Computational Statistics and Data Analysis 109 (2017) 231–246

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Hierarchical models: Local proposal variances for RWM-within-Gibbs and MALA-within-Gibbs

Mylène Bédard

Département de mathématiques et de statistique, Université de Montréal, C.P. 6128, H3C 3J7, Canada

HIGHLIGHTS

- Locally optimal proposal variances are introduced for RWM-within-Gibbs algorithms.
- These locally optimal tunings are shown to theoretically outperform constant ones.
- Similar state-dependent step sizes are discussed for MALA-within-Gibbs samplers.
- MALA-within-Gibbs constitutes an efficient, yet computationally affordable option.
- Efficiency of local tunings depends on the variability in the hierarchical target.

ARTICLE INFO

Article history:
Received 3 February 2016
Received in revised form 7 December 2016
Accepted 15 December 2016
Available online 24 December 2016

Keywords:
Efficiency
Gaussian proposal distribution
Inhomogeneous proposal variances
Langevin diffusion
Location/scale parameters
MALA

ABSTRACT

The performance of RWM- and MALA-within-Gibbs algorithms for sampling from hierarchical models is studied. For the RWM-within-Gibbs, asymptotically optimal tunings for Gaussian proposal distributions featuring a diagonal covariance matrix are developed using existing scaling analyses. This leads to locally optimal proposal variances that depend on the mixing components of the hierarchical model and that correspond to the classical asymptotically optimal acceptance rate of 0.234. Ignoring the local character of the optimal scaling is possible, leading to an optimal proposal variance that remains fixed for the duration of the algorithm; the corresponding asymptotically optimal acceptance rate is then shown to be lower than 0.234. Similar ideas are applied to MALA-within-Gibbs samplers, leading to efficient yet computationally affordable algorithms. Simplifications for location and scale hierarchies are presented, and findings are illustrated through numerical studies. The local and fixed approaches for the RWM- and MALA-within-Gibbs are compared to competitive samplers in the literature.

1. Introduction

The Random walk Metropolis algorithm (RWM) and the Metropolis-adjusted Langevin algorithm (MALA) are commonly used to produce samples from arbitrary distributions π that may be complex, high-dimensional, or both (Hastings, 1970). The idea is to build a Markov chain $\{\mathbf{X}[j], j \in \mathbb{N}\}$ on a state space $\mathcal{X}$ by proposing candidates to be included in the process according to some acceptance probability. The resulting Markov chain admits the n-dimensional target distribution π as its unique invariant distribution. Hereafter, π shall also be used for denoting the target density on a state space $\mathcal{X}$ with respect to Lebesgue measure.

Suppose that the time-j state of the Markov chain is $\mathbf{X}[j] = \mathbf{x}$. In a (symmetrical) RWM algorithm for instance, the proposal distribution selected to generate a candidate $\mathbf{Y}[j+1] = \mathbf{y}$ for the next state of the chain is assumed to have a density

E-mail address: bedard@dms.umontreal.ca.

http://dx.doi.org/10.1016/j.csda.2016.12.007
0167-9473/

$q_n(\mathbf{y};\mathbf{x}) = q_n(|\mathbf{y}-\mathbf{x}|)$ with respect to Lebesgue measure. A pragmatic choice, on which we focus in this article, is to draw candidates from a $\mathcal{N}(\mathbf{x}, \sigma^2 I_n)$ for some $\sigma > 0$, where I_n is the n-dimensional identity matrix (the specific normal proposal distribution used with MALA shall be described in Section 5). In implementing RWM and MALA samplers, one can update all n components simultaneously (classical RWM/MALA), or divide them into subgroups to be updated consecutively (RWM- or MALA-within-Gibbs). The latter are commonly preferred for sampling hierarchical models, as full conditional densities are usually available.

The variance of the normal proposal distribution (σ^2) has a significant impact on the speed at which the process travels across its state space (hereafter referred to as "efficiency"), with extremal variances leading to slow-mixing samplers. Simply put, large variances induce lazy processes (large candidate jumps that are refused), while small variances yield hyperactive processes (tiny candidate steps that are accepted). To optimize exploration of the state space, we aim for candidate steps that strike a balance, so that we have sizable steps that are still accepted a reasonable proportion of the time. Seeking for this intermediate proposal variance is called the optimal scaling problem.

There exist, in Markov chain Monte Carlo theory, different notions of efficiency. In this paper, the term efficiency is used as a measure of how rapidly the Markov chain explores its state space once stationarity has been reached. For finite-dimensional chains, this can be measured by the expected squared jumping distance (ESJD) to be introduced in (13). In an infinite-dimensional setting, the theoretical (or asymptotical) efficiency is measured through the speed function of the limiting Langevin diffusion, to be discussed in Sections 3 and 4. In the high-dimensional limit ($n \to \infty$), the ESJD is equivalent to the limiting speed measure.

This paper studies the optimal scaling theory for RWM-within-Gibbs with some heuristics for MALA-within-Gibbs, and then looks at the performance of both in practice. In particular, the theory exposed leads to the determination of proposal variances and acceptance rates producing optimally mixing RWM-within-Gibbs chains. The theoretical results are derived for high-dimensional hierarchical target densities with a large number of conditionally independent and identically distributed (i.i.d.) components. The principal difference with traditional optimal scaling results lies in the local character of the optimal proposal variances obtained, meaning that they vary from one iteration to the next. The concept of local proposal variances has been discussed in Girolami and Calderhead (2011) and Bédard (submitted for publication); in the latter, scaling analyses of the RWM algorithm for hierarchical target densities are performed. Although theoretically appealing, local proposal variances had to be obtained numerically in that context, which turned out to be rather impractical. With the RWM-within-Gibbs sampler (and even the MALA-within-Gibbs), these variances may now be found analytically in a large number of cases, leading to a personalized version of the proposal variance in a given iteration. The theoretical results derived thus stand on the work in Bédard (submitted for publication), and as such are expressed as a corollary of its main theorem.

The derivation of local proposal variances requires that certain expectations be obtained analytically from the hierarchical model considered. The new approach is thus predicated on the tractability of the distribution of the conditionally i.i.d. components, given the mixing parameters and (in practice) the observations. It is thus well suited to some hierarchical models; alternatively, we propose a fixed optimal proposal variance, which is shown to be less efficient than the local ones. In an attempt to quantify the benefit, in terms of efficiency, of using local proposal variances rather than a fixed one in the RWM- and MALA-within-Gibbs, we present numerical illustrations. To add some perspective, we compare these samplers to single-block RWM and MALA algorithms, along with some of their variants that include correlation among candidates. We also include the Adaptive Metropolis (AM) sampler of Haario et al. (2001), which tunes the proposal covariance matrix on the fly.

We shall realize that in tractable cases (and when there is not a strong correlation between mixing parameters and the remaining components), local versions of RWM- and MALA-within-Gibbs can outperform fancy variants included in the MCMC toolbox. Local MALA-within-Gibbs is the approach that provides the most convincing results, leading to net efficiency gains in a wide range of situations, compared to a large set of competitors. These gains are however largely influenced by the degree of variability present in the hierarchical model (a large variability sustaining the pursuit of local proposal variances). Even in cases where local samplers do not allow for large gains in terms of theoretical efficiency, the risk associated with these local variances is limited to the extra computational effort required for their implementation, which is usually insignificant compared to a fixed variance.

The next section sets up the framework, while Section 3 reviews optimal scaling notions for high-dimensional i.i.d. and hierarchical targets. Understanding these notions turns out to be useful in Section 4, where we derive optimal tunings for the RWM-within-Gibbs for sampling from hierarchical models; extensions to MALA-within-Gibbs are then discussed in Section 5. Section 6 focuses on single-level hierarchical models where the mixing parameter acts on the location or scale of the conditionally i.i.d. components; a simulation study illustrates the theoretical results. An extension to inhomogeneous proposal variances is introduced in Section 7, and we conclude by presenting a numerical study on a hierarchical target model that falls slightly outside the assumptions of the theory (Section 8).

2. Framework

Consider the following $(n+p)$-dimensional target density π with respect to Lebesgue measure

$$\pi(\mathbf{x}^{(n+p)}) = f_1(x_1,\ldots,x_p)\prod_{i=p+1}^{p+n} f(x_i|x_1,\ldots,x_p); \tag{1}$$