The Price of Competition: *Effect Size Heterogeneity* Matters in High Dimensions!

joint work with Yachong Yang and Weijie Su

Hua Wang

The Wharton School, University of Pennsylvania

June 2, 2020

Settings: Model selection in high dimensions

- High-dimensional linear regression

$$\underset{n \times 1}{\boldsymbol{y}} = \underset{n \times p}{\boldsymbol{X}} \underset{p \times 1}{\boldsymbol{\beta}} + \underset{n \times 1}{\boldsymbol{z}}$$

- An important question of great practical value is model selection.
- How hard is model selection?

Settings: Model selection in high dimensions

- High-dimensional linear regression

$$\underset{n \times 1}{\boldsymbol{y}} = \underset{n \times p}{\boldsymbol{X}} \underset{p \times 1}{\boldsymbol{\beta}} + \underset{n \times 1}{\boldsymbol{z}}$$

- An important question of great practical value is model selection.
- How hard is model selection?
  An intuitive answer: It depends on sparsity (as long as signals are large enough, e.g. beta-min).

Performance criteria: FDP and TPP

- Relevant variables (or signals).

$$S = \{j : \ \beta_j \neq 0\}$$

Performance criteria: FDP and TPP

- Relevant variables (or signals).

$$S = \{j : \ \beta_j \neq 0\}$$

- Discoveries, or model selected at λ

$$\widehat{S} = \{j : \ \hat{\beta}_j(\lambda) \neq 0\}$$

Performance criteria: FDP and TPP

- Relevant variables (or signals).

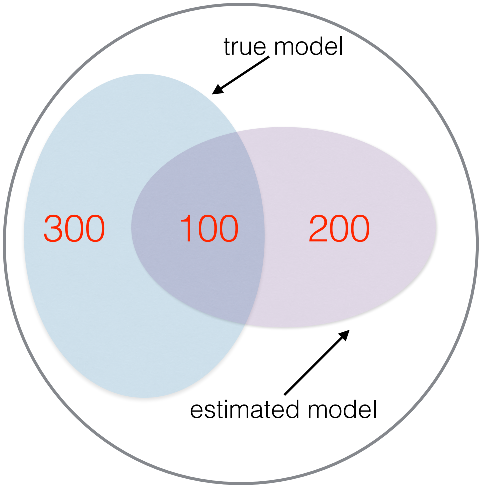
$$S = \{j : \ \beta_j \neq 0\}$$

- Discoveries, or model selected at λ

$$\widehat{S} = \{j : \ \hat{\beta}_j(\lambda) \neq 0\}$$

$$\mathrm{FDP}(\lambda) := \frac{\#\{j : j \in \widehat{S}, \beta_j = 0\}}{\#\widehat{S}} = \frac{200}{100 + 200}$$

$$\mathrm{TPP}(\lambda) := \frac{\#\{j : j \in \widehat{S}, \beta_j \neq 0\}}{\#\{j : \beta_j \neq 0\}} = \frac{100}{300 + 100}$$



Folklore theorem of signal strength

When $p > n$, Lasso is the popular method to do variable selection.

Folklore theorem of signal strength

When $p > n$, Lasso is the popular method to do variable selection.

Belief (Some folks, nowadays)

- *With $\|\beta\|_0$ fixed, the stronger all signals are, the better a model selector (e.g. Lasso) will perform.*

Folklore theorem of signal strength

When $p > n$, Lasso is the popular method to do variable selection.

BELIEF (SOME FOLKS, NOWADAYS)

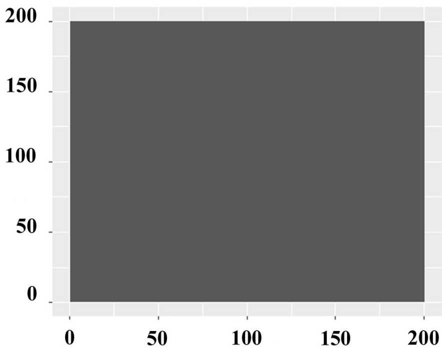
- *With $||\beta||_0$ fixed, the stronger all signals are, the better a model selector (e.g. Lasso) will perform.*

Is it really the case?

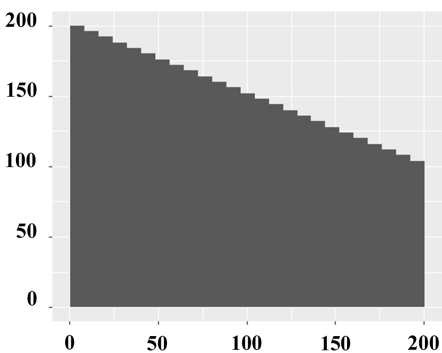
In which setting does Lasso perform best in?

$n = 1000, p = 1000, s = 200$, with weak noise $\sigma = 0.01$. The structure of signals:

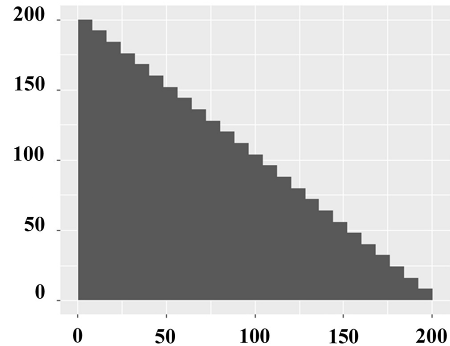
Setting 1: Strongest.



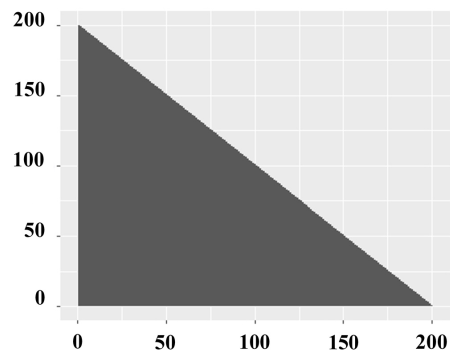
Setting 2: Strong.



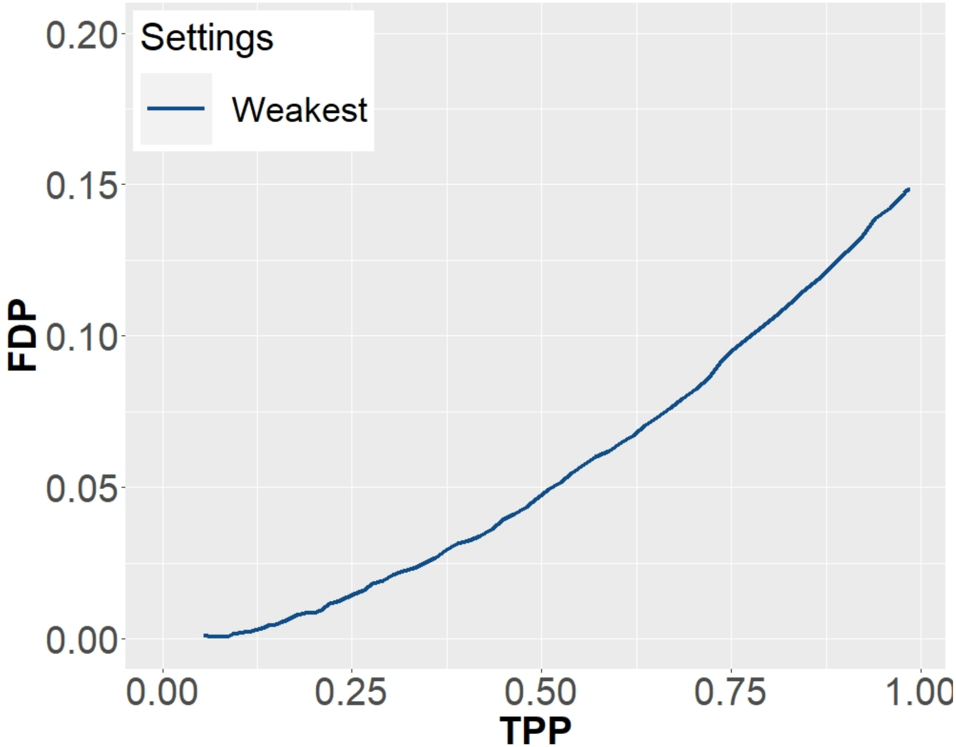
Setting 3: Weak.



Setting 4: Weakest.

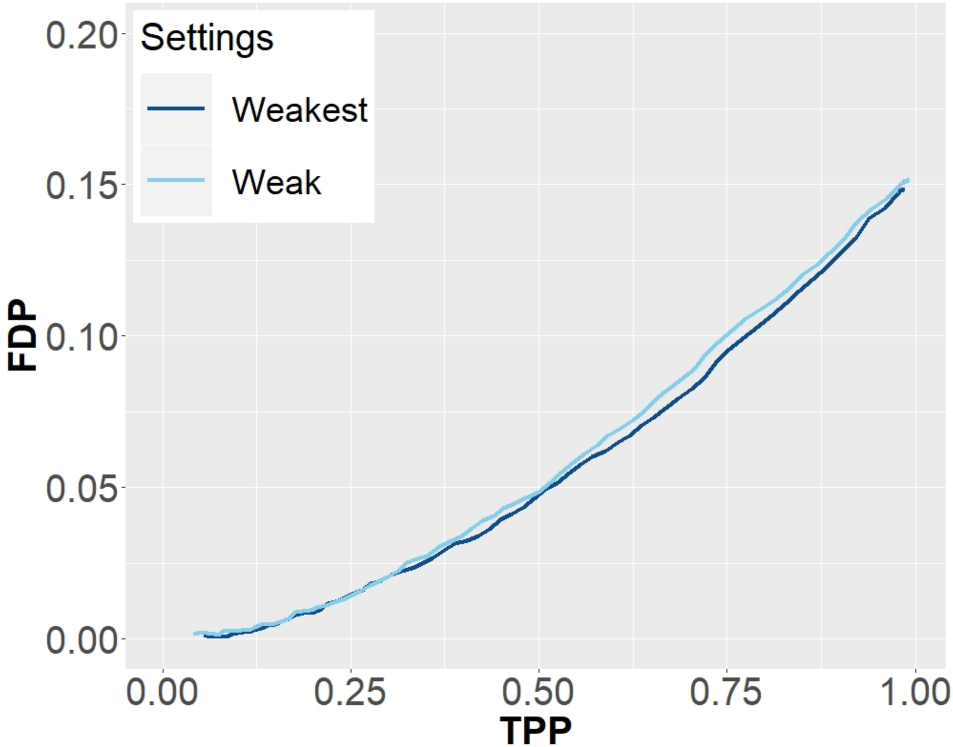


The result...



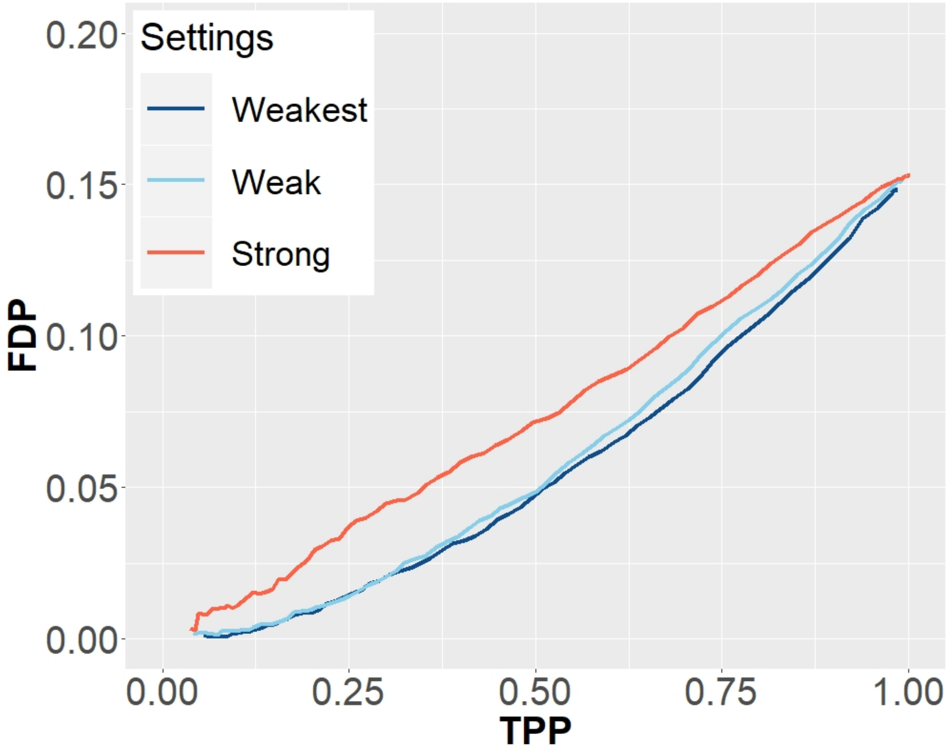
The tpp and fdp are calculated along Lasso path with λ varies from ∞ to 0.

Surprisingly...



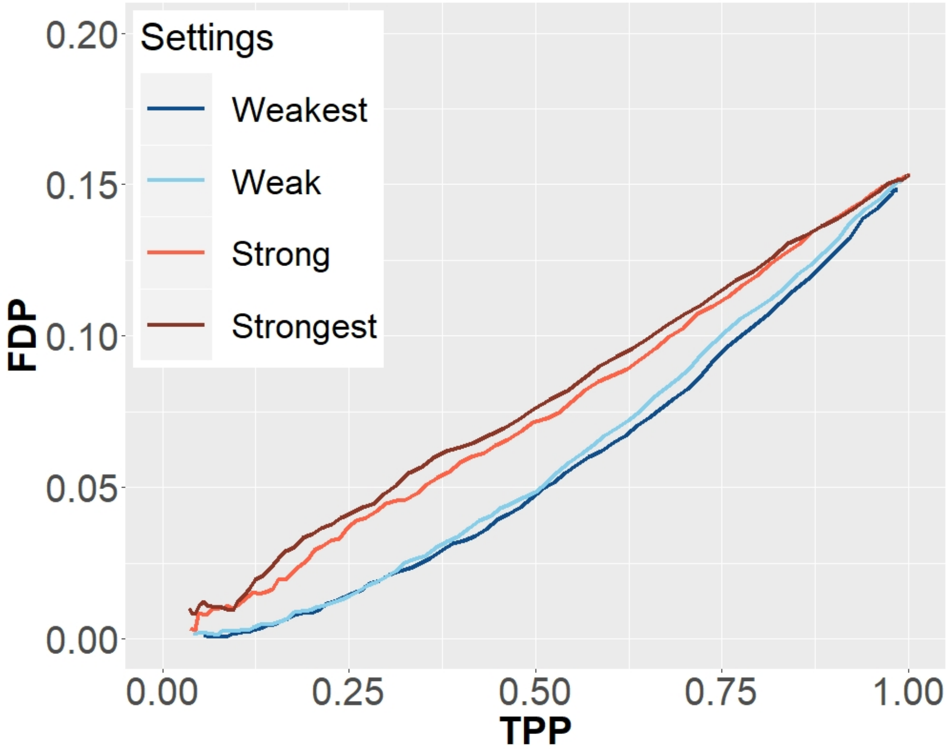
The tpp and fdp are calculated along Lasso path with λ varies from ∞ to 0.

Surprisingly...



The TPP and FDP are calculated along Lasso path with λ varies from ∞ to 0.

Surprisingly...



The TPP and FDP are calculated along Lasso path with λ varies from ∞ to 0.

Lasso prefers weak signals??

- Everything (including sparsity) except the strength of the signals are the same. The Lasso perform better with weaker signals!

Lasso prefers weak signals??

- Everything (including sparsity) except the strength of the signals are the same. The Lasso perform better with weaker signals!
- Our explanation: Lasso favors strong signals as we expected, but it "prefers" signals that are **wildly differing with each other**.

Lasso prefers weak signals??

- Everything (including sparsity) except the strength of the signals are the same. The Lasso perform better with weaker signals!
- Our explanation: Lasso favors strong signals as we expected, but it “prefers” signals that are **wildly differing with each other**.
- We term this diverse structure of signals as “Effect Size Heterogeneity”.

Lasso prefers weak signals??

- Everything (including sparsity) except the strength of the signals are the same. The Lasso perform better with weaker signals!
- Our explanation: Lasso favors strong signals as we expected, but it "prefers" signals that are **wildly differing with each other**.
- We term this diverse structure of signals as "Effect Size Heterogeneity".
- With everything else fixed, Lasso performs the best with the most heterogeneous signals.

Effect size heterogeneity matters!

- Everything (including sparsity) except the strength of the signals are the same. The Lasso perform better with weaker signals!
- Our explanation: Lasso favors strong signals as we expected, but it "prefers" signals that are **wildly differing with each other**.
- We term this diverse structure of signals as "Effect Size Heterogeneity".
- With everything else fixed, Lasso performs the best with the most heterogeneous signals.

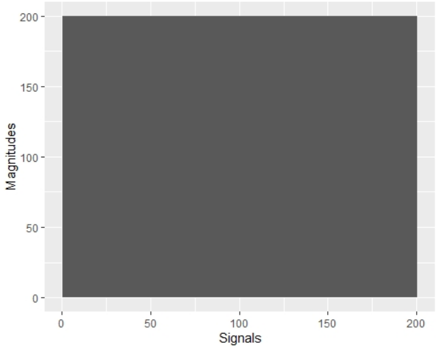
Effect size heterogeneity matters!

- Everything (including sparsity) except the strength of the signals are the same. The Lasso perform better with weaker signals!
- Our explanation: Lasso favors strong signals as we expected, but it “prefers” signals that are **wildly differing with each other**.
- We term this diverse structure of signals as “Effect Size Heterogeneity”.
- With everything else fixed, Lasso performs the best with the most heterogeneous signals.

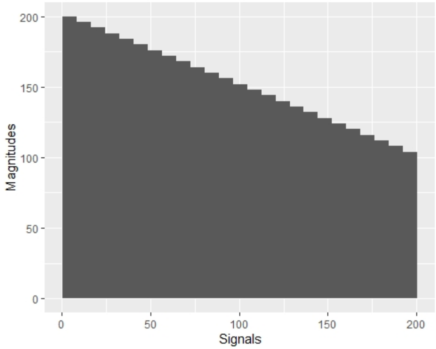
Effect Size Heterogeneity matters!

Which setting will Lasso perform best in? (Re-visit)

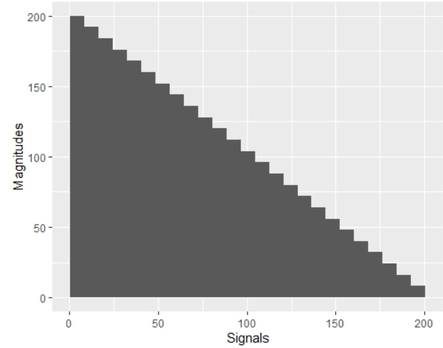
Setting 1: Most Homogeneous



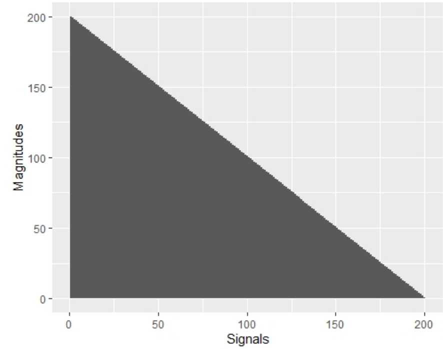
Setting 2: Homogeneous.



Setting 3: Heterogeneous.



Setting 4: Most Heterogeneous.



Theory of Lasso in literature

Belief (Literature[1], nowadays (informal))

Given the information of $k = \|\beta\|_0$, and the structure of X (n, p, RIP conditions, etc.), we can understand Lasso (as a model selector) well, especially if signals are sufficiently large (beta-min condition).

[1] e.g. E. Candes, T. Tao 2007; PJ. Bickel, Y. Ritov, AB. Tsybakov 2009; MJ. Wainwright 2009...

Theory of Lasso in literature

Belief (Literature[1], nowadays (informal))

Given the information of $k = \|\beta\|_0$, and the structure of X (n, p, RIP conditions, etc.), we can understand Lasso (as a model selector) well, especially if signals are sufficiently large (beta-min condition).

Theorem (W., Yang and Su, 2020 (informal))

The information of $(\|\beta\|_0, X)$ is not enough, we need to know more about the inner structure of β.

[1] e.g. E. Candes, T. Tao 2007; PJ. Bickel, Y. Ritov, AB. Tsybakov 2009; MJ. Wainwright 2009...

Main results

Assume

- X has iid $\mathcal{N}(0, 1/n)$ entries,
- $\sigma = 0$, i.e. noise $z_i = 0$,
- regression coefficients β_i are iid from prior Π with $\mathbb{E}\Pi^2 < \infty$ and $\mathbb{P}(\Pi \neq 0) = \epsilon \in (0, 1)$,
- $n/p \to \delta \in (0, \infty)$.

Then

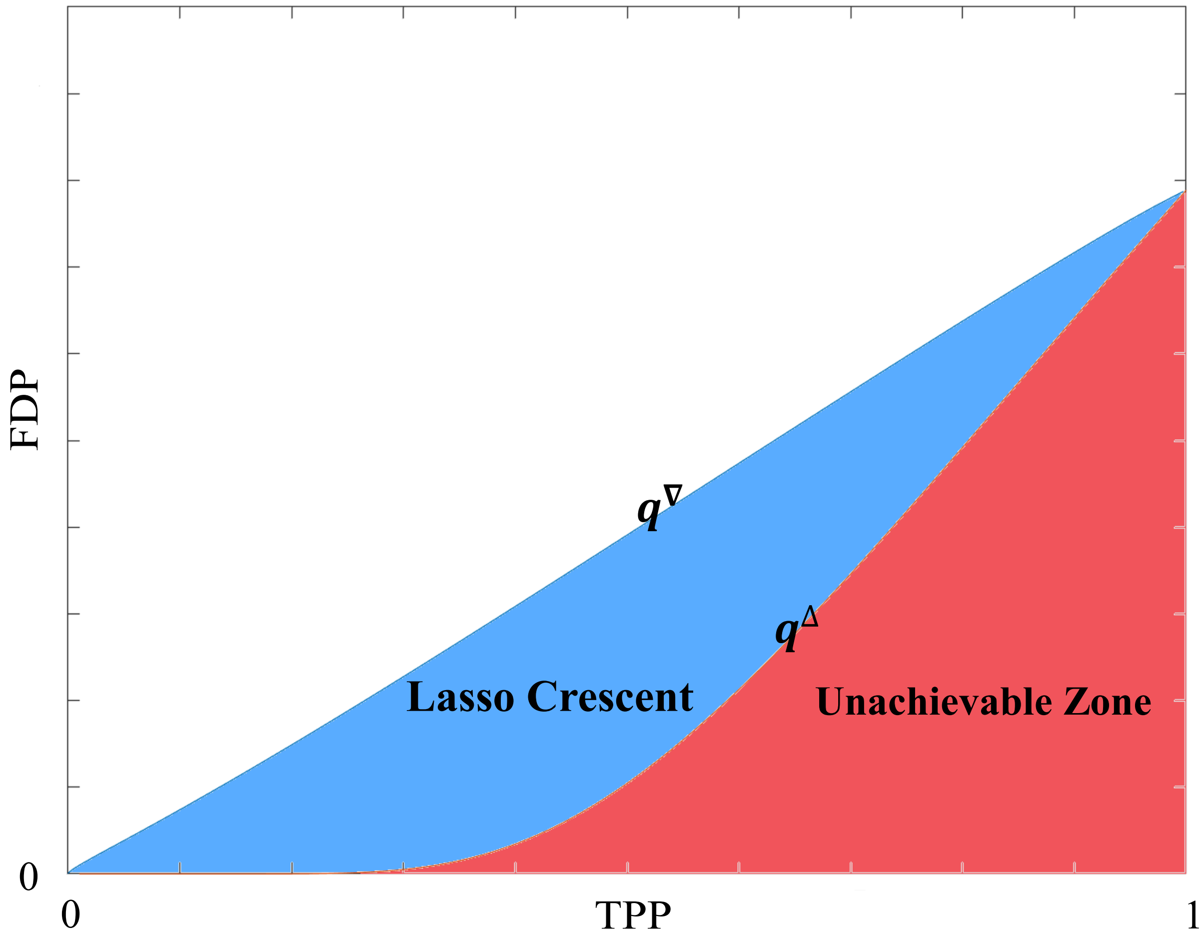
Theorem (W., Yang and Su, 2020+)

With probability tending to one,

$$q^{\triangle}(\mathrm{TPP}(\lambda)) - 0.001 \le \mathrm{FDP}(\lambda) \le q^{\triangledown}(\mathrm{TPP}(\lambda)) + 0.001$$

uniformly for all λ*, where* $q^{\triangle}(\cdot) = q^{\triangle}(\cdot; \delta, \epsilon) > 0$ *and* $q^{\triangledown}(\cdot) = q^{\triangledown}(\cdot; \delta, \epsilon) < 1$ *are two deterministic function.*

The Lasso Crescent



The sharpest of the Lasso Crescent

Definition (most favorable prior)

For $M > 0$ and an integer $m > 0$, we call the following the (ϵ, m, M)-prior:

$$\Pi^{\triangle} = \begin{cases} 0 & w.p.\ 1-\epsilon \\ M & w.p.\ \frac{\epsilon}{m} \\ M^2 & w.p.\ \frac{\epsilon}{m} \\ \cdots & \cdots \\ M^m & w.p.\ \frac{\epsilon}{m}. \end{cases}$$

Definition (least favorable prior)

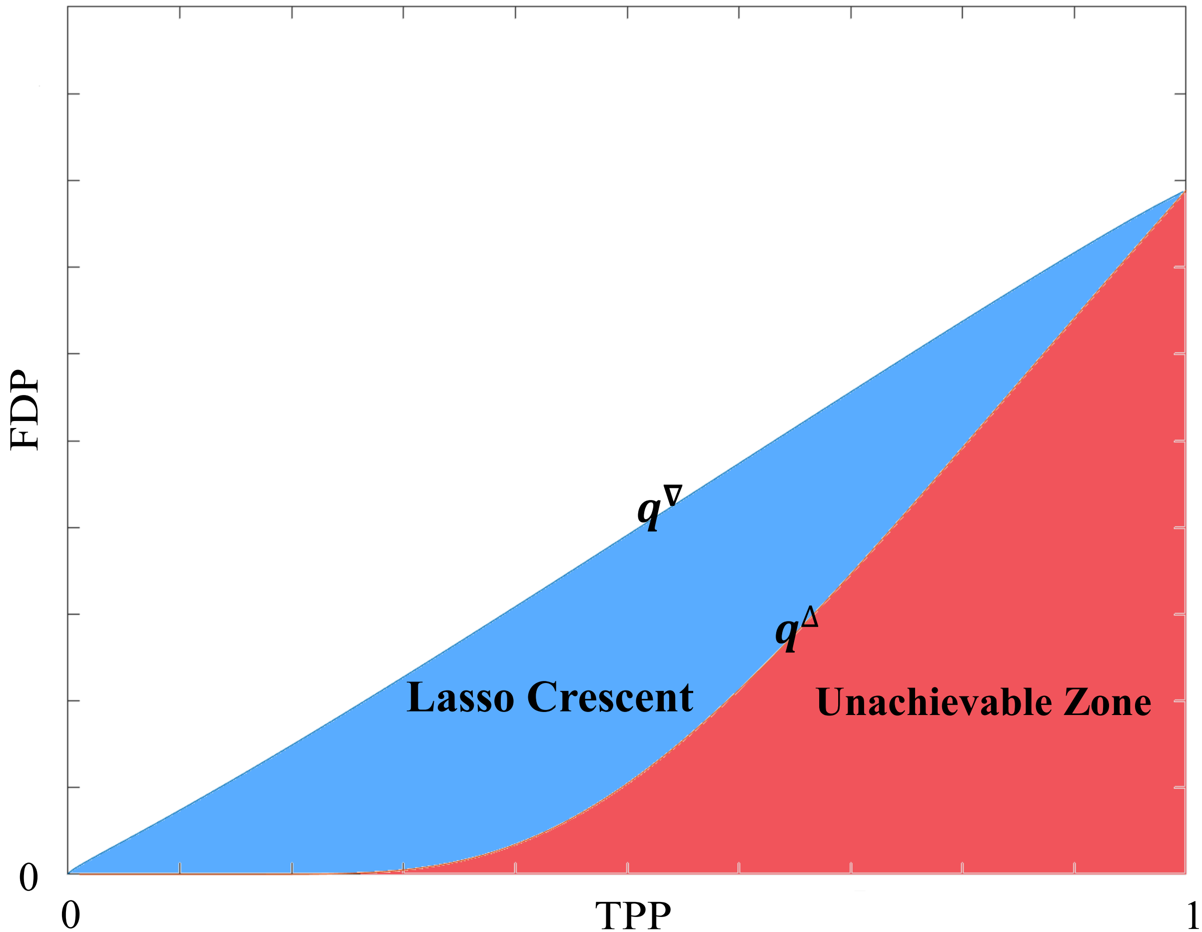
For $M > 0$, we call the following the (ϵ, M)-prior:

$$\Pi^{\nabla} = \begin{cases} 0 & w.p.\ 1-\epsilon \\ M & w.p.\ \epsilon. \end{cases}$$

Theorem (Effect Size Heterogeneity Matters!)

The Π^{∇} achieves q^{∇}, and $\Pi^{\triangle}$ achieves $q^{\triangle}$, as $M, m \to \infty$.

The Lasso Crescent (Re-visit)



Remarks on the results

Theorem (W., Yang and Su, 2020+)

With probability tending to one,

$$q^{\triangle}(\mathrm{TPP}(\lambda)) - 0.001 \leq \mathrm{FDP}(\lambda) \leq q^{\triangledown}(\mathrm{TPP}(\lambda)) + 0.001$$

for all $\lambda > 0.01$, where $q^{\triangle}(\cdot)$ and $q^{\triangledown}(\cdot)$ are two deterministic function. And the $\Pi^{\triangledown}$ (absolutely homogeneous) gives $q^{\triangledown}$, and $\Pi^{\triangle}$ (absolutely heterogeneous) gives $q^{\triangle}$.

Remarks on the results

Theorem (W., Yang and Su, 2020+)

With probability tending to one,

$$q^{\triangle}(\mathrm{TPP}(\lambda)) - 0.001 \le \mathrm{FDP}(\lambda) \le q^{\triangledown}(\mathrm{TPP}(\lambda)) + 0.001$$

for all $\lambda > 0.01$, *where* $q^{\triangle}(\cdot)$ *and* $q^{\triangledown}(\cdot)$ *are two deterministic function. And the* $\Pi^{\triangledown}$ *(absolutely homogeneous) gives* $q^{\triangledown}$, *and* $\Pi^{\triangle}$ *(absolutely heterogeneous) gives* $q^{\triangle}$.

- 0.001 can be any small number.

Remarks on the results

THEOREM (W., YANG AND SU, 2020+)

With probability tending to one,

$$q^{\triangle}(\mathrm{TPP}(\lambda)) - 0.001 \le \mathrm{FDP}(\lambda) \le q^{\triangledown}(\mathrm{TPP}(\lambda)) + 0.001$$

for all $\lambda > 0.01$, where $q^{\triangle}(\cdot)$ and $q^{\triangledown}(\cdot)$ are two deterministic function. And the $\Pi^{\triangledown}$ (absolutely homogeneous) gives $q^{\triangledown}$, and $\Pi^{\triangle}$ (absolutely heterogeneous) gives $q^{\triangle}$.

- 0.001 can be any small number.
- Lower curve $q^{\triangle}(\cdot)$ is first discovered in (W. Su, M. Bogdan, E. Candès 2017), yet here we first prove it is tight and uniformly achieveable by some prior.

Remarks on the results

Theorem (W., Yang and Su, 2020+)

With probability tending to one,

$$q^{\triangle}(\mathrm{TPP}(\lambda)) - 0.001 \leq \mathrm{FDP}(\lambda) \leq q^{\triangledown}(\mathrm{TPP}(\lambda)) + 0.001$$

for all $\lambda > 0.01$, where $q^{\triangle}(\cdot)$ and $q^{\triangledown}(\cdot)$ are two deterministic function. And the $\Pi^{\triangledown}$ (absolutely homogeneous) gives $q^{\triangledown}$, and $\Pi^{\triangle}$ (absolutely heterogeneous) gives $q^{\triangle}$.

- 0.001 can be any small number.
- Lower curve $q^{\triangle}(\cdot)$ is first discovered in (W. Su, M. Bogdan, E. Candès 2017), yet here we first prove it is tight and uniformly achieveable by some prior.
- Approximate message passing (Donoho et al, 2009).

The first false variable

Let T denotes the number of true variables before the first false variable (including itself). i.e.

$$T := \left\|\widehat{\beta}(\lambda^* - 0)\right\|_0 = \left\|\widehat{\beta}(\lambda^*)\right\|_0 + 1,$$

where λ^* is the first time along the Lasso path when a false variable is about to be selected:

$$\lambda^* = \sup\{\lambda : \text{there exists } 1 \le i \le p \text{ such that } \widehat{\beta}_i(\lambda) \neq 0, \beta_i = 0\}.$$

- Intuitively, the larger T is, the better the performance as a model selector.

The most favorable and least favorable prior (Re-visit)

Recall the most favorable and least favorable prior we defined.

Definition (most favorable prior)

For $M > 0$ and an integer $m > 0$, we call the following the (ϵ, m, M)-prior:

$$\Pi^{\triangle} = \begin{cases} 0 & w.p.\ 1-\epsilon \\ M & w.p.\ \frac{\epsilon}{m} \\ M^2 & w.p.\ \frac{\epsilon}{m} \\ \cdots & \cdots \\ M^m & w.p.\ \frac{\epsilon}{m}. \end{cases}$$

Definition (least favorable prior)

For $M > 0$, we call the following the (ϵ, M)-prior:

$$\Pi^{\nabla} = \begin{cases} 0 & w.p.\ 1-\epsilon \\ M & w.p.\ \epsilon. \end{cases}$$

The most favorable and least favorable prior (Re-visit)

Recall the most favorable and least favorable prior we defined. **Any critique?**

Definition (most favorable prior)

For $M > 0$ and an integer $m > 0$, we call the following the (ϵ, m, M)-prior:

$$\Pi^{\triangle} = \begin{cases} 0 & w.p.\ 1-\epsilon \\ M & w.p.\ \frac{\epsilon}{m} \\ M^2 & w.p.\ \frac{\epsilon}{m} \\ \cdots & \cdots \\ M^m & w.p.\ \frac{\epsilon}{m}. \end{cases}$$

Definition (least favorable prior)

For $M > 0$, we call the following the (ϵ, M)-prior:

$$\Pi^{\nabla} = \begin{cases} 0 & w.p.\ 1-\epsilon \\ M & w.p.\ \epsilon. \end{cases}$$

The most favorable and least favorable prior (Re-visit)

Recall the most favorable and least favorable prior we defined. **Any critique?** One might argue that the actually signal should be *fixed*!

Definition (most favorable prior)

For $M > 0$ and an integer $m > 0$, we call the following the (ϵ, m, M)-prior:

$$\Pi^{\triangle} = \begin{cases} 0 & w.p. \ 1-\epsilon \\ M & w.p. \ \frac{\epsilon}{m} \\ M^2 & w.p. \ \frac{\epsilon}{m} \\ \cdots & \cdots \\ M^m & w.p. \ \frac{\epsilon}{m}. \end{cases}$$

Definition (least favorable prior)

For $M > 0$, we call the following the (ϵ, M)-prior:

$$\Pi^{\nabla} = \begin{cases} 0 & w.p. \ 1-\epsilon \\ M & w.p. \ \epsilon. \end{cases}$$

The best T via heterogeneous signal

The following considered a *typical realization* of most favorable prior.

Proposition (The (fixed) most heterogeneous signal)

Consider fixed signal structure $\beta_j = M^{k+1-j}$ *for* $1 \leq j \leq k$*, and* $\beta_j = 0$ *for* $j > k$*. When* M *is sufficiently large, the rank* T *satisfies:*

$$T \geq (1 + o_p(1)) \frac{n}{2 \log p} \quad a.s.$$

The best T via heterogeneous signal

The following considered a *typical realization* of most favorable prior.

Proposition (The (fixed) most heterogeneous signal)

Consider fixed signal structure $\beta_j = M^{k+1-j}$ *for* $1 \le j \le k$*, and* $\beta_j = 0$ *for* $j > k$*. When* M *is sufficiently large, the rank* T *satisfies:*

$$T \ge (1 + o_p(1)) \frac{n}{2 \log p} \quad a.s.$$

Theorem (The most favorable is the most favorable)

For arbitrary regression coefficients β *with sparsity satisfying* $k \le \epsilon p$*, the rank* T *of the first false variable selected by the Lasso satisfies*

$$T \le (1 + o_{\mathbb{P}}(1)) \frac{n}{2 \log p} \quad a.s.$$

The homogeneous signal gives *early* false discovery

Proposition (W. Su 2018)

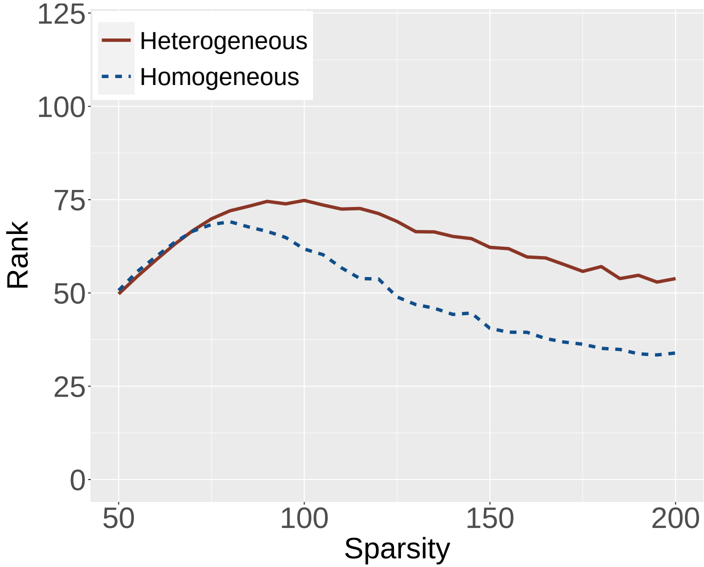
Consider the fixed signal structure as $\beta_j = M$ for $1 \leq j \leq k$, and $\beta_j = 0$ for $j > k$. The rank T satisfies

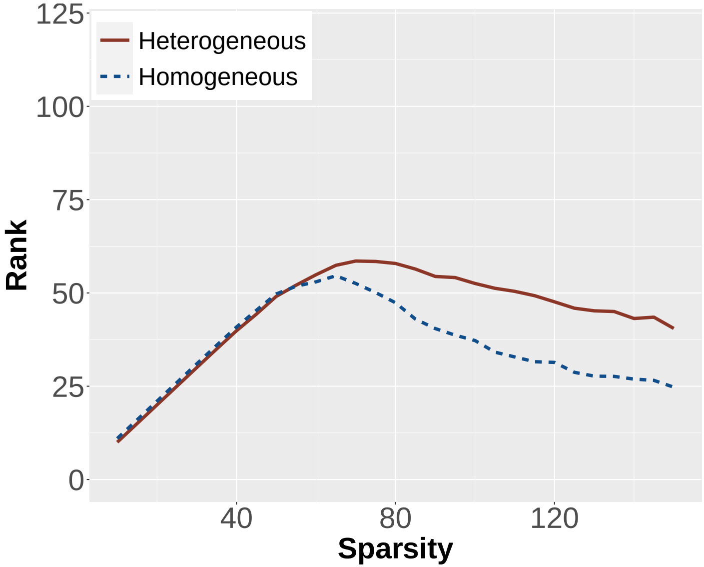
$$\log T = (1 + o_{\mathbb{P}}(1))\sqrt{\frac{2\delta \log p}{\epsilon}}.$$

- It is much earlier than that of heterogeneous signal:

$$e^{(1+o_{\mathbb{P}}(1))\sqrt{\frac{2\delta \log p}{\epsilon}}} \ll (1 + o_{\mathbb{P}}(1))\frac{n}{2 \log p},$$

Simulation: Rank of the first false discovery by Lasso





Left: $n = 1000, p = 1000, \sigma = 1$, Right: $n = 800, p = 1200, \sigma = 1$.
All averaged over 500 replicates.

Explanation: the price of competition

Well-matched (unselected) signals are the cause for bad performance.

Explanation: the price of competition

Well-matched (unselected) signals are the cause for bad performance.

- Consider $\hat{\beta}(\lambda)$, with support $\hat{S} \subset S$. The next variable $j \notin S$ will be falsely selected if it has the largest $X_j^T(y - X\hat{\beta})$ in absolute value.

Explanation: the price of competition

Well-matched (unselected) signals are the cause for bad performance.

- Consider $\hat{\beta}(\lambda)$, with support $\hat{S} \subset S$. The next variable $j \notin S$ will be falsely selected if it has the largest $X_j^T(y - X\hat{\beta})$ in absolute value.

$$X_j^T(y - X\hat{\beta})$$

Explanation: the price of competition

Well-matched (unselected) signals are the cause for bad performance.

- Consider $\hat{\beta}(\lambda)$, with support $\hat{S} \subset S$. The next variable $j \notin S$ will be falsely selected if it has the largest $X_j^T(y - X\hat{\beta})$ in absolute value.

$$X_j^T(y - X\hat{\beta}) \approx X_j^T(y - X\beta_{\hat{S}})$$

Explanation: the price of competition

Well-matched (unselected) signals are the cause for bad performance.

- Consider $\hat{\beta}(\lambda)$, with support $\hat{S} \subset S$. The next variable $j \notin S$ will be falsely selected if it has the largest $X_j^T(y - X\hat{\beta})$ in absolute value.

$$X_j^T(y - X\hat{\beta}) \approx X_j^T(y - X\beta_{\hat{S}}) = X_j^T X_{S\setminus\hat{S}}\beta_{S\setminus\hat{S}}$$

Explanation: the price of competition

Well-matched (unselected) signals are the cause for bad performance.

- Consider $\hat{\beta}(\lambda)$, with support $\hat{S} \subset S$. The next variable $j \notin S$ will be falsely selected if it has the largest $X_j^T(y - X\hat{\beta})$ in absolute value.

$$X_j^T(y - X\hat{\beta}) \approx X_j^T(y - X\beta_{\hat{S}}) = X_j^T X_{S\setminus\hat{S}}\beta_{S\setminus\hat{S}} \sim \mathcal{N}(0, \|\beta_{S\setminus\hat{S}}\|^2/n).$$

Explanation: the price of competition

Well-matched (unselected) signals are the cause for bad performance.

- Consider $\hat{\beta}(\lambda)$, with support $\hat{S} \subset S$. The next variable $j \notin S$ will be falsely selected if it has the largest $X_j^T(y - X\hat{\beta})$ in absolute value.

$$X_j^T(y - X\hat{\beta}) \approx X_j^T(y - X\beta_{\hat{S}}) = X_j^T X_{S\setminus\hat{S}}\beta_{S\setminus\hat{S}} \sim \mathcal{N}(0, \|\beta_{S\setminus\hat{S}}\|^2/n).$$

- When homogeneous, standard deviation is

$$1/\sqrt{n}\|\beta_{S\setminus\hat{S}}\| \approx \sqrt{k/n} \sup_{j\in S\setminus\hat{S}} \beta_j.$$

Explanation: the price of competition

Well-matched (unselected) signals are the cause for bad performance.

- Consider $\hat{\beta}(\lambda)$, with support $\hat{S} \subset S$. The next variable $j \notin S$ will be falsely selected if it has the largest $X_j^T(y - X\hat{\beta})$ in absolute value.

$$X_j^T(y - X\hat{\beta}) \approx X_j^T(y - X\beta_{\hat{S}}) = X_j^T X_{S\setminus\hat{S}}\beta_{S\setminus\hat{S}} \sim \mathcal{N}(0, \|\beta_{S\setminus\hat{S}}\|^2/n).$$

- When homogeneous, standard deviation is

$$1/\sqrt{n}\|\beta_{S\setminus\hat{S}}\| \approx \sqrt{k/n} \sup_{j\in S\setminus\hat{S}} \beta_j.$$

- When heterogeneous, standard deviation is

$$1/\sqrt{n}\|\beta_{S\setminus\hat{S}}\| \approx 1/\sqrt{n} \sup_{j\in S\setminus\hat{S}} \beta_j$$

Reflections on the assumptions

Assume

- X has iid $\mathcal{N}(0, 1/n)$ entries,
- $\sigma = 0$, i.e. noise $z_i = 0$,
- regression coefficients β_i are iid from prior Π with $\mathbb{E}\Pi^2 < \infty$ and $\mathbb{P}(\Pi \neq 0) = \epsilon \in (0, 1)$,
- $n/p \to \delta \in (0, \infty)$.

Then

Theorem (W., Yang and Su, 2020+)

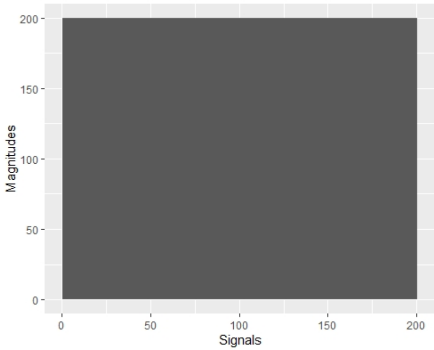
With probability tending to one,

$$q^{\triangle}(\mathrm{TPP}(\lambda)) - 0.001 \leq \mathrm{FDP}(\lambda) \leq q^{\triangledown}(\mathrm{TPP}(\lambda)) + 0.001$$

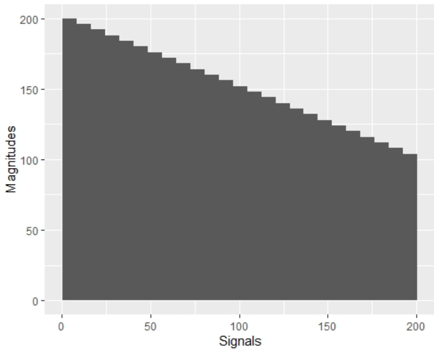
for all $\lambda > 0.01$, where $q^{\triangle}(\cdot) = q^{\triangle}(\cdot; \delta, \epsilon) > 0$ and $q^{\triangledown}(\cdot) = q^{\triangledown}(\cdot; \delta, \epsilon) < 1$ are two deterministic function.

The 4 settings (Re-visit)

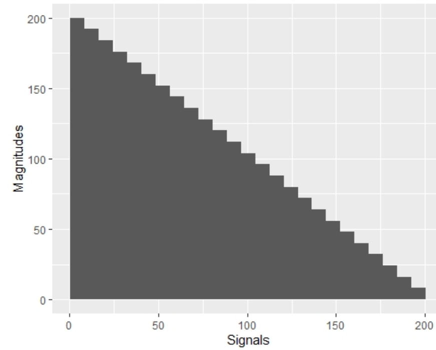
Setting 1: Most Homogeneous



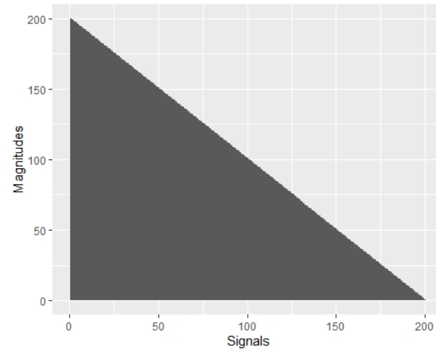
Setting 2: Homogeneous.



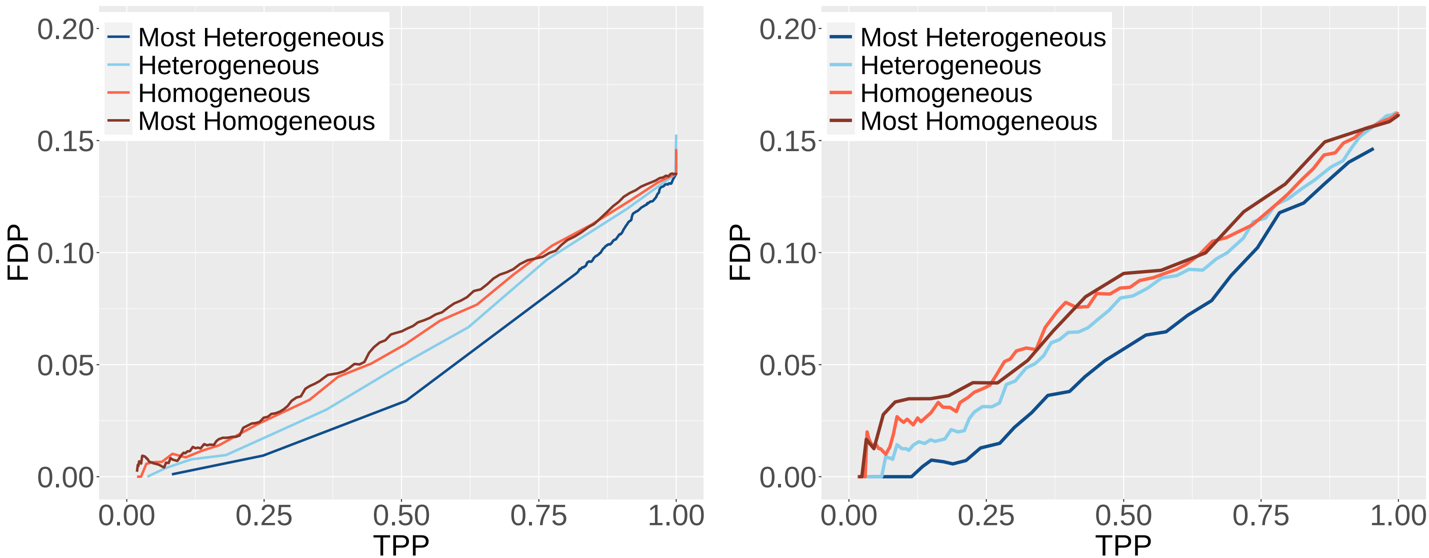
Setting 3: Heterogeneous.



Setting 4: Most Heterogeneous.



Non-Gaussian design matrix: The same phenomenon



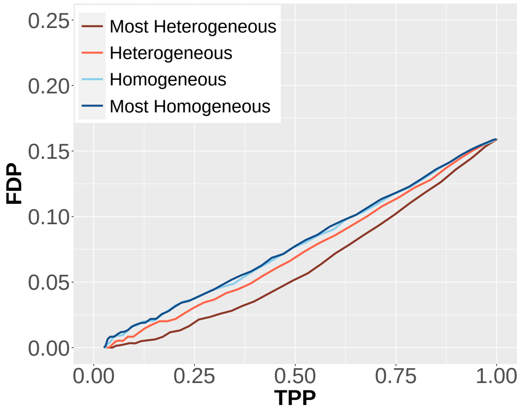
$n = 1000, p = 1000, k = 200, \sigma = 0$. Consider 4 different structure of signals.
Left: Autoregressive design matrix with $\rho = 0.5$;
Right: Bernoulli design matrix with success prob $= 0.5$

Real data as design matrix: Still the same phenomenon

Use the HIV real data. $n = 634$, $p = 463$, $\sigma = 0$, with 4 different signal structures.
Left: Original HIV data as X design matrix;
Right: perturbed X design matrix with unit Gaussian noise then re-normalize.

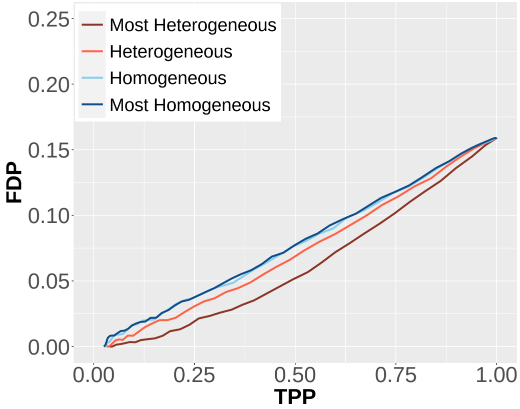
From noiseless to noisy: Still similar phenomenon!

Setting 1: No noise.

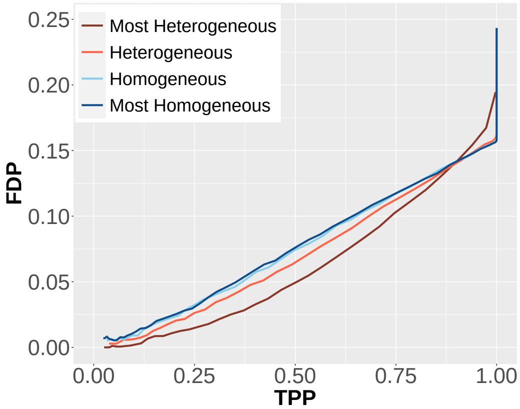


From noiseless to noisy: Still similar phenomenon!

Setting 1: No noise.

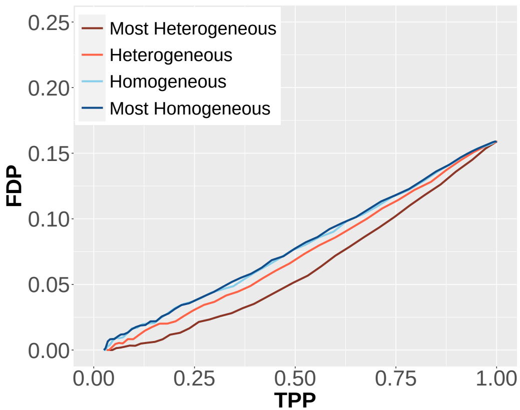


Setting 2: Small noise.

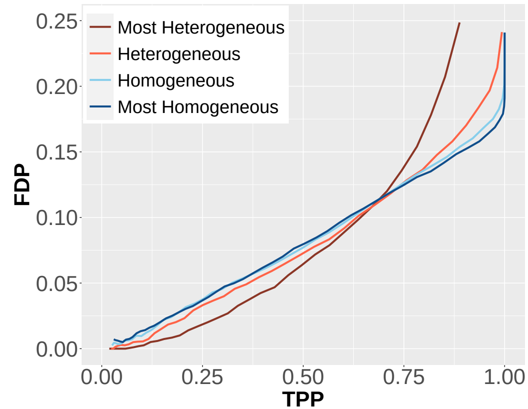


From noiseless to noisy: Still similar phenomenon!

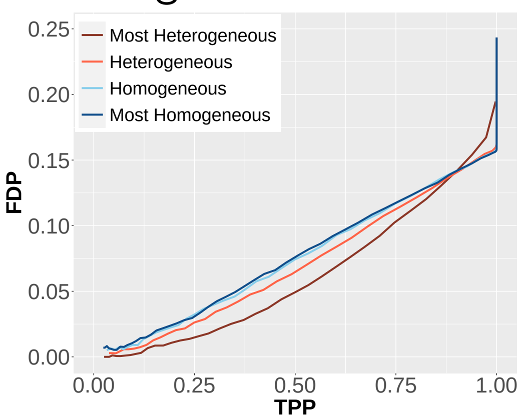
Setting 1: No noise.



Setting 3: Moderate Noise.

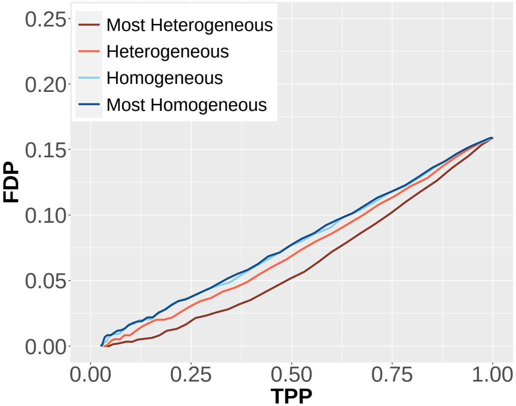


Setting 2: Small noise.

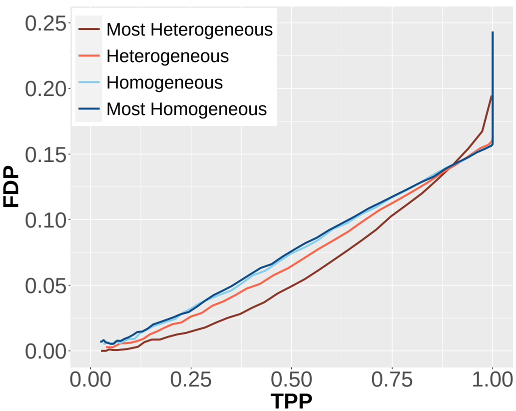


From noiseless to noisy: Still similar phenomenon!

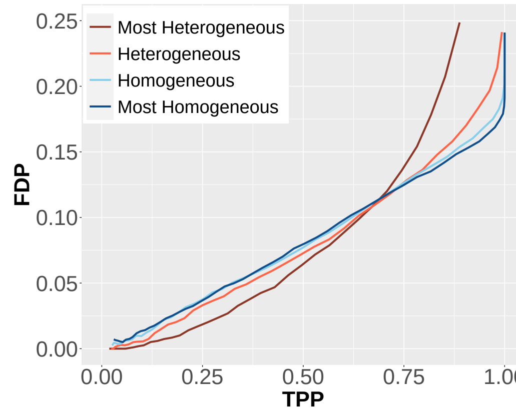
Setting 1: No noise.



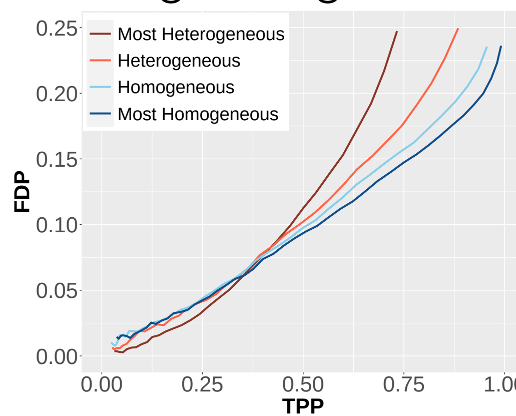
Setting 2: Small noise.



Setting 3: Moderate Noise.



Setting 4: Large Noise.



Take-home messages

Take-home messages

Our results: Always

- **Effect Size Heterogeneity** Matters

Take-home messages

Our results: Always

- **Effect Size Heterogeneity** Matters

Perspectives

- The TPP-FDP tradeoff curve
- The rank of the first false selection

Take-home messages

Our results: Always

- **Effect Size Heterogeneity** Matters

Perspectives

- The TPP-FDP tradeoff curve
- The rank of the first false selection

Occurs when

- Non-vanishing sparsity ratio
- Local orthogonal Gaussian design

Take-home messages

Our results: Always

- **Effect Size Heterogeneity** Matters

Perspectives

- The TPP-FDP tradeoff curve
- The rank of the first false selection

Occurs when

- Non-vanishing sparsity ratio
- Local orthogonal Gaussian design

Cause

- The **price of competition** of Lasso
- Inherent shrinkage of ℓ_1-methods

Take-home messages

Our results: Always

- **Effect Size Heterogeneity** Matters

Perspectives

- The TPP-FDP tradeoff curve
- The rank of the first false selection

Occurs when

- Non-vanishing sparsity ratio
- Local orthogonal Gaussian design

Cause

- The **price of competition** of Lasso
- Inherent shrinkage of ℓ_1-methods

Possible future work:

- Quantify its effect.

Take-home messages

Our results: Always

- **Effect Size Heterogeneity** Matters

Perspectives

- The TPP-FDP tradeoff curve
- The rank of the first false selection

Occurs when

- Non-vanishing sparsity ratio
- Local orthogonal Gaussian design

Cause

- The **price of competition** of Lasso
- Inherent shrinkage of ℓ_1-methods

Possible future work:

- Quantify its effect.
- Some Methodologies?

Take-home messages

Our results: Always

- **Effect Size Heterogeneity** Matters

Perspectives

- The TPP-FDP tradeoff curve
- The rank of the first false selection

Occurs when

- Non-vanishing sparsity ratio
- Local orthogonal Gaussian design

Cause

- The **price of competition** of Lasso
- Inherent shrinkage of ℓ_1-methods

Possible future work:

- Quantify its effect.
- Some Methodologies?
- General design?

Take-home messages

Our results: Always

- **Effect Size Heterogeneity** Matters

Perspectives

- The TPP-FDP tradeoff curve
- The rank of the first false selection

Occurs when

- Non-vanishing sparsity ratio
- Local orthogonal Gaussian design

Cause

- The **price of competition** of Lasso
- Inherent shrinkage of ℓ_1-methods

Possible future work:

- Quantify its effect.
- Some Methodologies?
- General design?
- The same phenomenon in other methods?

Take-home messages

Our results: Always

- **Effect Size Heterogeneity** Matters

Perspectives

- The TPP-FDP tradeoff curve
- The rank of the first false selection

Occurs when

- Non-vanishing sparsity ratio
- Local orthogonal Gaussian design

Cause

- The **price of competition** of Lasso
- Inherent shrinkage of ℓ_1-methods

Thanks!

Possible future work:

- Quantify its effect.
- Some Methodologies?
- General design?
- The same phenomenon in other methods?

Take-home messages

Our results: Always

- **Effect Size Heterogeneity** Matters

Perspectives

- The TPP-FDP tradeoff curve
- The rank of the first false selection

Occurs when

- Non-vanishing sparsity ratio
- Local orthogonal Gaussian design

Cause

- The **price of competition** of Lasso
- Inherent shrinkage of ℓ_1-methods

Thanks! :)

Possible future work:

- Quantify its effect.
- Some Methodologies?
- General design?
- The same phenomenon in other methods?