The Recycling Gibbs sampler for efficient and fast learning

L. Martino[†],V. Elvira[⋆], G. Camps-Valls[†]

[†] Image Processing Laboratory (IPL), Universitat de Valencia (Valencia), Spain.

[⋆] Department of Signal Processing, Universidad Carlos III de Madrid (Leganes), Spain.

Abstract

Monte Carlo methods are essential tools for Bayesian inference. Gibbs sampling is a well-known Markov chain Monte Carlo (MCMC) algorithm, extensively used in statistical signal processing, machine learning and statistics, employed to draw samples from complicated high-dimensional posterior distributions. The key point for the successful application of the Gibbs sampler is the ability to draw efficiently from the full-conditional pdfs. In the general case, this is not possible and it requires the generation of auxiliary samples that are wasted, since they are not used in the final estimators. In this work, we show that these auxiliary samples can be employed within the Gibbs estimators, improving their efficiency with no extra cost. This novel scheme arises naturally after pointing out the relationship between the Gibbs sampler and the chain rule used for sampling purpose. Numerical simulations confirm the excellent performance of the novel scheme.

1 Introduction

Monte Carlo algorithms [27, 42] have become very popular over the last decades. Many practical problems in statistical signal processing, machine learning and statistics, demand procedures for drawing from probability distributions with non-standard forms. One of most popular Monte Carlo methods are the Markov chain Monte Carlo (MCMC) algorithms [22, 8] and the particle filters [2, 6, 38]. The MCMC techniques generate a Markov chain (i.e., a sequence of samples) with a target probability density function (pdf) as invariant density [27, 26].

The Gibbs sampling technique is a well-known MCMC algorithm, extensively used in literature in order to generate samples from multivariate target densities, drawing each component of the samples from the full-conditional densities [23, 25, 44, 40, 42].[1] In order to draw from a multivariate target distribution, the key point for the successful application of the standard Gibbs sampler is the ability to draw efficiently from the univariate conditional pdfs [27, 42, 23]. The best scenario for Gibbs sampling occurs when specific direct samplers for each full-conditional are available (e.g., inversion method or more in general some transformations of a random variable [5, 42]). Otherwise, another Monte Carlo technique, like rejection sampling (RS) or an MH-type algorithm, is typically used *within* the Gibbs sampler to draw from the complicated full-conditionals. The performance of the resulting Gibbs sampler depends on the employed *internal* technique as pointed out, for instance, in [4, 12, 35, 36]).

With this purpose, some authors have suggested to use more steps of the MH method within the Gibbs sampler [39, 10, 9]. Moreover, other different algorithms has been proposed as alternative to the MH technique [4, 23, 30, 45]. For instance, several automatic and self-tuning samplers have been designed to be used primarly *within-Gibbs*: the adaptive rejection sampling (ARS) [11, 14], the griddy Gibbs sampler [41], the FUSS sampler [36], Adaptive Rejection Metropolis Sampling (ARMS) [12, 13, 37, 47], Independent Doubly Adaptive Rejection Metropolis Sampling (IA^{2}RMS) [35], etc. Most of these solutions require to perform several steps of MCMC steps for each full-conditional (in order to improve the performance, for instance allowing the adaptation of the parameters of the method) although only one of them is considered to produce the resulting Markov chain (the rest of samples plays

[1] In general these full-conditionals are univariate. However, blockwise Gibbs sampling approaches, where several random variables are updated simultaneously, have been proposed to speed up the convergence of the Gibbs sampler [43]. However, unless direct sampling from the multi-variate full-conditionals is feasible, these approaches result in an increased difficulty of drawing samples and a higher computational cost per iteration. Furthermore, the performance of the overall algorithm can decrease if the blocks are not properly chosen, especially when direct sampling from the multivariate full-conditionals is unfeasible [27, 26, 42]. The novel recycling scheme can also be used in the block approach.

the role of auxiliary variables). Namely, they require an increase in the computational cost that is not completely exploited: several samples are drawn from the full-conditionals, but only a subset of these generated samples is employed in the final estimators.

In this work, we show that also the rest of generated samples can be incorporated within the corresponding Gibbs estimator. We call this approach as *recycling Gibbs sampling* since the samples drawn from each full-conditional can be used also in order to provide a better estimation instead of wasting them. The novel scheme allows to obtain better performance without adding any extra computational cost.

The rest of the paper is organized as follows. Section 2 recall the problem statement and the Bayesian inference approach. The chain rule for sampling purpose is recalled in Section 4. In the same section, we introduce an alternative chain rule approach and highlight some connections with the Gibbs sampler. The novel technique is described in Section 5. Sections 6 provides several numerical results. Finally, Section 7 contains some conclusions.

2 Bayesian inference

In many applications, the goal is to infer a variable of interest, $\mathbf{x} = [x_1, \ldots, x_D] \in \mathcal{D} \subseteq \mathbb{R}^D$, given a set of observations or measurements, $\mathbf{y} \in \mathbb{R}^P$. In Bayesian inference all the statistical information is summarized by means of the posterior pdf, i.e.,

$$\bar{\pi}(\mathbf{x}) = p(\mathbf{x}|\mathbf{y}) = \frac{\ell(\mathbf{y}|\mathbf{x})g(\mathbf{x})}{Z(\mathbf{y})}, \tag{1}$$

where $\ell(\mathbf{y}|\mathbf{x})$ is the likelihood function, $g(\mathbf{x})$ is the prior pdf and $Z(\mathbf{y})$ is the Bayesian evidence (a.k.a., marginal likelihood). In general, $Z(\mathbf{y})$ is unknown, so we assume to be able to evaluate the unnormalized target function,

$$\pi(\mathbf{x}) = \ell(\mathbf{y}|\mathbf{x})g(\mathbf{x}). \tag{2}$$

The analytical study of the posterior density $\bar{\pi}(\mathbf{x}) \propto \pi(\mathbf{x})$ is often unfeasible. For instance, integrals involving $\bar{\pi}(\mathbf{x})$ are typically intractable), and numerical approximations are required, e.g.,

$$I = \int_{\mathcal{D}} f(\mathbf{x})\bar{\pi}(\mathbf{x})d\mathbf{x}, \tag{3}$$

where $f(\mathbf{x})$ is a square integrable function. Our goal is to approximate this integral by a Monte Carlo quadrature scheme. Namely, considering T independent samples from the posterior target pdf, $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(T)} \sim \bar{\pi}(\mathbf{x})$, then we can write

$$I \approx \widehat{I}_T = \frac{1}{T}\sum_{i=1}^{T} f(\mathbf{x}^{(t)}). \tag{4}$$

For the (weak) law of large numbers this estimator converges in probability to I ($\widehat{I}_T \xrightarrow{p} I$), i.e., $\lim_{T\to\infty} \Pr\left(|\widehat{I}_T - I| > \epsilon\right) = 0$. In general, a direct method for drawing independent samples from $\bar{\pi}(\mathbf{x})$ is not available and alternative approaches (e.g., MCMC algorithms) are needed. An MCMC method generate a Markov chain with invariant density $\bar{\pi}(\mathbf{x})$. Although, the generate samples, $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(T)}$, are then correlated in this case, $\widehat{I}_T = \frac{1}{T}\sum_{i=1}^{T} f(\mathbf{x}^{(t)})$ is still a valid consistent estimator.

Within the MCMC family of methods, we can consider a block approach, working directly in the space $\mathbb{R}^D$, e.g., using a Metropolis-Hastings (MH) algorithm [26, 27]), or a component-wise approach working iteratively in different unidimensional slices of the entire space, e.g., using a Gibbs sampler [27, 42].[2] In many applications, for different reasons, the component-wise approach is preferable. For instance, this is the case when the full-conditional distributions are directly provided or since the probability of accepting a new state with a complete block approach becomes negligible, as the dimension of the problem D grows. Below, we outline the standard Gibbs approach.

[2]There also exists intermediate strategies where same subset of variables are jointly updated (this is often called Blocked Gibbs sampler).

3 Standard Gibbs (SG) sampling

Let us denote as

$$\bar{\pi}_d(x_d|x_1,\ldots,x_{d-1},x_{d+1},\ldots,x_D) = \bar{\pi}_d(x_d|x_{1:d-1},x_{d+1:D}), \quad (5)$$
$$= \bar{\pi}_d(x_d|\mathbf{x}_{-d}), \quad (6)$$

the unidimensional full-conditional pdf of the component $x_d \in \mathbb{R}$, $d \in \{1,\ldots,D\}$, given the rest of variables ($\mathbf{x}_{-d} = [x_1,\ldots,x_{d-1},x_{d+1},\ldots,x_D]$). The Gibbs algorithm for generate a sequence of T samples is formed by the following steps:

For $t = 1,\ldots,T$:

1. For $d = 1,\ldots,D$:
   (a) Draw $x_d^{(t)} \sim \bar{\pi}_d(x_d|x_{1:d-1}^{(t)}, x_{d+1,D}^{(t-1)})$.
2. Set $\mathbf{x}^{(t)} = [x_1^{(t)}, x_2^{(t)},\ldots,x_D^{(t)}]$.

Note that the key point for the application of Gibbs sampling is being able to draw efficiently from these univariate full-conditional pdfs $\bar{\pi}_d$. However, in general, we are not able to draw directly from some full-conditional pdfs. Thus, another Monte Carlo technique is needed from drawing from all the $\bar{\pi}_d$ or some of them.

3.1 MCMC-within-SG schemes

In many cases, drawing directly from the full-conditional pdf is not possible. In some specific situations, rejection samplers [3, 19, 21, 29, 46] and their adaptive versions, *adaptive rejection samplers* (ARS), are employed to generate (one) sample from $\bar{\pi}_d$ [7, 14, 12, 15, 20, 17, 18, 31, 32, 33]. ARS are very appealing techniques since they construct a non-parametric proposal in order to mimic the shape of the target pdf, yielding in general excellent performance (i.e., independent samples from $\bar{\pi}_d$ with an high acceptance rate). However, their range of application is limited to some specific classes of densities [14, 33].

More generally, it is impossible to draw from a full-conditional pdf $\bar{\pi}_d$ (neither a rejection sampler can be applied), an additional MCMC sampler is required in order to draw from $\bar{\pi}_d$ [10]. Thus, in many practical scenarios, we have an MCMC (e.g., an MH sampler) inside another MCMC scheme (i.e., the Gibbs sampler). In the so-called *MH-within-Gibbs* approach, only one MH step is often performed within each Gibbs iteration, in order to drawn from each complicated full-conditionals. This hybrid approach preserves the ergodicity of the Gibbs sampler, and provide good performance in many cases. On the other hand, several authors have noticed that using a single MH step for the internal MCMC is not always the best solution in terms of performance (cf. [1]). Other approximated approaches has been also proposed, considering the application of the importance sampling within the Gibbs sampler [24].

Using of a larger number of iterations for the MH algorithm, there is more probability of avoiding the burn-in period so that the last sample be distributed as the full-conditional [39, 10, 9]. Thus, this case is closer to the ideal situation, i.e., sampling directly from the full-conditional pdf. However, unless the proposal is very well tailored to the target, a properly designed adaptive MCMC algorithm should provide less correlated samples than a standard MH algorithm. Several more sophisticated (adaptive or not) MH schemes for the application "within-Gibbs" have been proposed in literature [4, 12, 35, 36, 28, 37, 45, 47]. In general, these techniques employ a non-parametric proposal pdf in the same fashion of the ARS schemes. It is important to remark that performing more steps of a standard or adaptive MH within a Gibbs sampler can provide better performance than performing a longer Gibbs chain applying only one MH step (see, e.g., [12, 34, 35]). Table 1 describe a generic MCMC-within-Gibbs samplers considering to perform M iterations of the *internal* MCMC, at each Gibbs step.

Although, these algorithms are specifically designed to be applied "within-Gibbs" and provide very good performance, they require an increase in the computational cost that is not completely exploited: several samples are drawn from the full-conditionals and used (in some of these techniques, e.g., [12, 37, 35]) in order to adapt the proposal pdf, but only a subset of these generated samples is employed within the final Gibbs estimator.

Table 1: **Generic MCMC-within-SG sampler.**

Choose $\mathbf{x}^{(0)} = [x_1^{(0)}, \ldots, x_D^{(0)}]$.

1. For $t = 1, \ldots, T$:
   (a) For $d = 1, \ldots, D$:
      i. Perform M steps of an MCMC algorithm (like MH or a more advanced method) with target pdf $\bar{\pi}_d(x_d|x_{1:d-1}^{(t)}, x_{d+1,D}^{(t-1)})$, yielding the sequence of samples $v_{d,1}^{(t)}, \ldots, v_{d,M}^{(t)}$ (with starting point $v_{d,0}^{(t)} = x_d^{(t-1)}$).
      ii. Set $x_d^{(t)} = v_{d,M}^{(t)}$.
   (b) Set $\mathbf{x}^{(t)} = [x_1^{(t)}, x_2^{(t)}, \ldots, x_D^{(t)}]$.
2. Return $\{\mathbf{x}^{(t)}\}$ for $t = 1, \ldots, T$.

In this work, we show that also the rest of generated samples can be incorporated within the corresponding Gibbs estimator.

4 Chain rule and its connection with Gibbs sampling

For the sake of simplicity, let us consider a bivariate target pdf that can be factorized according to the chain rule, $\bar{\pi}(x_1, x_2) = \bar{\pi}_2(x_2|x_1)p_1(x_1)$ where we have denoted with p_1 the marginal pdf of x_1. Clearly, if we are able to draw from the marginal pdf $p_1(x_1)$ and from the conditional pdf $\bar{\pi}_2(x_2|x_1)$, we can draw samples from $\bar{\pi}((x_1, x_2)$ following the procedure below:

For $n = 1, \ldots, N$

1. Draw $x_1^{(n)} \sim p_1(x_1)$.
2. Draw $x_2^{(n)} \sim \bar{\pi}_2(x_2|x_1^{(n)})$.

The N independent random vectors $[x_1^{(n)}, x_2^{(n)}]$, with $n = 1, \ldots, N$, are all distributed as $\bar{\pi}(x_1, x_2)$.

4.1 Asymptotically equivalent formulation of a Gibbs sampler

Let us consider again the previous bivariate case where the target pdf is factorized as $\bar{\pi}(\mathbf{x}) = \bar{\pi}(x_1, x_2)$. Hence, the standard Gibbs sampler consists of the following steps:

For $t = 1, \ldots, T$:

1. Draw $x_2^{(t)} \sim \bar{\pi}_1(x_2|x_1^{(t-1)})$.
2. Draw $x_1^{(t)} \sim \bar{\pi}_2(x_1|x_2^{(t)})$.
3. Set $\mathbf{x}^{(t)} = [x_1^{(t)}, x_2^{(t)}]$.

After a certain number of iterations (the so-called "burn in" period [27, 42]), i.e., $t \geq t_b$, the chain converges so that $\mathbf{x}^{(t)} \sim \bar{\pi}(\mathbf{x})$. Moreover, each component of the vector $\mathbf{x}^{(t)} = [x_1^{(t)}, x_2^{(t)}]$ is distributed as the corresponding marginal pdf, i.e., $x_1^{(t)} \sim p_1(x_1)$ and $x_2^{(t)} \sim p_2(x_2)$ for $t \geq t_b$, for $t \geq t_b$ iterations. Therefore a procedure, asymptotically

equivalent for $t \geq t_b$ to the Gibbs sampler above, is:

For $t \geq t_b$, repeat:

1. Generate $x_1^{(t)} \sim p_1(x_1)$, and then draw $x_2^{(t)} \sim \bar{\pi}_1(x_2|x_1^{(t)})$.
2. Set $\mathbf{x}^{(t)} = [x_1^{(t)}, x_2^{(t)}]$.

Namely, after t_b iterations, the standard Gibbs sampler can be interpreted as the application of the chain rule procedure previously described.

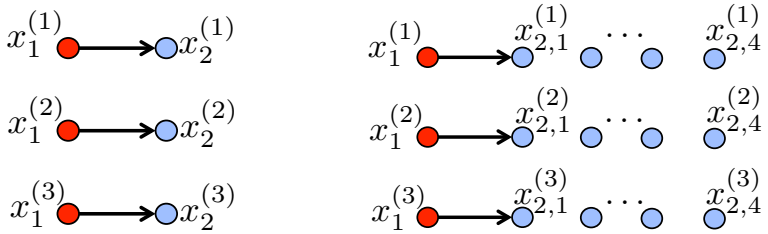
4.2 Alternative chain rule procedure

The following alternative procedure, drawing M samples from the full conditional $\pi_2(x_2|x_1)$ at each n-th iteration, is also valid:

For $n = 1, \ldots, N$

1. Draw $x_1^{(n)} \sim p_1(x_1)$.
2. Draw M samples $x_{2,m}^{(n)} \sim \bar{\pi}_2(x_2|x_1^{(n)})$, with $m = 1, \ldots .M$.

All the NM vectors, $[x_1^{(n)}, x_{2,m}^{(n)}]$, with $n = 1, \ldots, N$ and $m = 1, \ldots, M$, are samples from $\bar{\pi}(x_1, x_2)$. This scheme is valid and, in some cases, can present some benefit w.r.t. the traditional scheme in terms of performance, depending on the variances of the marginal pdfs $p_1(x_1)$ and $p_2(x_2)$. Figure 1 depicts the graphical representation of the standard chain rule sampling scheme (with $N = 3$, $M = 1$) and the alternative chain rule sampling procedure (with $N = 3$, $M = 4$), described above.



(a) Standard sampling. (b) Alternative sampling.

Figure 1: Graphical representation of the **(a)** standard chain rule sampling ($M = 1$) and **(b)** the alternative chain rule sampling ($M = 4$). In both cases, $N = 3$. The total number of drawn vectors $[x_1^{(n)}, x_{2,m}^{(n)}] \sim \bar{\pi}(x_1, x_2) = \bar{\pi}_2(x_2|x_1)p_1(x_1)$ is $NM = 3$ and $NM = 12$, respectively.

A question arises naturally: is it possible to design a Gibbs sampling scheme equivalent to the alternative chain rule scheme described above? In the next section, we introduce the Recycling Gibbs scheme which corresponds to the alternative chain rule procedure as summarized graphically in Figure 2.

5 The recycling Gibbs sampler

The previous considerations suggest that we can recycle some intermediate points, produced in the Gibbs procedure. More specifically, let us consider the following *recycling Gibbs* (RG) procedure

For $t = 1, \ldots, T$:

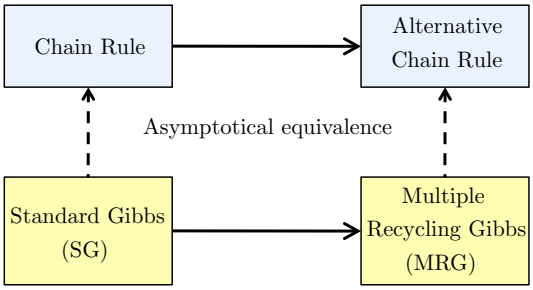


Figure 2: Graphical representation of the relationships between chain rules and Gibbs schemes.

1. For $d = 1, \ldots, D$:
   (a) Draw $x_d^{(t)} \sim \bar{\pi}_d(x_d | x_1^{(t)}, \ldots, x_{d-1}^{(t)}, x_{d+1}^{(t-1)}, \ldots, x_D^{(t-1)})$.
   (b) Set $\mathbf{x}_d^{(t)} = [x_1^{(t)}, \ldots, x_d^{(t)}, x_{d+1}^{(t-1)} \ldots, x_D^{(t-1)}]$.

The procedure above generate DT samples $\mathbf{x}_d^{(t)}$, with $d = 1, \ldots, D$ and $t = 1, \ldots, T$, shown in Figure 3(b) with circles and squares. Note that if we consider only the subset of generated vectors

$$\mathbf{x}_D^{(t)}, \quad t = 1, \ldots, T,$$

setting $d = D$, we obtain the outputs of the standard Gibbs (SG) approach. Namely, the samples generated by a SG procedure can be obtained by a subsampling of the samples obtained by RG. Figures 3(a) depicts with circles $T + 1$ vectors (considering also the starting point) corresponding to a run of SG with $T = 4$. Figures 3(b) shows with squares the additional points used in RG. However, after a closer inspection, we can note that both estimators, corresponding to SG and RG coincide[3], i.e.,

$$\widehat{I}_T = \frac{1}{DT} \sum_{t=1}^{T} \sum_{d=1}^{D} f(\mathbf{x}_d^{(t)}) = \frac{1}{T} \sum_{t=1}^{T} f(\mathbf{x}_D^{(t)}). \tag{7}$$

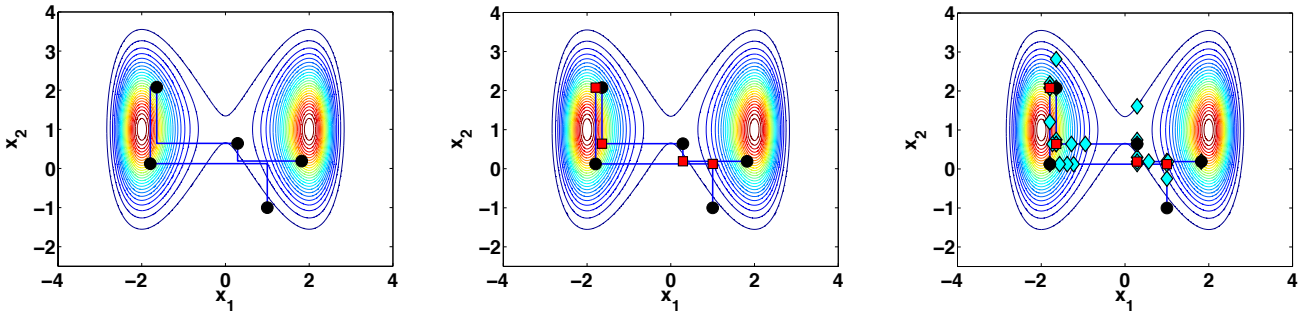
This is due to, in RG, each component $x_d^{(t)}$ is repeated exactly D times inside different consecutive samples in RG, and we have D times more samples in RG than in a SG sampler. Hence, there are no advantages in using RG w.r.t. a SG approach. However, we describe RG since it is a first theoretical step in order to design the Multiple Recycling Gibbs (MRG) sampler (see the next section). Indeed, if we draw M different (conditionally) independent samples from each full-conditionals, the resulting alternative MRG scheme presents interesting features and benefits, especially for the application of an adaptive MCMC-within-Gibbs approach (where an adaptive MCMC is employed in order to draw from the full-conditional pdfs).

5.1 The recycling Gibbs sampler with multiple points

Based on the previous considerations, we can design a more useful scheme called *Multiple Recycling Gibbs* (MRG) sampler, drawing M samples from each full conditional pdfs, as in Table 2. Observe that in order to go forward with the next iteration of the Gibbs sampler, we only consider the last generated component, i.e., we set $z_d^{(t)} = x_{d,M}^{(t)}$. However, any other sample $x_{d,m}^{(t)}$, with $m \in \{1, \ldots, M\}$, can be used. However, note that the last sample $x_{d,M}^{(t)}$ is the more convenient choice for an MCMC-within-Gibbs scheme. The final estimator (without removing any "burn-in" period) is formed by DMT samples,

$$\widehat{I}_T = \frac{1}{MDT} \sum_{t=1}^{T} \sum_{d=1}^{D} \sum_{m=1}^{M} f(\mathbf{x}_{d,m}^{(t)}). \tag{8}$$

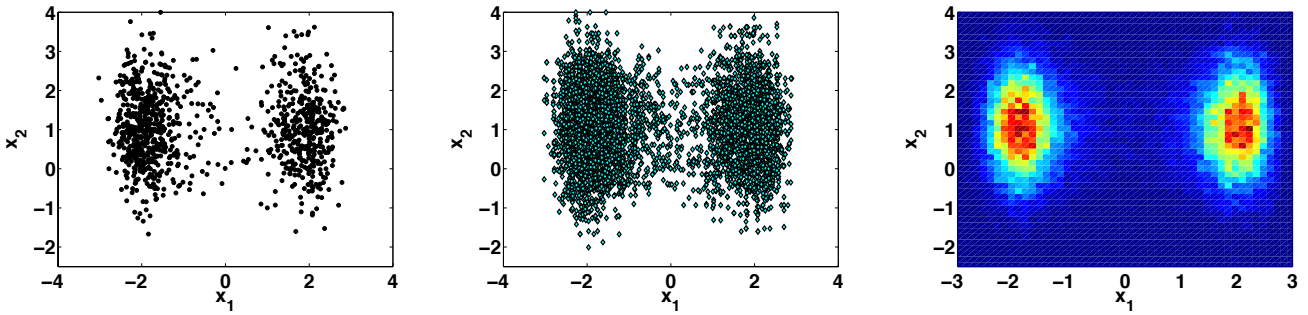
[3]Without removing any "burn-in" period.



(a) Standard Gibbs (SG) scheme. (b) Simple Recycling Gibbs (RG) scheme. (c) Multiple Recycling Gibbs (MRG) scheme.

Figure 3: We consider $T = 4$ iterations of a Gibbs sampler and $M = 5$ iterations of the MH for drawing from each full-conditional pdfs.**(a)** With the circles we denote the $T + 1$ points (considering the starting point) used in the standard Gibbs estimators. **(b)** The vectors (denoted with circles and squares) used in the RG estimators. **(c)** he vectors (denoted with circles, squares and rhombuses) used in the MRG estimators.

Note that with $M = 1$, we come back to the previous trivial RG scheme. MRG is particularly suitable when an adaptive MCMC is employed within a Gibbs scheme (i.e., when several MCMC steps are performed for sampling from each full-conditional). Indeed, we can employ all the sequence of samples generated by the internal MCMC algorithm in the final estimator. Table 3 shows the detailed steps of an MCMC-within-MRG algorithm.



(a) Points obtained with the SG. (b) Points obtained with MRG. (c) Histogram with MRG's outputs.

Figure 4: **(a)** Outputs of one MH-within-Gibbs run with $T = 10^3$ and $M = 5$, considering the target with contour plot shown in Fig. 3.**(b)** Outputs of one MH-within-MRG run with $T = 10^3$ and $M = 5$. **(c)** Histograms obtained using all the points in Figure (b), i.e., the MRG outputs with $T = 10^4$ and $M = 5$.

Figure 4(a) depicts the random vectors obtained with one run of an MH-within-Gibbs procedure, with $T = 10^3$ and $M = 5$. Figure 4(b) illustrates all the outputs of the previous run. Hence, these vectors are the samples obtained with a MH-within-MRG approach. Note that Figure 4(b) can be misleading since the MH-within-MRG scheme seems to collocate several points far from the two modes. However, The histogram of values reproduces adequately the shape the target pdf as shown in Figure 4(c) (this histogram has been obtained with one run of MH-within-MRG with $T = 10^4$ and $M = 5$).

Table 2: **Multiple Recycling Gibbs (MRG) sampler.**

- Choose a starting point $\mathbf{z}^{(0)} = [z_1^{(0)}, \ldots, z_D^{(0)}]$.

1. For $t = 1, \ldots, T$:
   (a) For $d = 1, \ldots, D$:
      i. For $m = 1, \ldots, M$:
         A. Draw $x_{d,m}^{(t)} \sim \bar{\pi}_d(x_d | z_1^{(t)}, \ldots, z_{d-1}^{(t)}, z_{d+1}^{(t-1)}, \ldots, z_D^{(t-1)})$.
         B. Set $\mathbf{x}_{d,m}^{(t)} = [z_1^{(t)}, \ldots, z_{d-1}^{(t)}, x_{d,m}^{(t)}, z_{d+1}^{(t-1)} \ldots, z_D^{(t-1)}]$.
      ii. Set $z_d^{(t)} = x_{d,M}^{(t)}$.[4]
2. Return $\{\mathbf{x}_{d,m}^{(t)}\}$ for all d, m and t.

Table 3: **Generic MCMC-within-MRG sampler.**

- Choose a starting point $\mathbf{z}^{(0)} = [z_1^{(0)}, \ldots, z_D^{(0)}]$.

1. For $t = 1, \ldots, T$:
   (a) For $d = 1, \ldots, D$:
      i. Perform M steps of an MCMC algorithm (as an MH scheme or a more advanced method) with target pdf $\bar{\pi}_d(x_d | x_{1:d-1}^{(t)}, x_{d+1,D}^{(t-1)})$, yielding the sequence of samples $x_{d,1}^{(t)}, \ldots, x_{d,M}^{(t)}$ (with starting point $x_{d,0}^{(t)} = x_d^{(t-1)}$).
      ii. Set $\mathbf{x}_{d,m}^{(t)} = [z_1^{(t)}, \ldots, z_{d-1}^{(t)}, x_{d,m}^{(t)}, z_{d+1}^{(t-1)} \ldots, z_D^{(t-1)}]$, for $m = 1, \ldots, M$.
      iii. Set $z_d^{(t)} = x_{d,M}^{(t)}$.
2. Return $\{\mathbf{x}_{d,m}^{(t)}\}$ for all d, m and t.

6 Simulations

We test the new MRG scheme in a simple numerical simulation, easily reproducible by any practitioner. More specifically, we consider a bi-dimensional target pdf

$$\bar{\pi}(x_1, x_2) \propto \exp\left(-\frac{(x_1^2 - \mu_1)^2}{2\delta_1^2} - \frac{(x_2 - \mu_2)^2}{2\delta_2^2} \right).$$

with $\mu_1 = 4$, $\mu_2 = 1$, $\delta_1 = \sqrt{\frac{5}{2}}$ and $\delta_2 = 1$. Figure 3 shows the contour plot of $\bar{\pi}(x_1, x_2)$. Our goal is to approximate via Monte Carlo the expected value, $E[\mathbf{X}]$ where $\mathbf{X} = [X_1, X_2] \sim \bar{\pi}(x_1, x_2)$, and the variances of component of $\mathbf{X}$. We test different Gibbs techniques, more specifically we consider the MH [42] and IA^{2}RMS [35] algorithms within SG and within MRG schemes. For the MH method we use a Gaussian random walk proposal, $q(x_{d,m}^{(t)}) \propto \exp\left(-\frac{(x_{d,m}^{(t)} - x_{d,m-1}^{(t)})^2}{2\sigma^2} \right)$ for $d \in \{1, 2\}$, $1 \leq m \leq M$ and $1 \leq t \leq T$. We test different values of σ. For IA^{2}RMS, we starts with the set of support points $\mathcal{S}_0 = \{-10, -6, -2, 2, 6, 10\}$ (see [35] for further details). When we consider the standard Gibbs (SG) scheme we have $M = 1$, whereas for MRG we have $M > 1$. We perform 10^5 independent runs for each Gibbs scheme. At each run and for each element of the expected value and the variances

(i.e., for 4 different values), we compute the error in estimation of the true values and calculate the Mean Square Error (MSE) averaged over the 10^5 independent runs (then, we obtained an averaged MSE over the different 4 estimated values).

Figure 5 shows the MSE (in log-scale) of the MH-within-SG scheme as function of the standard deviation σ of the proposal pdf (we set $M = 1$ and $T = 1000$, in this case). We can observe that the performance of the Gibbs samplers depends strongly on the choice of σ of the *internal* MH method. The optimal value is approximately $\sigma^* \approx 3$. The use of an adaptive proposal pdf is a possible solution (as shown in Figure 6(b)). Figure 6(a) depicts the MSE (in log-scale) as function of T with $M = 1$ and $M = 20$ (for MH-within-SG we also show the case $\sigma = 1$ and $\sigma = 3$). Again we observe that importance of using the optimal value $\sigma^* \approx 3$ and, as a consequence, an the use of an adaptive proposal pdf is recommended (e.g., see [16]). Moreover, the use $M = 20$ improves the results even without employing all the points in the estimators (i.e., in a SG scheme) since, as M increases due to the convergence of the internal chain, the last sample of the sequence becomes, roughly speaking, "statistically more similar" to a sample distributed as the full-conditional pdf. Moreover, the MH-within-MRG technique, i.e., the scheme where we recycle all the samples, provides the smallest MSE values. Namely, recycling the internal samples provides more efficient estimators.

These considerations are confirmed in Figure 6(b) (represented in log-scale). Here we fix again $T = 1000$ and vary M. As M increases. the MSE becomes smaller and smaller when the MRG technique is employed. When a standard Gibbs (SG) approach is used, the curves present an horizontal asymptote since after a certain value $M \geq M^*$, the internal chains converge and there is no benefit in increasing M (recall that in SG we do not recycle the internal samples). Within a MRG scheme the increase of M always generates a decrease of the MSE since we recycle the internal samples (recall that the figure is in log-scale). Clearly, the benefit of using MRG w.r.t. SG increases as M grows. Figure 6(b) also shows the advantage of using an adaptive MCMC scheme (in this case IA^{2}RMS [35]). The benefit is clearer when the scheme are used within MRG. More specifically, since for the MH method we have used the optimal scale $\sigma^* \approx 3$, Figure 6(b) illustrates the importance of a non-parametric construction of the proposal pdf employed in IA^{2}RMS. Indeed, the non-parametric construction allows the adaptation of the entire shape of the proposal (which becomes closer and closer to the target). The performance of IA^{2}RMS and MH within Gibbs becomes more similar as M increases since, in this case, with a great enough M, also the MH chain is able to exceed its "burn-in" period and then converges. Finally, note that the adaptation speeds up the convergence of the chain generated by IA^{2}RMS. The advantage of using the adaptation is more evident for intermediate values of M, e.g., $10 < M < 30$, where the difference with the use of a standard MH is greater. As M increases and the chain generated by MH converges, the difference on the performance between IA^{2}RMS and MH becomes smaller.

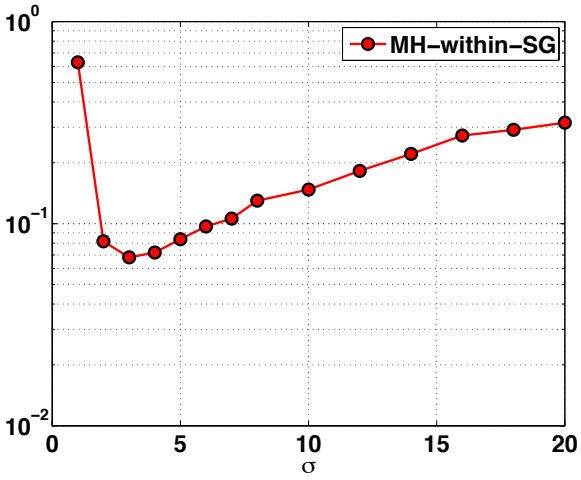
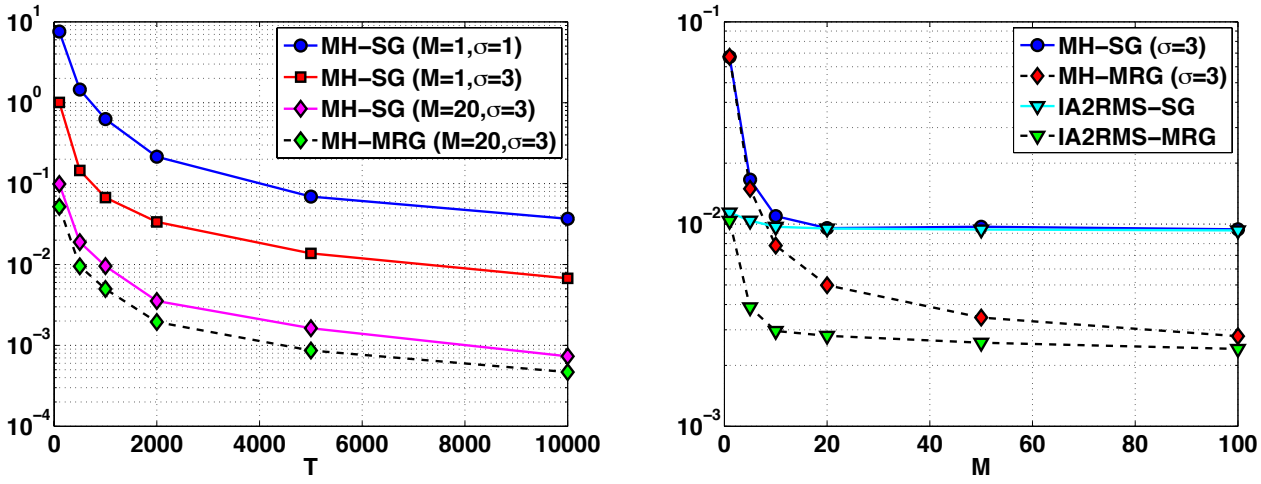


Figure 5: MSE (log-scale) as function of σ for MH-within-SG ($M = 1$ and $T = 1000$).



(a) MSE (log-scale) as function of T.

(b) MSE (log-scale) as function of M ($T = 1000$).

Figure 6: **(a)** MSE (log-scale) as function of T for MH-within-SG and MH-within-MRG schemes. We have tested $M \in \{1, 20\}$ and $\sigma = \{1, 3\}$ (we recall that $\sigma = 3$ is the optimal scale parameter for MH; see Figure 5). **(b)** MSE (log-scale) as function of M for different MCMC-within-Gibbs schemes (we set $T = 1000$).

7 Conclusions

The Gibbs sampler is a well-known Markov chain Monte Carlo (MCMC) algorithm extensively applied in literature. It is particularly useful to draw from complicated high-dimensional posterior distributions, by drawing from full-conditional pdfs (i.e., in the standard case, the random vector is drawn component by component). Hence, the key point for the successful application of the Gibbs sampler is the ability to draw efficiently from the full-conditional pdfs. For this purpose, several algorithms has been proposed to be employed within the Gibbs sampler. Most of them require the generation of auxiliary samples that are not included in the final estimators. In this work, we have shown that these auxiliary samples can be re-used within the Gibbs estimators improving their efficiency. Numerical simulations confirm the benefits of the novel scheme.

Acknowledgments

This work has been supported by the European Research Council (ERC) through the ERC Consolidator Grant SEDAL ERC-2014-CoG 647423.

References

[1] M. Brewer and C. Aitken. Discussion on the meeting on the Gibbs sampler and other Markov Chain Monte Carlo methods. *Journal of the Royal Statistical Society. Series B*, 55(1):69–70, 1993.

[2] M. F. Bugallo, S. Xu, and P. M. Djurić. Performance comparison of EKF and particle filtering methods for maneuvering targets. *Digital Signal Processing*, 17:774–786, October 2007.

[3] B. S. Caffo, J. G. Booth, and A. C. Davison. Empirical supremum rejection sampling. *Biometrika*, 89(4):745–754, December 2002.

[4] B. Cai, R. Meyer, and F. Perron. Metropolis-Hastings algorithms with adaptive proposals. *Statistics and Computing*, 18:421–433, 2008.

[5] L. Devroye. *Non-Uniform Random Variate Generation*. Springer, 1986.

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

[7] 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.

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

[9] C. Fox. A Gibbs sampler for conductivity imaging and other inverse problems. *Proc. of SPIE, Image Reconstruction from Incomplete Data VII*, 8500:1–6, 2012.

[10] A. E. Gelfand and T. M. Lee. Discussion on the meeting on the Gibbs sampler and other Markov Chain Monte Carlo methods. *Journal of the Royal Statistical Society. Series B*, 55(1):72–73, 1993.

[11] W. R. Gilks. Derivative-free adaptive rejection sampling for Gibbs sampling. *Bayesian Statistics*, 4:641–649, 1992.

[12] 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.

[13] W. R. Gilks, R.M. Neal, N. G. Best, and K. K. C. Tan. Corrigidum: Adaptive rejection Metropolis sampling within Gibbs sampling. *Applied Statistics*, 46(4):541–542, 1997.

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

[15] Dilan Görür and Yee Whye Teh. Concave convex adaptive rejection sampling. *Journal of Computational and Graphical Statistics*, 20(3):670–691, September 2011.

[16] H. Haario, E. Saksman, and J. Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7(2):223–242, April 2001.

[17] W. Hörmann. A rejection technique for sampling from T-concave distributions. *ACM Transactions on Mathematical Software*, 21(2):182–193, 1995.

[18] W. Hörmann. A universal generator for bivariate log-concave distributions. *Computing*, 52:89–96, 1995.

[19] W. Hörmann. A note on the performance of the Ahrens algorithm. *Computing*, 69:83–89, 2002.

[20] W. Hörmann, J. Leydold, and G. Derflinger. *Automatic nonuniform random variate generation*. Springer, 2003.

[21] 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.

[22] L. Jing and P. Vadakkepat. Interacting mcmc particle filter for tracking maneuvering target. *Digital Signal Processing*, 20:561–574, August 2010.

[23] K. R. Koch. Gibbs sampler by sampling-importance-resampling. *Journal of Geodesy*, 81(9):581–591, 2007.

[24] K. R. Koch. Gibbs sampler by sampling-importance-resampling. *Journal of Geodesy*, 81(9):581–591, 2007.

[25] 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.

[26] F. Liang, C. Liu, and R. Caroll. *Advanced Markov Chain Monte Carlo Methods: Learning from Past Samples*. Wiley Series in Computational Statistics, England, 2010.

[27] J. S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer-Verlag, 2004.

[28] F. Lucka. Fast Gibbs sampling for high-dimensional Bayesian inversion. *arXiv:1602.08595*, 2016.

[29] G. Marrelec and H. Benali. Automated rejection sampling from product of distributions. *Computational Statistics*, 19(2):301–315, May 2004.

[30] L. Martino, R. Casarin, and D. Luengo. Sticky proposal densities for adaptive MCMC methods. *IEEE Workshop on Statistical Signal Processing (SSP)*, 2016.

[31] L. Martino and F. Louzada. Adaptive rejection sampling with fixed number of nodes. *viXra:1509.0048*, (15), 2015.

[32] L. Martino and J. Míguez. Generalized rejection sampling schemes and applications in signal processing. *Signal Processing*, 90(11):2981–2995, November 2010.

[33] L. Martino and J. Míguez. A generalization of the adaptive rejection sampling algorithm. *Statistics and Computing*, 21(4):633–647, October 2011.

[34] L. Martino, J. Read, and D. Luengo. Independent doubly adaptive rejection Metropolis sampling. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2014.

[35] L. Martino, J. Read, and D. Luengo. Independent doubly adaptive rejection Metropolis sampling within Gibbs sampling. *IEEE Transactions on Signal Processing*, 63(12):3123–3138, June 2015.

[36] L. Martino, H. Yang, D. Luengo, J. Kanniainen, and J. Corander. A fast universal self-tuned sampler within Gibbs sampling. *Digital Signal Processing*, 47:68 – 83, 2015.

[37] 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.

[38] J. Míguez. Analysis of selection methods for cost-reference particle filtering with applications to maneuvering target tracking and dynamic optimization. *Digital Signal Processing*, 17:787–807, 2007.

[39] P. Müller. A generic approach to posterior integration and Gibbs sampling. *Technical Report 91-09, Department of Statistics of Purdue University*, 1991.

[40] K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.

[41] C. Ritter and M. A. Tanner. Facilitating the Gibbs sampler: The Gibbs stopper and the griddy-Gibbs sampler. *Journal of the American Statistical Association*, 87(419):861–868, 1992.

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

[43] G. O. Roberts and S. K. Sahu. Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *Journal of the Royal Statistical Society: Series B*, 59(2):291–317, 1997.

[44] J. K. O. Ruanaidh and W. J. Fitzgerald. *Numerical Bayesian Methods Applied to Signal Processing*. Springer, 1996.

[45] W. Shao, G. Guo, F. Meng, and S. Jia. An efficient proposal distribution for Metropolis-Hastings using a b-splines technique. *Computational Statistics and Data Analysis*, 53:465–478, 2013.

[46] H. Tanizaki. On the nonlinear and non-normal filter using rejection sampling. *IEEE Transaction on automatic control*, 44(3):314–319, February 1999.

[47] H. Zhang, Y. Wu, L. Cheng, and I. Kim. Hit and run ARMS: adaptive rejection Metropolis sampling with hit and run random direction. *Journal of Statistical Computation and Simulation*, 86(5):973–985, 2016.