Information and Inference: A Journal of the IMA (2023) **12**, 1193–1247
https://doi.org/10.1093/imaiai/iaad003
Advance Access publication on 20 March 2023

Sparse confidence sets for normal mean models

Yang Ning
Department of Statistics and Data Science, Cornell University, Ithaca, NY 14853, USA
*Corresponding author. Email: yn265@cornell.edu

and

Guang Cheng
Department of Statistics, University of California, Los Angeles, Los Angeles, CA 90095, USA

[Received on 05 December 2021; revised on 30 January 2023; accepted on 16 February 2023]

In this paper, we propose a new framework to construct confidence sets for a d-dimensional unknown sparse parameter $\boldsymbol{\theta}$ under the normal mean model $\boldsymbol{X} \sim N(\boldsymbol{\theta}, \sigma^2\mathbf{I})$. A key feature of the proposed confidence set is its capability to account for the sparsity of $\boldsymbol{\theta}$, thus named as *sparse* confidence set. This is in sharp contrast with the classical methods, such as the Bonferroni confidence intervals and other resampling-based procedures, where the sparsity of $\boldsymbol{\theta}$ is often ignored. Specifically, we require the desired sparse confidence set to satisfy the following two conditions: (i) uniformly over the parameter space, the coverage probability for $\boldsymbol{\theta}$ is above a pre-specified level; (ii) there exists a random subset S of $\{1, ..., d\}$ such that S guarantees the pre-specified true negative rate for detecting non-zero θ_j's. To exploit the sparsity of $\boldsymbol{\theta}$, we allow the confidence interval for θ_j to degenerate to a single point 0 for any $j \notin S$. Under this new framework, we first consider whether there exist sparse confidence sets that satisfy the above two conditions. To address this question, we establish a non-asymptotic minimax lower bound for the non-coverage probability over a suitable class of sparse confidence sets. The lower bound deciphers the role of sparsity and minimum signal-to-noise ratio (SNR) in the construction of sparse confidence sets. Furthermore, under suitable conditions on the SNR, a two-stage procedure is proposed to construct a sparse confidence set. To evaluate the optimality, the proposed sparse confidence set is shown to attain a minimax lower bound of some properly defined risk function up to a constant factor. Finally, we develop an adaptive procedure to the unknown sparsity. Numerical studies are conducted to verify the theoretical results.

Keywords: Adaptivity; confidence interval; high-dimensional statistics; minimax optimality; sparsity; true negative rate.

1. Introduction

Assume that we observe a d-dimensional random vector $\boldsymbol{X} = (X_1, ..., X_d)$ satisfying the following normal mean model, also known as Gaussian sequence model,

$$\boldsymbol{X} \sim N(\boldsymbol{\theta}, \sigma^2\mathbf{I}), \tag{1.1}$$

where $\boldsymbol{\theta} = (\theta_1, ..., \theta_d)$ is a d-dimensional unknown parameter, $\mathbf{I}$ is an identity matrix and σ^2 is the common variance which is assumed to be known. The mathematical simplicity of normal mean models is often exploited to discover the fundamental phenomena underlying more complicated statistical models. In particular, the normal mean model has attracted numerous interest in high-dimensional statistics. Among others, [1] proposed an adaptive procedure for estimating sparse $\boldsymbol{\theta}$ which is asymptotically minimax for ℓ_r loss, while [10] derived the minimax risk for the recovery of sparsity pattern under the Hamming loss. From a different perspective, the detection boundary for testing the null hypothesis $\boldsymbol{\theta} = 0$ has been well studied by [2, 17, 18], among many others. However, the uncertainty quantification in terms of confidence sets for $\boldsymbol{\theta}$ is less explored, partly because one can easily construct the following $(1-\alpha)$ level confidence sets:

$$\Big\{\boldsymbol{\theta} \in \mathbb{R}^d : \max_{1\le j\le d} |X_j - \theta_j| \le t_\alpha \sigma\Big\}, \tag{1.2}$$

where the cut-off t_α can be determined by the Gaussianity of $\boldsymbol{X}$ with Bonferroni (or Sidak) correction or resampling methods [3, 13]. With a slightly different goal, [6] proposed to construct confidence intervals for some randomly selected components of $\boldsymbol{\theta}$, known as selective confidence intervals; see also [7, 16, 31, 34] for some recent development.

Recently, there is a growing interest in developing confidence intervals for sparse linear regression and other regression models, for instance, [5, 11, 19, 24, 26, 27, 29, 32], a list that is far from exhaustive. The method is often termed as a debiased or desparsifying approach in the literature. Their main idea is to remove the bias of the penalized estimator, e.g. Lasso, so that the resulting estimator of the unknown regression coefficients is asymptotically linear. The confidence intervals for each component of the regression parameter are obtained by Gaussian approximation. Intuitively, the debiased estimator can be viewed as the random vector $\boldsymbol{X}$ in the normal mean model after the use of the central limit theorem and other asymptotic approximations. As a result, one can construct confidence sets for the whole vector of regression parameter in a similar way as (1.2) using the resampling method; see [33]. In another strand of research, [25] proposed honest and adaptive confidence sets for sparse linear models. A computationally feasible approach is developed by [12], in which the confidence set has the form $B_{C_n,\widehat{\boldsymbol{\theta}}} := \{\boldsymbol{\theta} \in \mathbb{R}^d : \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2 \le C_n\}$ for some suitable C_n, and $\widehat{\boldsymbol{\theta}}$ is the Lasso estimator in linear regression.

For such confidence set centred at the debiased estimator or Lasso estimator (e.g. $B_{C_n,\widehat{\boldsymbol{\theta}}}$), the points in this set are not necessarily sparse (especially in finite samples). If we know a priori that the true parameter is sparse, all non-sparse points in $B_{C_n,\widehat{\boldsymbol{\theta}}}$ are not the true parameter, which can be removed from $B_{C_n,\widehat{\boldsymbol{\theta}}}$. In other words, the confidence set $B_{C_n,\widehat{\boldsymbol{\theta}}}$ does not respect the sparsity structure of the parameter. Conceptually, we can improve $B_{C_n,\widehat{\boldsymbol{\theta}}}$ by taking all sparse points in $B_{C_n,\widehat{\boldsymbol{\theta}}}$ as the confidence set for $\boldsymbol{\theta}$. However, such a procedure is less intuitive and the result is hard to interpret. To the best of our knowledge, it is an open problem to formulate 'sparse confidence sets' and, if possible, construct them in a simple and optimal way.

1.1 *Formulation of sparse confidence sets*

To address this question, we propose a new framework to construct 'sparse confidence sets' for $\boldsymbol{\theta}$ under the normal mean model. The proposed method can simultaneously quantify the uncertainty of non-zero parameters and also account for the sparsity of $\boldsymbol{\theta}$. To be specific, we first consider the setting that the parameter $\boldsymbol{\theta} = (\theta_1, ..., \theta_d)$ belongs to a one-sided sparse set in $\mathbb{R}^d$, i.e. $\boldsymbol{\theta} \in \Theta^+(s, a)$, where

$$\Theta^+(s,a) = \{\boldsymbol{\theta} \in \mathbb{R}^d : \|\boldsymbol{\theta}\|_0 \le s, \min_{j:\theta_j \ne 0} \theta_j \ge a\}, \tag{1.3}$$

for some $s, a > 0$. Given $\boldsymbol{X} \sim N(\boldsymbol{\theta}, \sigma^2 \mathbf{I})$, a sparse confidence set $M(S, \boldsymbol{U}, \boldsymbol{L})$ for $\boldsymbol{\theta}$ is defined in the following form:

$$M(S, \boldsymbol{U}, \boldsymbol{L}) = \{\boldsymbol{\theta} \in \mathbb{R}^d : \boldsymbol{\theta}_{S^c} = 0 \text{ and } \theta_j \in [L_j, U_j] \text{ for any } j \in S\}, \tag{1.4}$$

where $S := S(\boldsymbol{X})$ is a random subset of $[d] = \{1, 2, ..., d\}$, S^c denotes the complement of S and $\boldsymbol{L} = (L_1, ..., L_d)$ and $\boldsymbol{U} = (U_1, ..., U_d)$ with $L_j := L_j(\boldsymbol{X})$ and $U_j := U_j(\boldsymbol{X})$, respectively, being the lower and upper confidence bounds for θ_j. If j belongs to S, $[L_j, U_j]$ is the confidence interval for θ_j, otherwise the confidence interval degenerates to a single point 0. The cardinality of the random set S determines the 'sparsity' level of $M(S, \boldsymbol{U}, \boldsymbol{L})$. Note that by setting $S = [d]$, $M(S, \boldsymbol{U}, \boldsymbol{L})$ reduces to the classical confidence intervals, such as (1.2). On the other hand, if the support set of $\boldsymbol{\theta}$ is known, one can take $S = \text{supp}(\boldsymbol{\theta})$ and $M(S, \boldsymbol{U}, \boldsymbol{L})$ reduces to the so-called oracle confidence intervals; see Remark 4. By exploiting the sparsity of $\boldsymbol{\theta}$, the oracle confidence interval degenerates to 0 for those θ_j not in the support, and therefore is an example of sparse confidence sets in (1.4). Since the support set of $\boldsymbol{\theta}$ is unknown, in regression models, [14, 15, 30] proposed to construct asymptotically valid oracle confidence intervals for the non-zero parameters under the assumption that the support set can be recovered with probability tending to 1.

Formally, we require that the desired sparse confidence set (1.4) should satisfy the following two conditions.

- $M(S, \boldsymbol{U}, \boldsymbol{L})$ has the desired coverage probability for $\boldsymbol{\theta}$ uniformly over $\Theta^+(s, a)$, that is for a given level $0 < \alpha < 1$,

$$\sup_{\boldsymbol{\theta} \in \Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin M(S, \boldsymbol{U}, \boldsymbol{L})) \leq \alpha. \tag{1.5}$$

  This is the typical requirement for the validity of the confidence set.

- $M(S, \boldsymbol{U}, \boldsymbol{L})$ is 'sparse.' Formally, for a given level $0 < \delta < 1$, we require

$$\text{FPR} \leq 1 - \delta, \quad \text{where } \text{FPR} := \sup_{j \in [d]} \sup_{\boldsymbol{\theta} \in \Theta^+(s,a), \theta_j = 0} \mathbb{P}_{\boldsymbol{\theta}}(j \in S(\boldsymbol{X})). \tag{1.6}$$

  This condition implies that the probability of a null signal with $\theta_j = 0$ being selected by some variable selection algorithm via the set $S(\boldsymbol{X})$ is no greater than $1 - \delta$. Thus, $1 - \delta$ corresponds to the false positive rate (FPR) of selecting non-zero θ_j's. Similarly, we can define the true negative rate (TNR) as $\text{TNR} = 1 - \text{FPR}$, and view δ as the desired TNR level. From this interpretation, we can see that a larger value of δ requires the confidence set to have less false positives. Finally, we note that δ also controls the expected cardinality of $S = S(\boldsymbol{X})$, where we use $|S(\boldsymbol{X})|$ to denote the cardinality of the set $S(\boldsymbol{X})$. Specifically, by (1.6) we obtain

$$\begin{aligned}\sup_{\boldsymbol{\theta} \in \Theta^+(s,a)} \mathbb{E}_{\boldsymbol{\theta}} |S(\boldsymbol{X})| &= \sup_{\boldsymbol{\theta} \in \Theta^+(s,a)} \Big[\sum_{j:\theta_j \neq 0} \mathbb{P}_{\boldsymbol{\theta}}(j \in S(\boldsymbol{X})) + \sum_{j:\theta_j = 0} \mathbb{P}_{\boldsymbol{\theta}}(j \in S(\boldsymbol{X})) \Big] \\ &\leq s + (d - s)(1 - \delta),\end{aligned}$$

where the inequality follows from $\|\boldsymbol{\theta}\|_0 \le s$ and (1.6). Recall that if $j \notin S(\boldsymbol{X})$, $[L_j, U_j]$ degenerates to a single point 0. Thus, the expected number of intervals $[\boldsymbol{L}, \boldsymbol{U}]$ that degenerate to 0 is at least $(d-s)\delta$, i.e. $\mathbb{E}_{\boldsymbol{\theta}}|S^c(\boldsymbol{X})| \ge d - s - (d-s)(1-\delta) = (d-s)\delta$.

Conceptually, it may be more intuitive to directly pre-specify the size of S when constructing $M(S, \boldsymbol{U}, \boldsymbol{L})$ as opposed to requiring (1.6). However, an appropriate choice of $|S|$ depends on the unknown sparsity of $\boldsymbol{\theta}$ and is often difficult to specify in practice. Therefore, we take the current approach which requires (1.6) together with (1.5).

1.2 *Main results*

Under this novel framework, our goal is to construct $M(S, \boldsymbol{U}, \boldsymbol{L})$ such that (1.5) and (1.6) hold. In view of the definition of the sparse confidence set (1.4), it is easily seen that if there exists some $j \in [d]$ such that $j \in \text{supp}(\boldsymbol{\theta})$ and $j \notin S$, then $\boldsymbol{\theta}$ would never be covered by $M(S, \boldsymbol{U}, \boldsymbol{L})$. Similarly, if $|S|$ is too large (e.g. $S = [d]$), there may exist too many false positives such that (1.6) is violated. Thus, the bottleneck is how to construct a set S for which $\text{supp}(\boldsymbol{\theta}) \subseteq S$ holds with some desired probability and (1.6) is valid. We first study the existence of such set S. To this end, a non-asymptotic minimax lower bound for $\mathbb{P}_{\boldsymbol{\theta}}(\text{supp}(\boldsymbol{\theta}) \nsubseteq S)$ is established in Theorem 1 over a suitable class of random sets S satisfying (1.6). More precisely, the class of the random sets is defined in (2.1). The lower bound details the conditions on the sparsity and minimum signal-to-noise ratio (SNR) in the construction. To match the lower bound, we further show in Theorem 3 that, under appropriate conditions on the SNR, a random set $\widehat{S}_{\alpha'}$ obtained by a simple thresholding procedure contains $\text{supp}(\boldsymbol{\theta})$ with probability greater than $1 - \alpha'$ and satisfies (1.6), where α' is a pre-specified tolerance level.

Given the set $\widehat{S}_{\alpha'}$, we proceed to construct the lower and upper confidence bounds $\boldsymbol{L}$ and $\boldsymbol{U}$. Since the parameter space $\Theta^+(s, a)$ in (1.3) is one-sided, we focus on the one-sided sparse confidence set with $U_j = +\infty$ for $j \in S$. In Section 2.2, we derive the lower confidence bound $\widehat{L}_j$ for those $j \in \widehat{S}_{\alpha'}$ using Bonferroni correction to account for the multiple comparisons and the randomness of the estimated set $\widehat{S}_{\alpha'}$. In Theorem 4, we show that the sparse confidence set constructed above satisfies the desired conditions (1.5) and (1.6).

Theorems 1, 3 and 4 together characterize the role of the minimum SNR, defined as a/σ, in the construction of sparse confidence sets. In particular, in the asymptotic regime $d, s \to \infty$, a phase transition phenomenon occurs when the SNR reaches the level $\Phi^{-1}(\delta) + \sqrt{2\log s}$, where $\Phi^{-1}(\cdot)$ is the inverse function of the Gaussian c.d.f. $\Phi(\cdot)$. To be specific, if $a/\sigma \le \Phi^{-1}(\delta) + (1-\epsilon)\sqrt{2\log s}$ for an arbitrarily small positive constant ϵ, it is impossible to construct sparse confidence sets with the set in (2.1). On the other hand, if $a/\sigma \ge \Phi^{-1}(\delta) + \sqrt{2\log s}$, the proposed sparse confidence set satisfies the conditions (1.5) and (1.6).

When the conditions on the SNR are fulfilled, there exist infinite number of sparse confidence sets of form (1.4) that meet (1.5) and (1.6). In Section 3, we further evaluate the optimality of the sparse confidence set. For the one-sided interval $M(S, \boldsymbol{U}, \boldsymbol{L})$, we formally define the following optimality criterion function

$$R(M(S, \boldsymbol{U}, \boldsymbol{L}), \Theta^+(s, a)) := \sup_{1 \le j \le d} \sup_{\boldsymbol{\theta} \in \Theta^+(s,a)} \mathbb{E}_{\boldsymbol{\theta}}(\theta_j - L_j), \tag{1.7}$$

which represents the maximum distance between θ_j and $\mathbb{E}_{\boldsymbol{\theta}}(L_j)$; see Section 3 for further details. Intuitively, $L_j \le \theta_j$ is expected in order for the one-sided confidence interval to cover the unknown parameter θ_j. As a result, the smaller $\mathbb{E}_{\boldsymbol{\theta}}(\theta_j - L_j)$ is, the more preferred the confidence interval is. However, the non-coverage probability of the confidence set $M(S, \boldsymbol{U}, \boldsymbol{L})$ can be inflated, if we force

$R(M(S,\boldsymbol{U},\boldsymbol{L}),\Theta^+(s,a))$ to be too small. This trade-off is formalized in Theorem 6. In particular, we establish the non-asymptotic minimax lower bound for the non-coverage probability of $M(S,\boldsymbol{U},\boldsymbol{L})$ over the class of confidence sets that satisfy (1.6) and $R(M(S,\boldsymbol{U},\boldsymbol{L}),\Theta^+(s,a)) \le m$ for some given m. Under the asymptotic regime $d,s \to \infty$, a direct implication of Theorem 6 is the minimax lower bound for $R(M(S,\boldsymbol{U},\boldsymbol{L}),\Theta^+(s,a))$. This result is shown in Corollary 8. We further show that the sparse confidence set $\bar{M}_{\alpha'}$ defined in (2.14) attains the above minimax lower bound up to a constant factor 2. Thus, the proposed sparse confidence set is optimal (up to a constant) with respect to $R(M(S,\boldsymbol{U},\boldsymbol{L}),\Theta^+(s,a))$.

While the proposed sparse confidence set $\bar{M}_{\alpha'}$ is optimal, the construction of $\bar{M}_{\alpha'}$ requires the knowledge of the unknown sparsity s and the minimum signal strength a. In Section 4, we propose a sparse confidence set that is adaptive to the unknown sparsity. In Theorem 9, we show that, under the asymptotic regime, the adaptive sparse confidence set attains the same minimax lower bound for $R(M(S,\boldsymbol{U},\boldsymbol{L}),\Theta^+(s,a))$ up to a constant.

Numerical studies are conducted in Section 6 to backup our results. The proofs are deferred to the Appendix.

1.3 *Comparison with the existing literature*

In selective inference, the goal is to provide valid confidence intervals for a set of selected parameters $\{\theta_i\}_{i\in\widehat{S}}$, where $\widehat{S}$ is a data-dependent subset of $[d]$ from some variable selection algorithm. Within this framework, there are different types of error rates one may want to control, such as simultaneous over all possible selection (SoP) error rate [8], conditional over selected error rate [21] and simultaneous over selected (SoS) error rate [7, 16]. Refer to [7] for the detailed literature review. Note that one requirement of our sparse confidence set is (1.5), which implies that the sparse confidence set controls the SoP and SoS errors at level α; see Section 5 in [7].

While our two-stage procedure is in similar spirit to selective confidence intervals, our goal is to provide a valid confidence set that covers the whole vector of $\boldsymbol{\theta}$ with the desired coverage probability. In particular, if $j \notin S(\boldsymbol{X})$ for some $S(\boldsymbol{X})$ constructed via our two-stage method, our confidence interval for θ_j is 0. The uncertainty of assigning 0 confidence intervals to θ_j is taken into account in our method. In contrast, selective inference makes no confidence statement about the parameters not selected in $\widehat{S}$ (or equivalently their confidence interval for θ_j is $(-\infty,+\infty)$ for $j \notin \widehat{S}$).

The confidence intervals from the debiased method can be viewed as the confidence intervals with Bonferroni correction under the Gaussian sequence model [5, 11, 19, 26, 29, 32]. Both the Bonferroni confidence intervals and our sparse confidence intervals have the desired coverage probability for the entire vector $\boldsymbol{\theta} \in \mathbb{R}^d$. However, our sparse confidence intervals can degenerate to the point 0 for some entries of $\boldsymbol{\theta}$, which means the sparse confidence intervals can also perform variable selection. Intuitively, if the practitioners are interested in both variable selection and confidence intervals for $\boldsymbol{\theta}$, the sparse confidence intervals can be more appropriate. Finally, we note that the optimality results established in Corollary 8 are not applicable to the Bonferroni confidence interval, as it does not belong to $\mathscr{M}_+$ in (3.4).

Notation. The following notations are used throughout the paper. For any $a,b \in \mathbb{R}$, denote $a \vee b = \max(a,b)$ and $a \wedge b = \min(a,b)$. Denote $(a)_+ = a$ if $a > 0$ and 0 otherwise. For any sequences a_n, b_n, we write $a_n \sim b_n$ if $a_n/b_n \to 1$ as $n \to \infty$.

2. Sparse Confidence Sets for One-sided Parameter Space

In this section, we consider how to construct sparse confidence sets $M(S,\boldsymbol{U},\boldsymbol{L})$ under the normal mean model $\boldsymbol{X} \sim N(\boldsymbol{\theta},\sigma^2\mathbf{I})$, where $\boldsymbol{\theta}$ belongs to the space $\Theta^+(s,a)$ defined in (1.3). In order to guarantee (1.5) and (1.6), the bottleneck is to construct the set S, if it is possible. In Section 2.1, we consider how to construct the set S as our first step. Once the set S is available, we construct appropriate lower and upper confidence bounds $\boldsymbol{L}$ and $\boldsymbol{U}$ in Section 2.2.

2.1 *Construction of the set S*

The first question concerns whether it is possible to construct an index set S with the desired properties. Define

$$\mathscr{F}(\delta) = \{S(\boldsymbol{X}) : \mathbb{P}_0(j \in S(\boldsymbol{X})) \leq 1 - \delta, \text{ and the event } \{j \in S(\boldsymbol{X})\} \text{ only depends on } X_j \text{ for any } j \in [d]\}, \quad (2.1)$$

where we define $\mathbb{P}_0(\mathscr{E}(X_j))$ as the probability of some event $\mathscr{E}(X_j)$ depending on X_j, where $X_j \sim N(0, \sigma^2)$ and δ is specified in (1.6). On top of (1.6), we focus on the separable rule: whether j is selected by $S(\boldsymbol{X})$ or not is independent of the data X_i for $i \neq j$.

The following theorem provides the non-asymptotic minimax lower bound for $\mathbb{P}_{\boldsymbol{\theta}}(\text{supp}(\boldsymbol{\theta}) \not\subseteq \widehat{S})$ over the class of separable rules $\mathscr{F}(\delta)$ for any given δ.

Theorem 1. (Minimax lower bound). For any $s \geq 1$ and $0 < \delta < 1$, we have

$$\inf_{\widehat{S} \in \mathscr{F}(\delta)} \sup_{\boldsymbol{\theta} \in \Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\text{supp}(\boldsymbol{\theta}) \not\subseteq \widehat{S}) \geq 1 - \frac{1}{(\Delta + 1)^s}, \quad (2.2)$$

where $\Delta = \Phi(\Phi^{-1}(\delta) - a/\sigma)$. Furthermore, consider the asymptotic setting that $s, d \to \infty$. Let c_s denote an arbitrary sequence $c_s \to \infty$ and $c_s/s \to 0$. When the SNR satisfies

$$a/\sigma \leq \kappa_* := \Phi^{-1}(\delta) - \Phi^{-1}(c_s/s), \quad (2.3)$$

we have

$$\liminf_{d,s\to\infty} \inf_{\widehat{S} \in \mathscr{F}(\delta)} \sup_{\boldsymbol{\theta} \in \Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\text{supp}(\boldsymbol{\theta}) \not\subseteq \widehat{S}) = 1. \quad (2.4)$$

A few remarks are in order. First, we note that the non-asymptotic lower bound in (2.2) depends on the TNR δ, the SNR a/σ and the sparsity level s. Since we are only interested in whether the non-zero parameters in $\boldsymbol{\theta}$ are selected by $\widehat{S}$ or not, the lower bound is free of the dimensionality d, which differs from the lower bounds for support recovery [10]. Second, the role of SNR and TNR becomes more transparent in the asymptotic regime as both $d, s \to \infty$. In particular, the asymptotic lower bound in (2.4) implies that when the SNR is finite or diverges slowly enough ($a/\sigma \leq \kappa_*$), we cannot construct sparse confidence sets $M(\widehat{S}, \boldsymbol{U}, \boldsymbol{L})$ where $\widehat{S} \in \mathscr{F}(\delta)$ contains $\text{supp}(\boldsymbol{\theta})$ uniformly over the parameter space $\Theta^+(s, a)$. Third, we comment that $\kappa_* > 0$ if and only if $\delta > c_s/s$. Thus, the negative result (2.4) is meaningful only if the pre-specified TNR is greater than c_s/s.

Recall that in view of the definition of the sparse confidence set, $\boldsymbol{\theta} \in M(S, \boldsymbol{U}, \boldsymbol{L})$ implies $\text{supp}(\boldsymbol{\theta}) \subseteq S$. Thus, Theorem 1 leads to the following simple corollary on the feasibility of sparse confidence sets. To avoid repetition, we only present the asymptotic result.

COROLLARY 2. Under the asymptotic setting $s, d \to \infty$, if (2.3) holds, then for any sparse confidence set $M(\widehat{S}, \boldsymbol{U}, \boldsymbol{L})$ with $\widehat{S} \in \mathscr{F}(\delta)$ we always have

$$\liminf_{d,s\to\infty} \sup_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin M(\widehat{S}, \boldsymbol{U}, \boldsymbol{L})) = 1.$$

As a result, the two requirements (1.5) and (1.6) cannot hold simultaneously unless the SNR is above the threshold κ_* defined in (2.3).

In the following text, we construct an index set that satisfies the desired coverage probability under certain signal strength condition. The estimator $\widehat{S}_{\alpha'}$ is defined as

$$\widehat{S}_{\alpha'} = \left\{ j \in [d] : \frac{X_j}{\sigma} \geq \left(\Phi^{-1}\left(\frac{\alpha'}{s}\right) + \frac{a}{\sigma} \right) \vee \Phi^{-1}(\delta) \right\}, \tag{2.5}$$

where α' denotes the tolerance level for the non-coverage probability of the index set. The following theorem shows that $\widehat{S}_{\alpha'}$ belongs to the set $\mathscr{F}(\delta)$ and the non-coverage probability of $\widehat{S}_{\alpha'}$ is no greater than α'.

THEOREM 3. (Upper bound). For any $0 < \alpha' < 1$, it holds that $\widehat{S}_{\alpha'} \in \mathscr{F}(\delta)$. In addition, if

$$a/\sigma \geq \kappa^* := \Phi^{-1}(\delta) - \Phi^{-1}(\alpha'/s) \tag{2.6}$$

holds, then

$$\sup_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\mathrm{supp}(\boldsymbol{\theta}) \not\subseteq \widehat{S}_{\alpha'}) \leq \alpha'. \tag{2.7}$$

REMARK 1. It is of interest to compare the two thresholds κ_* in (2.3) and κ^* in (2.6). Assume that $s \to \infty$ and α' is fixed. By the tail bound inequality for Gaussian random variables (e.g. Lemma 16), we can show that $\kappa^* \sim \Phi^{-1}(\delta) + \sqrt{2\log s}$. Similarly, we have $\kappa_* \sim \Phi^{-1}(\delta) + \sqrt{2\log(s/c_s)}$. Thus, Theorems 1 and 3 together imply a phase transition at the level $\Phi^{-1}(\delta) + \sqrt{2\log s}$, i.e.

- if $a/\sigma \leq \Phi^{-1}(\delta) + (1-\epsilon)\sqrt{2\log s}$ for some small positive constant ϵ, it is impossible to construct sparse confidence sets $M(\widehat{S}, \boldsymbol{U}, \boldsymbol{L})$ with $\widehat{S} \in \mathscr{F}(\delta)$, i.e. (2.4) holds.
- if $a/\sigma \geq \Phi^{-1}(\delta) + \sqrt{2\log s}$, $\widehat{S}_{\alpha'}$ has the desired coverage probability, i.e. (2.7) holds, which leads to a valid sparse confidence set as shown in the next subsection.

2.2 *Construction of one-sided confidence sets*

In this section, we are ready to construct the confidence set based on $\widehat{S}_{\alpha'}$ in (2.5). Recall that $\theta_j \geq 0$ in $\Theta^+(s, a)$. We are mainly interested in the one-sided confidence interval concerning the distance of the lower confidence bound to 0. Specifically, for $j \in \widehat{S}_{\alpha'}$, we would like to construct a one-sided confidence interval $[c_j, +\infty)$ with some $c_j \geq 0$. If c_j is strictly greater than 0 (i.e. 0 is not contained in the confidence interval), we can conclude that θ_j is non-zero with the desired confidence level. Thus, we define the one-sided sparse confidence set as

$$\widehat{M}_{\alpha'} = M(\widehat{S}_{\alpha'}, \widehat{\boldsymbol{U}}, \widehat{\boldsymbol{L}}), \text{ where } \widehat{L}_j = (X_j - \widehat{u}_{\alpha'}\sigma)_+, \ \widehat{U}_j = +\infty \tag{2.8}$$

for any $j \in \widehat{S}_{\alpha'}$ and $\widehat{u}_{\alpha'}$ is to be specified later to attain the desired coverage probability. To simplify the presentation, we treat α' as a given tuning parameter.

We partition the SNR region into low and high levels for constructing the sparse confidence set (2.8):

- Low SNR region: $R_L = \{\kappa : \kappa^* \leq \kappa < \kappa^* \vee \widehat{\kappa}\}$,
- High SNR region: $R_H = \{\kappa : \kappa \geq \kappa^* \vee \widehat{\kappa}\}$,

  where

$$\widehat{\kappa} = -\Phi^{-1}\left(\frac{\alpha - \alpha'}{d}\right) - \Phi^{-1}\left(\frac{\alpha'}{s}\right) \tag{2.9}$$

and α is the desired level specified in (1.5). In both regions, we require the SNR to be no smaller than κ^* in order to guarantee (2.7); see Remark 1. Under the asymptotic regime $d, s \to \infty$, provided that α, α' and δ are all taken to be constants, we have $\kappa^* < \widehat{\kappa}$ and R_L and R_H reduce to $\{\kappa : \kappa^* \leq \kappa < \widehat{\kappa}\}$ and $\{\kappa : \kappa \geq \widehat{\kappa}\}$, respectively. However, if the pre-specified TNR is sufficiently close to 1, i.e. $\delta > 1 - (\alpha - \alpha')/d$, we have $\kappa^* > \widehat{\kappa}$. In this case, R_L becomes an empty set and $R_H = \{\kappa : \kappa \geq \kappa^*\}$.

The following theorem shows that with a suitable choice of $\widehat{u}_{\alpha'}$ the sparse confidence set (2.8) satisfies (1.5) and (1.6).

THEOREM 4. For any $0 < \alpha' < \alpha$, provided (2.6) holds, we have

$$\sup_{\boldsymbol{\theta} \in \Theta^+(s,a), \theta_j = 0} \mathbb{P}_{\boldsymbol{\theta}}(j \in \widehat{S}_{\alpha'}) \leq 1 - \delta, \text{ and } \sup_{\boldsymbol{\theta} \in \Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin \widehat{M}_{\alpha'}) \leq \alpha,$$

where $\widehat{u}_{\alpha'}$ in (2.8) is given by

$$\widehat{u}_{\alpha'} = \begin{cases} \Phi^{-1}\left(1 - \frac{\alpha - \alpha'}{d}\right) & \text{if } a/\sigma \in R_L, \\ \Phi^{-1}\left(1 - \frac{\alpha - \alpha' - (d-s)(1-\eta^+)}{s}\right) & \text{if } a/\sigma \in R_H, \end{cases}$$

where $\eta^+ = \Phi(\frac{a}{\sigma} + \Phi^{-1}(\frac{\alpha'}{s}))$.

Theorem 4 implies that, when the SNR belongs to the low SNR region assuming it exists, the confidence interval for θ_j is either 0 if $j \notin \widehat{S}_{\alpha'}$ or $[(X_j - \sigma\Phi^{-1}(1 - \frac{\alpha - \alpha'}{d}))_+, +\infty)$ if $j \in \widehat{S}_{\alpha'}$. Note that the one-sided confidence interval for θ_j with Bonferroni correction (without accounting for sparsity) is given by

$$\left[\left(X_j - \sigma\Phi^{-1}\left(1 - \frac{\alpha}{d}\right)\right)_+, +\infty\right) \tag{2.10}$$

for $1 \le j \le d$. Thus, in the case of low SNR, our sparse confidence set for θ_j with $j \in \widehat{S}_{\alpha'}$ can be viewed as the Bonferroni correction at a reduced level $\alpha - \alpha'$ in order to account for the randomness of the estimated set $\widehat{S}_{\alpha'}$.

To better understand the choice of $\widehat{u}_{\alpha'}$ in the high SNR region R_H, we focus on the following subset of R_H,

$$\frac{a}{\sigma} \ge \Big(- \Phi^{-1}\Big(\frac{\alpha - \alpha'}{(1+\epsilon)d}\Big) - \Phi^{-1}\Big(\frac{\alpha'}{s}\Big)\Big) \vee \kappa^*, \tag{2.11}$$

where ϵ is an arbitrarily small positive constant. In this case, we can show that

$$(d-s)(1-\eta^+) \le \frac{\alpha - \alpha'}{1+\epsilon}\frac{d-s}{d} \le \frac{\alpha-\alpha'}{1+\epsilon}.$$

As a result, we have

$$\widehat{u}_{\alpha'} \le \Phi^{-1}\Big(1 - \frac{(\alpha-\alpha')}{s(1+\epsilon)/\epsilon}\Big). \tag{2.12}$$

Recall that the oracle confidence interval is defined as

$$\Big[\Big(X_j - \sigma\Phi^{-1}\Big(1-\frac{\alpha}{s}\Big)\Big)_+, +\infty\Big) \tag{2.13}$$

for $j \in \mathrm{supp}(\boldsymbol{\theta})$ and 0 otherwise. Thus, when (2.11) holds, our sparse confidence set with (2.12) is in similar spirit to the oracle interval with a multiplicity correction factor $s(1+\epsilon)/\epsilon$ at level $\alpha - \alpha'$.

While Theorem 4 shows that $\widehat{M}_{\alpha'}$ in (2.8) is a valid sparse confidence set, we show in Appendix C that $\widehat{M}_{\alpha'}$ is suboptimal in terms of the criterion function $R(M(S, \boldsymbol{U}, \boldsymbol{L}), \Theta^+(s,a))$ in (1.7) (as $d, s \to \infty$). To investigate the optimality of the sparse confidence set in the next section, we now focus on the asymptotic regime, where $d, s \to \infty$ and we treat the pre-specified levels α and δ as fixed. In the following text, we propose an asymptotic sparse confidence set $\bar{M}_{\alpha'}$, and establish its optimality in the next section:

$$\bar{M}_{\alpha'} = M(\bar{S}_{\alpha'}, \bar{\boldsymbol{U}}, \bar{\boldsymbol{L}}), \text{ where } \bar{L}_j = (X_j - \bar{u}_{\alpha'}\sigma)_+, \ \bar{U}_j = +\infty \tag{2.14}$$

for $j \in \bar{S}_{\alpha'}$. Here, $\bar{S}_{\alpha'}$ and $\bar{u}_{\alpha'}$ are defined as follows.

- When $\kappa^{**} \le a/\sigma < \bar{\kappa}$, define $j \in \bar{S}_{\alpha'}$ if and only if $X_j/\sigma \ge \Phi^{-1}(\delta)$, and

$$\bar{u}_{\alpha'} = \sqrt{2\log\Big(\frac{d}{(\alpha-\alpha')C_{d,\alpha-\alpha'}}\Big)}. \tag{2.15}$$

- When $a/\sigma \geq \bar{\kappa}$, define $j \in \bar{S}_{\alpha'}$ if and only if $X_j/\sigma \geq \sqrt{2\log(\frac{2(d-s)}{(\alpha-\alpha')C_{d-s,\alpha-\alpha'}})}$, and

$$\bar{u}_{\alpha'} = \sqrt{2\log\Big(\frac{2s}{(\alpha-\alpha')C_{s,\alpha-\alpha'}}\Big)}. \tag{2.16}$$

The cut-off points for the SNR are defined as

$$\kappa^{**} = \Phi^{-1}(\delta) + \sqrt{2\log\left(\frac{s}{C_{s,\alpha'}\alpha'}\right)},$$

and

$$\bar{\kappa} = \sqrt{2\log\left(\frac{2(d-s)}{(\alpha-\alpha')C_{d-s,\alpha-\alpha'}}\right)} + \sqrt{2\log\left(\frac{s}{C_{s,\alpha'}\alpha'}\right)},$$

where $C_{s,\alpha'} = 2(\pi \log(s/\alpha'))^{1/2}$. The cut-off points κ^{**} and $\bar{\kappa}$ are the asymptotic versions of κ^* in (2.6) and $\widehat{\kappa}$ in (2.9), respectively, by applying the tail bound inequality for Gaussian random variables (e.g. Lemma 16). Note that $\bar{\kappa}$ diverges to infinity faster than κ^* as $d, s \to \infty$. Thus, unlike the high SNR region R_H in the non-asymptotic setting, there is no need to take the maximum of $\bar{\kappa}$ and κ^*.

The following corollary shows that $\bar{M}_{\alpha'}$ is a valid sparse confidence set as $d, s \to \infty$.

COROLLARY 5. Assume that $d, s \to \infty$ and δ, α are pre-specified fixed constants. For any $0 < \alpha' < \alpha$, provided $\kappa^{**} \leq a/\sigma$, we have

$$\limsup_{d,s\to\infty} \sup_{\boldsymbol{\theta}\in\Theta^+(s,a),\theta_j=0} \mathbb{P}_{\boldsymbol{\theta}}(j \in \bar{S}_{\alpha'}) \leq 1-\delta, \quad \limsup_{d,s\to\infty} \sup_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin \bar{M}_{\alpha'}) \leq \alpha.$$

REMARK 2. (On the choice of α'). We note that in general κ^{**}, $\bar{\kappa}$ and $\bar{u}_{\alpha'}$ in (2.15) and (2.16) all depend on the choice of α'. However, in the asymptotic regime, if we set $\alpha' = \gamma\alpha$ for any fixed constant $0 < \gamma < 1$, then $\bar{u}_{\alpha'} \sim \sqrt{2\log d}$ in (2.15) and $\bar{u}_{\alpha'} \sim \sqrt{2\log s}$ in (2.16), and similarly, $\kappa^{**} \sim \Phi^{-1}(\delta) + \sqrt{2\log s}$ and $\bar{\kappa} \sim \sqrt{2\log(d-s)} + \sqrt{2\log s}$, which are all asymptotically independent of α'. From a theoretical perspective, when d, s are large enough, the choice of α' has little effect on the proposed confidence interval. Therefore, in the asymptotic analysis, we treat α' as a fixed small constant. We refer to the numerical studies in Section 6 for sensitivity analysis and further practical guidelines on choosing α'.

Finally, we note that when constructing the sparse confidence set, we treat α and δ as pre-specified. In other words, we do not consider δ as a tuning parameter in this work. In addition, Corollary 5 shows that the sparse confidence set $\bar{M}_{\alpha'}$ is valid when the minimum SNR satisfies $a/\sigma \geq \kappa^{**}$. Without knowing this information, the sparse confidence set should be interpreted with care.

3. Optimality of Sparse Confidence Sets

In this section, we will establish the optimality of the proposed sparse confidence sets with respect to the criterion function $R(M(S,\boldsymbol{U},\boldsymbol{L}),\Theta^+(s,a))$ in (1.7). We define a generic class of one-sided sparse confidence sets as

$$CI_+ = \{M(S,\boldsymbol{U},\boldsymbol{L}) : S \subseteq [d], \text{ and for any } j\in[d],\, L_j, U_j \text{ only depend on } X_j,$$
$$0\le L_j \le X_j \vee 0 \text{ and } U_j=+\infty \text{ if } j\in S, \text{ otherwise } L_j=U_j=0\}.$$

In the above definition, S in $M(S,\boldsymbol{U},\boldsymbol{L})$ can be any subset of $[d]$. For any confidence set $M(S,\boldsymbol{U},\boldsymbol{L})$ in CI_+, we first require that the construction of (L_j,U_j) is separable for $1\le j\le d$, which is compatible with the condition in the definition of $\mathscr{F}(\delta)$ in (2.1). In addition, we require $L_j\le X_j\vee 0$, a technical condition to control the tail of L_j. In order for the interval $[L_j,\infty)$ to cover θ_j, one would expect that the lower confidence bound L_j is smaller than X_j. Together with $L_j\ge 0$, this implies $0\le L_j\le X_j\vee 0$. It is easily seen that the one-sided Bonferroni confidence set with $S=[d]$ and $L_j=(X_j-\sigma\Phi^{-1}(1-\frac{\alpha}{d}))_+$ belongs to CI_+.

Within the class of confidence sets CI_+, we further define $\mathscr{M}_+(m,\delta)$ as a subset such that $R(M(S,\boldsymbol{U},\boldsymbol{L}),\Theta^+(s,a))\le m$ for some given $m>0$ and $S\in\mathscr{F}(\delta)$ holds as defined in (2.1). Formally, for any $m>0$ and δ in (2.1), we have

$$\mathscr{M}_+(m,\delta)=\Big\{M(S,\boldsymbol{U},\boldsymbol{L})\in CI_+ : R(M(S,\boldsymbol{U},\boldsymbol{L}),\Theta^+(s,a))\le m, \text{ and } S\in\mathscr{F}(\delta)\Big\}, \tag{3.1}$$

where the quantity m characterizes the maximum distance between θ_j and the expected value of the lower confidence bound $\mathbb{E}(L_j)$. Intuitively, given any two confidence sets in CI_+ both with the desired coverage probability, we would favour the one with a smaller value of $R(M(S,\boldsymbol{U},\boldsymbol{L}),\Theta^+(s,a))$, as it corresponds to a 'shorter' one-sided confidence interval and is more informative on the possible range of θ_j. However, if we set m to be too small, the non-coverage probability of any confidence sets in $\mathscr{M}_+(m,\delta)$ may go beyond the desired level α.

In the following theorem, we demonstrate this trade-off by showing the non-asymptotic lower bound for the non-coverage probability of any confidence set in $\mathscr{M}_+(m,\delta)$.

THEOREM 6. (Minimax lower bound). For any $s\ge 1$ and $M\in\mathscr{M}_+(m,\delta)$, it holds that

$$\sup_{\boldsymbol{\theta}\in\Theta^+(s,a)}\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta}\notin M)\ge \max\Big(\sup_{\rho\ge a, A\le s} G(d,A,\rho,m),\ \sup_{\rho\ge 0, B\le s} G(s,B,\rho,m),\ 1-\frac{1}{(\Delta+1)^s}\Big), \tag{3.2}$$

where Δ is defined in Theorem 1,

$$G(d,A,\rho,m)=\frac{A[g(d,A,\rho)-(m+R)/\rho]_+}{1+A[g(d,A,\rho)-(m+R)/\rho]_+},$$

with

$$g(d,A,\rho)=\frac{d-A}{A}\Phi\Big(-\frac{\rho}{2\sigma}-\frac{\sigma}{\rho}\log\Big(\frac{d}{A}-1\Big)\Big)+\Phi\Big(-\frac{\rho}{2\sigma}+\frac{\sigma}{\rho}\log\Big(\frac{d}{A}-1\Big)\Big),$$

and

$$R = \sqrt{\frac{1}{2\pi}}\sigma \exp\Big(-\frac{1}{2}\Big(\frac{\rho}{\sigma}\Big)^2\Big)\frac{\sqrt{1+4(\sigma/\rho)^2}-1}{\sqrt{1+4(\sigma/\rho)^2}+1},$$

and $G(s,B,\rho,m)$ is defined similarly.

We show that the non-asymptotic lower bound (3.2) is the maximum of three terms, which are obtained by reducing the supremum over $\boldsymbol{\theta} \in \Theta^+(s,a)$ to the maximum over different subsets of $\Theta^+(s,a)$. For example, the first term $G(d,A,\rho,m)$ is obtained by considering the set $\Theta(A) = \{\boldsymbol{\theta} \in \mathbb{R}^d : \|\boldsymbol{\theta}\|_0 = A, \theta_j = \rho, \text{ for any}\theta_j \neq 0\}$, where $0 < A \leq s$ and $\rho \geq a$. That is we consider all possible $\boldsymbol{\theta}$ with cardinality A and the parameter on the support set is fixed at ρ. Since the non-zero entry is no smaller than a and the sparsity level is no greater than s, we require $\rho \geq a$ and $A \leq s$. When a diverges slowly enough as $s,d \to \infty$, this term dominates and converges to 1 for some suitable m, see case (2) of the following Corollary 7. Similarly, the second term $G(s,B,\rho,m)$ is derived by reducing $\Theta^+(s,a)$ to a different subset; see the proof of Theorem 6 for more information. As seen in case (3) of Corollary 7, this term converges to 1 when a is sufficiently large. The last term is inherited from Theorem 1, see also the discussion of Corollary 2.

To simplify the results in Theorem 6, we consider the asymptotic regime in the following corollary.

COROLLARY 7. Assume that $s,d \to \infty$.

(1). If $a/\sigma \leq \kappa_*$ defined in (2.3), then

$$\liminf_{d,s\to\infty} \inf_{M\in\mathscr{M}_+(m,\delta)} \sup_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin M) = 1.$$

(2). If $\kappa_* < a/\sigma \leq \sqrt{2\log(d/A_d - 1)}$, then

$$\liminf_{d,s\to\infty} \inf_{M\in\mathscr{M}_+(m,\delta)} \sup_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin M) = 1,$$

for $m \leq \sigma(\frac{1}{2} - \frac{W_d}{A_d})\sqrt{2\log(d/A_d - 1)} + \frac{\sqrt{2}\sigma}{4\sqrt{\pi}}(1 - \frac{A_d}{d-A_d})$, where A_d, W_d are two arbitrary sequences satisfying

$$2W_d \leq A_d \leq s, \quad \frac{d}{A_d} \to \infty, \quad \text{and} \quad W_d \to \infty. \tag{3.3}$$

(3). If $a/\sigma \geq \sqrt{2\log(d/A - 1)}$ for some constant $0 < A \leq s$, then

$$\liminf_{d,s\to\infty} \inf_{M\in\mathscr{M}_+(m,\delta)} \sup_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin M) = 1,$$

for $m \leq \sigma(\frac{1}{2} - \frac{V_s}{B_s})\sqrt{2\log(s/B_s - 1)} + \frac{\sqrt{2}\sigma}{4\sqrt{\pi}}(1 - \frac{B_s}{s-B_s})$, where B_s, V_s are two arbitrary sequences satisfying $2V_s \leq B_s < s$ and $s/B_s \to \infty, V_s \to \infty$.

This corollary details the trade-off between the coverage probability of the confidence set and the magnitude of $R(M(S,\boldsymbol{U},\boldsymbol{L})$ in three regions depending on the value of a/σ. In particular, case (1) is inherited from Corollary 2. To understand case (2), we can first pick a sequence A_d that diverges to infinity sufficiently slow (e.g. slower than s), and then set $W_d = A_d^{1/2}$. The condition (3.3) holds. Thus, in case (2), m cannot be smaller than $\sigma(\frac{1}{2} - o(1))\sqrt{2\log d}$ in order to guarantee the desired coverage probability. Similarly, when the minimum SNR a/σ grows fast enough as in case (3), m cannot be smaller than $\sigma(\frac{1}{2} - o(1))\sqrt{2\log s}$.

Finally, we prove the optimality of the sparse confidence set $\bar{M}_{\alpha'}$ in Corollary 5. Consider the class of one-sided confidence sets for which the coverage probability is no smaller than $1-\alpha$ uniformly over $\Theta^+(s,a)$, defined as

$$\mathscr{M}_+ = \{M(S,\boldsymbol{U},\boldsymbol{L}) \in CI_+ : \liminf_{d,s\to\infty} \inf_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \in M(S,\boldsymbol{U},\boldsymbol{L})) \geq 1-\alpha, \text{ and } S \in \mathscr{F}(\delta)\}. \tag{3.4}$$

Recall that Corollary 5 implies $\bar{M}_{\alpha'} \in \mathscr{M}_+$. In the following corollary, we establish the optimality of $\bar{M}_{\alpha'}$ within the class $\mathscr{M}_+$ with respect to the criterion function $R(M,\Theta^+(s,a))$ defined in (1.7).

COROLLARY 8. Assume that $d,s\to\infty$ and δ,α are pre-specified fixed constants.

(1). If $\kappa^{**} \leq a/\sigma \leq \sqrt{2\log(d/A_d - 1)}$ for some sequence $A_d \leq s$ satisfying $A_d \to \infty$ and $d/A_d \to \infty$, then

$$\liminf_{d,s\to\infty} \inf_{M\in\mathscr{M}_+} \frac{R(M,\Theta^+(s,a))}{\sigma\sqrt{2\log d}/2} \geq 1. \tag{3.5}$$

Consider the sparse confidence set $\bar{M}_{\alpha'}$ in Corollary 5 with $\alpha' = \gamma\alpha$ for any constant $0<\gamma<1$. Then $\bar{M}_{\alpha'} \in \mathscr{M}_+$ and

$$\limsup_{d,s\to\infty} \frac{R(\bar{M}_{\alpha'},\Theta^+(s,a))}{\sigma\sqrt{2\log d}} \leq 1. \tag{3.6}$$

(2). If $a/\sigma \geq \tilde{\kappa}$ where $\tilde{\kappa} = \sqrt{2\log(d-s) - \log\log(d-s) + C'} + \sqrt{2\log s - \log\log s + C'} \vee \xi_d$ for some sufficiently large positive constant C' and $\xi_d = \sqrt{(\log\log(d-s) - \log\log s)_+}$, then

$$\liminf_{d,s\to\infty} \inf_{M\in\mathscr{M}_+} \frac{R(M,\Theta^+(s,a))}{\sigma\sqrt{2\log s}/2} \geq 1. \tag{3.7}$$

The sparse confidence set $\bar{M}_{\alpha'}$ satisfies $\bar{M}_{\alpha'} \in \mathscr{M}_+$ and

$$\limsup_{d,s\to\infty} \frac{R(\bar{M}_{\alpha'},\Theta^+(s,a))}{\sigma\sqrt{2\log s}} \leq 1. \tag{3.8}$$

REMARK 3. The inequalities (3.5) and (3.7) together lead to the asymptotic lower bound for $R(M,\Theta^+(s,a))$ over the class of one-sided confidence sets $\mathscr{M}_+$ in two different regimes. Furthermore, (3.6) and (3.8) imply that the sparse confidence set $\bar{M}_{\alpha'}$ developed in Corollary 5 matches the lower bounds up to a constant factor 2 in both regimes.

However, we note that there exists a gap on the minimum SNR between these two regimes. Let us consider the following setting. By taking A_d to be a sequence that diverges to infinity sufficiently slow, we have $\sqrt{2\log(d/A_d - 1)} \sim \sqrt{2\log d}$ in case (1). For case (2), assume that $s = d^{\beta}$ for some $0 < \beta \leq c < 1$, where c is a constant. Then $\log(d-s) = \beta \log d + \log(d^{1-\beta} - 1) = (1 + o(1)) \log d$, and $\log s = \beta \log d$. After some algebra, it can be shown that $\tilde{\kappa}$ in case (2) satisfies

$$1 \leq \lim_{d\to\infty} \frac{\tilde{\kappa}}{\sqrt{2\log d}} \leq 1 + \beta^{1/2}. \tag{3.9}$$

Thus, the ratio between the two cut-points $\tilde{\kappa}$ in case (2) and $\sqrt{2\log(d/A_d - 1)}$ in case (1) converges to 1 as $\beta \to 0$, which occurs if $\boldsymbol{\theta}$ is very sparse with $s = \log d$ (i.e. $\beta = \log\log d/\log d$). In this case, the gap between the two regimes diminishes to 0 asymptotically.

REMARK 4. (Support recovery and oracle confidence set). Recall that if we know the support of $\boldsymbol{\theta}$, we can construct the following one-sided oracle confidence interval $L_j^{oracle} = (X_j - \sigma\Phi^{-1}(1 - \frac{\alpha}{s}))_+$ and $U_j^{oracle} = +\infty$ for $j \in \text{supp}(\boldsymbol{\theta})$ and $L_j^{oracle} = U_j^{oracle} = 0$ otherwise. This implies $\mathbb{E}_{\boldsymbol{\theta}}(\theta_j - L_j^{oracle}) \sim \sigma\sqrt{2\log s}$ for $j \in \text{supp}(\boldsymbol{\theta})$. Intuitively, if the support set can be recovered exactly with high probability, i.e. $\widehat{S} = \text{supp}(\boldsymbol{\theta})$ for some estimator $\widehat{S}$, one would expect that (under some conditions) the same result holds for the plug-in interval

$$\left[\left(X_j - \sigma\Phi^{-1}\left(1 - \frac{\alpha}{|\widehat{S}|}\right)\right)_+, +\infty\right) \text{ for } j \in \widehat{S} \text{ and 0 otherwise.} \tag{3.10}$$

However, in the following text, we will show that the construction of oracle intervals (i.e. support recovery) is impossible even if the SNR satisfies the condition in case (2). In a recent work, [10] established sufficient and necessary conditions for exact (and almost full) support recovery under the Gaussian mean model. Using their notation, define the expected Hamming loss for variable selection as $\mathbb{E}_{\boldsymbol{\theta}}\|\widehat{\eta} - \eta\|_1$, where $\eta = (\eta_1, ..., \eta_d)$ with $\eta_j = I(\theta_j \neq 0)$ denotes the sparsity pattern of $\boldsymbol{\theta}$ and $\widehat{\eta}$ is an estimator of η. Consider the setting $a/\sigma = \tilde{\kappa}$ as in case (2). Theorem 4.2 (ii) of [10] implies that, for d large enough,

$$\inf_{\widehat{\eta}} \sup_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{E}_{\boldsymbol{\theta}}\|\widehat{\eta} - \eta\|_1 \geq s\Phi(-\Delta), \quad \text{where } \Delta = \frac{W}{2\sqrt{2\log(d-s) - 2\log s + W}}$$

and $W = 4\log s + 2\sqrt{(2\log(d-s) - \log\log(d-s) + C')(2\log s - \log\log s + C')}$ with C' given in case (2). To simplify the expression of $s\Phi(-\Delta)$, we consider the very sparse case with $s = \log d$. After some calculation, we can show that for d sufficiently large,

$$s\Phi(-\Delta) \geq s\Phi(-\sqrt{2\log s - \log\log s + C}) \geq \sqrt{\frac{2}{\pi}} \frac{\sqrt{\log s}\exp(-C/2)}{3\sqrt{2\log s - \log\log s + C}} \to \sqrt{\frac{1}{\pi}} \frac{\exp(-C/2)}{3} > 0,$$

where C is a constant. The above derivation shows that, when $a/\sigma = \tilde{\kappa}$ satisfies the SNR condition in case (2), it is impossible to recover the support of $\boldsymbol{\theta}$ no matter what estimators to use. Since the support recovery is impossible in this case, the plug-in interval (3.10) may not guarantee the desired coverage probability.

4. Adaptive Sparse Confidence Sets

In this section, we consider how to construct optimal sparse confidence sets which are adaptive to the unknown sparsity s. In particular, we will show that adaptation is feasible when the SNR satisfies the following condition:

a. $a/\sigma \geq \sqrt{2\log(d-s) - \log\log(d-s) + C'} + \sqrt{2\log s - \log\log s + C'} \vee \bar{\xi}_d$ for some sufficiently large positive constant C' and $\bar{\xi}_d = \sqrt{(2\log\log(d-s) - \log\log s)_+}$.

Note that (B) is slightly different from the case (2) in Corollary 8, where the quantity ξ_d is replaced with $\bar{\xi}_d$. Again, from the derivation in Remark 4, under the SNR condition in (B), it is impossible to recover the support of $\boldsymbol{\theta}$ no matter what estimators to use. Thus, inference after variable selection (e.g. hard thresholding) is still infeasible.

While Corollary 8 part (2) implies that $\bar{M}_{\alpha'}$ has the desired coverage probability and is asymptotically optimal, the construction of $\bar{M}_{\alpha'}$ with $\bar{u}_{\alpha'}$ given by (2.16) requires the knowledge of unknown sparsity s and therefore is not adaptive. In the following text, we propose to construct an adaptive sparse confidence set under condition (B). Define

$$\bar{S}^{ad}_{\alpha'} = \left\{ j \in [d] : X_j/\sigma \geq \sqrt{2\log\left(\frac{2d}{(\alpha-\alpha')C_{d,\alpha-\alpha'}}\right)} \right\},$$

where $C_{d,\alpha'} = 2\sqrt{\pi \log(\frac{d}{\alpha'})}$. Consider a grid of points $\{1, 2, 2^2, ..., 2^T\}$, where T is the largest integer such that $2^T \leq d$. Define $\widehat{s} = 2^{\widehat{m}}$, where $\widehat{m} = \{m \in [T] : 2^{m-1} \leq |\bar{S}^{ad}_{\alpha'}| < 2^m\}$. Finally, define the adaptive sparse confidence set as

$$\widehat{M}^{ad}_{\alpha'} = M(\bar{S}^{ad}_{\alpha'}, \widehat{\boldsymbol{U}}, \widehat{\boldsymbol{L}}_{\widehat{s}}), \text{ where } \widehat{L}_{j,\widehat{s}} = (X_j - u_{\alpha',\widehat{s}}\sigma)_+, \ \widehat{U}_j = +\infty \tag{4.1}$$

for $j \in \bar{S}^{ad}_{\alpha'}$ and $\widehat{L}_{j,\widehat{s}} = \widehat{U}_j = 0$ otherwise, and

$$u_{\alpha',\widehat{s}} = \sqrt{2\log\left(\frac{4\widehat{s}}{(\alpha-\alpha')C_{2\widehat{s},\alpha-\alpha'}}\right)}.$$

It is seen that the construction of the adaptive interval $\widehat{M}^{ad}_{\alpha'}$ is similar to $\bar{M}_{\alpha'}$, but there are several key differences. First, we use a slightly different cut-off for X_j/σ in $\bar{S}^{ad}_{\alpha'}$. When $2s \leq d$ and $s, d \to \infty$, both the cut-offs in $\bar{S}^{ad}_{\alpha'}$ and $\bar{S}_{\alpha'}$ are asymptotically equivalent to $\sqrt{2\log d}$. Second, we replace the unknown sparsity s in $\bar{u}_{\alpha'}$ in (2.16) with $2\widehat{s}$, where $\widehat{s}$ can be viewed as the rounding of the cardinality of the set $\bar{S}^{ad}_{\alpha'}$ to the grid $\{1, 2, 2^2, ..., 2^T\}$. The intuition is as follows. While the asymptotic exact recovery of the support set of $\boldsymbol{\theta}$ is infeasible under (B) (see Remark 4), $\bar{S}^{ad}_{\alpha'}$ is still a reasonable approximation of the unknown support set. In particular, we prove that the cardinality of $\bar{S}^{ad}_{\alpha'}$ is of an order s with high probability. We further round $|\bar{S}^{ad}_{\alpha'}|$ to the grid in order to rigorously control $\mathbb{E}_{\boldsymbol{\theta}}(\theta_j - \widehat{L}_{j,\widehat{s}})$ when $|\bar{S}^{ad}_{\alpha'}|$ is too large. The rounding step is similar to the peeling method in the empirical process [20, 28] and has been used in the Lepski's method for adaptive estimation [9, 22, 23].

The following theorem presents the main result in this section.

THEOREM 9. Assume that $2s \leq d$, $s, d \to \infty$ and δ and α are fixed. Let $\alpha' = \gamma\alpha$ for any constant $0 < \gamma < 1$. The adaptive sparse confidence set $\widehat{M}^{ad}_{\alpha'}$ satisfies $\liminf_{d,s\to\infty} \inf_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \in \widehat{M}^{ad}_{\alpha'}) \geq 1-\alpha$, $\bar{S}^{ad}_{\alpha'} \in \mathscr{F}(\delta)$ and

$$\limsup_{d,s\to\infty} \frac{R(\widehat{M}^{ad}_{\alpha'}, \Theta^+(s,a))}{\sigma\sqrt{2\log s}} \leq 1,$$

where s, a satisfy the condition (B).

Thus, the upper bound of $R(\widehat{M}^{ad}_{\alpha'}, \Theta^+(s,a))$ is asymptotically identical to the 'non-adaptive' confidence set $\bar{M}_{\alpha'}$ as shown in Corollary 8 part (2) and minimax optimal up to a constant. We note that when constructing $\widehat{M}^{ad}_{\alpha'}$, we borrow information from all the data to define $\widehat{L}_{j,\widehat{s}}$, and therefore $\widehat{M}^{ad}_{\alpha'} \notin CI_+$. In Appendix C, we extend our minimax results to a broader class of non-separable sparse confidence sets.

Empirically, the adaptive sparse confidence set $\widehat{M}^{ad}_{\alpha'}$ (with a given α') is fully data dependent and can be easily computed. The simulation studies in Section 6 demonstrate the favourable finite sample performance of this adaptive approach compared with several competing methods. Finally, we note that, as seen in the simulations, when the condition (B) fails (e.g. a/σ is small), the adaptive sparse confidence set $\widehat{M}^{ad}_{\alpha'}$ may not have the desired coverage probability.

5. Extension to Two-sided Sparse Confidence Sets

In this section, we consider $\boldsymbol{\theta} \in \Theta(s,a)$, where

$$\Theta(s,a) = \{\boldsymbol{\theta} \in \mathbb{R}^d : \|\boldsymbol{\theta}\|_0 \leq s, \min_{j:\theta_j\neq 0} |\theta_j| \geq a > 0\}$$

is a two-sided sparse set. The goal is to generalize the results in Sections 2 and 3 to two-sided sparse confidence intervals for $\boldsymbol{\theta}$ in $\Theta(s,a)$. To this end, consider the following estimator of the support set,

$$\widehat{S}^{TS}_{\alpha'} = \left\{ j \in [d] : |X_j|/\sigma \geq \left(\Phi^{-1}\left(\frac{\alpha'}{2s}\right) + a/\sigma\right)_+ \vee \Phi^{-1}\left(\frac{1+\delta}{2}\right) \right\}, \tag{5.1}$$

where α' is the tolerance level. Similarly, we require $|X_j|/\sigma \geq \Phi^{-1}((1+\delta)/2)$ to guarantee the resulting confidence interval is sparse, i.e. $\widehat{S}^{TS}_{\alpha'} \in \mathscr{F}(\delta)$, where $\mathscr{F}(\delta)$ is defined in (2.1). Similar to Theorems 1 and 3, we can establish the upper and lower bounds of the non-coverage probability $\mathbb{P}_{\boldsymbol{\theta}}(\mathrm{supp}(\boldsymbol{\theta}) \not\subseteq \widehat{S})$ over the set $\Theta(s,a)$. The detailed results are shown in Appendix A.

Given the index set $\widehat{S}^{TS}_{\alpha'}$, we define the two-sided sparse confidence set for $\boldsymbol{\theta} \in \Theta(s,a)$ as

$$\widehat{M}^{TS}_{\alpha'} = M(\widehat{S}^{TS}_{\alpha'}, \widehat{\boldsymbol{U}}^{TS}, \widehat{\boldsymbol{L}}^{TS}), \quad \text{where } \widehat{L}^{TS}_j = X_j - \widehat{u}^{TS}_{\alpha'}\sigma, \ \widehat{U}^{TS}_j = X_j + \widehat{u}^{TS}_{\alpha'}\sigma$$

for any $j \in \widehat{S}^{TS}_{\alpha'}$ and

$$\widehat{u}^{TS}_{\alpha'} = \begin{cases} \Phi^{-1}\Big(1-\frac{\alpha-\alpha'}{2d}\Big) & \text{if } \phi^* \le \frac{a}{\sigma} < \phi^* \vee \Big[-\Phi^{-1}(\frac{\alpha-\alpha'}{2d}) - \Phi^{-1}(\frac{\alpha'}{2s})\Big], \\ \Phi^{-1}\Big(1-\frac{\alpha-\alpha'-2(d-s)(1-\eta)}{2s}\Big) & \text{if } \frac{a}{\sigma} \ge \phi^* \vee \Big[-\Phi^{-1}(\frac{\alpha-\alpha'}{2d}) - \Phi^{-1}(\frac{\alpha'}{2s})\Big], \end{cases}$$

where $\eta = \Phi(a/\sigma + \Phi^{-1}(\alpha'/(2s)))$. We can prove that $\widehat{M}^{TS}_{\alpha'}$ satisfies the conditions (1.5) and (1.6),

$$\sup_{\boldsymbol{\theta}\in\Theta(s,a),\theta_j=0} \mathbb{P}_{\boldsymbol{\theta}}(j \in \widehat{S}^{TS}_{\alpha'}) \le 1-\delta, \quad \sup_{\boldsymbol{\theta}\in\Theta(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin \widehat{M}^{TS}_{\alpha'}) \le \alpha.$$

We also establish the optimality theory for the proposed two-sided sparse confidence sets in terms of the length of confidence intervals

$$\sup_{1\le j\le d} \sup_{\boldsymbol{\theta}\in\Theta(s,a)} \mathbb{E}_{\boldsymbol{\theta}}(U_j - L_j).$$

Due to space constraint, we defer the detailed results to Appendix A.

6. Numerical Results

In this section, we conduct simulation studies to evaluate the performance of the proposed sparse confidence sets and compare with several existing methods in terms of coverage probability, interval length and support recovery (sparsity). The sensitivity to the choice of α' is also examined empirically.

We generate $\boldsymbol{X}$ from the normal mean model with $d = 1000$, $\sigma = 1$ and $\boldsymbol{\theta} = (a, ..., a, 0, ..., 0)$, where the first $s = 100$ entries equal a, which is also the SNR, and the rest are 0. We set $\alpha = 0.05$, $\delta = 0.7$ and vary the value of SNR in the simulations. Recall that the proposed one-sided sparse confidence set $\widehat{M}_{\alpha'}$ in (2.8) and the asymptotic version $\bar{M}_{\alpha'}$ in (2.14) depend on the choice of α'. For simplicity, we set $\alpha' = \alpha/2$ in view of Remark 2. The sensitivity analysis of α' and further discussions will be shown subsequently.

We compare the sparse confidence set $\widehat{M}_{\alpha'}$ in (2.8), $\bar{M}_{\alpha'}$ in (2.14) and the adaptive version $\widehat{M}^{ad}_{\alpha'}$ in (4.1) with the following three methods: Bonferroni confidence interval (2.10), oracle interval (2.13) assuming the support of $\boldsymbol{\theta}$ is known and the plug-in interval (3.10), where $j \in \widehat{S}$ if and only if $X_j/\sigma > (2\log d)^{1/2}$. Provided the SNR is sufficiently large, the threshold $(2\log d)^{1/2}$ guarantees the exact support recovery as shown by [10]. The simulation was repeated 500 times. We report the empirical coverage probability of the above confidence sets for $\boldsymbol{\theta}$ and the average distance $(\theta_j - L_j)$ over $j \in \text{supp}(\boldsymbol{\theta})$ (which can be viewed as a version of interval length for one-sided intervals). For $j \notin \text{supp}(\boldsymbol{\theta})$, we often observe that the lower confidence bound is 0 and $\theta_j - L_j = \theta_j$. Hence, it is not very informative to look at the average distance $(\theta_j - L_j)$ over $j \notin \text{supp}(\boldsymbol{\theta})$, and thus we do not report these results.

Figure 1 shows the coverage probability and the average distance $(\theta_j - L_j)$ of the proposed sparse confidence set $\widehat{M}_{\alpha'}$ (hat M), $\bar{M}_{\alpha'}$ (bar M), oracle interval (oracle), plug-in interval (plug-in), Bonferroni confidence interval (Bonferroni) and our adaptive interval (adaptive) over 500 simulations. It is seen from the left panel that when SNR is small the sparse confidence sets ($\widehat{M}_{\alpha'}$, $\bar{M}_{\alpha'}$ and $\widehat{M}^{ad}_{\alpha'}$) all have considerably low coverage probability. This agrees with the minimax lower bound in Theorem 1, i.e. construction of sparse confidence sets is impossible if the SNR is too small. Provided the SNR exceeds 4,

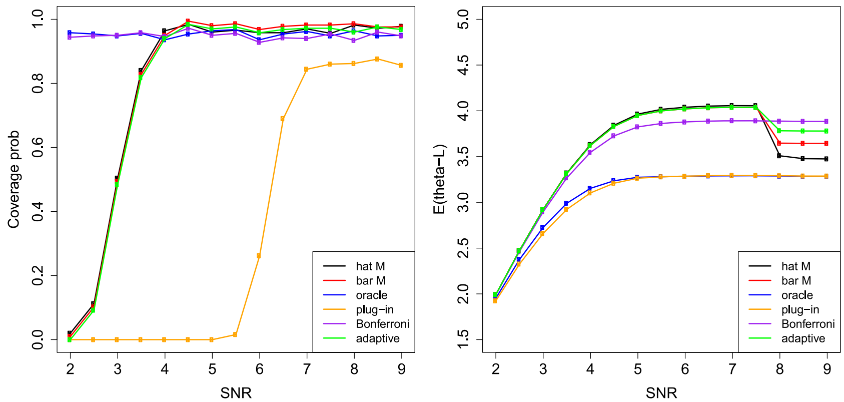


FIG. 1. Coverage probability and the average distance $(\theta_j - L_j)$ over 500 simulations.

all three versions of sparse confidence sets have very similar performance and their coverage probability becomes very close to the desired level. It is of interest to mention that the coverage probability of the plug-in intervals is only around 0.9 even if the SNR is sufficiently large. This is because in finite sample the set $\widehat{S}$ may still miss one or two non-zero signals so that the resulting confidence intervals fail to cover the target parameter $\boldsymbol{\theta}$.

From the right panel, we can see that when the SNR is moderate (say between 4 and 7) the average distance of our sparse confidence sets is comparable with the Bonferroni confidence interval, which is consistent with part (1) of Corollary 8. Once SNR exceeds 7, our sparse confidence sets have a smaller distance and outperform the Bonferroni confidence interval; see part (2) of Corollary 8. Among these three versions of sparse confidence sets, $\widehat{M}^{ad}_{\alpha'}$ is the most conservative one (with the largest average distance $(\theta_j - L_j)$). This can be viewed as the price to pay for not knowing the sparsity s when constructing the sparse confidence sets.

To better understand the sparsity of the proposed sparse confidence set, we can take a closer look at the estimators of the support set, that is $\widehat{S}_{\alpha'}$ in (2.5), $\bar{S}_{\alpha'}$ in (2.14) and $\widehat{S}$ for the plug-in interval. In particular, we plot $\log|\widehat{S}_{\alpha'}|$, $\log|\bar{S}_{\alpha'}|$ and $\log|\widehat{S}|$ in Fig. 2. When the SNR is relatively small, $\widehat{S}_{\alpha'}$ reduces to $\{j \in [d] : X_j/\sigma \geq \Phi^{-1}(\delta)\}$. This explains why the curve for $\widehat{S}_{\alpha'}$ (and similarly $\bar{S}_{\alpha'}$) is horizontal for small SNR. As SNR further grows, it becomes easier to separate the non-zero signals from the rest, and therefore, the size of $\widehat{S}_{\alpha'}$ and $\bar{S}_{\alpha'}$ decreases and eventually reduces to the true sparsity level. In contrast, the set $\widehat{S}$ for support recovery has completely different behaviours. When the SNR is small, very few non-zero θ_j can be identified via $\widehat{S}$ as $X_j \sim N(\theta_j, 1)$ tends to be below the threshold $(2\log d)^{1/2}$. This explains why the coverage probability of the plug-in interval is much lower than the desired level as seen in the left panel of Fig. 1.

Finally, we analyse how sensitive the coverage probability and the average distance $(\theta_j - L_j)$ of proposed sparse confidence set $\widehat{M}_{\alpha'}$ (hat M), $\bar{M}_{\alpha'}$ (bar M) is to the choice of α'. Figure 3 illustrates the results in two cases: $SNR = 3.8$ (moderate SNR) and $SNR = 9$ (high SNR). In panels (a) and (b), when we increase α', the coverage probability becomes closer to the desired level, with the price that the average distance $(\theta_j - L_j)$ is slightly inflated. For the case where the SNR is sufficiently large (panels (c) and (d)), the coverage probability is less dependent on α', whereas the average distance tends to be much larger when α' is close to $\alpha = 0.05$. While the effect of α' is asymptotically ignorable as seen in

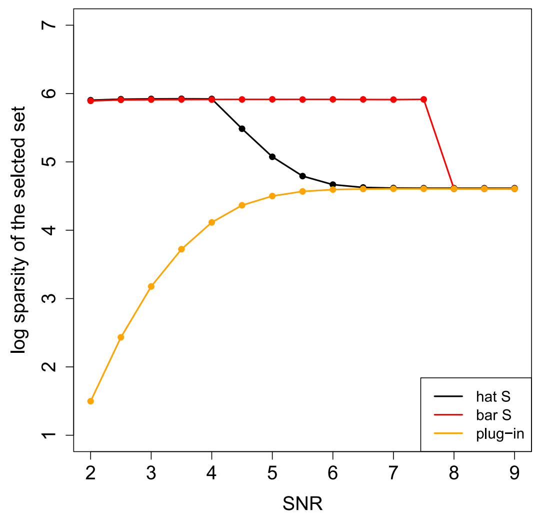


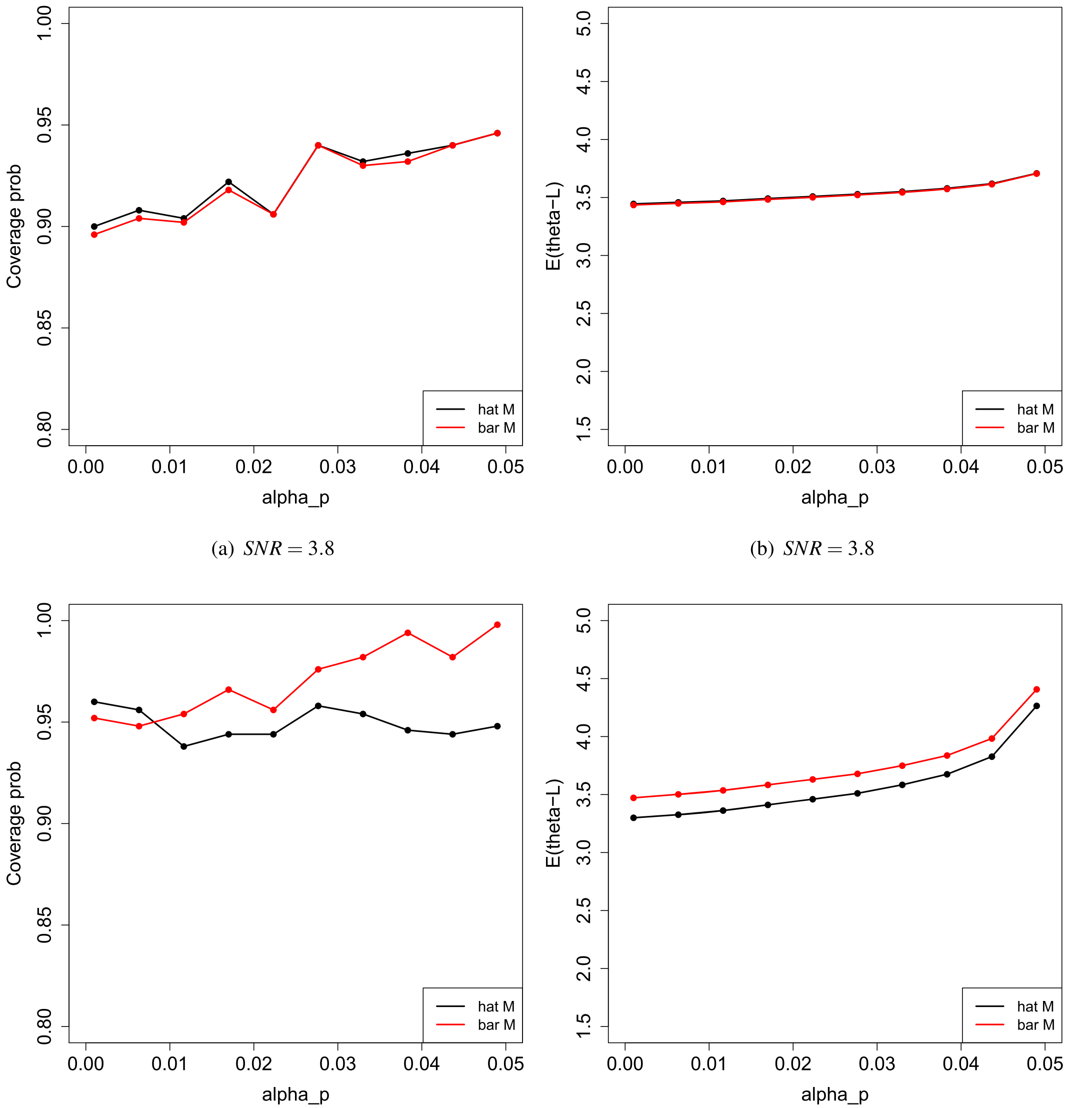
FIG. 2. Plot of $\log |\widehat{S}_{\alpha'}|$, $\log |\bar{S}_{\alpha'}|$ and $\log |\widehat{S}|$ averaged over 500 simulations.

Remark 2, in finite sample α' influences both the coverage probability and the distance $(\theta_j - L_j)$ of the proposed sparse confidence sets. As seen in Fig. 3, it seems they are not very sensitive to the choice of α'. For this reason, we simply take $\alpha' = \alpha/2$ in the previous simulations, leading to satisfactory numerical results.

7. Discussion

In this work, we propose a new framework to construct sparse confidence sets for the parameter $\boldsymbol{\theta}$ in the normal mean model. We first study the existence of such sparse confidence sets by establishing a non-asymptotic minimax lower bound for the non-coverage probability over a suitable class of sparse confidence sets. We further propose a two-step procedure to construct the sparse confidence set, and show that the resulting confidence set attains the minimax lower bound of the maximum expected length of confidence intervals up to a constant factor. Our optimality property is studied in the asymptotic regime as $d, s \to \infty$, and we treat the TNR level δ as a fixed parameter pre-specified by the users. If a high TNR level is desirable (e.g. $\delta = 0.999$), the asymptotic analysis can be conducted by further assuming $\delta \to 1$. However, as $\delta \to 1$, Theorem 1 implies that we have to require stronger minimum signal strength conditions to construct sparse confidence sets. When these conditions are not satisfied, the proposed confidence sets may not have the correct coverage probability.

One future research question is to study how to construct sparse confidence balls, which can be defined as $M(S, \boldsymbol{U}, R) = \{\boldsymbol{\theta} \in \mathbb{R}^d : \boldsymbol{\theta}_{S^c} = 0 \text{ and } \|\boldsymbol{\theta}_S - \boldsymbol{U}\|_2 \le R\}$, where $S \subseteq [d]$ is the selected set, $\boldsymbol{U}$ is the centre of the ball in $\mathbb{R}^{|S|}$ and R is the radius. In this case, it may be more appropriate to define the optimal sparse confidence ball based on the volume of the region. It is of interest to establish the minimax properties of sparse confidence balls. Another direction is to extend the minimax results to general Gaussian model $\boldsymbol{X} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ for some non-diagonal matrix $\boldsymbol{\Sigma}$. We show in Appendix C that the proposed sparse confidence sets still satisfy the desired conditions (1.5) and (1.6) under this more



(a) $SNR = 3.8$ (b) $SNR = 3.8$

(c) $SNR = 9$ (d) $SNR = 9$

FIG. 3. Sensitivity analysis of coverage probability and the average distance $(\theta_j - L_j)$ with respect to α' (alpha_p).

general model. However, the lower bound results in this work are valid only when the entries of $\boldsymbol{X}$ are independent. A rigorous minimax analysis is needed to close this gap.

Acknowledgments

The authors thank the anonymous reviewers for their valuable suggestions.

Funding

This work is supported in part by funds from the National Science Foundation (NSF: CAREER Award DMS-1941945, DMS-1854637, DMS-1811812, DMS-1821183), Office of Naval Research (N00014-22-1-2680) and Adobe Data Science Award.

Data Availability Statement

No new data were generated or analysed in support of this review.

1. ABRAMOVICH, F., BENJAMINI, Y., DONOHO, D. L. & JOHNSTONE, I. M. (2006) Adapting to unknown sparsity by controlling the false discovery rate. *Ann Stat*, **34**, 584–653.
2. ARIAS-CASTRO, E., CANDÈS, E. J. & PLAN, Y. (2011) Global testing under sparse alternatives: Anova, multiple comparisons and the higher criticism. *Ann Stat*, **39**, 2533–2556.
3. ARLOT, S., BLANCHARD, G. & ROQUAIN, E. (2010) Some nonasymptotic results on resampling in high dimension, i: confidence regions. *Ann Stat*, **38**, 51–82.
4. BARAUD, Y. (2002) Non-asymptotic minimax rates of testing in signal detection. *Bernoulli*, 577–606.
5. BELLONI, A., CHERNOZHUKOV, V. & KATO, K. (2014) Uniform post-selection inference for least absolute deviation regression and other z-estimation problems. *Biometrika*, asu056.
6. BENJAMINI, Y. & YEKUTIELI, D. (2005) False discovery rate–adjusted multiple confidence intervals for selected parameters. *J Am Stat Assoc*, **100**, 71–81.
7. BENJAMINI, Y., HECHTLINGER, Y. & STARK, P. B. (2019) Confidence intervals for selected parameters. *arXiv preprint arXiv:190600505*.
8. BERK, R., BROWN, L., BUJA, A., ZHANG, K. & ZHAO, L. (2013) Valid post-selection inference. *Ann Stat*, **41**, 802–837.
9. BIRGÉ, L. (2001) An alternative point of view on lepski's method. *Lecture Notes-Monograph Series*, 113–133.
10. BUTUCEA, C., NDAOUD, M., STEPANOVA, N. A. & TSYBAKOV, A. B. (2018) Variable selection with hamming loss. *Ann Stat*, **46**, 1837–1875.
11. CAI, T. T. & GUO, Z. (2017) Confidence intervals for high-dimensional linear regression: minimax rates and adaptivity. *Ann Stat*, **45**, 615–646.
12. CARPENTIER, A. (2015) Implementable confidence sets in high dimensional regression. *In Artificial Intelligence and Statistics*, 120–128.
13. CHERNOZHUKOV, V., CHETVERIKOV, D. & KATO, K. (2013) Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann Stat*, **41**, 2786–2819.
14. FAN, J. & LI, R. (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. *J Am Stat Assoc*, **96**, 1348–1360.
15. FAN, J., XUE, L. & ZOU, H. (2014) Strong oracle optimality of folded concave penalized estimation. *Ann Stat*, **42**, 819.
16. FUENTES, C., CASELLA, G. & WELLS, M. T. (2018) Confidence intervals for the means of the selected populations. *Electron J Stat*, **12**, 58–79.
17. HALL, P. & JIN, J. (2010) Innovated higher criticism for detecting sparse signals in correlated noise. *Ann Stat*, **38**, 1686–1732.
18. INGSTER, Y. I., TSYBAKOV, A. B. & VERZELEN, N. (2010) Detection boundary in sparse regression. *Electron J Stat*, **4**, 1476–1526.
19. JAVANMARD, A. & MONTANARI, A. (2013) Confidence intervals and hypothesis testing for high-dimensional regression. *arXiv preprint arXiv:13063171*.
20. KOSOROK, M. R. (2007) *Introduction to empirical processes and semiparametric inference*. Springer Science & Business Media.
21. LEE, J. D., SUN, D. L., SUN, Y. & TAYLOR, J. E. (2016) Exact post-selection inference, with application to the lasso. *Ann Stat*, **44**, 907–927.

22. Lepskii, O. (1991) On a problem of adaptive estimation in gaussian white noise. *Theory Probab Appl*, **35**, 454–466.
23. Lepskii, O. (1992) Asymptotically minimax adaptive estimation. I: upper bounds. Optimally adaptive estimates. *Theory Probab Appl*, **36**, 682–697.
24. Neykov, M., Ning, Y., Liu, J. S. & Liu, H. (2018) A unified theory of confidence regions and testing for high-dimensional estimating equations. *Stat Sci*, **33**, 427–443.
25. Nickl, R. & Van De Geer, S. (2013) Confidence sets in sparse regression. *Ann Stat*, **41**, 2852–2876.
26. Ning, Y. & Liu, H. (2017) A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *Ann Stat*, **45**, 158–195.
27. Ning, Y., Zhao, T. & Liu, H. (2017) A likelihood ratio framework for high-dimensional semiparametric regression. *Ann Stat*, **45**, 2299–2327.
28. Vaart, A. W. & Wellner, J. A. (1996) *Weak convergence and empirical processes: with applications to statistics*. Springer.
29. Geer van de, Bühlmann, P., Ritov, Y. & Dezeure, R. (2014) On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann Stat*, 1166–1202.
30. Wang, L., Kim, Y. & Li, R. (2013) Calibrating non-convex penalized regression in ultra-high dimension. *Ann Stat*, **41**, 2505.
31. Weinstein, A., Fithian, W. & Benjamini, Y. (2013) Selection adjusted confidence intervals with more power to determine the sign. *J Am Stat Assoc*, **108**, 165–176.
32. Zhang, C.-H. & Zhang, S. S. (2014) Confidence intervals for low dimensional parameters in high dimensional linear models. *J R Stat SocB*, **76**, 217–242.
33. Zhang, X. & Cheng, G. (2017) Simultaneous inference for high-dimensional linear models. *J Am Stat Assoc*, **112**, 757–768.
34. Zhao, H. & Cui, X. (2020) Constructing confidence intervals for selected parameters. *Biometrics*.

A. Extension to Two-sided Sparse Confidence Sets

In this section, we assume that $\boldsymbol{\theta} \in \Theta(s, a)$, where

$$\Theta(s,a) = \{\boldsymbol{\theta} \in \mathbb{R}^d : \|\boldsymbol{\theta}\|_0 \le s, \min_{j:\theta_j \neq 0} |\theta_j| \ge a > 0\}.$$

The goal is to generalize the results in Sections 2 and 3 to two-sided sparse confidence intervals for $\boldsymbol{\theta}$ in $\Theta(s, a)$. To this end, consider the following estimator of the support set:

$$\widehat{S}^{TS}_{\alpha'} = \left\{ j \in [d] : |X_j|/\sigma \ge \left(\Phi^{-1}\left(\frac{\alpha'}{2s}\right) + a/\sigma \right)_+ \vee \Phi^{-1}\left(\frac{1+\delta}{2}\right) \right\}, \tag{A.1}$$

where α' is the tolerance level. Similarly, we require $|X_j|/\sigma \ge \Phi^{-1}((1+\delta)/2)$ to guarantee the resulting confidence interval is sparse, i.e. $\widehat{S}^{TS}_{\alpha'} \in \mathcal{F}(\delta)$, where $\mathcal{F}(\delta)$ is defined in (2.1).

The following theorem, which is parallel to Theorems 1 and 3, establishes the upper and lower bounds of the non-coverage probability $\mathbb{P}_{\boldsymbol{\theta}}(\mathrm{supp}(\boldsymbol{\theta}) \not\subseteq \widehat{S})$ under $\Theta(s, a)$.

Theorem 10.

(1) For any $s \ge 1$ and $0 < \delta < 1$, we have

$$\inf_{\widehat{S} \in \mathcal{F}(\delta)} \sup_{\boldsymbol{\theta} \in \Theta(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\mathrm{supp}(\boldsymbol{\theta}) \not\subseteq \widehat{S}) \ge 1 - \frac{1}{(\Delta_{TS} + 1)^s}, \tag{A.2}$$

where $\Delta_{TS} = \Phi(\Phi^{-1}(\frac{1+\delta}{2}) + \frac{a}{\sigma}) - \Phi(-\Phi^{-1}(\frac{1+\delta}{2}) + \frac{a}{\sigma})$.

(2) Assume that $s, d \to \infty$. Let c_s be a sequence satisfying $c_s \to \infty$ and $c_s/s \to 0$. Assume that $\delta \geq c$ for some constant $c > 0$. If

$$a/\sigma \leq \phi_* := \Phi^{-1}\left(\frac{1+\delta}{2}\right) - \Phi^{-1}\left(\frac{c_s}{s}\right), \tag{A.3}$$

we have

$$\liminf_{d,s\to\infty} \inf_{\widehat{S}\in\mathcal{F}(\delta)} \sup_{\boldsymbol{\theta}\in\Theta(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\mathrm{supp}(\boldsymbol{\theta}) \not\subseteq \widehat{S}) = 1. \tag{A.4}$$

(3) For any $0 < \alpha' < 1$, it holds that $\widehat{S}^{TS}_{\alpha'} \in \mathcal{F}(\delta)$. In addition, if

$$\frac{a}{\sigma} \geq \phi^* := \Phi^{-1}\left(\frac{\delta+1}{2}\right) - \Phi^{-1}(\frac{\alpha'}{2s}) \tag{A.5}$$

holds, then

$$\sup_{\boldsymbol{\theta}\in\Theta(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\mathrm{supp}(\boldsymbol{\theta}) \not\subseteq \widehat{S}^{TS}_{\alpha'}) \leq \alpha'. \tag{A.6}$$

Note that in part (2), we require δ to be bounded away from 0 by a constant. To see the reason, consider the extreme case $\delta = 0$, which further implies $\Delta_{TS} = 0$. In this case, the lower bound in (A.2) becomes 0, which is no longer informative.

In view of (A.3) and (A.5), we observe a similar phase transition phenomenon under the parameter space $\Theta(s, a)$; see Remark 1 for details.

Given the index set $\widehat{S}^{TS}_{\alpha'}$, we define the two-sided sparse confidence set for $\boldsymbol{\theta} \in \Theta(s, a)$ as

$$\widehat{M}^{TS}_{\alpha'} = M(\widehat{S}^{TS}_{\alpha'}, \widehat{\boldsymbol{U}}^{TS}, \widehat{\boldsymbol{L}}^{TS}), \quad \text{where } \widehat{L}^{TS}_j = X_j - \widehat{u}^{TS}_{\alpha'}\sigma, \ \ \widehat{U}^{TS}_j = X_j + \widehat{u}^{TS}_{\alpha'}\sigma$$

for any $j \in \widehat{S}^{TS}_{\alpha'}$ and

$$\widehat{u}^{TS}_{\alpha'} = \begin{cases} \Phi^{-1}\Big(1 - \frac{\alpha-\alpha'}{2d}\Big) & \text{if } \phi^* \leq \frac{a}{\sigma} < \phi^* \vee \Big[-\Phi^{-1}(\frac{\alpha-\alpha'}{2d}) - \Phi^{-1}(\frac{\alpha'}{2s})\Big], \\ \Phi^{-1}\Big(1 - \frac{\alpha-\alpha'-2(d-s)(1-\eta)}{2s}\Big) & \text{if } \frac{a}{\sigma} \geq \phi^* \vee \Big[-\Phi^{-1}(\frac{\alpha-\alpha'}{2d}) - \Phi^{-1}(\frac{\alpha'}{2s})\Big], \end{cases}$$

where $\eta = \Phi(a/\sigma + \Phi^{-1}(\alpha'/(2s)))$.

The following theorem shows that $\widehat{M}^{TS}_{\alpha'}$ satisfies the conditions (1.5) and (1.6).

THEOREM 11. For any given level $0 < \alpha' < \alpha$, provided (A.5) holds, we have

$$\sup_{\boldsymbol{\theta}\in\Theta(s,a),\theta_j=0} \mathbb{P}_{\boldsymbol{\theta}}(j \in \widehat{S}^{TS}_{\alpha'}) \leq 1 - \delta, \qquad \sup_{\boldsymbol{\theta}\in\Theta(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin \widehat{M}^{TS}_{\alpha'}) \leq \alpha.$$

We can develop a similar framework as in Section 3 to study the optimality of the two-sided sparse confidence intervals. To this end, define the class of two-sided confidence sets as

$$CI = \{M(S, \boldsymbol{U}, \boldsymbol{L}) : L_j, U_j \text{ only depend on } X_j, L_j \leq U_j, \text{and for } j \notin S, L_j = U_j = 0\}.$$

To evaluate the optimality, it boils down to investigate the trade-off between the length of the interval $M(S,\boldsymbol{U},\boldsymbol{L}) \in CI$, i.e. $\sup_{1\le j\le d} \mathbb{E}_{\boldsymbol{\theta}}(U_j - L_j)$, and its coverage probability. Define

$$\mathcal{M}(m,\delta) = \Big\{M(S,\boldsymbol{U},\boldsymbol{L}) \in CI : \sup_{1\le j\le d} \sup_{\boldsymbol{\theta}\in\Theta(s,a)} \mathbb{E}_{\boldsymbol{\theta}}(U_j - L_j) \le m, \text{ and } S \in \mathcal{F}(\delta)\Big\},$$

to be the class of confidence sets such that the length is no greater than m uniformly over $1 \le j \le d$ and $\boldsymbol{\theta} \in \Theta(s,a)$ and $S \in \mathcal{F}(\delta)$ holds as defined in (2.1).

The following theorem, parallel to Theorem 6, provides the lower bound for the non-coverage probability of $M \in \mathcal{M}(m,\delta)$.

Theorem 12. (Minimax lower bound). For any $s \ge 1$ and $M \in \mathcal{M}(m,\delta)$, it holds that

$$\sup_{\boldsymbol{\theta}\in\Theta(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin M) \ge \max\Big(\sup_{\rho\ge a, A\le s} G_{TS}(d,A,\rho,m), \sup_{\rho\ge 0, B\le s} G_{TS}(s,B,\rho,m), 1 - \frac{1}{(\Delta_{TS}+1)^s}\Big), \quad \text{(A.7)}$$

where Δ_{TS} is defined in Theorem 10,

$$G_{TS}(d,A,\rho,m) = \frac{A[g_{TS}(d,A,\rho) - m/\rho]_+}{1 + A[g_{TS}(d,A,\rho) - m/\rho]_+},$$

with

$$g_{TS}(d,A,\rho) = \frac{2(d-A)}{A}\Phi(-D) + \Phi\Big(\frac{\rho}{\sigma} + D\Big) - \Phi\Big(\frac{\rho}{\sigma} - D\Big),$$

and

$$D = \frac{\sigma}{\rho}\cosh^{-1}\left(\frac{d-A}{A}\exp(\frac{\rho^2}{2\sigma^2})\right),$$

and $G_{TS}(s,B,\rho,m)$ is defined similarly. Note that $\cosh(x) = \exp(x)/2 + \exp(-x)/2$ and $\cosh^{-1}$ is the inverse function of $\cosh(x)$ on $\mathbb{R}^+$.

In practice, we usually pre-specify the coverage probability of the confidence set. Define

$$\mathcal{M} = \{M(S,\boldsymbol{U},\boldsymbol{L}) \in CI : \liminf_{d,s\to\infty} \inf_{\boldsymbol{\theta}\in\Theta(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \in M(S,\boldsymbol{U},\boldsymbol{L})) \ge 1-\alpha, \text{ and } S \in \mathcal{F}(\delta)\}$$

to be the two-sided sparse confidence sets with coverage probability no smaller than $1-\alpha$. We can similarly invert Theorem 12 to derive the lower bound for the length of confidence intervals $\sup_{1\le j\le d} \mathbb{E}_{\boldsymbol{\theta}}(U_j - L_j)$ of $M \in \mathcal{M}$. To match the lower bound, we consider the asymptotic version of $\widehat{M}^{TS}_{\alpha'}$. Define

$$\phi^{**} = \Phi^{-1}(\frac{1+\delta}{2}) + \sqrt{2\log\left(\frac{2s}{C_{2s,\alpha'}\alpha'}\right)},$$

and

$$\bar{\phi} = \sqrt{2\log\left(\frac{4(d-s)}{(\alpha-\alpha')C_{2(d-s),\alpha-\alpha'}}\right)} + \sqrt{2\log\left(\frac{2s}{C_{2s,\alpha'}\alpha'}\right)},$$

where $C_{s,\alpha'} = 2(\pi \log(s/\alpha'))^{1/2}$. Define

$$\bar{M}^{TS}_{\alpha'} = M(\bar{S}^{TS}_{\alpha'}, \bar{\boldsymbol{U}}^{TS}, \bar{\boldsymbol{L}}^{TS}), \text{ where } \bar{L}^{TS}_j = X_j - \bar{u}^{TS}_{\alpha'}\sigma,\ \bar{U}^{TS}_j = X_j + \bar{u}^{TS}_{\alpha'}\sigma \tag{A.8}$$

for $j \in \bar{S}^{TS}_{\alpha'}$, where $\bar{S}^{TS}_{\alpha'}$ and $\bar{u}^{TS}_{\alpha'}$ are given as follows:

- When $\phi^{**} \le a/\sigma < \bar{\phi}$, define $j \in \bar{S}^{TS}_{\alpha'}$ if and only if $|X_j/\sigma| \ge \Phi^{-1}((\delta+1)/2)$, and

$$\bar{u}^{TS}_{\alpha'} = \sqrt{2\log\left(\frac{2d}{(\alpha-\alpha')C_{2d,\alpha-\alpha'}}\right)}. \tag{A.9}$$

- When $a/\sigma \ge \bar{\phi}$, define $j \in \bar{S}^{TS}_{\alpha'}$ if and only if $|X_j/\sigma| \ge \sqrt{2\log(\frac{4(d-s)}{(\alpha-\alpha')C_{2(d-s),\alpha-\alpha'}})}$, and

$$\bar{u}^{TS}_{\alpha'} = \sqrt{2\log\left(\frac{4s}{(\alpha-\alpha')C_{2s,\alpha-\alpha'}}\right)}. \tag{A.10}$$

Similar to Corollary 5, we can show that

$$\limsup_{d,s\to\infty} \sup_{\boldsymbol{\theta}\in\Theta(s,a),\theta_j=0} \mathbb{P}_{\boldsymbol{\theta}}(j \in \bar{S}^{TS}_{\alpha'}) \le 1-\delta,\ \limsup_{d,s\to\infty} \sup_{\boldsymbol{\theta}\in\Theta(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin \bar{M}^{TS}_{\alpha'}) \le \alpha.$$

Finally, in the following corollary, we establish the optimality of $\bar{M}^{TS}_{\alpha'}$ within the class $\mathcal{M}$.

COROLLARY 13. Assume that $d, s \to \infty$ and $0 < \delta, \alpha < 1$ are fixed.

(1). If $\phi^{**} \le a/\sigma \le \sqrt{2\log(d/A_d - 1)}$ for some sequence $A_d \le s$ satisfying $A_d \to \infty$ and $d/A_d \to \infty$, then

$$\liminf_{d,s\to\infty} \inf_{M\in\mathcal{M}} \frac{\sup_{1\le j\le d}\sup_{\boldsymbol{\theta}\in\Theta(s,a)} \mathbb{E}_{\boldsymbol{\theta}}(U_j - L_j)}{\sigma\sqrt{2\log d}/2} \ge 1. \tag{A.11}$$

For $\bar{M}^{TS}_{\alpha'}$ with $\alpha' = \gamma\alpha$ for any constant $0 < \gamma < 1$, we have $\bar{M}^{TS}_{\alpha'} \in \mathcal{M}$ and

$$\limsup_{d,s\to\infty} \frac{\sup_{1\le j\le d}\sup_{\boldsymbol{\theta}\in\Theta(s,a)} \mathbb{E}_{\boldsymbol{\theta}}(\bar{U}^{TS}_j - \bar{L}^{TS}_j)}{2\sigma\sqrt{2\log d}} \le 1. \tag{A.12}$$

(2). If $a/\sigma \ge \sqrt{2\log(d-s) - \log\log(d-s) + C'} + \sqrt{2\log s - \log\log s + C'}$ for some sufficiently large positive constant C', then

$$\liminf_{d,s\to\infty} \inf_{M\in\mathcal{M}} \frac{\sup_{1\le j\le d}\sup_{\boldsymbol{\theta}\in\Theta(s,a)} \mathbb{E}_{\boldsymbol{\theta}}(U_j - L_j)}{\sigma\sqrt{2\log s}/2} \ge 1. \tag{A.13}$$

The sparse confidence set $\bar{M}^{TS}_{\alpha'}$ satisfies $\bar{M}^{TS}_{\alpha'} \in \mathcal{M}$ and

$$\limsup_{d,s\to\infty} \frac{\sup_{1\le j\le d} \sup_{\boldsymbol{\theta}\in\Theta(s,a)} \mathbb{E}_{\boldsymbol{\theta}}(\bar{U}^{TS}_j - \bar{L}^{TS}_j)}{2\sigma\sqrt{2\log s}} \le 1. \tag{A.14}$$

B. Proof

Note that, for notational simplicity, the constant C may differ from line to line in the proof. For a set S, we use $\boldsymbol{\theta}_S$ to be the subvector of $\boldsymbol{\theta}$ with index belonging to S. Under the separable rule, sometimes we simply use the notation $\mathbb{P}_{\theta_j}$ to denote the probability under θ_j.

B.1 *Proof of Theorem 1*

Denote $\partial\Theta^+(s,a) = \{\boldsymbol{\theta} \in \mathbb{R}^d : \|\boldsymbol{\theta}\|_0 = s, \theta_j = a \text{ for } \forall j, \theta_j \ne 0\}$. We know that

$$\sup_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\mathrm{supp}(\boldsymbol{\theta}) \not\subseteq \widehat{S}) \ge \sup_{\boldsymbol{\theta}\in\partial\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\mathrm{supp}(\boldsymbol{\theta}) \not\subseteq \widehat{S}) = 1 - \inf_{\boldsymbol{\theta}\in\partial\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\mathrm{supp}(\boldsymbol{\theta}) \subseteq \widehat{S}). \tag{B1}$$

Since $\mathrm{supp}(\boldsymbol{\theta}) \subseteq \widehat{S}$ is equivalent to the fact $j \in \widehat{S}$ for any $j \in \mathrm{supp}(\boldsymbol{\theta})$, we have $\mathbb{P}_{\boldsymbol{\theta}}(\mathrm{supp}(\boldsymbol{\theta}) \subseteq \widehat{S}) = \mathbb{P}_{\boldsymbol{\theta}}(\cap_{j\in\mathrm{supp}(\boldsymbol{\theta})}\{j \in \widehat{S}\}) = \prod_{j\in\mathrm{supp}(\boldsymbol{\theta})} \mathbb{P}_{\theta_j}(j \in \widehat{S})$, where the last step follows from the definition of the set $\mathcal{F}(\delta)$. For notational simplicity, we denote $S = \mathrm{supp}(\boldsymbol{\theta})$. We have

$$\begin{aligned}
\sup_{\boldsymbol{\theta}\in\partial\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(S \not\subseteq \widehat{S}) &\ge \frac{1}{|\partial\Theta^+(s,a)|} \sum_{\boldsymbol{\theta}\in\partial\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(S \not\subseteq \widehat{S}) \\
&= \frac{1}{|\partial\Theta^+(s,a)|} \sum_{\boldsymbol{\theta}\in\partial\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\cup_{j\in S}\{j \notin \widehat{S}\}) \\
&= \frac{1}{|\partial\Theta^+(s,a)|} \sum_{\boldsymbol{\theta}\in\partial\Theta^+(s,a)} \Big(\sum_{j_1\in S} \mathbb{P}_{\theta_{j_1}}(j_1 \notin \widehat{S}) \prod_{j\ne j_1\in S} \mathbb{P}_{\theta_j}(j \in \widehat{S}) \\
&\quad + \sum_{j_1\ne j_2\in S} \mathbb{P}_{\theta_{j_1}}(j_1 \notin \widehat{S})\mathbb{P}_{\theta_{j_2}}(j_2 \notin \widehat{S}) \prod_{j\ne j_1,j_2\in S} \mathbb{P}_{\theta_j}(j \in \widehat{S}) + \ldots + \prod_{j\in S} \mathbb{P}_{\theta_j}(j \notin \widehat{S}) \Big) \\
&\ge \frac{t}{|\partial\Theta^+(s,a)|} \sum_{\boldsymbol{\theta}\in\partial\Theta^+(s,a)} \Big(\sum_{j_1\in S} \mathbb{P}_{\theta_{j_1}}(j_1 \notin \widehat{S}) + \sum_{j_1\ne j_2\in S} \mathbb{P}_{\theta_{j_1}}(j_1 \notin \widehat{S})\mathbb{P}_{\theta_{j_2}}(j_2 \notin \widehat{S}) \\
&\quad + \ldots + \prod_{j\in S} \mathbb{P}_{\theta_j}(j \notin \widehat{S}) \Big),
\end{aligned} \tag{B2}$$

where $t = \inf_{\boldsymbol{\theta}\in\partial\Theta^+(s,a)} \prod_{j\in S} \mathbb{P}_{\theta_j}(j \in \widehat{S})$. Define $u = \mathbb{P}_a(j \notin \widehat{S})$, where $\mathbb{P}_a$ denotes the probability of $X_j \sim N(a, \sigma^2)$. Note that $|\partial\Theta^+(s,a)| = \binom{d}{s}$. Consider the kth term in (B.2) ($1 \le k \le s$),

$$\begin{aligned}
&\frac{t}{|\partial\Theta^+(s,a)|} \sum_{\boldsymbol{\theta}\in\partial\Theta^+(s,a)} \sum_{j_1\ne j_2..\ne j_k\in S} \prod_{m=1}^{k} \mathbb{P}_{\theta_{j_m}}(j_m \notin \widehat{S}) \\
&= t\binom{d}{s}^{-1}\binom{d}{k}\binom{d-k}{s-k}u^k = t\binom{s}{k}u^k.
\end{aligned}$$

Thus, by taking the infimum, (B.2) reduces to

$$\inf_{\widehat{S}\in\mathcal{F}(\delta)} \sup_{\boldsymbol{\theta}\in\partial\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(S \not\subseteq \widehat{S}) \geq \inf_{\widehat{S}\in\mathcal{F}(\delta)} t\sum_{k=1}^{s}\binom{s}{k}u^k = \inf_{\widehat{S}\in\mathcal{F}(\delta)} t[(1+u)^s - 1]. \tag{B.3}$$

Next, we consider the infimum of $(1+u)^s$ over all possible $\widehat{S}\in\mathcal{F}(\delta)$. Then

$$\inf_{\widehat{S}\in\mathcal{F}(\delta)} (1+\mathbb{P}_a(j\notin\widehat{S}))^s = (1+\inf_{\widehat{S}\in\mathcal{F}(\delta)} \mathbb{P}_a(j\notin\widehat{S}))^s.$$

Since $j\in\widehat{S}$ only depends on X_j, we can denote $j\in\widehat{S}$ by $T(X_j)=1$ for some function $T(\cdot)$. Then Neyman-Pearson lemma implies that the infimum of $\mathbb{P}_a(T(X_j)=0)$ over all possible $T(\cdot)$ such that $\mathbb{P}_0(T(X_j)=1)\leq 1-\delta$ is attained by the likelihood ratio test of $X_j\sim N(0,\sigma^2)$ versus $X_j\sim N(a,\sigma^2)$. After some simple calculation, we find that the optimal $T(X_j)$ is

$$T_{opt}(X_j) = I\Big(\frac{\phi(X_j-a)}{\phi(X_j)}\geq c\Big), \quad \text{where } c=\exp\Big\{\frac{a}{\sigma}\Phi^{-1}(\delta)-\frac{a^2}{2\sigma^2}\Big\}$$

and $\phi(\cdot)$ is the pdf of the standard normal distribution. With this $T_{opt}(X_j)$, $\inf_{\widehat{S}\in\mathcal{F}(\delta)}\mathbb{P}_a(j\notin\widehat{S})=\Delta$, where $\Delta=\Phi(\Phi^{-1}(\delta)-\frac{a}{\sigma})$. Plugging into (B.3), we obtain

$$\inf_{\widehat{S}\in\mathcal{F}(\delta)} \sup_{\boldsymbol{\theta}\in\partial\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\mathrm{supp}(\boldsymbol{\theta}) \not\subseteq \widehat{S}) \geq t[(1+\Delta)^s-1].$$

As $\inf_{\widehat{S}\in\mathcal{F}(\delta)}\sup_{\boldsymbol{\theta}\in\partial\Theta^+(s,a)}\mathbb{P}(\mathrm{supp}(\boldsymbol{\theta})\not\subseteq\widehat{S})\geq 1-t$ holds by (B.1), optimizing over t we obtain

$$\inf_{\widehat{S}\in\mathcal{F}(\delta)} \sup_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\mathrm{supp}(\boldsymbol{\theta}) \not\subseteq \widehat{S}) \geq 1-\frac{1}{(\Delta+1)^s}.$$

This completes the proof of (2.2). By (2.3), $\Delta\geq c_s/s$. When $c_s/s\to 0$, $\log(1+c_s/s)>(1-\epsilon)c_s/s$ for some constant $0<\epsilon<1$. Thus,

$$(\Delta+1)^s = \exp(s\log(1+\Delta)) \geq \exp(s\log(1+c_s/s)) > \exp((1-\epsilon)c_s)\to\infty,$$

as $c_s\to\infty$ and $c_s/s\to 0$. Clearly, (2.4) follows from the non-asymptotic bound (2.2).

B.2 *Proof of Theorem 3*

To show $\widehat{S}_{\alpha'}\in\mathcal{F}(\delta)$, notice that

$$\mathbb{P}_0(j\in\widehat{S}_{\alpha'}) = \mathbb{P}_0\Big(X_j/\sigma \geq \max(\Phi^{-1}(\frac{\alpha'}{s})+a/\sigma, \Phi^{-1}(\delta))\Big) \leq \mathbb{P}_0(X_j/\sigma\geq\Phi^{-1}(\delta)) = 1-\delta.$$

The event $\text{supp}(\boldsymbol{\theta}) \not\subseteq \widehat{S}_{\alpha'}$ is equivalent to that there exists $j \in [d]$ such that $j \in \text{supp}(\boldsymbol{\theta})$ and $j \notin \widehat{S}_{\alpha'}$. Then

$$
\begin{aligned}
\mathbb{P}_{\boldsymbol{\theta}}(\text{supp}(\boldsymbol{\theta}) \not\subseteq \widehat{S}_{\alpha'}) &= \mathbb{P}_{\boldsymbol{\theta}}(\exists j \in [d], j \in \text{supp}(\boldsymbol{\theta}), j \notin \widehat{S}_{\alpha'}) \\
&\leq \sum_{j:\theta_j \neq 0} \mathbb{P}_{\theta_j}(j \notin \widehat{S}_{\alpha'}) \\
&= \sum_{j:\theta_j \neq 0} \mathbb{P}_{\theta_j}(X_j \leq \sigma \Phi^{-1}(\frac{\alpha'}{s}) + a),
\end{aligned}
$$

where the last line follows from the condition that $a \geq \sigma(\Phi^{-1}(\delta) - \Phi^{-1}(\alpha'/s))$. Since $X_j \sim N(\theta_j, \sigma^2)$, we have $\mathbb{P}(X_j \leq t) = \Phi(\frac{t-\theta_j}{\sigma})$. Plugging into the above expression, we obtain

$$
\mathbb{P}_{\boldsymbol{\theta}}(\text{supp}(\boldsymbol{\theta}) \not\subseteq \widehat{S}_{\alpha'}) \leq \sum_{j:\theta_j \neq 0} \Phi(\Phi^{-1}(\frac{\alpha'}{s}) + \frac{a-\theta_j}{\sigma}) \leq \|\boldsymbol{\theta}\|_0 \frac{\alpha'}{s} = \alpha',
$$

as $\theta_j \geq a$ for $\theta_j \neq 0$. This completes the proof of Theorem 3.

B.3 *Proof of Theorem 4*

We first note that $\sup_{\boldsymbol{\theta} \in \Theta^+(s,a), \theta_j = 0} \mathbb{P}_{\boldsymbol{\theta}}(j \in \widehat{S}_{\alpha'}) \leq 1 - \delta$ holds by Theorem 3. In the following text, we bound $\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin \widehat{M}_{\alpha'})$ by intersecting with the event $\text{supp}(\boldsymbol{\theta}) \subseteq \widehat{S}_{\alpha'}$,

$$
\begin{aligned}
\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin \widehat{M}_{\alpha'}) &\leq \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin \widehat{M}_{\alpha'}, \text{supp}(\boldsymbol{\theta}) \subseteq \widehat{S}_{\alpha'}) + \mathbb{P}_{\boldsymbol{\theta}}(\text{supp}(\boldsymbol{\theta}) \not\subseteq \widehat{S}_{\alpha'}) \\
&= \mathbb{P}_{\boldsymbol{\theta}}(\exists j \in \widehat{S}_{\alpha'}, \theta_j < \widehat{L}_j, \text{supp}(\boldsymbol{\theta}) \subseteq \widehat{S}_{\alpha'}) + \mathbb{P}_{\boldsymbol{\theta}}(\text{supp}(\boldsymbol{\theta}) \not\subseteq \widehat{S}_{\alpha'}) \\
&\leq \mathbb{P}_{\boldsymbol{\theta}}(\exists j \in \widehat{S}_{\alpha'}, \theta_j < \widehat{L}_j) + \mathbb{P}_{\boldsymbol{\theta}}(\text{supp}(\boldsymbol{\theta}) \not\subseteq \widehat{S}_{\alpha'}). \qquad \text{(B.4)}
\end{aligned}
$$

By Theorem 3 and $a/\sigma \geq \kappa^*$, $\mathbb{P}_{\boldsymbol{\theta}}(\text{supp}(\boldsymbol{\theta}) \not\subseteq \widehat{S}_{\alpha'}) \leq \alpha'$. The first term can be further bounded as

$$
\mathbb{P}_{\boldsymbol{\theta}}(\exists j \in \widehat{S}_{\alpha'}, \theta_j < \widehat{L}_j) \qquad \text{(B.5)}
$$

$$
\leq \mathbb{P}_{\boldsymbol{\theta}}(\exists j \in \text{supp}(\boldsymbol{\theta}), \theta_j < \widehat{L}_j) + \mathbb{P}_{\boldsymbol{\theta}}(\exists j \in \widehat{S}_{\alpha'} \backslash \text{supp}(\boldsymbol{\theta}), \theta_j < \widehat{L}_j) := I_1 + I_2. \qquad \text{(B.6)}
$$

Write u for $\widehat{u}_{\alpha'}$. For I_1, by noting that $\theta_j < \max(X_j - u\sigma, 0)$ implies $Z_j > u$ where $Z_j = \frac{X_j - \theta_j}{\sigma} \sim N(0, 1)$, we have

$$
I_1 \leq \sum_{j \in \text{supp}(\boldsymbol{\theta})} \mathbb{P}_{\boldsymbol{\theta}}(Z_j > u) = s(1 - \Phi(u)). \qquad \text{(B.7)}
$$

To bound I_2, noting that $j \notin \text{supp}(\boldsymbol{\theta})$ implying $\theta_j = 0$, we have

$$
I_2 = \mathbb{P}(\exists j \notin \text{supp}(\boldsymbol{\theta}), Z_j \geq \Phi^{-1}(\frac{\alpha'}{s}) + \frac{a}{\sigma}, Z_j > u) \leq \sum_{j \notin \text{supp}(\boldsymbol{\theta})} \mathbb{P}(Z_j \geq \Phi^{-1}(\frac{\alpha'}{s}) + \frac{a}{\sigma}, Z_j > u).
$$

To bound the last probability, we now consider the following two cases.

(1). When $a/\sigma \in R_L$, by setting $u = \Phi^{-1}(1 - \frac{\alpha-\alpha'}{d})$, we can easily verify that $\Phi^{-1}(\frac{\alpha'}{s}) + \frac{a}{\sigma} \leq u$. Thus, $I_2 \leq (d-s)(1-\Phi(u))$. Together with (B.4), (B.6), (B.7), we have

$$\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin \widehat{M}_{\alpha'}) \leq d(1-\Phi(u)) + \alpha' = \alpha.$$

(2). When $a/\sigma \in R_H$, by setting $u = \Phi^{-1}(1 - \frac{\alpha-\alpha'-(d-s)(1-\eta^+)}{s})$, we can easily verify that $\Phi^{-1}(\frac{\alpha'}{s}) + \frac{a}{\sigma} > u$. Thus, it implies $I_2 \leq (d-s)(1-\eta^+)$, and finally we have

$$\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin \widehat{M}_{\alpha'}) \leq (d-s)(1-\eta^+) + s(1-\Phi(u)) + \alpha' = \alpha.$$

B.4 *Proof of Corollary 5*

When $a/\sigma < \bar{\kappa}$, it holds that

$$\sup_{\boldsymbol{\theta}\in\Theta^+(s,a),\theta_j=0} \mathbb{P}_{\boldsymbol{\theta}}(j \in \bar{S}_{\alpha'}) = \mathbb{P}_{\theta_j=0}(X_j/\sigma \geq \Phi^{-1}(\delta)) = 1-\delta,$$

and when $a/\sigma \geq \bar{\kappa}$ we have

$$\sup_{\boldsymbol{\theta}\in\Theta^+(s,a),\theta_j=0} \mathbb{P}_{\boldsymbol{\theta}}(j \in \bar{S}_{\alpha'}) = \mathbb{P}_{\theta_j=0}\Big(X_j/\sigma \geq \sqrt{2\log(\frac{2(d-s)}{(\alpha-\alpha')C_{d-s,\alpha-\alpha'}})}\Big) \leq 1-\delta.$$

So, it also holds that $\sup_{\boldsymbol{\theta}\in\Theta^+(s,a),\theta_j=0} \mathbb{P}_{\boldsymbol{\theta}}(j \in \bar{S}_{\alpha'}) \leq 1-\delta$.

In the following text, we first focus on the case $a/\sigma < \bar{\kappa}$. Note that

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\theta}}(\mathrm{supp}(\boldsymbol{\theta}) \not\subseteq \bar{S}_{\alpha'}) &\leq \sum_{j:\theta_j\neq 0} \mathbb{P}_{\theta_j}(X_j \leq \sigma\Phi^{-1}(\delta)) \\
&\leq \sum_{j:\theta_j\neq 0} \mathbb{P}_{\theta_j}(\frac{X_j-\theta_j}{\sigma} \leq \frac{\sigma\Phi^{-1}(\delta)-a}{\sigma}) \\
&\leq \sum_{j:\theta_j\neq 0} \mathbb{P}_{\theta_j}\Big(\frac{X_j-\theta_j}{\sigma} \leq -\sqrt{2\log(\frac{s}{C_{s,\alpha'}\alpha'})}\Big),
\end{aligned}$$

where the last step follows from $a/\sigma \geq \Phi^{-1}(\delta) + \sqrt{2\log(\frac{s}{C_{s,\alpha'}\alpha'})}$. By the tail probability in Lemma 16, it yields for any $0 < \alpha' < \alpha$

$$\begin{aligned}
&\lim_{d,s\to\infty} \sup_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\mathrm{supp}(\boldsymbol{\theta}) \not\subseteq \bar{S}_{\alpha'}) - \alpha' \\
&\leq \lim_{d,s\to\infty} s\sqrt{\frac{2}{\pi}} \frac{1}{2\sqrt{2\log(\frac{s}{C_{s,\alpha'}\alpha'})}} \exp\Big(-\log(\frac{s}{C_{s,\alpha'}\alpha'})\Big) - \alpha' \\
&= \lim_{d,s\to\infty} \sqrt{\frac{2}{\pi}} \frac{\alpha' C_{s,\alpha'}}{2\sqrt{2\log(\frac{s}{C_{s,\alpha'}\alpha'})}} - \alpha' = 0. \qquad \text{(B.8)}
\end{aligned}$$

By the proof of Theorem 4, we can similarly show that

$$\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin \bar{M}_{\alpha'}) \leq d\Big(1 - \Phi\Big(\sqrt{2\log(\frac{d}{(\alpha-\alpha')C_{d,\alpha-\alpha'}})}\Big)\Big) + \mathbb{P}_{\boldsymbol{\theta}}(\mathrm{supp}(\boldsymbol{\theta}) \not\subseteq \bar{S}_{\alpha'}). \quad \text{(B.9)}$$

By taking the limit $d, s \to \infty$, similar to (B.8), the tail bound in Lemma 16 implies

$$\lim_{d,s\to\infty} \sup_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin \bar{M}_{\alpha'}) \leq (\alpha - \alpha') + \alpha' = \alpha.$$

When $a/\sigma \geq \bar{\kappa}$, it is easily seen that

$$\begin{aligned}\mathbb{P}_{\boldsymbol{\theta}}(\mathrm{supp}(\boldsymbol{\theta}) \not\subseteq \bar{S}_{\alpha'}) &\leq \sum_{j:\theta_j\neq 0} \mathbb{P}_{\theta_j}\Big(X_j/\sigma \leq \sqrt{2\log(\frac{2(d-s)}{(\alpha-\alpha')C_{d-s,\alpha-\alpha'}})}\Big) \\ &\leq \sum_{j:\theta_j\neq 0} \mathbb{P}_{\theta_j}\Big(\frac{X_j-\theta_j}{\sigma} \leq -\sqrt{2\log(\frac{s}{C_{s,\alpha'}\alpha'})}\Big). \end{aligned} \quad \text{(B.10)}$$

As a result, (B.8) still holds. In the following text, we consider two cases separately.

Case (1) $d \geq 2s$. Recall the way of controlling the term I_2 in the proof of Theorem 4. In this case, j is selected if $Z_j \geq t$, where

$$t = \sqrt{2\log(\frac{2(d-s)}{(\alpha-\alpha')C_{d-s,\alpha-\alpha'}})}.$$

With the monotonicity of the function $\log x - \frac{1}{2}\log\log x$, this term is no smaller than $\bar{u}_{\alpha'}$ as $d - s \geq s$. Thus, we can show that by the proof of Theorem 4,

$$\begin{aligned}\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin \bar{M}_{\alpha'}) &\leq s(1-\Phi(\bar{u}_{\alpha'})) + \sum_{j\notin \mathrm{supp}(\boldsymbol{\theta})} \mathbb{P}_0\big(X_j/\sigma \geq t\big) + \mathbb{P}_{\boldsymbol{\theta}}(\mathrm{supp}(\boldsymbol{\theta}) \not\subseteq \bar{S}_{\alpha'}) \\ &= s\Big(1 - \Phi\Big(\sqrt{2\log(\frac{2s}{(\alpha-\alpha')C_{s,\alpha-\alpha'}})}\Big)\Big) + (d-s)(1-\Phi(t)) + \mathbb{P}_{\boldsymbol{\theta}}(\mathrm{supp}(\boldsymbol{\theta}) \not\subseteq \bar{S}_{\alpha'}).\end{aligned}$$

Similar to (B.8),

$$\begin{aligned}&\lim_{d,s\to\infty} s(1-\Phi(\bar{u}_{\alpha'})) - \frac{\alpha-\alpha'}{2} \\ &= \lim_{d,s\to\infty} s\sqrt{\frac{2}{\pi}} \frac{1}{2\sqrt{2\log(\frac{2s}{C_{s,\alpha-\alpha'}(\alpha-\alpha')})}} \exp\Big(-\log(\frac{2s}{C_{s,\alpha-\alpha'}(\alpha-\alpha')})\Big) - \frac{\alpha-\alpha'}{2} = 0,\end{aligned}$$

and

$$\lim_{d,s\to\infty}(d-s)(1-\Phi(t))-\frac{\alpha-\alpha'}{2}$$
$$=\lim_{d,s\to\infty}(d-s)\sqrt{\frac{2}{\pi}}\frac{1}{2\sqrt{2\log(\frac{2(d-s)}{C_{d-s,\alpha-\alpha'}(\alpha-\alpha')})}}\exp\Big(-\log(\frac{2(d-s)}{C_{d-s,\alpha-\alpha'}(\alpha-\alpha')})\Big)-\frac{\alpha-\alpha'}{2}=0.$$

We obtain

$$\lim_{d,s\to\infty}\sup_{\boldsymbol{\theta}\in\Theta^+(s,a)}\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta}\notin\bar{M}_{\alpha'})\le\frac{\alpha-\alpha'}{2}+\frac{\alpha-\alpha'}{2}+\alpha'=\alpha.$$

Case (2) $d<2s$. Unlike the previous case, we now have $t<\bar{u}_{\alpha'}$. Thus,

$$\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta}\notin\bar{M}_{\alpha'})\le d(1-\Phi(\bar{u}_{\alpha'}))+\mathbb{P}_{\boldsymbol{\theta}}(\mathrm{supp}(\boldsymbol{\theta})\not\subseteq\bar{S}_{\alpha'})$$
$$=d\Big(1-\Phi\Big(\sqrt{2\log(\frac{2s}{(\alpha-\alpha')C_{s,\alpha-\alpha'}})}\Big)\Big)+\mathbb{P}_{\boldsymbol{\theta}}(\mathrm{supp}(\boldsymbol{\theta})\not\subseteq\bar{S}_{\alpha'}).$$

Note that

$$\lim_{d,s\to\infty}d(1-\Phi(\bar{u}_{\alpha'}))-(\alpha-\alpha')$$
$$=\lim_{d,s\to\infty}\frac{d}{2s}\sqrt{\frac{2}{\pi}}\frac{C_{s,\alpha-\alpha'}(\alpha-\alpha')}{2\sqrt{2\log(\frac{2s}{C_{s,\alpha-\alpha'}(\alpha-\alpha')})}}-(\alpha-\alpha')\le 0,$$

where we use $d<2s$ in the last step. This implies $\lim_{d,s\to\infty}\sup_{\boldsymbol{\theta}\in\Theta^+(s,a)}\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta}\notin\bar{M}_{\alpha'})\le\alpha$.

B.5 *Proof of Theorem 6*

Define $\Theta(A)=\{\boldsymbol{\theta}\in\mathbb{R}^d:\|\boldsymbol{\theta}\|_0=A,\theta_j=\rho,\ \text{for any}\theta_j\ne 0\}$, where $0<A\le s$ and ρ is an arbitrary positive quantity that is $\rho\ge a$. Then, $\Theta(A)$ is contained in the parameter space $\Theta^+(s,a)$. For any M in $\mathcal{M}_+(m,\delta)$, we use CI_j to denote the confidence interval for θ_j. Following the similar arguments in the proof of Theorem 1, we have

$$\sup_{\boldsymbol{\theta}\in\Theta^+(s,a)}\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta}\notin M)\ge\binom{d}{A}^{-1}\sum_{\boldsymbol{\theta}\in\Theta(A)}\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta}\notin M)$$
$$\ge\binom{d}{A}^{-1}\sum_{\boldsymbol{\theta}\in\Theta(A)}\Big(\sum_{j=1}^d\mathbb{P}_{\theta_j}(\theta_j\notin CI_j)\prod_{j'\ne j}\mathbb{P}_{\theta_{j'}}(\theta_{j'}\in CI_{j'})\Big)$$
$$\ge t\binom{d}{A}^{-1}\sum_{j=1}^d\sum_{\boldsymbol{\theta}\in\Theta(A)}\mathbb{P}_{\theta_j}(\theta_j\notin CI_j)$$
$$=t\binom{d}{A}^{-1}\sum_{j=1}^d\Big(\sum_{\boldsymbol{\theta}\in\Theta(A),\theta_j=0}\mathbb{P}_0(0\notin CI_j)+\sum_{\boldsymbol{\theta}\in\Theta(A),\theta_j=\rho}\mathbb{P}_\rho(\rho\notin CI_j)\Big)$$
$$=t\sum_{j=1}^d\Big(\frac{d-A}{d}\mathbb{P}_0(0\notin CI_j)+\frac{A}{d}\mathbb{P}_\rho(\rho\notin CI_j)\Big),\qquad\text{(B.11)}$$

where $t = \inf_{\boldsymbol{\theta}\in\Theta(A)} \prod_{j=1}^{d} \mathbb{P}_{\theta_j}(\theta_j \in CI_j)$. Furthermore, we can control the last term in (B.11) as follows:

$$\mathbb{P}_\rho(\rho \notin CI_j) \geq \mathbb{P}_\rho(L_j > \rho) = \mathbb{P}_\rho(L_j = 0) + [\mathbb{P}_\rho(L_j > \rho) - \mathbb{P}_\rho(L_j = 0)]. \tag{B.12}$$

Since $M \in \mathcal{M}_+(m, \delta)$ implies $\sup_{1\leq j\leq d} \sup_{\boldsymbol{\theta}\in\Theta(s,a)} \mathbb{E}_{\boldsymbol{\theta}}(\theta_j - L_j) \leq m$, by taking $\theta_j = \rho$ we have

$$\begin{aligned}\rho - m \leq \mathbb{E}_\rho L_j &\leq \rho\mathbb{P}_\rho(0 < L_j \leq \rho) + 2\rho\mathbb{P}_\rho(\rho < L_j \leq 2\rho) + \mathbb{E}_\rho L_j I(L_j > 2\rho)\\ &= \rho\mathbb{P}_\rho(0 < L_j \leq \rho) + 2\rho\mathbb{P}_\rho(\rho < L_j) + \mathbb{E}_\rho(L_j - 2\rho)I(L_j > 2\rho).\end{aligned}$$

Then, we can plug $\mathbb{P}_\rho(0 < L_j \leq \rho) = 1 - \mathbb{P}_\rho(L_j = 0) - \mathbb{P}_\rho(L_j > \rho)$ into the above display, which can reduce to

$$\mathbb{P}_\rho(L_j > \rho) - \mathbb{P}_\rho(L_j = 0) \geq -\frac{m + \mathbb{E}_\rho(L_j - 2\rho)I(L_j > 2\rho)}{\rho}. \tag{B.13}$$

Our next step is to upper bound $\mathbb{E}_\rho(L_j - 2\rho)I(L_j > 2\rho)$. Recall that we assume $L_j \leq X_j$ whenever $X_j \geq 0$. Thus,

$$\begin{aligned}\mathbb{E}_\rho(L_j - 2\rho)I(L_j > 2\rho) &\leq \mathbb{E}_\rho(X_j \vee 0 - 2\rho)I(X_j \vee 0 > 2\rho)\\ &= \mathbb{E}_\rho(X_j - 2\rho)I(X_j > 2\rho) = \sigma\mathbb{E}NI(N > \frac{\rho}{\sigma}) - \rho\mathbb{P}(N > \frac{\rho}{\sigma}),\end{aligned}$$

where $N \sim N(0, 1)$. By the tail bound in Lemma 16 and some simple algebra,

$$\begin{aligned}\mathbb{E}_\rho(L_j - 2\rho)I(L_j > 2\rho) &\leq \sqrt{\frac{1}{2\pi}}\sigma \exp\Big(-\frac{1}{2}(\frac{\rho}{\sigma})^2\Big) - \sqrt{\frac{2}{\pi}} \exp\Big(-\frac{1}{2}(\frac{\rho}{\sigma})^2\Big)\frac{\rho}{\frac{\rho}{\sigma} + \sqrt{4 + \frac{\rho^2}{\sigma^2}}}\\ &\leq \sqrt{\frac{1}{2\pi}}\sigma \exp\Big(-\frac{1}{2}(\frac{\rho}{\sigma})^2\Big)\frac{\sqrt{1 + 4(\sigma/\rho)^2} - 1}{\sqrt{1 + 4(\sigma/\rho)^2} + 1} := R.\end{aligned}$$

Combining with (B.11), (B.12) and (B.13), we have shown that

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin M) &\geq t\Big\{\sum_{j=1}^{d}\Big(\frac{d-A}{d}\mathbb{P}_0(0 \notin CI_j) + \frac{A}{d}\mathbb{P}_\rho(0 \in CI_j)\Big) - A\frac{m+R}{\rho}\Big\}\\ &\geq tA\Big\{\inf_{T\in\{0,1\}}\Big(\frac{d-A}{A}\mathbb{E}_0(1-T) + \mathbb{E}_\rho T\Big) - \frac{m+R}{\rho}\Big\}.\end{aligned}$$

Here, T denotes an arbitrary test function from $\mathbb{R}$ to $\{0, 1\}$. By the Neyman–Pearson lemma, the optimal test function is given by

$$T_{opt}(x) = I\Big(x \leq \frac{\rho}{2} + \frac{\sigma^2}{\rho}\log(\frac{d-A}{A})\Big).$$

With the optimal test function, lower bound reduces to

$$\sup_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta}\notin M) \ge tA\Big\{g(d,A,\rho) - \frac{m+R}{\rho}\Big\}_+,$$

where

$$g(d,A,\rho) = \frac{d-A}{A}\Phi\Big(-\frac{\rho}{2\sigma} - \frac{\sigma}{\rho}\log(\frac{d}{A}-1)\Big) + \Phi\Big(-\frac{\rho}{2\sigma} + \frac{\sigma}{\rho}\log(\frac{d}{A}-1)\Big).$$

We also notice that

$$\sup_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta}\notin M) \ge 1 - \inf_{\boldsymbol{\theta}\in\Theta(A)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta}\in M) = 1 - \inf_{\boldsymbol{\theta}\in\Theta(A)} \prod_{j=1}^{d} \mathbb{P}_{\theta_j}(\theta_j \in CI_j) = 1-t.$$

We then optimize the lower bound with respect to t, which leads to

$$\sup_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta}\notin M) \ge G(d,A,\rho,m).$$

As the above lower bound holds for any $0 < A \le s$ and $\rho \ge a$, we obtain

$$\sup_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta}\notin M) \ge \sup_{\rho\ge a, A\le s} G(d,A,\rho,m).$$

The rest of the proof focuses on showing

$$\sup_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta}\notin M) \ge \sup_{\rho\ge 0, B\le s} G(s,B,\rho,m). \tag{B.14}$$

We first define an s-dimensional vector $\boldsymbol{a} = (a,a,...,a) \in \mathbb{R}^s$, and $\Theta_s(B) = \{\boldsymbol{\theta}\in\mathbb{R}^s : \|\boldsymbol{\theta}-\boldsymbol{a}\|_0 = B, \theta_j = a+\rho \text{ for } \theta_j \ne a\}$, where $0 < B \le s$ and ρ is an arbitrary positive quantity. Then, we define the parameter set $\Theta'(B) = \{(\boldsymbol{\theta},0,...,0)\in\mathbb{R}^d : \boldsymbol{\theta}\in\Theta_s(B)\}$, which is contained in the parameter space $\Theta^+(s,a)$. In this case, we only perturb the parameters that are non-zero. Let $\boldsymbol{\theta}_{[s]}$ and $M_{[s]}$ denote the first s entries of $\boldsymbol{\theta}$ and the confidence intervals for $\boldsymbol{\theta}_{[s]}$. Similar to the previous argument, we can show that

$$\begin{aligned}
\sup_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta}\notin M) &\ge \sup_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_{[s]}\notin M_{[s]}) \\
&\ge \binom{s}{B}^{-1} \sum_{\boldsymbol{\theta}\in\Theta'(B)} \Big(\sum_{j=1}^{s} \mathbb{P}_{\theta_j}(\theta_j\notin CI_j) \prod_{j'\ne j\in[s]} \mathbb{P}_{\theta_{j'}}(\theta_{j'}\in CI_{j'})\Big) \\
&\ge t'\binom{s}{B}^{-1} \sum_{j=1}^{s} \Big(\sum_{\boldsymbol{\theta}\in\Theta'(B),\theta_j=a} \mathbb{P}_a(a\notin CI_j) + \sum_{\boldsymbol{\theta}\in\Theta'(B),\theta_j=a+\rho} \mathbb{P}_{a+\rho}(a+\rho\notin CI_j)\Big) \\
&= t'\sum_{j=1}^{s} \Big(\frac{s-B}{s}\mathbb{P}_a(a\notin CI_j) + \frac{B}{s}\mathbb{P}_{a+\rho}(a+\rho\notin CI_j)\Big),
\end{aligned}$$

where $t' = \inf_{\boldsymbol{\theta}\in\Theta'(B)} \prod_{j=1}^{s} \mathbb{P}_{\theta_j}(\theta_j \in CI_j)$. In addition,

$$\mathbb{P}_{a+\rho}(a+\rho \notin CI_j) - \mathbb{P}_{a+\rho}(a \in CI_j) \geq \mathbb{P}_{a+\rho}(L_j > a+\rho) - \mathbb{P}_{a+\rho}(0 \leq L_j < a).$$

To further lower bound the right-hand side of the above display, we notice that

$$\begin{aligned}
a+\rho-m &\leq \mathbb{E}_{a+\rho}L_j \\
&\leq a\mathbb{P}_{a+\rho}(0 \leq L_j \leq a) + (a+\rho)\mathbb{P}_{a+\rho}(a < L_j \leq a+\rho) \\
&\quad + (a+2\rho)\mathbb{P}_{a+\rho}(a+\rho < L_j \leq a+2\rho) + \mathbb{E}_{a+\rho}L_j I(L_j > a+2\rho) \\
&\leq a\mathbb{P}_{a+\rho}(0 \leq L_j \leq a) + (a+\rho)\{1 - \mathbb{P}_{a+\rho}(0 \leq L_j \leq a) - \mathbb{P}_{a+\rho}(L_j > a+\rho)\} \\
&\quad + (a+2\rho)\mathbb{P}_{a+\rho}(a+\rho < L_j) + \mathbb{E}_{a+\rho}(L_j - (a+2\rho))I(L_j > a+2\rho) \\
&= (a+\rho) - \rho\mathbb{P}_{a+\rho}(0 \leq L_j \leq a) + \rho\mathbb{P}_{a+\rho}(a+\rho < L_j) + \mathbb{E}_{a+\rho}(L_j - (a+2\rho))I(L_j > a+2\rho).
\end{aligned}$$

Thus, we have

$$\mathbb{P}_{a+\rho}(L_j > a+\rho) - \mathbb{P}_{a+\rho}(0 \leq L_j < a) \geq -\frac{m + \mathbb{E}_{a+\rho}(L_j - a - 2\rho)I(L_j > a+2\rho)}{\rho}.$$

Finally, we also note that

$$\sup_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_{[s]} \notin M_{[s]}) \geq 1 - \inf_{\boldsymbol{\theta}\in\Theta'(B)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_{[s]} \in M_{[s]}) = 1 - t'.$$

The rest of the proofs are similar and therefore we omit the details. Together with Theorem 1, we complete the proof.

B.6 *Proof of Corollary 7*

Denote $a_* = \sigma\kappa_*$. When $a \leq a_*$, Corollary 2 holds. When $a_* < a \leq a_1 := \sigma\sqrt{2\log(d/A_d - 1)}$, Theorem 6 implies

$$\inf_{M\in\mathcal{M}_+(m,\delta)} \sup_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin M) \geq G(d, A_d, a_1, m) \geq G(d, A_d, a_1, m^*),$$

where $m^* = \sigma(\frac{1}{2} - \frac{W_d}{A_d})\sqrt{2\log(d/A_d - 1)} + \frac{\sqrt{2}\sigma}{4\sqrt{\pi}}(1 - \frac{A_d}{d-A_d})$ with A_d, W_d defined as in the Corollary. By Lemma 16

$$g(d, A_d, a_1) = \frac{1}{2} + \frac{d-A_d}{A_d}\Phi(-\sqrt{2\log(d/A_d - 1)}) \geq \frac{1}{2} + \frac{1}{2\sqrt{\pi}}\frac{1}{\sqrt{\log(d/A_d - 1) + 2}}. \tag{B.15}$$

In addition,

$$\frac{m^*}{a_1} = \frac{1}{2} - \frac{W_d}{A_d} + \frac{1}{4\sqrt{\pi}}\frac{1}{\sqrt{\log(d/A_d - 1)}}\Big(1 - \frac{A_d}{d-A_d}\Big),$$

and for d, s large enough

$$\frac{R}{a_1} = \frac{1}{\sqrt{2\pi}} \frac{A_d}{d - A_d} \frac{1}{\sqrt{2\log(d/A_d - 1)}} \frac{\sqrt{1 + 2/\log(d/A_d - 1)} - 1}{\sqrt{1 + 2/\log(d/A_d - 1)} + 1}$$
$$\leq \frac{1}{4\sqrt{\pi}} \frac{1}{\sqrt{\log(d/A_d - 1)}} \frac{A_d}{d - A_d}.$$

Thus, for d large enough and as $\frac{d}{A_d} \to \infty$,

$$A_d[g(d, A_d, a_1) - (m^* + R)/a_1]$$
$$\geq A_d[\frac{1}{2\sqrt{\pi}} \frac{1}{\sqrt{\log(d/A_d - 1) + 2}} + \frac{W_d}{A_d} - \frac{1}{4\sqrt{\pi}} \frac{1}{\sqrt{\log(d/A_d - 1)}} \Big(1 - \frac{A_d}{d - A_d}\Big) - \frac{R}{a_1}]$$
$$= A_d[\frac{1}{2\sqrt{\pi}} \frac{1}{\sqrt{\log(d/A_d - 1) + 2}} + \frac{W_d}{A_d} - \frac{1}{4\sqrt{\pi}} \frac{1}{\sqrt{\log(d/A_d - 1)}}]$$
$$> W_d.$$

Since $W_d \to \infty$, we have

$$\liminf_{d,s\to\infty} A_d[g(d, A_d, a_1) - (m^* + R)/a_1] = +\infty,$$

which further implies $\liminf_{d,s\to\infty} G(d, A_d, a_1, m^*) = 1$.

Similarly, when $a \geq \sigma\sqrt{2\log(d/A - 1)}$, Theorem 6 implies

$$\inf_{M\in\mathcal{M}_+(m,\delta)} \sup_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin M) \geq G(s, B_s, \rho^*, m^{**}),$$

where $m^{**} = \sigma(\frac{1}{2} - \frac{V_s}{B_s})\sqrt{2\log(s/B_s - 1)} + \frac{\sqrt{2}\sigma}{4\sqrt{\pi}}(1 - \frac{B_s}{s - B_s})$ and $\rho^* = \sigma\sqrt{2\log(s/B_s - 1)}$. Following a similar argument, it is shown that

$$B_s[g(s, B_s, \rho^*) - (m^{**} + R)/\rho^*] \geq B_s[\frac{1}{2\sqrt{\pi}} \frac{1}{\sqrt{\log(s/B_s - 1) + 2}} + \frac{V_s}{B_s} - \frac{R}{\rho^*}] \geq V_s.$$

This implies $\liminf_{d,s\to\infty} G(s, B_s, \rho^*, m^{**}) = 1$. This completes the proof.

B.7 *Proof of Corollary 8*

Proof of (3.5) and (3.7). For simplicity, we follow the the same notations in the proof of Corollary 7 and Theorem 6. By the proof of Theorem 6 and $M \in \mathcal{M}_+$, we have $\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin M) \geq 1 - t$ and

$$\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin M) \geq tA_d\Big\{g(d, A_d, \rho) - \frac{m + R}{\rho}\Big\}_+,$$

where $m = \sup_{1\leq j\leq d} \sup_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{E}_{\boldsymbol{\theta}}(\theta_j - L_j)$. If $g(d, A_d, \rho) - \frac{m+R}{\rho} \leq 0$, then

$$m \geq \rho g(d, A_d, \rho) - R. \tag{B.16}$$

Otherwise, we have

$$\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin M) \geq tA_d\Big\{g(d, A_d, \rho) - \frac{m+R}{\rho}\Big\} \geq \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \in M)A_d\Big\{g(d, A_d, \rho) - \frac{m+R}{\rho}\Big\},$$

which implies

$$m \geq \rho\Big\{g(d, A_d, \rho) - \frac{1}{A_d}\frac{\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin M)}{\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \in M)}\Big\} - R. \tag{B.17}$$

Clearly, (B.16) implies (B.17), and the lower bound reduces to (B.17) by combining these two cases.

When $a \leq a_1 := \sigma\sqrt{2\log(d/A_d - 1)}$, we can take $\rho = a_1$. By the proof of Corollary 7, e.g. (B.15),

$$\begin{aligned}
&\liminf_{d,s\to\infty} \inf_{M\in\mathcal{M}_+} \frac{m}{a_1} \\
&\geq \liminf_{d,s\to\infty}\Big[\frac{1}{2} + \frac{1}{2\sqrt{\pi}}\frac{1}{\sqrt{\log(d/A_d - 1) + 2}} - \frac{1}{A_d}\frac{\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin M)}{\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \in M)} - \frac{R}{a_1}\Big] \\
&\geq \liminf_{d,s\to\infty}\Big[\frac{1}{2} + \frac{1}{2\sqrt{\pi}}\frac{1}{\sqrt{\log(d/A_d - 1) + 2}} - \frac{1}{A_d}\frac{\alpha}{1-\alpha} - \frac{1}{4\sqrt{\pi}}\frac{1}{\sqrt{\log(d/A_d - 1)}}\frac{A_d}{d - A_d}\Big] \\
&= \frac{1}{2}.
\end{aligned}$$

We then obtain (3.5).

In case (2) it is easy to verify that for d, s large enough

$$\begin{aligned}
a/\sigma &\geq \sqrt{2\log(d-s) - \log\log(d-s) + C'} + \sqrt{2\log s - \log\log s + C'} \vee \xi_d \\
&\geq \sqrt{2\log(d-s) - \log\log(d-s) + C'} + \sqrt{2\log s - \log\log s + C'} \geq \sqrt{2\log d + C},
\end{aligned}$$

for some constant $C > 0$. This further implies (3.7) by the proof of Corollary 7 and the similar argument in case (1).

Proof of (3.6) and (3.8). We first note that Corollary 5 implies $\bar{M}_{\alpha'} \in \mathcal{M}_+$. For (3.6), it suffices to show that the following inequality holds regardless of the value of a,

$$\limsup_{d,s\to\infty} \frac{\sup_{1\leq j\leq d}\sup_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{E}_{\boldsymbol{\theta}}(\theta_j - L_j)}{\sigma\sqrt{2\log d}} \leq 1,$$

where L_j is defined in (2.14) and for notational simplicity we write L_j for $\bar{L}_j$. First, consider the case that $a^{**} \leq a < \bar{a}$, where $a^{**} = \kappa^{**}\sigma$ and $\bar{a}/\sigma = \sqrt{2\log(\frac{2(d-s)}{(\alpha-\alpha')C_{d-s,\alpha-\alpha'}})} + \sqrt{2\log(\frac{s}{C_{s,\alpha'}\alpha'})}$. Thus, for d large enough

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}(\theta_j - L_j) &= \mathbb{E}_{\boldsymbol{\theta}}(\theta_j - L_j)I(j \in \bar{S}_{\alpha'}) + \mathbb{E}_{\boldsymbol{\theta}}(\theta_j - L_j)I(j \notin \bar{S}_{\alpha'}) \\
&= \mathbb{E}_{\boldsymbol{\theta}}(\theta_j - L_j)I(X_j/\sigma > \Phi^{-1}(\delta)) + \theta_j\mathbb{P}_{\boldsymbol{\theta}}(X_j/\sigma \leq \Phi^{-1}(\delta)) \\
&= \mathbb{E}_{\boldsymbol{\theta}}(\theta_j - X_j + \bar{u}_{\alpha'}\sigma)I(X_j/\sigma > \bar{u}_{\alpha'}) + \theta_j\mathbb{P}_{\boldsymbol{\theta}}(X_j/\sigma \leq \bar{u}_{\alpha'}) \\
&\leq \sigma + \bar{u}_{\alpha'}\sigma\mathbb{P}_{\boldsymbol{\theta}}(X_j/\sigma > \bar{u}_{\alpha'}) + \theta_j\mathbb{P}_{\boldsymbol{\theta}}(X_j/\sigma \leq \bar{u}_{\alpha'}),
\end{aligned} \tag{B.18}$$

where $\bar{u}_{\alpha'}$ is defined in (2.15). If $\bar{u}_{\alpha'}\sigma \geq \theta_j$, the above display implies

$$\mathbb{E}_{\boldsymbol{\theta}}(\theta_j - L_j) \leq \sigma(1 + \bar{u}_{\alpha'}).$$

If $\bar{u}_{\alpha'}\sigma < \theta_j$, the above display and the Gaussian tail bound in Lemma 16 leads to

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\theta}}(\theta_j - L_j) &\leq \sigma(1 + \bar{u}_{\alpha'}) + (\theta_j - \bar{u}_{\alpha'}\sigma)\mathbb{P}(Z > \theta_j/\sigma - \bar{u}_{\alpha'}) \\ &\leq \sigma(1 + \bar{u}_{\alpha'}) + \frac{\sigma}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2}(\theta_j/\sigma - \bar{u}_{\alpha'})^2\Big) \\ &\leq \sigma(1 + \bar{u}_{\alpha'}) + \frac{\sigma}{\sqrt{2\pi}},\end{aligned}$$

where $Z \sim N(0, 1)$. Combining these two cases, we have

$$\limsup_{d,s\to\infty} \frac{\sup_{1\leq j\leq d}\sup_{\boldsymbol{\theta}\in\Theta^+(s,a)}\mathbb{E}_{\boldsymbol{\theta}}(\theta_j - L_j)}{\sigma\sqrt{2\log d}} \leq \limsup_{d,s\to\infty} \frac{1 + \bar{u}_{\alpha'} + 1/\sqrt{2\pi}}{\sqrt{2\log d}} = 1. \tag{B.19}$$

Now we consider the case $a \geq \bar{a}$. If $\bar{u}_{\alpha'} \leq w$, we have

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\theta}}(\theta_j - L_j) &= \mathbb{E}_{\boldsymbol{\theta}}(\theta_j - L_j)I(j \in \bar{S}_{\alpha'}) + \mathbb{E}_{\boldsymbol{\theta}}(\theta_j - L_j)I(j \notin \bar{S}_{\alpha'}) \\ &= \mathbb{E}_{\boldsymbol{\theta}}(\theta_j - X_j + \bar{u}_{\alpha'}\sigma)I(X_j/\sigma > w) + \theta_j\mathbb{P}_{\boldsymbol{\theta}}(X_j/\sigma \leq w) \\ &\leq \sigma + \bar{u}_{\alpha'}\sigma\mathbb{P}_{\boldsymbol{\theta}}(X_j/\sigma > w) + \theta_j\mathbb{P}_{\boldsymbol{\theta}}(X_j/\sigma \leq w),\end{aligned} \tag{B.20}$$

where $w = \sqrt{2\log(\frac{2(d-s)}{(\alpha-\alpha')C_{d-s,\alpha-\alpha'}})}$ and $\bar{u}_{\alpha'}$ is defined in (2.16). Note that it suffices to only consider the non-zero θ_j, since if $\theta_j = 0$, (B.20) can be trivially bounded by $\sigma(1 + \bar{u}_{\alpha'})$. Then, for $\theta_j \neq 0$ we have

$$\frac{\theta_j}{\sigma} \geq \frac{a}{\sigma} \geq \bar{\kappa} > w \geq \bar{u}_{\alpha'}.$$

By the same derivation, it can be shown that

$$\mathbb{E}_{\boldsymbol{\theta}}(\theta_j - L_j) \leq \sigma(1 + \bar{u}_{\alpha'}) + \frac{\sigma}{\sqrt{2\pi}}\frac{\theta_j/\sigma - \bar{u}_{\alpha'}}{\theta_j/\sigma - w} \exp\Big(-\frac{1}{2}(\theta_j/\sigma - w)^2\Big). \tag{B.21}$$

Note that

$$\lim_{d,s\to\infty} \frac{\sigma(1 + \bar{u}_{\alpha'})}{\sigma\sqrt{2\log d}} \leq \lim_{d,s\to\infty} \frac{1 + \sqrt{2\log s - \log\log s + C}}{\sqrt{2\log d}} \leq 1,$$

for some constant C (depending on α', α), and

$$\frac{\theta_j/\sigma - \bar{u}_{\alpha'}}{\theta_j/\sigma - w} = 1 + \frac{w - \bar{u}_{\alpha'}}{\theta_j/\sigma - w} \leq 1 + \frac{w - \bar{u}_{\alpha'}}{\sqrt{2\log(\frac{s}{C_{s,\alpha'}\alpha'})}} \leq 1 + \frac{w}{\sqrt{2\log(\frac{s}{C_{s,\alpha'}\alpha'})}}.$$

Plugging into (B.21) and notice that $w/\sqrt{2\log d} \le 1$ for d sufficiently large, we obtain

$$\frac{\sup_{1\le j\le d}\sup_{\boldsymbol\theta\in\Theta^+(s,a)}\mathbb{E}_{\boldsymbol\theta}(\theta_j - L_j)}{\sigma\sqrt{2\log d}} \le 1 + O(\frac{1}{\sqrt{\log s}}) \to 1.$$

However, if $\bar u_{\alpha'} > w$, (B.20) has a slightly different form as follows:

$$\begin{aligned}\mathbb{E}_{\boldsymbol\theta}(\theta_j - L_j) &= \mathbb{E}_{\boldsymbol\theta}(\theta_j - L_j)I(j\in\bar S_{\alpha'}) + \mathbb{E}_{\boldsymbol\theta}(\theta_j - L_j)I(j\notin\bar S_{\alpha'})\\ &= \mathbb{E}_{\boldsymbol\theta}(\theta_j - X_j + \bar u_{\alpha'}\sigma)I(X_j/\sigma > \bar u_{\alpha'}) + \theta_j\mathbb{P}_{\boldsymbol\theta}(X_j/\sigma \le \bar u_{\alpha'})\\ &\le \sigma + \bar u_{\alpha'}\sigma\mathbb{P}_{\boldsymbol\theta}(X_j/\sigma > \bar u_{\alpha'}) + \theta_j\mathbb{P}_{\boldsymbol\theta}(X_j/\sigma \le \bar u_{\alpha'}),\end{aligned} \tag{B.22}$$

which is identical to (B.18) expect that $\bar u_{\alpha'}$ is defined in (2.16) rather than (2.15). However, this does not change the proof of (B.19) (i.e. (B.19) still holds). This completes the proof of (3.6).

Finally, we focus on the last result (3.8). As $d, s \to \infty$,

$$\begin{aligned}a/\sigma &\ge \sqrt{2\log(d-s) - \log\log(d-s) + C'} + \sqrt{2\log s - \log\log s + C'} \vee \xi_d\\ &\ge \sqrt{2\log(\frac{2(d-s)}{(\alpha-\alpha')C_{d-s,\alpha-\alpha'}})} + \sqrt{2\log(\frac{s}{C_{s,\alpha'}\alpha'})} = \bar\kappa.\end{aligned}$$

Thus, by the definition of $\bar M_{\alpha'}$, when $a/\sigma \ge \bar\kappa$, $\bar u_{\alpha'}$ is defined in (2.16). By (B.21), we first note that

$$\lim_{d,s\to\infty}\frac{\sigma(1+\bar u_{\alpha'})}{\sigma\sqrt{2\log s}} \le \lim_{d,s\to\infty}\frac{1+\sqrt{2\log s - \log\log s + C}}{\sqrt{2\log s}} = 1.$$

For d, s large enough,

$$\theta_j/\sigma - w \ge a/\sigma - w \ge \sqrt{2\log s - \log\log s + C'} \vee \xi_d.$$

Recall that $\xi_d = \sqrt{(\log\log(d-s) - \log\log s)_+}$. Thus, uniformly over $\boldsymbol\theta$ we have

$$\begin{aligned}&\frac{\theta_j/\sigma - \bar u_{\alpha'}}{\theta_j/\sigma - w}\exp\Big(-\frac{1}{2}(\theta_j/\sigma - w)^2\Big)\\ &= \Big(1 + \frac{w - \bar u_{\alpha'}}{\theta_j/\sigma - w}\Big)\exp\Big(-\frac{1}{2}(\theta_j/\sigma - w)^2\Big)\\ &\le \Big(1 + \frac{w}{\sqrt{2\log s - \log\log s + C'}}\Big)\exp\Big(-\frac{1}{2}(\xi_d^2 \vee (2\log s - \log\log s + C'))\Big).\end{aligned}$$

If $d - s > s$ holds, we have that as $d - s \to \infty$ and $s \to \infty$,

$$\begin{aligned}&\frac{\theta_j/\sigma - \bar u_{\alpha'}}{\theta_j/\sigma - w}\exp\Big(-\frac{1}{2}(\theta_j/\sigma - w)^2\Big)\\ &\le \Big(1 + \frac{\sqrt{2\log(d-s) - \log\log(d-s) + C}}{\sqrt{2\log s - \log\log s + C'}}\Big)\sqrt{\frac{\log s}{\log(d-s)}} \le 2.\end{aligned}$$

Otherwise, under $d-s\le s$,

$$\frac{\theta_j/\sigma-\bar u_{\alpha'}}{\theta_j/\sigma-w}\exp\Big(-\frac{1}{2}(\theta_j/\sigma-w)^2\Big)$$
$$\le\Big(1+\frac{\sqrt{2\log(d-s)-\log\log(d-s)+C'}}{\sqrt{2\log s-\log\log s+C}}\Big)\frac{\exp(-C'/2)\sqrt{\log s}}{s}\to 0.$$

Thus, by (B.21) we have

$$\frac{\sup_{1\le j\le d}\sup_{\boldsymbol\theta\in\Theta^+(s,a)}\mathbb{E}_{\boldsymbol\theta}(\theta_j-L_j)}{\sigma\sqrt{2\log s}}\le 1+O(\frac{1}{\sqrt{\log s}})\to 1.$$

Similarly, if $\bar u_{\alpha'}>w$ holds, then we have (B.22). By the proof of (3.6) with $\bar u_{\alpha'}$ defined in (2.16), we still arrive at

$$\frac{\sup_{1\le j\le d}\sup_{\boldsymbol\theta\in\Theta^+(s,a)}\mathbb{E}_{\boldsymbol\theta}(\theta_j-L_j)}{\sigma\sqrt{2\log s}}\le 1.$$

This completes the proof of (3.8).

B.8 *Proof of Theorem 9*

Proof. Proof of Theorem 9Our proof relies on the following two lemmas.

LEMMA 14. Under the same condition in Theorem 9, we have

$$\sup_{\boldsymbol\theta\in\Theta^+(s,a)}\mathbb{P}_{\boldsymbol\theta}(|\bar S^{ad}_{\alpha'}|\ge 2s)\le\Big(\frac{C_2(d-s)}{sd}\Big)^s,\quad \sup_{\boldsymbol\theta\in\Theta^+(s,a)}\mathbb{P}_{\boldsymbol\theta}(|\bar S^{ad}_{\alpha'}|\le s/2)\le\Big(\frac{C_4}{s}\Big)^{s/2},$$

where C_2 and C_4 are two universal constants.

LEMMA 15. Under the same condition in Theorem 9, there exists a positive constant C such that for (t,s,d) sufficiently large,

$$\sup_{1\le j\le d}\sup_{\boldsymbol\theta\in\Theta^+(s,a)}\mathbb{E}_{\boldsymbol\theta}(\theta_j-\widehat L_{j,t})^2\le C\sigma^2\log t,$$

where s,a satisfy the scenario (B).

First, the proof of Corollary 5 implies $\sup_{\theta_j=0}\mathbb{P}_{\boldsymbol\theta}(j\in\bar S^{ad}_{\alpha'})\le 1-\delta$. Note that $2s\le d$ implies $d-s\ge d/2$. By the monotonicity of $\log(x/\sqrt{\log x})$, we can show that for any s,a satisfying the scenario (B) and $\theta_j\ne 0$, we have

$$\frac{\theta_j}{\sigma}\ge\frac{a}{\sigma}\ge\sqrt{2\log(d-s)-\log\log(d-s)+C'}+\sqrt{2\log s-\log\log s+C'}$$
$$\ge\sqrt{2\log(d/2)-\log\log(d/2)+C'}+\sqrt{2\log s-\log\log s+C'}$$
$$\ge\sqrt{2\log(2d)-\log\log d+C'/2}+\sqrt{2\log s-\log\log s+C'},$$

where C' is a sufficiently large constant. Thus, we can show that

$$\mathbb{P}_{\boldsymbol{\theta}}(\text{supp}(\boldsymbol{\theta}) \not\subseteq \bar{S}^{ad}_{\alpha'}) \le \sum_{j:\theta_j \neq 0} \mathbb{P}_{\theta_j}\Big(X_j/\sigma \le \sqrt{2\log(\frac{2d}{(\alpha-\alpha')C_{d,\alpha-\alpha'}})}\Big)$$
$$= \sum_{j:\theta_j \neq 0} \mathbb{P}_{\theta_j}\Big(\frac{X_j-\theta_j}{\sigma} \le \sqrt{2\log(\frac{2d}{(\alpha-\alpha')C_{d,\alpha-\alpha'}})} - \frac{\theta_j}{\sigma}\Big)$$
$$\le \sum_{j:\theta_j \neq 0} \mathbb{P}_{\theta_j}\Big(\frac{X_j-\theta_j}{\sigma} \le -\sqrt{2\log s - \log\log s + C'}\Big) \le \alpha', \qquad \text{(B.23)}$$

for s sufficiently large, where the last step holds as C' is a sufficiently large constant. Note that under the event $s/2 \le |\bar{S}^{ad}_{\alpha'}| \le 2s$, by the definition of $\widehat{s}$, we have $s/2 \le |\bar{S}^{ad}_{\alpha'}| < \widehat{s} \le 2|\bar{S}^{ad}_{\alpha'}| \le 4s$. That is, $s/2 \le \widehat{s} \le 4s$. Formally, the above argument and Lemma 14 imply

$$\sup_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin \widehat{M}^{ad}_{\alpha'}) \le \sup_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin \widehat{M}^{ad}_{\alpha'}, s/2 \le \widehat{s} \le 4s) + \Big(\frac{C_2(d-s)}{sd}\Big)^s + \Big(\frac{C_4}{s}\Big)^{s/2}. \qquad \text{(B.24)}$$

By the proof of Theorem 4, we have

$$\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin \widehat{M}^{ad}_{\alpha'}, s/2 \le \widehat{s} \le 4s)$$
$$\le s(1-\Phi(u_{\alpha',s/2})) + \sum_{j\notin \text{supp}(\boldsymbol{\theta})} \mathbb{P}_0\big(X_j/\sigma \ge \eta\big) + \mathbb{P}_{\boldsymbol{\theta}}(\text{supp}(\boldsymbol{\theta}) \not\subseteq \bar{S}^{ad}_{\alpha'})$$
$$= s\Big(1-\Phi\Big(\sqrt{2\log(\frac{2s}{(\alpha-\alpha')C_{s,\alpha-\alpha'}})}\Big)\Big) + (d-s)(1-\Phi(\eta)) + \mathbb{P}_{\boldsymbol{\theta}}(\text{supp}(\boldsymbol{\theta}) \not\subseteq \bar{S}^{ad}_{\alpha'}), \qquad \text{(B.25)}$$

where

$$\eta = \sqrt{2\log(\frac{2d}{(\alpha-\alpha')C_{d,\alpha-\alpha'}})}.$$

The first two terms in (B.25) are both upper bounded by $(\alpha-\alpha')/2$ for s, d sufficiently large, see the proof of Corollary 5. The upper bound for the last term in (B.25) is shown in (B.23). Thus, from (B.24) we obtain $\liminf_{d,s\to\infty} \inf_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \in \widehat{M}^{ad}_{\alpha'}) \ge 1-\alpha$.

Next, we are ready to show

$$\limsup_{d,s\to\infty} \frac{R(\widehat{M}^{ad}_{\alpha'}, \Theta^+(s,a))}{\sigma\sqrt{2\log s}} \le 1.$$

Note that

$$\mathbb{E}_{\boldsymbol{\theta}}(\theta_j - \widehat{L}_{j,\widehat{s}}) = \mathbb{E}_{\boldsymbol{\theta}}(\theta_j - \widehat{L}_{j,\widehat{s}})I(\widehat{s} \le 4s) + \mathbb{E}_{\boldsymbol{\theta}}(\theta_j - \widehat{L}_{j,\widehat{s}})I(\widehat{s} > 4s) := I_1 + I_2. \qquad \text{(B.26)}$$

We further decompose I_1 as follows:

$$I_1 = \mathbb{E}_{\boldsymbol{\theta}}(\theta_j - \widehat{L}_{j,\widehat{s}})I(\widehat{s} \le 4s)I(X_j/\sigma \ge \eta) + \mathbb{E}_{\boldsymbol{\theta}}(\theta_j - \widehat{L}_{j,\widehat{s}})I(\widehat{s} \le 4s)I(X_j/\sigma < \eta) := I_{11} + I_{12}.$$

Under the event $\widehat{s} \leq 4s$, it holds that

$$u_{\alpha',\widehat{s}} = \sqrt{2\log\Big(\frac{4\widehat{s}}{(\alpha-\alpha')C_{2\widehat{s},\alpha-\alpha'}}\Big)} \leq \sqrt{2\log\Big(\frac{16s}{(\alpha-\alpha')C_{8s,\alpha-\alpha'}}\Big)} := u_{\alpha',4s}.$$

Thus, for I_{11}, we have

$$\begin{aligned}
I_{11} &= \mathbb{E}_{\boldsymbol{\theta}}[\theta_j - (X_j - u_{\alpha',\widehat{s}}\sigma)_+]I(\widehat{s} \leq 4s)I(X_j/\sigma \geq \eta) \\
&\leq \mathbb{E}_{\boldsymbol{\theta}}[\theta_j - (X_j - u_{\alpha',4s}\sigma)_+]I(\widehat{s} \leq 4s)I(X_j/\sigma \geq \eta) \\
&= \mathbb{E}_{\boldsymbol{\theta}}[\theta_j - (X_j - u_{\alpha',4s}\sigma)_+]I(\widehat{s} \leq 4s)I(\theta_j > (X_j - u_{\alpha',4s}\sigma)_+)I(X_j/\sigma \geq \eta) \\
&\quad + \mathbb{E}_{\boldsymbol{\theta}}[\theta_j - (X_j - u_{\alpha',4s}\sigma)_+]I(\widehat{s} \leq 4s)I(\theta_j \leq (X_j - u_{\alpha',4s}\sigma)_+)I(X_j/\sigma \geq \eta) \\
&\leq \mathbb{E}_{\boldsymbol{\theta}}[\theta_j - (X_j - u_{\alpha',4s}\sigma)_+]I(\widehat{s} \leq 4s)I(\theta_j > (X_j - u_{\alpha',4s}\sigma)_+)I(X_j/\sigma \geq \eta) \\
&\leq \mathbb{E}_{\boldsymbol{\theta}}[\theta_j - X_j + u_{\alpha',4s}\sigma]I(\theta_j > (X_j - u_{\alpha',4s}\sigma)_+)I(X_j > u_{\alpha',4s}\sigma)I(X_j/\sigma \geq \eta) \\
&\quad + \theta_j\mathbb{E}_{\boldsymbol{\theta}}I(\theta_j > (X_j - u_{\alpha',4s}\sigma)_+)I(X_j \leq u_{\alpha',4s}\sigma)I(X_j/\sigma \geq \eta).
\end{aligned}$$

Following the proof of Corollary 8, we can further show that

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}}[\theta_j - X_j + u_{\alpha',4s}\sigma]I(\theta_j > (X_j - u_{\alpha',4s}\sigma)_+)I(X_j > u_{\alpha',4s}\sigma)I(X_j/\sigma \geq \eta) \\
&\leq \mathbb{E}_{\boldsymbol{\theta}}[\theta_j - X_j]I(\theta_j > (X_j - u_{\alpha',4s}\sigma)_+)I(X_j > u_{\alpha',4s}\sigma)I(X_j/\sigma \geq \eta) + u_{\alpha',4s}\sigma\mathbb{P}_{\boldsymbol{\theta}}(X_j > u_{\alpha',4s}\sigma) \\
&\leq \sigma + u_{\alpha',4s}\sigma.
\end{aligned}$$

After some simple calculation similar to the proof of Corollary 8, we can show that by the tail bound in Lemma 16,

$$\theta_j\mathbb{E}_{\boldsymbol{\theta}}I(\theta_j > (X_j - u_{\alpha',4s}\sigma)_+)I(X_j \leq u_{\alpha',4s}\sigma)I(X_j/\sigma \geq \eta) \leq \theta_j\mathbb{P}_{\boldsymbol{\theta}}(X_j \leq u_{\alpha',4s}\sigma) \leq C\sigma,$$

for any $\theta_j \neq 0$, where C is a positive constant. Combining the above inequalities, we obtain

$$I_{11} \leq (1+C)\sigma + u_{\alpha',4s}\sigma. \tag{B.27}$$

For I_{12}, recall that it suffices to only consider non-zero θ_j. The tail bound in Lemma 16 leads to

$$\begin{aligned}
I_{12} &\leq \theta_j\mathbb{P}_{\boldsymbol{\theta}}(X_j/\sigma < \eta) = \theta_j\mathbb{P}(Z/\sigma < -(\theta_j/\sigma - \eta)) \\
&\leq C\sigma(1 + \frac{\eta}{\theta_j/\sigma - \eta})\exp(-\frac{1}{2}\bar{\xi}_d^2) \\
&\leq C\sigma\Big(1 + \frac{\sqrt{2\log d - \log\log d + C''}}{\sqrt{2\log s - \log\log s + C'}}\Big)\frac{\sqrt{\log s}}{\log(d-s)} \leq 2C\sigma.
\end{aligned}$$

In the following text, we consider I_2. Let $t_0 = \{t \in [T] : 2^{t-1} \leq 4s < 2^t\}$. Then

$$I_2 = \sum_{t=t_0}^{T} \mathbb{E}_{\boldsymbol{\theta}}(\theta_j - \widehat{L}_{j,2^t}) I(\widehat{s} = 2^t) \leq \sum_{t=t_0}^{T} \{\mathbb{E}_{\boldsymbol{\theta}}(\theta_j - \widehat{L}_{j,2^t})^2\}^{1/2} \{\mathbb{P}_{\boldsymbol{\theta}}(\widehat{s} = 2^t)\}^{1/2}.$$

For any $t, \widehat{s} = 2^t$ implies $2^{t-1} \leq |\bar{S}^{ad}_{\alpha'}| < 2^t$. Thus,

$$\mathbb{P}_{\boldsymbol{\theta}}(\widehat{s} = 2^t) \leq \mathbb{P}_{\boldsymbol{\theta}}(2^{t-1} \leq |\bar{S}^{ad}_{\alpha'}| < 2^t) \leq \mathbb{P}_{\boldsymbol{\theta}}(|\bar{S}^{ad}_{\alpha'}| \geq 2^{t-1}).$$

Recall that $2^{t_0-1} > 2s$. Following the proof of Lemma 14, we have for any $t \geq t_0$

$$\mathbb{P}_{\boldsymbol{\theta}}(\widehat{s} = 2^t) \leq \binom{d-s}{2^{t-1}-s} \left(\frac{C}{d}\right)^{2^{t-1}-s} \leq \left(\frac{C'}{2^{t-1}-s}\right)^{2^{t-1}-s},$$

where C, C' are positive constants. In addition, Lemma 15 implies

$$\mathbb{E}_{\boldsymbol{\theta}}(\theta_j - \widehat{L}_{j,2^t})^2 \leq C(\log 2)\sigma^2 t.$$

Thus, for I_2, there exists a constant $C > 0$ such that

$$\begin{aligned} I_2 &\leq \sigma \sum_{t=t_0}^{T} t^{1/2} \left(\frac{C}{2^{t-1}-s}\right)^{\frac{2^{t-1}-s}{2}} \leq \sigma \sum_{t=t_0}^{T} \frac{Ct^{1/2}}{2^{t-1} - 2^{t_0-2}} = \sigma \sum_{q=0}^{T-t_0} \frac{C(t_0+q)^{1/2}}{2^{q+t_0-1} - 2^{t_0-2}} \\ &\leq \frac{C\sigma t_0^{1/2}}{2^{t_0-2}} \sum_{q=0}^{T-t_0} \frac{1}{2^{q+1}-1} + \frac{C\sigma}{2^{t_0-2}} \sum_{q=0}^{T-t_0} \frac{q^{1/2}}{2^{q+1}-1}. \end{aligned}$$

It is easily seen that the infinite sum $\sum_{q=0}^{\infty} \frac{1}{2^{q+1}-1}$ and $\sum_{q=0}^{\infty} \frac{q^{1/2}}{2^{q+1}-1}$ converges. Thus, there exists a constant $C' > 0$ such that

$$I_2 \leq \frac{C'\sigma t_0^{1/2}}{2^{t_0-2}} + \frac{C'\sigma}{2^{t_0-2}}.$$

Combining with (B.26) and (B.27), we obtain

$$\mathbb{E}_{\boldsymbol{\theta}}(\theta_j - \widehat{L}_{j,\widehat{s}}) \leq (1+3C)\sigma + u_{\alpha',4s}\sigma + \frac{C'\sigma t_0^{1/2}}{2^{t_0-2}} + \frac{C'\sigma}{2^{t_0-2}}.$$

Noting that

$$\lim_{d,s\to\infty} \frac{u_{\alpha',4s}}{\sqrt{2\log s}} = \lim_{d,s\to\infty} \frac{\sqrt{2\log s - \log\log s + C}}{\sqrt{2\log s}} = 1,$$

we complete the proof. □

Proof. Proof of Lemma 14.Consider the event $|\bar{S}^{ad}_{\alpha'}| \geq 2s$. It implies that there exist at least s number of θ_j such that $\theta_j = 0$ and $X_j/\sigma \geq \sqrt{2\log(\frac{2d}{(\alpha-\alpha')C_{d,\alpha-\alpha'}})}$. Thus, we have uniformly over $\boldsymbol{\theta} \in \Theta^+(s,a)$,

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\theta}}(|\bar{S}^{ad}_{\alpha'}| \geq 2s) &\leq \binom{d-s}{s}\Big[\mathbb{P}_{\theta_j=0}\Big(X_j/\sigma \geq \sqrt{2\log(\frac{2d}{(\alpha-\alpha')C_{d,\alpha-\alpha'}})}\Big)\Big]^s \\
&\leq \Big(\frac{(d-s)e}{s}\Big)^s\Big[\sqrt{\frac{2}{\pi}}\frac{1}{2\sqrt{2\log(\frac{2d}{C_{d,\alpha-\alpha'}(\alpha-\alpha')})}}\exp\Big(-\log(\frac{2d}{C_{d,\alpha-\alpha'}(\alpha-\alpha')})\Big)\Big]^s \\
&\leq \Big(\frac{(d-s)e}{s}\Big)^s\Big[\frac{C_1 d^{-1}\sqrt{\log d}}{\sqrt{\log(\frac{2d}{C_{d,\alpha-\alpha'}})}}\Big]^s \\
&\leq \Big(\frac{C_2(d-s)}{sd}\Big)^s,
\end{aligned}$$

for d large enough, where $C_1, C_2 > 0$ are two universal constants. Consider the event $|\bar{S}^{ad}_{\alpha'}| \leq s/2$. It implies that there exist at least $s/2$ number of θ_j such that $\theta_j > 0$ and $j \notin \bar{S}^{ad}_{\alpha'}$. Following the similar argument and the inequality (B.10), we can show that

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\theta}}(|\bar{S}^{ad}_{\alpha'}| \leq s/2) &\leq \binom{s}{s/2}\Big[\mathbb{P}_{\theta_j}\Big(X_j/\sigma \leq \sqrt{2\log(\frac{2d}{(\alpha-\alpha')C_{d,\alpha-\alpha'}})}\Big)\Big]^{s/2} \\
&\leq \binom{s}{s/2}\Big[\mathbb{P}\Big(N \leq -\sqrt{2\log s - \log\log s + C'}\Big)\Big]^{s/2} \\
&\leq (2e)^{s/2}\Big[\frac{C_3 s^{-1}\sqrt{\log s}}{\sqrt{2\log s - \log\log s + C'}}\Big]^{s/2} \\
&\leq \Big(\frac{C_4}{s}\Big)^{s/2},
\end{aligned}$$

for d, s large enough, where $N \sim N(0,1)$ and $C_3, C_4 > 0$ are two universal constants. □

Proof. Proof of Lemma 15.Following the proof of Corollary 8, we consider two cases $d \geq 2t$ and $d < 2t$. When the former condition holds, we can show that

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}(\theta_j - \widehat{L}_{j,t})^2 &= \mathbb{E}_{\boldsymbol{\theta}}(\theta_j - X_j + u_{\alpha',t}\sigma)^2 I(X_j/\sigma > w) + \theta_j^2\mathbb{P}_{\boldsymbol{\theta}}(X_j/\sigma \leq w) \\
&= \mathbb{E}_{\boldsymbol{\theta}}(\theta_j - X_j)^2 I(X_j/\sigma > w) + u^2_{\alpha',t}\sigma^2\mathbb{P}_{\boldsymbol{\theta}}(X_j/\sigma > w) \\
&\quad + 2u_{\alpha',t}\sigma\mathbb{E}_{\boldsymbol{\theta}}(\theta_j - X_j)I(X_j/\sigma > w) + \theta_j^2\mathbb{P}_{\boldsymbol{\theta}}(X_j/\sigma \leq w) \\
&\leq \sigma^2 + u^2_{\alpha',t}\sigma^2 + 2u_{\alpha',t}\sigma^2 + \theta_j^2\mathbb{P}_{\boldsymbol{\theta}}(X_j/\sigma \leq w), \qquad \text{(B.28)}
\end{aligned}$$

where $w = \sqrt{2\log(\frac{2d}{(\alpha-\alpha')C_{d,\alpha-\alpha'}})}$. Furthermore, the tail bound in Lemma 16 implies

$$\begin{aligned}\theta_j^2\mathbb{P}_{\boldsymbol{\theta}}(X_j/\sigma \le w) &\le \frac{\sigma^2}{\sqrt{2\pi}}\frac{\theta_j^2/\sigma^2}{\theta_j/\sigma - w}\exp\Big(-\frac{1}{2}(\theta_j/\sigma - w)^2\Big)\\ &\le \frac{\sigma^2}{\sqrt{2\pi}}\frac{2(\theta_j/\sigma - w)^2 + 2w^2}{\theta_j/\sigma - w}\exp\Big(-\frac{1}{2}(\theta_j/\sigma - w)^2\Big).\end{aligned}$$

Recall that if d, s large enough, we have

$$\theta_j/\sigma - w \ge a/\sigma - w \ge \sqrt{2\log s - \log\log s + C'} \vee \bar{\xi}_d,$$

and $w^2 \le 2\log d - \log\log d + C'$ for some constant C'. Thus, for d, s large enough,

$$\begin{aligned}\theta_j^2\mathbb{P}_{\boldsymbol{\theta}}(X_j/\sigma \le w) &\le \frac{2\sigma^2}{\sqrt{2\pi}}\sqrt{2\log s - \log\log s + C'}\exp\Big(-\frac{1}{2}(2\log s - \log\log s + C')^2\Big)\\ &\quad + \frac{2\sigma^2}{\sqrt{2\pi}}\frac{2\log d - \log\log d + C'}{\sqrt{2\log s - \log\log s + C'}}\exp\Big(-\frac{1}{2}\bar{\xi}_d^2\Big)\\ &\le \sigma^2 + \frac{2\sigma^2}{\sqrt{2\pi}}\frac{2\log d - \log\log d + C'}{\sqrt{2\log s - \log\log s + C'}}\frac{\sqrt{\log s}}{\log(d-s)}\\ &\le \sigma^2 + \frac{2\sigma^2}{\sqrt{2\pi}}\frac{2\log d - \log\log d + C'}{\sqrt{2\log s - \log\log s + C'}}\frac{\sqrt{\log s}}{\log(d/2)} \le C''\sigma^2,\end{aligned}$$

where C'' is a positive constant. Finally, we plug into the inequality (B.28),

$$\mathbb{E}_{\boldsymbol{\theta}}(\theta_j - \widehat{L}_{j,t})^2 \le \sigma^2 + u_{\alpha',t}^2\sigma^2 + 2u_{\alpha',t}\sigma^2 + C''\sigma^2.$$

For (t, s, d) sufficiently large, it can be easily verified that the following inequality holds:

$$\sup_{1\le j\le d}\sup_{\boldsymbol{\theta}\in\Theta^+(s,a)}\mathbb{E}_{\boldsymbol{\theta}}(\theta_j - \widehat{L}_{j,t})^2 \le C\sigma^2\log t. \tag{B.29}$$

In the following text, we consider the second case $d < 2t$. Similar to (B.28), we obtain that

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\theta}}(\theta_j - \widehat{L}_{j,t})^2 &= \mathbb{E}_{\boldsymbol{\theta}}(\theta_j - X_j + u_{\alpha',t}\sigma)^2 I(X_j/\sigma > u_{\alpha',t}) + \theta_j^2\mathbb{P}_{\boldsymbol{\theta}}(X_j/\sigma \le u_{\alpha',t})\\ &= \mathbb{E}_{\boldsymbol{\theta}}(\theta_j - X_j)^2 I(X_j/\sigma > u_{\alpha',t}) + u_{\alpha',t}^2\sigma^2\mathbb{P}_{\boldsymbol{\theta}}(X_j/\sigma > u_{\alpha',t})\\ &\quad + 2u_{\alpha',t}\sigma\mathbb{E}_{\boldsymbol{\theta}}(\theta_j - X_j)I(X_j/\sigma > u_{\alpha',t}) + \theta_j^2\mathbb{P}_{\boldsymbol{\theta}}(X_j/\sigma \le u_{\alpha',t})\\ &\le \sigma^2 + u_{\alpha',t}^2\sigma^2\mathbb{P}_{\boldsymbol{\theta}}(X_j/\sigma > u_{\alpha',t}) + 2u_{\alpha',t}\sigma^2 + \theta_j^2\mathbb{P}_{\boldsymbol{\theta}}(X_j/\sigma \le u_{\alpha',t}).\end{aligned}$$

If $2u_{\alpha',t}\sigma \ge \theta_j$, then $\mathbb{E}_{\boldsymbol{\theta}}(\theta_j - \widehat{L}_{j,t})^2 \le \sigma^2 + 5u^2_{\alpha',t}\sigma^2 + 2u_{\alpha',t}\sigma^2$. If $2u_{\alpha',t}\sigma < \theta_j$, then $\theta_j/\sigma - u_{\alpha',t} > \theta_j/(2\sigma)$, and we further have

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\theta}}(\theta_j - \widehat{L}_{j,t})^2 &\le (1+u_{\alpha',t})^2\sigma^2 + (\theta_j^2 - u^2_{\alpha',t}\sigma^2)\mathbb{P}(Z > \theta_j/\sigma - u_{\alpha',t})\\ &\le (1+u_{\alpha',t})^2\sigma^2 + \frac{\sigma^2(\theta_j/\sigma)}{\sqrt{2\pi}}\exp\Big(-\frac{1}{2}(\theta_j/(2\sigma))^2\Big)\\ &\le (1+u_{\alpha',t})^2\sigma^2 + \frac{\sigma^2}{\sqrt{2\pi}},\end{aligned}$$

as s, d tend to infinity, where $Z \sim N(0,1)$. For (t,s,d) sufficiently large, (B.29) holds as well for the second case $d < 2t$. This completes the proof. □

B.9 *Proof of Theorem 10*

Denote $A^+ = \{\boldsymbol{\theta} \in \mathbb{R}^d : \|\boldsymbol{\theta}\|_0 = s, \theta_j = a \text{ for } \forall j, \theta_j \ne 0\}$ and $A^- = \{\boldsymbol{\theta} \in \mathbb{R}^d : \|\boldsymbol{\theta}\|_0 = s, \theta_j = -a \text{ for } \forall j, \theta_j \ne 0\}$. Then

$$\sup_{\boldsymbol{\theta}\in\Theta(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\mathrm{supp}(\boldsymbol{\theta}) \not\subseteq \widehat{S}) \ge \sup_{\boldsymbol{\theta}\in A^+\cup A^-} \mathbb{P}_{\boldsymbol{\theta}}(\mathrm{supp}(\boldsymbol{\theta}) \not\subseteq \widehat{S}) = 1 - \inf_{\boldsymbol{\theta}\in A^+\cup A^-} \mathbb{P}_{\boldsymbol{\theta}}(\mathrm{supp}(\boldsymbol{\theta}) \subseteq \widehat{S}).$$

Following the proof of Theorem 1, we obtain that

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\Theta(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(S \not\subseteq \widehat{S}) &\ge \frac{1}{2}\Big[\sup_{\boldsymbol{\theta}\in A^+}\mathbb{P}_{\boldsymbol{\theta}}(S \not\subseteq \widehat{S}) + \sup_{\boldsymbol{\theta}\in A^-}\mathbb{P}_{\boldsymbol{\theta}}(S \not\subseteq \widehat{S})\Big]\\ &\ge \frac{1}{2|A^+|}\sum_{\boldsymbol{\theta}\in A^+}\mathbb{P}_{\boldsymbol{\theta}}(S \not\subseteq \widehat{S}) + \frac{1}{2|A^-|}\sum_{\boldsymbol{\theta}\in A^-}\mathbb{P}_{\boldsymbol{\theta}}(S \not\subseteq \widehat{S})\\ &\ge t\sum_{k=1}^{s}\binom{s}{k}\frac{u_+^k + u_-^k}{2},\end{aligned}$$

where $t = \inf_{\boldsymbol{\theta}\in A^+\cup A^-}\prod_{j\in S}\mathbb{P}_{\theta_j}(j \in \widehat{S})$, and $u_+ = \mathbb{P}_a(j \notin \widehat{S})$ and $u_- = \mathbb{P}_{-a}(j \notin \widehat{S})$, where $\mathbb{P}_a$ denotes the probability of $X_j \sim N(a,\sigma^2)$. Applying Jensen's inequality, the above display can be further bounded from below which yields

$$\sup_{\boldsymbol{\theta}\in\Theta(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(S \not\subseteq \widehat{S}) \ge t\sum_{k=1}^{s}\binom{s}{k}\Big(\frac{u_+ + u_-}{2}\Big)^k = t\Big[(1 + \frac{u_+ + u_-}{2})^s - 1\Big].$$

We denote $j \in \widehat{S}$ by $T(X_j) = 1$ for some function $T(\cdot)$. The Neyman–Pearson lemma implies that the infimum of $(u_+ + u_-)/2 = \mathbb{P}_a(T(X_j) = 0)/2 + \mathbb{P}_{-a}(T(X_j) = 0)/2$ over all possible $T(\cdot)$ such that $\mathbb{P}_0(T(X_j) = 1) \le 1 - \delta$ is attained by the likelihood ratio test of $X_j \sim N(0,\sigma^2)$ versus the mixture normal $X_j \sim \frac{1}{2}N(a,\sigma^2) + \frac{1}{2}N(-a,\sigma^2)$, which is

$$T(X) = I\Big(\cosh(aX/\sigma^2) \ge c^*\exp(\frac{a^2}{2\sigma^2})\Big), \tag{B.30}$$

where $\cosh(x) = (\exp(x) + \exp(-x))/2$ and c^* is chosen such that $\mathbb{P}_0(T(X_j) = 0) = \delta$. Since $\cosh(x)$ is symmetric and monotonically increasing for $x > 0$, we have

$$\delta = \mathbb{P}_0\Big(|X/\sigma| \le \frac{\sigma}{a}\cosh^{-1}(c^* \exp(\frac{a^2}{2\sigma^2}))\Big) = 1 - 2\Phi\Big(-\frac{\sigma}{a}\cosh^{-1}(c^* \exp(\frac{a^2}{2\sigma^2}))\Big).$$

Solving above equation, we obtain

$$c^* = \exp(-\frac{a^2}{2\sigma^2})\cosh\Big(\frac{a}{\sigma}\Phi^{-1}(\frac{1+\delta}{2})\Big).$$

Denote $\Delta_{TS} = \frac{u_+ + u_-}{2} = \mathbb{P}_a(T(X_j) = 0)/2 + \mathbb{P}_{-a}(T(X_j) = 0)/2$ with $T(X)$ defined in (B.30). Then,

$$\begin{aligned}\Delta_{TS} &= \frac{1}{2}\mathbb{P}_a\Big(\cosh(aX/\sigma^2) \le c^* \exp(\frac{a^2}{2\sigma^2})\Big) + \frac{1}{2}\mathbb{P}_{-a}\Big(\cosh(aX/\sigma^2) \le c^* \exp(\frac{a^2}{2\sigma^2})\Big)\\ &= \frac{1}{2}\mathbb{P}_a\Big(|X/\sigma| \le \Phi^{-1}(\frac{1+\delta}{2})\Big) + \frac{1}{2}\mathbb{P}_{-a}\Big(|X/\sigma| \le \Phi^{-1}(\frac{1+\delta}{2})\Big)\\ &= \Phi(\Phi^{-1}(\frac{1+\delta}{2}) + \frac{a}{\sigma}) - \Phi(-\Phi^{-1}(\frac{1+\delta}{2}) + \frac{a}{\sigma}).\end{aligned}$$

Following the same steps in the proof of Theorem 1, we can obtain (A.2).

To show (A.4), we consider the following two cases separately. Case (1): $a/\sigma \le \Phi^{-1}(\frac{1+\delta}{2})$. Then

$$\Delta_{TS} \ge \Phi\Big(\Phi^{-1}(\frac{1+\delta}{2})\Big) - \Phi(0) \ge \frac{c}{2},$$

since $\delta \ge c$ for some constant $c > 0$. As a result, $(1 + \Delta_{TS})^s \to \infty$ as $s \to \infty$. This yields (A.4).

Case (2): $\Phi^{-1}(\frac{1+\delta}{2}) < a/\sigma \le \Phi^{-1}(\frac{1+\delta}{2}) - \Phi^{-1}(c_s/s)$. Denote $g(x) = \Phi(\Phi^{-1}(\frac{1+\delta}{2}) + x) - \Phi(-\Phi^{-1}(\frac{1+\delta}{2}) + x)$. We have $\Delta_{TS} = g(a/\sigma)$. Note that the function $g(x)$ is monotonically decreasing for $x \ge \Phi^{-1}(\frac{1+\delta}{2})$. This implies that

$$\begin{aligned}\Delta_{TS} &\ge g\Big(\Phi^{-1}(\frac{1+\delta}{2}) - \Phi^{-1}(c_s/s)\Big)\\ &= \Phi\Big(2\Phi^{-1}(\frac{1+\delta}{2}) + \Phi^{-1}(1 - c_s/s)\Big) - \Phi(\Phi^{-1}(1 - c_s/s))\\ &\ge \Phi(c' + T(s/c_s)) - (1 - c_s/s), \qquad \text{(B.31)}\end{aligned}$$

where $c' = 2\Phi^{-1}(\frac{1+c}{2})$ and $T(x) = \sqrt{2\log x - \log\log x - C}$ and the last step holds by Lemma 16. Applying Lemma 16 again yields

$$\begin{aligned}\Phi(c' + T(s/c_s)) &\ge 1 - \exp\Big(-\frac{1}{2}(c' + T(s/c_s))^2\Big)\\ &\ge 1 - C\exp\Big(-\log(s/c_s) + \frac{1}{2}\log\log(s/c_s) - c'\sqrt{\log(s/c_s)}\Big)\\ &= 1 - C\frac{c_s}{s}\exp\Big(\frac{1}{2}\log\log(s/c_s) - c'\sqrt{\log(s/c_s)}\Big), \qquad \text{(B.32)}\end{aligned}$$

where C is a generic constant which may differ from line to line. Since s/c_s tends to infinity, $\log\log(s/c_s) \ll \sqrt{\log(s/c_s)}$. Combining (B.31) and (B.32), we obtain

$$\Delta_{TS} \geq \frac{c_s}{s}\left[1 - C\exp\left(\frac{1}{2}\log\log(s/c_s) - c'\sqrt{\log(s/c_s)}\right)\right] \geq \frac{c_s}{2s}.$$

Following the same steps in the proof of Theorem 1, we can obtain $(\Delta_{TS}+1)^s \to \infty$. This completes the proof of (A.4).

To show $\widehat{S}^{TS}_{\alpha'} \in \mathcal{F}(\delta)$, notice that

$$\mathbb{P}_0(j \in \widehat{S}^{TS}_{\alpha'}) \leq \mathbb{P}_0(|X_j|/\sigma \geq \Phi^{-1}(1/2+\delta/2)) = 1-\delta.$$

The event $\mathrm{supp}(\boldsymbol{\theta}) \not\subseteq \widehat{S}^{TS}_{\alpha'}$ is equivalent to that there exists $j \in [d]$ such that $j \in \mathrm{supp}(\boldsymbol{\theta})$ and $j \notin \widehat{S}^{TS}_{\alpha'}$. Then

$$\begin{aligned}\mathbb{P}_{\boldsymbol{\theta}}(\mathrm{supp}(\boldsymbol{\theta}) \not\subseteq \widehat{S}^{TS}_{\alpha'}) &= \mathbb{P}_{\boldsymbol{\theta}}(\exists j \in [d], j \in \mathrm{supp}(\boldsymbol{\theta}), j \notin \widehat{S}^{TS}_{\alpha'})\\ &\leq \sum_{j:\theta_j \neq 0} \mathbb{P}_{\theta_j}(j \notin \widehat{S}^{TS}_{\alpha'})\\ &= \sum_{j:\theta_j \neq 0} \mathbb{P}(|X_j| \leq (\sigma\Phi^{-1}(\frac{\alpha'}{2s}) + a)_+),\end{aligned}$$

where the last step holds since $a/\sigma \geq \Phi^{-1}(\frac{\delta+1}{2}) - \Phi^{-1}(\frac{\alpha'}{2s})$. If $\sigma\Phi^{-1}(\frac{\alpha'}{2s}) + a < 0$, the above probability is 0. Otherwise,

$$\mathbb{P}(\mathrm{supp}(\boldsymbol{\theta}) \not\subseteq \widehat{S}^{TS}_{\alpha'}) \leq \sum_{j:\theta_j \neq 0} \mathbb{P}(|\theta_j/\sigma| - |Z_j| \leq \Phi^{-1}(\frac{\alpha'}{2s}) + a/\sigma) \leq \sum_{j:\theta_j \neq 0} \mathbb{P}(|Z_j| \geq -\Phi^{-1}(\frac{\alpha'}{2s})) = \alpha',$$

where $Z_j = \frac{X_j - \theta_j}{\sigma} \sim N(0,1)$, and we use $\min_{j:\theta_j \neq 0} |\theta_j| \geq a$. This completes the proof.

B.10 *Proof of Theorem 11*

Similar to the proof of Theorem 4, we can bound $\mathbb{P}(\boldsymbol{\theta} \notin \widehat{M}^{TS}_{\alpha'})$ by

$$\mathbb{P}(\boldsymbol{\theta} \notin \widehat{M}^{TS}_{\alpha'}) \leq \mathbb{P}(\exists j \in \widehat{S}^{TS}_{\alpha'}, \theta_j \notin [\widehat{L}_j, \widehat{U}_j]) + \mathbb{P}(\mathrm{supp}(\boldsymbol{\theta}) \not\subseteq \widehat{S}^{TS}_{\alpha'}). \tag{B.33}$$

By part (3) of Theorem 10, $\mathbb{P}(\mathrm{supp}(\boldsymbol{\theta}) \not\subseteq \widehat{S}^{TS}_{\alpha'}) \leq \alpha'$. The first term can be further bounded as follows:

$$\mathbb{P}(\exists j \in \widehat{S}^{TS}_{\alpha'}, \theta_j \notin [\widehat{L}_j, \widehat{U}_j]) \tag{B.34}$$

$$\leq \mathbb{P}(\exists j \in \mathrm{supp}(\boldsymbol{\theta}), \theta_j \notin [\widehat{L}_j, \widehat{U}_j]) + \mathbb{P}(\exists j \in \widehat{S}^{TS}_{\alpha'} \backslash \mathrm{supp}(\boldsymbol{\theta}), \theta_j \notin [\widehat{L}_j, \widehat{U}_j]) := I_1 + I_2. \tag{B.35}$$

For I_1, by noting that $\theta_j \notin [\widehat{L}_j, \widehat{U}_j]$ is equivalent to $|Z_j| \geq u$, where $Z_j = \frac{X_j - \theta_j}{\sigma} \sim N(0,1)$ and $u = \widehat{u}^{TS}_{\alpha'}$, we have

$$I_1 \leq \sum_{j \in \text{supp}(\boldsymbol{\theta})} \mathbb{P}(|Z_j| \geq u) = 2s(1 - \Phi(u)). \tag{B.36}$$

To bound I_2, noting that $j \notin \text{supp}(\boldsymbol{\theta})$ implying $\theta_j = 0$, we have

$$I_2 = \mathbb{P}(\exists j \notin \text{supp}(\boldsymbol{\theta}), |Z_j| \geq \Phi^{-1}(\frac{\alpha'}{2s}) + \frac{a}{\sigma}, |Z_j| \geq u) \leq \sum_{j \notin \text{supp}(\boldsymbol{\theta})} \mathbb{P}(|Z_j| \geq \Phi^{-1}(\frac{\alpha'}{2s}) + \frac{a}{\sigma}, |Z_j| \geq u).$$

To bound the last probability, we now consider the following two cases.

(1). When $\frac{a}{\sigma} \leq -\Phi^{-1}(\frac{\alpha - \alpha'}{2d}) - \Phi^{-1}(\frac{\alpha'}{2s})$, by setting $u = \Phi^{-1}(1 - \frac{\alpha - \alpha'}{2d})$, we have $I_2 \leq 2(d-s)(1 - \Phi(u))$. Combining with (B.33), (B.35), (B.36), we have

$$\mathbb{P}(\boldsymbol{\theta} \notin \widehat{M}^{TS}_{\alpha'}) \leq 2d(1 - \Phi(u)) + \alpha' = \alpha.$$

(2). When $\frac{a}{\sigma} > -\Phi^{-1}(\frac{\alpha - \alpha'}{2d}) - \Phi^{-1}(\frac{\alpha'}{2s})$, we have $\eta > 1 - \frac{\alpha - \alpha'}{2d}$, where $\eta = \Phi(\frac{a}{\sigma} + \Phi^{-1}(\frac{\alpha'}{2s}))$. Recall that $u = \Phi^{-1}(1 - \frac{\alpha - \alpha' - 2(d-s)(1-\eta)}{2s})$. Then

$$\Phi(u) = 1 - \frac{\alpha - \alpha' - 2(d-s)(1-\eta)}{2s} < \eta,$$

which is equivalent to $\Phi^{-1}(\frac{\alpha'}{2s}) + \frac{a}{\sigma} \geq u$. Therefore, $I_2 \leq 2(d-s)(1-\eta)$, and finally we have

$$\mathbb{P}(\boldsymbol{\theta} \notin \widehat{M}^{TS}_{\alpha'}) \leq 2(d-s)(1-\eta) + 2s(1 - \Phi(u)) + \alpha' = \alpha.$$

B.11 *Proof of Theorem 12*

The idea of the proof is very similar to the proof of Theorem 6. For simplicity of presentation, we skip some intermediate steps. Denote $A^+ = \{\boldsymbol{\theta} \in \mathbb{R}^d : \|\boldsymbol{\theta}\|_0 = A, \theta_j = \rho, \text{ for any} \theta_j \neq 0\}$ and $A^- = \{\boldsymbol{\theta} \in \mathbb{R}^d : \|\boldsymbol{\theta}\|_0 = A, \theta_j = -\rho, \text{ for any} \theta_j \neq 0\}$, where $0 < A \leq s$ and ρ is an arbitrary positive quantity that is $\rho \geq a$. Then

$$\begin{aligned}
\sup_{\boldsymbol{\theta} \in \Theta(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin M) &\geq \frac{1}{2}\binom{d}{A}^{-1}\Big[\sum_{\boldsymbol{\theta} \in A^+} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin M) + \sum_{\boldsymbol{\theta} \in A^-} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin M)\Big] \\
&\geq \frac{t}{2}\binom{d}{A}^{-1}\Big[\sum_{j=1}^{d}\sum_{\boldsymbol{\theta} \in A^+} \mathbb{P}_{\theta_j}(\theta_j \notin CI_j) + \sum_{j=1}^{d}\sum_{\boldsymbol{\theta} \in A^-} \mathbb{P}_{\theta_j}(\theta_j \notin CI_j)\Big] \\
&= t\sum_{j=1}^{d}\Big(\frac{d-A}{d}\mathbb{P}_0(0 \notin CI_j) + \frac{A}{2d}\mathbb{P}_\rho(\rho \notin CI_j) + \frac{A}{2d}\mathbb{P}_{-\rho}(-\rho \notin CI_j)\Big),
\end{aligned} \tag{B.37}$$

where $t = \inf_{\boldsymbol{\theta}\in A^+\cup A^-} \prod_{j=1}^d \mathbb{P}_{\theta_j}(\theta_j \in CI_j)$. Since $\mathbb{P}_\rho(0, \rho \in CI_j) \le \mathbb{E}_\rho|U_j - L_j|/\rho$, we have

$$\mathbb{P}_\rho(\rho \notin CI_j) \ge \mathbb{P}_\rho(0 \in CI_j) - \mathbb{P}_\rho(0, \rho \in CI_j) \ge \mathbb{P}_\rho(0 \in CI_j) - m/\rho,$$

and similarly $\mathbb{P}_{-\rho}(-\rho \notin CI_j) \ge \mathbb{P}_{-\rho}(0 \in CI_j) - m/\rho$. Together with (B.37),

$$\sup_{\boldsymbol{\theta}\in\Theta(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin M) \ge t\Big\{\sum_{j=1}^d \Big(\frac{d-A}{d}\mathbb{P}_0(0 \notin CI_j) + \frac{A}{2d}\mathbb{P}_\rho(0 \in CI_j) + \frac{A}{2d}\mathbb{P}_{-\rho}(0 \in CI_j)\Big) - \frac{Am}{\rho}\Big\}$$
$$\ge tA\Big\{\inf_{T\in\{0,1\}} \Big(\frac{d-A}{A}\mathbb{E}_0(1-T) + \frac{1}{2}\mathbb{E}_\rho T + \frac{1}{2}\mathbb{E}_{-\rho}T\Big) - \frac{m}{\rho}\Big\}, \quad \text{(B.38)}$$

where T denotes a test function from $\mathbb{R}$ to $\{0, 1\}$. Note that

$$\frac{d-A}{A}\mathbb{E}_0(1-T) + \frac{1}{2}\mathbb{E}_\rho T + \frac{1}{2}\mathbb{E}_{-\rho}T = \mathbb{E}_0\Big(\frac{d-A}{A} + T(\frac{f_\rho(x) + f_{-\rho}(x)}{2f_0(x)} - \frac{d-A}{A})\Big),$$

where $f_\rho(x)$ denotes the pdf of $N(\rho, \sigma^2)$. Thus, the above function is minimized by

$$T^*(x) = I(\frac{f_\rho(x) + f_{-\rho}(x)}{2f_0(x)} - \frac{d-A}{A} \le 0) = I\Big(|x/\sigma| \le \frac{\sigma}{\rho}\cosh^{-1}(\frac{d-A}{A}\exp(\frac{\rho^2}{2\sigma^2}))\Big),$$

where $\cosh^{-1}(\cdot)$ is the inverse function of cosh. Plugging the definition of $T^*(x)$ into (B.38), after some calculation we obtain

$$\sup_{\boldsymbol{\theta}\in\Theta(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin M) \ge tA\Big\{\frac{2(d-A)}{A}\Phi(-D) + \Phi(\frac{\rho}{\sigma} + D) - \Phi(\frac{\rho}{\sigma} - D) - \frac{m}{\rho}\Big\},$$

where $D = \frac{\sigma}{\rho}\cosh^{-1}(\frac{d-A}{A}\exp(\frac{\rho^2}{2\sigma^2}))$. The rest of the proof is the same as Theorem 6. We omit the details.

B.12 *Proof of Corollary 13*

The proof of this corollary follows from the same line as in the proof of Corollary 8. We only highlight the main difference. By Theorem 12, we can obtain that

$$m \ge \rho\Big\{g_{TS}(d, A_d, \rho) - \frac{1}{A_d}\frac{\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin M)}{\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \in M)}\Big\}.$$

To show (A.11), denote $a_1 := \sigma\sqrt{2\log(d/A_d - 1)}$, and we can take $\rho = a_1$. The key step is to lower bound $g_{TS}(d, A_d, \rho)$. Recall that

$$g_{TS}(d, A_d, \rho) = \frac{2(d-A_d)}{A_d}\Phi(-D) + \Big\{\Phi\Big(\frac{\rho}{\sigma} + D\Big) - \Phi\Big(\frac{\rho}{\sigma} - D\Big)\Big\} := I_1 + I_2,$$

where

$$D = \frac{\sigma}{\rho}\cosh^{-1}(\frac{d-A_d}{A_d}\exp(\frac{\rho^2}{2\sigma^2})).$$

We now consider the two terms I_1, I_2. By the definition of cosh, we can easily verify the following inequality $\log y < \cosh^{-1}(y) < \log(2y)$ holds. Applying it to I_1 yields

$$
\begin{aligned}
I_1 &\geq \frac{2(d-A_d)}{A_d}\Phi\Big(-\frac{\sigma}{\rho}\log\frac{2(d-A_d)}{A_d}-\frac{\rho}{2\sigma}\Big)\\
&= \frac{2(d-A_d)}{A_d}\Phi\Big(-\sqrt{2\log(d/A_d-1)}-\frac{\log 2}{\sqrt{2\log(d/A_d-1)}}\Big)\\
&\geq \frac{C}{\sqrt{2\log(d/A_d-1)}},
\end{aligned}
$$

for some universal positive constant C. Similarly, we can show that

$$
\begin{aligned}
I_2 &\geq \Phi\Big(\frac{\sigma}{\rho}\log\frac{d-A_d}{A_d}+\frac{3\rho}{2\sigma}\Big)-\Phi\Big(-\frac{\sigma}{\rho}\log\frac{d-A_d}{A_d}+\frac{\rho}{2\sigma}\Big)\\
&= \Phi\Big(2\sqrt{2\log(d/A_d-1)}\Big)-\Phi(0)\\
&\geq 1/2 - C\Big(\frac{A_d}{d-A_d}\Big)^4,
\end{aligned}
$$

The same argument in the proof of Corollary 8 implies (A.11). The proof of (A.13) is similar. Finally, to show (A.12), notice that

$$
\mathbb{E}_{\boldsymbol{\theta}}(\bar{U}_j^{TS}-\bar{L}_j^{TS}) = \mathbb{E}_{\boldsymbol{\theta}}(\bar{U}_j^{TS}-\bar{L}_j^{TS})I(j\in\bar{S}_{\alpha'}^{TS}) \leq 2\sigma\bar{u}_{\alpha'}^{TS}.
$$

Regardless of the SNR, $\limsup \bar{u}_{\alpha'}^{TS}/\sqrt{2\log d} \leq 1$ always holds. Finally, for (A.14), we have

$$
a/\sigma \geq \sqrt{2\log(d-s)-\log\log(d-s)+C}+\sqrt{2\log s-\log\log s+C} \geq \bar{\phi},
$$

and therefore

$$
\bar{u}_{\alpha'}^{TS} = \sqrt{2\log\Big(\frac{4s}{(\alpha-\alpha')C_{2s,\alpha-\alpha'}}\Big)},
$$

which further implies $\limsup \bar{u}_{\alpha'}^{TS}/\sqrt{2\log s} \leq 1$. This completes the proof.

LEMMA 16. Tail bound for Gaussian distributionLet $N \sim N(0,1)$. Then for any $y > 0$

$$
\sqrt{\frac{2}{\pi}}\frac{\exp(-y^2/2)}{y+\sqrt{y^2+4}} < \mathbb{P}(N>y) \leq \sqrt{\frac{2}{\pi}}\frac{\exp(-y^2/2)}{y+\sqrt{y^2+8/\pi}}.
$$

Conversely, for any $t > 2$,

$$
\sqrt{(2\log t-\log\log t-C)_+} \leq \Phi^{-1}(1-\frac{1}{t}) \leq \sqrt{2\log t-\log\log t},
$$

where $C = 2\log 4+\log\pi$.

C. Additional Theoretical Results

C.1 $\bar{L}_j$ *in the sparse confidence set*

PROPOSITION 17. Assume that $d, s \to \infty$ and δ, α, α' are fixed. For $\theta_j = a$, we have the following result.

(2)If $\kappa^{**} \le a/\sigma < \bar{\kappa}$, and $a \le m^* := \sigma\sqrt{2\log d}/2$, then

$$\lim_{d,s\to\infty} \mathbb{P}(\bar{L}_j = 0) = 1.$$

(2)If $a/\sigma \ge \bar{\kappa}$, $d \ge 2s$, then

$$\lim_{d,s\to\infty} \mathbb{P}(\bar{L}_j = 0) = 0.$$

Proof. By the definition of $\bar{L}_j$, we know that $\bar{L}_j = 0$ is equivalent to $X_j/\sigma \le \bar{u}_{\alpha'}$, which is $N(0,1) \le \bar{u}_{\alpha'} - a/\sigma$. In case 1, we have

$$\mathbb{P}(\bar{L}_j = 0) = \Phi\Big(\sqrt{2\log\Big(\frac{d}{(\alpha-\alpha')C_{d,\alpha-\alpha'}}\Big)} - \frac{a}{\sigma}\Big)$$
$$\ge \Phi\Big(\sqrt{2\log\Big(\frac{d}{(\alpha-\alpha')C_{d,\alpha-\alpha'}}\Big)} - \frac{1}{2}\sqrt{2\log d}\Big) \ge \Phi\big((\frac{1}{2}-\epsilon)\sqrt{2\log d}\big),$$

where ϵ is an arbitrarily small constant. By the Gaussian tail bound in Lemma 16, we have $\bar{L}_j = 0$ with probability tending to 1.

Similarly, in case 2, under $d \ge 2s$, we have

$$\mathbb{P}(\bar{L}_j = 0) = \Phi\Big(\sqrt{2\log\Big(\frac{2s}{(\alpha-\alpha')C_{s,\alpha-\alpha'}}\Big)} - \frac{a}{\sigma}\Big)$$
$$\le \Phi\Big(-\sqrt{2\log\Big(\frac{s}{C_{s,\alpha'}\alpha'}\Big)}\Big) \le c\frac{C_s}{s},$$

for some constant $c > 0$, where we use the Gaussian tail bound in Lemma 16 again in the last step. Letting $s \to \infty$, we complete the proof. □

C.2 *Extension to* $X \sim N(\theta, \Sigma)$

In this section, we show that the proposed sparse confidence set is still valid under the more general Gaussian model $\boldsymbol{X} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ for some *known* PD matrix $\boldsymbol{\Sigma}$. Without loss of generality, we assume the diagonal entries of $\boldsymbol{\Sigma}$ are identical and equals to σ^2. Under this assumption, our sparse confidence sets are exactly identical to $\widehat{M}_{\alpha'}$. Recall that

$$\widehat{M}_{\alpha'} = M(\widehat{S}_{\alpha'}, \widehat{\boldsymbol{U}}, \widehat{\boldsymbol{L}}),\ \text{where}\ \widehat{L}_j = (X_j - \widehat{u}_{\alpha'}\sigma)_+,\ \widehat{U}_j = +\infty,$$

where

$$\widehat{S}_{\alpha'} = \Big\{ j \in [d] : \frac{X_j}{\sigma} \geq (\Phi^{-1}(\frac{\alpha'}{s}) + \frac{a}{\sigma}) \vee \Phi^{-1}(\delta) \Big\}.$$

THEOREM 18. For any $0 < \alpha' < \alpha$, provided (2.6) holds, under the model $\boldsymbol{X} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$, we have

$$\sup_{\boldsymbol{\theta} \in \Theta^+(s,a), \theta_j = 0} \mathbb{P}_{\boldsymbol{\theta}}(j \in \widehat{S}_{\alpha'}) \leq 1 - \delta, \quad \text{and} \quad \sup_{\boldsymbol{\theta} \in \Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \notin \widehat{M}_{\alpha'}) \leq \alpha,$$

where $\widehat{u}_{\alpha'}$ in (2.8) is given by

$$\widehat{u}_{\alpha'} = \begin{cases} \Phi^{-1}\Big(1 - \frac{\alpha - \alpha'}{d}\Big) & \text{if } a/\sigma \in R_L, \\ \Phi^{-1}\Big(1 - \frac{\alpha - \alpha' - (d-s)(1-\eta^+)}{s}\Big) & \text{if } a/\sigma \in R_H, \end{cases}$$

where $\eta^+ = \Phi(\frac{a}{\sigma} + \Phi^{-1}(\frac{\alpha'}{s}))$, and the low SNR region R_L and high SNR region R_H are defined in the same way as those in Theorem 4.

It is easily seen that the conclusion of this theorem is identical to Theorem 4. In other words, our sparse confidence set $\widehat{M}_{\alpha'}$ is robust to the correlation among $\boldsymbol{X}$. This result is not surprising, because the proof of Theorem 4 does not rely on the independence assumption of $\boldsymbol{X}$. We also note that, the proof of Theorem 9 does require the independence assumption of $\boldsymbol{X}$, see the proof of Lemma 14. Thus, the current adaptivity results are not valid under the general model $\boldsymbol{X} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$.

C.3 *Suboptimality of $\widehat{M}_{\alpha'}$*

Recall that

$$\widehat{M}_{\alpha'} = M(\widehat{S}_{\alpha'}, \widehat{\boldsymbol{U}}, \widehat{\boldsymbol{L}}), \text{ where } \widehat{L}_j = (X_j - \widehat{u}_{\alpha'}\sigma)_+, \ \widehat{U}_j = +\infty$$

for any $j \in \widehat{S}_{\alpha'}$, where

$$\widehat{S}_{\alpha'} = \Big\{ j \in [d] : \frac{X_j}{\sigma} \geq (\Phi^{-1}(\frac{\alpha'}{s}) + \frac{a}{\sigma}) \vee \Phi^{-1}(\delta) \Big\}.$$

We partition the SNR region into low and high levels:

- Low SNR region: $R_L = \big\{\kappa : \kappa^* \leq \kappa < \kappa^* \vee \widehat{\kappa}\big\}$,
- High SNR region: $R_H = \big\{\kappa : \kappa \geq \kappa^* \vee \widehat{\kappa}\big\}$,

  where

$$\widehat{\kappa} = -\Phi^{-1}(\frac{\alpha - \alpha'}{d}) - \Phi^{-1}(\frac{\alpha'}{s}),$$

and

$$\widehat{u}_{\alpha'} = \begin{cases} \Phi^{-1}\Big(1 - \frac{\alpha - \alpha'}{d}\Big) & \text{if } a/\sigma \in R_L, \\ \Phi^{-1}\Big(1 - \frac{\alpha - \alpha' - (d-s)(1-\eta^+)}{s}\Big) & \text{if } a/\sigma \in R_H. \end{cases}$$

PROPOSITION 19. For $\theta_j = a := C_d\sigma\sqrt{\log d}$, where $C_d/s \to \infty$, then as $d, s \to \infty$ (assuming α, δ are fixed)

$$\mathbb{E}(\theta_j - \widehat{L}_j) \asymp \sigma\frac{C_d}{s}\sqrt{\log d}.$$

Proof. Similar to the proof of Corollary 8, we can show that

$$\mathbb{E}(\theta_j - \widehat{L}_j) = \widehat{u}_{\alpha'}\sigma + \mathbb{E}(\theta_j - X_j)I(X_j/\sigma > \widehat{u}_{\alpha'})I(X_j/\sigma > B)$$
$$+ \mathbb{E}(\theta_j - \widehat{u}_{\alpha'}\sigma)I(X_j/\sigma < \widehat{u}_{\alpha'})I(X_j/\sigma > B) + \mathbb{E}(\theta_j - \widehat{u}_{\alpha'}\sigma)I(X_j/\sigma < B),$$

where $B = (\Phi^{-1}(\frac{\alpha'}{s}) + \frac{a}{\sigma}) \vee \Phi^{-1}(\delta)$. We denote these terms by $I_1, \ldots, I_4$. For I_2, we can apply Cauchy–Schwarz inequality to get $I_2 \le \sigma$. For I_3, we have

$$|I_3| \le (\theta_j - \widehat{u}_{\alpha'}\sigma)\mathbb{P}(Z > \theta_j/\sigma - \widehat{u}_{\alpha'}) \le \frac{\sigma}{\sqrt{2\pi}},$$

where the last error bound is from the proof of Corollary 8. For I_4, since $\theta_j = a := C_d\sigma\sqrt{\log d}$, we have

$$I_4 = \sigma(C_d\sqrt{\log d} - \widehat{u}_{\alpha'})\frac{\alpha'}{s}.$$

It is easily seen that $C_d\sqrt{\log d} \gg \widehat{u}_{\alpha'}$. As $C_d/s \to \infty$, we have

$$\mathbb{E}(\theta_j - \widehat{L}_j) \asymp \sigma\frac{C_d}{s}\sqrt{\log d}.$$

The proof is complete. □

Note that, since C_d in the above proposition can be arbitrarily large, the length of the confidence interval $\mathbb{E}(\theta_j - \widehat{L}_j) \asymp \sigma\frac{C_d}{s}\sqrt{\log d} \gg \sigma\sqrt{\log d}$, where the latter is the minimax rate in case 1 in Corollary 8. Thus, the sparse confidence interval $\widehat{M}_{\alpha'}$ is suboptimal.

C.4 *Extension to non-separable sparse confidence sets*

Recall that we define the class of separable selectors as

$$\mathcal{F}(\delta) = \{S(\boldsymbol{X}) : \mathbb{P}_0(j \in S(\boldsymbol{X})) \le 1 - \delta,$$
$$\text{and the event } \{j \in S(\boldsymbol{X})\} \text{ only depends on } X_j \text{ for any } j \in [d]\}.$$

In this section, we define non-separable sparse confidence intervals as

$$\bar{\mathcal{M}}_+ = \{M(S, \boldsymbol{U}, \boldsymbol{L}) : \inf_{\boldsymbol{\theta}\in\Theta^+(s,a)} \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \in M(S, \boldsymbol{U}, \boldsymbol{L})) \ge 1 - \alpha, \text{ and } S \in \mathcal{F}(\delta)\}, \tag{C1}$$

where $\mathcal{F}(\delta)$ is the class of separable selectors defined above. Recall that we consider one-sided confidence intervals for the parameters in $\Theta^+(s, a)$, so we still require $L_j \ge 0$ and $U_j = +\infty$ if $j \in S$, otherwise $L_j = U_j = 0$. In the definition of $\bar{\mathcal{M}}_+$, we remove the constraint from CI_+ that $L_j \le X_j \vee 0$ and L_j only depends on X_j.

Now we prove the following non-asymptotic lower bound over the class of non-separable confidence intervals in $\bar{\mathcal{M}}_+$, which is parallel to Corollary 3.3.

THEOREM 20. Assume that $\alpha < 1/2$ and let $\eta = 1 - 2\alpha$.

(1). If $\kappa^{**} \le a/\sigma \le \sqrt{\log(d\eta^2 + 1)}$, then

$$\inf_{M \in \bar{\mathcal{M}}_+} R(M, \Theta^+(s,a)) \ge (1/2 - \alpha)\sigma\sqrt{\log(d\eta^2 + 1)}. \tag{C2}$$

(2). If $a/\sigma \ge \sqrt{\log(d\eta^2 + 1)}$, then

$$\inf_{M \in \bar{\mathcal{M}}_+} R(M, \Theta^+(s,a)) \ge (1/2 - \alpha)\sigma\sqrt{\log(s\eta^2 + 1)}. \tag{C3}$$

We can see that, as $d, s \to \infty$, in low SNR (case 1), the lower bound over the class of non-separable confidence intervals is $O(\sigma\sqrt{\log d})$, whereas, in high SNR (case 2), the lower bound is $O(\sigma\sqrt{\log s})$. Compared with the asymtotic results in Corollary 3.3 (for separable confidence intervals) and the follow-up discussion, the lower bounds for separable and non-separable confidence intervals have the same order as $d, s \to \infty$.

Proof. The result is proved based on the following Le Cam's method [4]. Define $\mathbb{P}_\pi(A) = \int \mathbb{P}_{\boldsymbol{\theta}}(A) d\pi(\boldsymbol{\theta})$ to be the probability of the event A when $\boldsymbol{X}|\boldsymbol{\theta}$ follows the specified Gaussian mean model and $\boldsymbol{\theta} \sim \pi$ for some distribution π on $\Theta^+(s,a)$. For any fixed $\boldsymbol{\theta}' \in \Theta^+(s,a)$, we can show that

$$|\mathbb{P}_{\boldsymbol{\theta}'}(\boldsymbol{\theta}' \in \mathcal{M}(U,L)) - \mathbb{P}_\pi(\boldsymbol{\theta}' \in \mathcal{M}(U,L))| \le \sup_A |\mathbb{P}_{\boldsymbol{\theta}'}(A) - \mathbb{P}_\pi(A)| \le \frac{1}{2}[\mathbb{E}_{\boldsymbol{\theta}'}(L_\pi^2) - 1]^{1/2}, \tag{C4}$$

where $L_\pi = \frac{d\mathbb{P}_\pi}{d\mathbb{P}_{\boldsymbol{\theta}'}}(\boldsymbol{X})$ is the likelihood ratio of the data under the measure $\mathbb{P}_\pi$ versus $\mathbb{P}_{\boldsymbol{\theta}'}$. Since $\boldsymbol{\theta}' \in \Theta^+(s,a)$, we have $\mathbb{P}_{\boldsymbol{\theta}'}(\boldsymbol{\theta}' \in \mathcal{M}(U,L)) \ge 1 - \alpha$. Combining with (C4), we obtain

$$\mathbb{P}_\pi(\boldsymbol{\theta}' \in \mathcal{M}(U,L)) \ge 1 - \alpha - \frac{1}{2}[\mathbb{E}_{\boldsymbol{\theta}'}(L_\pi^2) - 1]^{1/2}. \tag{C5}$$

Finally, by noting that

$$\mathbb{P}_\pi(\boldsymbol{\theta} \in \mathcal{M}(U,L)) = \int \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta} \in \mathcal{M}(U,L)) d\pi(\boldsymbol{\theta}) \ge (1-\alpha)\int d\pi(\boldsymbol{\theta}) = 1 - \alpha,$$

(C5) further implies

$$\mathbb{P}_\pi(\boldsymbol{\theta}, \boldsymbol{\theta}' \in \mathcal{M}(U,L)) \ge 1 - 2\alpha - \frac{1}{2}[\mathbb{E}_{\boldsymbol{\theta}'}(L_\pi^2) - 1]^{1/2}. \tag{C6}$$

We pick $\boldsymbol{\theta}' = (0, ..., 0)$ and define π in the following way. Let $\widehat{m}$ be a random index set (with size 1) drawn uniformly in $\{1, 2, ..., d\}$. We define π to be the distribution of the random variable $\sum_{j \in \widehat{m}} \kappa \boldsymbol{e}_j$, where κ is a positive quantity to be determined. It is easy to verify the following lemma:

LEMMA 21. Under the construction of π above, it holds that

$$\mathbb{E}_0(L_\pi^2) = 1 + \frac{1}{d}\Big(\exp(\frac{\kappa^2}{\sigma^2}) - 1\Big).$$

By taking $\kappa^2 = \sigma^2 \log(d\eta^2 + 1)$, (C6) implies $\mathbb{P}_\pi(0, \boldsymbol{\theta} \in \mathcal{M}(U, L)) \geq 1/2 - \alpha$. Let $\bar{\boldsymbol{\theta}}_j = \kappa \boldsymbol{e}_j$. Then

$$\begin{aligned} 1/2 - \alpha \leq \mathbb{P}_\pi(\boldsymbol{\theta}, 0 \in \mathcal{M}(U, L)) &= \frac{1}{d}\sum_{j=1}^{d} \mathbb{P}_{\bar{\boldsymbol{\theta}}_j}(\bar{\boldsymbol{\theta}}_j, 0 \in \mathcal{M}(U, L)) \\ &\leq \frac{1}{d}\sum_{j=1}^{d} \mathbb{E}_{\bar{\boldsymbol{\theta}}_j} I(0, \kappa \in CI_j) \\ &\leq \frac{1}{d}\frac{1}{\kappa}\sum_{j=1}^{d} \mathbb{E}_{\bar{\boldsymbol{\theta}}_j}(\theta_j - L_j) \\ &\leq \frac{1}{\kappa}\sup_j \sup_{\boldsymbol{\theta}} \mathbb{E}_{\boldsymbol{\theta}}(\theta_j - L_j). \end{aligned}$$

Then we have $\sup_j \sup_{\boldsymbol{\theta}} \mathbb{E}_{\boldsymbol{\theta}}(\theta_j - L_j) \geq \kappa(1/2 - \alpha)$, which completes the proof of (C.2). The next result (C.3) follows the similar proof which is omitted. □