# Optimal Rates for Nonparametric Density Estimation under Communication Constraints: Appendix 

Jayadev Acharya<br>acharya@cornell.edu<br>Aditya Vikram Singh<br>clement.canonne@sydney.edu. au<br>adityavs@iisc.ac.in htyagi@iisc.ac.in

## Contents

A Preliminaries ..... 2
A. 1 Besov spaces ..... 2
A. 2 Useful facts about Besov spaces ..... 3
A. 3 Useful probabilistic inequalities ..... 4
A. 4 Assumptions ..... 5
B Analysis of single-level estimator ..... 6
B. 1 Coupling of simulated and ideal estimators ..... 6
B. 2 Key lemmas ..... 7
B. 3 Analyzing the error ..... 8
C Analysis of multi-level estimator ..... 10
C. 1 Coupling of simulated and ideal estimators ..... 10
C. 2 Setting parameters ..... 11
C. 3 Analysing the error ..... 12
C.3.1 Linear and bias terms ..... 12
C.3.2 Details term ..... 13
C.3.3 Total error ..... 18
D Lower Bounds ..... 20
D. 1 Upper bound on average discrepancy ..... 20
D. 2 Lower bound on average discrepancy ..... 20
D. 3 Lower bound on $\mathcal{L}_{r}^{*}(n, \ell, p, q, s)$ for $r<(s+1) p$ ..... 21
D. 4 Lower bound on $\mathcal{L}_{r}^{*}(n, \ell, p, q, s)$ for $r \geq(s+1) p$ ..... 24
D. 5 Concluding the proof of Theorem 1.1 ..... 27

Organization. In Appendix A, we define Besov spaces and state some useful facts about them. In Appendix B, we analyze the single-level estimator and prove Theorem 1.2. In Appendix C, we analyze the multi-level estimator and prove Theorem 1.3. Finally, in Appendix D, we give details of our lower bound technique and prove Theorem 1.1.

## A Preliminaries

In this section, we formally define Besov spaces and provide some standard facts about them. We then state two useful probabilistic inequalities. Finally, we recall the assumptions that we make on the density function $f$.

## A. 1 Besov spaces

Our exposition here mainly derives from $[4,5]$. We start with a discussion on wavelets.
Wavelets. A wavelet basis for $\mathcal{L}^{2}(\mathbb{R})$ is generated using two functions: $\phi$ (father wavelet) and $\psi$ (mother wavelet). The main feature that distinguishes wavelet basis from the Fourier basis is that the functions $\phi$ and $\psi$ can have compact support. More precisely, there exists a function $\phi: \mathbb{R} \rightarrow \mathbb{R}$ such that

1. $\{\phi(\cdot-k): k \in \mathbb{Z}\}$ forms an orthonormal family of $\mathcal{L}^{2}(\mathbb{R})$. Let $V_{0}=\operatorname{span}\{\phi(\cdot-k): k \in \mathbb{Z}\}$.
2. For $j \in \mathbb{Z}$, let $V_{j}=\operatorname{span}\left\{\phi_{j, k}: k \in \mathbb{Z}\right\}$, where $\phi_{j, k}(x)=2^{j / 2} \phi\left(2^{j} x-k\right)$. Then $V_{j} \subset V_{j+1}$.
3. $\phi \in \mathcal{L}^{2}(\mathbb{R}), \int \phi(x) d x=1$.

Remark: Conditions $1,2,3$ ensure that $\cap_{j \in \mathbb{Z}} V_{j}=\{0\}$ and $\cup_{j \in \mathbb{Z}} V_{j}=\mathcal{L}^{2}(\mathbb{R})$.
4. $\phi$ satisfies the following regularity conditions for a given $N \in \mathbb{Z}_{+}$:
(a) There exists a bounded non-increasing function $\Phi$ such that $\int \Phi(|x|)|x|^{N} d x<\infty$, and $|\phi(x)| \leq \Phi(|x|)$ almost everywhere.
(b) $\phi$ is $N+1$ times (weakly) differentiable and $\phi^{(N+1)}$ satisfies ess $\sup _{x} \sum_{k \in \mathbb{Z}}\left|\phi^{(N+1)}(x-k)\right|<$ $\infty$.
$\phi$ satisfying (a),(b) is said to be $N$-regular.
Let $W_{j} \subset \mathcal{L}^{2}(\mathbb{R})$ be a subspace such that $V_{j+1}=V_{j} \oplus W_{j}$ (i.e. $V_{j+1}=V_{j}+W_{j}$ and $V_{j} \cap W_{j}=\{0\}$ ). Then, there exists a function $\psi: \mathbb{R} \rightarrow \mathbb{R}$ such that

1. $\{\psi(\cdot-k): k \in \mathbb{Z}\}$ forms an orthonormal basis of $W_{0}$.
2. $\operatorname{span}\left\{\psi_{j, k}: j \in \mathbb{Z}, k \in Z\right\}=\mathcal{L}^{2}(\mathbb{R})$, where $\psi_{j, k}(x)=2^{j / 2} \psi\left(2^{j} x-k\right)$.
3. $\psi$ satisfies the same regularity conditions as $\phi$.

For any $L \in \mathbb{Z}$, we can decompose $\mathcal{L}^{2}(\mathbb{R})$ as

$$
\mathcal{L}^{2}(\mathbb{R})=V_{L} \bigoplus W_{L} \bigoplus W_{L+1} \bigoplus \cdots
$$

That is, for any $f \in \mathcal{L}^{2}(\mathbb{R})$

$$
\begin{equation*}
f=\sum_{k \in \mathbb{Z}} \alpha_{L, k} \phi_{L, k}+\sum_{j \geq L} \sum_{k \in \mathbb{Z}} \beta_{j, k} \psi_{j, k}, \tag{1}
\end{equation*}
$$

where

$$
\alpha_{L, k}=\int f(x) \phi_{L, k}(x) d x, \quad \beta_{j, k}=\int f(x) \psi_{j, k}(x) d x
$$

are called the wavelet coefficients of $f$, where the convergence is to be understood in the $\mathcal{L}^{2}$ sense in general; however, when $\phi, \psi$ satisfies the regularity condition, then the convergence holds in $\mathcal{L}^{p}$ sense for $p \in[1, \infty]$ (see Fact A.4). Moreover, for a father wavelet $\phi$, there is a canonical mother wavelet $\psi$ (Section 5.2 in [5]) corresponding to $\phi$.

Besov spaces. We now define Besov space $\mathcal{B}(p, q, s)$ with parameters $p, q, s$, where $1 \leq p, q \leq \infty$, $s>0$. Let $\phi$ be a father wavelet satisfying properties (1)-(4) above, with $N>s$. Then,

$$
\begin{equation*}
f \in \mathcal{B}(p, q, s) \Longleftrightarrow\left\|\alpha_{0} \cdot\right\|_{p}+\left(\sum_{j=0}^{\infty}\left(2^{s+\frac{1}{2}-\frac{1}{p}}\left\|\beta_{j} \cdot\right\|_{p}\right)^{q}\right)^{1 / q}<\infty \tag{2}
\end{equation*}
$$

where $\left\|\alpha_{0} \cdot\right\|_{p}$ is the $\ell_{p}$ norm of the sequence $\left\{\alpha_{0, k}\right\}_{k \in \mathbb{Z}},\left\|\beta_{j} \cdot\right\|_{p}$ is the $\ell_{p}$ norm of the sequence $\left\{\beta_{j, k}\right\}_{k \in \mathbb{Z}}$. The sequences $\left\{\alpha_{0, k}\right\}_{k \in \mathbb{Z}},\left\{\beta_{j, k}\right\}_{k \in \mathbb{Z}}$ come from the wavelet expansion of $f$ using the father wavelet $\phi$ and the corresponding mother wavelet $\psi$. The definition (2) of $\mathcal{B}(p, q, s)$ is invariant to the choice of $\phi$ as long as $N>s$. For the purposes of defining Besov norm, we fix a particular $\phi, \psi$, where $\phi$ is $N$-regular with $N>s$. Then, the Besov norm of a function $f$ is defined as

$$
\begin{equation*}
\|f\|_{p q s}:=\left\|\alpha_{0} \cdot\right\|_{p}+\left(\sum_{j=0}^{\infty}\left(2^{s+\frac{1}{2}-\frac{1}{p}}\left\|\beta_{j} \cdot\right\|_{p}\right)^{q}\right)^{1 / q} \tag{3}
\end{equation*}
$$

## A. 2 Useful facts about Besov spaces

We record a few facts about Besov spaces that will be used in our analysis. Throughout, we assume that $f \in \mathcal{B}(p, q, s)$ with $\|f\|_{s p q} \leq 1$ and $\operatorname{supp}(f) \subseteq[0,1]$. The following fact is apparent from our discussion on wavelets.

Fact A.1. Let the wavelet expansion of $f$ be as in (1). For $H \geq L$, define $f^{(H)}:=\sum_{k \in \mathbb{Z}} \alpha_{L, k} \phi_{L, k}+$ $\sum_{j=L}^{H-1} \sum_{k \in \mathbb{Z}} \beta_{j, k} \psi_{j, k}$. Then, $f^{(H)}=\sum_{k \in \mathbb{Z}} \alpha_{H, k} \phi_{H, k}$.

Since $\operatorname{supp}(f) \subseteq[0,1]$, and $\operatorname{supp}\left(\phi_{j, 0}\right), \operatorname{supp}\left(\psi_{j, 0}\right) \subseteq\left[-A 2^{-j}, A 2^{-j}\right]$, there is no overlap between $\operatorname{supp}(f)$ and $\operatorname{supp}\left(\phi_{j, k}\right), \operatorname{supp}\left(\psi_{j, k}\right)$ for all but a finite number of indices $k$. In particular, we have the following.

Fact A.2. Let the wavelet expansion of $f$ be as in (1). Then, for any given $j \in \mathbb{Z}_{+}$, there are $O\left(2^{j}\right)$ translation indices $k$ such that $\phi_{j, k}(x)$ or $\psi_{j, k}(x)$ is possibly non-zero, where $x \in \operatorname{supp}(f)$.

For $f \in \mathcal{B}(p, q, s)$, it is clear from the definition of Besov norm that the wavelet coefficients must decay sufficiently fast. More precisely, we have the following.

Fact A.3. If $f \in \mathcal{B}(p, q, s)$, then

$$
\lim _{j \rightarrow \infty} 2^{j p\left(s+\frac{1}{2}-\frac{1}{p}\right)} \sum_{k \in \mathbb{Z}}\left|\beta_{j, k}\right|^{p}=0
$$

and in particular there exists $C>0$ such that

$$
\left\|\beta_{j}\right\|_{p}^{p}=\sum_{k \in \mathbb{Z}}\left|\beta_{j, k}\right|^{p} \leq C \cdot 2^{-j p\left(s+\frac{1}{2}-\frac{1}{p}\right)}
$$

The next fact quantifies the approximation error when the wavelet expansion of $f$ is truncated.
Fact A.4. Let $f^{(H)}$ be as in Fact A.1. Then, for $r \geq 1$,

$$
\left\|f^{(H)}-f\right\|_{r} \leq C \begin{cases}2^{-H s} & \text { if } r \leq p \\ 2^{-H(s-1 / p+1 / r)} & \text { if } r>p\end{cases}
$$

The next fact (from equation (15) in [4]) gives a bound on $\|f\|_{\infty}$ when $\|f\|_{s p q} \leq 1$.
Fact A.5. Let $s>1 / p$. Then

$$
\|f\|_{\infty} \leq\left(1-2^{-(s-1 / p) q^{\prime}}\right)^{1 / q^{\prime}}
$$

where $1 / q+1 / q^{\prime}=1$.
Now, let $X_{1}, \ldots, X_{n}$ be independent samples from distribution with density $f$. For $j, k \in \mathbb{Z}$, define

$$
\begin{equation*}
\bar{\alpha}_{j, k}:=\frac{1}{n} \sum_{i=1}^{n} \phi_{j, k}\left(X_{i}\right), \quad \bar{\beta}_{j, k}:=\frac{1}{n} \sum_{i=1}^{n} \psi_{j, k}\left(X_{i}\right) . \tag{4}
\end{equation*}
$$

Observe that $\bar{\alpha}_{j, k}$ (resp., $\bar{\beta}_{j, k}$ ) is an unbiased estimate of $\alpha_{j, k}$ (resp., $\beta_{j, k}$ ). The following fact is from equation (16) in [4].

Fact A.6. Let $n \geq 2^{j}$. Then, for $r \geq 1$,

$$
\mathbb{E}\left[\left|\bar{\alpha}_{j, k}-\alpha_{j, k}\right|^{r}\right] \leq C n^{-r / 2}, \quad \mathbb{E}\left[\left|\bar{\beta}_{j, k}-\beta_{j, k}\right|^{r}\right] \leq C n^{-r / 2}
$$

where $C$ is a constant that depends on $p, q, s, r, \phi, \psi$.
We note another useful fact (obtained after setting $\beta=0$ in equation (21) in [4]).
Fact A.7. Let $g=\sum_{j=L}^{H} \sum_{k \in \mathbb{Z}} \widehat{g}_{j, k} \psi_{j, k}$, where $\widehat{g}_{j, k}$ is random. Then, for $r \geq 1$,

$$
\mathbb{E}\left[\|g\|_{r}^{r}\right] \leq C(H-L)^{(r / 2-1)_{+}} \sum_{j=L}^{H} 2^{j(r / 2-1)} \sum_{k \in \mathbb{Z}} \mathbb{E}\left[\left|\widehat{g}_{j, k}\right|^{r}\right]
$$

where $C$ is a constant that depends on $r$.

## A. 3 Useful probabilistic inequalities

Theorem A. 8 (Rosenthal's inequality [6]). Let $X_{1}, \ldots, X_{n}$ be independent random variables such that $\mathbb{E}\left[X_{i}\right]=0$. and $\mathbb{E}\left[\left|X_{i}\right|^{r}\right]<\infty$ for every $i$.

1. Suppose $\mathbb{E}\left[X_{i}^{2}\right]<\infty$ for every $i$. Then, for $1 \leq r \leq 2$,

$$
\mathbb{E}\left[\left|\sum_{i=1}^{n} X_{i}\right|^{r}\right] \leq\left(\sum_{i=1}^{n} \mathbb{E}\left[X_{i}^{2}\right]\right)^{r / 2}
$$

(This just follows from concavity of $f(x)=x^{r / 2}$ for $r \leq 2$.)
2. Suppose $\mathbb{E}\left[\left|X_{i}\right|^{r}\right]<\infty$ for every $i$. Then, for $r>2$, there exists a constant $K_{r}$ depending only on $r$ such that

$$
\mathbb{E}\left[\left|\sum_{i=1}^{n} X_{i}\right|^{r}\right] \leq K_{r}\left\{\sum_{i=1}^{n} \mathbb{E}\left[\left|X_{i}\right|^{r}\right]+\left(\sum_{i=1}^{n} \mathbb{E}\left[X_{i}^{2}\right]\right)^{r / 2}\right\}
$$

Theorem A. 9 (Bernstein's inequality). Let $X_{1}, \ldots, X_{n}$ be independent random variables such that $\left|X_{i}\right| \leq b$ almost surely, and $\mathbb{E}\left[X_{i}^{2}\right] \leq v_{i}$ for every $i$. Let $X:=\sum_{i=1}^{n} X_{i}$ and $V:=\sum_{i=1}^{n} v_{i}$. Then, for every $u \geq 0$,

$$
\operatorname{Pr}(|X-\mathbb{E}[X]| \geq u) \leq \exp \left(-\frac{u^{2}}{2\left(V+\frac{b u}{3}\right)}\right)
$$

## A. 4 Assumptions

We recall here the assumptions we make on the density $f$ :

1. $f$ is compactly supported: without loss of generality, $\operatorname{supp}(f) \subseteq[0,1]$.
2. Besov norm of $f$ is bounded: without loss of generality, $\|f\|_{p q s} \leq 1$.

Our algorithm works with any father and mother wavelets $\phi$ and $\psi$ satisfying the following conditions:

1. $\phi$ and $\psi$ are $N$-regular, where $N>s$, and
2. $\operatorname{supp}(\phi), \operatorname{supp}(\psi) \subseteq[-A, A]$ for some integer $A>0$ (which may depend on $N$ ).

As a concrete example, Daubechies' family of wavelets [3] satisfies these assumptions.

## B Analysis of single-level estimator

Our goal is to upperbound the worst-case $\mathcal{L}^{r} \operatorname{loss} \mathbb{E}\left[\|\widehat{f}-f\|_{r}^{r}\right]$, where $\widehat{f}$ is the estimate output by the referee. Arguments and results in Sections B. 1 and B. 2 will be used in the analysis of the multi-level estimator as well.

## B. 1 Coupling of simulated and ideal estimators

Denote by $\mathbf{p}$ the probability distribution corresponding to density $f$. Recall that $\mathbf{p}_{\mathcal{Z}(H)}$ is the distribution after quantization of samples from $\mathbf{p}$. Suppose the referee has $m$ samples from $\mathbf{p}_{\mathcal{Z}^{(H)}}$ using which it outputs the estimate $\widehat{f}$. To compute $\mathbb{E}\left[\|\widehat{f}-f\|_{r}^{r}\right]$, consider the following statistically equivalent situation:

- There are $m$ i.i.d. samples $X_{1}, \ldots, X_{m} \sim \mathbf{p}$.
- For each $i \in \llbracket m \rrbracket$, let

$$
\widehat{\phi}_{H, k}\left(X_{i}\right)= \begin{cases}0 & \text { if } k \notin \mathcal{A}_{B_{i}}^{(H)},  \tag{5}\\ 2^{H / 2} Q\left(V_{i}\right)(k) & \text { if } k \in \mathcal{A}_{B_{i}}^{(H)},\end{cases}
$$

where $B_{i}$ is the bin $X_{i}$ lies in, $Q\left(V_{i}\right)$ is obtained by quantizing $\left\{2^{-H / 2} \phi_{H, k}\left(X_{i}\right)\right\}_{k \in \mathcal{A}_{B_{i}}^{(H)}}$ using Algorithm 1, and $Q\left(V_{i}\right)(k)$ is the entry in $Q\left(V_{i}\right)$ corresponding to $k \in \mathcal{A}_{B_{i}}^{(H)}$. In other words, $\left\{\widehat{\phi}_{H, k}\left(X_{i}\right)\right\}_{k \in \mathbb{Z}}$ is the quantized version of $\left\{\phi_{H, k}\left(X_{i}\right)\right\}_{k \in \mathbb{Z}}$.

- Define $\hat{f}$ as

$$
\begin{equation*}
\widehat{f}=\sum_{k \in \mathbb{Z}} \widehat{\alpha}_{H, k} \phi_{H, k}, \quad \text { where } \widehat{\alpha}_{H, k}=\frac{1}{m} \sum_{i=1}^{m} \widehat{\phi}_{H, k}\left(X_{i}\right), k \in \mathbb{Z} . \tag{6}
\end{equation*}
$$

Then, computing the $\mathcal{L}^{r}$ loss for the single-level estimator is equivalent to computing the $\mathcal{L}^{r}$ loss for $\widehat{f}$ defined in (6). From here on, when we refer to $\widehat{f}$, we mean $\widehat{f}$ defined in (6). Now,

$$
\begin{equation*}
\mathbb{E}\left[\|\widehat{f}-f\|_{r}^{r}\right] \leq 2^{r-1}\left(\mathbb{E}\left[\|\widehat{f}-\bar{f}\|_{r}^{r}\right]+\mathbb{E}\left[\|\bar{f}-f\|_{r}^{r}\right]\right) \tag{7}
\end{equation*}
$$

where $\bar{f}$ is defined as

$$
\begin{equation*}
\bar{f}=\sum_{k \in \mathbb{Z}} \bar{\alpha}_{H, k} \phi_{H, k} \quad \text { where } \quad \bar{\alpha}_{H, k}=\frac{1}{m} \sum_{i=1}^{m} \phi_{H, k}\left(X_{i}\right), k \in \mathbb{Z} . \tag{8}
\end{equation*}
$$

Note that this is just the classical density estimate obtained using $m$ samples $X_{1}, \ldots, X_{m}$. The idea behind introducing this coupling is to facilitate analysis by bringing in the classical density estimate (8) and breaking up the $\mathcal{L}^{r}$ loss as in (7).

## B. 2 Key lemmas

Since these lemmas will be used in the analysis of the multi-level estimator as well, we discuss them in more generality than would be required for only analyzing the single-level estimator. For $j, k \in \mathbb{Z}$, define

$$
\begin{array}{ll}
\widehat{\alpha}_{j, k}:=\frac{1}{m} \sum_{i=1}^{m} \widehat{\phi}_{j, k}\left(X_{i}\right), & \widehat{\beta}_{j, k}:=\frac{1}{m} \sum_{i=1}^{m} \widehat{\psi}_{j, k}\left(X_{i}\right), \\
\bar{\alpha}_{j, k}:=\frac{1}{m} \sum_{i=1}^{m} \phi_{j, k}\left(X_{i}\right), & \bar{\beta}_{j, k}:=\frac{1}{m} \sum_{i=1}^{m} \psi_{j, k}\left(X_{i}\right),
\end{array}
$$

where $\left\{\widehat{\phi}_{j, k}\left(X_{i}\right)\right\}_{k \in \mathbb{Z}}$ and $\left\{\widehat{\psi}_{j, k}\left(X_{i}\right)\right\}_{k \in \mathbb{Z}}$ are quantized versions of $\left\{\phi_{j, k}\left(X_{i}\right)\right\}_{k \in \mathbb{Z}}$ and $\left\{\psi_{j, k}\left(X_{i}\right)\right\}_{k \in \mathbb{Z}}$, respectively (as in (5), with $H$ replaced by $j ; \phi$ replaced by $\psi$, and $\mathcal{A}_{B_{i}}^{(j)}$ replaced by $\mathcal{B}_{B_{i}}^{(j)}$ in the case of $\widehat{\beta}_{j, k}$ ). The following claim bounds the error between quantized and unquantized (classical) estimates of wavelet coefficients.
Claim B. 1 (Error between quantized and unquantized estimates). For $r \geq 1$, we have

$$
\mathbb{E}\left[\left|\widehat{\alpha}_{j, k}-\bar{\alpha}_{j, k}\right|^{r}\right] \leq C \begin{cases}\frac{1}{m^{r / 2}}, & \text { if } r \in[1,2], \\ \frac{2^{j(r / 2-1)}}{m^{r-1}}+\frac{1}{m^{r / 2}}, & \text { if } r>2,\end{cases}
$$

for a constant $C$. The same bound holds for $\mathbb{E}\left[\left|\widehat{\beta}_{j, k}-\bar{\beta}_{j, k}\right|^{r}\right]$ as well.
Proof. For a given $j, k$,

$$
\left.\mathbb{E}\left[\left|\widehat{\alpha}_{j, k}-\bar{\alpha}_{j, k}\right|^{r}\right]=\mathbb{E}\left[\left|\frac{1}{m} \sum_{i=1}^{m}\left(\widehat{\phi}_{j, k}\left(X_{i}\right)-\phi_{j, k}\left(X_{i}\right)\right) \mathbb{1}_{\left\{\mathcal{A}_{B_{i}}^{(j)} \ni k\right\}}\right|^{r}\right]\right]=\frac{1}{m^{r}} \mathbb{E}\left[\left|\sum_{i=1}^{m} Y_{i k}\right|^{r}\right]
$$

where

$$
Y_{i k}:=\left(\widehat{\phi}_{j, k}\left(X_{i}\right)-\phi_{j, k}\left(X_{i}\right)\right) \mathbb{1}_{\left\{k \in \mathcal{A}_{B_{i}}^{(j)}\right\}} .
$$

Note that, since the quantization is unbiased, we have $\mathbb{E}\left[Y_{i k}\right]=0$. Moreover, $\left|Y_{i k}\right| \lesssim 2^{j / 2}$ almost surely. We first consider the case $r>2$. Then, by Rosenthal's inequality (Theorem A.8),

$$
\begin{equation*}
\mathbb{E}\left[\left|\sum_{i=1}^{m} Y_{i k}\right|^{r}\right] \lesssim\left(2^{j / 2}\right)^{(r-2)} \sum_{i=1}^{m} \mathbb{E}\left[Y_{i k}{ }^{2}\right]+\left(\sum_{i=1}^{m} \mathbb{E}\left[Y_{i k}^{2}\right]\right)^{\frac{r}{2}}=2^{j\left(\frac{r}{2}-1\right)} m \mathbb{E}\left[Y_{1 k}^{2}\right]+m^{\frac{r}{2}} \mathbb{E}\left[Y_{1 k}^{2}\right]^{\frac{r}{2}} \tag{11}
\end{equation*}
$$

Moreover,

$$
\mathbb{E}\left[Y_{i k}^{2}\right]=\mathbb{E}\left[\left(\widehat{\phi}_{j, k}\left(X_{i}\right)-\phi_{j, k}\left(X_{i}\right)\right)^{2} \mathbb{1}_{\left\{k \in \mathcal{A}_{B_{i}}^{(j)}\right\}}\right] \lesssim 2^{j} \operatorname{Pr}\left(k \in \mathcal{A}_{B_{i}}^{(j)}\right) .
$$

Now, note that

$$
\operatorname{Pr}\left(k \in \mathcal{A}_{B_{i}}^{(j)}\right)=\operatorname{Pr}\left(X_{i} \in \operatorname{supp}\left(\phi_{j, k}\right)\right) \leq \frac{2 A}{2^{j}}\|f\|_{\infty} \lesssim \frac{1}{2^{j}} \quad \text { (using Fact A.5) }
$$

which gives

$$
\mathbb{E}\left[Y_{i k}^{2}\right] \lesssim 1
$$

Substituting this in (11), we get the desired result when $r>2$. For $r \in[1,2]$, using part (1) of Theorem A.8, only the second term in (11) remains. This gives the result for $r \in[1,2]$. The proof for $\mathbb{E}\left[\left|\widehat{\beta}_{j, k}-\bar{\beta}_{j, k}\right|^{r}\right]$ is analogous.

The claim above, combined with Fact A.6, lets us bound the error between quantized estimates and true coefficients as follows.

Claim B. 2 (Error between quantized estimates and true coefficients). Let $m \geq 2^{j}$. Then, for $r \geq 1$, we have

$$
\mathbb{E}\left[\left|\widehat{\alpha}_{j, k}-\alpha_{j, k}\right|^{r}\right] \leq \frac{C}{m^{r / 2}}, \quad \mathbb{E}\left[\left|\widehat{\beta}_{j, k}-\beta_{j, k}\right|^{r}\right] \leq \frac{C}{m^{r / 2}}
$$

for a constant $C$.
Proof. Note that

$$
\mathbb{E}\left[\left|\widehat{\alpha}_{j, k}-\alpha_{j, k}\right|^{r}\right] \leq 2^{r-1}\left(\mathbb{E}\left[\left|\widehat{\alpha}_{j, k}-\bar{\alpha}_{j, k}\right|^{r}\right]+\mathbb{E}\left[\left|\bar{\alpha}_{j, k}-\alpha_{j, k}\right|^{r}\right]\right) .
$$

The first term can be handled with Claim B.1. The second term can be bound using Fact A.6. Overall, for $r>2$, we get

$$
\mathbb{E}\left[\left|\widehat{\alpha}_{j, k}-\alpha_{j, k}\right|^{r}\right] \lesssim \frac{2^{j(r / 2-1)}}{m^{r-1}}+\frac{1}{m^{r / 2}} \leq \frac{2}{m^{r / 2}},
$$

since $m \geq 2^{j}$. We get the same bound for $r \in[1,2]$. The result follows. The bound on $\mathbb{E}\left[\left|\widehat{\beta}_{j, k}-\beta_{j, k}\right|^{r}\right]$ is obtained in the same way.

Since, for any $j$, there are $O\left(2^{j}\right)$ translations $k$ for which the coefficients are non-zero (Fact A.2), Claim B. 2 readily implies the corollary below.

Corollary B.3. Let $m \geq 2^{j}$. Then, for $r \geq 1$ and a constant $C$., we have

$$
\begin{gathered}
\sum_{k \in \mathbb{Z}} \mathbb{E}\left[\left|\widehat{\alpha}_{j, k}-\alpha_{j, k}\right|^{r}\right] \leq C \frac{2^{j}}{m^{r / 2}}, \\
\sum_{k \in \mathbb{Z}} \mathbb{E}\left[\left|\widehat{\beta}_{j, k}-\beta_{j, k}\right|^{r}\right] \leq C \frac{2^{j}}{m^{r / 2}}
\end{gathered}
$$

## B. 3 Analyzing the error

For the single-level estimator

$$
\begin{align*}
\mathbb{E}\left[\|\widehat{f}-f\|_{r}^{r}\right] & =\mathbb{E}\left[\left\|\widehat{f}-f^{(H)}+f^{(H)}-f\right\|_{r}^{r}\right] \\
& \leq 2^{r-1}\left(\mathbb{E}\left[\left\|\widehat{f}-f^{(H)}\right\|_{r}^{r}\right]+\left\|f-f^{(H)}\right\|_{r}^{r}\right), \tag{12}
\end{align*}
$$

where $f^{(H)}=\sum_{k \in \mathbb{Z}} \alpha_{H, k} \phi_{H, k}$. Now, from Fact A.4, we have $\left\|f-f^{(H)}\right\|_{r}^{r} \lesssim 2^{-H r \sigma}$, where

$$
\sigma= \begin{cases}s & \text { if }, r \leq p, \\ (s-1 / p+1 / r), & \text { if } r>p\end{cases}
$$

Moreover,

$$
\mathbb{E}\left[\left\|\widehat{f}-f^{(H)}\right\|_{r}^{r}\right]=\mathbb{E}\left[\left\|\sum_{k \in \mathbb{Z}}\left(\hat{\alpha}_{H, k}-\alpha_{H, k}\right) \phi_{H, k}\right\|_{r}^{r}\right]
$$

$$
\begin{align*}
& \lesssim 2^{H(r / 2-1)} \sum_{k \in \mathbb{Z}} \mathbb{E}\left[\left|\hat{\alpha}_{H, k}-\alpha_{H, k}\right|^{r}\right]  \tag{FactA.7}\\
& \lesssim 2^{H(r / 2-1)} \frac{2^{H}}{m^{r / 2}} \\
& =\left(\frac{2^{H}}{m}\right)^{r / 2}
\end{align*}
$$

(Corollary B.3)
(In our case, $m, H$ will be such that $m \geq 2^{H}$ holds, which is why we can use Corollary B.3.) Substituting these in (12), we get

$$
\mathbb{E}\left[\|\widehat{f}-f\|_{r}^{r}\right] \lesssim 2^{-H r \sigma}+\left(\frac{2^{H}}{m}\right)^{r / 2}
$$

Recall that $m$ is the number of quantized samples available with the referee, where, for a constant C,

$$
m= \begin{cases}n & \text { if } 2^{H} \lesssim 2^{\ell}(\text { no simulation required }) \\ C n 2^{\ell} / 2^{H} & \text { if } 2^{H} \gtrsim 2^{\ell} \text { (after simulation) } .\end{cases}
$$

In other words (ignoring constant $C$ ) $m=\frac{n 2^{\ell}}{2^{H} \vee 2^{\ell}}$, where $a \vee b=\max \{a, b\}$. Thus,

$$
\mathbb{E}\left[\|\widehat{f}-f\|_{r}^{r}\right] \lesssim 2^{-H r \sigma}+\left(\frac{2^{H}\left(2^{H} \vee 2^{\ell}\right)}{n 2^{\ell}}\right)^{r / 2} \leq 2^{-H r \sigma}+\left(\frac{2^{2 H}}{n 2^{\ell}}\right)^{r / 2}+\left(\frac{2^{H}}{n}\right)^{r / 2}
$$

Setting $H$ such that

$$
2^{H}=\left(n 2^{\ell}\right)^{\frac{1}{2 s+2}} \wedge n^{\frac{1}{2 s+1}}
$$

gives us

$$
\mathbb{E}\|\widehat{f}-f\|_{r}^{r} \lesssim\left(n 2^{\ell}\right)^{-\frac{r \sigma}{2 \sigma+2}} \vee n^{-\frac{r \sigma}{2 \sigma+1}} .
$$

For $r \leq p, \sigma=s$, which proves Theorem 1.2 in the main paper.

## C Analysis of multi-level estimator

Our goal is to upper bound the worst-case $\mathcal{L}^{r} \operatorname{loss} \mathbb{E}\left[\|\widehat{f}-f\|_{r}^{r}\right]$, where $\widehat{f}$ is the estimate output by the referee. We proceed by describing the coupling in the multi-estimator setting.

## C. 1 Coupling of simulated and ideal estimators

Denote by $\mathbf{p}$ the probability distribution corresponding to density $f$. Recall that we divide the players into $H-L+1$ groups. Suppose the referee obtains $m_{J}$ (quantized) samples from players in group- $J$. To compute $\mathbb{E}\left[\|\widehat{f}-f\|_{r}^{r}\right]$, consider the following statistically equivalent situation:

- For each $J \in \llbracket L, H \rrbracket$, there are $m_{J}$ i.i.d. samples $X_{1, J}, \ldots, X_{m_{J}, J} \sim \mathbf{p}$.
- For each $i \in \llbracket m_{L} \rrbracket$, let

$$
\widehat{\phi}_{L, k}\left(X_{i, L}\right)= \begin{cases}0 & \text { if } k \notin \mathcal{A}_{B_{i, L}}^{(L)},  \tag{13}\\ 2^{L / 2} Q\left(V_{i, L}\right)(k) & \text { if } k \in \mathcal{A}_{B_{i, L}}^{L()},\end{cases}
$$

where $B_{i, L}$ is the bin (out of $2^{L}$ bins) $X_{i, L}$ lies in, $Q\left(V_{i, L}\right)$ is obtained by quantizing $\left\{2^{-L / 2} \phi_{L, k}\left(X_{i, L}\right)\right\}_{k \in \mathcal{A}_{B_{i, L}}^{(L)}}$ using Algorithm 1, and $Q\left(V_{i, L}\right)(k)$ is the entry in $Q\left(V_{i, L}\right)$ corresponding to $k \in \mathcal{A}_{B_{i, L}}^{(L)}$. In other words, $\left\{\widehat{\phi}_{L, k}\left(X_{i, L}\right)\right\}_{k \in \mathbb{Z}}$ is the quantized version of $\left\{\phi_{L, k}\left(X_{i, L}\right)\right\}_{k \in \mathbb{Z}}$.

- Similarly, for each $J \in \llbracket L, H \rrbracket$, for each $i \in \llbracket m_{J} \rrbracket$, let

$$
\widehat{\psi}_{J, k}\left(X_{i, J}\right)= \begin{cases}0 & \text { if } k \notin \mathcal{B}_{B_{i, J}}^{(J)},  \tag{14}\\ 2^{J / 2} Q\left(V_{i, J}\right)(k) & \text { if } k \in \mathcal{B}_{B_{i, J}}^{(J)}\end{cases}
$$

That is, for each $J \in \llbracket L, H \rrbracket,\left\{\widehat{\psi}_{J, k}\left(X_{i, J}\right)\right\}_{k \in \mathbb{Z}}$ is the quantized version of $\left\{\psi_{J, k}\left(X_{i, J}\right)\right\}_{k \in \mathbb{Z}}$.

- Given thresholds $\left\{t_{J}\right\}_{J \in \llbracket L, H \rrbracket}$, define $\widehat{f}$ as

$$
\begin{equation*}
\widehat{f}=\sum_{k} \widehat{\alpha}_{L, k} \phi_{L, k}+\sum_{J=L}^{H} \sum_{k} \tilde{\beta}_{J, k} \psi_{J, k}, \tag{15}
\end{equation*}
$$

where

$$
\begin{align*}
& \widehat{\alpha}_{L, k}=\frac{1}{m_{L}} \sum_{i=1}^{m_{L}} \widehat{\phi}_{L, k}\left(X_{i, L}\right), \\
& \tilde{\beta}_{J, k}=\widehat{\beta}_{J, k} \mathbb{1}_{\left\{\widehat{\beta}_{J, k} \mid \geq t_{J}\right\}} \text { with } \widehat{\beta}_{J, k}=\frac{1}{m_{J}} \sum_{i=1}^{m_{J}} \widehat{\psi}_{J, k}\left(X_{i, J}\right) . \tag{16}
\end{align*}
$$

Then, computing $\mathcal{L}^{r}$ loss for the multi-level estimator is equivalent to computing $\mathcal{L}^{r}$ loss for $\widehat{f}$ defined in (15).

## C. 2 Setting parameters

Since we want our multi-level estimator to be adaptive, the parameters $L, H$, and $\left\{t_{J}\right\}_{J \in \llbracket L, H \rrbracket}$ should not depend explicitly on Besov parameters. We set $L, H$ as

$$
\begin{aligned}
2^{L} & :=C\left(\left(n 2^{\ell}\right)^{\frac{1}{2(N+1)+2}} \wedge n^{\frac{1}{2(N+1)+1}}\right), \\
2^{H} & :=C^{\prime}\left(\frac{\sqrt{n 2^{\ell}}}{\log n 2^{\ell}} \wedge \frac{n}{\log ^{2} n}\right)
\end{aligned}
$$

where $C, C^{\prime}>0$ are two constants, sufficiently large and small, respectively. Also, note that, since players in group- $J$ have alphabet size $O\left(2^{J}\right)$, we have

$$
m_{J}=\frac{n}{(H-L+1)} \cdot\left(\frac{2^{\ell}}{2^{J}} \wedge 1\right) \asymp \frac{n 2^{\ell}}{H\left(2^{J} \vee 2^{\ell}\right)}
$$

Note that the setting of $H$ implies both $H 2^{H} \ll \sqrt{n 2^{\ell}}$ and $H^{2} 2^{H} \ll n$, and consequently $m_{J} \geq J 2^{J}$.

Threshold values. How should we set the threshold values $\left\{t_{J}\right\}_{J \in \llbracket L, H \rrbracket}$ ? Since we will pay a cost for the coefficients we zero out (increase in bias), we would like to choose $t_{J}$ as small as possible. But, in order to have reasonable concentration, we also need $t_{J}$ to satisfy, for every (sufficiently large) $\gamma>0$,

$$
\operatorname{Pr}\left(\left|\widehat{\beta}_{J, k}-\beta_{J, k}\right| \geq \gamma t_{J}\right) \lesssim 2^{-\gamma J}
$$

so that our estimates concentrate well around their true value, and we only zero them out wrongly with very small probability. Now, a natural approach to choose $t_{J}$ according to the constraint above would be to use Hoeffding's inequality, as $\widehat{\beta}_{J, k}$ is the empirical mean of $m_{J}$ unbiased estimates of $\beta_{J, k}$, each with magnitude $\lesssim 2^{J / 2}$. One can check that this would lead to the setting of $t_{J} \asymp \sqrt{J 2^{J} / m_{J}}$, which, unfortunately, is too big (by a factor of $2^{J / 2}$ ) to give optimal rates. However, recall that the $m_{J}$ unbiased estimates, $\widehat{\psi}_{J, k}\left(X_{i, J}\right)$, are not only such that $\left|\widehat{\psi}_{J, k}\left(X_{i, J}\right)\right| \lesssim$ $2^{J / 2}$; in many cases, they are actually zero, since $\left|\widehat{\psi}_{J, k}\left(X_{i, J}\right)\right| \simeq 2^{J / 2} \mathbb{1}_{\left\{k \in \mathcal{B}_{B_{i, J}}^{(J)}\right\}}$. This allows us to derive the following, improving upon the naïve use of Hoeffding's inequality.
Lemma C.1. For $J \in \llbracket L, H \rrbracket$, setting $t_{J}:=\sqrt{J / m_{J}}$, we have

$$
\operatorname{Pr}\left(\left|\widehat{\beta}_{J, k}-\beta_{J, k}\right| \geq \gamma t_{J}\right) \leq 2^{-\gamma J}
$$

for every $\gamma \geq 6 A\|f\|_{\infty}$.
Proof. Fix $J, k$, and consider any $i \in \llbracket m_{J} \rrbracket$. Since $\left|\widehat{\psi}_{J, k}\left(X_{i, J}\right)\right| \leq b:=2^{J / 2}$ and

$$
\mathbb{E}\left[\widehat{\psi}_{J, k}\left(X_{i, J}\right)^{2}\right]=2^{J} \operatorname{Pr}\left(k \in \mathcal{B}_{B_{i, J}}^{(J)}\right) \leq 2^{J} \cdot\|f\|_{\infty} \cdot \frac{2 A}{2^{J}}=2 A\|f\|_{\infty}:=v
$$

where the inequality follows from our assumption that $\operatorname{supp}(\psi) \subseteq[-A, A]$. In particular, we have $v \asymp 1$. Recalling the definition of $\widehat{\beta}_{J, k}$ from (16), we can apply Bernstein's inequality (Theorem A.9) to obtain, for $t \geq 0$ and $\gamma \geq 3 v$,

$$
\operatorname{Pr}\left(\left|\widehat{\beta}_{J, k}-\beta_{J, k}\right| \geq \gamma t\right) \leq e^{-\frac{3 \gamma^{2} m_{J} t^{2}}{6 v+2 b \gamma t}}=e^{-\frac{3}{2} \cdot \frac{\gamma m_{J} t^{2}}{\frac{3 v}{\gamma}+2^{J / 2} t}} \leq e^{-\frac{3}{2} \cdot \frac{\gamma m_{J} t^{2}}{1+2^{J / 2} t}} \leq 2^{-\frac{2 \gamma m_{J t^{2}}}{1+2^{J / 2} t}} .
$$

Setting $t_{J}:=\sqrt{\frac{J}{m_{J}}} \vee \frac{J 2^{J / 2}}{m_{J}}$, we get $\operatorname{Pr}\left(\left|\widehat{\beta}_{J, k}-\beta_{J, k}\right|\right) \leq 2^{-\gamma J}$. Finally, our setting of $H$ and $m_{J}$ together imply that $t_{J}:=\sqrt{\frac{J}{m_{J}}}$, as (from our choice of parameters) $m_{J} \geq J 2^{J}$ for all $J \leq H$.

Conclusion. For constants $C, C^{\prime}, \kappa>0$, the values of parameters are summarized below.

$$
\begin{align*}
2^{L} & :=C\left(\left(n 2^{\ell}\right)^{\frac{1}{2(N+1)+2}} \wedge n^{\frac{1}{2(N+1)+1}}\right)  \tag{17}\\
2^{H} & :=C^{\prime}\left(\frac{\sqrt{n 2^{\ell}}}{\log n 2^{\ell}} \wedge \frac{n}{\log ^{2} n}\right)  \tag{18}\\
m_{J} & :=\frac{n}{(H-L+1)} \cdot\left(\frac{2^{\ell}}{2^{J}} \wedge 1\right) \asymp \frac{n 2^{\ell}}{H\left(2^{J} \vee 2^{\ell}\right)}  \tag{19}\\
t_{J} & :=\kappa \sqrt{\frac{J}{m_{J}}} \tag{20}
\end{align*}
$$

As previously mentioned choices imply both $H 2^{H} \ll \sqrt{n 2^{\ell}}$ and $H^{2} 2^{H} \ll n$, and consequently $m_{J} \geq J 2^{J}, J \in \llbracket L, H \rrbracket$.

## C. 3 Analysing the error

Following the outline of Theorem 3 of [4] and Theorem 5.1 of [2], we will bound $\mathcal{L}_{r}$ loss as

$$
\begin{equation*}
\mathbb{E}\left[\|f-\widehat{f}\|_{r}^{r}\right] \leq 3^{r-1}(\operatorname{bias}(f)+\operatorname{linear}(f)+\operatorname{details}(f)) \tag{21}
\end{equation*}
$$

where

$$
\begin{aligned}
\operatorname{bias}(f) & =\mathbb{E}\left[\left\|f-\sum_{k \in \mathbb{Z}} \alpha_{H, k} \phi_{H, k}\right\|_{r}^{r}\right] \\
\operatorname{linear}(f) & =\mathbb{E}\left[\left\|\sum_{k \in \mathbb{Z}}\left(\widehat{\alpha}_{L, k}-\alpha_{L, k}\right) \phi_{L, k}\right\|_{r}^{r}\right] \\
\operatorname{details}(f) & =\mathbb{E}\left[\left\|\sum_{J=L}^{H} \sum_{k \in \mathbb{Z}}\left(\widetilde{\beta}_{J, k}-\beta_{J, k}\right) \psi_{J, k}\right\|_{r}^{r}\right]
\end{aligned}
$$

and handle each of the three terms separately. Note that only the third term relates to thresholding.

## C.3.1 Linear and bias terms

Linear term. To bound linear $(f)$, we invoke Fact A. 7 and Corollary B. 3 as in the analysis of single-level estimator. This gives

$$
\begin{align*}
\mathbb{E}\left[\left\|\sum_{k \in \mathbb{Z}}\left(\widehat{\alpha}_{L, k}-\alpha_{L, k}\right) \phi_{L, k}\right\|_{r}^{r}\right] & \lesssim 2^{L\left(\frac{r}{2}-1\right)} \sum_{k \in \mathbb{Z}} \mathbb{E}\left[\left|\widehat{\alpha}_{L, k}-\alpha_{L, k}\right|^{r}\right] \lesssim 2^{L\left(\frac{r}{2}-1\right)} \cdot \frac{2^{L}}{m_{L}^{r / 2}}  \tag{22}\\
& =H^{\frac{r}{2}}\left(\left(\frac{2^{2 L}}{n 2^{\ell}}\right)^{r / 2} \vee\left(\frac{2^{L}}{n}\right)^{r / 2}\right)
\end{align*}
$$

$$
\begin{equation*}
\lesssim H^{\frac{r}{2}}\left(\left(n 2^{\ell}\right)^{-\frac{r(N+1)}{2(N+1)+2}} \vee n^{-\frac{r(N+1)}{2(N+1)+1}}\right) \leq H^{\frac{r}{2}}\left(\left(n 2^{\ell}\right)^{-\frac{r s}{2 s+2}} \vee n^{-\frac{r s}{2 s+1}}\right) \tag{23}
\end{equation*}
$$

where the second-to-last inequality relies on our choice of $L$.
Bias term. To bound bias $(f)$, we use Fact A. 4 to get, with $s^{\prime}=s-1 / p+1 / r$,

$$
\begin{equation*}
\operatorname{bias}(f) \leq C \cdot 2^{-H s^{\prime} r} \leq C^{\prime} \cdot\left(\sqrt{\frac{\log ^{2}\left(n 2^{\ell}\right)}{n 2^{\ell}}} \vee \frac{\log ^{2} n}{n}\right)^{r(s-1 / p+1 / r)} \tag{24}
\end{equation*}
$$

## C.3.2 Details term

To bound the term details $(f)$, we define, for $J \in \llbracket L, H \rrbracket$, the three sets of indices:

$$
\begin{array}{lr}
\widehat{\mathcal{I}}_{J}:=\left\{k \in \mathbb{Z}:\left|\widehat{\beta}_{J, k}\right|>\kappa t_{J}\right\} & \text { (estimate big: not thresholded) } \\
\mathcal{I}_{J}^{s}:=\left\{k \in \mathbb{Z}:\left|\beta_{J, k}\right| \leq \frac{1}{2} \kappa t_{J}\right\} & \text { (small coefficients) } \\
\mathcal{I}_{J}^{b}:=\left\{k \in \mathbb{Z}:\left|\beta_{J, k}\right|>2 \kappa t_{J}\right\} & \text { (big coefficients) }
\end{array}
$$

We will partition the error according to these sets of indices, and argue about them separately. Specifically, we write

$$
\begin{aligned}
\operatorname{details}(f)= & \mathbb{E}\left[\left\|\sum_{J=L}^{H} \sum_{k \in \widehat{\mathcal{I}}_{J} \cap \mathcal{I}_{J}^{s}}\left(\widetilde{\beta}_{J, k}-\beta_{J, k}\right) \psi_{J, k}\right\|_{r}^{r}\right]+\mathbb{E}\left[\left\|\sum_{J=L}^{H} \sum_{k \in \widehat{\mathcal{I}}_{J} \backslash \mathcal{I}_{J}^{s}}\left(\widetilde{\beta}_{J, k}-\beta_{J, k}\right) \psi_{J, k}\right\|_{r}^{r}\right] \\
& +\mathbb{E}\left[\left\|\sum_{J=L}^{H} \sum_{k \in \mathcal{I}_{J}^{b} \backslash \widehat{\mathcal{I}}_{J}} \beta_{J, k} \psi_{J, k}\right\|_{r}^{r}\right]+\mathbb{E}\left[\left\|\sum_{J=L}^{H} \sum_{k \notin \mathcal{I}_{J}^{b} \cup \widehat{\mathcal{I}}_{J}} \beta_{J, k} \psi_{J, k}\right\|_{r}^{r}\right] \\
= & E_{b s}+E_{b b}+E_{s b}+E_{s s}, w e
\end{aligned}
$$

the four errors coming from the "big-small," "big-big," "small-big," and"small-small" indices, respectively. Our analysis is along the lines of that in [2].

The term $E_{b s}$. We can write

$$
\begin{align*}
E_{b s} & \lesssim H^{r / 2} \sum_{J=L}^{H} 2^{J\left(\frac{r}{2}-1\right)} \sum_{k \in \mathbb{Z}} \mathbb{E}\left[\left|\widehat{\beta}_{J, k}-\beta_{J, k}\right|^{r} \mathbb{1}_{k \in \widehat{\mathcal{I}}_{J} \cap \mathcal{I}_{J}^{s}}\right]  \tag{FactA.7}\\
& \lesssim H^{r / 2} \sum_{J=L}^{H} 2^{J\left(\frac{r}{2}-1\right)} \sum_{k \in \mathbb{Z}} \mathbb{E}\left[\left|\widehat{\beta}_{J, k}-\beta_{J, k}\right|^{2 r}\right]^{\frac{1}{2}} \operatorname{Pr}\left(k \in \widehat{\mathcal{I}}_{J} \cap \mathcal{I}_{J}^{s}\right)^{\frac{1}{2}}  \tag{Cauchy-Schwarz}\\
& \lesssim H^{r / 2} \sum_{J=L}^{H} 2^{J\left(\frac{r}{2}-1\right)} \sum_{k \in \mathbb{Z}} \mathbb{E}\left[\left|\widehat{\beta}_{J, k}-\beta_{J, k}\right|^{2 r}\right]^{\frac{1}{2}} \operatorname{Pr}\left(\left|\widehat{\beta}_{J, k}-\beta_{J, k}\right|>\frac{\kappa}{2} t_{J}\right) \\
& \lesssim H^{r / 2} \sum_{J=L}^{H} 2^{J\left(\frac{r}{2}-1\right)} \sum_{k \in \mathbb{Z}} \mathbb{E}\left[\left|\widehat{\beta}_{J, k}-\beta_{J, k}\right|^{2 r}\right]^{\frac{1}{2}} 2^{-\frac{\kappa}{2} J}  \tag{LemmaC.1}\\
& \lesssim H^{r / 2} \sum_{J=L}^{H} 2^{J\left(\frac{r-\kappa}{2}-1\right)} \frac{2^{J}}{m_{J}^{r / 2}}
\end{align*}
$$

where the last inequality follows from Claim B. 2 and the $O\left(2^{J}\right)$-sparsity of coefficients (Fact A.2). Going forward, recalling our setting of $m_{J}$ we get

$$
\begin{align*}
E_{b s} & \lesssim H^{r / 2} \sum_{J=L}^{H} \frac{2^{J \frac{r-\kappa}{2}}}{m_{J}^{r / 2}} \lesssim H^{r / 2}\left(\frac{H}{n 2^{\ell}}\right)^{r / 2} \sum_{J=L}^{H} 2^{J \frac{r-\kappa}{2}}\left(2^{J} \vee 2^{\ell}\right)^{r / 2} \\
& \lesssim H^{r / 2}\left(\frac{H}{n 2^{\ell}}\right)^{r / 2} \sum_{J=L}^{H} 2^{J\left(r-\frac{\kappa}{2}\right)}+H^{r / 2}\left(\frac{H}{n}\right)^{r / 2} \sum_{J=L}^{H} 2^{J \frac{r-\kappa}{2}} \\
& \lesssim H^{r / 2}\left(\frac{H}{n 2^{\ell}}\right)^{r / 2} 2^{L(r-\kappa / 2)}+H^{r / 2}\left(\frac{H}{n}\right)^{r / 2} 2^{L^{\frac{r-\kappa}{2}}} \lesssim H^{r} n^{-r / 2} \tag{25}
\end{align*}
$$

where the inequalities hold for $\kappa>2 r$.
The term $E_{b b}$. Turning to the term $E_{b b}$, we have

$$
\begin{align*}
E_{b b} & =\left\|\sum_{J=L}^{H} \sum_{k \in \mathbb{Z}}\left(\widehat{\beta}_{J, k}-\beta_{J, k}\right) \psi_{J, k} \mathbb{1}_{k \in \widehat{\mathcal{I}}_{J} \backslash \mathbb{I}_{J}^{s}}\right\|_{r}^{r} \\
& \lesssim H^{r / 2} \sum_{J=L}^{H} 2^{J(r / 2-1)} \sum_{k \in \mathbb{Z}} \mathbb{E}\left[\left|\widehat{\beta}_{J, k}-\beta_{J, k}\right|^{r} \mathbb{1}_{k \in \widehat{\mathcal{I}}_{J} \backslash \mathcal{I}_{J}^{s}}\right]  \tag{FactA.7}\\
& \lesssim H^{r / 2} \sum_{J=L}^{H} \frac{2^{J(r / 2-1)}}{m_{J}^{r / 2}} \sum_{k \in \mathbb{Z}} \mathbb{1}_{k \notin \mathcal{I}_{J}^{s}}
\end{align*}
$$

using that $\mathbb{1}_{k \in \widehat{\mathcal{I}}_{J} \backslash \mathcal{I}_{J}^{s}} \leq \mathbb{1}_{k \notin \mathcal{I}_{J}^{s}}$ and Claim B.2. Using the definition of $\mathcal{I}_{J}^{s}$, for any nonnegative sequence $\left(\alpha_{J}\right)_{J}$ we can further bound this as

$$
\begin{align*}
E_{b b} & \lesssim H^{r / 2} \sum_{J=L}^{H} \frac{2^{J(r / 2-1)}}{m_{J}^{r / 2}} \sum_{k \in \mathbb{Z}} \mathbb{1}_{k \notin \mathcal{I}_{J}^{s}}\left|\beta_{J, k}\right|^{\alpha_{J}}\left(\kappa t_{J} / 2\right)^{-\alpha_{J}} \\
& \leq H^{r / 2} \sum_{J=L}^{H} \frac{2^{J(r / 2-1)}}{m_{J}^{r / 2}}\left(\kappa t_{J} / 2\right)^{-\alpha_{J}} \sum_{k \in \mathbb{Z}}\left|\beta_{J, k}\right|^{\alpha_{J}} \\
& =H^{r / 2} \sum_{J=L}^{H} 2^{\alpha_{J}} \kappa^{-\alpha_{J}} J^{-\frac{\alpha_{J}}{2}} \frac{2^{J(r / 2-1)}}{m_{J}^{\left(r-\alpha_{J}\right) / 2}} \sum_{k \in \mathbb{Z}}\left|\beta_{J, k}\right|^{\alpha_{J}} \\
& \leq H^{r / 2} \sum_{J=L}^{H} \frac{2^{J(r / 2-1)}}{m_{J}^{\left(r-\alpha_{J}\right) / 2}} \sum_{k \in \mathbb{Z}}\left|\beta_{J, k}\right|^{\alpha_{J}} \tag{26}
\end{align*}
$$

the last inequality using $\kappa / 2 \geq 1$ and $J \geq 1$ to simplify the expression a little. For now, we ignore the factor $H^{r / 2}$ (we will bring it back at the end), and look at two cases:

- If $p>\frac{r}{s+1}$, we continue by writing

$$
E_{b b} \lesssim \sum_{J=L}^{H} \frac{2^{J(r / 2-1)}}{m_{J}^{\left(r-\alpha_{J}\right) / 2}} 2^{-J \alpha_{J}\left(s+\frac{1}{2}-\frac{1}{\alpha_{J}}\right)} \asymp \sum_{J=L}^{H} m_{J}^{-\frac{r-\alpha_{J}}{2}} 2^{\frac{1}{2} J\left(r-(2 s+1) \alpha_{J}\right)}=\sum_{J=L}^{H} 2^{J \frac{r}{2}} m_{J}^{-\frac{r}{2}} \cdot 2^{\frac{\alpha_{J}}{2}\left(\log m_{J}-J(2 s+1)\right)}
$$

where the first inequality, which holds for any $\alpha_{J} \in[0, p]$, uses the following bound on $\sum_{k}\left|\beta_{J, k}\right|^{p}$ : For any $\alpha \in[0, p]$, we have from Fact A. 3 and Hölder's inequality along with
the sparsity of coefficients (Fact A.2), that

$$
\sum_{k \in \mathbb{Z}}\left|\beta_{J, k}\right|^{\alpha}=\sum_{k \in \mathbb{Z}}\left|\beta_{J, k}\right|^{\alpha} \mathbb{1}_{\left\{\beta_{J, k} \neq 0\right\}} \leq\left(\sum_{k \in \mathbb{Z}}\left|\beta_{J, k}\right|^{p}\right)^{\frac{\alpha}{p}}\left|\mathcal{B}_{J}\right|^{1-\frac{\alpha}{p}} \leq C^{\frac{\alpha}{p}}(2 A+1)^{1-\frac{\alpha}{p}} \cdot 2^{-J \alpha\left(s+\frac{1}{2}-\frac{1}{p}\right)} \cdot 2^{J\left(1-\frac{\alpha}{p}\right)}
$$

so that $\sum_{k \in \mathbb{Z}}\left|\beta_{J, k}\right|^{\alpha} \lesssim 2^{-J \alpha\left(s+\frac{1}{2}-\frac{1}{\alpha}\right)}$, as in [2, Section C.2.3]. To bound the resulting sum, we need to choose $\alpha_{J} \in[0, p]$ for all $J$ in order to minimize the result. Since $m_{J} \propto 2^{J-\ell} \wedge 1$, the quantity

$$
\log m_{J}-J(2 s+1)
$$

is decreasing in $J$, and thus becomes negative at some value $M$ (for simplicity, assumed to be an integer), such that ${ }^{1}$

$$
\begin{equation*}
2^{M} \asymp\left(\frac{n 2^{\ell}}{H}\right)^{\frac{1}{2 s+2}} \wedge\left(\frac{n}{H}\right)^{\frac{1}{2 s+1}} \tag{27}
\end{equation*}
$$

we see that we should set $\alpha_{J}:=0$ for $J \leq M$, and for $J>M$ set all $\alpha_{J}$ to some value $\alpha=\alpha(r, s)$ which will balance the remaining terms. With this choice, we can write

$$
\begin{aligned}
\sum_{J=L}^{H} m_{J}^{-\frac{r-\alpha}{2}} 2^{\frac{1}{2} J\left(r-(2 s+1) \alpha_{J}\right)} \leq & \sum_{J=L}^{M} m_{J}^{-\frac{r}{2}} 2^{J \frac{r}{2}}+\sum_{J=M}^{H} m_{J}^{-\frac{r-\alpha}{2}} 2^{\frac{1}{2} J(r-(2 s+1) \alpha)} \\
\leq & \left(\frac{H}{n 2^{\ell}}\right)^{\frac{r}{2}} \sum_{J=1}^{M} 2^{J r}+\left(\frac{H}{n}\right)^{\frac{r}{2}} \sum_{J=1}^{M} 2^{J \frac{r}{2}} \\
& +\left(\frac{H}{n 2^{\ell}}\right)^{\frac{r-\alpha}{2}} \sum_{J=M}^{H} 2^{J(r-(s+1) \alpha)}+\left(\frac{H}{n}\right)^{\frac{r-\alpha}{2}} \sum_{J=M}^{H} 2^{\frac{1}{2} J(r-(2 s+1) \alpha)}
\end{aligned}
$$

recalling for the second inequality that $m_{J}^{-1} \asymp \frac{H}{n}\left(\frac{2^{J}}{2^{\ell}} \vee 1\right)$. We can bound the first and second terms as

$$
\left(\frac{H}{n 2^{\ell}}\right)^{\frac{r}{2}} \sum_{J=1}^{M} 2^{J r} \leq \frac{2^{r}}{2^{r}-1}\left(\frac{H}{n 2^{\ell}}\right)^{\frac{r}{2}} 2^{r M}, \quad\left(\frac{H}{n}\right)^{\frac{r}{2}} \sum_{J=1}^{M} 2^{J \frac{r}{2}} \leq \frac{2^{r / 2}}{2^{r / 2}-1}\left(\frac{H}{n}\right)^{\frac{r}{2}} 2^{\frac{r}{2} M}
$$

and from (27) we get that their sum is then

$$
\left(\frac{H}{n 2^{\ell}}\right)^{\frac{r}{2}} \sum_{J=1}^{M} 2^{J r}+\left(\frac{H}{n}\right)^{\frac{r}{2}} \sum_{J=1}^{M} 2^{J \frac{r}{2}} \lesssim\left(\frac{H}{n 2^{\ell}}\right)^{\frac{r}{2}} 2^{r M} \vee\left(\frac{H}{n}\right)^{\frac{r}{2}} 2^{\frac{r}{2} M} \lesssim\left(\frac{H}{n 2^{\ell}}\right)^{\frac{r s}{2 s+2}} \vee\left(\frac{H}{n}\right)^{\frac{r s}{2 s+1}} .
$$

Thus, it only remains to handle the third and fourth terms by choosing a suitable value for $\alpha$. Recalling that we are in the case $p>\frac{r}{s+1}$, we pick any $\frac{r}{s+1}<\alpha \leq p$; for instance, $\alpha:=p$. Since $r-(s+1) p<0$ we then have

$$
\left(\frac{H}{n 2^{\ell}}\right)^{\frac{r-p}{2}} \sum_{J=M}^{H} 2^{J(r-(s+1) p)} \leq\left(\frac{H}{n 2^{\ell}}\right)^{\frac{r-p}{2}} \frac{2^{M(r-(s+1) p)}}{1-2^{r-(s+1) p}} \asymp\left(\frac{H}{n 2^{\ell}}\right)^{\frac{r-p}{2}} 2^{M(r-(s+1) p)} ;
$$

[^0]from our setting of $m_{J}$. Finding the value of $J$ for which $\log \frac{n}{H}-(J-\ell)_{+}-J(2 s+1)$ cancels gives the claimed relation.
note that $\frac{1}{1-2^{r-(s+1) p}}>0$ is a constant, depending only on $r, s, p$. Similarly, $r-(2 s+1) p<0$, and so
$$
\left(\frac{H}{n}\right)^{\frac{r-p}{2}} \sum_{J=M}^{H} 2^{\frac{1}{2} J(r-(2 s+1) p)} \leq\left(\frac{H}{n}\right)^{\frac{r-p}{2}} \frac{2^{\frac{1}{2} M(r-(2 s+1) p)}}{1-2^{\frac{1}{2}(r-(2 s+1) p)}} \asymp\left(\frac{H}{n}\right)^{\frac{r-p}{2}} 2^{\frac{1}{2} M(r-(2 s+1) p)} .
$$

From the setting of $M$ from (27), by a distinction of cases we again can bound their sum as

$$
\left(\frac{H}{n 2^{\ell}}\right)^{\frac{r-p}{2}} \sum_{J=M}^{H} 2^{J(r-(s+1) p)}+\left(\frac{H}{n}\right)^{\frac{r-p}{2}} \sum_{J=M}^{H} 2^{\frac{1}{2} J(r-(2 s+1) p)} \lesssim\left(\frac{H}{n 2^{\ell}}\right)^{\frac{r s}{2 s+2}} \vee\left(\frac{H}{n}\right)^{\frac{r s}{2 s+1}}
$$

Therefore, overall, in the case $p>\frac{r}{s+1}$ we have (bringing back the factor $H^{r / 2}$ we had ignored earlier)

$$
\begin{equation*}
E_{b b} \lesssim H^{r / 2}\left(\frac{H}{n 2^{\ell}}\right)^{\frac{r s}{2 s+2}} \vee H^{r / 2}\left(\frac{H}{n}\right)^{\frac{r s}{2 s+1}} \tag{28}
\end{equation*}
$$

- If $p \leq \frac{r}{s+1}$, we will choose $\alpha_{J} \geq p$ for all $J$. Under this constraint, we can use the monotonicity of $\ell_{p}$ norms (for every $x,\|x\|_{p} \leq\|x\|_{q}$ if $p \geq q$ ) to write

$$
\begin{align*}
E_{b b} & \lesssim \sum_{J=L}^{H} \frac{2^{J(r / 2-1)}}{m_{J}^{\left(r-\alpha_{J}\right) / 2}} \sum_{k \in \mathbb{Z}}\left|\beta_{J, k}\right|^{\alpha_{J}} \leq \sum_{J=L}^{H} \frac{2^{J(r / 2-1)}}{m_{J}^{\left(r-\alpha_{J}\right) / 2}}\left(\sum_{k \in \mathbb{Z}}\left|\beta_{J, k}\right|^{p}\right)^{\alpha_{J} / p} \\
& \lesssim \sum_{J=L}^{H} \frac{2^{J(r / 2-1)}}{m_{J}^{\left(r-\alpha_{J}\right) / 2}} 2^{-J \alpha_{J}\left(s+\frac{1}{2}-\frac{1}{p}\right)}  \tag{FactA.3}\\
& =\sum_{J=L}^{H} m_{J}^{-\frac{r-\alpha_{J}}{2}} 2^{J\left(\frac{r}{2}-1-\alpha_{J}\left(s+\frac{1}{2}-\frac{1}{p}\right)\right)}
\end{align*}
$$

As before, one can see that for there exists some $M$ such that the best choice is to set $\alpha_{J}=p$ for $J \leq M$ (as small as possible given our constraint $\alpha_{J} \geq p$ ). Moreover, proceeding as in the previous case, ${ }^{2}$ we can see that this $M$ is such that

$$
\begin{equation*}
2^{M} \asymp\left(\frac{n 2^{\ell}}{H}\right)^{\frac{1}{2(s-1 / p)+2}} \wedge\left(\frac{n}{H}\right)^{\frac{1}{2(s-1 / p)+1}} \tag{29}
\end{equation*}
$$

This part of the sum will then contribute

$$
\begin{aligned}
\sum_{J=L}^{M} m_{J}^{-\frac{r-p}{2}} 2^{J\left(\frac{r}{2}-1-p\left(s+\frac{1}{2}-\frac{1}{p}\right)\right)} & \asymp\left(\frac{H}{n 2^{\ell}}\right)^{\frac{r-p}{2}} \sum_{J=L}^{M} 2^{J\left(r-1-p\left(s+1-\frac{1}{p}\right)\right)}+\left(\frac{H}{n}\right)^{\frac{r-p}{2}} \sum_{J=L}^{M} 2^{J\left(\frac{r}{2}-1-p\left(s+\frac{1}{2}-\frac{1}{p}\right)\right)} \\
& \asymp\left(\frac{H}{n 2^{\ell}}\right)^{\frac{r-p}{2}} 2^{M(r-p(s+1))}+\left(\frac{H}{n}\right)^{\frac{r-p}{2}} 2^{M\left(\frac{r}{2}-1-p\left(s+\frac{1}{2}-\frac{1}{p}\right)\right)} \\
& \asymp\left(\frac{H}{n 2^{\ell}}\right)^{\frac{r(s-1 / p+1 / r)}{2(s-1 / p)+2}} \vee\left(\frac{H}{n}\right)^{\frac{r(s-1 / p+1 / r)}{2(s-1 / p)+1}}
\end{aligned}
$$

[^1]For $J>M$, we choose an arbitrary constant $\alpha \geq p$ such that $\alpha>\frac{r-1}{s+1-1 / p}$ (so that $r-1-\alpha(s+1-1 / p)<0$ ), and set $\alpha_{J}=\alpha$ for all $J>M .{ }^{3}$ Observe that this implies $\alpha>\frac{r / 2-1}{s+1 / 2-1 / p}$. This part of the sum will then contribute at most

$$
\begin{aligned}
\sum_{J=M}^{H} m_{J}^{-\frac{r-\alpha}{2}} 2^{J\left(\frac{r}{2}-1-\alpha\left(s+\frac{1}{2}-\frac{1}{p}\right)\right)} & \asymp\left(\frac{H}{n 2^{\ell}}\right)^{\frac{r-\alpha}{2}} \sum_{J=M}^{H} 2^{J\left(r-1-\alpha\left(s+1-\frac{1}{p}\right)\right)}+\left(\frac{H}{n}\right)^{\frac{r-\alpha}{2}} \sum_{J=M}^{H} 2^{J\left(\frac{r}{2}-1-\alpha\left(s+\frac{1}{2}-\frac{1}{p}\right)\right)} \\
& \asymp\left(\frac{H}{n 2^{\ell}}\right)^{\frac{r-\alpha}{2}} 2^{M\left(r-1-\alpha\left(s+1-\frac{1}{p}\right)\right)}+\left(\frac{H}{n}\right)^{\frac{r-\alpha}{2}} 2^{M\left(\frac{r}{2}-1-\alpha\left(s+\frac{1}{2}-\frac{1}{p}\right)\right)} \\
& \asymp\left(\frac{H}{n 2^{\ell}}\right)^{-\frac{r(s-1 / p+1 / r)}{2(s-1 / p)+2}} \vee\left(\frac{H}{n}\right)^{\frac{r(s-1 / p+1 / r)}{2(s-1 / p)+1}}
\end{aligned}
$$

as well. Thus, overall, in the case $p \leq \frac{r}{s+1}$ we have (bringing back the factor $H^{r / 2}$ we had ignored earlier)

$$
\begin{equation*}
E_{b b} \lesssim H^{r / 2}\left(\frac{H}{n 2^{\ell}}\right)^{\frac{r(s-1 / p+1 / r)}{2(s-1 / p)+2}} \vee H^{r / 2}\left(\frac{H}{n}\right)^{\frac{r(s-1 / p+1 / r)}{2(s-1 / p)+1}} \tag{30}
\end{equation*}
$$

The term $E_{s b}$. To handle the term $E_{s b}$, we will rely on the fact that, for any $r \geq p$, we have the inclusion $\mathcal{B}(p, q, s) \subseteq \mathcal{B}\left(r, q, s^{\prime}\right)$, for $s^{\prime}=s-\left(\frac{1}{p}-\frac{1}{r}\right)$. This will let us use Fact A. 3 on $\sum_{k \in \mathbb{Z}}\left|\beta_{J, k}\right|^{r}:$

$$
\begin{align*}
E_{s b} & \lesssim H^{r / 2} \sum_{J=L}^{H} 2^{J\left(\frac{r}{2}-1\right)} \sum_{k \in \mathbb{Z}} \mathbb{E}\left[\left|\beta_{J, k}\right|^{r} \mathbb{1}_{k \in \mathcal{I}_{J}^{b}} \widehat{\mathcal{I}}_{J}\right] \quad \text { (Fact A.7) }  \tag{FactA.7}\\
& \lesssim H^{r / 2} \sum_{J=L}^{H} 2^{J\left(\frac{r}{2}-1\right)} \sum_{k \in \mathbb{Z}}\left|\beta_{J, k}\right|^{r} \operatorname{Pr}\left(\left|\widehat{\beta}_{J, k}-\beta_{J, k}\right|>\kappa t_{J}\right) \\
& \lesssim H^{r / 2} \sum_{J=L}^{H} 2^{J\left(\frac{r}{2}-1\right)} \sum_{k \in \mathbb{Z}}\left|\beta_{J, k}\right|^{r} 2^{-\kappa J}  \tag{LemmaC.1}\\
& \lesssim H^{r / 2} \sum_{J=L}^{H} 2^{J\left(\frac{r}{2}-1-\kappa\right)} 2^{-J r\left(s^{\prime}+\frac{1}{2}-\frac{1}{r}\right)} \quad \text { (Lemma C.1) }  \tag{FactA.3}\\
& =H^{r / 2} \sum_{J=L}^{H} 2^{-J\left(r s^{\prime}+\kappa\right)} \leq H^{r / 2} \frac{2^{-L\left(r s^{\prime}+\kappa\right)}}{1-2^{-\left(r s^{\prime}+\kappa\right)}} \\
& \lesssim H^{r / 2} 2^{-L r(N+1)} \lesssim H^{r / 2}\left(n 2^{\ell}\right)^{-\frac{r(N+1)}{2(N+1)+2}} \vee H^{r / 2} n^{-\frac{r(N+1)}{2(N+1)+1}} \leq H^{r / 2}\left(n 2^{\ell}\right)^{-\frac{r s}{2 s+2}} \vee H^{r / 2} n^{-\frac{r s}{2 s+1}} \tag{31}
\end{align*}
$$

where for the third-to-last inequality we relied on our choice of $\kappa \geq r(N+1)$, and for the second-to-last, on our setting of $L$.

The term $E_{s s}$. Finally, we bound the last error term for details $(f), E_{s s}$. In view of proceeding as for $E_{b b}$, for any nonnegative sequence $\left(\alpha_{J}\right)_{J}$ with $0 \leq \alpha_{J} \leq r$, we can write

$$
\begin{equation*}
E_{s s} \lesssim H^{r / 2} \sum_{J=L}^{H} 2^{J\left(\frac{r}{2}-1\right)} \sum_{k \in \mathbb{Z}} \mathbb{E}\left[\left|\beta_{J, k}\right|^{r} \mathbb{1}_{k \in \mathcal{I}_{J}^{s} \backslash \widehat{\mathcal{I}}_{J}}\right] \tag{FactA.7}
\end{equation*}
$$

[^2]\[

$$
\begin{aligned}
& \leq H^{r / 2} \sum_{J=L}^{H} 2^{J\left(\frac{r}{2}-1\right)} \sum_{k \in \mathbb{Z}}\left|\beta_{J, k}\right|^{r} \mathbb{1}_{k \in \mathcal{I}_{J}^{s}} \\
& \leq H^{r / 2} \sum_{J=L}^{H} 2^{J\left(\frac{r}{2}-1\right)} \sum_{k \in \mathbb{Z}}\left|\beta_{J, k}\right|^{\alpha_{J}}\left(\kappa t_{J} / 2\right)^{r-\alpha_{J}} \mathbb{1}_{k \in \mathcal{I}_{J}^{s}} \\
& \leq H^{r / 2}(\kappa / 2)^{r} \sum_{J=L}^{H} 2^{J\left(\frac{r}{2}-1\right)} J^{\frac{r-\alpha_{J}}{2}} m_{J}^{-\left(r-\alpha_{J}\right) / 2} \sum_{k \in \mathbb{Z}}\left|\beta_{J, k}\right|^{\alpha_{J}} \\
& \leq(\kappa / 2)^{r} H^{r} \sum_{J=L}^{H} \frac{2^{J(r / 2-1)}}{m_{J}^{\left(r-\alpha_{J}\right) / 2}} \sum_{k \in \mathbb{Z}}\left|\beta_{J, k}\right|^{\alpha_{J}}
\end{aligned}
$$
\]

which is, except for the extra factor of $(\kappa / 2)^{r} H^{\frac{r}{2}}$, exactly the same expression as (26). We can thus continue the analysis of $E_{s s}$ the same way as we did $E_{b b}$, noting that since $r \geq p$ all the choices for $\alpha_{J}$ in that analysis are still possible; leading to the bound:

$$
E_{s s} \lesssim \begin{cases}H^{r} \cdot\left(\left(\frac{H}{n 2^{\ell}}\right)^{\frac{r s}{2 s+2}} \vee\left(\frac{H}{n}\right)^{\frac{r s}{2 s+1}}\right) & p>\frac{r}{s+1}  \tag{32}\\ H^{r} \cdot\left(\left(\frac{H}{n 2^{\ell}}\right)^{\frac{r(s-1 / p+1 / r)}{2(s-1 / p)+2}} \vee\left(\frac{H}{n}\right)^{\frac{r(s-1 / p+1 / r)}{2(s-1 / p)+1}}\right) & p \leq \frac{r}{s+1}\end{cases}
$$

## C.3.3 Total error

Defining, for $r \geq 1, s \geq 0$, and $p \geq 1$, the quantities

$$
\nu(r, p, s):=\frac{r s}{2 s+2} \mathbb{1}_{p>\frac{r}{s+1}}+\frac{r(s-1 / p+1 / r)}{2(s-1 / p)+2} \mathbb{1}_{p \leq \frac{r}{s+1}}
$$

and

$$
\mu(r, p, s):=\frac{r s}{2 s+1} \mathbb{1}_{p>\frac{r}{s+1}}+\frac{r(s-1 / p+1 / r)}{2(s-1 / p)+1} \mathbb{1}_{p \leq \frac{r}{s+1}}
$$

we can gather all the error terms from Eqs. (23), (24), (25), (28), (30), (31) and (32), to get
$\mathbb{E}\left[\|f-\widehat{f}\|_{r}^{r}\right] \lesssim H^{\kappa}\left(\left(n 2^{\ell}\right)^{-\frac{r(s-1 / p+1 / r)}{2}} \vee n^{-r(s-1 / p+1 / r)}+\left(n 2^{\ell}\right)^{-\frac{r s}{2 s+2}} \vee n^{-\frac{r s}{2 s+1}}+\left(n 2^{\ell}\right)^{-\nu(r, p, s)} \vee n^{-\mu(r, p, s)}\right)$
where $\kappa=\kappa(s, r, p)$ is a constant obtained for simplicity by taking the maximum of the exponent of $H$ in the previous bounds. To simplify this expression, we observe that the following holds for all $p, r, s \geq 1$ :

- $\nu(r, p, s) \leq \frac{r(s-1 / p+1 / r)}{2}$
- $\nu(r, p, s) \leq \frac{r s}{2 s+2}$
- $\mu(r, p, s) \leq r\left(s-\frac{1}{p}+\frac{1}{r}\right)$
- $\mu(r, p, s)>\frac{r s}{2 s+1}$ if, and only if, $r \in((s+1) p,(2 s+1) p)$ (and of course $\mu(r, p, s)=\frac{r s}{2 s+1}$ if $r \leq(s+1) p)$
(this follows from somewhat tedious algebraic manipulations and distinctions of cases). Given the above, we finally get the following bound (where we loosened the bound on the exponent of $H$ to make the result simpler to state):

$$
\mathbb{E}\left[\|f-\widehat{f}\|_{r}^{r}\right] \lesssim H^{\kappa}\left(\left(n 2^{\ell}\right)^{-\nu(r, p, s)} \vee n^{-\mu(r, p, s)} \vee n^{-\frac{r s}{2 s+1}}\right)
$$

$$
=\log ^{\kappa} n \cdot \begin{cases}\left(n 2^{\ell}\right)^{-\frac{r s}{2 s+2}} \vee n^{-\frac{r s}{2 s+1}} & \text { if } r<(s+1) p  \tag{33}\\ \left(n 2^{\ell}\right)^{-\frac{r(s-1 / p+1 / r)}{2(s-1 / p)+2}} \vee n^{-\frac{r s}{2 s+1}} & \text { if }(s+1) p \leq r<(2 s+1) p \\ \left(n 2^{\ell}\right)^{-\frac{r(s-1 / p+1 / r)}{2(s-1 / p)+2}} \vee n^{-\frac{r(s-1 / p+1 / r)}{2(s-1 / p)+1}} & \text { if } r \geq(2 s+1) p\end{cases}
$$

which proves Theorem 1.3 in the main paper.

## D Lower Bounds

Our lower bound construction will depend on whether $r<(s+1) p$ or $r \geq(s+1) p$. Before delving into these cases, we first (a) recall the result from [1] that we will use to upperbound average discrepancy; (b) discuss how the consideration of binary hypothesis testing problem gives a lower bound on average discrepancy.

## D. 1 Upper bound on average discrepancy

Consider the following assumptions on $\mathcal{P}=\left\{\mathbf{p}_{z}: z \in\{-1,1\}^{d}\right\}$, where $\mathbf{p}_{z}$ 's are probability distributions on $[0,1]$.
Assumption D. 1 (Densities exist). For every $z \in\{-1,1\}^{d}$ and $i \in \llbracket d \rrbracket$, there exist functions $\phi_{z, i}:[0,1] \rightarrow \mathbb{R}$ such that $\mathbb{E}_{\mathbf{p}_{z}}\left[\phi_{z, i}^{2}\right]=1$ and

$$
\frac{d \mathbf{p}_{z} \oplus i}{d \mathbf{p}_{z}}=1+\alpha \phi_{z, i}
$$

where $\alpha \in \mathbb{R}$ is a fixed constant independent of $z, i$.
Assumption D. 2 (Orthonormality). For all $z \in\{-1,1\}^{d}$ and $i, j \in \llbracket d \rrbracket, \mathbb{E}_{\mathbf{p}_{z}}\left[\phi_{z, i} \phi_{z, j}\right]=\mathbb{1}_{\{i=j\}}$.
Theorem D. 3 (corollary 2 in [1]). Suppose $\mathcal{P}$ satisfies Assumptions D. 1 and D.2. For some $\tau \in(0,1 / 2]$, let $\pi$ be a prior on $Z \in\{-1,1\}^{d}$ defined as $Z_{i} \sim \operatorname{Rademacher}(\tau)$ independently for each $i \in \llbracket d \rrbracket$. For $Z \sim \pi$, let $X_{1}, \ldots, X_{n}$ be i.i.d. samples from $\mathbf{p}_{Z}$. Then, for any interactive protocol generating $\ell$-bit messages $Y_{1}, \ldots, Y_{n}$, we have

$$
\left(\frac{1}{d} \sum_{i=1}^{d} \mathrm{~d}_{\mathrm{TV}}\left(\mathbf{p}_{-i}^{Y^{n}}, \mathbf{p}_{+i}^{Y_{n}^{n}}\right)\right)^{2} \leq \frac{7}{d} n \alpha^{2} 2^{\ell}
$$

## D. 2 Lower bound on average discrepancy

For $Z \sim \pi$, let $X_{1}, \ldots, X_{n}$ be i.i.d. samples from $\mathbf{p}_{Z}$ distributed across $n$ players, and let $Y_{1}, \ldots, Y_{n}$ be $\ell$-bit messages sent by the players (possibly interactively) to the referee. Based on the $\ell$-bit messages $Y_{1}, \ldots, Y_{n}$, suppose the referee outputs an estimate $\hat{Z}=\left(\hat{Z}_{1}, \ldots, \hat{Z}_{d}\right)$ of $Z=\left(Z_{1}, \ldots, Z_{d}\right)$. Then, an upper bound on $\sum_{i=1}^{d} \operatorname{Pr}\left\{\hat{Z}_{i} \neq Z_{i}\right\}$ gives a lower bound on $\sum_{i=1}^{d} \mathrm{D}\left(\mathbf{p}_{-i}^{Y_{n}^{n}} \| \mathbf{p}_{+i}^{Y^{n}}\right)$. To see this, note that, for a given $i \in \llbracket d \rrbracket$,

$$
\begin{aligned}
& \operatorname{Pr}\left\{\hat{Z}_{i} \neq Z_{i}\right\}=\operatorname{Pr}\left\{\hat{Z}_{i}=-1 \mid Z_{i}=1\right\} \operatorname{Pr}\left\{Z_{i}=1\right\}+\operatorname{Pr}\left\{\hat{Z}_{i}=1 \mid Z_{i}=-1\right\} \operatorname{Pr}\left\{Z_{i}=-1\right\} \\
&=\tau\left(1-\operatorname{Pr}\left\{\hat{Z}_{i}=1 \mid Z_{i}=1\right\}\right)+(1-\tau) \operatorname{Pr}\left\{\hat{Z}_{i}=1 \mid Z_{i}=-1\right\} \\
& \geq \tau\left(1-\operatorname{Pr}\left\{\hat{Z}_{i}=1 \mid Z_{i}=1\right\}\right)+\tau \operatorname{Pr}\left\{\hat{Z}_{i}=1 \mid Z_{i}=-1\right\} \\
&\quad \text { (since }(1-\tau) \geq \tau \text { for } \tau \leq 1 / 2) \\
&=\tau\left(1-\left(\operatorname{Pr}\left\{\hat{Z}_{i}=1 \mid Z_{i}=1\right\}-\operatorname{Pr}\left\{\hat{Z}_{i}=1 \mid Z_{i}=-1\right\}\right)\right) \\
& \geq \tau\left(1-\mathrm{d}_{\mathrm{TV}}\left(\mathbf{p}_{+i}^{Y_{i}^{n}}, \mathbf{p}_{-i}^{Y_{-i}^{n}}\right)\right) .
\end{aligned}
$$

Thus,

$$
\sum_{i=1}^{d} \operatorname{Pr}\left\{\hat{Z}_{i} \neq Z_{i}\right\} \geq \tau\left(d-\sum_{i=1}^{d} \mathrm{~d}_{\mathrm{TV}}\left(\mathbf{p}_{+i}^{\left.\left.Y_{+i}^{n}, \mathbf{p}_{-i}^{Y^{n}}\right)\right), ~}\right.\right.
$$

which gives

$$
\begin{equation*}
\frac{1}{d} \sum_{i=1}^{d} \mathrm{~d}_{\mathrm{TV}}\left(\mathbf{p}_{+i}^{Y n}, \mathbf{p}_{-i}^{Y_{n}^{n}}\right) \geq 1-\frac{1}{d \tau} \sum_{i=1}^{d} \operatorname{Pr}\left\{\hat{Z}_{i} \neq Z_{i}\right\} . \tag{34}
\end{equation*}
$$

In conclusion, to get a lower bound on average discrepancy, it suffices to upperbound $\sum_{i=1}^{d} \operatorname{Pr}\left\{\hat{Z}_{i} \neq Z_{i}\right\}$ for an estimator $\hat{Z}$ of $Z$.

## D. 3 Lower bound on $\mathcal{L}_{r}^{*}(n, \ell, p, q, s)$ for $r<(s+1) p$

Construction. The family of distributions $\mathcal{P}_{1}$ that we will use to derive lower bound when $r<(s+1) p$ has also been used in deriving lower bounds in the unconstrained setting [4,5] and in the LDP setting [2].

Let $g_{0}$ be a density function (see [5, p.157]) such that

1. $\operatorname{supp}\left(g_{0}\right) \subseteq[0,1] ;$
2. $\left\|g_{0}\right\|_{p q s} \leq 1 / 2$;
3. $g_{0} \equiv c_{0}>0$ on some interval $[a, b] \subseteq[0,1]$.

In what follows, $j$ is a free parameter that will be suitably chosen later in the proof. Let $\psi_{j, k}$ be defined as $\psi_{j, k}(x)=2^{j / 2} \psi\left(2^{j} x-k\right)$, where $\psi$ is the mother wavelet used to define $\|\cdot\|_{p q s}$ (see Section A). It is a fact that $\int \psi_{j, k}(x) d x=0$ for every $j, k[5]$.
For a given $z \in\{-1,1\}^{d}$, define

$$
\begin{equation*}
f_{z}:=g_{0}+\gamma \sum_{k \in \mathcal{I}_{j}} z_{k} \psi_{j, k} \tag{35}
\end{equation*}
$$

where

- $\mathcal{I}_{j}$ is the set of indices $k \in \mathbb{Z}$ such that
i. $\operatorname{supp}\left(\psi_{j, k}\right) \subseteq[a, b]$ for every $k \in \mathcal{I}_{j}$;
ii. for $k, k^{\prime} \in \mathcal{I}_{j}, k \neq k^{\prime}, \psi_{j, k}$ and $\psi_{j k^{\prime}}$ have disjoint support;
iii. $d:=\left|\mathcal{I}_{j}\right|=C 2^{j}$, for a constant $C$. Here on, we will assume for simplicity that $d=2^{j}$.
- $\gamma$ is chosen such that
i. $f_{z}(x) \geq c_{0} / 2$ for every $x \in[a, b]$; this condition is satisfied if $c_{0}-\gamma 2^{j / 2}\|\psi\|_{\infty} \geq c_{0} / 2$, i.e., $\gamma \leq\left(c_{0} / 2\|\psi\|_{\infty}\right) 2^{-j / 2}$.
ii. $\left\|f_{z}\right\|_{p q s} \leq 1$; since $\left\|f_{z}\right\|_{p q s} \leq\left\|g_{0}\right\|_{p q s}+\gamma\left\|\psi_{j, k}\right\|_{p q s} \leq 1 / 2+\gamma C 2^{j / p} 2^{j(s+1 / 2-1 / p)}$ (see pg. 160 in [5]), we get that $\left\|f_{z}\right\|_{p q s} \leq 1$ if $\gamma \leq(1 / 2 C) 2^{-j(s+1 / 2)}$.

Since $s>1 / p>0$, we get that for $j$ large enough, if $\gamma$ satisfies condition (ii), it automatically satisfies condition (i). Thus, we choose $\gamma=C 2^{-j(s+1 / 2)}$ for some constant $C$.

Finally, we define the family of distributions as

$$
\begin{equation*}
\mathcal{P}_{1}=\left\{\mathbf{p}_{z}: \mathbf{p}_{z} \text { has density } f_{z}=g_{0}+\gamma \sum_{k \in \mathcal{I}_{j}} z_{k} \psi_{j, k}, z \in\{-1,1\}^{d}\right\} . \tag{36}
\end{equation*}
$$

Prior on $Z$. We assume a uniform prior on $Z \in\{-1,1\}^{d}$, i.e., $Z_{i} \sim$ Rademacher ( $1 / 2$ ) independently for each $i \in \llbracket d \rrbracket$.

Upper bound on average discrepancy. To upperbound average discrepancy, we verify that $\mathcal{P}_{1}$ satisfies the three assumptions described in Section D.1, and then use Theorem D.3. For any $z \in\{-1,1\}^{d}, k \in \llbracket d \rrbracket$, we have

$$
\frac{d \mathbf{p}_{z \oplus k}}{d \mathbf{p}_{z}}(x)=1-\frac{2 \gamma z_{k} \psi_{j, k}(x)}{c_{0}+\gamma z_{k} \psi_{j, k}(x)} .
$$

Since $\operatorname{supp}\left(\psi_{j, k}\right) \cap \operatorname{supp}\left(\psi_{j, k^{\prime}}\right)$ is empty for $k \neq k^{\prime}$, it follows that Assumptions D. 1 and D. 2 hold. We now compute an upper bound on $\alpha^{2}:=\mathbb{E}_{\mathbf{p}_{z}}\left[\left(\frac{\gamma z_{k} \psi_{j, k}(X)}{c_{0}+\gamma z_{k} \psi_{j, k}(X)}\right)^{2}\right]$.

$$
\begin{aligned}
\mathbb{E}_{\mathbf{p}_{z}}\left[\left(\frac{2 \gamma z_{k} \psi_{j, k}(X)}{c_{0}+\gamma z_{k} \psi_{j, k}(X)}\right)^{2}\right] & =4 \gamma^{2} \int_{\operatorname{supp}\left(\psi_{j, k}\right)} \frac{\psi_{j, k}(x)^{2}\left(c_{0}+\gamma z_{k} \psi_{j, k}(x)\right)}{\left(c_{0}+\gamma z_{k} \psi_{j, k}(x)\right)^{2}} d x \\
& =4 \gamma^{2} \int_{\operatorname{supp}\left(\psi_{j, k}\right)} \frac{\psi_{j, k}(x)^{2}}{c_{0}+\gamma z_{k} \psi_{j, k}(x)} d x \\
& \leq 2 \gamma^{2} c_{0} \int_{\operatorname{supp}\left(\psi_{j, k}\right)}^{\psi_{j, k}(x)^{2} d x \quad\left(\text { as } c_{0}+\gamma z_{k} \psi_{j, k}(x) \geq c_{0} / 2\right)} \\
& \leq 2 \gamma^{2} c_{0} \times\left(2^{j / 2}\|\psi\|_{\infty}\right)^{2} \times \text { length }\left(\operatorname{supp}\left(\psi_{j, k}\right)\right) \\
& \leq 2 \gamma^{2} c_{0} \times\left(2^{j / 2}\|\psi\|_{\infty}\right)^{2} \times \frac{C^{\prime \prime}}{2^{j}} \quad \quad\left(\text { for a constant } C^{\prime \prime}>0\right) \\
& =C^{\prime} \gamma^{2} \\
& =C 2^{-j(2 s+1) .} \quad\left(\text { for a constant } C^{\prime}>0\right) \\
& \quad(\text { for a constant } C>0)
\end{aligned}
$$

Thus, using Theorem D.3, we get

$$
\begin{equation*}
\left(\frac{1}{2^{j}} \sum_{k \in \mathcal{I}_{j}} \mathrm{~d}_{\mathrm{TV}}\left(\mathbf{p}_{-k}^{Y^{n}}, \mathbf{p}_{+k}^{Y^{n}}\right)\right)^{2} \lesssim\left(n 2^{\ell}\right) 2^{-2 j(s+1)} . \tag{37}
\end{equation*}
$$

Lower bound on average discrepancy. To lower bound the average discrepancy, we will use the idea described in Section D.2. Consider a communication-constrained density estimation algorithm (possibly interactive) that outputs $\hat{f}$ satisfying $\sup _{f \in \mathcal{B}(p, q, s)} \mathbb{E}_{f}\left[\|\hat{f}-f\|_{r}^{r}\right] \leq \varepsilon^{r}$. Using this density estimator, we estimate $\hat{Z}$ as

$$
\hat{Z}=\underset{z}{\operatorname{argmin}}\left\|f_{z}-\hat{f}\right\|_{r} .
$$

Then

$$
\begin{equation*}
\mathbb{E}_{\mathbf{p}_{z}}\left[\left\|f_{z}-f_{\hat{Z}}\right\|_{r}^{r}\right] \leq 2^{r-1}\left(\mathbb{E}_{\mathbf{p}_{z}}\left[\left\|f_{z}-\hat{f}\right\|_{r}^{r}\right]+\mathbb{E}_{\mathbf{p}_{z}}\left[\left\|f_{z}-f_{\hat{Z}}\right\|_{r}^{r}\right]\right) \leq 2^{r} \varepsilon^{r} . \tag{38}
\end{equation*}
$$

Now, for $z \neq z^{\prime}$, we have

$$
\begin{aligned}
\left\|f_{z}-f_{z^{\prime}}\right\|_{r}^{r} & =\int_{0}^{1}\left|f_{z}(x)-f_{z^{\prime}}(x)\right|^{r} d x \\
& =\gamma^{r} \int_{0}^{1}\left|\sum_{k \in \mathcal{I}_{j}} \psi_{j, k}(x) \mathbb{1}_{\left\{z_{k} \neq z_{k}^{\prime}\right\}}\right|^{r} d x
\end{aligned}
$$

$$
\begin{array}{lr}
=\gamma^{r} \int_{0}^{1} \sum_{k \in \mathcal{I}_{j}}\left|\psi_{j, k}(x)\right|^{r} \mathbb{1}_{\left\{z_{k} \neq z_{k}^{\prime}\right\}} d x & \text { (since the } \psi_{j, k} \text { 's have disjoint supports) } \\
=\gamma^{r} \sum_{k \in \mathcal{I}_{j}} \int_{\operatorname{supp}\left(\psi_{j, k}\right)}\left|\psi_{j, k}(x)\right|^{r} \mathbb{1}_{\left\{z_{k} \neq z_{k}^{\prime}\right\}} d x & \\
=\gamma^{r}\left(2^{j / 2}\|\psi\|_{\infty}\right)^{r} \frac{C^{\prime \prime}}{2^{j}} \sum_{k \in \mathcal{I}_{j}} \mathbb{1}_{\left\{z_{k} \neq z_{k}^{\prime}\right\}} & \text { (for a constant } C^{\prime \prime}>0 \text { ) } \\
=C^{\prime} \gamma^{r} 2^{j(r / 2-1)} \sum_{k \in \mathcal{I}_{j}} \mathbb{1}_{\left\{z_{k} \neq z_{k}^{\prime}\right\}} & \text { (for a constant } C^{\prime}>0 \text { ) } \\
=C 2^{-j(r s+1)} \sum_{k \in \mathcal{I}_{j}} \mathbb{1}_{\left\{z_{k} \neq z_{k}^{\prime}\right\}} . & \text { (for a constant } C>0 \text { ) }
\end{array}
$$

which gives that, for an estimator $\hat{Z}$,

$$
\mathbb{E}_{\mathbf{p}_{z}}\left[\left\|f_{z}-f_{\hat{Z}}\right\|_{r}^{r}\right]=C 2^{-j(r s+1)} \sum_{k \in \mathcal{I}_{j}} \operatorname{Pr}\left\{Z_{k} \neq \hat{Z}_{k}\right\} .
$$

Combining this with (38), we get

$$
\begin{equation*}
\sum_{k \in \mathcal{I}_{j}} \operatorname{Pr}\left\{\hat{Z}_{k} \neq Z_{k}\right\} \lesssim \varepsilon^{r} 2^{j(r s+1)} . \tag{39}
\end{equation*}
$$

Thus, substituting $d=2^{j}$ and $\tau=1 / 2$ in (34) (and ignoring multiplicative constants), we get

$$
\frac{1}{2^{j}} \sum_{k \in \mathcal{I}_{j}} \mathrm{~d}_{\mathrm{TV}}\left(\mathbf{p}_{-k}^{Y^{n}}, \mathbf{p}_{+k}^{Y_{n}^{n}}\right) \gtrsim 1-\frac{2}{2^{j}} \varepsilon^{r} 2^{j(r s+1)} \simeq 1-\varepsilon^{r} 2^{j r s} .
$$

Now, observe that $j$ is a free parameter that we can choose. If we choose $j$ such that

$$
\begin{equation*}
\varepsilon^{r} 2^{j r s} \simeq 1 \tag{40}
\end{equation*}
$$

then we get

$$
\frac{1}{2^{j}} \sum_{k \in \mathcal{I}_{j}} \mathrm{~d}_{\mathrm{TV}}\left(\mathbf{p}_{-k}^{\left.Y_{-k}^{n}, \mathbf{p}_{+k}^{Y_{n}^{n}}\right) \gtrsim 1.10 .}\right.
$$

or

$$
\begin{equation*}
\left(\frac{1}{2^{j}} \sum_{k \in \mathcal{I}_{j}} \mathrm{~d}_{\mathrm{TV}}\left(\mathbf{p}_{-k}^{Y^{n}}, \mathbf{p}_{+k}^{Y^{n}}\right)\right)^{2} \gtrsim 1 \tag{41}
\end{equation*}
$$

Putting things together. From (37) and (41), we get that, for $j$ satisfying $\varepsilon^{r} 2^{j r s} \simeq 1$,

$$
1 \lesssim\left(n 2^{\ell}\right) 2^{-2 j(s+1)} .
$$

which gives

$$
2^{j} \lesssim\left(n 2^{\ell}\right)^{\frac{1}{2 s+2}} .
$$

Using $\varepsilon^{r} 2^{j r s} \simeq 1$, we finally get

$$
\varepsilon^{r} \gtrsim\left(n 2^{\ell}\right)^{-\frac{r s}{2 s+2}}
$$

which is our desired lower bound.

## D. 4 Lower bound on $\mathcal{L}_{r}^{*}(n, \ell, p, q, s)$ for $r \geq(s+1) p$

Construction. The family of distributions $\mathcal{P}_{2}$ that we will use to derive lower bound when $r \geq(s+1) p$ is not exactly the same as that in the unconstrained and in the LDP setting [4, 5, 2]; but, combined with the prior that we will choose on $Z$, it will essentially mimic that.
Let $g_{0}, \psi_{j, k}, \mathcal{I}_{j}$ be as in Section D.3. For a given $z \in\{-1,1\}^{d}$ (where $\left.d:=\left|\mathcal{I}_{j}\right|\right)$, define

$$
\begin{equation*}
f_{z}:=g_{0}+\gamma \sum_{k \in \mathcal{I}_{j}}\left(1+z_{k}\right) \psi_{j, k} . \tag{42}
\end{equation*}
$$

where we will choose $\gamma$ after we describe the prior on $Z$. Finally, we define the family of distributions as

$$
\begin{equation*}
\mathcal{P}_{2}=\left\{\mathbf{p}_{z}: \mathbf{p}_{z} \text { has density } f_{z}=g_{0}+\gamma \sum_{k \in \mathcal{I}_{j}}\left(1+z_{k}\right) \psi_{j, k}, z \in\{-1,1\}^{d}\right\} \tag{43}
\end{equation*}
$$

Prior on $Z$. We assume a "sparse" prior on $Z \in\{-1,1\}^{d}$, defined as $Z_{k} \sim \operatorname{Rademacher}(1 / d)$ independently for each $k \in \llbracket d \rrbracket$. We call it "sparse" because, with high probability, for $Z=$ $\left(Z_{1}, \ldots, Z_{d}\right)$ sampled from this prior, the number of indices $k$ with $Z_{k}=1$ will be small (we will quantify this soon). Now, since $f_{Z}=g_{0}+\gamma \sum_{k \in \mathcal{I}_{j}}\left(1+Z_{k}\right) \psi_{j, k}$, this means that with high probability $1+Z_{k}=0$ for a large number of $k$ 's, and thus there will be only a few "bumps" in $f_{Z}$.

Choosing $\gamma$. Define $\mathcal{G} \subset\{-1,1\}^{d}$ as

$$
\mathcal{G}:=\left\{z \in\{-1,1\}^{d}: \sum_{k=1}^{d} \mathbb{1}_{\left\{z_{k}=1\right\}} \leq 2 j\right\} .
$$

Then, by Bernstein's inequality

$$
\begin{equation*}
\operatorname{Pr}\{Z \in \mathcal{G}\} \geq 1-4 \cdot 2^{-2 j} \tag{44}
\end{equation*}
$$

We will choose $\gamma$ such that
i. $f_{z}(x) \geq c_{0} / 2$ for every $x \in[a, b]$; as seen in Section D.3, this condition is satisfied if $\gamma \leq\left(c_{0} / 2\|\psi\|_{\infty}\right) 2^{-j / 2}$.
ii. $\left\|f_{z}\right\|_{p q s} \leq 1$ for every $z \in \mathcal{G}$; argument similar to that in Section D. 3 gives that $\left\|f_{z}\right\|_{p q s} \leq 1$ for $z \in \mathcal{G}$ if $\gamma \lesssim 2^{-j(s+1 / 2-1 / p)} j^{-1 / p}$.

Since $s>1 / p$, we get that for $j$ large enough ( $j$ is a free parameter that we choose later), if $\gamma$ satisfies condition (ii), it automatically satisfies condition (i). Thus, we choose $\gamma=$ $C 2^{-j(s+1 / 2-1 / p)} j^{-1 / p}$ for some constant $C$.
Note that, for $z \notin \mathcal{G}$, this choice of $\gamma$ still results in $f_{z}$ being a density function (since $\int \psi_{j, k}(x) d x=$ 0 ), but it may be the case that $\left\|f_{z}\right\|_{s p q}>1$.

Upper bound on average discrepancy. To upperbound average discrepancy, we verify that $\mathcal{P}_{2}$ satisfies the three assumptions described in Section D.1. For any $z \in\{-1,1\}^{d}, k \in \llbracket d \rrbracket$, we have

$$
\begin{aligned}
\frac{d p_{z \oplus k}}{d p_{z}}(x) & =\frac{g_{0}+\gamma \sum_{k \in I_{j}}\left(1-z_{k}\right) \psi_{j, k}(x)}{g_{0}+\gamma \sum_{k \in I_{j}}\left(1+z_{k}\right) \psi_{j, k}(x)} \\
& =1-\frac{2 \gamma z_{k} \psi_{j, k}(x)}{c_{0}+\gamma \sum_{k \in I_{j}} z_{k} \psi_{j, k}(x)}
\end{aligned}
$$

which is same as what we had in Section D.3. Similar arguments lead to the conclusion that Assumptions D. 1 and D. 2 are satisfied. Moreover, an upper bound on $\alpha^{2}:=\mathbb{E}_{\mathbf{p}_{z}}\left[\left(\frac{2 \gamma z_{i} \psi_{j, k}(X)}{c_{0}+\gamma z_{i} \psi_{j, k}(X)}\right)^{2}\right]$ follows similarly (with different value of $\gamma$ ), and we get that

$$
\alpha^{2} \leq C 2^{-2 j(s+1 / 2-1 / p)} j^{-2 / p} . \quad(\text { for a constant } C>0)
$$

Thus, using Theorem D.3, we get

$$
\begin{equation*}
\left(\frac { 1 } { 2 ^ { j } } \sum _ { k \in \mathcal { I } _ { j } } \mathrm { d } _ { \mathrm { TV } } \left(\mathbf{p}_{-k}^{\left.\left.Y_{-k}^{n}, \mathbf{p}_{+k}^{Y_{n}^{n}}\right)\right)^{2} \lesssim\left(n 2^{\ell}\right) 2^{-2 j(s+1-1 / p)} j^{-2 / p} . . . . . . .}\right.\right. \tag{45}
\end{equation*}
$$

Lower bound on average discrepancy. To lowerbound average discrepancy, we proceed as in Section D.3. Consider a communication-constrained density estimation algorithm (possibly interactive) that outputs $\hat{f}$ satisfying $\sup _{f \in \mathcal{B}(p, q, s)} \mathbb{E}_{f}\left[\|\hat{f}-f\|_{r}^{r}\right] \leq \varepsilon^{r}$. Using this density estimator, we estimate $\hat{Z}$ as

$$
\hat{Z}=\underset{z}{\operatorname{argmin}}\left\|f_{z}-\hat{f}\right\|_{r} .
$$

Then, for $z \in \mathcal{G}$,

$$
\begin{equation*}
\mathbb{E}_{\mathbf{p}_{z}}\left[\left\|f_{z}-f_{\hat{Z}}\right\|_{r}^{r}\right] \leq 2^{r} \varepsilon^{r} . \tag{46}
\end{equation*}
$$

This only holds for $z \in \mathcal{G}$ because the estimator's guarantee only holds if samples come from a density $f$ satisfying $\|f\|_{s p q} \leq 1$. Now, for $z \neq z^{\prime}$, plugging in the value of $\gamma$ in the calculation done in Section D.3, we get

$$
\left.\left\|f_{z}-f_{z^{\prime}}\right\|_{r}^{r}=C j^{-r / p} 2^{-j(r(s-1 / p)+1)} \sum_{k \in \mathcal{I}_{j}} \mathbb{1}_{\left\{z_{k} \neq z_{k}^{\prime}\right\}} \quad \quad \text { (for a constant } C>0\right)
$$

which gives that, for any estimator $\hat{Z}$,

$$
\mathbb{E}_{\mathbf{p}_{z}}\left[\left\|f_{z}-f_{\hat{Z}}\right\|_{r}^{r}\right]=C j^{-r / p} 2^{-j(r(s-1 / p)+1)} \sum_{k \in \mathcal{I}_{j}} \operatorname{Pr}\left\{\hat{Z}_{k} \neq Z_{k}\right\} .
$$

Combining this with (46), we get that

$$
\begin{equation*}
\sum_{k \in \mathcal{I}_{j}} \operatorname{Pr}\left\{Z_{k} \neq \hat{Z}_{k}, Z \in \mathcal{G}\right\} \lesssim \varepsilon^{r} j^{r / p} 2^{j(r(s-1 / p)+1)} \tag{47}
\end{equation*}
$$

Thus,

$$
\sum_{k \in \mathcal{I}_{j}} \operatorname{Pr}\left\{\hat{Z}_{k} \neq Z_{k}\right\}=\sum_{k \in \mathcal{I}_{j}} \operatorname{Pr}\left\{Z_{k} \neq \hat{Z}_{k}, Z \in \mathcal{G}\right\}+\sum_{k \in \mathcal{I}_{j}} \operatorname{Pr}\left\{Z_{k} \neq \hat{Z}_{k}, Z \notin \mathcal{G}\right\}
$$

$$
\begin{align*}
& \leq \sum_{k \in \mathcal{I}_{j}} \operatorname{Pr}\left\{Z_{k} \neq \hat{Z}_{k}, Z \in \mathcal{G}\right\}+\left(\sum_{k \in \mathcal{I}_{j}} \operatorname{Pr}\left\{\hat{Z}_{k} \neq Z_{k} \mid Z \notin \mathcal{G}\right\}\right) \operatorname{Pr}\{Z \notin \mathcal{G}\} \\
& \lesssim \varepsilon^{r} j^{r / p_{2}} 2^{j(r(s-1 / p)+1)}+2^{j} 2^{-2 j} . \tag{47}
\end{align*}
$$

Thus, substituting $d=2^{j}$ and $\tau=1 / d=2^{-j}$ in (34) (and ignoring multiplicative constants), we get

$$
\begin{aligned}
\frac{1}{2^{j}} \sum_{k \in \mathcal{I}_{j}} \mathrm{~d}_{\mathrm{TV}}\left(\mathbf{p}_{-k}^{Y_{-k}^{n}}, \mathbf{p}_{+k}^{Y_{k}^{n}}\right) & \gtrsim 1-\varepsilon^{r} j^{r / p} 2^{j(r(s-1 / p)+1)}-2^{-j} \\
& \simeq 1-\varepsilon^{r} j^{r / p} 2^{j r(s-1 / p+1 / r)}
\end{aligned}
$$

Choosing $j$ such that

$$
\begin{equation*}
\varepsilon^{r} 2^{j r(s-1 / p+1 / r)} j^{r / p} \simeq 1 \tag{48}
\end{equation*}
$$

gives

$$
\frac{1}{2^{j}} \sum_{k \in \mathcal{I}_{j}} \mathrm{~d}_{\mathrm{TV}}\left(\mathbf{p}_{-k}^{Y_{k}^{n}}, \mathbf{p}_{+k}^{Y n}\right) \gtrsim 1
$$

or

$$
\begin{equation*}
\left(\frac{1}{2^{j}} \sum_{k \in \mathcal{I}_{j}} \mathrm{~d}_{\mathrm{TV}}\left(\mathbf{p}_{-k}^{Y^{n}}, \mathbf{p}_{+k}^{Y^{n}}\right)\right)^{2} \gtrsim 1 \tag{49}
\end{equation*}
$$

Putting things together. From (45) and (49), we get, for any $j$ satisfying $\varepsilon^{r} 2^{j r(s-1 / p+1 / r)} j^{r / p} \simeq$ 1 , that $1 \lesssim\left(n 2^{\ell}\right) 2^{-2 j(s+1-1 / p)} j^{-2 / p}$. This then yields

$$
\begin{equation*}
2^{2 j(s+1-1 / p)} j^{2 / p} \lesssim n 2^{\ell} \tag{50}
\end{equation*}
$$

To get a rough idea of the bound this will give, let us ignore $j^{2 / p}$ to get,

$$
\begin{equation*}
2^{j} \lesssim\left(n 2^{\ell}\right)^{\frac{1}{2(s+1-1 / p)}} \tag{51}
\end{equation*}
$$

Now, since $\varepsilon^{r} 2^{j r(s-1 / p+1 / r)} j^{r / p} \simeq 1$, we get, roughly, (ignoring $j^{r / p}$ )

$$
\begin{equation*}
2^{j} \simeq(1 / \varepsilon)^{\frac{1}{s-1 / p+1 / r}} \tag{52}
\end{equation*}
$$

Combining (51), (52), we get that (up to logarithmic factors)

$$
\varepsilon^{r} \gtrsim\left(n 2^{\ell}\right)^{-\frac{r(s-1 / p+1 / r)}{2(s+1-1 / p)}}
$$

which is the desired bound, again up to logarithmic factors. We now show how a slightly more careful analysis lets us obtain the tight bound.

Bringing in log factors. From (50), we get that

$$
\begin{equation*}
2^{j} \lesssim\left(n 2^{\ell}\right)^{\frac{1}{2(s+1-1 / p)}}\left(\log \left(n 2^{\ell}\right)\right)^{-\frac{2 / p}{2(s+1-1 / p)}} \tag{53}
\end{equation*}
$$

Now, since $\varepsilon^{r} 2^{j r(s-1 / p+1 / r)} j^{r / p} \simeq 1$, we get

$$
2^{j} \simeq(1 / \varepsilon)^{\frac{1}{s-1 / p+1 / r}}(\log (1 / \varepsilon))^{-\frac{1}{p(s-1 / p+1 / r)}}
$$

Substituting this in (53), we get

$$
(1 / \varepsilon)^{\frac{1}{s-1 / p+1 / r}}(\log (1 / \varepsilon))^{-\frac{1}{p(s-1 / p+1 / r)}} \lesssim\left(n 2^{\ell}\right)^{\frac{1}{2(s+1-1 / p)}}\left(\log \left(n 2^{\ell}\right)\right)^{-\frac{2 / p}{2(s+1-1 / p)}}
$$

or

$$
1 / \varepsilon(\log (1 / \varepsilon))^{-1 / p} \lesssim\left(n 2^{\ell}\right)^{\frac{s-1 / p+1 / r}{2(s+1-1 / p)}}\left(\log \left(n 2^{\ell}\right)\right)^{-(2 / p) \frac{(s-1 / p+1 / r)}{2(s+1-1 / p)}}
$$

This implies that

$$
\begin{aligned}
1 / \varepsilon & \lesssim\left(n 2^{\ell}\right)^{\frac{s-1 / p+1 / r}{2(s+1-1 / p)}}\left(\log \left(n 2^{\ell}\right)\right)^{-(2 / p) \frac{(s-1 / p+1 / r)}{2(s+1-1 / p)}}\left(\log \left(\left(n 2^{\ell}\right)^{\frac{s-1 / p+1 / r}{2(s+1-1 / p)}}\left(\log \left(n 2^{\ell}\right)\right)^{-(2 / p) \frac{(s-1 / p+1 / r)}{2(s+1-1 / p)}}\right)\right)^{1 / p} \\
& \simeq\left(n 2^{\ell}\right)^{\frac{s-1 / p+1 / r}{2(s+1-1 / p)}}\left(\log \left(n 2^{\ell}\right)\right)^{-(2 / p) \frac{(s-1 / p+1 / r)}{2(s+1-1 / p)}+\frac{1}{p}} \\
& =\left(n 2^{\ell}\right)^{\frac{s-1 / p+1 / r}{2(s+1-1 / p)}}\left(\log \left(n 2^{\ell}\right)\right)^{-\frac{1-1 / r}{p(s-1 / p+1)}}
\end{aligned}
$$

Thus

$$
\begin{equation*}
\varepsilon^{r} \gtrsim\left(n 2^{\ell}\right)^{\frac{r(s-1 / p+1 / r)}{2(s+1-1 / p)}}\left(\log \left(n 2^{\ell}\right)\right)^{-\frac{r-1}{p(s-1 / p+1)}} . \tag{54}
\end{equation*}
$$

## D.5 Concluding the proof of Theorem 1.1

Combining lower bounds from Sections D. 3 and D. 4 with lower bounds in the classical setting [4] (where the rate transition happens at $r=(2 s+1) p$ ), we get Theorem 1.1.

## References

[1] Jayadev Acharya, Clément L. Canonne, Ziteng Sun, and Himanshu Tyagi. Unified lower bounds for interactive high-dimensional estimation under information constraints. arXiv preprint arXiv:2010.06562, 2021.
[2] Cristina Butucea, Amandine Dubois, Martin Kroll, and Adrien Saumard. Local differential privacy: elbow effect in optimal density estimation and adaptation over Besov ellipsoids. Bernoulli, 26(3):1727-1764, 2020.
[3] Ingrid Daubechies. Ten lectures on wavelets. SIAM, 1992.
[4] David L. Donoho, Iain M. Johnstone, Gérard Kerkyacharian, and Dominique Picard. Density estimation by wavelet thresholding. Ann. Statist., 24(2):508-539, 1996.
[5] Wolfgang Härdle, Gerard Kerkyacharian, Dominique Picard, and Alexander Tsybakov. Wavelets, approximation, and statistical applications, volume 129. Springer Science \& Business Media, 2012.
[6] Haskell P. Rosenthal. On the subspaces of $L_{p}(p>2)$ spanned by sequences of independent random variables. Israel Journal of Mathematics, 8(3):273-303, 1970.


[^0]:    ${ }^{1}$ To see why, recall that

    $$
    \log m_{J}-J(2 s+1)=\log \frac{n}{H}-(J-\ell)_{+}-J(2 s+1)+O(1)
    $$

[^1]:    ${ }^{2}$ That is, find the value $J$ solving (approximately) the equation $\log \frac{n}{H}-(J-\ell)_{+}-J(2 s+1-2 / p)=0$ (note that the LHS is again decreasing in $J$ ).

[^2]:    ${ }^{3}$ It will be important later, when bounding $E_{s s}$, to note that $\frac{r-1}{s+1-1 / p}<r$, and thus one can also enforce $\alpha \leq r$.

