$ often denotes the joint posterior distribution of these variables.

There is another (broad) class of applications where we come across optimization problems [6, 102] that can be conveniently tackled by generating random samples. Indeed, let us consider the problem minimizing a cost function $V(\mathbf{x})$. This is equivalent to maximizing another function $p(x) = \exp\{-V(\mathbf{x})/T\}$, with $T > 0$. If $p(x)$ is integrable for all $T > 0$, we can define

the density

$$p_o(x) \propto p(x) = \exp\{-V(\mathbf{x})/T\},$$

and if we are able to draw from $p_o(x)$ when T is small enough, the generated samples are located (with high probability) in a region close to the global minimum of $V(\mathbf{x})$.

1.3.1 Random, pseudo-random, quasi-random

There are several types of random numbers:

- "Real" random numbers are generated using a physical device as a source of randomness. Examples found in the literature include coin flipping, roulette wheels, white noise and count of emitted particles.

- Pseudo-random numbers: we produce a deterministic sequence that passes tests of randomness. The standard procedures for generating sequences of pseudo-random numbers are based on recursive methods and yield sequences that can be periodic (with a very large period) [44, Chapter 1], [124, Chapter 7] or *chaotic* [4, 16].

- Quasi-random numbers: a deterministic number sequence that presents a low discrepancy with respect to a given distribution. Deterministic constructions have been studied to build low-discrepancy point sets and sequences (using, for example, digit and fractional expansions) [124, Chapter 3].

The general requirements that are usually placed on random number generation methods can be divided into four categories.

1. *Computational requirements*: they refer to the computational cost and the resources needed to generate the random numbers.

2. *Structural requirements*: they include, e.g., the period length (in period sequences of pseudo-random numbers) and the lattice structure.

3. *Statistical requirements*: the produced numbers are expected to pass statistical tests related essentially to their distribution and statistical independence properties.

9

4. *Complexity - theoretic requirements*: some definitions of "randomness" for a finite string of random digits have been presented in the literature [22, 23, 83, 84, 109]. These represent a collection of conditions that a finite sequence has to satisfy in order to be considered "random".

In computational statistics, random variate generation is usually divided into two steps:

(1) generating "imitations" of independent and identically distributed (i.i.d.) random numbers having a uniform distribution and

(2) applying some transformation and/or selection techniques such that these i.i.d. uniform samples are converted into variates from the target probability distribution.

These two steps are essentially independent. The expression *pseudo-random number generator* usually refers to an algorithm used for the first step while the term *sampling method* is usually associated to an algorithm used in the second step. In particular, a sampling technique assumes that some random number generator with known distribution (typically uniform) is available. Random sampling algorithms are also termed *non-uniform random variate generators*.

1.3.2 Random sampling methods

Non-uniform random numbers are also referred to simply as *random variates*. Non-uniform random variate generation is a field in the crossroad of mathematics, statistics and computer science. It is often considered a subarea of statistical computing and simulation methodology [51, 44, 64, 95, 97, 149].

The main bibliography landmark is the book by Devroye [30], nearly an encyclopedia of random variate generation techniques. However, most books on Monte Carlo methods include at least one chapter on random variate generation (see, e.g., [18, 134]). Sampling techniques can be classified in three large categories:

- *Direct methods*: these techniques use an adequate transformation to convert the samples provided by an available random source into samples with desired statistical properties [64, Chapter 2]. In general, the generation is fast and samples are independent. However, in many practical situations a suitable transformation is unknown.

- *Accept/reject methods*: given samples from an available random source, these algorithms accept or discard the samples by an adequate test. The accepted samples are independent with the desired distribution. The main drawback is that the acceptance rate can be very low and the computational cost very high. Many adaptive schemes have been proposed in order to improve the probability of accepting candidate samples [50], [64, Chapter 4].

- *Markov Chain Monte Carlo (MCMC)*: these techniques construct a Markov chain that converges to a prescribed stationary distribution. The major advantage of this methodology is that it can be applied almost universally. The main problem, on the other hand, is that MCMC algorithms produce sequences of highly correlated variates. Hence, the estimates resulting from these samples tend to have greater variance than those resulting from independent samples.

In Chapter 2, we describe, in detail, the first two classes. Many books about Monte Carlo methods add another class: the *importance sampling* techniques. Here, we do not include it because importance sampling is not exactly a random variate generation method. The importance procedure approximates a density measure with weighted samples. It is very useful for numerical integration (it is crucial in particle filtering, for example), but it does not generate random numbers. It produces an approximation of a probability measure using samples from a different probability distribution.

In the literature, the terms *universal*, *automatic* and *black-box* are usually associated with generators, almost equivalently. They mean that the generator can draw samples from a large family of distributions (if it is universal, virtually *all* the probability densities) and can be coded in a computer program that only needs to evaluate the target density.

Combination of methods

Many mixed strategies have been proposed that combine methods from different categories [17, 20, 25, 123, 97]. For instance, many authors have tried to integrate the MCMC techniques in particle filtering [10, 51, 72]. One example is the resample-move algorithm [11], which combines sequential importance resampling (SIR) with MCMC sampling.

Since this thesis is focused on rejection (accept/reject) sampling schemes, a most interesting mix of different classes is provided by algorithms that

combine rejection and importance sampling. Some examples are the so-called *weighted rejection sampling* and *rejection control* algorithms [17, 21, 99, 97]. Other methods combine rejection sampling with sequential importance sampling [99], as partial rejection control [97] and rejection particle filters [12, 68, 90, 144, 145]. All these techniques are accept/reject methods that re-incorporate (in some way) the discarded samples in the computed estimators. For a detailed comparison of the performance of rejection and importance sampling estimators see [20, 24, 96].

Finally, a deterministic version of rejection sampling has been investigated in the quasi-Monte Carlo framework [155].

1.4   Goal and organization

The goal of this thesis is to devise and analyze adaptive rejection sampling methods that can be applied to generate i.i.d. random variates from the broadest possible class of probability distributions.

Adaptive rejection sampling algorithms typically construct a sequence of proposal functions π_0, π_1,...,π_t,... such that

(a) it is easy to draw i.i.d. samples from them and

(b) converge, in some way, to the density p_o of the target probability distribution.

When surveying the literature, it is simple to identify several such algorithms [50, 48, 55, 36, 61, 65, 94] but virtually all of them present severe limitations in the class of target densities, p_o, for which they can be applied. The "standard" adaptive rejection sampler, for instance, only works when p_o is strictly log-concave. In this work, we try to make progress toward a "universal" adaptive rejection sampler.

The remaining of this thesis is organized as follows. In Chapter 2, we describe the main, and most popular, techniques to produce i.i.d. random samples. Using both the direct and accept/reject methods. Besides reviewing the state of art we also pinpoint some relationships among different techniques that had not been identified in the literature.

Chapter 3 is devoted to a novel adaptive rejection sampling scheme that can be used to draw exactly from a certain family of densities, not necessarily log-concave and possibly multimodal. The new method is a generalization

of the classical adaptive rejection sampling scheme of [50], and includes it as a particular case. The proposed algorithm constructs a sequence of proposal probability density functions (pdf's) that converge towards the target density and, therefore, can attain very high acceptance rates. This technique breaks down when the target pdf has log-convex tails.

For this reason, in Chapter 4 we propose two adaptive rejection sampling schemes that can be used to draw exactly from densities that have not-necessarily log-concave tails. It should be noticed that many (virtually *all*) strategies presented in the literature [36, 50, 55, 93] break down when the tails of the target pdf are log-convex in an infinite domain (or T-convex, in general).

In Chapter 5 we introduce several extensions of the basic technique presented in Chapter 3. We also provide an interpretation of the algorithm of Chapter 3 as a method to draw samples from a target density generated by a transformation of scale of another distribution [74]. We discuss how it is possible to extend the class of target densities that the methods of Chapter 3 and 4 can address and, finally, introduce an alternative "automatic" algorithm that demands only the ability to evaluate the target density.

In Chapter 6 we present a number of examples that illustrate the range of applicability of the adaptive rejection schemes introduced in this work. They include mostly synthetic examples but also an application to positioning using real data from a wireless sensor network.

Finally, some conclusions and a discussion about future research directions are presented in Chapter 7.

Chapter 2

Background

In this chapter we present some background material needed in the rest of this work. In particular, we describe a collection of methods used for random sampling. All of them consider available a random source with known distribution. Moreover, all of them are aimed to produce independent and identically distributed (i.i.d.) samples with the exception of the adaptive rejection Metropolis sampling (ARMS) method that provides correlated samples. The definitions and notations are as close as possible to the original materials but also connected to the structure and notation used in the next chapters. We also present and analyze some relevant connections and relationships among different algorithms that we have not found, so far, in the literature.

The chapter is organized as follows. Section 2.1 contains a summary of notations. Section 2.2 describes sampling techniques based on transformations and Section 2.3 explains how to draw from a discrete mixture of densities. The fundamental theorem of simulation is introduced in Section 2.4, while Section 2.5 is devoted to the rejection sampling techniques. The adaptive rejection sampling algorithm and its generalizations are presented in Section 2.6 and 2.7, respectively. Finally, we introduce the ratio of uniforms technique in Section 2.8.

2.1 Notation

2.1.1 Vectors, points and intervals

Scalar magnitudes are denoted using regular face letters, e.g., x, X, while vectors are displayed as bold-face letters, e.g., $\mathbf{x}$, $\mathbf{X}$. The scalar coordinates of a vector in n-dimensional space are denoted with square brackets, e.g., $\mathbf{x} = [x_1, ..., x_n]$. Often, it is more convenient to interpret $\mathbf{x}$ as a point in the space. When needed, we emphasize this representation with the alternative notation $\mathbf{x} = (x_1, ..., x_n)$.

We use a similar notation for the intervals in the real line. Specifically, for two boundary values $a \leq b$, we denote $[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$ for a closed interval, while

$$(a, b] = \{x \in \mathbb{R} : a < x \leq b\}, \quad [a, b) = \{x \in \mathbb{R} : a \leq x < b\},$$

are half-open intervals and finally $(a, b) = \{x \in \mathbb{R} : a < x < b\}$ is an open interval.

2.1.2 Random variables, distributions and densities

We indicate random variables (r.v.) with upper-case letters, e.g., X, $\mathbf{X}$, while we use lower-case letters to denote the corresponding realizations, e.g., x, $\mathbf{x}$. Often, when we draw a collection of samples of a r.v., we use the superscript notation $x^{(i)}$, $\mathbf{x}^{(i)}$ where i indicates the sample number.

We use lower-case letters, e.g., $q(\cdot)$, to denote the probability density function (pdf) of a random variable or vector, e.g., $q(y)$ is the pdf of Y. The conditional pdf of X given $Y = y$ is written $p(x|y)$. The cumulative distribution function (cdf) of a r.v. X is written as $F_X(\cdot)$. The probability of an event, e.g., $X \leq x$, is indicated as $\text{Prob}\{X \leq x\}$. In particular, $F_X(a) = \text{Prob}\{X \leq a\}$.

The target pdf from which we wish to draw samples is denoted as $p_o(x)$ while $p(x)$ is a function proportional to $p_o(x)$, i.e., $p(x) \propto p_o(x)$.

The uniform distribution in an interval $[a, b]$ is written $\mathcal{U}([a, b])$. The Gaussian distribution with mean μ and variance σ^2 is denoted $\mathcal{N}(\mu, \sigma^2)$. The symbol $\sim$ means that a r.v. X or a sample x' have the indicated distribution, e.g., $X \sim \mathcal{U}([a, b])$ or $x' \sim \mathcal{N}(\mu, \sigma^2)$. The expression $X \stackrel{d}{=} Z$ denotes that the two r.v.'s X and Z are "equal in distribution", i.e., they have the same cdf. Finally, $\mathcal{N}(x; \mu, \sigma^2)$ represents a Gaussian pdf with mean μ and variance σ^2.

2.1.3 Sets

Sets are denoted with calligraphic upper-case letters, e.g., $\mathcal{R}$. The support of the r.v. of interest X is denoted as $\mathcal{D} \subseteq \mathbb{R}$ (i.e., $\mathcal{D}$ is the domain of the target pdf $p_o(x)$). In some cases, without loss of generality, we may consider $\mathcal{D} = \mathbb{R}$ for convenience. When needed, we denote with $\mathcal{C}$ the support of auxiliary variables.

Finally, we write the indicator function on the set $\mathcal{S}$ as $\mathbb{I}_{\mathcal{S}}(x)$. It takes value 1 if $x \in \mathcal{S}$ and 0 otherwise, i.e.,

$$\mathbb{I}_{\mathcal{S}}(x) = \begin{cases} 1 & \text{if } x \in \mathcal{S} \\ 0 & \text{if } x \notin \mathcal{S} \end{cases}. \tag{2.1}$$

2.2 Methods based on transformations

Sampling methods assume that a random source is available that produces samples from a known distribution which, in general, differs from the target distribution. However, in some cases, we can find adequate transformations that convert the samples provided by the available random source into samples distributed according to the target pdf. For this reason, now we recall the links between the pdf's of two random variables related through a known invertible transformation.

Consider two random variables $\mathbf{Y} = [Y_1, Y_2, ..., Y_m] \in \mathbb{R}^m$ and $\mathbf{Z} = [Z_1, Z_2, ..., Z_m] \in \mathbb{R}^m$ with joint pdf's $p(y_1, y_2, ...y_m)$ and $q(z_1, z_2, ..., z_m)$, respectively, related through an invertible transformation $\mathbf{\Psi} = [\psi_1, ..., \psi_m]$, i.e., $\mathbf{Y} = \mathbf{\Psi}(\mathbf{Z})$. Specifically, we can write

$$\begin{cases} Y_1 = \psi_1(Z_1, Z_2, ..., Z_m) \\ \vdots \\ Y_m = \psi_m(Z_1, Z_2, ..., Z_m) \end{cases}, \tag{2.2}$$

and we can also indicate the inverse transformation as

$$\begin{cases} Z_1 = \psi_1^{-1}(Y_1, Y_2, ..., Y_m) \\ \vdots \\ Z_m = \psi_m^{-1}(Y_1, Y_2, ..., Y_m) \end{cases}, \tag{2.3}$$

i.e., $\mathbf{Z} = \mathbf{\Psi}^{-1}(\mathbf{Y})$, where $\mathbf{\Psi}^{-1} = [\psi_1^{-1}, ..., \psi_m^{-1}]$.

Hence, the two joint pdf's, p and q, are linked by the following relationship

$$p(y_1, y_2, ...y_m) = q\left(\psi_1^{-1}(y_1, y_2, ..., y_m), ..., \psi_m^{-1}(y_1, y_2, ..., y_m)\right) \left|\det J^{-1}\right|, \tag{2.4}$$

where J^{-1} is the Jacobian matrix of the inverse transformation, i.e.,

$$J^{-1} = \begin{bmatrix} \dfrac{d\psi_1^{-1}}{dy_1} & \dfrac{d\psi_1^{-1}}{dy_2} & .... & \dfrac{d\psi_1^{-1}}{dy_m} \\ .... & .... & .... & .... & .... \\ \dfrac{d\psi_m^{-1}}{dy_1} & \dfrac{d\psi_m^{-1}}{dy_2} & .... & \dfrac{d\psi_m^{-1}}{dy_m} \end{bmatrix}. \tag{2.5}$$

Clearly, if we are able to draw samples $\mathbf{z}' = (z_1', z_2', ...z_m')$ from $q(z_1, z_2, ...z_m)$, we can obtain samples $\mathbf{y}' = (y_1', y_2', ...y_m')$ from $p(y_1, y_2, ...y_m)$ simply computing $\mathbf{y}' = \mathbf{\Psi}(\mathbf{z}')$. Otherwise, if we are able to draw samples $\mathbf{y}' = (y_1', y_2', ...y_m')$ from $p(y_1, y_2, ...y_m)$ we can generate $\mathbf{z}' = (z_1', z_2', ..., z_m')$ from $q(z_1, z_2, ...z_m)$ using the inverse relationship $\mathbf{z}' = \mathbf{\Psi}^{-1}(\mathbf{y}')$. In the next sections, we analyze some useful transformations.

2.2.1 Inversion method

Let X be a random variable with pdf $p_o(x)$. Moreover, consider the related cdf

$$F_X(x) = \text{Prob}\{X \leq x\} = \int_{-\infty}^{x} p_o(v)dv, \tag{2.6}$$

which is a monotonic increasing function, and its generalized inverse function defined as

$$F_X^{-1}(y) \triangleq \inf\{x \in \mathcal{D} : F_X(x) \geq y\}. \tag{2.7}$$

The following theorem provides a very useful tool to generate samples distributed according to the target density $p_o(x)$ using a uniform random distribution $\mathcal{U}([0, 1])$ as the random source.

Theorem 1 *If $U \sim \mathcal{U}([0, 1])$, then the r.v. $Z = F_X^{-1}(U)$ has density $p_o(x)$ [30, Chapter 2].*

Proof: The generalized inverse function satisfies, by definition,

$$\{(u, x) \in \mathbb{R}^2 : F_X^{-1}(u) \leq x\} = \{(u, x) \in \mathbb{R}^2 : u \leq F_X(x)\}$$

Therefore, we have

$$F_Z(z) = \text{Prob}\{Z = F_X^{-1}(U) \le z\} = \text{Prob}\{U \le F_X(z)\} = F_X(z), \qquad (2.8)$$

so that X and Z have the same cdf, $X \overset{d}{=} Z$, and, therefore the same density $p_o(x)$. $\quad \square$

If we have an analytically expression for the inverse function $F_X^{-1}(\cdot)$, then we can first generate a sample $u' \sim \mathcal{U}([0, 1])$, uniformly in $[0, 1]$, transform it, $x' = F_X^{-1}(u')$, and the resulting sample x' is distributed according to $p_o(x)$.

In general, it is straightforward to show that, given a random variable Y with c.d.f. $F_Y(y)$, the random variable defined with the monotonic transformation

$$Z = F_X^{-1}(F_Y(Y)) \qquad (2.9)$$

is distributed according to $p_o(x)$, i.e., $X \overset{d}{=} Z$.

The inverse transform method allows to easily generate i.i.d. random numbers from a generic pdf $p_o(x)$, but we need to obtain the c.d.f. F_X and its inverse F_X^{-1} analytically. In many practical cases both steps are intractable. In other cases, we are able to find $F_X(x)$ but it is impossible to invert it [30, Chapters 3-7],[29, 106]. Therefore, we often have to resort to different classes of random sampling algorithms. Related topics are addressed in Sections 2.5.5 and 2.5.6.

2.2.2   Non-monotonic transformations

In this section we assume that the relationship between two random variables Y and X (with density $q(y)$ and $p_o(x)$, respectively) can be expressed with a non-monotonic transformation $Y = \psi(X)$ [121]. Consider a generalization of the inverse function defined as the set

$$\psi^{-1}(y) = \{x \in \mathbb{R} : \psi(x) = y\}.$$

Since ψ is not monotonic, for a generic value y the set $\psi^{-1}(y)$ contains more than one solution, i.e., in general $\psi^{-1}(y) = \{x_1, ..., x_n\}$, so that the inverse function is not uniquely defined. Therefore, if we are able to draw a sample y' from $q(y)$ then, in general, we have to choose adequately one solution x' of the n elements in $\psi^{-1}(y')$ such that x' be distributed according to $p_o(x)$.

For simplicity, let us consider the case $n = 2$, i.e., $\psi^{-1}(y) = \{x_1, x_2\}$ for all possible values of y. This means that the function $\psi(x)$ can be decomposed

19

into two monotonic (and invertible) parts. Namely, for each y it is possible to find a value $x_c \in \mathbb{R}$ such that $\psi_1(x) \triangleq \psi(x)$ is a monotonic function in the domain $x \in (-\inf, x_c]$ and $\psi_2(x) \triangleq \psi(x)$ is another monotonic function in the domain $x \in [x_c, +\inf)$. Hence, we use the notation

$$x_1 = \psi_1^{-1}(y), \quad x_2 = \psi_2^{-1}(y) \tag{2.10}$$

to indicate the two solutions of the equation $y = \psi(x)$. Moreover, the density $q(y)$ of the r.v. $Y = \psi(X)$ can be expressed as

$$q(y) = p_o(\psi_1^{-1}(y)) \left| \frac{d\psi_1^{-1}}{dy} \right| + p_o(\psi_2^{-1}(y)) \left| \frac{d\psi_2^{-1}}{dy} \right|, \tag{2.11}$$

where $p_o(x)$ is the pdf of X. Then, given a sample y' from $q(y)$, we can obtain a sample x' from $p_o(x)$ taking $x_1' = \psi_1^{-1}(y')$ with probability

$$
\begin{aligned}
w_1 &= \frac{p_o(\psi_1^{-1}(y')) \left| \frac{d\psi_1^{-1}(y')}{dy} \right|}{q(y')} \\
&= \frac{p_o(\psi_1^{-1}(y')) \left| \frac{d\psi_1^{-1}(y')}{dy} \right|}{p_o(\psi_1^{-1}(y')) \left| \frac{d\psi_1^{-1}(y')}{dy} \right| + p_o(\psi_2^{-1}(y')) \left| \frac{d\psi_2^{-1}(y')}{dy} \right|},
\end{aligned}
\tag{2.12}
$$

or choosing $x_2' = \psi_2^{-1}(y')$ otherwise with probability $w_2 = 1 - w_1$. Moreover, since

$$\left| \frac{d\psi_i^{-1}(y')}{dy} \right| = \left| \frac{1}{\frac{d\psi(\psi_i^{-1}(y'))}{dx}} \right| = \left| \frac{1}{\frac{d\psi(x_i')}{dx}} \right|,$$

$i = 1, 2$, we can rewrite the weight w_1 in Eq. (2.12) (that represents the probability of accepting x_1') as

$$
\begin{aligned}
w_1 &= \frac{p_o(x_1') \left| 1/\frac{d\psi(x_1')}{dx} \right|}{p_o(x_1') \left| 1/\frac{d\psi(x_1')}{dx} \right| + p_o(x_2') \left| 1/\frac{d\psi(x_2')}{dx} \right|} \\
&= \frac{p_o(x_1') \left| \frac{d\psi(x_2')}{dx} \right|}{p_o(x_1') \left| \frac{d\psi(x_2')}{dx} \right| + p_o(x_2') \left| \frac{d\psi(x_1')}{dx} \right|}.
\end{aligned}
\tag{2.13}
$$

Therefore, if we have a non-monotonic transformation $Y = \psi(X)$, and the equation $x = \psi^{-1}(y)$ has $n = 2$ possible solutions, we can use the algorithm suggested in [121] and summarized in Table 2.1 to draw from $p_o(x)$.

Table 2.1: Non-monotonic transformation

1. Set $i = 1$. Let N be the number of desired samples from $p_o(x)$.
2. Draw a sample y' from $q(y)$ in Eq. (2.11).
3. Set $x'_1 = \psi_1^{-1}(y')$ and $x'_2 = \psi_2^{-1}(y')$.
4. Draw u' from $\mathcal{U}([0, 1])$.
5. If $u' \leq \dfrac{p_o(x'_1) \left| \frac{d\psi(x'_2)}{dx} \right|}{p_o(x'_1) \left| \frac{d\psi(x'_2)}{dx} \right| + p_o(x'_2) \left| \frac{d\psi(x'_1)}{dx} \right|}$ then set $x^{(i)} = x'_1$, otherwise $x^{(i)} = x'_2$.
5. Update $i = i + 1$. If $i > N$ then stop, else go back to step 2.

An interesting possibly non-monotonic transformation is $Y = p_o(X)$, i.e., the transformation $\psi(x) = p_o(x)$ is exactly the pdf of X. In this case the pdf $q(y)$ of Y is called *vertical density* [87, 147, 148].

2.2.3 Transformations of many random variables

Given $Z_1, Z_2, ..., Z_m$ r.v.'s with joint pdf $q(z_1, z_2, ..., z_m)$, another possibility to design a random sampler based on a transformation is to find a function $\phi : \mathbb{R}^m \to \mathbb{R}$ such that we can write

$$X = \phi(Z_1, ..., Z_m), \tag{2.14}$$

where X is distributed according to our target pdf $p_o(x)$. Therefore, if we are able to draw $(z'_1, ..., z'_m)$ from $q(z_1, ..., z_m)$ then the sample $x' = \phi(z'_1, ..., z'_m)$ is distributed according to $p_o(x)$.

To analyze the relationship between $q(z_1, ..., z_m)$ and $p_o(x)$ we have to study the system of equations

$$\begin{cases} X = \phi(Z_1, ..., Z_m), \\ Y_1 = Z_2, \\ \vdots \\ Y_{m-1} = Z_m, \end{cases} \tag{2.15}$$

where the $m - 1$ random variables involved in the equations $Y_1 = Z_2, ..., Y_{m-1} = Z_m$ are arbitrarily chosen. We also assume that ϕ can be inverted with respect to (w.r.t.) z_1, i.e., $\frac{d\phi}{dz_1} \neq 0$. Hence, the inverse transformation is

$$\begin{cases} Z_1 = \phi^{-1}(X, Y_1, ..., Y_{m-1}), \\ Z_2 = Y_1, \\ \vdots \\ Z_m = Y_{m-1}, \end{cases} \tag{2.16}$$

where ϕ^{-1} represents the solution of the equation $x = \phi(z_1, z_2, ..., z_m)$ w.r.t. the variable z_1, i.e., $z_1 = \phi^{-1}(x, z_2, ..., z_m)$. The Jacobian matrix of the transformation in Eq. (2.16) is

$$J^{-1} = \begin{bmatrix} \dfrac{d\phi^{-1}}{dx} & \dfrac{d\phi^{-1}}{dy_1} & .... & \dfrac{d\phi^{-1}}{dy_{m-1}} \\ 0 & 1 & .... & 0 \\ .... & .... & .... & .... \\ 0 & 0 & .... & 1 \end{bmatrix}, \tag{2.17}$$

so that $|\det J^{-1}| = \left| \frac{d\phi^{-1}}{dx} \right|$. Hence, in this case Eq. (2.4) reduces to

$$p(x, y_2, ...y_m) = q\left(\phi^{-1}(x, y_1, ..., y_{m-1}), y_2, ..., y_m\right) \left| \frac{d\phi^{-1}}{dx} \right|, \tag{2.18}$$

where $p(x, y_2, ...y_m)$ is the joint density of the vector $[X, Y_2, ..., Y_m]$. Finally, we have to marginalize in order to obtain our target pdf $p_o(x)$, i.e.,

$$p_o(x) = \underbrace{\int_{\mathcal{C}_2} ... \int_{\mathcal{C}_{m-1}} \int_{\mathcal{C}_m}}_{m-1} p(x, y_2, ...y_m) dy_2 ... dy_m$$

$$= \int_{\mathcal{C}_2} ... \int_{\mathcal{C}_m} q\left(\phi^{-1}(x, y_1, ..., y_{m-1}), y_2, ..., y_m\right) \left| \frac{d\phi^{-1}}{dx} \right| dy_2 ... dy_m. \tag{2.19}$$

One well-known example, in which this method is applied, is the *Box-Muller transformation*,

$$X = \phi(U_1, U_2) = \sqrt{-2 \log U_1} \cos(2\pi U_2), \tag{2.20}$$

that converts two uniform r.v.'s U_1 and U_2, $U_i \sim \mathcal{U}([0, 1])$, $i = 1, 2$, into a standard Gaussian r.v. $X \sim N(0, 1)$. Other relevant transformations are:

1. The sum of random variables, i.e., for $m = 2$

$$X = Z_1 + Z_2, \tag{2.21}$$

where Z_1, Z_2 have joint pdf $q(z_1, z_2)$ and $p_o(x) = \int_{\mathcal{C}} q\,(x - y, y)\,dy$. If Z_1 are Z_2 independent, this technique is also named *convolution* method.

2. The product or ratio of two random variables (see, for example, Sections 2.4.1 and 2.8), i.e.,

$$X = Z_1 Z_2, \tag{2.22}$$

or

$$X = \frac{Z_1}{Z_2}, \tag{2.23}$$

where we have $p_o(x) = \int_{\mathcal{C}} \frac{1}{|y|} q(\frac{x}{y}, y)\,dy$ and $p_o(x) = \int_{\mathcal{C}} |y| q\,(xy, y)\,dy$, respectively.

2.2.4 Deconvolution method

Let us consider two r.v.'s X, Z with known joint pdf $f(x, z)$, where X has a density $p_o(x)$. We also know the relationship

$$Y = \varphi(X, Z), \tag{2.24}$$

where Y is distributed according to $q(y)$, and assume that we are able to evaluate $q(y)$ and to draw from it. Our goal is to generate samples from $p_o(x)$.

Assuming $\frac{d\varphi}{dz} \neq 0$, Eq. (2.18) can be rewritten as

$$p(x, y) = f(x, \varphi^{-1}(x, y)) \left| \frac{d\varphi^{-1}}{dy} \right|, \tag{2.25}$$

where we have substituted $z = \varphi^{-1}(x, y)$. Moreover, $q(y)$ is a marginal density of $p(x, y)$, i.e., $q(y) = \int_{\mathcal{D}} p(x, y)dx$ and, obviously, $p_o(x)$ is the other marginal pdf, i.e., $p_o(x) = \int_{\mathcal{C}} p(x, y)dy$.

Therefore, since we can rewrite the joint pdf as $p(x, y) = h(x|y)q(y)$, we can draw from $p_o(x)$ following the procedure[1]:

---

[1]This method was termed "deconvolution" because it was first developed for the specific transformation $\varphi(X, Z) = X + Z$. See [30].

1. Generate y' from $q(y)$.

2. Draw x' from $h(x|y')$, where the conditional pdf $h(x|y)$ is

$$h(x|y) = \frac{p(x,y)}{q(y)} = \frac{f(x, \varphi^{-1}(x,y)) \left| \frac{d\varphi^{-1}}{dy} \right|}{q(y)}.$$ (2.26)

This method is tightly connected with the class of *polar algorithms* [30, Chapter 5]. Moreover, this technique is a particular case of a more general idea described in Section 2.4.2.

2.3   Discrete mixture of densities

Let us assume that the target pdf can be expressed as

$$p_o(x) = \omega_1 h_1(x) + \omega_2 h_2(x) + ... + \omega_n h_n(x) = \sum_{i=1}^{n} \omega_i h_i(x),$$ (2.27)

where $\omega_i \geq 0$, $i = 1, ..., n$, $\sum_{i=1}^{n} \omega_i = 1$ and the functions $h_i(x)$, $i = 1, ..., n$, $x \in \mathcal{D}$, are densities that we assume easy to draw from. The sum in Eq. (2.27) is termed as discrete mixture of densities.

In order to draw from $p_o(x)$ we can follow these steps:

1. Draw an index $j' \in \{1, ..., n\}$, according to the weights ω_i, $i, ..., n$. Namely, we generate a random index j' with probability mass function (pmf) $\text{Prob}\{j' = i\} = \omega_i$.

2. Draw x' from the pdf $h_{j'}(x)$.

Given a partition of the domain $\mathcal{D} = \cup_{i=1}^{n} \mathcal{D}_i$, $\mathcal{D}_i \cap \mathcal{D}_j = \emptyset$ for $i \neq j$, if each density $h_i(x)$ is defined in a support $\mathcal{D}_i$, i.e., if the pdf's $h_i(x)$, $i = 1, .., n$, have disjoint (non-overlapping) domains, the algorithm is also known as composition method [64, Chapter 2]. This idea, jointly with the rejection technique described later in Section 2.5.1, is used in the so-called *patchwork algorithms* [77, 80, 140].

If the target density can be expressed as a series of densities, i.e.,

$$p_o(x) = \sum_{i=1}^{+\infty} \omega_i h_i(x),$$ (2.28)

it is also possible to draw from $p_o(x)$, but we need the ability to simulate the discrete pmf $\text{Prob}\{i\} = \omega_i$, $i = 1, ..., +\infty$ [44, Chapter 2], [64, Chapter 3] (note that this is always possible by the inversion method using a sequential search procedure).

Additionally, in some specific cases (see [30]) it is possible to assume that the weights ω_i in Eq. (2.27) be negative. However, there are no general methods to draw from $p_o(x)$ in such case.

Finally, it is also possible to define a *continuous* mixture of pdf's, as shown later in Section 2.4.2. Other related topics are tackled in Section 2.5.3.

2.4   The fundamental theorem of simulation

Many Monte Carlo techniques (inverse-of-density method, rejection sampling, slice sampling etc....) are based on a simple result that we enunciate below.

Theorem 2 *[134, Chapter 2] Drawing samples from a unidimensional r.v. X with density $p_o(x) \propto p(x)$ is equivalent to sample uniformly on the bidimensional region defined by*

$$\mathcal{A}_0 = \{(x, u) \in \mathbb{R}^2 : \ 0 \leq u \leq p(x)\}. \tag{2.29}$$

Namely, if (x', u') is uniformly distributed on $\mathcal{A}_0$, then x' is a sample from $p_o(x)$.

Proof: Let us consider a random pair (X, U) uniformly distributed on the region $\mathcal{A}_0$ in Eq. (2.29), and let $q(x, u)$ be its joint pdf, i.e.,

$$q(x, u) = \frac{1}{|\mathcal{A}_0|} \mathbb{I}_{\mathcal{A}_0}(x, u), \tag{2.30}$$

where $\mathbb{I}_{\mathcal{A}_0}(x, u)$ is the indicator function on $\mathcal{A}_0$ and $|\mathcal{A}_0|$ is the area of the region $\mathcal{A}_0$. Clearly, we can also write $q(x, u) = q(u|x)q(x)$. The theorem is proved if the marginal density $q(x)$ is exactly $p_o(x)$.

Since (X, U) is uniformly distributed on the set $\mathcal{A}_0$ defined in Eq. (2.29), we have $q(u|x) = 1/p(x)$ with $0 \leq u \leq p(x)$, or, in a more compact form,

$$q(u|x) = \frac{1}{p(x)} \mathbb{I}_{\mathcal{A}_0}(x, u). \tag{2.31}$$

Therefore, we can express the joint pdf as

$$q(x, u) = q(u|x)q(x) = \frac{1}{p(x)} \mathbb{I}_{\mathcal{A}_0}(x, u) \cdot q(x). \qquad (2.32)$$

Taking Eqs. (2.30) and (2.32) together and solving for $q(x)$ yields

$$q(x) = \frac{1}{|\mathcal{A}_0|} p(x) = p_o(x). \quad \square \qquad (2.33)$$

Therefore, if we are able to draw a pair (x', u') uniformly on the region $\mathcal{A}_0$, the coordinate x' is marginally distributed according to $p_o(x)$. Many Monte Carlo techniques simulate jointly the random variables (X, U) and then consider only the first sample x'. The variable U plays the role of an *auxiliary* variable.

Figure 2.1 depicts a target function $p(x) \propto p_o(x)$ and the green area $\mathcal{A}_0$ below.
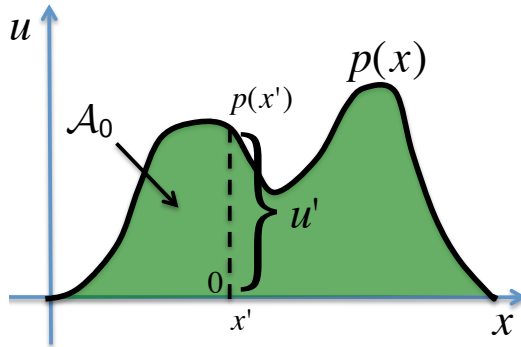
Figure 2.1: The area $\mathcal{A}_0$ below the target function $p(x) \propto p_o(x)$.

The next two techniques (the inverse-of-density method and the rejection sampling method) are clear examples of how this simple idea can be used to design Monte Carlo sampling algorithms. For instance, since one marginal density of $q(x, u)$ in Eq. (2.30) is exactly $p_o(x)$, the inverse-of-density method considers the possible employment of the other marginal density $q(u) = \int_{\mathcal{D}} q(x, u) dx$ in order to draw from $p_o(x)$.

2.4.1 Inverse-of-density method for monotonic pdf's

We present the inverse-of-density method [30, Chapter 4], [73], also known as *Khinchine's method* [78], for monotonic densities. However, it can be easily

extended for generic pdf's.

Given a monotonic target pdf $p_o(x)$ and the equation $u = p_o(x)$, we indicate with $p_o^{-1}(u)$ the corresponding inverse function of the target density. Note that $p_o^{-1}(u)$ is also a normalized density since it describes the same area $\mathcal{A}_0$ defined by $p_o(x)$, as shown in Figure 2.2.

(a)   (b)
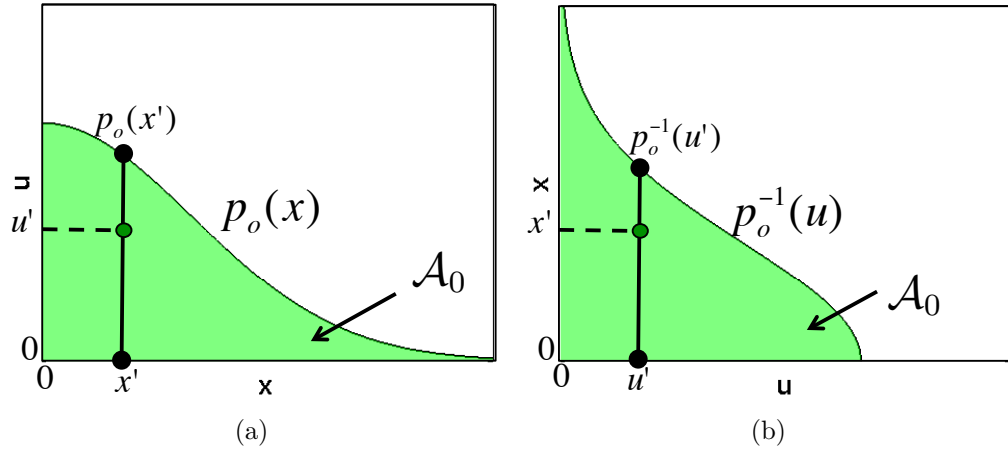
Figure 2.2: Two ways to draw a random point (x', u') uniformly in the area $\mathcal{A}_0$. **(a)** We can first draw x' from $p_o(x)$ and then $u' \sim \mathcal{U}([0, p_o(x')])$. **(b)** Otherwise, we can first draw u' from $p_o^{-1}(u)$ and then $x' \sim \mathcal{U}([0, p_o^{-1}(u')])$.

Therefore, we can write

$$\mathcal{A}_0 = \{(x, u) \in \mathbb{R}^2 : \ 0 \le u \le p_o(x)\}$$

or

$$\mathcal{A}_0 = \{(u, x) \in \mathbb{R}^2 : \ 0 \le x \le p_o^{-1}(u)\}.$$

Therefore, to generate samples (x', u') uniformly in $\mathcal{A}_0$ we can proceed in two alternative ways:

1. Draw x' from $p_o(x)$ and then u' uniformly in the interval $[0, p_o(x')]$, i.e, $u' \sim \mathcal{U}([0, p_o(x')])$ (see Figure 2.2(a)),

2. Draw u' from $p_o^{-1}(u)$ and x' uniformly in the interval $[0, p_o^{-1}(u')]$, i.e., $x' \sim \mathcal{U}([0, p_o^{-1}(u')])$ (see Figure 2.2(b)).

Both procedures generate points (x', u') uniformly distributed on the region $\mathcal{A}_0$. Moreover, from the fundamental theorem of simulation, the first coordinate x' is distributed according to the target pdf $p_o(x)$, while the second coordinate u' is distributed according to the inverse pdf $p_o^{-1}(u)$. Hence, if we are able to draw samples u' from $p_o^{-1}(u)$, we can use the second procedure to generate samples x' from $p_o(x)$.

Note that generating a sample x' uniformly in the interval $[0, a]$, i.e. $x' \sim \mathcal{U}([0, a])$, is equivalent to drawing a sample z' uniformly in $[0, 1]$ and then multiply it by a, i.e. $x' = z'a$. In the same way given a known value u', to draw a sample x' uniformly in the interval $[0, p_o^{-1}(u')]$, i.e. $x' \sim \mathcal{U}([0, p_o^{-1}(u')])$, is equivalent to generate a sample z' uniformly in $[0, 1]$ and then take $x' = z'p_o^{-1}(u')$. The algorithm described in Table 2.2 uses exactly the latter procedure. Obviously, to use this technique we need the ability to draw from the inverse pdf $p_o^{-1}(u)$.

Table 2.2: Inverse-of-density algorithm (Version 1).

1. Set $i = 1$. Let N be the number of desired samples from $p_o(x)$.
2. Draw a sample u' from $p_o^{-1}(u)$.
3. Draw z' uniformly in $[0, 1]$, i.e., $z' \sim \mathcal{U}([0, 1])$.
4. Then set $x^{(i)} = z'p_o^{-1}(u')$ and $i = i + 1$.
5. If $i > N$ then stop, else go back to step 2.

The inverse-of-density method can be summarized by the following relationship

$$X = Zp_o^{-1}(U), \tag{2.34}$$

where X has density $p_o(x)$, $Z \sim \mathcal{U}([0, 1])$ and U is distributed according to $p_o^{-1}(u)$. In the literature, it is possible to find this method in an alternative form. This second version considers the transformed variable $W = p_o^{-1}(U)$ where U has the pdf $p_o^{-1}(u)$. As a consequence, the density of W is

$$q(w) = p_o^{-1}(p_o(w)) \left| \frac{dp_o}{dw} \right| = w \left| \frac{dp_o}{dw} \right|. \tag{2.35}$$

The function in Eq. (2.35) is the *vertical density* associated to the inverse pdf $p_o^{-1}(u)$ [73, 147, 148].

Using the r.v. W, we can express the Eq. (2.34) as

$$X = ZW,$$

where $Z \sim \mathcal{U}([0, 1])$ and W is distributed according to $q(w)$ in Eq. (2.35). Table 2.3 outlines this alternative form of the inverse-of-density method.

Table 2.3: Inverse-of-density algorithm (Version 2).

1. Set $i = 1$. Let N be the number of desired samples from $p_o(x)$.

2. Draw a sample w' from $q(w)$ in Eq. (2.35).

3. Draw z' uniformly in $[0, 1]$, i.e., $z' \sim \mathcal{U}([0, 1])$.

4. Set $x^{(i)} = z'w'$ and $i = i + 1$.

5. If $i > N$ then stop, else go back to step 2.

Finally, it is important to notice that this technique coincides with the *vertical density representation (VDR) type 2* introduced in [38].

2.4.2 Continuous mixtures

The fundamental theorem of simulation represents the target density as a marginal density of a joint pdf, which is uniform on the region $\mathcal{A}_0$ in Eq. (2.29). From a general point of view, we can consider a joint pdf $p(x, y)$ such that

$$p_o(x) = \int_{\mathcal{C}} p(x, y) dy, \tag{2.36}$$

where $\mathcal{C}$ is the support of the variable y. This equation can be considered as an integral representation of the density $p_o(x)$. Moreover, since we can express the joint pdf as $p(x, y) = h(x|y)q(y)$, for some adequate conditional density h, we can also write

$$p_o(x) = \int_{\mathcal{C}} h(x|y)q(y) dy. \tag{2.37}$$

Therefore, if we are able to draw a sample y' from the marginal pdf $q(y)$ and then x' from $h(x|y')$, the sample x' has density $p_o(x)$. The variable y plays the role of an auxiliary variable.

This method is also known as *continuous mixture* of densities [30, Chapter 1]. Indeed, we can consider the marginal pdf $q(y)$ as a weight function and $h(x|y)$ as an uncountable collection of densities in the continuous mixture. Specifically, the weight $q(y^*)$ is associated to the density $h(x|y^*)$.

The methods in Sections 2.2.3 and 2.2.4 are specific examples of this idea.

2.5   Rejection sampling

The accept/reject method, also known as *rejection sampling* (RS), was suggested by John von Neumann in 1951 [152]. It is a classical Monte Carlo technique for "universal sampling". It can be used to generate samples from any target density $p_o(x)$ by drawing from a possibly simpler proposal density $\pi(x)$ [30, Chapter 2]. 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. Specifically, the RS algorithm can be viewed as choosing a subsequence of i.i.d. realizations from the proposal density $\pi(x)$ in such a way the elements of the subsequence have density $p_o(x)$.

This technique requires the ability to evaluate the density $p_o(x)$ of interest up to a multiplicative constant (which is often the case in practical applications). However, an important limitation of RS methods is the need to analytically establish a bound for the ratio of the target and proposal densities, since there is a lack of general procedures for the computation of such tight bounds.

In the rest of this section, we review the basic RS algorithm as well as some variations.

2.5.1   Rejection sampling algorithm

Consider a function $p(x) \propto p_o(x)$ and a proposal density $\pi(x)$ easy to simulate. Additionally, choose a constant L such that $L\pi(x)$ is an overbounding function for $p(x)$, i.e.,

$$L\pi(x) \geq p(x), \tag{2.38}$$

for all $x \in \mathcal{D}$. The constant L obviously is an upper bound for the ratio $p(x)/\pi(x)$, i.e.,

$$L \geq \frac{p(x)}{\pi(x)} \quad \forall x \in \mathcal{D}. \tag{2.39}$$

In the standard rejection sampling algorithm, we first draw a sample from the proposal pdf, $x' \sim \pi(x)$, and then accept it with probability

$$p_A(x') = \frac{p(x')}{L\pi(x')} \leq 1. \tag{2.40}$$

Otherwise, the proposed sample x' is discarded. In Figure 2.3, we can see a graphical representation of the rejection sampling technique. The RS procedure can also be outlined as:

1. Draw x' from $\pi(x)$.

2. Generate v' uniformly in the interval $[0, L\pi(x')]$, i.e., $v' \sim \mathcal{U}([0, L\pi(x')])$.

3. If the point (x', v') belongs to the area $\mathcal{A}_0$ (see Eq. (2.29)) below the target function $p(x)$, the sample x' is accepted.

4. Otherwise, when the point (x', v') falls into the region between the functions $L\pi(x)$ and $p(x)$, the sample x' is rejected.

We can also summarize this procedure in an equivalent way: first draw a sample x' from $\pi(x)$ and $u' \sim \mathcal{U}([0, 1])$. If $u'L\pi(x') \leq p(x')$, we accept x'. Table 2.4 describes how we can generate N samples from the target pdf $p_o(x) \propto p(x)$ according to the standard rejection sampling algorithm.

The RS technique is based on the following theorem.

Theorem 3 *Let the r.v.'s X_1 and X_2 have pdf's $\pi(x)$ and $p_o(x) \propto p(x)$, respectively, and let U have a uniform distribution $\mathcal{U}([0, 1])$. If there exists a bound $L \geq p(x)/\pi(x) \ \forall x \in \mathcal{D}$, then*

$$Prob\left\{ X_1 \leq y \,\middle|\, U \leq \frac{p(X_1)}{L\pi(X_1)} \right\} = Prob\{X_2 \leq y\}. \tag{2.41}$$

31
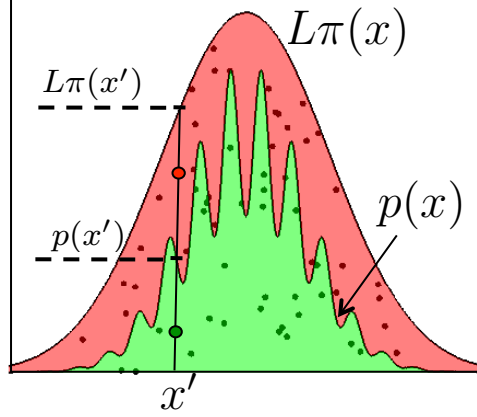
Figure 2.3: Graphical description of the RS procedure. The green region corresponds to the area $\mathcal{A}_0$ defined in Eq. (2.29), with $p(x) = \exp(-x^2/2)(\sin(6x)^2 + 1)$. The red region indicates the area between the functions $L\pi(x)$ and $p(x)$. First, a sample x' is generated from the proposal pdf $\pi(x)$ and another coordinate $v' \sim \mathcal{U}([0, L\pi(x')])$. If the point (x', v') belongs to the area $\mathcal{A}_0$ (green region) the sample x' is accepted. Otherwise, it is discarded.

Proof: Assuming $\mathcal{D} = \mathbb{R}$ (without lack of generality) and recalling that $U \sim \mathcal{U}([0,1])$ and X_1 is distributed according to $\pi(x)$, we can easily write

$$\text{Prob}\left\{X_1 \leq y \middle| U \leq \frac{p(X_1)}{L\pi(X_1)}\right\} = \frac{\text{Prob}\left\{X_1 \leq y, U \leq \frac{p(X_1)}{L\pi(X_1)}\right\}}{\text{Prob}\left\{U \leq \frac{p(X_1)}{L\pi(X_1)}\right\}}$$

$$= \frac{\int_{-\infty}^{y} \int_{0}^{\frac{p(x)}{L\pi(x)}} \pi(x)dudx}{\int_{-\infty}^{+\infty} \int_{0}^{\frac{p(x)}{L\pi(x)}} \pi(x)dudx}.$$

Then, integrating first w.r.t. u and after some trivial calculations, we arrive at the expression

$$\text{Prob}\left\{X_1 \leq y \middle| U \leq \frac{p(X_1)}{L\pi(X_1)}\right\} = \frac{\int_{-\infty}^{y} p(x)dx}{\int_{-\infty}^{+\infty} p(x)dx}.$$

32

Furthermore, since $p_o(x) \propto p(x)$, i.e., $p(x) = cp_o(x)$ where c is a normalization constant, we can rewrite the expression above as

$$\text{Prob}\left\{X_1 \le y \middle| U \le \frac{p(X_1)}{L\pi(X_1)}\right\} = \frac{\int_{-\infty}^{y} p(x)dx}{\int_{-\infty}^{+\infty} p(x)dx} = \frac{\int_{-\infty}^{y} cp_o(x)dx}{\int_{-\infty}^{+\infty} cp_o(x)dx} =$$
$$= \int_{-\infty}^{y} p_o(x)dx.$$

Finally, since the r.v. X_2 has density $p_o(x)$, we can write

$$\text{Prob}\left\{X_1 \le y \middle| U \le \frac{p(X_1)}{L\pi(X_1)}\right\} = \int_{-\infty}^{y} p_o(x)dx = P\{X_2 \le y\}, \qquad (2.42)$$

so that the expression in Eq. (2.41) is verified. $\quad\square$

Table 2.4: Rejection sampling algorithm.

1. Set $i = 1$. Let N be the number of desired samples from $p_o(x)$.

2. Draw a sample x' from $\pi(x)$ and u' from $\mathcal{U}([0,1])$.

3. If $u' \le \frac{p(x')}{L\pi(x')}$ then $x^{(i)} = x'$ and set $i = i + 1$.

4. Else if $u' > \frac{p(x')}{L\pi(x')}$ then discard x' and go back to step 2.

5. If $i > N$ then stop, else go back to step 2.

Let us assume, for a moment, that $p(x) = p_o(x)$. It is interesting to note that, if the inequality $L\pi(x) \ge p_o(x)$ is not guaranteed, the samples generated by a RS procedure are distributed according to $\min\{p_o(x), L\pi(x)\}$. Indeed given x' drawn from $\pi(x)$, if we have $L\pi(x') < p_o(x')$, the sample x' is accepted with probability 1. Hence, the samples that belong to the region where $L\pi(x) < p_o(x)$ are distributed according to $\pi(x)$. Otherwise when we have $L\pi(x') \ge p_o(x')$, the rejection sampler accepts x' if it is exactly distributed according to $p_o(x)$.

The most favorable scenario to use the RS algorithm occurs when $p(x)$ is bounded with bounded domain. Indeed, in this case, the proposal pdf $\pi(x)$

can be chosen as a uniform density (the easiest possible proposal). Moreover, the calculation of the bound L for the ratio $p(x)/\pi(x)$ is converted into the problem of finding an upper bound for the target function $p(x)$ (in general, it is an easier issue). Otherwise when $p(x)$ is unbounded or its domain is infinite, the proposal $\pi(x)$ cannot be a uniform density. Sections 2.5.6-2.8 are devoted to describe methods that transform $p(x)$ by embedding it in a finite region. Figure 2.4 illustrates these three possible cases[2].

(a)   (b)   (c)
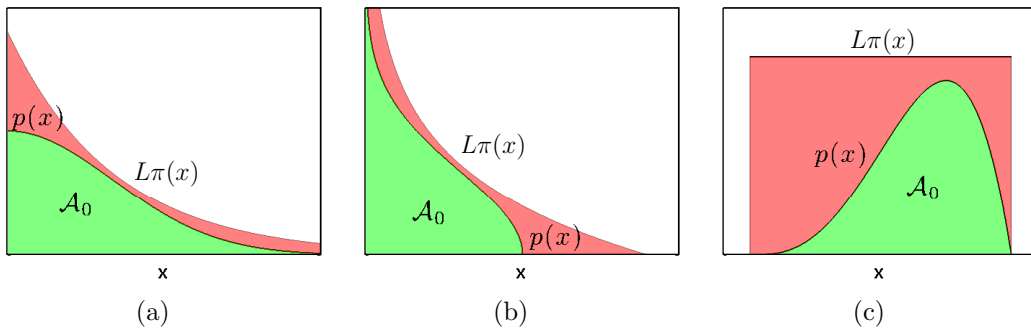
Figure 2.4: Three possible cases of density $p_o(x) \propto p(x)$ with three possible overbounding functions $L\pi(x)$: **(a)** bounded with infinite domain, **(b)** unbounded in a finite domain and **(c)** bounded with a finite domain. Only in the last case it is possible to use a uniform distribution as a proposal pdf $\pi(x)$.

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 upper bound L and, in general, a "good" overbounding function, i.e., $L\pi(x) \geq p(x)$, is crucial for the performance of a rejection sampling algorithm. We define the acceptance rate as

$$\text{Prob}\left\{U \leq \frac{p(X)}{L\pi(X)}\right\} = \int_{\mathcal{D}} \frac{p(x)}{L\pi(x)}\pi(x)dx = \frac{\int_{\mathcal{D}} p(x)dx}{L} = \frac{c}{L} \qquad (2.43)$$

where $1/c$ is the normalization constant of $p(x)$. Note that in cases where $p(x) = p_o(x)$, $c = 1$ and the constant L is necessarily larger than 1. Moreover,

---

[2]There exists a fourth possible scenario, when the function $p(x)$ is unbounded with infinite support. However, we can consider it as a combination of the cases in Figure 2.4(a) and Figure 2.4(b).

for the ratio $p(x)/\pi(x)$ to remain bounded, it is necessary that $\pi(x)$ has tails that decay to zero slower than those of $p(x)$.

It is important to remark that the RS method has a structural limitation. Indeed, even if we are able to find the optimal bound $L^* = \sup p(x)/\pi(x)$ the acceptance rate can be far from 1, depending on the difference in shape between $p(x)$ and $\pi(x)$. Specifically, the efficiency of the RS algorithm is a function of how the shape of $\pi(x)$ approaches the shape of $p(x)$. This is an important criticism because, in general, the accept-reject algorithms generate many "useless" samples when rejecting. For this reason, in many cases an *importance sampling* approach [97, Chapter 2] is used to bypass this problem. In [20, 24, 96] we can find a comparison of the performances obtained by rejection and importance sampling. Moreover, the *rejection control* algorithm [17, 21, 99, 97] mixes these two approaches.

In order to overcome this drawback, many adaptive accept-reject schemes (see Sections 2.6 and 2.7) have been designed. The basic idea is to build "good" proposal densities $\pi(x)$ close to the target density $p_o(x)$ and, as a consequence, improve the acceptance rate.

2.5.2 Squeeze principle

If we observe carefully the RS algorithm described in Table 2.4, it is obvious that testing the acceptance condition is a fundamental and, in some cases, expensive step. Therefore, in order to speed up the RS technique, we can find a function $\varphi(x)$ easy to evaluate, often called also *squeeze function*, such that

$$\varphi(x) \leq p(x). \tag{2.44}$$

The basic idea of the squeeze principle is to add a previous test involving the squeeze function $\varphi(x)$ in order to avoid the evaluation of $p(x)$. The method can be summarize in this way:

1. Draw x' from $\pi(x)$ and $u' \sim \mathcal{U}([0,1])$.

2. If $u'L\pi(x') \leq \varphi(x')$ then accept x' without evaluating the target function $p(x)$.

3. Otherwise, if $u'L\pi(x') \leq p(x')$ then accept x', else reject x'.

Figure 2.5 illustrates the squeeze principle. If the point $\big(x', u'L\pi(x')\big)$ belongs to the area below $\varphi(x)$ (darker green area), the sample x' is already accepted.

Otherwise, we check whether the point stays in the lighter green area and in this case we also accept x'. The sample x' is discarded if the point $\left(x', u'L\pi(x')\right)$ falls within the red region.
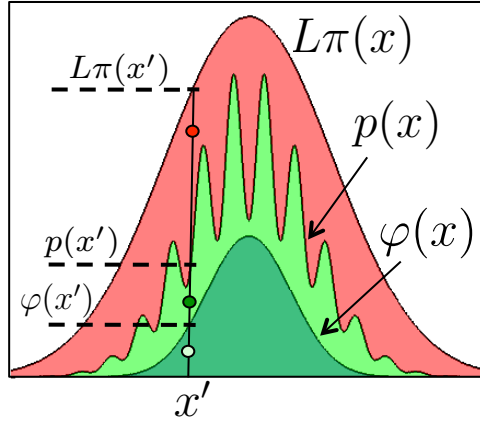
Figure 2.5: Squeeze principle: we first check whether the point $\left(x', u'L\pi(x')\right)$ falls within the darker green region, hence the sample x' is already accepted. Note that to check it we do not need to evaluate the function $p(x)$ but only the squeeze function $\varphi(x) \leq p(x)$.

A detailed step-by-step description is given in Table 2.5.

2.5.3   Acceptance-complement method

If we impose additional assumptions on the target $p_o(x)$ then we can construct variations of the rejection sampling algorithm that turn out more efficient for some problems. One example is the acceptance-complement method [88].

Let us assume that we are able to decompose our target pdf in two terms,

$$p_o(x) = g_1(x) + g_2(x), \tag{2.45}$$

where $\omega_1 = \int_{\mathcal{D}} g_1(x)dx$, $\omega_2 = \int_{\mathcal{D}} g_2(x)dx$ and, obviously, $\omega_1 + \omega_2 = 1$. Therefore, we can rewrite Eq. (2.45) as the discrete mixture

$$p_o(x) = \omega_1 h_1(x) + \omega_2 h_2(x), \tag{2.46}$$

where $h_1(x) \triangleq g_1(x)/\omega_1$ and $h_2(x) \triangleq g_2(x)/\omega_2$ are proper pdf's. Hence, if we are able to draw from $h_1(x)$ and $h_2(x)$, we can use the procedure described in Section 2.3.

Table 2.5: Rejection with Squeeze Algorithm.

1. Set $i = 1$. Let N be the number of desired samples from $p_o(x)$.
2. Draw samples x' from $\pi(x)$ and u' from $\mathcal{U}([0,1])$, where $\mathcal{U}([0,1])$ is the uniform pdf in $[0,1]$.
3. If $u' \leq \frac{\varphi(x')}{L\pi(x')}$ then $x^{(i)} = x'$, set $i = i+1$ and go to step 6.
4. Else if $u' \leq \frac{p(x')}{L\pi(x')}$ then $x^{(i)} = x'$, set $i = i+1$ and go to step 6.
5. Otherwise $u' > \frac{p(x')}{L\pi(x')}$, discard x' and go back to step 2.
6. If $i > N$ then stop, else go back to step 2.

Let us consider the case in which we are able to draw directly from $h_2(x)$ but not from $h_1(x)$. However, we assume to know a pdf $\pi(x)$ such that

$$\pi(x) \geq g_1(x) = \omega_1 h_1(x), \tag{2.47}$$

and, therefore, enables us to use rejection sampling. Note that $\pi(x)$ can be a proper density, different to $h_1(x)$, because $\int_{\mathcal{D}} g_1(x)dx = \omega_1 < 1$. In this scenario, the standard procedure to generate a sample x' from $p_o(x)$, as seen in Section 2.3, is:

1. Draw an index $j' \in \{1, 2\}$ according to the weights ω_1, ω_2.

2. If $j' = 2$ then generate x' from $h_2(x)$.

3. If $j' = 1$, then

   (a) Draw x' from $\pi(x)$ and u' from $\mathcal{U}([0,1])$.

   (b) If $u' \leq \frac{g_1(x')}{\pi(x')}$ then return x'.

   (c) Otherwise, if $u' > \frac{g_1(x')}{\pi(x')}$, repeat from step (a).

An interesting variant of this approach, that avoids rejection steps, is the so-called acceptance-complement method [88, 89]. In order to draw a sample x' from $p_o(x)$, the acceptance-complement algorithm performs the following steps.

1. Draw x' from $\pi(x)$ and u' from $\mathcal{U}([0,1])$.

2. If $u' \leq \frac{g_1(x')}{\pi(x')}$ then accept x'.

3. Otherwise, if $u' > \frac{g_1(x')}{\pi(x')}$, then draw x'' from $h_2(x)$ and accept it.

Clearly, since we avoid any rejection, this approach is computationally more "economic" and faster than the original procedure. Table 2.6 provides the algorithm, with more detail.

Table 2.6: Acceptance-complement algorithm.

1. Set $i = 1$. Let N be the number of desired samples from $p_o(x)$.
2. Find a suitable decomposition of the form of Eq. (2.46), i.e., $$p_o(x) = \omega_1 h_1(x) + \omega_2 h_2(x).$$
3. Draw samples x' from $\pi(x)$ and u' from $\mathcal{U}([0,1])$.
3. If $u' \leq \frac{\omega_1 h_1(x')}{\pi(x')}$ then accept $x^{(i)} = x'$ and set $i = i + 1$.
4. If $u' > \frac{\omega_1 h_1(x')}{\pi(x')}$ generate x'' from $h_2(x)$ and set $x^{(i)} = x''$.
5. If $i > N$ then stop, else go back to step 2.

The acceptance-complement technique is based on the following theorem.

Theorem 4 *If*

- X_1 *has pdf* $\pi(x)$,

- X_2 *has pdf* $h_2(x)$,

- X_3 *has density* $p_o(x) \propto p(x)$,

- $U \sim \mathcal{U}([0,1])$,

- *and* $\pi(x) \geq g_1(x) = \omega_1 h_1(x) \ \forall x \in \mathcal{D}$ *where* $\mathcal{D} \equiv \mathbb{R}$,

then

$$Prob\left\{X_1 \leq y \Big| U \leq \frac{\omega_1 h_1(X_1)}{\pi(X_1)}\right\} + Prob\left\{X_2 \leq y \Big| U > \frac{\omega_1 h_1(X_1)}{\pi(X_1)}\right\} = \qquad (2.48)$$
$$= Prob\{X_3 \leq y\}.$$

Proof: From Theorem 3, we can write

$$\mathrm{Prob}\left\{X_1 \leq y \Big| U \leq \frac{\omega_1 h_1(X_1)}{\pi(X_1)}\right\} = \omega_1 \int_{-\infty}^{y} h_1(x)dx. \qquad (2.49)$$

Moreover, the second term in (2.48) can be expressed as

$$\mathrm{Prob}\left\{X_2 \leq y \Big| U > \frac{\omega_1 h_1(X_1)}{\pi(X_1)}\right\} = \mathrm{Prob}\{X_2 \leq y\}\,\mathrm{Prob}\left\{U > \frac{\omega_1 h_1(X_1)}{\pi(X_1)}\right\}$$
$$(2.50)$$
$$= \left(\int_{-\infty}^{y} h_2(x)dx\right)\left(1 - \mathrm{Prob}\left\{U \leq \frac{\omega_1 h_1(X_1)}{\pi(X_1)}\right\}\right)$$
$$= \left(\int_{-\infty}^{y} h_2(x)dx\right)\left(1 - \int_{-\infty}^{+\infty} \frac{\omega_1 h_1(x)}{\pi(x)}\pi(x)dx\right) \qquad (2.51)$$
$$= \left(\int_{-\infty}^{y} h_2(x)dx\right)\left(1 - \omega_1 \int_{-\infty}^{+\infty} h_1(x)dx\right)$$
$$= \left(\int_{-\infty}^{y} h_2(x)dx\right)(1 - \omega_1),$$

where the equality (2.50) follows from the independence of X_1, X_2 and U, while Eq. (2.51) results from $\mathrm{Prob}\{U \leq Y\} = E[Y]$ with $Y \leq 1$, since $U \sim \mathcal{U}([0,1])$.

Furthermore, since $\omega_2 = 1 - \omega_1$, we have

$$\mathrm{Prob}\left\{X_2 \leq y \Big| U > \frac{\omega_1 h_1(X_1)}{\pi(X_1)}\right\} = \omega_2 \int_{-\infty}^{y} h_2(x)dx. \qquad (2.52)$$

Therefore, given Eqs. (2.49) and (2.52) we can write

$$\text{Prob}\left\{X_1 \le y \middle| U \le \frac{\omega_1 h_1(X_1)}{\pi(X_1)}\right\} + \text{Prob}\left\{X_2 \le y \middle| U > \frac{\omega_1 h_1(X_1)}{\pi(X_1)}\right\} =$$

$$= \omega_1 \int_{-\infty}^{y} h_1(x)dx + \omega_2 \int_{-\infty}^{y} h_2(x)dx \qquad (2.53)$$

$$= \int_{-\infty}^{y} p_o(x)dx = \text{Prob}\{X_3 \le y\},$$

since $p_o(x) = \omega_1 h_1(x) + \omega_2 h_2(x)$. □

Note that if we use the acceptance-complement procedure of Table 2.6 with a proposal $\pi(x)$ such that

$$L\pi(x) \ge g_1(x) = \omega_1 h_1(x),$$

where $L \ne 1$, then the generated samples x' are distributed according to the density

$$q(x) = \frac{\omega_1}{L} h_1(x) + \left(1 - \frac{\omega_1}{L}\right) h_2(x), \qquad (2.54)$$

that is different from our target pdf $p_o(x)$. In fact, if $L \to +\infty$ then $q(x) \to h_2(x)$, since we always discard the sample x' from $\pi(x)$ and generate x'' from $h_2(x)$ at each step.

2.5.4 Strip methods

The goal of this class of algorithms is to make the sampling method as simple and fast as possible [64, Chapter 5]. Indeed, the central idea is very straightforward: cover the area $\mathcal{A}_0$ in Eq. (2.29) below the target pdf $p_o(x)$ by a union of rectangles forming an overbounding region for $\mathcal{A}_0$, as depicted in Figure 2.6. Clearly when the number of rectangles increases, this "strip" approximation of $\mathcal{A}_0$ becomes tighter and tighter.

The resulting algorithms are quite simple and fast but, in the basic formulation, they only work with bounded pdf's with finite support (however, there exist some variations in the literature [125], [149, Chapter 4] that may extend the applicability of this technique). Moreover, it is necessary to know the stationary points of the target pdf $p_o(x)$, i.e., to know where $p_o(x)$ is increasing or decreasing. For this reason and to simplify the treatment, we consider a monotonic decreasing target pdf $p_o(x) \propto p(x)$.
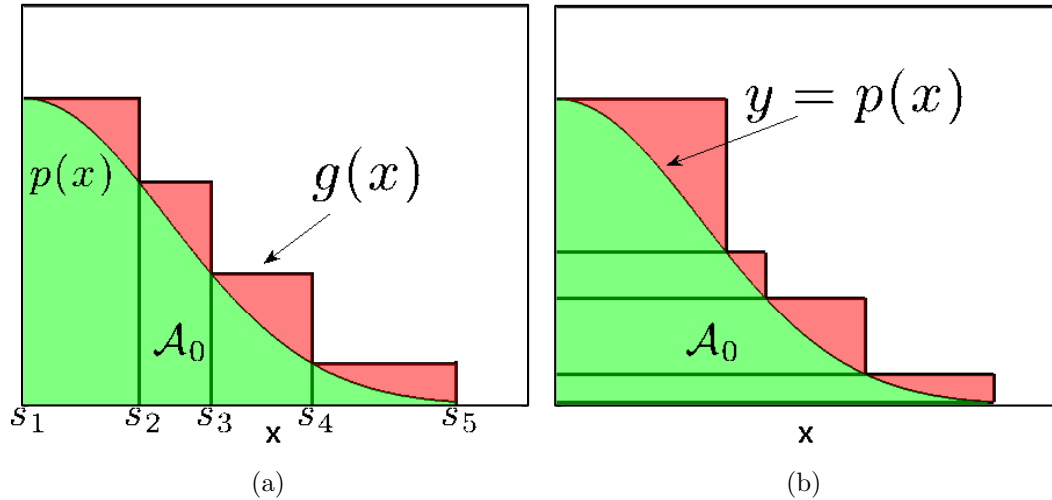
Figure 2.6: Examples of construction of a region $\mathcal{R}$ formed by 5 rectangles that covers the area $\mathcal{A}_0$ ($\mathcal{A}_0 \subseteq \mathcal{R}$) below $p(x)$. The bounded function $p(x) \propto p_o(x)$ is defined in a bounded domain. **(a)** The region $\mathcal{R}$ is composed by vertical bars. **(b)** The region $\mathcal{R}$ is composed by horizontal bars.

A strip method relies on building a stepwise overbounding function

$$g(x) \geq p(x).$$

There are two possibilities to construct it:

1. The first approach uses *vertical strips* (see Figure 2.6(a)) and is also named *Ahrens method* in honor of J. Ahrens [1, 2, 62].

2. The second way, that uses *horizontal strips* (see Figure 2.6(b)), is also called *ziggurat method* [107].

In the sequel, we focus our attention on the first approach (the development with horizontal strips is equivalent). Consider, for a moment, a bounded pdf $p_o(x)$ with bounded domain $\mathcal{D} = [a, b]$. We choose a set of support points

$$\mathcal{S} = \{s_1 = a, s_2, ..., s_{n-1}, s_{n+1} = b\} \tag{2.55}$$

where $s_i \in \mathcal{D}$, $\forall i$, and sort them in ascending order, $s_1 < s_2 < .... < s_{n+1}$. We can define the intervals $\mathcal{D}_i = [s_i, s_{i+1}]$, $i = 1, ..., n$, that form a partition

$\mathcal{D} = \cup_{i=1}^n \mathcal{D}_i$. Moreover, since we assume that the function $p(x) \propto p_o(x)$ is decreasing, i.e., $p(s_i) \geq p(s_{i+1})$, $i = 1, ..., n$, the rectangular set

$$\mathcal{R}_i \triangleq [s_i, s_{i+1}] \times [0, p(s_i)], \tag{2.56}$$

embeds the area below $p(x)$ for all $x \in \mathcal{D}_i$. Hence, the region composed by rectangular pieces $\mathcal{R} = \cup_{i=1}^n \mathcal{R}_i$ covers the area below $p(x)$, for $x \in \mathcal{D} = [a, b]$. Furthermore, the stepwise function

$$g(x) \triangleq \begin{cases} p(s_1), & \text{if } x \in \mathcal{D}_1, \\ \vdots \\ p(s_n), & \text{if } x \in \mathcal{D}_n. \end{cases} \tag{2.57}$$

yields an upper bound for $p(x)$, i.e., $p(x) \leq g(x) \; \forall x$. Therefore, we can easily define a proposal density $\pi(x) \propto g(x)$ as a mixture of uniform pdf's

$$\pi(x) \triangleq \sum_{i=1}^n \omega_i \mathbb{I}_{\mathcal{D}_i}(x), \tag{2.58}$$

where $\mathbb{I}_{\mathcal{D}_i}(x)$ is the indicator function for the set $\mathcal{D}_i$ (that yields 1 if $x \in \mathcal{D}_i$ and 0 if $x \notin \mathcal{D}_i$), and the weights are defined as

$$\omega_i = \frac{|\mathcal{R}_i|}{\sum_{i=1}^n |\mathcal{R}_i|}, \tag{2.59}$$

for $i = 1, ..., n$. Hence, if we first draw an index j' with $\text{Prob}\{j = j'\} = \omega_{j'}$ and a uniform sample x' from $\mathcal{U}([s_{j'}, s_{j'+1}])$, then x' is distributed as the proposal pdf $\pi(x)$ in Eq. (2.58) (this is explained in Section 2.3). Therefore, recalling also the overbounding function $g(x) \propto \pi(x)$ in Eq. (2.57), the algorithm to draw a sample x' can be summarized as follows.

1. Draw x' from $\pi(x)$ in Eq. (2.58) and u' from $\mathcal{U}([0, 1])$.

2. If $u' \leq \frac{p(x')}{g(x')}$ then accept x' otherwise discard it.

Table 2.7 provides a detailed description of the vertical strip algorithm. Recall that $g(x) = p(s_i) \; \forall x \in \mathcal{D}_i$.

Different strategies have been studied to choose the positions of the support points in order to improve the acceptance rate of the rejection

Table 2.7: Vertical strip algorithm for a decreasing pdf $p_o(x)$.

1. Set $i = 1$. Let N be the number of desired samples from $p_o(x)$.

2. Draw an index j' with $\text{Prob}\{j = j'\} = \omega_{j'}$, $j' = \{1, ..., n\}$, as defined in Eq. (2.59).

3. Generate a point (x', u_2') uniformly in the rectangle $\mathcal{R}_{j'}$, i.e., $x' \sim \mathcal{U}([s_{j'}, s_{j'+1}])$ and $u_2' \sim \mathcal{U}([0, p(s_{j'})])$.

4. If $u_2' \leq p(x')$ then accept $x^{(i)} = x'$ and set $i = i + 1$.

5. Otherwise, if $u_2' > p(x')$, discard x' and go back to step 2.

6. If $i > N$ then stop, else go back to step 2.

sampler [64, Chapter 5], [94]. Specifically, the equal area approach (i.e., rectangles with equal areas, so that $\omega_i = 1/n$, $i = 1, ..., n$) has been considered in [2] and [30, Chapter 8], and advocated because of its simplicity. In this case, the technique is called *grid method*, as well.

The next technique, called inversion-rejection method, can be seen an extension of the strip methods to deal with unbounded pdf's or densities with unbounded domain.

2.5.5 Inversion-rejection method

In many cases, we are able to calculate analytically the cdf $F_X(x)$ but it is not possible to invert it. Therefore, the inversion method in Section 2.2.1 cannot be applied. To overcome this problem, numerical inversion methods have been proposed in the literature [64, Chapter 7]. However, this approach can be only considered approximate, since the generated samples are not *exactly* drawn from $p_o(x)$. Furthermore, it can be computationally demanding.

Other approaches that start from the inversion principle and guarantee exact sampling have been studied. An example is the inversion-rejection method [29, 30], which is a combination of the inversion and rejection algorithms for the case that the cdf $F_X(x)$ is computable but not invertible.

For simplicity, we describe this technique for a bounded and decreasing

43

target pdf $p_o(x)$ with infinite support $\mathcal{D} = [a, +\infty)$. However, the algorithm can easily be extended to more general pdf's. Consider an infinite sequence of support points

$$\mathcal{S} = \{s_1 = a, s_2, ..., s_i, s_{i+1}....\},$$

$s_i \in \mathcal{D}$, $i = 1, ..., +\infty$, sorted in ascending order. The sequence is fixed but needs not be stored: for instance, we can compute s_i from i or the previous point s_{i-1}. Moreover, we assume that the cdf $F_X(x)$ can be evaluated and a global upper bound M is also available, i.e.,

$$M \geq p(x), \tag{2.60}$$

where $p(x) \propto p_o(x)$. The algorithm consists in the following steps.

1. Generate v' from $\mathcal{U}([0, 1])$.

2. Find the index j' such that

$$F_X(s'_j) \leq v' \leq F_X(s_{j'+1}), \tag{2.61}$$

thus the interval $\mathcal{D}_{j'} = [s_{j'}, s_{j'+1}] \subset \mathcal{D}$ is chosen with probability $\mathrm{Prob}\{j = j'\} = F_X(s_{j'+1}) - F_X(s_{j'})$ by an inversion method. This is always possible since we can use, e.g., a sequential search to find $s_{j'}$.

3. Draw a sample x' uniformly from $\mathcal{D}_{j'} = [s_{j'}, s_{j'+1}]$, i.e., $x' \sim \mathcal{U}([s_{j'}, s_{j'+1}])$.

4. Generate u' from $\mathcal{U}([0, 1])$.

5. If $Mu' \leq p(x')$ then accept x'.

6. Otherwise, if $Mu' > p(x')$, then discard x'.

Hence, we first select an interval $\mathcal{D}_i = [s_i, s_{i+1}]$ with probability $F_X(s_{i+1}) - F_X(s_i)$ by inversion. Since the interval $\mathcal{D}_i$ is closed and $p_o(x)$ is bounded, we can use a uniform proposal pdf in $\mathcal{D}_i$ as proposal density and draw a sample x' from $p_o(x)$ by rejection. Since we have assumed $p(x)$ to be a decreasing function, this procedure can be easily improved if we define a stepwise overbounding function

$$g(x) \triangleq \begin{cases} p(s_1) & \forall x \in \mathcal{D}_1, \\ \vdots \\ p(s_i) & \forall x \in \mathcal{D}_i, \\ \vdots \end{cases} \tag{2.62}$$

44

such that $g(x) \geq p(x)$, for all $x \in \mathcal{D}$ to be used in the rejection sampler. The improved version of the algorithm is described in Table 2.8.

Table 2.8: Inversion-rejection algorithm for a decreasing pdf $p_o(x)$.

1. Set $i = 1$. Let N be the number of desired samples from $p_o(x)$.

2. Draw a index j' with $\mathrm{Prob}\{j = i\} = F_X(s_{i+1}) - F_X(s_i)$, $i = 1, ..., +\infty$, by inversion.

3. Generate a pair $x' \sim \mathcal{U}([s_{j'}, s_{j'+1}])$ and $u' \sim \mathcal{U}([0, g(x')])$, where $g(x) = p(s_{j'})$ for all $x \in [s_{j'}, s_{j'+1}]$.

3. If $u' \leq p(x')$ then accept $x^{(i)} = x'$ and set $i = i + 1$.

4. Otherwise, if $u' > p(x')$, discard x' and go back to step 2.

5. If $i > N$ then stop, else go back to step 2.

Obviously, the algorithm in Table 2.8 is also a *strip* technique. However, unlike the strip algorithms in Section 2.5.4, this technique can be applied to unbounded pdf's and densities with infinite support. Hence, in this sense the inversion-rejection algorithm can be considered as an extension of the strip methods (when the cdf $F_X(x)$ can be evaluated). It can also be viewed as a numerical inversion method with the addition of a rejection step.

The next technique is also related to the inversion algorithm. It tries to replace a non-invertible cdf $F_X(x)$ with another invertible transformation $h(x)$ (as close as possible to $F_X(x)$).

2.5.6 Transformed rejection method

The transformed rejection method due to [63, 154] is also called *almost exact inversion method* in [30, Chapters 3] and *exact-approximation method* in [106].

This technique consists in transforming a bounded target pdf $p_o(x)$ with an unbounded domain into another bounded pdf with bounded domain. Indeed, as we have already explained, the simplest scenario to use the RS algorithm occurs when the target density is bounded with bounded support.

In this case, it is possible to use a uniform pdf as proposal function, as suggested in [63].

In this section, we study how it is possible to find a transformation in order to modify a pdf of the type displayed in Figure 2.4(a) to obtain a pdf of type depicted in Figure 2.4(c) (bounded with bounded support).

Let $p_o(x)$ be a bounded density with unbounded support $\mathcal{D}$. We assume $\mathcal{D} \equiv \mathbb{R}$ without loss of generality. Moreover, let us consider a monotonic continuous transformation

$$h(x) : \mathbb{R} \to [0, 1].$$

If X is a random variable with pdf $p_o(x)$, then the transformed random variable $Y = h(X)$ has density

$$q(y) = p_o\big(h^{-1}(y)\big) \left| \frac{dh^{-1}}{dy} \right|, \quad \text{for } 0 \leq y \leq 1, \tag{2.63}$$

where $h^{-1}(y)$ is the inverse function of $h(x)$. Clearly, to draw from $p_o(x)$ we can generate a sample y' from $q(y)$ and then take $x' = h^{-1}(y')$. The idea in [154] is to use a RS algorithm to draw from $q(y)$. Choosing an adequate transformation $h(x)$ such that $q(y)$ is also bounded, the advantage of this strategy is that the proposal pdf $\pi(x)$ can be an uniform density (see Figure 2.4 (c)).

Obviously, the domain of $q(y)$ is $\mathcal{C} = [0, 1]$, bounded. However, in general, the density $q(y)$ can be unbounded (it may have vertical asymptotes) depending on the choice of the transformation $h(x)$. Indeed, $q(y) = p_o\big(h^{-1}(y)\big) \left| \frac{dh^{-1}}{dy} \right|$ and, although the first term $p_o\big(h^{-1}(y)\big)$ is bounded (since $p_o(x)$ is assumed bounded), the second term $\left| \frac{dh^{-1}}{dy} \right|$ is, in general, unbounded, i.e.,

$$\lim_{y \to 0} \left| \frac{dh^{-1}}{dy} \right| = \lim_{y \to 1} \left| \frac{dh^{-1}}{dy} \right| = \infty.$$

Figures 2.7 (b) and (c) illustrate some examples of the transformations $h(x)$ and $h^{-1}(y)$.

It is clear from Eq. (2.63) that the density $q(y)$ resulting from the transformation $h(x)$ remains bounded when the tails of $p_o(x)$ decay to zero quickly enough, namely, faster than the derivative $\frac{dh^{-1}}{dy} = \left(\frac{dh}{dx} \right)^{-1}$ diverges when $y \to 0, 1$ or, equivalently, $x \to \pm\infty$.

$$p_o(x) \qquad \mathcal{A}_0$$

$$h(x)$$

$$h^{-1}(y)$$

(a)   (b)   (c)
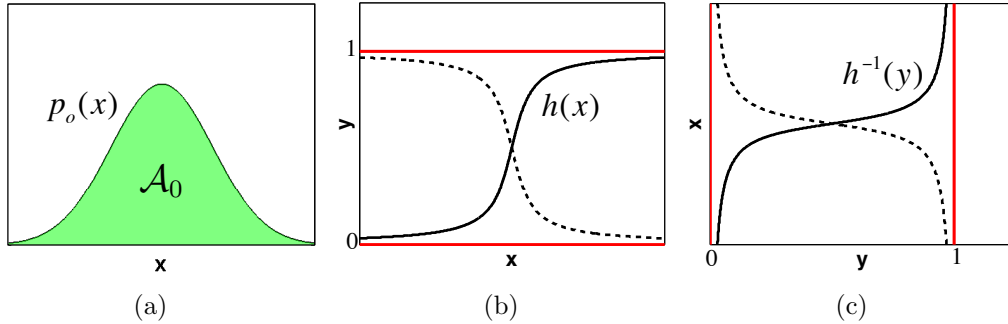
Figure 2.7: **(a)** A bounded target pdf $p_o(x)$ with an unbounded domain. **(b)** Two possible examples (solid and dashed line) of transformation $h(x)$. **(c)** The corresponding inverse transformations $h^{-1}(y)$.

It is also important to realize that better acceptance rates can be obtained by a suitable choice of the transformation function $h(x)$. Indeed, when $h(x)$ is "similar" to the cdf $F_X(x)$, the pdf $q(y)$ becomes flatter and "similar" to a uniform density, so that the acceptance rate is improved. In fact, if $h(x)$ is exactly the cdf of X, i.e., $h(x) = F_X(x)$, then $q(y)$ is the uniform density in $[0, 1]$ (for this reason, this technique is also termed "almost exact inversion method" in [30]).

Table 2.9 summarizes the algorithm.

Table 2.9: Transformed rejection sampling algorithm.

1. Set $i = 1$. Let N be the number of desired samples from $p_o(x)$.

2. Find a suitable monotonic transformation $h(x) : \mathcal{D} \to [0, 1]$.

3. Draw samples y' from $\pi(y)$ and u' from $\mathcal{U}([0, 1])$.

3. If $u' \leq \frac{q(y')}{L\pi(y')}$ then $x^{(i)} = h^{-1}(y')$ and set $i = i + 1$.

4. If $u' > \frac{q(y')}{L\pi(y')}$ discard $h^{-1}(y')$ and go back to step 2.

5. If $i > N$ then stop, else go back to step 2.

Transformed rejection method for unbounded pdf's

A similar methodology can also be applied when the target pdf $p_o(x)$ is unbounded, but has a bounded support $\mathcal{D} = (a, b]$.

Indeed, we can transform $p_o(x)$ into a bounded density with bounded domain $\mathcal{C} = [h(a), h(b)]$, where h is a suitable differentiable and monotonically increasing transformation in $\mathcal{D}$ (so that $h(a) < h(b) < +\infty$). For clarity of exposition, let us assume that $p_o(x)$ has only one asymptote at[3] $x^* = a$, i.e,

$$\lim_{x \to a} p_o(x) = +\infty, \qquad (2.64)$$

as shown, e.g., in Fig. 2.4(b) (where $x^* = 0$). Moreover, let us consider the r.v.'s X with pdf $p_o(x)$ and $Y = h(X)$. We know that Y has density

$$q(y) = p_o\big(h^{-1}(y)\big) \left| \frac{dh^{-1}}{dy} \right|,$$

and, since $h(x)$ is increasing, $h^{-1}(y)$ is also increasing, hence

$$q(y) = p_o\big(h^{-1}(y)\big) \frac{dh^{-1}}{dy}. \qquad (2.65)$$

In general, $q(y)$ is unbounded. Indeed,

$$\lim_{y \to h(a)} q(y) = \lim_{y \to h(a)} p_o\big(h^{-1}(y)\big) \frac{dh^{-1}}{dy} = +\infty \times \lim_{y \to h(a)} \frac{dh^{-1}}{dy}, \qquad (2.66)$$

and setting $L = \lim_{y \to h(a)} \frac{dh^{-1}}{dy}$, we have

$$\lim_{y \to h(a)} q(y) = +\infty \cdot L = \begin{cases} +\infty & \text{if } L \neq 0 \\ \text{undetermined} & \text{if } L = 0 \end{cases}. \qquad (2.67)$$

Therefore, in order to guarantee the applicability of the method it is necessary that we choose $h(x)$ such that $L = 0$. However, this is not sufficient. If we want to ensure that $\lim_{y \to h(a)} q(y) = c$ for some constant c (i.e., $q(y)$ remains bounded) we have to design h in such a way that $\frac{dh^{-1}}{dy}$ vanishes quickly

---

[3] In many cases, the asymptote is located at one of the extreme points of the support $\mathcal{D} = (a, b]$ (as often in the case of inverse density $p_o^{-1}(y)$). For this reason, we choose exactly $x^* = a$.

enough as $y \to h(a)$. Since $\frac{dh^{-1}}{dy} = \frac{dx}{dh}$, this is equivalent to guarantee that the derivative $\frac{dx}{dh}$ grows more slowly than $p_o(x)$ when $x \to a$. In particular, it is sufficient to impose the condition that, $\forall \epsilon > 0$, there exists a constant k such that

$$p_o(x) \geq k \left| \frac{dh}{dx} \right|, \tag{2.68}$$

whenever $|x - a| \leq \epsilon$. When $L = 0$ and the condition (2.68) holds, $q(y)$ is bounded on the bounded domain $[h(a), h(b)]$ and we can apply the procedure of Table 2.9.

Another technique that transforms the unbounded region $\mathcal{A}_0$ below the target pdf $p_o(x)$ into a bounded region $\mathcal{A}$ is the so-called *ratio of uniforms method*, described in Section 2.8. In Section 2.8.2 we explore the relationship between the transformed rejection technique and the ratio of uniforms method.

2.6 Adaptive rejection sampling

The main limitation of RS methods is that it is in general very hard to find a proposal function $\pi(x)$ and a bound $L \geq p(x)/\pi(x)$, such that the overbounding function $L\pi(x) \geq p(x)$ be actually "close" enough to the target density, as needed to attain good acceptance rates. One way to tackle this difficulty is to construct $\pi(x)$ adaptively.

The standard adaptive rejection sampling (ARS) [50], [30, Chapter 7] 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 stringent limitation for many practical applications.

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 standard ARS procedure can be applied when $\log[p(x)]$ is concave, i.e., when the potential function

$$V(x) \triangleq -\log[p(x)], \quad x \in \mathcal{D} \subseteq \mathbb{R}, \tag{2.69}$$

is convex.

The basic idea is to partition the domain $\mathcal{D}$ into several intervals and construct the overbounding function locally on each of these pieces. Let

$$\mathcal{S}_t \triangleq \{s_1, s_2, \ldots, s_{m_t}\} \subset \mathcal{D} \tag{2.70}$$

be a set of support points, sorted in ascending order $s_1 < \ldots < s_{m_t}$. 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 by 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 we can define

$$W_t(x) \triangleq \max\{w_1(x), \ldots, w_{m_t}(x)\} \leq V(x) \quad \forall x \in \mathcal{D}. \tag{2.71}$$

Figure 2.8 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.,

$$\exp\{-W_t(x)\} \geq p(x) = \exp\{-V(x)\}. \tag{2.72}$$

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

$$\pi_t(x) = c_t \exp[-W_t(x)], \tag{2.73}$$

where c_t is the proportionality constant. We can draw from $\pi(x)$ easily. First, we calculate the area ω_k, $k = 0, \ldots, m_t$, below each piece and obtain the normalized weights

$$\bar{\omega}_k = \frac{\omega_k}{\sum_{k=0}^{m_t} \omega_k}. \tag{2.74}$$

Note that the ω_k's can be calculated exactly, as we only need to integrate functions of the form $\exp\{-\lambda x\}$ (for a constant λ) in finite intervals.

In order to draw a sample from $\pi_t(x)$, we randomly choose a piece according to the probability masses $\bar{\omega}_k$, $k = 0, \ldots, m_t$, and then we generate a sample x' from the corresponding truncated exponential pdf using the inverse transform method (see Section 2.2.1). Since Eq. (2.72) can be rewritten as $\frac{1}{c_t}\pi_t(x) \geq p(x)$, we can readily apply the rejection sampling principle.

When a sample x' from $\pi_t(x)$ is rejected we incorporate it into the set of support points, i.e., $\mathcal{S}_{t+1} = S_t \cup \{x'\}$ and $m_{t+1} = m_t + 1$. Then, we compute a refined lower hull, $W_{t+1}(x)$, and a new proposal density

Figure 2.8: **(a)** 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 function $W_t(x) \triangleq \max[w_1(x), w_2(x), w_3(x)]$ is formed by segments of linear functions tangent to the potential $V(x) = x^2$ at the support points in $\mathcal{S}_t$. **(b)** The corresponding overbounding function $\exp\{-W_t(x)\}$ and the target function $p(x) = \exp\{-V(x)\}$.

$\pi_{t+1}(x) = c_{t+1} \exp\{-W_{t+1}(x)\}$ that is closer to the target pdf. Table 2.10 summarizes the ARS algorithm.

If the target density is computationally expensive to evaluate, it is also possible to construct a squeeze function

$$\exp\{-\hat{W}_t(x)\} \le p(x) = \exp\{-V(x)\},$$

for all $x \in \mathcal{D}$ (see Section 2.5.2). In order to construct $\hat{W}_t(x)$ in such a way that it is also piecewise linear, we can use the secant lines passing through the points $(s_k, V(s_k))$ and $(s_{k+1}, V(s_{k+1}))$ where $s_k, s_{k+1} \in \mathcal{S}_t$ are support points. Obviously, as illustrated in Figure 2.9, this construction is possible only in the finite domain $[\min(\mathcal{S}_t), \max(\mathcal{S}_t)]$, hence, we set $\hat{W}_t(x) \to +\infty$ for any $x \notin [\min(\mathcal{S}_t), \max(\mathcal{S}_t)]$. It is straightforward to see that with this construction

$$\exp\{-W_t(x)\} \ge \exp\{-V(x)\} \ge \exp\{-\hat{W}_t(x)\}, \qquad (2.75)$$

so that we can also apply the squeeze technique.

Table 2.10: 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 Figure 2.8, using the tangent lines to $V(x)$ at the support points in $\mathcal{S}_t$.

3. Draw 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 $x^{(i)} = x'$, set $\mathcal{S}_{t+1} = \mathcal{S}_t$, $m_{t+1} = m_t$ and $i = i + 1$.

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.

2.6.1 Derivative-free ARS

A variation of the standard ARS algorithm that avoids the need to compute derivatives of $V(x)$ and lends itself to a simpler automatic implementation has been proposed in [47].

Given the set of support points $\mathcal{S}_t = \{s_1, ..., s_{m_t}\}$, we denote with $w_k(x)$ the secant line passing through the points $(s_k, V(s_k))$ and $(s_{k+1}, V(s_{k+1}))$, for $k = 1, ..., m_t - 1$. Whereas for $k \in \{-1, 0, m_t, m_{t+1}\}$ we set

$$w_k(x) \to -\infty, \tag{2.76}$$

as infinite constant values. The piecewise linear function $W_t(x)$ is constructed as

$$W_t(x) \triangleq \max[w_{k-1}(x), w_{k+1}(x)] \ \ \text{for} \ \ x \in [s_k, s_{k+1}], \ k = 0, ..., m_t. \tag{2.77}$$

Figure 2.10 illustrates the construction of $W_t(x)$ using the derivative-free ARS algorithm with 4 and 5 support points.
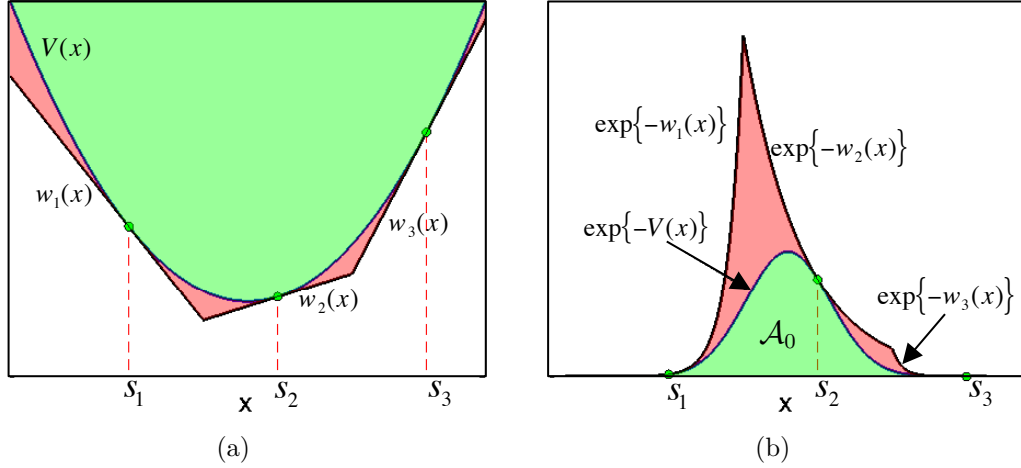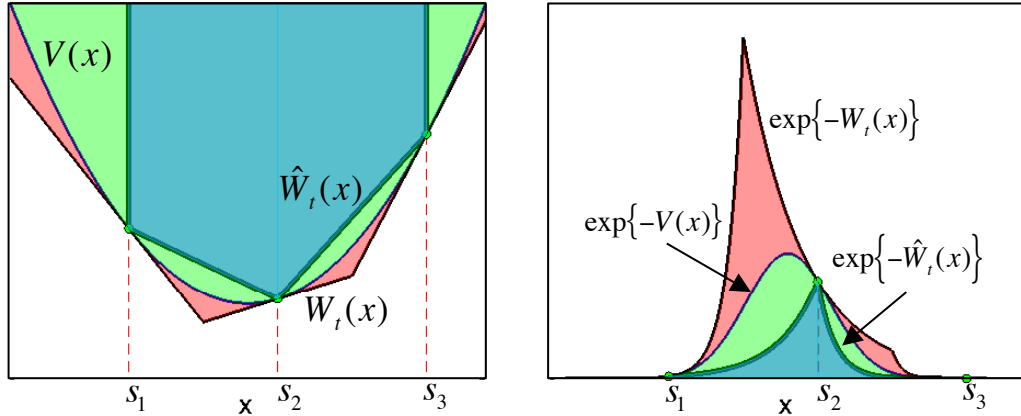
Figure 2.9: **(a)** Example of construction of the squeeze function $\exp\{-\hat{W}_t(x)\}$ with three support points $\mathcal{S}_t = \{s_1, s_2, s_{m_t=3}\}$. **(b)** The corresponding overbounding and squeeze functions $\exp\{-W_t(x)\}$, $\exp\{-\hat{W}_t(x)\}$, together with the target function $p(x) = \exp\{-V(x)\}$

2.7   Generalizations of the ARS algorithm

The condition of $\log[p(x)]$ being concave rules out many target pdf's of interest. Indeed, in many practical applications the target is *non*-log-concave or, in general, multimodal and the standard (or derivative-free) ARS techniques cannot be applied. In order to deal with these densities many generalizations of the standard ARS method have been proposed in the literature.

2.7.1   Adaptive rejection Metropolis sampling

This method, introduced in [48, 146], is a generalization of the standard ARS algorithm that includes a Metropolis-Hastings step and can be applied to any target pdf. Unfortunately, because of the incorporation of MCMC steps, the produced samples are correlated, i.e., they are not statistically independent.

The main idea is relatively simple. We first construct a function $\exp\{-W_t(x)\}$ following the procedure of Section 2.6.1 (derivative-free). Note that, since $V(x)$ is not convex, there is no guarantee that $\exp\{-W_t(x)\} \geq \exp\{-V(x)\}$. Therefore, when a sample drawn from $\pi(x) \propto \exp\{-W_t(x)\}$ is accepted by the rejection sampler, a Metropolis-Hastings control test is added
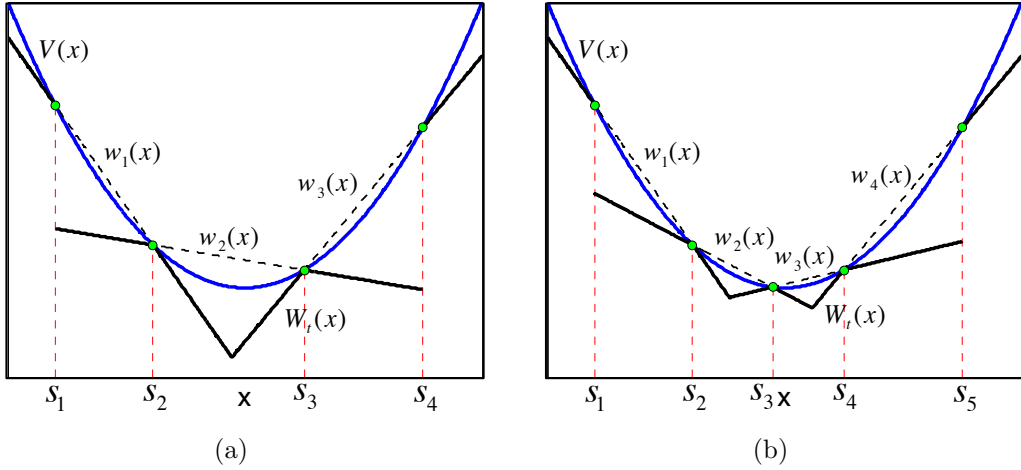
Figure 2.10: Example of construction of the piecewise linear function $W_t(x)$, as carried out by the derivative-free ARS technique. The function $W_t(x)$ is composed by pieces of secant lines passing through $(s_k, V(s_k))$ and $(s_{k+1}, V(s_{k+1}))$, $k = 1, ..., m_t - 1$, as described in Eqs. (2.76) and (2.77). **(a)** Construction with four support points $\mathcal{S}_t = \{s_1, s_2, s_3, s_{m_t=4}\}$. **(b)** Construction with five support points $\mathcal{S}_t = \{s_1, s_2, s_3, s_4, s_{m_t=5}\}$.

to ensure that the sampling is exact. Otherwise, when a sample drawn from $\pi(x) \propto \exp\{-W_t(x)\}$ is rejected, the MCMC step is avoided and the sample is discarded. Table 2.11 summarizes the algorithm.

Note that the Metropolis-Hastings step is strictly necessary for exact sampling. Since it cannot be guaranteed that $\exp\{-W_t(x)\} \geq p(x)$, the rejection sampler generates samples with density proportional to $\min[p(x), \exp\{-W_t(x)\}]$.

In [120] it is suggested to use functions $W_t(x)$ formed by polynomials of degree 2 (parabolic pieces), instead of linear functions (polynomial of degree 1) in order to make a better approximation of the real potential $V(x)$ and improve the acceptance rate.

2.7.2 Concave-convex ARS

The concave-convex ARS method [55, 60] is another generalization of the standard ARS algorithm where the potential $V(x)$ can be decomposed into

54

Table 2.11: Adaptive Rejection Metropolis Sampling Algorithm (ARMS). The generated samples are correlated.

---

1. Start with $i = 1$, $t = 0$, $m_0 = 2$, $\mathcal{S}_0 = \{s_1,\ s_2\}$ where $s_1 < s_2$, and
   Let N be the number of desired samples from $p_o(x)$.
   Choose an arbitrary abscissae x_c.

2. Build the piecewise-linear function $W_t(x)$ as explained in Section 2.6.1.

3. Draw x' from $\pi_t(x) \propto \exp\{-W_t(x)\}$, and u' from $\mathcal{U}([0,1])$.

4. If $u' > \frac{p(x')}{\exp\{-W_t(x')\}}$, then reject x', set $\mathcal{S}_{t+1} = \mathcal{S}_t \cup \{x'\}$ and update
   $m_{t+1} = m_t + 1$. Jump to the step 8.

5. Otherwise, if $u' \leq \frac{p(x')}{\exp\{-W_t(x')\}}$, draw v' from $\mathcal{U}([0,1])$.

6. If $v' \leq \min\left[\frac{p(x')\min[p(x_c),\exp\{-W_t(x_c)\}]}{p(x_c)\min[p(x'),\exp\{-W_t(x')\}]}\right]$ then accept $x^{(i)} = x'$, set $x_c = x'$
   and $\mathcal{S}_{t+1} = \mathcal{S}_t$, $m_{t+1} = m_t$ and $i = i + 1$.

7. If $v' > \min\left[\frac{p(x')\min[p(x_c),\exp\{-W_t(x_c)\}]}{p(x_c)\min[p(x'),\exp\{-W_t(x')\}]}\right]$ then reject x', set $x^{(i)} = x_c$, and
   $\mathcal{S}_{t+1} = \mathcal{S}_t$, $m_{t+1} = m_t$, $i = i + 1$.

8. Sort $\mathcal{S}_{t+1}$ in ascending order and increment $t = t + 1$.
   If $i > N$ then stop, else go back to step 2.

---

a sum of convex, $V_1(x)$, and concave, $V_2(x)$, functions, i.e.,

$$V(x) = V_1(x) + V_2(x). \tag{2.78}$$

The two parts can be analyzed separately in order to obtain two different piecewise linear functions, $W_{t,1}(x)$ and $W_{t,2}(x)$, such that

$$W_{t,i}(x) \leq V_i(x)$$

with $i = 1, 2$. Clearly, the overbounding function in this case is $\exp\{-W_{t,1}(x) - W_{t,2}(x)\} \geq \exp\{-V(x)\}$. Figure 2.11 illustrates the procedures to handle the potentials V_1 and V_2 with different concavity.
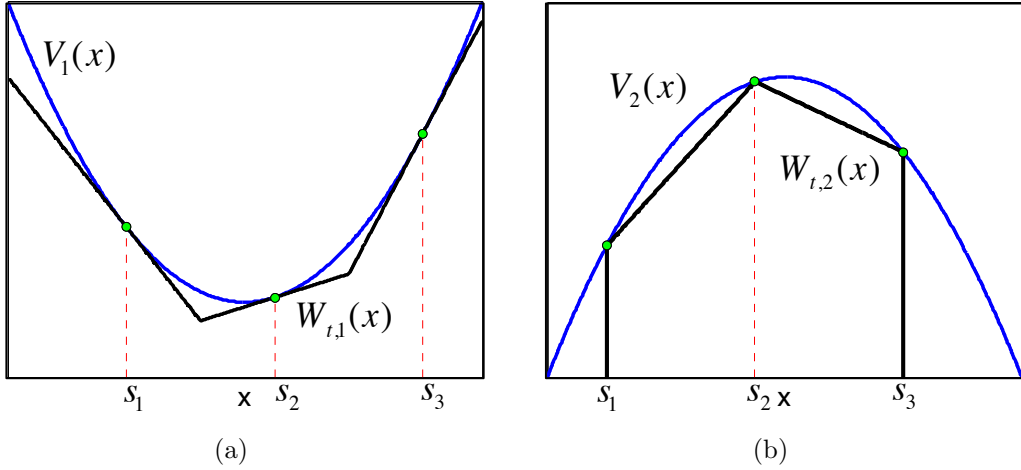
Figure 2.11: Application of the concave-convex ARS method. Example of construction with three support points $\mathcal{S}_t = \{s_1, s_2, s_{m_t=3}\}$ of the two piecewise linear functions $W_{t,i}(x)$, $i = 1, 2$, such that $W_{t,i}(x) \leq V_i(x)$. **(a)** Since $V_1(x)$ is convex, we can use pieces of tangent lines to $V_1(x)$ at the support points. **(b)** Since $V_1(x)$ is concave, the function $W_t(x)$ is composed by pieces of secant lines passing through $(s_k, V(s_k))$ and $(s_{k+1}, V(s_{k+1}))$, $k = 1, 2$, to fulfill the inequality $W_{t,2}(x) \leq V_2(x)$.

Given a set of support points $\mathcal{S}_t \triangleq \{s_1, s_2, \ldots, s_{m_t}\} \subset \mathcal{D}$, sorted in ascending order, we already know that we can use pieces of tangent lines at the support points to build a lowerbounding function $W_{t,1}(x)$ for the concave potential $V_1(x)$ as in the standard ARS technique.

For the concave potential $V_2(x)$, we can use the secant lines passing through $(s_k, V(s_k))$ and $(s_{k+1}, V(s_{k+1}))$, $k = 1, ..., m_t-1$ to obtain a piecewise linear function $W_{t,2}(x)$ such that $W_{t,2}(x) \leq V_2(x)$. It should be noticed that this procedure is possible only in a finite domain, precisely in the interval $[\min(\mathcal{S}_t), \max(\mathcal{S}_t)]$. Therefore, we can apply this technique to a target pdf with unbounded domain only if the tails of the entire potential $V(x)$ are convex. Indeed, in this situation we can handle the tails separately using tangent lines in order to build a lowerbounding function, as in the standard ARS method.

2.7.3 Transformed density rejection

In [61, 65, 94] it is suggested to replace the $\log(\vartheta)$ function with another monotonically increasing transformation $T(\vartheta) : \mathbb{R}^+ \to \mathbb{R}$ such that $T[p(x)]$ (with $p_o(x) \propto p(x)$) is concave or, equivalently, the corresponding potential function

$$V_T(x) \triangleq -T[p(x)], \tag{2.79}$$

is convex. It is important to notice that this method extends the standard ARS algorithm in [50] but it can be applied to unimodal target densities only, since T is a monotonic function. Eq. (2.79) above implies that the target pdf can be expressed as

$$p_o(x) \propto p(x) = T^{-1}[-V_T(x)]. \tag{2.80}$$

Obviously, we can go back to the standard ARS method by choosing $T(\vartheta) = \log(\vartheta)$.

Given a set of support points $\mathcal{S}_t \triangleq \{s_1, s_2, \ldots, s_{m_t}\} \subset \mathcal{D}$, the idea is, again, to replace the convex potential $V_T(x)$ with a piecewise-linear function $W_t(x)$, such that $W_t(x) \leq V_T(x)$ and formed by segments of linear functions tangent to $V_T(x)$ at the support points $s_k \in \mathcal{S}_t$. If we let $w_k(x)$ be the linear function tangent to $V_T(x)$ at s_k, then the piecewise linear function $W_t(x)$ is defined as $W_t(x) \triangleq \max\{w_1(x), \ldots, w_{m_t}(x)\}$, exactly as in the standard ARS method. Hence, the proposal pdf has the form

$$\pi_t(x) \propto T^{-1}[-W_t(x)]. \tag{2.81}$$

For this procedure, the key is the identification of an adequate transformation $T(\vartheta)$. To be useful, T has to satisfy the following conditions:

1. It has to be monotonically increasing.

2. Given the inverse transformation T^{-1}, the integral $\int_{-\infty}^{\vartheta} T^{-1}(t)dt$ must be bounded for all fixed values of ϑ in the image of T. Moreover, it must be possible to calculate analytically the integral

$$\int_a^b T^{-1}(t)dt. \tag{2.82}$$

3. The composition $(T \circ p)(x) = T[p(x)]$ has to be concave, i.e.,

$$\frac{d^2}{dx^2}T[p(x)] = \left[\frac{d^2T}{d\vartheta^2}\right]_{p(x)}\left(\frac{dp}{dx}\right)^2 + \left[\frac{dT}{d\vartheta}\right]_{p(x)}\frac{d^2p}{dx^2} \leq 0.$$

57

The satisfaction of the first condition guarantees that the inverse transformation $T^{-1} : \mathbb{R}^+ \to \mathbb{R}$ exists and is monotonically increasing as well. The second condition is required to ensure that the integral of the overbounding function $T^{-1}[-W_t(x)]$ in a bounded domain is finite, i.e.,

$$\int_a^b T^{-1}[-W_t(x)]dx \leq +\infty,$$

and we are able to calculate it analytically (recall that $W_t(x)$ is a piecewise linear function). If the transformation T fulfills this condition, we can use the inverse method to draw independent samples from $\pi_t(x)$. The third condition is necessary to allow the construction of $W_t(x)$ using tangent lines such that $W_t(x) \leq V_T(x)$ and, correspondingly, $T^{-1}[-W_t(x)] \geq p(x)$.

An example of a class of transformations with these properties, different from the logarithm, is the the family of power functions $T_c(\vartheta) = \text{sign}(c)\vartheta^c$. The most used transformation of this class is $T_{-1/2}(\vartheta) = -1/\sqrt{\vartheta}$.

Table 2.12 summarizes the technique.

Table 2.12: Transformed Density Rejection (TDR) Algorithm.

1. Start with $i = 1$, $t = 0$, $m_0 = 2$, $\mathcal{S}_0 = \{s_1, \ s_2\}$ where $s_1 < s_2$.
   Let N be the number of desired samples from $p_o(x)$.

2. Build the piecewise-linear function $W_t(x)$ using the lines tangent to $V_T(x) = -T[p(x)]$ at the support points in $\mathcal{S}_t$.

3. Draw x' from $\pi_t(x) \propto T^{-1}[-W_t(x)]$, and u' from $\mathcal{U}([0, 1])$.

4. If $u' \leq \frac{p(x')}{T^{-1}[-W_t(x')]}$, then accept $x^{(i)} = x'$, set $\mathcal{S}_{t+1} = \mathcal{S}_t$, $m_{t+1} = m_t$ and $i = i + 1$.

5. Otherwise, if $u' > \frac{p(x')}{T^{-1}[-W_t(x')]}$, 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.

We remark that, since $T(\vartheta)$ has to be a monotonic function and $T[p(x)]$

has to be concave, this procedure can be applied only to unimodal target pdf's. Let us also mention that, despite the similarity in the names, the transformation in the "Transformed rejection method" of Section 2.5.6 is applied to the random variable X, with pdf $p_o(x)$, while here the transformation is applied directly to the density $p_o(x)$.

Extensions

The TDR method is not necessarily restricted to the case in which T is monotonically increasing and V_T is convex, although it was originally proposed [61] in this setup.

Indeed, we can consider combinations of increasing and decreasing functions T with concave or convex potentials V_T. The procedure to construct the piecewise linear function $W_t(x)$, however, is different depending on the type of T and V_T at hand. This is briefly analyzed in this section.

Let us recall Eq. (2.80), $p_o(x) \propto p(x) = T^{-1}[-V_T(x)]$. So far, we have considered the combination of

1) a monotonically increasing function $\vartheta = T^{-1}(z)$,

2) with a concave function $z = -V_T(x)$.

In this case, a piecewise linear function $z = -W_t(x)$ formed by straight lines tangent to $z = -V_T(x)$ can be used to construct an overbounding function $T^{-1}[-W_t(x)]$ such that

$$T^{-1}[-W_t(x)] \geq T^{-1}[-V_T(x)].$$

The other cases of interest, depending on to the choice of T and the concavity of V_T, are listed below and summarized in Table 2.13.

- If $\vartheta = T^{-1}(z)$ is increasing and $z = -V_T(x)$ is convex, a suitable function $z = -W_t(x)$ can be constructed using secant lines. However, the construction is possible only in a bounded domain (see Figures 2.12(a) and 2.12(c)).

- If $\vartheta = T^{-1}(z)$ is decreasing and $z = -V_T(x)$ is concave, an adequate function $z = -W_t(x)$ has to be formed by secant lines. Also in this case, the construction is possible only in a bounded domain (see Figures 2.12(d) and 2.12(e)).
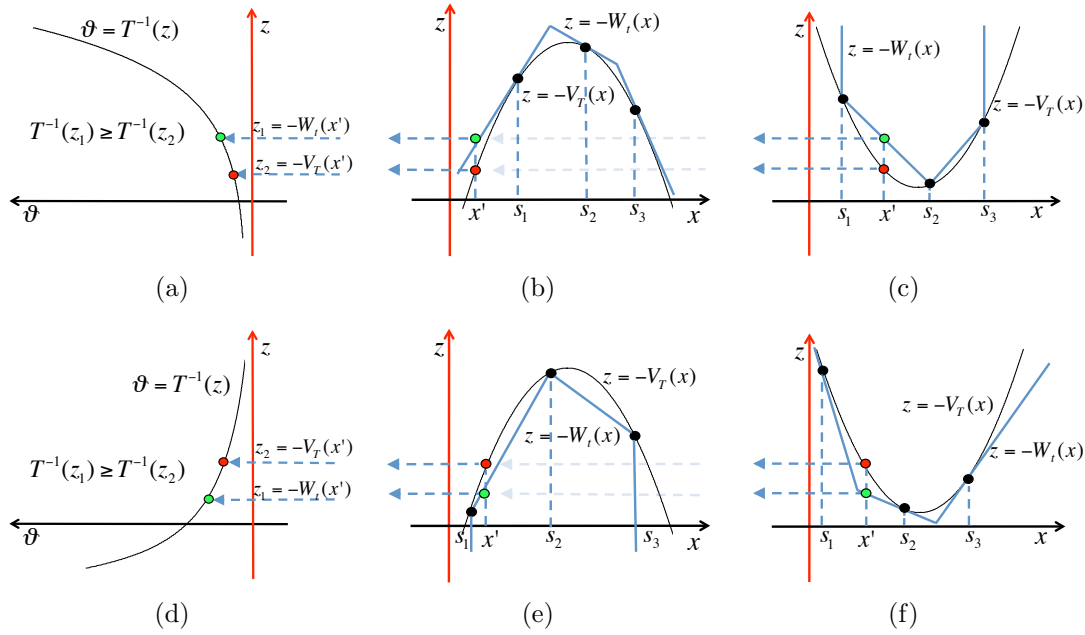
Figure 2.12: Example of construction of an adequate piecewise linear function $z = -W_t(x)$, with three support points $\mathcal{S}_t = \{s_1, s_2, s_3\}$, for different cases. Figures **(a)**-**(b)**-**(c)** consider an increasing function $\vartheta = T^{-1}(z)$, hence, $W_t(x)$ is built to ensure $-W_t(x) \geq -V_T(x)$. Figures **(d)**-**(e)**-**(f)** depict a decreasing function $\vartheta = T^{-1}(z)$, hence we need to build $W_t(x)$ to ensure that $-W_t(x) \leq -V_T(x)$. Given an arbitrary x', it is possible to see that the value $T^{-1}[-W_t(x')]$ (green point) is always greater than $T^{-1}[-V_T(x')]$ (red point), i.e., $T^{-1}[-W_t(x')] \geq T^{-1}[-V_T(x')]$. Note that the axes associated to the independent variable z in Figures **(a)**-**(d)** are vertical.

- If $\vartheta = T^{-1}(z)$ is decreasing and $z = -V_T(x)$ is convex, the straight lines tangent to $z = -V_T(x)$ can also be used to build $z = -W_t(x)$. The construction is also possible in an infinite domain (see Figures 2.12(d) and 2.12(f)).

Potential functions with known inflection points

Consider a target density $p_o(x)$ that can be written as

$$p_o(x) \propto p(x) = T^{-1}[-V_T(x)],$$

60

Table 2.13: Possible combinations.

$\vartheta = T^{-1}(z)$	$z = -V_T(x)$	$z = -W_t(x)$	Domain	Figure 2.12
increasing	concave	tangent lines	unbounded	**(a)**,**(b)**
increasing	convex	secant lines	bounded	**(a)**,**(c)**
decreasing	convex	tangent lines	unbounded	**(d)**,**(f)**
decreasing	concave	secant lines	bounded	**(d)**,**(e)**

where the potential $V_T(x)$ can be non-convex (it may present several minima) but we assume that the positions of all its inflection points are known. In this case, we can extend the TDR procedure.

Indeed, we can find a partition of the support $\mathcal{D} = \cup_{i=1}^{n}\mathcal{D}_i$, $\mathcal{D}_i \cap \mathcal{D}_j = \emptyset$, $i \neq j$, where in each $\mathcal{D}_i$ the function $V_T(x)$ has a second derivative with constant sign. Therefore, in each interval $\mathcal{D}_i$ where $V_T(x)$ is convex, we use tangent lines to build $W_t(x)$. Alternatively, if $V_T(x)$ is concave in $\mathcal{D}_i$, the function $W_t(x)$ is composed by secant lines.

Clearly, this procedure can be applied to non-convex potentials $V_T(x)$ and multimodal target pdf's. In general, however, for complicated target densities it is not straightforward to study analytically the second derivative of the potential $V_T(x)$. Furthermore, even if the inflection points are known, we need that the tails of the potential be convex, as in Section 2.7.2, in order to build a proper proposal.

Recently, another approach has been studied [14] that requires only knowledge of an interval where an inflection point is located, but not exactly the position of the inflection point. To apply this method the potential has to be three-times differentiable, though, and it can be used only with target pdf's with bounded domain.

2.8 Ratio of uniforms method

The standard ratio of uniforms (RoU) method [79, 122] is a sampling technique that relies on the following result.

Theorem 5 *Let $p(x) \geq 0$ be a pdf known only up to a proportionality constant ($p(x) \propto p_o(x)$). If (v, u) is a sample drawn from the uniform*

61

distribution on the set

$$\mathcal{A} = \left\{ (v, u) : 0 \le u \le \sqrt{p(v/u)} \right\}, \tag{2.83}$$

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

Proof: Given the transformation $(v, u) \to (x, y)$

$$\begin{cases} x = \dfrac{v}{u} \\ y = u \end{cases} \longrightarrow \begin{cases} v = xy \\ u = y \end{cases}, \tag{2.84}$$

and a pair of r.v.'s (V, U) uniformly distributed on $\mathcal{A}$, we can write the joint pdf $q(x, y)$ of the transformed r.v.'s (X, Y) as

$$q(x, y) = \frac{1}{|\mathcal{A}|} |J^{-1}| \quad \text{for all} \quad 0 \le y \le \sqrt{p(x)}, \tag{2.85}$$

where we have indicated with $|\mathcal{A}|$ the area of $\mathcal{A}$, and J^{-1} is the Jacobian of the inverse transformation, i.e.,

$$J^{-1} = \det \begin{bmatrix} y & x \\ 0 & 1 \end{bmatrix} = y. \tag{2.86}$$

Substituting (2.86) into (2.85) yields

$$q(x, y) = \begin{cases} \dfrac{1}{|\mathcal{A}|} y, & 0 \le y \le \sqrt{p(x)}, \\ 0, & \text{otherwise.} \end{cases} \tag{2.87}$$

Then, the marginal density of the r.v. X obtained by integrating the pdf $q(x, y)$ coincides with $p_o(x)$. Indeed,

$$\begin{aligned} \int_{-\infty}^{+\infty} q(x, y) dy &= \int_0^{\sqrt{p(x)}} \frac{y}{|\mathcal{A}|} dy = \\ &= \frac{1}{|\mathcal{A}|} \left[\frac{y^2}{2} \right]_0^{\sqrt{p(x)}} = \frac{1}{2|\mathcal{A}|} p(x) = p_o(x), \end{aligned} \tag{2.88}$$

where the first equality follows from Eq. (2.87) and the rest of the calculations are straightforward. $\square$

Furthermore, from Eq. (2.88) we can see that $\frac{1}{2|\mathcal{A}|}$ is the normalization constant of $p(x)$. Therefore, if we denote $c = \int_{\mathcal{D}} p(x)dx$, we have $\frac{1}{2|\mathcal{A}|} = \frac{1}{c}$ and we can obtain the measure of the region $\mathcal{A}$ as

$$|\mathcal{A}| = \frac{c}{2} = \frac{\int_{\mathcal{D}} p(x)dx}{2}. \tag{2.89}$$

In the case that $p(x) = p_o(x)$, then $c = 1$ and the area of $\mathcal{A}$ is $|\mathcal{A}| = 1/2$.

This theorem provides the means to draw from $p_o(x)$. Indeed, if we are able to draw uniformly a point (v', u') from $\mathcal{A}$, then the sample $x' = v'/u'$ is distributed according to $p_o(x)$. Therefore, the efficiency of the RoU method depends on the ease with which we can generate points uniformly within the region $\mathcal{A}$.

The cases of practical interest are those in which the region $\mathcal{A}$ is bounded. The set $\mathcal{A}$ is bounded if, and only if, both $\sqrt{p(x)}$ and $x\sqrt{p(x)}$ are bounded. Furthermore, the function $\sqrt{p(x)}$ is bounded if, and only if, the target density $p_o(x) \propto p(x)$ is bounded and, the function $x\sqrt{p(x)}$ is bounded if, and only if, the tails of $p(x)$ decay as $1/x^2$ or faster. However, some generalizations [26, 153] defining different type of regions $\mathcal{A}$ can be used for heavier tails (see Section 2.8.1).

Figure 2.13 (a) depicts a bounded two-dimensional 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 α yields the same value of x. From the definition of $\mathcal{A}$, it follows that $u_i \le p(x)$ and $v_i = xu_i \le x\sqrt{p(x)}$, hence, if we choose the point (v_2, u_2) that lies on the boundary of $\mathcal{A}$, we obtain that $u_2 = \sqrt{p(x)}$ and $v_2 = x\sqrt{p(x)}$. If the suprema of $\sqrt{p(x)}$ and $x\sqrt{p(x)}$, as well as the infima of $x\sqrt{p(x)}$, can be found, the we can embed the set $\mathcal{A}$ in the rectangular region

$$\mathcal{R} = \left\{ (v, u) : 0 \le u \le \sup_x \sqrt{p(x)}, \ \inf_x x\sqrt{p(x)} \le v \le \sup_x x\sqrt{p(x)} \right\}, \tag{2.90}$$

as depicted in Fig. 2.13(b).

Once the rectangle $\mathcal{R}$ 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 $\mathcal{A}$. Note that in this rejection procedure we do not need to know the analytical expression of the boundary of the region $\mathcal{A}$. Indeed, Eq. (2.83) provides a way to check
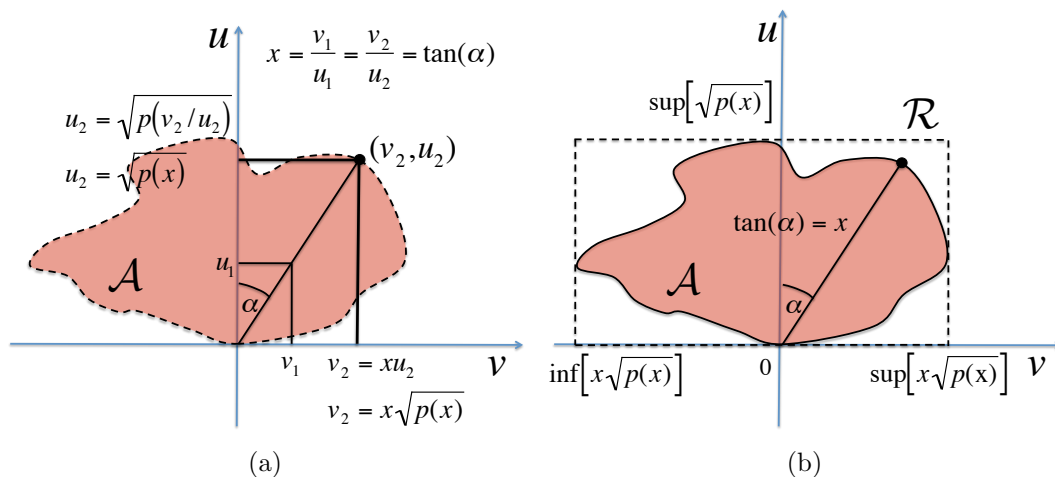
Figure 2.13: **(a)** A bounded region $\mathcal{A}$ and the straight line $v = xu$ corresponding to the sample $x = \tan(\alpha)$. Every point in the intersection of the line $v = xu$ and the set $\mathcal{A}$ yields the same sample x. The point on the boundary, (v_2, u_2), has coordinates $v_2 = x\sqrt{p(x)}$ and $u_2 = \sqrt{p(x)}$. **(b)** If the two functions $\sqrt{p(x)}$ and $x\sqrt{p(x)}$ are bounded, the set $\mathcal{A}$ is bounded and embedded in the rectangle $\mathcal{R}$.

whether a point (v, u) falls inside $\mathcal{A}$ or not. Table 2.14 summarizes this simple accept/reject scheme.

Figures 2.14 (b) and (d) provide two examples in which the region $\mathcal{A}$ corresponds to standard Gaussian and Cauchy densities (shown in Figures 2.14 (a) and (c), respectively). The pictures also illustrate different lines corresponding to x constant (dotted line), u constant (dashed line), v constant (solid line) in the domain $x - u$ and in the transformed domain $v - u$.

In some cases the equation $u = \sqrt{p(v/u)}$ can be solved analytically and the boundary $\mathcal{A}$ can be found explicitly. If we assume a monotonic function $p(x)$, and indicate with p^{-1} its inverse function, the boundary can be expressed with the equation

$$v = up^{-1}(u^2). \tag{2.91}$$

In particular, when

$$p_o(x) \propto \frac{\lambda^2}{(\delta x + \beta)^2} \tag{2.92}$$

64

Table 2.14: Rejection via RoU method.

1. Start with $j = 1$.

2. Construct the rectangle $\mathcal{R} \supseteq \mathcal{A}$.

3. Draw a point (v', u') uniformly from the rectangular region $\mathcal{R}$.

4. If $u' \leq \sqrt{p(v'/u')}$, then accept the sample $x^{(j)} = x' = \frac{v'}{u'}$ and set $j = j + 1$.

5. Otherwise, if $u' > \sqrt{p(v'/u')}$, then reject the sample $x' = \frac{v'}{u'}$,

6. If $j > N$ then stop, else go back to step 2.

with λ, δ, β, and a compact support, $x \in [a, b]$, the region $\mathcal{A}$ is a triangle, as depicted in Fig. 2.15(a), with one vertex at the origin, $\mathbf{v}_1 = (0, 0)$, and the opposite side, $\mathbf{v}_2 - \mathbf{v}_3$, given by the equation $\delta v + \beta u = \lambda$. Figure 2.15(b) illustrates the particular case with $\delta = 0$, when $p_o(x)$ becomes a uniform distribution and we obtain a triangular region with the side $\mathbf{v}_2 - \mathbf{v}_3$ parallel to the axis v. Moreover, if $\beta = 0$ the pdf $p_o(x) \propto \frac{1}{x^2}$, $x \in [a, b]$, is called *reciprocal* uniform density (because we can obtain it by taking the reciprocal of a uniform random variable U, i.e., $1/U$) and the corresponding region $\mathcal{A}$ is triangular with $\mathbf{v}_1 = (0, 0)$ and the side $\mathbf{v}_2 - \mathbf{v}_3$ parallel to the axis u, as shown in Figure 2.15(c).

Another example, for which $\mathcal{A}$ has a closed form occurs for the so-called *table mountain density* [44, 64]. In particular, if $\mathcal{A} = [b^-, b^+] \times [0, a]$ is a rectangular region in the $v - u$ domain, then the associated pdf is

$$q(x) \propto \begin{cases} (b^-)^2/x^2 & \text{for} \quad x \in (-\infty, b^-/a] \\ a^2 & \text{for} \quad x \in [b^-/a, b^+/a], \\ (b^+)^2/x^2 & \text{for} \quad x \in [b^+/a, +\infty) \end{cases} \tag{2.93}$$

plotted in Fig. 2.16(a) (up to a proportionally constant). If we divide the rectangular region $\mathcal{R}$ in three non-overlapping triangular parts, $\mathcal{R} = \mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T}_3$ as illustrated in Figure 2.16(b), then we can see that each part of $q(x)$ is related to each triangular part $\mathcal{T}_i$, $i = 1, 2, 3$, by comparing Figures
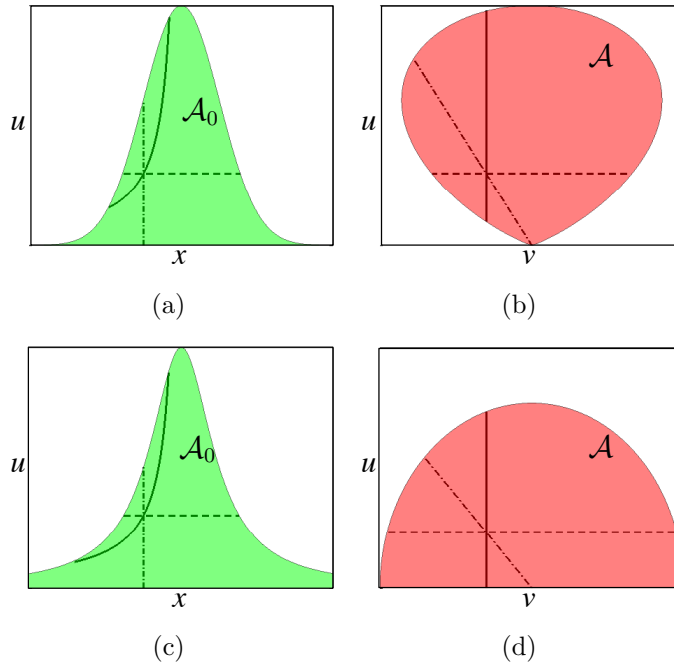
Figure 2.14: Examples of the regions $\mathcal{A}$ obtained applying the RoU transformation to standard Gaussian and Cauchy pdf's. Each figure shows lines corresponding to x constant (dotted line), u constant (dashed line) and v constant (solid line). **(a)** A standard Gaussian density $p_o(x) \propto \exp\{-x^2/2\}$. **(b)** The region $\mathcal{A}$ corresponding to a standard Gaussian pdf. **(c)** A standard Cauchy density $p_o(x) \propto 1/(1 + x^2)$. **(d)** The region $\mathcal{A}$ corresponding to a standard Cauchy pdf. It is a semi-circle with radius 1 and center in $(0,0)$.

2.16(a) and 2.16(b).

2.8.1 Generalized ratio of uniforms method

A more general version of the standard RoU method can be established with the following theorem [153].

Theorem 6 *Let $g(u)$ be a strictly increasing differentiable function on $\mathbb{R}^+$ such that $g(0) = 0$ and let $p(x) \geq 0$ be a pdf known only up to a proportionality constant. Assume that (u, v) is a sample drawn from the uniform distribution*
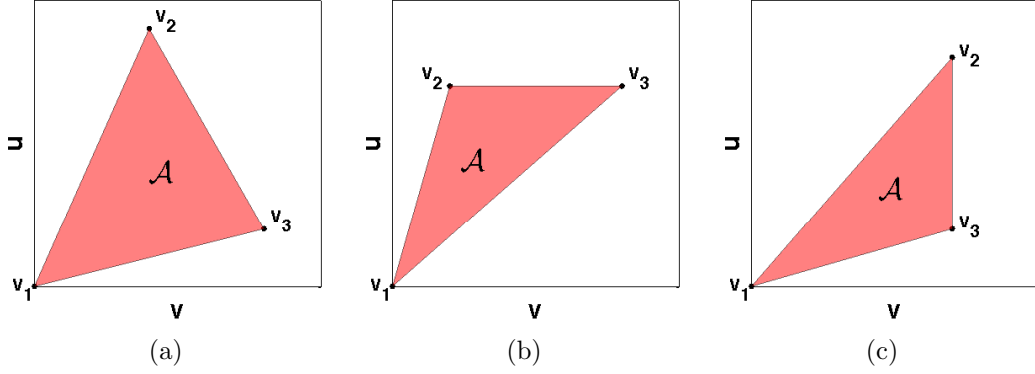
Figure 2.15: **(a)** A triangular region $\mathcal{A}$ with a vertex at the origin $\mathbf{v}_1 = (0,0)$ and where the side $\mathbf{v}_2 - \mathbf{v}_3$ has a generic slope. It corresponds to a density of the form $p_o(x) \propto 1/(\delta x + \beta)^2$ transformed via the RoU method. **(b)** A triangular region $\mathcal{A}$ obtained by transforming a uniform pdf by the RoU method. The side $\mathbf{v}_2 - \mathbf{v}_3$ is parallel to the axis v. **(c)** A triangular region $\mathcal{A}$ obtained transforming a reciprocal uniform pdf $p_o(x) \propto 1/x^2$, $x \in [a,b]$, by the RoU method. The side $\mathbf{v}_2 - \mathbf{v}_3$ is parallel to the axis u.

on the set

$$\mathcal{A}_g = \left\{ (v, u) : 0 \le u \le g^{-1}\left[c\, p\left(\frac{v}{\dot{g}(u)} \right) \right] \right\}, \tag{2.94}$$

where $c > 0$ is a positive constant and $\dot{g} = \frac{dg}{du}$. Then $x = \frac{v}{\dot{g}(u)}$ is a sample from $p_o(x)$.

Proof: The argument is the same used for the standard RoU method (see [79, 153]). Given the transformation $(v, u) \to (x, z)$

$$\begin{cases} x = \dfrac{v}{\dot{g}(u)} \\ z = u \end{cases} \longrightarrow \begin{cases} v = x\dot{g}(z) \\ u = z \end{cases}, \tag{2.95}$$

and a pair of r.v.'s (V, U) uniformly distributed on $\mathcal{A}_g$, we can write the joint pdf $q(x, y)$ of the transformed r.v.'s (X, Z) as

$$q(x, z) = \frac{1}{|\mathcal{A}_g|} |J^{-1}| \quad \text{for all} \quad 0 \le z \le g^{-1}[cp(x)], \tag{2.96}$$

67
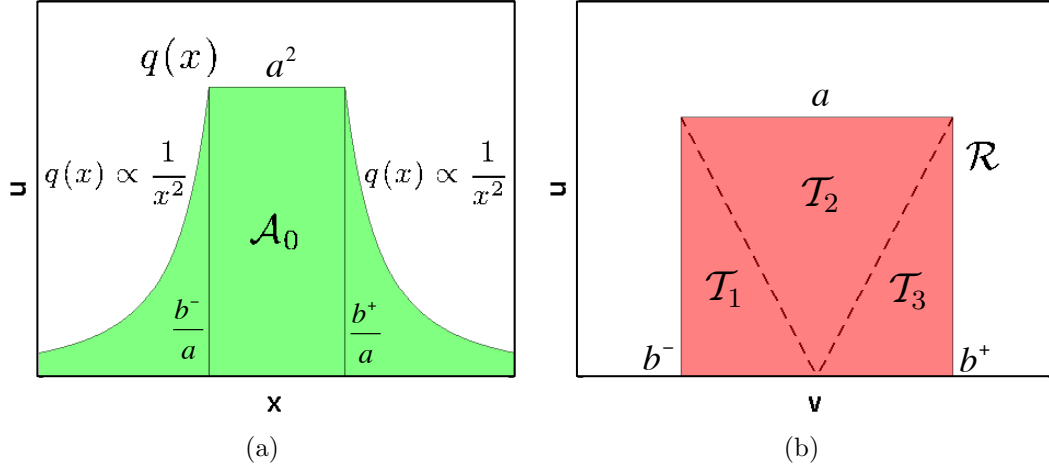
Figure 2.16: **(a)** The shape of a table mountain density $q(x)$ defined in Eq. (2.93). **(b)** The region $\mathcal{A}$ obtained with the RoU transformation of the table mountain density is a rectangle, i.e., $\mathcal{A} = \mathcal{R}$. The rectangular region $\mathcal{R}$ can be divided in three non-overlapping triangular parts $\mathcal{R} = \mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T}_3$.

where $|\mathcal{A}_g|$ denotes the area of $\mathcal{A}_g$, and J^{-1} is the Jacobian of the inverse transformation, namely,

$$J^{-1} = \det \begin{bmatrix} \dot{g}(z) & x\ddot{g}(z) \\ 0 & 1 \end{bmatrix} = \dot{g}(z). \tag{2.97}$$

Substituting (2.97) into (2.96) yields

$$q(x,z) = \begin{cases} \dfrac{1}{|\mathcal{A}_g|} \dot{g}(z) & \text{for} \quad 0 \leq z \leq g^{-1}[cp(x)], \\ 0, & \text{otherwise.} \end{cases} \tag{2.98}$$

Hence, integrating $q(x,z)$ w.r.t. z yields the marginal pdf of the r.v. X,

$$\int_{-\infty}^{+\infty} q(x,z)dz = \int_0^{g^{-1}[cp(x)]} \frac{1}{|\mathcal{A}_g|} \dot{g}(z) dz =$$

$$= \frac{1}{|\mathcal{A}_g|} \Big[g(z) \Big]_0^{g^{-1}[cp(x)]} = \frac{c}{|\mathcal{A}_g|} p(x) - \frac{1}{|\mathcal{A}_g|} g(0) = q(x)$$

where the first equality follows from Eq. (2.98) and the remaining calculations are trivial. Since we have also assumed $g(0) = 0$, it turns out

68

that

$$q(x) = \frac{c}{|\mathcal{A}_g|} p(x) = p_o(x). \quad \square$$

In the boundary of the region $\mathcal{A}_g$ we have $u = g^{-1}[cp(x)]$ and, since $v = x\dot{g}(u)$, we also have $v = x\dot{g}[g^{-1}(cp(x))]$. The contour of $\mathcal{A}_g$ is described parametrically by the latter two equations.

Furthermore, if the two functions $g^{-1}[cp(x)]$ and $x\dot{g}[g^{-1}(cp(x))]$ are bounded, the region $\mathcal{A}_g$ is embedded in the rectangular region

$$\mathcal{R}_g = \Big\{ (v, u) : \ 0 \le u \le \sup_x g^{-1}[cp(x)],$$
$$\inf_x x\dot{g}[g^{-1}(cp(x))] \le v \le \sup_x x\dot{g}[g^{-1}(cp(x))] \Big\}. \tag{2.99}$$

In the sequel, we obtain the conditions that the function $s = g(u)$ has to satisfy in order that $g^{-1}[cp(x)]$ and $x\dot{g}[g^{-1}(cp(x))]$ be bounded.

Since g is a monotonic and continuous function, $u = g^{-1}[cp(x)]$ is bounded if, and only if, $p(x)$ is bounded. Moreover, the function $v = x\dot{g}[g^{-1}(cp(x))]$ is bounded if $p(x)$ is bounded and if the limits

$$\lim_{x \to +\infty} x\dot{g}[g^{-1}(cp(x))] = L_1, \tag{2.100}$$

$$\lim_{x \to -\infty} x\dot{g}[g^{-1}(cp(x))] = L_2, \tag{2.101}$$

are both finite, $L_1, L_2 \le +\infty$. To obtain (2.100) and (2.101) we need that

1. the limits of the composition $\dot{g} \circ g^{-1} \circ cp$ vanish as $x \to \pm\infty$,

$$\lim_{x \to \pm\infty} \dot{g}[g^{-1}(cp(x))] = 0, \tag{2.102}$$

2. and, $\dot{g}[g^{-1}(cp(x))]$ decays to zero faster than x diverges to infinity.

Since $\lim_{x \to \pm\infty} p(x) = 0$ and $g^{-1}(0) = 0$ (recall that we have assumed $g(0) = 0$), the first condition is satisfied if

$$\lim_{u \to 0} \dot{g}(u) = 0. \tag{2.103}$$

Moreover, in order that $\dot{g}[g^{-1}(cp(x))]$ decays to zero faster than x diverges to infinity, the tails of $p(x)$ have to decay to zero as

$$\frac{1}{c} g\left[\dot{g}^{-1}\left(\frac{1}{x}\right) \right], \tag{2.104}$$

69

or faster (note that $\dot{g}^{-1}$ denotes the inverse function of the derivative $\frac{dg}{du}$). Furthermore, given $y = g(u)$ and since we know the relationship between the derivatives

$$w = \dot{g}(u) = \frac{1}{\frac{dg^{-1}}{dy}(g(u))},$$

we can obtain (inverting both sides) $u = \dot{g}^{-1}(w)$ and $u = g^{-1}\left[\left(\frac{dg^{-1}}{dy}\right)^{-1}\left(\frac{1}{w}\right)\right]$

where $\left(\frac{dg^{-1}}{dy}\right)^{-1}$ represents the inverse function of $\frac{dg^{-1}}{dy}$. Hence, we can also write

$$\dot{g}^{-1}(w) = g^{-1}\left[\left(\frac{dg^{-1}}{dy}\right)^{-1}\left(\frac{1}{w}\right)\right], \qquad (2.105)$$

and with the change of variable $w = 1/x$ we obtain

$$\dot{g}^{-1}\left(\frac{1}{x}\right) = g^{-1}\left[\left(\frac{dg^{-1}}{dy}\right)^{-1}(x)\right]. \qquad (2.106)$$

Substituting (2.106) into (2.104), we find that the tails of $p(x)$ have to decay to zero as

$$\frac{1}{c}\left(\frac{dg^{-1}}{dy}\right)^{-1}(x). \qquad (2.107)$$

Finally, the region $\mathcal{A}_g$ is bounded if

1. $\lim_{u \to 0} \frac{dg}{du} = 0$,

2. and the tails of $p(x)$ decays to zero as or faster than $\frac{1}{c}\left(\frac{dg^{-1}}{dy}\right)^{-1}(x)$.

Other generalizations of the RoU method can be found in the literature [75]. Further development involving ratio of r.v.'s can be found in [8, 27, 31, 105, 127, 141, 150]

Transformation using power functions

A suitable family of transformations $g(u)$ are the power functions [26], i.e.,

$$g(u) = \frac{u^{r+1}}{(r+1)}, \qquad u \geq 0, \qquad (2.108)$$

70

with $r \geq 0$ and $c = 1/(r+1)$. Note that the first derivative $\dot{g}(u)$ is strictly increasing for $u \geq 0$.

The region $\mathcal{A}_g$ defined in Equation (2.94) becomes

$$\mathcal{A}_g = \mathcal{A}_r = \left\{ (v,u) : 0 \leq u \leq \left[p(v/u^r) \right]^{1/(r+1)} \right\}, \qquad (2.109)$$

that we denote $\mathcal{A}_r$, since with $r = 1$ we obtain the set of the standard RoU method in Eq. (2.83) and the region $\mathcal{A}_0$ (delimited by the pdf $p_o(x)$, see Figure 2.1) defined in Eq. (2.29) with $r = 0$. The bounding rectangular region is defined in this case as

$$\mathcal{R}_r = \left\{ (v',u') : 0 \leq u' \leq \sup_x [p(x)]^{1/(r+1)}, \right.$$
$$\left. \inf_x x[p(x)]^{r/(r+1)} \leq v' \leq \sup_x x[p(x)]^{r/(r+1)} \right\}. \qquad (2.110)$$

In other words, the region $\mathcal{A}_r$ is bounded if the functions $[p(x)]^{1/(r+1)}$ and $x[p(x)]^{r/(r+1)}$ are both bounded. This occurs, in turn, when $p(x)$ is bounded and its tails decay as $1/x^{(r+1)/r}$ or faster. Hence, for $r > 1$ we can handle pdf's with heavier tails than with the standard RoU method.

It is interesting to analyze the probability of acceptance, $p_A(r)$, for a point drawn uniformly from the rectangle $\mathcal{R}_r$. This probability is given by the ratio between the two areas, i.e.,

$$p_A(r) = \frac{|\mathcal{A}_r|}{|\mathcal{R}_r|}, \qquad (2.111)$$

and it is easy to prove that

$$|\mathcal{A}_r| = \frac{\int_{\mathcal{D}} p(x) dx}{r + 1}, \qquad (2.112)$$

in a way similar to the derivation of Eq. (2.89). Moreover, defining $a(r) \triangleq \sup[p(x)]^{1/(r+1)}$, $b^-(r) \triangleq \inf x[p(x)]^{r/(r+1)}$ and $b^+(r) \triangleq \sup x[p(x)]^{r/(r+1)}$, the area of the bounding rectangle is $|\mathcal{R}_r| = a(r)[b^+(r) - b^-(r)]$. Substituting this expression into Eq. (2.111), we obtain

$$p_A(r) = \frac{\int_{\mathcal{D}} p(x) dx}{(r+1)a(r)[b^+(r) - b^-(r)]}. \qquad (2.113)$$

In some cases, it is possible to analytically obtain the optimum value of r in order to maximize the acceptance probability $p_A(r)$ in Eq. (2.113). We now show an example involving a standard Gaussian pdf.

Example 1 *Consider a standard Gaussian density, i.e., $p_o(x) \propto p(x) = \exp\{-x^2/2\}$ with $x \in \mathbb{R}$. In this case, we know that*

$$\int_{\mathbb{R}} p(x)dx = (2\pi)^{1/2}$$

and

$$a(r) = \sup[p(x)]^{1/(r+1)} = 1. \qquad (2.114)$$

Moreover, we can find the first derivative of the function $\phi(x) \triangleq x[p(x)]^{r/(r+1)} = x \exp\left\{-\frac{r}{2(r+1)}x^2\right\}$ w.r.t. x and then write

$$\frac{d\phi}{dx} = \left(1 - \frac{r}{r+1}x^2\right)\exp\left\{-\frac{r}{2(r+1)}x^2\right\}. \qquad (2.115)$$

The solutions of $\frac{d\phi}{dx} = 0$ are $x_{1,2} = \pm\sqrt{\frac{r+1}{r}}$, where $\phi(x_1) = b^-(r)$ and $\phi(x_2) = b^+(r)$ in Eq. (2.113). Namely, we obtain

$$b^-(r) = \inf_x x[p(x)]^{r/(r+1)} = -\left(\sqrt{\frac{r+1}{r}}\right)\exp\{-1/2\}, \qquad (2.116)$$

and

$$b^+(r) = \sup_x x[p(x)]^{r/(r+1)} = \left(\sqrt{\frac{r+1}{r}}\right)\exp\{-1/2\}. \qquad (2.117)$$

Substituting (2.114), (2.116) and (2.117) into (2.113) yields

$$p_A(r) = \frac{(2\pi)^{1/2}}{(r+1)\left[(\frac{r+1}{r})^{1/2}\exp\{-1/2\} + (\frac{r+1}{r})^{1/2}\exp\{-1/2\}\right]},$$

which reduces to

$$p_A(r) = \frac{(2\pi re)^{1/2}}{2(r+1)^{3/2}}, \qquad (2.118)$$

after some straightforward calculations. The maximization of $p_A(r)$ in (2.118) w.r.t. r yields $\min_r p_A(r) = 0.755$, which is attained for $r^* = \frac{1}{2}$. Note that, for the standard RoU method ($r = 1$), we have $p_A(1) = 0.731$.

RoU method for multidimensional densities

The ratio of uniforms technique can be easily generalized for multidimensional pdf's. We provide below a version using also the power transformation.

Theorem 7 *Consider the target pdf $p_o(\mathbf{x}) \propto p(\mathbf{x})$ with $\mathbf{x} = (x_1, ..., x_k) \in \mathbb{R}^k$ and assume that the point $(v_1, ..., v_k, u) \in \mathbb{R}^{k+1}$ is a sample drawn uniformly from the set*

$$\mathcal{A}_{rk} = \left\{ (v_1, ..., v_k, u) : 0 \le u \le \left[p\left(\frac{v_1}{u^r}, ..., \frac{v_k}{u^r} \right) \right]^{1/(rk+1)} \right\}, \qquad (2.119)$$

where $r \ge 0$. Then $\mathbf{x} = (x_1, ..., x_k)$, where $x_i = v_i/u^r$, is a sample from the distribution with density $p_o(\mathbf{x}) \propto p(\mathbf{x})$.

Proof: Assume that the r.v.'s $(V_1, ..., V_k, U)$ are distributed uniformly on $\mathcal{A}_{rk}$, and consider the direct and inverse transformations

$$\begin{cases} x_1 = \dfrac{v_1}{u^r} \\ \dots \\ x_i = \dfrac{v_i}{u^r} \\ \dots \\ y = u \end{cases} \longrightarrow \begin{cases} v_1 = x_1 y^r \\ \dots \\ v_i = x_i y^r \\ \dots \\ u = y \end{cases} \qquad (2.120)$$

Then, the joint pdf $q(\mathbf{x}, y)$ of the r.v.'s $(X_1, ..., X_k, Y)$ is

$$q(\mathbf{x}, y) = \frac{1}{|\mathcal{A}_{rk}|} |J^{-1}| \quad \text{for all} \quad 0 \le y \le [p(x_1, ..., x_k)]^{1/(rk+1)}. \qquad (2.121)$$

Moreover, we can calculate easily the Jacobian of the inverse transformation

$$J^{-1} = \det \begin{bmatrix} y^r & 0 & \dots & 0 & x_1 r y^{r-1} \\ 0 & y^r & \dots & 0 & x_2 r y^{r-1} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} = y^r \times y^r \times \overset{(k)}{\dots} \times y^r = y^{rk}, \quad (2.122)$$

so that

$$q(\mathbf{x}, y) = \frac{1}{|\mathcal{A}_{rk}|} y^{rk} \quad \text{for all} \quad 0 \le y \le [p(x_1, ..., x_k)]^{1/(rk+1)}. \qquad (2.123)$$

73

Finally, we integrate $q(\mathbf{x}, y)$ to obtain the marginal density $q(x)$,

$$
\int_{-\infty}^{+\infty} q(\mathbf{x}, y) dy = \int_0^{[p(\mathbf{x})]^{1/(rk+1)}} \frac{y^{rk}}{|\mathcal{A}_{rk}|} dy =
$$

$$
= \frac{1}{|\mathcal{A}_{rk}|} \left[\frac{y^{(rk+1)}}{rk+1} \right]_0^{[p(\mathbf{x})]^{1/(rk+1)}} = \frac{p(\mathbf{x})}{(rk+1)|\mathcal{A}_{rk}|} = p_o(\mathbf{x}),
$$

where the first equality follows from Eq. (2.123). $\square$

2.8.2 Relationship between the RoU, transformed rejection and inverse-of-density methods

Let us consider a monotonic bounded target density $p_o(x)$ with an unbounded support $\mathcal{D}$, so that the region $\mathcal{A}_0$ below $p_o(x)$ is unbounded as well. In the sequel, for simplicity we assume $p(x) = p_o(x)$. Since $y = p_o(x)$ is a bounded monotonic function, the set $\mathcal{A}_g$ given by Eq. (2.94) can also be rewritten as

$$
\mathcal{A}_g = \left\{ (v, u) : 0 \le v \le p_o^{-1}(g(u)) \dot{g}(u) \right\}, \tag{2.124}
$$

where $p_o^{-1}(y)$ is the inverse of the target density (we have considered $c = 1$). The RoU method asserts that if we are able to draw points (v', u') uniformly from $\mathcal{A}_g$ defined as in Eq. (2.94) or, equivalently, in Eq. (2.124), the sample $x' = v'/\dot{g}(u')$ is distributed according to $p_o(x)$.

Let us consider an increasing differentiable transformation $u = h(y)$. Moreover, consider the random variable Y with pdf $p_o^{-1}(y)$ and the transformed variable $U = h(Y)$ with density $q(u) \propto p_o^{-1}(h^{-1}(u)) \dot{h}^{-1}(u)$. The region below $q(u)$ is

$$
\mathcal{A}_h = \left\{ (v, u) : 0 \le v \le p_o^{-1}(h^{-1}(u)) \dot{h}^{-1}(u) \right\}, \tag{2.125}
$$

and we can note that Eq. (2.124) is equivalent to Eq. (2.125) when $g(u) = h^{-1}(u)$. Therefore, we can state the following result.

Proposition 1 *The region $\mathcal{A}_g$ can be obtained as a transformation $h = g^{-1}$ of a random variable Y distributed according to the inverse pdf $p_o^{-1}(y)$. Specifically, given a r.v. $U = h(Y) = g^{-1}(Y)$ with pdf indicated as $q(u)$, the region $\mathcal{A}_g$ coincides with the area $\mathcal{A}_h$ below the curve $q(u)$.*

Clearly, the cases of interest are those in which the region $\mathcal{A}_g$ is bounded, as seen in Section 2.8. Moreover, in Section 2.5.6 we have discussed the properties that a transformation $h(y)$ have to fulfill in order to obtain a bounded region $\mathcal{A}_h$, when the inverse density $p_o^{-1}(y)$ is itself unbounded but with bounded support, as shown in Figure 2.17(a). Specifically, we need that

$$\lim_{u \to h(0)} \frac{dh^{-1}}{du} = 0, \tag{2.126}$$

and $x = p_o^{-1}(y)$ goes to infinity faster than $\frac{dh}{dy}$ for $y \to 0$, i.e., $p_o(x)$ decays to zero as

$$\left(\frac{dh}{dy}\right)^{-1}(x), \tag{2.127}$$

when $x \to \pm\infty$.

Note that if we set $g(u) = h^{-1}(u)$ and $h(0) = 0$, the expressions (2.126) and (2.127) are exactly equivalent to the equations (2.103) and (2.104), respectively. Hence, the conditions that the function $g(u)$ in Section 2.8.1 must satisfy in order to guarantee the the region $\mathcal{A}_g$ be bounded are exactly the same conditions that have to be imposed on the function $h^{-1}(u)$ of Section 2.5.6 in order to apply the transformed rejection method.

Therefore, we can argue the set $\mathcal{A}_g$ defined by Eq. (2.94) or (2.124) is obtained by applying the transformed rejection method for unbounded pdf's to the inverse density $p_o^{-1}(y)$ (see Section 2.5.6). Figure 2.17(b) displays the region $\mathcal{A}_h$ (that coincides with $\mathcal{A}_g$ if $g = h^{-1}$) defined in Eq. (2.125). Figure 2.17(c) depicts the same region $\mathcal{A}_h$ rotated 90°.

Moreover, given two variables v' and u' uniformly distributed on the set $\mathcal{A}_h$ generated by *extended inverse-of-density method* described in Appendix A, we can assert that the variable defined as

$$x' = \frac{v'}{\left|\frac{dh^{-1}}{du}\right|_{u'}} = \frac{v'}{\left|\dot{h}^{-1}(u')\right|} \tag{2.128}$$

has density $p_o(x)$. If we set $h(y) = g^{-1}(y)$ (with g a monotonic function) we obtain $x' = v'/\dot{g}(u')$ that is exactly equivalent to the generalized RoU technique in Section 2.8.1.

Therefore, we can see the RoU method as a combination of the "transformed rejection method" applied to the inverse density $p_o^{-1}(y)$, described in Section 2.5.6, with the "extended inverse-of-density method" explained in Appendix A.
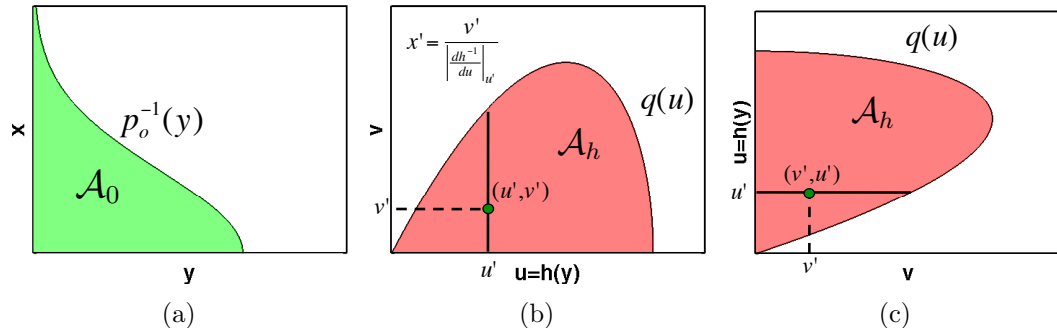
Figure 2.17: **(a)** Example of region $\mathcal{A}_0$ defined by the inverse density $p_o^{-1}(y)$. **(b)** The density $q(u) = 2 \left| \frac{dh^{-1}}{du} \right| p_o^{-1}(h^{-1}(u))$ obtained transforming the r.v. Y, i.e., $U = h(Y)$. Generating uniformly the point (u', v') in the area $\mathcal{A}_h$ we can obtain samples x' from $p_o(x)$ using Eq. (2.128). **(c)** The region $\mathcal{A}_h$ rotated 90° in order to show it how appears when we apply the RoU transformation.

2.8.3 Bounding polygons

The adaptive rejection sampling idea has been implemented jointly with the RoU method in [92, 93]. Indeed, if the region $\mathcal{A}$ is convex it is possible to construct adaptively a bounding region $\mathcal{P}_t$, such that $\mathcal{A} \subseteq \mathcal{P}_t$, with a polygonal boundary. The underlying idea is that drawing from the polygon $\mathcal{P}_t$ is easier than drawing from $\mathcal{A}$. This ability readily enables an accept/reject procedure to draw uniformly from $\mathcal{A}$. To be specific consider a set of support points

$$\mathcal{S}_t = \{ \mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_{m_t} \}$$

where $\mathbf{s}_i = [v_i, u_i]$, $i = 1, ..., m_t$, are point in the $v - u$ space. The boundary region $\mathcal{P}_t$ can be built using the straight lines tangent at $\mathbf{s}_i$ to the boundary of the convex region $\mathcal{A}$. Figure 2.18 shows an example of bounding set $\mathcal{P}_t$ with polygonal boundary built using $m_t = 5$ support points.

As a next step, note that it is always possible to calculate the first derivative of the boundary of $\mathcal{A}$ if the function $p(x)$ is differentiable, without knowing the explicit equation of the contour. Indeed, the boundary of $\mathcal{A}$ can

76
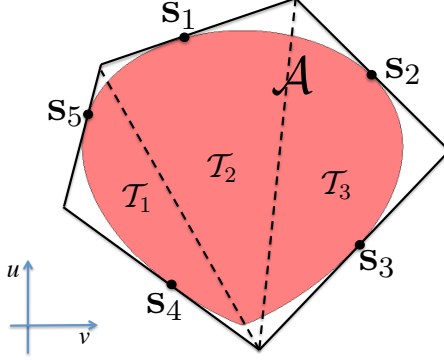
Figure 2.18: Example of construction of a bounding polygon $\mathcal{P}_t$ using the tangent lines at the support points $\mathbf{s}_i$, $i = 1, ..., m_t = 5$, to the boundary of the convex region $\mathcal{A}$. The polygon can be divided in $m_t - 2 = 3$ non-overlapping triangles, i.e., $\mathcal{P}_t = \cup_{k=1}^{3}\mathcal{T}_k$.

be described parametrically as

$$
\begin{cases}
u = u(x) = \sqrt{p(x)} \\
v = v(x) = x\sqrt{p(x)}
\end{cases}, \tag{2.129}
$$

in the case of the standard RoU method. Hence, we can use the chain rule for computing the derivative and write

$$
\frac{dv}{du} = \frac{dv}{dx}\frac{dx}{du} = \left(\sqrt{p(x)} + \frac{x}{2\sqrt{p(x)}}\frac{dp}{dx} \right)(du/dx)^{-1} \tag{2.130}
$$

$$
= \left(\sqrt{p(x)} + \frac{x}{2\sqrt{p(x)}}\frac{dp}{dx} \right)\left(\frac{1}{\frac{1}{2\sqrt{p(x)}}\frac{dp}{dx}} \right) = 2\frac{p(x)}{\dot{p}(x)} + x,
$$

where $x = \frac{v}{u}$ and $\dot{p} = \frac{dp}{dx}$.

Furthermore, it is straightforward to draw samples uniformly from the polygon $\mathcal{P}_t$ by dividing it into $m_t - 2$ non-overlapping triangular areas $\mathcal{T}_k$, i.e., $\mathcal{P}_t = \cup_{k=1}^{m_t-2}\mathcal{T}_k$ where $\mathcal{T}_i \cap \mathcal{T}_j = \emptyset$ whenever $i \neq j$. Note that, it is straightforward to sample uniformly from a triangle $\mathcal{T}_k$ using only two uniform random variables, as shown in Appendix B.

Therefore, to generate samples uniformly from $\mathcal{P}_t$, we first have to randomly select a triangle with proportional probabilities to the areas $|\mathcal{T}_k|$,

$k = 0, ..., m_t - 2$, and then draw from the selected triangular subset. For the first step, we define the normalized weights

$$w_k \triangleq \frac{|\mathcal{T}_k|}{\sum_{i=0}^{m_t-2} |\mathcal{T}_i|}, \tag{2.131}$$

and then we choose a triangular piece by drawing an index $k' \in \{0, ..., m_t - 2\}$ from the probability distribution $P(k) = w_k$. For the second step, we easily generate a point (v', u') uniformly in the selected triangular region $\mathcal{T}_{k'}$ using the procedure in Appendix B.

If this point (v', u') belongs to $\mathcal{A}$, we accept the sample $x' = v'/u'$ and set $m_{t+1} = m_t$, $\mathcal{S}_{t+1} = \mathcal{S}_t$ and $\mathcal{P}_{t+1} = \mathcal{P}_t$. Otherwise, we discard the sample $x' = v'/u'$ and incorporate it into the set of support points, $\mathcal{S}_{t+1} = \mathcal{S}_t \cup \{\mathbf{s}' = (v', u')\}$, so that $m_{t+1} = m_t + 1$ and the region $\mathcal{P}_{t+1}$ is improved by adding another tangent line.

We remark that this procedure is applicable if the region $\mathcal{A}$ is convex. It is possible to prove that $\mathcal{A}$ is convex if, and only if, the target pdf $p_o(x)$ is T-concave where $T(x) = -1/\sqrt{x}$ [92].

Moreover, in [61] it is proved that every log-concave density is also a T-concave pdf with $T(x) = -1/\sqrt{x}$. Therefore, this adaptive RoU technique can be applied to log-concave target pdf $p_o(x)$ as well.

An outline of the adaptive RoU algorithm is given in Table 2.15.

2.9   Summary

In this chapter, we have described a collection of random sampling methods [30, 44, 64, 149] that are relevant for the original material to be introduced in Chapters 3, 4 and 5. All the technique we have discussed assume the availability of a random source with known distribution (often uniform) and all of them, except for the ARMS algorithm, are designed to yield i.i.d. random samples.

The various techniques have been broadly classified by the methodology in which they are based. Hence, we have started with direct algorithm based on transformation of random variables. Then we have revisited the fundamental theorem of simulation and, from there, we have elaborated on the rejection sampling and adaptive rejection sampling families of methods. Finally, we have explored the ratio of uniforms approach, that can be seen both a transformation-based and as a rejection-based technique.

Table 2.15: Adaptive RoU scheme.

1. Start with $t = 0$, $j = 1$ and initialize the set of support points
   $\mathcal{S}_0 = \{\mathbf{s}_1, ..., \mathbf{s}_{m_0}\}$.
   For every $t \geq 0$:

2. Construct the enveloping polygon $\mathcal{P}_t$ using the tangent lines at $\mathbf{s}_i$,
   $i = 1, ..., m_t$, to the boundary of the convex region $\mathcal{A}$.

3. Construct the triangular regions $\mathcal{T}_k$, $k = 0, ..., m_t - 2$, as described
   in Figure 2.18.

4. Calculate the area $|\mathcal{T}_k|$ of every triangle, and compute the normalized
   weights
   $$w_k \triangleq \frac{|\mathcal{T}_k|}{\sum_{i=0}^{m_t-2} |\mathcal{T}_i|}, \quad \text{with} \quad k = 0, ..., m_t - 2.$$

5. Draw an index $k' \in \{0, ..., m_t - 2\}$ from the probability distribution
   $P(k) = w_k$.

6. Generate a point (v', u') uniformly from the region $\mathcal{T}_{k'}$ as explained
   in Appendix B.

7. If $u' \leq \sqrt{p(v'/u')}$, then accept the sample $x^{(j)} = x' = \frac{v'}{u'}$, set $j = j + 1$,
   $\mathcal{S}_{t+1} = \mathcal{S}_t$ and $m_{t+1} = m_t$.

8. Otherwise, if $u' > \sqrt{p(v'/u')}$, then reject the sample $x' = \frac{v'}{u'}$,
   set $\mathcal{S}_{t+1} = \mathcal{S}_t \cup \{\mathbf{s}' = (v', u')\}$, and sort $\mathcal{S}_{t+1}$ in ascending order.
   Finally, update $m_{t+1} = m_t + 1$.

9. If $j > N$ then stop, else go back to step 2.

An additional effort has been made to establish connections and relationships among the various techniques. Some of these links are well-known, but some others we have not to be able to find in the literature. The latter include the material on Section 2.2.4 (on the generic deconvolution method); the equivalent between the inverse-of-density algorithm and the VDR type 2 technique (see Section 2.4.1); the extension of the transformed

rejection method for unbounded pdf's (see Section 2.5.6); and the extensions of the transformed density rejection in Section 2.7.3.

Chapter 3

A generalization of the adaptive rejection sampling algorithm.

In this chapter, we introduce a generalization of the standard ARS method of [50]. The new algorithm can be applied to a large class of target pdf's, possibly not log-concave and possibly multimodal. In particular, we only assume that the log-density $\log[p_o(x)]$ can be expressed as a sum of composed functions, $\log[p_o(x)] = -\sum_i^n (\bar{V}_i \circ g_i)(x) + cst$, where the $\bar{V}_i$'s are convex and the g_i's are either convex or concave. These initial assumptions on the convexity of the $\bar{V}_i$'s and the g_i's can be relaxed, as shown in Chapter 5. Although this is not a universal decomposition that can be applied to every density of interest, indeed the freedom in the choice of the $\bar{V}_i$' and the g_i's enables to describe a large family of pdf's that includes, e.g., *a posteriori* distributions of random variables given a set of independent observations.

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

The proposed method includes the standard ARS of [47, 50] as a special case and brings other improvements over existing techniques. It can be applied to non-log-concave pdf's (unlike the standard ARS [50]), to multimodal pdf's (differently from the transformed density rejection of [61], described in Section 2.7.3) and does not need the knowledge of the

inflection points of the potential function (as required in [36]). Moreover, the proposed technique produces independent samples (differently from the ARMS algorithm of [48], described in Section 2.7.1).

The new generalized ARS (GARS) 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 [90]. There is another potential application in the implementation of the Gibbs sampler for systems in which the conditional densities are complicated.

The rest of the chapter is organized as follows. Some preliminary definitions and assumptions are presented in Section 3.1. The basic form of the new algorithm is introduced in Section 3.2 that contains the standard ARS as a special case, as shown in Section 3.3. In Section 3.4, we study the asymptotic convergence of sequence of the proposal pdf's built with the GARS method toward the target density. Some limitations are discussed in Section 3.6. Section 3.7 is devoted to describe the range of applicability of the proposed technique. We conclude with a brief summary and conclusions in Section 3.8.

3.1   Model

In this chapter we consider a target pdf $p_o(x)$, $x \in \mathcal{D} \subseteq \mathbb{R}$, that can be written as

$$p_o(x) \propto p(x) = \exp\left\{-V(x; \mathbf{g})\right\} = \exp\left\{-\sum_{i=1}^{n} \bar{V}_i(g_i(x))\right\}, \qquad (3.1)$$

where the potential function has the form

$$V(x; \mathbf{g}) \triangleq \sum_{i=1}^{n} \bar{V}_i(g_i(x)). \qquad (3.2)$$

We assume that

1. the functions $\bar{V}_i(\vartheta_i)$, for $i = 1, \ldots, n$ (hereafter called *marginal potentials*), are convex with a minimum at μ_i and

2. the nonlinearities $g_i(x)$, $i = 1, \ldots, n$, are either convex or concave (i.e., they have a second derivative with constant sign).

82

Note that this scenario can be related to a product of densities [104], i.e., $p_o(x)$ can be expressed as

$$p_o(x) \propto \prod_{i=1}^{n} q_i(x), \qquad (3.3)$$

where $q_i(x) \propto \exp\{-\bar{V}_i(g_i(x))\}$, $i = 1, ..., n$.

The potential $V(x; \mathbf{g})$ in Eq. (3.2) is, in general, a non-convex function. Moreover, in general it is impossible to study analytically the first and second derivatives of the potential $V(x; \mathbf{g})$ of Eq. (3.2) in order to calculate the stationary or inflection points. Therefore, the procedures (standard ARS, transformed density, concave convex ARS, etc...) [36, 50, 61, 54] in the literature cannot be applied, in general, but only for some specific choices of the $\bar{V}_i$'s and g_i's.

3.2 The GARS algorithm

3.2.1 Basics

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. We first recall the potential function $-\log[p(x)]$ in Eq.(3.2),

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

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

$$\mathbf{r}(x) \triangleq [r_1(x), r_2(x), \dots, r_n(x)],$$

that lies below the original one, i.e., $V(x, \mathbf{r}) \le V(x, \mathbf{g})$. 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

$$\bar{V}_i(r_i(x)) \le \bar{V}_i(g_i(x)) , \qquad \forall x \in \mathcal{I}, \qquad (3.4)$$

83

for every $i = 1, \ldots, n$ (see Section 3.2.3 for details). As a consequence, substituting $\mathbf{g}$ by $\mathbf{r}$ into the functional $V(x; \cdot)$, we obtain the inequality

$$
\begin{aligned}
V(x; \mathbf{r}) &\triangleq \sum_{i=1}^{n} \bar{V}_i(r_i(x)) \\
&\leq V(x; \mathbf{g}) = \sum_{i=1}^{n} \bar{V}_i(g_i(x)),
\end{aligned}
\tag{3.5}
$$

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

$$
\exp\{-V(x; \mathbf{r})\} \geq \exp\{-V(x; \mathbf{g})\} = p(x).
\tag{3.6}
$$

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,

$$
\begin{aligned}
\frac{d^2 \bar{V}_i(r_i(x))}{dx^2} &= \frac{d\bar{V}_i}{d\vartheta} \frac{d^2 r_i}{dx^2} + \left(\frac{dr_i}{dx}\right)^2 \frac{d^2 \bar{V}_i}{d\vartheta^2} \\
&= 0 + \left(\frac{dr_i}{dx}\right)^2 \frac{d^2 \bar{V}_i}{d\vartheta^2} \geq 0
\end{aligned}
\tag{3.7}
$$

where we have used that
$$
\frac{d^2 r_i}{dx^2} = 0,
$$

because r_i is linear, and the convexity of the marginal potentials $\bar{V}_i(\vartheta)$, $i = 1, \ldots, n$. Therefore, we can choose 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,

$$
\begin{aligned}
\exp\{-W(x)\} &\geq \exp\{-V(x; \mathbf{r})\} \\
&\geq \exp\{-V(x; \mathbf{g})\} = p(x)
\end{aligned}
\tag{3.8}
$$

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

$$
c^{-1} = \int_{x \in \mathcal{I}} \exp\{-W(x)\} dx
\tag{3.9}
$$

84

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 $\mathcal{I}$.

Figure 3.1 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 with a dashed line. The linear function $W(x)$ is tangent to the modified potential $V(x; \mathbf{r})$ in an arbitrarily chosen point $x^* \in \mathcal{I}$.
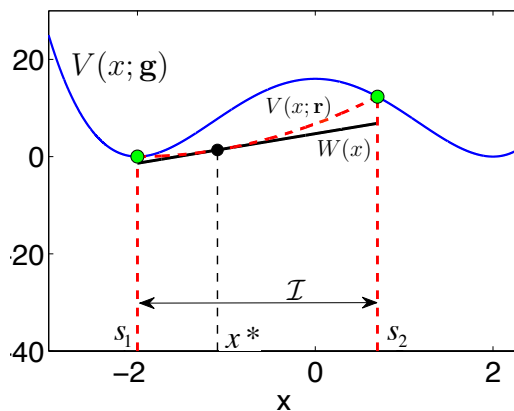
Figure 3.1: Example of construction of the linear function $W(x)$ inside a generic interval $\mathcal{I} = [s_1, s_2]$. The picture shows a non-convex potential $V(x; \mathbf{g})$ in solid line while the modified potential $V(x; \mathbf{r})$ is depicted in dashed line for $\forall x \in \mathcal{I}$. The linear function $W(x)$ is tangent to $V(x; \mathbf{r})$ at a arbitrary point $x^* \in \mathcal{I}$.

3.2.2 Adaptive algorithm

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

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

that converges to the target pdf $p_o(x)$. Similar to other ARS-like techniques, the proposed adaptive algorithm is based on a collection of support points from which the proposed densities are built. Let us denote the set of support points after the t-th iteration as

$$\mathcal{S}_t \triangleq \{s_1, s_2, \ldots, s_{m_t}\} \subset \mathcal{D} \tag{3.10}$$

and sort them in ascending order, $s_1 < \ldots < s_{m_t}$, 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 $k = 1, \ldots, m_t - 1$, 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 $\mathcal{I}_k$, $k = 0, \ldots, m_t$, we build suitable vectors of linear functions $\mathbf{r}_k(x) \triangleq [r_{1,k}(x), \ldots, r_{n,k}(x)]$, using the technique in Section 3.2.3, to comply with the inequality (3.4), i.e.,

$$\bar{V}_i(r_{i,k}(x)) \leq \bar{V}_i(g_{i,k}(x)), \tag{3.11}$$

for all $x \in \mathcal{I}_k$, $i = 1, \ldots, n$ and $k = 0, \ldots, m_t$. This implies that $V(x; \mathbf{r}_k) \leq V(x; \mathbf{g})$ when $x \in \mathcal{I}_k$.

Moreover, since the modified potential $V(x; \mathbf{r}_k)$ is convex, it is possible to build a piecewise-linear lower hull $W_t(x)$ such that

$$W_t(x) \leq V(x; \mathbf{r}_k) \leq V(x; \mathbf{g}), \tag{3.12}$$

for all $x \in \mathcal{I}_k$ and for every $k = 0, \ldots, m_t$. Indeed, let us build the straight lines tangent to the modified potential $V(x; \mathbf{r}_k)$ at arbitrary points $x_k^* \in \mathcal{I}_k = [s_k, s_{k+1}]$, and denote them $w_k(x)$. As a result, the piecewise linear function $W_t(x)$ at the t-iteration is

$$W_t(x) \triangleq \begin{cases} w_0(x), & \text{if} \quad x \in \mathcal{I}_0 \\ \vdots \\ w_{m_t}(x), & \text{if} \quad x \in \mathcal{I}_{m_t}, \end{cases} \tag{3.13}$$

Since $w_k(x) \leq V(x; \mathbf{g}) \ \forall x \in \mathcal{I}_k$, it follows that Eq. (3.12) is satisfied.

The t-th proposal density is

$$\pi_t(x) \propto \exp\{-W_t(x)\}. \tag{3.14}$$

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 3.2 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 a 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 $V(x; \mathbf{r}_k)$, $k = 0, \ldots, m_t = 3$. 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 $k = 0, \ldots, m_t = 3$.
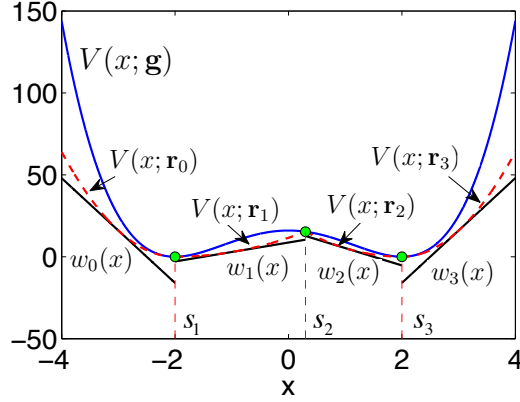
Figure 3.2: Example of construction of the piecewise linear function $W_t(x)$ with three support points $\mathcal{S}_t = \{s_1, s_2, s_{m_t=3}\}$, as carried out by the GARS technique. The potential is $V(x; \mathbf{g}) = 16 - 8x^2 + x^4 = (4 - x^2)^2$ (blue solid line), therefore we can express it as $V(x; \mathbf{g}) = (\bar{V}_1 \circ g_1)(x)$ 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 dashed red lines. The piecewise linear function $W_t(x)$ (depicted with solid black lines) 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 $k = 0, \ldots, m_t = 3$, where $\mathcal{I}_0 = [-\infty, s_1]$, $\mathcal{I}_1 = [s_1, s_2]$, $\mathcal{I}_2 = [s_2, s_3]$ and $\mathcal{I}_3 = [s_3, +\infty]$.

3.2.3 Construction of the linear functions $r_{i,k}(x)$

In this subsection we first define the set of *simple estimates*, needed in the sequel, and then describe in detail how to build adequate linear functions $r_{i,k}(x)$, $i = 1, ..., n$ and $k = 0, ..., m_t$, for each nonlinearity g_i and each interval $\mathcal{I}_k$.

Simple estimates

In order to build suitable linear functions $r_{i,k}(x)$, we need to introduce the set of *simple estimates* corresponding to the nonlinearity $g_i(x)$ as

$$\mathcal{X}_i \triangleq \{x_i \in \mathbb{R} : \quad g_i(x_i) = \mu_i\}, \tag{3.15}$$

where μ_i is the position of the minimum of the marginal potential $\bar{V}_i$. The reason of the name "simple estimates" is clarified in Section 3.7.1.

We 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 ($|\mathcal{X}_i| = 0$, i.e., it is empty), one ($|\mathcal{X}_i| = 1$) or two ($|\mathcal{X}_i| = 2$) simple estimates. Clearly, if $g_i(x)$ is a monotonic function then $|\mathcal{X}_i| \leq 1$. Figure 3.3 displays the three possible cases for a generic concave $g_i(x)$.
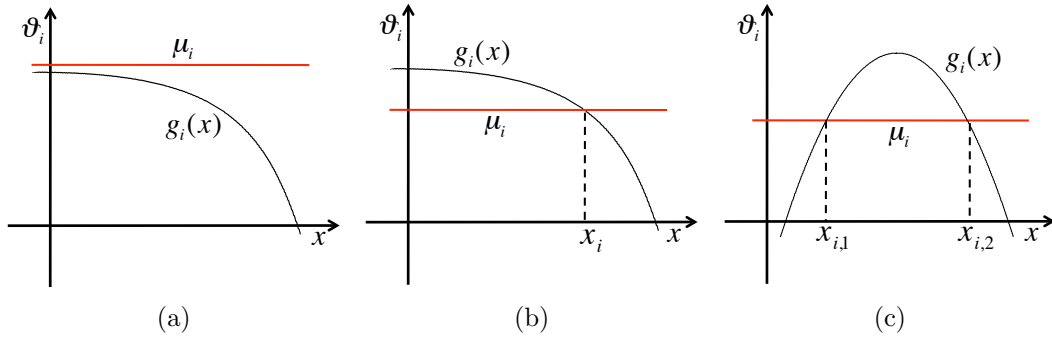
Figure 3.3: Example of the three possible cases for a concave nonlinearity $g_i(x)$. **(a)** The set $\mathcal{X}_i$ is empty, $|\mathcal{X}_i| = 0$. **(b)** There exists one simple estimate x_i, i.e., $|\mathcal{X}_i| = 1$. **(c)** The nonlinearity $g_i(x)$ is a non-monotonic function and $|\mathcal{X}_i| = 2$ ($\mathcal{X}_i = \{x_{i,1}, x_{i,2}\}$).

We assume that all the simple estimates in $\mathcal{X}_i$, $i = 1, ..., n$, are included in the initial set of support points $\mathcal{S}_0$ ($t = 0$), i.e.,

$$\mathcal{X}_i \subset \mathcal{S}_0, \quad \text{for } i = 1, ..., n. \tag{3.16}$$

This condition is needed for the construction of suitable linear functions $r_{i,k}(x)$, $i = 1, ..., n$ and $k = 0, ..., m_t$.

Construction

The GARS algorithm relies on the ability to obtain linear functions $r_{i,k}(x)$, for $i = 1, \ldots, n$ and $k = 0, \ldots, m_t$. It is easy to see that the inequality (3.4) is satisfied for the class of marginal potential functions $\bar{V}_i$ (convex with a minimum at μ_i) if

$$|\mu_i - r_{i,k}(x)| \leq |\mu_i - g_i(x)| \quad \text{and} \tag{3.17}$$

$$(\mu_i - r_{i,k}(x))(\mu_i - g_i(x)) \geq 0 \tag{3.18}$$

jointly, $\forall x \in \mathcal{I}_k$, where $\mu_i = \arg\min_{\vartheta} \bar{V}_i(\vartheta)$. Indeed, if $\mu_i \leq a \leq b$ then $\bar{V}_i(a) \leq \bar{V}_i(b)$ because $\bar{V}_i$ is increasing in $(\mu_i, +\infty)$ whereas for $b \leq a \leq \mu_i$ we have also $\bar{V}_i(a) \leq \bar{V}_i(b)$ because $\bar{V}_i$ is decreasing in $(-\infty, \mu_i)$. Figure 3.4 illustrates the latter inequalities. We can see that green points, closer to the minimum μ_i than the red points, have always a smaller potential value.
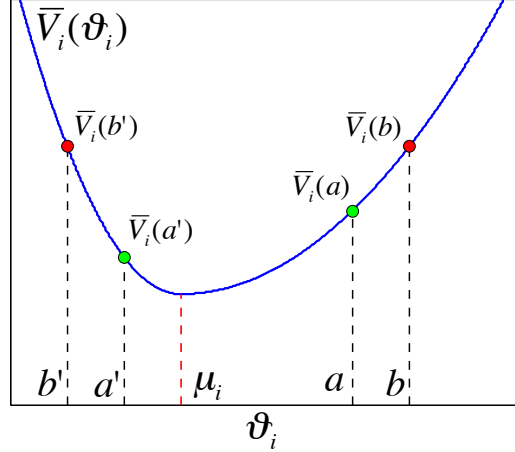
Figure 3.4: An example of marginal potential $\bar{V}_i(\vartheta_i)$. Since we assume that $\bar{V}_i$ is convex with a minimum at μ_i, we have always $\bar{V}_i(a) \leq \bar{V}_i(b)$ if $b \geq a \geq \mu_i$ or $\bar{V}_i(a') \leq \bar{V}_i(b')$ if $b' \leq a' \leq \mu_i$.

Figure 3.5 provides the basic idea of how to construct the linear functions $r_{i,k}(x)$, $k = 0, ..., 3$, for three support points $\mathcal{S}_t = \{s_1 = x_{i,1}, s_2, s_3 = x_{i,2}\}$ (s_1 and s_3 coincide with the two simple estimates). We seek a linear function $r_{i,k}(x)$ such that the absolute difference $d_r = |\mu_i - r_{i,k}(x)|$ is always less than the distance $d_g = |\mu_i - g_i(x)|$, i.e., $d_r \leq d_g$ in an interval $\mathcal{I}_k$. Therefore, in

the intervals $\mathcal{I}_0 = [-\infty, s_1]$ and $\mathcal{I}_3 = [s_3, +\infty]$ we should use tangent straight lines while in $\mathcal{I}_1 = [s_1, s_2]$ and $\mathcal{I}_2 = [s_2, s_3]$ we should use the linear functions passing through the two support points.
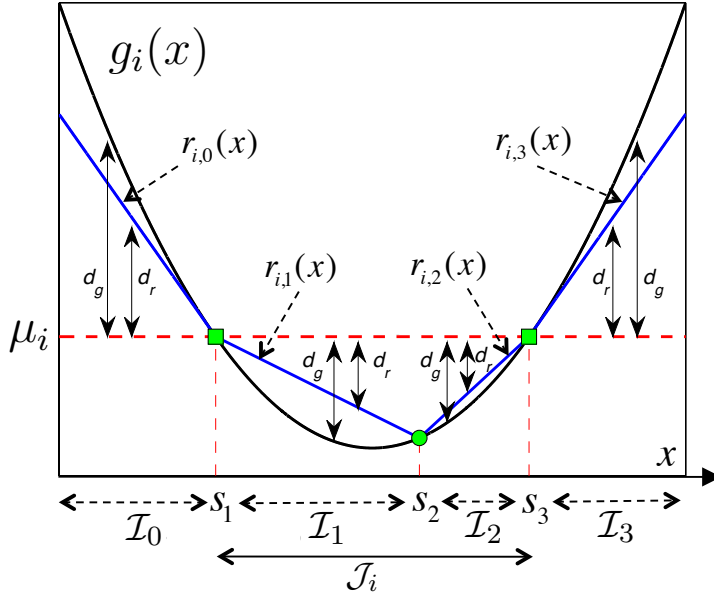
Figure 3.5: Example of construction of the linear function $r_{i,k}(x)$ in order to replace a convex nonlinearity $g_i(x)$ in different intervals $\mathcal{I}_k$, using $m_t = 3$ support points, $\mathcal{S}_t = \{s_1 = x_{i,1}, s_2, s_3 = x_{i,2}\}$. The absolute difference between the linear function $r_{i,k}(x)$ 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 $\mathcal{I}_k$, i.e., $d_r \leq d_g$ for all $x \in \mathcal{I}_k$. Hence, in $\mathcal{I}_0 = [-\infty, s_1]$ and $\mathcal{I}_3 = [s_3, +\infty]$ we use tangent straight lines while in $\mathcal{I}_1 = [s_1, s_2]$ and $\mathcal{I}_2 = [s_2, s_3]$ we use the linear functions passing through the two support points.

Note that if we denote as $\mathcal{J}_i = [s_1 = x_{i,1}, s_3 = x_{i,2}]$ the interval limited by the simple estimates associated to the function $g_i(x)$, the procedure in Figure 3.5 can be summarized as

1. if $\mathcal{I}_k \subset \mathcal{J}_i$ (i.e., $\mathcal{I}_k \cap \mathcal{J}_i = \mathcal{I}_k$), use secant lines,

2. otherwise, if $|\mathcal{I}_k \cap \mathcal{J}_i| = 0$, use tangent lines.

Figures 3.6 displays two examples of construction with four support points. Specifically, it shows the construction of the linear functions $r_{i,k}(x)$
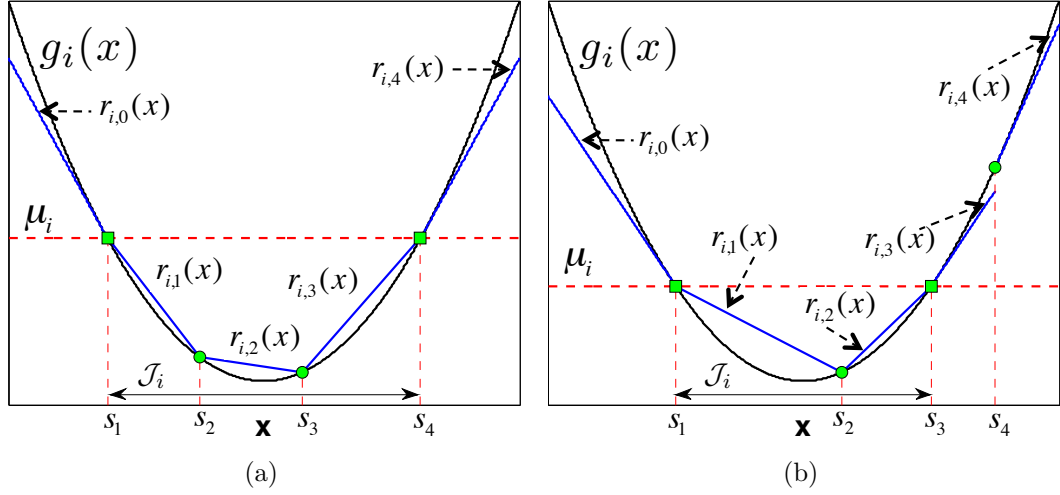
Figure 3.6: Example of construction of the appropriate linear functions $r_{i,k}(x)$, with $k = 0, \ldots, m_t = 4$ for a non-monotonic convex nonlinearity $g_i(x)$. The interval defined by the simple estimates $\mathcal{J}_i = [x_{i,1}, x_{i,2}]$ is indicated by solid double arrows. The set of support points $\mathcal{S}_t = \{s_1, s_2, s_3, s_4\}$ includes always the simple estimates $x_{i,1}$ and $x_{i,2}$ (indicated by squares) and $\mathcal{I}_k = [s_k, s_{k+1}]$, $k = 1, 2, 3$, $\mathcal{I}_0 = (-\infty, s_1]$ and $\mathcal{I}_4 = [s_4, +\infty)$. **(a)** Since $\mathcal{I}_0$, $\mathcal{I}_4$ are not contained in $\mathcal{J}_i$ we use tangent lines for $r_{i,0}(x)$ and $r_{i,4}(x)$. Since $\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3 \subseteq \mathcal{J}_i$, we use secant lines for $r_{i,1}(x)$, $r_{i,2}(x)$ and $r_{i,3}(x)$. **(b)** Since $\mathcal{I}_0$, $\mathcal{I}_3$ and $\mathcal{I}_4$ are not contained in $\mathcal{J}_i$ we use tangent lines for $r_{i,0}(x)$, $r_{i,3}(x)$ and $r_{i,4}(x)$. Since $\mathcal{I}_1, \mathcal{I}_2 \subseteq \mathcal{J}_i$, we use secant lines for $r_{i,1}(x)$ and $r_{i,2}(x)$.

when $g_i(x)$ is non-monotonic and convex with $m_t = 4$ support points. In Figure 3.6(a) the intervals $\mathcal{I}_0$ and $\mathcal{I}_4$ are not contained in $\mathcal{J}_i = [x_{i,1}, x_{i,2}]$, hence the use two tangent lines to build $r_{i,0}(x)$ and $r_{i,4}(x)$. Since $\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3 \subseteq \mathcal{J}_i$, we use secant lines for $r_{i,1}(x)$, $r_{i,2}(x)$ and $r_{i,3}(x)$. In Figure 3.6(b) the intervals $\mathcal{I}_0$, $\mathcal{I}_3$ and $\mathcal{I}_4$ are not contained in $\mathcal{J}_i$, hence we use tangent lines for $r_{i,0}(x)$, $r_{i,3}(x)$ and $r_{i,4}(x)$. Since $\mathcal{I}_1, \mathcal{I}_2 \subseteq \mathcal{J}_i$, we use secant lines for $r_{i,1}(x)$ and $r_{i,2}(x)$.

Computational procedure

Now, we introduce a general computational procedure that enables the computation of $r_{i,k}(x)$ for all cases of interest. For the adequate enumeration of the possible scenarios, we have to extend the definition of the interval $\mathcal{J}_i$

associated to the function $g_i(x)$. Recall that $x_{i,j}$ is a simple estimate of $g_i(x)$ if, and only if, $g_i(x_{i,j}) = \mu_i$. Since $\frac{d^2 g_i}{dx^2}$ is assumed to have constant sign in $\mathcal{I}_k$, there are three possibilities for the construction of $\mathcal{J}_i$:

- If $g_i(x)$ is non-monotonic then there may exist two, one or zero solutions to the equation $g_i(x) = \mu_i$. If there are two solutions, denoted $x_{i,1} < x_{i,2}$, we define $\mathcal{J}_i = [x_{i,1}, x_{i,2}]$. Otherwise, we define $\mathcal{J}_i = \emptyset$.

- If $g_i(x)$ is monotonic and

$$\frac{dg_i(x)}{dx} \times \frac{d^2 g_i(x)}{dx^2} \geq 0$$

  (increasing and convex or decreasing and concave, see Figure 3.8(f) and Figure 3.8(h)), then there may exist one or zero solutions to the equation $g_i(x) = \mu_i$. 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

$$\frac{dg_i(x)}{dx} \times \frac{d^2 g_i(x)}{dx^2} \leq 0$$

  (decreasing and convex or increasing and concave, see Figure 3.8(e) and Figure 3.8(i)), 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 $\mathcal{I}_k$, $k \in \{0, \ldots, m_t\}$. 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 $\mathcal{S}_0 \subseteq \ldots \subseteq \mathcal{S}_t$ contain all the simple estimates.

Now we provide a procedure for the construction of $r_{i,k}(x)$, $i \in \{1, \ldots, n\}$, $k \in \{0, \ldots, m_t\}$ with $x \in \mathcal{I}_0 = (-\infty, s_1]$ for $k = 0$, $x \in \mathcal{I}_k = [s_k, s_{k+1}]$ for $k = 1, \ldots, m_t - 1$ 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 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 set $r_{i,k}(x)$ as the tangent line to $g_i(x)$

92

2a) at s_k, if $\frac{dg_i(x)}{dx} \times \frac{d^2 g_i(x)}{dx^2} \geq 0$ in $\mathcal{I}_k$, or

2b) at s_{k+1}, if $\frac{dg_i(x)}{dx} \times \frac{d^2 g_i(x)}{dx^2} \leq 0$ 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 $g_i(x)$. Specifically, set

$$B_i \triangleq \begin{cases} \max\{\mu_i, \epsilon_i\}, & \text{if } \frac{d^2 g_i(x)}{dx^2} > 0 \\ \min\{\mu_i, \epsilon_i\}, & \text{if } \frac{d^2 g_i(x)}{dx^2} < 0 \end{cases}, \qquad (3.19)$$

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

(a)   (b)
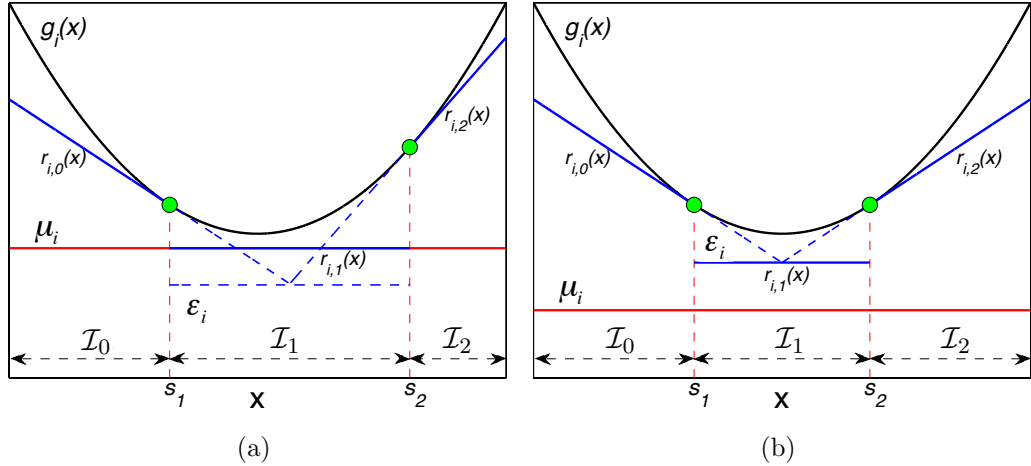
Figure 3.7: Example of construction of the linear functions $r_{i,k}(x)$ when $|\mathcal{J}_i| = 0$ and $g_i(x)$ is non-monotonic and convex. We use two support points $\mathcal{S}_t = \{s_1, s_2\}$. Since $|\mathcal{J}_i \cap \mathcal{I}_0| = |\mathcal{J}_i \cap \mathcal{I}_2| = 0$ ($|\mathcal{J}_i| = 0$), we construct $r_{i,0}(x)$ and $r_{i,2}(x)$ as tangent lines. In $\mathcal{I}_1 = [s_1, s_2]$, the nonlinearity $g_i(x)$ 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 $B_i = \epsilon_i$.

We have already illustrated in Figures 3.5 and 3.6 the construction of the linear functions $r_{i,k}(x)$ when $g_i(x)$ is non-monotonic, convex and $|\mathcal{J}_i| > 0$ (indeed, in this case there are two simple estimates $x_{i,1}$, $x_{i,2}$). Those figures illustrate the steps 1 and 2 of the procedure proposed above.

Figure 3.7 depicts the construction of $r_{i,k}(x)$ when $|\mathcal{J}_i| = 0$ and $g_i(x)$ is convex. The linear functions $r_{i,0}(x)$ and $r_{i,2}(x)$ are tangent to $g_i(x)$ and they have an intersection $r_{i,0}(x^*) = r_{i,2}(x^*) = \epsilon_i$, $x^* \in \mathcal{I}_1$. Since $g_i(x)$ is non-monotonic in $\mathcal{I}_1 = [s_1, s_2]$, $r_{i,1}(x)$ is a constant. Those figures illustrate step 3 of the procedure.

Finally, Figure 3.8 displays an example of construction of the linear functions $r_{i,k}(x)$ for all possible types of nonlinearities $g_i(x)$. Figures 3.8(a),(d),(g) show a marginal potential $\bar{V}_i(\vartheta_i)$ (convex with a minimum at μ_i) with rotated axes. Figures 3.8(b),(c) illustrate the construction of $r_{i,k}(x)$ with $m_t = 3$ support points for non-monotonic $g_i(x)$. Figures 3.8(e),(f) depict the construction with $m_t = 2$ support points for monotonic increasing $g_i(x)$, while Figures 3.8(h),(i) correspond to monotonic decreasing nonlinearities $g_i(x)$. For all cases, we can see that given an arbitrary x' the value $\bar{V}_i(r_{i,k}(x))$ (green point) is always less than $\bar{V}_i(g_i(x))$ (red point), i.e., $\bar{V}_i(r_{i,k}(x)) \leq \bar{V}_i(g_i(x))$. Moreover, 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, $r_{i,m_t}(x) = g_i(s_{m_t})$.

94
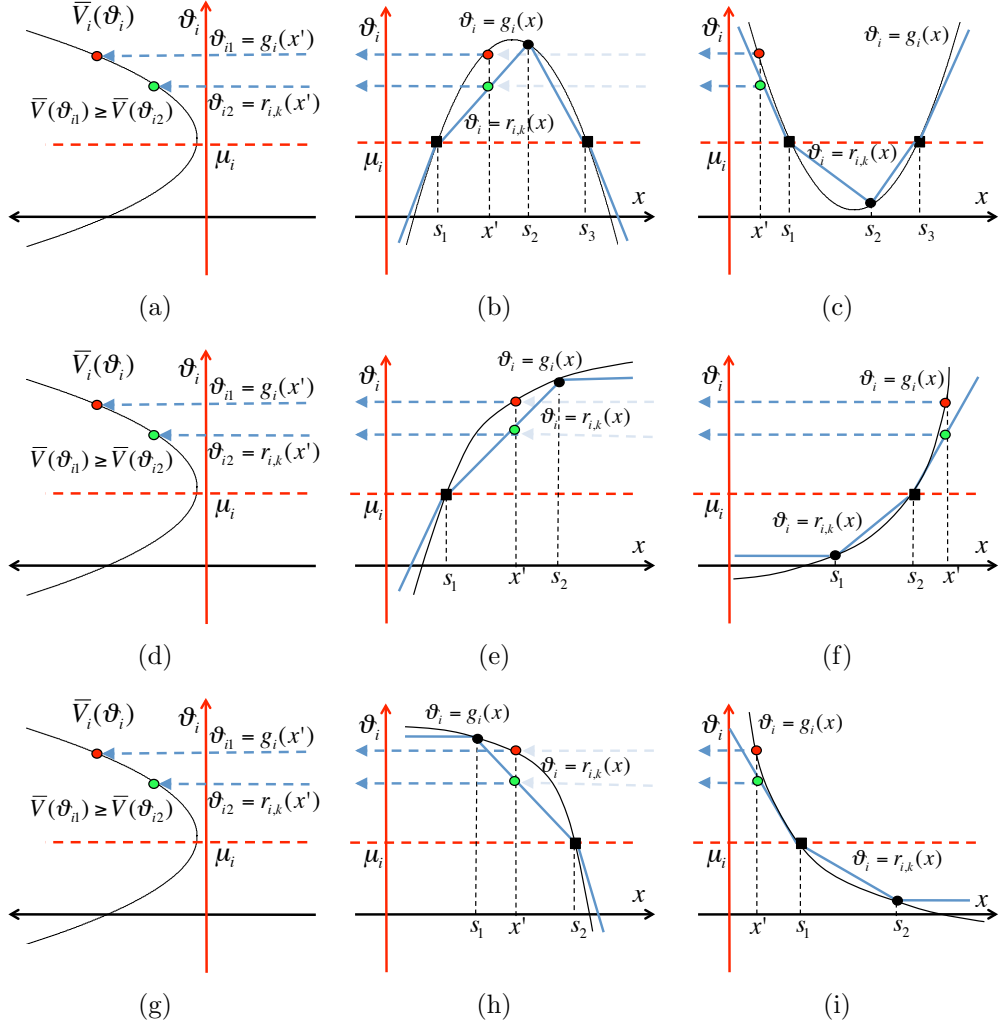
Figure 3.8: Example of construction of suitable linear functions $r_{i,k}(x)$, for different types of nonlinearities $g_i(x)$. **(a)-(d)-(g)** The corresponding marginal potential $\bar{V}_i(\vartheta_i)$ is convex, with a minimum at μ_i. Note that the axis of the independent variable ϑ_i is vertical. **(b)** Non-monotonic and concave $g_i(x)$, $m_t = 3$. **(c)** Non-monotonic and convex $g_i(x)$, $m_t = 3$. **(e)** Monotone increasing and concave $g_i(x)$, $m_t = 2$. **(f)** Monotone increasing and convex $g_i(x)$, $m_t = 2$. **(h)** Monotone decreasing and concave $g_i(x)$, $m_t = 2$. **(i)** Monotone decreasing and convex $g_i(x)$, $m_t = 2$.

3.2.4 Initialization and summary

Let us recall that the set of simple estimates corresponding to the nonlinearities $g_i(x)$ is defined as $\mathcal{X}_i = \{x_i \in \mathbb{R} : \quad g_i(x_i) = \mu_i\}$, where μ_i is the position of the minimum of the marginal potential $\bar{V}_i$. Since we have assumed nonlinearities $g_i(x)$ for which the second derivative $\frac{\partial^2 g_i}{\partial^2 x^2}$ has constant sign, each equation $\mu_i = g_i(x_i)$ can yield zero, one or two different simple estimates.

We initialize the algorithm with a set of support points $\mathcal{S}_0 \triangleq \{s_j\}_{j=1}^{m_0}$ such that all simple estimates are contained in $\mathcal{S}_0$, i.e., $\mathcal{X}_i \subset \mathcal{S}_0$, $i = 1, \ldots, n$.

The set $\mathcal{S}_0$ thus constructed enables us to build the linear functions $r_{i,k}(x)$ in the way described in the 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 3.1. Figures 3.9 illustrates how the sequence of proposal pdf's converges toward the target density as the number of support points increases.

(a)   (b)   (c)
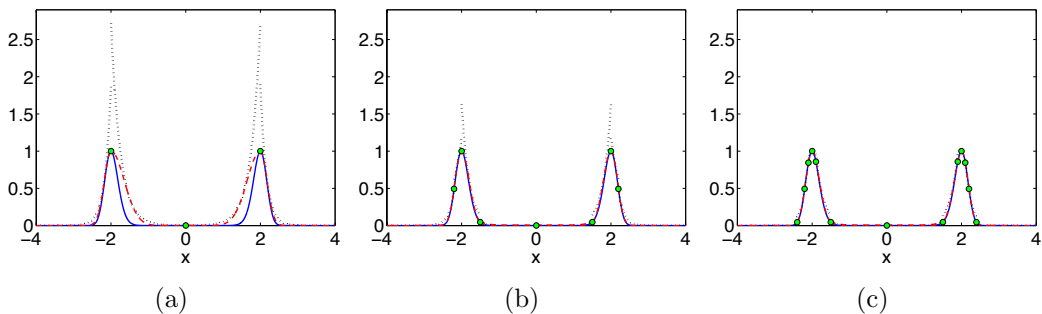
Figure 3.9: Convergence of the overbounding functions $\exp\{-V(x; \mathbf{r}_k)\}$ (red dashed line), and $\exp\{-W_t(x)\}$ (black dotted line) toward the function $p(x) = \exp\{-(4 - x^2)^2\}$ (black solid line), with the GARS technique. The points, corresponding to the support points $\{s_j\}_j^{m_t}$, are depicted with green circles. **(a)** Construction of the overbounding functions with $m_t = 3$ support points. **(b)** Construction with $m_t = 7$. **(c)** Construction with $m_t = 13$.

Table 3.1: Generalized Adaptive Rejection Sampling Algorithm.

1. Start with $q = 1$, $t = 0$ and $\mathcal{S}_0 \triangleq \{s_j\}_{j=1}^{m_0}$ such that $\cup_{i=1}^n \mathcal{X}_i \subset \mathcal{S}_0$. 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 $i = 1, \ldots, n$, $k = 0, \ldots, m_t$.

3. Build $W_t(x) \triangleq w_k(x) \ \forall x \in \mathcal{I}_k$, for every $k = 0, \ldots, m_t$, where $w_k(x)$ is a line tangent 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 $u' \leq \frac{p(x')}{\exp[-W_t(x')]}$ then 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.

3.3 The standard ARS algorithm as a special case

Note that if all the functions $g_i(x)$, $i = 1, ..., n$, are linear the procedure described above coincides with the standard ARS method [50]. Indeed, in this case, the potential $V(x; \mathbf{g})$ is already convex and we can use lines tangent to $V(x; \mathbf{g})$ at the support points in order to build the lower-hull $W_t(x)$. To be specific, if $r_{i,k}(x) = g_i(x)$ for all $i = 1, ..., n$ and $k = 0, ..., m_t$, then the modified and the true potentials coincides $V(x; \mathbf{r}_k) \equiv V(x; \mathbf{g})$ and the GARS method consists in constructing straight lines tangent to $V(x; \mathbf{g})$ exactly as in the standard ARS method.

3.4 Acceptance probabilities

Note that every time a sample x' drawn from $\pi_t(x)$ is rejected, x' is incorporated as a support point in the new set $\mathcal{S}_{t+1} = \mathcal{S}_t \cup \{x'\}$ and, as a consequence, a refined lower hull $W_{t+1}(x)$ is constructed yielding a better approximation of the system potential function. In this way, $\pi_{t+1}(x) \propto \exp\{-W_{t+1}(x)\}$ becomes "closer" to $p(x) \propto p_o(x)$ and it can be expected that the mean acceptance rate be higher. This is illustrated by numerical examples in Chapter 6 (in particular, see Section 6.1.2).

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

$$a_t(x) \triangleq \frac{\exp\{-V(x; \mathbf{g})\}}{\exp\{-W_t(x)\}} = \exp\{-[V(x; \mathbf{g}) - W_t(x)]\}, \qquad (3.20)$$

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

$$\pi_t(x) = c_t \exp\{-W_t(x)\},$$

i.e.,

$$\hat{a}_t \triangleq E[a_t(x)] = \int_{\mathcal{A}} a_t(x) \pi_t(x) dx = c_t \int_{\mathcal{A}} \exp\{-V(x; \mathbf{g})\} dx = \frac{c_t}{c_v} \leq 1, \qquad (3.21)$$

where c_t and c_v are the proportionality constants for $\pi_t(x)$ and $p(x)$,

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

and

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

respectively. Note that $\frac{c_t}{c_v} \leq 1$ because $W_t(x) \leq V(x; \mathbf{g}) \; \forall x \in \mathcal{D}$.

From the Eq. (3.21), we obtain that $\hat{a}_t = 1$ if, and only if, $c_t = c_v$ or, equivalently, $\hat{a}_t = 1$ if and only if the integral

$$e(t) \triangleq \int_{\mathcal{A}} [\exp\{-W_t(x)\} - \exp\{-V(x; \mathbf{g})\}] \, dx \qquad (3.22)$$

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

The error signal $e(t)$ can be interpreted as a divergence[1] between $\pi_t(x)$ and $p(x)$. In particular if $e(t)$ decreases, the acceptance rate $\hat{a}_t = \frac{c_t}{c_v}$ increases and, since $\exp\{-W_t(x)\} \geq \exp\{-V(x;\mathbf{g})\}\ \forall x \in \mathcal{D}$, $e(t) = 0$ if, and only if, $W_t(x) = V(x;\mathbf{g})$ almost everywhere. Equivalently, $\hat{a}_t = 1$ if, and only if, $\pi_t(x) = p_o(x)$ almost everywhere. The convergence of $\hat{a}_t$ toward 1 is illustrated numerically in Section 6.1.2.

3.5 Continuous proposals

The procedure to build the piecewise linear function $W_t(x)$, such that $W_t(x) \leq V(x;\mathbf{g})$, is not unique. For instance, we could build a continuous lower-hull $W_t(x)$ with little variations in the two steps composing the GARS technique. Figure 3.10 depicts how to modify first step of the GARS algorithm in order to obtain a continuous linear approximation of a nonlinearity $g_i(x)$. Figure 3.10(a) can be compared with Figure 3.6(b).

To be specific, it is necessary to calculate the intersection points (displayed with red points in Figure 3.10) $e_{i1},...,e_{ik_i}$, $i = 1,...n$, (the number k_i changes for each nonlinearity $g_i(x)$), of the straight lines tangent to the nonlinearity g_i at the support points $s_1,...s_{m_t}$ outside the interval $\mathcal{J}_i$ (defined in Section 3.2.3). In Figure 3.10(a) we have only one intersection point e_{i1} between the lines $r_{i,3}(x)$ and $r_{i,4}(x)$, while in Figure 3.10(b) we have two points, e_{i1}, e_{i2}, between the lines $r_{i,0}(x)$ and $r_{i,1}(x)$ and between the lines $r_{i,4}(x)$ and $r_{i,5}(x)$. We can denote the set of all intersection points associated to the nonlinearities as

$$\mathcal{E}_t = \{e_{ij}\}_{i=1,j=1}^{n,k_i}, \tag{3.23}$$

hence $|\mathcal{E}_t| = \gamma_t$ is the number of points in $\mathcal{E}_t$. Note that, in Figure 3.10 the straight lines $r_{i,k}$ form a continuous piecewise linear approximation of the nonlinearity $g_i(x)$. Thus, we ensure that the modified potential $V(x;\mathbf{r}_k)$ is also a continuos function. On the contrary, the procedure originally described in Section 3.2.3 (and illustrated in Fig. 3.6(b)) produces a discontinuity at s_4.

Recalling the set of support points $\mathcal{S}_t = \{s_1,...,s_{m_t}\}$, in this case we need to define a extended set

$$\bar{\mathcal{S}}_t \triangleq \mathcal{S}_t \cup \mathcal{E}_t = \{\bar{s}_1,...,\bar{s}_{\bar{m}_t}\}, \tag{3.24}$$

---

[1] Note that $\exp\{-W_t(x)\} - \exp\{-V(x;\mathbf{g})\} \geq 0$ for all $x \in \mathcal{D}$.
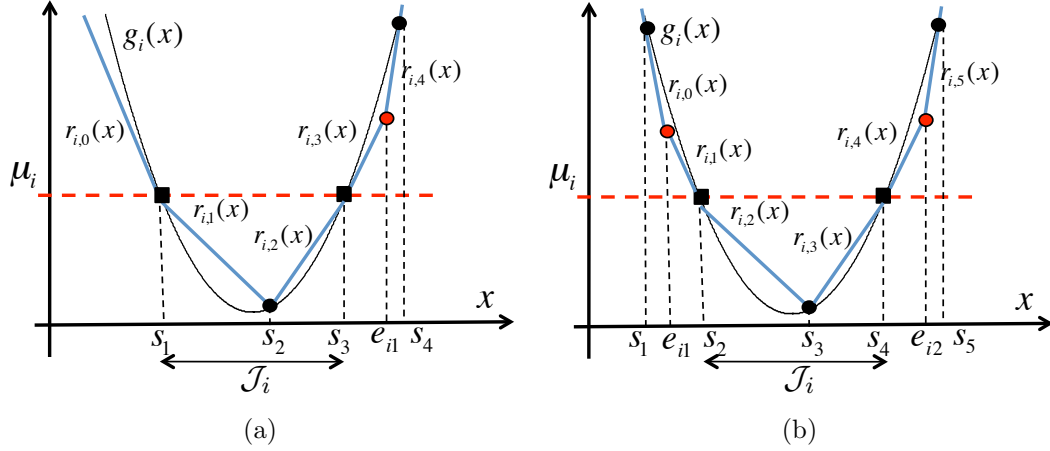
Figure 3.10: Construction of a continuous piecewise linear approximation of the nonlinearity $g_i(x)$. **a)** Example with $m_t = 4$ support points. In the figure, $r_{i,3}(x)$ is defined $\forall x \in [s_3, e_{i1}]$ while $r_{i,4}(x)$ is defined $\forall x \in [e_{i1}, +\infty)$. **b)** Example with $m_t = 5$ support points. The linear functions $r_{i,0}(x)$, $r_{i,1}(x)$ are defined $\forall x \in (-\infty, e_{i1}]$ and $\forall x \in [e_{i1}, s_1]$, respectively, while $r_{i,4}(x)$, $r_{i,5}(x)$ are defined $\forall x \in [s_4, e_{i2}]$ and $\forall x \in [e_{i2}, +\infty)$, respectively.

where $\bar{m}_t = m_t + \gamma_t$ and $\bar{s}_1 < ... < \bar{s}_{\bar{m}_t}$. The points in $\mathcal{S}_t$ are shared for all nonlinearities $g_i(x)$, $i = 1, ..., n$, in order to build the linear functions $r_{i,k}(x)$, while we incorporate the set $\mathcal{E}_t$ in $\bar{\mathcal{S}}_t$ to construct adequately the lower-hull $W_t(x)$.

Figure (3.11) shows an example of construction of the continuous piecewise linear function $W_t(x) \leq V(x; \mathbf{g})$ for a generic potential $V(x; \mathbf{g})$. The construction uses the extended set $\bar{\mathcal{S}}_t = \{\bar{s}_1, ..., \bar{s}_{\bar{m}_t}\}$. In order to make the piecewise function $W_t(x)$ continuous, it is built by lines tangent to the modified potential $V(x, \mathbf{r}_k)$ at the support points $\bar{s}_k$ and $\bar{s}_{k+1}$, for $k = 1, ..., \bar{m}_t$. Figure 3.11 can be compared with Figure 3.5.

Let us remark that, in order to draw from the proposal $\pi_t(x) \propto \exp\{-W_t(x)\}$, we need to calculate other intersection points $v_1, v_2, ...$, shown in Figure 3.11 with green points. Specifically, $v_1, v_2, ...$ are the intersection points for the straight lines tangent to the modified potential $V(x; \mathbf{r}_k)$ at the extended support points $\bar{s}_1, \bar{s}_2..., \bar{s}_{\bar{m}_t}$.

In particular, when a sample x', drawn from $\pi_t(x)$, is accepted we keep $\mathcal{S}_{t+1} = \mathcal{S}_t$, so that the construction of the linear functions $r_{i,k}$ remain the
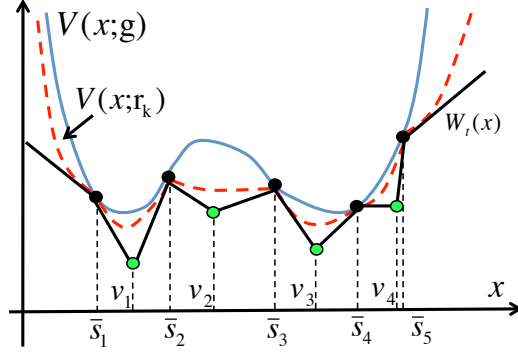
Figure 3.11: Example of construction of a continuous lower-hull $W_t(x)$ for a generic potential $V(x; \mathbf{g})$. In the figure, there are $\bar{m}_t = 5$ extended support points $\bar{\mathcal{S}}_t = \{\bar{s}_1, \bar{s}_2, \bar{s}_3, \bar{s}_4, \bar{s}_5\}$. The piecewise linear function $W_t(x)$ is formed by straight lines tangent to the modified potential $V(x; \mathbf{r}_k)$ (shown in dashed line) at the support points $\bar{s}_k$ and $\bar{s}_{k+1}$. To draw from $\pi_t(x) \propto \exp\{-W_t(x)\}$, it is necessary to compute the positions v_1, v_2, v_3 and v_4.

same, and then $\mathcal{E}_{t+1} = \mathcal{E}_t$ and $\bar{\mathcal{S}}_{t+1} = \bar{\mathcal{S}}_t$. Otherwise, if x' is rejected, we set $\mathcal{S}_{t+1} = \mathcal{S}_t \cup \{x'\}$ and sort it in ascending order. Moreover, we have to update the set of the intersection points $\mathcal{E}_{t+1}$ (after building the linear functions $r_{i,k}(x)$) and the extended set $\bar{\mathcal{S}}_{t+1} = \mathcal{S}_{t+1} \cup \mathcal{E}_{t+1}$.

This construction may improve the mean acceptance rate since possibly the continuous lower-hull $W_t(x)$ is closer to $V(x; \mathbf{g})$. However, the resulting procedure becomes more complicated and computationally more expensive than the original method in Section 3.2.

3.6 Improper proposals

The GARS algorithm summarized in Table 3.1 breaks down when the potential function $V(x; \mathbf{g})$ has both an infinite support ($x \in \mathcal{D} = \mathbb{R}$) and concave tails.

This occurs when all the nonlinearites $g_i(x)$, $i = 1, \ldots, n$, are monotonic and the product

$$\frac{dg_i(x)}{dx} \times \frac{d^2 g_i(x)}{dx^2},$$

has the same sign for all the functions $g_i(x)$, $i = 1, \ldots, n$.

In this case, the proposed procedure yields a piecewise lower hull $W_t(x)$ which is 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.

For instance, if all the nonlinearities $g_i(x)$, $i = 1, \ldots, n$ are monotonic functions with

$$\frac{dg_i(x)}{dx} \times \frac{d^2 g_i(x)}{dx^2} > 0,$$

i.e., increasing and convex or decreasing and concave (see Figures 3.8(f) and (h)), the left tail of $\pi_t(x) \propto \exp\{-W_t(x)\}$ is constant, while, if all the nonlinearities $g_i(x)$, $i = 1, \ldots, n$ are monotonic with

$$\frac{dg_i(x)}{dx} \times \frac{d^2 g_i(x)}{dx^2} < 0,$$

i.e., increasing and concave or decreasing and convex (see Figures 3.8(e) and (i)), the right tail of $\pi_t(x) \propto \exp\{-W_t(x)\}$ is constant.

This drawback is shared by all other adaptive techniques in the literature [36, 50, 54, 61]. We address this problem in Chapter 4.

3.7 Applicability

In this section we briefly describe three general classes of target densities that appear often in practice and can be easily handled with the proposed method. We do not imply that only these three types of pdf's can be sampled using our method. Many other classes of densities can also fit the structure of Eq. (3.1).

3.7.1 Class 1: posterior pdf's in Bayesian inference

Densities of the form of Eq. (3.1) appear naturally in Bayesian inference problems [7, 15, 45, 66] (see also the so-called *nonlinear hierarchical models* [51, Chapter 19], [37, 42, 43]) where it is desired to draw from the posterior pdf $p(x|\mathbf{y})$ with $\mathbf{y} = [y_1, y_2, ..., y_n] \in \mathbb{R}^n$, of a random variable X given a collection of observations

$$\begin{cases} Y_1 = \bar{g}_1(X) + \Theta_1, \\ \vdots \\ Y_n = \bar{g}_n(X) + \Theta_n, \end{cases} \tag{3.25}$$

where $\Theta_1, ..., \Theta_n$ are independent "noise" variables. In fact, writing the noise pdf's as $p(\vartheta_i) \propto \exp\{-\bar{V}_i(\vartheta_i)\}$ (with a mode at $\vartheta_i^* = \mu_i$), $i = 1, ..., n$, the likelihood function can be expressed as

$$p(\mathbf{y}|x) \propto \exp\left\{ -\sum_{i=1}^{n} \bar{V}_i(y_i - \bar{g}_i(x)) \right\}. \qquad (3.26)$$

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

$$
\begin{aligned}
V(x; \mathbf{g}) &= -\log[p(x|\mathbf{y})] \\
&= -\log[p(\mathbf{y}|x)p(x)] = \sum_{i=1}^{n+1} \bar{V}_i(g_i(x)).
\end{aligned}
\qquad (3.27)
$$

Since we are assuming that each $p(\vartheta_i) \propto \exp\{-\bar{V}_i(\vartheta_i)\}$ has only a mode at $\vartheta_i^* = \mu_i$, if we have only one observation ($n = 1$, hence we have only one equation, for instance, $Y_1 = \bar{g}_1(X) + \Theta_1$) the set of maximum likelihood estimators $\hat{\mathcal{X}}$ of the variable of interest x is

$$\hat{\mathcal{X}} = \{x \in \mathcal{D} : g_1(x) = \mu_1\}, \qquad (3.28)$$

where $g_1(x) = y_1 - \bar{g}_1(x)$. Note that Eq. (3.28) is exactly the definition of the "simple estimates" (see Eq. (3.15)) for the first nonlinearity $g_1(x)$, hence the name.

The GARS algorithm can be applied in a more general framework than the model with additive noise of Eq. (3.25), as shown below.

Non-additive noise models

Let us consider a generic collection of r.v.'s $[Y_1, ..., Y_n] \in \mathbb{R}^n$ obtained as

$$
\begin{cases}
Y_1 = G_1(X, \Theta_1), \\
\vdots \\
Y_n = G_n(X, \Theta_n),
\end{cases}
\qquad (3.29)
$$

where $G_i(X, \Theta_i)$ are nonlinear real functions of the r.v.'s X and Θ_i with pdf's $p(\vartheta_i) \propto \exp\{-\bar{V}_i(\vartheta_i)\}$, $i = 1, ..., n$. If every $G_i(x, \cdot)$ is invertible for

the second argument, then we can solve for the noise variables, i.e., we can transform the system of Eq. (3.29) into

$$
\begin{cases}
G_1^{-1}(X, Y_1) = \Theta_1, \\
\vdots \\
G_n^{-1}(X, Y_n) = \Theta_n.
\end{cases}
\tag{3.30}
$$

From (3.30), we obtain a collection of "null pseudo-observations" with additive noise,

$$
0 = -G_1^{-1}(X, Y_1) + \Theta_1, \ldots, 0 = -G_n^{-1}(X, Y_n) + \Theta_n.
\tag{3.31}
$$

Note that the collection of observations in Eq. (3.29) is also contained in the system of Eq. (3.31). Moreover, the model of (3.31) is equivalent to the model (3.25) if we define the nonlinearities as $g_i(x) = -G_i^{-1}(x, y_i)$ (y_i is a realization/sample of Y_i). This simple transformation enables us to work, e.g., with observations contaminated with multiplicative noise [34, Chapter 9],[69].

3.7.2 Class 2: marginal potential $\vartheta - \log[\vartheta]$

The standard ARS algorithm can be interpreted as a method for sampling from pdf's of the form $p_o(x) \propto \exp\{-h(x)\}$, where $h(x)$ is a convex function. From a similar perspective, the proposed GARS algorithm can handle target pdf's of the form

$$
\begin{aligned}
p_o(x) &\propto h(x) \exp\{-h(x)\} \\
&= \exp\{-h(x) + \log[h(x)]\},
\end{aligned}
\tag{3.32}
$$

where the function $h(x)$ can be either convex or concave. In fact, in this case we can write $-\log[p_o(x)]$ as a composition of two functions, $\bar{V}_1 \circ g_1$, where $\bar{V}_1(\theta_1) \triangleq \vartheta_1 - \log[\vartheta_1]$ (which is convex with a minimum at $\mu_1 = 1$) and $g_1(x) \triangleq h(x)$. In Chapter 5, we extend this class of pdf's (see Example 2).

3.7.3 Class 3: polynomial potentials

In this subsection, we provide guidelines to apply the GARS technique when $V(x; \mathbf{g})$ is a polynomial function. Indeed, in some cases, a polynomial can be decomposed in a suitable way that eases the algorithm derivation. Obviously, there are many possibilities and here we show just two simple examples.

4th order polynomials

Consider a generic polynomial potential of 4th order,

$$V(x; \mathbf{g}) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4, \tag{3.33}$$

with $a_4 > 0$. This can be always written as

$$\begin{aligned} V(x; \mathbf{g}) &= \kappa + (\alpha + \beta x + \gamma x^2)^2 + (\delta + \eta x)^2 \\ &= \kappa + \bar{V}_1(g_1(x)) + \bar{V}_2(g_2(x)), \end{aligned} \tag{3.34}$$

where κ, α, β, γ, η, δ are real constants, $\bar{V}_i(\vartheta_i) = \vartheta_i^2$, $i = 1, 2$, $g_1(x) \triangleq \alpha + \beta x + \gamma x^2$ 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 GARS procedure.

The constants κ, α, β, γ, η and δ have to satisfy the following equalities

$$\begin{cases} \gamma^2 = a_4, \\ 2\beta\gamma = a_3, \\ \beta^2 + 2\alpha\gamma + \eta^2 = a_2, \\ 2\alpha\beta + 2\delta\eta = a_1, \\ \alpha^2 + \delta^2 + \kappa = a_0. \end{cases} \tag{3.35}$$

this is a nonlinear system of 5 equations and 6 unknowns that can be always solved if we assume $a_4 > 0$.

8th order polynomials

Consider now a polynomial potential of 8th order,

$$V(x; \mathbf{g}) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_6 x^6 + a_8 x^8, \tag{3.36}$$

where the coefficients corresponding to the powers $5, 7$ are null, i.e., $a_5 = a_7 = 0$. Moreover, if $a_2, a_6, a_8 > 0$, we can rewrite the polynomial in Eq. (3.36) as

$$\begin{aligned} V(x; \mathbf{g}) &= \kappa + \left(\frac{a_1}{2\sqrt{a_2}} + \sqrt{a_2}x \right)^2 + \left(\frac{a_3}{2\sqrt{a_6}} + \sqrt{a_6}x^3 \right)^2 + \left(\frac{a_4}{2\sqrt{a_8}} + \sqrt{a_8}x^4 \right)^2 \\ &= \kappa + \bar{V}_1(g_1(x)) + \bar{V}_2(g_2(x)) + \bar{V}_3(g_3(x)), \end{aligned} \tag{3.37}$$

where

$$\kappa = a_0 - \frac{a_1^2}{2a_2} - \frac{a_3^2}{2a_6} - \frac{a_4^2}{2a_8}, \tag{3.38}$$

and $\bar{V}_i(\vartheta_i) = \vartheta_i^2$, $i = 1, 2, 3$, $g_1(x) \triangleq \frac{a_1}{2\sqrt{a_2}} + \sqrt{a_2}x$ is already linear, $g_2(x) \triangleq \frac{a_3}{2\sqrt{a_6}} + \sqrt{a_6}x^3$ is concave when $x < 0$ and convex when $x > 0$ and $g_3(x) \triangleq \frac{a_4}{2\sqrt{a_8}} + \sqrt{a_8}x^4$ is always convex.

It is important to remark that although the second derivative of $g_3(x)$ has not a constant sign, it is possible to apply the GARS procedure because we know the inflection points ($x = 0$ in this case). We discuss this issue in detail in Chapter 5 (Section 5.2.1).

3.8 Summary

We have proposed a novel adaptive rejection sampling scheme that can be used to draw exactly from a certain family of pdf's, not necessarily log-concave and possibly multimodal. The new method is a generalization of the classical adaptive rejection sampling scheme of [50], and includes it as a particular case as shown in Section 3.3. The proposed algorithm constructs a sequence of proposal pdf's that converge towards the target density and, therefore, can attain very high acceptance rates. We have discussed the asymptotic convergence of the constructed proposal pdf's in Section 3.4.

An alternative continuous, but more complicated, construction of the proposal pdf's has been introduced in Section 3.5. Finally, the applicability of the introduced GARS algorithm has been discussed in Section 3.7, where some examples are provided, including posterior pdf's for Bayesian inference and polynomial potentials.

The proposed technique can also be extended in different ways, as we show in Chapter 5. For example, it is also possible to build easily a lowerbounding function of the target pdf, in order to apply the squeeze principle (discussed in Section 2.5.2) and simplify the test for acceptance of candidate samples. Additionally, we have only tackled the log-transformation to introduce the potential function $V(x; \mathbf{g}) = -\log[p(x)]$ but, in general, we can also consider more general approaches, for instance, with monotonic T-transformations as described in [61]. Moreover, the assumptions over the nonlinearities g_i and $\bar{V}_i$ can be relaxed in order to extend the classes of target pdf's that can be tackled and to render more automatic as possible the the GARS method.

This problem is tackled in Chapter 5.

It has also been shown that in some specific cases the new GARS algorithm may yield improper proposal densities, namely when the tails of $p_o(x)$ are log-convex. We address this problem in the following chapter and introduce two alternative strategies to overcome it.

Chapter 4

GARS for densities with log-convex tails

As shown in Section 3.6, the GARS technique fails when the target pdf has log-convex tails and an infinite domain $\mathcal{D}$. In this chapter, 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 classical techniques such as the standard ARS algorithm of Section 2.6 [50], the TDR technique of Section 2.7.3 [61], the concave-convex ARS algorithm of Section 2.7.2 [55] and other methods [36, 113, 114] are invalid.

The first adaptive scheme introduced below, is easy to implement and provides good performance (as shown by the numerical examples of Sections 6.2.1 and 6.2.2). However, it presents some technical requirements that can prevent its use with certain densities.

The second proposed approach is more general and it is based on the ratio of uniforms (RoU) technique [30, 79, 153] described in Chapter 2. 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 $\mathcal{A}$. Assuming that $\mathcal{A}$ is bounded, 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 recall that a basic adaptive rejection sampling scheme based on the RoU technique was already introduced in [92, 93] but it only works when the region $\mathcal{A}$ is strictly

convex. It can be seen as a RoU-based counterpart of the original ARS algorithm in [50], that requires the log-density to be concave. The adaptive scheme that we introduce can also be used with non-convex sets.

4.1 The difficulty of handling log-convex tails

Given a target pdf of the form

$$p_o(x) \propto \exp\{-V(x; \mathbf{g})\}, \tag{4.1}$$

with $x \in \mathcal{D} \subseteq \mathbb{R}$. The GARS algorithm introduced in Chapter 3 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, $\int_{-\infty}^{+\infty} \exp\{-W_t(x)\}dx = +\infty$ and a proper pdf of the form $\pi_t(x) \propto \exp\{-W_t(x)\}$ does not exist.

This difficulty with the tails is actually shared by all adaptive rejection sampling techniques in the literature. A theoretical solution to the problem is to find an invertible transformation $G : \mathcal{D} \to \mathcal{D}^*$, where $\mathcal{D}^* \subset \mathbb{R}$ is a bounded set [30, 63, 106, 154]. Indeed, consider a r.v. X with pdf $p_o(x)$. Then, we can define a random variable $Y = G(X)$ with density $q(y)$, draw samples $y^{(1)}, \ldots, y^{(N)}$ and then convert them into samples $x^{(1)} = G^{-1}(y^{(1)}), \ldots, x^{(N)} = G^{-1}(y^{(N)})$ 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 [36], also described in Section 2.7.3. 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 $T_i : \mathcal{D}_i \to \mathbb{R}$, $i = 1, ..., m$, 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 [36] requires the ability to compute all the inflection points of the target function $p(x)$.

4.2   GARS with log-convex tails

4.2.1   Algorithm

In this section, we investigate a variant of standard GARS strategy of Chapter 3 to obtain an adaptive rejection sampling algorithm that remains valid when the tails of the potential function $V(x; \mathbf{g})$ are concave (i.e, when the target pdf $p_o(x)$ has log-convex tails).

Consider a target pdf $p_o(x)$ of the type in Eq. (4.1), with a potential function of the form

$$V(x; \mathbf{g}) = \sum_{i=1}^{n} \bar{V}_i(g_i(x)), \tag{4.2}$$

and also recall the set of support points

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

and the intervals $\mathcal{I}_0 = (-\infty, s_1]$, $\mathcal{I}_k = [s_k, s_{k+1}]$ for $k = 1, ..., m_t - 1$ and $\mathcal{I}_{m_t} = [s_{m_t}, +\infty)$. Let us assume that, for some $j \in \{1, \ldots, n\}$, the pdf defined as

$$q(x) \propto \exp\{-\bar{V}_j(g_j(x))\} \tag{4.3}$$

is such that:

(a) we can integrate $q(x)$ over the intervals $\mathcal{I}_0, \mathcal{I}_1, ..., \mathcal{I}_{m_t}$ 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

$$V_{-j}(x; \mathbf{g}) \triangleq \sum_{i=1, i \neq j}^{n} \bar{V}_i(g_i(x)), \tag{4.4}$$

attained by removing $\bar{V}_j(g_j(x))$ from $V(x; \mathbf{g})$ in Eq. (4.2). Assume for the moment that it is possible to calculate lower bounds for the reduced potential in every interval between two support points, i.e., we can compute bounds γ_k such that

$$\gamma_k \leq V_{-j}(x; \mathbf{g}),$$

for every $x \in \mathcal{I}_k$, $k = 0, ..., m_t$. Once these bounds are available, we set $L_k \triangleq \exp\{-\gamma_k\}$ and build the piecewise exponential proposal function

$$
\pi_t(x) \propto \begin{cases}
L_0 \exp\{-\bar{V}_j(g_j(x))\}, & \forall x \in \mathcal{I}_0, \\
\vdots \\
L_k \exp\{-\bar{V}_j(g_j(x))\}, & \forall x \in \mathcal{I}_k, \\
\vdots \\
L_{m_t} \exp\{-\bar{V}_j(g_j(x))\}, & \forall x \in \mathcal{I}_{m_t}.
\end{cases} \tag{4.5}
$$

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.

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

$$
\bar{\alpha}_k \triangleq L_k \int_{\mathcal{I}_k} q(x) dx \tag{4.6}
$$

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

$$
\pi_t(x) = \sum_{k=1}^{m_t} \alpha_k q(x) \mathbb{I}_{\mathcal{I}_k}(x) \tag{4.7}
$$

where $\mathbb{I}_{\mathcal{I}_k}(x)$ is an indicator function ($\mathbb{I}_{\mathcal{I}_k}(x) = 1$ if $x \in \mathcal{I}_k$ and $\mathbb{I}_{\mathcal{I}_k}(x) = 0$ if $x \notin \mathcal{I}_k$). The complete variant of the standard GARS algorithm, that uses the sequence of densities $\pi_t(x)$ of Eq. (4.5) as proposals, is summarized below.

1. **Initialization.** Set $i = 1$, $t = 0$ and choose m_1 support points, $\mathcal{S}_1 = \{s_1, ..., s_{m_1}\}$.

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 $j \in \{1, ..., n\}$.

- For $k = 0, \ldots, m_t$, 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 Section 4.2.2 for details).

- Calculate the weights $\bar{\alpha}_k$ in Eq. (4.6) and normalize them obtaining α_k, $k = 0, ..., m_t$.

- Draw x' from the proposal $\pi_t(x)$ in Eq. (4.7) and u' from $\mathcal{U}(0, 1)$. To generate x', it is necessary

  - draw an index k' with probability $\text{Prob}\{k' = k\} = \alpha_k$,
  - then draw x' from $q(x)$ restricted to the interval $\mathcal{I}_{k'} = [s_{k'}, s_{k'+1}]$.

- 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 the 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 4.1(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 4.1(b) depicts the target pdf $p_o(x)$ together 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 $j \in \{1, ..., n\}$, such that the pdf $q(x) \propto \exp\{-\bar{V}_j(g_j(x))\}$

- is analytically integrable in every interval $\mathcal{I}_k \subset \mathcal{D}$, since otherwise the weights $\alpha_1, ..., \alpha_{m_t}$ in Eq. (4.6)-(4.7) cannot be computed in general, and

- is easy to sample when truncated into a finite or an infinite interval, since otherwise we cannot draw from it.

The technique that we introduce in Section 4.3, based on the RoU method, overcomes these constraints. Before that Section 4.2.2 below describes the procedure to obtain the lower bounds γ_k, $k = 0, ..., m_t$.
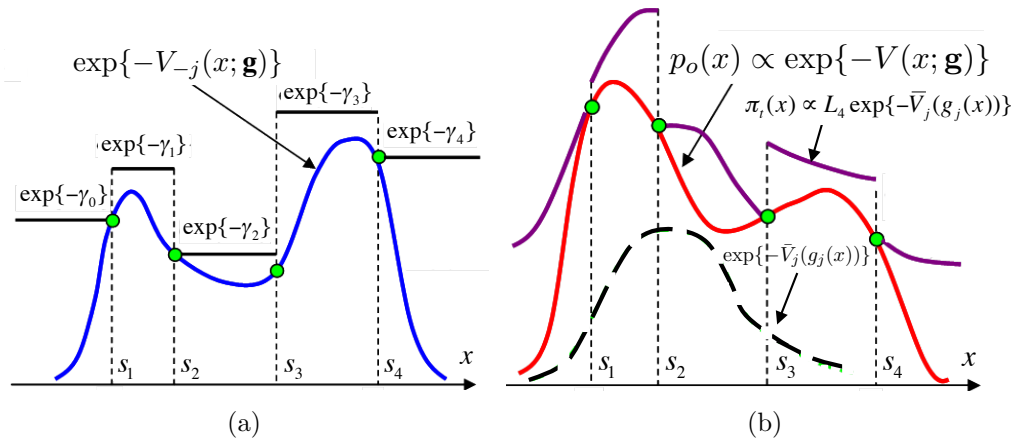
Figure 4.1: **(a)** Example of the function $\exp\{-V_{-j}(x; \mathbf{g})\}$ and its stepwise approximation $L_k = \exp\{-\gamma_k\}$, $k = 0, ..., m_t = 4$, constructed with the proposed technique using four support points $\mathcal{S}_t = \{s_1, s_2, s_3, s_4\}$. **(b)** Our target pdf $p_o(x) \propto \exp\{-V(x; \mathbf{g})\} = \exp\{-V_{-j}(x; \mathbf{g}) - \bar{V}_j(g_j(x))\}$ obtained by multiplying $\exp\{-V_{-j}(x; \mathbf{g})\} \times \exp\{-\bar{V}_j(g_j(x))\}$. 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 $L_k = \exp\{-\gamma_k\}$.

4.2.2 Calculation of lower bounds

We tackle the calculation of a lower bound

$$\gamma_k \leq \min_{x \in \mathcal{I}_k} V(x; \mathbf{g}), \tag{4.8}$$

in some interval $\mathcal{I}_k = [s_k, s_{k+1}]$. Note that if we are able to minimize analytically the modified potential $V(x; \mathbf{r}_k)$, then we trivially obtain

$$\gamma_k = \min_{x \in \mathcal{I}_k} V(x; \mathbf{r}_k) \leq \min_{x \in \mathcal{I}_k} V(x; \mathbf{g}), \tag{4.9}$$

and the problem is solved. Let us assume, however, that the analytical minimization of $V(x; \mathbf{r}_k)$ remains intractable. Since the modified potential

114

$V(x; \mathbf{r}_k)$ is convex $\forall x \in \mathcal{I}_k$, we can use a straight line $w_k(x)$ tangent 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

$$\gamma_k \triangleq \min\{w_k(s_k), w_k(s_{k+1})\}, \tag{4.10}$$

so that we can write

$$\gamma_k \leq \min_{x \in \mathcal{I}_k} V(x; \mathbf{r}_k) \leq \min_{x \in \mathcal{I}_k} V(x; \mathbf{g}). \tag{4.11}$$

Figure 4.2 sketches the procedure to obtain a lower bound in an interval $\mathcal{I}_k$ for a system potential $V(x; \mathbf{g})$ (blue) using a straight line tangent (black) to the modified potential $V(x; \mathbf{r}_k)$ (red dashed) at an arbitrary point $x^* \in \mathcal{I}_k$.
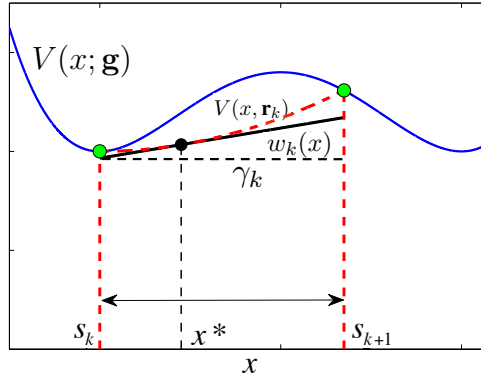
Figure 4.2: The picture shows the potential $V(x; \mathbf{g})$ (solid blue line), the modified potential $V(x; \mathbf{r}_k)$ (dashed red line) and the straight line $w_k(x)$ (solid black line) tangent to the modified potential at an arbitrary point x^* in $\mathcal{I}_k$. The lower bound γ_k is obtained as $\gamma_k = \min\{w_k(s_k), w_k(s_{k+1})\}$. In this picture, we have $\gamma_k = w_k(s_k)$.

Note that this procedure for the computation of a lower bound can be applied both to the complete potential $V(x; \mathbf{g})$ and to the reduced potential $V_{-j}(x; \mathbf{g})$, as needed in the previous section.

4.3 Adaptive RoU scheme

The RoU method described in Section 2.8 is a potentially useful tool to draw samples from a target pdf $p_o(x) \propto p(x) = \exp\{-V(x; \mathbf{g})\}$ with log-convex

tails. Indeed, we have seen that, as long as the tails of $p_o(x)$ decay as $1/x^2$ (or faster), the problem of drawing from $p_o(x)$ can be transformed into a problem of drawing uniformly from a bounded set $\mathcal{A} \subset \mathbb{R}^2$ that is obtained thorough a transformation of $p_o(x)$.

The main difficulty of this approach is that the boundary of the set $\mathcal{A}$ may take complicated forms, possibly intractable by analytical means, making it impossible to draw directly from this set. In this section, we show how to build a sequence of decreasing bounded sets $\mathcal{P}_1 \supseteq \mathcal{P}_2 \supseteq ...\mathcal{P}_t \supseteq ... \supseteq \mathcal{A}$ that contain and approximate $\mathcal{A}$. The sets $\mathcal{P}_t$ are regular (they can be partitioned into disjoint triangular regions) and, hence it is straightforward to draw uniformly from them. The draws (v, u) from $\mathcal{P}_t$ that belong to $\mathcal{A}$ are also uniform in $\mathcal{A}$, and hence $x = \frac{v}{u}$ are exact samples from $p_o(x)$. Moreover, as $\mathcal{P}_t \to \mathcal{A}$, the acceptance rate converges to 1.

In Section 4.3.1 we outline the proposed adaptive RoU (ARoU) algorithm. Some implementation issues are discussed in Section 4.3.2.

4.3.1   Algorithm

The notation for the description of the algorithm is very similar to the ARS method. After the $t - 1$-th iteration, we obtain a set of support points $\mathcal{S}_t = \{s_1, \ldots, s_{m_t}\} \subset \mathcal{D}$, sorted out in ascending order, $s_1 < s_2 < .... < s_{m_t}$. These points, in turn, yield a collection of intervals $\mathcal{I}_0 = (-\infty, s_1]$, $\mathcal{I}_{m_t} = [s_{m_t}, +\infty)$ and $\mathcal{I}_k = [s_k, s_{k+1}]$ for $k = 1, ..., m_t - 1$. We assume that the point 0 is contained in $\mathcal{S}_t$, i.e., there exists $j \in \{1, ..., m_t\}$ such that $s_j = 0$. As a consequence, for every interval $\mathcal{I}_k = [s_k, s_{k+1}]$ the points s_k and s_{k+1} are both either non-positive or non-negative. We also assume the ability to compute the upper bounds

1. $L_k^{(1)} \geq \sqrt{p(x)}$, for all $x \in \mathcal{I}_k$,

2. $L_k^{(2)} \geq x\sqrt{p(x)}$, for all $x \in \mathcal{I}_k \cap [0, +\infty)$,

3. $L_k^{(3)} \geq -x\sqrt{p(x)}$, for all $x \in \mathcal{I}_k \cap (-\infty, 0]$.

A computational procedure is given in Section 4.3.2. Below, we introduce the approximating sequence of the sets $\mathcal{P}_1 \supseteq \mathcal{P}_2 \supseteq ...\mathcal{P}_t \supseteq ... \supseteq \mathcal{A}$, then explain how to generate candidate samples using $\mathcal{P}_t$ and, finally, summarize the ARoU algorithm.

Construction of $\mathcal{P}_t$

Consider the construction in Figure 4.3(a). The consecutive support points $s_k, s_{k+1} \in \mathcal{S}_t$ yield the angles $\alpha_k \triangleq \arctan(s_k)$ and $\alpha_{k+1} \triangleq \arctan(s_{k+1})$, $k = 1, ..., m_t - 1$. We then 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 α_k and α_{k+1} w.r.t. the u axis. Additionally, the sets $\mathcal{A}_0$ and $\mathcal{A}_{m_t}$ are formed from the cones delimited by the angles $\alpha_0 = -\pi/2$, $\alpha_1 = \arctan(s_1)$, and $\alpha_{m_t} = \arctan(s_{m_t})$, $\alpha_{m_t+1} = \pi/2$. Note that, clearly, $\mathcal{A} = \cup_{k=0}^{m_t} \mathcal{A}_k$.

For each $k = 0, ..., m_t$ the subset $\mathcal{A}_k$ is contained in a piece of circle $\mathcal{C}_k$ ($\mathcal{A}_k \subseteq \mathcal{C}_k$) with center at $(0, 0)$, delimited by the angles α_k, α_{k+1} and radius

$$
r_k = \begin{cases} \sqrt{\left(L_k^{(1)}\right)^2 + \left(L_k^{(2)}\right)^2}, & \text{if } \quad s_k, s_{k+1} \geq 0 \\ \sqrt{\left(L_k^{(1)}\right)^2 + \left(L_k^{(3)}\right)^2}, & \text{if } \quad s_k, s_{k+1} \leq 0 \end{cases} \tag{4.12}
$$

also shown (with a dashed red line) in Fig. 4.3(a).

Unfortunately, it is not straightforward to generate samples uniformly from $\mathcal{C}_k$. However, we can easily draw samples uniformly from a triangle in the plane $\mathbb{R}^2$ (see Appendix B.1). Hence, we can choose an arbitrary point in the arc of circumference that delimits $\mathcal{C}_k$ (e.g., the point $(L_k^{(2)}, L_k^{(1)})$ in Fig. 4.3(a)) and calculate the straight line tangent 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 $k = 0, ..., m_t$ 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 4.3 summarizes the procedure to build the set $\mathcal{P}_t$. In Figure 4.3(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 the single interval $x \in \mathcal{I}_k = [s_k, s_{k+1}]$. Figure 4.3(b) illustrates the entire region $\mathcal{P}_t \triangleq \cup_{k=0}^{5} \mathcal{T}_k$ formed by $m_t + 1 = 6$ triangular subsets that covers completely the region $\mathcal{A}$, i.e., $\mathcal{A} \subseteq \mathcal{P}_t$.
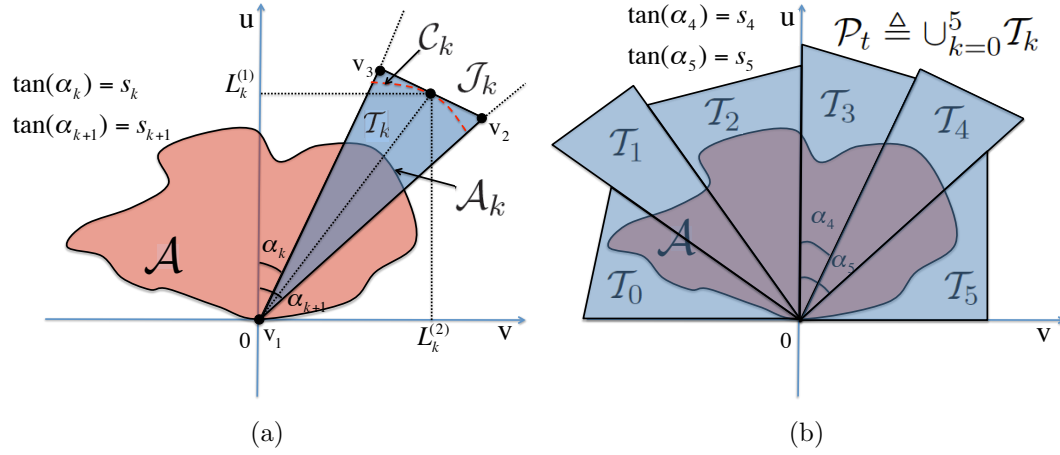
Figure 4.3: **(a)** A region $\mathcal{A}$ constructed by the RoU method and a triangular region $\mathcal{T}_k$ defining by the vertices $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$ built using the upper bounds $L_k^{(1)}$, $L_k^{(2)}$ for the functions $\sqrt{p(x)}$ and $x\sqrt{p(x)}$, $x \in \mathcal{I}_k = [s_k, s_{k+1}]$. The red 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 $\mathbf{v} = \theta_1 \mathbf{v}_1 + \theta_2 \mathbf{v}_2$ 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., $\mathcal{A} \subset \mathcal{P}_t$.

Adaptive sampling

To generate samples uniformly in $\mathcal{P}_t$, we first draw an index $k \in \{0, ..., m_t\}$ with probability proportional to the area $|\mathcal{T}_k|$, $k = 0, ..., m_t$. To be specific, we define the normalized weights

$$\delta_k \triangleq \frac{|\mathcal{T}_k|}{\sum_{i=0}^{m_t} |\mathcal{T}_i|}, \quad k = 0, ..., m_t, \tag{4.13}$$

and then choose a triangular piece by drawing an index $k' \in \{0, ..., m_t\}$ from the probability distribution $P(k) = \delta_k$. Using the procedure in Appendix B.1, we draw a point (v', u') uniformly from the triangular region $\mathcal{T}_{k'}$. If $(v', u') \in \mathcal{A}$, then we accept the sample $x' = v'/u'$ 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 we build $\mathcal{P}_{t+1} \subseteq \mathcal{P}_t$, that yields a tighter approximation of $\mathcal{A}$.

Summary of the algorithm

The ARoU algorithm to generate N samples from $p_o(x)$ is outlined below.

1. **Initialization.** Start with $i = 1$, $t = 0$ and choose m_1 support points, $\mathcal{S}_1 = \{s_1, ..., s_{m_1}\}$, with $0 \in \mathcal{S}_1$.

2. **Iteration.** For every $t \geq 1$, 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 $k = 0, ..., m_t$ (see Section 4.3.2).

   - Construct the triangular regions $\mathcal{T}_k$, $k = 0, ..., m_t$, as described in Section 4.3.1.

   - Compute the area $|\mathcal{T}_k|$ of every triangle and the normalized weights

     $$\delta_k \triangleq \frac{|\mathcal{T}_k|}{\sum_{i=0}^{m_t} |\mathcal{T}_i|}, \tag{4.14}$$

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

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

   - Generate a point (v', u') uniformly from the region $\mathcal{T}_{k'}$ (see Appendix B.1).

   - 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 $m_{t+1} = m_t$. If $i > N$, then stop the iteration.

   - 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 $t = t + 1$.

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

4.3.2 Computation of bounds

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 ARoU algorithm.

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

$$
\begin{aligned}
\sqrt{p(x)} &\propto \exp\big\{-V^{(1)}(x; \mathbf{g})\big\}, \\
x\sqrt{p(x)} &\propto \exp\big\{-V^{(2)}(x; \mathbf{g})\big\}, \quad (x > 0), \\
-x\sqrt{p(x)} &\propto \exp\big\{-V^{(3)}(x; \mathbf{g})\big\}, \quad (x < 0),
\end{aligned}
\tag{4.15}
$$

with

$$
\begin{aligned}
V^{(1)}(x; \mathbf{g}) &\triangleq \frac{1}{2} V(x; \mathbf{g}), \\
V^{(2)}(x; \mathbf{g}) &\triangleq \frac{1}{2} V(x; \mathbf{g}) - \log(x), \quad (x > 0), \\
V^{(3)}(x; \mathbf{g}) &\triangleq \frac{1}{2} V(x; \mathbf{g}) - \log(-x), \quad (x < 0),
\end{aligned}
\tag{4.16}
$$

respectively. 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 $V^{(j)}(x; \mathbf{g})$, $j = 1, 2, 3$, also w.r.t. x. As a consequence, we may focus on the calculation of lower bounds $\gamma_k^{(j)} \leq V^{(j)}(x; \mathbf{g})$, $x \in \mathcal{I}_k$, which are related to the upper bounds as $L_k^{(j)} = \exp\{-\gamma_k^{(j)}\}$, $j = 1, 2, 3$ and $k = 0, ..., m_t$. This problem is far from trivial, though. Even for very simple marginal potentials, $\bar{V}_i$, $i = 1, \ldots, n$, the potential functions, $V^{(j)}$, $j = 1, 2, 3$, can be highly multimodal w.r.t. x [114].

In Section 4.2.2 we describe a procedure to find a lower bound γ_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 potential function $V(x; \mathbf{g})$. 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 Section 4.2.2 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, we can use the modified potentials $V(x; \mathbf{r}_k)$ to build

$$
\begin{aligned}
V^{(2)}(x; \mathbf{r}_k) &\triangleq \frac{1}{2} V(x; \mathbf{r}_k) - \log(x), \quad \text{and} \\
V^{(3)}(x; \mathbf{r}_k) &\triangleq \frac{1}{2} V(x; \mathbf{r}_k) - \log(-x).
\end{aligned}
\tag{4.17}
$$

for all $x \in \mathcal{I}_k$. Clearly, both $V^{(2)}(x; \mathbf{r}_k)$ and $V^{(3)}(x; \mathbf{r}_k)$ are convex in $\mathcal{I}_k$. Hence, it is straightforward to obtain lower bounds $\gamma_k^{(2)} \leq V^{(2)}(x; \mathbf{r}_k) \leq V^{(2)}(x; \mathbf{g})$ and $\gamma_k^{(3)} \leq V^{(3)}(x; \mathbf{r}_k) \leq V^{(3)}(x; \mathbf{g})$, as explained in Section 4.2.2. The corresponding upper bounds are $L_k^{(j)} = \exp\{-\gamma_k^{(j)}\}$, $j = 1, 2, 3$, for all $x \in \mathcal{I}_k$.

4.3.3 Heavier tails

It is possible to generalize the RoU method [26] shown in Section 2.8.1, in order to state that if we draw a point (v', u') uniformly from the set

$$\mathcal{A}_\rho = \left\{ (v, u) : 0 \leq u \leq [p(v/u^\rho)]^{1/(\rho+1)} \right\}, \tag{4.18}$$

then $x = \frac{v'}{u'^\rho}$ is a sample form $p_o(x) \propto p(x)$. The cases of interest are those in which $\mathcal{A}_\rho$ is a bounded set, and it can be shown that $\mathcal{A}_\rho$ is bounded when the tails of $p(x)$ decay as $1/x^{(\rho+1)/\rho}$ or faster. Therefore, for $\rho > 1$ we can extend the RoU method to deal with heavy-tailed distributions.

The ARoU algorithm can be extended in a similar way. The potentials associated to the target pdf $p_o(x) \propto \exp\{-V(x; \mathbf{g})\}$ become

$$
\begin{aligned}
V^{(1)}(x; \mathbf{g}) &\triangleq \frac{1}{\rho + 1} V(x; \mathbf{g}), \\
V^{(2)}(x; \mathbf{g}) &\triangleq \frac{\rho}{\rho + 1} V(x; \mathbf{g}) - \log(x), \text{ for } x > 0 \\
V^{(3)}(x; \mathbf{g}) &\triangleq \frac{\rho}{\rho + 1} V(x; \mathbf{g}) - \log(-x), \text{ for } x < 0,
\end{aligned}
\tag{4.19}
$$

and we can use the technique of Section 4.2.2 to obtain the necessary lower bounds. Moreover, the constant parameter ρ can be selected to maximize the acceptance rate.

4.4 Position of the stationary points

In this section, we present a result that can be useful to initialize (i.e., to choose adequately and automatically the initial set of support points $\mathcal{S}_0$) the two strategies introduced in this chapter. Namely, it can be used to identify automatically the tails of $p_o(x)$.

Indeed, it is possible to prove that all the stationary points of the potential $V(x; \mathbf{g})$ and, as a consequence, all the stationary points of the target pdf $p_o(x) \propto \exp\{-V(x; \mathbf{g})\}$, are contained within the interval defined by the minimum and maximum value of the simple estimates, i.e.,

$$x_{min} \triangleq \min\left\{x \in \cup_{i=1}^{n} \mathcal{X}_i\right\}, \tag{4.20}$$

and

$$x_{max} \triangleq \max\left\{x \in \cup_{i=1}^{n} \mathcal{X}_i\right\}, \tag{4.21}$$

where

$$\mathcal{X}_i = \{x_i \in \mathbb{R} : \quad g_i(x_i) = \mu_i\}, \tag{4.22}$$

$i = 1, ..., n$, contains the simple estimates corresponding to the nonlinearity $g_i(x)$. Specifically, the potential $V(x; \mathbf{g})$ is strictly decreasing for $x \leq x_{min}$ and strictly increasing for $x \geq x_{max}$. The following proposition summarizes this result.

Proposition 2 *All the stationary points of the potential $V(x; \mathbf{g})$, i.e., every $\hat{x}$ that belongs to the set*

$$\hat{\mathcal{X}} \triangleq \left\{\hat{x} \in \mathcal{D} : \left.\frac{dV}{dx}\right|_{x=\hat{x}} = 0\right\}, \tag{4.23}$$

are contained in the interval $[x_{min}, x_{max}]$, where x_{min} and x_{max} are defined in Eqs. (4.20) and (4.21), respectively. Namely, we have

$$\hat{\mathcal{X}} \subseteq [x_{min}, x_{max}]. \tag{4.24}$$

Proof: We have to prove that the first derivative of the potential function $V(x; \mathbf{g})$ is

$$\frac{dV}{dx} < 0, \quad \text{for all} \quad x < x_{min}, \tag{4.25}$$

and

$$\frac{dV}{dx} > 0, \quad \text{for all} \quad x > x_{max}, \tag{4.26}$$

so that all stationary points of V stay inside $[x_{min}, x_{max}]$. Routine calculations yield the derivative of $V(x; \mathbf{g})$,

$$\frac{dV}{dx} = \sum_{i=1}^{n} \frac{d\vartheta_i}{dx} \left.\frac{d\bar{V}_i}{d\vartheta_i}\right|_{\vartheta_i = g_i(x)}, \tag{4.27}$$

and we aim to evaluate it outside the interval $[x_{min}, x_{max}]$.

To do it, first we recall that we assume nonlinearities $g_i(x)$, $i = 1, ..., n$, with constant concavity ($\frac{d^2 g_i}{dx^2}$ with constant sign $\forall x \in \mathcal{D}$). This implies that $g_i(x)$ can be monotonic, or non-monotonic with one minimum or one maximum. Then, we also have to notice that:

1. If $g_i(x)$ is a monotonic function then the first derivative $\frac{dg_i}{dx}$ has also constant sign $\forall x \in \mathcal{D}$.

2. If $g_i(x)$ is non-monotonic and we denote $x_{i,1}$ and $x_{i,2}$ ($x_{i,1} < x_{i,2}$) the two simple estimates (i.e., $\mu_i = g_i(x_{i,1})$ and $\mu_i = g_i(x_{i,2})$), the first derivative $\frac{dg_i}{dx}$ has constant sign for all $x < x_{i,1}$ and $x > x_{i,2}$ (i.e., $x \notin \mathcal{J}_i = [x_{i,1}, x_{i,2}]$). Indeed, the first derivative $\frac{dg_i}{dx}$ changes sign in the interval $\mathcal{J}_i = [x_{i,1}, x_{i,2}]$ that contains the minimum or the maximum of the non-monotonic $g_i(x)$.

Moreover, note that each interval $\mathcal{J}_i$ is included in $[x_{min}, x_{max}]$, i.e., $\mathcal{J}_i \subseteq [x_{min}, x_{max}]$, $i = 1, ..., n$. Hence, for any type of $g_i(x)$ we can assert that $\frac{dg_i}{dx}$ has constant sign $\forall x \notin [x_{min}, x_{max}]$. Let us consider the cases $\frac{dg_i}{dx} > 0$ and $\frac{dg_i}{dx} < 0$, separately, first for $x \leq x_{min}$ and then for $x \geq x_{max}$.

- $\frac{dg_i}{dx} > 0$: since for every simple estimate x_i we have $x_i \geq x_{min}$, $\mu_i = g_i(x_i)$, and since $\frac{dg_i}{dx} > 0$, we can write

$$\mu_i = g_i(x_i) \geq g_i(x_{min}) > g_i(x), \text{ for all } x < x_{min}.$$

Then $\mu_i > g_i(x)$, for all $x < x_{min}$, and due to the marginal potential $\bar{V}_i(\vartheta_i)$ being convex with minimum at μ_i, we have

$$\left. \frac{d\bar{V}_i}{d\vartheta_i} \right|_{\vartheta_i = g_i(x) < \mu_i} < 0,$$

for all $i = 1, ..., n$ and $\forall x < x_{min}$. As a consequence,

$$\frac{dV}{dx} = \sum_{i=1}^{n} \underbrace{\frac{dg_i}{dx}}_{>0} \underbrace{\left. \frac{d\bar{V}_i}{d\vartheta_i} \right|_{\vartheta_i = g_i(x)}}_{<0} < 0, \quad \forall x < x_{min}.$$

123

- $\frac{dg_i}{dx} < 0$: since for every simple estimate x_i we have $x_i \geq x_{min}$, $\mu_i = g_i(x_i)$, and since $\frac{dg_i}{dx} < 0$, we can write

$$\mu_i = g_i(x_i) \leq g_i(x_{min}) < g_i(x), \text{ for all } x < x_{min}.$$

Then $\mu_i < g_i(x)$, for all $x < x_{min}$, and since the marginal potential $\bar{V}_i(\vartheta_i)$ is convex with minimum at μ_i, we have

$$\frac{d\bar{V}_i}{d\vartheta_i}\bigg|_{\vartheta_i = g_i(x) > \mu_i} > 0,$$

for all $i = 1, ..., n$ and $\forall x < x_{min}$. As a consequence,

$$\frac{dV}{dx} = \sum_{i=1}^{n} \underbrace{\frac{dg_i}{dx}}_{<0} \underbrace{\frac{d\bar{V}_i}{d\vartheta_i}\bigg|_{\vartheta_i = g_i(x)}}_{>0} < 0, \quad \forall x < x_{min}.$$

Since in both cases $\frac{dV}{dx} < 0$, we can assert that

$$\frac{dV}{dx} < 0 \text{ for all } x < x_{min}. \tag{4.28}$$

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

This property can be used to identify the tails of the target pdf $p_o(x) \propto \exp\{-V(x; \mathbf{g})\}$ and to handle them separately. However, if no simple estimates exist, i.e., $|\mathcal{X}_i| = 0$ for all $i = 1, ..., n$, the proposition above becomes of no use for this purpose.

4.5   Summary

Many strategies presented in the literature (ARS, TDR, Concave-convex ARS, GARS, etc....) [36, 50, 55, 93, 113] break down when the tails of the target pdf are log-convex[1] in an infinite domain. In this chapter, we have proposed two adaptive rejection sampling schemes that can be used to draw exactly from a large family of pdf's, not necessarily with log-concave tails. Probability distributions of this class appear in many inference

---

[1] or T-convex, more in general.

problems as, for example, localization in sensor networks [90, 113, 114], stochastic volatility [34, Chapter 9], [69] or hierarchical models [51, Chapter 9], [37, 42, 43]. The new methods yield a sequence of proposal pdf's that converge towards the target density and, as a consequence, can attain high acceptance rates. Since they can be applied to multimodal densities also when the tails are log-convex, these new techniques have a broad applicability compared to the related methods in the literature.

The first adaptive approach in Section 4.2 is easier to implement. However, it needs to identify a suitable pdf $q(x) \propto \exp\{-\bar{V}_j(g_j(x))\}$ with $j \in \{1, ..., n\}$, that in general can be a difficult task. One type of pdf's for which the identification of $q(x)$ can turn out natural includes the *a posteriori* density of a random variable X given a collection of observations (see Section 3.7.1). In this case, it may often be simple to identify $q(x)$ with the prior of X and $\exp\{-V_{-j}(x; \mathbf{g})\}$ with the likelihood.

The proposed adaptive RoU technique is more general than the first scheme, as also shown in the application to a stochastic volatility model in Chapter 6 (Section 6.2.3). Indeed, it can be always applied for all the target pdf's of the form $p_o(x) \propto \exp\{-\sum_{i=1}^{n} \bar{V}_i(g_i(x))\}$. Moreover, using the adaptive RoU scheme to generate a candidate sample, we only need to draw two uniform random variates although, in exchange, we have to find bounds for three (similar) potential functions.

In Section 4.4 we have proved that all the stationary points of the target pdf $p_o(x)$ are included in an interval delimited by the minimum and the maximum value of the simple estimates. This result can be used to identify automatically the tails of $p_o(x)$.

In the next chapter, we extend the class of pdf's that can be tackled with the GARS algorithm introduced in the previous Chapter 3 and also with the two adaptive techniques presented in this chapter. For instance, we relax the assumptions about the marginal potentials $\bar{V}_i$ and the nonlinearities g_i.

Chapter 5

Extensions and enhancements

In Chapter 3, we have introduced the standard version of the GARS technique. The basic assumptions for the development of the method are that

- the target pdf has the form $p_o(x) \propto \exp\left\{-\sum_{i=1}^n \bar{V}_i(g_i(x))\right\}$ where

- the marginal potentials $\bar{V}_i(\vartheta_i)$ are all convex with a minimum μ_i, $i = 1, ..., n$, and

- the nonlinearities $g_i(x)$ are all either convex or concave.

In this chapter, we study techniques that enable us to relax some of these assumptions. Specifically, in Section 5.1, we apply the GARS method to target densities that may not belong of the exponential family. Then, in Section 5.2, we discuss how the assumptions about the marginal potentials $\bar{V}_i$ and the nonlinearities g_i can be relaxed.

Finally, we introduce a simpler automatic implementation of the basic GARS strategy in Section 5.3. This alternative GARS procedure requires only the ability to evaluate the potential function $V(x; \mathbf{g})$, i.e., the knowledge of the simple estimates and of the minima μ_i, $i = 1, ..., n$, is not necessary. The need to compute the first derivatives of the g_i's (and the $\bar{V}_i$'s) is also removed in Section 5.3.3.

5.1 Non-exponential target densities

5.1.1 GARS with T-transformation

So far, in Chapters 3 and 4 we have considered target pdf's of form

$$p_o(x) \propto \exp\left\{ -\sum_{i=1}^{n} \bar{V}_i(g_i(x)) \right\}.$$

However, the GARS procedure can also be applied to target pdf's of the more general type

$$p_o(x) \propto p(x) = T^{-1}\left[-\sum_{i=1}^{n} \bar{V}_i(g_i(x)) \right], \tag{5.1}$$

where T is some transformation that satisfies the following conditions (as shown in Section 2.7.3):

1. T is monotonically increasing, hence invertible.

2. Given the inverse transformation T^{-1}, the integral $\int_{-\infty}^{\vartheta} T^{-1}(t)dt$ must be bounded for ϑ in the image of T. Moreover, it must be possible to analytically obtain the integral

$$\int_{a}^{b} T^{-1}(t)dt, \quad a,b \in \mathbb{R}. \tag{5.2}$$

Then, we can use the two usual two steps of the GARS technique in order to build:

1. Suitable linear functions $r_{i,k}(x)$, $i = 1, ..., n$, in order to replace $g_i(x)$, $\forall x \in \mathcal{I}_k$, and to obtain the modified potentials $V(x; \mathbf{r}_k)$, $k = 0, ..., m_t$.

2. The piecewise linear function $W_t(x)$ composed by straight lines tangent to each potential $V(x; \mathbf{r}_k)$, $\forall x \in \mathcal{I}_k$ and $k = 0, ..., m_t$.

The resulting proposal pdf has the form

$$\pi_t(x) \propto T^{-1}[-W_t(x)]. \tag{5.3}$$

5.1.2 GARS and the TDR method

The transformed density rejection (TDR) technique of Section 2.7.3 enables us to draw from target pdf's of the type

$$p_o(x) \propto T^{-1}[g(x)] = (T^{-1} \circ g)(x),$$

where $T^{-1}(\vartheta)$ is monotonically increasing and $g(x)$ is concave.[1]

The GARS technique can be expressed in a similar form, that extends the applicability of the TDR method to non-monotonic transformations. Indeed, let us assume that we are able to draw from the pdf

$$f(\vartheta) \propto H(\vartheta), \tag{5.4}$$

$\forall \vartheta \in \mathcal{C} \subseteq \mathbb{R}$, with a single mode[2] at μ. Furthermore, consider a target density of the form

$$p_o(x) = f(g(x)) = (f \circ g)(x), \tag{5.5}$$

where $g(x)$ is either a concave or a convex function (or a general function with known inflection points, see Section 5.2.1). Hence, the target pdf is a distribution generated by a transformation of scale $g(x)$ of the pdf $f(\vartheta)$ [74].

The first step of the standard GARS procedure can be used to replace the nonlinearity $g(x)$ with suitable linear functions $r_k(x) = a_k x + b_k$, for all $x \in \mathcal{I}_k$ and $k = 0, ..., m_t$, such that

$$H(r_k(x)) = H(a_k x + b_k) \geq H(g(x)), \tag{5.6}$$

in order to use

$$\pi_t(x) \propto H(r_k(x))$$

as the proposal pdf in a RS scheme. Since $f(\vartheta) \propto H(\vartheta)$, it is important to remark that $\pi_t(x) \propto f(r_k(x)) \propto H(r_k(x))$ is a linearly-scaled version of the pdf $f(\vartheta)$, hence if we can draw from f then we can generate samples also from $\pi_t(x)$. We recall that if the function H is monotonic, then we go back

---

[1]Table 2.13 in Section 2.7.3 summarizes the other three possible cases where the technique is applicable: $T^{-1}(\vartheta)$ increasing and $g(x)$ convex, $T^{-1}(\vartheta)$ decreasing and $g(x)$ concave and finally $T^{-1}(\vartheta)$ decreasing and $g(x)$ convex. Note that in all the four cases $T^{-1}(\vartheta)$ is a monotonic function.

[2]However, these considerations can also be extended when $f(\vartheta)$ has several modes, following the ideas in Section 5.2.5.

to the scenario in Section 2.7.3, where $H = T^{-1}$ (with either $\mu \to +\infty$ if H is increasing, or $\mu \to -\infty$ if H is decreasing).

The construction of the necessary linear functions is very similar to the standard algorithm. Given a set of support points

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

we aim to find a linear function $r_k(x)$ in the interval $\mathcal{I}_k = [s_k, s_{k+1}]$ such that $H(r_k(x)) \geq H(g(x))$. An example with $m_t = 3$ support points and a convex nonlinearity $g(x)$ is shown in Figure 5.1. The simple estimates (shown with squares) are calculated as solutions of the equation $\mu = g(x)$. The straight lines $r_0(x)$ and $r_3(x)$ are tangent to $g(x)$ at s_1 and s_3, respectively, while $r_1(x)$ and $r_2(x)$ are secant lines.
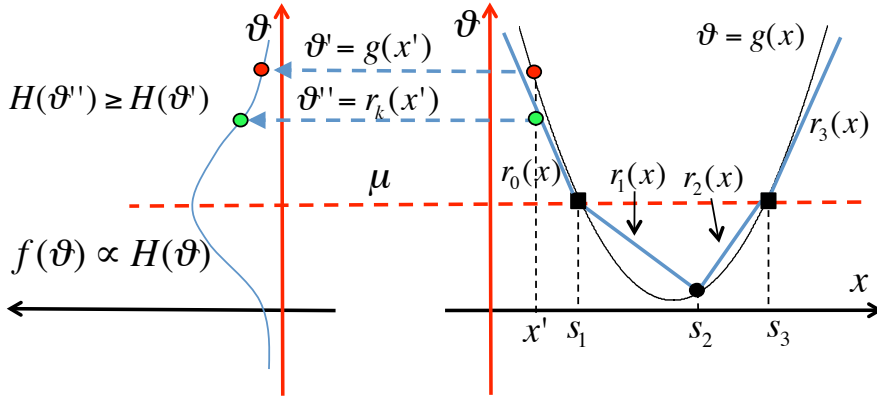
Figure 5.1: Example of construction of the linear functions $r_k(x)$, $k = 0, ..., m_t = 3$ for a convex nonlinearity $g(x)$ and a function $H(\vartheta)$ with a single mode at μ. The simple estimates s_1 and s_3 are solutions of the equation $\mu = g(x)$.

The construction is valid because the inequality in Eq. (5.6) is satisfied. Indeed, arbitrarily choosing x' and taking the values $\vartheta' = g(x')$ and $\vartheta'' = r_k(x')$ we always have that $H(\vartheta'') \geq H(\vartheta')$.

Therefore, the GARS method can be interpreted as a technique to use rejection sampling for target densities generated by a transformation of scale [74]. Here, we have considered target pdf's which are scaled versions of a pdf $f(\vartheta)$ with a single mode μ, but the GARS technique can also be applied when $f(\vartheta)$ has several modes (the underlying idea is the same used in Section 5.2.5).

130

5.2 Extensions of the standard GARS procedure

In Chapters 3 and 4, we have considered nonlinearities $g_i(x)$ with second derivatives with constant sign and we have assumed convex marginal potentials $\bar{V}_i(\vartheta_i)$, $i = 1, ..., n$. In the following sections, we address how we can relax these assumptions and use the GARS technique with more general nonlinearities and marginal potentials.

5.2.1 General nonlinearities $g_i(x)$

Let us consider a generic (almost everywhere) continuous nonlinearity $g_i(x)$, defined in $\mathcal{D}$, with second derivative with non-constant sign. In this case, it is not possible to apply directly the proposed GARS method of Chapter 3. However, we can extend the proposed methodology to this more general case if the inflection points of g_i are available.

Indeed, if we are able to calculate the inflection points $d_{i,j}$, $j = 1, ..., q_i$, of each nonlinearity $g_i(x)$ with $x \in \mathcal{D}_i$, $i = 1, ..., n$, then we can also find a partition $\mathcal{D}_i = \cup_{j=1}^{q_i}[\mathcal{D}_{i,j}]$ (where $[\cdot]$ denotes the closure of an interval) such that $\mathcal{D}_{i,j} \cap \mathcal{D}_{i,k} = \emptyset$, $\forall j \neq k$, and such that $\frac{d^2 g_i}{dx^2}$ has constant sign in every $\mathcal{D}_{i,j}$, $i = 1, ..., n$ and $j = 0, \ldots, q_i$. This information can be incorporated into the initial set of support points $\mathcal{S}_0$ and apply the standard GARS algorithm.

Specifically, let $\mathcal{D}_{i,j} = (d_{i,j}, d_{i,j+1})$. If we let $d_{i,j}, d_{i,j+1} \in \mathcal{S}_0 = \{s_1, \ldots, s_{m_0}\}$ (in addition to the simple estimates, as indicated in Section 3.2.4), 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 exactly as described in Section 3.2.

Figure 5.2(a) displays an example of a generic nonlinearity with 3 inflection points, shown with red triangles. In this case, the initial set of support points is $\mathcal{S}_0 = \{d_{i,1}, x_{i,1}, x_{i,2}, d_{i,2}, d_{i,3}\}$ ($m_0 = 4$), including the inflection points and the simple estimates. Figure 5.2(b) illustrates an example of construction of the suitable linear functions $r_{i,k}(x)$ with $m_t = 7$ support points.

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 [36]. For example, in the positioning application of Chapter 6 (Section 6.1.4) we can easily find the inflection points of each $g_i(x)$, $i = 1, \ldots, n$, but the analysis over the whole log-density function is intractable (so that the method in [36]

cannot applied).

(a) (b)
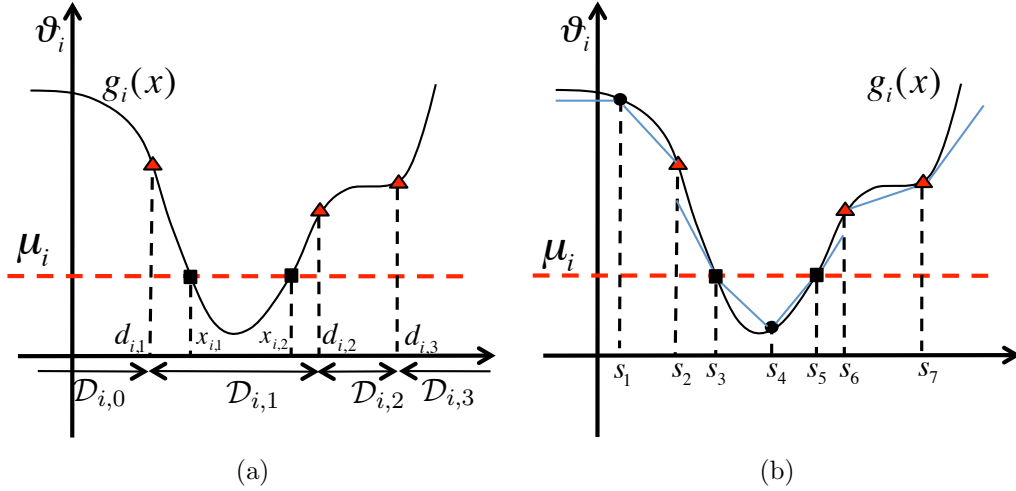
Figure 5.2: **(a)** Example of a generic nonlinearity $g_i(x)$ with three inflection points $d_{i,1}$, $d_{i,2}$ and $d_{i,3}$, shown with red triangles. In this case the initial set of support points $\mathcal{S}_0 = \{d_{i,1}, x_{i,1}, x_{i,2}, d_{i,2}, d_{i,3}\}$ ($m_0 = 5$) has to contain the two simple estimates $x_{i,1}, x_{i,2}$ (depicted with squares) and the inflection points. **(b)** Construction of the suitable linear functions $r_{i,k}(x)$, $k = 0, ..., 7$, with $m_t = 7$ support points, as described for the standard GARS procedure.

In the sequel, we discuss how we can relax some of the assumptions on the marginal potentials $\bar{V}_i(\vartheta_i)$, $i = 1, ..., n$.

5.2.2 Monotonic marginal potentials

The standard strategy in Chapter 3 can be easily extended for a monotonic (always convex) marginal potential $\bar{V}_i(\vartheta_i)$. Indeed, the procedure described in Section 3.2.3 to build the linear functions $r_{i,k}(x)$ can also be applied when some marginal potential $\bar{V}_i(\vartheta_i)$, $i \in \{1, ..., n\}$, is a monotonic function.

In particular, it is sufficient to assign either $\mu_i \rightarrow -\infty$ if $\bar{V}_i$ is monotonically increasing, or $\mu_i \rightarrow +\infty$ if $\bar{V}_i$ is monotonically decreasing.

Figures 5.3 illustrates how the procedure in Section 3.2.3 works in this case, considering $\mu_i \rightarrow \pm\infty$, for non-monotonic concave or convex nonlinearities $g_i(x)$. Clearly, the method can also be applied when the nonlinearity $g_i(x)$ is monotonic.
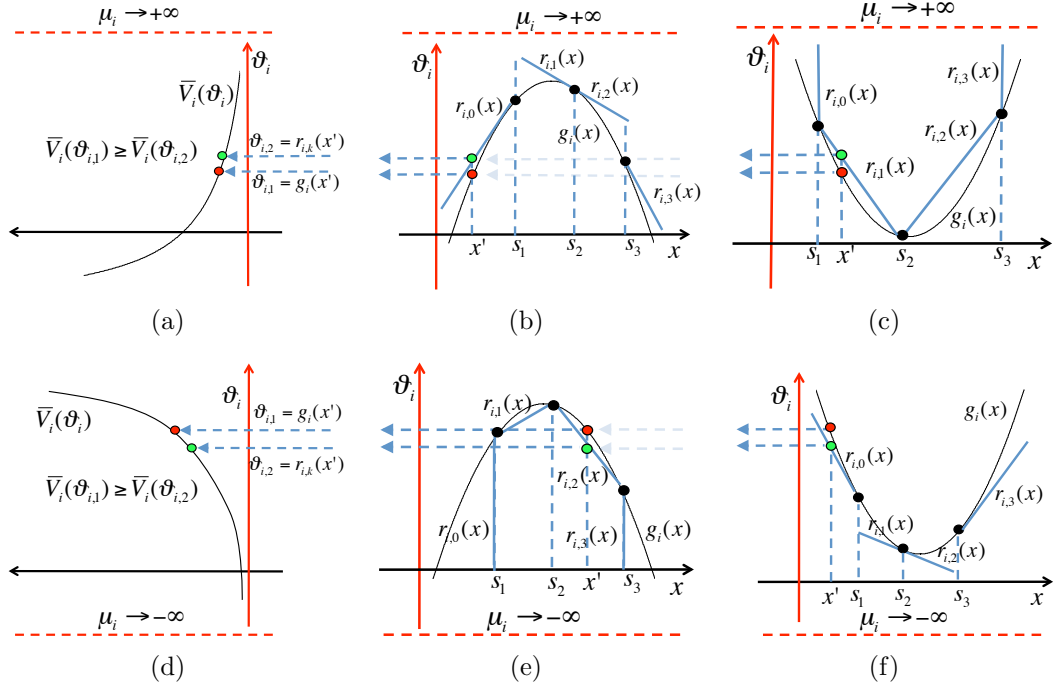
Figure 5.3: Example of construction of the linear functions $r_{i,k}(x)$ when $\bar{V}_i(\vartheta_i)$ is monotonic. **(a)** A decreasing marginal potential $\bar{V}_i(\vartheta_i)$. The axis associated to the independent variable ϑ_i is vertical. **(b)**,**(e)** Construction with a non-monotonic concave $g_i(x)$. **(c)**,**(f)** Construction with a non-monotonic convex $g_i(x)$. **(d)** An increasing marginal potential $\bar{V}_i(\vartheta_i)$. The axis associated to the independent variable ϑ_i is vertical.

Figure 5.3(a) shows a decreasing (convex) marginal potential $\bar{V}_i$ while Figure 5.3(d) depicts an increasing (convex) $\bar{V}_i$. In Figures 5.3(b),(f), the linear functions $r_{i,k}$ are all tangent at the support points to the the nonlinearities $g_i(x)$. Note that the linear functions $r_{i,1}$ and $r_{i,2}$ are represented by the same straight line, in both pictures.

On the contrary, in Figures 5.3(c),(e), the linear functions $r_{i,k}$ are all secant straight lines. However, it is important to remark that in Figures 5.3(c),(e), i.e. when $\bar{V}_i$ is decreasing and g_i is convex or $\bar{V}_i$ is increasing and g_i is concave, the construction is possible only in a finite domain. Indeed, the linear functions $r_{i,0}$ and $r_{i,3}$ have an infinite slope in both figures.

Finally, note that in Figures 5.3(a),(d) the axis associated to the

133

independent variable ϑ_i is vertical.

A simple analytical example is given below.

Example 2 *Consider the target pdf*

$$p_o(x) \propto w(x) \exp\{-h(x)\}, \tag{5.7}$$

where $w(x) > 0$ is a positive function, either convex or concave, and $h(x)$ is a convex function with a minimum at μ_1. We can rewrite $p_o(x)$ as

$$p_o(x) \propto \exp\{-h(x) + \log[w(x)]\}, \tag{5.8}$$

so that the associated potential function is

$$V(x; \mathbf{g}) = h(x) - \log[w(x)].$$

We decompose V as

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

where $\bar{V}_1(\vartheta_1) = h(\vartheta_1)$, $g_1(x) = x$, $\bar{V}_2(\vartheta_2) = -\log(\vartheta_2)$ and $g_2(x) = w(x)$. Note that $\bar{V}_2(\vartheta_2)$ is a monotonic (decreasing) marginal potential (hence, $\mu_2 \to +\infty$).

5.2.3   Quasi-convex marginal potentials

Let us now consider the case of quasi-convex marginal potentials, i.e., functions $\bar{V}_i(\vartheta_i)$, $i = 1, ...n$, such that

- $\bar{V}_i$ is increasing for $\vartheta_i > \mu_i$, and

- $\bar{V}_i$ is decreasing for $\vartheta_i < \mu_i$.

Figure 5.4(a) displays an example of a quasi-convex marginal potential $\bar{V}_i(\vartheta_i)$, where the independent variable ϑ_i is plotted in the vertical axis. In this case, the procedure to construct the linear functions $r_{i,k}(x)$, $x \in \mathcal{I}_k$, in Section 3.2.3 (i.e., the first step of the standard GARS technique) remains valid. Namely, we are able to find suitable linear functions such that

$$\bar{V}_i(r_{i,k}(x)) \le \bar{V}_i(g_i(x)), \tag{5.9}$$

134
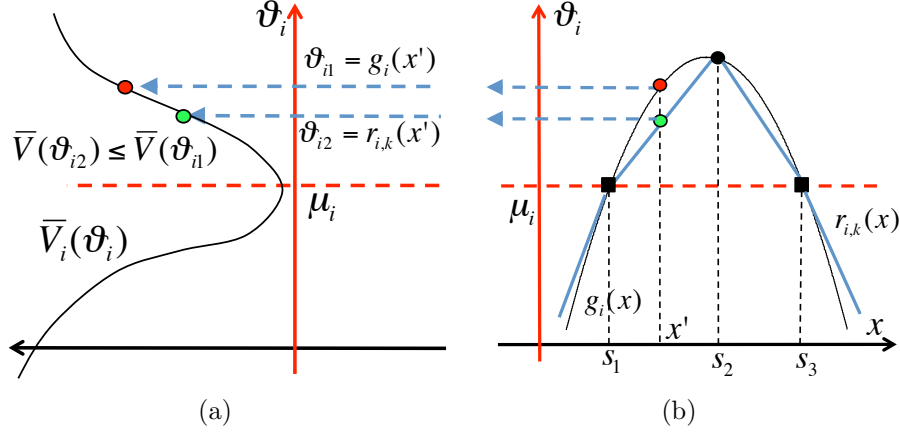
Figure 5.4: **(a)** Example of quasi-convex potential $\bar{V}_i(\vartheta_i)$. The axis of the independent variable ϑ_i is rotated by 90°. **(b)** Example of construction of the linear functions $r_{i,k}(x)$ such that $\bar{V}_i(r_{i,k}(x)) \leq \bar{V}_i(g_i(x))$, $\forall x \in \mathcal{D}$, when the nonlinearity $g_i(x)$ is non-monotonic and concave.

$i = 1, ..., n$ and $k = 0, ..., m_t$, using the procedure explained in Section 3.2.3. Figure 5.4(b) illustrates the construction of the linear functions $r_{i,k}$ when the nonlinearity $g_i(x)$ is non-monotonic and concave.

Actually, given an interval $\mathcal{I}_k = [s_k, s_{k+1}]$ and denoting

$$\mathcal{I}_{\vartheta_i}^{(k)} \triangleq \{\vartheta_i \in \mathbb{R} : \vartheta_i = g_i(x), \ \forall x \in \mathcal{I}_k\}, \tag{5.10}$$

the image set corresponding to the interval $\mathcal{I}_k$, the procedure to construct the linear functions $r_{i,k}(x)$, $x \in \mathcal{I}_k$, can be summarized as follows:

R1 If the marginal potential $\bar{V}_i(\vartheta_i)$ is increasing in the image interval $\mathcal{I}_{\vartheta_i}^{(k)}$ $(\vartheta_i > \mu_i)$, then use a straight line tangent to $g_i(x)$.

R2 Otherwise, if the marginal potential $\bar{V}_i(\vartheta_i)$ is decreasing in the image interval $\mathcal{I}_{\vartheta_i}^{(k)}$ $(\vartheta_i < \mu_i)$, use a line secant to $g_i(x)$.

We also use this construction procedure in the next Sections 5.2.4 and 5.2.5. With these rules, even with quasi-convex marginal potentials we can produce a modified potential that satisfies

$$V(x; \mathbf{r}_k) \leq V(x; \mathbf{g}), \quad \forall x \in \mathcal{D}. \tag{5.11}$$

135

If we are able to draw from $\pi_t^*(x) \propto \exp\{-V(x; \mathbf{r}_k)\}$, then we can simply use $\pi_t^*(x)$ as the proposal pdf in the RS scheme. However, if are not able to draw from $\pi_t^*(x)$, in this scenario we cannot ensure that the modified potential $V(x; \mathbf{r}_k)$ is convex and, therefore, the second step of the standard GARS algorithm cannot be implemented. Namely, in general we cannot build a piecewise-linear lower hull $W_t(x) \leq V(x; \mathbf{r}_k)$.

Next we consider a special kind of quasi-convex marginal potentials such that, if and only if all the marginal potentials are of this type, it is also possible to build an adequate lower hull $W_t(x)$. However, the resulting proposal pdf $\pi_t(x)$ can be improper.

A special kind of quasi-convex marginal potentials

Consider a target pdf $p_o(x) \propto \exp\{-V(x; \mathbf{g})\}$ where the potential function has the form $V(x; \mathbf{g}) = \sum_{i=1}^n \bar{V}_i(g_i(x))$ and each marginal potential $\bar{V}_i(\vartheta_i)$, $i = 1, ..., n$, is

- strictly increasing and concave for $\vartheta_i > \mu_i$, and

- strictly decreasing and concave for $\vartheta_i < \mu_i$.

Obviously, $\bar{V}_i(\vartheta_i)$ has a unique minimum at $\vartheta = \mu_i$ and we are also assuming $\bar{V}_i(\mu_i) > -\infty$. Figure 5.5 shows an example 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 [117, 131].

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 $\cup_{i=1}^n \mathcal{X}_i$ (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 Section 3.2.4), i.e., $\cup_{i=1}^n \mathcal{X}_i \subset \mathcal{S}_t = \{s_k\}_{k=1}^{m_t}$, a simple estimate x_i can belong to the interval $\mathcal{I}_k = [s_k, s_{k+1}]$ only as a border point. Therefore, replacing $g_i(x)$ with the linear function $r_{i,k}(x)$, where x belongs to the interval $\mathcal{I}_k = [s_k, s_{k+1}]$ defined by two support points, we can write

$$\frac{d^2 \bar{V}_i(r_{i,k}(x))}{dx^2} = \frac{d^2 r_{i,k}}{dx^2} \frac{d\bar{V}_i}{d\vartheta_i} + \left(\frac{dr_{i,k}}{dx}\right)^2 \frac{d^2 \bar{V}_i}{d\vartheta_i^2} = 0 + \left(\frac{dr_{i,k}}{dx}\right)^2 \frac{d^2 \bar{V}_i}{d\vartheta_i^2} \leq 0 \quad (5.12)$$

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
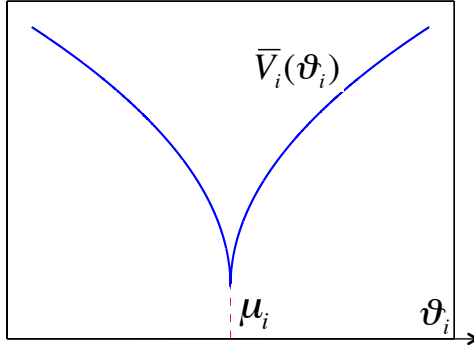
Figure 5.5: Example of generic marginal potential function $\bar{V}_i(\vartheta_i)$ strictly increasing and concave for $\vartheta_i > \mu_i$, strictly decreasing and concave for $\vartheta_i < \mu_i$.

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 $x \in \mathcal{I}_k = [s_k, s_{k+1}]$, $k = 1, \ldots, m_t - 1$, 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 an 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 the Chapter 4.

Figure 5.6 shows an example of construction of the lower hull $W_t(x)$ with three support points, $m_t = 3$.

5.2.4 Concave marginal potentials

Let us consider a target pdf $p_o(x) \propto \exp\left\{-\sum_{i=1}^n \bar{V}_i(g_i(x))\right\}$, $x \in \mathcal{D}$, where *all* the marginal potentials $\bar{V}_i$ are concave with maxima at μ_i, $i = 1, ..., n$. Note that, in this case, the domain $\mathcal{D}$ has to be *bounded* for $p_o(x)$ to be a proper pdf.

The construction of the linear functions $r_{i,k}(x)$ follows the rules R1 and R2 described in the previous Section 5.2.3. Therefore, given the intervals $\mathcal{I}_k$,
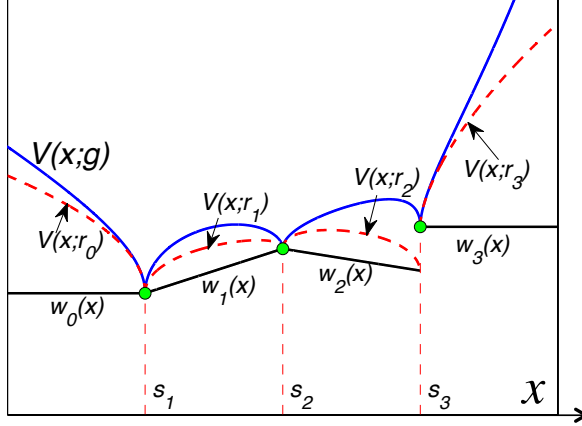
Figure 5.6: Example of construction of the modified potentials $V(x; \mathbf{r}_k)$ (dashed red lines) and the piecewise linear function $W_t(x)$ (black straight lines) when the potential function is $V(x; \mathbf{g}) = \sqrt{|x^2 - 4|} + \sqrt{|1 - \exp(x)|}$ (blue line). We can express it as $V(x; \mathbf{g}) = \bar{V}_1(x^2 - 4) + \bar{V}_2(1 - \exp(x))$, where $\bar{V}_1(\vartheta_i) = \bar{V}_2(\vartheta_i) = \sqrt{|\vartheta_i|}$, i.e., the marginal potentials are concave, strictly decreasing for $\vartheta_i < 0$ and strictly increasing for $\vartheta > 0$. The two nonlinearities $g_1(x) = x^2 - 4$ and $g_2(x) = 1 - \exp(x)$ generate the sets of simple estimates $\mathcal{X}_1 = \{-2, 2\}$ and $\mathcal{X}_2 = \{0\}$, that are contained in the set of support points $\mathcal{S}_t = \{s_1 = -2, s_2 = 0, s_3 = 2\}$.

and

$$\mathcal{I}_{\vartheta_i}^{(k)} = \{\vartheta_i \in \mathbb{R} : \; \vartheta_i = g_i(x), \; x \in \mathcal{I}_k\},$$

$k = 0, ..., m_t$, we use a straight line $r_{i,k}(x)$ tangent to $g_i(x)$ at an arbitrary point $x^* \in \mathcal{I}_k$ if the marginal potential $\bar{V}_i(\vartheta_i)$ is increasing in the interval $\mathcal{I}_{\vartheta_i}^{(k)}$. Otherwise, if the marginal potential $\bar{V}_i(\vartheta_i)$ is decreasing in the interval $\mathcal{I}_{\vartheta_i}^{(k)}$, we use a secant line. Note that this construction is proper because the domain $\mathcal{D}$ is bounded.

Figure 5.7 sketches an example of construction of the linear functions $r_{i,k}(x)$, $k = 0, ..., 5$, when $\bar{V}_i$ is concave and the nonlinearity g_i is convex. The straight lines $r_{i,1}(x)$ and $r_{i,2}(x)$ are tangent to $g_i(x)$ at arbitrary points within $\mathcal{I}_1$ and $\mathcal{I}_2$, respectively. The linear functions $r_{i,0}(x)$ and $r_{i,3}(x)$ are secant, passing through the support points. This construction ensures that $\bar{V}_i(r_{i,k}(x)) \leq \bar{V}_i(g_i(x))$ for all $x \in \mathcal{I}_k$, $k = 0, ..., m_t$. The value μ_i is indicated with a green dashed line.
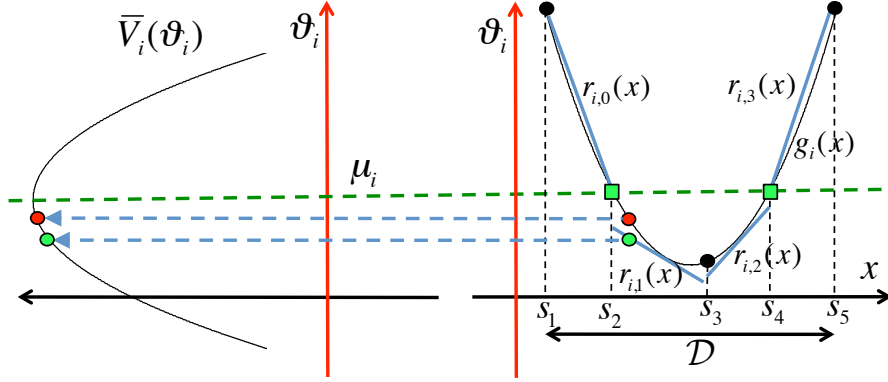
Figure 5.7: Example of construction of the linear functions $r_{i,k}(x)$, $k = 0, ..., m_t$, using $m_t = 5$ support points when $\bar{V}_i$ is concave with a maximum at μ_i and the nonlinearity g_i is convex. Note that, clearly, $\bar{V}_i(r_{i,k}(x)) \leq \bar{V}_i(g_i(x))$. The value μ_i is indicated with green dashed line.

Note that, unlike in the case of convex marginal potentials, when the function $\bar{V}_i$ is concave, the condition satisfied by the linear functions $r_{i,k}(x)$ is

$$|\mu_i - r_{i,k}(x)| \geq |\mu_i - g_i(x)|,$$

for all $x \in \mathcal{I}_k = [s_k, s_{k+1}]$.

Since we assume that all $\bar{V}_i$ are concave, the modified potential

$$V(x; \mathbf{r}_k) = \sum_{i=1}^{n} \bar{V}_i(r_{i,k}(x)), \quad \forall x \in \mathcal{I}_k,$$

is also concave. Then, the lower hull $W_t(x)$, $x \in \mathcal{I}_k = [s_k, s_{k+1}]$, $k = 0, ..., m_t$, can be easily built using the linear segments joining the points $(s_k, V(s_k; \mathbf{r}_k))$ and $(s_{k+1}, V(s_{k+1}; \mathbf{r}_{k+1}))$ as in Figure 5.6 (recall that the domain $\mathcal{D}$ is bounded in this case).

It is important to remark that if some marginal potentials are convex and others are concave, the substitution of the g_i's with suitable $r_{i,k}$'s is possible, but the modified potential $V(x; \mathbf{r}_k)$ has a second derivative with non-constant sign, in general. Therefore, in such scenario we cannot build a suitable piecewise linear function $W_t(x)$[3]. The only proposal function that

---

[3]However, the alternative procedure of Section 5.3.1 may be applied to overcome this limitation.

139

we can obtain is $\pi_t^*(x) \propto \exp\{-V(x; \mathbf{r}_k)\}$, but it may not be easy to draw directly from it.

5.2.5 Marginal potentials with several stationary points

So far, we have considered marginal potentials with only one minimum or one maximum (or without any stationary points, in Section 5.2.2). Here, we deal with marginal potentials $\bar{V}_i$ with more than one stationary point.

It turns out that we can still build useful linear functions $r_{i,k}(x)$ using the image sets $\mathcal{I}_{\vartheta_i}^{(k)}$ defined Eq. (5.10) and the construction rules R1 and R2 of Section 5.2.3. Figure 5.8(a) depicts an example of a marginal potential $\bar{V}_i$ with two minima, $\mu_{i,1}$, $\mu_{i,3}$, and a maximum, $\mu_{i,2}$. The vertical axis corresponds to the independent variable ϑ_i. Figure 5.8(b) illustrates an example of construction of the linear functions $r_{i,k}(x)$, $k = 0, ..., m_t$, using $m_t = 7$ support points $\mathcal{S}_t = \{s_1, ..., s_7\}$, with a non-monotonic convex nonlinearity $g_i(x)$. We use secant lines in the intervals $\mathcal{I}_1 = [s_1, s_2]$, $\mathcal{I}_3 = [s_3, s_4]$, $\mathcal{I}_4 = [s_4, s_5]$ and $\mathcal{I}_6 = [s_6, s_7]$, and we use tangent lines in the intervals $\mathcal{I}_0 = (-\infty, s_1]$, $\mathcal{I}_2 = [s_2, s_3]$, $\mathcal{I}_5 = [s_5, s_6]$ and $\mathcal{I}_7 = [s_7, +\infty)$. The value $\mu_{i,2}$ of the maximum is indicated with a green dashed line, while the values $\mu_{i,1}$ and $\mu_{i,3}$ of the minima are displayed with red dashed lines.

More details of the construction are provided in the sequel. Let q be the number of stationary points of the marginal potential $\bar{V}_i(\vartheta_i)$ (in Figure 5.8(a) we have $q = 3$). The set of simple estimates is extended as

$$\mathcal{X}_i = \{x \in \mathcal{D} : \mu_{i,j} = g_i(x) \ \ j = 1, ..., q\}, \tag{5.13}$$

hence, in Figure 5.8(b) we have $|\mathcal{X}_i| = 6$ simple estimates and $\mathcal{X}_i = \{s_1, s_2, s_3, s_5, s_6, s_7\}$ (only s_4 is not a simple estimate). Four of them are generated from the minima $\mu_{i,1}$ and $\mu_{i,3}$ while the two remaining ones come from the maximum $\mu_{i,2}$. All the simple estimates are contained in the set of support points, i.e., $\mathcal{X}_i \subseteq \mathcal{S}_t$, as described in Section 3.2.4.

The linear functions $r_{i,k}(x)$, for $k \in \{0, ..., 7\}$, are properly built. Namely, it is easily seen from Figure 5.8 that

$$|\mu_{i,1} - r_{i,k}(x)| \leq |\mu_{i,1} - g_i(x)|, \quad x \in \mathcal{I}_k,$$

for $k = 0$, $k = 1$, $k = 6$ and $k = 7$,s while

$$|\mu_{i,3} - r_{i,k}(x)| \leq |\mu_{i,3} - g_i(x)|, \quad x \in \mathcal{I}_k,$$
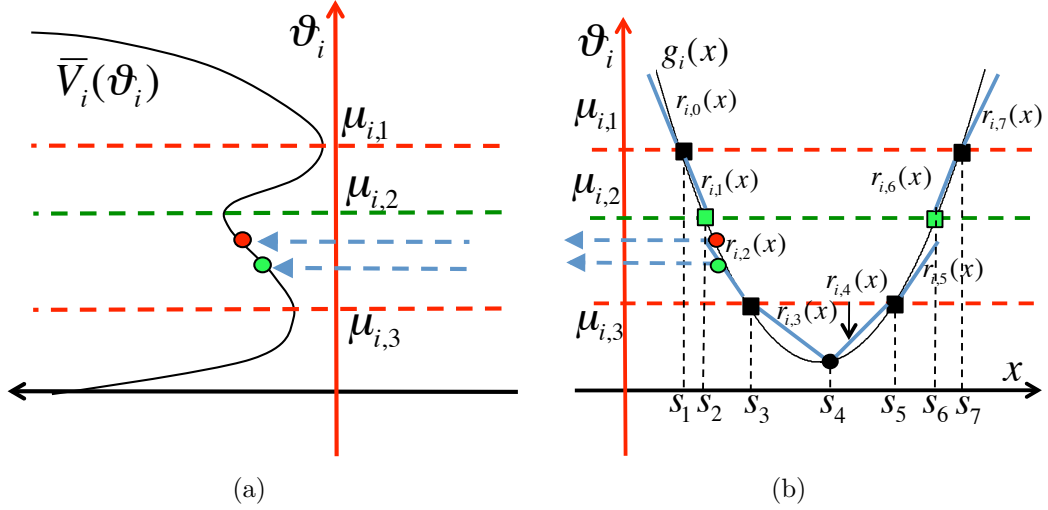
Figure 5.8: **(a)** Example of a marginal potential $\bar{V}_i(\vartheta_i)$ with two minima $\mu_{i,1}$, $\mu_{i,3}$, and a maximum $\mu_{i,2}$. The axis of the independent variable ϑ_i is vertical. **(b)** Example of construction of the linear functions $r_{i,k}(x)$, $k = 0, ..., m_t$, using $m_t = 7$ support points, with a non-monotonic convex nonlinearity $g_i(x)$. The simple estimates are indicated by squares s_1, s_2, s_3, s_5, s_6 and s_7 (s_4 is not a simple estimate).

for $k = 2, 3, 4, 5$. Additionally, the inequality

$$|\mu_{i,2} - r_{i,k}(x)| \geq |\mu_{i,2} - g_i(x)|, \quad x \in \mathcal{I}_k,$$

is verified for $k = 1, 2$ and $k = 5, 6$. As a consequence, if we define $\mathbf{r}_k = [r_{1,k}, ..., r_{n,k}]$, $k = 0, ..., m_t$, as usual then the condition $V(x; \mathbf{r}_k) \leq V(x; \mathbf{g})$ is satisfied.

If it is easy to draw from $\pi_t^*(x) \propto \exp\{-V(x; \mathbf{r}_k)\}$, then we can readily use $\pi_t^*(x)$ as a proposal pdf. Unfortunately, we cannot ensure that the modified potential $V(x; \mathbf{r}_k)$ is convex (or concave). Hence, in general, we cannot implement the second step of the standard GARS algorithm. Namely, we cannot build a piecewise linear lower hull $W_t(x) \leq V(x; \mathbf{r}_k) \leq V(x; \mathbf{g})$, in general. However, this problem may be overcome when the inflection points of $\bar{V}_i$ are available using the alternative GARS procedure in Section 5.3.1.

5.2.6 Summary

Table 5.1 provides a summary of the different possibilities that we have analyzed so far. The first column describes the type of the marginal potentials, the second column states whether the domain $\mathcal{D}$ of the target pdf $p_o(x)$ can be unbounded or not with the previous choice of the marginal potentials. The remaining columns indicate whether it is possible, in general, to build the modified potential $V(x; \mathbf{r}_k)$ and the piecewise lower hull $W_t(x)$.

5.3 Simplified GARS algorithm

The standard GARS procedure introduced in Chapter 3 relies on the ability to analytically obtain the minima μ_i of the marginal potentials $\bar{V}_i$ and the sets of simple estimates $\mathcal{X}_i = \{x \in \mathcal{D} : \mu_i = g_i(x)\}$, $i = 1, ..., n$.

The need for the (not necessarily straightforward) calculation of the μ_i's and $\mathcal{X}_i$'s may render the GARS algorithm non-intuitive for some problems and discourage many potential users. Ideally, the technique should be easy to code, requiring only the ability to evaluate some function related to the pdf.

Let us note that leaving aside the calculation of μ_i and $\mathcal{X}_i$, $i = 1, ..., n$, the application of GARS method demands:

- the evaluation of the marginal potentials $\bar{V}_i(\vartheta_i)$,

- the evaluation of the nonlinearities, $\vartheta_i = g_i(x)$, and

- the evaluation of the first derivatives $\frac{g_i}{dx}$,

for $i = 1, ..., n$. In the sequel, we propose extensions of the standard GARS algorithm that enable its use when the minima μ_i, $i = 1, ..., n$, are unknown (see Section 5.3.1) and the simple estimates $\mathcal{X}_i$, $i = 1, ..., n$, are unavailable (see Section 5.3.2). Finally, in Section 5.3.3, we explain how to remove the computation of the derivatives $\frac{dg_i}{dx}$, $i = 1, ..., n$, from the sampling procedure. The implementation of the resulting method is straightforward, nearly automatic, and interestingly enough, this simplified method turns out to be applicable in problems that cannot be tackled using the conventional GARS algorithm.

Table 5.1: Summary of the different cases.

Marginal potentials $\bar{V}_i$	Domain $\mathcal{D}$ of the target pdf $p_o(x)$	Construction of $V(x; \mathbf{r}_k)$	Construction of $W_t(x)$
all convex non-monotonic	unbounded	possible	possible
some $\bar{V}_i$ convex and increasing, with g_i convex	unbounded	possible	possible
some $\bar{V}_i$ convex and increasing, with g_i concave	bounded	possible	possible
some $\bar{V}_i$ convex and decreasing, with g_i convex	bounded	possible	possible
some $\bar{V}_i$ convex and decreasing, with g_i concave	unbounded	possible	possible
quasi-convex	unbounded	possible	impossible
quasi-convex specific case of subsection in 5.2.3	unbounded	possible	possible but $\pi_t(x)$ is improper
all concave	bounded	possible	possible
some convex and others concave	unbounded	possible	possible (see Section 5.3.1)
with several stationary points	unbounded	possible	possible (see Section 5.3.1)

5.3.1 Unknown μ_i

The standard GARS procedure described in Section 3.2 consists in the following two steps:

1. Replace the nonlinearities $g_i(x)$ with the linear functions $r_{i,k}(x)$ for all $x \in \mathcal{I}_k$, $i = 1, ..., n$, and $k = 0, ..., m_t$.

2. Build the piecewise linear function $W_t(x)$ using straight lines tangent to the modified potential $V(x; \mathbf{r}_k) = \sum_{i=1}^{n} \bar{V}_i(r_{i,k}(x))$, in such a way that $W_t(x) \leq V(x; \mathbf{r}_k)$.

In the first step, we need to know the position μ_i of the minima of the marginal potentials, $\bar{V}_i(\vartheta_i)$, $i = 1, ..., n$. If the minima μ_i are unknown, we can apply the alternative procedure described below.

Let us consider a convex marginal potential $\bar{V}_i(\vartheta_i)$, $\vartheta_i \in \mathcal{C} \subseteq \mathbb{R}$, with an unknown minimum at μ_i, and a set of q_t support points at time t,

$$\mathcal{F}_t = \{\theta_1, ..., \theta_{q_t}\} \subset \mathcal{C}, \tag{5.14}$$

sorted out in ascending order $\theta_1 < ... < \theta_{q_t}$. We can construct straight lines tangent to $\bar{V}_i$ at the support points in $\mathcal{F}_t$ and combine them to yield a piecewise linear lower hull $R_i(\vartheta_i)$, i.e.,

$$R_i(\vartheta_i) \leq \bar{V}_i(\vartheta_i), \quad \vartheta_i \in \mathcal{C}. \tag{5.15}$$

Figure 5.9 shows a simple example, where $R_i(\vartheta_i)$ is built using $q_t = 3$ support points.
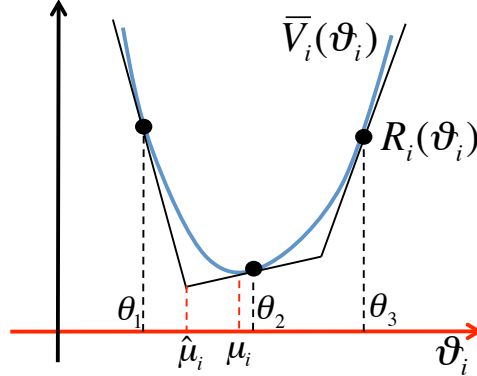
Figure 5.9: Example of construction of the lower hull $R_i(\vartheta_i) \leq \bar{V}_i(\vartheta_i)$ using straight lines tangent to $\bar{V}_i(\vartheta_i)$ with $q_t = 3$ support points, $\mathcal{F}_t = \{\theta_1, \theta_2, \theta_3\}$.

The lower hull $R_i(\vartheta_i)$ has a minimum at $\hat{\mu}_i$ that can be calculated in a straightforward manner. The value $\hat{\mu}_i$ can be used as a surrogate for μ_i in order to apply the GARS method. In particular, from $\hat{\mu}_i$ we can obtain an alternative set of simple estimates

$$\hat{\mathcal{X}}_i = \{x_i \in \mathcal{D} : \ g_i(x) = \hat{\mu}_i\}$$

144

and then use $\hat{\mathcal{X}}_i$, $i = 1, ..., n$, to construct the set of support points $\mathcal{S}_t = \{s_1, ..., s_{m_t}\} \subset \mathcal{D}$. Given $\mathcal{S}_t$, we compute the linear functions $r_{i,k}$, $i = 1, ..., n$, and $k = 0, ..., m_t$, in the usual way and, given these functions, we obtain a lower hull

$$W_t(x) \triangleq \sum_{i=1}^{n} R_i(r_{i,k}(x)) \leq V(x; \mathbf{g}) = \sum_{i=1}^{n} \bar{V}_i(g_i(x)). \qquad (5.16)$$

The main difference with the convention GARS techniques is that we have to build two sets of support points ($\mathcal{F}_t$ and $\mathcal{S}_t$) instead of one. Figure 5.10 shows how the $r_{i,k}(x)$, $k = 0, ..., m_t = 3$, are built, involving both $\mathcal{F}_t$ and $\mathcal{S}_t$.
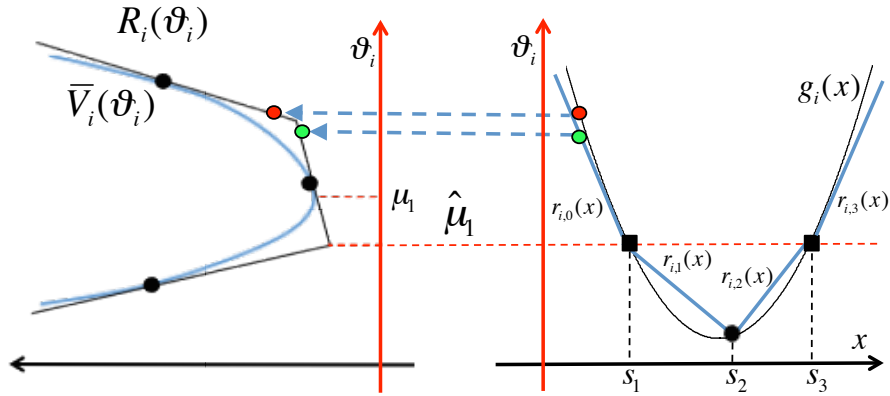
Figure 5.10: Example of construction of the linear functions $r_{i,k}$, $k = 0, ..., m_t = 3$, using $\hat{\mu}_i$ to calculate the simple estimates. The set of support points is $\mathcal{S}_t = \{s_1, s_2, s_3\}$ and the simple estimates are s_1 and s_3. The point s_2 is arbitrary (as long as $s_1 < s_2 < s_3$).

The alternative GARS procedure with unknown minima μ_i can be outlined as follows.

1. Choose arbitrarily $\mathcal{F}_0 = \{\theta_1, ..., \theta_{q_0}\} \in \mathcal{C}$ and an auxiliary set $\mathcal{A}_0 = \{s_1, ..., s_{a_0}\} \in \mathcal{D}$. Set $t = 0$.

2. Substitute each marginal potential $\bar{V}_i$ with a lower hull R_i, $i = 1, ..., n$, formed by straight lines tangent at the points $\mathcal{F}_t = \{\theta_1, ..., \theta_{q_t}\}$.

3. Find the minimum $\hat{\mu}_i$ of each R_i and calculate $\hat{\mathcal{X}}_i = \{x_i \in \mathcal{D} : g_i(x) = \hat{\mu}_i\}$, $i = 1, ..., n$.

4. Build the set of support points

$$\mathcal{S}_t = \{s_1, ..., s_{m_t}\} = \mathcal{A}_t \cup \hat{\mathcal{X}}_1 \cup \hat{\mathcal{X}}_2 \cup ... \cup \hat{\mathcal{X}}_n \subset \mathcal{D},$$

and sort it in ascending order, i.e., $s_1 < s_2 < .... < s_{m_t}$. Namely, the set $\mathcal{S}_t$ contains all points in $\mathcal{A}_t$ and all the simple estimates. For instance, in Figure 5.10 we have $\mathcal{A}_t = \{s_2\}$ and $\mathcal{S}_t = \{s_1, s_2, s_3\}$.

5. Replace the nonlinearities $g_i(x)$ with the linear functions $r_{i,k}(x)$ for all $x \in \mathcal{I}_k = [s_k, s_{k+1}]$, $i = 1, ..., n$, $k = 0, ..., m_t$, as described in Chapter 3.

6. Build the piecewise linear function $W_t(x) = \sum_{i=1}^{n} R_i(r_{i,k}(x))$, for all $x \in \mathcal{I}_k$.

7. Draw samples x' from $\pi_t(x) \propto \exp\{-W_t(x)\}$ and u' from $\mathcal{U}([0, 1])$.

8. If $u' \leq \frac{\exp\{-V(x';\mathbf{g})\}}{\exp\{-W_t(x')\}}$ then accept x' and set $\mathcal{A}_{t+1} = \mathcal{A}_t$ and $\mathcal{F}_{t+1} = \mathcal{F}_t$.

9. Otherwise, if $u' > \frac{\exp\{-V(x';\mathbf{g})\}}{\exp\{-W_t(x')\}}$, then discard x' and set $\mathcal{A}_{t+1} = \mathcal{A}_t \cup \{x'\}$ and $\mathcal{F}_{t+1} = \mathcal{F}_t \cup \{\theta' = g_i(x')\}$.

10. Sort $\mathcal{A}_{t+1}$ and $\mathcal{F}_{t+1}$ in ascending order, set $t = t + 1$ and go back to step 2.

This strategy actually enables us to address more general cases. Consider, for instance, a marginal potential with $\bar{V}_i$ with several stationary points. We have seen in Section 5.2.5 that the second step of the standard GARS procedure cannot be implemented. However, if the positions of the inflection points of $\bar{V}_i$ are available, we can construct adequately a lower hull $R_i(\vartheta_i) \leq \bar{V}_i(\vartheta_i)$ and then apply the alternative GARS procedure introduced above. Moreover, it can be directly applied when some marginal potentials $\bar{V}_i$ are convex and others are concave (also in this case the standard GARS procedure cannot be used, see Section 5.2.4).

5.3.2   Unknown $\mathcal{X}_i$

Consider a collection of marginal potentials and nonlinearities such that the minima μ_i, $i = 1, ..., n$, are available but it is not possible to solve the

equations $\mu_i = g_i(x_i)$ and, hence we cannot construct the sets $\mathcal{X}_i$ of simple estimates.

We need to devise a new method for the computation of suitable linear functions $r_{i,k}(x)$ that does not involve the sets $\mathcal{X}_i$. Recall that the conditions

$$|\mu_i - r_{i,k}(x)| \leq |\mu_i - g_i(x)| \quad \text{and} \tag{5.17}$$

$$(\mu_i - r_{i,k}(x))(\mu_i - g_i(x)) \geq 0. \tag{5.18}$$

should be satisfied in order to guarantee that $V(x; \mathbf{r}_k) \leq V(x; \mathbf{g})$, $\forall x \in \mathcal{I}_k$.

We start choosing an auxiliary set $\mathcal{E}_t$ of $e_t > 0$ support points, $\mathcal{E}_t = \{s_1, ..., s_{e_t}\}$. For each non-linearity g_i, $i = 1, ..., n$, we build a partition of $\mathcal{E}_t$ as

$$\mathcal{E}_t = \mathcal{E}_{t,1}^i \cup \mathcal{E}_{t,2}^i \cup \mathcal{E}_{t,3}^i,$$

where the disjoint subsets $\mathcal{E}_{t,1}^i$, $\mathcal{E}_{t,2}^i$ and $\mathcal{E}_{t,3}^i$ are constructed as follows:

For $i = 1, ..., n$

if g_i is convex then

$$\mathcal{E}_{t,2}^i = \{s \in \mathcal{E}_t : \ g_i(x) \leq \mu_i\},$$

$$\mathcal{E}_{t,1}^i = \{s \in \mathcal{E}_t : \ s < \min\{\mathcal{E}_{t,2}^i\} \text{ and } g_i(s) > \mu_i\},$$

$$\mathcal{E}_{t,3}^i = \{s \in \mathcal{E}_t : \ s > \max\{\mathcal{E}_{t,2}^i\} \text{ and } g_i(s) > \mu_i\},$$

otherwise if g_i is concave then

$$\mathcal{E}_{t,2}^i = \{s \in \mathcal{E}_t : \ g_i(x) \geq \mu_i\},$$

$$\mathcal{E}_{t,1}^i = \{s \in \mathcal{E}_t : \ s < \min\{\mathcal{E}_{t,2}^i\} \text{ and } g_i(s) < \mu_i\},$$

$$\mathcal{E}_{t,3}^i = \{s \in \mathcal{E}_t : \ s > \max\{\mathcal{E}_{t,2}^i\} \text{ and } g_i(s) < \mu_i\}.$$

It should be noticed that when the nonlinearity $g_i(x)$ is monotonically increasing and concave, or monotonically decreasing and convex, $\mathcal{E}_{t,1}^i = \emptyset$. Also, if the nonlinearity $g_i(x)$ is monotonically increasing and convex or monotonically decreasing and concave, $\mathcal{E}_{t,3}^i = \emptyset$.

In order to build the linear functions we need to add two supports points for each unknown simple estimate of a nonlinearity g_i. We denote the set of these new added points as $\mathcal{B}_t^i = \{b_1, ..., b_{b_t^i}\}$ and define

$$\mathcal{S}_t^i \triangleq \{s_1, ..., s_{m_t^i}\} = \mathcal{E}_t \cup \mathcal{B}_t^i,$$

with $s_1 < .... < s_{m_t^i}$ where $m_t^i = e_t + b_t^i$, $i = 1, ..., n$. Note that the set $\mathcal{S}_t^i$ is specific for each g_i. For example, if the nonlinearity $g_i(x)$ is non-monotonic and either concave or convex, in general we have to incorporate 4 support points, i.e., $m_t^i = e_t + b_t^i = e_t + 4$.

Figure 5.11 shows the construction of the linear functions $r_{i,k}(x)$, $k = 0, ..., m_t^i = 7$, without the knowledge of the positions of the simple estimates, when $g_i(x)$ is a non-monotonic convex nonlinearity. The $b_t^i = 4$ added points (shown with red circles) are

- the intersection point of the straight line tangent to $g_i(x)$ at $\max\{\mathcal{E}_{t,1}^i\}$ and the straight line $\vartheta_i = \mu_i$ (parallel to the axis x),

- the intersection point of the secant line passing through the points $(\max\{\mathcal{E}_{t,1}^i\}, g_i(\max\{\mathcal{E}_{t,1}^i\}))$, $(\min\{\mathcal{E}_{t,2}^i\}, g_i(\min\{\mathcal{E}_{t,2}^i\}))$, and the straight line $\vartheta_i = \mu_i$ (parallel to the axis x),

- the intersection point of the secant line passing through the points $(\max\{\mathcal{E}_{t,2}^i\}, g_i(\max\{\mathcal{E}_{t,2}^i\}))$, $(\min\{\mathcal{E}_{t,3}^i\}, g_i(\min\{\mathcal{E}_{t,3}^i\}))$, and the straight line $\vartheta_i = \mu_i$ (parallel to the axis x),

- and the intersection point of the straight line tangent to $g_i(x)$ at $\min\{\mathcal{E}_{t,3}^i\}$ and the straight line $\vartheta_i = \mu_i$.

Although the choice of these points may seen complicated, we have simple selected them to enclose the unknown solutions of the equation $g_i(x) = \mu_i$. To summarize, in the figure we have $\mathcal{E}_t = \{s_1, s_4, s_7\}$, $\mathcal{E}_{t,1}^i = \{s_1\}$, $\mathcal{E}_{t,2}^i = \{s_4\}$ and $\mathcal{E}_{t,3}^i = \{s_7\}$. The added points are $\mathcal{B}_t^i = \{s_2, s_3, s_5, s_6\}$ so that $\mathcal{S}_t^i = \{s_1, s_2, s_3, s_4, s_5, s_6, s_7\}$ ($m_t^i = 7$).

All the linear functions $r_{i,k}(x)$ in Figure 5.11 satisfy the conditions in Eq. (5.17) and (5.18). Furthermore, Figure 5.11 can be compared with Figure 3.5 in Chapter 3 that depicts the construction when the simple estimates are known. In particular, note that, as other support points are added the intervals $[s_2, s_3]$ and $[s_5, s_6]$, that contain the simple estimates, become smaller and smaller, and the construction here becomes equivalent to the standard procedure in Section 3.2.3.

Since $\mathcal{S}_t^i$ is different for each g_i, $i = 1, ..., n$, the number m_t^i of linear functions $r_{i,k}$ changes for each nonlinearity g_i. However, we can define the piecewise linear function

$$\hat{r}_i(x) = r_{i,k}(x), \quad \forall x \in \mathcal{I}_k \subset \mathcal{D}, \tag{5.19}$$

148

Figure 5.11: Example of construction of the linear functions $r_{i,k}(x)$, $k = 0, ..., m_t^i = 7$, without the knowledge of simple estimates, with a non-monotonic convex $g_i(x)$. We have $\mathcal{E}_{t,1}^i = \{s_1\}$, $\mathcal{E}_{t,2}^i = \{s_4\}$, $\mathcal{E}_{t,3}^i = \{s_7\}$, $\mathcal{E}_t = \{s_1, s_4, s_7\}$ and $\mathcal{B}_t^i = \{s_2, s_3, s_5, s_6\}$ while $\mathcal{S}_t^i = \{s_1, s_2, s_3, s_4, s_5, s_6, s_7\}$. The unknown simple estimates are shown with squares. The linear functions $r_{i,0}(x) = r_{i,1}(x)$ are tangent lines to $g_i(x)$ at s_1, $r_{i,2}(x) = r_{i,5}(x) = \mu_i$ are constant, $r_{i,3}(x)$ and $r_{i,4}(x)$ are secant lines and $r_{i,0}(x) = r_{i,1}(x)$ are tangent lines to $g_i(x)$ at s_7.

with $k = 0, ..., m_t^i$ for each $i \in \{1, ..., n\}$. Note that $\hat{r}_i(x)$ is defined for all $x \in \mathcal{D}$. Then, denoting as $\hat{\mathbf{r}} = [\hat{r}_1(x), ..., \hat{r}_n(x)]$, the modified potential is built as

$$V(x; \hat{\mathbf{r}}) = \sum_{i=1}^{n} \bar{V}_i(\hat{r}_i(x)) \leq V(x; \mathbf{g}) = \sum_{i=1}^{n} \bar{V}_i(g_i(x)), \qquad (5.20)$$

for all $x \in \mathcal{D}$ and we can construct the lower-hull $W_t(x)$ by tracing lines tangent to $V(x; \hat{\mathbf{r}})$ at arbitrary points. We recall that the proposal pdf is $\pi_t(x) \propto \exp\{-W_t(x)\}$.

Let us mention that the technique in this section can be combined with the procedure of Section 5.3.1 when the locations of the minima μ_i are not available.

5.3.3 A derivative-free procedure

To apply the procedure of Section 5.3.2, the position of the simple estimates is not needed. However, it is still necessary to use the first derivative of

$g_i(x)$ and also to build the lower hull $R_i(\vartheta_i)$ to substitute $\bar{V}_i$, $i = 1, .., n$, as described in Section 5.3.1.

However, we can use exactly the same procedure of Section 2.6.1 (for the standard ARS algorithm) to construct a piecewise linear lower hull $R_i(\vartheta_i)$ using only secant lines, i.e, without the need to compute derivatives.

Moreover, to avoid the calculation of the first derivative of the nonlinearity $g_i(x)$ (and only to evaluate it), we introduce here a derivative-free procedure using only secant lines, similar to the techniques of Section 2.6.1 for the standard ARS algorithm.

Figure 5.12 displays an example of construction of the linear functions $r_{i,k}(x)$, $k = 0, ..., m_t^i = 5$, without using the first derivative for a nonlinearity $g_i(x)$, which is non-monotonic and convex. In this case, we need to add only one support point for each unknown simple estimate (displayed as red circles). Figure 5.12 can be compared with Figure 5.11, where we use the first derivative and Figure 3.5, where we also know the position of the simple estimates.

The straight lines $r_{i,0}(x)$, $\forall x \in \mathcal{I}_0 = (-\infty, s_1]$, and $r_{i,2}(x)$, $\forall x \in \mathcal{I}_2 = [s_2, s_3]$, are formed by the secant line passing through $(s_3, g_i(s_3))$ and $(s_1, g_i(s_1))$. The straight lines $r_{i,3}(x)$, $\forall x \in \mathcal{I}_3 = [s_3, s_4]$, and $r_{i,5}(x)$, $\forall x \in \mathcal{I}_5 = [s_5, +\infty)$, are formed by secant lines passing through $(s_3, g_i(s_3))$ and $(s_5, g_i(s_5))$, respectively. The linear functions $r_{i,1}(x) = \mu_i$, $\forall x \in \mathcal{I}_1 = [s_1, s_2]$ and $r_{i,4}(x) = \mu_i$, $\forall x \in \mathcal{I}_4 = [s_4, s_5]$ are constant. Finally, note that $\mathcal{E}_{t,1}^i = \{s_1\}$, $\mathcal{E}_{t,2}^i = \{s_3\}$, $\mathcal{E}_{t,3}^i = \{s_5\}$ and $\mathcal{B}_t^i = \{s_2, s_4\}$.

Figure 5.13 depicts the construction incorporating one additional support point. In this case, we have $\mathcal{E}_{t,1}^i = \{s_1, s_2\}$, $\mathcal{E}_{t,2}^i = \{s_4\}$ and $\mathcal{E}_{t,3}^i = \{s_6\}$. The straight line $r_{i,0}(x)$, $\forall x \in \mathcal{I}_0 = (-\infty, s_1]$, is formed by a piece of the secant line passing through $(s_1, g_i(s_1))$ and $(s_2, g_i(s_2))$, while $r_{i,1}(x)$, $\forall x \in \mathcal{I}_1 = [s_1, s_2]$, is formed by a piece of the secant line passing through $(s_2, g_i(s_2))$ and $(s_4, g_i(s_4))$.

Finally, Figure 5.14 shows an example of the construction with one more point, i.e., $\mathcal{E}_{t,1}^i = \{s_1, s_2\}$, $\mathcal{E}_{t,2}^i = \{s_4, s_5\}$, $\mathcal{E}_{t,3}^i = \{s_7\}$ and $\mathcal{B}_t^i = \{s_3, s_6\}$. We recall that $\mathcal{E}_t = \mathcal{E}_{t,1}^i \cup \mathcal{E}_{t,2}^i \cup \mathcal{E}_{t,3}^i = \{s_1, s_2, s_4, s_5, s_7\}$ and

$$\mathcal{S}_t^i = \mathcal{E}_t \cup \mathcal{B}_t^i = \{s_1, s_2, s_3, s_4, s_5, s_6, s_7\}.$$

In this case, $r_{i,1}(x)$, $\forall x \in \mathcal{I}_1 = [s_1, s_2]$, is formed by a piece of the secant line passing through $(s_2, g_i(s_2))$ and $(s_4, g_i(s_4))$ and $r_{i,7}(x)$, $\forall x \in \mathcal{I}_7 = [s_7, +\infty)$, by a piece of the secant line passing through $(s_5, g_i(s_5))$ and $(s_7, g_i(s_7))$.
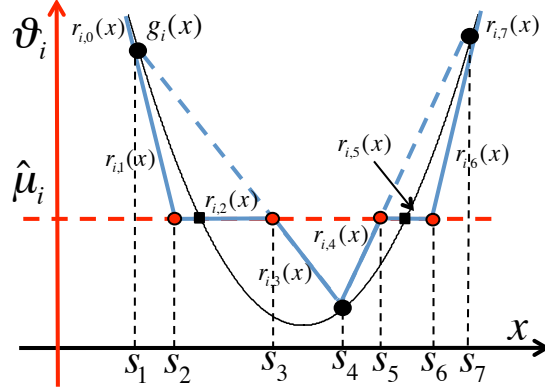
Figure 5.12: Example of construction of the linear functions $r_{i,k}(x)$, $k = 0, ..., \hat{m}_t = 5$, without the knowledge of simple estimates and without using the first derivative. All the linear functions $r_{i,k}(x)$ consist of pieces of secant lines passing through the support points. In this case, we have $\mathcal{E}_{t,1}^i = \{s_1\}$, $\mathcal{E}_{t,2}^i = \{s_3\}$, $\mathcal{E}_{t,3}^i = \{s_5\}$ and $\mathcal{B}_t^i = \{s_2, s_4\}$. The dashed straight lines represent the secant lines passing through the points $(s_1, g_i(s_1))$, $(s_3, g_i(s_3))$ and $(s_3, g_i(s_3))$, $(s_5, g_i(s_5))$.

From the sequence of plots in Figures 5.12-5.14 we can see how the approximation of g_i is improved when incorporating more support points ($m_t^i = 5$, $m_t^i = 6$ and $m_t^i = 7$, respectively). Indeed, the straight lines $r_{i,k}$ form the piecewise linear function

$$\hat{r}_i(x) = r_{i,k}(x), \quad \forall x \in \mathcal{I}_k \subset \mathcal{D} \tag{5.21}$$

with $k = 0, ..., m_t^i$, that becomes closer and closer to the nonlinearity g_i as we add support points (i.e. $\hat{r}_i(x) \to g_i(x)$ when $m_t^i \to +\infty$).

To summarize, the entire GARS technique can be implemented only evaluating $\bar{V}_i(\vartheta_i)$ (in order to build $R_i(\vartheta_i)$) and $g_i(x)$ (in order to build $\hat{r}_i(x)$). Specifically, the computation of the first derivatives can be skipped.

5.4 Summary

In this chapter, we have presented several extensions of the standard GARS technique presented in Chapter 3. We have first explained how to use the GARS algorithm when the target pdf does not belong to the exponential
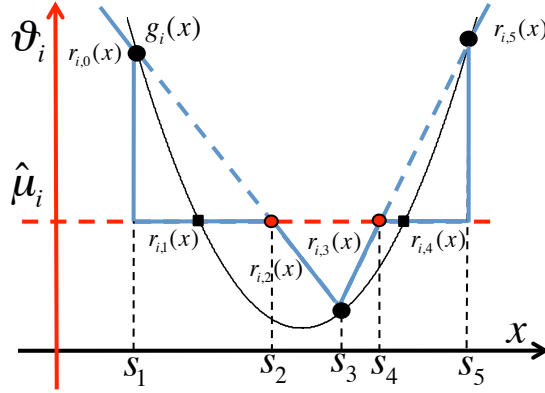
Figure 5.13: Example of construction of the linear functions $r_{i,k}(x)$, $k = 0, ..., \hat{m}_t = 6$, without the knowledge of simple estimates and without using the first derivative. All the linear functions $r_{i,k}(x)$ are composed for pieces of secant lines. In this case, we have $\mathcal{E}^i_{t,1} = \{s_1, s_2\}$, $\mathcal{E}^i_{t,2} = \{s_4\}$, $\mathcal{E}^i_{t,3} = \{s_6\}$ and $\mathcal{B}^i_t = \{s_3, s_5\}$. The dashed straight lines represent the secant lines passing through between the points $(s_1, g_i(s_1))$, $(s_2, g_i(s_2))$, between $(s_2, g_i(s_2))$, $(s_3, g_i(s_3))$ and finally $(s_5, g_i(s_5))$, $(s_6, g_i(s_6))$.

family (Section 5.1.1) and an interpretation of the GARS algorithm of Chapter 3 as a technique to draw samples from a target pdf generated by a transformation of scale of another density [74]. Note that the idea in Section 5.1.2 can be seen as a generalization of the concept of T-transformation, introduced in Section 2.7.3.

We have also discussed how it is possible to extend the class of target pdf's that the GARS method can address. In particular, we have shown that, with some minor modifications, the GARS methodology can be applied with a broader class of nonlinearities (not necessarily either convex or concave) and marginal potentials (not necessarily convex, with a minimum at μ_i).

In Section 5.3 we have introduced an alternative "automatic" algorithm that demands only the ability to evaluate the potential function $V(x; \mathbf{g})$, and does not assume knowledge of the simple estimates and/or the minima μ_i of the marginal potentials $\bar{V}_i$, $i = 1, ..., n$. Furthermore, a derivative-free algorithm has also been introduced in Section 5.3.3. Besides its simplicity, this alternative procedure expands the class of target pdf's that can be tackled with the GARS technique. Indeed, all the pdf's $p_o(x) \propto \exp\{-\sum_{i=1}^n \bar{V}_i(g_i(x))\}$, where the minima (or maxima) μ_i of $\bar{V}_i$ are unknown
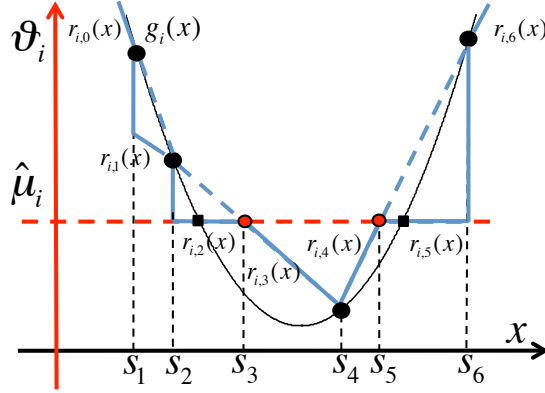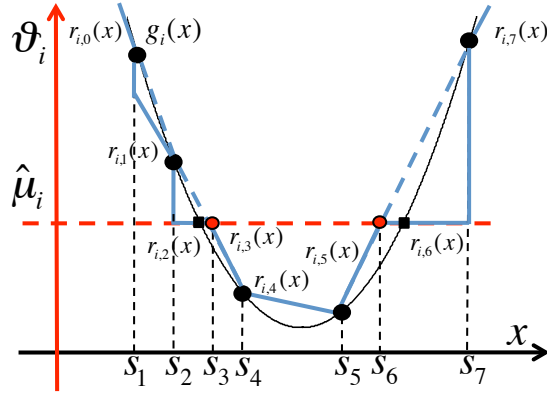
Figure 5.14: Example of construction of the linear functions $r_{i,k}(x)$, $k = 0, ..., \hat{m}_t = 7$, without the knowledge of simple estimates and without using the first derivative. All the linear functions $r_{i,k}(x)$ are composed for pieces of secant lines. In this case, we have $\mathcal{E}^i_{t,1} = \{s_1, s_2\}$, $\mathcal{E}^i_{t,2} = \{s_4, s_5\}$, $\mathcal{E}^i_{t,3} = \{s_7\}$ and $\mathcal{B}^i_t = \{s_3, s_6\}$. The dashed straight lines represent the secant lines passing through between the points $(s_1, g_i(s_1))$, $(s_2, g_i(s_2))$, between $(s_2, g_i(s_2))$, $(s_4, g_i(s_4))$ and finally $(s_6, g_i(s_6))$, $(s_7, g_i(s_7))$.

or the equations $\mu_i = g_i(x)$ cannot be solved, can be addressed with the alternative procedure.

We have described each extension separately for clarity, and in order to focus the attention on one specific technique at a time. However, almost all the extensions and improvements presented in this chapter can be combined and utilized together. For instance, all techniques described in Section 5.2 can be applied without the knowledge of the minima (or maxima) μ_i, without the set of simple estimates $\mathcal{X}_i$, $i = 1, ..., n$, and without using the first derivative, by combining them with the method of Section 5.3.

153

Chapter 6

Numerical Results

In this chapter we present a collection of examples where the proposed adaptive sampling methods can be applied. They are broadly classified as applications of the standard GARS algorithm of Chapter 3 (in Section 6.1), examples involving target distributions with log-convex tails (in Section 6.2) and an application of the simplified GARS procedure of Chapter 5 (in Section 6.3). The goal is not to carry out an extensive numerical study of all the proposed algorithms (as the possible variations and versions are very numerous) but to illustrate how the main techniques can be used in practice and how they compare to some relevant methods existing in the literature. Most of the examples involve computer simulations only, but we have also applied the GARS algorithm to a target localization problem using real data from a network of wireless sensors.

6.1 Standard GARS technique

All the examples of this Section can be addressed using the standard GARS methodology of Chapter 3. The first one, in Section 6.1.1, is a toy problem where both the GARS algorithm and Evans' method [36] can be applied. In Section 6.1.2, we compare the performance of the GARS technique and the ARMS algorithm of Section 2.7.1 [48] when sampling from a bimodal target pdf. In the third example, Section 6.1.3, we use the GARS method to improve the performance of a standard particle filter. Finally, in order to show how the proposed technique 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 in Section 6.1.4. In this case, we use the GARS technique jointly with a Gibbs sampler.

6.1.1   A toy example

We begin with a simple example in order to illustrate how to apply the GARS technique. Let us consider a target pdf

$$p_o(x) \propto p(x) = \exp\left\{-\sum_{i=0}^{4} a_i x^i\right\},$$

with $a_4 > 0$. 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

$$V(x; \mathbf{g}) = \kappa + (\alpha + \beta x + \gamma x^2)^2 + (\delta + \eta x)^2, \tag{6.1}$$

where κ, α, β, γ, η and δ are real constants and, as a consequence, we can rewrite $V(x; \mathbf{g})$ using our notation as

$$V(x; \mathbf{g}) = \kappa + \bar{V}_1(g_1(x)) + \bar{V}_2(g_2(x)) = \kappa + \sum_{i=1}^{n=2} \bar{V}_i(g_i(x)), \tag{6.2}$$

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 Section 3.2 to this problem.

In particular, let $\mathcal{S}_t = \{s_1, s_2, \ldots, s_{m_t}\}$ 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 $r_{1,k}(x) = a_{1,k}x + b_{1,k}$ (while $r_{2,k}(x) = g_2(x)$ for all x) using the method in Chapter 3 in order to obtain a lower bound for the potential,

$$\begin{aligned}
V(x; \mathbf{r}_k) &= \kappa + V(r_{1,k}(x)) + V(r_{2,k}(x)) \\
&\leq V(x; \mathbf{g}) = \kappa + V(g_1(x)) + V(g_2(x)),
\end{aligned} \tag{6.3}$$

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

$$\pi_t^*(x) \propto \begin{cases} \exp\{-V(x; \mathbf{r}_0)\}, & \text{if } x \in \mathcal{I}_0 \\ \vdots \\ \exp\{-V(x; \mathbf{r}_{m_t})\}, & \text{if } x \in \mathcal{I}_{m_t} \end{cases} \tag{6.4}$$
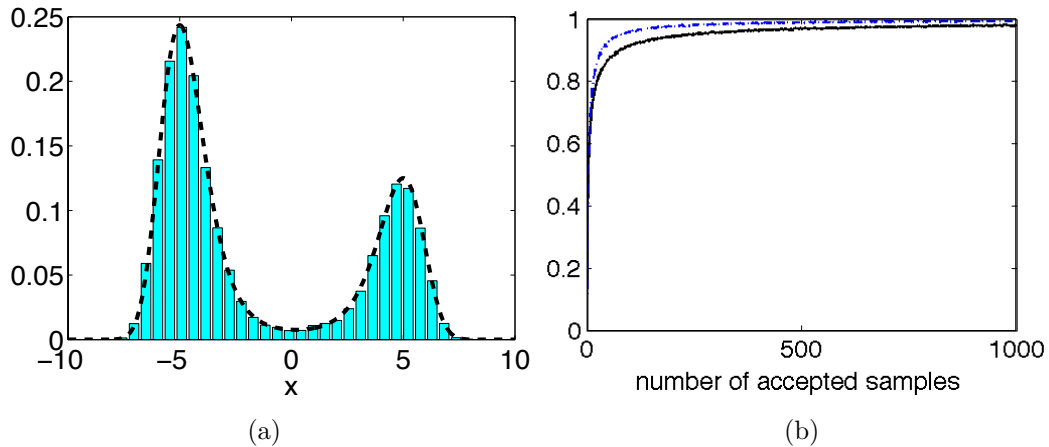
Figure 6.1: **(a)** The target density $p_o(x) \propto \exp\{-(\frac{1}{200}x^4 + \frac{1}{750}x^3 - \frac{1}{4}x^2 + \frac{1}{10}x)\}\}$ (dashed line) and the normalized histogram of $N = 10,000$ samples obtained using the GARS algorithm. **(b)** The curve of acceptance rates (averaged over 10,000 simulations) as a function of the accepted samples using the GARS algorithm with proposal function $\pi_t(x)$ (black solid line) and using the method of [36] with a log-transformation (blue dashed line).

is a piecewise pdf that could be used directly to draw candidate samples if an efficient method to sample from truncated Gaussian pdf's is at hand [85].

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 described in Section 2.2.1.

Figure 6.1 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. 6.1(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 of [36], described in Section 2.7.3, and compare it with the GARS technique.

Figure 6.1(b) shows the acceptance rates[1] (averaged over 10,000 independent simulation runs) versus the number of accepted samples for the GARS algorithm using the proposal functions $\pi_t(x) \propto \exp\{-W_t(x)\}$ (black solid line) and the method of [36] (blue 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 of [36] provides a tighter piecewise linear approximation of $V(x; \mathbf{g})$ and, as a result, attains a higher acceptance rate. Unfortunately, for more complicated distributions the calculation of the inflection points of $V(x; \mathbf{g})$ becomes analytically intractable, as shown in the experimental example of Section 6.1.4.

6.1.2 Comparison of the ARMS and GARS techniques

Consider the problem of sampling a scalar random variable X from a posterior bimodal density

$$p_o(x) = p(x|y) \propto p(y|x)q(x),$$

where the likelihood function is

$$p(y|x) \propto \exp\{-\cosh(y - x^2)\},$$

and the prior pdf is

$$q(x) \propto \exp\{-\alpha(\eta - \exp(|x|))^2\},$$

with constant parameters $\alpha > 0$ and η. Therefore, the posterior pdf is

$$p_o(x) = p(x|y) \propto \exp\{-V(x; \mathbf{g})\}, \tag{6.5}$$

where, $\mathbf{g}(x) = [g_1(x), g_2(x)] = [x^2, \exp(|x|)]$ and the potential function is

$$V(x; \mathbf{g}) = \cosh(y - x^2) + \alpha(\eta - \exp(|x|))^2. \tag{6.6}$$

The marginal potentials are $\bar{V}_1(\vartheta_1) = \cosh(\vartheta_1) = \frac{\exp\{\vartheta_1\} + \exp\{-\vartheta_1\}}{2}$ and $\bar{V}_2(\vartheta_2) = \alpha\vartheta_2^2$. Note that the density $p(x|y)$ is an even function, $p(x|y) =$

---

[1]In order to compute the (empirical) acceptance rate for the i-th sample of the adaptive methods, we have run $M = 10,000$ independent simulations and recorded the numbers $k_{i,j}$, $j = 1, ..., M$, of candidate samples generated in order to accept the i-th sample from $p_o(x)$ in the j-th simulation. The resulting empirical acceptance rate is $\hat{R}_i = \frac{1}{M} \sum_{j=1}^{M} \kappa_{i,j}^{-1}$.

$p(-x|y)$, hence it has a zero mean, $\mu = \int x p(x|y) dx = 0$. The constant α is a scale parameter that allows to control the variance of the random variable X, both *a priori* and *a posteriori*. The higher the value of α, the sharper the modes of $p(x|y)$ become.

There are no closed-form methods to sample directly from $p(x|y)$. Moreover, since the posterior density $p(x|y)$ is bimodal, the system potential is non-log-concave and the conventional ARS technique cannot be applied. However, we can easily use the GARS technique of Chapter 3. If, e.g., $y = 5$ and $\eta = 10$, then the simple estimates corresponding to $g_1(x)$ are $x_{1,1} = -\sqrt{5}$ and $x_{1,2} = \sqrt{5}$, so that $\mathcal{J}_1 = [-\sqrt{5}, \sqrt{5}]$ (see Section 3.2.3). In the same way, the simple estimates corresponding to $g_2(x)$ are $x_{2,1} = -\log(10)$ and $x_{2,2} = \log(10)$, therefore $\mathcal{J}_2 = [-\log(10), \log(10)]$.

An alternative possibility to draw from this density is to use the ARMS method [48]. Therefore, in this section we compare the two algorithms. Specifically, we look into the accuracy of the approximation of the posterior mean $\mu = 0$ by way of the sample mean estimate, $\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} x^{(i)}$, for different values of the scale parameter α.

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

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

Table 6.1 illustrates the estimated posterior mean $\hat{\mu}$ in 5 independent simulations, obtained with the two techniques when $\alpha = 5$. We can see that the ARMS algorithm remains trapped at the negative mode in the first simulation ($\hat{\mu} = -2.2981$) and at the positive mode ($\hat{\mu} = 2.2994$) in the last simulation.

Table 6.1: Estimated posterior mean, $\hat{\mu}$ (for $\alpha = 5$).

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

Figure 6.2(a) shows the posterior density

$$p(x|y) \propto \exp\left\{-\cosh(y_1 - x^2) - \alpha(\mu - \exp(|x|))^2\right\},$$

with $\alpha = 0.2$, depicted as a dashed line, and the normalized histogram obtained with the samples drawn using the GARS technique.

Figure 6.2(b) illustrates the acceptance rates (averaged over $20,000$ simulations) for the first 50 accepted samples drawn with the GARS algorithm. Every time a sample x' drawn from $\pi_t(x)$ is rejected, it is incorporated as a support point. Then, the proposal pdf $\pi_t(x)$ becomes closer to the target pdf $p(x|y)$ and, as a consequence, the acceptance rate becomes higher. For instance, the acceptance rate for the first sample is 16%, but for the second sample, it is already $\approx 53\%$. The acceptance rate for the 20-th sample is $\approx 93\%$ and for the 50-th sample is $\approx 96\%$.

In Figure 6.2(c), we also depicts the average acceptance probability $\hat{a}_t$ defined in Section 3.4 (averaged over $10,000$ simulations) as a function of the iteration index t. The acceptance probability for $t = 1$ is approximately 1.8%, but for $t = 10$, it is already $\approx 71\%$. The acceptance probability $\hat{a}_t$ for $t = 100$ is $\approx 95\%$.

6.1.3 Particle filtering

In filtering applications, the signal of interest (SoI) is a random sequence X_k and the aim is to compute (or approximate) the conditional densities $p(x_k|y_{1:k})$, where $y_{1:k}$ are the observations collected up to time k.

In many applications of signal processing, the SoI is generated by a dynamic system, described as a state-space model. Let us consider, for instance, the non-stationary growth model [34, Chapter 9] given by

$$X_k = 0.5X_{k-1} + 25X_{k-1}/(1 + X_{k-1}^2) + \cos(1.2(k-1)) + \Theta_{1,k}, \qquad (6.7)$$
$$Y_k = X_k^2/20 + \Theta_{2,k}, \qquad (6.8)$$
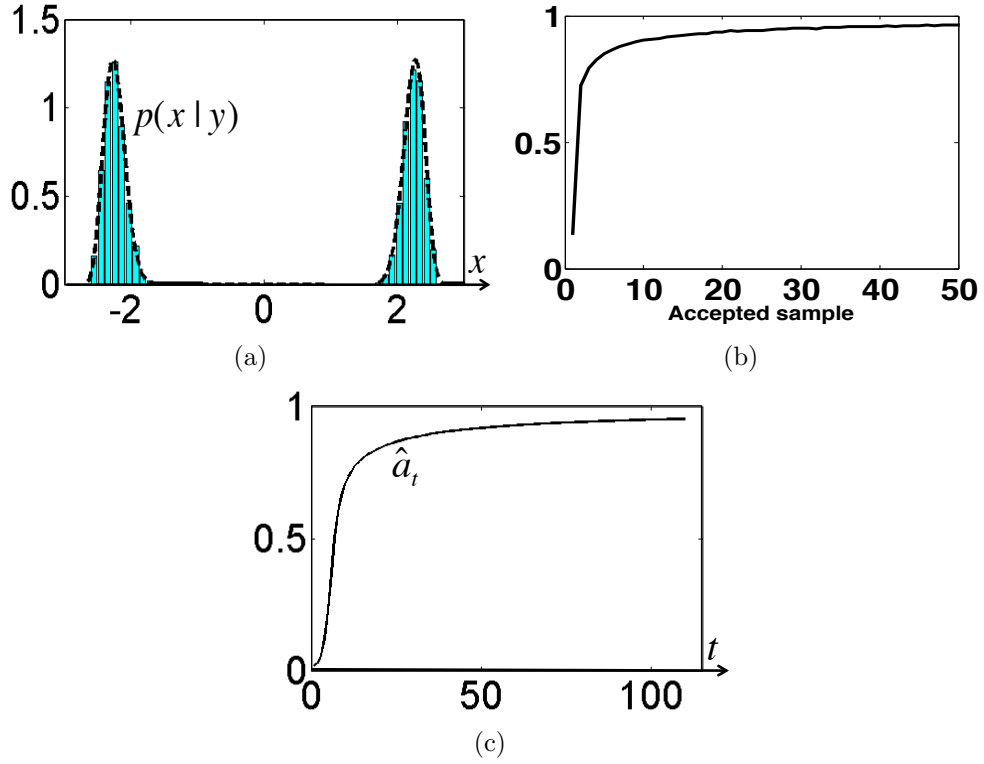
Figure 6.2: (a) The bimodal density $p(x|y) = p_o(x) \propto \exp\{-V(x; \mathbf{g})\}$ (dashed line) and the normalized histogram of $N = 5000$ samples obtained using GARS algorithm. (b) The curve of acceptance rates (averaged over $20,000$ simulations) as a function of the accepted samples. (c) The curve of acceptance rate $\hat{a}_t$ (averaged over $10,000$ simulations) as a function of the iteration index t.

where $k \in \mathbb{N}$ represents discrete time, X_k is the unknown signal of interest, Y_k is the observation, $\Theta_{1,k}$ and $\Theta_{2,k}$ are Gaussian noise variables with densities $\mathcal{N}(\vartheta_{1,k}; 0, 1) \propto \exp\{\vartheta_{1,k}/2\}$ and $\mathcal{N}(\vartheta_{2,k}; 0, 10) \propto \exp\{\vartheta_{2,k}/20\}$. Often Eq. (6.7) is termed "state equation" and Eq. (6.8) is termed "observation equation", meaning that X_k is the (unobserved) state of a system and Y_k is an observation, both at time k.

Due to the nonlinearities in the dynamic model (6.7)-(6.8), we need to apply Monte Carlo techniques to estimate X_k from the observations $Y_{1:k} = y_{1:k}$.

161

The posterior density of X_k given the available data can be expressed as

$$p(x_k|y_{1:k}) \propto p(y_k|x_k) \int p(x_k|x_{k-1})p(x_{k-1}|y_{1:k-1})dx_{k-1}, \qquad (6.9)$$

and we can approximate the integral (6.9) via Monte Carlo by sampling N times from the posterior at time $k-1$. As a result

$$p(x_k|y_{1:k}) \approx p_N(x_k|y_{1:k}) \propto p(y_k|x_k)\frac{1}{N}\sum_{i=1}^{N} p(x_k|x_{k-1}^{(i)}), \qquad (6.10)$$

where $x_{k-1}^{(i)}$, $i = 1, ..., N$, are samples (often called particles) from the distribution with pdf $p(x_{k-1}|y_{1:k-1})$.

One possible procedure is the so called accept-reject particle filter [12, 68, 90, 144, 145]. It entails to sample $x_k^{(i)}$ from the (approximate) density $\frac{1}{N}\sum_{i=1}^{N} p(x_k|x_{k-1}^{(i)})$ and to accept with probability $p(y_k|x_k)/L$, where L is an upper bound for the likelihood. Unfortunately the performance of the filter proposed in [90] is limited, depending on the shape of the likelihood. If the likelihood function is very sharp, the acceptance rate can be very low.

Here, we use a different approach. We can readily rewrite Eq. (6.10) as

$$p(x_k|y_{1:k}) \approx p_N(x_k|y_{1:k}) \propto \frac{1}{M}\sum_{i=1}^{M} p(y_k|x_k)p(x_k|x_{k-1}^{(i)}),$$

and then, in order to draw from $p(x_k|y_{1:k})$, we apply the following procedure:

1. sample uniformly a discrete index $j = 1, \ldots, M$ (each element has probability $1/M$), and then

2. draw the i-th particle $x_t^{(i)}$ from $p_{o,j}(x_t) \propto p(y_k|x_k)p(x_k|x_{k-1}^{(j)})$.

This approach is feasible if, and only if, we are able to draw from every factor $p_{o,j}(x_k) \propto p(y_k|x_k)p(x_k|x_{k-1}^{(j)})$. The potential function $V(x_k; \mathbf{g}) \triangleq -\log[p_{o,j}(x_k)]$ has the form

$$V(x_k; \mathbf{g}) = \frac{1}{2}\left(x_k - \mu_k^{(j)}\right)^2 + \frac{1}{20}\left(y_k - \frac{x_k^2}{20}\right)^2 \qquad (6.11)$$

where $\mu_k^{(j)} \triangleq 0.5x_{k-1}^{(i)} + 25x_{k-1}^{(j)}/(1 + (x_{k-1}^{(j)})^2) + \cos(1.2(k-1))$ and

$$\mathbf{g}(x_k) = [g_1(x_k) = x_k - \mu_k^{(j)}, g_2(x_k) = (y_k - x_k^2/20)/20].$$

The marginal potentials are both quadratic $\bar{V}_1(\vartheta_1) = \frac{1}{2}\vartheta_1^2$, $\bar{V}_2(\vartheta_2) = \frac{1}{20}\vartheta_2^2$, since the noise variables $\Theta_{1,k}$, $\Theta_{2,k}$ are Gaussian.

In general, the potential function in (6.11) is not convex, so that we cannot use the standard ARS method to draw from $p_{o,j}(x_t)$. However, since the marginal potentials and the nonlinearities are convex, we can apply the GARS technique. When a sample is discarded, the GARS procedure improves the proposal pdf in order to increase the probability of accepting the next proposed sample. Hence, in order to take advantage of this adaptive feature of the algorithm, one can draw first N indices $j_1, ..., j_N$, from the set $\{1, ..., N\}$, and then let N_r denote the number of times the index r has been drawn in such way that $N = N_1 + ... + N_N$. Then we generate N_r samples $x_k^{(m)}$, $m = 1, ..., N_r$, from the pdf $p_{o,r}(x_k)$. We repeat the latter step for $r = 1, ..., N$.

We have compared our approach, using the GARS technique, with a standard sequential importance resampling (SIR) filter. We have run $10,000$ independent simulations calculating the absolute error between the real and estimated trajectory of the dynamic system, with $T = 50$ time steps. With the same number of particles, the GARS-based particle filter provides a lower absolute error. However, the mean acceptance rate for one particle is $\approx 30\%$, hence the complexity is higher. To take this issue into account, we have also run simulations with a SIR filter with triple and quadruple number of particles. In these cases, the performance is similar for all the algorithms. Table 6.2 shows the numerical results.

Table 6.2: Absolute error between the real and estimated trajectory (10000 simulations).

Type of Filter	SIR	SIR	SIR	SIR	SIR	SIR-GARS
Number of Particles	30	60	90	120	1000	30
Absolute Error	5,2281	5,0837	4,9910	4,9326	4.7654	4,9660

6.1.4 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 many engineering applications involving sensor networks [3, 126].

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 Figure 6.3 (a). We consider a square monitored area of 4×4 m^2 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 $\mathbf{x} = [x_1 = 2.5, x_2 = 2]$.
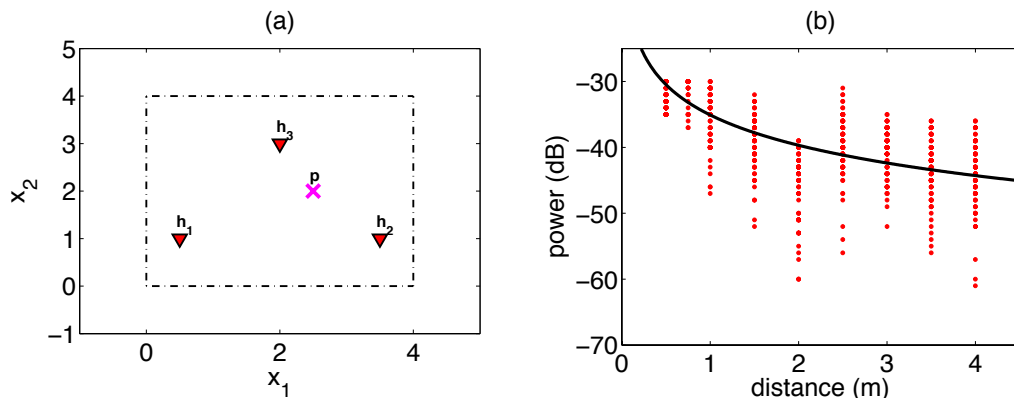
Figure 6.3: **(a)** Deployment of the experimental sensor network over a rectangular surveillance area of 4×4 m^2. The sensors are depicted with triangles while the target is depicted with a cross. **(b)** The least squares regression to adjust the parameters l and γ. The points indicate the measurements collected by the sensors at different distances, and the solid curve denotes the function $\hat{l} - 10\hat{\gamma} \log \left[\frac{d}{d_0} \right]$ with $d_0 = 0.3$, $\hat{l} = -27.08$ dB and $\hat{\gamma} = 1.52$.

The measurement provided by the i-th sensor is modeled as a random variable Y_i. To describe the relationship between the observed radio signal

strength, $Y_i = y_i$, and the target position, modeled by the random vector $\mathbf{X} = [X_1, X_2]$, we use the free space propagation model [130]

$$Y_i = l - 10\gamma \log \left[\frac{D_i}{d_0} \right] + \Theta_i \quad \text{(dB)}, \tag{6.12}$$

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 $\mathcal{N}(\vartheta_i; 0, \sigma^2) \propto \exp\left\{-\frac{\vartheta_i^2}{2\sigma^2}\right\}$ 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 have been fitted by least squares regression using 200 measurements with the target placed at known distances from each sensor. As a result, we have obtained $\hat{l} = -27.08$ dB, $\hat{\gamma} = 1.53$ and $\hat{\sigma} = 4.41$. Figure 6.3(b) depicts the measurements at several distances and the fitted curve $\hat{l} - 10\hat{\gamma} \log \left[\frac{d}{d_0} \right]$.

Algorithm

Assume we collect M independent measurements from each sensor using the experimental setup just described. Let

$$\mathbf{Y} = [Y_{1,1}, \ldots, Y_{1,M}, Y_{2,1}, \ldots, Y_{2,M}, Y_{3,1}, \ldots, Y_{3,M}]$$

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 Eq. (6.12), i.e.,

$$p(\mathbf{y}|\mathbf{x}) = \prod_{q=1}^{3} \prod_{m=1}^{M} N\left(y_{q,m}; \hat{l} - 10\hat{\gamma} \log(||\mathbf{X} - \mathbf{h}_q||/d_0), \hat{\sigma}^2\right). \tag{6.13}$$

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

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}), \tag{6.14}$$

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

$$p(x_k) = \mathcal{N}(x_k; 1.5, 1/2), \qquad k = 1, 2. \tag{6.15}$$

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 $\mathbf{x}^{(j)} = [x_1^{(j)}, x_2^{(j)}]$.

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

$$V(x_k; \mathbf{g}) = \sum_{i=1}^{3M} \bar{V}_i(g_i(x_k)) + \bar{V}_{3M+1}(g_{3M+1}(x_k)), \tag{6.16}$$

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

$$\bar{V}_i(g_i(x_k)) = \left[y_{q,m} - \hat{l} + 10\hat{\gamma} \log\left(||\mathbf{x}_k^{(j)} - \mathbf{h}_q||/d_0 \right) \right]^2, \tag{6.17}$$

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

$$\bar{V}_{3M+1}(g_{3M+1}(x_k)) = \left(x_k^{(j)} - 1.5 \right)^2. \tag{6.18}$$

166

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

$$g_{(q-1)M+m}(x_k) = y_{q,m} - \hat{l} + 10\hat{\gamma} \log \left(||\mathbf{x}_k^{(j)} - \mathbf{h}_q||/d_0 \right), \qquad (6.19)$$

for $q = 1, 2, 3$ (sensors), $m = 1, \ldots, M$ (measurements), $k = 1, 2$ and $j = 1, \ldots, N$, 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., $\bar{V}_i(\vartheta) = \frac{1}{\hat{\sigma}^2}\vartheta^2$ for $i = 1, \ldots, 3M$, 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 can not 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 $g_i(x_k)$, $i = 1, \ldots, 3M + 1$, can be partitioned as described in Section 5.2.1. As a consequence, we can use the proposed GARS technique to implement the Gibbs sampler. On the contrary, the form of Eqs. (6.16), (6.17) and (6.18) makes the calculation of the inflection points (with respect x_k) intractable and, therefore, the methods of [36, 61] (see Section 2.7.3) and [54] (see Section 2.7.2) are not applicable in this example.

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 $\mathcal{S}_0$.

Figure 6.4(a) shows the shape of the true target density $p(x_1, x_2|\mathbf{y})$ with $M = 1$. Figure 6.4(b) depicts the corresponding histograms using $N = 30,000$ samples.

Figure 6.5(a) displays the shape of the true target density $p(x_1, x_2|\mathbf{y})$ when we have $M = 3$ measurements. Figure 6.5(b) depicts the corresponding

Figure 6.4: **(a)** The shape of the true target density $p(x_1, x_2|\mathbf{y})$ with $M = 1$ measurement. **(b)** The normalized histogram, using $N = 20,000$ samples, corresponding to $p(x_1, x_2|\mathbf{y})$ with $M = 1$ measurement per sensor.

histograms using $N = 30,000$ samples. We can observe that Figures 6.4(b) and 6.5(b) approximate closely the shape of target pdf's.

Finally, Figures 6.6(a) and 6.6(b) illustrate the normalized histograms corresponding to the number of proposed candidates which are needed to accept one sample.

Note that we obtain an empirical approximation of the posterior distribution that is very accurate although in terms of the localization accuracy the performance is relatively poor, as there is a bias between the mode of $p(\mathbf{x}|\mathbf{y})$ and the actual target position.

6.2 Log-convex tails

In this section we illustrate the application of the techniques proposed in Chapter 4. The first example of Section 6.2.1 is devoted to compare the performance of the GARS technique of Section 4.2 and the ARoU algorithm of Section 4.3 using an artificial model.

In the second example, in Section 6.2.2, we implement an adaptive version of the accept/reject particle filter (ARPF) in [90] using the technique proposed in Section 4.2.

In Section 6.2.3, we apply the ARoU scheme as a building block in another type of accept/reject particle filter for inference in a financial volatility model. In this third example, the method of Section 4.2 cannot be implemented.
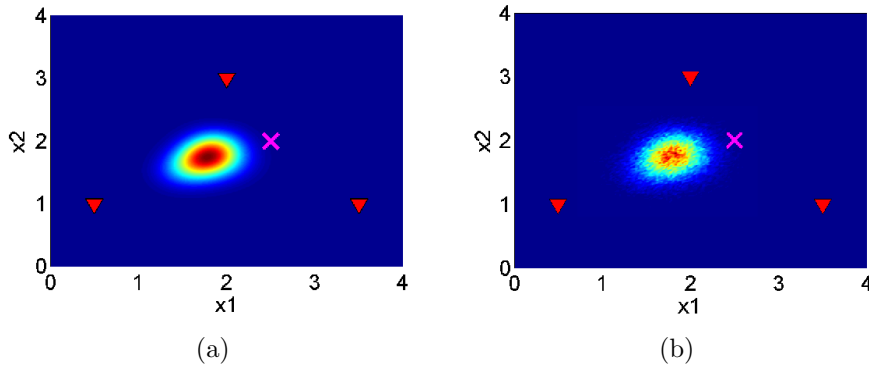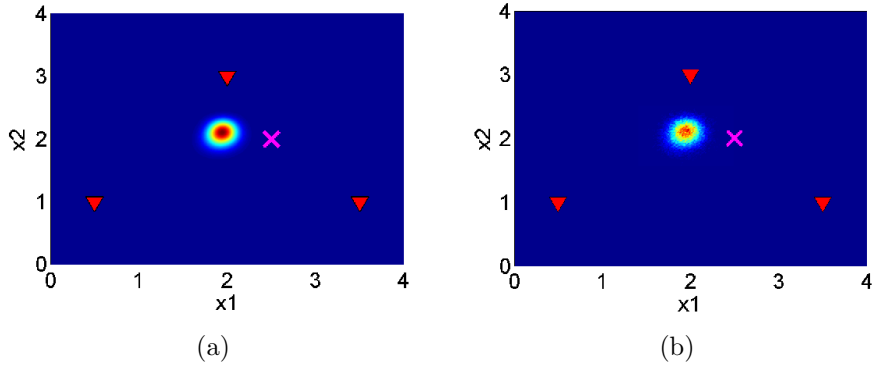
Figure 6.5: **(a)** The shape of the true target density $p(x_1, x_2|\mathbf{y})$ with $M = 3$ measurements. **(b)** The normalized histogram, using $N = 20,000$ samples, corresponding to $p(x_1, x_2|\mathbf{y})$ with $M = 3$ measurements per sensor.

It should be noted that the various generalizations of the standard ARS algorithm, such as the TDR technique of Section 2.7.3 [61], the concave-convex ARS algorithm of Section 2.7.2 [55] and other methods [36, 113, 114] are all invalid in the first and third examples.

6.2.1 Drawing from a posterior pdf

Let $X > 0$ be a scalar r.v. with exponential prior pdf $q(x) \propto \exp\{-\lambda x\}$, $\lambda > 0$. We collect three observations $\mathbf{Y} = [Y_1, Y_2, Y_3]$, related to X, of the form

$$\begin{cases} Y_1 = a\exp(-bX) + \Theta_1, \\ Y_2 = c\log(dX+1) + \Theta_2, \\ Y_3 = (X-e)^2 + \Theta_3, \end{cases} \qquad (6.20)$$

where Θ_1, Θ_2, Θ_3 are independent noise variables and a, b, c, d, e are constant parameters. Specifically, Θ_1 and Θ_2 have generalized gamma pdf's $\Gamma_g(\vartheta_i; \alpha_i, \beta_i) \propto \vartheta_i^{\alpha_i} \exp\{-\vartheta_i^{\beta_i}\}$, $i = 1, 2$, with parameters $\alpha_1 = 4$, $\beta_1 = 2$ and $\alpha_2 = 2$, $\beta_2 = 2$, respectively. The variable Θ_3 has a Gaussian density $\mathcal{N}(\vartheta_3; 0, 1/2) \propto \exp\{-\vartheta_3^2\}$.

For a fixed vector of observations $\mathbf{Y} = \mathbf{y} = [y_1, y_2, y_3]$, our goal is to draw from the posterior pdf $p(x|\mathbf{y}) \propto p(\mathbf{y}|x)q(x)$. Given Eq. (6.20), the target

(a)   (b)
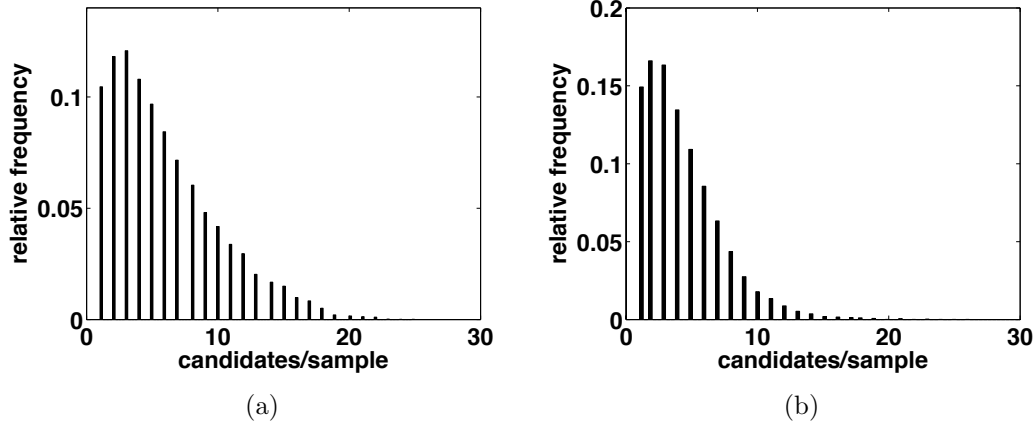
Figure 6.6: The normalized histogram corresponding to the number of proposed candidates which are needed to accept one sample, with **(a)** $M = 1$ measurement or **(b)** with $M = 3$ measurements.

density can be written as

$$p_o(x) = p(x|\mathbf{y}) \propto \exp\{-V(x; \mathbf{g})\}, \qquad (6.21)$$

where the potential function is

$$
\begin{aligned}
V(x; \mathbf{g}) =\ & (y_1 - a\exp(-bx))^2 + \\
& - \log[(y_1 - a\exp(-bx))^4] + (y_2 - c\log(dx + 1))^2 + \\
& - \log[(y_2 - c\log(dx + 1))^2] + (y_3 - (x - e)^2)^2 + \\
& + \lambda x.
\end{aligned}
\qquad (6.22)
$$

From (6.22), it is straightforward to identify four marginal potentials, i.e.,

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

where $\bar{V}_1(\vartheta) = \vartheta^2 - \log[\vartheta^4]$, $\bar{V}_2(\vartheta) = \vartheta^2 - \log[\vartheta^2]$, $\bar{V}_3(\vartheta) = \vartheta^2$ and $\bar{V}_4(\vartheta) = \lambda|\vartheta|$. The associated nonlinearities are

$$
\begin{aligned}
g_1(x) &= a\exp(-bx), \\
g_2(x) &= c\log(dx + 1), \\
g_3(x) &= (x - e)^2, \\
g_4(x) &= x.
\end{aligned}
\qquad (6.23)
$$

170

Since all the marginal potentials are convex and all the nonlinearities are either convex or concave, we can apply the ARoU algorithm of Section 4.3. Moreover, since the fourth marginal potential yields a simple pdf $\exp\{-\bar{V}_4(g_4(x))\} = \exp\{-\lambda x\}$ (actually the prior of the r.v. X) that can be easily integrated and sampled in arbitrary intervals, we can also apply the adaptive rejection sampler of Section 4.2.

Note, however, that

- the potential $V(x; \mathbf{g})$ in Eq. (6.22) is not convex,

- the target function $p(x|\mathbf{y}) \propto \exp\{-V(x; \mathbf{g})\}$ can be multimodal,

- the tails of the potential $V(x; \mathbf{g})$ are not convex (i.e., the tails of $\log p_o(x)$ are not concave) and

- it is not possible to obtain the zeros of the first and second derivatives of the potential $V(x; \mathbf{g})$, i.e., we cannot solve the equations $V'(x; \mathbf{g}) = 0$ and $V''(x; \mathbf{g}) = 0$ analytically.

Therefore, the techniques in [50, 36, 55, 61, 114] and the basic GARS method cannot be used in this problem.

In order to use the GARS procedure in Section 4.2 note that 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 it determines the form of the likelihood function, i.e.,

$$p(\mathbf{y}|x) = \exp\{-V_{-4}(x; \mathbf{g})\}.$$

To apply the ARoU algorithm, 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 have assumed $X \geq 0$, it is not necessary to study the potential $V^{(3)}$.

For the simulations, we set $a = -2$, $b = 1.1$, $c = -0.8$, $d = 1.5$, $e = 2$, $\lambda = 0.2$ and $\mathbf{y} = [2.314, 1.6, 2]$, and initialize the algorithms with the set of support points $\mathcal{S}_0 = \{0, 2 - \sqrt{2}, 2, 2 + \sqrt{2}\}$.

Figure 6.7(a) depicts, jointly, the likelihood function $p(\mathbf{y}|x) = \exp\{-V_{-4}(x; \mathbf{g})\}$ and its stepwise approximation $\exp\{-\gamma_k\} \ \forall x \in \mathcal{I}_k$, $k = 0, ..., m_t$. The computation of the lower bounds $\gamma_k \leq V_{-4}(x; \mathbf{g})$ has been carried out according to the procedure in Section 4.2.2. Figure 6.7(b) depicts the transformed set $\mathcal{A}$ (solid) that corresponds to the posterior density $p(x|\mathbf{y})$, and the region $\mathcal{P}_t$ formed by triangular pieces, generated by the ARoU algorithm with $m_t = 9$ support points.

(a)   (b)
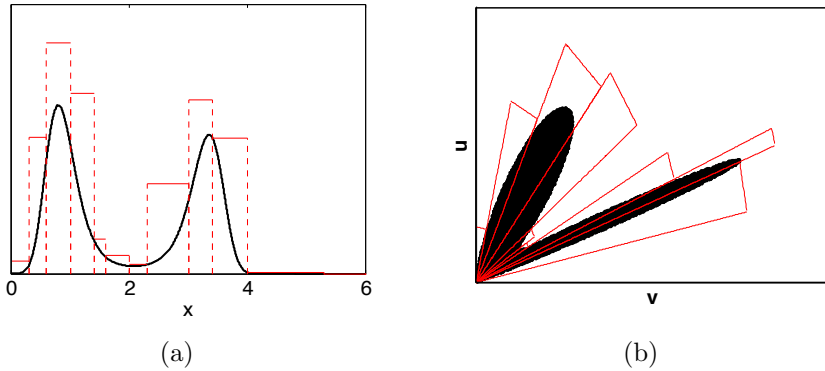
Figure 6.7: **(a)** The function $\exp\{-V_{-4}(x;\mathbf{g})\}$ obtained using the reduced potential $V_{-4}(x;\mathbf{g}) = V(x;\mathbf{g}) - \bar{V}_4(g_4(x))$ and its constant upper bounds $\exp\{-\gamma_k\}$, $k = 0, ..., m_t$. In this example, the function $p(\mathbf{y}|x) = \exp\{-V_{-4}(x;\mathbf{g})\}$ coincides with the likelihood. **(b)** The set $\mathcal{A}$ corresponding to the RoU transformation of the target pdf $p(x|\mathbf{y})$ and the region $\mathcal{P}_t = \cup_{k=0}^{m_t=9} \mathcal{T}_k$, formed by non-overlapping triangles, constructed using the ARoU sampling scheme.

The simulations show that both methods attain very similar acceptance rates. In Figure 6.8(a) and 6.8(b), we plot the empirical acceptance rates (averaged over 10,000 independent simulation runs) versus the first $1,000$ accepted samples, using the GARS algorithm of Section 4.2 and the ARoU method of Section 4.3, respectively. We can see that the two samplers are equally efficient and the rates quickly converge close to 1.

6.2.2 Stochastic volatility model 1

We address the implementation of an accept/reject particle filter [12, 68, 90, 144, 145] to track the volatility of a financial time series Y_k, $k \in \mathbb{N}$. In particular, we adopt the state space model [159]

$$\begin{cases} X_k = \beta X_{k-1} + \Theta_{2,k}, \\ Y_k = \exp(-X_k)\Theta_{1,k}, \end{cases} \tag{6.24}$$

where X_k is the state at time k, Y_k is the value of the financial series at time k, β is a constant value, $\Theta_{2,k} \sim \mathcal{N}(\mu, \sigma_1^2)$ and $\Theta_{1,k} \sim \mathcal{N}(0, \sigma_2^2)$ are Gaussian noise variables with pdf's $p(\vartheta_{1,k}) \propto \exp\{-(\vartheta_{1,k} - \mu)^2/2\sigma_1^2\}$ and
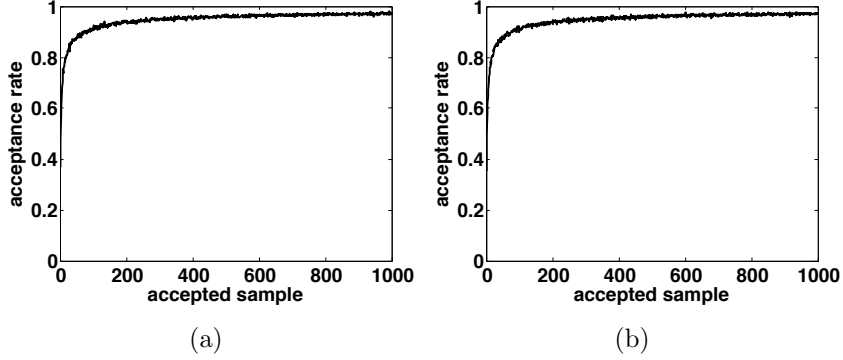
Figure 6.8: **(a)** The curve of acceptance rates (averaged over 10,000 simulations) as a function of the first 1,000 accepted samples using the ARS scheme of Section 4.2. **(b)** The curve of acceptance rates (averaged over 10,000 simulations) as a function of the first 1,000 accepted samples using the ARoU algorithm.

$p(\vartheta_{2,k}) \propto \exp\{-(\vartheta_{2,k})^2/2\sigma_2^2\}$, respectively. Note that the model in Eq. (6.24) can be rewritten in the equivalent form

$$\begin{cases} X_k = \beta X_{k-1} + \Theta_{2,k}, \\ 0 = -Y_k \exp(X_k) + \Theta_{1,k}, \end{cases} \tag{6.25}$$

following the argument in Section 3.7.1.

We want to implement a particle filter to make inference about x_k. 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 of X_k given $Y_{1:k-1} = y_{1:k-1}$ as

$$\begin{aligned} p(x_k|y_{1:k-1}) &= \int p(x_k|x_{k-1})p(x_{k-1}|y_{1:k-1})dx_{k-1} \\ &\approx \frac{1}{N}\sum_{i=1}^N p(x_k|x_{k-1}^{(i)}) \end{aligned} \tag{6.26}$$

and then the filtering pdf as [33]

$$p(x_k|y_{1:k}) \approx p_N(x_k|y_{1:k}) \propto p(y_k|x_k)\frac{1}{N}\sum_{i=1}^N p(x_k|x_{k-1}^{(i)}), \tag{6.27}$$

173

that we can also be written as

$$p_N(x_k|y_{1:k}) = \frac{1}{N}\sum_{i=1}^{N} p_{o,i}(x_k) \propto \frac{1}{N}\sum_{i=1}^{N} p(y_k|x_k)p(x_k|x_{k-1}^{(i)}),$$

where we have denoted $p_{o,i}(x_k) \propto p(y_k|x_k)p(x_k|x_{k-1}^{(i)})$,

If we can draw exact samples $\{x_k^{(i)}\}_{i=1}^{N}$ from (6.27) using a RS scheme, then the integrals of measurable functions w.r.t. to the filtering pdf $I(f) = \int f(x_k)p(x_k|y_{1:k})dx_k$ can be approximated as $I(f) \approx I_N(f) = \frac{1}{N}\sum_{i=1}^{N} f(x_k^{(i)})$. With this purpose, we can use the GARS scheme of Section 4.2 at each time step k in order to build an adaptive version of the accept/reject particle filter (ARPF) in [90].

Specifically, we first draw uniformly a random index $j \in \{1, ..., N\}$ and then we draw a particle, using the GARS scheme of Section 4.2, from

$$p_{o,j}(x_k) \propto \exp\{-V(x_k; \mathbf{g})\} = p(y_k|x_k)p(x_k|x_{k-1}^{(j)}),$$

where the potential function[2] is

$$V(x_k; \mathbf{g}) = \frac{1}{2\sigma_2^2}\left(y_k\exp(x_k) - \mu\right)^2 + \frac{1}{2\sigma_1^2}(x_k - \beta x_{k-1}^{(j)})^2. \qquad (6.28)$$

It obviously decomposes naturally into two marginal potentials,

$$V(x_k; \mathbf{g}) = \bar{V}_1(g_{1,k}(x_k)) + \bar{V}_2(g_{2,k}(x_k)),$$

where $\bar{V}_1(\vartheta_1) = \frac{1}{2\sigma_2^2}\vartheta_1^2$, $\bar{V}_2(\vartheta_2) = \frac{1}{2\sigma_1^2}\vartheta_2^2$ and

$$\mathbf{g}(x_k) = [g_{1,k} = y_k\exp(x_k) - \mu, g_{2,k} = x_k - \beta x_{k-1}^{(j)}].$$

Using this notation, the likelihood function

$$p(y_k|x_k) \propto \exp\left\{-\frac{1}{2\sigma_2^2}\left(y_k\exp(x_k) - \mu\right)^2\right\} \qquad (6.29)$$

can be denoted as

$$p(y_k|x_k) \propto \exp\{-\bar{V}_1(g_{1,k}(x_k))\} = \exp\{-V_{-2}(x_k; \mathbf{g})\}, \qquad (6.30)$$

---

[2]Note that the nonlinearities in $\mathbf{g}$ vary each time, we avoid explicitly showing it in the notation for simplicity.

where $V_{-2}(x_k; \mathbf{g}) = \bar{V}_1(g_{1,k}(x_k))$ is a *reduced* potential, following the terminology in Section 4.2. Moreover, we have

$$p(x_k|x_{k-1}^{(j)}) \propto \exp\left\{-\bar{V}_2(g_{2,k}(x_k))\right\} = \exp\left\{-\frac{1}{2\sigma_1^2}(x_k - \beta x_{k-1}^{(j)})^2\right\},$$

that is a standard Gaussian density $\mathcal{N}(x_k; \beta x_{k-1}, \sigma_1^2)$. Notice that we can draw from it when truncated within a finite interval [85].

Therefore, we can use the GARS procedure of Section 4.2, approximating the likelihood function with piecewise constant upper bounds $L_{t,k}(x_k)$ and generating samples from $p(x_k|x_{k-1}^{(j)})$, i.e., our proposal pdf is

$$\pi_t(x_k) = L_{t,k}(x_k)p(x_k|x_{k-1}^{(j)}). \tag{6.31}$$

We draw x' from $\pi_t(x_k)$ and then accept it, $x_k^{(i)} = x'$, with probability

$$p_a = \frac{p(y_k|x_k)p(x_k|x_{k-1}^{(j)})}{L_{t,k}(x_k)p(x_k|x_{k-1}^{(j)})} = \frac{p(y_k|x_k)}{L_{t,k}(x_k^{(i)})}.$$

If the sample is discarded, we add it to the set of support points

$$\mathcal{S}_{t+1,k} = \mathcal{S}_{t,k} \cup \{x'\}$$

in order to improve the stepwise approximation $L_{t+1,k}(x_k)$ of the likelihood function. Otherwise, if it is accepted, we set $\mathcal{S}_{t+1,k} = \mathcal{S}_{t,k}$ and $L_{t+1,k}(x_k) = L_{t,k}(x_k)$.

Note that, since in general the potential $V(x_k; \mathbf{g})$ has one concave tail (i.e., $p_{0,j}(x_k)$ has one log-convex tail), the standard ARS of Section 2.6, the concave-convex ARS of Section 2.7.2 and the GARS of Chapter 3 cannot be used. Figure 6.9(a) depicts an example of the reduced potential function $V_{-2}(x_k; \mathbf{g})$ (recall that $V(x_k; \mathbf{g}) = V_{-2}(x_k; \mathbf{g}) + \bar{V}_2(g_{2,k}(x_k))$).

In order to take advantage of the adaptive nature of the proposed technique, one can draw first N indices $j_1, ...j_N$, with $j_i \in \{1, ..., N\}$, and then we sample N_i particles $x_k^{(m)}$ from the same proposal, $p_{o,j}(x_k) \propto L_{t,k}(x_k)p(x_k|x_{k-1}^{(i)})$, $m = 1, ..., N_i$, where N_i is the number of times the index $i \in \{1, ..., N\}$ has been drawn. Obviously, $N_1 + N_2 + ... + N_N = N$.

We have used the set of parameters $\beta = 1$, $\mu = 1$, $\sigma_1 = 1$ and $\sigma_2 = 0.3$ to carry out some illustrative simulations. The acceptance rate, averaged

over 30 time steps and $10{,}000$ independent simulation runs, was $\approx 34\%^3$. Note that this is a relevant figure, since only N_i particles (a number typically small) are drawn from the same target pdf, hence only a few support points can be used. This acceptance rate can be improved adding support points in the initial set $\mathcal{S}_0$ to the detriment of the computational cost.

Table 6.3: Mean Square Error.

Number of particles N	10	20	50
Boostrap Filter	1.31	0.50	0.24
Adaptive ARPF	0.92	0.38	0.22

(a) (b)
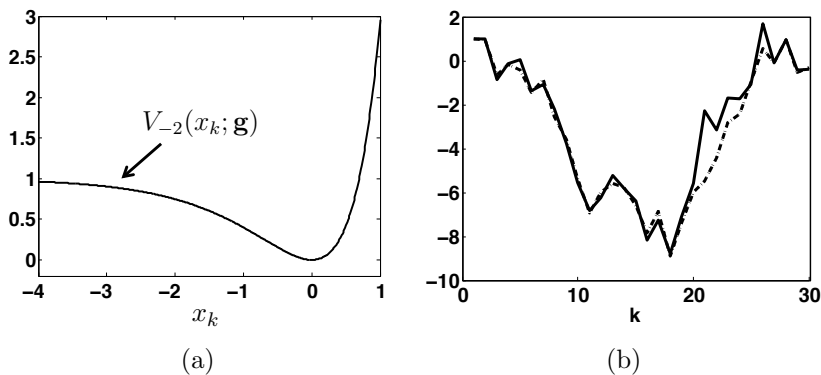
Figure 6.9: **(a)** An example of reduced potential $V_{-2}(x_k; \mathbf{g}) = \big(y_k \exp(x_k) - \mu\big)^2 / 2\sigma_1^2$ when $y_k = 1$, $\mu = 1$ and $\sigma_1^2 = 1/2$. The left tail is concave. **(b)** An example of trajectory of the signal of interest x_k with 30 time steps. The solid line illustrates the real trajectory while the dashed line shows the estimated trajectory obtained by the adaptive ARPF.

Figure 6.9(b) displays an example of real trajectory (solid line) and its estimation (dashed line) with the adaptive ARPF. For the sake of comparison, we have also applied the standard bootstrap filter with prior importance function [34] in the same set of simulations. Table 6.3 shows the mean square error (MSE) achieved in the estimation of x_k, using different

---

[3]Since the target pdf changes each time step k, the GARS restarts each k from the iteration index $t = 0$.

numbers of particles, with the standard bootstrap filter and with the ARPF. The ARPF obtains a better performances with a lesser number of particles but the computational effort is greater because we have to generate more candidate particles in order to accept one. However, it should be remarked that the ARPF is specially advantageous w.r.t. to the bootstrap filter when there is a significant discrepancy between the likelihood and prior functions.

6.2.3 Stochastic volatility model 2

In this example, we implement a particle filter for a stochastic volatility model, based on the ARoU scheme of Section 4.3. Let be $X_k \in \mathbb{R}^+$ the volatility of a financial time series at time k, and consider the following state space system [34, Chapter 9],[69] with a scalar observation, $Y_k \in \mathbb{R}$,

$$\begin{cases} \log(X_k^2) = \beta \log(X_{k-1}^2) + \Theta_{2,k}, \\ Y_k = \log(X_k^2) + \Theta_{1,k}, \end{cases} \tag{6.32}$$

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

Given the state-space system of Eq. (6.32), the prior pdf at time k is

$$p(x_k|x_{k-1}) \propto \exp\left\{-\frac{\log(x_k^2) - \beta \log(x_{k-1}^2)}{2\sigma^2}\right\}, \tag{6.33}$$

and the likelihood function is

$$p(y_k|x_k) \propto \exp\left\{\frac{y_k - \log(x_k^2)}{2} - \frac{\exp(y_k - \log(x_k^2))}{2}\right\}$$

$$= \exp\left\{\frac{y_k - \log(x_k^2)}{2} - \frac{\exp(y_k) - 1/x_k^2}{2}\right\}.$$

We can apply the proposed ARoU scheme of Section 4.3 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

$$\begin{aligned} p(x_k|y_{1:k-1}) &= \int p(x_k|x_{k-1})p(x_{k-1}|y_{1:k-1})dx_{k-1} \\ &\approx \frac{1}{N}\sum_{i=1}^N p(x_k|x_{k-1}^{(i)}) \end{aligned} \tag{6.34}$$

and then the filtering pdf as [33]

$$p(x_k|y_{1:k}) \approx p_N(x_k|y_{1:k}) \propto p(y_k|x_k) \frac{1}{N} \sum_{i=1}^{N} p(x_k|x_{k-1}^{(i)}). \qquad (6.35)$$

As in the previous example, if we can draw exact samples $\{x_k^{(i)}\}_{i=1}^{N}$ from (6.35) using a RS scheme, then the integrals of measurable functions w.r.t. to the filtering pdf $I(f) = \int f(x_k)p(x_k|y_{1:k})dx_k$ can be approximated as $I(f) \approx I_N(f) = \frac{1}{N} \sum_{i=1}^{N} f(x_k^{(i)})$.

To implement the sampler, we first recall that the problem of drawing $x_k^{(i)}$ from the pdf in Eq. (6.35) can be reduced to generate an index $j \in \{1, ..., N\}$ with uniform probabilities, $1/N$, and then draw from the pdf

$$p_{o,j}(x_k) \propto p(y_k|x_k)p(x_k|x_{k-1}^{(j)}). \qquad (6.36)$$

Using the ARoU technique, in order to take advantage of the adaptive feature of the algorithm, one can draw first N indices $j_1, ..., j_N$, from the set $\{1, ..., N\}$, and then let N_r denote the number of times the index r has been drawn in such way that $N = N_1 + ... + N_N$. Then we generate N_r samples $x_k^{(m)}$, $m = 1, ..., N_r$, from the pdf $p_{o,r}(x_k)$. We repeat the latter step for $r = 1, ..., N$.

The potential function associated with the pdf in Eq. (6.36) is

$$V(x_k; \mathbf{g}) = -\frac{y_k - \log(x_k^2)}{2} + \frac{\exp\{y_k\}}{2} + +\frac{1}{2x_k^2} + \frac{\left(\log(x_k^2) - \alpha_k\right)^2}{2\sigma^2}, \qquad (6.37)$$

where $\alpha_k = \beta \log\left[(x_{k-1}^{(j)})^2\right]$ is a constant and the nonlinearities are

$$\begin{aligned} g_1(x_k) &= \log(x_k^2), \\ g_2(x_k) &= -\log(x_k^2) + \alpha_k. \end{aligned} \qquad (6.38)$$

From Eq. (6.37) we identify the marginal potentials

$$V(x_k; \mathbf{g}) = \bar{V}_1\big(g_1(x_k)\big) + \bar{V}_2\big(g_2(x_k)\big), \qquad (6.39)$$

where $\bar{V}_1(\vartheta_1) = \frac{1}{2}\big(-\vartheta_1 + \exp\{\vartheta_1\}\big)$ and $\bar{V}_2(\vartheta_2) = \frac{1}{2\sigma^2}\vartheta_2^2$.

Let us remark that in this example:

178

- Since the potential $V(x_k; \mathbf{g})$ is in general a non-convex function and it is not possible to find the zeros of the first and second derivatives analytically, the methods in [36, 50] cannot be applied.

- Since the potential function $V(x_k; \mathbf{g})$ possibly has concave tails, as shown in Figure 6.10(a), the method in [55] and the basic adaptive technique described in [113] cannot be used either.

- Finally, we cannot apply the GARS procedure of Section 4.2 because there are no direct techniques to draw from (and to integrate) the functions $\exp\{-\bar{V}_1(g_1(x_k))\}$ or $\exp\{-\bar{V}_2(g_2(x_k))\}$.

However, 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 $g_2(x_k)$, it is possible to apply the ARoU scheme of Section 4.3.

(a)   (b)
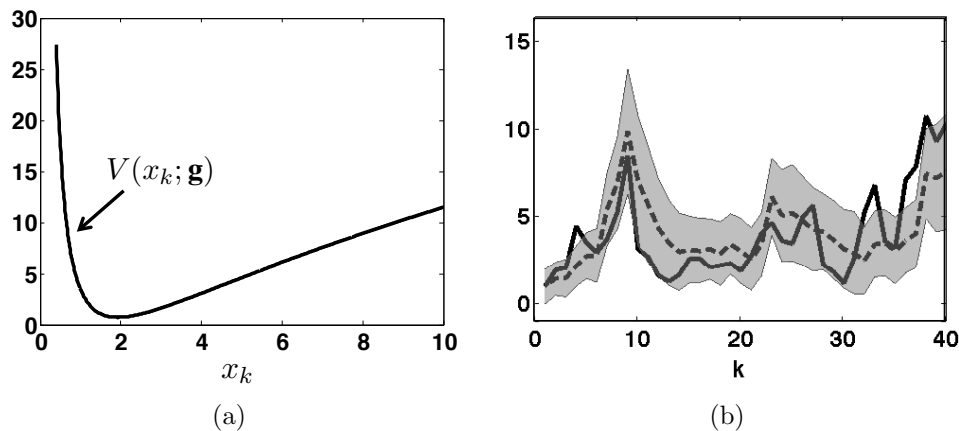
Figure 6.10: **(a)** A realization of the potential function $V(x_k; \mathbf{g})$ defined in Eq. (6.37), when $\alpha_k = 1$, $y_k = 2$ and $\sigma = 0.8$. The right tail is concave. **(b)** An example of a trajectory with 40 time steps of the signal of interest X_k. The solid line illustrates the true trajectory generated by the system in Eq. (6.32) with $\beta = 0.8$ and $\sigma = 0.9$, while the dashed line shows the estimated trajectory computed by the particle filter with $N = 1,000$ particles. The shadowed area shows one standard deviation from the estimates.

If we set the constant parameters as $\beta = 0.8$ and $\sigma = 0.9$, we obtain an empirical 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, because it is not straightforward to draw from the prior density in Eq. (6.33). Indeed, there are no direct methods to sample from this prior pdf and, in general, we need to use rejection sampling or MCMC techniques.

Figure 6.10(b) depicts 40 time steps of a trajectory (solid line) of the signal X_k generated by the system in Eq. (6.32), with $\beta = 0.8$ and $\sigma = 0.9$. In dashed line, we see the estimated trajectory obtained by the particle filter using the ARoU scheme with $N = 1,000$ particles. The shadowed area shows one standard deviation from the estimates obtained by the filter. The mean square error (MSE) achieved in the estimation of X_k using $N = 1,000$ particles is 1.48.

6.3   Automatic procedure

In this section, we implement a simple example in order to compare the performance of the alternative simplified algorithm in Section 5.3 with the standard GARS technique. Let us consider a target pdf of the form

$$p_o(x) \propto \exp\{-V(x; \mathbf{g})\} = \exp\{-(-4 - x + x^2)^2\}, \qquad (6.40)$$

where the potential is $V(x; \mathbf{g}) = \bar{V}_1(g_1(x)) = (-4 - x + x^2)^2$ with $\bar{V}_1(\vartheta_1) = \vartheta_1^2$ and $g_1(x) = -4 - x + x^2$.

In this simple case, we know both the minimum $\mu_1 = 0$ of $\bar{V}_1$ and the simple estimates $x_{1,1} = -1.56$ and $x_{1,2} = 2.56$ (solutions of the equation $g_1(x) = \mu_1 = 0$). Hence, we can apply the standard GARS technique. However, we have also implemented different simplified GARS (S-GARS) algorithms in order to assess the loss in the performance w.r.t. the standard procedure. Specifically, we have applied 4 different simulations:

1) the standard GARS method of Chapter 3, using the known values μ_1, $x_{1,1}$ and $x_{1,2}$;

2) the alternative procedure of Section 5.3.1 assuming the minimum μ_i is unknown but with calculating the corresponding simple estimates;

3) the alternative procedure of Section 5.3.2 assuming the value μ_1 is known but without calculating the simple estimates;

4) and, finally, an algorithm assuming that μ_1 is unknown and $g_1(x)$ can not be inverted.

The algorithms built under the assumptions 2), 3) and 4) are termed S-GARS-2, S-GARS-3 and S-GARS-4, respectively.

Figures 6.11(a), (b), (c), (d) depict the acceptance rates (averaged over 10,000 independent simulation runs) versus the number of the accepted samples attained by the four methods. The numerical results show that S-GARS-2 and S-GARS-4 techniques (in both the μ_1 is unknown) yield the worst performance in terms of acceptance rate. However, the procedure of Section 5.3.1 has some advantages:

1. The same strategy can also be applied to a broader family of densities, for instance, with marginal potentials $\bar{V}_i$ with several stationary points (see Section 5.2.5) or when some marginal potentials $\bar{V}_i$ are convex and others are concave (see in Section 5.2.4).

2. This procedure turns out to be numerically more robust than the standard algorithm in Chapter 3. Indeed, the choice of arbitrary points $x^* \in \mathcal{I}_0$ and $x^* \in \mathcal{I}_{m_t}$ (intervals that contain the tails of $p_o(x)$) is usually critical at the beginning, when the algorithm starts with few support points, because in some cases the slope of $W_t(x)$ can turn out very sharp and the value of $\exp\{-W_t(x)\}$ very high. This numerical problem does not appear in the alternative S-GARS-2 and S-GARS-4 procedures.

3. It can be used when μ_i is unknown.

We can observe that there is no apparent difference between Figures 6.11(a) and (c), and between Figures 6.11(b) and (d). These results indicate that the procedure S-GARS-3, without the knowledge of the simple estimates, does not cause a significant degradation of the acceptance rates. Clearly, the procedure S-GARS-3 of Section 5.3.2 is computationally more expensive than the standard GARS approach with knowledge of the positions of the simple estimates, but the acceptance rates remain virtually the same.

Figure 6.12 displays together the curves of the acceptance rates for the standard GARS method (solid line) and the S-GARS-4 alternative procedure with unknown μ_1 and unknown the simple estimates (dashed line). The acceptance rates for the first sample are $\approx 25\%$ and $\approx 9\%$ for the standard and the S-GARS-4 procedure, respectively. For the 20-th sample we have acceptance rates $\approx 85\%$ and $\approx 80\%$ while for the 500-th sample they are $\approx 98\%$ and $\approx 93\%$.
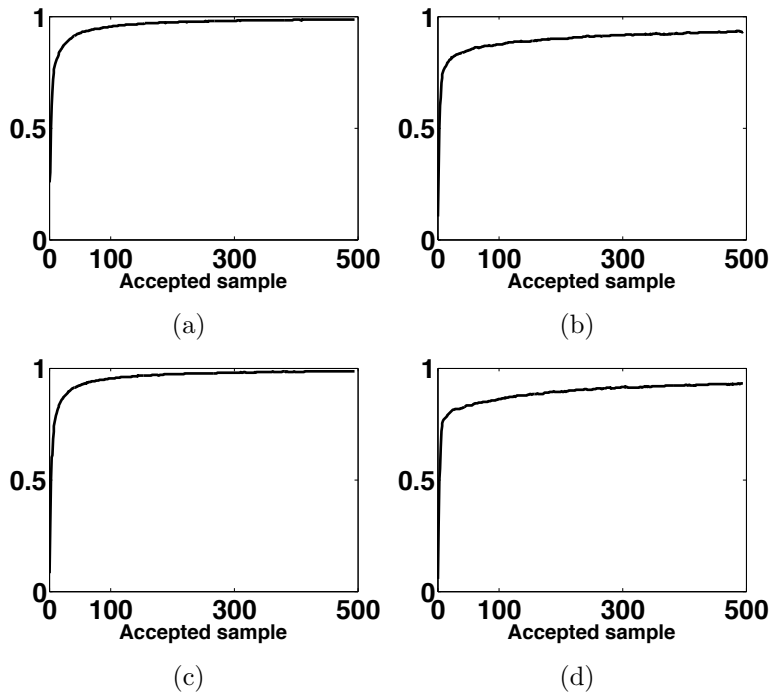
Figure 6.11: Curve of the acceptance rates of: **(a)** the standard GARS method. **(b)** The S-GARS-2 procedure assuming unknown the minimum μ_1 of $\bar{V}_1$. **(c)** The S-GARS-3 procedure assuming unknown the simple estimates. **(d)** The S-GARS-4 procedure assuming unknown both μ_1 and the simple estimates.

6.4   Summary

In this chapter, we have presented a collection of different applications of the proposed GARS techniques. We have illustrated how the new techniques can be used jointly with other Monte Carlo methods, such as the Gibbs sampler and the particle filter.

In most of these examples, the adaptive rejection sampling schemes in the literature (producing independent samples, i.e., except for the ARMS technique of Section 2.7.1) can not be applied. Specifically, the standard ARS of Section 2.6, the concave-convex ARS of Section 2.7.2 and the TDR of Section 2.7.3 cannot be used in any of the examples. The Evans' method [36] (an extension of the TDR method when the inflection points of $p_o(x)$ are available, see Section 2.7.3) can be applied in the examples of Sections
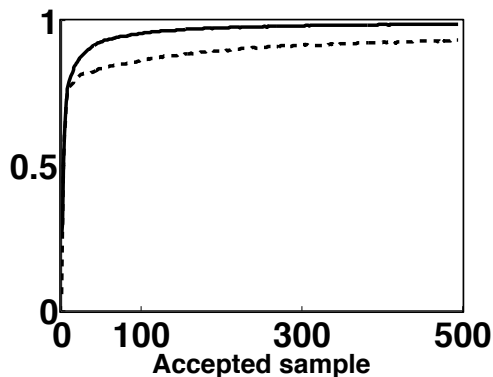
Figure 6.12: Curve of the acceptance rates of the standard GARS method (solid line) and the S-GARS-4 procedure without the knowledge of μ_1 and the impossibility to invert $g_1(x)$ (dashed line).

6.1.1, 6.1.3, 6.2.2 and 6.3. However, in the examples of Sections 6.1.2, 6.1.4, 6.2.1 and 6.2.3 only the proposed GARS techniques can be used in order to generate i.i.d. samples.

In the example of Section 6.1.4 we draw from a multidimensional target pdf, using GARS method jointly with the Gibbs sampling. The Gibbs sampler *needs* to be able to draw from one-dimensional conditional pdf's that, in general, can be very complicated. In this scenario, the GARS algorithms (introduced in the previous chapters) can be applied to draw from these conditional pdf's. However, the generated multidimensional samples are correlated.

We have also shown that it is possible to implement the GARS scheme without the knowledge of the minima μ_i and the sets of simple estimates $\mathcal{X}_i$. This simplified procedure need only to identify and to evaluate components of the target pdf. Although it exhibits some performance degradation when compared with the standard GARS method (with knowledge of both the μ_i's and the $\mathcal{X}_i$'s), this is compensated by the broader range of applicability and its superior numerical stability.

Chapter 7

Summary and future research

7.1 Summary

In this work has been aimed to the design, analysis and assessment of a family of generalized adaptive rejection samplers that can be applied to generate i.i.d. random variates from a broad class of probability distributions. The novel adaptive schemes can be applied to draw exactly from a certain family of pdf's, not necessarily log-concave and possibly multimodal, including pdf's with log-convex tails. The new methodology can be seen as a generalization of the classical adaptive rejection sampling scheme of [50], and includes it as a particular case. All the proposed algorithms are based on the accept/reject principle and the construction of a sequence of proposal pdf's that converge towards the target density and, therefore, can attain very high acceptance rates.

The basics of the methodology, that we have termed generalized adaptive rejection sampling (GARS), have been presented in Chapter 3. There, we have first made explicit the family of target pdf's that can be possibly handled with the new methods. Then, we have introduced the standard form of the GARS algorithm, discussed its applicability and found expressions for the probability of acceptance of the sampler and the divergence between the proposal densities and the target pdf.

One limitation that the basic GARS technique shares with virtually all adaptive rejection samplers in the literature [36, 50, 55, 93] is the impossibility to draw from a target pdf with log-convex tails in an infinite domain. To address this difficulty, in Chapter 4 we have introduced two

GARS schemes that can be used to draw exactly from densities that have not-necessarily log-concave tails. The first adaptive approach in Section 4.2 runs along the same lines as the basic GARS technique and is easier to implement. However, it needs to identify a suitable pdf $q(x) \propto \exp\{-\bar{V}_j(g_j(x))\}$ with $j \in \{1, ..., n\}$, from which samples can be drawn easily and this can be in general a difficult task. One type of pdf's for which the identification of $q(x)$ can turn out natural includes the *a posteriori* density of a random variable X given a collection of observations as those appearing in typical Bayesian inference problems (see Section 3.7.1). In this case, it may often be simple to identify $q(x)$ with the prior of X and $\exp\{-V_{-j}(x; \mathbf{g})\}$ with the likelihood.

The second approach in Chapter 4, the adaptive ratio-of-uniforms (ARoU) algorithm, is a more general scheme that combines techniques typical from transformation-based random sampling with the rejection sampling approach and the adaptation of proposal densities. The range of applicability of ARoU method is larger, and its computational implementation essentially requires being able to draw two independent samples from a uniform distribution. On the other hand, the *design* of the algorithm may be more demanding for the user, as it involves, e.g., the calculation of upper bounds for three potential functions.

Even with the techniques of Chapter 4, there are a number of probability distributions that can not be handled without further modifications of the GARS scheme. Such extensions are explored in Chapter 5. We have investigated how to relax various constrains on the nonlinearities and the marginal potential functions from which the target pdf is constructed. Most of these assumptions actually refer to the convexity and monotonicity of these functions. Then, we have followed a different direction to investigate simplifications of the GARS methodology that ease its software implementation. Interestingly enough, this simplifications also have benefits in term of expanding the class of tractable target pdf's (since less "knowledge" about them is required) and improving the numerical stability of the algorithms.

Several numerical examples have been implemented in Chapter 6, in order to illustrate the performance of the adaptive rejection schemes introduced in this work. Although most of the examples are synthetic, we have also included an application to positioning using real data from a wireless sensor network. In most of the examples only the proposed GARS techniques can be used in order to generate i.i.d. samples. We have also shown that it is possible to implement the GARS scheme without the knowledge of the

minima μ_i and the sets of simple estimates $\mathcal{X}_i$. Although it exhibits some performance degradation when compared with the standard GARS method (with knowledge of both the μ_i's and the $\mathcal{X}_i$'s), this is compensated by the broader range of applicability and its superior numerical stability.

7.2 Future research

The GARS techniques can be applied in a more general framework than the one in this thesis. However, the necessary mathematical formulation can easily become complicated and the resulting algorithms difficult to understand. In our opinion, the future research on the topic can be organized as shown below. We have already studied and analyzed the topics in Sections 7.2.1 and 7.2.2 but we still lack a formulation which is sufficiently concise and appealing.

7.2.1 Different classes of densities

The underlying ideas that support the techniques described in Chapters 3 and 4, can be used to handle the following classes of pdf's with different structure.

- *Composition of several nonlinearities*: The GARS schemes can be extended to draw form target pdf's of the form

$$p_o(x) \propto \exp\{-\bar{V}(g_1(g_2(...g_\tau(x))))\} = \exp\{-(\bar{V} \circ g_1 \circ g_2 \circ .... \circ g_\tau)(x)\},$$
(7.1)

  where $\bar{V}$, g_i, with $i = 1, ..., \tau$ are either convex or concave functions. Actually, we address target densities that can be written as

$$p_o(x) \propto (H \circ g_1 \circ g_2 \circ .... \circ g_\tau)(x)$$
(7.2)

  where $H(\vartheta)$ is an integrable function, i.e., there exists a pdf $f(\vartheta) \propto H(\vartheta)$.

- *Joint potentials*: It is also possible to deal with densities generated by more complicated potentials, that not necessary can be written as a sum of marginal potentials. Indeed, we can handled target pdf's of the form

$$p_o(x) \propto p(x) = \exp\left\{-\hat{V}_n\left(g_1(x), g_2(x), ..., g_n(x)\right)\right\},$$
(7.3)

where the function

$$\hat{V}_n(\vartheta_1, \vartheta_2, ..., \vartheta_n) : \mathbb{R}^n \to \mathbb{R},$$

is termed a *joint potential*, it has a Hessian matrix that is definite positive with a minimum at $\boldsymbol{\mu} = [\mu_1, ..., \mu_n]$ and the nonlinearities $g_i(x)$, $i = 1, ..., n$ are either convex or concave functions. When the joint potential can be expressed as

$$\hat{V}_n(\vartheta_1, \vartheta_2, ..., \vartheta_n) = \sum_{i=1}^{n} \bar{V}_i(\vartheta_i),$$

we come back to the basic framework of Chapter 3 (in this case, the Hessian is a diagonal matrix). Pdf's of Eq. (7.3) appear when we have a collection of measurements contaminated by correlated noise variables and we wish to draw samples from the posterior distribution of X.

7.2.2 Multidimensional random sampling

In this work, we have restricted our study to the case in which we have a scalar random variable, $X \in \mathbb{R}$, that we want to sample. We have seen that a first possibility to draw from multidimensional random variables is to use GARS method jointly with the Gibbs sampler. However, the resulting samples are correlated. Moreover, as shown in Section 6.1.2, the Markov chain produced by an MCMC algorithm can remain trapped at one of the (possibly several) modes of the target density.

Alternatively, we are studying the possibility to extend the approaches described in Chapters 3 and 4 to draw i.i.d. samples directly from a multidimensional pdf $p_o(\mathbf{x})$, with $\mathbf{x} \in \mathbb{R}^m$. The method involves the construction of certain m-dimensional linear manifolds embedded in an n-dimensional metric space $(\mathbb{R}^n, d)$, where d is a suitably defined distance. Unfortunately, a complete description of this extension requires the introduction of a considerable amount of new notations and mathematical tools. Specifically, given a pdf with the form

$$p_o(\mathbf{x}) \propto \exp\left\{-\hat{V}_n\left(g_1(\mathbf{x}), g_2(\mathbf{x}), ..., g_n(\mathbf{x})\right)\right\},$$

where $\mathbf{x} \in \mathbb{R}^m$, the function $\hat{V}_n(\vartheta_1, \vartheta_2, ..., \vartheta_n) : \mathbb{R}^n \to \mathbb{R}$ should have Hessian matrix definite positive with a minimum at $\boldsymbol{\mu} = [\mu_1, ..., \mu_n]$ and the

nonlinearities $g_i(\mathbf{x})$, $i = 1, ..., n$, either convex or concave functions. Consider the manifold

$$\mathcal{G} \triangleq \{[\vartheta_1, ..., \vartheta_n] \in \mathbb{R}^n : \vartheta_1 = g_1(\mathbf{x}), ..., \vartheta_n = g_n(\mathbf{x}) \text{ with } \mathbf{x} \in \mathcal{D} \subseteq \mathbb{R}^m\} \quad (7.4)$$

When $m \leq n$, in general, Eq. (7.4) yields a parametric description of a hypersurface (d-dimensional manifold with $d < m$), embedded in a space $\mathbb{R}^n$, where $\mathbf{x}$ plays the role of the parameter. Therefore, the potential

$$V(\mathbf{x}; \mathbf{g}) \triangleq \hat{V}_n\left(g_1(\mathbf{x}), g_2(\mathbf{x}), ..., g_n(\mathbf{x})\right),$$

is equivalent to $\hat{V}_n(\vartheta_1, ..., \vartheta_n)$ when $\vartheta_i = g_i(\mathbf{x})$, $i = 1, ..., n$. The problem to minimize the potential $V(\mathbf{x}; \mathbf{g})$, i.e.,

$$\min_{\mathbf{x} \in \mathcal{D}} V(\mathbf{x}; \mathbf{g}),$$

is equivalent to minimizing the joint potential

$$\min_{[\vartheta_1, ..., \vartheta_n] \in \mathcal{G}} \hat{V}_n(\vartheta_1, ..., \vartheta_n),$$

restricted to the manifold $\mathcal{G}$. This is an important consideration because, similar to method of Chapter 4, we need to find lower bounds $\gamma_k \leq V(\mathbf{x}; \mathbf{g})$ with $\mathbf{x} \in \mathcal{I}_k$ (where $\mathcal{I}_k$ is a cell in $\mathbb{R}^m$). The rest of the procedure can also related to the 1-dimensional GARS. Again, we wish to modify the potential $\hat{V}_n$ by substituting the nonlinearities g_i by linear functions $r_{i,k}$ properly defined in cells $\mathcal{I}_k \in \mathbb{R}^m$. The difficulty arises to compute an auxiliary manifold

$$\mathcal{R} \triangleq \{[\vartheta_1, ..., \vartheta_n] \in \mathbb{R}^n : \vartheta_1 = r_{1,k}(\mathbf{x}), ..., \vartheta_n = r_{n,k}(\mathbf{x}) \text{ with } \mathbf{x} \in \mathcal{I}_k \subseteq \mathbb{R}^m\}$$

such that

$$\min_{\mathbf{x} \in \mathcal{I}_k} V(\mathbf{x}; \mathbf{r}_k) = \min_{[\vartheta_1, ..., \vartheta_n] \in \mathcal{R}} \hat{V}_n(\vartheta_1, ..., \vartheta_n) \leq \min_{[\vartheta_1, ..., \vartheta_n] \in \mathcal{G}} \hat{V}_n(\vartheta_1, ..., \vartheta_n) = \min_{\mathbf{x} \in \mathcal{I}_k} V(\mathbf{x}; \mathbf{g}).$$

where $V(\mathbf{x}; \mathbf{r}_k) \triangleq \hat{V}_n(r_{1,k}(\mathbf{x}), ..., r_{n,k}(\mathbf{x}))$. The function $V(\mathbf{x}; \mathbf{r}_k)$ is built to have a definite positive Hessian matrix in $\mathcal{I}_k$, hence we can readily obtain

$$\gamma_k \leq V(\mathbf{x}; \mathbf{r}_k) \leq V(\mathbf{x}; \mathbf{g}),$$

for all $\mathbf{x} \in \mathcal{I}_k$.

7.2.3 Rejection control

Rejection sampling is a very useful tool to build random number generators, but for numerical approximations the importance sampling is often preferred. However, importance sampling methods usually produces samples with very small importance weights depending on the choice of the proposal pdf.

For this reason, many different techniques combine rejection and importance sampling [20, 24, 96] as the so-called *weighted rejection sampling* and *rejection control* [17, 21, 99, 97]. Other methods combine rejection sampling with sequential importance sampling [99] as partial rejection control [97], and rejection particle filters [12, 68, 90, 144, 145].

We can use the rejection control approach jointly with the GARS techniques to improve the performance of the estimators recycling the discarded samples. Indeed, one can use the rejected samples and adjust the bias by giving these samples appropriate weights.

In this way, we may obtain faster computation and better efficiency.

7.2.4 Implementation

In order to facilitate the use of the proposed techniques, we may produce Matlab code, C code or both implementing the simplified algorithm in Chapter 5, that can be run basically with the ability to evaluate different components (potentials and nonlinearities) of the target pdf $p_o(x)$.

Appendix A

Inverse-of-density method for a transformed random variable

Let us assume a monotonic pdf $p_o(x)$, so that we can readily define the inverse pdf $p_o^{-1}(y)$. In Chapter 2, Section 2.4.1, we have seen the relationship between a pdf $p_o(x)$ and its inverse pdf $p_o^{-1}(u)$ and how we can use samples from $p_o(x)$ or $p_o^{-1}(u)$ to draw from $p_o^{-1}(u)$ or $p_o(x)$, respectively.

Specifically, we have proved that given x' with pdf $p_o(x)$ and $z' \sim \mathcal{U}([0, 1])$ the sample

$$u' = z' p_o(x'),$$

has density $p_o^{-1}(y)$. While, given y' with pdf $p_o^{-1}(y)$ and $v' \sim \mathcal{U}([0, 1])$, the sample

$$u' = z' p_o^{-1}(y'),$$

is distributed as $p_o(x)$.

Now, we study the relationship between the r.v.'s X, with pdf $p_o(x)$, and Y, with pdf $p_o^{-1}(y)$, when one of them, X or Y, is transformed with a monotonic transformation, $u = h(x)$ or $u = h(y)$, respectively.

A.1   Transforming X: $U = h(X)$

Given a random variable X with pdf $p_o(x)$ and a transformed random variable $U = h(X)$, where h is a monotonic function, we know that the density of U is

$$q(u) = p_o(h^{-1}(u)) \left| \frac{dh^{-1}}{du} \right|. \tag{A.1}$$

Denoting as $\mathcal{A}_h$ the area below $q(u)$, our goal in this appendix is to find the relationship between the pair (U, V) uniformly distributed on $\mathcal{A}_h$ and the r.v. Y with density $p_o^{-1}(y)$. To obtain this relationship, we first investigate the links between the r.v.'s X, U and later Y, U.

Clearly, if we are able to draw a sample u' from $q(u)$, we can easily find a sample x' from $p_o(x)$ as

$$x' = h^{-1}(u').$$ (A.2)

Now consider a random variable Y with pdf $p_o^{-1}(y)$. Given a sample x' from $p_o(x)$, we can obtain a sample y' from $p_o^{-1}(y)$ using the inverse-of-density relationship, i.e,

$$y' = z'p_o(x'),$$ (A.3)

where $z' \sim \mathcal{U}([0, 1])$. Then, replacing x' of Eq. (A.2) into Eq. (A.3), we arrive at

$$y' = z'p_o(h^{-1}(u')).$$ (A.4)

The relationship in Eq. (A.4) connects the r.v. $U = h(X)$ with density $q(u)$ in Eq. (A.1) and the r.v. Y with pdf $p_o^{-1}(y)$, namely

$$Y = Zp_o(h^{-1}(U))$$

with $Z \sim \mathcal{U}([0, 1])$. Figure A.1(a) depicts the area $\mathcal{A}_0$ below $p_o(x)$ and a point (x', y') drawn uniformly from $\mathcal{A}_0$. As explained in Section 2.4.1, x' is distributed as $p_o(x)$ and y' is distributed as $p_o^{-1}(y)$.

Now, we denote as $\mathcal{A}_h$ the area below the density $q(u)$, i.e, the region delimited by the equation $v = q(u)$ and the axis u. Figure A.1(b) shows the pdf $q(u)$, the area $\mathcal{A}_h$ and a point (u', v') drawn uniformly from $\mathcal{A}_h$. To generate the pair of samples (u', v') uniformly in $\mathcal{A}_h$ we can draw u' from $q(u)$ and then draw v' uniformly in the interval $[0, q(u')]$, i.e., $v' \sim \mathcal{U}([0, q(u')])$. Moreover, given $z' \sim \mathcal{U}([0, 1])$, this procedure is equivalent to writing

$$v' = z'q(u').$$ (A.5)

Furthermore, we can substitute $q(u)$ of Eq. (A.1) into Eq. (A.5) to obtain

$$v' = z' \underbrace{p_o(h^{-1}(u'))\left|\frac{dh^{-1}}{du}\right|_{u'}}_{q(u')}.$$ (A.6)

192

Note that from Eq. (A.4) we have $y' = z'p_o(h^{-1}(u'))$, where y' has pdf $p_o^{-1}(y)$ and, since

$$v' = \underbrace{z'p_o(h^{-1}(u'))}_{y'} \left| \frac{dh^{-1}}{du} \right|_{u'}, \tag{A.7}$$

we can write

$$v' = y' \left| \frac{dh^{-1}}{du} \right|_{u'}. \tag{A.8}$$

We recall that the sample

$$y' = \frac{v'}{\left| \frac{dh^{-1}}{du} \right|_{u'}} = v'|\dot{h}(h^{-1}(u'))| \tag{A.9}$$

is distributed as $p_o^{-1}(y)$. We indicate with $\dot{h} = \frac{dh}{dx}$ the first derivative of h.

The Eq. (A.9) asserts that if we are able to draw points (u', v') uniformly in $\mathcal{A}_h$, then we can generate a sample y' from the inverse pdf $p_o^{-1}(y)$. We recall that the region $\mathcal{A}_h$ is the area below $q(u)$, that is the density of a transformed r.v. $U = h(X)$, where X has pdf $p_o(x)$.

Therefore, we can state the following result.

Proposition 3 *Let X be a r.v. with a monotonic pdf $p_o(x)$, and let $U = h(X)$ be another (transformed) r.v., where $h(x)$ is a monotonic transformation. Let us denote with $q(u)$ the density of U and let $\mathcal{A}_h$ be the area below $q(u)$. If we are able to draw a point (u', v') uniformly from the region $\mathcal{A}_h$, then*

$$y' = \frac{v'}{\left| \frac{dh^{-1}}{du} \right|_{u'}} = v'|\dot{h}(h^{-1}(u'))|,$$

is a sample from the inverse pdf $p_o^{-1}(y)$.

Moreover, since $x' = h^{-1}(u')$ is a sample from $p_o(x)$ (if u' is a sample from $q(u)$), we can also write

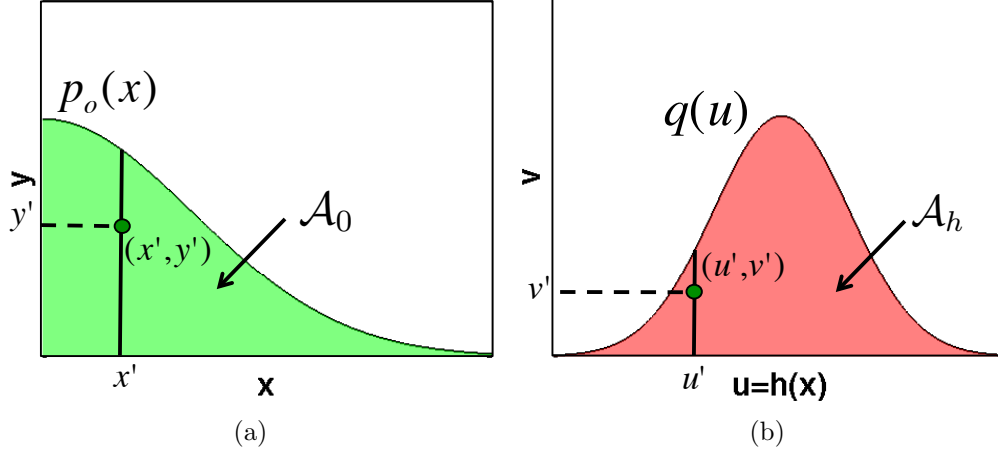$$y' = v'|\dot{h}(x')| = v' \left| \frac{dh}{dx} \right|_{x'}. \tag{A.10}$$

193

Figure A.1: **(a)** Given a point (x', y') uniformly distributed on $\mathcal{A}_0$, x' has pdf $p_o(x)$ while y' is distributed as $p_o^{-1}(y)$, as given by fundamental theorem of simulation and the inverse-of-density-method. **(b)** Given a transformation of the r.v. $U = h(X)$ with pdf $q(u)$, and a point (u', v') uniformly distributed on the area $\mathcal{A}_h$ below $q(u)$, then the sample $y' = v'\dot{h}(h^{-1}(u'))$ has density $p_o^{-1}(y)$.

A.1.1 Special cases

If we consider the identity function $h(x) = x$, we have $U = X$ and $q(u) = p_o(x)$. Since $h^{-1}(y) = y$ as well, the relationship in Eq. (A.9) becomes

$$y' = v'. \tag{A.11}$$

Indeed, in this case the region $\mathcal{A}_h$ is exactly the area $\mathcal{A}_0$ below $p_o(x)$. So that if we are to draw a point $(u' = x', v' = y')$ uniformly from $\mathcal{A}_0 \equiv \mathcal{A}_h$, the fundamental theorem of simulation and the inverse-of-density method, yield that $x' = u'$ has pdf $p_o(x)$ while $y' = v'$ has pdf $p_o^{-1}(y)$.

Moreover, if we choose $h(x) = \sqrt{x}$, $x \geq 0$, hence, since $h^{-1}(u) = u^2$, we have

$$y' = \frac{v'}{2u'}, \tag{A.12}$$

that, except for the constant value $1/2$, coincides with the standard RoU method applied to the inverse pdf $p_o^{-1}(y)$.

194

A.2 Extended inverse-of-density method

Given a random variable Y with pdf $p_o^{-1}(y)$ and a transformed random variable $U = h(Y)$, where h is a monotonic function, we know that the density of U is

$$q(u) = p_o^{-1}(h^{-1}(u)) \left| \frac{dh^{-1}}{du} \right|. \tag{A.13}$$

Denoting as $\mathcal{A}_h$ the area below $q(u)$, our goal is now to find the relationship between the pair (U, V) uniformly distributed on $\mathcal{A}_h$ and the r.v. X with density $p_o(x)$. To obtain this relationship, we first investigate the links between the r.v.'s Y, U and later X, U.

Obviously, if we are able to draw a sample u' from $q(u)$, we can easily compute a sample y' from $p_o^{-1}(y)$ as

$$y' = h^{-1}(u'). \tag{A.14}$$

Therefore, using the inverse-of-density method, we can obtain a sample x' from $p_o(x)$ as

$$x' = z' p_o^{-1}(y') = z' p_o^{-1}(h^{-1}(u')), \tag{A.15}$$

where $z' \sim \mathcal{U}([0,1])$, $u' \sim q(u)$ and $y' = h^{-1}(u')$ from Eq. (A.14). Figure A.2(a) depicts the area $\mathcal{A}_0$ delimited by $p_o(x)$ and a point (y', x') drawn uniformly from $\mathcal{A}_0$. As explained in Section 2.4.1, x' is distributed as $p_o(x)$ and y' is distributed as $p_o^{-1}(y)$.

Moreover, we denote as $\mathcal{A}_h$ the region delimited by the curve $v = q(u)$ and the axis u. Figure A.2(b) illustrates the pdf $q(u)$, the area $\mathcal{A}_h$ and a point (u', v') drawn uniformly from $\mathcal{A}_h$. To draw a point (u', v') uniformly from $\mathcal{A}_h$, we can first draw a sample u' from $q(u)$ and then v' uniformly the interval $[0, q(u')]$, i.e., $v' \sim \mathcal{U}([0, q(u')])$. Therefore, the sample v' can be also expressed as

$$v' = z' q(u'), \tag{A.16}$$

where $z' \sim \mathcal{U}([0,1])$. Substituting $q(u)$ in Eq. (A.13) into Eq. (A.16), we obtain

$$v' = z' \underbrace{p_o^{-1}(h^{-1}(u')) \left| \frac{dh^{-1}}{du} \right|_{u'}}_{q(u')}. \tag{A.17}$$

Furthermore, recalling Eq. (A.15) we can see that

$$v' = \underbrace{z' p_o^{-1}(h^{-1}(u'))}_{x'} \left| \frac{dh^{-1}}{du} \right|_{u'}, \tag{A.18}$$

hence

$$v' = x' \left| \frac{dh^{-1}}{du} \right|_{u'}. \tag{A.19}$$

Then, we can also write

$$x' = \frac{v'}{\left| \frac{dh^{-1}}{du} \right|_{u'}} = v' |\dot{h}(h^{-1}(u'))|, \tag{A.20}$$

that is a sample from $p_o(x)$. We indicate with $\dot{h} = \frac{dh}{dx}$ the first derivative of $h(x)$.

(a)   (b)
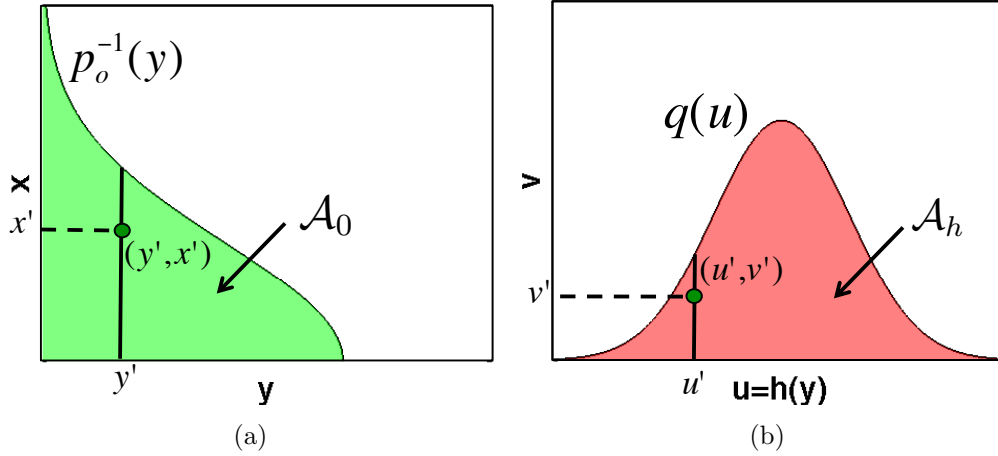
Figure A.2: **(a)** Given a point (x', y') uniformly distributed on $\mathcal{A}_0$, y' has pdf $p_o^{-1}(y)$ while x' is distributed as $p_o(x)$, as affirmed by fundamental theorem of simulation and the inverse-of-density-method. **(b)** Given a transformation of r.v. $U = h(Y)$ with pdf $q(u)$, and a point (u', v') uniformly distributed on the area $\mathcal{A}_h$ below $q(u)$, then the sample $x' = v' \dot{h}(h^{-1}(u'))$ has density $p_o(x)$.

Equation (A.20) can be seen as an extension of the inverse-of-density method when the r.v. Y with inverse pdf $p_o^{-1}(y)$ is transformed by $U = h(Y)$. Therefore, if we are able to draw points (u', v') uniformly from $\mathcal{A}_h$ we can

generate sample x' from the density $p_o(x)$ using Eq. (A.20), as formalized by the following proposition.

Proposition 4 *Let Y be a r.v. with a monotonic pdf $p_o^{-1}(y)$, and let $U = h(Y)$ be another (transformed) r.v., where $h(y)$ is a monotonic transformation. Let us denote with $q(u)$ the density of U and let $\mathcal{A}_h$ be the area below $q(u)$. If we are able to draw a point (u', v') uniformly from the region $\mathcal{A}_h$, then*

$$x' = \frac{v'}{\left|\frac{dh^{-1}}{du}\right|_{u'}} = v' |\dot{h}(h^{-1}(u'))|,$$

is a sample from the pdf $p_o(x)$, that is the inverse function of $p_o^{-1}(y)$.

Since $y' = h^{-1}(u')$ is a sample from $p_o^{-1}(y)$ (if u' is a sample from $q(u)$), we can also write

$$x' = v' |\dot{h}(y')| = v' \left|\frac{dh}{dy}\right|_{y'}. \tag{A.21}$$

A.2.1   Special cases

If we choose $h(y) = y$, we have $U = Y$ and as a consequence $q(u) = p_o^{-1}(y)$ and the region $\mathcal{A}_h$ is exactly $\mathcal{A}_0$, so that Eq. (A.20) becomes

$$x' = v', \tag{A.22}$$

i.e., we come back to the fundamental theorem of simulation and the standard inverse-of-density method. Indeed, if we are to draw a point $(u' = x', v' = y')$ uniformly from $\mathcal{A}_0 \equiv \mathcal{A}_h$, for the fundamental theorem of simulation and the inverse-of-density method, yield that $x' = u'$ has pdf $p_o(x)$ while $y' = v'$ has pdf $p_o^{-1}(y)$.

Moreover, if we take $h(y) = \sqrt{y}$, $y \geq 0$, since $h^{-1}(u) = u^2$, we have

$$x' = \frac{v'}{2u'}, \tag{A.23}$$

that, except for the constant value $1/2$, corresponds to the standard RoU method.

Appendix B

Sampling from triangles

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 [77, 142], [30, p. 570] with the following steps:

1. Sample u_1 from $\mathcal{U}([0,1])$ and u_2 from $\mathcal{U}([0,1])$.

2. The resulting sample is generated by

$$
\begin{aligned}
\mathbf{x}' = \mathbf{v}_1 \min[u_1, u_2] + \mathbf{v}_2 (1 - \max[u_1, u_2]) + \\
+ \mathbf{v}_3 (\max[u_1, u_2] - \min[u_1, u_2]).
\end{aligned}
\tag{B.1}
$$

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 $\mathbf{v}_3$.

Appendix C

Acronyms and abbrevations

- ARMS: Adaptive Rejection Metropolis Sampling.

- ARoU: Adaptive Ratio of Uniforms.

- ARPF: Accept/Reject Particle Filter.

- ARS: Adaptive Rejection Sampling.

- CCARS: Concave Convex Adaptive Rejection Sampling.

- cdf: cumulative distribution function.

- GARS: Generalized Adaptive Rejection Sampling.

- i.i.d.: identically and identical distributed.

- MCMC: Markov Chain Monte Carlo.

- pdf: probability density function.

- PF: particle Filter.

- RoU: Ratio of Uniforms.

- RS: Rejection Sampling.

- r.v.: random variable.

- SIR: sequential importance resampling.

- SoI: signal of interest.

- TDR: Transformed Density Rejections.

- VDR: Vertical Density Representation.

- w.r.t.: with respect to.

Appendix D

Publications

Most of the material related to the standard GARS method has been published in the journal papers [113, 114] with some preliminary results related to the computation of bounds for pdf's in [111, 112]. An initial version, of the first adaptive scheme in Chapter 4 was introduced in [110, 115]. The complete description of the latter method, together with the ARoU algorithm can be found in [116].

Bibliography

[1] J. H. Ahrens. Sampling from general distributions by suboptimal division of domains. *Grazer Math. Berichte*, 319:20, 1993.

[2] J. H. Ahrens. A one-table method for sampling from continuous and discrete distributions. *Computing*, 54(2):127–146, June 1995.

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

[4] K. Alligood, T. Sauer, and J. A. York. *Chaos: An Introduction to Dynamical Systems*. Springer-Verlag, 1997.

[5] C. Andrieu, N. de Freitas, and A. Doucet. Sequential MCMC for Bayesian model selection. In *Proceedings of the IEEE HOS Workshop*, June 1999.

[6] M. J. Appel, R. Labarre, and D. Radulovic. On accelerated random search. *SIAM Journal of Optimization*, 14(3):708–730, 2003.

[7] M. S. Arulumpalam, S. Maskell, N. Gordon, and T. Klapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions Signal Processing*, 50(2):174–188, February 2002.

[8] Gh. Barbu. On computer generation of random variables by transformations of uniform varaibles. *Soc. Sci. Math. R. S. Romania, Tome 26*, 74(2):129–139, 1982.

[9] R. Bellazzi, P. Magni, and G. De Nicolao. Bayesian analysis of blood glucose time series from diabetes home monitoring. *IEEE Transactions on Biomadical Engineering*, 47(7):971–975, July 2000.

[10] C. Berzuini, N. G. Best, W. Gilks, and C. Larizza. Dynamic conditional independence models and Markov chain Monte Carlo methods. *Journal of the American Statistical Association*, 92:1403–1412, 1996.

[11] C. Berzuini and W. Gilks. Resample-move filtering with cross-model jumps. In A. Doucet, N. de Freitas, and N. Gordon, editors, *Sequential Monte Carlo Methods in Practice*, chapter 6. Springer, 2001.

[12] E. Bolviken, P. J. Acklam, N. Christophersen, and J. M. Stordal. Monte Carlo filters for non-linear state estimation. *Automatica*, 37(2):177–183, February 2001.

[13] E. Bolviken and G. Storvik. Deterministic and stochastic particle filters in state-space models. In A. Doucet, N. de Freitas, and N. Gordon, editors, *Sequential Monte Carlo Methods in Practice*, chapter 5, pages 97–116. Springer, 2001.

[14] C. Botts. A modified adaptive accept-reject algorithm for univariate densities with bounded support. *Technical Report, http://williams.edu/Mathematics/cbotts/Research/paper3.pdf*, January 2010.

[15] G. E. P. Box and G. C. Tiao. *Bayesian Inference in Statistical Analysis*. Wiley & sons, 1973.

[16] A. Boyarsky and P. Góora. *Law of Chaos*. Birkhöuser, Boston (USA), 1997.

[17] B. S. Caffo, J. G. Booth, and A. C. Davison. Empirical supremum rejection sampling. *Biometrika*, 89(4):745–754, December 2002.

[18] J. Candy. *Bayesian signal processing: classical, modern and particle filtering methods*. John Wiley & Sons, England, 2009.

[19] O. Cappé, A. Gullin, J. M. Marin, and C. P. Robert. Population Monte Carlo. *Journal of Computational and Graphical Statistics*, 13(4):907–929, 2004.

[20] G. Casella and C. P. Robert. Rao-blackwellisation of sampling schemes. *Biometrika*, 83(1):81–94, 1996.

[21] G. Casella and C. P. Robert. Post-processing accept-reject samples: Recycling and rescaling. *Journal of Computational and Graphical Statistics*, 7(2):139–157, June 1998.

[22] G. Chaitin. On the length of programs for computing finite binary sequences. *Journal of the ACM*, 13:547–569, 1966.

[23] G. Chaitin. On the length of programs for computing finite binary sequences: statistical considerations. *Journal of the ACM*, 16:145–159, 1969.

[24] R. Chen. Another look at rejection sampling through importance sampling. *Statistics & Probability Letters*, 72:277–283, 2005.

[25] R. Chen and J. S. Liu. Mixture Kalman filters. *Journal of the Royal Statistics Society B*, 62:493–508, 2000.

[26] Y. Chung and S. Lee. The generalized ratio-of-uniform method. *Journal of Applied Mathematics and Computing*, 4(2):409–415, August 1997.

[27] J. H. Curtiss. On the distribution of the quotient of two chance variables. *The Annals of Mathematical Statistics*, 12(4):409–421, December 1941.

[28] M. H. DeGroot and M. J. Schervish. *Probability and Statistics, 3rd ed.* Addison-Wesley, New York, 2002.

[29] L. Devroye. Random varlate generation for unlmodal and monotone densltles. *Computing*, 32:43–68, 1984.

[30] L. Devroye. *Non-Uniform Random Variate Generation.* Springer, 1986.

[31] U. Dieter. Mathematical aspects of various methods for sampling from classical distributions. In *Proceedings of Winter Simulation Conference*, 1989.

[32] P. M. Djurić and S. J. Godsill, editors. *Special Issue on Monte Carlo Methods for Statistical Signal Processing.* IEEE Transactions Signal Processing 50 (3), February 2002.

[33] 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.

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

[35] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Physics Letters B*, 195(2):216–222, September 1987.

[36] 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.

[37] P. H. Everson. Exact Bayesian inference for normal hierarchical models. *Journal of Statistical Computation and Simulation*, 68(3):223 – 241, February 2001.

[38] K. T. Fang, Z. H. Yang, and S. Kotz. Generation of multivariate distributions by vertical density representation. *Statistics*, 35(3):281–293, 2001.

[39] P. Fearnhead. *Sequential Monte Carlo Methods in Filter Theory*. Ph. D. Thesis, Merton College, University of Oxford, 1998.

[40] Y. Fong, J. Wakefield, and K. Rice. An efficient markov chain monte carlo method for mixture models by neighborhood pruning. *Journal of Computational and Graphical Statistics (to appear)*, 2011.

[41] D. Frenkel and B. Smit. Understanding molecular simulation: from algorithms to applications. *Academic Press, San Diego*, 1996.

[42] 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.

[43] D. Gamerman. Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing*, 7(1):57–68, 1997.

[44] J. E. Gentle. *Random Number Generation and Monte Carlo Methods*. Springer, 2004.

[45] J. Geweke. Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, 24:1317–1399, 1989.

[46] C. J. Geyer. Markov Chain Monte Carlo maximum likelihood. *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, pages 156–163, 1991.

[47] W. R. Gilks. Derivative-free Adaptive Rejection Sampling for Gibbs Sampling. *Bayesian Statistics*, 4:641–649, 1992.

[48] 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.

[49] W. R. Gilks, N. G. O. Robert, and E. I. George. Adaptive Direction Sampling. *The Statistician*, 43(1):179–189, 1994.

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

[51] W.R. Gilks, S. Richardson, and D. Spiegelhalter. *Markov Chain Monte Carlo in Practice: Interdisciplinary Statistics.* Taylor & Francis, Inc., UK, 1995.

[52] J. Goodman and A. D. Sokal. Multigrid Monte Carlo method for lattice field theories. *Physical Review Letters*, 56(10):1015–1018, March 1986.

[53] N. Gordon, D. Salmond, and A. F. M. Smith. Novel approach to nonlinear and non-Gaussian Bayesian state estimation. *IEE Proceedings-F*, 140:107–113, 1993.

[54] Dilan Görür and Yee Whye Teh. Concave convex adaptive rejection sampling. *University College London, Technical Report*, 2009.

[55] Dilan Görür and Yee Whye Teh. Concave convex adaptive rejection sampling. *Journal of Computational and Graphical Statistics (to appear)*, 2010.

[56] R. C. Griths and S. Tavaré. Monte Carlo inference methods in population genetics. *Mathematical and Computer Modelling*, 23(8-9):141–158, April-May 1996.

[57] A. Hall. On an experimental determination of Pi. *journal Messenger of Mathematics*, 2:113–114, 1873.

[58] J. M. Hammersley and K. W. Morton. Poor man's Monte Carlo. *Journal of the Royal Statistical Society. Series B (Methodological)*, 16(1):23–38, 1954.

[59] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[60] H. Hirose and A.Todoroki. Random number generation for the generalized normal distribution using the modified adaptive rejection method. *International Information Institute*, 8(6):829–836, March 2005.

[61] W. Hörmann. A rejection technique for sampling from T-concave distributions. *ACM Transactions on Mathematical Software*, 21(2):182–193, 1995.

[62] W. Hörmann. A note on the performance of the Ahrens algorithm. *Computing*, 69:83–89, 2002.

[63] 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, Wirtschaftsuniversitat*, 1994.

[64] W. Hörmann, J. Leydold, and G. Derflinger. *Automatic nonuniform random variate generation*. Springer, 2003.

[65] W. Hörmann, J. Leydold, and G. Derflinger. Inverse transformed density rejection for unbounded monotone densities. *Research Report Series/ Department of Statistics and Mathematics (Economy and Business), Vienna University*, 2007.

[66] Y. Huang, J. Zhang, and P. M. Djurić. Bayesian detection for BLAST. *IEEE Transactions on Signal Processing*, 53(3):1086–1096, March 2005.

[67] K. Hukushima and K. Nemoto. Exchange Monte Carlo method and application to spin glass simulations. *Journal of Physical Society of Japan*, 65:1604–1608, 1996.

[68] M. Hürzeler and H. R. Künsch. Monte Carlo approximations for general state-space models. *Journal of Computational and Graphical Statistics*, 7(2):175– 193, June 1998.

[69] 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.

[70] P. Jaeckel. *Monte Carlo Methods in Finance*. Wiley, 2002.

[71] A. Jasra, D. A. Stephens, and C. C. Holmes. Population-based reversible jump Markov chain Monte Carlo. *Biometrika*, 94(4):787– 807, 2007.

[72] L. Jing and P. Vadakkepat. Interacting mcmc particle filter for tracking maneuvering target. *Digital Signal Processing*, 20:561–574, August 2010.

[73] M. C. Jones. On khintchine's theorem and its place in random variate generation. *The American Statistician*, 56(4):304–307, November 2002.

[74] M. C. Jones. Distributions generated by transformation of scale using an extended cauchy-schlömilch transformation. *The Indian Journal of Statistics*, 72-A(2):359–375, 2010.

[75] M. C. Jones and A. D. Lunn. Transformations and random variate generation: generalised ratio-of-uniforms methods. *Journal of Statistical Computation and Simulation*, 55(1):49 – 55, September 1996.

[76] S. Karlin and H. M. Taylor. *A First Course on Stochastic Processes*. Academic Press, 1975.

[77] A. W. Van Kemp. Patchwork rejection algorithms. *Journal of Computational and Applied Mathematics*, 31(1):127–131, July 1990.

[78] A. Y. Khintchine. On unimodal distributions. *Izvestiya NauchnoIssledovatelskogo Instituta Matematiki i Mekhaniki*, 2:1–7, 1938.

[79] 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.

[80] A. J. Kinderman and J. G. Ramage. Computer generation of normal random variables. *Journal of the American Statistical Association*, 71(356):893–898, December 1976.

[81] S. K. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, May 1983.

[82] J. Kohlas. *Monte Carlo simulation in operations research*. Springer-Verlag, Berlin, 1972.

[83] A. N. Kolmogorov. On tables of random numbers. *Sankhya, The Indian Journal of Statistics, serie A*, 25:369–376, 1963.

[84] A. N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems of Information Transmission*, 1(1):1–7, 1965.

[85] J. Kotecha and Petar M. Djurić. Gibbs sampling approach for generation of truncated multivariate gaussian random variables. *Proceedings of Acoustics, Speech, and Signal Processing, (ICASSP)*, 1999.

[86] J. Kotecha and Petar M. Djurić. Gaussian sum particle filtering. *IEEE Transactions Signal Processing*, 51(10):2602–2612, October 2003.

[87] S. Kotz and M. D. Troutt. On vertical density representation and ordering of distributions. *Statistics*, 28:241–247, 1996.

[88] R. A. Kronmal and A.V. Peterson. A variant of the acceptance-rejection method for computer generation of random variables. *Journal of the American Statistical Association*, 76(374):446–451, June 1981.

[89] R. A. Kronmal and A.V. Peterson. An acceptance-complement analogue of the mixture-plus-acceptance-rejection method for generating random variables. *ACM Transactions on Mathematical Software*, 10(3):271–281, September 1984.

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

[91] P. K. Kythe and M. R. Schaferkotter. *Handbook of Computational Methods for Integration*. Chapman and Hall/CRC, 2004.

[92] J. Leydold. Automatic sampling with the ratio-of-uniforms method. *ACM Transactions on Mathematical Software*, 26(1):78–98, 2000.

[93] J. Leydold. Short universal generators via generalized ratio-of-uniforms method. *Mathematics of Computation*, 72:1453–1471, 2003.

[94] J. Leydold, J. Janka, and W. Hörmann. Variants of transformed density rejection and correlation induction. *In Monte Carlo and Quasi-Monte Carlo methods 2000*, Springer-Verlag:345–356, 2002.

[95] F. Liang, C. Liu, and R. Caroll. *Advanced Markov Chain Monte Carlo Methods: Learning from Past Samples*. Wiley Series in Computational Statistics, England, 2010.

[96] J. S. Liu. Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Statistics and Computing*, 6(2):113–119, 1996.

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

[98] J. S. Liu and R. Chen. Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93(443):1032–1044, September 1998.

[99] J. S. Liu, R. Chen, and W. H. Wong. Rejection control and sequential importance sampling. *Journal of the American Statistical Association*, 93(443):1022–1031, September 1998.

[100] J. S. Liu, F. Liang, and W. H. Wong. The multiple-try method and local optimization in metropolis sampling. *Journal of the American Statistical Association*, 95(449):121–134, March 2000.

[101] M. Ljungberg, S. E. Strand, and M. A. King. *Monte Carlo Calculations in Nuclear Medicine*. Taylor & Francis, 1998.

[102] M. Locatelli. Convergence of a simulated annealing algorithm for continuous global optimization. *Journal of Global Optimization*, 18:219–234, 2000.

[103] E. Marinari and G. Parisi. Simulated tempering: a new Monte Carlo scheme. *Europhysics Letters*, 19(6):451–458, July 1992.

[104] G. Marrelec and H. Benali. Automated rejection sampling from product of distributions. *Computational Statistics*, 19(2):301–315, May 2004.

[105] G. Marsaglia. Ratios of normal variables and ratios of sums of uniform variables. *American Statistical Association*, 60(309):193–204, March 1965.

[106] G. Marsaglia. The exact-approximation method for generating random variables in a computer. *American Statistical Association*, 79(385):218–221, March 1984.

[107] G. Marsaglia and W. W. Tsang. The ziggurat method for generating random variables. *Journal of Statistical Software*, 8(5), 2000.

[108] A. Marshall. The use of multistage sampling schemes in Monte Carlo computations. *Symposium on Monte Carlo, Wiley New York*, pages 123–140, 1956.

[109] P. Martin-Iöf. Complexity of oscillations in infinite binary sequences. *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 19:225–230, 1971.

[110] L. Martino and J. Míguez. An adaptive accept/reject sampling algorithm for posterior probability distributions. *Proc. of the IEEE SSP*, August 2009.

[111] L. Martino and J. Míguez. New accept/reject methods for independent sampling from posterior probability distributions. *European Signal Processing Conference (EUSIPCO), Glasgow (Scotland)*, August 2009.

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

[113] L. Martino and J. Míguez. A generalization of the adaptive rejection sampling algorithm. *Statistics and Computing (to appear) DOI: 10.1007/s11222-010-9197-9*, July 2010.

[114] L. Martino and J. Míguez. Generalized rejection sampling schemes and applications in signal processing. *Signal Processing*, 90(11):2981–2995, November 2010.

[115] L. Martino and J. Míguez. A rejection sampling scheme for posterior probability distributions via the ratio-of-uniforms method. *European Signal Processing Conference (EUSIPCO), Aalborg (Denmark)*, August 2010.

[116] L. Martino and J. Míguez. Two adaptive rejection sampling methods for posterior probability distributions. *Statistics and Computing (submitted)*, December 2010.

[117] P. Mayo, F. Rodenas, and G. Verdú. Comparing methods to denoise mammographic images. In *Proceedings of the 26th IEEE EMBS*, volume 1, pages 247 – 250, September 2004.

[118] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1091, 1953.

[119] N. Metropolis and S. Ulam. The Monte Carlo method. *Journal of the American Statistical Association*, 44:335–341, 1949.

[120] R. Meyer, B. Cai, and F. Perron. Adaptive rejection Metropolis sampling using Lagrange interpolation polynomials of degree 2. *Computational Statistics and Data Analysis*, 52(7):3408–3423, March 2008.

[121] J.R. Michael, W.R. Schucany, and R.W. Haas. Generating random variates using transformations with multiple roots. *The American Statistician*, 30(2):88–90, May 1976.

[122] J. Monahan. An algorithm for generating chi random variables. *Transactions on Mathematical Software*, 13(2):168–172, June 1987.

[123] P. Del Moral. *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer, 2004.

[124] H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*. Society for Industrial Mathematics, 1992.

[125] W. K. Pang, Z.H. Yang, S. H. Hou, and P.K. Leung. Non-uniform random variate generation by the vertical strip method. *European Journal of Operational Research*, 142:595–609, 2002.

215

[126] N. Patwari, A. O. Hero, M. Perkins, N. S. Correal, and R. J. O'Dea. Relative location estimation in wireless sensor networks. *IEEE Transactions Signal Processing*, 51(5):2137–2148, 2003.

[127] C. J. Perez, J. Martín, C. Rojano, and F. J. Girón. Efficient generation of random vectors by using the ratio-uniforms method with ellipsoidal envelopes. *Statistics and Computing*, 18(4):209–217, January 2008.

[128] M. M. Pieri, H. Martel, and C. Grenón. Anisotropic galactic outflows and enrichment of the intergalactic Medium. I. Monte Carlo simulations. *The Astrophysical Journal*, 658(1):36–51, March 2007.

[129] S. B. Pope. A Monte Carlo method for the PDF equations of turbolent reactive flow. *Combustion Science and Technology*, 25:159–174, 1981.

[130] T. S. Rappaport. *Wireless Communications: Principles and Practice (2nd edition)*. Prentice-Hall, Upper Saddle River, NJ (USA), 2001.

[131] R. Reiss and M. Thomas. *Statistical analysis of extreme values: with applications to insurance, finance, hydrology and other fields*. Springer, 2007.

[132] D. Remondo, R. Srinivasan, V. F. Nicola, W. C. van Etten, and H. E. P. Tattje. Adaptive importance sampling for performance evaluation and parameter optimization of communication systems. *IEEE Transactions on communications*, 48(4):557–565, l 2000.

[133] B. Ristic, S. Arulampalam, and N. Gordon. *Beyond the Kalman Filter*. Artech House, Boston, 2004.

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

[135] M. Rosenbluth and A. Rosenbluth. Monte Carlo calculation of average extension of molecular chains. *Journal of Chemical Physics*, 23:356–359, 1955.

[136] D. B. Rubin. A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: the sir algorithm. *Journal of the American Statistical Association*, 82:543–546, 1987.

[137] J. I. Siepmann. A method for the direct calculation of chemical potentials for dense chain systems. *Molecular Physics*, 70(6):1145–1158, 1990.

[138] J. I. Siepmann and D. Frenkel. Configurational bias Monte Carlo: a new sampling scheme for flexible chains. *Molecular Physics*, 75(1):59–70, 1992.

[139] T. Siiskonen and R. Pollanen. Alpha-electron and alpha-photon coincidences in high-resolution alpha spectrometry. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 558(2):437–440, March 2006.

[140] E. Stadlober and H. Zechner. The patchwork rejection technique for sampling from unimodal distributions. *ACM Transactions on Modeling and Computer Simulation*, 9(1):59–80, 1999.

[141] S. Stefanescu and I. Vaduva. On computer generation of random vectors by transformations of uniformly distributed vectors. *Computing*, 39:141–153, September 1987.

[142] W. E. Stein and M. F. Keblis. A new method to simulate the triangular distribution. *Mathematical and Computer Modeling*, 49(5):1143–1147, 2009.

[143] R. H. Swendsen and J.S. Wang. Replica Monte Carlo simulation of spin glasses. *Physical Review Letters*, 57(21):2607–2609, November 1986.

[144] H. Tanizaki. On the nonlinear and non-normal filter using rejection sampling. *IEEE Transaction on automatic control*, 44(3):314–319, February 1999.

[145] H. Tanizaki. Nonlinear and non-Gaussian state space modeling using sampling techniques. *Annals of the Institute of Statistical Mathemathics*, 53(1):63–81, 2001.

[146] L. Tierney. Exploring posterior distributions using Markov Chains. In *Computer Science and Statistics: Proceedings of IEEE 23rd Symp. Interface*, pages 563–570, 1991.

[147] M. D. Troutt. A theorem on the density of the density ordinate and an alternative interpretation of the box-muller method. *Statistics*, 22:463–466, 1991.

[148] M. D. Troutt. Vertical density representation and a further remark on the box-muller method. *Statistics*, 24:81–83, 1993.

[149] M. D. Troutt, W. K. Pang, and S. H. Hou. *Vertical density representation and its applications*. World Scientific, 2004.

[150] I. Vaduva. Computer generation of random vectors based on transformations on uniform distributed vectors. In *Proceedings of Sixth Conf. on Probability Theory, Brasov*, pages 589–598, September 1982.

[151] J. P. Valleau. Density-scaling: A new Monte Carlo technique in statistical mechanics. *Journal of Computational Physics*, 96(1):193–216, September 1991.

[152] John von Neumann. Various techniques in connection with random digits. *National Bureau of Standard Applied Mathematics Series*, 12:36–38, 1951.

[153] 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.

[154] C.S. Wallace. Transformed rejection generators for gamma and normal pseudo-random variables. *Australian Computer Journal*, 8:103–105, 1976.

[155] X. Wang. Improving the rejection sampling method in quasi-Monte Carlo methods. *Journal of Computational and Applied Mathematics*, 114(2):231–246, Febrary 2000.

[156] E. M. Wijsman. Monte Carlo Markov chain methods and model selection in genetic epidemiology. *Computational Statistics & Data Analysis*, 32(3-4):349–360, January 2000.

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

[158] S. R. Williams and D. J. Evans. Nonequilibrium Dynamics and Umbrella Sampling. *Physical Review Letters*, 105(11):1–26, September 2010.

[159] J. Yu, Z. Yang, and X. Zhang. A class of nonlinear stochastic volatility models and its implications for pricing currency options. *Computational Statistics and Data Analysis*, 51(4):2218–2231, December 2006.

[160] P. Zanetti. New Monte Carlo scheme for simulating Lagranian particle diffusion with wind shear effects. *Applied Mathematical Modelling*, 8(3):188–192, June 1984.