Convex Optimization with Greedy Methods

Zheming Gao

Texas A&M University
Beihang University

gorgeous1992@tamu.edu

May 12, 2015
Overview

1. Introduction to Greedy Algorithms


3. Expected Results
Definition (Hilbert Space)

\( \mathcal{H} \) is a **Hilbert space** if it is a complete inner product space with its norm induced by the inner product:

\[
\| \cdot \| = (\cdot , \cdot )^{\frac{1}{2}}
\]

Definition (Dictionary)

A set of functions \( \mathcal{D} = \{ \varphi \} \) is called a **dictionary** if

\[
\| \varphi \| = 1 \quad \text{and} \quad \overline{\text{span}\{ \mathcal{D} \}} = \mathcal{H}
\]
**Used For:** finding a good approximant of a given function \( f \in \mathcal{H} \) via finite linear combinations of elements from the dictionary \( \mathcal{D} \) and at step \( m \) the algorithm generates the approximant (\textit{m-sparse approximation})

\[
f_m = \sum_{i=1}^{m} c_j \varphi_{ij} \quad (c_j \in \mathbb{R}, \varphi_{ij} \in \mathcal{D})
\]

We want:

\[
\lim_{m \to \infty} \|f_m - f\| = 0
\]

Investigate:

\[
\|f_m - f\| \sim m^{-\alpha}
\]
Definition (Approximation Error)

We define the **m-term Approximation Error** as

\[ \sigma_m(f, \mathcal{D}) := \inf_{s \in \sum_m(\mathcal{D})} \|f - s\| \]

\(\sum_m(\mathcal{D})\) is a collection of the linear combinations of \(m\) elements from \(\mathcal{D}\):

\[ \sum_m(\mathcal{D}) := \left\{ \sum_{j=1}^{m} c_j \varphi_{ij} \mid c_j \in \mathbb{R}, \varphi_{ij} \in \mathcal{D} \right\} \]
Definition

For a general dictionary $\mathcal{D}$, define the class of functions:

1

$A_0^0(\mathcal{D}, M) := \{ f \in \mathcal{H} | f = \sum_{\varphi_k \in \mathcal{D}, k \in \Lambda} c_k \varphi_k, |\Lambda| < \infty \text{ and } \sum_{k \in \Lambda} |c_k| \leq M \}$

and define $A_1(\mathcal{D}, M)$ as the closure (in $\mathcal{H}$) of $A_0^0(\mathcal{D}, M)$.

2

$A_1(\mathcal{D}) = \bigcup_{M > 0} A_1(\mathcal{D}, M)$

The semi-norm of $f$ in $A_1(\mathcal{D})$:

$|f|_{A_1(\mathcal{D})} := \inf_{f \in A_1(\mathcal{D}, M)} \{ M \}$
**Theorem** (DeVore-Temlyakov, 1996)

If $f \in A_1(D) \subset \mathcal{H}$, then

$$\sigma_m(f, D) \leq C |f|_{A_1(D)} m^{-\frac{1}{2}}$$

- Goal: If $f_m$ is the approximant generated by greedy algorithm, we want to get the rate of error $\|f_m - f\| \sim m^{-1/2}$. 
Types of Algorithms I

There are 3 main types of greedy algorithms:

- Pure Greedy Algorithm (PGA)
- Relaxed Greedy Algorithm (RGA)
- Orthogonal Greedy Algorithm (OGA)
Pure Greedy Algorithm (PGA):
Step $m = 0$: Define $f_0 := 0$
Step $m$: choose $\varphi_m \in \mathcal{D}$, such that

$$ |(\varphi_m , f - f_{m-1})| = \sup_{\varphi \in \mathcal{D}} |(\varphi , f - f_{m-1})| $$

Define the next approximant:

$$ f_m = f_{m-1} + (f - f_{m-1} , \varphi_m)\varphi_m $$

Theorem (DeVore-Temlyakov, 1996)
If $f \in A_1(\mathcal{D})$, $f_m$ is generated by PGA, then

$$ \|f_m - f\| \leq |f|_{A_1(\mathcal{D})} m^{-\frac{1}{6}} $$

Figure: RGA
Relaxed Greedy Algorithm (RGA):

Step $m = 0$: Define $f_0 := 0$

Step $m$: choose $\varphi_m \in \mathcal{D}$, such that

$$\| (\varphi_m, f - f_{m-1}) \| = \sup_{\varphi \in \mathcal{D}} | (\varphi, f - f_{m-1}) |$$

Define the next approximant:

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

**Theorem (DeVore-Temlyakov, 1996)**

If $f \in A_1(\mathcal{D})$, $f_m$ is generated by RGA, then

$$\| f_m - f \| \leq \| f \|_{A_1(\mathcal{D})} m^{-\frac{1}{2}}$$
Orthogonal Greedy Algorithm (OGA):

Step $m = 0$: Define $f_0 = 0$

Step $m$: Define

$$| (\varphi_m, f - f_{m-1}) | = \sup_{\varphi \in D} | (\varphi, f - f_{m-1}) |$$

and $H_m := \text{span}\{\varphi_1, \ldots, \varphi_m\}$

Define the next approximant:

$$f_m = P_{H_m}(f)$$

which is the orthogonal projection on $H_m$.

**Theorem (DeVore-Temlyakov, 1996)**

If $f \in A_1(D)$, $f_m$ is generated by OGA, then

$$\| f_m - f \| \leq | f |_{A_1(D)} m^{-\frac{1}{2}}$$
Figure: OGA
Interested in applying greedy techniques for solving the problem:

Given a convex function $E : \mathcal{H} \to \mathbb{R}$, try to find the global minimum of $E$:

$$E(x^*) = \inf_{x \in \mathcal{H}} E(x)$$

**Example**

To find the minimum of $f(x) = ax^2 + bx + c \ (a \geq 0)$

$$x^* = -\frac{b}{2a}$$
**Definition (Convex Function)**

$E : \mathcal{H} \rightarrow \mathbb{R}$ is a convex function if

$\forall u, v \in \mathcal{H}, t \in [0, 1]$

$E(tu + (1 - t)v) \leq tE(u) + (1 - t)E(v)$

**Figure:** Convex Function
**Definition (Frechet Derivative)**

We call that $E$ is Frechet differentiable at $x \in \mathcal{H}$ if there exists a bounded linear functional, denote by $E'(x)$, such that

$$\lim_{h \to 0} \frac{|E(x + h) - E(x) - \langle E'(x), h \rangle|}{\|h\|} = 0$$

**Definition (Solution Space)**

Given a Hilbert Space $\mathcal{H}$ and a convex function $E$, define

$$\Omega := \{x \in \mathcal{H} | E(x) \leq E(0)\}$$

i.e.

$$\inf_{x \in \mathcal{H}} E(x) = \inf_{x \in \Omega} E(x)$$
Example

Given a dictionary $\mathcal{D}$ from $\mathcal{H}$, use greedy algorithms to get a sequence $\{x_m\} \in \Omega$, such that

$$x_m \in \sum_m (\mathcal{D}) \text{ and } x^* = \lim_{m \to \infty} x_m$$

**Chebyshev Greedy Algorithm (CGA(co))**

Step $m = 0$: Set $x_0 := 0$.

Step $m$: choose $\varphi_m$ from $\mathcal{D}$, such that

$$\left| \langle E'(x_{m-1}) , \varphi_m \rangle \right| = \sup_{g \in \mathcal{D}} \left\{ \left| \langle E'(x_{m-1}) , g \rangle \right| \right\}$$

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

$$E(x_m) = \inf_{x \in \Phi_m} E(x)$$
Conditions on $E$

- **Condition 0:** $E$ has a Frechet derivative $E'(x) \in H$ at each point $x$ in $\Omega$, which is bounded, and

  \[ \|E'(x)\| \leq M_0, \quad x \in \Omega \]

- **Uniform Smoothness (US):** There are $0 < \alpha$, $1 < q \leq 2$ and $0 < M$, such that for all $x, x'$ with $\|x - x'\| \leq M$, $x \in \Omega$,

  \[ E(x') - E(x) - E'(x' - x) \leq \alpha \|x' - x\|^q \tag{1} \]

- **Uniform Convexity (UC):** There are $0 < \beta$, $2 \leq p < \infty$ and $0 < M$, such that for all $x, x'$ with $\|x - x'\| \leq M$, $x \in \Omega$,

  \[ E(x') - E(x) - E'(x' - x) \geq \beta \|x' - x\|^p \tag{2} \]

  Particularly, we say $E$ is **strongly convex** when $p = 2$. 
Theorem (Temlyakov)

Let $E$ be a uniformly smooth convex function with modulus of smoothness $\rho(E, u) \leq \gamma u^q$, where $1 < q \leq 2$. Then we have for CGA(co):

$$E(x_m) - \inf_{x \in \Omega} E(x) \leq C m^{1-q}$$
Theorem (H. Nguyen & G. Petrova)

Let \( E \) be a convex function that satisfies **Condition 0**, the **US**, and the **UC** conditions. Let the minimizer \( x^* = \sum_i c_i(x^*) \varphi_i \in \Omega \) with support \( |S| := |\{ i : c_i(x^*) \neq 0 \}| < \infty \), where \( \{ \varphi_i \} \) is an orthonormal basis for \( \mathcal{H} \).

Then, at step \( m \),

1. If \( p \neq q \),

\[
E(x_m) - E(x^*) \leq Cm^{-\frac{p(q-1)}{p-q}}
\]

where \( C = C(|S|, p, q, \alpha, \beta, E) \).

2. If \( p = q = 2 \),

\[
E(x_m) - E(x^*) \leq \tilde{C} \gamma^{m-1}
\]

where \( \tilde{C} = E(0) - E(x^*), \gamma = \gamma(\alpha, \beta, E) \).
Expected Results

Error Estimate

1. Greedy methods
2. Different Conditions
3. Better convergence under various conditions on $E$

Improve the results of the algorithms given in Dr. Temlyakov’s paper.
Thank you