STATISTICS 593C: Spring, 2007

Model Selection and Regularization

Jon A. Wellner

Lecture 14 (May 10): In this lecture we will continue discussing the paper by Candès and Tao (2007) on *the Dantzig selector*. In Lecture 13, we (almost) completed the proof of Theorem 2 (Theorem 1.1 of C & T) up to a proof of Lemma 1 (Lemma 3.1 of C & T). The next goal is to prove Lemma 3.1.

Proof of Lemma 1: Consider $\mathbf{X}_{T_{01}}$ as a map from $\mathbb{R}^{|T_{01}|} = \mathbb{R}^{2S}$ to $\mathbb{R}^n$ defined in the natural way by

$$\mathbf{X}_{T_{01}}\underline{c} = \sum_{j \in T_{01}} c_j \mathbf{x}^j .$$

Let $V \equiv [\mathbf{x}^j : \ j \in T_{01}] \equiv$ linear span of $\mathbf{x}^j, j \in T_{01} \subset \mathbb{R}^n$. Thus $V = \text{Range}(\mathbf{X}_{T_{01}}) \equiv R(\mathbf{X}_{T_{01}}) = N(\mathbf{X}_{T_{01}}^T)^{\perp}$. Thus $V \oplus V^{\perp} = \mathbb{R}^n$. Since $\delta < 1$ we know that $\mathbf{X}_{T_{01}}$ is a bijection from $R^{|T_{01}|}$ to V with singular values between $\sqrt{1-\delta}$ and $\sqrt{1+\delta}$:

$$\sqrt{1-\delta}\|c\|_{\ell_2(T_{01})} \le \|\mathbf{X}_{T_{01}} c\|_{\ell_2} \le \sqrt{1+\delta}\|c\|_{\ell_2(T_{01})}. \tag{1}$$

Let $P_V \equiv \Pi(\cdot|V) = \mathbf{X}_{T_{01}}(\mathbf{X}_{T_{01}}^T \mathbf{X}_{T_{01}})^{-1}\mathbf{X}_{T_{01}}^T$ denote the orthogonal projection operator onto V. Thus for each $w \in \mathbb{R}^n$

$$\mathbf{X}_{T_{01}}^T w = \mathbf{X}_{T_{01}}^T P_V w,$$

and it follows that

$$\sqrt{1-\delta}\|P_V w\|_2 \le \|\mathbf{X}_{T_{01}}^T w\|_2 \le \sqrt{1+\delta}\|P_V w\|_2.$$

Taking $w = \mathbf{X}h$ in this set of inequalities yields

$$\sqrt{1-\delta}\|P_V \mathbf{X}h\|_2 \le \|\mathbf{X}_{T_{01}}^T \mathbf{X}h\|_2 \le \sqrt{1+\delta}\|P_V \mathbf{X}h\|_2,$$

and, in particular,

$$\|P_V \mathbf{X}h\|_2 \le \frac{1}{\sqrt{1-\delta}}\|\mathbf{X}_{T_{01}}^T \mathbf{X}h\|_2. \tag{2}$$

Now enumerate T_0^c as $n_1, \ldots, n_{p-|T_0|}$ in decreasing order of magnitude of $h_{T_0^c}$, and thereby divide T_0^c into subsets of size S: set

$$T_j = \{n_l : \ (j-1)S + 1 \le l \le jS\}.$$

Thus T_1 is as in the statement of the lemma and contains the indices of the $S-$largest coefficients of $h_{T_0^c}$. Now decompose $P_V\mathbf{X}h$ as

$$\begin{aligned} P_V\mathbf{X}h &= P_V\mathbf{X}h_{T_{01}} + P_V\mathbf{X}\sum_{j\geq 2} h_{T_j} \\ &= P_V\mathbf{X}h_{T_{01}} + \sum_{j\geq 2} P_V\mathbf{X}h_{T_j} \\ &= \mathbf{X}h_{T_{01}} + \sum_{j\geq 2} P_V\mathbf{X}h_{T_j} \qquad (3) \end{aligned}$$

since $\mathbf{X}h_{T_{01}} \in V$ implies $P_V\mathbf{X}h_{T_{01}} = \mathbf{X}h_{T_{01}}$. Note that

$$P_V\mathbf{X}h_{T_j} = \sum_{l\in T_{01}} c_l \mathbf{x}^l \qquad \text{for some} \;\; c_l \equiv c_l(j),$$

and

$$\begin{aligned} \|P_V\mathbf{X}h_{T_j}\|_2^2 &= \langle P_V\mathbf{X}h_{T_j}, P_V\mathbf{X}h_{T_j}\rangle \\ &= \langle P_V^T P_V\mathbf{X}h_{T_j}, \mathbf{X}h_{T_j}\rangle \\ &= \langle P_V\mathbf{X}h_{T_j}, \mathbf{X}h_{T_j}\rangle \qquad \text{since} \;\; P_V^T = P_V, \;\; P_V^2 = P_V. \qquad (4) \end{aligned}$$

By restricted orthogonality followed by restricted isometry this yields

$$\langle P_V\mathbf{X}h_{T_j}, \mathbf{X}h_{T_j}\rangle \;\leq\; \theta\left(\sum |c_j|^2\right)^{1/2} \|h_{T_j}\|_2 \leq \frac{\theta}{\sqrt{1-\delta}}\|P_V\mathbf{X}h_{T_j}\|_2\|h_{T_j}\|_2. \qquad (5)$$

Combining (4) and (5) gives

$$\|P_V\mathbf{X}h_{T_j}\|_2^2 \leq \frac{\theta}{\sqrt{1-\delta}}\|P_V\mathbf{X}h_{T_j}\|_2\|h_{T_j}\|_2,$$

and, upon dividing through by $\|P_V\mathbf{X}h_{T_j}\|_2$,

$$\|P_V\mathbf{X}h_{T_j}\|_2 \leq \frac{\theta}{\sqrt{1-\delta}}\|h_{T_j}\|_2. \qquad (6)$$

Now we develop an upper bound for $\sum_{j\geq 2}\|h_{T_j}\|_2$. By construction, the magnitude of each coefficient in T_{j+1} is less than the average of the magnitudes in T_j:

$$|h_{T_{j+}}(t)| \leq \frac{\|h_{T_j}\|_1}{S}.$$

Thus

$$\|h_{T_{j+}}\|_2^2 \leq \frac{\|h_{T_j}\|_1^2}{S},$$

and hence

$$\sum_{j\geq 2}\|h_{T_j}\|_2 \leq \sum_{j\geq 1}\frac{\|h_{T_j}\|_1}{\sqrt{S}} = \frac{1}{\sqrt{S}}\|h\|_{\ell_1(T_0^c)}. \qquad (7)$$

Now we just need to put the pieces together: rearranging the equality in (3) yields

$$\mathbf{X}h_{T_{01}} = P_V \mathbf{X}h - \sum_{j \geq 2} P_V \mathbf{X}h_{T_j}. \tag{8}$$

Hence

$$\begin{aligned}
\sqrt{1-\delta}\|h_{T_0}\|_2 &\leq \|\mathbf{X}h_{T_{01}}\|_2 \quad \text{by (1)} \\
&\leq \|P_V\mathbf{X}h\|_2 + \|\sum_{j\geq 2} P_V \mathbf{X}h_{T_j}\|_2 \quad \text{by (8)} \\
&\leq \frac{1}{\sqrt{1-\delta}}\|\mathbf{X}_{T_{01}}^T\mathbf{X}h\|_{\ell_2} + \frac{\theta}{\sqrt{1-\delta}}\frac{1}{\sqrt{S}}\|h\|_{\ell_1(T_0^c)} \\
&\qquad \text{by (2) and (7).}
\end{aligned}$$

Dividing through by $\sqrt{1-\delta}$ yields the first inequality of the lemma.

To prove the second inequality, note that the $k-$th largest value of $h_{T_0^c}$ satisfies

$$|h_{T_0^c}|_{(k)} \leq \frac{\|h_{T_0^c}\|_1}{k},$$

and therefore

$$\|h_{T_{01}^c}\|_2^2 \leq \|h_{T_0^c}\|_1^2 \sum_{k\geq S+1} \frac{1}{k^2} \leq \frac{1}{S}\|h_{T_0^c}\|_1^2.$$

This inequality easily yields the second inequality of the lemma. □

Proof of Theorem 4. (C & T, Theorem 1.3): Let T_0 be the indices corresponding to the $S-$largest entries of β, and write $\hat{\beta} = \beta + h$. Let

$$\sup_{j\leq p}|Z_j| = \sup_{j\leq p}|\langle \epsilon, \mathbf{x}^j\rangle|.$$

On the set $\{Z^* \leq \lambda_p\}$ (which C & T describe as "ϵ satisfies the restricted orthogonality condition"), β is feasible (with high probability as described before), and hence

$$\|\beta_{T_0}\|_1 - \|h_{T_0}\|_1 + \|h_{T_0^c}\|_1 - \|\beta_{T_0^c}\|_1 \leq \|\beta + h\|_1 = \|\widehat{\beta}\|_1 \leq \|\beta\|_1;$$

and hence it follows that

$$\|h_{T_0^c}\|_1 \leq \|h_{T_0}\|_1 + 2\|\beta_{T_0^c}\|_1. \tag{9}$$

This is a rewrite of "Fact 1" ; the only difference is that now we have a second term on the right side. We conclude from Lemma 1 and Fact 2 that

$$\begin{aligned}
\|h\|_{\ell_2(T_{01})} &\leq \frac{1}{1-\delta}\sqrt{2S}2\lambda_p + \frac{\theta}{(1-\delta)\sqrt{S}}\left(\|h_{T_0}\|_1 + 2\|\beta_{T_0^c}\|_1\right) \\
&\leq \frac{1}{1-\delta}\sqrt{2S}2\lambda_p + \frac{\theta}{(1-\delta)\sqrt{S}}\|h_{T_{01}}\|_1 + \frac{2\theta}{(1-\delta)\sqrt{S}}\|\beta_{T_0^c}\|_1.
\end{aligned}$$

Moving the middle term to the left side and rewriting yields

$$\begin{aligned} \|h\|_{\ell_2(T_{01})} &\leq \frac{2\sqrt{2}\lambda_p\sqrt{S}}{1-\delta-\theta} + \frac{2\theta}{(1-\delta-\theta)\sqrt{S}}\|\beta_{T_0^c}\|_1 \\ &\leq \frac{2\sqrt{2}}{1-\delta-\theta}\left(\lambda_p\sqrt{S} + S^{-1/2}\|\beta_{T_0^c}\|_1\right). \qquad (10) \end{aligned}$$

Now from the second part of Lemma 1,

$$\begin{aligned} \|h\|_2^2 &\leq \|h\|_{\ell_2(T_{01})}^2 + S^{-1}\|h\|_{\ell_1(T_0^c)}^2 \\ &\leq \|h\|_{\ell_2(T_{01})}^2 + 2S^{-1}\left(\|h_{T_0}\|_1^2 + 4\|\beta_{T_0^c}\|_1^2\right) \\ &\qquad \text{by } (9) \text{ and } (a+b)^2 \leq 2(a^2+b^2) \\ &\leq 3\|h\|_{\ell_2(T_{01})}^2 + 8S^{-1}\|\beta_{T_0^c}\|_1^2 \\ &\leq \frac{24}{(1-\delta-\theta)^2}\left(\lambda_p^2 S + S^{-1}\|\beta_{T_0^c}\|_1^2\right) + 8S^{-1}\|\beta_{T_0^c}\|_1^2 \qquad \text{by } (10) \\ &= \frac{C}{(1-\delta-\theta)^2}\left(\lambda_p^2 S + S^{-1}\|\beta_{T_0^c}\|_1^2\right) \end{aligned}$$

where

$$\begin{aligned} S^{-1/2}\|\beta_{T_0^c}\|_1 &\leq S^{-1/2}\sum_{j\geq S+1} R j^{-1/s} \leq S^{-1/2} R\int_S^\infty x^{-1/s}dx \\ &= \frac{2R}{2r-1}S^{-r} \qquad \text{with } r = (1/s) - (1/2). \end{aligned}$$

Thus we conclude that

$$\|h\|_2^2 \leq \frac{C}{(1-\delta-\theta)^2}\left(\lambda_p^2 S + R^2 \cdot S^{-2r}\right)$$

for all $S \leq S^*$, and this yields the inequality claimed in Theorem 4 (C & T Theorem 1.3). □