Nonlinear tensor product approximation

Vladimir Temlyakov

ICERM; October 3, 2014

Best multilinear approximation

We are interested in approximation of a multivariate function $f(x_1, \dots, x_d)$ by linear combinations of products $u^1(x_1) \cdots u^d(x_d)$ of univariate functions $u^i(x_i)$, $i = 1, \dots, d$.

Best multilinear approximation

We are interested in approximation of a multivariate function $f(x_1,\dots,x_d)$ by linear combinations of products $u^1(x_1)\cdots u^d(x_d)$ of univariate functions $u^i(x_i)$, $i=1,\dots,d$.

Definition

For a function $f(x_1,\dots,x_d)$ denote

$$\Theta_M(f)_X := \inf_{\{u^i_j\},j=1,\dots,M,i=1,\dots,d} \|f(x_1,\dots,x_d) - \sum_{j=1}^{M}\prod_{i=1}^{d} u^i_j(x_i)\|_X$$

and for a function class F define

$$\Theta_M(F)_X := \sup_{f\in F}\Theta_M(f)_X.$$

Best multilinear approximation

We are interested in approximation of a multivariate function $f(x_1,\dots,x_d)$ by linear combinations of products $u^1(x_1)\cdots u^d(x_d)$ of univariate functions $u^i(x_i)$, $i=1,\dots,d$.

Definition

For a function $f(x_1,\dots,x_d)$ denote

$$\Theta_M(f)_X := \inf_{\{u^i_j\},j=1,\dots,M,i=1,\dots,d} \|f(x_1,\dots,x_d) - \sum_{j=1}^{M}\prod_{i=1}^{d} u^i_j(x_i)\|_X$$

and for a function class F define

$$\Theta_M(F)_X := \sup_{f\in F} \Theta_M(f)_X.$$

In the case $X=L_p$ we write p instead of L_p in the notation.

Multilinear dictionary

In other words we are interested in studying M-term approximations of functions with respect to the dictionary

$$\Pi^d := \{g(x_1,\dots,x_d) : g(x_1,\dots,x_d) = \prod_{i=1}^{d} u^i(x_i)\}$$

where $u^i(x_i)$ are arbitrary univariate functions. We discuss the case of 2π-periodic functions of d variables and approximate them in the L_p spaces. Denote by Π^d_p the normalized in L_p dictionary Π^d of 2π-periodic functions.

Mega problem

Problem

Find a simple algorithm such that for any $f \in L_p$ *it provides after* $M \leq m\phi(m)$ *iterations an* M*-term with respect to* Π^d *approximant* $A_M(f)$ *such that*

$$\|f - A_M(f)\|_p \leq C_1 \Theta_m(f)_p.$$

Mega problem

Problem

Find a simple algorithm such that for any $f \in L_p$ *it provides after* $M \leq m\phi(m)$ *iterations an* M*-term with respect to* Π^d *approximant* $A_M(f)$ *such that*

$$\|f - A_M(f)\|_p \leq C_1 \Theta_m(f)_p.$$

Simple – incremental, greedy.

Mega problem

Problem

Find a simple algorithm such that for any $f \in L_p$ *it provides after* $M \le m\phi(m)$ *iterations an* M*-term with respect to* Π^d *approximant* $A_M(f)$ *such that*

$$\|f - A_M(f)\|_p \le C_1 \Theta_m(f)_p.$$

Simple – incremental, greedy.

Ideally, $\phi(m) = C_2$. Otherwise, the slower growing $\phi(m)$ the better.

What does give a hope?

- In the case $d = 2$, $p = 2$ the Pure Greedy Algorithm and the Orthogonal Greedy Algorithm (Orthogonal Matching Pursuit) solve the problem with $\phi(m) = 1$, $C_1 = 1$.

What does give a hope?

- In the case $d=2$, $p=2$ the Pure Greedy Algorithm and the Orthogonal Greedy Algorithm (Orthogonal Matching Pursuit) solve the problem with $\phi(m)=1$, $C_1=1$.
- For the trigonometric system we have.

Theorem (T., 2014)

Let $\mathcal{D}$ be the normalized in L_p, $2 \le p < \infty$, real d-variate trigonometric system. Then for any $f \in L_p$ the Weak Chebyshev Greedy Algorithm with weakness parameter t gives

$$\|f_{C(t,p,d)m\ln(m+1)}\|_p \le C\sigma_m(f,\mathcal{D})_p. \tag{1}$$

Dictionary

Definition

A set of functions $\mathcal{D}$ from a Banach space X is a dictionary if each $g \in \mathcal{D}$ has norm one $(\|g\| := \|g\|_X = 1)$ and the closure of span$\mathcal{D}$ coincides with X.

Dictionary

Definition

A set of functions $\mathcal{D}$ from a Banach space X is a dictionary if each $g \in \mathcal{D}$ has norm one $(\|g\| := \|g\|_X = 1)$ and the closure of span$\mathcal{D}$ coincides with X.

Definition

For a nonzero element $f \in X$ we denote by F_f a norming (peak) functional for f:

$$\|F_f\| = 1, \qquad F_f(f) = \|f\|.$$

Dictionary

Definition

A set of functions $\mathcal{D}$ from a Banach space X is a dictionary if each $g \in \mathcal{D}$ has norm one $(\|g\| := \|g\|_X = 1)$ and the closure of span$\mathcal{D}$ coincides with X.

Definition

For a nonzero element $f \in X$ we denote by F_f a norming (peak) functional for f:

$$\|F_f\| = 1, \qquad F_f(f) = \|f\|.$$

The existence of such a functional is guaranteed by the Hahn-Banach theorem.

Chebyshev greedy algorithm

Weak Chebyshev Greedy Algorithm (WCGA)(t) Let $t \in (0,1]$. For a given f_0 we inductively define for each $m \geq 1$

- $\varphi_m \in \mathcal{D}$ is any satisfying

$$|F_{f_{m-1}}(\varphi_m)| \geq t \sup_{g \in \mathcal{D}} |F_{f_{m-1}}(g)|.$$

Chebyshev greedy algorithm

Weak Chebyshev Greedy Algorithm (WCGA)(t) Let $t \in (0,1]$. For a given f_0 we inductively define for each $m \geq 1$

- $\varphi_m \in \mathcal{D}$ is any satisfying

$$|F_{f_{m-1}}(\varphi_m)| \geq t \sup_{g \in \mathcal{D}} |F_{f_{m-1}}(g)|.$$

- Define

$$\Phi_m := \operatorname{span}\{\varphi_j\}_{j=1}^m,$$

and define G_m to be the best approximant to f_0 from Φ_m.

Chebyshev greedy algorithm

Weak Chebyshev Greedy Algorithm (WCGA)(t) Let $t \in (0,1]$. For a given f_0 we inductively define for each $m \geq 1$

- $\varphi_m \in \mathcal{D}$ is any satisfying

$$|F_{f_{m-1}}(\varphi_m)| \geq t \sup_{g \in \mathcal{D}} |F_{f_{m-1}}(g)|.$$

- Define

$$\Phi_m := \operatorname{span}\{\varphi_j\}_{j=1}^m,$$

  and define G_m to be the best approximant to f_0 from Φ_m.
- Denote

$$f_m := f - G_m.$$

Best m-term approximation

Definition

We let $\Sigma_m(\mathcal{D})$ denote the collection of all functions (elements) in X which can be expressed as a linear combination of at most m elements of $\mathcal{D}$. Thus each function $f \in \Sigma_m(\mathcal{D})$ can be written in the form

$$f = \sum_{g\in\Lambda} c_g g, \quad \Lambda \subset \mathcal{D}, \quad \#\Lambda \le m,$$

where the c_g are real or complex numbers.

Best m-term approximation

Definition

We let $\Sigma_m(\mathcal{D})$ denote the collection of all functions (elements) in X which can be expressed as a linear combination of at most m elements of $\mathcal{D}$. Thus each function $f \in \Sigma_m(\mathcal{D})$ can be written in the form

$$f = \sum_{g \in \Lambda} c_g g, \quad \Lambda \subset \mathcal{D}, \quad \#\Lambda \le m,$$

where the c_g are real or complex numbers.

Definition

For a function $f \in X$ we define its best m-term approximation error

$$\sigma_m(f) := \sigma_m(f, \mathcal{D}) := \inf_{a \in \Sigma_m(\mathcal{D})} \|f - a\|.$$

Nice bases

There are two fundamental systems: the d-variate trigonometric system $\mathcal{T}^d := \{e^{i(k,x)}\}$ and the prototype of the wavelets the $\mathcal{U}^d$ system. We define the system $\mathcal{U} := \{U_I\}$ in the univariate case.

Nice bases

There are two fundamental systems: the d-variate trigonometric system $\mathcal{T}^d := \{e^{i(k,x)}\}$ and the prototype of the wavelets the $\mathcal{U}^d$ system. We define the system $\mathcal{U} := \{U_I\}$ in the univariate case. Denote

$$U_n^+(x) := \sum_{k=0}^{2^n-1} e^{ikx} = \frac{e^{i2^n x} - 1}{e^{ix} - 1}, \quad n = 0, 1, 2, \ldots;$$

Nice bases

There are two fundamental systems: the d-variate trigonometric system $\mathcal{T}^d := \{e^{i(k,x)}\}$ and the prototype of the wavelets the $\mathcal{U}^d$ system. We define the system $\mathcal{U} := \{U_I\}$ in the univariate case. Denote

$$U_n^+(x) := \sum_{k=0}^{2^n-1} e^{ikx} = \frac{e^{i2^n x}-1}{e^{ix}-1}, \quad n = 0, 1, 2, \dots;$$

$$U_{n,k}^+(x) := e^{i2^n x} U_n^+(x - 2\pi k 2^{-n}), \quad k = 0, 1, \dots, 2^n - 1;$$

$$U_{n,k}^-(x) := e^{-i2^n x} U_n^+(-x + 2\pi k 2^{-n}), \quad k = 0, 1, \dots, 2^n - 1.$$

System $\mathcal{U}$

It will be more convenient for us to normalize in L_2 the system of functions $\{U^+_{m,k}, U^-_{n,k}\}$ and enumerate it by dyadic intervals. We write

$$U_I(x) := 2^{-n/2} U^+_{n,k}(x) \quad \text{with} \quad I = [(k+1/2)2^{-n}, (k+1)2^{-n});$$

$$U_I(x) := 2^{-n/2} U^-_{n,k}(x) \quad \text{with} \quad I = [k2^{-n}, (k+1/2)2^{-n});$$

and

$$U_{[0,1)}(x) := 1.$$

System $\mathcal{U}^d$

In the multivariate case of $x = (x_1, \dots, x_d)$ we define the system $\mathcal{U}^d$ as the tensor product of the univariate systems $\mathcal{U}$. Let $I = I_1 \times \cdots \times I_d$, $I_j \in D$, $j = 1, \dots, d$, then

$$U_I(x) := \prod_{j=1}^{d} U_{I_j}(x_j).$$

System $\mathcal{U}^d$

In the multivariate case of $x = (x_1, \dots, x_d)$ we define the system $\mathcal{U}^d$ as the tensor product of the univariate systems $\mathcal{U}$. Let $I = I_1 \times \cdots \times I_d$, $I_j \in D$, $j = 1, \dots, d$, then

$$U_I(x) := \prod_{j=1}^{d} U_{I_j}(x_j).$$

Both $\mathcal{T}^d$ and $\mathcal{U}^d$ have two fundamental features: orthogonality and tensor product structure.

Definition

We say that a dictionary $\mathcal{D}$ has a tensor product structure if all its elements have a form of products $u^1(x_1) \cdots u^d(x_d)$ of univariate functions $u^i(x_i)$, $i = 1, \dots, d$.

System $\mathcal{U}^d$

In the multivariate case of $x = (x_1, \dots, x_d)$ we define the system $\mathcal{U}^d$ as the tensor product of the univariate systems $\mathcal{U}$. Let $I = I_1 \times \cdots \times I_d$, $I_j \in D$, $j = 1, \dots, d$, then

$$U_I(x) := \prod_{j=1}^{d} U_{I_j}(x_j).$$

Both $\mathcal{T}^d$ and $\mathcal{U}^d$ have two fundamental features: orthogonality and tensor product structure.

Definition

We say that a dictionary $\mathcal{D}$ has a tensor product structure if all its elements have a form of products $u^1(x_1) \cdots u^d(x_d)$ of univariate functions $u^i(x_i)$, $i = 1, \dots, d$.

Any dictionary with tensor product structure is a subset of Π^d.

Function class W^r_q

Let

$$F_r(t) := 1 + 2\sum_{k=1}^{\infty} k^{-r}\cos(kt - \pi r/2)$$

be the univariate Bernoulli kernel and let

$$F_r(x) := F_r(x_1,\dots,x_d) := \prod_{i=1}^{d} F_r(x_i).$$

Function class W^r_q

Let

$$F_r(t) := 1 + 2\sum_{k=1}^{\infty} k^{-r}\cos(kt - \pi r/2)$$

be the univariate Bernoulli kernel and let

$$F_r(x) := F_r(x_1,\dots,x_d) := \prod_{i=1}^{d} F_r(x_i).$$

Definition

We define

$$W^r_q := \{f : f = F_r * \varphi, \quad \|\varphi\|_q \le 1\},$$

where $*$ denotes the convolution.

Best m-term approximation with respect to $\mathcal{U}^d$

Theorem (T, 2000; Th1)

For $1 < q, p < \infty$ *and big enough* r *we have*

$$\sigma_m(W^r_q, \mathcal{U}^d)_p \asymp m^{-r}(\ln m)^{(d-1)r}.$$

Best m-term approximation with respect to $\mathcal{U}^d$

Theorem (T, 2000; Th1)

For $1 < q, p < \infty$ and big enough r we have

$$\sigma_m(W^r_q, \mathcal{U}^d)_p \asymp m^{-r}(\ln m)^{(d-1)r}.$$

Theorem (T, 2000; Th2)

For any orthonormal system Ψ we have for $1 \leq q < \infty$

$$\sigma_m(W^r_q, \Psi)_2 \gg m^{-r}(\ln m)^{(d-1)r}.$$

Best m-term approximation with respect to $\mathcal{U}^d$

Theorem (T, 2000; Th1)

For $1 < q, p < \infty$ and big enough r we have

$$\sigma_m(W^r_q, \mathcal{U}^d)_p \asymp m^{-r}(\ln m)^{(d-1)r}.$$

Theorem (T, 2000; Th2)

For any orthonormal system Ψ we have for $1 \le q < \infty$

$$\sigma_m(W^r_q, \Psi)_2 \gg m^{-r}(\ln m)^{(d-1)r}.$$

Thus, the $\mathcal{U}^d$ is an ideal orthogonal system.

Bilinear approximation

In the case $d = 2$ the multilinear approximation problem is a classical problem of bilinear approximation. There are known results on the rate of decay of errors of best bilinear approximation in L_p under different smoothness assumptions on f. We only mention some known results for classes of functions W^r_q. The problem of estimating $\Theta_M(f)_2$ in case $d = 2$ (best M-term bilinear approximation in L_2) is a classical one and was considered for the first time by E. Schmidt in 1907. For many function classes F an asymptotic behavior of $\Theta_M(F)_p$ is known. For instance

Theorem (T, 1986)

In the case $d = 2$ for $r > 1$ and $1 \le q \le p \le \infty$ we have

$$\Theta_M(W^r_q)_p \asymp M^{-2r+(1/q-\max(1/2,1/p))_+} \qquad (2)$$

Multilinear approximation

In the case $d > 2$ almost nothing is known. There is an upper estimate in the case $q = p = 2$

Theorem (T, 1988)

For $r > 0$ we have

$$\Theta_M(W^r_2)_2 \ll M^{-rd/(d-1)}. \tag{3}$$

Multilinear approximation

In the case $d > 2$ almost nothing is known. There is an upper estimate in the case $q = p = 2$

Theorem (T, 1988)

For $r > 0$ we have

$$\Theta_M(W^r_2)_2 \ll M^{-rd/(d-1)}. \tag{3}$$

The above theorems show that the rate $M^{-r}(\log M)^{(d-1)r}$ of best M-term approximation with respect to the basis $\mathcal{U}^d$, which has a tensor structure, is not as good as best M-term approximation with respect to Π^d (we have exponent r for $\mathcal{U}^d$ instead of $\frac{rd}{d-1}$ for Π^d).

Known lower bounds

New results of this talk are around the bound (3). First of all we discuss the lower bound matching the upper bound (3). In the case $d = 2$ the lower bound

$$\Theta_M(W^r_p)_p \gg M^{-2r}, \qquad 1 \le p \le \infty. \tag{4}$$

follows from more general results in [T, 1986] (see (2) above).

Known lower bounds

New results of this talk are around the bound (3). First of all we discuss the lower bound matching the upper bound (3). In the case $d = 2$ the lower bound

$$\Theta_M(W^r_p)_p \gg M^{-2r}, \qquad 1 \le p \le \infty. \tag{4}$$

follows from more general results in [T, 1986] (see (2) above). A stronger result

$$\Theta_M(W^r_\infty)_1 \gg M^{-2r} \tag{5}$$

follows from Theorem 1.1 in [T, 1992].

New lower bounds

We could not prove the lower bound matching the upper bound (3) for $d > 2$. Instead, we prove a weaker lower bound.

New lower bounds

We could not prove the lower bound matching the upper bound (3) for $d > 2$. Instead, we prove a weaker lower bound. For a function $f(x_1, \dots, x_d)$ denote

$$\Theta_M^b(f)_X := \inf_{\{u_j^i\}, \|u_j^i\|_X \le b\|f\|_X^{1/d}} \|f(x_1, \dots, x_d) - \sum_{j=1}^{M} \prod_{i=1}^{d} u_j^i(x_i)\|_X$$

and for a function class F define $\Theta_M^b(F)_X := \sup_{f\in F} \Theta_M^b(f)_X$.

New lower bounds

We could not prove the lower bound matching the upper bound (3) for $d > 2$. Instead, we prove a weaker lower bound. For a function $f(x_1, \dots, x_d)$ denote

$$\Theta_M^b(f)_X := \inf_{\{u_j^i\}, \|u_j^i\|_X \le b\|f\|_X^{1/d}} \|f(x_1, \dots, x_d) - \sum_{j=1}^{M} \prod_{i=1}^{d} u_j^i(x_i)\|_X$$

and for a function class F define $\Theta_M^b(F)_X := \sup_{f\in F} \Theta_M^b(f)_X$.

Theorem (Bazarkhanov and T., 2014)

$$\Theta_M^b(W_\infty^r)_1 \gg (M \ln M)^{-\frac{rd}{d-1}}.$$

Upper bounds

Secondly, we discuss some upper bounds which extend the bound (3). The relation (2) shows that for $2 \leq p \leq \infty$ in the case $d = 2$ one has

$$\Theta_M(W_2^r)_p \ll M^{-2r}. \tag{6}$$

We extended (6) for $d > 2$.

Upper bounds

Secondly, we discuss some upper bounds which extend the bound (3). The relation (2) shows that for $2 \le p \le \infty$ in the case $d = 2$ one has

$$\Theta_M(W_2^r)_p \ll M^{-2r}. \tag{6}$$

We extended (6) for $d > 2$.

Theorem (Bazarkhanov and T., 2014)

Let $2 \le p < \infty$ and $r > (d-1)/d$. Then

$$\Theta_M(W_2^r)_p \ll \left(\frac{M}{(\log M)^{d-1}}\right)^{-\frac{rd}{d-1}}.$$

Upper bounds

Secondly, we discuss some upper bounds which extend the bound (3). The relation (2) shows that for $2 \leq p \leq \infty$ in the case $d = 2$ one has

$$\Theta_M(W_2^r)_p \ll M^{-2r}. \tag{6}$$

We extended (6) for $d > 2$.

Theorem (Bazarkhanov and T., 2014)

Let $2 \leq p < \infty$ and $r > (d-1)/d$. Then

$$\Theta_M(W_2^r)_p \ll \left(\frac{M}{(\log M)^{d-1}}\right)^{-\frac{rd}{d-1}}.$$

The proof of the above theorem is not constructive. It goes by induction and uses a nonconstructive bound in the case $d = 2$.

TGA

We define a well-known Thresholding Greedy Algorithm with respect to a basis. It is convenient for us to enumerate the basis functions by dyadic intervals. Assume a given system Ψ of functions ψ_I indexed by dyadic intervals can be enumerated in such a way that $\{\psi_{I^j}\}_{j=1}^\infty$ is a basis for L_p. Then we define the greedy algorithm $G^p(\cdot,\Psi)$ as follows. Let

$$f=\sum_{j=1}^\infty c_{I^j}(f,\Psi)\psi_{I^j}, \quad c_I(f,p,\Psi):=\|c_I(f,\Psi)\psi_I\|_p.$$

Then $c_I(f,p,\Psi)\to 0$ as $|I|\to 0$.

TGA continue

Denote Λ_m a set of m dyadic intervals I such that

$$\min_{I\in\Lambda_m} c_I(f,p,\Psi) \geq \max_{J\notin\Lambda_m} c_J(f,p,\Psi).$$

We define $G^p(\cdot,\Psi)$ by formula

$$G^p_m(f,\Psi) := \sum_{I\in\Lambda_m} c_I(f,\Psi)\psi_I.$$

TGA is good

It is proved in [T., 2000] that for $1 < q, p < \infty$ and big enough r

$$\sup_{f \in W^r_q} \|f - G^p_M(f, \mathcal{U}^d)\|_p \asymp \sigma_M(W^r_q, \mathcal{U}^d)_p$$

$$\asymp M^{-r}(\log M)^{(d-1)r}. \tag{7}$$

TGA is good

It is proved in [T., 2000] that for $1 < q, p < \infty$ and big enough r

$$\sup_{f \in W^r_q} \|f - G^p_M(f, \mathcal{U}^d)\|_p \asymp \sigma_M(W^r_q, \mathcal{U}^d)_p$$

$$\asymp M^{-r}(\log M)^{(d-1)r}. \tag{7}$$

The above relation (7) illustrates two phenomena: (I) for the class W^r_q the simple Thresholding Greedy Algorithm provides near best M-term approximation; (II) the rate $M^{-r}(\log M)^{(d-1)r}$ of best M-term approximation with respect to the basis $\mathcal{U}^d$, which has a tensor structure, is not as good as best M-term approximation with respect to Π^d (we have exponent r for $\mathcal{U}^d$ instead of $\frac{rd}{d-1}$ for Π^d).

Constructive multilinear approximation

We use two very different greedy-type algorithms to provide a constructive multilinear approximant. Surprisingly, these two algorithms gave the same error bound.

Theorem (Bazarkhanov and T., 2014)

For big enough r the following constructive upper bound for $2 \le p < \infty$ holds

$$\Theta_M(W^r_2)_p \ll \left(\frac{M}{(\ln M)^{d-1}}\right)^{-\frac{rd}{d-1}+\frac{\beta}{d-1}}.$$

Constructive multilinear approximation

We use two very different greedy-type algorithms to provide a constructive multilinear approximant. Surprisingly, these two algorithms gave the same error bound.

Theorem (Bazarkhanov and T., 2014)

For big enough r the following constructive upper bound for $2 \le p < \infty$ holds

$$\Theta_M(W_2^r)_p \ll \left(\frac{M}{(\ln M)^{d-1}}\right)^{-\frac{rd}{d-1}+\frac{\beta}{d-1}}.$$

This constructive upper bound has an extra term $\frac{\beta}{d-1}$ in the exponent compared to the best M-term approximation. It would be interesting to find a constructive way to obtain the near best approximation in this case.

Trigonometric system. Some history

De la Vallee Poussin (1908) and Bernstein (1912) proved

$$E_n(|\sin x|)_\infty \asymp n^{-1}.$$

Trigonometric system. Some history

De la Vallee Poussin (1908) and Bernstein (1912) proved

$$E_n(|\sin x|)_\infty \asymp n^{-1}.$$

Ismagilov (1974) and Maiorov (1986) proved

$$\sigma_m(|\sin x|, \mathcal{T})_\infty \asymp n^{-3/2}.$$

Trigonometric system. Some history

De la Vallee Poussin (1908) and Bernstein (1912) proved

$$E_n(|\sin x|)_\infty \asymp n^{-1}.$$

Ismagilov (1974) and Maiorov (1986) proved

$$\sigma_m(|\sin x|, \mathcal{T})_\infty \asymp n^{-3/2}.$$

Key: constructive, uses Gaussian sums to prove an inequality for trigonometric polynomials

$$\sigma_m(f, \mathcal{T})_\infty \le CN^{3/2}m^{-1}\|f\|_1, \quad f \in \mathcal{T}(N). \qquad (T1)$$

More history

It is easy to see that (T1) follows from

$$\sigma_m(f,\mathcal{T})_\infty \le CN^{1/2}m^{-1}\|f\|_A, \quad f\in \mathcal{T}(N). \qquad (T2)$$

where

$$\|f\|_A := \sum_k |\hat{f}(k)|.$$

More history

It is easy to see that (T1) follows from

$$\sigma_m(f,\mathcal{T})_\infty \le CN^{1/2}m^{-1}\|f\|_A, \quad f\in\mathcal{T}(N). \qquad (T2)$$

where

$$\|f\|_A := \sum_k |\hat{f}(k)|.$$

Devore and T. (1995)

$$\sigma_m(f,\mathcal{T})_\infty \le Cm^{-1/2}(\ln(1+N/m))^{1/2}\|f\|_A, \quad f\in\mathcal{T}(N). \quad (T3)$$

In a certain sense (T3) is much stronger than (T2). However, the proof was nonconstructive.

History continues

Dilworth, Kutzarova, and T. (2002)

$$\sigma_m(f,\mathcal{T})_p \le C(p)m^{-1/2}\|f\|_A, \quad 2 \le p < \infty, \quad f \in \mathcal{T}(N). \quad (T4)$$

History continues

Dilworth, Kutzarova, and T. (2002)

$$\sigma_m(f,\mathcal{T})_p \le C(p)m^{-1/2}\|f\|_A, \quad 2\le p<\infty, \quad f\in\mathcal{T}(N). \quad (T4)$$

T. (2005) gave a constructive proof of (T3). This proof and the proof of (T4) use the Weak Chebyshev Greedy Algorithm.

History continues

Dilworth, Kutzarova, and T. (2002)

$$\sigma_m(f,\mathcal{T})_p \le C(p)m^{-1/2}\|f\|_A, \quad 2\le p<\infty, \quad f\in\mathcal{T}(N). \quad (T4)$$

T. (2005) gave a constructive proof of (T3). This proof and the proof of (T4) use the Weak Chebyshev Greedy Algorithm.
In particular, (T4) implies

$$\sigma_m(f,\mathcal{T})_p \le C(p)N^{1/2}m^{-1/2}\|f\|_2, \quad 2\le p<\infty, \quad f\in\mathcal{T}(N). \quad (T5)$$

Best multilinear approximation

Lemma (Bazarkhanov and T., 2014)

Let $f \in T(\mathbf{N})$. *Denote* $v(\mathbf{N}) := \prod_{j=1}^{d} \bar{N}_j$. *Then for* $2 \le p < \infty$ *one has*

$$\Theta_M(f)_p \ll v(\mathbf{N})^{1-\frac{1}{d}} (\bar{M})^{-1} \|f\|_2, \quad \bar{M} = \max(M, 1).$$

TGA-type algorithm

Lemma (Bazarkhanov and T., 2014)

Suppose that $f \in T(\mathbf{N})$. *Denote* $v(\mathbf{N}) := \prod_{j=1}^{d} \bar{N}_j$. *Then for* $1 \le q \le p \le \infty$

$$\Theta_m(f)_p \ll v(\mathbf{N})^{\beta} (\bar{m})^{-\beta} \|f\|_q, \quad \beta := \frac{1}{q} - \frac{1}{p}, \quad \bar{m} := \max(1, m). \tag{8}$$

The bound (8) is realized by a simple greedy-type algorithm.

General greedy-type algorithms

Lemma (Bazarkhanov and T., 2014)

Let $f \in T(\mathbf{N})$. Then for $2 \le p < \infty$

$$\Theta_m(f)_p \ll v(\mathbf{N})^{\frac{1}{2}-\frac{1}{pd}}(\bar{m})^{-1/2}\|f\|_2.$$

The above bound is realized by the Weak Chebyshev Greedy Algorithm with weakness parameter t.