$$c_2 \le c_0(\sigma,2) - c_1\left(1 + \frac{1+\log(1/2(\sigma))}{\log(n(\varepsilon))}\right) \quad \square$$

$m \lesssim \log(n)$.

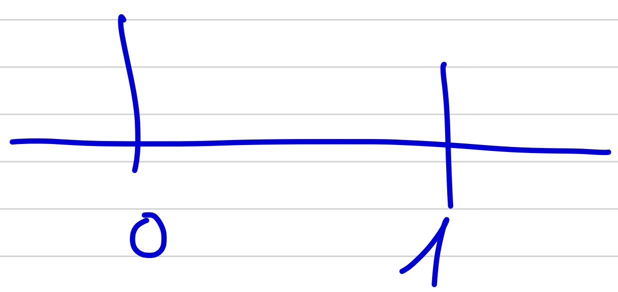
Chapter 10. Approximation Theory

Min-Max Kolmogorov Rate-Distortion Theory

Function class $\mathcal{C} \subset L^2(\Omega)$, $\Omega \subset \mathbb{R}^d$

$$\int_\Omega |f(x)|^2 dx < \infty$$

e.g. $[0,1] \Rightarrow \ell$ bits $\Rightarrow 2^\ell$ different values



set of binary encoders of length ℓ

$$\mathcal{E}^\ell := \{ E : C \to \{0,1\}^\ell \}$$

set of binary decoders of length ℓ

$$\mathcal{D}^\ell := \{ D : \{0,1\}^\ell \to L^2(\Omega) \}$$

uniform error ε over the function class C

$$\sup_{f \in C} \| D(E(f)) - f \|_{L^2(\Omega)} \leq \varepsilon .$$

Def. 10.1. Let $d \in \mathbb{N}$, $\Omega \subset \mathbb{R}^d$, and $C \subset L^2(\Omega)$. Then, for $\varepsilon > 0$, the minimax code length $L(\varepsilon, C)$ is

$$L(\varepsilon, C) := \min \{ \ell \in \mathbb{N} : \exists (E, D) \in \mathcal{E}^\ell \times \mathcal{D}^\ell : \sup_{f \in C} \| D(E(f)) - f \|_{L^2(\Omega)} \leq \varepsilon \}.$$

Moreover, the optimal exponent $p^*(C)$ is defined as

$$p^*(C) = \sup \{ p \in \mathbb{R} : L(\varepsilon, C) \in \Theta(\varepsilon^{-1/p}), \varepsilon \to 0 \}.$$

Larger $p \Rightarrow$ smaller growth rate!

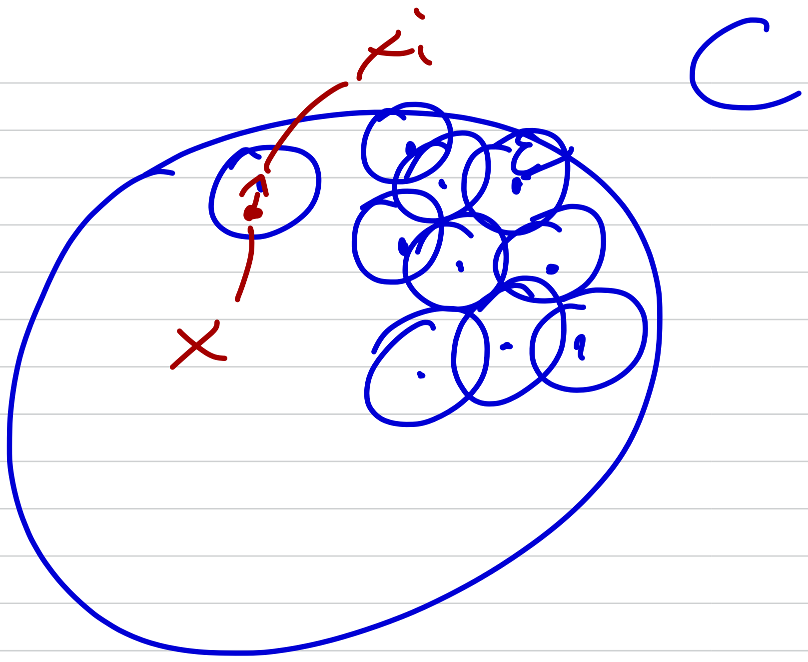
10.2. Metric entropy, covering, and packing

$(\mathcal{X}, \rho)$

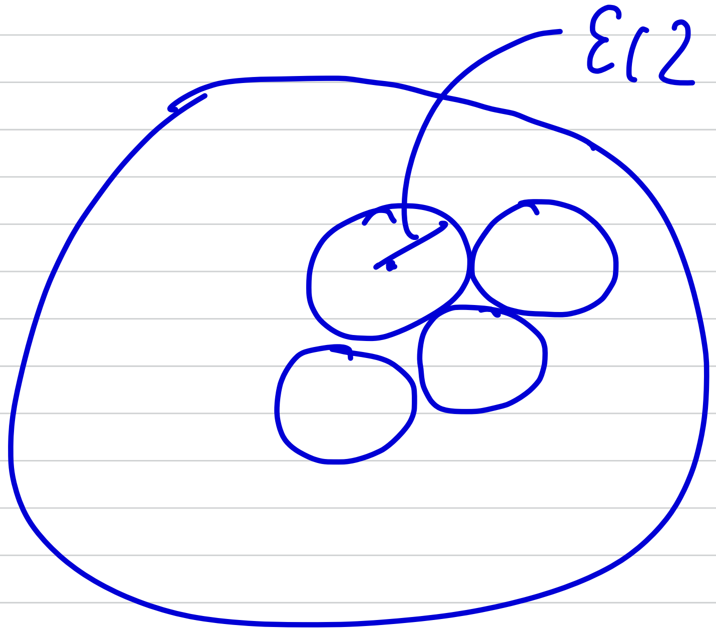
$\rho : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

1. $\rho(x, x') \geq 0, \forall x, x' \in \mathcal{X}$, with $=$ iff $x = x'$
2. $\rho(x, x') = \rho(x', x)$
3. Δ-inequ.: $\rho(x, \tilde{x}) \leq \rho(x, x') + \rho(x', \tilde{x})$, for all $x, x', \tilde{x}$

$L^2([0,1])$, $\|f - g\|_{L^2([0,1])} = \left[\int_0^1 (f(x) - g(x))^2 dx \right]^{1/2}$



$\rho(x, x_i) \leq \varepsilon$



covering

Def. 10.2. Let $(\mathcal{X}, \rho)$ be a metric space. An ε-covering of $C \subseteq \mathcal{X}$ w.r.t. the metric ρ is a set $\{x_1, \dots, x_N\} \subset C$ s.t. for each $x \in C$, there exists an $i \in [1, N]$ so that $\rho(x, x_i) \leq \varepsilon$. The ε-covering number $N(\varepsilon; C, \rho)$ is the cardinality of the smallest ε-covering.

$$C \subset \bigcup_{i=1}^{N} B(x_i, \varepsilon)$$

encoder: $x \in C$, map x to the closest (in ρ) ball center x_i, produce a binary representation of x by assigning it the binary representation of x_i

$N(\varepsilon; C, \rho)$ is the no. of ball centers

metric entropy $\longrightarrow$ $\log_2 N(\varepsilon; C, \rho)$ bits to label (in binary form) the ball centers

decoder: takes bit string of length $\log_2 N(\varepsilon; C, \rho)$ and maps it to the corr. x_i

$$D(E(x)) = x_i$$

Ex. $\log_2 (N; \varepsilon, \rho) \leq \log_2 \left(\frac{1}{\varepsilon} + 1 \right) \asymp \log_2 (\varepsilon^{-1})$

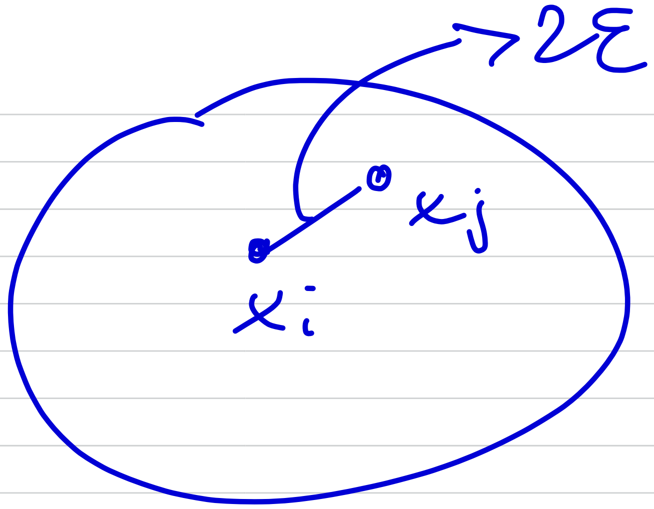
Def. 10.3. Let $(\mathcal{X}, \rho)$ be a metric space. An ε-packing for a compact set $C \subset \mathcal{X}$ w.r.t. the metric ρ is a set $\{x_1, \dots, x_N\} \subset C$ s.t. $\rho(x_i, x_j) > \varepsilon$, for all $i \neq j$. The ε-packing number $M(\varepsilon; \mathcal{X}, \rho)$ is the cardinality of the largest ε-packing.

Lemma 10.4. Let $(\mathcal{X}, \rho)$ be a metric space and C a compact set in $\mathcal{X}$. For all $\varepsilon > 0$, the packing and the covering number are related according to

$$M(2\varepsilon; C, \rho) \leq N(\varepsilon; C, \rho) \leq M(\varepsilon; C, \rho)$$

Proof. Choose minimal ε-covering and a maximal 2ε-packing of C.

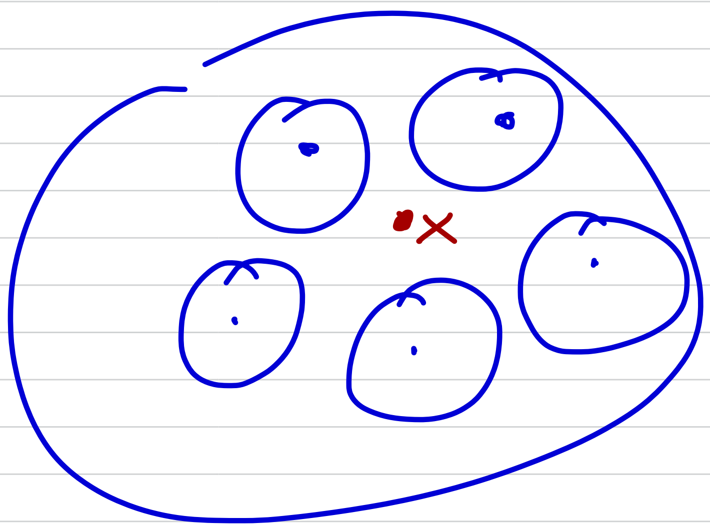
1. no two centers of the 2ε-packing can lie in the same ball of the ε-covering.



$$M(2\varepsilon; C, \ell) \le N(\varepsilon; C, \ell)$$

2. $N(\varepsilon; C, \ell) \le M(\varepsilon; C, \ell)$: given an ε-packing $M(\varepsilon; C, \ell)$

For given $x \in C$, we have the center of at least one of the balls of the ε-packing within dist. $< \varepsilon$.



$\Rightarrow$ ε- packing is also an ε- covering. $\square$

$$M(\varepsilon; C, \ell) \asymp \log_2\left(\frac{1}{\varepsilon}\right)$$

$$M(2\varepsilon) \le N(\varepsilon) \le M(\varepsilon) \Rightarrow N \asymp \log_2\left(\frac{1}{\varepsilon}\right)$$

Lemma 10.5. Consider $\|\cdot\|$, $\|\cdot\|'$ on $\mathbb{R}^d$, and let B & B' be their corr. unit balls, i.e.,

$$B = \{x \in \mathbb{R}^d \mid \|x\| \leq 1\}$$

$$B' = \{x \in \mathbb{R}^d \mid \|x\|' \leq 1\}$$

Then, the ε-covering number of B in the $\|\cdot\|'$-norm satisfies

$$\left(\frac{1}{\varepsilon}\right)^d \frac{\mathrm{vol}(B)}{\mathrm{vol}(B')} \leq N(\varepsilon; B, \|\cdot\|') \leq \frac{\mathrm{vol}\left(\frac{2}{\varepsilon}B + B'\right)}{\mathrm{vol}(B')}.$$

Proof. Let $\{x_1, \ldots, x_{N(\varepsilon; B, \|\cdot\|')}\}$ be an ε-covering of B in the $\|\cdot\|'$-norm. Then,

$$B \subseteq \bigcup_{j=1}^{N(\varepsilon; B, \|\cdot\|')} \{x_j + \varepsilon B'\} \quad \Big| \; \mathrm{vol}(\cdot)$$

$$\mathrm{vol}(B) \le N(\varepsilon; B, \|\cdot\|') \underbrace{\mathrm{vol}(\varepsilon B')}_{\varepsilon^d \, \mathrm{vol}(B')}$$

$$\Rightarrow \quad N(\varepsilon; B, \|\cdot\|') \ge \frac{\mathrm{vol}(B)}{\mathrm{vol}(B')} \left(\frac{1}{\varepsilon}\right)^d \quad ✓$$

maximal ε-packing $\{x_1, \dots, x_{M(\varepsilon; B, \|\cdot\|')}\}$ of B in the $\|\cdot\|'$-norm. The balls $\{x_j + \frac{\varepsilon}{2}B', j = 1, \dots, M(\varepsilon; B, \|\cdot\|')\}$ are disjoint and contained within $B + \frac{\varepsilon}{2}B'$.

Taking volumes, we get

$$\underbrace{\sum_{j=1}^{M(\varepsilon; B, \|\cdot\|')} \mathrm{vol}(x_j + \tfrac{\varepsilon}{2}B')}_{} \le \mathrm{vol}(B + \tfrac{\varepsilon}{2}B')$$

$$= M(\varepsilon; B, \|\cdot\|') \underbrace{\mathrm{vol}(\tfrac{\varepsilon}{2}B')}_{} \le \mathrm{vol}\left(\tfrac{\varepsilon}{2}\left(\tfrac{2}{\varepsilon}B + B'\right)\right)$$

$\left(\frac{\varepsilon}{2}\right)^d \mathrm{vol}\left(\frac{2}{\varepsilon}B+B'\right)$

$\left(\frac{\varepsilon}{2}\right)^d \mathrm{vol}(B')$

$$M(\varepsilon; B, \|\cdot\|') \geq N(\varepsilon; B, \|\cdot\|')$$

$$N(\varepsilon; B, \|\cdot\|') \leq \frac{\mathrm{vol}\left(\frac{2}{\varepsilon}B+B'\right)}{\mathrm{vol}(B')} \cdot D$$

metric entropy of unit ball in its own norm

$$B' = B : \mathrm{vol}\left(\frac{2}{\varepsilon}B + B'\right) = \mathrm{vol}\left(\left(\frac{2}{\varepsilon}+1\right)B\right)$$

$$= \left(\frac{2}{\varepsilon}+1\right)^d \mathrm{vol}(B)$$

$$\left(\frac{1}{\varepsilon}\right)^d \leq N(\varepsilon; B, \|\cdot\|) \leq \left(\frac{2}{\varepsilon}+1\right)^d \frac{\mathrm{vol}(B)}{\mathrm{vol}(B)}$$

$\mathcal{B}$: $\quad \underline{N(\varepsilon; \mathcal{B}, \|\cdot\|) \simeq \varepsilon^{-d}}$

$[0,1]^d$: $\quad N(\varepsilon; \mathcal{B}, \|\cdot\|) \simeq d \log(1/\varepsilon)$

Lipschitz functions: $\mathcal{F}_L := \{g : [0,1] \to \mathbb{R} \mid g(0) = 0,$

$|g(x) - g(x')| \leq L|x - x'|,$

$\forall x, x' \in [0,1]\}.$

$$\log_2(N(\varepsilon; \mathcal{F}_L, \|\cdot\|_\infty)) \simeq L/\varepsilon.$$

$\mathcal{F}_L([0,1]^d)$: $\quad \log_2(N(\varepsilon; \mathcal{F}_L, \|\cdot\|_\infty)) \simeq (L/\varepsilon)^d$

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

general scaling behavior: $\log_2(N(\varepsilon; C, \|\cdot\|)) \simeq \varepsilon^{-1/n} \log^{\beta}(\frac{1}{\varepsilon})$

— or some function scaling slower than $\varepsilon^{-1/p}$

$$\varepsilon^{-1/p} \log^{\beta}(\varepsilon^{-1}) \in \mathcal{O}\left(\varepsilon^{-1/(p^*(C))}\right)$$

10.3. Approximation with Representation Systems

$\mathcal{H}$ ... Hilbert space, $\langle \cdot, \cdot \rangle$, $\|\cdot\|$

$\{e_k\}_{k=1}^{\infty}$ an ONB for $\mathcal{H}$

1. Linear approximation

$\mathcal{H}_M := \operatorname{span}\{e_k : 1 \leq k \leq M\}$

$$E_M(f)_{\mathcal{H}} := \inf_{g \in \mathcal{H}_M} \|f - g\|_{\mathcal{H}}.$$

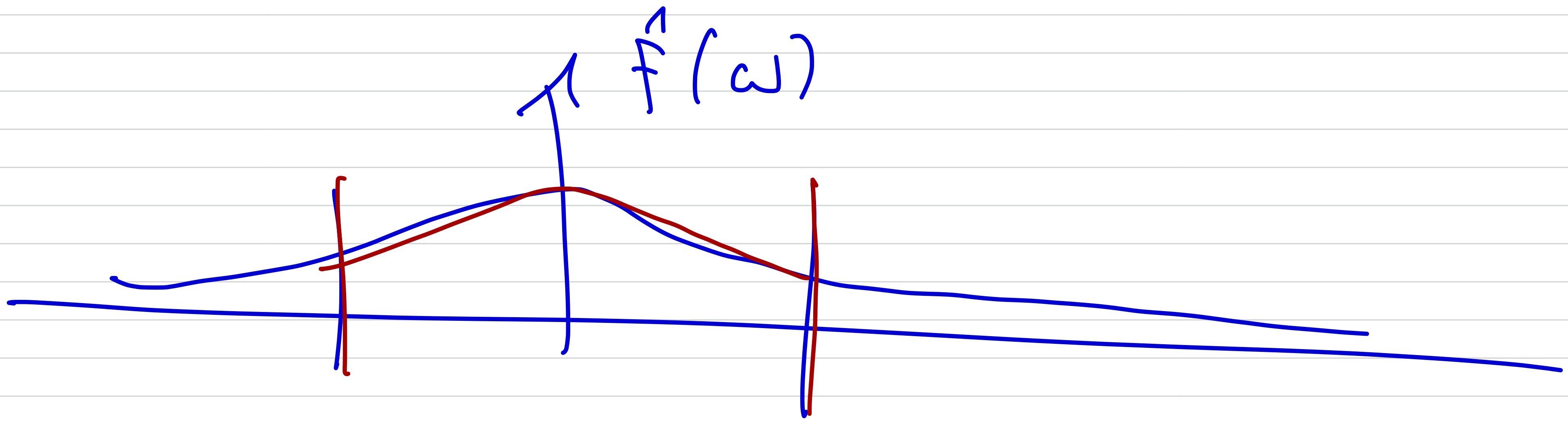
$g_1 \in \mathcal{H}_M,\ g_2 \in \mathcal{H}_M$:

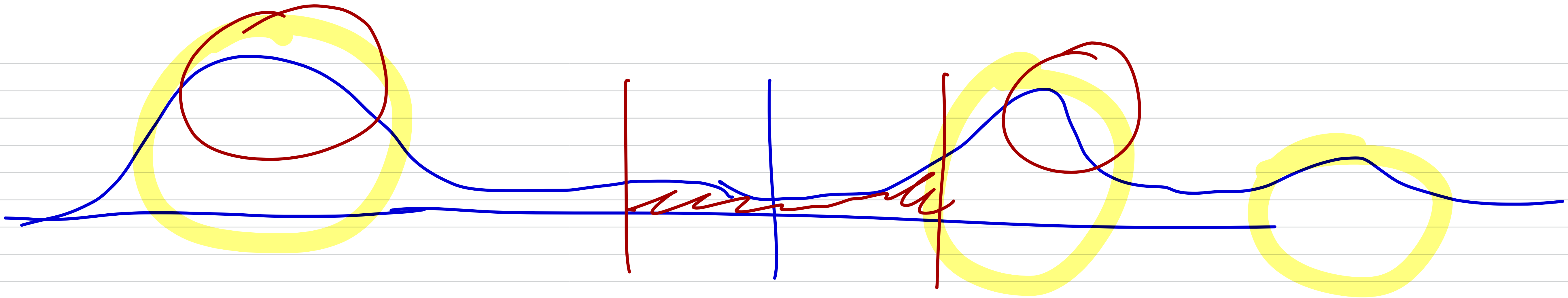
$$g_1 = \sum_{k=1}^{M} c_k^{(1)} g_k$$

$$g_2 = \sum_{k=1}^{M} c_k^{(2)} g_k$$

$$\alpha g_1 + \beta g_2 = \sum_{k=1}^{M} \underbrace{\left(\alpha c_k^{(1)} + \beta c_k^{(2)}\right)}_{\gamma_k} g_k$$

$\in \mathcal{H}_M$





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

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

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

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 $f \in C$, $M \in \mathbb{N}$,

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

The supremal $r > 0$ s.t.

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

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

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

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

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 $(c_i)_{i \in I_n}$

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

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

$$f_n = \sum_{i \in I_n} c_i \varphi_i = \sum_{i \in \widehat{I_n}} \widehat{c_i} \widehat{\varphi_i}$$

↑ G.S. orth.

$\widehat{n} \leq n$

$\widehat{\varphi_i}$ ONB for span $\{\varphi_i\}_{i \in I_n}$

$$e = f - \sum_{i \in \widehat{I_n}} \widehat{c_i} \widehat{\varphi_i} = f - \sum_{i \in I_n} c_i \varphi_i$$

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

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

$$\Big\| \sum_{i \in \widehat{I}_n} \widehat{c_i}\, \widehat{\varphi_i} \Big\|^2 = \sum_{i \in \widehat{I}_n} |\widehat{c_i}|^2$$

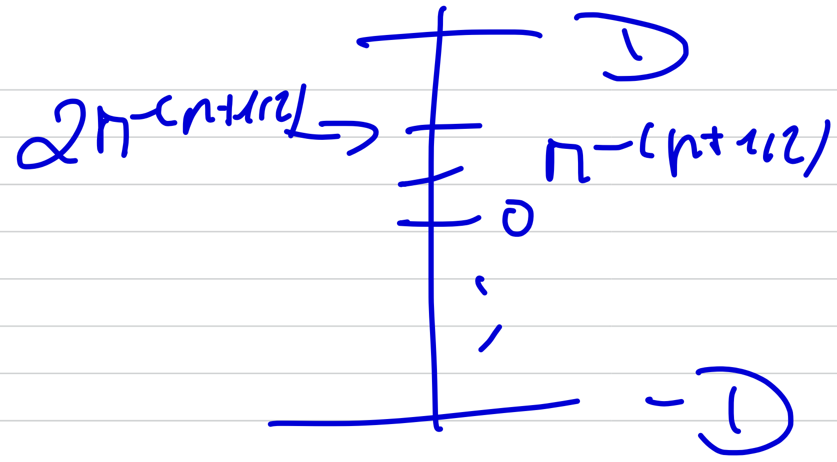
Parseval

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

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

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

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



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

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

$C = (n+1/2)$

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

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

$$\| f - \sum_{i \in \tilde{I}_M} \hat{c}_i \tilde{\varphi}_i \| = \| f - \sum_{i \in \tilde{I}_M} \tilde{c}_i \tilde{\varphi}_i + \sum_{i \in \tilde{I}_M} \tilde{c}_i \varphi_i - \sum_{i \in \tilde{I}_M} \hat{c}_i \tilde{\varphi}_i \|$$

$$\| D(E(f)) - f \|$$

$$\leq \| f - \sum_{i \in \hat{I}_n} \hat{c}_i \hat{\varphi}_i \| + \| \sum_{i \in \hat{I}_n} (\hat{c}_i - c_i) \hat{\varphi}_i \|$$

$$\underbrace{\| f - \sum_{i \in \hat{I}_n} \hat{c}_i \hat{\varphi}_i \|}_{\leq C M^{-r}}$$

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

$$\underbrace{\qquad}_{M \cdot M^{-2r+1} = M^{-2r}}$$

$$\underbrace{\qquad}_{\leq C' M^{-r}}$$

$$\leq \hat{C} M^{-r}$$

establishes achievability part of proof.

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

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

$$n \leq n^{*,\text{eff}}(C, D)$$

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

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

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

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

$$n^{*,\text{eff}}(C, D) \leq n^*(C).$$

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 $\mathcal{D}$.