Chapter 19
Projection Test with Sparse Optimal Direction for High-Dimensional One Sample Mean Problem

Wanjun Liu and Runze Li

Abstract Testing whether the mean vector from some population is zero or not is a fundamental problem in statistics. In the high-dimensional regime, where the dimension of data p is greater than the sample size n, traditional methods such as Hotelling's T^2 test cannot be directly applied. One can project the high-dimensional vector onto a space of low dimension and then traditional methods can be applied. In this paper, we propose a projection test based on a new estimation of the optimal projection direction $\Sigma^{-1}\mu$. Under the assumption that the optimal projection $\Sigma^{-1}\mu$ is sparse, we use a regularized quadratic programming with nonconvex penalty and linear constraint to estimate it. Simulation studies and real data analysis are conducted to examine the finite sample performance of different tests in terms of type I error and power.

19.1 Introduction

One-sample mean vector test or two-sample test on the equality of two means is a fundamental problem in high-dimensional data analysis. These tests are commonly encountered in genome-wide association studies. For instance, [6] performed a hypothesis testing to identify sets of genes which are significant with respect to certain treatments in a genetics research. Reference [21] applied various tests to the bipolar disorder dataset from a genome-wide association study collected by [7] in which one would like to test whether there is any association between a disease and a large number of genetic variants. In these applications, the dimension of the data p is often much larger than the sample size n. Traditional methods such as Hotelling's T^2 test [13] either cannot be directly applied or have low power against the alternative.

W. Liu (✉) · R. Li
Department of Statistics, Pennsylvania State University,
University Park 16802-2111, USA
e-mail: wxl204@psu.edu

R. Li
e-mail: rzli@psu.edu

J. Fan and J. Pan (eds.), *Contemporary Experimental Design, Multivariate Analysis and Data Mining*,
https://doi.org/10.1007/978-3-030-46161-4_19

Suppose that a random sample $x_1, \ldots, x_n$ from a p-dimensional population x with finite mean $\mathrm{E}(x) = \mu$ and positive definite covariance matrix $\mathrm{cov}(x) = \Sigma$. Of interest is to test the following hypothesis

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu \neq \mu_0, \tag{19.1}$$

for some known vector μ_0. This problem is typically referred to as the one-sample hypothesis testing problem in multivariate analysis and has been extensively studied when $p < n$ and p is fixed. Without loss of generality, we assume $\mu_0 = 0$ and the one-sample problem (19.1) becomes

$$H_0 : \mu = 0 \quad \text{versus} \quad H_1 : \mu \neq 0. \tag{19.2}$$

In most cases, the test statistic constructed for one-sample problem can be easily extended to two-sample problem and the theories hold as well. For this reason, we only focus on the one-sample problem (19.2) and assume $\mu_0 = 0$. Let $\bar{x}$ and S be the sample mean vector and the sample covariance matrix respectively,

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad S = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^\top. \tag{19.3}$$

The Hotelling's T^2 statistic [13] for problem (19.2) is $T^2 = n\bar{x}^\top S^{-1}\bar{x}$. If $x_1, \ldots, x_n$ are normally distributed, under H_0, then we have $(n-p)/\{(n-1)p\}T^2$ follows $F_{p,n-p}$, the F distribution with degrees of freedom p and $n-p$. The Hotelling's T^2 requires that the sample covariance matrix S is invertible and cannot be directly used in high-dimensional setting where $p > n$. Despite the singularity of S, it has been observed that the power of the Hotelling's T^2 test can be adversely affected even when $p < n$, if S is nearly singular; see [1, 17].

Several one-sample tests for high-dimensional data have been proposed recently. These tests can be roughly classified into three types. The first type is the sum-of-squares-type test which is based on the sum-of-squares of the sample mean and can be regarded as modified versions of Hotelling's T^2 test. These tests simply replace S by some invertible matrix such as identity matrix I or diagonal matrix, leading to a sum-of-squares test statistic. Bai and Saranadasa [1] proposed the following test statistic for one-sample problem, in which S is substituted by identity matrix I,

$$T_{BS} = \bar{x}^\top\bar{x} - \mathrm{tr}S/n.$$

The test statistic T_{BS} can be regarded as unscaled distance $\bar{x}^\top\bar{x}$ with offset $\mathrm{tr}S/n$. Bai and Saranadasa [1] established its asymptotic normal null distribution when $p/n \to c$ for some $c > 0$. Chen and Qin [6] further studied an equivalent form of T_{BS}:

$$T_{CQ} = \frac{1}{n(n-1)}\sum_{i\neq j}^{n} x_i^\top x_j.$$

under a different set of assumptions on population. Neither T_{BS} nor T_{CQ} is invariant under different scales. To get rid of the unit effect, [19] replaced S with diagonal matrix D, where $D = \text{diag}(S)$ is a diagonal matrix with diagonal elements from the sample covariance matrix S. The test statistic in [19] is defined as

$$T_{SD} = n\bar{x}^\top D^{-1}\bar{x} - (n-1)p/(n-3),$$

which is also asymptotically normally distributed under null hypothesis.

The second type is the maximum-type test. Cai et al. [4] introduced a test that is based on a linear transformation of the data by the precision matrix $\Omega = \Sigma^{-1}$ which incorporates the correlations among the variables. Given that the precision matrix $\Omega = (\omega_{ij})_{p\times p}$ is known, the test statistic is defined as

$$T_{CLX} = n \max_{1\le j\le p} (\Omega\bar{x})_j^2/\omega_{jj}. \tag{19.4}$$

If Ω is known to be sparse, then the CLIME estimator [3] can be used to estimate Ω directly. Otherwise, Ω can be estimated by the inverse of the adaptive thresholding estimator of Σ [2]. Under H_0, the test statistic T_{CLX} converges to the type I extreme value distribution. Chen et al. [5] proposed a test that removes components that are estimated to be zero via thresholding. The motivation is that zero components are expected to contribute little to the squared sample mean and those smaller than a given threshold can be ignored. The test statistic with index s is defined as

$$T_{CLZ}(s) = \sum_{j=1}^{p}\left\{\frac{n\bar{x}_j^2}{\sigma_{jj}} - 1\right\} I\left\{\frac{n\bar{x}_j^2}{\sigma_{jj}} > \lambda_p(s)\right\},$$

where the threshold level is set to be $\lambda_p(s) = 2s\log p$ for some $s \in (0, 1)$. Since the optimal choice of the threshold is unknown, [5] further proposed using s that results in the largest value of $T_{CLZ}(s)$ as the final test statistic,

$$T_{CLZ} = \max_{s\in(0,1-\eta)} \{T_{CLZ}(s) - \widehat{\mu}_{CLZ,0}(s)\}/\widehat{\sigma}_{CLZ,0}(s),$$

for some $\eta \in (0, 1)$, where $\widehat{\mu}_{CLZ,0}(s)$ and $\widehat{\sigma}_{CLZ,0}(s)$ are estimates of the mean and standard deviation of $T_{CLZ}(s)$ under H_0. The asymptotic null distribution of T_{CLZ} is the Gumbel distribution.

The third type is the projection test. The idea is to project the high-dimensional vector x onto a space of low dimension and then traditional methods such as Hotelling's T^2 can be applied. Lauter [14] proposed the following procedure for the one-sample normal mean problem based on left-spherical distribution theory [11, 12]. Consider the linear score $z = (z_1, \ldots, z_n)^\top = Xd$, where d is a $p \times 1$ weight vector depending on X only through $X^\top X$ and $d \neq 0$ with probability 1. Then one can perform the one-sample t-test using $z_1, \ldots, z_n$. Lauter [14] also proposed different ways to obtain the weight vector d. For example, d can take the form of $d = (\text{diag}(X^\top X))^{-1/2}$, or be the eigenvector corresponding to the largest eigenvalue $\lambda_{\max}$ for the following eigenvalue problem $(X^\top X)d = \text{diag}(X^\top X)d\lambda_{\max}$. Lopes et al. [16] proposed a test based on random projection. Let P_k be a $p \times k$ random matrix whose entries are randomly drawn from the $N(0, 1)$ distribution. Define $y_i = P_k^\top x_i$, $i = 1, \ldots, n$. The random projection test T_{RP} in [16] is defined as

$$T_{RP} = n\bar{y}^\top S_y^{-1}\bar{y} = n\bar{x}^\top P_k(P_k^\top S P_k)^{-1}P_k^\top \bar{x},$$

where $\bar{y}$ and S_y are the sample mean and sample covariance matrix of $y_1, \ldots, y_n$. As a result, this random projection test is the Hotelling's T^2 test with $y_1, \ldots, y_n$ and is an exact test if x_i's are normally distributed. Lopes et al. [16] also proposed a test that utilizes multiple projection to improve the power of random projection test. The idea is generating the projection matrix P_k multiple times and using their average as the final projection matrix.

These types of tests are powerful only against certain alternatives. For example, if the true mean μ is dense in the sense that there is a large proportion of small to moderate nonzero components, then sum-of-squares-type test is more powerful. In contrast, if the true mean μ is sparse in the sense that there are only few nonzero components with large magnitude in μ, then the maximum-type test is more powerful. In practice, since the true alternative hypothesis is unknown, it is unclear how to choose a powerful test. Furthermore, there are denser and intermediate situations in which neither type of test is powerful [21].

Li et al. [15] studied the projection test and derived the optimal projection direction which leads to the best power under alternative hypothesis. However, the estimation of the optimal projection direction has not been systematically studied yet. This paper aims to fill this gap by studying how to construct a sparse optimal projection test to achieve better power. We propose an estimation procedure of the sparse optimal projection direction by regularized quadratic programming with nonconvex penalty and linear constraint. We further examine the finite sample performance of the proposed procedure and illustrate it by an empirical analysis of a real data set.

The rest of this paper is organized as follows. In Sect. 19.2, we propose a new projection test with the optimal projection being estimated by the regularized quadratic programming. In Sect. 19.3, simulation studies are conducted to examine the finite sample performance of different tests in terms of type I error and power. In Sect. 19.4, we apply various tests to a real data example, which shows that the proposed projection test is more powerful than existing tests.

19.2 Projection Test with Sparse Optimal Direction

Li et al. [15] proposed an exact projection test using the optimal projection direction. They showed that the optimal choice of k in P_k is 1 and the optimal projection is $\Sigma^{-1}\mu$ in the sense that the power is maximized. Let $\theta = \Sigma^{-1}\mu$ and $y_i = \theta^\top x_i, i = 1, \ldots, n$. The projection Hotelling's T^2 test is

$$T_\theta^2 = n\bar{x}^\top \theta(\theta^\top S\theta)^{-1}\theta^\top \bar{x},$$

which follows the $F_{1,n-1}$ distribution under H_0 and normality assumption. It is equivalent to the one-sample t-test based on $y_1, \ldots, y_n$. In order to control the type I error, [15] also proposed a data-splitting strategy to estimate the optimal projection direction and obtained an exact t-test. The entire sample is randomly partitioned into two separate sets $\mathscr{S}_1 = \{x_1, \ldots, x_{n1}\}$ and $\mathscr{S}_2 = \{x_{n1+1}, \ldots, x_n\}$. Set $\mathscr{S}_1$ is used to estimate the projection direction θ and set $\mathscr{S}_2$ is used to construct the test statistic T_θ^2. To estimate θ, a ridge-type estimator is constructed $\widehat{\theta} = (S_1 + \lambda D_1)^{-1}\bar{x}_1$, where $\bar{x}_1$ and S_1 are the sample mean and the sample covariance matrix computed from $\mathscr{S}_1$ and $D_1 = \text{diag}(S_1)$, the diagonal matrix of S_1. Therefore, the estimator $\widehat{\theta}$ is independent of set $\mathscr{S}_2$. Then the data points from $\mathscr{S}_2$ are projected onto a 1-dimensional space by left-multiplying $\widehat{\theta}$. The one-sample t-test is performed based on the new data points $\widehat{\theta}^\top x_{n_1+1}, \ldots, \widehat{\theta}^\top x_n$. In order to have high power, [15] recommended to use $n_1 = \lfloor \kappa n \rfloor$ with $\kappa \in [0.4, 0.6]$ and $\lambda = n_1^{-1/2}$ in practice based on their empirical study. If κ is small, only a small portion of sample is used to estimate the optimal projection and the estimator is not accurate. If κ is large, only a small portion of sample is used to perform the test. As a result, a too small or too large κ leads to significant loss in the power of the test. The advantage of the data-splitting procedure is that we can obtain an exact t-test, meanwhile we may lose power since the sample in $\mathscr{S}_1$ is discarded when performing the test.

We propose a new estimation of the optimal projection under the assumption that the optimal projection $\Sigma^{-1}\mu$ is sparse. The assumption that the optimal projection direction is sparse is relatively mild and can be satisfied in different scenarios. For example, if Σ has the autocorrelation structure and μ is sparse and then the optimal projection $\Sigma^{-1}\mu$ is sparse. Another example is that if Σ has the compound symmetry structure and μ is sparse and then $\Sigma^{-1}\mu$ is approximately sparse in the sense that the first few entries in $\Sigma^{-1}\mu$ dominate the rest entries. Note that it is the direction rather than the magnitude of the projection $\Sigma^{-1}\mu$ that matters. In other words, $\Sigma^{-1}\mu$ and $a\Sigma^{-1}\mu$ have exactly the same performance for the one-sample problem (19.2), where a is some positive number. We observe that $\beta^* = \Sigma^{-1}\mu/\mu^\top\Sigma^{-1}\mu$, which is proportional to the optimal projection, is the solution to the following problem

$$\min_\beta \frac{1}{2}\beta^\top \Sigma\beta \text{ subject to } \mu^\top\beta = 1.$$

Based on the above observation, we propose the following estimation based on a regularized quadratic programming with nonconvex penalty and linear constraint,

$$\begin{aligned} &\min_{\beta} \frac{1}{2}\beta^\top S_1 \beta + \sum_{j=1}^{p} p_\lambda(|\beta_j|) \\ &\text{subject to } \bar{x}_1^\top \beta = 1, \end{aligned} \tag{19.5}$$

where $\bar{x}_1$ and S_1 are computed from set $\mathscr{S}_1$, $\beta = (\beta_1, \ldots, \beta_p)^\top \in \mathbb{R}^p$ and $p_\lambda(\cdot)$ is taken to be the smoothly clipped absolute deviation (SCAD) penalty [9]. Its first derivative is defined to be

$$p'_\lambda(|t|) = \lambda \left\{ I(|t| \leq \lambda) + \frac{(a\lambda - |t|)_+}{(a-1)\lambda} I(|t| > \lambda) \right\},$$

where $a = 3.7$, $I(\cdot)$ is the indicator function and b_+ stands for the positive part of b. To solve the high-dimensional nonconvex optimization problem (19.5), we apply the local linear approximation (LLA) algorithm proposed in [22]. The idea is to approximate the nonconvex penalty by its first order expansion. Given the current solution $\beta^{(k)}$, (19.5) can be approximated by

$$\begin{aligned} &\min_{\beta} \frac{1}{2}\beta^\top S_1 \beta + \sum_{j=1}^{p} p'_\lambda(|\beta_j^{(k)}|)|\beta_j|, \\ &\text{subject to } \bar{x}_1^\top \beta = 1. \end{aligned}$$

Let

$$Q(\beta|\beta^{(k)}, \lambda) = \frac{1}{2}\beta^\top S_1 \beta + \sum_{j=1}^{p} p'_\lambda(|\beta_j^{(k)}|)|\beta_j|.$$

Wang et al. [20] and Fan et al. [10] studied how to implement the LLA under high-dimensional regression settings to obtain a sparse solution with oracle property. Here we apply their strategy for the above problem. Starting with initial value 0, we propose a two-step LLA estimator, which consists of the following two steps:

$$\textbf{Step 1}: \widehat{\beta}^{(1)} = \underset{\{\beta: \bar{x}_1^\top \beta = 1\}}{\text{argmin}} \; Q(\beta|0, \tau\lambda);$$

$$\textbf{Step 2}: \widehat{\beta} = \underset{\{\beta: \bar{x}_1^\top \beta = 1\}}{\text{argmin}} \; Q(\beta|\widehat{\beta}^{(1)}, \lambda).$$

The solution $\widehat{\beta}$ in step 2 is our final estimator. Typically, we choose τ to be some small number such as $\tau = 1/\log n_1$ or $\tau = \lambda$. Instead of using the ridge-type estimator $\widehat{\theta} = (S_1 + \lambda D_1)^{-1}\bar{x}_1$, we use our two-step LLA estimator $\widehat{\beta}$ to carry out the projection test with data splitting. It can be shown that the resulting LLA estimator is consistent under relatively mild conditions and thus the asymptotic power is valid for our new test with the data-splitting procedure. We call this new test LLA projection test.

19.3 Simulation Studies

In this section, we conduct numerical studies to examine the finite sample performance of different tests including the proposed LLA projection test for the one-sample problem. The LLA projection test is the same as that in [15] except that we use the LLA estimator as the projection direction. More specifically, we compare the LLA projection test with the ones proposed by [1, 6, 8, 14, 15]. We denote them by Li2015, D1958, BS1996, CQ2010 and L1996, respectively. We also compare the new test with the tests proposed in [19]. The authors considered two versions of their test, one with modification and one without modification, denoted by SD2008w and SD2008wo, respectively. Lopes et al. [16] proposed a single random projection test, labeled as LWJ2011.

We generate a random sample of size n from $N(\mu, \Sigma)$ with $\mu = c \cdot (1_{s_0}^\top, 0_{p-s_0}^\top)^\top$ and $s_0 = 10$. We set $c = 0, 0.5$ and 1 to examine the type I error rate and the power of the tests. For $\rho \in (0, 1)$, we consider the following three covariance structures:

(1) Compound symmetry with $\Sigma_1 = (1-\rho)I + \rho 11^\top$;
(2) Autocorrelation with $\Sigma_2 = (\rho^{|i-j|})_{i,j}$;
(3) Composite structure with $\Sigma_3 = 0.5\Sigma_1 + 0.5\Sigma_2$.

We consider $\rho = 0.25, 0.5, 0.75$ and 0.95 to examine the influence of correlation on the power of the test. We set sample size $n = 40, 160$ and dimension $p = 400, 1600$. We split the data set by setting $n_1 = \lfloor n\kappa \rfloor$ with $\kappa = 0.4$, where $\lfloor \cdot \rfloor$ is the rounding operator. To this end, we replace sample covariance matrix S_1 by $S_\phi = S_1 + \phi I$ with a small positive number $\phi = \sqrt{\log p/n_1}$. Such a perturbation does not noticeably affect the computational accuracy of the final solution and all the theoretical properties hold as well when $\phi \le \sqrt{\log p/n_1}$. All simulation results are based on 10,000 independent replicates. These results are summarized in Tables 19.1, 19.2 and 19.3.

Tables 19.1, 19.2 and 19.3 clearly indicate that the LLA projection test and the tests in [14–16] keep the type I error very well. This is not surprising since all these tests are exact tests. All other tests do not keep the type I error rate well because their critical values are determined from the asymptotic distributions. Next we compare the power of the LLA projection test with other existing methods. It can be seen from Tables 19.1, 19.2 and 19.3, the power of the tests strongly relies on the covariance structure as well as the values of ρ and c.

Table 19.1 reports the results for the compound symmetry covariance structure Σ_1. We first compare the LLA projection test and the Li2015 test since these two tests are of the same flavor but using different methods to estimate the projection direction. When we have relatively large sample size $n = 160$, both of the tests have high power and the LLA projection test slightly improves the performance of Li2015

Table 19.1 Power comparison for $N(\mu, \Sigma_1)$ (values in table are in percentage)

	$c = 0$				$c = 0.5$				$c = 1$			
ρ	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95
$n = 40,\ p = 400$												
LLA	4.98	4.50	4.94	5.19	71.53	89.92	99.00	99.96	99.97	99.88	99.99	100.0
Li2015	5.16	4.47	4.88	4.90	50.22	70.74	94.04	100.0	98.61	99.53	100.0	100.0
D1958	6.77	6.22	5.71	5.49	12.63	8.13	6.98	6.46	80.44	22.71	13.06	10.23
BS1996	7.73	7.80	7.79	7.80	14.64	10.55	9.39	9.11	88.21	30.28	18.51	15.33
CQ2010	7.72	7.82	7.79	7.77	14.64	10.50	9.41	9.11	88.18	30.22	18.50	15.32
SD2008w	4.20	1.71	0.52	0.15	7.97	2.29	0.63	0.22	54.21	6.41	1.29	0.36
SD2008wo	8.48	8.21	7.87	7.71	16.34	11.15	9.53	8.96	90.25	32.69	18.93	15.06
L1996	5.18	5.18	5.17	5.15	5.66	5.21	5.17	5.11	6.25	5.59	5.31	5.22
LJW2011	5.01	4.99	4.86	5.03	13.80	20.65	40.58	98.34	54.05	74.46	95.94	100.0
$n = 40,\ p = 1600$												
LLA	5.22	4.99	5.21	5.08	50.43	79.97	98.51	99.98	99.92	99.94	99.99	100.0
Li2015	5.01	4.71	5.06	4.94	14.62	23.71	54.68	98.14	71.49	81.98	95.74	100.0
D1958	6.93	6.19	5.73	5.48	7.81	6.71	5.96	5.73	12.45	8.12	6.96	6.45
BS1996	7.74	7.79	7.78	7.79	8.85	8.30	8.12	8.06	14.47	10.42	9.35	8.92
CQ2010	7.76	7.80	7.77	7.77	8.88	8.32	8.14	8.05	14.49	10.41	9.34	8.92
SD2008w	2.76	0.69	0.14	0.00	3.22	0.73	0.17	0.00	5.26	0.97	0.20	0.01
SD2008wo	8.37	8.12	7.86	7.71	9.67	8.70	8.24	7.94	15.79	11.14	9.52	8.78
L1996	5.15	5.15	5.15	5.15	5.30	5.22	5.19	5.18	5.50	5.29	5.23	5.18
LJW2011	4.65	5.08	4.99	4.95	6.77	7.68	11.52	52.16	14.49	20.55	42.17	98.29
$n = 160,\ p = 400$												
LLA	4.77	5.10	4.96	4.83	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Li2015	4.97	4.89	4.80	4.99	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
D1958	5.74	5.26	4.89	4.77	87.20	19.91	12.09	9.73	100.0	100.0	99.95	89.50
BS1996	6.66	6.71	6.69	6.71	94.00	26.42	16.69	13.78	100.0	100.0	100.0	99.41
CQ2010	6.66	6.71	6.69	6.71	94.02	26.45	16.69	13.76	100.0	100.0	100.0	99.39
SD2008w	3.11	0.99	0.34	0.07	50.59	3.63	0.72	0.17	100.0	92.93	7.69	1.27
SD2008wo	6.87	6.83	6.71	6.65	94.39	26.76	16.83	13.72	100.0	100.0	100.0	99.36
L1996	4.76	4.74	4.74	4.73	5.98	5.23	4.95	4.87	7.08	5.88	5.32	5.11
LJW2011	4.81	5.15	4.99	4.84	98.07	99.92	100.0	100.0	100.0	100.0	100.0	100.0
$n = 160,\ p = 1600$												
LLA	4.63	4.99	4.79	4.96	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Li2015	4.91	5.43	5.40	4.74	98.84	99.92	100.0	100.0	100.0	100.0	100.0	100.0
D1958	5.76	5.22	4.87	4.77	11.15	7.32	6.11	5.68	93.07	19.49	11.60	9.41
BS1996	6.71	6.69	6.69	6.69	13.09	9.46	8.37	8.11	97.90	26.09	16.49	13.60
CQ2010	6.71	6.69	6.70	6.70	13.10	9.47	8.37	8.11	97.91	26.11	16.48	13.61
SD2008w	2.10	0.40	0.05	0.02	3.82	0.53	0.05	0.02	29.18	1.19	0.15	0.03
SD2008wo	6.90	6.82	6.71	6.66	13.48	9.46	8.43	8.09	98.05	26.51	16.39	13.53
L1996	4.72	4.73	4.73	4.74	4.76	4.71	4.69	4.71	4.93	4.73	4.73	4.71
LJW2011	5.23	4.83	4.80	4.70	34.28	55.48	91.85	100.0	98.27	99.95	100.0	100.0

Table 19.2 Power comparison for $N(\mu, \Sigma_2)$ (values in table are in percentage)

	$c = 0$				$c = 0.5$				$c = 1$			
ρ	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95
$n = 40$, $p = 400$												
LLA	5.18	5.19	5.26	4.78	61.15	50.17	36.46	27.19	100.0	99.99	99.67	97.72
Li2015	5.29	4.46	5.16	4.81	46.27	35.27	21.13	13.86	99.98	99.53	91.08	68.03
D1958	5.06	4.97	4.75	5.30	89.47	77.24	51.45	17.29	100.0	100.0	99.96	84.57
BS1996	5.57	5.57	5.46	6.86	90.19	78.40	53.88	20.81	100.0	100.0	99.99	88.16
CQ2010	5.59	5.57	5.44	6.85	90.16	78.39	53.83	20.81	100.0	100.0	99.99	88.15
SD2008w	3.75	3.68	3.30	2.72	84.86	70.93	44.71	9.94	100.0	100.0	99.85	68.93
SD2008wo	7.25	7.28	7.61	8.52	90.57	80.54	57.97	23.86	100.0	100.0	99.96	87.61
L1996	4.69	4.67	4.93	4.96	36.78	28.64	16.49	6.63	92.38	77.29	40.81	9.26
LJW2011	5.52	5.11	5.00	4.97	12.71	12.15	11.51	15.28	44.17	43.04	42.40	60.42
$n = 40$, $p = 1600$												
LLA	5.25	5.19	5.09	5.12	38.03	30.96	22.88	16.49	100.0	99.94	98.81	91.04
Li2015	4.61	4.95	5.30	4.92	17.85	14.57	9.55	6.10	94.90	84.59	58.09	22.43
D1958	4.91	5.14	4.88	4.74	48.45	37.63	23.47	9.96	99.99	99.91	94.58	42.28
BS1996	5.05	5.46	5.36	5.49	49.13	38.40	24.63	11.40	99.99	99.91	94.96	45.81
CQ2010	5.08	5.48	5.29	5.50	49.26	38.35	24.60	11.44	99.99	99.91	94.94	45.74
SD2008w	1.77	1.91	2.04	1.81	30.97	22.82	12.73	3.66	99.92	99.03	86.28	23.53
SD2008wo	7.04	7.13	7.19	7.55	53.38	43.79	29.11	14.45	99.98	99.79	95.07	50.80
L1996	4.92	5.11	5.08	4.99	15.69	13.57	9.77	6.17	45.99	34.47	19.55	7.99
LJW2011	4.61	4.99	4.87	4.89	6.04	6.47	6.17	6.68	11.71	12.12	11.46	13.14
$n = 160$, $p = 400$												
LLA	5.10	4.94	4.82	5.03	100.0	99.97	99.37	99.98	100.0	100.0	100.0	100.0
Li2015	5.33	4.68	5.03	5.16	99.99	99.43	89.97	96.04	100.0	100.0	100.0	100.0
D1958	4.61	4.97	5.12	5.34	100.0	100.0	100.0	85.83	100.0	100.0	100.0	100.0
BS1996	5.03	5.50	5.83	6.61	100.0	100.0	100.0	89.10	100.0	100.0	100.0	100.0
CQ2010	5.03	5.49	5.83	6.62	100.0	100.0	100.0	89.10	100.0	100.0	100.0	100.0
SD2008w	4.20	4.42	4.17	2.73	100.0	100.0	100.0	72.60	100.0	100.0	100.0	100.0
SD2008wo	5.41	5.78	6.19	6.93	100.0	100.0	100.0	88.85	100.0	100.0	100.0	100.0
L1996	4.87	4.71	4.70	5.00	89.99	71.70	34.04	7.34	100.0	100.0	71.60	10.28
LJW2011	4.65	4.95	4.75	5.27	89.44	85.36	80.43	98.54	100.0	100.0	100.0	100.0
$n = 160$, $p = 1600$												
LLA	4.85	5.04	4.92	4.79	100.0	99.94	97.61	93.27	100.0	100.0	100.0	100.0
Li2015	5.24	4.83	4.97	5.01	97.18	88.69	61.66	35.37	100.0	100.0	100.0	99.60
D1958	4.73	4.72	4.99	5.11	99.99	99.89	95.03	42.55	100.0	100.0	100.0	100.0
BS1996	4.86	5.00	5.30	5.98	100.0	99.90	95.35	45.67	100.0	100.0	100.0	100.0
CQ2010	4.86	4.98	5.29	5.99	100.0	99.90	95.35	45.66	100.0	100.0	100.0	100.0
SD2008w	3.47	3.46	3.57	2.70	100.0	99.83	93.02	29.91	100.0	100.0	100.0	99.99
SD2008wo	5.40	5.48	5.65	6.33	100.0	99.87	95.47	46.65	100.0	100.0	100.0	100.0
L1996	5.27	5.08	4.84	4.78	42.34	31.61	16.88	6.42	97.49	83.24	39.61	9.04
LJW2011	4.86	4.67	4.56	5.45	25.35	24.48	23.41	37.24	92.06	90.94	90.47	98.77

Table 19.3 Power comparison for $N(\mu, \Sigma_3)$ (values in table are in percentage)

	$c = 0$				$c = 0.5$				$c = 1$			
ρ	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95
$n = 40$, $p = 400$												
LLA	5.13	4.61	5.68	4.93	60.97	60.90	57.77	55.26	99.99	99.86	99.81	99.73
Li2015	5.08	4.76	4.76	4.73	43.34	42.40	35.76	28.98	98.94	98.15	97.05	94.31
D1958	6.82	6.68	6.38	6.02	26.75	12.64	9.38	8.05	99.93	78.81	38.79	21.82
BS1996	7.37	7.73	7.75	7.86	29.27	14.57	11.73	10.39	99.95	86.13	48.62	29.47
CQ2010	7.40	7.71	7.71	7.89	29.30	14.54	11.68	10.35	99.96	86.09	48.67	29.42
SD2008w	5.39	4.27	2.71	1.32	20.68	8.06	4.18	1.88	99.61	52.80	15.70	5.47
SD2008wo	8.15	8.36	8.34	8.27	33.33	15.93	12.72	10.85	99.96	88.48	52.41	32.24
L1996	5.07	5.17	5.13	5.19	6.12	5.57	5.33	5.22	8.10	6.30	5.80	5.57
LJW2011	4.86	5.20	4.95	5.06	12.83	14.60	15.63	26.31	48.15	53.40	60.76	86.04
$n = 40$, $p = 1600$												
LLA	4.96	5.13	5.06	4.62	40.72	42.46	42.69	36.21	99.95	99.82	99.79	99.21
Li2015	4.60	5.13	5.16	5.05	12.67	13.26	11.95	8.45	70.81	69.11	66.18	46.23
D1958	7.17	6.90	6.48	6.19	9.58	7.78	7.18	6.81	27.30	12.45	9.31	8.28
BS1996	7.74	7.80	7.80	7.78	10.34	9.01	8.46	8.30	30.00	14.44	11.59	10.42
CQ2010	7.76	7.81	7.78	7.72	10.28	8.95	8.46	8.29	30.05	14.45	11.54	10.40
SD2008w	4.19	2.70	1.44	0.73	5.79	3.18	1.61	0.77	15.53	5.26	2.19	1.04
SD2008wo	8.48	8.43	8.32	8.19	11.70	9.76	9.05	8.82	34.82	15.94	12.59	11.17
L1996	5.10	5.14	5.19	5.20	5.39	5.27	5.21	5.21	5.71	5.44	5.30	5.31
LJW2011	5.00	4.84	4.79	5.23	6.68	6.81	7.05	8.80	13.07	14.15	16.65	23.86
$n = 160$, $p = 400$												
LLA	5.35	5.05	4.75	5.42	100.0	99.99	100.0	100.0	100.0	100.0	100.0	100.0
Li2015	5.01	5.03	4.86	5.36	100.0	100.0	99.91	99.99	100.0	100.0	100.0	100.0
D1958	5.98	5.73	5.44	5.04	100.0	83.26	33.73	18.91	100.0	100.0	100.0	100.0
BS1996	6.45	6.67	6.73	6.72	100.0	90.99	43.97	25.76	100.0	100.0	100.0	100.0
CQ2010	6.47	6.67	6.72	6.72	100.0	91.00	43.98	25.74	100.0	100.0	100.0	100.0
SD2008w	4.92	3.04	1.81	0.82	99.99	48.74	10.03	2.98	100.0	100.0	99.99	78.09
SD2008wo	6.70	6.80	6.89	6.85	100.0	91.47	45.28	26.38	100.0	100.0	100.0	100.0
L1996	4.70	4.75	4.75	4.74	7.09	6.01	5.56	5.20	10.27	7.09	6.21	5.83
LJW2011	5.36	5.28	4.97	4.91	94.31	96.07	97.07	100.0	100.0	100.0	100.0	100.0
$n = 160$, $p = 1600$												
LLA	4.99	4.85	5.10	4.63	100.00	100.0	99.99	99.96	100.0	100.0	100.0	100.0
Li2015	5.22	5.02	5.16	4.38	97.38	97.63	93.89	80.36	100.0	100.0	100.0	100.0
D1958	6.12	5.77	5.45	5.25	23.85	11.17	8.51	7.42	100.0	91.94	33.55	20.31
BS1996	6.66	6.68	6.70	6.77	26.24	13.24	10.44	9.46	100.0	97.40	44.83	27.33
CQ2010	6.67	6.68	6.70	6.77	26.26	13.22	10.41	9.44	100.0	97.42	44.81	27.32
SD2008w	4.13	2.06	0.84	0.39	15.66	3.70	1.26	0.58	100.0	29.28	4.52	1.28
SD2008wo	6.83	6.80	6.85	6.86	27.30	13.49	10.70	9.61	100.0	97.42	46.28	28.33
L1996	4.69	4.71	4.75	4.74	4.93	4.78	4.71	4.70	5.44	4.91	4.79	4.74
LJW2011	4.97	5.30	5.41	5.20	28.38	33.95	40.01	68.67	94.88	97.86	99.36	99.99

test. When we have a relatively small sample size $n = 40$, LLA projection test can dramatically improve the performance of Li2015 test especially when the signal is not strong ($c = 0.5$). A weaker correlation ρ results in a more significant improvement. This is because a small correlation makes the optimal direction $\Sigma^{-1}\mu$ closer to a sparse direction. When $c = 0.5$, the power of both tests increases significantly as ρ increases. As the value of c increases from 0.5 to 1, the power of the two tests increases dramatically. As the dimension p increases, there is a downward trend for the two tests. Even in the most challenging case $(n, p, c) = (40, 1600, 0.5)$, our LLA projection test has high power as well and is much powerful than Li2015 test. These two tests outperform all other tests. Some of the tests, such as D1958, BS1996, CQ2010 and SD2008w, tend to become less powerful when ρ increases. This is because these methods ignore the correlation among the variables and therefore their overall performance is not satisfactory.

Table 19.2 reports the results for the autocorrelation covariance structure Σ_2. Under this setting, the LLA projection test improves the performance of Li2015 test in all the combinations of c and ρ. In particular, the LLA projection test improves the performance of Li2015 dramatically when the sample size is relatively small and the correlation is large. For example, when $(n, p, c, \rho) = (40, 1600, 1, 0.95)$, the LLA test improves the power from 22.42% to 91.04%. The D1958, BS1996, CQ2010 and SD2008wo have more satisfactory performance than LLA test when $(n, c) = (40, 0.5)$ and ρ is not 0.95. Notice that the D1958, BS1996, CQ2010 and SD2008wo tests ignore the correlation among variables and replace Σ^{-1} by diagonal matrix. When Σ has the autocorrelation structure, its inverse is a 3-banded matrix – only its diagonal and first off-diagonal elements are nonzero. As a result, replacing Σ^{-1} by identity matrix does not lose much information. This explains why tests of D1958, BS1996, CQ2010 and SD2008wo have more satisfactory performance when Σ has autocorrelation structure and ρ is low. It is also observed that the power of these four tests decreases significantly as the correlation increases and become less powerful than the LLA test when $\rho = 0.95$. This is not surprising since all the four tests ignore the correlations among the variables. In general, the proposed test is preferred if Σ^{-1} is far away from identity matrix.

Table 19.3 reports the results for Σ_3. The LLA test is more powerful than Li2015 test in all the combinations of n, p, c, ρ and improves the power dramatically when ρ is large. The LLA test outperforms all other tests. The patterns for D1958, BS1996, CQ2010, SD2008w and SD2008wo are similar to the first scenario where $\Sigma = \Sigma_1$.

We also investigate the finite sample performance of the LLA projection test without the normality assumption. To this end, we generate random samples from the multivariate t distribution with degrees of freedom 6. To examine the robustness of the LLA test, we use the same critical values as those used in settings with normality assumption. Simulation results for Σ_2 are summarized in Table 19.4, from which it can be seen that the LLA test and Li2015 test can still retain the type I error rate very well. This implies that these two projection tests are not very sensitive to the normality assumption. All other alternative tests except for CQ2010 test fail to retain the type I error. In terms of power, LLA projection test is more powerful than Li2015 test in all combinations of n, p, c, ρ. For this autocorrelation covariance (i.e., Σ_2)

case, the LLA test and the CQ2010 test have similar performance and these two tests outperform all other tests. The overall patterns for Σ_1 and Σ_3 are similar to those in Tables 19.1 and 19.3. Results are not presented in this paper to save space.

19.4 Real Data Example

In this section, we apply the LLA projection test to a real dataset of high resolution micro-computed tomography. This dataset contains the bone density of 58 mice's skull of three different genotypes ("T0A0", "T0A1", and "T1A1") measured at different bone density levels in a genetic mutation study. For each mouse, bone density is measured for 16 different areas of its skull. For each area, bone volume is measured at density levels from 130 to 249. This dataset was collected at Center for Quantitative X-Ray Imaging at the Pennsylvania State University. See [18] for a detailed description of protocols. In this empirical analysis, we are interested in comparing the bone density patterns of two different areas in mice's skull. We compare the performance of the proposed LLA projection test with several existing methods. To emphasize the high-dimensionality nature of this dataset, we only use half sample of the dataset. We select the mice of the genotype "T0A1" and there are 29 samples available in the dataset, i.e., sample size $n = 29$. The two areas of the skull "Mandible" and "Nasal" are selected. We use all density levels from 130 to 249 for our analysis, hence dimension $p = 120$. We first take the difference of the bone density of the two selected areas at the corresponding density level for each subject since the two areas come from the same mouse. Then we normalize the bone density in the sense that $\frac{1}{n}\sum_{i=1}^{n} x_{ij}^2 = 1$ for all $1 \leq j \leq 120$.

We apply LLA projection test and several other existing methods to this dataset. Due to the relatively small sample size ($n = 29$), we opt to use slightly more data points to estimate the projection direction such that the estimator is reasonably well. As a result, we set $\kappa = 0.6$. The p-values are reported in the first row in Table 19.5. The p-values of all methods are 0, implying that the bone volume is significantly different. To see which test is more powerful, we also compute the p-values of these tests when we decrease the signals. Let $\bar{x}$ be the sample mean and $r_i = x_i - \bar{x}$ is the residual for the ith subject. Then a new observation $z_i = \delta\bar{x} + r_i$ is constructed for the ith subject. By the construction, a smaller δ results in a weaker signal and would make the test more challenging. Table 19.5 reports the p-values of all these tests for the new data z_i with $\delta = 1, 0.8, \ldots, 0.2$. As expected, the p-values of all tests increase as δ decreases. When $\delta = 0.8$ or 0.6, all these tests perform well and reject the null hypothesis at level 0.05. When $\delta = 0.4$, the Lauter's test fails to reject the null hypothesis. When $\delta = 0.2$, all the tests except for our method fail to reject the null hypothesis, which suggests that our method would perform well even though the signal is weak. Among those tests that fail to reject H_0 when $\delta = 0.2$, Li2015 projection test has the smallest p-value.

Table 19.4 Power comparison for $t_6(\mu, \Sigma_2)$ (Values in table are in percentage)

	$c = 0$				$c = 0.5$				$c = 1$			
ρ	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95
$n = 40,\ p = 400$												
LLA	4.74	4.44	4.81	5.19	46.94	38.23	27.23	19.02	99.95	99.52	96.50	91.21
Li2015	4.77	5.00	4.53	5.52	36.00	26.34	16.44	11.40	99.27	95.97	79.71	55.13
D1958	0.05	0.17	0.80	3.27	10.04	11.00	11.39	8.56	92.77	91.15	83.76	48.47
BS1996	0.08	0.25	1.02	4.65	12.42	13.43	14.05	11.29	94.45	92.97	86.68	55.35
CQ2010	5.49	5.71	5.83	6.77	68.47	55.52	36.29	15.44	100.0	99.97	97.97	64.15
SD2008w	0.04	0.05	0.37	1.34	7.13	8.03	8.00	3.84	91.59	89.34	79.00	33.06
SD2008wo	0.16	0.48	1.57	5.99	20.35	20.70	19.99	14.67	97.78	96.67	91.34	61.40
L1996	0.46	0.76	1.66	4.00	1.53	2.31	3.49	4.98	5.90	6.81	7.87	6.73
LJW2011	3.85	4.40	4.34	4.12	10.16	10.22	10.12	13.08	37.12	36.14	35.35	51.57
$n = 40,\ p = 1600$												
LLA	4.89	4.61	4.85	4.77	28.90	24.90	17.90	12.35	99.89	99.19	94.29	78.77
Li2015	5.24	4.58	5.08	5.37	13.82	11.03	8.69	5.40	83.00	69.65	44.60	17.05
D1958	0.00	0.00	0.02	1.13	0.00	0.00	0.22	2.09	7.62	7.79	9.38	9.44
BS1996	0.00	0.00	0.06	1.58	0.00	0.00	0.30	2.72	9.59	9.90	11.96	11.86
CQ2010	5.07	5.16	5.23	5.93	30.83	24.57	16.62	9.58	98.44	94.23	75.89	29.10
SD2008w	0.00	0.00	0.00	0.11	0.00	0.00	0.02	0.30	1.74	2.05	2.91	2.41
SD2008wo	0.00	0.00	0.13	2.51	0.00	0.05	0.57	4.55	18.33	18.81	19.84	17.45
L1996	0.05	0.12	0.40	2.08	0.10	0.18	0.52	2.30	0.21	0.40	0.80	2.87
LJW2011	4.26	4.24	4.22	4.18	5.60	5.31	5.49	6.01	10.32	9.83	9.97	11.53
$n = 160,\ p = 400$												
LLA	4.68	4.94	4.29	4.69	100.00	99.58	94.11	99.66	100.0	100.0	99.98	100.0
Li2015	4.77	4.98	4.82	4.95	99.58	96.61	78.27	88.18	100.0	100.0	99.97	100.0
D1958	0.41	1.03	2.41	4.33	99.52	99.02	95.86	53.53	99.99	99.99	99.99	99.96
BS1996	0.52	1.30	2.97	5.57	99.61	99.27	96.64	59.70	99.99	100.0	99.99	99.98
CQ2010	5.32	5.68	5.75	6.38	99.99	99.92	98.91	62.70	100.0	100.0	100.0	100.0
SD2008w	0.31	0.81	1.70	2.14	99.59	99.00	94.47	37.44	99.99	100.0	99.99	99.93
SD2008wo	0.77	1.48	3.28	6.09	99.80	99.46	96.90	61.76	100.0	100.0	100.0	99.99
L1996	0.92	1.60	2.87	4.35	6.34	8.40	9.53	5.81	24.25	25.56	22.83	8.63
LJW2011	4.55	4.42	4.36	4.38	82.42	76.72	70.40	95.73	100.0	100.0	100.0	100.0
$n = 160,\ p = 1600$												
LLA	5.19	5.08	5.35	4.85	99.95	98.95	89.73	79.71	100.0	100.0	100.0	100.0
Li2015	5.19	4.68	4.39	4.81	88.97	75.32	45.68	27.00	100.0	100.0	99.90	97.58
D1958	0.00	0.03	0.40	3.04	43.74	40.56	33.32	17.30	99.72	99.67	99.64	95.47
BS1996	0.00	0.07	0.47	3.80	47.26	43.91	36.50	19.82	99.78	99.74	99.75	96.40
CQ2010	4.98	4.93	5.16	6.08	98.83	95.34	75.76	27.93	100.0	100.0	100.0	98.82
SD2008w	0.00	0.00	0.16	1.22	33.20	30.69	24.45	9.06	99.59	99.57	99.37	88.00
SD2008wo	0.00	0.05	0.57	4.16	53.10	49.61	40.86	21.79	99.92	99.86	99.90	96.36
L1996	0.15	0.26	0.65	3.31	0.29	0.66	1.22	4.07	0.95	1.56	2.51	5.19
LJW2011	4.48	3.98	4.32	4.05	19.83	19.19	18.86	30.07	84.29	84.13	82.51	96.48

Table 19.5 Bone density dataset: p-value of one-sample test

δ	LLA	Li2015	D1958	BS1996	CQ2010	SD2008w	SD2008wo	L1996	LJW2011
1	0	0	0	0	0	0	0	0	0
0.8	0	0	0	0	0	0	0	0	0
0.6	0	0	0	0	0	0	0	0.0005	0
0.4	0	4×10^{-5}	0.0015	0	0	0.0088	0	0.6775	2×10^{-5}
0.2	0.0390	0.0906	0.2145	0.2710	0.2714	0.4136	0.2999	0.8870	0.3073

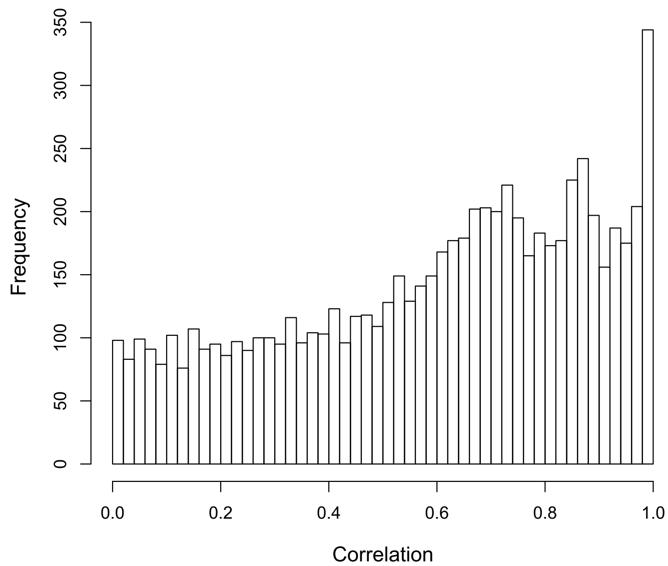

Fig. 19.1 Histogram of absolute values of paired sample correlations among bone densities at all different bone density levels

We plot the histogram of absolute values of paired sample correlations among all bone density levels in Fig. 19.1. It indicates that some bone density levels are highly correlated. This may explain why our method is more powerful than Dempster test, BS test and SD test since these methods do not take the dependence among variables into account.

Acknowledgments This work was supported by a NSF grant DMS 1820702 and a NIDA, NIH grant P50 DA039838. The content is solely the responsibility of the authors and does not necessarily represent the official views of NSF, NIH or NIDA.

References

1. Bai, Z., Saranadasa, H.: Effect of high dimension: by an example of a two sample problem. Statistica Sinica **6**, 311–329 (1996)
2. Cai, T., Liu, W.: Adaptive thresholding for sparse covariance matrix estimation. J. Am. Stat. Assoc. **106**(494), 672–684 (2011)
3. Cai, T., Liu, W., Luo, X.: A constrained ℓ_1 minimization approach to sparse precision matrix estimation. J. Am. Stat. Assoc. **106**(494), 594–607 (2011)

4. Cai, T., Liu, W., Xia, Y.: Two-sample test of high dimensional means under dependence. J. R. Stat. Soc. Series B (Statistical Methodology) **76**(2), 349–372 (2014)
5. Chen, S.X., Li, J., Zhong, P.-S.: Two-sample and ANOVA tests for high dimensional means. Ann. Stat. **47**(3), 1443–1474 (2019)
6. Chen, S.X., Qin, Y.-L.: A two-sample test for high-dimensional data with applications to gene-set testing. Ann. Stat. **38**(2), 808–835 (2010)
7. Consortium, W. T. C. C: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature **447**(7145), 661 (2007)
8. Dempster, A.P.: A high dimensional two sample significance test. Ann. Math. Stat. **29**(4), 995–1010 (1958)
9. Fan, J., Li, R.: Variable selection via nonconcave penalized likelihood and its oracle properties. J. Am. Stat. Assoc. **96**(456), 1348–1360 (2001)
10. Fan, J., Xue, L., Zou, H.: Strong oracle optimality of folded concave penalized estimation. Ann. Stat. **42**(3), 819 (2014)
11. Fang, K.-T., Kotz, S., Ng, K.: Symmetric Multivariate and Related Distributions. Chapman and Hall (1990)
12. Fang, K.-T., Zhang, Y.-T.: Generalized Multivariate Analysis. Science Press, Springer-Verlag, Beijing (1990)
13. Hotelling, H.: The generalization of student's ratio. Ann. Math. Stat. **2**(3), 360–378 (1931)
14. Lauter, J.: Exact t and F tests for analyzing studies with multiple endpoints. Biometrics **52**(3), 964–970 (1996)
15. Li, R., Huang, Y., Wang, L., Xu, C.: Projection Test for High-dimensional Mean Vectors with Optimal Direction (2015)
16. Lopes, M., Jacob, L., Wainwright, M.J.: A more powerful two-sample test in high dimensions using random projection. In: Advances in Neural Information Processing Systems, pp. 1206–1214 (2011)
17. Pan, G., Zhou, W.: Central limit theorem for hotellings T^2 statistic under large dimension. Ann. Appl. Probab. **21**(5), 1860–1910 (2011)
18. Percival, C.J., Huang, Y., Jabs, E.W., Li, R., Richtsmeier, J.T.: Embryonic craniofacial bone volume and bone mineral density in fgfr2+/p253r and nonmutant mice. Dev. Dyn. **243**(4), 541–551 (2014)
19. Srivastava, M.S., Du, M.: A test for the mean vector with fewer observations than the dimension. J. Multivar. Anal. **99**(3), 386–402 (2008)
20. Wang, L., Kim, Y., Li, R.: Calibrating nonconvex penalized regression in ultra-high dimension. Ann. Stat. **41**(5), 2505 (2013)
21. Xu, G., Lin, L., Wei, P., Pan, W.: An adaptive two-sample test for high-dimensional means. Biometrika **103**(3), 609–624 (2016)
22. Zou, H., Li, R.: One-step sparse estimates in nonconcave penalized likelihood models. Ann. Stat. **36**(4), 1509 (2008)