# Math 408A
# Convergence of Backtracking Methods

# 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 $x, y \in D$.

# 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:

# 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:

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

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 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 $D$.*

# 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 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.

# 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. For example, every norm is Lipschitz continuous, but not differentiable at the origin.

# The Quadratic Bound Lemma (QBL)

Let $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ 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 $D$.

# Lipschitz Continuity and the Quadratic Bound Lemma

Proof:

$$
\begin{aligned}
F(y) - F(x) - F'(x)(y-x) &= \int_0^1 F'(x + t(y-x))(y-x)dt - F'(x)(y-x) \\
&= \int_0^1 [F'(x + t(y-x)) - F'(x)](y-x)dt \\
\|F(y) - (F(x) + F'(x)(y-x))\| &= \| \int_0^1 [F'(x + t(y-x)) - F'(x)](y-x)dt\| \\
&\leq \int_0^1 \|(F'(x + t(y-x) - F'(x))(y-x)\|dt \\
&\leq \int_0^1 \|F'(x + t(y-x)) - F'(x)\| \, \|y-x\|dt \\
&\leq \int_0^1 Kt\|y-x\|^2 dt \\
&= \frac{K}{2}\|y-x\|^2.
\end{aligned}
$$

# Theorem: Convergence of Backtracking

Let $f : \mathbb{R}^n \to \mathbb{R}$ 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,

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

and $t_k$ is chosen by the backtracking stepsize selection method.

# Theorem: Convergence of Backtracking

Let $f : \mathbb{R}^n \to \mathbb{R}$ 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,

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

and $t_k$ is chosen by the backtracking stepsize selection method. Then one of thefollowing statements must be true:

# Theorem: Convergence of Backtracking

Let $f : \mathbb{R}^n \to \mathbb{R}$ 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,

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

and $t_k$ is chosen by the backtracking stepsize selection method. Then one of thefollowing statements must be true:

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

# Theorem: Convergence of Backtracking

Let $f : \mathbb{R}^n \to \mathbb{R}$ 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,

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

and $t_k$ is chosen by the backtracking stepsize selection method. Then one of thefollowing statements must be true:

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

(ii) $f(x^k) \searrow -\infty$

# Theorem: Convergence of Backtracking

Let $f : \mathbb{R}^n \to \mathbb{R}$ 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,

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

and $t_k$ is chosen by the backtracking stepsize selection method. Then one of thefollowing statements must be true:

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

(ii) $f(x^k) \searrow -\infty$

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

# Theorem: Convergence of Backtracking

Let $f : \mathbb{R}^n \to \mathbb{R}$ 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,

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

and $t_k$ is chosen by the backtracking stepsize selection method. Then one of thefollowing statements must be true:

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

(ii) $f(x^k) \searrow -\infty$

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

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

$$\lim_{k \in J} f'(x^k; d^k) = 0.$$

# Corollaries

Corollary 1:*If the sequences* $\{d^k\}$ *and* $\{f(x^k)\}$ *are bounded, then*

$$\lim_{k\to\infty} f'(x^k; d^k) = 0.$$

# Corollaries

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,* $\overline{x}$*, of the sequence* $\{x^k\}$ *satisfies* $\nabla f(\overline{x}) = 0$*.*

# Corollaries

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, $\overline{x}$, of the sequence $\{x^k\}$ satisfies $\nabla f(\overline{x}) = 0$.*