Lecture notes: Newton

Function $F: \mathbb{R}^N \to \mathbb{R}^N$ want to solve

$$F(x^*) = 0$$

Suppose we guess $x^{(0)}$. If $N=1$,

$$F(x^*) = \underbrace{F(x^{(0)}) + F'(x^{(0)})(x^* - x^{(0)})}_{T_F^1(x^*)} + \dots$$

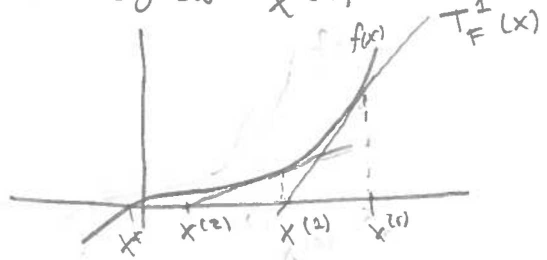
Linear problem: solve $T_F^1(x^*) = 0$

$N>1$: $T_F^1(x^*) = F(x^{(0)}) + J_F(x^{(0)})\underbrace{(x^* - x^{(0)})}_{\delta x}$

Solve $J_F(x^{(0)})\,\delta x = -F(x^{(0)})$

new iterate: $x^{(1)} = x^{(0)} + \delta x$

Repeat with guess $x^{(1)}$



10 min

$$\begin{cases} J_F(x^{(k)})\,\delta x = F(x^{(k)}) \\ x^{(k+1)} = x^{(k)} + \delta x \end{cases} \qquad (1)$$

<u>Preliminaries</u> Lemma 1

Let $B \in \mathbb{R}^{N\times N}$. If $\|B\| < 1$, then $I-B$ is invertible and $\|(I-B)^{-1}\| \leq \frac{1}{1-\|B\|}$

Proof: Define $S = \lim_{M\to\infty} I + \sum_{k=1}^{M} B^k$

$\|S\| \leq \|I\| + \sum_{k=1}^{\infty} \|B^k\| \leq \|I\| + \sum_{k=1}^{\infty} \|B\|^k < \infty$ since $\|B\| < 1$.

So S is well-defined

$$(I+B)S = \lim_{M\to\infty} (I-B)\left(I + \sum_{k=1}^{M} B^k\right) =$$

$$= \lim_{M\to\infty} \left(I + \sum_{k=1}^{M} B^k - B - \sum_{k=1}^{M} B^{k+1}\right) =$$

$$= \lim_{M\to\infty} \left(I + \sum_{k=1}^{M} - \sum_{k=1}^{M+1} B^k\right) = \lim_{M\to\infty} I - B^{M+1}$$

Since $\lim_{M\to\infty} \|B^{M+1}\| \leq \lim_{M\to\infty} \|B\|^{M+1} = 0$,

$$(I-B)S = I \Rightarrow S = (I-B)^{-1}$$

10 min

$$\|(I-B)^{-1}\| = \|S\| \leq \|I\| + \sum_{k=1}^{\infty} \|B\|^k = \sum_{k=0}^{\infty} \|B\|^k = \frac{1}{1-\|B\|}$$

$\square$

<u>Prel. 2</u> Lemma 2

We can write

$$F(y) - F(x) = \int_0^1 J_F(x + t(y-x))(y-x)\, dt$$

Proof: Define $g(t) = F(x + t(y-x))$

Then $F(y) - F(x) = g(1) - g(0) = \int_0^1 g'(t)\, dt$

Chain rule says:

$$\frac{dg}{dt} = \frac{d}{dt} F(x+t(y-x)) = J_F(x+t(y-x))(y-x)$$

$\square$

10 min

Theorem: Quadratic convergence of Newton's method

Let $F: \mathbb{R}^N \to \mathbb{R}^N$ be a C^1 function in an open convex set D and $\exists x^* \in D$ such that $F(x^*) = 0$. Suppose $J_F(x^*)^{-1}$ exists and $\|J_F(x^*)\|^{-1} \leq C$ for some $C > 0$. Suppose further $\exists R, L > 0$ s.t.

$$\|J_F(x) - J_F(y)\| \leq L\|x-y\|, \quad \forall x, y \in B(x^*, R).$$

Then, $\forall x^{(0)} \in B(x^*, r)$, where

$$r = \min\{R, \frac{1}{2CL}\}$$

the sequence defined by (1) is uniquely defined and converges to x^*, with

$$\|x^{(k+1)} - x^*\| \leq CL\|x^{(k)} - x^*\|^2.$$

10 min

Proof:

Step 1) We first show that $J_F^{-1}(x^{(0)})$ exists.

Denote $A_0 = J_F(x^{(0)})$, $A_* = J_F(x^*)$

What if we use A_*^{-1} as an inverse to A_0?

We want $I - A_*^{-1}A_0 \approx 0$

Define $B = I - A_*^{-1}A_0 = A_*^{-1}(A_* - A_0)$

Then

$$\|B\| \le \|A_*^{-1}\| \, \|A_* - A_0\| =$$

$$= \|A_*^{-1}\| \, \|J_F(x^*) - J_F(x^{(0)})\| \le$$

$$\le \underbrace{\|J_F^{-1}(x^*)\|}_{\le C} L \underbrace{\|x^* - x^{(0)}\|}_{\le r = \min\{R, \frac{1}{2CL}\}} \le CL \frac{1}{2CL} = \frac{1}{2}$$

By Lemma 1, $I - B$ is invertible

$$I - B = I - (I - A_*^{-1}A_0) = A_*^{-1}A_0$$

10 min

Since $0 \ne \det(I-B) = \det(A_*^{-1})\det(A_0)$,

$\det(A_0) \ne 0 \Rightarrow A_0$ invertible.

Also, $(I-B)^{-1} = (A_*^{-1}A_0)^{-1} = A_0^{-1}A_* \Rightarrow$

$$\Rightarrow A_0^{-1} = A_*^{-1}(I-B)^{-1}$$

Therefore,

$$\|A_0^{-1}\| \le \underbrace{\|A_*^{-1}\|}_{\le C} \underbrace{\|I-B\|^{-1}}_{\le \frac{1}{1-\|B\|}} \le \frac{C}{1-\|B\|} \le \frac{C}{1-\frac{1}{2}} = 2C$$

Step 2)

$$\begin{cases} J_F(x^{(0)})\delta x = -F(x^{(0)}) \\ x^{(1)} = x^{(0)} + \delta x \end{cases}$$

$$x^{(1)} = x^{(0)} + \delta x = x^{(0)} - J_F(x^{(0)})^{-1}F(x^{(0)})$$

Look at error:

$$x^{(1)} - x^* = x^{(0)} - x^* - J_F(x^{(0)})^{-1}\big(F(x^{(0)}) - \underbrace{F(x^*)}_{=0}\big) =$$

$$= J_F(x^{(0)})^{-1}\big(F(x^*) - F(x^{(0)}) - J_F(x^{(0)})(x^* - x^{(0)})\big)$$

10 min

Therefore

$$\|x^{(1)}-x^*\| \le \|J_F(x^{(0)})^{-1}\| \, \| F(x^*) - F(x^{(0)}) - J_F(x^{(0)})(x^*-x^{(0)}) \| \quad (2)$$

We write

$$F(x^*)-F(x^{(0)}) - J_F(x^{(0)})(x^*-x^{(0)}) =$$

$$= \int_0^1 J_F(x^{(0)}+t(x^*-x^{(0)}))(x^*-x^{(0)})\,dt - \int_0^1 J_F(x^{(0)})(x^*-x^{(0)})\,dt =$$

$$= \int_0^1 \left(J_F(x^{(0)}+t(x^*-x^{(0)})) - J_F(x^{(0)}) \right) dt \; (x^*-x^{(0)})$$

Inserting this into (2) we get

$$\|x^{(1)}-x^*\| \le \|J_F(x^{(0)})^{-1}\| \left\| \int_0^1 \left(J_F(x^{(0)}+t(x^*-x^{(0)})) - J_F(x^{(0)}) \right) dt \, (x^*-x^{(0)}) \right\|$$

$$\le \|J_F(x^{(0)})^{-1}\| \|x^*-x^{(0)}\| \int_0^1 \| J_F(x^{(0)}+t(x^*-x^{(0)})) - J_F(x^{(0)}) \| \, dt \quad (3)$$

Since $x^{(0)}+t(x^*-x^{(0)}) \in B(x^*, r) \;\; \forall t \in [0,1]$,

$$\| J_F(x^{(0)}+t(x^*-x^{(0)})) - J_F(x^{(0)}) \| \le$$

$$\le L \, \| x^{(0)} + t(x^*-x^{(0)}) - x^{(0)} \| \le L t \|x^*-x^{(0)}\|$$

10 mm

Hence, inserting into (3) we get

$$\|x^{(1)}-x^*\| \le \|J_F(x^{(0)})^{-1}\| \, \|x^*-x^{(0)}\| \int_0^1 L t \|x^*-x^{(0)}\| \, dt =$$

$$= 2C \frac{L}{2} \|x^*-x^{(0)}\|^2 = C L \|x^*-x^{(0)}\|^2$$

Step 3) Show $x^{(1)} \in B(x^*, r)$

Since $x^{(0)} \in B(x^*, r)$ and $r = \min\{R, \frac{1}{2CL}\}$,

$$\|x^*-x^{(0)}\| \le \frac{1}{2CL}$$

Therefore

$$\|x^{(1)}-x^*\| \le CL \cdot \frac{1}{2CL} \|x^*-x^{(0)}\| = \frac{1}{2} \|x^*-x^{(0)}\|$$

Therefore, we can now iterate again with $x^{(1)}$ as initial guess, instead of $x^{(0)}$.

$\square$

10 mm