

Lecture 12: The LASSO risk.

① The LASSO problem.

Signal: $x_0 \in \mathbb{R}^d$, Noise: $w \in \mathbb{R}^n$, Sensing matrix: $A \in \mathbb{R}^{n \times d}$.

Response: $y = Ax_0 + w \in \mathbb{R}^n$.

Assumption 1: $A_{ij} \sim \text{iid } N(0, \frac{1}{n})$, $w_i \sim \text{iid } N(0, \sigma^2)$, $\frac{1}{d} \sum_{j=1}^d \delta_{x_{0,j}} \rightarrow \mathbb{P}_0$ (potentially sparse).

Goal: Observe y and A , estimate x_0 .

LASSO estimator:

$$\hat{x} = \underset{x}{\operatorname{argmin}} \frac{1}{2n} \|y - Ax\|_2^2 + \frac{\lambda}{n} \|x\|_1.$$

② Main result (to be shown using the replica method today).

Theorem: Consider the asymptotic regime $n/d \rightarrow s \in (0, \infty)$ as $d \rightarrow \infty$. Let (A, x_0, w, y) be a sequence of problem instances satisfy Assumption 1. Let $\hat{x}$ be the LASSO estimator.

(A) We have $\mathbb{E}$ is w.r.t. A, w .

$$\lim_{\substack{d, n \rightarrow \infty \\ n/d \rightarrow s}} \mathbb{E} \|\hat{x} - x_0\|_2^2 / d = \mathbb{E}_{(x_0, G) \sim \mathbb{P}_0 \times N(0,1)} [(\eta(x_0 + \tau_* G; \tau_* \alpha_*) - x_0)^2]$$

where $\eta(x, s) = \operatorname{sign}(x) \cdot (|x| - s)_+$ is the soft thresholding function. $\leftarrow$ shrink to 0.

Here, we define a function $\tau(\alpha)$ to be the largest solution of

$$\tau^2 = \sigma^2 + s^{-1} \mathbb{E}_{(x_0, G) \sim \mathbb{P}_0 \times N(0,1)} [(\eta(x_0 + \tau G; \alpha \tau) - x_0)^2].$$

and denote α_* by the unique non-negative solution of

$$\lambda = \alpha \cdot \tau(\alpha) \cdot \left\{ 1 - s^{-1} \mathbb{E} [\eta'(x_0 + \tau(\alpha) G; \alpha \cdot \tau(\alpha))] \right\}.$$

Hereby we define $\tau_* = \tau(\alpha_*)$.

Self consistent equation.

$$\begin{aligned} x_0 &\sim \mathbb{P}_0, \\ Y &= x_0 + \tau_* G \\ \hat{x} &= \eta(Y; \alpha_* \tau_*) \end{aligned}$$

(B) Moreover, for any pseudo-Lipschitz function $\psi: \mathbb{R}^2 \rightarrow \mathbb{R}$, we have

$$\lim_{\substack{d \rightarrow \infty \\ n/d \rightarrow s}} \mathbb{E} \left[\frac{1}{d} \sum_{j=1}^d \psi(\hat{x}_j, x_{0,j}) \right] = \mathbb{E}_{(x_0, G) \sim \mathbb{P}_0 \times N(0,1)} [\psi(\eta(x_0 + \tau_* G; \alpha_* \tau_*); x_0)].$$

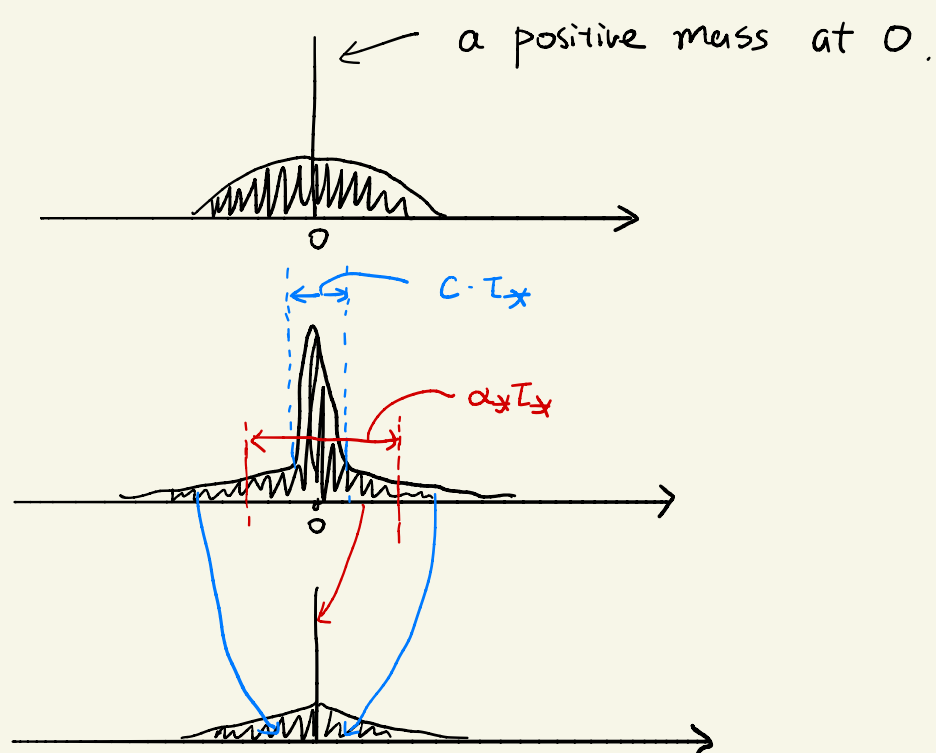
The solution of a one dim problem

Heuristic result: [Guo, Baron, Shamai, 2009], [Kabashima, Wadayama, Tanaka, 2009]

Rigorous proof: [Bayati, Montanari, 2010] $\leftarrow$ AMP

Alternative proof: CGMT, leave-one-out.

X_0 distribution



$$Y = X_0 + T_x G$$

$$\hat{X} = \eta(Y; \alpha T_x)$$

Application: Want to estimate the support of X_0

$$S = \{j : X_{0,j} \neq 0\}$$

using the support of Lasso solution

$$\hat{S}(\lambda) = \{j : \hat{x}_{\lambda,j} \neq 0\}.$$

False discovery proportion:

$$FDP(\lambda) = \frac{|\{j : \hat{x}_{\lambda,j} \neq 0, X_{0,j} = 0\}|}{|\{j : \hat{x}_{\lambda,j} \neq 0\}|}$$

True positive proportion:

$$TPP(\lambda) = \frac{|\{j : \hat{x}_{\lambda,j} \neq 0, X_{0,j} \neq 0\}|}{|\{j : X_{0,j} \neq 0\}|}$$

Goal: choose a λ , s.t. $\mathbb{E}[FDP(\lambda)] \leq \alpha$, $\mathbb{E}[TPP(\lambda)]$ as large as possible.

Under the model assumption ($A_{ij} \stackrel{iid}{\sim} N(0, \frac{1}{n})$, $w_i \stackrel{iid}{\sim} N(0, \sigma^2)$)

$$\frac{1}{d} |\{j : \hat{x}_{\lambda,j} \neq 0, X_{0,j} = 0\}| = \frac{1}{d} \sum_{j=1}^d \mathbb{1}\{\hat{x}_{\lambda,j} \neq 0, X_{0,j} = 0\}$$

$$\rightarrow \mathbb{P}(X_0 = 0, \eta(X_0 + T_x G; \alpha T_x) \neq 0).$$

$$X_0 \sim P_0 \perp G \sim N(0, 1)$$

$$\frac{1}{d} |\{j : \hat{x}_{\lambda,j} \neq 0\}| = \frac{1}{d} \sum_{j=1}^d \mathbb{1}\{\hat{x}_{\lambda,j} \neq 0\}$$

$$\rightarrow \mathbb{P}(\eta(X_0 + T_x G; \alpha T_x) \neq 0).$$

$$\Rightarrow \lim_{d \rightarrow \infty} \mathbb{E}[FDP(\lambda)] = \frac{\mathbb{P}(X_0 = 0, \eta(X_0 + T_x G; \alpha T_x) \neq 0)}{\mathbb{P}(\eta(X_0 + T_x G; \alpha T_x) \neq 0)}.$$

Similarly for $TPP(\lambda)$.

This gives a "simple" expression for the power.

③. The free energy trick and the replica trick

The configuration space: $\Omega = \mathbb{R}^d$, $\nu_0 = \text{Lebesgue measure}$

Hamiltonian: $H_{\lambda,h}(x) = \frac{1}{2} \|y - Ax\|_2^2 + \lambda \|x\|_1 + h \sum_{j=1}^d \psi(x_j, x_{0,j})$

Randomness: $y = \underline{A} x_0 + \underline{w}$ 2 source of randomness.

free entropy density: $\varphi(\beta, \lambda, h) \equiv \lim_{d \rightarrow \infty} \left[\frac{1}{d} \log \int_{\mathbb{R}^d} \exp\{-\beta H_{\lambda,h}(x)\} dx \right]$

Limiting ensemble average:

$$\lim_{d \rightarrow \infty} \mathbb{E} \left[\frac{1}{d} \sum_{j=1}^d \psi(x_j, x_{0,j}) \right] \stackrel{\text{The free energy trick.}}{=} \partial_h \lim_{\beta \rightarrow \infty} \left[-\frac{1}{\beta} \varphi(\beta, \lambda, h) \right] \Big|_{h=0}$$

(A) The $d \rightarrow \infty$ limit.

$$S(k, \beta, \lambda, h) \equiv \lim_{d \rightarrow \infty} \frac{1}{d} \log \mathbb{E} \left[\left(\int_{\mathbb{R}^d} \exp\{-\beta H_{\lambda,h}(x)\} dx \right)^k \right].$$

(B) The $k \rightarrow 0$ limit.

$$\varphi(\beta, \lambda, h) = \lim_{k \rightarrow 0} \left[\frac{1}{k} S(k, \beta, \lambda, h) \right].$$

(C) The h differentiation and $\beta \rightarrow \infty$ limit

$$m_x(\lambda) = \left\{ \partial_h \lim_{\beta \rightarrow \infty} \left[-\frac{1}{\beta} \varphi(\beta, \lambda, h) \right] \right\} \Big|_{h=0}.$$

④ The calculation

(A) The $d \rightarrow \infty$ limit. ($n/d \rightarrow 8$).

$$x^0 = x_0 \in \mathbb{R}^d$$

$$\begin{aligned} \mathbb{E}[Z_n^k] &= \mathbb{E} \int_{\mathbb{R}^{d \times k}} \exp \left\{ -\beta \sum_{a=1}^k \|y - Ax^a\|_2^2 / 2 - \beta \lambda \sum_{a=1}^k \|x^a\|_1 - \beta h \sum_{j=1}^d \psi(x_j^a, x_j^0) \right\} \prod_{a=1}^k dx^a \\ &= \int_{\mathbb{R}^{d \times k}} \underbrace{\mathbb{E} \left[\exp \left\{ -\beta \sum_{a=1}^k \|y - Ax^a\|_2^2 / 2 \right\} \right]}_{\mathbb{E}(x)} \times \exp \left\{ -\beta \lambda \sum_{a=1}^k \|x^a\|_1 - \beta h \sum_{j=1}^d \psi(x_j^a, x_j^0) \right\} \prod_{a=1}^k dx^a \end{aligned}$$

w.r.t. A, w

w.r.t. A, w

$$E(x) = \mathbb{E} \left[\exp \left\{ -\beta \sum_{a=1}^k \|y - Ax^a\|_2^2 / 2 \right\} \right]$$

$$= \mathbb{E} \left[\exp \left\{ -\beta \sum_{a=1}^k \|w - A(x^0 - x^a)\|_2^2 / 2 \right\} \right]$$

$$= \mathbb{E} \left[\exp \left\{ -\beta k \|w\|_2^2 / 2 + \beta \langle w, \sum_{a=1}^k A(x^0 - x^a) \rangle - \frac{\beta}{2} \sum_{a=1}^k \|A(x^0 - x^a)\|_2^2 / 2 \right\} \right]$$

density of w

$$= \mathbb{E} \int_{\mathbb{R}^n} \frac{1}{\sqrt{2\pi}^n \sigma^n} \exp \left\{ -\beta k \|\varepsilon\|_2^2 / 2 - \frac{\|\varepsilon\|_2^2}{2\sigma^2} + \beta \langle \varepsilon, \sum_{a=1}^k A(x^0 - x^a) \rangle \right\} d\varepsilon$$

w.r.t. A
 ε is dummy for w

$$\times \exp \left\{ -\frac{\beta}{2} \sum_{a=1}^k \|A(x^0 - x^a)\|_2^2 \right\}.$$

Gaussian integration

$$\bar{\sigma}^2 = \frac{\sigma^2}{1 + \beta k \sigma^2}.$$

C_n

$$= \underbrace{(k\beta\sigma^2 + 1)^{-\frac{n}{2}}}_{C_n} \times \mathbb{E} \left[\exp \left\{ \beta^2 \left\| \sum_{a=1}^k A(x^0 - x^a) \right\|_2^2 \bar{\sigma}^2 / 2 - \frac{\beta}{2} \sum_{a=1}^k \|A(x^0 - x^a)\|_2^2 / 2 \right\} \right].$$

Note $A = \begin{bmatrix} u_1^T \\ \vdots \\ u_n^T \end{bmatrix}$, $u_i \sim \text{iid } N(0, \frac{1}{n} \mathbf{I}_d) \Rightarrow \|A(x^0 - x^a)\|_2^2 = \sum_{i=1}^n (u_i^T (x^0 - x^a))^2$

w.r.t. $u \sim N(0, \frac{1}{n} \mathbf{I}_d)$

$$= C_n \cdot \mathbb{E} \left[\exp \left\{ \beta^2 \left(\sum_{a=1}^k u^T (x^0 - x^a) \right)^2 \bar{\sigma}^2 / 2 - \frac{\beta}{2} \sum_{a=1}^k [u^T (x^0 - x^a)]^2 \right\} \right]^n$$

Define $G_a = u^T (x^0 - x^a)$, Then $\mathbb{E}[G_a G_b] = (x^0 - x^a)^T (x^0 - x^b) / n \equiv \Sigma_{ab}(x)$

$$= C_n \cdot \left[\int \frac{1}{(2\pi)^{\frac{k}{2}} \det(\Sigma)^{\frac{1}{2}}} \exp \left\{ \beta^2 \left(\sum_{a=1}^k G_a \right)^2 \bar{\sigma}^2 / 2 - \frac{\beta}{2} \sum_{a=1}^k G_a^2 - \underbrace{G^T \Sigma^{-1} G / 2}_{\text{density of } G} \right\} dG \right]^n$$

Gaussian integration

$$= C_n \cdot \left[\det \left(\Sigma^{-1} + \beta \mathbf{I}_k - \beta^2 \bar{\sigma}^2 \mathbf{1}_k \mathbf{1}_k^T \right) \det(\Sigma) \right]^{-\frac{n}{2}}$$

$$= (1 + k\beta\sigma^2)^{-\frac{n}{2}} \left\{ \det \left[\mathbf{I}_k + \Sigma \left(\beta \mathbf{I}_k - \beta^2 \bar{\sigma}^2 \mathbf{1}_k \mathbf{1}_k^T \right) \right] \right\}^{-\frac{n}{2}}$$

Define $\bar{Q}(x) = (\langle x^a, x^b \rangle / d)_{a,b \in [k]}$.

$$\bar{\mu}(x) = (\langle x^a, x^0 \rangle / d)_{a \in [k]}.$$

$$p = \|x^0\|_2^2 / d$$

$$\Rightarrow \Sigma(x) = \bar{Q}(x) - \bar{\mu}(x) \mathbf{1}_k^T + \mathbf{1}_k \bar{\mu}(x)^T + p \mathbf{1}_k \mathbf{1}_k^T.$$

$$\Rightarrow E(x) = E(\bar{Q}(x), \bar{\mu}(x)).$$

← abuse notation

$$\Rightarrow \mathbb{E}[Z_n^n] = \int_{\mathbb{R}^{d \times k}} E(\bar{Q}(x), \bar{\mu}(x))$$

$$\times \exp \left\{ -\beta \lambda \sum_{a=1}^k \|x^a\|_1 - \beta h \sum_{j=1}^d \psi(x_j^a, x_j^0) \right\} \prod_{a=1}^k dx^a$$

$$= \int dQ d\mu E(Q, \mu) \times \text{Ent}(Q, \mu)$$

$$S(k, \beta, \lambda, h) \doteq \sup_{Q, \mu} \left[\frac{1}{d} \log E(Q, \mu) \right] \times \left[\frac{1}{d} \log \text{Ent}(Q, \mu) \right].$$

where $\text{Ent}(Q, \mu) = \int \delta(\bar{Q}(x) - Q) \delta(\bar{\mu}(x) - \mu) \times \exp\left\{-\beta\lambda \sum_{a=1}^k \|x^a\|_1 - \beta h \sum_{j=1}^d \psi(x_j^a, x_j^0)\right\} \prod_{a=1}^k dx^a$

$\delta = n/d$

$\frac{1}{d} \log E(Q, \mu) = -\frac{\delta}{2} \log(1 + k\beta\sigma^2) - \frac{\delta}{2} \log \det[I_k + (Q - \mu 1^T + 1 \mu^T + \rho 11^T)(\beta I_k - \beta^2 \bar{\sigma}^2 11^T)]$
 $e(Q, \mu)$ $\bar{\sigma}^2 = \frac{\sigma^2}{1 + k\beta\sigma^2}$

By delta identity formula + saddle point approx :

$\text{Ent}(Q, \mu) = \text{ext}_{r_{ab}, \xi_a} \int_{\mathbb{R}^{dk}} \exp\left\{-\sum_{ab} r_{ab} (d \cdot Q_{ab} - \langle x^a, x^b \rangle)/2 - \sum_a \xi_a (d \cdot \mu_a - \langle x^a, x^0 \rangle) - \beta\lambda \sum_{a=1}^k \|x^a\|_1 - \beta h \sum_{j=1}^d \psi(x_j^a, x_j^0)\right\} \prod_{a=1}^k dx^a$
 $= \text{ext}_{r_{ab}, \xi_a} \left[\frac{d}{\prod_{j=1}^d} \int_{\mathbb{R}^k} \exp\left\{-\sum_{ab} r_{ab} (Q_{ab} - x^a x^b)/2 - \sum_a \xi_a (\mu_a - x^a x_j^0) - \beta\lambda \sum_{a=1}^k |x^a| - \beta h \sum_{j=1}^d \psi(x^a, x_j^0)\right\} \prod_{a=1}^k dx^a \right]$

Law of large numbers,
uniform in r_{ab}, ξ_a

$\frac{1}{d} \log \text{Ent} = \text{ext}_r \frac{1}{d} \sum_{j=1}^d \log M(x_j^0, r) \rightarrow \text{ext}_r \mathbb{E}_{x^0 \sim p_0} [\log M(x^0, r)]$

$\Rightarrow \frac{1}{d} \log \text{Ent} \rightarrow \text{ext}_{r_{ab}, \xi_a} \mathbb{E}_{x^0} \left[\log \int_{\mathbb{R}^k} \exp\left\{-\sum_{ab} r_{ab} (Q_{ab} - x^a x^b)/2 - \sum_a \xi_a (\mu_a - x^a x^0) - \beta\lambda \sum_{a=1}^k |x^a| - \beta h \sum_{j=1}^d \psi(x^a, x_j^0)\right\} \prod_{a=1}^k dx^a \right]$

$\Rightarrow S(k, \beta, \lambda, h) = \text{ext}_{Q, \mu, r, \xi} [e(Q, \mu) + \text{ent}(Q, \mu, r, \xi)]$ $Q, r \in \mathbb{R}^{k \times k}$
 $\mu, \xi \in \mathbb{R}^k$

where

$e(\cdot) = -\frac{\delta}{2} \log(1 + k\beta\sigma^2) - \frac{\delta}{2} \log \det[I_k + (Q - \mu 1^T + 1 \mu^T + \rho 11^T)(\beta I_k - \beta^2 \bar{\sigma}^2 11^T)]$
 $\bar{\sigma}^2 = \frac{\sigma^2}{1 + k\beta\sigma^2}$

$\text{ent}(\cdot) = \mathbb{E}_{x^0} \left[\log \int_{\mathbb{R}^k} \exp\left\{-\sum_{ab} r_{ab} (Q_{ab} - x^a x^b)/2 - \sum_a \xi_a (\mu_a - x^a x^0) - \beta\lambda \sum_{a=1}^k |x^a| - \beta h \sum_{j=1}^d \psi(x^a, x_j^0)\right\} \prod_{a=1}^k dx^a \right]$

(B) The $k \rightarrow 0$ limit.

Replica symmetric ansatz.

$Q = \begin{bmatrix} q_1 & & q_0 \\ & \ddots & \\ q_0 & & q_1 \end{bmatrix}, \quad r = \begin{bmatrix} r_1 & & r_0 \\ & \ddots & \\ r_0 & & r_1 \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix}, \quad \xi = \begin{bmatrix} \xi \\ \vdots \\ \xi \end{bmatrix}$

Consider that later we'll take $\beta \rightarrow \infty$.

do reparametrization here:

will be clear later.

$$q_0 = q, \quad q_1 = q + \frac{w}{\beta}, \quad r_0 = \beta^2 \rho, \quad r_1 = \beta^2 \rho - \beta v$$

$$\mu = \mu, \quad \xi = \beta \zeta$$

$$\Rightarrow \lim_{k \rightarrow 0} \frac{1}{k} e = -\frac{\delta}{2} \log(1 + \frac{w}{\delta}) - \frac{\beta \delta}{2(\delta + w)} (\rho - 2\mu + q + \delta \sigma^2) \equiv \bar{e}$$

Need to use $\log \det (a I_k + b \mathbf{1} \mathbf{1}^T) = (k-1) \log b + \log(a + kb)$

and

$$\text{ent} = -[\beta^2 k^2 \rho q - \beta k v q + \beta k \rho w - k v w]/2 - k \beta \mu \zeta$$

$$+ \mathbb{E}_{x^0} \left[\log \int_{\mathbb{R}^k} \exp \left\{ \underbrace{\frac{\beta^2 \rho}{2} \left(\sum_{a=1}^k x^a \right)^2}_{\text{}} - \frac{\beta v}{2} \sum_{a=1}^k (x^a)^2 + \beta \zeta \sum_{a=1}^k x^a x^0 \right. \right.$$

$$\left. \left. - \beta \lambda \sum_{a=1}^k |x^a| - \beta h \varphi(x^a, x^0) \right\} \prod_{a=1}^k dx^a \right]$$

use $\mathbb{E} \left[e^{\lambda \left(\sum_{a=1}^k x^a \right) G} \right] = \exp \left\{ \frac{\lambda^2}{2} \left(\sum_{a=1}^k x^a \right)^2 \right\}$

$$= -[\beta^2 k^2 \rho q - \beta k v q + \beta k \rho w - k v w]/2 - k \beta \mu \zeta$$

$$+ \mathbb{E}_{x^0} \left\{ \log \mathbb{E}_G \left[\int_{\mathbb{R}^k} \exp \left\{ \underbrace{\beta \sqrt{\rho} \cdot \sum_{a=1}^k x^a G}_{\text{}} - \frac{\beta v}{2} \sum_{a=1}^k (x^a)^2 + \beta \zeta \sum_{a=1}^k x^a x^0 \right. \right. \right.$$

$$\left. \left. - \beta \lambda \sum_{a=1}^k |x^a| - \beta h \varphi(x^a, x^0) \right\} \prod_{a=1}^k dx^a \right] \right\}$$

$$= -[\beta^2 k^2 \rho q - \beta k v q + \beta k \rho w - k v w]/2 - k \beta \mu \zeta$$

$$+ \mathbb{E}_{x^0} \log \mathbb{E}_G \left[\left(\int_{\mathbb{R}} \exp \left\{ \beta \sqrt{\rho} \cdot x \cdot G - \frac{\beta v}{2} \cdot x^2 + \beta \cdot \zeta \cdot x \cdot x^0 - \beta \lambda |x| - \beta h \varphi(x, x^0) \right\} dx \right)^k \right]$$

$$\Rightarrow \lim_{k \rightarrow 0} \frac{1}{k} \text{ent} = -[-\beta \chi q + \beta \rho w - v w]/2 - \beta \mu \zeta$$

$$+ \mathbb{E}_{G, x^0} \left[\log \int_{\mathbb{R}} \exp \left\{ \beta \sqrt{\rho} \cdot x \cdot G - \frac{\beta v}{2} \cdot x^2 + \beta \cdot \zeta \cdot x \cdot x^0 - \beta \lambda |x| - \beta h \varphi(x, x^0) \right\} dx \right]$$

$$\varphi(\beta, \lambda, h) = \underset{\substack{q, w, \mu \\ \rho, v, \zeta}}{\text{ext}} \left[\bar{e} + \overline{\text{ent}} \right]$$

$\overline{\text{ent}}$
//

$$(C) \quad f(\lambda, h) \equiv \lim_{\beta \rightarrow \infty} -\frac{1}{\beta} \varphi(\beta, \lambda, h)$$

$$= \text{ext}_{\substack{q, w, \mu \\ \rho, v, \zeta}} \left\{ \lim_{\beta \rightarrow \infty} \left[-\frac{1}{\beta} \bar{e} \right] + \lim_{\beta \rightarrow \infty} \left[-\frac{1}{\beta} \bar{e}_{\text{ext}} \right] \right\}.$$

$$\lim_{\beta \rightarrow \infty} \left[-\frac{1}{\beta} \bar{e} \right] = \frac{\delta}{2(\delta+w)} (p - 2\mu + q + \delta \sigma^2)$$

$$\lim_{\beta \rightarrow \infty} \left[-\frac{1}{\beta} \bar{e}_{\text{ext}} \right] = (-vq + pw)/2 + \mu\zeta$$

$$\lim_{\beta \rightarrow \infty} \frac{1}{\beta} \log \int \exp\{\beta U(x)\} dx = \max_x U(x)$$

$$- \mathbb{E}_{G, x^0} \left\{ \max_x \left[\sqrt{p} \cdot x \cdot G - \frac{v}{2} \cdot x^2 + \zeta \cdot x \cdot x^0 - \lambda |x| - h \psi(x, x^0) \right] \right\}$$

$$\Rightarrow f(\lambda, h) = \text{ext}_{\substack{q, w, \mu, \rho, v, \zeta}} f_{mf}(\lambda, h; q, w, \mu, \rho, v, \zeta).$$

$$f_{mf}(\lambda, h; q, w, \mu, \rho, v, \zeta) = \frac{\delta}{2(\delta+w)} (p - 2\mu + q + \delta \sigma^2) + (-vq + pw)/2 + \mu\zeta$$

$$- \mathbb{E}_{G, x^0} \left\{ \max_x \left[\sqrt{p} \cdot x \cdot G - \frac{v}{2} \cdot x^2 + \zeta \cdot x \cdot x^0 - \lambda |x| - h \psi(x, x^0) \right] \right\}$$

$$\frac{\partial f_{mf}}{\partial q} = \frac{\delta}{2(\delta+w)} - \frac{v}{2} = 0, \quad \frac{\partial f_{mf}}{\partial \mu} = -\frac{\delta}{\delta+w} + \zeta = 0$$

$$\Rightarrow v = \zeta = \frac{\delta}{\delta+w}$$

$$\Rightarrow f(\lambda, h) = \text{ext}_{\rho, v} f_{mf}(\lambda, h; \rho, v)$$

$$f_{mf}(\lambda, h; \rho, v) = \frac{v}{2} (p + \delta \sigma^2) + \frac{\rho \delta}{2} \left(\frac{1}{v} - 1 \right)$$

$$- \mathbb{E}_{G, x^0} \left\{ \max_x \left[\sqrt{p} \cdot x \cdot G - \frac{v}{2} \cdot x^2 + v \cdot x \cdot x^0 - \lambda |x| - h \psi(x, x^0) \right] \right\}$$

$$\partial f(\lambda, h)|_{h=0} = \partial_h f_{mf}(\lambda, h; \rho, v)|_{h=0} = \mathbb{E}_{G, x^0} [\psi(\hat{x}, x^0)]$$

$$\hat{x} = \arg \max_x \left[-\frac{v}{2} x^2 + x(v x_0 + \sqrt{p} G) - \lambda |x| \right]$$

$$= \arg \min_x \left[\frac{1}{2} \left(x - \left(x_0 + \frac{\sqrt{p}}{v} G \right) \right)^2 + \frac{\lambda}{v} |x| \right]$$

$$= \eta \left(x_0 + \frac{\sqrt{p}}{v} G; \frac{\lambda}{v} \right). \quad \leftarrow \text{soft-thresholding function c.f. Theorem.}$$

where v, ρ is s.t. $f_{mf}(\lambda, 0; \rho, v)$ is stationary.

$$\partial_p f_{mf} = \frac{\delta}{2} \left(\frac{1}{\nu} - 1 \right) - \frac{1}{2\sqrt{p}} \mathbb{E}[\hat{x} G] \xrightarrow{\text{integration by part.}} \eta(x_0 + \frac{\sqrt{p}}{\nu} G; \frac{\lambda}{\nu})$$

$$= \frac{\delta}{2} \left(\frac{1}{\nu} - 1 \right) - \frac{1}{2\nu} \mathbb{E}[\partial \eta(x_0 + \frac{\sqrt{p}}{\nu} G; \frac{\lambda}{\nu})] = 0$$

$$\partial_\nu f_{mf} = \frac{1}{2} (p + \delta \sigma^2) - \frac{p\delta}{2\nu^2} + \mathbb{E} \left[\frac{\hat{x}^2}{2} - x_0 \hat{x} \right]$$

$$= \frac{1}{2} \mathbb{E}[(\hat{x} - x_0)^2] + \frac{1}{2} \sigma^2 - \frac{p}{2\nu^2} = 0$$

$$\Rightarrow \begin{cases} \frac{p}{\nu^2} = \sigma^2 + \delta^{-1} \mathbb{E}[(\eta(x_0 + \frac{\sqrt{p}}{\nu} G; \frac{\lambda}{\nu}) - x_0)^2] \\ \nu = 1 - \delta^{-1} \mathbb{E}[\partial \eta(x_0 + \frac{\sqrt{p}}{\nu} G; \frac{\lambda}{\nu}) - x_0] \end{cases} \quad \begin{cases} \frac{p}{\nu^2} = \tau^2 \\ \nu = \frac{\lambda}{\alpha \tau} \end{cases}$$

This recovers the self consistent equation of the theorem.