(1)

Approximate Message Passing (迭代算法)

$$y = \underset{n}{A}_{p}\, \beta_0 + \varepsilon \qquad \delta = \frac{n}{p}$$

<u>AMP</u> $\beta^{(0)} = 0$ random guess

$\mathbb{R}^p \ni \beta^0 = 0$

$\mathbb{R}^n \ni z^0 = y - A\beta^0 = y$

$\mathbb{R}^p \ni \beta^1 = \eta_0(A^T z^0 + \beta^0)$ $\quad\eta_0$ is a func.

$\mathbb{R}^n \ni z^1 = y - A\beta^1 + z^0 \cdot \boxed{\frac{1}{\delta}\overline{\mathrm{Avar}}\left(\eta_0'(A^T z^0 + \beta^0)\right)}$

average for vector

(*) correct term for us to track the dist.

$\beta^2 = \eta_1(A^T z^1 + \beta^1)$

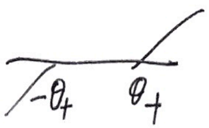
$\beta^0, z^0, \beta^1, z^1, \beta^2, \cdots, z^{t-1}, \beta^t$

~~$z^t = y - A\beta^t + z$~~

$z^t = y - A\beta^t + z^{t-1}\frac{1}{\delta}\overline{Av}\left(\eta_{t-1}'(A^T z^{t-1} + \beta^{t-1})\right)$

$\beta^{t+1} = \eta_t(A^T z^t + \beta^t)$

eg. Lasso $\quad \eta_t(x) = \eta_1(x; \theta_t)$



Abstract AMP iteration

$A \in \mathbb{R}^{n\times p}$ iid entry $N(0, \frac{1}{n})$

$\{f_t, g_t\}$

$\boxed{\text{Ite0}}$ $\quad \mathbb{R}^p \ni h^0 = 0$

$q^0 = f_0(h^0)$

$\mathbb{R}^n \ni b^0 = Aq^0$

$m^0 = g_0(b^0)$

$\boxed{\text{Ite1}}$ $\quad h^1 = A^T m^0 - \xi_0 q^0$

$q^1 = f_1(h^1)$

$b^1 = Aq^1 - \lambda_1 m^0$

$m^1 = g_1(b^1)$

$\xi_0 = \overline{Av}(g_0'(b^0))$

$\lambda_1 = \frac{1}{\delta}\overline{A_v}(f_1'(h^1))$

given iteration at $t-1$

$h^t = A^T m^{t-1} - \xi_{t-1} q^{t-1}$

$q^t = f_t(h^t)$

$b^t = Aq^t - \lambda_t m^{t-1}$ $\qquad \xi_{t-1} = \overline{Av}(g_{t-1}'(b^{t-1}))$

$m^t = g_t(b^t)$ $\qquad \lambda_t = \frac{1}{\delta}\overline{Av}(f_t'(h^t))$

Relationship: $\qquad$ def. $\eta_{-1} =$

$f_t(h) \triangleq \eta_{t-1}(\beta_0 - h) - \beta_0$

$g_t(b) \triangleq b - \varepsilon$

Then $\xi_t = \frac{1}{n}\sum_{i=1}^{n} g_t'(b_i^t) = 1$

$\lambda_t = \frac{1}{\delta}\frac{1}{p}\sum_{j=1}^{p} f_t'(h_j^t)$

$= -\frac{1}{\delta}\overline{Av}\left(\eta_{t-1}'(\beta_0 - h^t)\right)$

P14

Identification 对应

$h^t = \beta_0 - (A^T z^{t-1} + \beta^{t-1})$

$q^t = \beta^t - \beta_0$

$b^t = \varepsilon - z^t$

$m^t = -z^t$

Abstract AMP	AMP in linear model
(h^t, q^t, b^t, m^t)	(z^t, β^t)

$t=0$: $h^0 = 0$

$q^0 = f_0(h^0) = -\beta_0$ $\quad z^0 = y - A\overset{0}{\beta^0} = y$

$b^0 = Aq^0 = A(-\beta_0) = \varepsilon - y = \varepsilon - z^0$

$m^0 = -z^0$

$t=1$: $h^1 = A^T m^0 - \overset{\xi_0=1}{\xi_0} q^0 = -A^T z^0 - (-\beta_0)$

$= \beta_0 - A^T z^0$

$q^1 = f_1(h^1) = \eta_0(\beta_0 - h^1) - \beta_0$

$= \eta_0(A^T z^0) - \beta_0$

$= \eta_0(A^T z^0 + \underset{0}{\beta^0}) - \beta_0$

$= \beta^1 - \beta_0$

$b^1 = Aq^1 - \lambda_1 m^0$

$= A(\beta^1 - \beta_0) - \left(-\frac{1}{\delta}\mathrm{Av}\left(\eta_0'(\beta_0 - h^1)\right)\right)(-z^0)$

$= A\beta^1 - A\beta_0 - z^0\left(\frac{1}{\delta}\mathrm{Av}\left(\eta_0'(\beta_0 - h^1)\right)\right)$

note: $z^1 = \underbrace{y - X\beta^1 + z^0\left(\frac{1}{\delta}\mathrm{Av}(\eta_0'(\dots))\right)}$

$= y - z^1 - A\beta_0$

$= \varepsilon - z^1$

Master thm for AMP

$b^0 = Aq^0 = Af_0(h^0)$

$\overset{d}{\approx} \sigma_0 Z_n \cdots \quad \sigma_0^2 = \frac{1}{n}\sum_{j=1}^{p} f_0^2(h_j^0)$

独立的 matrix A $\quad \approx \frac{1}{\delta}\mathbb{E} f_0^2(0, \bar{\beta}_0)$

$h^1 \overset{d}{\approx} A_1^T m^0 = \tau_1 Z_p \quad \sigma_1^2 = \frac{1}{n}\sum_{j=1}^{n} g_0^2(b_{0,i})$

$\approx \mathbb{E}\, g_0^2(\sigma_0 Z)$

$b^1 \overset{d}{\approx} \tilde{A}_1 q^1 = \sigma_1 Z_n \quad \sigma_1^2 = \frac{\|q^1\|^2}{n} = \frac{\|f_1(h^1)\|^2}{n}$

$$\begin{cases} b^0 \overset{d}{\approx} \sigma_0 Z_n, & \sigma_0^2 = \frac{1}{\delta}\mathbb{E} f_0^2(0, \bar{\beta}_0) \\ h^1 \overset{d}{\approx} \tau_1 Z_p & \tau_1^2 = \mathbb{E}\, g_0^2(\sigma_0 Z) \\ b^1 \overset{d}{\approx} \sigma_1 Z_n & \sigma_1^2 = \frac{1}{\delta}\mathbb{E} f_1^2(\tau_1 Z, \bar{\beta}_0) \end{cases}$$

$\sigma_t^2 = \frac{1}{\delta}\mathbb{E} f_t^2(\tau_t Z, \bar{\beta}_0)$

$\tau_{t+1}^2 = \mathbb{E}\, g_t^2(\sigma_t Z)$ $\quad$ State evolution

(P15)

$$\beta^{t+1} \overset{d}{\approx} \eta_t(\beta_0 + \tau_t Z_p)$$

where
$$\begin{cases} \tau_0^2 = \sigma^2 + \frac{1}{\delta}\mathbb{E}\beta_0^2 \\ \tau_{t+1}^2 = \sigma^2 + \frac{1}{\delta}\mathbb{E}(\eta_t(\bar\beta_0 + \tau_t Z) - \bar\beta_0)^2 \end{cases}$$

$\because$
$$\begin{cases} h^{t+1} = \beta_0 - (A^T z^t + \beta^t) \text{ by identification} \\ \beta^{t+1} = \eta_t(A^T z^t + \beta^0) \end{cases}$$

by master thm

$$\beta^{t+1} = \eta_t(\beta_0 - h^{t+1}) \overset{d}{=} \eta_t(\beta_0 + \tau_{t+1} Z_p).$$

$$\begin{cases} f_t(h, \beta_0) = \eta_{t-1}(\beta_0 - h) - \beta_0 \\ g_t(b) = b - \varepsilon \end{cases}$$ 选取 f_t, g_t

by state evolution

$$\sigma_0^2 = \frac{1}{\delta}\mathbb{E} f_0^2(0, \bar\beta_0) = \frac{1}{\delta}\mathbb{E}\bar\beta_0^2$$

$$\tau_1^2 = \mathbb{E} g_1^2(\sigma_0 Z) = \mathbb{E}(\sigma_0 Z - \varepsilon_1)^2 = \sigma^2 + \sigma_0^2 = \sigma^2 + \frac{1}{\delta}\mathbb{E}\bar\beta_0^2$$

$$\sigma_1^2 = \frac{1}{\delta}\mathbb{E}(\eta_0(\bar\beta_0 + \tau_1 Z) - \bar\beta_0)^2$$

$$\tau_2^2 = \sigma^2 + \frac{1}{\delta}\mathbb{E}(\eta_0(\bar\beta_0 + \tau_1 Z) - \bar\beta_0)^2$$

Application to Lasso

取 η 为 soft-thresholding func.

$$\eta_t(x) = \eta_1(x; \theta_t).$$

Suppose (β, z) is a fixed point for linear AMP with threshold θ_t

fixed point
$$\begin{cases} \beta_* = \eta_1(\beta_* + A^T z_*; \theta_*) \\ z_* = y - A\beta_* + \rho z_*, \quad \rho \overset{\Delta}{=} \frac{1}{\delta}\overline{Av}[\eta_1'(\beta_t + A^T z_*; \theta_*)] \end{cases}$$

threshold func.

claim $\beta = \eta_1(r; \theta) \Leftrightarrow \exists v(\beta) \in \partial\|\cdot\|_1(\beta)$ (scalar)

s.t. $\beta + \theta \cdot v(\beta) = r$

$$\beta_* + \theta_* \cdot v(\beta_*) = \beta_* + A^T z_* = \beta_* + A^T\left(\frac{1}{1-\rho}(y - A\beta_*)\right)$$

$$\Leftrightarrow A^T(y - A\beta_*) = (1-\rho)\theta_* v(\beta_*)$$ $v(\cdot)$: sub-gradient

recall the F.O.C. of Lasso

$$\nabla_\beta\left(\frac{1}{2}\|y - A\beta\|^2 + \lambda\|\beta\|_1\right) = 0$$

$$\Leftrightarrow -A^T(y - A\beta^{lasso}) + \lambda v(\beta^{Lasso}) = 0$$

$$\boxed{\lambda = (1-\rho)\theta_* = \theta_*\left(1 - \frac{1}{\delta}\overline{Av}(\eta_1'(\beta_* + A^T z_*; \theta_*))\right)}$$

1

(P16) WT know how to choose θ_t

$$F(\tau^2,\theta) \triangleq \sigma^2 + \frac{1}{\delta}\mathbb{E}(\eta_1(\bar\beta_0+\tau Z;\theta)-\bar\beta_0)^2$$

$\theta_t = \alpha\tau_t$, α to be deterministic

~~Rec~~

Recurison for τ_t

$$\tau_{t+1}^2 = F(\tau_t^2, \alpha\tau_t)$$

Fact 1: $\alpha_{min}(\delta)$ = unique non-negative solution to

$$(1+\alpha^2)\Phi(-\alpha) - \alpha\varphi(\alpha) = \frac{\delta}{2}$$

normal dist.

唯一

Then $\forall \sigma^2>0$, $\alpha > \alpha_{min}(\delta)$, $\exists!\ \tau_* = \tau_*(\alpha)$

s.t. $\tau_*^2 = f(\tau_*^2, \alpha\tau_*)$, $\tau_t \to \tau_*(\alpha)$

(Donoho, Montonari, Maleki, PNAS 20xx)

$$\lambda(\alpha) \triangleq \alpha\tau_*(\alpha)\left(1-\frac{1}{\delta}\mathbb{E}\eta_1'(\bar\beta_0+\tau_*(\alpha)Z;\alpha\tau_*(\alpha))\right)$$

For a given λ, want to solve inverse map

$$\alpha(\lambda) \in \{\alpha\in(\alpha_{min}(\delta),\infty) : \lambda(\alpha)=\lambda\}$$

Fact 2: $(\lambda) \neq \phi$

For a given $\lambda>0$

choose $\alpha = \alpha(\lambda) \in (\alpha_{min}(\delta),\infty)$ (depend on $\bar\beta_0$)

$\theta_t \triangleq \alpha\tau_*$ this is a theoritical choice

① $$\frac{\lambda}{\alpha\tau_*} = 1 - \frac{1}{\delta}\mathbb{E}\eta_1'(\bar\beta_0+\tau_* Z;\alpha\tau_*)$$

Application to robust regression

(R) $$\hat\beta = \underset{\beta\in\mathbb{R}^p}{\arg\min}\sum_{i=1}^n \rho(y_i - \langle A_i, \beta\rangle)$$

score: $\psi = \rho'$

$$e_\rho(z) = \min\{\rho(x)+\tfrac{1}{2}\|x-z\|^2\}$$

effective score: $\Psi(z;b) \doteq \frac{d}{dz} e_{b\rho}(z)$

$$= z - \mathrm{prox}_{b\rho}(z)$$

$\beta^0 = 0$

for $t=0,1,\dots$

(R1)

$$r^t = y - A\beta^t + \Psi(r^{t-1}; b_{t-1})$$

b_t = unique solution to $\frac{1}{\delta} = \frac{1}{n}\sum_{i=1}^n \Psi'(r_i^t; b_t)$

$$\beta^{t+1} = \beta^t + \delta A^T\Psi(r^t; b_t)$$

$$r^* = y - A\beta_* + \Psi(r_*; b_*)$$

$$\Leftrightarrow y - A\beta_* = r_* - \Psi(r_*; b_*)$$

$$\Rightarrow (b_*\rho)'(y-A\beta_*) = \rho'(r_* - \Psi(r_*; b_*)) = (b_*\rho)'(\mathrm{prox}_{b_*\rho}(r_*))$$

$$F'(\mathrm{prox}_F(x)) = x - \mathrm{prox}_F(x) \qquad = r_* - \mathrm{prox}_{b_*\rho}(r_*)$$

$$\Rightarrow \Psi(r_*; b_*) = b_*\rho'(y - A\beta_*)$$

② $$0 = \delta A^T\Psi(r_*, b_*) = \delta b_* A^T\rho'(y - A\beta_*)$$

(17)

$$\frac{1}{\delta} = \mathbb{E}\Psi'(\varepsilon + \tau_t Z; \bar{b}_t)$$

RI'

$$\bar{r}^t = \varepsilon - A\bar{\beta}^t + \Psi(\bar{r}^{t-1}; \bar{b}_{t-1})$$

$$\bar{\beta}^{t+1} = \bar{\beta}^t + \delta A^T \Psi(\bar{r}^t; \bar{b}_t)$$

$\bar{r}^t - \varepsilon \to \bar{B}^t$, $\bar{\beta}^t \to \bar{\theta}^t$

R2

$$\bar{B}^t = -A\bar{\theta}^t + \Psi(\varepsilon + \bar{B}^{t-1}; \bar{b}_{t-1})$$

$$\bar{\theta}^{t+1} = \delta A^T \Psi(\varepsilon + \bar{B}^t, \bar{b}_t) + \bar{\xi}_t \bar{\theta}^t \qquad \bar{\xi}_t = \delta\left(\frac{1}{n}\sum_{i=1}^{n} \Psi'(\varepsilon_i + \bar{B}_i^t, \bar{b}_t)\right]$$

$$b^t = Af_t(h^t) - \lambda_t g_{t-1}(b^{t-1}), \ \lambda_t = \frac{1}{\delta}Av(f_t'(h^t))$$

$$h^{t+1} = A^T g_t(b^t) - \xi_t f_t(h^t) \qquad \xi_t = Av(g_t'(b^t))$$

Identification

$$\begin{cases} f_t(h) = -h \\ g_t(b) = \delta\Psi(\varepsilon + h; \bar{b}_t) \end{cases}$$

Then $\lambda_t = -1$ $\xi_t = \bar{\xi}_t$

state evolution

$$\bar{B}^t \overset{d}{=} \sigma_t Z_n \quad \sigma_t^2 = \frac{1}{\delta}\mathbb{E}f_t^2(\tau_t Z) = \frac{\tau_t^2}{\delta}$$

$$\bar{\theta}^{t+1} \overset{d}{=} \tau_{t+1} Z_p, \ \tau_{t+1}^2 = \mathbb{E}g_t^2(\sigma + Z) = \delta^2\mathbb{E}\Psi^2(\varepsilon + \sigma_t Z; \bar{b}_t)$$

$$\boxed{\frac{1}{\delta} = \mathbb{E}\Psi'(\varepsilon + \sigma_t Z; \bar{b}_t) \ ① \qquad \Rightarrow \sigma_t^2 = \delta\,\mathbb{E}\Psi^2(\varepsilon + \sigma_t Z; \bar{b}_t)] \ ②}$$

for robust regression

$$\xi_t = \delta\,\mathbb{E}\Psi'(\varepsilon + \sigma_t Z; \bar{b}_t) = 1$$

$$\bar{\theta}^t \xrightarrow[t\to\infty]{} \beta \simeq N(0, \sigma\sqrt{\delta} + x Z_p)$$

① $b^0 = Aq^0 \overset{d}{=} \frac{\|q^0\|}{\sqrt{n}} Z_n = \frac{\|f_0(h^0)\|}{\sqrt{n}} Z_n = \sigma_0 Z_n \quad \sigma_0^2 = \frac{1}{\delta}\mathbb{E}f_0^2(0, \bar{\beta}_0)$

② $h^1 = A^T m^0 - \xi_0 q^0 \left(\overset{d}{\approx} \tilde{A}^T m^0\right)$

$$= A^T g(Aq^0) - \xi_0 q^0$$

$$h^1 = (P_{q^0} + P_{q^0}^\perp)A^T m^0 - \xi_0 q^0 \left(P_{q^0} = \frac{q^0 (q^0)^T}{\|q^0\|^2}\right)$$

$$= P_{q^0}^\perp A^T m^0 + (P_{q^0}A^T m^0 - \xi_0 q^0)$$

$$\overset{d}{=} \underset{①}{P_{q^0}^\perp \tilde{A} m^0} + \underset{②}{\cdots} \quad (② \text{ can be ignore})$$

$$① = \tilde{A} g(b^0) = P_{q^0}\tilde{A}^T g(b^0)$$

$$q^0 \frac{\langle q^0, \tilde{A}^T m^0 \rangle}{\|q^0\|^2} \quad \langle q^0, \tilde{A}^T m^0\rangle = O_p(\sqrt{n}), \quad \|q^0\|^2 = O(p)$$

(note: $\mathbb{E}\langle a, Ab\rangle^2 = \frac{1}{n}\|a\|^2\|b\|^2$)

$$= \frac{\|g(b^0)\|}{\sqrt{n}} Z_p$$

$$= \tau_1 Z_p$$

(P18)

$$② = \frac{q_h^0 (q_h^0)^T}{\|q_h^0\|^2} A^T m^0 - \xi_0 q_h^0$$

$$= q_h^0 \left(\frac{(A q_h^0)^T m^0}{\|q_h^0\|^2} - \xi_0 \right)$$

$$= \frac{q_h^0}{\|q_h^0\|^2/n} \left(\frac{1}{n} \langle b^0, g_0(b^0) \rangle - \xi_0 \frac{\|q_h^0\|^2}{n} \right)$$

$$b^0 \approx \sigma_0 Z_n = \mathbb{E} g_0'(\sigma_0 Z_n) \sigma_0^2 \qquad \xi_0 = \mathrm{Av}(g_0'(b^0))$$

$$\mathbb{E} g_0'(\sigma_0 Z_n)$$

$$= q^0 \cdot O(n^{-\frac{1}{2}})$$