MEAN FIELD METHODS IN HIGH-DIMENSIONAL STATISTICS AND NON-CONVEX OPTIMIZATION

1. Motivation
2. Exact asymptotics via Gaussian comparison
3. First order algs (GFOM) and AMP
4. Optimal GFOMs for regression
5. Optimization of MF spin glasses: Background
6. AMP algorithms for optimization
7. Optimal AMP and connection to Parisi formula

$(P_\theta)_{\theta\in\Theta}$ $\quad \Theta \subseteq \mathbb{R}^d$ $\quad$ pr. distr on $\mathcal{X}$

$z_1, \dots z_n \sim_{iid} P_\theta \qquad \underline{z} = (z_1, \dots z_n)$

$\hat{\theta} : \mathcal{X}^n \to \mathbb{R}^d, \quad \underline{z} \mapsto \hat{\theta}(\underline{z})$

$R(\hat{\theta}, \theta_0) := \mathbb{E}_{\theta_0} \mathrm{dist}(\hat{\theta}(\underline{z}), \theta_0)$

fct $\quad \ell : \mathcal{X} \times \mathbb{R}^d \to \mathbb{R}_{\geq 0}$

define $\mathcal{L}(\theta, \theta_0) := \mathbb{E}_{\theta_0} \ell(\underline{z}, \theta)$

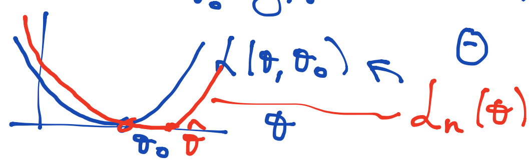
$$\theta_0 = \arg\min_{\theta} \mathcal{L}(\theta, \theta_0)$$

$$\mathcal{L}_n(\theta) := \frac{1}{n} \sum_{i \leq n} \ell(z_i, \theta) = \hat{\mathbb{E}}_n \ell(z, \theta)$$

$$\hat{\theta} := \arg\min_{\theta \in \Omega_n} \mathcal{L}_n(\theta)$$

$$\ell(z, \theta) = -\log p_\theta(z)$$

$$\mathcal{L}(\theta,\theta_0) = -\mathbb{E}_{\theta_0} \log p_\theta(z) = KL(p_{\theta_0} \| p_\theta) + cst.$$



Classically : d fixed, $n \to \infty$

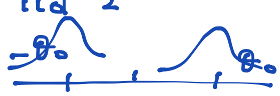
High. dim : $n, d \to \infty$, $n \ll d$, $\dim(\Theta) \ll n$

$\Theta = \{ s_0\text{-sparse vectors} \} \subseteq \mathbb{R}^d$

$s_0 \ll d \qquad s_0 \ll n.$

Noisy hidim : $\frac{d}{n}, \frac{\dim(\Theta)}{n} \asymp 1$

1) $z_i \underset{iid}{\sim} \frac{1}{2} N(\theta_0, I_d) + \frac{1}{2} N(-\theta_0, I_d)$



2) Sparse regression

$z_i = (y_i, x_i) \quad x_i \sim N(0, I_d) \quad y_i = \langle \theta_0, x_i \rangle + \epsilon_i \quad \epsilon_i \sim N(0, \sigma^2)$

θ_0 sparse

$$\mathcal{L}_n(\theta) = \underbrace{\frac{1}{2n} \| y - X\theta \|_2^2}_{-\log p_\theta(z)} + \underbrace{\frac{\lambda}{\sqrt{n}} \|\theta\|_1}_{\text{promotes sparse } \theta}.$$

3) Robust regression (M. estimation)

$$\mathcal{L}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \varphi(y_i - \langle x_i, \theta \rangle)$$

- Gaussian comparison (simple, elegant)
- AMP (algorithmic)

$\Leftarrow \; z_1 \ldots z_n \simeq \theta + \epsilon_i \quad |$

$\hat\theta = \frac{1}{n} \sum_{i=1}^{n} z_i \leftarrow \sum_i (z_i - \theta)^2$

$\hat\theta = \text{median}(z_i) \leftarrow \sum |z_i - \theta| = \sum \varphi(z_i - \theta)$

Theorem (Gordon) $(X_{s,t})_{s\in S, t\in T}$ $(Y_{s,t})_{s\in S, t\in T}$ cent. Gauss

Assume 1) $\mathbb{E}X_{st}^2 = \mathbb{E}Y_{st}^2$

2) $\mathbb{E}[(X_{st_1}-X_{st_2})^2] \geq \mathbb{E}[(Y_{st_1}-Y_{st_2})^2]$ $\forall t_1, t_2$

3) $\mathbb{E}[(X_{s_1t_1}-X_{s_2t_2})^2] \leq \mathbb{E}[(Y_{s_1t_1}-Y_{s_2t_2})^2]$ $\forall s_1 \neq s_2$

Then

$$\min_s \max_t (X_{st}-\xi_{st}) \succsim \min_s \max_t (Y_{st}-\xi_{st})$$

$\forall (\xi_{st})$

$$A \succsim B \iff \mathbb{P}(A \geq u) \geq \mathbb{P}(B \geq u) \quad \forall u \in \mathbb{R} \quad \square$$

Rmk Generalizes Fernique $|T|=1$ $\square$

Corollary $U \subseteq \mathbb{R}^d$, $V \subseteq \mathbb{R}^n$ compact, Q cont.

$(G_{ij}) \underset{iid}{\sim} N(0,1)$, $(g_i) \underset{iid}{\sim} N(0,1)$ $(h_j) \underset{iid}{\sim} N(0,1)$

$$L_*(G) := \min_{u\in U} \max_{v\in V} \{\langle v, Gu\rangle + Q(u,v)\}$$

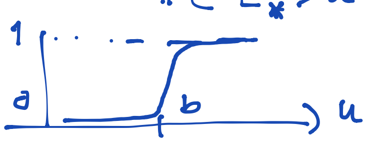
$$B_*(g,h) := \min_{u\in U} \max_{v\in V} \{\|v\|\langle g,u\rangle + \|u\|\langle h,v\rangle + Q(u,v)\}$$

Then

$-$ $\mathbb{P}(L_* \leq u) \leq 2\mathbb{P}(B_* \leq u)$ $\forall u \in \mathbb{R}$

Further if minmax is convex-concave

$-$ $\mathbb{P}(L_* \geq u) \leq 2\mathbb{P}(B_* \geq u)$ #



$\mathbb{P}(B_* \geq u)$

Proof $X_{u,v} = \|v\|\langle g,u\rangle + \|u\|\langle h,v\rangle$

$Y_{uv} = \langle v, Gu\rangle + z\|u\|\|v\|$ $\quad z \sim N(0,1)$

$$\mathbb{E}(Y_{u_1v_1}Y_{u_2v_2}) - \mathbb{E}[X_{u_1v_1}X_{u_2v_2}] = \langle u_1,u_2\rangle\cdot\langle v_1,v_2\rangle + \|u_1\|\|u_2\|\|v_1\|\|v_2\|$$
$$- \langle u_1,u_2\rangle\|v_1\|\|v_2\| - \langle v_1,v_2\rangle\|u_1\|\|u_2\|$$

$$= \left(\|u_1\|\|u_2\| - \langle u_1,u_2\rangle\right)\left(\|v_1\|\|v_2\| - \langle v_1,v_2\rangle\right)$$

$$\geq 0 \qquad \square$$

$$L_* = \min_u \max_v \left[\langle v, G u\rangle + \dots\right]$$
$$= \max_v \min_u \left[\langle v, G u\rangle + \dots\right]$$
$$= -\min_v \max_u \left[\langle v, \underbrace{(-G)}_{\overset{d}{=} G} u\rangle + \dots\right] \qquad \boxtimes$$

How to apply it?

$$\theta - \theta_0 = u$$

$$\mathcal{L}_n(\theta) = \frac{1}{2n}\|y - X\theta\|^2 + \lambda\|\theta\|_1 \,, \quad y = X\theta_0 + \sigma w$$

$$\min_\theta \mathcal{L}_n(\theta) = \min_{u \in \mathbb{R}^d} \frac{1}{2n}\|\sigma w - X u\|^2 + \lambda\|\theta_0 + u\|_1$$

$$= \min_{u\in\mathbb{R}^d} \max_{v \in \mathbb{R}^n} \left\{ \frac{1}{n}\langle v, X u\rangle - \frac{1}{2n}\left(\|v\|^2 - \sigma\langle w, v\rangle\right) + \lambda\|\theta_0 + u\|_1 \right\}$$

Q

$$\hat\theta = \arg\min_\theta \mathcal{L}_n(\theta)$$

$$\mathcal{L}_n^*(S) = \min_{\theta\in S} \mathcal{L}_n(\theta)$$

$$\mathcal{L}_n^*(S) > \mathcal{L}_n^* \Rightarrow \hat\theta \in S^c$$

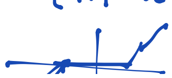
$$\|\hat\theta - \theta_0\|_2^2 \xrightarrow{P} a \qquad S = \left\{\theta : \left|\|\theta - \theta_0\|^2 - a\right| \geq \epsilon\right\}$$

$$n, d \to \infty\,, \quad \frac{n}{d} \to \delta \qquad \hat\mu_{\theta_0} = \frac{1}{n}\sum_{i=1}^n \delta_{\theta_{0,i}} \overset{W_2}{\Longrightarrow} P_\Theta$$

τ_*, β_* sol of

$$\begin{cases} \tau^2 = \sigma^2 + \frac{1}{\delta}\mathbb{E}\left\{\left(\eta_{\frac{\tau\lambda}{\beta}}(\Theta + \tau Z) - \Theta\right)^2\right\} \\ \beta = \tau\left(1 - \frac{1}{\delta}\mathbb{P}\left(|\Theta + \tau Z| \geq \frac{\tau\lambda}{\beta}\right)\right) \end{cases} \qquad \begin{matrix}\Theta \sim P_\Theta \\ Z \sim N(0,1)\end{matrix}$$

$$\eta(x; u) = (|x| - u)_+ \cdot \mathrm{sign}(x)$$



Thm $\hat{\mu}_\lambda = \frac{1}{d}\sum_{i=1}^{d} \delta_{\underline{\theta_{0i}, \hat{\theta}_i}}$ $\hat{\mu}^d_\lambda = \mathrm{Law}\left(\Theta, \eta_{\frac{\tau\lambda}{\rho}}(\Theta + \tau Z)\right)$

$$\mathbb{P}\left(W_2(\hat{\mu}, \hat{\mu}^d) \geq \epsilon\right) \leq C(\epsilon) e^{-n\bar{c}(\epsilon)}$$

$\|\hat{\theta} - \theta_0\|_2^2 \to 0$ Class

$$= \frac{d}{n}$$

$$= \frac{\dim(\Theta)}{n} \log d \to 0$$

$\|\hat{\theta} - \theta_0\|^2 \to \delta(\tau_*^2 - \sigma^2)$

$\psi(\hat{\theta}, \theta_0) = (\hat{\theta} - \theta_0)^2$

$\sum_{i=1}^{d} (\hat{\theta}_i - \theta_{0,i})^2 = \|\hat{\theta} - \theta_0\|^2 \to *$

$\mathcal{L}_n(\theta) = \hat{\mathbb{E}}_n \underline{\ell(z, \theta)}$

$\min \sum_i \ell(z_i, \theta)$ $\max_v \langle \ \rangle - \ell^*$