WEEK 2 SUMMARY

V. CHEN

1. Summary. Last week's experiments indicated that the original parameterization, treating u, τ, α as the variables, would not converge even after 100,000 iterations. First, we explain the different parameterization that we tried this week. We decided to try a different parameterization which treats ξ, τ, α as variables to sample, and we compare results between this and the old parameterization.

2. Learning τ and α with non-centered parameterization.

2.1. Algorithm. In the centered parameterization, the prior was $u|\tau, \alpha \sim N(0, C(\tau, \alpha))$, τ, α distributed independently and uniformly over intervals. We would take the posterior by Bayes' theorem to be

$$f(u, \tau, \alpha) \propto \exp\left(-\Phi(u) - \frac{1}{2}\langle u, C(\tau, \alpha)^{-1}u\rangle - \frac{1}{2}\log(\det(C(\tau, \alpha))) + \log(\pi_0(\tau, \alpha))\right)$$

In the new parameterization, we sample $\xi \sim N(0, I)$ and τ, α on uniform intervals as before. ξ is related to u by $T(\xi, \tau, \alpha) = u = \sum_{j=0}^{M}(\lambda_j + \tau^2)^{-\alpha/2}\xi_j q_j$. Recall that we are taking $M = N - 1$. The new joint posterior becomes:

$$g(\xi, \tau, \alpha) \propto \exp\left(-\Phi(T(\xi, \tau, \alpha)) - \frac{1}{2}\langle u, u\rangle + \log(\pi_0(\tau, \alpha))\right)$$

This algorithm works as follows:

Algorithm 1 Non-centered parameterization: sampling ξ, τ, α

Choose $\xi^{(0)} \in \mathbb{R}^N, \alpha^{(0)}, \tau^{(0)} > 0, \beta \in (0, 1]$ and $\epsilon_1, \epsilon_2 > 0$.
for $k = 0$ to S **do**
  Propose $\hat{\xi}^{(k)} = (1 - \beta^2)\xi^{(k)} + \beta\zeta^{(k)}$, $\zeta^{(k)} \sim N(0, I)$
  Make transition $\xi^{(k+1)} \to \hat{\xi}^{(k)}$ with probability

$$A(\xi^{(k)} \to \hat{\xi}^{(k)}) = \min\{1, \frac{g(\hat{\xi}^{(k)}, \tau^{(k)}, \alpha^{(k)})}{g(\xi^{(k)}, \tau^{(k)}, \alpha^{(k)})}\}$$

  Propose $\hat{\tau}^{(k)} = \tau^{(k)} + \epsilon_1\rho^{(k)}, \rho^{(k)} \sim N(0, I)$
  Make transition $\tau^{(k+1)} \to \hat{\tau}^{(k)}$ with probability

$$A(\tau^{(k)} \to \hat{\tau}^{(k)}) = \min\{1, \frac{g(\xi^{(k+1)}, \hat{\tau}^{(k)}, \alpha^{(k)})}{g(\xi^{(k+1)}, \tau^{(k)}, \alpha^{(k)})}\}$$

  Propose $\hat{\alpha}^{(k)} = \alpha^{(k)} + \epsilon_2\sigma^{(k)}, \sigma^{(k)} \sim N(0, I)$
  Make transition $\alpha^{(k+1)} \to \hat{\alpha}^{(k)}$ with probability

$$A(\alpha^{(k)} \to \hat{\alpha}^{(k)}) = \min\{1, \frac{g(\xi^{(k+1)}, \tau^{(k+1)}, \hat{\alpha}^{(k)})}{g(\xi^{(k+1)}, \tau^{(k+1)}, \alpha^{(k)})}\}$$

end for
return $\{T(\xi^{(k)}, \tau^{(k)}, \alpha^{(k)}), \tau^{(k)}, \alpha^k\}$

2.2. Simulation results. The relevant figures are Figure 1, Figure 2, Figure 3, Figure 4, Figure 5, Figure 6. The parameters used are as follows:

Iterations	100000
Burn in period	1000
Laplacian	self-tuning, unnormalized
β	0.1
γ	0.0001
Labeled +1	$280 - 290$
Labeled -1	$20 - 30$
ϵ_α	1
ϵ_τ	1
Initial τ	30
Initial α	5
τ range	$[0, 60]$
α range	$[0, 100]$
Average τ	3.74
Average α	63.5
Percent of correct classification	0.864407
Time elapsed	183.21 s

FIG. 1. *α acceptance probability*

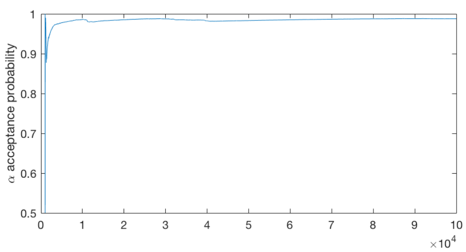


FIG. 2. *α trace*

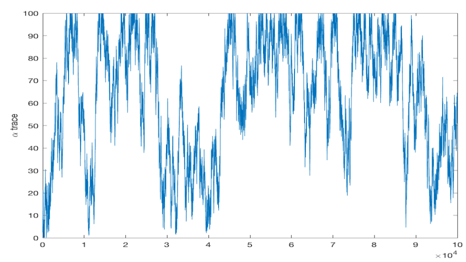


FIG. 3. *τ acceptance probability*

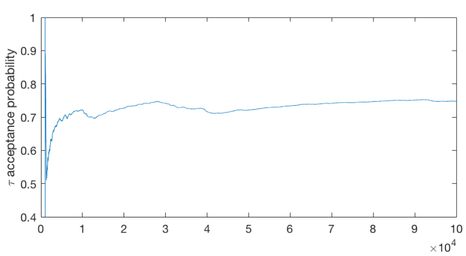


FIG. 4. *τ trace*

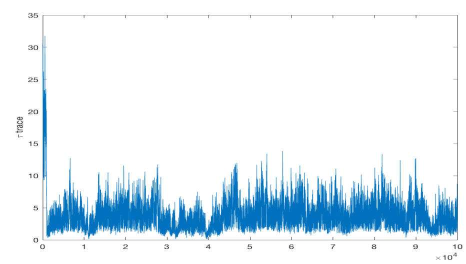


FIG. 5. *ξ acceptance probability*

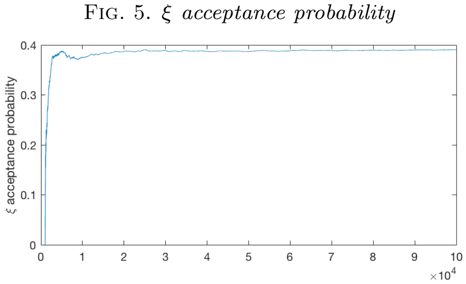
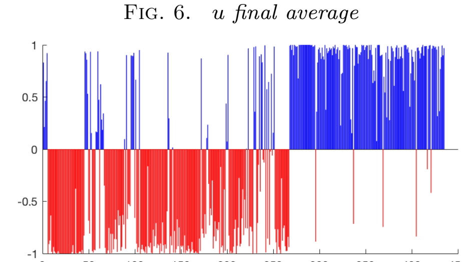


FIG. 6. *u final average*



3. Revisiting the centered parameterization.

4. Two moons data. The non-centered parameterization with the unnormalized Laplacian does not perform well on the two moons data set. One possibility is that the first eigenvector in the unnormalized Laplacian is constant and is throwing off the classification.
Initializing from $u = 0$ was also a bad idea because τ, α would blow up to their max limit.