Some inequalities between characteristics of function classes

Vladimir Temlyakov

Moscow, 2022

Kolmogorov width and entropy

For two sets A and B in a Banch space X define the best approximation of A by B (the deviation of A from B)

$$d(A, B) := \sup_{a\in A} \inf_{b\in B} \|a - b\|.$$

Kolmogorov width of a centrally symmetric compact set $W \subset X$:

$$d_n(W, X) := \inf_{B-linear\ subspace,\, \dim B \le n} d(W, B).$$

Entropy numbers of a compact set $W \subset X$:

$$\varepsilon_k(W, X) := \inf_{B-finite\ set\ of\ points\ of\ cardinality\ |B| \le 2^k} d(W, B).$$

Known inequalities

The following result from B. Kashin and V. Temlyakov, 1995, was basically obtained in G.G. Lorentz, 1966.

Lemma (L1; Kashin and VT, 1995)

Let A be centrally symmetric compact in a separable Banach space X and for two real numbers $r > 0$ and $a \in \mathbb{R}$ we have

$$d_m(A, X) \ll m^{-r}(\log m)^a$$

and

$$\varepsilon_m(A, X) \gg m^{-r}(\log m)^a.$$

Then the following relations

$$d_m(A, X) \asymp \varepsilon_m(A, X) \asymp m^{-r}(\log m)^a$$

hold.

Carl's inequality

Carl's inequality (Carl, 1981) states: For any $r > 0$ we have

$$\max_{1\le k\le n} k^r \varepsilon_k(F,X) \le C(r) \max_{1\le m\le n} m^r d_{m-1}(F,X).$$

Numerical integration

For a function class $W \subset \mathcal{C}(\Omega)$ consider the best error of numerical integration by cubature formulas with m knots:

$$\kappa_m(W) := \inf_{(\xi,\Lambda)} \sup_{f \in W} |I_\mu(f) - \Lambda_m(f,\xi)|,$$

$$I_\mu(f) := \int_\Omega f d\mu, \quad \Lambda_m(f,\xi) := \sum_{j=1}^{m} \lambda_j f(\xi^j).$$

Novak's inequality

The following inequality was proved by E. Novak, 1986,

$$\kappa_m(W) \le 2d_m(W, L_\infty),$$

where $d_m(W, L_\infty)$ is the Kolmogorov width of W in the uniform norm L_∞.

Let $W \subset L_q(\Omega, \mu)$, $1 \leq q < \infty$, be a class of continuous on Ω functions. We are interested in estimating the following optimal errors of discretization of the L_q norm of functions from W

Sampling discretization with absolute error

Let $W \subset L_q(\Omega, \mu)$, $1 \le q < \infty$, be a class of continuous on Ω functions. We are interested in estimating the following optimal errors of discretization of the L_q norm of functions from W

$$er_m(W, L_q) := \inf_{\xi^1,\dots,\xi^m} \sup_{f \in W} \left| \|f\|_q^q - \frac{1}{m} \sum_{j=1}^{m} |f(\xi^j)|^q \right|,$$

Sampling discretization with absolute error

Let $W \subset L_q(\Omega, \mu)$, $1 \le q < \infty$, be a class of continuous on Ω functions. We are interested in estimating the following optimal errors of discretization of the L_q norm of functions from W

$$er_m(W, L_q) := \inf_{\xi^1,\dots,\xi^m} \sup_{f\in W} \left| \|f\|_q^q - \frac{1}{m}\sum_{j=1}^{m} |f(\xi^j)|^q \right|,$$

$$er_m^o(W, L_q) := \inf_{\xi^1,\dots,\xi^m;\lambda_1,\dots,\lambda_m} \sup_{f\in W} \left| \|f\|_q^q - \sum_{j=1}^{m} \lambda_j |f(\xi^j)|^q \right|.$$

General theorem

Theorem (T1; VT, 2018, 2022)

Assume that a class of real functions W *is such that for all* $f \in W$ *we have* $\|f\|_\infty \le M$ *with some constant* M. *Also assume that the entropy numbers of* W *in the uniform norm* L_∞ *satisfy the condition*

$$\varepsilon_n(W) \le n^{-r}(\log(n+1))^b, \qquad r \in (0,1/2), \quad b \ge 0, \quad n \in \mathbb{N}.$$

Then

$$er_m(W) := er_m(W, L_2) \le C(M,r,b)m^{-r}(\log(m+1))^b, \quad m \in \mathbb{N}.$$

Comments

Theorem T1 is a rather general theorem, which connects the behavior of absolute errors of discretization with the rate of decay of the entropy numbers. This theorem is derived from known results in supervised learning theory. It is well understood in learning theory that the entropy numbers of the class of priors (regression functions) is the right characteristic in studying the regression problem.

Comments

Theorem T1 is a rather general theorem, which connects the behavior of absolute errors of discretization with the rate of decay of the entropy numbers. This theorem is derived from known results in supervised learning theory. It is well understood in learning theory that the entropy numbers of the class of priors (regression functions) is the right characteristic in studying the regression problem.

- We impose a restriction $r < 1/2$ in Theorem T1 because the probabilistic technique from the supervised learning theory has a natural limitation to $r \leq 1/2$.

Comments

Theorem T1 is a rather general theorem, which connects the behavior of absolute errors of discretization with the rate of decay of the entropy numbers. This theorem is derived from known results in supervised learning theory. It is well understood in learning theory that the entropy numbers of the class of priors (regression functions) is the right characteristic in studying the regression problem.

- We impose a restriction $r < 1/2$ in Theorem T1 because the probabilistic technique from the supervised learning theory has a natural limitation to $r \leq 1/2$.
- It would be interesting to understand if Theorem T1 holds for $r \geq 1/2$.

Quasi-Monte Carlo methods

Define the best error of numerical integration by Quasi-Monte Carlo methods with m knots as follows

$$\kappa_m^Q(W) := \inf_{\xi^1,\dots,\xi^m} \sup_{f\in W} \left| \int_\Omega f d\mu - \frac{1}{m}\sum_{j=1}^m f(\xi^j) \right|.$$

Obviously, $\kappa_m^Q(W) \geq \kappa_m(W)$.

Assume that a class of real functions $W \subset \mathcal{C}(\Omega)$ has the following extra property.

Property A. For any $f \in W$ we have $f^+ := (f+1)/2 \in W$ and $f^- := (f-1)/2 \in W$.

In particular, this property is satisfied if W is a convex set containing functions 1 and -1.

Theorem (T2; VT, 2018)

Suppose $W \subset \mathcal{C}(\Omega)$ has Property A. Then for any $m \in \mathbb{N}$ we have

$$er^o_m(W, L_2) \geq \frac{1}{2}\kappa_m(W), \quad er_m(W, L_2) \geq \frac{1}{2}\kappa^Q_m(W).$$

Theorem (T3; VT, 2022)

Assume that a class of real functions $W \subset \mathcal{C}(\Omega)$ has Property A and is such that for all $f \in W$ we have $\|f\|_\infty \le M$ with some constant M. Also assume that the entropy numbers of W in the uniform norm L_∞ satisfy the condition

$$\varepsilon_n(W) \le n^{-r}(\log(n+1))^b, \qquad r \in (0,1/2), \quad b \ge 0.$$

Then

$$\kappa_m(W) \le \kappa_m^Q(W) \le C(M,r,b)m^{-r}(\log(m+1))^b.$$

Open problem 1. Does Theorem 3 hold for $r \geq 1/2$?

Open problems

Open problem 1. Does Theorem 3 hold for $r \geq 1/2$?
Open problem 2. Does Carl's inequality hold for sequences $\{\kappa_m(W)\}$ and $\{\varepsilon_n(W)\}$?

Open problems

Open problem 1. Does Theorem 3 hold for $r \geq 1/2$?
Open problem 2. Does Carl's inequality hold for sequences $\{\kappa_m(W)\}$ and $\{\varepsilon_n(W)\}$?
Open problem 3. Does Carl's inequality hold for sequences $\{\kappa_m^Q(W)\}$ and $\{\varepsilon_n(W)\}$?

Let $X \subset \mathbb{R}^d$, $Y \subset \mathbb{R}$ be Borel sets, ρ be a Borel probability measure on a Borel set $Z \subset X \times Y$. For $f : X \to Y$ define *the error*

$$\mathcal{E}(f) := \int_Z (f(\mathbf{x}) - y)^2 d\rho.$$

Let $X \subset \mathbb{R}^d$, $Y \subset \mathbb{R}$ be Borel sets, ρ be a Borel probability measure on a Borel set $Z \subset X \times Y$. For $f : X \to Y$ define *the error*

$$\mathcal{E}(f) := \int_Z (f(\mathbf{x}) - y)^2 d\rho.$$

Let ρ_X be the marginal probability measure of ρ on X, i.e., $\rho_X(S) = \rho(S \times Y)$ for Borel sets $S \subset X$. Define

$$f_\rho(\mathbf{x}) := \mathbb{E}(y|\mathbf{x})$$

to be a conditional expectation of y.

Setting

- The function f_ρ is known in statistics as the *regression function* of ρ. In the sense of error $\mathcal{E}(\cdot)$ the regression function f_ρ is the best to describe the relation between inputs $\mathbf{x} \in X$ and outputs $y \in Y$.

Setting

- The function f_ρ is known in statistics as the *regression function* of ρ. In the sense of error $\mathcal{E}(\cdot)$ the regression function f_ρ is the best to describe the relation between inputs $\mathbf{x} \in X$ and outputs $y \in Y$.
- The goal is to find an estimator $f_\mathbf{z}$, on the base of given data $\mathbf{z} := ((\mathbf{x}^1, y_1), \dots, (\mathbf{x}^m, y_m))$ that approximates f_ρ well with high probability.

Setting

- The function f_ρ is known in statistics as the *regression function* of ρ. In the sense of error $\mathcal{E}(\cdot)$ the regression function f_ρ is the best to describe the relation between inputs $\mathbf{x} \in X$ and outputs $y \in Y$.
- The goal is to find an estimator $f_{\mathbf{z}}$, on the base of given data $\mathbf{z} := ((\mathbf{x}^1, y_1), \dots, (\mathbf{x}^m, y_m))$ that approximates f_ρ well with high probability.
- We assume that $(\mathbf{x}^i, y_i)$, $i = 1, \dots, m$ are independent and distributed according to ρ.

Setting

- The function f_ρ is known in statistics as the *regression function* of ρ. In the sense of error $\mathcal{E}(\cdot)$ the regression function f_ρ is the best to describe the relation between inputs $\mathbf{x} \in X$ and outputs $y \in Y$.
- The goal is to find an estimator $f_{\mathbf{z}}$, on the base of given data $\mathbf{z} := ((\mathbf{x}^1, y_1), \dots, (\mathbf{x}^m, y_m))$ that approximates f_ρ well with high probability.
- We assume that $(\mathbf{x}^i, y_i)$, $i = 1, \dots, m$ are independent and distributed according to ρ.
- We measure the error between $f_{\mathbf{z}}$ and f_ρ in the $L_2(\rho_X)$ norm.

Empirical error

We define the *empirical error* of f as

$$\mathcal{E}_{\mathbf{z}}(f) := \frac{1}{m}\sum_{i=1}^{m}(f(\mathbf{x}^i) - y_i)^2.$$

Empirical error

We define the *empirical error* of f as

$$\mathcal{E}_{\mathbf{z}}(f) := \frac{1}{m}\sum_{i=1}^{m}(f(\mathbf{x}^i) - y_i)^2.$$

Let $f \in L_2(\rho_X)$. The *defect function* of f is

$$L_{\mathbf{z}}(f) := L_{\mathbf{z},\rho}(f) := \mathcal{E}(f) - \mathcal{E}_{\mathbf{z}}(f); \quad \mathbf{z} = (z_1, \dots, z_m), \quad z_i = (\mathbf{x}^i, y_i).$$

Empirical error

We define the *empirical error* of f as

$$\mathcal{E}_{\mathbf{z}}(f) := \frac{1}{m}\sum_{i=1}^{m}(f(\mathbf{x}^i) - y_i)^2.$$

Let $f \in L_2(\rho_X)$. The *defect function* of f is

$$L_{\mathbf{z}}(f) := L_{\mathbf{z},\rho}(f) := \mathcal{E}(f) - \mathcal{E}_{\mathbf{z}}(f); \quad \mathbf{z} = (z_1, \dots, z_m), \quad z_i = (\mathbf{x}^i, y_i).$$

We are interested in estimating $L_{\mathbf{z}}(f)$ for functions f coming from a given class W. We assume that ρ and W satisfy the following condition: for all $f \in W$ and any $(\mathbf{x}, y) \in Z$

$$|f(\mathbf{x}) - y| \le M. \tag{1}$$

Estimate for the defect function

Theorem (T4; S. Konyagin and VT, 2004)

Assume that ρ, W *satisfy (1) and* W *is such that*

$$\sum_{n=1}^{\infty} n^{-1/2}\varepsilon_n(W) = \infty.$$

For $\eta > 0$ *define* $J := J(\eta/M)$ *as the minimal* j *satisfying* $\varepsilon_{2^j} \le \eta/(8M)$ *and*

$$S_J := \sum_{j=1}^{J} 2^{(j+1)/2}\varepsilon_{2^{j-1}}.$$

Then for m, η *satisfying* $m(\eta/S_J)^2 \ge 480M^2$ *we have*

$$\rho^m\{\mathbf{z} : \sup_{f\in W} |L_{\mathbf{z}}(f)| \ge \eta\} \le C(M,\varepsilon(W)) \exp(-c(M)m(\eta/S_J)^2).$$

Corollary (C1; S. Konyagin and VT, 2004)

Assume ρ, W *satisfy (1) and* $\varepsilon_n(W) \le Dn^{-r}$, $r \in (0, 1/2)$. *Then for* m, η *satisfying* $m\eta^{1/r} \ge C_1(M, D, r)$ *we have*

$$\rho^m\{\mathbf{z} : \sup_{f\in W} |L_{\mathbf{z}}(f)| \ge \eta\} \le C(M, D, r)\exp(-c(M, D, r)m\eta^{1/r}).$$

Corollaries

Corollary (C2; VT, 2022)

Assume ρ, W *satisfy (1) and*

$$\varepsilon_n(W) \le n^{-r}(\log(n+1))^b, \qquad r \in (0, 1/2), \quad b \ge 0.$$

Then there are three positive constants $C_i = C_i(M, r, b)$, $i = 1, 2$, $c = c(M, r, b)$ *such that for* m, $\eta \le M$, *satisfying* $m\eta^{1/r}(\log(8M/\eta))^{-b/r} \ge C_1$ *we have*

$$\rho^m\{\mathbf{z} : \sup_{f \in W} |L_{\mathbf{z}}(f)| \ge \eta\} \le C_2 \exp(-cm\eta^{1/r}(\log(8M/\eta))^{-b/r}).$$

We begin with a very simple general observation on a connection between norm discretization and numerical integration.

Quasi-algebra property. We say that a function class W has the quasi-algebra property if there exists a constant a such that for any $f, g \in W$ we have $fg/a \in W$.

We begin with a very simple general observation on a connection between norm discretization and numerical integration.

Quasi-algebra property. We say that a function class W has the quasi-algebra property if there exists a constant a such that for any $f, g \in W$ we have $fg/a \in W$.

The above property was introduced and studied in detail by H. Triebel. He introduced this property under the name *multiplication algebra*. Normally, the term *algebra* refers to the corresponding property with parameter $a = 1$. To avoid any possible confusions we call it *quasi-algebra*. We refer the reader to the very resent book of Triebel, 2018, which contains results on the multiplication algebra (quasi-algebra) property for a broad range of function spaces.

Upper bound

Proposition (P1; VT, 2018)

Suppose that a function class W *has the quasi-algebra property and for any* $f \in W$ *we have for the complex conjugate function* $\bar{f} \in W$. *Then for a cubature formula* $\Lambda_m(\cdot, \xi)$ *we have: for any* $f \in W$

$$|\|f\|_2^2 - \Lambda_m(|f|^2, \xi)| \leq a \sup_{g \in W} \left| \int_\Omega g d\mu - \Lambda_m(g, \xi) \right|.$$

Function classes

We discuss some classical classes of smooth periodic functions. We begin with a general scheme and then give a concrete example. Let $F \in L_1(\mathbb{T}^d)$ be such that $\hat{F}(\mathbf{k}) \neq 0$ for all $\mathbf{k} \in \mathbb{Z}^d$, where

$$\hat{F}(\mathbf{k}) := \mathcal{F}(F, \mathbf{k}) := (2\pi)^{-d} \int_{\mathbb{T}^d} F(\mathbf{x}) e^{-i(\mathbf{k},\mathbf{x})} d\mathbf{x}.$$

We discuss some classical classes of smooth periodic functions. We begin with a general scheme and then give a concrete example. Let $F \in L_1(\mathbb{T}^d)$ be such that $\hat{F}(\mathbf{k}) \neq 0$ for all $\mathbf{k} \in \mathbb{Z}^d$, where

$$\hat{F}(\mathbf{k}) := \mathcal{F}(F, \mathbf{k}) := (2\pi)^{-d} \int_{\mathbb{T}^d} F(\mathbf{x}) e^{-i(\mathbf{k},\mathbf{x})} d\mathbf{x}.$$

Consider the space

$$W_2^F := \{f : f(\mathbf{x}) = J_F(\varphi)(\mathbf{x}) := (2\pi)^{-d} \int_{\mathbb{T}^d} F(\mathbf{x}-\mathbf{y})\varphi(\mathbf{y}) d\mathbf{y},$$

$$\|\varphi\|_2 < \infty\}.$$

Quasi-algebra property for function classes

For $f \in W_2^F$ we have $\hat{f}(\mathbf{k}) = \hat{F}(\mathbf{k})\hat{\varphi}(\mathbf{k})$ and, therefore, our assumption $\hat{F}(\mathbf{k}) \neq 0$ for all $\mathbf{k} \in \mathbb{Z}^d$ implies that function φ is uniquely defined by f. Introduce a norm on W_2^F by

$$\|f\|_{W_2^F} := \|\varphi\|_2, \qquad f = J_F(\varphi).$$

Quasi-algebra property for function classes

For $f \in W_2^F$ we have $\hat{f}(\mathbf{k}) = \hat{F}(\mathbf{k})\hat{\varphi}(\mathbf{k})$ and, therefore, our assumption $\hat{F}(\mathbf{k}) \neq 0$ for all $\mathbf{k} \in \mathbb{Z}^d$ implies that function φ is uniquely defined by f. Introduce a norm on W_2^F by

$$\|f\|_{W_2^F} := \|\varphi\|_2, \qquad f = J_F(\varphi).$$

For convenience, with a little abuse of notation we use notation W_2^F for the unit ball of the space W_2^F. We are interested in the following question. Under what conditions on F the fact that $f, g \in W_2^F$ implies that $fg \in W_2^F$ and

$$\|fg\|_{W_2^F} \leq C_0 \|f\|_{W_2^F} \|g\|_{W_2^F} ?$$

Quasi-algebra property for function classes

For $f \in W_2^F$ we have $\hat{f}(\mathbf{k}) = \hat{F}(\mathbf{k})\hat{\varphi}(\mathbf{k})$ and, therefore, our assumption $\hat{F}(\mathbf{k}) \neq 0$ for all $\mathbf{k} \in \mathbb{Z}^d$ implies that function φ is uniquely defined by f. Introduce a norm on W_2^F by

$$\|f\|_{W_2^F} := \|\varphi\|_2, \qquad f = J_F(\varphi).$$

For convenience, with a little abuse of notation we use notation W_2^F for the unit ball of the space W_2^F. We are interested in the following question. Under what conditions on F the fact that $f, g \in W_2^F$ implies that $fg \in W_2^F$ and

$$\|fg\|_{W_2^F} \leq C_0 \|f\|_{W_2^F} \|g\|_{W_2^F}?$$

In other words: Which properties of F guarantee that the class W_2^F has the quasi-algebra property? We give a simple sufficient condition.

Sufficient condition for quasi-algebra prperty

Proposition (P2; VT, 2018)

Suppose that for each $\mathbf{n} \in \mathbb{Z}^d$ *we have*

$$\sum_{\mathbf{k}\in\mathbb{Z}^d} |\hat{F}(\mathbf{k})\hat{F}(\mathbf{n}-\mathbf{k})|^2 \le C_0^2|\hat{F}(\mathbf{n})|^2. \qquad (2)$$

Then, for any $f, g \in W_2^F$ *we have* $fg \in W_2^F$ *and*

$$\|fg\|_{W_2^F} \le C_0\|f\|_{W_2^F}\|g\|_{W_2^F}.$$

Classes with mixed smoothness

As an example consider the class $\mathbf{W}_2^r$ of functions with bounded mixed derivative. By the definition $\mathbf{W}_2^r := W_2^{F_r}$ with function $F_r(\mathbf{x})$ defined as follows. For a number $k \in \mathbb{Z}$ denote $k^* := \max(|k|, 1)$. Then for $r > 0$ we define F_r by its Fourier coefficients

$$\hat{F}_r(\mathbf{k}) = \prod_{j=1}^{d} (k_j^*)^{-r}. \tag{3}$$

As an example consider the class $\mathbf{W}_2^r$ of functions with bounded mixed derivative. By the definition $\mathbf{W}_2^r := W_2^{F_r}$ with function $F_r(\mathbf{x})$ defined as follows. For a number $k \in \mathbb{Z}$ denote $k^* := \max(|k|, 1)$. Then for $r > 0$ we define F_r by its Fourier coefficients

$$\hat{F}_r(\mathbf{k}) = \prod_{j=1}^{d} (k_j^*)^{-r}. \tag{3}$$

Lemma (L1)

Function $F = F_r$ with $r > 1/2$ satisfies condition (2) with $C_0 = C(r, d)$.

Classes with mixed smoothness

As an example consider the class $\mathbf{W}^r_2$ of functions with bounded mixed derivative. By the definition $\mathbf{W}^r_2 := W^{F_r}_2$ with function $F_r(\mathbf{x})$ defined as follows. For a number $k \in \mathbb{Z}$ denote $k^* := \max(|k|, 1)$. Then for $r > 0$ we define F_r by its Fourier coefficients

$$\hat{F}_r(\mathbf{k}) = \prod_{j=1}^{d} (k_j^*)^{-r}. \tag{3}$$

Lemma (L1)

Function $F = F_r$ with $r > 1/2$ satisfies condition (2) with $C_0 = C(r, d)$.

Lemma L1 and Proposition P1 imply that the class $\mathbf{W}^r_2$ has the quasi-algebra property.

Smoothness classes

For classes of smooth functions we obtained error bounds, which do not have a restriction on smoothness r. We proved the following bounds for the class $\mathbf{W}^r_2$ of functions on d variables with bounded in L_2 mixed derivative.

For classes of smooth functions we obtained error bounds, which do not have a restriction on smoothness r. We proved the following bounds for the class $\mathbf{W}^r_2$ of functions on d variables with bounded in L_2 mixed derivative.

Theorem (T5; VT, 2018)

Let $r > 1/2$ and μ be the Lebesgue measure on $[0, 2\pi]^d$. Then

$$er^o_m(\mathbf{W}^r_2, L_2) \asymp m^{-r}(\log m)^{(d-1)/2}.$$

An interesting phenomenon

There are results (see G.W. Wasilkowski, 1984) on optimal estimation of the $\|f\|$ under assumption that $f \in W$.

- At a first glance the problems of estimation of $\|f\|$ and, say, estimation of $\|f\|^2$, like in our case, are very close.

An interesting phenomenon

There are results (see G.W. Wasilkowski, 1984) on optimal estimation of the $\|f\|$ under assumption that $f \in W$.

- At a first glance the problems of estimation of $\|f\|$ and, say, estimation of $\|f\|^2$, like in our case, are very close.
- A simple inequality $|a^2 - b^2| \leq 2M|a-b|$ for numbers satisfying $|a| \leq M$ and $|b| \leq M$ shows that normally we can get an upper bound for estimation of $\|f\|^2$ in terms of the error of estimation of $\|f\|$.

However, it turns out that the above two problems are different.

An interesting phenomenon continue

- It is proved in G.W. Wasilkowski, 1984 that the error of optimal estimation of the $\|\cdot\|$ is of the same order as the optimal error of approximation. For instance, in case of the class $\mathbf{W}^r_2$ this error is of the order $m^{-r}(\log m)^{r(d-1)}$, which is larger than the corresponding error $er^o_m(\mathbf{W}^r_2, L_2)$ in Theorem T2.

An interesting phenomenon continue

- It is proved in G.W. Wasilkowski, 1984 that the error of optimal estimation of the $\|\cdot\|$ is of the same order as the optimal error of approximation. For instance, in case of the class $\mathbf{W}^r_2$ this error is of the order $m^{-r}(\log m)^{r(d-1)}$, which is larger than the corresponding error $er^o_m(\mathbf{W}^r_2, L_2)$ in Theorem T2.
- The above Example shows that the optimal error for estimation of the $\|f\|$ may be different from the optimal error of estimation of the $\|f\|^2$.

An interesting phenomenon continue

- It is proved in G.W. Wasilkowski, 1984 that the error of optimal estimation of the $\|\cdot\|$ is of the same order as the optimal error of approximation. For instance, in case of the class $\mathbf{W}_2^r$ this error is of the order $m^{-r}(\log m)^{r(d-1)}$, which is larger than the corresponding error $er_m^o(\mathbf{W}_2^r, L_2)$ in Theorem T2.
- The above Example shows that the optimal error for estimation of the $\|f\|$ may be different from the optimal error of estimation of the $\|f\|^2$.
- Detailed comparison of my paper with G.W. Wasilkowski, 1984 shows that the problems of optimal errors in estimation of $\|f\|$ and $\|f\|^2$ are different.

Thank you!