d_1 and d_2 are conjugate if

$d_1 Q d_2 = 0$

$Q = I \Rightarrow d_1 . d_2$ are orthogonal

Suppose : $Q = Q' > 0$ $\Rightarrow$ $d_i^T Q d_j = 0 \quad i \neq j$

: $d_1 , d_2 \dots d_k$ $\quad d_i^T Q d_i > 0$

- The directions of $d_1 \dots d_k$ are linearly independent.

Suppose $d_1 \dots d_k$ such that

$\alpha_1 d_1 + \cdots + \alpha_k d_k = 0$

$(\times Q)$ $\alpha_1 Q d_1 + \alpha_2 Q d_2 + \cdots + \alpha_k Q d_k = 0$

$(\times d_i')$ $\alpha_1 \underbrace{d_i' Q d_1}_{>0} + \alpha_2 \underbrace{d_i' Q d_2}_{=0} + \cdots + \alpha_k \underbrace{d_i' Q d_k}_{=0} = 0$

- There are at most n mutually independent directions

Let

$$f = \frac{1}{2}x'Qx + c'x + d \qquad Q = Q' > 0$$

f has a global min $x^* = -Q^{-1}c$

The conjugate direction method

$$x_0, \underbrace{d_0 \ldots d_{n-1}}_{Q\text{ conjugate}}$$

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

α_k performs a exact line search along d_k

$$\rightarrow \alpha_k = -\frac{\nabla' f(x_k) d_k}{d_k' Q d_k}$$

The sequence $\{x_k\} = \{x_0 \ldots x_n\}$ is such that

$$x_n = x^* = -Q^{-1}c$$

Define a new algorithm such that

$d_0, d_1 \dots d_{n-1}$ are Q-conjugate.

but not a-priori complicated.

$x_0, d_0 = -\nabla f(x_0)$

$x_{k+1} = x_k + \alpha_k d_k$ $\quad\quad$ $\alpha_k = -\frac{\nabla' f(x_k) d_k}{d_k' Q d_k}$ α^*

$d_{k+1} = -\nabla f(x_{k+1}) + \beta_k d_k$ $\quad$ pick d_k, d_{k+1} Q-conjugate

select β_k such that $\overbrace{d_k' Q d_{k+1}}^{\text{Q-conjugate}} = 0$

$d_k' Q d_{k+1} = -d_k' Q \nabla f(x_{k+1}) + \beta_k d_k' Q d_k$

$$\beta_k = \frac{d_k' Q \nabla f(x_{k+1})}{d_k' Q d_k}$$

$x_0 \cdot d_0 = -\nabla f(x_0)$

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

$d_{k+1} = -\nabla f(x_{k+1}) + \beta_k d_k$

exact line search

$\alpha_k = -\dfrac{\nabla f(x_k) d_k}{d_k' Q d_k}$

$\beta_k = \dfrac{\nabla f(x_{k+1})' Q d_k}{d_k' Q d_k}$

Q-conjugacy

This algorithm is such that, for any x_0, the sequence $\{x_k\}$ converges to $x_0 = -Q^{-1}c$ in at most n step, and the directions $d_0 \ldots d_{n-1}$ are mutually Q-conjugant.

Properties

$\underbrace{d_k' \nabla f(x_{k+1}) = 0}_{\text{result of exact line search}}$

$\nabla' f(x_{k+1}) d_k = 0$

$\nabla' f(x_k) \nabla f(x_k) = -\nabla f(x_k)' d_k$

$d_{k+1} = -\nabla f(x_{k+1}) + \beta_k d_k \quad \times (\nabla f'(x_{k+1}))$

$\nabla f(x_{k+1})' d_{k+1} = -\nabla f(_{k+1})$

$$f = \frac{1}{2} x' Q x + c' x + d \qquad Q = Q' > 0$$

$$\nabla f = Qx + c$$

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

$$Q x_{k+1} = Q x_k + \alpha_k Q d_k$$

$$\underbrace{Q x_{k+1} + c}_{\nabla f(x_{k+1})} = \underbrace{Q x_k + c}_{\nabla f(x_k)} + \alpha_k Q d_k$$

$$Q d_k = \frac{\nabla f(x_{k+1}) - \nabla f(x_k)}{\alpha_k}$$

$$\beta_k = \frac{\nabla f(x_{k+1}) \left[\frac{\nabla f(x_{k+1}) - \nabla f(x_k)}{\alpha_k} \right]}{d_k' \left[\frac{\nabla f(x_{k+1}) - \nabla f(x_k)}{\alpha_k} \right]}$$

$x_0 . d_0 = -\nabla f(x_0)$

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

$d_{k+1} = -\nabla f(x_{k+1}) + \beta_k d_k$

$$\beta_k = \nabla f(x_{k+1})' \frac{[\nabla f(x_{k+1}) - \nabla f(x_k)]}{d_k [\nabla f(x_{k+1}) - \nabla f(x_k)]}$$

α_k is a line search parameter obtained using a sufficiently accurate line search algorithm.

$$= \frac{\nabla f(x_{k+1})' [\nabla f(x_{k+1}) - \nabla f(x_k)]}{-d_k' \nabla f(x_k)}$$

Quasi-Newton

$$x_{k+1} = x_k \underbrace{- [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)}_{d_k} \quad \text{Newton}$$

$$x_{k+1} = x_k - \alpha_k \underbrace{[\nabla^2 f(x_k)]^{-1}}_{??} \nabla f(x_k) \quad + \text{line search}$$

$$x_{k+1} = x_k - \alpha_k H_k \nabla f(x_k)$$

$H_{k+1} = ??$ $\qquad \{H_k\} \to [\nabla^2 f(x_k)]^{-1}$

an approximate, in some sense, $[\nabla^2 f(x_k)]^{-1}_{x = x_k}$

$H_k > 0$ for all k

$$f = \frac{1}{2} x' Q x + c' x + d$$

$$\nabla f = Qx + c \qquad \nabla^2 f = Q$$

$$\nabla f(y) - \nabla f(x) = (Qy + c) - (Qx + c)$$

$$= \underbrace{Q}_{\nabla^2 f}(y - x)$$

$$Q^{-1}[\nabla f(y) - \nabla f(x)] = y - x$$

$$Q^{-1}[\nabla f(x_{k+1}) - \nabla f(x_k)] = (x_{k+1} - x_k)$$

$$\underline{H_{k+1}}[\underbrace{\nabla f(x_{k+1}) - \nabla f(x_k)}^{\sigma_k}] = [\underbrace{x_{k+1} - x_k}^{\delta_k}]$$

$$H_{k+1}\sigma_k = \delta_k$$

n equations in n^2 unknowns
however $H_{k+1} = H_{k+1}'$

The quasi-Newton eqs has several solutions.
Given H_0. we would like an update law for H
like $H_{k+1} = \cdots$

$H_{k+1}\sigma_k = \delta_k$

$H_0 = I$ outer product

$$H_{k+1} = H_k + \frac{\delta_k \delta_k'}{\delta_k' \sigma_k} - \frac{H_k \sigma_k \sigma_k' H_k}{\sigma_k' H_k \sigma_k}$$

inner product

H_0 is symmetric
if H_k is symmetric. then H_{k+1} is also symmetric.

If $H_k > 0$, it is always possible to select α_k such that $H_{k+1} > 0$

$$H_{k+1}\sigma_k = H_k\sigma_k + \frac{\delta_k \cancel{\delta_k' \sigma_k}}{\cancel{\delta_k'\sigma_k}} - \frac{H_k\sigma_k \cancel{\sigma_k' H_k \sigma_k}}{\cancel{\sigma_k' H_k \sigma_k}}$$

$$= H_k\sigma_k + \delta_k - H_k\sigma_k = \delta_k$$

Quasi-Newton (x_0, H_0 are given) $H_0 = H_0' > 0$

$$x_{k+1} = x_k - \alpha_k H_k \nabla f(x_k)$$

$$H_{k+1} = H_k + \frac{\delta_k \delta_k'}{\delta_k'\sigma_k} - \frac{H_k\sigma_k\sigma_k' H_k}{\sigma_k' H_k \sigma_k}$$

If $f = \frac{1}{2} x'Qx + c'x + d$ $\quad Q = Q' > 0$

then $\{x_k\} \to x_* = -Q^{-1}c$ in at most n steps

$\{H_k\} \to Q^{-1}$ in at most n steps

(provided $\alpha_k = \alpha^*$).

For non-quadratic functions under some assumptions. the line-search is suff-accurate & have a global convergence to $x_0 \in \Lambda$ with quadratic/superlinear speed.

Moreover if $\nabla^2 f(x_k) > 0$ then $\{H_k\} \to [\nabla^2 f(x_k)]^{-1}$

Methods without derivatives

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

α_k is selected using parabolic line search.

d_k: $d_0 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$. $d_1 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}$. $d_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}$ . . $d_{n-1} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$

Coordinate directions method

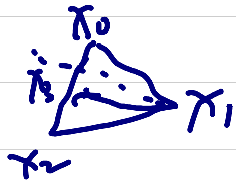
If $L(f(x_0))$ is compact then $\{x_k\}$ is such that $\lim_{k\to\infty}\{x_k\} \in \Omega$

$\{x_k\}$ has a limit if in addition $\lim_{k\to\infty} \|x_{k+1} - x_k\| = 0$

The stopping condition relies on the line search method.

$\min_x f(x) \quad x \in \mathbb{R}^n$

guess/pick 3 points



$f(x_1) > f(x_0) > f(x_2)$

$\{x_0, x_1, x_2\}$

$\downarrow$ projection

$\{x_0, x_2, x_3\}$ $\quad f(x_3) < f(x_1) \Rightarrow$

$\downarrow$

$\{x_2, x_3, x_4\}$

Simplex method: n points in n-dimensional space.

discard the worst.

can cycle between points.