EM Lasso

$$y = Xb + \varepsilon \leftarrow \varepsilon \sim N(0, s^2)$$

$$b_j \mid s_j^2 \sim N(0, s_j^2)$$

$$s_{b_j}^2 \sim \mathrm{Exp}\left(\tfrac{1}{\eta}\right)$$

↑ rate, so mean $= \eta$

$$\Rightarrow p(b_j) = \frac{1}{2}\sqrt{\frac{2}{\eta}}\exp\left(-\sqrt{\frac{2}{\eta}}|b_j|\right)$$

$$\left(E(|b_j|) = \sqrt{\frac{\eta}{2}}\right)$$

$$\log p(y, b, \underline{s}^2) = \frac{-1}{2s^2}\|y - Xb\|_2^2 - \frac{n}{2}\log(2\pi s^2)$$

$$- \sum_j \frac{b_j^2}{2 s_{b_j}^2} - \sum_j \frac{1}{2}\log 2\pi s_{b_j}^2$$

$$+ p \log\frac{1}{\eta} - \frac{1}{\eta}\sum_j s_{b_j}^2 \quad \leftarrow \text{prior on } s_{b_j}^2$$

$$p(s_j^2 \mid b_j) \propto \exp\left(-\frac{s_j^2}{\eta}\right)\frac{1}{\sqrt{s_j^2}}\exp\left(-\frac{b_j^2}{2s_j^2}\right)$$

$$= \frac{1}{\sqrt{s_j^2}}\exp\left[-\frac{s_j^2}{\eta} - \frac{b_j^2}{2s_j^2}\right]$$

ie if $w = s_j^2$

$$p(w \mid b) \propto \frac{1}{w^{\frac{1}{2}}}\exp\left[-\frac{w}{\eta} - \frac{b^2}{2w}\right]$$

$$E_q(\log p(y, b, \underline{s}^2)) + E_q(\log q(b)) + E_q(\log q(s))$$

$\hat{q}(b)$ = ridge, w/ prior precision $= E\left(\frac{1}{s_{b_j}^2}\right)$

$\dot{q}_j(s_j)$ = posterior given data $\sqrt{E(b_j^2)} \sim N(0, s_j)$

so for Lasso, posterior mean, just replace $|b|$ w/ $\sqrt{E(b^2)}$

$$E_q\left[\log p(y|b) + \log \frac{p(b|s)}{q(s)} + \log \frac{p(s|\eta)}{q(s)}\right] \cancel{- \log q(b)}$$

Optimize over η

$$F(\eta) = p \log \frac{1}{\eta} - \frac{1}{\eta} \underbrace{\sum_j E_q(s_{b_j}^2)}$$

$$F'(\tfrac{1}{\eta}) = p \frac{1}{(1/\eta)} - \sum_j E_q(s_{b_j}^2)$$

$$\Rightarrow \hat{\eta} = \frac{1}{p} \sum_j E_q(s_{b_j}^2) \quad \left[\text{intuition as prior mean is } \eta\right]$$

(Figueiredo gives $E_q\left(\frac{1}{s_{b_j}^2}\right)$, so we need $E(s_{b_j}^2)$ $= E(\tau)$ in Fig notation)

In Figueiredo's notation:

$$E(\tau) = \frac{\int_0^\infty \tau \, N(\beta; 0, \tau) \cancel{\tfrac{\gamma}{2}} \exp(-\tfrac{\gamma}{2}\tau) \, d\tau}{\int N(\beta; 0, \tau) \cancel{\tfrac{\gamma}{2}} \exp(-\tfrac{\gamma}{2}\tau) \, d\tau}$$

$(\gamma = \tfrac{1}{\eta})$

$$\text{top} = \int_0^\infty \tau \frac{1}{\sqrt{2\pi\tau}} \exp\left(-\frac{\beta^2 + \gamma\tau^2}{2\tau}\right)$$

$$= \int_0^\infty \tau^2 \left[\frac{1}{\sqrt{2\pi\tau^3}} \exp\left(-\frac{\beta^2 + \gamma\tau^2}{2\tau}\right)\right] d\tau$$

$E = \mu$
$Var = \mu^3/\lambda$
(densities)

$$\text{IG distn} = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left[-\frac{\lambda (x-\mu)^2}{2\mu^2 x}\right]$$

$$= \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left[-\frac{\lambda x}{2\mu^2} - \frac{\lambda \cancel{\mu^2}}{2\cancel{\mu^2} x}\right] \exp\left[+\frac{\lambda \cancel{2} x \cancel{\mu}}{\cancel{2}\mu^{\cancel{2}} \cancel{x}}\right]$$

$$= \sqrt{\lambda}\ \exp\left(\frac{\lambda}{\mu}\right)\left[\left(\frac{1}{\sqrt{2\pi x^3}}\exp\left[-\frac{\lambda x}{2\mu^2} - \frac{\lambda}{2x}\right]\right)\right]$$

We have: TOP

$$\int_0^\infty \tau^2 \left(\frac{1}{\sqrt{2\pi\tau^3}}\exp\left[-\frac{\gamma\tau}{2} - \frac{\beta^2}{2\tau}\right]\right) d\tau \qquad \begin{cases} \frac{\lambda}{\mu^2} = \gamma \\ \lambda = \beta^2 \end{cases}$$

$$\Rightarrow \mu = \sqrt{\frac{\lambda}{\gamma}} = \frac{\beta}{\sqrt{\gamma}}$$

$$= \int_0^\infty \tau^2 \frac{1}{\sqrt{\lambda}}\exp\left(-\frac{\lambda}{\mu}\right)\left[IG(\lambda, \mu)\right]$$

$$\lambda = \beta^2,\ \mu = |\beta|/\sqrt{\gamma}$$

$$= \frac{1}{\sqrt{\lambda}}\exp\left(-\frac{\lambda}{\mu}\right)\left[\mu^2 + \mu^3/\lambda\right]$$

$$\left(\frac{\gamma}{2} = \frac{1}{\eta}\right)$$

$$\mu = |\beta|\sqrt{\frac{\eta}{2}}$$

$$\frac{1}{\mu} = \frac{1}{|\beta|}\sqrt{2/\eta}$$

consistent with (33) in his notes.

$$\text{Bottom} = \int \frac{1}{\sqrt{2\pi\tau}}\exp\left(-\frac{\beta^2 + \gamma\tau^2}{2\tau}\right)$$

$$= \int \tau\left(\frac{1}{\sqrt{2\pi\tau^3}}\exp(\qquad)\right) d\tau$$

$$= \frac{1}{\sqrt{\lambda}}\exp\left(-\frac{\lambda}{\mu}\right)\left[\mu\right]$$

[but inconsistent with his (9) which says $\frac{\gamma}{|\beta|} = \frac{2}{\eta|\beta|}$]

$$\text{Ratio} = \mu + \frac{\mu^2}{\lambda} = \frac{|\beta|}{\sqrt{\gamma}} + \frac{1}{\gamma}$$

$$= |\beta|\sqrt{\frac{\eta}{2}} + \frac{\eta}{2}$$

[recall $\eta = E(s^{-1})$]

So $$\boxed{\hat{\eta}_{\text{new}} = \frac{1}{p}\sum_j\left(|\beta_j|\sqrt{\frac{\eta}{2}} + \frac{\eta}{2}\right)}$$

[note $\beta_j \sim \text{Exp}\left(\frac{\eta}{2}\right)$
$E(|\beta_j|) = \sqrt{\eta/2}$]

$$\sqrt{E(\beta^2)}$$

$$\hat{s}^2 = \frac{1}{n}E\left((y - Xb)^T(y - Xb)\right)$$

$$= \frac{1}{n}\left[y^Ty - 2b^TX^Ty + b^TX^TXb\right]$$

$$E(bb^T) = \Sigma + \mu\mu^T$$

$$b^T X^T X b = \sum_{ij} b_i (X^T X)_{ij} b_j$$

$$= \sum_{ijk} b_i b_j X_{ki} X_{kj}$$