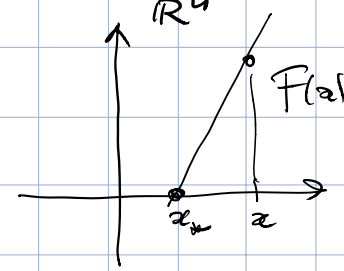
Newton method As we are almost always looking for critical points (i.e. solution to $\nabla f = 0$), it makes sense to look for general tool of finding roots, i.e. x_*: $F(x_*)=0$, $F: \mathcal{U} \to \mathbb{R}^n$, $\mathcal{U} \subset \mathbb{R}^n$



Easy case: F is linear: $F(x)=Ax+b$; then

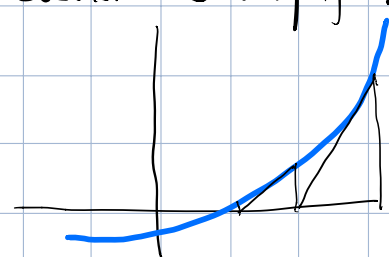
$$0 = F(x_*) = F(x) + A(x_* - x), \text{ and}$$

$$x_* = -A^{-1}F(x) + x.$$

If F is nonlinear, just pretend it is, using $D_x F$ as approx. for A: So, Newton method just iterates

$$x_{k+1} = x_k - \underbrace{(D_{x_k}F)^{-1}}_{\mathrm{Mat}(n\times n)} \underbrace{F(x_k)}_{\in \mathbb{R}^n}$$

From pictures, Newton method should be very fast:



Example: $n=1$, $f = x^3 - x$

$$N(x) \mapsto x - \frac{x^3-x}{3x^2-1} = \frac{2x^3}{3x^2-1}$$

$x_0 = 2$

$x_1 = \frac{16}{17}$

$x_2 = \frac{20842662324445349}{20717993079584776} \approx 1.0060+\ldots$

However, there might be some problems, too:

$x_0 = 1/\sqrt{5}$ $\quad x_1 = \frac{2}{5\sqrt{5}(3/5-1)} = -1/\sqrt{5}$

$x_2 = 1/\sqrt{5}$ etc...

Thm If $F: \mathcal{U} \to \mathbb{R}^n$ is continuously differentiable, $F(x_*)=0$ and $D_{x_*}F$ is invertible, then:

a) There is an open ball $B_\rho(x_*) =: B$: for any starting $x_0 \in B$, $\{x_k\} \subset B$; $x_k \to x_*$ &

$$\lim \frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} = 0 \quad \left(\text{i.e. } \frac{1}{k}\log \|x_k - x_*\| \longrightarrow -\infty\right)$$

b) If F is 2-Lipschitz, then

$$\|x_{k+1} - x_*\| \le K \|x_k - x_*\|^2$$

(i.e. $\|x_k - x_*\| = O(e^{-c2^k})$).

<u>Proof:</u> Compute $x_{k+1} - x_* = x_k - x_* - (D_{x_k}F)^{-1} F(x_k) =$

$$= (D_{x_k}F)^{-1}\left[D_{x_k}F(x_k - x_*) - F(x_k)\right] =$$

$$= (D_{x_k}F)^{-1}\left[D_{x_k}F(x_k - x_*) - \int_0^1 D_{x_t}F \cdot (x_k - x_*)\,dt\right] =$$

(here $x_t = x_* + t(x_k - x_*)$ —

— we use standard formula $F(x_k) - \underbrace{F(x_*)}_{=0} = \int_0^1 \frac{d}{dt} F(x_* + t(x_k - x_*))\,dt$

$$= \int_0^1 D_{x_t}F \cdot (x_* - x_k)\,dt \Big)$$

$$= (D_{x_k}F)^{-1}\left[\int_0^1 (D_{x_k}F - D_{x_t}F)(x_k - x_*)\,dt\right]$$

$$\Rightarrow \|x_{k+1} - x_*\| \le \|(D_{x_k}F)^{-1}\| \cdot \|x_k - x_*\| \cdot \underbrace{\int_0^1 \|D_{x_k}F - D_{x_t}F\|\,dt}_{\text{can be assumed small}} \quad (*)$$

As D_xF is continuous in x, for $\|x_t - x_k\|$ small, i.e. for $x_k \in B_\rho(x_*)$. E.g., if $\|D_{x_k}F\|^{-1} \le M$ on $B_\rho(x_*)$ we can shrink ρ if needed so that

$\|D_{x_k}F - D_{x_*}F\| \le 1/4M$ on B, and therefore $(*) \le$

$$\le M \cdot \frac{1}{2M} \cdot \|x_k - x_*\| \Rightarrow \|x_k - x_*\| \longrightarrow 0 \text{ geometrically.}$$

This $1/2$ is actually going done with ρ, so that a) holds.

Now, if DF is Lipschitz, then $(*) \le M\|x_k - x_*\| \cdot L\|x_k - x_*\|$

$$= ML\|x_k - x_*\|^2.$$

This implies b).

A useful (and visually interesting) class of mappings is given by $f:\mathbb{C}\to\mathbb{C}$, a polynomial. Julia sets emerging there are examples of fractals.

What happens when DF_{x_*} is not invertible? Quadratic convergence disappears. Say $F(x)=x^2$ $(x_*=0)$

Then $x_{k+1}=x_k-\frac{x_k^2}{2x_k}=x_k/2$; $\Rightarrow$ bland geometric convergence.

Question about the region for starting points to converge to x_* is important one. One typical result:

(valid, actually, for ∞-dim Banach spaces)

<u>Thm</u> (Kantorovich) $F:D\to\mathbb{R}^n$; D – open convex:

If for $x_0\in D$, $\|(D_{x_0}F)^{-1}F(x_0)\|\le\alpha$;

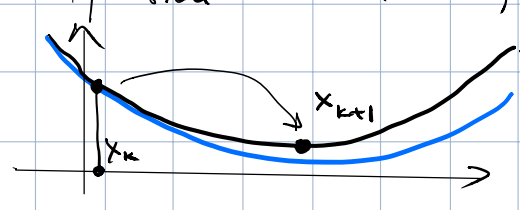
$\|(D_{x_0}F)^{-1}(D_xF-D_yF)\|\le k\|x-y\|$ for $x,y\in D$

$\alpha k<1/2$,

$B_{2\alpha}(x_0)\subset D$, then $B_{2\alpha}(x_0)\ni x_*$, unique root $F(x_*)=0$; and $\|x_k-x_*\|\le\frac{\alpha}{2^{n-1}}q^{2^n-1}$, $q=2\alpha k<1$

<u>No proof here.</u>

Use of Newton method in optimization requires finding DF, where $F(x)=\nabla f(x)$, i.e. $DF=Hf$. NM is therefore a <u>second order</u> method. Conceptually, one approximates f with its Taylor expansion to 2nd order, and finds minimum:



$f(x_k)+(\nabla f(x_k),x-x_k)+$
$+\frac{1}{2}(Hf(x-x_k),x-x_k)$