Supplementary File: Structured Lasso for Regression with Matrix Covariates

Junlong Zhao and Chenlei Leng *

July 4, 2013

1. Proof of Lemma 1

Let $A = \{\|\mathbb{V}\varepsilon\|_\infty/n < \lambda_n\}$ and $B = \{\|\mathbb{V}_j\|_2/\sqrt{n} \leq 1+\delta_0, j = 1, \cdots, pq\}$. Then $P(\mathcal{A}) = P(B)P(A|B)$. Then by the inequality of the tail probability of the normal distribution, we have

$$P(B) = 1 - P(B^c) \geq 1 - \sum_{j=1}^{pq} P\left(\|\mathbb{V}_j\|_2/\sqrt{n} \geq 1+\delta_0\right).$$

Denoting $\delta_1 = (1+\delta_0)^2 - 1$, we have

$$P\left(\|\mathbb{V}_j\|_2/\sqrt{n} \geq 1+\delta_0\right) \leq P\left(\|\mathbb{V}_j\|_2^2 - n \geq n[(1+\delta_0)^2 - 1]\right) = P\left(\|\mathbb{V}_j\|_2^2 - n \geq n\delta_1\right).$$

Note that $\|\mathbb{V}_j\|_2^2$ has a χ_n^2 distribution. Based on the tail probability bound of χ_n^2(that is, $P(\chi_n^2 > n + x) < \exp(-\frac{1}{8}(x, x^2/n)))$. We have $P\left(\|\mathbb{V}_j\|_2/\sqrt{n} \geq 1+\delta_0\right) \leq \exp(-\frac{1}{8}\min(n\delta_1, n\delta_1^2))$. Consequently,

$$P(B) > 1 - \exp(-\frac{1}{8}\min(n\delta_1, n\delta_1^2) + \log(pq)).$$

On the other hand by Lemma C.1 of Zhou (2009), it is easy to see that taking $\lambda_n = (1+\delta_0)\sigma\sqrt{2(1+a)\log(pq)/n}$ for any $a > 0$, we have $P(A|B) \geq 1 - [(pq)^a\sqrt{\pi\log(pq)}]^{-1}$. Therefore the conclusion holds. $\square$

2. Proof of Theorem 1

Step 1. We first show that $\|\hat{\alpha} - \alpha\|_1$ can be small.

Let $u = \hat{\theta} - \theta = \hat{\beta} \otimes \hat{\alpha} - \beta \otimes \alpha$. Note that

$$(\hat{\alpha}, \hat{\beta}) = \arg\min_{(\alpha,\beta)\in\mathcal{E}} \frac{1}{n}\|\mathbb{Y} - \mathbb{V}^T\theta\|_2^2 + \lambda_n P_\theta$$

*Zhao is with LMIB of the Ministry of Education, Beihang University, China (zhaojunlong928@126.com). Leng is with Department of Statistics, University of Warwick, UK and Department of Statistics and Applied Probability, National University of Singapore, Singapore (stalc@nus.edu.sg). Corresponding author: Chenlei Leng.

Conditioning on $\mathcal{A}$, similar to that of Bickel et al. (2009) and Lemma C.2 of Zhou (2009), for $\lambda_n = (1+\delta_0)\sigma\sqrt{2(1+a)\log(pq)/n}$, we have

$$\|u_{S_\theta^c}\|_1 < 3\|u_{S_\theta}\|_1.$$

Furthermore, by the similar procedure of the Proposition C.3 of Zhou (2009), we have

$$\|u\|_1 = \|\hat{\beta}\otimes\hat{\alpha} - \beta\otimes\alpha\|_1 \le B_0\lambda_n s_0. \tag{A.1}$$

For simplicity, we denote $\tilde{\lambda}_n = B_0\lambda_n s_0$. Recall that $\|\hat{\alpha}\|_1 = \|\alpha\|_1 = 1$ and that for any $\beta = (\beta_1,\cdots,\beta_q)^T$, $|\beta_{(1)}| \ge \cdots \ge |\beta_{(q)}|$ is the decreasing order of $|\beta_j|$. For $j = 1,\cdots,q$, define $\hat{a}_j = |\hat{\beta}_j|\|\hat{\alpha}\|_1 = |\hat{\beta}_j|$ and $a_j = |\beta_j|\|\alpha\|_1 = |\beta_j|$. By (A.1), we have

$$\max_{1\le j\le q}\|\hat{\beta}_j\hat{\alpha} - \beta_j\alpha\|_1 \le \tilde{\lambda}_n. \tag{A.2}$$

Therefore, by the triangular inequality, we have

$$\max_{1\le j\le q}\Big||\hat{\beta}_j| - |\beta_j|\Big| = \max_{1\le j\le q}\Big||\hat{\beta}_j|\|\hat{\alpha}\|_1 - |\beta_j|\|\alpha\|_1\Big| \le \tilde{\lambda}_n. \tag{A.3}$$

Let $d_0 = |\beta_{(1)}| - |\beta_{(2)}|$ and denote $k_0 = \arg\max_j|\beta_j|$. Recall our assumption that $\text{sign}(\beta_{k_0}) = 1$. As $\tilde{\lambda}_n < d_0/2$, by (A.3), we have $\hat{a}_{k0} = |\hat{\beta}_{k_0}| = \max_{1\le j\le q}|\hat{\beta}_j|$. By (2.2) and the fact $\text{sign}(\hat{\beta}_{(1)}) = 1$ in the algorithm, we have

$$\text{sign}(\hat{\beta}_{k0}) = 1. \tag{A.4}$$

Denote $\delta_{k0} = \hat{a}_{k0} - a_{k0}$. Then $|\delta_{k0}| \le \tilde{\lambda}_n$. Also by (A.2), we have

$$\begin{aligned}
\tilde{\lambda}_n &\ge \|\hat{\beta}_{k0}\hat{\alpha} - \beta_{k0}\alpha\|_1 = \|\hat{a}_{k0}\hat{\beta}_{k0}\hat{\alpha}/\hat{a}_{k0} - a_{k0}\beta_{k0}\alpha/a_{k0}\|_1 = \|(a_{k0}+\delta_{k0})\hat{\beta}_{k0}\hat{\alpha}/\hat{a}_{k0} - a_{k0}\beta_{k0}\alpha/a_{k0}\|_1\\
&= \|a_{k0}(\hat{\beta}_{k0}\hat{\alpha}/\hat{a}_{k0} - \beta_{k0}\alpha/a_{k0}) + \delta_{k0}\hat{\beta}_{k0}\hat{\alpha}/\hat{a}_{k0}\|_1\\
&\ge \Big|\|a_{k0}(\hat{\beta}_{k0}\hat{\alpha}/\hat{a}_{k0} - \beta_{k0}\alpha/a_{k0})\|_1 - \|\delta_{k0}\hat{\beta}_{k0}\hat{\alpha}/\hat{a}_{k0}\|_1\Big| = \Big|\|a_{k0}(\hat{\beta}_{k0}\hat{\alpha}/\hat{a}_{k0} - \beta_{k0}\alpha/a_{k0})\|_1 - |\delta_{k0}|\Big|.
\end{aligned}$$

Therefore, it follow that $\|a_{k0}(\hat{\beta}_{k0}\hat{\alpha}/\hat{a}_{k0} - \beta_{k0}\alpha/a_{k0})\|_1 \le \tilde{\lambda}_n + |\delta_{k0}| \le 2\tilde{\lambda}_n$. That is,

$$\|(\hat{\beta}_{k0}\hat{\alpha}/\hat{a}_{k0} - \beta_{k0}\alpha/a_{k0})\|_1 \le 2\tilde{\lambda}_n/a_{k0} = 2\tilde{\lambda}_n/|\beta_{k0}|. \tag{A.5}$$

Recalling that $\text{sign}(\beta_{k0}) = 1$, by (A.4), we have $\beta_{k0}/a_{k_0} = \hat{\beta}_{k0}/\hat{a}_{k0} = 1$. Combined with (A.5), we have

$$\|\hat{\alpha} - \alpha\|_1 \leq 2\tilde{\lambda}_n/|\beta_{k0}|. \tag{A.6}$$

Step 2. Based on (A.1) and the fact that $\|\hat{\alpha}\|_1 = \|\alpha\|_1 = 1$, we have

$$\begin{aligned}
\tilde{\lambda}_n &\geq \|\hat{\beta} \otimes \hat{\alpha} - \beta \otimes \alpha\|_1 = \|(\hat{\beta} - \beta) \otimes \hat{\alpha} - \beta \otimes (\alpha - \hat{\alpha})\|_1 \\
&\geq \Big| \|(\hat{\beta} - \beta) \otimes \hat{\alpha}\|_1 - \|\beta \otimes (\alpha - \hat{\alpha})\|_1 \Big| = \Big| \|\hat{\beta} - \beta\|_1 - \|\beta\|_1 \|(\alpha - \hat{\alpha})\|_1 \Big|.
\end{aligned}$$

Consequently, by (A.6), we have

$$\begin{aligned}
\|\hat{\beta} - \beta\|_1 &\leq \tilde{\lambda}_n + \|\beta\|_1 \|\alpha - \hat{\alpha}\|_1 = \tilde{\lambda}_n + \|\beta\|_1 \|\hat{\alpha} - \alpha\|_1 \\
&= \tilde{\lambda}_n (1 + 2\|\beta\|_1/|\beta_{k0}|) = B_0 \lambda_n s_0 (1 + 2\|\beta\|_1/|\beta_{k0}|).
\end{aligned}$$

This completes the proof.

3. Proof of Theorem 2

Let $\Psi = \mathbb{V}\Sigma^{-1/2} = (\Psi_1, \cdots, \Psi_n)^T$ be a n by pq matrix with *i.i.d.* rows $\Psi_i, i = 1, \cdots, n$ from $N(0, I_{pq})$.

Step 1. We first show that, for any $0 < \gamma < 1$, if (3.1) holds, then the following inequality holds with probability at least $1 - \exp\{-\bar{c}\gamma^2 n/\alpha_0^4\}$,

$$1 - \gamma \leq \frac{\|\Psi\Sigma^{1/2} u\|_2}{\sqrt{n}} \leq 1 + \gamma, \quad \text{for all} \ \ u \in \mathcal{J}_1. \tag{A.1}$$

Note that Ψ_i is an isotropic vector with the ψ_2 norm α_0. Let $S^{pq-1} := \{v : v \in R^{pq}, \|v\|_2 = 1\}$. By Theorem 2.5 of Zhou (2009) or Theorem 2.1 of Mendelson et al (2008), for any $0 < \gamma < 1$ and any set $\mathcal{V} \subset S^{pq-1}$, if $n > \frac{c'\alpha_0^4}{\gamma^2}[l_*(\mathcal{V})]^2$, then with probability at least $1 - \exp\{-\bar{c}\gamma^2 n/\alpha_0^4\}$, we have that

$$1 - \gamma \leq \|\Psi v\|_2/\sqrt{n} \leq 1 + \gamma \quad \text{for all} \ \ v \in \mathcal{V},$$

where $c', \bar{c} > 0$ are constants. Therefore, (A.1) can be proved by taking $\mathcal{V} = \mathcal{I}_1$ and computing the complexity measure $l_*(\mathcal{I}_1)$. Lemma 3 gives the following bound on $l_*(\mathcal{I}_1)$

$$l_*(\mathcal{I}_1) \leq C_{s_0,k_0}\sqrt{\log[c(\epsilon, s_0)\max(p^{2s_0}q, pq^{2s_0})]},$$

where $c(\epsilon, s_0)$ and C_{s_0,k_0} are defined in Lemma 3. Therefore, (A.1) holds with probability at least $1-\exp\{-\bar{c}\gamma^2 n/\alpha_0^4\}$ since

$$n > \{c'\alpha_0^4 C_{s_0,k_0}\log[c(\epsilon, s_0)\max(p^{2s_0}q, pq^{2s_0})]\}/\gamma^2.$$

Step 2. We prove that, as (A.1) and (4.1) hold, the structured RE condition $SRE(s_0, k_0)$ holds.

By Proposition 1.4 of Zhou (2009), for any $u \in R^{pq}$, such that $\|u_{S_0^c}\|_1 < k_0\|u_{S_0^c}\|_1$ for some $S_0 \subset \{1, \cdots, pq\}$ with $|S_0| \leq s_0$, we have $\|u_{T_0^c}\|_1 < k_0\|u_{T_0}\|_1$. Recall the definition of u_{T_0}. It follows that set $A_1 \triangleq \{u \in \mathcal{J}_0 : \exists S_0 \text{ with } |S_0| \leq s_0 \text{ such that } \|u_{S_0^c}\|_1 < k_0\|u_{S_0}\|_1\} = \{u \in \mathcal{J}_0 : \|u_{T_0^c}\|_1 < k_0\|u_{T_0}\|_1\} \triangleq A_2$. Recalling the definition of T_0, it follows that $\|u_{S_0}\|_2 \leq \|u_{T_0}\|_2$, for any $S_0 \subset \{1, \cdots, pq\}$ with $|S_0| \leq s_0$. Therefore, RE condition $RE(s_0, k_0, \Sigma)$ in (4.1) is equivalent to

$$\min_{u\in A_1}\frac{\|\Sigma^{1/2}u\|_2}{\|u_{S_0}\|_2} = \min_{u\in A_2}\frac{\|\Sigma^{1/2}u\|_2}{\|u_{T_0}\|_2} = K(s_0, k_0, \Sigma), \tag{A.2}$$

for some $K(s_0, k_0, \Sigma) > 0$. That is, for any $u \in A_2$, we have

$$\|\Sigma^{1/2}u\|_2 \geq \|u_{T_0}\|_2 K(s_0, k_0, \Sigma). \tag{A.3}$$

In addition, for any $u \in A_1$, then $u/\|\Sigma^{1/2}u\|_2 \in \mathcal{J}_1$. From (A.1), it is easy to see that

$$\frac{\|\Psi\Sigma^{1/2}u\|_2}{\sqrt{n}} \geq (1-\gamma)\|\Sigma^{1/2}u\|_2, \quad \text{for all } \ u \in A_1.$$

That is,

$$\frac{\|\mathbb{V}u\|_2}{\sqrt{n}} \geq (1-\gamma)\|\Sigma^{1/2}u\|_2, \quad \text{for all } \ u \in A_1.$$

This combined with (A.3) results in

$$\frac{\|\mathbb{V}u\|_2}{\sqrt{n}} \geq \|\Sigma^{1/2}u\|_2(1-\gamma) \geq (1-\gamma)\|u_{T_0}\|_2 K(s_0, k_0, \Sigma), \quad \text{for all } \ u \in A_1. \tag{A.4}$$

Similar to argument of (A.2), we have

$$\min_{u\in A_1}\frac{\|\mathbb{V}u\|_2}{\sqrt{n}\|u_{S_0}\|_2}=\min_{u\in A_1}\frac{\|\mathbb{V}u\|_2}{\sqrt{n}\|u_{T_0}\|_2}\geq(1-\gamma)K(s_0,k_0,\Sigma)>0,$$

where the last inequality follows from (A.4).

4. Proof of Theorem 3

Define the counterpart $\widetilde{\mathcal{J}}_1$ and $\widetilde{\mathcal{I}}_1$ of $\mathcal{J}_1$ and $\mathcal{I}_1$, respectively, as

$$\widetilde{\mathcal{J}}_1=\{u:u\in R^{pq},\|\Sigma^{1/2}u\|_2=1,u_{T_0^c}\leq k_0u_{T_0}\},\ \widetilde{\mathcal{I}}_1=\{v:v=\Sigma^{1/2}u,u\in\widetilde{\mathcal{J}}_1\}.$$

From Step 1 of Theorem 2, if $n>\frac{c'\alpha_0^4}{\gamma^2}l_*(\widetilde{\mathcal{I}}_1)^2$, then unstructured RE condition $RE(s_0,k_0)$ holds.

Define the counterpart $\widetilde{U}_{s_0}$ of U_{S_0} as $\widetilde{U}_{s_0}=\{u-\beta\otimes\alpha:u\in R^{pq},\|u-\beta\otimes\alpha\|_2=1,|\mathrm{supp}(u)|=s_0\}$ and let $\Pi_{\widetilde{U}_{s_0}}$ be the ϵ-cover of $\widetilde{U}_{s_0}$. Similar to the proof of Lemma 3, we have

$$l_*(\widetilde{\mathcal{I}}_1)\leq\frac{(k_0+2)}{K(s_0,k_0,\Sigma)}\mathbb{E}\sup_{u\in\widetilde{U}_{s_0}}|g^Tu|\leq\frac{6(k_0+2)}{K(s_0,k_0,\Sigma)}\sqrt{\rho_{\max}(s_0)\log|\Pi_{\widetilde{U}_{s_0}}|}.$$

Recalling the definition of $\widetilde{W_0}$ in Section 4.2, similar to the argument of the relation between U_{s_0} and W_0 in Section 4.1, due to $s_0\ll pq$, we have

$$|\Pi_{\widetilde{U}_{s_0}}|\leq|\Pi_{\widetilde{W_0}}|\leq 2s_0\left(\frac{5d_0}{2\epsilon}\right)^{2s_0}\binom{pq}{2s_0},$$

where the last inequality is derived from (4.4). Note that $\binom{pq}{2s_0}\leq(\frac{epq}{2s_0})^{2s_0}$. We have

$$|\Pi_{\widetilde{U}_{s_0}}|\leq 2s_0\left(\frac{5d_0}{2\epsilon}\right)^{2s_0}\left(\frac{epq}{2s_0}\right)^{2s_0}=c_1(s_0,\epsilon)(pq)^{2s_0}.$$

Consequently,

$$l_*(\widetilde{\mathcal{I}}_1)\leq\frac{6(k_0+2)}{K(s_0,k_0,\Sigma)}\sqrt{\rho_{\max}(s_0)\log[c_1(s_0,\epsilon)(pq)^{2s_0}]}.$$

This completes the proof. □

5. Proof of Lemma 3

Step 1. We first prove that

$$l_*(\mathcal{I}_1) \leq \frac{(k_0+2)}{K(s_0,k_0,\Sigma)} \mathbb{E} \sup_{u \in U_{s_0}} |g^T \Sigma^{1/2} u|.$$

By the definition of complexity measure, we have

$$l_*(\mathcal{I}_1) = \mathbb{E} \sup_{v \in \mathcal{I}_1} |g^T v| = \mathbb{E} \sup_{u \in \mathcal{J}_1} |g^T \Sigma^{1/2} u|,$$

where $g = (g_1, \cdots, g_{pq}) \sim N(0, I_{pq})$. For any $J \subset \{1, \cdots, pq\}$, We extend u_J into $u' \in R^{pq}$, such that $u'_{J^c} = 0, u'_J = u_J$. In the following argument, with some abuse of notations, we still use u_J to denote the extended vector u' in R^{pq}. The following argument is similar to that of Zhou (2009). We present here for completeness of the paper. For any $u \in \mathcal{J}_1$, we have

$$\begin{aligned} |g^T \Sigma^{1/2} u| &\leq |g^T \Sigma^{1/2} u_{T_0}| + \sum_{k \geq 1} |g^T \Sigma^{1/2} u_{T_k}| \leq \|u_{T_0}\|_2 \frac{|g^T \Sigma^{1/2} u_{T_0}|}{\|u_{T_0}\|_2} + \sum_{k \geq 1} \|u_{T_k}\|_2 \frac{|g^T \Sigma^{1/2} u_{T_k}|}{\|u_{T_k}\|_2} \\ &\leq (\|u_{T_0}\|_2 + \sum_{k \geq 1} \|u_{T_k}\|_2) \sup_{t \in U_{s_0}} |g^T \Sigma^{1/2} u|. \end{aligned} \tag{A.1}$$

It is easy to see that $\|u_{T_k}\|_2 \leq \sqrt{s_0} \|u_{T_k}\|_\infty \leq \|u_{T_{k-1}}\|_1 / \sqrt{s_0}$. Furthermore, by Lemma 2, we have

$$\sum_{k \geq 1} \|u_{T_k}\|_2 \leq (\|u_{T_0}\|_1 + \sum_{k \geq 1} \|u_{T_j}\|_1)/\sqrt{s_0} \leq (\|u_{T_0}\|_1 + \|u_{T_0^c}\|_1)/\sqrt{s_0} \leq (k_0+1)\|u_{T_0}\|_1/\sqrt{s_0}.$$

Combining with $\|u_{T_0}\|_1/\sqrt{s_0} \leq \|u_{T_0}\|_2$, we have

$$\sum_{k \geq 1} \|u_{T_k}\|_2 \leq (k_0+1)\|u_{T_0}\|_2. \tag{A.2}$$

Furthermore, by the definition of $K(s_0, k_0, \Sigma)$, it follows that

$$\|u_{T_0}\|_2 \leq \|\Sigma^{1/2} u\|_2 / K(s_0, k_0, \Sigma) = 1/K(s_0, k_0, \Sigma). \tag{A.3}$$

The last equation is due to the fact $\|\Sigma^{1/2} u\|_2 = 1$ for any $u \in \mathcal{J}_1$. By (A.1)–(A.3), we have

$$|g^T \Sigma^{1/2} u| \leq (k_0+2)\|u_{T_0}\|_2 \sup_{u \in U_{s_0}} |g^T \Sigma^{1/2} u| \leq \frac{(k_0+2)}{K(s_0,k_0,\Sigma)} \sup_{u \in U_{s_0}} |g^T \Sigma^{1/2} u|.$$

Therefore we have

$$l_*(\mathcal{I}_1) = E\left(\sup_{u\in\mathcal{J}_1} |g^T\Sigma^{1/2}u|\right) \le \frac{(k_0+2)}{K(s_0,k_0,\Sigma)}\mathbb{E}\sup_{u\in U_{s_0}} |g^Tu|.$$

Step 2. We show that $\mathbb{E}\sup_{t\in U_{s_0}} |g^Tu| \le 6\sqrt{\rho_{\max}(s_0)\log[c(\epsilon_0,s_0)\max(p^{2s_0}q, pq^{2s_0})]}$.

By Lemma 2.3 of Mendelson et al. (2008), there exists set $\Pi_{U_{s_0}} = \{u_i \in U_{s_0}, i = 1, \cdots\}$ of the cardinality $|\Pi_{U_{s_0}}|$, such that

$$U_{s_0} \subseteq 2\text{conv}\Pi_{U_{s_0}}.$$

Consequently, we have

$$\mathbb{E}\sup_{t\in U_{s_0}} |g^T\Sigma^{1/2}u| \le 2\mathbb{E}\sup_{u\in\text{conv}\Pi_{U_{s_0}}} |g^T\Sigma^{1/2}u| = 2\mathbb{E}\sup_{u\in\Pi_{U_{s_0}}} |g^T\Sigma^{1/2}u|. \quad (A.4)$$

By the results of Ledoux and Talagrand (1991)(See also Lemma B.5 of Zhou (2009)), we have

$$\mathbb{E}\sup_{u\in\Pi_{U_{s_0}}} |g^T\Sigma^{1/2}u| \le 3\sqrt{\log|\Pi_{U_{s_0}}|}\max_{1\le i\le|\Pi_{U_{s_0}}|}\sqrt{E(g^T\Sigma^{1/2}u_i)^2} \le 3\sqrt{\rho_{\max}(s_0)\log|\Pi_{U_{s_0}}|}, \quad (A.5)$$

where we have used the definition of $\sqrt{\rho_{\max}(s_0)}$ and the fact that $E(g^T\Sigma^{1/2}u_i)^2 = u_i^T\Sigma u_i$ and that $g = (g_1, \cdots, g_{pq}) \sim N(0, I_{pq})$. Also by (4.2), we have $|\Pi_{U_{s_0}}| \le |\Pi_{W_0}|$. Consequently, by the results of Lemma 4, and the argument on the leading term of (4.3) in Section 4.2, we have

$$|\Pi_{W_0}| \le 2s_0\left(\frac{15d_0}{2\epsilon}\right)^{2s_0+1}\max\{qC_p^{2s_0}, pC_q^{2s_0}\} \le c(\epsilon,s_0)\max(qp^{2s_0}, pq^{2s_0}), \quad (A.6)$$

where $c(\epsilon,s_0) = 2s_0\left(\frac{15d_0}{2\epsilon}\right)^{2s_0+1}(\frac{e}{2s_0})^{2s_0}$ and $C_p^{2s_0} = \begin{pmatrix} p \\ 2s_0 \end{pmatrix}$, which is less than $(\frac{ep}{2s_0})^{2s_0}$ and $C_q^{2s_0}$ is defined in analogy. Therefore, by (A.4)– (A.6) and the fact $|\Pi_{U_{s_0}}| \le |\Pi_{W_0}|$, we have

$$\mathbb{E}\sup_{t\in U_{s_0}} |g^T\Sigma^{1/2}u| \le 6\sqrt{\rho_{\max}(s_0)\log|\Pi_{W_0}|} \le 6\sqrt{\rho_{\max}(s_0)\log[c(\epsilon,s_0)\max(p^{2s_0}q, pq^{2s_0})]}.$$

The proof is completed.

6. Proof of Lemma 4

Step 1. For simplicity, we first compute the covering number of the W_1 defined below, which is a special case of W_0 with $d_0 = 1$

$$W_1 = \{w = v_1 \otimes v_2; v_1 \in R^q, v_2 \in R^p, \|v_1 \otimes v_2\|_2 \leq 1, |\text{supp}(v_1)| \cdot |\text{supp}(v_2)| \leq 2s_0, \}.$$

We will show that for $0 < \epsilon \leq 1/2$, there exists ϵ-cover Π_{W_1} of W_1 with

$$|\Pi_{W_1}| \leq \sum_{0<k_1,k_2\in\mathbb{Z}^+,k_1k_2\leq 2s_0} \left(\frac{15}{2\epsilon}\right)^{k_1+k_2} \binom{p}{k_1}\binom{q}{k_2}.$$

In fact, it is easy to see that

$$W_1 = \bigcup_{0<k_1,k_2\in\mathbb{Z}^+,k_1k_2\leq 2s_0} W_{k_1k_2},$$

where $W_{k_1k_2} = \{w = v_1 \otimes v_2; v_1 \in R^p, v_2 \in R^q, \|v_1 \otimes v_2\|_2 \leq 1, |\text{supp}(v_1)| = k_1, |\text{supp}(v_2)| = k_2, \}$. Since s_0 is fixed, here the summation involves only finite terms. For any $m \in \mathbb{Z}^+$, define $S^m = \{v : v \in R^m, \|v\|_2 = 1\}$ and $\mathbb{B}^m = \{v : v \in R^m, \|v\|_2 \leq 1\}$.

For any $v_1 \in R^q, v_2 \in R^p$, set $\tilde{v}_i = v_i/\|v_i\|_2, i = 1, 2$, then $\tilde{v}_1 \in S^{q-1}, \tilde{v}_2 \in S^{p-1}$. Noting that $\|v_1 \otimes v_2\|_2 \leq 1$, it follows that $\|v_1\|_2 \cdot \|v_2\|_2 \leq 1$. Consequently, for any $w \in W_{k_1k_2}$, we have

$$w = v_1 \otimes v_2 = (\|v_1\|_2 \cdot \|v_2\|_2) \cdot \tilde{v}_1 \otimes \tilde{v}_2 \triangleq \dot{v}_1 \otimes \tilde{v}_2.$$

Note that $\{\|v_1\|_2\|v_2\|_2 \cdot \tilde{v}_1 : \tilde{v}_1 \in S^{q-1}, \text{supp}(v_1) = k_1, \|v_1\|_2\|v_2\|_2 \leq 1\} = \{\dot{v}_1 : \dot{v}_1 \in \mathbb{B}^q, \text{supp}(\dot{v}_1) = k_1\}$. For any fixed k_1, k_2, by Lemma 2.3 of Mendelson et al. (2008), for the sets $\{\dot{v}_1 : \dot{v}_1 \in \mathbb{B}^q, \text{supp}(v_1) = k_1\}$ and $\{\tilde{v}_2 : \tilde{v}_2 \in S^{p-1}, \text{supp}(v_2) = k_2\}$, there exist respectively $\epsilon/3$-cover of $\Lambda_1 = \{\dot{v}_{1i}, i = 1, \cdots, \}$ and $\Lambda_2 = \{\tilde{v}_{2i}, i = 1, \cdots, \}$ with

$$|\Lambda_1| \leq \left(\frac{15}{2\epsilon}\right)^{k_1} \binom{q}{k_1}, \quad |\Lambda_2| \leq \left(\frac{15}{2\epsilon}\right)^{k_2} \binom{p}{k_2}.$$

Therefore the ϵ-cover of $W_{k_1k_2}$ can be taken as

$$\Pi_{k_1k_2} = \{\dot{v}_{1i} \otimes \tilde{v}_{2j}, i = 1, \cdots, |\Lambda_1|, j = 1, \cdots, |\Lambda_2|\}.$$

In fact, for any $w = v_1 \otimes v_2 \in W_{k_1k_2}$, we have $w = \dot{v}_1 \otimes \tilde{v}_2$ for some $\dot{v}_1 \in \mathbb{B}^q$ and $\tilde{v}_2 \in S^{p-1}$. There exists i_0, j_0 such that

$$\|\dot{v}_1 - \dot{v}_{1i_0}\|_2 \leq \epsilon/3, \qquad \|\tilde{v}_2 - \tilde{v}_{2j_0}\|_2 \leq \epsilon/3.$$

Consequently, letting $v_1^d = \dot{v}_1 - \dot{v}_{1i_0}$ and $v_2^d = \tilde{v}_2 - \tilde{v}_{2j_0}$ and noting that $\|\dot{v}_{1i_0}\|_2 \leq 1$ and $\|\tilde{v}_{2j_0}\|_2 = 1$, we have

$$\|v_1 \otimes v_2 - \tilde{v}_1 \otimes \tilde{v}_2\|_2 = \|v_1^d \otimes \tilde{v}_{2j_0} + \dot{v}_{1i_0} \otimes v_2^d + v_1^d \otimes v_2^d\|_2 \leq \|v_1^d\|_2 + \|v_2^d\|_2 + \|v_1^d\|_2 \|v_2^d\|_2 \leq \epsilon.$$

Therefore, $\Pi_{k_1k_2}$ is a ϵ-cover of $W_{k_1k_2}$. In addition, by the definition of $\Pi_{k_1k_2}$, we have

$$|\Pi_{k_1k_2}| \leq \left(\frac{15}{2\epsilon}\right)^{k_1+k_2} \binom{q}{k_1}\binom{p}{k_2}.$$

Thus, $\Pi_{W_1} = \bigcup\limits_{k_1,k_2} \Pi_{k_1k_2}$ is the ϵ-cover of W_1 with

$$|\Pi_{W_1}| \leq \sum_{0<k_1,k_2\in\mathbb{Z}^+, k_1k_2\leq 2s_0} \left(\frac{15}{2\epsilon}\right)^{k_1+k_2} \binom{q}{k_1}\binom{p}{k_2}.$$

Step 2. We compute the covering number of W_0. Replacing ϵ by $d_0\epsilon$, v_i by $\sqrt{d_0}v_i$, $i = 1, 2$, and making some small revision of the proof of step 1, we have the conclusion of Lemma 4. This completes the proof. $\square$