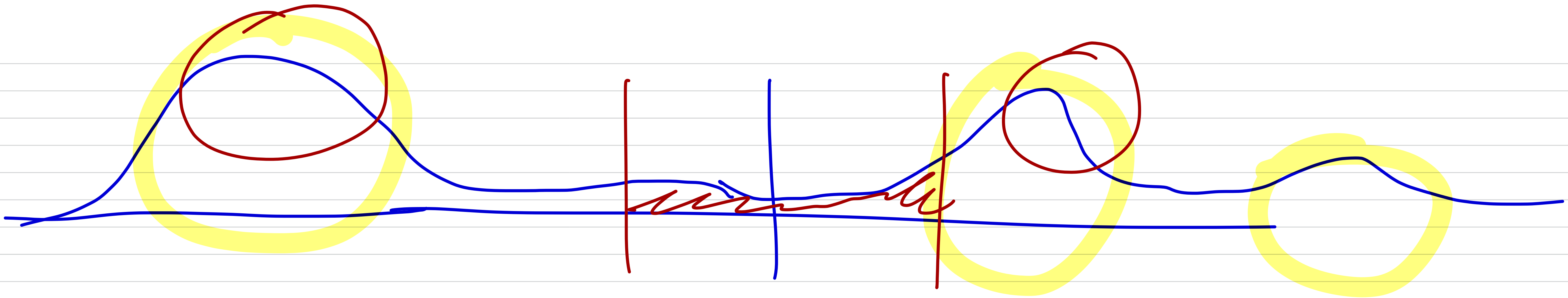


2. Nonlinear approximation

M-term approximation

replace $\mathcal{H}_M$ by Σ_M consisting of all $g \in \mathcal{H}$ that can be expressed as

$$g = \sum_{\lambda \in \Lambda} c_\lambda e_\lambda$$

with $\Lambda \subset \mathbb{N}$ s.t. $|\Lambda| \leq M$.

Def. 10.6. Given a function class $C \subset L^2(\Omega)$, a dictionary $\mathcal{D} = (\varphi_i)_{i \in I} \subset L^2(\Omega)$,

we define for $f \in C$, $M \in \mathbb{N}$,

$$\Pi_M^{\mathcal{D}}(f) = \inf_{\substack{I_M \subseteq I \\ \# I_M = M,\ (c_i)_{i \in I_M}}} \Big\| f - \sum_{i \in I_M} c_i \varphi_i \Big\|_{L^2(\Omega)}.$$

The supremal $r > 0$ s.t.

$$\sup_{f \in C} \Pi_M^{\mathcal{D}}(f) \in \mathcal{O}(M^{-r}),\ M \to \infty$$

will be denoted as $r^*(C, \mathcal{D})$.

Q: given the function class C, we are allowed to vary over $\mathcal{D}$, what is the largest $r^*(C, \mathcal{D})$ we can expect to get?

Take a dense (and countable) $\mathcal{D}$.

$M = 1 \Rightarrow$ approx. error ε can be made arbitrarily small $\Rightarrow$

$r^*(C, \mathcal{D}) = \infty$

$$\varepsilon \propto M^{-r} = \frac{1}{M^r}$$

Two issues: 1. would need to search a dictionary that has ∞ many elements

2. would need infinitely many bits to encode the optimal dictionary elements

---

polynomial depth search: $\pi(M)$

$\uparrow$

polynomial

e.g. M^3

Def. 10.7. $\mathcal{C}, \mathcal{D}$

$$\underbrace{\sup_{f \in \mathcal{C}} \inf_{\substack{I_M \subset \{1,2,\dots,\pi(M)\} \\ \# I_M = M,\ (c_i)_{i \in I_M}}} \| f - \sum_{i \in I_M} c_i \varphi_i \|_{L^2(\Omega)} \in \mathcal{O}(M^{-\gamma}),\quad M \to \infty}$$

$\gamma^{*,\mathrm{eff}}(\mathcal{C}, \mathcal{D})$ is the largest γ s.t.

Theorem 10.8. Let $d \in \mathbb{N}$ and $\Omega \subset \mathbb{R}^d$. The effective (polynomial depth-search) best M-term approximation rate of the function class $\mathcal{C} \subset L^2(\Omega)$ in the dictionary $\mathcal{D} \subset L^2(\Omega)$ satisfies

$$\gamma^{*,\mathrm{eff}}(\mathcal{C},\mathcal{D}) \leq \gamma^*(\mathcal{C}).$$

polynomial depth-search

Kolmogorov exponent

Proof. 1. encoding the indices of the dictionary elements participating in the best M-term approximation

no. of bits needed to describe an index in the set $\{1,2,\ldots,\pi(M)\}$ is given by $\log_2 \underbrace{\pi(M)}_{\leq M^p}$

$$\leq p \log_2(M)$$

\# bits is $\propto$ $n \log_2(n)$

2. encode the coefficients $(c_i)_{i \in I_n}$

$$\left(n \log_2(n), \quad n = \varepsilon^{-1/\gamma} \right.$$

$$\left. \quad \llcorner \ \varepsilon^{-1/\gamma} \log_2(\varepsilon^{-1/\gamma}) \in \Theta\left(\varepsilon^{-1/(\gamma-\delta)}\right) \right)$$

$$f_n = \sum_{i \in I_n} c_i \varphi_i \underset{\uparrow}{=} \sum_{i \in \hat{I}_n} \hat{c}_i \hat{\varphi}_i$$

G.S. orth.

$\hat{n} \leq n$

$\hat{\varphi}_i$ ONB for span $\{\varphi_i\}_{i \in I_n}$

$$e = f - \sum_{i \in \hat{I}_n} \hat{c}_i \hat{\varphi}_i = f - \sum_{i \in I_n} c_i \varphi_i$$

$$\Big\| \sum_{i \in \hat{I}_n} \hat{c_i}\, \hat{\varphi_i} \Big\| = \| f - e \| \le \|f\| + \|e\|$$

exploit that $\{\hat{\varphi_i}\}_{i \in \hat{I}_n}$ is an ONB

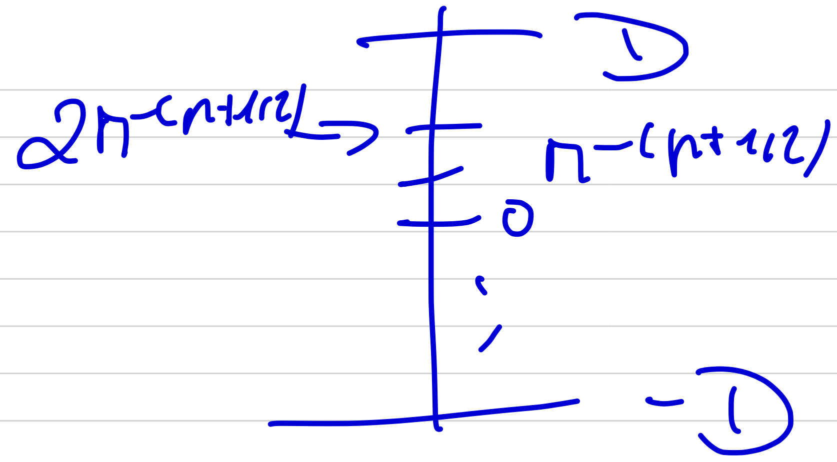
$$\Big\| \sum_{i \in \hat{I}_n} \hat{c_i}\, \hat{\varphi_i} \Big\|^2 \underset{\text{Parseval}}{=} \sum_{i \in \hat{I}_n} |\hat{c_i}|^2$$

$$\Big(\sum_{i \in \hat{I}_n} |\hat{c_i}|^2 \Big)^{1/2} \le \underbrace{\sup_{f \in C} \|f\|}_{< \infty} + \underbrace{\|e\|}_{\le C \eta^{-n}}$$

$\Rightarrow$ $\hat{c_i}$ are all bounded

$|\hat{c_i}| \le D < \infty$

quantize $\hat{c_i}$ to integer multiples of $\eta^{-(n+1/2)}$



$\Rightarrow$ total no. of qu. levels is $\propto M^{(n+1/2)}$

$$\log_2 M^{(n+1/2)} \propto \underset{(n+1/2)}{C} \log_2(M)$$

$$\#\text{bits} = C M \log_2(M)$$

$$\hat{c}_i = Q(\tilde{c}_i)$$

$$\underbrace{\Big\| f - \sum_{i\in \tilde{I}_n} \hat{c}_i \tilde{\varphi}_i \Big\|}_{\|D(E(f)) - f\|} = \Big\| f - \sum_{i\in \tilde{I}_n} \tilde{c}_i \tilde{\varphi}_i + \sum_{i\in \tilde{I}_n} \tilde{c}_i \tilde{\varphi}_i - \sum_{i\in \tilde{I}_n} \hat{c}_i \tilde{\varphi}_i \Big\|$$

$$\le \underbrace{\Big\| f - \sum_{i \in \hat{I}_n} \hat{c}_i \hat{\varphi}_i \Big\|}_{\le C n^{-r}} + \underbrace{\Big\| \sum_{i \in \hat{I}_n} (\hat{c}_i - c_i) \hat{\varphi}_i \Big\|}_{\le C' n^{-r}}$$

$$\Big\| \sum_{i \in \hat{I}_n} (\hat{c}_i - c_i) \hat{\varphi}_i \Big\| = \Big(\underbrace{\sum_{i \in \hat{I}_n}}_{\bar{n} \le n} \underbrace{|\hat{c}_i - c_i|^2}_{\le n^{-2r+1}} \Big)^{1/2}$$

$$\underbrace{\sum_{i \in \hat{I}_n} |\hat{c}_i - c_i|^2}_{n \cdot n^{-2r+1}} \qquad n \cdot n^{-2r+1} = n^{-2r}$$

$$\le \hat{C} n^{-r}$$

establishes achievability part of proof.

$$C \cdot M \log_2(M) = C\,\varepsilon^{-1/n} \log_2(\varepsilon^{-1/n}) \in \Theta(\varepsilon^{-1/(n-\delta)})$$

$$\varepsilon = M^{-n} \Rightarrow M = \varepsilon^{-1/n}$$

$$n \le n^{*,\mathrm{eff}}(C, D)$$

have constructed an encoder–decoder pair that achieves error behavior $\eta - (n^{*,\mathrm{eff}}(C,D) - \delta)$;

Converse: Couldn't we choose $n > n^{*}(C)$

$$L(\varepsilon, C) \in \Theta(\varepsilon^{-1/n})$$

$\Rightarrow$ $n > n^{*}(C)$ not possible

$$n^{*,\mathrm{eff}}(C, D) \le n^{*}(C).$$

Def. 10.9. If the effective best n-term approximation rate of $\mathcal{C}$ in $\mathcal{D}$ satisfies

$$\gamma^{*,\mathrm{eff}}(\mathcal{C},\mathcal{D}) = \gamma^{*}(\mathcal{C}),$$

then we say that $\mathcal{C}$ is optimally represented by $\mathcal{D}$.