Three inequalities for sampling numbers

David Krieg

Nonlinear Approximation for High-Dimensional Problems.
Workshop in honor of Albert Cohen.
Paris, 30 Jun – 4 Jul 2025

UNIVERSITY OF PASSAU
Faculty of Computer Science and Mathematics

The problem of sampling recovery

Given: A domain D and a function $f\colon D \to \mathbb{C}$.

Task: Find a good approximation of f. The error is measured in $L_p(\mu)$ for some p and μ.

Problem: We do not know f and can only obtain samples $f(x_1), \ldots, f(x_m)$. Each sample is costly!

- What are good sampling points? What is the best we can do with m samples?

This question only make sense if we have some a priori knowledge about f (smoothness, structure, ...). The model class F is the class of all functions that satisfy the a priori knowledge.

Sampling numbers

We want to study the worst-case error of the best possible algorithm that uses at most m samples:

$$g_m(F, L_p) := \underbrace{\inf_{\substack{x_1,\dots,x_m \in D \\ \Phi\colon \mathbb{C}^m \to L_p}} \underbrace{\sup_{f \in F} \underbrace{\Big\| f - \overbrace{\overbrace{\Phi}^{\text{recovery map}} \overbrace{(f(x_1),\dots,f(x_m))}^{\text{samples}}}^{\text{sampling algorithm}} \Big\|_p}_{\text{approximation error in } L_p(\mu)}}_{\text{worst case over the model class}}}_{\text{worst case error of best sampling algorithm}}.$$

What model classes F do we consider?

- There are many results for particular (smoothness) classes F: Sobolev, Korobov, Gaussian, Hölder, Wiener, Besov, ...
- We do not want to consider specific classes, but rather study general relations to other approximation benchmarks.
- We discuss three different approaches.
- All three approaches work nicely if:

Global assumption

Let D be a compact metric space, μ a Borel probability measure, and F a compact subset of $C(D)$, the space of continuous functions.

1. Comparison with best uniform approximation

The n-th Kolmogorov number of F in $C(D)$ is

$$d_n(F, C(D)) := \underbrace{\inf_{\substack{V_n \subset C(D) \\ \dim(V_n)=n}} \underbrace{\sup_{f \in F} \underbrace{\inf_{g \in V_n} \|f - g\|_\infty}_{\text{best approximation from } V_n}}_{\text{worst case over model class}}}_{\text{best approximation space } V_n}.$$

- It describes the best uniform approximation of F by an n-dimensional space.
- By the compactness, we have $\lim_{n \to \infty} d_n(F, C(D)) = 0$.

Theorem 1

For any D, μ and F,

$$g_{4n}(F, L_p(\mu)) \leq 6 \cdot n^{(1/2-1/p)_+} \cdot d_n(F, C(D)).$$

Remark (references)

The case $p = 2$ was obtained by V. Temlayakov (JoC, 2021).

The case for general p is joint work with K. Pozharska, M. Ullrich, and T. Ullrich (JMAA, to appear).

Constants: work in progress with M. Dolbeault and M. Ullrich.

Theorem 1

For any D, μ and F,

$$g_{4n}(F, L_p(\mu)) \leq 6 \cdot n^{(1/2-1/p)_+} \cdot d_n(F, C(D)).$$

Remark (local version)

In fact, for any space $V_n \subset C(D)$, we find a plain least-squares estimator $\hat{f} := \operatorname{argmin}_{g \in V_n} \sum_{i=1}^{4n} |g(x_i) - f(x_i)|^2$ such that

$$\forall f \in C(D): \quad \|f - \hat{f}\|_p \leq 6 \cdot n^{(1/2-1/p)_+} \cdot \inf_{g \in V_n} \|f - g\|_\infty.$$

Theorem 1

For any D, μ and F,

$$g_{4n}(F, L_p(\mu)) \leq 6 \cdot n^{(1/2-1/p)_+} \cdot d_n(F, C(D)).$$

Remark (sharpness)

In general, the bound cannot be improved:

If F is the unit ball of the Sobolev space $W_1^s[0,1]$ (for $p \geq 2$) or $W_\infty^s[0,1]$ (for $p \leq 2$), then

$$g_n(F, L_p(\mu)) \asymp n^{(1/2-1/p)_+} \cdot d_n(F, C(D)).$$

Theorem 1

For any D, μ and F,

$$g_{2n}(F, C(D)) \leq 4\sqrt{n} \cdot d_n(F, C(D)).$$

Remark (related bounds for $p = \infty$)

Two related bounds by E. Novak (Springer Lecture Notes, 1988) and B. Kashin, S. Konyagin, V. Temlyakov (CA, 2023):

$$g_n(F, C(D)) \leq (n+1) \cdot d_n(F, C(D))$$
$$g_{9^n}(F, C(D)) \leq 5 \cdot d_n(F, C(D))$$

2. Comparison with best L_2-approximation

The n-th Kolmogorov number of F in $L_2(\mu)$ is

$$d_n(F, L_2(\mu)) := \underbrace{\inf_{\substack{V_n \subset C(D) \\ \dim(V_n)=n}} \underbrace{\sup_{f \in F} \underbrace{\inf_{g \in V_n} \|f-g\|_2}_{\text{best approximation from } V_n}}_{\text{worst case over model class}}}_{\text{best approximation space } V_n} .$$

- It describes the best $L_2(\mu)$ approximation of F by an n-dimensional space. We could restrict to linear approximation (the best approximation is given by the orthogonal projection).

Theorem 2

For any D, μ and F,

$$g_{8n}(F, L_2) \le \frac{20}{\sqrt{n}} \cdot \sum_{k \ge n} \frac{d_k(F, L_2)}{\sqrt{k}}.$$

Remark (references)

The result is basically from joint work with M. Dolbeault and M. Ullrich (ACHA, 2023).

Small improvement in a paper with K. Pozharska, M. Ullrich, and T. Ullrich (preprint).

Constants: work in progress with M. Dolbeault and M. Ullrich.

Theorem 2

For any D, μ and F,

$$g_{8n}(F, L_2) \leq \frac{20}{\sqrt{n}} \cdot \sum_{k \geq n} \frac{d_k(F, L_2)}{\sqrt{k}}.$$

Remark (tractability)

Requires

$$R := \sum_{k=1}^{\infty} \frac{d_k(F, L_2)}{\sqrt{k}} < \infty.$$

But then already a simplified formula can be useful in high dimensions:

$$g_{8n}(F, L_2(D)) \leq 20R \cdot n^{-1/2}.$$

Theorem 2

For any D, μ and F,

$$g_{8n}(F, L_2) \;\leq\; \frac{20}{\sqrt{n}} \cdot \sum_{k \geq n} \frac{d_k(F, L_2)}{\sqrt{k}}.$$

Remark (convergence rates)

In particular, if $d_n(F, L_2) \lesssim n^{-s} \log^r n$, then

$$g_n(F, L_2) \lesssim \begin{cases} n^{-s} \log^r n & \text{if } s > 1/2, \\ n^{-s} \log^{r+1} n & \text{if } s = 1/2 \text{ and } r < -1, \\ 1 & \text{otherwise.} \end{cases}$$

Moreover, there exist classes F such that these bounds are sharp.

Theorem 2

For any D, μ and F,

$$g_{8n}(F, L_2) \;\leq\; \frac{20}{\sqrt{n}} \cdot \sum_{k \geq n} \frac{d_k(F, L_2)}{\sqrt{k}}.$$

Remark (other approximation spaces)

In fact, for any sequence of nested approximation spaces $(V_n)_{n \in \mathbb{N}}$, we find a weighted least squares estimator

$$\hat{f} := \underset{g \in V_n}{\operatorname{argmin}} \sum_{i=1}^{8n} w_i |g(x_i) - f(x_i)|^2 \quad \text{s.t.}$$

$$\forall f \in F: \quad \|f - \hat{f}\|_p \;\leq\; \frac{20}{\sqrt{n}} \cdot \sum_{k \geq n} k^{-1/2} \Big(\sup_{f \in F} \operatorname{dist}_2(f, V_n) \Big).$$

Theorem 2

For any D, μ and F,

$$g_{8n}(F, L_2) \;\leq\; \frac{20}{\sqrt{n}} \cdot \sum_{k \geq n} \frac{d_k(F, L_2)}{\sqrt{k}}.$$

Remark (Hilbert case)

There is a slightly better formula for Hilbert spaces (see ACHA):

$$g_{4n}(F, L_2) \;\leq\; \frac{12}{\sqrt{n}} \cdot \sqrt{\sum_{k \geq n} d_k(F, L_2)^2}.$$

This bound is sharp up to the constants for any possible sequence of Kolmogorov numbers (joint work with J. Vybíral (JFAA, 2023)).

Theorem 2

For any D, μ and F,

$$g_{8n}(F, L_2) \leq \frac{20}{\sqrt{n}} \cdot \sum_{k \geq n} \frac{d_k(F, L_2)}{\sqrt{k}}.$$

Remark ($p \neq 2$)

If the approximation spaces are "nice enough", this can be turned into a bound for sampling numbers in $L_p(\mu)$ or other error norms (see preprint with K. Pozharska, M. Ullrich, and T. Ullrich). "Nice enough" means a good behavior of

$$\Lambda_n := \sup_{f \in V_n \setminus \{0\}} \frac{\|f\|_p}{\|f\|_2}.$$

Theorem 2

For any D, μ and F,

$$g_{8n}(F, L_2) \le \frac{20}{\sqrt{n}} \cdot \sum_{k \ge n} \frac{d_k(F, L_2)}{\sqrt{k}}.$$

Remark (randomized algorithms)

The expected error of randomized algorithm behaves even nicer. Here we have

$$g_{2n}^{\mathsf{ran}}(F, L_2) \le 4\, d_n(F, L_2).$$

See the papers by A. Cohen and M. Dolbeault (JoC, 2022) and A. Chkifa and M. Dolbeault (SIAM JoNA, 2024).

3. Comparison with best sparse approximation

Let $\mathcal{B}$ be a finite orthonormal system in $L_2(\mu)$ which consists of bounded functions and let

$$\Sigma_n(\mathcal{B}) := \bigcup_{b_1,\dots,b_n \in \mathcal{B}} \mathrm{Span}\{b_1, \dots, b_n\}.$$

The best n-term widths of F w.r.t. $\mathcal{B}$ are defined by

$$\sigma_n(F, \mathcal{B}) := \underbrace{\sup_{f \in F} \underbrace{\inf_{g \in \Sigma_n(\mathcal{B})} \|f - g\|_\infty}_{\text{best } n\text{-sparse approximation}}}_{\text{worst case over model class}} .$$

Theorem 3

For any $n \geq 2\max_{b\in\mathcal{B}} \|b\|_\infty^2$, *we have*

$$g_{m(n)}(F, L_p(\mu)) \;\leq\; C \cdot n^{(1/2-1/p)_+} \cdot \sigma_n(F,\mathcal{B}).$$

where $m(n) := Cn\log^3(n)\log(\#\mathcal{B})$.

Remarks

Note the similarity to Theorem 1 for the special case $\#\mathcal{B} = n$, where $\sigma_n(F,\mathcal{B}) = d_n(F, \mathrm{Span}(\mathcal{B}))$.

Because of the condition on n, the bound is only useful for special orthonormal systems (like Fourier or Walsh functions). The constant C is universal.

Theorem 3

For any $n \geq 2\max_{b\in\mathcal{B}} \|b\|_\infty^2$, *we have*

$$g_{m(n)}(F, L_p(\mu)) \;\leq\; C \cdot n^{(1/2-1/p)_+} \cdot \sigma_n(F,\mathcal{B}).$$

where $m(n) := Cn\log^3(n)\log(\#\mathcal{B})$.

Remarks (reference)

The bound follows from classical compressed sensing algorithms (ℓ_1-minimization, square-root lasso) and their analysis, for instance, Rauhut & Ward (2016).

We only reinterpret the samples $f(x_i)$, $f \in F$, as noisy samples of the sparse function $g(x_i)$, $g \in \Sigma_n$.

Theorem 3

For any $n \geq 2\max_{b\in\mathcal{B}} \|b\|_\infty^2$, *we have*

$$g_{m(n)}(F, L_p(\mu)) \;\leq\; C \cdot n^{(1/2-1/p)_+} \cdot \sigma_n(F,\mathcal{B}).$$

where $m(n) := Cn\log^3(n)\log(\#\mathcal{B})$.

Remarks (linear vs. nonlinear)

The approach was used by T. Jahn, T. Ullrich and F. Voigtlaender (JoC, 2023) to obtain new asymptotic bounds for the sampling numbers $W_p^{s,\mathrm{mix}}([0,1]^d)$, $p < 2$.

Any linear algorithm has a worse order of convergence.

Theorem 3

For any $n \geq 2 \max_{b \in \mathcal{B}} \|b\|_\infty^2$, *we have*

$$g_{m(n)}(F, L_p(\mu)) \leq C \cdot n^{(1/2-1/p)_+} \cdot \sigma_n(F, \mathcal{B}).$$

where $m(n) := Cn \log^3(n) \log(\#\mathcal{B})$.

Remarks (linear vs. nonlinear)

The approach was used by K. (PAMS, 2024) to obtain tractability results for L_p-approximation in Wiener-type function classes, e.g., the unit ball of

$$C^\alpha(\mathbb{T}^d) \cap \mathcal{A}(\mathbb{T}^d).$$

All linear algorithms suffer from the curse of dimensions.

Thank you for your attention!