Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity

Eurekaimer

Nankai University
Department of Statistics and Data Science

December 4, 2025

Aos2017

The Annals of Statistics
2017, Vol. 45, No. 2, 615–646
DOI: 10.1214/16-AOS1461

CONFIDENCE INTERVALS FOR HIGH-DIMENSIONAL LINEAR REGRESSION: MINIMAX RATES AND ADAPTIVITY[1]

BY T. TONY CAI AND ZIJIAN GUO

University of Pennsylvania

Confidence sets play a fundamental role in statistical inference. In this paper, we consider confidence intervals for high-dimensional linear regression with random design. We first establish the convergence rates of the minimax expected length for confidence intervals in the oracle setting where the sparsity parameter is given. The focus is then on the problem of adaptation to sparsity for the construction of confidence intervals. Ideally, an adaptive confidence interval should have its length automatically adjusted to the sparsity of the unknown regression vector, while maintaining a pre-specified coverage probability. It is shown that such a goal is in general not attainable, except when the sparsity parameter is restricted to a small region over which the confidence intervals have the optimal length of the usual parametric rate. It is further demonstrated that the lack of adaptivity is not due to the conservativeness of the minimax framework, but is fundamentally caused by the difficulty of learning the bias accurately.

Outline

Outline

The Linear Model

Consider the high-dimensional linear regression model:

$$Y = X\beta^* + \varepsilon$$

where:

- $Y \in \mathbb{R}^n$ is the response vector
- $X \in \mathbb{R}^{n \times p}$ is the design matrix
- $\beta^* \in \mathbb{R}^p$ is the unknown regression coefficient vector
- $\varepsilon \in \mathbb{R}^n$ is the noise vector with i.i.d. components $\varepsilon_i \sim N(0, \sigma^2)$
- Key assumption: $p \gg n$ (high-dimensional regime)

Background: The Success of Estimation

Driven by a wide range of applications, high-dimensional linear regression, where the dimension p can be much larger than the sample size n, has received significant recent attention.

Several penalized/constrained ℓ_1 minimization methods:

Example

- Lasso [7] Robert Tibshirani 1996
- Dantzig Selector [4] Emmanuel J. Candès and Terence Tao 2005
- Square-root Lasso [1] A. BELLONI 2011
- scaled Lasso [6] Tingni Sun and Cun-Hui Zhang 2011

Prior Estimation Works

Under regularity conditions on the design matrix X, we have:

Optimal rate of convergence

We can achieve the optimal rate of convergence $k\frac{\log p}{n}$ for confidence intervals, under the squared error loss over the set of k-sparse regression coefficient vectors with $k \leq c\frac{n}{\log p}$ where $c > 0$ is a constant. That is, there exists some constant $C > 0$ such that

$$\sup_{\|\beta\|_0 \leqslant k} \mathbb{P}\left(\|\hat{\beta} - \beta\|_2^2 > Ck\frac{\log p}{n} \right) = o(1) \tag{1}$$

Prior Inference Works

- **De-biasing Approach**: Zhang and Zhang (2014) [9] introduced de-biasing for valid confidence intervals in high-dimensional regression
- **Subsequent Developments**: Van de Geer et al. (2014) [8], and Javanmard and Montanari (2014) [5] further refined the method
- **Current Limitations**: Existing methods focus on ultra-sparse case ($k \ll \frac{\sqrt{n}}{\log p}$)
- **Rate of Convergence**: Confidence intervals achieve parametric rate $\frac{1}{\sqrt{n}}$ under ultra-sparse conditions

Research Motivation and Gaps

Adaptive Inference

Construct confidence intervals without prior knowledge of sparsity k, whose lengths automatically adapt to the true sparsity while maintaining the pre-specified coverage probability.

- **Sparsity Gap**: Point estimation allows $k \ll \frac{n}{\log p}$, but inference requires stricter $k \ll \frac{\sqrt{n}}{\log p}$

Open Questions

- What happens in the region $\frac{\sqrt{n}}{\log p} \lesssim k \lesssim \frac{n}{\log p}$?
- Can we still construct valid confidence intervals for β_i in this case?
- Is adaptive honest inference possible without knowing k?

- **Our Goal**: Address these questions for high-dimensional linear regression with random design

Core Contributions

Main Contributions

- Establish minimax rates for adaptive honest inference in high-dimensional linear regression
- Develop different strategies for **sparse loading** (individual coefficients) and **dense loading** (sums of coefficients)
- Provide theoretical guarantees for adaptive confidence intervals without knowing the sparsity level k

Outline

Notations

ℓ_p norm

The ℓ_p norm of a vector $a = (a_1, \ldots, a_n)^T \in \mathbb{R}^n$, we define the ℓ_p norm

$$\|a\|_p = \left(\sum_{i=1}^{n} |a_i|^p \right)^{1/p}$$

so we can get the ℓ_0 norm $\|a\|_0 = \sum_{i=1}^{n} 1\{a_i \neq 0\}$, ℓ_∞ norm $\|a\|_\infty = \max_{1 \leqslant i \leqslant n} |a_i|$, and $a_{-j} \in \mathbb{R}^{n-1}$ stand for the subvector of a without the $j-th$ component.

Asymptotic notation

For positive sequences $\{a_n\}, \{b_n\}$, we write $a_n = o(b_n), a_n \ll b_n$ if $\lim_{n \to \infty} \frac{a_n}{b_n} = 0$ and $a_n = O(b_n), a_n \lesssim b_n / b_n \gtrsim a_n$ if there exists a C such that $a_n \leqslant C b_n$ for all n. If $a_n \lesssim b_n, b_n \lesssim a_n \implies a_n \asymp b_n$

Model Setting

Key Model Components

- **The Model (Gaussian Design)**:

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

- **High-Dimensional Regime:** Dimension p much larger than sample size n ($p \gg n$).
- **Random Design:** Rows of X are i.i.d. $N_p(0, \Sigma)$.
- **Parameter Vector θ:** $\theta = (\beta, \Omega, \sigma)$, where $\Omega = \Sigma^{-1}$ (Precision Matrix) is generally **unknown**, along with the noise level σ.
- **Datas $Z = (Z_1, \ldots, Z_n)$:** $Z_i = (y_i, X_i^T) \in \mathbb{R}^{p+1}$ for $i = 1, \ldots, n$

Linear Functional and Loading Vector

- We consider confidence intervals for a linear functional [3]:

$$T(\beta) = \xi^T \beta \tag{2}$$

- where $\xi \in \mathbb{R}^p$ is a given *loading vector*
- ξ determines which combination of coefficients we want to estimate

Two Important Cases

1. $T(\beta) = \beta_i$: When ξ has 1 at position i and 0 elsewhere (sparse loading)
2. $T(\beta) = \sum_{i=1}^{p} \beta_i$: When ξ is all ones (dense loading)

Minimax expected length of confidence interval over Θ

Confidence Interval Set $\mathcal{I}_\alpha(\Theta, T)$

The set of all $(1-\alpha)$ level CIs for $T(\beta)$ over Θ is defined as $\mathcal{I}_\alpha(\Theta, T) = \{\mathrm{CI}_\alpha(T, Z) = [l(Z), u(Z)] : \inf_{\theta\in\Theta} \mathbb{P}_\theta(l(Z) \leq T(\beta) \leq u(Z)) \geq 1-\alpha\}$

Maximum Expected Length

For any confidence interval $CI_\alpha(T, Z, \Theta, T) \in$, the maximum expected length over Θ is defined as

$$L(\mathrm{CI}_\alpha(T, Z), \Theta, T) = \sup_{\theta\in\Theta} \mathbb{E}_\theta L(\mathrm{CI}_\alpha(T, Z))$$

where $L(\mathrm{CI}_\alpha(T, Z)) = u(Z) - l(Z)$

Benchmark $L_\alpha^*(\Theta_1, \Theta, T)$

$$L_\alpha^*(\Theta_1, \Theta, T) = \inf_{\mathrm{CI}_\alpha(T,Z) \in \mathcal{I}_\alpha(\Theta, T)} L(\mathrm{CI}_\alpha(T, Z), \Theta_1, T)$$

$L_\alpha^*(\Theta, T) = L_\alpha^*(\Theta, \Theta, T)$ is minimax expected length

$$L_\alpha^*(\Theta_1, \Theta, T) = \inf_{\mathrm{CI}_\alpha(T,Z) \in \mathcal{I}_\alpha(\Theta, T)} \sup_{\theta \in \Theta_1} \mathbb{E}_\theta L(\mathrm{CI}_\alpha(T, Z))$$

Parameter Spaces and Key Questions

Parameter Space $\Theta(k)$

$\Theta(k) = \{\theta = (\beta, \Omega, \sigma) : \|\beta\|_0 \leq k,\ \ \frac{1}{M_1} \leq \lambda_{\min}(\Omega) \leq \lambda_{\max}(\Omega) \leq M_1,$
$0 < \sigma \leq M_2\}$, $M_1 > 1, M_2 > 0$

Main Research Questions

1. Minimax length $L^*_\alpha(\Theta(k), T)$ when sparsity k is known?
2. Is rate-optimal adaptation over different sparsity levels possible?

Adaptivity Definition

For $k_1 \ll k$, can we have $\mathrm{CI}_\alpha(T, Z)$ such that:

$$L(\mathrm{CI}_\alpha(T, Z), \Theta(k_1), T) \asymp L^*_\alpha(\Theta(k_1), T),$$

$$L(\mathrm{CI}_\alpha(T, Z), \Theta(k), T) \asymp L^*_\alpha(\Theta(k), T)?$$

Loading Vector ξ and Sparsity Regimes

Loading Vector Restriction

$$\xi \in \Xi(q, \bar{c}) = \{\xi \in \mathbb{R}^p : \|\xi\|_0 = q, \xi \neq \mathbf{0}, \ \frac{\max_{j \in \mathrm{supp}(\xi)} |\xi_j|}{\min_{j \in \mathrm{supp}(\xi)} |\xi_j|} \leq \bar{c}\}$$

Why we need $\frac{\max_{j \in \mathrm{supp}(\xi)} |\xi_j|}{\min_{j \in \mathrm{supp}(\xi)} |\xi_j|} \leq \bar{c}\}$?

1. To Prevent Imbalance and Masking in Estimation.
2. To establish a clear and consistent constraint for the theoretical analysis of minimax rates.
3. To lay the foundation for classifying and analyzing different scenarios of linear functionals.

Loading Vector ξ and Sparsity Regimes

Two Important Sparsity Regimes

1. **Sparse Loading:** $\xi \in \Xi(q, \bar{c})$ with $q \leq Ck$
2. **Dense Loading:** $\xi \in \Xi(q, \bar{c})$ with $q = cp^{\gamma_q}$, $2\gamma < \gamma_q \leq 1$

Note

Different behaviors in different regimes:

- Sparse loading: Section 3
- Dense loading: Section 4

Outline

Our Objectives in the Sparse Loading Regime

Research Goals

For sparse loading vectors ($\xi \in \Xi(q, \bar{c})$ with $q \leq Ck$), our objectives are to:

1. Characterize the minimax expected length of valid confidence intervals
2. Determine whether rate-optimal adaptation across different sparsity levels is possible
3. Establish theoretical foundations for constructing optimal confidence intervals

Theorem Organization

To systematically address these objectives, we present our results in the following order:

1. First, Theorem 2 (Non-adaptivity Result): Shows fundamental limits of adaptation
2. Then, Theorem 1 (Minimax Rate): Establishes the optimal rate of convergence
3. Finally, we provide construction methods and analyze adaptivity regions

Why This Order?

Starting with non-adaptivity results helps us understand the fundamental constraints before presenting optimal constructions and rates.

Theorem 2: Non-adaptivity Result

Theorem 2

Suppose $0 < \alpha < \frac{1}{2}$ and $k_1 \leq k \leq c \min\{p^\gamma, \frac{n}{\log p}\}$ for some constants $c > 0$ and $0 \leq \gamma < \frac{1}{2}$. If ξ belongs to the sparse loading regime, then

$$L_\alpha^*(\Theta(k_1), \Theta(k), \xi^T\beta) \geq c_1 \|\xi\|_2 \left(\frac{1}{\sqrt{n}} + k \frac{\log p}{n} \right).$$

Key Insight

When $\frac{\sqrt{n}}{\log p} \ll k \lesssim \frac{n}{\log p}$, we have
$L_\alpha^*(\Theta(k_1), \Theta(k), \xi^T\beta) \asymp L_\alpha^*(\Theta(k), \xi^T\beta) \gg L_\alpha^*(\Theta(k_1), \xi^T\beta)$.

This implies that rate-optimal adaptation is impossible in this regime.

Connection Between Theorem 2 and Theorem 1

Key Observation

Theorem 2 implies the minimax lower bound in Theorem 1 by taking $k_1 = k$.

This shows that the lower bound in Theorem 1 is tight and that our construction achieves the optimal rate.

Implication

For confidence intervals to achieve optimal performance, we need to design methods that:

- Account for the trade-off between $\frac{1}{\sqrt{n}}$ and $k\frac{\log p}{n}$
- Consider the sparsity regime ($k \leq \frac{\sqrt{n}}{\log p}$ or $k > \frac{\sqrt{n}}{\log p}$)

Theorem 1: Minimax Rate for Sparse Loading

Theorem 1

Suppose $0 < \alpha < \frac{1}{2}$ and $k \leq c \min\{p^{\gamma}, \frac{n}{\log p}\}$ for some constants $c > 0$ and $0 \leq \gamma < \frac{1}{2}$. If ξ belongs to the sparse loading regime, then

$$L_{\alpha}^{*}(\Theta(k), \xi^{T}\beta) \asymp \|\xi\|_2 \left(\frac{1}{\sqrt{n}} + k\frac{\log p}{n} \right).$$

Proof Strategy

The theorem is established in two steps:

1. Construct a confidence interval achieving the upper bound
2. Use Theorem 2 to derive the lower bound

Proof Strategy for Theorem 1

Key Components of the Proof

1. Construct a de-biased estimator to center our confidence interval
2. Derive the variance and bias bounds
3. Establish the coverage property and length optimality

The De-biasing Formula

The core of our method is the de-biased scaled Lasso estimator:

$$\tilde{\mu} = \xi^T \hat{\beta} + \hat{u}^T \frac{1}{n} X^T (y - X\hat{\beta})$$

where:

- $\hat{\beta}$ is the scaled Lasso estimator from (3.4)
- $\hat{u}$ is the constrained estimator minimizing variance while controlling bias
- The second term corrects the bias introduced by Lasso regularization

Definition and Rationale of $\hat{u}$

The Definition of $\hat{u}$

$$\hat{u} = \arg\min_{u \in \mathbb{R}^p} \{ u^T \hat{\Sigma} u : \|\hat{\Sigma} u - \xi\|_\infty \leq \lambda_n \}$$

where:

- $\hat{\Sigma} = \frac{1}{n} X^T X$ is the sample covariance matrix
- $\lambda_n = 12\|\xi\|_2 M_1^2 \sqrt{\frac{\log p}{n}}$ is the tuning parameter

Why This Particular Definition?

1. **Variance Minimization**: Minimizes the variance of debiasing term
2. **Bias Control**: The constraint $\|\hat{\Sigma} u - \xi\|_\infty \leq \lambda_n$ ensures that the bias introduced by using $\hat{u}$ instead of $\Sigma^{-1}\xi$ is properly bounded

This construction is critical for ensuring that our de-biased estimator achieves the minimax rate even when $k \gtrsim \frac{\sqrt{n}}{\log p}$, where standard asymptotic normality results break down.

Figure 1: Adaptivity Regions

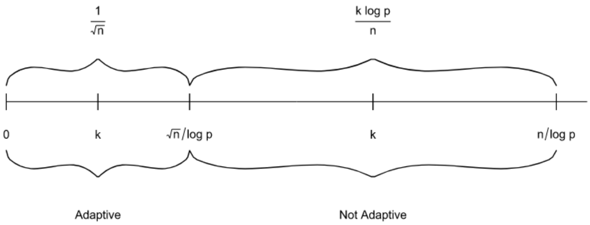

FIG. 1. *Illustration of adaptivity of confidence intervals for $\xi^\intercal \beta$ with a sparse loading ξ satisfying $\|\xi\|_0 \leq Ck_1$. For adaptation between $\Theta(k_1)$ and $\Theta(k)$ with $k_1 \ll k$, rate-optimal adaptation is possible if $k \lesssim \frac{\sqrt{n}}{\log p}$ and impossible otherwise.*

Key Insights

- **Ultra-sparse regime** ($k \leq \frac{\sqrt{n}}{\log p}$): Rate-optimal adaptation possible. Optimal length: $\frac{1}{\sqrt{n}}$.
- **Moderate-sparse regime** ($\frac{\sqrt{n}}{\log p} \ll k \lesssim \frac{n}{\log p}$): Adaptation impossible. Optimal length: $k\frac{\log p}{n}$.

Outline

Dense Loading Regime

Definition of Dense Loading

$\xi \in \Xi(q, \bar{c})$ with $q = cp^{\gamma_q}$, where $2\gamma < \gamma_q \leq 1$

- In this regime, the loading vector ξ has significant mass across many coordinates
- Examples include the sum of all coefficients $\sum_{i=1}^{p} \beta_i$
- Different behavior from sparse loading regime due to the dense nature

Theorem 4: Minimax Rate for Dense Loading

Theorem 4

Suppose $0 < \alpha < \frac{1}{2}$ and $k \leq c \min\{p^\gamma, \frac{n}{\log p}\}$ for some constants $c > 0$ and $0 \leq \gamma < \frac{1}{2}$. If ξ belongs to the dense loading regime, the minimax expected length satisfies:

$$L_\alpha^*(\Theta(k), \xi^T\beta) \asymp \|\xi\|_\infty k\sqrt{\frac{\log p}{n}}$$

Key Difference

This rate is significantly different from the sparse loading case:
$\|\xi\|_2\left(\frac{1}{\sqrt{n}} + k\frac{\log p}{n}\right)$

Construction of Confidence Intervals for Dense Loading

Key Components

Define the confidence interval $\mathrm{CI}_{\alpha}^{D}(\xi^T\beta, Z)$ as:

$$\mathrm{CI}_{\alpha}^{D}(\xi^T\beta, Z) = \begin{cases} [\xi^T\hat{\beta} - \|\xi\|_\infty\rho_2(k), \xi^T\hat{\beta} + \|\xi\|_\infty\rho_2(k)] & \text{on } A, \\ \{0\} & \text{on } A^c, \end{cases}$$

where:

- $A = \{\hat{\sigma} \leqslant \log p\}$
- $\hat{\beta}$ is the scaled Lasso estimator
- $\rho_2(k) = \min\left\{ C_2(X,k)k\sqrt{\frac{\log p}{n}}\hat{\sigma}, \log p\left(k\sqrt{\frac{\log p}{n}}\hat{\sigma}\right)\right\}$

Important Observation

Unlike sparse loading, this interval is not centered at a de-biased estimator. Using a de-biased estimator would result in a confidence interval with length much larger than the optimal.

Theorem 5: Non-adaptivity in Dense Loading Regime

Theorem 5

Suppose $0 < \alpha < \frac{1}{2}$ and $k_1 \leq k \leq c \min\{p^\gamma, \frac{n}{\log p}\}$ for some constants $c > 0$ and $0 \leq \gamma < \frac{1}{2}$. If ξ belongs to the dense loading regime, then:

$$L_\alpha^*(\Theta(k_1), \Theta(k), \xi^T \beta) \geq c_1 \|\xi\|_\infty k \sqrt{\frac{\log p}{n}}$$

Implication

For $k_1 \ll k$, we have:

$$L_\alpha^*(\Theta(k_1), \Theta(k), \xi^T \beta) \gg L_\alpha^*(\Theta(k_1), \xi^T \beta)$$

This shows that rate-optimal adaptation is impossible in the dense loading regime.

Theorem 6: Strong Non-adaptivity Result

Theorem 6

Suppose $0 < \alpha < \frac{1}{2}$ and $k \leq c \min\{p^\gamma, \frac{n}{\log p}\}$ for some constants $c > 0$ and $0 \leq \gamma < \frac{1}{2}$. Let $k_1 \leq (1-\zeta_0)k - 1$ for some positive constant $0 < \zeta_0 < 1$. Then for any $\theta^* = (\beta^*, I, \sigma) \in \Theta(k_1)$ and $\xi \in \Xi(q, \bar{c})$:

$$\inf_{\mathrm{CI}_\alpha(\xi^T\beta, Z) \in \mathcal{I}_\alpha(\Theta(k), \xi^T\beta)} \mathbb{E}_{\theta^*} L(\mathrm{CI}_\alpha(\xi^T\beta, Z)) \geq c_1 \|\xi\|_\infty k \sqrt{\frac{\log p}{n}} \sigma$$

Significance

This theorem establishes a stronger non-adaptivity result for the dense loading regime.

Special Case: Sum of Coefficients $\sum_{i=1}^{p} \beta_i$

Minimax Rate

Theorem 4 implies that the minimax expected length for confidence intervals of $\sum_{i=1}^{p} \beta_i$ satisfies:

$$L_\alpha^*(\Theta(k), \sum \beta_i) \asymp k\sqrt{\frac{\log p}{n}}$$

Comparison of Sparse vs Dense Loading Regimes

Adaptivity Result

A direct consequence of Theorem 6 is that adaptation is also impossible for the sum of coefficients:

$$\inf_{\mathrm{CI}_\alpha(\sum \beta_i, Z)} \mathbb{E}_{\theta^*} L(\mathrm{CI}_\alpha(\sum \beta_i, Z)) \geq c_1 k \sqrt{\frac{\log p}{n}} \sigma$$

Related Work

In the Gaussian sequence model, similar results for the sum of sparse means have been studied in [3], [2](T. Tony Cai,2004).

Comparison of Sparse vs Dense Loading Regimes (Table)

Aspect	Sparse Loading	Dense Loading
Loading	ξ with $q \leq Ck$	ξ with $q = cp^{\gamma_q}$
Minimax	$\|\xi\|_2(\frac{1}{\sqrt{n}} + k\frac{\log p}{n})$	$\|\xi\|_\infty k\sqrt{\frac{\log p}{n}}$
Ultra-sparse	Adaptation possible	No comparable regime
Moderate-sparse	Adaptation impossible	Adaptation impossible
Estimator	De-biased	Scaled Lasso

Outline

Leveraging Prior Knowledge

Setting with Known Ω and σ

- Previous results focused on unknown precision matrix Ω and noise level σ
- Now consider the theoretical setting where $\Omega = I$ and $\sigma = \sigma_0$ are known
- Parameter space: $\Theta(k, I, \sigma_0) = \{\theta = (\beta, I, \sigma_0) : \|\beta\|_0 \leq k\}$
- Will analyze both sparse and dense loading regimes separately

Sparse Loading Regime with Prior Knowledge

Theorem 7: Minimax Rate with Known Ω and σ

When ξ belongs to the sparse loading regime and $\Omega = I$, $\sigma = \sigma_0$ are known:

$$L_\alpha^*(\Theta(k, I, \sigma_0), \xi^T \beta) \asymp \frac{\|\xi\|_2}{\sqrt{n}}$$

Key Insights

- Significantly better than unknown Ω, σ case: $\frac{\|\xi\|_2}{\sqrt{n}} + \|\xi\|_2 k \frac{\log p}{n}$
- Rate is independent of sparsity k (parametric rate)
- Adaptation is possible over the full range $k \leq c \frac{n}{\log p}$
- Uses de-biasing method with sample splitting

Dense Loading Regime with Prior Knowledge

Theorem 8: Adaptivity Lower Bound

For ξ in dense loading regime with known $\Omega = I$, $\sigma = \sigma_0$:
$L^*_\alpha(\Theta(k_1, I, \sigma_0), \Theta(k, I, \sigma_0), \xi^T\beta) \geq$
$c_1\|\xi\|_\infty\sigma_0 \max\left\{\sqrt{kk_1}\sqrt{\frac{\log p}{n}}, \min\left\{k\sqrt{\frac{\log p}{n}}, \frac{\sqrt{k}}{n^{1/4}}\right\}\right\}$

Minimax Rate and Adaptation

- Minimax rate remains: $L^*_\alpha(\Theta(k, I, \sigma_0), \xi^T\beta) \asymp \|\xi\|_\infty k\sqrt{\frac{\log p}{n}}\sigma_0$
- Same as unknown Ω, σ case
- Adaptation still impossible (but cost of adaptation is reduced)
- For $k_1 \ll k$: $L^*_\alpha(\Theta(k_1, I, \sigma_0), \Theta(k, I, \sigma_0), \xi^T\beta) \gg L^*_\alpha(\Theta(k_1, I, \sigma_0), \xi^T\beta)$

Outline

Key Findings for Unknown Ω and σ

Adaptivity Challenges

- Negative adaptivity results in most practical scenarios
- Only ultra-sparse region allows adaptation for sparse loading ξ
- Knowledge of true sparsity k is crucial for honest confidence intervals
- Confidence interval construction is harder than estimation problem

Impact of Known Ω and σ

Dramatic Improvement in Sparse Loading Regime

- Minimax rate becomes parametric: $\frac{\|\xi\|_2}{\sqrt{n}}$ (independent of k)
- Adaptation possible over full range $k \lesssim \frac{n}{\log p}$
- Knowledge of Ω and σ is extremely valuable

Limited Impact in Dense Loading Regime

- Same minimax rate as unknown case
- Adaptation still impossible
- Only cost of adaptation is reduced

Different Construction Methods

Sparse vs Dense Loading

- **Sparse Loading**: De-biasing method is effective
  - Reduces bias without dramatically increasing variance
  - Works well when loading is concentrated
- **Dense Loading**: De-biasing is not applicable
  - Would significantly increase variance
  - Leads to unnecessarily long confidence intervals

Open Problems

Middle Loading Regime

- Construction of confidence intervals for $\xi^T\beta$ when loading ξ is in middle regime
- Specifically when $cp^{\gamma} \leq q \leq cp^{2\gamma+\delta}$ for $0 < \zeta < 1 - 2\gamma$
- No current method achieves minimax length in this challenging regime

Confidence Balls

- Extending results to confidence balls for the entire vector β
- Connections and differences with confidence intervals for linear functionals
- Similar impossibility results have been shown in sparse linear regression

Future Research Directions

General Adaptation Theory

- Developing general adaptation theory for confidence intervals in non-convex settings
- Current theory by Cai and Low requires convex parameter spaces
- High-dimensional linear regression parameter spaces are highly non-convex
- Geometric quantities beyond between-class modulus of continuity needed

Outline

Summary

Key Contributions

- Established minimax rates for confidence intervals in high-dimensional linear regression
- Analyzed two important regimes: sparse loading and dense loading
- Investigated the adaptivity of confidence intervals across different sparsity levels

Summary

Sparse Loading Findings

- Minimax rate: $\|\xi\|_2 \left(\frac{1}{\sqrt{n}} + k\frac{\log p}{n}\right)$
- Adaptation possible when $k \leq \frac{\sqrt{n}}{\log p}$ (ultra-sparse)
- Adaptation impossible when $\frac{\sqrt{n}}{\log p} \ll k \lesssim \frac{n}{\log p}$ (moderate)

Dense Loading Findings

- Minimax rate: $\|\xi\|_\infty k\sqrt{\frac{\log p}{n}}$ (different from sparse case)
- Adaptation impossible across all sparsity levels
- Uses scaled Lasso estimator instead of de-biased estimator

Ending

Thank you for listening!
Any questions?

References I

[1] A. BELLONI, *Square-root lasso: pivotal recovery of sparse signals via conic programming*, Biometrika **98** (2011), no. 4, 791–806.

[2] T. Tony Cai and Mark G. Low, *An adaptation theory for nonparametric confidence intervals*, The Annals of Statistics **32** (2004), no. 5, 1805–1840.

[3] ______, *Minimax estimation of linear functionals over nonconvex parameter spaces*, The Annals of Statistics **32** (2004), no. 2, 552–576.

[4] Emmanuel J. Candès and Terence Tao, *The dantzig selector: Statistical estimation when p is much larger than n*, Quality Engineering **54** (2005), 83–84.

[5] Adel Javanmard and Andrea Montanari, *Confidence intervals and hypothesis testing for high-dimensional regression*, Journal of Machine Learning Research **15** (2014), no. 82, 2869–2909.

[6] Tingni Sun and Cun-Hui Zhang, *Scaled sparse linear regression*, Biometrika **99** (2011), 879–898.

References II

[7] Robert Tibshirani, *Regression shrinkage and selection via the lasso*, Journal of the royal statistical society series b-methodological **58** (1996), 267–288.

[8] Sara van de Geer, Peter Bühlmann, Ya'acov Ritov, and Ruben Dezeure, *On asymptotically optimal confidence regions and tests for high-dimensional models*, The Annals of Statistics **42** (2014), no. 3, 1166 – 1202.

[9] Cun-Hui Zhang and Stephanie S. Zhang, *Confidence intervals for low dimensional parameters in high dimensional linear models*, Journal of the Royal Statistical Society Series B: Statistical Methodology **76** (2013), no. 1, 217–242.