MXB324
SHORT LECTURE 2

GLOBALLY CONVERGENT NEWTON METHODS

LINE SEARCHING

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:

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

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

Simple Line searching

The simple line searching algorithm is based on ensuring

$$\left\|\mathbf{F}(\mathbf{x}^{(k+1)})\right\| < \left\|\mathbf{F}(\mathbf{x}^{(k)})\right\|$$

or equivalently:

$$\left\|\mathbf{F}(\mathbf{x}^{(k)} + \lambda^{(k)}\boldsymbol{\delta}\mathbf{x}^{(k)})\right\| < \left\|\mathbf{F}(\mathbf{x}^{(k)})\right\| . \tag{12}$$

This almost always works well. However, there is a theoretical possibility that iterates satisfying (12) may still oscillate about the root without converging.

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

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

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

while $\|\mathbf{F}(\mathbf{x}^{\dagger})\| \geq \|\mathbf{F}(\mathbf{x}^{(k)})\|$

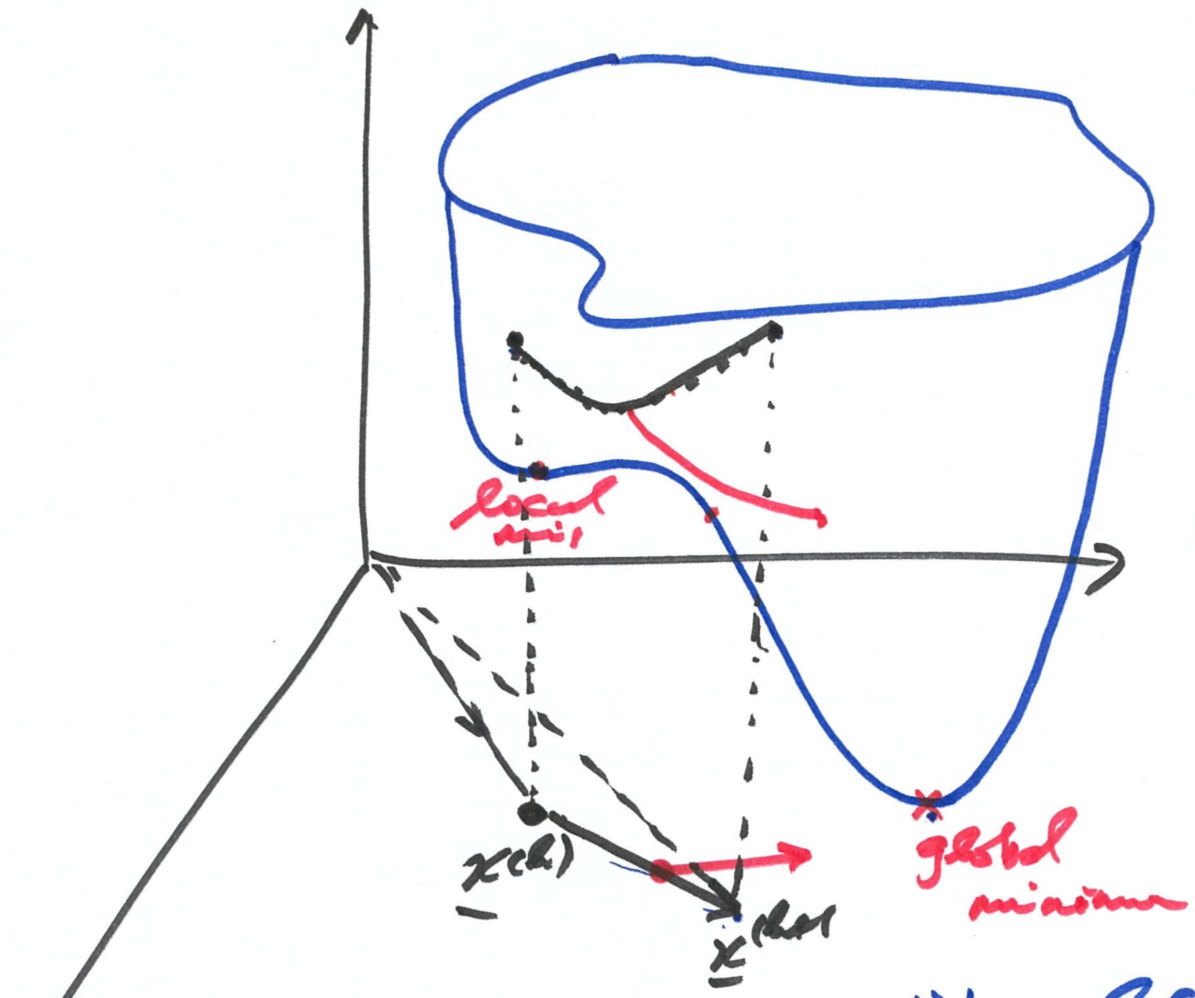
  $\lambda^{(k)} = \lambda^{(k)}/2$

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

end

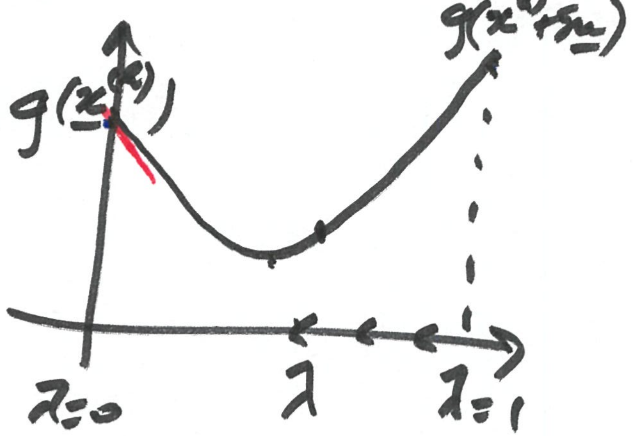
$\mathbf{x}^{(k+1)} = \mathbf{x}^{\dagger}$

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



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

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



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

Recall from WEEK 7 LECTURE :

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

$$= 2\, J^T(\underline{x})\, \underline{F}(\underline{x})$$

2. The Newton step is a "descent direction" for $g(\underline{x})$:

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

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

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

We could consider the unconstrained minimisation problem:

$$\min_{\underline{x} \in \mathbb{R}^N} g(\underline{x}) : \mathbb{R}^N \longrightarrow \mathbb{R} \; ;$$

$$g(\underline{x}) := \| \underline{E}(\underline{x}) \|_2^2 = \underline{F}^T(\underline{x}) \underline{F}(\underline{x})$$

$$= \sum_{k=1}^{N} f_k^2(\underline{x})$$

IDEA: Applying Newton's method to

$$\underline{\nabla} g(\underline{x}) = \underline{0} .$$

Hessian matrix for g.

By Taylor's Th^{m} for functions of several variables the value in the neighbourhood of $\underline{x}^{(k)}$ is:

$$g(\underline{x}) = g(\underline{x}^{(k)}) + (\underline{x} - \underline{x}^{(k)})^T \nabla g(\underline{x}^{(k)}) + \frac{1}{2} (\underline{x} - \underline{x}^{(k)})^T \boxed{H(\underline{x}^{(k)})} (\underline{x} - \underline{x}^{(k)}) + O(\| \underline{x} - \underline{x}^{(k)} \|^2) \quad \ldots (1)$$

$$[H(\underline{x})]_{ij} = \frac{\partial^2 g(\underline{x})}{\partial x_i \partial x_j} \quad \text{Symmetric matrix.}$$

We can differentiate (1) at any point $\underline{x}$ to obtain:

$$\underline{\nabla} g(\underline{x}) = \underline{\nabla} g(\underline{x}^{(k)}) + H(\underline{x}^{(k)})(\underline{x} - \underline{x}^{(k)}) + O(\|\underline{x} - \underline{x}^{(k)}\|) . \quad \ldots (2).$$

$$\underline{x}^{(k+1)} = \underline{x}^{(k)} + \delta \underline{x}_g$$

where $H(\underline{x}^{(k)}) \, \delta \underline{x}_g = - \underbrace{\underline{\nabla} g(\underline{x}^{(k)})}$.

$? \quad = -2 J^T(\underline{x}^{(k)}) \underline{F}(\underline{x}^{(k)})$.

Consider the i,j^{th} entry of $H(\underline{x}^{(k)})$:

$$\frac{\partial^2 g(\underline{x}^{(k)})}{\partial x_i \partial x_j} = \frac{\partial^2}{\partial x_i \partial x_j} \sum_{\cancel{k}\, m=1}^{N} f_{\cancel{k}\, m}^2(\underline{x}^{(k)}).$$

$$= 2 \frac{\partial}{\partial x_i} \left(\sum_{\cancel{k}\, m=1}^{N} \frac{\partial f_{\cancel{k}\, m}}{\partial x_j} f_{\cancel{k}\, m} \right)$$

$$= 2 \sum_{\cancel{k}\, m=1}^{N} \left(\frac{\partial f_{\cancel{k}\, m}}{\partial x_j} \frac{\partial f_{\cancel{k}\, m}}{\partial x_i} + f_{\cancel{k}\, m} \frac{\partial^2 f_{\cancel{k}\, m}}{\partial x_i \partial x_j} \right)$$

$$\therefore H(\underline{x}^{(k)}) = 2 \left\{ J^T(\underline{x}^{(k)}) J(\underline{x}^{(k)}) + \underbrace{\sum_{\cancel{k}\, m=1}^{N} f_{\cancel{k}\, m}(\underline{x}^{(k)}) H_{f_{\cancel{k}\, m}}(\underline{x}^{(k)})}_{B(\underline{x}^{(k)})} \right\}.$$

Suppose $J(\underline{x}^{(k)})$ is invertible.

Hence,

$$\cancel{2} \left\{ \cancel{J^{-T}} \cancel{J^T(\underline{x}^{(k)})} J(\underline{x}^{(k)}) + \cancel{\overset{0}{B(\underline{x}^{(k)})}} \right\} \delta \underline{x}_g = -\cancel{2}\, \cancel{J^{-T}}\, \cancel{J^T(\underline{x}^{(k)})}\, \underline{F}(\underline{x}^{(k)})$$

$$J(\underline{x}^{(k)}) \delta \underline{x}_g = -\underline{F}(\underline{x}^{(k)})$$

IDEA: Determine $\lambda^{(k)}$ in such a way that $g(\underline{x})$ is <u>minimised</u> along search direction. Too HARD!

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

i.e, find λ such that $\dfrac{dh}{d\lambda} = 0$.

. Instead of this approach, we use a weaker condition; choose a λ such that a 'sufficient decrease' in $g(\underline{x})$ (and also a significant change in $\underline{x}$).

" Reduce the overall computation at each iteration ".

Q: What do we mean by a 'sufficient decrease' in $g(\underline{x})$ has occurred?

USE ARMIJO, GOLDSTEIN & PRICE RULES.

1. $h(\lambda^+) \leq h(0) + \underline{\alpha} \lambda \underline{\underline{h'(0)}}$, $0<\alpha<1$

2. $\lambda^+ > \lambda_{min}$ (1×10^{-3}).

The condition 1. ensures the reduction in function value is at least some proportion of its value if $g(\underline{x})$ had been linear and continued to decrease at a rate given by the initial slope.

$$h(0) + \alpha\lambda\, h'(0) \quad ; \quad \frac{dh(\lambda)}{d\lambda} = \frac{d}{d\lambda}\, g(\underline{x}^{(k)} + \lambda\, \delta\underline{x}^{(k)})$$

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

$$\| $$

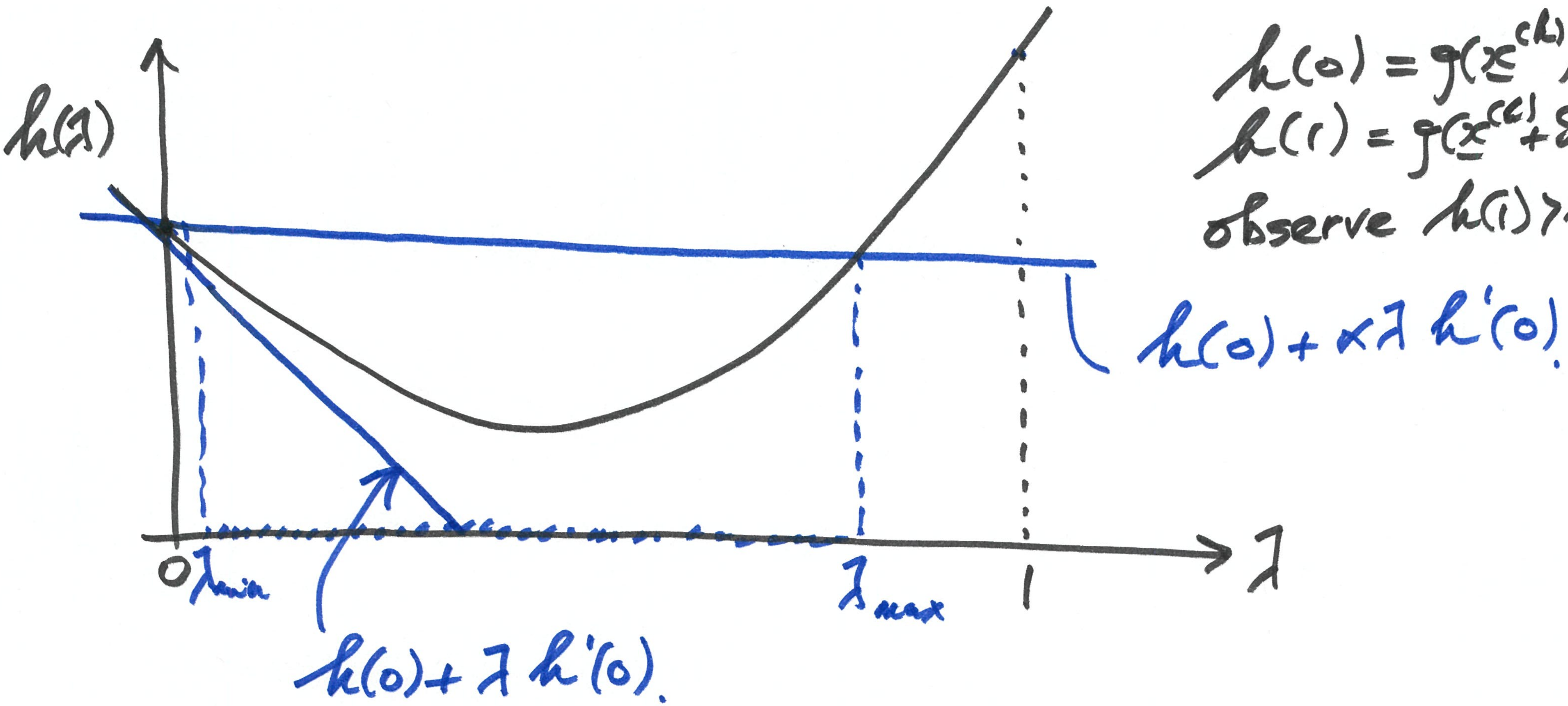
$$g(\underline{x}^{(k)}) + \alpha\lambda\left(-2\, g(x^{(k)})\right)$$

$$= (1 - 2\alpha\lambda)\, g(\underline{x}^{(k)}).$$

$h(0) = g(\underline{x}^{(k)})$

$h(1) = g(\underline{x}^{(k)} + \delta \underline{x}^{(k)})$;

observe $h(1) > h(0)$



Simple Line searching

To overcome this, a slightly stronger condition than (12) is used in practice:

$$\left\|\mathbf{F}(\mathbf{x}^{(k)} + \lambda^{(k)}\boldsymbol{\delta}\mathbf{x}^{(k)})\right\| < \left(1 - \alpha\lambda^{(k)}\right)\left\|\mathbf{F}(\mathbf{x}^{(k)})\right\| \tag{13}$$

or for a slightly different variant using the square of the nonlinear residual:

$$g(\mathbf{x}^{(k)} + \lambda^{(k)}\boldsymbol{\delta}\mathbf{x}^{(k)}) < \left(1 - 2\alpha\lambda^{(k)}\right) g(\mathbf{x}^{(k)}), \tag{14}$$

with $g(\mathbf{x}) = \|\mathbf{F}(\mathbf{x})\|^2 = \mathbf{F}(\mathbf{x})^T\mathbf{F}(\mathbf{x})$.

Here, α is a small, positive constant (say, $\alpha = 10^{-4}$). Condition (13) or (14) is known as the **Armijo rule**. When satisfied, we say that there has been a **sufficient decrease** of the nonlinear residual norm.

General Line searching

A prototypical algorithm is given as follows.

General line search

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

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

while $g(\mathbf{x}^{\dagger}) \geq (1 - 2\alpha\lambda^{(k)})\, g(\mathbf{x}^{(k)})$ & $\lambda^{(k)} > \lambda_{min}$.

  choose $\sigma \in [\overset{0.1}{\sigma_0}, \sigma_1 = 0.5]$

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

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

end

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