$$c_2 \le \underline{C_0(\sigma/2) - c_1\left(1 + \frac{1+\log(1/2(\sigma))}{\log(n(R))}\right)} \quad . \square$$

$m \lesssim s \log(n)$.

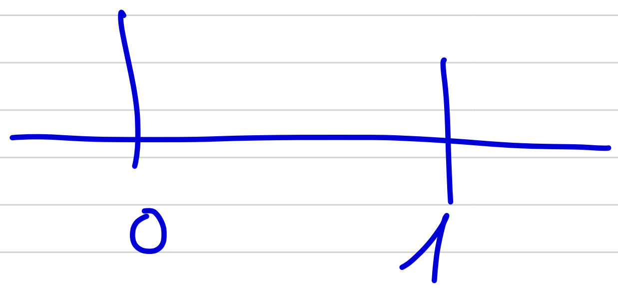
Chapter 10. Approximation Theory

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!

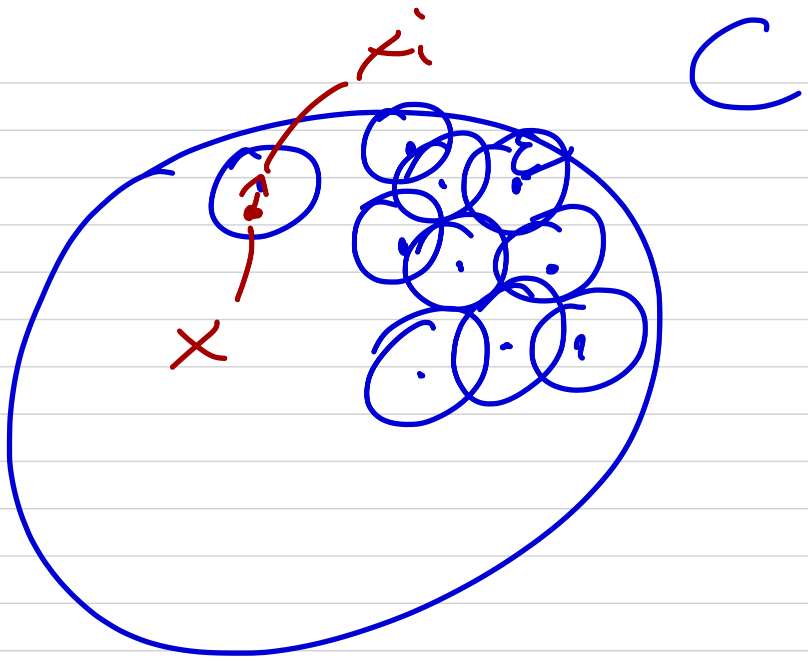
10.2. Metric entropy, covering, and packing

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

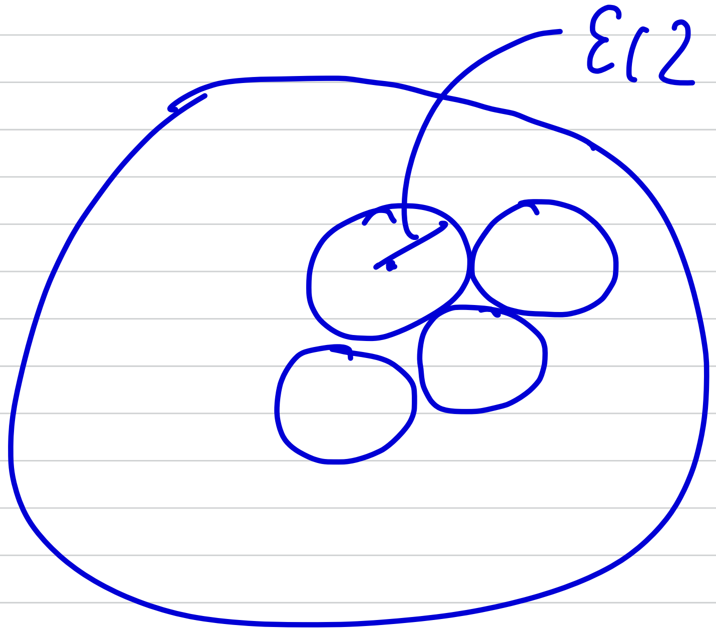
$\rho : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

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 $x, x', \tilde{x}$

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



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



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

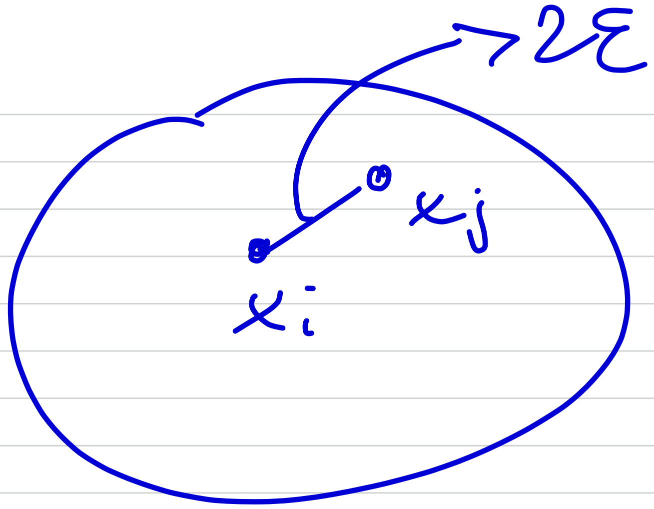
Def. 10.3. Let $(\mathcal{X}, \rho)$ be a metric space. An ε-packing for a compact set $C \subset \mathcal{X}$ w.r.t. the metric ρ is a set $\{x_1, \dots, x_N\} \subset C$ s.t. $\rho(x_i, x_j) > \varepsilon$, for all $i \neq j$. The ε-packing number $M(\varepsilon; \mathcal{X}, \rho)$ is the cardinality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

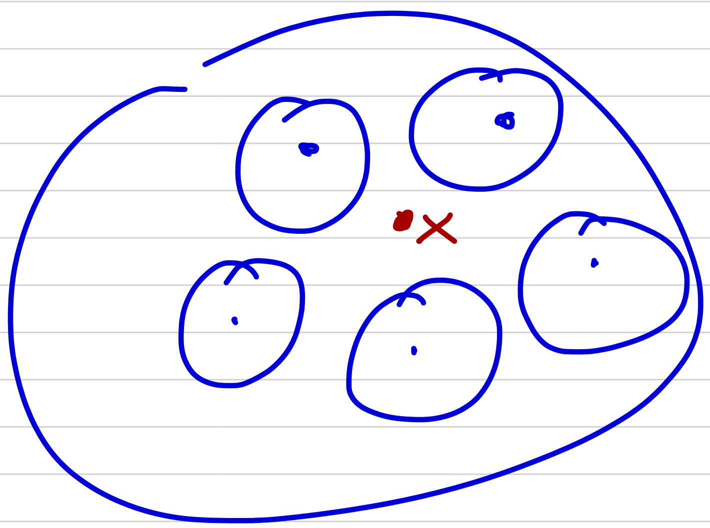
1. no two centers of the 2ε-packing can lie in the same ball of the ε-covering.



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

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

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

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

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

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

Taking volumes, we get

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

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

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

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

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

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

metric entropy of unit ball in its own norm

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

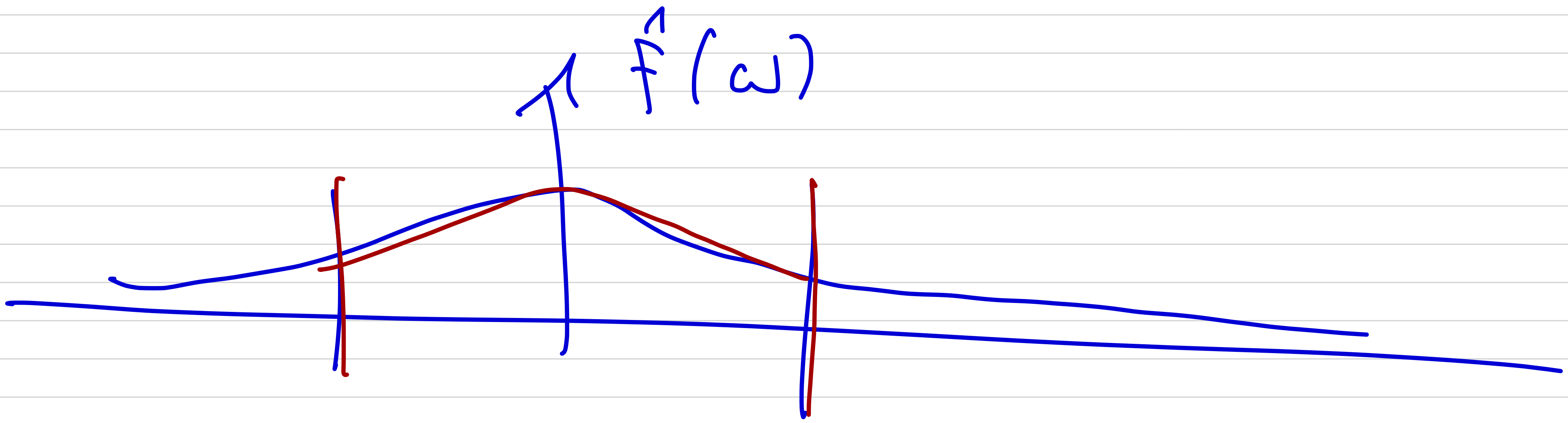
$g_1 \in \mathcal{H}_M,\ g_2 \in \mathcal{H}_M$:

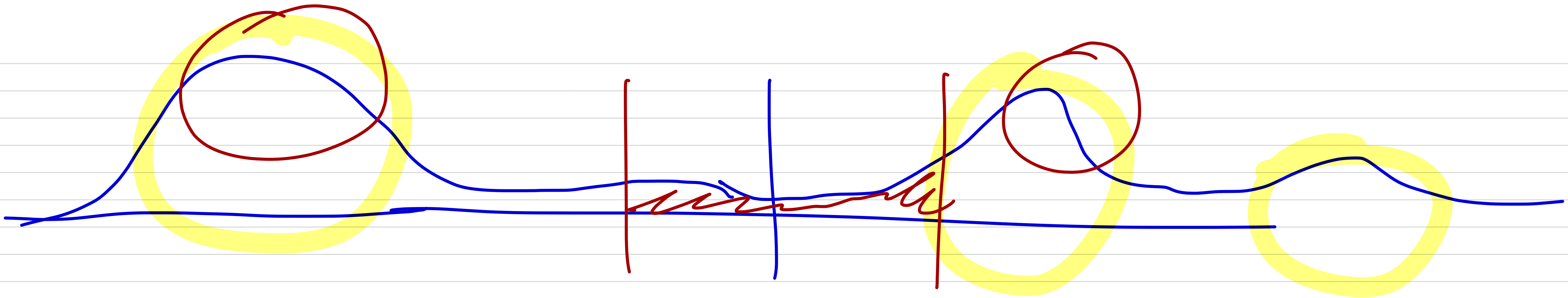
$$g_1 = \sum_{k=1}^{M} c_k^{(1)} g_k$$

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

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

$\in \mathcal{H}_M$





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

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

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