Statistica Sinica: Supplement

The Lasso under Poisson-like Heteroscedasticity

Jinzhu Jia[1], Karl Rohe[2] and Bin Yu[3]

LMAM, School of Mathematical Sciences and Center for Statistical Science,
Peking University, China [1]
Department of Statistics,
University of Wisconsin-Madison [2]
Department of Statistics and Department of EECS,
University of California, Berkeley [3]

Supplementary Material

This supplementary material gives the detailed proofs of the theorems in the paper: The Lasso under Poisson-like Heteroscedasticity.

0.1 Proofs

The Lasso problem is defined as

$$\hat{\beta}(\lambda) = \arg\min_{\beta} \frac{1}{2n}\|Y - \mathbf{X}\beta\|_2^2 + \lambda\|\beta\|_1, \tag{1}$$

where for some vector $x \in \mathbb{R}^k$, $\|x\|_r = (\sum_{i=1}^k |x_i|^r)^{1/r}$.

0.1.1 Proof of Theorem 1

To prove the theorem, we need the next Lemma which gives necessary and sufficient conditions for the Lasso's sign consistency. They are important to the asymptotic analysis. Wainwright (2009) gives this condition which follows from KKT conditions.

Lemma 1. *For linear model $Y = \mathbf{X}\beta^* + \epsilon$, assume that the matrix $X(S)^T X(S)$ is invertible. Then for any given $\lambda > 0$ and any noise term $\epsilon \in R^n$, there exists a Lasso estimate $\hat{\beta}(\lambda)$ which satisfies $\hat{\beta}(\lambda) =_s \beta^*$, if and only if the following two conditions hold*

$$\left| X(S^c)^T X(S)(X(S)^T X(S))^{-1}\left[\frac{1}{n}X(S)^T\epsilon - \lambda sign(\beta^*(S))\right] - \frac{1}{n}X(S^c)^T\epsilon \right| \le \lambda, \tag{2}$$

$$sign\left(\beta^*(S) + (\frac{1}{n}X(S)^T X(S))^{-1}\left[\frac{1}{n}X(S)^T\epsilon - \lambda sign(\beta^*(S))\right]\right) = sign(\beta^*(S)), \tag{3}$$

where the vector inequality and equality are taken elementwise. Moreover, if (2) *holds strictly, then*

$$\hat{\beta} = (\hat{\beta}^{(1)}, 0)$$

is the unique optimal solution to the Lasso problem (1)*, where*

$$\hat{\beta}^{(1)} = \beta^*(S) + (\frac{1}{n}X(S)^T X(S))^{-1}\left[\frac{1}{n}X(S)^T\epsilon - \lambda sign(\beta^*)\right]. \tag{4}$$

As in Wainwright (2009), we state sufficient conditions for (2) and (3). Define

$$\overrightarrow{b} = \mathrm{sign}(\beta^*(S)),$$

and denote by e_i the vector with 1 in the ith position and zeroes elsewhere. Define

$$U_i = e_i^T(\frac{1}{n}X(S)^T X(S))^{-1}\left[\frac{1}{n}X(S)^T\epsilon - \lambda\overrightarrow{b}\right],$$

$$V_j = X_j^T\left\{X(S)(X(S)^T X(S))^{-1}\lambda\overrightarrow{b} - \left[X(S)(X(S)^T X(S))^{-1}X(S)^T - I\right]\frac{\epsilon}{n}\right\}.$$

By rearranging terms, it is easy to see that (2) holds strictly if and only if

$$\mathcal{M}(V) = \left\{\max_{j\in S^c}|V_j| < \lambda\right\} \tag{5}$$

holds. If we define $M(\beta^*) = \min_{j\in S}|\beta_j^*|$ (recall that $S = \{j : \beta_j^* \neq 0\}$ is the sparsity index), then the event

$$\mathcal{M}(U) = \left\{\max_{i\in S}|U_i| < M(\beta^*)\right\}, \tag{6}$$

is sufficient to guarantee that condition (3) holds. Finally, a proof of Theorem 1.

Proof. This proof is divided into two parts. First we analysis the asymptotic probability of event $\mathcal{M}(V)$, and then we analysis the event of $\mathcal{M}(U)$.

Analysis of $\mathcal{M}(V)$: Note from (5) that $\mathcal{M}(V)$ holds if and only if $\frac{\max_{j\in S^c}|V_j|}{\lambda} < 1$. Each random variable V_j is Gaussian with mean

$$\mu_j = \lambda X_j^T X(S)(X(S)^T X(S))^{-1}\overrightarrow{b}.$$

Define $\tilde{V}_j = X_j^T\left[I - X(S)(X(S)^T X(S))^{-1}X(S)^T\right]\frac{\epsilon}{n}$, then $V_j = \mu_j + \tilde{V}_j$. Using the Irrepresentable condition (defined in Equation (5) in the paper), we have $|\mu_j| \le (1-\eta)\lambda$ for all $j \in S^c$, from which we obtain that

$$\frac{1}{\lambda}\max_{j\in S^c}|\tilde{V}_j| < \eta \Rightarrow \frac{\max_{j\in S^c}|V_j|}{\lambda} < 1.$$

By the Gaussian comparison result (17) stated in Lemma 5, we have

$$P\left[\frac{1}{\lambda}\max_{j\in S^c}|\tilde{V}_j|\geq\eta\right]\leq 2(p-q)\exp\{-\frac{\lambda^2\eta^2}{2\max_{j\in S^c}E(\tilde{V}_j^2)}\}.$$

Since

$$E(\tilde{V}_j^2)=\frac{1}{n^2}X_j^T H[VAR(\epsilon)]HX_j,$$

where $H=I-X(S)(X(S)^TX(S))^{-1}X(S)^T$ which has maximum eigenvalue equal to 1, and $VAR(\epsilon)$ is the variance-covariance matrix of ϵ, which is a diagonal matrix with the ith diagonal element equal to $\sigma^2\times|x_i^T\beta^*|$.

Since $|x_i^T\beta^*|\leq\sqrt{\|x_i(S)\|_2^2\|\beta^*\|_2^2}\leq\max_i\|x_i(S)\|_2\|\beta^*\|_2$, an operator bound yields

$$E(\tilde{V}_j^2)\leq\frac{\sigma^2}{n^2}\max_i\|x_i(S)\|_2\|\beta^*\|_2\|X_j\|_2^2=\frac{\sigma^2}{n}\max_i\|x_i(S)\|_2\|\beta^*\|_2.$$

Therefore

$$P\left[\frac{1}{\lambda}\max_j|\tilde{V}_j|\geq\eta\right]\quad\leq\quad 2(p-q)\exp\left\{-\frac{n\lambda^2\eta^2}{2\sigma^2\max_i\|x_i(S)\|_2\|\beta^*\|_2}\right\}.$$

So we have

$$\begin{aligned}P\left[\frac{1}{\lambda}\max_j|V_j|<1\right]&\geq 1-P\left[\frac{1}{\lambda}\max_j|\tilde{V}_j|\geq\eta\right]\\&\geq 1-2(p-q)\exp\left\{-\frac{n\lambda^2\eta^2}{2\sigma^2\|\beta^*\|_2\max_i\|x_i(S)\|_2}\right\}.\end{aligned}$$

Analysis of $\mathcal{M}(U)$:

$$\max_i|U_i|\leq\|(\frac{1}{n}X(S)^TX(S))^{-1}\frac{1}{n}X(S)^T\epsilon\|_\infty+\lambda\|(\frac{1}{n}X(S)^TX(S))^{-1}\overrightarrow{b}\|_\infty$$

Define $Z_i:=e_i^T(\frac{1}{n}X(S)^TX(S))^{-1}\frac{1}{n}X(S)^T\epsilon$. Each Z_i is a normal Gaussian with mean 0 and variance

$$\begin{aligned}var(Z_i)&=e_i^T(\frac{1}{n}X(S)^TX(S))^{-1}\frac{1}{n}X(S)^T[VAR(\epsilon)]\frac{1}{n}X(S)(\frac{1}{n}X(S)^TX(S))^{-1}e_i\\&\leq\frac{\sigma^2\|\beta^*\|_2\max_i\|x_i(S)\|_2}{nC_{\min}}.\end{aligned}$$

So, for any $t>0$, by (17)

$$P(\max_{i\in S}|Z_i|\geq t)\leq 2q\exp\{-\frac{t^2nC_{\min}}{2\sigma^2\|\beta^*\|_2\max_i\|x_i(S)\|_2}\},$$

by taking $t = \frac{\lambda\eta}{\sqrt{C_{\min}}}$, we have

$$P(\max_{i\in S} |Z_i| \geq \frac{\lambda\eta}{\sqrt{C_{\min}}}) \leq 2q \exp\left\{-\frac{n\lambda^2\eta^2}{2\sigma^2\|\beta^*\|_2 \max_i \|x_i(S)\|_2}\right\}.$$

Recall the definition of $\Psi(\mathbf{X}, \beta^*, \lambda) = \lambda\left[\eta\ (C_{\min})^{-1/2} + \left\|\left(\frac{1}{n}X(S)^T X(S)\right)^{-1}\overrightarrow{b}\right\|_\infty\right]$, we have

$$P(\max_i |U_i| \geq \Psi(\mathbf{X}, \beta^*, \lambda)) \leq 2q \exp\left\{-\frac{n\lambda^2\eta^2}{2\sigma^2\|\beta^*\|_2 \max_i \|x_i(S)\|_2}\right\}.$$

By condition $M(\beta^*) > \Psi(\mathbf{X}, \beta^*, \lambda)$, we have

$$P(\max_i |U_i| < M(\beta^*)) \geq 1 - 2q \exp\left\{-\frac{n\lambda^2\eta^2}{2\sigma^2\|\beta^*\|_2 \max_i \|x_i(S)\|_2}\right\}.$$

At last, we have

$$P\left[\mathcal{M}(V)\&\ \mathcal{M}(U)\right] \geq 1 - 2p \exp\left\{-\frac{n\lambda^2\eta^2}{2\sigma^2\|\beta^*\|_2 \max_i \|x_i(S)\|_2}\right\}$$

□

0.1.2 Proof of Corollary 1

Proof. Recall the definition of $\Gamma(\mathbf{X}, \beta^*, \sigma^2)$:

$$\Gamma(\mathbf{X}, \beta^*, \sigma^2) = \frac{\eta^2\ uSNR}{8 \max_i \|x_i(S)\|_2(\eta\ C_{\min}^{-1/2} + \sqrt{q}C_{min}^{-1})^2 \log(p+1)},$$

where $uSNR = \frac{n[M(\beta^*)]^2}{\sigma^2\|\beta^*\|_2}$. So,

$$\frac{n\eta^2}{2\sigma^2\|\beta^*\|_2 \max_i \|x_i(S)\|_2} = \frac{4\Gamma(\mathbf{X}, \beta^*, \sigma^2)(\eta\ C_{\min}^{-1/2} + \sqrt{q}\ C_{\min}^{-1})^2 \log(p+1)}{[M(\beta^*)]^2}$$

By taking

$$\lambda = \frac{M(\beta^*)}{2\left(\eta\ C_{\min}^{-1/2} + \sqrt{q}\ C_{\min}^{-1}\right)},$$

we have

$$\begin{aligned}
\Psi(\mathbf{X}, \beta^*, \lambda) &= \lambda\left[\eta\ (C_{\min})^{-1/2} + \left\|\left(\frac{1}{n}X(S)^T X(S)\right)^{-1}\overrightarrow{b}\right\|_\infty\right] \\
&\leq \lambda\left[\eta\ C_{\min}^{-1/2} + \sqrt{q}C_{\min}^{-1}\right] \\
&= \frac{M(\beta^*)}{2} \\
&< M(\beta^*),
\end{aligned}$$

and

$$\frac{n\lambda^2\eta^2}{2\sigma^2\|\beta^*\|_2 \max_i \|x_i(S)\|_2} = \Gamma(\mathbf{X}, \beta^*, \sigma^2)\log(p+1).$$

So, the probability bound in Theorem 1 greater than

$$1 - 2\exp\left\{-\left(\Gamma(\mathbf{X}, \beta^*, \sigma^2) - 1\right)\log(p+1)\right\},$$

which goes to one when $\Gamma(\mathbf{X}, \beta^*, \sigma^2, \alpha) \to \infty$.

□

0.1.3 Proof of Theorem 2

Proof. First prove (b). Without loss of generality, assume for some $j \in S^c$, $X_j^T X(S)\left(X(S)^T X(S)\right)^{-1}\overrightarrow{b} = 1+\zeta$ with $\zeta > 0$, then $V_j = \lambda(1+\zeta) + \tilde{V}_j$, where $\tilde{V}_j = -[X(S)\left(X(S)^T X(S)\right)^{-1} X(S)^T - I]\frac{\epsilon}{n}$ is a Gaussian random variable with mean 0, so $P(\tilde{V}_j > 0) = \frac{1}{2}$. So, $P(V_j > \lambda) \geq \frac{1}{2}$, which implies that for any λ, Condition (2) (a necessary condition) is violated with probability greater than 1/2.

For claim (a). Condition (3),

$$sign\left(\beta^*(S) + (\frac{1}{n}X(S)^T X(S))^{-1}\left[\frac{1}{n}X(S)^T\epsilon - \lambda sign(\beta^*(S))\right]\right) = sign(\beta^*(S))$$

is also a necessary condition for sign consistency. Since $\frac{1}{n}X(S)^T X(S) = I_{q\times q}$, (3) becomes

$$sign\left(\beta^*(S) + \left[\frac{1}{n}X(S)^T\epsilon - \lambda sign(\beta^*(S))\right]\right) = sign(\beta^*(S)),$$

which implies that

$$sign\left(\beta^*(S) + \frac{1}{n}X(S)^T\epsilon\right) = sign(\beta^*(S)). \tag{7}$$

Without loss of generality, assume for some $j \in S$, $\beta_j^* > 0$. Then (7) implies $\beta_j^* + Z_j > 0$, where $Z_j = e_j^T \frac{1}{n} X(S)^T \epsilon$ is a Gaussian random variable with mean 0, and variance

$$\begin{aligned} var(Z_j) &= e_j^T \frac{1}{n} X(S)^T VAR(\epsilon) \frac{1}{n} X(S) e_j \\ &= \frac{\sigma^2 e_j^T\left[X(S)^T diag(|X\beta^*|)X(S)\right]e_j}{n^2} \\ &= \frac{\beta_j^{*2}}{c_{n,j}^2}, \end{aligned}$$

where the last equality uses the definition of $c_{n,j}^2$ in Theorem 2. To summarize,

$$\begin{aligned}
P[\hat{\beta}(\lambda) =_s \beta^*] &\leq P[\beta_j^* + Z_j > 0] \\
&= P[Z_j > -\beta_j^*] \\
&= P[Z_j < \beta_j^*] \\
&= 1 - \int_{\beta_j^*}^{\infty} \frac{1}{\sqrt{2\pi var(Z_j)}} \exp\{-\frac{x^2}{2var(Z_j)}\} dx \\
&= 1 - \int_{\beta_j^*/\sqrt{var(Z_j)}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{x^2}{2}\} dx \\
&\leq 1 - \frac{1}{\sqrt{2\pi}} \int_{\beta_j^*/\sqrt{var(Z_j)}}^{\infty} (\frac{x}{1+x} + \frac{1}{(1+x)^2}) \exp\{-\frac{x^2}{2}\} dx \\
&= 1 - \frac{\exp\left\{-\frac{\beta_j^{*2}}{2var(Z_j)}\right\}}{\sqrt{2\pi}(1 + \frac{\beta_j^*}{\sqrt{var(Z_j)}})} \\
&= 1 - \frac{\exp\left\{-\frac{c_{n,j}^2}{2}\right\}}{\sqrt{2\pi}(1 + c_{n,j})}.
\end{aligned}$$

□

0.1.4 Proof of Theorem 3

To prove Theorem 3, we need some preliminary results.

Lemma 2. *Conditioned on $X(S)$ and ϵ, the random vector V is Gaussian. Its mean vector is upper bound as*

$$|\ E[V|\epsilon, X(S)]\ | \leq \lambda(1-\eta)\mathbf{1}. \tag{8}$$

Moreover, its conditional covariance takes the form

$$cov[V|\epsilon, X(S)] = M_n \Sigma_{2|1} = M_n[\Sigma_{22} - \Sigma_{21}(\Sigma_{11})^{-1}\Sigma_{12}], \tag{9}$$

where

$$M_n = \lambda^2 \overrightarrow{b}^T (X(S)^T X(S))^{-1} \overrightarrow{b} + \frac{1}{n^2}\epsilon^T[I - X(S)(X(S)^T X(S))^{-1} X(S)^T]\epsilon. \tag{10}$$

Lemma 3. *Let $M_1 = \lambda^2 \overrightarrow{b}^T (X(S)^T X(S))^{-1} \overrightarrow{b}$ and $M_2 = \frac{1}{n^2}\epsilon^T[I - X(S)(X(S)^T X(S))^{-1} X(S)^T]\epsilon$, then $M_n = M_1 + M_2$. We have when n is big enough*

$$P\left[\frac{\lambda^2 q}{2n\tilde{C}_{\max}} \leq M_1 \leq \frac{2\lambda^2 q}{n\tilde{C}_{\min}}\right] \geq 1 - \exp\{-0.01n\}, \tag{11}$$

$$P\left[M_2 \geq \frac{3\sigma^2\sqrt{\tilde{C}_{\max}}\|\beta^*\|_2}{n}\right] \leq \frac{1}{n}. \tag{12}$$

Lemma 4.

$$P\left[\max_{i=1,\dots n}\|x_i(S)\|_2^2 \geq 2\tilde{C}_{\max}\max\left(16q, 4\log n\right)\right] \leq \frac{1}{n}. \tag{13}$$

Proofs of these lemmas can be found in Appendix 0.1.7. Now, we prove Theorem 3.

Analysis of $M(V)$: Define the event $T = \{M_n \geq v^*\}$, where

$$v^* = \frac{2\lambda^2 q}{n\tilde{C}_{\min}} + \frac{3\sigma^2\sqrt{\tilde{C}_{\max}}\|\beta^*\|_2}{n}.$$

By Lemma 3, we have $P[T] \leq \exp\{-0.01n\} + \frac{1}{n}$, when n is big enough.

Let $\mu_j = E[V_j|\epsilon, X(S)]$, $Z_j = V_j - \mu_j$, and $Z = (Z_j)_{j\in S^c}$, then $E[Z|X(S),\epsilon] = 0$ and $cov(Z|X(S),\epsilon) = cov(V|X(S),\epsilon) = M_n\Sigma_{2|1}$.

$$\begin{aligned}\max_{j\in S^c}|V_j| &= \max_{j\in S^c}|\mu_j + Z_j| \\ &\leq \max_{j\in S^c}[|\mu_j| + |Z_j|] \\ &\leq (1-\eta)\lambda + \max_{j\in S^c}|Z_j|.\end{aligned}$$

From this inequality, we have

$$\{\max_{j\in S^c}|Z_j| < \eta\lambda\} \subset \{\max_{j\in S^c}|V_j| < \lambda\}.$$

Define $\tilde{Z}$ to be a zero-mean Gaussian with covariance $v^*\Sigma_{2|1}$. Since

$$\begin{aligned}P\left[\max_{j\in S^c}|Z_j| \geq \eta\lambda \mid T^c\right] &\leq \sum_{j\in S^c} P\left[|Z_j| > \eta\lambda \mid T^c\right] \\ &\leq (p-q)\max_{j\in S^c} P\left[|\tilde{Z}_j| > \eta\lambda\right] \\ &\leq 2(p-q)\exp\{-\frac{\eta^2\lambda^2}{2v^*\tilde{C}_{\max}}\},\end{aligned}$$

we have

$$\begin{aligned}P[\max_{j\in S^c}|V_j| \geq \lambda] &\leq P\left[\max_{j\in S^c}|Z_j| \geq \lambda \mid T^c\right] + P[T] \\ &\leq 2(p-q)\exp\{-\frac{\eta^2\lambda^2}{2v^*\tilde{C}_{\max}}\} + \exp\{-0.01n\} + \frac{1}{n},\end{aligned}$$

when n is big enough. This says that

$$P[\mathcal{M}(V)] \geq 1 - 2(p-q)\exp\{-\frac{\eta^2\lambda^2}{2v^*\tilde{C}_{\max}}\} - \exp\{-0.01n\} - \frac{1}{n}.$$

Analysis of $\mathcal{M}(U)$: Now we analyze $\max_{j\in S}|U_j|$.

$$\max_j |U_j| \leq \left\|(\frac{1}{n}X(S)^T X(S))^{-1}\frac{1}{n}X(S)^T\epsilon\right\|_\infty + \lambda\left\|(\frac{1}{n}X(S)^T X(S))^{-1}\overrightarrow{b}\right\|_\infty.$$

Define $\Lambda_i(\cdot)$ to be the ith largest eigenvalue of a matrix. Since

$$\lambda \left\| (\frac{1}{n} X(S)^T X(S))^{-1} \overrightarrow{b} \right\|_\infty \leq \frac{\lambda\sqrt{q}}{\Lambda_{\min}(\frac{1}{n} X(S)^T X(S))},$$

by Equation (20) in Corollary 1, we have when n is big enough

$$P\left[\lambda \left\| (\frac{1}{n} X(S)^T X(S))^{-1} \overrightarrow{b} \right\|_\infty \leq \frac{2\lambda\sqrt{q}}{\tilde{C}_{\min}} \right] \geq 1 - 2\exp(-0.01n).$$

Let

$$W_i = e_i^T (\frac{1}{n} X(S)^T X(S))^{-1} \frac{1}{n} X(S)^T \epsilon,$$

then conditioned on $X(S)$, W_i is a Gaussian random variable with mean 0, and variance

$$\begin{aligned} var(W_i|X(S)) &= e_i^T (\frac{1}{n} X(S)^T X(S))^{-1} \frac{1}{n} X(S)^T [VAR(\epsilon)] \frac{1}{n} X(S) (\frac{1}{n} X(S)^T X(S))^{-1} e_i \\ &\leq \frac{\sigma^2 \|\beta^*\|_2 \max_i \|x_i(S)\|_2}{n\Lambda_{\min}(\frac{1}{n} X(S)^T X(S))}. \end{aligned}$$

Using (20)

$$P\left[\Lambda_i(\frac{1}{n} X^T X) \geq \frac{1}{2}\tilde{C}_{\min} \right] \geq 1 - 2\exp(-0.01n),$$

and Lemma 4, we have

$$\frac{\sigma^2 \|\beta^*\|_2 \max_i \|x_i(S)\|_2}{n\Lambda_{\min}(\frac{1}{n} X(S)^T X(S))} \leq \frac{2\sigma^2 \|\beta^*\|_2 \sqrt{2\tilde{C}_{\max} \max(16q, 4\log n)}}{n\tilde{C}_{\min}}$$

with probability no less than $1 - 2\exp\{-0.01n\} - \frac{1}{n}$.

Define event

$$\mathcal{T} = \left\{ \frac{\sigma^2 \|\beta^*\|_2 \max_i \|x_i(S)\|_2}{n\Lambda_{\min}(\frac{1}{n} X(S)^T X(S))} \leq \frac{2\sigma^2 \|\beta^*\|_2 \sqrt{2\tilde{C}_{\max} \max(16q, 4\log n)}}{n\tilde{C}_{\min}} \right\},$$

then $P(\mathcal{T}) \geq 1 - 2\exp\{-0.01n\} - \frac{1}{n}$. From the proof of Lemma 5, for any $t > 0$,

$$P(|W_i| > t \mid X(S)) \leq 2\exp\left(-\frac{t^2}{2var(W_i \mid X(S))} \right).$$

The above is also true if we replace $var(W_i \mid X(S)$ with any upper bound. So we have

$$P(|W_i| > t \mid X(S), \mathcal{T}) \leq 2\exp\left\{ -\frac{t^2}{2\frac{2\sigma^2 \|\beta^*\|_2 \sqrt{2\tilde{C}_{\max} \max(16q, 4\log n)}}{n\tilde{C}_{\min}}} \right\}.$$

So

$$
\begin{aligned}
P(|W_i| > t) &\leq P(|W_i| > t \,|\mathcal{T}) + P(\mathcal{T}^c) \\
&\leq 2\exp\left\{ -\frac{t^2}{2\frac{2\sigma^2\|\beta^*\|_2\sqrt{2\tilde{C}_{\max}\max(16q, 4\log n)}}{n\tilde{C}_{\min}}} \right\} + 2\exp\{-0.01n\} + \frac{1}{n}.
\end{aligned}
$$

By takeing $t = A(n, \beta^*, \sigma^2) := \sqrt{\frac{4\sigma^2\|\beta^*\|_2 \log n\sqrt{2\tilde{C}_{\max}\max(16q, 4\log n)}}{n\tilde{C}_{\min}}}$, we have

$$
\begin{aligned}
P\left[\max_{i\in S}|W_i| > A(n, \beta^*, \sigma^2)\right] &\leq \frac{2q}{n} + 2q\exp\{-0.01n\} + \frac{q}{n} \\
&= \frac{3q}{n} + 2q\exp\{-0.01n\}.
\end{aligned}
$$

Summarize,

$$
\begin{aligned}
&P\left[\max_i |U_i| \geq A(n, \beta^*, \sigma^2) + \frac{2\lambda\sqrt{q}}{\tilde{C}_{\min}}\right] \\
\leq\ &\frac{3q}{n} + 2q\exp\{-0.01n\} + 2\exp\{-0.01n\} \quad .
\end{aligned}
$$

At last, if $M(\beta^*) > \tilde{\Psi}(n, \beta^*, \lambda, \sigma^2)$, we have when n is big enough

$$
P\left[\mathcal{M}(V) \ \&\ \mathcal{M}(U)\right] \leq 1 - 2(p-q)\exp\{-\frac{\eta^2\lambda^2}{2v^*\tilde{C}_{\max}}\} - (2q+3)\exp\{-0.01n\} - \frac{1+3q}{n}.
$$

0.1.5 Proofs of Corollary 3

Proof. By taking $\lambda = \frac{[M(\beta^*) - A(n, \beta^*, \sigma^2)]\tilde{C}_{\min}}{4\sqrt{q}}$, we have

$$
\begin{aligned}
\tilde{\Psi}(n, \beta^*, \lambda, \sigma^2) &= A(n, \beta^*, \sigma^2) + \frac{2\lambda\sqrt{q}}{\tilde{C}_{\min}} \\
&= \frac{M(\beta^*) + A(n, \beta^*, \sigma^2)}{2} \\
&< M(\beta^*),
\end{aligned}
$$

where the last inequality uses the assumption that $M(\beta^*) > A(n, \beta^*, \sigma^2)$.

$$
\begin{aligned}
\frac{\lambda^2}{V^*(n,\beta^*,\lambda,\sigma^2)} &= \frac{\lambda^2}{\frac{2\lambda^2 q}{n\tilde{C}_{\min}} + \frac{3\sigma^2\sqrt{\tilde{C}_{\max}}\|\beta^*\|_2}{n}} \\
&= \frac{1}{\frac{2q}{n\tilde{C}_{\min}} + \frac{3\sigma^2\sqrt{\tilde{C}_{\max}}\|\beta^*\|_2}{n\lambda^2}} \\
&= \frac{1}{\frac{2q}{n\tilde{C}_{\min}} + \frac{48\sigma^2 q\sqrt{\tilde{C}_{\max}}\|\beta^*\|_2}{n[M(\beta^*)-A(n,\beta^*,\sigma^2)]^2\tilde{C}^2_{\min}}}.
\end{aligned}
$$

By the definition of $\tilde{\Gamma}(n,\beta^*,\sigma^2)$, we have that

$$
\frac{\lambda^2\eta^2}{2V^*(n,\beta^*,\lambda,\sigma^2)\tilde{C}_{\max}} = \log(p-q+1)\tilde{\Gamma}(n,\beta^*,\sigma^2),
$$

so the probability bound in Theorem 3 now becomes,

$$
\begin{aligned}
&1 - 2\exp\left\{-\frac{\lambda^2\eta^2}{2V^*(n,\beta^*,\lambda,\sigma^2)\tilde{C}_{\max}} + \log(p-q)\right\} - (2q+3)\exp\{-0.01n\} - \frac{1+3q}{n} \\
&= 1 - 2\exp\left\{-\log(p-q+1)\tilde{\Gamma}(n,\beta^*,\sigma^2) + \log(p-q)\right\} - (2q+3)\exp\{-0.01n\} \\
&-\frac{1+3q}{n} \\
&\geq 1 - 2\exp\left\{-\log(p-q+1)[\tilde{\Gamma}(n,\beta^*,\sigma^2) - 1]\right\} - (2q+3)\exp\{-cn\} - \frac{1+3q}{n}
\end{aligned}
$$

If Condition (14) stated in Corollary 3 holds, then $\tilde{\Gamma}(n,\beta^*,\sigma^2,\alpha)\to\infty$ which guarantees $P[\hat{\beta}(\lambda) =_s \beta^*] \to 1$. □

0.1.6 Proof of Theorem 4

Proof. Without loss of generality, assume

$$
e_j^T\Sigma_{21}(\Sigma_{11})^{-1}sign(\beta^*(S)) = 1+\zeta,
$$

for some $j \in S^c$ and $\zeta > 0$. Since $E[V|X(S),\epsilon] = \lambda\Sigma_{21}(\Sigma_{11})^{-1}sign(\beta^*(S))$, V_j conditioned on $X(S)$ and ϵ is a Gaussian random variable with mean $\lambda(1+\zeta)$. So $P[V_j > \lambda(1+\zeta)|X(S),\epsilon] = \frac{1}{2}$, which implies $P[V_j > \lambda|X(S),\epsilon] \geq \frac{1}{2}$. Then we have $P(V_j > \lambda) \geq \frac{1}{2}$. So for any λ,

$$
P[\hat{\beta}(\lambda) =_s \beta^*] \leq P[\max_k V_k \leq \lambda] \leq \frac{1}{2}.
$$

□

0.1.7 Proofs of Lemma 2 – Lemma 4

Proof of Lemma 2

Proof. Conditioned on $X(S)$ and ϵ, the only random component in V_j is the column in the column vector X_j, $j \in S^c$. We know that $(X(S^c)|X(S), \epsilon) \sim (X(S^c)|X(S))$ is Gaussian with mean and covariance

$$E[X(S^c)^T|X(S), \epsilon] = \Sigma_{21}(\Sigma_{11})^{-1}X(S)^T, \tag{14}$$
$$var(X(S^c)|X(S)) = \Sigma_{2|1} = \Sigma_{22} - \Sigma_{21}(\Sigma_{11})^{-1}\Sigma_{12}. \tag{15}$$

Consequently, we have,

$$\begin{aligned} & |E[V|X(S), \epsilon]| \\ = & \Big|\Sigma_{21}(\Sigma_{11})^{-1}X(S)^T \Big\{X(S)(X(S)^T X(S))^{-1}\lambda\overrightarrow{b} \\ & - \Big[X(S)(X(S)^T X(S))^{-1}X(S)^T - I\Big]\frac{\epsilon}{n}\Big\}\Big| \\ = & |\Sigma_{21}(\Sigma_{11})^{-1}\lambda\overrightarrow{b}| \\ \leq & \lambda(1-\eta)\mathbf{1}, \end{aligned}$$

where the last inequality uses Irreprsentable Condition (defined in Equation (13) in the paper).

Now, we compute the elements of the conditional covariance

$$cov(V_j, V_k|\epsilon, X(S)).$$

Let $\vec{\alpha} = X(S)(X(S)^T X(S))^{-1}\lambda\overrightarrow{b} - \Big[X(S)(X(S)^T X(S))^{-1}X(S)^T - I)\Big]\frac{\epsilon}{n}$, then $V_j = X_j^T\vec{\alpha}$. So we have

$$cov(V_j, V_k|\epsilon, X(S)) = \vec{\alpha}^T cov(X_j^T, X_k^T|\epsilon, X(S))\vec{\alpha} = [var(X(S^c)|X(S))]_{jk}\,\vec{\alpha}^T\vec{\alpha}.$$

Consequently,

$$cov(V|\epsilon, X(S)) = \vec{\alpha}^T\vec{\alpha}\ var(X(S^c)|X(S)) = \vec{\alpha}^T\vec{\alpha}\Sigma_{2|1} = \vec{\alpha}^T\vec{\alpha}[\Sigma_{22} - \Sigma_{21}(\Sigma_{11})^{-1}\Sigma_{12}].$$

By careful calculation, we have $\vec{\alpha}^T\vec{\alpha} = M_n$. □

Proof of Lemma 3

Proof. Recall that $M_1 = \lambda^2\overrightarrow{b}^T(X(S)^T X(S))^{-1}\overrightarrow{b}$. So,

$$\frac{\lambda^2 q}{\Lambda_{\max}(X(S)^T X(S))} \leq M_1 \leq \frac{\lambda^2 q}{\Lambda_{\min}(X(S)^T X(S))}.$$

From (20) we have, when n is big enough

$$P\left[\frac{\lambda^2 q}{2n\tilde{C}_{\max}} \leq M_1 \leq \frac{2\lambda^2 q}{n\tilde{C}_{\min}}\right] \geq 1 - 2\exp(-0.01n).$$

Define $\varrho = E[|Z|]$, where $Z \sim N(0,1)$, then for any random variable $R \sim N(0,\sigma^2)$, $E[|R|] = \sigma\varrho$. Since $x_i^T\beta^* \sim N(0, \beta^*(S)^T\Sigma_{11}\beta^*(S))$, we have

$$E[|x_i^T\beta^*|] = \sqrt{\beta^*(S)^T\Sigma_{11}\beta^*(S)}\varrho.$$

We know that $M_2 \leq \frac{1}{n^2}\epsilon^T\epsilon$. Since $E[\epsilon_i^2] = E[E[\epsilon_i^2|X(S)]] = E[\sigma^2|x_i^T\beta^*|] = \sigma^2\sqrt{\beta^*(S)^T\Sigma_{11}\beta^*(S)}\varrho$, and $E[\epsilon_i^4] = E[E[\epsilon_i^4|X(S)]] = 3E[\sigma^4|x_i^T\beta|^2] = 3\sigma^4\beta^*(S)^T\Sigma_{11}\beta^*(S)$, we have

$$\begin{aligned}
& P\left[\frac{\sum_i \epsilon_i^2}{n^2} \geq \frac{\sigma^2(\varrho + \sqrt{3-\varrho^2})\sqrt{\beta^*(S)^T\Sigma_{11}\beta^*(S)}}{n}\right] \\
= \; & P\left[\sum_i \epsilon_i^2 - n\sigma^2\varrho\sqrt{\beta^*(S)^T\Sigma_{11}\beta^*(S)} \geq n\sigma^2\sqrt{3-\varrho^2}\sqrt{\beta^*(S)^T\Sigma_{11}\beta^*(S)}\right] \\
\leq \; & \frac{nvar(\epsilon_i^2)}{n^2\sigma^4(3-\varrho^2)\beta^*(S)^T\Sigma_{11}\beta^*(S)} \\
= \; & \frac{3\sigma^4\beta^*(S)^T\Sigma_{11}\beta^*(S) - \sigma^4\beta^*(S)^T\Sigma_{11}\beta^*(S)\varrho^2}{n\sigma^4(3-\varrho^2)\beta^*(S)^T\Sigma_{11}\beta^*(S)} \\
= \; & \frac{1}{n}
\end{aligned}$$

So

$$P\left[M_2 \geq \frac{\sigma^2(\varrho + \sqrt{3-\varrho^2})\sqrt{\beta^*(S)^T\Sigma_{11}\beta^*(S)}}{n}\right] \leq \frac{1}{n}.$$

While $\sqrt{\beta_1^T\Sigma_{11}\beta^*(S)} \leq \sqrt{\tilde{C}_{\max}}\|\beta\|_2$ and $\varrho = E(|Z|) \leq \sqrt{E(|Z|^2)} = 1$, where Z is a standard normal random variable, so

$$\frac{\sigma^2(\varrho + \sqrt{3-\varrho^2})\sqrt{\beta^*(S)^T\Sigma_{11}\beta^*(S)}}{n} \leq \frac{3\sigma^2\sqrt{\tilde{C}_{\max}}\|\beta^*\|_2}{n}.$$

Then we have

$$P[M_2 \geq \frac{3\sigma^2\sqrt{\tilde{C}_{\max}}\|\beta^*\|_2}{n}] \leq \frac{1}{n}.$$

$\square$

Proof of Lemma 4

Proof. By lemma 6, we have for any $t > q$,

$$P\left[\max_{i=1,\dots n}\|\Sigma_{11}^{-\frac{1}{2}}x_i(S)\|_2^2 \geq 2t\right] \leq n\exp(-t\left[1-2\sqrt{\frac{q}{t}}\right]).$$

Take $t = \max(16q, 4\log n)$, we have

$$\begin{aligned}\exp(-t\left[1-2\sqrt{\tfrac{q}{t}}\right]) &\leq \exp(-t\left[1-2\sqrt{\tfrac{1}{16}}\right]) \\ &= \exp(-\tfrac{t}{2}) \\ &\leq \tfrac{1}{n^2}.\end{aligned}$$

So

$$P\left[\max_{i=1,\dots n}\|\Sigma_{11}^{-\frac{1}{2}}x_i(S)\|_2^2 \geq 2\max(16q, 4\log n)\right] \leq \frac{1}{n}.$$

Since $\|\Sigma_{11}^{-\frac{1}{2}}x_i(S)\|_2^2 \geq \frac{1}{\tilde{C}_{\max}}\|x_i(S)\|_2^2$, we have

$$P\left[\max_{i=1,\dots n}\|x_i(S)\|_2^2 \geq 2\tilde{C}_{\max}\max(16q, 4\log n)\right] \leq \frac{1}{n}. \tag{16}$$

□

0.2 Some Gaussian Comparison Results

Lemma 5. *For any mean zero Gaussian random vector* $(X_1, \dots, X_n)$*, and* $t > 0$*, we have*

$$P(\max_{1\leq i\leq n}|X_i| \geq t) \leq 2n\exp\{-\frac{t^2}{2\max_i E(X_i^2)}\} \tag{17}$$

Proof. Note that the moment generating function of X_i is

$$E(e^{tX_i}) = \exp\{\frac{E(X_i^2)t^2}{2}\}.$$

So for any $t > 0$,

$$P(X_i \geq x) = P(e^{tX_i} \geq e^{tx}) \leq \frac{E(e^{tX_i})}{e^{tx}} = \exp\{\frac{E(X_i^2)t^2}{2} - xt\},$$

by taking $t = \frac{x}{E(X_i^2)}$, we have

$$P(X_i \geq x) \leq \exp\{-\frac{x^2}{2E(X_i^2)}\}.$$

So

$$P(|X_i| \geq t) = 2P(X_i \geq t) \leq 2\exp\{-\frac{t^2}{2E(X_i^2)}\} \leq 2\exp\{-\frac{t^2}{2\max_i E(X_i^2)}\}.$$

So

$$P(\max_{1\le i\le n} |X_i| \ge t) \le 2n \exp\{-\frac{t^2}{2\max_i E(X_i^2)}\}.$$

□

0.3 Large deviation for χ^2 distribution

Lemma 6. *Let $Z_1, \dots, Z_n$ be i.i.d. χ^2-variates with q degrees of freedom. Then for all $t > q$, we have*

$$P\left[\max_{i=1,\dots,n} Z_i > 2t\right] \le n \exp(-t\left[1 - 2\sqrt{\frac{q}{t}}\right]). \tag{18}$$

The proof of this lemma can be found in Obozinski et al. (2008).

0.4 Some useful random matrix results

In this appendix, we use some known concentration inequalities for the extreme eigenvalues of Gaussian random matrices (Davidson and Szarek, 2001) to bound the eigenvalues of a Gaussian random matrix. Although these results hold more generally, our interest here is on scalings (n, q) such that $q/n \to 0$.

Lemma 7 (Davidson and Szarek (2001))**.** *Let $\Gamma \in \mathbb{R}^{n\times q}$ be a random matrix whose entries are i.i.d. from $N(0, 1/n)$, $q \le n$. Let the singular values of Γ be $s_1(\Gamma) \ge \dots \ge s_q(\Gamma)$. Then for any $t > 0$,*

$$\max\left\{P\left[s_1(\Gamma) \ge 1 + \sqrt{\frac{q}{n}} + t\right], P\left[s_q(\Gamma) \le 1 - \sqrt{\frac{q}{n}} - t\right]\right\} < \exp\{-nt^2/2\}.$$

Using Lemma 7, we now have some useful results.

Lemma 8. *Let $U \in \mathbb{R}^{n\times q}$ be a random matrix with elements from the standard normal distribution (i.e., $U_{ij} \sim N(0, 1)$, i.i.d.) Assume that $q/n \to 0$. Let the eigenvalues of $\frac{1}{n}U^TU$ be $\Lambda_1(\frac{1}{n}U^TU) \ge \dots \ge \Lambda_q(\frac{1}{n}U^TU)$. Then when n is big enough,*

$$P\left[\frac{1}{2} \le \Lambda_i(\frac{1}{n}U^TU) \le 2\right] \ge 1 - 2\exp(-0.01n). \tag{19}$$

Proof. Let $\Gamma = \frac{1}{\sqrt{n}}U$, then $\Lambda_i(\frac{1}{n}U^TU) = s_i^2(\Gamma)$. By Lemma 7,

$$P\left[s_q(\Gamma) \le 1 - \sqrt{\frac{q}{n}} - t\right] < \exp\{-nt^2/2\},$$

by taking $t = t_0 = 1 - \frac{\sqrt{2}}{2} - 0.1$, we have

$$P\left[s_q(\Gamma) \leq \frac{\sqrt{2}}{2} + 0.1 - \sqrt{\frac{q}{n}}\right] < \exp\{-nt_0^2/2\}.$$

Since $q/n \to 0$ by assumption, we have when n is big enough, $q/n < 0.1$, then

$$P\left[s_q(\Gamma) < \frac{\sqrt{2}}{2}\right] < \exp\{-nt_0^2/2\},$$

which implies that, for any $i = 1, \ldots, q$,

$$P\left[\Lambda_i(\frac{1}{n}(U^TU)) < \frac{1}{2}\right] < \exp\{-nt_0^2/2\}.$$

Followed the same procedures,

$$P\left[\Lambda_i(\frac{1}{n}(U^TU)) > 2\right] < \exp\{-nt_1^2/2\},$$

for $t_1 = \sqrt{2} - 1.1$. Then inequality (19) holds immediately. □

Corollary 1. *Let $X \in R^{n\times q}$ be a random matrix, of which, the rows are i.i.d. from the normal distribution with mean 0 and covariance Σ. Assume that $\tilde{C}_{\min} \leq \Lambda_i(\Sigma) \leq \tilde{C}_{\max}$ and $q/n \to 0$, then there exist a constant c, when n is big enough,*

$$P\left[\frac{1}{2}\tilde{C}_{\min} \leq \Lambda_i(\frac{1}{n}X^TX) \leq 2\tilde{C}_{\max}\right] \geq 1 - 2\exp(-0.01n). \tag{20}$$

Proof. Let x_i' denote the ith row of X. Let $u_i' = x_i'\Sigma^{-\frac{1}{2}}$, then $var(u_i) = I_{q\times q}$ and matrix U with ith row u_i' satisfies the condition in Lemma 8. Then

$$P\left[\frac{1}{2} \leq \Lambda_i(\frac{1}{n}U^TU) \leq 2\right] \geq 1 - 2\exp(-0.01n).$$

Since

$$\tilde{C}_{\min}\Lambda_1(\frac{1}{n}U^TU) \leq \Lambda_i(\frac{1}{n}X^TX) \leq \tilde{C}_{\max}\Lambda_q(\frac{1}{n}U^TU),$$

result (20) is obtained immediately. □

References

Davidson, K. R. and Szarek, S. J. 2001. *Local operator theory, random matrices, and Banach spaces. In Handbook of Banach Spaces, volume 1, pages 317-366.* Elsevier, Amsterdan, NL.

Obozinski, G., Wainwright, M., Jordan, M. (2008). Union support recovery in high-dimensional multivariate regression. *46 th Annual Allerton Conference on Communication, Control, and Computing,* 21 - 26.

Wainwright, M. (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery using-constrained quadratic programming (Lasso). *IEEE Transactions on Information Theory* 55, 2183–2202.