# A Comparison of Hamming Errors of Representative Variable Selection Methods

**Zheng Tracy Ke**
Department of Statistics
Harvard University
Cambridge, MA 02138, USA
zke@fas.harvard.edu

**Longlin Wang**
Department of Statistics
Harvard University
Cambridge, MA 02138, USA
lwang2@fas.harvard.edu

## ABSTRACT

Lasso is a celebrated method for variable selection in linear models, but it faces challenges when the variables are moderately or strongly correlated. This motivates alternative approaches such as using a non-convex penalty, adding a ridge regularization, or conducting a post-Lasso thresholding. In this paper, we compare Lasso with 5 other methods: Elastic net, SCAD, forward selection, thresholded Lasso, and forward backward selection. We measure their performances theoretically by the expected Hamming error, assuming that the regression coefficients are *iid* drawn from a two-point mixture and that the Gram matrix is block-wise diagonal. By deriving the rates of convergence of Hamming errors and the phase diagrams, we obtain useful conclusions about the pros and cons of different methods.

## 1 INTRODUCTION

Variable selection is one of the core problems in high-dimensional data analysis. Consider a linear regression, where the response $y \in \mathbb{R}^n$ and the design matrix $X = [x_1, \ldots, x_p] \in \mathbb{R}^{n \times p}$ satisfy that

$$y = X\beta + z, \qquad \|x_j\| = 1, \qquad z \sim \mathcal{N}(0, \sigma^2 I_n). \tag{1}$$

The goal is estimating the support of $\beta$ ($\mathrm{Supp}(\beta)$). Lasso (Tibshirani, 1996) is a popular method:

$$\hat{\beta}^{\mathrm{lasso}} = \mathrm{argmin}_\beta \big\{ \|y - X\beta\|^2/2 + \lambda\|\beta\|_1 \big\}. \tag{2}$$

Lasso has good rates of convergence on the $L_q$-estimation error or prediction error (Bickel et al., 2009). However, it can be unsatisfactory for variable selection, especially when the columns in the design matrix are moderately or strongly correlated. Zhao & Yu (2006) showed that an *irrepresentable condition* on $X$ is necessary for Lasso to recover $\mathrm{Supp}(\beta)$ with high probability, and such a condition is restrictive when $p$ is large (Fan & Lv, 2010). Ji & Jin (2012) studied the Hamming error of Lasso and revealed Lasso's non-optimality by lower-bounding its Hamming error rate. Many alternative strategies were proposed for variable selection, such as using non-convex penalties (Fan & Li, 2001; Zhang, 2010; Shen et al., 2012), adding a ridge regularization (Zou & Hastie, 2005), post-processing on the Lasso estimator (Zou, 2006; Zhou, 2009), and iterative algorithms (Zhang, 2011; Donoho et al., 2012). In this paper, our main interest is to theoretically compare these different strategies.

Existing theoretical studies focused on 'model selection consistency' (e.g., Fan & Li (2001); Zhao & Yu (2006); Zou (2006); Meinshausen & Bühlmann (2010); Loh & Wainwright (2017)), which uses $\mathbb{P}(\mathrm{Supp}(\hat{\beta}) = \mathrm{Supp}(\beta))$ to measure the performance of variable selection. However, for many real applications, the study of the Hamming error (i.e., total number of false positives and false negatives) is in urgent need. For example, in genome-wide association studies (GWAS) or Genetic Regulatory Network, the goal is to identify the genes or SNPs that are truly associated with a given phenotype, and we hope to find a multiple testing procedure that simultaneously controls the FDR and maximizes the power (for multiple testing). This problem can be re-cast as minimizing the Hamming error in a special regression setting (Efron, 2004; Jin, 2012; Sun & Cai, 2007). This motivates us to study the *Hamming errors* of variable selection methods, which were rarely considered in the literature.

We adopt the *rare and weak signal model* (Donoho & Jin, 2004; Arias-Castro et al., 2011; Jin & Ke, 2016), which is often used in theoretical analysis of sparse linear models. Let $p$ be the asymptotic

parameter. Given constants $\vartheta \in (0, 1)$ and $r > 0$, we assume that $\beta_j$'s are iid generated such that

$$\beta_j = \begin{cases} \tau_p, & \text{with probability } \epsilon_p, \\ 0, & \text{with probability } 1 - \epsilon_p, \end{cases} \qquad \text{where} \qquad \epsilon_p = p^{-\vartheta}, \quad \tau_p = \sqrt{2r \log(p)}. \qquad (3)$$

As $p \to \infty$, $\|\beta\|_0 \approx p^{1-\vartheta}$, and a nonzero $\beta_j$ is at the critical order $\sqrt{\log(p)}$. [1] The two parameters $(\vartheta, r)$ capture the sparsity level and signal strength, respectively. We may generalize (3) to let nonzero $\beta_j$'s take different values in $[\tau_p, \infty)$, but the current form is more convenient for presentation.

The *blockwise covariance structure* is frequently observed in real applications. In genetic data, there may exist strong correlations between nearby genetic markers, but the long-range dependence is usually negligible; as a result, the sample covariance matrix is approximately blockwise diagonal (Dehman et al., 2015). In financial data, the sample covariance matrix of stock returns (after common factors are removed) is also approximately blockwise diagonal, where each block corresponds to an industry group (Fan et al., 2015). Motivated by these examples, we consider an idealized setting, where the Gram matrix $G = X'X$ is block-wise diagonal consisting of $2 \times 2$ blocks:

$$G = \text{diag}(B, B, \dots, B, B_0), \qquad \text{where} \quad B = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \text{ and } B_0 = \begin{cases} B, & \text{if } p \text{ is even,} \\ 1, & \text{if } p \text{ is odd.} \end{cases} \qquad (4)$$

This is an idealization of the blockwise covariance structures in real applications. We may generalize (4) to allow unequal block sizes and unequal off-diagonal entries, but we keep the current form for convenience of presentation. Model (4) is also closely connected to the random designs in compressed sensing (Donoho, 2006). Write $X = [X_1, X_2, \dots, X_n]'$. Suppose $X_1, X_2, \dots, X_n$ are iid generated from $\mathcal{N}(0, n^{-1}\Sigma)$, where $\Sigma$ has the same form as $G$ in (4). In a high-dimensional sparse setting, we have $\|\beta\|_0 \ll n \ll p$. Then, $G = X'X \approx \Sigma$, and due to the blessing of sparsity of $\beta$, $G\beta \approx \Sigma\beta$. As a result, $X'y$ (sufficient statistic of $\beta$) satisfies that $X'y = G\beta + \mathcal{N}(0, G) \approx \Sigma\beta + \mathcal{N}(0, \Sigma)$, and the right hand side reduces to Model (4) (Genovese et al., 2012). In Section 3.5, we formally show that this random design setting is asymptotically equivalent to Model (4).

Now, under model (3) and model (4), we have three parameters $(\vartheta, r, \rho)$. They capture the sparsity level, signal strength and design correlations, respectively. Our main results are the explicit convergence rates of Hamming error, as a function of $(\vartheta, r, \rho)$, for different methods. We will study six methods: (i) Lasso as in (2); (ii) Elastic net (Zou & Hastie, 2005), which adds an additional $L^2$-penalty to (2), (iii) smoothly clipped absolute deviation (SCAD) (Fan & Li, 2001), which replaces the $L^1$-penalty by a non-convex penalty, (iv) thresholded Lasso (Zhou, 2009), which further thresholds the Lasso solution, and two iterative algorithms, (v) forward selection and (vi) forward backward selection (Huang et al., 2016); see Section 3 for a precise description of each method. To our best knowledge, our results are the first that directly compare Hamming errors of these methods.

## 2 A PREVIEW OF MAIN RESULTS AND SOME DISCUSSIONS

For any $\hat{\beta}$, its Hamming error is $H(\hat{\beta}, \beta) = \sum_{j=1}^{p} 1\{\hat{\beta}_j \neq 0, \beta_j = 0\} + \sum_{j=1}^{p} 1\{\hat{\beta}_j = 0, \beta_j \neq 0\}$. As we shall show, for any of the six methods studied here, there exists a function $h(\vartheta, r, \rho) \in [0, 1]$ such that $\mathbb{E}[H(\hat{\beta}, \beta)] = L_p p^{1-h(\vartheta, r, \rho)}$, where $L_p$ is a *multi-$\log(p)$* term. (A multi-$\log(p)$ term is such that $L_p \cdot p^\epsilon \to \infty$ and $L_p \cdot p^{-\epsilon} \to 0$ for any $\epsilon > 0$.) Since the expected number of true relevant variables is $p^{1-\vartheta}$, we are interested in three cases:

- *Exact recovery:* $h(\vartheta, r, \rho) > 1$. In this case, the expected Hamming error is $o(1)$ as $p \to \infty$. It follows that model selection consistency holds.

- *Almost full recovery:* $\vartheta < h(\vartheta, r, \rho) < 1$. In this case, the expected Hamming error does not vanish as $p \to \infty$, but it is much smaller than the total number of true relevant variables. Variable selection is still satisfactory (although model selection consistency no longer holds).

- *No recovery:* $h(\vartheta, r, \rho) \leq \vartheta$. In this case, the expected Hamming error is comparable with or much larger than the total number of true relevant variables. Variable selection fails.

---

[1]In (1), we assume that each column of $X$ is standardized to have a unit $\ell^2$-norm and that the order of nonzero $\beta_j$ is $\sqrt{\log(n)}$. Alternatively, many works assume that each column of $X$ is standardized to have an $\ell^2$-norm of $\sqrt{n}$ and that the order for nonzero $\beta_j$ is $n^{-1/2}\sqrt{\log(p)}$. These are two equivalent parameterizations.
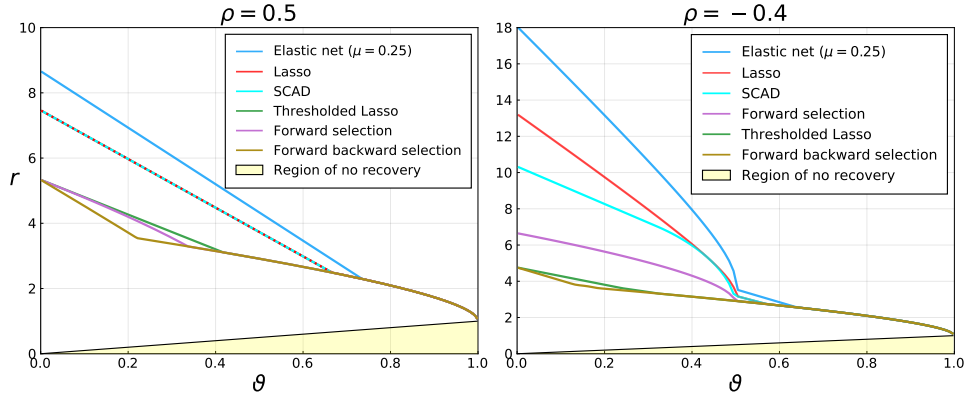
Figure 1: Phase diagrams of six variable selection methods for a block-wise diagonal design. The parameters $(\vartheta, r, \rho)$ characterize the sparsity, signal strength, and design correlations, respectively. For each method, we plot the curve $r = U(\vartheta)$ which separates Region of Almost Full Recovery and Region of Exact Recovery (the lower this curve, the better). Explicit forms of $U(\vartheta)$ are in Section 3. On the left panel, the curves for Lasso and SCAD overlap and are displayed as a dashed line. How to interpret these phase curves are discussed in Section 2.

We call the two-dimensional space $(\vartheta, r)$ the *phase space*. For each fixed $\rho$, the phase space is divided into three regions: *Region of Exact Recovery (ER)*, which is the subset $\{(\vartheta, r) : h(\vartheta, r, \rho) > 1\}$, and *Region of Almost Full Recovery (AFR)* and *Region of No Recovery (NR)* defined similarly. This gives rise to a *phase diagram* for each method. We denote the curve separating ER region and AFR region by $r = U(\vartheta)$ and the curve separating AFR region and NR region by $r = L(\vartheta)$; they are called the *upper* and *lower phase curves*, respectively. The phase diagram and phase curves are convenient ways to visualize the convergence rates of the Hamming error.

Figure 1 shows the phase curves for the six methods (with explicit expressions in the theorems in Section 3). These phase curves depend on the correlation parameter $\rho$. Under our model, for each diagonal block $(j, j+1)$, it holds that $\mathbb{E}[x_j' y | \beta] = \beta_j + \rho\beta_{j+1}$, where $\beta_j, \beta_{j+1} \in \{0, \tau_p\}$. Therefore, a *positive* $\rho$ boosts the signal at each individual site (i.e., $\mathbb{E}[x_j' y | \beta] \geq \beta_j$), while a *negative* $\rho$ leads to potential 'signal cancellation' (i.e., $\mathbb{E}[x_j' y | \beta] \leq \beta_j$). This is why the phase curves have different shapes for positive and negative $\rho$. In Figure 1, we plot the phase curves for $\rho = 0.5$ and $\rho = -0.4$. For other positive/negative value of $\rho$, the patterns are similar.

**Discussion of SCAD**. SCAD is a representative of non-convex penalization methods. There have been inspiring works that demonstrate the advantages of using a non-convex penalty (e.g., Fan & Peng (2004); Loh & Wainwright (2017)). Our results support their insights from a different angle: The phase curve of SCAD is strictly better than that of Lasso, when $\vartheta < 0.5$ and $\rho < 0$. Furthermore, our results illustrate where the advantage of SCAD comes from — compared with Lasso, it handles 'signal cancellation' better. To see this, we recall that 'signal cancellation' only happens for $\rho < 0$. Moreover, under our model (3), the expected number of signal pairs (a signal pair is a diagonal block $\{j, j+1\}$ where both $\beta_j$ and $\beta_{j+1}$ are nonzero) is $\asymp p\epsilon_p^2 = p^{1-2\vartheta}$. Therefore, 'signal cancellation' becomes problematic only when $\vartheta < 0.5$ and $\rho < 0$ both hold. This explains why the phase curves of SCAD and Lasso are the same for the other values of $\vartheta$ and $\rho$. We note that in the previous studies (e.g., Loh & Wainwright (2017)), the advantages of a non-convex penalty in handling 'signal cancellation' are reflected in the weaker conditions of $(X, \beta)$ for achieving model selection consistency. Our results support the advantage of using a non-convex penalty by directly studying the Hamming errors and phase diagrams.

The performance of SCAD can be further improved by adding an entry-wise thresholding on $\hat{\beta}$. We believe that the phase diagrams of *thresholded SCAD* are better than those of SCAD itself, although the extremely tedious analysis impedes us from specific results for now. Also, we are cautious about what to conclude from comparing SCAD and thresholded Lasso. In our settings, Lasso has no model selection consistency mainly because the signals are too weak (i.e., $r$ is not sufficiently large). In such settings, thresholded Lasso outperforms SCAD in terms of the Hamming error. However, there

3

are cases where Lasso has no model selection consistency no matter how large the signal strength is (Zhao & Yu, 2006). For those cases, it is possible that SCAD is better than thresholded Lasso (see Wainwright (2009) for a related study).

**Discussion of Elastic net**. The phase curve of Elastic net is worse than that of Lasso. As we will explain in Section 3.1, Elastic net is a 'bridge' between Lasso and marginal regression in our case. Since the phase curve of marginal regression is always worse than that of Lasso for the blockwise diagonal design, we do not benefit from using Elastic net in the current setting. We must note that Elastic net is motivated by genetic applications where several correlated variables are competing as predictors, and where it is implicitly assumed that groups of correlated variables tend to be all relevant or all irrelevant (Zou & Hastie, 2005). This is not captured by our model (3). Therefore, our results do not go against the benefits of Elastic net known in the literature, but rather our results support that the advantages of Elastic net come from 'group' appearance of signal variables.

**Discussion of thresholded Lasso**. Thresholded Lasso is a representative of improving Lasso by post-processing. There have been inspiring works that demonstrate the advantages of such a post-processing (van de Geer et al., 2011; Wang et al., 2020; Weinstein et al., 2020). Our results support these insights from a different angle. It is surprising (and very encouraging) that the improvement by post-Lasso thresholding is so significant. We note that Lasso is a 1-stage method, which solves a single optimization to obtain $\hat{\beta}$. By comparison, thresholded Lasso is a 2-stage method. Lasso has only one algorithm parameter $\lambda$, while thresholded Lasso has two algorithm parameters $\lambda$ and $t$ (the threshold). In Lasso, we control false positives and false negatives with the same algorithm parameter $\lambda$, and it is sometimes hard to find a value of $\lambda$ that simultaneously controls the two types of errors well. In thresholded Lasso, the two types of errors can be controlled separately by two algorithm parameters. This explains why thresholded Lasso enjoys such a big improvement upon Lasso. It inspires us to modify other 1-stage methods, such as SCAD, by adding a post-processing step of thresholding. For example, we conjecture that thresholded SCAD also has a strictly better phase diagram than that of Lasso, even for a positive $\rho$. On the other hand, thresholding is no free lunch. It leaves one more tuning parameter to be decided in practice. Our theoretical results are based on ideal tuning parameters. How to properly choose these tuning parameters in a data-driven way is an interesting question. Weinstein et al. (2020) proposes a promising approach, where they use cross-validation to select $\lambda$ and FDR control by knockoff to select $t$. We leave it for future work to study the phase diagrams with data-driven tuning parameters.

**Discussion of the two iterative algorithms**. We consider two iterative algorithms, forward selection ('Forward') and forward backward selection ('FB'). The FB algorithm we analyze is a simplified version in Huang et al. (2016), which has only one backward step (after all the forward steps have finished) by thresholding the refitted least-squares solution. Our results show that both methods outperform Lasso, and between these two methods, FB is strictly better than Forward. In the literature, there are very interesting theoretical works showing the advantages of iterative algorithms for variable selection (Donoho et al., 2012; Zhang, 2011). Our results support their insights from a different angle. We discover that, for a wide range of $\rho$, FB has the best phase diagram among all the six methods. This is a very encouraging result. Of course, it is as important to note that the performance of an iterative algorithm tends to be more sensitive to the form of the design, due to its sequential nature.

## 3 MAIN RESULTS

Consider model (1), (3), and (4), where we set $\sigma^2 = 1$ without loss of generality. Let $\mathbb{E}[H(\hat{\beta}, \beta)]$ be the expected Hamming error, where the expectation is with respect to the randomness of $\beta$ and $z$. Let $L_p$ denote a generic multi-$\log(p)$ term such that $L_p p^\epsilon \to \infty$ and $L_p p^{-\epsilon} \to 0$ for any $\epsilon > 0$.

**Theorem 1.** *Under Models (1), (3), and (4), for each of the methods considered in this paper (Lasso, Elastic net, SCAD, thresholded Lasso, forward selection, forward backward selection, as well as marginal regression in Section 3.1), there exists a function $h(\vartheta, r, \rho)$ such that $\mathbb{E}[H(\hat{\beta}, \beta)] = L_p p^{1-h(\vartheta, r, \rho)}$. The explicit expressions of $h(\vartheta, r, \rho)$, which may depend on the tuning parameters of a method, are given in Theorems B.1, C.1, D.1-D.3, F.1, G.1, H.1-H.4 of the supplement.*

In the main article, to save space, we only present the expressions of the upper phase curve $U(\vartheta) = U(\vartheta; \rho)$ and the lower phase curve $L(\vartheta) = L(\vartheta; \rho)$ for each method, which are defined as follows:

$$U(\vartheta; \rho) = \inf\{r > 0 : h(\vartheta, r, \rho) > 1\}, \qquad L(\vartheta; \rho) = \inf\{r > 0 : h(\vartheta, r, \rho) > \vartheta\}. \qquad (5)$$
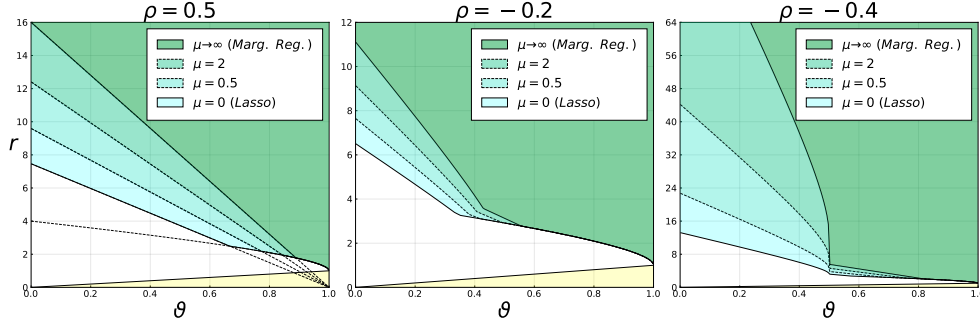
Figure 2: The phase diagrams of Elastic net and its comparison with Lasso (notation: $\eta = \rho/(1+\mu)$).

These two curves describe the phase diagram: The upper phase curve $U(\vartheta)$ separates the ER region and AFR region, and the lower phase curve $L(\vartheta)$ separates the AFR region and NR region.

## 3.1 ELASTIC NET AND LASSO

The Elastic net (Zou & Hastie, 2005) is a method that estimates $\beta$ by

$$\hat{\beta}^{\mathrm{EN}} = \mathrm{argmin}_\beta\{\|y - X\beta\|^2/2 + \lambda\|\beta\|_1 + (\mu/2)\|\beta\|^2\}. \tag{6}$$

Compared with Lasso, it adds an additional $L^2$-penalty to the objective function. Below, we fix $\mu > 0$ and re-parametrize $\lambda = \sqrt{2q\log(p)}$, for some constant $q > 0$. The choice of $q$ affects the exponent, $1 - h(\vartheta, r, \rho)$, in the expression of $\mathbb{E}[H(\hat{\beta}, \beta)]$. We choose the ideal $q$ that minimizes $1 - h(\vartheta, r, \rho)$. The next theorem is proved in the supplement.

**Theorem 2** (Elastic Net). *Under Models* (1), (3), *and* (4), *let* $\hat{\beta}^{\mathrm{EN}}$ *be the Elastic net estimator in* (6). *Fix* $\mu$ *and write* $\eta = \rho/(1 + \mu)$. *Let* $\lambda = \sqrt{2q\log(p)}$ *with an ideal choice of* $q$ *that minimizes the exponent of* $\mathbb{E}[H(\hat{\beta}, \beta)]$. *The phase curves are given by* $L(\vartheta) = \vartheta$, *and*

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

*where* $h_1(\vartheta) = (1 + \sqrt{1 - \vartheta})^2$, $h_2(\vartheta) = \left(\frac{1-|\eta|}{1-|\rho|} + \frac{\sqrt{1+\eta^2-2\rho\eta}}{1-|\rho|}\right)^2(1 - \vartheta)$, $h_3(\vartheta) = \frac{1}{(1-|\rho|)^2}\left(1 + \frac{\sqrt{1+\eta^2-2\rho\eta}}{1+|\eta|}\sqrt{1-2\vartheta}\right)^2$, *and* $h_4(\vartheta) = \frac{1+\eta^2-2\rho\eta}{(1-2|\rho|+\rho\eta)_+^2}\left(\sqrt{1-\vartheta} + \frac{1-|\eta|}{1+|\eta|}\sqrt{1-2\vartheta}\right)^2$.

Lasso is a special case with $\mu = 0$. By setting $\mu = 0$ (equivalently, $\eta = \rho$) in Theorem 2, we obtain the phase curves for Lasso. They agree with the results in Ji & Jin (2012) (but Ji & Jin (2012) does not cover Elastic net).

To see the effect of the $L^2$-penalty, we consider an extreme case where $\mu \to \infty$. Some elementary algebra shows that $(1 + \mu)\hat{\beta}^{\mathrm{EN}}$ converges to the soft-thresholding of $X'y$ at the threshold $\lambda$. In other words, as $\mu \to \infty$, Elastic net converges to marginal screening (i.e., select variables by thresholding the marginal regression coefficients). At the same time, when $\mu = 0$, $(1 + \mu)\hat{\beta}^{\mathrm{EN}}$ equals the Lasso estimate. Hence, Elastic net serves as a bridge between Lasso and marginal regression. In the setting here, the phase diagram of marginal regression is inferior to that of Lasso, and so the phase diagram of Elastic net is also inferior to that of Lasso. See the proposition below and Figure 2:

**Proposition 1.** *In Theorem 2, for each fixed* $\vartheta \in (0, 1)$, *as* $\mu \to 0$, $U(\vartheta)$ *is monotone decreasing and converges to* $U_{\mathrm{Lasso}}(\vartheta)$, *which is the upper phase curve of Lasso; as* $\mu \to \infty$, $U(\vartheta)$ *is monotone increasing and converges to* $U_{\mathrm{MR}}(\vartheta)$, *which is the upper phase curve of marginal regression. Furthermore, when* $\rho \leq -\frac{1}{2}$, $U_{\mathrm{MR}}(\theta) = \infty$ *for all* $0 < \vartheta \leq \frac{1}{2}$ *(i.e. exact recovery is impossible to achieve no matter how large $r$ is).*

## 3.2 SMOOTHLY CLIPPED ABSOLUTE DEVIATION PENALTY (SCAD)

SCAD (Fan & Li, 2001) is a non-convex penalization method. For any $a > 2$, it defines a penalty function $q_\lambda(\theta)$ on $(0, \infty)$ by $q_\lambda(\theta) = \int_0^\theta q_\lambda'(t)dt$, where $q_\lambda'(\theta) = \lambda\{I(\theta \leq \lambda) + \frac{(a\lambda-\theta)_+}{(a-1)\lambda}I(\theta > \lambda)\}$.
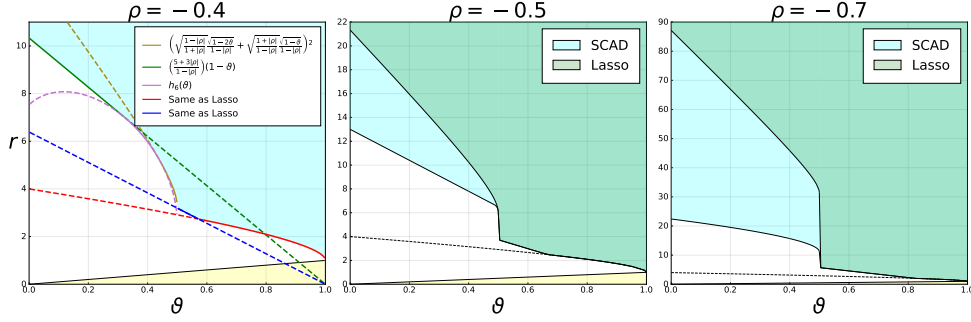
Figure 3: Left: Phase curves of SCAD. Middle and Right: Comparison of SCAD and Lasso.

The resulting penalty function $q_\lambda(\cdot)$ coincides with the $L^1$-penalty in $(0, \lambda]$ and becomes a constant in $[a\lambda, \infty)$. Let $Q_\lambda(\beta) = \sum_{j=1}^p q_\lambda(|\beta_j|)$. Then, SCAD estimates $\beta$ by

$$\hat{\beta}^{\mathrm{SCAD}} = \mathrm{argmin}_\beta \{ \|y - X\beta\|^2/2 + Q_\lambda(\beta) \}. \tag{7}$$

The following theorem is proved in the supplemental material (see Figure 3, left panel):

**Theorem 3** (SCAD). *Under Models* (1), (3), *and* (4), *let* $\hat{\beta}^{\mathrm{SCAD}}$ *be the SCAD estimator in* (7). *Fix* $a \in (2, \frac{2}{1-|\rho|})$. *Let* $\lambda = \sqrt{2q \log(p)}$ *with an ideal choice of $q$ that minimizes the rates of convergence of the expected Hamming error. The phase curves are given by* $L(\vartheta) = \vartheta$, *and*

$$U(\vartheta) = \begin{cases} \max \{h_1(\vartheta), h_2(\vartheta), h_3(\vartheta)\}, & \text{when } \rho \geq 0, \\ \max \{h_1(\vartheta), h_2(\vartheta), h_4(\vartheta), h_5(\vartheta)\}, & \text{when } \rho < 0, \end{cases}$$

*where* $h_1(\vartheta) = (1 + \sqrt{1-\vartheta})^2$, *and* $h_2(\vartheta) = \left(1 + \sqrt{\frac{1+|\rho|}{1-|\rho|}}\right)^2 (1 - \vartheta)$, $h_4(\vartheta) = \left(\sqrt{\frac{1-2\vartheta}{1-|\rho|^2}} + \frac{1}{1-|\rho|}\right)^2$, $h_3(\vartheta) = \left(\frac{3+\rho}{2(1-\rho^2)} \sqrt{\frac{1+\rho}{1-\rho}} \sqrt{1-\vartheta} + \frac{1}{2} \sqrt{\frac{2(1-2\vartheta)}{1+\rho} - \frac{(1-\vartheta)}{(1-\rho)^2}}\right)^2$, *and*

$$h_5(\vartheta) = \begin{cases} \left(\frac{5+3|\rho|}{1-|\rho|}\right)(1 - \vartheta), & \text{if } \sqrt{\frac{1-2\vartheta}{1-\vartheta}} \geq \frac{3-4|\rho|-3\rho^2}{(1-|\rho|)} \sqrt{\frac{1+|\rho|}{5+3|\rho|}}, \\ \frac{1}{(1-|\rho|)^2} \left(\sqrt{\frac{1+|\rho|}{1-|\rho|}} \sqrt{1-\vartheta} + \sqrt{\frac{1-|\rho|}{1+|\rho|}} \sqrt{1-2\vartheta}\right)^2, & \text{if } \sqrt{\frac{1-2\vartheta}{1-\vartheta}} \leq \frac{(1+|\rho|)(1-2|\rho|)}{1-|\rho|}, \\ h_6(\vartheta) & \text{other wise}, \end{cases}$$

$$h_6(\vartheta) = \left\{ \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{\frac{1-2|\rho|}{1-|\rho|} \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \sqrt{\left[\left(\frac{1-2|\rho|}{1-|\rho|}\right)^2 + \frac{1-|\rho|}{1+|\rho|}\right](1-\vartheta) - \frac{1-2\vartheta}{(1+|\rho|)^2}}}{(1-|\rho|)\left[\left(\frac{1-2|\rho|}{1-|\rho|}\right)^2 + \frac{1-|\rho|}{1+|\rho|}\right]} \right\}^2.$$

Note that the phase curves of Lasso are given in Theorem 2 by setting $\eta = \rho$. We compare SCAD with Lasso. When $\rho < 0$, the upper phase curve in Theorem 3 is strictly lower than that of Lasso (see Figure 3, middle and right panels). When $\rho \geq 0$, the upper phase curve in Theorem 3 is sometimes higher than that of Lasso. Note that we restrict $a < \frac{2}{1-|\rho|}$ in Theorem 3. In fact, a larger $a$ may be preferred for $\rho \geq 0$. The next proposition is about using an optimal $a$.

**Proposition 2.** *In the SCAD estimator, we choose $a = a^*$ and $\lambda = \sqrt{2q^* \log(p)}$ such that $(a^*, q^*) = (a^*(\vartheta, r, \rho), q^*(\vartheta, r, \rho))$ minimize the rates of convergence of the expected Hamming error among all choices of $(a, q)$. Let $U^*(\vartheta)$ be the resulting upper phase curve for SCAD. Then, $U(\vartheta) = U_{\mathrm{Lasso}}(\vartheta)$ when $\rho \geq 0$, and $U(\vartheta) < U_{\mathrm{Lasso}}(\vartheta)$ when $\rho < 0$.*

The phase curves of SCAD are insensitive to the choice of $a$. When $a < 0$, the optimal $a^*$ can be any value in $(2, \frac{2}{1-|\rho|})$. When $\rho \geq 0$, there exists a constant $c = c(\vartheta, \rho)$ such that the optimal $a^*$ is any value in $(c, \infty)$. As $a \to \infty$, the SCAD penalty reduces to the $L^1$-penalty. This explains why the phase curve of SCAD is the same as that of Lasso when $\rho \geq 0$.
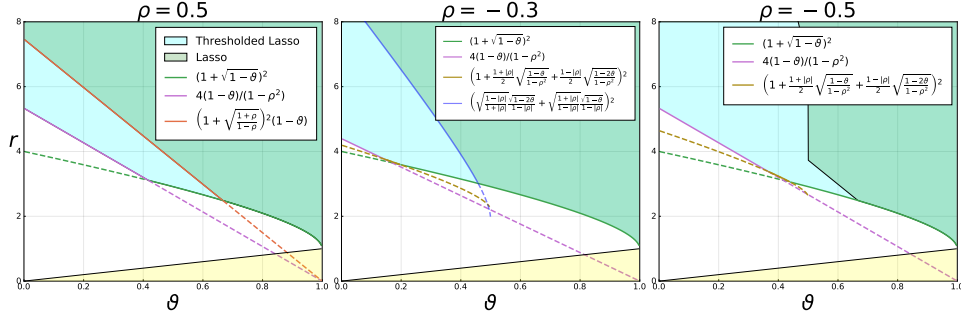
Figure 4: Comparison of the phase diagrams of thresholded Lasso and Lasso.

## 3.3 THRESHOLDED LASSO

Let $\hat{\beta}^{\text{Lasso}}$ be the Lasso estimator in (2). The thresholded Lasso estimator $\hat{\beta}^{\text{TL}}$ is obtained by applying coordinate-wise hard-thresholding to the Lasso estimator:

$$\hat{\beta}_j^{\text{TL}} = \hat{\beta}^{\text{Lasso}} \cdot 1\{|\hat{\beta}^{\text{Lasso}}| > t\}, \qquad 1 \le j \le p. \tag{8}$$

**Theorem 4** (Thresholded Lasso). *Under Models* (1), (3), *and* (4), *let* $\hat{\beta}^{\text{TL}}$ *be the thresholded Lasso estimator in* (8). *Let* $\lambda = \sqrt{2q \log(p)}$ *and* $t = \sqrt{2w \log(p)}$ *with the ideal* $(q, w)$ *that minimize the exponent of the expected Hamming error. The phase curves are given by* $L(\vartheta) = \vartheta$, *and*

$$U(\vartheta) = \begin{cases} \max\{h_1(\vartheta), h_2(\vartheta)\}, & \text{when } \rho \ge 0, \\ \max\{h_1(\vartheta), h_2(\vartheta), h_3(\vartheta)\}, & \text{when } \rho < 0, \end{cases}$$

*where* $h_1(\vartheta) = (1 + \sqrt{1-\vartheta})^2$, $h_2(\vartheta) = \frac{4(1-\vartheta)}{1-\rho^2}$, *and* $h_3(\vartheta) = \left(1 + \frac{1+|\rho|}{2}\sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{1-|\rho|}{2}\sqrt{\frac{1-2\vartheta}{1-\rho^2}}\right)^2$.

See Figure 4 for a comparison with Lasso (a special case of $t = 0$). With the flexibility of using an optimal $t$, the phase diagram of thresholded Lasso is always better than that of Lasso.

Theorem 4 also gives other interesting facts about thresholded Lasso. First, the shape of phase curves is much less affected by the sign of $\rho$. This differs from Lasso, Elastic net, and SCAD, for which the shape of phase curves is significantly different for positive and negative $\rho$. Second, the optimal $\lambda$ in thresholded Lasso is considerably smaller than the optimal $\lambda$ in Lasso (it can be seen from the proofs of Theorem 4 and Theorem 2). This is because the $\lambda$ in thresholded Lasso only serves to control false negatives, but the $\lambda$ in Lasso is used to simultaneously control false positives and false negatives, hence, cannot be too small. We observe the same phenomenon in simulations; see Section 4.

## 3.4 FORWARD SELECTION AND FORWARD BACKWARD SELECTION

Forward selection is a classical textbook method for variable selection. Write $X = [x_1, x_2, \ldots, x_p]$, where $x_i \in \mathbb{R}^n$ for $1 \le i \le p$. For any subset $A \subset \{1, 2, \ldots, p\}$, let $P_A^\perp$ be the projection onto the orthogonal complement of the linear space spanned by $\{x_i : i \in A\}$. Given a threshold $t > 0$, the forward selection algorithm initializes with $S_0 = \emptyset$ and $\hat{r}_0 = y$. At the $k$th iteration, compute

$$i^* = \text{argmax}_{i \notin S_{k-1}} |x_i' \hat{r}_{k-1}|, \qquad \delta = |x_{i^*}' \hat{r}_{k-1}| / \|P_{S_{k-1}}^\perp x_{i^*}\|.$$

If $\delta > t$, compute $S_k = S_{k-1} \cup \{i^*\}$ and $\hat{r}_k = P_{S_k}^\perp y$; otherwise, output $\hat{\beta}^{\text{forward}}$ as the least-squares estimator restricted to $S_{k-1}$. The stopping rule of $\delta \le t$ is equivalent to measuring the decrease of the residual sum of squares. The following theorem is proved in the supplemental material:

**Theorem 5** (Forward Selection). *Under Models* (1), (3), *and* (4), *let* $\hat{\beta}^{\text{forward}}$ *be the estimator from forward selection. Let* $t = \sqrt{2q \log(p)}$ *with the ideal* $q$ *that minimizes the exponent of the expected Hamming error. The phase curves are given by* $L(\vartheta) = \vartheta$, *and*

$$U(\vartheta) = \begin{cases} \max\{h_1(\vartheta), h_2(\vartheta), h_3(\vartheta)\}, & \text{when } \rho \ge 0, \\ \max\{h_1(\vartheta), h_2(\vartheta), h_3(\vartheta), h_4(\vartheta)\}, & \text{when } \rho < 0, \end{cases}$$

*with* $h_1(\vartheta) = (1 + \sqrt{1-\vartheta})^2$, $h_2(\vartheta) = \frac{2(1-\vartheta)}{1-|\rho|}$, $h_3(\vartheta) = \frac{(1+\sqrt{1-2\vartheta})^2}{1-\rho^2}$, $h_4(\vartheta) = \left(\sqrt{\frac{1-2\vartheta}{2(1-|\rho|)}} + \frac{1}{1-|\rho|}\right)^2$.
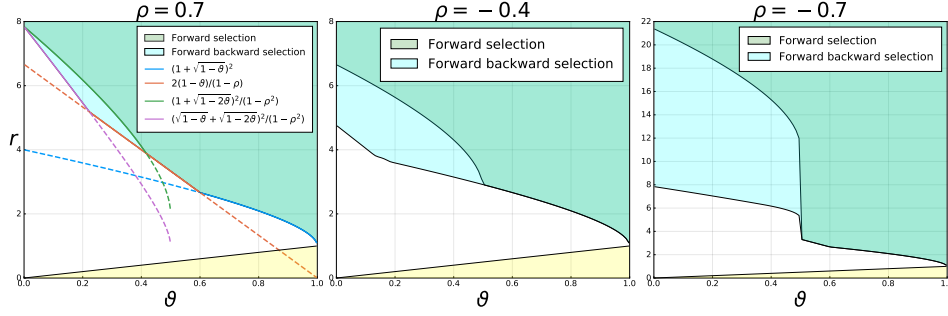
Figure 5: The phase diagrams of forward selection and forward backward selection.

Forward backward selection (FB) modifies forward selection by allowing to drop variables. We use the FB algorithm in Huang et al. (2016), where the backward step is conducted after all the forward steps are finished. For a threshold $v > 0$, it applies entry-wise thresholding on $\hat{\beta}^{\text{forward}}$:

$$\hat{\beta}_j^{\text{FB}} = \hat{\beta}_j^{\text{forward}} \cdot 1\{|\hat{\beta}_j^{\text{forward}}| > v\}, \qquad 1 \leq j \leq p. \tag{9}$$

**Theorem 6** (Forward Backward Selection). *Under Models* (1), (3), *and* (4), *let* $\hat{\beta}^{\text{FB}}$ *be the estimator from forward selection. Let* $t = \sqrt{2q \log(p)}$ *and* $v = \sqrt{2u \log(p)}$ *with the ideal* $(q, u)$ *that minimize the exponent of the expected Hamming error. When* $\rho \geq 0$, *the phase curves are given by* $L(\vartheta) = \vartheta$, *and*

$$U(\vartheta) = \max\{h_1(\vartheta), h_2(\vartheta), h_3^*(\vartheta)\},$$

*where* $h_1(\vartheta)$ *and* $h_2(\vartheta)$ *are the same as in Theorem 5 and* $h_3^*(\vartheta) = \frac{(\sqrt{1-\vartheta}+\sqrt{1-2\vartheta})^2}{1-\rho^2}$. *When* $\rho < 0$,

$$U(\vartheta) \leq \max\{g_1(\vartheta), g_2(\vartheta), g_3(\vartheta), g_4(\vartheta)\},$$

*where* $g_1(\vartheta) = (v_{\min}(\vartheta) + \sqrt{1-\vartheta})^2$, $g_2(\vartheta) = \frac{2(1-\vartheta)}{1-|\rho|}$, $g_3(\vartheta) = \left(\sqrt{\frac{1-2\vartheta}{1-\rho^2}} + v_{\min}(\vartheta)\right)^2$, $g_4(\vartheta) = \left(\sqrt{\frac{1-2\vartheta}{2(1-|\rho|)}} + \frac{t_{\min}(\vartheta)}{1-|\rho|}\right)^2$, $v_{\min}(\vartheta) = \max\{1, \sqrt{\frac{1-\vartheta}{1-\rho^2}}\}$, *and* $t_{\min}(\vartheta) = \max\{\frac{\sqrt{2}}{2}, \frac{v_{\min}(\vartheta)}{1+|\rho|/\sqrt{1-\rho^2}}\}$.

Theorem 6 gives $U(\vartheta)$ for $\rho \geq 0$ and an upper bound of it for $\rho < 0$. Combining it with Theorems 2 and 5, we conclude that the upper phase curve of FB is always better than those of Lasso and forward selection (for $\rho < 0$, the upper bound here is already better than $U(\vartheta)$ for the other two methods).

We remark that we did obtain the exact phase curve for $\rho < 0$ in the proof of Theorem 6. It is just too complicated and space-consuming to present it in the main text. However, given specific values of $(\vartheta, \rho)$, we can always plot the exact phase curve using the (complicated) formulas in the supplement. In Figures 1 and 5, the phase curves of FB are indeed the exact ones.

### 3.5 CONNECTION TO THE RANDOM DESIGN MODEL

Consider the random design as mentioned in Section 1. The *minimax Hamming error* is $H^*(\vartheta, r, \rho) = \inf_{\hat{\beta}} \mathbb{E}[H(\hat{\beta}, \beta)]$, where the infimum is taken over all methods $\hat{\beta}$ and the expectation is with respect to the randomness of $(X, \beta, z)$. We can define $H^*(\vartheta, r, \rho)$ in the same way for our current model (4). The minimax Hamming error is related to the statistical limit of the model setting, but not any specific method. The next theorem shows that, when $n \gg s_p = p^{1-\vartheta}$ (we allow both $p \leq n$ and $p > n$), the convergence rate of the minimax Hamming error is the same under two models.

**Theorem 7.** *Under Models* (1) *and* (3), *suppose* $X$ *is independent of* $(\beta, z)$ *and its rows are iid generated from* $\mathcal{N}(0, n^{-1}\Sigma)$, *with* $\Sigma$ *having the same form as* $G$ *in* (4). *Suppose* $n = p^\omega$, *with* $\omega > 1 - \vartheta$ *(note: this allows* $\omega < 1$, *which corresponds to* $n \ll p$). *There exists a number* $h^{**}(\vartheta, r, \rho)$ *such that the minimax Hamming error satisfies that* $H^*(\vartheta, r, \rho) = L_p p^{1-h^{**}(\vartheta, r, \rho)}$. *Furthermore, if we instead have* $X'X = \Sigma$ *(i.e., model* (4)), *then it also holds that* $H^*(\vartheta, r, \rho) = L_p p^{1-h^{**}(\vartheta, r, \rho)}$.

## 4 SIMULATIONS

In Experiments 1-3, $(n, p) = (1000, 300)$. In Experiment 4, $(n, p) = (500, 1000)$.

**Experiment 1** (block-wise diagonal designs). We generate $(X, \beta)$ as in (3)-(4). For each method, we select the ideal tuning parameters that minimize the average Hamming error over 50 repetitions. The averaged Hamming errors and its standard deviations under the ideal tuning parameters over 500 repetitions are reported below. The results are consistent with the theoretical phase diagrams (see Figure 1). E.g., thresholded Lasso and forward backward selection are the two methods that perform the best; Lasso is more unsatisfactory when $\rho < 0$; SCAD improves Lasso when $\rho < 0$.

| $\rho$ | $\vartheta$ | $r$ | Lasso | ThresLasso | ElasticNet | SCAD | Forward | FoBackward |
|---|---|---|---|---|---|---|---|---|
| 0.5 | 0.1 | 1.5 | 11.57 (3.59) | **10.48** (3.34) | 11.57 (3.31) | 11.72 (3.33) | 14.88 (4.12) | 13.35 (3.90) |
| 0.5 | 0.1 | 4 | 1.00 (1.00) | **0.42** (0.65) | 1.03 (1.00) | 1.00 (0.96) | 0.66 (0.84) | 0.51 (0.73) |
| -0.5 | 0.1 | 1.5 | 35.62 (5.09) | 15.62 (4.06) | 35.48 (5.64) | 25.87 (5.04) | 19.48 (4.61) | **14.82** (3.82) |

Table 1: Experiment 1 (block-diagonal designs). $(n, p) = (1000, 300)$.

**Experiment 2** (general designs). In the Toeplitz design, we let $(X'X)_{i,j} = 0.7^{|i-j|}$ and set $(\vartheta, r) = (0.1, 2.5)$. In the factor model design, we let $X'X = BB' - \text{diag}(BB') + I_p$, where entries of $B \in \mathbb{R}^{p \times 2}$ are *iid* from $\text{Unif}(0, 0.6)$, and set $(\vartheta, r) = (0.1, 1.5)$. Same as in Experiment 1, we use the ideal tuning parameters. The averaged Hamming errors and its standard deviations are reported below. The Toeplitz design is a setting where each variable is only highly correlated with a few other variables. The factor model design is a setting where a variable is (weakly) correlated with all the other variables. The results are quite similar to those in Experiment 1. This confirms that the insight gained in the study of the block-wise diagonal design continues to apply to more general designs.

| design | Lasso | ThresLasso | ElasticNet | SCAD | Forward | FoBackward |
|---|---|---|---|---|---|---|
| Toeplitz | 47.15 (6.32) | **22.02** (5.31) | 47.40 (6.41) | 24.61 (5.70) | 30.77 (6.18) | 22.93 (5.44) |
| Factor model | 21.14 (4.52) | **15.90** (3.87) | 21.20 (4.45) | 19.68 (4.23) | 20.04 (4.34) | 16.13 (3.76) |

Table 2: Experiment 2 (general designs). $(n, p) = (1000, 300)$.

**Experiment 3** (tuning parameters). Fix $(\vartheta, r) = (0.1, 1.5)$ and $\rho \in \{\pm 0.5\}$ in the block-wise diagonal design. We study the effect of tuning parameters in Lasso, thresholded Lasso (ThreshLasso), forward selection (ForwardSelect), and forward backward selection (FB). In (a)-(b), we show the heatmap of averaged Hamming error (over 50 repetitions) of ThreshLasso for a grid of $(t, \lambda)$; when $t = 0$, it reduces to Lasso. In (c)-(d), we show the Hamming error of FB for a grid of $(v, t)$; when $v = 0$, it reduces to ForwardSelect. Cyan points are theoretically optimal tuning parameters (formulas are in proofs of theorems). Red points are empirically optimal tuning parameters that minimize the averaged Hamming error. The theoretical tuning parameter values are quite close to the empirically optimal ones. Moreover, the optimal $\lambda$ in ThreshLasso is smaller than the optimal $\lambda$ in Lasso.



(a) Lasso and ThreshLasso ($\rho = 0.5$)  (b) Lasso and ThreshLasso ($\rho = -0.5$)  (c) ForwardSelect and FB ($\rho = 0.5$)  (d) ForwardSelect and FB ($\rho = -0.5$)
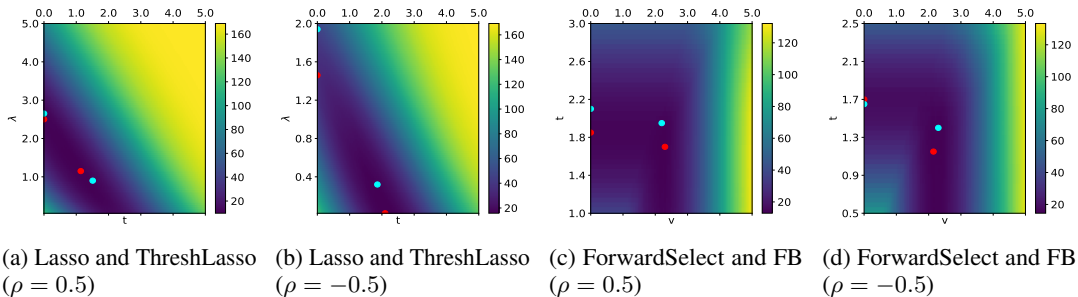
Figure 6: Experiment 3 (effects of tuning parameters). In all plots, cyan points are computed from the formulas in our theory, and red points are the empirically best tuning parameters (they minimize the average Hamming error over 500 repetitions). In (a)-(b), the cyan/red points with $t = 0$ correspond to Lasso, and the other two are for thresholded Lasso. In (c)-(d), the cyan/red points with $t = 0$ correspond to forward selection, and the other two are for forward backward selection.

**Experiment 4** ($p > n$ and random designs). Fix $(n, p, \vartheta, r) = (500, 1000, 0.5, 1.5)$. We simulate data from the random design setting in Theorem 7. We study the average Hamming error over 500

repetitions (tuning parameters are set in the same way as in Experiment 1). See Table 3. We have some similar observations as before: e.g., ThreshLasso and FoBackward are still the best two,

| $\rho$ | $\vartheta$ | $r$ | Lasso | ThresLasso | ElasticNet | SCAD | Forward | FoBackward |
|---|---|---|---|---|---|---|---|---|
| 0.5 | 0.5 | 1.5 | 16.02 (5.52) | **9.83** (4.08) | 13.92 (5.12) | 15.98 (6.28) | 11.74 (5.55) | 9.84 (4.93) |
| -0.5 | 0.5 | 1.5 | 18.49 (6.03) | 10.50 (4.23) | 15.18 (5.64) | 18.12 (6.00) | 12.00 (5.67) | **10.41** (5.03) |

Table 3: Experiment 4 ($p > n$ and random designs).

## 5 CONCLUSION

Most papers on variable selection focus on one method and study its properties in a relatively broad setting. In contrast, we focus on a relatively narrow setting but study a variety of different methods. Our motivation is to facilitate a direct comparison of main-stream approaches for variable selection. Although the model we use seems idealized, by varying the parameters, it already accommodates many different combinations of sparsity level, signal strength, and design correlation level. Under this model, we derive tractable forms of the Hamming error and phase diagram for each method, and we make notable discoveries out of these theoretical results.

## 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*, pp. 2533–2556, 2011.

Peter J Bickel, Ya'acov Ritov, and Alexandre B Tsybakov. Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, 37(4):1705–1732, 2009.

Alia Dehman, Christophe Ambroise, and Pierre Neuvial. Performance of a blockwise approach in variable selection using linkage disequilibrium information. *BMC Bioinformatics*, 16(1):1–14, 2015.

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

David L Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

David L Donoho, Yaakov Tsaig, Iddo Drori, and Jean-Luc Starck. Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 58(2):1094–1121, 2012.

Bradley Efron. Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. *Journal of the American Statistical Association*, 99(465):96–104, 2004.

Jianqing Fan and Runze Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360, 2001.

Jianqing Fan and Jinchi Lv. A selective overview of variable selection in high dimensional feature space. *Statistica Sinica*, 20(1):101, 2010.

Jianqing Fan and Heng Peng. Nonconcave penalized likelihood with a diverging number of parameters. *The Annals of Statistics*, 32(3):928–961, 2004.

Jianqing Fan, Yuan Liao, and Xiaofeng Shi. Risks of large portfolios. *Journal of Econometrics*, 186 (2):367–387, 2015.

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

Shiqiong Huang, Jiashun Jin, and Zhigang Yao. Partial correlation screening for estimating large precision matrices, with applications to classification. *The Annals of Statistics*, 44(5):2018–2057, 2016.

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. Comment on "Estimating false discovery proportion under arbitrary covariance dependence" by Fan et al. *Journal of the American Statistical Association*, 107(499):1042–1045, 2012.

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

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. Power of FDR control methods: The impact of ranking algorithm, tampered design, and symmetric statistic. *arXiv preprint arXiv:2010.08132*, 2020.

Po-Ling Loh and Martin J Wainwright. Support recovery without incoherence: A case for nonconvex regularization. *The Annals of Statistics*, 45(6):2455–2482, 2017.

Nicolai Meinshausen and Peter Bühlmann. Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4):417–473, 2010.

Xiaotong Shen, Wei Pan, and Yunzhang Zhu. Likelihood-based selection and sharp parameter estimation. *Journal of the American Statistical Association*, 107(497):223–232, 2012.

Wenguang Sun and T Tony Cai. Oracle and adaptive compound decision rules for false discovery rate control. *Journal of the American Statistical Association*, 102(479):901–912, 2007.

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

Sara van de Geer, Peter Bühlmann, and Shuheng Zhou. The adaptive and the thresholded lasso for potentially misspecified models (and a lower bound for the lasso). *Electronic Journal of Statistics*, 5:688–749, 2011.

Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing: Theory and Applications*, pp. 210–268. Cambridge Univ. Press, 2012.

Martin J Wainwright. Sharp thresholds for high-dimensional and noisy sparsity recovery using $L_1$-constrained quadratic programming (lasso). *IEEE Transactions on Information Theory*, 55(5): 2183–2202, 2009.

Shuaiwen Wang, Haolei Weng, and Arian Maleki. Which bridge estimator is optimal for variable selection? *The Annals of Statistics*, 48(5), 2020.

Asaf Weinstein, Weijie 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*, 2020.

Cun-Hui Zhang. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2):894–942, 2010.

Tong Zhang. Adaptive forward-backward greedy algorithm for learning sparse representations. *IEEE Transactions on Information Theory*, 57(7):4689–4708, 2011.

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

Shuheng Zhou. Thresholding procedures for high dimensional variable selection and statistical estimation. *Advances in Neural Information Processing Systems*, 22:2304–2312, 2009.

Hui Zou. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429, 2006.

Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.

# A    SKETCH OF THE PROOF IDEAS

We use a similar proof idea for every main theorem, which we explain as follows. To obtain the phase diagram, the key is deriving the rate of convergence of the expected Hamming error $\mathbb{E}[H(\hat{\beta}, \beta)]$. Let

$$\text{FP}_p = \sum_{j=1}^{p} \mathbb{P}(\beta_j = 0, \hat{\beta}_j \neq 0), \qquad \text{and} \qquad \text{FN}_p = \sum_{j=1}^{p} \mathbb{P}(\beta_j = \tau_p, \hat{\beta}_j = 0). \qquad (10)$$

By definition,

$$\mathbb{E}[H(\hat{\beta}, \beta)] = \text{FP}_p + \text{FN}_p.$$

Suppose $j$ is in the diagonal block $\{j, j+1\}$ of the Gram matrix $G$. For most methods (except for forward selection and forward backward selection, which we discuss separately), it is easy to see that $\hat{\beta}_j$ does not depend on any other $\beta_i$ with $i \notin \{j, j+1\}$. It follows that

$$
\begin{aligned}
\mathbb{P}(\beta_j = 0, \hat{\beta}_j \neq 0) &= \mathbb{P}(\beta_j = 0, \beta_{j+1} = 0, \hat{\beta}_j \neq 0) + \mathbb{P}(\beta_j = 0, \beta_{j+1} = \tau_p, \hat{\beta}_j \neq 0) \\
&= (1 - \epsilon_p)^2 \, \mathbb{P}(\hat{\beta}_j \neq 0 | \beta_j = 0, \beta_{j+1} = 0) \\
&\quad + (1 - \epsilon_p)\epsilon_p \cdot \mathbb{P}(\hat{\beta}_j \neq 0 | \beta_j = 0, \beta_{j+1} = \tau_p) \\
&= L_p \, \mathbb{P}_{00}(\hat{\beta}_j \neq 0) + L_p p^{-\vartheta} \, \mathbb{P}_{01}(\hat{\beta}_j \neq 0),
\end{aligned}
$$

where $\mathbb{P}_{00}$ is the conditional probability conditioning on $(\beta_j, \beta_{j+1}) = (0, 0)$ and $\mathbb{P}_{01}$ is the conditional probability conditioning on $(\beta_j, \beta_{j+1}) = (0, \tau_p)$. Similarly, we can derive

$$\mathbb{P}(\beta_j = \tau_p, \hat{\beta}_j = 0) = L_p p^{-\vartheta} \, \mathbb{P}_{10}(\hat{\beta}_j = 0) + L_p p^{-2\vartheta} \, \mathbb{P}_{11}(\hat{\beta}_j = 0),$$

where $\mathbb{P}_{10}$ is the conditional probability conditioning on $(\beta_j, \beta_{j+1}) = (\tau_p, 0)$ and $\mathbb{P}_{11}$ is the conditional probability conditioning on $(\beta_j, \beta_{j+1}) = (\tau_p, \tau_p)$. When $p$ is even, by symmetry in this design, the above expressions do not change with $j$. When $p$ is odd, this is true except for $j = p$; however, this single $j$ has a negligible effect on the expected Hamming error. We thus have

$$
\begin{aligned}
\mathbb{E}[H(\hat{\beta}, \beta)] =\ & L_p p \cdot \mathbb{P}_{00}(\hat{\beta}_j \neq 0) + L_p p^{1-\vartheta} \cdot \mathbb{P}_{01}(\hat{\beta}_j \neq 0) \\
& + L_p p^{1-\vartheta} \cdot \mathbb{P}_{10}(\hat{\beta}_j = 0) + L_p p^{1-2\vartheta} \cdot \mathbb{P}_{11}(\hat{\beta}_j = 0).
\end{aligned}
\qquad (11)
$$

It remains to study the probabilities in (11). Let $\tilde{y}_j = x_j' y / \sqrt{2 \log(p)}$ and $\tilde{y}_{j+1} = x_{j+1}' y / \sqrt{2 \log(p)}$. For most methods considered in this paper, $\hat{\beta}_j$ is determined by $(\tilde{y}_j, \tilde{y}_{j+1})$ only. Define

$$\mathcal{R} = \{(h_1, h_2) \in \mathbb{R}^2 : \ (\tilde{y}_j, \tilde{y}_{j+1}) = (h_1, h_2) \text{ implies that } \hat{\beta}_j \neq 0\}. \qquad (12)$$

Write $\tilde{y} = (\tilde{y}_1, \tilde{y}_2)'$. Then, we can re-write (11) as

$$
\begin{aligned}
\mathbb{E}[H(\hat{\beta}, \beta)] =\ & L_p p \cdot \mathbb{P}_{00}(\tilde{y} \in \mathcal{R}) + L_p p^{1-\vartheta} \cdot \mathbb{P}_{01}(\tilde{y} \in \mathcal{R}) \\
& + L_p p^{1-\vartheta} \cdot \mathbb{P}_{10}(\tilde{y} \notin \mathcal{R}) + L_p p^{1-2\vartheta} \cdot \mathbb{P}_{11}(\tilde{y} \notin \mathcal{R}).
\end{aligned}
\qquad (13)
$$

In the settings of interest in this paper, conditioning on each realization of $(\beta_j, \beta_{j+1})$, it can be shown that

$$\tilde{y} \ \sim \ \mathcal{N}_2\Big(\mu, \ \frac{1}{2\log(p)}\Sigma\Big), \qquad \text{for some fixed } \mu \in \mathbb{R}^2 \text{ and } \Sigma \in \mathbb{R}^{2 \times 2}.$$

For any $x \in \mathbb{R}^2$ and $S \subset \mathbb{R}^2$, define

$$d_\Sigma^2(x, S) = \inf_{v \in S}\big\{(x - v)'\Sigma^{-1}(x - v)\big\}. \qquad (14)$$

We apply Lemma 6.1 in Ke et al. (2020) to get that, as $p \to \infty$, ($L_p$ denotes a multi-$\log(p)$ term; see Section 3 or the notations below)

$$\mathbb{P}(\tilde{y} \in \mathcal{R}) = L_p p^{-d_\Sigma^2(\mu, \mathcal{R})}, \qquad \mathbb{P}(\tilde{y} \notin \mathcal{R}) = L_p p^{-d_\Sigma^2(\mu, \mathcal{R}^c)}. \qquad (15)$$

Combining (15) with (13), we can get the rate of convergence of the expected Hamming error, if we calculate the following quantities:

- The set $\mathcal{R}$ (we call it "rejection region"). The rejection region depends on the definition of the method and the choice of tuning parameters.

- The distances $d_\Sigma(\mu, \mathcal{R})$ and $d_\Sigma(\mu, \mathcal{R}^c)$. Note that $(\mu, \Sigma)$ depend on the realization of $(\beta_j, \beta_{j+1})$. Therefore, we need to calculate $(\mu, \Sigma)$ for each of the four possible realizations.

In the remaining of this supplemental material, we prove Theorems 2-6 and Proposition 1-2. For each theorem, the proof can be divided into three parts:

(a) Derive the rejection region $\mathcal{R}$.

(b) Apply (13)-(15) to calculate the rate of convergence of $\mathbb{E}[H(\hat\beta, \beta)]$.

(c) Calculate the phase diagram based on the result from (b).

Throughout the proof, we use $L_p$ to denote a generic multi-$\log(p)$ term, which satisfies that $L_p p^\epsilon \to \infty$ and $L_p p^{-\epsilon} \to 0$ for any $\epsilon > 0$. We also frequently use the notation:

**Definition 1.** *For $\rho \in (-1, 1)$ and $u, v \in \mathbb{R}^2$, define $d_\rho(u, v) > 0$ by $d_\rho^2(u, v) = (u_1 - v_1)^2 + (u_2 - v_2)^2 - 2\rho(u_1 - v_1)(u_2 - v_2)$.*

In our proofs, we also frequently calculate the infimum of $d_\rho^2(u, v)$, for $v$ a line in $\mathbb{R}^2$. The following lemma is very useful. Its proof is elementary and thus omitted.

**Lemma A.1.** *Fix $\rho \in (-1, 1)$. Given real numbers $A, B, C$ such that $AB \neq 0$, consider a constrained optimization over $x = (x_1, x_2)$ that minimizes $d_\rho^2(x, (0,0)) = x_1^2 + x_2^2 - 2\rho x_1 x_2$ subject to the constraint $Ax_1 + Bx_2 + C = 0$. The solution is $x_1^* = \frac{-C(A+\rho B)}{A^2+B^2+2\rho AB}$ and $x_2^* = \frac{-C(B+\rho A)}{A^2+B^2+2\rho AB}$, and the objective function evaluated at $x^* = (x_1^*, x_2^*)$ is*

$$d_\rho^2(x^*, (0,0)') = \frac{C^2(1-\rho^2)}{A^2 + B^2 + 2\rho AB}.$$

## B   PROOF OF THEOREM 2 (ELASTIC NET)

As described in Section A, our proof has three parts: (a) deriving the rejection region, (b) obtaining the rate of convergence of $\mathbb{E}[H(\hat\beta, \beta)]$, and (c) calculating the phase diagram.

**Part 1: Deriving the rejection region.**   Recall that the rejection region $\mathcal{R}$ is as defined in (12). Write $h_1 = x_j' y / \sqrt{2\log(p)}$, $h_2 = x_{j+1}' y / \sqrt{2\log(p)}$, and $\lambda = \sqrt{2q\log(p)}$. Consider a bivariate Elastic net problem, where $(\hat b_1, \hat b_2)$ minimizes

$$L(b) \equiv \frac{1}{2} b' \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} b + b'h + \sqrt{q}\|b\|_1 + \frac{1}{2}\mu\|b\|^2. \tag{16}$$

It is seen that $(\hat\beta_j, \hat\beta_{j+1}) = \sqrt{2\log(p)}(\hat b_1, \hat b_2)$. Hence, $\mathcal{R}$ consists of all values of $h$ such that $\hat b_1 \neq 0$.

Fix $\rho \geq 0$. The next lemma gives the explicit solution to (16) in the case of $h_1 > |h_2|$. It is proved in Section B.1.

**Lemma B.1** (Solution path of Elastic net). *Consider the optimization in (16). Suppose $h_1 > |h_2| \geq 0$. Write $\eta = \rho/(1+\mu)$.*

- *When $\sqrt{q} \geq h_1$, we have $\hat b_1 = \hat b_2 = 0$.*

- *If $h_2 \geq \eta h_1$, when $\frac{h_2 - \eta h_1}{1 - \eta} \leq \sqrt{q} < h_1$, we have $\hat b_1 = \frac{h_1 - \sqrt{q}}{1 + \mu}$, and $\hat b_2 = 0$;*

  *When $\sqrt{q} < \frac{h_2 - \eta h_1}{1 - \eta}$, we have*

  $$\hat b_1 = \frac{\frac{h_1 - \sqrt{q}}{1+\mu} - \eta \frac{h_2 - \sqrt{q}}{1+\mu}}{1 - \eta^2}, \qquad \hat b_2 = \frac{\frac{h_2 - \sqrt{q}}{1+\mu} - \eta \frac{h_1 - \sqrt{q}}{1+\mu}}{1 - \eta^2};$$

- *if $h_2 < \eta h_1$, when $\frac{-h_2 + \eta h_1}{1 + \eta} \leq \sqrt{q} < h_1$, we have $\hat b_1 = \frac{h_1 - \sqrt{q}}{1 + \mu}$, and $\hat b_2 = 0$;*

14

*When $\sqrt{q} < \frac{-h_2 + \eta h_1}{1 + \eta}$, we have*

$$\hat{b}_1 = \frac{\frac{h_1 - \sqrt{q}}{1 + \mu} - \eta \frac{h_2 + \sqrt{q}}{1 + \mu}}{1 - \eta^2}, \quad \hat{b}_2 = \frac{\frac{h_2 + \sqrt{q}}{1 + \mu} - \eta \frac{h_1 - \sqrt{q}}{1 + \mu}}{1 - \eta^2}.$$

We now use Lemma B.1 to derive $\mathcal{R}$. Partition $\mathbb{R}^2$ into 4 non-overlapping regions:

$$M_1 = \{(h_1, h_2) : h_1 > |h_2|\}, \qquad M_2 = \{(h_1, h_2) : h_1 < -|h_2|\},$$
$$M_3 = \{(h_1, h_2) : h_2 > |h_1|\}, \qquad M_4 = \{(h_1, h_2) : h_2 < -|h_1|\}.$$

First, we derive $\mathcal{R} \cap M_1$. By Lemma B.1, as $\sqrt{q}$ decreases from $\infty$ to $0$, $\hat{b}_1$ is initially zero and then becomes positive when $\sqrt{q}$ hits $h_1$ (second bullet point of this lemma). Then, if we further decrease $\sqrt{q}$, the value of $\hat{b}_1$ is always increasing (third bullet point of this lemma) and remains positive. Therefore, $\sqrt{q} < h_1$ is the sufficient and necessary condition for $\hat{b}_1$ to be nonzero. It follows that

$$\mathcal{R} \cap M_1 = M_1 \cap \{(h_1, h_2) : h_1 > \sqrt{q}\}.$$

Second, we consider $\mathcal{R} \cap M_2$. Note that $(h_1, h_2) \in \mathcal{R} \cap M_2$ if and only if $(-h_1, -h_2) \in \mathcal{R} \cap M_1$. Additionally, if we simultaneously flip the sign of $(h_1, h_2, b_1, b_2)$, the objective in (16) is unchanged. It follows that
$$\mathcal{R} \cap M_2 = \{(h_1, h_2) : (-h_1, -h_2) \in \mathcal{R} \cap M_1\}.$$

Next, we derive $\mathcal{R} \cap M_3$. Note that $(h_1, h_2) \in M_3$ if and only if $(h_2, h_1) \in M_1$. Moreover, if we swap $(h_1, b_1)$ with $(h_2, b_2)$, the objective in (16) is unchanged. Hence, we can obtain $\mathcal{R} \cap M_3$ as follows: We first find the collection of $(h_1, h_2) \in \mathcal{R} \cap M_1$ such that $\hat{b}_2 \neq 0$, and then switch the two coordinates $h_1$ and $h_2$ to get $\mathcal{R} \cap M_3$. To this end, by Lemma B.1, for $(h_1, h_2) \in \mathcal{R} \cap M_1$, $\hat{b}_2 \neq 0$ if either $h_2 - \eta h_1 > \sqrt{q}(1 - \eta)$ or $h_2 - \eta h_1 < -\sqrt{q}(1 + \eta)$. It follows that, for $(h_1, h_2) \in \mathcal{R} \cap M_3$, $\hat{b}_1 \neq 0$ if either $h_1 - \eta h_2 > \sqrt{q}(1 - \eta)$ or $h_1 - \eta h_2 < -\sqrt{q}(1 + \eta)$. It implies that

$$\mathcal{R} \cap M_3 = M_3 \cap \left( \{(h_1, h_2) : h_1 - \eta h_2 > \sqrt{q}(1 - \eta)\} \cup \{(h_1, h_2) : h_1 - \eta h_2 < -\sqrt{q}(1 + \eta)\} \right).$$

Last, we obtain $\mathcal{R} \cap M_4$ by

$$\mathcal{R} \cap M_4 = \{(h_1, h_2) : (-h_1, -h_2) \in \mathcal{R} \cap M_3\}.$$

Combining the above results gives

$$\begin{aligned}
\mathcal{R} = {} & \{(h_1, h_2) : h_1 - \eta h_2 > \sqrt{q}(1 - \eta), \; h_1 > \sqrt{q}\} \\
& \cup \{(h_1, h_2) : h_1 - \eta h_2 > \sqrt{q}(1 + \eta)\} \cup \{(h_1, h_2) : h_1 - \eta h_2 < -\sqrt{q}(1 + \eta)\} \\
& \cup \{(h_1, h_2) : h_1 - \eta h_2 < -\sqrt{q}(1 - \eta), \; h_1 < -\sqrt{q}\}.
\end{aligned} \tag{17}$$

See Figure 7 for a visualization of the rejection region (recall that $\eta = \rho/(1 + \mu)$).

Figure 7 only depicts the rejection region for $\rho \geq 0$. For $\rho < 0$, we can similarly draw the rejection region, but it is not necessary for the proof of this theorem. In Part 2, we will see that, by carefully utilizing the symmetry in our problem, we can derive the rate of convergence of the Hamming error for $\rho < 0$ without deriving the rejection region directly.

**Part 2. Analyzing the Hamming error.** We aim to use (13)-(15) to derive the rate of convergence of $\mathbb{E}[H(\hat{\beta}, \beta)]$. Recall that $\tilde{y}_1 = x_j' y / \sqrt{2 \log(p)}$ and $\tilde{y}_2 = x_{j+1}' y / \sqrt{2 \log(p)}$. It is easy to see that $\tilde{y} \sim \mathcal{N}_2(\mu, \frac{1}{2 \log(p)} \Sigma)$, where the covariance matrix $\Sigma$ is the $2 \times 2$ matrix with 1 in the diagonal and $\rho$ in the off-diagonal, and the vector $\mu$ is equal to

$$\mu_{00} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \mu_{01} = \begin{bmatrix} \rho \sqrt{r} \\ \sqrt{r} \end{bmatrix}, \quad \mu_{10} = \begin{bmatrix} \sqrt{r} \\ \rho \sqrt{r} \end{bmatrix}, \quad \mu_{11} = \begin{bmatrix} (1 + \rho)\sqrt{r} \\ (1 + \rho)\sqrt{r} \end{bmatrix},$$

when $(\beta_j, \beta_{j+1})$ takes the value of $(0, 0)$, $(0, \tau_p)$, $(\tau_p, 0)$, and $(\tau_p, \tau_p)$, respectively. By (13)-(15), $\mathbb{E}[H(\hat{\beta}, \beta)] = \text{FP}_p + \text{FN}_p$, where

$$\begin{aligned}
\text{FP}_p &= L_p p^{1 - d_\Sigma^2(\mu_{00}, \mathcal{R})} + L_p p^{1 - \vartheta - d_\Sigma^2(\mu_{01}, \mathcal{R})}, \\
\text{FN}_p &= L_p p^{1 - \vartheta - d_\Sigma^2(\mu_{10}, \mathcal{R}^c)} + L_p p^{1 - 2\vartheta - d_\Sigma^2(\mu_{11}, \mathcal{R}^c)}.
\end{aligned} \tag{18}$$
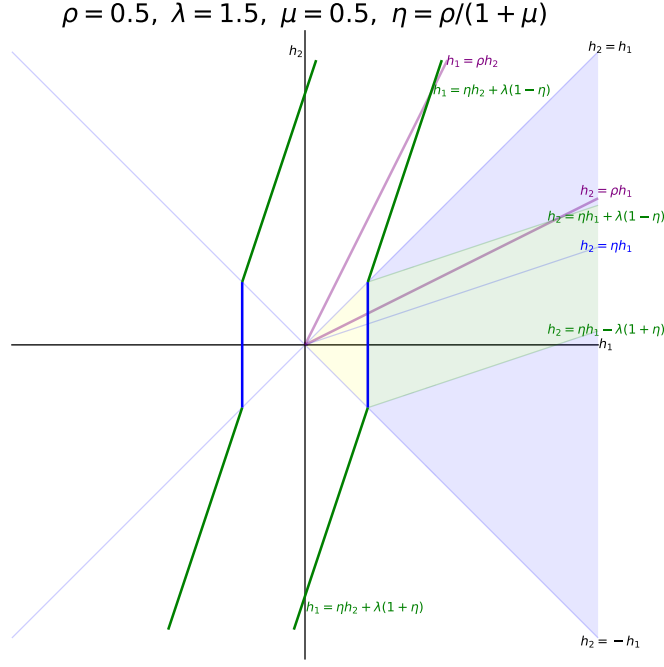
Figure 7: The rejection region of Elastic net for $\rho \geq 0$.

It suffices to calculate $d_\Sigma(\mu_{00}, \mathcal{R})$, $d_\Sigma(\mu_{01}, \mathcal{R})$, $d_\Sigma(\mu_{10}, \mathcal{R}^c)$, and $d_\Sigma(\mu_{11}, \mathcal{R}^c)$.

First, consider the case of $\rho \geq 0$. The expression of $\mathcal{R}$ is given explicitly in (17). By the definition in (14) and Definition 1, for any $S \subset \mathbb{R}^2$ and $\mu \notin S$,

$$d_\Sigma^2(\mu, S) = \frac{1}{1 - \rho^2} \inf_{\xi \in S} d_\rho^2(\mu, \xi). \tag{19}$$

If $S$ can be expressed as the interaction and union of finitely many half-planes, then the point $\xi^*$ that attains the infimum must be on the boundary line of one of these half-planes. We thus only need to calculate:

   (i) $\inf_{\xi \in \mathcal{L}} d_\rho^2(\mu, \xi)$ for the boundary line $\mathcal{L}$ of each half-plane in the definition of $S$ (with verification that the tangent point on $\mathcal{L}$ is achievable on the boundary);

   (ii) $d_\rho^2(\mu, \zeta)$ for each point $\zeta$ that is a vertex of $S$ (i.e., the intersection of two boundary lines).

For (i), we apply the formula given in Lemma A.1. For (ii), we apply Definition 1 directly. These calculations give a finite collection of values. In (i), the $\xi^*$ that attains the infimum may not belong to $S$; if that happens, we delete it from the collection. Finally, $\inf_{\xi \in S} d_\rho^2(\mu, \xi)$ is the minimum of the values in this collection.

Since $\mathcal{R}$ and $\mathcal{R}^c$ can be expressed via the interaction and union of finitely many half-planes, we follow the above routine to calculate the desired quantities. Take $d_\Sigma(\mu_{01}, \mathcal{R})$ for example. Recall that $\mu_{01} = (\rho\sqrt{r}, \sqrt{r})$. By (19), it suffices to calculate $\inf_{\xi \in \mathcal{R}} d_\rho^2(\mu_{01}, \xi)$. The region $\mathcal{R}$ has 6 boundary lines, but since $\rho \geq 0$, the infimum can only be attained in either of three cases:

    • on the line $\mathcal{L}_1 : h_1 - \eta h_2 = \sqrt{q}(1 - \eta)$;

    • on the line $\mathcal{L}_2 : h_1 = \sqrt{q}$;

    • on the the vertex $v^* = (\sqrt{q}, \sqrt{q})'$, which is an intersection of $\mathcal{L}_1 \cap \mathcal{L}_2$.

Let $x = (x_1, x_2)' = (h_1 - \rho\sqrt{r}, h_2 - \sqrt{r})'$. We can re-write $\mathcal{L}_1$ as a line $\mathcal{L}_1'$ for $x$, which expression is $x_1 - \eta x_2 = (1 - \eta)\sqrt{q} - (\rho - \eta)\sqrt{r}$. Similarly, we can re-write $\mathcal{L}_2$ as a line $\mathcal{L}_2': x_1 = \sqrt{q} - \rho\sqrt{r}$.

We apply Lemma A.1 to get

$$\inf_{\xi \in \mathcal{L}_1} d_\rho^2(\mu_{01}, \xi) = \inf_{x \in \mathcal{L}_1'} d_\rho^2(x, (0,0)') = \frac{\left[(1-\eta)\sqrt{q} - (\rho-\eta)\sqrt{r}\right]^2 (1-\rho^2)}{1+\eta^2 - 2\rho\eta},$$

$$\inf_{\xi \in \mathcal{L}_2} d_\rho^2(\mu_{01}, \xi) = \inf_{x \in \mathcal{L}_2'} d_\rho^2(x, (0,0)') = (\sqrt{q} - \rho\sqrt{r})^2(1-\rho^2),$$

$$d_\rho^2(\mu_{01}, v^*) = d_\rho^2\big((\sqrt{q}, \sqrt{q})', (\rho\sqrt{r}, \sqrt{r})'\big)$$

$$\equiv (\sqrt{q} - \rho\sqrt{r})^2 + (\sqrt{q} - \sqrt{r})^2 - 2\rho(\sqrt{q} - \rho\sqrt{r})(\sqrt{q} - \sqrt{r})^2.$$

The value of $\inf_{\xi \in \mathcal{R}} d_\rho^2(\mu_{01}, \xi)$ is the minimum of the above three values. In fact, the distance $d_\rho$ is related to the size of an ellipsoid that centers at $(\rho\sqrt{r}, \sqrt{r})'$ and hits the boundary of $\mathcal{R}$. As $\sqrt{r}$ increases from zero, the center of this ellipsoid moves upwards on the line of $h_1 = \rho h_2$. Consequently, the minimum of the above three values is initially (i) $\inf_{\xi \in \mathcal{L}_2} d_\rho^2(\mu_{01}, \xi)$ when $\sqrt{r}$ is appropriately small, then (ii) $d_\rho^2(\mu_{01}, v^*)$ when $\sqrt{r}$ is moderately large, and finally (iii) $\inf_{\xi \in \mathcal{L}_1} d_\rho^2(\mu_{01}, \xi)$ when $\sqrt{r}$ is sufficiently large; see Figure 7. We now figure out the range of $\sqrt{r}$ for each of the three cases. Recall that $v^* = (\sqrt{q}, \sqrt{q})'$. Let $\xi^* = (\xi_1^*, \xi_2^*)'$ be the vector that attains $\inf_{\xi \in \mathcal{L}_2} d_\rho^2(\mu_{01}, \xi)$. We have an explicit expression of $\xi^*$ from Lemma A.1. By equating it with $v^*$, we can solve the critical value of $\sqrt{r}$ at which case (i) transits to case (ii):

$$\sqrt{q} = v_2^* = \xi_2^* = \sqrt{r} + \rho(\sqrt{q} - \rho\sqrt{r}) \qquad \Longrightarrow \qquad \sqrt{r} = \frac{\sqrt{q}}{1+\rho}.$$

Similarly, let $\tilde{\xi}^* = (\tilde{\xi}_1^*, \tilde{\xi}_2^*)'$ be the vector that attains $\inf_{\xi \in \mathcal{L}_1} d_\rho^2(\mu_{01}, \xi)$. By equating $\tilde{\xi}^*$ with $v^*$, we can solve the critical value of $\sqrt{r}$ at which case (ii) transits to case (iii):

$$\sqrt{q} = v_2^* = \tilde{\xi}_2^* = \sqrt{r} + \frac{\left[(1-\eta)\sqrt{q} - (\rho-\eta)\sqrt{r}\right](\rho-\eta)}{1+\eta^2 - 2\rho\eta} \qquad \Longrightarrow \qquad \sqrt{r} = \frac{1+\eta}{1+\rho}\sqrt{q}.$$

We combine the above results to get

$$\inf_{\xi \in \mathcal{R}} d_\rho^2(\mu_{01}, \xi) = \begin{cases} (1-\rho^2)(\sqrt{q} - \rho\sqrt{r})_+^2, & \text{if } \sqrt{r} \le \frac{1}{1+\rho}\sqrt{q}, \\ d_\rho^2\big((\sqrt{q}, \sqrt{q})', (\rho\sqrt{r}, \sqrt{r})'\big), & \text{if } \frac{1}{1+\rho}\sqrt{q} < \sqrt{r} \le \frac{1+\eta}{1+\rho}\sqrt{q}, \\ \frac{(1-\rho^2)}{1+\eta^2 - 2\rho\eta}[(1-\eta)\sqrt{q} - (\rho-\eta)\sqrt{r}]_+^2, & \text{if } \sqrt{r} > \frac{1+\eta}{1+\rho}q. \end{cases} \quad (20)$$

Recall that $\eta = \rho/(1+\mu)$ is a shorthand notation. We plug (20) into (19), and then we insert it into (18). This gives the second term in $\mathrm{FP}_p$. We can follow the same routine to derive every term in $\mathrm{FP}_p$ and $\mathrm{FN}_p$. We omit the details but summarize the results in Theorem 8 below.

Next, consider the case of $\rho < 0$. We re-parametrize the linear model by replacing $(x_{j+1}, \beta_{j+1})$ with $(-x_{j+1}, -\beta_{j+1})$. After this re-parametrization, the $(j, j+1)$ block of the Gram matrix is a $2 \times 2$ matrix $\Sigma$ whose off-diagonal entries are $-\rho = |\rho|$. The rejection region is defined by the solution path of (16) associated with $|\rho| > 0$. This allows us to use the expression in (17) directly with a simple replacement of $\rho$ by $|\rho|$. There is no need to re-calculate the rejection region for a negative $\rho$.

Let $\tilde{y}_1 = x_j' y / \sqrt{2 \log(p)}$ and $\tilde{y}_2 = x_{j+1}' y / \sqrt{2 \log(p)}$. We still have $\tilde{y} \sim \mathcal{N}_2(\mu, \frac{1}{2 \log(p)}\Sigma)$. However, the four realizations of $(\beta_j, \beta_{j+1})$ become $(0, 0)$, $(0, -\tau_p)$, $(\tau_p, 0)$, and $(\tau_p, -\tau_p)$. Therefore, the mean vectors $\mu$ have changed to

$$\mu_{00} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \mu_{01} = \begin{bmatrix} -|\rho|\sqrt{r} \\ -\sqrt{r} \end{bmatrix}, \quad \mu_{10} = \begin{bmatrix} -\sqrt{r} \\ -|\rho|\sqrt{r} \end{bmatrix}, \quad \mu_{11} = \begin{bmatrix} (1-|\rho|)\sqrt{r} \\ -(1-|\rho|)\sqrt{r} \end{bmatrix}.$$

Similar to (18), it suffices to calculate $d_\Sigma(\mu_{00}, \mathcal{R})$, $d_\Sigma(\mu_{01}, \mathcal{R})$, $d_\Sigma(\mu_{10}, \mathcal{R}^c)$, and $d_\Sigma(\mu_{11}, \mathcal{R}^c)$. Here $\mathcal{R}$ is the same as in Figure 7, but the locations of the $\mu$ vectors have changed. Since $\mathcal{R}$ is centrosymmetric, $d_\Sigma(\mu_{00}, \mathcal{R})$, $d_\Sigma(\mu_{01}, \mathcal{R})$, and $d_\Sigma(\mu_{10}, \mathcal{R}^c)$ are actually the same as before. We only need to re-calculate $d_\Sigma^2(\mu_{11}, \mathcal{R}^c)$. The calculation routine is the same as that for (20). We omit the details but present the results directly in the theorem below.

To summarize, in this part, we have proved the following theorem:

**Theorem 8.** *Suppose the conditions of Theorem 2 hold. Let $\lambda = \sqrt{2q \log(p)}$ in Elastic net. Write $\eta = \rho/(1 + \mu)$. As $p \to \infty$,*

$$\mathrm{FP}_p = L_p p^{1 - \min\left\{ q,\ \vartheta + f_1(\sqrt{r}, \sqrt{q}) \right\}}, \qquad \mathrm{FN}_p = L_p p^{1 - \min\left\{ \vartheta + f_2(\sqrt{r}, \sqrt{q}),\ 2\vartheta + f_3(\sqrt{r}, \sqrt{q}) \right\}},$$

*where (below, $d_{|\rho|}^2(u, v)$ is as in Definition 1)*

$$f_1(\sqrt{r}, \sqrt{q}) = \begin{cases} (\sqrt{q} - |\rho|\sqrt{r})_+^2, & \text{if } \sqrt{r} \leq \frac{1}{1+|\rho|}\sqrt{q}, \\ \frac{1}{1-\rho^2} \cdot d_{|\rho|}^2\big((\sqrt{q}, \sqrt{q})', (|\rho|\sqrt{r}, \sqrt{r})'\big), & \text{if } \frac{1}{1+|\rho|}\sqrt{q} < \sqrt{r} \leq \frac{1+|\eta|}{1+|\rho|}\sqrt{q}, \\ \frac{[(1-|\eta|)\sqrt{q} - (|\rho| - |\eta|)\sqrt{r}]_+^2}{1 + \eta^2 - 2|\rho||\eta|}, & \text{if } \sqrt{r} \geq \frac{1+|\eta|}{1+|\rho|}\sqrt{q}, \end{cases}$$

$$f_2(\sqrt{r}, \sqrt{q}) = \min\left\{ (\sqrt{r} - \sqrt{q})_+^2,\ \frac{[(1-\rho\eta)\sqrt{r} - (1-|\eta|)\sqrt{q}]_+^2}{1 + \eta^2 - 2\rho\eta} \right\},$$

$$f_3(\sqrt{r}, \sqrt{q}) = \frac{(1-\eta)^2[(1+\rho)\sqrt{r} - \sqrt{q}]_+^2}{1 + \eta^2 - 2\rho\eta}.$$

**Part 3. Calculating the phase diagram.** By Theorem 8, the Hamming error is $\mathrm{FP}_p + \mathrm{FN}_p = L_p p^{1-h(q;\vartheta,r)}$, where

$$h(q; \vartheta, r) = \min\Big\{ \min\{ q,\ \vartheta + f_1(\sqrt{r}, \sqrt{q}) \},\ \min\{ \vartheta + f_2(\sqrt{r}, \sqrt{q}),\ 2\vartheta + f_3(\sqrt{r}, \sqrt{q}) \} \Big\}. \quad (21)$$

To calculate the phase diagram, we need to find $q^*$ that maximizes $h(q; \vartheta, r)$ and then investigate the conditions on $(r, \vartheta)$ such that $h(q^*; \vartheta, r) > 1$ or $\vartheta < h(q^*; \vartheta, r) < 1$ or $h(q^*; \vartheta, r) \leq \vartheta$.

We first prove that $r = \vartheta$ is the boundary between the Regions of Almost Full Recovery and No Recovery, i.e., the boundary separating $\vartheta < h(q^*; \vartheta, r) < 1$ and $h(q^*; \vartheta, r) \leq \vartheta$.

When $r < \vartheta$, we need to show $h(q; \vartheta, r) \leq \vartheta, \forall q$. If $q \leq \vartheta$, then $h(q; \vartheta, r) \leq q \leq \vartheta$. If $q > \vartheta$, then we look at $f_2(\sqrt{r}, \sqrt{q})$: Now we have $0 \leq f_2(\sqrt{r}, \sqrt{q}) \leq (\sqrt{r} - \sqrt{q})_+^2 = 0$. Thus $h(q; \vartheta, r) \leq \vartheta + f_2(\sqrt{r}, \sqrt{q}) = \vartheta$.

When $r > \vartheta$, we can always find suitable $q$ to make $h(q^*; \vartheta, r) \geq h(q; \vartheta, r) > \vartheta$. It is left for later discussion whether $h(q^*; \vartheta, r)$ is greater than 1. In fact, such $q$ can be any value satisfying $\max\{\vartheta, (\frac{|\rho| - |\eta|}{1-|\eta|})^2 r\} < q < r$, which always exists because $\frac{|\rho| - |\eta|}{1-|\eta|} < 1$. Since $r > q$, we know $f_2(\sqrt{r}, \sqrt{q})$ and $f_3(\sqrt{r}, \sqrt{q})$ are strictly positive from their definition; since $(1 - |\eta|)\sqrt{q} > (|\rho| - |\eta|)\sqrt{r}$, we also know $f_1(\sqrt{r}, \sqrt{q}) > 0$. Since all four components of $h(q; \vartheta, r)$ in (21) is greater than $\vartheta$, we have the desired result.

To sum up the discussion so far, we have shown that $r = \vartheta$ is the curve separating the regions of $\vartheta < h(q^*; \vartheta, r) < 1$ and $h(q^*; \vartheta, r) \leq \vartheta$.

For the rest of Part 3, we try to find the boundary between $h(q^*; \vartheta, r) > 1$ and $\vartheta < h(q^*; \vartheta, r) < 1$.

We need an important fact about such boundary, not only for the proof of Elastic net but also for all other methods. Recall the definition of $\mathrm{FP}_p$ and $\mathrm{FN}_p$ in (18), and we actually have the general form

$$\mathrm{FP}_p = L_p p^{1 - \min\left\{ d_\Sigma^2(\mu_{00}, \mathcal{R}),\ \vartheta + d_\Sigma^2(\mu_{01}, \mathcal{R}) \right\}}, \qquad \mathrm{FN}_p = L_p p^{1 - \min\left\{ \vartheta + d_\Sigma^2(\mu_{10}, \mathcal{R}^c),\ 2\vartheta + d_\Sigma^2(\mu_{11}, \mathcal{R}^c) \right\}},$$

and

$$h(q; \vartheta, r) = \min\Big\{ \min\{ d_\Sigma^2(\mu_{00}, \mathcal{R}),\ \vartheta + d_\Sigma^2(\mu_{01}, \mathcal{R}) \},\ \min\{ \vartheta + d_\Sigma^2(\mu_{10}, \mathcal{R}),\ 2\vartheta + d_\Sigma^2(\mu_{11}, \mathcal{R}) \} \Big\}.$$

As an important fact, we always have the following relationship at the boundary:

$$\min\big\{ d_\Sigma^2(\mu_{00}, \mathcal{R}),\ \vartheta + d_\Sigma^2(\mu_{01}, \mathcal{R}) \big\} = \min\big\{ \vartheta + d_\Sigma^2(\mu_{10}, \mathcal{R}),\ 2\vartheta + d_\Sigma^2(\mu_{11}, \mathcal{R}) \big\} = 1 \quad (22)$$

This is because: First, at the boundary $r = r(\vartheta)$, we must always have $h(q^*; \vartheta, r(\vartheta)) = 1$; otherwise, since $h(q; \vartheta, r)$ is continuous in $(q, r)$ for fixed $\vartheta$, it would contradict the definition of the boundary itself. Second, the exponents of $\mathrm{FP}_p$ and $FN_p$ has to be equal. This is because if we change the tuning parameters for fixed $(\vartheta, r)$, it can only enlarge or shrink the rejection region $\mathcal{R}$, and thus the effects

on $FP_p$ and $FN_p$ would always be in the opposite directions. As a result, if the exponents of $FP_p$ and $FN_p$ are not equal at the boundary, we can change the tuning parameters to make $h(q; \vartheta, r) > 1$.

With the important fact, we can proceed our discussion. By the definition of $h(q; \vartheta, r)$ in (21), it gives us $2 \times 2 = 4$ cases respectively for $\rho > 0$ and $\rho < 0$. We discuss them one by one and summarise the results when the full phase curves are complete. For brevity, we also denote $\lambda' = \sqrt{q}$ for the rest of Part 3.

When $\rho > 0$, we have four cases.

*First*, if $\lambda'^2 = \vartheta + f_2(\sqrt{r}, \lambda') = 1$ and $\vartheta + f_1(\sqrt{r}, \lambda') \geq 1$, $2\vartheta + f_4(\sqrt{r}, \lambda') \geq 1$: we have $\lambda' = 1$. From $\vartheta + f_2(\sqrt{r}, \lambda') = 1$, we know

$$\sqrt{r} = \max \left\{ 1 + \sqrt{1 - \vartheta}, \frac{\sqrt{1 + \eta^2 - 2\rho\eta}}{1 - \rho\eta} \sqrt{1 - \vartheta} + \frac{1 - \eta}{1 - \rho\eta} \right\}.$$

We also know from $\vartheta + f_1(\sqrt{r}, \lambda') \geq 1$, that since $\sqrt{r} > \lambda' = 1$, $\sqrt{r} \leq \frac{1-\eta}{\rho-\eta} - \frac{\sqrt{1+\eta^2-2\rho\eta}}{\rho-\eta}\sqrt{1-\vartheta}$; from $2\vartheta + f_4(\sqrt{r}, \lambda') \geq 1$, that $\sqrt{r} \geq \frac{\sqrt{1+\eta^2-2\rho\eta}}{(1-\eta)(1+\rho)}\sqrt{1-2\vartheta} + \frac{1}{1+\rho}$.

After roughly interpreting the requirements, we make two points: (i) we always have

$$\frac{\sqrt{1 + \eta^2 - 2\rho\eta}}{1 - \rho\eta} \sqrt{1 - \vartheta} + \frac{1 - \eta}{1 - \rho\eta} \geq \frac{\sqrt{1 + \eta^2 - 2\rho\eta}}{(1 - \eta)(1 + \rho)} \sqrt{1 - 2\vartheta} + \frac{1}{1 + \rho} \tag{23}$$

and thus the requirement from $2\vartheta + f_4(\sqrt{r}, \lambda') \geq 1$ is loose. This can be proven by showing $\frac{1-\eta}{1-\rho\eta} \geq \frac{1}{1+\rho}$ and $\frac{\sqrt{1+\eta^2-2\rho\eta}}{1-\rho\eta} \geq \frac{\sqrt{1+\eta^2-2\rho\eta}}{(1-\eta)(1+\rho)}$ respectively. (ii) Actually, we can eliminate the curve $\sqrt{r} = \frac{\sqrt{1+\eta^2-2\rho\eta}}{1-\rho\eta}\sqrt{1-\vartheta} + \frac{1-\eta}{1-\rho\eta}$ from this step, without using the requirement from $2\vartheta + f_4(\sqrt{r}, \lambda') \geq 1$. (As a result, the same proof holds for the corresponding case of $\rho < 0$.) This is because if we put together

$$\begin{cases} \frac{\sqrt{1+\eta^2-2\rho\eta}}{1-\rho\eta}\sqrt{1-\vartheta} + \frac{1-\eta}{1-\rho\eta} > 1 + \sqrt{1-\vartheta} \\ \frac{\sqrt{1+\eta^2-2\rho\eta}}{1-\rho\eta}\sqrt{1-\vartheta} + \frac{1-\eta}{1-\rho\eta} \leq \frac{1-\eta}{\rho-\eta} - \frac{\sqrt{1+\eta^2-2\rho\eta}}{\rho-\eta}\sqrt{1-\vartheta} \end{cases}$$

we will have no solution. To be more specific, the first inequality gives us $\sqrt{1 - \vartheta} > \frac{\eta(1-\rho)}{\sqrt{1+\eta^2-2\rho\eta}-1+\rho\eta}$, and the second equation will eventually give us $\sqrt{1 - \vartheta} \leq \frac{(1+\eta)(1-\rho)}{(1+\rho)\sqrt{1+\eta^2-2\rho\eta}}$. However, the upper and lower bounds on $\sqrt{1 - \vartheta}$ admits no solution, because we can prove $\frac{\eta(1-\rho)}{\sqrt{1+\eta^2-2\rho\eta}-1+\rho\eta} \geq \frac{(1+\eta)(1-\rho)}{(1+\rho)\sqrt{1+\eta^2-2\rho\eta}}$ jusy by simplifying it for a few steps.

To sum up, the *first* case gives us $\sqrt{r} = 1 + \sqrt{1 - \vartheta}$ with the requirement $\sqrt{r} \leq \frac{1-\eta}{\rho-\eta} - \frac{\sqrt{1+\eta^2-2\rho\eta}}{\rho-\eta}\sqrt{1-\vartheta}$.

*Second*, if $q = 2\vartheta + f_4(\sqrt{r}, \lambda') = 1$ and $\vartheta + f_1(\sqrt{r}, \lambda') \geq 1$, $\vartheta + f_2(\sqrt{r}, \lambda') \geq 1$, we will need

$$\sqrt{r} = \frac{\sqrt{1 + \eta^2 - 2\rho\eta}}{(1 - \eta)(1 + \rho)} \sqrt{1 - 2\vartheta} + \frac{1}{1 + \rho}$$

while requiring $\sqrt{r} \geq \max \left\{ 1 + \sqrt{1 - \vartheta}, \frac{\sqrt{1+\eta^2-2\rho\eta}}{1-\rho\eta}\sqrt{1-\vartheta} + \frac{1-\eta}{1-\rho\eta} \right\}$. We know this is impossible from Equation (23). No curve is produced in this case.

*Third*, if $\vartheta + f_1(\sqrt{r}, \lambda') = \vartheta + f_2(\sqrt{r}, \lambda') = 1$, and $\lambda' \geq 1$, $2\vartheta + f_3(\sqrt{r}, \lambda') \geq 1$, we will know from $\vartheta + f_2(\sqrt{r}, \lambda') = 1$ that

$$\sqrt{r} = \max \left\{ \lambda' + \sqrt{1 - \vartheta}, \frac{\sqrt{1 + \eta^2 - 2\rho\eta}}{1 - \rho\eta} \sqrt{1 - \vartheta} + \lambda' \frac{1 - \eta}{1 - \rho\eta} \right\}.$$

We can use the same method as in the first point of the *first* case to show that $2\vartheta + f_3(\sqrt{r}, \lambda') \geq 1$ is loose with $\lambda' \geq 1$.

Now the curve seems to have two choices, but the latter one is actually impossible. When $\sqrt{r} = \frac{\sqrt{1+\eta^2-2\rho\eta}}{1-\rho\eta}\sqrt{1-\vartheta} + \lambda'\frac{1-\eta}{1-\rho\eta}$, we have

$$\begin{cases} \sqrt{r} = & \lambda'\frac{1-\eta}{\rho-\eta} - \frac{\sqrt{1+\eta^2-2\rho\eta}}{\rho-\eta}\sqrt{1-\vartheta} = \frac{\sqrt{1+\eta^2-2\rho\eta}}{1-\rho\eta}\sqrt{1-\vartheta} + \lambda'\frac{1-\eta}{1-\rho\eta} \\ \lambda' \geq & 1 \end{cases}$$

which implies $\lambda' = \frac{(1+\rho)\sqrt{1+\eta^2-2\rho\eta}\sqrt{1-\vartheta}}{(1-\rho)(1+\eta)} \geq 1$. We can eliminate this case now, without considering the requirement of $2\vartheta + f_3(\sqrt{r}, \lambda') \geq 1$, because with such $\lambda' \geq 1$, $\frac{\sqrt{1+\eta^2-2\rho\eta}}{1-\rho\eta}\sqrt{1-\vartheta} + \lambda'\frac{1-\eta}{1-\rho\eta} \geq \lambda' + \sqrt{1-\vartheta}$ cannot hold. To see this, we can compare $\frac{\sqrt{1+\eta^2-2\rho\eta}}{1-\rho\eta}\sqrt{1-\vartheta} + \lambda'\frac{1-\eta}{1-\rho\eta}$; cancelling out "$\sqrt{1-\vartheta}$", we have

$$\frac{\sqrt{1+\eta^2-2\rho\eta}}{1-\rho\eta} - 1 \geq \frac{\eta(1-\rho)}{1-\rho\eta} \cdot \frac{(1+\rho)\sqrt{1+\eta^2-2\rho\eta}}{(1-\rho)(1+\eta)}$$

Simplifying this for a few steps, and we will arrive at "$\sqrt{1+\eta^2-2\rho\eta} \geq 1+\eta$" which gives a contradiction.

We can only have one case, where $\lambda' + \sqrt{1-\vartheta}$ is greater:

$$\begin{cases} \sqrt{r} = & \lambda'\frac{1-\eta}{\rho-\eta} - \frac{\sqrt{1+\eta^2-2\rho\eta}}{\rho-\eta}\sqrt{1-\vartheta} = \lambda' + \sqrt{1-\vartheta} \\ \lambda' \geq & 1 \end{cases}$$

To sum up, the *third* case gives us the curve $\sqrt{r} = \left[\frac{1-\eta}{1-\rho} + \frac{\sqrt{1+\eta^2-2\rho\eta}}{1-\rho}\right]\sqrt{1-\vartheta}$ with the requirement $\lambda' = \left[\frac{\rho-\eta}{1-\rho} + \frac{\sqrt{1+\eta^2-2\rho\eta}}{1-\rho}\right]\sqrt{1-\vartheta} \geq 1$.

*Fourth*, if $\vartheta + f_2(\sqrt{r}, \lambda') = 2\vartheta + f_3(\sqrt{r}, \lambda') = 1$ and $\lambda' \geq 1$, $\vartheta + f_1(\sqrt{r}, \lambda') \geq 1$, we have the same contradiction as the *second* case, that

$$\frac{\sqrt{1+\eta^2-2\rho\eta}}{(1-\eta)(1+\rho)}\sqrt{1-2\vartheta} + \frac{\lambda'}{1+\rho} \geq \frac{\sqrt{1+\eta^2-2\rho\eta}}{1-\rho\eta}\sqrt{1-\vartheta} + \lambda'\frac{1-\eta}{1-\rho\eta}$$

cannot hold.

*Summarising the cases of positive correlation,*, we have two curves:

$$\sqrt{r} = 1 + \sqrt{1-\vartheta}$$

$$\sqrt{r} = \left[\frac{1-\eta}{1-\rho} + \frac{\sqrt{1+\eta^2-2\rho\eta}}{1-\rho}\right]\sqrt{1-\vartheta}$$

and the intersection point of the two curves is exactly at $\left[\frac{\rho-\eta}{1-\rho} + \frac{\sqrt{1+\eta^2-2\rho\eta}}{1-\rho}\right]\sqrt{1-\vartheta} = 1$ so the two curves can be summarised as taking the maximum.

When $\rho < 0$, we also have four cases.

*First*, if $\lambda'^2 = \vartheta + f_2(\sqrt{r}, \lambda') = 1$ and $\vartheta + f_1(\sqrt{r}, \lambda') \geq 1$, $2\vartheta + f_4(\sqrt{r}, \lambda') \geq 1$: We already know from the same proof when the correlation is positive, that $\sqrt{r} = 1 + \sqrt{1-\vartheta}$ is the only admissible curve. Now we additionally need it to satisfy $\sqrt{r} \geq \frac{\sqrt{1+\eta^2-2\rho\eta}}{(1+|\eta|)(1-|\rho|)}\sqrt{1-2\vartheta} + \frac{1}{1-|\rho|}$.

*Second*, if $\lambda'^2 = 2\vartheta + f_4(\sqrt{r}, \lambda') = 1$ and $\vartheta + f_1(\sqrt{r}, \lambda') \geq 1$, $\vartheta + f_2(\sqrt{r}, \lambda') \geq 1$, we will need $\sqrt{r} = \frac{\sqrt{1+\eta^2-2\rho\eta}}{(1+|\eta|)(1-|\rho|)}\sqrt{1-2\vartheta} + \frac{1}{1-|\rho|}$ and we need $\sqrt{r} \geq 1 + \sqrt{1-\vartheta}$ from $\vartheta + f_2(\sqrt{r}, \lambda') \geq 1$.

*To sum up the first two cases,* they give us

$$\sqrt{r} = \max\left\{1 + \sqrt{1-\vartheta}, \frac{\sqrt{1+\eta^2-2\rho\eta}}{(1+|\eta|)(1-|\rho|)}\sqrt{1-2\vartheta} + \frac{1}{1-|\rho|}\right\}.$$

and we need $\sqrt{r} \leq \frac{1-|\eta|}{|\rho|-|\eta|} - \frac{\sqrt{1+\eta^2-2\rho\eta}}{|\rho|-|\eta|}\sqrt{1-\vartheta}$ from $\vartheta + f_1(\sqrt{r}, \lambda') \geq 1$.

*Third*, if $\vartheta + f_1(\sqrt{r}, \lambda') = \vartheta + f_2(\sqrt{r}, \lambda') = 1$, and $\lambda' \geq 1$, $2\vartheta + f_3(\sqrt{r}, \lambda') \geq 1$, we already know from the same proof when the correlation is positive, that we have only one curve $\sqrt{r} = \left[\frac{1-|\eta|}{1-|\rho|} + \frac{\sqrt{1+\eta^2-2\rho\eta}}{1-|\rho|}\right]\sqrt{1-\vartheta}$. Now we only need to update the requirement from $2\vartheta + f_3(\sqrt{r}, \lambda') \geq 1$, which is

$$\sqrt{r} \geq \frac{\sqrt{1+\eta^2-2\rho\eta}}{(1+|\eta|)(1-|\rho|)}\sqrt{1-2\vartheta} + \frac{\lambda'}{1-|\rho|}$$

Plug in $\lambda' = \left[\frac{|\rho|-|\eta|}{1-|\rho|} + \frac{\sqrt{1+\eta^2-2\rho\eta}}{1-|\rho|}\right]\sqrt{1-\vartheta}$, it is equivalent to

$$\left[\frac{1-|\eta|}{1-|\rho|} + \frac{\sqrt{1+\eta^2-2\rho\eta}}{1-|\rho|}\right]\sqrt{1-\vartheta} \geq \frac{\sqrt{1+\eta^2-2\rho\eta}}{1-2|\rho|+\rho\eta}\left[\sqrt{1-\vartheta} + \frac{1-|\eta|}{1+|\eta|}\sqrt{1-2\vartheta}\right]$$

where the RHS is another curve which will show up in the next case.

*Fourth*, if $\vartheta + f_2(\sqrt{r}, \lambda') = 2\vartheta + f_3(\sqrt{r}, \lambda') = 1$ and $\lambda' \geq 1$, $\vartheta + f_1(\sqrt{r}, \lambda') \geq 1$, we will have the most tedious case.

It can be implied by $\vartheta + f_2(\sqrt{r}, \lambda') = 2\vartheta + f_3(\sqrt{r}, \lambda') = 1$ that

$$\lambda'\left(\frac{1-|\eta|}{|\rho|-|\eta|} - \frac{1}{1-|\rho|}\right) = \sqrt{1-\vartheta}\frac{\sqrt{1+\eta^2-2\eta\rho}}{|\rho|-|\eta|} + \sqrt{1-2\vartheta}\frac{\sqrt{1+\eta^2-2\eta\rho}}{(1+|\eta|)(1-|\rho|)}$$

When $1 + \rho\eta - 2|\rho| \leq 0$, the equation admits no solution, because the coefficient of $\lambda'$ is not positive. In this case, if we look back at the curve in the *second* case $FP_1 = FN_2$, we will notice that

$$\frac{\sqrt{1+\eta^2-2\rho\eta}}{(1+|\eta|)(1-|\rho|)}\sqrt{1-2\vartheta} + \frac{1}{1-|\rho|} \leq \frac{1-|\eta|}{|\rho|-|\eta|} - \frac{\sqrt{1+\eta^2-2\rho\eta}}{|\rho|-|\eta|}\sqrt{1-\vartheta}$$

has no solution either. As a result, when $1 + \rho\eta - 2|\rho| \leq 0$, there is simply no Exact Recovery region in $\vartheta \in (0, \frac{1}{2})$.

When $1 + \rho\eta - 2|\rho| > 0$, we can proceed to solve for $\lambda'$ and then $\sqrt{r}$:

$$\lambda' = \frac{\sqrt{1+\eta^2-2\rho\eta}}{1-2|\rho|+\rho\eta}\left[(1-|\rho|)\sqrt{1-\vartheta} + \frac{|\rho|-|\eta|}{1+|\eta|}\sqrt{1-2\vartheta}\right]$$

$$\sqrt{r} = \frac{\sqrt{1+\eta^2-2\rho\eta}}{1-2|\rho|+\rho\eta}\left[\sqrt{1-\vartheta} + \frac{1-|\eta|}{1+|\eta|}\sqrt{1-2\vartheta}\right]$$

For all the requirements from $\lambda' \geq 1$ and $\vartheta + f_1(\sqrt{r}, \lambda') \geq 1$, we actually need $\lambda' \geq 1$ and $\sqrt{r} \geq \max\left\{\lambda' + \sqrt{1-\vartheta}, \frac{\sqrt{1+\eta^2-2\rho\eta}}{1-\rho\eta}\sqrt{1-\vartheta} + \lambda'\frac{1-|\eta|}{1-\rho\eta}\right\}$.

The requirement $\lambda' \geq 1$ is actually

$$\frac{\sqrt{1+\eta^2-2\rho\eta}}{1-2|\rho|+\rho\eta}\left[\sqrt{1-\vartheta} + \frac{1-|\eta|}{1+|\eta|}\sqrt{1-2\vartheta}\right] \geq \frac{\sqrt{1+\eta^2-2\rho\eta}}{(1+|\eta|)(1-|\rho|)}\sqrt{1-2\vartheta} + \frac{1}{1-|\rho|}$$

and $\sqrt{r} \geq \lambda' + \sqrt{1-\vartheta}$ is actually

$$\frac{\sqrt{1+\eta^2-2\rho\eta}}{1-2|\rho|+\rho\eta}\left[\sqrt{1-\vartheta} + \frac{1-|\eta|}{1+|\eta|}\sqrt{1-2\vartheta}\right] \geq \left[\frac{1-|\eta|}{1-|\rho|} + \frac{\sqrt{1+\eta^2-2\rho\eta}}{1-|\rho|}\right]\sqrt{1-\vartheta}.$$

So we can already conclude, that the diagram is

$$\sqrt{r} = \max\left\{1 + \sqrt{1-\vartheta}, \left[\frac{1-|\eta|}{1-|\rho|} + \frac{\sqrt{1+\eta^2-2\rho\eta}}{1-|\rho|}\right]\sqrt{1-\vartheta}, \right.$$
$$\left. \frac{\sqrt{1+\eta^2-2\rho\eta}}{(1+|\eta|)(1-|\rho|)}\sqrt{1-2\vartheta} + \frac{1}{1-|\rho|}, \frac{\sqrt{1+\eta^2-2\rho\eta}}{(1-2|\rho|+\rho\eta)_+}\left[\sqrt{1-\vartheta} + \frac{1-|\eta|}{1+|\eta|}\sqrt{1-2\vartheta}\right]\right\}$$

However, we are left with one last constraint $\frac{\sqrt{1+\eta^2-2\rho\eta}}{1-2|\rho|+\rho\eta}\left[\sqrt{1-\vartheta}+\frac{1-|\eta|}{1+|\eta|}\sqrt{1-2\vartheta}\right] \geq$

$\frac{\sqrt{1+\eta^2-2\rho\eta}}{1-\rho\eta}\sqrt{1-\vartheta}+\lambda'\frac{1-|\eta|}{1-\rho\eta}$. It is actually loose, but the proof may be tedious (and unimportant). We just need to prove that when $1-2|\rho|+\rho\eta > 0$ and

$$\lambda' = \frac{\sqrt{1+\eta^2-2\rho\eta}}{1-2|\rho|+\rho\eta}\left[(1-|\rho|)\sqrt{1-\vartheta}+\frac{|\rho|-|\eta|}{1+|\eta|}\sqrt{1-2\vartheta}\right] \geq 1,$$

we always have $\lambda' + \sqrt{1-\vartheta} \geq \frac{\sqrt{1+\eta^2-2\rho\eta}}{1-\rho\eta}\sqrt{1-\vartheta}+\lambda'\frac{1-|\eta|}{1-\rho\eta}$, which is equivalent to

$$\frac{(1-|\rho|)\sqrt{1+\eta^2-2\rho\eta}}{1-2|\rho|+\rho\eta}\sqrt{1-\vartheta}+\frac{|\rho|-|\eta|}{1+|\eta|}\frac{\sqrt{1+\eta^2-2\rho\eta}}{1-2|\rho|+\rho\eta}\sqrt{1-2\vartheta} \geq \frac{\sqrt{1+\eta^2-2\rho\eta}-(1-\rho\eta)}{|\eta|(1-|\rho|)}\sqrt{1-\vartheta}$$

We first look at one sufficient condition, $\frac{\sqrt{1+\eta^2-2\rho\eta}-(1-\rho\eta)}{|\eta|(1-|\rho|)} \leq 1$ By simplifying this inequality, we get $\sqrt{1+\eta^2-2\rho\eta} \leq 1+|\eta|-2\rho\eta$. The RHS is positive, because: (i) If $|\rho| \leq 0.5$, $1+|\eta|(1-2|\rho|) > 0$; (ii) If $|\rho| > 0.5$, recall $1-2|\rho|+\rho\eta > 0 \implies |\eta| \geq \frac{2|\rho|-1}{|\rho|} \implies 1+|\eta|(1-2|\rho|) \geq 1 - \frac{(2|\rho|-1)^2}{|\rho|} \geq 0$ for $|\rho| \geq 0.5$. Then we can square both sides and proceed, and finally getting $\eta\rho \leq \frac{1}{2}$.

As a result, when $\rho\eta \leq \frac{1}{2}$, we already have a sufficient condition for what we want to prove. When $\rho\eta > \frac{1}{2}$, we look at another sufficient condition:

We only need to prove another sufficient condition, by looking at the coefficients of $\sqrt{1-\vartheta}$,

$$\frac{(1-|\rho|)\sqrt{1+\eta^2-2\rho\eta}}{1-2|\rho|+\rho\eta} \geq \frac{\sqrt{1+\eta^2-2\rho\eta}-(1-\rho\eta)}{\eta(1-|\rho|)}$$

which is equivalent to verifying

$$(1-|\eta|-2|\rho|+3\eta\rho-|\eta|\rho^2)\sqrt{1+\eta^2-2\rho\eta} \leq (1-\rho\eta)(1-2|\rho|+\rho\eta)$$

It is elementary mathematics that $(RHS - LHS)$ is always positive as a function of $(|\rho|, |\eta|)$ under $\rho\eta \geq \frac{1}{2}$, $1-2|\rho|+\rho\eta > 0$ and $0 < |\eta| < |\rho|$.

*Summarising the cases of negative correlation:* The diagram is

$$\sqrt{r} = \max\Big\{ 1+\sqrt{1-\vartheta}, \left[\frac{1-|\eta|}{1-|\rho|}+\frac{\sqrt{1+\eta^2-2\rho\eta}}{1-|\rho|}\right]\sqrt{1-\vartheta},$$

$$\frac{\sqrt{1+\eta^2-2\rho\eta}}{(1+|\eta|)(1-|\rho|)}\sqrt{1-2\vartheta}+\frac{1}{1-|\rho|}, \frac{\sqrt{1+\eta^2-2\rho\eta}}{(1-2|\rho|+\rho\eta)_+}\left[\sqrt{1-\vartheta}+\frac{1-|\eta|}{1+|\eta|}\sqrt{1-2\vartheta}\right]\Big\}$$

## B.1 PROOF OF LEMMA B.1

Recall the optimization in (16); the solution $b = (b_1, b_2)$ has to set the sub-gradient of the objective function to zero. As a result, the equation of the sub-gradient for $b = (b_1, b_2)$ is:

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}+\sqrt{q}\begin{bmatrix} \text{sgn}(b_1) \\ \text{sgn}(b_2) \end{bmatrix}+\mu\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}$$

Now we begin to find out the solution path. Fixing $\mu$, we decrease $\sqrt{q}$ from a sufficiently large value to see when the variables enter the model. We assume $h_1 > 0$ and $0 < |h_2| < h_1$. Other cases can be computed in a similar way.

*Stage 1*: When $\sqrt{q}$ is so large that neither of $(x_j, x_{j+1})$ is in the model

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \end{bmatrix}+\sqrt{q}\begin{bmatrix} \text{sgn}(0) \\ \text{sgn}(0) \end{bmatrix}+\mu\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix},$$

it requires $\sqrt{q} \geq \sqrt{q_1} = \max\{|h_1|, |h_2|\} = h_1$.

*Stage 2*: When $\sqrt{q}$ crosses $\sqrt{q_1} = h_1$, we assert that variable $x_j$ has to enter the model, while $x_{j+1}$ does not. This is because:

- $b_1$ has to be positive. If it is negative, we have $b_1 - \sqrt{q} + \mu \cdot 2b_1 = h_1$, which implies $b_1$ has the same sign as $h_1 + \sqrt{q} > 0$, which is a contradiction.

- $b_2$ cannot enter the model at this point. Otherwise, we have $(1+\mu)b_2 + \sqrt{q}\,\mathrm{sgn}(b_2) = h_2$, for $|h_2| < \sqrt{q} < h_1$. Considering the sign of $b_2$, we have a contradiction.

So $b_1$ has to enter the model as a positive number. Now the equation is

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ 0 \end{bmatrix} + \sqrt{q} \begin{bmatrix} 1 \\ \mathrm{sgn}(0) \end{bmatrix} + \mu \begin{bmatrix} b_1 \\ 0 \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}$$

Thus $b_1 = \frac{h_1 - \sqrt{q}}{1+\mu}$ and $\rho b_1 + \sqrt{q}\,\mathrm{sgn}(0) = h_2$. Since $\mathrm{sgn}(0) \in [-1, 1]$, we need

$$|h_2 - \frac{\rho}{1+\mu}h_1 + \frac{\rho}{1+\mu}\sqrt{q}| \le \sqrt{q} \tag{24}$$

By discussing the sign of the content of the absolute value as $\sqrt{q}$ decreases, we have the following two cases:

*Stage 3, Case 1*: When $h_2 > \eta h_1$ (recall we define $\eta = \frac{\rho}{1+\mu}$ as a shorthand), and $\sqrt{q}$ crosses $\sqrt{q_2} = \frac{h_2 - \eta h_1}{1-\eta}$, then $x_{j+1}$ enters the model and $b_2$ is positive. This is because Equation (24) is now $h_2 - \eta h_1 + \eta\sqrt{q} \le \sqrt{q}$; as $\sqrt{q} = \sqrt{q_2} = \frac{h_2 - \eta h_1}{1-\eta}$, the sub-gradient of $|b_2|$ is taking the value of $1 \in \mathrm{sgn}(0)$, so $b_2$ has to enter the model as a positive number.

In this case, we solve

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \sqrt{q} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \mu \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}$$

and get

$$b_1 = \frac{\frac{h_1 - \sqrt{q}}{1+\mu} - \eta\frac{h_2 - \sqrt{q}}{1+\mu}}{1 - \eta^2}, \quad b_2 = \frac{\frac{h_2 - \sqrt{q}}{1+\mu} - \eta\frac{h_1 - \sqrt{q}}{1+\mu}}{1 - \eta^2}.$$

*Stage 3, Case 2*: When $h_2 < \eta h_1$, and $\sqrt{q}$ crosses $\sqrt{q_2} = \frac{\eta h_1 - h_2}{1+\eta}$, $x_{j+1}$ enters the model and $b_2$ is negative. This is because reviewing Equation 24, we always have $h_2 - \eta h_1 + \eta\sqrt{q} < \sqrt{q}$, and thus when $\sqrt{q}$ is small enough to make $|h_2 - \eta h_1 + \eta\sqrt{q}| = \sqrt{q}$, it has to be $-h_2 + \eta h_1 - \eta\sqrt{q} = \sqrt{q}$. As a result, when $\sqrt{q}$ crosses $\sqrt{q_2} = \frac{\eta h_1 - h_2}{1+\eta}$, $b_2$ enters the model as a negative number. Solving

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \sqrt{q} \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \mu \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}$$

we have

$$b_1 = \frac{\frac{h_1 - \sqrt{q}}{1+\mu} - \eta\frac{h_2 + \sqrt{q}}{1+\mu}}{1 - \eta^2}, \quad b_2 = \frac{\frac{h_2 + \sqrt{q}}{1+\mu} - \eta\frac{h_1 - \sqrt{q}}{1+\mu}}{1 - \eta^2}.$$

## C  PROOF OF PROPOSITION 1 (MARGINAL REGRESSION)

Proposition 1 is about the connection of Elastic net to Lasso and marginal regression.

To prove the assertion about Lasso, we only need to quote the results from the Corollary 4.2 of Ke et al. (2020) on the phase curves of Lasso. In fact, the phase curves of Lasso can be exactly obtained by setting $\mu = 0$ in Theorem 2. As $\mu$ decreases from some positive value to zero, the curves in Theorem 2 just converges downwards to the phase curves of Lasso.

To prove the assertion about marginal regression, we need to fully study the variable selection method based on marginal regression, which we will be devoted to for the rest of this section.

**Definition 2.** *Marginal regression refers to the variable selection method which ranks all the variabels according to $\{|X'y|_j : j \in [p]\}$ and sets some cutoff point $t$ for the ranking. Then the variables $\{j \in [p] : |X'y|_j > t\}$ is selected.*

Soft-thresholded marginal regression behaves the same as Definition 2 in terms of variable selection.

**Remark 1.** *If we focus on $(1 + \mu)\hat{\beta}^{\mathrm{EN}}$, then Lasso and soft-thresholded marginal regression are just two limits of $\mu = 0$ and $\mu = \infty$, in terms of the solution and its path, the shape of the rejection region, and phase curves.*

As described in Section A, our proof still consists of three parts: (a) deriving the rejection region, (b) obtaining the rate of convergence of $\mathbb{E}[H(\hat{\beta}, \beta)]$, and (c) calculating the phase diagram.

**Part 1: Deriving the rejection region.** According to Definition 2, the variable selection based on marginal regression can be decomposed to every correlated pair of variables, $(x_j, x_{j+1})$. It directly thresholds $(x'_j y, x'_{j+1} y)$ with $t$, and if we divide $(x'_j y, x'_{j+1} y, t)$ with $\sqrt{2 \log(p)}$, it is equivalent to thresholding $h_1 = x'_j y / \sqrt{2 \log(p)}$, $h_2 = x'_{j+1} y / \sqrt{2 \log(p)}$ with $t' = t / \sqrt{2 \log(p)}$.

The solution path of marginal regression is very straight forward, so we present it in Lemma 1 without proof.

**Lemma 1** (the solution path of marginal regression). *The definition of $(h_1, h_2, \hat{b}_1, \hat{b}_2)$ follows from that of Lemma B.1, and we still assume $h_1 > |h_2| \geq 0$. The solution path of marginal regression defined in 2 can be describes as:*

1. *When $t' \geq h_1$, we have $\hat{b}_1 = \hat{b}_2 = 0$.*

2. *When $|h_2| \leq t' < h_1$, we have $\hat{b}_1 \neq 0$ and $\hat{b}_2 = 0$.*

3. *When $t < |h_2|$, we have $\hat{b}_1 \neq 0$ and $\hat{b}_2 \neq 0$.*

We now use Lemma 1 to derive the rejection region $\mathcal{R}$ of marginal regression. Recall that $\mathcal{R}$ is the set of $h = (h_1, h_2)'$ such that $\hat{b}_1 \neq 0$. In fact, the same procedure of Elastic net can be copied here, except that the specific behavior of the variable selection method is different. Lemma 1 tells us it is just

$$\mathcal{R} = \{(h_1, h_2) : h_1 > t'\} \cup \{(h_1, h_2) : h_1 < -t'\} \tag{25}$$

**Part 2. Analyzing the Hamming error.** The first steps of analysing Elastic net applies here as well, and we just need to compute $d_\Sigma(\mu_{00}, \mathcal{R})$, $d_\Sigma(\mu_{01}, \mathcal{R})$, $d_\Sigma(\mu_{10}, \mathcal{R}^c)$, and $d_\Sigma(\mu_{11}, \mathcal{R}^c)$ given the different shape of $\mathcal{R}$. Then we can compute

$$\mathrm{FP}_p = L_p p^{1 - \min\left\{d_\Sigma^2(\mu_{00}, \mathcal{R}), \; \vartheta + d_\Sigma^2(\mu_{01}, \mathcal{R})\right\}}, \qquad \mathrm{FN}_p = L_p p^{1 - \min\left\{\vartheta + d_\Sigma^2(\mu_{10}, \mathcal{R}^c), \; 2\vartheta + d_\Sigma^2(\mu_{11}, \mathcal{R}^c)\right\}}.$$

Finally $\mathbb{E}[H(\hat{\beta}, \beta)] = \mathrm{FP}_p + \mathrm{FN}_p = L_p p^{1 - h(t'; \vartheta, r)}$.

**Theorem 9.** *Under the conditions of Theorem 2, let $t = t' \sqrt{2 \log(p)}$ in marginal regression defined in 2. As $p \to \infty$,*

$$\mathrm{FP}_p = L_p \cdot p^{1 - \min\left\{t'^2, \; \vartheta + (t' - |\rho|\sqrt{r})_+^2\right\}}$$

$$\mathrm{FN}_p = L_p \cdot p^{1 - \min\left\{\vartheta + (\sqrt{r} - t')_+^2, \; 2\vartheta + [(1+\rho)\sqrt{r} - t']_+^2\right\}}$$

**Part 3. Calculating the phase diagram.** By Theorem 9, the Hamming error is $\mathrm{FP}_p + \mathrm{FN}_p = L_p p^{1 - h(t'; \vartheta, r)}$, where

$$h(t'; \vartheta, r) = \min\left\{\min\left\{t'^2, \; \vartheta + (t' - |\rho|\sqrt{r})_+^2\right\}, \; \min\left\{\vartheta + (\sqrt{r} - t')_+^2, \; 2\vartheta + [(1+\rho)\sqrt{r} - t']_+^2\right\}\right\}. \tag{26}$$

The first steps of the proof of Elastic net can also be applied directly. We still try to find $t'^*$ that maximizes $h(t'; \vartheta, r)$ and then investigate the conditions on $(r, \vartheta)$ such that $h(t'^*; \vartheta, r) > 1$ or $\vartheta < h(t'^*; \vartheta, r) < 1$ or $h(t'^*; \vartheta, r) \leq \vartheta$.

We can still prove that $r = \vartheta$ is the boundary between the Regions of Almost Full Recovery and No Recovery, i.e., the boundary separating $\vartheta < h(q^*; \vartheta, r) < 1$ and $h(q^*; \vartheta, r) \leq \vartheta$. The proof just needs slight modification to the proof of Elastic net.

For the rest of Part 3, we try to find the boundary between the Regions of Exact Recovery and Almost Full Recovery, i.e., the boundary separating $h(t'^*; \vartheta, r) > 1$ and $\vartheta < h(t'^*; \vartheta, r) < 1$. We can still leverage the important fact that Equation 22 holds at the boundary:

$$\min\{t'^2,\ \vartheta + (t' - |\rho|\sqrt{r})_+^2\} = \min\{\vartheta + (\sqrt{r} - t')_+^2,\ 2\vartheta + [(1+\rho)\sqrt{r} - t']_+^2\} = 1$$

The above equation still gives us $2 \times 2 = 4$ cases respectively for $\rho > 0$ and $\rho < 0$. We discuss them one by one and summarise the results when the full phase curves are complete.

When $\rho > 0$, we set out to prove the final phase diagram is

$$\sqrt{r} = \max\left\{1 + \sqrt{1 - \vartheta},\ \frac{2}{1-\rho}\sqrt{1-\vartheta}\right\}$$

*First*, if $t'^2 = \vartheta + (\sqrt{r} - t')_+^2 = 1$ and $\vartheta + (t' - |\rho|\sqrt{r})_+^2 \geq 1$, $2\vartheta + [(1+\rho)\sqrt{r} - t']_+^2 \geq 1$, then $t' = 1$ and $\sqrt{r} = 1 + \sqrt{1-\vartheta}$.

We also need to meet two requirements: $\sqrt{r} \leq \frac{1-\sqrt{1-\vartheta}}{\rho}$ from $\vartheta + (t' - |\rho|\sqrt{r})_+^2 \geq 1$ and $\sqrt{r} \geq \frac{1+\sqrt{1-2\vartheta}}{1+\rho}$ from $2\vartheta + [(1+\rho)\sqrt{r} - t']_+^2 \geq 1$. The second requirement is not restrictive, and the first one is equivalent to $\sqrt{1-\vartheta} \leq \frac{1-\rho}{1+\rho}$.

*Second*, if $t'^2 = 2\vartheta + [(1+\rho)\sqrt{r} - t']_+^2 = 1$ and $\vartheta + (\sqrt{r} - t')_+^2 \geq 1$, $\vartheta + (t' - |\rho|\sqrt{r})_+^2 \geq 1$, then we have no admissible curve. This is because $\sqrt{r} = \frac{1+\sqrt{1-2\vartheta}}{1+\rho}$ and it is required that $\sqrt{r} \geq 1 + \sqrt{1-\vartheta}$ by $\vartheta + (\sqrt{r} - t')_+^2 \geq 1$. It gives us no admissible $\sqrt{r}$.

*Summarising the first two cases*, we have only one curve $\sqrt{r} = 1 + \sqrt{1-\vartheta}$ under the constraint $\sqrt{1-\vartheta} \leq \frac{1-\rho}{1+\rho}$.

*Third*, if $\vartheta + (t' - |\rho|\sqrt{r})_+^2 = \vartheta + (\sqrt{r} - t')_+^2 = 1$, and $t' \geq 1$, $2\vartheta + [(1+\rho)\sqrt{r} - t']_+^2 \geq 1$, the equality gives us $\sqrt{r} = \frac{2}{1-\rho}\sqrt{1-\vartheta}$ and $t' = \frac{1+\rho}{1-\rho}\sqrt{1-\vartheta}$. We are also required to have $t' \geq 1$ and $\sqrt{r} \geq \frac{\sqrt{1-2\vartheta}}{1+\rho} + \frac{\sqrt{1-\vartheta}}{1-\rho}$ (not restrictive).

*Fourth*, if $\vartheta + (t' - |\rho|\sqrt{r})_+^2 = 2\vartheta + [(1+\rho)\sqrt{r} - t']_+^2 = 1$, and $t' \geq 1$, $\vartheta + (\sqrt{r} - t')_+^2 \geq 1$, then we have no admissible curve. In fact, the equality gives us $\sqrt{r} = \frac{t' - \sqrt{1-\vartheta}}{\rho} = \frac{t' + \sqrt{1-2\vartheta}}{1+\rho}$, $t' = \rho\sqrt{1-2\vartheta} + (1+\rho)\sqrt{1-\vartheta}$. We are also required to have $\sqrt{r} \geq t' + \sqrt{1-\vartheta} \Leftrightarrow \sqrt{1-2\vartheta} \geq \frac{1+\rho}{1-\rho}\sqrt{1-\vartheta}$, which gives a contradiction.

*Summarising the last two cases*, we have only one curve $\sqrt{r} = \frac{2}{1-\rho}\sqrt{1-\vartheta}$ under the constraint $\sqrt{1-\vartheta} \geq \frac{1-\rho}{1+\rho}$.

*Summarising the cases of positive correlation*, we have proven the phase diagram is $\sqrt{r} = \max\left\{1 + \sqrt{1-\vartheta},\ \frac{2}{1-\rho}\sqrt{1-\vartheta}\right\}$.

When $\rho < 0$, we then prove that the phase diagram is

$$\sqrt{r} = \max\left\{1 + \sqrt{1-\vartheta},\ \frac{2}{1-|\rho|}\sqrt{1-\vartheta},\ \frac{\sqrt{1-\vartheta} + \sqrt{1-2\vartheta}}{1-2|\rho|},\ \frac{1+\sqrt{1-2\vartheta}}{1-|\rho|}\right\}$$

*First*, if $t'^2 = \vartheta + (\sqrt{r} - t')_+^2 = 1$ and $\vartheta + (t' - |\rho|\sqrt{r})_+^2 \geq 1$, $2\vartheta + [(1-|\rho|)\sqrt{r} - t']_+^2 \geq 1$, then $t' = 1$ and $\sqrt{r} = 1 + \sqrt{1-\vartheta}$.

We also have two requirements, $\sqrt{r} \leq \frac{1-\sqrt{1-\vartheta}}{|\rho|}$ and $\sqrt{r} \geq \frac{1+\sqrt{1-2\vartheta}}{1-|\rho|}$. We have studied the first one when $\rho > 0$, and know it is $1 + \sqrt{1-\vartheta} \geq \frac{2}{1-|\rho|}\sqrt{1-\vartheta}$. The RHS of the second requirement is actually a new curve we will see later.

*Second*, if $t'^2 = 2\vartheta + [(1-|\rho|)\sqrt{r} - t']_+^2 = 1$ and $\vartheta + (t' - |\rho|\sqrt{r})_+^2 \geq 1$, $\vartheta + (\sqrt{r} - t')_+^2 \geq 1$, we have $\sqrt{r} = \frac{1+\sqrt{1-2\vartheta}}{1-|\rho|}$ and $\sqrt{r} \leq \frac{1-\sqrt{1-\vartheta}}{|\rho|}$, $\sqrt{r} \geq 1 + \sqrt{1-\vartheta}$..

*Summarising the first two cases*, we have $\sqrt{r} = \max\left\{1 + \sqrt{1-\vartheta}, \frac{1+\sqrt{1-2\vartheta}}{1-|\rho|}\right\}$ and we need $\sqrt{r} \le \frac{1-\sqrt{1-\vartheta}}{|\rho|}$.

*Third*, if $\vartheta + (t' - |\rho|\sqrt{r})_+^2 = \vartheta + (\sqrt{r} - t')_+^2 = 1$ and $t' \ge 1$, $2\vartheta + [(1-|\rho|)\sqrt{r} - t']_+^2 \ge 1$, then $\sqrt{r} = \frac{2}{1-|\rho|}\sqrt{1-\vartheta}$, and the two other requirements are $t' = \frac{1+|\rho|}{1-|\rho|}\sqrt{1-\vartheta} \ge 1$ and $\sqrt{r} \ge \frac{\sqrt{1-2\vartheta}}{1-|\rho|} + \frac{1+|\rho|}{(1-|\rho|)^2}\sqrt{1-\vartheta}$.

In the next case, we will get another curve $\sqrt{r} = \frac{\sqrt{1-\vartheta}+\sqrt{1-2\vartheta}}{1-2|\rho|}$. In the above inequalities, the last one corresponds to $\frac{2}{1-|\rho|}\sqrt{1-\vartheta} \ge \frac{\sqrt{1-\vartheta}+\sqrt{1-2\vartheta}}{1-2|\rho|}$. $t' = \frac{1+|\rho|}{1-|\rho|}\sqrt{1-\vartheta} \ge 1$ in the above inequalities is just $1 + \sqrt{1-\vartheta} \le \frac{2}{1-|\rho|}\sqrt{1-\vartheta}$.

*Fourth*, if $\vartheta + (t' - |\rho|\sqrt{r})_+^2 = 2\vartheta + [(1-|\rho|)\sqrt{r} - t']_+^2 = 1$ and $t' \ge 1$, $\vartheta + (\sqrt{r} - t')_+^2 \ge 1$, then we know from the equality that

$$\sqrt{r} = \frac{t' - \sqrt{1-\vartheta}}{|\rho|} = \frac{t' + \sqrt{1-2\vartheta}}{1-|\rho|}$$

and we will get this when we solve for $t'$:

$$\frac{t'}{|\rho|} - \frac{t'}{1-|\rho|} = \frac{\sqrt{1-2\vartheta}}{1-|\rho|} + \frac{\sqrt{1-\vartheta}}{|\rho|}.$$

If $|\rho| \ge \frac{1}{2}$, this equation admits no positive solution for $t'$. Recall that in the *first and second* cases, we also needed $\sqrt{r} \le \frac{1-\sqrt{1-\vartheta}}{|\rho|}$ and $\sqrt{r} \ge \frac{1+\sqrt{1-2\vartheta}}{1-|\rho|}$ in the cases of $FP_1$ being tight. When $|\rho| \ge \frac{1}{2}$, $\frac{1+\sqrt{1-2\vartheta}}{1-|\rho|} \le \frac{1-\sqrt{1-\vartheta}}{|\rho|}$ has no solution either, so there cannot be any curve in the interval $0 < \vartheta < \frac{1}{2}$.

if $|\rho| < \frac{1}{2}$, we can proceed to have the two requirements:

$$\begin{cases} t' = & \frac{1}{1-2|\rho|}\left[|\rho|\sqrt{1-2\vartheta} + (1-|\rho|)\sqrt{1-\vartheta}\right] \ge 1 \\ \sqrt{r} = & \frac{\sqrt{1-\vartheta}+\sqrt{1-2\vartheta}}{1-2|\rho|} \ge \frac{|\rho|}{1-2|\rho|}\sqrt{1-2\vartheta} + \frac{2-3|\rho|}{1-2|\rho|}\sqrt{1-\vartheta} \end{cases}$$

In the above inequalities, $\sqrt{r} = \frac{\sqrt{1-\vartheta}+\sqrt{1-2\vartheta}}{1-2|\rho|} \ge \frac{|\rho|}{1-2|\rho|}\sqrt{1-2\vartheta} + \frac{2-3|\rho|}{1-2|\rho|}\sqrt{1-\vartheta}$ is equivalent to $\frac{\sqrt{1-\vartheta}+\sqrt{1-2\vartheta}}{1-2|\rho|} \ge \frac{2}{1-|\rho|}\sqrt{1-\vartheta}$. The requirement on $t'$ is equivalent to $\sqrt{r} = \frac{\sqrt{1-\vartheta}+\sqrt{1-2\vartheta}}{1-2|\rho|} \ge \frac{1+\sqrt{1-2\vartheta}}{1-|\rho|}$.

*Summarising the cases of negative correlation*: The final phase diagram is

$$\sqrt{r} = \max\left\{1 + \sqrt{1-\vartheta}, \frac{2}{1-|\rho|}\sqrt{1-\vartheta}, \frac{\sqrt{1-\vartheta}+\sqrt{1-2\vartheta}}{1-2|\rho|}, \frac{1+\sqrt{1-2\vartheta}}{1-|\rho|}\right\}$$

When $\rho \le -\frac{1}{2}$, $\sqrt{r}$ has no finite value for $\vartheta \in (0, 1/2)$, and we do not have Region of Exact Recovery or $h(t'; \vartheta, r) > 1$ at all.

## D   PROOF OF THEOREM 3 (SCAD)

As described in Section A, our proof for SCAD still consists of three parts: (a) deriving the rejection region, (b) obtaining the rate of convergence of $\mathbb{E}[H(\hat{\beta}, \beta)]$, and (c) calculating the phase diagram.

Before starting, we first recall the definition of SCAD. An alternative way to write the derivative of the penalty function is:

$$q'(\theta) = \begin{cases} \lambda \cdot \operatorname{sgn}(\theta) & \text{if } |\theta| < \lambda \\ \frac{a\lambda - \theta}{a-1} \cdot \operatorname{sgn}(\theta) & \text{if } \lambda < |\theta| < a\lambda \\ 0 & \text{if } |\theta| > a\lambda \end{cases} \tag{27}$$

for $\theta \in \mathbb{R}$, $a > 2\lambda > 0$.

**Part 1: Deriving the rejection region.** Just like the first steps of Elastic net, we define $h = (h_1, h_2)' \in \mathbb{R}^2$ where $h_1 = x'_j y / \sqrt{2 \log(p)}$, $h_2 = x'_{j+1} y / \sqrt{2 \log(p)}$; $\lambda = \lambda' \sqrt{2 \log(p)}$; $(\hat{\beta}_j, \hat{\beta}_{j+1}) = \sqrt{2 \log(p)} (\hat{b}_1, \hat{b}_2)$ are the entries of the estimator $\hat{\beta}^{\text{SCAD}}$ corresponding to $(x_j, x_{j+1})$.

The estimator of SCAD is $\hat{\beta}^{\text{SCAD}} = \text{argmin}_\beta \left\{ \|y - X\beta\|^2 / 2 + Q_\lambda(\beta) \right\}$. Like Elastic net, it can be decomposed into bivariate sub-problems of each pair of correlated variables. By dividing the bivariate sub-problem of $(x_j, x_{j+1})$ by $\sqrt{2 \log(p)}$, we have

$$L(b) \equiv \frac{1}{2} b' \begin{bmatrix} 1 & \rho \\ \rho' & 1 \end{bmatrix} b + b'h + \lambda'(q'(b_1) + q'(b_2)) \tag{28}$$

and the minimizer of the optimization (28) is $(\hat{b}_1, \hat{b}_2)$. The next lemma proves the solution to (28) when $h_1 > |h_2|$, and it is proven in Section D.1.

**Lemma 2** (the solution path of SCAD). *Consider the optimization in* (28). *Suppose* $h_1 > |h_2|$, *and suppose* $\rho \geq 0$.

- *When* $\lambda' \geq \lambda'_1 = \max\{|h_1|, |h_2|\}$, $\hat{b}_1 = \hat{b}2 = 0$.

- *If* $(\rho - \frac{1}{a})h_1 < h_2 < \begin{cases} (\rho + \frac{1}{a})h_1 & \text{if } a > \frac{2}{1-\rho}, \\ \frac{1+\rho}{2}h_1 & \text{if } a \leq \frac{2}{1-\rho}, \end{cases}$

    1. *When* $\lambda' < \lambda'_1$ *and* $\lambda' \geq |h_2 - \rho h_1|$, $\hat{b}_1 \neq 0$ *and* $\hat{b}_2 = 0$.

    2. *When* $\lambda' < |h_2 - \rho h_1|$, $\hat{b}_1 \neq 0$ *and* $\hat{b}_2 \neq 0$.

- *If* $a > \frac{2}{1-\rho}$ *and* $(\rho + \frac{1}{a})h_1 < h_2 < \frac{1+\rho}{2}h_1$:

    1. *When* $\lambda' < \lambda'_1$ *and* $\lambda' \geq \frac{(a-2)h_2 - \rho(a-1)h_1}{a-2-a\rho}$, $\hat{b}_1 \neq 0$ *and* $\hat{b}_2 = 0$.

    2. *When* $\lambda' < \frac{(a-2)h_2 - \rho(a-1)h_1}{a-2-a\rho}$, $\hat{b}_1 \neq 0$ *and* $\hat{b}_2 \neq 0$.

- *If* $h_2 < (\rho - \frac{1}{a})h_1$, $\forall a$:

    1. *When* $\lambda' < \lambda'_1$ *and* $\lambda' \geq \frac{\rho(a-1)h_1 - (a-2)h_2}{a+a\rho-2}$, $\hat{b}_1 \neq 0$ *and* $\hat{b}_1 = 0$.

    2. *When* $\lambda' < \frac{\rho(a-1)h_1 - (a-2)h_2}{a+a\rho-2}$, $\hat{b}_1 \neq 0$ *and* $\hat{b}_2 \neq 0$.

- *If* $h_2 \leq \frac{-1+\rho}{2}h_1$ *or* $h_2 \geq \frac{1+\rho}{2}h_1$:

    1. *When* $\lambda' < \lambda'_1$ *and* $\lambda' \geq \lambda'^{(2)}_2 = \frac{|h_2 - \rho h_1|}{1-\rho}$, $\hat{b}_1 \neq 0$ *and* $\hat{b}_1 = 0$.

    2. *When* $\lambda' < \lambda'^{(2)}_2$, $\hat{b}_1 \neq 0$ *and* $\hat{b}_2 \neq 0$.

We did not require $\rho \geq 0$ in the solution path of Elastic net, but here $\rho \geq 0$ is needed to cut down unnecessary discussion. The proof of Elastic net has shown that the solution path of $h_1 > |h_2|$ and $\rho \geq 0$ is enough to draw the whole rejection region.

Still requiring $\rho > 0$, the rejection region looks different for $a > \frac{2}{1-\rho}$ and $a \leq \frac{2}{1-\rho}$. The first steps are the same as those of Elastic net, and we only present the rejection region here:

When $a \geq \frac{2}{1-\rho}$:

$$\mathcal{R} = \{(h_1, h_2) : h_1 - \rho h_2 > \lambda'(1-\rho), h_1 > \lambda', h_1 - \frac{|\rho|(a-1)}{a-2}h_2 > \lambda'(1 - \frac{a\rho}{a-2})\}$$
$$\cup \{(h_1, h_2) : h_1 - \rho h_2 > \lambda', h_2 > a\lambda'\} \cup \{(h_1, h_2) : h_1 - \rho h_2 > \lambda'(1+\rho)\}$$
$$\cup \{(h_1, h_2) : h_1 - \frac{\rho(a-1)}{a-2}h_2 > \lambda'(1 + \frac{a\rho}{a-2}), h_1 - \rho h_2 > \lambda'\}$$
$$\cup \{(h_1, h_2) : -h_1 + \rho h_2 > \lambda'(1-\rho), h_1 < -\lambda', -h_1 + \frac{|\rho|(a-1)}{a-2}h_2 > \lambda'(1 - \frac{a\rho}{a-2})\}$$
$$\cup \{(h_1, h_2) : -h_1 + \rho h_2 > \lambda', h_2 < -a\lambda'\} \cup \{(h_1, h_2) : -h_1 + \rho h_2 > \lambda'(1+\rho)\}$$
$$\cup \{(h_1, h_2) : -h_1 + \frac{\rho(a-1)}{a-2}h_2 > \lambda'(1 + \frac{a\rho}{a-2}), -h_1 + \rho h_2 > \lambda'\} \tag{29}$$

When $a \leq \frac{2}{1-\rho}$:

$$\mathcal{R} = \{(h_1, h_2) : h_1 - \rho h_2 > \lambda'(1-\rho), h_1 > \lambda', h_1 > \frac{1+\rho}{2}h_2\}$$
$$\cup \{(h_1, h_2) : h_1 - \rho h_2 > \lambda', h_2 > \frac{2\lambda'}{1-\rho}\} \cup \{(h_1, h_2) : h_1 - \rho h_2 > \lambda'(1+\rho)\}$$
$$\cup \{(h_1, h_2) : h_1 - \frac{\rho(a-1)}{a-2}h_2 > \lambda'(1 + \frac{a\rho}{a-2}), h_1 - \rho h_2 > \lambda'\}$$
$$\cup \{(h_1, h_2) : -h_1 + \rho h_2 > \lambda'(1-\rho), h_1 < -\lambda', h_1 < \frac{1+\rho}{2}h_2\}$$
$$\cup \{(h_1, h_2) : -h_1 + \rho h_2 > \lambda', h_2 < -\frac{2\lambda'}{1-\rho}\} \cup \{(h_1, h_2) : -h_1 + \rho h_2 > \lambda'(1+\rho)\}$$
$$\cup \{(h_1, h_2) : -h_1 + \frac{\rho(a-1)}{a-2}h_2 > \lambda'(1 + \frac{a\rho}{a-2}), -h_1 + \rho h_2 > \lambda'\}$$
$$\tag{30}$$

When $\rho < 0$, we apply the same sign-flipping technique in the proof of Elastic net and still use the rejection region of positive correlation. Such technique requires considering one more case for $\rho < 0$, which is the elliptical distance from the center $\mu_{11} = ((1 - |\rho|)\sqrt{r}, -(1 - |\rho|)\sqrt{r})$ to $\mathcal{R}^c$ (plotted with positive correlation $|\rho| > 0$).

**Part 2. Analyzing the Hamming error.** we allow $\rho \in (-1, 1)$ from now on. The analysis of Hamming error follows the same procedure as that of Elastic net. It is only the shape of $\mathcal{R}$ which is different. For $a \geq \frac{2}{1-|\rho|}$ and $a \leq \frac{2}{1-|\rho|}$, we respectively present a theorem for the Hamming error rate.

**Theorem 10.** *Suppose the conditions of Theorem 3 hold. Let $\lambda = \lambda'\sqrt{2\log(p)}$ in SCAD. Define a few important points in the rejection region (as noted in Figure 8): $A(\lambda', \lambda')$, $B((1+|\rho|)\lambda', 2\lambda')$, $C((1+ a|\rho|)\lambda', a\lambda')$, $D((1 - |\rho|)\lambda', -2\lambda')$. As $p \to \infty$,*

$$\text{FP}_p = L_p p^{1-\min\{\lambda'^2, \ \vartheta+f_1(\sqrt{r},\lambda')\}}, \qquad \text{FN}_p = L_p p^{1-\min\{\vartheta+f_2(\sqrt{r},\lambda'), \ 2\vartheta+f_3(\sqrt{r},\lambda')\}},$$

*where (below, $d^2_{|\rho|}(u, v)$ is as in Definition 1)*

$$f_1(\sqrt{r}, \lambda') = \begin{cases} (\lambda' - |\rho|\sqrt{r})^2 & \text{if } \sqrt{r} \le \frac{\lambda'}{1+|\rho|} \\ \frac{1}{1-|\rho|^2} d^2_{|\rho|}(A, (|\rho|\sqrt{r}, \sqrt{r})) & \text{if } \frac{\lambda'}{1+|\rho|} \le \sqrt{r} \le \lambda' \\ \frac{1-|\rho|}{1+|\rho|}\lambda'^2 & \text{if } \lambda' \le \sqrt{r} \le 2\lambda' \\ \min\{\frac{\lambda'^2}{1-\rho^2}, \frac{1}{1-\rho^2} d^2(B, (|\rho|\sqrt{r}, \sqrt{r}))\} & \text{if } 2\lambda' \le \sqrt{r} \le \lambda'\left[2 + \frac{|\rho|-\rho^2}{(a-2)(1+|\rho|)}\right] \\ \min\left\{\frac{\lambda'^2}{1-\rho^2}, \frac{\left[\lambda'(1-\frac{a|\rho|}{a-2})+\frac{|\rho|\sqrt{r}}{a-2}\right]^2}{1+\frac{\rho^2(a-1)^2}{(a-2)^2}-\frac{2\rho^2(a-1)}{a-2}}\right\} & \text{if } \sqrt{r} \ge \lambda'\left[2 + \frac{|\rho|-\rho^2}{(a-2)(1+|\rho|)}\right] \end{cases}$$

$$f_2(\sqrt{r}, \lambda') = \begin{cases} \min \begin{cases} (\sqrt{r}-\lambda')^2_+ \\ \frac{1}{1-\rho^2}\left[(1-\rho^2)\sqrt{r}-(1-|\rho|)\lambda'\right] \\ \frac{1}{1+\frac{\rho^2(a-1)^2}{(a-2)^2}-\frac{2\rho^2(a-1)}{a-2}}\left[\lambda'\left(1-\frac{a|\rho|}{a-2}\right)-\sqrt{r}\cdot\left(1-\frac{\rho^2(a-1)}{a-2}\right)\right]^2 \end{cases} & \text{if } \sqrt{r} \le \frac{a(a-2)(1-\rho^2)+|\rho|}{|\rho|(a-1)(1-\rho^2)}\lambda' \\ \frac{1}{1-\rho^2}d^2_{|\rho|}(C, (\sqrt{r}, |\rho|\sqrt{r})) & \text{if } \frac{a(a-2)(1-\rho^2)+|\rho|}{|\rho|(a-1)(1-\rho^2)}\lambda' \le \sqrt{r} \le \frac{a\lambda'}{|\rho|} \\ \frac{1}{1-\rho^2}\left[(1-\rho^2)\sqrt{r}-\lambda'\right]^2 & \text{if } \sqrt{r} \ge \frac{a\lambda'}{|\rho|} \end{cases}$$

*The definition of $f_3(\sqrt{r}, \lambda')$ is different for different signs of $\rho$. When $\rho > 0$:*

$$f_3(\sqrt{r}, \lambda') = \frac{1}{1-\rho^2} \cdot \begin{cases} \min \begin{cases} [(1-\rho^2)\sqrt{r}-(1-\rho)\lambda']^2_+ \\ h(\sqrt{r}, \lambda') \end{cases} & \text{if } \sqrt{r} \le \frac{a\lambda'}{1+\rho} \\ [(1-\rho^2)\sqrt{r}-\lambda']^2 & \text{if } \sqrt{r} \ge \frac{a\lambda'}{1+\rho} \end{cases}$$

*where*

$$h(\sqrt{r}, \lambda') = \begin{cases} \frac{(1-\rho^2)}{1+\frac{\rho^2(a-1)^2}{(a-2)^2}-\frac{2\rho^2(a-1)}{a-2}}\left[\lambda'\left(1-\frac{a\rho}{a-2}\right)-(1+\rho)\sqrt{r}\cdot\left(1-\frac{\rho(a-1)}{a-2}\right)\right]^2 & \text{if } \sqrt{r} \le \frac{\lambda'}{1+\rho}\cdot\frac{a(a-2)(1-\rho^2)+\rho}{(a-2)(1-\rho^2)+\rho-\rho^2} \\ d^2(C, ((1+\rho)\sqrt{r}, (1+\rho)\sqrt{r})) & \text{if } \sqrt{r} \ge \frac{\lambda'}{1+\rho}\cdot\frac{a(a-2)(1-\rho^2)+\rho}{(a-2)(1-\rho^2)+\rho-\rho^2} \end{cases}$$

*When $\rho < 0$,*

$$f_3(\sqrt{r}, \lambda') = \frac{1}{1-\rho^2} \cdot \begin{cases} [(1-\rho^2)\sqrt{r}-(1+|\rho|)\lambda']^2 & \text{if } \sqrt{r} \le \frac{2\lambda'}{1-|\rho|} \\ \min \begin{cases} [(1-\rho^2)\sqrt{r}-\lambda']^2 \\ k(\lambda', a) \end{cases} & \text{if } \sqrt{r} \ge \frac{2\lambda'}{1-|\rho|} \end{cases}$$

*where*

$$k(\lambda', a) = \begin{cases} d^2\left(D, ((1-|\rho|)\sqrt{r}, -(1-|\rho|)\sqrt{r})\right) & \text{if } \frac{2\lambda'}{1-|\rho|} \le \sqrt{r} \le \frac{\lambda'}{1-|\rho|}\left[2+\frac{|\rho|+\rho^2}{(a-2)(1-\rho^2)-(|\rho|+\rho^2)}\right] \\ \frac{(1-\rho^2)}{1+\frac{\rho^2(a-1)^2}{(a-2)^2}-\frac{2\rho^2(a-1)}{a-2}}\left[-\lambda'\left(1+\frac{a|\rho|}{a-2}\right)+(1-|\rho|)\sqrt{r}\cdot\left(1+\frac{|\rho|(a-1)}{a-2}\right)\right]^2 \\ \qquad \text{if } \sqrt{r} \ge \frac{\lambda'}{1-|\rho|}\left[2+\frac{|\rho|+\rho^2}{(a-2)(1-\rho^2)-(|\rho|+\rho^2)}\right] \end{cases}$$

*Proof of Theorem 10.* See the rejection region in Figure 8 for a visualization of the rejection region.

*The ellipsoid centered at the point $\mu_{00}$:* Easy to see the rate is $L_p \cdot p^{1-\lambda'^2}$.

*The ellipsoid centered at the point $\mu_{01}$ :* We set out to find out $f_1(\sqrt{r}, \lambda')$. Similar to Lasso, we have: (i) When $\sqrt{r} \le \frac{\lambda'}{1+|\rho|}$, $f_1(\sqrt{r}, \lambda') = (\lambda' - |\rho|\sqrt{r})^2$. (ii) When $\frac{\lambda'}{1+|\rho|} \le \sqrt{r} \le \lambda'$, $f_1(\sqrt{r}, \lambda') = \frac{1}{1-\rho^2}d^2(A, (|\rho|\sqrt{r}, \sqrt{r}))$. The point $A$ is noted in Figure 8. (iii) When $\lambda' \le \sqrt{r} \le 2\lambda'$, $f_1(\sqrt{r}, \lambda') = \frac{1-|\rho|}{1+|\rho|}\lambda'^2$.

Then we need to investigate the green segment in Figure 8. When the ellipsoid is tangent to the green segment on the *right* side (i.e. $BC$), and the tangent point is above Point $B$, then using Lemma A.1,

$$\sqrt{r} + \frac{-\frac{|\rho|}{a-2}\left[\lambda'\left(1-\frac{a|\rho|}{a-2}\right)+\sqrt{r}\cdot\frac{|\rho|}{a-2}\right]}{1+\frac{\rho^2(a-1)^2}{(a-2)^2}-\frac{2\rho^2(a-1)}{a-2}} \ge 2\lambda'$$
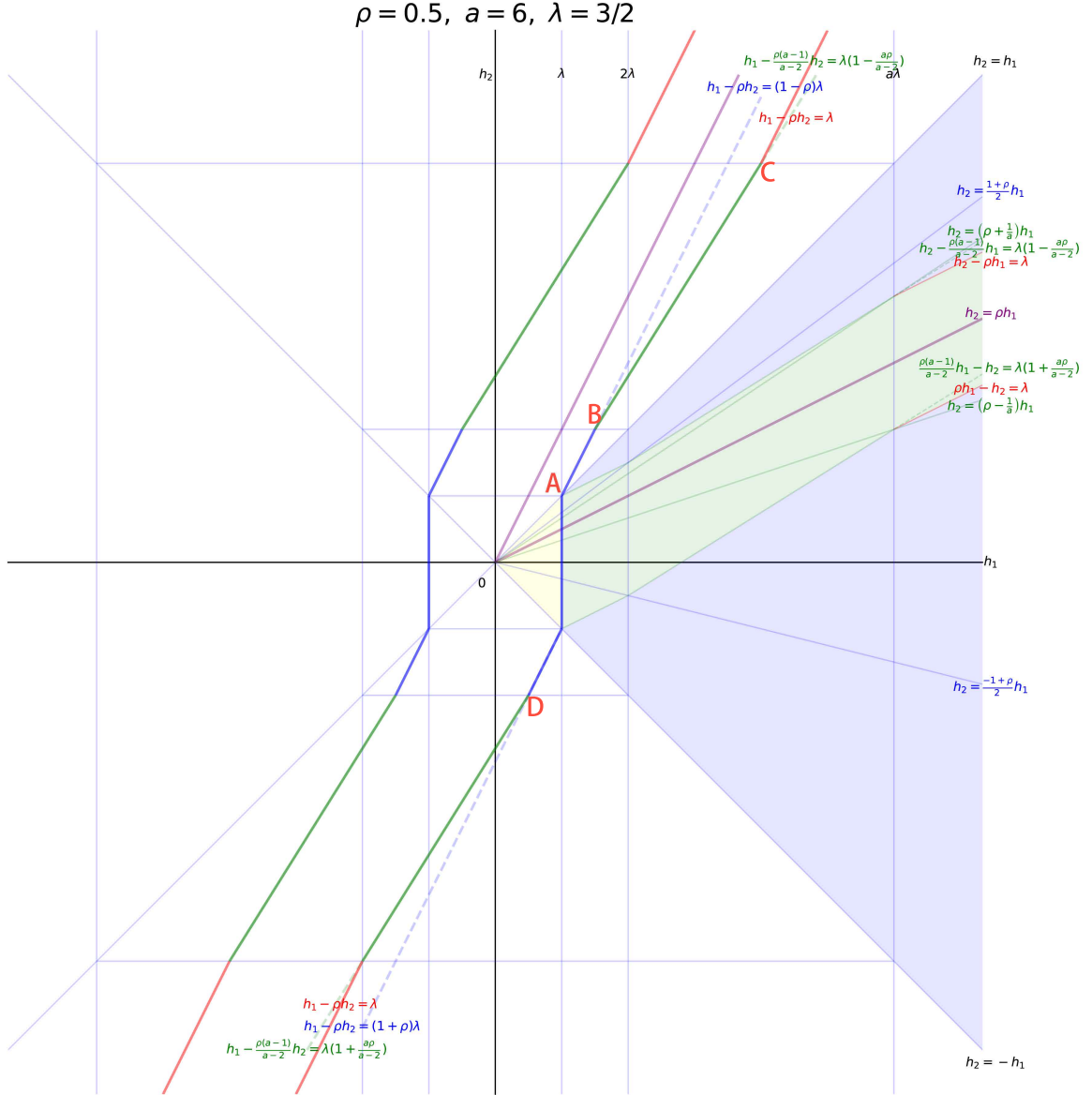
Figure 8: When $a > \frac{2}{1-|\rho|}$, the rejection region looks like this

which implies $\sqrt{r} \geq \lambda'\left[2 + \frac{|\rho|-\rho^2}{(a-2)(1+|\rho|)}\right]$

When $\sqrt{r} \geq \lambda'\left[2 + \frac{|\rho|-\rho^2}{(a-2)(1+\rho)}\right]$, the ellipsoid either intersects with the green line segment $BC$, or the red segment beyond $C$. However, we need to eliminate the possibility of the ellipsoid having a smaller radius when tangent to the segments on the left.

We will see that the line segments on the left can indeed be eliminated. This is because when the ellipsoid is tangent to both the green lines on the left and right,

$$\frac{(1-\rho^2)\left[\lambda'(1-\frac{a|\rho|}{a-2}) + \frac{|\rho|\sqrt{r}}{a-2}\right]^2}{1 + \frac{\rho^2(a-1)^2}{(a-2)^2} - \frac{2\rho^2(a-1)}{a-2}} = \frac{(1-\rho^2)\left[-\lambda'(1+\frac{a|\rho|}{a-2}) + \frac{|\rho|\sqrt{r}}{a-2}\right]^2}{1 + \frac{\rho^2(a-1)^2}{(a-2)^2} - \frac{2\rho^2(a-1)}{a-2}}$$

$$\implies \lambda'(1-\frac{a|\rho|}{a-2}) + \frac{|\rho|\sqrt{r}}{a-2} = \lambda'(1+\frac{a|\rho|}{a-2}) - \frac{|\rho|\sqrt{r}}{a-2} \implies \sqrt{r} = a\lambda$$

30

As a result, when $\sqrt{r} \leq a\lambda$, we can ignore the possibility of the ellipsoid intersecting the green or red segments on the left side. When $\sqrt{r} \geq a\lambda$, the right side still has the smaller distance.

when $2\lambda' \leq \sqrt{r} \leq \lambda' \left[2 + \frac{|\rho| - \rho^2}{(a-2)(1+|\rho|)}\right]$, $f_1(\sqrt{r}, \lambda') = \frac{1}{1-\rho^2} d^2(B, (|\rho|\sqrt{r}, \sqrt{r}))$. The point $B$ is noted in Figure 8.

when $\sqrt{r} \geq \lambda' \left[2 + \frac{|\rho| - \rho^2}{(a-2)(1+|\rho|)}\right]$,

$$f_1(\sqrt{r}, \lambda') = \frac{1}{1-\rho^2} \min \left\{ \lambda'^2, \frac{(1-\rho^2)\left[\lambda'(1 - \frac{a|\rho|}{a-2}) + \frac{|\rho|\sqrt{r}}{a-2}\right]^2}{1 + \frac{\rho^2(a-1)^2}{(a-2)^2} - \frac{2\rho^2(a-1)}{a-2}} \right\}.$$

*The ellipsoid centered at the $\mu_{10}$*: We only explain one tricky point: When the tangent point to the segment $BC$ is precisely Point $C$,

$$|\rho|\sqrt{r} + \frac{-\frac{|\rho|}{a-2}\left[\lambda'\left(1 - \frac{a|\rho|}{a-2}\right) - \sqrt{r} \cdot \left(1 - \frac{\rho^2(a-1)}{a-2}\right)\right]}{1 + \frac{\rho^2(a-1)^2}{(a-2)^2} - \frac{2\rho^2(a-1)}{a-2}} = a\lambda'$$

then $\sqrt{r} = \frac{a(a-2)(1-\rho^2) + |\rho|}{|\rho|(a-1)(1-\rho^2)}$.

*The ellipsoid centered at $\mu_{11} = ((1+\rho)\sqrt{r}, (1+\rho)\sqrt{r})$, when $\rho > 0$* : We still explain only one important point: When the ellipsoid is tangent to the green segment precisely at Point $C$,

$$(1+\rho)\sqrt{r} + \frac{-\frac{\rho}{a-2}\left[\lambda'\left(1 - \frac{a\rho}{a-2}\right) - (1+\rho)\sqrt{r} \cdot \left(1 - \frac{\rho(a-1)}{a-2}\right)\right]}{1 + \frac{\rho^2(a-1)^2}{(a-2)^2} - \frac{2\rho^2(a-1)}{a-2}} = a\lambda'$$

then $(1+\rho)\sqrt{r} = \frac{a(a-2)(1-\rho^2) + \rho}{(a-2)(1-\rho^2) + \rho - \rho^2} \lambda'$.

*The ellipsoid centered at $\mu_{11} = ((1-|\rho|)\sqrt{r}, -(1-|\rho|)\sqrt{r})$, when $\rho < 0$* : We explain one important point: when the ellipsoid is tangent to the green segment at the Point $D$. Now we have

$$-(1-|\rho|)\sqrt{r} + \frac{-\frac{|\rho|}{a-2}\left[\lambda'\left(1 + \frac{a|\rho|}{a-2}\right) - (1-|\rho|)\sqrt{r} \cdot \left(1 + \frac{|\rho|(a-1)}{a-2}\right)\right]}{1 + \frac{\rho^2(a-1)^2}{(a-2)^2} - \frac{2\rho^2(a-1)}{a-2}} = -2\lambda'$$

then $(1-|\rho|)\sqrt{r} = \lambda' \left[2 + \frac{|\rho| + \rho^2}{(a-2)(1-\rho^2) - (|\rho| + \rho^2)}\right]$.

Note that even when $\frac{2\lambda'}{1-|\rho|} \leq \sqrt{r} \leq \frac{\lambda'}{1-|\rho|}\left[2 + \frac{|\rho| + \rho^2}{(a-2)(1-\rho^2) - (|\rho| + \rho^2)}\right]$, and the ellipsoid intersects with the rejection region at Point $D$, it may be tangent to the red segment without being tangent to the green segment. This is especially true when $a < 2 + \frac{|\rho|}{1-|\rho|}$ (but this only happens when $a < \frac{2}{1-|\rho|}$, the next section.) $\qquad \square$

**Theorem 11.** *Suppose the conditions of Theorem 3 hold. Let $\lambda = \lambda'\sqrt{2\log(p)}$ in SCAD. Define a few important points (as noted in the rejection region in Figure 9): $A(\lambda', \lambda')$, $B((1+|\rho|)\lambda', 2\lambda')$, $C(\frac{1+|\rho|}{1-|\rho|}\lambda', \frac{2\lambda'}{1-|\rho|})$, $D((1-|\rho|)\lambda', -2\lambda')$. As $p \to \infty$,*

$$\text{FP}_p = L_p p^{1 - \min\left\{\lambda'^2, \ \vartheta + f_1(\sqrt{r}, \lambda')\right\}}, \qquad \text{FN}_p = L_p p^{1 - \min\left\{\vartheta + f_2(\sqrt{r}, \lambda'), \ 2\vartheta + f_3(\sqrt{r}, \lambda')\right\}},$$

where (below, $d^2_{|\rho|}(u, v)$ is as in Definition 1)

$$f_1(\sqrt{r}, \lambda') = \begin{cases} (\lambda' - |\rho|\sqrt{r})^2 & \text{if } \sqrt{r} \leq \frac{\lambda'}{1+|\rho|} \\ \frac{1}{1-\rho^2} d^2(A, (|\rho|\sqrt{r}, \sqrt{r})) & \text{if } \frac{\lambda'}{1+|\rho|} \leq \sqrt{r} \leq \lambda' \\ \frac{1-|\rho|}{1+|\rho|} \lambda'^2 & \text{if } \lambda' \leq \sqrt{r} \leq 2\lambda' \\ \min\left\{ \frac{\lambda'^2}{1-\rho^2}, \frac{1}{1-\rho^2} d^2(B, (|\rho|\sqrt{r}, \sqrt{r})) \right\} & \text{if } 2\lambda' \leq \sqrt{r} \leq \frac{5+3|\rho|}{2+2|\rho|} \lambda' \\ \min\left\{ \frac{\lambda'^2}{1-\rho^2}, \frac{(1-|\rho|)r}{(5+3|\rho|)} \right\} & \text{if } \sqrt{r} \geq \frac{5+3|\rho|}{2+2|\rho|} \lambda' \end{cases}$$

$$f_2(\sqrt{r}, \lambda') = \begin{cases} \min \begin{cases} (\sqrt{r} - \lambda')^2_+ \\ \frac{1}{1-\rho^2} \left[ (1-\rho^2)\sqrt{r} - (1-|\rho|)\lambda' \right]^2 \\ \frac{(1-|\rho|)(2+|\rho|)^2}{5+3|\rho|} r \end{cases} & \text{if } \sqrt{r} \leq \frac{5+3|\rho|}{(1-|\rho|)(1+|\rho|)^2} \lambda' \\ \frac{1}{1-\rho^2} d^2(C, (\sqrt{r}, |\rho|\sqrt{r})) & \text{if } \frac{5+3|\rho|}{(1-|\rho|)(1+|\rho|)^2} \lambda' \leq \sqrt{r} \leq \frac{2\lambda'}{|\rho|(1-|\rho|)} \\ \frac{1}{1-\rho^2} \left[ (1-\rho^2)\sqrt{r} - \lambda' \right]^2 & \text{if } \sqrt{r} \geq \frac{2\lambda'}{|\rho|(1-|\rho|)} \end{cases}$$

The definition of $f_3(\sqrt{r}, \lambda')$ is different for different signs of $\rho$. When $\rho > 0$:

$$f_3(\sqrt{r}, \lambda') = \frac{1}{1-\rho^2} \cdot \begin{cases} \min \begin{cases} \left[ (1-\rho^2)\sqrt{r} - (1-\rho)\lambda' \right]^2_+ \\ h(\sqrt{r}, \lambda') \end{cases} & \text{if } \sqrt{r} \leq \frac{2\lambda'}{1-\rho^2} \\ \left[ (1-\rho^2)\sqrt{r} - \lambda' \right]^2 & \text{if } \sqrt{r} \geq \frac{2\lambda'}{1-\rho^2} \end{cases}$$

where

$$h(\sqrt{r}, \lambda') = \begin{cases} \frac{(1-\rho^2)^2(1+\rho)}{5+3\rho} r & \text{if } \sqrt{r} \leq \frac{\lambda'}{1+\rho} \cdot \frac{5+3\rho}{(1-\rho)(3+\rho)} \\ d^2(C, ((1+\rho)\sqrt{r}, (1+\rho)\sqrt{r})) & \text{if } \sqrt{r} \geq \frac{\lambda'}{1+\rho} \cdot \frac{5+3\rho}{(1-\rho)(3+\rho)} \end{cases}$$

When $\rho < 0$,

$$f_3(\sqrt{r}, \lambda') = \frac{1}{1-\rho^2} \cdot \begin{cases} \left[ (1-\rho^2)\sqrt{r} - (1+|\rho|)\lambda' \right]^2 & \text{if } \sqrt{r} \leq \frac{2\lambda'}{1-|\rho|} \\ \min \begin{cases} \left[ (1-\rho^2)\sqrt{r} - \lambda' \right]^2 \\ k(\lambda', a) \end{cases} & \text{if } \sqrt{r} \geq \frac{2\lambda'}{1-|\rho|} \end{cases}$$

where

$$k(\lambda', a) = \begin{cases} d^2\left( D, ((1-|\rho|)\sqrt{r}, -(1-|\rho|)\sqrt{r}) \right) & \text{if } \frac{2\lambda'}{1-|\rho|} \leq \sqrt{r} \leq \frac{\lambda'}{1-|\rho|} \left[ 2 + \frac{|\rho|+\rho^2}{(a-2)(1-\rho^2)-(|\rho|+\rho^2)} \right] \\ \frac{(1-\rho^2)}{1+\frac{\rho^2(a-1)^2}{(a-2)^2} - \frac{2\rho^2(a-1)}{a-2}} \left[ -\lambda' \left(1 + \frac{a|\rho|}{a-2}\right) + (1-|\rho|)\sqrt{r} \cdot \left(1 + \frac{|\rho|(a-1)}{a-2}\right) \right]^2 \\ \quad \text{if } \sqrt{r} \geq \frac{\lambda'}{1-|\rho|} \left[ 2 + \frac{|\rho|+\rho^2}{(a-2)(1-\rho^2)-(|\rho|+\rho^2)} \right] \end{cases}$$

*Proof of Theorem 11.* See the different rejection region in Figure 9.

*The ellipsoid centered at $\mu_{00}$:* The rate is $L_p \cdot p^{1-\lambda'^2}$

*The ellipsoid centered at $\mu_{01}$:* Still similar to Lasso, we have

- when $\sqrt{r} \leq \frac{\lambda'}{1+|\rho|}$, $f_1(\sqrt{r}, \lambda') = (\lambda' - |\rho|\sqrt{r})^2$.

- when $\frac{\lambda'}{1+|\rho|} \leq \sqrt{r} \leq \lambda'$, $f_1(\sqrt{r}, \lambda') = \frac{1}{1-\rho^2} d^2(A, (|\rho|\sqrt{r}, \sqrt{r}))$. The point $A$ has been defined in Theorem 11 and noted in Figure 9.

- when $\lambda' \leq \sqrt{r} \leq 2\lambda'$, $f_1(\sqrt{r}, \lambda') = \lambda'^2$.

Then we need to investigate the green segment in Figure 9. When the ellipsoid is tangent to the green segment on the *right* side (i.e. $BC$), and the tangent point is above Point $B$, then using Lemma A.1,

$$\sqrt{r} - \frac{\frac{1}{4}(1-|\rho|)^2\sqrt{r}}{1 + (\frac{1+|\rho|}{2})^2 - |\rho|(1+|\rho|)} \geq 2\lambda'$$

Figure 9: When $a \leq \frac{2}{1-|\rho|}$, the rejection region looks like this

which implies $\sqrt{r} \geq \frac{5+3|\rho|}{2+2|\rho|} \lambda'$.

When $\sqrt{r} \geq \frac{5+3|\rho|}{2+2|\rho|} \lambda'$, the ellipsoid either intersects with the green line segment $BC$, or the red segment beyond $C$. However, we need to eliminate the possibility of the ellipsoid having a smaller radius when tangent to the segments on the left.

Actually, we will see that the line segments on the left can indeed be eliminated, without doing any computation. The case of $a \leq \frac{2}{1-|\rho|}$ is a degenerate case, as we have $\frac{|\rho|(a-1)}{a-2} \geq \frac{1+|\rho|}{2}$. when $a \leq \frac{2}{1-|\rho|}$. From the computation in the $a > \frac{2}{1-|\rho|}$ counterpart, the green and red segments on the left sides can be ignored.

So we can continue the discussion and present the rest two cases:

- when $2\lambda' \leq \sqrt{r} \leq \frac{5+3|\rho|}{2+2|\rho|} \lambda'$, rate $= d^2(B, (|\rho|\sqrt{r}, \sqrt{r}))$. The point $B$ is noted in Figure 9.

33

- when $\sqrt{r} \geq \frac{5+3|\rho|}{2+2|\rho|}\lambda'$,

$$\text{rate} = \min\left\{\lambda'^2,\ \frac{(1-|\rho|^2)(1-|\rho|)r}{(5+3|\rho|)}\right\}$$

*The ellipsoid centered at $\mu_{10}$* : Only one special point needs to be investigated. When the tangent point to the segment $BC$ is precisely Point $C\left(\frac{1+|\rho|}{1-|\rho|}\lambda',\ \frac{2\lambda'}{1-|\rho|}\right)$,

$$|\rho|\sqrt{r} + \frac{-\frac{1-|\rho|}{2}\left[\frac{|\rho|(1+|\rho|)}{2}-1\right]\sqrt{r}}{1+(\frac{1+|\rho|}{2})^2-|\rho|(1+|\rho|)} = \frac{2\lambda'}{1-|\rho|}$$

then $\sqrt{r} = \frac{5+3|\rho|}{(1-|\rho|)(1+|\rho|)^2}\lambda'$.

*The ellipsoid centered at $\mu_{11} = ((1+\rho)\sqrt{r}, (1+\rho)\sqrt{r})$, when $\rho > 0$.* We explain one special point: When the ellipsoid is tangent to the green segment precisely at Point $C$,

$$(1+\rho)\sqrt{r} + \frac{\frac{(1-\rho)(1-\rho^2)}{4}\sqrt{r}}{1+(\frac{1+\rho}{2})^2-\rho(1+\rho)} = \frac{2\lambda'}{1-\rho}$$

then $\sqrt{r} = \frac{5+3\rho}{(1-\rho^2)(3+\rho)}\lambda'$.

*The ellipsoid centered at $\mu_{11} = ((1-|\rho|)\sqrt{r}, -(1-|\rho|)\sqrt{r})$, only when $\rho < 0$:* This case is identical to the counterpart proof for $a \geq \frac{2}{1-|\rho|}$, and nothing needs to be changed. □

**Part 3. Calculating the phase diagram, for $a \leq \frac{2}{1-|\rho|}$.** The boundary between the Regions of Almost Full Recovery and No Recovery is still $r = \vartheta$, and it can be proven in the same manner as that of Elastic net. For the rest of this part, we focus on the boundary between Exact Recovery and Almost Full Recovery.

We focus on the case of $a < \frac{2}{1-|\rho|}$, because: First, the phase diagram of SCAD when $a < \frac{2}{1-|\rho|}$ is worse than Lasso's diagram when $\rho > 0$, and becomes the same as Lasso when $a$ is sufficiently larger than $\frac{2}{1-\rho}$. When $\rho < 0$, the phase diagram is better than that of Lasso when $a \leq \frac{2}{1-|\rho|}$, and numerical results show that when $a > \frac{2}{1-|\rho|}$, the diagram is monotonically moving upwards towards Lasso's diagram when $a$ is increasing. Second, when $a < \frac{2}{1-|\rho|}$, Theorem 11 does not depend on $a$ in its most part, and is much easier to compute. To sum up, the case of $a > \frac{2}{1-|\rho|}$ is much more tedious in computation but less informative.

**We start from the case of $\rho > 0$.** Before diving into the proof, we give an overall account for the diagram:

1. The diagram is the same as that of Lasso, only except that when $\rho < 0.179$, there is a tiny difference in a small neighborhood of $\vartheta = 0$, slightly worse than Lasso. See equation (32)
2. As long as $a < \frac{2}{1-\rho}$, the phase diagram does not depend on the specific value of $a$.

Then we move on to the proof, which has four cases just like the proof of Elastic net.

*First, $\lambda'^2 = \vartheta + f_2(\sqrt{r}, \lambda') = 1$ and $\vartheta + f_1(\sqrt{r}, \lambda') \geq 1, 2\vartheta + f_3(\sqrt{r}, \lambda') \geq 1$:* We know $\lambda' = 1$. From the definitin of $f_2(\sqrt{r}, \lambda')$, we know $\sqrt{r} \geq 1 + \sqrt{1-\vartheta} > 1$.

Then we start our discussion on the conditional expression of $f_2(\sqrt{r}, \lambda')$. (Note that numerically, $\min_\rho \frac{5+3\rho}{(1-\rho)(1+\rho)^2} = 4.848$.)

If $1 - \vartheta = \frac{1}{1-\rho^2}d^2(C, (\rho\sqrt{r}, \sqrt{r}))$: As we know $d^2(C, (\rho\sqrt{r}, \sqrt{r})) \geq \frac{(1-\rho^2)(1-\rho)(2+\rho)^2}{5+3\rho}r$, so $\sqrt{r} \leq \sqrt{1-\vartheta}\sqrt{\frac{5+3\rho}{(1-\rho)(2+\rho)^2}}$, which contradicts the pre-condition that $\sqrt{r} \geq \frac{5+3\rho}{(1-\rho)(1+\rho)^2}$. We have no curve in this case. If $1 - \vartheta = \frac{1}{1-\rho^2}[(1-\rho^2)\sqrt{r}-1]^2$, then $\sqrt{r} = \sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{1}{1-\rho^2}$. This again contradicts the pre-condition that $\sqrt{r} \geq \frac{2}{\rho(1-\rho)}$. We have no curve in this case.

As a result, we can only have $1 < \sqrt{r} < \frac{5+3\rho}{(1-\rho)(1+\rho)^2}$ and one of the following three:

$$\sqrt{r} = 1 + \sqrt{1 - \vartheta}$$

$$\sqrt{r} = \sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{1}{1+\rho}$$

$$\sqrt{r} = \sqrt{1-\vartheta}\sqrt{\frac{5+3\rho}{(1-\rho)(2+\rho)^2}}$$

We then discuss the three curves one by one, starting from the last one.

1. $\sqrt{r} = \sqrt{1-\vartheta}\sqrt{\frac{5+3\rho}{(1-\rho)(2+\rho)^2}}$. We need to look at $FP_2$ to eliminate this curve.

   When $1 \le \sqrt{r} \le 2$, we have $\sqrt{1-\vartheta} \le \sqrt{\frac{1-\rho}{1+\rho}}$. In $FN_1$, for the last term to be the minimum among the three

   $$\min \begin{cases} (1-\rho^2)(\sqrt{r} - \lambda')_+^2 \\ \left[(1-\rho^2)\sqrt{r} - (1-\rho)\lambda'\right]^2 \\ \frac{(1-\rho^2)(1-\rho)(2+\rho)^2}{5+3\rho}r \end{cases},$$

   we need $\sqrt{1-\vartheta}\sqrt{\frac{5+3\rho}{(1-\rho)(2+\rho)^2}} \ge \sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{1}{1+\rho}$, which implies

   $$\sqrt{\frac{1-\rho}{1+\rho}}\left[\sqrt{\frac{5+3\rho}{(1-\rho)(2+\rho)^2}} - \frac{1}{\sqrt{1-\rho^2}}\right] \ge \frac{1}{1+\rho}.$$

   Simplify this for a few steps and we can see the contradiction.

   When $\sqrt{r} \ge \frac{5+3\rho}{2+2\rho}$, we can see the contradiction by simplifying this inequality itself.

   When $2 < \sqrt{r} < \frac{5+3\rho}{2+2\rho}$, by looking at

   $$\sqrt{1-\vartheta}\sqrt{\frac{5+3\rho}{(1-\rho)(2+\rho)^2}} \le \frac{5+3\rho}{2+2\rho}$$

   $$\sqrt{1-\vartheta}\sqrt{\frac{5+3\rho}{(1-\rho)(2+\rho)^2}} \ge \frac{1}{1+\rho} + \sqrt{\frac{1-\vartheta}{1-\rho^2}}$$

   we can see that no $\rho \in (0,1)$ can admit a possible $\sqrt{1-\vartheta}$.

2. $\sqrt{r} = \frac{1}{1+\rho} + \sqrt{\frac{1-\vartheta}{1-\rho^2}}$. We need to look at $FP_2$ to eliminate this curve.

   We already have $\sqrt{r} \ge 1$; when $1 < \sqrt{r} \le 2$, from the rate of $FP_2$, we have $1 - \vartheta \le \frac{1-\rho}{1+\rho}$. However, In $FN_1$, for the middle term to be the minimum among the three

   $$\min \begin{cases} (1-\rho^2)(\sqrt{r} - \lambda')_+^2 \\ \left[(1-\rho^2)\sqrt{r} - (1-\rho)\lambda'\right]^2 \\ \frac{(1-\rho^2)(1-\rho)(2+\rho)^2}{5+3\rho}r \end{cases},$$

   we need

   $$\frac{1}{1+\rho} + \sqrt{\frac{1-\vartheta}{1-\rho^2}} > 1 + \sqrt{1-\vartheta}.$$

   The upper and lower bound of $\sqrt{1-\vartheta}$ would render this case impossible.

   When $2 < \sqrt{r} \le \frac{5+3\rho}{2+2\rho}$, using the expression of (34), we need

   $$\rho^2 r - 2(1+\rho)\sqrt{r} + 4 \ge 0$$

in which we use $\sqrt{r}$ to express $\sqrt{1-\vartheta}$. By letting $\sqrt{r} = 2$ or $\frac{5+3\rho}{2+2\rho}$, we can see they are both negative, so we have a contradiction.

When $\sqrt{r} > \frac{5+3\rho}{2+2\rho}$, we have

$$\frac{1}{1+\rho} + \sqrt{\frac{1-\vartheta}{1-\rho^2}} > \frac{5+3\rho}{2+2\rho}$$

$$\sqrt{\frac{5+3\rho}{1-\rho}}\sqrt{1-\vartheta} \leq \frac{1}{1+\rho} + \sqrt{\frac{1-\vartheta}{1-\rho^2}}$$

and the upper bound on $\sqrt{1-\vartheta}$ is even smaller than the lower bound. Contradiction.

Now we are only left with $\sqrt{r} = 1 + \sqrt{1-\vartheta}$. We need it to meet the following requirements:

$$\begin{cases} 1 + \sqrt{1-\vartheta} \geq \frac{1}{1+\rho} + \sqrt{\frac{1-\vartheta}{1-\rho^2}} & \text{for it to be the smallest among the three} \\ 1 + \sqrt{1-\vartheta} \geq \sqrt{1-\vartheta}\sqrt{\frac{5+3\rho}{(1-\rho)(2+\rho)^2}} & \text{for it to be the smallest among the three} \\ 1 + \sqrt{1-\vartheta} \leq \frac{5+3\rho}{(1-\rho)(1+\rho)^2} & \text{pre-condition; not restrictive} \\ \vartheta > \frac{2\rho}{1+\rho} & \text{for } \vartheta + f_1(\sqrt{r}, \lambda') \geq 1 \\ 2\vartheta + f_3(\sqrt{r}, \lambda') \geq 1 \end{cases}$$

Among the first 4 requirements, the fourth one is can imply all of the rest. Then we look at $FN_2$, and show it is always $o(1)$ when $\vartheta > \frac{2\rho}{1+\rho}$. When $\rho > \frac{1}{3}$, $\vartheta > \frac{1}{2}$; when $\rho < \frac{1}{3}$, we discuss as follows:

- When $\sqrt{r} \leq \frac{\lambda'}{1+\rho} \cdot \frac{5+3\rho}{(1-\rho)(3+\rho)}$,

  For the first term, we need $1 + \sqrt{1-\vartheta} > \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{1+\rho}$ so that the exponent is negative. $(LHS - RHS)$ is increasing in $\vartheta$, and verifying $\vartheta = \frac{2\rho}{1+\rho}$ is enough.

  For the next term, we need $\sqrt{\frac{5+3\rho}{(1-\rho^2)(1+\rho)}}\sqrt{1-2\vartheta} \leq 1 + \sqrt{1-\vartheta}$. Now we need to verify $\vartheta = \max\{\frac{2\rho}{1+\rho}, 1 - \left(\frac{5+3\rho}{(1-\rho^2)(\rho+3)} - 1\right)^2\}$, and it still holds.

  *As long as $k \leq 2$, $k\sqrt{1-\vartheta} - \sqrt{1-2\vartheta}$ is increasing in $\vartheta$.* $\sqrt{\frac{(1-\rho^2)(1+\rho)}{5+3\rho}} \leq 0.454167$ *numerically.*

- When $\sqrt{r} \geq \frac{2\lambda'}{1-\rho^2}$: impossible, because $\sqrt{r} = 1 + \sqrt{1-\vartheta}$.

- When $\frac{\lambda'}{1+\rho} \cdot \frac{5+3\rho}{(1-\rho)(3+\rho)} \leq \sqrt{r} \leq \frac{2\lambda'}{1-\rho^2}$: We know

$$d^2(C, ((1+\rho)\sqrt{r}, (1+\rho)\sqrt{r})) \geq \max\left\{\frac{(1-\rho^2)^2(1+\rho)}{5+3\rho}r, \left[(1-\rho^2)\sqrt{r} - \lambda'\right]^2\right\}$$

  So a sufficient condition is

$$\sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{1-\rho^2} \leq 1 + \sqrt{1-\vartheta}$$

  with $\vartheta = \frac{2\rho}{1+\rho}$. When $\rho \leq \frac{1}{3}$, this holds.

To conclude, we have verified that $FP_1 = FN_1$ can only admit $\sqrt{r} = 1 + \sqrt{1-\vartheta}$, for $\vartheta > \frac{2\rho}{1+\rho}$; and this curve indeed meets all the requirements.

*Second*, if $\vartheta + f_1(\sqrt{r}, \lambda') = \vartheta + f_2(\sqrt{r}, \lambda') = 1$, and $\lambda' \geq 1$, $2\vartheta + f_3(\sqrt{r}, \lambda') \geq 1$, we also need to discuss along the conditional expression of $f_1(\sqrt{r}, \lambda')$.

When $\lambda' \leq \sqrt{r} \leq 2\lambda'$ in $FP_2$, now we have $\lambda' = \sqrt{\frac{1+\rho}{1-\rho}}\sqrt{1-\vartheta}$. Because we want $\lambda' \geq 1$, all we need is to require $\lambda' > 1$, which implies $\vartheta < \frac{2\rho}{1+\rho}$.

Since $\min_\rho \frac{5+3\rho}{(1-\rho)(1+\rho)^2} = 4.848$, we only need to consider the first case in conditional expression of $f_2(\sqrt{r}, \lambda')$. This gives us three possible curves:

$$\sqrt{r} = (1 + \sqrt{\frac{1+\rho}{1-\rho}})\sqrt{1-\vartheta}$$

$$\sqrt{r} = 2\sqrt{\frac{1-\vartheta}{1-\rho^2}}$$

$$\sqrt{r} = \sqrt{\frac{5+3\rho}{(1-\rho)(2+\rho)^2}}\sqrt{1-\vartheta}$$

Actually, $\forall \rho \in (0,1)$, $1 + \sqrt{\frac{1+\rho}{1-\rho}} > \max\left\{\frac{2}{\sqrt{1-\rho^2}}, \sqrt{\frac{5+3\rho}{(1-\rho)(2+\rho)^2}}\right\}$, which means that in the expression of $f_2$ when

$$\min \begin{cases} (\sqrt{r} - \lambda')_+^2 \\ \frac{1}{1-\rho^2}\left[(1-\rho^2)\sqrt{r} - (1-\rho)\lambda'\right]^2 \\ \frac{(1-\rho)(2+\rho)^2}{5+3\rho}r \end{cases},$$

neither of the last two lines cannot produce a curve and be the minimum at the same time.

So we are only left with $\sqrt{r} = (1 + \sqrt{\frac{1+\rho}{1-\rho}})\sqrt{1-\vartheta}$. When $\vartheta < \frac{2\rho}{1+\rho}$, we already have $\lambda' \geq 1$, $\vartheta + f_1(\sqrt{r}, \lambda') \geq 1$ and $\vartheta + f_2(\sqrt{r}, \lambda') \geq 1$, and we are left to verify $2\vartheta + f_3(\sqrt{r}, \lambda') \geq 1$. Actually, this does not always hold when $\rho$ is very small, in which case we need one more curve.

When $\rho \geq 0.197$, $\sqrt{r} \leq \lambda' \cdot \frac{5+3\rho}{(1-\rho^2)(3+\rho)}$ always holds, and we only need to consider the first case. We need

$$\begin{cases} (1 + \sqrt{\frac{1+\rho}{1-\rho}})\sqrt{1-\vartheta} \geq \frac{1}{1+\rho}\sqrt{\frac{1+\rho}{1-\rho}}\sqrt{1-\vartheta} + \sqrt{\frac{1-2\vartheta}{1-\rho^2}} \\ (1 + \sqrt{\frac{1+\rho}{1-\rho}})\sqrt{1-\vartheta} \geq \sqrt{1-2\vartheta}\sqrt{\frac{5+3\rho}{(1+\rho)(1-\rho^2)}} \end{cases}$$

The first one always holds for $\rho \in (0,1)$. We can separate $\vartheta$ and $\rho$ into two sides of the inequality and see this. The second one always holds for $\rho \geq 0.183$. We can separate $\vartheta$ and $\rho$ into two sides of the inequality and see this.

As a result, when $\rho > 0.197$, $FN_2 = o(1)$.

When $0.183 \leq \rho < 0.197$, $\sqrt{r}$ is always in the second case, and $FN_2 = o(1)$ still holds, because $d^2(C, ((1+\rho)\sqrt{r}, (1+\rho)\sqrt{r})) \geq \frac{(1-\rho^2)^2(1+\rho)}{5+3\rho}r$.

When $\rho < 0.183$, we need to look at the expression of $d^2(C, ((1+\rho)\sqrt{r}, (1+\rho)\sqrt{r}))$.

$$d^2(C, ((1+\rho)\sqrt{r}, (1+\rho)\sqrt{r})) = 2(1-\rho)(1+\rho)^2 r - 2(1+\rho)(3+\rho)\lambda'\sqrt{r} + \frac{5+3\rho}{1-\rho}\lambda'^2 \quad (31)$$

In our case, we want $d^2(C, ((1+\rho)\sqrt{r}, (1+\rho)\sqrt{r})) \geq (1-\rho^2)(1-2\vartheta)$ which is

$$2(1+\rho)\left(1 + \sqrt{\frac{1+\rho}{1-\rho}}\right)^2 - \frac{2(3+\rho)}{1-\rho}\sqrt{\frac{1+\rho}{1-\rho}}\left(1 + \sqrt{\frac{1+\rho}{1-\rho}}\right) + \frac{3\rho+5}{(1-\rho)^3} \geq \frac{1-2\vartheta}{1-\vartheta}.$$

There is no simpler form even if we further break this down.

When $\rho > 0.179$, the LHS is always greater than 1, and not restrictive to $\vartheta$. When $\rho < 0.179$, this imposes a lower bound on $\vartheta$; when $\vartheta$ is very small, there will be another curve above $\sqrt{r} = (1 + \sqrt{\frac{1+\rho}{1-\rho}})\sqrt{1-\vartheta}$.

To sum up, now we have $\sqrt{r} = \max\left\{1 + \sqrt{1-\vartheta}, (1 + \sqrt{\frac{1+\rho}{1-\rho}})\sqrt{1-\vartheta}\right\}$, but when $\rho < 0.179$, we seem to need one more curve which is unknown for now.

When $\sqrt{r} \geq \frac{5+3\rho}{2+2\rho}\lambda'$ in $f_1(\sqrt{r}, \lambda')$, if $\lambda'^2 \leq \frac{(1-\rho^2)(1-\rho)r}{5+3\rho}$, then we need $\lambda^2 = (1-\vartheta)(1-\rho^2)$, which contradicts $\lambda > 1$ (for $FP_1$). So we must have $\lambda'^2 > \frac{(1-\rho^2)(1-\rho)r}{5+3\rho}$, and $\sqrt{r} = \sqrt{\frac{5+3\rho}{1-\rho}}\sqrt{1-\vartheta}$.

However, this curve $\sqrt{r} = \sqrt{\frac{5+3\rho}{1-\rho}}\sqrt{1-\vartheta}$ is always greater than the $\sqrt{r} = (1 + \sqrt{\frac{1+\rho}{1-\rho}})\sqrt{1-\vartheta}$ we have computed. By requiring $\sqrt{r} \geq \frac{5+3\rho}{2+2\rho}\lambda'$, we have $1 < \lambda' \leq \frac{2(1+\rho)\sqrt{1-\vartheta}}{\sqrt{5+3\rho}\sqrt{1-\rho}}$. Notice that $\sqrt{1-\vartheta} \geq \frac{\sqrt{(5+3\rho)(1-\rho)}}{2(1+\rho)}$ is a strictly tighter requirement than $\vartheta < \frac{2\rho}{1+\rho}$. As a result, even if this curve exists, it is not part of the boundary in the phase diagram.

When $2\lambda' < \sqrt{r} < \frac{5+3\rho}{2+2\rho}\lambda'$, $\lambda'^2$ cannot be smaller than $d^2(B, (\rho\sqrt{r}, \sqrt{r}))$; otherwise $\lambda'^2 = (1-\rho^2)(1-\vartheta)$ which contradicts $\lambda' > 1$ in $FP_1$.

Then we need the following things:

$$\begin{cases} 2\lambda' < \sqrt{r} < \frac{5+3\rho}{2+2\rho}\lambda' \\ d^2(B, (\rho\sqrt{r}, \sqrt{r})) = (1-\rho^2)(1-\vartheta) \\ \lambda' \geq 1 \\ 2\vartheta + f_3(\sqrt{r}, \lambda') \geq 1 \\ \sqrt{r} \geq \lambda' + \sqrt{1-\vartheta} \\ \sqrt{r} \geq \sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{\lambda'}{1+\rho} \\ \sqrt{r} \geq \sqrt{\frac{5+3\rho}{(1-\rho)(2+\rho)^2}}\sqrt{1-\vartheta} \end{cases}$$

and one of the last three inequalities must attain equality.

If $\sqrt{r} = \lambda' + \sqrt{1-\vartheta}$, then $\sqrt{r} > 2\lambda'$ would imply $\lambda' < \sqrt{1-\vartheta}$, contradicting $\lambda' > 1$.

If $\sqrt{r} = \sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{\lambda'}{1+\rho}$, we need to look at the expression of $d^2(B, (\rho\sqrt{r}, \sqrt{r}))$ computed in (34). We let $\lambda^* = \frac{1}{\sqrt{1-\vartheta}}\lambda'$, and $\sqrt{r} = \sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{\lambda^*\sqrt{1-\vartheta}}{1+\rho}$. Now we have $\frac{2+4\rho+3\rho^2}{(1+\rho)^2}\lambda^{*2} - \frac{2(1+2\rho)}{(1+\rho)\sqrt{1-\rho^2}}\lambda^* + \frac{\rho^2}{1-\rho^2} = 0$. However,

$$\sqrt{r} > 2\lambda' \implies \lambda^* < \frac{1+\rho}{1+2\rho}\frac{1}{\sqrt{1-\rho^2}}$$

$$\sqrt{r} < \frac{5+3\rho}{2+2\rho}\lambda' \implies \lambda^* > \frac{2}{3}\frac{1}{\sqrt{1-\rho^2}}$$

Plug such lower bound and upper bound into the quadratic equation, and the values are negative at both the upper and lower bounds. Thus we know that it has no solution for $\lambda'$ at all.

If $\sqrt{r} = \sqrt{\frac{5+3\rho}{(1-\rho)(2+\rho)^2}}\sqrt{1-\vartheta}$, then we have the following two requirements:

$$\begin{cases} \sqrt{r} \geq \sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{\lambda'}{1+\rho} \implies \lambda' \leq (1+\rho)\left[\sqrt{\frac{5+3\rho}{(1-\rho)(2+\rho)^2}} - \frac{1}{\sqrt{1-\rho^2}}\right]\sqrt{1-\vartheta} \\ \sqrt{r} \leq \frac{5+3\rho}{2+2\rho}\lambda' \implies \lambda' \geq \frac{2(1+\rho)}{5+3\rho}\sqrt{\frac{5+3\rho}{(1-\rho)(2+\rho)^2}}\sqrt{1-\vartheta} \end{cases}$$

and the upper bound is smaller than the lower bound, contradiction.

Tp sum up the *second* case, we have $\sqrt{r} = \max\left\{1 + \sqrt{1-\vartheta}, (1 + \sqrt{\frac{1+\rho}{1-\rho}})\sqrt{1-\vartheta}\right\}$, but when $\rho < 0.179$, we seem to need one more curve which is unknown for now.

*Third,* if $\lambda'^2 = 2\vartheta + f_3(\sqrt{r}, \lambda') = 1$ and $\vartheta + f_1(\sqrt{r}, \lambda') \geq 1$, $\vartheta + f_2(\sqrt{r}, \lambda') \geq 1$, we will eventually have now curve in this case. We start from some basic requirements:

Now we have $\lambda' = 1$. When $\lambda' = 1$ is fixed, the exponents of $FP$ and $FN$ are all decreasing in $\sqrt{r}$. For $FN_1$, we thus need

$$\sqrt{r} \geq \max\left\{1 + \sqrt{1-\vartheta}, \sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{1}{1+\rho}, \sqrt{\frac{5+3\rho}{(1-\rho)(2+\rho)^2}}\sqrt{1-\vartheta}\right\}$$

Even if we finally find an admissible curve with $\lambda'^2 = 2\vartheta + f_3(\sqrt{r}, \lambda') = 1$, it cannot be lower than $\sqrt{r} = 1 + \sqrt{1 - \vartheta}$, so we can require $\vartheta < \frac{2\rho}{1+\rho}$.

Now we look at $\vartheta + f_1(\sqrt{r}, \lambda') \geq 1$. If $\sqrt{r} \leq 2$, then $\vartheta < \frac{2\rho}{1+\rho}$ gives us a contradiction. Thus $\sqrt{r} \geq 2$. We still need $FPtwo \geq$ additionally to $\sqrt{r} \geq 2$.

Then we discuss the conditional expression of $f_3(\sqrt{r}, \lambda')$ one by one.

When $\sqrt{r} \geq \frac{2\lambda'}{1-\rho^2}$, $\sqrt{r} = \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{1-\rho^2}$ which contradicts $\sqrt{r} \geq \frac{2\lambda'}{1-\rho^2}$.

When $\frac{\lambda'}{1+\rho} \cdot \frac{5+3\rho}{(1-\rho)(3+\rho)} \leq \sqrt{r} < \frac{2\lambda'}{1-\rho^2}$, we have

$$d^2(C, ((1+\rho)\sqrt{r}, (1+\rho)\sqrt{r})) = (1 - \rho^2)(1 - 2\vartheta)$$

$$d^2(C, ((1+\rho)\sqrt{r}, (1+\rho)\sqrt{r})) \geq \frac{(1-\rho^2)^2(1+\rho)}{5+3\rho}r$$

and thus $\sqrt{r} \leq \sqrt{\frac{5+3\rho}{(1+\rho)^2(1-\rho)}}\sqrt{1 - 2\vartheta}$.

For $\vartheta + f_1(\sqrt{r}, \lambda') \geq 1$, we look at the conditional expression of $f_1$, which now can only take one of the last two cases.

- if it is the last case, then we need $\sqrt{r} \geq \sqrt{\frac{5+3\rho}{1-\rho}}\sqrt{1 - \vartheta}$, which does not hold.

- if it is the fourth case, then we need

$$d^2(B, (\rho\sqrt{r}, \sqrt{r})) \geq (1 - \rho^2)(1 - \vartheta).$$

We already have the expression of $d^2(B, (\rho\sqrt{r}, \sqrt{r}))$ in (34), and the expression of $d^2(C, ((1+\rho)\sqrt{r}, (1+\rho)\sqrt{r}))$ in (31):

$$r - 4\sqrt{r} + \frac{5 + 3\rho}{1 + \rho} \geq 1 - \vartheta$$

$$r - \frac{3 + \rho}{1 - \rho^2}\sqrt{r} + \frac{5 + 3\rho}{2(1 - \rho^2)^2} = \frac{1 - 2\vartheta}{2(1 + \rho)}$$

$$\implies \frac{1 - 2\vartheta}{2(1 + \rho)} + \frac{3 + \rho}{1 - \rho^2}\sqrt{r} - \frac{5 + 3\rho}{2(1 - \rho^2)^2} - 4\sqrt{r} + \frac{5 + 3\rho}{1 + \rho} \geq 1 - \vartheta$$

$$\implies \frac{1}{2(1 + \rho)} + \frac{3 + \rho}{1 - \rho^2}\sqrt{r} - \frac{5 + 3\rho}{2(1 - \rho^2)^2} - 4\sqrt{r} + \frac{5 + 3\rho}{1 + \rho} \geq 1 - \frac{\rho}{1 + \rho}\vartheta$$

It turns out that when we regard $\sqrt{r}$ as an independent variable, and let it vary in the interval $(2, 2.5)$ as in $FP_2$, we always have, $\forall \rho \in (0, 1)$,

$$\frac{1}{2(1 + \rho)} + \frac{3 + \rho}{1 - \rho^2}\sqrt{r} - \frac{5 + 3\rho}{2(1 - \rho^2)^2} - 4\sqrt{r} + \frac{5 + 3\rho}{1 + \rho} < 1 - \frac{\rho}{1 + \rho}$$

so we have a contradiction.

To sum up, no curve in this case.

When $\sqrt{r} \leq \frac{\lambda'}{1+\rho} \cdot \frac{5+3\rho}{(1-\rho)(3+\rho)}$, if $\left[(1 - \rho^2)\sqrt{r} - (1 - \rho)\lambda'\right]_+^2 \leq \frac{(1-\rho^2)^2(1+\rho)}{5+3\rho}r$, then we have $\sqrt{r} = \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{1+\rho}$, which contradicts $\sqrt{r} \geq \sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{1}{1+\rho}$ required by $FN_1$. If $\left[(1 - \rho^2)\sqrt{r} - (1 - \rho)\lambda'\right]_+^2 > \frac{(1-\rho^2)^2(1+\rho)}{5+3\rho}r$, then we have $\sqrt{r} = \sqrt{\frac{5+3\rho}{(1+\rho)^2(1-\rho)}}\sqrt{1 - 2\vartheta}$. This is a tedious case.

- Upper bound on $\vartheta$: $\sqrt{r} \geq 2$, which implies $\vartheta \leq \min\left\{\frac{2\rho}{1+\rho}, \frac{1}{2} - \frac{2(1+\rho)^2(1-\rho)}{5+3\rho}\right\}$.

- Lower bound on $\vartheta$: Even if the curve is admissible, it only makes a difference if it is smaller than $(1 + \sqrt{\frac{1+\rho}{1-\rho}})\sqrt{1 - \vartheta}$. This gives us $\sqrt{\frac{1-2\vartheta}{1-\vartheta}} < \left(1 + \sqrt{\frac{1+\rho}{1-\rho}}\right)\sqrt{\frac{(1+\rho)^2(1-\rho)}{5+3\rho}} =: \phi(\rho)$. Also, $\sqrt{r} \leq \frac{5+3\rho}{(2+2\rho)}$ implies $\vartheta \geq \frac{1}{2} - \frac{1}{8}(5 + 3\rho)(1 - \rho)$.

39

We then examine $\vartheta + f_1(\sqrt{r}, \lambda')$.

- When $\sqrt{r} \geq \frac{5+3\rho}{2+2\rho}\lambda' \geq \frac{5+3\rho}{2+2\rho}$, we easily have a contradiction because $\sqrt{r} = \sqrt{\frac{5+3\rho}{(1+\rho)^2(1-\rho)}}\sqrt{1-2\vartheta}$.

- When $\sqrt{r} < \frac{5+3\rho}{2+2\rho}\lambda'$, we need to look at $d^2(B, (\rho\sqrt{r}, \sqrt{r}))$. It now becomes $\sqrt{r} - 2 \geq \sqrt{\frac{2\rho}{1+\rho} - \vartheta}$. The derivative of $(LHS - RHS)$ w.r.t. $\vartheta$ is

$$\frac{1}{2\sqrt{\frac{2\rho}{1+\rho} - \vartheta}}\left[1 - \sqrt{\frac{2(5+3\rho)}{(1+\rho)^2(1-\rho)}}\sqrt{\frac{\frac{2\rho}{1+\rho} - \vartheta}{\frac{1}{2} - \vartheta}}\right]$$

from which we can see $(LHS - RHS)$ is either increasing, decreasing, or first-decreasing-then-increasing. If we evaluate $(LHS - RHS)$ at the smallest and largest $\vartheta$ and they are both negative ($\forall \rho$), then we have a contradiction.

  - When $\rho \geq 0.183$, $\phi(\rho) > 1$. The $(LHS - RHS)$ is below 0 at both $\max\{0, \frac{1}{2} - \frac{1}{8}(1-\rho)(5+3\rho)\}$ and $\min\left\{\frac{2\rho}{1+\rho}, \frac{1}{2} - \frac{2(1+\rho)^2(1-\rho)}{5+3\rho}\right\}$, for all $0.183 \leq \rho \leq 1$.
  - When $\rho < 0.183$, $0 < \phi(\rho) < 1$, and it poses a lower bound on $\vartheta$.
    * When $\rho \leq 0.091$, the lower bound is greater than the upper bound, so this case does not exist for any $\vartheta$.
    * When $\rho \in (0.091, 0.183)$, we can verify that the $(LHS - RHS)$ is below 0 at both $\max\{0, \frac{1}{2} - \frac{1}{8}(1-\rho)(5+3\rho)\}$ and $\min\left\{\frac{2\rho}{1+\rho}, \frac{1}{2} - \frac{2(1+\rho)^2(1-\rho)}{5+3\rho}, \frac{1-\phi^2(\rho)}{2-\phi^2(\rho)}\right\}$.

We finally know that $FP_1 = FN_2$ gives nothing.

*Fourth*, if $\vartheta + f_1(\sqrt{r}, \lambda') = 2\vartheta + f_3(\sqrt{r}, \lambda') = 1$ and $\lambda' \geq 1$, $\vartheta + f_2(\sqrt{r}, \lambda') \geq 1$, we will get the last curve.

From $\vartheta + f_2(\sqrt{r}, \lambda') \geq 1$, we know $\sqrt{r} \geq \lambda'$, so we start from the case $\lambda' \leq \sqrt{r} \leq 2\lambda'$ in $f_1(\sqrt{r}, \lambda')$.

When $\lambda' \leq \sqrt{r} \leq 2\lambda'$ in $f_1(\sqrt{r}, \lambda')$, we have $\lambda' = \sqrt{\frac{1+\rho}{1-\rho}}\sqrt{1-\vartheta}$. If $\sqrt{r} \leq \frac{\lambda'}{1+\rho} \cdot \frac{5+3\rho}{(1-\rho)(3+\rho)}$ in $f_3(\sqrt{r}, \lambda')$, we could have either of the following two,

$$\sqrt{r} = \frac{1}{\sqrt{1-\rho^2}}\left(\sqrt{1-\vartheta} + \sqrt{1-2\vartheta}\right)$$

$$\sqrt{r} = \sqrt{\frac{5+3\rho}{(1+\rho)^2(1-\rho)}}\sqrt{1-2\vartheta}$$

When it is the former, for $\vartheta + f_2(\sqrt{r}, \lambda') \geq 1$, we need $\sqrt{r} \geq (1 + \sqrt{\frac{1+\rho}{1-\rho}})\sqrt{1-\vartheta}$, which implies $\sqrt{\frac{1-2\vartheta}{1-\vartheta}} \geq \rho + \sqrt{1-\rho^2} > 1$, which is impossible.

When it is the latter, for $\vartheta + f_2(\sqrt{r}, \lambda') \geq 1$, we need $\sqrt{r} \geq (1 + \sqrt{\frac{1+\rho}{1-\rho}})\sqrt{1-\vartheta}$, which implies $\sqrt{\frac{1-2\vartheta}{1-\vartheta}} \geq \sqrt{\frac{(1+\rho)^2(1-\rho)}{5+3\rho}}\left(1 + \sqrt{\frac{1+\rho}{1-\rho}}\right)$. Also, $\sqrt{r} \leq \frac{\lambda'}{1+\rho} \cdot \frac{5+3\rho}{(1-\rho)(3+\rho)}$ implies $\sqrt{\frac{1-2\vartheta}{1-\vartheta}} \leq \frac{\sqrt{(5+3\rho)(1+\rho)}}{(1-\rho)(3+\rho)}$. However, now either the lower bound $\sqrt{\frac{(1+\rho)^2(1-\rho)}{5+3\rho}}\left(1 + \sqrt{\frac{1+\rho}{1-\rho}}\right) > 1 \geq \sqrt{\frac{1-2\vartheta}{1-\vartheta}}$, or the upper bound is smaller than the lower bound.

If $\frac{\lambda'}{1+\rho} \cdot \frac{5+3\rho}{(1-\rho)(3+\rho)} \leq \sqrt{r} \leq \frac{2\lambda'}{1-\rho^2}$ in $f_3(\sqrt{r}, \lambda')$, we need to solve a quadratic function of $\sqrt{r}$, i.e. $d^2(C, ((1+\rho)\sqrt{r}, (1+\rho)\sqrt{r})) = (1-\rho^2)(1-2\vartheta)$. The expression of the LHS is already in (31). Then we have

$$\sqrt{r} = \frac{3+\rho}{2(1-\rho^2)}\sqrt{\frac{1+\rho}{1-\rho}}\sqrt{1-\vartheta} + \frac{1}{2}\sqrt{\frac{2(1-2\vartheta)}{1+\rho} - \frac{(1-\vartheta)}{(1-\rho)^2}}. \tag{32}$$

We take the larger root, because when $\sqrt{r}$ takes the smaller one, the ellipsoid is actually still tangent to the green line segment in Figure 9. Thus the smaller root should be discarded.

Of course, we also list all the requirements it must meet. They are loose only except the first one.

$$
\begin{cases}
\sqrt{r} \geq (1 + \sqrt{\frac{1+\rho}{1-\rho}})\sqrt{1 - \vartheta} \\
\vartheta \leq \frac{2\rho}{1+\rho} \\
\frac{5+3\rho}{(1-\rho^2)(3+\rho)}\sqrt{\frac{1+\rho}{1-\rho}}\sqrt{1 - \vartheta} \leq \sqrt{r} \leq 2\sqrt{\frac{1+\rho}{1-\rho}}\sqrt{1 - \vartheta}
\end{cases}
$$

When $\sqrt{r} \geq \frac{5+3\rho}{2+2\rho}\lambda'$ in $f_1(\sqrt{r}, \lambda')$, We know from $FP_2$ that $\sqrt{r} = \sqrt{\frac{5+3\rho}{1-\rho}}\sqrt{1 - \vartheta}$. ( Because we need $\lambda' \geq 1$ for $FP_1$, the other term is not possible.) If $\sqrt{r} \leq \frac{\lambda'}{1+\rho} \cdot \frac{5+3\rho}{(1-\rho)(3+\rho)}$ in $f_3(\sqrt{r}, \lambda')$, it is only possible that $\sqrt{r} = \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{\lambda'}{1+\rho}$. However, $FN_1 = o(1)$ requires that $\sqrt{r} \geq \sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{\lambda'}{1+\rho}$, so this case is not possible. If $\sqrt{r} \geq \frac{2\lambda'}{1-\rho^2}$ in $f_3(\sqrt{r}, \lambda')$, we have $\sqrt{r} = \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{\lambda'}{1-\rho^2}$. Because $\sqrt{r} \geq 2\left(\sqrt{r} - \sqrt{\frac{1-2\vartheta}{1-\rho^2}}\right)$, we have $\sqrt{r} \leq 2\sqrt{\frac{1-2\vartheta}{1-\rho^2}} \implies \lambda' \leq \sqrt{(1 - 2\vartheta)(1 - \rho^2)}$, which contradicts $\lambda' > 1$. If $\frac{\lambda'}{1+\rho} \cdot \frac{5+3\rho}{(1-\rho)(3+\rho)} < \sqrt{r} < \frac{2\lambda'}{1-\rho^2}$ in $f_3(\sqrt{r}, \lambda')$, we know that

$$
(1 - \rho^2)(1 - 2\vartheta) = d^2(C, ((1 + \rho)\sqrt{r}, (1 + \rho)\sqrt{r})) \geq \frac{(1 - \rho^2)^2(1 + \rho)}{5 + 3\rho}r
$$

and thus $\sqrt{r} \leq \sqrt{\frac{5+3\rho}{(1-\rho)(1+\rho)^2}}\sqrt{1 - 2\vartheta}$ which contradicts $\sqrt{r} = \sqrt{\frac{5+3\rho}{1-\rho}}\sqrt{1 - \vartheta}$.

When $\sqrt{r} \in (2\lambda', \frac{5+3\rho}{2+2\rho}\lambda')$ in $f_1(\sqrt{r}, \lambda')$:

If $\sqrt{r} \leq \frac{\lambda'}{1+\rho} \cdot \frac{5+3\rho}{(1-\rho)(3+\rho)}$ in $f_3(\sqrt{r}, \lambda')$, and $\sqrt{r} = \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{\lambda'}{1+\rho}$, we have a contradiction. This is because $FN_1 = o(1)$ requires that $\sqrt{r} \geq \sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{\lambda'}{1+\rho}$, so this case is not possible.

If $\sqrt{r} \leq \frac{\lambda'}{1+\rho} \cdot \frac{5+3\rho}{(1-\rho)(3+\rho)}$ in $f_3(\sqrt{r}, \lambda')$, and $\sqrt{r} = \sqrt{\frac{5+3\rho}{(1+\rho)^2(1-\rho)}}\sqrt{1 - 2\vartheta}$, this turns out a very tedious case because we generally need to work with $(\vartheta, \lambda', \rho)$ at the same time. For completeness, we include a rigorous proof anyway.

Briefly speaking, we let $\lambda^* = \frac{1}{\sqrt{1-\vartheta}}\lambda'$. Because all we need is

$$
\begin{cases}
d^2(B, (\rho\sqrt{r}, \sqrt{r})) = (1 - \rho^2)(1 - \vartheta) & (\vartheta + f_1(\sqrt{r}, \lambda') \geq 1) \\
\sqrt{r} = \sqrt{\frac{5+3\rho}{(1+\rho)^2(1-\rho)}}\sqrt{1 - 2\vartheta} & (2\vartheta + f_3(\sqrt{r}, \lambda') \geq 1) \\
\lambda' \geq 1 \\
\sqrt{r} \geq \lambda' + \sqrt{1 - \vartheta} & (\vartheta + f_2(\sqrt{r}, \lambda') \geq 1) \\
\sqrt{r} \geq \sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{\lambda'}{1+\rho} & (\vartheta + f_2(\sqrt{r}, \lambda') \geq 1) \\
\sqrt{r} \geq \sqrt{\frac{5+3\rho}{(1-\rho)(2+\rho)^2}}\sqrt{1 - \vartheta} & (\vartheta + f_2(\sqrt{r}, \lambda') \geq 1) \\
2\lambda' \leq \sqrt{r} \leq \frac{5+3\rho}{2+2\rho}\lambda' & (\vartheta + f_1(\sqrt{r}, \lambda') \geq 1) \\
\sqrt{r} \leq \frac{5+3\rho}{(1-\rho^2)(3+\rho)}\lambda' & (2\vartheta + f_3(\sqrt{r}, \lambda') \geq 1)
\end{cases}
$$

by cleaning this up, we have, letting $x = \sqrt{\frac{1-2\vartheta}{1-\vartheta}}$,

$$
\begin{cases}
q(x) \overset{\text{def}}{=} \frac{5+3\rho}{(1+\rho)^2(1-\rho)}x^2 - 4\sqrt{\frac{5+3\rho}{(1+\rho)^2(1-\rho)}}\lambda^* x + \left(\frac{5+3\rho}{1+\rho}\right)(\lambda^*)^2 - 1 = 0 \\
\lambda^* \geq 1 \\
x \geq \max\left\{\frac{(1+\rho)^2(1-\rho)}{5+3\rho}\left(\frac{\lambda^*}{1+\rho} + \frac{1}{\sqrt{1-\rho^2}}\right), \frac{1+\rho}{2+\rho}, 2\lambda^*\sqrt{\frac{(1+\rho)^2(1-\rho)}{5+3\rho}}\right\} \\
x \leq \min\left\{1, \frac{\lambda^*}{2}\sqrt{(5 + 3\rho)(1 - \rho)}, \frac{\lambda^*}{3+\rho}\sqrt{\frac{5+3\rho}{1-\rho}}\right\}
\end{cases}
\tag{33}
$$

If we want some admissible $x$ to exist, we need the upper bound to be greater than the lower bound in the last two inequalities. This will give us $\underline{B}(\rho) \leq \lambda^* \leq \overline{B}(\rho)$. (The expressions of $\underline{B}(\rho)$ and $\overline{B}(\rho)$ can be explicitly written, but are omitted for brevity.)

- For some $\rho \in (0, 1)$ and $\lambda^* \in [\underline{B}(\rho), \overline{B}(\rho)]$, we try to plot $q(x)$ at a suitable $x$. If $q(x) < 0$ always holds, then $x$ has no solution, and this case is eliminated.

- $q(x)$ is a quadratic function of $x$. Its axis of symmetry is $2\lambda^* \sqrt{\frac{(1+\rho)^2(1-\rho)}{5+3\rho}}$, which we already know is smaller than $x$. Thus $q(x)$ is increasing in admissible $x$, and we only need to evaluate $q(x)$ at the maximum $x$: $x = \min\left\{1, \frac{\lambda^*}{2}\sqrt{(5+3\rho)(1-\rho)}, \frac{\lambda^*}{3+\rho}\sqrt{\frac{5+3\rho}{1-\rho}}\right\}$.

- Now we are left to prove a bivariate function is below zero, which can be easily verified given all the requirements on $\lambda'$ and $\forall\, \rho > 0$.

When $\sqrt{r} \geq \frac{2\lambda'}{1-\rho^2}$ in $f_3(\sqrt{r}, \lambda')$, we have $\sqrt{r} = \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{\lambda'}{1-\rho^2}$. Because $\sqrt{r} \geq 2\left(\sqrt{r} - \sqrt{\frac{1-2\vartheta}{1-\rho^2}}\right)$, we have $\sqrt{r} \leq 2\sqrt{\frac{1-2\vartheta}{1-\rho^2}} \implies \lambda' \leq \sqrt{(1-2\vartheta)(1-\rho^2)}$, which contradicts $\lambda' > 1$.

When $\frac{\lambda'}{1+\rho} \cdot \frac{5+3\rho}{(1-\rho)(3+\rho)} < \sqrt{r} < \frac{2\lambda'}{1-\rho^2}$ in $f_3(\sqrt{r}, \lambda')$, we have

$$d^2(C, ((1+\rho)\sqrt{r}, (1+\rho)\sqrt{r})) = (1-\rho^2)(1-2\vartheta)$$
$$d^2(B, (\rho\sqrt{r}, \sqrt{r})) = (1-\rho^2)(1-\vartheta).$$

This is an even more tedious case, and we eliminate this case as follows: We first list all of the requirements we have:

$$\begin{cases} d^2(B, (\rho\sqrt{r}, \sqrt{r})) = (1-\rho^2)(1-\vartheta) & (\vartheta + f_1(\sqrt{r}, \lambda') \geq 1) \\ d^2(C, ((1+\rho)\sqrt{r}, (1+\rho)\sqrt{r})) = (1-\rho^2)(1-2\vartheta) & (2\vartheta + f_3(\sqrt{r}, \lambda') \geq 1) \\ \lambda' \geq 1 & \\ 2\lambda' \leq \sqrt{r} \leq \frac{5+3\rho}{2+2\rho}\lambda' & (\vartheta + f_1(\sqrt{r}, \lambda') \geq 1) \\ \sqrt{r} \geq \max\left\{\sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{\lambda'}{1+\rho}, \sqrt{\frac{5+3\rho}{(1-\rho)(2+\rho)^2}}\sqrt{1-\vartheta}\right\} & (\vartheta + f_2(\sqrt{r}, \lambda') \geq 1) \\ \sqrt{r} \geq \frac{5+3\rho}{(1-\rho^2)(3+\rho)}\lambda' & (\vartheta + f_1(\sqrt{r}, \lambda') \geq 1) \end{cases}$$

Define $x = \sqrt{\frac{r}{1-\vartheta}}$ and $\lambda^* = \frac{1}{\sqrt{1-\vartheta}}\lambda'$. We know the upper and lower bounds of $x$, from the last three inequalities:

$$\max\left\{\frac{5+3\rho}{(1-\rho^2)(3+\rho)}\lambda^*, 2\lambda^*, \frac{1}{\sqrt{1-\rho^2}} + \frac{\lambda^*}{1+\rho}, \sqrt{\frac{5+3\rho}{(1-\rho)(2+\rho)^2}}\right\} \leq x \leq \frac{5+3\rho}{2+2\rho}\lambda^*.$$

For admissible $x$ to exist, we need $\rho \leq 0.415$ and $\lambda^* \geq \max\left\{\frac{2}{3\sqrt{1-\rho^2}}, \frac{2(1+\rho)}{(2+\rho)\sqrt{(5+3\rho)(1-\rho)}}\right\}$.

Then we want to know the upper and lower bounds of $\lambda^*$ given $\rho \leq 0.415$. In terms of the upper bound, we know from $d^2(B, (\rho\sqrt{r}, \sqrt{r})) = (1-\rho^2)(1-\vartheta) \geq (1-\rho)^2\lambda'^2$ that $\lambda^* \leq \sqrt{\frac{1+\rho}{1-\rho}}$. To sum up,

$$\max\left\{1, \frac{2}{3\sqrt{1-\rho^2}}, \frac{2(1+\rho)}{(2+\rho)\sqrt{(5+3\rho)(1-\rho)}}\right\} \leq \lambda^* \leq \sqrt{\frac{1+\rho}{1-\rho}}.$$

Having the upper and lower bound on $\lambda'$, we look at the first two quadratic functions, and we know that

$$\begin{cases} x^2 - 4x\lambda^* + \frac{5+3\rho}{1+\rho}\lambda^{*2} = 1 \\ 2(1+\rho)x^2 - \frac{2(3+\rho)}{1-\rho}x\lambda^* + \frac{5+3\rho}{(1-\rho)^2(1+\rho)}\lambda^{*2} = \frac{1-2\vartheta}{1-\vartheta} \leq 1 \end{cases}$$

The LHS of the second quadratic equation, as a function of $x$, has the axis of symmetry at $\frac{3+\rho}{2(1-\rho^2)}\lambda^* > \frac{5+3\rho}{2+2\rho}\lambda^*$, so it is decreasing in $x$ (fixing $\lambda^*$).

We look closer at $x^2 - 4x\lambda^* + \frac{5+3\rho}{1+\rho}\lambda^{*2} = 1$, which gives $x = 2\lambda^* + \sqrt{1 - \frac{1-\rho}{1+\rho}\lambda^{*2}} \leq 2\lambda^* + \sqrt{1 - \frac{1-\rho}{1+\rho}\lambda^*}$.

Let $y = \min\left\{\frac{5+3\rho}{2+2\rho}\lambda^*, 2\lambda^* + \sqrt{1 - \frac{1-\rho}{1+\rho}\lambda^*}\right\}$ be an upper bound on $x$. Both terms in $y$ are increasing in $\lambda^* \leq \sqrt{\frac{1+\rho}{1-\rho}}$, while $x$ has no such monoticity.

$$2(1+\rho)x^2 - \frac{2(3+\rho)}{1-\rho}x\lambda^* + \frac{5+3\rho}{(1-\rho)^2(1+\rho)}\lambda^{*2} \geq 2(1+\rho)y^2 - \frac{2(3+\rho)}{1-\rho}y\lambda^* + \frac{5+3\rho}{(1-\rho)^2(1+\rho)}\lambda^{*2}$$

The RHS of the above line, viewed as a function of $\lambda^*$ and fixing $y$, is also decreasing in $\lambda^*$, because the axis of symmetry

$$\frac{(1-\rho^2)(3+\rho)}{5+3\rho}y = \frac{(1-\rho^2)(3+\rho)}{5+3\rho}\min\left\{\frac{5+3\rho}{2+2\rho}\lambda^*, 2\lambda^* + \sqrt{1 - \frac{1-\rho}{1+\rho}\lambda^*}\right\} \geq \lambda^*.$$

(The above line can be proven for $\lambda^* < \sqrt{\frac{1+\rho}{1-\rho}}$ and $\rho < 0.415$ as we have required.)

Thus, as a whole, $\frac{\mathrm{d}RHS(y(\lambda^*),\lambda^*)}{\mathrm{d}\lambda^*} = \frac{\partial RHS(y,\lambda^*)}{\partial y} \cdot \frac{\partial y}{\partial \lambda^*} + \frac{\partial RHS(y,\lambda^*)}{\partial \lambda^*} \leq 0$. So the $RHS$ is decreasing in $\lambda^*$.

When we let $\lambda^* = \sqrt{\frac{1+\rho}{1-\rho}}$ which is the maximum, the $RHS$ is a univariate function of $\rho \in (0, 0.415)$, which is always greater than 1. It cannot be equal to $\frac{1-2\vartheta}{1-\vartheta}$, and now we have a contradiction.

**We then look at the case of $\rho < 0$.** Before diving into the proof, we first re-iterate the phase curves in Theorem 3 in an equivalent way. As the proof is tedious, the simplified form of the diagram in Theorem 3 may not be recognizable, so we describes the diagram in an equivalent way in Theorem 12 again, making it more consistent with what we will see in the proof.

To ease the notation, we recall in Theorem 3 we defined

$$\sqrt{h_6(\vartheta)} = \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{\frac{1-2|\rho|}{1-|\rho|}\sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \sqrt{\left[\left(\frac{1-2|\rho|}{1-|\rho|}\right)^2 + \frac{1-|\rho|}{1+|\rho|}\right](1-\vartheta) - \frac{1-2\vartheta}{(1+|\rho|)^2}}}{(1-|\rho|)\left[\left(\frac{1-2|\rho|}{1-|\rho|}\right)^2 + \frac{1-|\rho|}{1+|\rho|}\right]}$$

**Theorem 12** (Re-iterating Theorem 3 in an equivalent way for $\rho < 0$). *For $\rho < 0$ and $a < \frac{2}{1-|\rho|}$, the half of the phase diagram of SCAD when $\vartheta \in [\frac{1}{2}, 1)$ is the same as that of Lasso and SCAD for positive $\rho$. When $\vartheta < \frac{1}{2}$, from left to right:*

*When $|\rho| \geq 0.535$ (approximately), $\sqrt{r} = \max\left\{\sqrt{\frac{5+3|\rho|}{1-|\rho|}}\sqrt{1-\vartheta}, \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{1-|\rho|}\right\}$.*

*When $\frac{1}{2} \leq |\rho| < 0.535$,*

$$\sqrt{r} = \begin{cases} \sqrt{\frac{5+3|\rho|}{1-|\rho|}}\sqrt{1-\vartheta} & if\sqrt{\frac{1-2\vartheta}{1-\vartheta}} \geq \left(1 - \frac{2(1+|\rho|)}{(1-|\rho|)(5+3|\rho|)}\right)\sqrt{(5+3|\rho|)(1+|\rho|)} \\ \sqrt{h_6(\vartheta)} & if\sqrt{\frac{1-2\vartheta}{1-\vartheta}} < \left(1 - \frac{2(1+|\rho|)}{(1-|\rho|)(5+3|\rho|)}\right)\sqrt{(5+3|\rho|)(1+|\rho|)} \end{cases}$$

*When $0.3965 \leq |\rho| < \frac{1}{2}$,*

$$\sqrt{r} = \begin{cases} \sqrt{\frac{5+3|\rho|}{1-|\rho|}}\sqrt{1-\vartheta} & if\sqrt{\frac{1-2\vartheta}{1-\vartheta}} \geq \left(1 - \frac{2(1+|\rho|)}{(1-|\rho|)(5+3|\rho|)}\right)\sqrt{(5+3|\rho|)(1+|\rho|)} \\ \sqrt{h_6(\vartheta)} & if\sqrt{\frac{1-2\vartheta}{1-\vartheta}} < \left(1 - \frac{2(1+|\rho|)}{(1-|\rho|)(5+3|\rho|)}\right)\sqrt{(5+3|\rho|)(1+|\rho|)} \ and\sqrt{\frac{1-2\vartheta}{1-\vartheta}} > \frac{(1+|\rho|)(1-2|\rho|)}{1-|\rho|} \\ \sqrt{\frac{1-2\vartheta}{1-|\rho|^2}} + \sqrt{\frac{1+|\rho|}{1-|\rho|}}\frac{\sqrt{1-\vartheta}}{1-|\rho|} & if\sqrt{\frac{1-2\vartheta}{1-\vartheta}} \leq \frac{(1+|\rho|)(1-2|\rho|)}{1-|\rho|} \end{cases}$$

When $\frac{1}{3} \leq |\rho| < 0.3965$,

$$\sqrt{r} = \begin{cases} \sqrt{\frac{5+3|\rho|}{1-|\rho|}}\sqrt{1-\vartheta} & if \sqrt{\frac{1-2\vartheta}{1-\vartheta}} \geq \left(1 - \frac{2(1+|\rho|)}{(1-|\rho|)(5+3|\rho|)}\right)\sqrt{(5+3|\rho|)(1+|\rho|)} \\ \sqrt{h_6(\vartheta)} & if \sqrt{\frac{1-2\vartheta}{1-\vartheta}} < \left(1 - \frac{2(1+|\rho|)}{(1-|\rho|)(5+3|\rho|)}\right)\sqrt{(5+3|\rho|)(1+|\rho|)} \, and \sqrt{\frac{1-2\vartheta}{1-\vartheta}} > \frac{(1+|\rho|)(1-2|\rho|)}{1-|\rho|} \\ \sqrt{\frac{1-2\vartheta}{1-|\rho|^2}} + \sqrt{\frac{1+|\rho|}{1-|\rho|}}\sqrt{\frac{1-\vartheta}{1-|\rho|}} & if \sqrt{\frac{1-2\vartheta}{1-\vartheta}} \leq \frac{(1+|\rho|)(1-2|\rho|)}{1-|\rho|} \, and \sqrt{\frac{1-2\vartheta}{1-\vartheta}} \geq \sqrt{1-|\rho|^2} - \frac{|\rho|(1+|\rho|)}{(1-|\rho|)} \\ \left(1 + \sqrt{\frac{1+|\rho|}{1-|\rho|}}\right)\sqrt{1-\vartheta} & if \sqrt{\frac{1-2\vartheta}{1-\vartheta}} < \sqrt{1-|\rho|^2} - \frac{|\rho|(1+|\rho|)}{(1-|\rho|)} \end{cases}$$

When $0.311 \leq |\rho| < \frac{1}{3}$,

$$\sqrt{r} = \begin{cases} \sqrt{\frac{5+3|\rho|}{1-|\rho|}}\sqrt{1-\vartheta} & if \sqrt{\frac{1-2\vartheta}{1-\vartheta}} \geq \left(1 - \frac{2(1+|\rho|)}{(1-|\rho|)(5+3|\rho|)}\right)\sqrt{(5+3|\rho|)(1+|\rho|)} \\ \sqrt{h_6(\vartheta)} & if \sqrt{\frac{1-2\vartheta}{1-\vartheta}} < \left(1 - \frac{2(1+|\rho|)}{(1-|\rho|)(5+3|\rho|)}\right)\sqrt{(5+3|\rho|)(1+|\rho|)} \, and \sqrt{\frac{1-2\vartheta}{1-\vartheta}} > \frac{(1+|\rho|)(1-2|\rho|)}{1-|\rho|} \\ \sqrt{\frac{1-2\vartheta}{1-|\rho|^2}} + \sqrt{\frac{1+|\rho|}{1-|\rho|}}\sqrt{\frac{1-\vartheta}{1-|\rho|}} & if \sqrt{\frac{1-2\vartheta}{1-\vartheta}} \leq \frac{(1+|\rho|)(1-2|\rho|)}{1-|\rho|} \, and \sqrt{\frac{1-2\vartheta}{1-\vartheta}} \geq \sqrt{1-|\rho|^2} - \frac{|\rho|(1+|\rho|)}{(1-|\rho|)} \\ \max\{\left(1 + \sqrt{\frac{1+|\rho|}{1-|\rho|}}\right)\sqrt{1-\vartheta}, \quad 1 + \sqrt{1-\vartheta}\} \, if \sqrt{\frac{1-2\vartheta}{1-\vartheta}} < \sqrt{1-|\rho|^2} - \frac{|\rho|(1+|\rho|)}{(1-|\rho|)} \end{cases}$$

When $0.28832 \leq |\rho| < 0.311$,

$$\sqrt{r} = \begin{cases} \sqrt{h_6(\vartheta)} & if \sqrt{\frac{1-2\vartheta}{1-\vartheta}} > \frac{(1+|\rho|)(1-2|\rho|)}{1-|\rho|} \\ \sqrt{\frac{1-2\vartheta}{1-|\rho|^2}} + \sqrt{\frac{1+|\rho|}{1-|\rho|}}\sqrt{\frac{1-\vartheta}{1-|\rho|}} & if \sqrt{\frac{1-2\vartheta}{1-\vartheta}} \leq \frac{(1+|\rho|)(1-2|\rho|)}{1-|\rho|} \, and \sqrt{\frac{1-2\vartheta}{1-\vartheta}} \geq \sqrt{1-|\rho|^2} - \frac{|\rho|(1+|\rho|)}{(1-|\rho|)} \\ \max\{\left(1 + \sqrt{\frac{1+|\rho|}{1-|\rho|}}\right)\sqrt{1-\vartheta}, \quad 1 + \sqrt{1-\vartheta}\} \, if \sqrt{\frac{1-2\vartheta}{1-\vartheta}} < \sqrt{1-|\rho|^2} - \frac{|\rho|(1+|\rho|)}{(1-|\rho|)} \end{cases}$$

When $0 \leq |\rho| < 0.28832$,

$$\sqrt{r} = \begin{cases} \sqrt{h_6(\vartheta)} & if \sqrt{\frac{1-2\vartheta}{1-\vartheta}} > \frac{(1+|\rho|)(1-2|\rho|)}{1-|\rho|} \\ \sqrt{\frac{1-2\vartheta}{1-|\rho|^2}} + \sqrt{\frac{1+|\rho|}{1-|\rho|}}\sqrt{\frac{1-\vartheta}{1-|\rho|}} & if \sqrt{\frac{1-2\vartheta}{1-\vartheta}} \leq \frac{(1+|\rho|)(1-2|\rho|)}{1-|\rho|} \, and \sqrt{\frac{1-2\vartheta}{1-\vartheta}} \geq \sqrt{1-|\rho|^2} - \frac{|\rho|(1+|\rho|)}{(1-|\rho|)} \\ \max\{\sqrt{\frac{1-2\vartheta}{1-|\rho|^2}} + \frac{1}{1-|\rho|}, \quad 1 + \sqrt{1-\vartheta}\} \, if \sqrt{\frac{1-2\vartheta}{1-\vartheta}} < \sqrt{1-|\rho|^2} - \frac{|\rho|(1+|\rho|)}{(1-|\rho|)} \end{cases}$$

Wo note two things additionally: First, the exact computation of the numerical results (e.g. "0.28832") are covered in the rest of the proof. Second,

$$\left(1 - \frac{2(1+|\rho|)}{(1-|\rho|)(5+3|\rho|)}\right)\sqrt{(5+3|\rho|)(1+|\rho|)} \equiv \frac{3 - 4|\rho| - 3\rho^2}{(1-|\rho|)}\sqrt{\frac{1+|\rho|}{5+3|\rho|}}$$

always holds, and the shorter RHS is shown in Theorem 3.

Before proving Theorem 12 (and Theorem 3 at the same time). we put two important results here.

- In $f_1(\sqrt{r}, \lambda')$ (recall Theorem 11), when $2\lambda' < \sqrt{r} < \frac{5+3|\rho|}{2+2|\rho|}\lambda'$,

$$d^2(B, (|\rho|\sqrt{r}, \sqrt{r})) = (1 - |\rho|^2)\left[r - 4\sqrt{r}\lambda' + \frac{5+3|\rho|}{1+|\rho|}\lambda'^2\right]. \tag{34}$$

- In $f_3(\sqrt{r}, \lambda')$, when $\sqrt{r} \geq \frac{2\lambda'}{1-|\rho|}$,

$$d^2\left(D, ((1-|\rho|)\sqrt{r}, -(1-|\rho|)\sqrt{r})\right) = (1 - |\rho|^2)\left[2(1-|\rho|)r - 2(3-|\rho|)\lambda'\sqrt{r} + \frac{5-3|\rho|}{1-|\rho|}\lambda'^2\right] \tag{35}$$

Then we move on to the proof, which has four sections: $FP_1 = FN_1$, $FP_2 = FN_1$, $FP_1 = FN_2$, $FP_2 = FN_2$.

*First*, if $\lambda'^2 = \vartheta + f_2(\sqrt{r}, \lambda') = 1$ and $\vartheta + f_1(\sqrt{r}, \lambda') \geq 1$, $2\vartheta + f_3(\sqrt{r}, \lambda') \geq 1$, we will prove that $\sqrt{r} = 1 + \sqrt{1-\vartheta}$ is part of the diagram with the condition

$$\begin{cases} \vartheta \geq \frac{2|\rho|}{1+|\rho|} \\ 1 + \sqrt{1-\vartheta} \geq \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{1-|\rho|} \end{cases}$$

When $\vartheta > \frac{1}{2}$, the second condition can be ignored. Even when $\vartheta \leq \frac{1}{2}$, the second condition is restrictive (stronger than the first one) only when $|\rho| < 0.28832$ approximately. ($LHS - RHS$ is increasing in $\vartheta$.)

We know $\lambda' = 1$ and $\sqrt{r} > 1$. From the previous discussion in the case of positive correlation, we already know that only one curve is possible, which is

$$\sqrt{r} = 1 + \sqrt{1-\vartheta}$$

We still need it to meet the following requirements:

$$\begin{cases} 1 + \sqrt{1-\vartheta} \geq \frac{1}{1+|\rho|} + \sqrt{\frac{1-\vartheta}{1-\rho^2}} & \text{for it to be the smallest among the three in } f_2 \\ 1 + \sqrt{1-\vartheta} \geq \sqrt{1-\vartheta}\sqrt{\frac{5+3|\rho|}{(1-|\rho|)(2+|\rho|)^2}} & \text{for it to be the smallest among the three in } f_2 \\ 1 + \sqrt{1-\vartheta} \leq \frac{5+3|\rho|}{(1-|\rho|)(1+|\rho|)^2} & \text{pre-condition in } f_2; \text{ not restrictive} \\ \vartheta > \frac{2|\rho|}{1+|\rho|} & \text{for } \vartheta + f_1(\sqrt{r}, \lambda') \geq 1 \\ 2\vartheta + f_3(\sqrt{r}, \lambda') \geq 1 \end{cases}$$

Among the first 4 requiremenrs, the fourth one can imply the rest. Finally, we look at $2\vartheta + f_3(\sqrt{r}, \lambda') \geq 1$.

When $\vartheta \geq \frac{1}{2}$, we naturally have $2\vartheta + f_3(\sqrt{r}, \lambda') \geq 1$. When $|\rho| \geq \frac{1}{3}$, $\vartheta > \frac{2|\rho|}{1+|\rho|} \geq \frac{1}{2}$ always holds, and no more discussion is needed. When $|\rho| < \frac{1}{3}$, we proceed to the following discussion.

For $|\rho| < \frac{1}{3}$, since $\sqrt{r} = 1 + \sqrt{1-\vartheta} \leq 2 < \frac{2\lambda'}{1-|\rho|}$, we need $1 + \sqrt{1-\vartheta} \geq \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{1-|\rho|}$.

*Second*, if $\vartheta + f_1(\sqrt{r}, \lambda') = \vartheta + f_2(\sqrt{r}, \lambda') = 1$ and $\lambda' \geq 1$, $2\vartheta + f_3(\sqrt{r}, \lambda') \geq 1$, we will prove that we can have one curve at most, $\sqrt{r} = \left(1 + \sqrt{\frac{1+|\rho|}{1-|\rho|}}\right)\sqrt{1-\vartheta}$, and it exists in the interval: (Define $\psi(|\rho|) \stackrel{\text{def}}{=} \sqrt{1-\rho^2}\left(1 - \frac{|\rho|}{1-|\rho|}\sqrt{\frac{1+|\rho|}{1-|\rho|}}\right)$.)

- When $0.28832 \leq |\rho| \leq 0.3965$, the curve exists in the interval $\left[\frac{1-\psi(|\rho|)^2}{2-\psi(|\rho|)^2}, \frac{2|\rho|}{1+|\rho|}\right)$.
- When $|\rho| < 0.28832$, the curve does not exist.
- When $|\rho| > 0.3965$, the curve exists in the interval $[\frac{1}{2}, \frac{2|\rho|}{1+|\rho|})$.

We discuss the conditional expression of $f_1(\sqrt{r}, \lambda')$ to prove such result:

When $\lambda' \leq \sqrt{r} \leq 2\lambda'$ in $f_1(\sqrt{r}, \lambda')$, we have $\lambda' = \sqrt{\frac{1+|\rho|}{1-|\rho|}}\sqrt{1-\vartheta}$. Because we want $\lambda' > 1$, it implies $\vartheta < \frac{2|\rho|}{1+|\rho|}$.

We have discussed and eliminated several curves in the case of positive correlation in the case of positive correlation, so now we are only left with one curve: $\sqrt{r} = \left(1 + \sqrt{\frac{1+|\rho|}{1-|\rho|}}\right)\sqrt{1-\vartheta}$ and we only need to additionally verify $2\vartheta + f_3(\sqrt{r}, \lambda') \geq 1$.

Since $\vartheta < \frac{2|\rho|}{1+|\rho|}$, when $|\rho| > \frac{1}{3}$, $FN_2 = o(1)$ naturally holds for $\vartheta \in [\frac{1}{2}, \frac{2|\rho|}{1+|\rho|})$. We only need to discuss $\vartheta \in (0, \frac{1}{2})$; when $|\rho| \leq \frac{1}{3}$, we need to verify $FN_2 = o(1)$ for all $\vartheta \leq [0, \frac{2|\rho|}{1+|\rho|})$.

Now it can be verified that $\sqrt{r} \leq \frac{2}{1-|\rho|}\lambda'$ always holds, so we need

$$\left(1 + \sqrt{\frac{1+|\rho|}{1-|\rho|}}\right)\sqrt{1-\vartheta} \geq \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{1-|\rho|}\sqrt{\frac{1+|\rho|}{1-|\rho|}}\sqrt{1-\vartheta}$$

which implies

$$\psi(|\rho|) = \sqrt{1 - \rho^2} \left( 1 - \frac{|\rho|}{1 - |\rho|} \sqrt{\frac{1 + |\rho|}{1 - |\rho|}} \right) \geq \sqrt{\frac{1 - 2\vartheta}{1 - \vartheta}}. \tag{36}$$

By taking a close look at the function of $|\rho|$ on the LHS:

- When $\frac{1}{3} < |\rho| < 0.3965$ (approximately), $\psi(|\rho|)$ is positive and smaller than 1. If we define $\psi(|\rho|) \stackrel{\text{def}}{=} \sqrt{1 - \rho^2} \left( 1 - \frac{|\rho|}{1 - |\rho|} \sqrt{\frac{1 + |\rho|}{1 - |\rho|}} \right)$, then the curve $\sqrt{r} = \left( 1 + \sqrt{\frac{1 + |\rho|}{1 - |\rho|}} \right) \sqrt{1 - \vartheta}$ exists in the interval $\left[ \frac{1 - \psi(|\rho|)^2}{2 - \psi(|\rho|)^2}, \frac{2|\rho|}{1 + |\rho|} \right)$.

- When $|\rho| \geq 0.3965$, $\psi(|\rho|)$ is negative. The curve $\sqrt{r} = \left( 1 + \sqrt{\frac{1 + |\rho|}{1 - |\rho|}} \right) \sqrt{1 - \vartheta}$ exists only in the interval $[\frac{1}{2}, \frac{2|\rho|}{1 + |\rho|})$.

- When $0.28832 \leq |\rho| \leq \frac{1}{3}$, the curve $\sqrt{r} = \left( 1 + \sqrt{\frac{1 + |\rho|}{1 - |\rho|}} \right) \sqrt{1 - \vartheta}$ exists in the interval $\left[ \frac{1 - \psi(|\rho|)^2}{2 - \psi(|\rho|)^2}, \frac{2|\rho|}{1 + |\rho|} \right)$.

- When $|\rho| < 0.28832$, this curve does not exist at all, because on the RHS of the inequality (36), the smallest value it can take is $\sqrt{\frac{1 - 2(2|\rho|)/(1 + |\rho|)}{1 - (2|\rho|)/(1 + |\rho|)}}$. When $|\rho| < 0.28832$, even this smallest value is greater than the LHS, so the inequality cannot hold.

When $\sqrt{r} \geq \frac{5 + 3|\rho|}{2 + 2|\rho|} \lambda'$ in $FP_2$, or when $2\lambda' < \sqrt{r} < \frac{5 + 3|\rho|}{2 + 2|\rho|} \lambda'$, the same proof for positive correlation can be used to eliminate these cases.

To sum up the whole case of $\vartheta + f_1(\sqrt{r}, \lambda') = \vartheta + f_2(\sqrt{r}, \lambda') = 1$: We can have one curve at most, $\sqrt{r} = \left( 1 + \sqrt{\frac{1 + |\rho|}{1 - |\rho|}} \right) \sqrt{1 - \vartheta}$, which exists in the interval mentioned above.

Now we are left with the *third* and *fourth* cases, both requiring $2\vartheta + f_3(\sqrt{r}, \lambda') = 1$. We assume $\vartheta \leq \frac{1}{2}$ from now on, because the different definition of $f_3(\sqrt{r}, \lambda')$ makes no difference when $\vartheta \geq \frac{1}{2}$, and the proof for $\rho > 0$ can be copied.

*Third*, if $\lambda'^2 = 2\vartheta + f_3(\sqrt{r}, \lambda') = 1$ and $\vartheta + f_1(\sqrt{r}, \lambda') \geq 1$, $\vartheta + f_2(\sqrt{r}, \lambda') \geq 1$, since $\lambda' = 1$ is fixed, all of $(f_1(\sqrt{r}, \lambda'), f_2(\sqrt{r}, \lambda'), f_3(\sqrt{r}, \lambda'))$ are increasing in $\sqrt{r}$. As a result, the requirement from $\vartheta + f_2(\sqrt{r}, \lambda') \geq 1$ is just

$$\sqrt{r} \geq \max \left\{ 1 + \sqrt{1 - \vartheta}, \ \sqrt{\frac{1 - \vartheta}{1 - \rho^2}} + \frac{1}{1 + |\rho|}, \ \sqrt{\frac{5 + 3|\rho|}{(1 - |\rho|)(2 + |\rho|)^2}} \sqrt{1 - \vartheta} \right\} \tag{37}$$

When $\sqrt{r} \leq \frac{2\lambda'}{1 - |\rho|}$ in $f_3(\sqrt{r}, \lambda')$, we have $\sqrt{r} = \sqrt{\frac{1 - 2\vartheta}{1 - |\rho|^2}} + \frac{1}{1 - |\rho|}$, and we will see this curve does not exist in the diagram. We need to look at $\vartheta + f_1(\sqrt{r}, \lambda') \geq 1$:

If $\sqrt{r} \leq 2$ in $f_1(\sqrt{r}, \lambda')$, then we can limit $|\rho| < \frac{1}{2}$ because we have assumed $\vartheta < \frac{1}{2}$, and now we need $\frac{2|\rho|}{1 + |\rho|} \leq \vartheta < \frac{1}{2}$,

- When $0.28832 \leq |\rho| \leq \frac{1}{2}$, this case does not exist. This is because we also need $\sqrt{r} = \sqrt{\frac{1 - 2\vartheta}{1 - |\rho|^2}} + \frac{1}{1 - |\rho|} \geq 1 + \sqrt{1 - \vartheta}$. But we already know from the start of the *first* case, that $\vartheta \geq \frac{2|\rho|}{1 + |\rho|}$ implies $1 + \sqrt{1 - \vartheta} \geq \sqrt{\frac{1 - 2\vartheta}{1 - |\rho|^2}} + \frac{1}{1 - |\rho|}$.

- When $|\rho| < 0.28832$, this case indeed exists, though it is visually only a tiny segment. The curve exists in the interval defined by

$$\begin{cases} \vartheta \geq \frac{2|\rho|}{1 + |\rho|} \\ \sqrt{r} = \sqrt{\frac{1 - 2\vartheta}{1 - |\rho|^2}} + \frac{1}{1 - |\rho|} \geq 1 + \sqrt{1 - \vartheta} \end{cases}$$

in which the second line is an upper bound. (Actually, $\vartheta \geq \frac{2|\rho|}{1+|\rho|} \implies \sqrt{\frac{1-2\vartheta}{1-|\rho|^2}} + \frac{1}{1-|\rho|} \leq 2$)

In terms of (37) (the requirements of $\vartheta + f_2(\sqrt{r}, \lambda')$), we still need to verify $\sqrt{r} \geq \max\left\{\sqrt{\frac{1-\vartheta}{1-|\rho|^2}} + \frac{1}{1+|\rho|}, \sqrt{\frac{5+3|\rho|}{(1-|\rho|)(2+|\rho|)^2}}\sqrt{1-\vartheta}\right\}$. Using $\sqrt{1-\vartheta} \leq \sqrt{r} - 1$ to replace $\sqrt{1-\vartheta}$, we will find both of these are much weaker than $\sqrt{r} \leq 2$ and not restrictive.

If $\sqrt{r} \geq \frac{5+3|\rho|}{2+2|\rho|}$ in $f_1(\sqrt{r}, \lambda')$, then, in terms of $\vartheta + f_1(\sqrt{r}, \lambda') \geq 1$, we are required to have

$$\begin{cases} \sqrt{r} \geq \frac{5+3|\rho|}{2+2|\rho|} \\ \sqrt{r} \geq \sqrt{\frac{5+3|\rho|}{1-|\rho|}}\sqrt{1-\vartheta} \end{cases}$$

Now, by using $\frac{1}{2}\sqrt{r} \geq \frac{5+3|\rho|}{2(2+2|\rho|)}$ and $\frac{1}{2}\sqrt{r} \geq \frac{1}{2}\sqrt{\frac{5+3|\rho|}{1-|\rho|}}\sqrt{1-\vartheta}$, we can easily verify that the requirements of $\vartheta + f_2(\sqrt{r}, \lambda') \geq 1$ in (37) all holds. So we only need to focus on $\vartheta + f_1(\sqrt{r}, \lambda') \geq 1$.

We further argue that $\sqrt{r} \geq \sqrt{\frac{5+3|\rho|}{1-|\rho|}}\sqrt{1-\vartheta}$ implies $\sqrt{r} \geq \frac{5+3|\rho|}{2+2|\rho|}$.

- If $\sqrt{\frac{5+3|\rho|}{1-|\rho|}}\sqrt{1-\vartheta} < \frac{5+3|\rho|}{2+2|\rho|}$, this condition itself imposes a lower bound on $\vartheta$. Since $\vartheta < \frac{1}{2}$, we actually need $|\rho| \leq 0.3798$ for such lower bound to be smaller than $\frac{1}{2}$. However, when $|\rho|$ is too small, $\sqrt{r} \geq \sqrt{\frac{5+3|\rho|}{1-|\rho|}}\sqrt{1-\vartheta}$ admits no solution at all. Namely,

$$\sqrt{(5+3|\rho|)(1+|\rho|)}\sqrt{1-\vartheta} - \sqrt{1-2\vartheta} \leq \frac{\sqrt{1-|\rho|^2}}{1-|\rho|}$$

admits no solution. The LHS is not monotone, but its minimum in $\vartheta \in (0, \frac{1}{2})$ is taken at

$$\vartheta = \frac{(5+3|\rho|)(1+|\rho|) - 4}{2(5+3|\rho|)(1+|\rho|) - 4}.$$

At this point, the LHS is greater than the RHS, so it admits no solution.

- Since $\sqrt{\frac{5+3|\rho|}{1-|\rho|}}\sqrt{1-\vartheta} \geq \frac{5+3|\rho|}{2+2|\rho|}$, we have $\sqrt{r} \geq \sqrt{\frac{5+3|\rho|}{1-|\rho|}}\sqrt{1-\vartheta} \implies \sqrt{r} \geq \frac{5+3|\rho|}{2+2|\rho|}$.

As a result, $\sqrt{r} = \sqrt{\frac{1-2\vartheta}{1-|\rho|^2}} + \frac{1}{1-|\rho|}$ makes part of the boundary when it is greater than $\sqrt{\frac{5+3|\rho|}{1-|\rho|}}\sqrt{1-\vartheta}$. We will see later, that the latter is also part of the boundary, when it is greater than $\sqrt{\frac{1-2\vartheta}{1-|\rho|^2}} + \frac{1}{1-|\rho|}$.

If $2 < \sqrt{r} < \frac{5+3|\rho|}{2+2|\rho|}$ in $f_1(\sqrt{r}, \lambda')$, then in terms of $\vartheta + f_1(\sqrt{r}, \lambda') \geq 1$, we need $d^2(B, (|\rho|\sqrt{r}, \sqrt{r})) \geq (1-|\rho|^2)(1-\vartheta)$. Because $2 < \sqrt{\frac{1-2\vartheta}{1-|\rho|^2}} + \frac{1}{1-|\rho|} < \frac{5+3|\rho|}{2+2|\rho|}$, we need $|\rho| \leq 0.535$; otherwise no solution for $\vartheta$.

We now need to use (34). When $\vartheta \geq \frac{2|\rho|}{1+|\rho|}$, $d^2(B, (|\rho|\sqrt{r}, \sqrt{r})) \geq (1-\rho^2)(1-\vartheta)$ always holds. However, $\vartheta \geq \frac{2|\rho|}{1+|\rho|}$ also makes $\sqrt{r} < 2$, which contradicts $\sqrt{r} \geq 2$. When $\vartheta < \frac{2|\rho|}{1+|\rho|}$, $d^2(B, (|\rho|\sqrt{r}, \sqrt{r})) \geq (1-\rho^2)(1-\vartheta)$ implies $\sqrt{r} \geq 2 + \sqrt{\frac{2|\rho|}{1+|\rho|} - \vartheta}$. which is

$$\sqrt{1-\rho^2}\sqrt{\frac{2|\rho|}{1+|\rho|} - \vartheta} - \sqrt{1-2\vartheta} \leq \left(\frac{1}{1-|\rho|} - 2\right)\sqrt{1-\rho^2}.$$

We take a close look at the LHS, as a function of $\vartheta$.

- Using the lower bound of $\vartheta$ implied by $\sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{1-|\rho|} < \frac{5+3|\rho|}{2+2|\rho|}$, the LHS is greater than the RHS.

- Using the upper bound of $\vartheta$ implied by $\sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{1-|\rho|} > 2$, the LHS is greater than the RHS.

- The LHS is either increasing in $\vartheta$, or decreasing in $\vartheta$, or first-increasing-then-decreasing. Since $LHS > RHS$ holds at both ends of the interval, it holds for all $\vartheta$. Now we have a contradiction.

When $\sqrt{r} > \frac{2\lambda'}{1-|\rho|}$ in $f_3(\sqrt{r}, \lambda')$, this case produces no curve in the diagram, but there is no easy way to eliminate this case. We still need to use some tedious calculation.

If the smallest term is $\left[(1-\rho^2)\sqrt{r} - \lambda'\right]^2$ in $f_3(\sqrt{r}, \lambda')$, then we have $\sqrt{r} = \sqrt{\frac{1-2\vartheta}{1-\rho^2} + \frac{1}{1-\rho^2}}$ which contradicts $\sqrt{r} \geq \frac{2}{1-|\rho|}$.

If $\frac{2-|\rho|}{1-|\rho|} < a < \frac{2}{1-|\rho|}$, we recall the form of $k(\lambda', a)$:

$$k(\lambda', a) = \begin{cases} d^2\left(D, ((1-|\rho|)\sqrt{r}, -(1-|\rho|)\sqrt{r})\right) & \text{if } \frac{2\lambda'}{1-|\rho|} \leq \sqrt{r} \leq \frac{\lambda'}{1-|\rho|}\left[2 + \frac{|\rho|+\rho^2}{(a-2)(1-\rho^2)-(|\rho|+\rho^2)}\right] \\ \frac{(1-\rho^2)}{1+\frac{\rho^2(a-1)^2}{(a-2)^2}-\frac{2\rho^2(a-1)}{a-2}}\left[-\lambda'\left(1+\frac{a|\rho|}{a-2}\right) + (1-|\rho|)\sqrt{r}\cdot\left(1+\frac{|\rho|(a-1)}{a-2}\right)\right]^2 \\ \qquad if \sqrt{r} \geq \frac{\lambda'}{1-|\rho|}\left[2 + \frac{|\rho|+\rho^2}{(a-2)(1-\rho^2)-(|\rho|+\rho^2)}\right] \end{cases}$$

- When $\sqrt{r} \geq \frac{\lambda'}{1-|\rho|}\left[2 + \frac{|\rho|+\rho^2}{(a-2)(1-\rho^2)-(|\rho|+\rho^2)}\right]$, the expression of $\sqrt{r}$ is

$$\sqrt{r} = \frac{1 + \frac{a|\rho|}{a-2} + \sqrt{1-2\vartheta}\sqrt{1 + \frac{\rho^2(a-1)^2}{(a-2)^2} - \frac{2\rho^2(a-1)}{a-2}}}{(1-|\rho|)(1 + \frac{|\rho|(a-1)}{a-2})}$$

and actually $\sqrt{r} < \frac{\lambda'}{1-|\rho|}\left[2 + \frac{|\rho|+\rho^2}{(a-2)(1-\rho^2)-(|\rho|+\rho^2)}\right]$, which gives a contradiction. To prove this, we take $\vartheta = 0$, and re-arrange the terms:

$$\sqrt{1 + \frac{\rho^2(a-1)^2}{(a-2)^2} - \frac{2\rho^2(a-1)}{a-2}} < 1 + |\rho| + \frac{|\rho|(1 + \frac{|\rho|(a-1)}{a-2})}{(a-2)(1-|\rho|)-|\rho|}$$

$$= \frac{\rho^2 + (a-2)^2(1-\rho^2)}{(a-2)^2(1-|\rho|)-|\rho|(a-2)}$$

$$\Leftrightarrow (a-2)^2 + \rho^2(a-1)^2 - 2\rho^2(a-1)(a-2) < \frac{\left[\rho^2 + (a-2)^2(1-\rho^2)\right]^2}{\left[(a-2)(1-|\rho|)-|\rho|\right]^2}$$

Multiply each side with $\left[(a-2)(1-|\rho|)-|\rho|\right]^2$, and we have a polynomial. We can then factorize $LHS - RHS$ and get $-2(a-2)(a-1)|\rho|(1-|\rho|)\left[\rho^2 + (1-\rho^2)(a-2)^2\right] < 0$.

- When $\sqrt{r} < \frac{\lambda'}{1-|\rho|}\left[2 + \frac{|\rho|+\rho^2}{(a-2)(1-\rho^2)-(|\rho|+\rho^2)}\right]$, we calculate the distance related to Point $D$ and eventually get $2(1-|\rho|)\sqrt{r} - 2(3-|\rho|)\sqrt{r} + \frac{5-3|\rho|}{1-|\rho|} = 1 - 2\vartheta$ which implies $\sqrt{r} = \frac{3-|\rho|}{2(1-|\rho|)} + \frac{1}{2}\sqrt{1 - \frac{4\vartheta}{1-|\rho|}}$. This expression is not too complicated, and we can easily verify $\sqrt{r} < \frac{2}{1-|\rho|}$ which gives us a contradiction.

If $a \leq \frac{2-|\rho|}{1-|\rho|}$, we still have $\sqrt{r} = \frac{3-|\rho|}{2(1-|\rho|)} + \frac{1}{2}\sqrt{1 - \frac{4\vartheta}{1-|\rho|}} < \frac{2}{1-|\rho|}$ like the case above, which is still a contradiction.

*Fourth*, if $\vartheta + f_1(\sqrt{r}, \lambda') = 2\vartheta + f_3(\sqrt{r}, \lambda') = 1$ and $\lambda' \geq 1$, $\vartheta + f_2(\sqrt{r}, \lambda') \geq 1$, then:

For $\vartheta + f_2(\sqrt{r}, \lambda') \geq 1$: When $\lambda'$ is fixed, the exponents of $FP$ and $FN$ are all decreasing in $\sqrt{r}$. We thus need

$$\sqrt{r} \geq \max\left\{\lambda' + \sqrt{1-\vartheta}, \sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{\lambda'}{1+|\rho|}, \sqrt{\frac{5+3|\rho|}{(1-|\rho|)(2+|\rho|)^2}}\sqrt{1-\vartheta}\right\} \tag{38}$$

When $\sqrt{r} \leq \frac{2\lambda'}{1-|\rho|}$ in $2\vartheta + f_3(\sqrt{r}, \lambda') \geq 1$, for $f_1(\sqrt{r}, \lambda')$:

If $\lambda' \leq \sqrt{r} \leq 2\lambda'$ in $f_1(\sqrt{r}, \lambda')$, we have $\lambda' = \sqrt{\frac{1+|\rho|}{1-|\rho|}}\sqrt{1-\vartheta}$, and thus $\vartheta \leq \frac{2|\rho|}{1+|\rho|}$ and $\sqrt{r} = \sqrt{\frac{1-2\vartheta}{1-|\rho|^2}} + \frac{1}{1-|\rho|}\sqrt{\frac{1+|\rho|}{1-|\rho|}}\sqrt{1-\vartheta}$. Since $\sqrt{r} \leq 2\lambda'$, we need both $|\rho| < \frac{1}{2}$ and $\sqrt{\frac{1-2\vartheta}{1-\vartheta}} \leq \frac{(1+|\rho|)(1-2|\rho|)}{1-|\rho|}$. When we come to verify the conditions in (38), the first one still dominates the others. As a result, the curve

$$\sqrt{r} = \sqrt{\frac{1-2\vartheta}{1-|\rho|^2}} + \frac{1}{1-|\rho|}\sqrt{\frac{1+|\rho|}{1-|\rho|}}\sqrt{1-\vartheta}.$$

exists under the following conditions:

$$\begin{cases} |\rho| < \frac{1}{2} \\ \sqrt{\frac{1-2\vartheta}{1-\vartheta}} \leq \frac{(1+|\rho|)(1-2|\rho|)}{1-|\rho|} \\ \sqrt{\frac{1-2\vartheta}{1-\vartheta}} \geq \sqrt{1-|\rho|^2} - \frac{|\rho|(1+|\rho|)}{(1-|\rho|)} \end{cases}$$

(The last one implies $\vartheta \leq \frac{2|\rho|}{1+|\rho|}$.)

If $\sqrt{r} \geq \frac{5+3|\rho|}{2+2|\rho|}$ in $f_1(\sqrt{r}, \lambda')$, we have $\sqrt{r} = \sqrt{\frac{5+3|\rho|}{1-|\rho|}}\sqrt{1-\vartheta}$ and $\lambda'$ is computed with $\sqrt{r} = \sqrt{\frac{1-2\vartheta}{1-|\rho|^2}} + \frac{\lambda'}{1-|\rho|}$.

Now, $\sqrt{r} = \sqrt{\frac{5+3|\rho|}{1-|\rho|}}\sqrt{1-\vartheta} \geq \frac{5+3|\rho|}{2+2|\rho|}$ requires

$$\sqrt{\frac{1-2\vartheta}{1-\vartheta}} \geq \left(1 - \frac{2(1+|\rho|)}{(1-|\rho|)(5+3|\rho|)}\right)\sqrt{(5+3|\rho|)(1+|\rho|)}. \tag{39}$$

and $\vartheta + f_2(\sqrt{r}, \lambda') \geq 1$ requires $\sqrt{\frac{1-2\vartheta}{1-\vartheta}} \geq \frac{1+|\rho|}{1-|\rho|}\left(1 - 2|\rho|\sqrt{\frac{5+3|\rho|}{1+|\rho|}}\right)$ and $\sqrt{\frac{1-2\vartheta}{1-\vartheta}} \geq \frac{\sqrt{1-\rho^2}}{1-|\rho|} - \frac{|\rho|}{1-|\rho|}\sqrt{(5+3|\rho|)(1+|\rho|)}$. The conditions required by $\vartheta + f_2(\sqrt{r}, \lambda') \geq 1$ are actually even weaker than $\sqrt{r} \geq \frac{5+3|\rho|}{2+2|\rho|}$. $\lambda' \geq 1$ requires $\sqrt{\frac{5+3|\rho|}{1-|\rho|}}\sqrt{1-\vartheta} \geq \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{1-|\rho|}$.

- When $|\rho| \geq 0.535$, the RHS of

$$\sqrt{\frac{1-2\vartheta}{1-\vartheta}} \geq \left(1 - \frac{2(1+|\rho|)}{(1-|\rho|)(5+3|\rho|)}\right)\sqrt{(5+3|\rho|)(1+|\rho|)}.$$

  is negative, thus not restrictive.

  Taking the requirement $\sqrt{\frac{5+3|\rho|}{1-|\rho|}}\sqrt{1-\vartheta} \geq \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{1-|\rho|}$ into account, now the boundary consists of

$$\max\left\{\sqrt{\frac{5+3|\rho|}{1-|\rho|}}\sqrt{1-\vartheta},\ \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{1-|\rho|}\right\}.$$

- When $|\rho| < 0.535$: $\sqrt{\frac{5+3|\rho|}{1-|\rho|}}\sqrt{1-\vartheta} \geq \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{1-|\rho|}$ is not restrictive, because it can be re-written as

$$\sqrt{(5+3|\rho|)(1+|\rho|)}\sqrt{1-\vartheta} - \sqrt{1-2\vartheta} \geq \frac{\sqrt{1-\rho^2}}{1-|\rho|}$$

  which always holds, because we have shown the minimum of the LHS is taken at

$$\vartheta = \frac{(5+3|\rho|)(1+|\rho|) - 4}{2(5+3|\rho|)(1+|\rho|) - 4},$$

  and the minimum can be verified to be greater than the RHS. Thus

$$\sqrt{\frac{5+3|\rho|}{1-|\rho|}}\sqrt{1-\vartheta} \geq \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{1-|\rho|}.$$

  is not restrictive. We only need requirement (39).

If $2\lambda' < \sqrt{r} < \frac{5+3|\rho|}{2+2|\rho|}\lambda'$ in $f_1(\sqrt{r}, \lambda')$, this case is very tedious. We present the closed form of $\lambda'$ and $\sqrt{r}$:

$$\lambda' = \left[\left(\frac{1-2|\rho|}{1-|\rho|}\right)^2 + \frac{1-|\rho|}{1+|\rho|}\right]^{-1} \cdot \left[\frac{1-2|\rho|}{1-|\rho|}\sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \sqrt{\left[\left(\frac{1-2|\rho|}{1-|\rho|}\right)^2 + \frac{1-|\rho|}{1+|\rho|}\right](1-\vartheta) - \frac{1-2\vartheta}{(1+|\rho|)^2}}\right]$$

$$\sqrt{r} = \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{\lambda'}{1-|\rho|} = 2\lambda' + \sqrt{1-\vartheta-\frac{1-|\rho|}{1+|\rho|}\lambda'^2}$$

In terms of the requirements from $2\lambda' < \sqrt{r} < \frac{5+3|\rho|}{2+2|\rho|}\lambda'$: First, $\sqrt{r} > 2\lambda'$ will give us

$$\text{either } |\rho| \ge \frac{1}{2} \text{ or } \left(|\rho| < \frac{1}{2} \text{ and } \sqrt{\frac{1-2\vartheta}{1-\vartheta}} > \frac{(1+|\rho|)(1-2|\rho|)}{1-|\rho|}\right)$$

Second, $\sqrt{r} < \frac{5+3|\rho|}{2+2|\rho|}\lambda'$ will give us

$$|\rho| < 0.535 \text{ (the same numerical value which appreared before)}$$

$$\text{and } \sqrt{\frac{1-2\vartheta}{1-\vartheta}} < \left(1 - \frac{2(1+|\rho|)}{(1-|\rho|)(5+3|\rho|)}\right)\sqrt{(5+3|\rho|)(1+|\rho|)}.$$

In terms of other requirements, we show that they are can be implied by the two conditions we have just arrived at.

We first look at the requirement $\lambda' \ge 1$. We need $\sqrt{\frac{1-2\vartheta}{1-\vartheta}} \le (1+|\rho|)\sqrt{\left(\frac{1-2|\rho|}{1-|\rho|}\right)^2 + \frac{1-|\rho|}{1+|\rho|}}$ to make the content of the square root positive, but this is implied by $\sqrt{r} \le \frac{5+3|\rho|}{2+2|\rho|}\lambda'$. Then $\lambda' \ge 1$ is equivalent to: (let $x = \sqrt{\frac{1-2\vartheta}{1-\vartheta}}$)

$$\frac{1-2|\rho|}{1-|\rho|}\sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \sqrt{\left[\left(\frac{1-2|\rho|}{1-|\rho|}\right)^2 + \frac{1-|\rho|}{1+|\rho|}\right](1-\vartheta) - \frac{1-2\vartheta}{(1+|\rho|)^2}} \ge \left(\frac{1-2|\rho|}{1-|\rho|}\right)^2 + \frac{1-|\rho|}{1+|\rho|}$$

$$\Leftrightarrow \frac{1}{\sqrt{2-x^2}}\left[\frac{1-2|\rho|}{1-|\rho|}\frac{x}{\sqrt{1-\rho^2}} + \sqrt{\left[\left(\frac{1-2|\rho|}{1-|\rho|}\right)^2 + \frac{1-|\rho|}{1+|\rho|}\right] - \frac{x^2}{(1+|\rho|)^2}}\right] \ge \left(\frac{1-2|\rho|}{1-|\rho|}\right)^2 + \frac{1-|\rho|}{1+|\rho|}$$

$$\text{for } \max\left\{0, \frac{(1+|\rho|)(1-2|\rho|)}{1-|\rho|}\right\} \le x \le \min\left\{1, \left(1 - \frac{2(1+|\rho|)}{(1-|\rho|)(5+3|\rho|)}\right)\sqrt{(5+3|\rho|)(1+|\rho|)}\right\}$$

and $|\rho| < 0.535$

It always holds, because we can verify the graph of $(LHS - RHS)$ as a bi-variate function of $(|\rho|, x)$ is always above zero.

We then look at the requirement $\sqrt{r} \le \frac{2\lambda'}{1-|\rho|}$: This is equivalent to

$$\frac{1-2\vartheta}{1-\vartheta} \le \frac{\left(\frac{1-2|\rho|}{1-|\rho|}\right)^2 + \frac{1-|\rho|}{1+|\rho|}}{\frac{\left(\frac{1-|\rho|}{1+|\rho|} - \frac{2|\rho|(1-2|\rho|)}{(1-|\rho|)^2}\right)^2}{1-\rho^2} + \frac{1}{(1+|\rho|)^2}}$$

which is always weaker than $\sqrt{\frac{1-2\vartheta}{1-\vartheta}} \le \min\left\{1, \left(1 - \frac{2(1+|\rho|)}{(1-|\rho|)(5+3|\rho|)}\right)\sqrt{(5+3|\rho|)(1+|\rho|)}\right\}$.

We finally look at the requirement from $\vartheta + f_2(\sqrt{r}, \lambda') \ge 1$, or equivalently (38). First, we need to verify $\sqrt{r} \ge \lambda' + \sqrt{1-\vartheta}$. Since $\sqrt{r} = \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{\lambda'}{1-|\rho|}$, this is equivalent to $\sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{|\rho|}{1-|\rho|}\lambda' \ge$

$\sqrt{1-\vartheta}$. It naturally holds if $|\rho| \leq 0.5$, because $\lambda' \geq 1$. When $0.5 < |\rho| < 0.535$, still letting $x = \sqrt{\frac{1-2\vartheta}{1-\vartheta}}$, we have

$$\sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{|\rho|}{1-|\rho|}\lambda' \geq \sqrt{1-\vartheta}$$

$$\Leftrightarrow \frac{x}{\sqrt{1-\rho^2}} + \frac{|\rho|\lambda'}{(1-|\rho|)}\sqrt{2-x^2} \geq 1$$

$$\Leftarrow \frac{x}{\sqrt{1-\rho^2}} + \frac{|\rho|}{(1-|\rho|)}\sqrt{2-x^2} \geq 1$$

The LHS is either increasing in $x \in (0,1)$, or first-increasing-then-decreasing. When $x = 0$ or $1$, the inequality holds for $0.5 < |\rho| < 0.535$, so it always holds.

We still need to verify $\sqrt{r} \geq \max\left\{\sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{\lambda'}{1+|\rho|}, \sqrt{\frac{5+3|\rho|}{(1-|\rho|)(2+|\rho|)^2}}\sqrt{1-\vartheta}\right\}$. With $\sqrt{1-\vartheta} \leq \sqrt{r} - \lambda'$, we can get rid of $\sqrt{1-\vartheta}$ and arrange either of the requirements as an inequality between $\sqrt{r}$ and $\lambda'$. Such an inequality will be weaker than $\sqrt{r} \leq \frac{5+3|\rho|}{2+2|\rho|}\lambda'$.

When $\sqrt{r} > \frac{2\lambda'}{1-|\rho|}$ in $f_3(\sqrt{r}, \lambda')$, we will see this case does not produce any curve in the diagram. We still need to discuss two cases in terms of $f_1(\sqrt{r}, \lambda')$:

If $\sqrt{r} \geq \frac{5+3|\rho|}{2+2|\rho|}\lambda'$ in $f_1(\sqrt{r}, \lambda')$: We have $\sqrt{r} = \sqrt{\frac{5+3|\rho|}{1-|\rho|}}\sqrt{1-\vartheta}$. Because $\sqrt{r} > \frac{2\lambda'}{1-|\rho|}$, we have

$$\lambda' < \frac{1}{2}\sqrt{(5+3|\rho|)(1-|\rho|)}\sqrt{1-\vartheta}$$

To admit a solution for $\lambda' \geq 1$, we need $|\rho| < \frac{1}{3}$ so that the RHS is large enough.

- When $\sqrt{r} = \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{\lambda'}{1-\rho^2}$, $\lambda' = (1-\rho^2)\left(\sqrt{\frac{5+3|\rho|}{1-|\rho|}}\sqrt{1-\vartheta} - \sqrt{\frac{1-2\vartheta}{1-\rho^2}}\right)$. Combining this with $\lambda' < \frac{1}{2}\sqrt{(5+3|\rho|)(1-|\rho|)}\sqrt{1-\vartheta}$, we can get a requirement for $x = \sqrt{\frac{1-2\vartheta}{1-\vartheta}}$:

$$x > \sqrt{(5+3|\rho|)(1+|\rho|)} - \frac{1}{2}\sqrt{\frac{5+3|\rho|}{1+|\rho|}}$$

  but the RHS is always greater than 1. Thus we have a contradition.

- When $\sqrt{r} = \frac{3-|\rho|}{2(1-|\rho|)}\lambda' + \frac{1}{2}\sqrt{\frac{2(1-2\vartheta)}{1-|\rho|} - \frac{1+|\rho|}{1-|\rho|}\lambda'^2}$: We temporarily ignore the relationship between $\lambda'$ and $\vartheta$, and the $\lambda'$ which maximizes $\sqrt{r} = \frac{3-|\rho|}{2(1-|\rho|)}\lambda' + \frac{1}{2}\sqrt{\frac{2(1-2\vartheta)}{1-|\rho|} - \frac{1+|\rho|}{1-|\rho|}\lambda'^2}$ is $\lambda' = \sqrt{\frac{2(3-|\rho|)^2(1-2\vartheta)}{(3-|\rho|)^2(1+|\rho|)+(1-|\rho|)(1+|\rho|)^2}}$. Even with the maximizer $\lambda'$, we still have

$$\frac{3-|\rho|}{2(1-|\rho|)}\lambda' + \frac{1}{2}\sqrt{\frac{2(1-2\vartheta)}{1-|\rho|} - \frac{1+|\rho|}{1-|\rho|}\lambda'^2} < \sqrt{\frac{5+3|\rho|}{1-|\rho|}}\sqrt{1-\vartheta}$$

  for $|\rho| < \frac{1}{3}$. Thus we have no solution for $\lambda'$.

- When $\frac{2-|\rho|}{1-|\rho|} < a < \frac{2}{1-|\rho|}$ and $\sqrt{r} \geq \frac{\lambda'}{1-|\rho|}\left[2 + \frac{|\rho|+\rho^2}{(a-2)(1-\rho^2)-(|\rho|+\rho^2)}\right]$, the contradiction comes from the fact that

$$\sqrt{r} = \sqrt{\frac{5+3|\rho|}{1-|\rho|}}\sqrt{1-\vartheta} < \frac{\lambda'}{1-|\rho|}\left[2 + \frac{|\rho|+\rho^2}{(a-2)(1-\rho^2)-(|\rho|+\rho^2)}\right]$$

  for any $\frac{2-|\rho|}{1-|\rho|} < a < \frac{2}{1-|\rho|}$, $|\rho| < \frac{1}{3}$ and $\lambda' \geq 1$.

If $\frac{2\lambda'}{1-|\rho|} < \sqrt{r} < \frac{5+3|\rho|}{2+2|\rho|}\lambda'$ in $f_1(\sqrt{r}, \lambda')$: For this case to exist, we need $\frac{2}{1-|\rho|} < \frac{5+3|\rho|}{2+2|\rho|}$, which requires $|\rho| < 0.1547$. Solving $d^2(B, (|\rho|\sqrt{r}, \sqrt{r})) = (1-\rho^2)(1-\vartheta)$ in $FP_2$, we have $\sqrt{r} = 2\lambda' + \sqrt{1-\vartheta - \frac{1-|\rho|}{1+|\rho|}\lambda'^2}$

- When $\sqrt{r} = \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{\lambda'}{1-\rho^2}$, we need $\sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{\lambda'}{1-\rho^2} > \frac{2}{1-|\rho|}\lambda'$ which contradicts $\lambda' \geq 1$ when $|\rho| < 0.1547$.

- When $d^2\left(D, ((1-|\rho|)\sqrt{r}, -(1-|\rho|)\sqrt{r})\right)$, the $(\sqrt{r}, \lambda')$ pair are given by

$$\begin{cases} \sqrt{r} = & 2\lambda' + \sqrt{1 - \vartheta - \frac{1-|\rho|}{1+|\rho|}\lambda'^2} \\ \sqrt{r} = & \frac{3-|\rho|}{2(1-|\rho|)}\lambda' + \frac{1}{2}\sqrt{\frac{2(1-2\vartheta)}{1-|\rho|} - \frac{1+|\rho|}{1-|\rho|}\lambda'^2} \end{cases}$$

which implies $\frac{1-3|\rho|}{2(1-|\rho|)}\lambda' + \sqrt{1 - \vartheta - \frac{1-|\rho|}{1+|\rho|}\lambda'^2} = \frac{1}{2}\sqrt{\frac{2(1-2\vartheta)}{1-|\rho|} - \frac{1+|\rho|}{1-|\rho|}\lambda'^2}$.

However, the above equation has no solution for $\lambda'$, because the LHS is always greater than the RHS. We look at $(LHS - RHS)$ from now on, and prove it is positive:

Let $\lambda^* = \frac{\lambda'}{\sqrt{1-\vartheta}}$, and $\lambda^* \geq 1$ as well. Then

$$\frac{(LHS - RHS)}{\sqrt{1-\vartheta}} = \frac{1-3|\rho|}{2(1-|\rho|)}\lambda^* + \sqrt{1 - \frac{1-|\rho|}{1+|\rho|}\lambda^{*2}} - \frac{1}{2}\sqrt{\frac{2}{1-|\rho|}\frac{1-2\vartheta}{1-\vartheta} - \frac{1+|\rho|}{1-|\rho|}\lambda^{*2}}$$

$$\geq \frac{1-3|\rho|}{2(1-|\rho|)}\lambda^* + \sqrt{1 - \frac{1-|\rho|}{1+|\rho|}\lambda^{*2}} - \frac{1}{2}\sqrt{\frac{2}{1-|\rho|} - \frac{1+|\rho|}{1-|\rho|}\lambda^{*2}}$$

Thus we only need to prove

$$\frac{1-3|\rho|}{2(1-|\rho|)}\lambda^* + \sqrt{1 - \frac{1-|\rho|}{1+|\rho|}\lambda^{*2}} > \frac{1}{2}\sqrt{\frac{2}{1-|\rho|} - \frac{1+|\rho|}{1-|\rho|}\lambda^{*2}}$$

From the content of the square roots, we can see that $\lambda^* \leq \sqrt{\frac{1+|\rho|}{1-|\rho|}}$.

Square both sides, and ignore the cross term on the LHS, and we can actually prove a stronger result,

$$\left(\frac{1-3|\rho|}{2(1-|\rho|)}\right)^2 \lambda^{*2} + \left(1 - \frac{1-|\rho|}{1+|\rho|}\lambda^{*2}\right) \geq \frac{1}{2(1-|\rho|)} - \frac{1+|\rho|}{4(1-|\rho|)}\lambda^{*2}.$$

We only need to verify the two ends, $\lambda^* = 1$ and $\lambda^* = \sqrt{\frac{1+|\rho|}{1-|\rho|}}$, to see that this inequality holds for $0 < |\rho| < 0.1547$.

- When $\frac{2-|\rho|}{1-|\rho|} < a < \frac{2}{1-|\rho|}$ and $\sqrt{r} \geq \frac{\lambda'}{1-|\rho|}\left[2 + \frac{|\rho|+\rho^2}{(a-2)(1-\rho^2)-(|\rho|+\rho^2)}\right]$:

$\frac{\lambda'}{1-|\rho|}\left[2 + \frac{|\rho|+\rho^2}{(a-2)(1-\rho^2)-(|\rho|+\rho^2)}\right]$ is at least $\frac{3\lambda'}{1-|\rho|}$, which is still greater than $\sqrt{r} = 2\lambda' + \sqrt{1 - \vartheta - \frac{1-|\rho|}{1+|\rho|}\lambda'^2}$.

We have finished discussing the last case. To sum up the whole phase diagram, it is exactly Theorem 12.

## D.1 Proof of Lemma 2

(We have assumed $\rho > 0$ in Lemma 2.)

Recall optimization 28 and our assumption $b_1 > |h_2|$. The equation of the sub-gradient for $b = (b_1, b_2)'$ is:

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} q'(b_1) \\ q'(b_2) \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}. \tag{40}$$

When $\lambda'$ is sufficiently large, neither of $(b_1, b_2)$ is nonzero. We investigate the process of decreasing $\lambda'$ from $\infty$, and discuss the major stages along the way.

*Stage 1:* When $\lambda'$ is large, both $(b_j, b_{j+1})$ are zero, and SCAD behaves like Lasso. When $\lambda'$ is large, Equation (40) becomes $\lambda' \cdot \text{sgn}(0) = h_1$, $\lambda' \cdot \text{sgn}(0) = h_2$ and so we need

$$\lambda' \geq \lambda'_1 = \max\{|h_1|, |h_2|\}. \tag{41}$$

*Stage 2:* When $\lambda'$ crosses $\lambda_1' = \max\{|h_1|, |h_2|\}$, since we have assumed $h_1 > |h_2| \geq 0$, $b_1$ becomes positive. To see this, consider $\lambda'$ in a very small interval $(\lambda_1' - \delta, \lambda_1') = (|h_1| - \delta, |h_1|)$:

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ 0 \end{bmatrix} + \begin{bmatrix} \lambda' \cdot \mathrm{sgn}(b_1) \\ \lambda' \cdot \mathrm{sgn}\, 0 \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}.$$

Now we have $b_1 = h_1 - \lambda' \cdot \mathrm{sgn}(b_1)$. From this equation, we know $b_1$ has the same sign as $h_1$, so $b_1$ enters the model as a positive number. On the other hand, $b_2$ cannot enter the model before $b_1$, or we have a contradiction. This is because we would have $b_2 = h_2 - \lambda' \cdot \mathrm{sgn}(b_2)$, but the signs of the LHS and RHS can never agree.

Now that $b_1$ is positive, for the above system of equations to admit a solution, we also need

$$|h_2 + \rho\lambda' - \rho h_1| < \lambda' \tag{42}$$

*Stage 3:* When $\lambda'$ continues to decrease, we have two possible cases when the solution path enters the next stage. First, $b_1$ continues to increase and becomes larger than $\lambda'$; then its gradient will change according to the definition of SCAD (see (27)), while $b_2 = 0$ all along. Second, $b_2$ enters the model before $b_1$ gets larger than $\lambda'$.

We start from the first case mentioned above. In this case, at the next critical point $\lambda' = \lambda_2'$, we would have $b_1 = \lambda'$ while $b_2 = 0$ still. Then

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} \lambda' \\ 0 \end{bmatrix} + \begin{bmatrix} \lambda' \\ \lambda' \cdot \mathrm{sgn}\, 0 \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}.$$

Then we have $\lambda_2'^{(1)} = \frac{1}{2} h_1$.

For this equation to admit a solution, we need $|h_2 - \rho\lambda_2'| = |h_2 - \frac{1}{2}\rho h_1| < \lambda_2' = \frac{1}{2} h_1$ which is

$$\frac{-1 + \rho}{2} h_1 < h_2 < \frac{1 + \rho}{2} h_1. \tag{43}$$

With this constraint (43) and $\lambda' \in [\lambda_2'^{(1)}, \lambda_1']$, we can also go back to check the Condition (42) in the previous stage, and we can see it holds.

We then consider the next case in which $b_2$ enters the model first. This case is essentially Lasso. We solve

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} \lambda' \\ \lambda' \cdot \mathrm{sgn}(b_2) \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}.$$

when $\lambda' \in (\lambda_2'^{(2)} - \delta, \lambda_2'^{(2)})$. We then have two cases, depending on the sign of $b_2$ when it enters the model. If $b_2$ enters the model as a positive number, then eventually we have

$$0 < \rho h_1 < h_2 < h_1, \text{ and } \lambda' < \lambda_2' = \frac{h_2 - \rho h_1}{1 - \rho}$$

If $b_2$ enters the model as a negative number, then eventually we have

$$-h_1 < h_2 < \rho h_1, \text{ and } \lambda' < \lambda_2' = \frac{\rho h_1 - h_2}{1 + \rho}.$$

With these constraints and $\lambda' \in [\lambda'^{(2)}, \lambda_1']$, we can also go back to check the Condition (42) in the previous stage. It holds, and we omit the details for brevity.

We have discussed the two cases in the third stage, and we need to decide which one actually happens. When $h_2 \geq \frac{1+\rho}{2} h_1$ or $h_2 \leq \frac{-1+\rho}{2} h_1$, requirement 43 is not met, and thus the second case is the case that happens. The subset of rejection region when $h_2 \geq \frac{1+\rho}{2} h_1$ or $h_2 \leq \frac{-1+\rho}{2} h_1$ is the same as that of Lasso. Now both $b_1$ and $b_2$ are nonzero, and we need not discuss any further.

When $\frac{-1+\rho}{2} h_1 < h_2 < \frac{1+\rho}{2} h_1$, we can verify that

$$\lambda_2'^{(1)} = \frac{1}{2} h_1 > \lambda_2'^{(2)} = \frac{h_2 - \rho h_1}{1 - \rho} \text{ when } h_2 \geq \rho h_1$$

$$\lambda_2'^{(1)} = \frac{1}{2} h_1 > \lambda_2'^{(2)} = \frac{\rho h_1 - h_2}{1 + \rho} \text{ when } h_2 < \rho h_1$$

So $b_1$ becomes larger than $\lambda'$ before $b_2$ enters the model.

*Stage 3:* When $(h_1, h_2)$ satisfies $\frac{-1+\rho}{2}h_1 < h_2 < \frac{1+\rho}{2}h_1$ in the last stage, we stil need to find out when $b_2$ enters the model after $b_1$ becomes greater than $\lambda'$. When $\lambda' < \lambda_2'^{(1)} = \frac{1}{2}h_1$ we still have two possible cases to discuss: First, $b_1$ continues to grow larger than $(a\lambda')$, making the expression of $q'(b_1)$ different again, while $b_2$ is still at zero. Second, $b_2$ enters the model before $b_1$ hits $(a\lambda')$.

Before the discussion of the two cases, we look at the system of equations when $\lambda' \in (\lambda_2' - \delta, \lambda_2')$ for a very small $\delta$.

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{a\lambda' - b_1}{a-1} \\ \lambda' \cdot \mathrm{sgn}(0) \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}.$$

The solution of $b_1$ is $b_1 = \frac{(a-1)h_1 - a\lambda'}{a-2}$ and the sub-gradient for $b_2$ requires

$$|h_2 - \rho b_1| < \lambda' \tag{44}$$

We start from the first case in which $b_1$ reaches $(a\lambda)$ first. At the point $\lambda' = \lambda_3'^{(1)}$, we have $b_1 = a\lambda'$ and

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} a\lambda' \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \lambda' \cdot \mathrm{sgn}(0) \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}.$$

Thus $\lambda_3'^{(1)} = \frac{h_1}{a}$. In terms of the sub-gradient $\mathrm{sgn}(0)$, we need $|h_2 - \rho h_1| < \frac{h_1}{a}$, and thus $(\rho - \frac{1}{a})h_1 < h_2 < (\rho + \frac{1}{a})h_1$. Compare the above equation with Equation 43, and we get

$$(\rho - \frac{1}{a})h_1 < h_2 < \min\left\{ (\rho + \frac{1}{a})h_1, \frac{1+\rho}{2}h_1 \right\} = \begin{cases} (\rho + \frac{1}{a})h_1 & \text{if } a > \frac{2}{1-\rho} \\ \frac{1+\rho}{2}h_1 & \text{if } a \leq \frac{2}{1-\rho} \end{cases} \tag{45}$$

Then we consider the second case in which $b_2$ enters the model first. We look at the condition 44 to find $\lambda_3'^{(2)}$, because condition 44 would become tight at $\lambda = \lambda_3'^{(2)}$. It requires

$$-\lambda' + \frac{\rho(a-1)h_1 - a\rho\lambda'}{a-2} < h_2 < \lambda' + \frac{\rho(a-1)h_1 - a\rho\lambda'}{a-2}$$

The left half of the inequality is equivalent to (We have implicitly used $\rho > 0$.)

$$\lambda' > \frac{\rho(a-1)h_1 - (a-2)h_2}{a + a\rho - 2}$$

The right half of the inequality is

$$(a-2)h_2 - \rho(a-1)h_1 < (a-2-a\rho)\lambda'. \tag{46}$$

It turns out that we still need to discuss whether $a > \frac{2}{1-\rho}$:

When $a > \frac{2}{1-\rho}$, $a - 2 - a\rho > 0$ and requirement (46) is restrictive. We have

$$\lambda' > \lambda_3'^{(2)} = \max\left\{ \frac{(a-2)h_2 - \rho(a-1)h_1}{a-2-a\rho}, \frac{\rho(a-1)h_1 - (a-2)h_2}{a + a\rho - 2} \right\}.$$

Of course only one of the two terms will be positive, and it involves the discussion of whether $(a-2)h_2 > \rho(a-1)h_1$ or not.

When $a \leq \frac{2}{1-\rho}$, we can prove requirement (46) always holds without any requirement. To see this, just plug $\lambda' = \lambda_2' = \frac{h_1}{2}$ into requirement (46), and we will see $(a-2)h_2 - \rho(a-1)h_1 \leq (a-2-a\rho)\lambda_2' < 0$ because it is equivalent to $h_2 < \frac{1+\rho}{2}h_1$. In other words, requirement (46) is not restrictive, and we only need the left half:

$$\lambda' > \lambda_3'^{(2)} = \frac{\rho(a-1)h_1 - (a-2)h_2}{a + a\rho - 2}.$$

Finally, we need to decide how to choose between the two cases:

- When $a > \frac{2}{1-\rho}$ and $(\rho - \frac{1}{a})h_1 < h_2 < (\rho + \frac{1}{a})h_1$, we should choose $\lambda_3'^{(1)} = \frac{h_1}{a}$. This is because now $\frac{h_1}{a} > \max\left\{ \frac{(a-2)h_2 - \rho(a-1)h_1}{a-2-a\rho}, \frac{\rho(a-1)h_1 - (a-2)h_2}{a+a\rho-2} \right\}$ always holds.

- When $a \leq \frac{2}{1-\rho}$ and $(\rho - \frac{1}{a})h_1 < h_2 < \frac{1+\rho}{2}h_1$, we should choose $\lambda_3'^{(1)} = \frac{h_1}{a}$. This is because now $\frac{h_1}{a} > \frac{\rho(a-1)h_1 - (a-2)h_2}{a+a\rho-2}$ always holds.

- When $a > \frac{2}{1-\rho}$ and $h_2 > (\rho + \frac{1}{a})h_1$, we should choose $\lambda_3'^{(2)} = \frac{(a-2)h_2 - \rho(a-1)h_1}{a-2-a\rho} > 0$.

- When $h_2 < (\rho - \frac{1}{a})h_1$, for arbitrary $a$, we should choose $\lambda_3'^{(2)} = \frac{\rho(a-1)h_1 - (a-2)h_2}{a+a\rho-2} > 0$.

In the discussion above, if the Condition (45) is met and $\lambda' < \lambda_3'^{(1)} = \frac{h_1}{a}$, we still have not seen $b_2$ in the model when $b_1$ hits $a\lambda'$, and we need to discuss further.

*Stage 4:* When $\lambda'$ fall below $\lambda_3'^{(1)}$, $b_1$ is greater than $a\lambda'$. Now the system of equations:

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \lambda' \cdot \mathrm{sgn}(0) \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}.$$

Now we have $b_1 = h_1$, and $\lambda' > |h_2 - \rho h_1|$. Thus we know $\lambda_4' = |h_2 - \rho h_1|$. (The sign of $(h_2 - \rho h_1)$ is not determined yet.) When $\lambda'$ fall below $\lambda_4'$, $b_2$ inevitably enters the model.

# E  PROOF OF PROPOSITION 2 (COMPARING SCAD AND LASSO)

Proposition 2 makes two assertions, the first about positive $\rho$ and the second about negative $\rho$. We prove them respectively.

When $\rho < 0$, it would be obvious that the diagram of SCAD is lower than the diagram of Lasso. Reviewing the diagram of Lasso in the main text, it is the maximum of four curves:

$$U_{\mathrm{Lasso}}(\vartheta) = \begin{cases} \max\{h_1(\vartheta), h_2^*(\vartheta)\}, & \text{when } \rho \geq 0, \\ \max\{h_1(\vartheta), h_2^*(\vartheta), h_3^*(\vartheta), h_4^*(\vartheta)\}, & \text{when } \rho < 0, \end{cases} \tag{47}$$

where $h_1(\vartheta) = (1 + \sqrt{1-\vartheta})^2$, $h_2^*(\vartheta) = \left(1 + \sqrt{\frac{1+|\rho|}{1-|\rho|}}\right)^2(1-\vartheta)$, $h_3^*(\vartheta) = \frac{1}{(1-|\rho|)^2}\left(1 + \sqrt{\frac{1-|\rho|}{1+|\rho|}}\sqrt{1-2\vartheta}\right)^2$, and $h_4^*(\vartheta) = \frac{1}{(1-|\rho|)^2}\left(\sqrt{\frac{1+|\rho|}{1-|\rho|}}\sqrt{1-\vartheta} + \sqrt{\frac{1-|\rho|}{1+|\rho|}}\sqrt{1-2\vartheta}\right)^2$.

In terms of SCAD, $h_1(\vartheta) = (1 + \sqrt{1-\vartheta})^2$, $h_2^*(\vartheta) = \left(1 + \sqrt{\frac{1+|\rho|}{1-|\rho|}}\right)^2(1-\vartheta)$ and $h_3^*(\vartheta) = \frac{1}{(1-|\rho|)^2}\left(1 + \sqrt{\frac{1-|\rho|}{1+|\rho|}}\sqrt{1-2\vartheta}\right)^2$ are also present in Theorem 3. (The notation of the corresponding curve of $h_3^*(\vartheta)$ is different, $h_3^*(\vartheta) = \left(\frac{1}{1-\rho} + \sqrt{\frac{1-2\vartheta}{1-\rho^2}}\right)$.) The only different curve is the last one; we compare only the last curve below:

$$h_{\mathrm{Lasso}}(\vartheta) = \frac{1}{(1-|\rho|)^2}\left(\sqrt{\frac{1+|\rho|}{1-|\rho|}}\sqrt{1-\vartheta} + \sqrt{\frac{1-|\rho|}{1+|\rho|}}\sqrt{1-2\vartheta}\right)^2$$

$$h_{\mathrm{SCAD}}(\vartheta) = \begin{cases} \left(\frac{5+3|\rho|}{1-|\rho|}\right)(1-\vartheta), & \text{if } \sqrt{\frac{1-2\vartheta}{1-\vartheta}} \geq \frac{3-4|\rho|-3\rho^2}{(1-|\rho|)}\sqrt{\frac{1+|\rho|}{5+3|\rho|}} \\ \frac{1}{(1-|\rho|)^2}\left(\sqrt{\frac{1+|\rho|}{1-|\rho|}}\sqrt{1-\vartheta} + \sqrt{\frac{1-|\rho|}{1+|\rho|}}\sqrt{1-2\vartheta}\right)^2, & \text{if } \sqrt{\frac{1-2\vartheta}{1-\vartheta}} \leq \frac{(1+|\rho|)(1-2|\rho|)}{1-|\rho|} \\ h_6(\vartheta) & \text{other wise} \end{cases}$$

$$\leq \min\left\{\left(\frac{5+3|\rho|}{1-|\rho|}\right)(1-\vartheta), \frac{1}{(1-|\rho|)^2}\left(\sqrt{\frac{1+|\rho|}{1-|\rho|}}\sqrt{1-\vartheta} + \sqrt{\frac{1-|\rho|}{1+|\rho|}}\sqrt{1-2\vartheta}\right)^2\right\}$$

$$\leq h_{\mathrm{Lasso}}(\vartheta)$$

Thus the assertion when $\rho < 0$ is proven.

When $\rho > 0$, from the details proof in Section D, we know that when $a \leq \frac{2}{1-\rho}$, the diagram of SCAD is the same as that of Lasso except when $\rho < 0.179$ in a tiny neighborhood of $\vartheta = 0$. See Figure 10 for an example.
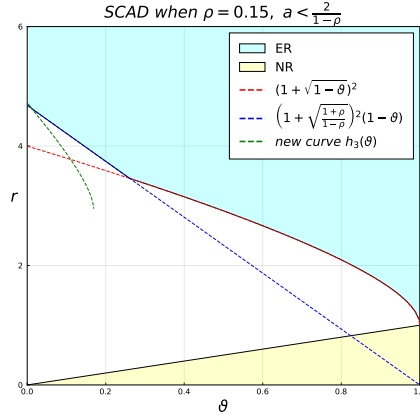
Figure 10: The phase diagram when $0 \leq \rho < 0.179$ with the newly added curve.

When $a > \frac{2}{1-\rho}$ and increases, the penalty function of SCAD converges to that of Lasso and so does the rejection region. Eventually the tiny corner will vanish, and the diagram of SCAD with optimal $(a^*, \lambda^*)$ and $\rho > 0$ will be the same as Lasso.

## F    PROOF OF THEOREM 4 (THRESHOLDED LASSO)

As described in Section A, our proof has three parts: (a) deriving the rejection region, (b) obtaining the rate of convergence of $\mathbb{E}[H(\hat{\beta}, \beta)]$, and (c) calculating the phase diagram.

**Part 1: Deriving the rejection region.**    Recall that the rejection region $\mathcal{R}$ is as defined in (12). Still use the scaled version of $(\lambda, t, x_j' y, x_{j+1}' y)$: Define $h_1 = x_j' y / \sqrt{2 \log(p)}$, $h_2 = x_{j+1}' y / \sqrt{2 \log(p)}$, $\lambda' = \lambda / \sqrt{2 \log(p)}$ and $t' = t / \sqrt{2 \log(p)}$. Consider Lasso decomposed into bivariate sub-problems, and for $(x_j, x_{j+1})$, $(\hat{b}_1, \hat{b}_2)$ minimizes

$$L(b) \equiv \frac{1}{2} b' \begin{bmatrix} 1 & \rho \\ \rho' & 1 \end{bmatrix} b + b'h + \lambda' \|b\|_1 \tag{48}$$

It is seen that $(\hat{\beta}_j, \hat{\beta}_{j+1}) = \sqrt{2 \log(p)}(\hat{b}_1, \hat{b}_2)$. Thresholded Lasso applies threshold $t$ to $(\hat{\beta}_j, \hat{\beta}_{j+1})$, which is equivalent to thresholding $(\hat{b}_1, \hat{b}_2)$ with $t'$.

Fix $\rho \geq 0$. The next lemma gives the explicit solution to (16) in the case of $h_1 > |h_2|$. It is proved in Section F.1.

**Lemma 3.** *Consider the variable selection method by solving the optimization 48 and then thresholding the solution with $t'$; if $(\hat{b}_1)$ (or $(\hat{b}_2)$) survives the thresholding, then variable $x_j$ (or $x_{j+1}$) is selected. Suppose $h_1 > |h_2|$ and $\rho \geq 0$, then*

- *When $\lambda' > h_1$, neither of $(x_j, x_{j+1})$ is selected.*

- *If $h_1 \geq \lambda'$ and $\rho h_1 - \lambda'(+\rho) \leq h_2 \leq \rho h_1 + \lambda'(1-\rho)$, then: When $h_1 \leq \lambda' + t'$, neither of $(x_j, x_{j+1})$ is selected. When $h_1 > \lambda' + t'$, only $x_j$ is selected.*

- *If $h_1 \geq \lambda'$ and $h_2 > \rho h_1 + \lambda'(1-\rho)$, then*

    1. *When $h_1 < \rho h_2 + \lambda'(1-\rho) + t'(1-\rho^2)$, neither of $(x_j, x_{j+1})$ is selected.*
    2. *When $h_1 \geq \rho h_2 + \lambda'(1-\rho) + t'(1-\rho^2)$ and $h_2 \leq \rho h_1 + \lambda'(1-\rho) + t'(1-\rho^2)$, only $x_j$ is selected.*
    3. *When $h_2 > \rho h_1 + \lambda'(1-\rho) + t'(1-\rho^2)$, both $(x_j, x_{j+1})$ are selected.*

- *If $h_1 \geq \lambda'$ and $h_2 < \rho h_1 - \lambda'(1+\rho)$, then*

56

1. *When $h_1 < \rho h_2 + \lambda'(1 + \rho) + t'(1 - \rho^2)$, neither of $(x_j, x_{j+1})$ is selected.*

2. *When $h_1 \geq \rho h_2 + \lambda'(1 + \rho) + t'(1 - \rho^2)$ and $h_2 \leq \rho h_1 + \lambda'(1 + \rho) + t'(1 - \rho^2)$, only $x_j$ is selected.*

3. *When $h_2 > \rho h_1 + \lambda'(1 + \rho) + t'(1 - \rho^2)$, both $(x_j, x_{j+1})$ are selected.*

Following similar reasoning to that of Elastic net, we can use Lemma 3 to write explicitely the rejection region $\mathcal{R}$, which is the region in $\mathbb{R}^2$ where the value of $(h_1, h_2)$ implies $x_j$ will get selected eventually. The rejection region of Thresholded Lasso and $\rho > 0$ is

$$\mathcal{R} = \{(h_1, h_2) : h_1 > \rho h_2 + \lambda'(1 - \rho) + t(1 - \rho^2), \ h_1 > \lambda' +' t'\}$$
$$\cup \{(h_1, h_2) : h_1 > \rho h_2 + \lambda'(1 + \rho) + t'(1 - \rho^2)\} \cup \{(h_1, h_2) : h_1 < \rho h_2 - \lambda'(1 + \rho) - t'(1 - \rho^2)\}$$
$$\cup \{(h_1, h_2) : h_1 < \rho h_2 - \lambda'(1 - \rho) - t(1 - \rho^2), \ h_1 < -\lambda' -' t'\}. \tag{49}$$

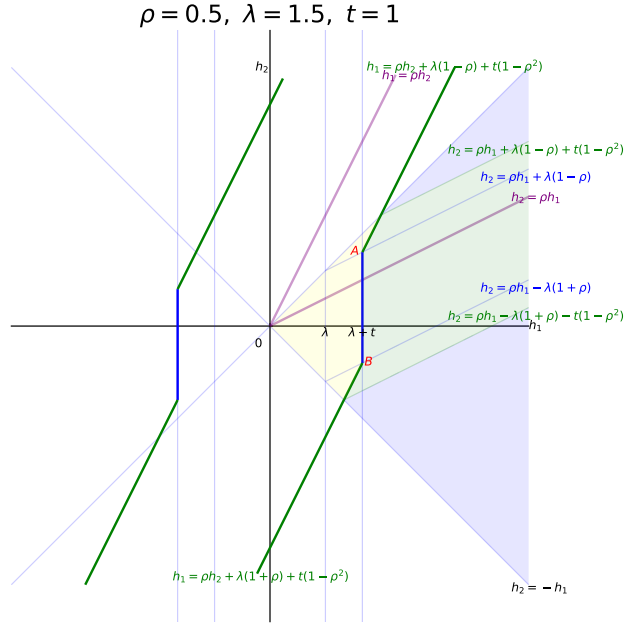See Figure 7 for a visualization of the rejection region.



Figure 11: The rejection region of Thresholded Lasso for $\rho \geq 0$.

**Part 2. Analyzing the Hamming error.** The discussion of Elastic net can be applied here as well, and we present Theorem 13 directly.

**Theorem 13.** *Suppose the conditions of Theorem 4 hold. Let $\lambda' = \lambda/\sqrt{2\log(p)}$ and $t' = t/\sqrt{2\log(p)}$ in Thresholded Lasso. The correlation $\rho \in (-1, 1)$. As $p \to \infty$,*

$$\mathrm{FP}_p = L_p p^{1 - \min\left\{f_1(\sqrt{r}, \lambda', t'), \ \vartheta + f_2(\sqrt{r}, \lambda', t')\right\}}, \qquad \mathrm{FN}_p = L_p p^{1 - \min\left\{\vartheta + f_3(\sqrt{r}, \lambda', t'), \ 2\vartheta + f_4(\sqrt{r}, \lambda', t')\right\}},$$

*where (below, $d^2_{|\rho|}(u,v)$ is as in Definition 1)*

$$f_1(\sqrt{r}, \lambda', t') = \min\left\{ \frac{1}{1-\rho^2} \left[ (1+|\rho|)\lambda' + (1-\rho^2)t' \right]^2, \ (\lambda'+t')^2 \right\}$$

$$f_2(\sqrt{r}, \lambda', t') = \begin{cases} (\lambda'+t'-|\rho|\sqrt{r})^2 & \text{if } \sqrt{r} \le \frac{\lambda'(1-|\rho|)}{1-\rho^2} \\ d^2_{|\rho|}((\lambda'+t', \lambda'+|\rho|t'), (|\rho|\sqrt{r}, \sqrt{r})) & \text{if } \sqrt{r} \in (\frac{\lambda'(1-|\rho|)}{1-\rho^2}, \lambda'+|\rho|t') \\ \frac{1}{1-\rho^2} \left[ \lambda'(1-|\rho|) + t'(1-\rho^2) \right]^2 & \text{if } \sqrt{r} \ge \lambda'+|\rho|t' \end{cases}$$

$$f_3(\sqrt{r}, \lambda', t') = \min\left\{ (\sqrt{r}-\lambda'-t')_+^2, \ \frac{1}{1-\rho^2} \left[ (1-\rho^2)\sqrt{r} - t'(1-\rho^2) - \lambda'(1-|\rho|) \right]^2 \right\},$$

$$f_4(\sqrt{r}, \lambda', t') = \frac{1}{1-\rho^2} \left[ (1-\rho^2)\sqrt{r} - t'(1-\rho^2) - \lambda'(1-\rho) \right]^2$$

**Remark 2.** *When $\rho > 0$ in Theorem 13, we notice that $f_3(\sqrt{r}, \lambda', t') \le f_4(\sqrt{r}, \lambda', t')$, and thus* $\text{FN}_p$ *can be simplified:*

$$\text{FN}_p = L_p p^{1-\vartheta - f_3(\sqrt{r}, \lambda', t')}.$$

*When $\rho < 0$, such simplification is not available.*

The proof of Theorem 13 is easy given the simple rejection region shown in Figure 11, and we omit it for brevity.

**Part 3. Calculating the phase diagram.** The boundary line between Almost Full Recovery and No Recovery is still $r = \vartheta$, and the proof is similar to that of Elastic net. The rest of this part calculates the curve between Almost Full Recovery and Exact Recovery.

In such calculation, thresholded Lasso has two tuning parameters, $(\lambda', t')$, and thus we need one more equality additional to the important fact noted in the Part 3 of Elastic net. In other words, we not only need

$$\min\{f_1(\sqrt{r}, \lambda', t'), \ \vartheta + f_2(\sqrt{r}, \lambda', t')\} = \min\{\vartheta + f_3(\sqrt{r}, \lambda', t'), \ 2\vartheta + f_4(\sqrt{r}, \lambda', t')\} = 1$$

but also need one more equation. This gives us more than 4 cases for other methods. (For brevity, we use $f_i$ ($i = 1, 2, 3, 4$) as shorthand of $f_i(\sqrt{r}, \lambda', t')$ for the rest of this part.)

For the rest of this section, we use a clearer way to discuss all the cases; that is, we discuss each possible curve and find out whether they can be present in some interval of $\vartheta$.

*We first talk about $\rho > 0$:* In this case, since we always need $\vartheta + f_3 \ge 1$, we much have $\sqrt{r} \ge \lambda' + t'$. As a result, for $\vartheta + f_2 \ge 1$, it can only be $\vartheta + \frac{1}{1-\rho^2} \left[ \lambda'(1-|\rho|) + t'(1-\rho^2) \right]^2 \ge 1$ since $\sqrt{r} \ge \lambda' + t'$. Also, according to Remark 2, we can ignore the requirement $2\vartheta + f_4 \ge 1$ for $\rho \ge 0$.

*First*, we study the curve $\sqrt{r} = 2\sqrt{\frac{1-\vartheta}{1-\rho^2}}$, which is the curve given by letting

$$\vartheta + \frac{1}{1-\rho^2} \left[ \lambda'(1-|\rho|) + t'(1-\rho^2) \right]^2 = \vartheta + \frac{1}{1-\rho^2} \left[ (1-\rho^2)\sqrt{r} - t'(1-\rho^2) - \lambda'(1-|\rho|) \right]^2 = 1$$

in $f_2$ and $f_3$. We also need $\vartheta + f_1, \vartheta + f_3 \ge 1$ and one more equality. One possible case is

$$\begin{cases} \frac{1}{1-\rho^2} \left[ (1+|\rho|)\lambda' + (1-\rho^2)t' \right]^2 \ge 1 \\ (\lambda'+t')^2 \ge 1 \\ \vartheta + \frac{1}{1-\rho^2} \left[ \lambda'(1-|\rho|) + t'(1-\rho^2) \right]^2 = 1 \\ \vartheta + (\sqrt{r}-\lambda'-t')_+^2 = 1 \\ \vartheta + \frac{1}{1-\rho^2} \left[ (1-\rho^2)\sqrt{r} - t'(1-\rho^2) - \lambda'(1-|\rho|) \right]^2 = 1 \end{cases}$$

which gives $\lambda' = \frac{1+\rho}{\rho} \left( \frac{1}{\sqrt{1-\rho^2}} - 1 \right) \sqrt{1-\vartheta}$ and $t' = \sqrt{\frac{1+\vartheta}{1-\rho^2}} - \frac{1}{\rho} \left( \frac{1}{\sqrt{1-\rho^2}} - 1 \right) \sqrt{1-\vartheta}$. In this case, $t' \ge 0$ always holds; $(\lambda'+t')^2 \ge 1 \implies \frac{1}{1-\rho^2} \left[ (1+|\rho|)\lambda' + (1-\rho^2)t' \right]^2 \ge 1$, and $(\lambda'+t')^2 \ge 1$ is equivalent to $\sqrt{r} = 2\sqrt{\frac{1-\vartheta}{1-\rho^2}} \ge 1 + \sqrt{1-\vartheta}$. This is a sufficient condition for this curve to show up in the diagram.

*Second*, we study the curve $\sqrt{r} = 1 + \sqrt{1-\vartheta}$, which is given by $(\lambda'+t')^2 = \vartheta + (\sqrt{r}-\lambda'-t')_+^2 = 1$. We also need $f_1, \vartheta + f_2, \vartheta + f_3 \geq 1$ and one more equality. Depending on which requirement to take equality, we discuss two possible cases:

If $\frac{1}{1-\rho^2}\left[(1+|\rho|)\lambda' + (1-\rho^2)t'\right]^2 = 1$, and $\vartheta + \frac{1}{1-\rho^2}\left[\lambda'(1-|\rho|) + t'(1-\rho^2)\right]^2 \geq 1$, $\vartheta + \frac{1}{1-\rho^2}\left[(1-\rho^2)\sqrt{r} - t'(1-\rho^2) - \lambda'(1-|\rho|)\right]^2 \geq 1$, then we will get $\lambda' = \frac{\sqrt{1-\rho^2}-(1-\rho^2)}{\rho(1+\rho)}$ and $t' = \frac{(1+\rho)-\sqrt{1-\rho^2}}{\rho(1+\rho)}$. From the two inequality requirements, we get

$$\sqrt{1-\vartheta} \leq \min\left\{\frac{-\sqrt{1-\rho^2} + 2\rho + 2}{(1+\rho)^2}\sqrt{1-\rho^2}, \frac{1-\rho}{1+\rho}\right\} = \frac{1-\rho}{1+\rho}$$

If $\vartheta + \frac{1}{1-\rho^2}\left[(1-\rho^2)\sqrt{r} - t'(1-\rho^2) - \lambda'(1-|\rho|)\right]^2 = 1$ and $\frac{1}{1-\rho^2}\left[(1+|\rho|)\lambda' + (1-\rho^2)t'\right]^2 \geq 1$, $\vartheta + \frac{1}{1-\rho^2}\left[\lambda'(1-|\rho|) + t'(1-\rho^2)\right]^2 \geq 1$, then we will get $\lambda' = \frac{1+\rho}{\rho}\left(\frac{1}{\sqrt{1-\rho^2}} - 1\right)\sqrt{1-\vartheta}$, $t' = 1 - \lambda'$. From $t' \geq 0$ and the two inequality requirements, we get

$$\frac{1-\rho}{1+\rho} \leq \sqrt{1-\vartheta} \leq \min\left\{\left(\frac{2}{\sqrt{1-\rho^2}} - 1\right)^{-1}, \frac{\rho}{1+\rho}\left(\frac{1}{\sqrt{1-\rho^2}} - 1\right)^{-1}\right\} = \left(\frac{2}{\sqrt{1-\rho^2}} - 1\right)^{-1}$$

Taking the intersection of the first two cases, we already know that $\sqrt{r} = 1 + \sqrt{1-\vartheta}$ exists as long as $\sqrt{r} = 1 + \sqrt{1-\vartheta} \geq 2\sqrt{\frac{1-\vartheta}{1-\rho^2}}$. There is one more case left, but the interval of $\vartheta$ for $\sqrt{r} = 1 + \sqrt{1-\vartheta}$ to exist will be a subset of what we already have, so we omit it.

Now we already seem to have the whole phase curve, but the tricky part of thresholded Lasso having two tunable parameters is that we might have multiple curves for the same $\vartheta$, and we need to take the minimum across all the curves. Thus we need to continue discussing all the other curves. For $\rho > 0$, we have three more to go.

*Third*, we study the curve $\sqrt{r} = \frac{2\sqrt{1-\rho^2}-(1+\rho)}{(1-\rho)\sqrt{1-\rho^2}}\sqrt{1-\vartheta} + \frac{1}{\sqrt{1-\rho^2}}$, given by

$$\frac{1}{1-\rho^2}\left[(1+|\rho|)\lambda' + (1-\rho^2)t'\right]^2 = \vartheta + (\sqrt{r}-\lambda'-t')_+^2 = \vartheta + \frac{1}{1-\rho^2}\left[(1-\rho^2)\sqrt{r} - t'(1-\rho^2) - \lambda'(1-|\rho|)\right]^2 = 1.$$

Now we have $\lambda' = \frac{1+\rho}{\rho}\left(\frac{1}{\sqrt{1-\rho^2}} - 1\right)\sqrt{1-\vartheta}$ and $t' = \frac{1}{\sqrt{1-\rho^2}} - \frac{1+\rho}{\rho(1-\rho)}\left(\frac{1}{\sqrt{1-\rho^2}} - 1\right)\sqrt{1-\vartheta}$. From $t' \geq 0$ and $f_1, \vartheta + \vartheta + f_2 \geq 1$, we get

$$\sqrt{1-\vartheta} \leq \min\left\{\left[\frac{(1+\rho)^2}{\rho\sqrt{1-\rho^2}}\left(\frac{1}{\sqrt{1-\rho^2}} - 1\right)\right]^{-1}, \frac{1-\rho}{1+\rho}, \left(2 - \frac{2\sqrt{1-\rho^2}}{1-\rho} + \frac{1+\rho}{1-\rho}\right)^{-1}\right\} = \frac{1-\rho}{1+\rho}$$

As we can see, when $\sqrt{1-\vartheta} \leq \frac{1-\rho}{1+\rho}$, now we have 2 curves, both of which seem to be the boundary. We must take the *lower one* then.

*Fourth*, we study the curve $\sqrt{r} = \left(1 + \frac{1+\rho}{2\sqrt{1-\rho^2}}\right)\sqrt{1-\vartheta} + \frac{1-\rho}{2\sqrt{1-\rho^2}}$ given by

$$\frac{1}{1-\rho^2}\left[(1+|\rho|)\lambda' + (1-\rho^2)t'\right]^2 = \vartheta + \frac{1}{1-\rho^2}\left[\lambda'(1-|\rho|) + t'(1-\rho^2)\right]^2 = \vartheta + (\sqrt{r}-\lambda'-t')_+^2 = 1.$$

Now we have $\lambda' = \frac{1}{2\rho}\left(1 - \sqrt{1-\vartheta}\right)\sqrt{1-\rho^2}$ and $t' = \frac{1}{2\rho\sqrt{1-\rho^2}}\left[(1+\rho)\sqrt{1-\vartheta} - (1-\rho)\right]$.

From $t' \geq 0$ and $f_1, \vartheta + f_3 \geq 1$, we need

$$\max\left\{\frac{1-\rho}{1+\rho}, \frac{2\sqrt{1-\rho^2}-(1-\rho)}{1+\rho}\right\} \leq \sqrt{1-\vartheta} \leq \frac{1-\rho}{2}\cdot\left(\frac{3-\rho}{2} - \sqrt{1-\rho^2}\right)^{-1}$$

This gives us an empty set, because actually

$$\frac{1-\rho}{2} \cdot \left(\frac{3-\rho}{2} - \sqrt{1-\rho^2}\right)^{-1} < \frac{2\sqrt{1-\rho^2} - (1-\rho)}{1+\rho}.$$

*Fifth*, we study the curve $\sqrt{r} = \sqrt{\frac{1-\vartheta}{1-\rho^2} + \frac{2\rho+2-\sqrt{1-\rho^2}}{(1+\rho)^2}}$ given by

$$\frac{1}{1-\rho^2}\left[(1+|\rho|)\lambda' + (1-\rho^2)t'\right]^2 = (\lambda'+t')^2 = \vartheta + \frac{1}{1-\rho^2}\left[(1-\rho^2)\sqrt{r} - t'(1-\rho^2) - \lambda'(1-|\rho|)\right]^2 = 1.$$

We get $\lambda' = \frac{\sqrt{1-\rho^2}-(1-\rho^2)}{\rho(1+\rho)}$ and $t' = \frac{(1+\rho)-\sqrt{1-\rho^2}}{\rho(1+\rho)}$. From $\vartheta + f_2, \vartheta + f_3 \geq 1$, we get

$$\frac{1-\rho}{1+\rho} \leq \sqrt{1-\vartheta} \leq \frac{2\sqrt{1-\rho^2}-(1-\rho)}{1+\rho}$$

*Summarising all the five curves when $\rho > 0$:* We will elimiate the *third* and *fifth* curve. For the *third* curve, when $\sqrt{1-\vartheta} \leq \frac{1-\rho}{1+\rho}$, it is always larger than the other curve $\sqrt{r} = 1 + \sqrt{1-\vartheta}$. In other words, when $\sqrt{1-\vartheta} \leq \frac{1-\rho}{1+\rho}$, we always have

$$\sqrt{r} = \frac{2\sqrt{1-\rho^2}-(1+\rho)}{(1-\rho)\sqrt{1-\rho^2}}\sqrt{1-\vartheta} + \frac{1}{\sqrt{1-\rho^2}} \geq \sqrt{1-\vartheta} + 1$$

In fact, $(LHS - RHS)$ takes its minimum at $\sqrt{1-\vartheta} = \frac{1-\rho}{1+\rho}$, which is exactly zero.

For the *fifth* curve, when $\frac{1-\rho}{1+\rho} \leq \sqrt{1-\vartheta} \leq \frac{2\sqrt{1-\rho^2}-(1-\rho)}{1+\rho}$, we have

$$\sqrt{r} = \sqrt{\frac{1-\vartheta}{1-\rho^2} + \frac{2\rho+2-\sqrt{1-\rho^2}}{(1+\rho)^2}} \geq \max\left\{1+\sqrt{1-\vartheta}, 2\sqrt{\frac{1-\vartheta}{1-\rho^2}}\right\}$$

which can be verified using $\frac{1-\rho}{1+\rho} \leq \sqrt{1-\vartheta} \leq \frac{2\sqrt{1-\rho^2}-(1-\rho)}{1+\rho}$ in a similar manner. To sum up, for $\rho \geq 0$, the phase curve of thresholded Lasso is

$$\sqrt{r} = \max\left\{1+\sqrt{1-\vartheta}, 2\sqrt{\frac{1-\vartheta}{1-\rho^2}}\right\}.$$

*We then talk about $\rho < 0$*, which now requires additionally $2\vartheta + f_4 \geq 1$, or

$$2\vartheta + \left[(1-\rho^2)\sqrt{r} - \lambda'(1+|\rho|) - (1-\rho^2)t'\right]_+^2 \geq 1.$$

When $\vartheta \geq \frac{1}{2}$, this newly added requirement has no effects, and the right half ($\vartheta \geq \frac{1}{2}$) of the phase diagram should be the same as that of $\rho \geq 0$. As a result, we can limit ourselve to consider $\vartheta \leq \frac{1}{2}$.

Also note that we used to ignore $2\vartheta + f_4 \geq 1$ for $\rho \geq 0$ because it is not restrictive; now we have added it, and it is the only difference between the cases of $\rho \geq 0$ and $\rho < 0$ (because $f_1, \vartheta + f_2, g_3$ only rely on $|\rho|$). As a result, we only need to discuss this additional requirement; since we have eliminated three curves when $\rho > 0$, there is no need to discuss them again.

*First*, we study the curve $\sqrt{r} = 2\sqrt{\frac{1-\vartheta}{1-\rho^2}}$, which is the curve given by letting

$$\vartheta + \frac{1}{1-\rho^2}\left[\lambda'(1-|\rho|) + t'(1-\rho^2)\right]^2 = \vartheta + \frac{1}{1-\rho^2}\left[(1-\rho^2)\sqrt{r} - t'(1-\rho^2) - \lambda'(1-|\rho|)\right]^2 = 1$$

in $\vartheta + f_2$ and $\vartheta + f_3$. We also need $f_1, \vartheta + f_3, 2\vartheta + f_4 \geq 1$ and one more quality. When $\rho > 0$, we used to consider only one case, but now we need to discuss all four cases and take the union to get the interval of $\vartheta$, because each case has different $(\lambda', t')$ and lead to different intervals of $\vartheta$.

If $\vartheta + \frac{1}{1-\rho^2}\left[\lambda'(1-|\rho|)+t'(1-\rho^2)\right]^2 = \vartheta + (\sqrt{r} - \lambda' - t')^2_+ = \vartheta + \frac{1}{1-\rho^2}\left[(1-\rho^2)\sqrt{r} - t'(1-\rho^2) - \lambda'(1-|\rho|)\right]^2 = 1$, and $\vartheta + f_1, 2\vartheta + f_4 \geq 1$, this is the case we covered when $\rho \geq 0$. Then $2\vartheta + \frac{1}{1-\rho^2}\left[(1-\rho^2)\sqrt{r} - t'(1-\rho^2) - \lambda'(1+|\rho|)\right]^2 \geq 1$ implies $\sqrt{\frac{1-2\vartheta}{1-\vartheta}} \leq \frac{2\sqrt{1-\rho^2}-(1+|\rho|)}{1-|\rho|}$ and other requirements imply $\sqrt{1-\vartheta}\left(\frac{2}{\sqrt{1-\rho^2}}-1\right) > 1$. (The first constraint implies $\vartheta \geq$ some value, and the second one implies $\vartheta \leq$ some value.) The overall requirements are:

$$\sqrt{1-\vartheta}\left(\frac{2}{\sqrt{1-\rho^2}}-1\right) > 1, \text{ and } \sqrt{\frac{1-2\vartheta}{1-\vartheta}} \leq \frac{2\sqrt{1-\rho^2}-(1+|\rho|)}{1-|\rho|}$$

If $\frac{1}{1-\rho^2}\left[(1+|\rho|)\lambda' + (1-\rho^2)t'\right]^2 = \vartheta + \frac{1}{1-\rho^2}\left[\lambda'(1-|\rho|)+t'(1-\rho^2)\right]^2 = \vartheta + \frac{1}{1-\rho^2}\left[(1-\rho^2)\sqrt{r} - t'(1-\rho^2) - \lambda'(1-|\rho|)\right]^2 = 1$, and other terms are greater than one, then we have $\lambda' = \frac{1-\rho^2}{2|\rho|} \cdot \frac{1-\sqrt{1-\vartheta}}{\sqrt{1-\rho^2}}, t' = \frac{1}{\sqrt{1-\rho^2}} - \frac{\lambda'}{1-|\rho|}$. In this case, $t' \geq 0$ and $f_1, \vartheta + f_3, 2\vartheta + f_4 \geq 1$ implies

$$\sqrt{1-\vartheta} \geq \max\left\{\frac{1-|\rho|}{1+|\rho|}, \frac{2\sqrt{1-\rho^2}-(1-|\rho|)}{1+|\rho|}, \frac{1-|\rho|}{(3-|\rho|)-2\sqrt{1-\rho^2}}\right\} = \frac{2\sqrt{1-\rho^2}-(1-|\rho|)}{1+|\rho|}$$

If $(\lambda' + t')^2 = \vartheta + \frac{1}{1-\rho^2}\left[\lambda'(1-|\rho|)+t'(1-\rho^2)\right]^2 = \vartheta + \frac{1}{1-\rho^2}\left[(1-\rho^2)\sqrt{r} - t'(1-\rho^2) - \lambda'(1-|\rho|)\right]^2 = 1$ and other terms are greater than one, then we have $\lambda' = \frac{1+|\rho|}{|\rho|}\left(1 - \sqrt{\frac{1-\vartheta}{1-\rho^2}}\right)$ and $t' = 1 - \lambda'$. In this case, $\lambda' \geq 0$ and $t' \geq 0$ implies $\frac{\sqrt{1-\rho^2}}{1+|\rho|} \leq \sqrt{1-\vartheta} \leq \sqrt{1-\rho^2}$, which is weaker than the requirements from $f_1, \vartheta + f_3$ and $2\vartheta + f_4$:

$$\left(\frac{2}{\sqrt{1-\rho^2}}-1\right)^{-1} \leq \sqrt{1-\vartheta} \leq \frac{2\sqrt{1-\rho^2}-(1-|\rho|)}{1+|\rho|}, \text{ and } \frac{2\sqrt{1-\rho^2}}{1-|\rho|}+\sqrt{1-2\vartheta} \leq \frac{3-|\rho|}{1-|\rho|}\sqrt{1-\vartheta}$$

If $\vartheta + \frac{1}{1-\rho^2}\left[\lambda'(1-|\rho|)+t'(1-\rho^2)\right]^2 = \vartheta + \frac{1}{1-\rho^2}\left[(1-\rho^2)\sqrt{r} - t'(1-\rho^2) - \lambda'(1-|\rho|)\right]^2 = 2\vartheta + \frac{1}{1-\rho^2}\left[(1-\rho^2)\sqrt{r} - t'(1-\rho^2) - \lambda'(1+|\rho|)\right]^2 = 1$ and other terms are greater than one, then we have $\lambda' = \frac{1-\rho^2}{2|\rho|}\frac{1}{\sqrt{1-\rho^2}}\left(\sqrt{1-\vartheta} - \sqrt{1-2\vartheta}\right)$ and $t' = \sqrt{\frac{1-\vartheta}{1-\rho^2}} - \frac{\lambda'}{1+|\rho|}$. In this case, $\lambda' \geq 0$ and $t' \geq 0$ implies $\sqrt{\frac{1-2\vartheta}{1-\vartheta}} \geq 1 - \frac{2|\rho|}{1-|\rho|}$, which is still weaker than the requirements from $f_1, \vartheta + f_3 \geq 1$:

$$\sqrt{\frac{1-2\vartheta}{1-\vartheta}} \geq \frac{2\sqrt{1-\rho^2}-(1+|\rho|)}{1-|\rho|}, \text{ and } \frac{2\sqrt{1-\rho^2}}{1-|\rho|}+\sqrt{1-2\vartheta} \leq \frac{3-|\rho|}{1-|\rho|}\sqrt{1-\vartheta}$$

Taking the union over all the four cases of curve $\sqrt{r} = 2\sqrt{\frac{1-\vartheta}{1-\rho^2}}$, it exists for $\vartheta$ satisfying $\sqrt{1-\vartheta} \geq \left(\frac{2}{\sqrt{1-\rho^2}}-1\right)^{-1}$ and $\frac{2\sqrt{1-\rho^2}}{1-|\rho|} + \sqrt{1-2\vartheta} \leq \frac{3-|\rho|}{1-|\rho|}\sqrt{1-\vartheta}$, which is equivalent to

$$\sqrt{r} = 2\sqrt{\frac{1-\vartheta}{1-\rho^2}} \geq \max\left\{1+\sqrt{1-\vartheta}, \; 1 + \frac{1+|\rho|}{2}\sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{1-|\rho|}{2}\sqrt{\frac{1-2\vartheta}{1-\rho^2}}\right\}$$

*Second*, we study the curve $\sqrt{r} = 1+\sqrt{1-\vartheta}$, which is given by $(\lambda'+t')^2 = \vartheta + (\sqrt{r}-\lambda'-t')^2_+ = 1$. We also need $f_1, \vartheta + f_2, \vartheta + f_3 \geq 1$ and one more equality. When $\rho \geq 0$, we have discussed two cases; now we discuss the two old cases with the additional requirement $2\vartheta + f_4 \geq 1$, and two additional cases.

If $\frac{1}{1-\rho^2}\left[(1+|\rho|)\lambda' + (1-\rho^2)t'\right]^2 = (\lambda' + t')^2 = \vartheta + (\sqrt{r} - \lambda' - t')^2_+ = 1$ and other terms are greater than one, we have already considered this in a previous section. Now we only add the requirement $2\vartheta + \frac{1}{1-\rho^2}\left[(1-\rho^2)\sqrt{r} - t'(1-\rho^2) - \lambda'(1+|\rho|)\right]^2 \geq 1$. The final requirement on $\vartheta$ is $\sqrt{1-\vartheta} \leq \frac{1-|\rho|}{1+|\rho|}$ and $(1 - \sqrt{1-\rho^2}) + \sqrt{1-2\vartheta} \leq \sqrt{1-\rho^2}\sqrt{1-\vartheta}$.

If $(\lambda' + t')^2 = \vartheta + (\sqrt{r} - \lambda' - t')^2_+ = \vartheta + \frac{1}{1-\rho^2}\left[(1-\rho^2)\sqrt{r} - t'(1-\rho^2) - \lambda'(1-|\rho|)\right]^2 = 1$ and other terms are greater than one, we have already considered this in a previous section. Now we only add the requirement $2\vartheta + \frac{1}{1-\rho^2}\left[(1-\rho^2)\sqrt{r} - t'(1-\rho^2) - \lambda'(1+|\rho|)\right]^2 \geq 1$. The final requirement on $\vartheta$ is $\frac{1-|\rho|}{1+|\rho|} \leq \sqrt{1-\vartheta} \leq \left(\frac{2}{\sqrt{1-\rho^2}} - 1\right)^{-1}$ and $\sqrt{\frac{1-2\vartheta}{1-\vartheta}} \leq \frac{2\sqrt{1-\rho^2}-(1+|\rho|)}{1-|\rho|}$

If $\vartheta + \frac{1}{1-\rho^2}\left[\lambda'(1-|\rho|) + t'(1-\rho^2)\right]^2 = (\lambda' + t')^2 = \vartheta + (\sqrt{r} - \lambda' - t')^2_+ = 1$ and other terms are greater than one, this is a new case, and we have $\lambda' = \frac{1+|\rho|}{|\rho|}\left(1 - \sqrt{\frac{1-\vartheta}{1-\rho^2}}\right)$, $t' = 1 - \lambda'$. $\lambda', t' \geq 0$ requires $\frac{\sqrt{1-\rho^2}}{1+|\rho|} \leq \sqrt{1-\vartheta} \leq \sqrt{1-\rho^2}$, and the requirements from $f_1, \vartheta + f_3, 2\vartheta + f_4 \geq 1$ are $\sqrt{1-\vartheta} \leq \min\left\{\frac{2\sqrt{1-\rho^2}-(1-|\rho|)}{1+|\rho|}, \left(\frac{2}{\sqrt{1-\rho^2}} - 1\right)^{-1}\right\}$ and $\frac{1+|\rho|}{1-|\rho|}\sqrt{1-\rho^2} + \sqrt{1-2\vartheta} \leq \left(\frac{1+|\rho|}{1-|\rho|} + \sqrt{1-\rho^2}\right)\sqrt{1-\vartheta}$. Taking the intersection, the overall requirement on $\vartheta$ is

$$\frac{\sqrt{1-\rho^2}}{1+|\rho|} \leq \sqrt{1-\vartheta} \leq \left(\frac{2}{\sqrt{1-\rho^2}} - 1\right)^{-1} \text{ and } \frac{1+|\rho|}{1-|\rho|}\sqrt{1-\rho^2} + \sqrt{1-2\vartheta} \leq \left(\frac{1+|\rho|}{1-|\rho|} + \sqrt{1-\rho^2}\right)\sqrt{1-\vartheta}$$

If $(\lambda' + t')^2 = \vartheta + (\sqrt{r} - \lambda' - t')^2_+ = 2\vartheta + \frac{1}{1-\rho^2}\left[(1-\rho^2)\sqrt{r} - t'(1-\rho^2) - \lambda'(1+|\rho|)\right]^2 = 1$ and other terms are greater than one, then we have $\lambda' = \frac{1-|\rho|}{|\rho|}\left(\sqrt{1-\vartheta} - \sqrt{\frac{1-2\vartheta}{1-\rho^2}}\right)$, $t' = 1 - \lambda'$. The requirements from $\lambda', t' \geq 0$ and $f_1, \vartheta + f_2, \vartheta + f_3 \geq 1$ are:

$$\sqrt{\frac{1-2\vartheta}{1-\vartheta}} \leq \frac{2\sqrt{1-\rho^2}-(1+|\rho|)}{1-|\rho|}$$

$$\frac{|\rho|\sqrt{1-\rho^2}}{1-|\rho|} + \sqrt{1-2\vartheta} \geq \sqrt{1-\rho^2} \cdot \sqrt{1-\vartheta}$$

$$(1 - \sqrt{1-\rho^2}) + \sqrt{1-2\vartheta} \leq \sqrt{1-\rho^2} \cdot \sqrt{1-\vartheta}$$

$$\frac{1+|\rho|}{1-|\rho|}\sqrt{1-\rho^2} + \sqrt{1-2\vartheta} \geq \left(\frac{1+|\rho|}{1-|\rho|} + \sqrt{1-\rho^2}\right)\sqrt{1-\vartheta}$$

Taking intersection, the first and the last can imply the other two, so the final requirements are just the first and the last one.

Taking the union over all the four cases of $\sqrt{r} = 1 + \sqrt{1-\vartheta}$, we have $\sqrt{1-\vartheta} \leq \left(\frac{2}{\sqrt{1-\rho^2}} - 1\right)^{-1}$ and $\sqrt{\frac{1-2\vartheta}{1-\vartheta}} \leq \frac{2\sqrt{1-\rho^2}-(1+|\rho|)}{1-|\rho|}$, which is equivalent to

$$\sqrt{r} = 1 + \sqrt{1-\vartheta} \geq \max\left\{2\sqrt{\frac{1-\vartheta}{1-\rho^2}}, 1 + \frac{1+|\rho|}{2}\sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{1-|\rho|}{2}\sqrt{\frac{1-2\vartheta}{1-\rho^2}}\right\}$$

*Third*, we study the curve $\sqrt{r} = \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{\sqrt{1-\rho^2}}$ given by $\frac{1}{1-\rho^2}\left[(1+|\rho|)\lambda' + (1-\rho^2)t'\right]^2 = 2\vartheta + \frac{1}{1-\rho^2}\left[(1-\rho^2)\sqrt{r} - t'(1-\rho^2) - \lambda'(1+|\rho|)\right]^2 = 1$. We need one more equality constraint and other terms greater than one. We will see this curve does not exist for ant $\vartheta$ at all.

If the additional equality is $(\lambda' + t')^2 = 1$, then we have $\lambda' = \frac{\sqrt{1-\rho^2}-(1-\rho^2)}{|\rho|(1+|\rho|)}$ and $t' = \frac{(1+|\rho|)-\sqrt{1-\rho^2}}{|\rho|(1+|\rho|)}$. However, this case admits no $\vartheta$, because $\vartheta + \frac{1}{1-\rho^2}\left[\lambda'(1-|\rho|) + t'(1-\rho^2)\right]^2 \geq 1$

requires $\sqrt{1-\vartheta} \leq \frac{2\sqrt{1-\rho^2}-(1-|\rho|)}{1+|\rho|}$ and $\vartheta + (\sqrt{r} - \lambda' - t')_+^2 \geq 1$ requires $(1 - \sqrt{1-\rho^2}) + \sqrt{1-2\vartheta} \geq \sqrt{1-\rho^2} \cdot \sqrt{1-\vartheta}$; these two requirements have no intersection.

If the additional equality is $\vartheta + \frac{1}{1-\rho^2}\left[\lambda'(1-|\rho|) + t'(1-\rho^2)\right]^2 = 1$, then we have $\lambda' = \frac{1-\rho^2}{2|\rho|}\frac{1}{\sqrt{1-\rho^2}}\left(1 - \sqrt{1-\vartheta}\right)$ and $t' = \frac{1}{\sqrt{1-\rho^2}} - \frac{\lambda'}{1-|\rho|}$. However, this case admits no $\vartheta$, because $t' \geq 0$ requires $\sqrt{1-\vartheta} > \frac{1-|\rho|}{1+|\rho|}$ and $\vartheta + \frac{1}{1-\rho^2}\left[(1-\rho^2)\sqrt{r} - t'(1-\rho^2) - \lambda'(1-|\rho|)\right]^2 \geq 1$ requires $1 + \sqrt{1-2\vartheta} \geq 2\sqrt{1-\vartheta}$; these two requirements have no intersection.

If the additional equality is $\vartheta + (\sqrt{r} - \lambda' - t')_+^2 = 1$, then $\lambda' = \frac{1-|\rho|}{|\rho|}\left(\sqrt{1-\vartheta} - \sqrt{\frac{1-2\vartheta}{1-\rho^2}}\right)$ and $t' = \frac{1}{\sqrt{1-\rho^2}} - \frac{\lambda'}{1-|\rho|}$. However, this case admits no $\vartheta$, because $(\lambda'+t')^2 \geq 2$ requires $(1-\sqrt{1-\rho^2}) + \sqrt{1-2\vartheta} \geq \sqrt{1-\rho^2}\sqrt{1-\vartheta}$ and $\vartheta + \frac{1}{1-\rho^2}\left[(1-\rho^2)\sqrt{r} - t'(1-\rho^2) - \lambda'(1-|\rho|)\right]^2 \geq 1$ requires $\sqrt{\frac{1-2\vartheta}{1-\vartheta}} \leq \frac{2\sqrt{1-\rho^2}-(1+|\rho|)}{1-|\rho|}$; these two requirements have no intersection.

If the additional equality is $\vartheta + \frac{1}{1-\rho^2}\left[(1-\rho^2)\sqrt{r} - t'(1-\rho^2) - \lambda'(1-|\rho|)\right]^2 = 1$, then $\lambda' = \frac{1-\rho^2}{2|\rho|}\frac{1}{\sqrt{1-\rho^2}}\left(\sqrt{1-\vartheta} - \sqrt{1-2\vartheta}\right)$ and $t' = \frac{1}{\sqrt{1-\rho^2}} - \frac{\lambda'}{1-|\rho|}$. However, this case admits no $\vartheta$, because $t' \geq 0$, $(\lambda' + t')^2 \geq 1$ and $\vartheta + \frac{1}{1-\rho^2}\left[\lambda'(1-|\rho|) + t'(1-\rho^2)\right]^2 \geq 1$ respectively requires

$$\frac{2|\rho|}{1+|\rho|} + \sqrt{1-2\vartheta} \geq \sqrt{1-\vartheta}$$

$$\frac{2(1-\sqrt{1-\rho^2})}{1+|\rho|} + \sqrt{1-2\vartheta} \geq \sqrt{1-\vartheta}$$

$$1 + \sqrt{1-2\vartheta} \geq 2\sqrt{1-\vartheta}$$

These three requirements admit no $\vartheta \in (0,1)$.

To sum up, the *third* surve does not show up in the phase diagram.

*Fourth*, we study the curve $\sqrt{r} = \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{2(1-|\rho|)-\sqrt{(1-\vartheta)(1-\rho^2)}}{(1-|\rho|)^2}$ given by

$$(\lambda'+t')^2 = \vartheta + \frac{1}{1-\rho^2}\left[\lambda'(1-|\rho|) + t'(1-\rho^2)\right]^2 = 2\vartheta + \frac{1}{1-\rho^2}\left[(1-\rho^2)\sqrt{r} - t'(1-\rho^2) - \lambda'(1+|\rho|)\right]^2 = 1.$$

We have $\lambda' = \frac{1+|\rho|}{|\rho|}\left(1 - \sqrt{\frac{1-\vartheta}{1-\rho^2}}\right)$ and $t' = 1 - \lambda'$. $\lambda', t' \geq 0$ requires $\frac{\sqrt{1-\rho^2}}{1+|\rho|} \leq \sqrt{1-\vartheta} \leq \sqrt{1-\rho^2}$, and we also need the requirements from $f_1, \vartheta + f_3 \geq 1$. The overall requirement on $\vartheta$ is:

$$\frac{2\sqrt{1-\rho^2}-(1-|\rho|)}{1+|\rho|} \leq \sqrt{1-\vartheta} \leq \sqrt{1-\rho^2}$$

$$\frac{1+|\rho|}{1-|\rho|}\sqrt{1-\rho^2} + \sqrt{1-2\vartheta} \geq \left(\frac{1+|\rho|}{1-|\rho|} + \sqrt{1-\rho^2}\right)\sqrt{1-\vartheta}$$

$$\frac{2\sqrt{1-\rho^2}}{1-|\rho|} + \sqrt{1-2\vartheta} \geq \frac{3-|\rho|}{1-|\rho|}\sqrt{1-\vartheta}$$

This is also an empty set, and this curve can never be present in the phase diagram.

To prove this, we note that $\phi(\vartheta) = \frac{2\sqrt{1-\rho^2}}{1-|\rho|} + \sqrt{1-2\vartheta} - \frac{3-|\rho|}{1-|\rho|}\sqrt{1-\vartheta}$ is a "first increasing, then decreasing" function of $\vartheta \in [0, \frac{1}{2}]$. When $|\rho| \geq \frac{1+2\sqrt{2}}{7}$, the maximum of $\frac{2\sqrt{1-\rho^2}}{1-|\rho|} + \sqrt{1-2\vartheta} - \frac{3-|\rho|}{1-|\rho|}\sqrt{1-\vartheta}$ is not positive, and does not admit a curve. When $|\rho| < \frac{1+2\sqrt{2}}{7}$, we add the requirement of $\sqrt{1-\vartheta} \geq \frac{2\sqrt{1-\rho^2}-(1-|\rho|)}{1+|\rho|}$, even the largest $\vartheta$ is still on the left side of the peak of the maximum point of $\phi(\vartheta)$ and still makes it negative.

*Fifth*, we study the curve $\sqrt{r} = \left(\frac{2}{1+|\rho|} + \frac{1}{\sqrt{1-\rho^2}}\right)\sqrt{1-\vartheta} - \frac{1-|\rho|}{1+|\rho|}\sqrt{\frac{1-2\vartheta}{1-\rho^2}}$, which is given by

$$\vartheta + \frac{1}{1-\rho^2}\left[\lambda'(1-|\rho|) + t'(1-\rho^2)\right]^2 = \vartheta + (\sqrt{r} - \lambda' - t')_+^2 = 2\vartheta + \frac{1}{1-\rho^2}\left[(1-\rho^2)\sqrt{r} - t'(1-\rho^2) - \lambda'(1+|\rho|)\right]^2 = 1.$$

Now we have $\lambda' = \frac{1-|\rho|}{|\rho|}\left(\sqrt{1-\vartheta} - \sqrt{\frac{1-2\vartheta}{1-\rho^2}}\right)$ and $t' = \frac{1}{\sqrt{1-\rho^2}} - \frac{\lambda'}{1-|\rho|}$. The requirements of $\lambda', t' \geq 0$ and $f_1, \vartheta + f_3 \geq 0$ are

$$\sqrt{1-\rho^2} - \frac{|\rho|(1+|\rho|)}{1-|\rho|} \leq \sqrt{\frac{1-2\vartheta}{1-\vartheta}} \leq \frac{2\sqrt{1-\rho^2} - (1+|\rho|)}{1-|\rho|}$$

$$\frac{1+|\rho|}{2} + \sqrt{1-2\vartheta} \leq \left(\frac{1+|\rho|}{2} + \sqrt{1-\rho^2}\right)\sqrt{1-\vartheta}$$

$$\frac{1+|\rho|}{1-|\rho|}\sqrt{1-\rho^2} + \sqrt{1-2\vartheta} \leq \left(\frac{1+|\rho|}{1-|\rho|} + \sqrt{1-\rho^2}\right)\sqrt{1-\vartheta}$$

This may not be an empty set. However, even when it is not an empty set, the curve is actually either greater than $\sqrt{r} = 1 + \sqrt{1-\vartheta}$ or $\sqrt{r} = 2\sqrt{\frac{1-\vartheta}{1-\rho^2}}$. To see this:

- In terms of the existence of $\sqrt{r} = 1 + \sqrt{1-\vartheta}$ and $\sqrt{r} = 2\sqrt{\frac{1-\vartheta}{1-\rho^2}}$: When $\frac{1-|\rho|}{1+|\rho|} \leq \sqrt{1-\vartheta} \leq \left(\frac{2}{\sqrt{1-\rho^2}} - 1\right)^{-1}$ and $\sqrt{\frac{1-2\vartheta}{1-\vartheta}} \leq \frac{2\sqrt{1-\rho^2} - (1+|\rho|)}{1-|\rho|}$, we have proven that $\sqrt{r} = 1 + \sqrt{1-\vartheta}$ is one segment of the phase diagram. When $\sqrt{1-\vartheta} \geq \left(\frac{2}{\sqrt{1-\rho^2}} - 1\right)^{-1}$ and $\sqrt{\frac{1-2\vartheta}{1-\vartheta}} \leq \frac{2\sqrt{1-\rho^2} - (1+|\rho|)}{1-|\rho|}$, we have proven that $\sqrt{r} = 2\sqrt{\frac{1-\vartheta}{1-\rho^2}}$ is one segment of the phase diagram. We can prove that

$$\sqrt{1-\rho^2} - \frac{|\rho|(1+|\rho|)}{1-|\rho|} \leq \sqrt{\frac{1-2\vartheta}{1-\vartheta}} \leq \frac{2\sqrt{1-\rho^2} - (1+|\rho|)}{1-|\rho|}$$

$$\implies \begin{cases} \text{either } \frac{1-|\rho|}{1+|\rho|} \leq \sqrt{1-\vartheta} \leq \left(\frac{2}{\sqrt{1-\rho^2}} - 1\right)^{-1} \text{ and } \sqrt{\frac{1-2\vartheta}{1-\vartheta}} \leq \frac{2\sqrt{1-\rho^2} - (1+|\rho|)}{1-|\rho|} \\ \text{or } \sqrt{1-\vartheta} \geq \left(\frac{2}{\sqrt{1-\rho^2}} - 1\right)^{-1} \text{ and } \sqrt{\frac{1-2\vartheta}{1-\vartheta}} \leq \frac{2\sqrt{1-\rho^2} - (1+|\rho|)}{1-|\rho|} \end{cases}$$

so one of $\sqrt{r} = 1 + \sqrt{1-\vartheta}$ and $\sqrt{r} = 2\sqrt{\frac{1-\vartheta}{1-\rho^2}}$ exists as long as the *fifth* curve exists.

- In the latter case, it is greater than $\sqrt{r} = 2\sqrt{\frac{1-\vartheta}{1-\rho^2}}$ which exists in the same region. (can be easily verified)

- In the former case, we can assume $|\rho| \leq 3 - 2\sqrt{2}$ because we need $\sqrt{1-\vartheta} \geq \frac{1-|\rho|}{1+|\rho|}$ to hold for some $\vartheta \in (0, \frac{1}{2})$. Using $|\rho| \leq 3 - 2\sqrt{2}$, we can prove $\sqrt{\frac{1-2\vartheta}{1-\vartheta}} \geq \sqrt{1-\rho^2} - \frac{|\rho|(1+|\rho|)}{1-|\rho|} \implies \sqrt{1-\vartheta} \geq \frac{1-|\rho|}{1+|\rho|}$. Now we can easily verify the curve is greater than $\sqrt{r} = 1 + \sqrt{1-\vartheta}$ which exists in the same region

As a result, this curve does not play a part in the final phase diagram either.

*Sixth*, we study the curve $\sqrt{r} = 1 + \frac{1+|\rho|}{2}\sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{1-|\rho|}{2}\sqrt{\frac{1-2\vartheta}{1-\rho^2}}$, given by

$$(\lambda' + t')^2 = \vartheta + \frac{1}{1-\rho^2}\left[(1-\rho^2)\sqrt{r} - t'(1-\rho^2) - \lambda'(1-|\rho|)\right]^2 = 2\vartheta + \frac{1}{1-\rho^2}\left[(1-\rho^2)\sqrt{r} - t'(1-\rho^2) - \lambda'(1+|\rho|)\right]^2 =$$

We get $\lambda' = \frac{1-\rho^2}{2|\rho|}\frac{1}{\sqrt{1-\rho^2}}\left(\sqrt{1-\vartheta} - \sqrt{1-2\vartheta}\right)$ and $t' = 1 - \lambda'$. The requirements from $t' \geq 0$, $\frac{1}{1-\rho^2}\left[(1+|\rho|)\lambda' + (1-\rho^2)t'\right]^2 \geq 1$, $\vartheta + \frac{1}{1-\rho^2}\left[\lambda'(1-|\rho|) + t'(1-\rho^2)\right]^2 \geq 1$ and $\vartheta + (\sqrt{r} - $

$\lambda' - t')_+^2 \geq 1$ are respectively

$$\frac{2|\rho|}{\sqrt{1-\rho^2}} + \sqrt{1-2\vartheta} \geq \sqrt{1-\vartheta}$$

$$\frac{2(1-\sqrt{1-\rho^2})}{1+|\rho|} + \sqrt{1-2\vartheta} \leq \sqrt{1-\vartheta}$$

$$\frac{2\sqrt{1-\rho^2}}{1-|\rho|} + \sqrt{1-2\vartheta} \geq \frac{3-|\rho|}{1-|\rho|}\sqrt{1-\vartheta}$$

$$\sqrt{\frac{1-2\vartheta}{1-\vartheta}} \geq \frac{2\sqrt{1-\rho^2}-(1+|\rho|)}{1-|\rho|}$$

Taking the intersection, the last two inequalities can imply the rest, and it is equivalent to

$$\sqrt{r} = 1 + \frac{1+|\rho|}{2}\sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{1-|\rho|}{2}\sqrt{\frac{1-2\vartheta}{1-\rho^2}} \geq \max\left\{1+\sqrt{1-\vartheta}, \ 2\sqrt{\frac{1-\vartheta}{1-\rho^2}}\right\}.$$

Now we have studied all the curves for $\rho < 0$. To sum up, the phase curve is

$$\sqrt{r} = \max\left\{1+\sqrt{1-\vartheta}, \ 2\sqrt{\frac{1-\vartheta}{1-\rho^2}}, \ 1 + \frac{1+|\rho|}{2}\sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{1-|\rho|}{2}\sqrt{\frac{1-2\vartheta}{1-\rho^2}}\right\}$$

### F.1 Proof of Lemma 3

Recall the optimization in (16); the solution $b = (b_1, b_2)$ has to set the sub-gradient of the objective function to zero. As a result, the equation of the sub-gradient for $b = (b_1, b_2)$ is:

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \lambda'\begin{bmatrix} \text{sgn}(b_1) \\ \text{sgn}(b_2) \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}$$

Now we begin to find out the solution path. Thresholded Lasso has two steps: First, we run Lasso to select variables from $(x_j, x_{j+1})$; second, a thresholding step with $t = t'\sqrt{2\log(p)}$ is further performed, and the surviving variables of the two steps are the finally selected ones. Also note that we have required $\rho \geq 0$.

First, we study the behavior of Lasso, and decrease $\lambda'$ from a sufficiently large value to see when the variables enter the model. We assume $h_1 > 0$ and $0 < |h_2| < h_1$.

The procedure is just setting $\mu = 0$ in the proof of Lemma B.1, and we summarise the results below:

- When $\lambda' \geq h_1$, we have $\hat{b}_1 = \hat{b}_2 = 0$.

- If $h_2 \geq \rho h_1$, when $\frac{h_2-\rho h_1}{1-\rho} \leq \lambda' < h_1$, we have $\hat{b}_1 = h_1 - \lambda'$, and $\hat{b}_2 = 0$;

  When $\lambda' < \frac{h_2-\rho h_1}{1-\rho}$, we have

  $$\hat{b}_1 = \frac{(h_1-\rho h_2)-(1-\rho)\lambda'}{1-\rho^2}, \qquad \hat{b}_2 = \frac{(h_2-\rho h_1)-(1-\rho)\lambda'}{1-\rho^2};$$

- if $h_2 < \rho h_1$, when $\frac{-h_2+\rho h_1}{1+\rho} \leq \lambda' < h_1$, we have $\hat{b}_1 = h_1 - \lambda'$, and $\hat{b}_2 = 0$;

  When $\lambda' < \frac{-h_2+\rho h_1}{1+\rho}$, we have

  $$\hat{b}_1 = \frac{(h_1-\rho h_2)-(1+\rho)\lambda'}{1-\rho^2}, \quad \hat{b}_2 = \frac{(h_2-\rho h_1)+(1+\rho)\lambda'}{1-\rho^2}.$$

When $\lambda' \geq h_1$, for any $t'$, we have $\hat{b}_1 = \hat{b}_2 = 0$.

For $h_2 \geq \rho h_1$, when $\frac{h_2 - \rho h_1}{1-\rho} \leq \lambda' < h_1$, if $h_1 \leq \lambda' + t'$, then we still selected neither of $(x_j, x_{j+1})$ in the end. If $h_1 > \lambda' + t'$, then we will select only $x_j$. When $\lambda' < \frac{h_2 - \rho h_1}{1-\rho}$, we have $\hat{b}_1 > \hat{b}_2 > 0$ and it depends on whether $t \geq \hat{b}_1$, $\hat{b}_2 \leq t < b_1$ or $\hat{b}_2 > t$ how $(x_j, x_{j+1})$ are selected in the end.

For $h_2 < \rho h_1$, when $\frac{-h_2 + \rho h_1}{1+\rho} \leq \lambda' < h_1$, if $h_1 \leq \lambda' + t'$, then we still selected neither of $(x_j, x_{j+1})$ in the end. If $h_1 > \lambda' + t'$, then we will select only $x_j$. When $\lambda' < \frac{-h_2 + \rho h_1}{1-\rho}$, we have $\hat{b}_1 > -\hat{b}_2 > 0$ and it depends on whether $t \geq \hat{b}_1$, $-\hat{b}_2 \leq t < b_1$ or $-\hat{b}_2 > t$ how $(x_j, x_{j+1})$ are selected in the end.

## G  PROOF OF THEOREM 5 (FORWARD SELECTION)

The proof for Forward Selection still consists of three parts: (a) deriving the rejection region, (b) obtaining the rate of convergence of $\mathbb{E}[H(\hat{\beta}, \beta)]$, and (c) calculating the phase diagram.

For Forward Selection, we first write $X = [x_1, x_2, \ldots, x_p]$, where $x_i \in \mathbb{R}^n$ for $1 \leq i \leq p$. For any subset $A \subset \{1, 2, \ldots, p\}$, let $P_A^\perp$ be the projection onto the orthogonal complement of the linear space spanned by $\{x_i : i \in A\}$. Before the first part, we formally define Forward Selection in Algorithm 1.

---

**Algorithm 1** forward selection

---

1: Input $X$ and $y$ (generated with our own setting.)
2: Fix $t > 0$.
3: Initialize $S^{(0)} = \emptyset, \hat{\beta}^{(0)} = 0, \hat{r}^{(0)} = y$.
4: Initialize $k = 0$.
5: **while** true **do**                                                                    ▷ forward step
6:     $k \leftarrow k + 1$
7:     $\hat{r}^{(k-1)} \leftarrow P_{k-1}^\perp y$        ▷ $\hat{r}^{(k-1)}$ is the residual of the OLS fit of $Y$ onto $X_{S^{(k-1)}}$.
8:     $i^* \leftarrow \arg\max_{i \notin S^{(k-1)}} |x_i' \hat{r}^{(k-1)}|$
9:     $\delta^+ \leftarrow \frac{|x_{i^*} \hat{r}^{(k-1)}|}{\|P_{k-1}^\perp x_{i^*}\|}$     ▷ the forward gain, equivalent to the decrease in the loss function.
10:     **if** $\delta^+ \leq t$ **then**
11:         Break.
12:     **end if**
13:     $S^{(k)} \leftarrow S^{(k-1)} \cup \{i^*\}$        ▷ Note this step is after checking $\delta^+ \leq t$
14: **end while**
15: $k \leftarrow k - 1$
             ▷ Because the $k$th variable hasn't been added when the "while" loop is broken.
16: $\hat{\beta} = \hat{\beta}^{ols}(S^{(k)})$        ▷ $S^{(k)}$ is the set of selected variables.

---

**Remark 3.** *The stopping rule is equivalent to measuring the decrease in the residual sum of squares. To see this, suppose $i \in \{1, 2, \ldots, p\}$ is enrolled at step $k$, ans $S^{(k)} = S^{(k-1)} \cup \{i\}$. Then $\|y - X_{S^{(k)}} \hat{\beta}^{ols}(S^{(k)})\|^2 = \|P_k^\perp y\|^2$ and $\|y - X_{S^{(k-1)}} \hat{\beta}^{ols}(S^{(k-1)})\|^2 = \|P_{k-1}^\perp y\|^2$. By adding variable $i$ into $S^{(k-1)}$, the decrease $\|P_{k-1}^\perp y\|^2 - \|P_k^\perp y\|^2$ is equal to the squared norm of the projection of $P_{k-1}^\perp y$ onto the direction of $P_{k-1}^\perp x_i$, which is $\left( \frac{|x_i' \hat{r}^{(k-1)}|}{\|P_{k-1}^\perp x_i\|} \right)^2$ where $\hat{r}^{(k-1)} = P_{k-1}^\perp y$.*

**Part 1: Deriving the rejection region.** Forward selection is a sequential method, and we first need to show it can be decomposed into bivariate sub-problems. The main reason is that whether some variable $x_j$ is selected in the end only depends on $(x_j, x_{j+1})$, and has nothing to do with other variables, or the number of steps $k$. We still use $(x_j, x_{j+1})$ to denote an arbitrary pair of correlated variables:

In terms of forward gain, whenever $x_{j+1}$ is not in $\hat{S}^{(k-1)}$ for arbitrary $k$, $P_{k-1}^\perp x_j = x_j$, and $\frac{|x_j' \hat{r}^{(k-1)}|}{\|P_{k-1}^\perp x_j\|} = |x_j' y| = |h_1| \sqrt{2 \log(p)}$. When $x_{j+1}$ is already in $\hat{S}^{(k-1)}$ for arbitrary $k$, $P_{k-1}^\perp x_j = x_j - \rho x_{j+1}$, and $\frac{|x_j' \hat{r}^{(k-1)}|}{\|P_{k-1}^\perp x_j\|} = \frac{|(x_j' - \rho x_{j+1}')y|}{\sqrt{1-\rho^2}} = \frac{|h_1 - \rho h_2|}{\sqrt{1-\rho^2}} \sqrt{2 \log(p)}$.

In terms of the entry rule ("$i^* \leftarrow \arg\max_{i \notin S^{(k-1)}} |x_i' \hat{r}^{(k-1)}|$" in Algorithm 1), since $x_j' \hat{r}^{(k-1)} = x_j' P_{k-1}^{\perp} y$, it is still $x_j' y$ or $(x_j' y - \rho x_{j+1} y)$ depending on whether $x_{j+1} \in S^{(k-1)}$. It has nothing to do with specific $k$ or other variables than $(x_j, x_{j+1})$.

As a result, Algorithm 1 under the block-wise diagonal design can be viewed as many bivariate sub-problems going on simultaneously for each pair of correlated variables. In each "while" loop, variables from different pairs may be enrolled, but (i) the order of $(x_j, x_{j+1})$ does not depend on $k$ or other variables, and (ii) the bivariate problem must have terminated when the whole algorithm terminates, and the result of the bivariate problem does not depend on $k$ or other variables.

Of course, we have assumed that Algorithm 1 will always terminate in finite steps, which is true, because each bivariate sub-problem always terminates as we will see in the proof of Lemma G.1.

Working on a bivariate problem, we can scale everything down by $\sqrt{2\log(p)}$ and define $t' = t/\sqrt{2\log(p)}$. Then the solution path can be described in Lemma G.1.

**Lemma G.1** (Solution path of Forward Selection). *Consider the bivariate problem of running Algorithm 1 with $y$ and $(x_j, x_{j+1})$. Suppose $h_1 > |h_2| \geq 0$, and $\rho > 0$.*

- *When $t' \geq h_1$, none of $(x_j, x_{j+1})$ will get selected when the algorithm ends.*

- *If $-h_1 < h_2 \leq (\rho - \sqrt{1-\rho^2})h_1$, when $t' < h_1$, both $(x_j, x_{j+1})$ will get selected when the algorithm ends.*

- *If $(\rho - \sqrt{1-\rho^2})h_1 < h_2 < h_1$, when $t \in \left[ \frac{|h_2 - \rho h_1|}{\sqrt{1-\rho^2}}, h_1 \right)$, only $x_j$ is in the model before the algorithm ends.*

- *If $(\rho - \sqrt{1-\rho^2})h_1 < h_2 < h_1$, when $t < \frac{|h_2 - \rho h_1|}{\sqrt{1-\rho^2}}$, both $(x_j, x_{j+1})$ will get selected when the algorithm ends.*

*Proof of Lemma G.1.* Note that we have required $\rho > 0$ to avoid unnecessary discussion. Since $h_1 > |h_2|$, at any step $k$ when neither of $(x_j, x_{j+1})$ is in the model, we have $|x_j' \hat{r}_{k-1}| = |h_1| > |x_{j+1}' \hat{r}_{k-1}| = |h_2|$. As a result, if $t \geq h_1$, then the algorithm will terminate without selecting either of $(x_j, x_{j+1})$. If $t > h_1$, it will select $x_j$ at some step and proceed to the next "while" loop.

After $x_j$ has been selected, if $h_2 \geq (\rho + \sqrt{1-\rho^2})h_1$ or $h_2 \leq (\rho - \sqrt{1-\rho^2})h_1$, we have $\frac{|x_{j+1}' \hat{r}^{(k-1)}|}{\|P_{k-1}^{\perp} x_{j+1}\|} = \frac{|h_2 - \rho h_1|}{\sqrt{1-\rho^2}} \geq h_1 > t$, and $x_{j+1}$ will be selected at some later step. However, since $\rho > 0$, we have $\rho + \sqrt{1-\rho^2} > 1$, so we can have $\frac{|x_{j+1}' \hat{r}^{(k-1)}|}{\|P_{k-1}^{\perp} x_{j+1}\|} = \frac{|h_2 - \rho h_1|}{\sqrt{1-\rho^2}} \geq h_1$ only when $h_2 \geq (\rho - \sqrt{1-\rho^2})h_1$.

If $(\rho - \sqrt{1-\rho^2})h_1 < h_2 < h_1$, we have $\frac{|h_2 - \rho h_1|}{\sqrt{1-\rho^2}} < h_1$. When $t < \frac{|h_2 - \rho h_1|}{\sqrt{1-\rho^2}}$, both of the variables will be selected; when $t \in \left[ \frac{|h_2 - \rho h_1|}{\sqrt{1-\rho^2}}, h_1 \right)$, only $x_j$ will be selected. □

We use Lemma G.1 to write down the rejection region, as in Figure 12, still for $\rho > 0$.

$$\mathcal{R} = \{(h_1, h_2) : h_1 - \rho h_2 > t'\sqrt{1-\rho^2}, h_1 > \frac{t'\sqrt{1-\rho^2}}{1-\rho}\}$$

$$\cup \{(h_1, h_2) : h_1 > t', h_1 > h_2\} \cup \{(h_1, h_2) : h_2 < -t, h_1 - \rho h_2 > t'\sqrt{1-\rho^2}\}$$

$$\cup \{(h_1, h_2) : -h_1 + \rho h_2 > t'\sqrt{1-\rho^2}, h_1 < -\frac{t'\sqrt{1-\rho^2}}{1-\rho}\}$$

$$\cup \{(h_1, h_2) : h_1 < -t', h_1 < h_2\} \cup \{(h_1, h_2) : h_2 > t, -h_1 + \rho h_2 > t'\sqrt{1-\rho^2}\} \quad (50)$$

$$\rho = 0.5, \ t = 1.5$$
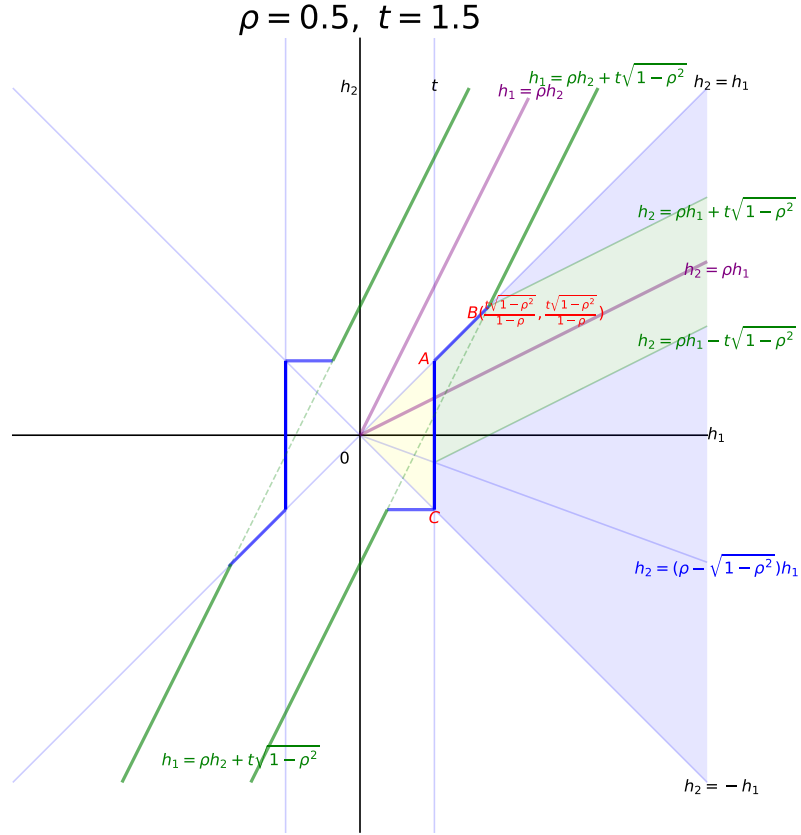


Figure 12: the rejection region of forward selection ($\rho \geq 0$)

**Part 2. Analyzing the Hamming error.**

**Theorem 14.** *Suppose the conditions of Theorem 5 holds. Let* $t' = t/\sqrt{2\log(p)}$ *and* $h_1 = x'_j y/\sqrt{2\log(p)}$, $h_2 = x'_{j+1} y/\sqrt{2\log(p)}$. *As* $p \to \infty$,

$$\mathrm{FP}_p = L_p p^{1 - \min\left\{t'^2, \ \vartheta + f_1(\sqrt{r}, t')\right\}}, \qquad \mathrm{FN}_p = L_p p^{1 - \min\left\{\vartheta + f_2(\sqrt{r}, t'), \ 2\vartheta + f_3(\sqrt{r}, t')\right\}},$$

*where (below,* $d_{|\rho|}^2(u, v)$ *is as in Definition 1),*

$$f_1(\sqrt{r}, t') = \begin{cases} (t' - |\rho|\sqrt{r})^2 & \text{if } \sqrt{r} \leq \frac{t'}{1+|\rho|} \\ \frac{1}{1-\rho^2} d_{|\rho|}^2((t', t'), (|\rho|\sqrt{r}, \sqrt{r})) & \text{if } \frac{t'}{1+|\rho|} < \sqrt{r} \leq \frac{2t'}{1+|\rho|} \\ \min\left\{\frac{1}{2}(1 - |\rho|)r, \ t'^2\right\} & \text{if } \sqrt{r} > \frac{2t'}{1+|\rho|} \end{cases}$$

$$f_2(\sqrt{r}, t') = \begin{cases} \min\left\{(\sqrt{r} - t')_+^2, \ \frac{1}{2}(1 - |\rho|)r\right\} & \text{if } \sqrt{r} \leq \frac{2t'}{\sqrt{1-\rho^2}} \\ \min\left\{(\sqrt{r} - t')_+^2, \ \frac{1}{1-\rho^2} d_{|\rho|}^2(B, (\sqrt{r}, |\rho|\sqrt{r}))\right\} & \text{if } \frac{2t'}{\sqrt{1-\rho^2}} < \sqrt{r} \leq \frac{t'\sqrt{1-\rho^2}}{|\rho|(1-|\rho|)} \\ \left[\sqrt{1-\rho^2}\sqrt{r} - t'\right]^2 & \text{if } \sqrt{r} > \frac{t'\sqrt{1-\rho^2}}{|\rho|(1-|\rho|)} \end{cases}$$

*The definition of* $f_3(\sqrt{r}, t')$ *depends on the sign of* $\rho$. *When* $\rho > 0$,

$$f_3(\sqrt{r}, t') = \left[\sqrt{1-\rho^2}\sqrt{r} - t\right]^2$$

*When* $\rho < 0$,

$$f_3(\sqrt{r}, t') = \min\left\{\left[\sqrt{1-\rho^2}\sqrt{r} - t\right]^2, \ d_{|\rho|}^2(C, ((1 - |\rho|)\sqrt{r}, -(1 - |\rho|)\sqrt{r}))\right\}$$

$$= \min\left\{\left[\sqrt{1-\rho^2}\sqrt{r} - t\right]^2, \ \frac{2}{1-|\rho|}\left[(1 - |\rho|)\sqrt{r} - t\right]^2\right\}$$

The calculation of the elliptical distances are easy given Lemma A.1.

**Part 3. Calculating the phase diagram.** The computation of the boundary between Alomst Full Recovery and No Recovery is almost the same for every method, so we omit the details and conclude that such boundary is $r = \vartheta$.

Then we set out to calculate the boundary between Alomst Full Recovery and Exact Recovery. As usual, we have four cases respectively for $\rho > 0$ and $\rho < 0$. However, unlike previous methods, forward selection is very easy, so we combine $\rho > 0$ and $\rho < 0$ in our discussion and use $|rho|$ all along.

When $\rho > 0$, the phase curve is

$$\sqrt{r} = \max\left\{ 1 + \sqrt{1-\vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-\rho}}, \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{\sqrt{1-\rho^2}} \right\}.$$

When $\rho < 0$, the phase curve is

$$\sqrt{r} = \max\left\{ 1 + \sqrt{1-\vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-|\rho|}}, \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{\sqrt{1-\rho^2}}, \sqrt{\frac{1-2\vartheta}{2(1-|\rho|)}} + \frac{1}{1-|\rho|} \right\}$$

*First*, if $t'^2 = \vartheta + f_2(\sqrt{r}, t') = 1$ and $\vartheta + f_1(\sqrt{r}, t') \geq 1$, $2\vartheta + f_3(\sqrt{r}, t') \geq 1$, we discuss the conditional expression of $f_2(\sqrt{r}, t')$:

When $\sqrt{r} \leq \frac{2t'}{\sqrt{1-\rho^2}}$ in $f_3(\sqrt{r}, t')$: Now $\sqrt{r} = \max\left\{ 1 + \sqrt{1-\vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-|\rho|}} \right\}$. We need the following requirements: First, $\sqrt{r} \leq \frac{2t'}{\sqrt{1-\rho^2}} = \frac{2}{\sqrt{1-\rho^2}}$ itself. It is not restrictive, because $1 + \sqrt{1-\vartheta} \leq 2 \leq \frac{2}{\sqrt{1-\rho^2}}$ and $\sqrt{\frac{2(1-\vartheta)}{1-|\rho|}} \leq \sqrt{\frac{2}{1-|\rho|}} \leq \frac{2}{\sqrt{1-\rho^2}}$. Second, $\vartheta + f_2(\sqrt{r}, t') \geq 1$, which is still not restrictive. When $\sqrt{r} \geq \frac{2t'}{1+|\rho|}$, we know $FP_2 = o(1)$ from $\sqrt{r} \geq \sqrt{\frac{2(1-\vartheta)}{1-|\rho|}}$ and $t' = 1$. When $1 < \sqrt{r} < \frac{2t'}{1+|\rho|}$, we know $FP_2 = o(1)$ from $d^2(B, (\sqrt{r}, |\rho|\sqrt{r})) \geq \frac{1}{2}(1+|\rho|)(1-|\rho|)^2 r$. Third, $2\vartheta + f_3(\sqrt{r}, t') \geq 1$, which requires $\sqrt{r} \geq \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{\sqrt{1-\rho^2}}$ when the correlation is positive, and $\sqrt{r} \geq \max\left\{ \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{\sqrt{1-\rho^2}}, \sqrt{\frac{1-2\vartheta}{2(1-|\rho|)}} + \frac{1}{1-|\rho|} \right\}$. when the correlation is negative.

When $\frac{2t'}{\sqrt{1-\rho^2}} < \sqrt{r} \leq \frac{t'\sqrt{1-\rho^2}}{|\rho|(1-|\rho|)}$ in $f_3(\sqrt{r}, t')$: If $(1-\vartheta)(1-\rho^2) = d^2(B, (\sqrt{r}, |\rho|\sqrt{r}))$, then since $d^2(B, (\sqrt{r}, |\rho|\sqrt{r})) \geq \frac{1}{2}(1+|\rho|)(1-|\rho|)^2 r$, we have $\sqrt{r} \leq \sqrt{\frac{2(1-\vartheta)}{1-|\rho|}}$. However, we also need $\sqrt{r} > \frac{2}{\sqrt{1-\rho^2}}$, which gives a contradiction. This case does not exist. If $\sqrt{r} = 1 + \sqrt{1-\vartheta}$, then it also contradicts $\sqrt{r} > \frac{2}{\sqrt{1-\rho^2}}$.

When $\sqrt{r} > \frac{t'\sqrt{1-\rho^2}}{|\rho|(1-|\rho|)}$ in $f_3(\sqrt{r}, t')$: $\sqrt{r} = \sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{1}{\sqrt{1-\rho^2}}$. It cannot meet the requirement $\sqrt{r} > \frac{t'\sqrt{1-\rho^2}}{|\rho|(1-|\rho|)}$, so this case does not exist.

To sum up, the *first* case gives

$$\sqrt{r} = \max\left\{ 1 + \sqrt{1-\vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-|\rho|}} \right\}.$$

which exists in the region:

- $\left(\frac{1}{2}, 1\right] \cup \left\{ \vartheta \leq \frac{1}{2} : \sqrt{r} \geq \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{\sqrt{1-\rho^2}} \right\}$ for $\rho > 0$.

- $\left(\frac{1}{2}, 1\right] \cup \left\{ \vartheta \leq \frac{1}{2} : \sqrt{r} \geq \max\left\{ \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{\sqrt{1-\rho^2}}, \sqrt{\frac{1-2\vartheta}{2(1-|\rho|)}} + \frac{1}{1-|\rho|} \right\} \right\}$ for $\rho < 0$.

*Second*, if $\vartheta + f_1(\sqrt{r}, t') = \vartheta + f_2(\sqrt{r}, t') = 1$ and $t' \geq 1$, $2\vartheta + f_3(\sqrt{r}, t') \geq 1$, then we will get nothing in this case. To see this, we first list a few requirements:

- We know that $t' \geq 1$;

- We know from $\vartheta + f_2(\sqrt{r}, t') = 1$ that $\sqrt{r} \geq \max\left\{t' + \sqrt{1-\vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-|\rho|}}\right\}$;

- We know from $2\vartheta + f_3(\sqrt{r}, t') \geq 1$ that, when $\vartheta \leq \frac{1}{2}$, we need

  - $\sqrt{r} \geq \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{t'}{\sqrt{1-\rho^2}}$ for positive correlation;

  - $\sqrt{r} \geq \max\left\{\sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{t'}{\sqrt{1-\rho^2}}, \sqrt{\frac{1-2\vartheta}{2(1-|\rho|)}} + \frac{t'}{1-|\rho|}\right\}$ for negative correlation.

From these requirements, even if this case does admit some curve, it can only be higher than the one in the previous *first* case, and exist in a smaller region.

*Third*, if $t'^2 = 2\vartheta + f_3(\sqrt{r}, t') = 1$ and $\vartheta + f_1(\sqrt{r}, t') \geq 1$, $\vartheta + f_2(\sqrt{r}, t') \geq 1$, then:

When the correlation is positive, we have only one possible curve $\sqrt{r} = \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{\sqrt{1-\rho^2}}$. From $\vartheta + f_2(\sqrt{r}, t') \geq 1$, we get the requirement $\sqrt{r} \geq \max\left\{1 + \sqrt{1-\vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-\rho}}\right\}$. For $\vartheta + f_1(\sqrt{r}, t') \geq 1$, since we already have $\sqrt{r} \geq \sqrt{\frac{2(1-\vartheta)}{1-\rho}}$, we know $\vartheta + f_1(\sqrt{r}, t') \geq 1$ always holds.

When the correlation is negative, we have $\sqrt{r} = \max\left\{\sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{\sqrt{1-\rho^2}}, \sqrt{\frac{1-2\vartheta}{2(1-|\rho|)}} + \frac{1}{1-|\rho|}\right\}$. From $\vartheta + f_2(\sqrt{r}, t') \geq 1$, we get the requirement $\sqrt{r} \geq \max\left\{1 + \sqrt{1-\vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-|\rho|}}\right\}$. For $\vartheta + f_1(\sqrt{r}, t') \geq 1$, since we already have $\sqrt{r} \geq \sqrt{\frac{2(1-\vartheta)}{1-|\rho|}}$, we know $FP_2 = o(1)$ always holds.

*Fourth*, if $\vartheta + f_1(\sqrt{r}, t') = 2\vartheta + f_3(\sqrt{r}, t') = 1$ and $t' \geq 1$, $\vartheta + f_2(\sqrt{r}, t') \geq 1$, then we will get nothing from this case. To see this, we still list a few requirement:

- We know that $t' \geq 1$.

- From $\vartheta + f_2(\sqrt{r}, t') \geq 1$, we know that $\sqrt{r} \geq \max\left\{t' + \sqrt{1-\vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-|\rho|}}\right\}$.

- From $2\vartheta + f_3(\sqrt{r}, t') = 1$, we know that $\sqrt{r} = \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{t'}{\sqrt{1-\rho^2}}$ for $\rho > 0$ and $\sqrt{r} = \max\left\{\sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{t'}{\sqrt{1-\rho^2}}, \sqrt{\frac{1-2\vartheta}{2(1-|\rho|)}} + \frac{t'}{1-|\rho|}\right\}$ for $\rho < 0$.

Even if this case admits any curve, that curve would be above the curve in the previous *third* case, and exist within a smaller region of $\vartheta$.

To sum up, we have the following results:

- Phae diagram when the correlation is positive:

$$\sqrt{r} = \max\left\{1 + \sqrt{1-\vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-\rho}}, \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{\sqrt{1-\rho^2}}\right\}.$$

- Phae diagram when the correlation is negative:

$$\sqrt{r} = \max\left\{1 + \sqrt{1-\vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-|\rho|}}, \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{\sqrt{1-\rho^2}}, \sqrt{\frac{1-2\vartheta}{2(1-|\rho|)}} + \frac{1}{1-|\rho|}\right\}$$

# H   PROOF OF THEOREM 6 (FORWARD BACKWARD SELECTION)

The proof for Forward Selection still has three tasks: (a) deriving the rejection region, (b) obtaining the rate of convergence of $\mathbb{E}[H(\hat{\beta}, \beta)]$, and (c) calculating the phase diagram. However, as we will see later, forward backward selection has six cases, each of which has a different shape of the rejection region. After deriving the rejection region, we consider the $\mathbb{E}[H(\hat{\beta}, \beta)]$ and phase curves of the six cases one by one, and summarise the results at the end.

Defore deriving the rejection region in the first part, we need some clarification about the definition of the forward backward selection we have investigated. More precisely, we have simplified the backward step into one thresholding step after the forward selection algorithm, so it is more precisely "thresholded forward selection".

The reason why we have not used a sequential algorithm with alternating forward and backward steps, like FoBa defined in Zhang (2011), is not compuational simplicity, but to avoid degeneration. We explain briefly why any sequential algorithm with alternating forward and backward steps will either have nonfunctional backward steps, or be unable to terminate at a finite step.

To see this, we review the setting of Lemma G.1 about the solution path of forward selection, in which $h_1 > |h_2|$ and we only consider a bivariate problem. Using the same argument, some version of FoBa can also be decomposed into bivariate subproblems, and it is equivalent to running the algorithm only on $y$ and $(x_j, x_{j+1})$.

In brief, in such a bivariate problem with $h_1 > |h_2|$, if either or both of $(x_j, x_{j+1})$ ever get selected and then deleted at some backward step, then they will be selected back again because they still meet the requirements for a variable to get enrolled. When $h_1 > |h_2|$, the case of deleting $x_j$ while leaving $x_{j+1}$ still in the model cannot happen, because no deletion rule based on $(x_j'y, x_{j+1}'y)$ can delete $x_j$ without touching $x_{j+1}$. As a result, the algorithm cannot terminate at a finite step.

If the algorithm termininates at a finite step, then the backward step much have not deleted any of $(x_j, x_{j+1})$, and such algorithm performs the same as forward selection.

We have explained the degeneration of Foba (Zhang, 2011), but we still want to implement some kind of backward step additional to forward selection, because the problem with forward selection is inability to correct the mistakes made in the early steps. Thus it is natural to use one thresholding step at the end.

**Part 1: Deriving the rejection region.**   We first work on the solution path, and then compute the rejection region. The forward selection part has been discussed before, and we recall the results in Lemma G.1 (re-iterated in an equivalent way):

1. When $t \geq h_1$, neither is selected.

2. When $t < h_1$, and $\rho h_1 - t\sqrt{1 - \rho^2} \leq h_2 \leq \rho h_1 + t\sqrt{1 - \rho^2}$, only $x_j$ is selected.

3. When $t < h_1$, and $\begin{cases} \text{either } h_2 > \rho h_1 + t\sqrt{1 - \rho^2} \\ \text{or } -h_1 < h_2 < \rho h_1 - t\sqrt{1 - \rho^2} \end{cases}$ , both $x_j$ and $x_{j+1}$ are selected.

Now this is followed by a thresholding step. Before using $v$ to threshold the results, we note that:

1. When $x_{j+1}$ is not selected, $\hat{\beta}_j = h_1$;

2. When both $(x_j, x_{j+1})$ are selected, $\hat{\beta}_j = \frac{h_1 - \rho h_2}{1 - \rho^2}$ and $\hat{\beta}_{j+1} = \frac{h_2 - \rho h_1}{1 - \rho^2}$.

When the thresholding is performed, we can naturally describe the solution path of thresholded forward selection as:

**Lemma 4.**   *Consider the bivariate problem of running Algorithm 1 with $y$ and $(x_j, x_{j+1})$, followed by thresholding $(\hat{\beta}_j, \hat{\beta}_{j+1})$ with $v$. Define $h_1 = x_j'y/\sqrt{2\log(p)}$, $h_2 = x_{j+1}'y/\sqrt{2\log(p)}$ and $t' = t/\sqrt{2\log(p)}$, $v' = v/\sqrt{2\log(p)}$. Suppose $h_1 > |h_2| \geq 0$, and $\rho > 0$. Then*

- *When $t' \geq h_1$, neither is selected.*

- *When $t' < h_1$ and $\rho h_1 - t'\sqrt{1-\rho^2} \leq h_2 \leq \rho h_1 + t'\sqrt{1-\rho^2}$,*

  - *If $v' \geq h_1$, neither is selected.*
  - *If $v' < h_1$, only $x_j$ is selected.*

- *When $t' < h_1$, and $\begin{cases} \text{either } h_2 > \rho h_1 + t'\sqrt{1-\rho^2} \\ \text{or } -h_1 < h_2 < \rho h_1 - t'\sqrt{1-\rho^2} \end{cases}$ ,*

  - *If $h_2 > \rho h_1 + t'\sqrt{1-\rho^2}$, and $h_2 - \rho h_1 > v'(1-\rho^2)$, both $(x_j, x_{j+1})$ are selected.*
  - *If $h_2 > \rho h_1 + t'\sqrt{1-\rho^2}$, and $h_2 - \rho h_1 \leq v'(1-\rho^2) < h_1 - \rho h_2$, only $x_j$ is selected.*
  - *If $h_2 > \rho h_1 + t'\sqrt{1-\rho^2}$, and $h_1 - \rho h_2 \leq v'(1-\rho^2)$, neither is selected.*
  - *If $h_2 < \rho h_1 - t'\sqrt{1-\rho^2}$, and $\rho h_1 - h_2 > v'(1-\rho^2)$, both $(x_j, x_{j+1})$ are selected.*
  - *If $h_2 < \rho h_1 - t'\sqrt{1-\rho^2}$, and $\rho h_1 - h_2 \leq v'(1-\rho^2) < h_1 - \rho h_2$, only $x_j$ is selected.*
  - *If $h_2 < \rho h_1 - t'\sqrt{1-\rho^2}$, and $h_1 - \rho h_2 \leq v'(1-\rho^2)$, neither is selected.*

The rejection region can be complicated, and it has many cases visually. See Figure 13.

$$
\begin{aligned}
\mathcal{R} = &\{(h_1, h_2) : h_1 - \rho h_2 > \max\{t'\sqrt{1-\rho^2}, v'(1-\rho^2)\}, h_2 > \rho h_1 + t\sqrt{1-\rho^2}\} \\
&\cup \{(h_1, h_2) : h_1 > \max\{t', v'\}, h_2 \leq \rho h_1 + t\sqrt{1-\rho^2}, h_2 \geq \rho h_1 - t\sqrt{1-\rho^2}\} \\
&\cup \{(h_1, h_2) : h_1 > h_2, h_1 > \max\{t', v'\}, h_2 > \rho h_1 - t\sqrt{1-\rho^2}\} \\
&\cup \{(h_1, h_2) : h_1 > t', h_2 \leq \rho h_1 - t\sqrt{1-\rho^2}, h_1 - \rho h_2 > v'(1-\rho^2)\} \\
&\cup \{(h_1, h_2) : h_2 < -t, h_1 - \rho h_2 > \max\{t'\sqrt{1-\rho^2}, v'(1-\rho^2)\}\} \\
&\cup \{(h_1, h_2) : -h_1 + \rho h_2 > \max\{t'\sqrt{1-\rho^2}, v'(1-\rho^2)\}, h_2 < \rho h_1 - t\sqrt{1-\rho^2}\} \\
&\cup \{(h_1, h_2) : h_1 < -\max\{t', v'\}, h_2 \geq \rho h_1 - t\sqrt{1-\rho^2}, h_2 \leq \rho h_1 + t\sqrt{1-\rho^2}\} \\
&\cup \{(h_1, h_2) : h_1 < h_2, h_1 > \max\{t', v'\}, h_2 < \rho h_1 + t\sqrt{1-\rho^2}\} \\
&\cup \{(h_1, h_2) : h_1 < -t', h_2 \geq \rho h_1 + t\sqrt{1-\rho^2}, -h_1 + \rho h_2 > v'(1-\rho^2)\} \\
&\cup \{(h_1, h_2) : h_2 > t, -h_1 + \rho h_2 > \max\{t'\sqrt{1-\rho^2}, v'(1-\rho^2)\}\}\}
\end{aligned}
\tag{51}
$$

Due to the many cases of thresholded forward selection, we structure the rest of the proof in a different way: We discuss the six cases shown in Figure 13 in the next six parts, and summarise the results for $\rho \geq 0$ and $\rho < 0$ respectively, at the end of the proof. In other words, each of the six cases has its phase curves, and we take the minimum of all the curves to be the final phase curve.

**Case 1: When $v' \leq t'$.** From the rejection region defined in Equation (51), and Figure 13a, we know that thresholding does not have any effects in this case. Everything can be copied from forward selection:

The curve between Almost Full Recovery and No Recovery is $r = \vartheta$. The curve between Almost Full Recovery and Exact Recovery is: When $\rho \geq 0$,

$$
\sqrt{r} = \max\left\{ 1 + \sqrt{1-\vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-\rho}}, \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{\sqrt{1-\rho^2}} \right\}.
\tag{52}
$$

When $\rho < 0$,

$$
\sqrt{r} = \max\left\{ 1 + \sqrt{1-\vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-|\rho|}}, \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{\sqrt{1-\rho^2}}, \sqrt{\frac{1-2\vartheta}{2(1-|\rho|)}} + \frac{1}{1-|\rho|} \right\}.
\tag{53}
$$

**Case 2: When $t' \leq v' \leq \frac{t'}{\sqrt{1-\rho^2}}$.**

(a) When $v' \leq t'$

(b) When $t' < v' \leq \frac{t}{\sqrt{1-\rho^2}}$

(c) When $\frac{t'}{\sqrt{1-\rho^2}} < v' \leq t'\left(1+\frac{\rho}{\sqrt{1-\rho^2}}\right)$

(d) When $t'\left(1+\frac{\rho}{\sqrt{1-\rho^2}}\right) \leq v' \leq \frac{t'\sqrt{1-\rho^2}}{1-\rho}$

(e) When $\frac{t'\sqrt{1-\rho^2}}{1-\rho} \leq v' \leq \frac{t'}{1-\rho}$
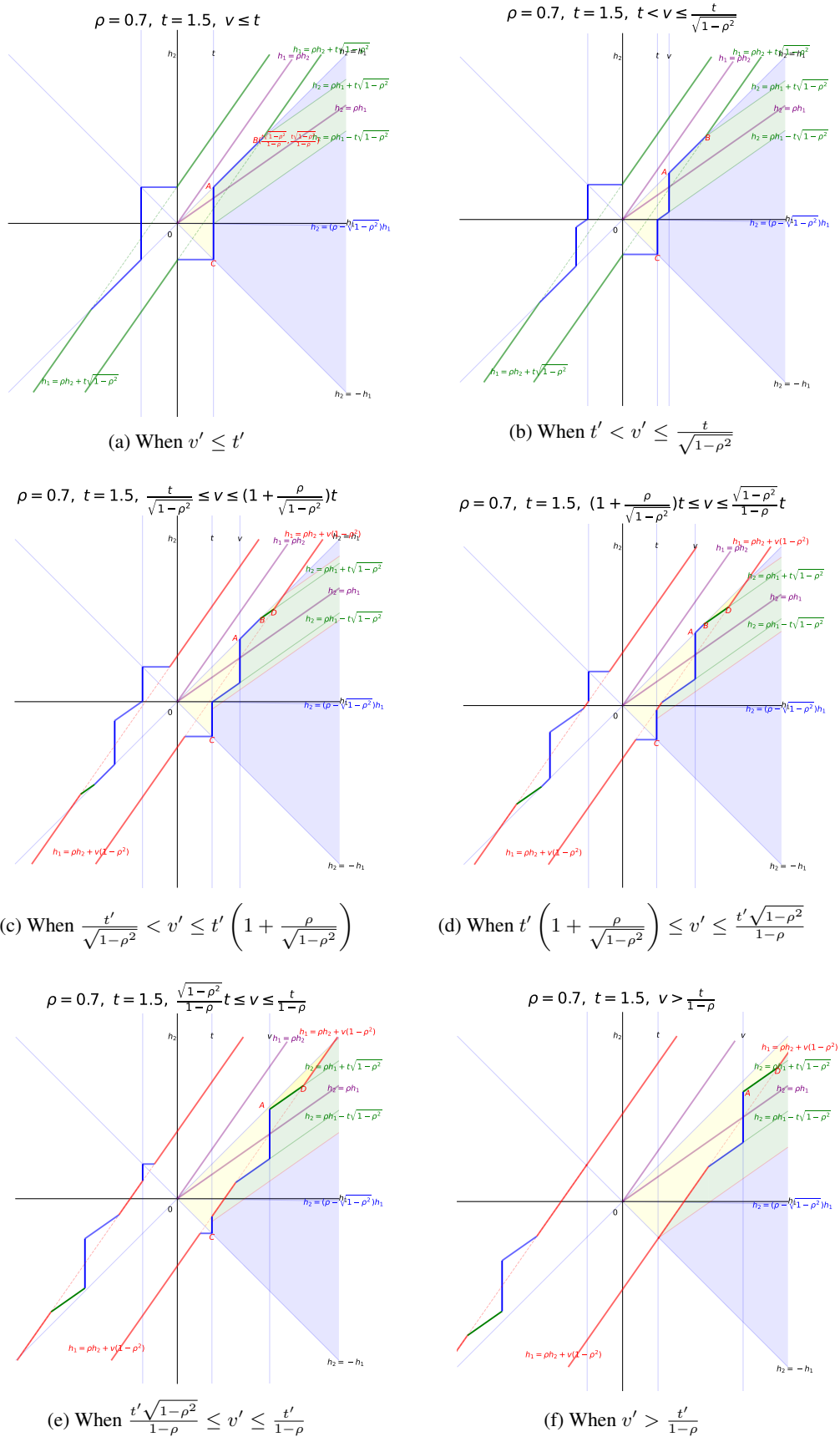
(f) When $v' > \frac{t'}{1-\rho}$

Figure 13: The rejection region of thresholded forward selection has many cases ($\rho \geq 0$).

**Theorem 15** (The Hamming error rate When $t' \leq v' \leq \frac{t'}{\sqrt{1-\rho^2}}$). *Suppose the conditions of Theorem 6 holds. Let $h_1 = x'_j y / \sqrt{2\log(p)}$, $h_2 = x'_{j+1} y / \sqrt{2\log(p)}$ and $v' = v/\sqrt{2\log(p)}$, $t' = t/\sqrt{2\log(p)}$. We require $t' \leq v' \leq \frac{t'}{\sqrt{1-\rho^2}}$. As $p \to \infty$,*

$$\mathrm{FP}_p = L_p p^{1-\min\left\{\min\{v'^2, 2t'^2\}, \ \vartheta + f_1(\sqrt{r}, t', v')\right\}}, \qquad \mathrm{FN}_p = L_p p^{1-\min\left\{\vartheta + f_2(\sqrt{r}, t', v'), \ 2\vartheta + f_3(\sqrt{r}, t', v')\right\}},$$

*where (below, $d^2_{|\rho|}(u, v)$ is as in Definition 1),*

$$f_1(\sqrt{r}, t', v') = \begin{cases} (v' - |\rho|\sqrt{r})^2 & \text{if } \sqrt{r} \leq \frac{v'}{1+|\rho|} \\ \frac{1}{1-\rho^2} d^2_{|\rho|}((v', v'), (|\rho|\sqrt{r}, \sqrt{r})) & \text{if } \frac{v'}{1+|\rho|} < \sqrt{r} \leq \frac{2v'}{1+|\rho|} \\ \min\left\{\frac{1}{2}(1-|\rho|)r, \ t'^2\right\} & \text{if } \sqrt{r} > \frac{2v'}{1+|\rho|} \end{cases}$$

$$f_2(\sqrt{r}, t', v') = \begin{cases} \min\left\{(\sqrt{r} - t')^2_+, \ \frac{1}{2}(1-|\rho|)r\right\} & \text{if } \sqrt{r} \leq \frac{2t'}{\sqrt{1-\rho^2}} \\ \min\left\{(\sqrt{r} - t')^2_+, \ \frac{1}{1-\rho^2} d^2_{|\rho|}(B, (\sqrt{r}, |\rho|\sqrt{r}))\right\} & \text{if } \frac{2t'}{\sqrt{1-\rho^2}} < \sqrt{r} \leq \frac{t'\sqrt{1-\rho^2}}{|\rho|(1-|\rho|)} \\ \left[\sqrt{1-\rho^2}\sqrt{r} - t'\right]^2 & \text{if } \sqrt{r} > \frac{t'\sqrt{1-\rho^2}}{|\rho|(1-|\rho|)} \end{cases}$$

*The definition of $f_3(\sqrt{r}, t')$ depends on the sign of $\rho$. When $\rho > 0$,*

$$f_3(\sqrt{r}, t') = \left[\sqrt{1-\rho^2}\sqrt{r} - t\right]^2$$

*When $\rho < 0$,*

$$f_3(\sqrt{r}, t') = \min\left\{\left[\sqrt{1-\rho^2}\sqrt{r} - t\right]^2, \ d^2_{|\rho|}(C, ((1-|\rho|)\sqrt{r}, -(1-|\rho|)\sqrt{r}))\right\}$$

$$= \min\left\{\left[\sqrt{1-\rho^2}\sqrt{r} - t\right]^2, \ \frac{2}{1-|\rho|}\left[(1-|\rho|)\sqrt{r} - t\right]^2\right\}$$

**Remark 4.** *The proof of Theorem 15 is easy, but we need to emphasize one thing: For $FP_1$, whose ellipsoid is centered at $(0,0)$, it may be tangent to $h_1 = v'$ at $(v', |\rho|v')$; or, it may intersect the rejection region at the corner $(t, t(\rho - \sqrt{1-\rho^2}))$.*

**Theorem 16** (The phase diagram when $t \leq v \leq \frac{t}{\sqrt{1-\rho^2}}$). *Suppose the conditions of Theorem 6 holds. The boundary between Exact Recovery and and Almost Full Recovery is Equation 54 when the correlation is positive, and Equation 55 when the correlation is negative. When $\rho \geq 0$,*

$$\sqrt{r} = \max\left\{1 + \sqrt{1-\vartheta}, \ \sqrt{\frac{2(1-\vartheta)}{1-\rho}}, \ \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \sqrt{\frac{1-\vartheta}{1-\rho^2}}\right\}. \tag{54}$$

*When $\rho < 0$,*

$$\sqrt{r} = \max\left\{1 + \sqrt{1-\vartheta}, \ \sqrt{\frac{2(1-\vartheta)}{1-|\rho|}}, \ \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \sqrt{\frac{1-\vartheta}{1-\rho^2}}, \right.$$
$$\left. \sqrt{\frac{1-2\vartheta}{2(1-|\rho|)}} + \frac{\sqrt{1-\vartheta}}{1-|\rho|}, \ \sqrt{\frac{1-2\vartheta}{2(1-|\rho|)}} + \frac{\sqrt{1-\rho^2}}{1-|\rho|}\right\}. \tag{55}$$

*Proof of Theorem 16.* Like the proof of forward selection and other methods, We still discuss the $2 \times 2 = 4$ cases. For brevity, we use $f_1, f_2, f_3$ as shorthand of $f_1(\sqrt{r}, t', v'), f_2(\sqrt{r}, t', v'), f_3(\sqrt{r}, t', v')$.

*First*, if $\min(v'^2, 2t'^2) = \vartheta + f_2 = 1$, we have $\sqrt{r} = \max\left\{v' + \sqrt{1-\vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-|\rho|}}\right\}$. To ensure $\vartheta + f_1 \geq 1$, we need $t' \geq \sqrt{1-\vartheta}$. For $f_2$, it need to meet the requirement $\sqrt{r} \leq \frac{2t'}{\sqrt{1-\rho^2}}$. This is not restrictive, because:

$$\begin{cases} t' \geq \sqrt{1-\vartheta} \\ 1 \leq v' \leq \frac{t'}{\sqrt{1-\rho^2}} \end{cases} \implies t' \geq v'\sqrt{1-\rho^2} \implies 2t' \geq \sqrt{1-\vartheta} + v'\sqrt{1-\rho^2}.$$

(For the conditional expression of $f_2$, $\sqrt{r} \leq \frac{2t'}{\sqrt{1-\rho^2}}$ is the only possibility here; to see this, we can just refer to the same part of proof for forward selection.) Finally, $2\vartheta + f_4 \geq 1$ requires $\sqrt{r} \geq \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{t'}{\sqrt{1-\rho^2}}$ when $\rho \geq 0$, and $\sqrt{r} \geq \max\left\{\sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{t'}{\sqrt{1-\rho^2}}, \sqrt{\frac{1-2\vartheta}{2(1-|\rho|)}} + \frac{t'}{1-|\rho|}\right\}$ when $\rho \leq 0$.

In the above discussion, the $(v', t')$ refers to any admissible $t'$ in this case, so we choose $v'_{\min} = 1$ and $t'_{\min} = \min\left\{\sqrt{1-\vartheta}, \sqrt{1-\rho^2}, \frac{\sqrt{2}}{2}\right\}$. To sum up, $\sqrt{r} = \max\left\{v'_{\min} + \sqrt{1-\vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-|\rho|}}\right\}$, and we require $\sqrt{r} \geq \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{t'_{\min}}{\sqrt{1-\rho^2}}$ when $\rho \geq 0$, and $\sqrt{r} \geq \max\left\{\sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{t'_{\min}}{\sqrt{1-\rho^2}}, \sqrt{\frac{1-2\vartheta}{2(1-|\rho|)}} + \frac{t'_{\min}}{1-|\rho|}\right\}$ when $\rho \leq 0$.

*Second*, if $\vartheta + f_1 = \vartheta + f_2 = 1$, this case will not give us any curve. We first need $v' \geq 1$ and $t' \geq \frac{\sqrt{2}}{2}$, and $\sqrt{r} = \max\left\{v' + \sqrt{1-\vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-\rho}}\right\}$. For $\vartheta + f_1 \geq 1$, we already have $\sqrt{r} \geq \sqrt{\frac{2(1-\vartheta)}{1-\rho}}$, so we only need $t' \geq \sqrt{1-\vartheta}$. The requirement from $2\vartheta + f_3 \geq 1$ is still the same as that of the *first* case.

We notice that even if this case admits any curve, it is strictly above the curve yielded by $FP_1 = FN_1$, and it exists in a smaller interval of $\vartheta$. As a result, we need not discuss this case any further.

*Third*, if $\min\left\{v'^2, 2t'^2\right\} = \vartheta + f_2 = 1$, then we immdiately have $v' \geq 1$, $t' \geq \frac{\sqrt{2}}{2}$, and we can limit ourselves to consider $\vartheta \leq \frac{1}{2}$. From $\vartheta + f_1 \geq 1$ and $\vartheta + f_2 \geq 1$, we have the requirement $\sqrt{r} \geq \max\left\{v' + \sqrt{1-\vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-\rho}}\right\}$ and $t' \geq \sqrt{1-\vartheta}$. Also, since $v' \leq t'/\sqrt{1-\rho^2}$, we need $t' \geq \sqrt{1-\rho^2}$.

When $\rho \geq 0$, we have $\sqrt{r} = \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{t'}{\sqrt{1-\rho^2}}$; when $\rho < 0$, we have $\sqrt{r} = \max\left\{\sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{t'}{\sqrt{1-\rho^2}}, \sqrt{\frac{1-2\vartheta}{2(1-\rho)}} + \frac{t'}{1-\rho}\right\}$. Just like the *first* case, we can take $v'_{\min} = 1$ and $t'_{\min} = \min\left\{\sqrt{1-\vartheta}, \sqrt{1-\rho^2}, \frac{\sqrt{2}}{2}\right\}$ in the expression of $\sqrt{r}$.

*Fourth*, if $\vartheta + f_1 = 2\vartheta + f_3 = 1$, this case does not give any curve. The discussion is exactly the same as the *second* case: even if this case gives us any curve, it would be strictly above the curve in the *third* case.

To sum up, define $v'_{\min} = 1$ and $t'_{\min} = \min\left\{\sqrt{1-\vartheta}, \sqrt{1-\rho^2}, \frac{\sqrt{2}}{2}\right\}$: When $\rho \geq 0$:

$$\sqrt{r} = \max\left\{v'_{\min} + \sqrt{1-\vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-|\rho|}}, \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{t'_{\min}}{\sqrt{1-\rho^2}}\right\}$$

and add $\sqrt{\frac{1-2\vartheta}{2(1-\rho)}} + \frac{t'_{\min}}{1-\rho}$ into the maximum when $\rho < 0$.

We can simplify the expression of the curve above, by deleting a few curves in the maximum:

- the curve $\sqrt{r} = \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{\sqrt{2}/2}{\sqrt{1-\rho^2}}$ is always below other curves, and can be omitted. This is because when $\vartheta \leq \frac{1}{2}$, $\sqrt{1-\vartheta} \geq \frac{\sqrt{2}}{2}$, which implies $\sqrt{r} = \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{\sqrt{2}/2}{\sqrt{1-\rho^2}} \leq \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{\sqrt{1-\vartheta}}{\sqrt{1-\rho^2}}$.

- for the same reason, the curve $\sqrt{r} = \sqrt{\frac{1-2\vartheta}{2(1-\rho)}} + \frac{\sqrt{2}/2}{1-\rho}$ is also always below other curves, and can be omitted.

- the curve $\sqrt{r} = \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{\sqrt{1-\rho^2}}{\sqrt{1-\rho^2}}$ is always below other curves, and can be omitted.

  - When $\rho \geq \frac{\sqrt{2}}{2}$, $\sqrt{1-\rho^2} \leq \sqrt{1-\vartheta}$ for all $\vartheta \leq \frac{1}{2}$. Thus $\sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{\sqrt{1-\rho^2}}{\sqrt{1-\rho^2}} \leq$
    $\sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{\sqrt{1-\vartheta}}{\sqrt{1-\rho^2}}$

  - When $\rho \geq \frac{\sqrt{2}}{2}$: If $\vartheta \leq \rho^2$, we still have $\sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{\sqrt{1-\rho^2}}{\sqrt{1-\rho^2}} \leq \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{\sqrt{1-\vartheta}}{\sqrt{1-\rho^2}}$. If
    $\rho^2 < \vartheta \leq \frac{1}{2}$, it can be verified that $\sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{\sqrt{1-\rho^2}}{\sqrt{1-\rho^2}} \leq 1 + \sqrt{1-\vartheta}$.

Now we have arrived at the conclusion of Theorem 16. $\qquad\square$

**Case 3: When $\frac{t'}{\sqrt{1-\rho^2}} \leq v' \leq t'(1 + \frac{|\rho|}{\sqrt{1-\rho^2}})$.**

**Theorem 17** (The Hamming error rate When $\frac{t'}{\sqrt{1-\rho^2}} \leq v' \leq t'(1+\frac{|\rho|}{\sqrt{1-\rho^2}})$). *Suppose the conditions of Theorem 6 holds. Let $h_1 = x'_j y / \sqrt{2\log(p)}$, $h_2 = x'_{j+1} y / \sqrt{2\log(p)}$ and $v' = v/\sqrt{2\log(p)}$, $t' = t/\sqrt{2\log(p)}$. As shorthand notation, define the points $A(v', v')$, $B(\frac{t'\sqrt{1-\rho^2}}{1-|\rho|}, \frac{t'\sqrt{1-\rho^2}}{1-|\rho|})$, and $D(v' + \frac{\rho t'}{\sqrt{1-\rho^2}}, \rho v' + \frac{t'}{\sqrt{1-\rho^2}})$ as marked in Figure 13c. We require $t'/\sqrt{1-\rho^2} \leq v' \leq t'(1 + |\rho|/\sqrt{1-\rho^2})$. As $p \to \infty$,*

$$\mathrm{FP}_p = L_p p^{1-\min\left\{\min\{v'^2, 2t'^2\}, \; \vartheta + f_1(\sqrt{r}, t', v')\right\}}, \qquad \mathrm{FN}_p = L_p p^{1-\min\left\{\vartheta + f_2(\sqrt{r}, t', v'), \; 2\vartheta + f_3(\sqrt{r}, t', v')\right\}},$$

*where (below, $d^2_{|\rho|}(u, v)$ is as in Definition 1),*

$$f_1(\sqrt{r}, t', v') = \begin{cases} (v' - |\rho|\sqrt{r})^2 & \text{if } \sqrt{r} \leq \frac{v'}{1+\rho} \\ \frac{1}{1-\rho^2} d^2_{|\rho|}(A, (|\rho|\sqrt{r}, \sqrt{r})) & \text{if } \frac{v'}{1+\rho} < \sqrt{r} \leq \frac{2v'}{1+\rho} \\ \min\{k(v', t'), \; v'^2(1-\rho^2)\} & \text{if } \sqrt{r} > \frac{2v'}{1+\rho} \end{cases}$$

*where $k(v', t')$ is defined like:*

$$k(v', t') \stackrel{\text{def}}{=} \begin{cases} \frac{1}{2}(1 - |\rho|)r & \text{if } \frac{2v'}{1+|\rho|} \leq \sqrt{r} \leq \frac{2t'}{\sqrt{1-\rho^2}} \\ \frac{1}{1-\rho^2} d^2_{|\rho|}(B, (|\rho|\sqrt{r}, \sqrt{r})) & \text{if } \frac{2t'}{\sqrt{1-\rho^2}} \leq \sqrt{r} \leq \frac{t\sqrt{1-\rho^2}}{|\rho|(1-|\rho|)} \\ \left[\sqrt{1-\rho^2}\sqrt{r} - t'\right]^2 & \text{if } \sqrt{r} \geq \frac{t\sqrt{1-\rho^2}}{|\rho|(1-|\rho|)} \end{cases}$$

$$f_2(\sqrt{r}, t', v') = \begin{cases} \min\left\{(\sqrt{r} - v')^2_+, \frac{1}{2}(1-|\rho|)r, \; t'^2\right\} & \text{if } \sqrt{r} \leq v' + |\rho|\frac{t'}{\sqrt{1-\rho^2}} \\ \min\{(\sqrt{r} - v')^2_+, \frac{1}{2}(1-|\rho|)r, \frac{1}{1-\rho^2} d^2_{|\rho|}(D, (\sqrt{r}, |\rho|\sqrt{r}))\} \\ \qquad\qquad \text{if } v' + |\rho|\frac{t'}{\sqrt{1-\rho^2}} \leq \sqrt{r} \leq \min\{\sqrt{r_2(v', t')}, v' + \frac{t'}{|\rho|\sqrt{1-\rho^2}}\} \\ \min\left\{(\sqrt{r} - v')^2_+, \frac{1}{1-\rho^2} d^2_{|\rho|}(D, (\sqrt{r}, |\rho|\sqrt{r}))\right\} \; \text{if } \min\{\sqrt{r_2}, v' + \frac{t'}{|\rho|\sqrt{1-\rho^2}}\} \leq \sqrt{r} \leq v' + \frac{t'}{|\rho|\sqrt{1-\rho^2}} \\ (1-\rho^2)\left[\sqrt{r} - v'\right]^2 & \text{if } \sqrt{r} \geq v' + \frac{t'}{|\rho|\sqrt{1-\rho^2}} \end{cases}$$

*where $r_2 = r_2(v', t')$ is the larger root of the quadratic equation*

$$\frac{1}{1-\rho^2} d^2_{|\rho|}(D, (\sqrt{r}, |\rho|\sqrt{r})) = \frac{1}{2}(1-|\rho|)r$$

$$\Leftrightarrow \frac{1+|\rho|}{2} r - 2\left(v' + \frac{|\rho|t'}{\sqrt{1-\rho^2}}\right)\sqrt{r} + \left(v'^2 + \frac{t'^2}{1-\rho^2} + \frac{2|\rho|v't'}{\sqrt{1-\rho^2}}\right) = 0 \qquad (56)$$

76

and the explicitly form of $r_2(v', t')$ is

$$\sqrt{r_2(v', t')} = \frac{1}{1 + |\rho|} \left[ 2 \left( v' + \frac{|\rho| t'}{\sqrt{1 - \rho^2}} \right) + \sqrt{2(1 - |\rho|) \left( v' - \frac{t'}{\sqrt{1 - \rho^2}} \right) \left( v' + (1 + 2|\rho|) \frac{t'}{\sqrt{1 - \rho^2}} \right)} \right].$$

The definition of $f_3(\sqrt{r}, t')$ depends on the sign of $\rho$. When $\rho > 0$,

$$f_3(\sqrt{r}, t') = \left[ \sqrt{1 - \rho^2} \sqrt{r} - t \right]^2$$

When $\rho < 0$,

$$f_3(\sqrt{r}, t') = \min \left\{ \left[ \sqrt{1 - \rho^2} \sqrt{r} - t \right]^2, \ d_{|\rho|}^2(C, ((1 - |\rho|)\sqrt{r}, -(1 - |\rho|)\sqrt{r})) \right\}$$

$$= \min \left\{ \left[ \sqrt{1 - \rho^2} \sqrt{r} - t \right]^2, \ \frac{2}{1 - |\rho|} \left[ (1 - |\rho|)\sqrt{r} - t \right]^2 \right\}$$

*Proof of Theorem 17.* We explain one detail, about why $\sqrt{r_2(v', t')}$ is introduced in $f_2(\sqrt{r}, t', v')$, which corresponds to the ellipsoid centered at $\mu_{10} = (\sqrt{r}, |\rho|\sqrt{r})$. Recall that the point $D$ as noted in in Figure 13c has cooredinate $x_D = v' + \frac{|\rho| t'}{\sqrt{1 - \rho^2}}, y_D = |\rho|v' + \frac{t'}{\sqrt{1 - \rho^2}}$. Suppose $\rho \geq 0$, as the case of $\rho < 0$ can be obtained with symmetry.

When $\sqrt{r} \leq x_D = v' + \frac{\rho t'}{\sqrt{1 - \rho^2}}$, the ellipsoid can be tangent to any one among the three line segments: (i) $h_1 = v'$, (ii) $h_2 = h_1$ or (iii) $h_2 = \rho h_1 + t'\sqrt{1 - \rho^2}$.

When $\rho\sqrt{r} \geq y_D = \rho v' + \frac{t'}{\sqrt{1 - \rho^2}}$, the ellipsoid can either be tangent to the red line $h_1 = \rho h_2 + v'(1 - \rho^2)$, or the blue line $h_1 = v'$ in Figure 13c.

When $v' + \frac{\rho t'}{\sqrt{1 - \rho^2}} \leq \sqrt{r} \leq v' + \frac{t'}{\rho\sqrt{1 - \rho^2}}$, thing are more tricky:

- The ellipsoid can possibly be tangent to $h_1 = v'$, or it may intersect point $D$ and rotate around it.

- However, it is uncertain whether we should include the segment $h_2 = h_1$ into the form of the Hamming error. This is because when $\sqrt{r}$ is large, the ellipsoid is at the upper right side of point $D$, where $h_2 = h_1$ does not form the boundary of the rejection region. If we still include it, the final phase diagram will be worse than it actually is.

- To exclude $h_2 = h_1$ when it is unwanted, we require $\sqrt{r} > \sqrt{r_2(v', t')}$.

- The place of $\sqrt{r_2(v', t')}$ is exchangable to the symmetric axis of the quadratic equation (56), which is $\frac{2}{1+\rho} v' + \frac{2\rho t'}{(1+\rho)\sqrt{1 - \rho^2}}$.
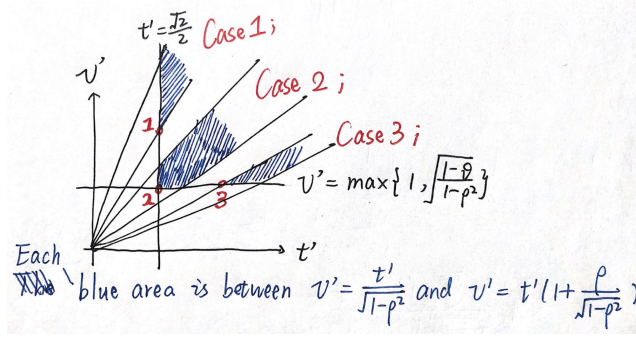
$\sqrt{r_2(v', t')}$ can be greater than $v' + \frac{t'}{\rho\sqrt{1 - \rho^2}}$, so it is taken minimum with $v' + \frac{t'}{\rho\sqrt{1 - \rho^2}}$ in the definition of $f_2$. $\qquad \square$

Before moving on to the phase diagram, we first introduce two terms to simplify notation:

**Definition 3.** *Define $v'_{\min}$ and $t'_{\min}$, both as functions of $\vartheta$ and $\rho$.* $v'_{\min} = \max \left\{ 1, \sqrt{\frac{1 - \vartheta}{1 - \rho^2}}, \frac{\sqrt{2}/2}{\sqrt{1 - \rho^2}} \right\}$

*and* $t'_{\min} = \max \left\{ \frac{\sqrt{2}}{2}, \frac{\max\{1, \sqrt{\frac{1 - \vartheta}{1 - \rho^2}}\}}{1 + |\rho|/\sqrt{1 - \rho^2}} \right\}.$

Figure 14 gives an explanation of how $v'_{\min}$ and $t'_{\min}$ are defined.

Figure 14: the definition of $v'_{\min}$ and $t'_{\min}$

**Theorem 18** (The phase diagram when $\frac{t'}{\sqrt{1-\rho^2}} \leq v' \leq t'(1 + \frac{|\rho|}{\sqrt{1-\rho^2}})$ ). *Suppose the conditions of Theorem 6 holds. The boundary between Exact Recovery and and Almost Full Recovery is Equation 54 when the correlation is positive, and Equation 55 when the correlation is negative. When $\rho \geq 0$,*

$$\sqrt{r} = \max\left\{ v'_{\min} + \sqrt{1-\vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-|\rho|}}, \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + v'_{\min} \right\} \tag{57}$$

*When the actual correlation is negative,*

$$\sqrt{r} = \max\left\{ v'_{\min} + \sqrt{1-\vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-|\rho|}}, \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + v'_{\min}, \sqrt{\frac{1-2\vartheta}{2(1-|\rho|)}} + \frac{t'_{\min}}{1-|\rho|} \right\}. \tag{58}$$

*Proof of Theorem 18.* We start by considering the case of $\rho \geq 0$. The other case of $\rho < 0$ can be prove in a very similar way by adding one more curve. In the discussion below, we do not work with $|\rho|$, but $\rho \geq 0$ itself.

In this particular proof, we do not limit ourselves to discuss the four cases as usual. Instead, we just think about the conditions for $\min\{v'^2, 2t'^2\}, \vartheta + f_1, \vartheta + f_2, 2\vartheta + f_3 \geq 1$, and take the smallest $\sqrt{r}$ possible. Also, remember the important fact that at least two of these requirements should be tight.

For fixed $(v', t')$, the necessary and sufficient condition for $\min\{v'^2, 2t'^2\} \geq 1$ is $v' \geq 1$ and $t' \geq \frac{\sqrt{2}}{2}$. The necessary and sufficient condition for $2\vartheta + f_3 \geq 0$ is $\sqrt{r} \geq \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + v'$. (If $\rho < 0$, the condition should be $\sqrt{r} \geq \max\left\{ \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + v', \sqrt{\frac{1-\vartheta}{1-\rho^2}} \right\}$.) From $\vartheta + f_2 \geq 1$, we know a necessary condition, $\sqrt{r} \geq v' + \sqrt{1-\vartheta}$. From the discussion above, we already know

$$\sqrt{r} \geq \max\left\{ v' + \sqrt{1-\vartheta}, \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + v' \right\}$$

In terms of admissible $(v', t')$, we already have $v' \geq 1$ and $t' \geq \frac{\sqrt{2}}{2}$. Additionally, note that $f_1(\sqrt{r}, t', v')$ as a function of $\sqrt{r}$ takes its maximum when $f_1 = (1-\rho^2)v'^2$, so $\vartheta + f_1 \geq 1$ implies $v' \geq \sqrt{\frac{1-\vartheta}{1-\rho^2}}$. Under the conditions $v' \geq \max\{1, \sqrt{\frac{1-\vartheta}{1-\rho^2}}\}$, $t' \geq \frac{\sqrt{2}}{2}$, and $\frac{t'}{\sqrt{1-\rho^2}} \leq v' \leq t'(1 + \frac{|\rho|}{\sqrt{1-\rho^2}})$, the smallest admissible $(v', t')$ are precisely defined by Definition 3 as $(v'_{\min}, t'_{\min})$.

We consider the two cases: $\rho \leq 0.576$ and $\rho \geq 0.576$. The point $0.576$ is important, because when $\rho \leq 0.576$, since $v'_{\min} \geq \max\left\{1, \sqrt{\frac{1-\vartheta}{1-\rho^2}}\right\}$, we can prove

$$\sqrt{r} \geq \max\left\{ v'_{\min} + \sqrt{1-\vartheta}, \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + v'_{\min} \right\} \geq \sqrt{\frac{2(1-\vartheta)}{1-\rho}}.$$

78

Since now $\sqrt{r} \geq \sqrt{\frac{2(1-\vartheta)}{1-\rho}}$ is already implied by other necessary conditions, we only need to check the sufficiency of $\sqrt{r} = \max\left\{v'_{\min} + \sqrt{1-\vartheta}, \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + v'_{\min}\right\}$ to prove Theorem 18 for $0 \leq \rho < 0.576$.

When $0 \leq \rho \leq 0.576$, we already have $\min\{v'^2, 2t^2\} \geq 1$, $2\vartheta + f_3 \geq 1$ and $\sqrt{r} \geq \sqrt{\frac{2(1-\vartheta)}{1-\rho}}$. we need to check $\vartheta + f_1 \geq 1$ and $\vartheta + f_2 \geq 1$.

For $\vartheta + f_1 \geq 1$, because $\frac{1}{1-\rho^2} d^2_{|\rho|}(A, (|\rho|\sqrt{r}, \sqrt{r})) \geq \frac{1}{2}(1-\rho)r$ and $\frac{1}{1-\rho^2} d^2_{|\rho|}(B, (|\rho|\sqrt{r}, \sqrt{r})) \geq \frac{1}{2}(1-\rho)r$, we only need to check $\sqrt{r} \geq \sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{t'}{\sqrt{1-\rho^2}}$ when $\sqrt{r} \geq \frac{t\sqrt{1-\rho^2}}{\rho(1-\rho)}$. Since $t' \geq \frac{v'}{1+\frac{\rho}{\sqrt{1-\rho^2}}} \geq \frac{\sqrt{1-\vartheta}}{\rho+\sqrt{1-\rho^2}} \geq \rho\sqrt{1-\vartheta}$ for all $\rho \leq \frac{\sqrt{2}}{2}$, we have $t' \geq \rho\sqrt{1-\vartheta}$, and now actually $\frac{t\sqrt{1-\rho^2}}{\rho(1-\rho)} \geq \sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{t'}{\sqrt{1-\rho^2}}$, so this case is not restrictive at all.

For $\vartheta + f_2 \geq 1$, when $\sqrt{r} \leq \sqrt{r_2(v', t')}$, it is sufficient to have $\sqrt{r} \geq \sqrt{\frac{2(1-\vartheta)}{1-\rho}}$. We only need to check the rest two cases in which $\sqrt{r}$ is large:

- When $\sqrt{r_2(v', t')} \leq \sqrt{r} \leq v' + \frac{t'}{\rho\sqrt{1-\rho^2}}$, we need $d^2(D, (\sqrt{r}, \rho\sqrt{r})) \geq (1-\rho^2)(1-\vartheta)$. This trivially holds, because $\sqrt{r} \geq \sqrt{r_2(v', t')}$. From Equation (56):

$$r - 2\left(v' + \frac{\rho t'}{\sqrt{1-\rho^2}}\right)\sqrt{r} + \left(v'^2 + \frac{t'^2}{1-\rho^2} + \frac{2\rho v' t'}{\sqrt{1-\rho^2}}\right) \geq \frac{1-\rho}{2}r \geq 1-\vartheta$$

  whose last inequality is because $\sqrt{r} \geq \sqrt{\frac{2(1-\vartheta)}{1-\rho}}$.

- When $\sqrt{r} \geq v' + \frac{t'}{\rho\sqrt{1-\rho^2}}$, we need $\sqrt{r} \geq \sqrt{\frac{1-\vartheta}{1-\rho^2}} + v'$. Since $t' \geq \rho\sqrt{1-\vartheta}$, it trivially holds.

We have proved Theorem 18 for $\rho \leq 0.576$. Now we move on to the next case of $\rho > 0.576$.

For $\rho > 0.576$, our task is to prove it is necessary and sufficient to have $\sqrt{r} \geq \sqrt{\frac{2(1-\vartheta)}{1-\rho}}$. Reviewing Theorem 18, as long as $\sqrt{r} \leq \sqrt{r_2(v', t')}$ and $\sqrt{r} \leq \frac{2t'}{\sqrt{1-\rho^2}}$ for fixed admissible $(v', t')$, then $\sqrt{r} \geq \sqrt{\frac{2(1-\vartheta)}{1-\rho}}$ is also sufficient for $\sqrt{r}$. In other words, we could simply set

$$\sqrt{r} = \max\left\{v'_{\min} + \sqrt{1-\vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-\rho}}, \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + v'_{\min}\right\}.$$

We are left to eliminate other cases, i.e. $\sqrt{r} \leq \sqrt{r_2(v', t')}$ or $\sqrt{r} \leq \frac{2t'}{\sqrt{1-\rho^2}}$ for fixed admissible $(v', t')$. They can either be impossible, or only produce a curve greater than the one in Equation 57. The rest of the proof focuses on the elimination of other cases.

To prepare for such work, we take a closer look at the definition of $f_2$. We point out that when $v' \geq \frac{\sqrt{2(1+\rho)}-1}{\rho} \cdot \frac{t}{\sqrt{1-\rho^2}}$, the case of

$$\begin{cases} (\sqrt{r} - v')^2_+ \\ \frac{1}{1-\rho^2} d^2_{|\rho|}(D, (\sqrt{r}, \rho\sqrt{r})) \end{cases} \quad \text{if } \min\{\sqrt{r_2}, v' + \frac{t'}{|\rho|\sqrt{1-\rho^2}}\} \leq \sqrt{r} \leq v' + \frac{t'}{|\rho|\sqrt{1-\rho^2}}$$

does not exist, and the degenerated $f_2$ is just

$$f_2(\sqrt{r}, t', v') = \begin{cases} \min\{(\sqrt{r} - v')^2_+, \frac{1}{2}(1-\rho)r, t'^2\} & \text{if } \sqrt{r} \leq v' + \rho\frac{t'}{\sqrt{1-\rho^2}} \\ \min\{(\sqrt{r} - v')^2_+, \frac{1}{2}(1-\rho)r, \frac{1}{1-\rho^2} d^2_{|\rho|}(D, (\sqrt{r}, \rho\sqrt{r}))\} & \text{if } v' + \rho\frac{t'}{\sqrt{1-\rho^2}} \leq \sqrt{r} \leq v' + \frac{t'}{\rho\sqrt{1-\rho^2}} \\ (1-\rho^2)\left[\sqrt{r} - v'\right]^2 & \text{if } \sqrt{r} \geq v' + \frac{t'}{\rho\sqrt{1-\rho^2}}. \end{cases}$$

When $v' < \frac{\sqrt{2(1+\rho)}-1}{\rho} \cdot \frac{t}{\sqrt{1-\rho^2}}$, the case of

$$\begin{cases} (1-\rho^2)(\sqrt{r}-v')_+^2 & \text{if } \sqrt{r_2(v',t')} \le \sqrt{r} \le v' + \frac{t'}{\rho\sqrt{1-\rho^2}} \\ d^2(D, (\sqrt{r}, \rho\sqrt{r})) & \end{cases}$$

in $f_2$ does exist.

Now we are ready to eliminate the unwanted cases.

When the case of $\frac{2t'}{\sqrt{1-\rho^2}} \le \sqrt{r} \le \frac{t\sqrt{1-\rho^2}}{\rho(1-\rho)}$ in $\vartheta + f_1 \ge 1$ is tight and active, we have $\sqrt{r} = \frac{(1+\rho)t'}{\sqrt{1-\rho^2}} - \sqrt{1-\vartheta-t'^2}$. There is one important fact: $v' + \frac{t'}{\rho\sqrt{1-\rho^2}} \ge \frac{2}{1+\rho}v' + \frac{2\rho t'}{(1+\rho)\sqrt{1-\rho^2}} \ge \frac{2t'}{\sqrt{1-\rho^2}}$. The middle term is the symmetric axis of Equation (56), and can be used exchangably with $\sqrt{r_2(v',t')}$ as we have noted in the proof of Theorem 17.

- If $\sqrt{r} \le v' + \frac{t'}{\rho\sqrt{1-\rho^2}}$, we need $d_{|\rho|}^2(D, (\sqrt{r}, \rho\sqrt{r})) \ge (1-\rho^2)(1-\vartheta)$, i.e. $\sqrt{r} \ge v' + \frac{\rho t'}{\sqrt{1-\rho^2}} + \sqrt{1-\vartheta-t'^2}$, which gives a contradiction.

- If $\sqrt{r} > v' + \frac{t'}{\rho\sqrt{1-\rho^2}}$, because $d_{|\rho|}^2(D, (\sqrt{r}, \rho\sqrt{r})) \ge (1-\rho^2)^2 \left[\sqrt{r}-v'\right]^2$, it is the same contradiction.

When the case of $\sqrt{r} > \frac{t\sqrt{1-\rho^2}}{\rho(1-\rho)}$ in $\vartheta + f_1 \ge 1$ is tight and active, we have $\sqrt{r} = \sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{t'}{\sqrt{1-\rho^2}}$. Then we discuss the conditional expression of $f_2$ in $\vartheta + f_2 \ge 1$; given the intractability of $\sqrt{r_2(\vartheta,\rho)}$, we work with the alternative $\frac{2}{1+\rho}v' + \frac{2\rho t'}{(1+\rho)\sqrt{1-\rho^2}}$ instead.

- When $\sqrt{r} \ge v' + \frac{t'}{\rho\sqrt{1-\rho^2}}$ in $f_2$, we need $\sqrt{r} \ge \sqrt{\frac{1-\vartheta}{1-\rho^2}} + v'$ which gives a contradiction.

- When $\frac{2}{1+\rho}v' + \frac{2\rho t'}{(1+\rho)\sqrt{1-\rho^2}} \le \sqrt{r} < v' + \frac{t'}{\rho\sqrt{1-\rho^2}}$ in $f_2$, we actually cannot have $\sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{t'}{\sqrt{1-\rho^2}} \ge \frac{2}{1+\rho}v' + \frac{2\rho t'}{(1+\rho)\sqrt{1-\rho^2}}$, because it means $\left(\sqrt{\frac{1-\vartheta}{1-\rho^2}} - v'\right) + \frac{1-\rho}{1+\rho}\frac{t'}{\sqrt{1-\rho^2}} \ge \frac{1-\rho}{1+\rho}v'$.

When the case of $\sqrt{r} \ge v' + \frac{t'}{\rho\sqrt{1-\rho^2}}$ in $FN_1$ is tight and active, we have $\sqrt{r} = \sqrt{\frac{1-\vartheta}{1-\rho^2}} + v'_{\min}$. This case does not have any problem or contradiction itself. However, with $v' \ge \max\{1, \sqrt{\frac{1-\vartheta}{1-\rho^2}}\}$, it is too large, much larger than $\sqrt{r} = \max\left\{v'_{\min} + \sqrt{1-\vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-\rho}}, \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + v'_{\min}\right\}$ whose sufficiency has been proven.

When $v \ge \frac{\sqrt{2(1+\rho)}-1}{\rho} \cdot \frac{t}{\sqrt{1-\rho^2}}$, there is no more cases in $f_2$, and our discussion is finished. When $v < \frac{\sqrt{2(1+\rho)}-1}{\rho} \cdot \frac{t}{\sqrt{1-\rho^2}}$, we need to look at the last case of $\sqrt{r_2(v',t')} \le \sqrt{r} \le v' + \frac{t'}{\rho\sqrt{1-\rho^2}}$ in $FN_1$. If this case is tight and active, we have $\sqrt{r} = v' + \frac{\rho t'}{\sqrt{1-\rho^2}} + \sqrt{1-\vartheta-t'^2}$.

- It can be verified that $\sqrt{r_2(v',t')} \ge \frac{t\sqrt{1-\rho^2}}{\rho(1-\rho)}$, which is equivalent to

$$-\frac{1+\rho}{2}v'^2 + \frac{2v't'}{\sqrt{1-\rho^2}}\left(1 + \frac{1-\rho^3}{2\rho}\right) \ge \frac{t'^2}{1-\rho^2}\left[\frac{3}{2} + \frac{1}{\rho} - \frac{\rho}{2} - \rho^2 - \frac{(1-\rho^2)^2}{4\rho^2}\right]$$

$(RHS - LHS)$ is decreasing in $\frac{t'}{v'}$, and the inequality holds as $\frac{t'}{v'} = \frac{\rho\sqrt{1-\rho^2}}{\sqrt{2(1+\rho)}-1}$.

- As a result, we need to verify $v' + \frac{\rho t'}{\sqrt{1-\rho^2}} + \sqrt{1 - \vartheta - t'^2} \geq \sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{t'}{\sqrt{1-\rho^2}}$ implied by $\vartheta + f_1 \geq 1$. This would give us a contradiction, because actually $v' + \frac{\rho t'}{\sqrt{1-\rho^2}} + 1.02 \cdot \sqrt{1 - \vartheta - t'^2} \leq \sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{t'}{\sqrt{1-\rho^2}}$. In fact, We only need to prove $\left[ \frac{\sqrt{2(1+\rho)}-1}{\rho} + \rho - 1 \right] \frac{t'}{\sqrt{1-\rho^2}} + 1.02 \cdot \sqrt{1 - \vartheta - t'^2} \leq \sqrt{\frac{1-\vartheta}{1-\rho^2}}$. The coefficient $1.02$ as to make the LHS decreasing in $t'$ for $\rho \geq 0.576$. Taking $t = \frac{\rho\sqrt{1-\vartheta}}{\sqrt{2(1+\rho)-1}}$ proves the inequality.

So far our discussion is finally finished, and we have proven the phase curve to be

$$\sqrt{r} = \max\left\{ v'_{\min} + \sqrt{1 - \vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-\rho}}, \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + v'_{\min} \right\}.$$

where $v'_{\min}$ is defined in Defition 3.

Reviewing the proof for $\rho \geq 0$, we notice that $2\vartheta + f_3(\sqrt{r}, t'v') \geq 1$ is only used at the very start of the proof, and does not change the bulk of the discussion. It can be proved with vitually the same proof, that when $\rho < 0$, the phase curve is

$$\sqrt{r} = \max\left\{ v'_{\min} + \sqrt{1 - \vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-|\rho|}}, \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + v'_{\min}, \sqrt{\frac{1-2\vartheta}{2(1-|\rho|)}} + \frac{t'_{\min}}{1-|\rho|} \right\}.$$

$\square$

*Summarising the first three cases:* When $\rho \geq 0$, among the first three cases, we can take the minimum over Equation (52),(54),(57). In fact, the minimum is just Equation (54), which is

$$\sqrt{r} = \max\left\{ 1 + \sqrt{1 - \vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-\rho}}, \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \sqrt{\frac{1-\vartheta}{1-\rho^2}} \right\}. \tag{59}$$

When $\rho < 0$, we also take the minimum over Equation (53),(55),(58). In fact, in the region $\vartheta \in (\frac{1}{2}, 1)$, Equation (55) is the minimum, but when $\vartheta \leq \frac{1}{2}$, Equation (58) is the minimum. As a result, we have an upper bound on the final phase curve, which can be expressed as:

$$\sqrt{r} = \max\left\{ v'_{\min} + \sqrt{1 - \vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-|\rho|}}, \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + v'_{\min}, \sqrt{\frac{1-2\vartheta}{2(1-|\rho|)}} + \frac{t'_{\min}}{1-|\rho|} \right\}. \tag{60}$$

where we define $v'_{\min} = \max\{1, \sqrt{\frac{1-\vartheta}{1-\rho^2}}\}$ and $t'_{\min} = \max\left\{ \frac{\sqrt{2}}{2}, \frac{v'_{\min}}{1+|\rho|/\sqrt{1-\rho^2}} \right\}$.

**Remark 5.** *Equation (59) and (60) is the result we presented as Theorem 6 in the main text.*

*When $\rho \geq 0$, since we used to define $v'_{\min} = \max\left\{ 1, \sqrt{\frac{1-\vartheta}{1-\rho^2}}, \frac{\sqrt{2}/2}{\sqrt{1-\rho^2}} \right\}$ in the third case, Equation (58) is strictly above Equation (55).*

*When $\rho < 0$, in Equation (58), we used to define $v'_{\min} = \max\left\{ 1, \sqrt{\frac{1-\vartheta}{1-\rho^2}}, \frac{\sqrt{2}/2}{\sqrt{1-\rho^2}} \right\}$. When $\vartheta \leq \frac{1}{2}$, it is equivalent to $v'_{\min} = \max\{1, \sqrt{\frac{1-\vartheta}{1-\rho^2}}\}$, which agrees with Equation (60) for $\vartheta \leq \frac{1}{2}$ and $\rho \leq 0$.*

*When $\rho < 0$ and $\vartheta > \frac{1}{2}$, Equation (58) is not the minimum among the three, mainly because $\frac{\sqrt{2}/2}{\sqrt{1-\rho^2}} > 1$ for $|\rho| > \frac{\sqrt{2}}{2}$. The lowest phase curve for $\vartheta > \frac{1}{2}$ and $\rho \in (-1, 1)$ should be $\max\{1 + \sqrt{1 - \vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-|\rho|}}\}$, but we can also write it equivalently as $\max\left\{ \max\{1, \sqrt{\frac{1-\vartheta}{1-\rho^2}}\} + \sqrt{1 - \vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-|\rho|}} \right\}$, because $1 + \sqrt{1 - \vartheta} \geq \sqrt{\frac{2(1-\vartheta)}{1-|\rho|}}$ and $\vartheta > \frac{1}{2}$ together imply $\sqrt{\frac{1-\vartheta}{1-\rho^2}} \leq 1$.*

**The last three cases:** We are still left to discuss the rest three cases: (i) $t'(1 + \frac{|\rho|}{\sqrt{1-\rho^2}}) \leq v' \leq$ $\frac{t'\sqrt{1-\rho^2}}{1-|\rho|}$, (ii) $\frac{t'\sqrt{1-\rho^2}}{1-|\rho|} \leq v' \leq \frac{t'}{1-|\rho|}$, (iii) $v' \geq \frac{t'}{1-|\rho|}$.

When $\rho \geq 0$, we can prove that Equation (59) is already the best, and there is no need to discuss the rest three cases for $\rho \geq 0$. This is because in Figure 13d, 13e and 13f, we need $v' \geq \max\left\{1, \sqrt{\frac{1-\vartheta}{1-\rho^2}}\right\}$. We also need at least that $\sqrt{r} \geq v' + \sqrt{1-\vartheta}$, $\sqrt{r} \geq \sqrt{\frac{2(1-\vartheta)}{1-\rho}}$ and $\sqrt{r} \geq \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + v'$, so it cannot be any better than Equation (59).

When $\rho < 0$, the optimal phase curve for $\vartheta \leq \frac{1}{2}$ may still be one of the last three cases, but the discussion is too difficult. Even the expression of the phase curves is very complicated. We present the phase curves of the rest three cases without proof:

*Case 4.* When $t'\left(1 + \frac{|\rho|}{\sqrt{1-\rho^2}}\right) \leq v' \leq \frac{t'\sqrt{1-\rho^2}}{1-|\rho|}$, the phase curve is

$$\sqrt{r} = \max\left\{ v'_{\min} + \sqrt{1-\vartheta}, \sqrt{\frac{2(1-\vartheta)}{1-|\rho|}}, \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + v'_{\min}(\vartheta), \sqrt{\frac{1-2\vartheta}{2(1-|\rho|)}} + \frac{t'_{\min}(\vartheta)}{1-|\rho|} \right\}.$$

(61)

and the definition of $(v'_{\min}, t'_{\min})$ is specific to this case.

$$v_{\min}(\vartheta) = \max\left\{1, \sqrt{\frac{1-\vartheta}{1-\rho^2}}, \frac{\sqrt{2}}{2}\left(1 + \frac{|\rho|}{\sqrt{1-\rho^2}}\right)\right\}, \quad t_{\min}(\vartheta) = \max\left\{\frac{1-|\rho|}{\sqrt{1-\rho^2}}v_{\min}(\vartheta), f(|\rho|), g(\vartheta)\right\}$$

where

$$f(|\rho|) = \begin{cases} \sqrt{\frac{1-\rho^2}{2-\rho^2}} & \text{if } |\rho| \leq (\sqrt{5}-1)/2 \\ \frac{1}{1+|\rho|} & \text{if } (\sqrt{5}-1)/2 \leq |\rho| \leq \frac{1}{3}\left[-2 + (19+3\sqrt{33})^{1/3} + (19-3\sqrt{33})^{1/3}\right] \\ \frac{|\rho|}{\sqrt{1+(1+|\rho|)^2 - 2(1+|\rho|)\sqrt{1-\rho^2}}} & \text{if } |\rho| \geq \frac{1}{3}\left[-2 + (19+3\sqrt{33})^{1/3} + (19-3\sqrt{33})^{1/3}\right] \end{cases}$$

and

$$g(\vartheta) = \begin{cases} g_1(\vartheta) & \vartheta \geq \vartheta^* \\ g_2(\vartheta) & \vartheta < \vartheta^* \end{cases}$$

where $\vartheta^*$, $g_1(\vartheta)$ and $g_2(\vartheta)$ are respectively the roots of $\vartheta$ of the following three equations:

$$\vartheta = \vartheta^* : \frac{|\rho|}{1-|\rho|} + \sqrt{\frac{1-2\vartheta}{2(1-|\rho|)}} - \frac{1+\rho^2}{\sqrt{1-\rho^2}} - \sqrt{1-\vartheta-\rho^2} = 0$$

$$t = g_1(\vartheta) : t\left(\frac{1}{1-|\rho|} - \frac{2|\rho|}{\sqrt{1-\rho^2}}\right) + \sqrt{\frac{1-2\vartheta}{2(1-|\rho|)}} - \sqrt{1-t^2} - \sqrt{1-\vartheta-t^2} = 0$$

$$t = g_2(\vartheta) : t\left(\frac{1}{1-|\rho|} - \frac{|\rho|}{\sqrt{1-\rho^2}}\right) + \sqrt{\frac{1-2\vartheta}{2(1-|\rho|)}} - \frac{1}{\sqrt{1-\rho^2}} - \sqrt{1-\vartheta-t^2} = 0$$

All three equations can be solved easily with bi-section methods.

*Case 5.* When $\frac{t'\sqrt{1-\rho^2}}{1-|\rho|} \leq v' \leq \frac{t'}{1-|\rho|}$: We first define $v_{\min}(\vartheta) = \max\left\{1, \sqrt{\frac{1-\vartheta}{1-\rho^2}}, \frac{\sqrt{1-\rho^2}}{1-|\rho|} \cdot f(|\rho|), \right\}$ in which $f(|\rho|)$ has the same definition from *Case 4*.

When $|\rho| \leq \frac{\sqrt{2}}{2}$, the boundary is

$$\sqrt{r} = \max\left\{v_{\min}(\vartheta) + \sqrt{1-\vartheta}, h_1(\vartheta)\right\}$$

where $h_1(\vartheta) = \min\left\{Slope(\vartheta) \cdot \sqrt{1-\vartheta}, h_2(\vartheta), \max\left\{\sqrt{\frac{1-2\vartheta}{2(1-|\rho|)}} + \frac{g_1(\vartheta)}{1-|\rho|}, \frac{1+2|\rho|}{\sqrt{2-\rho^2}} + \sqrt{1-\vartheta-\frac{1-\rho^2}{2-\rho^2}}\right\}\right\}$,

in which

$$Slope(\vartheta) = 1 + \frac{\sqrt{1-\rho^2}}{1-|\rho|} \cdot t^*, \text{ where } t^* \in (0,1) \text{ solves } \frac{|\rho|t}{\sqrt{1-\rho^2}} + \sqrt{1-t^2} = 1,$$

and

$$h_2(\vartheta) = \begin{cases} \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{\sqrt{1-\rho^2}} & \text{if } \vartheta \le 1 - \frac{1}{\rho^2(1+|\rho|)^2} \\ \frac{1+2|\rho|}{(1+|\rho|)\sqrt{1-\rho^2}} + \sqrt{1-\vartheta-\frac{1}{(1+|\rho|)^2}} & \text{if } \vartheta > 1 - \frac{1}{\rho^2(1+|\rho|)^2} \end{cases}$$

and $g_1(\vartheta)$ is the same one in *Case 4*.

We define a numerical special numerical value for $|\rho|$: $|\rho| = 0.7544$. It is the value which makes $\sqrt{r} = \frac{1+2|\rho|}{(1+|\rho|)\sqrt{1-\rho^2}} + \sqrt{1-\vartheta-\frac{1}{(1+|\rho|)^2}}$ and $\sqrt{r} = \left(1 + \frac{\sqrt{1-\rho^2}}{1-|\rho|}\right)\sqrt{1-\vartheta}$ intersect at $\vartheta = \frac{1}{2}$.

When $\frac{\sqrt{2}}{2} < |\rho| \le 0.7544$, the boundary is

$$\sqrt{r} = \max\left\{v_{\min}(\vartheta) + \sqrt{1-\vartheta}, h_3(\vartheta)\right\},$$

where the definiton of $v_{\min}$ is unchanged; $h_3(\vartheta) = \min\left\{\left(1 + \frac{\sqrt{1-\rho^2}}{1-|\rho|}\right)\sqrt{1-\vartheta}, h_2(\vartheta)\right\}$.

When $|\rho| > 0.7544$, the boundary is:

$$\sqrt{r} = \begin{cases} \max\left\{v_{\min}(\vartheta) + \sqrt{1-\vartheta}, \left(1 + \frac{\sqrt{1-\rho^2}}{1-|\rho|}\right)(1-\vartheta)\right\} & \text{if } \vartheta \le \frac{1}{2} \\ h_2(\vartheta) & \text{if } \vartheta > \frac{1}{2} \end{cases}$$

*Case 6.* When $v' \ge \frac{t'}{1-|\rho|}$

$$\sqrt{r} = \max\left\{\sqrt{1-\vartheta} + \frac{1}{\sqrt{1-\rho^2}}, \sqrt{\frac{1-2\vartheta}{1-\rho^2}} + \frac{1}{\sqrt{1-\rho^2}}, h(\vartheta)\right\}$$

where the curve $h(\vartheta)$ is defined as

$$h(\vartheta) = \begin{cases} \sqrt{\frac{1-\vartheta}{1-\rho^2}} + \frac{1}{\sqrt{1-\rho^2}} & \text{if } \vartheta \le 1 - \frac{1-|\rho|}{\rho^2(1+|\rho|)} \\ \frac{1}{\sqrt{1-\rho^2}} + \frac{|\rho|}{1+|\rho|} + \sqrt{1-\vartheta-\frac{1-|\rho|}{1+|\rho|}} & \text{if } \vartheta > 1 - \frac{1-|\rho|}{\rho^2(1+|\rho|)} \end{cases}.$$

To sum up all the six cases, the lowest phase curve over the six cases in given in Equation (59) when $\rho \ge 0$. When $\rho < 0$, the optimal curve is too complicated, but an upper bound is given by Equation (60).

## I  PROOF OF THEOREM 7

The key is to analyze the random-design setting and show that its minimax rate of Hamming error is only determined by $\mathbb{E}[X'X] = \Sigma$. Then, when we switch to the fixed-design case of $X'X = \Sigma$, the same minimax rate holds. For the random-design setting, we proceed by deriving a lower bound and an upper bound of the minimax Hamming error separately.

First, we derive a lower bound for the minimax Hamming error. Let $G = X'X$ denote the Gram matrix of the random-design model. Fixing any two subsets $V_0, V_1 \subset \{1, 2, \ldots, p\}$, we write $V = V_0 \cup V_1$. Let $\eta \in \{0, 1\}^p$ be an arbitrary binary vector. We consider two binary vectors $\mu^{(0)}, \mu^{(1)} \in \{0, 1\}^p$ where $\mu_j^{(0)} = \mu_j^{(1)} = \eta_j$, for $j \notin V$, and restricted on $V$, $\text{Supp}(\mu_V^{(0)}) = V_0$ and $\text{Supp}(\mu_V^{(1)}) = V_1$. Let $\tau_p = \sqrt{2r\log(p)}$. Consider the testing problem

$$H_0 : \beta = \tau_p \mu^{(0)}, \qquad v.s. \qquad H_1 : \beta = \tau_p \mu^{(1)}. \tag{62}$$

For a test $T$, let $R(T)$ be the sum of type I and type II errors. Any selector $\hat{\beta}$ can be converted to a test $T(\hat{\beta})$, where we reject the null hypothesis if $\mathrm{Supp}(\hat{\beta}) \neq V_0$. It is seen that $R(T(\hat{\beta})) = \mathbb{P}\{\beta = \tau_p \mu^{(0)}, \mathrm{Supp}(\hat{\beta}) \neq V_0\} + \mathbb{P}\{\beta = \tau_p \mu^{(1)}, \mathrm{Supp}(\hat{\beta}) = V_0)\} \leq \sum_{j \in V} \{\mathbb{P}(\beta_j = 0, \hat{\beta}_j \neq \tau_p) + \mathbb{P}(\beta_j = \tau_p, \hat{\beta}_j = 0)\}$. It follows that

$$\mathbb{E}[H(\hat{\beta}_V, \beta_V)|X] \geq R(T(\hat{\beta})) \geq \inf_T R(T) \equiv R^*(V_0, V_1; \eta, X). \tag{63}$$

We can compute the right hand side using the Neyman-Pearson lemma. Define

$$a = a(V_0, V_1, X) = (\mu^{(0)} - \mu^{(1)})' G (\mu^{(0)} - \mu^{(1)}). \tag{64}$$

The likelihood ratio test for (62) is equivalent to using the test statistic $Z = a^{-1/2}(\mu^{(1)} - \mu^{(0)})' X'(y - \tau_p X \mu^{(0)})$. Then, $Z \sim \mathcal{N}(0, 1)$ under $H_0$, and $Z \sim \mathcal{N}(a^{1/2}\tau_p, 1) = \mathcal{N}(\sqrt{2ar \log(p)}, 1)$, under $H_1$. By Neyman-Pearson lemma,

$$R^*(V_0, V_1; \eta, X) = \inf_t \left\{ \epsilon_p^{|V_0|} \cdot \mathbb{P}(\mathcal{N}(0, 1) > t) + \epsilon_p^{|V_1|} \cdot \mathbb{P}(\mathcal{N}(\sqrt{2ar \log(p)}, 1) < t) \right\}$$

$$= \inf_{t = \sqrt{2q \log(p)}} \left\{ L_p p^{-|V_0|\vartheta - q} + L_p p^{-|V_1|\vartheta - (\sqrt{ar} - \sqrt{q})_+^2} \right\}$$

$$= L_p p^{-h(V_0, V_1, X)}, \tag{65}$$

where

$$h(V_0, V_1, X) = \max_{q > 0} \left( \min \left\{ |V_0|\vartheta + q, \ |V_1|\vartheta + (\sqrt{ar} - \sqrt{q})_+^2 \right\} \right).$$

In the second line of (65), we have used the Mills' ratio of $N(0, 1)$ (e.g., see Ke et al. (2014) for a similar use of the Mills' ratio). Let $\Sigma$ be the covariance matrix, parameterized by $\rho$. We define the following quantities:

$$a^*(V_0, V_1, \rho) = (\mu^{(0)} - \mu^{(1)})' \Sigma (\mu^{(0)} - \mu^{(1)}),$$

$$h^*(V_0, V_1, \rho) = \max_{q > 0} \left( \min \left\{ |V_0|\vartheta + q, \ |V_1|\vartheta + (\sqrt{a^*r} - \sqrt{q})_+^2 \right\} \right). \tag{66}$$

Below, we show that $h(V_0, V_1, X)$ is sufficiently close to $h^*(V_0, V_1, \rho)$. The key is showing that $\Sigma$ and $G$ are sufficiently close on the diagonal block restricted to $V$. We use Theorem 5.39 and Remark 5.40 of Vershynin (2012) with $t = O(\sqrt{|V| \log(p)})$. It follows that, when $|V| \ll n$, with probability $1 - o(p^{-3 - |V|})$,

$$\|G_{V,V} - \Sigma_{V,V}\| \leq C \|\Sigma_{V,V}\| \sqrt{n^{-1}|V| \log(p)}; \quad \text{here, } C \text{ a constant independent of } |V|.$$

We note that $\|\Sigma_{V,V}\| \leq \|\Sigma\| \leq C$. For any finite integer $m \geq 1$, the total number of size-$m$ subset $V$ is $\binom{p}{m} = O(p^m)$. We then apply the probability union bound to get that, with probability $1 - O(p^{-3})$,

$$\max_{V: |V| \leq m} \|G_{V,V} - \Sigma_{V,V}\| \leq C\sqrt{n^{-1} \log(p)}. \tag{67}$$

Since $|a(V_0, V_1, X) - a^*(V_0, V_1, \rho)| \leq \|G_{V,V} - \Sigma_{V,V}\| \cdot \|\mu^{(1)} - \mu^{(0)}\|^2 \leq \|G_{V,V} - \Sigma_{V,V}\| \cdot |V|$, we immediately know that

$$|a(V_0, V_1, X) - a^*(V_0, V_1, \rho)| \leq C\sqrt{n^{-1} \log(p)} \quad \text{here, } C \text{ depends on } m. \tag{68}$$

Write $h = h(V_0, V_1, X)$ and $h^* = h^*(V_0, V_1, \rho)$ for short, and let $(h^*, a^*)$ be the shorthand notations defined similarly. Then, $h = \max_q g(q, a)$ and $h^* = \max_q f(q, a^*)$, for $f(q, a) = \min\{|V_0|\vartheta + q, |V_1|\vartheta + (\sqrt{ar} - \sqrt{q})_+^2\}$. Let $\tilde{q}$ and $\tilde{q}^*$ be the two maximizers. It is seen that $h = f(\tilde{q}, a) \leq f(\tilde{q}, a^*) + \max_q |f(q, a) - f(q, a^*)| \leq h^* + \max_q |f(q, a) - f(q, a^*)|$. Similarly, we can also derive that $h \leq h^* + \max_q |f(q, a^*) - f(q, a)|$. Combining them gives $|h - h^*| \leq \max_q |f(q, a^*) - f(q, a)|$. We plug in the expression of $f(q, a)$ to get

$$|h(V_0, V_1, X) - h^*(V_0, V_1, \rho)| \leq |\sqrt{ar} - \sqrt{a^*r}| \leq C\sqrt{n^{-1} \log(p)}.$$

We now combine all the results, and note that (67) has a maximum over all $V = V_0 \cup V_1$. It follows that, with probability $1 - O(p^{-3})$,

$$\max_{(V_0, V_1): |V_0 \cup V_1| \leq m} |h(V_0, V_1, X) - h^*(V_0, V_1, \rho)| \leq C\sqrt{n^{-1} \log(p)}. \tag{69}$$

We plug it into (65). Note that $L_p p^{-h} = L_p p^{-h^*} \cdot p^{h^* - h}$. In line of (69), $p^{h^* - h}$ is a multi-log$(p)$ term, i.e., $L_p p^{-h} = L_p p^{-h^*}$. We then combine it with (63). It yields that, with probability $1 - O(p^{-3})$,

$$\mathbb{E}[H(\hat{\beta}_V, \beta_V)|X] \geq L_p p^{-h^*(V_0, V_1, \rho)}, \text{ simultaneously for all } (V_0, V_1) \text{ with } |V_0 \cup V_1| \leq m. \quad (70)$$

Given $V$, we further take a maximum over $(V_0, V_1)$ on the right hand side. It follows that

$$\mathbb{E}[H(\hat{\beta}_V, \beta_V)|X] \geq L_p p^{-h^{**}(V, \rho)}, \qquad \text{where} \quad h^{**}(V, \rho) = \min_{\substack{(V_0, V_1): V_0 \neq V_1, \\ V_0 \cup V_1 = V}} h^*(V_0, V_1, \rho). \quad (71)$$

Write $\{1, 2, \ldots, p\} = \cup_{j=1}^{\lceil p/2 \rceil} V_j$, where $V_j = \{2j - 1, 2j\}$ for $j \leq p/2$ and $V_j = \{p\}$ for $j > p/2$ (this happens only if $p$ is odd). It follows that, with probability $1 - O(p^{-3})$,

$$\mathbb{E}[H(\hat{\beta}, \beta)|X] = \sum_{1 \leq j \leq \lceil p/2 \rceil} \mathbb{E}[H(\hat{\beta}_{V_j}, \beta_{V_j})|X] \geq \sum_{1 \leq j \leq \lceil p/2 \rceil} L_p p^{-h^{**}(V_j, \rho)}.$$

When $p$ is even, $h^{**}(V_j, \rho)$ are all equal. When $p$ is odd, $h^{**}(V_j, \rho)$ are all equal, except for one $V_j$; but this one has a negligible effect on the right hand side above. Let $h^{**}(\rho)$ be the common value of $h^{**}(V_j, \rho)$. Since $h^{**}(\rho)$ also depends on $(\vartheta, r)$, we write it as $h^{**}(\rho; \vartheta, r)$ to reflect this dependence. We immediately have that, with probability $1 - O(p^{-3})$,

$$\mathbb{E}[H(\hat{\beta}, \beta)|X] \geq L_p p^{1 - h^{**}(\rho; \vartheta, r)}.$$

On the event that the above inequality does not hold, the Hamming error is at most $p$. The contribution of this event to the expected Hamming error is at most $p \cdot O(p^{-3}) = O(p^{-2})$, which is negligible to $L_p p^{1 - h^*(\rho; \vartheta, r)}$. It follows that

$$\mathbb{E}[H(\hat{\beta}, \beta)] \geq L_p p^{1 - h^{**}(\rho; \vartheta, r)}, \qquad \text{for any method } \hat{\beta}. \quad (72)$$

This gives a lower bound for the minimax Hamming error.

Next, we give an upper bound for the minimax Hamming error. We will consider a specific $\hat{\beta}$. Let the partition $\{1, 2, \ldots, p\} = \cup_{j=1}^{\lceil p/2 \rceil} V_j$ be the same as above. For any subset $U \subset \{1, 2, \ldots, p\}$, let $\mathbf{1}_U$ be the binary vector such that its $j$th entry is 1 if $j \in U$ and 0 otherwise. Additionally, let $X_U$ be the submatrix of $X$ restricted to columns in $U$. For each $V_j$, define

$$\hat{U}_j = \arg\min_{U \subset V_j} \left\{ \frac{1}{2} \|y - \tau_p X \mathbf{1}_U\|^2 + \vartheta \log(p)|U| \right\}. \quad (73)$$

Define $\hat{\mu} \in \{0, 1\}^p$ such that for any $i \in V_j$, $\hat{\mu}_i = 1$ if $i \in \hat{U}_j$, and $\hat{\mu}_i = 0$ otherwise. The estimator is $\hat{\beta} = \tau_p \hat{\mu}$. We now calculate the expected Hamming error of this estimator. Let $S$ be the support of $\beta$. Fix $V_j$ and write $V = V_j$ for short. Given any two subsets $U_0$ and $U_1$ of $V$ such that $U_0 \neq U_1$, we consider the event

$$\text{Supp}(\beta_V) = U_0, \qquad \text{Supp}(\hat{\beta}_V) = U_1, \qquad |S| \leq 2p^{1 - \vartheta}. \quad (74)$$

On this event, it is true that

$$\frac{1}{2} \|y - \tau_p X \mathbf{1}_{U_0}\|^2 + \vartheta \log(p)|U_0| \geq \frac{1}{2} \|y - \tau_p X \mathbf{1}_{U_1}\|^2 + \vartheta \log(p)|U_1|. \quad (75)$$

Note that $y = X\beta + z = \tau_p X \mathbf{1}_{U_0} + \tau_p X \mathbf{1}_{S \cap V^c} + z$. We can re-write (75) as

$$\frac{1}{2} \|z + \tau_p X \mathbf{1}_{S \cap V^c}\|^2 + \vartheta \log(p)|U_0| \geq \frac{1}{2} \|(z + \tau_p X \mathbf{1}_{S \cap V^c}) - \tau_p X (\mathbf{1}_{U_1} - \mathbf{1}_{U_0})\|^2 + \vartheta \log(p)|U_1|.$$

Let $F = U_0 \cap U_1$, $E_0 = U_0 \setminus F$ and $E_1 = U_1 \setminus F$. Then, $|U_0| = |F| + |E_0|$ and $|U_1| = |F| + |E_1|$. We plug it into the above inequality and re-arrange the terms. It gives

$$z' X(\mathbf{1}_{U_1} - \mathbf{1}_{U_0}) \geq \frac{\tau_p}{2} (\mathbf{1}_{U_1} - \mathbf{1}_{U_0})' G(\mathbf{1}_{U_1} - \mathbf{1}_{U_0}) - \mathbf{1}'_{S \cap V^c} G(\mathbf{1}_{U_1} - \mathbf{1}_{U_0}) + \frac{\vartheta \log(p)}{\tau_p}(|E_1| - |E_0|).$$

Let $a = a(U_0, U_1, X) = (\mathbf{1}_{U_1} - \mathbf{1}_{U_0})' G(\mathbf{1}_{U_1} - \mathbf{1}_{U_0})$. We note that this definition is indeed the same as that in (64). Let $\tilde{z} = z' X(\mathbf{1}_{U_1} - \mathbf{1}_{U_0})/\sqrt{a}$. The above can be written equivalently as

$$\tilde{z} \geq \sqrt{2 \log(p)} \left[ \frac{\sqrt{ra}}{2} - \frac{\mathbf{1}'_{S \cap V^c} G(\mathbf{1}_{U_1} - \mathbf{1}_{U_0})}{\sqrt{a}} + \frac{\vartheta(|E_1| - |E_0|)}{2\sqrt{ra}} \right], \quad \text{where } \tilde{z}|(X, \beta) \sim N(0, 1). \quad (76)$$

We bound $|\mathbf{1}'_{S \cap V^c} G(\mathbf{1}_{U_1} - \mathbf{1}_{U_0})|$. Since $\Sigma_{V^c V}$ is a zero matrix, we immediately have $\mathbf{1}'_{S \cap V^c} \Sigma(\mathbf{1}_{U_1} - \mathbf{1}_{U_0}) = 0$. It follows by the triangle inequality that

$$|\mathbf{1}'_{S \cap V^c} G(\mathbf{1}_{U_1} - \mathbf{1}_{U_0})| \leq |\mathbf{1}'_{S \cap V^c}(G - \Sigma)(\mathbf{1}_{U_1} - \mathbf{1}_{U_0})|$$
$$\leq |V| \cdot \max_{k \in V} |e'_k(G - \Sigma)\mathbf{1}_{S \cap V^c}|. \tag{77}$$

For $k \in V$, $e'_k(G - \Sigma)\mathbf{1}_{S \cap V^c} = \sum_{i=1}^{n} \sum_{\ell \in S \cap V^c} \{X(i,k)X(i,\ell) - \mathbb{E}[X(i,k)X(i,\ell)]\}$. We recall that $\{1, 2, \ldots, p\} = \cup_{m=1}^{\lceil p/2 \rceil} V_m$ is a partition. It induces a partition on $S \cap V^c$, which we denote by $S \cap V^c = \cup_{m=1}^{N_p} S_m$. Each $S_m$ contains at most 2 indices and $|S \cap V^c|/2 \leq N_p \leq |S \cap V^c|$. Write

$$e'_k(G - \Sigma)\mathbf{1}_{S \cap V^c} = \sum_{i=1}^{n} \sum_{m=1}^{N_p} \Big[ \sum_{\ell \in S_m} \{X(i,k)X(i,\ell) - \mathbb{E}[X(i,k)X(i,\ell)]\} \Big].$$

The right hand side is a sum of $n N_p$ independent variables, where each variable has a zero mean and a sub-exponential norm bounded by $n^{-1} K$, for a constant $K > 0$. We apply the Bernstein inequality (Vershynin, 2012, Proposition 5.16) to get that, for every $t > 0$,

$$\mathbb{P}(|e'_k(G - \Sigma)\mathbf{1}_{S \cap V^c}| > t) \leq 2 \exp\Big(-c \min\Big\{\frac{nt^2}{K^2 N_p}, \frac{nt}{K}\Big\}\Big),$$

where $c > 0$ is a universal constant. By letting $t = C\sqrt{n^{-1} N_p \log(p)}$ for a properly large constant $C$, we have that, with probability $1 - O(p^{-3})$,

$$|e'_k(G - \Sigma)\mathbf{1}_{S \cap V^c}| \leq C\sqrt{n^{-1} N_p \log(p)} \leq C\sqrt{n^{-1}|S| \log(p)}.$$

We plug it into (76) and apply the probability union bound. We also note that $|S| = O(p^{1-\vartheta})$ on the event (74); also, $n = p^\omega$ with $\omega > 1 - \vartheta$. It follows that, on this event, with probability $1 - O(p^{-3})$,

$$|\mathbf{1}'_{S \cap V^c} G(\mathbf{1}_{U_1} - \mathbf{1}_{U_0})| \leq C p^{-(\omega-1+\vartheta)/2}\sqrt{\log(p)}. \tag{78}$$

We plug (78) into (76) to get:

$$\tilde{z} \geq \sqrt{2b_p \log(p)}, \quad \text{where } b_p = \Big[\frac{\sqrt{ra}}{2} + \frac{\vartheta(|E_1| - |E_0|)}{2\sqrt{ra}} - L_p p^{-\frac{\omega-1+\vartheta}{2}}\Big]^2.$$

Moreover, let $a^* = a^*(U_0, U_1, \rho) = (\mathbf{1}_{U_1} - \mathbf{1}_{U_0})'\Sigma(\mathbf{1}_{U_1} - \mathbf{1}_{U_0})$, which is the same as the definition in (66). By (68), the replacement of $a$ by $a^*$ only yields a difference of $L_p p^{-\frac{\omega-1+\vartheta}{2}}$ in the expression of $b_p$. We further have:

$$\tilde{z} \geq \sqrt{2b_p \log(p)}, \quad \text{where } \tilde{z}|(X, \beta) \sim N(0,1) \text{ and } b_p = \Big[\frac{\sqrt{ra^*}}{2} + \frac{\vartheta(|E_1| - |E_0|)}{2\sqrt{ra^*}} + L_p p^{-\frac{\omega-1+\vartheta}{2}}\Big]^2. \tag{79}$$

First, what (79) says is that, conditioning on $(X, \beta_V, \beta_{V^c})$, if $\|\beta\|_0 \leq C p^{1-\vartheta}$, then except for an event of probability $O(p^{-3})$, $\text{Supp}(\hat{\beta}_V) = U_1$ implies $\tilde{z} > \sqrt{2b_p \log(p)}$. Second, under our model, $(X, \beta_{V^c})$ are independent of $\beta_V$, and $\mathbb{P}(\text{Supp}(\beta_V) = U_0) = L_p p^{-\vartheta|U_0|} = L_p p^{-\vartheta(|F| + |E_0|)}$. Last, $\mathbb{P}(\|\beta_{V^c}\|_0 \leq 2p^{1-\vartheta}) = 1 - O(p^{-3})$ (this is seen by noticing that $\|\beta_{V^c}\|_0$ is the sum of independent Bernoulli variables and by applying the Bernstein's inequality). We combine the above to get

$$\mathbb{P}\big(\text{Supp}(\beta_V) = U_0, \text{Supp}(\hat{\beta}_V) = U_1, |S| \leq 2p^{1-\vartheta}\big)$$
$$\leq L_p p^{-\vartheta|U_0|} \cdot \mathbb{P}\Big(\tilde{z} \geq \sqrt{2b_p \log(p)}\Big) + O(p^{-3})$$
$$\leq L_p p^{-\vartheta(|F| + |E_0|) - \big[\frac{\sqrt{ra^*}}{2} + \frac{\vartheta(|E_1| - |E_0|)}{2\sqrt{ra^*}}\big]_+^2}.$$

By elementary calculations, we have

$$\vartheta(|F| + |E_0|) + \Big[\frac{\sqrt{ra^*}}{2} + \frac{\vartheta(|E_1| - |E_0|)}{2\sqrt{ra^*}}\Big]_+^2$$
$$\geq \vartheta|F| + \max\{|E_0|, |E_1|\}\vartheta + \frac{1}{4}\Big(\sqrt{ra^*} - \frac{|(|E_1| - |E_0|)|}{\sqrt{ra^*}}\Big)_+^2$$
$$= \max\{|U_0|, |U_1|\}\vartheta + \frac{1}{4}\Big(\sqrt{ra^*} - \frac{|(|U_1| - |U_0|)|}{\sqrt{ra^*}}\Big)_+^2 = h^*(U_0, U_1, \rho),$$

where $h^*(U_0, U_1, \rho)$ is the same as that defined in (66) (the last equality above follows by solving $q$ in (66)). We combine the above to get

$$\mathbb{P}\big(\mathrm{Supp}(\beta_V) = U_0,\ \mathrm{Supp}(\hat{\beta}_V) = U_1,\ |S| \leq 2p^{1-\vartheta}\big) \leq L_p p^{-h^*(U_0, U_1, \rho)}. \qquad (80)$$

On the above event, the Hamming error contributed by $\hat{\beta}_V$ is $|E_0| + |E_1| \leq |V| \leq 2$. Moreover, $h^*(U_0, U_1, \rho) \geq h^{**}(\rho; \vartheta, r)$, where the latter is defined in (71). It follows that

$$\mathbb{E}[H(\hat{\beta}_V, \beta_V)] = \sum_{(U_0, U_1)} 2 \cdot \mathbb{P}\big(\mathrm{Supp}(\beta_V) = U_0,\ \mathrm{Supp}(\hat{\beta}_V) = U_1,\ |S| \leq 2p^{1-\vartheta}\big) + O(p^{-3})$$

$$\leq L_p \sum_{(U_0, U_1)} p^{-h^*(U_0, U_1, \rho)} \leq L_p p^{-h^{**}(\rho; \vartheta, r)}.$$

The above is true for every $V = V_j$ in the partition $\{1, 2, \ldots, p\} = \cup_{j=1}^{\lceil p/2 \rceil}$ (except for the last $V_j$ in the case that $p$ is odd; but this single $V_j$ has a negligible effect on the rate of the Hamming error). We immediately have

$$\mathbb{E}[H(\hat{\beta}, \beta)] = \sum_{1 \leq j \leq \lceil p/2 \rceil} \mathbb{E}[H(\hat{\beta}_V, \beta_V)] \leq L_p p^{1-h^{**}(\rho; \vartheta, r)}, \quad \text{for the } \hat{\beta} \text{ in (73).} \qquad (81)$$

This gives an upper bound for the minimax Hamming error.

Last, we use (72) and (81) to show the claim. By combining these two inequalities, we know that, for the random design,

$$\inf_{\hat{\beta}} \mathbb{E}[H(\hat{\beta}, \beta)] = L_p p^{1-h^{**}(\rho; \vartheta, r)}.$$

A key observation is that the exponent $h^{**}(\rho; \vartheta, r)$ is only related to $\Sigma$, not the realization of $X'X$. Now, we can mimic all the above derivations to get the same conclusion when the Gram matrix is fixed at $\Sigma$ (the proof is very similar, except that we now have $G = \Sigma$). Therefore, the minimax rates of the Hamming errors under two settings are exactly the same. $\qquad \square$