Robust Multi-task Regression – Supplementary Material

October 6, 2011

We provide in this supplementary material proofs that are omitted in the submission.

A Proof of Lemma 5

Proof. By Condition 2 we have

$$\sqrt{\phi_{\min}(S+S')} \geq 4\sqrt{(\phi_{\max}(S'))(S/S')},$$

which leads to

$$\begin{aligned}
\kappa_1 &= \frac{\sqrt{\phi_{\min}(S+S')} - 3\sqrt{\phi_{\max}(S')S/S'}}{1+3\sqrt{S/S'}} \\
&\geq \frac{\sqrt{\phi_{\min}(S+S')}}{16} \geq \frac{\sqrt{(\phi_{\max}(S'))(S/S')}}{4}.
\end{aligned}$$

This established the first claim. Next, note that

$$\begin{aligned}
\kappa_2/\kappa_1 &= \frac{\frac{\sqrt{\phi_{\min}(S+S')}-3\sqrt{\phi_{\max}(S')S/S'}}{\sqrt{S'}+3\sqrt{S}} + \sqrt{\phi_{\max}(S')/S'}}{\frac{\sqrt{\phi_{\min}(S+S')}-3\sqrt{\phi_{\max}(S')S/S'}}{1+3\sqrt{S/S'}}} \\
&= \frac{1}{\sqrt{S'}} + \frac{(1+3\sqrt{S/S'})\sqrt{\phi_{\max}(S')/S'}}{\sqrt{\phi_{\min}(S+S')} - 3\sqrt{\phi_{\max}(S')S/S'}} \\
&\leq \frac{1}{\sqrt{S'}} + \frac{(1+3\sqrt{S/S'})\sqrt{\phi_{\max}(S')/S'}}{\sqrt{\phi_{\max}(S')S/S'}} \\
&= \frac{4}{\sqrt{S'}} + \frac{1}{\sqrt{S}} \leq \frac{5}{\sqrt{S}},
\end{aligned}$$

where the last inequality holds from $S' \geq S$. □

B Proof of Lemma 6

Proof. To prove the lemma, we need the following two results. Lemma B.1 is adapted from Theorem2.13 of [1], and Lemma B.2 is from [2].

Lemma B.1. *Let* Γ *be an* $n \times q$ *matrix, whose entries are all i.i.d.* $\mathcal{N}(0,1)$ *Gaussian variables; then*

$$\Pr\big(\|\Gamma\|_{op} > \sqrt{n} + \sqrt{q} + \epsilon\big) \leq \exp(-\epsilon^2/2).$$

Lemma B.2. *Let* U *be a* χ^2 *statistic with* D *degrees of freedom. For any positive* x*,*

$$Pr(U - D \geq 2\sqrt{Dx} + 2x) \leq \exp(-x).$$

We now proceed to prove Lemma 6. Note that $\|X^\top W\|_{\infty,2} \leq \max_{i=1,\cdots,p} \|X_i\|_2 \|W\|_{op} \leq \|W\|_{op}$. Note that Lemma B.1 implies that for W with i.i.d. $\mathcal{N}(0, \sigma^2/n)$ entries, we have with probability

$$Pr\big(\|W\|_{op} \geq 7\sigma(1 + \sqrt{q/n})\big) \leq \exp(-6(n+q)/2) \leq 1/(4n^3), \tag{B.1}$$

where the last inequality holds when $n \geq 2$. On the other hand, fix a $X_i^\top$, $X_i^\top W_j \sim \mathcal{N}(0, \|X_i\|_2^2 \sigma^2/n)$, and hence

$$\frac{\sqrt{n}}{\sigma\|X_i\|_2} \|X_i' W\|_2 \sim \chi_{q-1}.$$

By Lemma B.2, we have

$$\begin{aligned}
& Pr\left(\|X_i' W\|_2 \geq \frac{\sigma(\sqrt{q} + \sqrt{8\log(pn)})}{\sqrt{n}} \right) \\
\overset{(a)}{\leq} \quad & Pr\left(\|X_i' W\|_2 \geq \frac{\sigma\|X_i\|_2(\sqrt{q} + \sqrt{2\log(p^4 n^4)})}{\sqrt{n}} \right) \\
= \quad & Pr\left(\frac{n\|X_i' W\|_2^2}{\sigma^2\|X_i\|_2^2} \geq q + 2\sqrt{2q\log p^4 n^4} + 2\log(p^4 n^4) \right) \\
\leq \quad & Pr\left(\frac{n\|X_i' W\|_2^2}{\sigma^2\|X_i\|_2^2} - q \geq 2\sqrt{q\log(p^4 n^4)} + 2\log(p^4 n^4) \right) \\
\overset{(b)}{\leq} \quad & 1/(n^4 p^4) \leq 1/(4n^3 p),
\end{aligned}$$

which by union bound leads to

$$Pr(\|X^\top W\|_{\infty,2} \geq \frac{\sigma(\sqrt{q} + \sqrt{2\log(pn)})}{\sqrt{n}}) \leq 1/(4n^3). \tag{B.2}$$

Here, (a) holds since $\|X_i\|_2 \leq 1$ and (b) holds by applying to Lemma B.2 to χ^2 random variable $\frac{n\|X_i' W\|_2^2}{\sigma^2\|X_i\|_2^2}$. Lemma 6 follows by a union bound on Equation (B.1) and (B.2). ∎

C Proof of Example 1

In this appendix we show the validity of the claim in Example 1. Recall the following theorem from [1] (Thm. 2.13).

Theorem C.1. *Given* $m, n \in \mathbb{N}$ *with* $m \le n$, *put* $\Theta = m/n$ *and consider the* $n \times m$ *random matrix* Γ *whose entries are real, independent Gaussian random variable following* $\mathcal{N}(0, 1/n)$. *Let the singular value be* $\sigma_1(\Gamma) \ge, \cdots, \ge \sigma_m(\Gamma)$. *Then for any* $t > 0$,

$$Pr(\sigma_1(\Gamma) \ge 1+\sqrt{\Theta}+t) \le \exp(-nt^2/2); \quad Pr(\sigma_m(\Gamma) \le 1-\sqrt{\Theta}-t) \le \exp(-nt^2/2).$$

Then, fix any $I \subset [1:p]$ with $|I| = S + S'$ and denote $M = X_I$ we have that M is a $(S+S') \times n$ random matrix with IID entries following $\mathcal{N}(0, 1/n)$. Thus, for any $t > 0$,

$$\begin{aligned} & Pr(\phi_{\min}(S+S') < 1 - \sqrt{(S+S')/n} - t) \\ \le & \begin{pmatrix} p \\ S+S' \end{pmatrix} \times Pr(\sigma_{S+S'}(M) < 1 - \sqrt{(S+S')/n} - t) \\ \le & p^{S+S'} \exp(-nt^2/2) = \exp\left(\frac{2(S+S')\log p - nt^2}{2}\right), \end{aligned}$$

where the first inequality follows from a union bound on all subsets with cardinality $S + S'$. Similarly

$$Pr(\phi_{\max}(S') > 1 + \sqrt{S'/n} + t) \le \exp\left(\frac{2S'\log p - nt^2}{2}\right)$$

Set $t = \sqrt{4(S+S')\log p/n}$ we have that with probability at least $1 - 1/p^2$,

$$\phi_{\min}(S+S') \ge 1 - \sqrt{\frac{S+S'}{n}} - \sqrt{\frac{4(S+S')\log p}{n}} \ge 1 - \sqrt{\frac{9(S+S')\log p}{n}};$$
$$\phi_{\max}(S') \le 1 + \sqrt{\frac{S'}{n}} + \sqrt{\frac{4(S+S')\log p}{n}} \le 1 + \sqrt{\frac{9(S+S')\log p}{n}}.$$

When $S\log p \le n/1764$ and $S' = 48S$, we have that $(S+S')\log p/n \le 1/36$ which implies that $\phi_{\min}(S+S') \ge 1/2$ and $\phi_{\max}(S') \le 3/2$. Hence

$$S'\phi_{\min}(S+S') \ge 16S\phi_{\max}(S'),$$

as claimed.

References

[1] K. Davidson and S. Szarek. Local operator theory, random matrices and banach spaces. In W. Johnson and J. Lindenstrauss, editors, *Handbook on the Geometry of Banach Spaces*, pages 317–366. Elsvier, 2001.

[2] B. Laurent and P. Massart. Adaptive estimation of a quadratic functional of by model selection. *Annals of Statistics*, 28:1302–1338, 2000.