Line search methods: $x_{k+1} = x_k + \alpha_k p_k$.

- Theoretical considerations of the step length.
- Global convergence of line search methods.

Two desirable properties of a line search algorithm: a guaranteed *global convergence* and a rapid *rate of convergence*.
Step length
Main point: to describe *easily verifiable* theoretical conditions on step length, that *allow to prove* convergence of an algorithm.

· "Significantly" decrease the objective function
· Assess whether there is even more "significant" decrease at a different step length.

Sufficient decrease or *Armijo condition*

$$f(x_k + \alpha p_k) \leq f(x_k) + c_1 \alpha \langle \nabla f(x_k), p_k \rangle, \;\; c_1 \approx 10^{-4}.$$

Curvature condition

$$\langle \nabla f(x_k + \alpha p_k), p_k \rangle \geq c_2 \langle \nabla f(x_k), p_k \rangle, \;\; c_2 \geq c_1, c_2 \approx .9$$

for Newton or Quasi-Newton methods.
Note. For many methods, e.g. Newton or Quasi-Newton methods, it may be that $p_k \neq 1$.
Wolfe condition= sufficient decrease + curvature condition.
Lemma For the class of problems we consider there are always points that satisfy the Wolfe condition.
Proof: Mean-value theorem.
Other methods: *strong Wolfe condition* is when the curvature condition is replaced by

$$|\langle \nabla f(x_k + \alpha p_k), p_k \rangle| \leq c_2 |\langle \nabla f(x_k), p_k \rangle|,$$

Goldstein condition:

$$f(x_k) + (1-c)\alpha \langle \nabla f(x_k), p_k \rangle \leq f(x_k + \alpha p_k) \leq f(x_k) + c\alpha \langle \nabla f(x_k), p_k \rangle, \;\; 0 < c < 1/2.$$

Backtracking line search $\approx$ just *sufficient decrease*: Take $\alpha_0 = 1$, not good enough $\alpha_1 = \rho \alpha_0$, not good enough $\alpha_2 = \rho \alpha_1$, ... where $\rho < 1$.
Next goal is to capitalize on one of these conditions to **prove convergence**. Line search methods have only *global convergence*

$$|\nabla f(x_k)| \to 0$$

can be proved.
Question Why *only* global convergence is not good enough for us?
Theorem (Zoutendijk)
Denote by

$$\cos \Theta_k = -\frac{\langle \nabla f(x_k), p_k \rangle}{||\nabla f(x_k)|| \, ||p_k||},$$

the angle between the search directions and the steepest descent direction, for the class of problems we consider, a line search that satisfies the Wolfe condition also satisfies

$$\sum_k \cos^2 \Theta_k ||\nabla f(x_k)||^2 < \infty.$$

Corollary:

$$\cos^2 \Theta_k ||\nabla f(x_k)||^2 \to 0.$$

Question. How we can derive from here global convergence?
For a quasi-Newton method a good requirement (see homework this week) is to assume a uniform bound on the condition number

$$||B_k|| \, ||B_k^{-1}|| \le M, \text{ for all } k.$$

Some algorithms, such as *conjugate gradient method*, give a weaker global convergence result:

$$\liminf_k ||\nabla f(x_k)|| = 0.$$