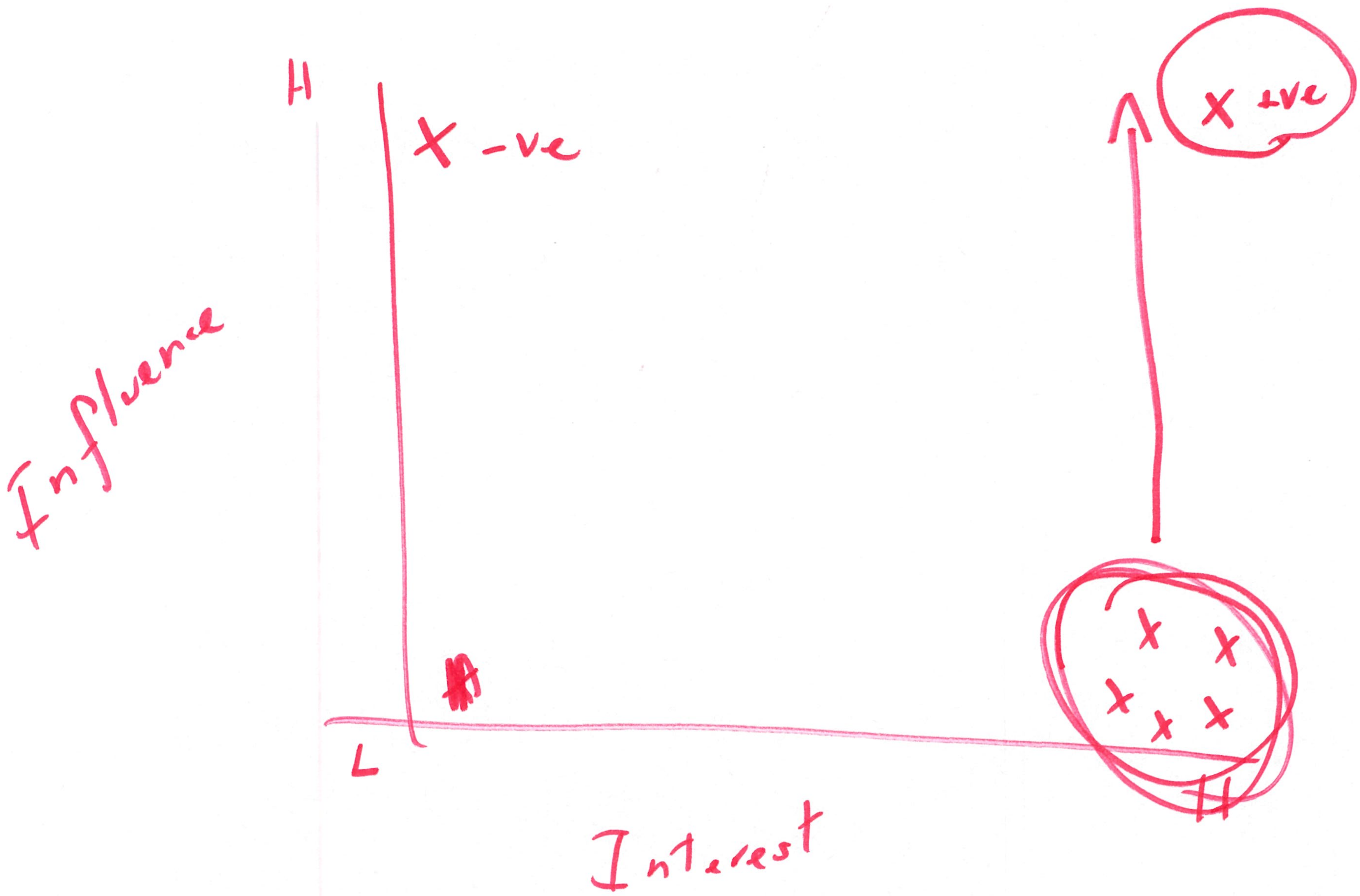

H
X -ve
X +ve
Influence
L
H
Interest

Convergence Analysis

Theorem

Let $\mathbf{F} : D \to \mathbb{R}^N$ *for some open subset* $D \subset \mathbb{R}^N$ *and let the standard assumptions hold. Then there exists some* δ *such that given* $\mathbf{x}^{(0)} \in D$ *with* $\|\mathbf{x}^{(0)} - \mathbf{x}^*\| < \delta$*, the Newton iteration*

$+ \ \underline{\delta x}_N^{(k)} \ ; \quad J(\underline{x}^{(k)})\,\underline{\delta x}_N^{(k)} = -\underline{F}(\underline{x}^{(k)})$

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \mathbf{J}\left(\mathbf{x}^{(k)}\right)^{-1} \mathbf{F}\left(\mathbf{x}^{(k)}\right) \quad \ldots\ldots (1)$$

satisfies $\|\mathbf{x}^{(k)} - \mathbf{x}^*\| < \delta$ *(for all* $k = 1, 2 \ldots$*) and converges quadratically to* $\mathbf{x}^*$*.*

Convergence Analysis

Definition (Rate of Convergence)

Let $\{\mathbf{x}^{(k)}\}_{k=0}^{\infty} \to \mathbf{x}^* \in \mathbb{R}^N$. We say

- $\mathbf{x}^{(k)} \to \mathbf{x}^*$ **quadratically** if there is $K > 0$ such that

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

- $\mathbf{x}^{(k)} \to \mathbf{x}^*$ **superlinearly** with order $\alpha > 1$ if there is $K > 0$ such that

$$\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\| \leq K\|\mathbf{x}^{(k)} - \mathbf{x}^*\|^\alpha$$

- $\mathbf{x}^{(k)} \to \mathbf{x}^*$ **linearly** with factor $\sigma \in (0, 1)$ if

$$\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\| \leq \sigma\|\mathbf{x}^{(k)} - \mathbf{x}^*\|$$

for k sufficiently large.

Convergence Analysis

Lemma

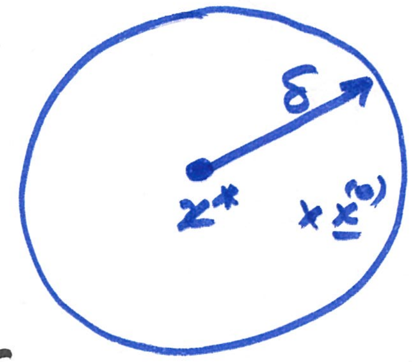
Let $\mathbf{B} : D \to \mathbb{R}^{N\times N}$ for some open convex set $D \subset \mathbb{R}^N$. If $\mathbf{B}$ is integrable on the closed line segment connecting $\mathbf{x}$ and $\mathbf{x}+\mathbf{h}$ denoted $[\mathbf{x}, \mathbf{x}+\mathbf{h}] \subset D$ then

$$\left\| \int_0^1 \mathbf{B}(\mathbf{x}+t\mathbf{h})\mathbf{h}\, dt \right\| \leq \int_0^1 \|\mathbf{B}(\mathbf{x}+t\mathbf{h})\mathbf{h}\| \; dt\,.$$

ROUGH OUTLINE OF PROOF.

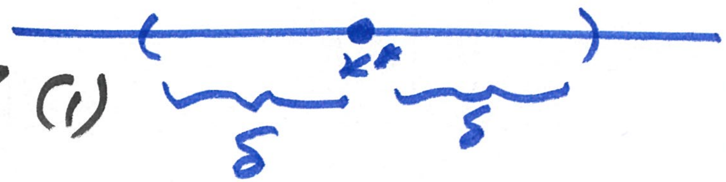
Note: ① $\|\underline{x}^{(0)} - \underline{x}^*\| < \delta$; $\underline{x}^{(0)} \in B_\delta(\underline{x}^*)$.

$$B_\delta(\underline{x}^*) = \{ \underline{x} \in \mathbb{R}^n \mid \|\underline{x} - \underline{x}^*\| < \delta \}.$$



② Define the error

$$\underline{\varepsilon}^{(k+1)} = \underline{x}^{(k+1)} - \underline{x}^* \, ; \quad \underline{\varepsilon}^{(k)} = \underline{x}^{(k)} - \underline{x}^*$$



Subtract $\underline{x}^*$ from both sides of (1)

$$\underline{\varepsilon}^{(k+1)} = \underline{\varepsilon}^{(k)} - J(\underline{x}^{(k)})^{-1} \underline{F}(\underline{x}^* + \underline{\varepsilon}^{(k)})$$

Recall $$\underline{F}(\underline{x}^* + \underline{\varepsilon}^{(k)}) = \overset{=\underline{0}}{\underline{F}(\underline{x}^*)} + \int_0^1 J(\underline{x}^* + t\underline{\varepsilon}^{(k)})\, \underline{\varepsilon}^{(k)}\, dt$$

STANDARD ASSUMPTION 1. $\underline{x}^*$ is a solution of $\underline{F}(\underline{x}) = \underline{0}$ on D.

Hence, $$\|\underline{\varepsilon}^{(k+1)}\| = J(\underline{x}^{(k)})^{-1}\left[J(\underline{x}^{(k)})\underline{\varepsilon}^{(k)} - \int_0^1 J(\underline{x}^* + t\underline{\varepsilon}^{(k)})\underline{\varepsilon}^{(k)}\,dt\right]$$

. introduce norm.

$$= \left\|J(\underline{x}^{(k)})^{-1}\int_0^1 \left[J(\underline{x}^{(k)}) - J(\underline{x}^* + t\underline{\varepsilon}^{(k)})\right]\underline{\varepsilon}^{(k)}\,dt\right\|$$

* Linear Algebra
. matrix-vector norm compatibility

$$\leqslant \|J(\underline{x}^{(k)})^{-1}\|\,\left\|\int_0^1 \left(J(\underline{x}^{(k)}) - J(\underline{x}^* + t\underline{\varepsilon}^{(k)})\right)\underline{\varepsilon}^{(k)}\,dt\right\|$$

LEMMA
$$\leqslant \|J(\underline{x}^{(k)})^{-1}\|\int_0^1 \|J(\underline{x}^{(k)}) - J(\underline{x}^* + t\underline{\varepsilon}^{(k)})\|\,\|\underline{\varepsilon}^{(k)}\|\,dt$$

STANDARD ASSUMPTION 2 : $J(\underline{x}^{(k)})$ is nonsingular on D and there exists $\beta > 0$ such that
$$\|J(\underline{x}^{(k)})^{-1}\| \leqslant \beta.$$

$$\|\underline{\varepsilon}^{(k+1)}\| \leqslant \beta\int_0^1 \|J(\underline{x}^{(k)}) - J(\underline{x}^* + t\underline{\varepsilon}^{(k)})\|\,\|\underline{\varepsilon}^{(k)}\|\,dt.$$

STANDARD ASSUMPTION 3 : $J(\underline{x}) \in Lip_\gamma(D)$

i.e., $$\|J(\underline{x}) - J(\underline{y})\| \leqslant \gamma\,\|\underline{x} - \underline{y}\|.$$

$$\therefore \quad \|\underline{\varepsilon}^{(k+1)}\| \leq \beta V \int_0^1 \| \underbrace{\underline{x}^{(k)} - \underline{x}^*}_{\underline{\varepsilon}^{(k)}} - t\,\underline{\varepsilon}^{(k)} \| \, dt \quad \|(1-t)\underline{\varepsilon}^{(k)}\|$$

$$\leq \beta V \|\underline{\varepsilon}^{(k)}\|^2 \int_0^1 |1-t| \, dt$$

$$\|\alpha \underline{v}\| = |\alpha| \|\underline{v}\|.$$

$$\leq \underbrace{\frac{\beta V}{2}}_{K} \|\underline{\varepsilon}^{(k)}\|^2$$

$$K = \frac{\beta V}{2}.$$

$$\boxed{\|\underline{\varepsilon}^{(k+1)}\| \leq K \|\underline{\varepsilon}^{(k)}\|^2 \quad \ldots (2)}$$

$\eta = K\delta < 1$

$\delta < \frac{1}{K}$

Convergence :

Base case : $k=0$, given $\underline{x}^{(0)} \in B_\delta(\underline{x}^*)$

i.e., $\|\underline{x}^{(0)} - \underline{x}^*\| < \delta$.

use (2) with $k=0$ and

$$\|\underline{\varepsilon}^{(1)}\| = \|\underline{x}^{(1)} - \underline{x}^*\| \leq K \|\underline{\varepsilon}^{(0)}\|^2 = \underbrace{K\|\underline{\varepsilon}^{(0)}\|}_{\eta} \|\underline{\varepsilon}^{(0)}\|$$

Reduce δ if necessary so that

$\eta < 1 \Rightarrow \|\underline{\varepsilon}^{(1)}\| < \|\underline{\varepsilon}^{(0)}\| < \delta$

i.e, $\|\underline{x}^{(1)} - \underline{x}^*\| < \delta$

$\therefore \ \underline{x}^{(1)} \in B_\delta(\underline{x}^*)$.

We continue this proof using mathematical induction to show that the entire sequence $\{\underline{x}^{(k)}\}_{k=1}^{\infty} \subset B_\delta(\underline{x}^*)$.

Consider now, $\|\underline{x}^{(k)} - \underline{x}^*\| \leq \eta \|\underline{x}^{(k-1)} - \underline{x}^*\|$

i.e.,
$$\begin{aligned}
\|\underline{x}^{(k)} - \underline{x}^*\| &\leq \eta \cdot \eta \|\underline{x}^{(k-2)} - \underline{x}^*\| \\
&\leq \eta \cdot \eta \cdot \eta \|\underline{x}^{(k-3)} - \underline{x}^*\| \\
&\ \ \vdots \\
&\leq \eta^k \|\underline{x}^{(0)} - \underline{x}^*\|.
\end{aligned}$$

$$\lim_{k\to\infty} \|\underline{x}^{(k)} - \underline{x}^*\| \leq \left(\lim_{k\to\infty} \eta^k\right) \|\underline{x}^{(0)} - \underline{x}^*\| = 0 \cdot \|\underline{x}^{(0)} - \underline{x}^*\| = 0.$$

$$\therefore \lim_{k\to\infty} \|\underline{x}^{(k)} - \underline{x}^*\| = 0$$

$$\underline{x}^{(k)} \longrightarrow \underline{x}^*. \quad \square$$

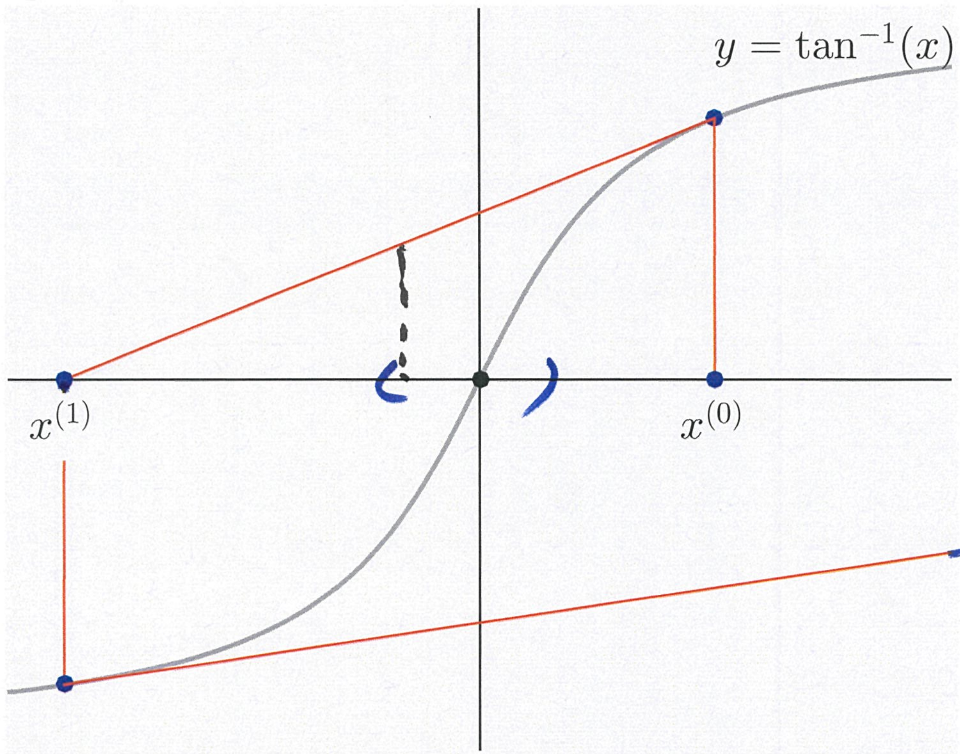
Globally Convergent Methods

- All of the methods presented so far suffer from the same weakness: there is no guarantee that convergence will be achieved for an arbitrary choice of initial iterate $\mathbf{x}^{(0)}$.

- **Globally convergent methods** seek to improve upon (but not, as the name might suggest, completely overcome) this weakness. In this section, we discuss a class of globally convergent Newton methods called **line search methods**.

- To illustrate the idea, consider Newton's method applied to the scalar function $f(x) = \tan^{-1}(x)$ with $x^{(0)} = 2.0$. The first few iterations are exhibited in Figure 2.

Globally Convergent Methods



$x^{(1)} = x^{(0)} + \lambda\, \delta x^{(0)}$

$\delta x^{(0)} = \dfrac{-f(x^{(0)})}{f'(x^{(0)})}$

$x^{(2)} = x^{(1)} + \delta x^{(1)}$

k	$x^{(k+1)}$	$\delta x^{(k)}$	$\lvert f(x^{(k+1)}) \rvert$
0	-3.54	-5.54	1.30
1	13.95	17.49	1.50
2	-279.34	-293.30	1.57
3	122017.00	122296.34	1.57
4	-23386004197.93	-23386126214.93	1.57

Figure: Newton iterations for $f(x) = \tan^{-1}(x)$ and $x^{(0)} = 2.0$ showing divergence.

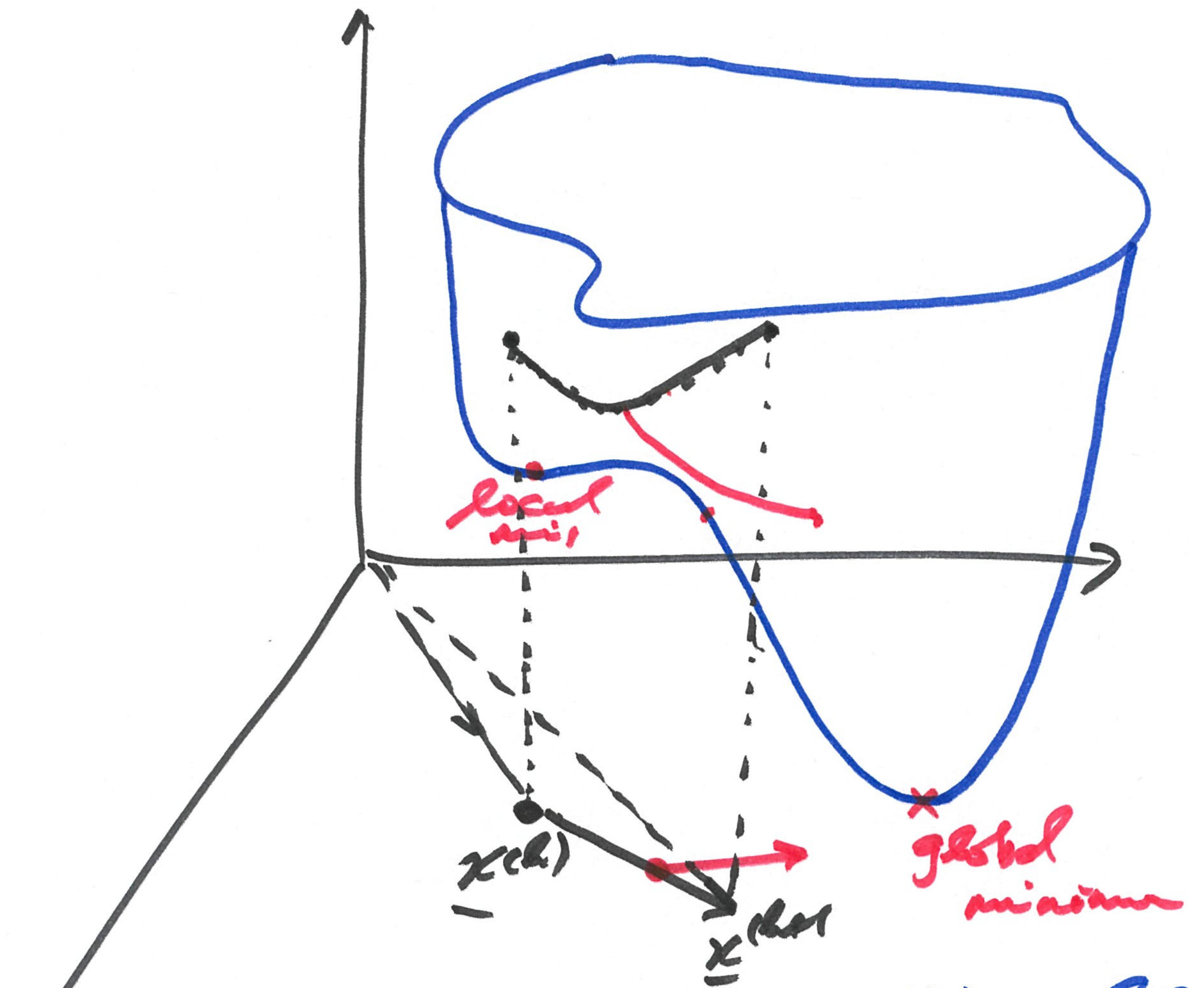
Globally Convergent Methods

The above example motivates the idea of limiting the size of the Newton step to ensure the nonlinear residual norm decreases. Thus, we consider an iterative scheme of the form:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \lambda^{(k)}\boldsymbol{\delta}\mathbf{x}^{(k)}, \quad \boldsymbol{\delta}\mathbf{x}^{(k)} = -\mathbf{J}\left(\mathbf{x}^{(k)}\right)^{-1}\mathbf{F}\left(\mathbf{x}^{(k)}\right) \qquad (11)$$

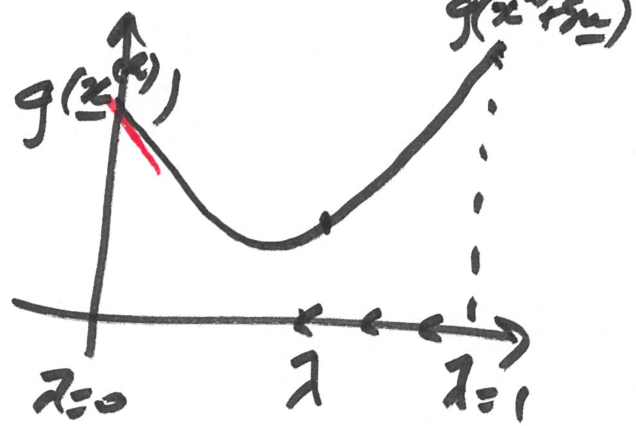
with $0 < \lambda^{(k)} \leq 1$.

$\underline{F}(\underline{x}) = \underline{0}$; $\underline{F}(\underline{x}) = (f_1(\underline{x}), f_2(\underline{x}) \ldots f_N(\underline{x}))^T$



$$g(\underline{x}) = \|\underline{F}(\underline{x})\|_2^2 = \sum_{i=1}^{N} f_i^2(\underline{x})$$

$$h(\lambda) = g(\underline{x} + \lambda\,\delta\underline{x})$$



The process of "backtracking" along the Newton search direction is called "line searching".

How do we know the Newton search direction is a "descent direction" for $g(\underline{x})$?

Usual process from M.V.C. is that we must show

$$\underline{\nabla} g^T \delta \underline{x}_N < 0$$

where $g(\underline{x}^{(k)}) = \|\underline{F}(\underline{x}^{(k)})\|_2^2 = \sum_{i=1}^{N} f_i^2(\underline{x})$

the j^{th} component of the gradient is

$$\underline{\nabla} g = \left(\frac{\partial g}{\partial x_1}, \frac{\partial g}{\partial x_2}, \ldots, \boxed{\frac{\partial g}{\partial x_j}}, \ldots \frac{\partial g}{\partial x_N} \right)^T$$

$$\left[\underline{\nabla} g\right]_j = \frac{\partial g}{\partial x_j} = \frac{\partial}{\partial x_j} \sum_{i=1}^{N} f_i^2(\underline{x}^{(k)}) = 2 \sum_{i=1}^{N} \frac{\partial f_i}{\partial x_j}(\underline{x}^{(k)}) f_i(\underline{x}^{(k)})$$

$$\underline{\nabla} g(\underline{x}^{(k)}) = 2 J^T(\underline{x}^{(k)}) \underline{F}(\underline{x}^{(k)}) \cdot \left(\frac{\partial f_1}{\partial x_j} \; \frac{\partial f_2}{\partial x_j} \; \cdots \; \frac{\partial f_N}{\partial x_j} \right) \begin{pmatrix} f_1(\underline{x}^{(k)}) \\ f_2(\underline{x}^{(k)}) \\ \vdots \\ f_N(\underline{x}^{(k)}) \end{pmatrix}$$

$$\delta \underline{x}_N = -J(\underline{x}^{(k)})^{-1} \underline{F}(\underline{x}^{(k)}).$$

$$\therefore \quad \nabla g(\underline{x}^{(k)})^T \delta \underline{x}_N = -2\left(J^T(\underline{x}^{(k)}) \underline{F}(\underline{x}^{(k)}) \right)^T J(\underline{x}^{(k)})^{-1} \underline{F}(\underline{x}^{(k)})$$

$$= -2 \underline{F}(\underline{x}^{(k)})^T \underbrace{J(\underline{x}^{(k)}) J(\underline{x}^{(k)})^{-1}}_{I} \underline{F}(\underline{x}^{(k)})$$

$$= -2g(\underline{x}^{(k)}) < 0.$$

Simple Line searching

A simple means of computing $\lambda^{(k)}$ is to use repeated halvings until the resulting step produces a reduction in the nonlinear residual norm. Such an algorithm is given below.

Simple line search

$\lambda^{(k)} = 1$

$\mathbf{x}^{\dagger} = \mathbf{x}^{(k)} + \lambda^{(k)} \delta \mathbf{x}^{(k)}$

while $\|\mathbf{F}(\mathbf{x}^{\dagger})\| \geq \|\mathbf{F}(\mathbf{x}^{(k)})\|$ & $\lambda > \lambda_{min}$.

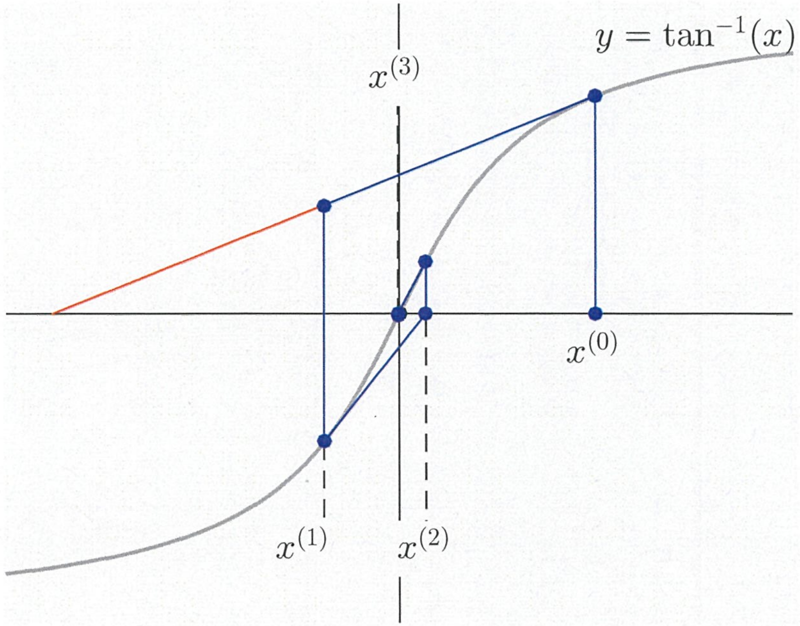
  $\lambda^{(k)} = \lambda^{(k)}/2$

  $\mathbf{x}^{\dagger} = \mathbf{x}^{(k)} + \lambda^{(k)} \delta \mathbf{x}^{(k)}$

end

$\mathbf{x}^{(k+1)} = \mathbf{x}^{\dagger}$

Simple Line searching



k	$x^{(k+1)}$	$\lambda^{(k)}$	$\delta x^{(k)}$	$\lvert f\left(x^{(k+1)}\right)\rvert$
0	-7.68e-01	0.50	-5.54e+00	6.55e-01
1	2.73e-01	1.00	1.04e+00	2.67e-01
2	-1.34e-02	1.00	-2.86e-01	1.34e-02
3	1.60e-06	1.00	1.34e-02	1.60e-06
4	-2.71e-18	1.00	-1.60e-06	2.71e-18

Figure: Newton iterations for $f(x) = \tan^{-1}(x)$ and $x^{(0)} = 2.0$ with simple line searching showing convergence to the root $x = 0$. Full Newton steps (equivalent to $\lambda^{(k)} = 1.0$) are shown in red.

Simple Line searching

- To again illustrate the idea, consider Newton's method applied to the scalar function $f(x) = \tan^{-1}(x)$ with $x^{(0)} = 2.0$. The first few iterations using simple line searching are summarised in Figure 3.
- Observe that during the first iteration, the size of the Newton step is limited to half of the full Newton step ($\lambda^{(0)} = 0.5$). This produces a decrease (and indeed sufficient decrease) in the nonlinear residual since $|f(x^{(1)})| = 0.65 < |f(x^{(0)})| = 1.11$ (Figure 3).
- On the other hand, this is not true for the full Newton step (equivalent to $\lambda^{(0)} = 1.0$) as $|f(x^{(1)})| = 1.30 > |f(x^{(0)})| = 1.11$ (Figure 2).