Algorithms for Wolfe Conditions

Qi Gong
Associate Professor
Dept. of Applied Math & Stats.
Baskin School of Engineering
University of California
Santa Cruz, CA

Wolfe Conditions

Condition 1

$$f(x_k + \alpha p_k) \le f(x_k) + c_1 \alpha \nabla f_k^T p_k,$$

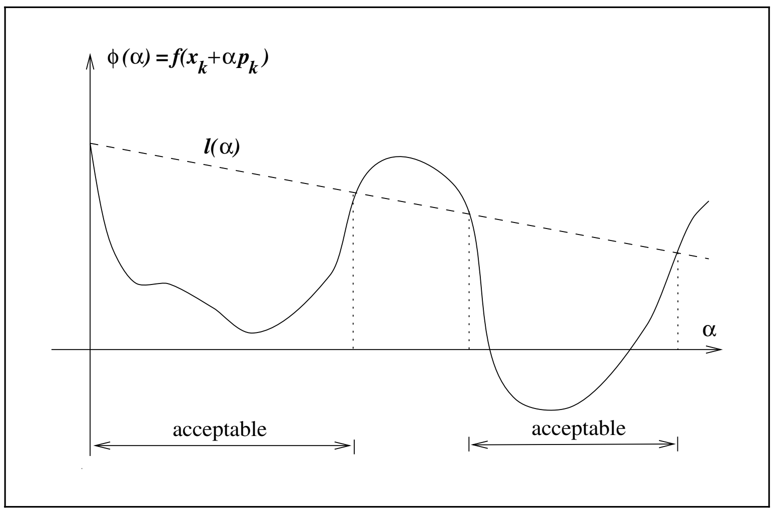


Figure 3.3 Sufficient decrease condition.

Condition 2

$$\nabla f(x_k + \alpha_k p_k)^T p_k \ge c_2 \nabla f_k^T p_k,$$

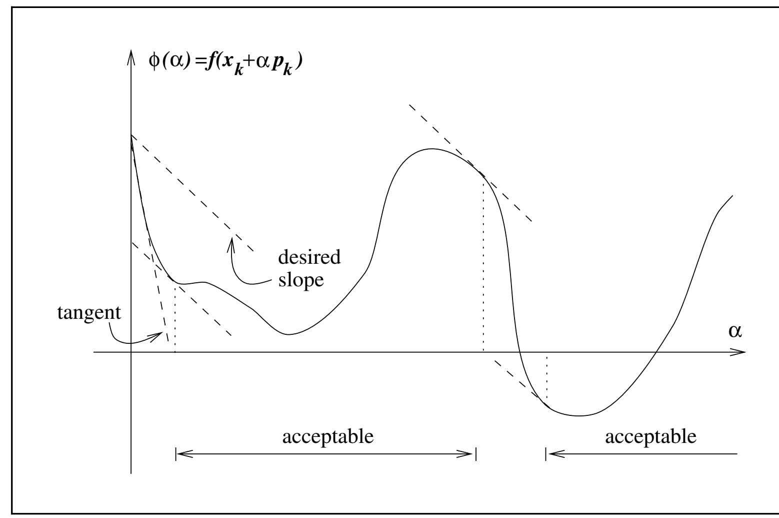


Figure 3.4 The curvature condition.

Wolfe Conditions

The Wolfe conditions:

$$f(x_k + \alpha_k p_k) \leq f(x_k) + c_1 \alpha_k \nabla f_k^T p_k,$$

$$\nabla f(x_k + \alpha_k p_k)^T p_k \geq c_2 \nabla f_k^T p_k,$$

The Strong Wolfe Conditions:

$$f(x_k + \alpha_k p_k) \leq f(x_k) + c_1 \alpha_k \nabla f_k^T p_k,$$

$$|\nabla f(x_k + \alpha_k p_k)^T p_k| \leq c_2 |\nabla f_k^T p_k|,$$

where $0 < c_1 < c_2 < 1$

Convergence – Zoutendijk Theorem

Theorem 3.2. (Zoutendijk Theorem)

Consider any iteration of the form (3.1), where p_k is a descent direction and α_k satisfies the Wolfe conditions (3.6). Suppose that f is bounded below in $\mathbb{R}^n$ and that f is continuously differentiable in an open set $\mathcal{N}$ containing the level set $\mathcal{L} \stackrel{\text{def}}{=} \{x : f(x) \le f(x_0)\}$, where x_0 is the starting point of the iteration. Assume also that the gradient ∇f is Lipschitz continuous on $\mathcal{N}$, that is, there exists a constant $L > 0$ such that

$$\|\nabla f(x) - \nabla f(\tilde{x})\| \le L\|x - \tilde{x}\|, \qquad \text{for all } x,\ \tilde{x} \in \mathcal{N}. \tag{3.13}$$

Then

$$\sum_{k \ge 0} \cos^2 \theta_k \, \|\nabla f_k\|^2 < \infty. \tag{3.14}$$

Wolfe Conditions – Existence Result

The Wolfe Conditions:

$$f(x_k + \alpha_k p_k) \le f(x_k) + c_1 \alpha_k \nabla f_k^T p_k,$$
$$\nabla f(x_k + \alpha_k p_k)^T p_k \ge c_2 \nabla f_k^T p_k,$$

The Strong Wolfe Conditions:

$$f(x_k + \alpha_k p_k) \le f(x_k) + c_1 \alpha_k \nabla f_k^T p_k,$$
$$|\nabla f(x_k + \alpha_k p_k)^T p_k| \le c_2 |\nabla f_k^T p_k|,$$

Lemma 3.1.

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable. Let p_k be a descent direction at x_k, and assume that f is bounded below along the ray $\{x_k + \alpha p_k | \alpha > 0\}$. Then if $0 < c_1 < c_2 < 1$, there exist intervals of step lengths satisfying the Wolfe conditions (3.6) and the strong Wolfe conditions (3.7).

An Algorithm for The Strong Wolfe Conditions

A two-phase algorithm based on Jorge More's 1994 paper

Algorithm 3.5 (Line Search Algorithm).

Set $\alpha_0 \leftarrow 0$, choose $\alpha_{\max} > 0$ and $\alpha_1 \in (0, \alpha_{\max})$;
$i \leftarrow 1$;
repeat
  Evaluate $\phi(\alpha_i)$;
  if $\phi(\alpha_i) > \phi(0) + c_1\alpha_i\phi'(0)$ or $[\phi(\alpha_i) \geq \phi(\alpha_{i-1})$ and $i > 1]$
    $\alpha_* \leftarrow \boxed{\mathbf{zoom}(\alpha_{i-1}, \alpha_i)}$ and **stop**;
  Evaluate $\phi'(\alpha_i)$;
  if $|\phi'(\alpha_i)| \leq -c_2\phi'(0)$
    set $\alpha_* \leftarrow \alpha_i$ and **stop**;
  if $\phi'(\alpha_i) \geq 0$
    set $\alpha_* \leftarrow \boxed{\mathbf{zoom}(\alpha_i, \alpha_{i-1})}$ and **stop**;
  Choose $\alpha_{i+1} \in (\alpha_i, \alpha_{\max})$;
  $i \leftarrow i + 1$;
end (repeat)

bracketing phase

Algorithm 3.6 (zoom).

repeat
  Interpolate (using quadratic, cubic, or bisection) to find
    a trial step length α_j between α_{lo} and α_{hi};
  Evaluate $\phi(\alpha_j)$;
  if $\phi(\alpha_j) > \phi(0) + c_1\alpha_j\phi'(0)$ or $\phi(\alpha_j) \geq \phi(\alpha_{lo})$
    $\alpha_{hi} \leftarrow \alpha_j$;
  else
    Evaluate $\phi'(\alpha_j)$;
    if $|\phi'(\alpha_j)| \leq -c_2\phi'(0)$
      Set $\alpha_* \leftarrow \alpha_j$ and **stop**;
    if $\phi'(\alpha_j)(\alpha_{hi} - \alpha_{lo}) \geq 0$
      $\alpha_{hi} \leftarrow \alpha_{lo}$;
    $\alpha_{lo} \leftarrow \alpha_j$;
end (repeat)

selection phase

Bracketing Phase (for Strong Wolfe Conditions)

Algorithm 3.5 (Line Search Algorithm).

Set $\alpha_0 \leftarrow 0$, choose $\alpha_{\max} > 0$ and $\alpha_1 \in (0, \alpha_{\max})$;
$i \leftarrow 1$;
repeat

  Evaluate $\phi(\alpha_i)$;

Cond #1 is NOT satisfied.

  if $\boxed{\phi(\alpha_i) > \phi(0) + c_1 \alpha_i \phi'(0)}$ or $[\phi(\alpha_i) \geq \phi(\alpha_{i-1})$ and $i > 1]$

    $\alpha_* \leftarrow$ **zoom**(α_{i-1}, α_i) and **stop**;

  Evaluate $\phi'(\alpha_i)$;

  if $|\phi'(\alpha_i)| \leq -c_2 \phi'(0)$

    set $\alpha_* \leftarrow \alpha_i$ and **stop**;

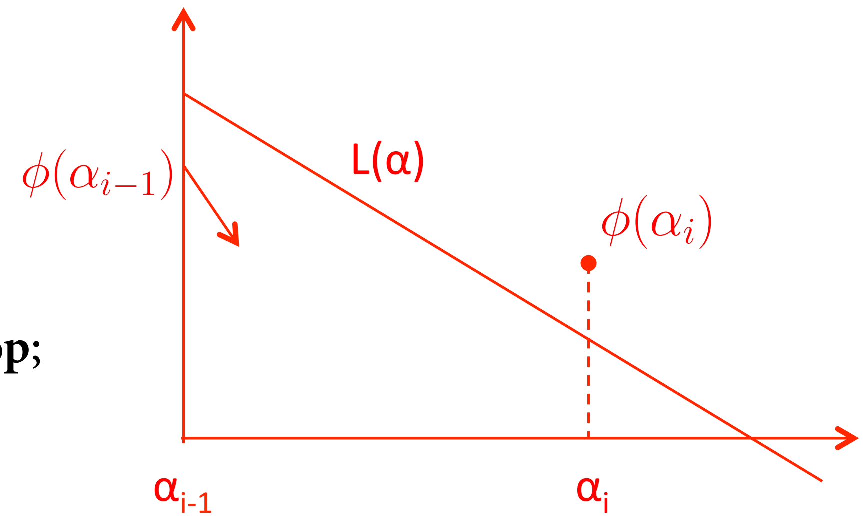
  if $\phi'(\alpha_i) \geq 0$

    set $\alpha_* \leftarrow$ **zoom**(α_i, α_{i-1}) and **stop**;

  Choose $\alpha_{i+1} \in (\alpha_i, \alpha_{\max})$;

  $i \leftarrow i + 1$;

end (repeat)



Note: the algorithm ensures that, at any iteration i, $$\begin{aligned} \phi(\alpha_{i-1}) &\leq L(\alpha) \\ \phi'(\alpha_{i-1}) &\leq c_1 \phi'(0) \end{aligned}$$

Bracketing Phase (for Strong Wolfe Conditions)

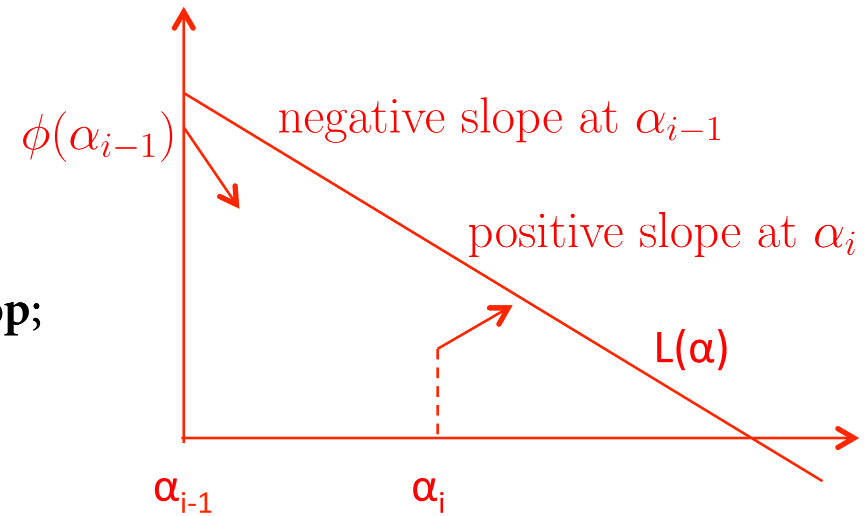
Algorithm 3.5 (Line Search Algorithm).

Set $\alpha_0 \leftarrow 0$, choose $\alpha_{\max} > 0$ and $\alpha_1 \in (0, \alpha_{\max})$;
$i \leftarrow 1$;
repeat
  Evaluate $\phi(\alpha_i)$;
  if $\phi(\alpha_i) > \phi(0) + c_1\alpha_i\phi'(0)$ or $[\phi(\alpha_i) \geq \phi(\alpha_{i-1})$ and $i > 1]$
    $\alpha_* \leftarrow$**zoom**(α_{i-1}, α_i) and **stop**;
  Evaluate $\phi'(\alpha_i)$;   Cond #1 is satisfied. We need to check Cond #2
  if $|\phi'(\alpha_i)| \leq -c_2\phi'(0)$
    set $\alpha_* \leftarrow \alpha_i$ and **stop**;
  if $\phi'(\alpha_i) \geq 0$
    set $\alpha_* \leftarrow$**zoom**(α_i, α_{i-1}) and **stop**;
  Choose $\alpha_{i+1} \in (\alpha_i, \alpha_{\max})$;
  $i \leftarrow i + 1$;
end (repeat)

Bracketing Phase (for Strong Wolfe Conditions)

Algorithm 3.5 (Line Search Algorithm).

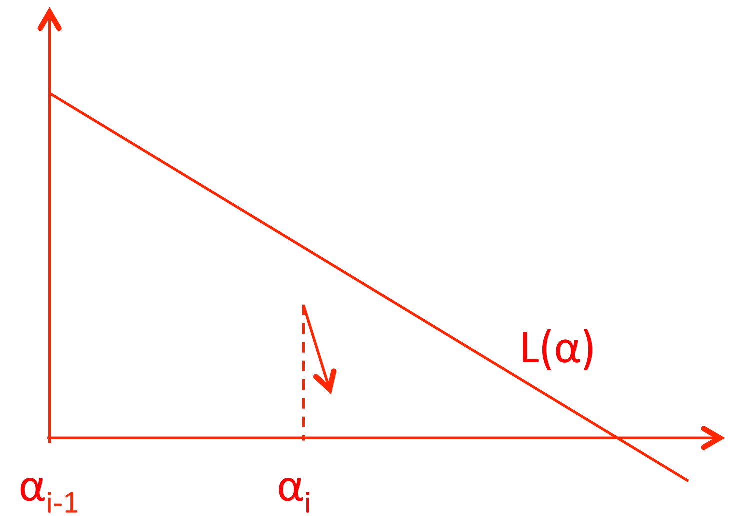
Set $\alpha_0 \leftarrow 0$, choose $\alpha_{\max} > 0$ and $\alpha_1 \in (0, \alpha_{\max})$;
$i \leftarrow 1$;
repeat
  Evaluate $\phi(\alpha_i)$;
  if $\phi(\alpha_i) > \phi(0) + c_1\alpha_i\phi'(0)$ or $[\phi(\alpha_i) \geq \phi(\alpha_{i-1})$ and $i > 1]$
    $\alpha_* \leftarrow$**zoom**(α_{i-1}, α_i) and **stop**;
  Evaluate $\phi'(\alpha_i)$;
  if $|\phi'(\alpha_i)| \leq -c_2\phi'(0)$
    set $\alpha_* \leftarrow \alpha_i$ and **stop**;
  if $\phi'(\alpha_i) \geq 0$
    set $\alpha_* \leftarrow$**zoom**(α_i, α_{i-1}) and **stop**;
  Choose $\alpha_{i+1} \in (\alpha_i, \alpha_{\max})$;
  $i \leftarrow i + 1$;
end (repeat)



Bracketing Phase (for Strong Wolfe Conditions)

Algorithm 3.5 (Line Search Algorithm).

Set $\alpha_0 \leftarrow 0$, choose $\alpha_{\max} > 0$ and $\alpha_1 \in (0, \alpha_{\max})$;
$i \leftarrow 1$;
repeat
  Evaluate $\phi(\alpha_i)$;
  if $\phi(\alpha_i) > \phi(0) + c_1\alpha_i\phi'(0)$ or $[\phi(\alpha_i) \geq \phi(\alpha_{i-1})$ and $i > 1]$
    $\alpha_* \leftarrow$**zoom**(α_{i-1}, α_i) and **stop**;
  Evaluate $\phi'(\alpha_i)$;
  if $|\phi'(\alpha_i)| \leq -c_2\phi'(0)$
    set $\alpha_* \leftarrow \alpha_i$ and **stop**;
  if $\phi'(\alpha_i) \geq 0$
    set $\alpha_* \leftarrow$**zoom**(α_i, α_{i-1}) and **stop**;
  Choose $\alpha_{i+1} \in (\alpha_i, \alpha_{\max})$;
  $i \leftarrow i + 1$;
end (repeat)

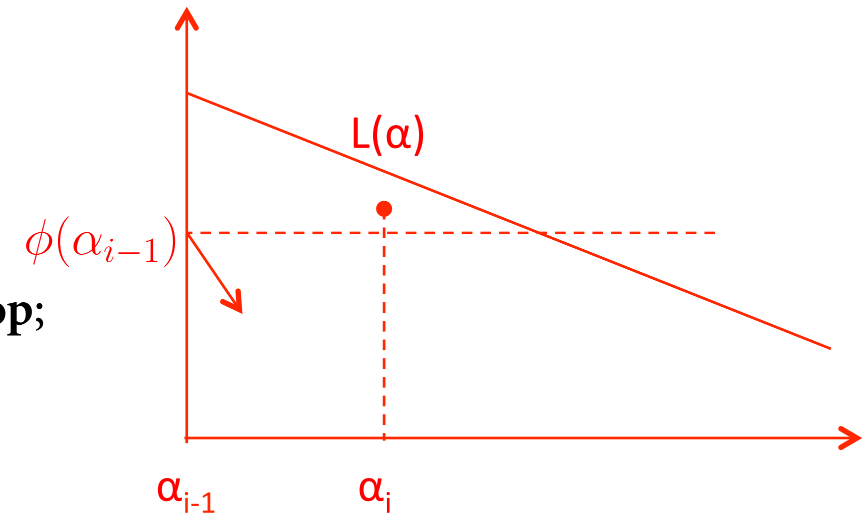


Base on the same argument used in the proof of Lemma 3.1, there must exist a step length in (α_i, α_{max}).

Bracketing Phase (for Strong Wolfe Conditions)

Algorithm 3.5 (Line Search Algorithm).

Set $\alpha_0 \leftarrow 0$, choose $\alpha_{\max} > 0$ and $\alpha_1 \in (0, \alpha_{\max})$;
$i \leftarrow 1$;
repeat
  Evaluate $\phi(\alpha_i)$;
  if $\phi(\alpha_i) > \phi(0) + c_1\alpha_i\phi'(0)$ or $\boxed{[\phi(\alpha_i) \geq \phi(\alpha_{i-1}) \text{ and } i > 1]}$
    $\alpha_* \leftarrow$**zoom**(α_{i-1}, α_i) and **stop**;
  Evaluate $\phi'(\alpha_i)$;
  if $|\phi'(\alpha_i)| \leq -c_2\phi'(0)$
    set $\alpha_* \leftarrow \alpha_i$ and **stop**;
  if $\phi'(\alpha_i) \geq 0$
    set $\alpha_* \leftarrow$**zoom**(α_i, α_{i-1}) and **stop**;
  Choose $\alpha_{i+1} \in (\alpha_i, \alpha_{\max})$;
  $i \leftarrow i + 1$;
end (repeat)



Bracketing Phase (for Strong Wolfe Conditions)

Algorithm 3.5 (Line Search Algorithm).

Set $\alpha_0 \leftarrow 0$, choose $\alpha_{\max} > 0$ and $\alpha_1 \in (0, \alpha_{\max})$;

$i \leftarrow 1$;

repeat

Evaluate $\phi(\alpha_i)$;

if $\phi(\alpha_i) > \phi(0) + c_1\alpha_i\phi'(0)$ or $[\phi(\alpha_i) \geq \phi(\alpha_{i-1})$ and $i > 1]$

$\alpha_* \leftarrow \boxed{\textbf{zoom}(\alpha_{i-1}, \alpha_i)}$ and **stop**;

Evaluate $\phi'(\alpha_i)$;

if $|\phi'(\alpha_i)| \leq -c_2\phi'(0)$

set $\alpha_* \leftarrow \alpha_i$ and **stop**;

if $\phi'(\alpha_i) \geq 0$

set $\alpha_* \leftarrow \boxed{\textbf{zoom}(\alpha_i, \alpha_{i-1})}$ and **stop**;

Choose $\alpha_{i+1} \in (\alpha_i, \alpha_{\max})$;

$i \leftarrow i + 1$;

end (repeat)

Selection Phase (for Strong Wolfe Conditions)

- zoom(a_{lo} , a_{hi}), a_{lo} may be larger than a_{hi}
- The interval between a_{lo} and a_{hi} contains an acceptable step length;
- At a_{lo} the first Wolfe condition is satisfied;
- At a_{hi}: $\phi'(\alpha_{\text{lo}})(\alpha_{\text{hi}} - \alpha_{\text{lo}}) < 0$.

Algorithm 3.6 (zoom).

repeat

Interpolate (using quadratic, cubic, or bisection) to find a trial step length α_j between α_{lo} and α_{hi};

Evaluate $\phi(\alpha_j)$;

if $\phi(\alpha_j) > \phi(0) + c_1\alpha_j\phi'(0)$ or $\phi(\alpha_j) \geq \phi(\alpha_{\text{lo}})$

$\alpha_{\text{hi}} \leftarrow \alpha_j$;

else

Evaluate $\phi'(\alpha_j)$;

if $|\phi'(\alpha_j)| \leq -c_2\phi'(0)$

Set $\alpha_* \leftarrow \alpha_j$ and **stop**;

if $\phi'(\alpha_j)(\alpha_{\text{hi}} - \alpha_{\text{lo}}) \geq 0$

$\alpha_{\text{hi}} \leftarrow \alpha_{\text{lo}}$;

$\alpha_{\text{lo}} \leftarrow \alpha_j$;

end (repeat)

Selection Phase (for Strong Wolfe Conditions)

- zoom(a_{lo} , a_{hi}), a_{lo} may be larger than a_{hi}
- The interval between a_{lo} and a_{hi} contains an acceptable step length;
- At a_{lo} the first Wolfe condition is satisfied;
- At a_{hi}: $\phi'(\alpha_{\text{lo}})(\alpha_{\text{hi}} - \alpha_{\text{lo}}) < 0.$

Algorithm 3.6 (zoom).

repeat

Interpolate (using quadratic, cubic, or bisection) to find a trial step length α_j between α_{lo} and α_{hi};

Evaluate $\phi(\alpha_j)$;

if $\phi(\alpha_j) > \phi(0) + c_1\alpha_j\phi'(0)$ or $\phi(\alpha_j) \geq \phi(\alpha_{\text{lo}})$

$\alpha_{\text{hi}} \leftarrow \alpha_j$;

else

Evaluate $\phi'(\alpha_j)$;

if $|\phi'(\alpha_j)| \leq -c_2\phi'(0)$

Set $\alpha_* \leftarrow \alpha_j$ and **stop**;

if $\phi'(\alpha_j)(\alpha_{\text{hi}} - \alpha_{\text{lo}}) \geq 0$

$\alpha_{\text{hi}} \leftarrow \alpha_{\text{lo}}$;

$\alpha_{\text{lo}} \leftarrow \alpha_j$;

end (repeat)