OPTIMAL ESTIMATION OF SIMULTANEOUS SIGNALS USING ABSOLUTE INNER PRODUCT WITH APPLICATIONS TO INTEGRATIVE GENOMICS

Rong Ma[1], T. Tony Cai[2] and Hongzhe Li[1]

Department of Biostatistics, Epidemiology and Informatics[1]

Department of Statistics[2]

University of Pennsylvania

Supplementary Material

The Supplementary Material includes proofs of the main theorems and the technical lemmas of the paper "Optimal Estimation of Simultaneous Signals Using Absolute Inner Product with Applications to Integrative Genomics." Supplementary notes, figures and tables are also included.

S1 Proofs of the Minimax Lower Bounds

S1.1 Proof of Theorem 1

By the definition of the rate function $\psi(s,n)$, it suffices to show the following two statements.

$$\inf_{\hat{T}} \sup_{(\theta,\mu,\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in D^\infty(s,L_n)} \mathbb{E}(\hat{T}-T(\theta,\mu))^2 \gtrsim L_n^2 s^2 \cdot \min\left\{ \log\left(1+\frac{n}{s^2}\right), L_n^2 \right\}, \tag{S1.1}$$

$$\inf_{\hat{T}} \sup_{(\theta,\mu,\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in D^\infty(s,L_n)} \mathbb{E}(\hat{T}-T(\theta,\mu))^2 \gtrsim \frac{L_n^2 s^2}{\log^2 s} \cdot \min\{\log s, L_n^2\}. \tag{S1.2}$$

Proof of (S1.1). Let $l(s,n)$ be the class of all subsets of $\{1,\dots,n\}$ of s elements. For $I\in l(s,n)$, we denote $\theta_I = \{\theta\in\mathbb{R}^n : \theta_i = 0, \forall i\notin I, \text{and } \theta_i = \rho, \forall i\in I\}$. Suppose that μ is fixed with $\mu=\mu^*$ where $\mu_i^* = L_n$ for all $1\le i\le n$. Denote

$$g_I(x_1,\dots,x_n,y_1,\dots,y_n) = \Pi_{i=1}^n \psi_{\theta_{I,i}}(x_i)\Pi_{i=1}^n \psi_{\mu_i^*}(y_i),$$

where ψ_{θ_i} denotes the density function of $N(\theta_i,1)$ and $\theta_{I,i}$ is the i-th component of θ_I. In this way, we are considering the class of probability measures for $\{x_i,y_i\}_{1\le i\le n}$ where the mean vector for $\{x_i\}$ is the s-sparse vector θ_I whereas the mean vector for $\{y_i\}$ is μ^*. By averaging over all the possible $I\in l(s,n)$, we have the mixture probability measure

$$g = \frac{1}{\binom{n}{s}} \sum_{I\in l(s,n)} g_I.$$

On the other hand, we consider the probability measure

$$f = \Pi_{i=1}^n \phi(x_i) \Pi_{i=1}^n \psi_{\mu_i^*}(y_i)$$

where ϕ is the normal density of $N(0,1)$. From the above construction, we consider $D_s^\theta(\rho) = \{(\theta, \mu, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2) : \theta = \theta_I, I \in l(s,n), \mu = \mu^*, \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = I\} \cup \{(\theta, \mu, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2) : \theta = 0, \mu = \mu^*, \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = I\}$. Apparently, for $\rho \leq L_n$, $D_s^\theta(\rho) \subset D^\infty(s, L_n)$. In the following, we will consider the χ^2-divergence between g and f and obtain the minimax lower bound over $D_s(\rho)$ using the constrained risk inequality obtained by Brown and Low (1996). Note that

$$\int \frac{g^2}{f} = \frac{1}{\binom{n}{s}^2} \sum_{I \in l(s,n)} \sum_{I' \in l(s,n)} \int \frac{g_I g_{I'}}{f}$$

and for any I and I',

$$\begin{aligned}
\int \frac{g_I g_{I'}}{f} &= \frac{1}{(2\pi)^{n/2}} \int \exp\left\{ - \frac{\sum_i (x_i - \theta_{I,i})^2 + \sum_i (x_i - \theta_{I',i})^2 - \sum_i x_i^2}{2} \right\} \\
&\quad \times \frac{1}{(2\pi)^{n/2}} \int \exp\left\{ - \frac{\sum_i (y_i - \mu_i^*)^2}{2} \right\} \\
&= \frac{1}{(2\pi)^{n/2}} \int \exp\left\{ - \frac{\sum_i (x_i - \theta_{I,i} - \theta_{I',i})^2 - 2\sum_{i=1}^n \theta_{I,i}\theta_{I',i}}{2} \right\} \\
&= \exp(\rho^2 j)
\end{aligned}$$

where j is the number of points in the set $I \cap I'$. It follows that

$$\begin{aligned}
\int \frac{g^2}{f} &= \frac{1}{\binom{n}{s}^2} \sum_{j=1}^{s} \binom{n}{s} \binom{s}{j} \binom{n-s}{s-j} \exp(2\rho^2 j) \\
&= \mathbb{E} \exp(2\rho^2 J)
\end{aligned}$$

where J has a hypergeometric distribution

$$\mathbb{P}(J=j)=\frac{\binom{s}{j}\binom{n-s}{s-j}}{\binom{n}{s}}.$$

As shown in p.173 of (Aldous, 1985), J has the same distribution as the random variable $\mathbb{E}(Z|\mathcal{B}_n)$ where Z is a binomial random variable of parameters $(s, s/n)$ and $\mathcal{B}_n$ some suitable σ-algebra. Thus, by Jensen's inequality we have

$$\mathbb{E}\exp(2J\rho^2)\le\left(1-\frac{s}{n}+\frac{s}{n}e^{2\rho^2}\right)^s. \tag{S1.3}$$

The rest of the proof will be separated into two parts, corresponding to $L_n\ge\sqrt{\log\left(1+\frac{n}{s^2}\right)}$ and $L_n<\sqrt{\log\left(1+\frac{n}{s^2}\right)}$, respectively.

Case I. $L_n\ge\sqrt{\log\left(1+\frac{n}{s^2}\right)}$. By taking $\rho=\sqrt{\log\left(1+\frac{n}{s^2}\right)}\le L_n$, we have

$$\int\frac{g^2}{f}=\mathbb{E}\exp(J\rho^2)\le e. \tag{S1.4}$$

Then if some estimator δ satisfies

$$\mathbb{E}_f(\delta-0)^2\le Cs^2\|\mu^*\|_\infty^2\log\left(1+\frac{n}{s^2}\right) \tag{S1.5}$$

then by the constrained risk inequality (Theorem 1 of Brown and Low (1996)),

$$\begin{aligned}\mathbb{E}_g\left(\delta-s\rho\|\mu^*\|_\infty\right)^2&\ge s^2\rho^2\|\mu^*\|_\infty^2-2\rho s\|\mu^*\|_\infty C^{1/2}s\|\mu^*\|_\infty\log^{1/2}\left(1+\frac{n}{s^2}\right)\\&=s^2\|\mu^*\|_\infty^2\log\left(1+\frac{n}{s^2}\right)-\sqrt{2C}s^2\|\mu^*\|_\infty^2\log\left(1+\frac{n}{s^2}\right),\end{aligned}$$

for any such estimator δ. Recall that $\|\mu^*\|_\infty = L_n$. By choosing C sufficiently small, we conclude that there exists some $I \in l(s,n)$ such that

$$\mathbb{E}_{g_I}\left(\delta - s\rho\|\mu^*\|_\infty\right)^2 \geq Cs^2L_n^2\log\left(1+\frac{n}{s^2}\right) \tag{S1.6}$$

for all δ. Therefore we have

$$\inf_{\hat{T}} \sup_{(\theta,\mu,\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in D_s^\theta(\rho)} \mathbb{E}(\hat{T}-T(\theta,\mu))^2 \geq Cs^2L_n^2\log\left(1+\frac{n}{s^2}\right). \tag{S1.7}$$

The lower bound (S1.1) then follows from the fact that $L_n \gtrsim \sqrt{\log n}$.

Case II. $L_n < \sqrt{\log\left(1+\frac{n}{s^2}\right)}$. By taking $\rho = L_n < \sqrt{\log\left(1+\frac{n}{s^2}\right)}$, again we have

$$\int \frac{g^2}{f} = \mathbb{E}\exp(J\rho^2) \leq e. \tag{S1.8}$$

Then if some estimator δ satisfies

$$\mathbb{E}_f(\delta-0)^2 \leq Cs^2\|\mu^*\|_\infty^2 L_n^2 \tag{S1.9}$$

then by the constrained risk inequality (Theorem 1 of Brown and Low (1996)),

$$\begin{aligned}\mathbb{E}_g\left(\delta - s\rho\|\mu^*\|_\infty\right)^2 &\geq s^2\rho^2\|\mu^*\|_\infty^2 - 2\rho s\|\mu^*\|_\infty C^{1/2}s\|\mu^*\|_\infty L_n \\ &= s^2\|\mu^*\|_\infty^2 L_n^2 - \sqrt{2C}s^2\|\mu^*\|_\infty^2 L_n^2,\end{aligned}$$

for any such estimator δ. Recall that $\|\mu^*\|_\infty = L_n$. By choosing C sufficiently small, we conclude that there exists some $I \in l(s,n)$ such that

$$\mathbb{E}_{g_I}\left(\delta - s\rho\|\mu^*\|_\infty\right)^2 \geq Cs^2L_n^4 \tag{S1.10}$$

for all δ. Therefore we have

$$\inf_{\hat{T}} \sup_{(\theta,\mu,\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in D_s^\theta(\rho)} \mathbb{E}(\hat{T}-T(\theta,\mu))^2 \geq Cs^2L_n^4. \tag{S1.11}$$

This proves the other part of (S1.1).

Proof of (S1.2). Now we prove the second part of the theorem. It follows by Lemma 1 in Cai and Low (2011) that there exist measures ν_i on $[-M_n, M_n]$ for $i = 0, 1$, such that:

1. ν_0 and ν_1 are symmetric around 0;
2. $\int t^l \nu_1(dt) = \int t^l \nu_0(dt)$, for $l = 0, 1, \dots, k_n$;
3. $\int |t|\nu_1(dt) - \int |t|\nu_0(dt) = 2M_n\delta_{k_n}$.
4. $\int |t|\nu_0(dt) > 0$.

where δ_{k_n} is the distance in the uniform norm on $[-1, 1]$ from the absolute value function $f(x) = |x|$ to the space of polynomials of no more than degree k_n. In addition, $\delta_{k_n} = \beta_* k_n^{-1}(1 + o(1))$ as $k_n \to \infty$. Now we consider product priors on the n-vector θ, which are supported on the first $s \leq n$ components. Let

$$\nu_{i1}^n = \Pi^{\otimes s}\nu_i \cdot \Pi^{\otimes n-s} 1_{\{0\}}, \qquad \nu_{i2}^n = \Pi^{\otimes n} 1_{\{\mu^*\}}$$

for $i = 0$ and 1. In other words, we put independent priors ν_i for the first s

components of the vector θ, while keeping the other coordinates as 0, and we fix $\mu = \mu^*$.

Following the above construction, we have

$$\mathbb{E}_{\nu_{11}^n} \frac{1}{n} \sum_{i=1}^n |\theta_i| - \mathbb{E}_{\nu_{01}^n} \frac{1}{n} \sum_{i=1}^n |\theta_i| = \frac{s}{n} \Big[\mathbb{E}_{\nu_{11}} |\theta_i| - \mathbb{E}_{\nu_{01}} |\theta_i| \Big] = 2sM_n \delta_{k_n}/n,$$

and

$$\mathbb{E}_{\nu_{12}^n} \frac{1}{n} \sum_{i=1}^n |\mu_i| - \mathbb{E}_{\nu_{02}^n} \frac{1}{n} \sum_{i=1}^n |\mu_i| = 0$$

Then we have

$$\begin{aligned}
&\mathbb{E}_{\nu_{11}^n \nu_{12}^n} \frac{1}{n} T(\theta, \mu) - \mathbb{E}_{\nu_{01}^n \nu_{02}^n} \frac{1}{n} T(\theta, \mu) \qquad \text{(S1.12)} \\
&= \frac{s}{n} \big(\mathbb{E}_{\nu_{11}} |\theta_i| \mathbb{E}_{\nu_{12}} |\mu_i| - \mathbb{E}_{\nu_{01}} |\theta_i| \mathbb{E}_{\nu_{02}} |\mu_i| \big) \\
&= \frac{s}{n} (\mathbb{E}_{\nu_{11}} |\theta_i| \mathbb{E}_{\nu_{12}} |\mu_i| - \mathbb{E}_{\nu_{01}} |\theta_i| \mathbb{E}_{\nu_{12}} |\mu_i|) \\
&\quad + \frac{s}{n} (\mathbb{E}_{\nu_{01}} |\theta_i| \mathbb{E}_{\nu_{12}} |\mu_i| - \mathbb{E}_{\nu_{01}} |\theta_i| \mathbb{E}_{\nu_{02}} |\mu_i|) \\
&= \mathbb{E}_{\nu_{12}} |\mu_i| \frac{s}{n} (\mathbb{E}_{\nu_{11}} |\theta_i| - \mathbb{E}_{\nu_{01}} |\theta_i|) + \mathbb{E}_{\nu_{01}} |\theta_i| \frac{s}{n} (\mathbb{E}_{\nu_{12}} |\mu_i| - \mathbb{E}_{\nu_{02}} |\mu_i|) \\
&= \frac{2sM_n \delta_{k_n}}{n} \mathbb{E}_{\nu_{12}} |\mu_i| \\
&= \frac{2sM_n L_n \delta_{k_n}^2}{n} \qquad \text{(S1.13)}
\end{aligned}$$

We further have

$$V_0^2 \equiv \frac{1}{n^2} \mathbb{E}_{\nu_{01}^n \nu_{02}^n} (T(\theta, \mu) - \mathbb{E}_{\nu_{01}^n \nu_{02}^n} T(\theta, \mu))^2 \leq \frac{sM_n^2 L_n^2}{n^2} \qquad \text{(S1.14)}$$

Set $f_{0,M_n}(y) = \int \phi(y-t) \nu_0(dt)$ and $f_{1,M_n}(y) = \int \phi(y-t) \nu_1(dt)$. Note that since $g(x) = \exp(-x)$ is a convex function of x, and ν_0 is symmetric,

$$\begin{aligned} f_{0,M_n}(y) &\geq \frac{1}{\sqrt{2\pi}} \exp\bigg(-\int \frac{(y-t)^2}{2} \nu_0(dt) \bigg) \\ &= \phi(y) \exp\bigg(-\frac{1}{2} M_n^2 \int t^2 v_0'(dt) \bigg) \\ &\geq \phi(y) \exp\bigg(-\frac{1}{2} M_n^2 \bigg). \end{aligned}$$

Let H_r be the Hermite polynomial defined by

$$\frac{d^r}{dy^r} \phi(y) = (-1)^r H_r(y) \phi(y) \tag{S1.15}$$

which satisfiy

$$\int H_r^2(y)\phi(y)dy = r! \qquad \text{and} \qquad \int H_r(y)H_l(y)\phi(y)dy = 0 \tag{S1.16}$$

when $r \neq l$. Then

$$\phi(y-t) = \sum_{k=0}^{\infty} H_k(y)\phi(y) \frac{t^k}{k!}$$

and it follows that

$$\begin{aligned}
&\int \frac{(f_{1,M_n}(y) - f_{0,M_n}(y))^2}{f_{0,M_n}(y)} dy \\
&\le \int (f_{1,M_n}(y) - f_{0,M_n}(y))^2 e^{M_n^2/2}/\phi(y) dy \\
&= e^{M_n^2/2} \int \left\{ \sum_{k=0}^{\infty} H_k(y) \frac{\phi(y)}{k!} \left[\int t^k \nu_1(dt) - \int t^k \nu_0(dt) \right] \right\}^2 \Big/ \phi(y) dy \\
&= e^{M_n^2/2} \int \sum_{k=k_n+1}^{\infty} H_k^2(y)\phi(y) \frac{M_n^{2k}}{(k!)^2} \left[\int t^k \nu_1(dt) - \int t^k \nu_0(dt) \right]^2 dy \\
&= e^{M_n^2/2} \sum_{k=k_n+1}^{\infty} \frac{M_n^{2k}}{k!} \left[\int t^k \nu_1(dt) - \int t^k \nu_0(dt) \right]^2 \\
&\le e^{M_n^2/2} \sum_{k=k_n+1}^{\infty} \frac{M_n^{2k}}{k!}.
\end{aligned}$$

It then follows

$$\begin{aligned}
I_n^2 &= \prod_{i=1}^{s} \int \frac{(f_{1,M_n}(x_i))^2}{f_{0,M_n}(x_i)} dx_i - 1 \\
&\le \left(1 + e^{M_n^2/2} \sum_{k=k_n+1}^{\infty} \frac{1}{k!} M_n^{2k} \right)^s - 1 \\
&\le \left(1 + e^{M_n^2/2} D \frac{1}{k_n!} M_n^{2k_n} \right)^s - 1,
\end{aligned}$$

for some $D > 0$. Since $k! > (k/e)^k$, we also have

$$I_n^2 \le \left(1 + e^{M_n^2/2} D \left(\frac{eM_n^2}{k_n} \right)^{k_n} \right)^s - 1 \tag{S1.17}$$

The rest of the proof is separated into two parts.

Case I. $L_n \ge \sqrt{\log s}$. Now let $M_n = \sqrt{\log s} \le L_n$ and $k_n \asymp \log s$. It then can be checked that $I_n < c$ for some sufficiently small constant $c > 0$.

Therefore, Corollary 1 in Cai and Low (2011) along with the fact that $L_n \geq \sqrt{\log s}$ yields

$$\inf_{\hat{T}} \sup_{(\theta,\mu,\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in D^\infty(s,L_n)} \frac{1}{n^2}\mathbb{E}(\hat{T}-T(\theta,\mu))^2 \geq \frac{(2sM_n\delta_{k_n}L_n/n-(s^{1/2}M_nL_n/n)I_n)^2}{(I_n+2)^2} \geq \frac{Cs^2L_n^2}{n^2\log s}. \tag{S1.18}$$

Case II. $L_n < \sqrt{\log s}$. In this case, we set $M_n = L_n < \sqrt{\log s}$ and $k_n \asymp \log s$. Then again $I_n < c$ for some sufficiently small constant $c > 0$. Therefore, Corollary 1 in Cai and Low (2011) along with the fact that $M_n \leq L_n$ yields

$$\inf_{\hat{T}} \sup_{(\theta,\mu,\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in D^\infty(s,L_n)} \frac{1}{n^2}\mathbb{E}(\hat{T}-T(\theta,\mu))^2 \geq \frac{(2sM_n\delta_{k_n}L_n/n-(s^{1/2}M_nL_n/n)I_n)^2}{(I_n+2)^2} \geq \frac{Cs^2L_n^4}{n^2\log^2 s}. \tag{S1.19}$$

S2 Proofs of the Risk Upper Bounds

S2.1 Proof of Theorem 2

For the hybrid estimators defined in Section 2.2 of the main paper, the key is to study the bias and variance of a single component. Let $x_1, x_2 \sim N(\theta, 1)$, $y_1, y_2 \sim N(\mu, 1)$. Denote $\delta(x) = \min\{S_K(x), n^2\}$. In the following, we analyse the two hybrid estimators separately.

Part I: Analysis of $\widehat{T}_K^S$. Let

$$\xi = \xi(x_1, x_2, y_1, y_2)$$
$$= [\delta(x_1)I(\sqrt{2\log n} < |x_2| \le 2\sqrt{2\log n}) + |x_1|I(|x_2| > 2\sqrt{2\log n})]$$
$$\times [\delta(y_1)I(\sqrt{2\log n} < |y_2| \le 2\sqrt{2\log n}) + |y_1|I(|y_2| > 2\sqrt{2\log n})] \tag{S2.1}$$

Note that

$$\mathbb{E}(\xi) = [\mathbb{E}\delta(x_1)P(\sqrt{2\log n} < |x_2| \le 2\sqrt{2\log n}) + \mathbb{E}|x_1|P(|x_2| > 2\sqrt{2\log n})]$$
$$\times [\mathbb{E}\delta(y_1)P(\sqrt{2\log n} < |y_2| \le 2\sqrt{2\log n}) + \mathbb{E}|y_1|P(|y_2| > 2\sqrt{2\log n})] \tag{S2.2}$$

We denote

$$\tilde{\sigma}_x^2 = \mathrm{Var}(\delta(x_1)I(\sqrt{2\log n} < |x_2| \le 2\sqrt{2\log n}) + |x_1|I(|x_2| > 2\sqrt{2\log n})),$$

$$\tilde{\sigma}_y^2 = \mathrm{Var}(\delta(y_1)I(\sqrt{2\log n} < |y_2| \le 2\sqrt{2\log n}) + |y_1|I(|y_2| > 2\sqrt{2\log n})),$$

and

$$\tilde{\theta}_x = \mathbb{E}\delta(x_1)P(\sqrt{2\log n} < |x_2| \le 2\sqrt{2\log n}) + \mathbb{E}|x_1|P(|x_2| > 2\sqrt{2\log n}),$$

$$\tilde{\mu}_y = \mathbb{E}\delta(y_1)P(\sqrt{2\log n} < |y_2| \le 2\sqrt{2\log n}) + \mathbb{E}|y_1|P(|y_2| > 2\sqrt{2\log n}).$$

Then we have

$$\mathrm{Var}(\xi) = \tilde{\sigma}_x^2\tilde{\sigma}_y^2 + \tilde{\sigma}_x^2\tilde{\mu}_y^2 + \tilde{\sigma}_y^2\tilde{\theta}_x^2$$

The following two propositions are key to our calculation of the estimation risk.

Proposition 1. *For all $\theta, \mu \in \mathbb{R}$, we have*

$$|B_x| \equiv |\tilde{\theta}_x - |\theta|| \lesssim \sqrt{\log n}, \qquad |B_y| \equiv |\tilde{\mu}_y - |\mu|| \lesssim \sqrt{\log n},$$

and

$$\tilde{\sigma}_x^2 \lesssim \log n, \qquad \tilde{\sigma}_y^2 \lesssim \log n.$$

In particular, when $\theta = 0$, we have $|B_x| \le n^{-2}\log n$ and $\tilde{\sigma}_x^2 \lesssim n^{-1}\log n$, whereas when $\mu = 0$, we have $|B_y| \le n^{-2}\log n$, and $\tilde{\sigma}_y^2 \lesssim n^{-1}\log n$.

Proposition 2. *For all $\theta, \mu \in \mathbb{R}$ such that $\theta, \mu \le L_n$ where $L_n \le (\sqrt{2} - 1)\sqrt{\log n}$, we have*

$$|B_x| \equiv |\tilde{\theta}_x - |\theta|| \lesssim L_n, \qquad |B_y| \equiv |\tilde{\mu}_y - |\mu|| \lesssim L_n,$$

and

$$\tilde{\sigma}_x^2 \lesssim \frac{\log n}{\sqrt{n}}, \qquad \tilde{\sigma}_y^2 \lesssim \frac{\log n}{\sqrt{n}}.$$

Now the bias of the estimator ξ satisfies

$$\begin{aligned} B(\xi) &= \mathbb{E}\xi - |\theta||\mu| = \tilde{\theta}_x\tilde{\mu}_y - |\theta||\mu| \\ &\le |\theta| \cdot |\tilde{\mu}_y - |\mu|| + |\mu| \cdot |\tilde{\mu}_x - |\theta|| + |\tilde{\mu}_y - |\mu|| \cdot |\tilde{\mu}_x - |\theta|| \\ &\le |\mu||B_y| + |\theta||B_x| + |B_x||B_y|. \end{aligned} \tag{S2.3}$$

and the variance

$$\begin{aligned} \text{Var}(\xi) &= \tilde{\sigma}_x^2\tilde{\sigma}_y^2 + \tilde{\sigma}_x^2\tilde{\mu}_y^2 + \tilde{\sigma}_y^2\tilde{\theta}_x^2 \\ &\lesssim \tilde{\sigma}_x^2\tilde{\sigma}_y^2 + |\mu|^2\tilde{\sigma}_x^2 + |\theta|^2\tilde{\sigma}_y^2 + B_x^2\tilde{\sigma}_y^2 + B_y^2\tilde{\sigma}_x^2. \end{aligned} \tag{S2.4}$$

Now let $(x_{1\ell}, x_{2\ell}, ..., x_{n\ell}) \sim N(\theta, \mathbf{I}_n)$ and $(y_{1\ell}, y_{2\ell}, ..., y_{n\ell}) \sim N(\mu, \mathbf{I}_n)$ for $\ell = 1, 2$, and let

$$\widehat{T}_K^S = \sum_{i=1}^n \xi(x_{i1}, x_{i2}, y_{i1}, y_{i2}).$$

Case I. $L \gtrsim \sqrt{\log n}$. It follows from (S2.3) and Cauchy-Schwartz inequality that the bias of $\widehat{T}_K^S$ is bounded by

$$|B(\widehat{T}_K^S)| \lesssim (\|\theta\|_\infty + \|\mu\|_\infty)s\sqrt{\log n} + s\log n$$

From (S2.4) we have the variance of $\widehat{T^S(\theta,\mu)}$ is bounded by

$$\begin{aligned}\mathrm{Var}(\widehat{T}_K^S) &\leq \sum_{i=1}^n \mathrm{Var}(\xi(x_{i1}, x_{i2}, y_{i1}, y_{i2})) \\ &\lesssim s\log^2 n + (\|\theta\|_\infty^2 + \|\mu\|_\infty^2)s\log n \qquad \text{(S2.5)}\end{aligned}$$

Therefore the mean squared error of $\widehat{T}_K^S$ satisfies

$$\begin{aligned}\mathbb{E}(\widehat{T}_K^S - T(\theta,\mu))^2 &\leq B^2(\widehat{T}_K^S) + \mathrm{Var}(\widehat{T}_K^S) \\ &\lesssim s^2 L_n^2 \log n.\end{aligned}$$

Case II. $L_n \lesssim \sqrt{\log n}$. It follows that

$$|B(\widehat{T}_K^S)| \lesssim (\|\theta\|_\infty + \|\mu\|_\infty)sL_n + sL_n^2$$

From (S2.4) we have the variance of $\widehat{T^S(\theta,\mu)}$ is bounded by

$$\begin{aligned}\mathrm{Var}(\widehat{T}_K^S) &\leq \sum_{i=1}^n \mathrm{Var}(\xi(x_{i1}, x_{i2}, y_{i1}, y_{i2})) \\ &\lesssim \log^2 n + \frac{sL_n^2\log n}{\sqrt{n}} + \frac{\log^2 n}{\sqrt{n}} + L_n^2\log n + \frac{\log^3 n}{n^3}. \qquad \text{(S2.6)}\end{aligned}$$

Therefore the mean squared error of $\widehat{T}_K^S$ satisfies

$$\begin{aligned}\mathbb{E}(\widehat{T}_K^S - T(\theta,\mu))^2 &\le B^2(\widehat{T}_K^S) + \text{Var}(\widehat{T}_K^S)\\ &\lesssim s^2 L_n^4 + \frac{\log^2 n}{\sqrt{n}} + L_n^2 \log n\end{aligned}$$

Part II: Analysis of $\widehat{T}_K$. With a slight abuse of notation, we denote

$$\begin{aligned}\xi = \xi(x_1,x_2,y_1,y_2) = &[\delta(x_1)I(|x_2| \le 2\sqrt{2\log n}) + |x_1|I(|x_2| > 2\sqrt{2\log n})]\\ &\times [\delta(y_1)I(|y_2| \le 2\sqrt{2\log n}) + |y_1|I(|y_2| > 2\sqrt{2\log n})]\end{aligned} \tag{S2.7}$$

Note that

$$\begin{aligned}\mathbb{E}(\xi) = &[\mathbb{E}\delta(x_1)P(|x_2| \le 2\sqrt{2\log n}) + \mathbb{E}|x_1|P(|x_2| > 2\sqrt{2\log n})]\\ &\times [\mathbb{E}\delta(y_1)P(|y_2| \le 2\sqrt{2\log n}) + \mathbb{E}|y_1|P(|y_2| > 2\sqrt{2\log n})]\end{aligned} \tag{S2.8}$$

We denote

$$\tilde{\sigma}_x^2 = \text{Var}(\delta(x_1)I(|x_2| \le 2\sqrt{2\log n}) + |x_1|I(|x_2| > 2\sqrt{2\log n})),$$

$$\tilde{\sigma}_y^2 = \text{Var}(\delta(y_1)I(|y_2| \le 2\sqrt{2\log n}) + |y_1|I(|y_2| > 2\sqrt{2\log n})),$$

and

$$\tilde{\theta}_x = \mathbb{E}\delta(x_1)P(|x_2| \le 2\sqrt{2\log n}) + \mathbb{E}|x_1|P(|x_2| > 2\sqrt{2\log n}),$$

$$\tilde{\mu}_y = \mathbb{E}\delta(y_1)P(|y_2| \le 2\sqrt{2\log n}) + \mathbb{E}|y_1|P(|y_2| > 2\sqrt{2\log n}).$$

Proposition 3. *Let* $B_x = \tilde{\theta}_x - |\theta|$, $B_y = \tilde{\mu}_y - |\mu|$. *For all* $\theta, \mu \in \mathbb{R}$ *and* $K = r \log n$ *for some* $0 < r < 1/3$, *we have*

$$|B_x| \lesssim \frac{1}{\sqrt{\log n}}, \qquad |B_y| \lesssim \frac{1}{\sqrt{\log n}},$$

and

$$\tilde{\sigma}_x^2 = O\big(n^{6r} \log^3 n\big), \qquad \tilde{\sigma}_y^2 = O\big(n^{6r} \log^3 n\big).$$

In particular, when $\theta = 0$, *we have* $|B_x| \leq n^{6r-2} \log^3 n$ *and* $\tilde{\sigma}_x^2 \lesssim n^{6r} \log^3 n$, *whereas when* $\mu = 0$, *we have* $|B_y| \leq n^{6r-2} \log^3 n$, *and* $\tilde{\sigma}_y^2 \lesssim n^{6r} \log^3 n$.

Again, let $(x_{1\ell}, x_{2\ell}, ..., x_{n\ell}) \sim N(\theta, \mathbf{I}_n)$ and $(y_{1\ell}, y_{2\ell}, ..., y_{n\ell}) \sim N(\mu, \mathbf{I}_n)$ for $\ell = 1, 2$, and let $\widehat{T}_K = \sum_{i=1}^n \xi(x_{i1}, x_{i2}, y_{i1}, y_{i2})$. It follows that, when $0 < r < 1/4$, the bias can be bounded by

$$|B(\widehat{T}_K)| \lesssim \frac{\|\theta\|_\infty s}{\sqrt{\log n}} + \frac{\|\mu\|_\infty s}{\sqrt{\log n}} + \frac{s}{\log n}.$$

On the other hand, the variance of $\widehat{T(\theta, \mu)}$ is bounded by

$$\begin{aligned}
\mathrm{Var}(\widehat{T}_K) &\leq \sum_{i=1}^n \mathrm{Var}(\xi(x_{i1}, x_{i2}, y_{i1}, y_{i2})) \\
&\lesssim n^{12r+1} \log^6 n + n^{6r+1} \log^3 n \cdot (\|\theta\|_\infty^2 + \|\mu\|_\infty^2) && \text{(S2.9)} \\
&\lesssim |B(\widehat{T}_K)|^2, && \text{(S2.10)}
\end{aligned}$$

as long as $0 < r < \frac{2\beta - 1}{12}$. In this case, we have

$$\mathbb{E}(\widehat{T}_K - T(\theta, \mu))^2 \lesssim \frac{s^2}{\log^2 n} + \frac{s^2 \|\theta\|_\infty^2}{\log n} + \frac{s^2 \|\mu\|_\infty^2}{\log n}.$$

The final result follows from the fact that $\max(\|\theta\|_\infty, \|\mu\|_\infty) \leq L_n$.

S2.2 Proof of Theorem 3

Let $x_1, x_2 \sim N(\theta, 1)$ and $y_1, y_2 \sim N(\mu, 1)$. Define

$$\xi = \xi(x_1, x_2, y_1, y_2) = [|x_1| I(|x_2| > 2\sqrt{2\log n})] \times [|y_1| I(|y_2| > 2\sqrt{2\log n})]. \tag{S2.11}$$

Note that

$$\mathbb{E}(\xi) = \mathbb{E}|x_1| P(|x_2| > 2\sqrt{2\log n}) \times \mathbb{E}|y_1| P(|y_2| > 2\sqrt{2\log n}) \tag{S2.12}$$

Denote

$$\tilde{\sigma}_x^2 = \text{Var}(|x_1| I(|x_2| > 2\sqrt{2\log n})), \qquad \tilde{\sigma}_y^2 = \text{Var}(|y_1| I(|y_2| > 2\sqrt{2\log n})),$$

and

$$\tilde{\theta}_x = \mathbb{E}|x_1| P(|x_2| > 2\sqrt{2\log n}), \qquad \tilde{\mu}_y = \mathbb{E}|y_1| P(|y_2| > 2\sqrt{2\log n}).$$

Then we have

$$\text{Var}(\xi) = \tilde{\sigma}_x^2 \tilde{\sigma}_y^2 + \tilde{\sigma}_x^2 \tilde{\mu}_y^2 + \tilde{\sigma}_y^2 \tilde{\theta}_x^2$$

Proposition 4. *Let $B_x = \tilde{\theta}_x - |\theta|$ and $B_y = \tilde{\mu}_y - |\mu|$. For all $\theta, \mu \in \mathbb{R}$, we have*

$$|B_x| \lesssim \sqrt{\log n}, \qquad |B_y| \lesssim \sqrt{\log n},$$

and

$$\tilde{\sigma}_x^2 \lesssim \log n, \qquad \tilde{\sigma}_y^2 \lesssim \log n.$$

In particular, when $\theta = 0$ *we have* $|B_x| \lesssim n^{-4}$ *and* $\tilde{\sigma}_x^2 \lesssim n^{-4}$*, whereas when* $\mu = 0$ *we have* $|B_y| \lesssim n^{-4}$ *and* $\tilde{\sigma}_y^2 \lesssim n^{-4}$*.*

Proposition 5. *For all* $\theta, \mu \in \mathbb{R}$ *such that* $\mu, \theta \leq L_n$ *where* $L_n \leq \sqrt{2\log n}$*, we have*

$$|B_x| \lesssim L_n, \qquad |B_y| \lesssim L_n,$$

and

$$\tilde{\sigma}_x^2 \lesssim \frac{\log n}{n}, \qquad \tilde{\sigma}_y^2 \lesssim \frac{\log n}{n}.$$

Now the bias of the estimator ξ is

$$\begin{aligned} B(\xi) &= \mathbb{E}\xi - |\theta||\mu| = \tilde{\theta}_x\tilde{\mu}_y - |\theta||\mu| \\ &\leq |\theta||B_y| + |\mu||B_x| + |B_x||B_y|, \end{aligned} \tag{S2.13}$$

whereas the variance is bounded by

$$\mathrm{Var}(\xi) = \tilde{\sigma}_x^2\tilde{\sigma}_y^2 + \tilde{\sigma}_x^2\tilde{\mu}_y^2 + \tilde{\sigma}_y^2\tilde{\theta}_x^2 \lesssim \tilde{\sigma}_x^2\tilde{\sigma}_y^2 + |\mu|^2\tilde{\sigma}_x^2 + |\theta|^2\tilde{\sigma}_y^2 + \log n(\tilde{\sigma}_y^2 + \tilde{\sigma}_x^2). \tag{S2.14}$$

Now let $(x_{1\ell}, x_{2\ell}, ..., x_{n\ell}) \sim N(\theta, \mathbf{I}_n)$ and $(y_{1\ell}, y_{2\ell}, ..., y_{n\ell}) \sim N(\mu, \mathbf{I}_n)$ for $\ell = 1, 2$, and let

$$\widetilde{T} = \sum_{i=1}^{n} \xi(x_{i1}, x_{i2}, y_{i1}, y_{i2}).$$

Case I. $L_n \geq \sqrt{2\log n}$. It then follows that

$$|B(\widetilde{T})| \leq s\log n + (\|\theta\|_\infty + \|\mu\|_\infty)s\sqrt{\log n},$$

and

$$\mathrm{Var}(\widetilde{T}) \le s\log^2 n + (\|\theta\|_\infty^2 + \|\mu\|_\infty^2)s\log n. \tag{S2.15}$$

Therefore the mean squared error of $\widetilde{T}$ satisfies

$$\begin{aligned}\mathbb{E}(\widetilde{T} - T(\theta,\mu))^2 &\le B^2(\widetilde{T}) + \mathrm{Var}(\widetilde{T})\\ &\lesssim s^2\log n(\|\theta\|_\infty^2 + \|\mu\|_\infty^2 + \log n).\end{aligned}$$

The final result follows from the fact that $\max(\|\theta\|_\infty, \|\mu\|_\infty) \le L_n$.

Case II. $L_n < \sqrt{2\log n}$. It then follows that

$$|B(\widetilde{T})| \le sL_n^2 + (\|\theta\|_\infty + \|\mu\|_\infty)sL_n,$$

and

$$\mathrm{Var}(\widetilde{T}) \le \frac{\log^2 n}{n} + L_n^2\log n. \tag{S2.16}$$

Therefore the mean squared error of $\widetilde{T}$ satisfies

$$\begin{aligned}\mathbb{E}(\widetilde{T} - T(\theta,\mu))^2 &\le B^2(\widetilde{T}) + \mathrm{Var}(\widetilde{T})\\ &\lesssim s^2L_n^4 + \frac{\log^2 n}{n} + L_n^2\log n.\end{aligned}$$

S2.3 Proof of Theorem 4

By our sample splitting argument, it suffices to obtain the mean squared risk bound for the estimator $\widetilde{T} = \sum_{i=1}^n \hat{U}_i(x_i)\hat{U}_i(y_i)$ of $T(\theta,\mu) = \sum_{i=1}^n |\theta_i||\mu_i|$

where $\hat{U}_i(x_i) = |x_{i1}|I(|x_{i2}| > 2\sqrt{2\log n})$, $\hat{U}_i(y_i) = |y_{i1}|I(|y_{i2}| > 2\sqrt{2\log n})$ and $(x_{1\ell}, x_{2\ell}, ..., x_{n\ell}) \sim N(\theta, \mathbf{\Sigma}_1)$ and $(y_{1\ell}, y_{2\ell}, ..., y_{n\ell}) \sim N(\mu, \mathbf{\Sigma}_2)$ for $\ell = 1, 2$. By Proposition 3, the bias of the estimator $\xi_i = \hat{U}_i(x_i)\hat{U}_i(y_i)$ satisfies

$$B(\xi_i) = \mathbb{E}\xi_i - |\theta_i||\mu_i| \leq |\mu_i||B_{ix}| + |\theta_i||B_{iy}| + |B_{ix}||B_{iy}|, \tag{S2.17}$$

where $B_{ix} = \hat{U}_i(x_i) - |\theta_i|$ and $B_{iy} = \hat{U}_i(y_i) - |\mu_i|$. The variance satisfies

$$\mathrm{Var}(\xi_i) \lesssim \tilde{\sigma}_{ix}^2\tilde{\sigma}_{iy}^2 + |\mu_i|^2\tilde{\sigma}_{ix}^2 + |\theta_i|^2\tilde{\sigma}_{iy}^2 + \log n(\tilde{\sigma}_{iy}^2 + \tilde{\sigma}_{ix}^2), \tag{S2.18}$$

where $\tilde{\sigma}_{ix}^2 = \mathrm{Var}(\hat{U}_i(x_i))$ and $\tilde{\sigma}_{iy}^2 = \mathrm{Var}(\hat{U}_i(y_i))$. The covariance between two copies

$$\begin{aligned} \mathrm{Cov}(\xi_i, \xi_j) &= \mathbb{E}\xi_i\xi_j - \mathbb{E}\xi_i\mathbb{E}\xi_j \\ &= \mathbb{E}\hat{U}_i(x_i)\hat{U}_i(y_i)\hat{U}_j(x_j)\hat{U}_j(y_j) - \mathbb{E}\hat{U}_i(x_i)\hat{U}_i(y_i)\mathbb{E}\hat{U}_j(x_j)\hat{U}_j(y_j) \\ &= \mathbb{E}\hat{U}_i(x_i)\hat{U}_j(x_j)\mathbb{E}\hat{U}_i(y_i)\hat{U}_j(y_j) - \mathbb{E}\hat{U}_i(x_i)\mathbb{E}\hat{U}_j(x_j)\mathbb{E}\hat{U}_i(y_i)\mathbb{E}\hat{U}_j(y_j) \end{aligned}$$

Thus, we have

$$\begin{aligned} |\mathrm{Cov}(\xi_i, \xi_j)| \leq |\mathrm{Cov}(\hat{U}_i(x_i), \hat{U}_j(x_j))| \cdot \tilde{\mu}_i\tilde{\mu}_j + |\mathrm{Cov}(\hat{U}_i(y_i), \hat{U}_j(y_j))| \cdot \tilde{\theta}_i\tilde{\theta}_j \\ + |\mathrm{Cov}(\hat{U}_i(x_i), \hat{U}_j(x_j))\mathrm{Cov}(\hat{U}_i(y_i), \hat{U}_j(y_j))|, \end{aligned}$$

where $\tilde{\theta}_j = \mathbb{E}\hat{U}_j(x_j)$ and $\tilde{\mu}_j = \mathbb{E}\hat{U}_j(y_j)$. Note that

$$|\mathrm{Cov}(\hat{U}_i(x_i), \hat{U}_j(x_j))| \leq |\mathbb{E}\hat{U}_i(x_i)\hat{U}_j(x_j)| + |\tilde{\theta}_i\tilde{\theta}_j|$$

where

$$\mathbb{E}\hat{U}_i(x_i)\hat{U}_j(x_j) = \mathbb{E}|x_{i2}||x_{j2}|P(|x_{i2}| > 2\sqrt{2\log n}, |x_{j2}| > 2\sqrt{2\log n}).$$

Suppose one of θ_i and θ_j is 0, and the other bounded by L_n. Then by the proof of Proposition 3, we have

$$\mathbb{E}\hat{U}_i(x_i)\hat{U}_j(x_j) = O(n^{-4}L_n^2)$$

and

$$|\tilde{\theta}_i\tilde{\theta}_j| = O(n^{-4}L_n).$$

So

$$|\text{Cov}(\hat{U}_i(x_i), \hat{U}_j(x_j))| = O(n^{-4}L_n^2).$$

As a result, since $\mu_i, \mu_j \lesssim L_n$, we have

$$|\text{Cov}(\xi_i, \xi_j)| \le O(n^{-4}L_n^4).$$

Hence, summation over $O(n^2)$ terms will be bounded by $O(n^{-2}L_n^4)$. On the other hand, if neither θ_i or θ_j is zero, we have the trivial bound from Proposition 3

$$|\text{Cov}(\xi_i, \xi_j)| \le \log^2 n,$$

and the summation over $O(s^2)$ terms will be bounded by $s^2 \log^2 n$. Thus, as long as $L_n \lesssim \sqrt{n}$, we have

$$\mathrm{Var}(\widetilde{T}) \leq \sum_{i=1}^{n} \mathrm{Var}(\xi(x_{i1}, x_{i2}, y_{i1}, y_{i2})) + O(n^{-2} L_n^4) + O(s^2 \log^2 n) \lesssim s^2 \log^2 n. \tag{S2.19}$$

Now note that $|B(\widetilde{T})| \lesssim s \log n + (\|\theta\|_\infty + \|\mu\|_\infty) s \sqrt{\log n}$, it follows that

$$\mathbb{E}(\widetilde{T} - T(\theta, \mu))^2 \lesssim s^2 \log^2 n + L_n^2 s^2 \log n.$$

S3 Proof of Propositions 1-5

S3.1 Proof of Proposition 1

In the following, we divide into four cases according to the value of $|\theta|$. When $|\theta| = 0$, we show that we are actually estimating $|\theta|$ by 0. When $0 \leq |\theta| \leq \sqrt{2 \log n}$, we show that the estimator ξ' behaves essentially like δ, which is a good estimators when $|\theta|$ is small. When $\sqrt{2 \log n} < |\theta| \leq 4\sqrt{2 \log n}$, we show that the hybrid estimator ξ uses either $\delta(x_1)$ or $|x_1|$ and in this case both are good estimators of $|\theta|$. When $|\theta|$ is large, the hybrid estimator is essentially the same as $|x_1|$. We need the following lemmas to facilitate our proof.

Lemma 1. *Consider $G_K(x)$ defined in the main paper. The constant term* *of* $\tilde{G}_K(x) = \sum_{l=0}^K \tilde{g}_{2l} x^{2l}$*, with* $\tilde{g}_{2l} = M_n^{-2l+1} g_{2l}$*, satisfies*

$$\tilde{g}_0 = M_n g_0 \leq \frac{2M_n}{\pi(2K+1)}. \tag{S3.1}$$

Lemma 2. *Let* $X \sim N(\theta, 1)$ *and* $S_K(x) = \sum_{k=1}^K g_{2k} M_n^{-2k+1} H_{2k}(x)$*. Then for all* $|\theta| \leq 4\sqrt{2\log n}$*, we have*

$$\left| \mathbb{E} S_K(X) - |\theta| \right| \leq \frac{4M_n}{\pi(2K+1)},$$

and for $M_n^2 \geq K$*, we have* $\mathbb{E} S_K^2(X) \lesssim 2^{8K} M_n^2 K^2$*.*

Lemma 3. *Suppose* $I(A)$ *and* $I(B)$ *are indicator random variables independent of* X *and* Y*, with* $A \cap B = \emptyset$ *then*

$$\begin{aligned} \text{Var}(XI(A) + YI(B)) = & \text{Var}(X)P(A) + \text{Var}(Y)P(B) + (\mathbb{E}X)^2 P(A)P(A^c) \\ & + (\mathbb{E}Y)^2 P(B)P(B^c) - 2\mathbb{E}X\mathbb{E}Y P(A)P(B). \end{aligned} \tag{S3.2}$$

In particular, if $A^c = B$*, then we have*

$$\text{Var}(XI(A)+YI(A^c)) = \text{Var}(X)P(A)+\text{Var}(Y)P(A^c)+(\mathbb{E}X-\mathbb{E}Y)^2 P(A)P(A^c). \tag{S3.3}$$

Applying Lemma 3, we have

$$\begin{aligned}\tilde{\sigma}_x^2 &= \mathrm{Var}(\delta(x_1))P(\sqrt{2\log n} < |x_2| \le 2\sqrt{2\log n}) + \mathrm{Var}(|x_1|)P(|x_2| > 2\sqrt{2\log n})\\ &+ (\mathbb{E}\delta(x_1))^2 P(\sqrt{2\log n} < |x_2| \le 2\sqrt{2\log n})(1 - P(\sqrt{2\log n} < |x_2| \le 2\sqrt{2\log n}))\\ &+ (\mathbb{E}|x_1|)^2 P(|x_2| > 2\sqrt{2\log n})(1 - P(|x_2| > 2\sqrt{2\log n}))\\ &- 2\mathbb{E}\delta(x_1)\mathbb{E}|x_1| P(\sqrt{2\log n} < |x_2| \le 2\sqrt{2\log n})P(|x_2| > 2\sqrt{2\log n}).\end{aligned}$$

$$\begin{aligned}\tilde{\sigma}_y^2 &= \mathrm{Var}(\delta(y_1))P(\sqrt{2\log n} < |x_2| \le 2\sqrt{2\log n}) + \mathrm{Var}(|y_1|)P(|y_2| > 2\sqrt{2\log n})\\ &+ (\mathbb{E}\delta(y_1))^2 P(\sqrt{2\log n} < |y_2| \le 2\sqrt{2\log n})(1 - P(\sqrt{2\log n} < |y_2| \le 2\sqrt{2\log n}))\\ &+ (\mathbb{E}|y_1|)^2 P(|y_2| > 2\sqrt{2\log n})(1 - P(|y_2| > 2\sqrt{2\log n}))\\ &- 2\mathbb{E}\delta(y_1)\mathbb{E}|y_1| P(\sqrt{2\log n} < |y_2| \le 2\sqrt{2\log n})P(|y_2| > 2\sqrt{2\log n}).\end{aligned}$$

Case 1. $\theta = 0$. Note that $\delta(x_1)$ can be written as

$$\delta(x_1) = S_K(x_1) - (S_K(x_1) - n^2)I(S_K(x_1) \ge n^2).$$

Consequently,

$$\begin{aligned}|B_x| &= \big|\mathbb{E}(\delta(x_1))P(\sqrt{2\log n} < |x_2| \le 2\sqrt{2\log n}) + \mathbb{E}(|x_1|)P(|x_2| > 2\sqrt{2\log n})\big|\\ &\le |\mathbb{E}S_K(x_1)| + \mathbb{E}\{(S_K(x_1) - n^2)I(S_K(x_1) \ge n^2)\} + \mathbb{E}|x_1| P(|x_2| > 2\sqrt{2\log n})\\ &\equiv B_1 + B_2 + B_3.\end{aligned}$$

By definition of $S_K(x_1)$ we have

$$B_1 = 0. \tag{S3.4}$$

It follows from the standard bound for normal tail probability $\Phi(-z) \leq z^{-1}\phi(z)$ for $z > 0$ that

$$P(|x_2| > 2\sqrt{2\log n}) = 2\Phi(-2\sqrt{2\log n}) \leq \frac{1}{2\sqrt{\pi\log n}}n^{-4}. \tag{S3.5}$$

And in this case

$$\mathbb{E}|x_1| = 2\phi(0). \tag{S3.6}$$

It then follows that

$$B_3 \leq 2\phi(0) \cdot \frac{1}{2\sqrt{\pi\log n}}n^{-4} = \frac{1}{\pi\sqrt{2\log n}}n^{-4}. \tag{S3.7}$$

Now consider B_2. Note that for any random variable X and any constant $\lambda > 0$,

$$\mathbb{E}(XI(X \geq \lambda)) \leq \lambda^{-1}\mathbb{E}(X^2I(X \geq \lambda) \leq \lambda^{-1}\mathbb{E}X^2.$$

This together with Lemma 2 yields that

$$B_2 \leq \mathbb{E}\{(S_K(x_1)I(S_K(x_1) \geq n^2)\} \leq n^{-2}\mathbb{E}(S_K^2(x_1)) \lesssim n^{-2}\log n. \tag{S3.8}$$

Combining the three pieces together, we have

$$|B_x| \leq B_1 + B_2 + B_3 \lesssim \frac{\log n}{n^2}.$$

We now consider the variance. It follows that

$$
\begin{aligned}
\tilde{\sigma}_x^2 &\le \mathrm{Var}(S_K(x_1))P(\sqrt{2\log n} < |x_2| \le 2\sqrt{2\log n}) + \mathrm{Var}(|x_1|)P(|x_2| > 2\sqrt{2\log n}) \\
&\quad + (\mathbb{E}\delta(x_1))^2 + (\mathbb{E}|x_1|)^2 P(|x_2| > 2\sqrt{2\log n}) - 2\mathbb{E}\delta(x_1)\mathbb{E}|x_1|P(|x_2| > 2\sqrt{2\log n}) \\
&\le \mathbb{E}S_K^2(x_1)n^{-1} + (\mathbb{E}\delta(x_1))^2 + [\mathbb{E}x_1^2 + (\mathbb{E}|x_1|)^2 - 2\mathbb{E}\delta(x_1)\mathbb{E}|x_1|] \cdot \frac{1}{2n^4\sqrt{\pi\log n}} \\
&\lesssim n^{-1}2^{8K}M_n^2K^2 + n^{-4}\log^4 n + \frac{1}{n^4\sqrt{\log n}} \\
&\lesssim n^{-1}\log n
\end{aligned}
$$

where we use the fact that

$$
P(\sqrt{2\log n} < |x_2| \le 2\sqrt{2\log n}) \le P(\sqrt{2\log n} \le |x|) \le \Phi(-\sqrt{2\log n}) \le n^{-1}.
$$

Case 2. $0 < |\theta| \le \sqrt{2\log n}$. In this case

$$
\begin{aligned}
|B_x| &= |\mathbb{E}(\delta(x_1))P(\sqrt{2\log n} < |x_2| \le 2\sqrt{2\log n}) + \mathbb{E}(|x_1|)P(|x_2| > 2\sqrt{2\log n}) - |\theta|| \\
&\le |\mathbb{E}S_K(x_1) - |\theta|| + \mathbb{E}\{(S_K(x_1) - n^2)I(S_K(x_1) \ge n^2)\} \\
&\quad + (\mathbb{E}|x_1|)P(|x_2| > 2\sqrt{2\log n}) + |\theta|(1 - P(\sqrt{2\log n} < |x_2| \le 2\sqrt{2\log n})) \\
&\equiv B_1 + B_2 + B_3 + B_4
\end{aligned}
$$

From Lemma 2 we have

$$
B_1 \lesssim \sqrt{\log n}.
$$

Again, the standard bound for normal tail probability yields

$$
P(|x_2| > 2\sqrt{2\log n}) \le 2\Phi(-\sqrt{2\log n}) \le \frac{1}{\sqrt{\pi\log n}}n^{-1}
$$

Note that

$$\mathbb{E}|x_1| = |\theta| + 2\phi(\theta) - 2|\theta|\Phi(-|\theta|) \leq |\theta| + 1 \leq \sqrt{2\log n} + 1.$$

Then we have

$$B_3 \leq \left(\sqrt{2\log n} + 1\right) \cdot \frac{1}{\sqrt{\pi \log n}} n^{-1} \leq 3n^{-1},$$

and

$$B_4 \leq |\theta| \leq \sqrt{2\log n}.$$

Note that B_2 follows (S3.8), and we have

$$|B_x| \leq B_1 + B_2 + B_3 + B_4 \lesssim \sqrt{\log n}.$$

For the variance, note that

$$\begin{aligned}
(\mathbb{E}\delta(x_1))^2 &\leq \mathbb{E}\delta^2(x_1) \\
&= \mathbb{E}(\min\{S_K^2(x_1), n^4\}) \\
&= \mathbb{E}\big[S_K^2(x_1) - (S_K^2(x_1) - n^4) I(S_K^2(x_1) > n^4)\big] \\
&\leq \mathbb{E}S_K^2(x_1) \\
&\lesssim \log n,
\end{aligned}$$

and

$$\mathbb{E}x_1^2 + (\mathbb{E}|x_1|)^2 - 2\mathbb{E}\delta(x_1)\mathbb{E}|x_1| \leq \mathrm{Var}(x_1) + (\mathbb{E}|x_1|)^2 + (\sqrt{2\log n})^2 \lesssim \log n$$

Then we have

$$
\begin{aligned}
\tilde{\sigma}_x^2 &\leq \mathrm{Var}(S_K(x_1)) + \mathrm{Var}(|x_1|)P(|x_2| > 2\sqrt{2\log n}) + (\mathbb{E}\delta(x_1))^2 \\
&\quad + (\mathbb{E}|x_1|)^2 P(|x_2| > 2\sqrt{2\log n}) + 2|\mathbb{E}\delta(x_1)| \cdot \mathbb{E}|x_1| P(|x_2| > 2\sqrt{2\log n}) \\
&\leq \mathbb{E}S_K^2(x_1) + (\mathbb{E}\delta(x_1))^2 + [\mathbb{E}x_1^2 + (\mathbb{E}|x_1|)^2 + 2|\mathbb{E}\delta(x_1)| \cdot \mathbb{E}|x_1|] \cdot \frac{1}{n\sqrt{\pi \log n}} \\
&\leq 2\log n + \frac{5\sqrt{\log n}}{\sqrt{\pi} n} \\
&\lesssim \log n.
\end{aligned}
$$

Case 3 $\sqrt{2\log n} \leq |\theta| \leq 4\sqrt{2\log n}$. In this case,

$$
\begin{aligned}
|B_x| &= |\mathbb{E}(\delta(x_1))P(\sqrt{2\log n} < |x_2| \leq 2\sqrt{2\log n}) + \mathbb{E}(|x_1|)P(|x_2| > 2\sqrt{2\log n}) - |\theta|| \\
&\leq |\mathbb{E}(\delta(x_1)) - |\theta|| + |\mathbb{E}|x_1| - |\theta|| + |\theta| P(|x_2| \leq 2\sqrt{2\log n}) \\
&\leq |\mathbb{E}S_K(x_1) - |\theta|| + \mathbb{E}\{(S_K(x_1) - n^2) I(S_K(x_1) \geq n^2)\} + 2\phi(\theta) + 4\sqrt{2\log n} \\
&\lesssim \sqrt{\log n} + n^{-2}\log n + n^{-1} \\
&\lesssim \sqrt{\log n}.
\end{aligned}
$$

For the variance, similarly since

$$
\begin{aligned}
\mathbb{E}x_1^2 + (\mathbb{E}|x_1|)^2 + 2|\mathbb{E}\delta(x_1)| \cdot \mathbb{E}|x_1| &\leq \mathrm{Var}(x_1) + (\mathbb{E}|x_1|)^2 + (4\sqrt{2\log n})^2 \\
&\quad + \left(4\sqrt{2\log n} + \frac{2M_n}{\pi K}\right) 4\sqrt{2\log n} \\
&\leq 1 + (32 + 32 + 32)\log n + \frac{64\sqrt{2}\log n}{\pi K} \\
&\lesssim \log n,
\end{aligned}
$$

and then

$$\begin{aligned}\tilde{\sigma}_x^2 &\le \text{Var}(S_K(x_1)) + \text{Var}(|x_1|) + (\mathbb{E}\delta(x_1))^2 + (\mathbb{E}|x_1|)^2 + 2|\mathbb{E}\delta(x_1)| \cdot \mathbb{E}|x_1| \\ &\le \mathbb{E}S_K^2(x_1) + (\mathbb{E}\delta(x_1))^2 + [\mathbb{E}x_1^2 + (\mathbb{E}|x_1|)^2 + 2\mathbb{E}\delta(x_1)\mathbb{E}|x_1|] \\ &\lesssim \log n.\end{aligned}$$

Case 4. $|\theta| > 4\sqrt{2\log n}$. In this case, the standard bound for normal tail probability yields that

$$P(|x_2| \le 2\sqrt{2\log n}) \le 2\Phi(-(|\theta|/-2\sqrt{2\log n})) \le 2\Phi\bigg(-\frac{|\theta|}{2}\bigg) \le \frac{4}{|\theta|}\phi\bigg(\frac{|\theta|}{2}\bigg).$$

In particular,

$$P(|x_2| \le 2\sqrt{2\log n}) \le 2\Phi(-2\sqrt{2\log n}) \le \frac{1}{2\sqrt{\pi\log n}}n^{-4}.$$

Also note that

$$\begin{aligned}\mathbb{E}\delta(x_1) &= \mathbb{E}\min\{S_K(x_1), n^2\} \\ &= \mathbb{E}(S_K(x_1)1\{S_K(x_1) \le n^2\} + n^2 1\{S_K(x_1) > n^2\}) \\ &\le n^2\end{aligned}$$

Hence,

$$
\begin{aligned}
|B_x| &\leq |\mathbb{E}(\delta(x_1))P(\sqrt{2\log n} < |x_2| \leq 2\sqrt{2\log n}) + \mathbb{E}(|x_1|)P(|x_2| > 2\sqrt{2\log n}) - |\theta|| \\
&\leq |\mathbb{E}(\delta(x_1))|P(\sqrt{2\log n} < |x_2| \leq 2\sqrt{2\log n}) + |\mathbb{E}|x_1| - |\theta|| \\
&\quad + \mathbb{E}|x_1|P(|x_2| \leq 2\sqrt{2\log n}) \\
&\leq |\mathbb{E}|x_1| - |\theta|| + (|\mathbb{E}\delta(x_1)| + \mathbb{E}|x_1|)P(|x_2| \leq 2\sqrt{2\log n}) \\
&\leq 2\phi(\theta) + (n^2 + |\theta| + 1)P(|x_2| \leq 2\sqrt{2\log n}) \\
&\leq 2\phi(\theta) + 4\phi\left(\frac{|\theta|}{2}\right) + \frac{1}{2}n^{-2} \\
&\leq 6\phi\left(\frac{\theta}{2}\right) + \frac{1}{2n^2} \\
&\leq \frac{1}{n^2}.
\end{aligned}
$$

For the variance, similarly we have

$$
\begin{aligned}
\tilde{\sigma}_x^2 &\leq \mathrm{Var}(\delta(x_1))P(\sqrt{2\log n} < |x_2| \leq 2\sqrt{2\log n}) + \mathrm{Var}(|x_1|) \\
&\quad + (\mathbb{E}|x_1|)^2 P(|x_2| \leq 2\sqrt{2\log n}) + (\mathbb{E}\delta(x_1))^2 P(\sqrt{2\log n} < |x_2| \leq 2\sqrt{2\log n}) \\
&\quad + 2|\mathbb{E}\delta(x_1)| \cdot \mathbb{E}|x_1| P(\sqrt{2\log n} < |x_2| \leq 2\sqrt{2\log n}) \\
&\leq 1 + [\mathrm{Var}(\delta(x_1)) + (\mathbb{E}\delta(x_1))^2 + (\mathbb{E}|x_1|)^2 + 2|\mathbb{E}\delta(x_1)| \cdot E|x_1|]P(|x_2| \leq 2\sqrt{2\log n}) \\
&\leq 1 + [\log n + (n^2 + |\theta| + 1)^2]P(|x_2| \leq 2\sqrt{2\log n}) \\
&= 1 + o(1).
\end{aligned}
$$

Obviously, the same argument holds for y_1, y_2 and $|\mu|$.

S3.2 Proof of Proposition 2

When $|\theta| \leq L_n \leq (\sqrt{2}-1)\sqrt{\log n}$, we have

$$
\begin{aligned}
|B_x| &= |\mathbb{E}(\delta(x_1))P(\sqrt{2\log n} < |x_2| \leq 2\sqrt{2\log n}) + \mathbb{E}(|x_1|)P(|x_2| > 2\sqrt{2\log n}) - |\theta|| \\
&\leq |\mathbb{E}S_K(x_1) - |\theta||P(\sqrt{2\log n} < |x_2| \leq 2\sqrt{2\log n}) \\
&\quad + \mathbb{E}\{(S_K(x_1) - n^2)I(S_K(x_1) \geq n^2)\} \\
&\quad + (\mathbb{E}|x_1|)P(|x_2| > 2\sqrt{2\log n}) + |\theta|(1 - P(\sqrt{2\log n} < |x_2| \leq 2\sqrt{2\log n})) \\
&\equiv B_1 + B_2 + B_3 + B_4
\end{aligned}
$$

Note that

$$
\begin{aligned}
P(\sqrt{2\log n} < |x_2| \leq 2\sqrt{2\log n}) &\leq P(\sqrt{2\log n} < |x_2|) \\
&\leq P(\sqrt{2\log n} - L_n < |z|) \\
&\leq 2\Phi(-\sqrt{2\log n} + L_n) \\
&\lesssim \frac{1}{n^{1/2}}.
\end{aligned}
$$

From Lemma 2 we have

$$
B_1 \lesssim \sqrt{\log n}P(\sqrt{2\log n} < |x_2| \leq 2\sqrt{2\log n}) \leq \sqrt{\frac{\log n}{n}}.
$$

Again, the standard bound for normal tail probability yields

$$
P(|x_2| > 2\sqrt{2\log n}) \leq 2\Phi(-\sqrt{2\log n}) \leq \frac{1}{\sqrt{\pi\log n}}n^{-1}
$$

Note that

$$\mathbb{E}|x_1| = |\theta| + 2\phi(\theta) - 2|\theta|\Phi(-|\theta|) \le |\theta| + 1 \le L_n + 1.$$

Then we have

$$B_3 \le \left(\sqrt{2\log n} + 1\right) \cdot \frac{1}{\sqrt{\pi \log n}} n^{-1} \le 3n^{-1},$$

and

$$B_4 \le |\theta| \le L_n.$$

Note that B_2 follows (S3.8), and we have

$$|B_x| \le B_1 + B_2 + B_3 + B_4 \lesssim L_n.$$

For the variance, note that

$$\begin{aligned}
(\mathbb{E}\delta(x_1))^2 &\le \mathbb{E}\delta^2(x_1) \\
&= \mathbb{E}(\min\{S_K^2(x_1), n^4\}) \\
&= \mathbb{E}\big[S_K^2(x_1) - (S_K^2(x_1) - n^4) I(S_K^2(x_1) > n^4)\big] \\
&\le \mathbb{E}S_K^2(x_1) \\
&\lesssim \log n,
\end{aligned}$$

and

$$\mathbb{E}x_1^2 + (\mathbb{E}|x_1|)^2 - 2\mathbb{E}\delta(x_1)\mathbb{E}|x_1| \le \mathrm{Var}(x_1) + (\mathbb{E}|x_1|)^2 + (\sqrt{2\log n})^2 \lesssim \log n$$

Then we have

$$\begin{aligned}\tilde{\sigma}_x^2 &\le \mathrm{Var}(S_K(x_1))P(\sqrt{2\log n} < |x_2| \le 2\sqrt{2\log n}) + \mathrm{Var}(|x_1|)P(|x_2| > 2\sqrt{2\log n}) \\ &\quad + (\mathbb{E}\delta(x_1))^2 + (\mathbb{E}|x_1|)^2 P(|x_2| > 2\sqrt{2\log n}) - 2\mathbb{E}\delta(x_1)\mathbb{E}|x_1|P(|x_2| > 2\sqrt{2\log n}) \\ &\le \mathbb{E}S_K^2(x_1)n^{-1} + (\mathbb{E}\delta(x_1))^2 + [\mathbb{E}x_1^2 + (\mathbb{E}|x_1|)^2 - 2\mathbb{E}\delta(x_1)\mathbb{E}|x_1|] \cdot \frac{1}{n\sqrt{\pi\log n}} \\ &\lesssim n^{-1/2}2^{8K}M_n^2K^2 \\ &\lesssim \frac{\log n}{\sqrt{n}}\end{aligned}$$

S3.3 Proof of Proposition 3

We only prove the proposition for θ. The argument for μ is the same. We need the following lemma for the proof.

Lemma 4. *Let $X \sim N(\theta, 1)$ and $S_K(x) = \sum_{k=1}^K g_{2k}M_n^{-2k+1}H_{2k}(x)$ with $M_n = 8\sqrt{\log n}$ and $K = r\log n$ for some $r > 0$. Then for all $|\theta| \le 4\sqrt{2\log n}$,*

$$\left|\mathbb{E}S_K(X) - |\theta|\right| \le \frac{4M_n}{\pi(2K+1)} \lesssim \frac{1}{\sqrt{\log n}},$$

and $\mathbb{E}S_K^2(X) \lesssim n^{6r}\log^3 n$.

Case 1. $\theta = 0$. Note that $\delta(x_1)$ can be written as

$$\delta(x_1) = S_K(x_1) - (S_K(x_1) - n^2)I(S_K(x_1) \ge n^2).$$

Consequently,

$$|B_x| \le |\mathbb{E}S_K(x_1)| + \mathbb{E}\{(S_K(x_1) - n^2)I(S_K(x_1) \ge n^2)\}$$
$$+ (|\mathbb{E}S_K(x_1)| + \mathbb{E}|x_1|)\mathbb{P}(|x_2| > 2\sqrt{2\log n})$$
$$\equiv B_1 + B_2 + B_3.$$

By definition of $S_K(x_1)$ we have

$$B_1 = 0. \tag{S3.9}$$

It follows from the standard bound for normal tail probability $\Phi(-z) \le z^{-1}\phi(z)$ for $z > 0$ that

$$\mathbb{P}(|x_2| > 2\sqrt{2\log n}) = 2\Phi(-2\sqrt{2\log n}) \le \frac{1}{2\sqrt{\pi\log n}}n^{-4}. \tag{S3.10}$$

And in this case

$$\mathbb{E}|x_1| = 2\phi(0). \tag{S3.11}$$

It then follows that

$$B_3 \le 2\phi(0)\cdot\frac{1}{2\sqrt{\pi\log n}}n^{-4} = \frac{1}{\pi\sqrt{2\log n}}n^{-4}. \tag{S3.12}$$

Now consider B_2. Note that for any random variable X and any constant $\lambda > 0$,

$$\mathbb{E}(XI(X \ge \lambda)) \le \lambda^{-1}\mathbb{E}(X^2I(X \ge \lambda) \le \lambda^{-1}\mathbb{E}X^2.$$

This together with Lemma 4 yields that

$$B_2 \le \mathbb{E}\{(S_K(x_1)I(S_K(x_1) \ge n^2)\} \le n^{-2}\mathbb{E}(S_K^2(x_1)) \lesssim n^{6r-2}\log^3 n. \tag{S3.13}$$

Combining the three pieces together, we have

$$|B_x| \leq B_1 + B_2 + B_3 \lesssim n^{6r-2} \log^3 n.$$

We now consider the variance. It follows from Lemma 3 and Lemma 4 that

$$\begin{aligned}
\tilde{\sigma}_x^2 &\leq \text{Var}(S_K(x_1))P(|x_2| \leq 2\sqrt{2\log n}) + \text{Var}(|x_1|)P(|x_2| > 2\sqrt{2\log n}) \\
&\quad + [(\mathbb{E}\delta(x_1))^2 + (\mathbb{E}|x_1|)^2]P(|x_2| > 2\sqrt{2\log n}) - 2\mathbb{E}\delta(x_1)\mathbb{E}|x_1|P(|x_2| > 2\sqrt{2\log n}) \\
&\leq \mathbb{E}S_K^2(x_1) + [(\mathbb{E}\delta(x_1))^2 + (\mathbb{E}|x_1|)^2 - 2\mathbb{E}\delta(x_1)\mathbb{E}|x_1|] \cdot \frac{1}{2n^4\sqrt{\pi \log n}} \\
&\lesssim n^{6r} \log^3 n
\end{aligned}$$

as long as $r < 3/4$.

Case 2. $0 < |\theta| \leq \sqrt{2\log n}$. In this case

$$\begin{aligned}
|B_x| &= |\mathbb{E}(\delta(x_1))P(|x_2| \leq 2\sqrt{2\log n}) + \mathbb{E}(|x_1|)P(|x_2| > 2\sqrt{2\log n}) - |\theta|| \\
&\leq |\mathbb{E}(\delta(x_1)) - |\theta|| + |\mathbb{E}|x_1| - |\theta||P(|x_2| > 2\sqrt{2\log n}) \\
&\lesssim |\mathbb{E}S_K(x_1) - |\theta|| + \mathbb{E}\{(S_K(x_1) - n^2)I(S_K(x_1) \geq n^2)\} + n^{-4} \\
&\lesssim 1/\sqrt{\log n} + n^{6r-2}\log^3 n + n^{-4} \\
&\lesssim 1/\sqrt{\log n}
\end{aligned}$$

as long as $r < 1/3$. Similarly, note that

$$\mathbb{E}\delta(x_1) = \mathbb{E}S_K(x_1) + B_2$$

then

$$
\begin{aligned}
(\mathbb{E}\delta(x_1) - \mathbb{E}|x_1|)^2 &\leq 2(\mathbb{E}\delta(x_1))^2 + 2(\mathbb{E}|x_1|)^2 \\
&\leq 2(\mathbb{E}S_K(x_1) + B_2)^2 + 2(|\theta| + 1)^2 \\
&\lesssim n^{6r}\log^3 n.
\end{aligned}
$$

Hence the variance can be bounded as follows.

$$
\begin{aligned}
\tilde{\sigma}_x^2 &\leq \mathbb{E}S_K^2(x_1) + \mathbb{E}x_1^2\mathbb{P}(|x_2| > 2\sqrt{2\log n}) \\
&\quad + (\mathbb{E}\delta(x_1) - \mathbb{E}|x_1|)^2\mathbb{P}(|x_2| > 2\sqrt{2\log n}) \\
&\lesssim n^{6r}\log^3 n.
\end{aligned}
$$

Case 3. $\sqrt{2\log n} < |\theta| \leq 4\sqrt{2\log n}$. In this case

$$
\begin{aligned}
|B_x| &= |\mathbb{E}(\delta(x_1))P(|x_2| \leq 2\sqrt{2\log n}) + \mathbb{E}(|x_1|)P(|x_2| > 2\sqrt{2\log n}) - |\theta|| \\
&\leq |\mathbb{E}(\delta(x_1)) - |\theta|| + |\mathbb{E}|x_1| - |\theta|| \\
&\lesssim |\mathbb{E}S_K(x_1) - |\theta|| + \mathbb{E}\{(S_K(x_1) - n^2)I(S_K(x_1) \geq n^2)\} + 2\phi(\theta) \\
&\lesssim 1/\sqrt{\log n} + n^{6r-2}\log^3 n + n^{-1} \\
&\lesssim 1/\sqrt{\log n}
\end{aligned}
$$

as long as $r < 1/3$. The variance can be bounded similar to the Case 2.

Case 4. $|\theta| > 4\sqrt{2\log n}$. In this case, same argument follows from the proof of Case 4 in Proposition 1.

Proof of Proposition 4. Similar to the proofs of previous propositions, we only prove the statements about θ. The argument for μ is the same.

Case 1. $\theta = 0$. Note that $|B_x| = \mathbb{E}(|x_1|)P(|x_2| > 2\sqrt{2\log n})$. It follows from the standard bound for normal tail probability $\Phi(-z) \le z^{-1}\phi(z)$ for $z > 0$ that

$$P(|x_2| > 2\sqrt{2\log n}) = 2\Phi(-2\sqrt{2\log n}) \le \frac{1}{2\sqrt{\pi\log n}} n^{-4}. \tag{S3.14}$$

And in this case $E|x_1| = 2\phi(0)$. It then follows that

$$|B_x| \le 2\phi(0)\cdot\frac{1}{2\sqrt{\pi\log n}} n^{-4} = \frac{1}{\pi\sqrt{2\log n}} n^{-4}. \tag{S3.15}$$

We now consider the variance. It follows that

$$\tilde{\sigma}_x^2 \le \mathrm{Var}(|x_1|)P(|x_2| > 2\sqrt{2\log n}) \le \mathbb{E}x_1^2\cdot\frac{1}{2n^4\sqrt{\pi\log n}} \lesssim \frac{1}{n^4\sqrt{\log n}}.$$

Case 2. $0 < |\theta| \le \sqrt{2\log n}$. In this case

$$|B_x| = \big|\mathbb{E}|x_1|P(|x_2| > 2\sqrt{2\log n}) - \theta\big| \le \mathbb{E}|x_1|P(|x_2| > 2\sqrt{2\log n}) + |\theta|.$$

The standard bound for normal tail probability yields

$$P(|x_2| > 2\sqrt{2\log n}) \le 2\Phi(-\sqrt{2\log n}) \le \frac{1}{\sqrt{\pi\log n}} n^{-1}$$

Note that

$$\mathbb{E}|x_1| = |\theta| + 2\phi(\theta) - 2|\theta|\Phi(-|\theta|) \le |\theta| + 1 \le \sqrt{2\log n} + 1.$$

Then we have

$$|B_x| \le \left(\sqrt{2\log n}+1\right)\cdot \frac{1}{\sqrt{\pi \log n}} n^{-1} + |\theta| \le 3n^{-1} + \sqrt{2\log n}.$$

On the other hand, since

$$\mathbb{E}x_1^2 \le \mathrm{Var}(x_1) + (\mathbb{E}|x_1|)^2 \le 1 + 2\log n.$$

we have

$$\tilde{\sigma}_x^2 \le \mathrm{Var}(|x_1|)P(|x_2| > 2\sqrt{2\log n}) \le \mathbb{E}x_1^2 \cdot \frac{1}{n\sqrt{\pi\log n}} \lesssim \frac{\sqrt{\log n}}{n}.$$

Case 3 $\sqrt{2\log n} \le |\theta| \le 4\sqrt{2\log n}$. In this case 2,

$$\begin{aligned} |B_x| &= |\mathbb{E}(|x_1|)P(|x_2| > 2\sqrt{2\log n}) - |\theta|| \\ &\le |\mathbb{E}|x_1| - |\theta|| + |\theta| P(|x_2| \le 2\sqrt{2\log n}) \\ &\lesssim \sqrt{\log n}. \end{aligned}$$

For the variance, similarly we have

$$\tilde{\sigma}_x^2 \le \mathbb{E}x_1^2 \lesssim \log n.$$

Case 4. $|\theta| > 4\sqrt{2\log n}$. In this case, the standard bound for normal tail probability yields that

$$P(|x_2| \le 2\sqrt{2\log n}) \le 2\Phi(-(|\theta| - 2\sqrt{2\log n})) \le 2\Phi\left(-\frac{|\theta|}{2}\right) \le \frac{4}{|\theta|}\phi\left(\frac{|\theta|}{2}\right).$$

In particular,

$$P(|x_2| \le 2\sqrt{2\log n}) \le 2\Phi(-2\sqrt{2\log n}) \le \frac{1}{2\sqrt{\pi \log n}} n^{-4}.$$

Hence,

$$|B_x| \le |\mathbb{E}(|x_1|)P(|x_2| > 2\sqrt{2\log n}) - |\theta|| \le |\mathbb{E}|x_1| - |\theta|| + \mathbb{E}|x_1|P(|x_2| \le 2\sqrt{2\log n})$$

$$\le 2\phi(\theta) + (|\theta|+1)P(|x_2| \le 2\sqrt{2\log n}) \le 2\phi(\theta) + 4\phi\left(\frac{|\theta|}{2}\right) \lesssim \frac{1}{n^2}.$$

For the variance, similarly we have

$$\begin{aligned}\tilde{\sigma}_x^2 &\le \mathrm{Var}(|x_1|) + (\mathbb{E}|x_1|)^2 P(|x_2| \le 2\sqrt{2\log n}) \\ &= 1 + o(1).\end{aligned}$$

Proof of Proposition 5. In this case

$$|B_x| = \left|\mathbb{E}|x_1|P(|x_2| > 2\sqrt{2\log n}) - \theta\right| \le \mathbb{E}|x_1|P(|x_2| > 2\sqrt{2\log n}) + |\theta|.$$

The standard bound for normal tail probability yields

$$P(|x_2| > 2\sqrt{2\log n}) \le 2\Phi(-\sqrt{2\log n}) \le \frac{1}{\sqrt{\pi \log n}} n^{-1}$$

Note that

$$\mathbb{E}|x_1| = |\theta| + 2\phi(\theta) - 2|\theta|\Phi(-|\theta|) \le |\theta| + 1 \le L_n + 1.$$

Then we have

$$|B_x| \le \left(L_n + 1\right) \cdot \frac{1}{\sqrt{\pi \log n}} n^{-1} + |\theta| \lesssim L_n.$$

On the other hand, since

$$\mathbb{E}x_1^2 \le \mathrm{Var}(x_1) + (\mathbb{E}|x_1|)^2 \le 1 + (1 + L_n)^2.$$

we have

$$\tilde{\sigma}_x^2 \le \mathrm{Var}(|x_1|)P(|x_2| > 2\sqrt{2\log n}) \le \mathbb{E}x_1^2 \cdot \frac{1}{n\sqrt{\pi \log n}} \lesssim \frac{\sqrt{\log n}}{n}.$$

S4 Proofs of Technical Lemmas

Proof of Lemma 1. By Lemma 2 of Cai and Low (2011), for $x \in [-1, 1]$, we have

$$\max_{x\in[-1,1]} |G_K(x) - |x|| \le \frac{2}{\pi(2K+1)}.$$

Then for $x' \in [-M_n, M_n]$, we have $x' = M_n x$, and thus

$$\max_{x'} |\tilde{G}_K(x') - |x'|| \le \frac{2M_n}{\pi(2K+1)}. \tag{S4.1}$$

Set $x' = 0$, we obtain the statement.

Proof of Lemma 2. The first statement follows from Lemma 2 in Cai and Low (2011) and that

$$\begin{aligned} \left|\mathbb{E}S_K(X) - |\theta|\right| &\le \left| \sum_{k=0}^{K} g_{2k} M_n^{-2k+1} \theta^{2k} - |\theta| \right| + \frac{2M_n}{\pi(2K+1)} &\text{(S4.2)} \\ &\le \frac{4M_n}{\pi(2K+1)}. &\text{(S4.3)} \end{aligned}$$

To bound $\mathbb{E}S_K^2(X)$, it follows from Lemma 3 in Cai and Low (2011) and that

$$\mathbb{E}S_K^2(X) \le \left(\sum_{k=1}^{K} |g_{2k}| M_n^{-2k+1} (\mathbb{E}H_{2k}^2(X))^{1/2}\right)^2 \tag{S4.4}$$
$$\le 2^{6K}\left(\sum_{k=1}^{K} M_n^{-2k+1} (2M_n^2)^k\right)^2 \tag{S4.5}$$
$$\le 2^{8K} M_n^2 K^2 \tag{S4.6}$$

Proof of Lemma 3. For equation (S3.2), since events A and B are independent of random variables X and Y, we have

$$\begin{aligned}
&\mathrm{Var}(XI(A) + YI(B)) \\
&= \mathbb{E}\big[X^2 I(A) + Y^2 I(B) + 2XYI(A)I(B)\big] - \big(\mathbb{E}XP(A) + \mathbb{E}YP(B)\big)^2 \\
&= \mathbb{E}X^2 P(A) + \mathbb{E}Y^2 P(B) - (\mathbb{E}X)^2 P^2(A) - (\mathbb{E}Y)^2 P^2(B) \\
&\quad - 2\mathbb{E}X\mathbb{E}YP(A)P(B) \\
&= \mathrm{Var}(X)P(A) + \mathrm{Var}(Y)P(B) + (\mathbb{E}X)^2(P(A) - P^2(A)) \\
&\quad + (\mathbb{E}Y)^2(P(B) - P^2(B)) - 2\mathbb{E}X\mathbb{E}YP(A)P(B)
\end{aligned}$$

Equation (S3.3) follows directly from the above derivation.

Proof of Lemma 4. It follows from Lemma 2 in Cai and Low (2011) that

$$\left|\mathbb{E}S_K(X) - |\theta|\right| = \left|\sum_{k=0}^{K} g_{2k} M_n^{-2k+1}\theta^{2k} - |\theta|\right| + \frac{2M_n}{\pi(2K+1)} \tag{S4.7}$$

$$\leq \frac{4M_n}{\pi(2K+1)}. \tag{S4.8}$$

To bound $\mathbb{E}S_K^2(X)$, it follows that when $K = r\log n$ for some $r > 0$,

$$\mathbb{E}S_K^2(X) \leq \left(\sum_{k=1}^{K} |g_{2k}| M_n^{-2k+1} (EH_{2k}^2(X))^{1/2}\right)^2 \tag{S4.9}$$

$$\leq 2^{6K}\left(\sum_{k=1}^{K} M_n^{-2k+1}(2M_n^2)^k\right)^2 \tag{S4.10}$$

$$\leq 2^{8K} M_n^2 K^2 \tag{S4.11}$$

$$\lesssim n^{6r}\log^3 n. \tag{S4.12}$$

S5 Supplementary Theoretical Discussions

S5.1 Minimax Optimal Rate When $\beta = 1/2$

When $\beta = 1/2$, we consider the following estimator

$$\widetilde{T}_0 = 2\sum_{i=1}^{n} \hat{U}_i(x_i)\hat{U}_i(y_i),$$

where

$$\hat{U}_i(x_i) = (|x_{i1}| - \alpha)\cdot I(|x_{i2}| > \sqrt{2\log 2}),$$

and

$$\alpha = \frac{\mathbb{E}[|\xi| I(\xi^2 > 2\log 2)]}{P(\xi^2 > 2\log 2)}, \qquad \text{for } \xi \sim N(0,1).$$

Following the similar arguments as in the proof of Theorem 4 of Collier et al. (2020) as well as the proof of Theorem 2 of the main paper, it can be shown that

$$\sup_{(\theta,\mu,\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in D^{\infty}(s,L_n)} \mathcal{R}(\widetilde{T}_0) \lesssim \frac{s^2(L_n^2+1)}{n^2}. \tag{S5.1}$$

Comparing the above risk upper bound to the minimax lower bound in Theorem 1, we have, for $\beta = 1/2$ and $L_n \ge \sqrt{\log(1+n/s^2)} \asymp 1$,

$$\inf_{\hat{T}} \sup_{(\theta,\mu,\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in D_0^{\infty}(s,L_n)} \mathcal{R}(\hat{T}) \asymp \frac{s^2 L_n^2}{n^2}. \tag{S5.2}$$

S5.2 Complexities from the Covariances

In the main paper, most of our theoretical results are obtained under the assumption $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{I}$, whereas our Theorem 4 essentially take the worst case over these covariances. In this section, we discuss the cases with known and unknown covariances and the corresponding estimators.

Known covariances. On the one hand, if the diagonals of the covariance matrices are all 1's, while the off-diagonal entries are known and possibly nonzero, then in principle our proposed estimators can still be applied, although the rate of convergence might not remain the same, nor does the minimax optimality property. Nevertheless, the analysis of these estimators can be technically challenging. For example, obtaining the risk upper bound of $\widehat{T}_K$ (or $\widehat{T}_K^S$) requires calculation of the covariances between the hybrid components $\hat{V}_{i,K}(x_i)$ (or $\hat{V}_{i,K}^S(x_i)$) for correlated x_i's, which relies on properties of Hermite polynomials. The final risk upper bounds will depend on the specific covariance structure.

On the other hand, if the diagonals are not 1, then an extra rescaling step is needed in our construction of the polynomial approximation based estimators. Specifically, suppose $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ have diagonal entries $\{\sigma_i^2\}_{i=1}^n$ and $\{\tau_i^2\}_{i=1}^n$, respectively, then we can define the adjusted version of $\widehat{T}_K$ as

$$\widehat{T}_K' = 2\sum_{i=1}^n \hat{V}_{i,K}'(x_i)\hat{V}_{i,K}'(y_i),$$

where

$$\hat{V}_{i,K}'(x_i) = \delta_K'(x_{i1})I(|x_{i2}| \le 2\sigma_i\sqrt{2\log n}) + |x_{i1}|I(|x_{i2}| > 2\sigma_i\sqrt{2\log n}),$$

$$\hat{V}_{i,K}'(y_i) = \delta_K'(y_{i1})I(|y_{i2}| \le 2\tau_i\sqrt{2\log n}) + |y_{i1}|I(|y_{i2}| > 2\tau_i\sqrt{2\log n}),$$

and

$$\delta_K'(x_i) = \sigma_i \min\{S_K(x_i/\sigma_i), n^2\}, \qquad \delta_K'(y_i) = \tau_i \min\{S_K(y_i/\tau_i), n^2\}.$$

Similarly, one can define the adjusted version of $\widehat{T}_K^S$ that takes into account the knowledge of the variances.

Unknown variances. When the covariances are completely unknown, the estimation problem will be extreme difficult since one may not be able to distinguish the mean from the variance components based on the observed data. Therefore, in the following, we only consider the cases where the diagonals of $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are unknown but identical, say, $\mathbf{\Sigma}_1 = \sigma_1^2\mathbf{I}$ and $\mathbf{\Sigma}_2 = \sigma_2^2\mathbf{I}$, respectively.

In this case, based on the previous discussions, especially the definitions of the adjusted estimators $\widehat{T}'_K$, it is important to estimate the variances σ_1^2 and σ_2^2. Toward this end, if in addition we assume that both θ and μ are sparse vectors in themselves, we may take advantage of such sparsity and estimate σ_k^2 by the smaller observations since they are likely to correspond to mean zero Gaussian random variables. Following the ideas in Collier et al. (2017, 2018), we may estimate σ_1 and σ_2 by

$$\hat{\sigma}_1 = 9\sqrt{\frac{1}{\lfloor n/2 \rfloor}\sum_{j\le n/2} x_{(j)}^2}, \qquad \hat{\sigma}_2 = 9\sqrt{\frac{1}{\lfloor n/2 \rfloor}\sum_{j\le n/2} y_{(j)}^2}, \tag{S5.3}$$

where $x_{(1)} \le x_{(2)} \le ... \le x_{(n)}$ and $y_{(1)} \le y_{(2)} \le ... \le y_{(n)}$ are ordered statistics associated to $\mathbf{x}_n$ and $\mathbf{y}_n$. With the above variance estimators, we can estimate the T-score using the above adjusted estimators by plugging-in the variance estimators.

Moreover, when $\beta \in (0, 1/2)$, there is an extra advantage of the simple thresholding estimator $\tilde{T}$. Specifically, if we define $D_\sigma^\infty(s, L_n) = \big\{(\theta, \mu, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2):$

$(\theta, \mu) \in D(s), \max(\|\theta\|_\infty, \|\mu\|_\infty) \le L_n, \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \sigma^2 \mathbf{I}_n\big\}$, then by the proofs of Theorems 1 and 3, it can be shown that

$$\inf_{\hat{T}} \sup_{(\theta,\mu,\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in D_\sigma^\infty(s,L_n)} \mathcal{R}(\hat{T}) \gtrsim \frac{s^2 L_n^2}{n^2} \cdot \min\{\sigma^2 \log n, L_n^2\}, \tag{S5.4}$$

and

$$\sup_{(\theta,\mu,\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in D_\sigma^\infty(s,L_n)} \mathcal{R}(\widetilde{T}) \lesssim \frac{\sigma^2(L_n^2 + \sigma^2 \log n)s^2 \log n}{n^2}. \tag{S5.5}$$

Again, if in addition $L_n \le \sigma\sqrt{2\log n}$, then

$$\sup_{(\theta,\mu,\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)\in D_\sigma^\infty(s,L_n)} \mathcal{R}(\widetilde{T}) \lesssim \frac{s^2 L_n^4}{n^2} + \frac{\sigma^4 \log^2 n}{n^3} + \frac{\sigma^2 L_n^2 \log n}{n^2}. \tag{S5.6}$$

Therefore, whenever $L_n \gtrsim \sigma$, $\widetilde{T}$ is minimax optimal with the optimal rate of convergence being $\frac{L_n^2 s^2}{n^2} \min\{\sigma^2 \log n, L_n^2\}$.

S6 Supplementary Figures and Tables from Numerical Studies

Supplementary simulation results. In our Section 3, to generate dependent observations from a non-identity covariance matrix, we considered the block-wise covariances where $\mathbf{\Sigma}$ is block diagonal where each block is either a 10×10 Toeplitz matrix or a 1000×1000 exchangeable covariance matrix whose off-diagonal elements are 0.5. In particular, the 10×10 Toeplitz matrix is given as follows

$$\begin{bmatrix} 1 & 0.9 & 0.8 & 0.7 & 0.6 & 0.5 & ... & 0.1 \\ 0.9 & 1 & 0.9 & 0.8 & 0.7 & 0.6 & ... & 0.2 \\ 0.8 & 0.9 & 1 & 0.9 & 0.8 & 0.7 & ... & 0.3 \\ ... \\ 0.1 & 0.2 & 0.3 & 0.4 & 0.5 & 0.6 & ... & 1 \end{bmatrix}.$$

Tables S6.1 and S6.2 include the empirical RMSE of the five estimators under covariance structures $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$, respectively.

Real data analysis. In our gene set enrichment analysis, 5,023 biological processes from Gene Ontology (GO) (Botstein et al. 2000) that contain at least 10 genes were tested. Figure S6.1 presents the directed acyclic graphs of the GO biological processes that linked to the most significant GO terms from the simultaneous signal GSEA analysis.

Bibliography

Aldous, D. J. (1985). Exchangeability and related topics. In *École d'Été de Probabilités de Saint-Flour XIII – 1983*, pp. 1–198. Springer.

Botstein, D., J. M. Cherry, M. Ashburner, et al. (2000). Gene ontology: tool for the unification of biology. *Nature Genetics 25*(1), 25–9.

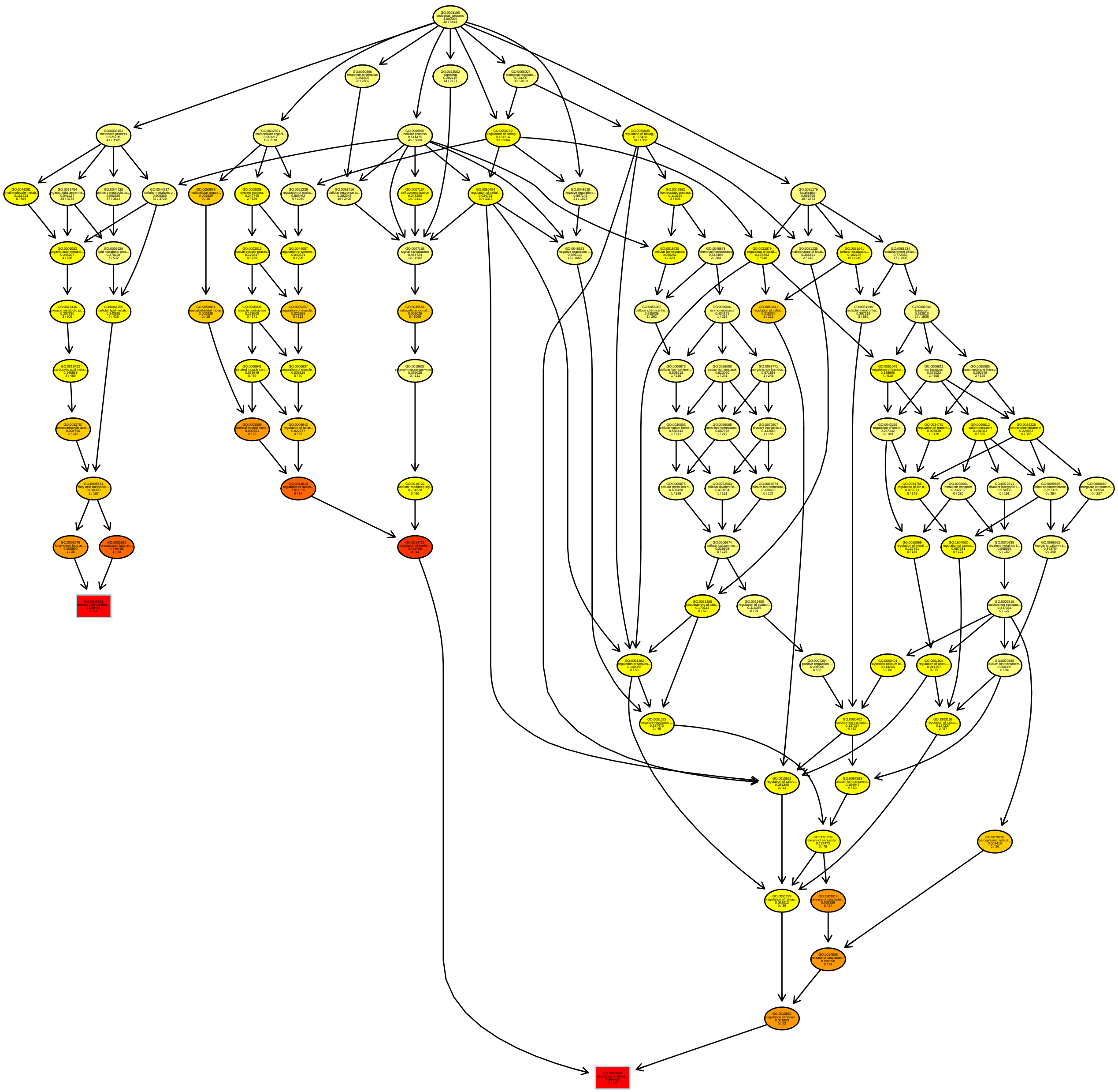

Figure S6.1: Directed acyclic graph of GO biological processes connected by some path to the most significant processes from the GSEA analysis. Yellow: least significant; Red: most significant; Rectangles: top GSEA results.

Table S6.1: Empirical RMSE under covariance $\boldsymbol{\Sigma}_1$.

$\frac{n}{10^4}$	s	$\widehat{T}_K^{S*}$	$\widehat{T}_K^S$	$\widetilde{T}$	$\widehat{T}_K$	$\overline{T}$	$\widehat{T}_K^{S*}$	$\widehat{T}_K^S$	$\widetilde{T}$	$\widehat{T}_K$	$\overline{T}$
		Sparse Pattern I					Sparse Pattern II				
	50	11.58	22.53	25.99	23.25	1909.3	7.89	26.97	36.25	31.72	1908.5
	100	10.16	21.44	28.93	26.60	954.7	7.82	26.55	36.19	31.98	953.9
15	200	10.74	23.10	31.72	28.66	476.0	7.23	25.19	34.72	30.5	477.4
	400	10.11	22.14	30.75	27.67	237.9	9.24	25.16	30.28	26.92	237.7
	800	10.24	21.56	28.53	25.76	118.8	8.44	26.04	34.06	30.19	118.6
	50	11.47	24.65	30.30	27.37	3821.8	10.53	28.01	37.42	32.88	3821.1
	100	10.82	22.60	31.32	27.92	1910.8	8.87	26.51	35.32	30.96	1909.7
30	200	11.01	23.15	28.85	26.01	954.0	8.63	28.06	35.27	30.83	953.5
	400	10.99	22.63	29.89	26.55	476.9	10.52	26.39	31.89	28.19	476.9
	800	10.79	22.71	30.20	26.85	237.9	9.57	26.66	33.2	29.25	238.0
	50	10.94	26.16	35.45	31.23	6364.9	9.19	30.27	39.04	34.17	6366.3
	100	11.62	23.04	28.53	25.68	3182.8	11.36	26.92	31.66	27.97	3184.7
50	200	11.25	22.29	29.01	25.83	1590.6	10.84	27.39	31.60	27.90	1591.0
	400	10.61	23.38	30.48	26.77	795.0	9.66	27.93	34.07	29.93	795.4
	800	11.38	23.25	29.63	26.35	397.4	10.25	27.14	32.64	28.76	397.2

Brown, L. D. and M. G. Low (1996). A constrained risk inequality with applications to nonparametric functional estimation. *Ann. Stat. 24*(6), 2524–2535.

Cai, T. T. and M. G. Low (2011). Testing composite hypotheses, hermite

Table S6.2: Empirical RMSE under covariance $\mathbf{\Sigma}_2$.

$\frac{n}{10^4}$	s	$\widehat{T}_K^{S*}$	$\widehat{T}_K^{S}$	$\widetilde{T}$	$\widehat{T}_K$	$\overline{T}$	$\widehat{T}_K^{S*}$	$\widehat{T}_K^{S}$	$\widetilde{T}$	$\widehat{T}_K$	$\overline{T}$
		Sparse Pattern I					Sparse Pattern II				
	50	10.59	23.56	31.26	28.15	1913.9	7.27	23.89	38.89	34.82	1913.7
	100	9.69	24.12	33.90	30.13	950.3	9.35	26.58	34.95	31.06	949.9
15	200	10.06	23.47	30.95	27.58	477.5	8.15	24.46	33.50	29.69	482.1
	400	10.12	21.66	29.35	26.38	239.5	7.84	25.80	34.89	30.93	238.2
	800	9.64	21.88	30.18	27.14	118.5	8.49	25.51	33.03	29.18	118.8
	50	9.23	25.31	35.58	31.36	3818.3	8.66	25.21	33.44	29.05	3826.3
	100	12.29	23.97	31.05	27.91	1908.5	9.57	26.84	35.01	30.96	1914.1
30	200	11.38	22.50	29.33	26.13	958.0	9.97	26.44	31.81	28.05	953.7
	400	11.63	23.15	29.64	26.57	477.1	8.88	26.18	34.72	30.77	476.8
	800	10.71	22.93	30.40	27.10	237.6	9.37	26.48	33.02	29.07	237.6
	50	8.97	25.34	33.31	29.20	6345.1	9.35	28.84	35.80	31.22	6360.9
	100	12.43	23.16	26.86	24.02	3178.4	8.55	28.53	37.41	32.49	3178.0
50	200	10.97	22.22	28.75	25.66	1586.2	8.90	28.15	35.82	31.19	1589.8
	400	11.07	22.40	28.48	25.34	793.8	9.91	26.84	32.92	28.94	797.0
	800	10.80	23.25	30.33	26.93	397.7	10.26	27.14	33.88	29.96	398.2

polynomials and optimal estimation of a nonsmooth functional. *Ann. Stat. 39*(2), 1012–1041.

Collier, O., L. Comminges, and A. B. Tsybakov (2017). Minimax estimation of linear and quadratic functionals on sparsity classes. *The Annals of*

Statistics 45(3), 923–958.

Collier, O., L. Comminges, and A. B. Tsybakov (2020). On estimation of nonsmooth functionals of sparse normal means. *Bernoulli 26*(3), 1989–2020.

Collier, O., L. Comminges, A. B. Tsybakov, and N. Verzelen (2018). Optimal adaptive estimation of linear functionals under sparsity. *The Annals of Statistics 46*(6A), 3130–3150.