Statistics 5444: MCMC for Finite Mixture Models

This is the final HW assignment of the semester. Please turn in by December 19th @ 5pm. Write each problem up *very* neatly (LaTeX is preferred). Provide no more than a 5 page report (including graphics/derivations/etc).

Consider the p-dimensional Gaussian Mixture Model:

$$(x_i, z_i) \sim \sum_{k}^{K} \pi_k p(x|\mu_k, \Sigma_k, z = k), \qquad \text{for } i = 1, \dots, N, \tag{1}$$

where $\pi_k = Pr(z = k)$, and $p(x|\mu_k, \Sigma_k) = \frac{1}{(2\pi)^{p/2}|\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x-\mu_k)'\Sigma_k^{-1}(x-\mu_k)}$.

State which priors you *should* use for μ_k, $\Phi_k = \Sigma_k^{-1}$, z_i, for $k = 1, \dots, K$ and $i = 1, \dots, N$, respectively. Provide at least one reason for why you have selected the priors you're using.

Find the full conditionals from (1) for μ_k, $\Phi_k = \Sigma_k^{-1}$, z_i, for $k = 1, \dots, K$ and $i = 1, \dots, N$, respectively. Specifically find:

1. $\pi(\mu_k| - - - -)$, for $k = 1, \dots, K$,
2. $\pi(\Phi_k| - - - -)$, for $k = 1, \dots, K$,
3. $\pi(z_i|x_i, - - -)$, for $i = 1, \dots, N$.

For $p = 1$, and $K = 2$, with values chosen for μ_k, $\Phi_k = \Sigma_k^{-1}$, π_k, simulate $N = 100$ values of x_i and z_i.

Implement a Gibbs Sampler for finding the joint posterior distribution of the μ_k, $\Phi_k = \Sigma_k^{-1}$ (for $k \in 1, 2$). Show how each of the marginal posterior distributions overlay the true values of the μ_k, $\Phi_k = \Sigma_k^{-1}$ values you used to simulate the data. State your burn-in period (this will depend on your starting values).

For various settings of μs, Σs, and πs, show how well the Gibbs Sampler works under several higher dimensional ($p \geq 2$) settings, with $K \geq 3$. The parameter settings are for you to pick, however, your simulation study should be thorough. Explore how well the algorithm works as both p and N increase.