Dual Control of Testing Errors in High-Dimensional Data Analysis *

X. Jessie Jeng and Yifei Hu
Department of Statistics, North Carolina State University
June 30, 2020

Abstract

False negative errors are of major concern in applications where missing a high proportion of true signals may cause serious consequences. False negative control, however, raises a bottleneck challenge in high-dimensional inference when signals are not identifiable at individual level. We propose a Dual Control of Errors (DCOE) method that regulates not only false positive but also false negative errors in measures that are highly relevant to high-dimensional data analysis. DCOE is developed under general covariance dependence with a new calibration procedure to measure the dependence effect. We specify how dependence co-acts with signal sparsity to determine the difficulty level of the dual control task and prove that DCOE is effective in retaining true signals that are not identifiable at individual level. Simulation studies are conducted to compare the new method with existing methods that focus on only one type of error. DCOE is shown to be more powerful than FDR methods and less aggressive than existing false negative control methods. DCOE is applied to a fMRI dataset to identify voxels that are functionally relevant to saccadic eye movements. The new method exhibits a nice balance in retaining signal voxels and avoiding excessive noise voxels.

*Address for correspondence: X. Jessie Jeng, Department of Statistics, North Carolina State University, SAS Hall, 2311 Stinson Dr., Raleigh, NC 27695-8203, USA. E-mail: xjjeng@ncsu.edu.
Keywords: False negative control; Sparse signal; Unnecessary false positives; Variable screening,

1 Introduction

In statistical hypothesis testing, a false positive/type I error occurs when a true null hypothesis is mistakenly rejected and a false negative/type II error occurs when a true alternative hypothesis is not accepted. It is well known that there exists a trade-off in controlling these two types of errors, and there is no guarantee to control both types of errors at predetermined levels at the same time. When multiple null hypotheses are tested simultaneously, the issue of multiplicity occurs, and important progress has been made to control family-wise error (FWE) or false discovery rate (FDR); see, e.g., Dudoit and Van Der Laan (2007), Chapters 1-7. Many recent developments focus on FDR control in various high-dimensional settings; see Fan et al. (2012), Barber and Candès (2015), Candès et al. (2018), Jeng and Chen (2019a), etc. Compared to the fast development in multiplicity adjustment, very limited progress has been made to address the inflating false negative errors in high-dimensional data analysis.

False negative errors are of major concern in applications where missing a large proportion of true signals may cause serious consequences. Examples can be found in medical, psychological, economic, and legal studies as reviewed in Petticrew et al. (2000) for their impact and implications. False negative control, however, raise a bottleneck challenge at the frontier of high-dimensional inference. In the framework of large-scale hypothesis testing, there are currently two major lines of research. One is global testing of mixture models, which addresses the problem of “detecting” the existence of signals without specifying their exact locations. The other is multiple testing, which considers the problem of “identifying” signals at individual levels. Figure 1 modified from Donoho and Jin (2015) illustrates the theoretical demarcation on the difficulty levels of these two problems in the setting where signal variables are relatively sparse compared to the number of noise variables. Given a sparsity level, signal intensity needs to be large enough (above the undetectable region) for
them to be detectable by global testing. For the signals to be identifiable by multiple testing, the signal intensity needs to be even larger (entering the identifiable region). It can be seen that signals in the middle white region of Figure 1 can only be detected for their existence by global testing but are not identifiable at the individual level by multiple testing. Such signals are highly relevant in high-dimensional data analysis and important in addressing the missing power and the issue of replicability in existing studies. New developments in theory and method are highly desirable to efficiently retain such signals.

Figure 1: Phase diagram for signal detection and identification. Signals in the middle white region are only detectable for their existence but not identifiable at individual level.

In this paper, we propose a novel analytic framework that features the so-called dual control property. A method possessing this property regulates false positive and false negative errors in criteria that are highly relevant to high-dimensional data analysis. More specifically, the method controls the false negative proportion (FNP = number of false negatives/number of alternative cases) at a user-specified level and, at the same time, controls the unnecessary false positives (UNFP), which are the variables ranked after all the signal variables in some importance measure, such as t statistic or p-value. Apparently, selecting more variables after the last signal variable does not help in reducing false negatives and should be avoided. Figure 2 demonstrates an example where 20 signal variables and a number of noise variables are ranked in some importance measure. Signal variables positioned relatively higher than some but not all noise variables. If we want to control FNP at the level of 0.1 and, at the same time, avoid selecting UNFP, an ideal cut-off point should be in between $t_3$ and $t_4$. A proce-
dure that can systematically provide such a cut-off point possesses the dual control property at the level of 0.1. Figure 2 also illustrates the ideal cut-off points for FWE control ($t_1$) and for false discovery proportion ($\text{FDP} = \text{number of false positives}/\text{number of rejections}$) control at the level of 0.1 ($t_2$). One can tell that a method with the dual control property is generally more powerful than the traditional FWE and FDP/FDR methods. Such method will be very useful in retaining the signals in the middle white region of Figure 1 in a principled way. Moreover, the adaptivity of the method to a pre-specified FNP level, such as 0.1 in Figure 2 allows the method to exclude a small percent of the weakest signals to reduce possibly a large number of false positives.

![Figure 2: Ideal cut-off points for different selection criteria (in bracket). FWE: family-wise error; FDP: false discovery proportion; FNP: false negative proportion; UNFP: unnecessary false positives.](image)

We propose to develop the Dual Control of Errors (DCOE) method in a setting with relatively sparse signals and arbitrary covariance dependence among variables. This setting is general enough to cover a wide range of high-dimensional applications. In theoretical analysis, we develop a new calibration procedure to quantify the dependence effect through a parameter whose scale is comparable to the scale of the parameter for signal sparsity. We find that dependence co-acts with signal sparsity to determine the difficulty level of the dual control task of DCOE. Moreover, when dependence is stronger than a certain level that depends on the signal sparsity, its effect on DCOE remains the same asymptotically. This interesting discovery is also supported in simulation examples with different types of dependence structures.

Additional simulation studies are conducted to compare the DCOE method with other
methods in multiple testing and false negative control. In both one-dimensional and two-dimensional examples, DCOE seems to be the only one possessing systematic dual control property. As a result, it has better power than the FDR procedure and less false positives than the other methods that mainly focus on false negative control.

We apply DCOE to analyze the fMRI data from the saccade experiment in Individual Brain Charting project [Pinho et al., 2018]. Compared to the other methods, results of the new method exhibit a nice balance in identifying brain regions that are functionally relevant to saccadic eye movement and avoiding the scattered or isolated voxels that are not functionally relevant to saccades.

The rest of the paper is organized as follows. Section 2 introduces the DCOE method and justify its dual control property in theory. Section 3 provides several illustrative examples under different dependence structures. DCOE with estimated number of signals is presented in Section 4 and additional simulation with one-dimensional and two-dimensional examples are conducted in Section 5. Section 6 applies DCOE to fMRI data analysis. Section 7 provides conclusion and further discussion. All the technical proofs are presented in the Appendix.

2 Method and Theory

We consider a model with continuous null distribution \(F_0\) and alternative distribution \(F_1\). We do not assume any specific forms for \(F_0\) and \(F_1\). Suppose that the test statistics

\[
x_j \sim F_0 \cdot 1\{j \in I_0\} + F_1 \cdot 1\{j \in I_1\}, \quad j = 1, ..., p,
\]

where \(I_0\) is the set of indices for noise variables and \(I_1\) is the set of indices for signal variables. All \(I_0\), \(I_1\), and \(F_1\) are unknown. One can perform inverse normal transformation as \(Z_j = \Phi(F_0^{-1}(X_j))\), where \(\Phi\) is the cumulative distribution function of standard normal distribution. Then, we have

\[
Z_j \sim \Phi \cdot 1\{j \in I_0\} + G \cdot 1\{j \in I_1\}, \quad j = 1, ..., p,
\]  

(1)
where \( G \) is some unknown distribution, representing signal distribution after the transformation. For presentation simplicity, we assume that \( G(t) < \Phi(t) \) for all \( t \in \mathbb{R} \). i.e., signal variables tend to show larger values than noise variables. This assumption can be easily generalized to signals with two-sided effects.

Consider a selection rule with threshold \( t \). Define the numbers of rejected cases, false positives, and false negatives as

\[
R(t) = \sum_{j=1}^{p} 1\{z_j > t\}, \quad \text{FP}(t) = \sum_{j \in I_0} 1\{z_j > t\}, \quad \text{FN}(t) = \sum_{j \in I_1} 1\{z_j \leq t\}.
\]

Note that \( R(t) \) can be directly observed from the data. \( \text{FN}(t) \) and \( \text{FP}(t) \) are unknown because \( I_0 \) and \( I_1 \) are unknown. A generalization to two-sided signal effects can be accommodated by replacing \( z_j \) with \(|z_j|\) and only allowing \( t > 0 \). Next, define FNP with respect to \( t \) as

\[
\text{FNP}(t) = \text{FN}(t)/s,
\]

where \( s \) denotes the number of signal variables, i.e. \( s = |I_1| \). \( \text{FNP}(t) \) may be regarded as the empirical type II error that is non-decreasing with respect to \( t \). For \( t \) sufficiently small, \( \text{FN}(t) \) and \( \text{FNP}(t) \) reach 0, and the corresponding noise variables not rejected are the UNFP as illustrated in Figure 2. Decreasing the threshold further and including more of such UNFP cannot trade for less false negatives as \( \text{FN}(t) \) already reaches 0. Define this critical value of threshold \( t \) as

\[
t^* = \sup \{ t : \text{FN}(t) = 0 \}
\]

and the corresponding number of false positives as

\[
\text{FP}^* = \text{FP}(t^*) = \sum_{j \in I_0} 1\{z_j > t^*\}.
\]

\( \text{FP}^* \) is the smallest/necessary number of false positives associated with total signal inclusion.

Both \( t^* \) and \( \text{FP}(t^*) \) are random variables that vary from sample to sample.

Our proposal for dual control of errors aims to determine a proper threshold \( \hat{t} \) that meets
two goals: (1) given a user-specified control level $\beta$ of FNP, the random quantity $\text{FNP}(\hat{t}) \leq \beta$ with high probability, and (2) the random quantity $\text{FP}(\hat{t}) \leq \text{FP}^*$ with high probability. Different from existing methods that focus on multiplicity adjustment through FWE or FDR control, we propose to control UNFP, a new criterion that is less stringent than FWE and FDR, but highly relevant to applications seeking more powerful signal discovery and tolerable to some more false positives if necessary. Moreover, our method has the flexibility to adapt to the user-specified $\beta$ level of FNP. Setting the $\beta$ level at a constant level, e.g. 0.1, allows the method to exclude $(100\beta)\%$ of the weakest signals to reduce possibly a large number of false positives intertwined with the weakest signals.

### 2.1 Estimation of False Negative Proportion

The new method is based on consistent estimation of FNP. Recall the definition of $\text{FNP}(t)$ in (2). By the fact that $s = \text{FN}(t) + \text{TP}(t)$ and $R(t) = \text{FP}(t) + \text{TP}(t)$, we have

$$\text{FNP}(t) = \frac{\text{FN}(t)}{s} = 1 - \frac{(R(t) - \text{FP}(t))}{s}, \tag{5}$$

where $R(t)$ is directly observable from the data. Because the noise distribution of $Z_j$ is $N(0,1)$ and there are $p - s$ noise variables, $\text{FP}(t)$ can be approximated by its mean value $E(\text{FP}(t)) = (p - s)\bar{\Phi}(t)$, where $\bar{\Phi}(t) = 1 - \Phi(t)$. For illustration purpose, we first assume that the true value of $s$ is known and construct an estimator for $\text{FNP}(t)$ as

$$\hat{\text{FNP}}(t) = \max\{1 - R(t)/s + (p - s)\bar{\Phi}(t)/s, \ 0\}. \tag{6}$$

$\hat{\text{FNP}}(t)$ with an estimated $s$ will be discussed later in Section 4. Note that if two-sided signal effects are under consideration, we can simply modify $\hat{\text{FNP}}(t)$ by replacing $(p - s)$ with $2(p - s)$.

Next, we study the estimation consistency of $\hat{\text{FNP}}(t)$. It is a challenging problem because the denominator $s$ is often much smaller than $p$, which can explode the scale of the ratio and hence the approximation error. This causes the fundamental difference between the analyses of FNP and FDP estimations. Specifically, we adopt the well-known calibration in sparse
inference for $s$ as
\[ s = s_p = p^{1-\gamma}, \quad \gamma \in [0, 1]. \tag{7} \]
The parameter $\gamma$ decreases with $s$, and when $\gamma > 0$, $s$ is of a smaller order than $p$.

We are particularly interested in FNP estimation under general covariance dependence, which has not been studied in literature. Given the correlation matrix $\Sigma$ of the transformed test statistics $Z_1, \ldots, Z_p$ in [1], let
\[ \bar{\rho} = \|\Sigma\|_1/p^2, \]
where $\|\Sigma\|_1 = \sum_{ij} |\sigma_{ij}|$; i.e., $\bar{\rho}$ represents the average absolute correlation of the test statistics. In many high-dimensional applications with large $p$, $\bar{\rho}$ is very close to zero because not every variable is correlated to all the other variables. For example, $\bar{\rho}$ of the $\Sigma_{p\times p}$ from an autoregressive model has the order of $p^{-1}$. In order to better calibrate the dependence effect in a wide range, we perform the re-parameterization
\[ \bar{\rho} = \bar{\rho}_p = p^{-\eta}, \quad \eta \in [0, 1]. \tag{8} \]
The parameter $\eta$ is in a constant scale and decreases with $\bar{\rho}$. $\eta = 0$ corresponds to the extremely dependent case where every variable is correlated to all the other variables, and, at the other end, $\eta = 1$ corresponds to the independent case.

With $\gamma$ and $\eta$ representing signal sparsity and covariance dependence respectively, we discover the lower bound of $t$ for the estimation consistency of $\hat{\text{FNP}}(t)$ as follows:
\[ \mu_{\text{min}} = \min\{\mu_1, \mu_2\}, \tag{9} \]
where
\[ \mu_1 = \sqrt{2\gamma \log p} \quad \text{and} \quad \mu_2 = \sqrt{(4\gamma - 2\eta) \log p + 4 \log \log p}. \]
The lower bound $\mu_{\text{min}}$ takes the value of either $\mu_1$ or $\mu_2$, depending on whichever is lower. It can be seen that $\mu_2 < \mu_1$ when $\eta$ is large enough or, in other words, when covariance dependence is weak enough. As dependence gets stronger and $\eta$ gets smaller, $\mu_1 < \mu_2$ occurs and the lower bound equals to $\mu_1$ and stops to change with $\eta$. The term $\log \log p$ is
a technical term for asymptotic analysis. This new estimation result is summarized in the following Lemma.

**Lemma 2.1.** Consider model (1). When \( t > \mu_{\text{min}} \), where \( \mu_{\text{min}} \) is defined in (9), \( \hat{FNP}(t) \) defined in (6) consistently estimates \( FNP(t) \), i.e.

\[
|\hat{FNP}(t) - FNP(t)| = o_p(1).
\]  

(10)

One can see that the consistency of \( \hat{FNP}(t) \) is achieved with \( t \) increasing with \( p \) in a \( \sqrt{\log p} \) scale. This is substantially different from the analysis in FDR studies, where the consistency of FDR/FDP estimation is studied with \( t \) as a constant. The \( \sqrt{\log p} \) scale and the sparsity parameter \( \gamma \) have been used in theoretical analysis for high-dimensional signal detection where the goal is to detect the existence of sparse signals without specifying their locations (see, e.g., [Donoho and Jin (2004)] and [Arias-Castro et al. (2011)]). Here, we adopt the sparsity parameter \( \gamma \) to analyze FNP estimation, which is a more difficulty task, and extend the framework to calibrate the effect of general dependence through a novel and compatible measure of \( \eta \). It is interesting to see how exactly \( \eta \) and \( \gamma \) play together to determine the estimation consistency of \( \hat{FNP}(t) \) for varying \( t \).

### 2.2 Controlling FNP at a User-Specified Level

In real studies, researchers may have different tolerance levels for false negative errors. Our proposed DCOE method allows a user-specified control level on FNP and efficiently selects a subset of candidates to achieve the control level. Specifically, given a user-specified constant \( \beta (>0) \), the method determines the selection threshold as

\[
\hat{t}(\beta) = \sup \{ t : \hat{FNP}(t) < \beta \},
\]  

(11)

and select all the candidates with \( z_j > \hat{t}(\beta) \). If two-sided signal effects are considered, all candidates with \( |z_j| > \hat{t}(\beta) \) will be selected. The following theorem presents the adaptivity of the DCOE method to the \( \beta \) level and its efficiency in selecting the smallest subset of
Theorem 2.1. Consider model (1) and a user-specified control level $\beta$ of FNP. Assume $G = G_p$ such that $G_p(\mu_{\text{min}}) = o(1)$, where $\mu_{\text{min}}$ is defined in (9), then the DCOE method with threshold $\hat{t}(\beta)$ efficiently controls the true FNP at the level of $\beta$, i.e.,

$$P(\text{FNP}(\hat{t}(\beta)) \leq \beta) \to 1 \quad (12)$$

and, for any threshold $\tilde{t} > \hat{t}(\beta)$,

$$P\{\text{FNP}(\tilde{t}) > \beta - \delta\} \to 1 \quad (13)$$

for arbitrarily small constant $\delta > 0$.

The $G_p(\mu_{\text{min}})$ condition in Theorem 2.1 implies that the signal variables are generally greater than $\mu_{\text{min}}$. Induced by the estimation consistency result in Section 2.1, this condition specifies the effect of covariance dependence on the FNP control of DCOE. It shows that the difficulty of FNP control increases as dependence gets stronger, but only to a certain level that depends on the signal sparsity. Figure 3 illustrates the $G_p(\mu_{\text{min}})$ condition in the phase diagram of signal detection and identification. Note that Figure 3 covers a wider range of sparsity than Figure 1. The $G_p(\mu_{\text{min}})$ condition is represented as the solid red line that moves with the dependence level $\eta$. Signals in the area above a solid red line can be retained by DCOE at a pre-specified level. Recall that signals in the region between the two dashed lines are only detectable for their existence but not identifiable at individual level. Results here demonstrate the capability of DCOE in retaining the unidentifiable signals under dependence.

Generally speaking, if we want to derive information of sparse signals from the data, a condition on the signal intensity is unfortunately unavoidable. When signal intensity is too low, no methods can even detect the existence of any sparse signals \cite{Donoho2004, Arias-Castro2011, Cai2011}, let along the more challenging tasks of FNP estimation and control. When the $G_p(\mu_{\text{min}})$ condition does not hold, i.e. when signals
are not all strong enough, threshold of DCOE tends to be higher than the ideal threshold that achieves the target level of $\beta$, which causes conservative results with inflated FNP and correspondingly less false positives. This tendency of DCOE is observed in simulation examples with low signal-to-noise ratio in Section 5.1.

### 2.3 Controlling Unnecessary False Positives

The UNFP control property of DCOE is very different from its FNP control property. Specifically, no condition on signal intensity is needed. Recall the threshold $\hat{t}(\beta)$ of DCOE in (11). Denote the true number of false positives associated with $\hat{t}(\beta)$ as

$$FP(\hat{t}(\beta)) = \sum_{j \in k_0} 1\{z_j > \hat{t}(\beta)\}. \quad (14)$$

The following result shows that $FP(\hat{t}(\beta))$ is no greater than $FP^*$ defined in (4), which is the number of necessary false positives associated with total signal inclusion. In other words, DCOE avoids UNFP with high probability.
Theorem 2.2. Consider model \([2]\) with \(s \to \infty\) as \(p \to \infty\). The DCOE method with threshold \(\hat{t}(\beta)\) asymptotically avoids the UNFP, i.e.,

\[
P(FP(\hat{t}(\beta)) \leq FP^*) \to 1,
\]

where \(FP(\hat{t}(\beta))\) and \(FP^*\) are defined in (14) and (4), respectively.

The conditions in Theorem 2.2 are very general. There is no need to assume signal intensity to be large enough, although details in the proof show that the convergence rates are different for large or small signal intensity. This result implies that the power gain of DCOE is achieved in a regulated way that does not incur excessive false positives, showing the potential of DCOE as a valuable complement to the existing signal detection framework.

3 Examples

In this section, we present three simulation examples with different dependence structures to illustrate the performance of DCOE. We first evaluate the estimator \(\hat{FNP}(t)\) from (6), then demonstrate the dual control property of DCOE. These examples help understand the foundation of the new method. Additional simulation studies are provided in Section 5.

We generate a series of test statistics

\[
(Z_1, \ldots, Z_p)^T \sim N((\mu_1, \ldots, \mu_p)^T, \Sigma),
\]

where \(p = 2000\), \(\mu_j = 3 \cdot 1\{j \in I_1\}\), and \(I_1\) is a set of indices randomly sampled from \(\{1, \ldots, p\}\) with cardinality \(s = |I_1| = p^{1-\gamma}\). We set \(\gamma = 0.3\), which corresponds to \(s = 205\), i.e. there are 205 signal variables with elevated mean values randomly located among 1795 noise variables.

We consider three different dependence structures:

- Model 1 [Autoregressive]: \(\Sigma = (\sigma_{ij}^{(1)})\), where \(\sigma_{ij}^{(1)} = \lambda^{|i-j|}\) for \(1 \leq i, j \leq p\).

- Model 2 [Block dependence]: \(\Sigma = I_{p/k} \otimes D\), where \(D\) is a \(k \times k\) matrix with diagonal entries 1 and off-diagonal entries \(r\).
• Model 3 [Factor model]: $\Sigma = (\sigma_{ij}^{(3)})$, where $\sigma_{ij}^{(3)} = V_{ij}/\sqrt{V_{ii}V_{jj}}$ for $1 \leq i, j \leq p$, $V = \tau hh^T + I_p$ with $\tau \in (0, 1)$ and $h \sim N(0, I_p)$.

In this section, we set $\lambda = 0.2, k = 40, r = 0.5, and \tau = 0.5$, so that the dependence parameter $\eta$ decreases from 0.95 in Model 1 to 0.57 in Model 2 to 0.23 in Model 3. Correspondingly, dependence among test statistics increases from very weak in Model 1 to moderately strong in Model 2 to very strong in Model 3.

It has been derived in Section 2.1 that $\mu_{\text{min}}$, as the boundary value for estimation consistency, is the minimum of $\mu_1$ and $\mu_2$, where $\mu_1$ depends on signal sparsity through $\gamma$ and $\mu_2$ depends on both signal sparsity and dependence through $\gamma$ and $\eta$ respectively. In these examples, $(\mu_1 = 2.14, \mu_2 = 1.69)$ for Model 1, $(\mu_1 = 2.14, \mu_2 = 2.92)$ for Model 2, and $(\mu_1 = 2.14, \mu_2 = 3.69)$ for Model 3. Consequently, $\mu_{\text{min}} = 1.69$ for Model 1 but remains the same at 2.14 for Model 2 and 3. These values of $\mu_{\text{min}}$ are illustrated as the solid vertical lines in Figure 4. The dotted vertical lines represent $\mu_1$ or $\mu_2$ whichever is larger. The dotted curves represent the absolute difference between the estimates $\hat{\text{FNP}}(t)$ and the true $\text{FNP}(t)$ from 100 replications. It can be seen that the estimation accuracy of $\hat{\text{FNP}}(t)$ increases with $t$, and the majority of the replicated differences are close to 0 after passing $\mu_{\text{min}}$, which support $\mu_{\text{min}}$ as the boundary value for estimation consistency in these examples with very different dependence structures.

Next, we demonstrate the dual control property of DCOE in finite sample. Table 1 presents the mean values and standard deviations of the realized FNP of DCOE with different $\beta$ levels from 100 replications. The frequencies of the realized event $\{\text{FP} < \text{FP}^*\}$ are also presented, where FP is the realized false positives of DCOE. The results of DCOE are compared with those of the classical BH-FDR procedure (Benjamini and Hochberg, 1995). For illustrative purposes, we also present the mean values of the realized false discovery proportion (FDP) of the two methods.

It can be seen that for Model 1 [Autoregressive], the mean values of FNP for DCOE with different nominal levels are 0.198 and 0.101, which are fairly close to their corresponding nominal levels of $\beta$. On the other hand, the frequencies of the event $\{\text{FP} < \text{FP}^*\}$ for DCOE are 100% for Model 1. In the more challenging cases generated by Model 2 [Block dependence]
Figure 4: Estimation differences from 100 replications. Plots from left to right are generated under Model 1 - 3. The solid vertical lines represent the boundary value $\mu_{\min} = \min\{\mu_1, \mu_2\}$. Model 1 has very weak dependence with $\eta = 0.95$ and $\mu_{\min} = \mu_2 = 1.69$. Model 2 has moderately strong dependence with $\eta = 0.57$ and $\mu_{\min} = \mu_1 = 2.14$. Model 3 has very strong dependence with $\eta = 0.23$ and $\mu_{\min} = \mu_1 = 2.14$.

Table 1: Mean values and standard deviations (in brackets) of the realized FNP and FDP of the proposed DCOE method with different nominal levels and the existing BH-FDR procedure. The frequencies of the event $\{FP < FP^*\}$ for different methods are also presented.

| Model Type          | Method          | FNP    | $\{FP < FP^*\}$ | FDP    |
|---------------------|-----------------|--------|------------------|--------|
| Autoregressive      | DCOE($\beta = 0.2$) | 0.198 (0.023) | 1.00             | 0.149 (0.035) |
|                     | DCOE($\beta = 0.1$) | 0.101 (0.037) | 1.00             | 0.307 (0.084) |
|                     | BH-FDR($\alpha = 0.05$) | 0.378 (0.040) | 1.00             | 0.044 (0.018) |
| Block dependence    | DCOE($\beta = 0.2$) | 0.169 (0.093) | 0.83             | 0.287 (0.279) |
|                     | DCOE($\beta = 0.1$) | 0.092 (0.077) | 0.73             | 0.449 (0.284) |
|                     | BH-FDR($\alpha = 0.05$) | 0.383 (0.068) | 1.00             | 0.047 (0.039) |
| Factor model        | DCOE($\beta = 0.2$) | 0.160 (0.099) | 0.89             | 0.262 (0.215) |
|                     | DCOE($\beta = 0.1$) | 0.084 (0.104) | 0.66             | 0.533 (0.227) |
|                     | BH-FDR($\alpha = 0.05$) | 0.400 (0.043) | 1.00             | 0.041 (0.049) |

and 3 [Factor model], the boundary value $\mu_{\min}$ required for estimation consistency increases from 1.69 to 2.14; and the differences between realized FNPs of DCOE and the nominal levels increase slightly. The frequencies of $\{FP < FP^*\}$ in Model 2 and 3 decrease a little due to the change of convergence rate when $\mu_{\min}$ is in different ranges as shown in the proof of Theorem 2.2. These observations generally agree with the theoretical results in Sections 2.2 and 2.3.

For illustrative purposes, we also present the results of the classical BH-FDR procedure. The nominal level of BH-FDR is set at $\alpha = 0.05$, which is a conventional choice, and its mean
values of FDP are fairly close to the nominal level and much lower than those of DCOE. However, the mean values of FNP for BH-FDR are 0.378, 0.383, and 0.4 for Model 1-3, which are much higher than those of DCOE. These results demonstrate the substantial differences between multiple testing procedures and the proposed DCOE method as they regulate errors differently for different purposes.

4 DCOE with estimated number of signals

In real applications, the number of signals, $s$, is often unknown. Existing studies for the estimation of $s$ often assume independence among variables (Genovese and Wasserman, 2004; Meinshausen and Rice, 2006; Cai et al., 2007; Cai and Jin, 2010), and most of them are for relatively dense signals. We are interested in finding an estimator of $s$ in less ideal settings with complex dependence structures. We find that if an estimator can provide a conservative result under general dependence, i.e., an estimated $\hat{s} < s$ with high probability, the UNFP control property of DCOE continues to hold. Moreover, if the estimator is consistent, i.e. $(1 - \epsilon)s < \hat{s} < s$ for any $\epsilon > 0$ with high probability, then the FNP control property of DCOE continues to hold.

Proposition 4.1. Replace $s$ in (11) with an estimator $\hat{s}$ and denote the selection threshold as $\hat{t}_s(\beta)$. If $\hat{s}$ is conservative, i.e. $P(\hat{s} < s) \to 1$, then under the condition in Theorem 2.2, we have

$$P(FP(\hat{t}_s(\beta)) \leq FP(t^*)) \to 1. \quad (17)$$

Further, if $\hat{s}$ is consistent, i.e. $P((1 - \epsilon) < \hat{s}/s < 1) \to 1$ for any constant $\epsilon > 0$, then under the condition in Theorem 2.1, we have

$$P(FNP(\hat{t}_s(\beta)) \leq \beta) \to 1 \quad (18)$$

and, for any threshold $\tilde{t} > \hat{t}_s(\beta)$,

$$P\{FNP(\tilde{t}) > \beta - \delta\} \to 1 \quad (19)$$
for arbitrarily small constant $\delta > 0$.

The estimation of $s$ under dependence is an on-going study. We adopt the estimator proposed in Meinshausen and Rice (2006) in our numerical analysis because it has been shown that this estimator provides a conservative result under general dependence (Meinshausen and Bühlmann, 2005). More recent study has investigated the consistency of this estimator under block dependence (Jeng et al., 2019). Further study on the estimation of $s$ under general dependence is beyond the scope of this paper but certainly of great interest.

5 Simulation

In this section, we implement the estimator $\hat{s}$ from Meinshausen and Rice (2006) to the DCOE procedure and compare the empirical performances of DCOE with existing methods that focus on false negative control. Such methods are relatively limited compared to multiple testing procedures and have only appeared recently. For example, the AFNC method in Jeng et al. (2016) was proposed to study rare genetic variants association through FNP control (Jeng et al., 2016); the MDR method in Cai and Sun (2017) was proposed to control the mean value of FNP using an empirical Bayesian approach (Cai and Sun, 2017); the AdSMR method in Jeng et al. (2019) focuses on controlling signal missing rate, but lacks the flexibility of adapting to a user-specified control level; and the FNC-Reg approach in Jeng and Chen (2019b) considers adaptive variable screening in linear regression. Among these methods, AFCN, MDR, and FNC-Reg are more comparable to DCOE because they all require the input of a user-specified control level. However, AFNC and MDR were developed under the independent assumption, and FNC-Reg considered specific dependence conditions among predictors to ensure good estimation of the precision matrix. Moreover, AFNC, MDR, and FNC-Reg control for only one type of error.

5.1 One-dimensional problems

In this section, we generate the test statistics by (16) with a covariance matrix that has 20 diagonal blocks with block sizes randomly generated from 10 to 100. The non-zero off-
diagonal correlations are set at 0.5. The dependence parameter $\eta$ varies from sample to sample due to the random block size. In the first set of examples, signal sparsity is fixed with $\gamma = 0.3$, and signal intensity ($\mu$) increases from 3 to 6. In the second set of examples, signal intensity is fixed and signal sparsity varies as $\gamma = 0.3$ and 0.5.

The performances of the methods are evaluated by four measures. The first three measures, $\{FP < FP^*\}$, FNP, and FDP are the same as in Table 1. The last measure is the Fowlkes-Mallows index (Fowlkes and Mallows, 1983; Halkidi et al., 2001), which summarizes the measures of FNP and FDP by calculating the geometric mean of $(1-FNP)$ and $(1-FDP)$, i.e.,

$$FM-index = \sqrt{(1-FNP) \times (1-FDP)}.$$ 

Higher values of the FM-index indicate better classification results. The FM-index is an appropriate summary measure in high-dimensional settings with sparse alternative cases because the scale of its FDP component is more comparable to that of its FNP component compared to the classical false positive proportion (FPP = number of false positives/number of null cases).

Results of the first set of examples are summarized in Table 2. It can be seen that DCOE continues to control UNFP well as the frequencies of $\{FP < FP^*\}$ are fairly high. This agrees with the claim in (17) in Proposition 4.1 as we know that the adopted estimator $\hat{s}$ is conservative under general dependence (Meinshausen and Bühlmann, 2005). As for FNP control, because the estimation boundary $\mu_{min}$ remains at 2.14 in these examples, the condition $G_p(\mu_{min}) = o(1)$ for (18) is not well supported for smaller $\mu$. Also, because the estimated $\hat{s}$ is generally less than the true $s$ when signal intensity is low, DCOE with $\hat{s}$ implemented selects less variables than actually needed to reach the nominal level of $\beta$. These result in inflated realized FNP as shown in Table 2. Further, as $\mu$ increases, the mean value of FNP of DCOE gets closer to the nominal level of $\beta$, which agrees with the claims in (18) and (19) in Proposition 4.1. Among the three methods presented in Table 2, DCOE shows the best ability to adapt to the nominal level of $\beta$ as $\mu$ increases and selects less variables than the other two methods. In terms of the FM-index, DCOE outperforms the other two methods in these examples.
Table 2: Effect of signal intensity on different FN control methods with estimated s. Mean values and standard deviations (in brackets) of the realized FNP, FDP, and the FM-index are presented for the proposed DCOE method and two existing methods, AFNC and MDR. The frequencies of the realized event \{FP < FP^\} of different methods are also presented from 100 replications.

| \(\mu = 3\) | \{FP < FP^\} | FNP | FDP | FM-index |
|-------------|----------------|-----|-----|----------|
| DCOE(\(\beta = 0.1\)) | 0.97 | 0.32 (0.11) | 0.11 (0.17) | 0.76 (0.08) |
| AFNC(\(\beta = 0.1\)) | 0.92 | 0.23 (0.11) | 0.20 (0.24) | 0.76 (0.14) |
| MDR(\(\beta = 0.1\)) | 0.90 | 0.09 (0.08) | 0.53 (0.26) | 0.62 (0.15) |

| \(\mu = 4\) | \{FP < FP^\} | FNP | FDP | FM-index |
|-------------|----------------|-----|-----|----------|
| DCOE(\(\beta = 0.1\)) | 0.97 | 0.18 (0.07) | 0.06 (0.17) | 0.87 (0.10) |
| AFNC(\(\beta = 0.1\)) | 0.90 | 0.09 (0.05) | 0.15 (0.27) | 0.86 (0.18) |
| MDR(\(\beta = 0.1\)) | 0.55 | 0.05 (0.05) | 0.37 (0.32) | 0.74 (0.19) |

| \(\mu = 5\) | \{FP < FP^\} | FNP | FDP | FM-index |
|-------------|----------------|-----|-----|----------|
| DCOE(\(\beta = 0.1\)) | 0.94 | 0.12 (0.04) | 0.03 (0.14) | 0.92 (0.09) |
| AFNC(\(\beta = 0.1\)) | 0.84 | 0.03 (0.02) | 0.13 (0.28) | 0.89 (0.20) |
| MDR(\(\beta = 0.1\)) | 0.45 | 0.04 (0.04) | 0.35 (0.33) | 0.76 (0.19) |

| \(\mu = 6\) | \{FP < FP^\} | FNP | FDP | FM-index |
|-------------|----------------|-----|-----|----------|
| DCOE(\(\beta = 0.1\)) | 0.94 | 0.10 (0.03) | 0.02 (0.11) | 0.94 (0.07) |
| AFNC(\(\beta = 0.1\)) | 0.69 | 0.01 (0.01) | 0.12 (0.27) | 0.91 (0.20) |
| MDR(\(\beta = 0.1\)) | 0.42 | 0.03 (0.04) | 0.37 (0.33) | 0.75 (0.19) |

The second set of examples have \(\gamma\) increased from 0.3 to 0.5, so that the number of signals deceases from 205 to 45. The nominal level of \(\beta\) also varies. The signal intensity \(\mu\) is fixed at 5. Results summarized in Table 3 show that when signals get sparser with larger \(\gamma\), the performances of all three methods deteriorate by including more noise variables. However, DCOE continues to outperform the other two methods in controlling UNFP and adapting to the nominal level.

5.2 A Two-dimensional example

In this example, we simulate a 100 × 100 two-dimensional grid graph with the signal region demonstrated in Figure 5a. The signal region covers 994 nodes, i.e. \(s = 994\). At each node, test statistic \(Z_{ij}\) is generated independently from \(N(A_{ij}, 1)\), where \(A_{ij} \sim \text{Uniform}[1, 2.5]\) if \((i, j)\) is in the signal region, and \(A_{ij} = 0\) otherwise.

We apply the proposed DCOE method with different nominal levels and compare the results with those of BH-FDR, AFNC, and MDR. Figure 5b shows the selected nodes of BH-FDR with \(\alpha = 0.05\) in a single trial, Figure 5c and 5d show the results of AFNC and MDR with \(\beta = 0.1\), respectively, and Figure 5e and 5f show the results of DCOE with
Table 3: Effect of nominal level and signal sparsity on different FN control methods with estimated s. Mean values and standard deviations (in brackets) of the realized FNP, FDP, and the FM-index are presented for the proposed DCOE method and two existing methods, AFNC and MDR. The frequencies of the realized event \{FP < FP^*\} of different methods are also presented from 100 replications.

| \(\gamma = 0.3\) | \(\{FP < FP^*\}\) | FNP \(\pm\) s | FDP \(\pm\) s | FM-index \(\pm\) s |
|-------------------|-------------------|----------------|----------------|-------------------|
| DCOE(\(\beta = 0.1\)) | 0.94 | 0.12 (0.04) | 0.03 (0.14) | 0.92 (0.09) |
| AFNC(\(\beta = 0.1\)) | 0.84 | 0.03 (0.02) | 0.13 (0.28) | 0.89 (0.20) |
| MDR(\(\beta = 0.1\)) | 0.45 | 0.04 (0.04) | 0.35 (0.33) | 0.76 (0.19) |
| DCOE(\(\beta = 0.2\)) | 0.98 | 0.21 (0.05) | 0.01 (0.03) | 0.88 (0.02) |
| AFNC(\(\beta = 0.2\)) | 0.85 | 0.03 (0.02) | 0.10 (0.24) | 0.92 (0.17) |
| MDR(\(\beta = 0.2\)) | 0.51 | 0.08 (0.09) | 0.27 (0.28) | 0.79 (0.14) |

| \(\gamma = 0.5\) | \(\{FP < FP^*\}\) | FNP \(\pm\) s | FDP \(\pm\) s | FM-index \(\pm\) s |
|-------------------|-------------------|----------------|----------------|-------------------|
| DCOE(\(\beta = 0.1\)) | 0.83 | 0.13 (0.08) | 0.12 (0.27) | 0.84 (0.18) |
| AFNC(\(\beta = 0.1\)) | 0.71 | 0.06 (0.05) | 0.23 (0.36) | 0.80 (0.28) |
| MDR(\(\beta = 0.1\)) | 0.33 | 0.03 (0.04) | 0.57 (0.44) | 0.54 (0.34) |
| DCOE(\(\beta = 0.2\)) | 0.86 | 0.21 (0.11) | 0.10 (0.24) | 0.82 (0.15) |
| AFNC(\(\beta = 0.2\)) | 0.76 | 0.07 (0.06) | 0.20 (0.33) | 0.83 (0.25) |
| MDR(\(\beta = 0.2\)) | 0.38 | 0.06 (0.08) | 0.54 (0.45) | 0.55 (0.33) |

\(\beta = 0.1\) and 0.5, respectively. It can be seen that DCOE identifies more true signals and delineates the signal region much better than BH-FDR even when \(\beta\) is as large as 0.5. On the other hand, it pays the price with some more false positives. Compared to AFNC and MDR, DCOE has less false positives and less true positives, which does not seem to impair its ability to delineate the signal region much.

6 Application

We obtained the fMRI data from the Individual Brain Charting (IBC) Project, which is a publicly available high-resolution fMRI dataset for cognitive mapping [Pinho et al., 2018]. The dataset refers to a cohort of 12 participants performing different tasks, addressing both low- and high-level cognitive functions. We focus on the data from the saccade experiment for spatial cognition, in which ocular movements were performed according to the displacement of a fixation cross from the center toward peripheral locations in the image displayed.

The data were collected using a Gradient-Echo (GE) pulse, whole-brain Multi-Band (MB) accelerated Echo-Planar Imaging (EPI) T2*-weighted sequence with Blood-Oxygenation-
Figure 5: Comparison results from a single trial. The clustered signals are shown in plot (a). Plots (b) - (f) demonstrate the results of different methods.
Level-Dependent (BOLD) contrasts, and preprocessed using PyPreprocess, a collection of python tools for preprocessing fMRI data. In order to assess the statistical significance of the differences among evoked BOLD responses, test statistics are computed at every voxel for each contrast using General Linear Model (GLM). All images are confined to an average mask of the gray matter across subjects, which yields 371,817 voxels at the chosen resolution. More details about the dataset can be found in Pinho et al. (2018).

We apply the proposed DCOE method, the popular BH-FDR procedure, and the existing AFNC and MDR methods to the statistical maps of 12 participants. The numbers of selected voxels for each participant are reported in Table 4. Note that sub-03 and sub-10 are not included in the saccade dataset. Among all the methods, BH-FDR selects the least voxels and MDR selects the most voxels. DCOE selects more than BH-FDR and less than AFNC and MDR. These results are consistent with what have been observed in simulation examples.

Table 4: Numbers of selected voxels by different methods.

| Subject | BH-FDR($\alpha = 0.05$) | DCOE($\beta = 0.1$) | AFNC($\beta = 0.1$) | MDR($\beta = 0.1$) |
|---------|--------------------------|----------------------|----------------------|----------------------|
| sub-1   | 18500                    | 19379                | 24837                | 27793                |
| sub-2   | 44902                    | 52614                | 70981                | 78288                |
| sub-4   | 15291                    | 31852                | 44893                | 60356                |
| sub-5   | 8623                     | 23236                | 37333                | 62976                |
| sub-6   | 17778                    | 29298                | 40537                | 47812                |
| sub-7   | 25011                    | 51405                | 76931                | 88407                |
| sub-8   | 35463                    | 47467                | 65238                | 74915                |
| sub-9   | 20846                    | 35662                | 50516                | 63989                |
| sub-11  | 28586                    | 32989                | 43147                | 48892                |
| sub-12  | 20469                    | 27388                | 36256                | 38353                |
| sub-13  | 21365                    | 44094                | 63192                | 74508                |
| sub-14  | 21067                    | 58494                | 85889                | 96651                |

Figure 6 illustrates the selected voxels of each method in the image of a single participant (sub-5) from posterior, superior, and left views. The figure is generated using the Multi-image Analysis GUI (http://ric.uthscsa.edu/mango/). We look into several regions that are known to be associated with saccadic eye movements. First, the visual cortex (VC) on occipital lobe in the posterior region of the brain, as indicated in Figure 6d is the primary cortical region that receives, integrates, and processes visual information (Bodis-Wollner et al., 1997). We
can see that all four methods have identified voxels in VC. However, the results of DCOE, AFNC and MDR seem to match the VC region much better than that of BH-FDR. From the superior and left views, it shows that BH-FDR has only a few or no discoveries in the frontal eye fields (FEF) located in Brodmann area 8, the supplementary eye fields (SEF) located in Brodmann area 6, and the posterior parietal cortex (PPC). The locations of FEF, SEF, and PPC are indicated in Figure 6e and 6f. FEF and SEF are believed to play important roles in visual attention and eye movements as electrical stimulation of these areas evokes eye movements (Bruce and Goldberg, 1985; Bruce et al., 1985). PPC, on the other hand, is related to decision making and saccades (Goldberg et al., 2002; Schluppeck et al., 2005). It can be seen that DCOE, AFNC, and MDR have better power for identifying signal voxels in FEF, SEF, and PPC. Among these three methods, DCOE selects the least scattered or isolated voxels that are not functionally relevant to saccades.

7 Conclusion and Discussion

In this paper, we propose a new dual control strategy to regulate both false positive and false negative errors in high-dimensional applications. The proposed DCOE method is built upon consistent estimation of false negative proportion. Its efficiency has been studied from two aspects in the paper. One is under the $G_p(\mu_{\mathrm{min}})$ condition on signal intensity (Theorem 2.1), the other assumes no intensity condition (Theorem 2.2). These results show that DCOE can effectively retain signals that are not identifiable at individual level without paying excessive price of false positives.

The dual control property of DCOE is theoretically justified under general covariance dependence. By utilizing a new calibration procedure on the dependence, we are able to quantify the dependence effect and find out how it co-acts with signal sparsity to affect the dual control ability of DCOE. As DCOE also relies on the information of the number of signals $s$, we demonstrate the performance of DCOE with an existing estimator that is conservative under general dependence. It will be interesting to investigate the estimation of $s$ under general dependence using the new calibration procedure in future research.
Figure 6: The selected voxels from BH-FDR, DCOE, AFNC, and MDR for a single participant (sub-05).
We demonstrate the finite sample performance of DCOE in simulation and compare DCOE with other existing methods. DCOE seems to be the only one possessing the dual control property. Moreover, its efficient adaptivity to the $\beta$ level helps to exclude false positives intertwined with the weakest $(100\beta)\%$ signals and select more concentrated candidate sets compared to the other false negative control methods.

We apply DCOE to identify functionally relevant regions for saccadic eye movement using fMRI data. Its results are compared with those of other methods and with regions that are known to be associated with saccade. DCOE seems to benefit from its dual control property and exhibits a good balance in identifying signal voxels in the functionally relevant regions and avoiding the scattered noise voxels. This new method would be useful in, for example, pre-surgical planning with fMRI data, where efficient false negative control is of vital importance because neurosurgical patients are likely to experience significant harm from mistakenly deeming a region to be functionally uninvolved and subsequently resecting critical tissues. In a broader sense, our proposal for dual control of testing errors has the potential to become a valuable complement to the existing sparse inference framework and provide new insights in real world studies.

8 Appendix

This section provides the proofs of Lemma 2.1, Lemma 8.1, Theorem 2.1, Theorem 2.2 and Proposition 4.1 as well as additional simulation results. We will frequently use the following result on Mill’s ratio:

$$\overline{\Phi}(x) \leq x^{-1}\phi(x) \text{ for any } x > 0.$$  

The symbol $C$ denotes a genetic, finite constant whose value can be different at different occurrences.
8.1 Proof of Lemma 2.1

For notation simplicity, let

\[ A(t) = 1 - \frac{R(t)}{s} + \frac{(p - s)\Phi(t)}{s}. \]  

Then \( \widehat{\text{FNP}}(t) = \max\{A(t), 0\} \), and it is sufficient to show that

\[ |A(t) - \text{FNP}(t)| = o_p(1) \quad \text{when} \quad A(t) \geq 0 \]  

and

\[ \text{FNP}(t) = o_p(1) \quad \text{when} \quad A(t) < 0. \]  

Consider (21) first. By the definitions of \( A(t) \) and \( \text{FNP}(t) \),

\[ |A(t) - \text{FNP}(t)| = |s^{-1}(R(t) - (p - s)\Phi(t)) - s^{-1}(R(t) - \text{FP}(t))| = s^{-1}|\text{FP}(t) - (p - s)\Phi(t)|. \]

Therefore, it is sufficient to show

\[ s^{-1}|\text{FP}(t) - (p - s)\Phi(t)| = o_P(1). \]  

Recall the definition of \( \mu_1 \) and \( \mu_2 \) in (9). The following proof is composed of two parts. The first part assumes \( t > \mu_1 \) and the second part assumes \( \mu_2 < t \leq \mu_1 \).

Consider the first part. It’s sufficient to show \( s^{-1}\text{FP}(t) = o_P(1) \) and \( s^{-1}(p - s)\Phi(t) = o(1) \) with \( t > \mu_1 \). By Mill’s ratio and the re-parameterization of \( s \) in \( \gamma \),

\[ s^{-1}(p - s)\Phi(t) \leq \frac{Cpe^{-t^2/2}}{ts} \leq \frac{C}{\sqrt{\log p}} = o(1). \]

On the other hand, for a fixed constant \( a > 0 \),

\[ P(s^{-1}\text{FP}(t) > a) \leq \frac{E(\text{FP}(t))}{as} \leq p \max_{j \in I_a} P(z_j > t) \leq \frac{Cp\Phi(t)}{s} = o(1). \]
Therefore, the claim in (23) is justified for $t > \mu_1 = \sqrt{2\gamma \log p}$.

Next we present the second part of the proof with $\mu_2 < t \leq \mu_1$. Define

$$D_p = s^{-2} e^{-t^2/2} \|\Sigma\|_1 \log p.$$ 

By the condition $t > \mu_2$ and the re-parameterizations of $\|\Sigma\|_1$ in $\eta$ and $s$ in $\gamma$, it can be shown that

$$D_p = p^{2\gamma-\eta} e^{-t^2/2} \log p = \log^{-1} p = o(1).$$

Then, (23) is implied by

$$s^{-1}|\text{FP}(t) - (p-s)\Phi(t)| = o_P(\sqrt{D_p}).$$ \hspace{1cm} (24)

Apply Chebyshev’s inequality,

$$P(s^{-1}|\text{FP}(t) - (p-s)\Phi(t)| > \sqrt{D_p}) \leq \frac{\text{Var}(\text{FP}(t))}{s^2 D_p}. \hspace{1cm} (25)$$

We have the following lemma for the order of $\text{Var}(\text{FP}(t))$ under dependence with the proof provided in Section 8.2.

**Lemma 8.1.** Consider the null version of model (1) with $G = \Phi$. We have

$$\text{Var}(\sum_{j=1}^{p} 1_{(z_j > t)}) = O(e^{-t^2/2}\|\Sigma\|_1).$$

Therefore, $\text{Var}(\text{FP}(t)) = \text{Var}(\sum_{j \in I_0} 1_{(z_j > t)}) \leq C e^{-t^2/2}\|\Sigma\|_1$, and it follows that

$$\frac{\text{Var}(\text{FP}(t))}{s^2 D_p} = o(1) \hspace{1cm} (26)$$

by the definition of $D_p$. Combining (25) and (26) gives (24), which justifies the claim in (23) for $\mu_2 < t \leq \mu_1$.

Finally, consider (22). It can be shown that $A(t) < 0$ implies $1 - \text{FP}(t)/s - TP(t)/s +$
\[(p - s)\Phi(t)/s < 0,\] which, combined with \(\text{FNP}(t) = 1 - \text{TP}(t)/s,\) further implies

\[\text{FNP}(t) < s^{-1}(\text{FP}(t) - (p - s)\Phi(t)).\]

Since the order of the right hand side has been derived in (23), (22) follows.

### 8.2 Proof of Lemma 8.1

For \(i \neq j,\) let \(\rho_{ij}\) be the correlation between \(z_i\) and \(z_j\) and \(C_{ij} = \text{Cov}(1_{\{z_j > t\}}, 1_{\{z_j > t\}}).\) Then,

\[\text{Var}(\sum_{j=1}^{p} 1_{\{z_j > t\}}) \leq \sum_{j=1}^{p} \text{Var}(1_{\{z_j > t\}}) + \sum_{i \neq j} C_{ij}. \tag{27}\]

By Mill’s ratio

\[\sum_{j=1}^{p} \text{Var}(1_{\{z_j > t\}}) \leq p\Phi(t)(1 - \Phi(t)) \leq Cpe^{-t^2/2} \leq Ce^{-t^2/2}\|\Sigma\|_1. \tag{28}\]

Thus, it remains to show

\[\sum_{i \neq j} C_{ij} \leq C\|\Sigma\|_1 e^{-t^2/2}.\]

Fix a pair of \((i, j)\) such that \(i \neq j\) and \(|\rho_{ij}| \neq 1,\)

\[C_{ij} = \int_{-\infty}^{t} \int_{-\infty}^{t} f_{\rho_{ij}}(x, y) \, dx \, dy - \int_{-\infty}^{t} \phi(x) \, dx \int_{-\infty}^{t} \phi(y) \, dy\]

For a nonnegative integer \(k,\) let \(H_k(x) = (-1)^k \frac{1}{\phi(x)} \frac{d^k}{dx^k} \phi(x)\) be the \(k\)th Hermite polynomial; see Feller (1971) for such a definition. Then Mehler’s expansion gives

\[f_{\rho_{ij}}(x, y) = \left(1 + \sum_{k=1}^{\infty} \frac{\rho_{ij}^k}{k!} H_k(x) H_k(y)\right) \phi(x) \phi(y),\]
and it follows that

\[ C_{ij} = \sum_{k=1}^{\infty} \frac{\rho_{ij}^k}{k!} \int_{-\infty}^{t} H_k(x) \phi(x) dx \int_{-\infty}^{t} H_k(y) \phi(y) dy. \]

Since \( H_{k-1}(t) \phi(t) = \int_{-\infty}^{t} H_k(y) \phi(y) dy \) for \( t \in \mathbb{R} \), then

\[ C_{ij} = \sum_{k=1}^{\infty} \frac{\rho_{ij}^k}{k!} [H_{k-1}(t) \phi(t)]^2. \]

Further, Lemma 3.1 of Chen and Doerge (2016) asserts

\[ |e^{-t^2/2} H_k(t)| \leq C_0 \sqrt{k!} k^{-1/12} e^{-t^2/4} \] for any \( t \in \mathbb{R} \)

for some constant \( C_0 > 0 \). Then

\[ [H_{k-1}(t) \phi(t)]^2 \leq C_0^2 (k - 1)! (k - 1)^{-1/6} e^{-t^2/2}. \]

Therefore,

\[ \left| \sum_{i \neq j} C_{ij} \right| \leq C \sum_{1 \leq i \leq j \leq p} |\rho_{ij}| \sum_{k=1}^{\infty} k^{-7/6} |\rho_{ij}| k^{-1} e^{-t^2/2} \leq C e^{-t^2/2} \sum_{1 \leq i \leq j \leq p} |\rho_{ij}| \leq C e^{-t^2/2} \| \Sigma \|_1. \] (29)

Combing (27) with (28) and (29) gives

\[ \text{Var} \left( \sum_{j=1}^{p} 1_{\{z_j > t\}} \right) = O(e^{-t^2/2} \| \Sigma \|_1). \]
8.3 Proof of Theorem 2.1

First, it can be shown that for a fixed constant $a > 0$,

$$P(\text{FNP}(\mu_{\text{min}}) > a) = P\left(s^{-1}\sum_{j \in I_1} 1\{z_j \leq \mu_{\text{min}}\} > a\right) \leq \frac{1}{as} \sum_{j \in I_1} G_p(\mu_{\text{min}}) \leq \frac{1}{a} \max_{j \in I_1} G_p(\mu_{\text{min}}) = o(1),$$

where the last step is by the condition $G_p(\mu_{\text{min}}) = o(1)$. Therefore, $\text{FNP}(\mu_{\text{min}}) = o(1)$. By Lemma 2.1, we have $|\hat{T}(\mu_{\text{min}}) - \text{FNP}(\mu_{\text{min}})| = o_P(1)$, which implies $\text{FNP}(\mu_{\text{min}}) = o_P(1)$. Then, by the construction of $\hat{t}(\beta)$, we have $P(\hat{t}(\beta) \geq \mu_{\text{min}}) \rightarrow 1$ and, consequently, by Lemma 2.1 again, $|\text{FNP}(\hat{t}(\beta)) - \text{FNP}(\hat{t}(\beta))| = o_P(1)$. Now, since $\text{FNP}(\hat{t}(\beta)) < \beta$ almost surely, the claim in (12) follows.

Next, let’s consider the claim in (13). Because $\hat{t} > \hat{t}(\beta)$, the construction of $\hat{t}(\beta)$ implies that $\text{FNP}(\hat{t}) \geq \beta$. On the other hand, because $\hat{t} > \hat{t}(\beta) > \mu_{\text{min}}$ with probability tending to 1, Lemma 2.1 implies that $|\text{FNP}(\hat{t}) - \text{FNP}(\hat{t})| = o_P(1)$. Therefore, the claim in (13) follows.

8.4 Proof of Theorem 2.2

Recall the definition of $t^*$ in (3) as $t^* = \sup\{t : \text{FN}(t) = \text{FNP}(t) = 0\}$. Define the cut-off value

$$t^*(\beta/2) = \sup\{t : \text{FNP}(t) \leq \beta/2\}. \quad (30)$$

Recall the quantity $\mu_{\text{min}}$ defined in (9). We prove (15) in two cases: $t^*(\beta/2) \geq \mu_{\text{min}}$ and $t^*(\beta/2) < \mu_{\text{min}}$. Note that $t^*(\beta/2) \geq \mu_{\text{min}}$ corresponds to the situation when signals are relatively stronger compared to the situation with $t^*(\beta/2) < \mu_{\text{min}}$.

In the first case with $t^*(\beta/2) \geq \mu_{\text{min}}$. Lemma 2.1 implies that $|\text{FNP}(t^*(\beta/2)) - \text{FNP}(t^*(\beta/2))| = o_p(1)$. Since $\text{FNP}(t^*(\beta/2)) = \beta/2$ almost surely, then $|\text{FNP}(t^*(\beta/2)) - \beta/2| = o_P(1)$. Recall the construction of $\hat{t}(\beta)$ and the fact that test statistics are decreasingly ordered, it can be shown that $P(\hat{t}(\beta) \geq t^*(\beta/2)) \rightarrow 1$. Since $t^*(\beta/2) > t^*$ almost surely, we have $P(\hat{t}(\beta) \geq t^*) \rightarrow 1$, and the claim in (15) follows.

Now, consider the second case with $t^*(\beta/2) < \mu_{\text{min}}$. Given that $\text{FP}(t)$ is a non-increasing
function of $t$, it is sufficient to show

$$P(\hat{t}(\beta) < t^*) \rightarrow 0.$$  \hspace*{1cm} (31)

By the definition of $\hat{t}(\beta)$, the following events are equivalent:

$$\{\hat{t}(\beta) < t^*\} \iff \{\forall z(j) > t^*, \; \widehat{\text{FNP}}(z(j)) > \beta\}.$$  

By the definition of $\widehat{\text{FNP}}(t)$, we have

$$\{\text{FNP}(z(j)) > \beta\} \iff \{s - R(z(j)) + (p - s)\Phi(z(j)) > \beta s\}.$$  

Since $R(t) = \text{TP}(t) + \text{FP}(t)$, then

$$\{s - R(z(j)) + (p - s)\Phi(z(j)) > \beta s\} \iff \{\text{FP}(z(j)) < (p - s)\Phi(z(j)) + (1 - \beta)s - \text{TP}(z(j))\}.$$  

Therefore, (31) is implied by

$$P(\text{FP}(z(j)) < (p - s)\Phi(z(j)) + (1 - \beta)s - \text{TP}(z(j)), \; \forall z(j) > t^*) \rightarrow 0.$$  \hspace*{1cm} (32)

Now, the definition of $t^*(\beta/2)$ in (30) implies that $\text{FNP}(t^*(\beta/2)) = \beta/2$ and $t^*(\beta/2) > t^*$ almost surely. Then for $z(j) \in (t^*, t^*(\beta/2)]$, the corresponding

$$\text{FNP}(z(j)) \in (0, \beta/2] \quad \text{and} \quad \text{TP}(z(j)) \in [(1 - \beta/2)s, s).$$
Therefore, the left hand side of (32)

\[ P(\text{FP}(z_{(j)}) < (p - s)\Phi(z_{(j)}) + (1 - \beta)s - TP(z_{(j)}), \quad \forall z_{(j)} > t^*) \]
\[ \leq P(\text{FP}(z_{(j)}) < (p - s)\Phi(z_{(j)}) + (1 - \beta)s - TP(z_{(j)}), \quad \forall z_{(j)} \in (t^*, t^*(\beta/2))] \]
\[ \leq P(\text{FP}(z_{(j)}) < (p - s)\Phi(z_{(j)}) + (1 - \beta)s - (1 - \beta/2)s, \quad \forall z_{(j)} \in (t^*, t^*(\beta/2))] \]
\[ = P(\text{FP}(z_{(j)}) < E(\text{FP}(z_{(j)})), \quad \forall z_{(j)} \in (t^*, t^*(\beta/2))] \]

Because there are at least \( \lceil \beta s/2 \rceil \) signal variables between \( t^* \) and \( t^*(\beta/2) \), then there are at least \( \lceil \beta s/2 \rceil \) test statistics \( z_{(j)} \) between \( t^* \) and \( t^*(\beta/2) \). Given that \( s \to \infty \), the above goes to 0 as \( p \to \infty \). This concludes the proof.

8.5 Proof of Proposition 4.1

It can be seen from (5) that \( \hat{\text{FNP}}(t) \) is a non-decreasing function of \( s \). Then given \( P(\hat{s} < s) \to 1 \), we have \( P(\text{FNP}_{\hat{s}}(t) \leq \text{FNP}_s(t)) \to 1 \), and consequently \( P(\hat{t}_{\hat{s}}(\beta) \geq \hat{t}_s(\beta)) \to 1 \). Combining this with the result in (31) gives

\[ P(\hat{t}_{\hat{s}}(\beta) < t^*) \to 0, \]

and the claim in (17) is proved.

Next consider the claims in (18) and (19). Similar arguments as in the proof of Theorem 2.1 can be applied, and we only need to show that Lemma 2.1 continues to hold with \( \hat{\text{FNP}}(t) \) replaced by \( \text{FNP}_{\hat{s}}(t) \). Given the result in (10), it is sufficient to show \( |\text{FNP}_{\hat{s}}(t) - \text{FNP}(t)| = o_P(1) \) for \( t \geq \mu_{min} \), which is implied by

\[ |A_{\hat{s}}(t) - A(t)| = o_P(1) \quad \text{for} \quad t \geq \mu_{min}, \quad (33) \]
given the definition of $A(t)$ in (20). By direct calculation,

$$
|A(\hat{s}(t)) - A(t)| = |(\hat{s}^{-1} - s^{-1})(R(t) - p\Phi(t))| \\
\leq |\hat{s}^{-1} - s^{-1}| \cdot TP(t) + |\hat{s}^{-1} - s^{-1}| \cdot |FP(t) - p\Phi(t)|.
$$

(34)

Given $P((1 - \delta) < \hat{s}/s < 1) \to 1$ for any $\delta > 0$, it can be shown that

$$
P(|\hat{s}^{-1} - s^{-1}| < \frac{\delta}{1 - \delta} s^{-1}) \to 1.
$$

On the other hand, $TP(t) \leq s$ almost surely. Then it follows that the first term in (34),

$$
|\hat{s}^{-1} - s^{-1}| \cdot TP(t) = o_p(1).
$$

For the second term in (34),

$$
|\hat{s}^{-1} - s^{-1}| \cdot |FP(t) - p\Phi(t)| < \frac{\delta}{1 - \delta} s^{-1} |FP(t) - p\Phi(t)| \leq \frac{\delta}{1 - \delta} (s^{-1} |FP(t) - (p - s)\Phi(t)| + \Phi(t))
$$

with probability tending 1, where $\Phi(t) = o(1)$ for $t \geq \mu_{\min}$, and it has been shown as for (23) that $s^{-1} |FP(t) - (p - s)\Phi(t)| = o_P(1)$ for $t \geq \mu_{\min}$. Then it follows that

$$
|\hat{s}^{-1} - s^{-1}| \cdot |FP(t) - p\Phi(t)| = o_P(1).
$$

Summing up the above gives (33).

References

Arias-Castro, E., E. J. Candès, and Y. Plan (2011). Global testing under sparse alternatives: Anova, multiple comparisons and the higher criticism. *The Annals of Statistics* 39(5), 2533–2556.

Barber, R. F. and E. J. Candès (2015). Controlling the false discovery rate via knockoffs.
Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. Ser. B* 57(1), 289–300.

Bodis-Wollner, I., S. Bucher, K. Seelos, W. Paulus, M. Reiser, and W. Oertel (1997). Functional mri mapping of occipital and frontal cortical activity during voluntary and imagined saccades. *Neurology* 49(2), 416–420.

Bruce, C. J. and M. E. Goldberg (1985). Primate frontal eye fields. i. single neurons discharging before saccades. *Journal of neurophysiology* 53(3), 603–635.

Bruce, C. J., M. E. Goldberg, M. C. Bushnell, and G. B. Stanton (1985). Primate frontal eye fields. ii. physiological and anatomical correlates of electrically evoked eye movements. *Journal of neurophysiology* 54(3), 714–734.

Cai, T., J. Jin, and M. Low (2007). Estimation and confidence sets for sparse normal mixtures. *The Annals of Statistics* 35(6), 2421–2449.

Cai, T. T., X. J. Jeng, and J. Jin (2011). Optimal detection of heterogeneous and heteroscedastic mixtures. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73(5), 629–662.

Cai, T. T. and J. Jin (2010). Optimal rates of convergence for estimating the null density and proportion of nonnull effects in large-scale multiple testing. *The Annals of Statistics*, 100–145.

Cai, T. T. and W. Sun (2017). Optimal screening and discovery of sparse signals with applications to multistage high-throughput studies. *Journal of the Royal Statistical Society: Series B* 79(1), 197–223.

Candes, E., Y. Fan, L. Janson, and J. Lv (2018). Panning for gold: Model-x knockoffs for high dimensional controlled variable selection. *J. R. Statist. Soc. Ser. B* 80(3), 551–577.
Chen, X. and R. Doerge (2016). A strong law of large numbers related to multiple testing normal means. arXiv:1410.4276v3.

Donoho, D. and J. Jin (2004). Higher criticism for detecting sparse heterogeneous mixtures. The Annals of Statistics 32(3), 962–994.

Donoho, D. and J. Jin (2015). Special invited paper: Higher criticism for large-scale inference, especially for rare and weak effects. Statistical Science, 1–25.

Dudoit, S. and M. J. Van Der Laan (2007). Multiple testing procedures with applications to genomics. Springer Science & Business Media.

Fan, J., X. Han, and W. Gu (2012). Estimating false discovery proportion under arbitrary covariance dependence. J. Amer. Statist. Assoc. 107(499), 1019–1035.

Feller, W. (1971). An Introduction to Probability Theory and its Applications, Volume II. Wiley, NewYork, NY.

Fowlkes, E. B. and C. L. Mallows (1983). A method for comparing two hierarchical clusterings. Journal of the American statistical association 78(383), 553–569.

Genovese, C. and L. Wasserman (2004). A stochastic process approach to false discovery control. Annals of Statistics 32, 1035–1061.

Goldberg, M. E., J. Bisley, K. D. Powell, J. Gottlieb, and M. Kusunoki (2002). The role of the lateral intraparietal area of the monkey in the generation of saccades and visuospatial attention. Annals of the New York Academy of Sciences 956(1), 205–215.

Halkidi, M., Y. Batistakis, and M. Vazirgiannis (2001). On clustering validation techniques. Journal of intelligent information systems 17(2-3), 107–145.

Jeng, X. J. and X. Chen (2019a). Predictor ranking and false discovery proportion control in high-dimensional regression. J. Multivariate Anal. 171, 163–175.

Jeng, X. J. and X. Chen (2019b). Variable selection via adaptive false negative control in linear regression. Electron. J. Statist. 13(2), 5306–5333.
Jeng, X. J., Z. J. Daye, W. Lu, and J.-Y. Tzeng (2016). Rare variants association analysis in large-scale sequencing studies at the single locus level. *PLoS computational biology* 12(6).

Jeng, X. J., T. Zhang, and J.-Y. Tzeng (2019). Efficient signal inclusion with genomic applications. *Journal of the American Statistical Association* 114, 1787–1799.

Meinshausen, N. and P. Bühlmann (2005). Lower bounds for the number of false null hypotheses for multiple testing of associations under general dependence structures. *Biometrika* 92(4), 893–907.

Meinshausen, N. and J. Rice (2006). Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. *Ann. Statist. 34*(1), 373–393.

Petticrew, M., A. Sowden, D. Lister-Sharp, and K. Wright (2000). False-negative results in screening programmes: systematic review of impact and implications. *Health technology assessment (Winchester, England)* 4(5), 1–120.

Pinho, A. L., A. Amadon, T. Ruest, M. Fabre, E. Dohmatob, I. Denghien, C. Ginisty, S. Becuwe-Desmidt, S. Roger, L. Laurier, et al. (2018). Individual brain charting, a high-resolution fmri dataset for cognitive mapping. *Scientific data* 5, 180105.

Schluppeck, D., P. Glimcher, and D. J. Heeger (2005). Topographic organization for delayed saccades in human posterior parietal cortex. *Journal of neurophysiology* 94(2), 1372–1384.