Optimality condition for quadratic problems

$\nabla f(x) = Qx + g$

$Hf(x) = Q$

$$\inf_{x \in \mathbb{R}^n} f(x) = \frac{1}{2}\langle x, Qx\rangle + \langle g, x\rangle + c$$

for x eigenvector of $\lambda < 0$

where Q is a symetric $n \times n$ matrix, $g \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

1 ▶ If Q is not positive semi definite then the problem has no solution : no $x \in \mathbb{R}^n$ realizes a local minimum.

$f(x) \to -\infty$ for some x

2 ▶ If Q is positive definite then $x^\star = -Q^{-1}g$ is the only global minimum.

$\nabla f(x^*) = 0$

3 ▶ if $Q \geq 0 \iff \exists \lambda = 0 \Rightarrow v / Qv = 0$

$\{\alpha v, \alpha \in \mathbb{R}\}$ $f(\alpha v) = \langle g, \alpha v\rangle + C = \alpha\langle g, v\rangle + C$

if $\langle g, v\rangle = 0$ $f(\alpha v) = C$ if not $f(\alpha v) \to -\infty$

Solving systems of non linear equations

a necessary condition for $x^* = \underset{x \in E}{\operatorname{argmin}} f(x)$

$\nabla f(x^*) = 0$

Fixed point method

$\psi(x) = g(x) - x$

Definition

A fixed point $x \in \mathbb{R}^N$ of a function $g : \mathbb{R}^N \to \mathbb{R}^N$ is a point such that $x = g(x)$ $\iff$ $g(x) - x = 0$

Definition

A fixed point $x \in \mathbb{R}^N$ of a function $g : \mathbb{R}^N \to \mathbb{R}^N$ is said to be attractive if there exists a neighborhood V of x such that for all x_0 in V, the sequence defined by $x_{n+1} = g(x_n)$ converges to x. Otherwise, the point is said to be repulsive.

if $x_n \to \ell$ (x_n converges to a limit)

at the limit $x_{n+1} = g(x_n)$

$\ell = g(\ell)$

Picard's Theorem

contractanti

Theorem (Picard's Theorem)

Let F be a closed ~~state~~ subset of $\mathbb{R}^N$ and let $g : F \subset \mathbb{R}^N \to \mathbb{R}^N$ be a map such that $g(F) \subset F$. We assume that g is contracting, i.e. there exists $k \in]0,1[$ such that:

$$\forall x, y \in F, \qquad \|g(x) - g(y)\| \leq k\|x - y\|. \tag{2}$$

Then there exists a unique $x^* \in F \subset \mathbb{R}^N$ such that $g(x^*) = x^*$ and, for all $x_0 \in F$, the sequence defined by $x_{n+1} = g(x_n)$ converges to x^* (i.e. $x^\star$ is an attractive fixed point). Furthermore, there exists a constant C (depending on the choice of x_0 and the function g) such that

$$\|x_n - x^*\| \leq Ck^n. \tag{3}$$

$C = \|x_1 - x_0\|$
$= \|g(x_0) - x_0\|$

Fixed point algorithm

$H: \exists k \in]0,1[\; \|g(x)-g(y)\| \le k\|x-y\|$

1) $x_{n+1} = g(x_n)$ converges

(x_n) de cauchy : $\|x_{n+1} - x_n\| = \|g(x_n) - g(x_{n-1})\|$

$\le k\|x_n - x_{n-1}\| \le k^{n+1}\|x_1 - x_0\|$

$n > p$

$\|x_n - x_p\| = \|x_n - x_{n-1} + x_{n-1} - x_{n-2} + \dots - x_{p+1} \cdot x_p\|$

$\le \sum_{i=0}^{n-p-1} \|x_{p+i+1} - x_{p+i}\| \le \sum_{i=0}^{n-p-1} k^{p+i+1}\|x_1 - x_0\| = k^{p+1}\|x_1 - x_0\| \sum_{i=0}^{n-p-1} k^i$

$\le \|x_1 - x_0\| k^{p+1} \frac{1-k^{n-p}}{1-k} \xrightarrow[n>p]{n,p\to\infty} 0$ $\quad (x_n)$ Cauchy $\Rightarrow x_n \xrightarrow[n\to\infty]{} \ell$

$\ell = g(\ell)$ ℓ is a fixed point $\quad \ell = x^* = g(x^*)$

$\|x_n - x^*\| = \|g(x_{n-1}) - g(x^*)\| \le k\|x_{n-1} - x^*\|$

$\le k^n \|x_1 - x_0\|$

$g(x) = 0$, scalar case

$\left(g(x) = f'(x) \right)$

Zeros of function $g : \mathbb{R} \longrightarrow \mathbb{R}$

$g(x) = g(x^*) + g'(x^*)(x - x^*) + o(\|x - x^*\|)$

$$0 = \cancel{g(x^\star)} = g(x) + g'(x)(x^\star - x) + \cancel{o(\|x^\star - x\|)}. \leftarrow$$

factor $g'(x) \neq 0$ $\quad g'(x)\left(\frac{g(x)}{g'(x)} + x^* - x \right) = 0$

Fixed point algorithm to solve a nonlinear equation

$\psi(x) = x - \frac{g(x)}{g'(x)} = x^* -$

$$\psi(x) = x, \text{ with } \psi(x) = x - g(x)/g'(x).$$

if $\psi(x) = x$ has a solution it converges to x^*

Approximation by a sequence x_n

$$x_{n+1} = x_n - g(x_n)/g'(x_n) = \psi(x_n)$$

Newton algorithm (scalar case)

equation of the tangent at $(x_0, g(x_0))$

$y = g(x_0) + g'(x_0)(x - x_0)$

$0 = g(x_0) + g'(x_0)(x - x_0)$

$\Leftrightarrow \quad x = x_0 - \frac{g(x_0)}{g'(x_0)} = x_1$

$|x_n - x_{n-1}| > \varepsilon_1$

program program

Data: Function $g(x)$, derivative $g'(x)$, tolerance ε, max number of iterations $k_{\max}$

Result: $x^\star$ such that $g(x^\star) = 0$ $\Rightarrow$ numerically $|g(x_k)| \le \varepsilon$

Initialisation : $k = 0$, x_0 initial guess for $g(x_0) \neq 0$.

while $\boxed{|g(x_k)| > \varepsilon}$ *and* $k \le k_{\max}$ **do**

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

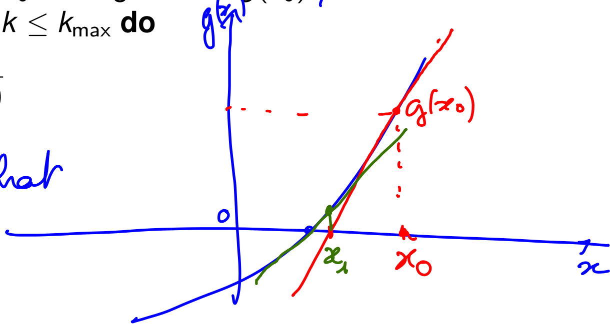
$$k \leftarrow k + 1$$

end

$x^\star \leftarrow x_k$

check that $|g(x^\star)| \le \varepsilon$

otherwise print a warning



Convergence of Newton algorithm

fixed point we had $\|x_k - x^*\| \leq C k^p$

Let g in C^2 on $I = [x^\star - r, x^\star + r]$ with $g(x^\star) = 0$ and $g' \neq 0$ on I. Let

$$M = \max_{x \in I} \left| \frac{g''(x)}{g'(x)} \right|, \quad \text{and } h = \min\left(r, \frac{1}{M}\right).$$

h cannot be computed in practice

Then for any $x_0 \in]x^\star - h, x^\star + h[$ we have

$$\boxed{|x_k - x^\star| \leq \frac{1}{M}\left(M|x_0 - x^\star|\right)^{2k},}$$

quadratic speed of convergence

from which we deduce $\lim_{k \to +\infty} |x_k - x^\star| = 0$.

Convergence speed (not specific to Newton)

suppose algorithm building a convergent sequence $x^{k+1} = \phi(x^k)$

$x^k \to x^\star$

Denote by $e_k = x^k - x^\star$ the error at iteration k. We say that

- the algorithm converges if $\lim_{k\to\infty} \|e_k\| = 0$
- the algorithm converges linearly if $c \in]0,1[$ tel que $\|e_k\| \leq c\|e_{k-1}\|$ for $k > K(c)$
- the algorithm converges supra-linearly if $(c_k)_{k\in\mathbb{N}}$ with $\lim_{k\to\infty} c_k = 0$ such that $\|e_k\| \leq c_k\|e_{k-1}\|$
- the algorithm converges geometrically if the sequence c_k is geometric
- the ~~algorithme~~ algorithm is of order p if there exists $c \in]0,1[$ such that $\|e_k\| \leq c\|e_{k-1}\|^p$ for $k > K(c)$

The convergence can be global or local

$\|e_k\| \leq c\,\|e_{k-1}\|^2$

for Newton quadratic or order 2

Newton method in dimension n

$x^* \in \mathbb{R}^n$

$\begin{cases} G_1(x^*) = 0 \\ \vdots \\ G_n(x^*) = 0 \end{cases}$

$G(x) = 0$ with $G : \mathbb{R}^n \longrightarrow \mathbb{R}^n$.
Let $JG(x) \in \mathbb{R}^{n \times n}$ the jacobian matrix of G in x,
$JG_{i,j}(x) = \dfrac{\partial G_i(x)}{\partial x_j}$.

$\boxed{\dfrac{1}{g'(x)} \longleftrightarrow JG(x)^{-1}}$

$$G(x^\star) = G(x) + JG(x)(x - x^\star) + o(||x - x^\star)||.$$

scalar case $\varphi(x) = x - g(x)/g'(x)$

Fixed point algorithm to solve a nonlinear equation

$$\Psi(X) = X, \text{ with } \Psi(x) = X - JG(X)^{-1}G(X).$$

Approximation by a fixed point sequence $X_{n+1} = \Psi(X_n)$

each $X_k \in \mathbb{R}^n$

$$X_{k+1} = X_k - JG(X_k)^{-1}G(X_k)$$

Except that in practice for n large, one never computes the inverse of a matrix

major rule in scientific computing

Newton-Ralphson algorithm

program flees 2 functions

Data: Function $G(x)$, jacobian matrix $JG(x)$, tolerance ε,
max number of iterations $k_{\max}$

Result: $x^\star$ such that $G(x^\star) = 0$

Initialisation : $k = 0$, $x_0 \in \mathbb{R}^n$

while $\|G(x_k)\| > \varepsilon$ *and* $k \leq k_{\max}$ **do**

  Solve $JG(x_k)d_k = -G(x_k)$

  $x_{k+1} = x_k + d_k$

  $k \leftarrow k + 1$

end

$x^\star \leftarrow x_k$

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

$X_{k+1} = X_k - JG(x_k)^{-1} G(X_k)$

numpy.linsolve

linear system numpy

Convergence of Newton-Ralphson algorithm

theoritical result. in practice the hard point is to choose x_0

Suppose :

$g'(x^*) \neq 0$

- G of class C^2
- $G(x^\star) \cancel{\neq} 0$
- the tangent linear map $JG(x^\star) \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^m)$ is inversible.

Then $x^\star$ is a superattractive fixed point of

$$\Psi(x) = x - (JG(x))^{-1} G(x).$$

$\exists V(x^*)$ s.t. $x_0 \in V(x^*)$

$x_{n+1} = \psi(x_n)$ $\quad x_n \to x^*$

$\|x_n - x^*\| \leq C k^{2n}$

Scalar case : the secant method

Data: Function $g(x)$, tolerance ε, max number of iterations $k_{\max}$

Result: $x^\star$ such that $g(x^\star) = 0$

Initialisation : $k = 0$, x_0 initial guess for $g(x) = 0$.

a_0 initial guess for $g'(x_0)$ (default =1)

while $|g(x_k)| > \varepsilon$ *and* $k < k_{\max}$ **do**

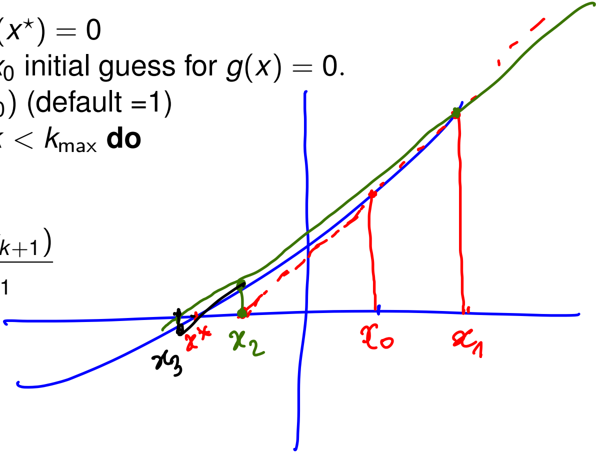
$$x_{k+1} = x_k - \frac{g(x_k)}{a_k}$$

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

$k \leftarrow k + 1$

end

$x^\star \leftarrow x_k$



Vector case : the quasi-Newton method

$x_{k+1} = x_k - JG(x_k)^{-1} G(x_k)$

$A_k^{-1} \; G(x_k)$

Data: $G : \mathbb{R}^n \longrightarrow \mathbb{R}^n$

$\varepsilon > 0$.

Result: $x^\star$ such that $g(x^\star) = 0$

Initialisation : a first approximation of $x_0 \in \mathbb{R}^n$

$A_0 \approx J(x_0)$ or $W_0 \approx J(x_0)^{-1}$ → start with $A_0 = Id$

$x_1 = x_0 - W_0 G(x_0)$

$d_0 = x_1 - x_0$,

$y_0 = G(x_1) - G(x_0)$,

$k = 1$

while $\|G(x_k)\| > \varepsilon$ *and* $k < k_{\max}$ **do**

Update : $W_k = W_{k-1} + B_{k-1}$

Compute d_k solution of $d_k = -W_k G(x_k)$

$x_{k+1} = x_k + d_k$

$y_k = G(x_{k+1}) - G(x_k)$

$k \leftarrow k + 1$

end

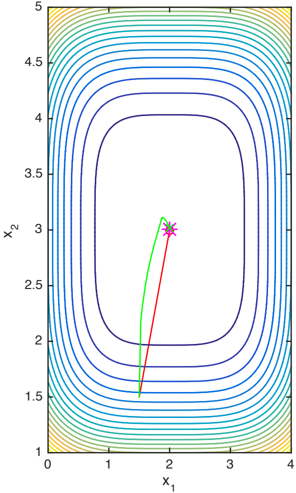
$x^\star \leftarrow x_k$

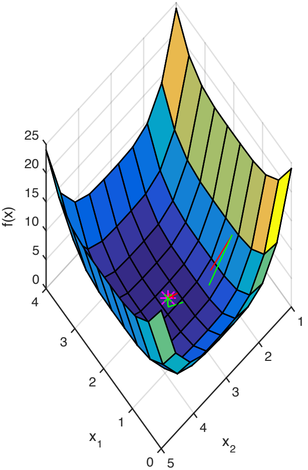
Comparison of Newton and quasi Newton methods

Minimum of the quadratic function $f(x) = ((x_1 - 2)^4 + (x_2 - 3)^4)/2$

Newton method : 12 iterations

quasi Newton (BFGS) method : 21 iterations





Update in the quasi Newton method

$W_k \simeq JG(x_k)^{-1}$
similar rules for
$A_k \simeq JG(x_k)$

Update :

$\boxed{W_k = W_{k-1} + B_{k-1}}$

Compute d_k solution of $d_k = -W_k G(x_k)$

$x_{k+1} = x_k + d_k$

Conditions on the W_k matrix

1. W_k should remain symetric positive definite for all k.
2. The quasi-Newton equation $W_k y_k = d_k$ is satisfied for each k
3. The difference between two consecutive approximations $W_{k+1} - W_k$ is minimum in some sense (for some norm), for example for the Frobenius norm

Examples of update rules satisfying the conditions

$$W_k = W_{k-1} + B_{k-1}$$
$$d_k = -W_k G(x_k)$$
$$x_{k+1} = x_k + d_k$$
$$y_k = G(x_{k+1}) - G(x_k)$$

- The Davidon-Fletcher-Powell method

$$(DFP) \quad W_{k+1} = W_k + \frac{d_k d_k^T}{\langle y_k, d_k \rangle} - \frac{W_k y_k y_k^T W_k}{\langle y_k, W_k y_k \rangle}.$$

- The Broyden-Fletcher-Goldfarb-Shanno method

$$(BFGS) \qquad W_{k+1} = W_k - \frac{d_k y_k^T W_k + W_k y_k d_k^T}{\langle y_k, d_k \rangle} + \left(1 + \frac{\langle y_k, W_k y_k \rangle}{\langle y_k, d_k \rangle}\right) \frac{d_k d_k^T}{\langle y_k, d_k \rangle}.$$

Exercice

Consider the minimisation problem

$$x^\star = \operatorname*{argmin}_{\mathbb{R}^2} f(x), \quad \text{with } f(x) = (x_1 - 1)^4 + (x_2 - 2)^4$$

1. Solve the problem exactly.
2. Write Newton algorithm for this problem. Show that it converges for any $x^0 \in \mathbb{R}^2$.

Consider now the gradient algorithm for this problem. In order to select the step α_k in the gradient direction

$$x^{k+1} = x^k - \alpha_k \nabla f(x^k)$$

we consider the function $\varphi^k(\rho) = f(x^k - \rho \nabla f(x^k))$ and $p^k(\rho)$ polynomial of degree 2 that interpolates $\varphi(0)$, $\varphi'(0)$ and $\varphi''(0)$.

3. Define $p^k(\rho)$ from the values $\varphi(0)$, $\varphi'(0)$ and $\varphi''(0)$
4. Compute $\rho^k = \operatorname{argmin}_{\rho > 0} p^k(\rho)$
5. Show that there exists $\alpha > 0$ s.t. if $\alpha_k = \alpha \rho_k$ then

$$\|x^{k+1} - x^\star\| \leq \mu \|x^k - x^\star\| \text{ with } 0 < \mu < 1$$

6. Conclude that $\lim_{k\to\infty} x^k = x^\star$ and compute the error estimate $\|x^0 - x^\star\|$.

Exercice

$f(x) = (x_1 - 1)^4 + (x_2 - 2)^4$ $\quad f : \mathbb{R}^2 \to \mathbb{R}$

find $f(x^*) = \min f(x)$ $\quad x^* = (1, 2) \Leftarrow (f(x) \geq 0$ and $f(x^*) = 0)$

$G(x) = \nabla f(x)$ $\quad JG(x) = Hf(x)$, write Newton: $X_{k+1} = X_k - JG(X_k)^{-1} G(X_k)$

$$G(x) = \begin{pmatrix} 4(x_1 - 1)^3 \\ 4(x_2 - 2)^3 \end{pmatrix} \qquad JG(x) = \begin{pmatrix} 12(x_1 - 1)^2 & 0 \\ 0 & 12(x_2 - 2)^2 \end{pmatrix}$$

$$JG(x)^{-1} = \begin{pmatrix} \frac{1}{12(x_1 - 1)^2} & 0 \\ 0 & \frac{1}{12(x_2 - 2)^2} \end{pmatrix} \qquad JG(x) \cdot JG(x)^{-1} = Id$$

Compute X_{k+1} knowing X_k

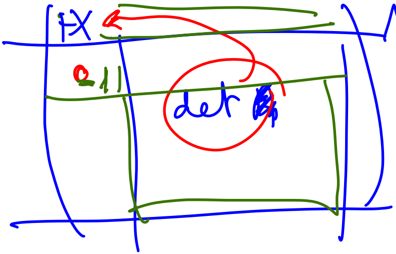
Exercice

inverse of a 2x2 matrix:

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^{-1} = \frac{1}{D} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$$

with $D = a_{11} a_{22} - a_{12} a_{21}$

General Case: use the cofactor:



M^{-1} = transpose of the cofactor matrix divided by the determinant.

Exercice

Exercice