ACCURACY ASSESSMENT FOR HIGH-DIMENSIONAL LINEAR REGRESSION†

BY T. TONY CAI, AND ZIJIAN GUO

University of Pennsylvania

This paper considers point and interval estimation of the \( \ell_q \) loss of an estimator in high-dimensional linear regression with random design. We establish the minimax rate for estimating the \( \ell_q \) loss and the minimax expected length of confidence intervals for the \( \ell_q \) loss of rate-optimal estimators of the regression vector, including commonly used estimators such as Lasso, scaled Lasso, square-root Lasso and Dantzig Selector. Adaptivity of the confidence intervals for the \( \ell_q \) loss is also studied. Both the setting of known identity design covariance matrix and known noise level and the setting of unknown design covariance matrix and unknown noise level are studied. The results reveal interesting and significant differences between estimating the \( \ell_2 \) loss and \( \ell_q \) loss with \( 1 \leq q < 2 \) as well as between the two settings.

New technical tools are developed to establish rate sharp lower bounds for the minimax estimation error and the expected length of minimax and adaptive confidence intervals for the \( \ell_q \) loss. A significant difference between loss estimation and the traditional parameter estimation is that for loss estimation the constraint is on the performance of the estimator of the regression vector, but the lower bounds are on the difficulty of estimating its \( \ell_q \) loss. The technical tools developed in this paper can also be of independent interest.

1. Introduction. In many applications, the goal of statistical inference is not only to construct a good estimator, but also to provide a measure of accuracy for this estimator. In classical statistics, when the parameter of interest is one-dimensional, this is achieved in the form of a standard error or a confidence interval. A prototypical example is the inference for a binomial proportion, where often not only an estimate of the proportion but also its margin of error are given. Accuracy measures of an estimation procedure have also been used as a tool for the empirical selection of tuning parameters. A well known example is Stein’s Unbiased Risk Estimate (SURE), which has been an effective tool for the construction of data-driven adaptive estimators in normal means estimation, nonparametric signal recovery, covariance

†The research was supported in part by NSF Grants DMS-1208982 and DMS-1403708, and NIH Grant R01 CA127334.

MSC 2010 subject classifications: Primary 62G15; secondary 62C20, 62H35

Keywords and phrases: Accuracy assessment, adaptivity, confidence interval, high-dimensional linear regression, loss estimation, minimax lower bound, minimaxity, sparsity.
matrix estimation, and other problems. See, for instance, [25, 21, 15, 11, 32].

The commonly used cross-validation methods can also be viewed as a useful tool based on the idea of empirical assessment of accuracy.

In this paper, we consider the problem of estimating the loss of a given estimator in the setting of high-dimensional linear regression, where one observes \((X, y)\) with \(X \in \mathbb{R}^{n \times p}\) and \(y \in \mathbb{R}^n\), and for \(1 \leq i \leq n\),

\[
y_i = X_i \beta + \epsilon_i.
\]

Here \(\beta \in \mathbb{R}^p\) is the regression vector, \(X_i \sim \text{iid} N_p(0, \Sigma)\) are the rows of \(X\), and the errors \(\epsilon_i \sim \text{iid} N(0, \sigma^2)\) are independent of \(X\). This high-dimensional linear model has been well studied in the literature, with the main focus on estimation of \(\beta\). Several penalized/constrained \(\ell_1\) minimization methods, including Lasso [28], Dantzig selector [12], scaled Lasso [26] and square-root Lasso [3], have been proposed. These methods have been shown to work well in applications and produce interpretable estimates of \(\beta\) when \(\beta\) is assumed to be sparse. Theoretically, with a properly chosen tuning parameter, these estimators achieve the optimal rate of convergence over collections of sparse parameter spaces. See, for example, [12, 26, 3, 23, 4, 5, 30].

For a given estimator \(\hat{\beta}\), the \(\ell_q\) loss \(\|\hat{\beta} - \beta\|_q^2\) with \(1 \leq q \leq 2\) is commonly used as a metric of accuracy for \(\hat{\beta}\). We consider in the present paper both point and interval estimation of the \(\ell_q\) loss \(\|\hat{\beta} - \beta\|_q^2\) for a given \(\hat{\beta}\). Note that the loss \(\|\hat{\beta} - \beta\|_q^2\) is a random quantity, depending on both the estimator \(\hat{\beta}\) and the parameter \(\beta\). For such a random quantity, prediction and prediction interval are usually used for point and interval estimation, respectively. However, we slightly abuse the terminologies in the present paper by using estimation and confidence interval to represent the point and interval estimators of the loss \(\|\hat{\beta} - \beta\|_q^2\). Since the \(\ell_q\) loss depends on the estimator \(\hat{\beta}\), it is necessary to specify the estimator in the discussion of loss estimation. Throughout this paper, we restrict our attention to a broad collection of estimators \(\hat{\beta}\) that perform well at least at one interior point or a small subset of the parameter space. This collection of estimators includes most state-of-art estimators such as Lasso, Dantzig selector, scaled Lasso and square-root Lasso.

High-dimensional linear regression has been well studied in two settings. One is the setting with known design covariance matrix \(\Sigma = I\) and known noise level \(\sigma = \sigma_0\) and sparse \(\beta\). See for example, [16, 2, 22, 30, 27, 20, 7, 1, 19]. Another commonly considered setting is sparse \(\beta\) with unknown \(\Sigma\) and \(\sigma\). We study point and interval estimation of the \(\ell_q\) loss \(\|\hat{\beta} - \beta\|_q^2\) in both settings. Specifically, we consider the parameter space \(\Theta_0(k)\) introduced in
HIGH-DIMENSIONAL ACCURACY ASSESSMENT

(2.3), which consists of $k$-sparse signals $\beta$ with known design covariance matrix $\Sigma = I$ and known noise level $\sigma = \sigma_0$, and $\Theta(k)$ defined in (2.4), which consists of $k$-sparse signals with unknown $\Sigma$ and $\sigma$.

1.1. Our contributions. The present paper studies the minimax and adaptive estimation of the loss $\|\hat{\beta} - \beta\|_2^2$ for a given estimator $\hat{\beta}$ and the minimax expected length and adaptivity of confidence intervals for the loss. A major step in our analysis is to establish rate sharp lower bounds for the minimax estimation error and the minimax expected length of confidence intervals for the $\ell_q$ loss over $\Theta_0(k)$ and $\Theta(k)$ for a broad class of estimators of $\beta$, which contains the subclass of rate-optimal estimators. We then focus on the estimation of the loss of rate-optimal estimators and take the Lasso and scaled Lasso estimators as generic examples. For these rate-optimal estimators, we propose procedures for point estimation as well as confidence intervals for their $\ell_q$ losses. It is shown that the proposed procedures achieve the corresponding lower bounds up to a constant factor. These results together establish the minimax rates for estimating the $\ell_q$ loss of rate-optimal estimators over $\Theta_0(k)$ and $\Theta(k)$. The analysis shows interesting and significant differences between estimating the $\ell_2$ loss and $\ell_q$ loss with $1 \leq q < 2$ as well as between the two parameter spaces $\Theta(k)$ and $\Theta_0(k)$.

- The minimax rate for estimating $\|\hat{\beta} - \beta\|_2^2$ over $\Theta_0(k)$ is $\min\left\{\frac{1}{\sqrt{n}}, \frac{k \log p}{n}\right\}$ and over $\Theta(k)$ is $k \frac{\log p}{n}$. So loss estimation is much easier with the prior information $\Sigma = I$ and $\sigma = \sigma_0$ when $\frac{\sqrt{n}}{\log p} \ll k \lesssim \frac{n}{\log p}$.

- The minimax rate for estimating $\|\hat{\beta} - \beta\|_q^2$ with $1 \leq q < 2$ over both $\Theta_0(k)$ and $\Theta(k)$ is $k^\frac{q}{2} \frac{\log p}{n}$.

In the regime $\frac{\sqrt{n}}{\log p} \ll k \lesssim \frac{n}{\log p}$, a practical loss estimator is proposed for estimating the $\ell_2$ loss and shown to achieve the optimal convergence rate $\frac{1}{\sqrt{n}}$ adaptively over $\Theta_0(k)$. We say estimation of loss is impossible if the minimax rate can be achieved by the trivial estimator $0$, which means that the estimation accuracy of the loss is at least of the same order as the loss itself. In all other considered cases, estimation of loss is shown to be impossible. These results indicate that loss estimation is difficult.

We then turn to the construction of confidence intervals for the $\ell_q$ loss. A confidence interval for the loss is useful even when it is “impossible” to estimate the loss, as a confidence interval can provide non-trivial upper and lower bounds for the loss. In terms of convergence rate over $\Theta_0(k)$ or $\Theta(k)$, the minimax rate of the expected length of confidence intervals for the $\ell_q$ loss, $\|\hat{\beta} - \beta\|_q^2$, of any rate-optimal estimator $\hat{\beta}$ coincides with
the minimax estimation rate. We also consider the adaptivity of confidence intervals for the \( \ell_q \) loss of any rate-optimal estimator \( \hat{\beta} \). (The framework for adaptive confidence intervals is discussed in detail in Section 3.1.) Regarding confidence intervals for the \( \ell_2 \) loss in the case of known \( \Sigma = I \) and \( \sigma = \sigma_0 \), a procedure is proposed and is shown to achieve the optimal length \( \frac{1}{\sqrt{n}} \) adaptively over \( \Theta_0(k) \) for \( \frac{\sqrt{n}}{\log p} \lesssim k \lesssim \frac{n}{\log p} \). Furthermore, it is shown that this is the only regime where adaptive confidence intervals exist, even over two given parameter spaces. For example, when \( k_1 \ll \frac{\sqrt{n}}{\log p} \) and \( k_1 \ll k_2 \), it is impossible to construct a confidence interval for the \( \ell_2 \) loss with guaranteed coverage probability over \( \Theta_0(k_2) \) (consequently also over \( \Theta_0(k_1) \)) and with the expected length automatically adjusted to the sparsity. Similarly, for the \( \ell_q \) loss with \( 1 \leq q < 2 \), adaptive confidence intervals is impossible over \( \Theta_0(k_1) \) and \( \Theta_0(k_2) \) for \( k_1 \ll k_2 \lesssim \frac{n}{\log p} \). Regarding confidence intervals for the \( \ell_q \) loss with \( 1 \leq q \leq 2 \) in the case of unknown \( \Sigma \) and \( \sigma \), the impossibility of adaptivity also holds over \( \Theta(k_1) \) and \( \Theta(k_2) \) for \( k_1 \ll k_2 \lesssim \frac{n}{\log p} \).

Establishing rate-optimal lower bounds requires the development of new technical tools. One main difference between loss estimation and the traditional parameter estimation is that for loss estimation the constraint is on the performance of the estimator \( \hat{\beta} \) of the regression vector \( \beta \), but the lower bound is on the difficulty of estimating its loss \( \| \hat{\beta} - \beta \|_q^2 \). We introduce useful new lower bound techniques for the minimax estimation error and the expected length of adaptive confidence intervals for the loss \( \| \hat{\beta} - \beta \|_q^2 \). In several important cases, it is necessary to test a composite null against a composite alternative in order to establish rate sharp lower bounds. The technical tools developed in this paper can also be of independent interest.

In addition to \( \Theta_0(k) \) and \( \Theta(k) \), we also study an intermediate parameter space where the noise level \( \sigma \) is known and the design covariance matrix \( \Sigma \) is unknown but of certain structure. Lower bounds for the expected length of minimax and adaptive confidence intervals for \( \| \hat{\beta} - \beta \|_q^2 \) over this parameter space are established for a broad collection of estimators \( \hat{\beta} \) and are shown to be rate sharp for the class of rate-optimal estimators. Furthermore, the lower bounds developed in this paper have wider implications. In particular, it is shown that they lead immediately to minimax lower bounds for estimating \( \| \beta \|_q^2 \) and the expected length of confidence intervals for \( \| \beta \|_q^2 \) with \( 1 \leq q \leq 2 \).

### 1.2. Comparison with other works.

Statistical inference on the loss of specific estimators of \( \beta \) has been considered in the recent literature. The papers [16, 2] established, in the setting \( \Sigma = I \) and \( n/p \to \delta \in (0, \infty) \), the limit of the normalized loss \( \frac{1}{p} \| \hat{\beta}(\lambda) - \beta \|_2^2 \) where \( \hat{\beta}(\lambda) \) is the Lasso estimator with a pre-specified tuning parameter \( \lambda \). Although [16, 2] provided an
exact asymptotic expression of the normalized loss, the limit itself depends on the unknown $\beta$. In a similar setting, the paper [27] established the limit of a normalized $\ell_2$ loss of the square-root Lasso estimator. These limits of the normalized losses help understand the properties of the corresponding estimators of $\beta$, but they do not lead to an estimate of the loss. Our results imply that although these normalized losses have a limit under some regularity conditions, such losses cannot be estimated well in most settings.

A recent paper, [20], constructed a confidence interval for $\|\hat{\beta} - \beta\|^2_2$ in the case of known $\Sigma = I$, unknown noise level $\sigma$, and moderate dimension where $n/p \rightarrow \xi \in (0, 1)$ and no sparsity is assumed on $\beta$. While no sparsity assumption on $\beta$ is imposed, their method requires the assumption of $\Sigma = I$ and $n/p \rightarrow \xi \in (0, 1)$. In contrast, in this paper, we consider both unknown $\Sigma$ and known $\Sigma = I$ settings, while allowing $p \gg n$ and assuming sparse $\beta$.

Honest adaptive inference has been studied in the nonparametric function estimation literature, including [8] for adaptive confidence intervals for linear functionals, [18, 10] for adaptive confidence bands, and [9, 24] for adaptive confidence balls, and in the high-dimensional linear regression literature, including [22] for adaptive confidence set and [7] for adaptive confidence interval for linear functionals. In this paper, we develop new lower bound tools, Theorems 8 and 9, to establish the possibility of adaptive confidence intervals for $\|\hat{\beta} - \beta\|^q_2$. The connection between $\ell_2$ loss considered in the current paper and the work [22] is discussed in more detail in Section 3.2.

1.3. Organization. Section 2 establishes the minimax lower bounds of estimating the loss $\|\hat{\beta} - \beta\|^2_q$ with $1 \leq q \leq 2$ over both $\Theta_0(k)$ and $\Theta(k)$ and shows that these bounds are rate sharp for the Lasso and scaled Lasso estimators, respectively. We then turn to interval estimation of $\|\hat{\beta} - \beta\|^2_q$. Sections 3 and 4 present the minimax and adaptive minimax lower bounds for the expected length of confidence intervals for $\|\hat{\beta} - \beta\|^2_q$ over $\Theta_0(k)$ and $\Theta(k)$. For Lasso and scaled Lasso estimators, we show that the lower bounds can be achieved and investigate the possibility of adaptivity. Section 5 considers the rate-optimal estimators and establishes the minimax convergence rate of estimating their $\ell_q$ losses. Section 6 presents new minimax lower bound techniques for estimating the loss $\|\hat{\beta} - \beta\|^2_q$. Section 7 discusses the minimaxity and adaptivity in another setting, where the noise level $\sigma$ is known and the design covariance matrix $\Sigma$ is unknown but of certain structure. Section 8 applies the newly developed lower bounds to establish lower bounds for a related problem, that of estimating $\|\beta\|^2_2$. Section 9 proves the main results and additional proofs are given in the supplemental material [6].
1.4. Notation. For a matrix $X \in \mathbb{R}^{n \times p}$, $X_i$, $X_j$, and $X_{i,j}$ denote respectively the $i$-th row, $j$-th column, and $(i,j)$ entry of the matrix $X$. For a subset $J \subset \{1, 2, \cdots, p\}$, $\lvert J \rvert$ denotes the cardinality of $J$, $J^c$ denotes the complement $\{1, 2, \cdots, p\} \setminus J$, $X_J$ denotes the submatrix of $X$ consisting of columns $X_j$ with $j \in J$ and for a vector $x \in \mathbb{R}^p$, $x_J$ is the subvector of $x$ with indices in $J$. For a vector $x \in \mathbb{R}^p$, $\text{supp}(x)$ denotes the support of $x$ and the $\ell_q$ norm of $x$ is defined as $\|x\|_q = (\sum_{i=1}^p |x_i|^q)^{\frac{1}{q}}$ for $q \geq 0$ with $\|x\|_0 = |\text{supp}(x)|$ and $\|x\|_\infty = \max_{1 \leq j \leq p} |x_j|$. For $a \in \mathbb{R}$, $a_+ = \max\{a, 0\}$. We use $\max\|X_{ij}\|_2$ as a shorthand for $\max_{1 \leq j \leq p} \|X_{ij}\|_2$ and $\min\|X_{ij}\|_2$ as a shorthand for $\min_{1 \leq j \leq p} \|X_{ij}\|_2$. For a matrix $A$, we define the spectral norm $\|A\|_2 = \sup_{\|x\|_2=1} \|Ax\|_2$ and the matrix $\ell_q$ norm $\|A\|_{q,1} = \sup_{1 \leq j \leq p} \sum_{i=1}^p |A_{ij}|^{\frac{q}{j}}$; For a symmetric matrix $A$, $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote respectively the smallest and largest eigenvalue of $A$. We use $c$ and $C$ to denote generic positive constants that may vary from place to place. For two positive sequences $a_n$ and $b_n$, $a_n \preceq b_n$ means $a_n \leq Cb_n$ for all $n$ and $a_n \succeq b_n$ if $b_n \preceq a_n$ and $a_n \asymp b_n$ if $a_n \preceq b_n$ and $b_n \preceq a_n$, and $a_n \ll b_n$ if $\limsup_{n \to \infty} \frac{a_n}{b_n} = 0$ and $a_n \gg b_n$ if $b_n \ll a_n$.

2. Minimax estimation of the $\ell_q$ loss. We begin by presenting the minimax framework for estimating the $\ell_q$ loss, $\|\hat{\beta} - \beta\|_q^2$, of a given estimator $\hat{\beta}$, and then establish the minimax lower bounds for the estimation error for a broad collection of estimators $\hat{\beta}$. We also show that such minimax lower bounds can be achieved for the Lasso and scaled Lasso estimators.

2.1. Problem formulation. Recall the high-dimensional linear model,

$$y_{n \times 1} = X_{n \times p} \beta_{p \times 1} + \epsilon_{n \times 1}, \quad \epsilon \sim N_n(0, \sigma^2 I).$$

We focus on the random design with $X_i \overset{iid}{\sim} N(0, \Sigma)$ and $X_i$ and $\epsilon_i$ are independent. Let $Z = (X, y)$ denote the observed data and $\hat{\beta}$ be a given estimator of $\beta$. Denoting by $\hat{L}_q(Z)$ any estimator of the loss $\|\hat{\beta} - \beta\|_q^2$, the minimax rate of convergence for estimating $\|\hat{\beta} - \beta\|_q^2$ over a parameter space $\Theta$ is defined as the largest quantity $\gamma_{\hat{\beta}, \ell_q}(\Theta)$ such that

$$\inf_{\hat{L}_q} \sup_{\theta \in \Theta} \mathbb{P}_{\theta} \left( \|\hat{L}_q(Z) - \|\hat{\beta} - \beta\|_q^2 \| \geq \gamma_{\hat{\beta}, \ell_q}(\Theta) \right) \geq \delta,$$

for some constant $\delta > 0$ not depending on $n$ or $p$. We shall write $\hat{L}_q$ for $\hat{L}_q(Z)$ when there is no confusion.

We denote the parameter by $\theta = (\beta, \Sigma, \sigma)$, which consists of the signal $\beta$, the design covariance matrix $\Sigma$ and the noise level $\sigma$. For a given
\( \theta = (\beta, \Sigma, \sigma) \), we use \( \beta(\theta) \) to denote the corresponding \( \beta \). Two settings are considered: The first is known design covariance matrix \( \Sigma = I \) and known noise level \( \sigma = \sigma_0 \) and the other is unknown \( \Sigma \) and \( \sigma \). In the first setting, we consider the following parameter space that consists of \( k \)-sparse signals,

\[
\Theta_0(k) = \{ (\beta, I, \sigma_0) : \| \beta \|_0 \leq k \},
\]

and in the second setting, we consider

\[
\Theta(k) = \left\{ (\beta, \Sigma, \sigma) : \| \beta \|_0 \leq k, \quad \frac{1}{M_1} \leq \lambda_{\min}(\Sigma) \leq \lambda_{\max}(\Sigma) \leq M_1, \quad 0 < \sigma \leq M_2 \right\},
\]

where \( M_1 \geq 1 \) and \( M_2 > 0 \) are constants. The parameter space \( \Theta_0(k) \) is a subset of \( \Theta(k) \), which consists of \( k \)-sparse signals with unknown \( \Sigma \) and \( \sigma \).

The minimax rate \( \gamma_{\hat{\beta}, \ell_q}(\Theta) \) for estimating \( \| \hat{\beta} - \beta \|_q^2 \) also depends on the estimator \( \hat{\beta} \). Different estimators \( \hat{\beta} \) could lead to different losses \( \| \hat{\beta} - \beta \|_q^2 \) and in general the difficulty of estimating the loss \( \| \hat{\beta} - \beta \|_q^2 \) varies with \( \hat{\beta} \). We first recall the properties of some state-of-art estimators and then specify the collection of estimators on which we focus in this paper. As shown in [12, 4, 3, 26], Lasso, Dantzig Selector, scaled Lasso and square-root Lasso satisfy the following property if the tuning parameter is properly chosen,

\[
\sup_{\theta \in \Theta(k)} P_{\theta} \left( \| \hat{\beta} - \beta \|_q^2 \geq C \frac{k^2 \log p}{n} \right) \rightarrow 0,
\]

where \( C > 0 \) is a constant. The minimax lower bounds established in [30, 23, 31] imply that \( k \frac{2 \log p}{n} \) is the optimal rate for estimating \( \beta \) over the parameter space \( \Theta(k) \). It should be stressed that all of these algorithms do not require knowledge of the sparsity \( k \) and is thus adaptive to the sparsity provided \( k \lesssim \frac{n}{\log p} \). We consider a broad collection of estimators \( \hat{\beta} \) satisfying one of the following two assumptions.

(A1) The estimator \( \hat{\beta} \) satisfies, for some \( \theta_0 = (\beta^*, I, \sigma_0) \in \Theta_0(k) \),

\[
P_{\theta_0} \left( \| \hat{\beta} - \beta^* \|_q^2 \geq C^* \| \beta^* \|_0 \frac{2 \log p}{n} \sigma_0^2 \right) \leq \alpha_0,
\]

where \( 0 \leq \alpha_0 < \frac{1}{4} \) and \( C^* > 0 \) are constants.

(A2) The estimator \( \hat{\beta} \) satisfies

\[
\sup_{\{ \theta = (\beta^*, I, \sigma), \sigma \leq 2\sigma_0 \}} P_{\theta} \left( \| \hat{\beta} - \beta^* \|_q^2 \geq C^* \| \beta^* \|_0 \frac{2 \log p}{n} \sigma_0^2 \right) \leq \alpha_0,
\]

where \( 0 \leq \alpha_0 < \frac{1}{4} \) and \( C^* > 0 \) are constants and \( \sigma_0 > 0 \) is given.
In view of the minimax rate given in (2.5), Assumption (A1) requires \( \hat{\beta} \) to be a good estimator of \( \beta \) at at least one point \( \theta_0 \in \Theta_0(k) \). Assumption (A2) is slightly stronger than (A1) and requires \( \hat{\beta} \) to estimate \( \beta \) well for a single \( \beta^* \) but over a range of noise levels \( \sigma \leq 2\sigma_0 \) while \( \Sigma = I \). Of course, any estimator \( \hat{\beta} \) satisfying (2.5) satisfies both (A1) and (A2). In addition to Assumptions (A1) and (A2), we also introduce the following sparsity assumptions that will be used in various theorems.

**(B1)** Let \( c_0 \) be the constant defined in (9.14). The sparsity levels \( k \) and \( k_0 \) satisfy \( k \leq c_0 \min\{p^{\gamma}, \frac{n}{\log p}\} \) for some constant \( 0 \leq \gamma < \frac{1}{2} \) and \( k_0 \leq c_0 \min\{k, \frac{\sqrt{n}}{\log p}\} \).

**(B2)** The sparsity levels \( k_1, k_2 \) and \( k_0 \) satisfy \( k_1 \leq k_2 \leq c_0 \min\{p^{\gamma}, \frac{n}{\log p}\} \) for some constant \( 0 \leq \gamma < \frac{1}{2} \) and \( c_0 > 0 \) and \( k_0 \leq c_0 \min\{k_1, \frac{\sqrt{n}}{\log p}\} \).

2.2. **Minimax estimation of the \( \ell_q \) loss over \( \Theta_0(k) \).** The following theorem establishes the minimax lower bounds for estimating the loss \( \|\hat{\beta} - \beta\|_2^2 \) over the parameter space \( \Theta_0(k) \).

**Theorem 1.** Suppose that the sparsity levels \( k \) and \( k_0 \) satisfy Assumption (B1). For any estimator \( \hat{\beta} \) satisfying Assumption (A1) with \( \|\beta^*\|_0 \leq k_0 \),

\[
\inf_{L_2} \sup_{\theta \in \Theta_0(k)} \mathbb{P}_\theta \left( |\hat{L}_2 - \|\hat{\beta} - \beta\|_2^2| \geq c \min \left\{ \frac{k \log p}{n}, \frac{1}{\sqrt{n}} \right\} \sigma_0^2 \right) \geq \delta.
\]

For any estimator \( \hat{\beta} \) satisfying Assumption (A2) with \( \|\beta^*\|_0 \leq k_0 \),

\[
\inf_{L_q} \sup_{\theta \in \Theta_0(k)} \mathbb{P}_\theta \left( |\hat{L}_q - \|\hat{\beta} - \beta\|_q^2| \geq c k^{\frac{2}{q}} \frac{\log p}{n} \sigma_0^2 \right) \geq \delta, \quad \text{for } 1 \leq q < 2,
\]

where \( \delta > 0 \) and \( c > 0 \) are constants.

**Remark 1.** Assumption (A1) restricts our focus to estimators that can perform well at at least one point \( (\beta^*, I, \sigma_0) \in \Theta_0(k) \). This weak condition makes the established lower bounds widely applicable as the benchmark for evaluating estimators of the \( \ell_q \) loss of any \( \hat{\beta} \) that performs well at a proper subset, or even a single point of the whole parameter space.

In this paper, we focus on estimating the loss \( \|\hat{\beta} - \beta\|_q^2 \) with \( 1 \leq q \leq 2 \). Similar results can be established for the loss in the form of \( \|\hat{\beta} - \beta\|_q^2 \) with \( 1 \leq q \leq 2 \); Under the same assumptions as those in Theorem 1, the lower bounds for estimating the loss \( \|\hat{\beta} - \beta\|_q^2 \) hold with replacing the convergence rates with their \( \frac{q}{2} \) power; that is, (2.8) remains the same while the convergence...
rate $k^{\frac{2}{q}}(\sqrt{\log p/n}\sigma_0)^2$ in (2.9) is replaced by $k(\sqrt{\log p/n}\sigma_0)^q$. Similarly, all the results established in the rest of the paper for $\|\widehat{\beta} - \beta\|_q^2$ hold for $\|\widehat{\beta} - \beta\|_q^q$ with corresponding convergence rates replaced by their $\frac{q}{2}$ power.

Theorem 1 establishes the minimax lower bounds for estimating the $\ell_2$ loss $\|\widehat{\beta} - \beta\|_2^2$ of any estimator $\widehat{\beta}$ satisfying Assumption (A1) and the $\ell_q$ loss $\|\widehat{\beta} - \beta\|_q^2$ with $1 \leq q < 2$ of any estimator $\widehat{\beta}$ satisfying Assumption (A2). We will take the Lasso estimator as an example and demonstrate the implications of the above theorem. We randomly split $Z = (y, X)$ into subsamples $Z^{(1)} = (y^{(1)}, X^{(1)})$ and $Z^{(2)} = (y^{(2)}, X^{(2)})$ with sample sizes $n_1$ and $n_2$, respectively. The Lasso estimator $\widehat{\beta}^L$ based on the first subsample $Z^{(1)} = (y^{(1)}, X^{(1)})$ is defined as

\begin{equation}
\widehat{\beta}^L = \arg \min_{\beta \in \mathbb{R}^p} \frac{1}{n_1} \|y^{(1)} - X^{(1)}\beta\|_2^2 + \lambda \sum_{j=1}^p \frac{\|X^{(1)}_{ij}\|}{\sqrt{n_1}} |\beta_j|,
\end{equation}

where $\lambda = A\sqrt{\log p/n_1}\sigma_0$ with $A > \sqrt{2}$ being a pre-specified constant. Without loss of generality, we assume $n_1 \approx n_2$. For the case $1 \leq q < 2$, (2.5) and (2.9) together imply that the estimation of the $\ell_q$ loss $\|\widehat{\beta}^L - \beta\|_q^2$ is impossible since the lower bound can be achieved by the trivial estimator of the loss, 0. That is, $\sup_{\beta \in \Theta_0(k)} \mathbb{P} \left( 0 - \|\widehat{\beta}^L - \beta\|_q^2 \right) \geq Ck^{\frac{2}{q}}(\log p/n) \rightarrow 0$.

For the case $q = 2$, in the regime $k \ll \sqrt{n}/\log p$, the lower bound $\frac{k \log p}{n}$ in (2.8) can be achieved by the zero estimator and hence estimation of the loss $\|\widehat{\beta}^L - \beta\|_2^2$ is impossible. However, the interesting case is when $\sqrt{n}/\log p \lesssim k \lesssim \frac{n}{\log p}$, the loss estimator $\tilde{L}_2$ proposed in (2.11) achieves the minimax lower bound $\frac{1}{\sqrt{n}}$ in (2.8), which cannot be achieved by the zero estimator. We now detail the construction of the loss estimator $\tilde{L}_2$. Based on the second half sample $Z^{(2)} = (y^{(2)}, X^{(2)})$, we propose the following estimator,

\begin{equation}
\tilde{L}_2 = \left( \frac{1}{n_2} \left\| y^{(2)} - X^{(2)} \widehat{\beta}^L \right\|_2^2 - \sigma_0^2 \right)_{+}.
\end{equation}

Note that the first subsample $Z^{(1)} = (y^{(1)}, X^{(1)})$ is used to produce the Lasso estimator $\widehat{\beta}^L$ in (2.10) and the second subsample $Z^{(2)} = (y^{(2)}, X^{(2)})$ is retained to evaluate the loss $\|\widehat{\beta}^L - \beta\|_2$. Such sample splitting technique is similar to cross-validation and has been used in [22] for constructing confidence sets for $\beta$ and in [20] for confidence intervals for the $\ell_2$ loss.

The following proposition establishes that the estimator $\tilde{L}_2$ achieves the minimax lower bound of (2.8) over the regime $\sqrt{n}/\log p \lesssim k \lesssim \frac{n}{\log p}$.
Proposition 1. Suppose that \( k \lesssim n \frac{\log p}{\log p} \) and \( \hat{\beta}^L \) is the Lasso estimator defined in (2.10) with \( A > \sqrt{2} \), then the estimator of loss proposed in (2.11) satisfies, for any sequence \( \delta_{n,p} \to \infty \),

\[
\limsup_{n,p} \sup_{\theta \in \Theta_0(k)} \mathbb{P}_\theta \left( \left| \hat{L}_2 - \|\hat{\beta}^L - \beta\|^2_2 \right| \geq \delta_{n,p} \frac{1}{\sqrt{n}} \right) = 0.
\]

2.3. Minimax estimation of the \( \ell_q \) loss over \( \Theta(k) \). We now turn to the case of unknown \( \Sigma \) and \( \sigma \) and establish the minimax lower bound for estimating the \( \ell_q \) loss over the parameter space \( \Theta(k) \).

Theorem 2. Suppose that the sparsity levels \( k \) and \( k_0 \) satisfy Assumption (B1). For any estimator \( \hat{\beta} \) satisfying Assumption (A1) with \( \|\beta^*\|_0 \leq k_0 \),

\[
\inf_{\hat{L}_q} \sup_{\theta \in \Theta(k)} \mathbb{P}_\theta \left( \hat{L}_q - \|\hat{\beta} - \beta\|^2_q \geq c k^q \frac{\log p}{n} \right) \geq \delta, \quad 1 \leq q \leq 2,
\]

where \( \delta > 0 \) and \( c > 0 \) are constants.

Theorem 2 provides a minimax lower bound for estimating the \( \ell_q \) loss of any estimator \( \hat{\beta} \) satisfying Assumption (A1), including the scaled Lasso estimator defined as

\[
\{\beta^{SL}, \hat{\sigma}\} = \arg \min_{\beta \in \mathbb{R}^p, \sigma \in \mathbb{R}^+} \frac{\|y - X\beta\|^2_2}{2n\sigma} + \frac{\sigma}{2} + \lambda_0 \sum_{j=1}^p \frac{\|X_j\|_2 \sqrt{n} |\beta_j|}{\sqrt{n}},
\]

where \( \lambda_0 = A \sqrt{\log p/n} \) with \( A > \sqrt{2} \). Note that for the scaled Lasso estimator, the lower bound in (2.13) can be achieved by the trivial loss estimator \( 0 \) in the sense, \( \sup_{\theta \in \Theta(k)} \mathbb{P}_\theta \left( 0 - \|\beta^{SL} - \beta\|^2_q \geq C k^q \frac{\log p}{n} \right) \to 0 \), and hence estimation of loss is impossible in this case.

3. Minimaxity and adaptivity of confidence intervals over \( \Theta_0(k) \).

We focused in the last section on point estimation of the \( \ell_q \) loss and showed the impossibility of loss estimation except for one regime. The results naturally lead to another question: Is it possible to construct “useful” confidence intervals for \( \|\hat{\beta} - \beta\|^2_q \) that can provide non-trivial upper and lower bounds for the loss? In this section, after introducing the framework for minimaxity and adaptivity of confidence intervals, we consider the case of known \( \Sigma = I \) and \( \sigma = \sigma_0 \) and establish the minimaxity and adaptivity lower bounds for the expected length of confidence intervals for the \( \ell_q \) loss of a broad collection of estimators over the parameter space \( \Theta_0(k) \). We also show that such
minimax lower bounds can be achieved for the Lasso estimator and then discuss the possibility of adaptivity using the Lasso estimator as an example. The case of unknown $\Sigma$ and $\sigma$ will be the focus of the next section.

3.1. Framework for minimaxity and adaptivity of confidence intervals. In this section, we introduce the following decision theoretical framework for confidence intervals of the loss $\|\hat{\beta} - \beta\|_q^2$. Given $0 < \alpha < 1$ and the parameter space $\Theta$ and the loss $\|\hat{\beta} - \beta\|_q^2$, denote by $I_\alpha(\Theta, \hat{\beta}, \ell)$ the set of all $(1 - \alpha)$ level confidence intervals for $\|\hat{\beta} - \beta\|_q^2$ over $\Theta$,

\[
I_\alpha(\Theta, \hat{\beta}, \ell) = \left\{ \text{CI}_\alpha(\hat{\beta}, \ell, Z) = [l(Z), u(Z)] : \inf_{\theta \in \Theta} \mathbb{P}_\theta\left(\|\hat{\beta} - \beta(\theta)\|_q^2 \in \text{CI}_\alpha(\hat{\beta}, \ell, Z)\right) \geq 1 - \alpha \right\}.
\]

We will write $\text{CI}_\alpha$ for $\text{CI}_\alpha(\hat{\beta}, \ell, Z)$ when there is no confusion. For any confidence interval $\text{CI}_\alpha(\hat{\beta}, \ell, Z) = [l(Z), u(Z)]$, its length is denoted by $L(\text{CI}_\alpha(\hat{\beta}, \ell, Z)) = u(Z) - l(Z)$ and the maximum expected length over a parameter space $\Theta_1$ is defined as

\[
L(\text{CI}_\alpha(\hat{\beta}, \ell, Z), \Theta_1) = \sup_{\theta \in \Theta_1} \mathbb{E}_\theta L(\text{CI}_\alpha(\hat{\beta}, \ell, Z)).
\]

For two nested parameter spaces $\Theta_1 \subseteq \Theta_2$, we define the benchmark $L^*_\alpha(\Theta_1, \Theta_2, \hat{\beta}, \ell_q)$, measuring the degree of adaptivity over the nested spaces $\Theta_1 \subset \Theta_2$,

\[
L^*_\alpha(\Theta_1, \Theta_2, \hat{\beta}, \ell_q) = \inf_{\text{CI}_\alpha(\hat{\beta}, \ell_q, Z) \in I_\alpha(\Theta_2, \hat{\beta}, \ell_q)} \sup_{\theta \in \Theta_1} \mathbb{E}_\theta L(\text{CI}_\alpha(\hat{\beta}, \ell_q, Z)).
\]

We will write $L^*_\alpha(\Theta_1, \hat{\beta}, \ell_q)$ for $L^*_\alpha(\Theta_1, \Theta_1, \hat{\beta}, \ell_q)$, which is the minimax expected length of confidence intervals of $\|\hat{\beta} - \beta\|_q^2$ over $\Theta_1$. The benchmark $L^*_\alpha(\Theta_1, \Theta_2, \hat{\beta}, \ell_q)$ is the infimum of the maximum expected length over $\Theta_1$ among all $(1 - \alpha)$-level confidence intervals over $\Theta_2$. In contrast, $L^*_\alpha(\Theta_1, \hat{\beta}, \ell_q)$ is considering all $(1 - \alpha)$-level confidence intervals over $\Theta_1$. In words, if there is prior information that the parameter lies in the smaller parameter space $\Theta_1$, $L^*_\alpha(\Theta_1, \hat{\beta}, \ell_q)$ measures the benchmark length of confidence intervals over the parameter space $\Theta_1$, which is illustrated in the left of Figure 1; however, if there is only prior information that the parameter lies in the larger parameter space $\Theta_2$, $L^*_\alpha(\Theta_1, \Theta_2, \hat{\beta}, \ell_q)$ measures the benchmark length of confidence intervals over the parameter space $\Theta_1$, which is illustrated in the right of Figure 1.
Rigorously, we define a confidence interval CI\(^\ast\) to be simultaneously adaptive over \(\Theta_1\) and \(\Theta_2\) if

\[
L(\text{CI}\(^\ast\), \Theta_1) \asymp L\(^\ast\)(\Theta_1, \hat{\beta}, \ell_q), \quad \text{and} \quad L(\text{CI}\(^\ast\), \Theta_2) \asymp L\(^\ast\)(\Theta_2, \hat{\beta}, \ell_q).
\]

The condition (3.4) means that the confidence interval CI\(^\ast\) has coverage over the larger parameter space \(\Theta_2\) and achieves the minimax rate over both \(\Theta_1\) and \(\Theta_2\). Note that if

\[
L\(^\ast\)(\Theta_1, \Theta_2, \hat{\beta}, \ell_q) \gg L\(^\ast\)(\Theta_1, \hat{\beta}, \ell_q),
\]

then the rate-optimal adaptation (3.4) is impossible to achieve for \(\Theta_1 \subset \Theta_2\). Otherwise, it is possible to construct confidence intervals simultaneously adaptive over parameter spaces \(\Theta_1\) and \(\Theta_2\). The possibility of adaptation over parameter spaces \(\Theta_1\) and \(\Theta_2\) can thus be answered by investigating the benchmark quantities

\[
L\(^\ast\)(\Theta_1, \hat{\beta}, \ell_q) \quad \text{and} \quad L\(^\ast\)(\Theta_1, \Theta_2, \hat{\beta}, \ell_q).
\]

Such framework has already been introduced in [7], which studies the minimaxity and adaptivity of confidence intervals for linear functionals in high-dimensional linear regression.

We will adopt the minimax and adaptation framework discussed above and establish the minimax expected length

\[
L\(^\ast\)(\Theta_0(k), \hat{\beta}, \ell_q)
\]

and the adaptation benchmark

\[
L\(^\ast\)(\Theta_0(k_1), \Theta_0(k_2), \hat{\beta}, \ell_q).
\]

In terms of the minimax expected length and the adaptivity behavior, there exist fundamental differences between the case \(q = 2\) and \(1 \leq q < 2\). We will discuss them separately in the following two sections.

3.2. Confidence intervals for the \(\ell_2\) loss over \(\Theta_0(k)\). The following theorem establishes the minimax lower bound for the expected length of confidence intervals of \(\|\hat{\beta} - \beta\|^2\) over the parameter space \(\Theta_0(k)\).

**Theorem 3.** Suppose that \(0 < \alpha < \frac{1}{4}\) and the sparsity levels \(k\) and \(k_0\) satisfy Assumption (B1). For any estimator \(\hat{\beta}\) satisfying Assumption (A1)
with $\|\beta^*\|_0 \leq k_0$, then there is some constant $c > 0$ such that

$$L^*_\alpha \left( \Theta_0(k), \hat{\beta}, \ell_2 \right) \geq c \min \left\{ \frac{k \log p}{n}, \frac{1}{\sqrt{n}} \right\} \sigma_0^2. \quad (3.5)$$

In particular, if $\hat{\beta}^L$ is the Lasso estimator defined in (2.10) with $A > \sqrt{2}$, then the minimax expected length for $(1 - \alpha)$ level confidence intervals of $\|\hat{\beta}^L - \beta\|_2^2$ over $\Theta_0(k)$ is

$$L^*_\alpha \left( \Theta_0(k), \hat{\beta}^L, \ell_2 \right) \approx \min \left\{ \frac{k \log p}{n}, \frac{1}{\sqrt{n}} \right\} \sigma_0^2. \quad (3.6)$$

We now consider adaptivity of confidence intervals for the $\ell_2$ loss. The following theorem gives the lower bound for the benchmark $L^*_\alpha \left( \Theta_0(k_1), \Theta_0(k_2), \hat{\beta}, \ell_2 \right)$. We will then discuss Theorems 3 and 4 together.

**Theorem 4.** Suppose that $0 < \alpha < \frac{1}{4}$ and the sparsity levels $k_1, k_2$ and $k_0$ satisfy Assumption (B2). For any estimator $\hat{\beta}$ satisfying Assumption (A1) with $\|\beta^*\|_0 \leq k_0$, then there is some constant $c > 0$ such that

$$L^*_\alpha \left( \Theta_0(k_1), \Theta_0(k_2), \hat{\beta}, \ell_2 \right) \geq c \min \left\{ \frac{k_2 \log p}{n}, \frac{1}{\sqrt{n}} \right\} \sigma_0^2. \quad (3.7)$$

In particular, if $\hat{\beta}^L$ is the Lasso estimator defined in (2.10) with $A > \sqrt{2}$, the above lower bound can be achieved.

The lower bound established in Theorem 4 implies that of Theorem 3 and both lower bounds hold for a general class of estimators satisfying Assumption (A1). There is a phase transition for the lower bound of the benchmark $L^*_\alpha \left( \Theta_0(k_1), \Theta_0(k_2), \hat{\beta}, \ell_2 \right)$. In the regime $k_2 \ll \frac{\sqrt{n}}{\log p}$, the lower bound in (3.7) is $\frac{k_2 \log p}{n} \sigma_0^2$; when $\frac{\sqrt{n}}{\log p} \lesssim k_2 \lesssim \frac{n}{\log p}$, the lower bound in (3.7) is $\frac{1}{\sqrt{n}} \sigma_0^2$. For the Lasso estimator $\hat{\beta}^L$ defined in (2.10), the lower bound $\frac{k \log p}{n} \sigma_0^2$ in (3.5) and $\frac{k \log p}{n} \sigma_0^2$ in (3.7) can be achieved by the confidence intervals $\text{CI}^0_\alpha (Z, k, 2)$ and $\text{CI}^0_\alpha (Z, k_2, 2)$ defined in (3.15), respectively. Such an interval estimator is also used for the $\ell_q$ loss with $1 \leq q < 2$. The minimax lower bound $\frac{1}{\sqrt{n}} \sigma_0^2$ in (3.6) and (3.7) can be achieved by the following confidence interval,

$$\text{CI}^0_\alpha (Z) = \left( \frac{\psi (Z)}{\frac{1}{n_2} \chi^2_{q-2} (n_2)} - \sigma_0^2 \right) \left( \frac{\psi (Z)}{\frac{1}{n_2} \chi^2_{q-2} (n_2)} - \sigma_0^2 \right), \quad (3.8)$$
where \( \chi_{1 - \frac{\alpha}{2}}^2 (n_2) \) and \( \chi_{\frac{\alpha}{2}}^2 (n_2) \) are the \( 1 - \frac{\alpha}{2} \) and \( \frac{\alpha}{2} \) quantiles of \( \chi^2 \) random variable with \( n_2 \) degrees of freedom, respectively, and

(3.9) \[
\psi (Z) = \min \left\{ \frac{1}{n_2} \| y^{(2)} - X^{(2)} \hat{\beta}^L \|_2^2, \sigma_0^2 \log p \right\}.
\]

Note that the two-sided confidence interval (3.8) is simply based on the observed data \( Z \), not depending on any prior knowledge of the sparsity \( k \). Furthermore, it is a two-sided confidence interval, which tells not only just an upper bound, but also a lower bound for the loss. The coverage property and the expected length of CI \( \alpha \) (3.8) are established in the following proposition.

**Proposition 2.** Suppose \( k \lesssim \frac{n}{\log p} \) and \( \hat{\beta}^L \) is the estimator defined in (2.10) with \( A > \sqrt{2} \). Then CI \( \alpha \) (3.8) satisfies,

(3.10) \[
\lim_{n,p \to \infty} \inf_{\theta \in \Theta_0 (k)} \mathbb{P} \left( \| \hat{\beta}^L - \beta \|_2^2 \in \text{CI} \alpha (Z) \right) \geq 1 - \alpha,
\]

and

(3.11) \[
L (\text{CI} \alpha (Z), \Theta_0 (k)) \lesssim \frac{1}{\sqrt{n}} \sigma_0^2.
\]

Fig 2. Illustration of \( L^* \alpha (\Theta_0 (k_1), \hat{\beta}^L, \ell_2) \) (top) and \( L^* \alpha (\Theta_0 (k_1), \Theta_0 (k_2), \hat{\beta}^L, \ell_2) \) (bottom) over regimes \( k_1 \leq k_2 \lesssim \frac{\sqrt{n}}{\log p} \) (leftmost), \( k_1 \lesssim \frac{\sqrt{n}}{\log p} \lesssim k_2 \lesssim \frac{n}{\log p} \) (middle) and \( \frac{\sqrt{n}}{\log p} \lesssim k_1 \leq k_2 \lesssim \frac{n}{\log p} \) (rightmost).

Regarding the Lasso estimator \( \hat{\beta}^L \) defined in (2.10), we will discuss the possibility of adaptivity of confidence intervals for \( \| \hat{\beta}^L - \beta \|_2^2 \). The adaptivity behavior of confidence intervals for \( \| \hat{\beta}^L - \beta \|_2^2 \) is demonstrated in Figure 2. As illustrated in the rightmost plot of Figure 2, in the regime \( \frac{\sqrt{n}}{\log p} \lesssim k_1 \leq k_2 \lesssim \frac{n}{\log p} \), we obtain \( L^* \alpha (\Theta_0 (k_2), \Theta_0 (k_2), \hat{\beta}^L, \ell_2) \times L^* \alpha (\Theta_0 (k_1), \hat{\beta}^L, \ell_2) \times \)
which implies that adaptation is possible over this regime. As shown in Proposition 2, the confidence interval $\text{CI}_1^\alpha (Z)$ defined in (3.8) is fully adaptive over the regime $\frac{\sqrt{n}}{\log p} \leq k \lesssim \frac{n}{\log p}$ in the sense of (3.4).

Illustrated in the leftmost and middle plots of Figure 2, it is impossible to construct an adaptive confidence interval for $\|\hat{\beta} - \beta\|_2^2$ over regimes $k_1 \leq k_2 \lesssim \frac{\sqrt{n}}{\log p}$ and $k_1 \ll \frac{n}{\log p}$ since $L^*_\alpha \left( \Theta_0(k_1), \Theta_0(k_2), \hat{\beta}^L, \ell_2 \right) \gg L^*_\alpha \left( \Theta_0(k_1), \hat{\beta}^L, \ell_2 \right)$ if $k_1 \ll \frac{n}{\log p}$ and $k_1 \ll k_2$. To sum up, adaptive confidence intervals for $\|\hat{\beta} - \beta\|_2^2$ is only possible over the regime $\frac{\sqrt{n}}{\log p} \lesssim k \lesssim \frac{n}{\log p}$.

Comparison with confidence balls. We should note that the problem of constructing confidence intervals for $\|\hat{\beta} - \beta\|_2^2$ is related to but different from that of constructing confidence sets for $\beta$ itself. Confidence balls constructed in [22] are of form $\{ \beta : \|\beta - \hat{\beta}\|_2^2 \leq u_n (Z) \}$, where $\hat{\beta}$ can be the Lasso estimator and $u_n (Z)$ is a data dependent squared radius. See [22] for further details. A naive application of this confidence ball leads to a one-sided confidence interval for the loss $\|\hat{\beta} - \beta\|_2^2$, (3.12) $\text{CI}_{\text{induced}}^\alpha (Z) = \left\{ \|\beta - \hat{\beta}\|_2^2 : \|\beta - \hat{\beta}\|_2^2 \leq u_n (Z) \right\}$.

Due to the reason that confidence sets for $\beta$ were sought for in Theorem 1 in [22], confidence sets in the form $\{ \beta : \|\beta - \hat{\beta}\|_2^2 \leq u_n (Z) \}$ will suffice to achieve the optimal length. However, since our goal is to characterize $\|\hat{\beta} - \beta\|_2^2$, we apply the unbiased risk estimation discussed in Theorem 1 of [22] and construct the two-sided confidence interval in (3.8). Such a two-sided confidence interval is more informative than the one-sided confidence interval (3.12) since the one-sided confidence interval does not contain the information whether the loss is close to zero or not. Furthermore, as shown in [22], the length of confidence interval $\text{CI}_{\text{induced}}^\alpha (Z)$ over the parameter space $\Theta_0(k)$ is of order $\frac{1}{\sqrt{n}} + \frac{k \log p}{n}$. The two-sided confidence interval $\text{CI}_1^\alpha (Z)$ constructed in (3.8) is of expected length $\frac{1}{\sqrt{n}}$, which is much shorter than $\frac{1}{\sqrt{n}} + \frac{k \log p}{n}$ in the regime $k \gg \frac{n}{\log p}$. That is, the two-sided confidence interval (3.8) provides a more accurate interval estimator of the $\ell_2$ loss. This is illustrated in Figure 3.

The lower bound technique developed in the literature of adaptive confidence sets [22] can also be used to establish some of the lower bound results for the case $q = 2$ given in the present paper. However, new techniques are needed in order to establish the rate sharp lower bounds for the minimax estimation error (2.9) in the region $\frac{\sqrt{n}}{\log p} \leq k \lesssim \frac{n}{\log p}$ and for the
expected length of the confidence intervals (3.18) and (7.3) in the region \(\frac{\sqrt{n}}{\log p} \lesssim k_1 \leq k_2 \lesssim \frac{n}{\log p}\), where it is necessary to test a composite null against a composite alternative in order to establish rate sharp lower bounds.

3.3. Confidence intervals for the \(\ell_q\) loss with \(1 \leq q < 2\) over \(\Theta_0(k)\). We now consider the case \(1 \leq q < 2\) and investigate the minimax expected length and adaptivity of confidence intervals for \(\|\hat{\beta} - \beta\|_q^2\) over the parameter space \(\Theta_0(k)\). The following theorem characterizes the minimax convergence rate for the expected length of confidence intervals.

**Theorem 5.** Suppose that \(0 < \alpha < \frac{1}{4}, 1 \leq q < 2\) and the sparsity levels \(k\) and \(k_0\) satisfy Assumption (B1). For any estimator \(\hat{\beta}\) satisfying Assumption (A2) with \(\|\beta^*\|_0 \leq k_0\), then there is some constant \(c > 0\) such that

\[
L^\star_{\alpha} \left(\Theta_0(k), \hat{\beta}, \ell_q\right) \geq ck^2 \frac{\log p}{n} \sigma_0^2.
\]

In particular, if \(\hat{\beta}^L\) is the Lasso estimator defined in (2.10) with \(A > 4\sqrt{2}\), then the minimax expected length for \((1 - \alpha)\) level confidence intervals of \(\|\hat{\beta}^L - \beta\|_q^2\) over \(\Theta_0(k)\) is

\[
L^\star_{\alpha} \left(\Theta_0(k), \hat{\beta}^L, \ell_q\right) \sim k^2 \frac{\log p}{n} \sigma_0^2.
\]

We now construct the confidence interval achieving the minimax convergence rate of (3.14),

\[
CL^0_{\alpha} (Z, k, q) = \left(0, C^* (A, k) k^{\frac{1}{q}} \frac{\log p}{n}\right),
\]

where \(C^* (A, k) = \max\left\{\frac{(22A\sigma_0)^2}{n^2}, \left(\frac{3\eta_0}{\eta_0 + 2}ight) \frac{A^2}{n}, \left(\frac{3\eta_0}{\eta_0 + 2}ight) \frac{A^2}{n}\right\}\) with \(\eta_0 = 1.01 \frac{\sqrt{A} + \sqrt{2}}{\sqrt{A} - \sqrt{2}}\). The following proposition establishes the coverage property and the expected length of \(CL^0_{\alpha} (Z, k, q)\).
Proposition 3. Suppose $k \lesssim \frac{n}{\log p}$ and $\hat{\beta}^L$ is the estimator defined in (2.10) with $A > 4\sqrt{2}$. For $1 \leq q \leq 2$, the confidence interval $\text{CI}_\alpha^0 (Z, k, q)$ defined in (3.15) satisfies

$$\lim inf_{n,p \to \infty} \inf_{\theta \in \Theta_0(k)} \mathbb{P}_\theta \left( \| \hat{\beta} - \beta \|_q \in \text{CI}_\alpha^0 (Z, k, q) \right) = 1,$$

and

$$L \left( \text{CI}_\alpha^0 (Z, k, q), \Theta_0 (k) \right) \lesssim k^\frac{1}{2} \frac{\log p}{n} \sigma_0^2.$$

In particular, for the case $q = 2$, (3.16) and (3.17) also hold for the estimator $\hat{\beta}^L$ defined in (2.10) with $A > \sqrt{2}$.

This result shows that the confidence interval $\text{CI}_\alpha^0 (Z, k, q)$ achieves the minimax rate given in (3.14). In contrast to the $\ell_2$ loss where the two-sided confidence interval (3.8) is significantly shorter than the one-sided interval and achieves the optimal rate over the regime $\sqrt{n} \log p \lesssim k \lesssim \frac{n}{\log p}$, for the $\ell_q$ loss with $1 \leq q < 2$, the one-sided confidence interval achieves the optimal rate given in (3.14).

We now consider adaptivity of confidence intervals. The following theorem establishes the lower bound for $L^*_\alpha \left( \Theta_0(k_1), \Theta_0(k_2), \hat{\beta}^L, \ell_q \right)$ with $1 \leq q < 2$.

**Theorem 6.** Suppose $0 < \alpha < \frac{1}{4}$, $1 \leq q < 2$ and the sparsity levels $k_1, k_2$ and $k_0$ satisfy Assumption (B2). For any estimator $\hat{\beta}$ satisfying Assumption (A2) with $\| \hat{\beta}^* \|_0 \leq k_0$, then there is some constant $c > 0$ such that

$$L^*_\alpha \left( \Theta_0(k_1), \Theta_0(k_2), \hat{\beta}, \ell_q \right) \geq \begin{cases} c k_1^2 \frac{\log p}{n} \sigma_0^2 & \text{if } k_1 \leq k_2 \lesssim \frac{\sqrt{n}}{\log p}, \\ c k_2^2 \frac{\log p}{n} \sigma_0^2 & \text{if } k_1 \lesssim \frac{\sqrt{n}}{\log p} \lesssim k_2 \lesssim \frac{n}{\log p}, \\ c k_2^2 \frac{1}{\sqrt{n}} \sigma_0^2 & \text{if } \frac{\sqrt{n}}{\log p} \lesssim k_1 \lesssim k_2 \lesssim \frac{n}{\log p}, \end{cases}$$

In particular, if $p \geq n$ and $\hat{\beta}^L$ is the Lasso estimator defined in (2.10) with $A > 4\sqrt{2}$, the above lower bounds can be achieved.
\( \hat{\beta}^L \) defined in (2.10), by comparing Theorem 5 and Theorem 6, we obtain \( \mathbf{L}_\alpha^* (\Theta_0(k_1), \Theta_0(k_2), \hat{\beta}^L, \ell_q) \gg \mathbf{L}_\alpha^* (\Theta_0(k_1), \hat{\beta}^L, \ell_q) \) if \( k_1 \ll k_2 \), which implies the impossibility of constructing adaptive confidence intervals for the case \( 1 \leq q < 2 \). There exists marked difference between the case \( 1 \leq q < 2 \) and the case \( q = 2 \), where it is possible to construct adaptive confidence intervals over the regime \( \frac{\sqrt{n}}{\log p} < k \lesssim \frac{n}{\log p} \).

For the Lasso estimator \( \hat{\beta}^L \) defined in (2.10), it is shown in Proposition 3 that the confidence interval \( \text{CI}_0^\alpha(Z, k_2, q) \) defined in (3.15) achieves the lower bound \( k_2^{\frac{2}{n}} \log p \sigma_0^2 \) of (3.18). The lower bounds \( k_2^{\frac{2}{n}} k_1^{\log p} \sigma_0^2 \) and \( k_2^{\frac{2}{n}} \frac{1}{\sqrt{n}} \sigma_0^2 \) of (3.18) can be achieved by the following proposed confidence interval, (3.19)

\[
\text{CI}_\alpha^\beta(Z, k_2, q) = \left( \left( \frac{\psi(Z)}{(1/2)^{n/2}} (n/2) - \sigma_0^2 \right)^2 + (16k_2)^{\frac{2}{n}} (1/2)^{n/2} (n/2) - \sigma_0^2 \right),
\]

where \( \psi(Z) \) is given in (3.9). The following result verifies the above claim.

**Proposition 4.** Suppose \( p \geq n, k_1 \leq k_2 \leq \frac{p}{\log p} \) and \( \hat{\beta}^L \) is defined in (2.10) with \( A > 4\sqrt{2} \). Then \( \text{CI}_\alpha^\beta(Z, k_2, q) \) defined in (3.19) satisfies,

\[
\liminf_{n,p \to \infty} \inf_{\Theta \subseteq \Theta_0(k_2)} \mathbb{P}_\theta \left( \| \hat{\beta} - \beta \|_q^2 \in \text{CI}_\alpha^\beta(Z, k_2, q) \right) \geq 1 - \alpha,
\]

and

\[
\mathbf{L} \left( \text{CI}_\alpha^\beta(Z, k_2, q), \Theta_0(k_1) \right) \lesssim k_2^{\frac{2}{n}} \left( k_1^{\log p} + \frac{1}{\sqrt{n}} \right) \sigma_0^2;
\]

**4. Minimaxity and adaptivity of confidence intervals over \( \Theta(k) \).**

In this section, we focus on the case of unknown \( \Sigma \) and \( \sigma \) and establish the rates of convergence for the minimax expected length of confidence intervals for \( \| \hat{\beta} - \beta \|_q^2 \) with \( 1 \leq q \leq 2 \) over \( \Theta(k) \) defined in (2.4). We also study the possibility of adaptivity of confidence intervals for \( \| \hat{\beta} - \beta \|_q^2 \). The following theorem establishes the lower bounds for the benchmark quantities \( \mathbf{L}_\alpha^* \left( \Theta(k_i), \hat{\beta}, \ell_q \right) \) with \( i = 1, 2 \) and \( \mathbf{L}_\alpha^* \left( \Theta(k_1), \Theta(k_2), \hat{\beta}, \ell_q \right) \).

**Theorem 7.** Suppose that \( 0 < \alpha < \frac{1}{4}, 1 \leq q \leq 2 \) and the sparsity levels \( k_1, k_2 \) and \( k_0 \) satisfy Assumption (B2). For any estimator \( \hat{\beta} \) satisfying Assumption (A1) at \( \theta_0 = (\beta^*, I, \sigma_0) \) with \( \| \beta^* \|_0 \leq k_0 \), there is a constant \( c > 0 \) such that

\[
\mathbf{L}_\alpha^* \left( \Theta(k_i), \hat{\beta}, \ell_q \right) \geq ck_i^{\frac{2}{n}} \log p, \quad \text{for} \quad i = 1, 2;
\]
In particular, if $\hat{\beta}^{SL}$ is the scaled Lasso estimator defined in (2.14) with $A > 2\sqrt{2}$, then the above lower bounds can be achieved.

The lower bounds (4.1) and (4.2) hold for any $\hat{\beta}$ satisfying Assumption (A1) at an interior point $\theta_0$, including the scaled Lasso estimator as a special case. We demonstrate the impossibility of adaptivity of confidence intervals for the $\ell_q$ loss of the scaled Lasso estimator $\hat{\beta}^{SL}$ defined in (2.14). Since $L^*_\alpha \left( \Theta (k_1), \Theta (k_2), \hat{\beta}^{SL}, \ell_q \right) \geq L^*_\alpha \left( \{\theta_0\}, \Theta (k_2), \hat{\beta}^{SL}, \ell_q \right)$, by (4.2), we have $L^*_\alpha \left( \Theta (k_1), \Theta (k_2), \hat{\beta}^{SL}, \ell_q \right) \gg L^*_\alpha \left( \Theta (k_1), \hat{\beta}^{SL}, \ell_q \right)$ if $k_1 < k_2$.

The comparison of $L^*_\alpha \left( \Theta (k_1), \hat{\beta}^{SL}, \ell_q \right)$ and $L^*_\alpha \left( \Theta (k_1), \Theta (k_2), \hat{\beta}^{SL}, \ell_q \right)$ is illustrated in Figure 4. Referring to the adaptivity defined in (3.4), it is impossible to construct adaptive confidence intervals for $\|\hat{\beta}^{SL} - \beta\|_q^2$.

**Theorem 7** shows that for any confidence interval $CI_\alpha \left( \hat{\beta}, \ell_q, Z \right)$ for the loss of any estimator $\hat{\beta}$ satisfying Assumption (A1), under the coverage constraint that $CI_\alpha \left( \hat{\beta}, \ell_q, Z \right) \in I_\alpha \left( \Theta (k_2), \hat{\beta}, \ell_q \right)$, its expected length at any given $\theta_0 = (\beta^*, I, \sigma) \in \Theta (k_0)$ must be of order $k_2^2 \frac{\log p}{n}$. In contrast to Theorem 4 and 6, Theorem 7 demonstrates that confidence intervals must be long at a large subset of points in the parameter space, not just at a small number of “unlucky” points. Therefore, the lack of adaptivity for confidence intervals is not due to the conservativeness of the minimax framework.

In the following, we detail the construction of confidence intervals for $\|\hat{\beta}^{SL} - \beta\|_q^2$. The construction of confidence intervals is based on the following...
definition of restricted eigenvalue, which is introduced in [4],

\[
\kappa(X, k_s, \alpha_0) = \min_{J_0 \subset \{1, \ldots, p\}, \delta \neq 0, \|\delta\|_0 \leq k_s} \frac{\|X\delta\|_2}{\sqrt{n}\|\delta_{J_0}\|_2},
\]

where \(J_1\) denotes the subset corresponding to the \(s\) largest in absolute value coordinates of \(\delta\) outside of \(J_0\) and \(J_{01} = J_0 \cup J_1\). Define the event \(B = \{\hat{\sigma} \leq \log p\}\). The confidence interval for \(\|\hat{\beta}^{SL} - \beta\|_q\) is defined as

\[
\text{CI}_\alpha(Z,k,q) = \begin{cases} [0, \varphi(Z,k,q)] & \text{on } B, \\ \{0\} & \text{on } B^c, \end{cases}
\]

where

\[
\varphi(Z,k,q) = \min \left\{ \left( \frac{16 \text{Amx}_j \|X_{.j}\|_2}{n \kappa^2 (X,k,k,3 \max_{j} \|X_{.j}\|_2)} \right)^2 \frac{\hat{\sigma}^2}{k_0^2 q \log p}, \left( \frac{2 \log p}{n} \right)^2 \right\}.
\]

Properties of \(\text{CI}_\alpha(Z,k,q)\) are established as follows.

**Proposition 5.** Suppose \(k \lesssim \frac{n}{\log p}\) and \(\hat{\beta}^{SL}\) is the estimator defined in (2.14) with \(A > 2\sqrt{2}\). For \(1 \leq q \leq 2\), then \(\text{CI}_\alpha(Z,k,q)\) defined in (4.4) satisfies the following properties,

\[
\lim_{n,p \to \infty} \inf_{\theta \in \Theta(k)} \mathbb{P}_{\theta} \left( \|\hat{\beta} - \beta\|_q^2 \in \text{CI}_\alpha(Z,k,q) \right) = 1,
\]

and

\[
\mathcal{L}(\text{CI}_\alpha(Z,k,q), \Theta(k)) \lesssim k_0^2 \frac{\log p}{n}.
\]

Proposition 5 shows that the confidence interval \(\text{CI}_\alpha(Z,k_1,q)\) defined in (4.4) achieves the lower bound in (4.1), for \(i = 1, 2\), and the confidence interval \(\text{CI}_\alpha(Z,k_2,q)\) defined in (4.4) achieves the lower bound in (4.2).

5. Estimation of the \(\ell_q\) loss of rate-optimal estimators. We have established the minimax lower bounds for the estimation accuracy of the loss of a broad class of estimators \(\hat{\beta}\) satisfying the weak assumptions (A1) or (A2) and also demonstrated that such minimax lower bounds are sharp for the Lasso estimator or scaled Lasso estimator. In this section, we will show that the minimax lower bounds are sharp for the class of rate-optimal estimators satisfying the following Assumption (A).
(A) The estimator $\hat{\beta}$ satisfies, for $k \ll \frac{n}{\log p}$,

$$
\sup_{\theta \in \Theta(k)} P_{\theta}\left(\|\hat{\beta} - \beta\|_q^2 \geq C^*\|\beta\|_0^2 \frac{\frac{1}{2}\log p}{n}\right) \leq Cp^{-\delta},
$$

for constants $\delta > 0$, $C^* > 0$ and $C > 0$.

We say an estimator $\hat{\beta}$ is rate-optimal if it satisfies Assumption (A). As shown in [12, 4, 3, 26], Lasso, Dantzig Selector, scaled Lasso and square-root Lasso are rate-optimal when the tuning parameter is chosen properly. We shall stress that Assumption (A) implies Assumptions (A1) and (A2). Assumption (A) requires the estimator $\hat{\beta}$ to perform well over the whole parameter space $\Theta(k)$ while Assumptions (A1) and (A2) only require $\hat{\beta}$ to perform well at a single point or over a proper subset. The following proposition shows that the minimax lower bounds established in Theorem 1 to Theorem 7 can be achieved for the class of rate-optimal estimators.

**Proposition 6.** Let $\hat{\beta}$ be an estimator satisfying Assumption (A).

1. There exist estimators of the loss $\|\hat{\beta} - \beta\|_q^2$ with $1 \leq q < 2$ achieving, up to a constant factor, the minimax lower bounds (2.9) in Theorem 1 and (3.13) in Theorem 5 and estimators of loss $\|\hat{\beta} - \beta\|_q^2$ with $1 \leq q \leq 2$ achieving, up to a constant factor, the minimax lower bounds (2.13) in Theorem 2 and (4.1) in Theorem 7.

2. Suppose that the estimator $\hat{\beta}$ is constructed based on the subsample $Z^{(1)} = (y^{(1)}, X^{(1)})$, then there exist estimators of the loss $\|\hat{\beta} - \beta\|_2^2$ achieving, up to a constant factor, the minimax lower bounds (2.8) in Theorem 1, (3.5) in Theorem 3 and (3.7) in Theorem 4.

3. Suppose the estimator $\hat{\beta}$ is constructed based on the subsample $Z^{(1)} = (y^{(1)}, X^{(1)})$ and it satisfies Assumption (A) with $\delta > 2$ and the assumption $\|\hat{\beta} - \beta\|_1 \leq C^*\|\hat{\beta} - \beta\|_1$, where $S = \text{supp}(\beta)$. Then for $p \geq n$ there exist estimators of the loss $\|\hat{\beta} - \beta\|_q^2$ with $1 \leq q < 2$ achieving the minimax lower bounds (3.18) in Theorem 6.

For reasons of space, we do not discuss the detailed construction of confidence intervals achieving these minimax lower bounds here and postpone the construction to the proof of Proposition 6.

**Remark 2.** Sample splitting has been widely used in the literature. For example, the condition that $\hat{\beta}$ is constructed based on the subsample $Z^{(1)} = (y^{(1)}, X^{(1)})$ has been introduced in [22] for constructing confidence sets for $\beta$ and in [20] for constructing confidence intervals for the $\ell_2$ loss. Such a
condition is imposed purely for technical reasons to create the independence between the estimator $\hat{\beta}$ and the subsample $Z^{(2)} = (y^{(2)}, X^{(2)})$, which is used to evaluate the $\ell_q$ loss of the estimator $\hat{\beta}$. As shown in [4], the assumption $\| (\hat{\beta} - \beta)_{S^c} \|_1 \leq c^* \| (\hat{\beta} - \beta)_{S} \|_1$ is satisfied for Lasso and Dantzig Selector.

6. General tools for minimax lower bounds. A major step in our analysis is to establish rate sharp lower bounds for the estimation error and expected length of confidence intervals for the $\ell_q$ loss. We introduce in this section new technical tools that are needed to establish these lower bounds.

A significant distinction of the lower bound results given in the previous sections from those for the traditional parameter estimation problems is that the constraint is on the performance of the estimator $\hat{\beta}$ of the regression vector $\beta$, but the lower bounds are on the difficulty of estimating its loss $\| \hat{\beta} - \beta \|_q$. It is necessary to develop new lower bound techniques to establish rate-optimal lower bounds for the estimation error and the expected length of confidence intervals for the loss $\| \hat{\beta} - \beta \|_q$. These technical tools may also be of independent interest.

We begin with notation. Let $Z$ denote a random variable whose distribution is indexed by some parameter $\theta \in \Theta$ and let $\pi$ denote a prior on the parameter space $\Theta$. We will use $f_\theta(z)$ to denote the density of $Z$ given $\theta$ and $f_\pi(z)$ to denote the marginal density of $Z$ under the prior $\pi$. Let $P_\pi$ denote the distribution of $Z$ corresponding to $f_\pi(z)$, i.e., $P_\pi(A) = \int 1_{z \in A} f_\pi(z) dz$, where $1_{z \in A}$ is the indicator function. For a function $g$, we write $E_\pi(g(Z))$ for the expectation under $f_\pi$. More specifically, $f_\pi(z) = \int f_\theta(z) \pi(\theta) d\theta$ and $E_\pi(g(Z)) = \int g(z) f_\pi(z) dz$. The $L_1$ distance between two probability distributions with densities $f_0$ and $f_1$ is given by $L_1(f_1, f_0) = \int |f_1(z) - f_0(z)| dz$. The following theorem establishes the minimax lower bounds for the estimation error and the expected length of confidence intervals for the $\ell_q$ loss, under the constraint that $\hat{\beta}$ is a good estimator at at least one interior point.

**Theorem 8.** Suppose $0 < \alpha, \alpha_0 < \frac{1}{4}$, $1 \leq q \leq 2$, $\Sigma_0$ is positive definite, $\theta_0 = (\beta^*, \Sigma_0, \sigma_0) \in \Theta$, and $F \subset \Theta$. Define $d = \min_{\theta \in F} \| \beta(\theta) - \beta^* \|_q$. Let $\pi$ denote a prior over the parameter space $F$. If an estimator $\hat{\beta}$ satisfies

\[
\mathbb{P}_{\theta_0} \left( \| \hat{\beta} - \beta^* \|_q^2 \leq \frac{1}{16} d^2 \right) \geq 1 - \alpha_0,
\]

then

\[
\inf_{\hat{L}_q} \sup_{\theta \in \{\theta_0\} \cup F} \mathbb{P}_\theta \left( \left| \hat{L}_q - \| \hat{\beta} - \beta \|_q \right| \geq \frac{1}{4} d^2 \right) \geq \bar{c}_1,
\]
and
\[ L^*_\alpha \left( \{ \theta_0 \} \ , \Theta, \tilde{\beta}, \ell_q \right) = \inf_{\text{CI}_\alpha \left( \tilde{\beta}, \ell_q, Z \right) \in L_\alpha \left( \Theta, \tilde{\beta}, \ell_q \right)} E_{\theta_0} L \left( \text{CI}_\alpha \left( \tilde{\beta}, \ell_q, Z \right) \right) \geq c_2^* d^2, \]

where \( c_1 = \min \left\{ \frac{1}{10}, \left( \frac{9}{10} - \alpha_0 - L_1 \left( f_\pi, f_{\theta_0} \right) \right)_+ \right\} \) and \( c_2^* = \frac{1}{2} \left( 1 - 2\alpha - \alpha_0 - 2L_1 \left( f_\pi, f_{\theta_0} \right) \right)_+ \).

**Remark 3.** The minimax lower bound (6.2) for the estimation error and (6.3) for the expected length of confidence intervals hold as long as the estimator \( \tilde{\beta} \) estimates \( \beta \) well at an interior point \( \theta_0 \). Besides Condition (6.1), another key ingredient for the lower bounds (6.2) and (6.3) is to construct the least favorable space \( F \) with the prior \( \pi \) such that the marginal distributions \( f_\pi \) and \( f_{\theta_0} \) are non-distinguishable. For the estimation lower bound (6.2), constraining that \( \| \tilde{\beta} - \beta^* \|_q^2 \) can be well estimated at \( \theta_0 \), due to the non-distinguishability between \( f_\pi \) and \( f_{\theta_0} \), we can establish that the loss \( \| \tilde{\beta} - \beta \|_q^2 \) cannot be estimated well over \( F \). For the lower bound (6.3), by Condition (6.1) and the non-distinguishability between \( f_\pi \) and \( f_{\theta_0} \), we will show that \( \| \tilde{\beta} - \beta \|_q^2 \) over \( F \) is much larger than \( \| \tilde{\beta} - \beta^* \|_q^2 \) and hence the honest confidence intervals must be sufficiently long.

Theorem 8 is used to establish the minimax lower bounds for both the estimation error and the expected length of confidence intervals of the \( \ell_q \) loss over \( \Theta(k) \). By taking \( \theta_0 \in \Theta(k_0) \) and \( \Theta = \Theta(k) \), Theorem 2 follows from (6.2) with a properly constructed subset \( F \subset \Theta(k) \). By taking \( \theta_0 \in \Theta(k_0) \) and \( \Theta = \Theta(k_2) \), Theorem 7 follows from (6.3) with a properly constructed \( F \subset \Theta(k_2) \). In both cases, Assumption (A1) implies Condition (6.1).

Several minimax lower bounds over \( \Theta_0(k) \) can also be implied by Theorem 8. For the estimation error, the minimax lower bounds (2.8) and (2.9) over the regime \( k \gtrsim \frac{\sqrt{n}}{\log p} \) in Theorem 1 follow from (6.2). For the expected length of confidence intervals, the minimax lower bounds (3.7) in Theorem 4 and (3.18) in the regions \( k_1 \leq k_2 \gtrsim \frac{\sqrt{n}}{\log p} \) and \( k_1 \leq \frac{\sqrt{n}}{\log p} \lesssim k_2 \lesssim \frac{n}{\log p} \) in Theorem 6 follow from (6.3). In these cases, Assumption (A1) or (A2) can guarantee that Condition (6.1) is satisfied. However, the minimax lower bound for estimation error (2.9) in the region \( \frac{\sqrt{n}}{\log p} \leq k \lesssim \frac{n}{\log p} \) and for the expected length of confidence intervals (3.18) in the region \( \frac{\sqrt{n}}{\log p} \lesssim k_1 \leq k_2 \lesssim \frac{n}{\log p} \) cannot be established using the above theorem. The following theorem, which requires testing a composite null against a composite alternative, establishes the refined minimax lower bounds over \( \Theta_0(k) \).
**THEOREM 9.** Let $0 < \alpha, \alpha_0 < \frac{1}{4}$, $1 \leq q \leq 2$, and $\theta_0 = (\beta^*, \Sigma_0, \sigma_0)$ where $\Sigma_0$ is a positive definite matrix. Let $k_1$ and $k_2$ be two sparsity levels. Assume that for $i = 1, 2$ there exist parameter spaces $\mathcal{F}_i \subset \{ (\beta, \Sigma_0, \sigma_0) : ||\beta||_0 \leq k_i \}$ such that for given $i$

\[
\| \Sigma_0 (\beta(\theta) - \beta^*) \|_2 = \text{dist}_i \quad \text{and} \quad \| \beta(\theta) - \beta^* \|_q = d_i, \quad \text{for all } \theta \in \mathcal{F}_i.
\]

Let $\pi_i$ denote a prior over the parameter space $\mathcal{F}_i$ for $i = 1, 2$. Suppose that for $\theta_1 = (\beta^*, \Sigma_0, \sigma_0^2 + \text{dist}_1^2)$ and $\theta_2 = (\beta^*, \Sigma_0, \sigma_0^2 + \text{dist}_2^2)$, there exist constants $c_1, c_2 > 0$ such that

\[
\mathbb{P}_{\pi_i} \left( \| \hat{\beta} - \beta^* \|_q^2 \leq c_i^2 d_i^2 \right) \geq 1 - \alpha_0, \quad \text{for } i = 1, 2.
\]

Then we have

\[
\inf \sup_{\mathcal{F}_1 \cup \mathcal{F}_2} \mathbb{P}_\theta \left( \| \hat{\beta} - \beta^* \|_q^2 \geq c_3^2 d_3^2 \right) \geq \tilde{c}_3,
\]

and

\[
\mathbf{L}_q^* \left( \Theta_0(k_1), \Theta_0(k_2), \hat{\beta}, \ell_q \right) \geq c_4^2 \left( (1 - c_2)^2 d_2^2 - (1 + c_1)^2 d_1^2 \right)_+,
\]

where $c_3 = \min \left\{ \frac{1}{4}, \left( 1 - c_2 \right)^2 - \frac{1}{4} - \left( 1 + c_1 \right)^2 \frac{d_2^2}{d_1^2} \right\}_+$, $c_4 = (1 - 2\alpha_0 - 2\alpha - \sum_{i=1}^2 L_1(f_{\pi_1}, f_{\theta_i}) - 2L_1(f_{\pi_2}, f_{\pi_1}))_+$ and $\tilde{c}_3 = \min \left\{ \frac{1}{10}, \left( \frac{1}{10} - 2\alpha_0 - \sum_{i=1}^2 L_1(f_{\pi_1}, f_{\theta_i}) - 2L_1(f_{\pi_2}, f_{\pi_1}) \right)_+ \right\}.$

**REMARK 4.** As long as the estimator $\hat{\beta}$ performs well at two points, $\theta_1$ and $\theta_2$, the minimax lower bounds (6.5) for the estimation error and (6.6) for the expected length of confidence intervals hold. Note that $\theta_i$ in the above theorem does not belong to the parameter space $\{ (\beta, \Sigma_0, \sigma_0) : ||\beta||_0 \leq k_i \}$, for $i = 1, 2$. In contrast to Theorem 8, Theorem 9 compares composite hypotheses $\mathcal{F}_1$ and $\mathcal{F}_2$, which will lead to a sharper lower bound than comparing the simple null $\{ \theta_0 \}$ with the composite alternative $\mathcal{F}$. For simplicity, we construct least favorable parameter spaces $\mathcal{F}_i$ such that the points in $\mathcal{F}_i$ is of fixed $\ell_2$ distance and fixed $\ell_q$ distance to $\beta^*$, for $i = 1, 2$, respectively. More importantly, we construct $\mathcal{F}_1$ with the prior $\pi_1$ and $\mathcal{F}_2$ with the prior $\pi_2$ such that $f_{\pi_1}$ and $f_{\pi_2}$ are not distinguishable, where $\theta_1$ and $\theta_2$ are introduced to facilitate the comparison. By Condition (6.4) and the construction of $\mathcal{F}_1$ and $\mathcal{F}_2$, we establish that the $\ell_q$ loss cannot be simultaneously estimated well over $\mathcal{F}_1$ and $\mathcal{F}_2$. For the lower bound (6.6), under the same conditions, it is shown that the $\ell_q$ loss over $\mathcal{F}_1$ and $\mathcal{F}_2$ are far apart and any confidence interval with guaranteed coverage probability over $\mathcal{F}_1 \cup \mathcal{F}_2$ must be
sufficiently long. Due to the prior information $\Sigma = I$ and $\sigma = \sigma_0$, the lower bound construction over $\Theta_0(k)$ is more involved than that over $\Theta(k)$. We shall stress that the construction of $F_1$ and $F_2$ and the comparison between composite hypotheses are of independent interest.

The minimax lower bound (2.9) in the region $\frac{\sqrt{n}}{\log p} \lesssim k \lesssim \frac{n}{\log p}$ follows from (6.5) and the minimax lower bound (3.18) in the region $\frac{\sqrt{n}}{\log p} \lesssim k_1 \leq k_2 \lesssim \frac{n}{\log p}$ for the expected length of confidence intervals follows from (6.6). In these cases, $\Sigma_0$ is taken as I and Assumption (A2) implies Condition (6.4).

7. An intermediate setting with known $\sigma = \sigma_0$ and unknown $\Sigma$.

The results given in Sections 3 and 4 show the significant difference between $\Theta_0(k)$ and $\Theta(k)$ in terms of minimaxity and adaptivity of confidence intervals for $\|\hat{\beta} - \beta\|_q^2$. $\Theta_0(k)$ is for the simple setting with known design covariance matrix $\Sigma = I$ and known noise level $\sigma = \sigma_0$, and the $\Theta(k)$ is for unknown $\Sigma$ and $\sigma$. In this section, we further consider minimaxity and adaptivity of confidence intervals for $\|\hat{\beta} - \beta\|_q^2$ in an intermediate setting where the noise level $\sigma = \sigma_0$ is known and $\Sigma$ is unknown but of certain structure. Specifically, we consider the following parameter space,

$$\Theta_{\sigma_0}(k, s) = \left\{ (\beta, \Sigma, \sigma_0) : \|\beta\|_0 \leq k, \frac{1}{M_1} \leq \lambda_{\min} (\Sigma) \leq \lambda_{\max} (\Sigma) \leq M_1, \|\Sigma^{-1}\|_{L_1} \leq M, \max_{1 \leq i \leq p} \|(\Sigma^{-1})_i\|_0 \leq s \right\},$$

for some constants $M_1 \geq 1$ and $M > 0$. $\Theta_{\sigma_0}(k, s)$ basically assumes known noise level $\sigma$ and imposes sparsity conditions on the precision matrix of the random design. This parameter space is similar to those used in the literature of sparse linear regression with random design [29, 13, 14]. $\Theta_{\sigma_0}(k, s)$ has two sparsity parameters where $k$ represents the sparsity of $\beta$ and $s$ represents the maximum row sparsity of the precision matrix $\Sigma^{-1}$. Note that $\Theta_0(k) \subset \Theta_{\sigma_0}(k, s) \subset \Theta(k)$ and $\Theta_0(k)$ is a special case of $\Theta_{\sigma_0}(k, s)$ with $M_1 = 1$.

Under the assumption $s \ll \sqrt{n/\log p}$, the minimaxity and adaptivity lower bounds for the expected length of confidence intervals for $\|\hat{\beta} - \beta\|_q^2$ with $1 \leq q < 2$ over $\Theta_{\sigma_0}(k, s)$ are the same as those over $\Theta_0(k)$. That is, Theorems 5 and 6 hold with $\Theta_0(k_1), \Theta_0(k_2),$ and $\Theta_0(k)$ replaced by $\Theta_{\sigma_0}(k_1, s), \Theta_{\sigma_0}(k_2, s),$ and $\Theta_{\sigma_0}(k, s),$ respectively. For the case $q = 2$, the following theorem establishes the minimaxity and adaptivity lower bounds for the expected length of confidence intervals for $\|\hat{\beta} - \beta\|_2^2$ over $\Theta_{\sigma_0}(k, s).$
Theorem 10. Suppose $0 < \alpha, \alpha_0 < 1/4$, $M_1 > 1$, $s \ll \sqrt{n/\log p}$ and the sparsity levels $k_1, k_2$ and $k_0$ satisfy Assumption (B2) with the constant $c_0$ replaced by $c_0^*$ defined in (9.14). For any estimator $\hat{\beta}$ satisfying

$$\sup_{\theta \in \Theta(k_0)} \mathbb{P}_\theta \left( \| \hat{\beta} - \beta^* \|_q^2 \geq C^* \| \beta^* \|_0^{\frac{2}{n}} \sigma_0^2 \frac{\log p}{n} \right) \leq \alpha_0,$$

with a constant $C^* > 0$, then there is some constant $c > 0$ such that

$$L_\alpha^* \left( \Theta_{\sigma_0}(k_1, s), \Theta_{\sigma_0}(k_2, s), \hat{\beta}, \ell_2 \right) \geq \frac{c}{n} \left\{ k_2 \frac{\log p}{n}, \max \left\{ k_1 \frac{\log p}{n}, \frac{1}{\sqrt{n}} \right\} \right\} \sigma_0^2$$

and

$$L_\alpha^* \left( \Theta_{\sigma_0}(k_1, s), \Theta_{\sigma_0}(k_2, s), \hat{\beta}, \ell_2 \right) \geq \frac{c}{n} \left\{ k_1 \frac{\log p}{n}, \sigma_0^2 \right\} \text{ and } i = 1, 2.$$

In particular, if $p \geq n$ and $\hat{\beta}$ is constructed based on the subsample $Z^{(1)} = (y^{(1)}, X^{(1)})$ and satisfies Assumption (A) with $\delta > 2$, the above lower bounds can be attained.

In contrast to Theorems 3 and 4, the lower bounds for the case $q = 2$ change in the absence of the prior knowledge $\Sigma = I$ but the possibility of adaptivity of confidence intervals over $\Theta_{\sigma_0}(k, s)$ is similar to that over $\Theta_0(k)$. Since the Lasso estimator $\beta^L$ defined in (2.10) with $A > 4\sqrt{2}$ satisfies Assumption (A) with $\delta > 2$, by Theorem 10, the minimax lower bounds (7.3) and (7.4) can be attained for $\beta^L$. For $\beta^L$, only when $\frac{\sqrt{n}}{\log p} \lesssim k_1 \leq k_2 \lesssim \frac{n}{\log p}$, $L_\alpha^* \left( \Theta_{\sigma_0}(k_1, s), \beta^L, \ell_2 \right) \asymp L_\alpha^* \left( \Theta_{\sigma_0}(k_1, s), \Theta_{\sigma_0}(k_2, s), \hat{\beta}, \ell_2 \right) \asymp \frac{k_1 \log p}{n}$ and adaptation between $\Theta_{\sigma_0}(k_1, s)$ and $\Theta_{\sigma_0}(k_2, s)$ is possible. In other regimes, if $k_1 \ll k_2$, then $L_\alpha^* \left( \Theta_{\sigma_0}(k_1, s), \beta^L, \ell_2 \right) \asymp L_\alpha^* \left( \Theta_{\sigma_0}(k_1, s), \Theta_{\sigma_0}(k_2, s), \beta, \ell_2 \right)$ and adaptation between $\Theta_{\sigma_0}(k_1, s)$ and $\Theta_{\sigma_0}(k_2, s)$ is impossible. For reasons of space, more discussion on $\Theta_{\sigma_0}(k, s)$, including the construction of adaptive confidence intervals over the regime $\frac{\sqrt{n}}{\log p} \lesssim k_1 \leq k_2 \lesssim \frac{n}{\log p}$, is postponed to the supplement [6].

8. Minimax lower bounds for estimating $\| \beta \|_q^2$ with $1 \leq q \leq 2$. The lower bounds developed in this paper have broader implications. In particular, the established results imply the minimax lower bounds for estimating $\| \beta \|_q^2$ and the expected length of confidence intervals for $\| \beta \|_q^2$ with $1 \leq q \leq 2$. To build the connection, it is sufficient to note that the trivial estimator $\hat{\beta} = 0$ satisfies Assumptions (A1) and (A2) with $\beta^* = 0$. 
Then we can apply the lower bounds \((2.8), (2.9)\) and \((2.13)\) to the estimator \(\hat{\beta} = 0\) and establish the minimax lower bounds of estimating \(\|\beta\|_2^q\),

\[
\inf_{L_2} \sup_{\theta \in \Theta_0(k)} \mathbb{P}_\theta \left( |\hat{L}_2 - \|\beta\|_2^2| \geq c \min \left\{ k \frac{\log p}{n}, \frac{1}{\sqrt{n}} \sigma_0^2 \right\} \right) \geq \delta;
\]

\[
\inf_{L_q} \sup_{\theta \in \Theta_0(k)} \mathbb{P}_\theta \left( |\hat{L}_q - \|\beta\|_q^2| \geq c k^{\frac{q}{2}} \frac{\log p}{n} \sigma_0^2 \right) \geq \delta, \quad \text{for } 1 \leq q < 2,
\]

\[
\inf_{L_q} \sup_{\theta \in \Theta_0(k)} \mathbb{P}_\theta \left( |\hat{L}_q - \|\beta\|_q^2| \geq c k^{\frac{q}{2}} \frac{\log p}{n} \right) \geq \delta, \quad \text{for } 1 \leq q \leq 2,
\]

for some constants \(\delta > 0\) and \(c > 0\). Similarly, all the lower bounds for the expected length of confidence intervals for \(\|\beta\|_2^q\) established in Theorem 3 to Theorem 7 imply corresponding lower bounds for \(\|\beta\|_2^q\). The lower bound \(\min \{ k \frac{\log p}{n}, \frac{1}{\sqrt{n}} \sigma_0^2 \} \) in \((8.1)\) is the same as the detection boundary in the sparse linear regression for the case \(\Sigma = I\) and \(\sigma = 1\); See [19] and [1] for more details. Estimation of \(\|\beta\|_2^2\) in high-dimensional linear regression has been considered in [17] under the general setting where \(\Sigma\) and \(\sigma\) are unknown and the lower bound \((8.3)\) with \(q = 2\) leads to one key component of the lower bound \(ck \frac{\log p}{n}\) for estimating \(\|\beta\|_2^2\).

9. Proofs. In this section, we present the proofs of the lower bound results. In Section 9.1, we establish the general lower bound result, Theorem 8. By applying Theorem 8 and Theorem 9, we establish Theorems 4 and 6 in Section 9.2. For reasons of space, the proofs of Theorems 1, 2, 3, 5, 7, 9, 10, the upper bound results, including Propositions 1, 2, 3, 4, 5, 6 and the proofs of technical lemmas are postponed to the supplement [6].

We define the \(\chi^2\) distance between two density functions \(f_1\) and \(f_0\) by

\[
\chi^2(f_1, f_0) = \int \frac{(f_1(z) - f_0(z))^2}{f_0(z)} dz = \int f_1^2(z) f_0(z) dz - 1,
\]

and it is well known that

\[
L_1(f_1, f_0) \leq \sqrt{\chi^2(f_1, f_0)}.
\]

Let \(P_{Z, \theta \sim \pi}\) denote the joint probability of \(Z\) and \(\theta\) with the joint density function \(f(\theta, z) = f_\theta(z) \pi(\theta)\). We introduce the following lemma, which is used in the proofs of Theorem 8 and Theorem 9. The proof of this lemma can be found in the supplement [6].

\textbf{Lemma 1. For any event } \mathcal{A}, \text{ we have}

\[
P_\pi(Z \in \mathcal{A}) = P_{Z, \theta \sim \pi}(Z \in \mathcal{A}),
\]

\[
\left| P_{\pi_2}(Z \in \mathcal{A}) - P_{\pi_1}(Z \in \mathcal{A}) \right| \leq L_1(f_{\pi_2}, f_{\pi_1}).
\]
We will write $\mathbb{P}_\pi(A)$ and $\mathbb{P}_{Z,\theta \sim \pi}(A)$ for $\mathbb{P}_\pi(Z \in A)$ and $\mathbb{P}_{Z,\theta \sim \pi}(Z \in A)$ respectively. Recall that $\tilde{L}_q(Z)$ denotes a data-dependent loss estimator and $\beta(\theta)$ denotes the corresponding $\beta$ of the parameter $\theta$.

9.1. Proof of Theorem 8. We set $c_0 = \frac{1}{4}$ and $\alpha_1 = \frac{1}{10}$.

Proof of (6.2)

We assume

$$P_{\theta_0}\left(\left| \tilde{L}_q(Z) - \|\hat{\beta}(Z) - \beta^*\|_q^2 \right| \leq \frac{1}{4}d^2 \right) \geq 1 - \alpha_1. \tag{9.4}$$

Otherwise, we have

$$P_{\theta_0}\left(\left| \tilde{L}_q(Z) - \|\hat{\beta}(Z) - \beta^*\|_q^2 \right| \geq \frac{1}{4}d^2 \right) \geq \alpha_1. \tag{9.5}$$

and hence (6.2) follows. Define the event

$$A_0 = \left\{ z : \|\hat{\beta}(z) - \beta^*\|_q^2 \leq c_0^2d^2, \ |\tilde{L}_q(z) - \|\hat{\beta}(z) - \beta^*\|_q^2| \leq \frac{1}{4}d^2 \right\}. \tag{9.6}$$

By (6.1) and (9.4), we have $P_{\theta_0}(A_0) \geq 1 - \alpha_0 - \alpha_1$. By (9.3), we obtain

$$P_{\pi}(A_0) \geq 1 - \alpha_0 - \alpha_1 - \int |f_{\theta_0}(z) - f_{\pi}(z)| \, dz. \tag{9.7}$$

For $z \in A_0$ and $\theta \in \mathcal{F}$, by triangle inequality,

$$\|\hat{\beta}(z) - \beta(\theta)\|_q \geq \|\beta(\theta) - \beta^*\|_q - \|\hat{\beta}(z) - \beta^*\|_q \geq (1 - c_0) \, d. \tag{9.8}$$

For $z \in A_0$ and $\theta \in \mathcal{F}$, then $|\tilde{L}_q(z) - \|\hat{\beta}(z) - \beta(\theta)\|_q^2| \geq \|\hat{\beta}(z) - \beta(\theta)\|_q^2 - \|\hat{\beta}(z) - \beta^*\|_q^2 - |\tilde{L}_q(z) - \|\hat{\beta}(z) - \beta^*\|_q^2| \geq (1 - 2c_0 - \frac{1}{4})d^2$, where the first inequality follows from triangle inequality and the last inequality follows from (9.6) and (9.8).

Hence, for $z \in A_0$, we obtain

$$\inf_{\theta \in \mathcal{F}} |\tilde{L}_q(z) - \|\hat{\beta}(z) - \beta(\theta)\|_q^2| \geq (1 - 2c_0 - \frac{1}{4})d^2. \tag{9.9}$$

Note that $\sup_{\theta \in \mathcal{F}} P_{\theta}\left(\left| \tilde{L}_q(Z) - \|\hat{\beta}(Z) - \beta(\theta)\|_q^2 \right| \geq (1 - 2c_0 - \frac{1}{4})d^2 \right) \geq \sup_{\theta \in \mathcal{F}} P_{\theta}\left(\left| \inf_{\theta \in \mathcal{F}} \tilde{L}_q(Z) - \|\hat{\beta}(Z) - \beta(\theta)\|_q^2 \right| \geq (1 - 2c_0 - \frac{1}{4})d^2 \right)$. Since the max risk is lower bounded by the Bayesian risk, we can further lower bound
the last term by \( P \left( \inf_{\theta \in \mathcal{F}} \left| \widehat{L}_q(Z) - \| \beta \|_q^2 \right| \right) \geq (1 - 2c_0 - \frac{1}{4})d^2 \).

Combined with (9.9), we establish

\[
(9.10) \quad \sup_{\theta \in \mathcal{F}} P_{\theta} \left( \left| \widehat{L}_q(Z) - \| \beta \|_q^2 \right| \geq (1 - 2c_0 - \frac{1}{4})d^2 \right) \geq P_\pi (\mathcal{A}_0).
\]

Combining (9.5), (9.7) and (9.10), we establish (6.2).

**Proof of (6.3)**

For \( \text{CI}_\alpha \left( \hat{\beta}, \ell, Z \right) \in \mathcal{I}_\alpha \left( \Theta, \hat{\beta}, \ell, \tau \right) \), we have

\[
(9.11) \quad \inf_{\theta \in \Theta} P_{\theta} \left( \| \beta \|_q^2 \in \text{CI}_\alpha \left( \hat{\beta}, \ell, Z \right) \right) \geq 1 - \alpha.
\]

Define the event \( \mathcal{A} = \left\{ z : \| \beta(z) - \beta^* \|_q < c_0d, \| \beta(z) - \beta^* \|_q^2 \in \text{CI}_\alpha \left( \hat{\beta}, L, z \right) \right\} \).

By (6.1) and (9.11), we have \( P_{\theta_0} (\mathcal{A}) \geq 1 - \alpha - \alpha_0 \). (9.2) and (9.3) imply

\[
(9.12) \quad P_{Z, \theta \sim \pi} (\mathcal{A}) = P_\pi (\mathcal{A}) \geq 1 - \alpha - \alpha_0 - L_1 (f_\pi, f_{\theta_0}).
\]

Define the event \( \mathcal{B}_\theta = \left\{ z : \| \beta(z) - \beta (\theta) \|_q^2 \in \text{CI}_\alpha \left( \hat{\beta}, \ell, z \right) \right\} \) and \( \mathcal{M} = \cup_{\theta \in \mathcal{F}} \mathcal{B}_\theta \). By (9.11), we have

\[
P_{Z, \theta \sim \pi} (\mathcal{M}) = \int \left( \int \left( \int 1_{\mathcal{M}} f_\theta(z) dz \right) \right) \pi (\theta) d\theta \geq \int \left( \int \left( \int 1_{\mathcal{B}_\theta} f_\theta(z) dz \right) \right) \pi (\theta) d\theta \geq 1 - \alpha.
\]

Combined with (9.12), we have \( P_{Z, \theta \sim \pi} (\mathcal{A} \cap \mathcal{M}) \geq 1 - 2\alpha - \alpha_0 - L_1 (f_\pi, f_{\theta_0}). \)

For \( z \in \mathcal{M} \), there exists \( \hat{\theta} \in \mathcal{F} \) such that \( \| \beta(z) - \beta(\hat{\theta}) \|_q^2 \in \text{CI}_\alpha \left( \hat{\beta}, \ell, z \right) \);

For \( z \in \mathcal{A} \), we have \( \| \beta(z) - \beta^* \|_q^2 \in \text{CI}_\alpha \left( \hat{\beta}, \ell, z \right) \) and \( \| \beta(z) - \beta^* \|_q < c_0d \).

Hence, for \( z \in \mathcal{A} \cap \mathcal{M} \), we have\( \| \beta(z) - \beta(\hat{\theta}) \|_q^2, \| \beta(z) - \beta^* \|_q^2 \in \text{CI}_\alpha \left( \hat{\beta}, \ell, z \right) \)

and \( \| \beta(z) - \beta(\hat{\theta}) \|_q \geq \| \beta(\hat{\theta}) - \beta^* \|_q - \| \beta(z) - \beta^* \|_q \geq (1 - c_0) d \) and hence

\[
L \left( \text{CI}_\alpha \left( \hat{\beta}, \ell, z \right) \right) \geq (1 - 2c_0) d^2.
\]

Define the event \( \mathcal{C} = \left\{ z : \mathbf{L} \left( \text{CI}_\alpha \left( \hat{\beta}, \ell, z \right) \right) \geq (1 - 2c_0) d^2 \right\} \). By (9.13), we have \( P_\pi (\mathcal{C}) = P_{Z, \theta \sim \pi} (\mathcal{C}) \geq P_{Z, \theta \sim \pi} (\mathcal{A} \cap \mathcal{M}) \geq 1 - 2\alpha - \alpha_0 - L_1 (f_\pi, f_{\theta_0}) \). By (9.3), we establish \( P_{\theta_0} (\mathcal{C}) \geq 1 - 2\alpha - \alpha_0 - 2L_1 (f_\pi, f_{\theta_0}) \) and hence (6.3).

**9.2. Proof of Theorems 4 and 6.** We first specify some constants used in the proof. Let \( C^* \) be given in (2.6). Define \( \epsilon_1 = \frac{1 - 2\alpha - 2\alpha_0}{12} \) and

\[
(9.14) \quad c_0 = \min \left\{ \frac{1}{2}, 32 \log (1 + \epsilon_1^2), \frac{2}{3} \sqrt{\log(1 + \epsilon_1)}, \frac{1 - 2\gamma}{16C^*}, \left( \frac{1 - 2\gamma}{16C^*} \right)^2 \right\}, \quad \gamma = \min \left\{ c_0, \frac{\sqrt{M_1} - 1}{C^* M_1 + \sqrt{M_1} - 1} \right\}.
\]

Theorems 4 and 6 follow from Theorem 11 below.
Theorem 11. Suppose $0 < \alpha < \frac{1}{4}$, $1 \leq q \leq 2$ and the sparsity levels $k_1, k_2$ and $k_0$ satisfy Assumption (B2). Suppose that $\hat{\beta}$ satisfies Assumption (A2) with $\|\beta^*\|_0 \leq k_0$.

1. If $k_2 \lesssim \sqrt{\frac{n}{\log p}}$, then there is some constant $c > 0$ such that

$$
(9.15) \quad L^*_\alpha \left( \Theta_0 (k_1), \Theta_0 (k_2), \hat{\beta}, \ell_q \right) \geq c k_2^{\frac{3}{2} - 1} \frac{\log p}{n} + k_2^{\frac{3}{2} - 1} \frac{1}{\sqrt{n}} \sigma_0^2.
$$

2. If $\sqrt{\frac{n}{\log p}} \lesssim k_2 \lesssim \frac{n}{\log p}$, then there is some constant $c > 0$ such that

$$
(9.16) \quad L^*_\alpha \left( \Theta_0 (k_1), \Theta_0 (k_2), \hat{\beta}, \ell_q \right) \geq c \max \left\{ \left( (1 - c_2)^2 k_2^{\frac{3}{2} - 1} k_1 \frac{\log p}{n} - (1 + c_1)^2 k_1^{\frac{3}{2} - 1} \frac{\log p}{n} \right) + \frac{k_2^{\frac{3}{2} - 1}}{\sqrt{n}} \right\} \sigma_0^2,
$$

where $c_1 = \frac{C^{*} M_1 k_1^2}{(k_1 - k_0)^{\frac{1}{2}}}$ and $c_2 = \frac{C^{*} k_0^2}{M_1 (k_2 - k_0)^{\frac{1}{2}} (k_1 - k_0)^{\frac{1}{2}}}$.

In particular, the minimax lower bound (9.15) and the term $\frac{k_2^{\frac{3}{2} - 1}}{\sqrt{n}} \sigma_0^2$ in (9.16) can be established under the weaker assumption (A1) with $\|\beta^*\|_0 \leq k_0$.

By Theorem 11, we establish (3.7) in Theorem 4 and (3.18) in Theorem 6. In the regime $k_2 \lesssim \sqrt{\frac{n}{\log p}}$, the lower bound (3.7) for $q = 2$ and (3.18) for $1 \leq q < 2$ follow from (9.15). For the case $q = 2$, in the regime $\sqrt{\frac{n}{\log p}} \lesssim k_2 \lesssim \frac{n}{\log p}$, the first term of the right hand side of (9.16) is 0 while the second term is $\frac{k_2^{\frac{3}{2} - 1}}{\sqrt{n}}$, which leads to (3.7). For $1 \leq q < 2$, let $k_1^* = \min \{k_1, \zeta_0 k_2 \}$ for some constant $0 < \zeta_0 < 1$, an application of (9.16) leads to $L^*_\alpha \left( \Theta_0 (k_1^*), \Theta_0 (k_2), \hat{\beta}, \ell_q \right) \geq c \max \left\{ \frac{k_2^{\frac{3}{2} - 1} k_1^{\frac{3}{2} - 1} \log p}{n}, \frac{k_2^{\frac{3}{2} - 1}}{\sqrt{n}} \right\} \sigma_0^2$. By this result, if $k_1 \leq \zeta_0 k_2$, the lower bounds (3.18) in the regions $k_1 \lesssim \sqrt{\frac{n}{\log p}} \leq k_2 \lesssim \frac{n}{\log p}$ and $\sqrt{\frac{n}{\log p}} \lesssim k_1 \leq k_2 \lesssim \frac{n}{\log p}$ follow; if $\zeta_0 k_2 < k_1 \leq k_2$, by the fact that $L^*_\alpha \left( \Theta_0 (k_1), \Theta_0 (k_2), \hat{\beta}, \ell_q \right) \geq L^*_\alpha \left( \Theta_0 (k_1^*), \Theta_0 (k_2), \hat{\beta}, \ell_q \right)$ and $k_1^* = \zeta_0 k_2 \geq \zeta_0 k_1$, the lower bounds (3.18) over the regions $k_1 \lesssim \sqrt{\frac{n}{\log p}} \lesssim k_2 \lesssim \frac{n}{\log p}$ and $\sqrt{\frac{n}{\log p}} \lesssim k_1 \leq k_2 \lesssim \frac{n}{\log p}$ follow. The following lemma shows that (3.7) holds for $\hat{\beta}^L$ defined in (2.10) with $A > \sqrt{2}$ by verifying Assumption (A1) and (3.18) holds for $\hat{\beta}^L$ defined in (2.10) with $A > 4\sqrt{2}$ by verifying Assumption (A2). Its proof can be found in the supplement [6].
Lemma 2. If $A > 4\sqrt{2}$, then we have

$$\inf_{\theta=(\beta^*,\lambda,\sigma)\geq 2\sigma_0} \mathbb{P}_\theta \left( \left\| \hat{\beta}_L - \beta^* \right\|_q^2 \leq C\left\| \beta^* \right\|_0^2 \frac{\log p}{n \sigma^2} \right) \geq 1 - c \exp \left( -c' n \right) - p^{-c}.$$  

In particular, the above result holds for $q = 2$ under the assumption $A > \sqrt{2}$.

References.


SUPPLEMENT TO “ACCURACY ASSESSMENT FOR HIGH-DIMENSIONAL LINEAR REGRESSION” †

University of Pennsylvania

BY T. TONY CAI, AND ZIHAN GUO

This note summarizes the supplementary materials to the paper “Accuracy Assessment for High-dimensional Linear Regression” [3]. We first discuss the differences between the two parameter spaces $\Theta(k)$ and $\Theta_0(k)$ and then present the minimaxity and adaptivity lower bounds of confidence intervals over the parameter space $\Theta_{\sigma_0}(k, s)$. We present the proofs of extra lower bound results, Theorem 1, Theorem 2, Theorem 3, Theorem 5, Theorem 7, Theorem 9, Theorem 10, Theorem 11 and Theorem 12. We also present the proofs of upper bound results, including Proposition 1, 2, 3, 4, 5, 6 and 7 and the proofs of technical lemmas.

1. Difference between $\Theta(k)$ and $\Theta_0(k)$. In the main paper [3], we have investigated the minimax estimation rate, minimax expected length and adaptivity of confidence intervals for the loss $\|\hat{\beta} - \beta\|_q^2$ with $1 \leq q \leq 2$ over the parameter spaces $\Theta_0(k)$ and $\Theta(k)$. It is interesting to compare the minimaxity and adaptivity behaviors between loss estimation over the parameter spaces $\Theta(k)$ and $\Theta_0(k)$. The comparison shows significant differences between estimating the $\ell_2$ loss and the $\ell_q$ loss with $1 \leq q < 2$ as well as the differences between the two parameter spaces $\Theta(k)$ and $\Theta_0(k)$.

In terms of the minimax estimation rate and minimax expected length of confidence intervals, the prior information $\Sigma = I$ and $\sigma = \sigma_0$ reduces the convergence rate for the $\ell_2$ loss $\|\hat{\beta} - \beta\|_2^2$ from $\frac{k \log p}{n}$ to $\min \left\{ \frac{k \log p}{n}, \frac{1}{\sqrt{n}} \right\}$. With this prior information, the adaptive estimation of $\ell_2$ loss is made possible over the regime $\sqrt{n \log p} \ll k \lesssim \frac{n}{\log p}$. In contrast, even with such prior information, the minimax convergence rate remains unchanged for the case $1 \leq q < 2$.

Regarding adaptivity of confidence intervals, the prior knowledge $\Sigma = I$ and $\sigma = \sigma_0$ is extremely useful for the construction of adaptive confidence intervals for the $\ell_2$ loss in the regime $\frac{\sqrt{n}}{\log p} \ll k \lesssim \frac{n}{\log p}$. Though adaptivity is still impossible outside this regime, we have seen that, with this prior knowledge, the expected length of optimal confidence intervals over the regime

†The research was supported in part by NSF Grants DMS-1208982 and DMS-1403708, and NIH Grant R01 CA127334.
k_1 \lesssim \sqrt{n} \lesssim k_2 \lesssim \frac{n}{\log p} is reduced from \( k_2 \frac{\log p}{n} \) to \( \frac{1}{\sqrt{n}} \).

In contrast, for \( \ell_q \) loss with \( 1 \leq q < 2 \), even with this prior information, it is still impossible to construct adaptive confidence intervals for \( \| \hat{\beta} - \beta \|_2^q \) with \( 1 \leq q < 2 \). However, a comparison of Theorem 6 in the main paper [3] with Theorem 7 in the main paper [3] reveals that the expected length of confidence intervals is reduced with such prior knowledge, from \( k_2^2 \lambda_{\min}(\Sigma) \frac{\log p}{n} \) to \( k_2^2 \frac{\log p}{n} \) in the regime \( k_1 \lesssim \sqrt{n} \) and from \( k_2^2 \lambda_{\min}(\Sigma) \frac{\log p}{n} \) to \( k_2^2 \frac{\log p}{n} \) in the regime \( \sqrt{n} \frac{\log p}{n} \lesssim k_1 \leq k_2 \lesssim \frac{n}{\log p} \).

2. Minimaxity and adaptivity of confidence intervals for \( \| \hat{\beta} - \beta \|_q^2 \) over \( \Theta_{\sigma_0}(k, s) \). In the main paper [3], we have shown that there is significant difference between \( \Theta_{\sigma_0}(k) \) and \( \Theta(k) \) in terms of the minimax convergence rates and the adaptivity behaviors. As discussed in Section 7 in [3], the parameter space \( \Theta_{\sigma_0}(k) \) is relatively simple and in this section, we consider a more general parameter space for \((\beta, \Sigma, \sigma)\),

\[
\Theta_{\sigma_0}(k, s) = \left\{ (\beta, \Sigma, \sigma_0) : \begin{aligned}
\| \beta \|_0 &\leq k, \\
\frac{1}{M} &\leq \lambda_{\min}(\Sigma) \leq \lambda_{\max}(\Sigma) \leq M \\
\| \Sigma^{-1} \|_{L_1} &\leq M, \\
\max_{1 \leq i \leq p} \| (\Sigma^{-1})_i \|_0 &\leq s
\end{aligned} \right\},
\]

for some positive constant \( M_1 \geq 1 \) and \( M > 0 \). We will present the lower bound results over \( \Theta_{\sigma_0}(k, s) \) in Section 2.1 and the upper bound results in Section 2.2.

2.1. Minimax Lower Bound. In this section, we first establish the minimax lower bounds over the parameter \( \Theta_{\sigma_0}(k, s) \).

**Theorem 12.** Suppose \( 0 < \alpha, \alpha_0 < 1/4 \), \( s \ll \sqrt{n/\log p} \) and the sparsity levels \( k_1, k_2 \) and \( k_0 \) satisfy Assumption (B2) in the main paper [3] with the constant \( c_0 \) defined in (9.14) in the main paper [3]. For any estimator \( \hat{\beta} \) satisfying

\[
\sup_{\theta \in \Theta(k_0)} \mathbb{P}_\theta \left( \| \hat{\beta} - \beta^* \|_q^2 \geq C^* \| \beta^* \|_0^2 \frac{\log p}{n} \sigma^2 \right) \leq \alpha_0,
\]

with a constant \( C^* > 0 \).

1. If \( k_2 \lesssim \sqrt{n} \), then there is some constant \( c > 0 \) such that

\[
\mathcal{L}^*_\alpha \left( \Theta_{\sigma_0}(k_1, s), \Theta_{\sigma_0}(k_2, s), \hat{\beta}, \ell_q \right) \geq c k_2^2 \frac{\log p}{n} \sigma_0^2.
\]
2. If \( \frac{\sqrt{n}}{\log p} \leq k_2 \leq \frac{n}{\log p} \), then there is some constant \( c > 0 \) such that

\[
\mathbf{L}_\alpha^* \left( \Theta_{\sigma_0}(k_1, s), \Theta_{\sigma_0}(k_2, s), \hat{\beta}, \ell_2 \right) \geq c \max \left\{ \left(1 - c_2\right)^2 M_1^2 k_2^{\frac{2}{k_2} - 1} k_1 \log p \frac{1}{n} - (1 + c_1^2) \frac{1}{M_1^2 k_1^{\frac{2}{k_1} - 1}} \frac{\log p}{n} + , \frac{k_2^{\frac{2}{k_2} - 1}}{\sqrt{n}} \right\} \sigma_0^2,
\]

where \( c_1 = C^* M_1 k_1^\frac{1}{k} \) and \( c_2 = \frac{C^* k_0^\frac{1}{k} - k_0^\frac{1}{k} (k_1 - k_0)^\frac{1}{k}}{M_1 (k_2 - k_0)^\frac{1}{k} - k_0^\frac{1}{k} (k_1 - k_0)^\frac{1}{k}} \).

Consequently, we have

\[
(2.5) \quad \mathbf{L}_\alpha^* \left( \Theta_{\sigma_0}(k_1, s), \hat{\beta}, \ell_2 \right) \geq c \frac{k_0^\frac{2}{k} \log p \sigma_0^2}{n} \quad \text{for} \quad 1 \leq q \leq 2, \ i = 1, 2.
\]

In particular, for the case \( 1 \leq q < 2 \), if \( p \geq n \) and \( \hat{\beta} \) is constructed based on the subsample \( Z^{(1)} = (y^{(1)}, X^{(1)}) \) and satisfies Assumption (A) in the main paper [3] with \( \delta > 2 \) and the assumption \( \| (\hat{\beta} - \beta)_S \|_1 \leq c \| (\beta - \beta)_S \|_1 \) where \( S = \text{supp}(\beta) \), the above lower bounds for \( 1 \leq q < 2 \) can be attained; for the case \( q = 2 \), if \( p \geq n \) and \( \hat{\beta} \) is constructed based on the subsample \( Z^{(1)} = (y^{(1)}, X^{(1)}) \) and satisfies Assumption (A) in the main paper [3] with \( \delta > 2 \), the above lower bounds for \( q = 2 \) can be attained.

Remark 5. The minimax lower bound (2.5) with \( i = 1 \) follows from (2.3) and (2.4) by taking \( k_2 = k_1 \). The minimax lower bound (2.5) with \( i = 2 \) follows from (2.3) and (2.4) by taking \( k_1 = \frac{1}{2} k_2 \) and the fact that \( \mathbf{L}_\alpha^* \left( \Theta_{\sigma_0}(k_2, s), \hat{\beta}, \ell_2 \right) \geq \mathbf{L}_\alpha^* \left( \Theta_{\sigma_0}(k_1, s), \Theta_{\sigma_0}(k_2, s), \hat{\beta}, \ell_2 \right) \). Theorem 4 and 6 follow from the above theorem with plugging in \( M_1 = 1 \) and the corresponding \( q \). Theorem 10 is the special case of the above with \( q = 2 \).

2.2. Minimax upper bound. In this section, we will focus on the estimator \( \hat{\beta} \) constructed based on the subsample \( Z^{(1)} = (y^{(1)}, X^{(1)}) \) and satisfying Assumption (A) in the main paper [3] with \( \delta > 2 \) and demonstrate the lower bounds (2.4) and (2.3) in Theorem 12 can be achieved. And the Lasso estimator \( \hat{\beta}^L \) defined in (2.10) in the main paper [3] with \( A > 4 \sqrt{2} \) is an example of such estimators. To simply the notation, we use \( \Omega \) to denote \( \Sigma^{-1} \). Let \( \hat{\Omega} \) denote the CLIME estimator [6] of \( \Omega \) and \( \lambda_{\max}^2(\hat{\Omega}) \) and \( \lambda_{\min}^2(\hat{\Omega}) \) denote the maximum and minimum eigenvalue of the estimator \( \hat{\Omega} \).

The constructions of confidence intervals are very similar to the confidence intervals, \( \text{CI}_\alpha^2(Z) \) in (3.8), and \( \text{CI}_\alpha^2(Z, k_2, q) \) in (3.19) in the main paper [3].
The only difference here is that there is no prior knowledge $\Sigma = I$ and we need to estimate $\Omega = \Sigma^{-1}$ based on the data $X$. In the following, we will detail the modification of $\text{CI}_1^\alpha(Z)$ in (3.8) in the main paper [3] and $\text{CI}_2^\alpha(Z, k_2, q)$ in (3.19) in the main paper [3].

We modify the construction of $\text{CI}_1^\alpha(Z)$ proposed in (3.8) in the main paper [3] as follows,

$$\text{CI}_3^\alpha(Z) = \left(0.99\lambda_2^\text{min} \times \left(\frac{\psi_1(Z)}{n^2 \chi_1^2(n_2)} - \sigma_0^2\right)_+, 1.01\lambda_2^\text{max} \times \left(\frac{\psi_2(Z)}{n^2 \chi_2^2(n_2)} - \sigma_0^2\right)_+\right),$$

where $\lambda_2^\text{max} = \max\left\{\lambda_2^\text{max}(\hat{\Omega}), \log p\right\}$ and $\lambda_2^\text{min} = \max\left\{\lambda_2^\text{min}(\hat{\Omega}), \log p\right\}$ and

$$\psi(Z) = \min\left\{\frac{1}{n^2} \left\|y^{(2)} - X^{(2)} \hat{\beta}\right\|^2_2, \sigma_0^2 \log p\right\}.$$

We modify the construction of $\text{CI}_2^\alpha(Z, k_2, q)$ in (3.19) in the main paper [3] as follows,

$$\text{CI}_4^\alpha(Z, k_2, q) = \left(0.99\lambda_2^\text{min} \times \left(\frac{\psi(Z)}{n^2 \chi_1^2(n_2)} - \sigma_0^2\right)_+, 1.01\lambda_2^\text{max} \times \left(1 + c^* k_2\right)^{\frac{q}{2} - 1} \left(\frac{\psi(Z)}{n^2 \chi_2^2(n_2)} - \sigma_0^2\right)_+\right).$$

The following proposition shows that the minimax lower bound (2.4) can be achieved by the confidence interval $\text{CI}_3^\alpha(Z)$ in (2.6) for the case $q = 2$ and $\text{CI}_4^\alpha(Z, k_2, q)$ in (2.8) for the case $1 \leq q < 2$.

**Proposition 7.** Suppose $p \geq n$, $k_1 \leq k_2 \lesssim \frac{n}{\log p}$, $s \ll \sqrt{\frac{n}{\log p}}$ and $\hat{\beta}$ is constructed based on the subsample $Z^{(1)} = (y^{(1)}, X^{(1)})$ and satisfies Assumption (A) in the main paper [3] with $\delta > 2$. Then $\text{CI}_3^\alpha(Z)$ defined in (2.6) satisfies,

$$\liminf_{n, p \to \infty} \inf_{\theta \in \Theta_{\sigma_0}(k_2, s)} \mathbb{P}\left(\|\hat{\beta} - \beta\|_2^2 \in \text{CI}_3^\alpha(Z)\right) \geq 1 - \alpha,$$

and

$$\mathbb{L}\left(\text{CI}_3^\alpha(Z), \Theta_{\sigma_0}(k_1, s)\right) \leq \left(k_1 \log p \frac{n}{n} + \frac{1}{\sqrt{n}}\right) \sigma_0^2.$$

Then $\text{CI}_4^\alpha(Z, k_2, q)$ defined in (2.8) satisfies,

$$\liminf_{n, p \to \infty} \inf_{\theta \in \Theta_{\sigma_0}(k_2, s)} \mathbb{P}_\theta\left(\|\hat{\beta} - \beta\|_q^2 \in \text{CI}_4^\alpha(Z, k_2, q)\right) \geq 1 - \alpha,$$
and

(2.12) \[ L(C_{\alpha}^4(Z, k_2, q), \Theta_{\sigma_0}(k_1, s)) \lesssim k_2^{\frac{2}{n}} - 1 \left( k_1 \frac{\log p}{n} + \frac{1}{\sqrt{n}} \right) \sigma_0^2. \]

Similar to the construction in (2.8) and (2.6), we can modify the construction of confidence interval \( C_{\alpha}^0(Z, k, q) \) in (3.15) in the main paper [3] by estimating the minimum and maximum eigenvalues \( \lambda_{\min}(\hat{\Omega}) \) and \( \lambda_{\max}(\hat{\Omega}) \). The minimax lower bounds (2.5) and (2.3) can be achieved by such construction of confidence intervals.

3. Additional lower bound analysis. In this Section, we prove the lower bound results, Theorem 1, Theorem 2, Theorem 3, Theorem 5, Theorem 7, Theorem 9, Theorem 10, Theorem 11 and Theorem 12.

3.1. Proof of Theorem 9. Proof of (6.6)

By (6.4) and (9.3) in the main paper [3], we have \( P_{\pi_i} \left( \| \hat{\beta}(Z) - \beta^* \|_q \leq c_i d_i \right) \geq 1 - \alpha_0 - L_1(f_{\pi_i}, f_{\theta_i}), \) for \( i = 1, 2 \). Then by (9.2) in the main paper [3], we have

(3.1) \[ P_{Z, \theta \sim \pi_i} \left( \| \hat{\beta}(Z) - \beta^* \|_q \leq c_i d_i \right) \geq 1 - \alpha_0 - L_1(f_{\pi_i}, f_{\theta_i}), \] for \( i = 1, 2 \).

Define the following events

(3.2) \[ \mathcal{A}_i = \left\{ z : (1 - c_i) d_i \leq \inf_{\theta \in \mathcal{F}_i} \| \hat{\beta}(z) - \beta(\theta) \|_q \leq \sup_{\theta \in \mathcal{F}_i} \| \hat{\beta}(z) - \beta(\theta) \|_q \leq (1 + c_i) d_i \right\}, \] for \( i = 1, 2 \).

If \( \| \hat{\beta}(z) - \beta^* \|_q \leq c_i d_i \) and \( \theta \in \mathcal{F}_i \) where \( i = 1, 2 \), then

(3.3) \[ \| \hat{\beta}(z) - \beta(\theta) \|_q \geq \| \beta(\theta) - \beta^* \|_q - \| \hat{\beta}(z) - \beta^* \|_q \geq (1 - c_i) d_i, \]

and

(3.4) \[ \| \hat{\beta}(z) - \beta(\theta) \|_q \leq \| \beta(\theta) - \beta^* \|_q + \| \hat{\beta}(z) - \beta^* \|_q \leq (1 + c_i) d_i. \]

By (3.1), (3.3) and (3.4), we have \( P_{Z, \theta \sim \pi_i} (\mathcal{A}_i) \geq 1 - \alpha_0 - L_1(f_{\pi_i}, f_{\theta_i}), \) for \( i = 1, 2 \). Applying (9.2) in the main paper [3], we obtain

(3.5) \[ P_{\pi_1}(\mathcal{A}_1) \geq 1 - \alpha_0 - L_1(f_{\pi_1}, f_{\theta_1}), \] and \( P_{\pi_2}(\mathcal{A}_2) \geq 1 - \alpha_0 - L_1(f_{\pi_2}, f_{\theta_2}). \)

By the second inequality of (3.5) and (9.3) in the main paper [3], we have \( P_{\pi_1}(\mathcal{A}_2) \geq 1 - \alpha_0 - L_1(f_{\pi_2}, f_{\theta_2}) - L_1(f_{\pi_2}, f_{\pi_1}). \) Combined with the first
By the covering property, we have

\[
\inf_{\theta \in F_1 \cup F_2} \mathbb{P}_{\theta} \left( \| (\beta_{\hat{\beta}} (Z) - \beta (\theta)) \|^2 \in \text{Cl}_{\alpha} \left( \beta_{\hat{\beta}}, \ell_q, z \right) \right) \geq 1 - \alpha.
\]

Define the following event indexed with \( \theta \), \( B_{\theta} = \left\{ z : \| (\beta_{\hat{\beta}} (z) - \beta (\theta)) \|^2 \in \text{Cl}_{\alpha} \left( \beta_{\hat{\beta}}, \ell_q, z \right) \right\} \).

Define \( M_1 = \bigcup_{\theta \in F_1} B_{\theta} \) and \( M_2 = \bigcup_{\theta \in F_2} B_{\theta} \). We then obtain

\[
\mathbb{P}_{Z, \theta \sim \pi_1} (M_1) = \int \left( \int 1_{M_1} f_{\theta} (z) dz \right) \pi_1 (\theta) d\theta \geq \int \left( \int 1_{B_{\theta}} f_{\theta} (z) dz \right) \pi_1 (\theta) d\theta \geq 1 - \alpha,
\]

where the last inequality follows from (3.7). Similarly, we can establish

\[
\mathbb{P}_{Z, \theta \sim \pi_2} (M_2) \geq 1 - \alpha. By (9.2) in the main paper [3], we further have
\]

(3.8)

\[
\mathbb{P}_{\pi_1} (M_1) = \mathbb{P}_{Z, \theta \sim \pi_1} (M_1) \geq 1 - \alpha, \quad \text{and} \quad \mathbb{P}_{\pi_2} (M_2) = \mathbb{P}_{Z, \theta \sim \pi_2} (M_2) \geq 1 - \alpha.
\]

By the second inequality of (3.8) with (9.3) in the main paper [3], we have

\[
\mathbb{P}_{\pi_1} (M_2) \geq 1 - \alpha - L_1 (f_{\pi_2}, f_{\pi_1}).\]

Combined with the first inequality of (3.8), we have

(3.9)

\[
\mathbb{P}_{\pi_1} (M_1 \cap M_2) \geq 1 - 2\alpha - L_1 (f_{\pi_2}, f_{\pi_1}).
\]

Combining (3.6) and (3.9), we obtain

(3.10)

\[
\mathbb{P}_{\pi_1} (A_1 \cap A_2 \cap M_1 \cap M_2) \geq 1 - 2\alpha_0 - 2\alpha - \sum_{i=1}^{2} L_1 (f_{\pi_i}, f_{\theta_i}) - 2L_1 (f_{\pi_2}, f_{\pi_1}).
\]

For \( z \in M_1 \cap M_2 \), there exists \( \tilde{\theta}_1 \in F_1 \) and \( \tilde{\theta}_2 \in F_2 \) such that

(3.11)

\[
\| \beta_{\hat{\beta}} (z) - \beta (\tilde{\theta}_1) \|_q \in \text{Cl}_{\alpha} \left( \beta, \ell_q, z \right) \quad \text{and} \quad \| \beta_{\hat{\beta}} (z) - \beta (\tilde{\theta}_2) \|_q \in \text{Cl}_{\alpha} \left( \beta, \ell_q, z \right).
\]

Since \( z \in A_1 \cap A_2 \), we have

(3.12)

\[
(1-c_1)d_1 \leq \| \beta_{\hat{\beta}} (z) - \beta (\tilde{\theta}_1) \|_q \leq (1+c_1)d_1, \quad \text{and} \quad (1-c_2)d_2 \leq \| \beta_{\hat{\beta}} (z) - \beta (\tilde{\theta}_2) \|_q \leq (1+c_2)d_2.
\]

For \( z \in A_1 \cap A_2 \cap M_1 \cap M_2 \), (3.11) and (3.12) lead to

\[
L \left( \text{Cl}_{\alpha} \left( \beta, \ell_q, z \right) \right) \geq (1-c_2)^2 d_2^2 - (1+c_1)^2 d_1^2.
\]
Combined with (3.10), we establish
\[ E_{x_1} L \left( \text{CL}_\alpha \left( \hat{\beta}, \ell_q, Z \right) \right) \geq c_4^* \left( (1 - c_2)^2 d_2^2 - (1 + c_1)^2 d_2^2 \right). \]

Since the maximum risk is lower bounded by the Bayesian risk, we establish (6.6) in the main paper [3].

Proof of (6.5)
The proof of (6.5) in the main paper [3] combines the proof ideas of (6.2) and (6.6) in the main paper [3]. Assume that
\[ (3.13) \sup_{\theta \in F_1} \mathbb{P}_\theta \left( \left| \tilde{L}_q(Z) - \| \hat{\beta}(Z) - \beta(\theta) \|_q^2 \right| \geq \frac{1}{4} d_2^2 \right) \leq \alpha_1, \text{ with } \alpha_1 = \frac{1}{10}. \]

Otherwise, we can establish (6.5) in the main paper [3] by having
\[ (3.14) \sup_{\theta \in F_1} \mathbb{P}_\theta \left( \left| \tilde{L}_q(Z) - \| \hat{\beta}(Z) - \beta(\theta) \|_q^2 \right| \geq \frac{1}{4} d_2^2 \right) \geq \alpha_1. \]

By (3.13), we have \( \sup_{\theta \in F_1} \mathbb{P}_\theta \left( \min_{\theta \in F_1} \left| \tilde{L}_q(Z) - \| \hat{\beta}(Z) - \beta(\theta) \|_q^2 \right| \leq \frac{1}{4} d_2^2 \right) \leq \alpha_1 \) and hence
\[ \mathbb{P}_{\pi_1} \left( \min_{\theta \in F_1} \left| \tilde{L}_q(Z) - \| \hat{\beta}(Z) - \beta(\theta) \|_q^2 \right| \leq \frac{1}{4} d_2^2 \right) \geq 1 - \alpha_1. \]

Define \( M_0 = \left\{ z : \min_{\theta \in F_1} \left| \tilde{L}_q(z) - \| \hat{\beta}(z) - \beta(\theta) \|_q^2 \right| \leq \frac{1}{4} d_2^2 \right\} \). By (9.3) in the main paper [3], we obtain
\[ (3.15) \mathbb{P}_{\pi_2} (M_0) \geq 1 - \alpha_1 - L_1( f_{\pi_2}, f_{\pi_1} ). \]

Define \( A_i \) with \( i = 1, 2 \) as in (3.2). Similar to (3.6), we can establish the following control of probability
\[ (3.16) \mathbb{P}_{\pi_2} (A_1 \cap A_2) \geq 1 - 2\alpha_0 - \sum_{i=1}^{2} L_1( f_{\pi_i}, f_{\theta_i} ) - L_1( f_{\pi_2}, f_{\pi_1} ). \]

By combining (3.16) and (3.15), we establish that
\[ (3.17) \mathbb{P}_{\pi_2} (A_1 \cap A_2 \cap M_0) \geq 1 - \alpha_1 - 2\alpha_0 - \sum_{i=1}^{2} L_1( f_{\pi_i}, f_{\theta_i} ) - 2L_1( f_{\pi_2}, f_{\pi_1} ). \]

For \( z \in M_0 \), there exist \( \hat{\theta} \in F_1 \) such that
\[ (3.18) \left| \tilde{L}_q(z) - \| \hat{\beta}(z) - \beta(\hat{\theta}) \|_q^2 \right| \leq \frac{1}{4} d_2^2. \]
For \( z \in \mathcal{A}_1 \cap \mathcal{A}_2 \), we have the following results for \( \bar{\theta} \) and any \( \theta \in \mathcal{F}_2 \),
\[(3.19)
(1-c_1)d_1 \leq \| \tilde{\beta}(z) - \beta(\bar{\theta}) \|_q \leq (1+c_1)d_1, \quad \text{and} \quad (1-c_2)d_2 \leq \| \tilde{\beta}(z) - \beta(\theta) \|_q \leq (1+c_2)d_2.
\]
Hence, for \( z \in \mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{M}_0 \) and \( \theta \in \mathcal{F}_2 \), we have
\[(3.20)
\|
\bar{L}_q (z) - \| \tilde{\beta}(z) - \beta(\bar{\theta}) \|_q^2
\| \geq \| \| \tilde{\beta}(z) - \beta(\theta) \|_q^2 - \| \tilde{\beta}(z) - \beta(\bar{\theta}) \|_q^2 \| - \| \bar{L}_q (z) - \| \tilde{\beta}(z) - \beta(\theta) \|_q^2 \| \geq (1-c_2)^2 d_2^2 - (1 + c_1)^2 d_1^2,
\]
where the last inequality follows from (3.18) and (3.19). That is, for \( z \in \mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{M}_0 \), we obtain
\[(3.21)
\min_{\theta \in \mathcal{F}_2} \| \bar{L}_q (z) - \| \tilde{\beta}(z) - \beta(\theta) \|_q^2 \| \geq ((1-c_2)^2 - \frac{1}{4})d_2^2 - (1 + c_1)^2 d_1^2.
\]
Note that
\[
\sup_{\theta \in \mathcal{F}_2} \mathbb{P}_\theta \left( \bar{L}_q (Z) - \| \tilde{\beta}(Z) - \beta(\theta) \|_q^2 \right) \geq \sup_{\theta \in \mathcal{F}_2} \mathbb{P}_\theta \left( \min_{\theta \in \mathcal{F}_2} \| \bar{L}_q (Z) - \| \tilde{\beta}(Z) - \beta(\theta) \|_q^2 \right) \geq ((1-c_2)^2 - \frac{1}{4})d_2^2 - (1 + c_1)^2 d_1^2
\]
Since the max risk is lower bounded by the Bayesian risk, the final term of the last inequality can be further bounded by
\[(3.22)
\mathbb{P}_{\pi_2} \left( \min_{\theta \in \mathcal{F}_2} \| \bar{L}_q (Z) - \| \tilde{\beta}(Z) - \beta(\theta) \|_q^2 \right) \geq \mathbb{P}_{\pi_2} (\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{M}_0) \geq 1 - \alpha_1 - 2\alpha_0 - \sum_{i=1}^{2} L_1 (f_{\pi_i}, f_{\theta_i}) - 2L_1 (f_{\pi_2}, f_{\pi_1})
\]
where the inequality follows from (3.21) and (3.17). Combining (3.14) and (3.22), we establish (6.5) in the main paper [3].

**3.2. Proof of Theorem 11 and Theorem 12.** To establish the theorems, we need to establish the following three lower bounds,
\[(3.23)
\mathbf{L}_\alpha^* \left( \Theta_0 (k_1), \Theta_0 (k_2), \tilde{\beta}, \ell_q \right) \geq c_k \frac{2^{\frac{p}{2}}}{{\log p}} \sigma_0^2, \quad \text{for} \quad k_2 \leq \sqrt{\frac{n}{\log p}}
\]
\[(3.24)
\mathbf{L}_\alpha^* \left( \Theta_0 (k_1), \Theta_0 (k_2), \tilde{\beta}, \ell_q \right) \geq c_k \frac{2^{\frac{q}{2}} - 1}{{\sqrt{n}}} \sigma_0^2, \quad \text{for} \quad \sqrt{\frac{n}{\log p}} \leq k_2 \leq \frac{n}{\log p}
\]
and for $\frac{\sqrt{\pi}}{\log p} \leq k_2 \leq \frac{n}{\log p}$,

$$(3.25)$$

$$L^*_\alpha \left( \Theta_{\sigma_0} (k_1), \Theta_{\sigma_0} (k_2), \hat{\beta}, \ell_q \right)$$

$$\geq c \left( (1 - c_2)^2 M_1^2 (k_2 - k_0) \frac{2}{3} - 1 (k_1 - k_0) \rho^2 - \frac{1}{M_1^2} (1 + c_1)^2 (k_1 - k_0) \frac{2}{3} \rho^2 \right) + .$$

By the fact that $L^*_\alpha \left( \Theta_{\sigma_0} (k_1), \Theta_{\sigma_0} (k_2), \hat{\beta}, \ell_q \right) \geq L^*_\alpha \left( \Theta_0 (k_1), \Theta_0 (k_2), \hat{\beta}, \ell_q \right)$, Theorem 12 follows from (3.23), (3.24) and (3.25). The proof of Theorem 11 follows from (3.23), (3.24) and (3.25) with $M_1 = 1$. In the following, we will prove (3.23), (3.24) and (3.25), separately.

**Proof of (3.23)**

The proof of (3.23) is an application of (6.3) of Theorem 8 in the main paper [3]. For $\theta_0 = (\beta^*, I, \sigma_0) \in \Theta_0 (k_0)$, we construct

$$(3.26) \quad \mathcal{F} = \{ (\beta^* + \delta, I, \sigma_0) : \delta \in \ell (\beta^*, k_2 - k_0, \rho) \} \subset \Theta_0 (k_2) ,$$

where

$$(3.27) \quad \ell (\beta^*, k_2 - k_0, \rho) = \{ \delta : \text{supp} (\delta) \subset \text{supp} (\beta^*)^c, \| \delta \|_0 = k_2 - k_0, \delta_i \in \{ 0, \rho \} \} .$$

Let $S = \text{supp} (\beta^*)$. Without loss of generality, we assume $S = \{ 1, 2, \ldots, k_0 \}$. Let $p_1$ denote the size of $S^c$ and hence $p_1 = p - k_0$. Let $\pi$ denote the uniform prior on the parameter space $\mathcal{F}$, which is induced by the uniform prior of $\delta$ on $\ell (\beta^*, k_2 - k_0, \rho)$. Under the Gaussian random design model, $Z_i = (y_i, X_i) \in \mathbb{R}^{p+1}$ follows a joint Gaussian distribution with mean 0. Let $\Sigma^z$ denote the covariance matrix of $Z_i$. For the indices of $\Sigma^z$, we use 0 as the index of $y$, and $\{ 1, \cdots, p \}$ as the indices for $(X_{i1}, \cdots, X_{ip}) \in \mathbb{R}^p$. Decompose $\Sigma^z$ into blocks $\begin{pmatrix} \Sigma^z_{yy} & \Sigma^z_{xy} \\ \Sigma^z_{yx} & \Sigma^z_{xx} \end{pmatrix}^T$, where $\Sigma^z_{yy}, \Sigma^z_{xx}$ and $\Sigma^z_{xy}$ denote the variance of $y$, the variance of $X$ and the covariance of $y$ and $X$, respectively. There exists a bijective function $h : \Sigma^z \rightarrow (\beta, \Sigma, \sigma)$ and the inverse mapping $h^{-1} : (\beta, \Sigma, \sigma) \rightarrow \Sigma^z$, where

$$h^{-1} ((\beta, \Sigma, \sigma)) = \begin{pmatrix} \beta^T \Sigma \beta + \sigma^2 \\ \Sigma \beta \end{pmatrix}$$

and

$$(3.28) \quad h (\Sigma^z) = \begin{pmatrix} (\Sigma^z_{xx})^{-1} & \Sigma^z_{xy} \\ \Sigma^z_{yx} & \Sigma^z_{yy} - (\Sigma^z_{xy})^T (\Sigma^z_{xx})^{-1} \Sigma^z_{xy} \end{pmatrix} .$$

Based on the bijection, the control of $\chi^2 (f_x, f_{\theta_0})$ is reduced to the control of $\chi^2$ distance between two multivariate Gaussian distributions.

The parameter spaces for $\Sigma^z$ corresponding to $\{ \theta_0 \}$ and $\mathcal{F}$ are

$$\mathcal{H}_1 = \{ \Sigma^z_0 \} , \quad \text{where} \quad \Sigma^z_0 = \begin{pmatrix} \frac{\| \beta^* \|_0^2 + \sigma_0^2}{\beta^T_0 \Sigma \beta_0} & \beta^T_0 \\ \beta_0 & 1_{k_0 \times k_0} \\ 0_{k_0 \times p_1} & 0_{p_1 \times p_1} \end{pmatrix} ,$$

$$\mathcal{H}_1 = \{ \Sigma^z_0 \} , \quad \text{where} \quad \Sigma^z_0 = \begin{pmatrix} ||\beta^*||^2_0 + \sigma_0^2 & (\beta^T_0)^T \\ \beta^T_0 & 1_{k_0 \times k_0} \\ 0_{k_0 \times p_1} & 0_{p_1 \times p_1} \end{pmatrix} ,$$

and
The following two lemmas are useful to control By (9.1) in the main paper [3], it is sufficient to establish \( \chi(\beta^*, k_2 - k_0, \rho) \). Define \( \theta_1 = (\beta^*, I, \sigma_0^2 + \|\delta\|_2^2) \). For \( \delta \in \ell(\beta^*, k_2 - k_0, \rho) \), we have \( \|\delta\|_2^2 = (k_2 - k_0)^2 \) and hence \( \theta_1 = (\beta^*, I, \sigma_0^2 + (k_2 - k_0)^2) \). By \( L_1(\pi, f_{\theta_1}) \leq L_1(\pi, f_{\theta_0}) + L_1(f_{\theta_1}, f_{\theta_0}) \), it is sufficient to control \( L_1(\pi, f_{\theta_1}) \) and \( L_1(f_{\theta_1}, f_{\theta_0}) \). By (9.1) in the main paper [3], it is sufficient to establish \( \chi^2(\pi, f_{\theta_1}) \leq \epsilon_1^2 \) and \( \chi^2(f_{\theta_1}, f_{\theta_0}) \leq \epsilon_1^2 \).

Let \( E_{\delta, \tilde{\delta}} \) denote the expectation with respect to the independent random variables \( \delta, \tilde{\delta} \) with a uniform prior over the parameter space \( \ell(\beta^*, k_2 - k_0, \rho) \).

The following two lemmas are useful to control \( \chi^2(\pi, f_{\theta_1}) \) and \( \chi^2(f_{\theta_1}, f_{\theta_0}) \). The proof of Lemma 3 is given in Section 5.2.

**Lemma 3.**

\[
\begin{align*}
(3.29) \quad & \chi^2(\pi, f_{\theta_1}) + 1 = E_{\delta, \tilde{\delta}} \left( 1 - \frac{\delta^T \tilde{\delta}}{\|\delta\|_2^2 + \sigma_0^2} \right)^{-\frac{n}{2}} \left( 1 - \frac{\delta^T \tilde{\delta}}{\|\delta\|_2^2 + \sigma_0^2} \right)^{-\frac{n}{2}}. \\
(3.30) \quad & \chi^2(f_{\theta_1}, f_{\theta_0}) + 1 = E_{\delta, \tilde{\delta}} \left( 1 - \frac{\|\delta\|_2^2 \|\tilde{\delta}\|_2^2}{\sigma_0^4} \right)^{-\frac{n}{2}}. 
\end{align*}
\]

The following lemma (Lemma 3 in [4]) controls the right hand side of (3.29).

**Lemma 4.** Suppose the random variable \( J \) follows Hypergeometric \((p, k, k)\) with \( P(J = j) = \binom{k}{j} \binom{n-k}{n-j} \binom{n}{k} \), then we have

\[
(3.31) \quad E \exp(t J) \leq e^{\frac{k^2}{p}} \left( 1 - \frac{k}{p} + \frac{k}{p} \exp(t) \right)^k.
\]

By the construction (3.26), we have \( \|\delta\|_2^2 = \|\tilde{\delta}\|_2^2 = (k_2 - k_0)^2 \) and \( \delta^T \tilde{\delta} \leq (k_2 - k_0)^2 \). By the inequality \( \frac{1}{1-x} \leq \exp(2x) \) for \( x \in [0, \log \frac{2}{2}] \), if
By \((6.3)\) in the main paper [3] and the fact \(C \left( \frac{k_0}{k_2 - k_0} \right)^{\frac{1}{q}} \leq \frac{1}{16}\), we establish \((3.23)\).

**Proof of \((3.24)\)**
The proof of (3.24) is based on the exactly same argument with (3.23) by taking \( \rho = (\log (1 + \epsilon_1^2))^{\frac{1}{2}} \frac{1}{n} \frac{1}{\sqrt{k_2 - k_0}} \sigma_0 \). Since \( k_2 \geq C \frac{\sqrt{n}}{\log p} \) and \( k_0 \leq c_0 k_2 \), then

\[
(\log (1 + \epsilon_1^2))^{\frac{1}{2}} \frac{1}{n} \frac{1}{\sqrt{k_2 - k_0}} \sigma_0 \leq \frac{1}{2} \sqrt{\log \frac{\nu_1}{(k_2 - k_0)^2}} \sigma_0 \text{ and hence (3.33) holds. It is sufficient to control the following term}
\]

\[
E_{\delta, \tilde{\delta}} \left( 1 - \frac{\|\delta\|_2^2 \|	ilde{\delta}\|_2^2}{\sigma_0^4} \right)^{-\frac{3}{2}} \leq \exp \left( n \left( \frac{(k_2 - k_0)\rho^2}{\sigma_0^2} \right)^2 \right) \leq 1 + \epsilon_1^2.
\]

In this case, \( d^2 = \sqrt{\log (1 + \epsilon_1^2)}(k_2 - k_0)^{\frac{3}{2} - 1} \frac{1}{\sqrt{n}} \sigma_0^2 \) and \( \mathbb{P}_{\theta_0} \left( \|\hat{\beta} - \beta^*\|_q^2 \leq C \frac{k_0^2 \log p}{(k_2 - k_0)^{\frac{1}{2} - 1} \sqrt{n}} d^2 \right) \geq 1 - \alpha_0 \). Since \( k_0 \leq c_0 \min \{k_1, \frac{\nu_1}{\log p} \} \) and \( k_1 \leq k_2 \), we have \( C \frac{k_0^2 \log p}{(k_2 - k_0)^{\frac{1}{2} - 1} \sqrt{n}} \leq \frac{1}{16} \) and the lower bound (3.24) follows from (6.3) of Theorem 8 in the main paper [3].

**Proof of (3.25)** The proof of (3.25) is an application of (6.6) of Theorem 9 in the main paper [3]. The key is to construct parameter spaces \( \mathcal{F}_1, \mathcal{F}_2 \) and the points \( \theta_1 \) and \( \theta_2 \) and then control the distribution distances between the density functions. Let \( S = \text{supp} (\beta^*) \). Without loss of generality, we assume \( S = \{1, 2, \cdots, k_0\} \). Define \( p_0 \) denote the largest integer smaller than \( \frac{p - k_0}{2} \), \( \mathcal{I}_1 = \{k_0 + 1, k_0 + 2, \cdots, k_0 + p_0\} \) and \( \mathcal{I}_2 = \{k_0 + p_0 + 1, k_0 + p_0 + 2, \cdots, p\} \). Define \( \Sigma_0 = \begin{pmatrix} I_{S \times S} & 0 & 0 \\ 0 & M_1 I_{I_1 \times I_1} & 0 \\ 0 & 0 & M_1 I_{I_2 \times I_2} \end{pmatrix} \).

For \( \theta_0 = (\beta^*, I, \sigma_0) \), we construct

\[
\mathcal{F}_1 = \left\{ (\beta^* + \nu, \Sigma_0, \sigma_0) : \nu \in \ell \left( \mathcal{I}_1, k_1 - k_0, \frac{1}{M_1} \rho \right) \right\} \subset \Theta_0 (k_1),
\]

\[
\mathcal{F}_2 = \left\{ (\beta^* + \delta, I, \sigma_0) : \delta \in \ell \left( \mathcal{I}_2, k_2 - k_0, M_1 \sqrt{\frac{k_1 - k_0}{k_2 - k_0} \rho} \right) \right\} \subset \Theta_0 (k_2),
\]

where

\[
\ell (\mathcal{I}_1, k_2 - k_0, \rho) = \left\{ \nu : \text{supp} (\nu) \subset \mathcal{I}_1, \|\nu\|_0 = k_1 - k_0, \nu_i \in \frac{1}{M_1} \{0, \rho\} \right\},
\]

and

\[
\ell (\mathcal{I}_2, k_2 - k_0, \rho) = \left\{ \delta : \text{supp} (\delta) \subset \mathcal{I}_2, \|\delta\|_0 = k_2 - k_0, \delta_i \in M_1 \sqrt{\frac{k_1 - k_0}{k_2 - k_0} \{0, \rho\}} \right\}.
\]
Let $\pi_i$ denote the uniform prior on the parameter space $\mathcal{F}_i$ for $i = 1, 2$. The corresponding parameter spaces for $\Sigma^z$ corresponding to $\mathcal{F}_1$ and $\mathcal{F}_2$ are

$$\mathcal{H}_1 = \left\{ \Sigma^z_{\nu} : \nu \in \ell \left( I_1, k_1 - k_0, \frac{1}{M_1} \rho \right) \right\} \text{ where } \Sigma^z_{\nu} = \left( \frac{\| \beta^* \|^2_2 + (k_1 - k_0) \rho^2 + \sigma_0^2}{\beta^* + \nu} \right) \left( \frac{(\beta^* + \nu)^T}{\Sigma_0} \right),$$

and

$$\mathcal{H}_2 = \left\{ \Sigma^z_{\delta} : \delta \in \ell \left( I_2, k_2 - k_0, M_1 \sqrt{\frac{k_1 - k_0}{k_2 - k_0}} \rho \right) \right\} \text{ where } \Sigma^z_{\delta} = \left( \frac{\| \beta^* \|^2_2 + (k_1 - k_0) \rho^2 + \sigma_0^2}{\beta^* + \delta} \right) \left( \frac{(\beta^* + \delta)^T}{\Sigma_0} \right).$$

Since $\text{dist}_1^2 = M_1^2 \| \nu \|^2_2 = (k_1 - k_0) \rho^2$ and $\text{dist}_2^2 = \frac{1}{M_1^2} \| \delta \|^2_2 = (k_1 - k_0) \rho^2$, we have $\theta_1 = \theta_2 = (\beta^*, \Sigma_0, \sigma_0^2 + (k_1 - k_0) \rho^2)$. In this case, we have $L_1 (f_{\theta_2}, f_{\theta_1}) = 0$. By the same argument of (3.29) in Lemma 3, we have

$$\chi^2 (f_{\pi_1}, f_{\theta_1}) + 1 = E_{\nu, \tilde{\nu}} \left( 1 - \frac{M_1 \nu^T \tilde{\nu}}{(k_1 - k_0) \rho^2 + \sigma_0^2} \right)^{-\frac{\nu}{2}} \left( 1 - \frac{M_1 \nu^T \tilde{\nu}}{(k_1 - k_0) \rho^2 + \sigma_0^2} \right)^{-\frac{\sigma_0^2}{2}},$$

and

$$\chi^2 (f_{\pi_2}, f_{\theta_2}) + 1 = E_{\delta, \tilde{\delta}} \left( 1 - \frac{1}{M_1} \delta^T \tilde{\delta} \right)^{-\frac{\nu}{2}} \left( 1 - \frac{1}{M_1} \delta^T \tilde{\delta} \right)^{-\frac{\sigma_0^2}{2}}.$$

Taking $\rho = \frac{1}{2(M_1^2 \log \frac{\rho}{\sqrt{2} M_1^2 \sigma_0^2})} \sigma_0$, a similar argument to (3.32) and (3.33) leads to $L_1 (f_{\pi_i}, f_{\theta_i}) \leq \epsilon_1$ for $i = 1, 2$. Note that $d_1^2 = \frac{1}{M_1^2} (k_1 - k_0)^2 \rho^2$, $d_2^2 = M_1^2 (k_2 - k_0)^2 \rho^2$. The assumption (2.7) leads to $P_{\theta_i} \left( \| \beta - \beta^* \|^2_2 \leq c_2^2 d_2^2 \right) \geq 1 - \alpha_0$, for $i = 1, 2$, where $c_1 = \frac{C^* M_1 k_0^{\frac{1}{2}}}{(k_1 - k_0)^{\frac{1}{2}}} \in \Theta_0 (k_1), \Theta_0 (k_2), \tilde{\beta}, \ell_q$ and $c_2 = \frac{C^* k_0^{\frac{1}{2}}}{M_1 (k_2 - k_0)^{\frac{1}{2}} (k_1 - k_0)^{\frac{1}{2}}}$. By (6.6) in the main paper [3], we obtain (3.39)

$$L^* \left( \Theta_0 (k_1), \Theta_0 (k_2), \tilde{\beta}, \ell_q \right) \geq c \left[ 1 - (1 - c_2)^2 M_1^2 (k_2 - k_0)^{\frac{1}{2}} \rho^2 \right] \left( 1 + c_1^2 (1 + c_1^2 (k_1 - k_0)^{\frac{1}{2}} \rho^2) \right).$$

3.3. Proof of Theorem 10. Theorem 10 is implied by (2.4) in Theorem 12, since $k_0 \leq c_0^2 k_1$, we have $(1 - c_2)^2 \geq \frac{1}{M_1^2}$ and $(1 + c_1^2) \leq M_1$. Combined with $k_0 \leq c_0 \min \{ k_1, \sqrt{\frac{n}{\log \rho}} \}$, we can establish (7.3) in the main paper [3].
3.4. Proof of Theorems 3 and 5. The minimax lower bound of Theorem 3 follows from Theorem 4. We take $k_1 = k_2 = k$ and (3.5) in the main paper [3] follows from (3.7) in the main paper [3]. The minimax lower bound (3.6) in the main paper [3] follows from (3.5) in the main paper [3] and Lemma 2. The minimax lower bound of Theorem 5 follows from Theorem 6. We take $k_1 = k_2 = k$ and (3.13) follows from (3.18). The minimax lower bound (3.14) in the main paper [3] follows from (3.13) in the main paper [3] and Lemma 2.

3.5. Proof of Theorems 2 and 7. The proofs of Theorem 2 and 7 are applications of the minimax lower bounds (6.2) and (6.3) of Theorem 8 in the main paper [3], respectively. To apply Theorem 8, it is sufficient to construct the least favorable set $\mathcal{F}$ corresponding to the point $\theta_0 = (\beta^*, I, \sigma_0)$ such that the distribution distance $L_1(f, f_0)$ or $\chi^2(f, f_0)$ is controlled and the functional distance $d = \min_{\theta \in \mathcal{F}} L(\theta, \beta^*)$ is maximized. In the following, we first establish Theorem 7 by constructing $\mathcal{F}$ with the prior $\pi$ and control the distance $\chi^2(f, f_0)$. Taking $\theta_0 = (\beta^*, I, \sigma_0)$ with $\|\beta^*\|_0 \leq k_0$ and $\Theta = \Theta (k_2)$, we define $\mathcal{F} = \mathcal{F}(\theta_0, k_2 - k_0, \rho)$ as

$$
\mathcal{F}(\theta_0, k_2 - k_0, \rho) = \left\{ \theta = (\beta^* + \delta, I, \sigma) : \delta \in \ell (\beta^*, k_2 - k_0, \rho), \sigma = \sqrt{\sigma_0^2 + (k_2 - k_0) \rho^2} \right\},
$$

where $\ell (\beta^*, k_2 - k_0, \rho)$ is defined in (3.27). Note that $\mathcal{F}(\theta_0, k_2 - k_0, \rho) \subset \Theta (k_2)$. The prior $\pi$ on $\mathcal{F}(\theta_0, k_2 - k_0, \rho)$ is induced by the uniform prior of $\delta$ on $\ell (\beta^*, k_2 - k_0, \rho)$. In the following, we will control $\chi^2(f, f_0)$.

Let $\Sigma_0^0$ denote the covariance matrix of $(y_i, X_i)$ corresponding to $\theta_0 = (\beta^*, I, \sigma) \in \Theta (k_0)$. Let $S = \text{supp} (\beta^*)$. Without loss of generality, we assume $S = \{1, 2, \cdots, k_0\}$. Let $p_1$ denote the size of $S^c$ and hence $p_1 = p - k_0$. We have the expression of the covariance matrix $\Sigma_0^0$ corresponding to $\theta_0$,

$$
\Sigma_0^0 = \begin{pmatrix}
\|\beta^*\|_0^2 + \sigma_0^2 & (\beta^*_S)^T \\
\beta^*_S & I_{k_0 \times k_0} & 0_{k_0 \times p_1} \\
0_{p_1 \times 1} & 0_{p_1 \times k_0} & I_{p_1 \times p_1}
\end{pmatrix}.
$$

The corresponding set of $\Sigma^*$ to $\mathcal{F} = \mathcal{F}(\theta_0, k_2 - k_0, \rho)$ is

$$
\mathcal{H} = \{ \Sigma_0^\delta : \delta \in \ell (\beta^*, k_2 - k_0, \rho) \}, \quad \text{with} \quad \Sigma_0^\delta = \begin{pmatrix}
\|\beta^*\|_0^2 + \sigma_0^2 & (\beta^*_S)^T \\
\beta^*_S & I_{k_0 \times k_0} & 0_{k_0 \times p_1} \\
\delta & 0_{p_1 \times k_0} & I_{p_1 \times p_1}
\end{pmatrix}.
$$

The following lemma (Lemma 7 in [5]) controls the $\chi^2$ distance between $f_\pi$ and $f_{\theta_0}$. 
By the construction (3.40), we have \( \| \delta \|^2_2 = \| \hat{\delta} \|^2_2 = (k_2 - k_0) \rho^2 \) and \( \delta^\top \hat{\delta} \leq (k_2 - k_0) \rho^2 \). By the inequality \( \frac{1}{2x} \leq \exp(2x) \) for \( x \in \left[ 0, \frac{\log 2}{2} \right] \), if \( \frac{(k_2 - k_0) \rho^2}{\sigma^2_0} < \frac{\log 2}{2} \), we have \( (1 - \frac{1}{2} \delta^\top \hat{\delta})^{-n} \leq \exp \left( \frac{2}{\sigma^2_0} n \delta^\top \hat{\delta} \right) \). Let \( J \) denote the hypergeometric distribution with parameters \((p_1, k_2 - k_0, k_2 - k_0)\). We further have

\[
\mathbb{E} \exp \left( \frac{2}{\sigma^2_0} n \delta^\top \hat{\delta} \right) = \mathbb{E} \exp \left( \frac{1}{\sigma^2_0} \frac{2J \rho^2}{n} \right) \leq e^{\frac{(k_2 - k_0)^2}{\sigma^2_0} \frac{2J \rho^2}{n}} \left( 1 - \frac{k_2 - k_0}{p_1} + \frac{k_2 - k_0}{p_1} \exp \left( \frac{1}{\sigma^2_0} \frac{2J \rho^2}{n} \right) \right)^{k_2 - k_0} \leq e^{\frac{(k_2 - k_0)^2}{\sigma^2_0} \frac{2J \rho^2}{n}} \left( 1 + \frac{1}{\sqrt{p_1}} \right)^{k_2 - k_0},
\]

where the first inequality applies Lemma 4 and the second inequality follows by plugging \( \rho = \frac{1}{2} \sqrt{\frac{\log \left( \frac{p_1}{(k_2 - k_0)^2} \right)}{n}} \sigma_0 \). If \( k_2 \leq c_0 \min \left\{ \frac{n}{\log p}, \frac{p}{\gamma} \right\} \), we have \( \frac{(k_2 - k_0)^2}{\sigma^2_0} \frac{2J \rho^2}{n} < \frac{\log 2}{2} \) and establish \( \chi^2(f_\pi, f_\theta_0) \leq \epsilon_1 \) by (3.44) and \( L_1(f_\pi, f_\theta_0) \leq \epsilon_1 \) by (9.1) in the main paper [3], where \( \epsilon_1 = \frac{1 - 2\alpha - 2\alpha_0}{12} \).

To establish Theorem 7, we apply Theorem 8 and compute \( d = \frac{1}{2} (k_2 - k_0) \sqrt{\frac{\log \left( \frac{p_1}{(k_2 - k_0)^2} \right)}{n}} \sigma_0 \). By (2.6) in the main paper [3] and \( \| \beta^* \|_0 \leq k_0 \), we establish

\[
\mathbb{P}_{\theta_0} \left( \| \hat{\beta} - \beta^* \|_q^2 \leq C^* \left( \frac{k_0}{k_2 - k_0} \right)^{\frac{2}{q}} d^2 \right) \geq 1 - \alpha_0.
\]

By the fact \( C^* \left( \frac{k_0}{k_2 - k_0} \right)^{\frac{1}{q}} \leq \frac{1}{4} \), we establish (6.1) in the main paper [3]. By applying (6.3) of Theorem 8, we establish (4.2) in the main paper [3]. Since \( \theta_0 \in \Theta(k_2) \) and \( L^*_\alpha \left( \Theta(k_2), \Theta, \ell_q \right) \geq L^*_\alpha \left( \{ \theta_0 \}, \Theta(k_2), \hat{\beta}, \ell_q \right) \), the lower bound (4.1) in the main paper [3] with \( i = 2 \) follows from (4.2) in the main paper [3]. For (4.1) in the main paper [3] with \( i = 1 \), the lower bound is established using the above argument with \( k_2 \) replaced by \( k_1 \). The following lemma shows that \( \hat{\beta}^SL \) with \( A > 2\sqrt{2} \) satisfying the assumption (A2) and hence the lower bounds (4.1) and (4.2) in the main paper [3] hold for \( \hat{\beta}^SL \) with \( A > 2\sqrt{2} \).
Lemma 6. If $A > 2\sqrt{2}$, then we have
\[ P_{\theta_0} \left( \| \hat{\beta}^{SL} - \beta^* \|_q^2 \leq C \| \beta^* \|_0 \frac{2 \log p}{n \sigma_0^2} \right) \geq 1 - c \exp (-c' n) - p^{-c}. \]

To establish Theorem 2, we apply the same argument between (3.40) and (3.45) by replacing $k_2$ with $k$ and the general lower bound (6.2) in the main paper [3].

3.6. Proof of Theorem 1. The proof of Theorem 1 is similar to the proof of Theorem 11, which is presented in Section 3.2. For the case $k \lesssim \sqrt{n \log p}$, the proof is similar to (3.23). Taking $\theta_0$, $F$ and $\rho$ as defined in the proof of (3.23), with $k_2 = k$, we apply (6.2) in Theorem 8, we establish (2.8) and (2.9) in the regime $k \lesssim \sqrt{n \log p}$ in the main paper [3].

The proof of (2.8) in the case $\sqrt{n \log p} \ll k \sim n \log p$ is similar to that of (3.24). Taking $\theta_0$, $F$ and $\rho$ as defined in the proof of (3.24), with $k_2 = k$, we establish (2.8) in the main paper [3] for $\sqrt{n \log p} \ll k \lesssim n \log p$. The proof of (2.9) in the main paper [3] in the case $\sqrt{n \log p} \ll k \lesssim n \log p$ is similar to that of (3.25). Taking $\theta_0, \theta_1, \theta_2, F_1, F_2$ and $\rho$ as defined in the proof of (3.25), with $k_1 = \frac{1}{3} k$ and $k_2 = k$, we apply (6.5) of Theorem 9 to establish (2.9) in the main paper [3] for $\sqrt{n \log p} \ll k \lesssim n \log p$.

4. Upper bound analysis. In Section 4.1, we establish minimax upper bounds of Theorem 3, 4, 5, 6 and 7 based on Proposition 2, 3, 4 and 5. In later sections, we establish Proposition 1, 2, 3, 4, 5, 6 and 7.

4.1. Proof of upper bounds of Theorems. In the following, we will establish the minimax upper bounds in the main paper based on Proposition 2, 3, 4 and 5.

Proof of the upper bound of Theorem 3 By Proposition 3, the minimax convergence rate (3.6) in the main paper [3] over $k \lesssim \frac{\sqrt{n}}{\log p}$ is achieved by the confidence interval $\text{CI}_0^0 (Z, k, 2)$ defined in (3.15) in the main paper [3]. By Proposition 2, the minimax convergence rate (3.6) in the main paper [3] over $\frac{\sqrt{n}}{\log p} \ll k \lesssim \frac{n}{\log p}$ is achieved by the confidence interval $\text{CI}_1^1 (Z)$ defined in (3.8) in the main paper [3].

Proof of the upper bound of Theorem 4 By Proposition 3, the minimax lower bound (3.7) in the main paper [3] in the region $k_2 \lesssim \frac{\sqrt{n}}{\log p}$ is achieved by the confidence interval $\text{CI}_0^0 (Z, k_2, 2)$. By proposition 2, the minimax lower bound (3.7) in the main paper [3] over $\frac{\sqrt{n}}{\log p} \ll k_2 \lesssim \frac{n}{\log p}$ is
achieved by the confidence interval $\text{CI}_1^\alpha (Z)$.

**Proof of the upper bound of Theorem 5** By Proposition 3, the minimax convergence rate (3.14) in the main paper [3] is achieved by the confidence interval $\text{CI}_0^\alpha (Z, k, q)$.

**Proof of the upper bound of Theorem 6** By Proposition 3, the minimax lower bound (3.18) in the main paper [3] in the regime $k_1 \leq k_2 \lesssim \sqrt{n \log p}$ is achieved by the confidence interval $\text{CI}_0^\alpha (Z, k_2, q)$. By Proposition 4, the minimax lower bounds (3.18) in the main paper [3] in the regime $k_1 \lesssim \frac{n}{\log p} \lesssim k_2 \leq k_2 \lesssim \frac{n}{\log p}$ are achieved by the confidence interval $\text{CI}_2^\alpha (Z, k_2, q)$ defined in (3.19) in the main paper [3].

**Proof of the upper bound of Theorem 7** By Proposition 5, the minimax lower bounds (4.1) in the main paper [3] are achieved by the confidence interval $\text{CI}_\alpha (Z, k_i, q)$ defined in (4.4) in the main paper [3] for $i = 1, 2$ and the lower bound in (4.2) in the main paper [3] is achieved by the confidence interval $\text{CI}_\alpha (Z, k_2, q)$.

### 4.2. Proof of Proposition 5

The following argument is similar to the upper bound argument in [4, 5], which also relies on the results from [1, 8, 9]. We first normalize the columns of $X$ and the true sparse vector $\beta$ and the linear regression model can be expressed as

\[(4.1) \quad y = Wd + \epsilon, \quad \text{with} \quad W = XD, \quad d = D^{-1} \beta \quad \text{and} \quad \epsilon \sim N(0, \sigma^2 I),\]

where $D = \text{diag} \left( \frac{\sqrt{n}}{\|X_j\|_2} \right)_{j \in [p]}$ denotes the $p \times p$ diagonal matrix with $(j, j)$ entry to be $\frac{\sqrt{n}}{\|X_j\|_2}$. Setting $\delta_0 = \frac{4}{\sqrt{2}}$ and $\eta_0 = (\frac{4}{\sqrt{2}})^\frac{1}{2} - 1$, we have $\lambda_0 = (1 + \eta_0) \sqrt{\frac{2k_0 \log p}{n}}$. Take $\epsilon_0 = \frac{2.01}{\eta_0} + 1$, $\nu_0 = 0.01$ and $C_1 = 2.25$. Rather than use the constants directly in the following discussion, we use $\delta_0, \eta_0, \epsilon_0, \nu_0$ and $C_1$ to represent the above fixed constants in the following discussion.

We also assume that $\frac{\log p}{n} \leq \frac{1}{25}$ and $\delta_0 \log p > 2$.

Define the $l_1$ cone invertibility factor ($CIF_1$) as follows,

\[(4.2) \quad CIF_1 (\alpha_0, K, W) = \inf \left\{ \frac{|K| \|\frac{W^TW}{n} u\|_\infty}{\|u_K\|_1} : \|u_{K^c}\|_1 \leq \alpha_0 \|u_K\|_1, \ u \neq 0 \right\} ,\]
where $K$ is an index set. Define $\sigma_{ora} = \frac{1}{\sqrt{n}} \|y - X\beta\|_2 = \frac{1}{\sqrt{n}} \|y - Wd\|_2$,

$$T = \{ k : |d_k| \geq \lambda_0 \sigma_{ora} \}, \quad \tau = (1 + \epsilon_0) \lambda_0 \max \left\{ \frac{4}{\sigma_{ora}} \|d_T\|_1, \frac{8\lambda_0 |T|}{CIF_1 (2\epsilon_0 + 1, T, W)} \right\}.$$

To facilitate the proof, we define the following events for the random design $X$ and the error $\epsilon$,

\begin{align*}
G_1 &= \left\{ \frac{2}{5} \frac{1}{\sqrt{M_1}} < \frac{\|X_j\|_2}{\sqrt{n}} < \frac{7}{5} \sqrt{M_1} \text{ for } 1 \leq j \leq p \right\}, \\
G_2 &= \left\{ \frac{(\sigma_{ora})^2}{\sigma^2} - 1 \leq 2 \frac{\log p}{n} + 2 \frac{\log p}{n} \right\}, \\
G_3 &= \left\{ \kappa(X, k, k, \alpha) \geq \frac{1}{4\sqrt{\lambda_{\max} (\Omega)}} - \frac{9}{\sqrt{\lambda_{\min} (\Omega)}} (1 + \alpha) \sqrt{\frac{k \log p}{n}} \right\}, \\
G_4 &= \left\{ \frac{\|W^\top \epsilon\|_\infty}{n} \leq \sigma \sqrt{\frac{2\delta_0 \log p}{n}} \right\}, \\
S_1 &= \left\{ \frac{\|W^\top \epsilon\|_\infty}{n} \leq \sigma \sigma_{ora} \lambda_0 \frac{\epsilon_0 - 1}{\epsilon_0 + 1} (1 - \tau) \right\}, \\
S_2 &= \left\{ (1 - \nu_0) \hat{\sigma} \leq \sigma \leq (1 + \nu_0) \hat{\sigma} \right\}.
\end{align*}

Define $G = \cap_{i=1}^5 G_i$ and $S = \cap_{i=1}^2 S_i$. We introduce the following lemma to control the probability of events $G$ and $S$. Lemma 7 was established as Lemma 4 in [4], which relies on the results in [8].

**Lemma 7.**

\begin{equation}
\Pr_{\theta} (G) \geq 1 - \frac{6}{p} - 2p^{1-C_1} - \frac{1}{2\sqrt{\pi \delta_0 \log p}} p^{1-\delta_0} - c' \exp \left( -cn \right),
\end{equation}

where $c$ and $c'$ are universal positive constants. If $k \leq c \frac{n}{\log p}$, then

\begin{equation}
\Pr_{\theta} (G \cap S) \geq \Pr_{\theta} (G) - 2 \exp \left( - \left( g_0 + 1 - \sqrt{2g_0 + 1} \right) n \right) - c'' \frac{1}{\sqrt{\log p}} p^{1-\delta_0},
\end{equation}

where $c, c'$ and $c''$ are universal positive constants and $g_0 = \frac{\nu_0}{2 + 3\nu_0}$.

The following lemma establishes a data-dependent upper bound for the term $\|\hat{\beta} - \beta\|_q^2$ with $1 \leq q \leq 2$. The proof of this lemma is in Section 5.3.
Lemma 8. On the event $\mathcal{G} \cap \mathcal{S}$,

\begin{equation}
\|\hat{\beta}^{SL} - \beta\|_q^2 \leq \left( \frac{16A_{\max} \|X_j\|_2^2 \hat{\sigma}}{nk^2 \left( X, k, k, 3 \left( \frac{\max \|X_j\|_2}{\min \|X_j\|_2} \right) \right)} \right)^2 \frac{k^2 \log p}{n}.
\end{equation}

On the event $\mathcal{G} \cap \mathcal{S}$, we have $\hat{\sigma} \leq (1 + \nu_0)\sigma < \log p$ and there exists $p_0$ such that if $p \geq p_0$, then we have

\begin{equation}
\left( \frac{16A_{\max} \|X_j\|_2^2 \hat{\sigma}}{nk^2 \left( X, k, k, 3 \left( \frac{\max \|X_j\|_2}{\min \|X_j\|_2} \right) \right)} \right)^2 \frac{k^2 \log p}{n} \leq C \left( k^\frac{3}{2} \log p \right) \left( p^{1-\min\{\delta_0, C_1\}} + c' \exp (-cn) \right) (\log p)^3
\end{equation}

By Lemma 8, we have $P_\theta \left( \|\hat{\beta} - \beta\|_q^2 \in \text{CI}_\alpha (Z, k, q) \right) \geq P_\theta (\mathcal{G} \cap \mathcal{S})$. Then the coverage property (4.5) in the main paper [3] follows from Lemma 7. The expected length is controlled as follows,

\begin{equation}
\mathbf{E}_\theta L (\text{CI}_\alpha (Z, k, q)) = \mathbf{E}_\theta L (\text{CI}_\alpha (Z, k, q)) 1_B = \mathbf{E}_\theta L (\text{CI}_\alpha (Z, k, q)) 1_{G\cap(S\cap\mathcal{G})} + \mathbf{E}_\theta L (\text{CI}_\alpha (Z, k, q)) 1_{G\cap(S\cap\mathcal{G})^c}
\end{equation}

where the first inequality follows from (4.7) and second inequality follows from Lemma 7. If $\log p \leq c$, then $(p^{1-\min\{\delta_0, C_1\}} + c' \exp (-cn)) (\log p)^3 \to 0$ and hence (4.6) in the main paper [3] follows.

4.3. Proof of Proposition 3. For the split samples $y^{(1)} = X^{(1)} \beta + \epsilon^{(1)}$ and $y^{(2)} = X^{(2)} \beta + \epsilon^{(2)}$, we define the following events

\begin{align*}
\bar{\mathcal{G}}_1 &= \left\{ 0.9 < \frac{\|X_j^{(1)}\|_2}{\sqrt{n_1}} < 1.1 \text{ for } 1 \leq j \leq p \right\}, \\
\bar{\mathcal{G}}_2 &= \left\{ \kappa(X^{(1)}) \geq 4\sqrt{\lambda_{\max} (\Omega)} - \frac{9}{\sqrt{\lambda_{\min} (\Omega)}} (1 + \alpha) \sqrt{2k \log p} \frac{n_1}{n_1} \right\}, \\
\bar{\mathcal{G}}_3 &= \left\{ \frac{2 \| (W^{(1)})^T \epsilon^{(1)} \|_\infty}{n_1} \leq A \sqrt{\frac{\log p}{n_1}} \sigma_0 \right\}, \\
\bar{\mathcal{G}}_4 &= \left\{ \frac{\| (W^{(1)})^T \epsilon^{(1)} \|_\infty}{n_1} \leq \eta_0 - 1 \frac{\log p}{\eta_0 + 1} A \sqrt{\frac{\log p}{n_1}} \sigma_0 \right\},
\end{align*}

where $\eta_0$ is defined in Proposition 3.
\( \mathcal{G}_5 = \left\{ \| \hat{\beta}^L - \beta \|_q^2 \leq C_1^* (A, k) \frac{k \log p}{n} \right\} \),
\( \mathcal{G}_6 = \left\{ \| \hat{\beta}^L - \beta \|_2^2 \leq C_2^* (A, k) \frac{k \log p}{n} \right\} \),
\( \mathcal{G}_7 = \left\{ \frac{1}{n^2} \| y^{(2)} - X^{(2)} \hat{\beta}^L \|_2^2 \leq \left( \frac{\sigma_0^2}{\sqrt{2}} + \| \hat{\beta}^L - \beta \|_2^2 \right) \left( 1 + 2 \sqrt{\frac{2 \log p}{n_2} + \frac{2 \log p}{n_2}} \right) \right\} \),
\( \mathcal{G}_8 = \left\{ \frac{1}{n^2} \| y^{(2)} - X^{(2)} \hat{\beta}^L \|_2^2 \leq \sigma_0^2 \left( 1 + \frac{k \log p}{n_1} \right) \left( 1 + 2 \sqrt{\frac{2 \log p}{n_2} + \frac{2 \log p}{n_2}} \right) \right\} \),

where \( W^{(1)}_j = X^{(1)}_j \frac{\sqrt{\eta_0}}{\| X^{(1)}_j \|_2} \) for \( j = 1, \ldots, p \) and \( \eta_0 = 1.01 \frac{\sqrt{\eta_0} + \sqrt{2}}{\sqrt{A - \sqrt{2}}} \), \( C_1^* (A, k) = \frac{(22 A \sigma_0)^2}{\left( \frac{1}{4} - 42 \sqrt{2} \frac{2 \log p}{n_1} \right)^4} \) and \( C_2^* (A, k) = \frac{(\frac{\eta_0}{\sqrt{2}} + A \sigma_0)^2}{\left( \frac{1}{4} - (9 + 11 \eta_0) \sqrt{\frac{2 \log p}{n_1}} \right)^4} \). We introduce the following lemma to control the probability \( P_{\theta} (\mathcal{G}_i) \) for \( 1 \leq i \leq 8 \). The lemma is proved in Section 5.4.

**Lemma 9.** Suppose \( k \leq c \frac{n}{\log p} \) and \( \theta \in \Theta_0 (k) \). If \( A > 4 \sqrt{2} \), we have

\[ P_{\theta} (\mathcal{G}_5) \geq P_{\theta} (\mathcal{G}_1 \cap \mathcal{G}_2 \cap \mathcal{G}_3) \geq 1 - c \exp (-c' n) - cp^{1 - \frac{4 \log p}{n_2}}. \]

If \( A > \sqrt{2} \), we have

\[ P_{\theta} (\mathcal{G}_6) \geq P_{\theta} (\mathcal{G}_1 \cap \mathcal{G}_2 \cap \mathcal{G}_4) \geq 1 - c \exp (-c' n) - p^{-c}, \]

and

\[ P_{\theta} (\mathcal{G}_7) \geq 1 - p^{-2} \quad \text{and} \quad P_{\theta} (\mathcal{G}_8) \geq P_{\theta} (\mathcal{G}_6) (1 - cp^{-2}), \]

where \( c \) and \( c' \) are positive constants.

The coverage property (3.16) in the main paper [3] follows from the fact that \( P_{\theta} \left( \| \hat{\beta}^L - \beta \|_q^2 \in \text{CI}_0^* (Z, k, q) \right) \geq P_{\theta} (\mathcal{G}_5) \). For the case \( q = 2 \), the coverage property follows from \( P_{\theta} \left( \| \hat{\beta}^L - \beta \|_2^2 \in \text{CI}_0^* (Z, k, 2) \right) \geq P_{\theta} (\mathcal{G}_6) \). The expected length (3.17) in the main paper [3] follows from the definition of \( \text{CI}_0^* (Z, k, q) \).
4.4. Proof of Proposition 1. We first introduce the following lemma (Theorem 2.3 in [2]) about concentration of $\chi^2$ random variable.

**Lemma 10.** Let $\chi^2_n$ denote the $\chi^2$ random variable with $n$ degrees of freedom, then we have the following concentration inequality,

$$
\mathbb{P}\left( |\chi^2_n - E\chi^2_n| > 2\sqrt{nt} + 2t \right) \leq 2 \exp(-t).
$$

Conditioning on $(y^{(1)}, X^{(1)})$, we have $y_i^{(2)} - X_i^{(2)}\hat{\beta}^L = X_i^{(2)}(\beta - \hat{\beta}^L) + \epsilon_i^{(2)} \sim N\left(0, \|\beta - \hat{\beta}^L\|^2_2 + \sigma_0^2\right)$ and hence

$$
\|y^{(2)} - X^{(2)}\hat{\beta}^L\|^2_2 \sim \left(\|\beta - \hat{\beta}^L\|^2_2 + \sigma_0^2\right) \chi^2(n_2).
$$

Since $\|\hat{\beta}^L - \beta\|^2_2$ is non-negative, we have

$$
\mathbb{P}_\theta\left(\left|\frac{1}{n_2}\|y^{(2)} - X^{(2)}\hat{\beta}^L\|^2_2 - \sigma_0^2 - \|\hat{\beta}^L - \beta\|^2_2\right| \geq \delta_{n,p} \frac{1}{\sqrt{n}}\right) \leq \mathbb{P}_\theta\left(\left|\frac{1}{n_2}\|y^{(2)} - X^{(2)}\hat{\beta}^L\|^2_2 - \sigma_0^2 - \|\hat{\beta}^L - \beta\|^2_2\right| \geq \delta_{n,p} \frac{1}{\sqrt{n}}\right).
$$

By Lemma 10 and (4.12), we establish that

$$
\mathbb{P}_\theta\left(\left|\frac{1}{n_2}\|y^{(2)} - X^{(2)}\hat{\beta}^L\|^2_2 - \sigma_0^2 - \|\hat{\beta}^L - \beta\|^2_2\right| \geq \delta_{n,p} \frac{1}{\sqrt{n}}\right) \leq \mathbb{P}_\theta\left(\left|\frac{1}{n_2}\|y^{(2)} - X^{(2)}\hat{\beta}^L\|^2_2 - \sigma_0^2 - \|\hat{\beta}^L - \beta\|^2_2\right| \geq \delta_{n,p} \frac{1}{\sqrt{n}}\right).
$$

Taking supremum on both sides of (4.13), we establish (2.12) in the main paper [3].

4.5. Proof of Proposition 2. We first introduce the following confidence intervals,

$$
\bar{\text{CI}}_\alpha^1(Z) = \left(\frac{\psi(Z)}{\frac{1}{n_2}\chi^2_{1 - \frac{\alpha}{2}}(n_2)} - \sigma_0^2, \frac{\psi(Z)}{\frac{1}{n_2}\chi^2_{\frac{\alpha}{2}}(n_2)} - \sigma_0^2\right).
$$

Since the loss $\|\hat{\beta} - \beta\|^2_2$ is positive, the coverage property of $\text{CI}_\alpha^1$ is the same with that of $\bar{\text{CI}}_\alpha^1$. We also have the following property of the expected length of confidence intervals,

$$
\mathbb{E}L\left(\text{CI}_\alpha^1(Z)\right) \leq \mathbb{E}L\left(\bar{\text{CI}}_\alpha^1(Z)\right).
$$
On the event $\tilde{G}_8$, we have $\psi(Z) = \frac{1}{n^2} \| y^{(2)} - X^{(2)} \beta L \|_2^2$. Since

$$\mathbb{P}_\theta \left( \chi^2_2 (n_2) \leq \frac{\| y^{(2)} - X^{(2)} \beta L \|_2^2}{\| \beta - \beta L \|_2^2 + \sigma_0^2} \leq \chi^2_{1 - \frac{\alpha}{2}} (n_2) \right) \geq 1 - \alpha,$$

we have

$$\mathbb{P}_\theta \left( \| \beta L - \beta \|_2^2 \in \mathcal{C}_1^1 (Z) \right) \geq \mathbb{P}_\theta \left( \left\{ \chi^2_2 (n_2) \leq \frac{\| y^{(2)} - X^{(2)} \beta L \|_2^2}{\| \beta - \beta L \|_2^2 + \sigma_0^2} \leq \chi^2_{1 - \frac{\alpha}{2}} (n_2) \right\} \cap \tilde{G}_8 \right) \geq \mathbb{P}_\theta \left( \left\{ \chi^2_2 (n_2) \leq \frac{\| y^{(2)} - X^{(2)} \beta L \|_2^2}{\| \beta - \beta L \|_2^2 + \sigma_0^2} \leq \chi^2_{1 - \frac{\alpha}{2}} (n_2) \right\} \right) + \mathbb{P}_\theta (\tilde{G}_8) - 1 \geq 1 - \alpha - \exp(-c'n) - cp^{-c}.$$

The coverage property (3.10) in the main paper [3] over $\Theta_0(k)$ follows from taking infimum over both sides of above inequality. In the following, we are going to control the length of the confidence interval $\mathcal{C}_1^1$ over $\Theta_0(k)$. Note that

$$\frac{\psi(Z)}{\frac{1}{n^2} \chi^2_2 (n_2)} - \frac{\psi(Z)}{\frac{1}{n^2} \chi^2_{1 - \frac{\alpha}{2}} (n_2)} = \psi(Z) f(n_2),$$

where $f(n_2) = \left( \frac{1}{n^2} \chi^2_2 (n_2) - \frac{1}{n^2} \chi^2_{1 - \frac{\alpha}{2}} (n_2) \right)$. On the event $\tilde{G}_8$, we have $\psi(Z) = \frac{1}{n^2} \| y^{(2)} - X^{(2)} \beta L \|_2^2$ and then obtain

$$\mathbb{E} \left| \frac{\psi(Z)}{\frac{1}{n^2} \chi^2_2 (n_2)} - \frac{\psi(Z)}{\frac{1}{n^2} \chi^2_{1 - \frac{\alpha}{2}} (n_2)} \right| \mathbf{1}_{\tilde{G}_6 \cap \tilde{G}_8} = \mathbb{E} \left( \frac{1}{n^2} \| y^{(2)} - X^{(2)} \beta L \|_2^2 \mathbf{1}_{\tilde{G}_6 \cap \tilde{G}_8} \right) f(n_2) \leq \left( \mathbb{E} \frac{1}{n^2} \| y^{(2)} - X^{(2)} \beta L \|_2^2 \mathbf{1}_{\tilde{G}_6} \right) f(n_2).$$

Conditioning on $(X^{(1)}, y^{(1)})$, by taking expectation of right hand side of above equation with respect to $(X^{(2)}, y^{(2)})$, the right hand side is equal to $\mathbb{E} \left( \| \beta - \beta L \|_2^2 + \sigma_0^2 \right) 1_{\tilde{G}_6} f(n_2) \leq \left( C \frac{\log p}{n} + 1 \right) \delta_0^2 f(n_2)$. Based on Lemma 10, we have $1 - 2 \sqrt{\frac{2 \log \frac{1}{n}}{n}} - 4 \frac{\log \frac{1}{n}}{n} \leq \frac{1}{n^2} \chi^2_2 (n_2) \leq \frac{1}{n^2} \chi^2_{1 - \frac{\alpha}{2}} (n_2) \leq 1 + 2 \sqrt{\frac{2 \log \frac{1}{n}}{n}} + 4 \frac{\log \frac{1}{n}}{n}$ and hence $f(n_2) \leq \frac{C}{\sqrt{n}}$. Hence, we have

$$\mathbb{E} \left| \frac{\psi(Z)}{\frac{1}{n^2} \chi^2_2 (n_2)} - \frac{\psi(Z)}{\frac{1}{n^2} \chi^2_{1 - \frac{\alpha}{2}} (n_2)} \right| \mathbf{1}_{\tilde{G}_6 \cap \tilde{G}_8} \leq \frac{C}{\sqrt{n}} \left( C \frac{\log p}{n} + 1 \right) \delta_0^2.$$
Note that

\[
E \left| \frac{\psi(Z)}{\frac{1}{n_2} \chi^2_\alpha(n_2)} - \frac{\psi(Z)}{\frac{1}{n_2} \chi^2_{1 - \frac{\alpha}{2}}(n_2)} \right| \mathbf{1}_{(\tilde{G}_6 \cap \tilde{G}_8)^c} \leq \sigma_0^2 \log p \times f(n_2) \mathbb{P}_\theta ((\tilde{G}_6 \cap \tilde{G}_8)^c) \\
\leq \frac{1}{\sqrt{n}} \sigma_0^2 \log p (c p^{-c} + c \exp (-c'n)) .
\]

Combined with (4.18), we establish (3.11) in the main paper [3].

4.6. Proof of Proposition 4. We first introduce the following lemma, which establishes an upper bound for \( \|a\|_q^q \) where \( 1 \leq q \leq 2 \) and \( a \in \mathbb{R}^p \).

**Lemma 11.**

\[
\|a\|_q^q \leq \left( \sum_{j=1}^{p} |a_j| \right)^{2-q} \left( \sum_{j=1}^{p} a_j^2 \right)^{q-1}.
\]

The above lemma is established in the proof of Theorem 7.1 in [1]. We introduce the following confidence intervals,

\[
\text{CI}_\alpha^2(Z, k_2, q) = \left( \frac{\psi(Z)}{\frac{1}{n_2} \chi^2_{1 - \frac{\alpha}{2}}(n_2)} - \sigma_0^2, (16k_2)^{\frac{q}{2}-1} \left( \frac{\psi(Z)}{\frac{1}{n_2} \chi^2_\alpha(n_2)} - \sigma_0^2 \right) \right).
\]

Since the loss \( \|\hat{\beta} - \beta\|_q^2 \) is positive, the coverage property of \( \text{CI}_\alpha^2 \) is the same with that of \( \text{CI}_\alpha^2 \). We also have the following property of the expected length of confidence intervals,

\[
\text{EL} (\text{CI}_\alpha^2(Z, k_2, q)) \leq \text{EL} (\text{CI}_\alpha^2(Z, k_2, q)).
\]

Define the events

\[
\tilde{G}_6' = \left\{ \|\hat{\beta}^L - \beta\|_2^2 \leq C_2^* (A, k_1) k_1 \frac{\log p}{n} \right\},
\]

\[
\tilde{G}_8' = \left\{ \frac{1}{n_2} \left\| y^{(2)} - X^{(2)} \hat{\beta}^L \right\|_2^2 \leq \sigma_0^2 \left( 1 + \frac{k_1 \log p}{n_1} \right) \left( 1 + 2 \sqrt{\frac{\log p}{n_2} + \frac{2 \log p}{n_2}} \right) \right\},
\]

and similar to Lemma 9, we have

\[
\min_{\theta \in \Theta_0(k_1)} \mathbb{P}_\theta (\tilde{G}_6' \cap \tilde{G}_8') \geq 1 - c \exp (-c'n) - c p^{1 - \frac{A^2}{8} - c p^{-2}}.
\]
Note that

\[
\|\hat{\beta}^L - \beta\|_2^2 \leq \|\hat{\beta}^L - \beta\|_q^2, \quad \text{for } 1 \leq q < 2.
\]

For the Lasso estimator $\hat{\beta}^L$ with $A > 4\sqrt{2}$, let $S$ denote the support of $\beta$, then we have

\[
\|\hat{\beta}^L - \beta\|_1^2 \leq 16\|\hat{\beta}^L - \beta\|_2^2 \leq 16k_2\|\hat{\beta}^L - \beta\|_2^2.
\]

By Lemma 11, we have $\|\hat{\beta}^L - \beta\|_q^2 \leq (16k_2)^{\frac{q}{2}-1}\|\hat{\beta}^L - \beta\|_2^2$. Combined with (4.17) and (4.23), we establish the coverage property (3.20) in the main paper [3]. In the following, we control the length of $\hat{C}_1^2(Z, k_2, q)$ over $\Theta_0(k_1)$. We decompose the length as

\[
L \left( \hat{C}_1^2(Z, k_2, q) \right) = L \left( \hat{C}_1^1(Z) \right) + \left( (16k_2)^{\frac{2}{q}-1} - 1 \right) \left( \frac{\psi(Z)}{\frac{1}{n_2} \chi^2_2(n_2)} - \sigma_0^2 \right).
\]

By (3.11) in the main paper [3] and (4.21), it is sufficient to control the term $E \left[ \frac{\psi(Z)}{\frac{1}{n_2} \chi^2_2(n_2)} - \sigma_0^2 \right]$. On the event $G'_6$, we have $\psi(Z) = \frac{1}{n_2} \|y^{(2)} - X^{(2)} \hat{\beta}^L\|_2^2$ and

\[
\frac{\psi(Z)}{\frac{1}{n_2} \chi^2_2(n_2)} - \sigma_0^2 = \frac{\frac{1}{n_2} \|y^{(2)} - X^{(2)} \hat{\beta}^L\|_2^2 - \sigma_0^2 - \left( \frac{1}{n_2} \chi^2_2(n_2) - 1 \right) \sigma_0^2}{\frac{1}{n_2} \chi^2_2(n_2)} = \frac{\frac{1}{n_2} \|y^{(2)} - X^{(2)} \hat{\beta}^L\|_2^2 - \sigma_0^2 - \|\hat{\beta}^L - \beta\|_2^2}{\frac{1}{n_2} \chi^2_2(n_2)} - \left( \frac{1}{n_2} \chi^2_2(n_2) - 1 \right) \sigma_0^2 - \|\hat{\beta}^L - \beta\|_2^2.
\]

Hence, we have

\[
E \left[ \frac{\psi(Z)}{\frac{1}{n_2} \chi^2_2(n_2)} - \sigma_0^2 \right] 1_{G'_6} \leq E \left[ \frac{\frac{1}{n_2} \|y^{(2)} - X^{(2)} \hat{\beta}^L\|_2^2 - \sigma_0^2 - \|\hat{\beta}^L - \beta\|_2^2}{\frac{1}{n_2} \chi^2_2(n_2)} \right] 1_{G'_6} + E \left[ \frac{\frac{1}{n_2} \chi^2_2(n_2) - 1}{\frac{1}{n_2} \chi^2_2(n_2)} \right] 1_{G'_6} \leq C \left( \frac{1}{\sqrt{n}} + k_1 \frac{\log p}{n} \right) \sigma_0^2,
\]
and (4.28)
\[
\frac{\psi(Z)}{\sqrt{n_2} \chi_2^2/2(n_2)} - \sigma_0^2 \leq \sigma_0^2 \log p \mathbb{P}(s_0 < (\bar{G}_6' \cap \bar{G}_8'))
\]
\[
\leq C \sigma_0^2 \log p \left( c p^{-1/4} + c p^{-2} + c \exp \left(-c'n\right) \right) \leq C \left( \frac{k_1 \log p}{n} + \frac{1}{\sqrt{n}} \right) \sigma_0^2,
\]
where the last inequality follows from the equality (4.22) and the fact that \( A > 4 \sqrt{2} \) and \( n \leq p \). The control of length (3.21) in the main paper [3] follows from (4.25), (4.27) and (4.28).

4.7. Proof of Proposition 6. We will restrict to the estimators \( \hat{\beta} \) satisfying Assumption (A) introduced in (5.1) in the main paper [3]. The minimax lower bounds (2.9) in Theorem 1, (2.13) in Theorem 2 and the minimax lower bound \( \frac{k \log p}{n} \sigma_0^2 \) of (2.8) in Theorem 1 in the main paper [3] can be achieved by the trivial estimator 0.

For estimators \( \hat{\beta} \) constructed using the subsample \( Z^{(1)} = (y^{(1)}, X^{(1)}) \), the minimax lower bound \( \frac{1}{\sqrt{n}} \sigma_0^2 \) of (2.8) in Theorem 1 in the main paper [3] can be achieved by the estimator as defined in (2.11) in the main paper [3],
\[
\bar{L}_2 = \left( \frac{1}{n_2} \left\| Y^{(2)} - X^{(2)} \hat{\beta} \right\|^2_2 - \sigma_0^2 \right)^{+}.
\]
Applying the proof of Proposition 1, we can establish, for any sequence \( \delta_{n,p} \to \infty \),
\[
\limsup_{n,p \to \infty} \sup_{\theta \in \Theta_0(k)} \mathbb{P}_\theta \left( \left| \bar{L}_2 - \| \hat{\beta} - \beta \|^2_2 \right| \geq \delta_{n,p} \frac{1}{\sqrt{n}} \right) = 0.
\]

The minimax lower bound \( \frac{k \log p}{n} \sigma_0^2 \) in (3.5) in Theorem 3 in the main paper [3] can be achieved by the confidence interval \( (0, C^* \frac{k \log p}{n} \sigma_0^2) \). The minimax lower bound \( \frac{k_2 \log p}{n} \sigma_0^2 \) of (3.7) in Theorem 4 in the main paper [3] can be achieved by the confidence interval \( (0, C^* \frac{k_2 \log p}{n} \sigma_0^2) \); for estimators \( \hat{\beta} \) constructed using the subsample \( Z^{(1)} = (y^{(1)}, X^{(1)}) \), the minimax lower bound \( \frac{1}{\sqrt{n}} \sigma_0^2 \) of (3.5) in Theorem 3 in the main paper [3] and (3.7) in Theorem 4 in the main paper [3] can be achieved by the confidence interval \( \text{CI}^1_\alpha(Z) \) as defined in (3.8) in the main paper [3] with \( \psi(Z) = \min \left\{ \frac{1}{n_2} \left\| Y^{(2)} - X^{(2)} \hat{\beta} \right\|^2_2, \sigma_0^2 \log p \right\} \). Applying the proof of Proposition 2, we can establish that the constructed confidence interval satisfies
\[
\liminf_{n,p \to \infty} \inf_{\theta \in \Theta_0(k)} \mathbb{P}_{\theta} \left( \| \hat{\beta}^L - \beta \|^2_2 \in \text{CI}^1_\alpha(Z) \right) \geq 1 - \alpha,
\]
and

\[ L(\text{CI}_\alpha^1(Z), \Theta_0(k)) \lesssim \frac{1}{\sqrt{n}} \sigma_0^2. \]

The minimax lower bounds (3.13) in Theorem 5 in the main paper [3] can be achieved by the confidence interval \((0, C^*k_2^2 \frac{\log p}{n}\sigma_0^2)\). The minimax lower bound \(ck_2^2 \frac{\log p}{n}\sigma_0^2\) of (3.18) in Theorem 6 in the main paper [3] can be achieved by the confidence interval \((0, C^*k_2^2 \frac{\log p}{n}\sigma_0^2)\). For estimators \(\hat{\beta}\) constructed using the subsample \(Z^{(1)} = (y^{(1)}, X^{(1)})\) and satisfying the assumption \(\|\hat{\beta} - \beta\|_1 \leq c^*\|\hat{\beta} - \beta\|_1\) where \(S\) is the support of the true beta, the minimax lower bounds \(ck_2^2 \frac{\log p}{n}\sigma_0^2\) of (3.18) and \(ck_2^2 \frac{\log p}{n}\sigma_0^2\) of (3.18) in Theorem 6 in the main paper [3] can be achieved by the confidence interval \(\text{CI}_\alpha^2(Z, k_2, q)\) as defined in (3.19) in the main paper [3]

\[
\left(\frac{\psi(Z)}{n_2 \chi^2_{1-\alpha} (n_2)} - \sigma_0^2\right)^+, \left((1 + c^*)^2 k_2 \frac{\log p}{n}\sigma_0^2\right)^+ \right),
\]

with \(\psi(Z) = \min \left\{ \frac{1}{n_2} \left\| y^{(2)} - X^{(2)} \hat{\beta}\right\|_2^2 + \eta_0^2 \log p \right\} \). Applying the proof of Proposition 4, we can establish that

\[
\liminf_{n,p \to \infty} \inf_{\theta \in \Theta_0(k_2)} \mathbb{P}_\theta \left( \|\hat{\beta} - \beta\|_q^2 \in \text{CI}_\alpha^2(Z, k_2, q) \right) \geq 1 - \alpha,
\]

and

\[
L(\text{CI}_\alpha^2(Z, k_2, q), \Theta_0(k_1)) \lesssim k_2^{\frac{q}{2}} \left( k_1 \frac{\log p}{n} + \frac{1}{\sqrt{n}} \right) \sigma_0^2.
\]

The minimax lower bound (4.1) in Theorem 7 in the main paper [3] can be achieved by the confidence interval \(\text{CI}_\alpha(Z, k, q)\) as defined in (4.4) in the main paper [3] with \(\varphi(Z, k, q) = C^*k_2^2 \frac{\log p}{n}\sigma^2\).

4.8. Proof of Proposition 7. The proof is a generalization of that of Proposition 2. We also define the following extra event to facilitate the discussion.

\[
(4.29) \quad \mathcal{G}_0 = \left\{ \max \left\{ \frac{\lambda^2_{\min}(\hat{\Omega})}{\lambda^2_{\min}(\hat{\Omega}) - 1}, \frac{\lambda^2_{\max}(\hat{\Omega})}{\lambda^2_{\max}(\hat{\Omega}) - 1} \right\} \leq 0.01 \right\}.
\]

By Theorem 1 in [6], with a proper chosen tuning parameter, we have \(\mathbb{P}(\mathcal{G}_0) \geq 1 - p^{-2}\).
We first introduce the following confidence intervals,
\begin{equation}
\bar{CI}_\alpha^3 (Z) = \left(0.99\lambda_{\min}^2 (\hat{\Omega}) \left(\frac{\psi_1 (Z)}{\frac{1}{n_2} \chi^2_{1-\frac{1}{q}} (n_2)} - \sigma_0^2\right), 1.01\lambda_{\max}^2 (\hat{\Omega}) \left(\frac{\psi_2 (Z)}{\frac{1}{n_2} \chi^2_{1-\frac{1}{q}} (n_2)} - \sigma_0^2\right)\right).
\end{equation}

Since the loss \(\|\beta - \beta\|_2^2\) is positive, the coverage property of \(CI_\alpha^1\) is the same with that of \(CI_\alpha^3\). We also have the following property of the expected length of confidence intervals,
\begin{equation}
\text{EL} \left(CI_\alpha^3 (Z)\right) \leq \text{EL} \left(\bar{CI}_\alpha^3 (Z)\right).
\end{equation}

On the event \(\bar{\mathcal{G}}_8\), we have \(\psi (Z) = \frac{1}{n_2} \|y^{(2)} - X^{(2)} \beta\|_2^2\). On the event \(\mathcal{G}_0\), we have \(\lambda_{\max}^2 = \lambda_{\max}^2 (\hat{\Omega})\) and \(\lambda_{\min}^2 = \lambda_{\min}^2 (\hat{\Omega})\). Since
\begin{equation}
P_\theta \left(\chi^2_\frac{q}{2} (n_2) \leq \frac{\|y^{(2)} - X^{(2)} \beta\|_2^2}{\|\Sigma (\beta - \beta)\|_2^2 + \sigma_0^2} \leq \chi^2_\frac{1}{2} (n_2)\right) \geq 1 - \alpha,
\end{equation}
we have
\begin{align*}
P_\theta \left(\|\beta - \beta\|_2^2 \in CI_\alpha^3 (Z)\right) &\geq P_\theta \left(\left\{\chi^2_\frac{q}{2} (n_2) \leq \frac{\|y^{(2)} - X^{(2)} \beta\|_2^2}{\|\Sigma (\beta - \beta)\|_2^2 + \sigma_0^2} \leq \chi^2_\frac{1}{2} (n_2)\right\} \cap \bar{\mathcal{G}}_8 \cap \mathcal{G}_0\right) \\
&\geq P_\theta \left(\left\{\chi^2_\frac{q}{2} (n_2) \leq \frac{\|y^{(2)} - X^{(2)} \beta\|_2^2}{\|\Sigma (\beta - \beta)\|_2^2 + \sigma_0^2} \leq \chi^2_\frac{1}{2} (n_2)\right\}\right) + P_\theta \left(\bar{\mathcal{G}}_8 \cap \mathcal{G}_0\right) - 1 \geq 1 - \alpha - cp^{-c}.
\end{align*}

The coverage property (2.9) over \(\Theta_{k, s}\) follows from taking infimum over both sides of above inequality.

Combing (2.9) and the fact that
\begin{equation}
\|\beta - \beta\|_q^2 \leq \|\beta - \beta\|_2^2 \leq ((1 + c^*)^2 k_2)^{\frac{2}{2} - 1} \|\beta - \beta\|_q^2, \quad \text{for} \quad 1 \leq q < 2,
\end{equation}
we can establish the coverage property (2.11).

In the following, we are going to control the expected length of the confidence interval \(CI_\alpha^3\) and \(CI_\alpha^1\) over \(\Theta_{k, s}\). Applying the similar argument as (4.27) and (4.28), we have
\begin{align*}
\text{EL} \lambda_{\max}^2 \left(\frac{\psi (Z)}{\frac{1}{n_2} \chi^2_{\frac{q}{2}} (n_2)} - \sigma_0^2 \right) 1_{\mathcal{G}_0} &\leq M_1 \text{E} \left[\frac{1}{n_2} \|y^{(2)} - X^{(2)} \beta\|_2^2 - \sigma_0^2 - \|\beta - \beta\|_2^2\right] 1_{\mathcal{G}_0} \\
+ M_1 \text{E} \left(\frac{1}{n_2} \chi^2_{\frac{q}{2}} (n_2) - 1\right) \sigma_0^2 &\leq C \left(\frac{1}{\sqrt{n}} + k_1 \frac{\log p}{n}\right) \sigma_0^2,
\end{align*}
and
\[
\mathbf{E} \psi^2 \left\{ \frac{\psi (Z)}{\frac{\chi_2^2}{2} (n_2)} - \sigma_0^2 \right\} 1_{(\bar{G}_6' \cap \bar{G}_8' \cap \bar{G}_9')^c} \leq \sigma_0^2 \log p^2 \mathbb{P}_\theta ((\bar{G}_6' \cap \bar{G}_8')^c) \\
\leq C \sigma_0^2 \log p \left( cp^{-\delta} + cp^{-2} + c \exp \left( -c' n \right) \right) \leq C \left( \frac{k_1 \log p}{n} + \frac{1}{\sqrt{n}} \right) \sigma_0^2,
\]
where the last inequality follows from the equality (4.22) and the fact that \( \delta > 2 \) and \( n \leq p \). The above two inequalities lead to the control of length (2.10) and (2.12).

5. Proof of extra lemmas. In this section, we prove Lemma 1, 2, 3, 6, 8 and 9.

5.1. Proof of Lemma 1. The equality (9.2) in the main paper [3] follows from

\[
\mathbb{P}_{Z, \theta \sim \pi} (Z \in A) = \int \int 1_{z \in A} f_\theta (z) \pi (\theta) d\theta d\theta = \int 1_{z \in A} \left( \int f_\theta (z) \pi (\theta) d\theta \right) dz \\
= \int 1_{z \in A} f_\pi (z) dz = \mathbb{P}_\pi (A),
\]
where the second equality follows from the Fubini’s theorem. The inequality (9.3) in the main paper [3] follows from

\[
| \mathbb{P}_{\pi_1} (Z \in A) - \mathbb{P}_{\pi_2} (Z \in A) | = \left\| \int 1_{z \in A} f_{\pi_1} (z) dz - \int 1_{z \in A} f_{\pi_2} (z) dz \right\| \leq \int | f_{\pi_1} (z) - f_{\pi_2} (z) | dz,
\]
where the equality follows from the definition of \( \mathbb{P}_{\pi_i} \) and the inequality follows from the triangle inequality.

5.2. Proof of Lemma 3. In the following, we prove Lemma 3 in the main paper [3]. The following lemmas are useful in controlling the \( \chi^2 \) distance between the null and the alternative hypothesis. The first lemma is established in [7, 8].

**Lemma 12.** Let \( g_i \) be the density function of \( N(0, \Sigma_i) \) for \( i = 0, 1, 2, \) respectively. Then

\[
\int \frac{g_1 g_2}{g_0} = \left( \det \left( I - \Sigma_0^{-1} (\Sigma_1 - \Sigma_0) \Sigma_0^{-1} (\Sigma_2 - \Sigma_0) \right) \right)^{-\frac{1}{2}}.
\]
\textbf{Supplement.} Let $\delta$ and $\tilde{\delta}$ be independent random variables with the prior distribution $\pi$, then

\begin{equation}
\chi^2(f_\pi, f_0) + 1 = \int \left( \det \left( I - (\Sigma_0^z)^{-1} \left( \Sigma_{1,\delta}^z - \Sigma_0^z \right) (\Sigma_0^z)^{-1} (\Sigma_{1,\delta}^z - \Sigma_0^z) \right) \right)^{-\frac{n}{2}} \pi(\delta) \pi(\tilde{\delta}) \ d\delta d\tilde{\delta}.
\end{equation}

The proof of the above lemma can be found in [4]. We first introduce the covariance matrix of $(y_i, X_i)$ corresponding to the parameter $\theta_1$,

\begin{equation}
\Sigma_{1,\delta} = \begin{pmatrix}
\|\beta_1\|_2^2 + \|\delta\|_2^2 + \sigma_0^2 & (\beta_0^*)^T \\
\beta_0^* & 0_{1 \times k_0}
\end{pmatrix}
\end{equation}

where $\Sigma_{1,\delta}$ only depends on $\delta$ through the $\ell_2$ norm $\|\delta\|_2^2 = (k_2 - k_0)^2$. To control $\chi^2(f_\pi, f_\theta_1)$, we have the following expression for the main term of (5.1),

\begin{equation}
(\Sigma_{1,\delta}^{-1} (\Sigma_{1,\delta} - \Sigma_1^{-1}) (\Sigma_{1,\delta} - \Sigma_1^{-1}) (\Sigma_{1,\delta} - \Sigma_1^{-1}) = \begin{pmatrix}
\frac{1}{\|\delta\|_2^2 + \sigma_0^2} & \frac{\delta^T \tilde{\delta}}{\|\delta\|_2^2 + \sigma_0^2} \\
\frac{\delta^T \tilde{\delta}}{\|\delta\|_2^2 + \sigma_0^2} & 0_{k_0 \times k_0}
\end{pmatrix}
\end{equation}

By applying Lemma 13, we have

\begin{equation}
\chi^2(f_\pi, f_\theta_1) + 1 = E_{\delta, \tilde{\delta}} \left( 1 - \frac{\delta^T \tilde{\delta}}{\|\delta\|_2^2 + \sigma_0^2} \right)^{-\frac{n}{2}} \left( 1 - \frac{\delta^T \tilde{\delta}}{\|\delta\|_2^2 + \sigma_0^2} \right)^{-\frac{n}{2}}.
\end{equation}

To control $\chi^2(f_\theta_1, f_0)$, we have the following expression for the main term of (5.1),

\begin{equation}
(\Sigma_0^{-1} (\Sigma_{1,\delta} - \Sigma_0^{-1}) (\Sigma_{1,\delta} - \Sigma_0^{-1}) (\Sigma_{1,\delta} - \Sigma_0^{-1}) = \begin{pmatrix}
\|\delta\|_2^2 \|\delta\|_2^2 & \|\delta\|_2^2 \|\delta\|_2^2 \\
\|\delta\|_2^2 \|\delta\|_2^2 & 0_{k_0 \times k_0}
\end{pmatrix}
\end{equation}

By applying Lemma 13, we have

\begin{equation}
\chi^2(f_\theta_1, f_0) + 1 = E_{\delta, \tilde{\delta}} \left( 1 - \frac{\|\delta\|_2^2 \|\delta\|_2^2}{\sigma_0^4} \right)^{-\frac{n}{2}}.
\end{equation}
5.3. Proof of Lemma 8. By the normalization (4.1) in the main paper [3], the scaled Lasso algorithm can be expressed as

\[(5.3) \quad \{\hat{d}, \hat{\sigma}\} = \arg \min_{d \in \mathbb{R}^p, \sigma \in \mathbb{R}} \frac{\|y - Wd\|_2^2}{2n\sigma} + \frac{\sigma}{2} + \lambda_0 \sum_{j=1}^p |d_j|.\]

The following lemma from an arXiv version of [8] is useful to control the estimation of the noise level \(\sigma\).

**Lemma 14.** Let \(\{\hat{d}, \hat{\sigma}\}\) be the solution of the scaled Lasso (5.3). For any \(\epsilon_0 > 1\), on the event \(\mathcal{S}_1 = \left\{ \frac{\|W\|^\infty}{n} \leq \sigma \epsilon_0 \lambda_0 \frac{\epsilon_0 - 1}{\epsilon_0 + 1} (1 - \tau) \right\}\), we have

\[(5.4) \quad \left| \frac{\hat{\sigma}}{\sigma \epsilon_0} - 1 \right| \leq \tau.\]

where \(\tau\) is defined in (4.3).

For fixed \(\mu\), we also define \(\hat{d}(\mu) = \arg \min_{d \in \mathbb{R}^p} \frac{\|y - Wd\|_2^2}{2n} + \mu \sum_{j=1}^p |d_j|\). Note that

\[(5.5) \quad \hat{d} = \hat{d}(\lambda_0 \hat{\sigma}) \quad \text{and} \quad \hat{d}_j = \hat{\beta}_{j}^{\text{SL}} \|X_j\|_2 \sqrt{n} \quad \text{for} \quad j \in [p].\]

Setting \(A > 2\sqrt{2}\), on event \(\mathcal{G} \cap \mathcal{S}\), we have the following events

\[(5.6) \quad \mathcal{B}_1 = \left\{ \frac{1}{n}\|W^TW (\hat{d} - d)\|_\infty \leq \lambda_0 \hat{\sigma} \right\} \quad \text{and} \quad \mathcal{B}_2 = \left\{ 2 \sqrt{\frac{2 \log p}{n}} \sigma_0 \leq \lambda_0 \hat{\sigma} \right\}.\]

Based on the proof of Theorem 7.2 in [1], on the event \(\mathcal{G} \cap \mathcal{S}\), we have

\[\|\hat{\beta}^{\text{SL}} - \beta\|_2^2 \leq \frac{\max \|X_j\|_2^2}{2n} \left( \frac{16A\sigma}{\kappa^2(W, k, k, 3)} \right) \left( \frac{1}{\kappa^2(W, k, k, 3)} \right) \frac{2 \log p}{n}.\]

Similar to the proof of Lemma 13 in [4], we have

\[(5.7) \quad \kappa^2 (W, k, k, 3) \geq \frac{n}{\max \|X_j\|_2^2} \kappa^2 \left( X, k, k, 3 \left( \frac{\max \|X_j\|_2}{\min \|X_j\|_2} \right) \right),\]

and then establish (4.6). By the definitions of \(\mathcal{G}_1\) and \(\mathcal{G}_3\), we establish (4.7).
5.4. Proof of Lemma 9. The control of events $\bar{G}_1, \bar{G}_2, \bar{G}_3$ and $\bar{G}_4$ is similar to that of Lemma 7 and will be omitted here. We will present the proofs of $\mathbb{P}_\theta(\bar{G}_5)$ and $\mathbb{P}_\theta(\bar{G}_6)$. We establish $\mathbb{P}_\theta(\bar{G}_5) \geq \mathbb{P}_\theta(\bar{G}_1 \cap \bar{G}_2 \cap \bar{G}_3)$ by showing that $\bar{G}_5$ holds on the event $\bar{G}_1 \cap \bar{G}_2 \cap \bar{G}_3$. Based on the proof of Theorem 7.2 in [1] and (5.7), on the event $\bar{G}_1 \cap \bar{G}_2 \cap \bar{G}_3$, we have

\[ \|\hat{\beta}^L - \beta\|^2 \leq \frac{(22A\sigma_0)^2}{\left(\frac{1}{4} - 42\sqrt{\frac{2k\log p}{n^2}}\right)^4 k^4 \log p}. \]

For the case $q = 2$, we can establish $\|\hat{\beta}^L - \beta\|^2_2$ for the case $A > \sqrt{2}$ by applying the finer results established in [10]. By Theorem 3 and (27) in [10], on the event $\bar{G}_1 \cap \bar{G}_2 \cap \bar{G}_4$, we have

\[ \|\hat{\beta}^L - \beta\|^2_2 \leq \frac{\left(\frac{30n}{w_0 + 1}A\sigma_0\right)^2}{\left(\frac{1}{4} - (9 + 11\eta_0)\sqrt{\frac{2k\log p}{n^2}}\right)^4 k^4 \log p}. \]

Hence, $\mathbb{P}_\theta(\bar{G}_6) \geq \mathbb{P}_\theta(\bar{G}_1 \cap \bar{G}_2 \cap \bar{G}_4)$. Let $\mathbb{P}_\theta(\cdot|(X^{(1)}, y^{(1)})|$ denote the conditional probability of $(X^{(2)}, y^{(2)})$ on $(X^{(1)}, y^{(1)})$. Note that conditioning on $(y^{(1)}, X^{(1)})$, we have

\[ y_i^{(2)} - X_i^{(2)}\hat{\beta}^L = X_i^{(2)}(\beta - \hat{\beta}^L) + \epsilon_i^{(2)} \sim N\left(0, \|\beta - \hat{\beta}^L\|^2_2 + \sigma_0^2\right) \]

and $\|y^{(2)} - X^{(2)}\hat{\beta}^L\|^2_2 \sim \left(\|\beta - \hat{\beta}^L\|^2_2 + \sigma_0^2\right) \chi^2(n_2)$. By Lemma 10, we have

$\mathbb{P}_\theta(\bar{G}_7|(X^{(1)}, y^{(1)})|) \geq 1 - cp^{-2}$ and hence $\mathbb{P}_\theta(\bar{G}_7) \geq 1 - cp^{-2}$. The control of $\mathbb{P}_\theta(\bar{G}_8)$ follows from the fact $\mathbb{P}_\theta(\bar{G}_8) \geq \mathbb{P}_\theta(\bar{G}_7 \cap \bar{G}_6) \geq (1 - cp^{-2})\mathbb{P}_\theta(\bar{G}_6)$.

5.5. Proof of Lemma 2 and 6. The proof of Lemma 2 follows from (4.9) and (4.10) of Lemma 9 by taking $k = \|\beta^*\|_0$. The proof of Lemma 6 follows from Lemma 7 and 8 by taking $k = \|\beta^*\|_0$.

References.


Department of Statistics
The Wharton School
University of Pennsylvania
Philadelphia, Pennsylvania 19104
USA
E-mail: tcai@wharton.upenn.edu
ziujuo@wharton.upenn.edu
URL: http://www-stat.wharton.upenn.edu/~tcai/