OPTIMAL ESTIMATION OF A QUADRATIC FUNCTIONAL UNDER THE GAUSSIAN TWO-SEQUENCE MODEL

T. Tony Cai and Xin Lu Tan

University of Pennsylvania

Supplementary Material

This supplement contains two parts: supplement S1 presents the estimation results for $Q(\mu, \theta)$ when μ and θ have different signal strengths, whereas supplement S2 presents the proofs for Theorems 3 and 4 given in the main text.

S1 Estimation of $Q(\mu, \theta)$ with Different Signal Strengths

We consider in Section 2 the estimation of $Q(\mu, \theta) = \frac{1}{n}\sum_{i=1}^{n} \mu_i^2 \theta_i^2$ over the parameter space (2.7) where $j_n = k_n = n^\beta$ and $r_n = s_n = n^b$, with $0 < \epsilon \leq \beta < \frac{1}{2}$ and $b \in \mathbb{R}$. In this section, we present the estimation result for $Q(\mu, \theta)$ with $j_n = k_n = n^\beta$ but allow r_n and s_n to differ. Specifically, we consider the following parameter space

$$\Omega(\beta, \epsilon, a, b) = \{(\mu, \theta) \in \mathbb{R}^n \times \mathbb{R}^n : \|\mu\|_0 \leq k_n, \|\mu\|_\infty \leq r_n, \|\theta\|_0 \leq k_n, \|\theta\|_\infty \leq s_n,$$
$$\|\mu \star \theta\|_0 \leq q_n\}, \tag{S1.1}$$

where $k_n = n^\beta, q_n = n^\epsilon$ with $0 < \epsilon \leq \beta < \frac{1}{2}$, and $r_n = n^a, s_n = n^b$ with $a, b \in \mathbb{R}$.

Similar as before, the estimation problem can be divided into three regimes: the sparse regime ($0 < \epsilon < \frac{\beta}{2}$), the moderately dense regime ($\frac{\beta}{2} \leq \epsilon \leq \frac{3\beta}{4}$), and the strongly dense regime ($\frac{3\beta}{4} < \epsilon \leq \beta$). When μ and θ have different signal strengths, the minimax rates of convergence for $Q(\mu,\theta)$ exhibit more elaborate phase transitions, though they still bear the familiar form

$$R^*(n,\Omega(\beta,\epsilon,a,b)) := \inf_{\widehat{Q}} \sup_{(\mu,\theta)\in\Omega(\beta,\epsilon,a,b)} E_{(\mu,\theta)}(\widehat{Q}-Q(\mu,\theta))^2 \asymp \gamma_n(\beta,\epsilon,a,b),$$

where $\gamma_n(\beta,\epsilon,a,b)$ is a function of n indexed by β,ϵ,a, and b. For readability, we summarize the corresponding $\gamma_n(\beta,\epsilon,a,b)$ in Table 1 (sparse regime), Table 2 (moderately dense regime), and Table 3 (strongly dense regime), respectively. The minimax rates of convergence are attained by the same estimators as before over the respective regimes, as stated in Theorem 5 and Theorem 6 given below.

Although we do not present the result here due to its lengthiness, estimation of $Q(\mu,\theta)$ for the case where no equality constraint is imposed on either sparsity or signal strength of μ and θ can be analyzed analogously provided that the magnitude of the simultaneous sparsity ϵ is compared to α if $a \geq b$, and to β if $b \geq a$, for the characterization of the sparse and dense regimes.

Theorem 5 (Sparse Regime)**.** *Let* $0<\epsilon<\frac{\beta}{2}$ *and* $0<\beta<\frac{1}{2}$*. Then* $\widehat{Q}_2$ *defined in (2.12) with* $\tau_n=\log n$ *attains the minimax rate of convergence over* $\Omega(\beta,\epsilon,a,b)$ *for* $(a,b)\in\{(a,b): a\wedge b>0\}$*. On the other hand,* $\widehat{Q}_0=0$ *attains the minimax rate of convergence over* $\Omega(\beta,\epsilon,a,b)$ *for* $(a,b)\in\{(a,b): a\wedge b\leq 0\}$*.*

Theorem 6 (Dense Regime)**.** *Let* $\frac{\beta}{2}\leq\epsilon\leq\beta$ *and* $0<\beta<\frac{1}{2}$*. Then* $\widehat{Q}_4$ *defined in (2.18) with* $\tau_n=4\log n$ *attains the minimax rate of convergence over* $\Omega(\beta,\epsilon,a,b)$ *for* $(a,b)\in\{(a,b): a\vee b>0$ *and* $a\wedge b>\frac{\beta-2\epsilon}{4}\}$*. On the other hand,* $\widehat{Q}_0=0$ *attains the minimax rate of convergence over* $\Omega(\beta,\epsilon,a,b)$ *for* $(a,b)\in\{(a,b): a\vee b\leq 0$ *or* $a\wedge b\leq\frac{\beta-2\epsilon}{4}\}$*.*

The shaded regions in the three tables represent the region where $\widehat{Q}_0$ attains the minimax rate of convergence. Thus, $\{(a,b): a\wedge b\leq 0\}$ is shaded in Table 1, while $\{(a,b): a\vee b\leq$

0 or $a \wedge b \leq \frac{\beta - 2\epsilon}{4}\}$ is shaded in Tables 2 and 3.

Note that the estimation result for the dense regime turns out to be interesting (and more inspiring) when r_n and s_n can differ. It seems that estimation is desirable whenever the signal strengths of both sequences barely exceed some small threshold ($a \wedge b > \frac{\beta - 2\epsilon}{4}$, but $\beta - 2\epsilon \leq 0$ in this case) and at least one sequence has sufficiently strong signal ($a \vee b > 0$). This is in contrast to the sparse regime where estimation is desirable only when the signal strength of both sequences are sufficiently strong ($a \wedge b > 0$). The intuitive explanation is that in the dense regime, knowing that $\mu_i \neq 0$ (because of large X_i^2) most often suggests that $\theta_i \neq 0$ too (even if Y_i^2 is small), and vice versa, so we cannot afford to estimate $\mu_i^2 \theta_i^2$ by 0 with this additional information. On the contrary, in the sparse regime, knowing that $\mu_i \neq 0$ does not entail much about whether $\theta_i \neq 0$ due to the sparseness of simultaneous nonzero coordinates. Therefore it is better to estimate $\mu_i^2 \theta_i^2$ by 0 unless both X_i^2 and Y_i^2 are large.

In fact, the minimax rates of convergence for the sparse regime are relatively simple to describe, when r_n is not necessarily equal to s_n:

$$\gamma_n(\beta, \epsilon, a, b) = \begin{cases} n^{2\epsilon + 4a + 4b - 2} & \text{if } a \wedge b \leq 0, \\ n^{2\epsilon + 4a \vee b - 2} (\log n)^2 & \text{if } 0 < a \wedge b \leq \frac{\epsilon}{2}, \\ n^{\epsilon + 4a \vee b + 2a \wedge b - 2} & \text{if } a \wedge b > \frac{\epsilon}{2}. \end{cases}$$

Unfortunately, we do not have such an easy representation for the minimax rates of convergence in the dense regime. Nonetheless, due to the two-dimensional nature of the estimation problem, we find tables useful not only in presenting the minimax rates of convergence but also in illustrating the regions with weak signals (i.e., the shaded regions).

	$b \le 0$	$0 < b \le \frac{\epsilon}{2}$	$b > \frac{\epsilon}{2}$
$a \le 0$	$n^{2\epsilon+4a+4b-2}$	$n^{2\epsilon+4a+4b-2}$	$n^{2\epsilon+4a+4b-2}$
$0 < a \le \frac{\epsilon}{2}$	$n^{2\epsilon+4a+4b-2}$	$n^{2\epsilon+4a\vee b-2}(\log n)^2$	$n^{2\epsilon+4b-2}(\log n)^2$
$a > \frac{\epsilon}{2}$	$n^{2\epsilon+4a+4b-2}$	$n^{2\epsilon+4a-2}(\log n)^2$	$n^{\epsilon+4a\vee b+2a\wedge b-2}$

Table 1: Minimax rates of convergence in the sparse regime: $0 < \epsilon < \frac{\beta}{2}$.

	$b \le \frac{\beta-2\epsilon}{4}$	$\frac{\beta-2\epsilon}{4} < b \le 0$	$0 < b \le \frac{2\epsilon-\beta}{4}$	$\frac{2\epsilon-\beta}{4} < b \le \frac{\beta-\epsilon}{2}$	$b > \frac{\beta-\epsilon}{2}$
$a \le \frac{\beta-2\epsilon}{4}$	$n^{2\epsilon+4a+4b-2}$	$n^{2\epsilon+4a+4b-2}$	$n^{2\epsilon+4a+4b-2}$	$n^{2\epsilon+4a+4b-2}$	$n^{2\epsilon+4a+4b-2}$
$\frac{\beta-2\epsilon}{4} < a \le 0$	$n^{2\epsilon+4a+4b-2}$	$n^{2\epsilon+4a+4b-2}$	$\max\{n^{\beta+4b-2}, n^{2\epsilon+4a-2}(\log n)^2\}$	$n^{\beta+4b-2}$	$n^{\beta+4b-2}$
$0 < a \le \frac{2\epsilon-\beta}{4}$	$n^{2\epsilon+4a+4b-2}$	$\max\{n^{\beta+4a-2}, n^{2\epsilon+4b-2}(\log n)^2\}$	$n^{2\epsilon-2}(\log n)^4$	$n^{\beta+4b-2}$	$n^{\beta+4b-2}$
$\frac{2\epsilon-\beta}{4} < a \le \frac{\beta-\epsilon}{2}$	$n^{2\epsilon+4a+4b-2}$	$n^{\beta+4a-2}$	$n^{\beta+4a-2}$	$n^{\beta+4a\vee b-2}$	$n^{\beta+4b-2}$
$a > \frac{\beta-\epsilon}{2}$	$n^{2\epsilon+4a+4b-2}$	$n^{\beta+4a-2}$	$n^{\beta+4a-2}$	$n^{\beta+4a-2}$	$n^{\epsilon+4a\vee b+2a\wedge b-2}$

Table 2: Minimax rates of convergence in the moderately dense regime: $\frac{\beta}{2} \le \epsilon \le \frac{3\beta}{4}$. In this case, we have $\frac{2\epsilon-\beta}{4} \le \frac{\beta-\epsilon}{2}$.

	$b \le \frac{\beta-2\epsilon}{4}$	$\frac{\beta-2\epsilon}{4} < b \le 0$	$0 < b \le \frac{\beta-\epsilon}{2}$	$\frac{\beta-\epsilon}{2} < b \le \frac{2\epsilon-\beta}{4}$	$b > \frac{2\epsilon-\beta}{4}$
$a \le \frac{\beta-2\epsilon}{4}$	$n^{2\epsilon+4a+4b-2}$	$n^{2\epsilon+4a+4b-2}$	$n^{2\epsilon+4a+4b-2}$	$n^{2\epsilon+4a+4b-2}$	$n^{2\epsilon+4a+4b-2}$
$\frac{\beta-2\epsilon}{4} < a \le 0$	$n^{2\epsilon+4a+4b-2}$	$n^{2\epsilon+4a+4b-2}$	$\max\{n^{\beta+4b-2}, n^{2\epsilon+4a-2}(\log n)^2\}$	$\max\{n^{\beta+4b-2}, n^{2\epsilon+4a-2}(\log n)^2\}$	$n^{\beta+4b-2}$
$0 < a \le \frac{\beta-\epsilon}{2}$	$n^{2\epsilon+4a+4b-2}$	$\max\{n^{\beta+4a-2}, n^{2\epsilon+4b-2}(\log n)^2\}$	$n^{2\epsilon-2}(\log n)^4$	$n^{2\epsilon-2}(\log n)^4$	$n^{\beta+4b-2}$
$\frac{\beta-\epsilon}{2} < a \le \frac{2\epsilon-\beta}{4}$	$n^{2\epsilon+4a+4b-2}$	$\max\{n^{\beta+4a-2}, n^{2\epsilon+4b-2}(\log n)^2\}$	$n^{2\epsilon-2}(\log n)^4$	$\max\{n^{2\epsilon-2}(\log n)^4, n^{\epsilon+4a\vee b+2a\wedge b-2}\}$	$n^{\epsilon+2a+4b-2}$
$a > \frac{2\epsilon-\beta}{4}$	$n^{2\epsilon+4a+4b-2}$	$n^{\beta+4a-2}$	$n^{\beta+4a-2}$	$n^{\epsilon+4a+2b-2}$	$n^{\epsilon+4a\vee b+2a\wedge b-2}$

Table 3: Minimax rates of convergence in the strongly dense regime: $\frac{3\beta}{4} < \epsilon \le \beta$. In this case, we have $\frac{\beta-\epsilon}{2} < \frac{2\epsilon-\beta}{4}$.

S2 Additional Proofs

In this section, we present the proofs of Theorems 3 and 4. Hereinafter, we omit the subscripts n in k_n, q_n, s_n and τ_n that signifies their dependence on the sample size. We denote by ψ_μ the density of a Gaussian distribution with mean μ and variance σ^2, and we denote by $\ell(n,k)$ the class of all subsets of $\{1,\ldots,n\}$ of k distinct elements. For a standard normal random variable Z, the expressions $\phi(z), \Phi(z) = P(Z \leq z)$, and $\tilde{\Phi}(z) = 1 - \Phi(z)$ represent its density, cumulative distribution function, and survival function, respectively. We let c and C be constants whose values may vary for each occurrence.

S2.1 Proof of Theorem 3

The proof is based on Lemmas 1 and 2 which bound, respectively, the bias and variance of one term in the estimator $\widehat{Q}_4$ (given in (2.18)). For clarity, we defer the proofs of Lemma 1 and Lemma 2 to Section S2.3.

Lemma 1. *Let $X \sim N(\mu, \sigma^2)$ and $Y \sim N(\theta, \sigma^2)$ be independent. Set $\eta = E[(Z_1^2 - \sigma^2)(Z_2^2 - \sigma^2)\mathbb{1}(Z_1^2 \vee Z_2^2 > \sigma^2\tau)]$, where $Z_1, Z_2 \overset{i.i.d.}{\sim} N(0, \sigma^2)$. Then*

$$\eta = -4\sigma^4\tau\phi^2(\tau^{1/2}),$$

and for $\tau \geq 1$,

$$\begin{aligned}
&\left|E[(X^2-\sigma^2)(Y^2-\sigma^2)\mathbb{1}(X^2 \vee Y^2 > \sigma^2\tau)] - \eta - \mu^2\theta^2\right| \\
&\leq \min\{\mu^2, 3\sigma^2\tau\}\min\{\theta^2, 3\sigma^2\tau\} + 2\sigma^2\tau^{1/2}\phi(\tau^{1/2})\min\{\mu^2, 3\sigma^2\tau\} \\
&\quad + 2\sigma^2\tau^{1/2}\phi(\tau^{1/2})\min\{\theta^2, 3\sigma^2\tau\}.
\end{aligned}$$

Lemma 2. *Let $X \sim N(\mu, \sigma^2)$ and $Y \sim N(\theta, \sigma^2)$ be independent. Then for $\tau \geq 1$,*

$$\begin{aligned}
&\mathrm{Var}[(X^2-\sigma^2)(Y^2-\sigma^2)\mathbb{1}(X^2 \vee Y^2 > \sigma^2\tau)] \\
&\leq \begin{cases} 2d^{1/2}\tilde{\Phi}(\tau^{1/2})^{1/2} & \text{if } \mu=\theta=0, \\ 4\sigma^2\mu^4\theta^2 + 4\sigma^2\mu^2\theta^4 + 16\sigma^4\mu^2\theta^2 + 2\sigma^4\mu^4 + 2\sigma^4\theta^4 & \\ \quad +8\sigma^6\mu^2 + 8\sigma^6\theta^2 + 4\sigma^8 + 8\sigma^4\mu^2\theta^2\tau^2 & \text{otherwise,} \end{cases}
\end{aligned}$$

where $d = E[(Z_1^2-\sigma^2)^4(Z_2^2-\sigma^2)^4]$ and $Z_1, Z_2 \overset{i.i.d.}{\sim} N(0,\sigma^2)$.

Proof of Theorem 3. We first compute the bias of $\widehat{Q}_4$. It follows from Lemma 1 that for all $(\mu,\theta) \in \Omega(\beta,\epsilon,b)$ and $\tau \geq 1$, we have

$$\begin{aligned}
&\left|E_{(\mu,\theta)}(\widehat{Q}_4) - Q(\mu,\theta)\right| \\
&\leq \frac{1}{n}\sum_{i=1}^{n}\left|E_{(\mu_i,\theta_i)}[(X_i^2-\sigma^2)(Y_i^2-\sigma^2)\mathbb{1}(X_i^2 \vee Y_i^2 > \sigma^2\tau)] - \eta - \mu_i^2\theta_i^2\right| \\
&\leq \frac{1}{n}\sum_{i=1}^{n}\Big[\min\{\mu_i^2, 3\sigma^2\tau\}\min\{\theta_i^2, 3\sigma^2\tau\} + 2\sigma^2\tau^{1/2}\phi(\tau^{1/2})\min\{\mu_i^2, 3\sigma^2\tau\} \\
&\qquad + 2\sigma^2\tau^{1/2}\phi(\tau^{1/2})\min\{\theta_i^2, 3\sigma^2\tau\}\Big] \\
&\leq \frac{1}{n}\Big[\min\{qs^4, 3\sigma^2 qs^2\tau, 9\sigma^4 q\tau^2\} + 4\sigma^2\tau^{1/2}\phi(\tau^{1/2})\min\{ks^2, 3\sigma^2 k\tau\}\Big],
\end{aligned}$$

the last inequality follows from the fact that for $(\mu,\theta) \in \Omega(\beta,\epsilon,b)$, there are at most k nonzero entries for either μ or θ, and there are at most q entries that are simultaneously nonzero for both μ and θ.

On the other hand, by Lemma 2, for all $(\mu,\theta) \in \Omega(\beta,\epsilon,b)$ and $\tau \geq 1$, the variance of $\widehat{Q}_4$

satisfies

$$
\begin{aligned}
&\mathrm{Var}_{(\mu,\theta)}(\widehat{Q}_4)\\
&= \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}_{(\mu_i,\theta_i)}[(X_i^2-\sigma^2)(Y_i^2-\sigma^2)\mathbb{1}(X_i^2 \vee Y_i^2 > \sigma^2\tau)]\\
&\le \frac{1}{n^2}\Big[\sum_{i:\mu_i=\theta_i=0} 2d^{1/2}\tilde{\Phi}(\tau^{1/2})^{1/2}\\
&\qquad + \sum_{i:\mu_i\neq 0 \text{ or } \theta_i\neq 0}\big(4\sigma^2\mu_i^4\theta_i^2 + 4\sigma^2\mu_i^2\theta_i^4 + 16\sigma^4\mu_i^2\theta_i^2 + 2\sigma^4\mu_i^4 + 2\sigma^4\theta_i^4\\
&\qquad\qquad + 8\sigma^6\mu_i^2 + 8\sigma^6\theta_i^2 + 4\sigma^8 + 8\sigma^4\mu_i^2\theta_i^2\tau^2\big)\Big]\\
&\le \frac{1}{n^2}\Big[2d^{1/2}n\tilde{\Phi}(\tau^{1/2})^{1/2} + 8\sigma^2 qs^6 + 16\sigma^4 qs^4 + 4\sigma^4 ks^4 + 16\sigma^6 ks^2 + 8\sigma^8 k + 8\sigma^4 qs^4\tau^2\Big]\\
&\le \frac{C}{n^2}\max\{n\tilde{\Phi}(\tau^{1/2})^{1/2}, qs^4, qs^6, k, ks^2, ks^4, qs^4\tau^2\}.
\end{aligned}
$$

Again, the second to the last inequality follows from the fact that for $(\mu,\theta)\in\Omega(\beta,\epsilon,b)$, there are at most k nonzero entries for either μ or θ, and there are at most q entries that are simultaneously nonzero for both μ and θ.

Combining the bias and variance term, we have

$$
\begin{aligned}
&\sup_{(\mu,\theta)\in\Omega(\beta,\epsilon,b)} E_{(\mu,\theta)}(\widehat{Q}_4 - Q(\mu,\theta))^2\\
&\le \frac{C}{n^2}\Big[\min\{q^2s^8, q^2s^4\tau^2, q^2\tau^4\} + \tau\phi^2(\tau^{1/2})\min\{k^2s^4, k^2\tau^2\}\\
&\qquad + \max\{n\tilde{\Phi}(\tau^{1/2})^{1/2}, qs^4, qs^6, k, ks^2, ks^4, qs^4\tau^2\}\Big]\\
&= \frac{C}{n^2}\Big[\min\{n^{2\epsilon+8b}, n^{2\epsilon+4b}\tau^2, n^{2\epsilon}\tau^4\} + \tau\phi^2(\tau^{1/2})\min\{n^{2\beta+4b}, n^{2\beta}\tau^2\}\\
&\qquad + \max\{n\tilde{\Phi}(\tau^{1/2})^{1/2}, n^{\epsilon+4b}, n^{\epsilon+6b}, n^{\beta}, n^{\beta+2b}, n^{\beta+4b}, n^{\epsilon+4b}\tau^2\}\Big].
\end{aligned}
$$

Let $\tau = 4\log n$, then we have $\tilde{\Phi}(\tau^{1/2}) \le C\phi(\tau^{1/2}) = O(n^{-2})$ for some constant C. It follows that for $b > 0$,

$$\sup_{(\mu,\theta)\in\Omega(\beta,\epsilon,b)} E_{(\mu,\theta)}(\widehat{Q}_4 - Q(\mu,\theta))^2 \leq C \max\left\{n^{2\epsilon-2}(\log n)^4, n^{\epsilon+6b-2}, n^{\beta+4b-2}\right\}.$$

$\square$

S2.2 Proof of Theorem 4

In this section, we prove Theorem 4, which constitutes the lower bound for the estimation rate of $Q(\mu,\theta)$ in the dense regime. We begin with some technical tools for establishing lower bounds.

General Tools

Let $\mathcal{P}$ be a set of probability measures on a measurable space $(\mathcal{X},\mathcal{A})$, and let $\theta : \mathcal{P} \longrightarrow \mathbb{R}$. For $P_f, P_g \in \mathcal{P}$, let $\theta_f = \theta(P_f), \theta_g = \theta(P_g)$, and let f, g denote the density of P_f, P_g with respect to some dominating measure u. The chi-square affinity between P_f and P_g is defined as

$$\xi = \xi(P_f, P_g) = \int \frac{g^2}{f}\, du.$$

In particular, for Gaussian distributions, we have

$$\xi(N(\theta_0,\sigma^2), N(\theta_1,\sigma^2)) = e^{(\theta_1-\theta_0)^2/\sigma^2}.$$

Throughout, the proof of lower bounds is established by the construction of two priors which have small chi-square distance but a large difference in the expected values of the resulting quadratic functionals, followed by an application of the Constrained Risk Inequality (CRI) in Brown and Low (1996). Essentially, CRI says that if P_f and P_g are such that $\theta_f, \theta_g \in \Theta$, the parameter space of estimation, with $\xi = \xi(P_f, P_g) < \infty$, then for any estimator δ of $\theta = \theta(P) \in \Theta$ based on the random variable X with distribution P, we have

$$\sup_{\theta\in\Theta} E_\theta(\delta(X) - \theta)^2 \geq \frac{(\theta_g - \theta_f)^2}{(1 + \xi^{1/2})^2}.$$

It follows that to establish lower bound for estimation rate, it suffices to find P_f and P_g such that $(\theta_g - \theta_f)^2$ is as large as possible subject to $\xi(P_f, P_g) < \infty$.

Proof of Theorem 4

To prove Theorem 4, it is sufficient to show that for $0 < \beta < \frac{1}{2}$,

$$\gamma_n(\beta, \epsilon, b) \geq \begin{cases} n^{2\epsilon+8b-2} & \text{if } b \leq 0, \quad \text{for } 0 < \epsilon \leq \beta, & \text{(Case 2)} \\ n^{\epsilon+6b-2} & \text{if } b > 0, \quad \text{for } 0 < \epsilon \leq \beta, & \text{(Case 3)} \\ n^{\beta+4b-2} & \text{if } b > 0, \quad \text{for } \frac{\beta}{2} \leq \epsilon \leq \beta, & \text{(Case 4)} \\ n^{2\epsilon-2}(\log n)^4 & \text{if } b > 0, \quad \text{for } 0 < \epsilon \leq \beta. & \text{(Case 5)} \end{cases}$$

The proof of Case 2 and Case 3 can be found in Section 5.2 in the main text, hence we will only provide proofs of Case 4 and Case 5 below. For individual regions in $\{(\beta, \epsilon, b) : \frac{\beta}{2} \leq \epsilon \leq \beta < \frac{1}{2}, b \in \mathbb{R}\}$, the minimax rate of convergence is obtained as the sharpest rate among all cases in which the region belongs to. For instance, the region $\{(\beta, \epsilon, b) : \frac{3\beta}{4} < \epsilon \leq \beta < \frac{1}{2}, b > \frac{\epsilon}{6}\}$ is included in Case 3, Case 4 and Case 5, hence $\gamma_n(\beta, \epsilon, b) \geq \max\{n^{\epsilon+6b-2}, n^{\beta+4b-2}, n^{2\epsilon-2}(\log n)^4\} = n^{\epsilon+6b-2}$.

Proof of Case 4. The proof of Case 4 is very similar to the proof of Case 1, besides that a slightly different mixture prior g is employed. Let

$$f(x_1,\ldots,x_n,y_1,\ldots,y_n) = \prod_{i=1}^{k} \psi_s(x_i) \prod_{i=k+1}^{n} \psi_0(x_i) \prod_{i=1}^{n} \psi_0(y_i).$$

For $I \in \ell(k,q)$, let

$$\begin{aligned} &g_I(x_1,\ldots,x_n,y_1,\ldots,y_n) \\ &= \prod_{i=1}^{k} \psi_s(x_i) \prod_{i=k+1}^{n} \psi_0(x_i) \prod_{i=1}^{k} \left[\frac{1}{2}\psi_{\theta_i}(y_i) + \frac{1}{2}\psi_{-\theta_i}(y_i)\right] \prod_{i=k+1}^{n} \psi_0(y_i), \end{aligned}$$

where $\theta_i = \rho \mathbb{1}(i \in I)$ with $\rho > 0$, and let

$$g = \frac{1}{\binom{k}{q}} \sum_{I \in \ell(k,q)} g_I.$$

Note that in constructing g, mixing is done not only over all possible subsets $\ell(k,q)$ but also over the signs of θ_i's. This has largely to do with the intuition that when signal is abundant, uncertainty about the signs of θ_i's further increase the difficulty of the estimation problem. That being said, mixing without sign flips (i.e., simply use the priors f and g as given in the proof of Case 1) does not give us the tightest lower bound. Similar to Case 1, keeping $\mu = (s,\ldots,s,0,\ldots,0)$ the same in both f and g essentially reduces the two-sequence problem to a one-sequence problem. Our choice of priors is equivalent to having only one Gaussian mean sequence of length k with q nonzero entries — thus the correspondence between the dense regime in the two-sequence case ($q \gg \sqrt{k}$) and the dense regime in the one-sequence case ($k \gg \sqrt{n}$).

Again, the chi-square affinity between f and g has the form (5.2), where for $I, J \in \ell(k,q)$

with $m = \mathrm{Card}(I \cap J)$,

$$\begin{aligned}
\int \frac{g_I g_J}{f} &= \prod_{i=1}^{k} \int \frac{[\frac{1}{2}\psi_{\rho \mathbb{1}(i \in I)}(y_i) + \frac{1}{2}\psi_{-\rho \mathbb{1}(i \in I)}(y_i)][\frac{1}{2}\psi_{\rho \mathbb{1}(i \in J)}(y_i) + \frac{1}{2}\psi_{-\rho \mathbb{1}(i \in J)}(y_i)]}{\psi_0(y_i)}\, dy_i \\
&= \prod_{i=1}^{k} \int \frac{1}{4}\Bigg\{ \frac{\psi_{\rho \mathbb{1}(i \in I)}(y_i)\psi_{\rho \mathbb{1}(i \in J)}(y_i)}{\psi_0(y_i)} + \frac{\psi_{-\rho \mathbb{1}(i \in I)}(y_i)\psi_{-\rho \mathbb{1}(i \in J)}(y_i)}{\psi_0(y_i)} \\
&\qquad + \frac{\psi_{\rho \mathbb{1}(i \in I)}(y_i)\psi_{-\rho \mathbb{1}(i \in J)}(y_i)}{\psi_0(y_i)} + \frac{\psi_{-\rho \mathbb{1}(i \in I)}(y_i)\psi_{\rho \mathbb{1}(i \in J)}(y_i)}{\psi_0(y_i)} \Bigg\}\, dy_i \\
&= \prod_{i \in I \cap J} \frac{1}{4}\left[\int \frac{\psi_\rho^2(y_i)}{\psi_0(y_i)} + \int \frac{\psi_{-\rho}^2(y_i)}{\psi_0(y_i)} + 2\int \frac{\psi_\rho(y_i)\psi_{-\rho}(y_i)}{\psi_0(y_i)} \right] \prod_{i \in I^c \cup J^c} 1 \\
&= \prod_{i \in I \cap J} \frac{1}{2}\left[\exp(\rho^2/\sigma^2) + \exp(-\rho^2/\sigma^2) \right] \\
&= \cosh(\rho^2/\sigma^2)^m.
\end{aligned}$$

It follows that

$$\int \frac{g^2}{f} = E[\cosh(\rho^2/\sigma^2)^M],$$

where M follows hypergeometric distribution as in (6.4). Since M coincides in distribution with the conditional expectation $E(\tilde{M}|\mathcal{B})$ where $\tilde{M}$ is a Binomial$(q,\, \frac{q}{k})$ random variable and $\mathcal{B}$ is a suitable σ-algebra (Aldous, 1985), with Jensen's inequality, we get

$$\int \frac{g^2}{f} \leq E[\cosh(\rho^2/\sigma^2)^{\tilde{M}}] = \left(1 + \frac{q}{k}[\cosh(\rho^2/\sigma^2) - 1]\right)^q.$$

Since $\cosh(x) = \frac{1}{2}(e^x + e^{-x}) = 1 + \frac{x^2}{2} + o(x^2)$ when $x \approx 0$, taking $x = \rho^2/\sigma^2$ with $\rho = (\frac{k}{q^2})^{1/4}$ yields

$$\int \frac{g^2}{f} \leq \left(1 + \frac{1}{2\sigma^4 q}\right)^q < \infty.$$

Since $Q(\mu,\theta)=0$ under f and $Q(\mu,\theta)=\frac{1}{n}qs^2\rho^2$ under g, it follows from CRI that

$$R^*(n,\Omega(\beta,\epsilon,b)) \geq c\left(\frac{1}{n}qs^2\rho^2\right)^2 = cn^{\beta+4b-2}.$$

□

Proof of Case 5. Let f and g be as given in the proof of Case 2 in Section 6.2, and take $\rho=\sigma\sqrt{\frac{1}{2}(1-2\epsilon)\log n}$ in (6.5). It follows that when n is sufficiently large,

$$e^{2\rho^2/\sigma^2}=n^{1-2\epsilon}=\frac{n}{q^2},$$

hence

$$\int\frac{g^2}{f}\leq\left(1+\frac{1}{q}\right)^q\leq e.$$

Since $Q(\mu,\theta)=0$ under f, and $Q(\mu,\theta)=\frac{1}{n}q\rho^4$ under g, it follows from CRI that

$$R^*(n,\Omega(\beta,\epsilon,b)) \geq c\left(\frac{1}{n}q\rho^4\right)^2 = cn^{2\epsilon-2}(\log n)^4.$$

□

S2.3 Proofs of Supporting Lemmas

In this section, we provide the proofs of technical lemmas that are used to establish Theorem 3 in Section 2.1.

Proof of Lemma 1

The proof of Lemma 1 is built on Lemma 3 and Lemma 4.

Lemma 3. *Let $Y \sim N(\theta, \sigma^2)$. Then for $\tau \geq 1$,*

$$E[(Y^2 - \sigma^2)\mathbb{1}(Y^2 \leq \sigma^2\tau)] = \theta^2\Big[\tilde{\Phi}(-\tau^{1/2} - \frac{\theta}{\sigma}\Big) - \tilde{\Phi}\Big(\tau^{1/2} - \frac{\theta}{\sigma}\Big)\Big]$$
$$+ \phi\Big(\tau^{1/2} + \frac{\theta}{\sigma}\Big)[-\sigma^2\tau^{1/2} + \sigma\theta] + \phi\Big(\tau^{1/2} - \frac{\theta}{\sigma}\Big)[-\sigma^2\tau^{1/2} - \sigma\theta].$$

In particular, when $\theta = 0$,

$$E[(Y^2 - \sigma^2)\mathbb{1}(Y^2 \leq \sigma^2\tau)] = -2\sigma^2\tau^{1/2}\phi(\tau^{1/2}).$$

Proof. Let $\lambda = \tau^{1/2}$. We have

$$\begin{aligned} E[Y^2\mathbb{1}(Y^2 \leq \sigma^2\tau)] &= \int_{-\sigma\lambda}^{\sigma\lambda} y^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\theta)^2/2\sigma^2}\ dy \\ &= \int_{-\lambda-\theta/\sigma}^{\lambda-\theta/\sigma} (\theta + \sigma z)^2 \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\ dz \\ &= \theta^2 \int_{-\lambda-\theta/\sigma}^{\lambda-\theta/\sigma} \phi(z)\ dz + 2\sigma\theta \int_{-\lambda-\theta/\sigma}^{\lambda-\theta/\sigma} z\phi(z)\ dz + \sigma^2 \int_{-\lambda-\theta/\sigma}^{\lambda-\theta/\sigma} z^2\phi(z)\ dz. \end{aligned}$$

Using the fact that

$$\int_a^\infty \phi(z)\ dz = \tilde{\Phi}(a), \quad \int_a^\infty z\phi(z)\ dz = \phi(a), \quad \int_a^\infty z^2\phi(z)\ dz = a\phi(a) + \tilde{\Phi}(a),$$

we have

$$
\begin{aligned}
&E[Y^2 \mathbb{1}(Y^2 \le \sigma^2\tau)] \\
&= \theta^2[\tilde{\Phi}(-\lambda - \theta/\sigma) - \tilde{\Phi}(\lambda - \theta/\sigma)] + 2\sigma\theta[\phi(-\lambda - \theta/\sigma) - \phi(\lambda - \theta/\sigma)] \\
&\qquad + \sigma^2[(-\lambda - \theta/\sigma)\phi(-\lambda - \theta/\sigma) + \tilde{\Phi}(-\lambda - \theta/\sigma) - (\lambda - \theta/\sigma)\phi(\lambda - \theta/\sigma) - \tilde{\Phi}(\lambda - \theta/\sigma)] \\
&= (\theta^2 + \sigma^2)[\tilde{\Phi}(-\lambda - \theta/\sigma) - \tilde{\Phi}(\lambda - \theta/\sigma)] + \phi(\lambda + \theta/\sigma)[-\sigma^2\lambda + \sigma\theta] + \phi(\lambda - \theta/\sigma)[-\sigma^2\lambda - \sigma\theta],
\end{aligned}
$$

the last equality due to $\phi(-\lambda - \theta/\sigma) = \phi(\lambda + \theta/\sigma)$. The proof is complete since $\sigma^2 E[\mathbb{1}(Y^2 < \sigma^2\tau)] = \sigma^2[\tilde{\Phi}(-\lambda - \theta/\sigma) - \tilde{\Phi}(\lambda - \theta/\sigma)]$. □

Lemma 4. *Let $Y \sim N(\theta, \sigma^2)$ and set $\theta_0 = E[(Z^2 - \sigma^2)\mathbb{1}(Z^2 \le \sigma^2\tau)]$, where $Z \sim N(0, \sigma^2)$. Then for $\tau \ge 1$,*

$$\left|E[(Y^2 - \sigma^2)\mathbb{1}(Y^2 \le \sigma^2\tau)] - \theta_0\right| \le \min\{\theta^2, 3\sigma^2\tau\}.$$

Proof. Let $B(\theta) = E[(Y^2 - \sigma^2)\mathbb{1}(Y^2 \le \sigma^2\tau)] - \theta_0$. We first show that $|B(\theta)| \le 3\sigma^2\tau$. Define $\lambda = \tau^{1/2}$. Then

$$E[(Y^2 - \sigma^2)\mathbb{1}(Y^2 \le \sigma^2\tau)] \le E[Y^2 \mathbb{1}(Y^2 \le \sigma^2\tau)] \le \sigma^2\lambda^2,$$

and

$$
\begin{aligned}
E[(Y^2 - \sigma^2)\mathbb{1}(Y^2 \le \sigma^2\tau)] &= E(Y^2 - \sigma^2) - E[(Y^2 - \sigma^2)\mathbb{1}(Y^2 > \sigma^2\tau)] \\
&\ge \theta^2 - E(Y^2) = -\sigma^2 \ge -\sigma^2\lambda^2.
\end{aligned}
$$

By Lemma 3, $\theta_0 = -2\sigma^2\lambda\phi(\lambda)$. It follows that

$$|B(\theta)| \le \left|E[(Y^2-\sigma^2)\mathbb{1}(Y^2 \le \sigma^2\tau)]\right| + |\theta_0| \le \sigma^2\lambda^2 + 2\sigma^2\lambda\phi(\lambda) \le 3\sigma^2\lambda^2 = 3\sigma^2\tau.$$

We now show that $|B(\theta)| \le \theta^2$. Straightforward calculation yields for $\theta \ge 0$,

$$B'(\theta) = \sigma(1+\lambda^2)[\phi(\lambda+\theta/\sigma) - \phi(\lambda-\theta/\sigma)] + 2\theta[\tilde{\Phi}(-\lambda-\theta/\sigma) - \tilde{\Phi}(\lambda-\theta/\sigma)], \tag{S2.1}$$

$$\begin{aligned} B''(\theta) &= \phi(\lambda+\theta/\sigma)[-\lambda^2(\lambda+\theta/\sigma) - \lambda + \theta/\sigma] \\ &\quad + \phi(\lambda-\theta/\sigma)[-\lambda^2(\lambda-\theta/\sigma) - \lambda - \theta/\sigma] + 2[\tilde{\Phi}(-\lambda-\theta/\sigma) - \tilde{\Phi}(\lambda-\theta/\sigma)]. \end{aligned} \tag{S2.2}$$

It suffices to only consider $\theta \ge 0$ since $B(\theta) = B(-\theta)$. It follows from (S2.1) that for all $\theta \ge 0$, $B'(\theta) \le 2\theta$. Since $B(0) = 0$, this implies that

$$B(\theta) \le \theta^2, \qquad \forall\theta \ge 0. \tag{S2.3}$$

On the other hand, $\theta_0 \le 0$ immediately gives $B(\theta) \ge -\sigma^2 \ge -\theta^2$ for $\theta \ge \sigma$. For $0 \le \theta < \sigma$, we have $\sigma(1+\lambda^2) \ge 2\theta$. For $x > 0$, we have $\tilde{\Phi}(x) < x^{-1}\phi(x)$, so $\tilde{\Phi}(-\lambda-\theta/\sigma) = 1 - \tilde{\Phi}(\lambda+\theta/\sigma) \ge 1 - (\lambda+\theta/\sigma)^{-1}\phi(\lambda+\theta/\sigma)$. It then follows from (S2.1) that for $0 \le \theta < \sigma$,

$$\begin{aligned} B'(\theta) &\ge 2\theta[\phi(\lambda+\theta/\sigma) - \phi(\lambda-\theta/\sigma) + \tilde{\Phi}(-\lambda-\theta/\sigma) - \tilde{\Phi}(\lambda-\theta/\sigma)] \\ &\ge 2\theta[1 + (1-(\lambda+\theta/\sigma)^{-1})\phi(\lambda+\theta/\sigma) - \phi(\lambda-\theta/\sigma) - \tilde{\Phi}(\lambda-\theta/\sigma)] \\ &\ge 2\theta\left[1 + (1-(\lambda+\theta/\sigma)^{-1})\phi(\lambda+\theta/\sigma) - \frac{1}{\sqrt{2\pi}} - \frac{1}{2}\right] \ge 0. \end{aligned}$$

Coupled with $B(0) = 0$, this implies that $B(\theta) \geq 0 \geq -\theta^2$ for $0 \leq \theta < \sigma$. Hence,

$$B(\theta) \geq -\theta^2, \qquad \forall \theta \geq 0. \tag{S2.4}$$

Since $B(-\theta) = B(\theta)$, combining (S2.3) and (S2.4), we obtain $|B(\theta)| \leq \theta^2$ for all $\theta \in \mathbb{R}$. $\square$

Proof of Lemma 1. Let $Z \sim N(0, \sigma^2)$, and let $\theta_0 = E[(Z^2 - \sigma^2)\mathbb{1}(Z^2 \leq \sigma^2\tau)] = -2\sigma^2\tau^{1/2}\phi(\tau^{1/2})$, the second equality due to Lemma 3. It follows from the expression

$$\begin{aligned}
&E[(X^2 - \sigma^2)(Y^2 - \sigma^2)\mathbb{1}(X^2 \vee Y^2 > \sigma^2\tau)] \\
&\quad = \mu^2\theta^2 - E[(X^2 - \sigma^2)\mathbb{1}(X^2 \leq \sigma^2\tau)]E[(Y^2 - \sigma^2)\mathbb{1}(Y^2 \leq \sigma^2\tau)]
\end{aligned}$$

and

$$\begin{aligned}
\eta &= E[(Z_1^2 - \sigma^2)(Z_2^2 - \sigma^2)\mathbb{1}(Z_1^2 \vee Z_2^2 > \sigma^2\tau)] \\
&= -E[(Z_1^2 - \sigma^2)\mathbb{1}(Z_1^2 \leq \sigma^2\tau)]E[(Z_2^2 - \sigma^2)\mathbb{1}(Z_2^2 \leq \sigma^2\tau)] = -\theta_0^2
\end{aligned}$$

that we have

$$\begin{aligned}
&\left|E[(X^2 - \sigma^2)(Y^2 - \sigma^2)\mathbb{1}(X^2 \vee Y^2 > \sigma^2\tau)] - \eta - \mu^2\theta^2\right| \qquad \text{(S2.5)}\\
&= \left|E[(X^2 - \sigma^2)\mathbb{1}(X^2 \leq \sigma^2\tau)]E[(Y^2 - \sigma^2)\mathbb{1}(Y^2 \leq \sigma^2\tau)] - \theta_0^2\right|.
\end{aligned}$$

Using the decomposition $AB - ab = (A - a)(B - b) + a(B - b) + b(A - a)$ and the triangle inequality, we get

$$
\begin{aligned}
&\big|E[(X^2-\sigma^2)\mathbb{1}(X^2\le\sigma^2\tau)]E[(Y^2-\sigma^2)\mathbb{1}(Y^2\le\sigma^2\tau)]-\theta_0^2\big| \\
&\le \big|E[(X^2-\sigma^2)\mathbb{1}(X^2\le\sigma^2\tau)]-\theta_0\big|\big|E[(Y^2-\sigma^2)\mathbb{1}(Y^2\le\sigma^2\tau)]-\theta_0\big| \\
&\qquad + |\theta_0|\big|E[(X^2-\sigma^2)\mathbb{1}(X^2\le\sigma^2\tau)]-\theta_0\big| + |\theta_0|\big|E[(Y^2-\sigma^2)\mathbb{1}(Y^2\le\sigma^2\tau)]-\theta_0\big| \\
&\le \min\{\mu^2,3\sigma^2\tau\}\min\{\theta^2,3\sigma^2\tau\} + 2\sigma^2\tau^{1/2}\phi(\tau^{1/2})\min\{\mu^2,3\sigma^2\tau\} \\
&\qquad + 2\sigma^2\tau^{1/2}\phi(\tau^{1/2})\min\{\theta^2,3\sigma^2\tau\},
\end{aligned}
$$

the last inequality follows from Lemma 4 and substitution of the value of θ_0. □

Proof of Lemma 2

We have

$$
\begin{aligned}
&\mathrm{Var}[(X^2-\sigma^2)(Y^2-\sigma^2)\mathbb{1}(X^2\vee Y^2>\sigma^2\tau)] \\
&= E[(X^2-\sigma^2)^2(Y^2-\sigma^2)^2\mathbb{1}(X^2\vee Y^2>\sigma^2\tau)] \\
&\qquad - \big\{E[(X^2-\sigma^2)(Y^2-\sigma^2)\mathbb{1}(X^2\vee Y^2>\sigma^2\tau)]\big\}^2 \\
&= E[(X^2-\sigma^2)^2(Y^2-\sigma^2)^2] - E[(X^2-\sigma^2)^2\mathbb{1}(X^2\le\sigma^2\tau)(Y^2-\sigma^2)^2\mathbb{1}(Y^2\le\sigma^2\tau)] \\
&\qquad - \big\{E[(X^2-\sigma^2)(Y^2-\sigma^2)] - E[(X^2-\sigma^2)\mathbb{1}(X^2\le\sigma^2\tau)(Y^2-\sigma^2)\mathbb{1}(Y^2\le\sigma^2\tau)]\big\}^2 \\
&= \mathrm{Var}[(X^2-\sigma^2)(Y^2-\sigma^2)] - E[(X^2-\sigma^2)^2\mathbb{1}(X^2\le\sigma^2\tau)]E[(Y^2-\sigma^2)^2\mathbb{1}(Y^2\le\sigma^2\tau)] \\
&\qquad - \big\{E[(X^2-\sigma^2)\mathbb{1}(X^2\le\sigma^2\tau)]E[(Y^2-\sigma^2)\mathbb{1}(Y^2\le\sigma^2\tau)]\big\}^2 \\
&\qquad + 2\mu^2\theta^2E[(X^2-\sigma^2)\mathbb{1}(X^2\le\sigma^2\tau)]E[(Y^2-\sigma^2)\mathbb{1}(Y^2\le\sigma^2\tau)] \\
&\le \mathrm{Var}[(X^2-\sigma^2)(Y^2-\sigma^2)] + 2\mu^2\theta^2E[(X^2-\sigma^2)\mathbb{1}(X^2\le\sigma^2\tau)]E[(Y^2-\sigma^2)\mathbb{1}(Y^2\le\sigma^2\tau)] \\
&\le \mathrm{Var}[(X^2-\sigma^2)(Y^2-\sigma^2)] + 8\sigma^4\mu^2\theta^2\tau^2.
\end{aligned}
$$

Straightforward calculation yields

$$
\begin{aligned}
&\mathrm{Var}[(X^2-\sigma^2)(Y^2-\sigma^2)] \\
&= \mathrm{Var}(X^2-\sigma^2)\mathrm{Var}(Y^2-\sigma^2) + [E(X^2-\sigma^2)]^2\mathrm{Var}(Y^2-\sigma^2) + \mathrm{Var}(X^2-\sigma^2)[E(Y^2-\sigma^2)]^2 \\
&= [4\sigma^2\mu^2+2\sigma^4][4\sigma^2\theta^2+2\sigma^4] + \mu^4[4\sigma^2\theta^2+2\sigma^4] + \theta^4[4\sigma^2\mu^2+2\sigma^4] \\
&= 4\sigma^2\mu^4\theta^2 + 4\sigma^2\mu^2\theta^4 + 16\sigma^4\mu^2\theta^2 + 2\sigma^4\mu^4 + 2\sigma^4\theta^4 + 8\sigma^6\mu^2 + 8\sigma^6\theta^2 + 4\sigma^8.
\end{aligned}
$$

Let $d = E[(Z_1^2-\sigma^2)^4(Z_2^2-\sigma^2)^4] < \infty$. Then

$$
\begin{aligned}
&\mathrm{Var}[(X^2-\sigma^2)(Y^2-\sigma^2)\mathbb{1}(X^2 \vee Y^2 > \sigma^2\tau)] \\
&\le E[(X^2-\sigma^2)^2(Y^2-\sigma^2)^2\mathbb{1}(X^2 \vee Y^2 > \sigma^2\tau)] \\
&\le \Big(E[(X^2-\sigma^2)^4(Y^2-\sigma^2)^4]P(X^2 \vee Y^2 > \sigma^2\tau)\Big)^{1/2} \\
&= d^{1/2}\Big(1-P(|Z| \le \tau^{1/2})^2\Big)^{1/2}, \qquad \text{where } Z \sim N(0,1) \\
&\le (2d)^{1/2}\Big(1-P(|Z| \le \tau^{1/2})\Big)^{1/2} \\
&= 2d^{1/2}\tilde{\Phi}(\tau^{1/2})^{1/2},
\end{aligned}
$$

the second inequality follows from the Cauchy-Schwarz inequality.

Bibliography

Aldous, D. J. (1985). Exchangeability and related topics. In *École d'été de probabilités de Saint-Flour, XIII—1983*, Volume 1117 of *Lecture Notes in Math.*, pp. 1–198. Springer, Berlin.

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