Math 408A
Bisection Method for the Weak Wolfe Conditions

Wolfe Conditions

$$f(x^k + t_k d^k) \approx \min_{t \in \mathbb{R}} f(x^k + td^k) \ .$$

Wolfe Conditions

$$f(x^k + t_k d^k) \approx \min_{t \in \mathbb{R}} f(x^k + td^k) \ .$$

$$\nabla f(x^k + t_k d^k)^T d^k \approx 0$$

Wolfe Conditions

$$f(x^k + t_k d^k) \approx \min_{t \in \mathbb{R}} f(x^k + td^k) \ .$$

$$\nabla f(x^k + t_k d^k)^T d^k \approx 0$$

Choose $0 < c_1 < c_2 < 1$.

Wolfe Conditions

$$f(x^k + t_k d^k) \approx \min_{t \in \mathbb{R}} f(x^k + t d^k) \ .$$

$$\nabla f(x^k + t_k d^k)^T d^k \approx 0$$

Choose $0 < c_1 < c_2 < 1$.

Strong Wolfe Conditions:

$$\begin{aligned} f(x^k + t_k d^k) &\leq f(x^k) + c_1 t_k f'(x^k; d^k) \\ |f'(x^k + t_k d^k; d^k)| &\leq c_2 |f'(x^k; d^k)| \ . \end{aligned}$$

Wolfe Conditions

Weak Wolfe Conditions:

$$\begin{aligned} f(x^k + t_k d^k) &\le f(x^k) + c_1 t_k f'(x^k; d^k) \\ c_2 f'(x^k; d^k) &\le f'(x^k + t_k d^k; d^k)\ . \end{aligned}$$

Wolfe Conditions

Lemma: Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable and suppose that $x, d \in \mathbb{R}^n$ are such that the set $\{f(x + td) \; : \; t \geq 0\}$ is bounded below and $f'(x; d) < 0$, then for each $0 < c_1 < c_2 < 1$ the set

$$\left\{ t \;\middle|\; \begin{array}{c} t > 0, f'(x + td; d) \geq c_2 f'(x; d), \text{ and} \\ f(x + td) \leq f(x) + c_1 t f'(x; d) \end{array} \right\}$$

has non–empty interior.

Wolfe Conditions

Set $\phi(t) = f(x + td) - (f(x) + c_1 t f'(x; d))$. Then

$$\phi(0) = 0 \quad \text{and} \quad \phi'(0) = (1 - c_1) f'(x; d) < 0.$$

Wolfe Conditions

Set $\phi(t) = f(x + td) - (f(x) + c_1 t f'(x; d))$. Then

$$\phi(0) = 0 \quad \text{and} \quad \phi'(0) = (1 - c_1) f'(x; d) < 0.$$

So $\exists\ \bar{t} > 0$ such that $\phi(t) < 0$ for $t \in (0, \bar{t})$.

Wolfe Conditions

Set $\phi(t) = f(x + td) - (f(x) + c_1 t f'(x; d))$. Then

$$\phi(0) = 0 \quad \text{and} \quad \phi'(0) = (1 - c_1) f'(x; d) < 0.$$

So $\exists\ \bar{t} > 0$ such that $\phi(t) < 0$ for $t \in (0, \bar{t})$.
Since $f'(x; d) < 0$ and $\{f(x + td) \ : \ t \geq 0\}$ is bounded below, we have $\phi(t) \to +\infty$ as $t \uparrow \infty$.

Wolfe Conditions

Set $\phi(t) = f(x + td) - (f(x) + c_1 t f'(x; d))$. Then

$$\phi(0) = 0 \quad \text{and} \quad \phi'(0) = (1 - c_1) f'(x; d) < 0.$$

So $\exists\ \bar{t} > 0$ such that $\phi(t) < 0$ for $t \in (0, \bar{t})$.
Since $f'(x; d) < 0$ and $\{f(x + td) \ : \ t \geq 0\}$ is bounded below, we have $\phi(t) \to +\infty$ as $t \uparrow \infty$.
Continuity $\implies\ \exists\ \hat{t} > 0$ such that $\phi(\hat{t}) = 0$:

Wolfe Conditions

Set $\phi(t) = f(x+td) - (f(x) + c_1 t f'(x;d))$. Then

$$\phi(0) = 0 \quad \text{and} \quad \phi'(0) = (1-c_1)f'(x;d) < 0.$$

So $\exists\ \bar{t} > 0$ such that $\phi(t) < 0$ for $t \in (0, \bar{t})$.
Since $f'(x;d) < 0$ and $\{f(x+td) \ :\ t \geq 0\}$ is bounded below, we have $\phi(t) \to +\infty$ as $t \uparrow \infty$.
Continuity $\implies\ \exists\ \hat{t} > 0$ such that $\phi(\hat{t}) = 0$:

$$t^* = \inf\left\{\hat{t}\ \middle|\, 0 < t,\ \phi(\hat{t}) = 0\right\}.$$

Since $\phi(t) < 0$ for $t \in (0, \bar{t})$, $t^* > 0$ and by continuity $\phi(t^*) = 0$.

Wolfe Conditions

Set $\phi(t) = f(x+td) - (f(x) + c_1 t f'(x;d))$. Then

$$\phi(0) = 0 \quad \text{and} \quad \phi'(0) = (1-c_1)f'(x;d) < 0.$$

So $\exists\ \bar{t} > 0$ such that $\phi(t) < 0$ for $t \in (0, \bar{t})$.
Since $f'(x;d) < 0$ and $\{f(x+td) \,:\, t \geq 0\}$ is bounded below, we have $\phi(t) \to +\infty$ as $t \uparrow \infty$.
Continuity $\implies$ $\exists\ \hat{t} > 0$ such that $\phi(\hat{t}) = 0$:

$$t^* = \inf \left\{ \hat{t} \;\middle|\; 0 < t,\ \phi(\hat{t}) = 0 \right\}.$$

Since $\phi(t) < 0$ for $t \in (0, \bar{t})$, $t^* > 0$ and by continuity $\phi(t^*) = 0$.
MVT $\implies$ $\exists\ \tilde{t} \in (0, t^*)$ with $\phi'(\tilde{t}) = 0$. That is,

$$\nabla f(x + \tilde{t}d)^T d = c_1 \nabla f(x)^T d$$

Wolfe Conditions

Set $\phi(t) = f(x+td) - (f(x) + c_1 t f'(x;d))$. Then

$$\phi(0) = 0 \quad \text{and} \quad \phi'(0) = (1-c_1)f'(x;d) < 0.$$

So $\exists\ \bar{t} > 0$ such that $\phi(t) < 0$ for $t \in (0, \bar{t})$.
Since $f'(x;d) < 0$ and $\{f(x+td) \ :\ t \geq 0\}$ is bounded below, we have $\phi(t) \to +\infty$ as $t \uparrow \infty$.
Continuity $\implies$ $\exists\ \hat{t} > 0$ such that $\phi(\hat{t}) = 0$:

$$t^* = \inf\left\{\hat{t}\ \middle|\ 0 < t,\ \phi(\hat{t}) = 0\right\}.$$

Since $\phi(t) < 0$ for $t \in (0, \bar{t})$, $t^* > 0$ and by continuity $\phi(t^*) = 0$.
MVT $\implies$ $\exists\ \tilde{t} \in (0, t^*)$ with $\phi'(\tilde{t}) = 0$. That is,

$$\nabla f(x + \tilde{t}d)^T d = c_1 \nabla f(x)^T d > c_2 \nabla f(x)^T d.$$

We also have

$$f(x+td) - (f(x) + c_1 \tilde{t} \nabla f(x)^T d) < 0\ .$$

A Bisection Method for the Weak Wolfe Conditions

INITIALIZATION: Choose $0 < c_1 < c_2 < 1$, and set $\alpha = 0$, $t = 1$, and $\beta = +\infty$.

REPEAT

If $f(x + td) > f(x) + c_1 t f'(x; d)$,
set $\beta = t$ and reset $t = \frac{1}{2}(\alpha + \beta)$.
Else if $f'(x + td; d) < c_2 f'(x; d)$,
set $\alpha = t$ and reset

$$t = \begin{cases} 2\alpha, & \text{if } \beta = +\infty \\ \frac{1}{2}(\alpha + \beta), & \text{otherwise.} \end{cases}$$

Else, STOP.

END REPEAT

Weak Wolf Decsent Algorithm

Let $\{x^\nu\}$ be a sequence initiated at x^0 and generated by the following algorithm:

Weak Wolf Decsent Algorithm

Let $\{x^\nu\}$ be a sequence initiated at x^0 and generated by the following algorithm:

Step 0: Set $k = 0$.

Weak Wolf Decsent Algorithm

Let $\{x^\nu\}$ be a sequence initiated at x^0 and generated by the following algorithm:

Step 0: Set $k = 0$.

Step 1: Choose $d^k \in \mathbb{R}^n$ such that $f'(x^k; d^k) < 0$.

Weak Wolf Decsent Algorithm

Let $\{x^\nu\}$ be a sequence initiated at x^0 and generated by the following algorithm:

Step 0: Set $k = 0$.

Step 1: Choose $d^k \in \mathbb{R}^n$ such that $f'(x^k; d^k) < 0$.
If no such d^k exists, then STOP.
(First-order necessary conditions satisfied at x^k.)

Weak Wolf Decsent Algorithm

Let $\{x^\nu\}$ be a sequence initiated at x^0 and generated by the following algorithm:

Step 0: Set $k = 0$.

Step 1: Choose $d^k \in \mathbb{R}^n$ such that $f'(x^k; d^k) < 0$.
If no such d^k exists, then STOP.
(First-order necessary conditions satisfied at x^k.)

Step 2: Let t^k be a stepsize satisfying the Weak Wolfe conditions.

Weak Wolf Decsent Algorithm

Let $\{x^\nu\}$ be a sequence initiated at x^0 and generated by the following algorithm:

Step 0: Set $k = 0$.

Step 1: Choose $d^k \in \mathbb{R}^n$ such that $f'(x^k; d^k) < 0$.
If no such d^k exists, then STOP.
(First-order necessary conditions satisfied at x^k.)

Step 2: Let t^k be a stepsize satisfying the Weak Wolfe conditions.
If no such t^k exists, then STOP.
(The function f is unbounded below.)

Weak Wolf Decsent Algorithm

Let $\{x^\nu\}$ be a sequence initiated at x^0 and generated by the following algorithm:

Step 0: Set $k = 0$.

Step 1: Choose $d^k \in \mathbb{R}^n$ such that $f'(x^k; d^k) < 0$.
If no such d^k exists, then STOP.
(First-order necessary conditions satisfied at x^k.)

Step 2: Let t^k be a stepsize satisfying the Weak Wolfe conditions.
If no such t^k exists, then STOP.
(The function f is unbounded below.)

Step 3: Set $x^{k+1} = x^k + t_k d^k$ and reset $k = k + 1$.
Return to Step 1.

Weak Wolf Convergence

Let $f : \mathbb{R}^n \to \mathbb{R}$, $x^0 \in \mathbb{R}^n$, and $0 < c_1 < c_2 < 1$.

Weak Wolf Convergence

Let $f : \mathbb{R}^n \to \mathbb{R}$, $x^0 \in \mathbb{R}^n$, and $0 < c_1 < c_2 < 1$.
Assume $\nabla f(x)$ exists and is Lipschitz continuous on an open set containing the set $\{x \mid f(x) \leq f(x^0)\}$.

Weak Wolf Convergence

Let $f : \mathbb{R}^n \to \mathbb{R}$, $x^0 \in \mathbb{R}^n$, and $0 < c_1 < c_2 < 1$.
Assume $\nabla f(x)$ exists and is Lipschitz continuous on an open set containing the set $\{x \,|\, f(x) \leq f(x^0)\}$.
If $\{x^\nu\}$ be a sequence initiated at x^0 and generated by the Weak Wolfe Descent Algorithm, then one of the following must occur:

Weak Wolf Convergence

Let $f : \mathbb{R}^n \to \mathbb{R}$, $x^0 \in \mathbb{R}^n$, and $0 < c_1 < c_2 < 1$.
Assume $\nabla f(x)$ exists and is Lipschitz continuous on an open set containing the set $\left\{x \;\middle|\; f(x) \leq f(x^0)\right\}$.
If $\{x^\nu\}$ be a sequence initiated at x^0 and generated by the Weak Wolfe Descent Algorithm, then one of the following must occur:

(i) The algorithm terminates finitely at a 1^{st}-order stationary point for f.

Weak Wolf Convergence

Let $f : \mathbb{R}^n \to \mathbb{R}$, $x^0 \in \mathbb{R}^n$, and $0 < c_1 < c_2 < 1$.
Assume $\nabla f(x)$ exists and is Lipschitz continuous on an open set containing the set $\{x \mid f(x) \leq f(x^0)\}$.
If $\{x^\nu\}$ be a sequence initiated at x^0 and generated by the Weak Wolfe Descent Algorithm, then one of the following must occur:

(i) The algorithm terminates finitely at a 1st-order stationary point for f.

(ii) For some k the stepsize selection procedure generates a sequence of trial stepsizes $t_{k\nu} \uparrow +\infty$ such that $f(x^k + t_{k\nu} d^k) \to -\infty$.

Weak Wolf Convergence

Let $f : \mathbb{R}^n \to \mathbb{R}$, $x^0 \in \mathbb{R}^n$, and $0 < c_1 < c_2 < 1$.
Assume $\nabla f(x)$ exists and is Lipschitz continuous on an open set containing the set $\left\{x \,\middle|\, f(x) \leq f(x^0)\right\}$.
If $\{x^\nu\}$ be a sequence initiated at x^0 and generated by the Weak Wolfe Descent Algorithm, then one of the following must occur:

(i) The algorithm terminates finitely at a 1^{st}-order stationary point for f.

(ii) For some k the stepsize selection procedure generates a sequence of trial stepsizes $t_{k\nu} \uparrow +\infty$ such that $f(x^k + t_{k\nu} d^k) \to -\infty$.

(iii) $f(x^k) \downarrow -\infty$.

Weak Wolf Convergence

Let $f : \mathbb{R}^n \to \mathbb{R}$, $x^0 \in \mathbb{R}^n$, and $0 < c_1 < c_2 < 1$.
Assume $\nabla f(x)$ exists and is Lipschitz continuous on an open set containing the set $\left\{x \,\middle|\, f(x) \le f(x^0)\right\}$.
If $\{x^\nu\}$ be a sequence initiated at x^0 and generated by the Weak Wolfe Descent Algorithm, then one of the following must occur:

(i) The algorithm terminates finitely at a 1^{st}-order stationary point for f.

(ii) For some k the stepsize selection procedure generates a sequence of trial stepsizes $t_{k\nu} \uparrow +\infty$ such that $f(x^k + t_{k\nu} d^k) \to -\infty$.

(iii) $f(x^k) \downarrow -\infty$.

(iv) $\sum_{k=0}^{\infty} \|\nabla f(x^k)\|^2 \cos^2\theta_k < +\infty$, where $\cos\theta_k = \dfrac{\nabla f(x^k)^T d^k}{\|\nabla f(x^k)\| \|d^k\|}$ for all $k = 1, 2, \ldots$.

Weak Wolf Convergence: Corollary

Let f and $\{x^k\}$ be as in the Theorem, and let $\{H_k\}$ be a sequence of symmetric positive definite matrices for which there exists $\overline{\lambda} > \underline{\lambda} > 0$ such that

$$\underline{\lambda}\|u\|^2 \leq u^T H_k u \leq \overline{\lambda}\|u\|^2 \ \forall\, u \in \mathbb{R}^n \text{ and } k = 1, 2, \dots \ .$$

Weak Wolf Convergence: Corollary

Let f and $\{x^k\}$ be as in the Theorem, and let $\{H_k\}$ be a sequence of symmetric positive definite matrices for which there exists $\overline{\lambda} > \underline{\lambda} > 0$ such that

$$\underline{\lambda}\|u\|^2 \le u^T H_k u \le \overline{\lambda}\|u\|^2 \;\forall\, u \in \mathbb{R}^n \text{ and } k = 1, 2, \ldots \;.$$

Let us further assume that f is bounded below.

Weak Wolf Convergence: Corollary

Let f and $\{x^k\}$ be as in the Theorem, and let $\{H_k\}$ be a sequence of symmetric positive definite matrices for which there exists $\overline{\lambda} > \underline{\lambda} > 0$ such that

$$\underline{\lambda}\|u\|^2 \leq u^T H_k u \leq \overline{\lambda}\|u\|^2 \;\forall\, u \in \mathbb{R}^n \text{ and } k = 1, 2, \ldots \;.$$

Let us further assume that f is bounded below.
If the search directions d^k are given by

$$d^k = -H_k \nabla f(x^k) \;\forall\, k = 1, 2, \ldots \;,$$

then $\nabla f(x^k) \to 0$.