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.

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

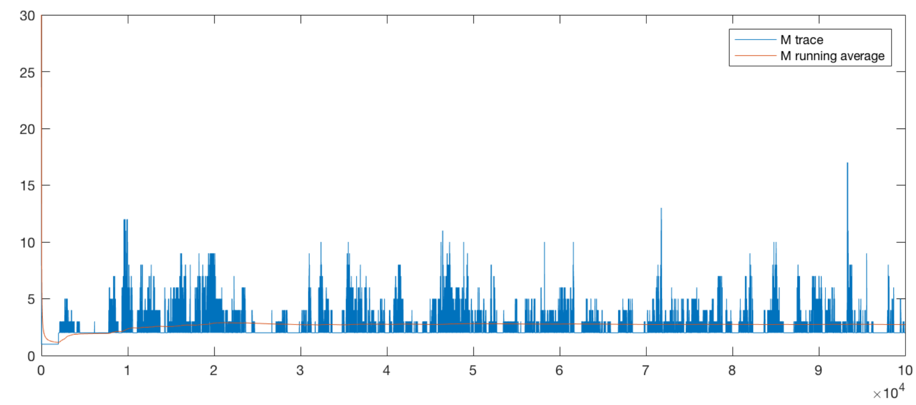
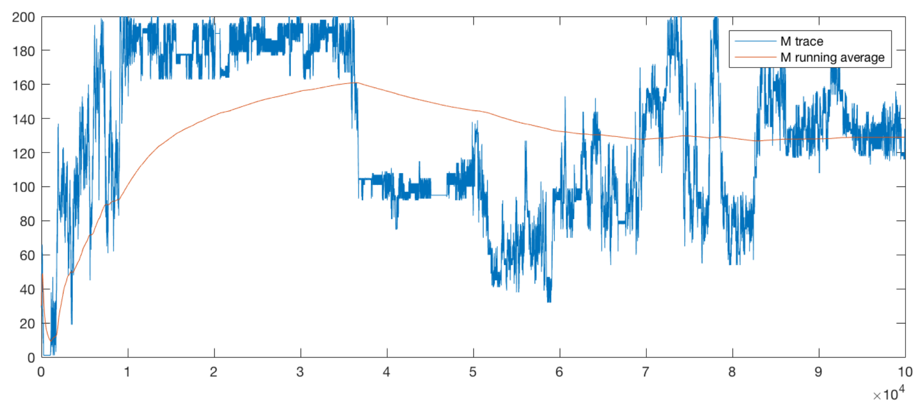


FIG. 2. $\sigma = 0.2$, *trace of* M



REFERENCES