Statistics and Probability Letters 90 (2014) 11–16

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Convergence bound in total variation for an image restoration model

Oliver Jovanovski *

Department of Mathematics and Statistics, 4700 Keele Street, York University, M3J 1P3, Canada

ARTICLE INFO

Article history:
Received 20 January 2014
Received in revised form 12 March 2014
Accepted 12 March 2014
Available online 21 March 2014

Keywords:
Markov chain
Gibbs sampler
MCMC
Image restoration

ABSTRACT

We consider a stochastic image restoration model proposed by A. Gibbs (2004), and give an upper bound on the time it takes for a Markov chain defined by this model to be ϵ-close in total variation to equilibrium. We use Gibbs' result for convergence in the Wasserstein metric to arrive at our result. Our bound for the time to equilibrium of similar order to that of Gibbs.

1. Introduction

A.L. Gibbs (Gibbs, 2004) introduced a stochastic image restoration model for an N pixel greyscale image $x = \{x_i\}_{i=1}^N$. More specifically, in this model each pixel x_i corresponds to a real value in $[0, 1]$, where a black pixel is represented by 0 and a white pixel is represented by the value 1. It is assumed that in the real-world space of such images, each pixel tends to be like its nearest neighbours (in the absence of any evidence otherwise). This assumption is expressed in the prior probability density of the image, which is given by

$$\pi_\gamma(x) \propto \exp\left\{-\sum_{\langle i,j\rangle} \frac{1}{2}\left[\gamma\left(x_i - x_j\right)\right]^2\right\} \tag{1.1}$$

on the state space $[0, 1]^N$, and is equal to 0 elsewhere. The sum in (1.1) is over all pairs of pixels that are considered to be neighbours, and the parameter γ represents the strength of the assumption that neighbouring pixels are similar. Here images are assumed to have an underlying graph structure. The familiar 2-dimensional digital image is a special case, where usually one might assume that the neighbours of a pixel x_i in the interior of the image (i.e. x_i not on the boundary of the image) are the 4 or 8 pixels surrounding x_i, depending on whether or not we decide to consider the 4 pixels diagonal to x_i.

The actual observed image $y = \{y_i\}_{i=1}^N$ is assumed to be the result of the original image subject to distortion by random noise, with every pixel modified independently through the addition of a *Normal* $\left(0, \sigma^2\right)$ random variable (hence $y_i \in \mathbb{R}$). The resulting posterior probability density for the original image is given by

$$\pi_{posterior}(x|y) \propto \exp\left\{-\sum_{i=1}^{N} \frac{1}{2\sigma^2}\left(x_i - y_i\right)^2 - \sum_{\langle i,j\rangle} \frac{1}{2}\left[\gamma\left(x_i - x_j\right)\right]^2\right\} \tag{1.2}$$

supported on $[0, 1]$.

* Tel.: +1 416 736 2100x40616.
E-mail address: olijovan@mathstat.yorku.ca.

http://dx.doi.org/10.1016/j.spl.2014.03.007

Samples from (1.2) can be approximately obtained by means of a Gibbs sampler. In this instance, the algorithm works as follows: at every iteration the sampler chooses a site i uniformly at random, and replaces the value x_i at this location according to the full conditional density at that site. This density is given by

$$\pi_{FC}\left(x_i|y, x_{k\neq i}\right) \propto \exp\left\{\frac{\left(\sigma^{-2}+n_i\gamma^2\right)}{2}\cdot\left[x_i-\left(\sigma^{-2}+n_i\gamma^2\right)^{-1}\left(\sigma^{-2}y_i+\gamma^2\sum_{j\sim i}x_j\right)\right]^2\right\} \tag{1.3}$$

on $[0, 1]$ and 0 elsewhere. Here n_i is the number of neighbours the ith pixel has, and $j \sim i$ indicates that the jth pixel is one of them. It follows that (1.3) is a restriction of a *Normal* $\left(\left(\sigma^{-2}+n_i\gamma^2\right)^{-1}\left(\sigma^{-2}y_i+\gamma^2\sum_{j\sim i}x_j\right), \left(\sigma^{-2}+n_i\gamma^2\right)^{-1}\right)$ distribution to the set $[0, 1]$.

The bound on the rate of convergence to equilibrium given in Gibbs (2004) is stated in terms of the Wasserstein metric d_W. This is defined as follows: if μ_1 and μ_2 are two probability measures on the same state space which is endowed with some metric d, then

$$d_W\left(\mu_1, \mu_2\right) := \inf \mathbb{E}\left[d\left(\xi_1, \xi_2\right)\right]$$

where the infimum is taken over all joint distributions (ξ_1, ξ_2) such that $\xi_1 \sim \mu_1$ and $\xi_2 \sim \mu_2$.

Another commonly used metric for measuring the distance of a Markov chain from its equilibrium distribution is the total variation metric, defined for two probability measures μ_1 and μ_2 on the state space Ω by

$$d_{TV}\left(\mu_1, \mu_2\right) := sup\left|\mu_1(A) - \mu_2(A)\right|$$

where the supremum is taken over all measurable $A \subseteq \Omega$.

The underlying metric on the state space used throughout (Gibbs, 2004) (and hence used implicitly in the statement of Theorem 1) is defined by $d(x, y) := \sum_i n_i |x_i - z_i|$. This is a non-standard choice for a metric on $[0, 1]^N$, however it is comparable to the more usual l_1 taxicab metric $\hat{d}(x, y) := \sum_i |x_i - z_i|$ since

$$n_{\min}\cdot\hat{d}(x, y) \leq d(x, y) \leq n_{\max}\cdot\hat{d}(x, y)$$

where $n_{\max} := \max_i\{n_i\}$ and $n_{\min} := \min_i\{n_i\}$. Hence, for two probability measures μ_1 and μ_2 on $[0, 1]^N$, it follows immediately that

$$n_{\min}\cdot d_{\hat{W}}\left(\mu_1, \mu_2\right) \leq d_W\left(\mu_1, \mu_2\right) \leq n_{\max}\cdot d_{\hat{W}}\left(\mu_1, \mu_2\right)$$

where $d_{\hat{W}}$ and d_W are the Wasserstein metrics associated with $\hat{d}$ and d respectively.

If Θ_1 and Θ_2 are two random variables on the same state space with probability measures m_1 and m_2 respectively, then we shall write

$$d_W\left(\Theta_1, \Theta_2\right) := d_W\left(m_1, m_2\right) \quad \text{and} \quad d_{TV}\left(\Theta_1, \Theta_2\right) := d_{TV}\left(m_1, m_2\right).$$

Gibbs (2004) shows that

Theorem 1 (*Gibbs, 2004*)**.** *Let X^t be a copy of the Markov chain evolving according to the Gibbs sampler, and let Z^t be a chain in equilibrium, distributed according to $\pi_{posterior}$. Then if $[0, 1]^N$ is given the metric $d(x, y) := \sum_i n_i |x_i - z_i|$, it follows that $d_W\left(X^t, Z^t\right) \leq \epsilon$ whenever*

$$t > \vartheta(\epsilon) := \frac{\log\left(\frac{\epsilon}{n_{\max}N}\right)}{\log\left(1 - N^{-1}\left(1 + n_{\max}\gamma^2\sigma^2\right)^{-1}\right)}. \tag{1.4}$$

By the comments preceding the statement of this theorem, (1.4) remains true with the standard l_1 metric on the state space, if we replace ϵ by $n_{\min}\cdot\epsilon$ on the right-hand side of this inequality.

Remark. Eq. (1.4) appears in Gibbs (2004) with the denominator being $\log\left(N - 1/N + n_{\max}N^{-1}\gamma^2\left(\sigma^{-2} + n_{\max}\gamma^2\right)^{-1}\right)$. It is obvious from their proof that this is a typographical error, and that the term $N - 1/N$ was intended to be $(N-1)/N$.

It is not difficult to see that d_{TV} is a special case of d_W when the underlying metric is given by $d(x, z) = 1$ if $x \neq z$. In general however, convergence in d_W does not imply convergence in d_{TV}, and vice versa (see Madras and Sezer (2010) for examples where convergence fails, as well as some conditions under which convergence in one of d_W, d_{TV} implies convergence in the other). The purpose of this paper is to obtain a bound in d_{TV} by making use of (1.4) and simple properties of the Markov chain, without specifically engaging in a new study of the mixing time.

Let X_t be a copy of the Markov chain, and let μ^t be its probability distribution. Furthermore, define $\zeta_i := \left(\sigma^{-2}+n_i\gamma^2\right)^{-1}\left(\sigma^{-2}y_i+\gamma^2 n_{\max}\right)$, $\zeta := \max\{|\zeta_i|\}$ and $\tilde{\sigma}_i^2 = \left(\sigma^{-2}+n_i\gamma^2\right)^{-1}$. If π is the posterior distribution with density function $\pi_{posterior}$, we show that