Tutorial 3

Fundamentals

CS/SWE 4/6TE3, CES 722/723 September 28, 2010

Rate of Convergence

Let $\alpha_1, \alpha_2, ..., \alpha_n \to \alpha$ be a convergent sequence, the order of convergence is p^* if

$$p^* = \sup\left\{p : \lim_{k\to\infty} \frac{|\alpha_{k+1} - \alpha|}{|\alpha_k - \alpha|^p} < \infty\right\}$$

If $p^* = 1$, then let

$$\beta = \lim_{k\to\infty} \frac{|\alpha_{k+1} - \alpha|}{|\alpha_k - \alpha|}$$

If $p^* = 1$ and $0 < \beta < 1 \to$ *linear* convergence
If $p^* = 1$ and $\beta = 0 \to$ convergence rate is *super-linear* (fast convergence)
If $p^* = 1$ and $\beta = 1 \to$ convergence rate is *sub-linear* (slow)
If $p^* = 2 \to$ *quadratic* convergence
To derive the order p^* of a sequence, we must show that for $p < \epsilon$, the sequence converges to a finite value, however for $p > \epsilon$, the sequence diverges, hence p^* is the sup.

Examples

1) Linear convergence: $\alpha_k = a^k$ where $0 < a < 1 \to p = 1$ and $\beta = \frac{a^{k+1}}{(a^k)^1} = a$
As well, $\lim_{k\to\infty} \frac{a^{k+1}}{(a^k)^{1+\epsilon}} = \frac{a}{a^{k\epsilon}} = \infty \quad \forall\epsilon > 0$
2) Quadratic convergence: $\alpha_k = a^{(2^k)} \to p = 2$ and $\lim_{k\to\infty} \frac{a^{(2^{k+1})}}{(a^{(2^k)})^2} = \frac{a^{(2^{k+1})}}{a^{(2^k)2}} = 1$
And, $\lim_{k\to\infty} \frac{a^{(2^{k+1})}}{(a^{(2^k)})^{2+\epsilon}} = \frac{1}{(a^{2^k})^\epsilon} = \infty \quad \forall\epsilon > 0$
3) Sub linearly: $\alpha_k = \frac{1}{k} \to p = 1$ and $\beta = \frac{1/(k+1)}{(1/k)^1} = \lim_{k\to\infty} \frac{k}{k+1} = 1$
And, $\lim_{k\to\infty} \frac{1/(k+1)}{(1/k)^{1+\epsilon}} = \frac{kk^\epsilon}{k+1} = \infty \quad \forall\epsilon > 0$
4) Super linearly: $\alpha_k = (\frac{1}{k})^k \to p = 1$ and $\beta = \lim_{k\to\infty} \left(\frac{k}{k+1}\right)^k \left(\frac{1}{k+1}\right) = 0$
And, $\lim_{k\to\infty} \left(\frac{k}{k+1}\right)^k \left(\frac{1}{k+1}\right) k^{k\epsilon} = \infty \quad \forall\epsilon > 0$

Pattern Search (Hooke and Jeeves)

Pattern Search Overview Given a set of n directions, check forward/backward of the first direction, if $f(y^{j-1} \pm d^j) \not< f(y^{j-1})$ then take the next direction. Continue checking this inequality for each of the n directions until one is true. After which, if $f(y^{j-1} \pm d^j) \not< f(y^{j-1})$ *still* holds then shrink the direction, $d_j = \alpha d_j \; \forall \; \alpha \in (0, 1)$. Thus, since we check forward/backward, we have checked $2n$ directions. However, if we use the cone direction search, then $n+1$ directions is needed and we only need to check $n+1$ directions, which is faster then previous.

The Algorithm

Input:

x^0	- given initial point
$d^1, d^2, ..., d^n$	- n linearly independent search directions
$0 < \alpha < 1$	- damping factor
$\epsilon > 0$	- accuracy parameter

Step 0: $k = 0$
Step 1: $y^0 = x^k$; Call PS
Step 2: If $y^n = y^0$ Then ("couldn't do better so shrink the steps") $d^j = \alpha d^j$; Go to **Step 6**
Step 2a: $z = y^n$
Step 3: [We have $f(z) < f(x^k)$]; $y^0 = y^n + (y^n - x^k)$; Call PS
Step 4: If $f(y^n) < f(z)$ Then $z = y^n$
Step 5: [We have $f(z) < f(x^k)$]; $x^{k+1} = z$; $k = k + 1$
Step 6: If $\|d_1\| < \epsilon$ **Stop**, Else go to **Step 1**.

SUBROUTINE (PS): (Pattern search)

For $j = 1 : n$ Do
If $f(y^{j-1} + d^j) < f(y^{j-1})$ Then $y^j = y^{j-1} + d^j$;
Else If $f(y^{j-1} - d^j) < f(y^{j-1})$
Then $y^j = y^{j-1} - d^j$;
Else $y^j = y^{j-1}$;
End

Simplex Method (Nelder and Mead)

Sort the points, then take an average. Compare the average with all the other points and try to remove one, if not, shrink the points by $\alpha \in (0, 1)$.

Input:

$x^0, x^1, ..., x^n$	- $n + 1$ points in a general position
$0 < \alpha < 1$	- damping factor

$0 < \beta < 1$ - contradiction factor
$\gamma > 1$ - extension factor
$\epsilon > 0$ - accuracy parameter

Step 0: Sort the points in the order of ascending functional value;

$$f(x^0) \leq f(x^1) \leq \cdots \leq f(x^n)$$

Step 1: Let

$$\overline{x} = \frac{1}{n}\sum_{j=0}^{n-1} x^j; \quad x^r = \overline{x} + (\overline{x} - x^n);$$

Step 1a:

If $f(x^r) < f(x^0)$ "Better"
Then $x^e = \overline{x} + \gamma(\overline{x} - x^n)$ "Take Better"
If $f(x^e) \leq f(x^r)$
Then DROP x^n and ADD x^e go to 1.
Else DROP x^n and ADD x^r go to 1.

Step 1b:

If $f(x^0) \leq f(x^r) < f(x^{n-1})$ "Better than Second Worst"
Then DROP x^n and ADD x^r go to 1.

Step 1c:

If $f(x^{n-1}) \leq f(x^r) < f(x^n)$ "Between Second Worst and Worst"
Then $x^c = \overline{x} + \beta(\overline{x} - x^n)$
If $f(x^c) \leq f(x^{n-1})$
Then DROP x^n and ADD x^c go to 1.

Step 1d:

If $f(x^r) \geq f(x^n)$
Then $x^c = \overline{x} - \beta(\overline{x} - x^n)$
If $f(x^c) \leq f(x^{n-1})$
Then DROP x^n and ADD x^c go to 1.

Step 2: If $\max\{\|x^i - x^j\| : 1 \leq i, j\} < \epsilon$ OR $f(x^n) - f(x^0) < \epsilon$ Then **Stop**.
Step 3: Contract the simplex to x^0; $x^j = x^0 + \alpha(x^j - x^0),\ j = 1, 2, ..., n$ go to 1.

Nedler-Mead Example

Example for the banana function:

$$f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

$\alpha = 1/2$ damping factor
$\beta = 2/3$ contraction factor

$\gamma = 2$ extension factor

$$x^0 = (3,3);\ x^1 = (1,2);\ x^2 = (2,1);$$

Iteration one:

Step 1: Sort the points in the order of ascending function value:

$$f(x^0) = f(1,2) = 100 \le f(x^1) = f(2,1) = 901 \le f(x^2) = f(3,3) = 3604$$

Step 2:

$$\bar{x} = \frac{1}{2}\sum_{j=0}^{1} x^j = (1.5, 1.5);\ x^r = \bar{x} + (\bar{x} - x^2) = (0,0)$$

Step 2a:

$$f(x^r) = f(0,0) = 1 < f(x^0) = f(1,2) = 100 \quad \Rightarrow \quad x^e = \bar{x} + \gamma(\bar{x} - x^2) = (-1.5, -1.5)$$

$$f(x^e) = f(-1.5,-1.5) = 1412.5 > f(x^r) = f(0,0) = 1 \quad \Rightarrow$$

$$\text{DROP } x^2, \text{ ADD } x^r = (0,0), \text{ GOTO 1}$$

Iteration two:

Step 1: Sort the points in the order of ascending function value:

$$f(x^0) = f(0,0) = 1 \le f(x^1) = f(1,2) = 100 \le f(x^2) = f(2,1) = 901$$

Step 2:

$$\bar{x} = \frac{1}{2}\sum_{j=0}^{1} x^j = (0.5, 1);\ x^r = \bar{x} + (\bar{x} - x^2) = (-1,1)$$

Step 2b:

$$f(x^0) = f(0,0) = 1 \le f(x^r) = f(-1,1) = 4 \le f(x^1) = f(1,2) = 100 \quad \Rightarrow$$

$$\text{DROP } x^2, \text{ ADD } x^r = (-1,1), \text{ GOTO 1}$$

Iteration three:

Step 1: Sort the points in the order of ascending function value:

$$f(x^0) = f(0,0) = 1 \le f(x^1) = f(-1,1) = 4 \le f(x^2) = f(1,2) = 100$$

Step 2:

$$\bar{x} = \frac{1}{2}\sum_{j=0}^{1} x^j = (-0.5, 0.5);\ x^r = \bar{x} + (\bar{x} - x^2) = (-2,-1)$$

Step 2d:

$$f(x^r) = f(-2,-1) = 2509 \ge f(x^2) = f(1,2) = 100 \quad \Rightarrow \quad x^c = \bar{x} - \beta(\bar{x} - x^2) = (0.5, 1.5)$$

$$f(x^c) = f(0.5, 1.5) = 156.5 > f(x^1) = f(-1,1) = 4 \quad \Rightarrow \ \text{GOTO 3,4}$$

Step 4: Contract the simplex to x^0:

$$x^j = x^0 + \alpha(x^j - x^0),\ j = 1, 2$$

$$x^1 = x^0 + \alpha(x^1 - x^0) = (-0.5, 0.5)\quad x^2 = x^0 + \alpha(x^2 - x^0) = (0.5, 1),\ \text{GOTO 1}$$

Iteration four:

Step 1: Sort the points in the order of ascending function value:

$$f(x^0) = f(0, 0) = 1 \leq f(x^1) = f(-0.5, 0.5) = 8.5 \leq f(x^2) = f(0.5, 1) = 56.5$$

Step 2:

$$\bar{x} = \frac{1}{2}\sum_{j=0}^{1} x^j = (-0.25, 0.25);\ x^r = \bar{x} + (\bar{x} - x^2) = (-1, -0.5)$$

Step 2d:

$$f(x^r) = f(-1, -0.5) = 229 \geq f(x^2) = f(0.5, 1) = 56.6 \quad \Rightarrow \quad x^c = \bar{x} - \beta(\bar{x} - x^2) = (0.25, 0.75)$$

$$f(x^c) = f(0.25, 0.75) = 47.8 > f(x^1) = f(-0.5, 0.5) = 8.5 \quad \Rightarrow \ \text{GOTO 3}$$

Step 3: The Nedler-Mead Simplex method produced the following results after 4 iterations: $x^0 = (0, 0);\ f(x^0) = 1$.

Step 4: Contract the simplex to x^0:

$$x^j = x^0 + \alpha(x^j - x^0),\ j = 1, 2$$

$$x^1 = x^0 + \alpha(x^1 - x^0) = (-0.25, 0.25)\quad x^2 = x^0 + \alpha(x^2 - x^0) = (0.25, 0.5),\ \text{GOTO 1}$$

Iteration five: ...

Golden Section Line Search

Finding a minimum of f $\rightarrow$ min f(x)

Unimodal Function

A univariate function is unimodal in $[a, b]$ if there exists a unique $x^* \in [a, b]$ such that for any $x_1, x_2 \in [a, b]$ where, $x_1 < x_2$

$$\begin{array}{lll} \text{If } x_2 < x^*, & \text{then } f(x_1) > f(x_2) & \text{(slope down)} \\ \text{If } x_1 > x^*, & \text{then } f(x_2) > f(x_1) & \text{(slope up)}\ . \end{array}$$

We can then reduce the interval of a unimodal function,

If $f(x_1) < f(x_2) \quad \rightarrow \quad [a, x_2]$

If $f(x_2) < f(x_1) \quad \rightarrow \quad [x_1, b]$

If $f(x_1) = f(x_2) \quad \rightarrow \quad [x_1, x_2]$.

Golden Section Search

Gold Ratio: Given a rectangle with sides 1 and ϕ, ϕ is defined such that partitioning the original rectangle into a square and new rectangle results in a new rectangle having sides with a ratio $1, \phi$. Such a rectangle is called a golden rectangle and we have,

$$\frac{1}{\phi - 1} = \phi \quad \rightarrow \quad \phi^2 - \phi - 1 = 0$$
$$\phi = \frac{1 \pm \sqrt{5}}{2} \approx \; -0.618033, 1.618033 \; .$$

We can use this approach for minimization, where we take $x_1, x_2 \in [0, 1]$ and a decreasing ratio τ such that we let $x_1 = 1 - \tau,\; x_2 = \tau$. Assuming $[0, \tau]$ is our new interval, $x_1 = 1 - \tau \in [0, \tau]$ and the decreasing rate for the next interval $\left(\frac{1-\tau}{\tau}\right)$ should be equal to the first τ.

$$\tau^2 + \tau - 1 = 0 \quad \rightarrow \quad \tau = \frac{-1 \pm \sqrt{5}}{2} \approx \; 0.618033, -1.618033 \; .$$

The Golden Section Search takes $\tau \approx 0.618$, where for an interval $[a, b]$,

$$\begin{aligned} x_1 &= a + (1 - \tau)(b - a) \\ x_2 &= a + \tau(b - a) \; . \end{aligned}$$

The method used to reduce the interval of a unimodal function can then be applied to find an approximation of a minimum. Here we need 4 points at the beginning to find the minimum of a function. Note that when we generate new points we keep the same ratio between them.