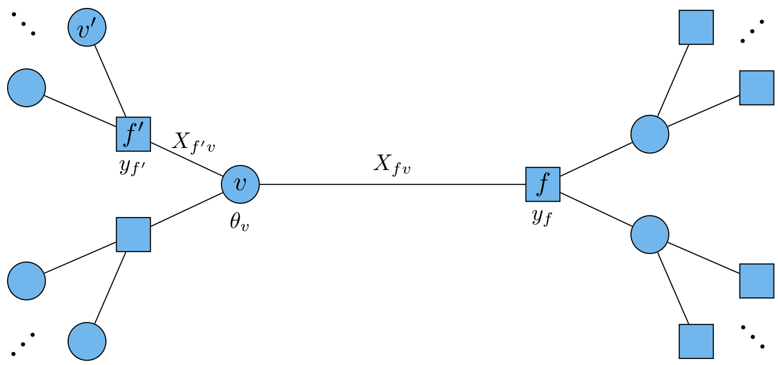
OOPS 2020
Mean field methods in high-dimensional statistics and nonconvex optimization
Lecturer: Andrea Montanari
Problem session leader: Michael Celentano
July 6, 2020

Problem Session 2

Problem 1: from belief propagation to Bayes AMP state evolution

Below I have depicted the computation tree.



We observe the edge weights X_{fv} and for each factor node the outcome

$$y_f = \sum_{v' \in \partial f} X_{fv'}\theta_{v'} + w_f.$$

Recall

$$X_{fv} \overset{\text{iid}}{\sim} \mathsf{N}(0, 1/n), \quad \theta_v \overset{\text{iid}}{\sim} \mu_\Theta, \quad \text{and } w_f \overset{\text{iid}}{\sim} \mathsf{N}(0, \sigma^2).$$

The belief propagation algorithm on the computation tree exactly computes the posterior $p_v(\vartheta|\mathcal{T}_{v,2t})$, where $\mathcal{T}_{v,2t}$ is the σ-algebra generated by the obervations corresponding to nodes and edges within a $2t$-radius ball of v. The iteration is

$$m^0_{v\to f}(\vartheta) = 1,$$

$$\tilde{m}^s_{f\to v}(\vartheta) \propto \int \exp\left(-\frac{1}{2\sigma^2}\Big(y_f - X_{fv}\vartheta - \sum_{v'\in\partial f\setminus v} X_{fv'}\vartheta_{v'}\Big)^2\right) \prod_{v'\in\partial f\setminus v} m^s_{v'\to f}(\vartheta_{v'}) \prod_{v'\in\partial f\setminus v} \mu_\Theta(\mathrm{d}\vartheta_{v'}),$$

$$m^{s+1}_{v\to f}(\vartheta) \propto \prod_{f'\in\partial v\setminus f} \tilde{m}^s_{f'\to v}(\vartheta),$$

with normalization $\int \tilde{m}^t_{f\to v}(\vartheta)\mu_{\Theta|V}(v_v, \mathrm{d}\vartheta) = \int m^t_{v\to f}(\vartheta)\mu_{\Theta|V}(v_v,\mathrm{d}\vartheta) = 1$. One can show that for any variable node v, the posterior density with respect to measure μ_Θ is

$$p_v(\vartheta|\mathcal{T}_{v,2t}) \propto \prod_{f\in\partial v} \tilde{m}^{t-1}_{f\to v}(\vartheta).$$

This equation is exact. Our goal is to show that when $n, d \to \infty$, $n/d \to \delta$

$$p_v(\vartheta|\mathcal{T}_{v,2t}) \propto \exp\left(-\frac{1}{2\tau_t^2}(\chi^t_v - \vartheta)^2 + o_p(1)\right),$$

where $(\chi_v^t, \theta_v) \xrightarrow{\mathrm{d}} (\Theta + \tau_t Z, \Theta)$, $\Theta \sim \mu_\Theta$, $G \sim \mathsf{N}(0,1)$ independent of Θ, and τ_t is given by the Bayes AMP state evolution equations

$$\tau_{t+1}^2 = \sigma^2 + \frac{1}{\delta}\mathsf{mmse}_\Theta(\tau_t^2),$$

initialized by $\tau_0^2 = \infty$. In fact, this follows without too much work once we show that

$$m_{v\to f}^s(\vartheta) \propto \exp\left(-\frac{1}{2\tau_s^2}(\chi_{v\to f}^s - \vartheta)^2 + o_p(1)\right), \tag{1}$$

where $(\chi_{v\to f}^s, \theta_v) \xrightarrow{\mathrm{d}} (\Theta + \tau_s Z \mu_{v'\to f}^s -, \Theta)$. This problem focuses on establishing (1). We do so inductively. The base case more-or-less follows the standard inductive step, except that we need to pay some attention to the infinite variance $\tau_0^2 = \infty$. We do not consider the base case here. Throughout, we assume μ_Θ has compact support. We do not carefully verify the validity of all approximations. See *Celentano, Montanari, Wu. "The estimation error of general first order methods." COLT 2020*, for complete details.

(a) Define

$$\mu^s_{v\to f} = \int \vartheta m^s_{v\to f}(\vartheta)\mu_\Theta(\mathrm{d}\vartheta), \quad (\tau^s_{v\to f})^2 = \int \vartheta^2 m^s_{v\to f}(\vartheta)\mu_\Theta(\mathrm{d}\vartheta) - (\mu^s_{v'\to f})^2,$$

and

$$\tilde{\mu}^s_{f\to v} = \sum_{v'\in\partial f\setminus v} X_{fv'}\mu^s_{v'\to f}, \quad (\tilde{\tau}^s_{f\to v})^2 = \sum_{v'\in\partial f\setminus v} X^2_{fv'}(\tau^s_{v'\to f})^2.$$

Argue (non-rigorously) that we may approximate (up to normalization)

$$\tilde{m}^s_{f\to v}(\vartheta) \approx \mathbb{E}_G\left[p(X_{fv}\vartheta + \tilde{\mu}^s_{f\to v} + \tilde{\tau}^s_{f\to v}G - y_f)\right],$$

where $G \sim \mathsf{N}(0,1)$ and $p(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}x^2}$ is the normal density at variance σ^2.

Remark: The quantities $\mu^s_{v\to f}$ and $(\tau^s_{v\to f})^2$ have a simple statistical interpretation: they are the posterior mean and variance for θ_v given observations in the computation tree within distance $2s$ of node v and excluding the branch in the direction of f.

(b) Using the inductive hypothesis, show that as $n, d \to \infty$, $n/d \to \delta$

$$(\tilde{\tau}^s_{f\to v})^2 \xrightarrow{\mathrm{p}} \frac{1}{\delta}\mathsf{mmse}_\Theta(\tau^2_s) =: \tilde{\tau}^2_s.$$

Further, note $y_f - \tilde{\mu}^s_{f\to v} = X_{fv}\theta_v + \tilde{Z}^s_{f\to v}$, where

$$\tilde{Z}^s_{f\to v} = w_f + \sum_{v'\in\partial f\setminus v} X_{fv'}(\theta_{v'} - \mu^s_{v'\to f}).$$

Argue

$$\tilde{Z}^s_{f\to v} \xrightarrow{\mathrm{d}} \mathsf{N}\left(0, \sigma^2 + \frac{1}{\delta}\mathsf{mmse}_\Theta(\tau^2_s)\right)$$

and is independent of X_{fv} and θ_v.

Hint: The (random) functions $m^s_{v'\to f}(\vartheta_{v'})$ as v' varies in ∂f are iid and independent of the edge weights $X_{fv'}$. Why?

(c) For any smooth probability density $f : \mathbb{R} \to \mathbb{R}_{>0}$, $\mu \in \mathbb{R}$, and $\tau > 0$, show that

$$\frac{\mathrm{d}}{\mathrm{d}\tilde{\mu}}\log \mathbb{E}_G[f(\tilde{\mu} + \tilde{\tau}G)] = -\frac{1}{\tilde{\tau}}\mathbb{E}[G|S + \tilde{\tau}G = \tilde{\mu}],$$

$$\frac{\mathrm{d}^2}{\mathrm{d}\tilde{\mu}^2}\log \mathbb{E}_G[f(\tilde{\mu} + \tilde{\tau}G)] = -\frac{1}{\tilde{\tau}^2}(1 - \mathrm{Var}[G|S + \tilde{\tau}G = \tilde{\mu}]),$$

where $S \sim f(s)\mathrm{d}s$ independent of $G \sim \mathsf{N}(0,1)$.

(d) We Taylor expand

$$\log \tilde{m}^s_{f\to v}(\vartheta) \approx \mathsf{const} + X_{fv}\tilde{a}^s_{f\to v}\vartheta - \frac{1}{2}X^2_{fv}\tilde{b}^s_{f\to v}\vartheta^2 + O_p(n^{-3/2}).$$

(We take this to be the definition of $\tilde{a}^s_{f\to v}$ and $\tilde{b}^s_{f\to v}$). Taking the approximation in part (a) to hold with equality, argue

$$\tilde{a}^s_{f\to v} = \frac{1}{\tau^2_{s+1}}(y_f - \tilde{\mu}^s_{f\to v}) + o_p(1), \qquad \tilde{b}^s_{f\to v} = \frac{1}{\tau^2_{s+1}} + o_p(1).$$

(e) Taking the approximations in part (d) to hold with equality and using part (b) to subsitute for $y_f - \tilde{\mu}^s_{f\to v}$, Taylor expand $\log m^{s+1}_{v\to f}(\vartheta)$ to conclude

$$\log m^{s+1}_{v\to f}(\vartheta) = \mathsf{const} + \frac{1}{\tau^2_{s+1}}\chi^{s+1}_{v\to f}\vartheta - \frac{1}{\tau^2_{s+1}}\vartheta^2 + o_p(1),$$

where $(\chi^{s+1}_{v\to f}, \Theta) \xrightarrow{\mathrm{d}} (\Theta + \tau_{s+1}Z, \Theta)$. Why do we expect this Taylor expansion to be valid for all $\vartheta = O(1)$? Conclude Eq. (1).