COMPARING HIERARCHICAL METHODS

V. CHEN

1. Algorithm for hierarchical towards τ, α, M.

$$T(\xi,\tau,\alpha,M)=\sum_{i=1}^{M}\frac{1}{(\lambda_i+\tau^2)^{\alpha/2}}\xi_i q_i = u \tag{1}$$

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

Algorithm 1 Non-centered parameterization, hierarchical with τ, α, M

Choose $\xi^{(0)}\in\mathbb{R}^N, \tau^{(0)}, \alpha^{(0)}, M^{(0)}>0, \beta\in(0,1]$ and $\epsilon_1,\epsilon_2>0$.

for $k=0$ to S **do**

Propose $\hat{\xi}^{(k)}=(1-\beta^2)^{\frac{1}{2}}\xi^{(k)}+\beta\zeta^{(k)}$, $\zeta^{(k)}\sim\mathsf{N}(0,I)$

Make transition $\xi^{(k)}\to\hat{\xi}^{(k)}$ with probability

$$A(\xi^{(k)}\to\hat{\xi}^{(k)})=\min\left\{1,\exp\left(\Phi(T(\xi^{(k)},\tau^{(k)},\alpha^{(k)},M^{(k)}))-\Phi(T(\hat{\xi}^{(k)},\tau^{(k)},\alpha^{(k)},M^{(k)}))\right)\right\}$$

▷ T defined in (1)

Propose $\hat{\tau}^{(k)}=\tau^{(k)}+\epsilon_1\rho^{(k)}, \rho^{(k)}\sim\mathsf{N}(0,1)$

Make transition $\tau^{(k)}\to\hat{\tau}^{(k)}$ with probability

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

▷ g defined in (2)

Propose $\hat{\alpha}^{(k)}=\alpha^{(k)}+\epsilon_2\sigma^{(k)}, \sigma^{(k)}\sim\mathsf{N}(0,1)$

Make transition $\alpha^{(k)}\to\hat{\alpha}^{(k)}$ with probability

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

Propose $\hat{M}^{(k)}=M^{(k)}+Q$, with jump Q distributed as $\mathbb{P}(Q=k)\propto\frac{1}{1+|k|}$, $|Q|$ bounded.

Make transition $M^{(k)}\to\hat{M}^{(k)}$ with probability

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

end for

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

---

2. M and σ relation. We tested our algorithm on the two moons dataset with varying σ, fixing 3% labeled nodes. We also fixed $\tau=1,\alpha=35$ by setting $\epsilon_1=\epsilon_2=0$, so the algorithm is only learning ξ and M. We initialized $M^{(0)}=30$. We can see that for small $\sigma=0.02$, M is very small as only the first few eigenvectors are necessary (see Figure 1). However, when we choose a larger $\sigma=0.2$, M needs to be larger (see Figure 2). For $\sigma=0.02$, the classification accuracy was 100%. For $\sigma=0.2$, the classification accuracy was 90.21%.

3. Nonhierarchical vs. Hierarchical. We will compare pCN with fixed τ and α against Algorithm 1, using the same fixed τ and α and learning M.

FIG. 1. $\sigma = 0.02$, *trace of* M

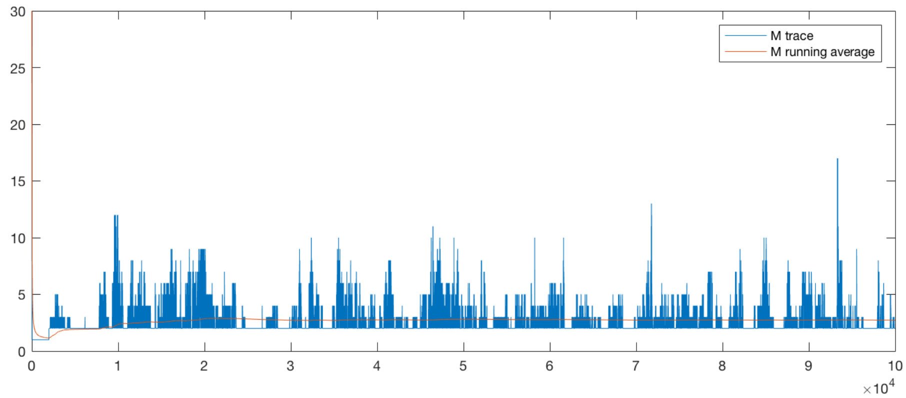
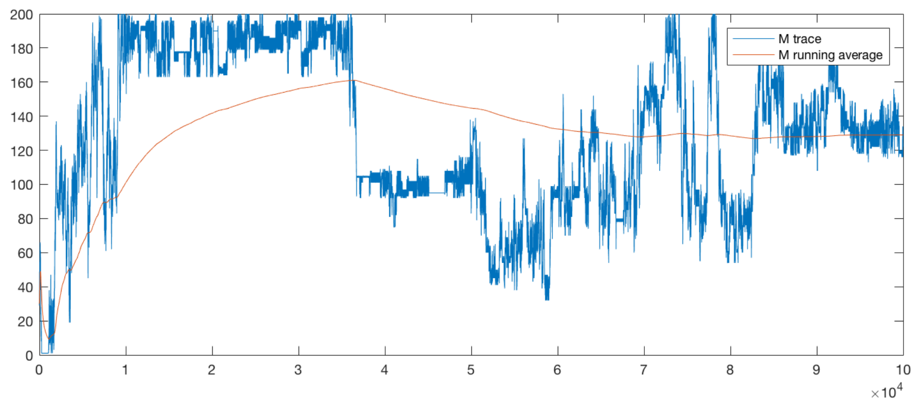


FIG. 2. $\sigma = 0.2$, *trace of* M



3.1. Voting records. The nonhierarchical algorithm for learning τ and α gave expected values of τ and α as $\tau \approx 2$ and $\alpha \approx 35$, so I chose those parameters for the pCN and for Algorithm 1. For the pCN, all 435 eigenvectors were used. I narrowed the range of M allowed in Algorithm 1 to 1 to 70. 5 labeled nodes were selected, consistent across the two methods. The results are not very concrete yet. Tentatively, it seems that being hierarchical does have some benefits, but more experiments are needed. In this particular realization (seeded with rng(5)), the hierarchical algorithm achieves 90.70% while the nonhierarchical algorithm achieves 87.67% classification accuracy. Figure 3, Figure 4, Figure 5, Figure 6, Figure 7, Figure 8, Figure 9 show the results of this experiment. Notice in Figure 6 that there seems to be an important eigenvector for classification indexed around 30, as M seldom drops below 30. Looking at the eigenvectors around index 30, I plotted the 34th eigenvector in Figure 10. From a purely visual perspective, it does appear that this eigenvector corresponds decently to the final classification in Figure 8 by looking at where some of the "spikes" are.

FIG. 3. *Average of* $S(u)$

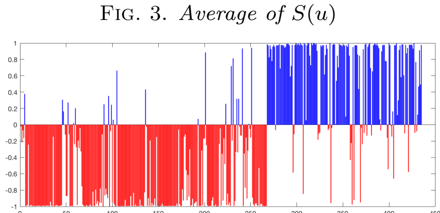


FIG. 4. u *acceptance probability*

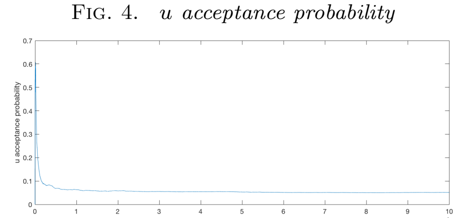


FIG. 5. *Running average of $S(u(i))$ for select i*

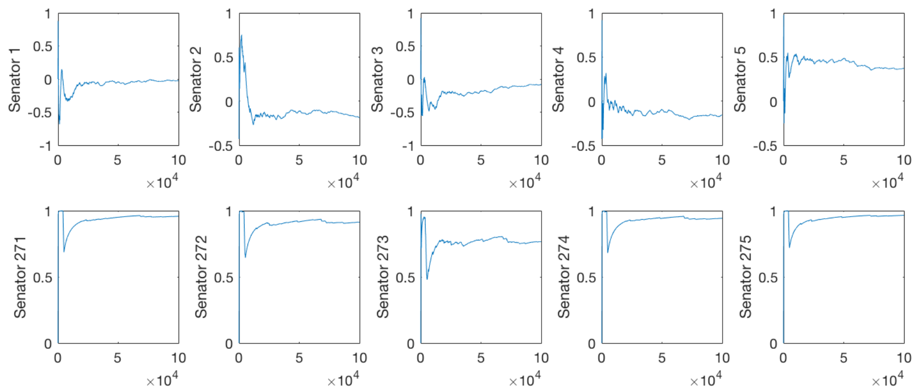


FIG. 6. *Trace of M*

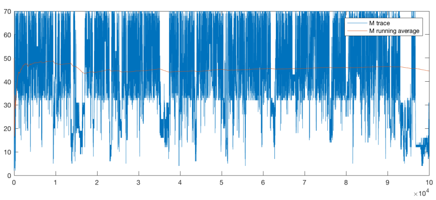


FIG. 7. *M acceptance probability*

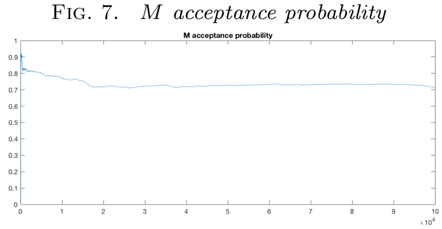


FIG. 8. *Average of $S(u)$*

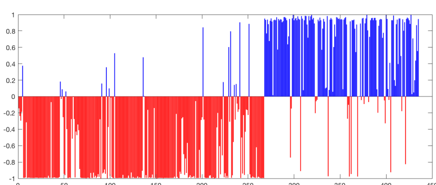


FIG. 9. *ξ acceptance probability*

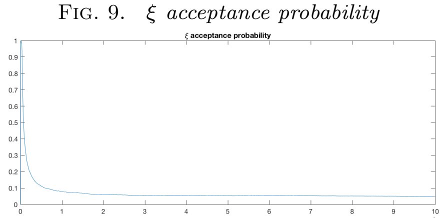
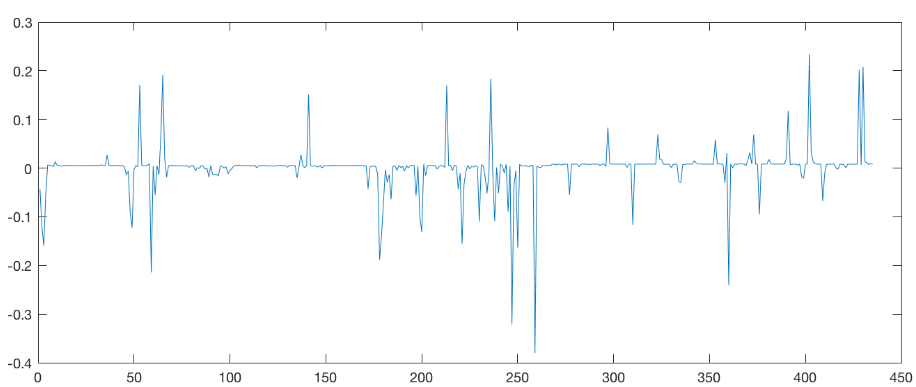


FIG. 10. *34th eigenvector of unnormalized fixed length-scale L*



3.2. Two moons. I tried a similar set of experiments on the two moons data. I fixed $\tau = 2, \alpha = 35$, again following the expected values of these hyperparameters from the noncentered algorithm that learns τ, α.

REFERENCES