Modified Newton's method

Newton: $d_k = -H(x_k)^{-1} \nabla f(x_k)$

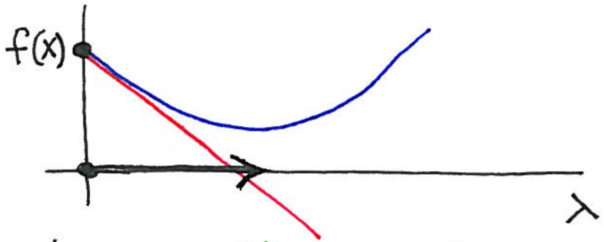
Troubles:
- $H(x_k)$ not invertible,
- d_k not descent direction.

Def $d \overset{\text{def}}{=}$ descent direction at x if $\nabla f(x)^T d < 0$.

$\varphi'(0)$

Remark: $\varphi(\lambda) = f(x + \lambda d)$

$\varphi'(0) < 0$



One can find smaller values of f along d.

Ex SD direction $d = -\nabla f(x)$ is descent.

$$\nabla f(x)^T d = -\|\nabla f(x)\|^2 < 0.$$

Newtons direction is not descent in general, but

B pos. def. $\overset{\text{Ex. A.2}}{\Longrightarrow} -B^{-1}\nabla f$ is descent

We want to modify $H(x_k)$:

- $H_{\varepsilon_k} \approx H(x_k)$
- H_{ε_k} positive definite

The simplest: $\boxed{H_{\varepsilon_k} = H(x_k) + \varepsilon_k \cdot I}$

for some good $\varepsilon_k \geq 0$.

- Start with small and increase.

Modified Newton: $\boxed{d_k = -(H(x_k) + \varepsilon_k I)^{-1} \nabla f(x_k) \text{ + line search}}$

How to: 1) check if $H + \varepsilon I$ is pos.def.?

2) solve $(H + \varepsilon I) d = -\nabla f$?

Some facts about matrices

$A = n\lceil\overline{}^{\,n}$, $b = n\lceil\overline{}^{\,1}$. How to solve $\underline{Ax=b}$?

LinAlg: Gauß elimination!

$$[A|b] \sim [U|\tilde{b}], \quad U = \begin{bmatrix} \diagdown & * \\ 0 & \diagdown \end{bmatrix}$$

What if we need to solve many equations with the same A?

Then it is good to remember the transformations.

Ex. Gauß elimination "without b"

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 3 \\ 0 & & \\ 0 & & * \end{bmatrix} \quad \left(\text{the first step.}\right)$$

"$d_1 \neq 0$"

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}}_{T_1} \underbrace{\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 1 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \\ 0 & -2 & -8 \end{bmatrix}}_{T_1 A}$$

$d_2 \neq 0$

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix}}_{T_2} \underbrace{\begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \\ 0 & -2 & -8 \end{bmatrix}}_{T_1 A} = \underbrace{\begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \\ 0 & 0 & -4 \end{bmatrix}}_{U} = \begin{bmatrix} \diagdown & * \\ 0 & \diagdown \end{bmatrix}$$

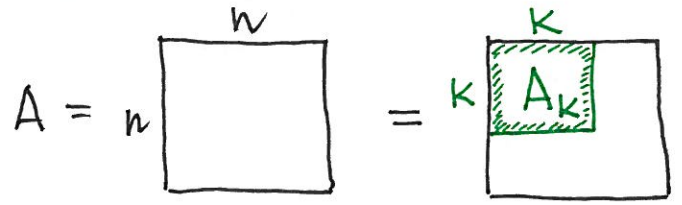
d_k - pivot elements

$$T_2 T_1 A = U \iff A = (T_2 T_1)^{-1} U = T_1^{-1} T_2^{-1} U = L\,U$$

$$L = T_1^{-1} T_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 1 \end{bmatrix}$$

Def. $A = LU \overset{\text{def}}{=}$ LU factorization of A if $L = \begin{bmatrix} 1 & & 0 \\ & 1 & \\ * & & \ddots\ 1 \end{bmatrix}$, $U = \begin{bmatrix} \diagdown & * \\ 0 & \diagdown \end{bmatrix}$

Remark: LU factorization exists if $d_k \neq 0$, $k = 1, 2, \dots, n-1$



<u>Lemma</u>: All $d_k \neq 0$, $k = 1, \dots, n-1$ $\iff$

$\iff$ all $\det A_k \neq 0$, $k = 1, 2, \dots, n-1$

<u>Proof</u>: look at LU factorization (Ex.)

$$\underbrace{\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 1 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 1 \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \\ 0 & 0 & -4 \end{bmatrix}}_{U}$$

$= A_1$, $= A_2$; $= L_1$, $= L_2$; $= U_1$, $= U_2$

We have $A_k = L_k U_k \implies$

$$\implies \det A_k = \underbrace{\det L_k}_{=1} \cdot \det U_k = d_1 \cdot d_2 \cdot \dots d_k$$

$\boxed{\implies}$: trivial (LU exists + above).

$\boxed{\impliedby}$: $d_1 = \det A_1 \neq 0$, $d_2 = \dfrac{\det A_2}{\det A_1} \neq 0$,

$$d_k = \frac{d_1 d_2 \dots d_{k-1} d_k}{d_1 d_2 \dots d_{k-1}} = \frac{\det A_k}{\det A_{k-1}} \neq 0$$

$\blacksquare$

In our example: $A = LU$

$$U = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \\ 0 & 0 & -4 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -4 \end{bmatrix}}_{D} \underbrace{\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}}_{L^T} \implies$$

$$\implies A = L D L^T, \quad D = \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{bmatrix}$$

<u>Remark</u>: possible since $A = A^T$.

<u>Fact 1</u>: if $\det A_k \neq 0$, $k = 1, 2, \dots, n-1$

a) $A = LU$, $L = \begin{bmatrix} 1 & & 0 \\ & 1 & \\ * & \ddots & 1 \end{bmatrix}$, $U = \begin{bmatrix} \diagdown & * \\ 0 & \end{bmatrix}$

b) $A = A^T \implies A = L D L^T$, D - diag.

(see Th. 4, p. 351)

<u>Remark</u>: we did LDL^T-factorization in FlerDim by completing squares in the corresponding quadratic form.

Ex. $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 1 \end{bmatrix}$, $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

$f(x) = x^T A x =$

$= x_1^2 + 4x_1x_2 + 6x_1x_3 + \underbrace{3x_2^2 + 8x_2x_3 + x_3^2}$ $=$

$= (x_1 + 2x_2 + 3x_3)^2 - (2x_2+3x_3)^2 + \text{the rest} =$

$= (x_1 + 2x_2 + 3x_3)^2 - x_2^2 - 4x_2x_3 - 8x_3^2 =$

$= \underbrace{(x_1 + 2x_2 + 3x_3)}_{\hat{x}_1}{}^2 - (\underbrace{x_2 + 2x_3}_{\hat{x}_2})^2 - 4\underbrace{x_3}_{\hat{x}_3}{}^2 =$

$$\left.\begin{aligned} &= \hat{x}^T \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -4 \end{bmatrix} \hat{x} = \hat{x}^T D \hat{x}. \\ &\hat{x} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} x = L^T x \end{aligned}\right\} \Rightarrow$$

$\Rightarrow x^T A x = x^T L D L^T x$, $\forall x \in \mathbb{R}^3 \Rightarrow$

$\Rightarrow$ $A = L D L^T$

Def $H = H^T = n \,\overset{n}{\ulcorner}$

$H \overset{\text{def}}{=}$ positive definite $\Leftrightarrow x^T H x > 0, \forall x \neq 0$.

Fact 2: (Sylvester criterion)

H pos. def. $\Longleftrightarrow \det H_k > 0, \forall k$

(see Th. 7, p. 359)

Idea of proof:

$\boxed{\Rightarrow}$ Let's take $x = \begin{bmatrix} x_1 \\ \vdots \\ x_k \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} z \\ 0 \end{bmatrix} \Rightarrow$

$\Rightarrow x^T H x = z^T H_k z$.

H pos. def. $\Rightarrow z^T H_k z > 0, \forall z \neq 0 \Rightarrow$

$\Rightarrow H_k$ pos. def $\Rightarrow$ eigenvalues of H_k are positive (Th. 3, p. 349) $\Rightarrow \det H_k > 0$.

$\boxed{\Leftarrow}$ Use Fact 1 b) : $H = LDL^T$.

H pos. def. $\Leftrightarrow D$ pos. def. $\Leftrightarrow \underset{\forall k}{d_k} = \dfrac{\det H_k}{\det H_{k-1}} > 0$

$\Rightarrow \det H_k > 0, \forall k$.

Remark: in Ch. 6 we will need to check if H is positive semidefinite

Def $H = H^T \overset{def}{=}$ pos. semidefinite if

$$x^T H x \geq 0, \quad \forall x \geq 0$$

Ex Zero matrix is pos. semidefinite, but not pos. def.

NB Sylvester does not work for pos. semidef.

Ex. $H = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix} \Rightarrow \det H_1 = \det H_2 = 0$,

but H is not pos. semidefinite.

Thus H pos. semidef. $\cancel{\Longleftrightarrow}$ $\det H_k \geq 0, \forall k$

However, it is true that

$$\boxed{H \text{ pos. semidef.} \Longleftrightarrow H + \varepsilon I \text{ pos. def.}, \forall \varepsilon > 0}$$

Fact 3: (Cholesky factorization)

H pos. def. $\Longleftrightarrow H = \hat{L}\hat{L}^T$, $\hat{L} = \begin{bmatrix} \ast & 0 \end{bmatrix}$ (lower triangular) and $\hat{L}_{ii} > 0, \forall i$

(see Th. 9, p. 361)

Idea of proof:

$\boxed{\Leftarrow}$ easy: $x^T H x = x^T \hat{L} \cdot \hat{L}^T x = \|\hat{L}^T x\|^2 \geq 0$.

Moreover, if $x^T H x = 0 \Rightarrow \|\hat{L}^T x\| = 0 \Rightarrow$

$\Rightarrow \hat{L}^T x = 0 \Rightarrow x = (\hat{L}^T)^{-1} 0 = 0$.

(invertible, since $\det \hat{L}^T = \hat{L}_{11} \cdot \hat{L}_{22} \cdot \ldots \cdot \hat{L}_{nn} \neq 0$)

$\boxed{\Rightarrow}$ Use Fact 1 b): $H = LDL^T$, $d_k > 0$.

Define $D^{1/2} = \begin{bmatrix} \sqrt{d_1} & & & \\ & \sqrt{d_2} & & \\ & & \ddots & \\ & & & \sqrt{d_n} \end{bmatrix}$. Then

$D = D^{1/2} \cdot D^{1/2} \Rightarrow H = L D^{1/2} \cdot D^{1/2} L^T = \hat{L}\hat{L}^T$.

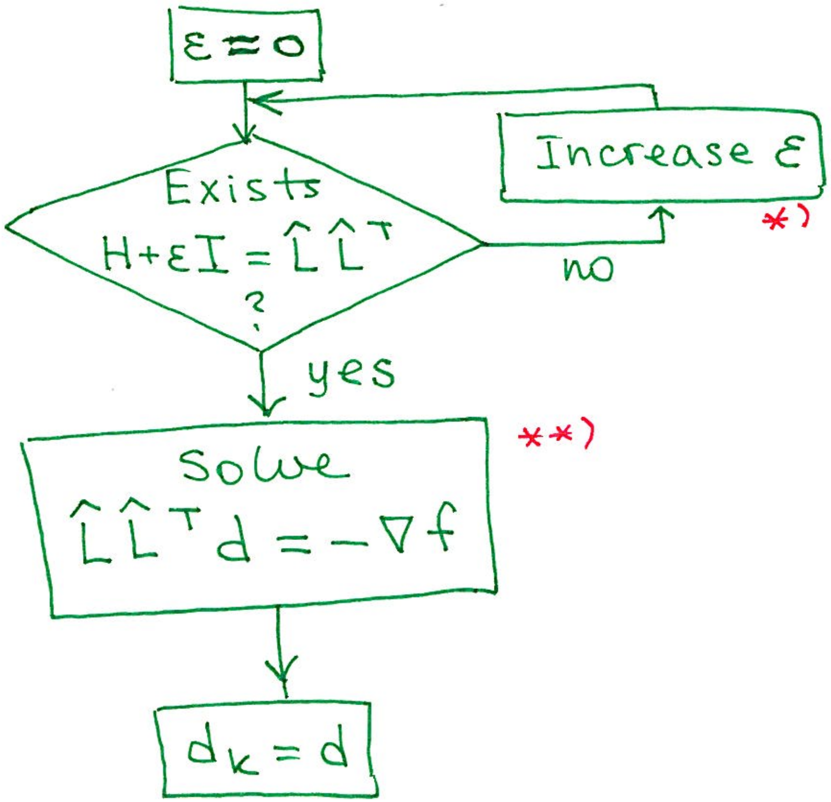
and $\hat{L}_{ii} = L_{ii} \cdot \sqrt{d_i} = 1 \cdot \sqrt{d_i} > 0$. ∎

Back to Modified Newton's method:

- $d_k = -(H(x_k) + \varepsilon_k I)^{-1} \nabla f(x_k)$

  where $H(x_k) + \varepsilon_k I$ pos. def.
  (ε_k not very large)
- line search on λ_k, $x_{k+1} = x_k + \lambda_k d_k$

At each step k: to find ε_k



Remark:

*) to increase ε:

- first time: ε = some small number
- next time: $\varepsilon := \varepsilon \times 4$.

**) to solve $\hat{L}\hat{L}^T d = -\nabla f$:

$$\hat{L}\underbrace{\hat{L}^T d}_{y} = -\nabla f$$

a) Solve $\hat{L} y = -\nabla f$.

The matrix is triangular $\Rightarrow$
$\Rightarrow$ No Gauß eliminations.

b) solve $\hat{L}^T d = y$.

The matrix is again triangular $\Rightarrow$
$\Rightarrow$ no eliminations, just substitutions.