Derivative Free Optimization

Topic 4: Nelder-Mead

Warren Hare

Nelder-Mead

A triangle moving downhill

- In 1965, J. Nelder and R. Mead published a 6 page paper
- The paper outlined an algorithm they called *a simplex method*
- Dantzig was already using the name *simplex method* for linear programming
- The optimization community renamed the Nelder-Mead method
- In 2020, google scholar lists over 30,000 citations to Nelder and Mead's paper

NELDER-MEAD philosophy

Reflect a *bad point* through the *good face* of a simplex to seek better solutions

Nelder-Mead

- Easy to implement
- Tends to adjust to local geometry of a problem
- Works well in many cases
- But, can be proven to fail for a 'easy' function

Simplex

Simplex

Definition: A **simplex** in $\mathbb{R}^n$ is a *bounded convex polytope* with non-empty *interior* and exactly $n+1$ vertices

Convex

Definition: A set Ω is **convex** if and only if

- given any two points $x, y \in \Omega$ and
- any $\theta \in [0, 1]$

we *always* have $\theta x + (1 - \theta)y \in \Omega$

Definition: The **convex hull** of Ω, $\mathrm{conv}(\Omega)$, is the smallest convex set containing Ω

Bounded convex polytope

Theorem: Let $Y = \{y^0, y^1, \ldots, y^m\} \subseteq \mathbb{R}^n$
Then

$$\mathrm{conv}(Y) = \left\{ y \in \mathbb{R}^n \ : \ y = \sum_{i=0}^{m} \theta_i y^i, \ y^i \in Y, \sum_{i=0}^{m} \theta_i = 1, \theta_i \geq 0 \right\}$$

Definition: Ω is a **bounded convex polytope** if and only if Ω is the convex hull of a finite set of points

Interior

Definition: A set S is **open** if for any $x \in S$ there exists $r > 0$ such that $B_r(x) \subseteq S$, where $B_r(x) = \{y \;:\; \|y - x\| < r\}$

Definition: The **interior** of a set S, $\mathrm{int}(S)$, is the largest open set that is contained in S

Simplex

Theorem: Let $\mathtt{Y} = \{\mathtt{y}^0, \mathtt{y}^1, \ldots, \mathtt{y}^n\} \subseteq \mathbb{R}^n$

Then the following are equivalent

1. $\mathrm{conv}(\mathtt{Y})$ is a simplex
2. The set $\{(y^1 - y^0), (y^2 - y^0), \ldots, (y^n - y^0)\}$ is linearly independent
3. The matrix
$$L = \begin{bmatrix}(y^1 - y^0) & (y^2 - y^0) & \ldots & (y^n - y^0)\end{bmatrix}$$
is invertible
4. $\det(L) \neq 0$

If these hold, then we say $\mathtt{Y}$ *forms* a simplex

If these fail, then we say $\mathtt{Y}$ forms a *degenerate* simplex

Volume of a simplex

Definition: Let $\mathtt{Y} = \{\mathtt{y}^0, \mathtt{y}^1, \ldots, \mathtt{y}^\mathtt{n}\} \subseteq \mathbb{R}^\mathtt{n}$
The **volume** of $\mathrm{conv}(\mathtt{Y})$ is defined

$$\mathrm{vol}(\mathrm{conv}(\mathtt{Y})) = \frac{|\det(\mathtt{L})|}{\mathtt{n}!}$$

The **diameter** of $\mathrm{conv}(\mathtt{Y})$ is defined

$$\mathrm{diam}(\mathrm{conv}(\mathtt{Y})) = \max\{\|\mathtt{y}^\mathtt{i} - \mathtt{y}^\mathtt{j}\| \ : \ \mathtt{y}^\mathtt{i} \in \mathtt{Y}, \mathtt{y}^\mathtt{j} \in \mathtt{Y}\}$$

The **normalized volume** of $\mathrm{conv}(\mathtt{Y})$ is

$$\mathrm{von}(\mathrm{conv}(\mathtt{Y})) = \frac{\mathrm{vol}(\mathtt{Y})}{(\mathrm{diam}(\mathtt{Y}))^\mathtt{n}}$$

Notation

For ease of writing we use

Definition:

- $\mathrm{vol}(\mathtt{Y}) = \mathrm{vol}(\mathrm{conv}(\mathtt{Y}))$
- $\mathrm{diam}(\mathtt{Y}) = \mathrm{diam}(\mathrm{conv}(\mathtt{Y}))$
- $\mathrm{von}(\mathtt{Y}) = \mathrm{von}(\mathrm{conv}(\mathtt{Y}))$

Examples

1. Let
$$\mathtt{Y} = \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$
Find $\mathrm{vol}(\mathtt{Y})$, $\mathrm{diam}(\mathtt{Y})$, $\mathrm{von}(\mathtt{Y})$

2. Let
$$\mathtt{Y} = \left\{ \begin{bmatrix} 6 \\ 2 \\ -1 \end{bmatrix} \begin{bmatrix} 7 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 5 \\ 2 \\ 0 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ -1 \end{bmatrix} \right\}$$
Find $\mathrm{vol}(\mathtt{Y})$, $\mathrm{diam}(\mathtt{Y})$, $\mathrm{von}(\mathtt{Y})$

Approximate diameter

Definition: Let $\mathtt{Y} = \{\mathtt{y}^0, \mathtt{y}^1, \ldots, \mathtt{y}^\mathtt{m}\} \subseteq \mathbb{R}^\mathtt{n}$
The **approximate diameter** of $\mathrm{conv}(\mathtt{Y})$ is

$$\overline{\mathrm{diam}}(\mathrm{conv}(\mathtt{Y})) = \max\{\|\mathtt{y}^\mathtt{i} - \mathtt{y}^0\| \; : \; \mathtt{y}^\mathtt{i} \in \mathtt{Y}\}$$

Definition: $\overline{\mathrm{diam}}(\mathtt{Y}) = \overline{\mathrm{diam}}(\mathrm{conv}(\mathtt{Y}))$

Theorem: $\overline{\mathrm{diam}}(\mathtt{Y}) \leq \mathrm{diam}(\mathtt{Y}) \leq 2\overline{\mathrm{diam}}(\mathtt{Y})$

Examples

1. Let
$$Y = \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$
Find $\overline{\mathrm{diam}}(Y)$

2. Let
$$Y = \left\{ \begin{bmatrix} 6 \\ 2 \\ -1 \end{bmatrix} \begin{bmatrix} 7 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 5 \\ 2 \\ 0 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ -1 \end{bmatrix} \right\}$$
Find $\overline{\mathrm{diam}}(Y)$

The Nelder-Mead Algorithm

Algorithm: Nelder-Mead (NM)

Given $f : \mathbb{R}^n \mapsto \mathbb{R}$ and the vertices of an initial simplex $\mathtt{Y}^0 = \{\mathtt{y}^0, \mathtt{y}^1, \ldots, \mathtt{y}^n\}$

0. Initialize:
   - $\delta^e, \delta^{oc}, \delta^{ic}, \gamma$ parameters
   - $k \leftarrow 0$ iteration counter
1. Order and create centroid:
   - reorder $\mathtt{Y}^k$ so $f(y^0) \leq f(y^1) \leq \ldots \leq f(y^n)$
   - set $x^c = \frac{1}{n} \sum_{i=0}^{n-1} y^i$, the centroid of all except the worst point
2. Reflect:
   - test reflection point $x^r = x^c + (x^c - y^n)$
     - if $f(y^0) \leq f(x^r) < f(y^{n-1})$, then accept x^r and goto 1
3. Expand:
   - if $f(x^r) < f(y^0)$, then test expansion point $x^e = x^c + \delta^e(x^c - y^n)$

4a). Outside Contraction:
   - if $f(y^{n-1}) \leq f(x^r) < f(y^n)$, then test outside contraction $x^{oc} = x^c + \delta^{oc}(x^c - y^n)$

4b). Inside Contraction:
   - if $f(x^r) \geq f(y^n)$, then test inside contraction point $x^{ic} = x^c + \delta^{ic}(x^c - y^n)$

5. Shrink:
   - if all tests fail, then shrink
     - $\mathtt{Y}^{k+1} = \{\mathtt{y}^0, \mathtt{y}^0 + \gamma(\mathtt{y}^1 - \mathtt{y}^0), \mathtt{y}^0 + \gamma(\mathtt{y}^2 - \mathtt{y}^0), \ldots, \mathtt{y}^0 + \gamma(\mathtt{y}^n - \mathtt{y}^0)\}$

Algorithm: Nelder-Mead (NM)

Given $f : \mathbb{R}^n \mapsto \mathbb{R}$ and the vertices of an initial simplex $\mathtt{Y}^0 = \{\mathtt{y}^0, \mathtt{y}^1, \ldots, \mathtt{y}^n\}$

0. Initialize:

$\delta^e \in (1, \infty)$	expansion parameter
$\delta^{oc} \in (0, 1)$	outside contraction parameter
$\delta^{ic} \in (-1, 0)$	inside contraction parameter
$\gamma \in (0, 1)$	shrink parameter
$k \leftarrow 0$	iteration counter

Algorithm: Nelder-Mead (NM)

1. Order and create centroid:

   reorder $\mathrm{Y^k}$ so $f(y^0) \leq f(y^1) \leq \ldots \leq f(y^n)$
   set $f^k_{\texttt{best}} = f(y^0)$
   set $x^c = \frac{1}{n} \sum_{i=0}^{n-1} y^i$, the centroid of all except the worst point

---

When ordering points, you can break ties

- at random
- by point age
- by distant to y^0
- etc

Algorithm: Nelder-Mead (NM)

2. Reflect:

   set $x^r = x^c + (x^c - y^n)$ and $f^r = f(x^r)$
   if $f^k_{\texttt{best}} \le f^r < f(y^{n-1})$, then
   set $\mathtt{Y}^{\mathtt{k+1}} = \{\mathtt{y}^0, \mathtt{y}^1, \dots, \mathtt{y}^{\mathtt{n}-1}, \mathtt{x}^{\mathtt{r}}\}$
   increment $k \leftarrow k+1$ and go to 1

Algorithm: Nelder-Mead (NM)

3. Expand:

   if $f^r < f^k_{\texttt{best}}$, then test expansion point
   set $x^e = x^c + \delta^e(x^c - y^n)$, $f^e = f(x^e)$
   if $f^e < f^r$, then set $\mathtt{Y}^{\mathtt{k+1}} = \{\mathtt{y}^0, \mathtt{y}^1, \ldots, \mathtt{y}^{\mathtt{n-1}}, \mathtt{x}^{\mathtt{e}}\}$
   else $(f^r \le f^e)$, then set $\mathtt{Y}^{\mathtt{k+1}} = \{\mathtt{y}^0, \mathtt{y}^1, \ldots, \mathtt{y}^{\mathtt{n-1}}, \mathtt{x}^{\mathtt{r}}\}$
   increment $k \leftarrow k + 1$ and go to 1

Algorithm: Nelder-Mead (NM)

4a). Outside Contraction:

if $f(y^{n-1}) \leq f^r < f(y^n)$, then test outside contraction point
 set $x^{oc} = x^c + \delta^{oc}(x^c - y^n)$, $f^{oc} = f(x^{oc})$
 if $f^{oc} < f^r$, then set $\mathtt{Y}^{k+1} = \{\mathtt{y}^0, \mathtt{y}^1, \ldots, \mathtt{y}^{n-1}, \mathtt{x}^{oc}\}$
 else $(f^r \leq f^{oc})$, then set $\mathtt{Y}^{k+1} = \{\mathtt{y}^0, \mathtt{y}^1, \ldots, \mathtt{y}^{n-1}, \mathtt{x}^{r}\}$
 increment $k \leftarrow k+1$ and go to 1

4b). Inside Contraction:

if $f^r \geq f(y^n)$, then test inside contraction point
 set $x^{ic} = x^c + \delta^{ic}(x^c - y^n)$, $f^{ic} = f(x^{ic})$
 if $f^{ic} < f(y^n)$, then
  set $\mathtt{Y}^{k+1} = \{\mathtt{y}^0, \mathtt{y}^1, \ldots, \mathtt{y}^{n-1}, \mathtt{x}^{ic}\}$
  increment $k \leftarrow k+1$ and go to 1

Algorithm: Nelder-Mead (NM)

5. Shrink:

   set $\mathtt{Y}^{\mathtt{k}+1} = \{\mathtt{y}^0, \mathtt{y}^0 + \gamma(\mathtt{y}^1 - \mathtt{y}^0), \mathtt{y}^0 + \gamma(\mathtt{y}^2 - \mathtt{y}^0), \dots, \mathtt{y}^0 + \gamma(\mathtt{y}^{\mathtt{n}} - \mathtt{y}^0)\}$
   increment $k \leftarrow k + 1$ and go to 1

Nelder-Mead and Simplex

Theorem: Let $\mathtt{Y}^{\mathtt{k}}$ form a simplex and suppose NM is used to create $\mathtt{Y}^{\mathtt{k}+1}$
Then, $\mathtt{Y}^{\mathtt{k}+1}$ forms a simplex

Volumes and Normalized Volumes

Theorem: Let $\mathtt{Y}^{\mathtt{k}}$ form a simplex and suppose NM is used to create $\mathtt{Y}^{\mathtt{k+1}}$

1. If a shrink step occurred, then
$$\mathrm{vol}(\mathtt{Y}^{\mathtt{k+1}}) = \gamma^{\mathtt{n}}\mathrm{vol}(\mathtt{Y}^{\mathtt{k}}) \quad \text{and} \quad \mathrm{von}(\mathtt{Y}^{\mathtt{k+1}}) = \mathrm{von}(\mathtt{Y}^{\mathtt{k}})$$
2. If an nonshrink step occurred, then
$$\mathrm{vol}(\mathtt{Y}^{\mathtt{k+1}}) = |\delta|\mathrm{vol}(\mathtt{Y}^{\mathtt{k}})$$
where $\delta \in \{\delta^r, \delta^e, \delta^{ic}, \delta^{oc}\}$ as appropriate

Convergence

Convergence

Theorem: Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be bounded below
Suppose NM is used to seek $\min\{f(x) : x \in \mathbb{R}^n\}$
Denote $\mathtt{Y}^{\mathtt{k}} = \{\mathtt{y}^{0,\mathtt{k}}, \mathtt{y}^{1,\mathtt{k}}, \ldots, \mathtt{y}^{\mathtt{n},\mathtt{k}}\}$ with $\mathtt{Y}^{\mathtt{k}}$ ordered
Denote $x^k_{\mathtt{best}} = y^{0,k}$ and $f^k_{\mathtt{best}} = f(y^{0,k})$
Then

1. $f^k_{\mathtt{best}}$ converges to some value $\hat{f}$
2. If a *finite number* of *non-shrink steps* occur, then $x^k_{\mathtt{best}}$ converges to some point $\hat{x}$, and all vertices $y^{i,k} \in \mathtt{Y}^{\mathtt{k}}$ also converge to $\hat{x}$
3. If a *finite number* of *shrink steps* occur, then for each fixed $i = 0, 1, \ldots, n$ the values $f(y^{i,k})$ converge to some value $\hat{f}^i$
   Moreover, these values satisfy $\hat{f}^0 \leq \hat{f}^1 \leq \ldots \leq \hat{f}^n$

Remarks

Notice

- The theorem does not say $\hat{f}$ is the minimum function value
- The theorem does not say $\hat{x}$ is a minimizer
- The theorem does not consider the case of infinite shrink and infinite non-shrink steps

In 1998, McKinnon demonstrated that NM can misbehave

The McKinnon Example

In search of a example

- 1987, Dennis and Wood create an example where 'a minor adaptation' to NM failed to converge
- 1994, Strasser gave an example where NM failed to converge, but the objective function was nonconvex
- 1998, McKinnon gave an example of a convex $\mathcal{C}^2$ objective function where NM failed to converge

McKinnon's idea was to create an example where an infinite sequence of inside contradictions occurred

McKinnon Example

Use the standard parameters

$$\delta^e = 2, \quad \delta^{oc} = 1/2, \quad \delta^{ic} = -1/2, \quad \gamma = 1/2$$

To create infinite inside contractions, we need

$$\begin{array}{rcll} f(x^r) & \geq & f(y^n) & \text{inside contraction checked} \\ f(x^{ic}) & < & f(y^n) & \text{inside contraction accepted} \end{array}$$

To ensure simplex shape remains similar, we need

$$f(x^{ic}) \quad < \quad f(y^1)$$

McKinnon Example

Without loss of generality

$$y^0 = 0$$

Define

$$y^{0,k} = 0, y^{1,k} = v^k, \quad y^{2,k} = w^k$$

We want

$$\begin{array}{rcllcl} y^{0,k+1} & = & y^{0,k} & & = & 0 \\ y^{1,k+1} & = & v^{k+1} & & = & x^{ic} \\ y^{2,k+1} & = & y^{1,k} = v^k & & & \end{array}$$

Recurrence relations

This leads to the recurrence relation

$$\begin{aligned} v^{k+1} &= x^{ic} \\ &= \dots \\ &= \frac{1}{4}v^k + \frac{1}{2}v^{k-1} \end{aligned}$$

Alternately

$$4v^{k+2} - v^{k+1} - 2v^k = 0$$

Recurrence relations

Theorem: Let A, B, C be real numbers
Consider the **second-order recurrence relation** of x^k given by

$$Ax^{k+2} + Bx^{k+1} + Cx^k = 0 \tag{1}$$

Then $\{1, t, (t)^2, (t)^3, \dots, \}$ satisfies (1)
if and only if

$$A(t)^2 + Bt + C = 0$$

Definition: The equation $A(t)^2 + Bt + C = 0$ is called the **characteristic equation** of the recurrence relation

McKinnon Example

The recurrence relation

$$4v^{k+2} - v^{k+1} - 2v^k = 0$$

links to solution

$$t = \frac{1 \pm \sqrt{33}}{8}$$

Define

$$\lambda = \frac{1 + \sqrt{33}}{8} \quad \mu = \frac{1 - \sqrt{33}}{8}$$

and set

$$y^{0,0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad v^1 = y^{1,0} = \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \quad v^0 = y^{2,0} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

McKinnon Example

$$\lambda = \frac{1+\sqrt{33}}{8} \quad \mu = \frac{1-\sqrt{33}}{8}$$

$$\mathrm{Y}^0 = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \quad \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$$

$$\mathrm{Y}^1 = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \begin{bmatrix} \lambda^2 \\ \mu^2 \end{bmatrix} \quad \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \right\}$$

$$\mathrm{Y}^2 = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \begin{bmatrix} \lambda^3 \\ \mu^3 \end{bmatrix} \quad \begin{bmatrix} \lambda^2 \\ \mu^2 \end{bmatrix} \right\}$$

...

McKinnon Points

Using this, we have

$$\begin{aligned} x^c &= \frac{1}{2}\begin{bmatrix}\lambda^{k+1}\\ \mu^{k+1}\end{bmatrix} \\ x^r &= \begin{bmatrix}\lambda^k(\lambda-1)\\ \mu^k(\mu-1)\end{bmatrix} \\ x^{ic} &= \begin{bmatrix}\lambda^{k+2}\\ \mu^{k+2}\end{bmatrix} \end{aligned}$$

So all we need now is f such that

$$\begin{aligned} f(x^r) &\geq f(v^k) \quad \text{inside contraction checked} \\ f(x^{ic}) &< f(v^k) \quad \text{inside contraction accepted} \end{aligned}$$

where

$$v^k = \begin{bmatrix}\lambda^{k+1}\\ \mu^{k+1}\end{bmatrix}$$

McKinnon Function Form

Consider the *functional form*

$$f(x_1, x_2) = \begin{cases} \theta_{big}(x_1)^2 + g(x_2) & \text{if } x_1 \leq 0 \\ \theta_{small}(x_1)^2 + g(x_2) & \text{if } x_1 > 0, \end{cases}$$

We want

- g to be smooth, convex, and fairly flat
- θ_{big} to ensure x^r is rejected
- θ_{small} to ensure x^{ic} is accepted
- the minimizer to **not** be $[0 \;\; 0]^\top$

McKinnon magic values

Consider the *functional form*

$$f(x_1, x_2) = \begin{cases} \theta_{big}(x_1)^2 + g(x_2) & \text{if } x_1 \leq 0 \\ \theta_{small}(x_1)^2 + g(x_2) & \text{if } x_1 > 0, \end{cases}$$

We want

- $g(x_2) = x_2 + (x_2)^2$ – gives g smooth, convex
- $\theta_{big} = 360$ – ensures x^r is rejected
- $\theta_{small} = 6$ – ensures x^{ic} is accepted
- $g(x_2) = x_2 + (x_2)^2$ – gives the minimizer $[0 \quad -1/2]^\top$

Final Remarks

Nelder-Mead

Pros:

- Easy to implement
- Often very effective

Cons:

- Convergence not satisfying
- Stopping condition not clear*

* Stopping condition will become clear in topic 7: building models

Conclusion

- Not really a DFO method*
- A very easy and effective heuristic for BBO

* see next slide

Fortified Nelder-Mead

1999, Tseng adapted NM to create a FORTIFIED NELDER-MEAD method that was proven to converge

- For failed reflection, replace shrink with a rotation step

$$\mathtt{Y}^{\mathtt{rot}} = \{\mathtt{y}^0, \mathtt{y}^0 - (\mathtt{y}^1 - \mathtt{y}^0), \mathtt{y}^0 - (\mathtt{y}^2 - \mathtt{y}^0), \ldots, \mathtt{y}^0 - (\mathtt{y}^n - \mathtt{y}^0)\}$$

- At each iteration, check $\mathrm{von}(\mathtt{Y}^{\mathtt{k}+1})$ and $\mathrm{diam}(\mathtt{Y}^{\mathtt{k}+1})$

  if getting too small, then take corrective action

Assignment 4

Assignment 4

MATH 462

1. Is it possible for $\mathrm{von}(\mathtt{Y}^{\mathtt{k}})$ to be the same as $\mathrm{von}(\mathtt{Y}^{\mathtt{k+1}})$ if an expansion step took place. (justify your answer)

- Textbook # 2.15, 5.1, 5.4, 5.6, 5.9(ab)

COSC 419K

2. Write a MATLAB code that takes an input n and numerically solves Textbook # 5.8(b) in $\mathbb{R}^n$
3. Write a MATLAB code that takes an input k and numerically solves the non-limit values in Textbook # 5.12

- Textbook # 2.14, 5.2, 5.3

MATH 562

- All MATH 462 and COSC 419K questions
- Textbook # 5.7, 5.12