Summary from previous lecture

Data: $z_1, \dots z_n \sim_{iid} P_\theta$; $(P_\theta)_{\theta \in \Theta}$

Estimation: $\hat{\theta}(z)$

loss

$$\text{minimize } \mathcal{L}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \ell(z_i, \theta) + \Lambda(\theta)$$

regularizer

Example: (Sparse regression / Lasso)

$$z_i = (y_i, x_i) \ , \ x_i \sim N(0, I_d) \quad y_i = \theta_0^T x_i + \varepsilon_i$$

$$\mathcal{L}_n(\theta) = \frac{1}{2n} \|y - X\theta\|^2 + \underline{\lambda \|\theta\|_1} .$$

$$\sum_i (y_i - x_i^T\theta)^2 \qquad n \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}}^{d}$$

Algorithms?

Gradient descent

$$\theta^{t+1} = \theta^t - s_t \nabla \mathcal{L}_n(\theta^t)$$

step size

- Prox gradient
- Acc. gradient
- Mirror descent

$\}$ FOM

Common structure? $\quad z_i = (y_i; x_i) \quad y_i \in \mathbb{R}, x_i \in \mathbb{R}^d$

$$\ell(z_i; \theta) = \ell(y_i, \theta^T x_i)$$

$$\ell_n(\theta) = \frac{1}{n} \sum_i \ell(y_i; \theta^T x_i) = \frac{1}{n} \ell(y; X\theta)$$

$$\nabla \ell_n(\theta) = X^T f(y; X\theta)$$

$$f(y; X\theta) = \begin{pmatrix} f(y_1, x_1^T\theta) \\ \vdots \\ f(y_n, x_n^T\theta) \end{pmatrix} \qquad f(y; \hat{y}) = \partial_{\hat{y}} \ell(y, \hat{y})$$

$$\theta^{t+1} = \theta^t - s_t \underline{X}^T f(y; \underline{X\theta^t})$$

* Mult. by X, X^T
* Apply separable fct

GFOM

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

$u^t \in \mathbb{R}^n$, $\theta^t \in \mathbb{R}^d$

$F_t^{(1)} : \mathbb{R}^{t+1} \to \mathbb{R}$

Can we analyze GFOMs?

Find the optimal one (statist)?

Setting $(X_{ij}) \sim_{iid} N(0, 1/n)$ $\quad \frac{n}{d} \to \delta \in (0,\infty$

$y_i = h(x_i^T\theta_0 ; w_i)$ $\quad w_i \sim \mu_W$, h suff reg.

$(\theta_0, \tilde{\theta}) \sim_{iid} \mu_{\theta_0\tilde{\theta}}$

$\dot{\theta}_t = -\nabla \mathcal{L}(\theta_t)$

$\ddot{\theta}_t = a\dot{\theta}_t - \nabla\mathcal{L}(\theta_t) \leftarrow$

Thm For any GFOM

$$\liminf_{n,d\to\infty} \frac{1}{d}\|\theta^t - \theta_0\|^2 \geq \hat{\tau}_t^2 \quad \text{where } \hat{\tau}_t \text{ explicit}$$

Further, there exists a special GFOM (Bayes AMP)

$$\lim_{n,d\to\infty} \frac{1}{d}\|\theta^t_{BAMP} - \theta_0\|^2 = \hat{\tau}_t^2 .$$

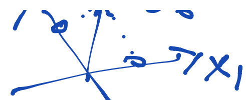
Proof (1) Reduction GFOM $\to$ AMP

(2) Sharp analysis, $n,d\to\infty$

(3) BAMP optimal among AMP $\square$

1) example Phase retrieval.

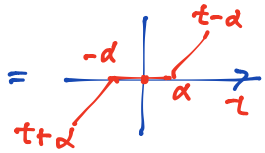
θ_0 $\quad y_i = \langle x_i, \theta_0\rangle^2 + \epsilon_i$ $\quad \mathcal{L}_n(\theta) = \sum_i (y_i - \langle x_i,\theta_0\rangle^2)^2$



$$\mathcal{L}_n(\theta) = \frac{1}{2}\|y - X\theta\|^2 + \lambda\|\theta\|_1$$

$$\begin{cases} \theta^{t+1} = \eta\left(\hat\theta^t + \frac{1}{L}X^T u^t ; \frac{\lambda}{L}\right) \\ u^t = y - X\hat\theta^t \end{cases}$$

$\eta(t;\alpha) =$



AMP

$$\begin{cases} \theta^{t+1} = \eta(\overbrace{\theta^t + X^T u^t}^{\tilde\theta_i^t} ; \alpha_t) \\ u^t = y - X\theta^t + \xi_t u^{t-1} \end{cases}$$

$$\xi_t = \frac{\|\theta^t\|_0}{n}$$

Thm $\frac{1}{d}\sum_{i=1}^{d} \delta_{\theta_{0,i}, \tilde\theta_i^t} \overset{W_2}{\Longrightarrow} \mathrm{Law}(\Theta, \Theta + \tau_t Z)$ $\quad Z \sim N(0,1)$ ind of Θ $\sim p_\Theta$

$$\tau_{t+1}^2 = \sigma^2 + \frac{1}{\delta}\mathbb{E}\{[\eta(\Theta + \tau_t Z; \alpha_t) - \Theta]^2\} \leftarrow \mathrm{II}$$

$\uparrow$ SE

$$\frac{1}{d}\|\tilde\theta^t - \theta_0\|_2^2 = \frac{1}{d}\sum_{i=1}^{d}(\tilde\theta_i^t - \theta_{0,i})^2 \to \mathbb{E}[(\Theta + \tau_t Z - \Theta)^2] = \tau_t^2$$

$$\frac{1}{d}\|\theta^t - \theta_0\|_2^2 \to \mathbb{E}\{[\eta(\Theta + \tau_t Z; \alpha_t) - \Theta]^2\}$$

$$\frac{1}{d}\sum_{i=1}^{d}\psi(\tilde\theta_i^t, \theta_{0i}) \to \mathbb{E}\,\psi(\Theta + \tau_t Z, \Theta)$$

$$\begin{cases} \tau^2 = \sigma^2 + \frac{1}{\delta}\mathbb{E}\{[\eta(\Theta + \tau Z; \alpha) - \Theta]^2\} & * \\ \lambda = \alpha\left(1 - \frac{1}{\delta}\mathbb{P}(|\Theta + \tau Z| \geq \alpha)\right) \end{cases}$$

$$\tilde\theta^t \approx \theta_0 + N(0, \tau_t^2 I_d) \quad \leftarrow \text{GAUSSIAN}$$

Can improve over soft thr AMP ?

$$\begin{cases} \theta^{t+1} = h_t(\theta^t + X^T u^t) \\ u^t = y - Xu^t + \xi_t u^{t-1} \end{cases}$$

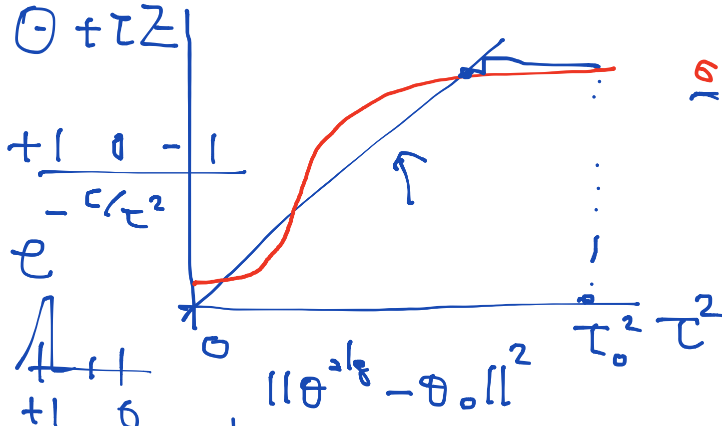
$$\tau_{t+1}^2 = \sigma^2 + \frac{1}{\delta}\mathbb{E}\{[h_t(\Theta + \tau_t Z) - \Theta]^2\} \leftarrow$$ Bayes AMP

$$h_t(y) = \mathbb{E}\{\Theta \mid \Theta + \tau_t Z = y\}$$

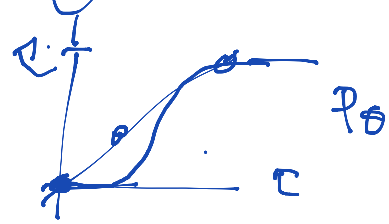
$$\tau_{t+1}^2 = \sigma^2 + \frac{1}{\delta}\,\mathrm{mmse}_{\Theta}(\tau_t^2)$$

$$\Theta + \tau_t Z$$

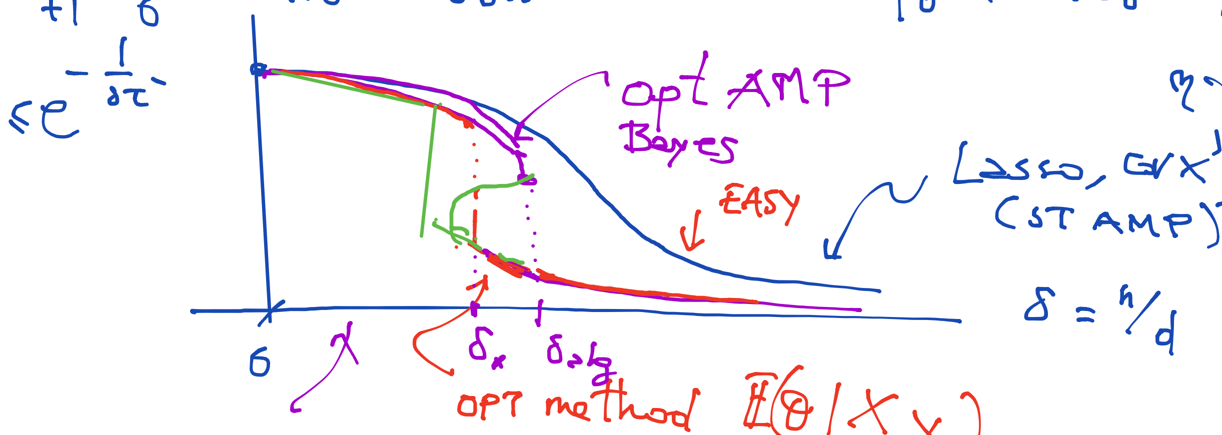
$$\mathrm{mmse}(\tau^2) \equiv \mathbb{E}\{[\Theta - \mathbb{E}(\Theta \mid \Theta + \tau Z)]^2\}$$



$$\sigma^2 + \frac{1}{\delta}\,\mathrm{mmse}_{\Theta}(\tau^2) = \tau^2$$



$$p_\Theta = (1-\varepsilon)\,\delta_0 + \varepsilon\,\delta_1$$



$0 < \delta < \delta_*$ $\qquad\qquad \uparrow \qquad\qquad \hat{\theta}$ ''

HARD

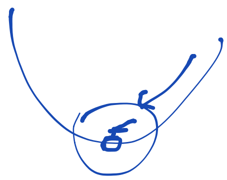
t fixed $\qquad O(nd) \qquad$ linear cplx.

$$\mathcal{L}(\theta) = \sum_i \left(y_i - \langle x_i, \theta\rangle^2\right)^2$$

$\theta = 0 \quad /\quad$ spectral init

$\lfloor + O(1)$ iterations

$\left[\begin{array}{c} O(\log n) \\ \text{random init} \end{array} \right]$



10^{-10}

$$M = \sum_{i=1}^{n} \varphi(y_i)\, x_i x_i^T$$

$v_1(M)$ $\qquad$ $\theta^0 = c \cdot v_1(M)$