Week 5, Lecture 2

Plan:

- Finish Chapter 6
- Exercises #2. These are:
  - Theoretical 1–4
  - Coding 5

  } examinable.

Convergence Rate for the Newton Method §6.3

Suppose that f is twice-differentiable and the Hessian $B(\underline{x})$ is Lipschitz in a neighbourhood of the minimizer $\underline{x}_*$ at which the sufficient conditions for optimality hold ($\nabla f(x_*)=0$, $B(\underline{x}_*)$ pos. definite). Suppose that the starting-point $\underline{x}_0$ is sufficiently close to $\underline{x}_*$ and consider the iterates

$$\underline{x}_{k+1} = \underline{x}_k + p_k^N .$$

Then:

1. The iterates converge: $\underline{x}_k \to \underline{x}_*$ as $k \to \infty$.
2. The rate of convergence is quadratic:

$$\|\underline{x}_{k+1} - \underline{x}_*\|_2 \leq C\,\|\underline{x}_k - \underline{x}_*\|_2^2 .$$

No proof!

Remark: By $B(\underline{x})$ Lipschitz in a neighbourhood $\mathcal{N}$ of $\underline{x}_*$, we mean, that there exists a positive constant K such that

$$\| B(\underline{y}_2) - B(\underline{y}_1) \|_2 \leq K \| \underline{y}_2 - \underline{y}_1 \|_2$$

for all $\underline{y}_1$ and $\underline{y}_2$ in $\mathcal{N}$.

Remark: As you go through the different steps in the proof, we get down to:

$$\frac{\| \underline{x}_1 - \underline{x}_0 \|_2}{\| \underline{x}_0 - \underline{x}_* \|_2} \leq \underbrace{C \| \underline{x}_0 - \underline{x}_* \|_2}$$

"Sufficiently close" means the combination on the RHS is $\leq 1/2$.

More about this in Exercises # 2, Question 1.

<u>One more remark</u>: A similar result holds for quasi-Newton methods (e.g. BFGS):

$$\| \underline{x}_{k+1} - \underline{x}_* \|_2 \leq C \| \underline{x}_k - \underline{x}_* \|_2^{1+\epsilon}$$

where $\epsilon > 0$. I.e. better than SD.
For an example of how S.D. ($\epsilon = 0$) can fail, see Exercises # 2, Question 5.

Week 5, Lecture 3

- Question 1
- One other question

Quadratic convergence, Newton Method:

$$\boxed{\frac{\|\underline{x}_k - \underline{x}_*\|_2}{\|\underline{x}_0 - \underline{x}_*\|_2} \le \frac{1}{2^{2^k - 1}}}$$

Question 1: To show this.

Given (from lecture notes):

$$\|\underline{x}_{k+1} - \underline{x}_*\|_2 \le C\|\underline{x}_k - \underline{x}_*\|_2^2 \quad \text{(Eq. 6.13)}$$

Work through this for different values of k.

$\underline{k=0}$: $\|\underline{x}_1 - \underline{x}_*\|_2 \le C\|\underline{x}_0 - \underline{x}_*\|_2^2$

$$\Rightarrow \frac{\|\underline{x}_1 - \underline{x}_*\|_2}{\|\underline{x}_0 - \underline{x}_*\|_2} \le \underbrace{C\|\underline{x}_0 - \underline{x}_*\|_2}_{\le 1/2}$$

The RHS is $\le 1/2$ for $\underline{x}_0$ sufficiently close to $\underline{x}_*$. Hence:

$$\boxed{\frac{\|\underline{x}_1 - \underline{x}_*\|_2}{\|\underline{x}_0 - \underline{x}_*\|_2} \le 1/2}.$$

$k = 1$:

$$\|\underline{x}_2 - \underline{x}_*\|_2 \le C \underbrace{\|\underline{x}_1 - \underline{x}_*\|_2}_{\|\underline{x}_1 - \underline{x}_*\|_2 \le \frac{1}{2}\|\underline{x}_0 - \underline{x}_*\|_2}{}^2$$

Sub in:

$$\|\underline{x}_2 - \underline{x}_*\|_2 \le C \left(\frac{1}{2}\|\underline{x}_0 - \underline{x}_*\|_2\right)\left(\frac{1}{2}\|\underline{x}_0 - \underline{x}_*\|_2\right)$$

$$\Rightarrow \|\underline{x}_2 - \underline{x}_*\|_2 \le \frac{1}{2^2}\|\underline{x}_0 - \underline{x}_*\|_2 \underbrace{\left(C\|\underline{x}_0 - \underline{x}_*\|_2\right)}_{\le 1/2}$$

$$\Rightarrow \boxed{\|\underline{x}_2 - \underline{x}_*\|_2 \le \frac{1}{2^2 \cdot 2}\|\underline{x}_0 - \underline{x}_*\|_2}$$

$k = 2$:

$$\|\underline{x}_3 - \underline{x}_*\|_2 \le C\|\underline{x}_2 - \underline{x}_*\|_2^2$$

$$\le \left(\frac{1}{2^3}\|\underline{x}_0 - \underline{x}_*\|_2\right)\left(\frac{1}{2^3}\|\underline{x}_0 - \underline{x}_*\|_2\right) \cdot C$$

$$= \frac{1}{2^6}\|\underline{x}_0 - \underline{x}_*\|_2 \underbrace{\left(C\|\underline{x}_0 - \underline{x}_*\|_2\right)}_{\le 1/2}$$

$$\Rightarrow \boxed{\|\underline{x}_3 - \underline{x}_*\|_2 \le \frac{1}{2^6 \cdot 2}\|\underline{x}_0 - \underline{x}_*\|_2}$$

So far: $\|\underline{x}_k - \underline{x}_*\|_2 \le \frac{1}{2^{p_k}}\|\underline{x}_0 - \underline{x}_*\|_2$.

$k = 1$: $\quad p_k = 1 \qquad (2^1)$.

$k = 2$ $\quad p_k = 3$

$k = 3$ $\quad p_k = 7$.

Guess the pattern:

$$\boxed{p_k = 2p_{k-1} + 1.}$$ PART (B)

This is a first-order inhomogeneous difference equation. Trial solution:

$$p_k = A + B\lambda^k.$$

Sub into the difference equation:

$$A + B\lambda^k = 2\left(A + B\lambda^{k-1}\right) + 1.$$

$$\Rightarrow A + B\lambda^k = (2A+1) + 2B\lambda^{k-1}$$

Equate coefficients of different powers of λ:

$$A = 2A+1 \Rightarrow A = -1.$$

$$B\lambda^k = 2B\lambda^{k-1}.$$

Take $B = 1$, $\lambda = 2$.

Hence:

$$p_k = 2^k - 1 \quad \left(\begin{array}{l}\text{Back-substitution into} \\ p_k = A + B\lambda^k\end{array}\right)$$

This solves PART (C), This is valid for $k=1, k=2, \ldots$.

In part (d) we put everything together.

Hence:

$$\cancel{a} + b\alpha + c\alpha^2 < \cancel{a} + c_1 \alpha b .$$

Hence, $\alpha = 0$ <u>OR</u> :

$$b + c\alpha < c_1 \cancel{\alpha} . b .$$

Extreme case with equality:

$$\cancel{b + c\alpha = c_1 \alpha b .}$$

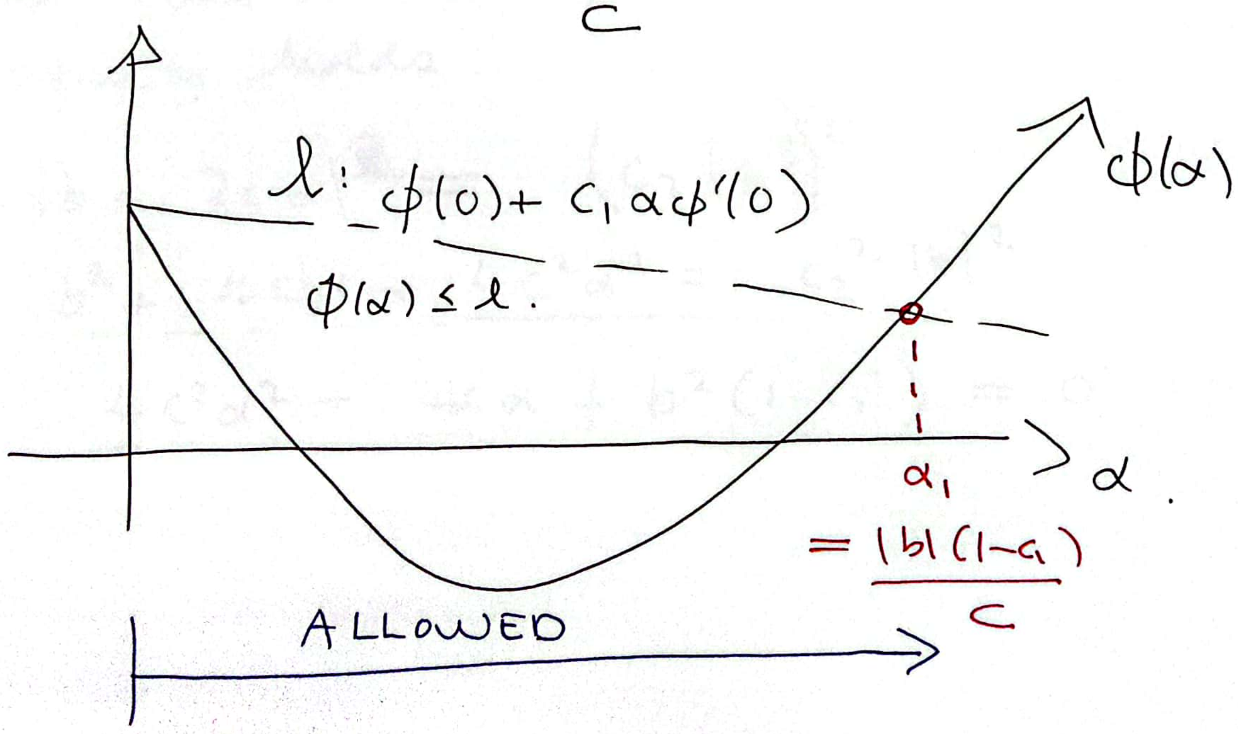
$$\cancel{\Rightarrow b = \alpha . c_1 b - c . \alpha}$$

$$\cancel{= \alpha (c_1 b - c)} .$$

$$b + c\alpha = c_1 b .$$

$$\Rightarrow \quad \alpha = \frac{(c_1 - 1) b}{c}$$

<u>OR</u> $\quad \alpha = \dfrac{|b| (1 - c_1)}{c}$ (positive).



Thus, any $\alpha \in (0, \alpha_1)$ satisfies SWC 1.

Here:

$$\alpha_1 = \frac{|b|(1-c_1)}{c}.$$

We now look at satisfying SWC 2.

SWC 2 says:

$$|\phi'(\alpha)| \leq c_2 |\phi'(0)|$$

$\phi(\alpha) = a + b\alpha + c\alpha^2$

$\phi'(\alpha) = b + 2c\alpha.$

$\phi'(0) = b.$

<u>OR</u> $|\phi'(\alpha)| \leq c_2 |b|.$

~~OR~~ $|\cancel{b+2c\alpha}|$

<u>OR</u>

$$|b + 2c\alpha| \leq c_2 |b|.$$

This is a quadratic inequality in α. To solve, we look at the extreme case, where equality holds:

$$|b + 2c\alpha|^2 = (c_2|b|)^2.$$

$$b^2 + 4cb\alpha + 4c^2\alpha^2 = c_2^2|b|^2.$$

$$\Rightarrow \quad 4c^2\alpha^2 + 4bc\alpha + b^2(1-c_2^2) = 0.$$

CamScanner

Hence:

$$\cancel{\alpha = -4bc \pm \sqrt{16c^2} - 4}$$

$$\alpha = \frac{-4bc \pm \sqrt{16b^2c^2 - 16c^2b^2(1-c_2^2)}}{2 \times 4c^2}$$

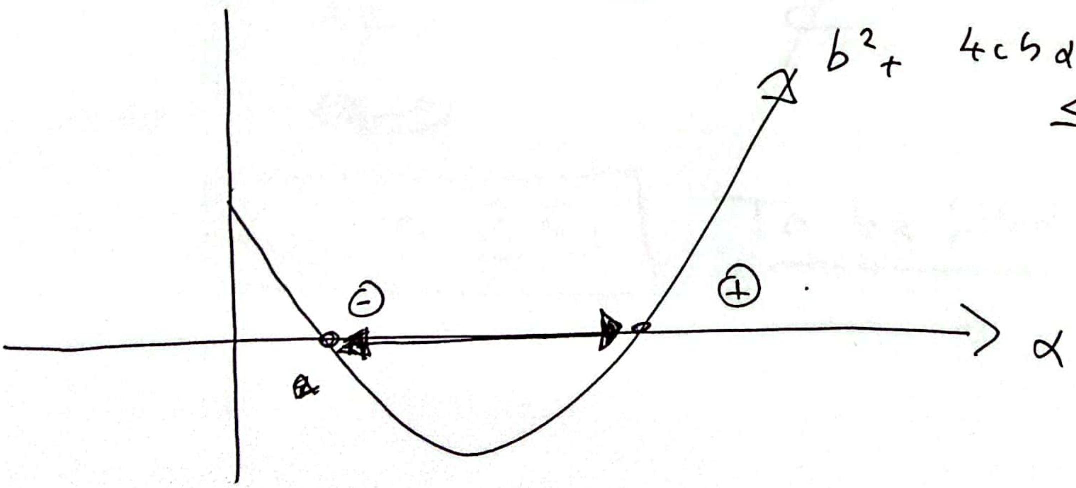
$$\Rightarrow \alpha = \frac{-4bc \pm \sqrt{\cancel{16b^2c^2} - \cancel{16b^2c^2} + 16b^2c^2c_2^2}}{2 \times 4c^2}$$

$$\Rightarrow \alpha = \frac{-4bc \pm 4|b|c\,c_2}{2 \times 4c^2}.$$

$$= \frac{\cancel{4}|b|c \pm \cancel{4}|b|c\,c_2}{2 \times \cancel{4}c^2}.$$

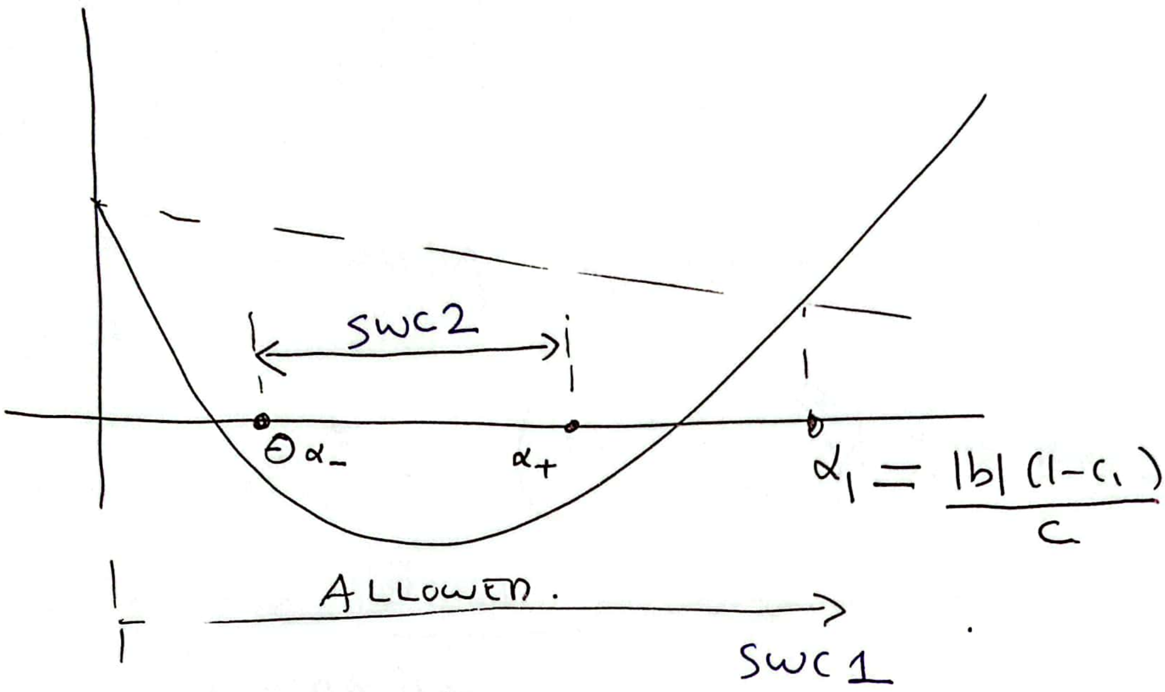
$$\Rightarrow \alpha = \frac{|b|\cancel{c}(1 \pm c_2)}{2 \cdot c \cdot \cancel{c}}$$

$$\Rightarrow \alpha = \frac{|b|(1 \pm c_2)}{2c}.$$



$b^2 + 4cb\alpha + 4c^2\alpha^2 \leq c_2^2 b^2.$

Combine the two figures:



We need α_- to be to the left of α_1. for SWC 1 and SWC 2 to be satisfied simultaneously. So, to produce a contradiction, we would put α_- to the right of α_1:

$$\frac{\cancel{|b|}(1-c_2)}{2\cancel{c}} > \frac{\cancel{|b|}(1-c_1)}{\cancel{c}}.$$

$\Rightarrow$

$$\Rightarrow \boxed{c_2 < 2c_1}$$ To be filled in later.

Question 5.

$$f(\underline{x}) = \langle \underline{g}, \underline{x} \rangle + \frac{1}{2} \langle \underline{x}, B\underline{x} \rangle .$$

B is manifestly symmetric and positive definite.

$$\| \underline{x}_{k+1} - \underline{x}_* \|_2 \leq \left(1 - \frac{1}{\kappa(B)}\right)^{\frac{1}{2}} \| \underline{x}_k - \underline{x}_* \|_2^{1}$$

This is the linear convergence rate for the SD algorithm.

$\kappa(B)$ is the condition number of the matrix.

$$\kappa(B) = \frac{\lambda_{max}}{\lambda_{min}} \text{ for a P.D. symmetric matrix.}$$

For B in Question(5), cond(B) $> 10^5$.

$$1 - \frac{1}{\kappa(B)} \simeq 1 .$$

Hence,

$$\| \underline{x}_{k+1} - \underline{x}_* \|_2 \lesssim 1 \, \| \underline{x}_k - \underline{x}_* \|_2 .$$

No shrinkage of the distance in going from one iteration to the next. SD method fails here.

"SD is to optimization what Euler is to ODEs."