Computational Statistics and Data Analysis 91 (2015) 4–19

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Convergent stochastic Expectation Maximization algorithm with efficient sampling in high dimension. Application to deformable template model estimation

Stéphanie Allassonnière [a], Estelle Kuhn [b,*]

[a] CMAP Ecole Polytechnique, Route de Saclay, 91128 Palaiseau, France
[b] MIA INRA, Domaine de Vilvert, 78352 Jouy-en-Josas, France

A R T I C L E   I N F O

Article history:
Received 23 May 2014
Received in revised form 20 April 2015
Accepted 21 April 2015
Available online 23 May 2015

Keywords:
Deformable template
Geometric variability
Maximum likelihood estimation
Missing variable
High dimension
Stochastic EM algorithm
MCMC
Anisotropic MALA

A B S T R A C T

Estimation in the deformable template model is a big challenge in image analysis. The issue is to estimate an atlas of a population. This atlas contains a template and the corresponding geometrical variability of the observed shapes. The goal is to propose an accurate estimation algorithm with low computational cost and with theoretical guaranties of relevance. This becomes very demanding when dealing with high dimensional data, which is particularly the case of medical images. The use of an optimized Monte Carlo Markov Chain method for a stochastic Expectation Maximization algorithm, is proposed to estimate the model parameters by maximizing the likelihood. A new Anisotropic Metropolis Adjusted Langevin Algorithm is used as transition in the MCMC method. First it is proven that this new sampler leads to a geometrically uniformly ergodic Markov chain. Furthermore, it is proven also that under mild conditions, the estimated parameters converge almost surely and are asymptotically Gaussian distributed. The methodology developed is then tested on handwritten digits and some 2D and 3D medical images for the deformable model estimation. More widely, the proposed algorithm can be used for a large range of models in many fields of applications such as pharmacology or genetic. The technical proofs are detailed in an appendix.[1]

1. Introduction

We consider here the deformable template model introduced for Computational Anatomy in Grenander and Miller (1998). This model, which has demonstrated great impact in image analysis, was developed and analyzed later on by many groups (among other Miller et al., 2002, Marsland and Twining, 2004, Vercauteren et al., 2009 and Su et al., 2013). It offers several major advantages. First, it enables to describe the population of interest by a digital anatomical template. It also captures the geometric variability of the population shapes through the modeling of deformations of the template which match it to the observations. Moreover, the metric on the space of deformations is specified in the model as a quantification of the deformation cost. This generative model not only describes the population, but also allows to sample synthetic data, using both the template and the geometrical metric of the deformation space which together define the atlas. Nevertheless, the key statistical issue is how to estimate efficiently and accurately these parameters of the deformable template model from an observed population of images.

* Corresponding author. Tel.: +33 134652701; fax: +33 34652217.
E-mail address: estelle.kuhn@jouy.inra.fr (E. Kuhn).

[1] The appendix is available as supplementary material (see Appendix A).

http://dx.doi.org/10.1016/j.csda.2015.04.011

Several numerical methods have been developed mainly for the estimation of the template image (for example Cootes et al., 1995 and Joshi et al., 2004). Even if these methods lead to visual interesting results on some training samples, they suffer from a lack of theoretical properties raising the question of the relevance of the output. Moreover they are not robust to noisy data. Another important contribution towards the statistical formulation of the template estimation issue was proposed in Glasbey and Mardia (2001). However interesting this approach, it is not entirely satisfactory since the deformations are applied to discrete observations requiring some interpolation. Moreover it does not formulate the analysis in terms of a generative model which appears very attractive as mentioned above. To overcome these lacks, a coherent statistical generative model was formulated in Allassonnière et al. (2007). For estimating all the model parameters, the template image together with the geometrical metric, the authors proposed a deterministic algorithm based on an approximation of the well-known Expectation Maximization (EM) algorithm (see Dempster et al., 1977), where the conditional distribution is replaced by a Dirac measure on its mode (called FAM-EM). However, such an approximation leads to the non-convergence of the estimates, which is highlighted when considering noisy observations.

One solution to face this problem is to consider a stochastic approximation of the EM (SAEM) algorithm which was proposed and proved to converge in Delyon et al. (1999). An extension using Monte Carlo Markov Chain (MCMC) methods was developed and studied in Kuhn and Lavielle (2004, 2005) and Allassonnière et al. (2010b), allowing for wider applications. To apply this extension to the deformable template model, the authors in Allassonnière et al. (2010b) chose a Metropolis Hastings within Gibbs sampler (also called hybrid Gibbs) as MCMC method since the variables to sample were of large dimension (the usual Metropolis Hastings algorithm providing low acceptation rates). This estimation algorithm has been proved to converge in Allassonnière et al. (2010b). Moreover it performs very well on very different kind of data as presented in Allassonnière et al. (2010a). Nevertheless, the hybrid Gibbs sampler becomes computationally very expensive when sampling very high dimensional variables. Although it reduces the dimension of the sampling to one, which enables to stride easier the target density support, it loops over the sampling variable coordinates, which becomes computationally unusable as soon as the dimension is very large or as the acceptation ratio involves heavy computations. To overcome the problem of computational cost of this estimation algorithm, some authors propose to simplify the statistical model constraining the correlations of the deformations (see Richard et al., 2009 and Maire et al., 2011).

Our purpose in this paper is to propose an efficient and convergent estimation algorithm for the deformable template model in high dimension without any constrains. With regard to the above considerations, the computational cost of the estimation algorithm can be reduced by optimizing the sampling scheme in the MCMC method.

The sampling of high dimensional variables is a well-known difficult challenge. In particular, many authors have proposed to use the Metropolis Adjusted Langevin Algorithm (MALA) (see Roberts and Tweedie, 1996 and Stramer and Tweedie, 1999a). This algorithm is a particular random walk Metropolis Hastings sampler. Starting from the current iterate of the Markov chain, one simulates a candidate with respect to a Gaussian proposal with an expectation equal to the sum of this current iterate and a drift related to the target distribution. The covariance matrix is diagonal and isotropic. This candidate is accepted or rejected with a probability given by the Metropolis Hastings acceptance ratio.

Some modifications have been proposed, in particular to optimize the covariance matrix of the proposal in order to better stride the support of the target distribution (see Stramer and Tweedie, 1999b, Atchadé, 2006, Marshall and Roberts, 2012 and Girolami and Calderhead, 2011). In Atchadé (2006) and Marshall and Roberts (2012), the authors proposed to construct adaptive MALA chains for which they prove the geometric ergodicity of the chain uniformly on any compact subset of its parameters. Unfortunately, this technique does not take the whole advantage of changing the proposal using the target distribution. In particular, the covariance matrix of the proposal is given by a stochastic approximation of the empirical covariance matrix. This choice seems completely relevant as soon as the convergence towards the stationary distribution is reached. However, it does not provide a good guess of the variability during the first iterations of the chain since it is still very dependent on the initialization. This leads to chains that may be numerically trapped. Moreover, this particular algorithm may require a lot of tuning parameters. Although the theoretical convergence is proved, this algorithm may be very difficult to optimize in practice *into* an estimation process.

Recently, the authors in Girolami and Calderhead (2011) proposed the Riemann manifold Langevin algorithm in order to sample from a target density in high dimensional setting with strong correlations. This algorithm is also a MALA based one for which the choice of the proposal covariance is guided by the metric of the underlying Riemann manifold. The quantities required to implement this method are not tractable in the deformable template model. Note that this is common with other application fields such as genetic or pharmacology, where models are often complex.

For these reasons, we propose to adapt the MALA algorithm in the spirit of both works in Atchadé (2006) and Girolami and Calderhead (2011) to get an efficient sampler into the stochastic EM algorithm. Therefore, we propose to sample from a proposal distribution which has the same expectation as the MALA but using a full anisotropic covariance matrix based on the anisotropy and correlations of the target distribution. This sampler will be called AMALA in the sequel. The expectation is obtained as the sum of the current iterate plus a drift which is proportional to the gradient of the logarithm of the target distribution. We construct the covariance matrix as a regularization of the Gram matrix of this drift. We prove the geometric ergodicity uniformly on any compact set of the AMALA assuming some regularity conditions on the target distribution. We also prove the almost sure convergence of the parameter estimated sequence generated by the coupling of AMALA and SAEM algorithms (AMALA–SAEM) towards the maximum likelihood estimate under some regularity assumptions on the model. Moreover, we prove a Central Limit Theorem for this sequence under usual conditions on the model.