Journal of Machine Learning Research 25 (2024) 1-67 Submitted 9/21; Revised 11/23; Published 1/24

Power of Knockoff: The Impact of Ranking Algorithm, Augmented Design, and Symmetric Statistic

Zheng Tracy Ke ZKE@FAS.HARVARD.EDU
Department of Statistics
Harvard University
Cambridge, MA 02138, USA

Jun S. Liu JLIU@STAT.HARVARD.EDU
Department of Statistics
Harvard University
Cambridge, MA 02138, USA

Yucong Ma YUCONGMA@G.HARVARD.EDU
Department of Statistics
Harvard University
Cambridge, MA 02138, USA

Editor: Ji Zhu

Abstract

The knockoff filter is a recent false discovery rate (FDR) control method for high-dimensional linear models. We point out that knockoff has three key components: ranking algorithm, augmented design, and symmetric statistic, and each component admits multiple choices. By considering various combinations of the three components, we obtain a collection of variants of knockoff. All these variants guarantee finite-sample FDR control, and our goal is to compare their power. We assume a Rare and Weak signal model on regression coefficients and compare the power of different variants of knockoff by deriving explicit formulas of false positive rate and false negative rate. Our results provide new insights on how to improve power when controlling FDR at a targeted level. We also compare the power of knockoff with its *propotype* - a method that uses the same ranking algorithm but has access to an ideal threshold. The comparison reveals the additional price one pays by finding a data-driven threshold to control FDR.

Keywords: CI-knockoff, Hamming error, phase diagram, Rare/Weak signal model, SDP-knockoff, variable ranking, variable selection

1 Introduction

We consider a linear regression model, where $y \in \mathbb{R}^n$ is the vector of responses and $X \in \mathbb{R}^{n\times p}$ is the design matrix. We assume

$$y = X\beta + e, \qquad X = [X_1, X_2, \ldots, X_n]' \in \mathbb{R}^{n\times p}; \quad \beta \in \mathbb{R}^p, \quad e \sim N(0, \sigma^2 I_n). \tag{1.1}$$

Driven by the interests of high-dimensional data analysis, we assume p is large and β is a sparse vector (i.e., many coordinates of β are zero). Variable selection is the problem of estimating the true support of β. Let $\hat{S} \subset \{1, 2, \ldots, p\}$ denote the set of selected variables.

The false discovery rate (FDR) is defined to be

$$\mathbb{E}\Big[\frac{\#\{j:\beta_j=0, j\in\hat{S}\}}{\#\{j:j\in\hat{S}\}\vee 1}\Big].$$

Controlling FDR is a problem of great interest. When the design is orthogonal (i.e., $X'X$ is a diagonal matrix), the BH-procedure (Benjamini and Hochberg, 1995) can be employed to control FDR at a targeted level. When the design is non-orthogonal, the BH-procedure faces challenges, and several recent FDR control methods were proposed, such as the knockoff filter (Barber and Candès, 2015), model-X knockoff (Candes et al., 2018), Gaussian mirror (Xing et al., 2023), and multiple data splits (Dai et al., 2022). All these methods are shown to control FDR at a targeted level, but their power is less studied. This paper aims to provide a theoretical understanding to the power of FDR control methods.

We introduce a unified framework that captures the key ideas behind recent FDR control methods. Starting from the seminal work of Barber and Candès (2015), this framework has been implicitly used in the literature, but it is the first time that we abstract it out:

(a) There is a *ranking algorithm*, which assigns an importance metric to each variable.

(b) An FDR control method creates an *augmented design matrix* by adding fake variables.

(c) The augmented design and the response vector y are supplied to the ranking algorithm as input, and the output is converted to a (signed) importance metric for each original variable through a *symmetric statistic*.

The three components, ranking algorithm, augmented design, and symmetric statistic, should coordinate so that the resulting importance metrics for null variables ($\beta_j = 0$) have symmetric distributions and that the importance metrics for non-null variables ($\beta_j \neq 0$) are positive with high probability. When these requirements are satisfied, one can mimic the BH procedure (Benjamini and Hochberg, 1995) to control FDR at a targeted level.

The choices of the three components are not unique. For example, we may use any linear regression method $\hat{\beta}$ as the ranking algorithm, where we assign $|\hat{\beta}_j|$ as the importance metric for variable j. Similarly, the other two components also admit multiple choices. This leads to many different combinations of the three components. The literature has revealed insights on how to choose these components to get a valid FDR control, but there is little understanding on how to design them to boost power. The main contribution of this paper is dissecting and detailing the impact of each component on the power.

1.1 Main results and discoveries

We start from the orthodox knockoff in Barber and Candès (2015), which uses Lasso as the ranking algorithm, a semi-definite programming (SDP) procedure to construct the augmented design, and the signed maximum function as the symmetric statistic. We then replace each component of the orthodox knockoff by a popular alternative choice in the literature. We compare the power of the resulting variant of knockoff with the power of the orthodox one. This serves to reveal the impact of each component on power.

Our results lead to some noteworthy discoveries: (i) For the choice of symmetric statistic, the signed maximum is better than a popular alternative - the difference statistic; (ii) For

the choice of the augmented design, the SDP approach in orthodox knockoff is less favored than a recent alternative - the conditional independence approach (Liu and Rigollet, 2019); (iii) For the choice of ranking algorithm, we compare Lasso and least-squares and find that Lasso has an advantage when the signals are extremely sparse and least-squares has an advantage when the signals are moderately sparse.

For each variant of knockoff, we also consider its *prototype*, which applies the ranking algorithm to the original design X and selects variables by applying an ideal threshold on the importance metrics output by the ranking algorithm. We note that the core idea of knockoff is hinged on the other two components - augmented design and symmetric statistic, as these two components serve to find a data-driven threshold on importance metrics. Therefore, the comparison of knockoff and its prototype reveals the key difference between FDR control and variable selection - we need to pay an extra price to find a data-driven threshold. If an FDR control method is designed effectively, it should have a negligible power loss compared with its prototype. In the knockoff framework, when the design is orthogonal or blockwise diagonal and when the ranking algorithm is Lasso, we can show that knockoff (with proper choices of augmented design and symmetric statistic) indeed yields a negligible power loss compared with its own prototype. On the other hand, this is not true for a general design or when the ranking algorithm is not Lasso.

1.2 The theoretical framework and criteria of power comparison

Let $G = X'X \in \mathbb{R}^{p\times p}$ be the Gram matrix. Without loss of generality, we assume that each column of X has been normalized so that the diagonal entries of G are all equal to 1.[1] We study a challenging regime of "Rare and Weak signals" (Donoho and Jin, 2015; Jin and Ke, 2016), where for some constants $\vartheta \in (0,1)$ and $r > 0$, we consider settings where

$$\#\{j : \beta_j \neq 0\} \ \sim \ p^{1-\vartheta}, \qquad |\beta_j| \ \sim \ \sqrt{2r\log(p)} \text{ if } \beta_j \neq 0.$$

The two parameters, ϑ and r, characterize the signal rarity and signal weakness, respectively. Here, $\sqrt{\log(p)}$ is the minimax order for a successful inference of the support of β (Genovese et al., 2012), and the constant factor r drives subtle phase transitions. This model is widely used in multiple testing (Donoho and Jin, 2004; Arias-Castro et al., 2011; Barnett et al., 2017) and variable selection (Ji and Jin, 2012; Jin et al., 2014; Ke et al., 2014).

The power of an FDR control method depends on the target FDR level q. Instead of fixing q, we derive a trade-off diagram between FDR and the true positive rate (TPR) as q varies. This trade-off diagram provides a full characterization of power, for any given model parameters (ϑ, r). We also derive a phase diagram (Jin and Ke, 2016) for each FDR control method. The phase diagram is a partition of the two-dimensional space (ϑ, r) into different regions, according to the asymptotic behavior of the Hamming error (i.e., the expected sum of false positives and false negatives). The phase diagram provides a visualization of power for all (ϑ, r) together. Both the FDR-TPR trade-off diagram and phase diagram can be used as criteria of power comparison. We prefer the phase diagram, because a single phase diagram covers the whole parameter range (in contrast, the FDR-TPR trade-off diagram is

1. We use a conventional normalization of X in the study of Rare/Weak signal models. It is different from the standard normalization where the diagonal entries of G are assumed to be n. We note that our β is actually $\sqrt{n}\beta$ in the standard normalization. This is why n disappears in the order of nonzero β_j.

tied to a specified (ϑ, r)). Throughout the paper, we use phase diagram to compare different variants of knockoff. At the same time, we also give explicit forms of false positive rate and false negative rate, from which the FDR-TPR trade-off diagram can be deduced easily.

1.3 Related literature

Power analysis of FDR control methods is a small body of literature. Su et al. (2017) set up a framework for studying the trade-off between false positive rate and true positive rate across the lasso solution path. Weinstein et al. (2017) and Weinstein et al. (2021) extended this framework to find a trade-off for the knockoff filter, when the ranking algorithm is the Lasso and thresholded Lasso, respectively. These trade-off diagrams are for linear sparsity (number of nonzero coefficients of β is a constant fraction of p) and independent Gaussian designs ($X(i,j)$ are iid $N(0, n^{-1/2})$ variables). However, their analysis and results do not apply to our setting: In our setting, β is much sparser, and the overall signal strength as characterized by $\|\beta\|$ is much smaller. Furthermore, we are primarily interested in correlated designs, but their study is mostly focused on iid Gaussian designs.

For correlated designs, Liu and Rigollet (2019) gave sufficient and necessary conditions on X such that knockoff has a full power, but they did not provide an explicit trade-off diagram. Moreover, their analysis does not apply to the orthodox knockoff but only to a variant of knockoff that uses de-biased Lasso as the ranking algorithm. Beyond linear sparsity, Fan et al. (2019) studied the power of model-X knockoff for arbitrary sparsity, under a stronger signal strength: we assume $|\beta_j| \asymp \sqrt{\log(p)}$, while they assumed $|\beta_j| \gg \sqrt{\log(p)}$. In a similar setting, Javanmard and Javadi (2019) studied the power of using de-biased Lasso for FDR control. Our paper differs from these works because we study the regime of weaker signals and also derive the explicit FDR-TPR trade-off diagrams and phase diagrams.

Wang and Janson (2022) and Spector and Janson (2022) studied the power of model-X knockoff and conditional randomization tests. They considered linear sparsity and iid Gaussian designs, and found a disadvantage of power by constructing augmented design as in the orthodox knockoff (with least-squares as the ranking algorithm). This qualitatively agrees with some of our conclusions in Section 5, but it is for a different setting with uncorrelated variables and linear sparsity. We also study more variants of knockoff than those considered in aforementioned works. Recently, Li and Fithian (2021) recast the fix-X knockoff as a conditional post-selection inference method and studied its power.

In a sequel of papers (Ji and Jin, 2012; Jin et al., 2014; Ke et al., 2014), the Rare/Weak signal model was used to study variable selection. They focused on the class of Screen-and-Clean methods for variable selection and proved its optimality under various design classes. We borrowed the notion of phase diagram from these works. However, they did not consider any FDR control method and their methods do not apply to knockoff. Different from the proof techniques employed by the aforementioned work, our proof is based on a geometric approach, where the key is studying the geometric properties of the "rejection region" induced by knockoff (see Section 7 for details).

1.4 Organization

The remainder of this paper is organized as follows. Section 2 reviews the idea of knockoff. Section 3 introduces the Rare/Weak signal model and explains how to use it as a theoretical

platform to study and compare the power of FDR control methods. Sections 4-6 contain the main results, where we study the impact of symmetric statistic, augmented design, and ranking algorithm, respectively. Section 7 sketches the proof and explains the geometrical insight behind the proof. Section 8 contains simulation results. Section 9 concludes with a short discussion. Detailed proofs are relegated to the Appendix.

2 The knockoff filter, its variants and prototypes

Let us first review the orthodox knockoff filter (Barber and Candès, 2015). Write $G = X'X$ and let $\mathrm{diag}(s)$, with $s \in \mathbb{R}^p$, be a nonnegative diagonal matrix (to be chosen by the user) such that $\mathrm{diag}(s) \preceq 2G$. The knockoff first creates a design matrix $\tilde{X} \in \mathbb{R}^{n\times p}$ such that

$$\tilde{X}'\tilde{X} = G, \qquad X'\tilde{X} = G - \mathrm{diag}(s). \tag{2.1}$$

Let x_j and $\tilde{x}_j$ be the jth column of X and $\tilde{X}$, respectively, $1 \le j \le p$. Here, $\tilde{x}_j$ is called a *knockoff* of variable j. For any $\lambda > 0$, let $\hat{\beta}(\lambda) \in \mathbb{R}^{2p}$ be the solution of Lasso (Tibshirani, 1996) on the expanded design matrix $[X, \tilde{X}]$ with a tuning parameter λ:

$$\hat{\beta}(\lambda) = \mathrm{argmin}_b\big\{\|y - [X, \tilde{X}]b\|^2/2 + \lambda\|b\|_1\big\}. \tag{2.2}$$

For each $1 \le j \le p$, let $Z_j = \sup\{\lambda > 0 : \hat{\beta}_j(\lambda) \neq 0\}$ and $\tilde{Z}_j = \sup\{\lambda > 0 : \hat{\beta}_{p+j}(\lambda) \neq 0\}$. The importance of variable j is measured by a *symmetric statistic*

$$W_j = f(Z_j, \tilde{Z}_j), \tag{2.3}$$

where $f(\cdot,\cdot)$ is a bivariate function satisfying $f(v,u) = -f(u,v)$. Here $\{W_j\}_{j=1}^p$ are (signed) importance metrics for variables. Under some regularity conditions, it can be shown that W_j has a symmetric distribution when $\beta_j = 0$ and that W_j is positive with high probability when $\beta_j \neq 0$. Given a threshold $t > 0$, the number of false discoveries is equal to

$$\#\{j : \beta_j = 0, W_j > t\} \;\approx\; \#\{j : \beta_j = 0, W_j < -t\} \;\approx\; \#\{j : W_j < -t\},$$

where the first approximation is based on the symmetry of the distribution of W_j for null variables and the second approximation comes from the sparsity of β. The right hand side gives an estimate of the number of false discoveries. Hence, a data-driven threshold to control FDR at q is

$$\hat{T}(q) = \min\left\{t > 0 : \frac{\#\{j : W_j < -t\}}{\#\{j : W_j > t\} \vee 1} \le q\right\}. \tag{2.4}$$

The set of selected variables is $\hat{S}(q) = \{j : W_j > \hat{T}(q)\}$. As long as $\tilde{X}$ in (2.1) exists, it can be shown that the FDR associated with $\hat{S}(q)$ is guaranteed to be $\le q$.

2.1 Variants of knockoff

The idea of knockoff provides a general framework for FDR control. It consists of three key components, as summarized below:

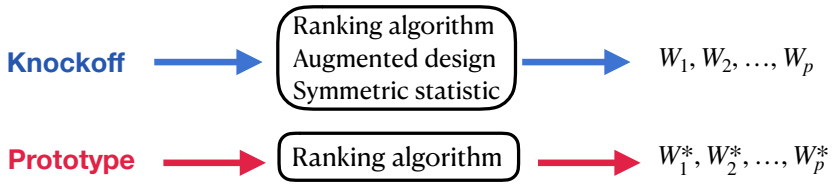

Figure 1: An illustration of the knockoff and its prototype.

(a) A ranking algorithm, which takes y and an arbitrary design and assigns an importance metric Z_j to each variable in the design. In (2.2), it uses a particular ranking algorithm based on the solution path of Lasso.

(b) An augmented design, which is the $n \times (2p)$ matrix $[X, \tilde{X}]$, where a knockoff $\tilde{x}_j$ is created for each original x_j. We supply the augmented design to the ranking algorithm to get importance metrics Z_j and $\tilde{Z}_j$ for each variable j and its knockoff.

(c) A symmetric statistic $f(\cdot, \cdot)$, which combines the two importance metrics Z_j and $\tilde{Z}_j$ to an ultimate importance metric W_j for variable j.

The choice of each component is non-unique. For (c), f can be any anti-symmetric function. Two popular choices are the *signed maximum* statistic and the *difference* statistic:

$$f^{\text{sgm}}(u, v) = \text{sgn}(u - v) \cdot \max\{u, v\}, \qquad \text{and} \qquad f^{\text{dif}}(u, v) = u - v. \tag{2.5}$$

For (b), the freedom comes from choosing $\text{diag}(s)$ and constructing $\tilde{X}$. In fact, once $\text{diag}(s)$ is given, it can be shown that any $\tilde{X}$ satisfying (2.1) yields the same asymptotic performance for knockoff. Hence, the choice of the augmented design boils down to choosing $\text{diag}(s)$. A popular option is the *SDP-knockoff*, which solves $\text{diag}(s)$ from a semi-definite programming:

$$\min \sum_j (1 - s_j), \qquad \text{subject to} \quad 0 \le s_j \le 1, \ \text{diag}(s) \preceq 2G. \tag{2.6}$$

Another option is the *CI-knockoff* (Liu and Rigollet, 2019):

$$\text{diag}(s) = c \cdot [\text{diag}(G^{-1})]^{-1}, \qquad \text{where } c = \sup\big\{0 < \tilde{c} \le 1 : c[\text{diag}(G^{-1})]^{-1} \preceq 2G\big\}. \tag{2.7}$$

For (a), Lasso is currently used as the ranking algorithm (see (2.2)), but it can be replaced by other linear regression methods. Take the least-squares $\hat{\beta}^{ols} = (X'X)^{-1}X'y$ for example. We can define a ranking algorithm that outputs $|\hat{\beta}_j^{ols}|$ as the importance metric. If we supply the augmented design $[X, \tilde{X}]$ to this ranking algorithm, then we have

$$\hat{\beta} = ([X, \tilde{X}]'[X, \tilde{X}])^{-1}[X, \tilde{X}]'y. \tag{2.8}$$

We can set $Z_j = |\hat{\beta}_j|$ and $\tilde{Z}_j = |\hat{\beta}_{j+p}|$ and plug them into (2.3).

In summary, the flexibility of the three components gives rise to many different variants of knockoff. For example, we can use (2.2) or (2.8) as the ranking algorithm, (2.6) or (2.7) as the augmented design, and either one in (2.5) as the symmetric statistic; this already

gives $2 \times 2 \times 2 = 8$ variants of knockoff. By the theory in Barber and Candès (2015), each variant guarantees the finite-sample FDR control. When the FDR is under control, the user always wants to select as many true signals as possible, i.e., to maximize the power. In this paper, one of our goals is to understand and compare the power of different variants of knockoff.

To this end, we start from the default choices of the three components in the orthodox knockoff, where the ranking algorithm is Lasso as in (2.2), the augmented design is the SDP-knockoff as in (2.6), and the symmetric statistic is the signed maximum in (2.5). In Sections 4-6, we successively alter each component and study its impact on the power.

2.2 Prototypes of knockoff

Given a variant of knockoff (where a specific choice of the ranking algorithm is applied to the augmented design $[X, \tilde{X}]$), we define the corresponding *prototype* method as follows: It runs the same ranking algorithm on the original design matrix X, and outputs $W_1^*, W_2^*, \ldots, W_p^*$ as importance metrics. The method selects variables by thresholding W_j^* at T_q^*, where

$$T_q^* = \min\left\{t > 0 : \mathbb{E}\left[\frac{\#\{j : \beta_j = 0, W_j^* > t\}}{\#\{j : W_j^* > t\} \vee 1}\right] \le q\right\}. \tag{2.9}$$

Compared with knockoff, the prototype ranks variables by W_j^*, whose induced ranking may be different from the one by W_j. Additionally, the prototype has access to an ideal threshold T_q^* that guarantees an exact FDR control but is practically infeasible. In contrast, knockoff has to find a data-driven threshold from (2.4). See Figure 1.

We look at two examples. Consider the orthodox knockoff, where the ranking algorithm is Lasso (see (2.2)). Its prototype runs Lasso on X to get $\hat{\beta}^{\text{lasso}}(\lambda) = \text{argmin}_b\{\|y - Xb\|^2/2 + \lambda\|b\|_1\}$ and assigns an importance metric to variable j as

$$W_j^* = \sup\big\{\lambda > 0 : \hat{\beta}_j^{\text{lasso}}(\lambda) \neq 0\big\}. \tag{2.10}$$

It then selects variables by thresholding W_j^* using the ideal threshold in (2.9). We call this method the *Lasso-path*. It is the prototype of all the variants of knockoff that use Lasso as ranking algorithm. For all variants of knockoff that use least-squares (see (2.8)) as the ranking algorithm, they share the same prototype, which computes $\hat{\beta}^{\text{ols}} = (X'X)^{-1}X'y$ and assigns an importance metric to variable j as

$$W_j^* = |\hat{\beta}_j^{\text{ols}}| = |e_j' G^{-1} X' y|. \tag{2.11}$$

It then selects variables by thresholding W_j^* using the ideal threshold in (2.9). We call this method the *least-squares*.

In this paper, besides comparing different variants of knockoff, we also aim to compare each variant with its prototype. Here is the motivation: FDR control splits into two tasks: (1) ranking the importance of variables and (2) finding a data-driven threshold. When the FDR is under control, the power depends on how well Task 1 is performed. Importantly, in knockoff, although the augmented design and the ranking algorithm are meant to carry Task 2 only, they do affect the final ranking of variables because the ranking by W_j's is usually different from the ranking by W_j^*'s. This yields a potential power loss compared

with its prototype – a price we pay for finding a data-driven threshold. Hence, a power comparison between knockoff and its prototype helps us understand how large this price is.

Remark 1. In this paper, we focus on two ranking algorithms, Lasso-path and least-squares. They are both tuning-free. When the ranking algorithm has tuning parameters, we should not set the tuning parameters in the prototype the same as in the original knockoff. For example, we can use $|\hat{\beta}_j^{\text{lasso}}(\lambda)|$ to rank variables, treating λ as a tuning parameter. The prototype runs lasso on the original design with p variables, while knockoff runs lasso on the augmented design with $2p$ variables. The optimal λ that minimizes the expected Hamming error is different in the two scenarios. A reasonable approach of comparing knockoff and its prototype is to use their respective optimal λ. This will require computing the expected Hamming error of lasso for an arbitrary λ (Ji and Jin, 2012; Ke and Wang, 2021).

3 Rare/Weak signal model and criteria of power comparison

We introduce our theoretical framework of power comparison. Recall that we consider a linear model $y = X\beta + e$, where $y \in \mathbb{R}^n$, $X = [X_1, X_2, \ldots, X_n]' \in \mathbb{R}^{n\times p}$, and $e \sim N(0, \sigma^2 I_n)$. Without loss of generality, fix $\sigma = 1$. Given p, we allow n to be any integer such that $n \geq 2p$. This is from the requirement of knockoff (it needs $n \geq 2p$ to guarantee the existence of $\tilde{X}$ in (2.1)) and should not be viewed as a limitation of our theory. Our results are extendable to $n < 2p$, provided that knockoff is replaced by its extension in this case (see Section 9 for a discussion). The Gram matrix is

$$G := X'X \in \mathbb{R}^{p\times p}, \qquad \text{where we assume } G_{jj} = 1, \text{ for all } 1 \leq j \leq p. \tag{3.1}$$

We adopt the Rare/Weak signal model (Donoho and Jin, 2004) to assume that β satisfies:

$$\beta_j \overset{iid}{\sim} (1-\epsilon_p)\nu_0 + \epsilon_p \nu_{\tau_p}, \qquad 1 \leq j \leq p, \tag{3.2}$$

where ν_a denotes a point mass at a. Here, $\epsilon_p \in (0,1)$ is the expected fraction of signals, and $\tau_p > 0$ is the signal strength. We let p be the driving asymptotic parameter and tie (ϵ_p, τ_p) with p through fixed constants $\vartheta \in (0,1)$ and $r > 0$:

$$\epsilon_p = p^{-\vartheta}, \qquad \tau_p = \sqrt{2r\log(p)}. \tag{3.3}$$

The parameters, ϑ and r, characterize the signal rarity and the signal strength, respectively. Here, n does not appear in the order of nonzero β_j, because we have already re-parameterized (X, β) such that the diagonals of G are 1 (see Footnote 1).

Under the Rare/Weak signal model (3.2)-(3.3), we define two diagrams for characterizing the power of knockoff. Let W_j be the ultimate importance metric (2.3) assigned to variable j, and consider the set of selected variables at a threshold $\sqrt{2u\log(p)}$:

$$\hat{S}(u) = \big\{1 \leq j \leq p : W_j > \sqrt{2u\log(p)}\big\}.$$

Let $S = \{1 \leq j \leq p : \beta_j \neq 0\}$. Define $\text{FP}_p(u) = \mathbb{E}(|\hat{S}(u)\backslash S|)$, $\text{FN}_p(u) = \mathbb{E}(|S\backslash\hat{S}(u)|)$, and $\text{TP}_p(u) = \mathbb{E}(|S \cap \hat{S}(u)|)$, where the expectation is taken with respect to the randomness of both β and y. Write $s_p = p\epsilon_p$, and define

$$\text{Hamm}_p(u) = \text{FP}_p(u) + \text{FN}_p(u), \quad \text{FDR}_p(u) = \frac{\text{FP}_p(u)}{\text{FP}_p(u) + \text{TP}_p(u)}, \quad \text{TPR}_p(u) = \frac{\text{TP}_p(u)}{s_p}.$$

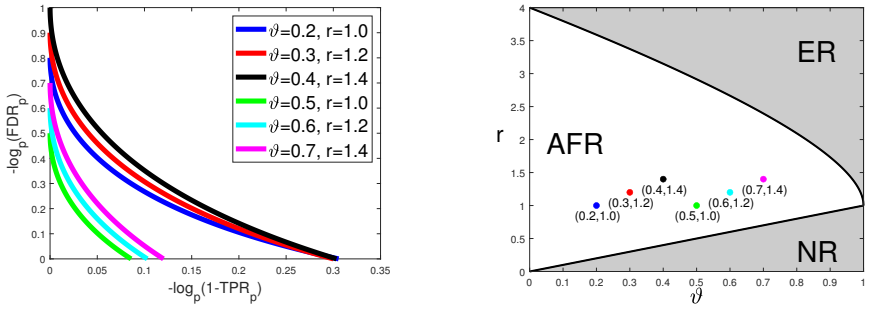

Figure 2: Left: the FDR-TPR trade-off diagram for a few values of (ϑ, r). Right: the phase diagram. The design is orthogonal, and the importance metric is as in (3.6). Each FDR-TPR trade-off diagram corresponds to one point in the phase diagram.

The first quantity is the expected Hamming error. The last two quantities are proxies of the false discovery rate and true positive rate, respectively.[2]

The following definition is conventional in the study of Rare/Weak signal models (Genovese et al., 2012; Ji and Jin, 2012) and will be used frequently in our theoretical results:

Definition 3.1 (Multi-$\log(p)$ term) *Consider a sequence $\{a_p\}_{p=1}^{\infty}$. If for any fixed $\delta > 0$, $a_p p^{\delta} \to \infty$ and $a_p p^{-\delta} \to 0$, we call a_p a multi-$\log(p)$ term and write $a_p = L_p$ (L_p is a generic notion for all multi-$\log(p)$ terms). If there is a constant $b_0 \in \mathbb{R}$ such that $a_p p^{-b_0} = L_p$, we write $a_p = L_p p^{b_0}$, which means for any fixed $\delta > 0$, $a_p p^{-b_0+\delta} \to \infty$ and $a_p p^{-b_0-\delta} \to 0$.*

In the Rare/Weak signal model, for many classes of designs of interest, $\mathrm{FDR}_p(u)$ and $\mathrm{TPR}_p(u)$ satisfy a property: There exist two fixed functions $g_{\mathrm{FDR}}(u;\vartheta,r)$ and $g_{\mathrm{TPR}}(u;\vartheta,r)$ such that, for any (ϑ, r, u), as $p \to \infty$,

$$\mathrm{FDR}_p(u) = L_p p^{-g_{\mathrm{FDR}}(u;\vartheta,r)}, \qquad 1 - \mathrm{TPR}_p(u) = L_p p^{-g_{\mathrm{TPR}}(u;\vartheta,r)}. \tag{3.4}$$

The two functions g_{FDR} and g_{TPR} depend on the choice of the three components in knockoff and the design class. We propose the FDR-TPR trade-off diagram as follows:

Definition 3.2 (FDR-TPR trade-off diagram) *Given a variant of knockoff and a sequence of designs indexed by p, if $\mathrm{FDR}_p(u)$ and $\mathrm{TPR}_p(u)$ satisfy (3.4), the FDR-TPR trade-off diagram associated with (ϑ, r) is the plot with $g_{\mathrm{FDR}}(u;\vartheta,r)$ in the y-axis and $g_{\mathrm{TPR}}(u;\vartheta,r)$ in the x-axis, as u varies.*

An FDR-TPR trade-off diagram depends on (ϑ, r). To compare the performance of two variants of knockoff, we have to draw many curves for different values of (ϑ, r). Hence, we introduce another diagram, which characterizes the power simultaneously at all (ϑ, r). Define $\mathrm{Hamm}_p^* \equiv \min_u\{\mathrm{FP}_p(u) + \mathrm{FN}_p(u)\}$. This is the minimum expected Hamming selection

2. The $\mathrm{FDR}_p(u)$ we consider here is the ratio of expectations of false positives and total discoveries (called mFDR in some literature), not the original definition of FDR, which is the expectation of ratios of false positives and total discoveries. According to Javanmard and Montanari (2018), mFDR and FDR can be different in situations with high variability, but it is not the case here. In our setting, the expectation of total discoveries grows to infinity as a power of p, hence, mFDR and FDR have a negligible difference.

error when the threshold u is chosen optimally. For each variant of knockoff and each class of designs of interest, there exists a bivariate function $f^*_{\text{Hamm}}(\vartheta, r) \in [0,1]$ such that

$$\text{Hamm}^*_p = L_p p^{f^*_{\text{Hamm}}(\vartheta, r)}. \tag{3.5}$$

The phase diagram is defined as follows:

Definition 3.3 (Phase diagram) *When* Hamm^*_p *satisfies* (3.5)*, the phase diagram is defined to the partition of the two-dimensional space* (ϑ, r) *into three regions:*

- *Region of Exact Recovery (ER):* $\{(\vartheta, r) : f^*_{\text{Hamm}}(\vartheta, r) < 0\}$.
- *Region of Almost Full Recovery (AFR):* $\{(\vartheta, r) : 0 < f^*_{\text{Hamm}}(\vartheta, r) < 1 - \vartheta\}$.
- *Region of No Recovery (NR):* $\{(\vartheta, r) : f^*_{\text{Hamm}}(\vartheta, r) \geq 1 - \vartheta\}$.

The curves separating different regions are called phase curves. We use $h_{\text{AFR}}(\vartheta)$ *to denote the curve between NR and AFR, and* $h_{\text{ER}}(\vartheta)$ *the curve between AFR and ER.*

In the ER region, the expected Hamming error, Hamm^*_p, tends to zero. Therefore, with high probability, the support of β is exactly recovered. In the AFR region, Hamm^*_p does not tend to zero but is much smaller than $p\epsilon_p$ (which is the expected number of signals). As a result, with high probability, the majority of signals are correctly recovered. In the region of NR, Hamm^*_p is comparable with the number of signals, and variable selection fails. The phase diagram was introduced in the literature (Genovese et al., 2012; Ji and Jin, 2012) but has never been used to study FDR control methods.

We illustrate these definitions with an example. Both the FDR-TPR trade-off diagram and phase diagram only depend on the importance metrics assigned to variables. Therefore, they are also well-defined for the prototypes in Section 2.2. We consider a special class of designs, where $X'X = I_p$, and a prototype that assigns the importance metrics

$$W^*_j = |x_j' y|, \qquad 1 \leq j \leq p. \tag{3.6}$$

The next proposition is adapted from literature (Donoho and Jin, 2004; Ji and Jin, 2012) and proved in the Appendix. We use a_+ to denote $\max\{a, 0\}$, for any $a \in \mathbb{R}$.

Proposition 3.1 *Suppose* $X'X = I_p$ *and consider the importance metric in* (3.6)*. When* $r > \vartheta$*, the FDR-TPR trade-off diagram is given by* $g_{\text{FDR}}(u; \vartheta, r) = (u - \vartheta)_+$ *and* $g_{\text{TPR}}(u; \vartheta, r) = (\sqrt{r} - \sqrt{u})^2_+$*. The phase diagram is given by* $h_{\text{AFR}}(\vartheta) = \vartheta$ *and* $h_{\text{ER}}(\vartheta) = (1 + \sqrt{1 - \vartheta})^2$*.*

These diagrams are visualized in Figure 2.

As we have mentioned in Section 2.2, FDR control is composed of the task of ranking variables and the task of finding a data-driven threshold. It is the variable ranking that determines the power of an FDR control method simultaneously at all FDR levels q. The two diagrams in Definitions 3.2-3.3 only depend on the importance metrics assigned to variables, hence, they measure the *quality of variable ranking*, which is fundamental for power comparison of different FDR control methods when they all control FDR at the same level.

4 Impact of the symmetric statistic

We fix the choice of ranking algorithm and augmented design in the orthodox knockoff and compare using the two symmetric statistics in (2.5). For simplicity, we consider the orthogonal design where $X'X = I_p$. In this special case, the ranking algorithm in (2.2) reduces to calculating marginal regression coefficients and the output Z_j and $\tilde{Z}_j$ reduce to

$$Z_j = |x_j'y|, \qquad \text{and} \qquad \tilde{Z}_j = |\tilde{x}_j'y|, \qquad 1 \le j \le p. \tag{4.1}$$

In the augmented design, the choice of $\mathrm{diag}(s)$ in (2.6) reduces to $\mathrm{diag}(s) = I_p$. We consider a slightly more general form:

$$\mathrm{diag}(s) = (1-a)I_p, \qquad \text{where } -1 < a < 1 \text{ is a fixed constant.} \tag{4.2}$$

Given $\mathrm{diag}(s)$, the construction of $\tilde{X}$ is not unique, but all constructions lead to exactly the same FDR-TPR trade-off diagram and the same phase diagram. For this reason, we only specify $\mathrm{diag}(s)$ as in (4.2), but not the actual $\tilde{X}$. Fixing the above choices (4.1)-(4.2), we consider the two symmetric statistics in (2.5), which lead to the importance metrics of

$$W_j^{\mathrm{sgm}} = (Z_j \vee \tilde{Z}_j) \cdot \begin{cases} +1, & \text{if } Z_j > \tilde{Z}_j \\ -1, & \text{if } Z_j \le \tilde{Z}_j \end{cases}, \qquad \text{and} \qquad W_j^{\mathrm{dif}} = Z_j - \tilde{Z}_j. \tag{4.3}$$

We call the two variants of knockoff *knockoff-sgm* and *knockoff-diff*, respectively. The next theorem gives the explicit forms of $\mathrm{FP}_p(u)$ and $\mathrm{FN}_p(u)$ associated with these two variants. Its proof can be found in the Appendix.

Theorem 4.1 *Consider a linear regression model where* (3.1)-(3.3) *hold. Suppose* $n \ge 2p$ *and* $G = I_p$. *We construct* $\tilde{X}$ *in knockoff as in* (4.2), *for a constant* $a \in (-1, 1)$, *and let* Z_j *and* $\tilde{Z}_j$ *be as in* (4.1). *For any constant* $u > 0$, *let* $\mathrm{FP}_p(u)$ *and* $\mathrm{FN}_p(u)$ *be the expected numbers of false positives and false negatives, by selecting variables with* $W_j > \sqrt{2u\log(p)}$. *When* W_j *is the signed maximum statistic in* (4.3), *as* $p \to \infty$,

$$\mathrm{FP}_p(u) = L_p p^{1-u}, \qquad \mathrm{FN}_p(u) = L_p p^{1-\vartheta-\min\left\{\frac{(1-|a|)r}{2},\ (\sqrt{r}-\sqrt{u})_+^2\right\}}.$$

When W_j *is the difference statistic in* (4.3), *as* $p \to \infty$,

$$\mathrm{FP}_p(u) = L_p p^{1-u}, \qquad \mathrm{FN}_p(u) = L_p p^{1-\vartheta-\frac{(1-|a|)}{2}(\sqrt{r}-\sqrt{u})_+^2}.$$

Here, L_p is the generic multi-$\log(p)$ notion in Definition 3.1. For Theorem 4.1, using Mills' ration, we actually know that $L_p \asymp 1/\sqrt{\log(p)}$. For other theorems, we do not always know the exact order of L_p, but we can intuitively regard it as a polynomial of $\log(p)$.

Given Theorem 4.1 and Definitions 3.2-3.3, we can derive the explicit FDR-TPR tradeoff diagrams and the phase diagrams:

Corollary 4.1 *In the setting of Theorem 4.1, when* $r > \vartheta$, *the FDR-TPR trade-off diagram is given by*

$$g_{\mathrm{FDR}}(u;\vartheta,r) = (u-\vartheta)_+, \quad g_{\mathrm{TPR}}(u) = \begin{cases} \min\left\{\frac{(1-|a|)r}{2},\ (\sqrt{r}-\sqrt{u})_+^2\right\}, & \text{if } W_j = W_j^{\mathrm{sgn}}, \\ \frac{(1-|a|)}{2}(\sqrt{r}-\sqrt{u})_+^2 & \text{if } W_j = W_j^{\mathrm{dif}}. \end{cases}$$

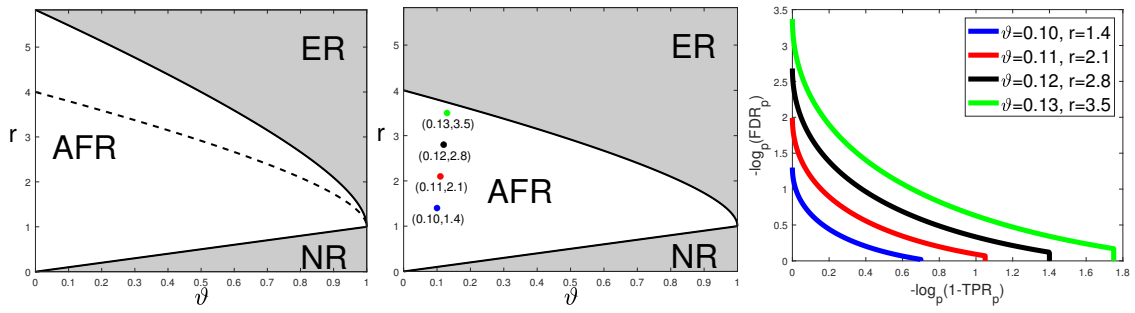

Figure 3: Power comparison of knockoff with different symmetric statistics (orthogonal design; ranking algorithm is Lasso, and augmented design is such that $\mathrm{diag}(s) = I_p$). The left two panels are the phase diagrams of knockoff-diff (left) and knockoff-sgm (middle), where the dashed lines are the phase curves of their common prototype. The right panel is the FDR-TPR trade-off diagram of knockoff-sgm, where each trade-off curve corresponds to one point in the phase diagram.

The phase diagram is given by

$$h_{\mathrm{AFR}}(\vartheta) = \vartheta, \qquad h_{\mathrm{ER}}(\vartheta) = \begin{cases} \max\big\{\frac{2-2\vartheta}{1-|a|},\ (1+\sqrt{1-\vartheta})^2\big\}, & \text{if } W_j = W_j^{\mathrm{sgn}}, \\ \Big(1+\sqrt{\frac{2-2\vartheta}{1-|a|}}\Big)^2, & \text{if } W_j = W_j^{\mathrm{dif}}. \end{cases}$$

For both knockoff-sgm and knockoff-diff, by Corollary 4.1, the best choice of a is $a = 0$. In the remaining of this section, we fix $a = 0$. Figure 3 gives visualizations of the FDR-TPR trade-off diagrams and phase diagrams. The prototype is Lasso-path (see (2.10)). In the orthogonal design $X'X = I_p$, Lasso-path reduces to the prototype in (3.6), whose FDR-TPR trade-off diagram and phase diagram are given in Figure 2.

Comparison of two symmetric statistics: First, we compare the phase diagrams in Figures 2-3 and find that (i) knockoff-sgm has a strictly better phase diagram than knockoff-diff, and (ii) knockoff-sgm has the same phase diagram as the prototype. It suggests that signed maximum is a better choice of symmetric statistic. It also suggests that knockoff-sgm yields a negligible power loss relative to its prototype.

We also point out that knockoff-sgm is already "optimal" among all symmetric statistics, in this orthogonal design. The reason is that, when $X'X = I_p$, the Hamming error (Hamm_p^*) has an information-theoretical lower bound (Genovese et al., 2012; Ji and Jin, 2012), whose induced phase diagram coincides with the phase diagram of the prototype, which is also the phase diagram of knockoff-sgm. This is the optimal phase diagram any method can achieve (including all variants of knockoff with other symmetric statistics).

Next, we compare the FDR-TPR trade-off diagrams of knockoff and the prototype. We focus on knockoff-sgm, whose trade-off diagram is in Figure 3 (right panel). The trade-off diagram of the prototype is in Figure 2 (right panel). We find that the trade-off diagram of knockoff-sgm is slightly different from the one of the prototype. By Theorem 4.1, $(1 - \mathrm{TPR}_p) = \mathrm{FN}_p/s_p \geq L_p p^{-r/2}$; hence, the FDR-TRP trade-off curve is truncated at $r/2$ in the x-axis. For large ϑ, the curve hits zero before the x-axis reaches $r/2$, and the truncation

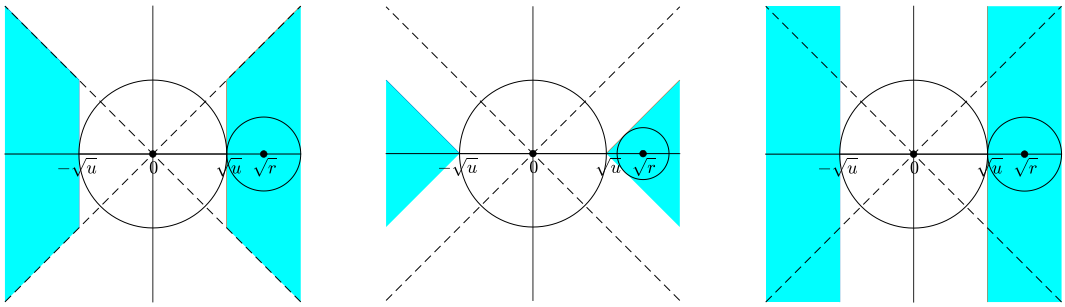

Figure 4: The rejection region of the symmetric statistics (orthogonal design, $a = 0$ in the construction of knockoff variables). Left: the signed maximum statistic. Middle: the difference statistic. Right: the prototype.

has no impact. However, for small ϑ, the curve has changed due to the truncation (Figure 3, right panel, all but the blue curve).

Some geometric insights, especially why signed maximum is "optimal". By (4.1) and (4.3), the importance metrics produced by knockoff can be written as $W_j = I(x_j'y, \tilde{x}_j'y)$, where x_j and $\tilde{x}_j$ are the jth variable and its knockoff, and $I(\cdot,\cdot)$ is a bivariate function depending on the choice of symmetric statistic. Define the "rejection region" as

$$\mathcal{R} = \Big\{(h_1, h_2) \in \mathbb{R}^2 : \ I\Big(h_1\sqrt{2\log(p)},\, h_2\sqrt{2\log(p)}\Big) > \sqrt{2u\log(p)}\Big\}.$$

Figure 4 shows the rejection region induced by knockoff-sgm, knockoff-diff, and their prototype. Write $\hat{h}_1 = x_j'y/\sqrt{2\log(p)}$ and $\hat{h}_2 = \tilde{x}_j'y/\sqrt{2\log(p)}$. The random vector $(\hat{h}_1, \hat{h}_2)'$ follows the bivariate normal distribution with covariance matrix $\frac{1}{\log(p)}I_2$. Its mean vector is $(0,0)'$ when $\beta_j = 0$ and $(\sqrt{r}, 0)'$ when $\beta_j = \tau_p$. By Lemma 7.1 (to be introduced in Section 7), the exponent in FP_p is determined by the Euclidean distance from $(0,0)'$ to $\mathcal{R}$ and the exponent in FN_p is determined by the Euclidean distance from $(\sqrt{r}, 0)'$ to $\mathcal{R}^c$. From Figure 4, it is clear that the difference statistic is inferior to the signed maximum statistic because the distance from $(\sqrt{r}, 0)'$ to $\mathcal{R}^c$ is strictly smaller in the former.

As we have mentioned earlier, signed maximum is "optimal" among all symmetric statistics, because its phase diagram already matches with the information-theoretic lower bound. Now, we use Figure 4 to provide a geometric interpretation of why signed maximum is "optimal". We call a subset $\mathcal{R}$ an eligible rejection region if there exists a symmetric statistic $f(\cdot,\cdot)$ whose induced rejection region is $\mathcal{R}$. It is not hard to see that any eligible $\mathcal{R}$ should be symmetric with respect to both x-axis and y-axis. In addition, by the anti-symmetry requirement $f(v,u) = -f(u,v)$, an eligible rejection region also needs to satisfy the following necessary condition: $\mathcal{R} \cap \mathcal{R}_\pm = \emptyset$, *where* $\mathcal{R}_\pm$ *is the reflection of* $\mathcal{R}$ *with respect to the line* $y = \pm x$. The prototype has the optimal phase diagram, but its rejection region $\mathcal{R}_0$ (Figure 4, right panel) does not satisfy this condition. We can see that the rejection region of knockoff-sgm (left panel) is a *minimal* modification of $\mathcal{R}_0$ to tailor to this condition. This partially explains why signed maximum is already the best choice in this setting.

Remark 2. We consider a non-stochastic threshold $\sqrt{2u\log(p)}$ in Theorem 4.1. For a data-driven threshold $\sqrt{2\hat{u}\log(p)}$, if there is $\epsilon_p = o(1)$ such that $|\hat{u}-u| \leq \epsilon_p$ with probability $1 - o(p^{-2})$, then Theorem 4.1 continues to hold.

Remark 3. In this section, we conduct power comparison only on the orthogonal design. Our rationale is as follows: A good method has to at least perform well in the simplest case. If a method is inferior to others in the orthogonal design, then we do not expect it to have a good potential in real applications where the designs can be much more complicated, i.e., the results on orthogonal designs help us filter out those methods that have little potential in practice. Such insights are valuable to users.

Remark 4. Recently, Weinstein et al. (2021) showed a remarkable result: Under linear sparsity and random Gaussian design, they found that knockoff with the difference symmetric statistic has great power. We note that the prototype of their knockoff is *thresholded Lasso*, not Lasso, and so the gain of power is primarily from prototype instead of symmetric statistic. Also, see Ke and Wang (2021) for a comparison of thresholded-Lasso v.s. Lasso.

5 Impact of the augmented design

We fix the choice of ranking algorithm and symmetric statistic as in the orthodox knockoff and compare using two augmented designs, the SDP-knockoff in (2.6) and the CI-knockoff in (2.7). We also call the two respective variants of knockoff the *SDP-knockoff* and *CI-knockoff*, so that "SDP-knockoff" (say) has two meanings, an augmented design or a variant of knockoff, depending on the context.

When $X'X = I_p$, SDP-knockoff and CI-knockoff are the same and reduce to $\mathrm{diag}(s) = I_p$, so it is impossible to tell their difference in power. We must consider non-orthogonal designs. However, since there is no explicit form of the Lasso solution path, the results for a general X are difficult to obtain; despite technical challenges, the phase diagrams may be too messy to provide any useful insight. We hope to find a class of non-orthogonal designs such that (i) it is mathematically tractable, (ii) it is considerably different from the orthogonal design and allows G to have some "large" off-diagonal entries, and (iii) it captures some key features of real applications. We start from a class of row-wise sparse designs, which approximate the designs in many real applications (e.g., in bioinformatics and in compressed sensing). The next proposition is adapted from Lemma 1 of Jin et al. (2014), whose proof is omitted.

Proposition 5.1 *Consider a linear model where* (3.1)-(3.3) *hold. Suppose each row of G has at most L_p nonzero entries, where L_p is a multi-*$\log(p)$ *term as in Definition 3.1. Let S be the support of β. There exists a constant integer $m_0 = m_0(\vartheta)$ such that with probability $1-o(1)$, G_{SS} is a blockwise diagonal matrix afte a permutation of indices, and the maximum block size is bounded by m_0.*

Proposition 5.1 is a consequence of the interplay between design sparsity and signal sparsity: Under the Rare/Weak signal model (3.2) and a sparse design, the true signals in S appear in groups, where each group contains only a small number of variables and distinct groups are mutually uncorrelated. This motivates us to consider a simpler setting where G is blockwise diagonal *by itself*. While these two settings look so different from each other, the asymptotic behavior of Hamming error is closely related. For example, when G is a tridiagonal matrix with equal values in the sub-diagonal, the optimal phase diagram is the same as in the case

where G is blockwise diagonal with 2×2 blocks (Ji and Jin, 2012; Jin et al., 2014). Inspired by these observations, we study a class of blockwise diagonal designs (Jin and Ke, 2016): For some $\rho\in(-1,1)$ and a $p\times p$ permutation matrix $\mathcal{T}$,

$$G=\mathcal{T}\mathrm{diag}(B,B,\ldots,B,B_1)\mathcal{T},\quad \text{where } B=\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix},\ B_1=\begin{cases}B, & \text{if } p \text{ is even},\\ 1, & \text{if } p \text{ is odd}.\end{cases} \tag{5.1}$$

This design serves the aforementioned purposes (i)-(iii): It has only one parameter ρ, so is mathematically tractable. The nonzero off-diagonal entries of G are at the constant order, so this design is sufficiently different from the orthogonal design (in contrast, many literature consider the independent random Gaussian design, for which the maximum absolute off-diagonal entry of G is only $o(1)$). Also, as we have argued above, studying this design helps us draw useful insights that will likely continue to hold for general sparse designs.

5.1 The prototype, Lasso-path

Before studying SDP-knockoff and CI-knockoff, we first study their prototype, Lasso-path. The next theorem characterizes $\mathrm{FP}_p(u)$ and $\mathrm{FN}_p(u)$ for Lasso-path and is proved in the Appendix.

Theorem 5.1 *Consider a linear regression model where* (3.1)-(3.3) *hold. Suppose $n\geq 2p$ and G is as in* (5.1) *with a correlation parameter $\rho\in(-1,1)$. Let W_j^* be as in* (2.10). *For any constant $u>0$, let $\mathrm{FP}_p(u)$ and $\mathrm{FN}_p(u)$ be the expected numbers of false positives and false negatives, by selecting variables with $W_j^*>\sqrt{2u\log(p)}$. As $p\to\infty$,*

$$\mathrm{FP}_p(u)=L_p p^{1-\min\{u,\ \vartheta+(\sqrt{u}-|\rho|\sqrt{r})^2+(\xi_\rho\sqrt{r}-\eta_\rho\sqrt{u})_+^2-(\sqrt{r}-\sqrt{u})_+^2\}},$$

and

$$\mathrm{FN}_p(u)=\begin{cases}L_p p^{1-\vartheta-\{(\sqrt{r}-\sqrt{u})_+-[(1-\xi_\rho)\sqrt{r}-(1-\eta_\rho)\sqrt{u}]_+\}^2}, & \rho\geq 0,\\ L_p p^{1-\min\{\vartheta+\{(\sqrt{r}-\sqrt{u})_+-[(1-\xi_\rho)\sqrt{r}-(1-\eta_\rho)\sqrt{u}]_+\}^2,\ 2\vartheta+(\xi_\rho\sqrt{r}-\eta_\rho^{-1}\sqrt{u})_+^2\}}, & \rho<0,\end{cases}$$

where $\xi_\rho=\sqrt{1-\rho^2}$ and $\eta_\rho=\sqrt{(1-|\rho|)/(1+|\rho|)}$.

Using Theorem 5.1, we can deduce the FDR-TPR tradeoff diagram and phase diagram. To save space, we only present the phase diagram:

Corollary 5.1 (Phase diagram of Lasso-path) *In the setting of Theorem 5.1, the phase diagram of Lasso-path is given by*

$$h_{\mathrm{AFR}}(\vartheta)=\vartheta,\qquad h_{\mathrm{ER}}(\vartheta)=\begin{cases}\max\{h_1(\vartheta),h_2(\vartheta)\}, & \text{when } \rho\geq 0,\\ \max\{h_1(\vartheta),h_2(\vartheta),h_3(\vartheta),h_4(\vartheta)\}, & \text{when } \rho<0,\end{cases}$$

where $h_1(\vartheta)=(1+\sqrt{1-\vartheta})^2$, $h_2(\vartheta)=\big(1+\sqrt{\frac{1+|\rho|}{1-|\rho|}}\big)^2(1-\vartheta)$, $h_3(\vartheta)=\frac{1}{(1+\rho)^2}\big(\sqrt{\frac{1+\rho}{1-\rho}}\sqrt{1-2\vartheta}+\sqrt{\frac{1-\rho}{1+\rho}}\sqrt{1-\vartheta}\big)^2\cdot 1\{\vartheta<1/2\}$, and $h_4(\vartheta)=\frac{1}{(1+\rho)^2}\big(1+\sqrt{\frac{1+\rho}{1-\rho}}\cdot\sqrt{1-2\vartheta}\big)^2\cdot 1\{\vartheta<1/2\}$.

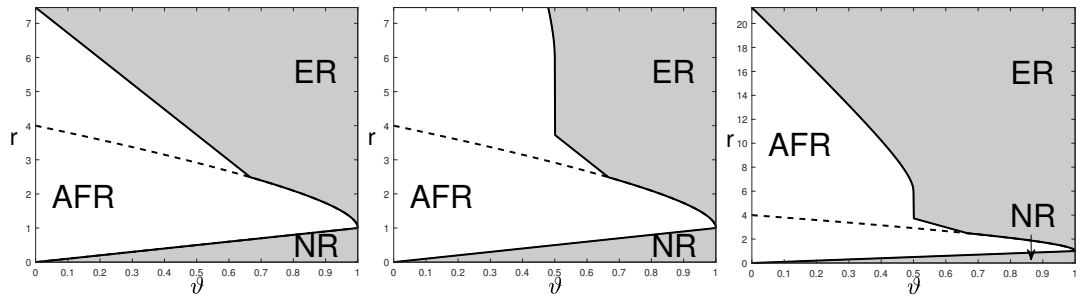

Figure 5: The phase diagrams of Lasso-path (block-wise diagonal designs). Left: $\rho = 0.5$. Middle: $\rho = -0.5$. Right: zoom-out of the middle panel. In all three panels, the dashed lines are the phase curves for orthogonal designs ($\rho = 0$), as a reference.

A visualization of the phase diagram for $\rho = \pm 0.5$ is in Figure 5.

When $\rho = 0$, the blockwise diagonal design reduces to the orthogonal design, Lasso-path reduces to (3.6), and the phase diagram reduces to the one in Proposition 3.1. Comparing Figure 5 with Figure 2, the phase diagram of Lasso-path is inferior to the one for orthogonal designs. This suggests that the strength of design correlations can have a significant impact on the performance of variable selection.

Another observation from Figure 5 is that the sign of ρ plays a crucial role. This is related to the "signal cancellation" phenomenon (Ke et al., 2014; Ke and Yang, 2017). Suppose $\{j, j+1\}$ is a block and both β_j and β_{j+1} are signals. It is seen that $\mathbb{E}[x_j' y|\beta] = (1+\rho)\tau_p$, whose absolute value is strictly smaller than τ_p for a negative ρ. Hence, when ρ is negative, the signal at $j+1$ creates a "cancellation effect" and makes x_j marginally less correlated with y. Lasso is known to be quite vulnerable to "signal cancellation" (Zhao and Yu, 2006). This is why the phase diagram becomes worse when the sign of ρ is flipped. We will provide a more rigorous explanation in Section 7 using geometric properties of the Lasso solution.

5.2 SDP-knockoff

We now study the SDP-knockoff, where $\mathrm{diag}(s)$ is as in (2.6). For the block-wise diagonal design parameterized by ρ, we have an explicit form of $\mathrm{diag}(s)$:

$$\mathrm{diag}(s) = (1-a)I_p, \qquad \text{where} \quad a = \begin{cases} 2|\rho| - 1, & |\rho| \geq 1/2, \\ 0, & |\rho| < 1/2. \end{cases} \tag{5.2}$$

The value of a controls the correlation between x_j and $\tilde{x}_j$. SDP-knockoff aims to minimize this correlation, subject to the eligibility constraints. We first study the case $|\rho| \geq 1/2$.

Theorem 5.2 (The case of $|\rho| \geq 1/2$) *Consider a linear model where* (3.1)-(3.3) *hold. Suppose $n \geq 2p$ and G is as in* (5.1)*, where $|\rho| \geq 1/2$. We construct $\tilde{X}$ in knockoff with* $\mathrm{diag}(s)$ *as in* (5.2)*. Let Z_j, $\tilde{Z}_j$ and W_j be as in* (2.2)-(2.3)*, where f is the signed maximum in* (2.5)*. For any constant $u > 0$, let $\mathrm{FP}_p(u)$ and $\mathrm{FN}_p(u)$ be the expected numbers of false positives and false negatives, by selecting variables with $W_j > \sqrt{2u\log(p)}$. As $p \to \infty$,*

$$\mathrm{FP}_p(u) = L_p p^{1-\min\left\{u,\ \vartheta + (\sqrt{u} - |\rho|\sqrt{r})^2 + (\xi_\rho \sqrt{r} - \eta_\rho \sqrt{u})_+^2 - (\sqrt{r} - \sqrt{u})_+^2\right\}},$$

and for $\rho \geq 1/2$,

$$\mathrm{FN}_p(u) = L_p p^{1-\vartheta-\left\{(\sqrt{r}-\sqrt{u})_+-[(1-\xi_\rho)\sqrt{r}-(1-\eta_\rho)\sqrt{u}]_+-(\lambda_\rho\sqrt{r}-\eta_\rho\sqrt{u})_+\right\}^2},$$

and for $\rho \leq -1/2$,

$$\mathrm{FN}_p(u) = L_p p^{1-\min\left\{\vartheta+\left\{(\sqrt{r}-\sqrt{u})_+-[(1-\xi_\rho)\sqrt{r}-(1-\eta_\rho)\sqrt{u}]_+-(\lambda_\rho\sqrt{r}-\eta_\rho\sqrt{u})_+\right\}^2,\ 2\vartheta\right\}},$$

where $\xi_\rho = \sqrt{1-\rho^2}$, $\eta_\rho = \sqrt{(1-|\rho|)/(1+|\rho|)}$, *and* $\lambda_\rho = \sqrt{1-\rho^2}-\sqrt{1-|\rho|}$.

When $|\rho| < 1/2$, listing the separate forms of $\mathrm{FP}_p(u)$ and $\mathrm{FN}_p(u)$ is tedious. We instead present the form of $\mathrm{FP}_p(u) + \mathrm{FN}_p(u)$, which is sufficient for deriving the phase diagram.

Theorem 5.3 (The case of $|\rho| < 1/2$) *Consider a linear model where* (3.1)-(3.3) *hold. Suppose* $n \geq 2p$ *and* G *is as in* (5.1), *where* $|\rho| < 1/2$. *We construct* $\tilde{X}$ *in knockoff with* $\mathrm{diag}(s)$ *as in* (5.2). *Let* Z_j, $\tilde{Z}_j$ *and* W_j *be as in* (2.2)-(2.3), *where* f *is the signed maximum in* (2.5). *For any constant* $u > 0$, *let* $\mathrm{FP}_p(u)$ *and* $\mathrm{FN}_p(u)$ *be the expected numbers of false positives and false negatives, by selecting variables with* $W_j > \sqrt{2u\log(p)}$. *As* $p \to \infty$,

$$\mathrm{FP}_p(u) + \mathrm{FN}_p(u) = \begin{cases} L_p p^{1-f^+_{Hamm}(u,r,\vartheta)}, & 0 \leq \rho < 1/2, \\ L_p p^{1-\min\left\{f^+_{Hamm}(u,r,\vartheta),\ 2\vartheta+(\xi_\rho\sqrt{r}-\eta_\rho^{-1}\sqrt{u})_+^2,\ 2\vartheta+\frac{(1+2\rho)^2(1-\rho)}{2(1+\rho)}r\right\}}, & -1/2 < \rho < 0, \end{cases}$$

where

$$f^+_{Hamm}(u,r,\vartheta) = \min\big\{u,\ \ \vartheta + (\sqrt{u}-|\rho|\sqrt{r})^2 + ((\xi_\rho\sqrt{r}-\eta_\rho\sqrt{u})_+)^2 - ((\sqrt{r}-\sqrt{u})_+)^2, \\ \vartheta + [(\sqrt{r}-\sqrt{u})_+ - ((1-\xi_\rho)\sqrt{r}-(1-\eta_\rho)\sqrt{u})_+]^2\big\},$$

and ξ_ρ, η_ρ *are the same as those in Theorem 5.2.*

Corollary 5.2 (Phase diagram of SDP-knockoff) *Suppose conditions of Theorems 5.2-5.3 hold, where* ρ *can be any value in* $(-1,1)$. *Let* $h^{\mathrm{lasso}}_{AFR}(\vartheta)$ *and* $h^{\mathrm{lasso}}_{ER}(\vartheta)$ *be the phase curves in Corollary 5.1. Define*

$$\rho_0 = \sqrt{2}-1-\sqrt{2-\sqrt{2}} \qquad \textit{(note: } \rho_0 \approx -0.35\textit{)}.$$

For SDP-knockoff, $h_{AFR}(\vartheta) = h^{\mathrm{lasso}}_{AFR}(\vartheta)$, *and* $h_{ER}(\vartheta)$ *has three cases:*

- *When* $\rho \in [\rho_0, 1)$, $h_{ER}(\vartheta) = h^{\mathrm{lasso}}_{ER}(\vartheta)$.
- *When* $\rho \in (-0.5, \rho_0)$, $h_{ER}(\vartheta) = \max\{h^{\mathrm{lasso}}_{ER}(\vartheta),\ h_5(\vartheta)\}$, *where* $h_5(\vartheta) = \frac{2(1-2\vartheta)(1+\rho)}{(1+2\rho)^2(1-\rho)}$.
- *When* $\rho \in (-1, -0.5]$, $h_{ER}(\vartheta) = h^{\mathrm{lasso}}_{ER}(\vartheta)$ *if* $\vartheta > 1/2$, *and* $h_{ER}(\vartheta) = \infty$ *otherwise.*

A visualization of the phase diagram for three values of ρ is in Figure 6.

Comparing Corollary 5.2 and Corollary 5.1, We have the following observations:

- When $\rho \in [\rho_0, 1)$, SDP-knockoff shares the same phase diagram as Lasso-path, i.e., SDP-knockoff yields a negligible power loss compared with its prototype.
- When $\rho \in (-1, \rho_0)$, the phase diagrams of SDP-knockoff is inferior to that of Lasso-path. Especially, when $\rho \in (-1, -0.5]$, the AFR region of SDP-knockoff is infinite: For any $\vartheta < 1/2$, no matter how large r is, SDP-knockoff never achieves Exact Recovery.

We give an explanation of the discrepancy of the phase diagram between SDP-knockoff and Lasso-path for $\rho \in (-1, \rho_0)$. First, consider $\rho \in (-0.5, \rho_0)$. By (5.2), $a = 0$ and $\mathrm{diag}(s) = I_p$. It follows that the jth knockoff is uncorrelated with the jth original variable. However, this knockoff is still highly correlated with the $(j+1)$th original variable. Suppose j is a true signal variable. Then, a true signal at $(j+1)$ will increase the absolute correlation between y and $\tilde{x}_j$ but decrease the absolute correlation between y and x_j (since $\rho < 0$), making it more difficult for x_j to stand out. Next, consider $\rho \in (-1, -0.5]$. Suppose $\{j, j+1\}$ has two 'nested' signals, i.e., $(\beta_j, \beta_{j+1}) = (\tau_p, \tau_p)$. By (5.2) and an elementary calculation,

$$\mathbb{E}[x_j' y|\beta] = (1+\rho)\tau_p, \qquad \mathbb{E}[\tilde{x}_j' y|\beta] = \begin{cases} \rho\tau_p, & \text{when } -0.5 < \rho < 0, \\ -(1+\rho)\tau_p, & \text{when } -1 < \rho \le -0.5. \end{cases}$$

When $\rho \le -0.5$, variable j and its knockoff have the same absolute correlation with y. Consequently, there is a non-diminishing probability that the true signal variable fails to dominate its knockoff variable, making it impossible to select j consistently. In the Rare/Weak signal model, 'nested' signals appear with a non-diminishing probability if $\vartheta < 1/2$. This explains why $h_{ER}(\vartheta) = \infty$ when $\rho \le -0.5$ and $\vartheta < 1/2$.

The rationale of SDP-knockoff is to minimize the correlation between a variable and its own knockoff, but this is not necessarily the best strategy for constructing knockoff variables when the original variables are highly correlated. In the next subsection, we will see that a proper increase of the correlation between a variable and its knockoff can boost power.

5.3 CI-knockoff

We study the CI-knockoff, where $\mathrm{diag}(s)$ is as in (2.7). Liu and Rigollet (2019) showed that when $c = 1$ in (2.7), the resulting $\tilde{X}$ satisfies $x_j'(I - P_{-j})\tilde{x}_j = 0$, where P_{-j} is the projection matrix to the linear span of $\{x_k : k \ne j\}$. It means x_j and $\tilde{x}_j$ are *conditionally uncorrelated*, conditioning on the other $(p-1)$ original variables. For the block-wise diagonal design (5.1), $\mathrm{diag}(s)$ has an explicit form:

$$\mathrm{diag}(s) = (1 - \rho^2)I_p, \qquad \text{for all } \rho \in (-1, 1). \tag{5.3}$$

Compared with (5.1), the value of a has changed. We recall that a controls the correlation between an original variable and its knockoff. In SDP-knockoff, a is chosen as the minimum eligible value, but in CI-knockoff, a is set at ρ^2.

Theorem 5.4 *Consider a linear model where* (3.1)-(3.3) *hold. Suppose* $n \ge 2p$ *and* G *is as in* (5.1)*, with a correlation parameter* $\rho \in (-1, 1)$*. We construct* $\tilde{X}$ *in knockoff with* $\mathrm{diag}(s)$ *as in* (5.3)*. Let* Z_j*,* $\tilde{Z}_j$ *and* W_j *be as in* (2.2)-(2.3)*, where* f *is the signed maximum*

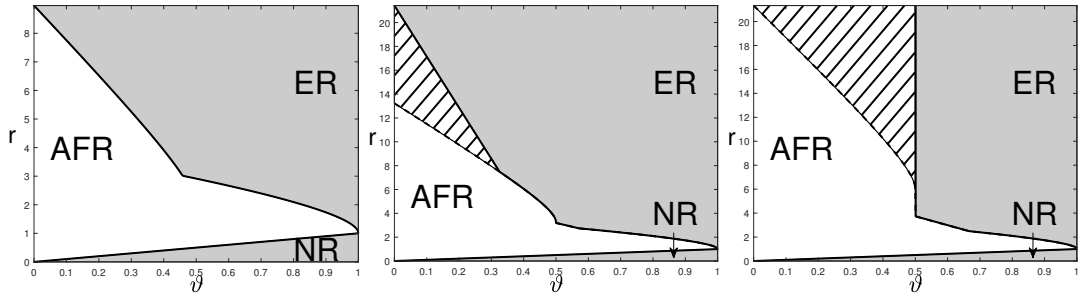

Figure 6: The phase diagrams of SDP-knockoff (blockwise diagonal designs; ranking algorithm is Lasso, and symmetric statistic is signed maximum). From left to right, the correlation parameter in the design is $\rho = -0.3$, $\rho = -0.4$, and $\rho = -0.5$, respectively. They correspond to the three cases in Corollary 5.2. The shadowed area is the Almost Full Recovery region for SDP-knockoff but Exact Recovery region for the prototype Lasso-path. If SDP-knockoff is replaced by CI-knockoff, then in each of three cases the phase diagram is the same as that of Lasso-path.

in (2.5). *For any constant* $u > 0$, *let* $\mathrm{FP}_p(u)$ *and* $\mathrm{FN}_p(u)$ *be the expected numbers of false positives and false negatives, by selecting variables with* $W_j > \sqrt{2u\log(p)}$. *As* $p \to \infty$,

$$\mathrm{FP}_p(u) + \mathrm{FN}_p(u) = \begin{cases} L_p p^{1-f^+_{Hamm}(u,r,\vartheta)}, & \rho \geq 0, \\ L_p p^{1-\min\left\{f^+_{Hamm}(u,r,\vartheta),\ 2\vartheta+(\xi_\rho\sqrt{r}-\eta_\rho^{-1}\sqrt{u})_+^2\right\}}, & \rho < 0, \end{cases}$$

where $f^+_{Hamm}(u,r,\vartheta)$ *is the same as that in Theorem 5.3.*

The exponent in Theorem 5.4 is in fact the same as that in Theorem 5.1. We immediately conclude that CI-knockoff yields the same phase diagram as its prototype, Lasso-path.

Corollary 5.3 (Phase diagram of CI-knockoff) *In the setting of Theorem 5.4, for any* $\rho \in (-1, 1)$, *the phase curves of CI-knockoff are the same as those in Corollary 5.1.*

The result of CI-knockoff is very encouraging. We now explain how CI-knockoff improves SDP-knockoff for $\rho \in (-1, \rho_0)$. Comparing (5.3) with (5.2), we find that the correlation between x_j and $\tilde{x}_j$ increases from $\max\{0, 2|\rho| - 1\}$ to ρ^2. We revisit the scenario of two 'nested' signals, i.e., $(\beta_j, \beta_{j+1}) = (\tau_p, \tau_p)$. By direct calculations,

$$\mathbb{E}[x_j' y|\beta] = (1+\rho)\tau_p, \qquad \mathbb{E}[\tilde{x}_j' y|\beta] = \rho(1+\rho)\tau_p.$$

It always holds that $|\mathbb{E}[x_j' y|\beta]| > |\mathbb{E}[\tilde{x}_j' y|\beta]|$. As long as r is sufficiently large, the original variable x_j can standard out. This resolves the previous issue of SDP-knockoff.

Going beyond the block-wise design, it is an interesting question whether CI-knockoff still improves SDP-knockoff. We study it numerically in Section 8, where we consider designs such as *Factor models*, *Exponential decay*, and *Normalized Wishart*; see Experiment 4.

Remark 5. Our theory is focused on the 2×2 blockwise design in (5.1). Using similar techniques, we can study other blockwise designs, such as $k \times k$ blocks or varying-size blocks.

Take $k \times k$ blocks for example. In knockoff, solving Lasso in (2.2) reduces to solving many $2k$-dimensional problems separately. Let $J = \{j, j+1, \ldots, j+k\text{-}1\}$ be a block. The sufficient statistic for W_j is $\tilde{y} = [X_J, \tilde{X}_J]'y \in \mathbb{R}^{2k}$. By Lemma 7.1, the Hamming error at j depends on the interplay between probability contour of $\tilde{y}$ and geometry of the $2k$-dimensional Lasso problem. We make such analysis for $k = 2$ in the proofs of Theorems 5.2-5.4, which can be extended to a general k.

6 Impact of the ranking algorithm

We consider two options of the ranking algorithm, Lasso and least-squares. As the ranking algorithm changes, the prototype is different. In Section 6.1, we first compare two prototypes. In Section 6.2, we further compare the associated versions of knockoff.

In the orthodox knockoff, ranking algorithm is Lasso, augmented design is SDP-knockoff, and symmetric statistic is signed maximum. We re-name it *SDP-knockoff-Lasso*. If ranking algorithm is changed to least-squares (with the other two components unchanged), we call it *SDP-knockoff-OLS*. In each method, if augmented design is changed to CI-knockoff (with the other two components unchanged), we call them *CI-knockoff-Lasso* and *CI-knockoff-OLS*, respectively. SDP-knockoff-Lasso, CI-knockoff-Lasso, and their prototype, Lasso-path, have been studied in Section 5. In this section, we study SDP-knockoff-OLS, CI-knockoff-OLS, and their prototype, least-squares, and compare the results with those in Section 5.

We consider the general design, where $G = X'X$ can be any positive definite matrix. We then restrict ourselves to the special case of 2×2 blockwise design in (5.1). The reason we can study general designs is that the least-squares solution has a simple and explicit form (but the Lasso solution does not).

6.1 The prototype, least-squares

Before studying SDP-knockoff-OLS and CI-knockoff-OLS, we first study their common prototype, the least-squares (see (2.11)).

Theorem 6.1 *Consider a linear regression model where* (3.1)-(3.3) *hold and* $n \geq 2p$. *Let* ω_j *be the* j*-th diagonal element of* G^{-1}. *Suppose* $\min_{1\leq j\leq p}\{\omega_j\} \leq C_0$, *for a constant* $C_0 > 0$. *Let* W_j^* *be as in* (2.11). *For any constant* $u > 0$, *let* $\mathrm{FP}_p(u)$ *and* $\mathrm{FN}_p(u)$ *be the expected numbers of false positives and false negatives, by selecting variables with* $W_j^* > \sqrt{2u\log(p)}$. *As* $p \to \infty$, *As* $p \to \infty$,

$$\mathrm{FP}_p(u) \leq L_p \sum_{j=1}^{p} p^{-\omega_j^{-1}u}, \qquad \mathrm{FN}_p(u) \leq L_p p^{-\vartheta} \sum_{j=1}^{p} p^{-\omega_j^{-1}(\sqrt{r}-\sqrt{u})_+^2}.$$

Corollary 6.1 (Phase diagram of OLS) *In the setting of Theorem 6.1, suppose* G *is as in* (5.1), *with a correlation parameter* $\rho \in (-1, 1)$. *Then,* $\omega_j = (1-\rho^2)^{-1}$ *for* $1 \leq j \leq p$. *As* $p \to \infty$, $\mathrm{FP}_p(u) = L_p p^{1-(1-\rho^2)u}$, *and* $\mathrm{FN}_p(u) = L_p p^{1-\vartheta-(1-\rho^2)(\sqrt{r}-\sqrt{u})_+^2}$. *The phase diagram of least-squares is given by* $h_{\mathrm{AFR}}(\vartheta) = \frac{\vartheta}{1-\rho^2}$ *and* $h_{\mathrm{ER}}(\vartheta) = \frac{(1+\sqrt{1-\vartheta})^2}{1-\rho^2}$.

Figure 7 (left panel) shows the phase diagram of least-squares for $|\rho| = 0.5$; as a reference, in the right two panels, we plot again the phase diagrams of Lasso-path for $\rho = \pm 0.5$. For the comparison between least-squares and Lasso-path, we have the following observations:

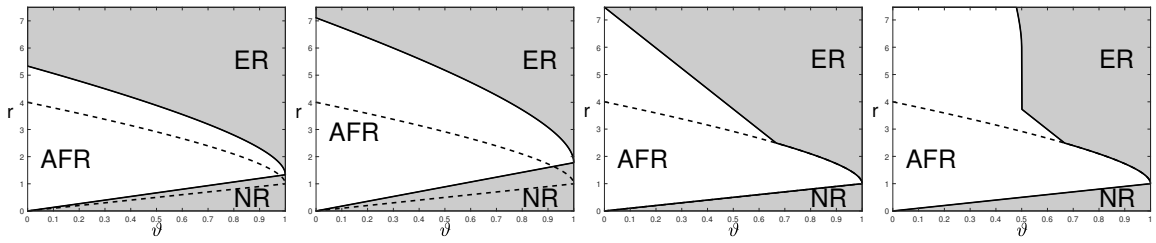

Figure 7: The phase diagrams of least-squares (left) and CI-knockoff-OLS (middle left), for blockwise designs with $|\rho| = 0.5$. For reference, we also plot the phase diagrams of Lasso-path for $\rho = 0.5$ (middle right) and $\rho = -0.5$ (right), which are also the phase diagrams of CI-knockoff-Lasso.

- In terms of $h_{\mathrm{AFR}}(\vartheta)$, Lasso-path is always better than least-squares. To attain Almost Full Recovery, Lasso-path requires $r > \vartheta$, but least-squares requires $r > \vartheta/(1-\rho^2)$.
- In terms of $h_{\mathrm{ER}}(\vartheta)$, Lasso-path is better than least-squares when ϑ is relatively large (i.e., β is comparably sparser), and least-squares is better than Lasso-path when ϑ is relatively small (i.e., β is comparably denser).
- The sign of ρ also matters. For small ϑ, the advantage of least-squares over Lasso-path on $h_{\mathrm{ER}}(\vartheta)$ is much more obvious when ρ is negative.

We give an intuitive explanation to the above phenomena. We say a signal variable (i.e., $\beta_j \neq 0$) is 'isolated' if it is the only signal variable in the 2×2 block, and we say two signals are 'nested' if they are in the same 2×2 block. In the sparser regime (i.e., ϑ is large), least-squares has a disadvantage because it is inefficient in discovering an 'isolated' signal. In the less sparse regime (i.e., ϑ is small), Lasso-path has a disadvantage because it suffers from signal cancellation when estimating a pair of 'nested' signals ('signal cancellation' means a signal variable has a weak marginal correlation with y due to the effect of other signals correlated with this one). A more rigorous explanation is given in Section 7, using geometry of solutions of least-squares and Lasso; see Lemma 7.2, Figure 8, and discussions therein.

6.2 Knockoff-OLS

We now study SDP-knockoff-OLS and CI-knockoff-OLS. The next theorem provides a general result that applies to all augmented designs:

Theorem 6.2 *Consider a linear model where* (3.1)-(3.3) *hold. Suppose* $n \geq 2p$. *We construct* $\tilde{X}$ *in knockoff as in* (2.1)*, with some choice of* $\mathrm{diag}(s)$. *Write* $G^* = [X, \tilde{X}]'[X, \tilde{X}] \in \mathbb{R}^{2p \times 2p}$. *Suppose* $\mathrm{diag}(s)$ *is chosen such that* G^* *is non-singular. Let* $A_j \in \mathbb{R}^{2\times 2}$ *be the submatrix of* $(G^*)^{-1}$ *restricted to the* j*th and* $(j+p)$*th rows and columns. Denote* $\omega_{1j} = A_j(1,1)$ *and* $\omega_{2j} = A_j(1,2)$. *Suppose* $\max_{1\leq j\leq p}\{\omega_{1j}\} \leq C_0$, *for a constanat* $C_0 > 0$. *Let* Z_j, $\tilde{Z}_j$ *and* W_j *be as in* (2.8) *and* (2.3)*, where* f *is the signed maximum in* (2.5). *For any constant* $u > 0$, *let* $\mathrm{FP}_p(u)$ *and* $\mathrm{FN}_p(u)$ *be the expected numbers of false positives and false negatives,*

by selecting variables with $W_j > \sqrt{2u\log(p)}$. *As* $p \to \infty$,

$$\mathrm{FP}_p(u) \le L_p \sum_{j=1}^{p} p^{-\omega_{1j}^{-1}u}, \qquad \mathrm{FN}_p(u) \le L_p p^{-\vartheta} \sum_{j=1}^{p} p^{-\omega_{1j}^{-1}\min\left\{(\sqrt{r}-\sqrt{u})_+^2,\ \frac{\omega_{1j}}{\omega_{1j}+|\omega_{2j}|}\cdot\frac{1}{2}r\right\}}.$$

The phase diagram of knockoff-OLS is governed by the quantities $\{\omega_{1j}\}_{1\le j\le p}$. We now consider the special case of the 2×2 blockwise design in (5.1), where the augmented design is such that $\mathrm{diag}(s) = (1-a)I_p$, where $a = \max\{0, 2|\rho|-1\}$ in SDP-knockoff and $a = \rho^2$ in CI-knockoff. We note that in SDP-knockoff, the matrix G^* is singular when $|\rho| \ge 1/2$. In other words, SDP-knockoff-OLS is well defined only for $|\rho| < 1/2$.

Corollary 6.2 (Phase diagram of knockoff-OLS) *In the same setting of Theorem 6.2, suppose* G *is as in* (5.1)*, with a correlation parameter* $\rho \in (-1, 1)$.

- *SDP-knockoff-OLS (only defined for* $|\rho| < 1/2$*):* $\omega_{1j} = (1-4\rho^2)^{-1}(1-2\rho^2)$ *for* $1 \le j \le p$. *The phase diagram is given by* $h_{\mathrm{AFR}}(\vartheta) = \frac{\vartheta(1-2\rho^2)}{1-4\rho^2}$ *and* $h_{\mathrm{ER}}(\vartheta) = \frac{(1+\sqrt{1-\vartheta})^2(1-2\rho^2)}{1-4\rho^2}$.
- *CI-knockoff-OLS:* $\omega_{1j} = (1-\rho^2)^{-2}$ *for* $1 \le j \le p$. *The phase diagram is given by* $h_{\mathrm{AFR}}(\vartheta) = \frac{\vartheta}{(1-\rho^2)^2}$ *and* $h_{\mathrm{ER}}(\vartheta) = \frac{(1+\sqrt{1-\vartheta})^2}{(1-\rho^2)^2}$.

Figure 7 (second left panel) shows the phase diagram of CI-knockoff-OLS for $|\rho| = 0.5$. In this figure, the right two panels are the phase diagrams of CI-knockoff-Lasso for $\rho = \pm 0.5$.

From Corollary 6.2 and Figure 7, we draw two conclusions: First, for both SDP-knockoff-OLS and and CI-knockoff-OLS, whenever $\rho \ne 0$, their phase diagrams are strictly inferior to the phase diagram of the least-squares (prototype). This is different from the case of using Lasso as ranking algorithm, where the phase diagrams of CI-knockoff-Lasso and Lasso-path (prototype) are the same in the blockwise design for all $\rho \in (-1, 1)$. Second, the comparison of CI-knockoff-OLS and CI-knockoff-Lasso is largely similar to the comparison between the least-squares and Lasso-path (see Section 6.1).

Remark 6. When we use the least-squares as the ranking algorithm, such a gap between knockoff and its prototype always exists, for a general design. To see this, note that by Theorem 6.1 and Theorem 6.2, the phase diagrams of knockoff-OLS and its prototype are governed by the quantities $\{\omega_{1j}\}_{1\le j\le p}$ and $\{\omega_j\}_{1\le j\le p}$, respectively. Since ω_{1j} and ω_j are the jth diagonal elements of G^{-1} and $(G^*)^{-1}$, respectively, and G is a principal submatrix of G^*, it follows by elementary linear algebra that $\omega_j \le \omega_{1j}$ is always true (and this inequality is often strict). Unfortunately, it is impossible to mitigate this gap by using the augmented design in (2.1), no matter how we choose $\mathrm{diag}(s)$. Xing et al. (2023) proposed a new idea of constructing an augmented design, called the Gaussian mirror, which is tailored to using the least-squares as the ranking algorithm. In a companion paper (Ke et al., 2022), we show that the Gaussian mirror attains the same phase diagram as the least-squares.

Remark 7. Besides Lasso and least-squares, we may consider other ranking algorithms, such as the thresholded Lasso, non-convex penalization methods, and the forward-backward selection. See Ke and Wang (2021) about the phase diagrams of these methods.

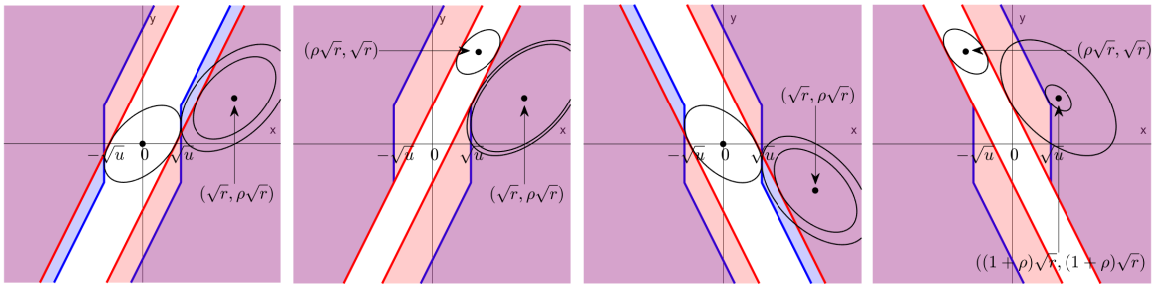

Figure 8: Rejection regions and 'most-likely' cases in block-wise diagonal designs (x-axis: $x_j'y/\sqrt{2\log(p)}$; y-axis: $x_{j+1}'y/\sqrt{2\log(p)}$). From left to right: (i) positive ρ and large ϑ, (ii) positive ρ and small ϑ, (iii) negative ρ and large ϑ, (iv) negative ρ and small ϑ. In each plot, the blue solid lines define rejection region of Lasso-path, and the red solid lines define rejection region of least-squares. For each method, FP_p is determined by the largest FP-ellipsoid in $\mathcal{R}^c$, and FN_p is determined by the largest FN-ellipsoid in $\mathcal{R}$, where the centers of these ellipsoids are determined by (β_j, β_{j+1}) in the 'most-likely' case. In each plot, the largest FP-ellipsoid is controlled to be the same for both Lasso-path and least-squares, and so the method with a larger FN-ellipsoid is better.

7 The proof ideas and some geometric insights

A key technical tool in the proof is the following lemma, which is proved in the Appendix. Recall that L_p is a generic notation of multi-$\log(p)$ terms; see Definition 3.1. For a vector v, $\|v\|$ denotes the ℓ^2-norm; for a matrix M, $\|M\|$ denotes the spectral norm.

Lemma 7.1 *Fix an integer $d \geq 1$, a vector $\mu \in \mathbb{R}^d$, a covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$, and an open set $S \subset \mathbb{R}^d$ such that $\mu \notin S$. The quantities (d, μ, Σ, S) do not change with p. Suppose $b \equiv \inf_{x\in S}\{(x-\mu)'\Sigma^{-1}(x-\mu)\} < \infty$. Consider a sequence of random vectors $X_p \in \mathbb{R}^d$, indexed by p, satisfying that*

$$X_p|(\mu_p, \Sigma_p) \sim \mathcal{N}_d\Big(\mu_p, \ \frac{1}{2\log(p)}\Sigma_p\Big),$$

where $\mu_p \in \mathbb{R}^d$ is a random vector and $\Sigma_p \in \mathbb{R}^{d\times d}$ is a random covariance matrix. As $p \to \infty$, suppose for any fixed $\gamma > 0$ and $L > 0$, $\mathbb{P}(\|\mu_p - \mu\| > \gamma) \leq p^{-L}$ and $\mathbb{P}(\|\Sigma_p - \Sigma\| > \gamma) \leq p^{-L}$. Then, as $p \to \infty$,

$$\mathbb{P}(X_p \in S) = L_p p^{-b},$$

or equivalently, $p^{b+\delta}\mathbb{P}(X_p \in S) \to \infty$ and $p^{b+\delta}\mathbb{P}(X_p \in S) \to 0$ for any constant $\delta > 0$.

This lemma connects the rate of convergence of $\mathbb{P}(X_p \in S)$ with the geometric property of the set S. The exponent b is the "radius" of the largest ellipsoid that centers at μ and is fully contained in the complement of S.

Proof sketch. We illustrate how to use Lemma 7.1 to prove the theorems in Sections 4-6. Take the proof of Theorem 5.1 for example. Consider the block-wise design in (5.1). Under this design, the objective of Lasso is separable, and it reduces to solving many 2-dimensional Lasso problems separably. Fix j and suppose $\{j, j+1\}$ is a block. Let W_j^* be as in (2.10). Write

$$\hat{h} = \big(x_j'y,\ x_{j+1}'y\big)'/\sqrt{2\log(p)} \ \in \ \mathbb{R}^2. \tag{7.1}$$

Since the Lasso objective is separable, (W_j^*, W_{j+1}^*) are purely determined by $\hat{h}$. Particularly, there exists $\mathcal{R}_u \subset \mathbb{R}^2$, such that $W_j^* > \sqrt{2u\log(p)}$ if and only if $\hat{h} \in \mathcal{R}_u$. We call $\mathcal{R}_u$ the "rejection region" of Lasso-path. The probabilities of a false positive and a false negative occurring at j are respectively

$$\mathbb{P}\big(\hat{h} \in \mathcal{R}_u,\ \beta_j = 0\big) \qquad \text{and} \qquad \mathbb{P}\big(\hat{h} \in \mathcal{R}_u^c,\ \beta_j = \tau_p\big).$$

Conditioning on β, the random vector $\hat{h}$ has a bivariate normal distribution, whose mean is a constant vector and whose covariance matrix is $\frac{1}{2\log(p)}B$, where B is the same as in (5.1). Applying Lemma 7.1, we reduce the proof into two steps: In Step 1, we derive the rejection region $\mathcal{R}_u$. In Step 2, for each possible realization of β with $\beta_j = 0$, we calculate $b(\beta) \equiv \inf_{x\in\mathcal{R}_u}\{(x-\mu(\beta))'B^{-1}(x-\mu(\beta))\}$, and for each possible realization of β with $\beta_j \neq 0$, we calculate $b(\beta) \equiv \inf_{x\in\mathcal{R}_u^c}\{(x-\mu(\beta))'B^{-1}(x-\mu(\beta))\}$, where $\mu(\beta) \equiv \mathbb{E}[\hat{h}|\beta]$. Both steps can be carried out by direct calculations. We use a similar strategy to prove other theorems. The proof is sometimes complicated. For example, to analyze knockoff for block-wise diagonal designs, we have to consider the random vector $\hat{h} = (x_j'y, x_{j+1}'y, \tilde{x}_j'y, \tilde{x}_{j+1}'y)'/\sqrt{2\log(p)} \in \mathbb{R}^4$. The proof requires deriving a 4-dimensional rejection region and calculating $b(\beta)$, for an arbitrary $\rho \in (-1, 1)$. The calculations are very tedious.

The geometric insight about two prototypes. We use the geometric interpretation of our proofs to give more insights about Lasso-path versus least-squares (see Corollary 5.1 and Corollary 6.1). Under the blockwise design (5.1), for each method, the objective is separable, so that the event $W_j^* > \sqrt{2u\log(p)}$ can be described via a 2-dimensional rejection region. The next lemma gives the rejection regions of Lasso-path and least-squares:

Lemma 7.2 *Consider a linear model, where the Gram matrix satisfies* (5.1)*, with a correlation parameter $\rho \in (-1, 1)$. Let $W_j^{*,path}$ and $W_j^{*,old}$ be as in* (2.10) *and* (2.11)*, respectively. Suppose $\{j, j+1\}$ is a block. Write $\hat{h} = (x_j'y, x_{j+1}'y)'/\sqrt{2\log(p)}$. Define*

$$\begin{aligned}
\mathcal{R}_u^{\text{path}}(\rho) &= \{(h_1,h_2) : h_1 - \rho h_2 > (1-\rho)\sqrt{u},\ h_1 > \sqrt{u}\} \\
&\quad \cup \{(h_1,h_2) : h_1 - \rho h_2 > (1+\rho)\sqrt{u}\} \\
&\quad \cup \{(h_1,h_2) : h_1 - \rho h_2 < -(1-\rho)\sqrt{u},\ h_1 < -\sqrt{u}\} \\
&\quad \cup \{(h_1,h_2) : x - \rho y < -(1+\rho)u\}, \quad \textit{for } \rho \geq 0, \\
\mathcal{R}_u^{\text{path}}(\rho) &= \{(h_1,h_2) : (h_1,-h_2) \in \mathcal{R}_u^{\text{path}}(-\rho)\}, \qquad \textit{for } \rho < 0, \\
\mathcal{R}_u^{\text{ols}}(\rho) &= \{(h_1,h_2) : h_1 - \rho h_2 > (1-\rho^2)\sqrt{u}\} \\
&\quad \cup \{(h_1,h_2) : h_1 - \rho h_2 < -(1-\rho^2)\sqrt{u}\}.
\end{aligned}$$

Then, for Lasso-path, $W_j^{*,path} > \sqrt{2u\log(p)}$ if and only if $\hat{h} \in \mathcal{R}_u^{\text{path}}(\rho)$; for least-squares, $W_j^{*,ols} > \sqrt{2u\log(p)}$ if and only if $\hat{h} \in \mathcal{R}_u^{\text{ols}}(\rho)$.

These rejection regions are shown in Figure 8. Their geometric properties are different for positive and negative ρ. Fix j. Let $\hat{h}$ be as in (7.1), and write $\mu(\beta) = \mathbb{E}[\hat{h}|\beta]$.

- The rate of convergence of $\mathrm{FP}_p(u)$ is determined by the largest ellipsoid that centers at $\mu(\beta)$ and is contained in $\mathcal{R}_u^c$. We call this ellipsoid the *FP-ellipsoid*.
- The rate of convergence of $\mathrm{FN}_p(u)$ is determined by the largest ellipsoid that centers at $\mu(\beta)$ and is contained in $\mathcal{R}_u$. We call this ellipsoid the *FN-ellipsoid*.

By direct calculations, $\mu(\beta) = \big(\beta_j + \rho\beta_{j+1},\ \rho\beta_j + \beta_{j+1}\big)'/\sqrt{2\log(p)}$. Under our model, (β_j, β_{j+1}) has 4 possible values $\{(0,0), (0,\tau_p), (\tau_p,0), (\tau_p,\tau_p)\}$, where the first two correspond to a null at j and the last two correspond to a non-null at j. The probability of having a selection error at j thus splits into 4 terms, and which term is dominating depends on the values of ϑ and ρ. The realization of (β_j, β_{j+1}) that plays a dominating role is called the 'most-likely' case. For example, when ϑ is large (i.e., β is sparser), the most-likely case of a false positive occuring at j is when $(\beta_j, \beta_{j+1}) = (0,0)$; when ϑ is small (i.e., β is less sparse), the most-likely case of a false positive is when $(\beta_j, \beta_{j+1}) = (0, \tau_p)$. Table 1 summarizes the 'most-likely' cases. We also visualize the 'most-likely' cases for different (ρ, ϑ) in Figure 8. In each plot of Figure 8, we have coordinated the thresholds u in two methods so that the FP-ellipsoid is exactly the same. It suffices to compare the FN-ellipsoid: The method with a larger FN-ellipsoid has a faster rate of convergence on the Hamming error. It is clear that, when ϑ is large, the FN-ellipsoid of Lasso-path is larger; when ϑ is small, the FN-ellipsoid of least-squares is larger. This explains the different performances of two methods. Moreover, when ϑ is small, comparing the case of a positive ρ with the case of a negative ρ, we find that the difference between FN-ellipsoids of two methods are much more prominent in the case of a negative ρ. This explains why the sign of ρ matters.

Sparsity	Correlation	Error type	Most-likely case	Center of ellipsoid
large ϑ	positive/negative ρ	FP	$\beta_j = 0,\ \beta_{j+1} = 0$	$(0,\ 0)$
		FN	$\beta_j = \tau_p,\ \beta_{j+1} = 0$	$(\sqrt{r},\ \rho\sqrt{r})$
small ϑ	positive ρ	FP	$\beta_j = 0,\ \beta_{j+1} = \tau_p$	$(\rho\sqrt{r},\ \sqrt{r})$
		FN	$\beta_j = \tau_p,\ \beta_{j+1} = 0$	$(\sqrt{r},\ \rho\sqrt{r})$
small ϑ	negative ρ	FP	$\beta_j = 0,\ \beta_{j+1} = \tau_p$	$(\rho\sqrt{r},\ \sqrt{r})$
		FN	$\beta_j = \tau_p,\ \beta_{j+1} = \tau_p$	$\big((1+\rho)\sqrt{r},\ (1+\rho)\sqrt{r}\big)$

Table 1: The 'most-likely' cases and the corresponding ellipsoid center $\mu(\beta)$

8 Simulations

We use numerical experiments to support and exemplify the theoretical results in Sections 4-6. In Experiments 1 and 2, we consider orthogonal designs and block-wise diagonal designs, respectively. In Experiments 3 and 4, we consider other design classes, including block-wise diagonal designs with larger blocks, factor models, exponentially decaying designs, and normalized Wishart designs. We consider four methods, Lasso-path (Lasso), least-squares (OLS), knockoff with Lasso-path ranking (KF.Lasso) and with least-squares ranking (KF.OLS). We use either the signed maximum or the difference as the symmetric statistic,

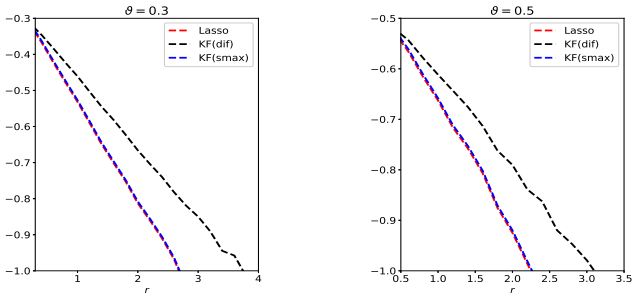

Figure 9: Experiment 1 (orthogonal designs). The y-axis is $\log_p(H_p/p)$, where H_p is the average Hamming error over 200 repetitions.

and for KF we choose $\text{diag}(s) = \min\{1,\, 2\lambda_{\min}(G)\}\cdot I_p$, unless specified otherwise. It is called the equi-correlated knockoff (EC-KF), and is the same as the SDP-knockoff for orthogonal designs and the 2×2 block-wise diagonal designs. In Experiments 1-3, this is the only $\text{diag}(s)$ we use, and so we write EC-KF as KF for short. In Experiments 4, we also consider the conditional independence knockoff (CI-KF). For most experiments, fixing a parameter setting, we generate 200 data sets and record the averaged Hamming selection error among these 200 repetitions.

Experiment 1. We investigate the performance of different methods for orthogonal designs. Given $(n, p) = (2000, 1000)$, $\vartheta \in \{0.3, 0.5\}$ and r ranging on a grid from 0 to 6 with step size 0.2, we generate data y from $N(X\beta, I_n)$ where X is an $n \times p$ matrix with unit length columns that are orthogonal to each other and β is generated from (3.2). We implemented Lasso and KF.Lasso using both the signed maximum and the difference as the symmetric statistic. Under the orthogonal design, Lasso and OLS yield the same importance metric thus OLS and KF.OLS are neglected in this experiment. Each method outputs p importance statistics, and we threshold these importance statistics at $\sqrt{2u^* \log(p)}$ where u^* minimizes $\text{FN}_p(u) + \text{FN}_p(u)$ in theory. The results are in Figure 9, where the y-axis is $\log_p(H_p/p)$, and H_p is the averaged Hamming selection error over 200 repetitions.

The theory in Sections 4-5 suggests the following for orthogonal designs: (i) Regarding the choice of symmetric statistic for KF, the signed maximum outperforms the difference. (ii) With signed maximum as the symmetric statistic, KF.Lasso has a similar performance as Lasso. These theoretical results are perfectly validated by simulations (see Figure 9).

Experiment 2. We consider the block-wise diagonal design with 2×2 blocks, where we take $\rho = 0.5$ and $\rho = 0.7$. In the data generation, we fix an $n \times p$ matrix X such that $X'X$ has the desirable form. We then generate (β, y) in the same way as before. For each ρ, we fix $(n, p, \vartheta) = (2000, 1000, 0.2)$, and let r range on a grid from 0 to 8 with a step size 0.2. For KF.Lasso and KF.OLS, we now fix the symmetric statistic as signed maximum and the default choice of $\text{diag}(s)$ yields that $\text{diag}(s) = (1 - a)I_p$ with $a = 2\rho - 1$. In this case, $G^* = [X, \tilde{X}]'[X, \tilde{X}]$ is degenerated, thus an $\epsilon = 10^{-5}$ was subtracted from each elements of $\text{diag}(s)$ to ensure KF.OLS is applicable. The results are in the first two panels of Figure 10.

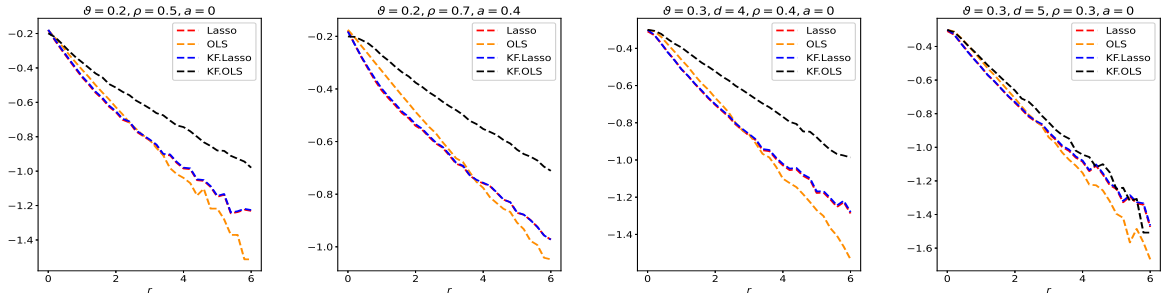

Figure 10: Experiments 2 and 3 (block-wise diagonal designs, d: block size, ρ: off-diagonal entries). The y-axis is $\log_p(H_p/p)$, where H_p is the average Hamming error. The parameter a controls the construction of knockoff.

The theory in Section 5 suggests that since the two values of ρ considered here are in $(\rho_0, 1)$, KF.Lasso has a similar performance as its prototype, Lasso. While according to Section 6, KF.OLS has a inferior performance comparing to its prototype, OLS. The simulation results are consistent with these theoretical predictions. Moreover, we can see that, for the current ϑ value, OLS has a smaller Hamming error than that of Lasso when r is large, and the opposite is true when r is small. These also agree with our theory.

Experiment 3. We further consider blockwise diagonal designs with larger-size blocks. Given $d \geq 2$ and p that is a multiple of d, we generate $X \in \mathbb{R}^{n\times p}$ such that $X'X$ is block-wise diagonal with $d \times d$ diagonal blocks, where the off-diagonal elements of each block are all equal to ρ. Other steps of the data generation are the same as in Experiment 2. We consider $(d, \rho) = (4, 0.4)$ and $(d, \rho) = (5, 0.3)$. For each choice of (d, ρ), we set $(n, p, \vartheta) = (2000, 1000, 0.3)$ and let r range on a grid from 0 to 6 with a step size 0.2. We use signed maximum as symmetric statistic in KF and use the equi-correlated knockoff described above. The results are in the last two panels of Figure 10.

One noteworthy observation is that KF.Lasso still has a similar performance as its own prototype. Meanwhile, KF.OLS can get close to its prototype in the case where ρ is close to 0. Another observation is that OLS outperforms Lasso when r is large, and Lasso slightly outperforms OLS when r is small. While our theory is only derived for $d = 2$, the simulations suggest that similar insight continues to apply when the block size gets larger.

Experiment 4. In Section 5, we studied variants of knockoff with different augmented designs. The theory for 2×2 block-wise designs suggests that using CI-knockoff to construct $\tilde{X}$ yields a higher power than using EC-knockoff (for 2×2 block-wise design, EC-knockoff is the same as SDP-knockoff). In this experiment, we investigate whether using CI-knockoff still yields a power boost for other design classes. We consider 4 types of designs:

- *Factor models*: $X'X = (BB' + I_p)/2$, where B is a $p \times 2$ matrix whose j-th row is equal to $[\cos(\alpha_j), \sin(\alpha_j)]$ with $\{\alpha_j\}_{j=1,\cdots,p}$ *iid* drawn from Uniform$[0, 2\pi]$;
- *Block diagonal*: Same as in Experiment 2, where $\rho = 0.5$.
- *Exponential decay*: The (i, j)-th element of $X'X$ is $0.6^{|i-j|}$, for $1 \leq i, j \leq p$.

- *Normalized Wishart*: $X'X$ is the sample correlation matrix of n *iid* samples of $N(0, I_p)$.

In the normalized Wishart design, the CI-knockoff in (2.7) may not satisfy $\mathrm{diag}(s) \preceq 2G$. We modify it to $\mathrm{diag}(s) = \alpha[\mathrm{diag}(G^{-1})]^{-1}$, where α is the maximum value in $[0, 1]$ such that $\mathrm{diag}(s) \preceq 2G$. For each design, we fix $(n, p) = (1000, 300)$, let ϑ take values in $\{0.2, 0.4\}$ and let r range on a grid from 0 to 6 with a step size 0.2. Different from previous experiments, we generate β from $\beta_j \overset{iid}{\sim} (1-\epsilon_p)\nu_0 + \frac{1}{2}\epsilon_p\nu_{\tau_p} + \frac{1}{2}\epsilon_p\nu_{-\tau_p}$, for $1 \le j \le p$. The motivation of using this model is to allow for negative entries in β. Even when $X'X$ contains only nonnegative elements, this signal model can still reveal the effect of having negative correlations in the design. We compare two versions of knockoff, EC-knockoff and CI-knockoff, along with the prototype, Lasso. The results are in Figure 11.

For the 2×2 block-wise diagonal design, the simulations suggest that CI-KF significantly outperforms EC-KF, and that CI-KF has a similar performance as the prototype, Lasso. This is consistent with the theory in Section 5.2 and Section 5.3. CI-KF also yields a significant improvement over EC-KF in the factor design, and the two methods perform similarly in the exponentially decaying design and the normalized Wishart design. We notice that the Gram matrix of the normalized Wishart design has uniformly small off-diagonal entries for the current (n, p), which is similar to the orthogonal design and explains why EC-KF and CI-KF do not have much difference. Combining these simulation results, we recommend CI-KF for practical use. Additionally, in some settings (e.g., factor design, $\vartheta = 0.4$; exponentially decaying design, $\vartheta = 0.2$), CI-KF even outperforms its prototype Lasso. One possible reason is that the ideal threshold we use is derived by ignoring the multi-$\log(p)$ term, but this term can have a non-negligible effect for a moderately large p, so the Hamming error of Lasso presented here may be larger than the actual optimal one.

9 Discussions

How to maximize the power when controlling FDR at a targeted level is a problem of great interest. We focus on the FDR control method, knockoff, and point out that it has three key components: ranking algorithm, augmented design, and symmetric statistic. Since each component admits multiple choices, knockoff has many different variants. All the variants guarantee finite-sample FDR control. Our goal is to understand which variants enjoy good power. In a Rare/Weak signal model, for each variant of knockoff under consideration, we derive explicit forms of false positive rate and false negative rate, and obtain the theoretical phase diagram. The results provide useful guidelines of choosing the version of knockoff to use in practice. We also define the prototype of knockoff, which uses only one component, ranking algorithm, and has access to an ideal threshold. We compare the phase diagram of knockoff with the phase diagram of prototype. The results help us understand the extra price we pay for finding a data-driven threshold to control FDR.

We have several notable discoveries: (i) For the choice of symmetric statistic, signed maximum is better than difference, because the latter has an inferior phase diagram in the orthogonal design. (ii) For the choice of augmented design, CI-knockoff is better than SDP-knockoff, because the latter has an inferior phase diagram in a simple blockwise diagonal design. (ii) For the choice of ranking algorithm, roughly, Lasso is better than least-squares when the signals are extremely sparse and the design correlations are moderate; and least-squares is better than Lasso when the signals are only moderately sparse and the design

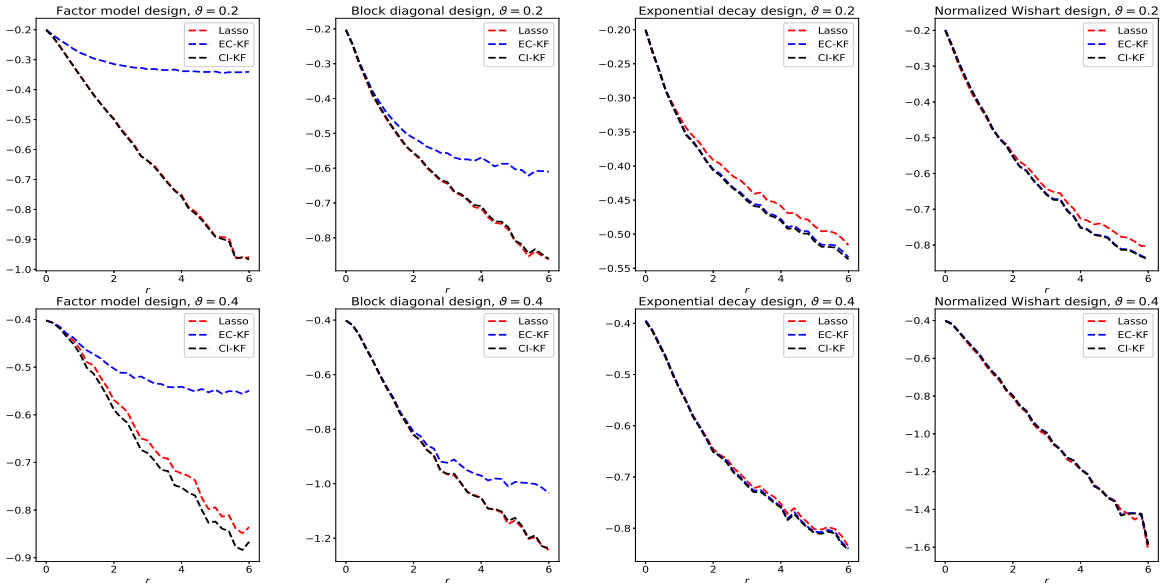

Figure 11: Experiment 4 (general designs). The y-axis is $\log_p(H_p/p)$, where H_p is the average Hamming error. We focus on comparing two constructions of knockoff's, EC-KF and CI-KF, and include Lasso as the benchmark.

correlations are more severe. (iv) In a simple blockwise diagonal design, when knockoff uses Lasso as ranking algorithm, with proper choices of two other components, knockoff has the same phase diagram as its prototype (i.e., we pay a negligible price for finding a data-driven threshold). This is however not true when knockoff uses least-squares as ranking algorithm.

There are several directions to extend our current results. First, we focus on the regime where FDR and TPR converge to either 0 or 1 and characterize the rates of convergence. The more subtle regime where FDR and TPR converge to constants between 0 and 1 is not studied. We leave it to future work. Second, the study of knockoff here is only for block-wise diagonal designs. For general designs, it is very tedious to derive the precise phase diagram, but some cruder results may be less tedious to derive, such as an upper bound for the Hamming error. This kind of results will help shed more insights on how to construct the knockoff variables (e.g., how to choose diag(s)). Third, we only investigate Lasso-path or the least-squares as options of the ranking algorithm. It is interesting to study the power of FDR control methods based on other ranking algorithms, such as the marginal screening and iterative sure screening (Fan and Lv, 2008) and the covariance assisted screening (Ke et al., 2014; Ke and Yang, 2017). The covariance assisted screening was shown to yield optimal phase diagrams for a broad class of sparse designs; whether it can be developed into an FDR control method with "optimal" power remains unknown and is worth future study. Last, some FDR control methods may not fit exactly the unified framework here. For instance, the multiple data splits (Dai et al., 2022) is a method that controls FDR through data splitting. We can similarly assess its power using the Rare/Weak signal model and

phase diagram, except that we need to assume the rows of X are $i.i.d.$ generated. We leave such study to future work.

Acknowledgments and Disclosure of Funding

The research of Jun S. Liu was partially supported by the NSF grant DMS-201541 and the NIH R01grant HG011485-01. The research of Zheng T. Ke was partially supported by the NSF CAREER grant DMS-1943902.

Appendix A. Proof of Lemma 7.1

By definition of the multi-log(p) term, it suffices to show that, for every $\epsilon > 0$, as $p \to \infty$,

$$p^{-\epsilon+b}\ \mathbb{P}(X_p \in S) \to 0, \qquad \text{and} \qquad p^{\epsilon+b}\ \mathbb{P}(X_p \in S) \to \infty. \tag{A.1}$$

We introduce two sets $\underline{S}$ and $\overline{S}$ such that

$$\underline{S} \subset S \subset \overline{S}.$$

Define $m(x) = (x-\mu)'\Sigma^{-1}(x-\mu)$ for any $x \in \mathbb{R}^d$. By definition, $b = \inf_{x\in S} m(x)$. As a result, $m(x) \geq b$ for all $x \in S$. Define

$$\overline{S} = \{x \in \mathbb{R}^p : m(x) \geq b\}. \tag{A.2}$$

Then, $S \subset \overline{S}$. Furthermore, since $m(x)$ is a quadratic function and $b = \inf_{x\in S} m(x)$, given any $\epsilon > 0$, there exists $x_0 \in S$ such that

$$m(x_0) \leq b + \epsilon/8. \tag{A.3}$$

Note that (A.3) guarantees that $\|x_0 - \mu\|$ is bounded. For any $x \in S$ and $\|x - x_0\| \leq 1$,

$$\begin{aligned} |m(x) - m(x_0)| &\leq 2|(x-\mu)'\Sigma^{-1}(x-x_0)| + |(x-x_0)'\Sigma^{-1}(x-x_0)| \\ &\leq 2\|x-\mu\|\|\Sigma^{-1}\| \cdot \|x-x_0\| + \|\Sigma^{-1}\|\|x-x_0\|^2 \\ &\leq C_1\|x-x_0\| + C_2\|x-x_0\|^2, \end{aligned}$$

where C_1 and C_2 are positive constants that only depend on $(\mu, \Sigma, b, \epsilon)$. It follows that there exists a constant $\delta_1 > 0$ such that

$$x \in S, \quad \|x - x_0\| \leq \delta_1 \qquad \Longrightarrow \qquad |m(x) - m(x_0)| \leq \epsilon/8. \tag{A.4}$$

Additionally, since S is an open set and $x_0 \in S$, there exists $\delta_2 > 0$, such that

$$\{x \in \mathbb{R}^d : \|x - x_0\| \leq \delta_2\} \subset S.$$

Define

$$\underline{S} = \{x \in \mathbb{R}^d : \|x - x_0\| \leq \delta\}, \qquad \text{where} \quad \delta = \min\{\delta_1, \delta_2\}. \tag{A.5}$$

It is easy to see that $\underline{S} \subset S$. Additionally, in light of (A.3) and (A.4),

$$m(x) \leq b + \epsilon/4, \qquad \text{for all } x \in \underline{S}. \tag{A.6}$$

Since $\underline{S} \subset S \subset \overline{S}$, to show (A.1), it suffices to show that

$$p^{\epsilon+b}\ \mathbb{P}\big(X_p \in \underline{S}\big) \to \infty \tag{A.7}$$

and

$$p^{-\epsilon+b}\ \mathbb{P}\big(X_p \in \overline{S}\big) \to 0. \tag{A.8}$$

First, we show (A.7). Let $f_p(x)$ denote the density of $\mathcal{N}_d\big(\mu_p,\ \frac{1}{2\log(p)}\Sigma_p\big)$. Write $m_p(x) = (x-\mu_p)'\Sigma_p^{-1}(x-\mu_p)$. It is seen that

$$f_p(x) = \frac{[2\log(p)]^{d/2}}{(2\pi)^{d/2}|\det(\Sigma_d)|^{1/2}} \cdot p^{-m_p(x)}. \tag{A.9}$$

By direct calculations,

$$\begin{aligned}\mathbb{P}\big(X_p \in \underline{S} \mid \mu_p, \Sigma_p\big) &= \frac{[2\log(p)]^{d/2}}{(2\pi)^{d/2}|\det(\Sigma_p)|^{1/2}} \int_{x\in\underline{S}} p^{-m_p(x)}dx \\ &\geq \frac{[\log(p)]^{d/2}}{\pi^{d/2}|\det(\Sigma_p)|^{1/2}} \cdot \text{Volume}(\underline{S}) \cdot p^{-\sup_{x\in\underline{S}}\{m_p(x)\}}. \end{aligned}\tag{A.10}$$

The assumptions on (μ_p, Σ_p) imply that, for any constant $\gamma > 0$,

$$\lim_{p\to\infty} \mathbb{P}\big(\|\mu_p - \mu\| > \gamma \text{ or } \|\Sigma_p - \Sigma\| > \gamma\big) = 0.$$

Let E be the event that $\|\mu_p - \mu\| \leq \gamma_*$ and $\|\Sigma_p - \Sigma\| \leq \gamma_*$, for some γ_* to be decided. On this event, for any $x \in \underline{S}$,

$$\begin{aligned}|m(x) - m_p(x)| &\leq |(x-\mu)'\Sigma^{-1}(x-\mu) - (x-\mu)'\Sigma_p^{-1}(x-\mu)| \\ &\quad + |(x-\mu)'\Sigma_p^{-1}(x-\mu) - (x-\mu_p)'\Sigma_p^{-1}(x-\mu_p)| \\ &\leq |(x-\mu)'(\Sigma^{-1} - \Sigma_p^{-1})(x-\mu)| + 2|(x-\mu)'\Sigma_p^{-1}(\mu-\mu_p)| \\ &\quad + (\mu-\mu_p)'\Sigma_p^{-1}(\mu-\mu_p) \\ &\leq \|x-\mu\|^2\|\Sigma^{-1}\|\|\Sigma_p^{-1}\| \cdot \|\Sigma_p - \Sigma\| + 2\|x-\mu\|\|\Sigma_p^{-1}\| \cdot \|\mu-\mu_p\| \\ &\quad + \|\Sigma_p^{-1}\| \cdot \|\mu-\mu_p\|^2 \\ &\leq C_3\gamma_* + C_4\gamma_*^2,\end{aligned}$$

where C_3 and C_4 are positive constants that do not depend on γ_*, and in the last line we have used the fact that $\underline{S}$ is a bounded set so that $\|x-\mu\|$ is bounded. It follows that we can choose an appropriately small γ_* such that

$$|m(x) - m_p(x)| \leq \epsilon/4, \qquad \text{for all } x \in \underline{S}. \tag{A.11}$$

Combining (A.11) with (A.6) gives

$$\sup_{x\in\underline{S}} m_p(x) \leq b + \epsilon/2, \qquad \text{on the event } E.$$

Moreover, since $\underline{S}$ is a ball with radius δ,

$$\text{Volume}(\underline{S}) = \delta^d \cdot \text{Volume}(B_d),$$

where B_d is the unit ball in $\mathbb{R}^d$, whose volume is a constant. We plug the above results into (A.10) and notice that $|\det(\Sigma_p)| \geq |\det(\Sigma)| - C_5\delta$ on the event E, for a constant $C_5 > 0$. It yields that, when (μ_p, Σ_p) satisfies the event E,

$$\mathbb{P}\big(X_p \in \underline{S} \mid \mu_p, \Sigma_p\big) \geq c_0[\log(p)]^{d/2} \cdot p^{-(b+\epsilon/2)}, \tag{A.12}$$

for some constant $c_0 > 0$. It follows that

$$\mathbb{P}\big(X_p \in \underline{S}\big) \geq \mathbb{P}(E) \cdot c_0[\log(p)]^{d/2} p^{-(b+\epsilon/2)}.$$

We plug it into the left hand side of (A.7) and note that $\mathbb{P}(E) \to 1$ as $p \to \infty$. This gives the desirable claim in (A.7).

Next, we show (A.8). We define a counterpart of the set $\overline{S}$ by

$$\overline{S}_p = \{x \in \mathbb{R}^d : m_p(x) \geq b\}.$$

Define $Y_p = \sqrt{2\log(p)} \cdot \Sigma_p^{-1/2}(X_p - \mu_p)$. Then, $Y_p \sim \mathcal{N}_d(0, I_d)$ and

$$X_p \in \overline{S}_p \qquad \text{if and only if} \qquad \|Y_p\|^2 \geq 2b\log(p).$$

The distribution of $\|Y_p\|^2$ is a χ^2_d distribution, which does not depend on (μ_p, Σ_p). We have

$$\begin{aligned}\mathbb{P}\big(X_p \in \overline{S}_p\big) &= \mathbb{E}\big[\mathbb{P}\big(X_p \in \overline{S}_p \mid \mu_p, \Sigma_p\big)\big] \\ &= \mathbb{E}\big[\mathbb{P}\big(\|Y_p\|^2 \geq 2b\log(p)\big)\big] \\ &= \mathbb{P}\big(\chi^2_d \geq 2b\log(p)\big). \end{aligned} \tag{A.13}$$

For chi-square distribution, the tail probability has an explicit form:

$$\mathbb{P}\big(\chi^2_d \geq 2b\log(p)\big) = \frac{\Gamma(d/2,\ b\log(p))}{\Gamma(d/2)},$$

where $\Gamma(s,x) \equiv \int_x^\infty t^{s-1}\exp(-t)dt$ is the upper incomplete gamma function and $\Gamma(s) \equiv \Gamma(s,0)$ is the ordinary gamma function. By property of the upper incomplete gamma function, $\Gamma(s,x)/(x^{s-1}\exp(-x)) \to 1$ as $x \to \infty$. It follows that

$$\frac{\Gamma(d/2,\ b\log(p))}{[b\log(p)]^{d/2-1}p^{-b}} \to 1, \qquad \text{as} \quad p \to \infty.$$

In particular, when p is sufficiently large, the left hand side is $\geq 1/2$. We plug these results into (A.13) to get

$$\mathbb{P}\big(X_p \in \overline{S}_p\big) \geq \frac{[b\log(p)]^{d/2-1}}{2\Gamma(d/2)} \cdot p^{-b}. \tag{A.14}$$

It remains to study the difference caused by replacing $\overline{S}_p$ by $\overline{S}$. Let

$$U_p = (\overline{S} \backslash \overline{S}_p) \cup (\overline{S}_p \backslash \overline{S}).$$

Then,

$$\big|\mathbb{P}\big(X_p \in \overline{S}\big) - \mathbb{P}\big(X_p \in \overline{S}_p\big)\big| \leq \mathbb{P}\big(X_p \in U_p\big). \tag{A.15}$$

Similar to (A.10), we have

$$\begin{aligned}\mathbb{P}\big(X_p \in U_p \mid \mu_p, \Sigma_p\big) &= \frac{[2\log(p)]^{d/2}}{(2\pi)^{d/2}|\det(\Sigma_p)|^{1/2}} \int_{x \in U_p} p^{-m_p(x)} dx \\ &\leq \frac{[\log(p)]^{d/2}}{\pi^{d/2}|\det(\Sigma_p)|^{1/2}} \cdot \text{Volume}(U_p) \cdot p^{-\inf_{x \in U_p}\{m_p(x)\}}. \end{aligned} \tag{A.16}$$

For a constant $\gamma > 0$ to be decided, let F be the event that

$$\|\mu_p - \mu\| \le \gamma, \qquad \text{and} \quad \|\Sigma_p - \Sigma\| \le \gamma. \tag{A.17}$$

On this event, we study both Volume(U_p) and $\inf_{x\in U_p} m_p(x)$. Re-write

$$U_p = (\overline{S}^c \backslash \overline{S}_p^c) \cup (\overline{S}_p^c \backslash \overline{S}^c).$$

By definition, $\overline{S}^c = \{x \in \mathbb{R}^d : m(x) \le b\} = \{x \in \mathbb{R}^d : \|\Sigma^{-1/2}(x-\mu)\| \le \sqrt{b}\}$, and $\overline{S}_p^c = \{x \in \mathbb{R}^d : \|\Sigma_p^{-1/2}(x-\mu_p)\| \le \sqrt{b}\}$. On the event F, for any $x \in \overline{S}_p^c$,

$$\begin{aligned}
\|\Sigma^{-1/2}(x-\mu)\| &\le \sqrt{b} + \|\Sigma^{-1/2}(x-\mu) - \Sigma_p^{-1/2}(x-\mu_p)\| \\
&\le \sqrt{b} + \|\Sigma^{-1/2}(\mu_p - \mu)\| + \|(\Sigma^{-1/2} - \Sigma_p^{-1/2})(x-\mu_p)\| \\
&\le \sqrt{b} + \|\Sigma^{-1/2}\| \cdot \|\mu_p - \mu\| + \|\Sigma^{1/2}\Sigma_p^{-1/2} - I_d\| \cdot \|\Sigma_p^{-1/2}(x-\mu_p)\| \\
&\le \sqrt{b} + \|\Sigma^{-1/2}\| \cdot \|\mu_p - \mu\| + \sqrt{b} \cdot \|\Sigma^{1/2}\Sigma_p^{-1/2} - I_d\| \\
&\le \sqrt{b} + C_5\gamma,
\end{aligned}$$

for a constant $C_5 > 0$ that does not depend on γ. Choosing $\gamma < C_5^{-1}\sqrt{b}$, we have $\|\Sigma^{-1/2}(x-\mu)\| \le 2\sqrt{b}$ for all $x \in \overline{S}_p^c$. Additionally, by definition, $\|\Sigma^{-1/2}(x-\mu)\| \le \sqrt{b}$ for all $x \in \overline{S}^c$. Combining the above gives

$$U_p \ \subset \ (\overline{S}^c \cup \overline{S}_p^c) \ \subset \ \big\{x \in \mathbb{R}^d : \|\Sigma^{-1/2}(x-\mu)\| \le 2\sqrt{b}\big\}.$$

Recall that B_d is the unit ball in $\mathbb{R}^d$. It follows immediately that

$$\text{Volume}(U_p) \le (2\sqrt{b})^d \cdot \text{Volume}(B_d), \qquad \text{on the event } F. \tag{A.18}$$

At the same time, for any $x \in \overline{S}$, on the event F,

$$\begin{aligned}
\|\Sigma_p^{-1/2}(x-\mu_p)\| &\ge \|\Sigma^{-1/2}(x-\mu)\| - \|\Sigma_p^{-1/2}(x-\mu_p) - \Sigma^{-1/2}(x-\mu)\| \\
&\ge \|\Sigma^{-1/2}(x-\mu)\| - \|\Sigma_p^{-1/2}(\mu_p - \mu)\| - \|(\Sigma^{-1/2} - \Sigma_p^{-1/2})(x-\mu)\| \\
&\ge \|\Sigma^{-1/2}(x-\mu)\| - \|\Sigma_p^{-1/2}\| \cdot \|\mu_p - \mu\| - \|\Sigma_p^{-1/2}\Sigma^{1/2} - I_d\| \cdot \|\Sigma^{-1/2}(x-\mu)\| \\
&= \|\Sigma^{-1/2}(x-\mu)\|\big(1 - \|\Sigma_p^{-1/2}\Sigma^{1/2} - I_d\|\big) - \|\Sigma^{-1/2}\| \cdot \|\mu_p - \mu\| \\
&\ge \|\Sigma^{-1/2}(x-\mu)\|(1 - C_6\gamma) - \|\Sigma^{-1/2}\|\gamma \\
&\ge \sqrt{b}(1 - C_6\gamma) - \|\Sigma^{-1/2}\|\gamma,
\end{aligned}$$

where $C_6 > 0$ is a constant that does not depend on γ and in the last line we have used the fact that $\|\Sigma^{-1/2}(x-\mu)\| \ge \sqrt{b}$ for $x \in \overline{S}$. We choose γ properly small so that $\sqrt{b}(1 - C_6\gamma) - \|\Sigma^{-1/2}\|\gamma \ge \sqrt{b - \epsilon/2}$. It follows that

$$m_p(x) = \|\Sigma_p^{-1/2}(x-\mu_p)\|^2 \ge b - \epsilon/2, \qquad \text{for all } x \in \overline{S}. \tag{A.19}$$

Additionally, the definition of $\overline{S}_p$ already guarantees that $m_p(x) \ge b$ for all $x \in \overline{S}_p$. Consequently,

$$\inf_{x\in U_p} m_p(x) \ge \inf_{x \in \overline{S}\cup\overline{S}_p} \{m_p(x)\} \ge b - \epsilon/2, \qquad \text{on the event } F. \tag{A.20}$$

We plug (A.18) and (A.20) into (A.16). It yields that, on the event F,

$$\mathbb{P}\big(X_p \in U_p \mid \mu_p, \Sigma_p\big) \leq C_7[\log(p)]^{d/2} \cdot p^{-(b-\epsilon/2)}, \tag{A.21}$$

for a constant $C_7 > 0$. Then,

$$\mathbb{P}\big(X_p \in U_p\big) \leq \mathbb{P}(F) \cdot C_7[\log(p)]^{d/2} \cdot p^{-(b-\epsilon/2)} + \mathbb{P}(F^c).$$

By our assumption, for any $\gamma > 0$ and $L > 0$, $\mathbb{P}(\|\mu_p - \mu\| > \gamma) \leq p^{-L}$ and $\mathbb{P}(\|\Sigma_p - \Sigma\| > \gamma) \leq p^{-L}$. In particular, we can choose $L = b$. It gives

$$\mathbb{P}(F^c) \leq p^{-b}.$$

We combine the above results and plug them into (A.15). It follows that

$$\big|\mathbb{P}\big(X_p \in \overline{S}\big) - \mathbb{P}\big(X_p \in \overline{S}_p\big)\big| \leq C_7[\log(p)]^{d/2} \cdot p^{-(b-\epsilon/2)} + p^{-b}. \tag{A.22}$$

Combining (A.14) and (A.22) gives

$$\mathbb{P}\big(X_p / \in \overline{S}\big) \leq [1 + o(1)] \cdot C_7[\log(p)]^{d/2} \cdot p^{-(b-\epsilon/2)}.$$

This gives the claim in (A.8). The proof of this lemma is complete.

Appendix B. Proof of Lemma 7.2

First, we study the least-squares. Note that $\hat{\beta}$ has an explicit solution: $\hat{\beta} = G^{-1}X^T y$. Since G is a block-wise diagonal matrix, we immediately have

$$\begin{bmatrix} \hat{\beta}_j \\ \hat{\beta}_{j+1} \end{bmatrix} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}^{-1} \begin{bmatrix} x_j^T y \\ x_{j+1}^T y \end{bmatrix} = \frac{1}{1-\rho^2} \begin{bmatrix} x_j^T y - \rho x_{j+1}^T y \\ x_{j+1}^T y - \rho x_{j+1}^T y \end{bmatrix}.$$

Recall that $\tilde{y} = X'y/\sqrt{2\log(p)}$. Then, $|\hat{\beta}_j| > \sqrt{2u\log(p)}$ if and only if

$$\frac{1}{1-\rho^2}|\tilde{y}_j - \rho\tilde{y}_{j+1}| > \sqrt{u}.$$

It immediately gives the rejection region for least-squares.

Next, we study the Lasso-path. We write $W_j^{*,\mathrm{path}}$ as W_j^* for notation simplicity. The lasso estimate $\hat{\beta}(\lambda)$ minimizes the objective

$$Q(b) = \frac{1}{2}\|y - Xb\|^2 + \lambda\|b\|_1 = \frac{1}{2}\|y\|^2 - y^T Xb + \frac{1}{2}b^T Gb + \lambda\|b\|_1.$$

When G is a block-wise diagonal matrix, the objective $Q(b)$ is separable, and we can optimize over each pair of (b_j, b_{j+1}) separately. It reduces to solving many bi-variate problems:

$$(\hat{\beta}_j(\lambda), \hat{\beta}_{j+1}(\lambda))^T = \mathrm{argmin}_b\Big\{\frac{1}{2}||y - [x_j, x_{j+1}]b||_2^2 + \lambda||b||_1\Big\}. \tag{B.1}$$

Write $\hat{b} = (\hat{\beta}_j(\lambda), \hat{\beta}_{j+1}(\lambda))^T$ and let

$$B = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \qquad \text{and} \qquad h = \begin{bmatrix} x_j^T y \\ x_{j+1}^T y \end{bmatrix}.$$

Then, the optimization (B.1) can be written as

$$\hat{b} = \text{argmin}_b\big\{-h^T b + b^T B b/2 + \lambda\|b\|_1\big\}. \tag{B.2}$$

Recall that W_j^* is the value of λ at which $\hat{b}_1$ becomes nonzero for the first time. Our goal is to find a region of (h_1, h_2) such that $W_j^* > t_p(u) \equiv \sqrt{2u\log(p)}$.

It suffices to consider the case of $\rho \geq 0$. To see this, we consider changing ρ to $-\rho$ in the matrix B. The objective remains unchanged if we also change b_2 to $-b_2$ and h_2 to $-h_2$. Note that the change of b_2 to $-b_2$ has no impact on W_j^*; this means W_j^* is unchanged if we simultaneously flip the sign of ρ and h_2. Consequently, once we know the rejection region for $\rho > 0$, we can immediately obtain that for $\rho < 0$ by a reflection of the region with respect to the x-axis.

Below, we fix $\rho \geq 0$. We first derive the explicit form of the whole solution path and then use it to decide the rejection region. Taking sub-gradients of (B.1), we find that $\hat{b}$ has to satisfy

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \end{bmatrix} + \lambda \begin{bmatrix} \text{sgn}(\hat{b}_1) \\ \text{sgn}(\hat{b}_2) \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}, \tag{B.3}$$

where $\text{sgn}(x) = 1$ if $x > 0$, $\text{sgn}(x) = -1$ if $x < 0$, and $\text{sgn}(x)$ can be equal to any value in $[-1, 1]$ if $x = 0$. Let $\lambda_1 > \lambda_2 > 0$ be the values at which variables enter the solution path. When $\lambda \in (\lambda_1, \infty)$, $\hat{b}_1 = 0$ and $\hat{b}_2 = 0$. Plugging them into (B.3) gives $\text{sgn}(\hat{b}_1) = \lambda^{-1} h_1$. The definition of $\text{sgn}(\hat{b}_1)$ implies that $|h_1| \leq \lambda$, for any $\lambda > \lambda_1$. We then have $|h_1| \leq \lambda_1$. Similarly, it is true that $|h_2| \leq \lambda_1$. It gives

$$\lambda_1 = \max\{|h_1|, |h_2|\}. \tag{B.4}$$

We first assume $|h_1| > |h_2|$. By (B.3) and continuity of solution path, there exists a sufficiently small constant $\delta > 0$ such that, for $\lambda \in (\lambda_2 - \delta, \lambda_2)$, the following equation holds.

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} \hat{b}_1(\lambda) \\ \hat{b}_2(\lambda) \end{bmatrix} + \lambda \begin{bmatrix} \text{sgn}(\hat{b}_1) \\ \text{sgn}(\hat{b}_2) \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}. \tag{B.5}$$

The sign vector of $\hat{b}$ for $\lambda \in (\lambda_2 - \delta, \lambda_2)$ has four different cases: $(1,1)^T$, $(1,-1)^T$, $(-1,1)^T$, $(-1,-1)^T$. For these four different cases, we can use (B.5) to solve $\hat{b}$. The solutions in four cases are respectively

$$\frac{1}{1-\rho^2}\begin{bmatrix} (h_1 - \rho h_2) - (1-\rho)\lambda \\ (h_2 - \rho h_1) - (1-\rho)\lambda \end{bmatrix}, \qquad \frac{1}{1-\rho^2}\begin{bmatrix} (h_1 - \rho h_2) - (1+\rho)\lambda \\ (h_2 - \rho h_1) + (1+\rho)\lambda \end{bmatrix},$$
$$\frac{1}{1-\rho^2}\begin{bmatrix} (h_1 - \rho h_2) + (1+\rho)\lambda \\ (h_2 - \rho h_1) - (1+\rho)\lambda \end{bmatrix}, \qquad \frac{1}{1-\rho^2}\begin{bmatrix} (h_1 - \rho h_2) + (1-\rho)\lambda \\ (h_2 - \rho h_1) + (1-\rho)\lambda \end{bmatrix}.$$

The solution $\hat{b}$ has to match the sign assumption on $\hat{b}$. For each of the four cases, the requirement becomes

- Case 1: $(h_1 - \rho h_2) - (1-\rho)\lambda > 0, \quad (h_2 - \rho h_1) - (1-\rho)\lambda > 0.$
- Case 2: $(h_1 - \rho h_2) - (1+\rho)\lambda > 0, \quad (h_2 - \rho h_1) + (1+\rho)\lambda < 0.$
- Case 3: $(h_1 - \rho h_2) + (1+\rho)\lambda < 0, \quad (h_2 - \rho h_1) - (1+\rho)\lambda > 0.$
- Case 4: $(h_1 - \rho h_2) + (1-\rho)\lambda < 0, \quad (h_2 - \rho h_1) + (1-\rho)\lambda < 0.$

Note that we have assumed $|h_1| > |h_2|$. Then, Case k is possible only in the region $\mathcal{A}_k$, where

$$\mathcal{A}_1 = \{(h_1,h_2) : h_1 > 0, \ \ \rho h_1 < h_2 < h_1\}, \quad \mathcal{A}_2 = \{(h_1,h_2) : h_1 > 0, \ \ -h_1 < h_2 < \rho h_1\},$$
$$\mathcal{A}_3 = \{(h_1,h_2) : h_1 < 0, \ \ \rho h_1 < h_2 < -h_1\}, \quad \mathcal{A}_4 = \{(h_1,h_2) : h_1 < 0, \ \ h_1 < h_2 < \rho h_1\}.$$

In each case, $\lambda_1 = |h_1|$. To get the value of λ_2, we use the continuity of the solution path. It implies that $\hat{b}_2(\lambda) = 0$ at $\lambda = \lambda_2$. As a result, the value of λ_2 in Case k is

$$\lambda_2^{(1)} = \frac{h_2 - \rho h_1}{1-\rho}, \qquad \lambda_2^{(2)} = \frac{\rho h_1 - h_2}{1+\rho}, \qquad \lambda_2^{(3)} = \frac{h_2 - \rho h_1}{1+\rho}, \qquad \lambda_2^{(4)} = \frac{\rho h_1 - h_2}{1-\rho}. \quad \text{(B.6)}$$

It is easy to verify that $\lambda_2 < \lambda_1$ in each case. We also need to check that in the region $\mathcal{A}_k$, the KKT condition (B.3) can be satisfied with $\hat{b}_2 = 0$ for all $\lambda \in (\lambda_2^{(k)}, \lambda_1)$. For example, in Case 1, (B.3) becomes

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} \hat{b}_1 \\ 0 \end{bmatrix} + \lambda \begin{bmatrix} 1 \\ c \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}, \qquad \text{for some } |c| \le 1.$$

We can solve the equations to get $\hat{b}_1 = h_1 - \lambda$ and $\lambda c = h_2 - \rho\hat{b}_1 = (h_2 - \rho h_1) - \lambda$. It can be verified that $|(h_2 - \rho h_1) - \lambda| \le \lambda$ for $(h_1, h_2) \in \mathcal{A}_1$ and $\lambda \in (\lambda_2^{(1)}, \lambda_1)$. The verification for other cases is similar and thus omitted. We then assume $|h_2| > |h_1|$. By symmetry, we will have the same result, except that (h_1, h_2) are switched in the expression of $\mathcal{A}$ and (λ_1, λ_2). This gives the other four cases:

$$\mathcal{A}_5 = \{(h_1,h_2) : h_2 > 0, \ \ \rho h_2 < h_1 < h_2\}, \quad \mathcal{A}_6 = \{(h_1,h_2) : h_2 > 0, \ \ -h_2 < h_1 < \rho h_2\},$$
$$\mathcal{A}_7 = \{(h_1,h_2) : h_2 < 0, \ \ \rho h_2 < h_1 < -h_2\}, \quad \mathcal{A}_8 = \{(h_1,h_2) : h_2 < 0, \ \ h_2 < h_1 < \rho h_2\}.$$

In these four cases, we similarly have $\lambda_1 = |h_2|$ and

$$\lambda_2^{(5)} = \frac{h_1 - \rho h_2}{1-\rho}, \qquad \lambda_2^{(6)} = \frac{\rho h_2 - h_1}{1+\rho}, \qquad \lambda_2^{(7)} = \frac{h_1 - \rho h_2}{1+\rho}, \qquad \lambda_2^{(8)} = \frac{\rho h_2 - h_1}{1-\rho}. \quad \text{(B.7)}$$

These eight regions are shown in Figure 12.

We then compute W_j^* and the associated rejection region. Note that $W_j^* = \lambda_1$ in Case 1-Case 4, and $W_j^* = \lambda_2$ in Case 5-Case 8. It follows directly that

$$W_j^* = \begin{cases} |h_1|, & \text{if } (h_1,h_2) \in \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3 \cup \mathcal{A}_4, \\ |h_1 - \rho h_2|/(1-\rho), & \text{if } (h_1,h_2) \in \mathcal{A}_5 \cup \mathcal{A}_8, \\ |h_1 - \rho h_2|/(1+\rho), & \text{if } (h_1,h_2) \in \mathcal{A}_6 \cup \mathcal{A}_7. \end{cases} \quad \text{(B.8)}$$

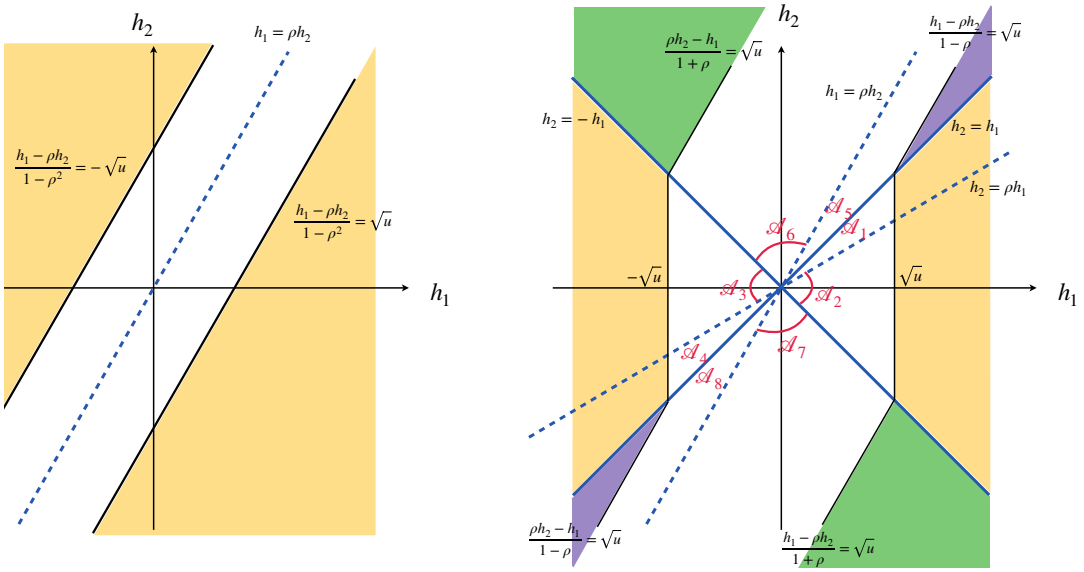

Figure 12: The rejection region of least-squares (left) and Lasso-path (right). On the right panel, the regions $\mathcal{A}_1$-$\mathcal{A}_8$ are the same as those defined in the proof. In the regions $\mathcal{A}_1$-$\mathcal{A}_4$, $W_j^* = |h_1|$, and the rejection region is colored by yellow. In the regions $\mathcal{A}_5$ and $\mathcal{A}_8$, $W_j^* = |h_1 - \rho h_2|/(1-\rho)$, and the rejection region is colored by purple. In the regiions $\mathcal{A}_6$ and $\mathcal{A}_7$, $W_j^* = |h_1 - \rho h_2|/(1+\rho)$, and the rejection region is colored by green.

As a result, the region $W_j^* > \sqrt{2u\log(p)}$ if and only if the vector $(x_j^T y, x_{j+1}^T y)/\sqrt{2\log(p)}$ is in the following set:

$$\begin{aligned}\mathcal{R} = &\big[(\mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3 \cup \mathcal{A}) \cap \{|h_1| > \sqrt{u}\}\big] \\ &\cup \big[(\mathcal{A}_5 \cup \mathcal{A}_8) \cap \{|h_1 - \rho h_2| > (1-\rho)\sqrt{u}\}\big] \\ &\cap \big[(\mathcal{A}_6 \cup \mathcal{A}_7) \cap \{|h_1 - \rho h_2| > (1+\rho)\sqrt{u}\}\big].\end{aligned}$$

In Figure 12, the 3 subsets are colored by yellow, purple, and green, respectively. This gives the rejection region for Lasso-path.

Appendix C. Proof of Theorem 4.1

By definition of $(\mathrm{FP}_p, \mathrm{FN}_p)$ and the Rare/Weak signal model (3.2)-(3.3), we have

$$\mathrm{FP}_p = \sum_{j=1}^{p}(1-\epsilon_p)\mathbb{P}(W_j > t_p(u)|\beta_j = 0), \quad \mathrm{FN}_p = \sum_{j=1}^{p}\epsilon_p\,\mathbb{P}(W_j < t_p(u)|\beta_j = \tau_p), \tag{C.1}$$

where $\epsilon_p = p^{-\vartheta}$, $\tau_p = \sqrt{2r\log(p)}$, and $t_p(u) = \sqrt{2u\log(p)}$. Therefore, it suffices to study $\mathbb{P}(W_j > t_p(u)|\beta_j = 0)$ and $\mathbb{P}(W_j < t_p(u)|\beta_j = \tau_p)$.

Fix $1 \le j \le p$. The knockoff filter applies Lasso to the design matrix $[X, \tilde{X}]$. This design is belongs to the block-wise diagonal design (5.1) with a dimension $2p$ and $\rho = a$. The variable j and its own knockoff are in one block. Fix j and write

$$h_1 = x_j' y / \sqrt{2\log(p)}, \qquad \text{and} \qquad h_2 = \tilde{x}_j' y / \sqrt{2\log(p)}. \tag{C.2}$$

It is easy to see that $(x_j' y, \tilde{x}_j' y)'$ follows a distribution $\mathcal{N}_2(\mathbf{0}_2, \Sigma)$ when $\beta_j = 0$, and it follows a distribution $\mathcal{N}_2(\mu\sqrt{2\log(p)},\ \Sigma)$, when $\beta_j = \tau_p$, where

$$\mu = \begin{bmatrix} \sqrt{r} \\ a\sqrt{r} \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix}.$$

Let $\mathcal{R}$ be the region of (h_1, h_2) corresponding to the event that $\{W_j > t_p(u)\}$. It follows from Lemma 7.1 that

$$\begin{aligned} \mathbb{P}(W_j > t_p(u) | \beta_j = 0) &= L_p p^{-\inf_{h\in\mathcal{R}}\{h'\Sigma^{-1}h\}}, \\ \mathbb{P}(W_j < t_p(u) | \beta_j = \tau_p) &= L_p p^{-\inf_{h\in\mathcal{R}^c}\{(h-\mu)'\Sigma^{-1}(h-\mu)\}}. \end{aligned} \tag{C.3}$$

Below, we first derive the rejection region $\mathcal{R}$, and then compute the exponents in (C.3).

Recall that Z_j and $\tilde{Z}_j$ are the same as in (4.3). They are indeed the values of λ at which the variable j and its knockoff enter the solution path of a bivariate lasso as in (B.1). We can apply the solution path derived in the proof of Lemma 7.2, with $\rho = a$. Before we proceed to the proof, we argue that it suffices to consider the case of $a \ge 0$. If $a < 0$, we can simultaneously flip the signs of a and h_2, so that the objective (B.1) remains unchanged; as a result, the values of $(Z_j, \tilde{Z}_j)$ remain unchanged, so does the symmetric statistic W_j. It implies that, if we flip the sign of a, the rejection region is reflected with respect to the x-axis. At the same time, in light of the exponents in (C.3), we consider two ellipsoids

$$\mathcal{E}_{\text{FP}}(t) = \{h \in \mathbb{R}^2 : h'\Sigma^{-1}h \le t\}, \qquad \mathcal{E}_{\text{FN}}(t) = \{h \in \mathbb{R}^2 : (h-\mu)'\Sigma^{-1}(h-\mu)' \le t\}. \tag{C.4}$$

Similarly, if we simultaneously flip the signs of a and h_2, these ellipsoids remain unchanged. It implies that, if we flip the sign of a, these ellipsoids are reflected with respect to the x-axis. Combining the above observations, we know that the exponents in (C.3) are unchanged with a sign flip of a, i.e., they only depend on $|a|$. We assume $a \ge 0$ without loss of generality.

Fix $a \ge 0$. Write $z = Z_j/\sqrt{2\log(p)}$ and $\tilde{z} = \tilde{Z}_j/\sqrt{2\log(p)}$. The symmetric statistics in (4.3) can be re-written as

$$W_j^{\text{sgm}} = (z \vee \tilde{z})\sqrt{2\log(p)} \cdot \begin{cases} +1, & \text{if } z > \tilde{z} \\ -1, & \text{if } z \le \tilde{z} \end{cases}, \qquad W_j^{\text{dif}} = (z - \tilde{z})\sqrt{2\log(p)}.$$

Recall that h_1 and h_2 are as in (C.2). Let $\lambda_1 > \lambda_2 > 0$ be the values of λ at which variables enter the solution path of a bivariate lasso. In the proof of Lemma 7.2, we have derived the formula of (λ_1, λ_2); see (B.6) and (B.7) (with ρ replaced by a). It follows that

$$(z, \tilde{z}) = \begin{cases} (\lambda_1, \lambda_2), & \text{in the regions } \mathcal{A}_1\text{-}\mathcal{A}_4, \\ (\lambda_2, \lambda_1), & \text{in the regions } \mathcal{A}_5\text{-}\mathcal{A}_8, \end{cases}$$

where regions $\mathcal{A}_1$-$\mathcal{A}_8$ are the same as those on the right panel of Figure 12 (with ρ replaced by a). Plugging in (B.6) and (B.7) gives the following results:

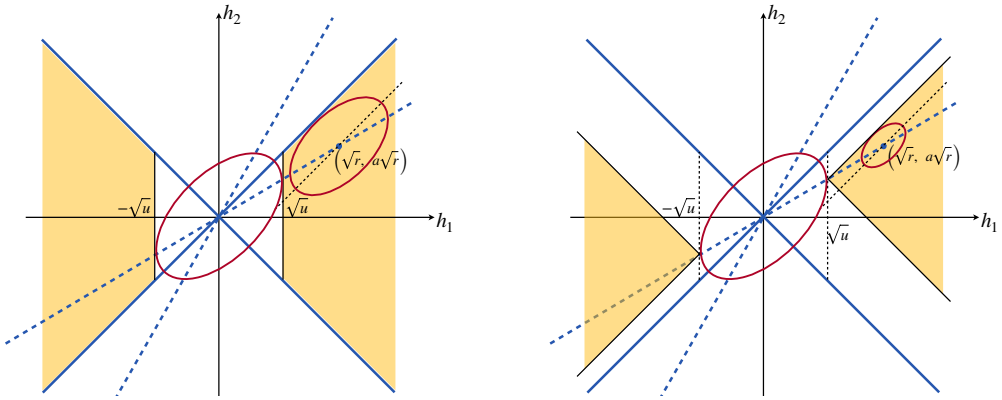

Figure 13: The rejection region of knockoff in the orthogonal design, where the symmetric statistic is signed maximum (left) and difference (right). The rate of convergence of FP_p is captured by an ellipsoid centered at $(0,0)$, and the rate of convergence of FN_p is captured by an ellipsoid centered at $(\sqrt{r}, a\sqrt{r})$.

- Region $\mathcal{A}_1$: $z = h_1$, $\tilde{z} = \frac{h_2-\rho h_1}{1-a}$, $W_j^{\mathrm{sgm}} = h_1\sqrt{2\log(p)}$, $W_j^{\mathrm{dif}} = \frac{h_1-h_2}{1-a}\sqrt{2\log(p)}$.
- Region $\mathcal{A}_2$: $z = h_1$, $\tilde{z} = \frac{\rho h_1-h_2}{1+a}$, $W_j^{\mathrm{sgm}} = h_1\sqrt{2\log(p)}$, $W_j^{\mathrm{dif}} = \frac{h_1+h_2}{1+a}\sqrt{2\log(p)}$.
- Region $\mathcal{A}_3$: $z = -h_1$, $\tilde{z} = \frac{h_2-\rho h_1}{1+a}$, $W_j^{\mathrm{sgm}} = -h_1\sqrt{2\log(p)}$, $W_j^{\mathrm{dif}} = -\frac{h_1+h_2}{1+a}\sqrt{2\log(p)}$.
- Region $\mathcal{A}_4$: $z = -h_1$, $\tilde{z} = \frac{\rho h_1-h_2}{1-a}$, $W_j^{\mathrm{sgm}} = -h_1\sqrt{2\log(p)}$, $W_j^{\mathrm{dif}} = \frac{h_2-h_1}{1-a}\sqrt{2\log(p)}$.
- Regions $\mathcal{A}_5$-$\mathcal{A}_8$: $|Z_j| < |\tilde{Z}_j|$, $W_j^{\mathrm{sgm}} < 0$, $W_j^{\mathrm{dif}} < 0$.

The event that $W_j^{\mathrm{sgm}} > \sqrt{2u\log(p)}$ corresponds to that (h_1, h_2) is in the region of

$$\begin{aligned}\mathcal{R}_u^{\mathrm{sgm}} &= (\mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3 \cup \mathcal{A}_4) \cap \{|h_1| > \sqrt{u}\} \\ &= \{|h_1| > |h_2|,\ |h_1| > \sqrt{u}\}. \end{aligned} \tag{C.5}$$

The event that $W_j^{\mathrm{dif}} > \sqrt{2u\log(p)}$ corresponds to that (h_1, h_2) is in the region of

$$\begin{aligned}\mathcal{R}_u^{\mathrm{dif}} = \big(\mathcal{A}_1 \cap \{h_1 - h_2 > (1-a)\sqrt{u}\}\big) \cup \big(\mathcal{A}_2 \cap \{h_1 + h_2 > (1+a)\sqrt{u}\}\big) \\ \cup \big(\mathcal{A}_3 \cap \{h_1 + h_2 < -(1+a)\sqrt{u}\}\big) \cup \big(\mathcal{A}_4 \cap \{h_1 - h_2 < -(1-a)\sqrt{u}\}\big). \end{aligned} \tag{C.6}$$

These two regions are shown in Figure 13.

We are now ready to compute the exponents in (C.3). First, we compute $\inf_{h\in\mathcal{R}}\{h'\Sigma^{-1}h\}$. Let $\mathcal{E}_{\mathrm{FP}}(t)$ be the same as in (C.4). Then,

$$\inf_{h\in\mathcal{R}}\{h'\Sigma^{-1}h\} = \sup\big\{t > 0 : \mathcal{E}_{\mathrm{FP}}(t) \cap \mathcal{R} \neq \emptyset\big\}.$$

When the rejection region is $\mathcal{R}_u^{\mathrm{sgm}}$, from Figure 13, we can increase t until $\mathcal{E}_{\mathrm{FP}}(t)$ intersects with the line of $h_1 = \pm\sqrt{u}$. For any h on the surface of this ellipsoid, the perpendicular

vector of its tangent plane is proportional to $'\Sigma^{-1}h$. When the ellipsoid intersects with the line of $h_1 = \pm\sqrt{u}$, the perpendicular vector should be proportional to $(1,0)'$. Therefore, we need to find h such that

$$h_1 = \pm\sqrt{u}, \quad h'\Sigma^{-1}h = t, \quad \text{and} \quad \Sigma^{-1}h \propto (1,0)'.$$

The second equation requires that $h_2 = ah_1$. Combining it with the first equation gives $h = (\pm\sqrt{u}, \pm a\sqrt{u})$. We then plug it into the second equation to obtain $t = u$. This gives

$$\inf_{h\in\mathcal{R}_u^{\text{sgm}}}\{h'\Sigma^{-1}h\} = u. \tag{C.7}$$

When the rejection region is $\mathcal{R}_u^{\text{dif}}$, there are 3 possible cases:

(i) The ellipsoid intersects with the line $h_1 - h_2 = (1-a)\sqrt{u}$,

(ii) The ellipsoid intersects with the line $h_1 + h_2 = (1+a)\sqrt{u}$,

(iii) The ellipsoid intersects with the point $h = (\sqrt{u}, a\sqrt{u})$.

In Case (i), we can compute the intersection point by solving h for $h_1 - h_2 = (1-a)\sqrt{u}$ and $\Sigma^{-1}h \propto (1,-1)'$. The second relationship gives $h_2 = -h_1$. Together with the first relationship, we have $h = (\frac{1-a}{2}\sqrt{u}, \frac{1-a}{2}\sqrt{u})$. It is not in $\mathcal{R}_u^{\text{dif}}$. Similarly, for Case (ii), we can show that the intersection point is $h = (\frac{1+a}{2}\sqrt{u}, \frac{1+a}{2}\sqrt{u})$, which is not in $\mathcal{R}_u^{\text{dif}}$ either. The only possible case is Case (iii), where the intersection point is $(\sqrt{u}, a\sqrt{u})$ and the associated $t = h'\Sigma^{-1}t = u$. We have proved that

$$\inf_{h\in\mathcal{R}_u^{\text{dif}}}\{h'\Sigma^{-1}h\} = u. \tag{C.8}$$

Next, we compute $\inf_{h\in\mathcal{R}^c}\{(h-\mu)'\Sigma^{-1}(h-\mu)\}$. Let $\mathcal{E}_{\text{FN}}(t)$ be the same as in (C.4). Then,

$$\inf_{h\in\mathcal{R}^c}\{(h-\mu)'\Sigma^{-1}(h-\mu)\} = \sup\big\{t>0 : \mathcal{E}_{\text{FN}}(t)\cap\mathcal{R}^c \neq \emptyset\big\}.$$

Note that the center of the ellipsoid is $\mu = (\sqrt{r}, a\sqrt{r})$. When either $\mathcal{R} = \mathcal{R}_u^{\text{sgm}}$ or $\mathcal{R} = \mathcal{R}_u^{\text{dif}}$, $\mu \notin \mathcal{R}^c$ if and only if $r > u$. In other words, the above is well defined only if $r > u$. We now fix $r > u$. When the rejection region is $\mathcal{R}_u^{\text{sgm}}$, the ellipsoid intersects with either the line of $h_1 = \sqrt{u}$ or the line of $h_1 = h_2$. Since the perpendicular vector of the tangent plane of the ellipsoid at h is proportional to $'\Sigma^{-1}(h-\mu)$, we can solve the intersection points from

$$\begin{cases} h_1 = \sqrt{u}, \\ \Sigma^{-1}(h-\mu) \propto (1,0)', \end{cases} \quad \text{and} \quad \begin{cases} h_1 = h_2, \\ \Sigma^{-1}(h-\mu) \propto (1,-1)'. \end{cases}$$

By calculations, the two intersection points are $h = (\sqrt{u},\ a\sqrt{u})$ and $h = (\frac{1+a}{2}\sqrt{r},\ \frac{1+a}{2}\sqrt{r})$. The associated value of $(h-\mu)'\Sigma^{-1}(h-\mu)$ is $t = (\sqrt{r}-\sqrt{u})^2$ and $t = (1-a)r/2$, respectively. When we increase the ellipsoid until it interacts with $(\mathcal{R}_u^{\text{sgm}})^c$, the corresponding t is the smaller of the above two values. This gives

$$\inf_{h\in(\mathcal{R}_u^{\text{sgm}})^c}\{(h-\mu)'\Sigma^{-1}(h-\mu)\} = \min\Big\{(\sqrt{r}-\sqrt{u})_+^2,\ \frac{1-a}{2}r\Big\}. \tag{C.9}$$

When the rejection region is $\mathcal{R}_u^{\rm dif}$, the ellipsoid intersects with either the line of $h_1 - h_2 = (1-a)\sqrt{u}$ or the line of $h_1 + h_2 = (1+a)\sqrt{u}$. We can solve the intersection points from

$$\begin{cases} h_1 - h_2 = (1-a)\sqrt{u}, \\ \Sigma^{-1}(h-\mu) \propto (1,-1)', \end{cases} \qquad \text{and} \qquad \begin{cases} h_1 + h_2 = (1+a)\sqrt{u}, \\ \Sigma^{-1}(h-\mu) \propto (1,1)'. \end{cases}$$

Solving these equations gives the two intersection points: $h = (\frac{1+a}{2}\sqrt{r} + \frac{1-a}{2}\sqrt{u},\ \frac{1+a}{2}\sqrt{r} - \frac{1-a}{2}\sqrt{u})$ and $h = (\frac{1-a}{2}\sqrt{r} + \frac{1+a}{2}\sqrt{u},\ -\frac{1-a}{2}\sqrt{r} + \frac{1+a}{2}\sqrt{u})$. The corresponding value of $(h-\mu)'\Sigma^{-1}(h-\mu)$ is $t = \frac{1-a}{2}(\sqrt{r}-\sqrt{u})^2$ and $t = \frac{1+a}{2}(\sqrt{r}-\sqrt{u})^2$, respectively. The smaller of these two values is $\frac{1-a}{2}(\sqrt{r}-\sqrt{u})^2$. We have proved that

$$\inf_{h \in (\mathcal{R}_u^{\rm dif})^c} \{(h-\mu)'\Sigma^{-1}(h-\mu)\} = \frac{1-a}{2}(\sqrt{r}-\sqrt{u})_+^2. \tag{C.10}$$

We plug (C.7)-(C.10) into (C.3), and we further plug it into (C.1). This gives the claim for $a \geq 0$. As we have argued, the results for $a < 0$ only requires replacing a by $|a|$.

Appendix D. Proof of Theorem 5.1

Without loss of generality, we assume p is even. Then, for block-wise diagonal designs as in (5.1), the Lasso objective is separable. Therefore, for each W_j^*, it is not affected by any β_k outside the block. Additionally, by symmetry, the distribution of W_j^* is the same for all $1 \leq j \leq p$. It follows that

$$\begin{aligned} \mathrm{FP}_p(u) = {} & L_p p \cdot \mathbb{P}\big\{W_j^* > t_p(u) \,\big|\, (\beta_j, \beta_{j+1}) = (0,0)\big\} \\ & + L_p p^{1-\vartheta} \cdot \mathbb{P}\big\{W_j^* > t_p(u) \,\big|\, (\beta_j, \beta_{j+1}) = (0,\tau_p)\big\}, \end{aligned} \tag{D.1}$$

where j can be odd index. Similarly, we can derive that

$$\begin{aligned} \mathrm{FN}_p(u) = {} & L_p p^{1-\vartheta} \cdot \mathbb{P}\big\{W_j^* < t_p(u) \,\big|\, (\beta_j, \beta_{j+1}) = (\tau_p,0)\big\} \\ & + L_p p^{1-2\vartheta} \cdot \mathbb{P}\big\{W_j^* < t_p(u) \,\big|\, (\beta_j, \beta_{j+1}) = (\tau_p,\tau_p)\big\}. \end{aligned} \tag{D.2}$$

Fix variables $\{j, j+1\}$, and consider the random vector $\hat{h} = (x_j'y, x_{j+1}'y)'/\sqrt{\log(p)}$. Then,

$$\hat{h} \sim \mathcal{N}_2\Big(\mu,\ \frac{1}{\log(p)}\Sigma\Big), \qquad \text{where} \quad \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

The vector μ is equal to

$$\mu^{(1)} \equiv \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \mu^{(2)} \equiv \begin{bmatrix} \rho\sqrt{r} \\ \sqrt{r} \end{bmatrix}, \quad \mu^{(3)} \equiv \begin{bmatrix} \sqrt{r} \\ \rho\sqrt{r} \end{bmatrix}, \quad \mu^{(4)} \equiv \begin{bmatrix} (1+\rho)\sqrt{r} \\ (1+\rho)\sqrt{r}) \end{bmatrix}, \tag{D.3}$$

in the four cases where $(\beta_j, \beta_{j+1})'$ is $(0,0)'$, $(0,\tau_p)'$, $(\tau_p,0)'$, and $(\tau_p,\tau_p)'$, respectively. Let $\mathcal{R}_u$ be the rejection region induced by Lasso-path, given explicitly in Lemma 7.2. By Lemma 7.1, the probabilities in (D.1) and (D.2) are related to the following quantities:

$$\alpha_k = \begin{cases} \inf\inf_{h \in \mathcal{R}_u}\{(h-\mu^{(k)})'\Sigma^{-1}(h-\mu^{(k)})\}, & k = 1,2, \\ \inf_{h \in \mathcal{R}_u^c}\{(h-\mu^{(k)})'\Sigma^{-1}(h-\mu^{(k)})\}, & k = 3,4. \end{cases}$$

and plug it into (D.1) and (D.2). It gives

$$\mathrm{FP}_p(u) = L_p p^{1-\min\{\alpha_1,\, \vartheta+\alpha_2\}}, \quad \mathrm{FN}_p(u) = L_p p^{1-\min\{\vartheta+\alpha_3,\, 2\vartheta+\alpha_4\}}. \tag{D.4}$$

It remains to compute the exponents α_1-α_4.

First, we consider the case that $\rho \geq 0$. The rejection region in Figure 12 is defined by the following lines:

- Line 1: $h_1 - \rho h_2 = (1-\rho)\sqrt{u}$.
- Line 2: $h_1 = \sqrt{u}$.
- Line 3: $h_1 - \rho h_2 = (1+\rho)\sqrt{u}$.
- Line 4: $h_1 - \rho h_2 = -(1-\rho)\sqrt{u}$.
- Line 5: $h_1 = -\sqrt{u}$.
- Line 6: $h_1 - \rho h_2 = -(1+\rho)\sqrt{u}$.

Consider a general ellipsoid:

$$\mathcal{E}(t;\mu) = \{h \in \mathbb{R}^2 : (h-\mu)'\Sigma^{-1}(h-\mu)' \leq t\}.$$

Given any line $h_1 + bh_2 = c$, as t increases, this ellipsoid eventually intersects with this line. The intersection point is computed by the following equations:

$$h_1 + bh_2 = c, \qquad \Sigma^{-1}(h-\mu) \propto (1,b)'.$$

The second equation (it is indeed a linear equation on h) says that the perpendicular vector of the tangent plane is orthogonal to the line. Solving the above equations gives the intersection point and the value of t: As long as $b^2 \neq 1$, we have

$$h^* = \mu + \frac{c - (\mu_1 + b\mu_2)}{1 + b^2 + 2b\rho}\begin{bmatrix} 1+b\rho \\ b+\rho \end{bmatrix}, \qquad t^* = \frac{[c-(\mu_1+b\mu_2)]^2}{1+b^2+2b\rho}. \tag{D.5}$$

Using the expressions of lines 1-6, we can obtain the corresponding t^* for 6 lines:

$$t_1^* = \frac{[(1-\rho)\sqrt{u} - (\mu_1 - \rho\mu_2)]^2}{1-\rho^2}, \qquad t_2^* = (\sqrt{u} - \mu_1)^2, \qquad t_3^* = \frac{[(1+\rho)\sqrt{u} - (\mu_1 - \rho\mu_2)]^2}{1-\rho^2},$$
$$t_4^* = \frac{[(1-\rho)\sqrt{u} + (\mu_1 - \rho\mu_2)]^2}{1-\rho^2}, \qquad t_5^* = (\sqrt{u} + \mu_1)^2, \qquad t_6^* = \frac{[(1+\rho)\sqrt{u} + (\mu_1 - \rho\mu_2)]^2}{1-\rho^2}.$$

We first look at the ellipsoid $\mathcal{E}(t;\mu^{(1)})$ and study when it intersects with $\mathcal{R}_u$. Note that $\mu^{(1)} = (0,0)'$. The above t^* values become

$$t_2^* = t_5^* = u, \qquad t_1^* = t_4^* = \frac{u}{1+\rho}, \qquad t_3^* = t_6^* = \frac{u}{1-\rho}.$$

Therefore, as we increase t, this ellipsoid first intersects with line 1 and line 4. For line 1, the intersection point is $((1-\rho)\sqrt{u}, 0)'$, but it is outside the rejection region (see Figure 12);

the situation for line 4 is similar. We then further increase t, and the ellipsoid intersects with line 2 and line 5, where the intersection point is $(\sqrt{u}, \rho\sqrt{u})'$; this point is indeed on the boundary of the rejection region. We thus conclude that

$$\inf_{h\in\mathcal{R}_u}\{(h-\mu^{(1)})'\Sigma^{-1}(h-\mu^{(1)})\} = u. \tag{D.6}$$

We then look at the the ellipsoid $\mathcal{E}(t;\mu^{(2)})$, with $\mu^{(2)} = (\rho\sqrt{r}, \sqrt{r})'$. The t^* values for 6 lines are:

$$t_1^* = t_4^* = \frac{1-\rho}{1+\rho}u, \qquad t_2^* = (\sqrt{u}-\rho\sqrt{r})^2, \qquad t_3^* = t_6^* = \frac{1+\rho}{1-\rho}u, \qquad t_5^* = (\sqrt{u}+\rho\sqrt{r})^2.$$

The smallest t^* is among $\{t_1^*, t_2^*, t_4^*\}$. Since $\mu^{(2)}$ is in the positive orthant, the intersection point of the ellipsoid with line 4 must be outside the rejection region, so we further restrict to t_1^* and t_2^*. The ellipsoid intersects with line 1 at $(\rho\sqrt{r}+(1-\rho)\sqrt{u},\ \sqrt{r})'$. This point is on the boundary of $\mathcal{R}_u$ if and only if its second coordinate is $\geq \sqrt{u}$ (see Figure 12), i.e., $u \leq r$. The ellipsoid intersects with line 2 at $(\sqrt{u},\ \rho\sqrt{u}+(1-\rho^2)\sqrt{r})'$. This point is on the boundary of $\mathcal{R}_u$ if and only if its second coordinate is $\leq \sqrt{u}$ (see Figure 12), i.e., $u \geq (1+\rho)^2 r$. In the range of $r < u < (1+\rho)^2 r$, the ellipsoid intersects with $\mathcal{R}_u$ at the corner point $(\sqrt{u}, \sqrt{u})'$, with the corresponding

$$t^* = r + \frac{2}{1+\rho}u - 2\sqrt{ru} = \begin{cases} \frac{1-\rho}{1+\rho}u + (\sqrt{u}-\sqrt{r})^2, \\ (\sqrt{u}-\rho\sqrt{r})^2 + \frac{1-\rho}{1+\rho}\big(\sqrt{u}-(1+\rho)\sqrt{r}\big)^2. \end{cases}$$

This t^* has two equivalent expressions. Comparing them with t_1^* and t_2^*, we can see that the smallest t^* is a continuous function of u, given (ρ, r). It follows that

$$\begin{aligned} &\inf_{h\in\mathcal{R}_u}\{(h-\mu^{(2)})'\Sigma^{-1}(h-\mu^{(2)})\} \\ =\ & \frac{1-\rho}{1+\rho}u + (\sqrt{u}-\sqrt{r})_+^2 - \frac{1-\rho}{1+\rho}\big(\sqrt{u}-(1+\rho)\sqrt{r}\big)_+^2. \end{aligned} \tag{D.7}$$

We plug (D.6) and (D.7) into (D.4). It gives the expression of $\mathrm{FP}_p(u)$ for $\rho \geq 0$.

We then look at the ellipsoid $\mathcal{E}(t;\mu^{(3)})$, with $\mu^{(3)} = (\sqrt{r}, \rho\sqrt{r})'$. Note that we now investigate its distance to the complement of $\mathcal{R}_u$. In order for $\mu^{(3)}$ to outside $\mathcal{R}_u^c$ (i.e., in the interior of $\mathcal{R}_u$), we require that $u < r$; furthermore, when $u < r$, the ellipsoid can only intersect with lines 1-2 (see Figure 12). Using the formula of t^* in the equation below (D.5), we have

$$t_1^* = \frac{1-\rho}{1+\rho}\big((1+\rho)\sqrt{r}-\sqrt{u}\big)^2, \qquad t_2^* = (\sqrt{r}-\sqrt{u})^2.$$

By (D.5), the ellipsoid intersects with line 1 at $\big(\sqrt{r}-(1-\rho)[(1+\rho)\sqrt{r}-\sqrt{u}],\ \rho\sqrt{r}\big)'$. To guarantee that this point is on the boundary of $\mathcal{R}_u$, we need its second coordinate to be $\geq \sqrt{u}$ (see Figure 12), i.e., $u \leq \rho^2 r$; furthermore, when $u > \rho^2 r$, it can be easily seen from Figure 12 that the ellipsoid must have already crossed line 2. By (D.5) again, the ellipsoid intersects with line 2 at $(\sqrt{u}, \rho\sqrt{u})'$. This point is always on the boundary of $\mathcal{R}_u$. It follows that

$$\inf_{h\in\mathcal{R}_u^c}\{(h-\mu^{(3)})'\Sigma^{-1}(h-\mu^{(3)})\} = \min\Big\{\frac{1-\rho}{1+\rho}\big((1+\rho)\sqrt{r}-\sqrt{u}\big)^2,\ (\sqrt{r}-\sqrt{u})_+^2\Big\}. \tag{D.8}$$

We then look at the ellipsoid $\mathcal{E}(t;\mu^{(4)})$, with $\mu^{(4)} = \big((1+\rho)\sqrt{r},(1+\rho)\sqrt{r}\big)'$. It follows from figure 12 that $\mu^{(4)}$ is in the interior of the ellipsoid if and only if $(1+\rho)\sqrt{r} > \sqrt{u}$. We restrict to $(1+\rho)\sqrt{r} > \sqrt{u}$. Then, this ellipsoid can only touch lines 1-2 first. The t^* values are

$$t_1^* = \frac{1-\rho}{1+\rho}\big((1+\rho)\sqrt{r}-\sqrt{u}\big)^2, \qquad t_2^* = \big((1+\rho)\sqrt{r}-\sqrt{u}\big)^2.$$

Since $t_1^* < t_2^*$, the ellipsoid touches line 1 first, at the intersection point $\big((1-\rho)\sqrt{u}+\rho(1+\rho)\sqrt{r},\ (1+\rho)\sqrt{r}\big)'$. In order for this point to be on the boundary of $\mathcal{R}_u$, we need that its second coordinate is $\geq \sqrt{u}$, which translates to $\sqrt{u} \leq (1+\rho)\sqrt{r}$. This is always true when $r > u$ and $\rho > 0$. It follows that

$$\inf_{h\in\mathcal{R}_u^c}\{(h-\mu^{(4)})'\Sigma^{-1}(h-\mu^{(4)})\} = \frac{1-\rho}{1+\rho}\big((1+\rho)\sqrt{r}-\sqrt{u}\big)_+^2. \tag{D.9}$$

We plug (D.8) and (D.9) into (D.4). It gives the expression of $\mathrm{FN}_p(u)$ for $\rho \geq 0$.

Next, we consider the case that $\rho < 0$. By Lemma 7.2, $\mathcal{R}_u(\rho)$ is a reflection of $\mathcal{R}_u(|\rho|)$ with respect to the x-axis. As a result, if we re-define $\hat{h} = (x_j'y,\ -x_{j+1}'y)/\sqrt{2\log(p)}$, then the rejection region becomes $\mathcal{R}_u(|\rho|)$, which has the same shape as that in Figure 12. At the same time, the distribution of $\hat{h}$ becomes

$$\hat{h} \sim \mathcal{N}_2\Big(\mu,\ \frac{1}{\log(p)}\Sigma\Big), \qquad \text{where} \quad \Sigma = \begin{bmatrix} 1 & |\rho| \\ |\rho| & 1 \end{bmatrix}.$$

The vector μ is equal to

$$\mu^{(1)} \equiv \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \mu^{(2)} \equiv \begin{bmatrix} -|\rho|\sqrt{r} \\ -\sqrt{r} \end{bmatrix}, \quad \mu^{(3)} \equiv \begin{bmatrix} \sqrt{r} \\ |\rho|\sqrt{r} \end{bmatrix}, \quad \mu^{(4)} \equiv \begin{bmatrix} (1-|\rho|)\sqrt{r} \\ -(1-|\rho|)\sqrt{r}) \end{bmatrix}, \tag{D.10}$$

when $(\beta_j,\beta_{j+1})'$ is $(0,0)'$, $(0,\tau_p)'$, $(\tau_p,0)'$, and $(\tau_p,\tau_p)'$, respectively. Therefore, the calculations are similar, except that the expressions of $\mu^{(1)}$ to $\mu^{(4)}$ have changed to (D.10).

Below, for a negative ρ, we calculate the exponents in (D.4) as follows: We pretend that $\rho > 0$ and calculate the exponents using the same $\mathcal{R}_u$ and Σ as before, with $\mu^{(1)}$ to $\mu^{(4)}$ replaced by those in (D.10). Finally, we replace ρ by $|\rho|$ in all four exponents.

We now pretend that $\rho > 0$. Then, for each ellipsoid $\mathcal{E}(t;\mu^{(k)})$, its intersection point with a line $h_1 + bh_2 = c$ still obeys the formula in (D.5), and the corresponding t_* values associated with line 1-line 6 are still the same as those in the equation below (D.5) (but the vector μ has changed). Comparing (D.10) with (D.3), we notice that $\mu^{(1)}$ and $\mu^{(3)}$ are unchanged. Therefore, the expressions of exponents in (D.6) and (D.8) are still correct. The current $\mu^{(2)}$ is a sign flip (on both x-axis and y-axis) of the $\mu^{(2)}$ in (D.3); also, it can be seen from Figure 12 that the rejection region remains unchanged subject to a sign flip. Therefore, the expression in (D.8) is also valid. We only need to re-calculate the exponent in (D.9). The current $\mu^{(4)}$ is in the 4-th orthant. It is in the interior of $\mathcal{R}_u$ only if $(1-\rho)\sqrt{r} > \sqrt{u}$, i.e., $u < (1-\rho)^2 r$. As we increase t, the ellipsoid $\mathcal{E}(t;\mu^{(4)})$ will first intersect with either line 2 or line 3. Using the formula of t^* in the equation below (D.5), we have

$$t_2^* = \big(\sqrt{u}-(1-\rho)\sqrt{r}\big)^2, \qquad t_3^* = \frac{1+\rho}{1-\rho}\big((1-\rho)\sqrt{r}-\sqrt{u}\big)^2.$$

While t_2^* is the smaller one, the intersection point of the ellipsoid with line 2 is $(\sqrt{u}, -(1-\rho)\sqrt{r})'$, which by Figure 12 is in the interior of $\mathcal{R}_u$. Hence, the ellipsoid hits line 3 first. We conclude that

$$\inf_{h\in\mathcal{R}_u^c}\{(h-\mu^{(4)})'\Sigma^{-1}(h-\mu^{(4)})\} = \frac{1+\rho}{1-\rho}\big((1-\rho)\sqrt{r}-\sqrt{u}\big)_+^2. \tag{D.11}$$

Finally, we plug (D.6), (D.7), (D.8) and (D.11) into (D.4), and then change ρ to $|\rho|$. This gives the expressions of $\mathrm{FP}_p(u)$ and $\mathrm{FN}_p(u)$ for a negative ρ.

Appendix E. Proof of Theorem 5.2

We assume $\rho \geq 1/2$ throughout the proof. The calculation for the case where $\rho \leq -1/2$ is similar. By the design of the gram matrix X^TX and the construction of the knockoff variables, we know Lasso regression problem with $2p$ variables can be reduced to $(p/2)$ independent four-variate Lasso regression problems:

$$(\hat\beta_j, \hat\beta_{j+1}, \hat\beta_{j+p}, \hat\beta_{j+p+1})(\lambda) = \mathrm{argmin}_b\Big\{\frac{1}{2}||y-(x_j, x_{j+1}, \tilde x_j, \tilde x_{j+1})b||_2^2 + \lambda||b||_1\Big\} \tag{E.1}$$

for $j = 1, 3, \cdots, p-1$. By taking the sub-gradients of the objective function in (E.1), we know $(\hat\beta_j, \hat\beta_{j+1}, \hat\beta_{j+p}, \hat\beta_{j+p+1})$ should satisfy:

$$\begin{aligned}(\hat\beta_j, \hat\beta_{j+1}, \hat\beta_{j+p}, \hat\beta_{j+p+1})G + \lambda(\mathrm{sgn}(\hat\beta_j), \mathrm{sgn}(\hat\beta_{j+1}), \mathrm{sgn}(\hat\beta_{j+p}), \mathrm{sgn}(\hat\beta_{j+p+1}))\\ = (y^Tx_j, y^Tx_{j+1}, y^T\tilde x_j, y^T\tilde x_{j+1})\end{aligned} \tag{E.2}$$

where $G = ((1,\rho,2\rho-1,\rho)^T, (\rho,1,\rho,2\rho-1)^T, (2\rho-1,\rho,1,\rho)^T, (\rho,2\rho-1,\rho,1)^T)$ and $\mathrm{sgn}(x) = 1$ if $x>0$; -1 if $x<0$; any value in $[-1,1]$ if $x=0$. We have choose the correlation between a true variable and its knockoff to be $2\rho-1$, which is the smallest value such that $(X,\tilde X)^T(X,\tilde X)$ is semi-positive definite. In this case, G is degenerated and has rank 3. As λ is decreasing from infinity, we recognize that the first two variables (assume these two features are linear independent) entering the model will not leave before the third variable enters the model, which is obviously true from the close form solution of the bi-variate Lasso problem. We then show that the first two variables enter the Lasso path, individually. Furthermore, if the first two variables are a true variable and its knockoff variable, then the third and fourth variable enter the Lasso path simultaneously.

Since $(y^Tx_j, y^Tx_{j+1}, y^T\tilde x_j, y^T\tilde x_{j+1})^T \sim \mathcal{N}(G(\beta_j,\beta_{j+1},0,0)^T, G)$ is a degenerated normal random variable, we reparametrize it as $(m+d_1, m+d_2, m-d_1, m-d_2)$ with $(m,d_1,d_2)^T \sim \mathcal{N}((\rho\beta_j+\rho\beta_{j+1}, (1-\rho)\beta_j, (1-\rho)\beta_{j+1})^T, \mathrm{diag}(\rho, 1-\rho, 1-\rho))$. We intend to give the Lasso solution path (or $Z_j, \tilde Z_j$) as a function of m, d_1 and d_2. We only present the result in the case where $d_1 > d_2 > 0$. Results from other cases are immediate by permuting the rows in equation set (E.2) and transforming to the $d_1 > d_2 > 0$ case. Lasso solution path are obtained by the KKT condition (E.2) and summarized in the table below.

Here we explain the third row of the table as an example, $b_1 = (\epsilon, 0, 0, 0)^T$ is a solution of the KKT condition (E.2) when $\lambda = m + d_1 - \epsilon$ for $\epsilon \in (0, \frac{m}{\rho}]$, so sign_1 is expressed as $(0^+, 0, 0, 0)$. By property of the Lasso solution, if b_1 and b_2 are both Lasso solutions, then $G(b_1-b_2) = 0$ and $||b_1||_1 = ||b_2||_1$. $G(b_1-b_2)=0$ implies $b_1 - b_2 = \delta \times (1,-1,1,-1)^T$

range of m	λ_1	sign_1	λ_2	sign_2	λ_3	sign_3
$(-\infty, \frac{\rho}{1-\rho}(d_2-d_1))$	$-m+d_1$	$(0,0,0^-,0)$	$-m-\frac{\rho}{1-\rho}d_1+\frac{1}{1-\rho}d_2$	$(0,0,-,0^-)$		
$(\frac{\rho}{1-\rho}(d_2-d_1),0)$	$-m+d_1$	$(0,0,0^-,0)$	$\frac{1-\rho}{\rho}m+d_1$	$(0^+,0,-,0)$	d_2	$(+,0^+,-,0^-)$
$(0,\frac{\rho}{1-\rho}(d_1-d_2))$	$m+d_1$	$(0^+,0,0,0)$	$\frac{\rho-1}{\rho}m+d_1$	$(+,0,0^-,0)$	d_2	$(+,0^+,-,0^-)$
$(\frac{\rho}{1-\rho}(d_1-d_2),\infty)$	$m+d_1$	$(0^+,0,0,0)$	$m-\frac{\rho}{1-\rho}d_1+\frac{1}{1-\rho}d_2$	$(+,0^+,0,0)$		

Table 2: Summary of solution path of the Lasso problem (E.1). λ_i record the critical value of λ where a new variable enters the model and sign_i records the sign and the limiting behavior of $(\hat{\beta}_j, \hat{\beta}_{j+p})$ as $\lambda \to \lambda_i^-$. Value of λ_3 is omitted in row 1 and 4 since it will not affect the value of W_j and W_{j+1}.

for some $\delta \neq 0$. Therefore, $b_2 = (\epsilon-\delta, \delta, -\delta, \delta)^T$ and $||b_2||_1 \geq ||b_1||_1 + 2|\delta|$. This means the Lasso solution is unique with $\lambda = m + d_1 - \epsilon$ and variable 1 is the only one entering the model when λ gets below λ_1. When $\lambda = \frac{\rho-1}{\rho}m + d_1 - \epsilon$ for $\epsilon \in (0, \frac{\rho-1}{\rho}m + d_1 - d_2]$, $b_1 = (\frac{m}{\rho} + \frac{\epsilon}{2-2\rho}, 0, -\frac{\epsilon}{2-2\rho}, 0)^T$ is a solution of the KKT conditions. If there is another Lasso solution b_2, then $b_2 = (\frac{m}{\rho} + \frac{\epsilon}{2-2\rho} - \delta, \delta, -\frac{\epsilon}{2-2\rho} - \delta, \delta)^T$ and $||b_2||_1 \geq ||b_1||_1 + 2|\delta|$. So b_2 does not exist and variable 3 is the only one entering the model when λ gets below λ_2. When $\lambda = d_2 - \epsilon$ for sufficient small positive ϵ, $b_1 = (\frac{m}{2\rho} + \frac{d_1}{2-2\rho}, \frac{\epsilon}{2-2\rho}, \frac{m}{2\rho} - \frac{d_1}{2-2\rho}, -\frac{\epsilon}{2-2\rho})^T$ satisfies the KKT condition, thus variable 2 and 4 enters the model simultaneously. At this point, the Lasso solution is not unique and all solutions can be expressed as $b_1 - \delta \times (1,-1,1,-1)^T$ with $\delta \in [-\frac{\epsilon}{2-2\rho}, \frac{\epsilon}{2-2\rho}]$. Other rows from the table can be analyzed similarly.

Table E implicitly expresses $Z_j, Z_{j+1}, \tilde{Z}_j$ and $\tilde{Z}_{j+1}$ as a function of d_1, d_2 and m. By examining all possible ordinal relationship of d_1, d_2 and 0, we record the region in the space of (d_1, d_2, m) such that $\hat{\beta}_j(u) > 0$ and denote it as $R(u)$. $R(u)$ is the union of 4 disjoint sub-regions $\{R_i(u)\}_{i=1,\cdots,4}$, defined as following:

$$
\begin{aligned}
R_1(u) =& \{(x,y,z) : x>0, y>0, x>y, z>0, x+z>T\} \\
&\cup \frac{1}{2}\{(x,y,z) : x>0, y>0, x<y, z<0, z>x-y, x>T\} \\
&\cup \frac{1}{2}\{(x,y,z) : x>0, y>0, x<y, z>0, z<\frac{\rho}{1-\rho}(y-x), x>T\} \\
&\cup \{(x,y,z) : x>0, y>0, x<y, z>0, z>\max\big(\frac{\rho}{1-\rho}(y-x), T+\frac{\rho}{1-\rho}y-\frac{1}{1-\rho}x\big)\},
\end{aligned} \tag{E.3}
$$

$R_2(u) = \{(x,y,z) : (-x,y,-z) \in R_1(u)\}$, $R_3(u) = \{(x,y,z) : (x,-y,z) \in R_1(u)\}$ and $R_4(u) = \{(x,y,z) : (-x,-y,-z) \in R_1(u)\}$, where $T = \sqrt{2u\log(p)}$ and the $\frac{1}{2}$ ahead of a certain region means when (d_1, d_2, m) is in this region, $\hat{\beta}_j(u) > 0$ happens with 1/2 probability. Let the four disjoint regions that composes $R_1(u)$ in (E.3) be denoted by $R_{1,j}(u)$ for $j = 1, \cdots, 4$. We can similarly define $R_{i,j}(u)$ for $i = 2, 3, 4$. By Lemma 1, as

$p \to \infty$,

$$\begin{aligned}\mathbb{P}(\beta_j = 0, \hat{\beta}_j(u) \neq 0) =& \mathbb{P}(\hat{\beta}_j(u) \neq 0 | \beta_j = 0, \beta_{j+1} = 0) \times \mathbb{P}(\beta_j = 0, \beta_{j+1} = 0) \\ &+ \mathbb{P}(\hat{\beta}_j(u) \neq 0 | \beta_j = 0, \beta_{j+1} = \tau_p) \times \mathbb{P}(\beta_j = 0, \beta_{j+1} = \tau_p) \\ =& L_p p^{-\inf_{R(u)}[(z^2/\rho + x^2/(1-\rho) + y^2/(1-\rho))/(2\log(p))]} \\ &+ L_p p^{-\vartheta - \inf_{R(u)}[((z-\rho\tau_p)^2/\rho + x^2/(1-\rho) + (y-(1-\rho)\tau_p)^2/(1-\rho))/(2\log(p))]},\end{aligned} \tag{E.4}$$

$$\begin{aligned}\mathbb{P}(\beta_j \neq 0, \hat{\beta}_j(u) = 0) =& \mathbb{P}(\hat{\beta}_j(u) = 0 | \beta_j = \tau_p, \beta_{j+1} = 0) \times \mathbb{P}(\beta_j = \tau_p, \beta_{j+1} = 0) \\ &+ \mathbb{P}(\hat{\beta}_j(u) = 0 | \beta_j = \tau_p, \beta_{j+1} = \tau_p) \times \mathbb{P}(\beta_j = \tau_p, \beta_{j+1} = \tau_p) \\ =& L_p p^{-\vartheta - \inf_{R(u)^C}[((z-\rho\tau_p)^2/\rho + (x-(1-\rho)\tau_p)^2/(1-\rho) + y^2/(1-\rho))/(2\log(p))]} \\ &+ L_p p^{-2\vartheta - \inf_{R(u)^C}[((z-2\rho\tau_p)^2/\rho + (x-(1-\rho)\tau_p)^2/(1-\rho) + (y-(1-\rho)\tau_p)^2/(1-\rho))/(2\log(p))]}.\end{aligned} \tag{E.5}$$

Define the ρ-distance function of two sets A and B in $\mathbb{R}^3$ as

$$d_\rho(A, B) = \inf_{a \in A, b \in B}[(a_1 - b_1)^2/(1-\rho) + (a_2 - b_2)^2/(1-\rho) + (a_3 - b_3)^2/\rho]$$

where a_k, b_k denote the k-th coordinate of vector a and b. An immediate property of the ρ-distance function would be

$$d_\rho(\cup_{i=1,\cdots,M} A_i, \cup_{j=1,\cdots,N} B_j) = \min_{i,j} d_\rho(A_i, B_j).$$

Utilizing the symmetry of the regions, we can compute the region distances involved in (E.4) and (E.5) explicitly. Take the second exponent in (E.4) as an example, it can be simplified as

$$\begin{aligned}&-\vartheta - d_\rho(R(u), \{(0, (1-\rho)\tau_p, \rho\tau_p)\})/(2\log(p)) \\ =& -\vartheta - d_\rho(R_1(u) \cup R_2(u) \cup R_3(u) \cup R_4(u), \{(0, (1-\rho)\tau_p, \rho\tau_p)\})/(2\log(p)) \\ =& -\vartheta - d_\rho(R_{1,1}(u) \cup R_{1,3}(u) \cup R_{1,4}(u) \cup R_{2,2}(u), \{(0, (1-\rho)\tau_p, \rho\tau_p)\}))/(2\log(p)).\end{aligned}$$

Define $\tilde{R}_{1,2}(u) = \{(x,y,z) : x > 0, y > 0, z > 0, x < y, x > T, z < y - x\}$, $\tilde{R}_{1,3}(u) = \{(x,y,z) : x > 0, y > 0, z > 0, x < y, x > T\}$ and $\tilde{R}_{1,4}(u) = \{(x,y,z) : x > 0, y > 0, z > 0, x < y, x < T, z > T + \frac{\rho}{1-\rho}y - \frac{1}{1-\rho}x\}$. Then $\tilde{R}_{1,2}(u) \subset \tilde{R}_{1,3}(u)$ and $R_{1,3}(u) \cup R_{1,4}(u) = \tilde{R}_{1,3}(u) \cup \tilde{R}_{1,4}(u)$. Since $\tilde{R}_{1,2}$(u) and $R_{2,2}(u)$ are symmetric about the plane $x = 0$, we know

$$d_\rho(R_{2,2}(u), \{(0, (1-\rho)\tau_p, \rho\tau_p)\}) = d_\rho(\tilde{R}_{1,2}(u), \{(0, (1-\rho)\tau_p, \rho\tau_p)\}).$$

Therefore,

$$\begin{aligned}
&d_\rho(R(u), \{(0, (1-\rho)\tau_p, \rho\tau_p)\}) \\
&= \min\{d_\rho(R_{1,1}(u), \{(0, (1-\rho)\tau_p, \rho\tau_p)\}), d_\rho(\tilde{R}_{1,3}(u), \{(0, (1-\rho)\tau_p, \rho\tau_p)\}), \\
&\qquad d_\rho(\tilde{R}_{1,4}(u), \{(0, (1-\rho)\tau_p, \rho\tau_p)\})\} \\
&= \min\Big\{\frac{1-\rho}{2} \times \tau_p^2 + \frac{2}{1+\rho} \times [(T - (1+\rho)\tau_p/2)_+]^2 - \frac{1-\rho}{1+\rho} \times [(T-(1+\rho)\tau_p)_+]^2, \\
&\qquad \frac{1}{1-\rho} \times T^2 + \frac{1}{1-\rho} \times [(T-(1-\rho)\tau_p)_+]^2, \ \ \frac{1-\rho}{1+\rho} \times T^2 + ((T-\tau_p)_+)^2\Big\} \\
&= (T - \rho\tau_p)^2 + (\xi_\rho \tau_p - \eta_\rho T)_+^2 - (\tau_p - T)_+^2,
\end{aligned}$$

where $\xi_\rho = \sqrt{1-\rho^2}$ and $\eta_\rho = \sqrt{(1-\rho)/(1+\rho)}$.

Let $\tau_p = 0$, we know $d_\rho(R(u), \{(0,0,0)\}) = T^2$. By (E.4) we immediately have

$$\mathbb{P}(\beta_j = 0, \hat{\beta}_j(u) \neq 0) = L_p p^{-\min\left\{u,\ \vartheta + (\sqrt{u} - \rho\sqrt{r})^2 + (\xi_\rho\sqrt{r} - \eta_\rho\sqrt{u})_+^2 - (\sqrt{r} - \sqrt{u})_+^2\right\}}. \tag{E.6}$$

We can see the false positive rate is exactly the same when using the Lasso filter and the Knockoff filter when $\rho > 0$. For $\rho \geq 1/2$, we can similarly compute $d_\rho(R(u)^C, \{((1-\rho)\tau_p, 0, \rho\tau_p)\})$ to be

$$[(\tau_p - T)_+ - ((1-\xi_\rho)\tau_p - (1-\eta_\rho)T)_+ - (\lambda_\rho\tau_p - \eta_\rho\ T)_+]^2,$$

and $d_\rho(R(u)^C, \{((1-\rho)\tau_p, (1-\rho)\tau_p, 2\rho\tau_p)\})$ to be

$$[(\xi_\rho\ \tau_p - \eta_\rho T)_+ - (\lambda_\rho\tau_p - \eta_\rho T)_+]^2,$$

where $\xi_\rho = \sqrt{1-\rho^2}$, $\eta_\rho = \sqrt{(1-\rho)/(1+\rho)}$, and $\lambda_\rho = \sqrt{1-\rho^2} - \sqrt{1-\rho}$.

Plug these results in to (E.5), we have

$$\mathbb{P}(\beta_j \neq 0, \hat{\beta}_j(u) = 0) = L_p p^{-\vartheta - \left\{(\sqrt{r} - \sqrt{u})_+ - [(1-\xi_\rho)\sqrt{r} - (1-\eta_\rho)\sqrt{u}]_+ - (\lambda_\rho\sqrt{r} - \eta_\rho\sqrt{u})_+\right\}^2}. \tag{E.7}$$

From here we have prove the result for $\rho \geq 1/2$ case.

In the case where $\rho \leq -1/2$, the exponent of false negative rate is additionally lower bounded by -2ϑ. One can verify the rate given in the theorem through similar calculations. This is somehow more straight forwards since in the case where $\beta_j = \beta_{j+1} = \tau$, $(y^T x_j, y^T x_{j+1}, y^T \tilde{x}_j, y^T \tilde{x}_{j+1})^T \sim \mathcal{N}((1+\rho)\tau \cdot (1,1,-1,-1)^T, G)$, meaning there is no way to distinguish the true variable from its knockoff variable.

Appendix F. Proof of Theorem 5.3

In the following proofs, we only consider $\rho \geq 0$ case, since $\rho < 0$ case can be transformed to the positive $|\rho|$ case by flipping the sign of either β_j or β_{j+1} for $j = 1, 3, \cdots, p-1$. By the block diagonal structure of the gram matrix, the Lasso problem with $2p$ features can be reduced to $(p/2)$ independent four-variate Lasso regression problems:

$$\hat{b}(\lambda) = \text{argmin}_b\Big\{\frac{1}{2}||y - (x_j, x_{j+1}, \tilde{x}_j, \tilde{x}_{j+1})b||_2^2 + \lambda||b||_1\Big\} \tag{F.1}$$

for $j = 1, 3, \cdots, p-1$. Before we turn to the proof of the theorem, we first analysis the solution path of the following four-variate Lasso problem:

$$\hat{b} = \text{argmin}_b\big\{-h^T b + b^T B b/2 + \lambda\|b\|_1\big\}. \tag{F.2}$$

with $B = ((1,\rho,a,\rho)^T, (\rho,1,\rho,a)^T, (a,\rho,1,\rho)^T, (\rho,a,\rho,1)^T)$ and $a \in [2|\rho|-1, 1]$. By taking the sub-gradients, we know $\hat{b}$ should satisfy

$$B\ \hat{b} + \lambda\ \text{sgn}(\hat{b}) = h. \tag{F.3}$$

Let $\hat{b}_i$ and h_i denotes the i-th coordinate of $\hat{b}$ and h. Let $\lambda_1 > \lambda_2 > \lambda_3 > \lambda_4$ be the values at which variables enter the solution path. As discussed in the proof of Lemma 7.2, $\lambda_1 = \max\{|h_1|, |h_2|, |h_3|, |h_4|\}$. Without loss of generality, assume $\lambda_1 = |h_1|$ and variable 1 is the first variable entering the model in solution path. We know for one variate Lasso problem, the only feature will not leave the model after its entry as λ is decreasing. So in the four-variate Lasso (F.2), variable 1 will stay in the model until the second variable enters the model. Consider three bi-variate Lasso problems ($k = 2, 3, 4$):

$$\hat{b}^{(k)} = \text{argmin}_{b^{(k)}}\big\{-(h^{(k)})^T b^{(k)} + (b^{(k)})^T B^{(k)} b^{(k)}/2 + \lambda\|b^{(k)}\|_1\big\} \tag{F.4}$$

with

$$B^{(2)} = B^{(4)} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \qquad \text{and} \qquad B^{(3)} = \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix},$$

$h^{(2)} = (h_1, h_2)$, $h^{(3)} = (h_1, h_3)$ and $h^{(4)} = (h_1, h_4)$. Now, we claim $\lambda_2 = \max_k\{\lambda_2^{(k)}\}$ where $\lambda_2^{(k)}$ is the value at which the second variables enter the solution path in the k-th bi-variate Lasso problems. Suppose $\lambda_2^{(i)} > \lambda_2^{(k)}$ for $i \neq k \in \{2,3,4\}$, when $\lambda \in [\lambda_2^{(i)}, \lambda_1]$, we know the KKT condition (F.3) is satisfied with $h_2 = h_3 = h_4 = 0$ by looking at the KKT conditions of the bi-variate Lasso problems. When $\lambda \in [\lambda_2^{(i)} - \epsilon, \lambda_2^{(i)})$, a second variable i must have entered the four-variate Lasso path, since the objective function of (F.2) is smaller when including variable 1 and i than including variable 1 alone (this is because the second variable have entered the model in the i-th bi-variate Lasso path when $\lambda \in [\lambda_2^{(i)} - \epsilon, \lambda_2^{(i)})$). We are ready to prove the theorem now, using what we have shown regarding λ_1 and λ_2. We next compute the false positive rate and false negative rate given $(\beta_j, \beta_{j+1}) = (0,0), (0,\tau_p), (\tau_p, 0), (\tau_p, \tau_p), (-\tau_p, \tau_p)$ by deriving upper and lower bounds for those rates.

We first establish some noatations. For the four-variate Lasso problem (F.1), let A_i denotes the event that variable i is the first one entering the model, A_{i_1,i_2} denotes the event that variable i_1 and i_2 are the first two entering the model (ignoring the order between i_1 and i_2) and $A_{i_1 \to i_2}$ denotes the event that variable i_1 is the first one and variable i_2 is the second one entering the model. Let L_{i_1,i_2} denote the bi-variate Lasso problem with y as the response and x_{i_1}, x_{i_2} as the variables. Let $h \equiv (y^T x_j, y^T x_{j+1}, y^T \tilde{x}_j, y^T \tilde{x}_{j+1})$, then $h \sim \mathcal{N}(\mu, G)$ with $\mu = G(\beta_j, \beta_{j+1}, 0, 0)^T$ and $G = ((1,\rho,0,\rho)^T, (\rho,1,\rho,0)^T, (0,\rho,1,\rho)^T, (\rho,0,\rho,1)^T)$. When not causing any confusing, we write t_p in place of $t_p(u)$ for simplicity.

- When $(\beta_j, \beta_{j+1}) = (0,0)$,

$$\mathbb{P}\big\{W_j > t_p \big| (\beta_j, \beta_{j+1}) = (0,0)\big\} = L_p p^{-u}. \tag{F.5}$$

To derive a lower bound for $\mathbb{P}\big\{W_j > t_p \big| (\beta_j, \beta_{j+1}) = (0,0)\big\}$, we look for a point in the region (or on the boundary of the region) that choose variable j as a signal and apply Lemma 7.1. The point we choose is $p_1 = (t_p, \rho t_p, 0, \rho t_p)^T$ where $t_p = \sqrt{2u\log(p)}$. It's obvious that when $h = p_1$, variable j is the first one entering the Lasso path. Though $h = p_1$ is in the rejection region, it is also on the boundary of the region that choose variable j as a signal because slight increasing the first coordinate will result in variable j being selected. Since $h \sim \mathcal{N}(\mu_1, G)$ with $\mu_1 = \mathbf{0}$, by Lemma 7.1,

$$\mathbb{P}\big\{W_j > t_p \big| (\beta_j, \beta_{j+1}) = (0,0)\big\} \geq L_p p^{-(p_1-\mu_1)^T G^{-1}(p_1-\mu_1)/2\log(p)} = L_p p^{-u}.$$

The upper bound is straight forward by considering the first variable-i entering the model and notice that $W_i \sim \mathcal{N}(0,1)$:

$$\begin{aligned}\mathbb{P}\big\{W_j > t_p \big| (\beta_j, \beta_{j+1}) = (0,0)\big\} &= \sum_i \mathbb{P}\big\{W_j > t_p, A_i \big| (\beta_j, \beta_{j+1}) = (0,0)\big\} \\ &\leq \sum_i \mathbb{P}\big\{W_i > t_p \big| (\beta_j, \beta_{j+1}) = (0,0)\big\} = L_p p^{-u}.\end{aligned} \tag{F.6}$$

- When $(\beta_j, \beta_{j+1}) = (0, \tau_p)$,

$$\mathbb{P}\big\{W_j > t_p \big| (\beta_j, \beta_{j+1}) = (0,\tau_p)\big\} \geq L_p p^{-(\sqrt{u}-\rho\sqrt{r})^2 - (\xi_\rho\sqrt{r} - \eta_\rho\sqrt{u})_+^2 + (\sqrt{r}-\sqrt{u})_+^2}, \tag{F.7}$$

$$\mathbb{P}\big\{W_j > t_p, A \big| (\beta_j, \beta_{j+1}) = (0,\tau_p)\big\} \leq L_p p^{-u} \tag{F.8}$$

for $A = A_{j+p+1\to j}, A_{j+1,j+p+1}$ and

$$\mathbb{P}\big\{W_j > t_p, A \big| (\beta_j, \beta_{j+1}) = (0,\tau_p)\big\} \leq L_p p^{-(\sqrt{u}-\rho\sqrt{r})^2 - (\xi_\rho\sqrt{r} - \eta_\rho\sqrt{u})_+^2 + (\sqrt{r}-\sqrt{u})_+^2} \tag{F.9}$$

for $A = A_{j,j+1}, A_{j,j+p}, A_{j\to j+p+1}$.

This time we choose

$$p_2^T = \begin{cases} (t_p, \rho t_p + (1-\rho^2)\tau_p, \rho\tau_p, \rho t_p - \rho^2\tau_p), & (1+\rho)\tau_p \leq t_p, \\ (t_p, t_p, \frac{\rho}{1+\rho}t_p, \frac{\rho}{1+\rho}t_p), & \tau_p \leq t_p < (1+\rho)\tau_p, \\ (t_p + \rho(\tau_p - t_p), \tau_p, \rho(\tau_p - t_p) + \frac{\rho}{1+\rho}t_p, \frac{\rho}{1+\rho}t_p), & t_p < \tau_p. \end{cases}$$

When $h = p_2$ and $t_p \geq \tau_p$, variable j is the first variable entering the four-variate Lasso path with $W_j = t_p$; when $h = p_2$ and $t_p < \tau_p$, variable $j+1$ is the first and j is the second variable entering the Lasso path with $W_j = t_p$ and $W_{j+1} = \tau_p$. $h = p_2$ is on the boundary of the region that chooses variable j as a signal. Since $h \sim \mathcal{N}(\mu_2, G)$ with $\mu_2 = (\rho\tau_p, \tau_p, \rho\tau_p, 0)^T$, by Lemma 7.1,

$$\begin{aligned}\mathbb{P}\big\{W_j > t_p \big| (\beta_j, \beta_{j+1}) = (0,\tau_p)\big\} &\geq L_p p^{-(p_2-\mu_2)^T G^{-1}(p_2-\mu_2)/2\log(p)} \\ &= L_p p^{-(\sqrt{u}-\rho\sqrt{r})^2 - (\xi_\rho\sqrt{r} - \eta_\rho\sqrt{u})_+^2 + (\sqrt{r}-\sqrt{u})_+^2}.\end{aligned}$$

When $A_{j+1,j+p+1}$ occurs, since by our argument on λ_1 and λ_2, Z_{j+1} and Z_{j+p+1} are the λ value at which the variables enter the solution path in the bi-variate Lasso problem $L_{j+1,j+p+1}$. Therefore, $Z_{j+1} = |y^T x_{j+1}|, Z_{j+p+1} = |y^T \tilde{x}_{j+1}|$. We notice that $Z_{j+p+1} > Z_j > t_p$ and marginally $y^T \tilde{x}_{j+1} \sim \mathcal{N}(0,1)$, so

$$\begin{aligned}&\mathbb{P}\big\{W_j > t_p, A_{j+1,j+p+1}\big|(\beta_j, \beta_{j+1}) = (0, \tau_p)\big\}\\ &\le \mathbb{P}\big\{|y^T \tilde{x}_{j+1}| > t_p\big|(\beta_j, \beta_{j+1}) = (0, \tau_p)\big\} = L_p p^{-u}.\end{aligned}$$

Above inequality also holds for $A_{j+p+1\to j}$ since if variable $j+p+1$ is the first entering the Lasso path, then we must have $|y^T \tilde{x}_{j+1}| = Z_{j+p+1} > Z_j > t_p$.

When any one of $A_{j,j+1}, A_{j,j+p}, A_{j\to j+p+1}$ occurs, it implies in the bi-variate Lasso problem $L_{j,j+1}$, the largest λ such that variable 1 enters the model for the first time is equal to W_j, thus larger than t_p. In other words, if variable j is a false positive using Knockoff for variable selection, then it is also a false positive when using bi-variate Lasso $L_{j,j+1}$. This means $\mathbb{P}\big\{W_j > t_p, A\big|(\beta_j, \beta_{j+1}) = (0, \tau_p)\big\}$ is upper bounded by the corresponding false positive rate of Lasso, which is $L_p p^{-(\sqrt{u}-\rho\sqrt{r})^2 - (\xi_\rho\sqrt{r} - \eta_\rho\sqrt{u})_+^2 + (\sqrt{r}-\sqrt{u})_+^2}$, for $A = A_{j,j+1}, A_{j,j+p}, A_{j\to j+p+1}$.

Since $A_{j+1,j+p}$ and $A_{j+p,j+p+1}$ can never occur when $W_j > 0$, (F.8) and (F.9) implies

$$\mathbb{P}\big\{W_j > t_p\big|(\beta_j, \beta_{j+1}) = (0, \tau_p)\big\} \le L_p p^{-\min\{u, (\sqrt{u}-\rho\sqrt{r})^2 + (\xi_\rho\sqrt{r} - \eta_\rho\sqrt{u})_+^2 - (\sqrt{r}-\sqrt{u})_+^2\}}. \tag{F.10}$$

Further coupled with (F.5) and (F.7), we have

$$\mathbb{P}\big\{W_j > t_p, \beta_j = 0\big\} = L_p p^{-\min\{u, \vartheta + (\sqrt{u}-\rho\sqrt{r})^2 + (\xi_\rho\sqrt{r} - \eta_\rho\sqrt{u})_+^2 - (\sqrt{r}-\sqrt{u})_+^2\}}. \tag{F.11}$$

- When $(\beta_j, \beta_{j+1}) = (\tau_p, 0)$,

$$\mathbb{P}\big\{W_j \le t_p\big|(\beta_j, \beta_{j+1}) = (\tau_p, 0)\big\} \ge L_p p^{-[(\sqrt{r}-\sqrt{u})_+]^2}, \tag{F.12}$$

and

$$\mathbb{P}\big\{W_j \le t_p\big|(\beta_j, \beta_{j+1}) = (\tau_p, 0)\big\} \le L_p p^{\vartheta - f^+_{\mathrm{Hamm}}(u,r,\vartheta)}. \tag{F.13}$$

Let $p_3 = (t_p, \rho t_p, 0, \rho t_p)^T$. when $h = p_3$, variable j is the first variable entering the Lasso path and p_3 is in the region of rejecting variable j as a signal. Since $h \sim \mathcal{N}(\mu_3, G)$ with $\mu_3 = (\tau_p, \rho\tau_p, 0, \rho\tau_p)^T$, by Lemma 7.1,

$$\begin{aligned}\mathbb{P}\big\{W_j \le t_p\big|(\beta_j, \beta_{j+1}) = (\tau_p, 0)\big\} &\ge L_p p^{-(p_3-\mu_3)^T G^{-1}(p_3-\mu_3)/2\log(p)}\\ &= L_p p^{-[(\sqrt{r}-\sqrt{u})_+]^2}.\end{aligned}$$

Before we prove (F.13), we first analysis $f^+_{\mathrm{Hamm}}(u,r,\vartheta)$. By simply calculation, we find the optimal value of u that maximize $f^+_{\mathrm{Hamm}}(u,r,\vartheta)$ given r, ϑ is

$$u^* = \begin{cases} \frac{1+\rho}{(\sqrt{1+\rho}+\sqrt{1-\rho})^2} r, & \vartheta \le \frac{2\rho}{(\sqrt{1+\rho}+\sqrt{1-\rho})^2} r,\\ \frac{(r+\vartheta)^2}{4r}, & \frac{2\rho}{(\sqrt{1+\rho}+\sqrt{1-\rho})^2} r \le \vartheta < r,\\ \vartheta, & r < \vartheta. \end{cases}$$

This implies $u^* \geq \frac{1+\rho}{(\sqrt{1+\rho}+\sqrt{1-\rho})^2} r$ regardless of the relationship of ϑ and r. Consider r, ϑ as fixed, $f^+_{\text{Hamm}}(r, u, \vartheta)$ as a function of u is monotonically non-decreasing in $[0, u^*]$ and monotonically non-increasing in $[u^*, \infty)$. $f^+_{\text{Hamm}}(r, \vartheta) = \vartheta + [(\sqrt{r} - \sqrt{u})_+ - ((1 - \xi_\rho)\sqrt{r} - (1 - \eta_\rho)\sqrt{u})_+]^2$ if and only if $u > u^*$. Since $(1 - \xi_\rho)\sqrt{r} - (1 - \eta_\rho)\sqrt{u^*} < 0$, $(1 - \xi_\rho)\sqrt{r} - (1 - \eta_\rho)\sqrt{u} < 0$ for all $u > u^*$, which implies $f^+_{\text{Hamm}}(r, \vartheta) = \vartheta + [(\sqrt{r} - \sqrt{u})_+]^2$ when $u > u^*$. Therefore,

$$f^+_{\text{Hamm}}(r, u, \vartheta) = \min\{u, \vartheta + (\sqrt{u} - |\rho|\sqrt{r})^2 + ((\xi_\rho\sqrt{r} - \eta_\rho\sqrt{u})_+)^2 - ((\sqrt{r} - \sqrt{u})_+)^2, \\ \vartheta + [(\sqrt{r} - \sqrt{u})_+]^2\}.$$

Now, we show that (F.13) holds for $u \geq u^*$. This would implies (F.13) for all $u \geq 0$, since the false negative rate $\mathbb{P}\big\{W_j \leq t_p(u)\big|(\beta_j, \beta_{j+1}) = (\tau_p, 0)\big\}$ is monotone non-decreasing with u, so for $u < u^*$, $\mathbb{P}\big\{W_j \leq t_p(u)\big|(\beta_j, \beta_{j+1}) = (\tau_p, 0)\big\} \leq \mathbb{P}\big\{W_j \leq t_p(u^*)\big|(\beta_j, \beta_{j+1}) = (\tau_p, 0)\big\} \leq L_p p^{\vartheta - f^+_{\text{Hamm}}(r, u^*, \vartheta)} \leq L_p p^{\vartheta - f^+_{\text{Hamm}}(r, u, \vartheta)}$.

Assume $u \geq u^*$, so $u \geq \frac{1+\rho}{(\sqrt{1+\rho}+\sqrt{1-\rho})^2} r$ and

$$-[(\sqrt{r} - \sqrt{u})_+]^2 \geq -\Big(\frac{\sqrt{1-\rho}}{\sqrt{1+\rho} + \sqrt{1-\rho}}\Big)^2 r \geq -(2 - \sqrt{3})(1 - \rho)r \geq -\frac{1-\rho}{2} r \geq -\frac{1}{2} r. \tag{F.14}$$

We next prove (F.13) by showing that

$$\mathbb{P}\big\{W_j \leq t_p, A\big|(\beta_j, \beta_{j+1}) = (\tau_p, 0)\big\} \leq L_p p^{-[(\sqrt{r} - \sqrt{u})_+]^2} \tag{F.15}$$

holds for $A = A_j, A_{j+1}, A_{j+p}, A_{j+p+1}$ and $u \geq u^*$. Respectively,

$$\mathbb{P}\big\{W_j \leq t_p, A_j\big|(\beta_j, \beta_{j+1}) = (\tau_p, 0)\big\} \leq \mathbb{P}\big\{|y^T x_j| \leq t_p\big|(\beta_j, \beta_{j+1}) = (\tau_p, 0)\big\} \\ = L_p p^{-[(\sqrt{r} - \sqrt{u})_+]^2},$$

and by symmetry and (F.14),

$$\mathbb{P}\big\{W_j \leq t_p, A_{j+1}\big|(\beta_j, \beta_{j+1}) = (\tau_p, 0)\big\} = \mathbb{P}\big\{W_j \leq t_p, A_{j+p+1}\big|(\beta_j, \beta_{j+1}) = (\tau_p, 0)\big\} \\ \leq \mathbb{P}\big\{|y^T x_j| \leq |y^T x_{j+p+1}|\big|(\beta_j, \beta_{j+1}) = (\tau_p, 0)\big\} \leq L_p p^{-\frac{1-\rho}{2} r} \leq L_p p^{-[(\sqrt{r} - \sqrt{u})_+]^2},$$

$$\mathbb{P}\big\{W_j \leq t_p, A_{j+p}\big|(\beta_j, \beta_{j+1}) = (\tau_p, 0)\big\} \leq \mathbb{P}\big\{|y^T x_j| \leq |y^T x_{j+p}|\big|(\beta_j, \beta_{j+1}) = (\tau_p, 0)\big\} \\ \leq L_p p^{-\frac{1}{2} r} \leq L_p p^{-[(\sqrt{r} - \sqrt{u})_+]^2}.$$

(F.13) is immediate by $[(\sqrt{r} - \sqrt{u})_+]^2 \geq f^+_{\text{Hamm}}(r, u, \vartheta) - \vartheta$.

- When $(\beta_j, \beta_{j+1}) = (\tau_p, \tau_p)$,

$$\mathbb{P}\big\{W_j \leq t_p\big|(\beta_j, \beta_{j+1}) = (\tau_p, \tau_p)\big\} \leq L_p p^{\vartheta - f^+_{\text{Hamm}}(u, r, \vartheta)}. \tag{F.16}$$

More precisely, we will prove that

$$\mathbb{P}\big\{W_j \leq t_p\big|(\beta_j, \beta_{j+1}) = (\tau_p, \tau_p)\big\} \leq L_p p^{-[(\sqrt{r} - \sqrt{u})_+]^2} \tag{F.17}$$

holds for $u \geq u^*$, thus implies (F.16). We prove (F.17) by showing

$$\mathbb{P}\big\{W_j \leq t_p, A\big|(\beta_j, \beta_{j+1}) = (\tau_p, \tau_p)\big\} \leq L_p p^{-[(\sqrt{r}-\sqrt{u})_+]^2} \tag{F.18}$$

holds for $A = A_j, A_{j+1\to j}, A_{j+1\to j+p}, A_{j+1\to j+p+1}, A_{j+p}, A_{j+p+1}$ and $u \geq u^*$, which cover all possibilities. Respectively,

$$\begin{aligned}\mathbb{P}\big\{W_j \leq t_p, A_j\big|(\beta_j, \beta_{j+1}) = (\tau_p, \tau_p)\big\} &\leq \mathbb{P}\big\{|y^T x_j| \leq t_p\big|(\beta_j, \beta_{j+1}) = (\tau_p, \tau_p)\big\} \\ &\leq L_p p^{-[((1+\rho)\sqrt{r}-\sqrt{u})_+]^2} \leq L_p p^{-[(\sqrt{r}-\sqrt{u})_+]^2},\end{aligned}$$

$$\begin{aligned}\mathbb{P}\big\{W_j \leq t_p, A_{j+p}\big|(\beta_j, \beta_{j+1}) = (\tau_p, \tau_p)\big\} &\leq \mathbb{P}\big\{|y^T x_j| \leq |y^T \tilde{x}_j|\big|(\beta_j, \beta_{j+1}) = (\tau_p, \tau_p)\big\} \\ &\leq L_p p^{-\frac{1}{2}r} \leq L_p p^{-[(\sqrt{r}-\sqrt{u})_+]^2},\end{aligned}$$

$$\begin{aligned}\mathbb{P}\big\{W_j \leq t_p, A_{j+p+1}\big|(\beta_j, \beta_{j+1}) = (\tau_p, \tau_p)\big\} &\leq \mathbb{P}\big\{|y^T x_j| \leq |y^T \tilde{x}_{j+1}|\big|(\beta_j, \beta_{j+1}) = (\tau_p, \tau_p)\big\} \\ &\leq L_p p^{-\frac{1}{2(1-\rho)}r} \leq L_p p^{-[(\sqrt{r}-\sqrt{u})_+]^2}.\end{aligned}$$

When $A_{j+1\to j}$ occurs, the bi-variate Lasso problem $L_{j,j+1}$ shares the same λ_1 and λ_2 with the four-variate Lasso problem. So variable j is a false negative when doing variable selection using the bi-variate Lasso $L_{j,j+1}$ given $W_j \leq t_p$, which implies $\mathbb{P}\big\{W_j \leq t_p, A_{j+1\to j}\big|(\beta_j, \beta_{j+1}) = (\tau_p, \tau_p)\big\}$ is upper bounded by the corresponding false negative rate of Lasso, which is $L_p p^{-(\xi_\rho\sqrt{r}-\eta_\rho\sqrt{u})_+^2} \leq L_p p^{-[(\sqrt{r}-\sqrt{u})_+]^2}$. The last inequality is equivalent to

$$(1-\sqrt{1-\rho^2})\sqrt{r} \leq \Big(1-\sqrt{\frac{1-\rho}{1+\rho}}\Big)\sqrt{u}.$$

By (F.14), the right hand side is no smaller than $\sqrt{r}$, thus no smaller than the left hand side.

When $A_{j+1\to j+p}$ occurs, we know variable $j+p$ instead of variable j is the second one entering the Lasso path. This means the λ_2 (the λ value when the second variable entering Lasso path) of the bi-variate Lasso problem $L_{j+1,j+p}$ is larger than the λ_2 of the bi-variate Lasso problem $L_{j,j+1}$. Since we have derived the explicit expression of λ_2 in bi-variate Lasso problems, when $y^T x_{j+1} \geq 0$, we must have

$$\max\{\frac{y^T x_j - \rho y^T x_{j+1}}{1-\rho}, \frac{y^T x_j - \rho y^T x_{j+1}}{-1-\rho}\} < \max\{\frac{y^T \tilde{x}_j - \rho y^T x_{j+1}}{1-\rho}, \frac{y^T \tilde{x}_j - \rho y^T x_{j+1}}{-1-\rho}\}.$$

Therefore, $A_{j+1\to j+p}$ implies one the three following events must occur:

$$y^T x_{j+1} < 0, \frac{y^T x_j - \rho y^T x_{j+1}}{1-\rho} < \frac{y^T \tilde{x}_j - \rho y^T x_{j+1}}{1-\rho}, \frac{y^T x_j - \rho y^T x_{j+1}}{1-\rho} < \frac{y^T \tilde{x}_j - \rho y^T x_{j+1}}{-1-\rho}$$

The probability of these three events given $(\beta_j, \beta_{j+1}) = (\tau_p, \tau_p)$ are $L_p p^{-(1+\rho)^2 r}$, $L_p p^{-\frac{r}{2}}$ and $L_p p^{-\frac{(1+2\rho)^2(1-\rho)}{2(1+\rho)}r}$, all of which are upper bounded by $L_p p^{-[(\sqrt{r}-\sqrt{u})_+]^2}$ when $u \geq u^*$.

When $A_{j+1\to j+p+1}$ occurs, the λ_2 of the bi-variate Lasso problem $L_{j+1,j+p+1}$ is larger than the λ_2 of the bi-variate Lasso problem $L_{j,j+1}$. When $y^T x_{j+1} \geq 0$, we must have

$$\max\{\frac{y^T x_j - \rho y^T x_{j+1}}{1-\rho}, \frac{y^T x_j - \rho y^T x_{j+1}}{-1-\rho}\} < |y^T \tilde{x}_{j+1}|.$$

Therefore, $A_{j+1\to j+p+1}$ implies one the three following events must occur:

$$y^T x_{j+1} < 0, \frac{y^T x_j - \rho y^T x_{j+1}}{1-\rho} < y^T \tilde{x}_{j+1}, \frac{y^T x_j - \rho y^T x_{j+1}}{1-\rho} < -y^T \tilde{x}_{j+1}.$$

Respectively, the probability of these three events are $L_p p^{-(1+\rho)^2 r}$, $L_p p^{-\frac{r}{2}}$ and $L_p p^{-\frac{(1+2\rho)^2(1-\rho)}{2(1+\rho)}r}$, all of which are upper bounded by $L_p p^{-[(\sqrt{r}-\sqrt{u})_+]^2}$ when $u \geq u^*$. From here we have verified (F.18), thus implies (F.16).

From (F.11) and (F.12), we have

$$\mathbb{P}\big\{W_j > t_p, \beta_j = 0\big\} + \mathbb{P}\big\{W_j \leq t_p, \beta_j = \tau_p\big\} \geq L_p p^{-f^+_{\text{Hamm}}(r,u,\vartheta)}. \tag{F.19}$$

$$\begin{aligned}\mathbb{P}\big\{W_j \leq t_p, \beta_j = \tau_p\big\} =& p^{-\vartheta} \times \mathbb{P}\big\{W_j \leq t_p \big| (\beta_j, \beta_{j+1}) = (\tau_p, 0)\big\} \\ &+ p^{-2\vartheta} \times \mathbb{P}\big\{W_j \leq t_p \big| (\beta_j, \beta_{j+1}) = (\tau_p, \tau_p)\big\} \\ \leq& L_p p^{-f^+_{\text{Hamm}}(r,u,\vartheta)}\end{aligned} \tag{F.20}$$

Since (F.11) also implies $\mathbb{P}\big\{W_j > t_p, \beta_j = 0\big\} \leq L_p p^{-f^+_{\text{Hamm}}(r,u,\vartheta)}$, we know

$$\mathbb{P}\big\{W_j > t_p, \beta_j = 0\big\} + \mathbb{P}\big\{W_j \leq t_p, \beta_j = \tau_p\big\} = L_p p^{-f^+_{\text{Hamm}}(r,u,\vartheta)}. \tag{F.21}$$

- When $(\beta_j, \beta_{j+1}) = (-\tau_p, \tau_p)$,

$$\mathbb{P}\big\{W_j \leq t_p \big| (\beta_j, \beta_{j+1}) = (-\tau_p, \tau_p)\big\} \geq L_p p^{-((\xi_\rho\sqrt{r} - \eta_\rho^{-1}\sqrt{u})_+)^2}, \tag{F.22}$$

$$\mathbb{P}\big\{W_j \leq t_p \big| (\beta_j, \beta_{j+1}) = (-\tau_p, \tau_p)\big\} \geq L_p p^{-\frac{(1-2\rho)^2(1+\rho)}{2(1-\rho)}r}, \tag{F.23}$$

and

$$\mathbb{P}\big\{W_j \leq t_p \big| (\beta_j, \beta_{j+1}) = (-\tau_p, \tau_p)\big\} \leq L_p p^{-\min\{((\xi_\rho\sqrt{r} - \eta_\rho^{-1}\sqrt{u})_+)^2, \frac{(1-2\rho)^2(1+\rho)}{2(1-\rho)}r, -\vartheta + f^+_{\text{Hamm}}(u,r,\vartheta)\}}. \tag{F.24}$$

Let

$$p_4^T = \begin{cases} (-(1-\rho)\tau_p, (1-\rho)\tau_p, \rho\tau_p, -\rho\tau_p), & (1-\rho)\tau_p \leq t_p, \\ (\rho(1-\rho)\tau_p - (1+\rho)t_p, (1-\rho)\tau_p, \rho(1-\rho)\tau_p + \frac{\rho^2}{1-\rho}t_p, -\frac{\rho}{1-\rho}t_p), & (1-\rho)\tau_p > t_p. \end{cases}$$

When $h = p_4$ and $(1-\rho)\tau_p \leq t_p$, variable j is the first variable entering the Lasso path with $W_j = (1-\rho)\tau_p \leq t_p$; when $h = p_4$ and $(1-\rho)\tau_p > t_p$, $j+1$ is the first and j is the second variable entering the Lasso path with $W_j = t_p$. Regardless of the

relationship between τ_p and t_p, $h = p_4$ is always in the region of rejecting j as a signal. Since $h \sim \mathcal{N}(\mu_4, G)$ with $\mu_4 = (-(1-\rho)\tau_p, (1-\rho)\tau_p, \rho\tau_p, -\rho\tau_p)^T$, by Lemma 7.1,

$$\begin{aligned}\mathbb{P}\big\{W_j \le t_p \big| (\beta_j, \beta_{j+1}) = (-\tau_p, \tau_p)\big\} &\ge L_p p^{-(p_4-\mu_4)^T G^{-1}(p_4-\mu_4)/2\log(p)} \\ &= L_p p^{-((\xi_\rho\sqrt{r} - \eta_\rho^{-1}\sqrt{u})_+)^2}.\end{aligned}$$

Let $p_5^T = \Big(\frac{4\rho^3-2\rho^2+\rho-1}{2(1-\rho)}\tau_p, (1-\rho)\tau_p, \frac{1+2\rho-4\rho^2}{2}\tau_p, -\frac{\rho^2}{1-\rho}\tau_p\Big)$. When $h = p_5$, variable $j+1$ is the first one entering the Lasso path with $W_{j+1} = (1-\rho)\tau_p$, if we slightly increase the value of the third coordinate of p_5, then it falls in the region of rejecting j as a signal since variable $j+p$ is the second variable entering the Lasso path. This implies $h = p_5$ in on the boundary of the region that rejects j as a signal, by Lemma 7.1,

$$\begin{aligned}\mathbb{P}\big\{W_j \le t_p \big| (\beta_j, \beta_{j+1}) = (-\tau_p, \tau_p)\big\} &\ge L_p p^{-(p_5-\mu_4)^T G^{-1}(p_5-\mu_4)/2\log(p)} \\ &= L_p p^{-\frac{(1-2\rho)^2(1+\rho)}{2(1-\rho)}r}.\end{aligned}$$

Next, we show that

$$\mathbb{P}\big\{W_j \le t_p, A \big| (\beta_j, \beta_{j+1}) = (-\tau_p, \tau_p)\big\} \le L_p p^{-\min\{((\xi_\rho\sqrt{r} - \eta_\rho^{-1}\sqrt{u})_+)^2, \frac{(1-2\rho)^2(1+\rho)}{2(1-\rho)}r, -\vartheta + f^+_{\mathrm{Hamm}}(u,r,\vartheta)\}}. \tag{F.25}$$

holds for $A = A_j, A_{j+1\to j}, A_{j+1\to j+p}, A_{j+1\to j+p+1}, A_{j+p}, A_{j+p+1}$, which cover all possibilities.

When $A = A_j$ or $A_{j+1\to j}$ occurs, as previously discussed, variable j is a false negative when doing variable selection using the bi-variate Lasso $L_{j,j+1}$ given $W_j \le t_p$, which implies $\mathbb{P}\big\{W_j \le t_p, A \big| (\beta_j, \beta_{j+1}) = (-\tau_p, \tau_p)\big\}$ is upper bounded by the corresponding false negative rate of Lasso, which is $L_p p^{-((\xi_\rho\sqrt{r} - \eta_\rho^{-1}\sqrt{u})_+)^2}$.

When $A_{j+1\to j+p}$ occurs, the λ_2 of the bi-variate Lasso problem $L_{j+1,j+p}$ is larger than the λ_2 of the bi-variate Lasso problem $L_{j,j+1}$. When $y^T x_{j+1} \ge 0$, we must have

$$\max\{\frac{y^T x_j - \rho y^T x_{j+1}}{1-\rho}, \frac{y^T x_j - \rho y^T x_{j+1}}{-1-\rho}\} < \max\{\frac{y^T \tilde{x}_j - \rho y^T x_{j+1}}{1-\rho}, \frac{y^T \tilde{x}_j - \rho y^T x_{j+1}}{-1-\rho}\}.$$

Therefore, $A_{j+1\to j+p}$ implies one of the three following events must occur:

$$y^T x_{j+1} < 0, \frac{y^T x_j - \rho y^T x_{j+1}}{-1-\rho} < \frac{y^T \tilde{x}_j - \rho y^T x_{j+1}}{1-\rho}, \frac{y^T x_j - \rho y^T x_{j+1}}{-1-\rho} < \frac{y^T \tilde{x}_j - \rho y^T x_{j+1}}{-1-\rho}$$

The probability of these three events are $L_p p^{-(1-\rho)^2 r}$, $L_p p^{-\frac{(1-2\rho)^2(1+\rho)}{2(1-\rho)}r}$ and $L_p p^{-\frac{r}{2}}$, all of which are upper bounded by $L_p p^{-\min\{\frac{(1-2\rho)^2(1+\rho)}{2(1-\rho)}r, -\vartheta + f^+_{\mathrm{Hamm}}(u,r,\vartheta)\}}$.

When $A_{j+1\to j+p+1}$ occurs, the λ_2 of the bi-variate Lasso problem $L_{j+1,j+p+1}$ is larger than the λ_2 of the bi-variate Lasso problem $L_{j,j+1}$. When $y^T x_{j+1} \ge 0$, we must have

$$\max\{\frac{y^T x_j - \rho y^T x_{j+1}}{1-\rho}, \frac{y^T x_j - \rho y^T x_{j+1}}{-1-\rho}\} < |y^T \tilde{x}_{j+1}|.$$

Therefore, $A_{j+1\to j+p+1}$ implies one of the three following events must occur:

$$y^T x_{j+1} < 0, \frac{y^T x_j - \rho y^T x_{j+1}}{-1-\rho} < y^T \tilde{x}_{j+1}, \frac{y^T x_j - \rho y^T x_{j+1}}{-1-\rho} < -y^T \tilde{x}_{j+1}.$$

The probability of these three events are $L_p p^{-(1-\rho)^2 r}$, $L_p p^{-\frac{r}{2}}$ and $L_p p^{-\frac{(1-2\rho)^2(1+\rho)}{2(1-\rho)} r}$, all of which are upper bounded by $L_p p^{-\min\{\frac{(1-2\rho)^2(1+\rho)}{2(1-\rho)} r, -\vartheta + f^+_{\text{Hamm}}(u,r,\vartheta)\}}$.

When A_{j+p} occurs, then $|y^T\tilde{x}_j| > |y^T x_j|$ and $|y^T\tilde{x}_j| > |y^T x_{j+1}|$. If $y^T\tilde{x}_j > 0$, we further have $(y^T\tilde{x}_j - y^T x_{j+1}) + \frac{1}{2\rho+1}(y^T\tilde{x}_j + y^T x_j) > 0$; if $y^T\tilde{x}_j \le 0$, we further have $y^T\tilde{x}_j + y^T x_{j+1} < 0$. Therefore,

$$\begin{aligned}
&\mathbb{P}\big\{W_j \le t_p, A_{j+p} \big| (\beta_j, \beta_{j+1}) = (-\tau_p, \tau_p)\big\} \\
\le &\mathbb{P}\big\{(y^T\tilde{x}_j - y^T x_{j+1}) + \frac{1}{2\rho+1}(y^T\tilde{x}_j + y^T x_j) > 0 \big| (\beta_j, \beta_{j+1}) = (-\tau_p, \tau_p)\big\} \\
&+ \mathbb{P}\big\{y^T\tilde{x}_j + y^T x_{j+1} < 0 \big| (\beta_j, \beta_{j+1}) = (-\tau_p, \tau_p)\big\} \\
\le &L_p p^{-\frac{2(1-2\rho)^2(1+\rho)}{3-4\rho^2} r} + L_p p^{-\frac{r}{2}} \le L_p p^{-\min\{\frac{(1-2\rho)^2(1+\rho)}{2(1-\rho)} r, -\vartheta + f^+_{\text{Hamm}}(u,r,\vartheta)\}}.
\end{aligned}$$

For $A = A_{j+p+1}$, (F.25) is immediate due to the symmetry between variable $j+p$ and $j+p+1$.

Now consider the case where β_j takes value in $\{0, -\tau_p\}$ and β_{j+1} takes value in $\{0, \tau_p\}$, this corresponds to the $\rho < 0$ case (we flipped the sign of ρ and β_j simultaneously). By (F.5), (F.7), (F.12), (F.22) and (F.23), we know

$$\begin{aligned}
&\mathbb{P}\big\{W_j > t_p, \beta_j = 0\big\} + \mathbb{P}\big\{W_j \le t_p, \beta_j = -\tau_p\big\} \\
\ge\ & L_p p^{-\min\{f^+_{\text{Hamm}}(u,r,\vartheta), 2\vartheta + ((\xi_\rho\sqrt{r} - \eta_\rho^{-1}\sqrt{u})_+)^2, 2\vartheta + \frac{(1-2|\rho|)^2(1+|\rho|)}{2(1-|\rho|)} r\}}.
\end{aligned} \tag{F.26}$$

Meanwhile, (F.5), (F.8), (F.9), (F.13) and (F.24) gives

$$\begin{aligned}
&\mathbb{P}\big\{W_j > t_p, \beta_j = 0\big\} + \mathbb{P}\big\{W_j \le t_p, \beta_j = -\tau_p\big\} \\
\le\ & L_p p^{-\min\{f^+_{\text{Hamm}}(u,r,\vartheta), 2\vartheta + ((\xi_\rho\sqrt{r} - \eta_\rho^{-1}\sqrt{u})_+)^2, 2\vartheta + \frac{(1-2|\rho|)^2(1+|\rho|)}{2(1-|\rho|)} r\}}.
\end{aligned} \tag{F.27}$$

Therefore,

$$\begin{aligned}
&\mathbb{P}\big\{W_j > t_p, \beta_j = 0\big\} + \mathbb{P}\big\{W_j \le t_p, \beta_j = -\tau_p\big\} \\
=\ & L_p p^{-\min\{f^+_{\text{Hamm}}(u,r,\vartheta), 2\vartheta + ((\xi_\rho\sqrt{r} - \eta_\rho^{-1}\sqrt{u})_+)^2, 2\vartheta + \frac{(1-2|\rho|)^2(1+|\rho|)}{2(1-|\rho|)} r\}}.
\end{aligned} \tag{F.28}$$

(F.21) and (F.28) complete the proof for Theorem 5.3.

Appendix G. Proof of Theorem 5.4

The only difference of the conditional knockoff from the Equal-correlated knockoff construction is that $x_j^T\tilde{x}_j$ is changed from 0 to ρ^2 for $j = 1, \cdots, p$. Therefore, $G = ((1, \rho, \rho^2, \rho)^T, (\rho, 1, \rho, \rho^2)^T, (\rho^2, \rho, 1, \rho)^T, (\rho, \rho^2, \rho, 1)^T)$ is the new gram matrix for the four-variate Lassos (F.1). We follow the same notations and workflow from the previous proof.

- When $(\beta_j, \beta_{j+1}) = (0, 0)$,

$$\mathbb{P}\big\{W_j > t_p \big| (\beta_j, \beta_{j+1}) = (0,0)\big\} = L_p p^{-u}. \tag{G.1}$$

Let $p_1 = (t_p, \rho t_p, \rho^2 t_p, \rho t_p)^T$ where $t_p = \sqrt{2u\log(p)}$. When $h = p_1$, variable j is the first one entering the Lasso path. Though $h = p_1$ is in the rejection region, it is also on the boundary of the region that choose variable j as a signal. Since $h \sim \mathcal{N}(\mu_1, G)$ with $\mu_1 = \mathbf{0}$, by Lemma 7.1,

$$\mathbb{P}\big\{W_j > t_p \big| (\beta_j, \beta_{j+1}) = (0,0)\big\} \geq L_p p^{-(p_1-\mu_1)^T G^{-1}(p_1-\mu_1)/2\log(p)} = L_p p^{-u}.$$

The upper bound is derived exactly the same as (F.6).

- When $(\beta_j, \beta_{j+1}) = (0, \tau_p)$,

$$\mathbb{P}\big\{W_j > t_p \big| (\beta_j, \beta_{j+1}) = (0,\tau_p)\big\} = L_p p^{-(\sqrt{u}-\rho\sqrt{r})^2 - (\xi_\rho\sqrt{r}-\eta_\rho\sqrt{u})_+^2 + (\sqrt{r}-\sqrt{u})_+^2}. \tag{G.2}$$

This time we choose

$$p_2^T = \begin{cases} (t_p, \rho t_p + (1-\rho^2)\tau_p, \rho^2 t_p + \rho(1-\rho^2)\tau_p, \rho t_p)^T, & (1+\rho)\tau_p \leq t_p, \\ (t_p, t_p, \rho t_p, \rho t_p)^T, & \tau_p \leq t_p < (1+\rho)\tau_p, \\ ((1-\rho)t_p + \rho\tau_p, \tau_p, \rho\tau_p, \rho(1-\rho)t_p + \rho^2\tau_p)^T, & t_p < \tau_p. \end{cases}$$

When $h = p_2$ and $t_p \geq \tau_p$, variable j is the first variable entering the four-variate Lasso path with $W_j = t_p$; when $h = p_2$ and $t_p < \tau_p$, variable $j+1$ is the first and j is the second variable entering the Lasso path with $W_j = t_p$ and $W_{j+1} = \tau_p$. $h = p_2$ is on the boundary of the region that chooses variable j as a signal. Since $h \sim \mathcal{N}(\mu_2, G)$ with $\mu_2 = (\rho\tau_p, \tau_p, \rho\tau_p, \rho^2\tau_p)^T$, by Lemma 7.1,

$$\begin{aligned} \mathbb{P}\big\{W_j > t_p \big| (\beta_j, \beta_{j+1}) = (0,\tau_p)\big\} &\geq L_p p^{-(p_2-\mu_2)^T G^{-1}(p_2-\mu_2)/2\log(p)} \\ &= L_p p^{-(\sqrt{u}-\rho\sqrt{r})^2 - (\xi_\rho\sqrt{r}-\eta_\rho\sqrt{u})_+^2 + (\sqrt{r}-\sqrt{u})_+^2}. \end{aligned} \tag{G.3}$$

Next we show that

$$\mathbb{P}\big\{W_j > t_p, A \big| (\beta_j, \beta_{j+1}) = (0,\tau_p)\big\} \leq L_p p^{-(\sqrt{u}-\rho\sqrt{r})^2 - (\xi_\rho\sqrt{r}-\eta_\rho\sqrt{u})_+^2 + (\sqrt{r}-\sqrt{u})_+^2} \tag{G.4}$$

holds for $A = A_{j,j+1}, A_{j,j+p}, A_{j\to j+p+1}, A_{j+p+1\to j}, A_{j+1,j+p+1}$, which covers all possibilities.

When any one of $A_{j,j+1}, A_{j,j+p}, A_{j\to j+p+1}$ occurs, same as for EC-knockoff, it implies if variable j is a false positive using Knockoff for variable selection, then it is also a false positive when using bi-variate Lasso $L_{j,j+1}$. So $\mathbb{P}\big\{W_j > t_p, A \big| (\beta_j, \beta_{j+1}) = (0,\tau_p)\big\}$ is upper bounded by the corresponding false positive rate of Lasso, which is $L_p p^{-(\sqrt{u}-\rho\sqrt{r})^2 - (\xi_\rho\sqrt{r}-\eta_\rho\sqrt{u})_+^2 + (\sqrt{r}-\sqrt{u})_+^2}$, for $A = A_{j,j+1}, A_{j,j+p}, A_{j\to j+p+1}$.

When $A = A_{j+p+1\to j}$, $j+p+1$ is the first variable entering the model in the four-variate Lasso problem, thus it's also the first variable entering the model in the bi-variate Lasso problem $L_{j+1,j+p+1}$ and $L_{j,j+p+1}$. Variable $j+p+1$ gets picked up as a signal in $L_{j+1,j+p+1}$ implies

$$\begin{aligned}\mathbb{P}\big\{W_j > t_p, A_{j+p+1\to j}\big|(\beta_j,\beta_{j+1}) = (0,\tau_p)\big\} &\le L_p p^{-(\sqrt{u}-|\rho^2|\sqrt{r})^2-(\xi_{\rho^2}\sqrt{r}-\eta_{\rho^2}\sqrt{u})_+^2+(\sqrt{r}-\sqrt{u})_+^2}\\ &\le L_p p^{-(\sqrt{u}-|\rho|\sqrt{r})^2-(\xi_{\rho}\sqrt{r}-\eta_{\rho}\sqrt{u})_+^2+(\sqrt{r}-\sqrt{u})_+^2}\end{aligned}$$

when $u \ge (1+\rho)^2 r$ or $u \le (1+\rho^2)^2 r$.

Now consider bi-variate Lasso problem $L_{j,j+p+1}$ given $(1+\rho^2)^2 r < u < (1+\rho)^2 r$. Variable $j, j+p+1$ both get picked up as signals with $j+p+1$ entering the model first given $W_j > t_p$. This implies $(y^T x_j, y^T \tilde{x}_{j+1})$ falls in the purple or green region of the right panel of Figure 12. Marginally, $(y^T x_j, y^T \tilde{x}_{j+1}) \sim \mathcal{N}((\rho\tau_p, \rho^2\tau_p)^T, [(1,\rho),(\rho,1)])$. The point in purple or green region that has the smallest ellipsoid distance to $(\rho\tau_p, \rho^2\tau_p)^T$ is (t_p, t_p) when $(1+\rho^2)^2 r < u < (1+\rho)^2 r$, thus by Lemma 7.1,

$$\begin{aligned}\mathbb{P}\big\{W_j > t_p, A_{j+p+1\to j}\big|(\beta_j,\beta_{j+1}) = (0,\tau_p)\big\} &\le L_p p^{-(\sqrt{u}-\rho\sqrt{r})^2-\frac{1-\rho}{1+\rho}u}\\ &\le L_p p^{-r+2\sqrt{ru}-\frac{2}{1+\rho}u}\\ &= L_p p^{-(\sqrt{u}-|\rho|\sqrt{r})^2-(\xi_{\rho}\sqrt{r}-\eta_{\rho}\sqrt{u})_+^2+(\sqrt{r}-\sqrt{u})_+^2}\end{aligned}$$

for $u \in ((1+\rho^2)^2 r, (1+\rho)^2 r)$, which completes the proof of (G.4) for $A = A_{j+p+1\to j}$.

When $A_{j+1,j+p+1}$ occurs, consider the bi-variate Lasso problem $L_{j+1,j+p+1}$. In this bi-variate Lasso problem, $\{\lambda_1, \lambda_2\} = \{Z_{j+1}, Z_{j+p+1}\}$, both of which are larger than W_j. Thus in this bi-variate Lasso problem, both variables will be picked up as signals given $W_j > t_p$. So $(y^T x_{j+1}, y^T \tilde{x}_{j+1})/\sqrt{2\log(p)}$ falls in one of the four regions in the right panel of Figure 12 (with $x_{j+1}^T \tilde{x}_{j+1} = \rho^2$ instead of ρ): the purple region, the mirror of purple region against $x = y$, the green region and the mirror of green region against $x = -y$. Since $(y^T x_{j+1}, y^T \tilde{x}_{j+1}) \sim \mathcal{N}((\tau_p, \rho^2\tau_p)^T, [(1,\rho^2),(\rho^2,1)])$. By Lemma 7.1, we need to find the point in those regions that has the smallest ellipsoid distance to the center-$(\tau_p, \rho^2\tau_p)^T$. When $\tau_p \le t_p$, this critical point is $(y^T x_{j+1}, y^T \tilde{x}_{j+1}) = (t_p, t_p)$; when $\tau_p > t_p$, this critical point is $(y^T x_{j+1}, y^T \tilde{x}_{j+1}) = (\tau_p, t_p + \rho(\tau_p - t_p))$. So Lemma 7.1 gives the probability for λ_1 and λ_2 in $L_{j+1,j+p+1}$ to be both larger than t_p is

$$L_p p^{-(\sqrt{u}-\sqrt{r})_+^2-\frac{1-\rho^2}{1+\rho^2}u} \le L_p p^{-(\sqrt{u}-|\rho|\sqrt{r})^2-(\xi_{\rho}\sqrt{r}-\eta_{\rho}\sqrt{u})_+^2+(\sqrt{r}-\sqrt{u})_+^2}.$$

Since $A_{j+1,j+p+1} \cap \{W_j > t_p\}$ implies $\{\lambda_1 > t_p\} \cap \{\lambda_2 > t_p\}$ in $L_{j+1,j+p+1}$, we know

$$\mathbb{P}\big\{W_j > t_p, A_{j+1,j+p+1}\big|(\beta_j,\beta_{j+1}) = (0,\tau_p)\big\} \le L_p p^{-(\sqrt{u}-|\rho|\sqrt{r})^2-(\xi_{\rho}\sqrt{r}-\eta_{\rho}\sqrt{u})_+^2+(\sqrt{r}-\sqrt{u})_+^2}.$$

Now, we have verified (G.4). Further coupled with (G.3), we have (G.2).

- When $(\beta_j, \beta_{j+1}) = (\tau_p, 0)$,

$$\mathbb{P}\big\{W_j \le t_p\big|(\beta_j,\beta_{j+1}) = (\tau_p,0)\big\} \ge L_p p^{-[(\sqrt{r}-\sqrt{u})_+]^2}, \tag{G.5}$$

and

$$\mathbb{P}\big\{W_j \le t_p \big| (\beta_j, \beta_{j+1}) = (\tau_p, 0)\big\} \le L_p p^{\vartheta - f^+_{\mathrm{Hamm}}(u,r,\vartheta)}. \tag{G.6}$$

Let $p_3 = (t_p, \rho t_p, \rho^2 t_p, \rho t_p)^T$. when $h = p_3$, variable j is the first variable entering the Lasso path and p_3 is in the region of rejecting variable j as a signal. Since $h \sim \mathcal{N}(\mu_3, G)$ with $\mu_3 = (\tau_p, \rho\tau_p, \rho^2\tau_p, \rho\tau_p)^T$, by Lemma 7.1,

$$\begin{aligned}\mathbb{P}\big\{W_j \le t_p \big| (\beta_j, \beta_{j+1}) = (\tau_p, 0)\big\} &\ge L_p p^{-(p_3-\mu_3)^T G^{-1}(p_3-\mu_3)/2\log(p)} \\ &= L_p p^{-[(\sqrt{r}-\sqrt{u})_+]^2}.\end{aligned}$$

Now, we show that (G.6) holds for $u \ge u^*$, which implies (G.6) for all $u \ge 0$ as discussed in the proof of EC-knockoff. We prove (G.6) by showing that

$$\mathbb{P}\big\{W_j \le t_p, A \big| (\beta_j, \beta_{j+1}) = (\tau_p, 0)\big\} \le L_p p^{-[(\sqrt{r}-\sqrt{u})_+]^2} \tag{G.7}$$

holds for $A = A_j, A_{j+1}, A_{j+p}, A_{j+p+1}$ given $u \ge u^*$. Respectively,

$$\begin{aligned}\mathbb{P}\big\{W_j \le t_p, A_j \big| (\beta_j, \beta_{j+1}) = (\tau_p, 0)\big\} &\le \mathbb{P}\big\{|y^T x_j| \le t_p \big| (\beta_j, \beta_{j+1}) = (\tau_p, 0)\big\} \\ &= L_p p^{-[(\sqrt{r}-\sqrt{u})_+]^2},\end{aligned}$$

and by symmetry and (F.14),

$$\begin{aligned}&\mathbb{P}\big\{W_j \le t_p, A_{j+1} \big| (\beta_j, \beta_{j+1}) = (\tau_p, 0)\big\} = \mathbb{P}\big\{W_j \le t_p, A_{j+p+1} \big| (\beta_j, \beta_{j+1}) = (\tau_p, 0)\big\} \\ &\le \mathbb{P}\big\{|y^T x_j| \le |y^T x_{j+p+1}| \big| (\beta_j, \beta_{j+1}) = (\tau_p, 0)\big\} \le L_p p^{-\frac{1-\rho}{2} r} \le L_p p^{-[(\sqrt{r}-\sqrt{u})_+]^2},\end{aligned}$$

$$\begin{aligned}\mathbb{P}\big\{W_j \le t_p, A_{j+p} \big| (\beta_j, \beta_{j+1}) = (\tau_p, 0)\big\} &\le \mathbb{P}\big\{|y^T x_j| \le |y^T x_{j+p}| \big| (\beta_j, \beta_{j+1}) = (\tau_p, 0)\big\} \\ &\le L_p p^{-\frac{1-\rho^2}{2} r} \le L_p p^{-[(\sqrt{r}-\sqrt{u})_+]^2}.\end{aligned}$$

(G.6) is immediate by $[(\sqrt{r} - \sqrt{u})_+]^2 \ge f^+_{\mathrm{Hamm}}(r, u, \vartheta) - \vartheta$.

- When $(\beta_j, \beta_{j+1}) = (\tau_p, \tau_p)$,

$$\mathbb{P}\big\{W_j \le t_p \big| (\beta_j, \beta_{j+1}) = (\tau_p, \tau_p)\big\} \le L_p p^{\vartheta - f^+_{\mathrm{Hamm}}(u,r,\vartheta)}. \tag{G.8}$$

We prove (G.8) by showing

$$\mathbb{P}\big\{W_j \le t_p, A \big| (\beta_j, \beta_{j+1}) = (\tau_p, \tau_p)\big\} \le L_p p^{-[(\sqrt{r}-\sqrt{u})_+]^2} \tag{G.9}$$

holds for $A = A_j, A_{j+1\to j}, A_{j+1\to j+p}, A_{j+1\to j+p+1}, A_{j+p}, A_{j+p+1}$ given $u \ge u^*$, which cover all possibilities. Respectively,

$$\begin{aligned}\mathbb{P}\big\{W_j \le t_p, A_j \big| (\beta_j, \beta_{j+1}) = (\tau_p, \tau_p)\big\} &\le \mathbb{P}\big\{|y^T x_j| \le t_p \big| (\beta_j, \beta_{j+1}) = (\tau_p, \tau_p)\big\} \\ &\le L_p p^{-[((1+\rho)\sqrt{r}-\sqrt{u})_+]^2} \le L_p p^{-[(\sqrt{r}-\sqrt{u})_+]^2},\end{aligned}$$

$$\begin{aligned}\mathbb{P}\big\{W_j \le t_p, A_{j+p} \big| (\beta_j, \beta_{j+1}) = (\tau_p, \tau_p)\big\} &\le \mathbb{P}\big\{|y^T x_j| \le |y^T \tilde{x}_j| \big| (\beta_j, \beta_{j+1}) = (\tau_p, \tau_p)\big\} \\ &\le L_p p^{-\frac{1-\rho^2}{2} r} \le L_p p^{-[(\sqrt{r}-\sqrt{u})_+]^2},\end{aligned}$$

$$\mathbb{P}\big\{W_j \le t_p, A_{j+p+1}\big|(\beta_j,\beta_{j+1})=(\tau_p,\tau_p)\big\} \le \mathbb{P}\big\{|y^Tx_j| \le |y^T\tilde{x}_{j+1}|\big|(\beta_j,\beta_{j+1})=(\tau_p,\tau_p)\big\}$$
$$\le L_p p^{-\frac{(1-\rho)(1+\rho)^2}{2}r} \le L_p p^{-[(\sqrt{r}-\sqrt{u})_+]^2}.$$

When $A_{j+1\to j}$ occurs, the bi-variate Lasso problem $L_{j,j+1}$ has variable j is a false negative given $W_j \le t_p$, which implies $\mathbb{P}\big\{W_j \le t_p, A_{j+1\to j}\big|(\beta_j,\beta_{j+1})=(\tau_p,\tau_p)\big\}$ is upper bounded by the corresponding false negative rate of Lasso, which is $L_p p^{-(\xi_\rho\sqrt{r}-\eta_\rho\sqrt{u})_+^2} \le L_p p^{-[(\sqrt{r}-\sqrt{u})_+]^2}$ for $u \ge u^*$.

When $A_{j+1\to j+p}$ occurs, we know variable $j+p$ instead of variable j is the second one entering the Lasso path. This means the λ_2 (the λ value when the second variable entering Lasso path) of the bi-variate Lasso problem $L_{j+1,j+p}$ is larger than the λ_2 of the bi-variate Lasso problem $L_{j,j+1}$. When $y^Tx_{j+1} \ge 0$, we must have

$$\max\{\frac{y^Tx_j-\rho y^Tx_{j+1}}{1-\rho}, \frac{y^Tx_j-\rho y^Tx_{j+1}}{-1-\rho}\} < \max\{\frac{y^T\tilde{x}_j-\rho y^Tx_{j+1}}{1-\rho}, \frac{y^T\tilde{x}_j-\rho y^Tx_{j+1}}{-1-\rho}\}.$$

Therefore, $A_{j+1\to j+p}$ implies one the three following events must occur:

$$y^Tx_{j+1} < 0, \frac{y^Tx_j-\rho y^Tx_{j+1}}{1-\rho} < \frac{y^T\tilde{x}_j-\rho y^Tx_{j+1}}{1-\rho}, \frac{y^Tx_j-\rho y^Tx_{j+1}}{1-\rho} < \frac{y^T\tilde{x}_j-\rho y^Tx_{j+1}}{-1-\rho}$$

The probability of these three events given $(\beta_j,\beta_{j+1})=(\tau_p,\tau_p)$ are $L_p p^{-(1+\rho)^2 r}$, $L_p p^{-\frac{1-\rho^2}{2}r}$ and $L_p p^{-\frac{(1+\rho)^3(1-\rho)}{2(1+\rho^2)}r}$, all of which are upper bounded by $L_p p^{-[(\sqrt{r}-\sqrt{u})_+]^2}$ when $u \ge u^*$.

When $A_{j+1\to j+p+1}$ occurs, the λ_2 of the bi-variate Lasso problem $L_{j+1,j+p+1}$ is larger than the λ_2 of the bi-variate Lasso problem $L_{j,j+1}$. When $y^Tx_{j+1} \ge 0$, we must have

$$\max\{\frac{y^Tx_j-\rho y^Tx_{j+1}}{1-\rho}, \frac{y^Tx_j-\rho y^Tx_{j+1}}{-1-\rho}\} < \max\{\frac{y^T\tilde{x}_{j+1}-\rho^2 y^Tx_{j+1}}{1-\rho^2}, \frac{y^T\tilde{x}_{j+1}-\rho^2 y^Tx_{j+1}}{-1-\rho^2}\}.$$

Therefore, $A_{j+1\to j+p+1}$ implies one the three following events must occur:

$$y^Tx_{j+1} < 0, \frac{y^Tx_j-\rho y^Tx_{j+1}}{1-\rho} < \frac{y^T\tilde{x}_{j+1}-\rho^2 y^Tx_{j+1}}{1-\rho^2}, \frac{y^Tx_j-\rho y^Tx_{j+1}}{1-\rho} < -\frac{y^T\tilde{x}_{j+1}-\rho^2 y^Tx_{j+1}}{-1-\rho^2}.$$

Respectively, the probability of these three events are $L_p p^{-(1+\rho)^2 r}$, $L_p p^{-\frac{1-\rho^2}{2}r}$ and $L_p p^{-\frac{(1+\rho)^3(1-\rho)}{2(1+\rho^2)}r}$, all of which are upper bounded by $L_p p^{-[(\sqrt{r}-\sqrt{u})_+]^2}$ when $u \ge u^*$. From here we have verified (G.9), thus implies (G.8).

From (G.1), (G.2), (G.5), (G.6) and (G.8), we have

$$\mathbb{P}\big\{W_j > t_p, \beta_j = 0\big\} + \mathbb{P}\big\{W_j \le t_p, \beta_j = \tau_p\big\} = L_p p^{-f^+_{\text{Hamm}}(r,u,\vartheta)}, \tag{G.10}$$

which completes the proof for positive ρ.

- When $(\beta_j,\beta_{j+1})=(-\tau_p,\tau_p)$,

$$\mathbb{P}\big\{W_j \le t_p\big|(\beta_j,\beta_{j+1})=(-\tau_p,\tau_p)\big\} \ge L_p p^{-((\xi_\rho\sqrt{r}-\eta_\rho^{-1}\sqrt{u})_+)^2} \tag{G.11}$$

and

$$\mathbb{P}\big\{W_j \le t_p \big| (\beta_j, \beta_{j+1}) = (-\tau_p, \tau_p)\big\} \le L_p p^{-\min\{((\xi_\rho\sqrt{r}-\eta_\rho^{-1}\sqrt{u})_+)^2, \frac{(1-\rho)^3(1+\rho)}{2(1+\rho^2)}r\}}. \quad \text{(G.12)}$$

Let

$$p_4^T = \begin{cases} (-(1-\rho)\tau_p, (1-\rho)\tau_p, \rho(1-\rho)\tau_p, -\rho(1-\rho)\tau_p), & (1-\rho)\tau_p \le t_p, \\ (\rho(1-\rho)\tau_p - (1+\rho)t_p, (1-\rho)\tau_p, \rho(1-\rho)\tau_p, \rho^2(1-\rho)\tau_p - \rho(1+\rho)t_p), & (1-\rho)\tau_p > t_p. \end{cases}$$

When $h = p_4$ and $(1-\rho)\tau_p \le t_p$, variable j is the first variable entering the Lasso path with $W_j = (1-\rho)\tau_p \le t_p$; when $h = p_4$ and $(1-\rho)\tau_p > t_p$, $j+1$ is the first and j is the second variable entering the Lasso path with $W_j = t_p$. Regardless of the relationship between τ_p and t_p, $h = p_4$ is always in the region of rejecting j as a signal. Since $h \sim \mathcal{N}(\mu_4, G)$ with $\mu_4 = (-(1-\rho)\tau_p, (1-\rho)\tau_p, \rho(1-\rho)\tau_p, -\rho(1-\rho)\tau_p)^T$, by Lemma 7.1,

$$\begin{aligned}\mathbb{P}\big\{W_j \le t_p \big| (\beta_j, \beta_{j+1}) = (-\tau_p, \tau_p)\big\} &\ge L_p p^{-(p_4-\mu_4)^T G^{-1}(p_4-\mu_4)/2\log(p)} \\ &= L_p p^{-((\xi_\rho\sqrt{r}-\eta_\rho^{-1}\sqrt{u})_+)^2}.\end{aligned}$$

Next, we show that

$$\mathbb{P}\big\{W_j \le t_p, A \big| (\beta_j, \beta_{j+1}) = (-\tau_p, \tau_p)\big\} \le L_p p^{-\min\{((\xi_\rho\sqrt{r}-\eta_\rho^{-1}\sqrt{u})_+)^2, \frac{(1-\rho)^3(1+\rho)}{2(1+\rho^2)}r\}}. \quad \text{(G.13)}$$

holds for $A = A_j, A_{j+1\to j}, A_{j+1\to j+p}, A_{j+1\to j+p+1}, A_{j+p}, A_{j+p+1}$, which cover all possibilities.

When $A = A_j$ or $A_{j+1\to j}$ occurs, as previously discussed, variable j is a false negative in the bi-variate Lasso $L_{j,j+1}$ given $W_j \le t_p$, which implies $\mathbb{P}\big\{W_j \le t_p, A \big| (\beta_j, \beta_{j+1}) = (-\tau_p, \tau_p)\big\}$ is upper bounded by the corresponding false negative rate of Lasso, which is $L_p p^{-((\xi_\rho\sqrt{r}-\eta_\rho^{-1}\sqrt{u})_+)^2}$.

When $A_{j+1\to j+p}$ occurs, the λ_2 of the bi-variate Lasso problem $L_{j+1,j+p}$ is larger than the λ_2 of the bi-variate Lasso problem $L_{j,j+1}$. When $y^T x_{j+1} \ge 0$, we must have

$$\max\{\frac{y^T x_j - \rho y^T x_{j+1}}{1-\rho}, \frac{y^T x_j - \rho y^T x_{j+1}}{-1-\rho}\} < \max\{\frac{y^T \tilde{x}_j - \rho y^T x_{j+1}}{1-\rho}, \frac{y^T \tilde{x}_j - \rho y^T x_{j+1}}{-1-\rho}\}.$$

Therefore, $A_{j+1\to j+p}$ implies one of the three following events must occur:

$$y^T x_{j+1} + y^T \tilde{x}_j < 0, \frac{y^T x_j - \rho y^T x_{j+1}}{-1-\rho} < \frac{y^T \tilde{x}_j - \rho y^T x_{j+1}}{1-\rho}, \frac{y^T x_j - \rho y^T x_{j+1}}{-1-\rho} < \frac{y^T \tilde{x}_j - \rho y^T x_{j+1}}{-1-\rho}$$

The probability of these three events are $L_p p^{-\frac{(1+\rho)(1-\rho)^2}{2}r}$, $L_p p^{-\frac{(1-\rho)^3(1+\rho)}{2(1+\rho^2)}r}$ and $L_p p^{-\frac{1-\rho^2}{2}r}$, all of which are upper bounded by $L_p p^{-\frac{(1-\rho)^3(1+\rho)}{2(1+\rho^2)}r}$.

When $A_{j+1\to j+p+1}$ occurs, the λ_2 of the bi-variate Lasso problem $L_{j+1,j+p+1}$ is larger than the λ_2 of the bi-variate Lasso problem $L_{j,j+1}$. When $y^T x_{j+1} \ge 0$, we must have

$$\max\{\frac{y^T x_j - \rho y^T x_{j+1}}{1-\rho}, \frac{y^T x_j - \rho y^T x_{j+1}}{-1-\rho}\} < \max\{\frac{y^T \tilde{x}_{j+1} - \rho^2 y^T x_{j+1}}{1-\rho^2}, \frac{y^T \tilde{x}_{j+1} - \rho^2 y^T x_{j+1}}{-1-\rho^2}\}.$$

Therefore, $A_{j+1\to j+p+1}$ implies one of the three following events must occur:

$$y^T x_{j+1}+y^T\tilde{x}_j<0, \frac{y^T x_j-\rho y^T x_{j+1}}{-1-\rho}<\frac{y^T\tilde{x}_{j+1}-\rho^2 y^T x_{j+1}}{1-\rho^2}, \frac{y^T x_j-\rho y^T x_{j+1}}{-1-\rho}<\frac{y^T\tilde{x}_{j+1}-\rho^2 y^T x_{j+1}}{-1-\rho^2}.$$

The probability of these three events are $L_p p^{-\frac{(1+\rho)(1-\rho)^2}{2}r}$, $L_p p^{-\frac{1-\rho^2}{2}r}$, $L_p p^{-\frac{(1-\rho)^3(1+\rho)}{2(1+\rho^2)}r}$, all of which are upper bounded by $L_p p^{-\frac{(1-\rho)^3(1+\rho)}{2(1+\rho^2)}r}$.

When A_{j+p} occurs, if $y^T\tilde{x}_j<0$, then $y^T x_{j+1}+y^T\tilde{x}_j\le 0$, which happens with probability $L_p p^{-\frac{(1+\rho)(1-\rho)^2}{2}r}\le L_p p^{-\frac{(1-\rho)^3(1+\rho)}{2(1+\rho^2)}r}$. If $y^T\tilde{x}_j\ge 0$, then $y^T\tilde{x}_j+\frac{1-\rho}{2}y^T x_j-\frac{1+\rho}{2}y^T x_{j+1}\ge 0$, which happens with probability $L_p p^{-\frac{2(1-\rho)^3}{3+\rho^2}r}\le L_p p^{-\frac{(1-\rho)^3(1+\rho)}{2(1+\rho^2)}r}$. Therefore, (G.13) holds for A_{j+p} and also for A_{j+p+1} due to symmetry. We thus complete the proof for (G.13).

Now consider the case where β_j takes value in $\{0,-\tau_p\}$ and β_{j+1} takes value in $\{0,\tau_p\}$, this corresponds to the $\rho<0$ case (we flipped the sign of ρ and β_j simultaneously). By (G.1), (G.2), (G.5) and (G.11), we know

$$\begin{aligned}&\mathbb{P}\big\{W_j>t_p,\beta_j=0\big\}+\mathbb{P}\big\{W_j\le t_p,\beta_j=-\tau_p\big\}\\&\ge L_p p^{-\min\{f^+_{\mathrm{Hamm}}(u,r,\vartheta),2\vartheta+((\xi_\rho\sqrt{r}-\eta_\rho^{-1}\sqrt{u})_+)^2\}}.\end{aligned}\tag{G.14}$$

Meanwhile, (G.1), (G.2), (G.6) and (G.12) gives

$$\begin{aligned}&\mathbb{P}\big\{W_j>t_p,\beta_j=0\big\}+\mathbb{P}\big\{W_j\le t_p,\beta_j=-\tau_p\big\}\\&\le L_p p^{-\min\{f^+_{\mathrm{Hamm}}(u,r,\vartheta),2\vartheta+((\xi_\rho\sqrt{r}-\eta_\rho^{-1}\sqrt{u})_+)^2,2\vartheta+\frac{(1-|\rho|)^3(1+|\rho|)}{2(1+|\rho|^2)}r\}}.\end{aligned}\tag{G.15}$$

The proof is complete once we show that

$$\min\{f^+_{\mathrm{Hamm}}(u,r,\vartheta),2\vartheta+((\xi_\rho\sqrt{r}-\eta_\rho^{-1}\sqrt{u})_+)^2\}\le 2\vartheta+\frac{(1-|\rho|)^3(1+|\rho|)}{2(1+\rho^2)}r.\tag{G.16}$$

Otherwise, there exists a tuple of (ϑ,r,ρ,u,r) such that

$$2\vartheta+\frac{(1-|\rho|)^3(1+|\rho|)}{2(1+\rho^2)}r<2\vartheta+((\xi_\rho\sqrt{r}-\eta_\rho^{-1}\sqrt{u})_+)^2\tag{G.17}$$

and

$$2\vartheta+\frac{(1-|\rho|)^3(1+|\rho|)}{2(1+\rho^2)}r<\vartheta+(\sqrt{u}-|\rho|\sqrt{r})^2+((\xi_\rho\sqrt{r}-\eta_\rho\sqrt{u})_+)^2-((\sqrt{r}-\sqrt{u})_+)^2\tag{G.18}$$

are satisfied simultaneously.

By (G.17), $\xi_\rho\sqrt{r}-\eta_\rho^{-1}\sqrt{u}>0$, which implies $(1-|\rho|)\sqrt{r}>\sqrt{u}$. Therefore, the right hand side of (G.18) simplifies to $\vartheta+\frac{1-|\rho|}{1+|\rho|}u$. By (G.18), we know

$$\frac{(1-|\rho|)^3(1+|\rho|)}{2(1+\rho^2)}r\le\vartheta+\frac{(1-|\rho|)^3(1+|\rho|)}{2(1+\rho^2)}r<\frac{1-|\rho|}{1+|\rho|}u.$$

Plug this into the right hand side of (G.17), we have

$$\begin{aligned} 2\vartheta + \frac{(1-|\rho|)^3(1+|\rho|)}{2(1+\rho^2)} r &< 2\vartheta + ((\xi_\rho\sqrt{r} - \eta_\rho^{-1}\sqrt{u})_+)^2 \\ &\le 2\vartheta + \Big(\sqrt{1-\rho^2} - \sqrt{\frac{(1-|\rho|)(1+|\rho|)^3}{2(1+\rho^2)}}\Big)^2 r, \end{aligned} \tag{G.19}$$

which can only be true when $\rho^2 > 1$. By reduction, we proved (G.16).

Appendix H. Proof of Theorem 6.1

The least-squares estimator satisfies that $\hat\beta \sim \mathcal{N}_p(\beta, G^{-1})$. It gives $\hat\beta_j \sim \mathcal{N}(\beta_j, \omega_j)$. Applying Lemma 7.1 to $X_p = \hat\beta_j$ and $S = \{x \in \mathbb{R} : x \ge \sqrt{u}\}$, we have

$$\mathbb{P}(|\hat\beta_j| > t_p(u)|\beta_j = 0) = L_p p^{-\omega_j^{-1}u}, \qquad \mathbb{P}(|\hat\beta_j| \le t_p(u)|\beta_j = \tau_p) = L_p p^{-\omega_j^{-1}(\sqrt{r}-\sqrt{u})_+^2}.$$

It follows that

$$\begin{aligned} \mathrm{FP}_p(u) &= \sum_{j=1}^p (1-\epsilon_p)\cdot \mathbb{P}(W_j^* > t_p(u)|\beta_j = 0) = L_p \sum_{j=1}^p p^{-\omega_j^{-1}u}, \\ \mathrm{FN}_p(u) &= \sum_{j=1}^p \epsilon_p \cdot \mathbb{P}(W_j^* < t_p(u)|\beta_j = \tau_p) = L_p p^{-\vartheta} \sum_{j=1}^p p^{-\omega_j^{-1}(\sqrt{r}-\sqrt{u})_+^2}. \end{aligned}$$

For the block-wise diagonal design (5.1), $\omega_j = (1-\rho^2)^{-1}$ for all $1 \le j \le p-1$.

Appendix I. Proof of Theorem 6.2

By the property of least-square coefficients,

$$(\hat\beta_1, \cdots, \hat\beta_p, \tilde\beta_1, \cdots, \tilde\beta_p) \sim \mathcal{N}_{2p}\big((\beta_1, \cdots, \beta_p, 0, \cdots, 0), (G^*)^{-1}\big).$$

Consider the joint distribution of $\hat\beta_j$ and $\tilde\beta_j$ which are the regression coefficient of x_j and $\tilde x_j$, we know that $(\hat\beta_j, \tilde\beta_j) \sim \mathcal{N}_2\big((\beta_j, 0), A_j\big)$ where A_j has ω_{1j} as its diagonal element and ω_{2j} as its off-diagonal elements. Then theorem 6.2 is immediate from the following lemma:

Lemma I.1 *If $(Z_j, \tilde Z_j)$ follows $\mathcal{N}_2\Big((\beta_j, 0)^T,\ \Sigma\Big)$ with $\Sigma = ((\sigma_1, \sigma_2), (\sigma_2, \sigma_1))$, then*

$$\mathbb{P}(|Z_j| > \sqrt{2u\log(p)}, |Z_j| \ge |\tilde Z_j||\beta_j = 0) = L_p p^{-u/\sigma_1} \tag{I.1}$$

and

$$\begin{aligned} &\mathbb{P}(|Z_j| \le \sqrt{2u\log(p)} \ \text{ or } |Z_j| < |\tilde Z_j||\beta_j = \sqrt{2r\log(p)}) \\ =&L_p p^{-\min\{(\sqrt{r}-\sqrt{u})_+^2/\sigma_1, r/(2\max\{\sigma_1+\sigma_2, \sigma_1-\sigma_2\})\}}. \end{aligned} \tag{I.2}$$

Next, we prove Lemma I.1. To compute the left hand side of (I.1), we only need to find the t such that ellipsoid $(x,y)\Sigma^{-1}(x,y)^T = t^2$ is tangent with $x = \pm\sqrt{2u\log(p)}$. This is because when we increase the radius of the ellipsoid, it must intersect with $x = \pm\sqrt{2u\log(p)}$ first amongst the boundaries of the region that pick variable j as a signal. When they intersect,

$$t^2 = \frac{1}{\sigma_1^2-\sigma_2^2}(\sigma_1 x^2 - 2\sigma_2 xy + \sigma_1 y^2) = \frac{1}{\sigma_1^2-\sigma_2^2}\Big(\sigma_1\Big(y-\frac{\sigma_2}{\sigma_1}x\Big)^2 + \Big(\sigma_1 - \frac{\sigma_2^2}{\sigma_1}\Big)x^2\Big) \geq \frac{2u\log(p)}{\sigma_1}.$$

When $t^2 = \frac{2u\log(p)}{\sigma_1}$, the tangent points are $(\pm\sqrt{2u\log(p)}, \pm\frac{\sigma_2}{\sigma_1}\sqrt{2u\log(p)})$. By Lemma 7.1, we verified (I.1).

For (I.2), when $r < u$, the center of the bi-variate normal is in the region of rejecting variable j as a signal thus the false positive rate is L_p. When $r > u$, we need to find the t such that ellipsoid $(x-\beta_j, y)\Sigma^{-1}(x-\beta_j, y)^T = t^2$ is tangent with either $x = \pm\sqrt{2u\log(p)}$ or $y = \pm x$. When the ellipsoid intersects with $x = \pm\sqrt{2u\log(p)}$,

$$t^2 = \frac{1}{\sigma_1^2-\sigma_2^2}\Big(\sigma_1\Big(y-\frac{\sigma_2}{\sigma_1}(x-\beta_j)\Big)^2 + \Big(\sigma_1 - \frac{\sigma_2^2}{\sigma_1}\Big)(x-\beta_j)^2\Big) \geq \frac{2(\sqrt{u}-\sqrt{r})^2\log(p)}{\sigma_1},$$

therefore, they are tangent at $(\pm\sqrt{2u\log(p)}, \frac{\sigma_2}{\sigma_1}(\pm\sqrt{2u\log(p)}-\beta_j))$ when $t^2 = \frac{2(\sqrt{u}-\sqrt{r})^2\log(p)}{\sigma_1}$.

Meanwhile, since the long/short shaft of the ellipsoid are paralleled with $y = \pm x$, the tangent points of ellipsoid with $y = \pm x$ must be $(\beta_j/2, \beta_j/2)$ and $(\beta_j/2, -\beta_j/2)$, which gives $t^2 = \frac{r\log(p)}{\sigma_1+\sigma_2}$ and $\frac{r\log(p)}{\sigma_1-\sigma_2}$. From here we can conclude the "distance" between the center of the normal distribution and the region that reject variable j as a signal is

$$\min\{\frac{2(\sqrt{r}-\sqrt{u})_+^2\log(p)}{\sigma_1}, \frac{r\log(p)}{\sigma_1+\sigma_2}, \frac{r\log(p)}{\sigma_1-\sigma_2}\}.$$

By Lemma 7.1, we know

$$\mathbb{P}(|Z_j| \leq \sqrt{2u\log(p)}\big|\beta_j = \sqrt{2r\log(p)}) = L_p p^{-\min\{(\sqrt{r}-\sqrt{u})_+^2/\sigma_1, r/(2\max\{\sigma_1+\sigma_2, \sigma_1-\sigma_2\})\}}.$$

References

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

Rina Foygel Barber and Emmanuel J Candès. Controlling the false discovery rate via knockoffs. *The Annals of Statistics*, 43(5):2055–2085, 2015.

Ian Barnett, Rajarshi Mukherjee, and Xihong Lin. The generalized higher criticism for testing snp-set effects in genetic association studies. *Journal of the American Statistical Association*, 112(517):64–76, 2017.

Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995.

Emmanuel Candes, Yingying Fan, Lucas Janson, and Jinchi Lv. Panning for gold: 'model-x' knockoffs for high dimensional controlled variable selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(3):551–577, 2018.

Chenguang Dai, Buyu Lin, Xin Xing, and Jun S Liu. False discovery rate control via data splitting. *Journal of the American Statistical Association*, pages 1–18, 2022.

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

David Donoho and Jiashun Jin. Special invited paper: Higher criticism for large-scale inference, especially for rare and weak effects. *Statistical Science*, pages 1–25, 2015.

Jianqing Fan and Jinchi Lv. Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(5): 849–911, 2008.

Yingying Fan, Emre Demirkaya, Gaorong Li, and Jinchi Lv. Rank: large-scale inference with graphical nonlinear knockoffs. *Journal of the American Statistical Association*, pages 1–43, 2019.

Christopher R Genovese, Jiashun Jin, Larry Wasserman, and Zhigang Yao. A comparison of the lasso and marginal regression. *Journal of Machine Learning Research*, 13(Jun): 2107–2143, 2012.

Adel Javanmard and Hamid Javadi. False discovery rate control via debiased lasso. *Electronic Journal of Statistics*, 13(1):1212–1253, 2019.

Adel Javanmard and Andrea Montanari. Online rules for control of false discovery rate and false discovery exceedance. *The Annals of statistics*, 46(2):526–554, 2018.

Pengsheng Ji and Jiashun Jin. UPS delivers optimal phase diagram in high-dimensional variable selection. *The Annals of Statistics*, 40(1):73–103, 2012.

Jiashun Jin and Zheng Tracy Ke. Rare and weak effects in large-scale inference: methods and phase diagrams. *Statistica Sinica*, pages 1–34, 2016.

Jiashun Jin, Cun-Hui Zhang, and Qi Zhang. Optimality of graphlet screening in high dimensional variable selection. *The Journal of Machine Learning Research*, 15(1):2723–2772, 2014.

Zheng Tracy Ke and Longlin Wang. A comparison of Hamming errors of representative variable selection methods. In *International Conference on Learning Representations*, 2021.

Zheng Tracy Ke and Fan Yang. Covariate assisted variable ranking. *arXiv preprint arXiv:1705.10370*, 2017.

Zheng Tracy Ke, Jiashun Jin, and Jianqing Fan. Covariance assisted screening and estimation. *The Annals of statistics*, 42(6):2202, 2014.

Zheng Tracy Ke, Jun S. Liu, and Yucong Ma. The de-randomized Gaussian mirror and its power analysis. *Manuscript*, 2022.

Xiao Li and William Fithian. Whiteout: when do fixed-x knockoffs fail? *arXiv preprint arXiv:2107.06388*, 2021.

Jingbo Liu and Philippe Rigollet. Power analysis of knockoff filters for correlated designs. In *Advances in Neural Information Processing Systems*, pages 15420–15429, 2019.

Asher Spector and Lucas Janson. Powerful knockoffs via minimizing reconstructability. *The Annals of Statistics*, 50(1):252–276, 2022.

Weijie Su, Małgorzata Bogdan, and Emmanuel Candes. False discoveries occur early on the lasso path. *The Annals of statistics*, 45(5):2133–2150, 2017.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

Wenshuo Wang and Lucas Janson. A high-dimensional power analysis of the conditional randomization test and knockoffs. *Biometrika*, 109(3):631–645, 2022.

Asaf Weinstein, Rina Barber, and Emmanuel Candes. A power and prediction analysis for knockoffs with lasso statistics. *arXiv preprint arXiv:1712.06465*, 2017.

Asaf Weinstein, Weijie J Su, Małgorzata Bogdan, Rina F Barber, and Emmanuel J Candès. A power analysis for knockoffs with the lasso coefficient-difference statistic. *arXiv preprint arXiv:2007.15346*, 2021.

Xin Xing, Zhigen Zhao, and Jun S Liu. Controlling false discovery rate using gaussian mirrors. *Journal of the American Statistical Association*, 118(541):222–241, 2023.

Peng Zhao and Bin Yu. On model selection consistency of lasso. *The Journal of Machine Learning Research*, 7:2541–2563, 2006.