Linear information versus function evaluations for L_2-approximation

Erich Novak

Mainly based on a paper with A. Hinrichs and J. Vybíral from 2008.

May 3, 2021

Intro 1

Algorithms for L_2-approximation (recovery) of functions.

Assume that f is from a *general* RKHS H of functions.

Worst case setting for the unit ball F of H, error of S_n is

$e(S_n) = \sup_{f \in F} \|f - S_n(f)\|_2$.

Two kinds of information and algorithms: $S_n(f) = \sum_{i=1}^n L_i(f) g_i$ can use general linear information $N(f) = (L_1(f), \dots, L_n(f)) \in \mathbb{R}^n$.

Often only $\widetilde{S}_n(f) = \sum_{i=1}^n f(x_i) g_i$ based on function values.

Question: Are these special algorithms $\widetilde{S}_n$ always, i.e., for any H, "almost as good" as general algorithms S_n?

Intro 2

First paper (for general H) seems to be Wasilkowski and Woźniakowski (2001), they prove *upper bounds*.

Study the optimal (polynomial) order of convergence of S_n and of $\widetilde{S}_n$. Can they be different?

When we started our work it was not known whether the two orders can be different.

Some notation

Assume that μ is a measure on a set D, consider $L_2 = L_2(D, \mu)$.

Hilbert space H of functions defined on D such that

function values $f \mapsto f(x)$ are continuous;

assume that the identity (embedding) $I : H \to L_2$ is compact.

Let F be the unit ball of H. Then the approximation numbers or linear widths $a_n(F)$ are defined as follows. For a continuous linear algorithm $S_n(f) = \sum_{i=1}^n L_i(f) g_i$ the (worst case) error is defined by $e(S_n) = \sup_{f \in F} \|f - S_n(f)\|_2$. Then $a_n(F)$ is given by

$$a_n(F) = \inf_{S_n} e(S_n) = \sigma_{n+1}$$

with the singular values σ_k.

Some notation

Usually algorithms $\widetilde{S}_n(f) = \sum_{i=1}^n f(x_i) g_i$ based on function values are preferred. Sampling numbers $g_n(F)$ are defined by $g_n(F) = \inf_{\widetilde{S}_n} e(\widetilde{S}_n)$. Often the $g_n(F)$ are only "slightly" larger than the $a_n(F)$. Define the "order of convergence" by

$$\mathrm{ord}(\mathrm{all}) = \sup \left\{ \alpha \ge 0 \, : \, \lim_{n\to\infty} a_n(F) \cdot n^\alpha = 0 \right\},$$

$$\mathrm{ord}(\mathrm{std}) = \sup \left\{ \alpha \ge 0 \, : \, \lim_{n\to\infty} g_n(F) \cdot n^\alpha = 0 \right\}.$$

We knew no example from the literature with $\mathrm{ord}(\mathrm{all}) > \mathrm{ord}(\mathrm{std})$.

Kuo, Wasilkowski, Woźniakowski (2009): $\mathrm{ord}(\mathrm{all}) > 1/2$ implies $\mathrm{ord}(\mathrm{std}) \ge \mathrm{ord}(\mathrm{all}) \cdot \frac{2\,\mathrm{ord}(\mathrm{all})}{2\,\mathrm{ord}(\mathrm{all})+1}$.

This lecture

We construct an artificial H such that

$$\text{ord}(\mathrm{all}) = \sup \left\{ \alpha \geq 0 \; : \; \lim_{n\to\infty} a_n(F) \cdot n^\alpha = 0 \right\} = \frac{1}{2}$$

and

$$\text{ord}(\mathrm{std}) = \sup \left\{ \alpha \geq 0 \; : \; \lim_{n\to\infty} g_n(F) \cdot n^\alpha = 0 \right\} = 0.$$

Later in the week, Mario Ullrich will report about his result with David Krieg (2019): If $\text{ord}(\mathrm{all}) > \frac{1}{2}$ then $\text{ord}(\mathrm{std}) = \text{ord}(\mathrm{all})$.

Deep recent *upper bounds* for the g_n by Cohen, Krieg, M. Ullrich, T. Ullrich, Temlyakov and others; we concentrate on *lower bounds*.

Finite dimensional case

Start with the finite dimensional case, $L_2 = \mathbb{R}^m = \ell_2^m$. Function evaluations correspond to scalar products with respect to a particular orthonormal system.

We assume that D has m elements and consider the mapping

$$I : \mathbb{R}^m \to \mathbb{R}^m = \ell_2^m$$

on an ellipsoid $F \subset \mathbb{R}^m$ of the form

$$F = \{f \in \mathbb{R}^m \mid f = \sum_{i=1}^m x_i e_i,\ \sum_{i=1}^m \frac{x_i^2}{\sigma_i^2} \leq 1\}.$$

Assume that the e_i form a complete ON-system and the singular values are ordered, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m \geq 0$.

Finite dimensional case

Then $x_i = \langle f, e_i \rangle$, with the scalar product in ℓ_2^m. The x_i are the coordinates of f with respect to the complete ON-system $\{e_i\}$.

Each f is a mapping from $D = \{1, 2, \dots, m\}$ to $\mathbb{R}$ and the function evaluations are the mappings

$$f \mapsto f_i = \langle f, b_i \rangle .$$

The $\{b_i\}$ form the standard basis of $\mathbb{R}^m = \ell_2^m$.

It holds $a_n(F) = \sigma_{n+1}$ and the optimal algorithm is

$$S_n^*(f) = \sum_{i=1}^{n} \langle f, e_i \rangle \ e_i .$$

Finite dimensional example

In the case of standard information we use functionals of the form $f \mapsto \langle f, b_i \rangle$. It seems that the numbers g_n are large if the b_i are "almost orthogonal" to the e_k. Hence we consider the following example.

We assume that the matrix which transforms $b_1, b_2, \ldots, b_m$ into $e_1, e_2, \ldots, e_m$ is a Hadamard matrix. Then we have formulas of the form

$$b_k = m^{-1/2} \cdot (\pm e_1 \pm e_2 \cdots \pm e_m)$$

and

$$e_k = m^{-1/2} \cdot (\pm b_1 \pm b_2 \cdots \pm b_m).$$

Some notation

Assume that m is of the form $m = 2^r$ and that the transformation $\{e_k\}_k \to \{b_k\}_k$ (and vice versa) is given by a Walsh-Hadamard matrix. Let

$$H^0 = (1), \qquad H^1 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \qquad H^{\ell+1} = \begin{pmatrix} H^\ell & H^\ell \\ H^\ell & -H^\ell \end{pmatrix}.$$

Then we have

$$m^{-1/2}\, H^r\, e_k = b_k \quad \text{and} \quad m^{-1/2}\, H^r\, b_k = e_k, \quad k = 1, 2, \ldots, m.$$

We need to do computations with those matrices ...

Lemma

Assume that the information consists in the function values with the numbers $k_1, k_2, \dots, k_n$ between 1 and m, where $m = 2^{n+t}$. Then there exist $2^t - 1$ numbers $\ell_1, \ell_2, \dots, \ell_{2^t-1}$ (different from 1) such that the information evaluated for e_1 coincides with the information evaluated for e_{ℓ_i} for each i. We can arrange that

$$\ell_1 \le 2^{n+1}, \qquad \ell_2, \ell_3 \le 2^{n+2}, \qquad \ell_4, \ell_5, \ell_6, \ell_7 \le 2^{n+3}$$

and so on. We get the zero information for a vector of the form

$$f = e_1 - c \sum_{i=1}^{2^t-1} \sigma_{\ell_i}^2 e_{\ell_i},$$

if c is chosen such that $c \sum_{i=1}^{2^t-1} \sigma_{\ell_i}^2 = 1$.

Result in finite dimension

Theorem: Assume that a sequence

$$1 = \sigma_1 \geq \sigma_2 \geq \dots$$

is given with $\sum_{k=1}^{\infty} \sigma_k^2 = \infty$. Assume further that a number n_0 and $\varepsilon > 0$ are given. Then there exists a (finite dimensional) example with $a_n(F) = \sigma_{n+1}$ for all $n \leq n_0$ and

$$g_{n_0}(F) \geq 1 - \varepsilon.$$

In this sense there does not exist any non-trivial upper bound for the $g_n(F)$ if $(\sigma_i)_i \notin \ell_2$.

Result in finite dimension: Upgrade 2021

Using results from HKNV21, JoC

Theorem: Assume that a sequence $1 = \sigma_1 \geq \sigma_2 \geq \dots$ is given with $\sum_{k=1}^{\infty} \sigma_k^2 = \infty$. Assume further that a number n_0 and $\varepsilon > 0$ are given. Then there exists a (finite dimensional) example with $a_{n_0}(F) \leq \sigma_{n_0+1}$ and $g_{n_0}(F) \geq 1 - \varepsilon$.

One can prove this also with standard tensor product spaces of trigon. polynomials of degree one: 1, $\sin(2\pi \cdot)$, $\cos(2\pi \cdot)$, or for larger (Sobolev) spaces.

In this sense: Function values are not enough for L_2-approximation, even for trigon. polyn. of degree 1.
For $\sigma \in \ell_2$ obtain examples with $g_n(F)^2 \geq 1 - \frac{n}{\|\sigma\|_2^2}$ for all n.

An infinite-dimensional example

Assume $\sigma_k = k^{-1/2}$. Use induction. We set $m_1 = 1$ and consider the 1-dimensional Hadamard example. Now assume that the first j building blocks with dimensions $m_1, m_2, \dots, m_j$ have already been constructed. We denote by H_j the corresponding sequence spaces. We set $D_j = m_1 + m_2 + \cdots + m_j$ and $n = 2^{D_j}$ and $t = \frac{2(D_j + 2^n)}{D_j}$ and $m_{j+1} = 2^{t+n}$.

The infinite-dimensional sequence space H is then defined as a direct sum of all the Hilbert spaces H_j. Very roughly

$$D_{j+1} \approx 2^{2^{2^{D_j}}}.$$

We obtain $g_n \geq (1 + \log_2 n)^{-1/2}$ for infinitely many n.

Main Result

Assume that $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$ with $\sum_n \sigma_n^2 = \infty$ and $\tau_1 \geq \tau_2 \geq \cdots \geq 0$ with $\lim_{n\to\infty} \tau_n = 0$.
Then there exists a RKHS such that for its unit ball F the following holds:

- $a_n(F) = \sigma_{n+1}$ for all natural numbers n,
- $g_n(F) \geq \tau_n$ for infinitely many natural numbers n.

On the curse of dimension

Assume that a sequence of spaces F_d is given, functions on $[0,1]^d$ or other domains.

Possible: using only function values the curse of dimension is present while the problem is tractable for general linear information.

This happens for certain (periodic and nonperiodic) Sobolev spaces with $\sum_k \sigma_k^2 < \infty$. Hence the g_n can be much worse than the a_n.

Only “small” n (such as $n < 10^{20}$) are practically relevant.

See Volume 3 (NW12) and a recent paper (2016) with Henryk.

L_q approximation for $q \neq 2$

Tandetzky (2012) studied embeddings of a RKHS $H \to L_q$ and proved:

For each q with $1 \leq q < \infty$ there exists an $H \subset L_q$ with

$$\mathrm{ord}(\mathrm{all}) = \min(1/q, 1/2)$$

and

$$\mathrm{ord}(\mathrm{std}) = 0.$$

Open (and once open) Problems on L_2-approximation

1. Is it true that $\mathrm{ord}(\mathrm{std}) = \mathrm{ord}(\mathrm{all})$ whenever $(\sigma_k)_k \in \ell_2$?
   From HNV08, also posed in NW12 as OP 126.

   The answer is yes, see Krieg and M. Ullrich (preprint 2019).

2. Assume $a_n \asymp n^{-r}(\log(n+1))^\beta$ with $r \geq 1/2$ and $\beta < -1/2$ if $r = 1/2$. Is it always true that $a_n \asymp g_n$?
   [OP 140 from NW12]

3. Assume $g_n \asymp n^{-r}(\log(n+1))^\beta$ with $r > 0$. Is it always true that $a_n \asymp g_n$?
   [OP 140 from NW12]

One may guess "yes" in 2) if $r > 1/2$ and "no" if $r = 1/2$.

See the new upper bounds of Krieg, M. Ullrich, T. Ullrich and Temlyakov.

References

HKNV21: Hinrichs, Krieg, Novak, Vybíral, J. Complexity.

HNV08: Hinrichs, Novak, Vybíral, J. Approx. Th.

KU19: Krieg, M. Ullrich, just appeared in FoCM.

KWW09: Kuo, Wasilkowski, Woźniakowski, J. Approx. Th.

NW12: Novak, Woźniakowski, Tractability of Multivariate Problems, Volume 3, EMS.

NW16: Novak, Woźniakowski, J. Approx. Th.

Ta12: Tandetzky, in the MCQMC 2010 Proceedings.

WW01: Wasilkowski, Woźniakowski, FoCM.