

Unconstrained minimization

Lectures for PHD course on
Unconstrained Numerical Optimization

Enrico Bertolazzi

DIMS – Università di Trento

May 2008

- 1 General iterative scheme
 - Descent direction failure
- 2 Backtracking Armijo line-search
 - Global convergence of backtracking Armijo line-search
 - Global convergence of steepest descent
- 3 Wolfe–Zoutendijk global convergence
 - The Wolfe conditions
 - The Armijo-Goldstein conditions
- 4 Algorithms for line-search
 - Armijo Parabolic-Cubic search
 - Wolfe linesearch



Given $f : \mathbb{R}^n \mapsto \mathbb{R}$:

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \quad f(\mathbf{x})$$

the following regularity about $f(\mathbf{x})$ is assumed in the following:

Assumption (Regularity assumption)

We assume $f \in C^1(\mathbb{R}^n)$ with Lipschitz continuous gradient, i.e. there exists $\gamma > 0$ such that

$$\|\nabla f(\mathbf{x})^T - \nabla f(\mathbf{y})^T\| \leq \gamma \|\mathbf{x} - \mathbf{y}\|, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$



Definition (Global minimum)

Given $f : \mathbb{R}^n \mapsto \mathbb{R}$ a point $\mathbf{x}_\star \in \mathbb{R}^n$ is a *global minimum* if

$$f(\mathbf{x}_\star) \leq f(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

Definition (Local minimum)

Given $f : \mathbb{R}^n \mapsto \mathbb{R}$ a point $\mathbf{x}_\star \in \mathbb{R}^n$ is a *local minimum* if

$$f(\mathbf{x}_\star) \leq f(\mathbf{x}), \quad \forall \mathbf{x} \in B(\mathbf{x}_\star; \delta).$$

Obviously a global minimum is a local minimum. Find a global minimum in general is not an easy task. The algorithms presented in the sequel will approximate *local minima's*.

Definition (Strict global minimum)

Given $f : \mathbb{R}^n \mapsto \mathbb{R}$ a point $\mathbf{x}_\star \in \mathbb{R}^n$ is a *strict global minimum* if

$$f(\mathbf{x}_\star) < f(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{x}_\star\}.$$

Definition (Strict local minimum)

Given $f : \mathbb{R}^n \mapsto \mathbb{R}$ a point $\mathbf{x}_\star \in \mathbb{R}^n$ is a *strict local minimum* if

$$f(\mathbf{x}_\star) < f(\mathbf{x}), \quad \forall \mathbf{x} \in B(\mathbf{x}_\star; \delta) \setminus \{\mathbf{x}_\star\}.$$

Obviously a strict global minimum is a strict local minimum.

First order Necessary condition

Lemma (First order Necessary condition for local minimum)

Given $f : \mathbb{R}^n \mapsto \mathbb{R}$ satisfying the regularity assumption. If a point $\mathbf{x}_\star \in \mathbb{R}^n$ is a **local minimum** then

$$\nabla f(\mathbf{x}_\star)^T = \mathbf{0}.$$

Proof.

Consider a generic direction \mathbf{d} , then for δ small enough we have

$$\lambda^{-1}(f(\mathbf{x}_\star + \lambda\mathbf{d}) - f(\mathbf{x}_\star)) \leq 0, \quad 0 < \lambda < \delta$$

so that

$$\lim_{\lambda \rightarrow 0} \lambda^{-1}(f(\mathbf{x}_\star + \lambda\mathbf{d}) - f(\mathbf{x}_\star)) = \nabla f(\mathbf{x}_\star)\mathbf{d} \leq 0,$$

because \mathbf{d} is a generic direction we have $\nabla f(\mathbf{x}_\star)^T = \mathbf{0}$. □



Remark

- ① *The first order necessary condition do not discriminate maximum, minimum, or saddle points.*
- ② *To discriminate maximum and minimum we need more information, e.g. second order derivative of $f(\mathbf{x})$.*
- ③ *With second order derivative we can build **necessary** and **sufficient** condition for a minima.*
- ④ *In general using only first and second order derivative at the point \mathbf{x}_* it is not possible to deduce a **necessary and sufficient** condition for a minima.*

Second order Necessary condition

Lemma (Second order Necessary condition for local minimum)

Given $f \in \mathcal{C}^2(\mathbb{R}^n)$ if a point $\mathbf{x}_\star \in \mathbb{R}^n$ is a *local minimum* then $\nabla f(\mathbf{x}_\star)^T = \mathbf{0}$ and $\nabla^2 f(\mathbf{x}_\star)$ is *semi-definite positive*, i.e.

$$\mathbf{d}^T \nabla^2 f(\mathbf{x}_\star) \mathbf{d} \geq 0, \quad \forall \mathbf{d} \in \mathbb{R}^n$$

Example

This condition is only, necessary, in fact consider $f(\mathbf{x}) = x_1^2 - x_2^3$,

$$\nabla f(\mathbf{x}) = (2x_1, -3x_2^2), \quad \nabla^2 f(\mathbf{x}) = \begin{pmatrix} 2 & 0 \\ 0 & -6x_2 \end{pmatrix}$$

for the point $\mathbf{x}_\star = \mathbf{0}$ we have $\nabla f(\mathbf{0}) = \mathbf{0}$ and $\nabla^2 f(\mathbf{0})$ semi-definite positive, but $\mathbf{0}$ is a saddle point not a minimum.

Proof.

The condition $\nabla f(\mathbf{x}_\star)^T = \mathbf{0}$ comes from first order necessary conditions. Consider now a generic direction \mathbf{d} , and the finite difference:

$$\frac{f(\mathbf{x}_\star + \lambda\mathbf{d}) - 2f(\mathbf{x}_\star) + f(\mathbf{x}_\star - \lambda\mathbf{d})}{\lambda^2} \geq 0$$

by using Taylor expansion for $f(\mathbf{x})$

$$f(\mathbf{x}_\star \pm \lambda\mathbf{d}) = f(\mathbf{x}_\star) \pm \nabla f(\mathbf{x}_\star)\lambda\mathbf{d} + \frac{\lambda^2}{2}\mathbf{d}^T\nabla^2 f(\mathbf{x}_\star)\mathbf{d} + o(\lambda^2)$$

and from the previous inequality

$$\mathbf{d}^T\nabla^2 f(\mathbf{x}_\star)\mathbf{d} + 2o(\lambda^2)/\lambda^2 \geq 0$$

taking the limit $\lambda \rightarrow 0$ and from the arbitrariness of \mathbf{d} we have that $\nabla^2 f(\mathbf{x}_\star)$ must be semi-definite positive. □

Second order sufficient condition

Lemma (Second order sufficient condition for local minimum)

Given $f \in \mathcal{C}^2(\mathbb{R}^n)$ if a point $\mathbf{x}_\star \in \mathbb{R}^n$ satisfy:

- 1 $\nabla f(\mathbf{x}_\star)^T = \mathbf{0}$;
- 2 $\nabla^2 f(\mathbf{x}_\star)$ is *definite positive*; i.e.

$$\mathbf{d}^T \nabla^2 f(\mathbf{x}_\star) \mathbf{d} > 0, \quad \forall \mathbf{d} \in \mathbb{R}^n \setminus \{\mathbf{x}_\star\}$$

then $\mathbf{x}_\star \in \mathbb{R}^n$ is a *strict local minimum*.

Remark

Because $\nabla^2 f(\mathbf{x}_\star)$ is symmetric we can write

$$\lambda_{\min} \mathbf{d}^T \mathbf{d} \leq \mathbf{d}^T \nabla^2 f(\mathbf{x}_\star) \mathbf{d} \leq \lambda_{\max} \mathbf{d}^T \mathbf{d}$$

If $\nabla^2 f(\mathbf{x}_\star)$ is positive definite we have $\lambda_{\min} > 0$.



Proof.

Consider now a generic direction \mathbf{d} , and the Taylor expansion for $f(\mathbf{x})$

$$\begin{aligned} f(\mathbf{x}_* + \mathbf{d}) &= f(\mathbf{x}_*) + \nabla f(\mathbf{x}_*)\mathbf{d} + \frac{1}{2}\mathbf{d}^T \nabla^2 f(\mathbf{x}_*)\mathbf{d} + o(\|\mathbf{d}\|^2) \\ &\geq f(\mathbf{x}_*) + \frac{1}{2}\lambda_{\min} \|\mathbf{d}\|^2 + o(\|\mathbf{d}\|^2) \\ &\geq f(\mathbf{x}_*) + \frac{1}{2}\lambda_{\min} \|\mathbf{d}\|^2 \left(1 + o(\|\mathbf{d}\|^2)/\|\mathbf{d}\|^2\right) \end{aligned}$$

choosing \mathbf{d} small enough we can write

$$f(\mathbf{x}_* + \mathbf{d}) \geq f(\mathbf{x}_*) + \frac{1}{4}\lambda_{\min} \|\mathbf{d}\|^2 > f(\mathbf{x}_*), \quad \mathbf{d} \neq \mathbf{0}, \|\mathbf{d}\| \leq \delta.$$

i.e. \mathbf{x}_* is a strict minimum. □

Outline

- 1 General iterative scheme
 - Descent direction failure
- 2 Backtracking Armijo line-search
 - Global convergence of backtracking Armijo line-search
 - Global convergence of steepest descent
- 3 Wolfe–Zoutendijk global convergence
 - The Wolfe conditions
 - The Armijo-Goldstein conditions
- 4 Algorithms for line-search
 - Armijo Parabolic-Cubic search
 - Wolfe linesearch

How to find a minimum

Given $f : \mathbb{R}^n \mapsto \mathbb{R}$: minimize $_{\mathbf{x} \in \mathbb{R}^n}$ $f(\mathbf{x})$.

- 1 We can solve the problem by solving the **necessary condition**.
i.e by solving the nonlinear systems

$$\nabla f(\mathbf{x})^T = \mathbf{0}.$$

- 2 Using such an approach we loses the information about $f(\mathbf{x})$.
- 3 Moreover such an approach can find solution corresponding to a maximum or saddle points.
- 4 A better approach is to use all the information and try to build **minimizing procedure**, i.e. procedures that, starting from a point \mathbf{x}_0 build a sequence $\{\mathbf{x}_k\}$ such that $f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k)$. In this way, at least, we avoid to converge to a **strict maximum**.



Iterative Methods

- In practice, rarely we are able to provide an explicit minimizer.
- Iterative method: given starting **guess** x_0 , generate the sequence,

$$\{x_k\}, \quad k = 1, 2, \dots$$

- **AIM:** ensure that (a subsequence) has some favorable limiting properties:
 - satisfies first-order necessary conditions
 - satisfies second-order necessary conditions



Line-search Methods

A generic iterative minimization procedure can be sketched as follows:

- calculate a **search direction** \mathbf{p}_k from \mathbf{x}_k
- ensure that this direction is a **descent direction**, i.e.

$$\nabla f(\mathbf{x}_k)\mathbf{p}_k < 0, \quad \text{whenever } \nabla f(\mathbf{x}_k)^T \neq \mathbf{0}$$

so that, at least for small steps along \mathbf{p}_k , the objective function $f(\mathbf{x})$ will be reduced

- use **line-search** to calculate a suitable step-length $\alpha_k > 0$ so that

$$f(\mathbf{x}_k + \alpha_k\mathbf{p}_k) < f(\mathbf{x}_k).$$

- Update the point:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k\mathbf{p}_k$$



Generic minimization algorithm

Written with a pseudo-code the minimization procedure is the following algorithm:

Generic minimization algorithm

Given an initial guess \mathbf{x}_0 , let $k = 0$;

while not converged **do**

 Find a descent direction \mathbf{p}_k at \mathbf{x}_k ;

 Compute a step size α_k using a line-search along \mathbf{p}_k .

 Set $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$ and increase k by 1.

end while

The crucial points which differentiate the algorithms are:

- 1 The computation of the direction \mathbf{p}_k ;
- 2 The computation of the step size α_k .

Practical Line-search methods

- The first developed minimization algorithms try to solve

$$\alpha_k = \arg \min_{\alpha > 0} f(\mathbf{x}_k + \alpha \mathbf{p}_k)$$

- performing **exact line-search** by univariate minimization;
 - rather expensive and certainly not cost effective.
-
- Modern methods implements **inexact** line-search:
 - ensure steps are neither too long nor too short
 - try to pick *useful* initial step size for fast convergence
 - best methods are based on:
 - backtracking–Armijo search;
 - Armijo–Goldstein search;
 - Franke–Wolfe search;



backtracking line-search

To obtain a monotone decreasing sequence we can use the following algorithm:

Backtracking line-search

```

Given  $\alpha_{\text{init}}$  (e.g.,  $\alpha_{\text{init}} = 1$ );
Given  $\tau \in (0, 1)$  typically  $\tau = 0.5$ ;
Let  $\alpha^{(0)} = \alpha_{\text{init}}$ ;
while not  $f(\mathbf{x}_k + \alpha^{(\ell)} \mathbf{p}_k) < f(\mathbf{x}_k)$  do
    set  $\alpha^{(\ell+1)} = \tau \alpha^{(\ell)}$ ;
    increase  $\ell$  by 1;
end while
Set  $\alpha_k = \alpha^{(\ell)}$ .

```

To be effective the previous algorithm should terminate in a finite number of steps. In the following we prove that if \mathbf{p}_k is a descent direction then a slight modification of the algorithm will terminate.



Existence of a descent step

(1/3)

Lemma (Descent Lemma)

Suppose that $f(x)$ satisfy the standard assumptions and that \mathbf{p}_k is a descent direction at \mathbf{x}_k , i.e. $\nabla f(\mathbf{x}_k)\mathbf{p}_k < 0$. Then we have

$$f(\mathbf{x}_k + \alpha\mathbf{p}_k) \leq f(\mathbf{x}_k) + \alpha\nabla f(\mathbf{x}_k)\mathbf{p}_k + \frac{\gamma}{2}\alpha^2 \|\mathbf{p}_k\|^2$$

for all $\alpha \in [0, \alpha_k^*]$ where $\alpha_k^* = \frac{-2\nabla f(\mathbf{x}_k)\mathbf{p}_k}{\gamma \|\mathbf{p}_k\|^2} > 0$

Assumption (Regularity assumption)

We assume $f \in \mathcal{C}^1(\mathbb{R}^n)$ with Lipschitz continuous gradient, i.e. there exists $\gamma > 0$ such that

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq \gamma \|\mathbf{x} - \mathbf{y}\|, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$



Existence of a descent step

(2/3)

Proof.

Let be $g(\alpha) = f(\mathbf{x}_k + \alpha \mathbf{p}_k)$ then we can write:

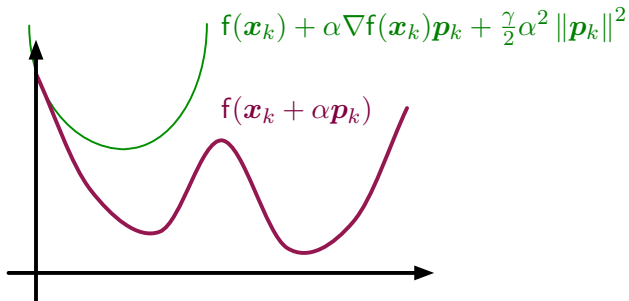
$$\begin{aligned} g(\alpha) - g(0) &= \int_0^\alpha g'(\xi) d\xi = \alpha g'(0) + \int_0^\alpha (g'(\xi) - g'(0)) d\xi \\ &= \alpha \nabla f(\mathbf{x}_k) \mathbf{p}_k + \int_0^\alpha (\nabla f(\mathbf{x}_k + \xi \mathbf{p}_k) - \nabla f(\mathbf{x}_k)) \mathbf{p}_k d\xi \\ &\leq \alpha \nabla f(\mathbf{x}_k) \mathbf{p}_k + \int_0^\alpha \|\nabla f(\mathbf{x}_k + \xi \mathbf{p}_k) - \nabla f(\mathbf{x}_k)\| \|\mathbf{p}_k\| d\xi \\ &\leq \alpha \nabla f(\mathbf{x}_k) \mathbf{p}_k + \|\mathbf{p}_k\|^2 \int_0^\alpha \gamma \xi d\xi \\ &\leq \alpha \nabla f(\mathbf{x}_k) \mathbf{p}_k + \frac{\gamma \alpha^2}{2} \|\mathbf{p}_k\|^2 = \alpha \left[\nabla f(\mathbf{x}_k) \mathbf{p}_k + \frac{\gamma \alpha}{2} \|\mathbf{p}_k\|^2 \right]. \end{aligned}$$

now the lemma follows trivially. □

Existence of a descent step

(3/3)

- The **descent lemma** means that there is a parabola that is entirely over the function $f(\mathbf{x})$ in the direction \mathbf{p}_k if this is a descent direction.
- The second part of the lemma permits to ensure a **minimal** reduction if the step length is chosen to be $\alpha_k = \alpha_k^*/2$.



Descent direction failure

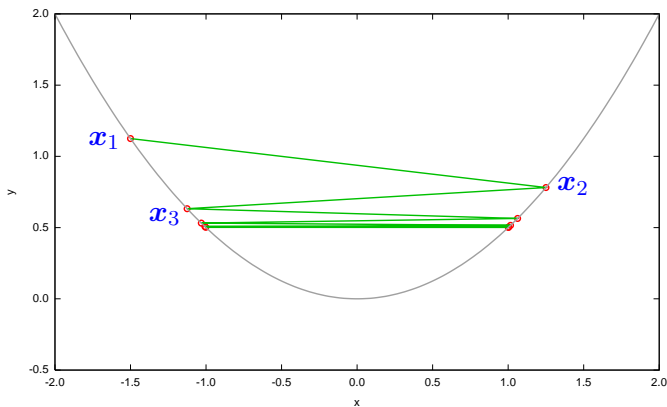
- The simple request to have a descent direction may be not enough.
- In fact, step length may be asymptotically **too short**
- Or step length may be asymptotically **too long**



Steps may be too long

The objective function is $f(x) = x^2$ and the iterates are generated by the descent directions $p_k = (-1)^{k+1}$ from $x_0 = 2$ with:

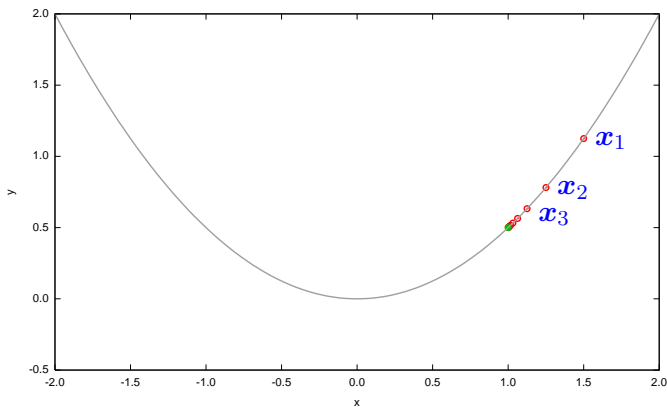
$$x_{k+1} = x_k + \alpha_k p_k, \quad \alpha_k = 2 + 3 \cdot 2^{-(k+1)}$$



Steps may be too short

The objective function is $f(x) = x^2$ and the iterates are generated by the descent directions $p_k = -1$ from $x_0 = 2$ with:

$$x_{k+1} = x_k + \alpha_k p_k, \quad \alpha_k = 2^{-(k+1)}$$



Outline

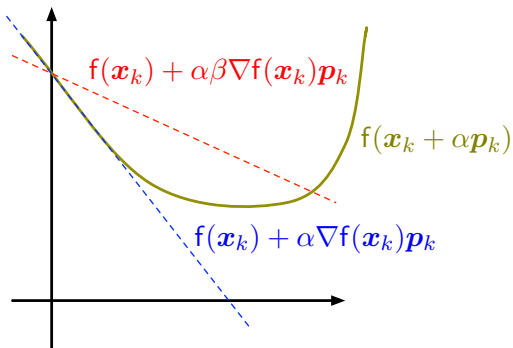
- 1 General iterative scheme
 - Descent direction failure
- 2 Backtracking Armijo line-search
 - Global convergence of backtracking Armijo line-search
 - Global convergence of steepest descent
- 3 Wolfe–Zoutendijk global convergence
 - The Wolfe conditions
 - The Armijo-Goldstein conditions
- 4 Algorithms for line-search
 - Armijo Parabolic-Cubic search
 - Wolfe linesearch

Armijo condition

To prevent large steps relative to the decreasing of $f(\mathbf{x})$ we require that

$$f(\mathbf{x}_k + \alpha_k \mathbf{p}_k) \leq f(\mathbf{x}_k) + \alpha_k \beta \nabla f(\mathbf{x}_k) \mathbf{p}_k$$

for some $\beta \in (0, 1)$. Typical values of β ranges from 10^{-4} to 0.1.



Backtracking Armijo line-search

Given α_{init} (e.g., $\alpha_{\text{init}} = 1$);

Given $\tau \in (0, 1)$ typically $\tau = 0.5$;

Let $\alpha^{(0)} = \alpha_{\text{init}}$;

while not $f(\mathbf{x}_k + \alpha^{(\ell)}\mathbf{p}_k) \leq f(\mathbf{x}_k) + \alpha^{(\ell)}\beta\nabla f(\mathbf{x}_k)\mathbf{p}_k$ **do**

 set $\alpha^{(\ell+1)} = \tau\alpha^{(\ell)}$;

 increase ℓ by 1;

end while

Set $\alpha_k = \alpha^{(\ell)}$.

- Backtracking Armijo line-search prevents the step from getting too large.
- Now the question is: will the backtracking Armijo line-search **terminate** in a finite number of steps ?

Finite termination of Armijo line-search

Theorem (Finite termination of Armijo linesearch)

Suppose that $f(x)$ satisfy the standard assumptions and $\beta \in (0, 1)$ and that \mathbf{p}_k is a descent direction at \mathbf{x}_k . Then the Armijo condition

$$f(\mathbf{x}_k + \alpha_k \mathbf{p}_k) \leq f(\mathbf{x}_k) + \alpha_k \beta \nabla f(\mathbf{x}_k) \mathbf{p}_k$$

is satisfied when $\alpha_k \in [0, \omega_k]$ where
$$\omega_k = \frac{2(\beta - 1) \nabla f(\mathbf{x}_k) \mathbf{p}_k}{\gamma \|\mathbf{p}_k\|^2}$$

Assumption (Regularity assumption)

We assume $f \in C^1(\mathbb{R}^n)$ with Lipschitz continuous gradient, i.e. there exists $\gamma > 0$ such that

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq \gamma \|\mathbf{x} - \mathbf{y}\|, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$



Finite termination of Armijo line-search

To prove finite termination we need the following Taylor expansion due to the regularity assumption:

$$f(\mathbf{x} + \alpha\mathbf{p}) = f(\mathbf{x}) + \alpha\nabla f(\mathbf{x})\mathbf{p} + E \quad \text{where} \quad |E| \leq \frac{\gamma}{2}\alpha^2 \|\mathbf{p}\|^2$$

Proof.

If $\alpha \leq \omega_k$ we have $\alpha\gamma \|\mathbf{p}_k\|^2 \leq 2(\beta - 1)\nabla f(\mathbf{x}_k)\mathbf{p}_k$ and by using Taylor expansion

$$\begin{aligned} f(\mathbf{x}_k + \alpha\mathbf{p}_k) &\leq f(\mathbf{x}_k) + \alpha\nabla f(\mathbf{x}_k)\mathbf{p}_k + \frac{\gamma}{2}\alpha^2 \|\mathbf{p}_k\|^2 \\ &\leq f(\mathbf{x}_k) + \alpha\nabla f(\mathbf{x}_k)\mathbf{p}_k + \alpha(\beta - 1)\nabla f(\mathbf{x}_k)\mathbf{p}_k \\ &\leq f(\mathbf{x}_k) + \alpha\beta\nabla f(\mathbf{x}_k)\mathbf{p}_k \end{aligned}$$



Finite termination of Armijo line-search

Corollary (Finite termination of Armijo linesearch)

Suppose that $f(x)$ satisfy the standard assumptions and $\beta \in (0, 1)$ and that \mathbf{p}_k is a descent direction at \mathbf{x}_k . Then the step-size generated by then backtracking-Armijo line-search terminates with

$$\alpha_k \geq \min \{ \alpha_{\text{init}}, \tau \omega_k \}, \quad \omega_k = 2(\beta - 1) \nabla f(\mathbf{x}_k) \mathbf{p}_k / (\gamma \|\mathbf{p}_k\|^2)$$

Proof.

Line-search will terminate as soon as $\alpha^{(\ell)} \leq \omega_k$:

- 1 May be that α_{init} satisfies the Armijo condition $\Rightarrow \alpha_k = \alpha_{\text{init}}$.
- 2 Otherwise in the last line-search iteration we have

$$\alpha^{(\ell-1)} > \omega_k, \quad \alpha_k = \alpha^{(\ell)} = \tau \alpha^{(\ell-1)} > \tau \omega_k.$$

Combining these 2 cases gives the required result. □



Backtracking-Armijo line-search

- 1 The previous analysis permit to say that Backtracking-Armijo line-search ends in a finite number of steps.
- 2 The line-search produce a step length **not too long** due to the condition

$$f(\mathbf{x}_k + \alpha_k \mathbf{p}_k) \leq f(\mathbf{x}_k) + \alpha_k \beta \nabla f(\mathbf{x}_k) \mathbf{p}_k$$

- 3 The line-search produce a step length **not too short** due to the **finite termination** theorem.
- 4 Armijo line-search can be improved by adding some further requirements on the step length acceptance criteria.



Global convergence

Theorem (Global convergence)

Suppose that $f(\mathbf{x})$ satisfy the standard assumptions, then, for the iterates generated by the **Generic minimization algorithm** with **backtracking Armijo line-search** either:

- 1 $\nabla f(\mathbf{x}_k)^T = \mathbf{0}$ for some $k \geq 0$;
- 2 *or* $\lim_{k \rightarrow \infty} f(\mathbf{x}_k) = -\infty$;
- 3 *or* $\lim_{k \rightarrow \infty} |\nabla f(\mathbf{x}_k) \mathbf{p}_k| \min \left\{ 1, \|\mathbf{p}_k\|^{-1} \right\} = 0$.

Remark

If the theorem, point 1 means that we found a stationary point in a finite number of steps. Point 2 means that function $f(\mathbf{x})$ is unbounded below, so that a minimum does not exist. Point 3 alone do not imply convergence, but if $\nabla f(\mathbf{x}_k)$ and \mathbf{p}_k do not become orthogonal and $\|\mathbf{p}_k\| \not\rightarrow 0$ then $\|\nabla f(\mathbf{x}_k)\| \rightarrow 0$.



Proof.

(1/3).

Assume points 1 and 2 are not satisfied, then we prove point 3.

Consider

$$f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) + \alpha_k \beta \nabla f(\mathbf{x}_k) \mathbf{p}_k \leq f(\mathbf{x}_0) + \sum_{j=0}^k \alpha_j \beta \nabla f(\mathbf{x}_j) \mathbf{p}_j$$

by the fact that \mathbf{p}_k is a descent direction we have that the series:

$$\sum_{j=0}^{\infty} \alpha_j |\nabla f(\mathbf{x}_j) \mathbf{p}_j| \leq \beta^{-1} \lim_{k \rightarrow \infty} [f(\mathbf{x}_0) - f(\mathbf{x}_{k+1})] < \infty$$

and then

$$\lim_{j \rightarrow \infty} \alpha_j |\nabla f(\mathbf{x}_j) \mathbf{p}_j| = 0$$



Proof.

(2/3).

Recall that from finite termination Armijo theorem (slide n.28)

$$\alpha_k \geq \min \{ \alpha_{\text{init}}, \tau \omega_k \}, \quad \omega_k = 2(\beta - 1) \nabla f(\mathbf{x}_k) \mathbf{p}_k / (\gamma \|\mathbf{p}_k\|^2)$$

and consider the two index set:

$$\mathcal{K}_1 = \{ k \mid \alpha_k \geq \alpha_{\text{init}} \}, \quad \mathcal{K}_2 = \{ k \mid \alpha_k < \alpha_{\text{init}} \},$$

Obviously $\mathbb{N} = \mathcal{K}_1 \cup \mathcal{K}_2$ and from $\lim_{k \rightarrow \infty} \alpha_k |\nabla f(\mathbf{x}_k) \mathbf{p}_k| = 0$ we have

$$\lim_{k \in \mathcal{K}_1 \rightarrow \infty} \alpha_k |\nabla f(\mathbf{x}_k) \mathbf{p}_k| = 0, \quad (\text{A})$$

$$\lim_{k \in \mathcal{K}_2 \rightarrow \infty} \alpha_k |\nabla f(\mathbf{x}_k) \mathbf{p}_k| = 0, \quad (\text{B})$$

Proof.

(3/3).

For $k \in \mathcal{K}_1$ we have $\alpha_k = \alpha_{\text{init}}$ and

$\alpha_k |\nabla f(\mathbf{x}_k) \mathbf{p}_k| = \alpha_{\text{init}} |\nabla f(\mathbf{x}_k) \mathbf{p}_k|$ and from (A) we have

$$\lim_{k \in \mathcal{K}_1 \rightarrow \infty} |\nabla f(\mathbf{x}_k) \mathbf{p}_k| = 0 \quad (\star)$$

For $k \in \mathcal{K}_2$ we have $\tau \omega_k \leq \alpha_k \leq \omega_k$ so

$$\alpha_k |\nabla f(\mathbf{x}_k) \mathbf{p}_k| \geq \tau \omega_k |\nabla f(\mathbf{x}_k) \mathbf{p}_k| \geq 2\tau(1 - \beta) \frac{|\nabla f(\mathbf{x}_k) \mathbf{p}_k|^2}{\gamma \|\mathbf{p}_k\|^2}$$

and from (B) we have

$$\lim_{k \in \mathcal{K}_1 \rightarrow \infty} \frac{|\nabla f(\mathbf{x}_k) \mathbf{p}_k|}{\|\mathbf{p}_k\|} = 0 \quad (\star\star)$$

Combining (\star) and $(\star\star)$ gives the required result. □

Steepest descent algorithm

Steepest descent algorithm

Given an initial guess \mathbf{x}_0 , let $k = 0$;

while not converged **do**

 Compute a step-size α_k using a line-search along $-\nabla f(\mathbf{x}_k)^T$.

 Set $\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \nabla f(\mathbf{x}_k)^T$ and increase k by 1.

end while

- The steepest descent algorithm is simply the **generic minimization algorithm** with search direction the opposite of the gradient in \mathbf{x}_k .
- The search direction $-\nabla f(\mathbf{x}_k)^T$ is always a **descent direction** unless the point \mathbf{x}_k is a stationary point.



Global convergence of steepest descent

Corollary (Global convergence of steepest descent)

Suppose that $f(x)$ satisfy the standard assumptions, then, for the iterates generated by the *steepest descent algorithm* with *backtracking Armijo line-search* either:

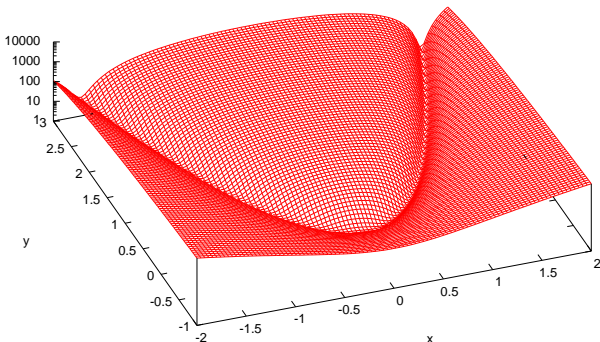
- 1 $\nabla f(\mathbf{x}_k)^T = \mathbf{0}$ for some $k \geq 0$;
- 2 *or* $\lim_{k \rightarrow \infty} f(\mathbf{x}_k) = -\infty$;
- 3 *or* $\lim_{k \rightarrow \infty} \nabla f(\mathbf{x}_k)^T = \mathbf{0}$.

The Rosenbrock example

(1/3)

- Although the **steepest descent** scheme is globally convergent it can be very slow!
- A classical example is the Rosenbrock function:

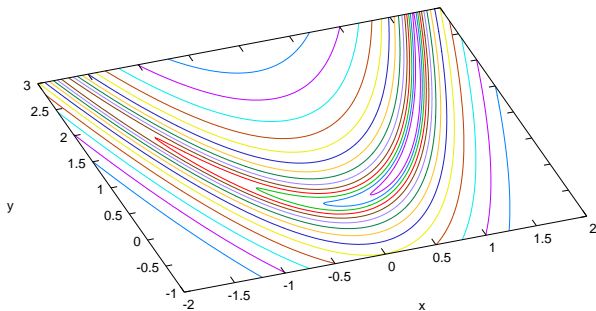
$$f(x, y) = 100(y - x^2)^2 + (x - 1)^2$$



The Rosenbrock example

(2/3)

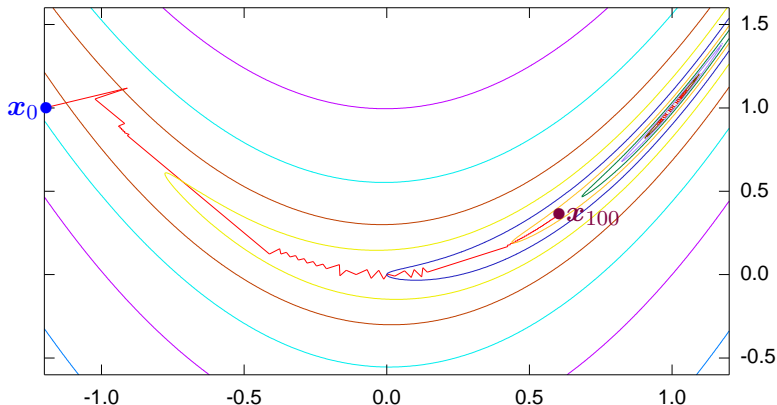
- This function has a unique minimum at $(1, 1)^T$ inside a **banana shaped** valley.



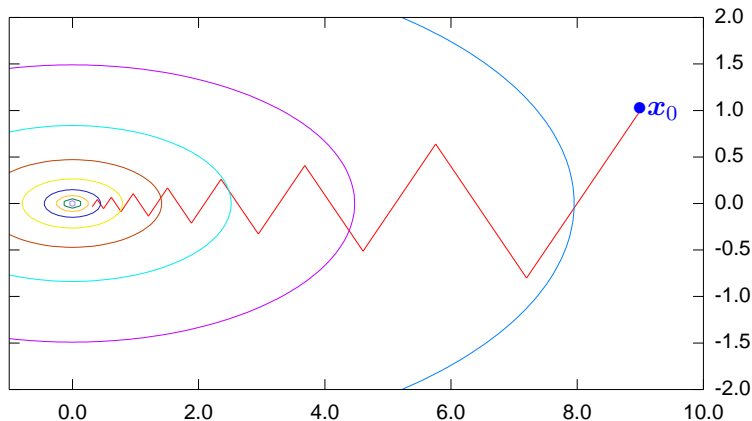
The Rosenbrock example

(3/3)

- After 100 iteration starting from $(-1.2, 1)^T$ the approximate minimum is **far** from the solution.



- The steepest descent is a slow method, not only on a difficult test case like the Rosenbrock example.
- Given the function $f(x, y) = \frac{1}{2}x^2 + \frac{9}{2}y^2$ starting from $\mathbf{x}_0 = (9, 1)^T$ we have the zig-zag pattern toward $(0, 0)^T$.



Outline

- 1 General iterative scheme
 - Descent direction failure
- 2 Backtracking Armijo line-search
 - Global convergence of backtracking Armijo line-search
 - Global convergence of steepest descent
- 3 Wolfe–Zoutendijk global convergence
 - The Wolfe conditions
 - The Armijo-Goldstein conditions
- 4 Algorithms for line-search
 - Armijo Parabolic-Cubic search
 - Wolfe linesearch

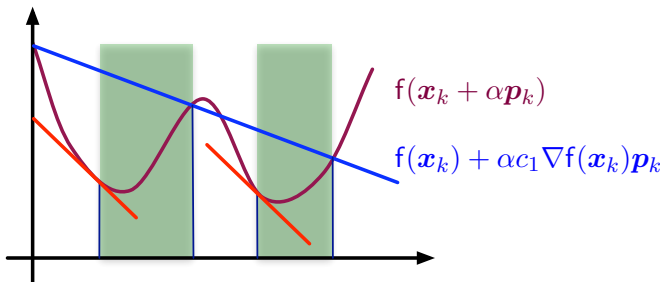
The Wolfe and Armijo Goldstein conditions

- 1 The simple condition of **descent step** is in general not enough for the convergence of a iterative minimization scheme.
- 2 The condition of **sufficient decrease** of backtracking Armijo line-search may be insufficient on general inexact line-search algorithm.
- 3 Adding another condition to the **sufficient decrease** condition such that we avoid **too short** step length we obtain **globally convergent** numerical procedure.
- 4 Depending on which additional condition is added we obtain the:
 - 1 Wolfe conditions;
 - 2 Armijo Goldstein conditions.

The Wolfe conditions

Let c_1 and c_2 two constant such that $0 < c_1 < c_2 < 1$. We say that the step length α_k satisfy the Wolfe conditions if α_k satisfy:

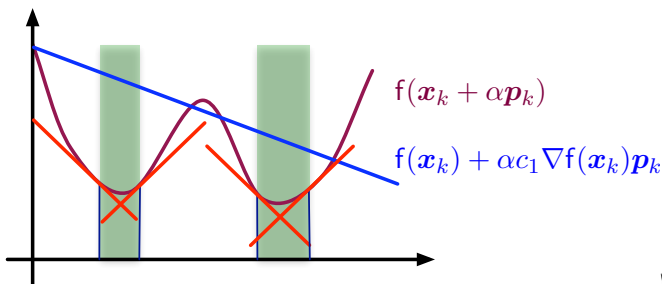
- ① **sufficient decrease:** $f(\mathbf{x}_k + \alpha_k \mathbf{p}_k) \leq f(\mathbf{x}_k) + c_1 \alpha_k \nabla f(\mathbf{x}_k) \mathbf{p}_k$;
- ② **curvature condition:** $\nabla f(\mathbf{x}_k + \alpha_k \mathbf{p}_k) \mathbf{p}_k \geq c_2 \nabla f(\mathbf{x}_k) \mathbf{p}_k$.



The strong Wolfe conditions

Let c_1 and c_2 two constant such that $0 < c_1 < c_2 < 1$. We say that the step length α_k satisfy the strong Wolfe conditions if α_k satisfy:

- 1 **sufficient decrease:** $f(\mathbf{x}_k + \alpha_k \mathbf{p}_k) \leq f(\mathbf{x}_k) + c_1 \alpha_k \nabla f(\mathbf{x}_k) \mathbf{p}_k$;
- 2 **curvature condition:** $|\nabla f(\mathbf{x}_k + \alpha_k \mathbf{p}_k) \mathbf{p}_k| \leq c_2 |\nabla f(\mathbf{x}_k) \mathbf{p}_k|$.



Existence of "Wolfe" step length

- The Wolfe condition seems quite restrictive.
- The next lemma answer to the question if a step length satisfying Wolfe conditions does exists.

Lemma (strong Wolfe step length)

Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ satisfying the regularity assumption. If the following condition are satisfied:

- 1 \mathbf{p}_k is a descent direction for the point \mathbf{x}_k , i.e. $\nabla f(\mathbf{x}_k)\mathbf{p}_k < 0$;
- 2 $f(\mathbf{x}_k + \alpha\mathbf{p}_k)$ is bounded from below, i.e.
$$\lim_{\alpha \rightarrow \infty} f(\mathbf{x}_k + \alpha\mathbf{p}_k) > -\infty.$$

then for any $0 < c_1 < c_2 < 1$ there exists an interval $[a, b]$ such that all $\alpha_k \in [a, b]$ satisfy the strong Wolfe conditions.

Proof.

Define $\ell(\alpha) = f(\mathbf{x}_k) + \alpha c_1 \nabla f(\mathbf{x}_k) \mathbf{p}_k$ and $g(\alpha) = f(\mathbf{x}_k + \alpha \mathbf{p}_k)$. From $\lim_{\alpha \rightarrow \infty} \ell(\alpha) = -\infty$ and from condition 1 it follows that there exists $\alpha_\star > 0$ such that

$$\ell(\alpha_\star) = g(\alpha_\star) \quad \text{and} \quad \ell(\alpha) > g(\alpha), \quad \forall \alpha \in (0, \alpha_\star)$$

so that all step length $\alpha \in (0, \alpha_\star)$ satisfy strong Wolfe condition 1. Because $\ell(0) = g(0)$ from Cauchy-Rolle theorem there exists $\alpha_{\star\star} \in (0, \alpha_\star)$ such that

$$g'(\alpha_{\star\star}) = \ell'(\alpha_{\star\star}) \quad \Rightarrow$$

$$0 > \nabla f(\mathbf{x}_k + \alpha_{\star\star} \mathbf{p}_k) \mathbf{p}_k = c_1 \nabla f(\mathbf{x}_k) \mathbf{p}_k > c_2 \nabla f(\mathbf{x}_k) \mathbf{p}_k$$

by continuity we find an interval around $\alpha_{\star\star}$ with step lengths satisfying strong Wolfe conditions. □

The Zoutendijk condition

Theorem (Zoutendijk)

Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ satisfying the regularity assumption and bounded from below, i.e.

$$\inf_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) > -\infty$$

Let $\{\mathbf{x}_k\}$, $k = 0, 1, \dots, \infty$ generated by a **generic minimization algorithm** where line-search satisfy **Wolfe conditions**, then

$$\sum_{k=1}^{\infty} (\cos \theta_k)^2 \|\nabla f(\mathbf{x}_k)^T\|^2 < +\infty$$

where

$$\cos \theta_k = \frac{-\nabla f(\mathbf{x}_k) \mathbf{p}_k}{\|\nabla f(\mathbf{x}_k)^T\| \|\mathbf{p}_k\|}$$



Proof.

(1/3).

Using the second condition of Wolfe (with $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$)

$$\nabla f(\mathbf{x}_{k+1}) \mathbf{p}_k \geq c_2 \nabla f(\mathbf{x}_k) \mathbf{p}_k$$

$$(\nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k)) \mathbf{p}_k \geq (c_2 - 1) \nabla f(\mathbf{x}_k) \mathbf{p}_k$$

by using Lipschitz regularity

$$\begin{aligned} \|\nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k)\| \|\mathbf{p}_k\| &\leq \gamma \|\mathbf{x}_{k+1} - \mathbf{x}_k\| \|\mathbf{p}_k\| \\ &= \alpha_k \gamma \|\mathbf{p}_k\|^2 \end{aligned}$$

and using both inequality we obtain the estimate for α_k :

$$\alpha_k \geq \frac{c_2 - 1}{\gamma \|\mathbf{p}_k\|^2} \nabla f(\mathbf{x}_k) \mathbf{p}_k$$



Proof.

(2/3).

Using the first condition of Wolfe and lower bound estimate of α_k

$$\begin{aligned} f(\mathbf{x}_{k+1}) &\leq f(\mathbf{x}_k) + \alpha_k c_1 \nabla f(\mathbf{x}_k) \mathbf{p}_k \\ &\leq f(\mathbf{x}_k) - \frac{c_1(1-c_2)}{\gamma \|\mathbf{p}_k\|^2} (\nabla f(\mathbf{x}_k) \mathbf{p}_k)^2 \end{aligned}$$

setting $A = c_1(1-c_2)/\gamma$ and using the definition of $\cos \theta_k$

$$f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) - A(\cos \theta_k)^2 \|\nabla f(\mathbf{x}_k)^T\|^2$$

and by induction

$$f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_1) - A \sum_{j=1}^k (\cos \theta_j)^2 \|\nabla f(\mathbf{x}_j)^T\|^2$$

Proof.

(3/3).

The function $f(\mathbf{x})$ is bounded from below, i.e.

$$\inf_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) > -\infty$$

so that

$$A \sum_{j=1}^k (\cos \theta_j)^2 \|\nabla f(\mathbf{x}_j)^T\|^2 \leq f(\mathbf{x}_1) - f(\mathbf{x}_{k+1})$$

and

$$A \sum_{j=1}^{\infty} (\cos \theta_j)^2 \|\nabla f(\mathbf{x}_j)^T\|^2 \leq f(\mathbf{x}_1) - \lim_{k \rightarrow \infty} f(\mathbf{x}_{k+1}) < +\infty$$

□



Corollary (Zoutendijk condition)

Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ satisfying the regularity assumption and bounded from below. Let $\{\mathbf{x}_k\}$, $k = 0, 1, \dots, \infty$ generated by a generic minimization algorithm where line-search satisfy **Wolfe conditions**, then

$$\cos \theta_k \|\nabla f(\mathbf{x}_k)^T\| \rightarrow 0 \quad \text{where} \quad \cos \theta_k = \frac{-\nabla f(\mathbf{x}_k) \mathbf{p}_k}{\|\nabla f(\mathbf{x}_k)^T\| \|\mathbf{p}_k\|}$$

Remark

If $\cos \theta_k \geq \delta > 0$ for all k from the Zoutendijk condition we have:

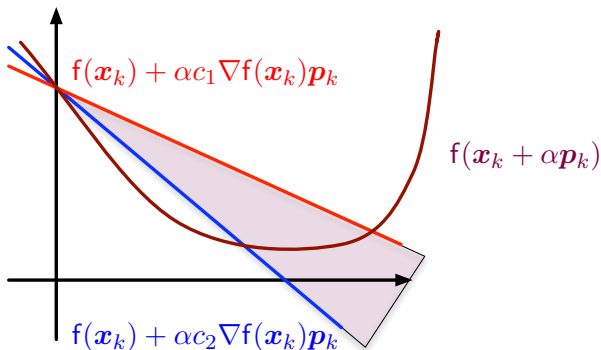
$$\|\nabla f(\mathbf{x}_k)^T\| \rightarrow 0$$

i.e. the **generic minimization algorithm** where line-search satisfy **Wolfe conditions** converge to a stationary point.

The Armijo-Goldstein conditions

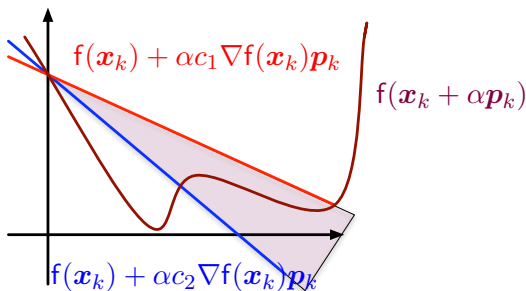
Let c_1 and c_2 two constant such that $0 < c_1 < c_2 < 1$. We say that the step length α_k satisfy the Wolfe conditions if α_k satisfy:

- 1 $f(\mathbf{x}_k + \alpha_k \mathbf{p}_k) \leq f(\mathbf{x}_k) + c_1 \alpha_k \nabla f(\mathbf{x}_k) \mathbf{p}_k$;
- 2 $f(\mathbf{x}_k + \alpha_k \mathbf{p}_k) \geq f(\mathbf{x}_k) + c_2 \alpha_k \nabla f(\mathbf{x}_k) \mathbf{p}_k$;



The Armijo-Goldstein conditions

- ① Armijo-Goldstein conditions has very similar theoretical properties like the Wolfe conditions.
- ② Global convergence theorems can be established.
- ③ The weakness of Armijo-Goldstein conditions respect to Wolfe conditions is that the former can exclude **local minima's** from the step length as you can see in the figure below.



Outline

- 1 General iterative scheme
 - Descent direction failure

- 2 Backtracking Armijo line-search
 - Global convergence of backtracking Armijo line-search
 - Global convergence of steepest descent

- 3 Wolfe–Zoutendijk global convergence
 - The Wolfe conditions
 - The Armijo-Goldstein conditions

- 4 Algorithms for line-search
 - Armijo Parabolic-Cubic search
 - Wolfe linesearch

Armijo Parabolic-Cubic search

- 1 Backtracking-Armijo line-search can be slow if a large number of reduction must be performed to satisfy Armijo condition.
- 2 A better performance is obtained if instead of reducing by a fixed factor we use polynomial interpolation to estimate the location of the minimum.
- 3 Assuming that that $f(\mathbf{x}_k)$ and $\nabla f(\mathbf{x}_k)\mathbf{p}_k$ are known at the first step we know also $f(\mathbf{x}_k + \lambda\mathbf{p}_k)$ if λ is the first trial step.
- 4 In this case a parabolic interpolation can be used to estimate the minimum.
- 5 If we store the last trial step length, in the successive iteration we can use cubic interpolation to estimate the minima's.
- 6 The resulting algorithm is in the following slides.



Algorithm (Armijo Parabolic-Cubic search)

```

1: armijo_linesearch( $f, \mathbf{x}, \mathbf{p}, \tau$ )
2:  $f_0 \leftarrow f(\mathbf{x}); \nabla f_0 \leftarrow \nabla f(\mathbf{x})\mathbf{p}; \lambda \leftarrow 1;$ 
3: while  $\lambda \geq \lambda_{\min}$  do
4:    $f_\lambda \leftarrow f(\mathbf{x} + \lambda\mathbf{p});$ 
5:   if  $f_\lambda \leq f_0 + \lambda\tau\nabla f_0$  then
6:     return  $\lambda$  ; successful search
7:   else
8:     if  $\lambda = 1$  then
9:        $\lambda_{tmp} \leftarrow \nabla f_0 / [2(f_0 + \nabla f_0 - f_\lambda)];$ 
10:    else
11:       $\lambda_{tmp} \leftarrow \text{cubic}(f_0, \nabla f_0, f_\lambda, \lambda, f_p, \lambda_p);$ 
12:    end if
13:     $\lambda_p \leftarrow \lambda; f_p \leftarrow f_\lambda; \lambda \leftarrow \text{range}(\lambda_{tmp}, \lambda/10, \lambda/2);$ 
14:  end if
15: end while
16: return  $\lambda_{\min}$  ; failed search

```



Algorithm (Armijo Parabolic-Cubic search)

```
17: range( $\lambda, a, b$ )
18: if  $\lambda < a$  then
19:   return  $a$ ;
20: else if  $\lambda > b$  then
21:   return  $b$ ;
22: else
23:   return  $\lambda$  ;
24: end if
```

Algorithm (Armijo Parabolic-Cubic search)

25: *cubic*($f_0, \nabla f_0, f_\lambda, \lambda, f_p, \lambda_p$)

26: *Evaluate*:

$$\begin{pmatrix} a \\ b \end{pmatrix} = \frac{1}{\lambda^2 \lambda_p^2 (\lambda - \lambda_p)} \begin{pmatrix} \lambda_p^2 & -\lambda^2 \\ -\lambda_p^3 & \lambda^3 \end{pmatrix} \begin{pmatrix} f_\lambda - f_0 - \lambda \nabla f_0 \\ f_p - f_0 - \lambda_p \nabla f_0 \end{pmatrix}$$

27: **if** $a = 0$ **then**

28: **return** $-\nabla f_0 / (2b)$;

cubic is a quadratic

29: **else**

30: $d \leftarrow b^2 - 3a \nabla f_0$;

discriminant

31: **return** $(-b + \sqrt{d}) / (3a)$;

legitimate cubic

32: **end if**

Wolfe linesearch

- 1 Wolfe linesearch is identical to the Armijo Parabolic-Cubic search, until a point satisfying the first condition is found.
- 2 At this point the Armijo algorithm stop while Wolfe search try to refine the search until the second condition is satisfied.
- 3 If the step estimated is too short then is is enlarged until it contains a minimum.
- 4 If the step estimated is too long it is reduced until the second condition is satisfied.



Algorithm (Wolfe linesearch)

```

1: wolfe_linesearch( $f, \mathbf{x}, \mathbf{p}, c_1, c_2$ )
2:  $f_0 \leftarrow f(\mathbf{x}); \nabla f_0 \leftarrow \nabla f(\mathbf{x})\mathbf{p}; \lambda \leftarrow 1;$ 
3: while  $\lambda \geq \lambda_{\min}$  do
4:    $f_\lambda \leftarrow f(\mathbf{x} + \lambda\mathbf{p});$ 
5:   if  $f_\lambda \leq f_0 + \lambda c_1 \nabla f_0$  then
6:     go to ZOOM; found a  $\lambda$  satisfying condition 1
7:   else
8:     if  $\lambda = 1$  then
9:        $\lambda_{tmp} \leftarrow \nabla f_0 / [2(f_0 + \nabla f_0 - f_\lambda)];$ 
10:    else
11:       $\lambda_{tmp} \leftarrow \text{cubic}(f_0, \nabla f_0, f_\lambda, \lambda, f_p, \lambda_p);$ 
12:    end if
13:     $\lambda_p \leftarrow \lambda; f_p \leftarrow f_\lambda; \lambda \leftarrow \text{range}(\lambda_{tmp}, \lambda/10, \lambda/2);$ 
14:  end if
15: end while
16: return  $\lambda_{\min}$  ; failed search

```



Algorithm (Wolfe linesearch)

```

17: ZOOM:
18:  $\nabla f_\lambda \leftarrow \nabla f(\mathbf{x} + \lambda \mathbf{p})\mathbf{p}$ ;
19: if  $\nabla f_\lambda \geq c_2 \nabla f_0$  then return  $\lambda$ ;           found Wolfe point!
20: if  $\lambda = 1$  then
21:   forward search of an interval bracketing a minimum
22:   while  $\lambda \leq \lambda_{\max}$  do
23:      $\{\lambda_p, f_p\} \leftarrow \{\lambda, f_\lambda\}$ ;           save values
24:      $\lambda \leftarrow 2\lambda$ ;  $f_\lambda \leftarrow f(\mathbf{x} + \lambda \mathbf{p})$ ;
25:     if not  $f_\lambda \leq f_0 + \lambda c_1 \nabla f_0$  then
26:        $\{\lambda_p, f_p\} \rightleftharpoons \{\lambda, f_\lambda\}$ ; go to REFINE;   swap values
27:     end if
28:      $\nabla f_\lambda \leftarrow \nabla f(\mathbf{x} + \lambda \mathbf{p})\mathbf{p}$ ;
29:     if  $\nabla f_\lambda \geq c_2 \nabla f_0$  then return  $\lambda$ ;   found Wolfe point!
30:   end while
31:   return  $\lambda_{\max}$  ; failed search
32: end if

```



Algorithm (Wolfe linesearch)



```

33: REFINE:
34:  $\{\lambda_{lo}, f_{lo}, \nabla f_{lo}\} \leftarrow \{\lambda, f_\lambda, \nabla f_\lambda\}; \Delta \leftarrow \lambda_p - \lambda_{lo};$ 
35: while  $\Delta > \epsilon$  do
36:    $\delta\lambda \leftarrow \Delta^2 \nabla f_{lo} / [2(f_{lo} + \nabla f_{lo} \Delta - f_p)];$ 
37:    $\delta\lambda \leftarrow \text{range}(\delta\lambda, 0.2\Delta, 0.8\Delta);$ 
38:    $\lambda \leftarrow \lambda_{lo} + \delta\lambda; f_\lambda \leftarrow f(\mathbf{x} + \lambda\mathbf{p});$ 
39:   if  $f_\lambda \leq f_0 + \lambda c_1 \nabla f_0$  then
40:      $\nabla f_\lambda \leftarrow \nabla f(\mathbf{x} + \lambda\mathbf{p})\mathbf{p};$ 
41:     if  $\nabla f_\lambda \geq c_2 \nabla f_0$  then return  $\lambda;$       found Wolfe point!
42:      $\{\lambda_{lo}, f_{lo}, \nabla f_{lo}\} \leftarrow \{\lambda, f_\lambda, \nabla f_\lambda\}; \Delta \leftarrow \Delta - \delta\lambda;$ 
43:   else
44:      $\{\lambda_p, f_p\} \leftarrow \{\lambda, f_\lambda\}; \Delta \leftarrow \delta\lambda;$ 
45:   end if
46: end while
47: return  $\lambda;$  failed search

```



References

-  J. Stoer and R. Bulirsch
Introduction to numerical analysis
Springer-Verlag, Texts in Applied Mathematics, **12**, 2002.
-  J. E. Dennis, Jr. and Robert B. Schnabel
Numerical Methods for Unconstrained Optimization and
Nonlinear Equations
SIAM, Classics in Applied Mathematics, **16**, 1996.