Quick recap:

$$\underline{x}_{k+1} = \underline{x}_k + \alpha_k \underline{p}_k \qquad (1)$$

We want to know the conditions under which

$$\underline{x}_k \to \underline{x}_* \quad \text{as } k \to \infty .$$

Notation:

$$\cos\theta_k = - \frac{\langle p_k, \nabla f_k \rangle}{\|p_k\|_2 \|\nabla f_k\|_2}$$

Theorem (Zoutendijk's Theorem)

Suppose that $\nabla f_k \cdot p_k < 0$ and that the α_k's in (1) satisfy the SWCs. Suppose also that the cost function is sufficiently well behaved, in the sense that:

1. f is bounded below in $\mathbb{R}^n$
2. f is continuously differentiable in an open set $\mathcal{N}$ containing the level sets.

   $$\mathcal{L} = \{ \underline{x} \in \mathbb{R}^n \mid f(\underline{x}) \leq f(\underline{x}_0) \}$$

   where $\underline{x}_0$ is the starting-value in the iterative method (1).
3. ∇f is Lipschitz in $\mathcal{N}$, i.e. there exists a constant $L > 0$ such that

   $$\|\nabla f(\underline{x}) - \nabla f(\underline{y})\|_2 \leq L \|\underline{x} - \underline{y}\|_2$$

   $\forall \underline{x}, \underline{y} \in \mathcal{N}$.

Then:
$$\sum_{k=0}^{\infty} \cos^2\theta_k \, \|\nabla f_k\|_2^2 < \infty . \qquad (2)$$

Corollary 6.1 : Since the series (2) is convergent, the general term goes to zero:
$$\cos^2\theta_k \, \|\nabla f_k\|_2^2 \to 0 \text{ as } k \to \infty .$$

Consequence : If we can keep $\cos^2\theta_k$ away from zero, then $\|\nabla f_k\|_2 \to 0$ as $k \to \infty$, hence x_k tends to a stationary point ($\nabla f = 0$) as $k \to \infty$.

Application : For SD, $\cos\theta_k = 1$, so by Zoutendijk's Theorem, $\|\nabla f_k\|_2 \to 0$ as $k \to \infty$.

So convergence of the SD method is guaranteed, once the conditions in (1) are satisfied.

Today - we look briefly at how to apply Zoutendijk's Theorem to Quasi-Newton methods,
$$p_k = - B_k^{-1} \nabla f_k . \qquad (3)$$

Theorem 6.2 (Examinable) Consider the iterative method (1) where the search direction is given by (3), where B_k is symmetric positive-definite.

by (3), where B_k is symmetric positive-definite.

Suppose B_k satisfies

$$\|B_k\|_2 \|B_k^{-1}\|_2 \leq M$$

where M is a k-independent constant. Then

$$\cos\theta_k \geq \frac{1}{M}.$$

Remark:

$$\kappa(B_k) = \|B_k\|_2 \|B_k^{-1}\|_2$$

is the condition number of the matrix B_k.

Here $\|\circ\|_2$ is the matrix L^2 norm:

$$\|B_k\|_2 = \sup_{\|\underline{u}\|_2=1} \|B_k \underline{u}\|_2$$

When B_k is symmetric positive definite like in the theorem, with eigenvalues $\{\lambda_1, \dots, \lambda_n\}$ and $\lambda_{max} = \max\{\lambda_1, \dots, \lambda_n\}$, then

$$\|B_k\|_2 = \lambda_{max}$$

Similarly, if $\lambda_{min} = \min\{\lambda_1, \dots, \lambda_n\}$, then

$$\|B_n^{-1}\|_2 = \frac{1}{\lambda_{min}}.$$

Hence,

$$\kappa(B_k) = \|B_k\|_2 \|B_k^{-1}\|_2 = \frac{\lambda_{max}}{\lambda_{min}},$$

for B_k symmetric positive-definite.

for B_k symmetric positive-definite.

Back to proof of theorem:

$$\cos\theta_k = - \frac{\langle p_k, \nabla f_k \rangle}{\|p_k\|_2 \, \|\nabla f_k\|_2}$$

$$= + \frac{\langle B_k^{-1}\nabla f_k, \nabla f_k \rangle}{\|B_k^{-1}\nabla f_k\|_2 \, \|\nabla f_k\|_2} \quad \ldots \text{Quasi-Newton methods}$$

Since B_k is symmetric, its eigenvectors form an orthonormal basis for $\mathbb{R}^n$:

$$B_k \underline{u}_i = \lambda_i \underline{u}_i$$

$$\nabla f_k = \sum_{i=1}^{n} x_i \underline{u}_i, \qquad x_i = \langle \underline{u}_i, \nabla f_k \rangle$$

Also,

$$B_k^{-1} \underline{u}_i = \frac{1}{\lambda_i} \underline{u}_i, \quad \text{since } B_k \text{ is positive-definite.}$$

Sub back into expression for $\cos\theta_k$:

$$\cos\theta_k = \frac{\langle B_k^{-1}\nabla f_k, \nabla f_k \rangle}{\|B_k^{-1}\nabla f_k\|_2 \, \|\nabla f_k\|_2}$$

$$= \frac{\left\langle \sum_i \frac{1}{\lambda_i} x_i \underline{u}_i, \sum_j x_j \underline{u}_j \right\rangle}{\left\| \sum_i \frac{1}{\lambda_i} x_i \underline{u}_i \right\|_2 \left\| \sum_i x_i \underline{u}_i \right\|_2}$$

$$= \frac{\sum_i \sum_j \frac{1}{\lambda_i} x_i x_j \langle \underline{u}_i, \underline{u}_j \rangle}{\ldots}$$

$$= \frac{\sum_i \sum_j \frac{1}{\lambda_i} x_i x_j \delta_{ij}}{\cdots}$$

Here, $\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$

Hence,

$$\cos\theta_k = \frac{\sum_i \frac{1}{\lambda_i} x_i^2}{\left(\sum_i x_i^2\right)^{1/2} \left(\sum_i \frac{1}{\lambda_i^2} x_i^2\right)^{1/2}}$$

$$\geq \frac{\frac{1}{\lambda_{max}} \sum_i x_i^2}{\left(\sum_i x_i^2\right)^{1/2} \left(\sum_i \frac{x_i^2}{\lambda_i^2}\right)^{1/2}}$$

$$\geq \frac{\frac{1}{\lambda_{max}} \cancel{\sum x_i^2}}{\cancel{\left(\sum_i x_i^2\right)^{1/2}} \frac{1}{\lambda_{min}} \cancel{\left(\sum_i x_i^2\right)^{1/2}}}$$

$$= \frac{1/\lambda_{max}}{1/\lambda_{min}}$$

$$= \frac{\lambda_{min}}{\lambda_{max}}$$

$$= \frac{1}{\kappa(B_k)}$$

$$\geq \frac{1}{M}.$$

$\kappa(B_k) \leq M$

$\frac{1}{\kappa(B_k)} \geq \frac{1}{M}$

Thus,

$$\cos\theta_k \geq \frac{1}{M}. \qquad \blacksquare$$

Corollary again: By Zoutendijk's Theorem, the Quasi-Newton methods (with conditions outlined in Theorem 6.2) has the property that

$$\|\nabla f_k\|_2 \to 0 \text{ as } k \to \infty.$$