

Minimax estimation of functionals in sparse vector model with correlated observations

Yuhao Wang¹, Pengkun Yang² and Alexandre Tsybakov³

¹*Institute for Interdisciplinary Information Sciences, Tsinghua University and Shanghai Qi Zhi Institute, e-mail: yuhaow@tsinghua.edu.cn*

²*Department of Statistics and Data Science, Tsinghua University, e-mail: yangpengkun@tsinghua.edu.cn*

³*CREST, ENSAE, Institut Polytechnique de Paris, e-mail: Alexandre.Tsybakov@ensae.fr*

Abstract: We consider the observations of an unknown s -sparse vector θ corrupted by Gaussian noise with zero mean and unknown covariance matrix Σ . We propose minimax optimal methods of estimating the ℓ_2 norm of θ and testing the hypothesis $H_0 : \theta = 0$ against sparse alternatives when only partial information about Σ is available, such as an upper bound on its Frobenius norm and the values of its diagonal entries to within an unknown scaling factor. We show that the minimax rates of the estimation and testing are leveraged not by the dimension of the problem but by the value of the Frobenius norm of Σ .

MSC2020 subject classifications: [Primary] 62J05, 62G05.

Keywords and phrases: Correlated noise, Gaussian sequence model, Nonasymptotic minimax estimation, Sparsity, Unknown noise level.

1. Introduction

In this paper, we consider the Gaussian vector model

$$Y = \theta + \sigma \varepsilon, \quad (1)$$

where $\varepsilon \sim \mathcal{N}(0, \Sigma)$ follows a d -dimensional zero-mean Gaussian distribution with covariance matrix Σ , and $\sigma > 0$ is the noise level. We assume throughout that the diagonal elements σ_i^2 of matrix Σ are known (the main example is $\sigma_i^2 = 1$ for all i , that is, we deal with a normalized matrix), while its off-diagonal elements are not, and σ is either known or unknown. Based on the data $Y = (Y_1, \dots, Y_d)$, we consider the problem of estimating the quadratic functional and the ℓ_2 -norm of the parameter $\theta = (\theta_1, \dots, \theta_d)$:

$$Q(\theta) = \sum_{i=1}^d \theta_i^2, \quad \|\theta\|_2 = \sqrt{Q(\theta)}.$$

We assume that θ belongs to the class $B_0(s, d)$ of all vectors $\theta \in \mathbb{R}^d$ such that $\|\theta\|_0 \leq s$ (s -sparse vectors), where s is an integer in $[1, d]$. Here, $\|\theta\|_0$ denotes the number of non-zero components of θ . When σ is unknown, we also study the problem of estimating σ .

Most of the existing literature on estimation of functionals and on the related problem of testing under sparse vector models is focused on the case of uncorrelated data (Σ is a diagonal matrix). However, in many applications such as time series analysis one needs to deal with

dependent observations. In this paper, we study the setting where only partial information on Σ is available, namely, either the Frobenius norm of Σ or an upper bound is known. We propose estimators of $Q(\theta)$ and $\|\theta\|_2$ that are adaptive to the noise level and are rate optimal in a minimax sense for $\|\theta\|_2$. As a consequence, we construct minimax optimal tests for signal detection. Furthermore, we propose minimax optimal estimators of the noise level σ .

When $\Sigma = I$, where I is the identity matrix, and σ is known, Collier, Comminges and Tsybakov [6] proved that the rate of the minimax quadratic risk for estimating the ℓ_2 -norm $\|\theta\|_2$ of a s -sparse vector is equal to $\sigma^2\phi(s, d)$, where for any $t > 0$,

$$\phi(s, t) := \begin{cases} s \log(1 + t/s^2) & \text{if } s \leq \sqrt{t}, \\ \sqrt{t} & \text{otherwise.} \end{cases} \quad (2)$$

Moreover, Collier, Comminges and Tsybakov [6] derived the minimax rates for estimation of $Q(\theta)$, and for testing the hypothesis $H_0 : \theta = \mathbf{0}$ against sparse alternatives separated from $\mathbf{0}$ in the ℓ_2 -norm (improving the result of Baraud [2] by matching the upper and lower bounds). Estimation of $\|\theta\|_2$ is related to such a testing problem in the sense that upper bounds on the risks for the problem of estimating $\|\theta\|_2$ imply upper bounds for testing, while lower bounds for the testing problem imply lower bounds for estimating $\|\theta\|_2$. Comminges et al. [7] considered the model (1) with unknown σ and uncorrelated noise (not necessarily Gaussian), and derived the minimax rates of estimating the functionals $Q(\theta)$ and $\|\theta\|_2$ as well as of the noise level σ on the class of s -sparse vectors. In particular, Comminges et al. [7] proved that, when the noise is i.i.d. Gaussian and σ is unknown, the rate of the minimax quadratic risk for estimating the ℓ_2 -norm of a s -sparse vector is equal to $\sigma^2\phi^*(s, d)$, where for any $t > 0$,

$$\phi^*(s, t) = \begin{cases} s \log(1 + t/s^2) & \text{if } s \leq \sqrt{t}, \\ \frac{s}{1 \vee \log(s/\sqrt{t})} & \text{otherwise.} \end{cases} \quad (3)$$

Carpentier et al. [3, 4] further extended these results to the model of high-dimensional sparse regression with i.i.d. noise. Model (1) in the case of diagonal matrix Σ with known entries and $\sigma = 1$ was analyzed in Laurent, Loubes and Marteau [15], Chhor, Mukherjee and Sen [5]. Under such particular conditions, Laurent, Loubes and Marteau [15] studied the minimax rates for the problem of testing mentioned above. In Chhor, Mukherjee and Sen [5] these results are extended to testing of $H_0 : \theta = \mathbf{0}$ against sparse alternatives separated from $\mathbf{0}$ in the ℓ_p -norm with $1 \leq p < \infty$ and further to estimation of the corresponding norms. Another recent development due to George and Canonne [9] shows that introducing an arbitrary contamination in model (1) with $\Sigma = I$ leads to slower minimax rates (for the problem of testing with ℓ_2 -separation) than those given in [6]. Finally, there exist works allowing for correlated noise, again focused on the problem of testing, see Hall and Jin [10], Liu, Gao and Samworth [16], Kotekal and Gao [13]. The paper [16] deals with a change-point framework and establishes the minimax rates of testing for the particular case $s = d$ (no sparsity) assuming that matrix Σ is fixed and positive definite. In [13] the authors consider model (1) with $\sigma = 1$, equicorrelation matrix $\Sigma = (1 - \gamma)I + \gamma\mathbf{1}\mathbf{1}^\top$, where $\gamma > 0$, and $\mathbf{1}$ is a d -dimensional column vector with all entries equal to 1, and establish the minimax rates of testing the hypothesis $H_0 : \theta = \mathbf{0}$ against sparse alternatives separated from $\mathbf{0}$ in the ℓ_2 -norm assuming that γ is known. They show that, for some choices of γ , the detection boundary can be smaller than its value when the noise is uncorrelated. In a more recent work [14], the same authors find the minimax rates of estimating linear functionals under the equicorrelated Gaussian noise. A

different asymptotic instance of the same testing problem is considered in Hall and Jin [10], where it is assumed that all non-zero components of θ have the same value $r\sqrt{\log d}$ for some $r > 0$, the sparsity parameter satisfies $s = d^a$ for some $a \in (0, 1/2)$, and the positions of non-zero components are randomly generated. In this model, it is proved in Hall and Jin [10] that if Σ is a Toeplitz matrix, then the asymptotic (as $d \rightarrow \infty$) detection boundary can be smaller than its value when the noise is uncorrelated. The improvement is not in the rate but in the constant factor.

In summary, the existing literature dealing with our non-asymptotic setting (mostly focused on the problem of testing rather than estimation of functionals) allows one to tackle the problem under sparsity when Σ is either a diagonal matrix with known entries or the equicorrelation matrix, and in the latter case all parameters of the model except θ need to be known. However, under general correlation structure, the question of how to estimate functionals of a sparse vector θ in a minimax optimal way and how to construct the associated tests remains open. In this paper, we focus on the problem of estimating the functionals $Q(\theta)$ and $\|\theta\|_2$ and we consider the general setting, with arbitrary structure of Σ and with the noise level σ that can be either known or unknown. Moreover, the estimators that we propose do not require knowing the off-diagonal entries of Σ . They need only partial information about Σ such as the relative scale of the diagonal entries, the value of its Frobenius norm or an upper bound on it. We construct estimators of $\|\theta\|_2$ that are optimal in a minimax sense and are adaptive to the noise level σ . We also derive the associated minimax rates of testing. The minimax rates that we obtain for estimating $\|\theta\|_2$ under known and unknown σ are of the form (2) and (3), respectively, with $t = \|\Sigma\|_F^2$, where $\|\Sigma\|_F$ is the Frobenius norm of Σ . If $\Sigma = I$ we recover the rates $\sigma^2\phi(s, d)$ and $\sigma^2\phi^*(s, d)$ obtained in [6, 7]. For general Σ , we arrive at an interesting conclusion that the possibility of faster or slower rates is leveraged not by the dimension d but by the value of the Frobenius norm of Σ .

Obtaining the results for the model with correlated noise is challenging. The tools used in [7] for the setting with unknown σ , including exponential inequalities and sample splitting to obtain multiple independent estimators, break down when the observations are correlated. We develop new proof techniques allowing us to obtain a slight improvement upon [7] even in the case of i.i.d. observations, which consists in the fact that our estimator does not need sample splitting.

This paper is organized as follows. In Section 2, we discuss estimation of $Q(\theta)$ and $\|\theta\|_2$ when σ is known. In Section 3, we introduce an estimator of σ . Section 4 deals with the problem of estimating $\|\theta\|_2$ when σ is unknown. In Section 5, we consider testing of the null hypothesis $\|\theta\|_2 = 0$. Section 6 is devoted to the proofs.

Notation. Let $[n] := \{1, \dots, n\}$. For $1 \leq q \leq \infty$, let $\|\cdot\|_q$ denote the ℓ_q -norm in $\mathbb{R}^d$, and let $B_q(r) := \{x \in \mathbb{R}^d : \|x\|_q \leq r\}$ be the ℓ_q -ball of radius r . We denote by Φ the cumulative distribution function of the standard Gaussian distribution. For $x, y \in \mathbb{R}$, we set $x \wedge y := \min\{x, y\}$ and $x \vee y := \max\{x, y\}$. We use standard big- O notations, e.g., for two positive sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n = O(b_n)$ or $a_n \lesssim b_n$ if $a_n \leq Cb_n$ for some absolute constant $C > 0$; $a_n = \Omega(b_n)$ or $a_n \gtrsim b_n$ if $b_n = O(a_n)$; $a_n = \Theta(b_n)$ or $a_n \asymp b_n$ if $a_n = O(b_n)$ and $b_n = O(a_n)$. Given a matrix $A \in \mathbb{R}^{d \times d}$ and $\mathcal{I}, \mathcal{J} \subseteq [d]$, we denote by $A_{\mathcal{I}, \mathcal{J}}$ the submatrix of A with rows from $\mathcal{I}$ columns from $\mathcal{J}$, by A_{ij} the entries of A , and by $\|A\|_F$, $\lambda_{\max}(A)$, and $\text{tr}[A]$, the Frobenius norm, the maximal eigenvalue, and the trace of A , respectively. We denote by $\mathbf{E}_{\theta, \Sigma}$ the expectation with respect to the probability measure $\mathbf{P}_{\theta, \Sigma}$.

of random vector Y satisfying (1). We denote by c, C, C' absolute positive constants that can vary from line to line.

2. Estimating the quadratic functional and the ℓ_2 -norm when σ is known

In this section, we consider the case of known noise level σ . Throughout the paper, we assume the following condition on the diagonal entries of the covariance matrix.

Assumption 1. The diagonal elements $\sigma_i^2, i = 1, \dots, d$, of the covariance matrix matrix Σ satisfy $\max_{1 \leq i \leq d} \sigma_i^2 \leq 1$.

Assumption 1 is just a normalization condition. It assures that σ can be considered as the noise level. As Σ is a covariance matrix Assumption 1 implies that all entries of Σ are bounded by 1. Note also that we can replace the condition $\max_{1 \leq i \leq d} \sigma_i^2 \leq 1$ by $\max_{1 \leq i \leq d} \sigma_i^2 \leq c$, where $c > 0$ is an absolute constant, with no incidence on the results.

Assume first that the Frobenius norm of the covariance matrix Σ is given. Then we define the estimator $\hat{Q}$ of the quadratic functional $Q(\theta)$ as

$$\hat{Q} = \begin{cases} \sum_{i=1}^d (Y_i^2 - \sigma^2 \sigma_i^2 \beta_s) \mathbb{1}\{|Y_i| > \sigma \sigma_i \tau\}, & s \leq \|\Sigma\|_F, \\ \sum_{i=1}^d (Y_i^2 - \sigma^2 \sigma_i^2), & \text{otherwise,} \end{cases}$$

where $\beta_s := \mathbf{E} [Z^2 | |Z| \geq \tau]$ with $Z \sim \mathcal{N}(0, 1)$, and $\tau = 3\sqrt{\log(1 + \|\Sigma\|_F^2/s^2)}$. The associated estimator of the ℓ_2 -norm $\|\theta\|_2$ is defined as

$$\hat{N} = \sqrt{|\hat{Q}|}.$$

Set

$$\psi(s, \Sigma) := \phi(s, \|\Sigma\|_F^2) = \begin{cases} s \log(1 + \|\Sigma\|_F^2/s^2) & \text{if } s \leq \|\Sigma\|_F, \\ \|\Sigma\|_F & \text{otherwise,} \end{cases}$$

and

$$\bar{\psi}(s, \Sigma) := \lambda_{\max}(\Sigma) \wedge s.$$

Let $\mathcal{S} := \{i : \theta_i \neq 0\}$ denote the support of θ . The following theorem provides upper bounds for the mean square risks of the estimators $\hat{Q}$ and $\hat{N}$.

Theorem 1. *There exists an absolute constant $C > 0$ such that, for any $\theta \in B_0(s, d)$ and any Σ satisfying Assumption 1,*

$$\mathbf{E}_{\theta, \Sigma}[(\hat{Q} - \|\theta\|_2^2)^2] \leq C \left(\sigma^2 \|\theta\|_2^2 \lambda_{\max}(\Sigma_{\mathcal{S}, \mathcal{S}}) + \sigma^4 \psi^2(s, \Sigma) \right). \quad (4)$$

Furthermore, for any Σ satisfying Assumption 1,

$$\sup_{\theta \in B_0(s, d) \cap B_2(r)} \mathbf{E}_{\theta, \Sigma}[(\hat{Q} - \|\theta\|_2^2)^2] \leq C \left(\sigma^2 r^2 \bar{\psi}(s, \Sigma) + \sigma^4 \psi^2(s, \Sigma) \right) \quad (5)$$

and

$$\sup_{\theta \in B_0(s, d)} \mathbf{E}_{\theta, \Sigma}[(\hat{N} - \|\theta\|_2)^2] \leq C \sigma^2 \psi(s, \Sigma). \quad (6)$$

It is interesting to compare (6) to the bound for the case $\Sigma = \mathbf{I}$ [6], that is, $\sigma^2\phi(s, d)$, where $\phi(s, d)$ is defined in (2). Since Σ is a covariance matrix, Assumption 1 implies that $\|\Sigma\|_F$ is at most of order d . Particularly, in the extreme case $\|\Sigma\|_F \asymp d$, there is no elbow in the rate and $\psi(s, \Sigma) \asymp s \log(1 + d^2/s^2)$ for all $1 \leq s \leq d$. Thus, we get a substantial deterioration in the zone $s > \sqrt{d}$ compared to the rate $\sigma^2\phi(s, d)$ with uncorrelated data. On the contrary, for the other extreme case $\|\Sigma\|_F \asymp \sqrt{d}$, we have $\psi(s, \Sigma) \asymp \phi(s, d)$ and we recover the rate obtained in [6] for $\Sigma = \mathbf{I}$. Finally, if Σ is a very sparse matrix, with very small diagonal elements (or some diagonal elements equal to 0, that is, no noise in some components of θ), then the rate can be further improved.

If only an upper bound on $\|\Sigma\|_F$ is available, we can modify the definition of $\hat{N}$ by using this upper bound in place of $\|\Sigma\|_F$. We now show that such an estimator is minimax optimal in the following sense. Consider the class of parameters

$$\Theta(s, d, \rho) := \{(\theta, \Sigma) \in \mathbb{R}^d \times \mathbb{R}^{d \times d} : \theta \in B_0(s, d), \|\Sigma\|_F \leq \rho \text{ and } \Sigma_{ii} = 1 \text{ for } i = 1, \dots, d\},$$

where $\rho \in (0, d]$ is a given value. Note that it makes no sense to consider $\rho > d$, since the bound $\|\Sigma\|_F \leq d$ is trivially satisfied if $\Sigma_{ii} = 1$ for $i = 1, \dots, d$. The following lower bound holds.

Theorem 2. *Let $w : [0, \infty) \rightarrow [0, \infty)$ be a monotone non-decreasing function such that $w(0) = 0$ and $w \not\equiv 0$. Then there exists a constant $c > 0$ depending only on $w(\cdot)$ such that*

$$\inf_{\hat{T}} \sup_{(\theta, \Sigma) \in \Theta(s, d, \rho)} \mathbf{E}_{\theta, \Sigma} w(\sigma^{-1}(\phi(s, \rho^2))^{-1/2} |\hat{T} - \|\theta\|_2|) \geq c,$$

where $\inf_{\hat{T}}$ denotes the infimum over all estimators.

Let $\tilde{N}$ be an estimator defined in the same way as $\hat{N}$ with the only difference that $\|\Sigma\|_F$ is replaced by its upper bound ρ . The following theorem shows that $\tilde{N}$ is minimax optimal over the set $\Theta(s, d, \rho)$ with respect to the mean square risk.

Theorem 3. *We have*

$$\sup_{(\theta, \Sigma) \in \Theta(s, d, \rho)} \mathbf{E}_{\theta, \Sigma} [(\tilde{N} - \|\theta\|_2)^2] \lesssim \sigma^2 \phi(s, \rho^2), \quad (7)$$

and

$$\inf_{\hat{T}} \sup_{(\theta, \Sigma) \in \Theta(s, d, \rho)} \frac{\mathbf{E}_{\theta, \Sigma} [(\hat{T} - \|\theta\|_2)^2]}{\phi(s, \rho^2)} \asymp \sigma^2. \quad (8)$$

We omit the proof of (7) as it follows exactly the same lines as the proof of (6) with $\|\Sigma\|_F$ replaced by ρ . The minimax property (8) follows by combining (7) and Theorem 2 with $w(u) = u^2$.

3. Estimating the noise level

In what follows, we consider the model with unknown σ and we impose the following additional assumption on the diagonal elements of the matrix Σ .

Assumption 2. There exists a constant $c_* > 0$ such that the diagonal elements σ_i^2 of Σ satisfy the condition $\min_{1 \leq i \leq d} \sigma_i^2 \geq c_*$.

Thus, the minimum of σ_i^2 's is controlled by a positive constant that does not depend on d . It implies that $\|\Sigma\|_F \gtrsim \sqrt{d}$. On the other hand, in this section, we do not need Assumption 1.

Under Assumption 2, it is convenient to work with the normalized observations $\tilde{Y}_i = \tilde{\theta}_i + \sigma \tilde{\varepsilon}_i$ for $i \in [d]$, where $\tilde{\theta}_i = \theta_i/\sigma_i$, and $\tilde{\varepsilon}_i = \varepsilon_i/\sigma_i$ is a standard Gaussian random variable. Thus, we replace the covariance matrix Σ by the correlation matrix $\tilde{\Sigma}$ with entries

$$\tilde{\Sigma}_{ij} = \Sigma_{ij}/(\sigma_i \sigma_j), \quad i, j \in [d].$$

In this section, we propose estimators for the unknown noise level σ^2 that will be further used to construct the estimators of functionals adaptive to the noise level. We first provide an estimator $\hat{\sigma}_S^2$, which is rate-optimal in the sparse regime $s \leq \|\tilde{\Sigma}\|_F$. Next, we construct a refined estimator $\hat{\sigma}_D^2$ that attains the optimal rate for all $s \lesssim d$.

3.1. Estimating σ^2 in the sparse regime

The first estimator $\hat{\sigma}_S^2$, designed for the sparse regime, is inspired by the techniques of robust estimation [12]. Note that all observations $\tilde{Y}_i$ with $i \notin \mathcal{S}$ are distributed as $\mathcal{N}(0, \sigma^2)$, and thus estimating σ^2 can be equivalently viewed as a robust estimation problem, where the observations with non-zero $\tilde{\theta}_i$'s play the role of outliers. In particular, let $\hat{F}_{Y^2}$ and $\hat{F}_{\varepsilon^2}$ be the empirical cumulative distribution functions of $\{\tilde{Y}_i^2 : i \in [d]\}$ and $\{\tilde{\varepsilon}_i^2 : i \in [d]\}$, respectively. Then,

$$\sup_{t>0} |\hat{F}_{Y^2}(t) - \hat{F}_{\varepsilon^2}(t/\sigma^2)| \leq \frac{\|\theta\|_0}{d}. \quad (9)$$

Considering here some fixed $t > 0$, approximating $\hat{F}_{\varepsilon^2}$ by the cumulative distribution function of the chi-squared distribution with one degree of freedom χ_1^2 (that we denote by F) and solving $\hat{F}_{Y^2}(t) \approx F(t/\sigma^2)$ leads to the following estimator of σ^2 :

$$\hat{\sigma}_S^2(t) := \frac{t}{F^{-1}(\hat{F}_{Y^2}(t))}. \quad (10)$$

A relevant choice of t in (10) is to be determined. Notice that for any deterministic $t > 0$ this estimator is not well-defined since the denominator in (10) can be 0 with positive probability. Indeed, for any $t > 0$ there exist $\theta \in B_0(s, d)$ such that $\mathbf{P}_{\theta, \Sigma}[\hat{F}_{Y^2}(t) = 0] > 0$. Thus, we will only define $\hat{\sigma}_S^2(t)$ by (10) if the denominator in (10) is non-zero. Otherwise, we set $\hat{\sigma}_S^2(t) = \sum_{i=1}^d Y_i^2/d$; we note that the conclusion remains valid for any other choice of this value. The next lemma provides a bound on the estimation error of $\hat{\sigma}_S^2(t)$ for any fixed $t > 0$.

Lemma 1. *There exists an absolute constant $C > 0$ such that for any $\theta \in B_0(s, d)$ and any Σ satisfying Assumption 2 the following holds.*

(i) *For any $t > 0$ and $0 \leq u \leq 1/2$ we have*

$$\mathbf{P}_{\theta, \Sigma} \left[\left| \frac{\hat{\sigma}_S^2(t)}{\sigma^2} - 1 \right| \geq u \right] \leq \frac{C}{u^2} \left(\frac{s^2}{d^2} + \frac{\|\tilde{\Sigma}\|_F^2 \left(\left(\frac{t}{\sigma^2} \right)^2 + 1 \right) e^{-\frac{t}{3\sigma^2}}}{d^2} \right) \frac{e^{2t/\sigma^2}}{t/\sigma^2}.$$

(ii) If, in addition, $|t/\sigma^2 - 1| \leq \frac{2}{3}$ and $s/d \leq 0.01$ then

$$\mathbf{P}_{\theta, \Sigma} \left[\left| \frac{\hat{\sigma}_S^2(t)}{\sigma^2} - 1 \right| \geq \frac{1}{2} \right] \leq C \frac{\|\tilde{\Sigma}\|_F^2}{d^2}.$$

Lemma 1 suggests that the estimator $\hat{\sigma}_S^2(t)$ achieves a reasonable accuracy for $t \asymp \sigma^2$. However, such choice of t is not realizable since σ is unknown. Moreover, for any deterministic $t > 0$ the expected error of $\hat{\sigma}_S^2(t)$ is not bounded uniformly over θ since $\hat{\sigma}_S^2(t)$ can be $+\infty$ with some positive probability as pointed out above. Therefore, we turn to an alternative solution that consists in choosing a random $t = \hat{t}$ based on the sample. Note that the standard sample splitting strategy, under which $\hat{t}$ can be independently chosen, is not a viable option since we have a model with dependent observations. We avoid sample splitting and propose a data-driven threshold $\hat{t}$ determined by scanning $\hat{F}_{Y^2}$ over a dyadic grid, where $\hat{F}_{Y^2}$ is based on the same set of observations as $\hat{\sigma}_S^2(\cdot)$ ¹:

$$\hat{t} := \min \{t = 2^\ell : \ell \in \mathbb{Z}, \hat{F}_{Y^2}(t) \geq 0.5\}. \quad (11)$$

We consider now the following estimator of σ^2 :

$$\hat{\sigma}_S^2 := \hat{\sigma}_S^2(\hat{t}).$$

The next theorem gives a bound on the estimation error of $\hat{\sigma}_S^2$.

Theorem 4. *There exist absolute constants $C > 0$ and $c > 0$ such that for any $s \leq cd$, any $\theta \in B_0(s, d)$ and any Σ satisfying Assumption 2 we have*

$$\mathbf{E}_{\theta, \Sigma} \left[\left| \frac{\hat{\sigma}_S^2 - \sigma^2}{\sigma^2} \right| \right] \leq C \left(\frac{s}{d} + \frac{\|\tilde{\Sigma}\|_F}{d} \right).$$

It can be seen from the proof of Theorem 4 that the threshold (11) obtained by a search on a dyadic grid satisfies $\hat{t} \asymp \sigma^2$ with high probability. An alternative approach could be to use the sample median $\hat{t} = \min\{t : \hat{F}_{Y^2}(t) \geq 0.5\}$, which is also of the order of σ^2 with high probability. The main limitation of this approach is that, under the dependent data, the resulting estimator $\hat{\sigma}^2(\hat{t})$ becomes difficult to analyze. On the other hand, as we prove below, the value $\hat{t}$ defined in (11) belongs with high probability to a *finite set*

$$\mathcal{T} := \left\{ t = 2^\ell : \ell \in \mathbb{Z}, \frac{\sigma^2}{3} \leq t \leq \frac{3}{2}\sigma^2 \right\}.$$

This allows us to control $\mathbf{E}_{\theta, \Sigma} \left[\left| \frac{\hat{\sigma}_S^2 - \sigma^2}{\sigma^2} \right| \right]$ by analyzing the convergence of $\hat{F}_{Y^2}(t)$ for every fixed $t \in \mathcal{T}$. In other words, with our dyadic grid search we restrict the “degrees of freedom” in choosing $\hat{t}$, which makes the resulting estimator more robust against data dependence. In the next section, we provide another estimator of σ^2 that outperforms $\hat{\sigma}_S^2$ in the dense regime.

¹The choice of constant 0.5 in the definition of $\hat{t}$ is not crucial. Replacing 0.5 by any other constant within $(0, 1)$ only changes the resulting rates by a constant factor.

3.2. Estimating σ^2 in the dense regime

Note that the estimator (10) can be viewed as a method of moments estimator based on the empirical moments of indicator functions of $\tilde{Y}_i^2$'s. We can develop a similar argument with the empirical moments of another family of bounded functions, namely, the cosine functions of $\tilde{Y}_i$'s. We consider the following empirical moments:

$$\hat{\varphi}_d(u) := \frac{1}{d} \sum_{i=1}^d \cos(u\tilde{Y}_i), \quad u > 0.$$

Analogously to (9), when s is small compared to d we can approximate $|\hat{\varphi}_d(u)|$ by the term $|\frac{1}{d} \sum_{i=1}^d \cos(u\sigma\tilde{\epsilon}_i)|$, which in turn is approximated by the value of the normal characteristic function $\varphi(u) = \exp(-u^2\sigma^2/2)$. Solving $\exp(-u^2\sigma^2/2) \approx |\hat{\varphi}_d(u)|$ suggests $-\frac{2}{u^2} \log |\hat{\varphi}_d(u)|$ as an estimator of σ^2 , under some choice of u .

We set $\lambda = \frac{1 \vee \log(s/\|\tilde{\Sigma}\|_F)}{6}$ and consider the estimator

$$\tilde{\sigma}_D^2 := -\frac{2\hat{t}}{\lambda} \log \left| \hat{\varphi}_d(\sqrt{\lambda/\hat{t}}) \right|,$$

where $\hat{t}$ is defined in (11). Our final estimator of the noise level is $\hat{\sigma}_D^2 := \min\{\tilde{\sigma}_D^2, 2\hat{\sigma}_S^2\}$.

Theorem 5. *There exist absolute constants $C > 0$ and $c > 0$ such that for any $s \leq cd$, any $\theta \in B_0(s, d)$ and any Σ satisfying Assumption 2 we have*

$$\mathbf{E}_{\theta, \Sigma} \left[\frac{|\hat{\sigma}_D^2 - \sigma^2|}{\sigma^2} \right] \leq C \left(\frac{s}{d(1 \vee \log(s/\|\tilde{\Sigma}\|_F))} + \frac{\|\tilde{\Sigma}\|_F}{d} \right).$$

If we specify the difference between sparse and dense regimes, the bound of Theorem 5 exhibits the rate

$$\tilde{\psi}(s, \Sigma) := \begin{cases} \frac{\|\tilde{\Sigma}\|_F}{d} & \text{if } s \leq \|\tilde{\Sigma}\|_F, \\ \frac{s}{d(1 \vee \log(s/\|\tilde{\Sigma}\|_F))} & \text{otherwise.} \end{cases}$$

Theorem 5 provides a generalization of the upper bound proved in [7, Proposition 7] for the case of i.i.d. noise. Compared to the techniques used in [7], our correlated noise setting brings fundamental challenges. The analysis in [7] relies on exponential concentration inequalities for the uniform deviations that we are short of as the observations are dependent and the spectral norm of Σ can be as large as $O(\sqrt{d})$. The sample splitting technique for obtaining multiple independent estimates used in [7] also breaks down when the observations are correlated. It turns out that for our estimator, that does not use any sample splitting, second-order moment inequalities suffice to achieve the desired rate. Note that we obtain an improvement upon Proposition 7 in [7] even in the case of i.i.d. observations since our estimator does not need sample splitting.

4. Estimating the ℓ_2 -norm when the noise level is unknown

In this section, we use the estimators of σ^2 proposed in Section 3 in order to construct estimators of the ℓ_2 -norm $\|\theta\|_2$ when the noise level σ is unknown.

Assume first that the Frobenius norm of the covariance matrix $\|\Sigma\|_F$ is known. We start by considering an estimator of $\|\theta\|_2$ defined as follows:

$$\hat{N}^* := \sqrt{|\hat{Q}^*|},$$

where

$$\hat{Q}^* := \begin{cases} \sum_{i=1}^d Y_i^2 \mathbb{1}\{|Y_i| > \hat{\sigma}_S \sigma_i \tau\} - \hat{\sigma}_S^2 \alpha_s \sum_{i=1}^d \sigma_i^2, & \text{if } s \leq \|\Sigma\|_F, \\ \sum_{i=1}^d (Y_i^2 - \hat{\sigma}_D^2 \sigma_i^2), & \text{otherwise.} \end{cases}$$

Here, $\alpha_s := \mathbf{E}[Z^2 \mathbb{1}\{|Z| \geq \tau\}]$ with $Z \sim \mathcal{N}(0, 1)$, and $\tau = 3\sqrt{\log(1 + \|\Sigma\|_F^2/s^2)}$. Set

$$\psi^*(s, \Sigma) := \phi^*(s, \|\Sigma\|_F^2) = \begin{cases} s \log(1 + \|\Sigma\|_F^2/s^2) & \text{if } s \leq \|\Sigma\|_F, \\ \frac{s}{1 \vee \log(s/\|\Sigma\|_F)} & \text{otherwise.} \end{cases}$$

Our aim in this section is to show that $\sigma^2 \psi^*(s, \Sigma)$ is the minimax optimal rate for the squared error of estimating $\|\theta\|_2$ on $\Theta(s, d)$ if Assumptions 1 and 2 hold, and that the estimator $\hat{N}^*$ or its modifications attain this rate. The following theorem provides bounds on the accuracy of the estimator $\hat{N}^*$.

Theorem 6. *There exists an absolute constant $\alpha \in (0, 1)$ such that, for any $\theta \in B_0(s, d)$ with s satisfying that $s \leq \alpha d$ and any Σ satisfying Assumptions 1 and 2, the following holds.*

(i) *In the regime $s > \|\Sigma\|_F$, there exists an absolute constant $c_1 > 0$ such that*

$$\mathbf{E}_{\theta, \Sigma}[(\hat{N}^* - \|\theta\|_2)^2] \leq c_1 \sigma^2 \psi^*(s, \Sigma).$$

(ii) *In the regime $s \leq \|\Sigma\|_F$, there exists an absolute constant $c_2 > 0$ such that, for any $\eta \in (c_2 \|\Sigma\|_F/d, 1)$, the following inequality holds with some constant $c_\eta > 0$ depending only on η :*

$$\mathbf{P}_{\theta, \Sigma} \left((\hat{N}^* - \|\theta\|_2)^2 \leq c_\eta \sigma^2 \psi^*(s, \Sigma) \right) \geq 1 - \eta. \quad (12)$$

(iii) *If, in addition, Σ is a positive definite matrix there exists an absolute constant $c_3 > 0$ such that*

$$\begin{aligned} \mathbf{E}_{\theta, \Sigma}[(\hat{N}^* - \|\theta\|_2)^2] &\leq c_3 \left(\sqrt{\lambda_{\max}(\Sigma^{-1})} + 1 \right) \sigma^2 \psi^*(s, \Sigma) \\ &\quad + c_3 \sigma^2 \frac{\|\Sigma\|_F^2}{d} \log(1 + \|\Sigma\|_F^2/s^2). \end{aligned} \quad (13)$$

Theorem 6 grants an expected squared risk bound with the desired rate $\sigma^2 \psi^*(s, \Sigma)$ in the dense regime $s > \|\Sigma\|_F$. However, in the sparse regime $s \leq \|\Sigma\|_F$ Theorem 6 implies such a conclusion only under additional conditions. The bound (12) provides a bound in probability under an additional (Σ -dependent) condition on η . Part (iii) of Theorem 6 implies that, under the additional assumption that Σ is positive definite with minimum eigenvalue bounded below by some positive constant and $\|\Sigma\|_F^2 \leq sd$, the proposed estimator achieves the rate $\sigma^2 \psi^*(s, \Sigma)$ for the expected squared risk also in the sparse regime.

The bound (12) is valid only under the condition that η is bigger than $c_2 \|\Sigma\|_F/d$, which is not negligible when $\|\Sigma\|_F \asymp d$. We now show that one can construct an estimator dependent

of η that is free of this constraint. Namely, for any given $\eta \in (0, 1)$, we propose an estimator depending on η that achieves (12). Fix $\eta \in (0, 1)$ and let q_η be the η -th quantile of a chi-square random variable with one degree of freedom. Define

$$\hat{\sigma}_\eta^2 := \frac{\hat{M}}{q_{1-\eta/20}}, \quad \text{where} \quad \hat{M} := \min\{t : \hat{F}_Y^2(t) \geq 0.5\}.$$

We now modify $\hat{Q}^*$ in the sparse regime $s \leq \|\Sigma\|_F$ as follows:

$$\hat{Q}_\eta^* := \sum_{i=1}^d Y_i^2 \mathbb{1}\{|Y_i| > \max\{\hat{\sigma}_S, \hat{\sigma}_\eta\} \sigma_i \tau_\eta\} - \hat{\sigma}_S^2 \alpha_{s,\eta} \sum_{i=1}^d \sigma_i^2,$$

where

$$\tau_\eta = 3 \sqrt{\frac{q_{1-\eta/20}}{q_{\eta/20}} \log(1 + \|\Sigma\|_F^2/s^2)} \quad \text{and} \quad \alpha_{s,\eta} = \mathbf{E}[Z^2 \mathbb{1}\{|Z| \geq \tau_\eta\}].$$

The corresponding estimator of the ℓ_2 -norm $\hat{N}_\eta^* := \sqrt{|\hat{Q}_\eta^*|}$ achieves the following bound in probability in the sparse regime.

Theorem 7. *There exists a constant $\alpha \in (0, 1)$ such that for any $\eta \in (0, 1)$, any $\theta \in B_0(s, d)$ with s satisfying that $s \leq \min\{\alpha d, \|\Sigma\|_F\}$, and any Σ satisfying Assumptions 1 and 2 we have*

$$\mathbf{P}_{\theta, \Sigma} \left((\hat{N}_\eta^* - \|\theta\|_2)^2 \leq c_\eta \sigma^2 \psi^*(s, \Sigma) \right) \geq 1 - \eta,$$

where $c_\eta > 0$ is a constant depending only on η .

Next, we consider the setting where the Frobenius norm $\|\Sigma\|_F$ is not available and we only know an upper bound ρ on $\|\Sigma\|_F$. In this setting, we show that estimators that are based on slight modifications of $\hat{N}^*$ and $\hat{N}_\eta^*$ can achieve minimax optimality on the parameter spaces $\Theta(s, d, \rho)$, $\rho > 0$, where the restriction $\|\Sigma\|_F \leq \rho$ is active. Let $\tilde{N}^*$ and $\tilde{N}_\eta^*$ be the estimators defined in the same way as $\hat{N}^*$ and $\hat{N}_\eta^*$, respectively, with the only difference that $\|\Sigma\|_F$ is replaced by its upper bound ρ . Let $\phi^*(s, \cdot)$ be defined in (3). Then $\phi^*(s, \rho^2)$ follows the same definition as $\psi^*(s, \Sigma)$ except that we replace $\|\Sigma\|_F$ by ρ .

Theorem 8. *Let $w : [0, \infty) \rightarrow [0, \infty)$ be a monotone non-decreasing function such that $w(0) = 0$ and $w \not\equiv 0$. Assume that there exists a constant $\alpha \in (0, 1)$ such that $s \leq \alpha d$. Then there exists a constant $c > 0$ depending only on $w(\cdot)$ such that*

$$\inf_{\hat{T}} \sup_{(\theta, \Sigma) \in \Theta(s, d, \rho)} \sup_{\sigma > 0} \mathbf{E}_{\theta, \Sigma} w(\sigma^{-1}(\phi^*(s, \rho^2))^{-1/2} |\hat{T} - \|\theta\|_2|) \geq c.$$

Moreover,

(i) *In the regime $s > \rho$,*

$$\sup_{(\theta, \Sigma) \in \Theta(s, d, \rho)} \sup_{\sigma > 0} \frac{\mathbf{E}_{\theta, \Sigma}[(\tilde{N}^* - \|\theta\|_2)^2]}{\sigma^2} \lesssim \phi^*(s, \rho^2).$$

(ii) *In the regime $s \leq \rho$, for any $\eta \in (0, 1)$ there exists a constant $c_\eta > 0$ depending only on η such that*

$$\sup_{(\theta, \Sigma) \in \Theta(s, d, \rho)} \sup_{\sigma > 0} \mathbf{P}_{\theta, \Sigma} \left((\tilde{N}_\eta^* - \|\theta\|_2)^2 \geq c_\eta \sigma^2 \phi^*(s, \rho^2) \right) \leq \eta.$$

5. Optimal testing for signal detection

In this section, we provide some corollaries of the above results for the problem of testing for signal detection. We now introduce the notation $\mathcal{M}_\rho := \{\Sigma : \Sigma_{ii} = 1, i \in [d], \text{ and } \|\Sigma\|_F \leq \rho\}$, where $\rho > 0$; then problem of testing hypotheses can be defined as follows:

$$H_0 : \theta = 0, \Sigma \in \mathcal{M}_\rho \quad \text{against} \quad H_1 : (\theta, \Sigma) \in \Theta(s, d, \rho) : \|\theta\|_2 \geq r,$$

where the parameter $r > 0$ characterizes a separation between the two hypotheses. When r is too small there is no hope to construct a good test for this problem. The main question is to find the smallest, in a minimax sense, separation radius r , under which one can obtain a valid test and to provide such a test.

For any test Ψ , that is, any measurable function of $\mathbf{Y}$ with values in $\{0, 1\}$, we define its risk as the sum of type I and type II error probabilities:

$$\mathcal{R}(\Psi, s, d, \rho, r) = \sup_{\Sigma \in \mathcal{M}_\rho} \mathbf{P}_{0, \Sigma}(\Psi = 1) + \sup_{(\theta, \Sigma) \in \Theta(s, d, \rho) : \|\theta\|_2 \geq r} \mathbf{P}_{\theta, \Sigma}(\Psi = 0).$$

The minimax risk is the smallest risk among all tests Ψ , that is,

$$\mathcal{R}^*(s, d, \rho, r) = \inf_{\Psi} \mathcal{R}(\Psi, s, d, \rho, r).$$

Assume that we are given a tolerance level $\eta \in (0, 1)$ for the minimax risk. The *minimax rate of testing* $r_\eta^*(s, d, \rho)$ is defined as

$$r_\eta^*(s, d, \rho) = \inf \{r > 0 : \mathcal{R}^*(s, d, \rho, r) \leq \eta\}.$$

A natural approach to testing H_0 against H_1 is to reject H_0 when some estimator of the ℓ_2 -norm $\|\theta\|_2$ exceeds a suitably chosen threshold. We use this approach with the estimators of the ℓ_2 -norm proposed above and the thresholds derived from the optimal rates of estimation that we have obtained. Thus, consider the test

$$\Psi^* = \mathbb{1}(\hat{N} > \gamma\sigma\sqrt{\phi(s, \rho^2)}),$$

where $\gamma > 0$ is a constant. Introduce the separation radius

$$\bar{r}_\gamma = 2\gamma\sigma\sqrt{\phi(s, \rho^2)}.$$

Using the upper bound (7) and Chebyshev's inequality we can control the risk of Ψ^* for the separation radius $\bar{r}_\gamma$ as follows:

$$\begin{aligned} \mathcal{R}(\Psi^*, s, d, \rho, \bar{r}_\gamma) &\leq \sup_{\Sigma \in \mathcal{M}_\rho} \mathbf{P}_{0, \Sigma}(\hat{N} > \gamma\sigma\sqrt{\phi(s, \rho^2)}) \\ &\quad + \sup_{(\theta, \Sigma) \in \Theta(s, d, \rho)} \mathbf{P}_{\theta, \Sigma}(\hat{N} - \|\theta\|_2 \leq -\gamma\sigma\sqrt{\phi(s, \rho^2)}) \\ &\leq 2 \sup_{(\theta, \Sigma) \in \Theta(s, d, \rho)} \frac{\mathbf{E}_{\theta, \Sigma}[(\hat{N} - \|\theta\|_2)^2]}{\gamma^2\sigma^2\phi(s, \rho^2)} \leq C^*\gamma^{-2}, \end{aligned}$$

where $C^* > 0$ is an absolute constant. Choosing $\gamma = \sqrt{\eta/C^*}$ implies that the last expression does not exceed η , so that

$$r_\eta^*(s, d, \rho) \leq 2\sqrt{\eta/C^*}\sigma\sqrt{\phi(s, \rho^2)}. \quad (14)$$

On the other hand, a matching lower bound on $r_\eta^*(s, d, \rho)$ follows from the next proposition.

Proposition 1. *There exists an absolute constant $c_4 > 0$ such that given any $\gamma > 0$, we have*

$$\mathcal{R}^*(s, d, \rho, \bar{r}_\gamma) \geq \exp(1 - \exp(c_4 \gamma^2))/2.$$

For any $\eta \in (0, 1/2)$, the equation $\exp(1 - \exp(c_4 \gamma^2))/2 = \eta$ has the solution

$$\gamma_\eta = \sqrt{(c_4)^{-1} \log \log \left(\frac{e}{2\eta} \right)}.$$

Thus, for $\eta \in (0, 1/2)$, we get $\mathcal{R}^*(s, d, \rho, \gamma_\eta \sigma \sqrt{\phi(s, \rho^2)}) \geq \eta$, which implies the bound

$$r_\eta^*(s, d, \rho) \geq \gamma_\eta \sigma \sqrt{\phi(s, \rho^2)}, \quad \forall \eta \in (0, 1/2). \quad (15)$$

Combining (14) and (15) we obtain the following corollary.

Corollary 1. *For any $\eta \in (0, 1/2)$, there exist constants c_η, C_η depending only on η such that the minimax rate of testing satisfies*

$$c_\eta \sigma \sqrt{\phi(s, \rho^2)} \leq r_\eta^*(s, d, \rho) \leq C_\eta \sigma \sqrt{\phi(s, \rho^2)}.$$

The results of this section deal with minimax optimality on a class of matrices, while the prior work on sparse testing discussed in the Introduction was analyzing minimax optimality for a fixed matrix Σ . In those papers, H_0 was a simple hypothesis (with a given matrix Σ , which was assumed diagonal, except for [10, 13, 16]), while here we deal with minimax optimality over classes of matrices Σ , and in our case H_0 is a composite hypothesis. Of course, for specific fixed matrices in the class one gets, in general, better rates than those achieved uniformly over a class. This can be seen in the setting with equicorrelation matrix detailed in Kotekal and Gao [13], and in Hall and Jin [10] for a different instance of sparse testing problem in asymptotics as $d \rightarrow \infty$.

In [16, Theorem 7] the authors provide the minimax rate for a testing problem in a change-point framework, which implies the minimax rate of the order $\sqrt{\|\Sigma\|_F}$ for testing in model (1) under the no sparsity scenario $s = d$ with fixed positive definite matrix Σ and $\sigma = 1$. Note that it agrees with our conclusions for this particular setting. Indeed, when Σ is fixed with entries independent of d we have $\|\Sigma\|_F \lesssim d$, so that for $s = d$ our Theorem 1 gives the rate $\psi(s, \Sigma) = \|\Sigma\|_F$ for the squared loss leading to the $\sqrt{\|\Sigma\|_F}$ rate of testing for this particular case.

Remark 1. An interesting phenomenon described in Hall and Jin [10] and further noticed in other context by Kotekal and Gao [13] consists in the fact that, for some specific families of matrices Σ , the minimax rate of testing can be better than in the case of independent noise, and the hardest subproblem corresponds to the identity covariance matrix. In contrast, in our setting this is not the case. Indeed, inspection of the proof shows that our lower bound is obtained on correlated data that correspond to block-diagonal matrices and not to the identity covariance matrix. The independent case appears in the proof only as a technical element – after a suitable reduction we apply the lower bounds for the independent case from Collier, Comminges and Tsybakov [6]. Our contribution is a reduction argument that leads to a contraction of the original problem to testing two composite hypotheses on a smaller scale. Importantly, the reduction step itself does not rely on the hypotheses being composite. The intuition is that, with a stronger correlation on Σ , the effective sample size of the noise vector ϵ becomes smaller, with the effective sample size being directly governed by the parameter

ρ . In our lower bound construction, this reduction is realized through a block-wise covariance structure, where the block size is also determined by ρ . Consequently, the estimation error of the mean parameter within each block is amplified by the block size. This construction provides a convenient and analytically tractable instance that captures the essential difficulty of the problem.

6. Proofs

6.1. Properties of correlated Gaussian variables

In this section, we collect some facts about correlated Gaussian variables that will be useful in the proofs. We show that several functions of Gaussian random variables scale quadratically with the correlation coefficients, which will be crucial to obtain the optimal dependence of the rates of convergence on the Frobenius norm of the covariance matrix.

Lemma 2. *There exists an absolute constant $C > 0$ such that the following statements hold for two standard Gaussian random variables ζ and η that are jointly normally distributed with correlation coefficient ν .*

(i) $\text{Cov}[\zeta^2, \eta^2] = 2\nu^2$.

(ii) For any $\tau \geq 1$, let $\beta = \mathbf{E}[Z^2 | |Z| > \tau]$ where $Z \sim \mathcal{N}(0, 1)$. Then

$$\text{Cov}[(\zeta^2 - \beta)\mathbb{1}(|\zeta| > \tau), (\eta^2 - \beta)\mathbb{1}(|\eta| > \tau)] \leq C\nu^2\tau^4 \exp(-\tau^2/2).$$

(iii) For any $\tau \geq 0$,

$$|\text{Cov}[\mathbb{1}(\zeta^2 \leq \tau), \mathbb{1}(\eta^2 \leq \tau)]| \leq C\nu^2(\tau^2 + 1)e^{-\tau/3}.$$

(iv) For any $t, \mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$ we have

$$\begin{aligned} & |\text{Cov}[\cos(t(\mu_1 + \sigma_1\zeta)), \cos(t(\mu_2 + \sigma_2\eta))]| \\ & \leq C(|\nu^2 \cos(t\mu_1) \cos(t\mu_2)| + |\nu \sin(t\mu_1) \sin(t\mu_2)|). \end{aligned}$$

Proof. Given $\zeta = u$, the conditional distribution of η is $\mathcal{N}(\nu u, 1 - \nu^2)$. Therefore, η has the same distribution as $\nu\zeta + \sqrt{1 - \nu^2}Z$ where Z is a standard Gaussian random variable independent of ζ .

(i) Let $\eta = \nu\zeta + \sqrt{1 - \nu^2}Z$. We obtain

$$\begin{aligned} \text{Cov}[\zeta^2, \eta^2] &= \mathbf{E}[(\zeta^2 - 1)(\eta^2 - 1)] = \mathbf{E}[(\zeta^2 - 1)((\nu\zeta + \sqrt{1 - \nu^2}Z)^2 - 1)] \\ &= \mathbf{E}[(\zeta^2 - 1)(\nu^2\zeta^2 + (1 - \nu^2)Z^2 + 2\nu\sqrt{1 - \nu^2}\zeta Z - 1)] \\ &= \nu^2\mathbf{E}[(\zeta^2 - 1)\zeta^2] + 2\nu\sqrt{1 - \nu^2}\mathbf{E}[(\zeta^2 - 1)\zeta Z] + \mathbf{E}[(\zeta^2 - 1)((1 - \nu^2)Z^2 - 1)] \\ &= 2\nu^2. \end{aligned}$$

(ii) See [3, Lemma 5].

- (iii) We consider separately the cases $|\nu| \geq \frac{1}{5}$ and $|\nu| < \frac{1}{5}$. For $|\nu| \geq \frac{1}{5}$, using the Cauchy-Schwarz inequality we get

$$|\mathbf{Cov}[\mathbb{1}(\zeta^2 \leq \tau), \mathbb{1}(\eta^2 \leq \tau)]| \leq \mathbf{Var}[\mathbb{1}(\eta^2 \leq \tau)] \leq \mathbf{P}[\eta^2 > \tau] \leq C\nu^2 e^{-\tau/2},$$

where $C > 0$ is an absolute constant.

Next we prove the result for $|\nu| < \frac{1}{5}$. Given $\zeta = u$, the conditional distribution of η is $\mathcal{N}(\nu u, 1 - \nu^2)$. Then,

$$\begin{aligned} \mathbf{Cov}[\mathbb{1}(\zeta^2 \leq \tau), \mathbb{1}(\eta^2 \leq \tau)] &= \mathbf{P}[\zeta^2 \leq \tau, \eta^2 \leq \tau] - \mathbf{P}[\zeta^2 \leq \tau]\mathbf{P}[\eta^2 \leq \tau] \\ &= \int_{-\sqrt{\tau}}^{\sqrt{\tau}} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \left(\Phi\left(\frac{\sqrt{\tau} - \nu u}{\sqrt{1 - \nu^2}}\right) - \Phi\left(\frac{-\sqrt{\tau} - \nu u}{\sqrt{1 - \nu^2}}\right) - \Phi(\sqrt{\tau}) + \Phi(-\sqrt{\tau}) \right) du. \end{aligned} \quad (16)$$

For the expression under the integral in (16) we have

$$\begin{aligned} &\left| \Phi\left(\frac{\sqrt{\tau} - \nu u}{\sqrt{1 - \nu^2}}\right) - \Phi\left(\frac{-\sqrt{\tau} - \nu u}{\sqrt{1 - \nu^2}}\right) - \Phi(\sqrt{\tau}) + \Phi(-\sqrt{\tau}) \right| \\ &= \left| \Phi\left(\frac{\sqrt{\tau} - \nu u}{\sqrt{1 - \nu^2}}\right) + \left(1 - \Phi\left(\frac{-\sqrt{\tau} - \nu u}{\sqrt{1 - \nu^2}}\right)\right) - \Phi(\sqrt{\tau}) - (1 - \Phi(-\sqrt{\tau})) \right| \\ &= \left| \Phi\left(\frac{\sqrt{\tau} + \nu u}{\sqrt{1 - \nu^2}}\right) + \Phi\left(\frac{\sqrt{\tau} - \nu u}{\sqrt{1 - \nu^2}}\right) - 2\Phi\left(\frac{\sqrt{\tau}}{\sqrt{1 - \nu^2}}\right) + 2\Phi\left(\frac{\sqrt{\tau}}{\sqrt{1 - \nu^2}}\right) - 2\Phi(\sqrt{\tau}) \right| \\ &\leq \max_{x \in I_1} |\Phi''(x)| \left(\frac{\nu u}{\sqrt{1 - \nu^2}} \right)^2 + 2 \max_{x \in I_2} |\Phi'(x)| \left(\frac{\sqrt{\tau}}{\sqrt{1 - \nu^2}} - \sqrt{\tau} \right), \end{aligned} \quad (17)$$

where I_1 is the interval with endpoints $\frac{\sqrt{\tau} - \nu u}{\sqrt{1 - \nu^2}}$ and $\frac{\sqrt{\tau} + \nu u}{\sqrt{1 - \nu^2}}$, $I_2 = [\sqrt{\tau}, \frac{\sqrt{\tau}}{\sqrt{1 - \nu^2}}]$, and we have used the triangle inequality and the relation between divided differences and derivatives (see, e.g., [17, (2.1.4.3)]). Since $|\nu| \leq \frac{1}{5}$ and $|u| \leq \sqrt{\tau}$, we have $I_1 \subseteq [\sqrt{\frac{2\tau}{3}}, \sqrt{\frac{3\tau}{2}}]$. Note that $\Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ and $\Phi''(x) = -\frac{x}{\sqrt{2\pi}} e^{-x^2/2}$, and $\frac{1}{\sqrt{1 - \nu^2}} - 1 = \frac{\nu^2}{\sqrt{1 - \nu^2} + 1 - \nu^2}$. Therefore, the upper bound (17) yields

$$\left| \Phi\left(\frac{\sqrt{\tau} - \nu u}{\sqrt{1 - \nu^2}}\right) - \Phi\left(\frac{-\sqrt{\tau} - \nu u}{\sqrt{1 - \nu^2}}\right) - \Phi(\sqrt{\tau}) + \Phi(-\sqrt{\tau}) \right| \leq C' \nu^2 (\tau^{3/2} + \tau^{1/2}) e^{-\tau/3},$$

where $C' > 0$ is an absolute constant. The conclusion follows from (16).

- (iv) Taking the real part of the characteristic function we obtain that, for $X \sim \mathcal{N}(\mu, \sigma^2)$,

$$\mathbf{E}[\cos(tX)] = \cos(t\mu) \exp\left(-\frac{1}{2}t^2\sigma^2\right). \quad (18)$$

Set $\delta = \nu \frac{2\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2}$, $\tilde{\zeta} = \mu_1 + \sigma_1\zeta$, and $\tilde{\eta} = \mu_2 + \sigma_2\eta$. Then, $\tilde{\zeta} + \tilde{\eta} \sim \mathcal{N}(\mu_1 + \mu_2, (\sigma_1^2 + \sigma_2^2)(1 + \delta))$ and $\tilde{\zeta} - \tilde{\eta} \sim \mathcal{N}(\mu_1 - \mu_2, (\sigma_1^2 + \sigma_2^2)(1 - \delta))$. Let $b = \frac{t^2(\sigma_1^2 + \sigma_2^2)}{2}$. We get

$$\begin{aligned} \mathbf{Cov}[\cos(t\tilde{\zeta}), \cos(t\tilde{\eta})] &= \mathbf{E}[\cos(t\tilde{\zeta}) \cos(t\tilde{\eta})] - \mathbf{E}[\cos(t\tilde{\zeta})]\mathbf{E}[\cos(t\tilde{\eta})] \\ &= \frac{1}{2}\mathbf{E}[\cos t(\tilde{\zeta} + \tilde{\eta})] + \frac{1}{2}\mathbf{E}[\cos t(\tilde{\zeta} - \tilde{\eta})] - \mathbf{E}[\cos(t\tilde{\zeta})]\mathbf{E}[\cos(t\tilde{\eta})] \\ &= \frac{1}{2} \cos(t\mu_1) \cos(t\mu_2) \left(e^{-b(1+\delta)} + e^{-b(1-\delta)} - 2e^{-b} \right) \\ &\quad - \frac{1}{2} \sin(t\mu_1) \sin(t\mu_2) \left(e^{-b(1+\delta)} - e^{-b(1-\delta)} \right). \end{aligned}$$

It remains to show that

$$\left| e^{-b(1+\delta)} + e^{-b(1-\delta)} - 2e^{-b} \right| \lesssim v^2, \quad \left| e^{-b(1+\delta)} - e^{-b(1-\delta)} \right| \lesssim |v|.$$

For $|v| \geq \frac{1}{2}$ these upper bounds hold trivially. For $|v| \leq \frac{1}{2}$, using the fact that $|\delta| \leq |v| \leq \frac{1}{2}$ and applying the relation between divided differences and derivatives (see, e.g., [17, (2.1.4.3)]) we obtain

$$\begin{aligned} \left| e^{-b(1+\delta)} + e^{-b(1-\delta)} - 2e^{-b} \right| &\leq \max_{\xi \in [\frac{1}{2}, \frac{3}{2}]} b^2 \delta^2 e^{-b\xi} \lesssim e^{-b/4} \delta^2 \leq v^2, \\ \left| e^{-b(1+\delta)} - e^{-b(1-\delta)} \right| &\leq \max_{\xi \in [\frac{1}{2}, \frac{3}{2}]} 2b|\delta| e^{-b\xi} \lesssim e^{-b/4} |\delta| \leq |v|. \end{aligned}$$

The conclusion follows. $\square$

As a corollary, we obtain the following proposition on the properties of various statistics of the correlated Gaussian random vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_d) \sim \mathcal{N}(0, \boldsymbol{\Sigma})$. Recall that we denote by $\sigma_1^2, \dots, \sigma_d^2$ the diagonal elements of $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_{ij})_{i,j}$ and by $\tilde{\boldsymbol{\Sigma}}$ be the matrix of correlation coefficients with entries $\tilde{\Sigma}_{ij} = \boldsymbol{\Sigma}_{ij} / (\sigma_i \sigma_j)$.

Proposition 2. *Let $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$. There exists an absolute constant $C > 0$ such that the following holds.*

- (i) $\text{Var}(\|\boldsymbol{\varepsilon}\|_2^2) = 2\|\boldsymbol{\Sigma}\|_F^2$.
- (ii) Let $\tau \geq 1$ and $\beta = \mathbf{E}[Z^2 | |Z| > \tau]$, where $Z \sim \mathcal{N}(0, 1)$. Then, for any subset $\mathcal{U} \subseteq [d]$,

$$\begin{aligned} \text{Var} \left[\sum_{i \in \mathcal{U}} (\varepsilon_i^2 - \sigma_i^2 \beta) \mathbb{1}(|\varepsilon_i| > \sigma_i \tau) \right] &\leq C \|\boldsymbol{\Sigma}\|_F^2 \tau^4 \exp(-\tau^2/2), \\ \text{Var} \left[\sum_{i \in \mathcal{U}} \varepsilon_i^2 \mathbb{1}(|\varepsilon_i| > \sigma_i \tau) \right] &\leq C \|\boldsymbol{\Sigma}\|_F^2 \tau^8 \exp(-\tau^2/3). \end{aligned}$$

- (iii) Let $\tau \geq 0$. Then

$$\text{Var} \left[\sum_{i \in \mathcal{U}} \mathbb{1}(\varepsilon_i^2 \leq \sigma_i^2 \tau) \right] \leq C \|\tilde{\boldsymbol{\Sigma}}\|_F^2 (\tau^2 + 1) e^{-\tau/3},$$

- (iv) Let $t \in \mathbb{R}$, and let $\mu_1, \dots, \mu_d \in \mathbb{R}$ be such that $\sum_{i=1}^d \mathbb{1}(\mu_i \neq 0) \leq s$. Then,

$$\text{Var} \left[\sum_{i=1}^d \cos(t(\mu_i + \varepsilon_i)) \right] \leq C \left(\|\tilde{\boldsymbol{\Sigma}}\|_F^2 + s \|\tilde{\boldsymbol{\Sigma}}\|_F \right).$$

Proof. Items (i) and (iii) follow directly from Lemma 2.

(ii) Consider the standard Gaussian random variables $\tilde{\varepsilon}_i = \varepsilon_i / \sigma_i$, $i \in [d]$. The correlation coefficient between $\tilde{\varepsilon}_i$ and $\tilde{\varepsilon}_j$ is $\tilde{\Sigma}_{ij} / (\sigma_i \sigma_j)$ for $i, j \in [d]$. Thus, we have

$$\begin{aligned} &\text{Var} \left[\sum_{i \in \mathcal{U}} (\varepsilon_i^2 - \sigma_i^2 \beta) \mathbb{1}(|\varepsilon_i| > \sigma_i \tau) \right] \\ &= \sum_{i \in \mathcal{U}, j \in \mathcal{U}} \sigma_i^2 \sigma_j^2 \text{Cov}((\tilde{\varepsilon}_i^2 - \beta) \mathbb{1}(|\varepsilon_i| > \sigma_i \tau), (\tilde{\varepsilon}_j^2 - \beta) \mathbb{1}(|\varepsilon_j| > \sigma_j \tau)) \\ &\lesssim \|\boldsymbol{\Sigma}\|_F^2 \tau^4 \exp(-\tau^2/2), \end{aligned} \tag{19}$$

where the last inequality follows from Lemma 2(ii). Next we prove the second inequality in (ii). Note that

$$\begin{aligned} \mathbf{Var} \left[\sum_{i \in \mathcal{U}} \varepsilon_i^2 \mathbb{1}(|\varepsilon_i| > \sigma_i \tau) \right] &\leq 2 \mathbf{Var} \left[\sum_{i \in \mathcal{U}} (\varepsilon_i^2 - \sigma_i^2 \beta) \mathbb{1}(|\varepsilon_i| > \sigma_i \tau) \right] \\ &\quad + 2 \mathbf{Var} \left[\sum_{i \in \mathcal{U}} \sigma_i^2 \beta \mathbb{1}(|\varepsilon_i| > \sigma_i \tau) \right]. \end{aligned}$$

Applying Lemma 2(iii) and the fact $\beta \lesssim \tau^2$ (see, e.g., [3, Lemma 4] or (24) below) yields that

$$\begin{aligned} \mathbf{Var} \left[\sum_{i \in \mathcal{U}} \sigma_i^2 \beta \mathbb{1}(|\varepsilon_i| > \sigma_i \tau) \right] &= \beta^2 \sum_{i \in \mathcal{U}, j \in \mathcal{U}} \sigma_i^2 \sigma_j^2 \mathbf{Cov}(\mathbb{1}(|\tilde{\varepsilon}_i| > \tau), \mathbb{1}(|\tilde{\varepsilon}_j| > \tau)) \\ &\lesssim \|\Sigma\|_F^2 \tau^8 \exp(-\tau^2/3). \end{aligned} \quad (20)$$

The conclusion follows by combining (19) and (20).

(iv) Introducing the notation $\mathcal{S} = \{i \in [d] : \mu_i \neq 0\}$ we get

$$\begin{aligned} \mathbf{Var} \left[\sum_{i=1}^d \cos(t(\mu_i + \varepsilon_i)) \right] &= \sum_{i,j} \mathbf{Cov} [\cos(t(\mu_i + \varepsilon_i)), \cos(t(\mu_j + \varepsilon_j))] \\ &\stackrel{(a)}{\lesssim} \sum_{i,j} \tilde{\Sigma}_{ij}^2 + \sum_{i,j \in \mathcal{S}} |\tilde{\Sigma}_{ij}| \\ &\stackrel{(b)}{\leq} \|\tilde{\Sigma}\|_F^2 + s \|\tilde{\Sigma}\|_F, \end{aligned}$$

where (a) follows from Lemma 2(iv), and (b) is due to the Cauchy-Schwarz inequality. $\square$

6.2. Proofs for Section 2 (known σ)

6.2.1. Proof of Theorem 1

Consider first the case $s > \|\Sigma\|_F$. Then

$$\hat{Q} - \|\theta\|_2^2 = \sum_{i=1}^d Y_i^2 - \sigma^2 \sigma_i^2 - \theta_i^2 = \sigma^2 \sum_{i=1}^d (\varepsilon_i^2 - \sigma_i^2) + 2\sigma \sum_{i \in \mathcal{S}} \theta_i \varepsilon_i,$$

where $\mathcal{S}$ denotes the support of θ . Consequently,

$$\mathbf{E}_{\theta, \Sigma} [(\hat{Q} - \|\theta\|_2^2)^2] \leq 2\sigma^4 \mathbf{Var}(\|\varepsilon\|_2^2) + 8\sigma^2 \theta_S^\top \Sigma_{\mathcal{S}, \mathcal{S}} \theta_S, \quad (21)$$

where $\theta_S \in \mathbb{R}^{\mathcal{S}}$ is the vector of nonzero entries of θ , and $\Sigma_{\mathcal{S}, \mathcal{S}} \in \mathbb{R}^{\mathcal{S} \times \mathcal{S}}$ is the submatrix of Σ with columns and rows in $\mathcal{S}$. Applying Proposition 2(i) we immediately obtain (4) for $s > \|\Sigma\|_F$.

Consider now the case $s \leq \|\Sigma\|_F$. Then we have

$$\begin{aligned}\hat{Q} - \|\theta\|_2^2 &= \sum_{i=1}^d (Y_i^2 - \sigma^2 \sigma_i^2 \beta_s) \mathbb{1}(|Y_i| > \sigma \sigma_i \tau) - \theta_i^2 \\ &= \sum_{i \notin S} \sigma^2 (\varepsilon_i^2 - \sigma_i^2 \beta_s) \mathbb{1}(|\varepsilon_i| > \sigma_i \tau) + \sum_{i \in S} (Y_i^2 - \sigma^2 \sigma_i^2 \beta_s - \theta_i^2) \\ &\quad - \sum_{i \in S} (Y_i^2 - \sigma^2 \sigma_i^2 \beta_s) \mathbb{1}(|Y_i| \leq \sigma \sigma_i \tau) \\ &= \sum_{i=1}^d \sigma^2 (\varepsilon_i^2 - \sigma_i^2 \beta_s) \mathbb{1}(|\varepsilon_i| > \sigma_i \tau) + \sum_{i \in S} 2\sigma \theta_i \varepsilon_i\end{aligned}\quad (22)$$

$$+ \sum_{i \in S} \sigma^2 (\varepsilon_i^2 - \sigma_i^2 \beta_s) \mathbb{1}(|\varepsilon_i| \leq \sigma_i \tau) - \sum_{i \in S} (Y_i^2 - \sigma^2 \sigma_i^2 \beta_s) \mathbb{1}(|Y_i| \leq \sigma \sigma_i \tau). \quad (23)$$

Notice that for any $\tau \geq 0$ we have

$$\beta_s = 1 + \frac{\tau \frac{1}{\sqrt{2\pi}} e^{-\tau^2/2}}{\int_{\tau}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx} < 1 + \frac{\tau(\tau + \sqrt{\tau^2 + 4})}{2} \leq \tau^2 + \tau + 1, \quad (24)$$

where the first inequality uses formula 7.1.13 in [1]. Since $\tau \geq 1$, by the triangle inequality, we get that almost surely, the absolute value of the expression in (23) satisfies:

$$\left| \sum_{i \in S} (\varepsilon_i^2 - \sigma^2 \sigma_i^2 \beta_s) \mathbb{1}(|\varepsilon_i| \leq \sigma_i \tau) - \sum_{i \in S} (Y_i^2 - \sigma^2 \sigma_i^2 \beta_s) \mathbb{1}(|Y_i| \leq \sigma \sigma_i \tau) \right| \lesssim s \sigma^2 \tau^2 \max_{i \in S} \sigma_i^2.$$

Next, the mean squared errors of the two terms in (22) can be controlled by Proposition 2(ii) and an argument similar to (21), respectively. It follows that

$$\begin{aligned}\mathbf{E}_{\theta, \Sigma} \left[\left(\sum_{i=1}^d \sigma^2 (\varepsilon_i^2 - \sigma_i^2 \beta_s) \mathbb{1}(|\varepsilon_i| > \sigma_i \tau) \right)^2 \right] &+ \mathbf{E}_{\theta, \Sigma} \left[\left(\sum_{i \in S} 2\sigma \theta_i \varepsilon_i \right)^2 \right] \\ &\lesssim \sigma^4 \|\Sigma\|_F^2 \tau^4 e^{-\tau^2/2} + \sigma^2 \theta_S^\top \Sigma_{S,S} \theta_S.\end{aligned}$$

Combining the above bounds we find that

$$\mathbf{E}_{\theta, \Sigma} \left[(\hat{Q} - \|\theta\|_2^2)^2 \right] \lesssim \sigma^2 \|\theta\|_2^2 \lambda_{\max}(\Sigma_{S,S}) + \sigma^4 \tau^4 \left(s^2 + \|\Sigma\|_F^2 e^{-\tau^2/2} \right).$$

Recalling that $\tau = 3\sqrt{\log(1 + \|\Sigma\|_F^2/s^2)}$ finishes the proof of (4) for $s \leq \|\Sigma\|_F$.

Inequality (5) follows immediately from (4) using the bounds

$$\lambda_{\max}(\Sigma_{S,S}) \leq \lambda_{\max}(\Sigma) \wedge \|\Sigma_{S,S}\|_F \leq \lambda_{\max}(\Sigma) \wedge s \max_{i \in S} \sigma_i^2 \leq \lambda_{\max}(\Sigma) \wedge s.$$

Finally, we deduce (6) from (5). Notice first that $\bar{\psi}(s, \Sigma) \leq \psi(s, \Sigma)$. Introduce the threshold $t := \sigma \sqrt{\psi(s, \Sigma)}$. Consider first the case $\|\theta\|_2 \leq t$. Then, using the inequality $(a - b)^2 \leq$

$2(a^2 - b^2) + 4b^2$, and (5) we obtain

$$\begin{aligned} \mathbf{E}_{\theta, \Sigma}[(\hat{N} - \|\theta\|_2)^2] &\leq 2\mathbf{E}_{\theta, \Sigma}[|\hat{Q}| - \|\theta\|_2^2] + 4\|\theta\|_2^2 \leq 2\sqrt{\mathbf{E}_{\theta, \Sigma}[(\hat{Q} - \|\theta\|_2^2)^2]} + 4t^2 \\ &\lesssim \sigma^2\psi(s, \Sigma) + \sigma t\sqrt{\bar{\psi}(s, \Sigma)} + t^2 \\ &\lesssim \sigma^2\psi(s, \Sigma). \end{aligned}$$

Next, consider the case $\|\theta\|_2 > t$. Then, using the inequality $(a - b)^2 \leq (a^2 - b^2)^2/a^2$ for $a > 0, b \geq 0$, and (5) we get

$$\begin{aligned} \mathbf{E}_{\theta, \Sigma}[(\hat{N} - \|\theta\|_2)^2] &\leq \frac{\mathbf{E}_{\theta, \Sigma}[(|\hat{Q}| - \|\theta\|_2^2)^2]}{\|\theta\|_2^2} \leq \frac{\mathbf{E}_{\theta, \Sigma}[(\hat{Q} - \|\theta\|_2^2)^2]}{t^2} \\ &\lesssim \sigma^2\bar{\psi}(s, \Sigma) + \frac{\sigma^4\psi^2(s, \Sigma)}{t^2} \\ &\lesssim \sigma^2\psi(s, \Sigma). \end{aligned}$$

The proof of (6) is complete.

6.2.2. Proof of Theorem 2

Given any loss function $w(\cdot)$ with $w(0) = 0$ and $w \not\equiv 0$ and any $\phi > 0$, we define the scaled minimax risk as

$$R^*(s, d, \rho, \phi) := \inf_{\hat{T}} \sup_{(\theta, \Sigma) \in \Theta(s, d, \rho)} \mathbf{E}_{\theta, \Sigma}[w(\sigma^{-1}\phi^{-1/2}|\hat{T} - \|\theta\|_2|)].$$

Since the diagonal elements of Σ are all equal to 1, we only need to consider the case $\rho \geq \sqrt{d}$. The proof is based on a reduction of the minimax lower bound with correlated observations to that with independent observations. In particular, we define

$$\tilde{R}^*(s, d, \phi) := \inf_{\hat{T}} \sup_{\theta \in B_0(s, d), \Sigma = I_{d \times d}} \mathbf{E}_{\theta, \Sigma}[w(\sigma^{-1}\phi^{-1/2}|\hat{T} - \|\theta\|_2|)].$$

It is known from [6] that, for any integers d', s' such that $s' \in [d']$, we have the lower bound

$$\tilde{R}^*(s', d', \phi(s', d')) \gtrsim 1. \quad (25)$$

We will use this bound with suitably chosen d', s' to prove that $R^*(s, d, \rho, \phi(s, \rho^2)) \gtrsim 1$, which is the required result. We consider separately the following two cases.

Case I: $\sqrt{d} \leq \rho \leq 2\sqrt{d}$. Since the identity matrix satisfies the Frobenius norm constraint in the definition of $\Theta(s, d, \rho)$ we have that, for any $w(\cdot)$ and ϕ ,

$$R^*(s, d, \rho, \phi) \geq \tilde{R}^*(s, d, \phi).$$

Set here $\phi = \phi(s, d)$ and note that, in the considered case, $\phi(s, d) \asymp \phi(s, \rho^2)$. It follows immediately from (25) with $s' = s$ and $d' = d$ that $R^*(s, d, \rho, \phi(s, \rho^2)) \gtrsim 1$.

Case II: $\rho > 2\sqrt{d}$. Without loss of generality we assume that $s \geq 2$. Given any two positive integers r and p such that

$$rp \leq d, \quad (26)$$

we define the following block diagonal covariance matrix:

$$\Sigma_0 = \text{diag}(\underbrace{\mathbf{1}_{r \times r}, \dots, \mathbf{1}_{r \times r}}_{p \text{ blocks}}, \mathbf{I}_{(d-rp) \times (d-rp)}). \quad (27)$$

Here $\mathbf{1}_{r \times r}$ is a $r \times r$ dimensional matrix with all entries equal to one, and $\mathbf{I}_{(d-rp) \times (d-rp)}$ is a $(d-rp) \times (d-rp)$ dimensional identity matrix. We also define a class of vectors θ as follows:

$$\tilde{\Theta} := \left\{ (\underbrace{\tilde{\theta}_1, \dots, \tilde{\theta}_1}_{r \text{ entries}}, \dots, \underbrace{\tilde{\theta}_p, \dots, \tilde{\theta}_p}_{r \text{ entries}}, 0, \dots, 0)^\top \in \mathbb{R}^d : \sum_{i=1}^p \mathbb{1}(\tilde{\theta}_i \neq 0) \leq 1 \right\}.$$

If the pair (r, p) is such that

$$r \leq s \quad \text{and} \quad r^2 p + d \leq \rho^2, \quad (28)$$

then for any $\theta \in \tilde{\Theta}$, we have $(\theta, \Sigma_0) \in \Theta(s, d, \rho)$. Moreover, for $\theta \in \tilde{\Theta}$, $\Sigma = \Sigma_0$, model (1) is equivalent to the model with p independent observations $\tilde{Y}_i \sim \mathcal{N}(\tilde{\theta}_i, \sigma^2)$, $i \in [p]$, such that at most one of $\tilde{\theta}_i$'s is nonzero. Since $\|\theta\|_2 = \sqrt{r} \sqrt{\sum_{i=1}^p \tilde{\theta}_i^2} = \sqrt{r} \|\tilde{\theta}\|_2$, where $\tilde{\theta} = (\tilde{\theta}_1, \dots, \tilde{\theta}_p)$, we obtain that, for any $\phi > 0$,

$$\begin{aligned} R^*(s, d, \rho, \phi) &\geq \inf_{\hat{T}} \sup_{\tilde{\theta} \in B_0(1, p), \Sigma = \mathbf{I}_{p \times p}} \mathbf{E}_{\tilde{\theta}, \Sigma} \left[w \left(\frac{|\hat{T} - \sqrt{r} \|\tilde{\theta}\|_2|}{\sigma \phi^{1/2}} \right) \right] \\ &= \inf_{\hat{T}} \sup_{\tilde{\theta} \in B_0(1, p), \Sigma = \mathbf{I}_{p \times p}} \mathbf{E}_{\tilde{\theta}, \Sigma} \left[w \left(\frac{|\hat{T} - \|\tilde{\theta}\|_2|}{\sigma (r^{-1} \phi)^{1/2}} \right) \right] = \tilde{R}^*(1, p, r^{-1} \phi). \end{aligned} \quad (29)$$

Thus, to complete the proof it remains to find a pair (r, p) satisfying the inequalities in (26), (28) and such that $\tilde{R}^*(1, p, r^{-1} \phi(s, \rho^2)) \gtrsim 1$. We will do it separately for two cases.

First, consider the case $s \leq \rho$. Then we choose r and p as the largest integers such that $r \leq s/2$ and $p \leq \frac{\rho^2}{s^2} \wedge \frac{d}{s}$. Recalling that $\rho^2 > 4d$ one can check that such r and p satisfy (26) and (28). Furthermore,

$$\begin{aligned} r \phi(1, p) &\asymp s \log \left(1 + \frac{\rho^2}{s^2} \wedge \frac{d}{s} \right) \asymp s \left(\log \left(1 + \frac{\rho^2}{s^2} \right) \wedge \log \left(1 + \frac{d}{s} \right) \right) \\ &\asymp s \log \left(1 + \frac{\rho^2}{s^2} \right) = \phi(s, \rho^2), \end{aligned} \quad (30)$$

where we have used the facts that $\rho \leq d$, and $\log \left(1 + \frac{\rho^2}{s^2} \right) \asymp \log \left(1 + \frac{\rho}{s} \right)$ for $s \leq \rho$. Combining the last two displays, we obtain that there exists an absolute constant $c > 0$ such that

$$R^*(s, d, \rho, \phi(s, \rho^2)) \geq \tilde{R}^*(1, p, c \phi(1, p)) \gtrsim 1.$$

Next, we consider the case $s > \rho$. Then $\phi(s, \rho^2) = \rho$. We choose the pair (r, p) satisfying (26) and (28) as follows: $r = \lfloor \rho/2 \rfloor$ and $p = 1$. Clearly, $\tilde{R}^*(1, 1, r^{-1} \rho) \gtrsim 1$, which implies that $R^*(s, d, \rho, \phi(s, \rho^2)) \gtrsim 1$ also in this case.

6.3. Proofs for Section 3 (estimation of σ^2)

6.3.1. Analysis of $\hat{\sigma}_S^2$

Proposition 2(iii) allows us to control the variance of the empirical distribution function $\hat{F}_{\varepsilon^2}(x)$ for any fixed $x \geq 0$. In addition, we have the bound (9). Consequently, we obtain that, for a deterministic $t \asymp \sigma^2$, the estimator $\hat{F}_{Y^2}(t)$ converges to $F(t/\sigma^2)$ with the desired rate $\frac{s+\|\tilde{\Sigma}\|_F}{d}$. In our construction, we do not operate with fixed t but rather with $\hat{t}$, which is data-driven and depends on the empirical distribution function, cf. (11). However, since we use a dyadic grid in (11), it is enough to control the estimation error for a finite number of values of t to obtain the desired error guarantee for $\hat{\sigma}_S^2$. We first prove the bound on the error $\hat{\sigma}_S^2(t)$ for fixed t as stated in Lemma 1, then derive the properties of $\hat{t}$ in Lemma 4, and finally obtain the error guarantee for $\hat{\sigma}_S^2$.

Proof of Lemma 1. (i) We prove only the bound on the upper tail probability. The proof of the lower tail bound is entirely analogous. The definition of $\hat{\sigma}_S^2(t)$ and the monotonicity of F imply that

$$\begin{aligned} \mathbf{P}_{\theta, \Sigma} \left[\frac{\hat{\sigma}_S^2(t)}{\sigma^2} \geq 1 + u \right] &= \mathbf{P}_{\theta, \Sigma} \left[\hat{F}_{Y^2}(t) \leq F \left(\frac{t}{(1+u)\sigma^2} \right) \right] \\ &\leq \mathbf{P}_{\theta, \Sigma} \left[F \left(\frac{t}{\sigma^2} \right) - \hat{F}_{\varepsilon^2} \left(\frac{t}{\sigma^2} \right) \geq F \left(\frac{t}{\sigma^2} \right) - F \left(\frac{t}{(1+u)\sigma^2} \right) - \frac{s}{d} \right], \end{aligned} \quad (31)$$

where the last inequality uses (9). Recall that $F(x) = 2\Phi(\sqrt{x}) - 1$ for $x > 0$. Then, for $|u| \leq 1/2$,

$$|F(x) - F(x/(1+u))| \geq \frac{2}{\sqrt{2\pi}} e^{-\frac{x}{2(1+|u|)}} \sqrt{x} \left| 1 - \frac{1}{\sqrt{1+u}} \right| \geq \frac{1}{4} e^{-x} |u| \sqrt{x}. \quad (32)$$

Applying Chebyshev's inequality and Proposition 2(iii) we get

$$\mathbf{P}_{\theta, \Sigma} \left[\frac{\hat{\sigma}_S^2(t)}{\sigma^2} \geq 1 + u \right] \leq \frac{C \frac{\|\tilde{\Sigma}\|_F^2}{d^2} ((\frac{t}{\sigma^2})^2 + 1) e^{-\frac{t}{3\sigma^2}}}{[(\frac{1}{4}|u|e^{-\frac{t}{\sigma^2}}(\frac{t}{\sigma^2})^{1/2} - \frac{s}{d})_+]^2}.$$

Part (i) of the lemma follows from this inequality by noticing that $1 \wedge \frac{x}{[(y-a)_+]^2} \leq 1 \wedge \frac{2(x+a^2)}{y^2}$ for any $x, y, a \geq 0$. Indeed, $1 \wedge \frac{x}{[(y-a)_+]^2} = (1 \wedge \frac{\sqrt{x}}{(y-a)_+})^2$, and $1 \wedge \frac{\sqrt{x}}{(y-a)_+} = \mathbb{1}\{y \leq \sqrt{x} + a\} + \frac{\sqrt{x}}{y-a} \mathbb{1}\{y > \sqrt{x} + a\} \leq 1 \wedge \frac{\sqrt{x}+a}{y}$.

We now prove part (ii) of the lemma. For any t satisfying $|t/\sigma^2 - 1| \leq 2/3$ and $u = 1/2$ we obtain from (32) that

$$|F(t/\sigma^2) - F(2t/3\sigma^2)| \geq \frac{2}{\sqrt{2\pi}} e^{-\frac{s}{3}} \left(\sqrt{\frac{1}{3}} - \sqrt{\frac{2}{9}} \right) \geq 0.015.$$

In view of this remark, part (ii) of the lemma follows from (31) by applying Chebyshev's inequality and Proposition 2(iii). $\square$

Lemma 3. For $\alpha \in (0, 1)$, let $\hat{q}_\alpha$ be the empirical α -th quantile of $\tilde{\varepsilon}_i^2$'s, that is,

$$\hat{q}_\alpha := \min\{u : \hat{F}_{\varepsilon^2}(u) \geq \alpha\}.$$

Then

$$\mathbf{E}_{\theta, \Sigma}(\hat{q}_\alpha^2) \leq 1 + \frac{8}{(1 - \alpha)\sqrt{2\pi}}.$$

Proof. Since $\hat{F}_{\varepsilon^2}(\cdot)$ is the empirical cdf of $\tilde{\varepsilon}_i^2$'s, where $\tilde{\varepsilon}_i$'s are standard normal variables, for any $t > 0$ we have

$$\mathbf{P}_{\theta, \Sigma}[1 - \hat{F}_{\varepsilon^2}(t)] = \mathbf{P}_{\theta, \Sigma}(\tilde{\varepsilon}_1^2 > t) \leq \frac{2e^{-\frac{t}{2}}}{\sqrt{2\pi t}}.$$

By Markov's inequality we obtain that, for any $t > 0$,

$$\mathbf{P}_{\theta, \Sigma}(\hat{q}_\alpha > t) \leq \mathbf{P}_{\theta, \Sigma}(\hat{F}_{\varepsilon^2}(t) \leq \alpha) = \mathbf{P}_{\theta, \Sigma}(1 - \hat{F}_{\varepsilon^2}(t) \geq 1 - \alpha) \leq \frac{2e^{-\frac{t}{2}}}{(1 - \alpha)\sqrt{2\pi t}}.$$

Hence

$$\mathbf{E}_{\theta, \Sigma}[\hat{q}_\alpha^2] \leq 1 + \int_1^\infty \mathbf{P}_{\theta, \Sigma}(\hat{q}_\alpha^2 > t) dt \leq 1 + \int_1^\infty \frac{2e^{-\frac{\sqrt{t}}{2}}}{(1 - \alpha)\sqrt{2\pi}t^{1/4}} dt \leq 1 + \frac{8}{(1 - \alpha)\sqrt{2\pi}}.$$

□

Lemma 4. Let $s/d \leq 0.01$. Then there exists an absolute constant $c > 0$ such that

$$\mathbf{P}_{\theta, \Sigma} \left[\frac{1}{3}\sigma^2 \leq \hat{t} \leq \frac{3}{2}\sigma^2 \right] \geq 1 - \frac{c\|\tilde{\Sigma}\|_F^2}{d^2}. \quad (33)$$

Furthermore, $\mathbf{E}_{\theta, \Sigma}[\hat{t}^2] \lesssim \sigma^4$.

Proof. Since F is the cdf of the χ_1^2 distribution we have $F(\alpha) + \tau < 0.5 < F(\beta) - \tau$ for $\alpha = \frac{1}{3}$, $\beta = \frac{3}{4}$ and $\tau = 0.01$. By Chebyshev's inequality and Proposition 2(iii),

$$\begin{aligned} \mathbf{P}_{\theta, \Sigma}[\hat{F}_{\varepsilon^2}(\beta) - \tau \leq 0.5] &\leq \frac{\mathbf{E}_{\theta, \Sigma}[|\hat{F}_{\varepsilon^2}(\beta) - F(\beta)|^2]}{|F(\beta) - \tau - 0.5|^2} \lesssim \frac{\|\tilde{\Sigma}\|_F^2}{d^2}, \\ \mathbf{P}_{\theta, \Sigma}[\hat{F}_{\varepsilon^2}(\alpha) + \tau \geq 0.5] &\leq \frac{\mathbf{E}_{\theta, \Sigma}[|\hat{F}_{\varepsilon^2}(\alpha) - F(\alpha)|^2]}{|F(\alpha) + \tau - 0.5|^2} \lesssim \frac{\|\tilde{\Sigma}\|_F^2}{d^2}. \end{aligned}$$

By the definition of $\hat{t}$, we have $\hat{F}_{Y^2}(\hat{t}) \geq 0.5$ and $\hat{F}_{Y^2}(\hat{t}/2) < 0.5$. Combining the above remarks and (9) we obtain that, with probability at least $1 - c\|\tilde{\Sigma}\|_F^2/d^2$, where $c > 0$ is an absolute constant,

$$\begin{aligned} \hat{F}_{Y^2}(\hat{t}) &\geq 0.5 > \hat{F}_{\varepsilon^2}(\alpha) + \tau \geq \hat{F}_{Y^2}(\alpha\sigma^2), \\ \hat{F}_{Y^2}(\hat{t}/2) &< 0.5 < \hat{F}_{\varepsilon^2}(\beta) - \tau \leq \hat{F}_{Y^2}(\beta\sigma^2). \end{aligned}$$

The bound (33) follows from these relations and the monotonicity of $\hat{F}_{Y^2}$. Next, from (9) and the assumption $s/d \leq \tau$ we obtain that $\hat{F}_{\varepsilon^2}(\frac{\hat{t}}{2\sigma^2}) \leq \hat{F}_{Y^2}(\hat{t}/2) + \tau < 0.6$. It means that $\frac{\hat{t}}{2\sigma^2} \leq \hat{q}_{0.6}$, where $\hat{q}_{0.6}$ is the 0.6-th empirical quantile of $\tilde{\varepsilon}_i^2$'s. Applying Lemma 3 we conclude that $\mathbf{E}_{\theta, \Sigma}[\hat{t}^2] \lesssim \sigma^4$. □

Consider the set

$$\mathcal{T} := \left\{ t = 2^\ell : \ell \in \mathbb{Z}, \frac{\sigma^2}{3} \leq t \leq \frac{3}{2}\sigma^2 \right\}.$$

Note that this set contains not more than 3 elements. Indeed, if $\mathcal{T}$ is non-empty and ℓ^* is the smallest value in $\mathbb{Z}$ such that $\frac{\sigma^2}{3} \leq 2^{\ell^*} \leq \frac{3}{2}\sigma^2$ then $2^{\ell^*+3} \geq 2^4\sigma^2/3 > \frac{3}{2}\sigma^2$.

Lemma 4 implies that there is an absolute constant $c > 0$ such that $\hat{t} \in \mathcal{T}$ with probability at least $1 - c\|\tilde{\Sigma}\|_F^2/d^2$. Before proving Theorem 4, we first show that there exists an absolute constant $c > 0$ such that $\hat{\sigma}_S^2 \asymp \sigma^2$ with probability at least $1 - c\|\tilde{\Sigma}\|_F^2/d^2$.

Lemma 5. *There exists an absolute constant $c > 0$ such that*

$$\mathbf{P}_{\theta, \Sigma} \left[\frac{1}{2}\sigma^2 \leq \hat{\sigma}_S^2 \leq \frac{3}{2}\sigma^2 \right] \geq 1 - c \frac{\|\tilde{\Sigma}\|_F^2}{d^2}.$$

Proof. Recall that, by definition, $\hat{\sigma}_S^2 = \hat{\sigma}_S^2(\hat{t})$. We have

$$\begin{aligned} \mathbf{P}_{\theta, \Sigma} \left[\left| \frac{\hat{\sigma}_S^2}{\sigma^2} - 1 \right| > \frac{1}{2} \right] &\leq \mathbf{P}_{\theta, \Sigma} \left[\left| \frac{\hat{\sigma}_S^2(\hat{t})}{\sigma^2} - 1 \right| > \frac{1}{2}, \hat{t} \in \mathcal{T} \right] + c \frac{\|\tilde{\Sigma}\|_F^2}{d^2} \\ &\leq \sum_{t \in \mathcal{T}} \mathbf{P}_{\theta, \Sigma} \left[\left| \frac{\hat{\sigma}_S^2(t)}{\sigma^2} - 1 \right| > \frac{1}{2} \right] + c \frac{\|\tilde{\Sigma}\|_F^2}{d^2}. \end{aligned}$$

The conclusion follows from Lemma 1(ii) and the fact that $\mathcal{T}$ contains not more than 3 elements. $\square$

Proof of Theorem 4. Consider the event $E := \{\frac{1}{2}\sigma^2 \leq \hat{\sigma}_S^2 \leq \frac{3}{2}\sigma^2, \hat{t} \in \mathcal{T}\}$. Lemmas 4 and 5 imply that $\mathbf{P}_{\theta, \Sigma}(E) \geq 1 - c \frac{\|\tilde{\Sigma}\|_F^2}{d^2}$, where $c > 0$ is an absolute constant. We have

$$\mathbf{E}_{\theta, \Sigma} \left| \frac{\hat{\sigma}_S^2}{\sigma^2} - 1 \right| = \mathbf{E}_{\theta, \Sigma} \left(\left| \frac{\hat{\sigma}_S^2}{\sigma^2} - 1 \right| \mathbb{1}(E) \right) + \mathbf{E}_{\theta, \Sigma} \left(\left| \frac{\hat{\sigma}_S^2}{\sigma^2} - 1 \right| \mathbb{1}(E^c) \right).$$

By the definition of $\hat{t}$ we have $\hat{F}_{Y^2}(\hat{t}) \geq 0.5$. Thus,

$$\frac{\hat{\sigma}_S^2}{\sigma^2} = \frac{\hat{t}}{\sigma^2 F^{-1}(\hat{F}_{Y^2}(\hat{t}))} \lesssim \frac{\hat{t}}{\sigma^2}. \quad (34)$$

Applying the Cauchy-Schwarz inequality and Lemma 4 yields

$$\mathbf{E}_{\theta, \Sigma} \left[\left| \frac{\hat{\sigma}_S^2}{\sigma^2} - 1 \right| \mathbb{1}(E^c) \right] \lesssim \frac{\|\tilde{\Sigma}\|_F}{d} \sqrt{1 + \mathbf{E}_{\theta, \Sigma} [(\hat{t}/\sigma^2)^2]} \lesssim \frac{\|\tilde{\Sigma}\|_F}{d}. \quad (35)$$

Next, note that

$$|F(\hat{t}/\sigma^2) - F(\hat{t}/\hat{\sigma}_S^2)| \mathbb{1}(E) \geq \min_{\xi \in [\frac{2}{9}, 3]} |F'(\xi)| \left| \frac{\hat{t}}{\sigma^2} - \frac{\hat{t}}{\hat{\sigma}_S^2} \right| \mathbb{1}(E) \gtrsim \left| \frac{\hat{t}}{\sigma^2} - \frac{\hat{t}}{\hat{\sigma}_S^2} \right| \mathbb{1}(E).$$

Using this inequality and (34) we find

$$\begin{aligned}
\mathbf{E}_{\theta, \Sigma} \left[\left| \frac{\hat{\sigma}_S^2}{\sigma^2} - 1 \right| \mathbb{1}(E) \right] &= \mathbf{E}_{\theta, \Sigma} \left[\left| \frac{\hat{\sigma}_S^2}{\hat{t}} \left| \frac{\hat{t}}{\sigma^2} - \frac{\hat{t}}{\hat{\sigma}_S^2} \right| \mathbb{1}(E) \right| \right] \\
&\lesssim \mathbf{E}_{\theta, \Sigma} \left[\left| F(\hat{t}/\sigma^2) - F(\hat{t}/\hat{\sigma}_S^2) \right| \mathbb{1}(E) \right] = \mathbf{E}_{\theta, \Sigma} \left[\left| F(\hat{t}/\sigma^2) - \hat{F}_{Y^2}(\hat{t}) \right| \mathbb{1}(E) \right] \\
&\leq \sum_{t \in \mathcal{T}} \mathbf{E}_{\theta, \Sigma} \left| F(t/\sigma^2) - \hat{F}_{Y^2}(t) \right| \lesssim \frac{s + \|\tilde{\Sigma}\|_F}{d},
\end{aligned} \tag{36}$$

where the last step follows from Proposition 2(iii), (9) and the fact that $\mathcal{T}$ contains at most 3 elements. Combining (35) and (36) concludes the proof. $\square$

6.3.2. Analysis of $\hat{\sigma}_D^2$

In this subsection, we prove Theorem 5. We first evaluate the accuracy of $\hat{\varphi}_d(z)$ considered as an estimator of the normal characteristic function $e^{-\frac{z^2 \sigma^2}{2}}$. The following lemma gives a bound on the relative error of $\hat{\varphi}_d(z)$.

Lemma 6. *For any $z \in \mathbb{R}$ we have*

$$\mathbf{E}_{\theta, \Sigma} \left[\left(\frac{\hat{\varphi}_d(z)}{e^{-\frac{z^2 \sigma^2}{2}}} - 1 \right)^2 \right] \lesssim \frac{s^2}{d^2} + \frac{\|\tilde{\Sigma}\|_F^2}{d^2} \left(\frac{s}{\|\tilde{\Sigma}\|_F} \vee 1 \right) \exp(z^2 \sigma^2). \tag{37}$$

Furthermore, if $s \leq d/8$ there exists an absolute constant $C > 0$ such that, for any $z \in \mathbb{R}$,

$$\mathbf{P}_{\theta, \Sigma} \left[\frac{1}{2} e^{-\frac{z^2 \sigma^2}{2}} \leq \hat{\varphi}_d(z) \leq \frac{3}{2} e^{-\frac{z^2 \sigma^2}{2}} \right] \geq 1 - C \frac{\|\tilde{\Sigma}\|_F^2}{d^2} \left(\frac{s}{\|\tilde{\Sigma}\|_F} \vee 1 \right) \exp(z^2 \sigma^2). \tag{38}$$

Proof. Applying (18) yields

$$\left| \frac{\mathbf{E}_{\theta, \Sigma}[\hat{\varphi}_d(z)]}{e^{-\frac{z^2 \sigma^2}{2}}} - 1 \right| = \left| \frac{1}{d} \sum_{i=1}^d \cos(z\theta_i) - 1 \right| \leq \frac{2s}{d}. \tag{39}$$

Consequently,

$$\mathbf{E}_{\theta, \Sigma} \left[\left(\frac{\hat{\varphi}_d(z)}{e^{-\frac{z^2 \sigma^2}{2}}} - 1 \right)^2 \right] \leq 2 \mathbf{E}_{\theta, \Sigma} \left[\left(\frac{\hat{\varphi}_d(z) - \mathbf{E}_{\theta, \Sigma}[\hat{\varphi}_d(z)]}{e^{-\frac{z^2 \sigma^2}{2}}} \right)^2 \right] + 2 \left(\frac{2s}{d} \right)^2.$$

The bound (37) follows this inequality and Proposition 2(iv).

If $s \leq d/8$ then (39) implies that

$$\left| \mathbf{E}_{\theta, \Sigma}[\hat{\varphi}_d(z)] - e^{-\frac{z^2 \sigma^2}{2}} \right| \leq e^{-\frac{z^2 \sigma^2}{2}} / 4.$$

Under this condition, the left hand side of (38) is not smaller than the probability

$$\mathbf{P}_{\theta, \Sigma} \left[|\hat{\varphi}_d(z) - \mathbf{E}_{\theta, \Sigma}[\hat{\varphi}_d(z)]| \leq e^{-\frac{z^2 \sigma^2}{2}} / 4 \right] \geq 1 - 16 e^{z^2 \sigma^2} \mathbf{Var}[\hat{\varphi}_d(z)].$$

Using this remark and Proposition 2(iv) we obtain (38). $\square$

Proof of Theorem 5. Recall that $\hat{\sigma}_D^2 = \min\{\tilde{\sigma}_D^2, 2\hat{\sigma}_S^2\}$, where

$$\tilde{\sigma}_D^2 = -\frac{2\hat{t}}{\lambda} \log |\hat{\varphi}_d((\lambda/\hat{t})^{1/2})|$$

and $\lambda = \frac{1 \vee \log(s/\|\tilde{\Sigma}\|_F)}{6}$.

Define the event $E = \{\hat{\sigma}_S^2 \geq \frac{1}{2}\sigma^2, \hat{t} \in \mathcal{T}, \frac{1}{2}e^{-\frac{\lambda\sigma^2}{2\hat{t}}} \leq \hat{\varphi}_d((\lambda/\hat{t})^{1/2}) \leq \frac{3}{2}e^{-\frac{\lambda\sigma^2}{2\hat{t}}}\}$. Then,

$$\mathbf{E}_{\theta, \Sigma} \left[\left| \frac{\hat{\sigma}_D^2}{\sigma^2} - 1 \right| \right] = \mathbf{E}_{\theta, \Sigma} \left[\left| \frac{\hat{\sigma}_D^2}{\sigma^2} - 1 \right| \mathbb{1}(E) \right] + \mathbf{E}_{\theta, \Sigma} \left[\left| \frac{\hat{\sigma}_D^2}{\sigma^2} - 1 \right| \mathbb{1}(E^c) \right]. \quad (40)$$

Next, we bound separately the two terms on the right hand side of (40).

On the event E we have $\sigma^2 \leq 2\hat{\sigma}_S^2$, which together with the fact that $\hat{\sigma}_D^2 = \min\{\tilde{\sigma}_D^2, 2\hat{\sigma}_S^2\}$ implies that $|\hat{\sigma}_D^2 - \sigma^2| \mathbb{1}(E) \leq |\tilde{\sigma}_D^2 - \sigma^2| \mathbb{1}(E)$. Using the inequality $|\log x - \log y| \leq \frac{|x-y|}{x \wedge y}$, $\forall x, y > 0$, we obtain:

$$\begin{aligned} \left| \frac{\hat{\sigma}_D^2}{\sigma^2} - 1 \right| \mathbb{1}(E) &\leq \left| \frac{\hat{\sigma}_D^2(\hat{t})}{\sigma^2} - 1 \right| \mathbb{1}(E) = \frac{1}{\sigma^2} \left| -\frac{2\hat{t}}{\lambda} \left(\log |\hat{\varphi}_d((\lambda/\hat{t})^{1/2})| - \log e^{-\frac{\lambda\sigma^2}{2\hat{t}}} \right) \right| \mathbb{1}(E) \\ &\leq \frac{2\hat{t}}{\lambda\sigma^2} \frac{|\hat{\varphi}_d((\lambda/\hat{t})^{1/2}) - e^{-\frac{\lambda\sigma^2}{2\hat{t}}}|}{\hat{\varphi}_d((\lambda/\hat{t})^{1/2}) \wedge e^{-\frac{\lambda\sigma^2}{2\hat{t}}}} \mathbb{1}(E) \\ &\leq \frac{6}{\lambda} \left| \frac{\hat{\varphi}_d((\lambda/\hat{t})^{1/2})}{e^{-\frac{\lambda\sigma^2}{2\hat{t}}}} - 1 \right| \mathbb{1}(E), \end{aligned}$$

where the last inequality follows from the conditions $\hat{t}/\sigma^2 \leq 3/2$ and $\hat{\varphi}_d((\lambda/\hat{t})^{1/2}) \geq e^{-\frac{\lambda\sigma^2}{2\hat{t}}}/2$ that hold on E . Using the fact that $\hat{t} \in \mathcal{T}$ under E , where $\mathcal{T}$ contains at most 2 elements, and applying Lemma 6 we find:

$$\begin{aligned} \mathbf{E}_{\theta, \Sigma} \left[\left| \frac{\hat{\sigma}_D^2}{\sigma^2} - 1 \right| \mathbb{1}(E) \right] &\leq \sum_{t \in \mathcal{T}} \frac{6}{\lambda} \mathbf{E}_{\theta, \Sigma} \left[\left| \frac{\hat{\varphi}_d((\lambda/t)^{1/2})}{e^{-\frac{\lambda\sigma^2}{2t}}} - 1 \right| \right] \\ &\lesssim \frac{1}{\lambda} \max_{t \in \mathcal{T}} \left(\frac{s}{d} + \frac{\|\tilde{\Sigma}\|_F}{d} \exp\left(\frac{\lambda\sigma^2}{2t}\right) \sqrt{\frac{s}{\|\tilde{\Sigma}\|_F} \vee 1} \right) \\ &\leq \frac{s}{\lambda d} + \frac{e^{3\lambda/2} \|\tilde{\Sigma}\|_F}{\lambda d} \sqrt{\frac{s}{\|\tilde{\Sigma}\|_F}} \vee 1. \end{aligned} \quad (41)$$

We now consider the second term on the right hand side of (40). Note first that since $\mathcal{T}$ contains at most 3 elements and $\sigma^2/3 \leq t \leq 3\sigma^2/2$ for $t \in \mathcal{T}$, we get by applying Lemma 6 that

$$\mathbf{P}_{\theta, \Sigma} \left[\frac{1}{2}e^{-\frac{\lambda\sigma^2}{2t}} \leq \hat{\varphi}_d((\lambda/t)^{1/2}) \leq \frac{3}{2}e^{-\frac{\lambda\sigma^2}{2t}}, \forall t \in \mathcal{T} \right] \geq 1 - C \frac{\|\tilde{\Sigma}\|_F^2}{d^2} \left(\frac{s}{\|\tilde{\Sigma}\|_F} \vee 1 \right) \exp(3\lambda).$$

Combining this inequality with Lemmas 4 and 5 we get

$$\mathbf{P}_{\theta, \Sigma}[E^c] \lesssim \frac{\|\tilde{\Sigma}\|_F^2}{d^2} \left(\frac{s}{\|\tilde{\Sigma}\|_F} \vee 1 \right) \exp(3\lambda). \quad (42)$$

Using the fact that $\hat{\sigma}_D^2 \leq 2\hat{\sigma}_S^2$ and the Cauchy-Schwarz inequality we find:

$$\begin{aligned} \mathbf{E}_{\theta, \Sigma} \left[\left| \frac{\hat{\sigma}_D^2}{\sigma^2} - 1 \right| \mathbb{1}(E^c) \right] &\leq \sqrt{\mathbf{E}_{\theta, \Sigma} \left[\left(\frac{\hat{\sigma}_D^2}{\sigma^2} - 1 \right)^2 \right]} \sqrt{\mathbf{P}_{\theta, \Sigma}[E^c]} \\ &\leq \sqrt{2 + 2\mathbf{E}_{\theta, \Sigma} [(2\hat{\sigma}_S^2/\sigma^2)^2]} \sqrt{\mathbf{P}_{\theta, \Sigma}[E^c]}. \end{aligned} \quad (43)$$

From Lemma 4 and (34) we have $\mathbf{E}_{\theta, \Sigma} [(\hat{\sigma}_S^2/\sigma^2)^2] \lesssim 1$. Using this remark and combining (40) (41), (42), and (43) we obtain

$$\mathbf{E}_{\theta, \Sigma} \left[\left| \frac{\hat{\sigma}_D^2}{\sigma^2} - 1 \right| \right] \lesssim \frac{s}{\lambda d} + \left(1 + \frac{1}{\lambda} \right) \frac{e^{3\lambda/2} \|\tilde{\Sigma}\|_F}{d} \left(\frac{s}{\|\tilde{\Sigma}\|_F} \vee 1 \right)^{1/2}. \quad (44)$$

Recall that $\lambda = (1 \vee \log \frac{s}{\|\tilde{\Sigma}\|_F})/6$. If $s \leq e\|\tilde{\Sigma}\|_F$ the right hand side of (44) does not exceed $C\|\tilde{\Sigma}\|_F/d$ for an absolute constant $C > 0$. In the case $s > e\|\tilde{\Sigma}\|_F$ we have $\lambda = (\log \frac{s}{\|\tilde{\Sigma}\|_F})/6$ and the right hand side of (44) is of the order $\frac{s}{\lambda d}$ since

$$\frac{e^{3\lambda/2} \|\tilde{\Sigma}\|_F}{d} \left(\frac{s}{\|\tilde{\Sigma}\|_F} \vee 1 \right)^{1/2} \leq \frac{s}{(s/\|\tilde{\Sigma}\|_F)^{1/4} d} \leq \frac{s}{\lambda d}.$$

This completes the proof. $\square$

6.4. Proofs for Section 4 (unknown σ)

In this section, we prove Theorem 6 using the estimators of the noise level σ^2 constructed in Section 3. Under Assumptions 1 and 2, we have $\Sigma_{ij}/(\sigma_i \sigma_j) \asymp \Sigma_{ij}$ and thus $\|\Sigma\|_F \asymp \|\tilde{\Sigma}\|_F$. Hence, we can replace $\|\tilde{\Sigma}\|_F$ by $\|\Sigma\|_F$ in the theorems of Section 3 and the bounds on estimation errors remain valid to within a constant factor.

6.4.1. Proof of part (i) of Theorem 6

For $s > \|\Sigma\|_F$ we have

$$\begin{aligned} |\hat{N}^* - \|\theta\|_2| &= \left| \sqrt{\sum_{i=1}^d Y_i^2 - \hat{\sigma}_D^2 \sigma_i^2} - \|\theta\|_2 \right| \\ &\leq \left| \sqrt{\|\theta\|_2^2 + 2\sigma \theta^\top \varepsilon} - \|\theta\|_2 \right| + \sqrt{\sum_{i=1}^d \sigma^2 \varepsilon_i^2 - \hat{\sigma}_D^2 \sigma_i^2}, \end{aligned} \quad (45)$$

where the upper bound is due to the inequality $|\sqrt{|x+y|} - z| \leq |\sqrt{|x|} - z| + \sqrt{|y|}$. Due to the inequality $|\sqrt{|1+x|} - 1| \leq |x|$, the first term on the right hand side of (45) admits the bound

$$\left| \sqrt{\|\theta\|_2^2 + 2\sigma \theta^\top \varepsilon} - \|\theta\|_2 \right| \leq \frac{2\sigma |\theta^\top \varepsilon|}{\|\theta\|_2}.$$

Since $\mathbf{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top) = \boldsymbol{\Sigma}$ we obtain

$$\mathbf{E} \left| \sqrt{\|\boldsymbol{\theta}\|_2^2 + 2\sigma\boldsymbol{\theta}^\top\boldsymbol{\varepsilon}} - \|\boldsymbol{\theta}\|_2 \right|^2 \leq 4\sigma^2\lambda_{\max}(\boldsymbol{\Sigma}) \leq 4\sigma^2\|\boldsymbol{\Sigma}\|_F. \quad (46)$$

To handle the second term on the right hand side of (45) we first write

$$\left| \sum_{i=1}^d \left(\varepsilon_i^2 - \frac{\hat{\sigma}_D^2}{\sigma^2} \sigma_i^2 \right) \right| \leq \left| \sum_{i=1}^d (\varepsilon_i^2 - \sigma_i^2) \right| + \left| \sum_{i=1}^d \sigma_i^2 \left(1 - \frac{\hat{\sigma}_D^2}{\sigma^2} \right) \right|.$$

Assumption 1 implies that $\sum_{i=1}^d \sigma_i^2 \leq d$. Therefore, we can use Theorem 5 to get

$$\mathbf{E}_{\boldsymbol{\theta}, \boldsymbol{\Sigma}} \left| \sum_{i=1}^d \sigma_i^2 \left(1 - \frac{\hat{\sigma}_D^2}{\sigma^2} \right) \right| \lesssim \|\boldsymbol{\Sigma}\|_F + \frac{s}{1 \vee \log \frac{s}{\|\boldsymbol{\Sigma}\|_F}}.$$

Furthermore, $\mathbf{E}_{\boldsymbol{\theta}, \boldsymbol{\Sigma}} \left| \sum_{i=1}^d (\varepsilon_i^2 - \sigma_i^2) \right| \leq \sqrt{2}\|\boldsymbol{\Sigma}\|_F$ by Proposition 2 (i). It follows that

$$\mathbf{E}_{\boldsymbol{\theta}, \boldsymbol{\Sigma}} \left| \sum_{i=1}^d \left(\varepsilon_i^2 - \frac{\hat{\sigma}_D^2}{\sigma^2} \sigma_i^2 \right) \right| \lesssim \|\boldsymbol{\Sigma}\|_F + \frac{s}{1 \vee \log \frac{s}{\|\boldsymbol{\Sigma}\|_F}} \lesssim \psi^*(s, \boldsymbol{\Sigma}). \quad (47)$$

Combining (45), (46), and (47) we complete the proof.

6.4.2. Proof of parts (ii) and (iii) of Theorem 6

Using the definition of $\hat{N}^*$ for $s \leq \|\boldsymbol{\Sigma}\|_F$ we obtain

$$\begin{aligned} |\hat{N}^* - \|\boldsymbol{\theta}\|_2| &\leq \left| \sqrt{\sum_{i \in S} Y_i^2 \mathbb{1}(|Y_i| > \hat{\sigma}_S \sigma_i \tau)} - \|\boldsymbol{\theta}\|_2 \right| \\ &\quad + \sigma \left| \sum_{i \notin S} \varepsilon_i^2 \mathbb{1} \left(|\varepsilon_i| > \frac{\hat{\sigma}_S}{\sigma} \sigma_i \tau \right) - \alpha_s \frac{\hat{\sigma}_S^2}{\sigma^2} \sum_{i=1}^d \sigma_i^2 \right|^{1/2}. \end{aligned} \quad (48)$$

Here

$$\begin{aligned} \left| \sqrt{\sum_{i \in S} Y_i^2 \mathbb{1}(|Y_i| > \hat{\sigma}_S \sigma_i \tau)} - \|\boldsymbol{\theta}\|_2 \right| &\leq \sqrt{\sum_{i \in S} (Y_i \mathbb{1}(|Y_i| > \hat{\sigma}_S \sigma_i \tau) - \theta_i)^2} \\ &= \sqrt{\sum_{i \in S} (\sigma \varepsilon_i - Y_i \mathbb{1}(|Y_i| \leq \hat{\sigma}_S \sigma_i \tau))^2}. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbf{E}_{\boldsymbol{\theta}, \boldsymbol{\Sigma}} \left[\left| \sqrt{\sum_{i \in S} Y_i^2 \mathbb{1}(|Y_i| > \hat{\sigma}_S \sigma_i \tau)} - \|\boldsymbol{\theta}\|_2 \right|^2 \right] &\leq 2\mathbf{E}_{\boldsymbol{\theta}, \boldsymbol{\Sigma}} \left(\sum_{i \in S} \sigma^2 \varepsilon_i^2 + \hat{\sigma}_S^2 \sigma_i^2 \tau^2 \right) \\ &= 2 \left(\sigma^2 + \mathbf{E}_{\boldsymbol{\theta}, \boldsymbol{\Sigma}}(\hat{\sigma}_S^2) \tau^2 \right) \sum_{i \in S} \sigma_i^2. \end{aligned}$$

Note that since $\|\tilde{\Sigma}\|_F \leq d$ it follows from Theorem 4 that $\mathbf{E}_{\theta, \Sigma}(\hat{\sigma}_S^2) \lesssim \sigma^2$. Moreover, Assumption 1 implies that $\sum_{i \in S} \sigma_i^2 \leq s$. It follows that

$$\mathbf{E}_{\theta, \Sigma} \left[\left| \sqrt{\sum_{i \in S} Y_i^2 \mathbb{1}(|Y_i| > \hat{\sigma}_S \sigma_i \tau)} - \|\theta\|_2 \right|^2 \right] \lesssim s \sigma^2 \tau^2 \lesssim \sigma^2 \psi^*(s, \Sigma). \quad (49)$$

The remaining proof is devoted to the control of the second term on the right hand side of (48). The major difficulty arises from the dependence between the variance estimator $\hat{\sigma}_S$ and ε_i 's. To begin with, we note that the following holds:

$$\begin{aligned} \mathbf{E}_{\theta, \Sigma} \left| \sum_{i \notin S} \varepsilon_i^2 \mathbb{1}(|\varepsilon_i| > \sigma_i \tau) - \alpha_s \frac{\hat{\sigma}_S^2}{\sigma^2} \sum_{i=1}^d \sigma_i^2 \right| &\leq \mathbf{E}_{\theta, \Sigma} \left| \sum_{i \notin S} \varepsilon_i^2 \mathbb{1}(|\varepsilon_i| > \sigma_i \tau) - \alpha_s \sum_{i=1}^d \sigma_i^2 \right| \\ &\quad + \alpha_s \sum_{i=1}^d \sigma_i^2 \mathbf{E}_{\theta, \Sigma} \left| \frac{\hat{\sigma}_S^2}{\sigma^2} - 1 \right| \\ &\stackrel{(a)}{\leq} \mathbf{E}_{\theta, \Sigma} \left| \sum_{i \notin S} (\varepsilon_i^2 \mathbb{1}(|\varepsilon_i| > \sigma_i \tau) - \alpha_s \sigma_i^2) \right| \\ &\quad + C \alpha_s (s + \|\Sigma\|_F) \\ &\stackrel{(b)}{\lesssim} \|\Sigma\|_F \tau^4 \exp(-\tau^2/6) \lesssim \psi^*(s, \Sigma), \end{aligned} \quad (50)$$

where (a) follows from Assumption 1 and Theorem 4 while (b) is due to the second inequality in Proposition 2(ii), the assumption that $s \leq \|\Sigma\|_F$, the inequality $\alpha_s \lesssim \tau^2 e^{-\tau^2/2}$ and the fact that $\tau = 3\sqrt{\log(1 + \|\Sigma\|_F^2/s^2)}$.

Given (50), in order to evaluate the second term on the right hand side of (48) it remains to control the difference

$$\left| \sum_{i \notin S} \varepsilon_i^2 \mathbb{1} \left(|\varepsilon_i| > \frac{\hat{\sigma}_S}{\sigma} \sigma_i \tau \right) - \sum_{i \notin S} \varepsilon_i^2 \mathbb{1}(|\varepsilon_i| > \sigma_i \tau) \right|, \quad (51)$$

where $\tau = 3\sqrt{\log(1 + \|\Sigma\|_F^2/s^2)}$. To this end, we use the following two lemmas. The first lemma provides a bound for the expression in (51) in probability. The second lemma establishes a bound on its mean squared error under the additional condition that Σ is positive definite. The proofs of the lemmas will be given at the end of this subsection.

Lemma 7. *There exist absolute constants $c > 0$ and $C > 0$ such that, for any $\eta \in (c(s + \|\Sigma\|_F)/d, 1)$,*

$$\begin{aligned} \mathbf{P}_{\theta, \Sigma} \left(\left| \sum_{i \notin S} \varepsilon_i^2 \left(\mathbb{1} \left(|\varepsilon_i| > \frac{\hat{\sigma}_S}{\sigma} \sigma_i \tau \right) - \mathbb{1}(|\varepsilon_i| > \sigma_i \tau) \right) \right| \leq C \frac{s + \|\Sigma\|_F}{\eta^2} \tau^3 e^{-\tau^2/4} \right) \\ \geq 1 - \eta. \end{aligned}$$

Lemma 8. *If Σ is positive definite there exists an absolute constant $c > 0$ such that, for any $\tau \geq 1$,*

$$\begin{aligned} & \mathbf{E}_{\theta, \Sigma} \left[\varepsilon_i^2 \left| \mathbb{1} \left(|\varepsilon_i| > \frac{\hat{\sigma}_S}{\sigma} \sigma_i \tau \right) - \mathbb{1} (|\varepsilon_i| > \sigma_i \tau) \right| \right] \\ & \leq c\tau^2 \left(1 + \sqrt{(\Sigma^{-1})_{ii}} \right) \frac{s + \|\Sigma\|_F}{d} e^{-\frac{\tau^2}{16}} + c\tau^2 \left(\frac{s + \|\Sigma\|_F}{d} \right)^2. \end{aligned}$$

Since $s \leq \|\Sigma\|_F$ and $\tau = 3\sqrt{\log(1 + \|\Sigma\|_F^2/s^2)}$ we obtain that

$$(s + \|\Sigma\|_F)\tau^3 e^{-\tau^2/4} \lesssim s \log(1 + \|\Sigma\|_F^2/s^2) = \psi^*(s, \Sigma).$$

Applying Markov's inequality to the random variable under the expectation in (50) and using Lemma 7 we obtain that for any $\eta > c\|\Sigma\|_F/d$ there exists a constant c_η such that

$$\mathbf{P}_{\theta, \Sigma} \left(\left| \sum_{i \notin S} \varepsilon_i^2 \mathbb{1} \left(|\varepsilon_i| > \frac{\hat{\sigma}_S}{\sigma} \sigma_i \tau \right) - \alpha_s \frac{\hat{\sigma}_S^2}{\sigma^2} \sum_{i=1}^d \sigma_i^2 \right| \leq \frac{c_\eta}{4} \psi^*(s, \Sigma) \right) \geq 1 - \frac{\eta}{2}. \quad (52)$$

Applying now Markov's inequality to the random variable under the expectation in (49) and invoking (48) and (52) we complete the proof of (12).

We now prove (13). Under the condition that Σ is positive definite, we apply Lemma 8 with $s \leq \|\Sigma\|_F$ and $\tau = 3\sqrt{\log(1 + \|\Sigma\|_F^2/s^2)}$. This yields:

$$\begin{aligned} & \mathbf{E}_{\theta, \Sigma} \left| \sum_{i \notin S} \varepsilon_i^2 \left(\mathbb{1} \left(|\varepsilon_i| > \frac{\hat{\sigma}_S}{\sigma} \sigma_i \tau \right) - \mathbb{1} (|\varepsilon_i| > \sigma_i \tau) \right) \right| \\ & \leq \sum_{i \notin S} \mathbf{E}_{\theta, \Sigma} \left(\varepsilon_i^2 \left| \mathbb{1} \left(|\varepsilon_i| > \frac{\hat{\sigma}_S}{\sigma} \sigma_i \tau \right) - \mathbb{1} (|\varepsilon_i| > \sigma_i \tau) \right| \right) \\ & \lesssim \frac{1}{d} \sum_{i=1}^d \left(\sqrt{(\Sigma^{-1})_{ii}} + 1 \right) \|\Sigma\|_F \frac{\log(1 + \|\Sigma\|_F^2/s^2)}{(1 + \|\Sigma\|_F^2/s^2)^{9/16}} + \frac{\|\Sigma\|_F^2}{d} \log(1 + \|\Sigma\|_F^2/s^2) \\ & \lesssim \frac{s}{d} \sum_{i=1}^d \left(\sqrt{(\Sigma^{-1})_{ii}} + 1 \right) + \frac{\|\Sigma\|_F^2}{d} \log(1 + \|\Sigma\|_F^2/s^2). \end{aligned}$$

Notice that $(\Sigma^{-1})_{ii} \leq \lambda_{\max}(\Sigma^{-1})$. Hence,

$$\begin{aligned} & \mathbf{E}_{\theta, \Sigma} \left| \sum_{i \notin S} \varepsilon_i^2 \left(\mathbb{1} \left(|\varepsilon_i| > \frac{\hat{\sigma}_S}{\sigma} \sigma_i \tau \right) - \mathbb{1} (|\varepsilon_i| > \sigma_i \tau) \right) \right| \\ & \lesssim s \left(\sqrt{\lambda_{\max}(\Sigma^{-1})} + 1 \right) + \frac{\|\Sigma\|_F^2}{d} \log(1 + \|\Sigma\|_F^2/s^2). \end{aligned}$$

Combining this bound with (48), (49) and (50) concludes the proof of (13).

Proof of Lemma 7. Given any η such that $\frac{c(s + \|\Sigma\|_F)}{d} < \eta < 1$, define a random event

$$\mathcal{A} := \left\{ \frac{|\hat{\sigma}_S^2 - \sigma^2|}{\sigma^2} \leq \delta := \frac{c(s + \|\Sigma\|_F)}{2\eta d} \right\},$$

where $c > 0$ is a constant to be chosen. Applying the Markov inequality to Theorem 4 yields that by choosing c sufficiently large, $\mathbf{P}_{\theta, \Sigma}(\mathcal{A}^c) \leq \eta/2$. Note that $\delta \leq \frac{1}{2}$. For any $\tau > 0$,

$$\begin{aligned} & \mathbf{E}_{\theta, \Sigma} \left(\varepsilon_i^2 \left| \mathbb{1} \left(|\varepsilon_i| > \frac{\hat{\sigma}_S}{\sigma} \sigma_i \tau \right) - \mathbb{1} (|\varepsilon_i| > \sigma_i \tau) \right| \mathbb{1}(\mathcal{A}) \right) \\ & \leq \mathbf{E}_{\theta, \Sigma} \left(\varepsilon_i^2 \mathbb{1} \left(\tau \sqrt{1 - \delta} \leq \frac{|\varepsilon_i|}{\sigma_i} \leq \tau \sqrt{1 + \delta} \right) \mathbb{1}(\mathcal{A}) \right) \\ & \leq \sigma_i^2 \mathbf{E}_{\theta, \Sigma} \left(\frac{\varepsilon_i^2}{\sigma_i^2} \mathbb{1} \left(\tau \sqrt{1 - \delta} \leq \frac{|\varepsilon_i|}{\sigma_i} \leq \tau \sqrt{1 + \delta} \right) \right) \\ & \lesssim \sigma_i^2 \delta \tau^3 e^{-\tau^2/4}, \end{aligned}$$

where in the last step we have used the fact that, for $b > a > 0$,

$$\int_a^b x^2 e^{-x^2/2} dx \leq b^2 e^{-a^2/2} (b - a).$$

Consequently, since $\sigma_i^2 \leq 1$ by Assumption 1, we have

$$\mathbf{E}_{\theta, \Sigma} \left(\left| \sum_{i \notin S} \varepsilon_i^2 \left(\mathbb{1} \left(|\varepsilon_i| > \frac{\hat{\sigma}_S}{\sigma} \sigma_i \tau \right) - \mathbb{1} (|\varepsilon_i| > \sigma_i \tau) \right) \right| \mathbb{1}(\mathcal{A}) \right) \lesssim \frac{s + \|\Sigma\|_F}{\eta} \tau^3 e^{-\tau^2/4}.$$

Applying the Markov inequality yields that

$$\mathbf{P}_{\theta, \Sigma} \left[\left\{ \left| \sum_{i \notin S} \varepsilon_i^2 \left(\mathbb{1} \left(|\varepsilon_i| > \frac{\hat{\sigma}_S}{\sigma} \sigma_i \tau \right) - \mathbb{1} (|\varepsilon_i| > \sigma_i \tau) \right) \right| \geq \frac{C(s + \|\Sigma\|_F)}{\eta^2} \tau^3 e^{-\tau^2/4} \right\} \cap \mathcal{A} \right] \leq \eta/2,$$

where C is a constant. In light of the above and that $\mathbf{P}_{\theta, \Sigma}(\mathcal{A}^c) \leq \eta/2$, we obtain the desired result. $\square$

The rest of this subsection is devoted to the proof of Lemma 8, where Σ is assumed positive definite. The idea of the proof is to use a leave-one-out argument. For $i \in [d]$, set $\varepsilon_{-i} = (\varepsilon_1, \dots, \varepsilon_{i-1}, \varepsilon_{i+1}, \dots, \varepsilon_d)$ and $\mathbf{Y}_{-i} = (Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_d)$. Since only one observation is dropped, the corresponding variance estimator using $\mathbf{Y}_{-i}$ is close to the original $\hat{\sigma}_S^2$ with high probability. On the other hand, since Σ is positive definite, all its principal submatrices are positive definite, the conditional distribution of ε_i given ε_{-i} is non-degenerate and has the form (see, e.g., [8])

$$\varepsilon_i | \varepsilon_{-i} \sim \mathcal{N}(\mathbf{Cov}(\varepsilon_i, \varepsilon_{-i}) \mathbf{Cov}(\varepsilon_{-i})^{-1} \varepsilon_{-i}, \sigma_i^2 - \mathbf{Cov}(\varepsilon_i, \varepsilon_{-i}) \mathbf{Cov}(\varepsilon_{-i})^{-1} \mathbf{Cov}(\varepsilon_{-i}, \varepsilon_i)), \quad (53)$$

where, by the Schur complement (see, e.g., [11, (0.7.3.1)]),

$$\sigma_i^2 - \mathbf{Cov}(\varepsilon_i, \varepsilon_{-i}) \mathbf{Cov}(\varepsilon_{-i})^{-1} \mathbf{Cov}(\varepsilon_{-i}, \varepsilon_i) = \frac{1}{(\Sigma^{-1})_{ii}} \in [0, \sigma_i^2].$$

We will use these remarks to derive the following preliminary lemma.

Lemma 9. *Let $t \asymp \sigma^2$. Then for any $\tau > 0$,*

$$\mathbf{P}_{\theta, \Sigma} \left[\frac{\hat{\sigma}_S(t)}{\sigma} \wedge 1 \leq \frac{|\varepsilon_i|}{\sigma_i \tau} \leq \frac{\hat{\sigma}_S(t)}{\sigma} \vee 1 \right] \lesssim \sqrt{(\Sigma^{-1})_{ii}} \frac{s + \|\Sigma\|_F}{d} e^{-\frac{\tau^2}{16}} + \left(\frac{s + \|\Sigma\|_F}{d} \right)^2.$$

Proof. Let $\hat{F}_{-i}(t)$ denote the empirical cumulative distribution function of $\{Y_j^2/\sigma_j^2 : j \neq i\}$. Similar to the estimator $\hat{\sigma}_S^2$ in (10), define the corresponding estimator and its relative error as

$$\hat{\sigma}_{-i}^2(t) := \frac{t}{F^{-1}(\hat{F}_{-i}(t))}, \quad \delta_{-i}(t) := \left| \frac{\hat{\sigma}_{-i}^2(t)}{\sigma^2} - 1 \right|,$$

respectively. Let $E_1 := \{\delta_{-i}(t) \leq 1/100\}$. It follows from Lemma 1 that $\mathbf{P}_{\theta, \Sigma}[E_1^c] \lesssim (\frac{s + \|\Sigma\|_F}{d})^2$ since $t \asymp \sigma^2$. We get

$$\begin{aligned} \mathbf{P}_{\theta, \Sigma} \left[\frac{\hat{\sigma}_S(t)}{\sigma} \wedge 1 \leq \frac{|\varepsilon_i|}{\sigma_i \tau} \leq \frac{\hat{\sigma}_S(t)}{\sigma} \vee 1 \right] \\ \lesssim \mathbf{P}_{\theta, \Sigma} \left[\frac{\hat{\sigma}_S(t)}{\sigma} \wedge 1 \leq \frac{|\varepsilon_i|}{\sigma_i \tau} \leq \frac{\hat{\sigma}_S(t)}{\sigma} \vee 1, E \right] + \left(\frac{s + \|\Sigma\|_F}{d} \right)^2. \end{aligned} \quad (54)$$

Additionally, we get under event E_1 that $\hat{F}_{-i}(t)$ is strictly bounded away from zero and from one. Using the fact that $|\hat{F}_{-i}(t) - \hat{F}(t)| \leq \frac{1}{d}$ and the mean value theorem yields that, under the event E_1 ,

$$\left| \frac{\hat{\sigma}_{-i}^2(t)}{\sigma^2} - \frac{\hat{\sigma}_S^2(t)}{\sigma^2} \right| \leq c/d,$$

for an absolute constant $c > 0$. Without loss of generality, we assume that d is sufficiently large such that $c/d \leq 0.01$. Then,

$$\begin{aligned} \mathbf{P}_{\theta, \Sigma} \left[\frac{\hat{\sigma}_S^2(t)}{\sigma^2} \wedge 1 \leq \frac{\varepsilon_i^2}{\sigma_i^2 \tau^2} \leq \frac{\hat{\sigma}_S^2(t)}{\sigma^2} \vee 1, E_1 \right] \\ \leq \mathbf{P}_{\theta, \Sigma} \left[\left(\frac{\hat{\sigma}_{-i}^2(t)}{\sigma^2} - \frac{c}{d} \right) \wedge 1 \leq \frac{\varepsilon_i^2}{\sigma_i^2 \tau^2} \leq \left(\frac{\hat{\sigma}_{-i}^2(t)}{\sigma^2} + \frac{c}{d} \right) \vee 1, E_1 \right] \\ \leq \mathbf{P}_{\theta, \Sigma} \left[1 - \delta_{-i}(t) - \frac{c}{d} \leq \frac{\varepsilon_i^2}{\sigma_i^2 \tau^2} \leq 1 + \delta_{-i}(t) + \frac{c}{d}, E_1 \right]. \end{aligned} \quad (55)$$

Since $\delta_{-i}(t)$ is measurable with respect to ε_{-i} we can use the conditional distribution (53) to obtain an upper bound on the expression in (55). Specifically, the conditional distribution of $\frac{\varepsilon_i}{\sigma_i}$ given ε_{-i} is $\mathcal{N}(\mu_i(\varepsilon_{-i}), \iota_i^2)$, where

$$\iota_i^2 = \frac{1}{\sigma_i^2 (\Sigma^{-1})_{ii}} \in [0, 1], \quad \mu_i(\varepsilon_{-i}) \sim \mathcal{N}(0, 1 - \iota_i^2).$$

Let $l := \tau\sqrt{1 - \delta_{-i}(t) - c/d}$, $u := \tau\sqrt{1 + \delta_{-i}(t) + c/d}$. The length of the interval $[l, u]$ is at most $L = 2(\delta_{-i}(t) + c/d)$. Then we get

$$\mathbf{P}_{\theta, \Sigma} [|\varepsilon_i|/\sigma_i \in [l, u], E_1] \leq \mathbf{E}_{\theta, \Sigma} \left[\mathbb{1}(E_1) \sqrt{\frac{2}{\pi}} \frac{L}{\iota_i} \exp \left(-\frac{(l - |\mu_i(\varepsilon_{-i})|)_+^2}{2\iota_i^2} \right) \right], \quad (56)$$

where the last inequality uses the fact that, for $0 \leq a \leq b$ and $X \sim \mathcal{N}(\mu, \sigma^2)$,

$$\begin{aligned} \mathbf{P}_{\theta, \Sigma} [|X| \in [a, b]] &= \int \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathbb{1}(a \leq |x| \leq b) dx \\ &\leq 2 \frac{b-a}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2} \left(\min_{x: a \leq |x| \leq b} |x - \mu|\right)^2\right) \\ &\leq \sqrt{\frac{2}{\pi}} \frac{b-a}{\sigma} \exp\left(-\frac{(a-|\mu|)_+^2}{2\sigma^2}\right). \end{aligned}$$

It remains to bound from above the right-hand side of (56). Define $E_2 = \{|\mu_i(\boldsymbol{\varepsilon}_{-i})| \leq \tau/2\}$. Then,

$$\mathbf{P}_{\theta, \Sigma}[E_2^c] \leq \exp\left(-\frac{(\tau/2)^2}{2(1-t_i^2)}\right) \leq e^{-\tau^2/8}.$$

Under the events E_1 and E_2 we have $l - |\mu_i(\boldsymbol{\varepsilon}_{-i})| \geq 0.48\tau$. Therefore,

$$\begin{aligned} &\mathbf{E}_{\theta, \Sigma} \left[\mathbb{1}(E_1) L \exp\left(-\frac{(l - |\mu_i(\boldsymbol{\varepsilon}_{-i})|)_+^2}{2t_i^2}\right) \right] \\ &\leq \mathbf{E}_{\theta, \Sigma} \left[\mathbb{1}(E_1) \mathbb{1}(E_2) L \exp\left(-\frac{(l - |\mu_i(\boldsymbol{\varepsilon}_{-i})|)_+^2}{2}\right) \right] + \mathbf{E}_{\theta, \Sigma} [\mathbb{1}(E_1) \mathbb{1}(E_2^c) L] \\ &\leq \mathbf{E}_{\theta, \Sigma} [L \mathbb{1}(E_1)] \exp(-\tau^2/9) + \sqrt{\mathbf{P}_{\theta, \Sigma}[E_2^c] \mathbf{E}_{\theta, \Sigma} [L^2 \mathbb{1}(E_1)]}. \end{aligned}$$

Recall that $L \leq \delta_{-i}(t) + \frac{c}{d}$. By an argument similar to (36) we obtain that

$$\mathbf{E}_{\theta, \Sigma} [L^2 \mathbb{1}(E_1)] \lesssim \left(\frac{s + \|\Sigma\|_F}{d} \right)^2.$$

Using the above inequalities in (56) we get the following upper bound on the probability in (55):

$$\mathbf{P}_{\theta, \Sigma} \left[\frac{\hat{\sigma}_S^2(t)}{\sigma^2} \wedge 1 \leq \frac{\varepsilon_i^2}{\sigma_i^2 \tau^2} \leq \frac{\hat{\sigma}_S^2(t)}{\sigma^2} \vee 1, E_1 \right] \lesssim \sigma_i \sqrt{(\Sigma^{-1})_{ii}} \frac{s + \|\Sigma\|_F}{d} \exp\left(-\frac{\tau^2}{16}\right).$$

The conclusion follows from this inequality and (54). $\square$

Equipped with Lemma 9, we are now in a position to prove Lemma 8.

Proof of Lemma 8. Define the following two events:

$$E_1 = \{\hat{t} \in \mathcal{T}\}, \quad E_2 = \left\{ \hat{\sigma}_S^2 \leq \frac{3}{2} \sigma^2 \right\},$$

where $\mathbf{P}_{\theta, \Sigma}[E_1^c] \lesssim \|\Sigma\|_F^2/d^2$ and $\mathbf{P}_{\theta, \Sigma}[E_2^c] \lesssim \|\Sigma\|_F^2/d^2$ according to Lemmas 4 and 5,

respectively. Set $E = E_1 \cap E_2$. We have

$$\begin{aligned} \mathbf{E}_{\theta, \Sigma} \left(\varepsilon_i^2 \left| \mathbb{1} \left(|\varepsilon_i| > \frac{\hat{\sigma}_S}{\sigma} \sigma_i \tau \right) - \mathbb{1}(|\varepsilon_i| > \sigma_i \tau) \right| \right) &\leq \mathbf{E}_{\theta, \Sigma} \left(\varepsilon_i^2 \mathbb{1} \left(\frac{\hat{\sigma}_S}{\sigma} \wedge 1 \leq \frac{|\varepsilon_i|}{\sigma_i \tau} \leq \frac{\hat{\sigma}_S}{\sigma} \vee 1 \right) \right) \\ &\leq \frac{3}{2} \sigma_i^2 \tau^2 \mathbf{E}_{\theta, \Sigma} \left(\mathbb{1} \left(\frac{\hat{\sigma}_S}{\sigma} \wedge 1 \leq \frac{|\varepsilon_i|}{\sigma_i \tau} \leq \frac{\hat{\sigma}_S}{\sigma} \vee 1 \right) \mathbb{1}(E) \right) \\ &\quad + \mathbf{E}_{\theta, \Sigma} \left(\varepsilon_i^2 \mathbb{1} \left(\frac{\hat{\sigma}_S}{\sigma} \wedge 1 \leq \frac{|\varepsilon_i|}{\sigma_i \tau} \leq \frac{\hat{\sigma}_S}{\sigma} \vee 1 \right) \mathbb{1}(E^c) \right). \end{aligned} \quad (57)$$

Applying Lemma 9 we get

$$\begin{aligned} &\sigma_i^2 \tau^2 \mathbf{E}_{\theta, \Sigma} \left(\mathbb{1} \left(\frac{\hat{\sigma}_S}{\sigma} \wedge 1 \leq \frac{|\varepsilon_i|}{\sigma_i \tau} \leq \frac{\hat{\sigma}_S}{\sigma} \vee 1 \right) \mathbb{1}(E) \right) \\ &\leq \sigma_i^2 \tau^2 \sum_{t \in \mathcal{T}} \mathbf{P}_{\theta, \Sigma} \left[\frac{\hat{\sigma}_S}{\sigma} \wedge 1 \leq \frac{|\varepsilon_i|}{\sigma_i \tau} \leq \frac{\hat{\sigma}_S}{\sigma} \vee 1 \right] \\ &\lesssim \sigma_i^2 \tau^2 \left(\sqrt{(\Sigma^{-1})_{ii}} \frac{s + \|\Sigma\|_F}{d} e^{-\frac{\tau^2}{16}} + \left(\frac{s + \|\Sigma\|_F}{d} \right)^2 \right). \end{aligned} \quad (58)$$

Next we bound from above the second term in (57), which can be further decomposed into two terms using the fact that $E^c = (E_2 \cap E_1^c) \cup E_2^c$. For the term corresponding to the event $E_2 \cap E_1^c$ we have

$$\mathbf{E}_{\theta, \Sigma} \left(\varepsilon_i^2 \mathbb{1} \left(\frac{\hat{\sigma}_S}{\sigma} \wedge 1 \leq \frac{|\varepsilon_i|}{\sigma_i \tau} \leq \frac{\hat{\sigma}_S}{\sigma} \vee 1 \right) \mathbb{1}(E_1^c \cap E_2) \right) \leq \frac{3}{2} \sigma_i^2 \tau^2 \mathbf{P}_{\theta, \Sigma}[E_1^c] \lesssim \sigma_i^2 \tau^2 \frac{\|\Sigma\|_F^2}{d^2}. \quad (59)$$

Next, for the term corresponding to the event E_2^c we have

$$\begin{aligned} \mathbf{E}_{\theta, \Sigma} \left(\varepsilon_i^2 \mathbb{1} \left(\frac{\hat{\sigma}_S}{\sigma} \wedge 1 \leq \frac{|\varepsilon_i|}{\sigma_i \tau} \leq \frac{\hat{\sigma}_S}{\sigma} \vee 1 \right) \mathbb{1}(E_2^c) \right) &\leq \mathbf{E}_{\theta, \Sigma} \left(\varepsilon_i^2 \mathbb{1}(|\varepsilon_i| \geq \sigma_i \tau) \mathbb{1}(E_2^c) \right) \\ &\leq \sqrt{\mathbf{E}(\varepsilon_i^4 \mathbb{1}(|\varepsilon_i| \geq \sigma_i \tau))} \frac{\|\Sigma\|_F}{d} \lesssim \sigma_i^2 \sqrt{(\tau^3 \vee 1) e^{-\tau^2/2}} \frac{\|\Sigma\|_F}{d}, \end{aligned} \quad (60)$$

where we have used first the Cauchy-Schwarz inequality and then the fact that

$$\int_{\tau}^{\infty} x^4 e^{-x^2/2} dx = e^{-\tau^2/2} (\tau^3 + 3\tau) + 3 \int_{\tau}^{\infty} e^{-x^2/2} dx.$$

Plugging (58), (59), and (60) into (57) we complete the proof. $\square$

6.4.3. Proof of Theorem 7

We will use the following lemma about the deviations of $\hat{\sigma}_{\eta}^2$, the proof of which will be given at the end of this subsection.

Lemma 10. *Let $\eta \in (0, 1)$. If $s \leq d/100$ then*

$$\mathbf{P}_{\theta, \Sigma}(\hat{\sigma}_{\eta}^2 / \sigma^2 \in (q_{\eta/20} / q_{1-\eta/20}, 1)) \geq 1 - \eta/4.$$

On the event defined in Lemma 10, we have that, almost surely, $|\max\{\hat{\sigma}_S, \hat{\sigma}_\eta\} - \sigma| \leq |\hat{\sigma}_S - \sigma|$. Using this and following the same argument as in the proof of Theorem 6, we only need to show that there exists some constant $c_\eta > 0$ depending on $\eta \in (0, 1)$ such that

$$\mathbf{P}_{\theta, \Sigma} \left(\left| \sum_{i \notin S} \varepsilon_i^2 \mathbb{1} \left(|\varepsilon_i| > \frac{\max\{\hat{\sigma}_S, \hat{\sigma}_\eta\}}{\sigma} \sigma_i \tau_\eta \right) - \sum_{i \in S} \varepsilon_i^2 \mathbb{1} (|\varepsilon_i| > \sigma_i \tau_\eta) \right| \geq c_\eta \psi^*(s, \Sigma) \right) \geq 1 - \eta/2. \quad (61)$$

To achieve this goal, we consider the event

$$\mathcal{A}(\eta) := \left\{ \frac{|\hat{\sigma}_S^2 - \sigma^2|}{\sigma^2} \leq \delta := c \frac{\|\Sigma\|_F}{\eta d}, \hat{\sigma}_\eta^2 / \sigma^2 \in (q_{\eta/20} / q_{1-\eta/20}, 1) \right\}.$$

Using Lemma 10 and Theorem 4 we obtain that, for a constant $c > 0$ large enough, $\mathbf{P}_{\theta, \Sigma}(\mathcal{A}(\eta)) \geq 1 - \frac{3\eta}{8}$. Arguing as in the proof of Lemma 7 we get

$$\begin{aligned} & \mathbf{E}_{\theta, \Sigma} \left(\varepsilon_i^2 \left| \mathbb{1} \left(|\varepsilon_i| > \frac{\max\{\hat{\sigma}_S, \hat{\sigma}_\eta\}}{\sigma} \sigma_i \tau_\eta \right) - \mathbb{1} (|\varepsilon_i| > \sigma_i \tau_\eta) \right| \mathbb{1}(\mathcal{A}(\eta)) \right) \\ & \leq \mathbf{E}_{\theta, \Sigma} \left(\varepsilon_i^2 \mathbb{1} \left(\max \left\{ \tau_\eta \sqrt{1 - \min(1, \delta)}, \tau_\eta \sqrt{\frac{q_{\eta/20}}{q_{1-\eta/20}}} \right\} \leq \frac{|\varepsilon_i|}{\sigma_i} \leq \tau_\eta \sqrt{1 + \delta} \right) \mathbb{1}(\mathcal{A}(\eta)) \right) \\ & \leq \mathbf{E}_{\theta, \Sigma} \left(\varepsilon_i^2 \mathbb{1} \left(\max \left\{ \tau_\eta \sqrt{1 - \min(1, \delta)}, \tau \right\} \leq \frac{|\varepsilon_i|}{\sigma_i} \leq \tau_\eta \sqrt{1 + \delta} \right) \mathbb{1}(\mathcal{A}(\eta)) \right) \\ & \lesssim \sigma_i^2 \delta (1 + \delta) \tau_\eta^3 e^{-\tau^2/2}. \end{aligned}$$

Recalling that $\tau = 3\sqrt{\log(1 + \|\Sigma\|_F^2/s^2)}$ and acting in the same way as in Lemma 7 we obtain (61) for any $\eta \in (0, 1)$.

Proof of Lemma 10. Without loss of generality, we assume that all σ_i 's are equal to 1. Then we have

$$\begin{aligned} \mathbf{P}_{\theta, \Sigma}(\hat{M} \geq \sigma^2 q_{1-\eta/20}) & \leq \mathbf{P}_{\theta, \Sigma} \left(\hat{F}_{Y^2} \left(\sigma^2 q_{1-\eta/20} \right) \leq \frac{1}{2} \right) \leq \mathbf{P}_{\theta, \Sigma} \left(\hat{F}_{\varepsilon^2} (q_{1-\eta/20}) \leq 0.51 \right) \\ & = \mathbf{P}_{\theta, \Sigma} \left(1 - \hat{F}_{\varepsilon^2} (q_{1-\eta/20}) \geq 0.49 \right) \leq \eta/8, \end{aligned}$$

where the last bound uses Markov's inequality. Similarly, we have

$$\mathbf{P}_{\theta, \Sigma}(\hat{M} \leq \sigma^2 q_{\eta/20}) \leq \mathbf{P}_{\theta, \Sigma} \left(\hat{F}_{Y^2} \left(\sigma^2 q_{\eta/20} \right) \geq \frac{1}{2} \right) \leq \mathbf{P}_{\theta, \Sigma} \left(\hat{F}_{\varepsilon^2} (q_{\eta/20}) \geq 0.49 \right) \leq \eta/8.$$

The desired result follows from the above inequalities and the union bound. $\square$

6.4.4. Proof of Theorem 8

We only prove the lower bound of Theorem 8. The proof of the upper bounds (items (i) and (ii) of Theorem 8) follows from arguments quite analogous to the proofs of Theorems 6 and 7. It is therefore omitted.

Notice that for $s \leq \rho$ the rates stated in the lower bounds of Theorems 2 and 8 are the same while the result of Theorem 2 is stronger since it holds for a smaller minimax risk. This proves the lower bound of Theorem 8 for $s \leq \rho$.

It remains only to prove the lower bound for the case $s > \rho$. Similarly to the proof of Theorem 2, we proceed by a reduction to the lower bound for the model with independent observations. Without causing confusion, we will use the same notation as in the proof of Theorem 2 for somewhat different but analogous quantities. Given a fixed loss function $w(\cdot)$ satisfying the assumptions of Theorem 8 and any $\phi > 0$, we now define $R^*(s, d, \rho, \phi)$ as the minimax risk

$$R^*(s, d, \rho, \phi) := \inf_{\hat{T}} \sup_{(\theta, \Sigma) \in \Theta(s, d, \rho)} \sup_{\sigma > 0} \mathbf{E}_{\theta, \Sigma} \left[w \left(\sigma^{-1} \phi^{-1/2} |\hat{T} - \|\theta\|_2| \right) \right].$$

With this notation, our aim is to prove that $R^*(s, d, \rho, \phi^*(s, \rho^2)) \gtrsim 1$ for $s > \rho$. Moreover, arguing analogously to the proof of Theorem 2 we see that it suffices to obtain this result for $\rho \geq 2\sqrt{d}$. We also assume without loss of generality that $s \geq 2$.

For a positive integer r , we define $p = \lfloor d/r \rfloor$. Let Σ_0^* be the block diagonal covariance matrix defined as in (27) with the only difference that we replace $\mathbf{I}_{(d-rp) \times (d-rp)}$ by $\mathbf{1}_{(d-rp) \times (d-rp)}$, that is, by the $(d-rp) \times (d-rp)$ dimensional matrix with all entries equal to 1. For a positive integer $k \leq p$, we define a class of vectors $\theta \in \mathbb{R}^d$ as follows:

$$\bar{\Theta} := \left\{ \underbrace{(\tilde{\theta}_1, \dots, \tilde{\theta}_1)}_{r \text{ entries}}, \dots, \underbrace{(\tilde{\theta}_p, \dots, \tilde{\theta}_p)}_{r \text{ entries}}, 0, \dots, 0 \right\}^\top \in \mathbb{R}^d : \sum_{i=1}^p \mathbb{1}(\tilde{\theta}_i \neq 0) \leq k \Bigg\}.$$

By the choice of p we have $d - rp < r$. If the triplet (r, p, k) is such that

$$k \leq p, \quad rk \leq s, \quad \text{and} \quad r^2(p+1) \leq \rho^2, \quad (62)$$

then for any $\theta \in \bar{\Theta}$, we have $(\theta, \Sigma_0^*) \in \Theta(s, d, \rho)$. Moreover, for $\theta \in \bar{\Theta}$, $\Sigma = \Sigma_0^*$, model (1) is equivalent to the model with $p+1$ independent observations $(\tilde{Y}_1, \dots, \tilde{Y}_{p+1})$ such that $\tilde{Y}_i \sim \mathcal{N}(\tilde{\theta}_i, \sigma^2)$, $i \in [p]$, where at most k of $\tilde{\theta}_i$'s are nonzero, and $\tilde{Y}_{p+1} \sim \mathcal{N}(0, \sigma^2)$. We denote by $\mathbf{E}_{\tilde{\theta}, \sigma}^*$, where $\tilde{\theta} = (\tilde{\theta}_1, \dots, \tilde{\theta}_p)$, the expectation with respect to the joint distribution of $(\tilde{Y}_1, \dots, \tilde{Y}_p, \tilde{Y}_{p+1})$ and we define, for any $\phi > 0$,

$$\tilde{R}^*(k, p, \phi) := \inf_{\hat{T}} \sup_{\tilde{\theta} \in B_0(k, p)} \sup_{\sigma > 0} \mathbf{E}_{\tilde{\theta}, \sigma}^* \left[w \left(\sigma^{-1} \phi^{-1/2} |\hat{T} - \|\tilde{\theta}\|_2| \right) \right]. \quad (63)$$

Since $\|\theta\|_2 = \sqrt{r} \sqrt{\sum_{i=1}^p \tilde{\theta}_i^2}$ we obtain following the same lines as in (29) that, for any $\phi > 0$,

$$R^*(s, d, \rho, \phi) \geq \tilde{R}^*(k, p, r^{-1}\phi). \quad (64)$$

In what follows, we set $r = \lfloor \frac{\rho^2}{2d} \rfloor$ and $k = \lfloor \frac{s}{r} \rfloor$. Then the triplet $(r, p = \lfloor d/r \rfloor, k)$ satisfies the inequalities in (62). Since $\rho \geq 2\sqrt{d}$ and $s > \rho > r$ we have that r and k are positive integers, so the above triplet is a feasible choice. Moreover, with such a choice, we have that $r \leq \frac{\rho^2}{2d} < \frac{s}{2} \leq \frac{d}{2}$ as $\rho < s \leq d$. Therefore,

$$\begin{aligned} k &\leq \frac{s}{r} \leq \frac{4sd}{\rho^2}, \\ k &\geq \frac{s}{r} - 1 \implies rk \geq s - r \geq s/2, \\ p &\geq \frac{d}{r} - 1 \implies rp \geq d - r \geq \frac{d}{2} \implies p \geq \frac{d}{2r} \geq \frac{d^2}{\rho^2}. \end{aligned}$$

Next, recall that $s \leq \alpha d$ for $\alpha > 0$ small enough by the assumption of the theorem. Thus, in the following we assume without loss of generality that $\alpha \leq 1/2$, so that $\rho \leq d/2$. It follows that

$$\begin{aligned} \frac{k^2}{p} &\geq \frac{(s/r - 1)^2}{(d/r)} = \frac{(s - r)^2}{dr} \\ &\geq \frac{(2s\rho - \rho^2)^2}{2d^2\rho^2} \geq \frac{(2d - \rho)^2}{2d^2} \geq 1. \end{aligned}$$

Consequently,

$$r\phi^*(k, p) = \frac{rk}{1 \vee \log(k/\sqrt{p})} \gtrsim \frac{s}{1 \vee \log(s/\rho)} = \phi^*(s, \rho^2).$$

Note also that, by choosing α small enough the condition $p \geq \bar{C}$ of Lemma 11 below is met. Indeed, $p \geq d/(2r) \geq d/s \geq 1/\alpha$. Using these remarks, (64) and Lemma 11 we obtain that, for some absolute constant $c > 0$,

$$R^*(s, d, \rho, \phi^*(s, \rho^2)) \geq R^*(s, d, \rho, cr\phi^*(k, p)) \geq \tilde{R}^*(k, p, c\phi^*(k, p)) \gtrsim 1,$$

This concludes the proof.

Lemma 11. *Let $\tilde{R}^*(k, p, \phi)$ be defined in (63), where w is a loss function satisfying the assumptions of Theorem 8. Then there exists an absolute constant $\bar{C} > 0$ such that for any $p \geq \bar{C}$, any $k \geq \sqrt{p}$, and any absolute constant $c > 0$ we have $\tilde{R}^*(k, p, c\phi^*(k, p)) \gtrsim 1$.*

Proof. It suffices to prove the lemma for $c = 1$ since the function $w(\cdot/c^{1/2})$ satisfies the assumptions of Theorem 8. The proof follows the argument in [7, Proposition 3(ii)], with the only difference is that, in our notation, [7, Proposition 3(ii)] proves the result when the sample is $(\tilde{Y}_1, \dots, \tilde{Y}_p)$, while in our case the sample contains the additional observation $\tilde{Y}_{p+1}$. Therefore, we only focus on the way to handle this difference and do not reproduce the other details of the proof.

Given a p -dimensional probability distribution $\mathbf{P}$ and $v > 0$, we denote by $\mathbf{P} \otimes \mathcal{N}(0, v)$ the distribution of a $(p + 1)$ -dimensional random vector $(X_1, \dots, X_{p+1})$ such that the first p coordinates $(X_1, \dots, X_p)$ are distributed as $\mathbf{P}$ and independent from the last coordinate $X_{p+1} \sim \mathcal{N}(0, v)$.

Let the mixture probability measures $\tilde{\mathbf{P}}_1, \tilde{\mathbf{P}}_2$ and the constant $\tilde{\varphi} > 0$ be defined in the same way as $\mathbf{P}_1, \mathbf{P}_2$ and φ in the proof of [7, Proposition 3(ii)] with the only difference that s and d are replaced by k and p , respectively. We now follow the same argument as in the proof of [7, Proposition 3(ii)] but considering the distributions $\tilde{\mathbf{P}}_1 \otimes \mathcal{N}(0, 1)$ and $\tilde{\mathbf{P}}_2 \otimes \mathcal{N}(0, 1 + \tilde{\varphi})$ instead of $\mathbf{P}_1$ and $\mathbf{P}_2$, respectively. It follows that, in order to prove the lemma, it suffices to show that

$$V(\tilde{\mathbf{P}}_1 \otimes \mathcal{N}(0, 1), \tilde{\mathbf{P}}_2 \otimes \mathcal{N}(0, 1 + \tilde{\varphi})) \leq \frac{5}{8}, \quad (65)$$

where $V(\cdot, \cdot)$ denotes the total variation distance. All other elements of the proof remain the same as in [7, Proposition 3(ii)]. Standard properties of the total variation distance yield that

$$\begin{aligned} &V(\tilde{\mathbf{P}}_1 \otimes \mathcal{N}(0, 1), \tilde{\mathbf{P}}_2 \otimes \mathcal{N}(0, 1 + \tilde{\varphi})) \\ &\leq V(\tilde{\mathbf{P}}_1 \otimes \mathcal{N}(0, 1), \tilde{\mathbf{P}}_2 \otimes \mathcal{N}(0, 1)) + V(\tilde{\mathbf{P}}_2 \otimes \mathcal{N}(0, 1), \tilde{\mathbf{P}}_2 \otimes \mathcal{N}(0, 1 + \tilde{\varphi})) \\ &= V(\tilde{\mathbf{P}}_1, \tilde{\mathbf{P}}_2) + V(\mathcal{N}(0, 1), \mathcal{N}(0, 1 + \tilde{\varphi})). \end{aligned}$$

From the proof of [7, Proposition 3(ii)] we have $V(\mathbf{P}_1, \mathbf{P}_2) \leq 1/2$, so that $V(\tilde{\mathbf{P}}_1, \tilde{\mathbf{P}}_2) \leq 1/2$. Moreover, $V(\mathcal{N}(0, 1), \mathcal{N}(0, 1 + \tilde{\varphi})) \leq c' \tilde{\varphi}$ for some absolute constant $c' > 0$. Recalling the definition of φ in the proof of [7, Proposition 3(ii)] we have that $\tilde{\varphi} \leq \frac{c_0 k}{2p \log(ek^2/p)} \leq c_0/2$, where the constant $c_0 > 0$ can be chosen as small as needed. By choosing $c_0 \leq \frac{1}{4c'}$ we obtain (65). $\square$

6.5. Proof of Proposition 1

As in the proof of Theorem 2 we consider two cases. In the first case corresponding to $\sqrt{d} \leq \rho \leq 2\sqrt{d}$ we have $\phi(s, \rho^2) \leq 2\phi(s, d)$. Then the result follows by reduction to the model with $\Sigma = \mathbf{I}$ by noticing that

$$\mathcal{R}^*(s, d, \rho, \bar{r}_\gamma) \geq \inf_{\Psi} \left[\mathbf{P}_{\mathbf{0}, \mathbf{I}}(\Psi = 1) + \sup_{\theta \in B_0(s, d): \|\theta\|_2 \geq 4\gamma\sigma\sqrt{\phi(s, d)}} \mathbf{P}_{\theta, \mathbf{I}}(\Psi = 0) \right] \geq c(\gamma) \quad (66)$$

for all $\gamma > 0$, where $c(\gamma) > 0$ depends only on γ . This is done following the same lines as in the proof of [6, Theorem 12]. Note that [6, Theorem 12] obtains (66) for sufficiently small γ , while we need it here for all $\gamma > 0$. Extending the proof in [6] to all $\gamma > 0$ is straightforward by using the inequality $V(P, Q) \leq 1 - \exp(-\chi^2(P, Q))/2$ instead of $V(P, Q) \leq \sqrt{\chi^2(P, Q)}$, where $V(P, Q)$ and $\chi^2(P, Q)$ are the total variation distance and the chi-square divergence between two probability measures P and Q (cf. [18, Theorem 2.2(iii)]). Following the lines of [6, Theorem 12] with this modification we obtain that the bound (66) holds with

$$c(\gamma) = \exp(1 - \exp(16\gamma^2))/2. \quad (67)$$

We now focus on the second case $\rho > 2\sqrt{d}$. We consider only the sparse regime $s \leq \rho$ since the result in the regime $s > \rho$ follows from an analogous argument. Repeating the same steps as in the proof of Theorem 2 we find:

$$\begin{aligned} \mathcal{R}^*(s, d, \rho, \bar{r}_\gamma) &\geq \inf_{\Psi} \left[\mathbf{P}_{\mathbf{0}, \Sigma}(\Psi = 1) + \sup_{\theta \in \tilde{\Theta}: \|\theta\|_2 \geq 2\gamma\sigma\sqrt{\phi(s, \rho^2)}} \mathbf{P}_{\theta, \Sigma}(\Psi = 0) \right] \\ &\geq \inf_{\Psi} \left[\mathbf{P}_{\mathbf{0}, \mathbf{I}_{p \times p}}(\Psi = 1) + \sup_{\tilde{\theta} \in B_0(1, p): \|\tilde{\theta}\|_2 \geq 2\gamma\sigma\sqrt{r^{-1}\phi(s, \rho^2)}} \mathbf{P}_{\tilde{\theta}, \mathbf{I}_{p \times p}}(\Psi = 0) \right], \end{aligned}$$

where in the last line the distributions are considered over a p -dimensional Euclidean space. Since according to (30) there exists an absolute constant $c' > 0$ such that $\phi(s, \rho^2) \leq c' r \phi(1, p)$ we further have

$$\begin{aligned} \mathcal{R}^*(s, d, \rho, \bar{r}_\gamma) &\geq \inf_{\Psi} \left[\mathbf{P}_{\mathbf{0}, \mathbf{I}_{p \times p}}(\Psi = 1) + \sup_{\tilde{\theta} \in B_0(1, p): \|\tilde{\theta}\|_2 \geq 2\gamma\sigma\sqrt{c' r \phi(1, p)}} \mathbf{P}_{\tilde{\theta}, \mathbf{I}_{p \times p}}(\Psi = 0) \right] \\ &\geq c(\sqrt{c' r} \gamma / 2), \end{aligned}$$

where the last inequality follows from the second inequality in (66) that we use with s and d taking values 1 and p , respectively. Combining the above lower bounds for both considered cases and using the expression for $c(\gamma)$ in (67) we obtain the desired result.

Acknowledgments

The authors would like to thank the anonymous referees, an Associate Editor and the Editor for their constructive comments that improved the quality of this paper.

Funding

The work of Y. Wang was supported by National Key R&D Program (No. 2022YFA1008100). The work of P. Yang was supported by Tsinghua University Dushi Program and the grant of National Natural Science Foundation of China (NSFC) 12101353. The work of A.B. Tsybakov was supported by the grants of French National Research Agency (ANR) “Investissements d’Avenir” LabEx Ecodec/ANR-11-LABX-0047 and ANR MaLip.

References

- [1] ABRAMOWITZ, M. and STEGUN, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* **55**. US Government printing office.
- [2] BARAUD, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli* **8** 577–606.
- [3] CARPENTIER, A., COLLIER, O., COMMINGES, L., TSYBAKOV, A. B. and WANG, Y. (2019). Minimax rate of testing in sparse linear regression. *Automation and Remote Control* **80** 1817–1834.
- [4] CARPENTIER, A., COLLIER, O., COMMINGES, L., TSYBAKOV, A. B. and WANG, Y. (2022). Estimation of the ℓ_2 -norm and testing in sparse linear regression with unknown variance. *Bernoulli* **28** 2744–2787.
- [5] CHHOR, J., MUKHERJEE, R. and SEN, S. (2024). Sparse Signal Detection in Heteroscedastic Gaussian Sequence Models: Sharp Minimax Rates. *Bernoulli* **30** 2127–2153.
- [6] COLLIER, O., COMMINGES, L. and TSYBAKOV, A. B. (2017). Minimax estimation of linear and quadratic functionals on sparsity classes. *The Annals of Statistics* **45** 923–958.
- [7] COMMINGES, L., COLLIER, O., NDAOUD, M. and TSYBAKOV, A. B. (2021). Adaptive robust estimation in sparse vector model. *The Annals of Statistics* **49** 1347–1377.
- [8] EATON, M. L. (1983). *Multivariate Statistics: a Vector Space Approach*. Wiley.
- [9] GEORGE, A. J. and CANONNE, C. L. (2022). Robust testing in high-dimensional sparse models. In *Advances in Neural Information Processing Systems* **35** 16469–16480.
- [10] HALL, P. and JIN, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *The Annals of Statistics* **38** 1686–1732.
- [11] HORN, R. A. and JOHNSON, C. R. (2012). *Matrix Analysis*, Second ed. Cambridge University Press.
- [12] HUBER, P. J. and RONCHETTI, E. M. (2009). *Robust Statistics*, Second ed. Wiley.
- [13] KOTEKAL, S. and GAO, C. (2023). Minimax rates for sparse signal detection under correlation. *Information and Inference: A Journal of the IMA* **12** 2873–2969.
- [14] KOTEKAL, S. and GAO, C. (2025). Sparsity meets correlation in Gaussian sequence model. *The Annals of Statistics* **53** 1095–1122.

- [15] LAURENT, B., LOUBES, J.-M. and MARTEAU, C. (2012). Non asymptotic minimax rates of testing in signal detection with heterogeneous variances. *Electronic Journal of Statistics* **6** 91–122.
- [16] LIU, H., GAO, C. and SAMWORTH, R. J. (2021). Minimax rates in sparse, high-dimensional change point detection. *The Annals of Statistics* **49** 1081–1112.
- [17] STOER, J. and BULIRSCH, R. (2002). *Introduction to Numerical Analysis*, Third ed. Springer.
- [18] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer.