Sampling discretization of integral norms. Lecture 3

Vladimir Temlyakov

Chemnitz; September, 2019

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

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)

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*

$$\varepsilon_n(W, L_\infty) \le Cn^{-r}, \qquad r \in (0, 1/2).$$

Then

$$er_m(W) := er_m(W, L_2) \le Km^{-r}.$$

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 $r \geq 1/2$.

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 $r \le 1/2$.
- It would be interesting to understand if Theorem T1 holds for $r \ge 1/2$.
- Also, it would be interesting to obtain an analog of Theorem T1 for discretization in the L_q, $1 \le q < \infty$, norm.

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 (T2; VT, 2018)

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

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

Marcinkiewicz problem

Let Ω be a compact subset of $\mathbb{R}^d$ with the probability measure μ. We say that a linear subspace X_N of the $L_q(\Omega)$, $1 \le q < \infty$, admits the Marcinkiewicz-type discretization theorem with parameters m and q if there exist a set $\{\xi^\nu \in \Omega, \nu = 1, \dots, m\}$ and two positive constants $C_j(d,q)$, $j = 1,2$, such that for any $f \in X_N$ we have

$$C_1(d,q)\|f\|_q^q \le \frac{1}{m}\sum_{\nu=1}^{m} |f(\xi^\nu)|^q \le C_2(d,q)\|f\|_q^q. \tag{1}$$

Marcinkiewicz problem

Let Ω be a compact subset of $\mathbb{R}^d$ with the probability measure μ. We say that a linear subspace X_N of the $L_q(\Omega)$, $1 \le q < \infty$, admits the Marcinkiewicz-type discretization theorem with parameters m and q if there exist a set $\{\xi^\nu \in \Omega, \nu = 1, \dots, m\}$ and two positive constants $C_j(d,q)$, $j = 1, 2$, such that for any $f \in X_N$ we have

$$C_1(d,q)\|f\|_q^q \le \frac{1}{m}\sum_{\nu=1}^{m} |f(\xi^\nu)|^q \le C_2(d,q)\|f\|_q^q. \qquad (1)$$

In the case $q = \infty$ we define L_∞ as the space of continuous on Ω functions and ask for

$$C_1(d)\|f\|_\infty \le \max_{1\le\nu\le m} |f(\xi^\nu)| \le \|f\|_\infty. \qquad (2)$$

Marcinkiewicz problem

Let Ω be a compact subset of $\mathbb{R}^d$ with the probability measure μ. We say that a linear subspace X_N of the $L_q(\Omega)$, $1 \le q < \infty$, admits the Marcinkiewicz-type discretization theorem with parameters m and q if there exist a set $\{\xi^\nu \in \Omega, \nu = 1, \dots, m\}$ and two positive constants $C_j(d,q)$, $j = 1, 2$, such that for any $f \in X_N$ we have

$$C_1(d,q)\|f\|_q^q \le \frac{1}{m}\sum_{\nu=1}^{m} |f(\xi^\nu)|^q \le C_2(d,q)\|f\|_q^q. \tag{1}$$

In the case $q = \infty$ we define L_∞ as the space of continuous on Ω functions and ask for

$$C_1(d)\|f\|_\infty \le \max_{1\le\nu\le m} |f(\xi^\nu)| \le \|f\|_\infty. \tag{2}$$

We will also use a brief way to express the above property: the $\mathcal{M}(m,q)$ theorem holds for a subspace X_N or $X_N \in \mathcal{M}(m,q)$.

We say that a linear subspace X_N of the $L_q(\Omega)$, $1 \le q < \infty$, admits the weighted Marcinkiewicz-type discretization theorem with parameters m and q if there exist a set of knots $\{\xi^\nu \in \Omega\}$, a set of weights $\{\lambda_\nu\}$, $\nu = 1, \dots, m$, and two positive constants $C_j(d, q)$, $j = 1, 2$, such that for any $f \in X_N$ we have

$$C_1(d,q)\|f\|_q^q \le \sum_{\nu=1}^{m} \lambda_\nu |f(\xi^\nu)|^q \le C_2(d,q)\|f\|_q^q. \qquad (3)$$

We say that a linear subspace X_N of the $L_q(\Omega)$, $1 \le q < \infty$, admits the weighted Marcinkiewicz-type discretization theorem with parameters m and q if there exist a set of knots $\{\xi^\nu \in \Omega\}$, a set of weights $\{\lambda_\nu\}$, $\nu = 1, \dots, m$, and two positive constants $C_j(d,q)$, $j = 1,2$, such that for any $f \in X_N$ we have

$$C_1(d,q)\|f\|_q^q \le \sum_{\nu=1}^{m} \lambda_\nu |f(\xi^\nu)|^q \le C_2(d,q)\|f\|_q^q. \tag{3}$$

Then we also say that the $\mathcal{M}^w(m,q)$ theorem holds for a subspace X_N or $X_N \in \mathcal{M}^w(m,q)$. Obviously, $X_N \in \mathcal{M}(m,q)$ implies that $X_N \in \mathcal{M}^w(m,q)$.

We write $X_N \in \mathcal{M}(m,q,\varepsilon)$ if (1) holds with $C_1(d,q) = 1-\varepsilon$ and $C_2(d,q) = 1+\varepsilon$. Respectively, we write $X_N \in \mathcal{M}^w(m,q,\varepsilon)$ if (3) holds with $C_1(d,q) = 1-\varepsilon$ and $C_2(d,q) = 1+\varepsilon$.

We write $X_N \in \mathcal{M}(m,q,\varepsilon)$ if (1) holds with $C_1(d,q) = 1-\varepsilon$ and $C_2(d,q) = 1+\varepsilon$. Respectively, we write $X_N \in \mathcal{M}^w(m,q,\varepsilon)$ if (3) holds with $C_1(d,q) = 1-\varepsilon$ and $C_2(d,q) = 1+\varepsilon$.
We note that the most powerful results are for $\mathcal{M}(m,q,0)$, when the L_q norm of $f \in X_N$ is discretized exactly by the formula with equal weights $1/m$.

Sampling discretization is a natural way of estimating the quantity of interest $\|f\|_q^q$. Certainly, one can ask a question of optimal estimation of $\|f\|_q^q$ using m function values or, even more general, using m linear functionals. It is an interesting problem but we do not address it in this talk. We only point out on a simple example that we obtain very different results when we allow arbitrary linear functionals to be used.

Example

Consider the class $\mathbf{W}^r_2$ of periodic functions with bounded in L_2 the rth mixed derivative. For $N \in \mathbb{N}$ define the hyperbolic cross

$$\Gamma(N) := \{\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{Z}^d \ : \ \prod_{j=1}^{d} \max(1, |k_j|) \le N\}$$

and for $f \in L_1$

$$S_N(f) := \sum_{\mathbf{k} \in \Gamma(N)} \hat{f}(\mathbf{k}) e^{i(\mathbf{k},\mathbf{x})}, \quad \hat{f}(\mathbf{k}) := (2\pi)^{-d} \int_{[0,2\pi]^d} f(\mathbf{x}) e^{-i(\mathbf{k},\mathbf{x})} d\mathbf{x}.$$

Then, it is well known and easy to prove that for any $f \in \mathbf{W}_2^r$ we have

$$0 \leq \|f\|_2^2 - \|S_N(f)\|_2^2 \leq N^{-2r}. \tag{4}$$

With this algorithm we use $m \asymp N(\log N)^{d-1}$ linear functionals $\hat{f}(\mathbf{k})$, $\mathbf{k} \in \Gamma(N)$. The bound (4) is very different from the asymptotic behavior in Theorem T2.

A remark on general inequalities

- The sampling discretization errors $er_m(W, L_q)$ and $er_m^o(W, L_q)$ are new asymptotic characteristics of a function class W.

A remark on general inequalities

- The sampling discretization errors $er_m(W, L_q)$ and $er^o_m(W, L_q)$ are new asymptotic characteristics of a function class W.
- It is natural to try to compare these characteristics with other classical asymptotic characteristics.

A remark on general inequalities

- The sampling discretization errors $er_m(W, L_q)$ and $er^o_m(W, L_q)$ are new asymptotic characteristics of a function class W.
- It is natural to try to compare these characteristics with other classical asymptotic characteristics.
- Theorem T1 addresses this issue. It is known that the sequence of entropy numbers is one of the smallest sequences of asymptotic characteristics of a class. For instance, by Carl's inequality it is dominated, in a certain sense, by the sequence of the Kolmogorov widths.

A remark on general inequalities continue

- Theorem T1 shows that the sequence $\{\varepsilon_n(W)\}$ dominates, in a certain sense, the sequence $\{er_m(W)\}$.

A remark on general inequalities continue

- Theorem T1 shows that the sequence $\{\varepsilon_n(W)\}$ dominates, in a certain sense, the sequence $\{er_m(W)\}$.
- Clearly, alike the Carl's inequality, one tries to prove the corresponding relations in as general situation as possible.

A remark on general inequalities continue

- Theorem T1 shows that the sequence $\{\varepsilon_n(W)\}$ dominates, in a certain sense, the sequence $\{er_m(W)\}$.
- Clearly, alike the Carl's inequality, one tries to prove the corresponding relations in as general situation as possible.
- We derive Theorem T1 from known results in learning theory. Our proof is a probabilistic one. The use of that kind of technique results in the limitation $r \in (0, 1/2)$ for the power in the rate of decay of the entropy numbers. As we pointed out above, we do not know if one can prove an analog of Theorem T1 in the case $r > 1/2$.

Known results on the asymptotic characteristics of the univariate class $\mathbf{W}^r_p$ show that Theorem T1 cannot be improved. Namely, on one hand it is known that

$$\varepsilon_n(\mathbf{W}^r_p, L_\infty) \asymp n^{-r}, \quad r > 1/p, \quad 1 \le p \le \infty. \tag{5}$$

Known results on the asymptotic characteristics of the univariate class $\mathbf{W}^r_p$ show that Theorem T1 cannot be improved. Namely, on one hand it is known that

$$\varepsilon_n(\mathbf{W}^r_p, L_\infty) \asymp n^{-r}, \quad r > 1/p, \quad 1 \le p \le \infty. \tag{5}$$

For $2 < p < \infty$ and $r \in (1/p, 1/2)$ relation (5) and Theorem T1 imply

$$er_m(\mathbf{W}^r_p) \le C(r,p)m^{-r}.$$

On the other hand, 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 $f^- := (f-1)/2 \in W$.

T1 is optimal. Lower bound

On the other hand, 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 function 1.

T1 is optimal. Lower bound

On the other hand, 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 function 1.

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

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

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

Connection to numerical integration

Theorem (T3; VT, 2018)

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

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

It is known that

$$\kappa_n(\mathbf{W}^r_p) \asymp n^{-r}, \quad r > 1/p, \quad 1 \le p \le \infty. \tag{6}$$

Theorem T3 and relation (6) imply

$$er_m(\mathbf{W}^r_p) \ge C(r,p)m^{-r}.$$

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.

- 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.

- 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.

For a compact subset Θ of a Banach space B we define the entropy numbers as follows

$$\varepsilon_n(\Theta, B) := \inf\{\varepsilon : \exists f_1, \dots, f_{2^n} \in \Theta : \Theta \subset \cup_{j=1}^{2^n} (f_j + \varepsilon U(B))\}$$

where $U(B)$ is the unit ball of a Banach space B.

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

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

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 $(\mathbf{x}, y) \in Z$

$$|f(\mathbf{x}) - y| \leq M. \tag{7}$$

Estimate for the defect function

Theorem (S. Konyagin and VT, 2004)

Assume ρ, W *satisfy (7) and*

$$\varepsilon_n(W, L_\infty) \le Dn^{-r}, \quad r \in (0, 1/2).$$

Then for m, η *satisfying* $m\eta^{1/r} \ge C_1(M, D, r)$ *we have*

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

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.

- 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.

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 $\mathbf{W}^r_2$ 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 $\|f\|^2$.
- 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.

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 $f \in W$

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

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

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

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

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

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

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

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*

$$\sum_{\mathbf{k}\in\mathbb{Z}^d} |\hat{F}(\mathbf{k})\hat{F}(\mathbf{n}-\mathbf{k})|^2 \le C_0^2 |\hat{F}(\mathbf{n})|^2. \tag{8}$$

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

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

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

$$\hat{F}_r(\mathbf{k}) = \prod_{j=1}^{d} (k_j^*)^{-r}. \tag{9}$$

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

$$\hat{F}_r(\mathbf{k}) = \prod_{j=1}^{d} (k_j^*)^{-r}. \tag{9}$$

Lemma (L1)

Function $F = F_r$ with $r > 1/2$ satisfies condition (8) with $C_0 = C(r, d)$.

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

$$\hat{F}_r(\mathbf{k}) = \prod_{j=1}^{d} (k_j^*)^{-r}. \tag{9}$$

Lemma (L1)

Function $F = F_r$ with $r > 1/2$ satisfies condition (8) with $C_0 = C(r, d)$.

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

We now illustrate how a combination of Proposition P1 and known results on numerical integration gives results on discretization. We discuss classes of periodic functions of two variables. Let $\{b_n\}_{n=0}^{\infty}$, $b_0 = b_1 = 1$, $b_n = b_{n-1} + b_{n-2}$, $n \geq 2$, – be the Fibonacci numbers.

We now illustrate how a combination of Proposition P1 and known results on numerical integration gives results on discretization. We discuss classes of periodic functions of two variables. Let $\{b_n\}_{n=0}^{\infty}$, $b_0 = b_1 = 1$, $b_n = b_{n-1} + b_{n-2}$, $n \geq 2$, – be the Fibonacci numbers. For continuous functions of two variables, which are 2π-periodic in each variable, we define cubature formulas

$$\Phi_n(f) := b_n^{-1} \sum_{\mu=1}^{b_n} f\left(2\pi\mu/b_n, 2\pi\{\mu b_{n-1}/b_n\}\right),$$

which are called the Fibonacci cubature formulas. In this definition $\{a\}$ is the fractional part of the number a.

Known result for numerical integration

For a function class W denote

$$\Phi_n(W) := \sup_{f\in W} |\Phi_n(f) - \hat{f}(\mathbf{0})|.$$

The following result is known

$$\Phi_n(\mathbf{W}^r_2) \asymp b_n^{-r}(\log b_n)^{1/2}, \qquad r > 1/2. \tag{10}$$

Combining (10) with Proposition P1 we obtain the following discretization result.

Discretization result

Theorem (T4; VT, 2018)

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

$$er_m(\mathbf{W}_2^r, L_2) \le C(r) m^{-r} (\log m)^{1/2}.$$

Thank you!