Article

Approximation Properties of the Vector Weak Rescaled Pure Greedy Algorithm

Xu Xu [1], Jinyu Guo [2], Peixin Ye [2] and Wenhui Zhang [2,*]

[1] School of Science, China University of Geosciences, Beijing 100083, China; xuxu@cugb.edu.cn
[2] School of Mathematics and LPMC, Nankai University, Tianjin 300071, China;
   2120210078@mail.nankai.edu.cn (J.G.); yepx@nankai.edu.cn (P.Y.)
* Correspondence: 1120200017@mail.nankai.edu.cn

Abstract: We first study the error performances of the Vector Weak Rescaled Pure Greedy Algorithm for simultaneous approximation with respect to a dictionary $\mathcal{D}$ in a Hilbert space. We show that the convergence rate of the Vector Weak Rescaled Pure Greedy Algorithm on $\mathcal{A}_1(\mathcal{D})$ and the closure of the convex hull of the dictionary $\mathcal{D}$ is optimal. The Vector Weak Rescaled Pure Greedy Algorithm has some advantages. It has a weaker convergence condition and a better convergence rate than the Vector Weak Pure Greedy Algorithm and is simpler than the Vector Weak Orthogonal Greedy Algorithm. Then, we design a Vector Weak Rescaled Pure Greedy Algorithm in a uniformly smooth Banach space setting. We obtain the convergence properties and error bound of the Vector Weak Rescaled Pure Greedy Algorithm in this case. The results show that the convergence rate of the VWRPGA on $\mathcal{A}_1(\mathcal{D})$ is sharp. Similarly, the Vector Weak Rescaled Pure Greedy Algorithm is simpler than the Vector Weak Chebyshev Greedy Algorithm and the Vector Weak Relaxed Greedy Algorithm.

Keywords: greedy algorithm; vector approximation; Hilbert spaces; modulus of smoothness; uniformly smooth Banach spaces; convergence; error bound

MSC: 41A50; 41A46; 41A65; 46A20

1. Introduction

Approximation using a sparse linear combination of elements from a fixed redundant family is actively used because of its concise representations and increased computational efficiency. It has been applied widely to signal processing, image compression, machine learning and PDE solvers (see [1–10]). Among others, simultaneous sparse approximation has been utilized in signal vector processing and multi-task learning (see [11–14]). It is well known that the greedy-type algorithms are powerful tools for generating such sparse approximations (see [15–19]). In particular, vector greedy algorithms are very efficient at approximating a given finite number of target elements simultaneously(see [20–23]). In this article, we propose a new vector greedy algorithm—the Vector Weak Rescaled Pure Greedy Algorithm (VWRPGA)—for simultaneous approximation. We estimate the error of the VWRPGA and show that its convergence rate on the convex hull of the dictionary is optimal.

Let X be a real Banach space with norm $\|\cdot\|$. We say a set of elements $\mathcal{D} \subset X$ is a dictionary, if $\|\varphi\| = 1$ for each $\varphi \in \mathcal{D}$ and $\overline{\mathrm{span}}(\mathcal{D}) = X$. We assume that every dictionary $\mathcal{D}$ is symmetric, i.e.,

$$\varphi \in \mathcal{D} \ \text{ implies } \ -\varphi \in \mathcal{D}.$$

If f_m is the output of a greedy algorithm after m iterations, then the efficiency of the approximation can be measured by the decay of the error $\|f - f_m\|$ as $m \to \infty$. We are mainly concerned with the error $\|f - f_m\|$. We want to know whether it tends to zero, as

$m \to \infty$. If it indeed converges to zero, then what is the convergence rate? To solve these problems, we need the following classes of elements.

For a general dictionary $\mathcal{D}$, we define the class of elements

$$\mathcal{A}_1^o(\mathcal{D}, M) := \left\{ f : f = \sum_{k \in \Lambda} c_k(f) \varphi_k, \varphi_k \in \mathcal{D}, |\Lambda| < \infty, \sum_{k \in \Lambda} |c_k(f)| \leq M \right\}$$

and $\mathcal{A}_1(\mathcal{D}, M)$ as the closure of $\mathcal{A}_1^o(\mathcal{D}, M)$. Let $\mathcal{A}_1(\mathcal{D})$ be the union of the classes $\mathcal{A}_1(\mathcal{D}, M)$ over all $M > 0$. Denote $A_1(\mathcal{D}) := \mathcal{A}_1(\mathcal{D}, 1)$. For $f \in \mathcal{A}_1(\mathcal{D})$, we define its norm as

$$\|f\|_{\mathcal{A}_1(\mathcal{D})} := \inf\{M : f \in \mathcal{A}_1(\mathcal{D}, M)\}.$$

We recall some related results in a Hilbert space for the reason that this kind of space has priority in geometric features and practical applications. Let H be a real Hilbert space with an inner product $\langle \cdot, \cdot \rangle$ and the norm $\|x\| := \langle x, x \rangle^{1/2}$.

The most natural greedy algorithm in a Hilbert space is the Pure Greedy Algorithm (PGA). This algorithm is also known as the Matching Pursuit in signal processing [24]. We recall its definition from [15].

PGA(H, $\mathcal{D}$):
- **Step** 0: Define $f_0 = 0$.
- **Step** m:
- If $f = f_{m-1}$, stop the algorithm and define $f_k = f_{m-1} = f$ for $k \geq m$.
- If $f \neq f_{m-1}$, choose an element $\varphi_m \in \mathcal{D}$ such that

$$|\langle f - f_{m-1}, \varphi_m \rangle| = \sup_{\varphi \in \mathcal{D}} |\langle f - f_{m-1}, \varphi \rangle|.$$

Define the next approximant to be

$$f_m = f_{m-1} + \langle f - f_{m-1}, \varphi_m \rangle \varphi_m,$$

and proceed to Step $m + 1$.

The first upper bound on the rate of convergence of the PGA for $f \in \mathcal{A}_1(\mathcal{D})$ was obtained in [15] as follows:

$$\|f - f_m\| \leq \|f\|_{\mathcal{A}_1(\mathcal{D})} m^{-\frac{1}{6}}, \quad m = 1, 2, \cdots.$$

Later, the above estimate of the PGA was improved in [25,26] to $\mathcal{O}(m^{-\frac{11}{62}})$ and $\mathcal{O}(m^{-\frac{s}{2(s+2)}})$, where s is the root of the equation

$$(1+x)^{\frac{1}{2+x}} \left(1 + \frac{1}{1+x}\right) - 1 - \frac{1}{x} = 0$$

on the closed interval $[1, 1.5]$. It is known that $\frac{s}{2(s+2)} > \frac{11}{62}$.

Note that when $\mathcal{D}$ is an ortho-normal basis of H, it is not difficult to prove that for any $f \in \mathcal{A}_1(\mathcal{D})$, there holds

$$\|f - f_m\| \leq c\|f\|_{\mathcal{A}_1(\mathcal{D})} m^{-\frac{1}{2}}, \ m = 1, 2, \cdots. \tag{1}$$

In addition, there exists an element $f^* \in \mathcal{A}_1(\mathcal{D})$ (see [27]) such that

$$\|f^* - f_m^*\| = c \cdot m^{-\frac{1}{2}}, \ m = 1, 2, \cdots.$$

Thus, inequality (1) cannot be improved for ortho-normal bases. A natural question arises: does inequality (1) hold for any dictionary $\mathcal{D} \subset H$? Unfortunately, the answer is negative.

In fact, Livshitz and Temlyakov [28] proved that there exists a dictionary $\mathcal{D} \subset H$, a positive constant C and an element $f \in A_1(\mathcal{D})$ such that

$$\|f - f_m\| \geq Cm^{-0.27}, \ \ m = 1, 2, \cdots.$$

This lower bound on the convergence rate of the PGA indicates that this algorithm does not attain the rate $O(m^{-\frac{1}{2}})$ for all $\mathcal{D}$.

In [15], the idea of the best approximation was introduced into the greedy algorithm, which formed the original idea of the Orthogonal Greedy Algorithm (OGA). In order to construct an approximation, the OGA takes the orthogonal projection of f on the subspace generated by all the chosen $\varphi_1, \ldots, \varphi_m$. We recall its definition from [15].

OGA(H, $\mathcal{D}$):
　　• **Step** 0: Define $f_0 = 0$.
　　• **Step** m:
　　- If $f = f_{m-1}$, stop the algorithm and define $f_k = f_{m-1} = f$ for $k \geq m$.
　　- If $f \neq f_{m-1}$, choose an element $\varphi_m \in \mathcal{D}$ such that

$$|\langle f - f_{m-1}, \varphi_m \rangle| = \sup_{\varphi \in \mathcal{D}} |\langle f - f_{m-1}, \varphi \rangle|.$$

Define the next approximant to be

$$f_m = P_m(f),$$

and proceed to Step $m + 1$, where P_m is the orthogonal projection onto $V_m := \mathrm{span}\{\varphi_1, \varphi_2, \cdots, \varphi_m\}$

In [15], it is shown that for any $\mathcal{D}$, the output of the OGA(H, $\mathcal{D}$) satisfies

$$\|f - f_m\| \leq c\|f\|_{\mathcal{A}_1(\mathcal{D})} m^{-\frac{1}{2}}, \ m = 1, 2, \cdots.$$

Note that when $\mathcal{D}$ is an ortho-normal basis of H, the OGA(H, $\mathcal{D}$) coincides with the PGA(H, $\mathcal{D}$). So, the rate $\mathcal{O}(m^{-\frac{1}{2}})$ is sharp.

The Relaxed Greedy Algorithm (RGA) is also a modification of PGA. We recall its definition from [15].

RGA(H, $\mathcal{D}$):
　　• **Step** 0: Define $f_0 = 0$.
　　• **Step** m:
　　- If $f = f_{m-1}$, stop the algorithm and define $f_k = f_{m-1} = f$ for $k \geq m$.
　　- If $f \neq f_{m-1}$, choose an element $\varphi_m \in \mathcal{D}$ such that

$$|\langle f - f_{m-1}, \varphi_m \rangle| = \sup_{\varphi \in \mathcal{D}} |\langle f - f_{m-1}, \varphi \rangle|.$$

For $m = 1$, define

$$f_1 = \langle f, \varphi_1 \rangle \varphi_1.$$

For $m \geq 2$, define the next approximant to be

$$f_m = \left(1 - \frac{1}{m}\right) f_{m-1} + \frac{1}{m} \varphi_m,$$

and proceed to Step $m + 1$.

It is shown in [15] that the RGA also achieves the rate $O(m^{-\frac{1}{2}})$ on $\mathcal{A}_1(\mathcal{D})$.

The Rescaled Pure Greedy Algorithm (RPGA) [17] is another kind of greedy algorithm which makes a modification to the rescaling process to replace the original output of the PGA with $f_m = s_m \hat{f}_m$ at each iteration. It is defined as follows.

RPGA($H, \mathcal{D}$):
  • **Step** 0: Define $f_0 = 0$.
  • **Step** m:
  - If $f = f_{m-1}$, stop the algorithm and define $f_k = f_{m-1} = f$ for $k \geq m$.
  - If $f \neq f_{m-1}$, choose an element $\varphi_m \in \mathcal{D}$ such that

$$|\langle f - f_{m-1}, \varphi_m \rangle| = \sup_{\varphi \in \mathcal{D}} |\langle f - f_{m-1}, \varphi \rangle|.$$

With

$$\lambda_m = \langle f - f_{m-1}, \varphi_m \rangle, \quad \hat{f}_m := f_{m-1} + \lambda_m \varphi_m, \quad s_m = \frac{\langle f, \hat{f}_m \rangle}{\|\hat{f}_m\|^2},$$

define the next approximant to be

$$f_m = s_m \hat{f}_m,$$

and proceed to Step $m + 1$.

In [17], the convergence rate of the RPGA was obtained as follows:

$$\|f - f_m\| \leq \|f\|_{\mathcal{A}_1(\mathcal{D})} (m+1)^{-\frac{1}{2}}, \quad m = 0, 1, 2, \cdots.$$

It is worth noting that the supremum of the inner product might not be attainable. To remedy this problem, the original condition on the selection of φ_m is replaced by

$$|\langle f - f_{m-1}, \varphi_m \rangle| \geq t_m \sup_{\varphi \in \mathcal{D}} |\langle f - f_{m-1}, \varphi \rangle|,$$

where $0 < t_m \leq 1$. This is often referred to as the "weak" condition. The study on the weak version of the above algorithms can be found in [15,25,29,30].

Meanwhile, building simultaneous approximations for a given vector of elements brings about the so-called vector-type greedy algorithms. Instead of running the algorithm for a finite collection of elements $f^1, \ldots, f^N$ each time separately, the vector greedy algorithm manages to obtain a simultaneous approximation of all elements with a single run. Hence, the complexity of calculation and the storage of information can be reduced greatly. Now, it comes to the question of how well this type of algorithm can perform. Namely, we need to measure its efficiency via its error bound. The Vector Weak Pure Greedy Algorithm (VWPGA, which is also referred to as the Vector Weak Greedy Algorithm (VWGA)) and the Vector Weak Orthogonal Greedy Algorithm (VWOGA) have been introduced and studied in [21–23].

We recall the definitions of the VWPGA and VWOGA from [23] as follows. Let $\tau = \{t_m\}_{m=1}^{\infty}$ and $0 < t_m \leq 1$ be a given sequence.

VWPGA($H, \mathcal{D}$): $\quad f^i \in H, i = 1, \ldots, N$ is the target element.
  • **Step** 0: Define $f_0^i := 0, i = 1, \ldots N$.
  • **Step** m:
  - If $f^i = f_{m-1}^i$, stop the algorithm and define $f_k^i = f_{m-1}^i = f$ for $k \geq m$.
  - If $f \neq f_{m-1}$, choose an element $\varphi_m \in \mathcal{D}$ such that

$$\max_i |\langle f^i - f_{m-1}^i, \varphi_m \rangle| \geq t_m \max_i \sup_{\varphi \in \mathcal{D}} |\langle f^i - f_{m-1}^i, \varphi \rangle|.$$

Define the next approximant to be

$$f_m^i := f_{m-1}^i + \langle f^i - f_{m-1}^i, \varphi_m \rangle \varphi_m, \quad i = 1, 2, \ldots, N,$$

and proceed to Step $m + 1$.

VWOGA($H, \mathcal{D}$): $f^i \in H, i = 1, \ldots, N$ is the target element.
  • **Step** 0: Define $f_0^i := 0, i = 1, \ldots N$.
  • **Step** m:
  - If $f^i = f_{m-1}^i$, stop the algorithm and define $f_k^i = f_{m-1}^i = f$ for $k \geq m$.
  - If $f^i \neq f_{m-1}^i$. Let i_m be such that

$$\|f^{i_m} - f_{m-1}^{i_m}\| \geq \|f^i - f_{m-1}^i\|, \;\; i = 1, \ldots N.$$

Choose an element $\varphi_m \in \mathcal{D}$ such that

$$|\langle f^{i_m} - f_{m-1}^{i_m}, \varphi_m \rangle| \geq t_m \sup_{\varphi \in \mathcal{D}} |\langle f^{i_m} - f_{m-1}^{i_m}, \varphi \rangle|.$$

Define the next approximant to be

$$f_m^i = P_m(f^i), \;\; i = 1, 2, \ldots, N,$$

and proceed to Step $m + 1$, where P_m is the orthogonal projection onto $V_m := \mathrm{span}\{\varphi_1, \varphi_2, \cdots, \varphi_m\}$.

We list the results on the convergence rate of the VWPGA and VWOGA in [23] as follows.

Theorem 1. *Let $\tau := \{t_k\}_{k=1}^\infty$, $t_k = t$, $0 < t \leq 1$ be a given real sequence. Then, for any $f^1, \ldots, f^N$, $f^i \in A_1(\mathcal{D})$, the output $\{f_m^i\}_{m \geq 0}$ of the VWPGA satisfies*

$$\sum_{i=1}^N \|f^i - f_m^i\|^2 \leq \left(1 + \frac{mt^2}{N}\right)^{-t/(2N+t)} N^{\frac{2N+2t}{2N+t}}.$$

Theorem 2. *Let $\tau := \{t_k\}_{k=1}^\infty$, $t_k = t$, $0 < t \leq 1$ be a given real sequence. Then, for any $f^i \in A_1(\mathcal{D})$, the output $\{f_m^i\}_{m \geq 0}$ of the VWOGA satisfies*

$$\|f^i - f_m^i\| \leq \min\left\{1, \left(\frac{N}{mt^2}\right)^{1/2}\right\}, \;\; i = 1, \ldots, N.$$

Improvements to the above estimates are made in [19,21,22]. The results indicate that the VWOGA achieves a better convergence rate on $A_1(\mathcal{D})$ than that of the VWPGA.

In [23], the authors gave a sufficient condition of convergence for the VWPGA.

Theorem 3. *Assume that $\sum_{m=1}^\infty \dfrac{t_m}{m} = \infty$. Then, for any dictionary and any finite elements $f^i \in H, i = 1, \ldots, N$, the VWPGA satisfies*

$$\lim_{m \to \infty} \|f^i - f_m^i\| = 0.$$

Motivated by these studies, we design the Vector Weak Rescaled Pure Greedy Algorithm (VWRPGA) and study its efficiency. The remainder of the paper is organized as follows. In Section 2, we deal with the case of Hilbert spaces. In Section 3, we deal with the case of Banach spaces. In Section 4, we draw the conclusions. Below, we provide more details.

In Section 2, we define the VWRPGA in Hilbert spaces and study its approximation properties. We first prove that

$$\sum_{m=1}^\infty t_m^2 = \infty$$

is the sufficient convergence condition of the VWRPGA for any $f^i \in H$ and $\mathcal{D} \subset H$, $i = 1, \cdots, N$. This convergence condition is weaker than that of the VWPGA. Then, we prove that the error bound of the VWRPGA on $A_1(\mathcal{D})$ satisfies

$$\|f^i - f_m^{i,v,\tau,r}\| \leq \min\left\{1, \left(\frac{1}{N}\sum_{k=1}^m t_k^2\right)^{-\frac{1}{2}}\right\}.$$

When $t_1 = t_2 = \cdots = t_m = 1$, we show that convergence rate of the VWRPGA on $\mathcal{A}_1(\mathcal{D})$ is $O(m^{-\frac{1}{2}})$, which is sharp. This convergence rate is better than that of the VWPGA. In particular, this advantage is more obvious when N is large. The VWRPGA is more efficient than VWOGA from the viewpoint of computational complexity. This is because, for N target elements, the VWRPGA only needs to solve N one-dimensional optimization problems, while the VWOGA involves N m-dimensional optimization problems.

In Section 3, we define the VWRPGA for some uniformly smooth Banach spaces. The sufficient condition of the convergence of the VWRPGA is obtained in this case. It seems that this is the first result on the convergence analysis of the vector greedy algorithms in the Banach space setting. Then, we derive the error bound of the VWRPGA. The results show that the convergence rate of the VWRPGA on $\mathcal{A}_1(\mathcal{D})$ is sharp. We compare the approximation properties of the VWRPGA with those of the Vector Weak Chebyshev Greedy Algorithm (VWCGA) and the Vector Weak Relaxed Greedy Algorithm (VWRGA). We show that the VWRPGA has better convergence properties than the VWRGA. Similarly, the computational complexity of the VWRPGA is essentially smaller than those of the VWCGA and VWRGA.

In Section 4, we draw the conclusions of our study. Our results show that the VWRPGA is the simplest vector greedy algorithm for simultaneous approximation with the best convergence property and the optimal convergence rate. We also discuss the possible applications of the VWRPGA in multi-task learning and signal vector processing.

2. The VWRPGA for Hilbert Spaces

In this section, we define the VWRPGA in Hilbert spaces and obtain its sufficient condition of convergence together with an estimate of its error bound. Based on these results, we compare the VWRPGA with the VWPGA and the VWOGA.

Firstly, we recall the definition of the Weak Rescaled Pure Greedy Algorithm (WRPGA) in Hilbert spaces from [17]. Let $\tau = \{t_m\}_{m=1}^\infty, 0 < t_m \leq 1$ be a given sequence. The WRPGA consists of the following stages:

WRPGA($H, \mathcal{D}$):
 • **Step 0:** Define $f_0 = 0$.
 • **Step m:**
- If $f = f_{m-1}$, stop the algorithm and define $f_k = f_{m-1} = f$ for $k \geq m$.
- If $f \neq f_{m-1}$, choose an element $\varphi_m \in \mathcal{D}$ such that

$$|\langle f - f_{m-1}, \varphi_m \rangle| \geq t_m \sup_{\varphi \in \mathcal{D}} |\langle f - f_{m-1}, \varphi \rangle|.$$

With

$$\lambda_m = \langle f - f_{m-1}, \varphi_m \rangle, \ \hat{f}_m := f_{m-1} + \lambda_m \varphi_m, \ s_m = \frac{\langle f, \hat{f}_m \rangle}{\|\hat{f}_m\|^2},$$

define the next approximation to be

$$f_m = s_m \hat{f}_m,$$

and proceed to Step $m + 1$.

The error bound of the WRPGA has been obtained as follows.

Theorem 4 (see Theorem 4.1 in [17]). *If $f \in \mathcal{A}_1(\mathcal{D}) \subset H$, then the output $\{f_m\}_{m \geq 0}$ of the WRPGA satisfies the error estimate*

$$\|f - f_m\| \leq \|f\|_{\mathcal{A}_1(\mathcal{D})} \left(\sum_{k=1}^m t_k^2 \right)^{-\frac{1}{2}}.$$

Based on the WRPGA, we can define the VWRPGA. Let $\tau = \{t_m\}_{m=1}^{\infty}, 0 < t_m \leq 1$ be a given sequence. Now, we define the VWRPGA using the following steps:

VWRPGA$(H, \mathcal{D})$:　Given $f^i \in H, i = 1, \ldots, N$.
- **Step 0:** Define $f_0^{i,v,\tau,r} := 0, i = 1, \ldots N$.
- **Step m:**
- If $f^i = f_{m-1}^{i,v,\tau,r}$, stop the algorithm and define $f_k^{i,v,\tau,r} = f_{m-1}^{i,v,\tau,r} = f$ for $k \geq m$.
- If $f^i \neq f_{m-1}^{i,v,\tau,r}$, let i_m be such that

$$\|f^{i_m} - f_{m-1}^{i_m,v,\tau,r}\| = \max_{1 \leq i \leq N} \|f^i - f_{m-1}^{i,v,\tau,r}\|.$$

Choose an element $\varphi_m \in \mathcal{D}$ such that

$$\langle f^{i_m} - f_{m-1}^{i_m,v,\tau,r}, \varphi_m \rangle \geq t_m \sup_{\varphi \in \mathcal{D}} \langle f^{i_m} - f_{m-1}^{i_m,v,\tau,r}, \varphi \rangle.$$

With

$$\lambda_m^i = \langle f^i - f_{m-1}^{i,v,\tau,r}, \varphi_m \rangle, \quad \hat{f}_m^i := f_{m-1}^{i,v,\tau,r} + \lambda_m^i \varphi_m, \quad s_m^i = \frac{\langle f^i, \hat{f}_m^i \rangle}{\|\hat{f}_m^i\|^2},$$

define the next approximation to be

$$f_m^{i,v,\tau,r} := s_m^i \hat{f}_m^i,$$

and proceed to Step $m + 1$.

We establish in this section two typical results on the approximation properties of the VWRPGA $(H, \mathcal{D})$. We first give the sufficient condition for the convergence of the VWRPGA for any dictionary $\mathcal{D}$ and any $f^i, i = 1, \cdots, N$.

Theorem 5. *Assume $\sum_{m=1}^{\infty} t_m^2 = \infty$. Then, the VWRPGA converges for any dictionary $\mathcal{D}$ and any $f^i \in H, i = 1, \ldots, N$.*

In the proof of Theorem 5, we will reduce the approximation of the general element to that of the element from $A_1(\mathcal{D})$. To this end, we recall from [31] the following lemmas on the approximation properties of $A_1(\mathcal{D})$.

Lemma 1. *Let X be a Banach space and $\mathcal{D} \subset X$ be a dictionary. Then, for any $\epsilon > 0$ and any $f \in X$, there exists $f^\epsilon \in X$ such that*

$$\|f - f^\epsilon\| < \epsilon$$

and

$$\frac{f^\epsilon}{A(\epsilon)} \in A_1(\mathcal{D}),$$

with some number $A(\epsilon) > 0$.

Lemma 2. *For any $f \in H$ and any dictionary $\mathcal{D}$, we have*

$$\sup_{\varphi \in \mathcal{D}} \langle f, \varphi \rangle = \sup_{g \in A_1(\mathcal{D})} \langle f, g \rangle.$$

Proof of Theorem 5. Note that $f_m^{i,v,\tau,r}$ is the orthogonal projection of f^i onto the one-dimensional space $\text{span}\{\hat{f}_m^i\}$. Thus, it is the best approximation of $f_m^{i,v,\tau,r}$ from $\text{span}\{\hat{f}_m^i\}$.

Let $r_m^i := \|f^i - f_m^{i,v,\tau,r}\|$, $i = 1, \cdots, N$, be the residual of $f_m^{i,v,\tau,r}$. By the definition of $\hat{f}_m^i$ and the choice of λ_m^i, we have

$$
\begin{aligned}
\|r_m^i\|^2 &= \|f^i - f_m^{i,v,\tau,r}\|^2 \\
&= \|f^i - s_m^i \hat{f}_m^i\|^2 \\
&\leq \|f^i - \hat{f}_m^i\|^2 \\
&\leq \langle f^i - f_{m-1}^{i,v,\tau,r} - \lambda_m^i \varphi_m, f^i - f_{m-1}^{i,v,\tau,r} - \lambda_m^i \varphi_m \rangle \\
&= \|f^i - f_{m-1}^{i,v,\tau,r}\|^2 - 2\lambda_m^i \langle f^i - f_{m-1}^{i,v,\tau,r}, \varphi_m \rangle + (\lambda_m^i)^2 \\
&= \|r_{m-1}^i\|^2 - \langle f^i - f_{m-1}^{i,v,\tau,r}, \varphi_m \rangle^2.
\end{aligned}
\tag{2}
$$

The latter inequality implies that $\{\|r_m^i\|\}_{m=0}^\infty$ is a decreasing sequence. According to the Monotone Convergence Theorem, we know that $\lim_{m\to\infty} \|r_m^i\|$ exists, $i = 1, \ldots, N$.

We prove that $\lim_{m\to\infty} \|r_m^i\| = 0$ by contradiction. Assume $\lim_{m\to\infty} \|r_m^i\| \geq a > 0$, $i = 1, \ldots, N$. Then, for any m, we have $\|r_m^i\| \geq a$. By (2), we obtain that

$$
\begin{aligned}
\sum_{i=1}^N \|r_m^i\|^2 &\leq \sum_{i=1}^N \|r_{m-1}^i\|^2 - \sum_{i=1}^N \langle f^i - f_{m-1}^{i,v,\tau,r}, \varphi_m \rangle^2 \\
&= \sum_{i=1}^N \|r_{m-1}^i\|^2 \left(1 - \frac{\sum_{i=1}^N \langle f^i - f_{m-1}^{i,v,\tau,r}, \varphi_m \rangle^2}{\sum_{i=1}^N \|r_{m-1}^i\|^2} \right) \\
&\leq \sum_{i=1}^N \|r_{m-1}^i\|^2 \left(1 - \frac{\langle f^{i_m} - f_{m-1}^{i_m,v,\tau,r}, \varphi_m \rangle^2}{N \|r_{m-1}^{i_m}\|^2} \right) \\
&\leq \sum_{i=1}^N \|f^i\|^2 \prod_{j=1}^m \left(1 - \frac{\langle f^{i_j} - f_{j-1}^{i_j,v,\tau,r}, \varphi_j \rangle^2}{N \|r_{j-1}^{i_j}\|^2} \right).
\end{aligned}
$$

Denote

$$x_j = \frac{\langle f^{i_j} - f_{j-1}^{i_j,v,\tau,r}, \varphi_j \rangle^2}{N \|r_{j-1}^{i_j}\|^2}.$$

By the inequality $1 - x \leq \frac{1}{1+x}$, $0 \leq x \leq 1$, we obtain that

$$
\begin{aligned}
\sum_{i=1}^N \|r_m^i\|^2 &\leq \sum_{i=1}^N \|f^i\|^2 \prod_{j=1}^m \left(\frac{1}{1+x_j} \right) \\
&\leq \sum_{i=1}^N \|f^i\|^2 \frac{1}{1 + \sum_{j=1}^m x_j}.
\end{aligned}
\tag{3}
$$

Then, we come to obtain a lower estimate for x_j, $j = 1, \ldots, m$.

Set $\epsilon = \frac{a}{2}$. In view of Lemma 1, we can find f_j^ϵ such that

$$\|f^{i_j} - f_j^\epsilon\| < \epsilon$$

and

$$\frac{f_j^\epsilon}{A(\epsilon)} \in A_1(\mathcal{D}),$$

with some number $A(\epsilon) > 0$.

Using Lemma 2, we have

$$
\begin{aligned}
\langle f^{i_j} - f_{j-1}^{i_j,v,\tau,r}, \varphi_j \rangle &\geq t_j \sup_{\varphi \in \mathcal{D}} \langle f^{i_j} - f_{j-1}^{i_j,v,\tau,r}, \varphi \rangle \\
&= t_j \sup_{g \in A_1(\mathcal{D})} \langle f^{i_j} - f_{j-1}^{i_j,v,\tau,r}, g \rangle \\
&\geq t_j A(\epsilon)^{-1} \langle f^{i_j} - f_{j-1}^{i_j,v,\tau,r}, f_j^\epsilon \rangle.
\end{aligned}
\tag{4}
$$

Since $f_m^{i,v,\tau,r}$ is the orthogonal projection of f^i onto $\mathrm{span}\{\hat{f}_m^i\}$, we have

$$\langle f^i - f_m^{i,v,\tau,r}, f_m^{i,v,\tau,r} \rangle = 0.$$

Then,

$$
\begin{aligned}
\langle f^{i_j} - f_{j-1}^{i_j,v,\tau,r}, f_j^\epsilon \rangle &= \langle r_{j-1}^{i_j}, f^{i_j} + f_j^\epsilon - f^{i_j} \rangle \\
&= \langle r_{j-1}^{i_j}, r_{j-1}^{i_j} + f_{j-1}^{i_j,v,\tau,r} \rangle - \langle r_{j-1}^{i_j}, f^{i_j} - f_j^\epsilon \rangle \\
&> \|r_{j-1}^{i_j}\|^2 - \|r_{j-1}^{i_j}\| \cdot \epsilon.
\end{aligned}
\tag{5}
$$

Combining (4) and (5) with $\epsilon = \frac{a}{2}$, we obtain

$$
\begin{aligned}
x_j &\geq \frac{1}{N\|r_{j-1}^{i_j}\|^2} \cdot \left(\frac{\langle f^{i_j} - f_{j-1}^{i_j,v,\tau,r}, f_j^\epsilon \rangle}{A(\epsilon)} t_j \right)^2 \\
&\geq \frac{1}{N} \cdot \left(\frac{\|r_{j-1}^{i_j}\| - \epsilon}{A(\epsilon)} t_j \right)^2 \\
&\geq \frac{a^2}{4NA(\epsilon)^2} t_j^2.
\end{aligned}
\tag{6}
$$

Combining (3) with (6), we can obtain that

$$\sum_{i=1}^N \|r_m^i\|^2 \leq \sum_{i=1}^N \|f^i\|^2 \frac{1}{1 + \dfrac{a^2}{4NA(\epsilon)^2} \displaystyle\sum_{j=1}^m t_j^2}.$$

The assumption $\displaystyle\sum_{m=1}^\infty t_m^2 = \infty$ implies that $\displaystyle\sum_{i=1}^N \|r_m^i\|^2 \to 0$ as $m \to \infty$.

It is obvious that $\lim_{m \to \infty} \|r_m^i\| = 0$ for $i = 1, \ldots, N$. Hence, we obtain a contradiction, which proves this theorem. $\quad\square$

Remark 1. *It is known from Theorem 2.1 in [32] that $\displaystyle\sum_{m=1}^\infty t_m^2 = \infty$ is also the necessary condition for the convergence of the VWRPGA.*

Remark 2. *According to the Cauchy–Schwartz inequality, we know that*

$$\sum_{m=1}^{\infty} \frac{t_m}{m} \le \left(\sum_{m=1}^{\infty} t_m^2 \right)^{\frac{1}{2}} \left(\sum_{m=1}^{\infty} \frac{1}{m^2} \right)^{\frac{1}{2}}.$$

Hence,

$$\sum_{m=1}^{\infty} \frac{t_m}{m} = \infty \quad \text{implies} \quad \sum_{m=1}^{\infty} t_m^2 = \infty.$$

On the other hand, taking $t_m = m^{-\frac{1}{2}}$, $m = 1, 2, \cdots$, we notice that

$$\sum_{m=1}^{\infty} t_m^2 = \infty, \quad \sum_{m=1}^{\infty} \frac{t_m}{m} < \infty.$$

Therefore, the convergence condition of the VWPRGA is weaker than that of the VWPGA.

The following theorem gives the error bound of the VWRPGA $(H, \mathcal{D})$ for $f^i \in A_1(\mathcal{D})$, $i = 1, \ldots, N$.

Theorem 6. *Let $\tau := \{t_k\}_{k=1}^{\infty}$, $0 < t_k \le 1$ be a weakness sequence. If $f^i \in A_1(\mathcal{D}) \subset H$, $i = 1, \ldots, N$. Then, we have for the VWRPGA $(H, \mathcal{D})$*

$$\|f^i - f_m^{i,v,\tau,r}\| \le \min \left\{ 1, \left(\frac{1}{N} \sum_{k=1}^{m} t_k^2 \right)^{-\frac{1}{2}} \right\}.$$

Proof. We establish the approximation error of the VWRPGA based on the methods of [25]. The main idea of this proof is that the VWRPGA can be seen as a realization of the WRPGA with a particular weakness sequence.

Let $i \in \{1, \ldots, N\}$. Under the assumption that $f^i \in A_1(\mathcal{D})$ and the fact that the sequence $\{\|f^i - f_m^{i,v,\tau,r}\|\}_{m=0}^{\infty}$ is decreasing, we have

$$\|f^i - f_m^{i,v,\tau,r}\| \le 1.$$

Thus, we only need to prove the estimate below:

$$\|f^i - f_m^{i,v,\tau,r}\| \le \left(\frac{1}{N} \sum_{k=1}^{m} t_k^2 \right)^{-\frac{1}{2}}, \quad i = 1, \ldots, N.$$

At step k, the VWRPGA chooses φ_k from the $\mathcal{D}$ in terms of only one remainder from $\{r_{k-1}^1, \ldots, r_{k-1}^N\}$. Then, each f^i, $i = 1, \cdots, N$ has been used different times to choose φ_k when the VWRPGA carries on to step m. Now, we record the usage of f^i, $i = 1, \ldots, N$.

For every l, $l = 1, \ldots, N$, denote $E_l := \{k | i_k = l, 1 \le k \le m\}$ (i_k is defined in the definition of VWRPGA). Then, we have

$$E_1 \cup E_2 \cup \ldots \cup E_N = \{1, \ldots, m\}, \quad E_i \cap E_j = \varnothing \text{ if } i \ne j.$$

Hence,

$$\|f^l - f_{k-1}^{l,v,\tau,r}\| = \|f^{i_k} - f_{k-1}^{i_k,v,\tau,r}\| = \max_{1 \le i \le N} \|f^i - f_{k-1}^{i,v,\tau,r}\|, \quad k \in E_l.$$

Using $\sum_{k=1}^{m} t_k^2 = \sum_{l=1}^{N} \sum_{k \in E_l} t_k^2$, we can find $l_0 \in [1, N]$ such that

$$\sum_{k \in E_{l_0}} t_k^2 \ge \frac{1}{N} \sum_{k=1}^{m} t_k^2.$$

Next, let $k_0 = \max\{k | k \in E_{l_0}\}, k_0 \le m$. We have

$$\max_{1 \le i \le N} \|f^i - f^{i,v,\tau,r}_m\| \le \max_{1 \le i \le N} \|f^i - f^{i,v,\tau,r}_{k_0-1}\| = \|f^{l_0} - f^{l_0,v,\tau,r}_{k_0-1}\|. \tag{7}$$

Now, we only consider element $f^{l_0} \in H$. For $f^{l_0} \in H$, we can obtain $f^{l_0,v,\tau,r}_1, \dots, f^{l_0,v,\tau,r}_m$ as an application of the WRPGA with the weakness sequence $\tau^{l_0} := \{t^{l_0}_k\}$ given by

$$t^{l_0}_k = \begin{cases} t_k, & k \in E_{l_0} \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, by Theorem 4, we obtain

$$\|f^{l_0} - f^{l_0,v,\tau,r}_{k_0-1}\| \le \left(1 + \sum_{k \in E_{l_0} - \{k_0\}} t^2_k\right)^{-\frac{1}{2}}$$

$$\le \left(1 + \sum_{k \in E_{l_0}} t^2_k - 1\right)^{-\frac{1}{2}}$$

$$\le \left(\frac{1}{N} \sum_{k=1}^{m} t^2_k\right)^{-\frac{1}{2}}.$$

Together with (7), we complete the proof of Theorem 6. $\quad\square$

We recall the theorem in [21] about the error estimate of the VWPGA.

Theorem 7. *Let $\tau := \{t_k\}_{k=1}^{\infty}, 0 < t_k \le 1$ be a decreasing sequence. Then, for any $f^1, \dots, f^N, f^i \in A_1(\mathcal{D})$, the output $\{f^i_m\}_{m \ge 0}$ of the VWPGA satisfies*

$$\sum_{i=1}^{N} \|f^i - f^i_m\|^2 \le N^2 \left(1 + \frac{1}{N} \sum_{i=1}^{m} t_k{}^2\right)^{\frac{-t_m}{2N^{1/2}+t_m}}.$$

We observe from Theorem 7, for a fixed m, the error of the VWPGA increases as the number of the target elements increases. The exponent is close to zero as long as N is sufficiently large.

Taking $t_k = 1, k = 1, 2, \dots$ in Theorem 7, we yield the following theorem, which gives the convergence rate of the VWPGA.

Theorem 8. *Let $\tau := \{t_k\}_{k=1}^{\infty}, t_k = 1$. Then, for any $f^1, \dots, f^N, f^i \in A_1(\mathcal{D})$, the output $\{f^i_m\}_{m \ge 0}$ of the VWPGA satisfies*

$$\sum_{i=1}^{N} \|f^i - f^i_m\| \le N^2 \left(1 + \frac{m}{N}\right)^{-\frac{1}{2N^{1/2}+1}}.$$

Again, by taking $t_k = 1, k = 1, 2, \dots$ in Theorem 6, we obtain the following theorem.

Theorem 9. *Let $\tau := \{t_k\}_{k=1}^{\infty}, t_k = 1$. If $f^i \in A_1(\mathcal{D}) \subset H, i = 1, \dots, N$. Then, for the VWRPGA $(H, \mathcal{D})$,*

$$\|f^i - f^{i,v,\tau,r}_m\| \le \min\left\{1, \left(\frac{m}{N}\right)^{-\frac{1}{2}}\right\}.$$

Remark 3. *From Theorems 8 and 9, we see that the VWRPGA provides a significantly better convergence rate than the VWPGA. In particular, this advantage is more obvious when N is large.*

Remark 4. *It is known from Theorems 2 and 9 that the approximation property of the VWRPGA is almost the same as that of the VWOGA. While the VWRPGA is simpler than the VWOGA from the viewpoint of computational complexity. For N target elements, from the definitions of the algorithms, one can see that the VWOGA needs to solve N m-dimensional optimization problems. However, the VWRPGA only needs to solve N one-dimensional optimization problems. Then, this makes the VWRPGA easier to implement than the VWOGA in practical applications.*

3. The VWRPGA for Banach Spaces

In this section, we consider the VWRPGA in the context of Banach spaces. We remark that there are two natural generations of the PGA in the case of Banach space X: the X-greedy algorithm and the dual greedy algorithm. However, there are no general results on convergence and error bound of these two algorithms, cf. [29]. On the other hand, the WOGA, WRGA, WRPGA and VWOGA have been successfully generalized to the case of Banach spaces. We first recall from [31] the definition of the Weak Chebyshev Greedy Algorithm (WCGA), which is a natural generalization of the WOGA.

For any non-zero element $f \in X$, we denote by F_f a norming functional for f:

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

The existence of such a functional is guaranteed by the Hahn–Banach theorem. Let $\tau = \{t_m\}_{m=1}^{\infty}, 0 < t_m \leq 1$ be a given sequence. The WCGA is defined as follows.

WCGA($X, \mathcal{D}$):
    • **Step 0:** Define $f_0 = 0$.
    • **Step m:**
    - If $f = f_{m-1}$, stop the algorithm and define $f_k = f_{m-1} = f$ for $k \geq m$.
    - If $f \neq f_{m-1}$, choose an element $\varphi_m \in \mathcal{D}$ such that

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

Set

$$\Phi_m := \text{span}\{\varphi_1, \varphi_2, \cdots, \varphi_m\}.$$

Define f_m to be the best approximant to f from Φ_m and proceed to Step $m + 1$.

To estimate the error of WCGA, we shall utilize some geometric aspects of Banach spaces. For a Banach space X, we define $\rho(u)$, the modulus of smoothness of X, as

$$\rho(u) := \sup_{f,g \in X, \|f\|=\|g\|=1} \left\{ \frac{\|f + ug\| + \|f - ug\|}{2} - 1 \right\}, \quad u > 0.$$

A uniformly smooth Banach space is one with the property

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

We shall only consider Banach spaces whose modulus of smoothness satisfies the inequality

$$\rho(u) \leq \gamma u^q, \quad 1 < q \leq 2,$$

where γ is a constant independent of u.

A typical example of a uniformly smooth Banach space is the Lebesgue space L_p, $1 < p < \infty$. It is known from [33] that

$$\rho(u) \leq \begin{cases} u^p/p & \text{if } 1 < p \leq 2, \\ (p-1)u^2/2 & \text{if } 2 \leq p < \infty. \end{cases} \tag{8}$$

Moreover, we obtain from [34] that for any X with $dim X = \infty$,

$$\rho(u) \geq (1 + u^2)^{\frac{1}{2}} - 1$$

and for any X with $dim X \geq 2$,

$$\rho(u) \geq C u^2, \ \ C > 0.$$

The following error bound of the WCGA on $A_1(\mathcal{D})$ has been established in [31].

Theorem 10. *Let X be a Banach space with modulus of smoothness $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. If $f \in A_1(\mathcal{D}) \subset X$, then the output $\{f_m\}_{m \geq 0}$ of the WCGA$(X, \mathcal{D})$ satisfies the inequality*

$$\|f - f_m\| \leq C_1(q, \gamma) \left(1 + \sum_{k=1}^{m} t_k^{\frac{q}{q-1}} \right)^{-1 + \frac{1}{q}},$$

where the constant $C_1(q, \gamma)$ depends only on q and γ.

Taking $\{t_k\}_{k=1}^{\infty}$, $t_k = 1, k = 1, 2, \cdots$, Theorem 10 implies the following corollary, which can be seen in [31].

Corollary 1. *Let X be a Banach space with modulus of smoothness $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. Then, for any $f \in A_1(\mathcal{D})$, the output $\{f_m\}_{m \geq 0}$ of the WCGA$(X, \mathcal{D})$ satisfies the inequality*

$$\|f - f_m\| \leq c \cdot m^{-1 + \frac{1}{q}}.$$

In order to show the convergence rate $O(m^{-1 + \frac{1}{q}})$ cannot be improved, we now take $L_p, 1 < p < \infty$ as an example.

Let $1 < p \leq 2$ be fixed. Combining Corollary 1 with inequality (8), we have for any $\mathcal{D} \subset L_p$ and any $f \in L_p$

$$\|f - f_m\| \leq c \cdot m^{-1 + \frac{1}{p}}. \tag{9}$$

When $\mathcal{D}$ is a wavelet basis of L_p, it is known from [35] that there is a $f \in A_1(\mathcal{D})$ such that

$$\|f - f_m\| \geq c \cdot m^{-1 + \frac{1}{p}}.$$

Thus, inequality (9) could not be improved.

Similarly, let $p > 2$ be fixed. Combining Corollary 1 with inequality (8), we have for any $\mathcal{D} \subset L_p$ and any $f \in L_p$

$$\|f - f_m\| \leq c \cdot m^{-\frac{1}{2}}. \tag{10}$$

When $\mathcal{D}$ is the trigonometric system of L_p, it is known from [36] that there is a $f \in A_1(\mathcal{D})$ such that

$$\|f - f_m\| \geq c \cdot m^{-\frac{1}{2}}.$$

Thus, inequality (10) could not be improved.

Then, the convergence rate $O(m^{-1 + \frac{1}{q}})$ in Corollary 1 serves as a benchmark for the performance of greedy algorithms in uniformly smooth Banach spaces.

Next, we recall the definition of the WRGA in the Banach space setting from [31]. Let $\tau = \{t_m\}_{m=1}^{\infty}$, $0 < t_m \leq 1$, be a given sequence. The WRGA is defined as follows.

WRGA$(X, \mathcal{D})$:
- **Step** 0: Define $f_0 = 0$.
- **Step** m:
- If $f = f_{m-1}$, stop the algorithm and define $f_k = f_{m-1} = f$ for $k \geq m$.

- If $f \neq f_{m-1}$, choose a element $\varphi_m \in \mathcal{D}$ such that

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

Find $0 \leq \lambda_m \leq 1$ such that

$$\|f - ((1 - \lambda_m)f_{m-1} + \lambda_m \varphi_m)\| = \inf_{0 \leq \lambda \geq 1} \|f - ((1 - \lambda)f_{m-1} + \lambda \varphi_m)\|.$$

Define $f_m := (1 - \lambda_m)f_{m-1} + \lambda_m \varphi_m$, and proceed to Step $m + 1$.
The following error bound of the WRGA on $A_1(\mathcal{D})$ has been established in [31].

Theorem 11. *Let X be a Banach space with a modulus of smoothness $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. If $f \in A_1(\mathcal{D}) \subset X$, then the output $\{f_m\}_{m \geq 0}$ of the WRGA $(X, \mathcal{D})$ satisfies the inequality*

$$\|f - f_m\| \leq C_2(q, \gamma) \left(1 + \sum_{k=1}^{m} t_k^p\right)^{-\frac{1}{p}}, \quad p = \frac{q}{q-1},$$

where the constant $C_2(q, \gamma)$ depends only on q and γ.

Now, we turn to the vector greedy algorithms. Let $\tau = \{t_m\}_{m=1}^{\infty}$, $0 < t_m \leq 1$ be a given sequence. The Vector Weak Chebyshev Greedy Algorithm (VWCGA) [22] is defined as follows.

VWCGA(X, $\mathcal{D}$):
- **Step 0:** Define $f_0^{i,v,\tau,c} := 0, i = 1, \ldots N$.
- **Step m:**
- If $f^i = f_{m-1}^{i,v,\tau,c}$, stop the algorithm and define $f_k^{i,v,\tau,c} = f_{m-1}^{i,v,\tau,c} = f$ for $k \geq m$.
- If $f^i \neq f_{m-1}^{i,v,\tau,c}$, let i_m be such that

$$\|f^{i_m} - f_{m-1}^{i_m,v,\tau,c}\| = \max_{1 \leq i \leq N} \|f^i - f_{m-1}^{i,v,\tau,c}\|.$$

Choose an element $\varphi_m \in \mathcal{D}$ such that

$$F_{f^{i_m} - f_{m-1}^{i_m,v,\tau,c}}(\varphi_m) \geq t_m \sup_{\varphi \in \mathcal{D}} F_{f^{i_m} - f_{m-1}^{i_m,v,\tau,c}}(\varphi).$$

Set

$$\Phi_m := \text{span}\{\varphi_1, \varphi_2, \cdots, \varphi_m\}.$$

Define f_m^i to be the best approximant to f^i from Φ_m, $i = 1, \cdots, N$ and proceed to Step $m + 1$.

Let $\tau = \{t_m\}_{m=1}^{\infty}$, $0 < t_m \leq 1$ be a given sequence. The Vector Weak Relaxed Greedy Algorithm (VWRGA) [22] is defined as follows.

VWRGA(X, $\mathcal{D}$):
- **Step 0:** Define $f_0^{i,v,\tau,c} := 0, i = 1, \ldots N$.
- **Step m:**
- If $f^i = f_{m-1}^{i,v,\tau,r}$, stop the algorithm and define $f_k^{i,v,\tau,r} = f_{m-1}^{i,v,\tau,r} = f$ for $k \geq m$.
- If $f^i \neq f_{m-1}^{i,v,\tau,r}$, let i_m be such that

$$\|f^{i_m} - f_{m-1}^{i_m,v,\tau,r}\| = \max_{1 \leq i \leq N} \|f^i - f_{m-1}^{i,v,\tau,r}\|.$$

Choose an element $\varphi_m \in \mathcal{D}$ such that

$$F_{f^{im} - f^{im,v,\tau,r}_{m-1}}(\varphi_m) \geq t_m \sup_{\varphi \in \mathcal{D}} F_{f^{im} - f^{im,v,\tau,r}_{m-1}}(\varphi).$$

Find $0 \leq \lambda^i_m \leq 1$ such that

$$\|f^i - ((1 - \lambda^i_m)f^{im,v,\tau,r}_{m-1} + \lambda^i_m \varphi_m)\| = \inf_{0 \leq \lambda \geq 1} \|f^i - ((1 - \lambda)f^{im,v,\tau,r}_{m-1} + \lambda \varphi_m)\|.$$

Define $f^{i,v,\tau,r}_m := (1 - \lambda^i_m)f^{i,v,\tau,r}_{m-1} + \lambda^i_m \varphi_m$, and proceed to Step $m + 1$.

The error bounds of the VWCGA and VWRGA on $A_1(\mathcal{D})$ have been established in [22].

Theorem 12. *Let X be a Banach space with a modulus of smoothness $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. Then, for a sequence $\tau := \{t_k\}^\infty_{k=1}$, $0 < t_k \leq 1$ and any $f^i \in A_1(\mathcal{D}) \subset X$, $i = 1, \ldots, N$, we have*

$$\|f^i - f^{i,v,\tau,c}_m\| \leq C_1(q, \gamma) \min\left\{1, \left(\frac{1}{N}\sum_{k=1}^m t_k^p\right)^{-\frac{1}{p}}\right\}, \quad p = \frac{q}{q - 1},$$

$$\|f^i - f^{i,v,\tau,r}_m\| \leq C_2(q, \gamma) \min\left\{1, \left(\frac{1}{N}\sum_{k=1}^m t_k^p\right)^{-\frac{1}{p}}\right\}, \quad p = \frac{q}{q - 1}.$$

Now, we start to define the VWRPGA $(X, \mathcal{D})$. To accomplish this, we recall the definition of the WRPGA from [17]. Let X be a Banach space with a modulus of smoothness $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. Let $\tau = \{t_m\}^\infty_{m=1}$, $0 < t_m \leq 1$ be a given sequence.

WRPGA$(X, \mathcal{D})$:
  • **Step 0:** Define $f_0 = 0$.
  • **Step m:**
  - If $f = f_{m-1}$, stop the algorithm and define $f_k = f_{m-1} = f$ for $k \geq m$.
  - If $f \neq f_{m-1}$, choose an element $\varphi_m \in \mathcal{D}$ such that

$$|F_{f - f_{m-1}}(\varphi_m)| \geq t_m \sup_{\varphi \in \mathcal{D}} F_{f - f_{m-1}}(\varphi).$$

With

$$\lambda_m = \text{sign}\{F_{f - f_{m-1}}(\varphi_m)\}\|f - f_{m-1}\|(2\gamma q)^{\frac{1}{1-q}}|F_{f - f_{m-1}}(\varphi_m)|^{\frac{1}{q-1}},$$

$$\widehat{f}_m := f_{m-1} + \lambda_m \varphi_m,$$

choose s_m such that
$$\|f - s_m\widehat{f}_m\| = \min_{s \in \mathbb{R}} \|f - s\widehat{f}_m\|.$$

Define the next approximant to be $f_m = s_m\widehat{f}_m$, and proceed to Step $m + 1$.

The sufficient conditions for the convergence of the WRPGA in terms of the weakness sequence and the modulus of smoothness can be found in [17]. Moreover, the following theorem gives the error bound of the WRPGA on $A_1(\mathcal{D})$.

Theorem 13 (see Theorem 6.1 in [17]). *Let X be a Banach space with modulus of smoothness $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. If $f \in A_1(\mathcal{D}) \subset X$, then the output $\{f_m\}_{m \geq 0}$ of the WRPGA$(X, \mathcal{D})$ satisfies the inequality*

$$\|f - f_m\| \leq C_3(q, \gamma)\left(1 + \sum_{k=1}^m t_k^p\right)^{-\frac{1}{p}}, \quad p = \frac{q}{q - 1},$$

where the constant $C_3(q, \gamma)$ depends only on q and γ.

Let X be a Banach space with modulus of smoothness $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. Let $\tau = \{t_m\}_{m=1}^{\infty}$, $0 < t_m \leq 1$ be a given sequence. We define the VWRPGA $(X, \mathcal{D})$ as follows.

VWRPGA$(X, \mathcal{D})$: Given $f^i \in X, i = 1, \ldots, N$.
- **Step 0:** Define $f_0^{i,v,\tau,R} := 0, i = 1, \ldots N$.
- **Step m:**
- If $f^i = f_{m-1}^{i,v,\tau,R}$, stop the algorithm and define $f_k^{i,v,\tau,R} = f_{m-1}^{i,v,\tau,R} = f$ for $k \geq m$.
- If $f^i \neq f_{m-1}^{i,v,\tau,R}$, let i_m be such that

$$\|f^{i_m} - f_{m-1}^{i_m,v,\tau,R}\| = \max_{1 \leq i \leq N} \|f^i - f_{m-1}^{i,v,\tau,R}\|.$$

Choose an element $\varphi_m \in \mathcal{D}$ such that

$$F_{f^{i_m} - f_{m-1}^{i_m,v,\tau,R}}(\varphi_m) \geq t_m \sup_{\varphi \in \mathcal{D}} F_{f^{i_m} - f_{m-1}^{i_m,v,\tau,R}}(\varphi).$$

With

$$\lambda_m^i = \mathrm{sign}\{F_{f^i - f_{m-1}^{i,v,\tau,R}}(\varphi_m)\} \|f^i - f_{m-1}^{i,v,\tau,R}\| (2\gamma q)^{\frac{1}{1-q}} |F_{f^i - f_{m-1}^{i,v,\tau,R}}(\varphi_m)|^{\frac{1}{q-1}},$$

$$\hat{f}_m^i := f_{m-1}^{i,v,\tau,R} + \lambda_m^i \varphi_m,$$

choose s_m such that

$$\|f^i - s_m^i \hat{f}_m^i\| = \min_{s \in \mathbb{R}} \|f^i - s\hat{f}_m^i\|.$$

Define the next approximant to be $f_m^{i,v,\tau,R} = s_m^i \hat{f}_m^i$, and proceed to Step $m + 1$.

In this section, we obtain the convergence properties and error bound of the VWRPGA.

Firstly, we establish the theorem on the convergence of the VWRPGA. It seems that this theorem is the first result on the convergence property of the vector greedy algorithms in the Banach space setting.

Theorem 14. *Let X be a Banach space with modulus of smoothness $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. Assume*

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

Then, for any $f^i \in X, i = 1, \ldots, N$ and any dictionary $\mathcal{D}$ the VWRPGA converges.

The idea of the proof of Theorem 14 is similar to that of Theorem 5. However, because of the complexity of Banach spaces, a series of arguments in the subsequence analysis must be modified, replaced, and generalized. Some useful results in the case of Hilbert spaces have been generalized to the case of Banach spaces, as shown in the following lemmas.

Lemma 3. *Let X be a Banach space with modulus of smoothness $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. For any two nonzero elements $f, g \in X$ and any $h > 0$, we have*

$$\|f - hg\| \leq \|f\| + 2\|f\|\gamma \cdot \left(\frac{h\|g\|}{\|f\|}\right)^q - hF_f(g).$$

Proof. The proof of this lemma follows from the proof of Lemma 6.1 in [29] and the fact that the modulus of smoothness of X satisfies $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. $\square$

Lemma 4. *Let X be a Banach space with modulus of smoothness $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. Let $f_{m-1}^{i,v,\tau,R}$ be the output of the VWRPGA at Step $m-1$ for $f^i, i = 1, \ldots, N$. If $f^i \neq f_{m-1}^{i,v,\tau,R}$, then we have*

$$F_{f^i - f_{m-1}^{i,v,\tau,R}} \left(f_{m-1}^{i,v,\tau,R} \right) = 0.$$

Proof. Denote $L := \mathrm{span}\{\hat{f}_m^i\} \subset X$. By the definition of the VWRPGA$(X, \mathcal{D})$, $f_{m-1}^{i,v,\tau,R}$ is the best approximant to f^i from L for $i = 1, \ldots, N$. Thus, the conclusion of the lemma follows from Lemma 2.1 in [31]. $\square$

Lemma 5 (see Lemma 2.2 in [31]). *For any bounded linear functional F and any dictionary $\mathcal{D}$ from a Banach space, we have*

$$\sup_{\varphi \in \mathcal{D}} F(\varphi) = \sup_{g \in A_1(\mathcal{D})} F(g).$$

Now, we prove Theorem 14.

Proof of Theorem 14. Let r_m^i, $i = 1, \cdots, N$ be the residual of $f_m^{i,v,\tau,R}$. It is known from the definition of the VWRPGA(X, D) that r_m^i satisfies

$$
\begin{aligned}
\|r_m^i\| &= \|f^i - f_m^{i,v,\tau,R}\| \\
&= \|f^i - s_m^i \hat{f}_m^i\| \\
&\leq \|f^i - \hat{f}_m^i\| \\
&= \|r_{m-1}^i - \lambda_m^i \varphi_m\|.
\end{aligned}
$$

We apply Lemma 3 to the latter inequality with $f = r_{m-1}^i$, $g = \mathrm{sign}(\lambda_m^i)\varphi_m$, $h = |\lambda_m^i|$, and obtain

$$\|r_m^i\| \leq \|r_{m-1}^i\| + 2\|r_{m-1}^i\|\gamma \cdot \left(\frac{|\lambda_m^i|}{\|r_{m-1}^i\|} \right)^q - \lambda_m^i F_{f^i - f_{m-1}^{i,v,\tau,R}}(\varphi_m).$$

By the choice of λ_m^i, we have

$$\|r_m^i\| \leq \|r_{m-1}^i\|\left\{ 1 - \frac{q-1}{q}(2\gamma q)^{\frac{1}{1-q}} \cdot |F_{f^i - f_{m-1}^{i,v,\tau,R}}(\varphi_m)|^{\frac{q}{q-1}} \right\}. \tag{11}$$

Thus, it is easy to see that $\{\|r_m^i\|\}_{m=0}^{\infty}$ is a decreasing sequence. According to the Monotone Convergence Theorem, we know that $\lim_{m\to\infty} \|r_m^i\|$ for $i = 1, \ldots, N$ exists.

Next, we prove that $\lim_{m\to\infty} \|r_m^i\| = 0$ by contradiction. Assume $\lim_{m\to\infty} \|r_m^i\| \geq a > 0$, $i = 1, \ldots, N$. Then, for any m, we have $\|r_m^i\| \geq a$. By (11), we obtain that

$$\sum_{i=1}^{N} \|r_m^i\| \leq \sum_{i=1}^{N} \|r_{m-1}^i\| - \frac{q-1}{q}(2\gamma q)^{\frac{1}{1-q}} \cdot \sum_{i=1}^{N} \|r_{m-1}^i\| \cdot |F_{f^i - f_{m-1}^{i,v,\tau,R}}(\varphi_m)|^{\frac{q}{q-1}}$$

$$\leq \sum_{i=1}^{N} \|r_{m-1}^i\| \left\{ 1 - \frac{q-1}{q}(2\gamma q)^{\frac{1}{1-q}} \cdot \frac{\|r_{m-1}^{i_m}\| \cdot (F_{f^{i_m} - f_{m-1}^{i_m,v,R,\tau}}(\varphi_m))^{\frac{q}{q-1}}}{\sum_{i=1}^{N} \|r_{m-1}^i\|} \right\}$$

$$\leq \sum_{i=1}^{N} \|r_{m-1}^i\| \left\{ 1 - \frac{q-1}{q}(2\gamma q)^{\frac{1}{1-q}} \cdot \frac{\|r_{m-1}^{i_m}\| \cdot (F_{f^{i_m} - f_{m-1}^{i_m,v,R,\tau}}(\varphi_m))^{\frac{q}{q-1}}}{N\|r_{m-1}^{i_m}\|} \right\}$$

$$\leq \sum_{i=1}^{N} \|f^i\| \prod_{j=1}^{m} \left\{ 1 - \frac{1}{N} \cdot \frac{q-1}{q}(2\gamma q)^{\frac{1}{1-q}} \cdot (F_{f^{i_j} - f_{j-1}^{i_j,v,R,\tau}}(\varphi_j))^{\frac{q}{q-1}} \right\}.$$

Denote

$$x_j = \frac{1}{N} \cdot \frac{q-1}{q}(2\gamma q)^{\frac{1}{1-q}} \cdot (F_{f^{i_j} - f_{j-1}^{i_j,v,R,\tau}}(\varphi_j))^{\frac{q}{q-1}}). \tag{12}$$

By the inequality $1 - x \leq \frac{1}{1+x}$, $0 \leq x \leq 1$, we obtain

$$\sum_{i=1}^{N} \|r_m^i\| \leq \sum_{i=1}^{N} \|f^i\| \prod_{j=1}^{m} \frac{1}{1+x_j}$$

$$\leq \sum_{i=1}^{N} \|f^i\| \frac{1}{1 + \sum_{j=1}^{m} x_j}. \tag{13}$$

Then, we proceed with a lower estimate for $x_j, j = 1, \ldots, m$.

By Lemma 1, we set $\epsilon = \frac{a}{2}$ and find f_j^ϵ such that

$$\|f^{i_j} - f_j^\epsilon\| < \epsilon$$

and

$$\frac{f_j^\epsilon}{A(\epsilon)} \in A_1(\mathcal{D}),$$

with some number $A(\epsilon) > 0$.

We obtain from Lemma 5 that

$$F_{f^{i_j} - f_{j-1}^{i_j,v,R,\tau}}(\varphi_j) \geq t_j \sup_{\varphi \in \mathcal{D}} F_{f^{i_j} - f_{j-1}^{i_j,v,R,\tau}}(\varphi)$$

$$= t_j \sup_{g \in A_1(\mathcal{D})} F_{f^{i_j} - f_{j-1}^{i_j,v,R,\tau}}(g)$$

$$\geq t_j A(\epsilon)^{-1} F_{f^{i_j} - f_{j-1}^{i_j,v,R,\tau}}(f_j^\epsilon). \tag{14}$$

By Lemma 4, we obtain

$$
\begin{aligned}
F_{f^{i_j} - f^{i_j,v,R,\tau}_{j-1}}(f^\epsilon_j) &= F_{f^{i_j} - f^{i_j,v,R,\tau}_{j-1}}(f^{i_j} - (f^{i_j} - f^\epsilon_j)) \\
&= F_{f^{i_j} - f^{i_j,v,R,\tau}_{j-1}}(f^{i_j}) - F_{f^{i_j} - f^{i_j,v,R,\tau}_{j-1}}(f^{i_j} - f^\epsilon_j) \\
&> F_{f^{i_j} - f^{i_j,v,R,\tau}_{j-1}}(f^{i_j} - f^{i_j,v,R,\tau}_{j-1} + f^{i_j,v,R,\tau}_{j-1}) - \epsilon \\
&= F_{f^{i_j} - f^{i_j,v,R,\tau}_{j-1}}(f^{i_j} - f^{i_j,v,R,\tau}_{j-1}) - \epsilon \\
&= \|r^{i_j}_{j-1}\| - \epsilon.
\end{aligned}
\tag{15}
$$

Inequalities (14), (15) and $\epsilon = \frac{a}{2}$ result in

$$
\begin{aligned}
F_{f^{i_j} - f^{i_j,v,R,\tau}_{j-1}}(\varphi_j) &\geq t_j A(\epsilon)^{-1}(\|r^{i_j}_{j-1}\| - \epsilon) \\
&\geq t_j A(\epsilon)^{-1} \cdot \frac{a}{2}.
\end{aligned}
\tag{16}
$$

Combining (12) with (16), we obtain

$$
\begin{aligned}
x_j &\geq \frac{1}{N} \cdot \frac{q-1}{q} (2\gamma q)^{\frac{1}{1-q}} \cdot \left(t_j A(\epsilon)^{-1} \cdot \frac{a}{2} \right)^{\frac{q}{q-1}} \\
&= \frac{1}{N} \cdot \frac{q-1}{q} (2\gamma q)^{\frac{1}{1-q}} \cdot \left(\frac{a}{2A(\epsilon)} \right)^p \cdot t^p_j.
\end{aligned}
\tag{17}
$$

Combining (17) with (13), we have

$$
\sum_{i=1}^N \|r^i_m\| \leq \sum_{i=1}^N \|f^i\| \cdot \frac{1}{1 + \frac{1}{N} \cdot \frac{q-1}{q} (2\gamma q)^{\frac{1}{1-q}} \cdot \left(\frac{a}{2A(\epsilon)} \right)^p \sum_{j=1}^m t^p_j}.
$$

The assumption $\sum_{m=1}^\infty t^p_m = \infty$ implies that $\sum_{i=1}^N \|r^i_m\| \to 0$ as $m \to \infty$.

Thus, $\lim_{m \to \infty} \|r^i_m\| = 0$ for $1 = 1, \ldots, N$. We obtain a contradiction which proves this theorem. $\square$

Remark 5. *According to Theorem 3.1 in [32], we know that $\sum_{m=1}^\infty t^p_m = \infty$ is also a necessary condition for the convergence of the VWRPGA.*

Remark 6. *Since the WRGA only converges for the target elements from $\mathcal{A}_1(\mathcal{D})$, see [31], then the VWRGA also only converges for the target elements from $\mathcal{A}_1(\mathcal{D})$. Thus, the convergence property of the VWRPGA is better than that of the VWRGA.*

Remark 7. *For $f^i \in A_1(\mathcal{D})$, $i = 1, \ldots, N$, the sufficient condition for the convergence of the VWCGA follows from Theorem 12. Theorem 15 gives the convergence condition for any $f^i \in X, i = 1, \ldots, N$.*

By using the same method, it is not difficult to prove the following theorem on the convergence of the VWCGA.

Theorem 15. *Let X be a Banach space with modulus of smoothness $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. Assume*

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

Then, for any $f^i \in X, i = 1, \ldots, N$ and any dictionary $\mathcal{D}$ the VWCGA converges.

Next, we give the theorem about the error bound of the VWRPGA$(X, \mathcal{D})$ on $A_1(\mathcal{D})$.

Theorem 16. *Let X be a Banach space with modulus of smoothness $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. Then, for a sequence $\tau := \{t_k\}_{k=1}^{\infty}$, $0 < t_k \leq 1$ and any $f^i \in A_1(\mathcal{D}) \subset X$, $i = 1, \ldots, N$, we have*

$$\|f^i - f_m^{i,v,\tau,R}\| \leq C_3(q, \gamma) \min\left\{1, \left(\frac{1}{N}\sum_{k=1}^{m} t_k^p\right)^{-\frac{1}{p}}\right\}, \quad p = \frac{q}{q-1}.$$

Proof. It is known from (11) that the sequences $\{\|f^i - f_m^{i,v,\tau,R}\|\}_{m=0}^{\infty}$, $i = 1, \cdots, N$ are decreasing. Fix i. The inequality

$$\|f^i - f_m^{i,v,\tau,R}\| \leq 1$$

follows from the assumption $f^i \in A_1(\mathcal{D})$ and the fact that $\{\|f^i - f_m^{i,v,\tau,R}\|\}_{m=0}^{\infty}$ is decreasing. Thus, we only need to prove the following estimate:

$$\|f^i - f_m^{i,v,\tau,R}\| \leq C(q, \gamma) \left(\frac{1}{N}\sum_{k=1}^{m} t_k^p\right)^{-\frac{1}{p}}, \quad i = 1, \ldots, N.$$

We define the set $E_l := \{k | i_k = l, 1 \leq k \leq m\}$ just as we did in proof of Theorem 6. It is obvious that

$$\sum_{k=1}^{m} t_k^p = \sum_{l=1}^{N} \sum_{k \in E_l} t_k^p.$$

Thus, there exists $l_0 \in [1, N]$ such that

$$\sum_{k \in E_{l_0}} t_k^p \geq \frac{1}{N}\sum_{k=1}^{m} t_k^p.$$

As in proof of Theorem 6, for $f^{l_0} \in X$, the sequences $f_1^{l_0,v,R,\tau}, \ldots, f_m^{l_0,v,R,\tau}$ are the outputs of the WRPGA with the weakness sequence $\tau^{l_0} := \{t_k^{l_0}\}$. Therefore, using Theorem 13, we obtain

$$\|f^{l_0} - f_{k_0-1}^{l_0,v,R,\tau}\| \leq C(q, \gamma)\left(1 + \sum_{k \in E_{l_0} - \{k_0\}} t_k^p\right)^{-\frac{1}{p}}$$

$$\leq C(q, \gamma)\left(1 + \sum_{k \in E_{l_0}} t_k^p - 1\right)^{-\frac{1}{p}}$$

$$\leq C(q, \gamma)\left(\frac{1}{N}\sum_{k=1}^{m} t_k^p\right)^{-\frac{1}{p}}.$$

The proof of Theorem 16 is completed. $\square$

Remark 8. *We know from Theorems 12 and 16 that the error bound of the VWRPGA is almost the same as those of the VWCGA and VWRGA. Similarly, the computational complexity of the VWRPGA is essentially smaller than those of the VWCGA and VWRGA.*

4. Conclusions

In this paper, we consider the use of vector greedy algorithms for simultaneous approximation. We first work in a Hilbert space H. We propose a new vector greedy algorithm—the Vector Weak Rescaled Pure Greedy Algorithm (VWRPGA)—for simultaneous approximation with respect to a dictionary $\mathcal{D}$ in H. Then, we study the error performances of the VWRPGA. We show that the convergence rate of the VWRPGA on $\mathcal{A}_1(\mathcal{D})$ is optimal. The VWRPGA has a weaker convergence condition than the VWPGA. The convergence rate of the VWRPGA is better than that of the VWPGA. In particular, this advantage is more obvious when N is large. Moreover, the error performances of the VWRPGA are similar to those of the VWOGA. However, from the viewpoint of computational complexity, the VWRPGA is simpler than the VWOGA. For N target elements, from the definitions of the algorithms, one can see that the VWOGA needs to solve N m-dimensional optimization problems. However, the VWRPGA only needs to solve N one-dimensional optimization problems.

Then, we design the Vector Weak Rescaled Pure Greedy Algorithm (VWRPGA) in a uniformly smooth Banach space setting. We obtain the convergence properties and error bound of the VWRPGA in this case. We also show that the convergence condition of the VWCGA is the same as that of the VWRPGA. We show that when the Banach space is a Lebesgue space, the convergence rate of the VWRPGA on $\mathcal{A}_1(\mathcal{D})$ is sharp. As for the convergence properties, the VWRGA converges only for the target elements from $\mathcal{A}_1(\mathcal{D})$, while the VWRPGA converges for any element. Therefore, the VWRPGA has better convergence properties than the VWRGA. The error bounds of the VWRPGA are similar to those of the VWCGA and VWRGA. From the viewpoint of computational complexity, the VWRPGA is simpler than the VWCGA and the VWRGA.

In conclusion, the VWRPGA is the simplest vector greedy algorithm for simultaneous approximation with the best convergence property and the optimal convergence rate.

The VWRGA is more efficient than the WRPGA, since the complexity of its calculation and the storage of information can be reduced greatly by the VWRPGA instead of the N-fold WRPGA. If $\tau := \{t_k\}_{k=1}^{\infty}, t_k = 1, k = 1, 2, \cdots$ and $N = 1$, then the VWRPGA degenerates into the RPGA. In [5], the authors applied the RPGA to a kernel-based regression. They defined the Rescaled Pure Greedy Learning Algorithm (RPGLA) and studied its efficiency. They showed that the computational complexity of the RPGLA is less than the Orthogonal Greedy Learning Algorithm (OGLA) [37] and Relaxed Greedy Learning Algorithm (RGLA) [38]. When the kernel is infinitely smooth, the learning rate can be arbitrarily close to the best rate $O(m^{-1})$ under a mild assumption of the regression function. Since the VWRPGA is more efficient than the RPGA, the VWRPGA can be used to solve the problems of multi-task learning more efficiently. Moreover, it is natural to consider the applications of the VWRPGA to vector signal processing. We will study these applications of the VWRPGA in the future.

References

1. Barron, A.; Cohen, A.; Dahmen, W.; DeVore, R. Approximation and learning by greedy algorithms. *Ann. Stat.* **2008**, *36*, 64–94. [CrossRef]
2. Cohen, A.; Dahmen, W.; DeVore, R. Compressed sensing and best *k*-term approximation. *J. Am. Math. Soc.* **2009**, *22*, 211–231. [CrossRef]
3. Yang, B.; Yang, C.; Huang, G. Efficient image fusion with approximate sparse representation. *Int. J. Wavelets Multiresolut. Inf. Process.* **2016**, *14*, 1650024.
4. Zhang, W.H.; Ye, P.X.; Xing, S.; Xu, X. Optimality of the approximation and learning by the rescaled pure super greedy algorithms. *Axioms* **2022**, *11*, 437. [CrossRef]
5. Zhang, W.H.; Ye, P.X.; Xing, S. Optimality of the rescaled pure greedy learning algorithms. *Int. J. Wavelets Multiresolut. Inf. Process.* **2023**, *21*, 2250048. [CrossRef]
6. Nguyen, H.; Petrova, G. Greedy strategies for convex optimization. *Calcolo* **2017**, *54*, 207–224. [CrossRef]
7. Huang, A.T.; Feng, R.Z.; Wang, A.D. The sufficient conditions for orthogonal matching pursuit to exactly reconstruct sparse polynomials. *Mathematics* **2022**, *10*, 3703. [CrossRef]
8. Liu, Z.Y.; Xu, Q.Y. A multiscale RBF collocation method for the numerical solution of partial differential equations. *Mathematics* **2019**, *7*, 964. [CrossRef]
9. Jin, D.F.; Yang, G.; Li, Z.H.; Liu, H.D. Sparse recovery algorithm for compressed sensing using smoothed l_0 norm and randomized coordinate descent. *Mathematics* **2019**, *7*, 834. [CrossRef]
10. Natsiou, A.A.; Gravvanis, G.A.; Filelis-Papadopoulos, C.K.; Giannoutakis, K.M. An aggregation-based algebraic multigrid method with deflation techniques and modified generic factored approximate sparse inverses. *Mathematics* **2023**, *11*, 640. [CrossRef]
11. Argyriou, A.; Evgeniou, T.; Pontil, M. Convex multitask feature learning. *Mach. Learn.* **2008**, *73*, 243–272. [CrossRef]
12. Schmidt, E. Zur Theorie der linearen und nichtlinearen Integralgleichungen. *I Math. Annalen.* **1906–1907**, *63*, 433–476. [CrossRef]
13. Tropp, J.A.; Gilbert, A.C.; Strauss, M.J. Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit. *Signal. Process.* **2006**, *86*, 572–588. [CrossRef]
14. Wirtz, D.; Haasdonk, B. A vectorial kernel orthogonal greedy algorithm. *Proc. DWCAA* **2013**, *6*, 83–100.
15. DeVore, R.A.; Temlyakov, V.N. Some remarks on greedy algorithms. *Adv. Comput. Math.* **1996**, *5*, 173–187. [CrossRef]
16. Gao, Z.; Petrova, G. Rescaled pure greedy algorithm for convex optimization. *Calcolo* **2019**, *56*, 15. [CrossRef]
17. Petrova, G. Rescaled pure greedy algorithm for Hilbert and Banach spaces. *Appl. Comput. Harmon. Anal.* **2016**, *41*, 852–866. [CrossRef]
18. Jiang, B.; Ye, P.; Zhang, W. Unified error estimate for weak biorthogonal greedy algorithms. *Int. J. Wavelets Multiresolut. Inform. Process.* **2022**, *5*, 2150001. [CrossRef]
19. Dereventsov, A.V.; Temlyakov, V.N. A unified way of analyzing some greedy algorithms. *J. Funct. Anal.* **2019**, *12*, 1–30. [CrossRef]
20. Temlyakov, V.N. A remark on simultaneous greedy approximation. *East J. Approx.* **2004**, *10*, 17–25.
21. Leviatan, D.; Temlyakov, V.N. Simultaneous approximation by greedy algorithms. *Adv. Comput. Math.* **2006**, *25*, 73–90. [CrossRef]
22. Leviatan, D.; Temlyakov, V.N. Simultaneous greedy approximation in Banach spaces. *J. Complex.* **2005**, *21*, 275–293. [CrossRef]
23. Lutoborski, A.; Temlyakov, V.N. Vector greedy algorithms. *J. Complex.* **2003**, *19*, 458–473. [CrossRef]
24. Mallat, S.; Zhang, Z. Matching pursuit with time-frequency dictionaries. *IEEE Trans. Signal Pross.* **1993**, *41*, 3397–3415. [CrossRef]
25. Konyagin, S.V.; Temlyakov, V.N. Rate of convergence of pure greedy algorithm. *East. J. Approx.* **1996**, *5*, 493–499.
26. Sil′nichenko, A.V. Rates of convergence of greedy algorithms. *Mat. Zametki.* **2004**, *76*, 628–632.
27. Burusheva, L.; Temlyakov, V. Sparse approximation of individual functions. *J. Approx. Theory* **2020**, *259*, 105471. [CrossRef]
28. Livshitz, D.; Temlyakov, V.N. Two lower estimates in greedy approximation. *Constr. Approx.* **2003**, *19*, 509–524. [CrossRef]
29. Temlyakov, V.N. *Greedy Approximation*; Cambridge University Press: Cambridge, UK, 2011.
30. Temlyakov, V.N. Weak greedy algorithms. *Adv. Comput. Math.* **2000**, *12*, 213–227. [CrossRef]
31. Temlyakov, V.N. Greedy algorithms in Banach spaces. *Adv. Comput. Math.* **2001**, *14*, 277–292. [CrossRef]
32. Jiang, B.; Ye, P.X. Efficiency of the weak rescaled pure greedy algorithm. *Int. J. Wavelets Multiresolut. Inform. Process.* **2021** *4*, 2150001. [CrossRef]
33. Donahue, M.; Gurvits, L.; Darken, C.; Sontag, E. Rate of convex approximation in non-Hilbert spaces. *Constr. Approx.* **1997**, *13*, 187–220. [CrossRef]
34. Lindenstrauss, J.; Tzafriri, L. *Classical Banach Spaces I*; Springer: Berlin/Heidelberg, Germany, 1977.
35. Temlyakov, V.N.; Yang, M.R.; Ye, P.X. Greedy approximation with regard to non-greedy bases. *Adv. Comput. Math.* **2011**, *34*, 319–337. [CrossRef]
36. Ye, P.X.; Wei, X.J. Efficiency of weak greedy algorithms for *m*-term approximations. *Sci. China Math.* **2016**, *59*, 697–714. [CrossRef]

37. Chen, H.; Zhou, Y.C.; Tang, Y.Y.; Li, L.Q.; Pan, Z.B. Convergence rate of the semi-supervised greedy algorithm. *Neural Netw.* **2013**, *44*, 44–50. [CrossRef]
38. Lin, S.B.; Rong, Y.H.; Sun, X.P.; Xu, Z.B. Learning capability of the relaxed greedy algorithms. *IEEE Trans. Neural Netw. Learn. Syst.* **2013**, *24*, 1598–1608.