On Greedy Algorithms with restricted depth search

Vladimir Temlyakov
(South Carolina)

1. Introduction

Let X be a Banach space with norm $\|\cdot\|$. We say that a set of elements (functions) $\mathcal{D}$ from X is a symmetric dictionary if each $g \in \mathcal{D}$ has norm less than or equal to one ($\|g\| \le 1$),

$$g \in \mathcal{D} \quad \text{implies} \quad -g \in \mathcal{D},$$

and closure of span$\mathcal{D} = X$.
For an 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.

Let $\tau := \{t_k\}_{k=1}^{\infty}$ be a given sequence of nonnegative numbers $t_k \leq 1$, $k = 1, \ldots$. We define first the Weak Chebyshev Greedy Algorithm (WCGA) that is a generalization for Banach spaces of Weak Orthogonal Greedy Algorithm defined for Hilbert spaces.

WCGA We define $f_0^c := f_0^{c,\tau} := f$. Then for each $m \geq 1$ we inductively define

1). $\varphi_m^c := \varphi_m^{c,\tau} \in \mathcal{D}$ is any satisfying

$$F_{f_{m-1}^c}(\varphi_m^c) \geq t_m \sup_{g \in \mathcal{D}} F_{f_{m-1}^c}(g).$$

2). Define

$$\Phi_m := \Phi_m^\tau := \text{span}\{\varphi_j^c\}_{j=1}^m,$$

and define $G_m^c := G_m^{c,\tau}$ to be the best approximant to f from Φ_m.

3). Denote

$$f_m^c := f_m^{c,\tau} := f - G_m^c.$$

In the case $t_k = 1$, $k = 1, 2, \dots$ we call the WCGA the Chebyshev Greedy Algorithm (CGA).

Let three sequences $\tau = \{t_k\}_{k=1}^{\infty}$, $\delta = \{\delta_k\}_{k=0}^{\infty}$, $\eta = \{\eta_k\}_{k=1}^{\infty}$ of numbers from $[0, 1]$ be given. We define the Approximate Weak Chebyshev Greedy Algorithm (AWCGA) as follows.

AWCGA We define $f_0 := f_0^{\tau,\delta,\eta} := f$. Then for each $m \geq 1$ we inductively define

1). F_{m-1} is a functional with properties

$$\|F_{m-1}\| \leq 1,$$

$$F_{m-1}(f_{m-1}) \geq \|f_{m-1}\|(1 - \delta_{m-1});$$

and $\varphi_m := \varphi_m^{\tau,\delta,\eta} \in \mathcal{D}$ is any satisfying

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

2). Define

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

and denote

$$E_m(f) := \inf_{\varphi \in \Phi_m} \|f - \varphi\|.$$

Let $G_m \in \Phi_m$ be such that

$$\|f - G_m\| \leq E_m(f)(1 + \eta_m).$$

3). Denote

$$f_m := f_m^{\tau,\delta,\eta} := f - G_m.$$

The term *approximate* in this definition means that we use a functional F_{m-1} that is an approximation to the norming (peak) functional $F_{f_{m-1}}$ and also we use an approximant $G_m \in \Phi_m$ which satisfies a weaker assumption than being a best approximant of f from Φ_m.

We now consider a countable dictionary $\mathcal{D} = \{\pm\psi_j\}_{j=1}^{\infty}$. We denote $\mathcal{D}(N) := \{\pm\psi_j\}_{j=1}^{N}$. Let $\mathcal{N} := \{N_j\}_{j=1}^{\infty}$ be a sequence of natural numbers. We define the Restricted Weak Chebyshev Greedy Algorithm (RWCGA) as follows.

RWCGA We define $f_0 := f_0^{c,\tau,\mathcal{N}} := f$. Then for each $m \ge 1$ we inductively define

1). $\varphi_m := \varphi_m^{c,\tau,\mathcal{N}} \in \mathcal{D}(N_m)$ is any satisfying

$$F_{f_{m-1}}(\varphi_m) \ge t_m \sup_{g \in \mathcal{D}(N_m)} F_{f_{m-1}}(g).$$

2). Define

$$\Phi_m := \Phi_m^{\tau,\mathcal{N}} := \text{span}\{\varphi_j\}_{j=1}^m,$$

and define $G_m := G_m^{c,\tau,\mathcal{N}}$ to be the best approximant to f from Φ_m.

3). Denote

$$f_m := f_m^{c,\tau,\mathcal{N}} := f - G_m.$$

2. Convergence of the RWCGA

We consider here approximation in uniformly smooth Banach spaces. For a Banach space X we define the modulus of smoothness

$$\rho(u) := \sup_{\|x\|=\|y\|=1} (\frac{1}{2}(\|x + uy\| + \|x - uy\|) - 1).$$

The uniformly smooth Banach space is the one with the property

$$\lim_{u \to 0} \rho(u)/u = 0.$$

Theorem 2.1 (V.T., 2001) *Let a Banach space* X *have modulus of smoothness* $\rho(u)$ *of power type* $1 < q \leq 2$; $(\rho(u) \leq \gamma u^q)$. *Assume that*

$$\sum_{m=1}^{\infty} t_m^p = \infty, \quad p = \frac{q}{q-1}.$$

Then the WCGA converges for any $f \in X$.

Theorem 2.2 *Let a Banach space* X *have modulus of smoothness* $\rho(u)$ *of power type* $1 < q \leq 2$; $(\rho(u) \leq \gamma u^q)$. *Assume that* $\lim_{m\to\infty} N_m = \infty$ *and*

$$\sum_{m=1}^{\infty} t_m^p = \infty, \quad p = \frac{q}{q-1}.$$

Then the RWCGA converges for any $f \in X$.

We now proceed to study the rate of convergence of the RWCGA. We denote the closure of the convex hull of $\mathcal{D}$ by $\mathcal{A}_1(\mathcal{D})$.

Theorem 2.3 (V.T., 2001) *Let X be a uniformly smooth Banach space with the modulus of smoothness $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. Then for a sequence $\tau := \{t_k\}_{k=1}^{\infty}$, $t_k \leq 1$, $k = 1, 2, \dots,$ we have for any $f \in \mathcal{A}_1(\mathcal{D})$ that*

$$\|f_m^{c,\tau}\| \leq C(q,\gamma)(1 + \sum_{k=1}^{m} t_k^p)^{-1/p}, \quad p := \frac{q}{q-1},$$

with a constant $C(q, \gamma)$ which may depend only on q and γ.

For $b > 0$, $K > 0$ we define the class

$$\mathcal{A}_1^b(K, \mathcal{D}) :=$$

$$\{f : d(f, \mathcal{A}_1(\mathcal{D}(n)) \leq Kn^{-b}, \quad n = 1, 2, \dots\}.$$

Here, $\mathcal{A}_1(\mathcal{D}(n))$ is a convex hull of $\{\pm\psi_j\}_{j=1}^n$ and for a compact set F

$$d(f, F) := \inf_{\phi \in F} \|f - \phi\|.$$

Theorem 2.4 *Let* X *be a uniformly smooth Banach space with the modulus of smoothness* $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. *Then for* $t \in (0, 1]$ *there exist* $C_1(t, \gamma, q, K)$, $C_2(t, \gamma, q, K)$ *such that for* $\mathcal{N}$ *with* $N_m \geq C_1(t, \gamma, q, K) m^{r/b}$, $m = 1, 2, \dots$ *we have for any* $f \in \mathcal{A}_1^b(K, \mathcal{D})$

$$\|f_m^{c,\tau,\mathcal{N}}\| \leq C_2(t, \gamma, q, K) m^{-r},$$

$$\tau = \{t\}, \quad r := 1 - 1/q.$$

We note that we can choose an algorithm from Theorem 2.4 that satisfies the *polynomial depth search* condition $N_m \leq C m^a$.

Example

We give an example of performance of the RWCGA. The problem concerns the trigonometric m-term approximation in the L_p-norm. Let $\mathcal{T}(N)$ be the subspace of real trigonometric polynomials of order N and let $\mathcal{T}$ be the real trigonometric system

$$\frac{1}{2}, \sin x, \cos x, \sin 2x, \cos 2x, \dots \quad .$$

Denote for $f \in L_p(\mathcal{T})$

$$\sigma_m(f, \mathcal{T})_p := \inf_{c_1,\dots,c_m;\phi_1,\dots,\phi_m \in \mathcal{T}} \|f - \sum_{j=1}^{m} c_j \phi_j\|_p$$

the best m-term trigonometric approximation of f in the L_p-norm. It is clear that one can get an upper estimate for $\sigma_{2m+1}(f,\mathcal{T})_p$ by approximating f by trigonometric polynomials of order m. Denote

$$E_m(f,\mathcal{T})_p := \inf_{u\in\mathcal{T}(m)} \|f-u\|_p.$$

Let

$$\mathcal{A}_1 := \mathcal{A}_1(\mathcal{T}) := \{f : \sum_{k=0}^{\infty} (|a_k(f)| + |b_k(f)|) \leq 1\}$$

where $a_k(f)$, $b_k(f)$ are the corresponding Fourier coefficients. From the general results on convergence rate of the WCGA (see Theorem 2.3 above) it follows that for $f \in \mathcal{A}_1$, $t_k = t \in (0, 1)$, $k = 1, 2, \dots$,

$$\|f_m^{c,\tau}\|_p \leq C(p, t) m^{-1/2}, \quad 2 \leq p < \infty.$$

Let us apply Theorem 2.4 in the same situation. Now, in addition to $f \in \mathcal{A}_1$ we require

$$E_n(f, \mathcal{T})_p \le D n^{-b}, \quad n = 1, 2, \dots, \tag{2.1}$$

with some $b > 0$.
Then it is easy to derive from Theorem 2.4 that there exist two constants $C_1(p, t, D)$, $C_2(p, t, D)$ such that for $\tau = \{t\}$ and $\mathcal{N}$ with $N_m \ge C_1(p, t, D) m^{-1/(2b)}$, $m = 1, 2, \dots$ we have for any $f \in \mathcal{A}_1$ satisfying (2.1) that

$$\|f_m^{c,\tau,\mathcal{N}}\|_p \le C_2(p, t, D) m^{-1/2}. \tag{2.2}$$

We note that for the above class one cannot obtain an esimate better than (2.2) (clearly, for $b \leq 1/2$). Indeed, let m be given. Consider

$$f(x) := (2m)^{-1} R(x), \quad R(x) = \sum_{k=1}^{2m} \pm \cos kx,$$

where $R(x)$ is the Rudin-Shapiro polynomial such that

$$\|R\|_\infty \leq C m^{1/2}.$$

Then $f \in \mathcal{A}_1$ and

$$E_n(f, \mathcal{T})_\infty \leq D n^{-1/2}, \quad n = 1, 2, \ldots.$$

Also,

$$\sigma_m(f, \mathcal{T})_2 \geq m^{-1/2}/2.$$

We now make some general remarks on m-term approximation with the depth search constraint. The depth search constraint means that for a given m we restrict ourselves to systems of elements (subdictionaries) containing at most $N := N(m)$ elements.

Let X be a linear metric space and for a set $\mathcal{D} \subset X$, let $\mathcal{L}_m(\mathcal{D})$ denote the collection of all linear spaces spanned by m elements of $\mathcal{D}$. For a linear space $L \subset X$, the ϵ-neighborhood $U_\epsilon(L)$ of L is the set of all $x \in X$ which are at a distance not exceeding ϵ from L (i.e. those $x \in X$ which can be approximated to an error not exceeding ϵ by the elements of L).

For any compact set $F \subset X$ and any integers $N, m \geq 1$, we define the (N, m)-entropy numbers (V.T., 1998)

$$\epsilon_{N,m}(F, X) :=$$

$$\inf_{\#\mathcal{D}=N} \inf\{\epsilon : F \subset \cup_{L \in \mathcal{L}_m(\mathcal{D})} U_\epsilon(L)\}.$$

We can express $\sigma_m(F, \mathcal{D})$ as

$$\sigma_m(F, \mathcal{D}) = \inf\{\epsilon : F \subset \cup_{L \in \mathcal{L}_m(\mathcal{D})} U_\epsilon(L)\}.$$

It follows therefore that

$$\inf_{\#\mathcal{D}=N} \sigma_m(F, \mathcal{D}) = \epsilon_{N,m}(F, X).$$

In other words, finding best dictionaries consisting of N elements for m-term approximation of F is the same as finding sets $\mathcal{D}$ which attain the (N, m)-entropy numbers $\epsilon_{N,m}(F, X)$. It is easy to see that $\epsilon_{m,m}(F, X) = d_m(F, X)$ where $d_m(F, X)$ is the Kolmogorov width of F in X. This establishes a connection between (N, m)-entropy numbers and the Kolmogorov widths.

3. Convergence of RAWCGA

Let three sequences $\tau = \{t_k\}_{k=1}^{\infty}$, $\delta = \{\delta_k\}_{k=0}^{\infty}$, $\eta = \{\eta_k\}_{k=1}^{\infty}$ of numbers from $[0,1]$ be given. Let $\mathcal{N} := \{N_j\}_{j=1}^{\infty}$ be a sequence of natural numbers. We define the Restricted Approximate Weak Chebyshev Greedy Algorithm (RAWCGA) as follows.

RAWCGA We define $f_0 := f_0^{\tau,\delta,\eta,\mathcal{N}} := f$. Then for each $m \geq 1$ we inductively define

1). F_{m-1} is a functional with properties

$$\|F_{m-1}\| \leq 1,$$

$$F_{m-1}(f_{m-1}) \geq \|f_{m-1}\|(1-\delta_{m-1});$$

and $\varphi_m := \varphi_m^{\tau,\delta,\eta,\mathcal{N}} \in \mathcal{D}(N_m)$ is any satisfying

$$F_{m-1}(\varphi_m) \geq t_m \sup_{g \in \mathcal{D}(N_m)} F_{m-1}(g).$$

2). Define

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

and denote

$$E_m(f) := \inf_{\varphi \in \Phi_m} \|f - \varphi\|.$$

Let $G_m \in \Phi_m$ be such that

$$\|f - G_m\| \le E_m(f)(1 + \eta_m).$$

3). Denote

$$f_m := f_m^{\tau,\delta,\eta,\mathcal{N}} := f - G_m.$$

We begin with the convergence theorem.

Theorem 3.1 (V.T., 2002) *Let a Banach space* X *have modulus of smoothness* $\rho(u)$ *of power type* $1 < q \leq 2$; $(\rho(u) \leq \gamma u^q)$. *Assume that*

$$\sum_{m=1}^{\infty} t_m^p = \infty, \quad p = \frac{q}{q-1};$$

and

$$\delta_m = o(t_m^p), \qquad \eta_m = o(t_m^p).$$

Then the AWCGA converges for any $f \in X$.

Theorem 3.2 (V.T., 2002) *Let* X *be a uniformly smooth Banach space. Assume that* $\tau = \{t\}$, $t \in (0, 1]$. *Then for any two sequences* $\delta, \eta \in c_0$ *the corresponding AWCGA converges for any* $f \in X$.

We got the following convergence result for the RAWCGA.

Theorem 3.3 *Let a Banach space* X *have modulus of smoothness* $\rho(u)$ *of power type* $1 < q \leq 2$; $(\rho(u) \leq \gamma u^q)$. *Assume that* $\lim_{m\to\infty} N_m = \infty$,

$$\sum_{m=1}^{\infty} t_m^p = \infty, \quad p = \frac{q}{q-1},$$

and

$$\delta_m = o(t_m^p), \qquad \eta_m = o(t_m^p).$$

Then the RAWCGA converges for any $f \in X$.

Theorem 3.4 *Let* X *be a uniformly smooth Banach space. Assume that* $\tau = \{t\}$, $t \in (0, 1]$. *Then for any two sequences* $\delta, \eta \in c_0$ *the corresponding RAWCGA converges for any* $f \in X$ *provided* $\lim_{m\to\infty} N_m = \infty$.