Unconstrainted optimization

First order necessary condition and second order sufficient condition to check for the local maximum or local minimum:

$$f : R^n \to R$$

$$\max_{x \in R^n} f(x) \quad or \quad \min_{x \in R^n} f(x)$$

Theorem 1 **(First order necessary condition)**

Let f be continuously differentiable.

If x^* is a local maximizer or minimizer of f, then

$$\boxed{g(x^*) = \nabla f(x^*) = 0.}$$

We call x^* a stationary point if $g(x^*) = 0$.

Sufficient conditions for local maximum & minimum

a) **If** $g(x^*) = 0$ and $H(x^*)$ is negative definite, then

$$f(x^* + x) < f(x^*), \quad \text{for all } x \in B_\varepsilon(x^*).$$

x^* is a local maximizer.

b) **If** $g(x^*) = 0$ and $H(x^*)$ is positive definite, then

$$f(x^* + x) > f(x^*), \quad \text{for all } x \in B_\varepsilon(x^*).$$

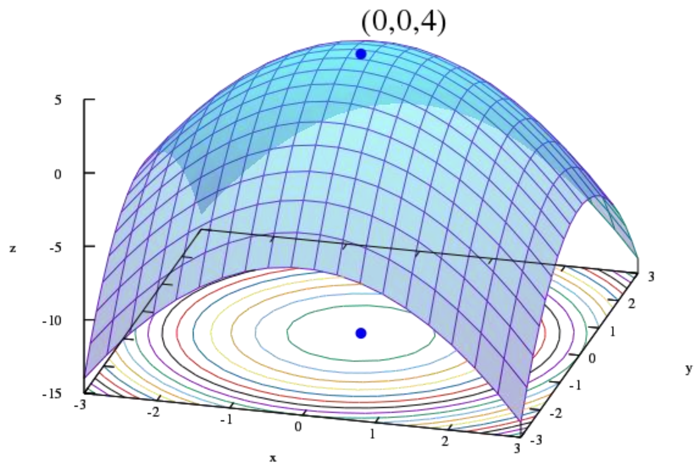
x^* is a local minimzer.

where the **gradient** $\nabla f^T(x^*)$ is the vector

$$g(x^*)^T = \nabla f(x^*)^T = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}^T (x^*),$$

and the **Hessian** $H(x^*)$ is the symmetric matrix

$$H(x^*) = \nabla^2 f(x^*) = \begin{bmatrix} \frac{\partial f^2}{\partial x_1^2} & \frac{\partial f^2}{\partial x_1 \partial x_2} & \cdots & \frac{\partial f^2}{\partial x_1 \partial x_n} \\ \frac{\partial f^2}{\partial x_2 \partial x_1} & \frac{\partial f^2}{\partial x_2^2} & \cdots & \frac{\partial f^2}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f^2}{\partial x_n \partial x_1} & \frac{\partial f^2}{\partial x_n \partial x_2} & \cdots & \frac{\partial f^2}{\partial x_n^2} \end{bmatrix} (x^*).$$



Graph of a paraboloid given by $z = f(x, y) = -(x^2 + y^2) + 4$.

The global maximum at $(x, y, z) = (0, 0, 4)$ is indicated by a blue dot.

In practice, there is no straight forward method to solve system of nonlinear equation. Therefore, there are many iterative methods developed to approximate the solution to system of nonlinear equation. Some methods to be introduced are Newton's method and fixed point method

$$g(x) = 0. \qquad (1)$$

The procedure is start with an initial guessing point $\mathbf{X_0}$ for the solution of (1).

Then the iterative algorithm will generate sequence of $\mathbf{X_1, X_2, X_3}$, …. so on

And hopefully the sequence converge to the limit that give $g(\mathbf{X_k}) = 0$.

Usually, for most iterative algorithm, the iteration process will stop when either one of the following satisfied.

(*i*) $\| x_{k+1} - x_k \| \leq \varepsilon$

(*ii*) $\| x_{k+1} - x_k \| \leq \varepsilon \| x_k \|$

(*iii*) $\| g(x_k) \| \leq \varepsilon$,

where $\varepsilon > 0$ is a given small value, which is called a stopping constant.

Worked Example 1: Fixed point method

Solve $g(x) = x^2 - 3x + 2 = 0$

Transform the problem by setttig x as the subject and the other side as h(x)

$$x = h(x) = \frac{x^2 + 2}{3}$$

Apply the fixed point method with $x_{k+1} = h(x_k)$ and initial starting point to be $x_0 = 0.0, 1.5, 1.6, 3.0$

$$x_{k+1} = (x_k^2 + 2)/3$$

x_0	0.0	1.5	1.6	3.0
x_1	0.6667	1.4167	1.5200	3.6667
x_2	0.8148	1.3356	1.4368	5.1481
x_3	0.8880	1.2613	1.3548	9.5011
x_4	0.9295	1.1970	1.2785	30.7572

We see that for the first three starting point, the sequence converges to 1.

***Theorem 1*: Convergence of fixed point methods**

Let x^* be a solution of $x = h(x)$ in R.

If $h(x)$ has a continuous derivative with $|h'(x)| \le \lambda < 1$ in an interval $I = [x^* - d, x^* + d] (d > 0)$, then

(i) $x_{k+1} = h(x_k)$ converges to x^* for any x_0 in I;

(ii) x^* is the unique solution of $x = h(x)$ in I;

(iii) x_k satisfies

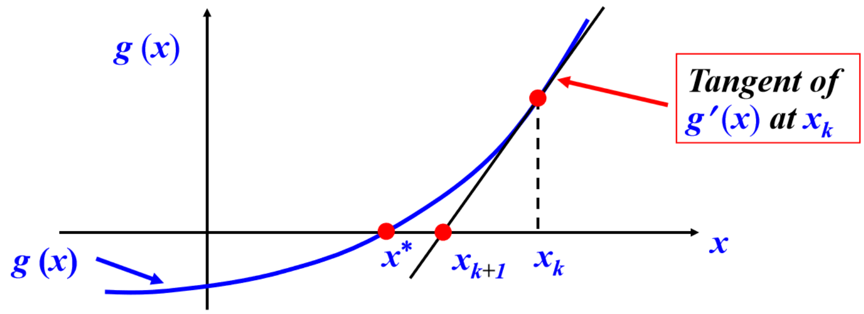
$$\frac{\varepsilon_k}{1+\lambda} \le |x_k - x^*| \le \frac{\varepsilon_k}{1-\lambda} \le \frac{\lambda^k \varepsilon_0}{1-\lambda},$$

where $\varepsilon_k = |x_{k+1} - x_k|$.

Worked Example 2: Newton method

$$x_{k+1} = h(x_k) = x_k - \frac{g(x_k)}{g'(x_k)}, \quad k = 0, 1, 2, \cdots$$

This method approximates $g'(x)$ as a straight line at x_k and obtains a new x_{k+1}, which is used to approximate the function at the next iteration.



This is carried on until the new point is sufficiently close to x^*.

Solve $g(x) \ = \ 9 - x(x-10) = 0,$

$g'(x) \ = \ -2x + 10$

Apply the Newton's method with $x_{n+1} = x_n - \frac{9 - x_n(x_n - 10)}{-2x_n + 10}$ and initial starting point to be $x_0 = 10.0$

Solve: $g(x) \ = \ 9 - x(x-10) = 0$

$g'(x) \ = \ -2x + 10$

```
-3.6333333333333337
-1.285778635778636
-0.8474070849064451
-0.8309750481229856
-0.8309518948912684
-0.8309518948453006
-0.8309518948453005
-0.8309518948453005
-0.8309518948453005
```

N - Dimensional Fixed point method

Generally, a nonlinear system for $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T$ has the form

$$\boldsymbol{g}(\boldsymbol{x}) = \begin{bmatrix} g_1(x_1, x_2, \ldots, x_n) \\ g_2(x_1, x_2, \ldots, x_n) \\ \vdots \\ g_n(x_1, x_2, \ldots, x_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = 0$$

To use the fixed point method, we reformulate $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{0}$ into

$$\boldsymbol{h}(\boldsymbol{x}) = \begin{bmatrix} h_1(x_1, x_2, \ldots, x_n) \\ h_2(x_1, x_2, \ldots, x_n) \\ \vdots \\ h_n(x_1, x_2, \ldots, x_n) \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \boldsymbol{x},$$

Fixed point method $\quad \boldsymbol{x}_{k+1} = \boldsymbol{h}(\boldsymbol{x}_k).$

Newton's method

The ***Newton's method*** is usully performed as the following procedure:

Given $\boldsymbol{x}_0$, for $k = 0, 1, \cdots$

1) **Solve** the linear system

$$\boldsymbol{H}(\boldsymbol{x}_k)\Delta\boldsymbol{x}_k = -\boldsymbol{g}(\boldsymbol{x}_k) \; for \; \Delta\boldsymbol{x}_k:$$

2) **Put** $\quad \boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \Delta\boldsymbol{x}_k:$

3) **Increase** k to $k + 1$ and trun to *1)*.