Supplementary Materials: Penalized Linear Regression with Pairwise Screening

Siliang Gong, Kai Zhang and Yufeng Liu*

Additional simulation results

In this section, we show some additional simulation results for Simulated Examples 1 and 2, where we set $\sigma = 6$ and keep all the other set ups the same. The results are shown in Tables S1 and S2. The information we obtain is similar to that from the scenarios with $\sigma = 2$.

Additional sensitivity study

In this section, we investigate how the performance of our method depends on the sample size, dimensionality, and noise level for Simulated Example 2, as a supplement to Section 5.2. In particular, we consider $n = 100$ or 500, $p = 500, 1000, 2000$ or 5000 and $\sigma = 2$ or 6 in the Simulated Example 2. We illustrate the MSE, $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2$, FN and FP against different values of p for each configuration of sample size and noise level in Figure S1.

One can see from the plots that the performance of PCS does not change much as the dimensionality p increases from 500 to 5000. In general, it is robust as sample size,

*Siliang Gong is Ph.D. candidate (E-mail: siliang@live.unc.edu), Department of Statistics and Operations Research; Kai Zhang is Assistant Professor (E-mail: zhangk@email.unc.edu), Department of Statistics and Operations Research; and Yufeng Liu is Professor (E-mail: yfliu@email.unc.edu), Department of Statistics and Operations Research, Department of Genetics, Department of Biostatistics, Carolina Center for Genome Science, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599.

Table S1: Results for simulated example 1. For each method, we report the average MSE, l2 distance, FN and FP over 100 replications (with standard errors given in parentheses).

Method	MSE	$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2$	FN	FP
		$p = 1000, \quad \sigma = 6$		
Elnet	52.11 (0.59)	3.31 (0.07)	0.81 (0.09)	1.85 (0.26)
SIS-Elnet	50.68 (0.53)	3.15 (0.07)	0.63 (0.08)	1.81 (0.20)
LASSO	52.52 (0.57)	3.96 (0.06)	1.50 (0.10)	1.13 (0.17)
SIS-LASSO	50.88 (0.54)	3.91 (0.06)	1.44 (0.10)	1.03 (0.13)
SIS-Ridge	119.9 (1.01)	4.59 (0.01)	0.00 (0.00)	12.00 (0.00)
SIS-PACS	52.50 (0.67)	3.40 (0.06)	0.00 (0.00)	4.86 (0.04)
PCS	41.68 (0.38)	1.67 (0.07)	0.06 (0.04)	0.00 (0.00)
PRCS	43.12 (0.37)	2.04 (0.08)	0.06 (0.04)	2.05 (0.14)
		$p = 5000, \quad \sigma = 6$		
Enet	55.57 (0.64)	3.55 (0.06)	0.99 (0.11)	2.47 (0.29)
SIS-Enet	53.86 (0.60)	3.45 (0.07)	0.99 (0.10)	1.83 (0.19)
LASSO	55.95 (0.64)	4.16 (0.06)	1.77 (0.12)	1.55 (0.17)
SIS-LASSO	53.78 (0.61)	4.02 (0.06)	1.68 (0.10)	1.22 (0.13)
SIS-Ridge	123.29 (1.03)	4.68 (0.01)	0.00 (0.00)	12.00 (0.00)
SIS-PACS	56.45 (0.74)	3.80 (0.04)	0.00 (0.00)	4.94 (0.03)
PCS	42.76 (0.42)	1.96 (0.11)	0.25 (0.07)	0.04 (0.02)
PRCS	43.16 (0.47)	2.11 (0.11)	0.25 (0.07)	0.80 (0.09)

dimensionality or signal to noise ratio (SNR) vary.

Additional technical proofs

Proof of Corollary ??. First note that $\frac{W_{pn}^2 - a_{p,n}}{b_{p,n}} \geq x$ is equivalent to

$$\log(1 - W_{pn}^2) \leq \log(1 - a_{p,n} - b_{p,n}x), \tag{1}$$

where $\log(1 - W_{pn}^2) = T_{pn}$. The RHS of (1) can be further expressed as

$$\begin{aligned} \log(1 - a_{p,n} - b_{p,n}x) &= \log\big(1 - \frac{2}{n-2}p^{-4/(n-2)}c_{p,n}x - (1 - p^{-4/(n-2)}c_{p,n})\big) \\ &= \log\big(p^{-4/(n-2)}(1 - \frac{2}{n-2}x)c_{p,n}\big) \\ &= -\frac{4\log p}{n-2} + \log(1 - \frac{2}{n-2}x) + \log c_{p,n}. \end{aligned} \tag{2}$$

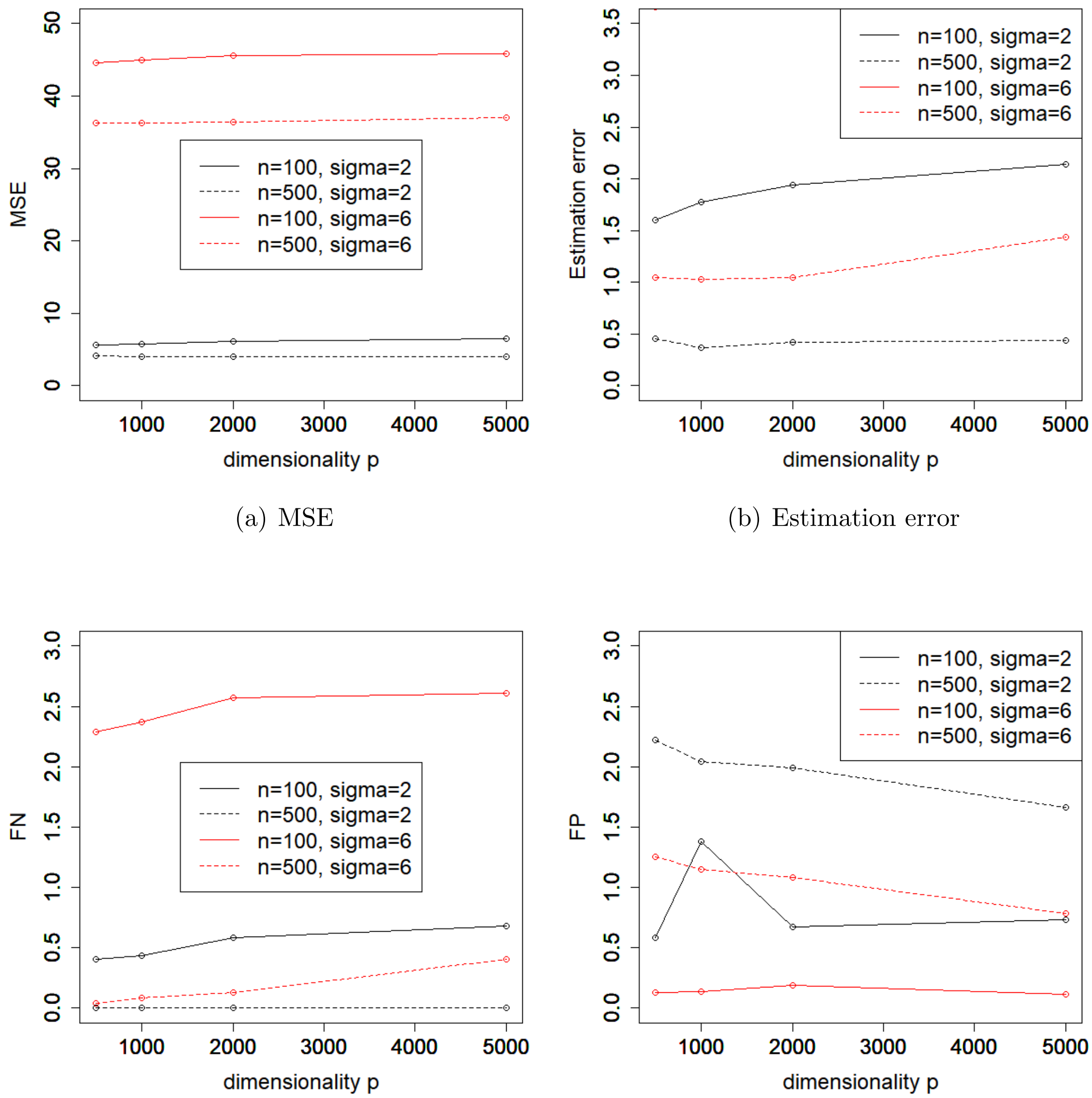
Table S2: Results for simulated example 2. The format of this table is the same as Table S1.

Method	MSE	$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2$	FN	FP
		$p = 1000, \quad \sigma = 6$		
Elnet	45.03 (0.35)	3.73 (0.03)	2.28 (0.07)	1.30 (0.59)
SIS-Elnet	45.08 (0.35)	3.75 (0.02)	2.31 (0.07)	1.53 (0.51)
LASSO	45.03 (0.36)	3.74 (0.03)	2.35 (0.06)	0.12 (0.04)
SIS-LASSO	45.09 (0.35)	3.75 (0.02)	2.43 (0.06)	0.12 (0.04)
SIS-Ridge	46.08 (0.30)	3.90 (0.00)	1.07 (0.07)	20.07 (0.07)
SIS-PACS	45.45 (0.34)	3.91 (0.02)	1.07 (0.07)	4.03 (0.06)
PCS	44.01 (0.46)	3.51 (0.06)	2.2 (0.08)	0.24 (0.05)
PRCS	44.98 (0.35)	3.73 (0.03)	2.37 (0.07)	0.14 (0.04)
		$p = 5000, \quad \sigma = 6$		
Elnet	45.78 (0.35)	3.84 (0.01)	2.48 (0.07)	1.09 (0.67)
SIS-Elnet	45.77 (0.35)	3.84 (0.02)	2.47 (0.05)	0.77 (0.36)
LASSO	45.78 (0.35)	3.84 (0.01)	2.57 (0.05)	0.20 (0.04)
SIS-LASSO	45.75 (0.35)	3.83 (0.02)	2.50 (0.05)	0.15 (0.04)
SIS-Ridge	46.14 (0.35)	3.90 (0.00)	1.42 (0.06)	20.42 (0.06)
SIS-PACS	45.76 (0.38)	3.85 (0.02)	2.46 (0.06)	0.76 (0.06)
PCS	45.80 (0.36)	3.85 (0.01)	2.61 (0.05)	0.12 (0.04)
PRCS	45.79 (0.36)	3.84 (0.02)	2.62 (0.05)	0.13 (0.05)

(i) **Sub-Exponential Case**

If $\log(p)/n \to 0$ as $n \to \infty$, then we have

$$
\begin{aligned}
c_{p,n} &= \big(\frac{2}{n-2} B(\frac{1}{2}, \frac{n-2}{2}) \sqrt{1 - p^{-4/(n-2)}}\big)^{\frac{2}{n-2}} \\
&= \left(\sqrt{\Big(\frac{(n-2)\pi}{2} + o(1) \Big) (1 - e^{-\frac{4 \log p}{n-2}})} \right)^{\frac{2}{n-2}} \\
&= \left(\frac{(n-2)\pi}{2} \cdot \frac{4 \log p}{n-2} (1 + o(1)) \right)^{\frac{2}{n-2}} \\
&= \exp\left\{ \frac{1}{n-2} \Big(\log(2\pi \log p) + o(1) \Big) \right\} \text{ for large enough } n.
\end{aligned}
$$



(a) MSE (b) Estimation error

(c) FN (d) FP

Figure S1: Performance of PCS against different dimensionality p.

Hence for large enough n,

$$
\begin{aligned}
n\log(1-a_{p,n}-b_{p,n}x) &= -\frac{4n\log p}{n-2} + n\log(1-\frac{2}{n-2}x) + \log 2\pi + \log\log p + o(1) \\
&= \log\log p - 4\log p + n\log(1-\frac{2}{n-2}x) + \log 2\pi + o(1)
\end{aligned} \tag{3}
$$

Let $y = n\log(1-\frac{2}{n-2}x)+\log 2\pi$, then the RHS of (2) becomes $\log\log p-4\log p+y+o(1)$. Combing with (1) we get

$$\begin{aligned}\lim_{n\to\infty} \mathrm{P}\left(\frac{W_{pn}^2-a_{p,n}}{b_{p,n}} \geq x\right) &= \lim_{n\to\infty} \mathrm{P}\left(nT_{pn} \leq n\log(1-a_{p,n}-b_{p,n}x)\right) \\ &= \lim_{n\to\infty} \mathrm{P}\left(nT_{pn} \leq \log\log p - 4\log p + y\right)\end{aligned} \tag{4}$$

As $p = p_n \to \infty$ as $n \to \infty$, we have

$$\begin{aligned}&\lim_{n\to\infty} \mathrm{P}\left(\frac{W_{pn}^2-a_{p,n}}{b_{p,n}} \geq x\right) = \lim_{n\to\infty, p\to\infty} \mathrm{P}\left(\frac{W_{pn}^2-a_{p,n}}{b_{p,n}} \geq x\right)\\ &= \lim_{n\to\infty}\lim_{p\to\infty} \mathrm{P}\left(\frac{W_{pn}^2-a_{p,n}}{b_{p,n}} \geq x\right) (\text{ as the convergence is uniform in } n)\\ &=1-\lim_{n\to\infty} G_n(x),\end{aligned}$$

where $G_n(x) = I(x \leq \frac{n-2}{2})\exp\big\{-\frac{1}{2}\big(1-\frac{2}{n-2}x\big)^{\frac{n-2}{2}}\big\} + I(x > \frac{n-2}{2})$.

Note that $1-\frac{2}{n-2}x = \exp\{\frac{1}{n}(y-\log 2\pi)\}$, plugging it into $G_n(x)$ yields

$$\begin{aligned}&\lim_{n\to\infty} G_n(x) = \lim_{n\to\infty} \exp\Big\{-\frac{1}{2}\exp\big\{\frac{n-2}{2n}(y-\log 2\pi)\big\}\Big\}\\ &=\exp\big\{-\frac{1}{\sqrt{8\pi}}\exp(\frac{1}{2}y)\big\}.\end{aligned}$$

Hence part (i) of Corollary **??** follows.

- **Exponential Case**

  When $(\log p)/n \to \beta \in (0,\beta)$ as $n \to \infty$, we have

$$\begin{aligned}c_{p,n} &= \big(\frac{2}{n-2}B(\frac{1}{2},\frac{n-2}{2})\sqrt{1-p^{-4/(n-2)}}\big)^{\frac{2}{n-2}}\\ &= \left(\frac{(n-2)\pi}{2}(1-e^{-4\beta})+o(1)\right)^{\frac{2}{n-2}}\\ &= \exp\left\{\frac{1}{n-2}\log\big(\frac{(n-2)\pi(1-e^{-4\beta})}{2}\big)+o(1)\big)\right\} \text{ for large enough } n.\end{aligned}$$

It follows that for large enough n,

$$\begin{aligned} n\log(c_{p,n}) &= \frac{n}{n-2}\log(n-2) + \log\big(\frac{\pi(1-e^{-4\beta})}{2}\big) + o(1) \\ &= \log\log p - \log\beta + \log\big(\frac{\pi(1-e^{-4\beta})}{2}\big) + o(1) \end{aligned}$$

Together with (2) we have

$$\begin{aligned} & n\log(1-a_{p,n}-b_{p,n}x) \\ =& \log\log p - \log\beta + \log\big(\frac{\pi(1-e^{-4\beta})}{2}\big) - \frac{4\log p}{n-2} + n\log(1-\frac{2}{n-2}x) \\ =& \log\log p - 4\log p - 8\beta + n\log(1-\frac{2}{n-2}x) + \log\big(\frac{\pi(1-e^{-4\beta})}{2\beta}\big) + o(1) \end{aligned} \tag{5}$$

Let $y = -8\beta + n\log(1-\frac{2}{n-2}x) + \log\big(\frac{\pi(1-e^{-4\beta})}{2\beta}\big)$, then the RHS of (5) becomes $\log\log p - 4\log p + y + o(1)$. Again combing with (1), we can still get (4).

Moreover,

$$\begin{aligned} \lim_{n\to\infty} G_n(x) &= \lim_{n\to\infty} \exp\Big\{ -\frac{1}{2}\exp\Big\{\frac{n-2}{2n}\Big(y + 8\beta - \log\big(\frac{\pi(1-e^{-4\beta})}{2\beta}\big)\Big)\Big\}\Big\} \\ &= \exp\big\{ -\big(\frac{\beta}{\pi(1-e^{-4\beta})}\big)^{1/2} e^{(y+8\beta)/2}\big\}, \end{aligned}$$

which leads to the convergence result in part (ii).

- **Super-Exponential Case**

  If $\log p/n \to \infty$ as $n\to\infty$, then for large enough n,

$$c_{p,n} = \big(\frac{2}{n-2}B(\frac{1}{2},\frac{n-2}{2})\sqrt{1-p^{-4/(n-2)}}\big)^{\frac{2}{n-2}} = \exp\Big\{\frac{1}{n-2}\log\big(\frac{(n-2)\pi}{2}\big)\Big\}.$$

Combing with (2) we obtain

$$
\begin{aligned}
& n\log(1-a_{p,n}-b_{p,n}x) \\
= & -\frac{4n\log p}{n-2}+n\log(1-\frac{2}{n-2}x)+\frac{n}{n-2}\log 2\pi-\frac{n}{n-2}\log(n-2)+o(1) \qquad (6)\\
= & -\frac{4n\log p}{n-2}+\log n+n\log(1-\frac{2}{n-2}x)+\log\frac{\pi}{2}+o(1).
\end{aligned}
$$

Let $y=n\log(1-\frac{2}{n-2}x)+\log\frac{\pi}{2}$, then the RHS of (5) becomes $-\frac{4n\log p}{n-2}+\log n+y+o(1)$. Moreover,

$$
\lim_{n\to\infty}G_n(x)=1-\lim_{n\to\infty}\exp\Big\{-\frac{1}{2}\exp\Big\{\frac{n-2}{2n}\big(y-\log\frac{\pi}{2}\big)\Big\}\Big\}=\exp\big\{-\frac{1}{\sqrt{2\pi}}e^{y/2}\big\}.
$$

$\square$

Proof of Theorem 2. If Y is normally distributed, then conditioning on X_i and X_j, $R^2_{ij}|X_i,X_j$ is distributed as Beta$(1,\frac{n-3}{2})$ [**?**], which is independent of X_i,X_j. Therefore, the unconditional distribution of R^2_{ij} is also Beta$(1,\frac{n-3}{2})$.

$$
\begin{aligned}
P(R^2_{pn}\geq 1-p^{-(4+\delta)/(n-3)}) &= P\big(\max_{1\leq i<j\leq p}R^2_{ij}\geq 1-p^{-(4+\delta)/(n-3)}\big)\\
&= P\big(\cup_{1\leq i<j\leq p}\{R^2_{ij}\geq 1-p^{-(4+\delta)/(n-3)}\}\big)\\
&\leq \frac{p(p-1)}{2}P(\{R^2_{ij}\geq 1-p^{-(4+\delta)/(n-3)}\})\\
&= \frac{p(p-1)}{2}\big(p^{-(4+\delta)/(n-3)}\big)^{\frac{(n-3)}{2}}\\
&= O(p^{-\delta/2})\to 0,
\end{aligned}
$$

as $p\to\infty$. $\square$