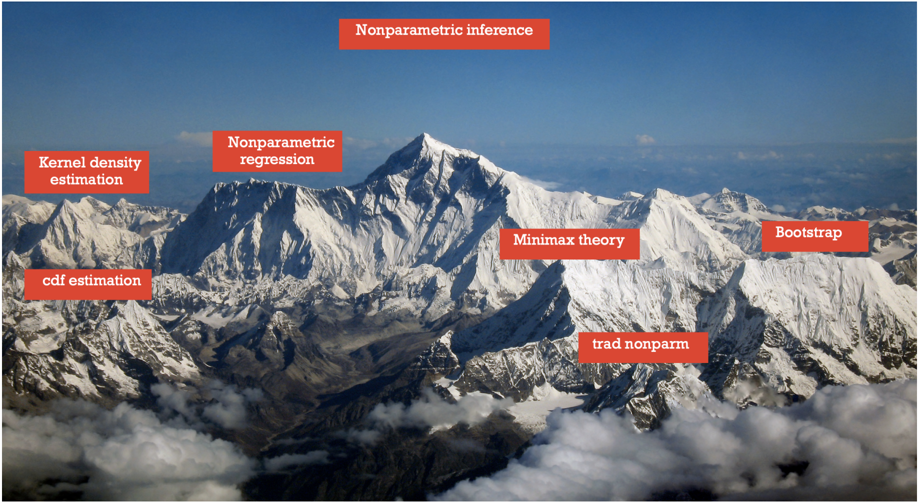
STAT 824 sp 2025 Lec 7 slides

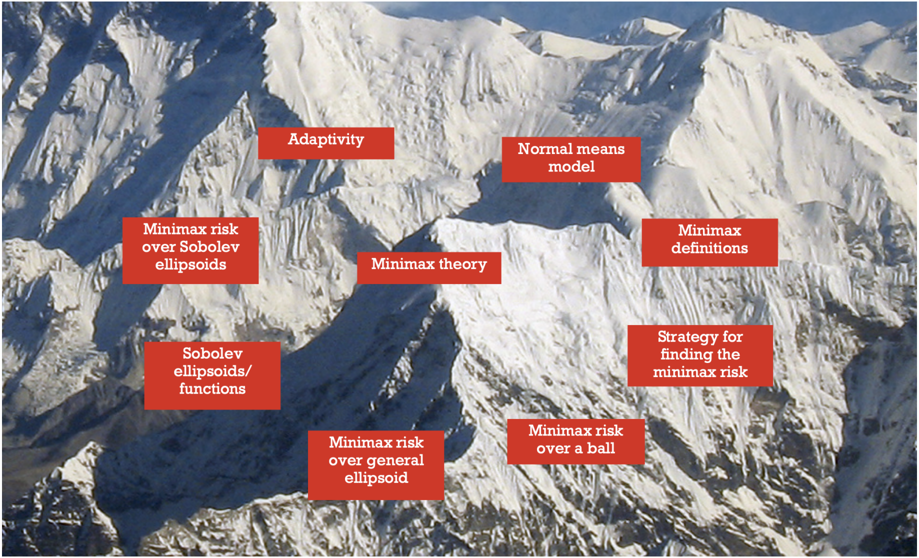
Minimax theory

Karl Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.


Nonparametric inference
Nonparametric regression
Kernel density estimation
Minimax theory
Bootstrap
cdf estimation
trad nonparm


Adaptivity
Normal means model
Minimax risk over Sobolev ellipsoids
Minimax theory
Minimax definitions
Sobolev ellipsoids/ functions
Strategy for finding the minimax risk
Minimax risk over general ellipsoid
Minimax risk over a ball

A nonparametric regression model

Suppose we observe $Y_1, \dots, Y_n$ arising as

$$Y_i = m(i/n) + \varepsilon_i, \quad i = 1, \dots, n$$

where $m \in L_2([0,1])$ and $\varepsilon_1, \dots, \varepsilon_n \overset{\text{ind}}{\sim} \mathcal{N}(0, \sigma^2)$, for some $\sigma^2 > 0$.

In above $L_2([0,1]) = \left\{ g : [0,1] \to \mathbb{R} : \int_0^1 |g(x)|^2 dx < \infty \right\}$.

Equally spaced design points $x_i = i/n$ for $i = 1, \dots, n$ for convenience.

We want to connect estimation of m in nonparametric regression to estimation in the Normal means model:

The Normal means model

Observe $Z_1, \dots, Z_n$ from

$$Z_i = \theta_i + n^{-1/2}\sigma\xi_i, \quad i = 1, \dots, n,$$

where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n) \in \Theta \subset \mathbb{R}^n$, $\xi_1, \dots, \xi_n \overset{\text{ind}}{\sim} \mathcal{N}(0, 1)$, and $\sigma > 0$.

These notes closely follow [1].

- Let $\varphi_1, \varphi_2, \ldots : [0,1] \to \mathbb{R}$ be a collection of functions such that

$$\int_0^1 \varphi_i(x)\varphi_j(x)dx = \begin{cases} 1 & i = j \\ 0 & i \neq j, \end{cases}$$

  for all i and j and such that for any function $g \in L_2([0,1])$

$$\int_0^1 g(x)\varphi_j(x)dx = 0 \quad \forall j \iff g(x) = 0.$$

- Then $\varphi_1, \varphi_2, \ldots$ comprise an orthonormal basis for $L_2([0,1])$.

- For $g \in L_2([0,1])$ we may write

$$g(x) = \sum_{j=1}^{\infty} \theta_j \varphi_j(x),$$

  where $\theta_j = \int_0^1 g(x)\varphi_j(x)dx, \quad j = 1, 2, \ldots$

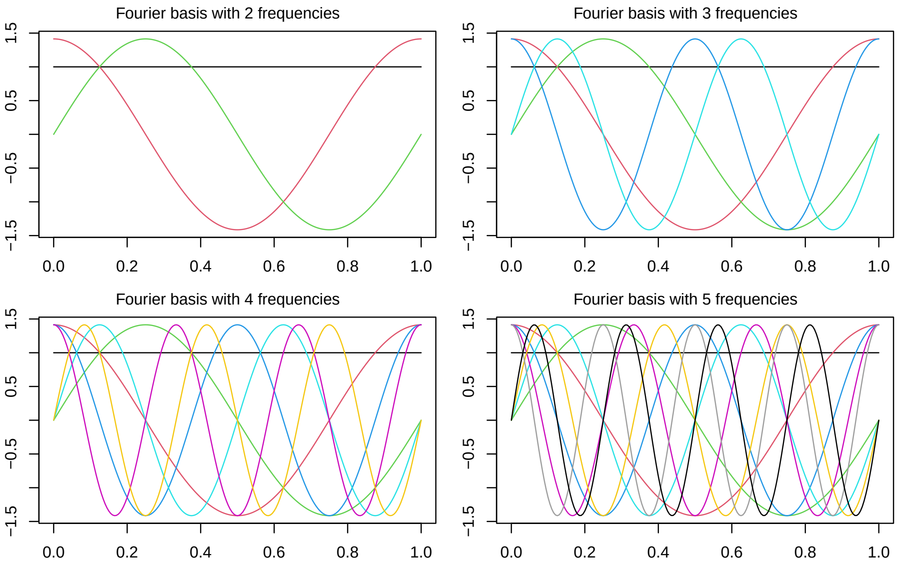
- Moreover we have $\int_0^1 |g(x)|^2 dx = \sum_{j=1}^{\infty} \theta_j^2$, which is called Parseval's identity.

Fourier basis

The Fourier basis on $[0,1]$ is the collection of functions $\phi_1(x) = 1$ and

$$\begin{aligned} \phi_{2k}(x) &= \sqrt{2}\cos(2\pi kx) \\ \phi_{2k+1}(x) &= \sqrt{2}\sin(2\pi kx) \end{aligned}$$

for $k = 1, 2, \ldots$



Fourier basis with 2 frequencies
Fourier basis with 3 frequencies
Fourier basis with 4 frequencies
Fourier basis with 5 frequencies
1.5
0.5
−0.5
−1.5
0.0
0.2
0.4
0.6
0.8
1.0

- Assume $m \in L_2([0,1])$ and suppose

$$m(x) \approx \sum_{j=1}^{n} \theta_j \varphi_j(x)$$

for some $\theta_1, \dots, \theta_n$ and orth. $\varphi_1, \dots, \varphi_n$.

- Consider estimating $\theta_1, \dots, \theta_n$ based on $Y_1, \dots, Y_n$ with

$$\hat{\theta}_j = \frac{1}{n} \sum_{i=1}^{n} Y_i \varphi_j(i/n), \quad j = 1, \dots, n.$$

- Then set $\hat{m}_n(x) = \sum_{j=1}^{n} \hat{\theta}_j \varphi_j(x)$.

Exercise: Show that estimating m is approximately like estimating $\theta_1, \dots, \theta_n$ in the Normal means model. Consider estimating each θ_j and consider MISE $\hat{m}_n$.

The Normal means model

Observe $Z_1, \dots, Z_n$ from

$$Z_i = \theta_i + n^{-1/2}\sigma\xi_i, \quad i = 1, \dots, n,$$

where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n) \in \Theta \subset \mathbb{R}^n$, $\xi_1, \dots, \xi_n \overset{\text{ind}}{\sim} \mathcal{N}(0, 1)$, and $\sigma > 0$.

We ask how well we can estimate $\boldsymbol{\theta}$ depending on the space Θ.

- *Loss (squared error)*: Represents the cost of estimation error.

$$L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 = \sum_{i=1}^{n} (\hat{\theta}_i - \theta_i)^2.$$

- *Risk*: The risk is the expected value of the loss.

$$R(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \mathbb{E}L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \sum_{i=1}^{n} \mathbb{E}(\hat{\theta}_i - \theta_i)^2.$$

- *Minimax risk*: The best worst performance of any estimator of $\boldsymbol{\theta} \in \Theta$.

$$M(\Theta) = \inf_{\hat{\boldsymbol{\theta}}} \sup_{\boldsymbol{\theta} \in \Theta} R(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}),$$

where the infimum is taken over all estimators.

- *Integrated risk*: with respect to a prior π is defined as

$$I_\pi(\hat{\boldsymbol{\theta}}) = \int R(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) d\pi(\boldsymbol{\theta}).$$

- *Bayes estimator*: under the prior π is defined as

$$\hat{\boldsymbol{\theta}}_\pi = \underset{\hat{\boldsymbol{\theta}}}{\operatorname{argmin}} \int R(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) d\pi(\boldsymbol{\theta}),$$

- *Integrated Bayes risk*: The integrated risk of the Bayes estimator

$$I_\pi = I_\pi(\hat{\boldsymbol{\theta}}_\pi) = \int R(\hat{\boldsymbol{\theta}}_\pi, \boldsymbol{\theta}) d\pi(\boldsymbol{\theta}).$$

- For any estimator $\hat{\boldsymbol{\theta}}$ we may write

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\Theta} R(\hat{\boldsymbol{\theta}},\boldsymbol{\theta}) &\geq \int_{\Theta} R(\hat{\boldsymbol{\theta}},\boldsymbol{\theta})d\pi(\boldsymbol{\theta})\\ &= \int R(\hat{\boldsymbol{\theta}},\boldsymbol{\theta})d\pi(\boldsymbol{\theta}) - \int_{\Theta^c} R(\hat{\boldsymbol{\theta}},\boldsymbol{\theta})d\pi(\boldsymbol{\theta})\\ &\geq I_\pi - \int_{\Theta^c} R(\hat{\boldsymbol{\theta}},\boldsymbol{\theta})d\pi(\boldsymbol{\theta}),\end{aligned}$$

- Taking the infimum of both sides of the above over all estimators $\hat{\boldsymbol{\theta}}$ gives

$$M(\Theta) \geq I_\pi - \sup_{\hat{\boldsymbol{\theta}}} \int_{\Theta^c} R(\hat{\boldsymbol{\theta}}_\pi,\boldsymbol{\theta})d\pi(\boldsymbol{\theta}),$$

- Second term vanishes under a seq. of priors that concentrates on Θ.

Strategy for finding the minimax risk

To find the minimax risk $M(\Theta)$, propose a candidate value M^* and then:

1. Find estimator with worst-case risk equal to (or $\leq$) M^*. Shows $M(\Theta) \leq M^*$.
2. Find a prior (or a sequence of priors) such that the integrated Bayes risk over Θ is equal to (or converges to) M^*. Shows $M(\Theta) \geq M^*$.

Steps 1 and 2 give $M(\Theta) = M^*$.

Define the ball with radius c centered at the origin

$$\Theta_n(c) = \left\{ \boldsymbol{\theta} \in \mathbb{R}^n : \|\boldsymbol{\theta}\| \leq c \right\}.$$

Two results

In the Normal means model we have the minimax risks

1. $M(\mathbb{R}^n) = \sigma^2$
2. $\liminf_{n \to \infty} M(\Theta_n(c)) = \dfrac{\sigma^2 c^2}{\sigma^2 + c^2}$

Exercise: Work through the proofs of the above.

To bring out minimax risk results which look like the nonparametric rates of convergence we have seen, e.g. $n^{-2\beta/(2\beta+1)}$, we consider Sobolev functions...

Sobolev class and periodic Sobolev class

For β a positive integer and $L > 0$, define the *Sobolev class* of functions as

$$\mathcal{W}(\beta, L) = \left\{ m : [0,1] \to \mathbb{R} : \begin{array}{c} m^{(\beta-1)} \text{ is absolutely continuous} \\ \text{and } \int_0^1 (m^{(\beta)}(x))^2 dx \leq L^2 \end{array} \right\}.$$

Moreover, define the *periodic Sobolev class* of functions $W_{\text{per}}(\beta, L)$ as

$$\mathcal{W}_{\text{per}}(\beta, L) = \{m \in W(\beta, L) : m^{(\ell)}(0) = m^{(\ell)}(1) \text{ for } \ell = 0, \dots, \beta - 1\}.$$

Sobolev ellipsoid

Define the *Sobolev ellipsoid* $\Theta(\beta, c)$ as

$$\Theta_{\text{Sob}}(\beta, L) = \left\{(\theta_1, \theta_2, \dots) : \sum_{j=1}^{\infty} \theta_j^2 < \infty \text{ and } \sum_{j=1}^{\infty} a_j^2 \theta_j^2 \leq L^2/\pi^{2\beta}\right\},$$

where $a_1 = 0$, $a_{2m} = a_{2m+1} = (2m)^\beta$, $m = 1, 2, \dots$

Definition of $a_1, a_2, \dots$ equivalent to

$$a_j = \begin{cases} j^\beta, & j \text{ even} \\ (j-1)^\beta, & j \text{ odd.} \end{cases}$$

Result: Fourier basis as a basis for periodic Sobolev functions

We have

$$\mathcal{W}_{\text{per}}(\beta, L) = \left\{ m : [0,1] \to \mathbb{R} : m(x) = \sum_{j=1}^{\infty} \theta_j \phi_j(x),\ (\theta_1, \theta_2, \dots) \in \Theta_{\text{Sob}}(\beta, L) \right\}.$$

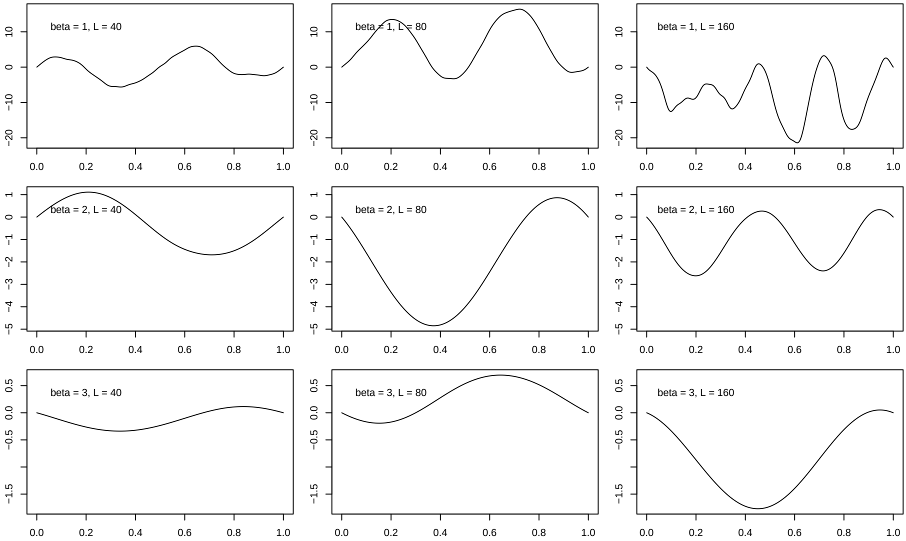
- So we can construct any periodic Sobolev function from the Fourier basis.
- Estimating $m \in \mathcal{W}_{\text{per}}(\beta, L)$ is essentially like estimating $\boldsymbol{\theta} \in \Theta_{\text{Sob}}(\beta, L)$ in the normal Means model.

What do periodic Sobolev functions look like?

Exercise: Can generate some functions belonging to $\mathcal{W}_{\text{per}}(\beta, L)$ in these steps:

1. Draw $\theta_1, \dots \theta_N \overset{\text{ind}}{\sim} \text{Normal}(0, 1)$ with $N = 50$, say.
2. Minimize $\sum_{i=1}^N (\theta_j - w_j)^2$ subject to $\sum_{j=1}^N a_j^2 w_j^2 = L^2/\pi^{2\beta}$, where the a_j are those which define the Sobolev ellipsoid $\Theta_{\text{Sob}}(\beta, L)$.
3. Set $m(x) = \sum_{j=1}^N \hat{w}_j \phi_j(x)$, where $\hat{w}_1, \dots, \hat{w}_N$ are from step 2 and $\{\phi_1, \phi_2, \dots\}$ is the Fourier basis.

Next slide plots several after subtracting $m(0)$ (to aid comparison).


beta = 1, L = 40
beta = 1, L = 80
beta = 1, L = 160
beta = 2, L = 40
beta = 2, L = 80
beta = 2, L = 160
beta = 3, L = 40
beta = 3, L = 80
beta = 3, L = 160

Infinite-dimensional Normal means model

Let

$$Z_j = \theta_j + \sigma\xi_j, \quad j = 1, 2, \dots,$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots)$ is unknown, $\xi_1, \xi_2, \dots, \overset{\text{ind}}{\sim} \mathcal{N}(0, 1)$, and $\sigma > 0$.

Want to find minimax risk for estimating $\boldsymbol{\theta}$ in the above model when

1. $\boldsymbol{\theta}$ lies in a general ellipsoid

$$\Theta(c, a_1, a_2, \dots) = \left\{(\theta_1, \theta_2, \dots) \in \mathbb{R} : \sum_{j=1}^{\infty} \theta_j^2 < \infty \text{ and } \sum_{j=1}^{\infty} a_j^2 \theta_j^2 \leq c^2\right\}.$$

2. $\boldsymbol{\theta} \in \Theta_{\text{Sob}}(\beta, L)$.

- Consider only linear estimators, i.e. of the form

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{\lambda}} = (\lambda_1 Z_1, \lambda_2 Z_2, \dots) \quad \text{for } \lambda = (\lambda_1, \lambda_2, \dots).$$

- Let

$$M_{\text{lin}}(\Theta(c, a_1, a_2, \dots)) = \inf_{\boldsymbol{\lambda}} \sup_{\boldsymbol{\theta} \in \Theta(c, a_1, a_2, \dots)} R(\hat{\boldsymbol{\theta}}_{\boldsymbol{\lambda}}, \boldsymbol{\theta})$$

denote the *linear minimax risk* over $\Theta(c, a_1, a_2, \dots)$, where

$$R(\hat{\boldsymbol{\theta}}_{\boldsymbol{\lambda}}, \boldsymbol{\theta}) = \sum_{j=1}^{\infty} \mathbb{E}(\lambda_j Z_j - \theta_j)^2.$$

Linear minimax risk over a general ellipsoid

Let $a_1, a_2, \dots$ be an increasing seq. such that $|\{j : a_j = 0\}| < \infty$ and $a_j \to +\infty$. Then a unique solution to

$$\eta^{-1}\sigma^2 \sum_{i=1}^{\infty} a_j(1 - \eta a_j)_+ = c^2 \tag{1}$$

over $\eta > 0$ exists such that, setting $\ell_j = (1 - \eta a_j)_+$ for $j = 1, 2, \dots$ and $\boldsymbol{\ell} = (\ell_1, \ell_2, \dots)$, we have

$$M_{\text{lin}}(\Theta(c, a_1, a_2, \dots)) = \sup_{\boldsymbol{\theta} \in \Theta(c, a_1, a_2, \dots)} R(\hat{\boldsymbol{\theta}}_{\boldsymbol{\ell}}, \boldsymbol{\theta}) = \sigma^2 \sum_{j=1}^{\infty} \ell_j,$$

provided the sum is finite.

The values $\ell_1, \ell_2, \dots$ are called the *Pinsker weights*.

Exercise: To see why one should consider $\sigma^2 \sum_{j=1}^\infty \ell_j$ as a candidate for the linear minimax risk, show the following:

1. We have
$$\inf_{\boldsymbol{\lambda}} R(\hat{\boldsymbol{\theta}}_{\boldsymbol{\lambda}}, \boldsymbol{\theta}) = \sum_{j=1}^\infty \frac{\sigma^2\theta_j^2}{\sigma^2 + \theta_j^2}.$$

2. An equation like that in (1) arises if we solve
$$\text{maximize} \sum_{j=1}^\infty \frac{\sigma^2\theta_j^2}{\sigma^2 + \theta_j^2} \quad \text{subject to} \sum_{j=1}^\infty a_j^2\theta_j^2 = c^2.$$

3. The maximum above is of the same form as $\sigma^2 \sum_{j=1}^\infty \ell_j$.

- Prove the linear minimax risk result by establishing

$$\sup_{\boldsymbol{\theta} \in \Theta(c, a_1, a_2, \dots)} R(\hat{\boldsymbol{\theta}}_{\boldsymbol{\ell}}, \boldsymbol{\theta}) \leq \sigma^2 \sum_{j=1}^{\infty} \ell_j$$

and

$$\sup_{\boldsymbol{\theta} \in \Theta(c, a_1, a_2, \dots)} \inf_{\boldsymbol{\lambda}} R(\hat{\boldsymbol{\theta}}_{\boldsymbol{\lambda}}, \boldsymbol{\theta}) \geq \sigma^2 \sum_{j=1}^{\infty} \ell_j.$$

- Explain why this is sufficient.

Adapted from Lemma 3.3 on pages 144–145 of Tsybakov [1]

Let $a_1 = 0$, $a_{2m} = a_{2m+1} = (2m)^\beta$, $m = 1, 2, \dots$ and let η be the solution to

$$\eta^{-1}\sigma^2 \sum_{i=1}^\infty a_j(1 - \eta a_j)_+ = L^2/\pi^{2\beta}$$

over $\eta > 0$ and set $\ell_j = (1 - \eta a_j)_+$, $j = 1, 2, \dots$ Then

1. $\eta = (\beta^{-1}\pi^{-2\beta}L(\beta+1)(2\beta+1))^{-\frac{\beta}{2\beta+1}}\sigma^{\frac{2\beta}{2\beta+1}}(1 + o(1))$
2. $\sigma^2 \sum_{j=1} \ell_j = C\sigma^{\frac{4\beta}{2\beta+1}}(1 + o(1))$

as $\sigma \to 0$, where $C = L^{\frac{2}{2\beta+1}}(\beta^{-1}\pi(\beta+1))^{-\frac{2\beta}{2\beta+1}}(2\beta+1)^{\frac{1}{2\beta+1}}$.

Linear minimax risk over a Sobolev ellipsoid

Under the infinite-dimensional Normal means model we have

$$M_{\text{Lin}}(\Theta_{\text{Sob}}(\beta, L)) = \sup_{\boldsymbol{\theta} \in \Theta_{\textbf{Sob}}(\beta, L)} R(\hat{\boldsymbol{\theta}}_{\boldsymbol{\ell}}, \boldsymbol{\theta}) = C\sigma^{\frac{4\beta}{2\beta+1}}(1 + o(1))$$

as $\sigma \to 0$, where ℓ and C are as on the previous slide.

Think of replacing σ^2 with σ^2/n. Then we obtain a minimax risk like

$$\tilde{C} n^{-\frac{2\beta}{2\beta+1}}(1 + o(1)) \quad \text{as } n \to \infty,$$

which resembles the nonparametric rates we have encountered before.

Alexandre B Tsybakov.
Introduction to nonparametric estimation.
Springer Science & Business Media, 2008.