Digital Signal Processing 47 (2015) 68–83

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

A fast universal self-tuned sampler within Gibbs sampling

L. Martino [a,*], H. Yang [b], D. Luengo [c], J. Kanniainen [b], J. Corander [a]

[a] Dep. of Mathematics and Statistics, University of Helsinki, 00014 Helsinki, Finland
[b] Dep. of Industrial Engineering, Tampere University of Technology, Tampere, Finland
[c] Dep. of Signal Theory and Communications, Universidad Politécnica de Madrid, 28031 Madrid, Spain

ARTICLE INFO

Article history:
Available online 16 April 2015

Keywords:
Markov Chain Monte Carlo (MCMC)
Gibbs sampling
Bayesian inference
Metropolis within Gibbs
Adaptive rejection Metropolis sampling

ABSTRACT

Bayesian inference often requires efficient numerical approximation algorithms, such as sequential Monte Carlo (SMC) and Markov chain Monte Carlo (MCMC) methods. The Gibbs sampler is a well-known MCMC technique, widely applied in many signal processing problems. Drawing samples from univariate full-conditional distributions efficiently is essential for the practical application of the Gibbs sampler. In this work, we present a simple, self-tuned and extremely efficient MCMC algorithm which produces virtually independent samples from these univariate target densities. The proposal density used is self-tuned and tailored to the specific target, but it is not adaptive. Instead, the proposal is adjusted during an initial optimization stage, following a simple and extremely effective procedure. Hence, we have named the newly proposed approach as FUSS (Fast Universal Self-tuned Sampler), as it can be used to sample from any bounded univariate distribution and also from any bounded multi-variate distribution, either directly or by embedding it within a Gibbs sampler. Numerical experiments, on several synthetic data sets (including a challenging parameter estimation problem in a chaotic system) and a high-dimensional financial signal processing problem, show its good performance in terms of speed and estimation accuracy.

1. Introduction

Bayesian methods, and their implementations by means of sophisticated Monte Carlo techniques [1,2], have become very popular over the last two decades. Indeed, many practical statistical signal processing problems demand procedures for drawing from probability distributions with non-standard forms, such as Markov chain Monte Carlo (MCMC) methods [3,4] and particle filters [5–7]. MCMC techniques generate samples from a target probability density function (pdf) by drawing from a simpler proposal pdf [1,8] and generating a Markov chain. The two most widely applied MCMC approaches are the Metropolis–Hastings (MH) algorithm and the Gibbs sampler [1,2].

The Gibbs sampling technique is extensively used in Bayesian inference [9] to generate samples from multivariate target densities, drawing each component of the samples from univariate full-conditional densities [10–12].[1] When the multivariate target can be easily factorized into univariate conditional pdfs, the key point for the successful application of the Gibbs sampler is the ability to draw efficiently from these univariate pdfs [1,2,10]. The best scenario for Gibbs sampling occurs when exact samplers for each full-conditional are available. Otherwise, another exact sampling technique, like rejection sampling (RS) or an MH-type algorithm, is typically used *within* the Gibbs sampler to draw from the complicated full-conditionals. In the first case, samples generated from the RS algorithm are independent, but the acceptance rate can be very low. In the second case, we have an approach where an internal MCMC (the MH method) is applied *inside* another external MCMC (the Gibbs sampler). Therefore, the typical problems of the *external*-MCMC (long "burn-in" period, large correlation, etc.) could raise dramatically if the *internal*-MCMC is not extremely efficient. Indeed, although the Gibbs sampler needs only one sample from

* Corresponding author.
E-mail address: lukatotal@gmail.com (L. Martino).

[1] Blockwise Gibbs sampling approaches, where several random variables are updated simultaneously, have been proposed to speed up the convergence of the Gibbs sampler [13]. 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 multi-variate full-conditionals is unfeasible.

http://dx.doi.org/10.1016/j.dsp.2015.04.005

each full-conditional, several iterations are typically performed to avoid the "burn-in" period of the *internal*-MCMC.[2]

In order to avoid these problems, several automatic and self-tuning samplers have been proposed: *adaptive rejection sampling* (ARS) [14,15], *Adaptive Rejection Metropolis Sampling* (ARMS) [16–18], *Independent Doubly Adaptive Rejection Metropolis Sampling* (IA^{2}RMS) [19,20], *Adaptive Sticky Metropolis* (ASM) [21], etc. ARS builds a piecewise linear proposal on the target's log-domain, starting with a reduced number of support points and incorporating new points whenever a candidate sample is rejected. Unfortunately, since it is based on the rejection sampling technique, the proposal must be always above the target, a requirement which is only fulfilled by log-concave targets. In order to solve this issue, ARMS introduces a Metropolis–Hastings step, thus obtaining a universal sampler which is able to draw virtually from any target pdf. However, the adaptive structure in ARMS has an important restriction: support points cannot be added inside regions where the proposal is below the target. Recently, the IA^{2}RMS and ASM algorithms have been proposed to overcome this drawback, introducing more flexibility in the mechanism used to add points to the support set and decoupling it from the proposal construction.

All the previous methods build an adaptive sequence of proposal pdfs via some interpolation procedure given a set of support points. The proposal is updated when a new support point is incorporated, according to some statistical criterion. However, although these methods can attain a very good performance, the results show a dependence on the initial set of support points. Another drawback is the difficulty of ensuring their ergodicity, especially in applications within Gibbs sampling [2,17], as the Markovian nature of the chain is lost due to the adaptive nature of the proposal, which may depend on all the previous samples. Other related works, where a non-adaptive proposal pdf is built via interpolation procedures can be found in literature [22–24]. Furthermore, different types of generic adaptive MH schemes based on independent proposals have also been studied [25,26]. However, in general, the considered proposal pdf has a fixed parametric form so that the complete adaptation of the proposal is not possible.

In this work, we present a novel algorithm that follows a complementary strategy: start with a large number of support points and remove many of them following some pruning strategy.[3] The key idea is starting with a thin uniform grid that covers the effective support of the target and discard those support points that do not provide relevant information according to some pre-defined criterion. The idea of using a grid and a piecewise linear constant function to approximate a uni-variate full conditional was already proposed by the griddy Gibbs sampler [27]. However, their approach is substantially different from ours. On the one hand, they simply select a few support points heuristically, instead of starting with a large set of points and selecting the best ones in a principled way. On the other hand, the proposal in the griddy Gibbs sampler does not take into account the tails of the target (the proposal is set to zero outside of the interval covered by the grid), thus providing a poor performance for slowly decaying tails (e.g., heavy tailed distributions). Furthermore, this griddy approximation of the uni-variate full conditionals is not embedded within another *inner* Monte Carlo method, thus leading to an approximate sampler, unlike our scheme, which results in an exact sampler.

The resulting method is fast and extremely efficient (it yields virtually independent samples), even for highly multimodal and complicated targets. The dependence on the initial set of points is also drastically reduced, since the algorithm only requires an approximate knowledge of the effective support of the target pdf. Moreover, unlike previous approaches, the proposal is self-tuned during the initialization stage, without any adaptation afterwards. Hence, ergodicity is not an issue and the convergence of the chain to the target distribution is always guaranteed. For these reasons, we call the new method FUSS ("Fast Universal Self-tuned Sampler") since, with this sampler, there is no "fuss" about convergence or tuning. The FUSS algorithm is particularly well suited for multimodal and spiky target densities (i.e., densities with several sharp and narrow modes), where virtually all of the existing MCMC techniques fail. This kind of target pdfs often appears in practical applications, e.g., in ecology, bioinformatics and financial inference problems (see Sections 6 and 7).

The rest of the paper is organized as follows. Sections 2 and 3 are devoted to recalling the general framework and describing the structure of the novel technique. Details about the proposal construction and generation are given in Section 4. Different pruning algorithms are then introduced in Section 5. Sections 6 and 7 provide numerical results on several uni-variate and multi-variate pdfs, including a challenging parameter estimation problem in a chaotic system, as well as a multi-dimensional and multi-modal inference problem in financial signal processing. Finally, Section 8 contains some brief final remarks.

2. Problem statement

Bayesian inference often requires drawing samples from complicated multivariate posterior pdfs, $\pi(\mathbf{x}|\mathbf{y})$ with $\mathbf{x} \in \mathcal{X}^D \subseteq \mathbb{R}^D$. A common approach, when direct sampling from $\pi(\mathbf{x}|\mathbf{y})$ is unfeasible, is using a Gibbs sampler [2]. At the i-th iteration, a Gibbs sampler obtains the d-th component ($d = 1, \dots, D$) of $\mathbf{x}$, x_d, by drawing from the full conditional pdf of x_d given all the previously generated components [2,9,28], i.e.,

$$x_d^{(i)} \sim \bar{\pi}(x_d|\mathbf{x}_{1:d-1}^{(i)}, \mathbf{x}_{d:D}^{(i-1)}) = \bar{\pi}(x_d) \propto \pi(x_d), \tag{1}$$

where $\bar{\pi}(x_d)$ is the normalized target pdf, $\pi(x_d)$ denotes its unnormalized counterpart (note that we have dropped the dependence on $\mathbf{x}_{1:d-1}^{(i)}$ and $\mathbf{x}_{d:D}^{(i-1)}$ to simplify the notation), $x_d \in \mathcal{X}$ and the initial vector is typically drawn from the prior (i.e., $\mathbf{x}^{(0)} \sim \bar{\pi}_0(\mathbf{x})$), but can also be set to some fixed value when no prior information is available or it is unreliable.

However, even sampling from the univariate pdfs in Eq. (1) can often be complicated. In these cases, a common approach is to use another Monte Carlo technique (e.g., rejection sampling (RS) or the Metropolis–Hastings (MH) algorithm) within the Gibbs sampler, drawing candidates from a simpler proposal pdf,

$$\bar{p}(x) \propto p(x) = e^{W(x)},$$

where $\bar{p}(x)$ and $p(x)$ denote the normalized and unnormalized proposal respectively, $W(x)$ is a "potential" function and $x \in \mathbb{R}$. The best case occurs when an RS technique can be applied, since it yields independent and identically distributed (i.i.d.) samples. However, RS requires $p(x) \geq \pi(x)$ for all $x \in \mathcal{X}$, which may be hard to guarantee in practice. For instance, the adaptive rejection sampling (ARS) technique can be applied only to log-concave target pdfs [15]. Thus, the use of another MCMC method becomes almost mandatory in practical applications. In this case, the performance of this approach depends strictly on the choice of $p(x)$. Our aim

[2] Note that, from a theoretical point of view, performing a single iteration of the internal MH is enough to guarantee the ergodicity of the Gibbs sampler. However, the convergence of the chain can be very slow. Namely, the performance of the resulting estimator built from those samples can be very poor if the proposal pdf is not very similar to the full-conditionals. Hence, several iterations of the internal MH algorithm are typically required in order to achieve the desired level of performance (more as the proposal differs more from the target). See the numerical examples in Section 6 for further details on this issue.

[3] Note that all of the previous methods typically follow the opposite strategy: start with a reduced number of support points and keep adding points in order to improve the proposal adaptively.