Derivative Free Optimization

Lecture 2 : Exhaustive & Grid Search

Warren Hare

DFO framework

DFO Framework

We seek to solve

$$(\arg)\min\{f(x) : x \in \Omega\}$$

where $f : \mathbb{R}^n \mapsto \mathbb{R}$

The DFO/BBO framework assumes, given a point $x \in \mathbb{R}^n$

- we can check if $x \in \Omega$ or $x \notin \Omega$
- we can acquire a function value $f(x)$
- but, we cannot 'look' at f

Exhaustive Search

Countable Dense subsets

Definition: A set $S \subseteq \mathbb{R}^n$ is **countable** if it can be ordered: i.e., $S = \{x^0, x^1, \ldots\}$.

Definition: A set $S \subseteq \mathbb{R}^n$ is a **dense subset** of Ω if

- $S \subseteq \Omega$, and
- given any $x \in \Omega$ and $\delta > 0$,

  there exists a point $\tilde{x} \in S$ such that $\|x - \tilde{x}\| < \delta$.

Countable Dense subsets

Example: Show that the set

$$S = \left\{ x : x = \frac{a}{b}, a \in \mathcal{Z}, b \in \mathcal{Z}, 0 < a < b \right\}$$

is countable and dense in $[0, 1] \subseteq \mathbb{R}^1$

Example: Create a countable dense subset of $[0, 1]^2 \subseteq \mathbb{R}^2$

Algorithm: Exhaustive Search (ES)

Given $f : \mathbb{R}^n \mapsto \mathbb{R}$, and $\Omega \subseteq \mathbb{R}^n$

0. Initialization

$S = \{x^0, x^1, x^2, \ldots\}$	a countable dense subset of Ω
$k \leftarrow 0$	iteration counter
$x_{\texttt{best}} = x^0 \in S$	best point found so far
$f_{\texttt{best}} = f(x_{\texttt{best}})$	best objective function value found so far

1. Update

   evaluate $f(x^{k+1})$

   if $f(x^{k+1}) < f_{\texttt{best}}$, then

   update $x_{\texttt{best}} \leftarrow x^{k+1}$, $f_{\texttt{best}} \leftarrow f(x_{\texttt{best}})$

   increment $k \leftarrow k + 1$ and go to 1

Convergence of ES

Theorem: Suppose $f \in \mathcal{C}^0$, $\Omega \subseteq \mathbb{R}^n$ and

$$X^* = \mathrm{argmin}_x \{f(x) : x \in \Omega\} \neq \emptyset.$$

Denote the value $f_{\texttt{best}}$ at iteration k of ES by $f^k_{\texttt{best}}$ and the point $x_{\texttt{best}}$ at iteration k by $x^k_{\texttt{best}}$. Then

$$\lim_{k\to\infty} f^k_{\texttt{best}} = f^* := \min_x \{f(x) \ : \ x \in \Omega\}$$

and moreover, if Ω is *compact*, then

$$\liminf_{k\to\infty} \mathrm{dist}(x^k_{\texttt{best}}, X^*) = 0.$$

Exhaustive Search

Pros:

- It works

Cons:

- Need to construct S
- Requires infinite time
- No way to know when you can stop

Conclusion

- Debatable if it is a DFO method
- Not a good approach for BBO

Grid Search

Grid search

Example: Let $\Omega = [0,1]^2$

Suppose S is an $p \times p$ *grid* of equally spaced points in Ω

1. How many points are in S?
2. What is the farthest any point in Ω is from S?

Algorithm: Grid Search (GS)

Given $f : \mathbb{R}^n \mapsto \mathbb{R}$, and $[\ell, u] \subset \mathbb{R}^n$

0. Initialization

   $p_i \in \mathbb{N}$ with $p_i \geq 2$ number of grid points for each variable

1. Sample f at all grid points

   define $\delta_i = (u_i - \ell_i)/(p_i - 1)$ for each $i \in \{1, 2, \ldots, n\}$

   create $G = \left\{ \begin{array}{lcl} x \in [\ell, u] & : & x_i = \ell_i + z\delta_i, \\ & & i = 1, 2, \ldots, n, z = 0, 1, \ldots, p - 1 \end{array} \right\}$

   choose $x_{\text{best}} \in \operatorname{argmin}\{f(t) : t \in G\}$

Convergence of GS

Suppose $f \in \mathcal{C}^{0+}$ with constant K
Suppose GS is used to minimize f over $x \in [\ell, u] \subseteq \mathbb{R}^n$
The optimal value $f^* = \min_x \{f(x) \ : \ x \in [\ell, u]\}$ satisfies

$$f_{\text{best}} - K\sqrt{n}\frac{\delta}{2} \leq f^* \leq f_{\text{best}}$$

where $f_{\text{best}} = f(x_{\text{best}})$ and $\delta = \max \delta_i$

GS example

Example: Let $\Omega = [0,5]^4$. Suppose $f \in \mathcal{C}^{0+}$ with constant $K = 0.25$.

1. What value of p is required to ensure $f_{\text{best}} - f^* \leq 10^{-3}$?
2. How many function evaluations are required to ensure $f_{\text{best}} - f^* \leq 10^{-3}$?

GS example

Example: Suppose $\Omega = [-1, 1]^n$. Suppose $f \in \mathcal{C}^{0+}$ with constant $K = 3$. As a function of n,

1. What value of p is required to ensure $f_{\text{best}} - f^* \leq 10^{-3}$?
2. How many function evaluations are required to ensure $f_{\text{best}} - f^* \leq 10^{-3}$?
3. If you can do 1000 function evaluation per second, what is the largest value of n that you can solve before this course is over ($\approx 8 \times 10^6$ seconds)?

Grid Search

Pros:

- Finite time
- Quality of result is known

Cons:

- The value of K is seldom known
- In $\mathbb{R}^n$ the number of points evaluated grows at p^n
- Most function evaluations are a waste of time

Conclusion

- A DFO method
- Generally not a good idea for BBO

DFO

DFO Framework

The basic approach

1. Evaluate f at some points
2. Use the information we gathered to decide where to evaluate next

---

Example: Suppose $f : \mathbb{R}^2 \mapsto \mathbb{R}$ and $\Omega = [0, 10]^2$
Suppose

$$\begin{aligned} f(0,0) &= 4 \\ f(0,10) &= 4 \\ f(10,0) &= 5 \\ f(10,10) &= 7 \\ f(5,5) &= 4 \end{aligned}$$

Where should we evaluate next?

Assignment 2

Assignment 2

MATH 462

- Textbook # 3.2(abc), 3.4
- Textbook # 3.5 (this requires reading Example 3.1)

COSC 419K

- Textbook # 3.2(a), 3.4
- Write a MATLAB script that numerical solves Textbook # 3.2(bc)
- Textbook # 3.5 (this requires reading Example 3.1)

MATH 562

- All MATH 462 and COSC 419K questions
- Write a MATLAB script that numerical completes Textbook # 3.6
- Textbook # 3.8