The Annals of Statistics
2020, Vol. 48, No. 1, 464–490
https://doi.org/10.1214/19-AOS1816

EFFICIENT ESTIMATION OF LINEAR FUNCTIONALS OF PRINCIPAL COMPONENTS

BY VLADIMIR KOLTCHINSKII[1], MATTHIAS LÖFFLER[2,*] AND RICHARD NICKL[2,**]

[1]*School of Mathematics, Georgia Institute of Technology, vladimir.koltchinskii@math.gatech.edu*
[2]*Statistical Laboratory, University of Cambridge, *m.loffler@statslab.cam.ac.uk; **r.nickl@statslab.cam.ac.uk*

We study principal component analysis (PCA) for mean zero i.i.d. Gaussian observations $X_1, \dots, X_n$ in a separable Hilbert space $\mathbb{H}$ with unknown covariance operator Σ. The complexity of the problem is characterized by its effective rank $\mathbf{r}(\Sigma) := \frac{\mathrm{tr}(\Sigma)}{\|\Sigma\|}$, where $\mathrm{tr}(\Sigma)$ denotes the trace of Σ and $\|\Sigma\|$ denotes its operator norm. We develop a method of bias reduction in the problem of estimation of linear functionals of eigenvectors of Σ. Under the assumption that $\mathbf{r}(\Sigma) = o(n)$, we establish the asymptotic normality and asymptotic properties of the risk of the resulting estimators and prove matching minimax lower bounds, showing their semiparametric optimality.

1. Introduction. Principal Component Analysis (PCA) is commonly used as a dimension reduction technique for high-dimensional data sets. Assuming a general framework where the data lies in a Hilbert space $\mathbb{H}$, PCA can be applied to a wide range of problems such as functional data analysis [24, 29] or machine learning [4].

The parametric setting has been well understood since the 1960s (e.g., [1] and [8]) and the asymptotic distribution of sample eigenvalues and sample eigenvectors is well known. For high-dimensional data, where the dimension $p = p(n) \to \infty$ with the sample size n, the spiked covariance model introduced by Johnstone in [17] has been the most common framework to study the asymptotic properties of principal components. In this model, it is assumed that the covariance matrix is given by a "spike" and a noise part, that is

$$\Sigma = \sum_{j=1}^{l} s_i(\theta_i \otimes \theta_i) + \sigma^2 I_p,$$

where $\sum_{j=1}^{l} s_i(\theta_i \otimes \theta_i)$ is a low rank covariance matrix involving several orthonormal components ("spike") θ_i and $\sigma^2 I_p$ is the covariance of the noise. Error bounds in this model, based on perturbation analysis, were studied in [25]. Moreover, if $\frac{p}{n} \to c \in (0, 1]$ the asymptotic distribution of sample eigenvectors was derived in [2, 28] and in more general asymptotic regimes in [38]. Assuming sparsity of the eigenvectors (sparse PCA), inference is possible even when $\frac{p}{n} \to \infty$. This model has recently received substantial attention; see, for example, [3, 7, 11, 36, 37].

More recently, a so-called "effective rank" setting for PCA has been considered, for example, in [20–22, 26, 30, 35]. In this dimension-free setting, it is assumed that the covariance Σ is an operator acting in a Hilbert space $\mathbb{H}$, no structural assumptions are made about Σ and its "complexity" is characterized by the *effective rank* $\mathbf{r}(\Sigma) := \mathrm{tr}(\Sigma)/\|\Sigma\|$, $\mathrm{tr}(\Sigma)$ denoting the trace and $\|\Sigma\|$ denoting the operator (spectral) norm of Σ. In a series of papers [20–23], Koltchinskii and Lounici derived sharp bounds on the spectral norm loss of estimation of Σ by the sample covariance $\hat{\Sigma}$ that provide complete characterization of the size of $\|\hat{\Sigma} - \Sigma\|$

Received September 2017; revised January 2019.
MSC2010 subject classifications. Primary 62H25; secondary 62E17.
Key words and phrases. Principal component analysis, spectral projections, asymptotic normality, semiparametric efficiency.

in terms of $\|\Sigma\|$ and $\mathbf{r}(\Sigma)$, and obtained error bounds and limiting results for empirical spectral projection operators and eigenvectors of $\hat{\Sigma}$ under the assumption that $\mathbf{r}(\Sigma) = o(n)$ as $n \to \infty$. In a recent paper [26], Naumov et al. constructed bootstrap confidence sets for spectral projections in a lower dimensional regime where $\mathbf{r}(\Sigma) = o(n^{1/3})$. In [30], Reiss and Wahl considered the reconstruction error for spectral projections.

In this paper, we further develop the results of [20] and [22] in the direction of semiparametric statistics. In particular, we develop a bias reduction method in the problem of estimation of linear functionals of principal components (eigenvectors of Σ) and show asymptotic normality of the resulting debiased estimators under the assumption that $\mathbf{r}(\Sigma) = o(n)$. We prove a nonasymptotic risk lower bound that asymptotically exactly matches our upper bounds, thus establishing rigorously the semiparametric optimality of our estimator in a general dimension-free setting (as long as $\mathbf{r}(\Sigma) = o(n)$).

The problem of $\sqrt{n}$-consistent estimation of low-dimensional functionals of high-dimensional parameters has received increased attention in recent years, and in various models semiparametric efficiency of regularisation-based estimators has been studied; see, for instance, [10, 15, 27, 31, 32]. Moreover, the paper [12] develops Bernstein–von-Mises (BvM) results for functionals of covariance matrices in situations where bias is asymptotically negligible. While formal calculations of the Fisher information in such models indicate optimality of these procedures, a rigorous interpretation of such efficiency claims requires some care: the standard asymptotic setting for semiparametric efficiency [33] can not be straightforwardly applied because parameters in high-dimensional models are not fixed but vary with sample size n, so that establishing LAN expansions to apply Le Cam theory is not always possible or even desirable. In [15] some nonasymptotic techniques have been suggested under conditions that ensure asymptotic negligibility of the bias of candidate estimators. We take here a different approach, based on using the van Trees' inequality [13] to construct nonasymptotic lower bounds for the minimax risk in our estimation problem that match the upper bound *exactly* in the large sample limit.

2. Preliminaries.

2.1. *Some notation and conventions.* Let $\mathbb{H}$ be a separable Hilbert space. In what follows, $\langle \cdot, \cdot \rangle$ denotes the inner product of $\mathbb{H}$ and also, with a little abuse of notation, the Hilbert–Schmidt inner product between Hilbert–Schmidt operators acting on $\mathbb{H}$. Similarly, the notation $\|\cdot\|$ is used both for the norm of vectors in $\mathbb{H}$ and for the operator (spectral) norm of bounded linear operators in $\mathbb{H}$. For a nuclear operator A, $\operatorname{tr}(A)$ denotes its trace. We use the notation $\|\cdot\|_p$, $1 \le p \le \infty$ for the Schatten p-norms of operators in $\mathbb{H}$: $\|A\|_p := (\operatorname{tr}(|A|^p))^{1/p}$, where $|A| = \sqrt{A^*A}$, A^* being the adjoint operator of A. For $p = 1$, $\|A\|_1$ is the nuclear norm; for $p = 2$, $\|A\|_2$ is the Hilbert–Schmidt norm; for $p = \infty$, $\|A\|_\infty = \|A\|$ is the operator norm.

Given vectors $u, v \in \mathbb{H}$, $u \otimes v$ denotes the tensor product of u and v:

$$(u \otimes v) : \mathbb{H} \mapsto \mathbb{H}, \qquad (u \otimes v)w := \langle v, w \rangle u.$$

Given bounded linear operators $A, B : \mathbb{H} \mapsto \mathbb{H}$, $A \otimes B$ denotes their tensor product:

$$(A \otimes B)(u \otimes v) = Au \otimes Bv, \quad u, v \in \mathbb{H}.$$

Note that $A \otimes B$ can be extended (by linearity and continuity) to a bounded operator in the Hilbert space $\mathbb{H} \otimes \mathbb{H}$, which could be identified with the space of Hilbert–Schmidt operators in $\mathbb{H}$. It is easy to see that, for a Hilbert–Schmidt operator C, we have $(A \otimes B)C = ACB^*$ (in the finite-dimensional case, this defines the so called Kronecker product of matrices). On a couple of occasions, we might need to use the tensor product of Hilbert–Schmidt operators

A, B, viewed as vectors in the space of Hilbert–Schmidt operators. For this tensor product, we use the notation $A \otimes_v B$.

Throughout the paper, the following notation will be used: for nonnegative a, b, $a \lesssim b$ means that there exists a numerical constant $c > 0$ such that $a \le cb$; $a \gtrsim b$ is equivalent to $b \lesssim a$; finally, $a \asymp b$ is equivalent to $a \lesssim b$ and $b \lesssim a$. Sometimes, constant c in the above relationships could depend on some parameter γ. In this case, we provide signs $\lesssim$, $\gtrsim$ and $\asymp$ with subscript γ. For instance, $a \lesssim_\gamma b$ means that there exists a constant $c_\gamma > 0$ such that $a \le c_\gamma b$.

In many places in the proofs, we use exponential bounds for some random variables, say, ξ of the following form: for all $t \ge 1$ with probability at least $1 - e^{-t}$, $\xi \le Ct$. In some cases, it would follow from our arguments that the inequality holds with a slightly different probability, say, at least $1 - 3e^{-t}$. In such cases, it is easy to rewrite the bound again as $1 - e^{-t}$ by adjusting the value of constant C. Indeed, for $t \ge 1$ with probability at least $1 - e^{-t} = 1 - 3e^{-t-\log(3)}$, we have $\xi \le C(t + \log(3)) \le 2\log(3)Ct$. We will use such an adjustment of the constants in many proofs, often, without further notice.

2.2. *Bounds on sample covariance.* Let X be a Gaussian vector in $\mathbb{H}$ with mean $\mathbb{E}X = 0$ and covariance operator $\Sigma := \mathbb{E}(X \otimes X)$. Given i.i.d. observations $X_1, \dots, X_n$ of X, let $\hat{\Sigma} = \hat{\Sigma}_n$ be the sample (empirical) covariance operator defined as follows:

$$\hat{\Sigma} := n^{-1} \sum_{j=1}^{n} X_j \otimes X_j.$$

DEFINITION 2.1. The effective rank of the covariance operator Σ is defined as

$$\mathbf{r}(\Sigma) := \frac{\operatorname{tr}(\Sigma)}{\|\Sigma\|}.$$

The role of the effective rank as a complexity parameter in covariance estimation is clear from the following result proved in [21].

THEOREM 2.1. *Let X be a mean zero Gaussian random vector in $\mathbb{H}$ with covariance operator Σ and let $\hat{\Sigma}$ be the sample covariance based on i.i.d. observations $X_1, \dots, X_n$ of X. Then*

$$\mathbb{E}\|\hat{\Sigma} - \Sigma\| \asymp \|\Sigma\| \left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \frac{\mathbf{r}(\Sigma)}{n} \right). \tag{2.1}$$

This result shows that the size of the properly rescaled operator norm deviation of $\hat{\Sigma}$ from Σ, $\frac{\mathbb{E}\|\hat{\Sigma} - \Sigma\|}{\|\Sigma\|}$, is characterized up to numerical constants by the ratio $\frac{\mathbf{r}(\Sigma)}{n}$. In particular, the condition $\mathbf{r}(\Sigma) = o(n)$ is necessary and sufficient for operator norm consistency of $\hat{\Sigma}$ as an estimator of Σ. In addition to this, the following concentration inequality for $\|\hat{\Sigma} - \Sigma\|$ around its expectation was also proved in [21].

THEOREM 2.2. *Under the conditions of the previous theorem, for all $t \ge 1$ with probability at least $1 - e^{-t}$*

$$\big| \|\hat{\Sigma} - \Sigma\| - \mathbb{E}\|\hat{\Sigma} - \Sigma\| \big| \lesssim \|\Sigma\| \left(\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee 1 \right) \sqrt{\frac{t}{n}} \vee \frac{t}{n} \right). \tag{2.2}$$

It immediately follows from the bounds (2.1) and (2.2) that, for all $t \geq 1$ with probability at least $1 - e^{-t}$

$$\|\hat{\Sigma} - \Sigma\| \lesssim \|\Sigma\| \biggl(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \frac{\mathbf{r}(\Sigma)}{n} \vee \sqrt{\frac{t}{n}} \vee \frac{t}{n}\biggr). \tag{2.3}$$

2.3. *Perturbation theory and empirical spectral projections.* The covariance operator Σ is self-adjoint, positively semidefinite and nuclear. It has spectral decomposition

$$\Sigma = \sum_{r \geq 1} \mu_r P_r,$$

where μ_r are distinct strictly positive eigenvalues of Σ arranged in decreasing order and P_r are the corresponding spectral projection operators. For $r \geq 1$, P_r is an orthogonal projection on the eigenspace of the eigenvalue μ_r. The dimension of this eigenspace is finite and will be denoted by m_r. The eigenspaces corresponding to different eigenvalues μ_r are mutually orthogonal. Denote by $\sigma(\Sigma)$ the spectrum of operator Σ and let $\lambda_j = \lambda_j(\Sigma)$, $j \geq 1$ be the eigenvalues of Σ arranged in a nonincreasing order and repeated with their multiplicities. Denote $\Delta_r := \{j : \lambda_j = \mu_r\}$, $r \geq 1$. Then $\mathrm{card}(\Delta_r) = m_r$. The r-th spectral gap is defined as

$$g_r = g_r(\Sigma) := \mathrm{dist}\bigl(\mu_r; \sigma(\Sigma) \setminus \{\mu_r\}\bigr).$$

Let $\bar{g}_r = \bar{g}_r(\Sigma) := \min_{1 \leq s \leq r} g_s$.

We turn now to the definition of empirical spectral projections of sample covariance $\hat{\Sigma}$ that could be viewed as estimators of the true spectral projections $P_r, r \geq 1$. In [20], the following definition was used: let $\hat{P}_r$ be the orthogonal projection on the direct sum of eigenspaces of $\hat{\Sigma}$ corresponding to its eigenvalues $\{\lambda_j(\hat{\Sigma}) : j \in \Delta_r\}$. This is not a perfect definition of a statistical estimator since the set Δ_r is unknown and it has to be recovered from the spectrum $\sigma(\hat{\Sigma})$ of $\hat{\Sigma}$.

When $\hat{\Sigma}$ is close to Σ in the operator norm, the spectrum $\sigma(\hat{\Sigma})$ of $\hat{\Sigma}$ is a small perturbation of the spectrum $\sigma(\Sigma)$ of Σ. This could be quantified by the following inequality that goes back to H. Weyl:

$$\sup_{j \geq 1} \bigl|\lambda_j(\hat{\Sigma}) - \lambda_j(\Sigma)\bigr| \leq \|\hat{\Sigma} - \Sigma\|. \tag{2.4}$$

It easily follows from this inequality that, if $\|\hat{\Sigma} - \Sigma\|$ is sufficiently small, then the eigenvalues $\lambda_j(\hat{\Sigma})$ of $\hat{\Sigma}$ form well separated clusters around the eigenvalues $\mu_1, \mu_2, \ldots$ of Σ. To make the last claim more precise, consider a finite or countable bounded set $A \subset \mathbb{R}_+$ such that $0 \in A$ and 0 is the only limit point (if any) of A. Given $\delta > 0$, define $\lambda_\delta(A) := \max\{\lambda \in A : (\lambda - \delta, \lambda) \cap A = \varnothing\}$ and let $T_\delta(A) := A \setminus [0, \lambda_\delta(A))$. The set $T_\delta(A)$ will be called the top δ-cluster of A. Let $A_1 := T_\delta(A)$, $A_2 := T_\delta(A \setminus A_1)$, $A_3 := T_\delta(A \setminus (A_1 \cup A_2)), \ldots$ and $\nu = \nu_\delta := \min\{j : A_{j+1} = \varnothing\}$. Obviously, $\nu < \infty$. We will call the sets $A_1, \ldots, A_\nu$ the δ-clusters of A. They provide a partition of A into sets separated by the gaps of length at least δ and such that the gaps between the points inside each of the clusters are smaller than δ.

The next lemma easily follows from inequality (2.4).

LEMMA 2.1. *Let $\delta > 0$ be such that, for some $r \geq 1$,*

$$\|\hat{\Sigma} - \Sigma\| < \delta/2 \quad \text{and} \quad \delta < \frac{\bar{g}_r}{2}.$$

Let $\hat{A}_1^\delta, \ldots, \hat{A}_\nu^\delta$ be the δ-clusters of the set $\sigma(\hat{\Sigma})$. Then $\nu \geq r$ and, for all $1 \leq s \leq r$

$$\hat{A}_s^\delta \subset (\mu_s - \delta/2, \mu_s + \delta/2) \quad \text{and} \quad \bigl\{j : \lambda_j(\hat{\Sigma}) \in \hat{A}_s^\delta\bigr\} = \Delta_s.$$

Given $\delta > 0$ and δ-clusters $\hat{A}_1^\delta, \dots, \hat{A}_\nu^\delta$ of $\sigma(\hat{\Sigma})$, define, for $1 \le s \le \nu$, the empirical spectral projection $\hat{P}_s^\delta$ as the orthogonal projection on the direct sum of eigenspaces of $\hat{\Sigma}$ corresponding to its eigenvalues from the cluster $\hat{A}_s^\delta$. It immediately follows from Lemma 2.1 that, under its assumptions on δ, $\hat{P}_s^\delta = \hat{P}_s$, $s = 1, \dots, r$.

In the following sections, we will be interested in the problem of estimation of spectral projections in the case when the true covariance Σ belongs to certain subsets of the following class of covariance operators:

$$\mathcal{S}^{(r)}(\mathfrak{r}; a) := \left\{\Sigma : \mathbf{r}(\Sigma) \le \mathfrak{r}, \frac{\|\Sigma\|}{\bar{g}_r(\Sigma)} \le a\right\}, \quad a > 1, \mathfrak{r} > 1.$$

We will allow the effective rank to be large, $\mathfrak{r} = \mathfrak{r}_n \to \infty$, but not too large such that $\mathfrak{r}_n = o(n)$ as $n \to \infty$. For $\Sigma \in \mathcal{S}^{(r)}(\mathfrak{r}; a)$, we take $\delta := \tau \|\hat{\Sigma}\|$ for a sufficiently small value of the constant $\tau > 0$ in the definition of spectral projections $\hat{P}_s^\delta$.

The following lemma is an easy consequence of the exponential bound (2.3).

LEMMA 2.2. *Suppose $a > 1$ and $\mathfrak{r}_n = o(n)$ as $n \to \infty$. Take $\tau \in (0, \frac{1}{4a} \wedge 2)$ and $\delta := \tau \|\hat{\Sigma}\|$. Then, there exists a numerical constant $\beta > 0$ such that, for all large enough n,*

$$\sup_{\Sigma \in \mathcal{S}^{(r)}(\mathfrak{r}; a)} \mathbb{P}_\Sigma\{\exists s = 1, \dots, r : \hat{P}_s^\delta \ne \hat{P}_s\} \le e^{-\beta\tau^2 n}.$$

PROOF. By (2.3) with $t := \beta\tau^2 n$, we obtain that

$$\sup_{\Sigma \in \mathcal{S}^{(r)}(\mathfrak{r}; a)} \mathbb{P}_\Sigma\left\{\|\hat{\Sigma} - \Sigma\| \ge C\|\Sigma\|\left(\sqrt{\frac{\mathfrak{r}_n}{n}} \vee \sqrt{\frac{\beta\tau^2 n}{n}}\right)\right\} \le e^{-\beta\tau^2 n},$$

where $C > 0$ is a numerical constant. Take $\beta = \frac{1}{16C^2}$ and note that, for all large enough n, $C\sqrt{\frac{\mathfrak{r}_n}{n}} \le \tau/4$ to obtain that

$$\sup_{\Sigma \in \mathcal{S}^{(r)}(\mathfrak{r}; a)} \mathbb{P}_\Sigma\{\|\hat{\Sigma} - \Sigma\| \ge (\tau/4)\|\Sigma\|\} \le e^{-\beta\tau^2 n}.$$

Since $\tau/4 \le 1/2$, we easily obtain that, for all $\Sigma \in \mathcal{S}^{(r)}(\mathfrak{r}; a)$ and for all n large enough with probability at least $1 - e^{-\beta\tau^2 n}$, $(1/2)\|\Sigma\| \le \|\hat{\Sigma}\| \le 2\|\Sigma\|$. This implies that with the same probability (and on the same event)

$$\|\hat{\Sigma} - \Sigma\| < (\tau/4)\|\Sigma\| \le (\tau/2)\|\hat{\Sigma}\| = \delta/2.$$

On the other hand, for all $\Sigma \in \mathcal{S}^{(r)}(\mathfrak{r}; a)$,

$$\delta = \tau\|\hat{\Sigma}\| \le 2\tau\|\Sigma\| < \frac{1}{2a}\|\Sigma\| \le \frac{\bar{g}_r(\Sigma)}{2}.$$

It remains to use Lemma 2.1 to complete the proof. □

In the proofs of the main results of the paper, we deal for the most part with spectral projections $\hat{P}_r$ that were studied in detail in [20]. We use Lemma 2.2 to reduce the results for $\hat{P}_r^\delta$ to the results for $\hat{P}_r$.

3. Main results. Our main goal is to develop an efficient estimator of the linear functional $\langle\theta_r, u\rangle$, where $u \in \mathbb{H}$ is a given vector and $\theta_r = \theta_r(\Sigma)$ is a unit eigenvector of the unknown covariance operator Σ corresponding to its rth eigenvalue μ_r, which is assumed to be simple (that is, of multiplicity $m_r = 1$). The corresponding spectral projection P_r is one-dimensional: $P_r = \theta_r \otimes \theta_r$. A "naive" plug-in estimator of P_r is the empirical spectral projection $\hat{P}_r^\delta$ with $\delta = \tau\|\hat{\Sigma}\|$ for a suitable choice of a small constant τ, as described in Lemma 2.2. According to this lemma and under its assumptions, $\hat{P}_r^\delta$ coincides with a high probability with the one-dimensional empirical spectral projection $\hat{P}_r := \hat{\theta}_r \otimes \hat{\theta}_r$, where $\hat{\theta}_r$ is the corresponding unit eigenvector of $\hat{\Sigma}$. As an estimator of θ_r, we can use an arbitrary unit vector $\hat{\theta}_r^\delta$ from the eigenspace $\operatorname{Im}(\hat{P}_r^\delta)$, which with a high-probability coincides with $\pm\hat{\theta}_r$ (under conditions of Lemma 2.2). In case $r = 1$, when the top eigenvalue $\mu_1 = \|\Sigma\|$ of Σ is simple and the goal is to estimate a linear functional of the top principal component θ_1, there is no need to use δ-clusters to define an estimator of θ_1 since $\hat{\theta}_1$ (a unit eigenvector in the eigenspace of the top eigenvalue $\|\hat{\Sigma}\|$ of $\hat{\Sigma}$) is already a legitimate estimator.

Note that both θ_r and $-\theta_r$ are unit eigenvectors of Σ, so, strictly speaking, $\langle\theta_r, u\rangle$ can be estimated only up to its sign. In what follows, we assume that $\hat{\theta}_r^\delta$ and θ_r (or, whenever is needed, $\hat{\theta}_r$ and θ_r) are *properly aligned* in the sense that $\langle\hat{\theta}_r^\delta, \theta_r\rangle \geq 0$ (which is always the case either for θ_r, or for $-\theta_r$). This allows us to view $\langle\hat{\theta}_r^\delta, u\rangle$ as an estimator of $\langle\theta_r, u\rangle$.

It was shown in [20] that "naive" plug-in estimators of the functional $\langle\theta_r, u\rangle$, such as $\langle\hat{\theta}_r^\delta, u\rangle$ or $\langle\hat{\theta}_r, u\rangle$, are biased with the bias becoming substantial enough to affect the efficiency of the estimator or even its convergence rates as soon as the effective rank is large enough, namely, $\mathbf{r}(\Sigma) \gtrsim n^{1/2}$. Moreover, it was shown that the quantity

$$b_r = b_r(\Sigma) := \mathbb{E}_\Sigma \langle\hat{\theta}_r, \theta_r\rangle^2 - 1 \in [-1, 0]$$

plays the role of a bias parameter. In particular, the results of [20] imply that the random variable $\langle\hat{\theta}_r, u\rangle$ concentrates around $\sqrt{1 + b_r}\langle\theta_r, u\rangle$ (rather than around $\langle\theta_r, u\rangle$) with the size of the deviations of order $O(n^{-1/2})$ provided that $\mathbf{r}(\Sigma) = o(n)$ as $n \to \infty$. Thus, the bias of $\langle\hat{\theta}_r, u\rangle$ as an estimator of $\langle\theta_r, u\rangle$ is of the order $(\sqrt{1 + b_r} - 1)\langle\theta_r, u\rangle \asymp b_r\langle\theta_r, u\rangle$. It was shown in [20] that $|b_r| \lesssim \frac{\mathbf{r}(\Sigma)}{n}$ and it will be proved below in this paper that, in fact, $|b_r| \asymp \frac{\mathbf{r}(\Sigma)}{n}$ (see Lemma 1.1 in the Supplementary Material [19]). This fact implies that, indeed, the bias of $\langle\hat{\theta}_r, u\rangle$ (and of $\langle\hat{\theta}_r^\delta, u\rangle$) is not negligible and affects the convergence rate as soon as $\frac{\mathbf{r}(\Sigma)}{n^{1/2}} \to \infty$. This resembles the situation in sparse regression (see, e.g., [16, 32, 39]): If p denotes the dimension of the model and s its sparsity and if $s\log(p) = o(n^{1/2})$, the bias of a desparsified LASSO estimator for the regressor β is negligible, which makes it possible to prove asymptotic normality of linear forms of β. On the other hand, if $s\log(p) \gg n^{1/2}$, Cai and Guo [6] proved that adaptive confidence sets for linear forms do not exist in general. This implies that any attempt to further debias the desparsified LASSO or any other estimator to prove asymptotic normality is deemed to fail. Contrary to this, in our case estimation of the bias parameter b_r is possible (as will be shown below).

We will state a uniform (and somewhat stronger) version of some of the results of [20] on asymptotic normality of linear forms

$$\sqrt{n}\big(\langle\hat{\theta}_r^\delta, u\rangle - \sqrt{1 + b_r(\Sigma)}\langle\theta_r(\Sigma), u\rangle\big), \quad u \in \mathbb{H}$$

under the assumption that $\mathbf{r}(\Sigma) = o(n)$. To this end, define the following operator:

$$C_r := \sum_{s \neq r} \frac{1}{\mu_r - \mu_s} P_s,$$

which is bounded with $\|C_r\| = \frac{1}{g_r}$. Denote

$$\sigma_r^2(\Sigma; u) := \langle\Sigma\theta_r, \theta_r\rangle\langle\Sigma C_r u, C_r u\rangle = \mu_r\langle\Sigma C_r u, C_r u\rangle.$$

Clearly,

$$\sigma_r^2(\Sigma;u) \le \frac{\|\Sigma\|^2}{g_r^2}\|u\|^2. \tag{3.1}$$

Note that, if $\mathbb{H}$ is finite-dimensional (with a fixed dimension) and Σ is nonsingular, then the Fisher information for the model $X \sim N(0;\Sigma)$ is $\mathbb{I}(\Sigma) = \frac{1}{2}(\Sigma^{-1} \otimes \Sigma^{-1})$ (see, e.g., [9]). The maximum likelihood estimator $\hat{\Sigma}$ based on n i.i.d. observations of X (the sample covariance) is then asymptotically normal with $\sqrt{n}$-rate and limit covariance $\mathbb{I}(\Sigma)^{-1} = 2(\Sigma \otimes \Sigma)$. An application of the Delta Method to the smooth function $g(\Sigma) := \langle \theta_r(\Sigma), u\rangle$ shows that $g(\hat{\Sigma})$ is also asymptotically normal with limiting variance $\langle (\mathbb{I}(\Sigma)^{-1} g'(\Sigma), g'(\Sigma)\rangle$, which turns out to be equal to $\sigma_r^2(\Sigma;u)$.

For $u \in \mathbb{H}$, $\mathfrak{r} > 1$, $a > 1$ and $\sigma_0 > 0$, consider the following class of covariance operators in $\mathbb{H}$:

$$\mathcal{S}^{(r)}(\mathfrak{r}, a, \sigma_0, u) := \left\{\Sigma : \mathbf{r}(\Sigma) \le \mathfrak{r}, \frac{\|\Sigma\|}{\bar{g}_r(\Sigma)} \le a, \sigma_r^2(\Sigma;u) \ge \sigma_0^2\right\}.$$

We emphasize here that we regard a and σ_0 as fixed constants, but $\mathfrak{r}$, $\|\Sigma\|$ and $\bar{g}_r$ may all possibly depend on n. For example, this allows that $\|\Sigma\| \to \infty$ as long as $\bar{g}_r \to \infty$ at the same rate as it is the case in factor models as considered in [38]. Note that some additional conditions on $\mathfrak{r}, a, \sigma_0, u$ are needed for the class $\mathcal{S}^{(r)}(\mathfrak{r}, a, \sigma_0, u)$ to be nonempty. Say, bound (3.1) implies that it is necessary for this that $\sigma_0^2 \le a^2\|u\|^2$. It is also obvious that there should be $a > r$ (since $\|\Sigma\| \ge r g_r(\Sigma)$).

We will also need the following assumption on the loss function ℓ.

ASSUMPTION 3.1. Let $\ell : \mathbb{R} \mapsto \mathbb{R}_+$ be a loss function satisfying the following conditions: $\ell(0) = 0$, $\ell(u) = \ell(-u)$, $u \in \mathbb{R}$, ℓ is nondecreasing and convex on $\mathbb{R}_+$ and, for some constants $c_1, c_2 > 0$

$$\ell(u) \le c_1 e^{c_2 u}, \quad u \ge 0.$$

The proofs to all our theorems are in fact nonasymptotic and often can be expressed by Berry–Esseen type bounds. However, for a more concise presentation we present asymptotic statements.

In what follows, Z denotes a standard Gaussian random variable and Φ denotes its distribution function.

THEOREM 3.1. *Let $u \in \mathbb{H}$, $a > 1$ and $\sigma_0 > 0$. Suppose that $\mathfrak{r}_n > 1$ and $\mathfrak{r}_n = o(n)$ as $n \to \infty$. Let $\delta = \tau\|\hat{\Sigma}\|$ for some $\tau \in (0, \frac{1}{4a} \wedge 2)$. Then*

$$\sup_{\Sigma \in \mathcal{S}^{(r)}(\mathfrak{r}_n, a, \sigma_0, u)} \sup_{x \in \mathbb{R}} \left| \mathbb{P}_\Sigma\left\{ \frac{\sqrt{n}(\langle \hat{\theta}_r^\delta, u\rangle - \sqrt{1 + b_r(\Sigma)}\langle \theta_r(\Sigma), u\rangle)}{\sigma_r(\Sigma;u)} \le x \right\} - \Phi(x) \right| \overset{n\to\infty}{\to} 0. \tag{3.2}$$

Moreover, under Assumption 3.1,

$$\sup_{\Sigma \in \mathcal{S}^{(r)}(\mathfrak{r}_n, a, \sigma_0, u)} \left| \mathbb{E}_\Sigma \ell\left(\frac{\sqrt{n}(\langle \hat{\theta}_r^\delta, u\rangle - \sqrt{1 + b_r(\Sigma)}\langle \theta_r(\Sigma), u\rangle)}{\sigma_r(\Sigma;u)} \right) - \mathbb{E}\ell(Z) \right| \overset{n\to\infty}{\to} 0.$$

The proof of this theorem will be given in Section 4 that also includes a number of auxiliary statements used in the proofs of our main results on efficient estimation of linear functionals.

COROLLARY 3.1. *Let $u \in \mathbb{H}$, $a > 1$ and $\sigma_0 > 0$. Suppose that $\mathfrak{r}_n > 1$ and $\mathfrak{r}_n = o(\sqrt{n})$ as $n \to \infty$. Let $\delta = \tau \|\hat{\Sigma}\|$ for some $\tau \in (0, \frac{1}{4a} \wedge 2)$. Then*

$$\sup_{\Sigma \in \mathcal{S}^{(r)}(\mathfrak{r}_n, a, \sigma_0, u)} \sup_{x \in \mathbb{R}} \left| \mathbb{P}_\Sigma \left\{ \frac{\sqrt{n}(\langle \hat{\theta}_r^\delta, u\rangle - \langle \theta_r(\Sigma), u\rangle)}{\sigma_r(\Sigma; u)} \le x \right\} - \Phi(x) \right| \overset{n\to\infty}{\to} 0.$$

Moreover, under Assumption 3.1,

$$\sup_{\Sigma \in \mathcal{S}^{(r)}(\mathfrak{r}_n, a, \sigma_0, u)} \left| \mathbb{E}_\Sigma \ell\left(\frac{\sqrt{n}(\langle \hat{\theta}_r^\delta, u\rangle - \langle \theta_r(\Sigma), u\rangle)}{\sigma_r(\Sigma; u)} \right) - \mathbb{E}\ell(Z) \right| \overset{n\to\infty}{\to} 0.$$

Our next goal is to provide a minimax lower bound on the risk of an arbitrary estimator of the linear functional $\langle \theta_r(\Sigma), u\rangle$ in the case of quadratic loss $\ell(t) = t^2$, $t \in \mathbb{R}$. The proof is based on van Trees' inequality and will be given in Section 3 in the Supplementary Material [19]. For $\mathfrak{r} > 1$, $a > 1$ and $\sigma_0^2 > 0$ define

$$\mathring{\mathcal{S}}^{(r)}(\mathfrak{r}, a, \sigma_0, u) := \left\{ \Sigma : \mathbf{r}(\Sigma) < \mathfrak{r}, \frac{\|\Sigma\|}{\bar{g}_r(\Sigma)} < a, \sigma_r^2(\Sigma; u) > \sigma_0^2 \right\},$$

the interior of the set $\mathcal{S}^{(r)}(\mathfrak{r}, a, \sigma_0, u)$.

THEOREM 3.2. *Let $\mathfrak{r} > 1$, $a > 1$ and $\sigma_0 > 0$. Suppose $\mathring{\mathcal{S}}^{(r)}(\mathfrak{r}, a, \sigma_0, u) \neq \emptyset$. Then, for all statistics $T_n(X_1, \dots, X_n)$,*

$$\liminf_{n\to\infty} \inf_{T_n} \sup_{\Sigma \in \mathring{\mathcal{S}}^{(r)}(\mathfrak{r}, a, \sigma_0, u)} \frac{n\mathbb{E}_\Sigma (T_n(X_1, \dots, X_n) - \langle \theta_r(\Sigma), u\rangle)^2}{\sigma_r^2(\Sigma; u)} \ge 1.$$

Moreover, for any $\Sigma_0 \in \mathring{\mathcal{S}}^{(r)}(\mathfrak{r}, a, \sigma_0, u)$

$$\lim_{\varepsilon \to 0} \liminf_{n\to\infty} \inf_{T_n} \sup_{\substack{\Sigma \in \mathring{\mathcal{S}}^{(r)}(\mathfrak{r}, a, \sigma_0, u), \\ \|\Sigma - \Sigma_0\|_1 \le \varepsilon}} \frac{n\mathbb{E}_\Sigma (T_n(X_1, \dots, X_n) - \langle \theta_r(\Sigma), u\rangle)^2}{\sigma_r^2(\Sigma; u)} \ge 1.$$

It follows from Corollary 3.1 and Theorem 3.2 that the estimator $\langle \hat{\theta}_r^\delta, u\rangle$ is efficient in a semiparametric sense for quadratic loss under the assumption that $\mathfrak{r}_n = o(n^{1/2})$. It turns out, however, that if $\frac{\mathfrak{r}_n}{n^{1/2}} \to \infty$, then not only the efficiency, but even the $\sqrt{n}$-convergence rate of this estimator fails in the class of covariance operators $\mathcal{S}^{(r)}(\mathfrak{r}_n, a, \sigma_0, u)$. The proof of Proposition 3.1 is given in Section 1 in the Supplementary Material [19].

PROPOSITION 3.1. *Let $a > r$ and let σ_0^2 be sufficiently small, say,*

$$\sigma_0^2 \le \frac{1}{2}\left[\frac{a^2}{(r-1)^2} - \frac{a}{r-1} \right].$$

Let $\mathfrak{r}_n = o(n)$ and $\frac{\mathfrak{r}_n}{n^{1/2}} \to \infty$ as $n \to \infty$. Then, for some constant $c = c(r; a; \sigma_0) > 0$

$$\lim_{n\to\infty} \sup_{\Sigma \in \mathcal{S}^{(r)}(\mathfrak{r}_n, a, \sigma_0, u)} \mathbb{P}_\Sigma \left\{ \left| \langle \hat{\theta}_r^\delta, u\rangle - \langle \theta_r(\Sigma), u\rangle \right| \ge c\|u\| \frac{\mathfrak{r}_n}{n} \right\} = 1.$$

The reason for the loss of the $\sqrt{n}$-convergence rate of plug-in estimators of linear functionals of principal components is their large bias in the case when the complexity of the problem is even moderately high ($\frac{\mathfrak{r}_n}{n^{1/2}} \to \infty$). In [20], a method of bias reduction in this problem was suggested that led to $\sqrt{n}$-consistent estimation of linear functionals. The estimator is, however, not efficient, since the basic sample split employed in its construction gives a limiting variance that is twice as large as the optimal one. Since the bias parameter depend itself on sample size in a subtle way, modifying the algorithm in [20] to obtain an efficient estimator is not straightforward, and we describe below a construction that yields an asymptotically normal estimator of $\langle \theta_r(\Sigma), u\rangle$ with optimal variance in the class of covariance operators $\mathcal{S}^{(r)}(\mathfrak{r}_n, a, \sigma_0, u)$ with $\mathfrak{r}_n = o(n)$. The idea is to use only a small portion of the data (of size $o(n)$) to estimate the bias parameters and to use most of the data for the estimator of the target eigenvector.

For some $m < n/3$, we split the sample $X_1, \dots, X_n$ into three disjoint subsamples, one of size $n' := n - 2m > n/3$ and two others of size m each. In Theorem 3.3 below, we choose $m = m_n = o(n)$ as $n \to \infty$, which implies $n' = n'_n = (1 + o(1))n$ as $n \to \infty$. Denote by $\hat{\Sigma}^{(1)}$, $\hat{\Sigma}^{(2)}$, $\hat{\Sigma}^{(3)}$ the sample covariances based on these three subsamples and let $\hat{\theta}_r^{\delta_j, j}$, $j = 1, 2, 3$ be the corresponding empirical eigenvectors with parameters $\delta_j = \tau \|\hat{\Sigma}^{(j)}\|$ for a proper choice of τ (see Lemma 2.2). Let

$$\check{d}_r := \frac{\langle \hat{\theta}_r^{\delta_1,1}, \hat{\theta}_r^{\delta_2,2}\rangle}{\langle \hat{\theta}_r^{\delta_2,2}, \hat{\theta}_r^{\delta_3,3}\rangle^{1/2}} \quad \text{and} \quad \check{\theta}_r := \frac{\hat{\theta}_r^{\delta_1,1}}{\check{d}_r \vee (1/2)}.$$

Our main goal is to prove the following result showing the efficiency of the estimator $\langle \check{\theta}_r, u\rangle$ of the linear functional $\langle \theta_r(\Sigma), u\rangle$. Its proof will be given in Section 5.

THEOREM 3.3. *Let $u \in \mathbb{H}$, $a > 1$ and $\sigma_0 > 0$. Suppose that $\mathfrak{r}_n > 1$ and $\mathfrak{r}_n = o(n)$ as $n \to \infty$. Take $m = m_n$ such that $m_n = o(n)$ and $n\mathfrak{r}_n = o(m_n^2)$ as $n \to \infty$. Then*

$$\sup_{\Sigma \in \mathcal{S}^{(r)}(\mathfrak{r}_n, a, \sigma_0, u)} \sup_{x \in \mathbb{R}} \left| \mathbb{P}_\Sigma \left\{ \frac{\sqrt{n}(\langle \check{\theta}_r, u\rangle - \langle \theta_r(\Sigma), u\rangle)}{\sigma_r(\Sigma; u)} \le x \right\} - \Phi(x) \right| \overset{n\to\infty}{\to} 0. \tag{3.3}$$

Moreover, under Assumption 3.1 on the loss ℓ,

$$\sup_{\Sigma \in \mathcal{S}^{(r)}(\mathfrak{r}_n, a, \sigma_0, u)} \left| \mathbb{E}_\Sigma \ell\left(\frac{\sqrt{n}(\langle \check{\theta}_r, u\rangle - \langle \theta_r(\Sigma), u\rangle)}{\sigma_r(\Sigma; u)} \right) - \mathbb{E}\ell(Z) \right| \overset{n\to\infty}{\to} 0.$$

REMARK 3.1. The assumption $\mathfrak{r}_n = o(n)$ is not necessary for the existence of a $\sqrt{n}$-consistent estimator of $\langle \theta_r(\Sigma), u\rangle$. In fact, the estimator $\langle \check{\theta}_r, u\rangle$ (say, with $m = n/4$) is $\sqrt{n}$-consistent provided that $\mathfrak{r}_n \le cn$ for a sufficiently small constant $c > 0$. This fact easily follows from (5.24) of Corollary 5.1 in Section 5. This is also the case for a somewhat simpler estimator (based on splitting the sample into two parts) considered earlier by Koltchinskii and Lounici [20] (see Proposition 3). However, it is not clear whether asymptotically efficient estimators (in the sense of Theorem 3.3) of linear functionals $\langle \theta_r(\Sigma), u\rangle$ of the eigenvector $\theta_r(\Sigma)$ with $\sqrt{n}$-rate and optimal limit variance $\sigma_r(\Sigma; u)$ exist when the condition $\mathfrak{r}_n = o(n)$ does not hold. In this case, the linear term of the perturbation series, that determines the limit variance $\sigma_r(\Sigma; u)$, is no longer dominant, which makes the existence of such estimators unlikely. However, asymptotically normal estimators of functionals $\langle \theta_r(\Sigma), u\rangle$ might still exist (but with a larger limit variance). It could be easier to develop such estimators in the case of spiked covariance models rather than in the more general framework of the current paper. The solution of this problem would rely on the tools of random matrix theory (see [28] as well as the more recent paper [5]) rather than perturbation theory, and, possibly, it would require the development of minimax lower bound techniques different from those employed in the present paper.

REMARK 3.2. It is not hard to develop similar asymptotically efficient estimators for l-dimensional "functionals" of the form $A\theta_r(\Sigma)$, where A is a linear operator from $\mathbb{H}$ into $\mathbb{R}^l$ for a fixed (small) dimension l. This is equivalent to the problem of estimation of $(\langle\theta_r(\Sigma), u_1\rangle, \dots, \langle\theta_r(\Sigma), u_l\rangle)$ for several linear functionals $u_1, \dots, u_l \in \mathbb{H}$. The bias reduction method developed in this paper can be extended to this case and the proof of asymptotic normality of the resulting estimators follows along the same lines as in the case when $l = 1$ with asymptotic covariance matrix equal to

$$\big(\mu_r\langle\Sigma C_r u_i, C_r u_j\rangle\big)_{i,j=1,\dots,p}.$$

Similarly, our approach can be extended to linear functionals of multiple eigenvectors of multiplicity 1 each, see, for example, $(\langle\theta_r(\Sigma), u\rangle, \langle\theta_s(\Sigma), v\rangle)$, $u, v \in \mathbb{H}$. In this case, the asymptotic covariance equals

$$-\frac{\mu_r\mu_s}{(\mu_r - \mu_s)^2}\langle\theta_r(\Sigma), v\rangle\langle\theta_s(\Sigma), u\rangle.$$

In this case, the debiasing strategy in Theorem 3.3 can be adjusted by using the second and third part oft the sample to estimate the bias for both $\theta_r(\Sigma)$ and $\theta_s(\Sigma)$.

However, note that when $\mathbf{r}(\Sigma)$ is large, the asymptotic normality of random vectors $n^{1/2}(\check{\theta}_r - \theta_r(\Sigma))$ holds only in the sense of finite-dimensional distributions, not in the sense of weak convergence in the Hilbert space $\mathbb{H}$ (indeed, the norm $\|\check{\theta}_r - \theta_r(\Sigma)\|$ is of order $\sqrt{\mathbf{r}(\Sigma)/n} \gg 1/\sqrt{n}$).

REMARK 3.3. Our method of bias reduction does not seem to have an easy extension to the problem of estimation of linear functionals of spectral projections P_r for an eigenvalue of multiplicity > 1. In part, this was a motivation for the first author to develop a more general approach to bias reduction (a so called "bootstrap chain" method) and to study the problem of efficient estimation for more general smooth functionals of covariance of the form $\langle f(\Sigma), B\rangle$, where f is a smooth function on the real line (see [18]). So far, the asymptotic efficiency for the resulting "bootstrap chain" estimators has been proved under more restrictive assumptions on the underlying covariance Σ. In particular, it was assumed that $\mathbb{H}$ is a space of finite (high) dimension p and that the spectrum of Σ is both upper and lower bounded away from 0 by constants which implies that $\mathbf{r}(\Sigma) \asymp p$.

REMARK 3.4. Lemma 5.3 of Section 5 provides explicit bounds on the accuracy of the normal approximation in Theorem 3.3. Using these bounds, it is possible to state somewhat more complicated conditions under which the normal approximation holds if $a = a_n \to \infty$ or $\sigma_0 = \sigma_0^{(n)} \to 0$ as $n \to \infty$. In particular, the normal approximation (3.3) still holds uniformly in $\mathcal{S}^{(r)}(\mathfrak{r}_n, a_n, \sigma_0^{(n)}, u)$ provided that $m_n = o(n)$ and

$$\frac{a_n^2}{\sigma_0^{(n)}}\left(\sqrt{\frac{n\mathfrak{r}_n}{m_n^2}}\log\frac{m_n^2}{n\mathfrak{r}_n} \vee \sqrt{\frac{n\log^2\frac{m_n^2}{n\mathfrak{r}_n}}{m_n^2}}\right) \to 0 \quad \text{as } n \to \infty.$$

Finally, we show that $\sigma_r(\Sigma; u)$ can be consistently estimated by $\sigma_r(\hat{\Sigma}; u)$, which allows us to replace the standard deviation $\sigma_r(\Sigma; u)$ in the normal approximation (3.3) by its empirical version. This yields the following result that can be used for hypotheses testing of linear functionals of θ_r. See Section 2 in the Supplementary Material [19] for its proof.

COROLLARY 3.2. *Under the conditions of Theorem 3.3,*

$$\sup_{\Sigma\in\mathcal{S}^{(r)}(\mathfrak{r}_n, a, \sigma_0, u)}\sup_{x\in\mathbb{R}}\left|\mathbb{P}_\Sigma\left\{\frac{\sqrt{n}(\langle\check{\theta}_r, u\rangle - \langle\theta_r(\Sigma), u\rangle)}{\sigma_r(\hat{\Sigma}; u)} \le x\right\} - \Phi(x)\right| \overset{n\to\infty}{\to} 0.$$

4. Proof of Theorem 3.1. We will prove the result for empirical eigenvectors $\hat{\theta}_r$ rather than for $\hat{\theta}_r^{\delta}$. The reduction to this case is based on Lemma 2.2 which immediately implies that

$$\sup_{\Sigma\in\mathcal{S}^{(r)}(\mathfrak{r}_n,a,\sigma_0,u)} \mathbb{P}_{\Sigma}\{\hat{\theta}_r^{\delta}\neq\hat{\theta}_r\}\le e^{-\beta\tau^2 n}.$$

Therefore, denoting

$$\xi_n(\Sigma):=\frac{\sqrt{n}(\langle\hat{\theta}_r^{\delta},u\rangle-\sqrt{1+b_r(\Sigma)}\langle\theta_r(\Sigma),u\rangle)}{\sigma_r(\Sigma;u)}$$

and

$$\eta_n(\Sigma):=\frac{\sqrt{n}(\langle\hat{\theta}_r,u\rangle-\sqrt{1+b_r(\Sigma)}\langle\theta_r(\Sigma),u\rangle)}{\sigma_r(\Sigma;u)},$$

we obtain

$$\sup_{\Sigma\in\mathcal{S}^{(r)}(\mathfrak{r}_n,a,\sigma_0,u)}\sup_{x\in\mathbb{R}}\big|\mathbb{P}_{\Sigma}\{\xi_n(\Sigma)\le x\}-\mathbb{P}_{\Sigma}\{\eta_n(\Sigma)\le x\}\big|\le e^{-\beta\tau^2 n}\overset{n\to\infty}{\to}0.$$

Also, since $\xi_n(\Sigma)\le\frac{2\sqrt{n}\|u\|}{\sigma_r(\Sigma;u)}$ and $\eta_n(\Sigma)\le\frac{2\sqrt{n}\|u\|}{\sigma_r(\Sigma;u)}$, we obtain that

$$\begin{aligned}&\sup_{\Sigma\in\mathcal{S}^{(r)}(\mathfrak{r}_n,a,\sigma_0,u)}\big|\mathbb{E}_{\Sigma}\ell(\xi_n(\Sigma))-\mathbb{E}_{\Sigma}\ell(\eta_n(\Sigma))\big|\\ &\quad\le\sup_{\Sigma\in\mathcal{S}^{(r)}(\mathfrak{r}_n,a,\sigma_0,u)}\mathbb{E}_{\Sigma}\big|\ell(\xi_n(\Sigma))-\ell(\eta_n(\Sigma))\big|I(\hat{\theta}_r^{\delta}\neq\hat{\theta}_r)\\ &\quad\le 2\ell\Big(\frac{2\sqrt{n}\|u\|}{\sigma_0}\Big)e^{-\beta\tau^2 n}\to 0,\end{aligned}$$

under Assumption 3.1.

We will prove more explicit bounds for the estimator $\hat{\theta}_r$ stated below in Lemma 4.8 that immediately implies the result.

Our starting point is the first order perturbation expansion of the empirical spectral projection operator $\hat{P}_r$:

$$\hat{P}_r=P_r+L_r(E)+S_r(E) \tag{4.1}$$

with a linear term $L_r(E)=P_rEC_r+C_rEP_r$ and a remainder $S_r(E)$, where $E:=\hat{\Sigma}-\Sigma$. It was proved in [20] that, under the assumption

$$\mathbb{E}\|\hat{\Sigma}-\Sigma\|\le\frac{(1-\gamma)g_r}{2} \tag{4.2}$$

for some $\gamma\in(0,1)$, the bilinear form of the remainder $S_r(E)$ satisfies the following concentration inequality: for all $u,v\in\mathbb{H}$ and for all $t\ge 1$ with probability at least $1-e^{-t}$

$$\big|\langle(S_r(E)-\mathbb{E}S_r(E))u,v\rangle\big|\lesssim_{\gamma}\frac{\|\Sigma\|^2}{g_r^2}\Big(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}}\vee\sqrt{\frac{t}{n}}\vee\frac{t}{n}\Big)\sqrt{\frac{t}{n}}\|u\|\|v\|. \tag{4.3}$$

Under the same assumption, it was also proved in [20] that the following representation holds for the bias $\mathbb{E}\hat{P}_r-P_r$ of empirical spectral projections $\hat{P}_r$:

$$\mathbb{E}\hat{P}_r-P_r=P_r(\mathbb{E}\hat{P}_r-P_r)P_r+T_r, \tag{4.4}$$

where the main term $P_r(\mathbb{E}\hat{P}_r-P_r)P_r$ is aligned with the spectral projection P_r and is of order

$$\|P_r(\mathbb{E}\hat{P}_r-P_r)P_r\|\lesssim\frac{\|\Sigma\|^2}{g_r^2}\frac{\mathbf{r}(\Sigma)}{n} \tag{4.5}$$

and the remainder T_r satisfies the bound

$$\|T_r\| \lesssim_\gamma \frac{m_r\|\Sigma\|^2}{g_r^2}\sqrt{\frac{\mathbf{r}(\Sigma)}{n}}\frac{1}{\sqrt{n}}. \tag{4.6}$$

Representation (4.4) is especially simple in the case when P_r is of rank 1 ($m_r = 1$), which also implies that $\hat{P}_r$ is of rank 1. In this case, $P_r = \theta_r \otimes \theta_r$, $\hat{P}_r = \hat{\theta}_r \otimes \hat{\theta}_r$ for unit eigenvectors $\theta_r, \hat{\theta}_r$ of covariance operators $\Sigma, \hat{\Sigma}$, respectively, and

$$P_r(\mathbb{E}\hat{P}_r - P_r)P_r = b_r P_r$$

for a "bias parameter" $b_r = b_r(\Sigma) = \mathbb{E}\langle\hat{\theta}_r, \theta_r\rangle^2 - 1 \in [-1, 0]$. Thus, it follows from (4.4) that

$$\mathbb{E}\hat{P}_r = (1 + b_r)P_r + T_r. \tag{4.7}$$

We obtain from (4.1) and (4.7) that

$$\hat{P}_r - (1 + b_r)P_r = L_r(E) + S_r(E) - \mathbb{E}S_r(E) + T_r. \tag{4.8}$$

Denote

$$\rho_r(u) := \big\langle\big(\hat{P}_r - (1 + b_r)P_r\big)\theta_r, u\big\rangle, \quad u \in \mathbb{H}.$$

As in [20], the function $\rho_r(u), u \in \mathbb{H}$ will be used in what follows to control the linear forms $\langle\hat{\theta}_r - \sqrt{1 + b_r}\theta_r, u\rangle, u \in \mathbb{H}$. First, we need to derive some bounds on $\rho_r(u)$.

The following lemma is an immediate consequence of (4.8), (4.3) and (4.6).

LEMMA 4.1. *Suppose condition* (4.2) *holds for some* $\gamma \in (0, 1)$. *Then, for all* $u \in \mathbb{H}$ *and for all* $t \geq 1$ *with probability at least* $1 - e^{-t}$

$$\big|\rho_r(u) - \big\langle L_r(E)\theta_r, u\big\rangle\big| \lesssim_\gamma \frac{\|\Sigma\|^2}{g_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}} \vee \frac{t}{n}\right)\sqrt{\frac{t}{n}}\|u\|. \tag{4.9}$$

We will need simple concentration and normal approximation bounds for $\langle L_r(E)\theta_r, u\rangle$ given in the next lemma.

LEMMA 4.2. *For all* $t \geq 1$ *with probability at least* $1 - e^{-t}$

$$\big|\big\langle L_r(E)\theta_r, u\big\rangle\big| \lesssim \sigma_r(\Sigma; u)\left(\sqrt{\frac{t}{n}} \vee \frac{t}{n}\right). \tag{4.10}$$

Moreover, if $\sigma_r(\Sigma; u) > 0$, *then*

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left\{\frac{\sqrt{n}\langle L_r(E)\theta_r, u\rangle}{\sigma_r(\Sigma; u)} \leq x\right\} - \Phi(x)\right| \lesssim \frac{1}{\sqrt{n}}, \tag{4.11}$$

where Φ *is the distribution function of standard normal r.v.*

PROOF. Without loss of generality, assume that the space $\mathbb{H}$ is finite-dimensional (the general case follows by a simple approximation argument). Since $L_r(E) = P_rEC_r + C_rEP_r$ and $C_r\theta_r = 0$, we have

$$\big\langle L_r(E)\theta_r, u\big\rangle = \langle C_rEP_r\theta_r, u\rangle = \langle E\theta_r, C_ru\rangle = \langle E, \theta_r \otimes C_ru\rangle.$$

Since E is self-adjoint, we obtain that

$$\big\langle L_r(E)\theta_r, u\big\rangle = \frac{1}{2}\langle E, \theta_r \otimes C_ru + C_ru \otimes \theta_r\rangle.$$

Let $Z, Z_1, \dots, Z_n$ be i.i.d. standard normal vectors in $\mathbb{H}$ such that $X_j = \Sigma^{1/2} Z_j$. Then

$$E = \Sigma^{1/2}\left(n^{-1}\sum_{j=1}^{n} Z_j \otimes Z_j - \mathbb{E}(Z \otimes Z)\right)\Sigma^{1/2}.$$

Defining

$$\begin{aligned} D &:= \frac{1}{2}\Sigma^{1/2}(\theta_r \otimes C_r u + C_r u \otimes \theta_r)\Sigma^{1/2} \\ &= \frac{1}{2}\big(\Sigma^{1/2}\theta_r \otimes \Sigma^{1/2} C_r u + \Sigma^{1/2} C_r u \otimes \Sigma^{1/2}\theta_r\big), \end{aligned}$$

we obtain that

$$\langle L_r(E)\theta_r, u\rangle = n^{-1}\sum_{j=1}^{n}\big(\langle DZ_j, Z_j\rangle - \mathbb{E}\langle DZ, Z\rangle\big).$$

Clearly, $\langle DZ, Z\rangle \stackrel{d}{=} \sum_k \lambda_k g_k^2$, where $\{\lambda_k\}$ are the eigenvalues of D and $\{g_k\}$ are i.i.d. standard normal r.v. It follows that

$$\mathbb{E}\langle DZ, Z\rangle = \operatorname{tr}(D) = 0$$

and

$$\operatorname{Var}\big(\langle DZ, Z\rangle\big) = 2\sum_k \lambda_k^2 = 2\|D\|_2^2 = \sigma_r^2(\Sigma; u).$$

We can now represent $\langle L_r(E)\theta_r, u\rangle$ as follows:

$$\langle L_r(E)\theta_r, u\rangle \stackrel{d}{=} n^{-1}\sum_{j=1}^{n}\sum_k \lambda_k\big(g_{k,j}^2 - 1\big),$$

where $\{g_{k,j}\}$ are i.i.d. standard normal r.v. Using standard exponential bounds for sums of independent ψ_1 r.v. (see, e.g., [35], Proposition 5.16 or Theorem 3.1.9 in [14]), we obtain that with probability at least $1 - e^{-t}$

$$\left|n^{-1}\sum_{j=1}^{n}\sum_k \lambda_k\big(g_{k,j}^2 - 1\big)\right| \lesssim \left(\sum_k \lambda_k^2\right)^{1/2}\sqrt{\frac{t}{n}} \vee \sup_k |\lambda_k|\frac{t}{n},$$

which implies that with the same probability

$$\big|\langle L_r(E)\theta_r, u\rangle\big| \lesssim \|D\|_2\sqrt{\frac{t}{n}} \vee \|D\|\frac{t}{n}.$$

Since $\|D\| \le \|D\|_2 = \frac{1}{2}\sigma_r^2(\Sigma; u)$, bound (4.10) follows.

To prove (4.11), we use the Berry–Esseen bound that implies

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left\{\frac{\sum_{j=1}^{n}\sum_k \lambda_k(g_{k,j}^2 - 1)}{\sqrt{n}(2\sum_k \lambda_k^2)^{1/2}} \le x\right\} - \Phi(x)\right| \lesssim \frac{\sum_k |\lambda_k|^3}{(\sum_k \lambda_k^2)^{3/2}}\frac{1}{\sqrt{n}},$$

and therefore

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left\{\frac{\sqrt{n}\langle L_r(E)\theta_r, u\rangle}{\sigma_r(\Sigma; u)} \le x\right\} - \Phi(x)\right| \lesssim \frac{\|D\|_3^3}{\|D\|_2^3}\frac{1}{\sqrt{n}} \lesssim \frac{\|D\|}{\|D\|_2}\frac{1}{\sqrt{n}} \lesssim \frac{1}{\sqrt{n}}. \qquad \square$$

The following bounds on $\rho_r(u)$ immediately follow from (4.9) and (4.10).

LEMMA 4.3. *Suppose condition* (4.2) *holds for some* $\gamma \in (0,1)$. *Then, for all* $u \in \mathbb{H}$ *and for all* $t \geq 1$ *with probability at least* $1 - e^{-t}$

$$|\rho_r(u)| \lesssim_\gamma \sigma_r(\Sigma; u)\biggl(\sqrt{\frac{t}{n}} \vee \frac{t}{n}\biggr) + \frac{\|\Sigma\|^2}{g_r^2}\biggl(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}}\biggr)\sqrt{\frac{t}{n}}\|u\|. \tag{4.12}$$

Moreover, with the same probability

$$|\rho_r(u)| \lesssim_\gamma \frac{\|\Sigma\|}{g_r}\sqrt{\frac{t}{n}}\|u\| + \frac{\|\Sigma\|^2}{g_r^2}\biggl(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}}\biggr)\sqrt{\frac{t}{n}}\|u\| \tag{4.13}$$

and, for $u = \theta_r$,

$$|\rho_r(\theta_r)| \lesssim_\gamma \frac{\|\Sigma\|^2}{g_r^2}\biggl(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}}\biggr)\sqrt{\frac{t}{n}}. \tag{4.14}$$

Note that we dropped the term $\frac{t}{n}$ in some of the expressions on the right hand side of the above bounds (compare with (4.9)). This term is dominated by $\sqrt{\frac{t}{n}}$ for $t \leq n$. Moreover, it follows from the definition of $\rho_r(u)$ that it is upper bounded by $2\|u\|$. Since $\frac{\|\Sigma\|}{g_r} \geq 1$, this easily implies that, for $t \geq n$, the right hand side of bound (4.13) (with a proper constant) is larger than $|\rho_r(u)|$. Bound (4.14) follows from (4.9) since $\langle L_r(E)\theta_r, \theta_r\rangle = 0$.

To study concentration and normal approximation of the linear form

$$\langle \hat{\theta}_r - \sqrt{1+b_r}\theta_r, u\rangle, \quad u \in \mathbb{H},$$

it remains to prove that it can be approximated by $\langle L_r(E)\theta_r, u\rangle$.

LEMMA 4.4. *Suppose that for some* $\gamma \in (0,1)$ *condition* (4.2) *holds and, in addition,*

$$1 + b_r \geq \gamma. \tag{4.15}$$

Then, for all $u \in \mathbb{H}$ *and for all* $t \geq 1$, *with probability at least* $1 - e^{-t}$

$$\begin{aligned} &\bigl|\langle \hat{\theta}_r - \sqrt{1+b_r}\theta_r, u\rangle - \langle L_r(E)\theta_r, u\rangle\bigr| \\ &\quad\lesssim_\gamma \frac{\|\Sigma\|^2}{g_r^2}\biggl(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}} \vee \frac{t}{n}\biggr)\sqrt{\frac{t}{n}}\|u\|. \end{aligned} \tag{4.16}$$

PROOF. We use the following representation obtained in [20] (see (6.7) in [20]), which holds provided that $\hat{\theta}_r$ and θ_r are properly aligned so that $\langle \hat{\theta}_r, \theta_r\rangle \geq 0$:

$$\begin{aligned} &\langle \hat{\theta}_r - \sqrt{1+b_r}\theta_r, u\rangle \\ &\quad= \frac{\rho_r(u)}{\sqrt{1+b_r+\rho_r(\theta_r)}} \\ &\qquad- \frac{\sqrt{1+b_r}}{\sqrt{1+b_r+\rho_r(\theta_r)}(\sqrt{1+b_r+\rho_r(\theta_r)} + \sqrt{1+b_r})}\rho_r(\theta_r)\langle \theta_r, u\rangle \end{aligned} \tag{4.17}$$

(it is clear from the proof given in [20] that $1 + b_r + \rho_r(\theta_r) \geq 0$). Denote

$$\nu_r := \frac{\rho_r(\theta_r)}{1+b_r}.$$

Then, it is easy to see that

$$
\begin{aligned}
\langle \hat{\theta}_r - \sqrt{1+b_r}\theta_r, u\rangle = \rho_r(u) - \frac{b_r/(1+b_r)+\nu_r}{1+\nu_r+\sqrt{(1+\nu_r)/(1+b_r)}}\rho_r(u) \\
- \frac{\nu_r\sqrt{1+b_r}}{1+\nu_r+\sqrt{1+\nu_r}}\langle \theta_r, u\rangle.
\end{aligned}
\tag{4.18}
$$

Recall that (4.2) and (4.15) hold for some $\gamma \in (0, 1)$. If $|\nu_r| \le 1/2$, then (4.18) easily implies that

$$
\bigl|\langle \hat{\theta}_r - \sqrt{1+b_r}\theta_r, u\rangle - \rho_r(u)\bigr| \le \frac{1}{\gamma}\bigl(|b_r|+|\nu_r|\bigr)\bigl|\rho_r(u)\bigr| + |\nu_r|\bigl|\langle\theta_r, u\rangle\bigr|. \tag{4.19}
$$

It also follows from (4.14) that, under condition (4.15),

$$
|\nu_r| \lesssim_\gamma \frac{\|\Sigma\|^2}{g_r^2}\biggl(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}}\biggr)\sqrt{\frac{t}{n}} \tag{4.20}
$$

with probability at least $1 - e^{-t}$. On the other hand, bound (4.5) implies that

$$
|b_r| \lesssim \frac{\|\Sigma\|^2}{g_r^2}\frac{\mathbf{r}(\Sigma)}{n}. \tag{4.21}
$$

It follows from (4.20) that for the condition $|\nu_r| \le 1/2$ to hold with probability at least $1-e^{-t}$, it is enough to have

$$
\frac{\|\Sigma\|^2}{g_r^2}\biggl(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}}\biggr)\sqrt{\frac{t}{n}} \le c_\gamma \tag{4.22}
$$

for a small enough constant $c_\gamma > 0$. Assume that (4.22) holds. Note also that it implies that $t \lesssim n$ and condition (4.2) and Theorem 2.1 imply that $\frac{\|\Sigma\|}{g_r}\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \lesssim 1$. It follows from (4.19), (4.13), (4.20) and (4.21) that with probability at least $1 - 3e^{-t}$:

$$
\begin{aligned}
&\bigl|\langle \hat{\theta}_r - \sqrt{1+b_r}\theta_r, u\rangle - \rho_r(u)\bigr| \\
&\quad \lesssim_\gamma \biggl[\frac{\|\Sigma\|^2}{g_r^2}\frac{\mathbf{r}(\Sigma)}{n} + \biggl(\frac{\|\Sigma\|^2}{g_r^2}\biggl(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}}\biggr)\sqrt{\frac{t}{n}}\biggr) \wedge 1\biggr] \\
&\qquad \times \biggl[\frac{\|\Sigma\|}{g_r}\sqrt{\frac{t}{n}}\|u\| + \frac{\|\Sigma\|^2}{g_r^2}\biggl(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}}\biggr)\sqrt{\frac{t}{n}}\|u\|\biggr] \\
&\qquad + \frac{\|\Sigma\|^2}{g_r^2}\biggl(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}}\biggr)\sqrt{\frac{t}{n}}\|u\|.
\end{aligned}
\tag{4.23}
$$

Using the facts that

$$
\frac{\|\Sigma\|^2}{g_r^2}\frac{\mathbf{r}(\Sigma)}{n} \lesssim \frac{\|\Sigma\|}{g_r}\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \lesssim 1,
$$

that

$$
\frac{\|\Sigma\|^2}{g_r^2}\frac{t}{n} \lesssim \frac{\|\Sigma\|}{g_r}\sqrt{\frac{t}{n}} \lesssim 1
$$

and that

$$
\frac{\|\Sigma\|^2}{g_r^2}\sqrt{\frac{\mathbf{r}(\Sigma)}{n}}\sqrt{\frac{t}{n}} \lesssim \frac{\|\Sigma\|}{g_r}\biggl(\frac{\mathbf{r}(\Sigma)}{n}\biggr)^{1/4}\biggl(\frac{t}{n}\biggr)^{1/4} \le \frac{\|\Sigma\|}{g_r}\biggl(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}}\biggr)
$$

(that follow from condition (4.22)), we conclude that the last term

$$\frac{\|\Sigma\|^2}{g_r^2}\Big(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}}\Big)\sqrt{\frac{t}{n}}\|u\|$$

in the right-hand side of bound (4.23) is dominant. Hence, with probability at least $1-e^{-t}$

$$\big|\langle\hat{\theta}_r - \sqrt{1+b_r}\theta_r, u\rangle - \rho_r(u)\big| \lesssim_\gamma \frac{\|\Sigma\|^2}{g_r^2}\Big(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}}\Big)\sqrt{\frac{t}{n}}\|u\| \tag{4.24}$$

provided that condition (4.22) holds. On the other hand, if

$$\frac{\|\Sigma\|^2}{g_r^2}\Big(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}}\Big)\sqrt{\frac{t}{n}} > c_\gamma,$$

then

$$\begin{aligned}
&\big|\langle\hat{\theta}_r - \sqrt{1+b_r}\theta_r, u\rangle - \rho_r(u)\big| \\
&\quad\le \big|\langle\hat{\theta}_r - \sqrt{1+b_r}\theta_r, u\rangle\big| + \big|\rho_r(u)\big| \\
&\quad\le \big(\|\hat{\theta}_r\| + \sqrt{1+b_r}\|\theta_r\|\big)\|u\| + \big(\|\hat{P}_r\| + (1+b_r)\|P_r\|\big)\|\theta_r\|\|u\| \\
&\quad\le 4\|u\| \\
&\quad\lesssim_\gamma \frac{\|\Sigma\|^2}{g_r^2}\Big(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}}\Big)\sqrt{\frac{t}{n}}\|u\|.
\end{aligned}$$

Thus, we proved that with probability at least $1-e^{-t}$

$$\big|\langle\hat{\theta}_r - \sqrt{1+b_r}\theta_r, u\rangle - \rho_r(u)\big| \lesssim_\gamma \frac{\|\Sigma\|^2}{g_r^2}\Big(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}}\Big)\sqrt{\frac{t}{n}}\|u\|. \tag{4.25}$$

It remains to combine this with the bound (4.9) to complete the proof. □

The following result is a slightly improved version of Theorem 6 in [20].

LEMMA 4.5. *Under conditions* (4.2) *and* (4.15) *for some* $\gamma \in (0,1)$, *the following bounds hold for all* $t \ge 1$ *with probability at least* $1-e^{-t}$:

$$\big|\langle\hat{\theta}_r - \sqrt{1+b_r}\theta_r, u\rangle\big| \lesssim_\gamma \frac{\|\Sigma\|}{g_r}\sqrt{\frac{t}{n}}\|u\| \tag{4.26}$$

and

$$\big|\langle\hat{\theta}_r - \sqrt{1+b_r}\theta_r, \theta_r\rangle\big| \lesssim_\gamma \frac{\|\Sigma\|^2}{g_r^2}\Big(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}}\Big)\sqrt{\frac{t}{n}}. \tag{4.27}$$

PROOF. Indeed, it follows from (4.16) and (4.10) that, for some constants $C, C_\gamma > 0$ with probability at least $1-e^{-t}$

$$\begin{aligned}
&\big|\langle\hat{\theta}_r - \sqrt{1+b_r}\theta_r, u\rangle\big| \\
&\qquad\le C\sigma_r(\Sigma;u)\Big(\sqrt{\frac{t}{n}} \vee \frac{t}{n}\Big) + C_\gamma\frac{\|\Sigma\|^2}{g_r^2}\Big(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}}\Big)\sqrt{\frac{t}{n}}\|u\|.
\end{aligned}$$

Since $\sigma_r(\Sigma; u) \lesssim \frac{\|\Sigma\|}{g_r}\|u\|$, with the same probability

$$\left|\langle \hat{\theta}_r - \sqrt{1+b_r}\theta_r, u\rangle\right| \le C\frac{\|\Sigma\|}{g_r}\sqrt{\frac{t}{n}}\|u\| + C_\gamma \frac{\|\Sigma\|^2}{g_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}}\right)\sqrt{\frac{t}{n}}\|u\|.$$

We dropped the term $\frac{t}{n}$ present in bounds (4.16) and (4.10) since for $t \ge n$ (the only case when it is needed), the right-hand side already dominates the left hand side (which is smaller than $2\|u\|$). Note that condition (4.2) and Theorem 2.1 imply that $\frac{\|\Sigma\|}{g_r}\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \le c_\gamma$ for some constant $c_\gamma > 0$. Assuming that also $\frac{\|\Sigma\|}{g_r}\sqrt{\frac{t}{n}} \le c_\gamma$, which implies that $t \lesssim n$, we obtain that for some constant $C_\gamma > 0$ with probability at least $1 - e^{-t}$ bound (4.26) holds. On the other hand, if $\frac{\|\Sigma\|}{g_r}\sqrt{\frac{t}{n}} > c_\gamma$, then

$$\left|\langle \hat{\theta}_r - \sqrt{1+b_r}\theta_r, u\rangle\right| \le \left(\|\hat{\theta}_r\| + \sqrt{1+b_r}\|\theta_r\|\right)\|u\| \le 2\|u\| \lesssim_\gamma \frac{\|\Sigma\|}{g_r}\sqrt{\frac{t}{n}}\|u\|,$$

implying again (4.26). For $u = \theta_r$, $\langle L_r(E)\theta_r, u\rangle = 0$ and bound (4.16) implies that with probability at least $1 - e^{-t}$ (4.27) holds. □

The following two lemmas will be used to derive normal approximation bounds for $\langle \hat{\theta}_r - \sqrt{1+b_r}\theta_r, u\rangle$ from the corresponding bounds for $\langle L_r(E)\theta_r, u\rangle$ as well as to control the risk for loss functions satisfying Assumption 3.1. We state them without proofs (which are elementary).

LEMMA 4.6. *For random variables ξ, η, denote*

$$\Delta(\xi; \eta) := \sup_{x\in\mathbb{R}}\left|\mathbb{P}\{\xi \le x\} - \mathbb{P}\{\eta \le x\}\right|$$

and

$$\delta(\xi; \eta) := \inf\left\{\delta > 0 : \mathbb{P}\left\{|\xi - \eta| \ge \delta\right\} + \delta\right\}.$$

Then, for a standard normal r.v. Z,

$$\Delta(\xi; Z) \le \Delta(\eta; Z) + \delta(\xi; \eta).$$

Under Assumption 3.1, for all $A > 0$

$$\left|\mathbb{E}\ell(\xi) - \mathbb{E}\ell(\eta)\right| \le 4\ell(A)\Delta(\xi;\eta) + \mathbb{E}\ell(\xi)I\left(|\xi| \ge A\right) + \mathbb{E}\ell(\eta)I\left(|\eta| \ge A\right).$$

LEMMA 4.7. *Let ξ be a random variable such that for some $\tau_1 \ge 0$ and $\tau_2 \ge 0$ and for all $t \ge 1$ with probability at least $1 - e^{-t}$*

$$|\xi| \le \tau_1\sqrt{t} \vee \tau_2 t.$$

Let ℓ be a loss function satisfying Assumption 3.1. If $2c_2\tau_2 < 1$, then

$$\mathbb{E}\ell^2(\xi) \le 2e\sqrt{2\pi}c_1^2 e^{2c_2^2\tau_1^2} + \frac{ec_1^2}{1 - 2c_2\tau_2}. \tag{4.28}$$

Next we prove the normal approximation bounds for linear forms $\langle \hat{\theta}_r - \sqrt{1+b_r}\theta_r, u\rangle$.

LEMMA 4.8. *Suppose that conditions* (4.2) *and* (4.15) *hold for some* $\gamma \in (0,1)$ *and also that* $n \ge 2\mathbf{r}(\Sigma)$. *Assume that, for some* $u \in \mathbb{H}$, $\sigma_r(\Sigma; u) > 0$. *Let* $\alpha \ge 1$. *Then the following bound holds*: *for some constants* $C, C_{\gamma,\alpha} > 0$,

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left\{\frac{\sqrt{n}\langle\hat{\theta}_r - \sqrt{1+b_r}\theta_r, u\rangle}{\sigma_r(\Sigma; u)} \le x\right\} - \Phi(x)\right|$$

$$\le Cn^{-1/2} + \frac{C_{\gamma,\alpha}}{\sigma_r(\Sigma;u)}\frac{\|\Sigma\|^2}{g_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}\log\frac{n}{\mathbf{r}(\Sigma)}} \vee \frac{\log\frac{n}{\mathbf{r}(\Sigma)}}{\sqrt{n}}\right)\|u\| + \left(\frac{\mathbf{r}(\Sigma)}{n}\right)^{\alpha}. \tag{4.29}$$

Moreover, under Assumption 3.1 *on the loss* ℓ, *there exist constants* $C, C_\gamma, C_{\gamma,\alpha} > 0$ *such that*

$$\left|\mathbb{E}\ell\left(\frac{\sqrt{n}\langle\hat{\theta}_r - \sqrt{1+b_r}\theta_r, u\rangle}{\sigma_r(\Sigma;u)}\right) - \mathbb{E}\ell(Z)\right|$$

$$\le c_1 e^{c_2 A}\Bigg[Cn^{-1/2} + \frac{C_{\gamma,\alpha}}{\sigma_r(\Sigma;u)}\frac{\|\Sigma\|^2}{g_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}\log\frac{n}{\mathbf{r}(\Sigma)}} \vee \frac{\log\frac{n}{\mathbf{r}(\Sigma)}}{\sqrt{n}}\right)\|u\| + \left(\frac{\mathbf{r}(\Sigma)}{n}\right)^{\alpha}\Bigg] + 2e^{3/2}(2\pi)^{1/4}c_1 e^{c_2^2\tau^2}e^{-A^2/2\tau^2} + c_1 e^{c_2^2}e^{-A^2/4}, \tag{4.30}$$

where

$$\tau := C_\gamma \frac{\|\Sigma\|\|u\|}{g_r\sigma_r(\Sigma;u)}.$$

PROOF. We will use the first claim of Lemma 4.6 with

$$\xi := \frac{\sqrt{n}\langle\hat{\theta}_r - \sqrt{1+b_r}\theta_r, u\rangle}{\sigma_r(\Sigma;u)} \quad \text{and} \quad \eta := \frac{\sqrt{n}\langle L_r(E)\theta_r, u\rangle}{\sigma_r(\Sigma;u)}.$$

It follows from bound (4.16) that, under conditions (4.2) and (4.15), for some $C_\gamma > 0$

$$\delta(\xi;\eta) \le \inf_{t\ge 1}\left\{\frac{C_\gamma}{\sigma_r(\Sigma;u)}\frac{\|\Sigma\|^2}{g_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \vee \sqrt{\frac{t}{n}} \vee \frac{t}{n}\right)\sqrt{t}\|u\| + e^{-t}\right\}.$$

Taking $t := \alpha \log\frac{n}{\mathbf{r}(\Sigma)}$ with some $\alpha \ge 1$ easily yields an upper bound

$$\delta(\xi;\eta) \le \frac{C_{\gamma,\alpha}}{\sigma_r(\Sigma;u)}\frac{\|\Sigma\|^2}{g_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}\log\frac{n}{\mathbf{r}(\Sigma)}} \vee \frac{\log\frac{n}{\mathbf{r}(\Sigma)}}{\sqrt{n}}\right)\|u\| + \left(\frac{\mathbf{r}(\Sigma)}{n}\right)^{\alpha}.$$

Using bound (4.11) to control $\Delta(\eta; Z)$, we obtain from Lemma 4.6 that bound (4.29) holds with some constants $C, C_{\gamma,\alpha} > 0$. To prove the second statement, we use the second bound of Lemma 4.6 with the random variable $\xi := \frac{\sqrt{n}\langle\hat{\theta}_r - \sqrt{1+b_r}\theta_r, u\rangle}{\sigma_r(\Sigma;u)}$ and $\eta = Z$. The following exponential bound on ξ is an easy corollary of bound (4.26): for some constant $C_\gamma > 0$ and for all $t \ge 1$ with probability at least $1 - e^{-t}$

$$|\xi| \le C_\gamma \frac{\|\Sigma\|}{g_r\sigma_r(\Sigma;u)}\sqrt{t}\|u\| = \tau\sqrt{t}. \tag{4.31}$$

Using bound (4.28) with $\tau_1 = \tau$ and $\tau_2 = 0$, we obtain

$$\mathbb{E}\ell^2(\xi) \le 2e\sqrt{2\pi}c_1^2 e^{2c_2^2\tau_1^2} + ec_1^2 \le 4e\sqrt{2\pi}c_1^2 e^{2c_2^2\tau_1^2}.$$

Therefore,

$$\mathbb{E}\ell(\xi)I(|\xi| \ge A) \le \mathbb{E}^{1/2}\ell^2(\xi)\mathbb{P}^{1/2}\{|\xi| \ge A\} \le 2e^{3/2}(2\pi)^{1/4}c_1 e^{c_2^2\tau^2}e^{-A^2/2\tau^2}.$$

We also have

$$\mathbb{E}\ell(Z)I(|Z| \ge A) \le c_1 e^{c_2^2}e^{-A^2/4}.$$

Using bound (4.29), we can now deduce bound (4.30) from the second statement of Lemma 4.6. □

Lemma 4.8 immediately implies Theorem 3.1 (by passing to the limit as $n \to \infty$ in (4.29) and as $n \to \infty$ and then $A \to \infty$ in (4.30)).

5. Proof of Theorem 3.3. Recall that the estimator $\check{\theta}_r$ is based on empirical eigenvectors $\hat{\theta}_r^{\delta_j,j}$, $j = 1, 2, 3$ with parameters $\delta_j = \tau\|\hat{\Sigma}^{(j)}\|$ and with a proper choice of τ (as in Lemma 2.2). These eigenvectors are in turn defined in terms of empirical spectral projections $\hat{P}_r^{\delta_j,j}$ of sample covariances $\hat{\Sigma}^{(j)}$ (based on δ_j-clusters of its spectrum $\sigma(\hat{\Sigma}^{(j)})$). We will, however, replace $\check{\theta}_r$ by the estimator $\tilde{\theta}_r$ defined in terms of empirical spectral projections $\hat{P}_r^{(j)}$, $j = 1, 2, 3$, $\hat{P}_r^{(j)}$ being the orthogonal projection onto direct sum of eigenspaces of $\hat{\Sigma}^{(j)}$ corresponding to its eigenvalues $\lambda_k(\hat{\Sigma}^{(j)})$, $k \in \Delta_r$. Since $\operatorname{card}(\Delta_r) = m_r = 1$, $\hat{P}_r^{(j)} = \hat{\theta}_r^{(j)} \otimes \hat{\theta}_r^{(j)}$ and we can define

$$\hat{d}_r := \frac{\langle\hat{\theta}_r^{(1)}, \hat{\theta}_r^{(2)}\rangle}{\langle\hat{\theta}_r^{(2)}, \hat{\theta}_r^{(3)}\rangle^{1/2}} \quad \text{and} \quad \tilde{\theta}_r := \frac{\hat{\theta}_r^{(1)}}{\hat{d}_r \vee (1/2)}.$$

The reduction to this case is based on Lemma 2.2 (implying that $\hat{P}_r^{\delta_j,j} = \hat{P}_r^{(j)}$ with a high probability) and is straightforward (as in the proof of Theorem 3.1).

The rest of the proof is based on several lemmas stated and proved below.

LEMMA 5.1. *Suppose that for some $\gamma \in (0, 1)$ condition (4.2) holds for the sample covariance $\hat{\Sigma}^{(2)}$ based on m observations*:

$$\mathbb{E}\|\hat{\Sigma}^{(2)} - \Sigma\| \le \frac{(1-\gamma)g_r}{2}. \tag{5.1}$$

Then, for all $t \ge 1$ with probability at least $1 - e^{-t}$

$$\left|\langle\hat{\theta}_r^{(1)}, \hat{\theta}_r^{(2)}\rangle - \sqrt{1 + b_r^{(n')}}\sqrt{1 + b_r^{(m)}}\right| \lesssim_\gamma \frac{\|\Sigma\|^2}{g_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{m}} \vee \sqrt{\frac{t}{m}}\right)\sqrt{\frac{t}{m}} \tag{5.2}$$

and with the same probability

$$|\langle\hat{\theta}_r^{(2)}, \hat{\theta}_r^{(3)}\rangle - (1 + b_r^{(m)})| \lesssim_\gamma \frac{\|\Sigma\|^2}{g_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{m}} \vee \sqrt{\frac{t}{m}}\right)\sqrt{\frac{t}{m}}. \tag{5.3}$$

PROOF. Obviously, condition (5.1) holds also for the sample covariance $\hat{\Sigma}^{(2)}$ (which is based on a sample of the same size m). Moreover, it also holds for the sample covariance $\hat{\Sigma}^{(1)}$ based on $n' \ge m$ observations since the sequence $n \mapsto \mathbb{E}\|\hat{\Sigma}_n - \Sigma\|$ is nonincreasing (see, e.g., Lemma 2.4.5 in [34]).

The following representation is obvious:

$$
\begin{aligned}
\langle\hat{\theta}_r^{(1)}, \hat{\theta}_r^{(2)}\rangle = {} & \sqrt{1+b_r^{(n')}}\sqrt{1+b_r^{(m)}}\langle\theta_r, \theta_r\rangle \\
& + \sqrt{1+b_r^{(m)}}\Big\langle\hat{\theta}_r^{(1)} - \sqrt{1+b_r^{(n')}}\theta_r, \theta_r\Big\rangle \\
& + \sqrt{1+b_r^{(n')}}\Big\langle\hat{\theta}_r^{(2)} - \sqrt{1+b_r^{(m)}}\theta_r, \theta_r\Big\rangle \\
& \times \Big\langle\hat{\theta}_r^{(1)} - \sqrt{1+b_r^{(n')}}\theta_r, \hat{\theta}_r^{(2)} - \sqrt{1+b_r^{(m)}}\theta_r\Big\rangle.
\end{aligned}
\tag{5.4}
$$

By bound (4.27), with probability at least $1-e^{-t}$

$$
\Big|\Big\langle\hat{\theta}_r^{(1)} - \sqrt{1+b_r^{(n')}}\theta_r, \theta_r\Big\rangle\Big| \lesssim_\gamma \frac{\|\Sigma\|^2}{g_r^2}\bigg(\sqrt{\frac{\mathbf{r}(\Sigma)}{n'}} \vee \sqrt{\frac{t}{n'}}\bigg)\sqrt{\frac{t}{n'}}. \tag{5.5}
$$

Similarly, with probability at least $1-e^{-t}$

$$
\Big|\Big\langle\hat{\theta}_r^{(2)} - \sqrt{1+b_r^{(m)}}\theta_r, \theta_r\Big\rangle\Big| \lesssim_\gamma \frac{\|\Sigma\|^2}{g_r^2}\bigg(\sqrt{\frac{\mathbf{r}(\Sigma)}{m}} \vee \sqrt{\frac{t}{m}}\bigg)\sqrt{\frac{t}{m}}. \tag{5.6}
$$

To bound the last term in the right-hand side of (5.4), we apply bound (4.26) to $\hat{\theta}_r^{(1)}$ conditionally on the second sample (similarly to the proof of Theorem 6 in [20]). This yields that with probability at least $1-e^{-t}$

$$
\begin{aligned}
& \Big|\Big\langle\hat{\theta}_r^{(1)} - \sqrt{1+b_r^{(n')}}\theta_r, \hat{\theta}_r^{(2)} - \sqrt{1+b_r^{(m)}}\theta_r\Big\rangle\Big| \\
& \quad \lesssim_\gamma \frac{\|\Sigma\|}{g_r}\sqrt{\frac{t}{n'}}\Big\|\hat{\theta}_r^{(2)} - \sqrt{1+b_r^{(m)}}\theta_r\Big\|.
\end{aligned}
\tag{5.7}
$$

On the other hand, under the assumption that $\langle\hat{\theta}_r, \theta_r\rangle \geq 0$,

$$
\begin{aligned}
\Big\|\hat{\theta}_r^{(2)} - \sqrt{1+b_r^{(m)}}\theta_r\Big\| & \leq \|\hat{\theta}_r^{(2)} - \theta_r\| + \Big|\sqrt{1+b_r^{(m)}} - 1\Big| \\
& = \sqrt{2 - 2\langle\hat{\theta}_r^{(2)}, \theta_r\rangle} + \frac{|b_r^{(m)}|}{\sqrt{1+b_r^{(m)}}+1} \\
& \leq \sqrt{2 - 2\langle\hat{\theta}_r^{(2)}, \theta_r\rangle^2} + |b_r^{(m)}| \\
& = \sqrt{2 - 2\langle\hat{P}_r^{(2)}, P_r\rangle} + |b_r^{(m)}| \\
& = \|\hat{P}_r^{(2)} - P_r\|_2 + |b_r^{(m)}|. \\
& \leq \sqrt{2}\|\hat{P}_r^{(2)} - P_r\| + |b_r^{(m)}|.
\end{aligned}
$$

By a standard perturbation bound (see, e.g., [20]),

$$
\|\hat{P}_r^{(2)} - P_r\| \leq 4\frac{\|\hat{\Sigma}^{(2)} - \Sigma\|}{g_r}.
$$

Thus,

$$
\Big\|\hat{\theta}_r^{(2)} - \sqrt{1+b_r^{(m)}}\theta_r\Big\| \leq 4\sqrt{2}\frac{\|\hat{\Sigma}^{(2)} - \Sigma\|}{g_r} + |b_r^{(m)}|. \tag{5.8}
$$

Using the exponential bound (2.3) on $\|\hat{\Sigma}^{(2)} - \Sigma\|$ and bound (4.21), we obtain that with probability at least $1 - e^{-t}$

$$\begin{aligned}&\left\|\hat{\theta}_r^{(2)} - \sqrt{1 + b_r^{(m)}}\theta_r\right\| \\ &\quad \lesssim \frac{\|\Sigma\|}{g_r}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{m}} \vee \frac{\mathbf{r}(\Sigma)}{m} \vee \sqrt{\frac{t}{m}} \vee \frac{t}{m}\right) + \frac{\|\Sigma\|^2}{g_r^2}\frac{\mathbf{r}(\Sigma)}{m}.\end{aligned} \tag{5.9}$$

Under assumption (5.1), we have $\frac{\|\Sigma\|}{g_r}\sqrt{\frac{\mathbf{r}(\Sigma)}{m}} \lesssim 1$, which implies $\frac{\|\Sigma\|^2}{g_r^2}\frac{\mathbf{r}(\Sigma)}{m} \lesssim \frac{\|\Sigma\|}{g_r}\sqrt{\frac{\mathbf{r}(\Sigma)}{m}}$. Thus, the first term in the right-hand side of bound (5.9) is dominant. Moreover, we can drop the term $\frac{\mathbf{r}(\Sigma)}{m}$ and, for $t \le m$, we can also drop the term $\frac{\|\Sigma\|}{g_r}\frac{t}{m}$ in the right-hand side. Since the left-hand side of (5.9) is not larger than 2, for $t > m$, the term $\frac{\|\Sigma\|}{g_r}\sqrt{\frac{t}{m}}$ is larger (up to a constant) than the left hand side. Thus, the term $\frac{\|\Sigma\|}{g_r}\frac{t}{m}$ can be dropped for all the values of t and the bound (5.9) simplifies as follows

$$\left\|\hat{\theta}_r^{(2)} - \sqrt{1 + b_r^{(m)}}\theta_r\right\| \lesssim \frac{\|\Sigma\|}{g_r}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{m}} \vee \sqrt{\frac{t}{m}}\right) \tag{5.10}$$

and it still holds with probability at least $1 - e^{-t}$. It follows from bound (5.7) and (5.10) that for all $t \ge 1$ with probability at least $1 - 2e^{-t}$

$$\begin{aligned}&\left|\left\langle\hat{\theta}_r^{(1)} - \sqrt{1 + b_r^{(n')}}\theta_r, \hat{\theta}_r^{(2)} - \sqrt{1 + b_r^{(m)}}\theta_r\right\rangle\right| \\ &\quad \lesssim_\gamma \frac{\|\Sigma\|^2}{g_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{m}} \vee \sqrt{\frac{t}{m}}\right)\sqrt{\frac{t}{n'}}.\end{aligned} \tag{5.11}$$

Taking into account that $n' \ge m$, it easily follows from representation (5.4) and bounds (5.5), (5.6) and (5.11) that with probability at least $1 - e^{-t}$

$$\left|\langle\hat{\theta}_r^{(1)}, \hat{\theta}_r^{(2)}\rangle - \sqrt{1 + b_r^{(n')}}\sqrt{1 + b_r^{(m)}}\right| \lesssim_\gamma \frac{\|\Sigma\|^2}{g_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{m}} \vee \sqrt{\frac{t}{m}}\right)\sqrt{\frac{t}{m}},$$

which proves (5.2). The proof of bound (5.3) is similar. □

Define

$$\Delta_1 := \frac{\langle\hat{\theta}_r^{(1)}, \hat{\theta}_r^{(2)}\rangle}{\sqrt{1 + b_r^{(n')}}\sqrt{1 + b_r^{(m)}}} - 1$$

and

$$\Delta_2 := \frac{\langle\hat{\theta}_r^{(2)}, \hat{\theta}_r^{(3)}\rangle}{1 + b_r^{(m)}} - 1.$$

Assuming that

$$1 + b_r^{(n')} \ge (3/4)^2 \quad \text{and} \quad 1 + b_r^{(m)} \ge (3/4)^2, \tag{5.12}$$

we obtain that, for some constant $C_\gamma > 0$ and for $t \ge 1$ on an event E of probability at least $1 - e^{-t}$

$$|\Delta_1| \vee |\Delta_2| \le C_\gamma \frac{\|\Sigma\|^2}{g_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{m}} \vee \sqrt{\frac{t}{m}}\right)\sqrt{\frac{t}{m}}. \tag{5.13}$$

Next, we have

$$\begin{aligned}\hat{d}_r &= \frac{\langle \hat{\theta}_r^{(1)}, \hat{\theta}_r^{(2)} \rangle}{\langle \hat{\theta}_r^{(2)}, \hat{\theta}_r^{(3)} \rangle^{1/2}} \\ &= \frac{\langle \hat{\theta}_r^{(1)}, \hat{\theta}_r^{(2)} \rangle / ((1+b_r^{(n')})^{1/2}(1+b_r^{(m)})^{1/2})}{\langle \hat{\theta}_r^{(2)}, \hat{\theta}_r^{(3)} \rangle^{1/2} / (1+b_r^{(m)})^{1/2}} \sqrt{1+b_r^{(n')}} \\ &= \frac{1+\Delta_1}{\sqrt{1+\Delta_2}} \sqrt{1+b_r^{(n')}} \\ &= \sqrt{1+b_r^{(n')}} + \frac{1+\Delta_1 - \sqrt{1+\Delta_2}}{\sqrt{1+\Delta_2}} \sqrt{1+b_r^{(n')}},\end{aligned}$$

which implies

$$\begin{aligned}\Big|\hat{d}_r - \sqrt{1+b_r^{(n')}}\Big| &\le \sqrt{1+b_r^{(n')}} \frac{|(1+\Delta_1)^2 - (1+\Delta_2)|}{\sqrt{1+\Delta_2}(1+\Delta_1+\sqrt{1+\Delta_2})} \\ &\le \frac{2|\Delta_1| + \Delta_1^2 + |\Delta_2|}{\sqrt{1+\Delta_2}(1+\Delta_1+\sqrt{1+\Delta_2})}.\end{aligned} \tag{5.14}$$

Under the assumption that

$$\frac{\|\Sigma\|^2}{g_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{m}} \vee \sqrt{\frac{t}{m}}\right)\sqrt{\frac{t}{m}} \le c_\gamma \tag{5.15}$$

for a sufficiently small constant $c_\gamma > 0$, bounds (5.14) and (5.13) imply that on the event E

$$\left|\frac{\hat{d}_r}{\sqrt{1+b_r^{(n')}}} - 1\right| \lesssim_\gamma \frac{\|\Sigma\|^2}{g_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{m}} \vee \sqrt{\frac{t}{m}}\right)\sqrt{\frac{t}{m}}. \tag{5.16}$$

Moreover, on the same event E,

$$\begin{aligned}\hat{d}_r &\ge \sqrt{1+b_r^{(n')}} - \frac{2|\Delta_1| + \Delta_1^2 + |\Delta_2|}{\sqrt{1+\Delta_2}(1+\Delta_1+\sqrt{1+\Delta_2})} \\ &\ge \frac{3}{4} - \frac{2|\Delta_1| + \Delta_1^2 + |\Delta_2|}{\sqrt{1+\Delta_2}(1+\Delta_1+\sqrt{1+\Delta_2})} \ge \frac{1}{2},\end{aligned} \tag{5.17}$$

$$\left|\frac{\sqrt{1+b_r^{(n')}}}{\hat{d}_r} - 1\right| \lesssim_\gamma \frac{\|\Sigma\|^2}{g_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{m}} \vee \sqrt{\frac{t}{m}}\right)\sqrt{\frac{t}{m}} \tag{5.18}$$

and also, using bound (4.21), we obtain that

$$\begin{aligned}|\hat{d}_r - 1| &\le \Big|\sqrt{1+b_r^{(n')}} - 1\Big| + \frac{2|\Delta_1| + \Delta_1^2 + |\Delta_2|}{\sqrt{1+\Delta_2}(1+\Delta_1+\sqrt{1+\Delta_2})} \\ &\le |b_r^{(n')}| + \frac{2|\Delta_1| + \Delta_1^2 + |\Delta_2|}{\sqrt{1+\Delta_2}(1+\Delta_1+\sqrt{1+\Delta_2})} \\ &\lesssim_\gamma \frac{\|\Sigma\|^2}{g_r^2}\frac{\mathbf{r}(\Sigma)}{n'} + \frac{\|\Sigma\|^2}{g_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{m}} \vee \sqrt{\frac{t}{m}}\right)\sqrt{\frac{t}{m}}\end{aligned} \tag{5.19}$$

and

$$\left|\frac{1}{\hat{d}_r} - 1\right| \lesssim_\gamma \frac{\|\Sigma\|^2}{g_r^2}\frac{\mathbf{r}(\Sigma)}{n'} + \frac{\|\Sigma\|^2}{g_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{m}} \vee \sqrt{\frac{t}{m}}\right)\sqrt{\frac{t}{m}}. \tag{5.20}$$

The key ingredient of the proof of Theorem 3.3 is the following lemma.

LEMMA 5.2. *Suppose that, for some $\gamma \in (0,1)$, conditions* (5.1) *and* (5.12) *hold. Then, for all $t \geq 1$ with probability at least* $1-e^{-t}$

$$\begin{aligned}&\big|\langle\tilde{\theta}_r-\theta_r,u\rangle-\langle L_r(\hat{\Sigma}^{(1)}-\Sigma)\theta_r,u\rangle\big|\\&\quad\lesssim_\gamma \frac{\|\Sigma\|^2}{g_r^2}\biggl(\sqrt{\frac{\mathbf{r}(\Sigma)}{m}}\vee\sqrt{\frac{t}{m}}\vee\frac{t}{m}\biggr)\sqrt{\frac{t}{m}}\|u\|.\end{aligned}\tag{5.21}$$

PROOF. We use the following simple representation:

$$\begin{aligned}&\langle\tilde{\theta}_r-\theta_r,u\rangle\\&\quad=\Bigl\langle\hat{\theta}_r^{(1)}-\sqrt{1+b_r^{(n')}}\theta_r,u\Bigr\rangle\\&\qquad+\biggl(\frac{1}{\hat{d}_r}-1\biggr)\Bigl\langle\hat{\theta}_r^{(1)}-\sqrt{1+b_r^{(n')}}\theta_r,u\Bigr\rangle+\biggl(\frac{\sqrt{1+b_r^{(n')}}}{\hat{d}_r}-1\biggr)\langle\theta_r,u\rangle\end{aligned}\tag{5.22}$$

that holds on the event E (where $\hat{d}_r \geq 1/2$). Using bounds (5.18) and (5.20) that both hold under assumption (5.15) on the event E as well as bound (4.26) (applied to $\hat{\theta}_r^{(1)}$ with $n=n'$), we obtain that with probability at least $1-2e^{-t}$

$$\begin{aligned}&\Bigl|\langle\tilde{\theta}_r-\theta_r,u\rangle-\Bigl\langle\hat{\theta}_r^{(1)}-\sqrt{1+b_r^{(n')}}\theta_r,u\Bigr\rangle\Bigr|\\&\quad\lesssim_\gamma \frac{\|\Sigma\|^2}{g_r^2}\frac{\mathbf{r}(\Sigma)}{n'}\frac{\|\Sigma\|}{g_r}\sqrt{\frac{t}{n'}}\|u\|\\&\qquad+\frac{\|\Sigma\|^2}{g_r^2}\biggl(\sqrt{\frac{\mathbf{r}(\Sigma)}{m}}\vee\sqrt{\frac{t}{m}}\biggr)\sqrt{\frac{t}{m}}\frac{\|\Sigma\|}{g_r}\sqrt{\frac{t}{n'}}\|u\|\\&\qquad+\frac{\|\Sigma\|^2}{g_r^2}\biggl(\sqrt{\frac{\mathbf{r}(\Sigma)}{m}}\vee\sqrt{\frac{t}{m}}\biggr)\sqrt{\frac{t}{m}}\|u\|.\end{aligned}$$

It is easy to check that the last term in the right-hand side is dominant yielding the simpler bound

$$\begin{aligned}&\Bigl|\langle\tilde{\theta}_r-\theta_r,u\rangle-\Bigl\langle\hat{\theta}_r^{(1)}-\sqrt{1+b_r^{(n')}}\theta_r,u\Bigr\rangle\Bigr|\\&\quad\lesssim_\gamma \frac{\|\Sigma\|^2}{g_r^2}\biggl(\sqrt{\frac{\mathbf{r}(\Sigma)}{m}}\vee\sqrt{\frac{t}{m}}\biggr)\sqrt{\frac{t}{m}}\|u\|\end{aligned}\tag{5.23}$$

that holds under assumption (5.15) with probability at least $1-e^{-t}$. Since the left-hand side is bounded by $5\|u\|$, bound (5.23) also holds trivially when

$$\frac{\|\Sigma\|^2}{g_r^2}\biggl(\sqrt{\frac{\mathbf{r}(\Sigma)}{m}}\vee\sqrt{\frac{t}{m}}\biggr)\sqrt{\frac{t}{m}}>c_\gamma.$$

It remains to combine (5.23) with the bound (4.16) (applied to $\hat{\theta}_r^{(1)}$) to complete the proof. □

The following statement is an immediate consequence of Lemma 5.2 and Lemma 4.2. As always, we dropped the terms $\frac{t}{n'}$, $\frac{t}{m}$ from the bounds since the left-hand side is smaller that $3\|u\|$ and, for $t \geq n'$ or $t \geq m$ (the only cases when these terms might be needed), it is dominated by the expression with $\sqrt{\frac{t}{n'}}$, $\sqrt{\frac{t}{m}}$ only.

COROLLARY 5.1. *Suppose that, for some $\gamma \in (0,1)$, conditions (5.1) and (5.12) hold. Then, for all $t \geq 1$ with probability at least $1 - e^{-t}$*

$$|\langle \tilde{\theta}_r - \theta_r, u\rangle| \lesssim_\gamma \frac{\|\Sigma\|}{g_r}\sqrt{\frac{t}{n'}}\|u\| + \frac{\|\Sigma\|^2}{g_r^2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{m}} \vee \sqrt{\frac{t}{m}}\right)\sqrt{\frac{t}{m}}\|u\|. \tag{5.24}$$

Lemma 5.2 implies the following statement. This, in turn, implies Theorem 3.3.

LEMMA 5.3. *Suppose that $m^2 \geq 2n\mathbf{r}(\Sigma)$ and conditions (5.1) and (5.12) hold for some $\gamma \in (0,1)$. For a given $u \in \mathbb{H}$, suppose that $\sigma_r(\Sigma; u) > 0$. Let $\alpha \geq 1$. Then the following bounds holds: for some constants $C, C_{\gamma,\alpha} > 0$,*

$$\begin{aligned}
&\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left\{\frac{\sqrt{n}\langle\tilde{\theta}_r - \theta_r, u\rangle}{\sigma_r(\Sigma; u)} \leq x\right\} - \Phi(x)\right| \\
&\quad\leq C(n')^{-1/2} + \frac{C_{\gamma,\alpha}}{\sigma_r(\Sigma;u)}\frac{\|\Sigma\|^2}{g_r^2} \\
&\quad\quad\times\left(\sqrt{\frac{n\mathbf{r}(\Sigma)}{m^2}\log\frac{m^2}{n\mathbf{r}(\Sigma)}} \vee \sqrt{\frac{n\log^2\frac{m^2}{n\mathbf{r}(\Sigma)}}{m^2}}\right)\|u\| \\
&\quad\quad+\left(\frac{n\mathbf{r}(\Sigma)}{m^2}\right)^{\alpha}.
\end{aligned} \tag{5.25}$$

Moreover, denote

$$\tau_1 := C_\gamma\left(\frac{\|\Sigma\|}{g_r} \vee \frac{\|\Sigma\|^2}{g_r^2}\sqrt{\frac{n\mathbf{r}(\Sigma)}{m^2}}\right)\|u\|$$

and

$$\tau_2 := C_\gamma\frac{\|\Sigma\|^2}{g_r^2}\sqrt{\frac{n}{m^2}}\|u\|.$$

Suppose that Assumptions 3.1 on the loss ℓ holds and $c_2\tau_2 \leq 1/4$. There exist constants $C, C_\gamma, C_{\gamma,\alpha} > 0$ such that

$$\begin{aligned}
&\left|\mathbb{E}\ell\left(\frac{\sqrt{n}\langle\tilde{\theta}_r - \theta_r, u\rangle}{\sigma_r(\Sigma;u)}\right) - \mathbb{E}\ell(Z)\right| \\
&\quad\leq c_1e^{c_2A}\left[\frac{C_{\gamma,\alpha}}{\sigma_r(\Sigma;u)}\frac{\|\Sigma\|^2}{g_r^2}\right. \\
&\quad\quad\times\left(\sqrt{\frac{n\mathbf{r}(\Sigma)}{m^2}\log\frac{m^2}{n\mathbf{r}(\Sigma)}} \vee \sqrt{\frac{n\log^2\frac{m^2}{n\mathbf{r}(\Sigma)}}{m^2}}\right)\|u\| \\
&\quad\quad\left.+\left(\frac{n\mathbf{r}(\Sigma)}{m^2}\right)^{\alpha} + C(n')^{-1/2}\right] \\
&\quad\quad+2e^{3/2}(2\pi)^{1/4}c_1e^{c_2^2\tau_1^2}\left(e^{-A^2/2\tau_1^2} \vee e^{-A/2\tau_2}\right) \\
&\quad\quad+c_1e^{c_2^2}e^{-A^2/4}.
\end{aligned} \tag{5.26}$$

PROOF. The proof is similar to that of Lemma 4.8. To prove (5.25), we apply the first bound of Lemma 4.6 to the random variables

$$\xi := \frac{\sqrt{n}\langle\tilde{\theta}_r - \theta_r, u\rangle}{\sigma_r(\Sigma; u)}, \qquad \eta := \frac{\langle L_r(\hat{\Sigma}^{(1)} - \Sigma)\theta_r, u\rangle}{\sigma_r(\Sigma; u)}$$

and use the bound of Lemma 5.2 with $t = \alpha \log(\frac{m^2}{n\mathbf{r}(\Sigma)})$ to control $\delta(\xi, \eta)$.

To prove the bound (5.26), observe that, by bound (5.24), for all $t \geq 1$ with probability at least $1 - e^{-t}$

$$|\xi| \leq \tau_1\sqrt{t} \vee \tau_2 t.$$

Under assumption $c_2\tau_2 \leq 1/4$, bound (4.28) implies that

$$\mathbb{E}\ell^2(\xi) \leq 2e\sqrt{2\pi}c_1^2 e^{2c_2^2\tau_1^2} + \frac{ec_1^2}{1 - 2c_2\tau_2} \leq 4e\sqrt{2\pi}c_1^2 e^{2c_2^2\tau_1^2}.$$

Therefore,

$$\begin{aligned}\mathbb{E}\ell(\xi)I(|\xi| \geq A) &\leq \mathbb{E}^{1/2}\ell^2(\xi)\mathbb{P}^{1/2}\{|\xi| \geq A\} \\ &\leq 2e^{3/2}(2\pi)^{1/4}c_1 e^{c_2^2\tau_1^2}(e^{-A^2/2\tau_1^2} \vee e^{-A/2\tau_2}).\end{aligned}$$

It remains to repeat the rest of the proof of the second statement of Lemma 4.8. □

Acknowledgments. The authors thank an Associate Editor and two referees for careful reading of and critical remarks on the manuscript. The first author thanks the Department of Pure Mathematics and Mathematical Statistics of the University of Cambridge for its hospitality during visits in Fall 2016 and Summer 2017.

The first author was supported in part by NSF Grants DMS-1810958, DMS-1509739 and CCF-1523768.

The second and third authors were supported by ERC Grant UQMSI/647812.

SUPPLEMENTARY MATERIAL

Supplement to "Efficient estimation of linear functionals of principal components" (DOI: 10.1214/19-AOS1816SUPP; .pdf). Supplementary information.

REFERENCES

[1] ANDERSON, T. W. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Stat.* **34** 122–148. MR0145620 https://doi.org/10.1214/aoms/1177704248

[2] BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227** 494–521. MR2782201 https://doi.org/10.1016/j.aim.2011.02.007

[3] BERTHET, Q. and RIGOLLET, P. (2013). Optimal detection of sparse principal components in high dimension. *Ann. Statist.* **41** 1780–1815. MR3127849 https://doi.org/10.1214/13-AOS1127

[4] BLANCHARD, G., BOUSQUET, O. and ZWALD, L. (2007). Statistical properties of kernel principal component analysis. *Mach. Learn.* **66** 259–294.

[5] BLOEMENDAL, A., KNOWLES, A., YAU, H.-T. and YIN, J. (2016). On the principal components of sample covariance matrices. *Probab. Theory Related Fields* **164** 459–552. MR3449395 https://doi.org/10.1007/s00440-015-0616-x

[6] CAI, T. T. and GUO, Z. (2017). Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity. *Ann. Statist.* **45** 615–646. MR3650395 https://doi.org/10.1214/16-AOS1461

[7] CAI, T. T., MA, Z. and WU, Y. (2013). Sparse PCA: Optimal rates and adaptive estimation. *Ann. Statist.* **41** 3074–3110. MR3161458 https://doi.org/10.1214/13-AOS1178

[8] DAUXOIS, J., POUSSE, A. and ROMAIN, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *J. Multivariate Anal.* **12** 136–154. MR0650934 https://doi.org/10.1016/0047-259X(82)90088-4

[9] EATON, M. L. (1983). *Multivariate Statistics*: *A Vector Space Approach*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, New York. MR0716321

[10] FAN, J., RIGOLLET, P. and WANG, W. (2015). Estimation of functionals of sparse covariance matrices. *Ann*. *Statist*. **43** 2706–2737. MR3405609 https://doi.org/10.1214/15-AOS1357

[11] GAO, C. and ZHOU, H. H. (2015). Rate-optimal posterior contraction for sparse PCA. *Ann*. *Statist*. **43** 785–818. MR3325710 https://doi.org/10.1214/14-AOS1268

[12] GAO, C. and ZHOU, H. H. (2016). Bernstein–von Mises theorems for functionals of the covariance matrix. *Electron*. *J*. *Stat*. **10** 1751–1806. MR3522660 https://doi.org/10.1214/15-EJS1048

[13] GILL, R. D. and LEVIT, B. Y. (1995). Applications of the Van Trees inequality: A Bayesian Cramér–Rao bound. *Bernoulli* **1** 59–79. MR1354456 https://doi.org/10.2307/3318681

[14] GINÉ, E. and NICKL, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge Univ. Press, New York. MR3588285 https://doi.org/10.1017/CBO9781107337862

[15] JANKOVÁ, J. and VAN DE GEER, S. (2018). Semiparametric efficiency bounds for high-dimensional models. *Ann*. *Statist*. **46** 2336–2359. MR3845020 https://doi.org/10.1214/17-AOS1622

[16] JAVANMARD, A. and MONTANARI, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J*. *Mach*. *Learn*. *Res*. **15** 2869–2909. MR3277152

[17] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann*. *Statist*. **29** 295–327. MR1863961 https://doi.org/10.1214/aos/1009210544

[18] KOLTCHINSKII, V. (2017). Asymptotically efficient estimation of smooth functionals of covariance operators. Available at https://arxiv.org/abs/1710.09072.

[19] KOLTCHINSKII, V., LÖFFLER, M. and NICKL, R. (2019). Supplement to "Efficient estimation of linear functionals of principal components." https://doi.org/10.1214/19-AOS1816SUPP.

[20] KOLTCHINSKII, V. and LOUNICI, K. (2016). Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance. *Ann*. *Inst*. *Henri Poincaré Probab*. *Stat*. **52** 1976–2013. MR3573302 https://doi.org/10.1214/15-AIHP705

[21] KOLTCHINSKII, V. and LOUNICI, K. (2017). Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli* **23** 110–133. MR3556768 https://doi.org/10.3150/15-BEJ730

[22] KOLTCHINSKII, V. and LOUNICI, K. (2017). Normal approximation and concentration of spectral projectors of sample covariance. *Ann*. *Statist*. **45** 121–157. MR3611488 https://doi.org/10.1214/16-AOS1437

[23] KOLTCHINSKII, V. and LOUNICI, K. (2017). New asymptotic results in principal component analysis. *Sankhya A* **79** 254–297. MR3707422 https://doi.org/10.1007/s13171-017-0106-6

[24] LILA, E., ASTON, J. A. D. and SANGALLI, L. M. (2016). Smooth principal component analysis over two-dimensional manifolds with an application to neuroimaging. *Ann*. *Appl*. *Stat*. **10** 1854–1879. MR3592040 https://doi.org/10.1214/16-AOAS975

[25] NADLER, B. (2008). Finite sample approximation results for principal component analysis: A matrix perturbation approach. *Ann*. *Statist*. **36** 2791–2817. MR2485013 https://doi.org/10.1214/08-AOS618

[26] NAUMOV, A., SPOKOINY, V. and ULYANOV, V. (2018). Confidence sets for spectral projectors of covariance matrices. *Dokl*. *Math*. **98** 511–514.

[27] NING, Y. and LIU, H. (2017). A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *Ann*. *Statist*. **45** 158–195. MR3611489 https://doi.org/10.1214/16-AOS1448

[28] PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist*. *Sinica* **17** 1617–1642. MR2399865

[29] RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2168993

[30] REISS, M. and WAHL, M. (2016). Non-asymptotic upper bounds for the reconstruction error of PCA. Available at arXiv:1609.03779.

[31] REN, Z., SUN, T., ZHANG, C.-H. and ZHOU, H. H. (2015). Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *Ann*. *Statist*. **43** 991–1026. MR3346695 https://doi.org/10.1214/14-AOS1286

[32] VAN DE GEER, S., BÜHLMANN, P., RITOV, Y. and DEZEURE, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann*. *Statist*. **42** 1166–1202. MR3224285 https://doi.org/10.1214/14-AOS1221

[33] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

[34] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. *Springer Series in Statistics*. Springer, New York. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2

[35] VERSHYNIN, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* 210–268. Cambridge Univ. Press, Cambridge. MR2963170

[36] VU, V. Q. and LEI, J. (2013). Minimax sparse principal subspace estimation in high dimensions. *Ann. Statist.* **41** 2905–2947. MR3161452 https://doi.org/10.1214/13-AOS1151

[37] WANG, T., BERTHET, Q. and SAMWORTH, R. J. (2016). Statistical and computational trade-offs in estimation of sparse principal components. *Ann. Statist.* **44** 1896–1930. MR3546438 https://doi.org/10.1214/15-AOS1369

[38] WANG, W. and FAN, J. (2017). Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *Ann. Statist.* **45** 1342–1374. MR3662457 https://doi.org/10.1214/16-AOS1487

[39] ZHANG, C.-H. and ZHANG, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 217–242. MR3153940 https://doi.org/10.1111/rssb.12026