# Supplementary Material
# Distributed Nonparametric Regression under Communication Constraints

## A. Proof of lemmas

### A.1. Proof of Lemma 3.1

. Write

$$\Theta_\ell(\alpha, c) = \left\{ \theta : \sum_{i=1}^{\ell} i^{2\alpha}\theta_i^2 \le c^2, \theta_i = 0 \text{ for } i \ge \ell + 1 \right\} \subset \Theta(\alpha, c).$$

For $\tau \in (0,1)$, write $s_i^2 = (1-\tau)\sigma_i^2$, and denote by $\pi_\tau(\theta)$ the prior distribution on $\theta$ such that $\theta_i \sim N(0, s_i^2)$ for $i = 1, \dots, \ell$, and $\mathbb{P}(\theta_i = 0) = 1$ for $i \ge \ell + 1$. For an estimator $\widehat{\theta}$ and its corresponding communication protocol, we observe that

$$\begin{aligned}
\sup_{\theta \in \Theta(\alpha,c)} \|\widehat{\theta} - \theta\|^2 &\ge \sup_{\theta \in \Theta_\ell(\alpha,c)} \|\widehat{\theta} - \theta\|^2 \\
&\ge \int_{\Theta_\ell(\alpha,c)} \|\widehat{\theta} - \theta\|^2 \mathrm{d}\pi_\tau(\theta) \\
&\ge I_\tau - r_\tau
\end{aligned}$$

where $I_\tau$ is the integrated risk of the estimator

$$I_\tau = \int_{\mathbb{R}^\ell \otimes \{0\}^\infty} \|\widehat{\theta} - \theta\|^2 \mathrm{d}\pi_\tau(\theta)$$

and $r_\tau$ is the residual

$$r_\tau = \int_{\overline{\Theta(\alpha,c)}} \|\widehat{\theta} - \theta\|^2 \mathrm{d}\pi_\tau(\theta)$$

where $\overline{\Theta(\alpha,c)} = (\mathbb{R}^\ell \otimes \{0\}^\infty) \backslash \Theta_\ell(\alpha, c)$. As $\lim_{\tau \to 0} I_\tau = \int_\Theta \mathbb{E}_\theta[\|\widehat{\theta} - \theta\|^2] \mathrm{d}\pi(\theta)$, it suffices to show that $r_\tau = o(I_\tau)$ as $\ell \to \infty$ for $\tau \in (0,1)$. Let $B_\ell = \sup_{\theta \in \Theta_\ell(\alpha,c)} \|\theta\|$, which is bounded since for any $\theta \in \Theta_\ell(\alpha, c)$

$$\|\theta\| = \sqrt{\sum_{i=1}^{\ell} \theta_i^2} = \sqrt{\sum_{i=1}^{\ell} i^{2\alpha}\theta_i^2} \le \sqrt{c^2} = c.$$

We have

$$\begin{aligned}
r_\tau &= \int_{\overline{\Theta_\ell(\alpha,c)}} \mathbb{E}_\theta \left[ \|\widehat{\theta} - \theta\|^2 \right] \mathrm{d}\pi_\tau(\theta) \\
&\le 2 \int_{\overline{\Theta_\ell(\alpha,c)}} \left( B_\ell^2 + \|\theta\|^2 \right) \mathrm{d}\pi_\tau(\theta) \\
&\le 2 \left( B_\ell^2 \mathbb{P}\left(\theta \notin \Theta_\ell(\alpha,c)\right) + \left( \mathbb{P}\left(\theta \notin \Theta_\ell(\alpha,c)\right) \mathbb{E}\left[\|\theta\|^4\right] \right) \right)
\end{aligned}$$

where we have used the Cauchy-Schwarz inequality. Noticing that

$$\begin{aligned}\mathbb{E}\left[\|\theta\|^4\right] &= \mathbb{E}\left[\left(\sum_{i=1}^{\ell}\theta_i^2\right)^2\right] \\ &= \sum_{i_1\neq i_2}\mathbb{E}\left[\theta_{i_1}^2\right]\mathbb{E}\left[\theta_{i_2}^2\right] + \sum_{i=1}^{\ell}\mathbb{E}\left[\theta_i^4\right] \\ &\le \sum_{i_1\neq i_2} s_{i_1}^2 s_{i_2}^2 + 3\sum_{i=1}^{\ell} s_i^4 \\ &\le 3(\sum_{i=1}^{\ell} s_i^2)^2 \le 3B_\ell^4,\end{aligned}$$

we obtain

$$r_\tau \le 2B_\ell^2\left(\mathbb{P}\left(\theta\notin\Theta_\ell(\alpha,c)\right) + \sqrt{3\mathbb{P}\left(\theta\notin\Theta_\ell(\alpha,c)\right)}\right) \le 6B_\ell^2\sqrt{3\mathbb{P}\left(\theta\notin\Theta_\ell(\alpha,c)\right)}.$$

Thus, we only need to show that $\sqrt{\mathbb{P}\left(\theta\notin\Theta_\ell(\alpha,c)\right)} = o(I_\tau)$. In fact,

$$\begin{aligned}\mathbb{P}\left(\theta\notin\Theta_\ell(\alpha,c)\right) &= \mathbb{P}\left(\sum_{i=1}^{\ell} i^{2\alpha}\theta_i^2 > c^2\right) \\ &= \mathbb{P}\left(\sum_{i=1}^{\ell} i^{2\alpha}(\theta_i^2 - \mathbb{E}[\theta_i^2]) > c^2 - (1-\tau)\sum_{i=1}^{\ell} i^{2\alpha}\sigma_i^2\right) \\ &= \mathbb{P}\left(\sum_{i=1}^{\ell} i^{2\alpha}(\theta_i^2 - \mathbb{E}[\theta_i^2]) > \tau c^2\right) \\ &= \mathbb{P}\left(\sum_{i=1}^{\ell} i^{2\alpha}s_i^2(Z_i^2-1) > \frac{\tau}{1-\tau}\sum_{i=1}^{\ell} i^{2\alpha}s_i^2\right)\end{aligned}$$

where $Z_i \sim N(0,1)$. By Lemma A.1, we get

$$\mathbb{P}(\theta\notin\Theta_\ell(m,c)) \le \exp\left(-\frac{\tau^2}{8(1-\tau)^2}\frac{\sum_{i=1}^{\ell} i^{2\alpha}s_i^2}{\max_{1\le i\le\ell} i^{2\alpha}s_i^2}\right) = \exp\left(-\frac{\tau^2}{8(1-\tau)^2}\frac{\sum_{i=1}^{\ell} i^{2\alpha}\sigma_i^2}{\max_{1\le i\le\ell} i^{2\alpha}\sigma_i^2}\right).$$

By the assumption that $\frac{\sum_{i=1}^{\ell} i^{2\alpha}\sigma_i^2}{\max_{1\le i\le\ell} i^{2\alpha}\sigma_i^2} = O(\ell)$, and that $\int_\Theta \mathbb{E}_\theta[\|\widehat{\theta}-\theta\|^2]\mathrm{d}\pi(\theta) = O(\ell^\delta)$, we conclude that $r_\tau = o(I_\tau)$ as $\ell\to\infty$. □

**Lemma A.1** (Lemma 3.5 in (Tsybakov, 2008))**.** *Suppose that $Z_1,\dots,Z_n \sim N(0,1)$ independently. For $t\in(0,1)$ and $\omega_i > 0$, $i=1,\dots,n$, we have*

$$\mathbb{P}\left(\sum_{i=1}^{n}\omega_i(Z_i^2-1) > t\sum_{i=1}^{n}\omega_i\right) \le \exp\left(-\frac{t^2\sum_{i=1}^{n}\omega_i}{8\max_{1\le i\le n}\omega_i}\right).$$

### A.2. Proof of Lemma 3.2

. Recall that we have $\theta_i \sim N(0,\sigma_i^2)$ and $X_{ij}|\theta_i \sim N(\theta_i,\varepsilon^2)$ for $i=1,\dots,\ell$, and $j=1,\dots,m$. For convenience, write $\theta=(\theta_1,\dots,\theta_n)$, $X_j=(X_{1j},\dots,X_{\ell j})$ and $X=(X_1,\dots,X_j)$. Suppose that we have a set of encoding functions $\Pi_j : \mathbb{R}^\ell \to \{1,\dots,M_j\}$ for $j=1,\dots,m$ satisfying that $\sum_{j=1}^{m}\log M_j \le mb$. Let $W_j = \Pi_j(X_j)$ be the message generated from the $j$th machine, and write $W=(W_1,\dots,W_m)$. Furthermore, we write $d_i = \mathbb{E}(\theta_i - \mathbb{E}(\theta_i|W))^2$ and

$d_{ij} = \mathbb{E}(X_{ij}|\theta, W_j)^2$. We then have

$$\begin{aligned}
\sum_{j=1}^m \log M_j &\geq H(W) \\
&\geq I(\theta, X; W) \\
&= I(\theta; W) + \sum_{j=1}^m I(X_j; W_j|\theta) \\
&= h(\theta) - h(\theta|W) + \sum_{j=1}^m (h(X_j|\theta) - h(X_j|\theta, W)) \\
&= \sum_{i=1}^\ell h(\theta_i) - \sum_{i=1}^\ell h(\theta_i|\theta_{1:(i-1)}, W) + \sum_{j=1}^m \left( \sum_{i=1}^\ell h(X_{ij}|\theta) - \sum_{i=1}^\ell h(X_{ij}|X_{1:(i-1),j}, \theta, W_j) \right) \\
&\geq \sum_{i=1}^\ell h(\theta_i) - \sum_{i=1}^\ell h(\theta_i|W) + \sum_{j=1}^m \left( \sum_{i=1}^\ell h(X_{ij}|\theta) - \sum_{i=1}^\ell h(X_{ij}|\theta, W_j) \right) \\
&\geq \sum_{i=1}^\ell \left( \frac{1}{2} \log \frac{\sigma_i^2}{d_i} + \sum_{j=1}^m \frac{1}{2} \log \frac{\varepsilon^2}{d_{ij}} \right). \qquad \text{(A.1)}
\end{aligned}$$

In order to obtain the relationship between $d_i$'s and $d_{ij}$'s, we consider the random vector $Y = \mathbb{E}(\theta|X)$, i.e., $Y_i = \mathbb{E}(\theta_i|X)$ for $i = 1, \ldots, n$. In fact, $Y_i$ takes the form

$$Y_i = \frac{\frac{1}{\varepsilon^2}}{\frac{1}{\sigma_i^2} + \frac{m}{\varepsilon^2}} \sum_{j=1}^m X_{ij}.$$

We first calculate the optimal mean squared error of estimating $Y_i$ based on $\theta$ and $W$

$$\mathbb{E}\left[(Y_i - \mathbb{E}(Y_i|\theta, W))^2\right] = \left( \frac{\frac{1}{\varepsilon^2}}{\frac{1}{\sigma_i^2} + \frac{m}{\varepsilon^2}} \right)^2 \mathbb{E}\left[ \left( \sum_{j=1}^m (X_{ij} - \mathbb{E}[X_{ij}|\theta, W_j]) \right)^2 \right] = \left( \frac{\frac{1}{\varepsilon^2}}{\frac{1}{\sigma_i^2} + \frac{m}{\varepsilon^2}} \right)^2 \sum_{j=1}^m d_{ij}$$

where we have used the equality that

$$\begin{aligned}
&\mathbb{E}\left[(X_{ij} - \mathbb{E}[X_{ij}|\theta, W_j])(X_{ij'} - \mathbb{E}[X_{ij'}|\theta, W_{j'}])\right] \\
&= \mathbb{E}\left[X_{ij} - \mathbb{E}[X_{ij}|\theta, X_{ij'}, W_j, W_{j'}]\right] \mathbb{E}\left[X_{ij'} - \mathbb{E}[X_{ij'}|\theta, W_{j'}]\right] = 0
\end{aligned}$$

for $j \neq j'$.

We then calculate the mean squared error of best linear estimator of $Y_i$ using $\theta_i$ and $T_i = \mathbb{E}(\theta_i|W)$. In particular, we search for $\beta_1$ and $\beta_2$ such that

$$\mathbb{E}\left[(Y_i - \beta_1\theta_i - \beta_2 T_i)^2\right]$$

is minimized. Towards that end, we calculate

$$\mathbb{E}\left[Y_i^2\right] = \left( \frac{\frac{1}{\varepsilon^2}}{\frac{1}{\sigma_i^2} + \frac{m}{\varepsilon^2}} \right)^2 \mathbb{E}\left[ \left( \sum_{j=1}^m X_{ij} \right)^2 \right] = \left( \frac{\frac{1}{\varepsilon^2}}{\frac{1}{\sigma_i^2} + \frac{m}{\varepsilon^2}} \right)^2 \left(m^2\varepsilon^2 + m\sigma_i^2\right) = \frac{\frac{m}{\varepsilon^2}}{\frac{1}{\sigma_i^2} + \frac{m}{\varepsilon^2}} \sigma_i^2 = \sigma_i^2 - \sigma_0^2$$

where we write $\sigma_0^2 = \frac{1}{\frac{1}{\sigma_i^2} + \frac{m}{\varepsilon^2}}$ to ease our notation. In addition, we have

$$\mathbb{E}\left[\theta_i^2\right] = \sigma_i^2 \text{ and } \mathbb{E}\left[Y_i\theta_i\right] = \frac{\frac{1}{\varepsilon^2}}{\frac{1}{\sigma_i^2} + \frac{m}{\varepsilon^2}} \mathbb{E}\left[ \sum_{j=1}^m \theta_i X_{ij} \right] = \sigma_i^2 - \sigma_0^2.$$

Furthermore, we notice that since $T_i = \mathbb{E}\left[\theta_i | W\right]$,

$$\mathbb{E}\left[T_i(\theta_i - T_i)\right] = \mathbb{E}\left[\mathbb{E}\left[T_i(\theta_i - T_i)\right] | W\right] = \mathbb{E}\left[T_i\right]\mathbb{E}\left[T_i - T_i\right] = 0,$$

and hence

$$d_i = \mathbb{E}\left[(\theta_i - T_i)^2\right] = \mathbb{E}\left[\theta_i(\theta_i - T_i) - T_i(\theta_i - T_i)\right] = \mathbb{E}\left[\theta_i(\theta_i - T_i)\right] = \mathbb{E}\left[\theta_i^2\right] - \mathbb{E}\left[\theta_i T_i\right],$$

from which we obtain

$$\mathbb{E}\left[T_i^2\right] = \mathbb{E}\left[\theta_i T_i\right] = \sigma_i^2 - d_i.$$

Finally, we have

$$\mathbb{E}\left[Y_i T_i\right] = \mathbb{E}\left[(\theta_i + (Y_i - \theta_i))T_i\right] = \mathbb{E}\left[\theta_i T_i\right] + \mathbb{E}[Y_i - \theta_i]\,\mathbb{E}[T_i] = \mathbb{E}\left[\theta_i T_i\right] = \sigma_i^2 - d_i$$

where the equality follows from the fact that $\theta_i$ and $\theta_i - Y_i$ are independent. To sum up, the covariance matrix of $(Y_i, \theta_i, T_i)$ is

$$\begin{pmatrix} \sigma_i^2 - \sigma_0^2 & \sigma_i^2 - \sigma_0^2 & \sigma_i^2 - d_i \\ \sigma_i^2 - \sigma_0^2 & \sigma_i^2 & \sigma_i^2 - d_i \\ \sigma_i^2 - d_i & \sigma_i^2 - d_i & \sigma_i^2 - d_i \end{pmatrix}.$$

Getting back to $\beta_1$ and $\beta_2$, they should satisfy

$$\mathbb{E}\left[\theta_i(Y_i - \beta_1\theta_i - \beta_2 T_i)\right] = 0, \quad \mathbb{E}\left[T_i(Y_i - \beta_1\theta_i - \beta_2 T_i)\right] = 0.$$

Solving the equations, we get

$$\beta_1 = \frac{d_i - \sigma_0^2}{d_i}, \quad \beta_2 = \frac{\sigma_0^2}{d_i},$$

and

$$\mathbb{E}\left[(Y_i - \beta_1\theta_i - \beta_2 T_i)^2\right] = \sigma_0^2 - \frac{\sigma_0^4}{d_i}.$$

Since conditional means minimize mean squared errors, we have

$$\mathbb{E}\left[(Y_i - \mathbb{E}(Y_i|\theta, W))^2\right] \le \mathbb{E}\left[(Y_i - \beta_1\theta_i - \beta_2 T_i)^2\right]$$

and therefore,

$$\left(\frac{\frac{1}{\varepsilon^2}}{\frac{1}{\sigma_i^2} + \frac{m}{\varepsilon^2}}\right)^2 \sum_{j=1}^m d_{ij} \le \frac{1}{\frac{1}{\sigma_i^2} + \frac{m}{\varepsilon^2}} - \left(\frac{1}{\frac{1}{\sigma_i^2} + \frac{m}{\varepsilon^2}}\right)^2 \frac{1}{d_i},$$

which gives

$$\sum_{j=1}^m d_{ij} \le \varepsilon^4 \left(\frac{1}{\sigma_i^2} + \frac{m}{\varepsilon^2} - \frac{1}{d_i}\right).$$

Now we plug this into (A.1), and obtain by applying Jensen's inequality that

$$\begin{aligned} mb &\ge \sum_{i=1}^{\ell} \left(\frac{1}{2}\log\frac{\sigma_i^2}{d_i} + \sum_{j=1}^m \frac{1}{2}\log\frac{\varepsilon^2}{d_{ij}}\right) \\ &\ge \sum_{i=1}^{\ell} \left(\frac{1}{2}\log\frac{\sigma_i^2}{d_i} + \frac{m}{2}\log\frac{\varepsilon^2}{\frac{1}{m}\sum_{j=1}^m d_{ij}}\right) \\ &\ge \sum_{i=1}^{\ell} \left(\frac{1}{2}\log\frac{\sigma_i^2}{d_i} + \frac{m}{2}\log\frac{\frac{m}{\varepsilon^2}}{\frac{1}{\sigma_i^2} + \frac{m}{\varepsilon^2} - \frac{1}{d_i}}\right), \end{aligned}$$

which completes the proof. □