Ordered Weighted ℓ_1 Regularized Regression with Strongly Correlated Covariates: Theoretical Aspects (Supplementary Material)

Mário A. T. Figueiredo
Instituto de Telecomunicações
Instituto Superior Técnico
Universidade de Lisboa, Portugal

Robert D. Nowak
Depart. of Electrical and Computer Engineering
University of Wisconsin, Madison, USA

Proofs of the Lemmas in Section 2

Proof of Lemma 2.1

Recall that x_i and x_j are non-negative and let l and m be their respective rank orders, *i.e.*, $x_i = x_{[l]}$ and $x_j = x_{[m]}$; of course, $l < m$, because $x_i = x_{[l]} > x_{[m]} = x_j$. Now let $l + a$ and $m - b$ be the rank orders of z_i and z_j, respectively, *i.e.*, $x_i - \varepsilon = z_i = z_{[l+a]}$ and $x_j + \varepsilon = z_j = z_{[m-b]}$. Of course, it may happen that a or b (or both) are zero, if ε is small enough not to change the rank orders of one (or both) of the affected components of $\boldsymbol{x}$. Furthermore, the condition $\varepsilon < (x_i - x_j)/2$ implies that $x_i - \varepsilon > x_j + \varepsilon$, thus $l + a < m - b$. A key observation is that $\boldsymbol{x}_\downarrow$ and $\boldsymbol{z}_\downarrow$ only differ in positions l to $l + a$ and $m - b$ to m, thus we can write

$$\Omega_{\boldsymbol{w}}(\boldsymbol{x}) - \Omega_{\boldsymbol{w}}(\boldsymbol{z}) = \sum_{k=l}^{l+a} w_k \left(x_{[k]} - z_{[k]}\right) + \sum_{k=m-b}^{m} w_k \left(x_{[k]} - z_{[k]}\right). \tag{i}$$

In the range from l to $l + a$, the relationship between $\boldsymbol{z}_\downarrow$ and $\boldsymbol{x}_\downarrow$ is

$$z_{[l]} = x_{[l+1]},\; z_{[l+1]} = x_{[l+2]},\; \dots\;,\; z_{[l+a-1]} = x_{[l+a]},\; z_{[l+a]} = x_{[l]} - \varepsilon,$$

whereas in the range from $m - b$ to m, we have

$$z_{[m-b]} = x_{[m]} + \varepsilon,\; z_{[m-b+1]} = x_{[m-b]},\; \dots\;,\; z_{[m]} = x_{[m-1]}.$$

Plugging these equalities into (i) yields

$$\begin{aligned}
\Omega_{\boldsymbol{w}}(\boldsymbol{x}) - \Omega_{\boldsymbol{w}}(\boldsymbol{z}) &= \sum_{k=l}^{l+a-1} w_k \underbrace{\left(x_{[k]} - x_{[k+1]}\right)}_{\geq 0} + \sum_{k=m-b+1}^{m} w_k \underbrace{\left(x_{[k]} - x_{[k-1]}\right)}_{\leq 0} \\
&\quad + w_{l+a}\left(x_{[l+a]} - x_{[l]} + \varepsilon\right) + w_{m-b}\left(x_{[m-b]} - x_{[m]} - \varepsilon\right) \\
&\stackrel{(a)}{\geq} w_{l+a} \sum_{k=l}^{l+a-1} \left(x_{[k]} - x_{[k+1]}\right) + w_{m-b} \sum_{k=m-b+1}^{m} \left(x_{[k]} - x_{[k-1]}\right) \\
&\quad + w_{l+a}\left(x_{[l+a]} - x_{[l]} + \varepsilon\right) + w_{m-b}\left(x_{[m-b]} - x_{[m]} - \varepsilon\right) \\
&= w_{l+a} \left(\sum_{k=l}^{l+a-1} \left(x_{[k]} - x_{[k+1]}\right) + \left(x_{[l+a]} - x_{[l]} + \varepsilon\right) \right) \\
&\quad + w_{m-b} \left(\sum_{k=m-b+1}^{m} \left(x_{[k]} - x_{[k-1]}\right) + \left(x_{[m-b]} - x_{[m]} - \varepsilon\right) \right) \\
&\stackrel{(c)}{=} \varepsilon \left(w_{l+a} - w_{m-b}\right) \stackrel{(c)}{\geq} \varepsilon\, \Delta_{\boldsymbol{w}},
\end{aligned}$$

where inequality (a) results from $x_{[k]} - x_{[k+1]} \geq 0$, $x_{[k]} - x_{[k-1]} \leq 0$, and the components of $\boldsymbol{w}$ forming a non-increasing sequence, thus $w_{l+a} \leq w_k$, for $k = l, ..., l+a-1$, and $w_{m-b} \geq w_k$, for $k = m-b+1, ..., m$; equality (c) is a consequence of the cancellation of the remains of the telescoping sums with the two other terms; inequality (c) results from the fact that (see above) $l + a < m - b$ and the definition of $\Delta_{\boldsymbol{w}}$ given in Section 1 of the paper.

Proof of Lemma 2.2

Let l and m be the rank orders of x_i and x_j, respectively, *i.e.*, $x_i = x_{[l]}$ and $x_j = x_{[m]}$; without loss of generality, assume that $m > l$. Furthermore, let $l + a$ and $m + b$ be the rank orders of s_i and s_j in $\boldsymbol{s}$ (*i.e.*, $s_i = s_{[l+a]}$ and $s_j = s_{[m+b]}$); naturally, $a, b \geq 0$. Then,

$$\begin{aligned} \Omega_{\boldsymbol{w}}(\boldsymbol{x}) - \Omega_{\boldsymbol{w}}(\boldsymbol{s}) &\geq w_l\, x_i + w_m\, x_j - w_{l+a}(x_i - \varepsilon) - w_{m+b}(x_j - \varepsilon) \\ &\geq \underbrace{(w_l - w_{l+a})}_{\geq 0} x_i + \underbrace{(w_m - w_{m+b})}_{\geq 0} x_j + \underbrace{(w_{l+a} + w_{m+b})}_{\geq \Delta_{\boldsymbol{w}}} \varepsilon \;\geq\; \Delta_{\boldsymbol{w}}\, \varepsilon, \end{aligned}$$

where the inequality $w_{l+a} + w_{m+b} \geq \Delta_{\boldsymbol{w}}$ results from the definition of $\Delta_{\boldsymbol{w}}$, which implies that $w_1, ..., w_{p-1} \geq \Delta_{\boldsymbol{w}}$ (only w_p can be less than $\Delta_{\boldsymbol{w}}$, maybe even zero).

Proof of Lemma 2.4

The proof is a direct consequence of the triangle inequality. Letting $\boldsymbol{g} = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}$, we have

$$\begin{aligned} L_1(\boldsymbol{v}) - L_1(\boldsymbol{x}) &= \|\boldsymbol{g} - \varepsilon \boldsymbol{a}_i + \varepsilon \boldsymbol{a}_j\|_1 - \|\boldsymbol{g}\|_1 \\ &\leq \|\boldsymbol{g}\|_1 + |\varepsilon| \|\boldsymbol{a}_i - \boldsymbol{a}_j\|_1 - \|\boldsymbol{g}\|_1 \\ &= |\varepsilon| \|\boldsymbol{a}_i - \boldsymbol{a}_j\|_1. \end{aligned}$$

Proof of Theorem 3.2

The bound stated in Theorem 3.2 follows from the deviation inequality

$$\mathbb{E} \sup_{\boldsymbol{u} \in \mathcal{T}} \left| \frac{1}{n} \sum_{i=1}^{n} |\langle \boldsymbol{a}_i, \boldsymbol{u} \rangle| - \sqrt{\frac{2}{\pi}} \left(\boldsymbol{u}^T \boldsymbol{C}^T \boldsymbol{C} \boldsymbol{u} \right)^{1/2} \right| \leq \frac{4}{\sqrt{n}}\, \mathbb{E} \sup_{\boldsymbol{u} \in \mathcal{T}} |\langle \boldsymbol{C}^T \boldsymbol{g}, \boldsymbol{u} \rangle| \,, \tag{ii}$$

where $\boldsymbol{a}_i$ denotes the ith row of $\boldsymbol{A}$. To see this, note that the inequality holds if we replace the set $\mathcal{T}$ on the left hand side by the smaller set $\mathcal{T}_\varepsilon$. For $\boldsymbol{u} \in \mathcal{T}_\varepsilon$ we have by assumption that

$$\frac{1}{n} \sum_{i=1}^{n} |\langle \boldsymbol{a}_i, \boldsymbol{u} \rangle| = \frac{1}{n} \|\boldsymbol{A}\boldsymbol{u}\|_1 \leq \varepsilon,$$

and the bound in the theorem follows by the triangle inequality.

To prove (ii), the first thing to note is that

$$\mathbb{E}|\langle \boldsymbol{a}_i, \boldsymbol{u} \rangle| = \mathbb{E}|\langle \boldsymbol{C}^T \boldsymbol{b}_i, \boldsymbol{u} \rangle| \;=\; \mathbb{E}|\langle \boldsymbol{b}_i, \boldsymbol{C}\boldsymbol{u} \rangle| \,,$$

where $\boldsymbol{b}_i$ is the ith row of $\boldsymbol{B}$. Because the Gaussian distribution of $\boldsymbol{b}_i$ is rotationally invariant, it follows that

$$\mathbb{E}|\langle \boldsymbol{b}_i, \boldsymbol{C}\boldsymbol{u} \rangle| \;=\; \sqrt{\frac{2}{\pi}} \left(\boldsymbol{u}^T \boldsymbol{C}^T \boldsymbol{C} \boldsymbol{u} \right)^{1/2} \,.$$

Using the symmetrization and contraction inequalities from a proposition by Vershynin (2014, Proposition 5.2), we have the bound

$$\begin{aligned} \mathbb{E} \sup_{\boldsymbol{u} \in T} \left| \frac{1}{n} \sum_{i=1}^{n} |\langle \boldsymbol{a}_i, \boldsymbol{u} \rangle| - \sqrt{\frac{2}{\pi}} \left(\boldsymbol{u}^T \boldsymbol{C}^T \boldsymbol{C} \boldsymbol{u} \right)^{1/2} \right| &\leq 4\, \mathbb{E} \sup_{\boldsymbol{u} \in T} \left| \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i \langle \boldsymbol{b}_i, \boldsymbol{C}\boldsymbol{u} \rangle \right| \\ &= 4\, \mathbb{E} \sup_{\boldsymbol{u} \in T} \left| \left\langle \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i \boldsymbol{b}_i, \boldsymbol{C}\boldsymbol{u} \right\rangle \right| \,, \end{aligned}$$

where each ε_i independently takes values -1 and $+1$ with probabilities $1/2$. Note that vector

$$\boldsymbol{g} := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i \boldsymbol{b}_i \sim \mathcal{N}(0, \mathbf{I}_q),$$

thus,

$$4\,\mathbb{E} \sup_{\boldsymbol{u} \in T} \left| \left\langle \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i \boldsymbol{b}_i, \boldsymbol{C}\boldsymbol{u} \right\rangle \right| = \frac{4}{\sqrt{n}} \mathbb{E} \sup_{\boldsymbol{u} \in T} |\langle \boldsymbol{g}, \boldsymbol{C}\boldsymbol{u} \rangle| = \frac{4}{\sqrt{n}} \mathbb{E} \sup_{\boldsymbol{u} \in T} |\langle \boldsymbol{C}^T \boldsymbol{g}, \boldsymbol{u} \rangle|,$$

which completes the proof.

References

R. Vershynin. Estimation in high dimensions: A geometric perspective. Technical report, http://arxiv.org/abs/1405.5103, 2014.