Abstract

The integrated nested Laplace approximation (INLA) [Rue et al., 2009] is now a well-known functional approximation algorithm for implementing Bayesian inference in latent Gaussian models but has some limitations: it is unable to handle a high dimensional model parameter $\theta$, and makes a poor approximation when the posterior is multi-modal and the likelihood is highly non-Gaussian. Two types of algorithm are proposed to address these limitations: (a) a combination of INLA and Monte Carlo methods and (b) analytic approximations with higher order moments. We test the performance of algorithms on a non-linear stochastic velocity model for exchange current data between Euros and Dollars and a factor analysis model with both synthetic datasets and real data from multi-spectral extra-terrestrial microwave maps.

1 Introduction

Latent Gaussian models, a subclass of structured additive regression models, have a wide range of application domains in statistics, signal processing and machine learning. Examples are regression with an additive mixed linear model [Fahrmeir and Lang, 1999; McCulloch et al., 2008] and random walk model [Rue and Held, 2005], dynamic linear models [West and Harrison, 1997], spatial and spatio-temporal models [Besag et al., 1991; Best et al., 2005; Brix and Diggle, 2001].

The integrated nested Laplace approximation (INLA) [Rue et al., 2009] is a fast and accurate functional approximation algorithm for implementing Bayesian inference with such models when compared to conventional Monte Carlo simulation.

The contribution of this paper is to propose several variants of Laplace Approximation (LA) to overcome some of limitations of INLA: inability to deal with high dimensionality of the model parameters, posterior multi-modality and highly skewed likelihoods. This is illustrated through application to both a stochastic velocity model and a factor analysis model.

2 Statistical Background

The latent Gaussian model has the following structure for observations $y$ in terms of latent variables $x$ and hyperparameters $\theta$: $p(y \mid x, \theta) = \prod_j p(y_j \mid x_j, \theta)$ (so observations are conditionally independent), $p(x \mid \theta)$ is a Gaussian Markov random field (GMRF) with precision matrix $Q$, or at least a Gaussian with sparse precision matrix, and there is a some prior $p(\theta)$ on the model parameters.

A sparse precision matrix as the prior of the latent Gaussian model speeds up computation, of which the most popular is through the Gaussian Markov random field (GMRF) model [Weir, 1997; Hanson and Wecksung, 1983; Hunt, 1977; Therrien, 1993].

2.1 Integrated Nested Laplace Approximation

The integrated nested Laplace approximation (INLA) [Rue et al., 2009] approximates the marginal posterior $p(x \mid y)$ by

$$p(x \mid y) = \int p(x \mid y, \theta)p(\theta \mid y)d\theta \approx \int \tilde{p}(x \mid y, \theta)d\theta \approx \sum_{\theta_j} \tilde{p}(x \mid y, \theta_j)\Delta_{\theta_j}$$

(1)

where

$$\tilde{p}(\theta \mid y) = \frac{p(x, y, \theta)}{p_G(x, y, \theta)} \bigg|_{x = x^*(\theta)}$$.

(2)

Here, $p_G(x \mid y, \theta)$ denotes a Gaussian approximation and $x^*(\theta)$ is its mode. The mode $x^*(\theta)$ is typically calculated by a numerical optimization of the log posterior.

A set of discrete values $\theta_j$ is defined over which $\tilde{p}(\theta \mid y)$ is computed by first finding a mode $\mu^*_\theta$ of $\log \tilde{p}(\theta \mid y)$ by a quasi-Newton optimization and computing the Hessian $H_\theta$ at that mode. A grid search is then conducted from the mode in all directions until $\log \tilde{p}(\mu^*_\theta) - \log \tilde{p}(\theta_j) > \delta_\tau$ where $\delta_\tau$ is a given threshold. This gives a region over which a grid of $\theta_j$ values may be defined.

3 Proposed algorithms

Problems with INLA occur if the dimension of $\theta$ is even modestly large e.g. much above 5 because the grid over $\theta$ space becomes too large. There are also problems if $p(\theta \mid y)$
or \( p(x|y, \theta) \) are multi-modal or severely skewed (i.e. far from Gaussian). In order to address the above problems, we propose two classes of approach. The first addresses a high dimension \( \theta \) through a Monte Carlo approach to generate samples of \( \theta \) but not \( x \). The other approach is still a functional approximation, where higher order moments of the function are used. In the multi-modal case, the Gaussian approximation results in large error in the approximated distribution, i.e. given a reasonable threshold \( \delta \), \( |p(x'|y, \theta) - p_G(x'|y, \theta)| > \delta \). Several variant LA algorithms are proposed in Algorithm 1.

### 3.1 IS-LA

The first approach is a hybrid approach between importance sampling and LA, named IS-LA. It uses:

\[
p(x|y) = \int p(x|y, \theta)p(\theta|y)d\theta \\
\approx \int p(x|y, \theta)p(\theta|y)d\theta \\
\approx \int p(x|y, \theta)\tilde{p}(\theta|y)q(\theta|y)d\theta \\
\approx \frac{1}{\sum_i w_i} \sum_i p(x|y, \theta_i)w_i,
\]

where \( \theta_i \sim q(\theta|y) \) and its weights is defined as

\[
w_i = \frac{\tilde{p}(\theta_i|y)}{q(\theta_i|y)} = \frac{p(y|x, \theta_i)p(x|\theta_i)}{p(y|x|\theta_i)}|_{x=x^*(\theta_i)}.
\]

The proposal function \( q(\theta_i|y) \) can be defined in different ways.

#### The prior distribution (IS-LA\(^{(1)}\))

The proposal distribution is the prior distribution: \( q(\theta|y) = p(\theta) \). In general importance sampling has a serious problem in obtaining incorrect weights in the tail area when the tail of the proposal distribution is too small. This problem is avoided since the terms are analytically cancelled off as

\[
w_i = \frac{p(y|x, \theta)p(x|\theta)}{p_G(x|y, \theta)}|_{x=x^*(\theta)}.
\]

In addition, using the prior as the proposal distribution does not require Newton-style optimization.

#### An approximation to the posterior distribution (IS-LA\(^{(2)}\))

Although the prior distribution has many benefits, it may not be close to the posterior. It may be better to use a proposal that depends on the data; ideally \( q(\theta|y) = p(\theta|y) \). However, it is rather difficult to obtain a reasonable \( p(\theta|y) \) if \( p(y|x, \theta) \) is either non-linear or non-Gaussian. A Laplace approximation can be used: \( q(\theta|y) = p_G(\theta|y) \) via the quasi-Newton approach of Section 2.1. Although this approach works efficiently, it can suffer from incorrect weights when values are sampled in the tail.

### 3.2 MH-LA

Another hybrid approach combines LA and the Metropolis-Hastings (MH) algorithm. Samples are generated from \( p(\theta|y) \) using the Laplace approximation as the MH proposal distribution. The accept probability is \( A = \min \left( 1, \frac{\tilde{p}(\theta'|y)q(\theta'|y)}{\tilde{p}(\theta|y)q(\theta|y)} \right) \) where \( \theta \) and \( \theta' \) denote current and new samples respectively, and \( \tilde{p}(\theta'|y) = \frac{p(y|x, \theta')}{p(y|x, \theta)} |_{x=x^*(\theta')} \).

There are several possibilities for the proposal function, as below.

#### A random walk

One of the simplest proposal functions is a random walk \( q(\theta'|\theta) = N(\theta'; \theta, \Sigma_\theta) \) where \( \theta \) and \( \theta' \) denote the previous and current samples respectively. In addition, \( \Sigma_\theta \) can be chosen either manually or systematically.

#### An approximation to the posterior distribution

The proposal function used in IS-LA can be used in MH-LA. For instance, the optimal proposal function is designed as \( q(\theta'|\theta) = p_G(\theta'|y) \), as by the quasi-Newton approach of Section 2.1.

### 3.3 Marginalisation based LA (MLA)

Multi-modality and high skewness cannot be dealt with by these hybrid approaches since the proposal function itself depends on the Laplace approximation. A functional approximation based on the marginal likelihood with higher order moments can be used to address the problem. The marginal posterior of \( \theta \) has the form:

\[
p(\theta|y) \propto p(\theta) \int_x \exp \{ \log p(y|x, \theta) + \log p(x|\theta) \} \, dx \\
= p(\theta) \int_x \exp \{ R(x; \theta) \} \, dx \\
= \frac{p(\theta)}{|Q_\theta^*(\theta)|^{1/2}} \exp \left\{ \tilde{f}(x; \theta) \right\}
\]

where \( \mu_\theta^*(\theta) \) and \( Q_\theta^*(\theta) \) are a mode and the negative Hessian matrix of \( R(x; \theta) \). And the \( \tilde{f}(x; \theta) \) is as following:

\[
R(x; \theta) = -\frac{1}{2} (x - \mu_\theta^*(\theta))^T Q_\theta^*(\theta) (x - \mu_\theta^*(\theta)) + f(x; \theta) \\
\approx -\frac{1}{2} (x - \mu_\theta^*(\theta))^T Q_\theta^*(\theta) (x - \mu_\theta^*(\theta)) + \tilde{f}(x; \theta).
\]

Here, it is not straightforward to obtain the approximation \( \tilde{f}(x; \theta) \) so that we propose an approximation to the mean estimate of \( E_{p(x)}(\tilde{f}(x)) \):

\[
\tilde{f}(x; \theta) = E_{p(x)}(f(x; \theta)) = \int f(x; \theta) p(x) \, dx \\
= \frac{1}{2} \sum_{x' \in S} \exp(R(x'^{(i)}; \theta)) \int R(x'^{(i)}; \theta) \\
+ \frac{1}{2} Q_\theta^*(\theta) (x'^{(i)} - \mu_\theta^*(\theta))^T 
\]

where \( S = \{ (x^j; \mu_\theta^*(\theta) + j \sqrt{Q_\theta^*(\theta)} 1^{1/2}, j = -2, -1, 0, 1, 2 \} \) and \( Q_\theta^*(\theta) = (\sqrt{Q_\theta^*(\theta)} 1^{1/2})^{-1} \) and 1 is a
vector with ones for simplicity. $\mathbf{V}$ and $\Lambda$ are obtained by eigen-decomposition. Note that if we use $K = 1$ the proposed MLA becomes exactly the conventional INLA since $\tilde{f}(x; \theta) = R(\mu_0^*(\theta); \theta)$. In this point of view, MLA is a super-position of INLA with varying $K$.

The mode $\mu_0^*$ and its negative Hessian matrix $H_0^*$ of $\tilde{p}(\theta|y)$ with respect to $\theta$ are found by a quasi-Newton method. A greedy search is used in the same manner as conventional INLA to obtain a discrete grid of points $\theta_i$ over which $\tilde{p}(\theta|y)$ is evaluated. Of course, if the model is linear and Gaussian, $p(y|\theta)$ is in closed form and Eq. (2) is unnecessary.

### 3.4 Multiple initial points

An alternative approach to the multiple mode problem is to repeat the search for modes of $p(x|y, \theta)$ from different initial points. Each search yields a mode and Hessian. Of course there is no guarantee that all modes will be discovered, even if the number of them is known.

### 4 Stochastic velocity model

The financial time series are often modelled by stochastic velocity models that are highly nonlinear and non-Gaussian. We tested our proposed algorithm on two different datasets: one from artificial data plotted in Fig. 1-(a) and the other from Euro-dollar exchange currency rate in Fig. 1-(b). The artificial dataset were drawn from the Eq. (5) and the real exchange currency rate dataset is obtained between 1st September 2010 and 20th January 2011. In this manuscript, we set the ground truth for the synthetic observations as $h = 0$, $\phi = 0.4$, and $\tau = 2.2$.

$$
\begin{align*}
  y_1|\eta_1 & \sim \mathcal{N}(0, \exp(\eta_1)) \\
  \eta_1 & = h + f_t \\
  f_t|\mu_{t-1}, \tau, \phi & \sim \sum_{i=0}^{1} \frac{1}{2} \mathcal{N}(\mu_i, 1/\tau)
\end{align*}
$$

where $\mu_i = \phi^{1-i}(1-\phi)^i \mu_{t-1} + 3i$ for a mixture model, $x_{1:T} = (\eta_1, \mu, \theta)$ and $\theta = (\phi, \tau)$. We used the following priors: $h \sim \mathcal{N}(0, 1)$, $\phi \sim B(2, 2)$ and $\tau \sim \mathcal{G}(1, 0.1)$ where $\mathcal{N}$, $\mathcal{B}$ and $\mathcal{G}$ denote the normal, beta and gamma distributions respectively. Two top sub-graphs of Fig. 2 demonstrate the reconstructed distributions of $\tilde{p}(\theta|y)$ of the synthetic dataset via INLA and MLA. As we can see in the figures, the approximated distribution via MLA is rather different from that via INLA since the $p(x|y, \theta)$ has multimodality and skewness due to the mixture model.

In order to monitor the performance at the unimodal case, the mixture model of Eq. (5) is replaced to a unimodal model for the real dataset, i.e., $\mu_i = \phi \mu_{t-1}$ for all is. The results of the real exchange currency rates are also processed in Fig. 2-(c, d). It founds that MLA and INLA have the identical results from this figure given the unimodal model.

### 5 Application to multi-spectral image source separation

In the multi-spectral source separation problem, the data consists of $n_f$ images of $J$ pixels, that usually correspond to images at different frequencies $v_1, \ldots, v_n$. The data at pixel $j$ are denoted $y_j \in \mathbb{R}^{n_v}$, $j = 1, 2, \ldots, J$, while $Y_k = (y_{1,k}, \ldots, y_{J,k})^T$ denotes the image at frequency $v_k$. The observed images are believed to be built up of linear combinations of $n_s$ sources, represented by the vectors $x_j \in \mathbb{R}^{n_s}$. We assume that the $y_j$ follow a standard statistical independent components analysis model, so that they can be represented as a linear combination of the $x_j$, such that $y_j = Ax_j + e_j$ where $A$ is an $n_f \times n_s$ mixing matrix and $e_j$ is a vector of $n_f$ independent Gaussian error

---

**Algorithm 1 Variant Laplace Approximations**

1. Choose an approach, $\text{type} \in \{\text{INLA, IS-LA, MH-LA, MLA}\}$.

2. Obtain a mode and its negative hessian matrix by a quasi newton approach for $\tilde{p}(\theta|y)$.

3. $$(\mu_0^*, H_0^*) = \arg \max_{\theta} \log \frac{p(y|x, \theta)}{p(y|x, \theta)} |\mathbf{x} = x^{\star}(\theta)$$

   where $x^{\star}(\theta)$ be a mode of $p_C(y|x, \theta)$.

4. else

   5. Calculate $\mu_0^*(\theta)$ and $Q_0^*(\theta)$ be a mode and Hessian matrix of $\log p(y|x, \theta)$.

   6. Obtain $V$ and $\Lambda$ by eigendecomposition of $Q_0^*(\theta)$.

   7. Build a set $S$ with $K$ elements by using $\mu_0^*(\theta)$ and $Q_0^*(\theta)$, i.e. $x^{(j)} = \mu_0^*(\theta) + jVA^{1/2}1$ for $j = -K^{-1}, \ldots, 0, \ldots, K^{-1}$.

   8. $$(\mu_0^*, H_0^*) = \arg \max_{\theta} \log \frac{p(\theta)}{Q_0^*(\theta)^{1/2}} \exp \left\{ \tilde{f}(x; \theta) \right\}$$

   where $f(x; \theta) = E_{x} f((x; \theta))$ with $S$ from Eq. (4).

9. end if

   Obtain $\theta_s$ given the following strategies:

10. if type is INLA then

11. After finding $(\mu_0^*, H_0^*)$, do a grid search from the mode in all directions until the log of $p(\mu_0^*|y) - \log \tilde{p}(\theta|y) > \delta$ where $\delta$ is a given threshold.

12. else if IS-LA then

13. Draw samples from the optimal proposal function, $q(\theta|y) = p(\theta|\mu_0^*, H_0^{-1})$.

14. Calculate the weights by

   $$
   w_i = \frac{\tilde{p}(\theta^{(i)}|y)}{q(\theta^{(i)}|y)} = \frac{p(y|x, \theta^{(i)})}{p(y|x, \theta^{(i)}|\mu_0^*, H_0^{-1})}.
   $$

15. else if MH-LA then

16. Draw samples from the proposal function such as optimal proposal function $q(\theta|y) = p(\theta|\mu_0^*, H_0^{-1})$.

17. Calculate the acceptance ratio by

   $$
   \min \left\{1, \frac{\tilde{p}(\theta^{(i)}|y)}{\tilde{p}(\theta^{(i)}|y)} \right\}.
   $$

18. else if MLA then

19. After finding $(\mu_0^*, H_0^*)$, do a grid search from the mode in all directions until the log of $p(\mu_0^*|y) - \log \tilde{p}(\theta|y) > \delta$ where $\delta$ is a given threshold.

20. end if

21. Estimate $p(x|y) = \sum_{\theta} p(x|y, \theta_s)\tilde{p}(\theta_s|y)\Delta \theta_s$. 

---
with the latent variables; see precision matrix is not of full rank; this is the intrinsic GMRF with precision \( \mathbf{Q}^* = \mathbf{Q} + \mathbf{B}^T \mathbf{CB} \) and mean \( \mathbf{\mu}^* = \mathbf{Q}^{-1} \mathbf{B}^T \mathbf{C} \mathbf{Y} \). Hence the marginalized likelihood is

\[
p(y|\theta) = \frac{1}{2\pi} \left| \mathbf{Q} \right|^{1/2} \exp \left\{ -\frac{1}{2} (\mathbf{\mu}^T \mathbf{Q}^{-1} \mathbf{\mu}^* - y^T \mathbf{C} \mathbf{Y}) \right\}.
\]

5.1 Simulated data example

The first example is a set of 9 images of size 16 × 16 constructed from 3 synthetic sources.

![Observed signals for the synthetic data](image)

The extracted sources obtained by the different algorithms, and other common separation approaches, are plotted in Fig. 4. Except for fastICA, the ground truth was well recovered in most approaches. Tables 1 and 2 demonstrate the performance based on a measure of goodness of fit Peak Signal-to-Inference Ratio (PSIR) and running time respectively. It can be seen that the proposed variant LAs have similar performance to MCMC in accuracy while they are considerably faster. In addition, we can check the linearity of the model by checking the similarity between results from INLA and MLA, which should give an identical result in that case; it is seen that that happens for our linear model example.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>PSIR</th>
<th>Running Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS</td>
<td>28.13 ± 6.26</td>
<td>45.10 ± 6.66</td>
</tr>
<tr>
<td>FastICA</td>
<td>10.95 ± 18.4</td>
<td>4.780 ± 95.3</td>
</tr>
<tr>
<td>SMICA</td>
<td>16.33 ± 130</td>
<td>15.11 ± 136</td>
</tr>
<tr>
<td>MCMC</td>
<td>41.47 ± 18.0</td>
<td>45.17 ± 7.51</td>
</tr>
<tr>
<td>INLA(1)</td>
<td>40.87 ± 16.3</td>
<td>44.90 ± 6.25</td>
</tr>
<tr>
<td>INLA(2)</td>
<td>40.93 ± 17.2</td>
<td>44.90 ± 6.23</td>
</tr>
<tr>
<td>IS-LA(1)</td>
<td>41.19 ± 16.4</td>
<td>44.94 ± 6.29</td>
</tr>
<tr>
<td>IS-LA(2)</td>
<td>41.11 ± 15.1</td>
<td>44.93 ± 6.37</td>
</tr>
<tr>
<td>MH-LA</td>
<td>41.13 ± 16.1</td>
<td>44.92 ± 6.27</td>
</tr>
<tr>
<td>MLA</td>
<td>41.77 ± 12.3</td>
<td>45.05 ± 6.49</td>
</tr>
</tbody>
</table>

Table 1: PSIR comparison for the synthetic data: INLA(1) and INLA(2) used GF and GMRF priors respectively. Also, LS is acronym of (General) Least Square.

5.2 Example of separation of the cosmic microwave background from all-sky microwave maps

The discovery of the cosmic microwave background (CMB) is strong evidence for the big bang theory of the formation
and development of the universe. According to the theory, the early universe was smaller and hotter but cooled as it expanded. Once the temperature cooled to about 3000K, photons were free to propagate without being scattered off ionized matter; the CMB is an image of this event and is visible across the entire sky. Three satellites have been launched to measure the CMB: the cosmic background explorer (COBE), Wilkinson microwave anisotropy probe (WMAP) and most recently the Planck surveyor. Planck is the highest resolution data to date, of the order of $10^7$ pixels across the sky measured at 9 channels.

Unfortunately, the signals measured by these satellites as in Fig. 5 contain radiation not only from CMB but also contributions from a number of other sources, namely foreground radiations and extragalactic sources in addition to antenna receiver noise. Foreground sources from our galaxy include synchrotron, dust and free-free emission. Therefore, the separation of the CMB signal from other sources is an important stage in the production of CMB maps. [Kuruoglu, 2010].

![Image](image1.png)

**Figure 4:** Comparison of restored three signals for the synthetic data: GT (Ground Truth) and LS (Least Square)

<table>
<thead>
<tr>
<th>LS</th>
<th>fastICA</th>
<th>SMICA</th>
<th>MCMC</th>
<th>INLA$^{(1)}$</th>
<th>INLA$^{(2)}$</th>
<th>IS-LA$^{(2)}$</th>
<th>MH-LA</th>
<th>MLA</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00785</td>
<td>9.8226</td>
<td>0.6961</td>
<td>1729.2</td>
<td>2.3283</td>
<td>3.1743</td>
<td>23.824</td>
<td>26.836</td>
<td>14.634</td>
</tr>
<tr>
<td>±0.0315</td>
<td>±18.506</td>
<td>0.1081</td>
<td>±140.42</td>
<td>±0.7104</td>
<td>±0.6661</td>
<td>±3.4309</td>
<td>±5.25</td>
<td>±3.5447</td>
</tr>
</tbody>
</table>

Table 2: Run time comparison for the synthetic example

Mixing Matrix Structure
In this application, A is parameterized and denoted A$^\theta$. Each column of A$(\theta)$ is the contribution to the observation of a source at different frequencies, which is written as a function of the frequencies and $\theta$. These parameterizations are approximations that come from the current state of knowledge about how the sources are generated. Here, we merely state the parameterization that we are going to use, and refer to [Eriksen et al., 2007] for a more detailed exposition on the background to them. It is assumed that the CMB is the first source and therefore, it corresponds to the first column of A$(\theta)$. It is modelled as a blackbody at a temperature, and its contribution is a known constant at each frequency. The parameterization of the mixing matrix is given as

\[ A_{i1}(\theta) = \frac{g(v_i)}{g(v_1)}, \]
\[ A_{i2}(\theta) = \left( \frac{v_i}{v_1} \right)^{\kappa_s} \]
\[ A_{i3}(\theta) = \frac{\exp(\eta v_i / k_B T_1) - 1}{\exp(\eta v_1 / k_B T_1) - 1} \left( \frac{v_i}{v_1} \right)^{1+\kappa_d} \]

where $g(v_i) = \left( \frac{\eta v_i}{k_B T_0} \right)^2 \frac{\exp(\eta v_i / k_B T_0)}{(\exp(\eta v_i / k_B T_0) - 1)^2}$. $T_0 = 2.725$ is the average CMB temperature in Kelvin, $T_1 = 18.1$, $\eta$ is the Plank constant and $k_B$ is Boltzmann’s constant. The ratio $g(v_i)/g(v_1)$ is designed to ensure that A$_{i1}(\theta) = 1$ as we constrain the fourth row of A$(\theta)$ to be ones. There are two unknown model parameters for A, for synchrotron $\kappa_s \in \{\kappa_s : -3.0 \leq \kappa_s \leq -2.3\}$ and the spectral indexes for dust $\kappa_d \in \{\kappa_d : 1 \leq \kappa_d \leq 2\}$.

Prior
There are two types of hidden variables: sources x and model parameters $\theta$, which consists of $\kappa_s$, $\kappa_d$ and a precision parameter $\phi$ for the GMRF prior of each source. The prior for $\theta$ is designed as $p(\tau_1, \kappa_s, \kappa_d) = \prod_{i=1}^n p(\tau_i) p(\kappa_s) p(\kappa_d)$ where $p(\tau_i) = \mathcal{G}(\tau_i; \alpha_i, \beta_i)$, $p(\kappa_s) = \mathcal{U}(\kappa_s; \alpha_{\kappa_s}, \beta_{\kappa_s})$ and $p(\kappa_d) = \mathcal{U}(\kappa_d; \alpha_{\kappa_d}, \beta_{\kappa_d})$. Let $\mathcal{G}$ and $\mathcal{U}$ denote the Gamma and uniform distributions respectively.
distribution and the uniform distribution and $\alpha$ and $\beta$ are hyper-parameters which are fixed in this paper: we assign this hyper parameter to generate flat prior.

**Results**

We analysed the seven year WMAP 7 year data by using the MLA. WMAP data consist of 5 images of $J = 3 \times 2^{20} = 3,145,728$ pixels (see Fig. 5) which were divided into 6144 blocks of 512 pixels for the analysis (we used blocking technique for WMAP data). Assuming 3 sources, the processed images via MLA are as shown in Fig. 6.

![Figure 6: Estimated Sources of WMAP by MLA](image)

6 Conclusion

The well-known integrated nested Laplace approximation (INLA) has several limitations. In this paper, several variants of Laplace Approximation (LA) are proposed to tackle such serious limitation of the conventional INLA. In order to solve the high dimensionality over the model parameter space, Monte Carlo (MC) simulation are hybridized with LA. This approach still practically fast and efficient since MC draws samples only from model parameter space. The other approach based on approximated marginalized likelihood provides the potential solution for the multi-modality and skewness problem.

**References**


