Appendix to A note on cross-validation for Lasso under measurement errors

Abhirup Datta
Department of Biostatistics, Johns Hopkins University
and
Hui Zou
Department of Statistics, University of Minnesota

September 9, 2019

A Appendix: Proofs

For any matrix M, $\|M\|_{\max}$ denotes the entrywise maximum norm, $\|M\|_2$ denote the matrix ℓ_2 norm and $\rho_{\min}(M)$ denote its minimum eigenvalue. Finally let X_j and Z_j denote the j^{th} columns of X and Z respectively.

Proof of Lemma 1. Recall that $X^\top X/n \to \Sigma$. Let $S_{1n} = \{\omega : 1/(2\rho_{\max}(\Sigma)) \leq \rho_{\min}(Z^\top Z/n - \Gamma)^{-1} \leq 2/\rho_{\min}(\Sigma)\}$. On S_{1n}, $Z^\top Z/n - \Gamma$ is invertible and the CoCoLasso estimate in (2) can be disguised as a Lasso equation as

$$\hat{\beta}_C(\lambda) = \arg\min_{\beta} \quad (2n)^{-1}\|\tilde{y} - \tilde{Z}\beta\|_2^2 + \lambda\|\beta\|_1. \tag{1}$$

where $\tilde{Z}/n$ is the Cholesky factor of $Z^\top Z/n - \Gamma$ and $\tilde{y}$ is the solution of $\tilde{Z}^\top \tilde{y} = Z^\top y$. The solution path is hence piecewise linear in λ (Tibshirani 2013, Lemma 8). If $R = R(\lambda)$ denote the active set (set of indices corresponding to the non-zero entries of $\hat{\beta}(\lambda)$), then the slope is given by $\|(\tilde{Z}^\top \tilde{Z}/n)^{-1}_{R,R}\|_2 = \|((\frac{1}{n}Z^\top Z - \Gamma)_{R,R})^{-1}\|_2$. So, the maximal slope is bounded by

$$\sup_{R\subset\{1,2,\ldots,p\}} \| \left((\frac{1}{n}Z^\top Z - \Gamma)_{R,R}\right)^{-1} \|_2 \leq \|(\frac{1}{n}Z^\top Z - \Gamma)^{-1}\|_2 \leq 2/\rho_{\min}(\Sigma) \tag{2}$$

From Lemma A2, $p(S^c_{in}) = o(n^{-1})$ which proves the Lipschitz continuity for the CocoLasso estimate $\hat{\beta}_{C,0}(\lambda)$. Similar, result will hold if Z is replaced by Z_{-i} (i.e., the i^{th} observation is removed). Since $no(n^{-1}) \to 0$, the result holds uniformly for $\hat{\beta}^{(i)}_C(\lambda)$ over $i = 0, 1, \ldots, n$. $\square$

Proof of Lemma 2. If X was observed, $\hat{\beta}_L(\lambda) = 0$ for all $\lambda \geq \|X^\top y/n\|_\infty$. Equivalently, $\hat{\beta}_C(\lambda) = 0$ if $\lambda \geq \|\tilde{Z}^\top \tilde{y}/n\|_\infty = \|Z^\top y/n\|_\infty$. We can expand $Z^\top y = X^\top X\beta^* + X^\top w + A^\top X\beta^* + A^\top w$. Using Lemma A1, $P(\|Z^\top y/n\|_\infty \geq \|\Sigma\beta^*\|_\infty + 1) = o(n^{-1})$ and hence

$$p(\|\beta_{C,0}(\lambda)\|_2 \geq 2(\|\Sigma\beta^*\|_\infty + 1)/\rho_{\min}(\Sigma)) = o(n^{-1}). \tag{3}$$

Again, similar result will hold by replacing Z with Z_{-i} which proves the lemma. $\square$

Before proving Lemma 3, we introduce some additional notation. Let $\beta(\lambda)$ denote the minimizer of $g(\beta)$ where $g(\beta) = \frac{1}{2}(\beta - \beta^*)^\top\Sigma(\beta - \beta^*) + \lambda\|\beta\|_1 + \sigma^2$. First note that as $g(\beta)$ can be viewed as a noiseless version of the Lasso loss function (1). Hence, $\|\beta(\lambda)\|_2$ is Lipshitz and uniformly stochastically bounded in $\lambda \in [0, \Lambda]$. For any β, let $R(\beta, \epsilon)$ denote the closed ℓ_2-ball of radius ϵ and centered around β. Since $g(\beta)$ is strictly convex, there is a $\delta(\epsilon, \lambda) > 0$ such that $\beta \notin R(\beta(\lambda), \epsilon) \Rightarrow g(\beta) > g(\beta(\lambda)) + \delta(\epsilon, \lambda)$. Clearly, $\delta(\epsilon, \lambda)$ is decreasing in ϵ and goes to zero as $\epsilon \to 0$. We first prove the following result:

Lemma 1. *For any fixed λ, and $p(\|\hat{\beta}_C(\lambda) - \beta(\lambda)\|_2 > \epsilon) = o(n^{-1})$.*

Proof. The CoCoLasso estimate $\hat{\beta}_C^{(n)}(\lambda)$ is obtained by minimizing $g_n(\beta) = \frac{1}{2}(y^\top y/n - 2y^\top Z\beta/n + \beta^\top(Z^\top Z/n - \Gamma)_+\beta) + \lambda\|\beta\|_1$.

$$\begin{aligned}|g_n(\beta) - g(\beta)| \leq\ & |\beta^\top((Z^\top Z/n - \Gamma)_+ - \Sigma)\beta|/2 + |\beta^{*\top}(X^\top X/n - \Sigma)\beta^*|/2 + \\ & |\beta^\top(Z^\top X/n - \Sigma)\beta^*| + |\beta^\top Z^\top w/n| + |w^\top X^\top\beta^*/n| + |w^\top w/n|\end{aligned}$$

Using Lemmas A1 and A2, we have $p(|g_n(\beta) - g(\beta)| > \delta(\epsilon, \lambda)/4) = o(n^{-1})$.

Let K be the compact hypercube in $\mathbb{R}^p$, centered at zero and having edges of length L such that $L = 5(\|\Sigma\beta^*\|_\infty + 1)/\rho_{\min}(\Sigma)$. Using the bound of Equation (3) one can see that this choice of L ensures that K contains $\cup_{\lambda\in[0,\Lambda]}R(\beta(\lambda), \epsilon)$ for small enough ϵ and contains $\{\hat{\beta}_{C,0}(\lambda) : \lambda \in [0, \Lambda]\}$ with probability $1 - o(n^{-1})$.

As $\lambda \in [0, \Lambda]$, $g(\beta)$ is also Lipschitz in β on K with constant $\kappa = ||\Sigma\|_2 L + \sqrt{p}\Lambda$. We use Lemma A3 (Equation 6) and conclude that for any $\epsilon > 0$,

$$p(S) \leq (8\sqrt{2}L\kappa/\delta(\epsilon, \lambda) + 3)^p o(n^{-1}) \text{ where } S = \{\sup_{\beta\in K}|g_n(\beta) - g(\beta)| > \delta(\epsilon, \lambda)/4\}.$$

On S^c, we have $g_n(\beta(\lambda)) < g(\beta(\lambda)) + \delta(\epsilon, \lambda)/4$. Therefore on S^c, for $\beta \in K \setminus R(\beta(\lambda), \epsilon)$, we have $g_n(\beta) > g(\beta) - \delta(\epsilon, \lambda)/4 > g(\beta(\lambda)) + 3\delta(\epsilon, \lambda)/4 > g_n(\beta(\lambda))$. Hence,

$$\begin{aligned}p(\|\hat{\beta}_C(\lambda) - \beta(\lambda)\|_2 > \epsilon) &\leq p(S) + p(\hat{\beta}_C(\lambda) \in K^c) \\ &\leq (4\sqrt{2}L\kappa/\delta(\epsilon, \lambda) + 3)^p o(n^{-1}) + o(n^{-1}).\end{aligned}$$

Since $\delta(\epsilon, 4)$ does not depend on n, the right hand side is $o(n^{-1})$. $\square$

Proof of Lemma 3. We emulate the proofs of Theorems 21.9 and 21.10 in Davidson (1994) with a more careful tracking of the probability bounds throughout to ensure that the results hold uniformly for all the leave-one-out estimators $\hat{\beta}_C^{(i)}(\lambda)$.

Let $Q_n(\lambda) = \|\hat{\beta}_C(\lambda) - \beta(\lambda)\|_2$. Since, $|Q_n(\lambda) - Q_n(\lambda')| \leq \|\hat{\beta}_C(\lambda) - \hat{\beta}_C(\lambda')\|_2 + \|\beta(\lambda) - \beta(\lambda')\|_2$, from Lemma 1, $|Q_n(\lambda) - Q_n(\lambda')| \leq C_n|\lambda - \lambda'|$ with $p(S_n) = o(n^{-1})$ for set $S_n = \{|C_n| > 4/\rho_{\min}(\Sigma)\}$. For any $\epsilon > 0$, let $\lambda_0 = 0, \lambda_1, \ldots, \lambda_m = \Lambda$ denote an increasing sequence of points such that $\lambda_i - \lambda_{i-1} = \mu \leq \epsilon\rho_{\min}(\Sigma)/8$. On S_n for any $\lambda \in [0, \Lambda]$, $Q_n(\lambda) \leq \epsilon/2 + \max_{i=1,\ldots,m} Q_n(\lambda_i)$. Then, we have

$$p(\sup_{\lambda\in[0,\Lambda]} |Q_n(\lambda)| > \epsilon) \leq \sum_{i=1}^{m} p(Q_n(\lambda_i) > \epsilon/2) + p(S_n^c) = \sum_{i=1}^{m} o(n^{-1}) + o(n^{-1}) = o(n^{-1}) \,.$$

The analogous result also holds if we replace $\hat{\beta}_C(\lambda)$ with the clean Lasso estimate $\hat{\beta}_L(\lambda)$ in $Q_n(\lambda)$ (a version of the clean Lasso result is provided in Theorem 1 of Knight & Fu (2000)). Hence, using triangular inequality we have,

$$p(\sup_{\lambda\in[0,\Lambda]} \|\hat{\beta}_C(\lambda) - \hat{\beta}_L(\lambda)\|_2 > \epsilon)) = o(n^{-1})\,.$$

Summing over the n probabilities for all the leave-one-out estimates, Lemma 3 is proved. $\square$

Proof of Lemma 4. $R_X(\lambda)$ and $\widetilde{R}_X(\lambda)$ are identical quadratic forms with the only difference being $\hat{\beta}_L(\lambda)$ in $R_X(\lambda)$ gets replaced by $\hat{\beta}_C(\lambda)$ in $\widetilde{R}_X(\lambda)$. Hence, part (a) follows immediately from Lemmas 2 and 3. Similarly, to prove part (b) we once use Lemmas 2 and 3 to show that the difference of the quadratic forms $\hat{\beta}_C(\lambda)^\top\Gamma\hat{\beta}_C(\lambda) - \hat{\beta}_L(\lambda)^\top\Gamma\hat{\beta}_L(\lambda)$ is $o_p(1)$. For part (c), we expand $y_i - z_i^\top\hat{\beta}_C^{(i)}(\lambda) = \sum_{j=1}^{4} t_{ij}$ where $t_{i1} = x_i^\top(\beta^* - \hat{\beta}_L(\lambda))$, $t_{i2} = w_i$, $t_{i3} = -a_i^\top\hat{\beta}_L(\lambda)$ and $t_{i4} = z_i^\top(\hat{\beta}_L(\lambda) - \hat{\beta}_C^{(i)}(\lambda))$. Let

$$L_{n,X}(\lambda) = n^{-1}\|X(\beta^* - \hat{\beta}_L(\lambda))\|_2^2\,. \tag{4}$$

Hence $|\widehat{L}_{n,Z}(\lambda) - L_{n,X}(\lambda) - \sigma^2 - \hat{\beta}_L(\lambda)^\top\Gamma\hat{\beta}_L(\lambda)\|_2^2|$ is less than

$$\begin{aligned}&|\left(\frac{1}{n}w^\top w - \sigma^2\right)| + |\hat{\beta}_L(\lambda)^\top\left(\frac{1}{n}A^\top A - \Gamma\right)\hat{\beta}_L(\lambda)| + |\frac{1}{n}w^\top X(\beta^* - \hat{\beta}_L(\lambda))| + |\frac{1}{n}w^\top A\hat{\beta}_L(\lambda)| + \\ &\quad |\frac{1}{n}\hat{\beta}_L(\lambda)^\top A^\top X(\beta^* - \hat{\beta}_L(\lambda))| + \max_{i=1,\ldots,n} t_{i4}^2 + 2\sum_{j=1}^{3}\left(\max_{i=1,\ldots,n}|t_{ij}|\right)\left(\max_{i=1,\ldots,n}|t_{i4}|\right)\end{aligned}$$

Since from Lemma 3, t_{i4} is $o_p(1)$ and t_{ij}'s, for $j \leq 3$, are $O_p(1)$ uniformly over i and λ, the last two terms in the equation above are $o_p(1)$. The other terms are $o_p(1)$ using Lemma A1. $\square$

Proof of Theorem 1. Using triangular inequality, we have

$$\begin{aligned}|\widehat{L}_{n,Z}(\lambda) - \widetilde{R}_Z(\lambda)| \leq &|\widehat{L}_{n,Z}(\lambda) - L_{n,X}(\lambda) - \hat{\beta}_L(\lambda)^\top\Gamma\hat{\beta}_L(\lambda) - \sigma^2| + \\ &|L_{n,X}(\lambda) + \sigma^2 - R_X(\lambda)| + |R_X(\lambda) - \widetilde{R}_X(\lambda)| + \\ &|\hat{\beta}_L(\lambda)^\top\Gamma\hat{\beta}_L(\lambda) - \hat{\beta}_C(\lambda)^\top\Gamma\hat{\beta}_C(\lambda)|\end{aligned}$$

From Lemma 4, the first and third terms in the right hand side are $o_p(1)$ uniformly in λ. Similarly, using Lemmas 2 and that $\frac{1}{2}nX'X \to \Sigma$, the second terms is $o_p(1)$ uniformly in λ. Finally, Lemmas 2 and 3 ensure that the fourth term is uniformly $o_p(1)$. Hence, part (a) is proved.
For part (b), note that

$$\begin{aligned}|R_X(\lambda) - \widehat{L}_{n,Z}(\lambda)| >& |R_X(\lambda) - \sigma^2 - n^{-1}\|X(\beta^* - \hat{\beta}_L(\lambda))\|_2^2 - \hat{\beta}_L(\lambda)^\top \Gamma \hat{\beta}_L(\lambda)| - \\ & |\widehat{L}_{n,Z}(\lambda) - \sigma^2 - n^{-1}\|X(\beta^* - \hat{\beta}_L(\lambda))\|_2^2 - \hat{\beta}_L(\lambda)^\top \Gamma \hat{\beta}_L(\lambda)| \\ >& |\hat{\beta}_L(\lambda)^\top \Gamma \hat{\beta}_L(\lambda)| - |R_X(\lambda) - \sigma^2 - L_{n,X}(\lambda)| - \\ & |\widehat{L}_{n,Z}(\lambda) - \sigma^2 - n^{-1}\|X(\beta^* - \hat{\beta}_L(\lambda))\|_2^2 - \hat{\beta}_L(\lambda)^\top \Gamma \hat{\beta}_L(\lambda)|\end{aligned}$$

The second term in the right hand side has already been shown to be uniformly $o_p(1)$ in part (a). Using Lemma 4, the third term in the right hand side is also $o_p(1)$ uniformly in λ. So, we only work with the first term $t(\lambda) = \hat{\beta}_L(\lambda)^\top \Gamma \hat{\beta}_L(\lambda)$. Note that

$$t(0) = y^\top X(X^\top X)^{-1}\Gamma(X^\top X)^{-1}X^\top y = \beta^{*\top}\Gamma\beta^* + v^\top \Gamma v \ ,$$

where $v = (X^\top X)^{-1}X^\top w = o_p(1)$. Hence, $t(0) = \beta^{*\top}\Gamma\beta^* + o_p(1)$ and consequently,

$$\sup_{\lambda \in [0,\Lambda]} |R_X(\lambda) - \widehat{L}_{n,Z}(\lambda)| > \beta^{*\top}\Gamma\beta^* + o_p(1) \ .$$

So part (b) is proved for any $\epsilon_0 < \beta^{*\top}\Gamma\beta^*$. □

Proof of Theorem 2. Since, $\widetilde{R}_X(\lambda)$ and $R_X(\lambda)$ are asymptotically equivalent uniformly in λ (Lemma 4 part (a)), it is enough to prove only one of the statements.

$$\begin{aligned}|\widetilde{R}_X(\lambda) - \widetilde{L}_{n,Z}(\lambda)| \leq& |\widetilde{R}_Z(\lambda) - \widehat{L}_{n,Z}(\lambda)| + \frac{1}{n}\sum_{i=1}^{n} |\hat{\beta}_C^{(i)}(\lambda)^\top \Gamma \hat{\beta}_C^{(i)}(\lambda) - \hat{\beta}_L(\lambda)^\top \Gamma \hat{\beta}_L(\lambda)| \\ \leq& |\widetilde{R}_Z(\lambda) - \widehat{L}_{n,Z}(\lambda)| + \|\Gamma\|_2 \max_{i=0,\ldots,n} \|\hat{\beta}_C^{(i)}(\lambda) - \hat{\beta}_L(\lambda)\|_2 (\max_{i=0,\ldots,n} \|\hat{\beta}_C^{(i)}(\lambda)\|_2 + \|\hat{\beta}_L(\lambda)\|_2)\end{aligned}$$

The first term is $o_p(1)$ uniformly in λ from Theorem 1, while an immediate consequence of Lemmas 2 and 3 is that the second term is also $o_p(1)$ uniformly in λ, proving the result. □

Proof of Theorem 3. Combining Lemmas 3 and 4, we get $\widetilde{L}_{n,Z}(\lambda) \to \frac{1}{n}||X(\beta - \hat{\beta}_C(\lambda))||_2^2 + \sigma^2$ uniformly over λ. Hence, we have

$$P(\frac{1}{n}||X(\beta - \hat{\beta}_C(\widetilde{\lambda}))||_2^2 - \widetilde{L}_{n,Z}(\widetilde{\lambda}) > \delta - \sigma^2) \to 0 \tag{5}$$

Let λ^* denote the value of λ for which the risk function (4) is minimized. Then, $R_{n,X}(\lambda^*) \leq R_{n,X}(0)$. As $\hat{\beta}_L(0)$ is the OLS estimate, $R_{n.X}(0) \to \sigma^2$ and so does $R_{n,X}(\lambda^*)$. Using Theorem 2, we have $\widetilde{L}_{n,Z}(\lambda^*) \to \sigma^2$ and hence $P(\widetilde{L}_{n,Z}(\widetilde{\lambda}) > \sigma^2 + \delta) \to 0$, as $\widetilde{\lambda}$ minimizes $\widetilde{L}_{n,Z}(\lambda)$.

Combining, this with (5), we have

$$\begin{aligned}P(\frac{1}{n}||X(\beta-\hat{\beta}_C(\widetilde{\lambda}))||_2^2>2\delta)\leq & P(\frac{1}{n}||X(\beta-\hat{\beta}_C(\widetilde{\lambda}))||_2^2-\widetilde{L}_{n,Z}(\widetilde{\lambda})>\delta-\sigma^2)\\ &+P(\widetilde{L}_{n,Z}(\widetilde{\lambda})>\sigma^2+\delta)\\ &\to 0\end{aligned}$$

Since, $X^\top X/n\to\Sigma>0$, this implies $||\hat{\beta}_C(\widetilde{\lambda})-\beta||_2\to 0$. □

Proof of Proposition 1. Since $\tau>\|\beta^*\|_2/\|\beta^*\|_\infty$, without loss of generality we can pick $1.001>c>1$ and $\epsilon_0>0$ such that $\tau>c\|\beta^*\|_2/(\|\beta^*\|_\infty-\epsilon_0)$. Let $\delta>0$ be such that $0<5\delta<c-1$ and let $c_1=c-2\delta$. The choice of δ ensures that $c_1>c_1-3\delta>1$.

Fix $\epsilon>0$. For an orthogonal design the i^{th} component of the Lasso estimate is given by $\hat{\beta}_L(\lambda)[i]=sgn(\frac{1}{n}|x_i^\top y|)(\frac{1}{n}|x_i^\top y|-\lambda)_+$. As $\frac{1}{n}\|X^\top y\|_\infty\xrightarrow{p}\|\beta^*\|_\infty$ in probability, given $\epsilon>0$ there exists n_0 such that for all $n\geq n_0$, we have $P(\mathcal{F}_\delta)\geq 1-\epsilon/4$ where $\mathcal{F}_\delta=\{\|\beta^*\|_\infty+\delta\|\beta^*\|_2/\tau\geq\frac{1}{n}\|X^\top y\|_\infty\geq\|\beta^*\|_\infty-\delta\|\beta^*\|_2/\tau\}$.

Let $\lambda_0=\|\beta^*\|_\infty-c_1\|\beta^*\|_2/\tau$. The choice of c_1 ensures that $\lambda_0>0$. For any $\lambda\leq\lambda_0$, we now have on $\mathcal{F}_\delta$, $|\hat{\beta}_L(\lambda)[i]|\geq(c_1-\delta)\|\beta^*\|_2/\tau$ for some i, and hence $\|\hat{\beta}_L(\lambda)\|_\infty\geq(c_1-\delta)\|\beta^*\|_2/\tau$.

Also, on $\mathcal{F}_\delta$, $\hat{\beta}_L(\lambda^*)=0$ and $L_{n,X}(\lambda^*)=\|\beta^*\|_2^2$ where $\lambda^*=\|\beta^*\|\infty+\delta\|\beta^*\|_2/\tau$ and $L_{n,X}(\lambda)$ is defined in (4). Now choose $\mathcal{G}_\delta=\{\sup\limits_{\lambda\in[0,\Lambda]}|\widehat{L}_{n,Z}(\lambda)-L_{n,X}(\lambda)-\tau^2\|\hat{\beta}_L(\lambda)\|_2^2|\leq\delta^2\|\beta^*\|_2^2\}$. By Lemma 4, there exists n_1 such that for $n\geq n_1$, $P(\mathcal{G}_\delta)\geq 1-\epsilon/4$. Hence, on $\mathcal{F}_\delta\cap\mathcal{G}_\delta$, for any $\lambda\leq\lambda_0$,

$$\begin{aligned}\widehat{L}_{n,Z}(\lambda)-\widehat{L}_{n,Z}(\lambda^*)\geq & L_{n,X}(\lambda)+\tau^2\|\hat{\beta}_L(\lambda)\|_2^2-L_{n,X}(\lambda^*)-\tau^2\|\hat{\beta}_L(\lambda^*)\|_2^2-\\ &-2\sup_{\lambda\in[0,\Lambda]}|\widehat{L}_{n,Z}(\lambda)-L_{n,X}(\lambda)-\tau^2\|\hat{\beta}_L(\lambda)\|_2^2|\\ \geq &((c_1-\delta)^2-1-2\delta^2)\|\beta^*\|_2^2>0 \qquad (\text{ as } 3\delta<c_1-1)\end{aligned}$$

Hence, on $\mathcal{F}_\delta\cap\mathcal{G}_\delta$, $\inf\limits_{\lambda\leq\lambda_0}\widehat{L}_{n,Z}(\lambda)>\widehat{L}_{n,Z}(\lambda^*)$ which implies $\widehat{\lambda}_Z>\lambda_0$. This means $\|\hat{\beta}_L(\widehat{\lambda}_Z)\|_\infty\leq(c_1+\delta)\|\beta^*\|_2/\tau)$. Letting, $\mathcal{H}_\delta=\{\sup\limits_{\lambda\in[0,\Lambda]}\|\hat{\beta}_C(\lambda)-\hat{\beta}_L(\lambda)\|_\infty\leq\delta\|\beta^*\|_2/\tau\}$, by Lemma 3, we can choose n_2 such that $n\geq n_2$ implies $P(\mathcal{H}_\delta)\geq 1-\epsilon/4$.

For large n, on $\mathcal{F}_\delta\cap\mathcal{G}_\delta\cap\mathcal{H}_\delta$, $\|\hat{\beta}_C(\widehat{\lambda}_Z)\|_\infty\leq(c_1+2\delta)\|\beta^*\|_2/\tau\leq c\|\beta^*\|_2/\tau<1.001\|\beta^*\|_2/\tau$. This proves part (a). Now, also observing that $c\|\beta^*\|_2/\tau\leq\|\beta^*\|_\infty-\epsilon_0$, we have for large n,

$$P(\|\hat{\beta}_C(\widehat{\lambda}_Z)-\beta^*\|_\infty\geq\|\beta^*\|_\infty-\|\hat{\beta}_C(\widehat{\lambda}_Z)\|_\infty\geq\epsilon_0)\geq 1-\epsilon$$

This proves the inconsistency result of Part (b). □

B Appendix: Sub-Gaussian random variables

Lemma A1. *If X denote a $n\times p_1$ fixed matrix with $\|x_i\|_2\leq C_x$, A and B respectively denote independent $n\times p_2$ and $n\times p_3$ random matrices whose rows are iid sub-Gaussian with zero* *means and covariance M_A and M_B. Then for any $\epsilon > 0$ and fixed $p_i, i = 1,2,3$, we have $p(\|A^\top A/n - M_A\|_{\max} > \epsilon)$, $p(\|A^\top B/n\|_{\max} > \epsilon)$ and $p(\|A^\top X/n\|_{\max} > \epsilon)$ are all $o(n^{-1})$.*

The proof follows directly from Lemma B.1 of Datta & Zou (2017). Note that, since p is fixed the result in Lemma A1 remains true if we replace the $\max$ norm with the ℓ_2 norm.

Lemma A2. *Let $M = (Z^\top Z/n - \Gamma)$, then $p(\rho_{\min}(M) \le \rho_{\min}(\Sigma)/2)$, $p(\rho_{\max}(M) \ge 2\rho_{\max}(\Sigma))$ and $p(\|M_+ - \Sigma\|_{\max} > \epsilon)$ are $o(n^{-1})$.*

Proof. We expand $M = \Sigma + B$ where $B = (X^\top X/n - \Sigma) + A^\top X/n + X^\top A/n + (A^\top A/n - \Gamma)$. Then we have

$$\rho_{\min}(M) = \inf_{\{u:\|u\|_2=1\}} u^\top M u \ge \rho_{\min}(\Sigma) - \sup_{\{u:\|u\|_2=1\}} |u^\top B u| \ge \rho_{\min}(\Sigma) - \sqrt{p}\|B\|_{\max}$$

Since from Lemma A1, $P(\|B\|_{\max} > \epsilon) = o(n^{-1})$, for large enough n we have with probability $1 - o(n^{-1})$, $\sqrt{p}\|B\|_{\max} \le \rho_{\min}(\Sigma)/2$ and hence $\rho_{\min}(M) \ge \rho_{\min}(\Sigma) - \sqrt{p}\|B\|_{\max} \ge \rho_{\min}(\Sigma)/2$ and the first result is proved. Similarly we can write $\rho_{\max}(M) \le \rho_{\max}(\Sigma) + \sqrt{p}\|B\|_{\max}$ and prove $p(\rho_{\max}(M) \ge 2\rho_{\max}(\Sigma)) = o(n^{-1})$. Also, as $p(\|M_+ - \Sigma\|_{\max} > \epsilon) \le p(\|M - \Sigma\|_{\max} > \epsilon) + p(\rho_{\min}(M) \le \rho_{\min}(\Sigma)/2)$, the third result follows immediately. □

Lemma A3. *Let a sequence of random convex functions $g_n(\beta)$ satisfy $p(|g_n(\beta) - g(\beta)| > \epsilon) = o(n^{-1})$ for any $\epsilon > 0$, pointwise for every β. Also, let $K \subset \mathbb{R}^p$ be any hypercube with edges of length L and assume that $g(\beta)$ is Lipschitz continuous on K with Lipschitz constant κ. Then $p(\sup_{\beta\in K} |g_n(\beta) - g(\beta)| > \epsilon) = o(n^{-1})$.*

Proof. This lemma is simply a more detailed version of the Convexity Lemma in Pollard (1991). The proof is identical with the only additional task being tracking the tail probabilities carefully throughout. We break K up into small hypercubes with edge length $\epsilon/(\sqrt{2}\kappa)$. We cover the boundary of K with an additional layer of such small hypercubes. Due to Lipschitz continuity, within each of these smaller boxes, $g(\beta)$ differs at most by ϵ. Let V denote the set of vertices of all these small boxes. Now, from the proof of the Convexity Lemma in Pollard (1991) we see that

$$-(p+1)(M_n + \epsilon) \le \sup_{\beta\in K} |g_n(\beta) - g(\beta| \le (M_n + \epsilon)$$

where $M_n = \sup_{\beta\in V} |g_n(\beta) - g(\beta|$. Since, V has $(\sqrt{2}L\kappa/\epsilon + 3)^p$ points,

$$p(M_n > \epsilon) \le (\sqrt{2}L\kappa/\epsilon + 3)^p o(n^{-1}) \tag{6}$$

and the result follows. □

References

Datta, A. & Zou, H. (2017), 'Cocolasso for high-dimensional error-in-variables regression', Annals of Statistics **45**(6), 2400–2426.
URL: *https://doi.org/10.1214/16-AOS1527*

Davidson, J. (1994), Stochastic Limit Theory: An Introduction for Econometricians, Oxford University Press.
URL: *http://EconPapers.repec.org/RePEc:oxp:obooks:9780198774037*

Knight, K. & Fu, W. (2000), 'Asymptotics for lasso-type estimators', Ann. Statist. **28**(5), 1356–1378.
URL: *http://dx.doi.org/10.1214/aos/1015957397*

Pollard, D. (1991), 'Asymptotics for least absolute deviation regression estimators', Econometric Theory **7**, 186–199.

Tibshirani, R. J. (2013), 'The lasso problem and uniqueness', Electron. J. Statist. **7**, 1456–1490.
URL: *http://dx.doi.org/10.1214/13-EJS815*