Math 408A: Non-Linear Optimization

Lecture 10

January 27
We say that $F : \mathbb{R}^n \to \mathbb{R}^m$ is **Lipschitz continuous** relative to a set $D \subset \mathbb{R}^n$ if there exists a constant $K \geq 0$ such that

$$\|F(x) - F(y)\| \leq K\|x - y\|$$

for all $x, y \in D$.  

Fact: Lipschitz continuity implies uniform continuity.

Examples:

1. $f(x) = x - 1$ is continuous on $(0, 1)$, but it is not uniformly continuous on $(0, 1)$.

2. $f(x) = \sqrt{x}$ is uniformly continuous on $[0, 1]$, but it is not Lipschitz continuous on $[0, 1]$. 
Lipschitz Continuity

We say that $F : \mathbb{R}^n \to \mathbb{R}^m$ is **Lipschitz continuous** relative to a set $D \subset \mathbb{R}^n$ if there exists a constant $K \geq 0$ such that

$$\|F(x) - F(y)\| \leq K\|x - y\|$$

for all $x, y \in D$.

Fact: Lipschitz continuity implies uniform continuity.
Lipschitz Continuity

We say that $F : \mathbb{R}^n \to \mathbb{R}^m$ is **Lipschitz continuous** relative to a set $D \subset \mathbb{R}^n$ if there exists a constant $K \geq 0$ such that

$$\|F(x) - F(y)\| \leq K\|x - y\|$$

for all $x, y \in D$.

Fact: Lipschitz continuity implies uniform continuity.

Examples:
We say that $F : \mathbb{R}^n \to \mathbb{R}^m$ is **Lipschitz continuous** relative to a set $D \subset \mathbb{R}^n$ if there exists a constant $K \geq 0$ such that

$$
\|F(x) - F(y)\| \leq K\|x - y\|
$$

for all $x, y \in D$.

Fact: Lipschitz continuity implies uniform continuity.

Examples:

1. $f(x) = x^{-1}$ is continuous on $(0, 1)$, but it is not uniformly continuous on $(0, 1)$. 

We say that $F : \mathbb{R}^n \to \mathbb{R}^m$ is **Lipschitz continuous** relative to a set $D \subset \mathbb{R}^n$ if there exists a constant $K \geq 0$ such that

$$
\|F(x) - F(y)\| \leq K\|x - y\|
$$

for all $x, y \in D$.

Fact: Lipschitz continuity implies uniform continuity.

Examples:

1. $f(x) = x^{-1}$ is continuous on $(0, 1)$, but it is not uniformly continuous on $(0, 1)$.

2. $f(x) = \sqrt{x}$ is uniformly continuous on $[0, 1]$, but it is not Lipschitz continuous on $[0, 1]$. 
Fact: If $F'$ exists and is continuous on a compact convex set $D \subset \mathbb{R}^m$, then $F$ is Lipschitz continuous on $D$. 
Fact: If $F'$ exists and is continuous on a compact convex set $D \subset \mathbb{R}^m$, then $F$ is Lipschitz continuous on $D$. Proof: By the MVT

$$\|F(x) - F(y)\| \leq \left( \sup_{z \in [x,y]} \|F'(z)\| \right) \|x - y\|.$$
Lipschitz Continuity and the Derivative

Fact: If $F'$ exists and is continuous on a compact convex set $D \subset \mathbb{R}^m$, then $F$ is Lipschitz continuous on $D$. Proof: By the MVT

\[ \|F(x) - F(y)\| \leq \left( \sup_{z \in [x, y]} \|F'(z)\| \right) \|x - y\|. \]

Lipschitz continuity is almost but not quite a differentiability hypothesis. The Lipschitz constant provides bounds on rate of change.
Fact: If \( F' \) exists and is continuous on a compact convex set \( D \subset \mathbb{R}^m \), then \( F \) is Lipschitz continuous on \( D \). Proof: By the MVT

\[
\| F(x) - F(y) \| \leq \left( \sup_{z \in [x,y]} \| F'(z) \| \right) \| x - y \|.
\]

Lipschitz continuity is almost but not quite a differentiability hypothesis. The Lipschitz constant provides bounds on rate of change. For example, every norm is Lipschitz continuous, but not differentiable at the origin.
Lipschitz Continuity and the Quadratic Bound Lemma

**QBL:** Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be such that $F'$ is Lipschitz continuous on the convex set $D \subset \mathbb{R}^n$. Then

$$\|F(y) - (F(x) + F'(x)(y - x))\| \leq \frac{K}{2}\|y - x\|^2$$

for all $x, y \in D$ where $K$ is a Lipschitz constant for $F'$ on $D$. 
Proof:

\[ F(y) - F(x) - F'(x)(y - x) = \int_0^1 F'(x + t(y - x))(y - x)dt - F'(x)(y - x) \]

\[ = \int_0^1 [F'(x + t(y - x)) - F'(x)](y - x)dt \]

\[ \|F(y) - (F(x) + F'(x)(y - x))\| = \| \int_0^1 [F'(x + t(y - x)) - F'(x)](y - x)dt \| \]

\[ \leq \int_0^1 \|[F'(x + t(y - x)) - F'(x)](y - x)\|dt \]

\[ \leq \int_0^1 \|F'(x + t(y - x)) - F'(x)\| \|y - x\|dt \]

\[ \leq \int_0^1 Kt\|y - x\|^2 dt \]

\[ = \frac{K}{2} \|y - x\|^2. \]
Theorem: Convergence of Backtracking

Let \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) and \( x_0 \in \mathbb{R} \) be such that \( f \) is differentiable on \( \mathbb{R}^n \) with \( \nabla f \) Lipschitz continuous on an open convex set containing the set \( \{ x : f(x) \leq f(x_0) \} \). Let \( \{ x^k \} \) be the sequence satisfying \( x^{k+1} = x^k \) if \( \nabla f(x^k) = 0 \); otherwise,

\[
x^{k+1} = x_k + t_k d^k,
\]
where \( d^k \) satisfies \( f'(x^k; d^k) < 0 \),

and \( t_k \) is chosen by the backtracking stepsize selection method. Then one of the following statements must be true:

(i) There is a \( k_0 \) such that \( \nabla f'(x^{k_0}) = 0 \).

(ii) \( f(x^k) \downarrow -\infty \)

(iii) The sequence \( \{ \|d^k\| \} \) diverges \( (\|d^k\| \rightarrow \infty) \).

(iv) For every subsequence \( J \subset \mathbb{N} \) for which \( \{d^k : k \in J\} \) is bounded, we have

\[
\lim_{k \in J} f'(x^k; d^k) = 0.
\]
Corollary 1: If the sequences \( \{d^k\} \) and \( \{f(x^k)\} \) are bounded, then

\[
\lim_{k \to \infty} f'(x^k; d^k) = 0.
\]
Corollary 1: If the sequences \( \{d^k\} \) and \( \{f(x^k)\} \) are bounded, then
\[
\lim_{k \to \infty} f'(x^k; d^k) = 0.
\]

Corollary 2: If \( d^k = -\nabla f'(x^k)/\|\nabla f(x^k)\| \) is the Cauchy direction for all \( k \), then every accumulation point, \( \bar{x} \), of the sequence \( \{x^k\} \) satisfies \( \nabla f(\bar{x}) = 0 \).
Corollary 3: Let us further assume that $f$ is twice continuously differentiable and that there is a $\beta > 0$ such that, for all $u \in \mathbb{R}^n$, $\beta \|u\|^2 < u^T \nabla^2 f(x) u$ on $\{x : f(x) \leq f(x^0)\}$. If the Basic Backtracking algorithm is implemented using the Newton search directions,

$$d^k = -\nabla^2 f(x^k)^{-1} \nabla f(x^k),$$

then every accumulation point, $\bar{x}$, of the sequence \{\(x^k\)\} satisfies $\nabla f(\bar{x}) = 0$. 


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

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

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

Choose \( 0 < c_1 < c_2 < 1. \)

Strong Wolfe Conditions:

\[ f(x^k + t_k d^k) \leq f(x^k) + c_1 t_k f'(x^k; d^k) \]

\[ |f'(x^k + t_k d^k; d^k)| \leq c_2 |f'(x^k; d^k)|. \]
Wolfe Conditions

Weak Wolfe Conditions:

\[ f(x^k + t_k d^k) \leq f(x^k) + c_1 t_k f'(x^k; d^k) \]
\[ c_2 f'(x^k; d^k) \leq f'(x^k + t_k d^k; d^k). \]
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

$$\left\{ t \mid t > 0, f'(x + td; d) \geq c_2 f'(x; d), \text{ and } f(x + td) \leq f(x) + c_1 tf'(x; d) \right\}$$

has non–empty interior.
Wolfe Conditions

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

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

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

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

So \( \exists \tilde{t} > 0 \) such that \( \phi(t) < 0 \) for \( t \in (0, \tilde{t}) \).
Wolfe Conditions

Set \( \phi(t) = f(x + td) - (f(x) + c_1 tf'(x; d)) \). 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 \( \phi(t) \rightarrow +\infty \) as \( t \uparrow \infty \).
Wolfe Conditions

Set $\phi(t) = f(x + td) - (f(x) + c_1 tf'(x; d))$. 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 $\iff$ $\exists \hat{t} > 0$ such that $\phi(\hat{t}) = 0$:
Wolfe Conditions

Set $\phi(t) = f(x + td) - (f(x) + c_1 tf'(x; d))$. 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 $\phi(\hat{t}) = 0$:

$$t^* = \inf \{\hat{t} \mid 0 \leq t, \phi(\hat{t}) = 0\}.$$  

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

Lecture 10  
Math 408A: Non-Linear Optimization
Wolfe Conditions

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

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

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

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 $\phi(\hat{t}) = 0$:

$$
t^* = \inf \{\hat{t} \mid 0 \leq t, \ \phi(\hat{t}) = 0 \}.
$$

Since $\phi(t) < 0$ for $t \in (0, \tilde{t})$, $t^* > 0$ 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 \( \phi(t) = f(x + td) - (f(x) + c_1 tf'(x; d)) \). 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 \( t \to \infty \).

Continuity \( \implies \exists \hat{t} > 0 \) such that \( \phi(\hat{t}) = 0 \):

\[
t^* = \inf \{ \hat{t} \mid 0 \leq t, \; \phi(\hat{t}) = 0 \}.
\]

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 $f(x + td) > f(x) + c_1 tf'(x; d)$,

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**