Necessary Conditions

First-order : $\nabla f(x) = 0$

Second-Order : $\nabla^2 f(x)$ positive semidefinite

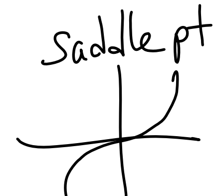
$\nabla^2 f(x) \succeq 0$

<u>Def:</u> $Q \in \mathbb{R}^{n \times n}$ is a positive semidefinite matrix iff $y^T Q y \geq 0$ for all $y \in \mathbb{R}^n$.

"Necessary" conditions because every local min satisfies these conditions.

Not sufficient to determine that a point is a local min.

$f(x) = x^3 \qquad f'(x) = 3x^2 \qquad f''(x) = 6x$



$g(x) = x^4 \qquad g'(x) = 4x^3 \qquad g''(x) = 12x^2$

$h(x) = -x^4 \qquad h'(x) = -4x^3 \qquad h''(x) = -12x^3$

$f'(x) = g'(x) = h'(x) = 0$ when $x = 0$

$f''(x) = g''(x) = h''(x) = 0$ when $x = 0$

$\nabla^2 f(x)$ is positive semidefinite

- Definition may be too hard to check

$$f(x) = 2x_1^2 - \cancel{x_1 x_2} + 8x_2^2$$

$$\nabla f(x) = \begin{bmatrix} 4x_1 \cancel{- x_2} \\ \cancel{-x_1} + 16x_2 \end{bmatrix}$$

$$\nabla^2 f(x) = \begin{bmatrix} 4 & \cancel{-1} \; 0 \\ \cancel{-1} \; 0 & 16 \end{bmatrix}$$

$$\begin{bmatrix} y_1 & y_2 \end{bmatrix} \begin{bmatrix} 4 & \cancel{-1} \; 0 \\ \cancel{-1} \; 0 & 16 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 4y_1^2 \cancel{- 2y_1y_2} + 16y_2^2$$

≥ 0

- Use the different properties of the matrix that guarantee its positive semidefiniteness:

1. all principal minors need to be positive semidefinite
2. all eigenvalues are ≥ 0
3. $Q = LDL^T$ - L is lower triangular, D is diagonal

$D_{ii} \geq 0$ for all i

$Q = LL^T$ - L exists

Sufficient Conditions: If $\nabla f(x) = 0$ and $\nabla^2 f(x)$ is positive definite, then x is a local min (strict local min).

Positive definite: $y^T Q y > 0$ for all $y \in \mathbb{R}^n$, $y \neq 0$.

---

x_* is a local min iff

$$f(x_* + a) - f(x_*) > 0$$

$$a \in (-\varepsilon, \varepsilon)$$

Taylor Series Expansion

$$f(x_* + a) \approx f(x_*) + \nabla f(x_*)^T a + \frac{1}{2} a^T \nabla^2 f(x_*) a$$

$$f(x_* + a) = f(x_*) + \nabla f(x_*)^T a + \frac{1}{2} a^T \nabla^2 f(\xi) a$$

where ξ is some point between x_* and $x_* + a$.

$$f(x_* + a) - f(x_*) \geq 0$$

$$\underbrace{Df(x_*)^T a}_{\overset{\shortparallel}{0}} + \frac{1}{2} \underbrace{a^T D^2 f(\xi)\, a}_{\substack{\succeq 0 \text{ (necessary)} \\ \succ 0 \text{ (sufficient)}}} \geq 0$$

Ex. $f(x) = x_1^4 + 3x_1^2 x_2 + 2x_2^3 + 3x_2^4 + 4x_1 x_2$

$$Df(x) = \begin{bmatrix} 4x_1^3 + 6x_1 x_2 + 4x_2 \\ 3x_1^2 + 6x_2^2 + 12x_2^3 + 4x_1 \end{bmatrix} = 0$$

$x_1 = (0, 0)$

$$D^2 f(x) = \begin{bmatrix} 12x_1^2 + 6x_2 & 6x_1 + 4 \\ 6x_1 + 4 & 12x_2 + 36x_2^2 \end{bmatrix}$$

$$D^2 f(0,0) = \begin{bmatrix} 0 & 4 \\ 4 & 0 \end{bmatrix}$$

$$\det\left(\begin{bmatrix} 0 & 4 \\ 4 & 0 \end{bmatrix} \right) = 0 - 16 = -16 \text{ saddle pt.}$$

Ex. $f(x) = (x_2 + x_1^2)(x_2 + 2x_1^2)$

$f(x) = x_2^2 + 3x_1^2 x_2 + 2x_1^4$

$$Df(x) = \begin{bmatrix} 8x_1^3 + 6x_1 x_2 \\ 2x_2 + 3x_1^2 \end{bmatrix} = 0$$

$2x_2 + 3x_1^2 = 0$

$x_2 = -\frac{3}{2}x_1^2$

$8x_1^3 + 6x_1\left(-\frac{3}{2}x_1^2\right) = 0$

$x_1 = 0, \ x_2 = 0$

$$D^2 f(x) = \begin{bmatrix} 24x_1^2 + 6x_2 & 6x_1 \\ 6x_1 & 2 \end{bmatrix}$$

$$D^2 f(0,0) = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix} \succeq 0 \ (\text{not} \succ 0)$$

Satisfies the necessary conditions

does not satisfy the sufficient conditions

$$[y_1 \ y_2]\begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 2y_2^2 \geq 0$$

Newton's Method

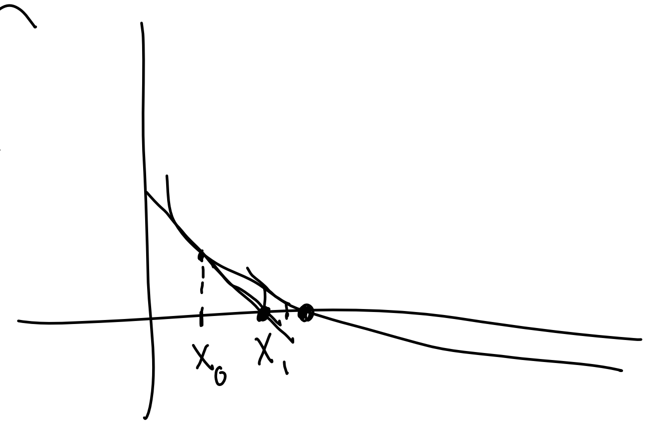
$$f(x+a) \approx f(x) + \nabla f(x)^T a$$

Start at x, find a such that $f(x+a)=0$

$$\boxed{f(x) + \nabla f(x)^T a \approx 0}$$

In one dimension

$$a = -\frac{f(x)}{f'(x)}$$



Instead of $f(x+a)=0$, find $\nabla f(x+a)=0$

$$\nabla f(x) + \nabla^2 f(x)\, a = 0$$

$$\nabla^2 f(x)\, a = -\nabla f(x)$$

Solve this system of equations to find a.

If $\nabla^2 f(x) \succ 0$, $a = -\left[\nabla^2 f(x)\right]^{-1} \nabla f(x)$

Find L and D such that

$$D^2 f(x) = LDL^T$$

(Harwell routines MA27, MA47, MA57)

free   free $

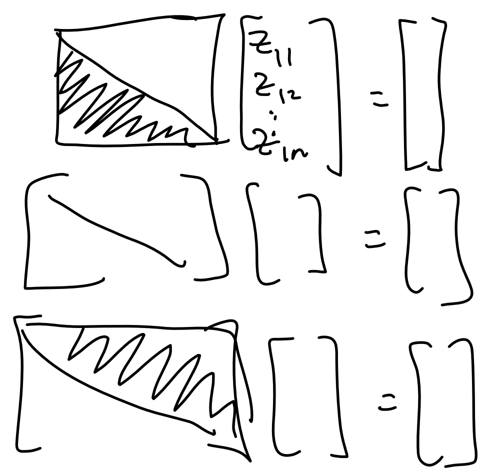
(fortran code)

$$(LDL^T)\, a = -Df(x)$$

→ ① $L z_1 = -Df(x)$

② $D z_2 = z_1$

③ $L^T a = z_2$



---

a is a diection of descent if

$$Df(x)^T a < 0$$

If $a = -Df(x)$

$$\left(Df(x)^T\right)\left(-Df(x)\right) = -\|Df(x)\| < 0$$

Steepest Descent can also be thought of as approximating $D^2 f(x)$ by I.

Quasi-Newton Methods

- Make a "nice" approximation to the Hessian
  - easy to evaluate
  - easy to store
  - easy to invert / factor

Easy to evaluate: Start with the identity and update it at each iteration

Easy to store: Sparse or if dense, use partitions

Easy to invert/factor: Positive definite

BFGS - secant method

B → Broyden
F → Fletcher
G → Goldfarb
S → Shanno

$$\Delta x_k = -\underbrace{B_k^{-1}}_{\text{approximation to the Hessian}} \nabla f(x_k)$$

www.pages.drexel.edu/~hvb22/OPR992