Extreme values: Iterative methods.

Discrete Mathematics and Optimization
Bioinformatics

1. Iterative methods

How to find extreme values?

Definition

An iterative method to find an extreme value of a function $f : D \subset \mathbb{R}^n \to \mathbb{R}$ is a procedure to find a sequence

$$\mathbf{x}_k = h(\mathbf{x}_{k-1}),\ k \geq 1,$$

from some initial value $\mathbf{x}_0$ such that $\mathbf{x}_k$ converges to the optimal value.

Key issues:

- Convergence: Find a region D such that, for $\mathbf{x}_0 \in D$, $\mathbf{x}_k$ converges to the optimal value of f.
- Efficiency: $\epsilon_k = \|\mathbf{x}_k - \mathbf{r}\|$ the error at k–th iteration.
  Convergence is of the order p if $\frac{\epsilon_{k+1}}{\epsilon_k^p} \leq M < 1$.
- Robustness: range of functions where the method is efficient
- Stability: convergence and efficiency are stable for small changes in starting point.

1. Iterative methods

How to find extreme values?

Definition

An iterative method to find an extreme value of a function $f : D \subset \mathbb{R}^n \to \mathbb{R}$ is a procedure to find a sequence

$$\mathbf{x}_k = h(\mathbf{x}_{k-1}),\ k \geq 1,$$

from some initial value $\mathbf{x}_0$ such that $\mathbf{x}_k$ converges to the optimal value.

Examples of iterative methods

- Newton method.
- Method of Steepest Descent.

Most modern iterative methods are variations of one of the two improving their features.

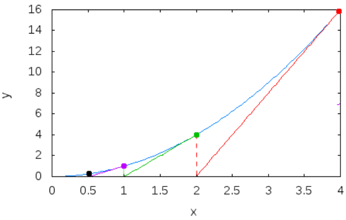
2. Newton method

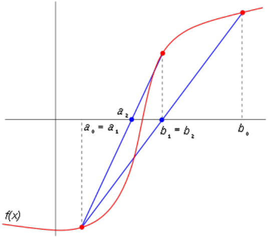
Recall the univariate Newton method for finding zeroes of a differentiable function.

Definition

Let $g : \mathbb{R} \to \mathbb{R}$ be differentiable and $x_0 \in \mathbb{R}$. The Newton method consists in, starting at x_0, defining the sequence

$$x_{k+1} = x_k - \frac{g(x_k)}{g'(x_k)}$$





2. Newton method

Recall the univariate Newton method for finding zeroes of a differentiable function.

Definition

Let $g : \mathbb{R} \to \mathbb{R}$ be differentiable and $x_0 \in \mathbb{R}$. The Newton method consists in, starting at x_0, defining the sequence

$$x_{k+1} = x_k - \frac{g(x_k)}{g'(x_k)}$$

- There are various sufficient conditions which guarantee that the method converges to a solution r of $g(x) = 0$:
  - If $g(a) < 0, g(b) > 0$, $|g'(x)| \geq m$, $|g''(x)| \leq M$ (smoothness of g)
  - $x_0 \in (r - c, r + c)$ (closedness of x_0 to the solution).
- When it converges it does so efficiently $|x_{k+1} - r| \leq C|x_k - r|^2$ (quadratic convergence).

2. Newton method

Recall the univariate Newton method for finding zeroes of a differentiable function.

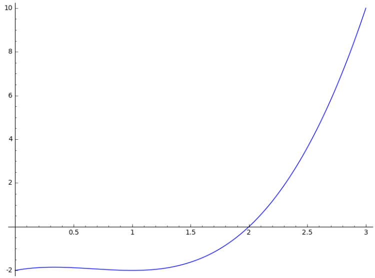
Definition

Let $g : \mathbb{R} \to \mathbb{R}$ be differentiable and $x_0 \in \mathbb{R}$. The Newton method consists in, starting at x_0, defining the sequence

$$x_{k+1} = x_k - \frac{g(x_k)}{g'(x_k)}$$

Example: $g(x) = x^3 - 2x^2 + x - 2$

x_0	1.2	4	1
x_1	3.57730	2.35849	crack!
x_2	2.70966	2.07345	
x_3	2.22393	2.00399	
x_4	2.03212	2.00001	



2. Newton method

Recall the univariate Newton method for finding zeroes of a differentiable function.

Definition

Let $g : \mathbb{R} \to \mathbb{R}$ be differentiable and $x_0 \in \mathbb{R}$. The Newton method consists in, starting at x_0, defining the sequence

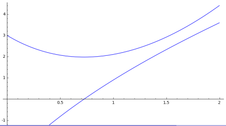
$$x_{k+1} = x_k - \frac{g(x_k)}{g'(x_k)}$$

Example in optimization: Minimum of $g(x) = x^2 + 3e^{-x}$.

- $g''(x) = 2 + 3e^{-x} > 0$: convex function (has a global minimum).
- $g'(x) = 2x - 3e^{-x} = 0$ (critical points)

$$x_{k+1} = x_k - \frac{2x_k - 3e^{-x_k}}{2 + 3e^{-x_k}}.$$

x_0	x_1	x_2	x_3
1	0.711	0.726	0.726



3. Newton method: several variables

Find critical points of f: find zeroes of $\nabla f = \mathbf{0}$.

$$\begin{aligned} \nabla f : \mathbb{R}^n &\rightarrow \mathbb{R}^n \\ \mathbf{x} = (x_1, \ldots, x_n) &\mapsto \left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \ldots, \frac{\partial f}{\partial x_n}(\mathbf{x}) \right) \end{aligned}$$

The gradient of $g = \nabla f$ is the Hessian $\nabla^2 f$ of f

Definition

Let $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be differentiable and $\mathbf{x}_0 \in \mathbb{R}$. The Newton method consists in defining the sequence

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \left(\nabla \mathbf{g}(\mathbf{x}_k)\right)^{-1} \mathbf{g}(\mathbf{x}_k)$$

Alternatively,

$$\nabla \mathbf{g}(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k) = -\mathbf{g}(\mathbf{x}_k).$$

If $\mathbf{x}_0$ is 'close' to a zero of $\mathbf{g}$ and $\mathbf{g}$ is sufficiently smooth, then $\mathbf{x}_k$ tends to $\mathbf{r}$ with $\mathbf{g}(\mathbf{r}) = \mathbf{0}$.

3. Newton method in several variables

Definition

Let $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^n$ be differentiable and $\mathbf{x}_0 \in \mathbb{R}$. The Newton method consists in defining the sequence

$$\nabla \mathbf{g}(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k) = -\mathbf{g}(\mathbf{x}_k).$$

If $\mathbf{g}(\mathbf{x}) = (g_1(\mathbf{x}), \dots, g_n(\mathbf{x}))$, each step requires solving the linear system

$$\begin{aligned} g_1(\mathbf{x}_k) - \nabla g_1(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k) &= 0, \\ &\vdots \\ g_n(\mathbf{x}_k) - \nabla g_n(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k) &= 0 \end{aligned}$$

3. Newton method in several variables

Definition

Let $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^n$ be differentiable and $\mathbf{x}_0 \in \mathbb{R}$. The Newton method consists in defining the sequence

$$\nabla \mathbf{g}(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k) = -\mathbf{g}(\mathbf{x}_k).$$

Example: Zero of $g(x, y) = (4x^3 + 4xy^2, 4x^2y + 4y^3)$:

- Starting point (a, a): $g(a, a) = (8a^3, 8a^3)$.
- Compute the gradients of components of g

$$\nabla g_1(x, y) = (12x^2 + 4y^2, 8xy), \nabla g_1(a, a) = (16a^2, 8a^2)$$
$$\nabla g_2(x, y) = (8xy, 4x^2 + 12y^2), \nabla g_2(a, a) = (8a^2, 16a^2)$$

- First iteration: $(a, a) \to (x, y)$

$$16a^2(x - a) + 8a^2(y - a) = -8a^3$$
$$8a^2(x - a) + 16a^2(y - a) = -8a^3$$

3. Newton method in several variables

Definition

Let $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^n$ be differentiable and $\mathbf{x}_0 \in \mathbb{R}$. The Newton method consists in defining the sequence

$$\nabla \mathbf{g}(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k) = -\mathbf{g}(\mathbf{x}_k).$$

Example: Zero of $g(x, y) = (4x^3 + 4xy^2, 4x^2y + 4y^2)$:

- Starting point (a, a): $g(a, a) = (8a^3, 8a^3)$.
- First iteration: $(a, a) \to (x, y)$

$$16a^2(x - a) + 8a^2(y - a) = -8a^3$$
$$8a^2(x - a) + 16a^2(y - a) = -8a^3$$

- Solution is $(2a/3, 2a/3)$. Iteration gives $((2/3)^k a, (2/3)^k a) \to (0, 0)$.

(Obvious in this case)

3. Newton method in several variables

Definition

Let $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^n$ be differentiable and $\mathbf{x}_0 \in \mathbb{R}$. The Newton method consists in defining the sequence

$$\nabla \mathbf{g}(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k) = -\mathbf{g}(\mathbf{x}_k).$$

Use in optimization: Minimum of $f(x, y) = x^4 + 2x^2y^2 + y^4$.

3. Newton method in several variables

Definition

Let $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^n$ be differentiable and $\mathbf{x}_0 \in \mathbb{R}$. The Newton method consists in defining the sequence

$$\nabla \mathbf{g}(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k) = -\mathbf{g}(\mathbf{x}_k).$$

Use in optimization: Minimum of $f(x, y) = x^4 + 2x^2y^2 + y^4$.

- f is convex: global minimum.

3. Newton method in several variables

Definition

Let $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^n$ be differentiable and $\mathbf{x}_0 \in \mathbb{R}$. The Newton method consists in defining the sequence

$$\nabla \mathbf{g}(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k) = -\mathbf{g}(\mathbf{x}_k).$$

Use in optimization: Minimum of $f(x, y) = x^4 + 2x^2y^2 + y^4$.

- f is convex: global minimum.
- Solve $g(x, y) = \nabla f(x, y) = (4x^3 + 4xy^2, 4x^2y + 4y^3) = (0, 0)$.

3. Newton method in several variables

Definition

Let $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^n$ be differentiable and $\mathbf{x}_0 \in \mathbb{R}$. The Newton method consists in defining the sequence

$$\nabla \mathbf{g}(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k) = -\mathbf{g}(\mathbf{x}_k).$$

Use in optimization: Minimum of $f(x, y) = x^4 + 2x^2y^2 + y^4$.

- f is convex: global minimum.
- Solve $g(x, y) = \nabla f(x, y) = (4x^3 + 4xy^2, 4x^2y + 4y^3) = (0, 0)$.
- $Hf(x, y) = \begin{pmatrix} 12x^2 + 4y^2 & 8xy \\ 8xy & 4x^2 + 12y^2 \end{pmatrix}$

3. Newton method in several variables

Definition

Let $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^n$ be differentiable and $\mathbf{x}_0 \in \mathbb{R}$. The Newton method consists in defining the sequence

$$\nabla \mathbf{g}(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k) = -\mathbf{g}(\mathbf{x}_k).$$

Use in optimization: Minimum of $f(x,y) = x^4 + 2x^2y^2 + y^4$.

- f is convex: global minimum.
- Solve $g(x,y) = \nabla f(x,y) = (4x^3 + 4xy^2, 4x^2y + 4y^3) = (0,0)$.
- $Hf(x,y) = \begin{pmatrix} 12x^2 + 4y^2 & 8xy \\ 8xy & 4x^2 + 12y^2 \end{pmatrix}$
- If starting point is $(1,1)$, the next point (x,y) is the solution of the system

$$\begin{aligned} 16(x-1) + 8(y-1) &= -8 \\ 8(x-1) + 16(y-1) &= -8 \end{aligned}$$

giving $(2/3, 2/3)$. Iteration gives $((2/3)^k, (2/3)^k) \to (0,0)$.

3. Newton method in several variables

Definition

Let $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^n$ be differentiable and $\mathbf{x}_0 \in \mathbb{R}$. The Newton method consists in defining the sequence

$$\nabla \mathbf{g}(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k) = -\mathbf{g}(\mathbf{x}_k).$$

Optimization of quadratic functions:

- $f(\mathbf{x}) = \mathbf{x}^T Q\mathbf{x} + \mathbf{x}^T \mathbf{b} + c$, Q invertible.
- $\mathbf{g}(\mathbf{x}) = \nabla f(\mathbf{x}) = 2Q\mathbf{x} + \mathbf{b}$
- $\nabla \mathbf{g}(\mathbf{x}) = Hf(\mathbf{x}) = 2Q$.
- Given $\mathbf{x}_0$,

$$\mathbf{x}_1 = \mathbf{x}_0 - \frac{1}{2}Q^{-1}(2Q(\mathbf{x}_0) + \mathbf{b}) = -\frac{1}{2}Q^{-1}(\mathbf{b}),$$

  which is the critical point of f.

For quadratic functions with Q invertible the Newton method reaches the critical point in one step.

3. Newton method in several variables

Definition

Let $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^n$ be differentiable and $\mathbf{x}_0 \in \mathbb{R}$. The Newton method consists in defining the sequence

$$\nabla \mathbf{g}(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k) = -\mathbf{g}(\mathbf{x}_k).$$

Optimization of quadratic functions: Example

- $f(x,y) = 2x^2 + 2xy + y^2 - x + 2y + 3 = (x,y)\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + (x,y)\begin{pmatrix} -1 \\ 2 \end{pmatrix} + 3.$
- Critical point $(x_c, y_c) = -\frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} -1 \\ 2 \end{pmatrix} = \begin{pmatrix} 3/2 \\ -5/2 \end{pmatrix}.$
- Since Q is positive definite, $(3/2, -5/2)$ is a global minimum of f.

3. Newton method in several variables

Definition

Let $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^n$ be differentiable and $\mathbf{x}_0 \in \mathbb{R}$. The Newton method consists in defining the sequence

$$\nabla \mathbf{g}(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k) = -\mathbf{g}(\mathbf{x}_k).$$

Newton method has quadratic convergence under some conditions:

Theorem

Let $f \in \mathcal{C}^3(\mathbb{R}^n)$. If f has a critical point at $\mathbf{x}_c$ and $Hf(\mathbf{x}_c)$ is invertible then, for every $\mathbf{x}_0$ close to $\mathbf{x}_c$ the Newton method converges to $\mathbf{x}_c$ with quadratic order.

The key fact is that Newton method uses second order approximation to $\mathbf{g}(\mathbf{x}) = \nabla f(\mathbf{x})$ in Taylor expansion:

$$\nabla f(\mathbf{x}) = \nabla f(\mathbf{x}_0) + \mathbf{x}_0^T Hf(\mathbf{x}_0)\mathbf{x}_0 + o(\|\mathbf{x} - \mathbf{x}_0\|^2).$$

Newton method uses the quadratic term (a quadratic function) and takes its minimum as approximation to the minimum of f.

Summary

- Iterative methods define successive approximations to an optimal value of a function.
- Key aspects of iterative methods are convergence, efficiency, stability and robustness.
- Newton method is an iterative method based on second order approximation of sufficiently smooth functions.
- With a good guess for initial point, it has quadratic order of convergence.
- For quadratic functions it converges in one step.
- There are variations of the Newton method:
  - Quasi Newton methods (trying to avoid computation of the Hessian).
  - Levenberg-Marquardt algorithm (modification when Hessian is not definite in nonlinear least squares minimization)
  - ....