# Chapter 7

# Line Search Methods

# The strategy and the key objects

**Problem.** Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be a $\mathbb{C}^1$ function. To solve

$$\min_{x \in \mathbb{R}^n} f(x)$$

it is necessary to find out points (vectors) $x^\star$ such that $\nabla f(x^\star) = 0$.

**Strategy (Line Search Methods).** A possible strategy for doing so is to start at a given vector $x_0 \in D$ and construct a sequence

$$x_k = \min_{\alpha_k \in \mathbb{R}} f(x_{k-1} + \alpha_k p_k), \quad \text{with } p_k \in \mathbb{R}^n$$

such that $x_k \to x^\star$ with $\nabla f(x^\star) = 0$. We want to choose $\alpha_k$ (the step) and $p_k$ (the line direction) at each step so that the convergence is optimal.

# The direction

**Theorem.** Let $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ be a differentiable function and let $a \in D$ and $\boldsymbol{u} \in \mathbb{R}^n$ be an unitary vector. Suppose that $\theta$ is the angle between $\boldsymbol{u}$ and $\nabla f(\boldsymbol{a})$. Then

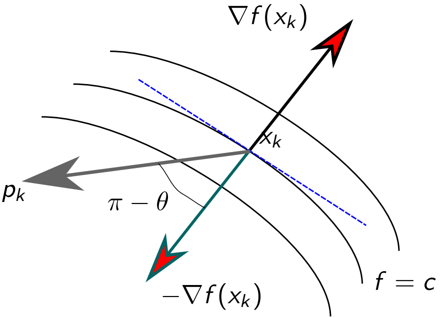
$$D_{\boldsymbol{u}} f(\boldsymbol{a}) =< (\nabla f(\boldsymbol{a})), \boldsymbol{u} >= \boldsymbol{u}^T \nabla f(\boldsymbol{a}) = \|\nabla f(\boldsymbol{a})\| \cos\theta.$$

In particular the vector $-\nabla f(\boldsymbol{a})$ gives the maximum descent direction of $f$ at the point $\boldsymbol{a}$.

# The direction $p_k$

**Definition.** We say that $p_k$ is a descent direction if $p_k^T \nabla f(\boldsymbol{x}_k) < 0$.
More generically (in line search methods) we consider

$$p_k = -B_k^{-1} \nabla f(\boldsymbol{x}_k) \qquad \text{with } B_k \text{ positive definite.}$$




- $B_k = \mathrm{Id}$ (descent method)
- $B_k = \nabla^2 f(\boldsymbol{x}_k)$ (Newton method)
- $B_k \approx \nabla^2 f(\boldsymbol{x}_k)$ (quasi Newton method)

# The step size $\alpha_k$

Formally at each $k$-step we are finding a a solution of

$$\min_{\alpha \in \mathbb{R}^+} \; f\left(x_k + \alpha p_k\right).$$

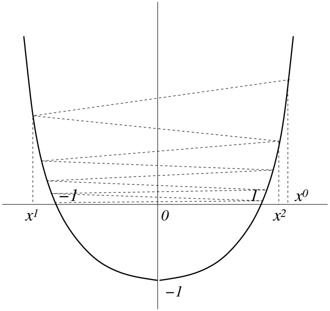
But we want to decide the value of $\alpha$ as fast as possible at each step. We are looking for a minimal cost to choose $\alpha$. In other words we want to have a easy way to terminate our finding of $\alpha$, and move forward to the next step.

A philosophical approach would be to (a) find an interval containing the desirable steps and (b) use a bisection method to conclude the desires $\alpha$.

# The step size $\alpha_k$

First tentative. We want to terminate the process at each step $k$ when we find $\alpha_k$ such that

$$f\left(x_k + \alpha_k p_k\right) < f\left(x_k\right).$$



# The step size $\alpha_k$: Sufficient decrease condition

Second tentative. We impose the following condition for $\alpha_k$

$$\phi(\alpha_k) := f(x_k + \alpha_k p_k) < f(x_k) + c_1 \alpha_k (\nabla f(x_k))^T p_k, \; c_1 \in (0,1).$$

The condition is called (sufficient decrease condition).

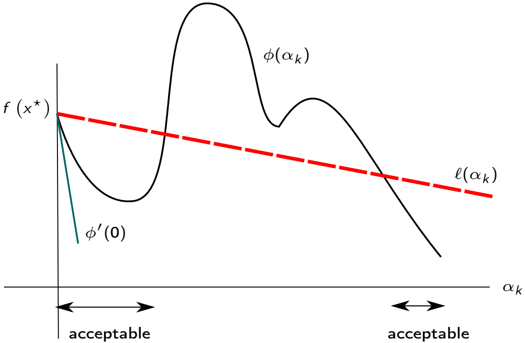
Remarks.

- $\ell(\alpha_k) := f(x_k) + c_1 \alpha_k \nabla f^T(x_k) p_k$ is a linear function.
- For small values of $\alpha_k > 0$ we have $\phi(\alpha_k) < \ell(\alpha_k)$. This is so because $c_1 \in (0,1)$ and then

$$\phi'(0) = (\nabla f(x_k))^T p_k < c_1 (\nabla f(x_k))^T p_k = \ell'(0) < 0.$$

# The step size $\alpha_k$

Sufficient decrease. We ask for a decrease proportional to $\alpha$ and $\phi'(0) = \nabla f^T(x_k)\, p_k$. Usually $c_1 \approx 0.1$.

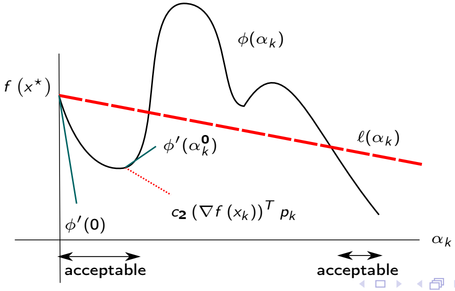


# The step size $\alpha_k$: curvature condition

Curvature condition. Since the previous condition is always satisfied for small values of $\alpha_k$ we need to add further conditions for termination. We use the so called curvature condition

$$(\nabla f(x_k + \alpha_k p_k))^T p_k \geq c_2 (\nabla f(x_k))^T p_k, \; c_2 \in (c_1, 1)$$

In other words if $\phi'(\alpha_k)$ is not negative enough we terminate the $k$-step.



# The step size $\alpha_k$: (strong) Wolfe Conditions

**Definition.** The conditions (together) to terminate the $k$-step given by

$$f\left(x_k+\alpha_k p_k\right) < f\left(x_k\right)+c_1\alpha_k\left(\nabla f\left(x_k\right)\right)^T p_k,$$
$$\left(\nabla f\left(x_k+\alpha_k p_k\right)\right)^T p_k \geq c_2\left(\nabla f\left(x_k\right)\right)^T p_k,$$

with $0 < c_1 < c_2 < 1$ are usually called Wolfe conditions.

**Definition.** The conditions (together) to terminate the $k$-step given by (we do not allow $\phi'\left(\alpha_k\right)$ to be too positive).

$$f\left(x_k+\alpha_k p_k\right) < f\left(x_k\right)+c_1\alpha_k\left(\nabla f\left(x_k\right)\right)^T p_k,$$
$$\left|\left(\nabla f\left(x_k+\alpha_k p_k\right)\right)^T p_k\right| \leq \left|c_2\left(\nabla f\left(x_k\right)\right)^T p_k\right|,$$

with $0 < c_1 < c_2 < 1$ are usually called strong Wolfe conditions.

## The step size $\alpha_k$: Existence

**Lemma.** Suppose $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be a $\mathcal{C}^1$ function. Let $p_k$ a descent direction at the point $x_k \in D$ and assume $f|L_{p_k}$ is bounded below where $L_{p_k} = \{x \in \mathbb{R}^n \mid x = x_k + \alpha p_k,\ \alpha > 0\}$. Then if $0 < c_1 < c_2 < 1$ there exist intervals of step lengths satisfying the (strong) Wolfe conditions

**Proof.** Since $\ell'(\alpha_k) < 0$ (and constant) there exists a first intersection, $\hat{\alpha}_k > 0$, between $\ell(\alpha_k)$ and $\phi(\alpha_k)$:

$$f(x_k + \hat{\alpha}_k p_k) = f(x_k) + c_1 \hat{\alpha}_k (\nabla f(x_k))^T p_k. \tag{1}$$

The sufficient decrease condition it is satisfied for all $\alpha_k \in [0, \hat{\alpha}_k]$. By the Mean Value Theorem we have that there exists $\tilde{\alpha}_k \in [0, \hat{\alpha}_k]$ such that

$$f(x_k + \hat{\alpha}_k p_k) - f(x_k) = \hat{\alpha}_k (\nabla f(x_k + \tilde{\alpha}_k p_k))^T p_k$$

All together imply

$$(\nabla f(x_k + \tilde{\alpha}_k p_k))^T p_k = c_1 \hat{\alpha}_k (\nabla f(x_k))^T p_k > c_2 \hat{\alpha}_k (\nabla f(x_k))^T p_k.$$

Therefore $\tilde{\alpha}_k$ satisfies the Wolfe conditions and smoothness gives the desired interval.

# Convergence of line search methods

**Remark.** Until this moment we just consider the definition of the process, that is the election of $p_k$ and $\alpha_k$. But we need to study if the process converge to somewhere.

Let $p_k$ be a descent direction, and let $\theta_k$ the angle of $p_k$ and $-\nabla f(x^\star)$

$$\cos(\theta_k) = -\frac{1}{||\nabla f(x_k)||\ ||p_k||} \left(\nabla f(x_k)\right)^T p_k$$

**Theorem.** Assume notation above with $p_k$ a descent direction and $\alpha_k$ satisfying Wolfe's conditions. Suppose $f$ is $\mathcal{C}^2$ and bounded below in $\mathbb{R}^n$. Then

$$\sum_{k=0}^{\infty} \cos^2(\theta_k)||\nabla f(x_k)|| < \infty. \tag{2}$$

# Convergence of line search methods

**Corollary.** Under the above notation and assumptions we have

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

Moreover if there exists $\delta > 0$ such that $\cos(\theta) > \delta$ then

$$\lim_{k\to\infty} ||\nabla f(x_k)|| = 0 \quad \text{(globally convergent algorithms)}$$

**Remark.** The final $\delta$-condition basically means that $p_k$ do not get arbitrarily orthogonal to the gradient vector. This is, for instance, the case of the steepest descent method.

# Convergence of line search methods: Newton's like methods

Assume that the matrices $B_k, \ k \geq 0$ which define the (Newton-like) direction $p_k = -B_k^{-1}\, \nabla f\,(\boldsymbol{x}_k)$ are uniformly positively definite

$$||B_k||\ ||B_k^{-1}|| \leq M, \quad \forall k \geq 0.$$

Lemma. Under the assumptions we have that

$$\cos(\theta_k) \geq \frac{1}{M},$$

and so

$$\lim_{k\to\infty} ||\nabla f\,(x_k)\,|| = 0.$$

# Convergence of line search methods: Final comments

Remark. We have shown that under the above hypothesis the line search method converge to an stationary point: $\nabla f(x^\star) = 0$. But this is not a guarantee that $x^\star$ is a minimizer. For this we need to add other conditions on the Hessian of $f$ at $x = x^\star$.

Remark. Another consideration is about the speed or rate of convergence. The asymptotic behaviour is the desired one but what about the number of iterates?

# Rate of convergence