Last lecture: generalized regression problem

$$y_i = h(\theta_0^T x_i; w_i) \quad i \le n, \quad (\theta_{0i})_{i\le d} \sim_{iid} P_\Theta \qquad x_i \sim N(0, I)$$

minimize $\mathcal{L}_n(\theta) := \frac{1}{n}\sum_{i=1}^{n} \ell(y_i; x_i^T\theta) + \Lambda(\theta)$

GFOM:

$$\begin{cases} \theta^{t+1} = X^T F_t^{(1)}(u^1, \dots u^t; y) + F_t^{(2)}(\theta^1, \dots \theta^t) \\ u^t = X G_t^{(1)}(\theta^1, \dots \theta^t) + G_t^{(2)}(u^1, \dots u^{t-1}; y) \end{cases}$$

Thm | Assume $(X_{ij}) \sim_{iid} N(0, 1/n)$. For any GFOM $(\theta^t)_{t\ge 0}$

$$\liminf_{n,d\to\infty} \frac{1}{d}\|\theta^t - \theta_0\|^2 \geq \lim_{n,d\to\infty} \frac{1}{d}\|\theta^t_{BAMP} - \theta_0\|^2 = \underbrace{\hat{\tau}_t^2}_{\text{explicit}}$$

Will explain this thm in a simpler setting

Rank-one matrix estimation

$$X = \frac{1}{n}\theta_0\theta_0^T + W \quad ; \quad W \sim GOE(n), \ (\theta_{0,i})_{i\le n} \sim_{iid} P_\Theta$$

$$[\, W = W^T, \ (W_{ij})_{i<j} \sim_{iid} N(0, 1/n), \ (W_{ii})_{i\le n} \sim_{iid} N(0, 2/n) \,]$$

minimize $\mathcal{L}_n(\theta) := \|X - \frac{1}{n}\theta\theta^T\|_F^2 + \Lambda(\theta)$

$-\log P_\theta(X)$

GFOM

$$\theta^{t+1} = X F_t^{(1)}(\theta^1, \dots, \theta^t) + F_t^{(2)}(\theta^1 \dots \theta^t) \quad *$$

AMP

$$\theta^{t+1} = X f_t(\theta^1, \dots \theta^t) - \sum_{s=0}^{t} b_{t,s} f_{s-1}(\theta^1 \dots \theta^{s-1}) \quad ; \quad f_t : \mathbb{R}^t \to \mathbb{R}$$

$$b_{t,s} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial f_t}{\partial \theta_i^s}(\theta_i^1 \dots \theta_i^t)$$

$$f_t(\theta^1, \dots \theta^t) = \left[f_t(\theta_i^1 \dots \theta_i^t) \right]_{i\le n}$$

Bayes AMP

$$f_t(x_1 \dots x_t) = \mathbb{E}\{\Theta \mid \xi_t \Theta + \sqrt{\xi_t} Z = x_t\}$$

$$\xi_{t+1} = \mathbb{E}\{\mathbb{E}(\Theta \mid \xi_t\Theta + \sqrt{\xi_t}Z)^2\}$$

$$\Theta \sim p_\Theta \perp Z \sim N(0,1)$$

$$\hat{\theta}^t_{BAMP} = f_t(\theta^t_{BAMP})$$

Thm

$$\limsup_{n\to\infty} \frac{\langle \theta^t, \theta_0\rangle}{\|\theta^t\| \|\theta_0\|} \leq \lim_{n\to\infty} \frac{\langle \hat{\theta}^t_{BAMP}, \theta_0\rangle}{\|\hat{\theta}_{BAMP}\|_2 \|\theta_0\|} = \sqrt{\xi_t}$$

□

① **Reduction**

Lem $\forall$ GFOM $(\bar{\theta}^t)_{t\geq 0}$ $\exists$ AMP $(\theta^t)_{t\geq 0}$, $\phi_t : \mathbb{R}^t \to \mathbb{R}$ st

$$\bar{\theta}^t = \phi_t(\theta^1, \dots \theta^t)$$

□

" GFOM = AMP + Post processing "

② **Analysis of AMP**

Thm If f_t Lipschitz $\forall t$

$$\frac{1}{n}\sum_{i\leq n} \delta_{(\theta_{0i}, \theta^1_i, \dots \theta^t_i)} \overset{W_2}{\Longrightarrow} \text{Law}(\Theta, \mu_1\Theta + Z_1, \dots \mu_t\Theta + Z_t)$$

$$(\Theta, Z_1, \dots Z_t) \sim p_\Theta \otimes N(0, Q_{\leq t})$$

$$\mu_{t+1} = \mathbb{E}\{\Theta f_t(\mu_1\Theta + Z_1, \dots \mu_t\Theta + Z_t)\}$$

$$Q_{s+1,t+1} = \mathbb{E}(F_s F_t)$$

$$F_s := f_s(\mu_1\Theta + Z_1, \dots \mu_s\Theta + Z_s)$$

□

Interpret (#1) $\Theta = 0$

v^1, v^2 vect indep of $X = W$

$$z^i := X v^i$$

$$\frac{1}{n}\begin{pmatrix} \|v^1\|^2 & \langle v^1 v^2\rangle \\ \langle v^1 v^2\rangle & \|v^2\|^2 \end{pmatrix} \to Q \quad \text{then} \quad \frac{1}{n}\sum_{i=1}^n \delta_{z^1_i, z^2_i} \overset{W_2}{\Longrightarrow} N(0,Q)$$

$$\theta^{t+1} = X f(\theta^1, \dots \theta^t) - \boxed{}$$

are dependent
act "as if independent"

Interpret (#2)

$$P_\Theta = \frac{1}{2}\delta_{+1} + \frac{1}{2}\delta_{-1} \qquad \|$$

$$P_\Theta = \left(\tfrac{1}{2}+\epsilon\right)\delta_{+1} + \left(\tfrac{1}{2}-\epsilon\right)\delta_{-1}$$

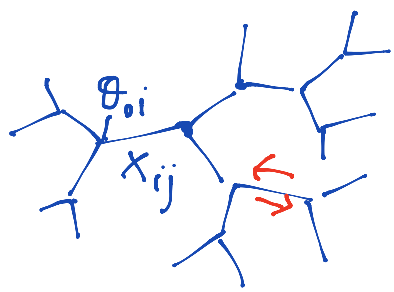
$$v^0 = (+a, \dots +a)$$

$$\theta^0 = c \cdot v_1(X)$$

$T_n = (V_n, E_n)$ infinite tree of degree n

$(\theta_{0i})_{i\in V_n} \overset{iid}{\sim} P_\Theta \qquad X_{ij} = \frac{\theta_{0i}\theta_{0j}}{n} + W_{ij}$

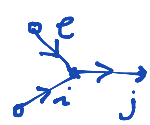
$(W_{ij})_{ij\in E_n} \overset{iid}{\sim} N(0, 1/n)$



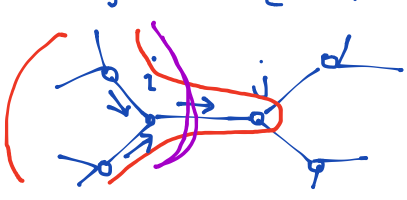
Given (X_{ij})
estimate θ_0

Message passing

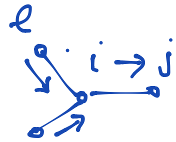
$$u^{t+1}_{i\to j} = \sum_{e\in\partial i\setminus j} X_{ie}\, f_t(u^1_{e\to i}, \dots u^t_{e\to i})$$



$$\mathcal{F}_{i\to j} := \sigma\left(\{W_{em} : (em)\in E(T_{i\to j}),\ \theta_{0e}\ e\in V(T_{i\to j})\}\right)$$



(1) $u^t_{i\to j} \in m\, \mathcal{F}^t_{i\to j}$



(2) $(\theta_{0i}, u^1_{i\to j}, \dots, u^t_{i\to j}) \overset{W_2}{\Rightarrow} (\Theta, \mu_1\Theta + Z_1, \dots \mu_t\Theta + Z_t)$

③ Optimality of Bayes AMP

- Sufficient to prove LB for estimation on T_n
- Message passing is a local alg

  $u^t_{i\to j} \in m\sigma(\{ X_{\ell m} : (\ell m) \in B_i(t) \})$
- Optimal local algorithm

  $$\hat{\theta}_i = \mathbb{E}(\theta_{0i} \mid (X_{\ell m}) : (\ell m) \in B_i(t))$$

  -- Can be implemented as a message passing algorithm!
  (Belief Propagation!)

Why should we expect state evolution?

$\theta_0 = 0$

$$\theta^{t+1} = W \underbrace{f(\theta^1, \dots \theta^t)}_{t} - \sum_{s=1}^{t} d_{t,s} f_{s-1}(\theta^1, \dots \theta^s)$$

$$\mathcal{G}_t = \sigma(\theta^1, \dots \theta^t) \to f_t \qquad x^{t+1} = W f_t$$

$$x^{t+1} = \theta^{t+1} + \sum_{s=1}^{t} d_{t,s} f_{s-1}(\theta^1, \dots \theta^s)$$

Conditioning on $\mathcal{G}_t$

$\equiv$ Condition on

$$x^1 = W f^0, \; x^2 = W f^1, \dots \; x^t = W f^{t-1}$$

$$X_t = W F_t \qquad X_t = [x^1, \dots x^t], \quad F_t = [f^0, \dots f^{t-1}]$$

$$W|_{\mathcal{G}_t} \stackrel{d}{=} P^\perp_{F_t} W^{new} P^\perp_{F_t} + \underbrace{\mathbb{E}[W \mid \mathcal{G}_t]}_{M(X_t, F_t)}$$

rank $= n - t$

$$\theta^{t+1} \stackrel{d}{=} P^\perp_{F_t} \underbrace{W^{new} P^\perp_{F_t}} f^t + \underbrace{M(X_t, F_t) f^{t-1}} - \sum_{s=1}^{t} d_{t,s} f^{s-1}$$

$$\approx W^{new} P^\perp_{F_t} f^t + \sum_{s \le t} c_{s,t} \theta^s + \sum d'_{t,s} f^{s-1} - \sum d_{t,s} f^{s-1}$$