

Article

# Minimax Rates of $\ell_p$ -Losses for High-Dimensional Linear Errors-in-Variables Models over $\ell_q$ -Balls

Xin Li <sup>1</sup> and Dongya Wu <sup>2,\*</sup>

<sup>1</sup> School of Mathematics, Northwest University, Xi'an 710069, China; lixin@nwu.edu.cn

<sup>2</sup> School of Information Science and Technology, Northwest University, Xi'an 710069, China

\* Correspondence: wudongya@nwu.edu.cn

**Abstract:** In this paper, the high-dimensional linear regression model is considered, where the covariates are measured with additive noise. Different from most of the other methods, which are based on the assumption that the true covariates are fully obtained, results in this paper only require that the corrupted covariate matrix is observed. Then, by the application of information theory, the minimax rates of convergence for estimation are investigated in terms of the  $\ell_p$  ( $1 \leq p < \infty$ )-losses under the general sparsity assumption on the underlying regression parameter and some regularity conditions on the observed covariate matrix. The established lower and upper bounds on minimax risks agree up to constant factors when  $p = 2$ , which together provide the information-theoretic limits of estimating a sparse vector in the high-dimensional linear errors-in-variables model. An estimator for the underlying parameter is also proposed and shown to be minimax optimal in the  $\ell_2$ -loss.

**Keywords:** sparse linear regression; errors-in-variables model; minimax rate; Kullback–Leibler divergence; information-theoretic limitations



**Citation:** Li, X.; Wu, D. Minimax Rates of  $\ell_p$ -Losses for High-Dimensional Linear Errors-in-Variables Models over  $\ell_q$ -Balls. *Entropy* **2021**, *23*, 722. <https://doi.org/10.3390/e23060722>

Academic Editor: Leandro Pardo

Received: 23 April 2021

Accepted: 4 June 2021

Published: 5 June 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In various fields of applied sciences and engineering, such as machine learning [1], a fundamental problem is to estimate an underlying parameter  $\beta^* \in \mathbb{R}^d$  of a linear regression model as follows

$$y_i = \langle X_{i\cdot}, \beta^* \rangle + e_i, \quad \text{for } i = 1, 2, \dots, n, \quad (1)$$

where  $\{(X_{i\cdot}, y_i)\}_{i=1}^n$  are i.i.d. observations ( $X_{i\cdot} \in \mathbb{R}^d$ ) and  $e \in \mathbb{R}^n$  is the random noise. In matrix form, model (1) can be written as  $y = X\beta^* + e$ , where  $X = (X_{1\cdot}, \dots, X_{n\cdot})^\top \in \mathbb{R}^{n \times d}$  and  $y, e \in \mathbb{R}^n$ . The covariates  $X_{i\cdot}$  ( $i = 1, 2, \dots, n$ ) are always assumed to be fully observed in standard formulations. However, this assumption is far away from reality since, in general, the measurement error cannot be avoided. In many real-world applications, due to the lack of observation or the instrumental constraint, the collected data, such as remote sensing data, may always be perturbed and tend to be noisy [2]. It has been shown in [3] that misleading inference results may be obtained if the method for clean data is applied to the noisy data naively. Therefore, it is more realistic to explore the case where only the corrupted covariates of the corresponding true covariates  $X_{i\cdot}$ 's are obtained; see, e.g., [4]. This is known as the measurement error model in the literature.

Estimation in the presence of measurement errors has attracted a lot of interest for a long time. Bickel and Ritov [5] first studied the linear measurement error models and proposed an efficient estimator. Then, Stefanski and Carroll [6] investigated the generalized linear measurement error models and constructed consistent estimators. Extensive results have also been established on parameter estimation and variable selection for both parametric or nonparametric settings; see [7,8] and references therein. It should be noted that these results are only applicable to classical low-dimensional (i.e.,  $n \geq d$ ) statistical models.

In the past two decades, high-dimensional statistical models, where the number of observations is much less than the number of predictors (i.e.,  $n \ll d$ ), have been paid much

attention and have achieved fruitful results in a wide range of research areas; see [9,10] for a detailed review. Most of the existing results are only suitable for models with clean data, while some researchers have begun to focus on the measurement error case. For example, Loh and Wainwright studied the high-dimensional sparse linear regression model with corrupted covariates. Though the proposed estimator involves solving a nonconvex optimization problem, they proved that the global and stationary points are statistically consistent; see [11,12]. Datta and Zou [13] proposed the Convex Conditioned Lasso (CoCoLasso), which enjoys the convex benefit of the Lasso in both estimation and algorithm and can handle a class of corrupted datasets, including the cases of additive or multiplicative measurement error. Li et al. [14] investigated a general nonconvex estimation method from statistical and computational aspects and the results can be immediately applied to the corrupted errors-in-variables linear regression.

Apart from the study on statistical convergence rates and designing efficient algorithms to solve certain estimators, it is also fundamental to information-theoretic limitations of statistical inference to understand the computationally efficient procedures. Such fundamental limits are usually studied by virtue of the minimax rates, which aim to find an estimator that minimizes the worst-case loss and, thus, can reveal gaps between the performance of some computationally efficient algorithm and that of an optimal algorithm. The minimax rate is always analyzed from two aspects, namely the informational lower bounds and statistical upper bounds. In the information-theoretic aspect, the Kullback–Leibler (KL) divergence is always used to provide lower bounds [15]. Recently, in [16], Loh provides a detailed review of a variety of techniques utilized to derive information-theoretic lower bounds for minimax estimation and learning, focusing on the problem settings with community recovery, parameter and function estimation, and online learning for multi-armed bandits. In the statistical aspect, a special estimator is always constructed to derive upper bounds; see, e.g., [17,18]. For the high-dimensional linear regression with additive errors, Loh and Wainwright [19] established minimax rates of convergence for estimating the unknown parameter in the  $\ell_2$ -loss. The proposed estimator was also shown to be minimax optimal in the additive error case under the  $\ell_2$ -loss, assuming that the true parameter is exact sparse, that is,  $\beta^*$  has at most  $s \ll d$  nonzero elements, which is also known as the exact sparsity assumption.

However, this exact assumption may be sometimes too restrictive to be satisfied in some real applications. For example, in the field of image processing, it is a standard phenomenon that wavelet coefficients for images usually exhibit an exponential decay, but do not need to be almost 0 (see, e.g., [20]). Other applications under high-dimensional scenarios include compressed sensing [21], genomic analysis [22], signal processing [23], and so on, where it is not suitable to impose an exact sparsity assumption on the underlying parameter. Hence, it is necessary to investigate minimax rates of estimation when the exact sparse assumption does not hold.

Our main purpose in the present study is to investigate the more general situation that coefficients of the true parameter are not almost zeros and then provide minimax rates of convergence for estimation in sparse linear regression with additive errors. More precisely speaking, we assume that for  $q \in [0, 1]$  fixed, the  $\ell_q$ -norm of  $\beta^*$  defined as  $\|\beta^*\|_q := (\sum_{j=1}^p |\beta_j^*|^q)^{1/q}$  is bounded from above. Note that this assumption is reduced to the exact sparsity assumption when  $q = 0$ . When  $q \in (0, 1]$ , this type of sparsity is known as the soft sparsity. The exact sparsity assumption has been widely used for statistical inference, while the soft sparsity assumption attracts relatively little attention apart from the work [24–26]. Specifically, under both exact and soft sparsity assumptions, Raskutti et al. [24] and Ye and Zhang [26] provided minimax rates of convergence for estimation in high-dimensional linear regression, respectively; Wang et al. [25] developed the optimal rates of convergence and proposed an adaptive  $\ell_q$ -aggregation strategy via model mixing which attains the established optimal rate automatically. It is worth noting that results in [24–26] are all obtained for clean data and cannot be applied to the errors-in-variables model. This is a fundamental difference from our present study.

The main contributions of this paper are as follows. By assuming that the regression parameter is of soft sparsity, in the information-theoretic aspect we establish lower bounds on the minimax risks for  $\ell_p$  ( $1 \leq p < \infty$ )-losses by virtue of the mutual information which hold for any arbitrary estimator for the model regardless of the specific method. In the statistical aspect, we propose an estimator which can be solved efficiently and then provide upper bounds on the  $\ell_2$ -loss between the estimator and the true parameter. Moreover, the lower and upper bounds when  $p = 2$  agree up to constant factors, implying that proposed estimator is minimax optimal in the  $\ell_2$ -loss.

The remainder of this paper is organized as follows. In Section 2, we provide background on the errors-in-variables linear regression model and some regularity conditions on the observed covariate matrix. In Section 3, we establish our main results on lower and upper bounds on minimax risks for  $\ell_p$  ( $1 \leq p < \infty$ )-losses over  $\ell_q$ -balls. Conclusions and future work are discussed in Section 4.

We end this section by introducing some notations for future reference. We use Greek lowercase letter  $\beta$  to denote the vectors. All vectors are column vectors following classical mathematical convention. A vector  $\beta$  is supported on  $S$  if and only if  $S = \{i \in \{1, 2, \dots, d\} : \beta_i \neq 0\}$ , and  $S$  is the support of  $\beta$  denoted by  $\text{supp}(\beta)$ , namely  $\text{supp}(\beta) = S$ . For  $d \geq 1$ , let  $\mathbb{I}_d$  stand for the  $d \times d$  identity matrix. For a matrix  $X \in \mathbb{R}^{n \times d}$ , let  $X_{ij}$  ( $i = 1, \dots, n, j = 1, 2, \dots, d$ ) denote its  $ij$ -th entry,  $X_i$  ( $i = 1, \dots, n$ ) denote its  $i$ -th row,  $X_{\cdot j}$  ( $j = 1, 2, \dots, d$ ) denote its  $j$ -th column.

## 2. Problem Setup

In this section, we begin with a precise formulation of the problem and then impose some regularity assumptions on the observed matrix.

Recall the standard linear regression model (1). One of the main types of measurement errors is the additive error. Specifically, for each  $i = 1, 2, \dots, n$ , we observe  $Z_i = X_i + W_i$ , where  $W_i \in \mathbb{R}^d$  is a random vector independent of  $X_i$  with mean 0 and known covariance matrix  $\Sigma_w$ . When the noise covariance  $\Sigma_w$  is unknown, there are some method to estimate it from the observed data; see, e.g., [4]. For example, a simple method is to estimate  $\Sigma_w$  from blank independent observations of the noise. Specifically, suppose that one independently observes a matrix  $W_0 \in \mathbb{R}^{n \times d}$  with  $n$  i.i.d. vectors of noise. Then we use  $\Sigma_w = \frac{1}{n} W_0^\top W_0$  as the estimate of  $\Sigma_w$ . Some other sophisticated variant of this method in are also provided in [4].

Throughout this paper, we assume that for  $i = 1, 2, \dots, n$ , the vectors  $X_i$ ,  $W_i$ , and  $e$  are Gaussian with mean 0 and covariance matrices  $\sigma_x^2 \mathbb{I}_d$  ( $\sigma_x > 0$ ),  $\sigma_w^2 \mathbb{I}_d$ , and  $\sigma_e^2 \mathbb{I}_n$ , respectively, and we write  $\sigma_z^2 = \sigma_x^2 + \sigma_w^2$  for simplicity.

According to the previous works of [11,12], we fix  $i \in \{1, 2, \dots, n\}$  and write  $\Sigma_x$  to denote the covariance matrix of  $X_i$  (i.e.,  $\Sigma_x = \text{cov}(X_i) = \sigma_x^2 \mathbb{I}_d$ ). Let  $(\hat{\Gamma}, \hat{Y})$  stand for the estimators for  $(\Sigma_x, \Sigma_x \beta^*)$  which only depend on the observed data  $\{(Z_i, y_i)\}_{i=1}^n$ . As has been discussed in [11], an unbiased and suitable choice of the surrogate pair  $(\hat{\Gamma}, \hat{Y})$  for the additive error case is given by

$$\hat{\Gamma} := \frac{Z^\top Z}{n} - \Sigma_w \quad \text{and} \quad \hat{Y} := \frac{Z^\top y}{n}.$$

Under the high-dimensional scenario ( $n \ll d$ ), the matrix  $\hat{\Gamma}$ , which is the estimator of  $\Sigma_x$  in the corrupted case, is always negative definite. To be specific, the matrix  $Z^\top Z$  has rank at most  $n$ , and then the positive definite matrices  $\Sigma_w$  are subtracted to obtain  $\hat{\Gamma}$ . Consequently,  $\hat{\Gamma}$  cannot be guaranteed to be positive definite regardless of the amount of noise. However, this does not affect the current result. Particularly, though the negative definiteness of  $\hat{\Gamma}$  leads to a nonconvex optimization problem in estimating  $\beta^*$  (cf. (14)) as well as the upper bound, a weaker condition (cf. Assumption 2) allows further analysis.

Instead of assuming the regression parameter  $\beta^*$  is exact sparse (i.e.,  $\text{supp}(\beta) \ll d$ ), we use a general notion to characterize the sparsity of  $\beta^*$ . Specifically, we assume that for  $q \in [0, 1]$ , and a radius  $R_q > 0$ ,  $\beta^* \in \mathbb{B}_q(R_q)$ , where

$$\mathbb{B}_q(R_q) := \{\beta \in \mathbb{R}^d : \|\beta\|_q^q = \sum_{j=1}^d |\beta_j|^q \leq R_q\}.$$

The use of  $\ell_q$ -ball is a common and popular way to measure the degree of sparsity (accurately, the above sets are not real “balls”, as they fail to be convex when  $q \in [0, 1)$ ). Note that  $\beta \in \mathbb{B}_0(R_0)$  corresponds to the case that  $\beta$  is exact sparse, while for  $q \in (0, 1]$ ,  $\beta \in \mathbb{B}_q(R_q)$  corresponds to the case of weak sparsity, which endows a certain decay rate on the ordered entries of  $\beta$ . Throughout this paper, let  $q \in [0, 1]$  be fixed, and we assume that  $\beta^* \in \mathbb{B}_q(R_q)$  unless otherwise specified. Moreover, without loss of generality, we assume that  $\|\beta^*\|_2 = 1$  and define  $\mathbb{S}_2(1) := \{\beta \in \mathbb{R}^d \mid \|\beta\|_2 = 1\}$ , i.e., the  $\ell_2$  unit sphere. Then it follows that  $\beta^* \in \mathbb{B}_q(R_q) \cap \mathbb{S}_2(1)$ .

In order to estimate the regression parameter, one usually considers an estimator  $\hat{\beta} : \mathbb{R}^{n \times d} \times \mathbb{R}^n \rightarrow \mathbb{R}^d$ , which is a measurable function of the observed data  $\{(Z_{i\cdot}, y_i)\}_{i=1}^n$ . Then, for the purpose of assessing the estimation quality of  $\hat{\beta}$ , it is typical to introduce a loss function  $\mathcal{L}(\hat{\beta}, \beta^*)$ , which represents the loss incurred by the estimator  $\hat{\beta}$  when the true parameter  $\beta^* \in \mathbb{B}_q(R_q) \cap \mathbb{S}_2(1)$ . Finally, in the minimax formalism, we aim to choose an estimator that minimizes the following worst-case loss

$$\min_{\hat{\beta}} \max_{\beta^* \in \mathbb{B}_q(R_q) \cap \mathbb{S}_2(1)} \mathcal{L}(\hat{\beta}, \beta^*).$$

Specifically, in this paper, we shall consider the  $\ell_p$ -losses for  $p \in [1, +\infty)$  as follows

$$\mathcal{L}_p(\hat{\beta}, \beta^*) := \|\hat{\beta} - \beta^*\|_p^p.$$

We then impose some regularity conditions on the observed matrix  $Z$ , which are beneficial to analyze the minimax rates. The first assumption requires that the columns of  $Z$  are bounded from above in  $\ell_2$ -norm.

**Assumption 1** (Column normalization). *There exists a constant  $0 < \kappa_c < +\infty$  such that*

$$\frac{1}{\sqrt{n}} \max_{j=1,2,\dots,d} \|Z_{\cdot j}\|_2 \leq \kappa_c.$$

The second assumption imposes a lower bound on the restricted eigenvalue of the surrogate gram matrix  $\hat{\Gamma}$ , which in other words is a lower bound for the restricted curvature.

**Assumption 2** (Restricted eigenvalue condition). *There exists a constant  $\kappa_l > 0$  and a function  $\tau_l(n, d)$  such that for all  $\beta \in \mathbb{B}_q(2R_q)$ ,*

$$\beta^\top \hat{\Gamma} \beta \geq \kappa_l \|\beta\|_2^2 - \tau_l(n, d).$$

**Remark 1.** (i) Note that though we focused on the random design case in this article, Assumptions 1 and 2 are stated in deterministic form. This choice is to make them universal to both fixed and random design matrices. Specifically, previous studies have shown that Assumptions 1 and 2 can be satisfied by a wide range of random matrices with high probability; see, e.g., [11,14,27]. Meanwhile, Assumptions 1 and 2 provide the possibility to analyze the fixed design case in which the matrices are usually chosen by researchers with suitable constants, i.e.,  $\kappa_c$  in Assumption 1 and  $\kappa_l, \tau_l$  in Assumption 2. This deterministic form of the regularity condition on the design matrix is also adopted in the field of modern high-dimensional statistics and machine learning; see, e.g., [11,12,14,28].

(ii) For the Gaussian model we assumed that for  $i = 1, 2, \dots, n$ , the vectors  $X_i$  and  $W_i$  are independently Gaussian with mean 0 and covariance matrices  $\sigma_x^2 \mathbb{I}_d$  and  $\sigma_w^2 \mathbb{I}_d$ , respectively, and the observed covariate  $Z_i$  is also Gaussian with mean 0 and covariance matrix  $\Sigma_z = (\sigma_x^2 + \sigma_w^2) \mathbb{I}_d$ . Recall that  $\sigma_z^2 = \sigma_x^2 + \sigma_w^2$ , then one has that  $Z_i \sim \mathcal{N}(0, \sigma_z^2 \mathbb{I}_d)$ . Furthermore, since the observations are i.i.d., each column  $Z_{\cdot j}$  ( $j = 1, \dots, d$ ) has i.i.d. elements, and thus  $\|Z_{\cdot j}\|_2^2$  ( $j = 1, \dots, d$ ) obeys

the  $\chi^2$  distribution with freedom  $n$ . Then, for Assumption 1, it follows immediately from [27] [Appendix I] on standard tail bounds for  $\chi^2$ -variates and union bounds that there exist universal positive constants  $(c_0, c_1)$  and that Assumption 1 holds with  $\kappa_c = c_0\sigma_z\sqrt{\frac{\log d}{n}}$  with probability at least  $1 - c_1 \exp(-\log d)$ .

(iii) As for Assumption 2, it follows from [29] [Lemma 1] that there exist universal positive constants  $(c_0, c_1, c_2)$ , such that Assumption 2 holds with  $\kappa_l = \frac{\sigma_x^2}{2}$  and  $\tau_l = c_0\sigma_x^2 \max(\frac{\sigma_z^4}{\sigma_x^4}, 1) \frac{\log d}{n}$  with probability at least  $1 - c_1 \exp(-c_2n \min(\frac{\sigma_x^4}{\sigma_z^4}, 1))$ .

### 3. Main Results

In this section, we turn to our main results on lower and upper bounds on minimax risks. We first begin with deriving intermediate results under Assumptions 1 and 2, then we will turn to our main results on probabilistic consequences by virtue of Remark 1(ii), (iii) on the conditions to guarantee Assumptions 1 and 2.

Let  $\mathbb{P}_\beta$  denote the distribution of  $y$  in the linear regression model with additive errors, when  $\beta$  is given and  $Z$  is observed. The following lemma tells us the KL divergence between the distributions induced by two different parameters  $\beta, \beta' \in \mathbb{B}_q(\mathbb{R}_q)$ . The KL divergence plays a key role in establishing the information-theoretic related lower bound. Recall that for two distributions  $\mathbb{P}$  and  $\mathbb{Q}$  which have densities  $d\mathbb{P}$  and  $d\mathbb{Q}$  with respect to some base measure  $\mu$ , the KL divergence is defined by  $D(\mathbb{P}||\mathbb{Q}) = \int \log \frac{d\mathbb{P}}{d\mathbb{Q}} \mathbb{P}(d\mu)$ .

**Lemma 1.** *In the additive error setting, the KL divergence between the distributions induced by any  $\beta, \beta' \in \mathbb{B}_q(\mathbb{R}_q) \cap \mathbb{S}_2(1)$  is equal to*

$$D(\mathbb{P}_\beta || \mathbb{P}_{\beta'}) = \frac{\sigma_x^4}{2\sigma_z^2(\sigma_x^2\sigma_w^2 + \sigma_z^2\sigma_e^2)} \|Z(\beta - \beta')\|_2^2.$$

**Proof.** For each  $i = 1, 2, \dots, n$  fixed, by the model setting,  $(y_i, Z_i)$  is jointly Gaussian with mean  $\mathbf{0}$ . Then by some elementary algebra to compute the covariances, one has that

$$\begin{bmatrix} y_i \\ Z_i \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \beta^\top \Sigma_x \beta + \sigma_e^2 & \beta^\top \Sigma_x \\ \Sigma_x \beta & \Sigma_x + \Sigma_w \end{bmatrix} \right).$$

Then, it follows from standard results on the conditional distribution of Gaussian variables that

$$y_i | Z_i \sim \mathcal{N}(\beta^\top \Sigma_x \Sigma_z^{-1} Z_i, \beta^\top (\Sigma_x - \Sigma_x \Sigma_z^{-1} \Sigma_x) \beta + \sigma_e^2). \tag{2}$$

Now assume that  $\sigma_e$  and  $\sigma_w$  are not both 0; otherwise, the conclusion holds trivially. Since  $\mathbb{P}_\beta$  is a product distribution of  $y_i | Z_i$  over all  $i = 1, 2, \dots, n$ , it follows from (2) that

$$\begin{aligned} D(\mathbb{P}_\beta || \mathbb{P}_{\beta'}) &= \mathbb{E}_{\mathbb{P}_\beta} \left[ \log \frac{\mathbb{P}_\beta(y)}{\mathbb{P}_{\beta'}(y)} \right] \\ &= \mathbb{E}_{\mathbb{P}_\beta} \left[ \frac{n}{2} \log \left( \frac{\sigma_{\beta'}^2}{\sigma_\beta^2} \right) - \frac{\|y - Z \Sigma_z^{-1} \Sigma_x \beta\|_2^2}{2\sigma_\beta^2} + \frac{\|y - Z \Sigma_z^{-1} \Sigma_x \beta'\|_2^2}{2\sigma_{\beta'}^2} \right] \tag{3} \\ &= \frac{n}{2} \log \left( \frac{\sigma_{\beta'}^2}{\sigma_\beta^2} \right) + \frac{n}{2} \left( \frac{\sigma_\beta^2}{\sigma_{\beta'}^2} - 1 \right) + \frac{1}{2\sigma_{\beta'}^2} \|Z \Sigma_z^{-1} \Sigma_x (\beta - \beta')\|_2^2, \end{aligned}$$

where  $\sigma_\beta^2 := \beta^\top (\Sigma_x - \Sigma_x \Sigma_z^{-1} \Sigma_x) \beta + \sigma_e^2$ , and  $\sigma_{\beta'}^2$  is given analogously. Since  $\Sigma_x = \sigma_x^2 \mathbb{I}_n$ ,  $\Sigma_w = \sigma_w^2 \mathbb{I}_n$ , and  $\|\beta\|_2 = 1$  by the assumptions, we immediately arrive at that

$$\sigma_\beta^2 = \left( \sigma_x^2 - \frac{\sigma_x^4}{\sigma_z^2} \right) \|\beta\|_2^2 + \sigma_e^2 = \frac{\sigma_x^2 \sigma_w^2}{\sigma_z^2} + \sigma_e^2.$$

Substituting this equality into (3) yields that

$$D(\mathbb{P}_\beta || \mathbb{P}_{\beta'}) = \frac{\sigma_x^4}{2\sigma_z^2(\sigma_x^2\sigma_w^2 + \sigma_z^2\sigma_\epsilon^2)} \|Z(\beta - \beta')\|_2^2.$$

The proof is completed.  $\square$

**Proposition 1.** *In the additive error setting, suppose that the observed matrix  $Z$  satisfies Assumption 1 with  $0 < \kappa_c < +\infty$ . Then, for any  $p \in [1, +\infty)$ , there exists a constant  $c_{q,p}$  depending only on  $q$  and  $p$  such that with probability at least  $1/2$ , the minimax  $\ell_p$ -loss over the  $\ell_q$ -ball is lower bounded as*

$$\min_{\hat{\beta}} \max_{\beta^* \in \mathbb{B}_q(R_q) \cap \mathbb{S}_2(1)} \|\hat{\beta} - \beta^*\|_p^p \geq c_{q,p} \left[ \frac{\sigma_z^2(\sigma_x^2\sigma_w^2 + \sigma_z^2\sigma_\epsilon^2)}{\sigma_x^4\kappa_c^2} \right]^{\frac{p-q}{2}} R_q \left( \frac{\log d}{n} \right)^{\frac{p-q}{2}}. \tag{4}$$

**Proof.** For positive numbers  $\delta > 0$  and  $\epsilon > 0$ , let  $M_p(\delta)$  denote the cardinality of a maximal  $\delta$ -packing of the ball  $\mathbb{B}_q(R_q)$  in the  $\ell_p$  metric with elements  $\{\beta^1, \beta^2, \dots, \beta^M\}$ , and  $N_2(\epsilon)$  denote the minimal cardinality of an  $\epsilon$ -covering of  $\mathbb{B}_q(R_q)$  in  $\ell_2$ -norm. We follow the standard technique in [30] to transform the estimation on lower bound into a multi-way hypothesis testing problem as follows

$$\mathbb{P} \left( \min_{\hat{\beta}} \max_{\beta^* \in \mathbb{B}_q(R_q) \cap \mathbb{S}_2(1)} \|\hat{\beta} - \beta^*\|_p^p \geq \frac{1}{2^p} \delta^p \right) \geq \min_{\tilde{\beta}} \mathbb{P}(B \neq \tilde{\beta}), \tag{5}$$

where  $B \in \mathbb{R}^d$  is a random variable uniformly distributed over the packing set  $\{\beta^1, \beta^2, \dots, \beta^M\}$ , and  $\tilde{\beta}$  is an estimator taking values in the packing set. It then follows from Fano's inequality [30] that

$$\mathbb{P}(B \neq \tilde{\beta}) \geq 1 - \frac{I(y; B) + \log 2}{\log M_p(\delta)}, \tag{6}$$

where  $I(y; B)$  is the mutual information between the random variable  $B$  and the observation vector  $y \in \mathbb{R}^n$ . It now remains to upper bound the mutual information  $I(y; B)$ . Based on the procedure of [30], the mutual information is upper bounded as

$$I(y; B) \leq \log N_2(\epsilon) + D(\mathbb{P}_\beta || \mathbb{P}_{\beta'}). \tag{7}$$

Let  $\text{absconv}_q(Z/\sqrt{n})$  denote the  $q$ -convex hull of the rescaled columns of the observed matrix  $Z$ , that is,

$$\text{absconv}_q(Z/\sqrt{n}) := \left\{ \frac{1}{\sqrt{n}} \sum_{j=1}^n \theta_j Z_{\cdot j} \mid \theta \in \mathbb{B}_q(R_q) \right\},$$

where the normalization factor  $1/\sqrt{n}$  is used for convenience. Since  $Z$  satisfies Assumption 1 [31], [Lemma 4] is applicable to concluding that there exists a set  $\{Z\tilde{\beta}^1, Z\tilde{\beta}^2, \dots, Z\tilde{\beta}^N\}$  such that for all  $Z\beta \in \text{absconv}_q(Z)$ , there exists some index  $i$  and some constant  $c > 0$  such that  $\|Z(\beta - \tilde{\beta}^i)\|_2/\sqrt{n} \leq c\kappa_c\epsilon$ . Combining this inequality with Lemma 1 and (7), one has that the mutual information is upper bounded as

$$I(y; B) \leq \log N_2(\epsilon) + \frac{\sigma_x^4}{\sigma_z^2(\sigma_x^2\sigma_w^2 + \sigma_z^2\sigma_\epsilon^2)} nc^2\kappa_c^2\epsilon^2.$$

Thus, we obtain by (6) that

$$\mathbb{P}(B \neq \tilde{\beta}) \geq 1 - \frac{\log N_2(\epsilon) + \frac{\sigma_x^4}{\sigma_z^2(\sigma_x^2\sigma_w^2 + \sigma_z^2\sigma_\epsilon^2)} nc^2\kappa_c^2\epsilon^2 + \log 2}{\log M_p(\delta)}. \tag{8}$$

It remains to choose the packing and covering set radii (i.e.,  $\delta$  and  $\epsilon$ , respectively) such that (8) is strictly above zero, say bounded below by  $1/2$ . For the sake of simplicity, denote  $\sigma^2 := \frac{\sigma_z^2(\sigma_x^2\sigma_w^2 + \sigma_z^2\sigma_\epsilon^2)}{\sigma_x^4}$ . Suppose that we choose the pair  $(\delta, \epsilon)$  such that

$$\frac{c^2 n}{\sigma^2} \kappa_c^2 \epsilon^2 \leq \log N_2(\epsilon), \text{ and} \tag{9a}$$

$$\log M_p(\delta) \geq 6 \log N_2(\epsilon). \tag{9b}$$

As long as  $N_2(\epsilon) \geq 2$ , it is guaranteed that

$$\mathbb{P}(B \neq \tilde{\beta}) \geq 1 - \frac{2 \log N_2(\epsilon) + \log 2}{6 \log N_2(\epsilon)} \geq \frac{1}{2}, \tag{10}$$

as desired. It remains to determine the values of the pair  $(\delta, \epsilon)$  satisfying (9). By [31] [Lemma 3], we know that if  $\frac{c^2 n}{\sigma^2} \kappa_c^2 \epsilon^2 = L_{q,2} \left[ R_q^{\frac{2}{2-q}} \left( \frac{1}{\epsilon} \right)^{\frac{2q}{2-q}} \log d \right]$  for some constant  $L_{q,2}$  depending only on  $q$ , then (9a) is satisfied. Thus, we can choose  $\epsilon$  satisfying

$$\epsilon^{\frac{4}{2-q}} = L_{q,2} R_q^{\frac{2}{2-q}} \frac{\sigma^2 \log d}{c^2 \kappa_c^2 n}. \tag{11}$$

In addition, it follows from [31] [Lemma 3] that if  $\delta$  is chosen to satisfy

$$U_{q,p} \left[ R_q^{\frac{p}{p-q}} \left( \frac{1}{\delta} \right)^{\frac{pq}{p-q}} \log d \right] \geq 6 L_{q,2} \left[ R_q^{\frac{2}{2-q}} \left( \frac{1}{\epsilon} \right)^{\frac{2q}{2-q}} \log d \right], \tag{12}$$

for some constant  $U_{q,p}$  depending only on  $q$  and  $p$ , then (9b) holds. Combining (11) and (12), one has that

$$\delta^p \leq \left[ \frac{U_{q,p}}{6 L_{q,2}} \right]^{\frac{p-q}{q}} \left( \epsilon^{\frac{4}{2-q}} \right)^{\frac{p-q}{2}} R_q^{\frac{2-p}{2-q}} = L_{q,2}^{\frac{p-q}{2}} \left[ \frac{U_{q,p}}{6 L_{q,2}} \right]^{\frac{p-q}{q}} R_q \left[ \frac{\sigma^2 \log d}{c^2 \kappa_c^2 n} \right]^{\frac{p-q}{2}}.$$

Combining this inequality with (10) and (5), we obtain that there exists a constant  $c_{q,p}$  depending only on  $q$  and  $p$  such that

$$\mathbb{P} \left( \min_{\hat{\beta}} \max_{\beta^* \in \mathbb{B}_q(R_q) \cap \mathbb{S}_2(1)} \|\hat{\beta} - \beta^*\|_p^p \geq c_{q,p} R_q \left[ \frac{\sigma_z^2(\sigma_x^2\sigma_w^2 + \sigma_z^2\sigma_\epsilon^2) \log d}{\sigma_x^4 \kappa_c^2 n} \right]^{\frac{p-q}{2}} \right) \geq \frac{1}{2}.$$

The proof is complete.  $\square$

Note that the probability  $1/2$  in Proposition 1 is just a standard convention, and it may be made arbitrarily close to  $2/3$  by choosing the universal constants suitably. Specifically, noting from Equation (10) that as long as  $N_2(\epsilon) \geq 2$  is sufficiently large, the probability can be made sufficiently close to  $2/3$ . The requirement on the sufficiently large value of  $N_2(\epsilon)$  can be satisfied by choosing the universal constants  $L_{q,2}$  and  $c$  in view of Equation (11).

**Proposition 2.** *In the additive error setting, suppose that for a universal constant  $c_1$ ,  $\hat{\Gamma}$  satisfies Assumption 2 with  $\kappa_l > 0$  and  $\tau_l(n, d) \leq c_1 R_q \left( \frac{\log d}{n} \right)^{1-q/2}$ . Then there exist universal constants  $(c_2, c_3)$  and a constant  $c_q$  depending only on  $q$  such that, with probability at least  $1 - c_2 \exp(-c_3 \log d)$ , the minimax  $\ell_2$ -loss over the  $\ell_q$ -ball is upper bounded as*

$$\min_{\hat{\beta}} \max_{\beta^* \in \mathbb{B}_q(R_q) \cap \mathbb{S}_2(1)} \|\hat{\beta} - \beta^*\|_2^2 \leq c_q \left[ \frac{\sigma_z^{2-q} (\sigma_w + \sigma_\epsilon)^{2-q} + \kappa_l^{1-q}}{\kappa_l^{2-q}} \right] R_q \left( \frac{\log d}{n} \right)^{1-q/2}. \tag{13}$$

**Proof.** It suffices to find an estimator for  $\beta^*$ , which has a small  $\ell_2$ -norm estimation error with high probability. We consider the estimator formulated as follows

$$\hat{\beta} \in \arg \min_{\beta \in \mathbb{B}_q(R_q) \cap \mathbb{S}_2(1)} \left\{ \frac{1}{2} \beta^\top \hat{\Gamma} \beta - \hat{Y}^\top \beta \right\}. \tag{14}$$

It is worth noting that (14) involves solving a nonconvex optimization problem when  $q \in [0, 1)$ , while a near-global solution can be obtained efficiently by the algorithm proposed in [14]. Since  $\beta^* \in \mathbb{B}_q(R_q) \cap \mathbb{S}_2(1)$ , it follows from the optimality of  $\hat{\beta}$  that  $1/2 \hat{\beta}^\top \hat{\Gamma} \hat{\beta} - \hat{Y}^\top \hat{\beta} \leq 1/2 \beta^{*\top} \hat{\Gamma} \beta^* - \hat{Y}^\top \beta^*$ . Define  $\hat{\Delta} := \hat{\beta} - \beta^*$ , and thus one has that  $\hat{\Delta} \in \mathbb{B}_q(2R_q)$ . Then it follows that

$$\hat{\Delta}^\top \hat{\Gamma} \hat{\Delta} \leq 2 \langle \hat{\Delta}, \hat{Y} - \hat{\Gamma} \beta^* \rangle.$$

This inequality, together with the assumption that  $\hat{\Gamma}$  satisfies Assumption 2, implies that

$$\kappa_l \|\hat{\Delta}\|_2^2 - \tau_l(R_q, n, d) \leq 2 \langle \hat{\Delta}, \hat{Y} - \hat{\Gamma} \beta^* \rangle \leq 2 \|\hat{\Delta}\|_1 \|\hat{Y} - \hat{\Gamma} \beta^*\|_\infty. \tag{15}$$

It then follows from [11] [Lemma 2] that there exist universal constants  $(c_2, c_3, c_4)$  such that, with probability at least  $1 - c_2 \exp(-c_3 \log d)$ ,

$$\|\hat{Y} - \hat{\Gamma} \beta^*\|_\infty \leq c_4 \sigma_z (\sigma_w + \sigma_e) \|\beta^*\|_2 \sqrt{\frac{\log d}{n}} = c_4 \sigma_z (\sigma_w + \sigma_e) \sqrt{\frac{\log d}{n}}. \tag{16}$$

Combining (15) and (16), one has that

$$\kappa_l \|\hat{\Delta}\|_2^2 \leq 2c_4 \sigma_z (\sigma_w + \sigma_e) \sqrt{\frac{\log d}{n}} \|\hat{\Delta}\|_1 + \tau_l(R_q, n, d).$$

Introduce the shorthand  $\sigma := \sigma_z (\sigma_w + \sigma_e)$ . Recall that  $\hat{\Delta} \in \mathbb{B}_q(2R_q)$ . It then follows from [24] [Lemma 5] (with  $\tau = \frac{2c_4 \sigma}{\kappa_l} \sqrt{\frac{\log d}{n}}$ ) and the assumption  $\tau_l(R_q, n, d) \leq c_1 R_q \left(\frac{\log d}{n}\right)^{1-q/2}$  that

$$\|\hat{\Delta}\|_2^2 \leq \sqrt{2R_q} \left(\frac{2c_4 \sigma}{\kappa_l} \sqrt{\frac{\log d}{n}}\right)^{1-q/2} \|\hat{\Delta}\|_2 + 2R_q \left(\frac{2c_4 \sigma}{\kappa_l} \sqrt{\frac{\log d}{n}}\right)^{2-q} + \frac{c_1}{\kappa_l} R_q \left(\frac{\log d}{n}\right)^{1-q/2}.$$

Therefore, by solving this inequality with the indeterminate viewed as  $\|\hat{\Delta}\|_2$ , we arrive at the conclusion that there exists a constant  $c_q$  depending only on  $q$  such that (13) holds with probability at least  $1 - c_2 \exp(-c_3 \log d)$ . The proof is complete.  $\square$

**Remark 2.** (i) The lower and upper bounds on minimax risks are dependent on the triple  $(R_q, n, d)$ , the error level, and structural properties of the observed matrix  $Z$ , as shown in Propositions 1 and 2. Specifically, by setting  $p = 2$  in Proposition 1, the lower and upper bounds agree up to constant factors independent of the triple  $(R_q, n, d)$ , showing the optimal minimax rate in the additive error case.

(ii) Note that when  $p = 2$  and  $q = 0$  (i.e., the exact sparse case), the minimax rate scales as  $\Theta\left(R_0 \frac{\log d}{n}\right)$ . In the high-dimensional regime when  $d/R_0 \sim d^\gamma$  for some constant  $\gamma > 0$ , this rate is equivalent to  $R_0 \frac{\log(d/R_0)}{n}$  (up to constant factors), which re-captures the same scaling as in [19].

(iii) The assumption that  $\tau_l(R_q, n, d) \leq c_1 R_q \left(\frac{\log d}{n}\right)^{1-q/2}$  in Proposition 2 is not unreasonable. It has been shown in [11] [Lemma 1] that it can be satisfied with high probability for the high-dimensional linear errors-in-variables model.

The following two theorems are on probabilistic consequences in view of conditions to ensure Assumptions 1 and 2. The proofs are obtained by applying Propositions 1 and 2 together with Remark 1(ii),(iii), respectively, as well as the elementary probability theory.

**Theorem 1** (Lower bound on  $\ell_p$ -loss). *In the additive error setting, for any  $p \in [1, +\infty)$ , there exist universal positive constants  $(c_0, c_1)$  and a constant  $c_{q,p}$  depending only on  $q$  and  $p$  such that, with probability at least  $1/2(1 - c_1 \exp(-\log d))$ , the minimax  $\ell_p$ -loss over the  $\ell_q$ -ball is lower bounded as*

$$\min_{\hat{\beta}} \max_{\beta^* \in \mathbb{B}_q(R_q) \cap \mathbb{S}_2(1)} \|\hat{\beta} - \beta^*\|_p^p \geq c_{q,p} \left[ \frac{\sigma_x^2 \sigma_w^2 + \sigma_z^2 \sigma_e^2}{c_0^2 \sigma_x^4} \right]^{\frac{p-q}{2}} R_q. \tag{17}$$

**Proof.** (17) follows from substituting  $\kappa_c = c_0 \sigma_z \sqrt{\frac{\log d}{n}}$  for a universal positive constant  $c_0$  to (4). As for the probability, we define the following two events  $\mathcal{A} = \{\text{Assumption 1 happens}\}$  and  $\mathcal{B} = \{(17) \text{ happens}\}$ . Then Proposition 1 is applicable to conclude that  $\mathbb{P}(\mathcal{B} | \mathcal{A}) \geq 1/2$ . Note from Remark 1(ii) that  $\mathbb{P}(\mathcal{A}) \geq 1 - c_1 \exp(-\log d)$  for a universal positive constant  $c_1$ . Then it follows from the elementary probability that  $\mathbb{P}(\mathcal{B}) = \mathbb{P}(\mathcal{B} | \mathcal{A}) \mathbb{P}(\mathcal{A}) + \mathbb{P}(\mathcal{B} | \mathcal{A}^c) \mathbb{P}(\mathcal{A}^c) \geq \mathbb{P}(\mathcal{B} | \mathcal{A}) \mathbb{P}(\mathcal{A}) \geq 1/2(1 - c_1 \exp(-\log d))$ , which completes the proof.  $\square$

**Theorem 2** (Upper bound on  $\ell_2$ -loss). *In the additive error setting, for a universal constant  $c_0$ , suppose that  $\sigma_x^2 \max(\frac{\sigma_z^4}{\sigma_x^4}, 1) \leq c_0 R_q \left(\frac{\log d}{n}\right)^{-q/2}$ . Then there exist universal constants  $(c_1, c_2, c_3, c_4)$  and a constant  $c_q$  depending only on  $q$  such that, with probability at least  $(1 - c_1 \exp(-c_2 n \min(\frac{\sigma_x^4}{\sigma_z^4}, 1)))(1 - c_3 \exp(-c_4 \log d))$ , the minimax  $\ell_2$ -loss over the  $\ell_q$ -ball is upper bounded as*

$$\min_{\hat{\beta}} \max_{\beta^* \in \mathbb{B}_q(R_q) \cap \mathbb{S}_2(1)} \|\hat{\beta} - \beta^*\|_2^2 \leq c_q \left[ \frac{\sigma_z^{2-q} (\sigma_w + \sigma_e)^{2-q} + \sigma_x^{2-2q}}{\sigma_x^{4-2q}} \right] R_q \left(\frac{\log d}{n}\right)^{1-q/2}. \tag{18}$$

**Proof.** (17) follows from substituting  $\kappa_l = \frac{\sigma_z^2}{2}$  to (13). As for the probability, we define the following two events  $\mathcal{C} = \{\text{Assumption 2 happens}\}$  and  $\mathcal{D} = \{(18) \text{ happens}\}$ . Then Proposition 2 is applicable to conclude that  $\mathbb{P}(\mathcal{D} | \mathcal{C}) \geq 1 - c_3 \exp(-c_4 \log d)$  for universal positive constants  $(c_3, c_4)$ . Note from Remark 1(iii) that  $\mathbb{P}(\mathcal{C}) \geq 1 - c_1 \exp(-c_2 n \min(\frac{\sigma_x^4}{\sigma_z^4}, 1))$  for universal positive constant  $(c_1, c_2)$ . Then it follows from the elementary probability that  $\mathbb{P}(\mathcal{D}) = \mathbb{P}(\mathcal{D} | \mathcal{C}) \mathbb{P}(\mathcal{C}) + \mathbb{P}(\mathcal{D} | \mathcal{C}^c) \mathbb{P}(\mathcal{C}^c) \geq \mathbb{P}(\mathcal{D} | \mathcal{C}) \mathbb{P}(\mathcal{C}) \geq (1 - c_1 \exp(-c_2 n \min(\frac{\sigma_x^4}{\sigma_z^4}, 1)))(1 - c_3 \exp(-c_4 \log d))$ , which completes the proof.  $\square$

#### 4. Conclusions

We focused on the information-theoretic limitations of estimation for sparse linear regression with additive measurement errors under the high-dimensional scaling. The minimax rates of convergence were analyzed by virtue of lower and upper bounds for  $\ell_p$ -losses over  $\ell_q$ -balls based on information theory. The derived lower and upper bounds together revealed the influence of corruption in the observed covariates on parameter estimation. Note that the assumed Gaussian random design matrices are of particular interest and widely applied in the field where the design matrix can be chosen by researchers, such as compressed sensing and signal processing [32]. However, the independent Gaussian assumption is still somewhat restrictive, while our earlier work [14] provides upper bounds on estimation to sub-Gaussian matrices with nondiagonal covariances. Further research may generalize the current result, especially the estimation on lower bounds, to sub-Gaussian matrices with nondiagonal covariances or other types of measurement errors, such as the multiplicative errors or errors with dependent structures. In addition, due to the modern high-dimensional challenge, it is of great significance for a method to be adaptive in learning problems. Hence, it would be a prospective direction to analyze minimax optimal rates of convergence without knowledge of the sparsity degree, such as  $q$  and the  $\ell_q$ -radius.

**Author Contributions:** Conceptualization, X.L. and D.W.; methodology, X.L.; validation, X.L. and D.W.; formal analysis, X.L.; investigation, X.L. and D.W.; writing—original draft preparation, X.L.; writing—review and editing, X.L. and D.W.; supervision, D.W.; project administration, X.L. Both authors have read and agreed to the published version of the manuscript.

**Funding:** The APC was funded by The APC was funded by the Research Achievement Cultivation Fund of School of Mathematics, Northwest University.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Ahmed, S.E.; Amiri, S.; Doksum, K. Ensemble linear subspace analysis of high-dimensional data. *Entropy* **2021**, *23*, 324. [[CrossRef](#)] [[PubMed](#)]
2. Song, X.; Wang, C.Y. GMM nonparametric correction methods for logistic regression with error-contaminated covariates and partially observed instrumental variables. *Scand. J. Stat.* **2019**, *46*, 898–919. [[CrossRef](#)] [[PubMed](#)]
3. Sørensen, Ø.; Frigessi, A.; Thoresen, M. Measurement error in Lasso: Impact and likelihood bias correction. *Stat. Sci.* **2015**, *25*, 809–829. [[CrossRef](#)]
4. Carroll, R.J.; Ruppert, D.; Stefanski, L.A.; Crainiceanu, C.M. *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd ed.; Chapman & Hall/CRC: Boca Raton, FL, USA, 2006.
5. Bickel, P.J.; Ritov, Y. Efficient estimation in the errors in variables model. *Ann. Stat.* **1987**, *15*, 513–540. [[CrossRef](#)]
6. Stefanski, L.A.; Carroll, R.J. Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika* **1987**, *74*, 703–716.
7. Torabi, M.; Datta, G.S.; Rao, J. Empirical Bayes estimation of small area means under a nested error linear regression model with measurement errors in the covariates. *Scand. J. Stat.* **2009**, *36*, 355–369. [[CrossRef](#)]
8. Xu, Y.H.; Li, Y.H.; Song, X. Locally efficient semiparametric estimators for proportional hazards models with measurement error. *Scand. J. Stat.* **2016**, *43*, 558–572. [[CrossRef](#)] [[PubMed](#)]
9. Bühlmann, P.; Van De Geer, S. *Statistics for High-Dimensional Data: Methods, Theory And Applications*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2011.
10. Wainwright, M.J. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*; Cambridge University Press: Cambridge, UK, 2019; Volume 48.
11. Loh, P.L.; Wainwright, M.J. High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *Ann. Stat.* **2012**, *40*, 1637–1664. [[CrossRef](#)]
12. Loh, P.L.; Wainwright, M.J. Regularized M-estimators with nonconvexity: Statistical and algorithmic theory for local optima. *J. Mach. Learn. Res.* **2015**, *16*, 559–616.
13. Datta, A.; Zou, H. Cocolasso for high-dimensional error-in-variables regression. *Ann. Stat.* **2017**, *45*, 2400–2426. [[CrossRef](#)]
14. Li, X.; Wu, D.Y.; Li, C.; Wang, J.H.; Yao, J.C. Sparse recovery via nonconvex regularized M-estimators over  $\ell_q$ -balls. *Comput. Stat. Data Anal.* **2020**, *152*, 107047. [[CrossRef](#)]
15. Bu, Y.H.; Zou, S.F.; Liang, Y.B.; Veeravalli, V.V. Estimation of KL divergence: Optimal minimax rate. *IEEE Trans. Inform. Theory* **2018**, *64*, 2648–2674. [[CrossRef](#)]
16. Loh, P.L. On lower bounds for statistical learning theory. *Entropy* **2017**, *19*, 617. [[CrossRef](#)]
17. Cai, T.T.; Zhang, A. Minimax rate-optimal estimation of high-dimensional covariance matrices with incomplete data. *J. Multivar. Anal.* **2016**, *150*, 55–74. [[CrossRef](#)] [[PubMed](#)]
18. Wu, J.B. Quasi-minimax estimation in the partial linear model. *Commun. Stat. Theory Methods* **2016**, *46*, 2982–2989. [[CrossRef](#)]
19. Loh, P.L.; Wainwright, M.J. Corrupted and missing predictors: Minimax bounds for high-dimensional linear regression. In Proceedings of the IEEE International Symposium on Information Theory Proceedings, Cambridge, MA, USA, 1–6 July 2012; pp. 2601–2605.
20. Joshi, R.L.; Crump, V.J.; Fischer, T.R. Image subband coding using arithmetic coded trellis coded quantization. *IEEE Trans. Circuits Syst. Video Technol.* **1995**, *5*, 515–523. [[CrossRef](#)]
21. Yu, J.Y.; Li, C.; Song, X.M.; Guo, S.Y.; Wang, E. Parallel mixed image encryption and extraction algorithm based on compressed sensing. *Entropy* **2021**, *23*, 278. [[CrossRef](#)]
22. Li, X.; Wu, D.Y.; Cui, Y.; Liu, B.; Walter, H.; Schumann, G.; Li, C.; Jiang, T.Z. Reliable heritability estimation using sparse regularization in ultrahigh dimensional genome-wide association studies. *BMC Bioinform.* **2019**, *20*, 1–11. [[CrossRef](#)]
23. Pourasad, Y.; Ranjbarzadeh, R.; Mardani, A. A new algorithm for digital image encryption based on chaos theory. *Entropy* **2021**, *23*, 341. [[CrossRef](#)] [[PubMed](#)]

24. Raskutti, G.; Wainwright, M.J.; Yu, B. Minimax rates of estimation for high-dimensional linear regression over  $\ell_q$ -balls. *IEEE Trans. Inform. Theory* **2011**, *57*, 6976–6994. [[CrossRef](#)]
25. Wang, Z.; Paterlini, S.; Gao, F.C.; Yang, Y.H. Adaptive minimax regression estimation over sparse  $\ell_q$ -hulls. *J. Mach. Learn. Res.* **2014**, *15*, 1675–1711.
26. Ye, F.; Zhang, C.H. Rate minimaxity of the Lasso and Dantzig selector for the  $\ell_q$  loss in  $\ell_r$  balls. *J. Mach. Learn. Res.* **2010**, *11*, 3519–3540.
27. Raskutti, G.; Wainwright, M.J.; Yu, B. Restricted eigenvalue properties for correlated Gaussian designs. *J. Mach. Learn. Res.* **2010**, *11*, 2241–2259.
28. Agarwal, A.; Negahban, S.N.; Wainwright, M.J. Fast global convergence of gradient methods for high-dimensional statistical recovery. *Ann. Stat.* **2012**, *40*, 2452–2482. [[CrossRef](#)]
29. Loh, P.L.; Wainwright, M.J. Supplementary material: High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *Ann. Stat.* **2012**, *40*, 1637–1664. [[CrossRef](#)]
30. Yang, Y.H.; Barron, A. Information-theoretic determination of minimax rates of convergence. *Ann. Stat.* **1999**, *27*, 1564–1599. [[CrossRef](#)]
31. Raskutti, G.; Wainwright, M.J.; Yu, B. *Minimax Rates of Estimation for High-Dimensional Linear Regression over  $\ell_q$ -Balls*; Technology Report; IEEE: Piscataway, NJ, USA, 2009.
32. Candès, E.J.; Romberg, J.K.; Tao, T. Stable signal recovery from incomplete and inaccurate measurements. *Commun. Pure Appl. Math.* **2006**, *59*, 410–412. [[CrossRef](#)]