2.2 Algorithmic Efficiency

We first consider MCMC algorithmic efficiency, independent of any computational requirements. This measure of efficiency solely represents the best mixing, or equivalently the least autocorrelation, or the highest effective sample size, without consideration for the computational (time) requirements of generating a set of samples. After reviewing the definition of MCMC algorithmic efficiency which is based upon integrated autocorrelation time, we study the use of Ψ_{scalar} or Ψ_{block} for particular choices of $\mathcal{M}$, and quantify the effects on this measure of efficiency.

As in Roberts and Rosenthal (2001), we define MCMC algorithmic efficiency as the effective sample size divided by the chain length. This represents the rate of production of effectively independent samples per MCMC sample. The effective sample size (ESS) of an MCMC chain is defined as $\text{ESS} = N/\tau$, where N is the chain length and τ is the integrated autocorrelation time. For a scalar chain of samples $X_0, X_1, \ldots$, which is assumed to have converged to its stationary distribution, Straatsma, Berendsen, and Stam (1986) define the integrated autocorrelation time as $\tau = 1 + 2\sum_{i=1}^{\infty} \text{cor}(X_0, X_i)$. τ may be interpreted as the number of MCMC samples required, on average, for an independent sample to be drawn. Our measure of algorithmic efficiency is thus τ^{-1}, the number of effective samples per actual sample (Thompson, 2010). τ^{-1} also characterizes the speed at which expectations of arbitrary functions of the sample values approach their stationary values (Roberts and Sahu, 1997), and no less satisfies the natural intuition that larger values indicate better performance.

For MCMC algorithm Ψ acting on model $\mathcal{M}$ with parameters Θ, we define the algorithmic efficiency of each $\theta \in \Theta$ as $A(\Psi, \theta) = \tau^{-1}$, where τ is the integrated autocorrelation time of the samples of θ generated from repeated application of Ψ. Overloading notation, we define the algorithmic efficiency of MCMC algorithm Ψ as $A(\Psi) = \min_{\theta \in \Theta} A(\Psi, \theta)$. This definition is motivated by noting that often an MCMC produces seemingly good mixing of many model dimensions but poor mixing of just a few dimensions. In this case, the poorly mixing dimensions will limit the validity of the entire posterior sample (although this is not universally true of all model structures). Therefore, we take the conservative approach, and our general aim is to maximize the algorithmic efficiency for the parameter exhibiting the slowest mixing.