Homework Week 13

5. We divide the data set $(\mathbf{y}, \mathbf{X})$ into m blocks $\{(\mathbf{y}_j, \mathbf{X}_j)\}_{j=1}^m$ such that

$$\frac{1}{2}||\mathbf{y} - \mathbf{X}\boldsymbol{\beta}||_2^2 = \frac{1}{2}\sum_{j=1}^m ||\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta}||_2^2$$

and reformulate the original problem as

$$\begin{array}{ll} \min\limits_{\boldsymbol{\theta}'_j s, \boldsymbol{\beta}} & \frac{1}{2}\sum\limits_{j=1}^m ||\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\theta}_j||_2^2 + \lambda||\boldsymbol{\beta}||_1 \\ \text{subject to} & \boldsymbol{\theta}_j = \boldsymbol{\beta} \text{ for } j = 1, 2, \cdots, m. \end{array}$$

We use the following iterative scheme to obtain the optimizer:

$$\begin{aligned} \boldsymbol{\theta}_i^{r+1} &= \arg\min_{\boldsymbol{\theta}_i}\left\{\frac{1}{2}||\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\theta}_j||_2^2 + \frac{\rho}{2}||\boldsymbol{\theta}_i - \boldsymbol{\beta}^r + \boldsymbol{\alpha}_i^r||_2^2\right\} \\ &= (\mathbf{X}_j^T\mathbf{X}_j + \rho\mathbf{I}_{p\times p})^{-1}[\mathbf{X}_j^T\mathbf{y}_j + \rho(\boldsymbol{\beta}^r - \boldsymbol{\alpha}_i^r)] \text{ for } i = 1, 2, \cdots, m \\ \boldsymbol{\beta}^{r+1} &= \arg\min_{\boldsymbol{\beta}}\left\{\lambda||\boldsymbol{\beta}||_1 + \frac{m\rho}{2}\left|\left|\boldsymbol{\beta} - \frac{1}{m}\sum_{i=1}^m\boldsymbol{\theta}_i^{r+1} - \frac{1}{m}\sum_{i=1}^m\boldsymbol{\alpha}_i^r\right|\right|_2^2\right\} \\ &= \mathrm{ST}_{\lambda/(m\rho)} \circ \left[\frac{1}{m}\sum_{i=1}^m(\boldsymbol{\theta}_i^{r+1} + \boldsymbol{\alpha}_i^r)\right] \\ \boldsymbol{\alpha}_i^{r+1} &= \boldsymbol{\alpha}_i^r + \boldsymbol{\theta}_j^{r+1} - \boldsymbol{\beta}^{r+1} \text{ for } i = 1, 2, \cdots, m, \end{aligned} \tag{1}$$

where $\mathrm{ST}_{\lambda/(m\rho)}(a) = \mathrm{sign}(a)(|a| - \lambda/(m\rho))_+$ is a soft-thresholding operator. We use the following criteron to stop the iterative scheme (1):

$$\sum_{j=1}^m\left(\frac{||\boldsymbol{\theta}_j^r - \boldsymbol{\beta}^r||_2}{\sqrt{mp}}\right) + \frac{\rho||\boldsymbol{\beta}^r - \boldsymbol{\beta}^{r-1}||_2}{\sqrt{p}} \le 5\times 10^{-3} \qquad \text{or} \qquad r > 500.$$

For numerical results, please see Figures 1 and 2.

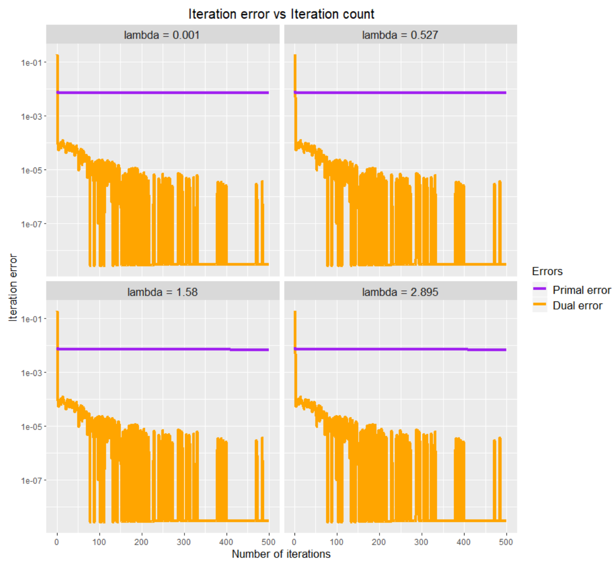

Figure 1: Simulation results.

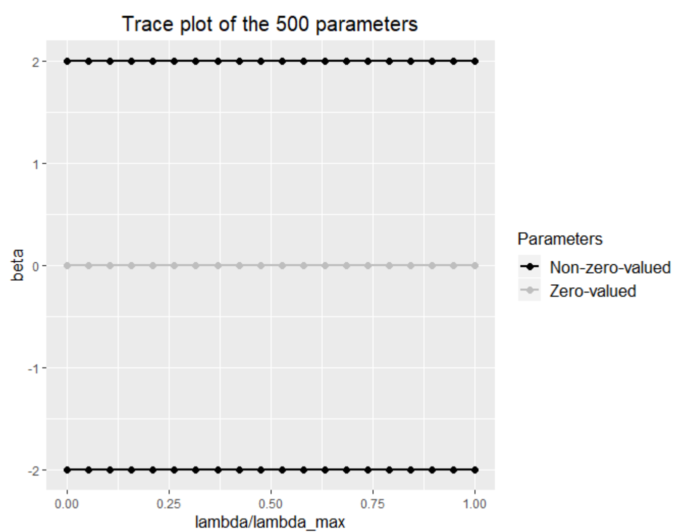

Figure 2: Simulation results.

Quizzes Week 13

pp. 12: We consider the following two cases:

Case 1: Obviously for $\mathbf{x} \neq \mathbf{0}$ we have

$$\frac{\partial ||\mathbf{x}||_\infty}{\partial \mathbf{x}} = \frac{\partial \max_j |x_j|}{\partial \mathbf{x}} = \text{ sign}(x_{j^*})\mathbf{e}_{j^*},$$

where $j^* = \arg\max_j |x_j|$. The above result implies

$$\mathbf{u}^T\mathbf{x} = \text{ sign}(x_{j^*})\mathbf{e}_{j^*}^T\mathbf{x} = \text{ sign}(x_{j^*})x_{j^*} = |x_{j^*}| = \max_j |x_j| = ||\mathbf{x}||_\infty.$$

Case 2: For $\mathbf{x} = \mathbf{0}$, if $\mathbf{u}$ is a subgradient of $||\mathbf{x}||_\infty$, we must have

$$\begin{aligned} \mathbf{y}^T\mathbf{u} \leq ||\mathbf{y}||_\infty \quad &\Rightarrow \quad \max_{\mathbf{u}} \mathbf{y}^T\mathbf{u} \leq ||\mathbf{y}||_\infty \\ &\Rightarrow \quad ||\mathbf{y}||_\infty ||\mathbf{u}||_1 \leq ||\mathbf{y}||_\infty \\ &\Rightarrow \quad ||\mathbf{u}||_1 \leq 1 \end{aligned}$$

for any $\mathbf{y} \in \mathbb{R}^p$.

From the above results we conclude for $\mathbf{x} \in \mathbb{R}^p$, we have

$$\partial ||\mathbf{x}||_\infty = \{\mathbf{u} \in \mathbb{R}^p : ||\mathbf{u}||_1 \leq 1 \text{ and } \mathbf{u}^T\mathbf{x} = ||\mathbf{x}||_\infty\}.$$

pp. 20: For **a.**, by definition we have

$$\begin{aligned} g^*(\mathbf{u}) &= \max_{\mathbf{x}} \left\{ \mathbf{u}^T\mathbf{x} - ||\mathbf{x}||_1 \right\} \\ &\leq \max_{\mathbf{x}} \left\{ ||\mathbf{u}||_\infty ||\mathbf{x}||_1 - ||\mathbf{x}||_1 \right\} \\ &= \begin{cases} 0 & \text{if } ||\mathbf{u}||_\infty \leq 1 \\ \infty & \text{otherwise} \end{cases}, \end{aligned}$$

which implies

$$g^*(\mathbf{u}) = \iota\{\mathbf{u} \in \{\mathbf{v} \in \mathbb{R}^p : ||\mathbf{v}||_\infty \leq 1\}\}.$$

For **b.**, by definition we have

$$\begin{aligned} g^*(\mathbf{u}) &= \max_{\mathbf{x}} \left\{ \mathbf{u}^T\mathbf{x} - ||\mathbf{x}||_2 \right\} \\ &\le \max_{\mathbf{x}} \left\{ ||\mathbf{u}||_2||\mathbf{x}||_2 - ||\mathbf{x}||_2 \right\} \\ &= \begin{cases} 0 & \text{if } ||\mathbf{u}||_2 \le 1 \\ \infty & \text{otherwise} \end{cases}, \end{aligned}$$

which implies

$$g^* = \iota\{\mathbf{u} \in \{\mathbf{v} \in \mathbb{R}^p : ||\mathbf{v}||_2 \le 1\}\}.$$

c. is wrong according to b..

pp. 49: For **a.**, the solution must satisfy the following equation:

$$(x-a) + \lambda = 0 \Rightarrow x = a - \lambda.$$

For **b.**, the solution must satisfy the following equation:

$$\lambda(x-a) + s = 0 \Rightarrow x = a - \frac{s}{\lambda},$$

where $s \in \partial|x|$. Obvious $S_\lambda(a)$ is not the right answer.

For **c.**, the solution must satisfy the following equation:

$$(x-a) + \lambda s = 0 \quad \Rightarrow \quad x = \begin{cases} a - \lambda & \text{if } a > \lambda \\ 0 & \text{if } -\lambda \le a \le \lambda \\ a + \lambda & \text{if } a < -\lambda \end{cases}.$$

The above representation is exactly the same as $S_\lambda(a)$.