

# MANIFOLD SAMPLING FOR MOMENT CLOSURE

BC

## 1. EXPECTED FISHER INFORMATION FOR MOMENT CLOSURE

We start by considering the likelihood induced by the moment closure approach using a Gaussian approximation. Following from Equation 7 in the Stats and Computing paper, the pseudo-likelihood follows as

$$\begin{aligned}
 \hat{L}(\boldsymbol{\theta}|\mathbf{Z}) &= \prod_{i=1}^{T-1} \hat{p}(\mathbf{Z}(t_{i+1})|\mathbf{Z}(t_i)) \\
 (1) \qquad \qquad &= \prod_{i=1}^{T-1} N_{\mathbf{Z}(t_{i+1})}(\varphi_i, \boldsymbol{\Sigma}_i)
 \end{aligned}$$

where

$$(2) \qquad \qquad \varphi_i = \hat{\mathbf{E}}(\mathbf{Z}(t_{i+1})|\mathbf{Z}(t_i))$$

$$(3) \qquad \qquad \boldsymbol{\Sigma}_i = \hat{\mathbf{V}}\text{ar}(\mathbf{Z}(t_{i+1})|\mathbf{Z}(t_i))$$

such that  $\varphi_i$  and  $\boldsymbol{\Sigma}_i$  are the mean and covariance, respectively, under the moment closure approximation. We now need to calculate the metric tensor for this pseudo-likelihood, which is given by the Expected Fisher Information. Using the Gaussian approximation above, the Expected Fisher Information is easy to calculate and takes a standard form; due to the Markov property it turns out to be the sum of Expected Fisher Informations for each  $\mathbf{Z}_{t_i}$ . For ease of notation, let us denote  $\hat{L}_i = \hat{p}(\mathbf{Z}(t_{i+1})|\mathbf{Z}(t_i))$ . The overall Expected Fisher Information follows as

$$\begin{aligned}
 \text{FI}_{\hat{L}} &= \mathbf{E}_{\hat{L}} \left( \frac{\partial \log \hat{L}}{\partial \boldsymbol{\theta}}^T \frac{\partial \log \hat{L}}{\partial \boldsymbol{\theta}} \right) \\
 (4) \qquad &= \sum_{i=1}^{T-1} \mathbf{E}_{\hat{L}_i} \left( \frac{\partial \log \hat{L}_i}{\partial \boldsymbol{\theta}}^T \frac{\partial \log \hat{L}_i}{\partial \boldsymbol{\theta}} \right) \\
 (5) \qquad &= \sum_{i=1}^{T-1} \text{FI}_{\hat{L}_i}
 \end{aligned}$$

The  $\hat{L}_i$ s are all Gaussian, and so the standard expressions for the Expected Fisher Information of multivariate Gaussians may be used for each component of the sum. In particular, the  $(m, n)$ th component of the  $i$ th FI is given by

$$(6) \quad \text{FI}_{\hat{L}_i(m,n)} = \frac{\partial \varphi_i}{\partial \theta_m}^T \Sigma_i^{-1} \frac{\partial \varphi_i}{\partial \theta_n} + \frac{1}{2} \text{tr} \left( \Sigma_i^{-1} \frac{\partial \Sigma_i}{\partial \theta_m} \Sigma_i^{-1} \frac{\partial \Sigma_i}{\partial \theta_n} \right)$$

## 2. SIMPLIFIED MMALA SAMPLER

We may construct a simple Metropolis-Hastings proposal mechanism that makes use of the local geometry around each point in the parameter space. We note that we are working with the log-likelihood and its associated geometry here. In particular, our proposal will have a mean based on the current parameters plus a small move in the direction of the natural gradient, and a covariance equal to the scaled inverse of the Expected Fisher Information at the current point.

The natural gradient is defined as the inverse of the Expected Fisher Information (the metric tensor) multiplied by the standard gradient of the log-likelihood. The mean of the proposal at  $\theta$  is therefore given by,

$$(7) \quad \mu(\theta) = \theta + \frac{\epsilon^2}{2} \text{FI}_{\hat{L}}^{-1}(\theta) \frac{\partial \hat{L}(\theta)}{\partial \theta}$$

and the covariance of the proposal is given by,

$$(8) \quad \Sigma(\theta) = \epsilon^2 \text{FI}_{\hat{L}}^{-1}(\theta)$$

The metric tensor, in this case given by the Expected Fisher Information, is usually represented by  $G$ , and the log-likelihood by  $L$ . Using this notation, the general simplified mMALA algorithm is given in full in Algorithm 1.

---

### Algorithm 1 Simplified Manifold MALA

---

- 1: Initialise current  $\theta$
  - 2: **for** IterationNum = 1 to NumSamples **do**
  - 3:   Sample  $\theta^{\text{new}} \sim p(\theta^{\text{new}}|\theta) = N(\mu(\theta, \epsilon), \Sigma(\theta, \epsilon))$ ,  
       where  $\mu(\theta, \epsilon) = \theta + \frac{\epsilon^2}{2} G^{-1}(\theta) \nabla_{\theta} L(\theta)$  and  $\Sigma(\theta, \epsilon) = \epsilon^2 G^{-1}(\theta)$
  - 4:   Calculate current log-likelihood  $L(\theta)$  and proposed log-likelihood  $L(\theta^{\text{new}})$
  - 5:   Calculate  $\log(p(\theta^{\text{new}}|\theta))$ ,  $\log(p(\theta|\theta^{\text{new}}))$ ,  $\log(\text{Prior}(\theta))$ ,  $\log(\text{Prior}(\theta^{\text{new}}))$
  - 6:   LogRatio =  $L(\theta^{\text{new}}) + \log(\text{Prior}(\theta^{\text{new}})) + \log(p(\theta|\theta^{\text{new}})) - L(\theta) - \log(\text{Prior}(\theta)) - \log(p(\theta^{\text{new}}|\theta))$
  - 7:   **if** LogRatio > 0 **or** LogRatio > log(rand) **then**
  - 8:     Set  $\theta = \theta^{\text{new}}$
  - 9:   **end if**
  - 10: **end for**
-