

An Improvement on the Hotelling T^2 Test Using the Ledoit-Wolf Nonlinear Shrinkage Estimator

Benjamin Robinson
Sensors Directorate
US Air Force Research Laboratory
Dayton, Ohio
ORCID: 0000-0002-9391-4543

Robert Malinas
EECS
University of Michigan
Ann Arbor, MI, USA
ORCID: 0000-0002-1318-5564

Van Latimer
Mathematics
UCLA
Los Angeles, California, USA
ORCID: 0000-0001-9789-7739

Beth Bjorkman Morrison
Sensors Directorate
US Air Force Research Laboratory
Dayton, Ohio
ORCID: 0000-0001-7406-0311

Alfred O. Hero, III
EECS
University of Michigan
Ann Arbor, MI, USA
ORCID: 0000-0002-2531-9670

Abstract—Hotelling’s T^2 test is a classical approach for discriminating the means of two multivariate normal samples that share a population covariance matrix. Hotelling’s test is not ideal for high-dimensional samples because the eigenvalues of the estimated sample covariance matrix are inconsistent estimators for their population counterparts. We replace the sample covariance matrix with the nonlinear shrinkage estimator of Ledoit and Wolf 2020. We observe empirically for sub-Gaussian data that the resulting algorithm dominates past methods (Bai and Saranadasa 1996, Chen and Qin 2010, and Li et al. 2020) for a family of population covariance matrices that includes matrices with high or low condition number and many or few nontrivial—i.e., spiked—eigenvalues.

Index Terms—Two-sample testing, high-dimensional limit, shrinkage covariance estimation, Ledoit-Wolf estimator, Hotelling T^2 test

I. INTRODUCTION

A fundamental problem in statistics and signal processing is determining whether two independent samples have the same mean. For multivariate samples that are Gaussian and have a shared population covariance matrix, classical methods like the Hotelling T^2 test apply [1]. However, there is no standard technique if the sample dimension is substantial compared to the sample sizes: the so-called large-dimensional regime.

Hotelling’s T^2 test relies on estimating the shared population covariance matrix using the sample covariance matrix. In the large-dimensional regime, the sample covariance matrix’s eigenvalues are inconsistent estimators for their population counterparts [2], leading to poor performance of Hotelling’s test. Additionally, the sample covariance matrix can be considerably more ill-conditioned than the population covariance matrix, resulting in numerical instabilities [3]. As a result, several authors have proposed alternatives to the standard T^2 test. Bai and Saranadasa [4] and Chen and Qin [5] propose tests (BS96 and CQ10) that are an improvement for a well-conditioned population covariance matrix, and Li et al. [6]

propose a test (LAPPW20) that is an improvement if the population covariance follows the spiked model of Johnstone [3]. To the best of our knowledge, until now no test improves upon of all these methods under more general assumptions.

In this paper, we propose a replacement for Hotelling’s T^2 test and present some simulations in which it dominates BS96, CQ10, and LAPPW20. We do not restrict the condition number or the number of spiked eigenvalues of the population covariance matrix in our testing. Our method, similar to LAPPW20, is to replace the sample covariance matrix in Hotelling’s T^2 with a matrix that has a smaller condition number. More precisely, we use the covariance matrix estimator of Ledoit and Wolf [7], which belongs to Stein’s shrinkage class [8], [9]. By contrast, LAPPW20 uses a diagonal-loading estimator, which also belongs to Stein’s class, but is simply a sum of the sample covariance matrix and a scalar multiple of the identity. We argue for our test’s optimality within Stein’s class using ideas similar to those some of us have applied to other detection problems [10]. We note that we are not the first to have thought of applying Ledoit-Wolf-type estimators to two-sample testing, but are developing the idea toward maturity [11].

In Section II, we explore past improvements on Hotelling’s T^2 test. In Section III, we give the Ledoit-Wolf estimator utilized in our approach, present the proposed test, and argue for the test’s asymptotic optimality and predict a precise asymptotically constant false-alarm rate. In Section IV, we provide simulations that show empirical improvement over past methods. Finally, in Section V, we present our conclusions.

II. BACKGROUND

Suppose we have independent random p -dimensional column vectors $\mathbf{x}_{ij} \sim \mathcal{N}(\boldsymbol{\mu}_i, \mathbf{R})$ for $j = 1, 2, \dots, n_i$ and $i \in \{1, 2\}$, where $\boldsymbol{\mu}_i$ are unknown means and \mathbf{R} is an unknown

$p \times p$ symmetric positive-definite population covariance matrix. We want to test the following hypotheses

$$\begin{cases} \mathcal{H}_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2, \\ \mathcal{H}_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2. \end{cases}$$

If the covariance matrix \mathbf{R} is known, a reasonable detector is the Mahalanobis-distance detector

$$(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{R}^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \tau, \quad (1)$$

where $(\cdot)'$ denotes the transpose, $\tau \in (0, \infty)$, and $\bar{\mathbf{x}}_i$ is the sample mean of the \mathbf{x}_{ij} 's. For unknown \mathbf{R} , the classical replacement of (1) is Hotelling's T^2 test:

$$(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_n^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \tau, \quad (2)$$

where $n = n_1 + n_2 - 2$ and \mathbf{S}_n is the "pooled" sample covariance matrix, given by

$$\mathbf{S}_n = \frac{1}{n} \sum_{i=1}^2 \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)'. \quad (3)$$

The performance of the above test is well-characterized when $n_1, n_2 \rightarrow \infty$ and p is fixed [1]; however, this test can become inadmissible in the regime where $p, n_1, n_2 \rightarrow \infty$ and $p/n_1 \rightarrow \gamma_1$ and $p/n_2 \rightarrow \gamma_2$ for some $\gamma_1, \gamma_2 \in (0, \infty)$ [4]. This latter limit is known as the *high-dimensional asymptotic regime*.

Bai and Saranadasa [4] suggest an improvement, BS96, under the assumption that $\|\mathbf{R}\| = o(\sqrt{\text{tr}(\mathbf{R}^2)})$. This assumption yields that the matrix \mathbf{R} is well-conditioned, for otherwise the two referenced quantities would be similar in size. Their proposed test statistic is

$$\frac{\frac{n_1 n_2}{n_1 + n_2} \|\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2\|^2 - \text{tr} \mathbf{S}_n}{\sqrt{\frac{2(n+1)}{n} B_n}}, \quad (4)$$

where

$$B_n = \frac{n^2}{(n+2)(n-1)} \left(\text{tr}(\mathbf{S}_n^2) - \frac{1}{n} (\text{tr} \mathbf{S}_n)^2 \right).$$

The corresponding detector is desirable in that, asymptotically, it has known constant false-alarm rate and predictable detection power.

In response to the need to analyze ultra-high dimensional data, Chen and Qin [5] propose a test, CQ10, that is similar to BS96 but is additionally well suited even to values of n much smaller than p . Their test possesses desirable provable optimality properties, provided again an assumption that implies a low condition number: $\text{tr}(\mathbf{R}^4) = o\{\text{tr}^2(\mathbf{R}^2)\}$. Their test statistic is

$$\frac{\sum_{i \neq j}^{n_1} \mathbf{x}'_{1i} \mathbf{x}_{1j} + \sum_{i \neq j}^{n_2} \mathbf{x}'_{2i} \mathbf{x}_{2j} - \frac{2 \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \mathbf{x}'_{1i} \mathbf{x}_{2j}}{n_1 n_2}}{n_1(n_1 - 1) + n_2(n_2 - 1)}.$$

Li et al. [6] assume instead that \mathbf{R} follows the well-known *spiked covariance model* of Johnstone [3], i.e., that all the eigenvalues of \mathbf{R} are equal to one except for a fixed finite number that are larger than one. This assumption potentially

contrasts with the well-conditionedness assumption of BS96 and CQ10. Li et al. propose a test, LAPPW20, that replaces \mathbf{S}_n in (2) by the *diagonal-loading* estimator $\mathbf{S}_n + \lambda \mathbf{I}$, for some optimal *loading factor* $\lambda > 0$. Like BS96 and CQ10, LAPPW20 has a known asymptotically constant false-alarm rate and predictable detection power. Further, Li et al. remark that LAPPW20 reduces to BS96 for population covariance matrices that are both well-conditioned and spiked.

The covariance estimator of LAPPW20 belongs to the shrinkage class of Stein [8], [9]. These are estimators that differ from \mathbf{S}_n only in their eigenvalues. That is, all of the eigenvectors of the estimator are eigenvectors of \mathbf{S}_n , but the eigenvalues may be different from \mathbf{S}_n 's. Shrinkage covariance estimators have been studied in the high-dimensional regime in the spiked model by Donoho et al. [12], and in a more general model by Ledoit and Wolf [13]–[15]. In particular, Ledoit and Wolf have devised closed-form expressions for shrinkage eigenvalues that are asymptotically optimal with respect to many criteria, including Stein's loss, inverse Stein's loss, Frobenius loss, inverse Frobenius loss, and so-called minimum-variance loss [7]. In what follows, we will propose again replacing \mathbf{S}_n in Hotelling's T^2 test, but rather than a diagonal-loading estimator, we utilize a shrinkage estimator of Ledoit and Wolf.

III. PROPOSED TEST

In this section, we define our proposed test, give an indication of its asymptotic false-alarm rate, and argue for a type of asymptotic detection-theoretic optimality.

A. Definition

In order to define our proposed test, we must first describe Ledoit and Wolf's nonlinear shrinkage estimator from [7]. Throughout this paper, we will consider the high-dimensional limit in which $n, p \rightarrow \infty$ and $p/n = p/(n_1 + n_2 - 2) \rightarrow \gamma \in (0, \infty)$. The notation \rightarrow will always refer to the limit in which $n, p \rightarrow \infty$ and $p/n \rightarrow \gamma$. For notational convenience, we will follow the convention of [16] that quantities that are not explicitly deemed constant are varying with n and p . Identifying \mathbf{S}_n with \mathbf{S} , then, we let $\mathbf{S} = \mathbf{U} \text{diag}(\boldsymbol{\lambda}) \mathbf{U}'$ be an eigendecomposition, where for emphasis \mathbf{U} and $\boldsymbol{\lambda}$ depend on n and p . We list the eigenvalues $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_p)$ in non-increasing order, with corresponding eigenvectors $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p]$.

The Ledoit-Wolf nonlinear shrinkage estimator can be described as follows. Let $h_j = n^{-1/3} \lambda_j$ and $[y]^+ = \max\{y, 0\}$. Define

$$a(\lambda, \boldsymbol{\lambda}) := \sum_{j=[p-n]^++1}^p \left\{ -\frac{3(\lambda - \lambda_j)}{10\pi h_j^2} + \frac{3}{4\sqrt{5}\pi h_j} \times \left[1 - \frac{1}{5} \left(\frac{\lambda - \lambda_j}{h_j} \right)^2 \right] \log \left| \frac{\sqrt{5}h_j - \lambda + \lambda_j}{\sqrt{5}h_j + \lambda - \lambda_j} \right| \right\},$$

and

$$b(\lambda, \boldsymbol{\lambda}) := \sum_{j=[p-n]^++1}^p \frac{3}{4\sqrt{5}h_j} \left[1 - \frac{1}{5} \left(\frac{\lambda - \lambda_j}{h_j} \right)^2 \right]^+,$$

and

$$s(\lambda, \boldsymbol{\lambda}) = \pi(a(\lambda, \boldsymbol{\lambda}) + ib(\lambda, \boldsymbol{\lambda}))/\min\{n, p\}.$$

Next, the shrunken eigenvalues \hat{d}_i are defined as

$$\hat{d}_i := \begin{cases} \frac{\lambda_i}{|1-p/n-(p/n)\lambda_i s(\lambda_i, \boldsymbol{\lambda})|^2}, & \lambda_i > 0 \\ \frac{1}{(p/n-1)a(0, \boldsymbol{\lambda})/n}, & \lambda_i = 0. \end{cases}$$

(We assume $p \neq n$.) Finally, the Ledoit-Wolf estimator, denoted $\hat{\mathbf{R}}_{\text{LW}}$, is given by $\hat{\mathbf{R}}_{\text{LW}} = \mathbf{U}\text{diag}(\hat{d}_1, \hat{d}_2, \dots, \hat{d}_p)\mathbf{U}'$.

We define our test to be the Hotelling T^2 test with $\hat{\mathbf{R}}_{\text{LW}}$ substituted for \mathbf{S} :

$$T_{\text{LW}}^2 := \frac{n_1 n_2}{n_1 + n_2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \hat{\mathbf{R}}_{\text{LW}}^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau.$$

For the false-alarm rate analysis that follows, we shift and re-scale T_{LW}^2 for the equivalent test

$$Z = \frac{1}{\sqrt{2p}} (T_{\text{LW}}^2 - p) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau. \quad (5)$$

B. False-Alarm Rate

A key consideration about any test statistic is whether its null distribution can be characterized. In the following, we argue intuitively that the null distribution of Z should be asymptotically standard normal for Gaussian data and provide empirical support for this assertion. We expect a similar result holds for sub-Gaussian data.

Let $\sigma_i^2 = \mathbf{u}_i' \mathbf{R} \mathbf{u}_i$. The test in subsection III-A is related to the test statistic \tilde{Z} , defined as follows. Let $\hat{\mathbf{R}}$ be a shrinkage estimator $\mathbf{U}\mathbf{D}\mathbf{U}'$, where without loss of generality,

$$\mathbf{D} = \text{diag} \left(\frac{\sigma_1^2}{c_1}, \frac{\sigma_2^2}{c_2}, \dots, \frac{\sigma_p^2}{c_p} \right),$$

for some positive coefficients $\{c_i\}$. Define

$$\tilde{Z} := \frac{1}{\sqrt{2 \sum_i c_i^2}} \left(\tilde{T}^2 - \sum_{i=1}^p c_i \right), \quad (6)$$

where \tilde{T} is given by

$$\tilde{T}^2 := \frac{n_1 n_2}{n_1 + n_2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \hat{\mathbf{R}}^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2). \quad (7)$$

By the Berry-Esseen theorem, the conditional distribution of \tilde{Z} given \mathbf{S} is almost surely asymptotically standard normal.

The shrunken eigenvalues \hat{d}_i of Ledoit and Wolf are designed to approximate σ_i^2 . A weak form of this approximation that holds under a general matrix model laid out [7], [13] is that for all fixed intervals $I \subset [0, \infty)$

$$\epsilon := \frac{1}{p} \sum_{\lambda_j \in I} (\hat{d}_j - \sigma_j^2) \xrightarrow{a.s.} 0. \quad (8)$$

To borrow notation from [17], we conjecture further that, under the conditions of [7], [13], $\epsilon = O_{\prec}(p^{-d})$ uniformly in intervals I for some $d > 0$, where, $a_p = O_{\prec}(b_p)$ means that for any $\delta, D > 0$, we have for sufficiently large p that

$$\Pr[|a_p| > p^\delta | b_p] < p^{-D}. \quad (9)$$

Histogram of Proposed Detector Under \mathcal{H}_0

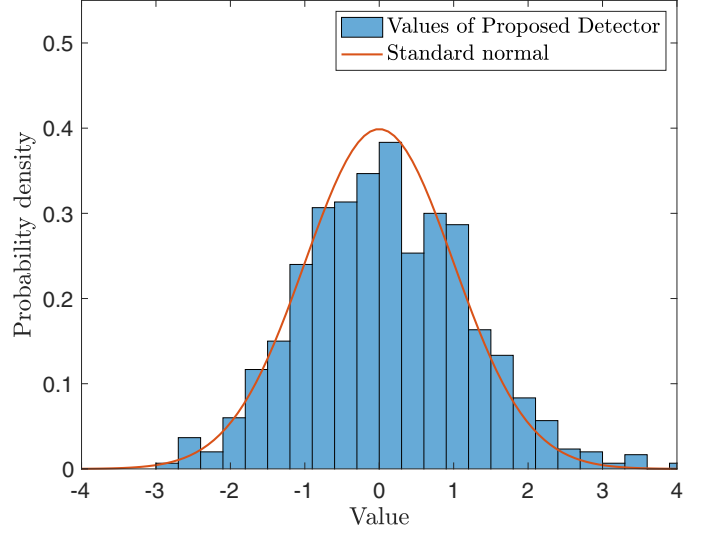


Fig. 1. A plot showing Z from (5) to be roughly standard normal. Here, the population covariance matrix follows the model \mathbf{R}_4 in Section IV.

In other words, with polynomially high confidence, a_p does not grow much more quickly than b_p , if at all. We further conjecture that

$$\max_{i,j} |\mathbf{u}_i' \mathbf{R} \mathbf{u}_j - \delta_{ij} \hat{d}_j| = O_{\prec}(p^{-d}) \quad (10)$$

for some $d > 0$. This conjecture is motivated by an analogous result involving quadratic forms of Wigner-matrix eigenvectors [18, Equation 1].

Taking $c_i = \sigma_i^2 / \hat{d}_i$, the test statistic in (6) becomes

$$\frac{1}{\sqrt{2 \sum_i \sigma_i^4 / \hat{d}_i^2}} \left(T_{\text{LW}}^2 - \sum_{i=1}^p \frac{\sigma_i^2}{\hat{d}_i} \right).$$

Using the fact that the above is asymptotically almost surely standard normal given \mathbf{S} , and given the approximations of the last paragraph, we expect Z to be asymptotically almost surely standard normal as well. In particular, we expect $Z \gtrsim \tau$ has an asymptotically constant false-alarm rate.

We present a simulation comparing a finite-sample distribution of Z to a standard normal in Figure 1. Using 1000 Monte-Carlo values of Z with \mathbf{R} being the 200×200 matrix \mathbf{R}_4 defined in the Section IV and $n_1 = n_2 = 200$, one can see that the finite-sample distribution of Z approximates a standard normal, as predicted.

C. Maximizing Detection Power

We now provide an intuitive argument that $\hat{\mathbf{R}}_{\text{LW}}$ in (7) optimizes conditional detection power given \mathbf{S} among shrinkage estimators $\hat{\mathbf{R}}$, for Gaussian data. It can be shown using (8) that $t = \text{tr}(\mathbf{R}\hat{\mathbf{R}}^{-1})/p$ depends in the limit only on the limiting sample spectrum, so $\hat{\mathbf{R}} \leftarrow t\hat{\mathbf{R}}$ can be considered to be a shrinkage estimator, and we assume without loss of generality that t converges to 1.

Consider the conditional false-alarm rate of the proposed detector given \mathbf{S} : $\Pr[\mathbf{d}'\hat{\mathbf{R}}^{-1}\mathbf{d} > \tau \mid \mathbf{S}]$, where $\mathbf{d} = \mathbf{R}^{1/2}\mathbf{w}$ and $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Using the result [19, Theorem 5.1.4] regarding concentration of Lipschitz functions on the sphere and the fact that $p \rightarrow \infty$, this probability can be approximated by

$$\Pr\left[\|\mathbf{w}\|^2 \text{tr}(\mathbf{R}^{1/2}\hat{\mathbf{R}}^{-1}\mathbf{R}^{1/2}) > \tau \mid \mathbf{S}\right]. \quad (11)$$

Since $\hat{\mathbf{R}}$ depends only on \mathbf{S} , the probability above is a function of t , τ , and p alone. Since by assumption t converges to 1, (11) is asymptotically dependent on τ and p alone, and the same can be said of the corresponding unconditional probability. Thus, we wish to maximize detection power for each τ and large p .

Consider the conditional probability of detection given \mathbf{S} :

$$\Pr\left[(\boldsymbol{\mu} + \mathbf{d})'\hat{\mathbf{R}}^{-1}(\boldsymbol{\mu} + \mathbf{d}) > \tau \mid \mathbf{S}\right], \quad (12)$$

where $\boldsymbol{\mu} = \sqrt{n_1 n_2 / (n_1 + n_2)}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$. Expanding (12), we get

$$\Pr\left[\alpha + \boldsymbol{\beta}'\mathbf{v} + \|\mathbf{w}\|^2 > \tau \mid \mathbf{S}\right], \quad (13)$$

where $\alpha = \boldsymbol{\mu}'\hat{\mathbf{R}}^{-1}\boldsymbol{\mu}$, $\boldsymbol{\beta} = 2\boldsymbol{\mu}'\hat{\mathbf{R}}^{-1}\mathbf{R}^{1/2}\|\mathbf{w}\|$, and $\mathbf{v} = \mathbf{w}/\|\mathbf{w}\|$. If we further condition on $\|\mathbf{w}\|$, the argument of the probability in (13) depends only on the randomness of \mathbf{v} , and the probability itself depends only on α and $\|\boldsymbol{\beta}\|$. More precisely, the conditional probability of detection given \mathbf{S} and $\|\mathbf{w}\|$ is maximized when $\alpha/\|\boldsymbol{\beta}\|$ is maximized.

For any $\|\mathbf{w}\|$ and \mathbf{S} , maximizing $\alpha/\|\boldsymbol{\beta}\|$ is equivalent to maximizing the *signal-to-noise ratio*:

$$\frac{(\boldsymbol{\mu}'\hat{\mathbf{R}}^{-1}\boldsymbol{\mu})^2}{\boldsymbol{\mu}'\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\boldsymbol{\mu}}. \quad (14)$$

Using [19, Theorem 5.1.4] again, (14) is well-approximated in probability for almost all $\boldsymbol{\mu}$ by

$$\frac{\text{tr}(\hat{\mathbf{R}}^{-1})^2}{p\text{tr}(\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1})}, \quad (15)$$

so that the ideal choice of $\hat{\mathbf{R}}$ subject to the constraint $t \approx 1$ is the Ledoit-Wolf estimator discussed in Section III-A. This justifies our choice of detector.

An additional feature of equation (15) is that it can be used to numerically find the optimum in the class of diagonal-loading estimators, which is a class of shrinkage estimators that includes LAPPW20. This can be done by simply plugging $\hat{\mathbf{R}} = \mathbf{S} + \lambda\mathbf{I}$ in (15) and iteratively solving for the optimal $\lambda > 0$ using, for example, `fminsearch` in Matlab. Although this estimator is not technically the same as LAPPW20, we refer to it henceforth as LAPPW20 since we can calculate it and its performance is an upper bound for LAPPW20's.

IV. SIMULATIONS

A key attribute of our proposed algorithm is its performance in simulation. In this section, we show performance that is consistent with average-case dominance over BS96, CQ10, and

LAPPW20, under the model that the difference in population means of the two samples is drawn uniformly from the sphere.

In this Section, the dimension p is 200, n_1 and n_2 are taken to be 150, and we consider diagonal population covariance matrices $\mathbf{R} = \mathbf{R}_P$ parametrized by an integer $P \in \{0, 2, 4\}$. The eigenvalues of \mathbf{R}_P are defined for $1 \leq j \leq 40$ by

$$(\mathbf{R}_P)_{jj} = 10^{(41-j)P/40} + \epsilon_j, \quad (1 \leq j \leq 40),$$

where ϵ_j is chosen i.i.d. uniformly at random from $[0, 1]$. On the other hand, for $j > 40$, we set $(\mathbf{R}_P)_{jj} = 1$. Thus, largest eigenvalue of \mathbf{R}_P is 10^P , the next 39 decrease exponentially, and the remaining 160 are equal to unity. This choice of spectrum mirrors the generalized spiked structure widely encountered in sensing applications, such as radar, and P corresponds to the order of the \mathbf{R} 's condition number.

For each detector and each $P \in \{0, 2, 4\}$, we generate 100,000 detection scores as follows. We generate a sub-Gaussian $p \times n_1$ and a $p \times n_2$ data matrix \mathbf{X}_1 and \mathbf{X}_2 , both by coloring a matrix of i.i.d. uniform random variables with mean zero and variance one. (Simulations appear to be similar for Gaussian data matrices.) We perturb the columns of one of these data matrices by a random vector uniformly drawn from a sphere of radius 1 and treat the mean of the other as $\mathbf{0}$. We then plug the data into the Proposed, LAPPW20, BS96, CQ10, and Hotelling detectors. The resulting 100,000 detection scores for each detector are used to generate the ROC curves in Figures 2, 3, and 4, which correspond to $P = 0, 2, 4$, respectively. Figure 2 shows the case $P = 0$, where BS96, LAPPW20, CQ10, and the proposed method all perform similarly, and all dominate Hotelling. In Figures 3 and 4, the proposed method dominates BS96, LAPPW20, CQ10, and Hotelling, with the advantage being more pronounced for increasing values of P . This is likely due to the fact that the number of large eigenvalues in these cases challenges both the spiked assumption of LAPPW20 and the well-conditioned assumption of BS96 and CQ10, mentioned in Section II.

V. CONCLUSION

In this paper, we have proposed an alternative to Hotelling's T^2 test with desirable theoretical and empirical properties in the high-dimensional setting. In simulation, the proposed test appears to dominate the state-of-the-art alternatives LAPPW20, CQ10, and BS96, likely due in part to the comparatively less forgiving assumptions they require. We provided an intuitive argument that our test has an asymptotically constant false-alarm rate and maximal detection power among shrinkage-based Hotelling-type tests under mild assumptions. We conjecture that a more rigorous foundation for this theory can be established by finding conditions under which (10) and strengthenings of (8) hold, which we intend to investigate in future work.

ACKNOWLEDGMENTS

This work was supported by the United States Air Force Sensors Directorate, AFOSR grant 19RYCOR036, ARO grant W911NF-15-1-0479, and the DoD SMART Scholarship SEED

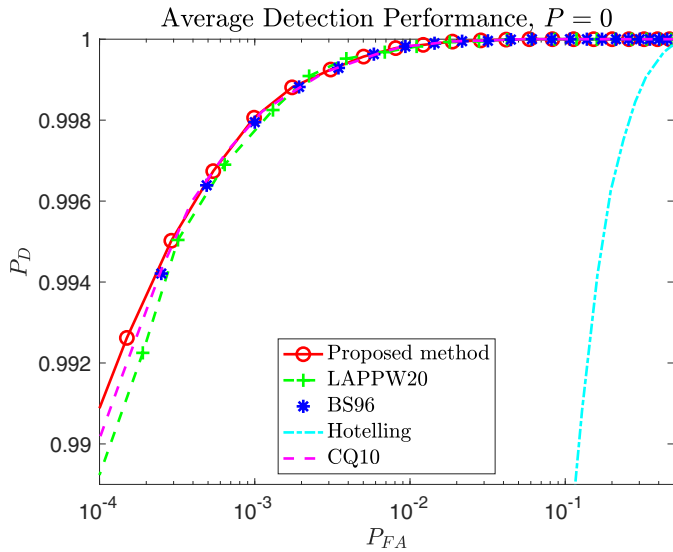


Fig. 2. For $\mathbf{R} = \mathbf{R}_0$ with $O(10^0)$ condition number, all methods perform similarly, except Hotelling.

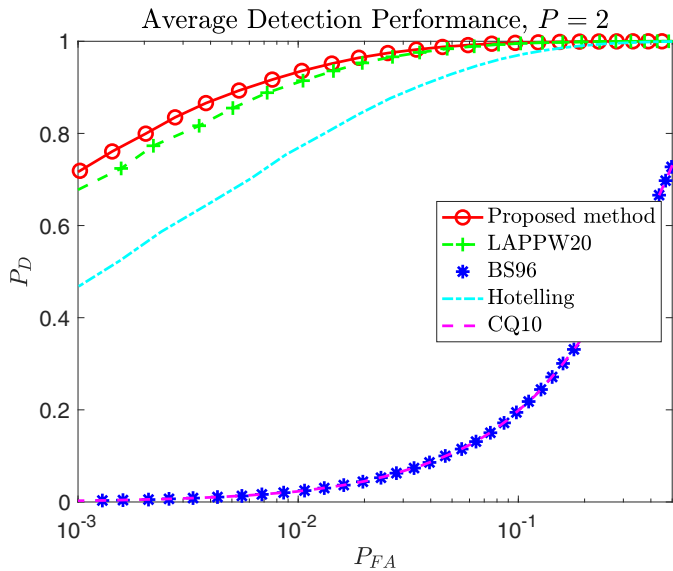


Fig. 3. For $\mathbf{R} = \mathbf{R}_2$ with $O(10^2)$ condition number, Proposed method performs similarly to LAPPW20, which outperforms Hotelling, which outperforms BS96 and CQ10, which lie on the chance line.

grant. However, the views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of any agency of the U.S. government. Examples of analysis performed within this article are only examples. Assumptions made within the analysis are also not reflective of the position of any U.S. Government entity. The Public Affairs approval number of this document is AFRL-2022-2605.

REFERENCES

- [1] T. W. Anderson, "Asymptotic theory for principal component analysis," *The Annals of Mathematical Statistics*, vol. 34, no. 1, pp. 122–148, 1963.
- [2] D. Paul, "Asymptotics of sample eigenstructure for a large dimensional spiked covariance model," *Statistica Sinica*, pp. 1617–1642, 2007.

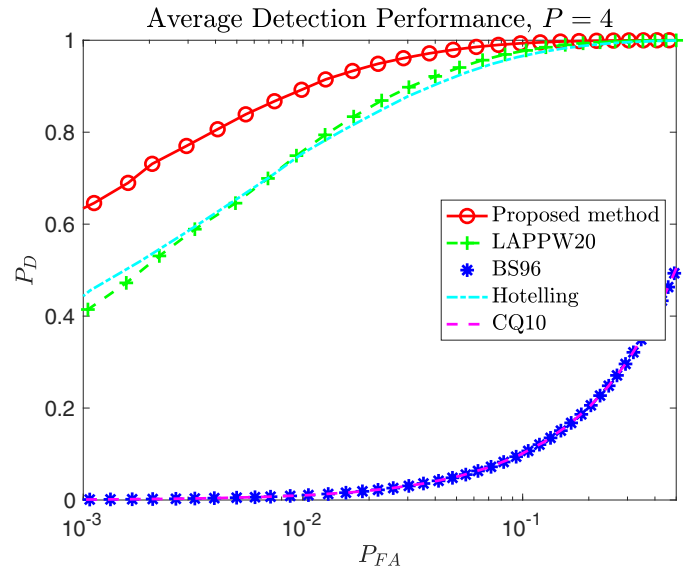


Fig. 4. For $\mathbf{R} = \mathbf{R}_4$ with $O(10^4)$ condition number, Proposed method outperforms LAPPW20 and Hotelling, which outperform BS96 and CQ10, which lie on the chance line.

- [3] I. M. Johnstone, "On the distribution of the largest eigenvalue in principal components analysis," *Annals of Statistics*, pp. 295–327, 2001.
- [4] Z. Bai and H. Saranadasa, "Effect of high dimension: By an example of a two sample problem," *Statistica Sinica*, pp. 311–329, 1996.
- [5] S. X. Chen and Y.-L. Qin, "A two-sample test for high-dimensional data with applications to gene-set testing," *The Annals of Statistics*, vol. 38, no. 2, pp. 808–835, 2010.
- [6] H. Li, A. Aue, D. Paul, J. Peng, and P. Wang, "An adaptable generalization of Hotelling's T^2 test in high dimension," *The Annals of Statistics*, vol. 48, no. 3, pp. 1815–1847, 2020.
- [7] O. Ledoit and M. Wolf, "Analytical nonlinear shrinkage of large-dimensional covariance matrices," *The Annals of Statistics*, vol. 48, no. 5, pp. 3043–3065, 2020.
- [8] C. Stein, "Estimation of a covariance matrix, Rietz lecture," in *39th Annual Meeting IMS, Atlanta, GA, 1975*, 1975.
- [9] —, "Lectures on the theory of estimation of many parameters," *Journal of Soviet Mathematics*, vol. 34, no. 1, pp. 1373–1403, 1986.
- [10] B. D. Robinson, R. Malinas, and A. O. Hero, "Space-time adaptive detection at low sample support," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2939–2954, 2021.
- [11] J. Namdari, D. Paul, and L. Wang, "High-dimensional linear models: A random matrix perspective," *Sankhya A*, vol. 83, no. 2, pp. 645–695, 2021.
- [12] D. L. Donoho, M. Gavish, and I. M. Johnstone, "Optimal shrinkage of eigenvalues in the spiked covariance model," *Annals of Statistics*, vol. 46, no. 4, p. 1742, 2018.
- [13] O. Ledoit and S. Péché, "Eigenvectors of some large sample covariance matrix ensembles," *Probability Theory and Related Fields*, vol. 151, no. 1-2, pp. 233–264, 2011.
- [14] O. Ledoit and M. Wolf, "Nonlinear shrinkage estimation of large-dimensional covariance matrices," *The Annals of Statistics*, vol. 40, no. 2, pp. 1024–1060, 2012.
- [15] —, "Optimal estimation of a large-dimensional covariance matrix under Stein's loss," *Bernoulli*, vol. 24, no. 4B, pp. 3791–3832, 2018.
- [16] F. Benaych-Georges and A. Knowles, "Lectures on the local semicircle law for Wigner matrices," *arXiv preprint arXiv:1601.04055*, 2016.
- [17] L. Erdős, A. Knowles, and H.-T. Yau, "Averaging fluctuations in resolvents of random band matrices," in *Annales Henri Poincaré*, vol. 14, no. 8. Springer, 2013, pp. 1837–1926.
- [18] G. Cipolloni, L. Erdős, and D. Schröder, "Eigenstate thermalization hypothesis for Wigner matrices," *Communications in Mathematical Physics*, vol. 388, no. 2, pp. 1005–1048, 2021.
- [19] R. Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, 2018, vol. 47.