Discussion of Girolami and Calderhead

Krzysztof Łatuszyński, Gareth O. Roberts, Alexandre Thiéry, Katarzyna Wolny
(University of Warwick)

We congratulate the authors for the exciting and thought provoking paper. They approach the fundamental problem of scaling MCMC algorithms by introducing a metric on probability distributions and altering the dynamics of the underlying SDE accordingly to obtain more efficient, state dependent proposals. In context of the MALA algorithm we note that any choice of $\sigma(\theta)$ in

$$d\theta(t) = \left(\frac{\sigma^2(\theta(t))}{2} \mathcal{L}(\theta(t))' + \sigma(\theta(t))\sigma(\theta(t))'\right) dt + \sigma(\theta(t)) db(t),$$

(1)

yields a Langevin diffusion with the correct stationary distribution $p(\theta)$. By choosing $\sigma^2(\theta) = |1/(\mathcal{L}(\theta))''|$ we arrive at their manifold MALA with the metric tensor $G(\theta)$ defined as the observed Fisher information matrix plus the negative Hessian of the log-prior (c.f. Section 4.2 and 5).

To have an insight into ergodicity and robustness of the above version of manifold MALA, we analyze it for the family of targets $p(\theta) \propto \exp\{-|\theta|^\beta\}$. It provides a good benchmark for MCMC algorithms: for the random walk Metropolis and for the standard MALA, it is well known for which values of $\beta$ (or $\beta$ and $\varepsilon$ respectively), geometrical ergodicity holds, and for which it fails, [1, 2]. This is summarized in Table 1 (see [1, 2] for details), together with the respective properties of manifold MALA.

These results indicate very impressive theoretical properties. It appears that for $\beta > 1$ manifold MALA can inherit the stability properties of the random walk Metropolis algorithm together with the speed up of MALA. Moreover, it is applicable for $\beta < 1$, where the other two algorithms fail. We did our calculations by establishing drift conditions and analyzing acceptance rates for $d = 1$, and expect to observe the same phenomenon also in higher dimensions. A full report will follow.

<table>
<thead>
<tr>
<th>algorithm</th>
<th>$0 &lt; \beta &lt; 1$</th>
<th>$\beta = 1$</th>
<th>$1 &lt; \beta &lt; 2$</th>
<th>$\beta = 2$</th>
<th>$2 &lt; \beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RWM</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>MALA</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y*</td>
<td>N</td>
</tr>
<tr>
<td>MMALA</td>
<td>Y*</td>
<td>x</td>
<td>Y*</td>
<td>Y*</td>
<td>Y*</td>
</tr>
</tbody>
</table>

Table 1. Geometric Ergodicity of random walk Metropolis (RWM), MALA and manifold MALA (MMALA) for target $\pi(\theta) \propto \exp\{-|\theta|^\beta\}$. N = geometric ergodicity fails, Y = geometric ergodicity holds, $Y^*$ = geometric ergodicity holds for $\varepsilon$ small enough, x = not applicable.

References
