Optimal recovery meets minimax estimation

By DeVore, Nowak, Parhi, Petrova, and Siegel.

Jiyoung Park

October 10, 2025

Outline

Introduction

- Fundamental problem in Statistics: Regression.
  - Let $\Omega \subset \mathbb{R}^d$: bounded set with sufficient regularity (open, simply connected, Lipschitz boundary $\partial\Omega$).
  - An unknown function $f \in K$, where K is some compact set on $L_q(\Omega)$. K is dubbed model space.
  - Given: Domain data $\mathcal{X} = \{x_i\}_{i=1,\dots,m}$ and corresponding noisy function data $y = \{f(x_i) + \eta_i\}_{i=1,\dots,m}$, where $\eta_i \overset{i.i.d}{\sim} N(0, \sigma^2)$ is a noise.
  - Goal: With the fixed the domain data $\mathcal{X}$, find an algorithm A which finds the 'best' approximator of $f \in K$.
$$A : \mathbb{R}^m \to L_q(\Omega).$$
  - The performance criterion is based on the worst-case $L_q(\Omega)$ risk:
$$E_A(K; \sigma, \mathcal{X})_q := \sup_{f \in K} \mathbb{E}_{\eta_i} \|f - A(y)\|_{L_q(\Omega)}.$$
  - Minimax risk: the optimal worst-case risk over all possible choices of $\mathcal{X}$ and A:
$$\mathcal{R}_m(K; \sigma) := \inf_{A, \mathcal{X}} E_A(K; \sigma, \mathcal{X})_q.$$
    Indicates the information theoretical lower bound; one cannot make a better algorithm.
  - Optimal recovery: $\mathcal{R}_m(K; 0)$; purely deterministic setting.

TEXAS A&M UNIVERSITY

- Consider the case $\Omega = [0,1]^d$, $K = B^s_\tau(L_p(\Omega))$.
- Known minimax rate (DJ98; GN15):

$$\mathcal{R}_m(K;\sigma)_2 \asymp m^{-\frac{s}{2s+d}}.$$

- Known (non-adaptive, *i.e.*, fixed $\mathcal{X}$) optimal recovery rate (NT06; KNS21; BDPS25):

$$\mathcal{R}_m(K;0)_2 \asymp m^{-\frac{s}{d}+(\frac{1}{p}-\frac{1}{2})_+}.$$

- Observe one should have $\mathcal{R}_m(K;\sigma) \to \mathcal{R}_m(K;0)$ as $\sigma \to 0$, but $m^{-\frac{s}{2s+d}} \nrightarrow m^{-\frac{s}{d}+(\frac{1}{p}-\frac{1}{2})_+}$.
- Is the theory wrong?

- Consider the case $\Omega = [0,1]^d$, $K = B^s_\tau(L_p(\Omega))$.
- Known minimax rate (DJ98; GN15):

$$\mathcal{R}_m(K;\sigma)_2 \asymp m^{-\frac{s}{2s+d}}.$$

- Known (non-adaptive, *i.e.*, fixed $\mathcal{X}$) optimal recovery rate (NT06; KNS21; BDPS25):

$$\mathcal{R}_m(K;0)_2 \asymp m^{-\frac{s}{d}+(\frac{1}{p}-\frac{1}{2})_+}.$$

- Observe one should have $\mathcal{R}_m(K;\sigma) \to \mathcal{R}_m(K;0)$ as $\sigma \to 0$, but $m^{-\frac{s}{2s+d}} \nrightarrow m^{-\frac{s}{d}+(\frac{1}{p}-\frac{1}{2})_+}$.
- Is the theory wrong?
- A: The effect of σ is hidden in the constant.
- Conclusion of the paper (DNP$^+$25):

$$\mathcal{R}_m(K;\sigma)_q \asymp m^{-\frac{s}{d}+(\frac{1}{p}-\frac{1}{q})_+} + \min\left\{1, (\sigma^2 m^{-1})^{\frac{s}{2s+d}}\right\}.$$

Preliminaries

- Besov space: characterizes a function space with some 'smoothness'.
  - Fix the domain of X by Ω (which will be $(0,1)^d$, a unit hypercube later).
  - For $f \in L^p(\Omega)$, define the 'rth-modulus of smoothness':

$$w_{r,p}(f,t) := \sup_{\|h\| \leq t} \left\| \Delta_h^r(f) \right\|_{L^p(\Omega)},$$

where $\Delta_h^r(f)(x) := \sum_{j=0}^r \binom{r}{j}(-1)^{r-j} f(x+jh)$ if $x, x+h \in \Omega$, and 0 otherwise.
    - E.g. $\Delta_h^1(f)(x) = f(x+h) - f(x)$; $\Delta_h^2(f)(x) = f(x+2h) - 2f(x+h) + f(x)$.
  - Let $r = \lfloor s \rfloor + 1$, and define a Besov semi-norm

$$|f|_{B_\tau^s(L_p(\Omega))} := \begin{cases} \left[\int_\Omega \left(\frac{w_{r,p}(f,t)}{t^s} \right)^\tau \frac{dt}{t} \right]^{\frac{1}{\tau}} & 0 < \tau < \infty, \\ \sup_{t>0} \frac{w_{r,p}(f,t)}{t^s} & \tau = \infty. \end{cases}$$

  - If f satisfies $\|f\|_{B_\tau^s(L_p(\Omega))} := \|f\|_{L_p(\Omega)} + |f|_{B_\tau^s(L_p(\Omega))} < \infty$, then f is said to be in a Besov space $B_\tau^s(L_p(\Omega))$.
  - Remark: $f \in B_{p,\infty}^s \Leftrightarrow w_r(f, 2^{-k})_{L_p(\Omega)} \lesssim 2^{-ks}$.

- $\mathcal{P}_r$: A set of algebraic polynomial of degree $r-1$.
- Some property of $\mathcal{P}_r$:
  - for all $P \in \mathcal{P}_r$ $w_r(P,t)_{L_p(\Omega)} = 0$.
  - $\dim(\mathcal{P}_r) = \binom{d+r-1}{d} := \rho$ (by classical combinatorics argument).
- Polynomial approximation rate: For $I \subset \Omega$ a cube with the sidelength ℓ_I, write the error $E_r(g,I)_p := \inf_{P\in\mathcal{P}_r} \|g-P\|_{L_p(I)}$. Then, Whitney's theorem (Jackson's theorem type result):

$$c_{r,d,p}E_r(g,I)_p \leq w_r(g,\ell_I)_{L_p(I)} \leq C_{r,d,p}E_r(g,I)_p.$$

- If $Q \in \mathcal{P}_r$ satisfies $\|g-Q\|_{L_p(I)} \leq c_0 E_r(g,I)$, it is called near approximation.
- If $Q \in \mathcal{P}_r$ is near best $L_p(I)$ approximation, then it is near best approximation on the larger cube J and larger $\bar{p} \geq p$.
  - ($\because$) Application of quasi-norm version triangular inequality, relationship between L_p and $L_{\bar{p}}$, Q being polynomial, and Hölder.

- Normalized L_p norm:
$$\|g\|^*_{L_p(I)} := |I|^{-\frac{1}{p}} \|g\|_{L_p(I)}.$$
  - Using the equiavlence between L_p and L_q in finite dim space $\mathcal{P}_r$,
  $$\|P\|^*_{L_q(I)} \leq C \|P\|^*_{L_p(I)}.$$
  - In particular, $\|P\|_{L_q(I)} \leq C |I|^{\frac{1}{q}-\frac{1}{p}} \|P\|_{L_p(I)}$ for $q \geq p$.

- $\mathcal{D}_k$: set of dyadic cubes $I \subset \Omega$ of sidelength 2^{-k}.
- $\mathcal{S}_k(r)$: A space of r order $\mathcal{D}_k$-piecewise polynomials.
- Then,

$$f \in B^s_{p,\infty}(\Omega) \Leftrightarrow d(f, \mathcal{S}_k(r))_{L_p(\Omega)} \leq |f|_{B^s_{p,\infty}(\Omega)} 2^{-ks}.$$

  (Finite element method type argument)
- Least square approximation $S_k f \in \mathcal{S}_k(r)$ for the target f from the observation on the grid:
  - Suppose we once more decompose each I into ℓ_I/N side length grid Λ for some $N > \rho^{1/d}$;
  - e.g., for $I = [0, \ell_I)^d$, $\Lambda = \left\{0, \frac{\ell_I}{N}, \ldots, \ell_I(1 - \frac{1}{N})\right\}^d$.
  - Can define a Hilbert space $L^2(\mu_N)$ with $\mu_N := \frac{1}{N^d} \sum_{z_i \in \Lambda} \delta_{z_i}$, the empirical probability measure.
  - $\mathcal{P}_r \subset L^2(\mu_N) \Rightarrow Q_1, \ldots, Q_\rho \in \mathcal{P}_r$: Orthonormal system of $L^2(\mu_N)$.
  - $P_I f := \sum_{j=1}^{\rho} \langle f, Q_j \rangle_{L^2(\mu_N)} Q_j$: Least-square approximation of $f|_I$ to $\mathcal{P}_r$ (classical result from Hilbert space theory).
  - For $k \leq n - r$, $S_k f := \sum_{I \in \mathcal{D}_k} (P_I f)\chi_I$.

TEXAS A&M UNIVERSITY

- (DNP+25)[Lemma 2.2]: $\|f - S_k f\|_{L_q(\Omega)} \le C\,|f|_{B^s_{p,\infty}(\Omega)}\,2^{-ks+kd(\frac{1}{p}-\frac{1}{q})_+}$.
  - The proof is generalization of interpolating polynomial proof discussed in the class.
  - $f = S_0 f + \sum_{k=1}^{\infty}(S_k f - S_{k-1} f) \Rightarrow \|f - S_k f\| \le \sum_{j>k}\|S_j f - S_{j-1} f\|_{L_q(\Omega)}$.
  - Use the L_p, L_q-norm equivalence and triangular inequality to write the bound w.r.t. $d(f, \mathcal{S}_k(r))_{L_p(\Omega)}$.
  - $|f|_{B^s_{p,\infty}}\,2^{-ks}$ term comes from $d(f, \mathcal{S}_k(r))_{L_p(\Omega)} \le |f|_{B^s_{p,\infty}}\,w_r(f, 2^{-k})_{L_p(\Omega)}$.
  - $2^{kd(\frac{1}{p}-\frac{1}{q})_+}$ term comes from $\|P\|_{L_q(I)} \le C\,|I|^{\frac{1}{q}-\frac{1}{p}}\,\|P\|_{L_p(I)} = C2^{kd(\frac{1}{p}-\frac{1}{q})}\,\|P\|_{L_Q(I)}$ for $P \in \mathcal{P}_r$.
- The noisy version counterpart will be denoted $\widetilde{\cdot}$ (e.g., $\widetilde{S}_k y$).

Main result

TEXAS A&M UNIVERSITY

- $\Omega = (0,1)^d$.
- $\eta_i \overset{i.i.d}{\sim} (0, \sigma^2)$ be a sub-Gaussian distribution.
- The model space $K = U(B^s_{p,\tau}(\Omega))$.
- $0 < p, \tau \leq \infty$, $p \leq q \leq \infty$, $s > 0$, $s > d/p$, $q < p + 2\frac{sp}{d}$.
  - $s > d/p$ is required to make $f \in K$ to be continuous (Sobolev embedding).
  - $q < p + 2\frac{sp}{d}$ ensures the minimax rate to be $m^{-\frac{s}{2s+d}}$ (*primary case*); but this can be dropped (DNP+25)[Remark 8.1].
  - $p \geq q$ case falls down to $p = q$ case.

TEXAS A&M UNIVERSITY

Theorem

(DNP$^+$25)[Theorem 1.3]

$$\mathcal{R}_m(K;\sigma) \gtrsim m^{-\frac{s}{d}+(\frac{1}{p}-\frac{1}{q})_+} + \min\left\{1, (\sigma^2 m^{-1})^{\frac{s}{2s+d}}\right\}.$$

- This indicates the information-theoretic lower bound; no algorithm can achieve the faster rate than this.
- Typical approach: create the 'worst-case' f_A for each algorithm f_A.

TEXAS A&M UNIVERSITY

Theorem

(DNP[+] 25)[Theorem 1.1, 1.3] If $\mathcal{X}_m = G_n := \{0, 2^{-n}, \ldots, 1-2^{-n}\}^d$ *for* $m = 2^{nd}$*, then for any* $\alpha \in (0, 2 - \frac{d(q-p)_+}{sp})$ *there exists an algorithm* A *such that*

$$\mathbb{P}\left[\|f - A(y)\|_{L_q} \geq C\left(m^{-\frac{s}{d}+(\frac{1}{p}-\frac{1}{q})_+} + t(\sigma^2 m^{-1})^{\frac{s}{2s+d}}\right)\right] \leq C\exp(-ct^{\alpha})$$

for some constant c, C *which does not depend on* m, σ*.*
Accordingly,

$$\mathcal{R}_m(K;\sigma,\mathcal{X}_m) \lesssim m^{-\frac{s}{d}+(\frac{1}{p}-\frac{1}{q})_+} + \min\left\{1, (\sigma^2 m^{-1})^{\frac{s}{2s+d}}\right\}.$$

- To convert probability bound to expectation bound, there is a canonical formula: $\mathbb{E}(X) = \int_0^\infty P(X > s)ds$ for X: non-negative random variable.
  - Remark: For the opposite direction, one can use Markov inequality (in probability theory).
- Two theorems together,

$$\mathcal{R}_m(K;\sigma) \asymp m^{-\frac{s}{d}+(\frac{1}{p}-\frac{1}{q})_+} + \min\left\{1, (\sigma^2 m^{-1})^{\frac{s}{2s+d}}\right\}.$$

- For the lower bound, can assume $\sigma^2 \leq m$.
- Let $\epsilon := (\sigma^2 m^{-1})^{\frac{s}{2s+d}}$ and n be the smallest integer such that $n^{-s} \leq \epsilon$ (think as an equal). $\therefore \epsilon \asymp \sigma n^{\frac{d}{2}}/\sqrt{m}$.
- For fixed $\mathcal{X}_m$, let $y(f) := \{f(x_i)\}_{i=1,\ldots,m}$, $\widetilde{y}(f) := \{f(x_i) + \eta_i\}_{i=1,\ldots,m}$.
- Want to show: For any algorithm $A : \mathbb{R}^m \to L_q(\Omega)$ there exists $f_A \in K$ s.t.

$$\mathbb{E}_\eta \left\| f_A - A(\widetilde{y}(f)) \right\|_{L_q} \gtrsim \epsilon$$

whenever $\epsilon \gtrsim m^{-\frac{s}{d}+(\frac{1}{p}-\frac{1}{q})_+}$.
- Known fact from OR literature: For any $\mathcal{X}_m$, there exists f, g such that they coincides on $\mathcal{X}_m$, *i.e.*, $y(f) = y(g)$, but $\|f - g\|_{L_q} \gtrsim m^{-\frac{s}{d}+(\frac{1}{p}-\frac{1}{q})_+}$.
- Then,

$$m^{-\frac{s}{d}+(\frac{1}{p}-\frac{1}{q})_+} \lesssim \left\| f - A(\widetilde{y}(f)) \right\|_q + \underbrace{\left\| A(\widetilde{y}(f)) - A(\widetilde{y}(g)) \right\|}_{=0\ (\because y(f)=y(g))} + \left\| g - A(\widetilde{y}(g)) \right\|.$$

- Implying

$$\max\left\{ \mathbb{E} \left\| f - A(\widetilde{y}(f)) \right\|_q, \mathbb{E} \left\| f - A(\widetilde{y}(g)) \right\|_q \right\} \gtrsim m^{-\frac{s}{d}+(\frac{1}{p}-\frac{1}{q})_+}.$$

Meaning when $\epsilon \leq m^{-\frac{s}{d}+(\frac{1}{p}-\frac{1}{q})_+}$ the optimal recovery rate appears.

- Typical minimax bound methods: Le Cam, Fano, Assouad.
- Idea: Estimation and choosing the closest one among sufficiently discretized set is almost similar.
- Goal: Construct a $c_0\epsilon$-separated covering $\mathcal{F}_N := \{f_1, \dots, f_N\} \subset K$, *i.e.*, $\left\|f_i - f_j\right\|_q \geq c_0\epsilon$ for all $i, j \leq N \Rightarrow$ we choose f_A among f_i's.
- Let $\phi \in C_c^\infty([0,1]^d)$ s.t. $\|\phi\|_\infty = 1$, $\|\phi\|_{B^s_{p,\tau}} := M < \infty$.
- Shift and scale this function to the cubes of side length $1/n$ $Q_1, \dots, Q_{n^d}$ (n defined in the above). $\phi_i(\cdot) = \gamma n^{-s}\phi(n(\cdot - \text{Bottom left corner of } Q_i))$. γ will be chosen later to match quantities.
- Result from combinatorics: there exists $S \subset \{\pm 1\}^{n^d}$ such that $|S| \geq 2^{cn^d}$ and for all $a \neq b \in S$ $\|a - b\|_{\ell^1} \geq cn^d$; cn^d-separating.
  - Meaning: Many but still well-separated.
  - e.g., $n^d = 2$, then $\|(1,1) - (1,0)\|_{\ell_1} \geq 2$.

Proof of lower bound III

- $\mathcal{F}_N := \left\{ f = \sum_{i=1}^{n^d} \kappa_i \phi_i \mid (\kappa_1, \ldots, \kappa_{n^d}) \in S \right\}$ with $N = |S| \geq 2^{cn^d}$. γ is chosen to make $\mathcal{F}_N \subset K$ (see Figure 1).

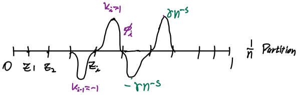

Figure 1: Example of $f \in \mathcal{F}_N$.

- Now, all $f_i \neq f_j \in \mathcal{F}_N$ satisfy
  - $\|f_i\|_\infty \leq c\gamma n^{-s} \leq c\gamma\epsilon$. $\Rightarrow \|y(f_i)\|_{\ell^2} \leq c\sqrt{m}\gamma\epsilon$.
  - $\|f_i - f_j\|_q \geq c\gamma\epsilon$ (any $c \leq n^d$ will work).
- By setting γ small enough,
  - Known metric entropy $\epsilon_{\log_2 N}(K)_{L_q} \asymp (\log_2 N)^{-\frac{s}{d}} \geq cn^{-s} \geq c\epsilon$ indicates $\mathcal{F}_N$ is a $c\epsilon$-separated covering of K.
  - $\|y(f_i)\|_{\ell^2} < \sigma\sqrt{\log(N/5)}$.

- We have $c\epsilon$-separated covering $\mathcal{F}_N$. Now, consider

$$B_i := A^{-1}\left(\left\{g \in L_q \mid \|f_i - g\|_q < \frac{c\epsilon}{2}\right\}\right) \subset \mathbb{R}^m.$$

  B_i means a region where the algorithm A chooses f_i among $\mathcal{F}_N$.
- Since $\mathcal{F}_N$ is a covering, $\coprod_i B_i = \mathbb{R}^m$.
- We will show B_i has a small measure.
- There exists i such that $\mathbb{P}_{N(0,\sigma^2)}(B_i) \le \frac{1}{N}$.
  - $\because \sum_{i=1}^N \mathbb{P}_{N(0,\sigma^2)}(B_i) = \mathbb{P}_{N(0,\sigma^2)}(\coprod_i B_i) = 1$.
- Known Lemma: If $\mathbb{P}_{N(0,\sigma^2)}(B_i) \le \frac{1}{N}$ and $\|y(f_i)\| < \sigma\sqrt{\log(N/5)}$, then $\mathbb{P}_{N(y(f_i),\sigma^2)}(B_i) < \frac{1}{2}$.
- This means any algorithm A has the corresponding f_i; while data is from f_i, the algorithm A is not likely to choose f_i. This is our f_A.
- Then, $\mathbb{E}_{\eta_i}\|f_i - A(\widetilde{y}(f_i))\| = \int_{B_i \coprod B_i^c} \|f_i - A(\widetilde{y}(f))\| \ge \mathbb{P}_{N(y(f_i),\sigma^2)}(B_i^c)\frac{c\epsilon}{2} \gtrsim c\epsilon$. □

- Main idea: construct an explicit algorithm that achieves the rate.
- Construction of the algorithm: Assume $\mathcal{X}_m$ be the regular grid.
  1. If $\sigma^2 \geq m$, then $A(y) = 0$ (too noisy regime).
  2. Otherwise, partition Ω by $\mathcal{D}_k$, a dyadic cube of side length of $I \in \mathcal{D}_k$ being 2^{-k}.
  3. Conduct a least square approximation by $\mathcal{P}_r$, r-degree polynomial with $r \leq n - k$ (defined above).
  4. Consider the orthonormal basis expression of $\widetilde{R}_I y := \widetilde{P}_I y - \widetilde{P}_{\text{parent}(I)} y = \sum_{j=1}^{\rho} c^*_{I,j}(y) Q_{I,j}$. Compute $c^*_{I,j}(y)$.
  5. Let $\widehat{c}_{I,j}(y) := \text{Thresholding}_{\lambda_k}(c^*_{I,j}(f))$. Construction of λ_k is as follows:
     - Let k^* be an integer s.t. $2^{k^*-1} < \epsilon^{-1/s} < 2^{k^*}$.
     - $\lambda_k := 0$ if $k \leq k^*$ and $2^{\beta k - \beta k^* - sk^*}$ otherwise. β is chosen to satisfy $\alpha = \frac{2\beta - d}{\beta + \delta}$ for some $\delta \in (0, \frac{d}{2})$ (Recall: α is tail prob we control).
  6. $\widehat{T}_k y = \sum_{I \in \mathcal{D}_k} \left(\sum_{j=1}^{\rho} \widehat{c}_{I,j}(y) Q_{I,j} \right) \chi_I$, and $\widehat{f} = A(y) = \sum_{k=0}^{n-r} \widehat{T}_k y$.
- Remark: s, σ^2 must be known a priori (to find k^* and therefore construct λ_k).

- One can view $T_k = S_k f - S_{k-1} f$.
- $\left\|\widehat{f} - f\right\|_q = \left\|f - \sum_{k=0}^{n-r} \widehat{T} y\right\| = \left\|\sum_{k=0}^{n-r} T_k f + S_{n-r} f - \sum_{k=0}^{n-r} \widehat{T}_k y\right\| \leq$ $\|f - S_{n-r} f\|_q + \sum_{k=0}^{n-r} \left\|T_k f - \widehat{T}_k y\right\|$.
- First term comes directly from approximation by least square: $\lesssim 2^{-(n-r)(s-d(\frac{1}{p}-\frac{1}{q})_+)} \lesssim m^{-(s-d(\frac{1}{p}-\frac{1}{q})_+)}$.
- The second term, as $Q_{I,j}$ is an orthonormal system,

$$\left\|T_k f - \widehat{T}_k y\right\|_q^q \lesssim \sum_{I \in \mathcal{D}_k} \left(\sum_{j=1}^{\rho} \left|c_{I,j}(f) - \widehat{c}_{I,j}(y)\right|^q \right) |I| \lesssim 2^{-kd} \sum_{I \in \mathcal{D}_k} \left(\sum_{j=1}^{\rho} \left|c_{I,j}(f) - \widehat{c}_{I,j}(y)\right|^q \right).$$

- For notational simpliicty, we aggregate $\nu_k := (c_{I,j})_{I \in \mathcal{D}_k, j \in \rho}$. Then, one can show

$$2^{-kd} \sum_{I \in \mathcal{D}_k} \left(\sum_{j=1}^{\rho} \left|c_{I,j} - \widehat{c}_{I,j}\right| \right) \leq \|\nu_k - \widehat{\nu}_k\|_q^* := \left(\frac{1}{L_k} \sum_{l=1}^{L_k} |(\nu_k)_l - (\widehat{\nu}_k)_l|^q \right)^{1/q}.$$

- In sum, we have

$$\left\|\widehat{f} - f\right\|_q \lesssim m^{-(s-d(\frac{1}{p}-\frac{1}{q})_+)} + \sum_{k=0}^{n-r} \|\nu_k - \widehat{\nu}_k\|_q^*.$$

- Goal: bound $\sum_{k=0}^{n-r} \|\nu_k - \widehat{\nu}_k\|_q^*$ with high probability.
- Using the fact that $T_k f$ is piecewise polynomial and the approximation rate of piecewise polynomial to Besov space, one can get $\|\nu_k\|_p^* \lesssim 2^{-ks}$ (DNP$^+$25)[Lemma 2.3].
- Then, using the known properties about thresholding on Gaussian noise ,

$$\|\nu_k - \widehat{\nu}_k\|_q^* \lesssim \underbrace{2^{-\frac{ksp}{q}} \lambda_k^{1-\frac{p}{q}}}_{\text{Deterministic}} + \underbrace{\|\eta_{\lambda_k}\|_q^*}_{\text{Stochastic}}.$$

Here, η_{λ_k} is a 0-mean $\sigma_{I,j}^2 \in (0, C2^{-(n-k)d}\sigma^2)$-variance normal variable thresholded by $\lambda_k/2$.

- Sum of determinstic term: interpret as low-signal $\Rightarrow$ error is at most λ_k. It is $\lesssim 2^{-k^* s} \lesssim \epsilon$ by the construction of λ_k (In fact, λ_k is chosen to bound this).
  - The part where λ_k and s are interacting, via $\|\nu_k\|_p^*$.
- Sum of stochastic term: interpret as a high noise level. Bounding this with high probability can be done with classical thresholding analysis result.
  - Remark: Typical Thresholding analysis was on expectation bound, but (DNP$^+$25) did a probability bound; this induces the analysis resulting in probability bound rather than expectation.

Conclusion

- Interpolating the result between optimal recovery and minimax bound.
  - Upper bound: by proposing explicit algorithm: Thresholding + least square piecewise polynomial fit.
  - Lower bound: Converting the estimation problem to choosing one among sufficiently well-discretized set.
- Pros: Explicit algorithm. Simple to implement.
- Cons: Need to know s, σ^2 a priori to determine the optimal threshold.

- Theoretical side: More general function space setting.
- Algorithmic side: improving algorithms? Or does other existing algorithm acheives the adaptive tight rate w/o knowledge of σ?
  - Adaptive to unknown parameters s, σ?
  - Neural network?

Thank You For Your Attention!

[BDPS25] Andrea Bonito, Ronald DeVore, Guergana Petrova, and Jonathan W Siegel, *Convergence and error control of consistent pinns for elliptic pdes*, IMA Journal of Numerical Analysis (2025), draf008.

[DJ98] David L. Donoho and Iain M. Johnstone, *Minimax estimation via wavelet shrinkage*, The Annals of Statistics **26** (1998), no. 3, 879 – 921.

[DNP+25] Ronald DeVore, Robert D. Nowak, Rahul Parhi, Guergana Petrova, and Jonathan W. Siegel, *Optimal recovery meets minimax estimation*, 2025.

[GN15] Evarist Giné and Richard Nickl, *Mathematical foundations of infinite-dimensional statistical models*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2015.

[KNS21] David Krieg, Erich Novak, and Mathias Sonnleitner, *Recovery of sobolev functions restricted to iid sampling*, Math. Comput. **91** (2021), 2715–2738.

[NT06] E Novak and H Triebel, *Function spaces in lipschitz domains and optimal rates of convergence for sampling*, Constr Approx **23** (2006).