Optimization in Machine Learning

Lecture 11: Lipschitz Continuity and Smoothness, Algorithms for Optimization and their analysis

Ganesh Ramakrishnan

Department of Computer Science
Dept of CSE, IIT Bombay
https://www.cse.iitb.ac.in/~ganesh

February, 2025

Convex Optimization Problem

- Formally, a convex optimization problem is an optimization problem of the form

$$\begin{aligned} &\textit{minimize } f(\mathbf{x}) \\ &\textit{subject to } \mathbf{x} \in \mathcal{X} \end{aligned}$$

  where f is a convex function, $\mathcal{X}$ is a convex set, and $\mathbf{x}$ is the optimization variable.
- if $\mathcal{X} = dom(f)$, this becomes unconstrained optimization.
- A special case (f is a convex function, g_i are convex functions, and h_i are affine functions, and $\mathbf{x}$ is the vector of optimization variables):

$$\begin{aligned} \textit{minimize } f(\mathbf{x}) & \\ \textit{subject to } g_i(\mathbf{x}) &\leq 0, \; i = 1, ..., m \\ h_i(\mathbf{x}) &= 0, i = 1, ..., p \end{aligned}$$

Convex Optimization Problem

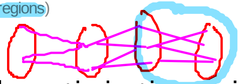
- Formally, a convex optimization problem is an optimization problem of the form

if $x^* \in int(\mathcal{X})$ then $0 \in \partial f(x^*)$ is sufficient

$$\begin{aligned} &\textit{minimize } f(\mathbf{x}) \\ &\textit{subject to } \mathbf{x} \in \mathcal{X} \end{aligned}$$

But the convexity of the set (convexity of the constraint set) is typically not relaxed

Relaxations are typically on the function f (need not be convex everywhere though it might be convex in some regions)

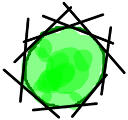


where f is a convex function, $\mathcal{X}$ is a convex set, and $\mathbf{x}$ is the optimization variable.

- if $\mathcal{X} = dom(f)$, this becomes unconstrained optimization.

Eg: Region could be defined as the region of last layer weights for a deep NN (that with all other layer weights frozen)

- A special case (f is a convex function, g_i are convex functions, and h_i are affine functions, and $\mathbf{x}$ is the vector of optimization variables):



$$\begin{aligned} &\textit{minimize } f(\mathbf{x}) \\ &\textit{subject to } g_i(\mathbf{x}) \leq 0, \; i = 1, \ldots, m \\ &\qquad\qquad\quad h_i(\mathbf{x}) = 0, \; i = 1, \ldots, p \end{aligned}$$

Intersection of convex 0-sublevel sets of g_i's

Typical convex set. In fact the dual description for convex sets should allow them to be specified as intersections of such inequalities

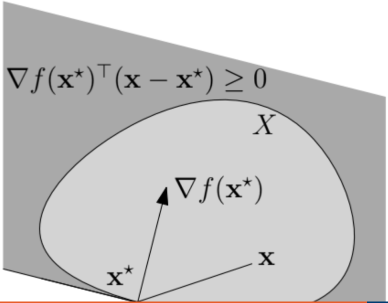
Either or both of m and p could be 0

Optimality Condition for Constrained Optimization

- Lemma: Suppose that f is convex and differentiable over an open domain $dmn(f)$. Let $\mathcal{X} \subseteq dom(f)$ be a convex set. A point $\mathbf{x}^*$ is a minimizer of f over $\mathcal{X}$ if and only if

$$\nabla f(\mathbf{x}^*)^T(\mathbf{x} - \mathbf{x}^*) \geq 0, \forall \mathbf{x} \in \mathcal{X}$$

- Note that the Condition for Unconstrained minimization becomes a special case.
- Nice geometric interpretation: The gradient $\nabla f(\mathbf{x}^*)$ at $\mathbf{x}^*$ is in the same direction as $\mathbf{x} - \mathbf{x}^*$ for any $\mathbf{x} \in dmn(f)$.

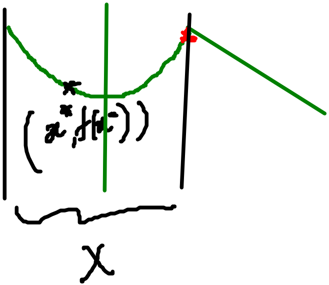


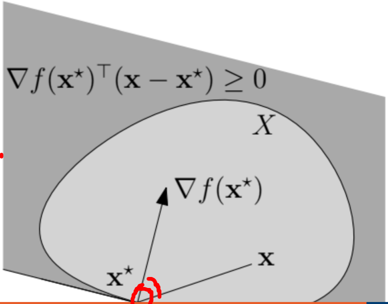
Optimality Condition for Constrained Optimization

- Lemma: Suppose that f is convex and differentiable over an open domain $dmn(f)$. Let $\mathcal{X} \subseteq dom(f)$ be a convex set. A point $\mathbf{x}^*$ is a minimizer of f over $\mathcal{X}$ if and only if

$$\nabla f(\mathbf{x}^*)^T(\mathbf{x} - \mathbf{x}^*) \geq 0, \forall \mathbf{x} \in \mathcal{X}$$

- Note that the Condition for Unconstrained minimization becomes a special case.
- Nice geometric interpretation: The gradient $\nabla f(\mathbf{x}^*)$ at $\mathbf{x}^*$ is in the same direction as $\mathbf{x} - \mathbf{x}^*$ for any $\mathbf{x} \in dmn(f)$.





Case 1: $x^* \in \text{int}(\mathcal{X})$
$0 \in \partial f(x^*)$ or $\nabla f(x^*) = 0$ is sufficient condition

Case 2: $x^* \in \text{Bnd}(\mathcal{X})$

Example: Linear and Quadratic Programs

- Linear Program (LP) is a special case of a convex optimization problem:

$$\begin{aligned} \textit{minimize } & \mathbf{c}^T\mathbf{x} \\ \textit{subject to } & A\mathbf{x} \leq b \end{aligned}$$

- Another special case is Quadratic Programs (QP):

$$\begin{aligned} \textit{minimize } & 1/2\mathbf{x}^T Q\mathbf{x} \\ \textit{subject to } & A\mathbf{x} \leq b \end{aligned}$$

- The QP is a convex optimization problem only if

Example: Linear and Quadratic Programs

- Linear Program (LP) is a special case of a convex optimization problem:

$$\begin{aligned} \textit{minimize } & \mathbf{c}^T\mathbf{x} \\ \textit{subject to } & A\mathbf{x} \leq b \end{aligned}$$

- Another special case is Quadratic Programs (QP):

$$\begin{aligned} \textit{minimize } & 1/2\mathbf{x}^T Q\mathbf{x} \\ \textit{subject to } & A\mathbf{x} \leq b \end{aligned}$$

- The QP is a convex optimization problem only if Q is positive semi-definite,

Outline of next few topics

- Local and Global Minimum [Done]
- Sufficient Subgradient condition for Global Minimum [Done]
- Convexity: Local & Global Minimum [Done]
- Strict Convexity and Unique minimum [Done]
- General Constrained Convex Problem and Optimality Condition [Preliminarily Done]
- Toward Algorithms for Optimization
  - Lipschitz continuity and (Lipschitz) Smoothness, Strong convexity
    - See lectures 18-19 of my Convex Optimization course
  - Gradient Descent and Analysis for Lipschitz Continuous, Smooth and Strongly Convex functions
  - Nestorov's accelerated gradient descent...

Reading on Lipschitz Continuity

- Good Notes on Lipshcitz Algebra: https://moodle.iitb.ac.in/mod/resource/view.php?id=33825
- Juha Heinonen, *Lectures on Lipschitz Analysis*, http://www.math.jyu.fi/research/reports/rep100.pdf
- https://ljk.imag.fr/membres/Anatoli.Iouditski/cours/convex/chapitre_3.pdf
- Nice Blog on Lipschitz Continuity: https://xingyuzhou.org/blog/notes/Lipschitz-gradient. The author has a similar blog on Strong Convexity: http://xingyuzhou.org/blog/notes/strong-convexity
- Papers attempting to compute the local Lipschitz constant of various deep networks
  - https://arxiv.org/abs/2002.03657
  - https://papers.nips.cc/paper/2020/hash/5227fa9a19dce7ba113f50a405dcaf09-Abstract.html

Lipschitz Continuity

- A function f is Lipschitz continuous with Lipschitz constant L if

$$|f(x) - f(y)| \leq L||x - y||$$

Lipschitz Continuity

- A function f is Lipschitz continuous with Lipschitz constant L if

$$|f(x) - f(y)| \leq L||x - y||$$

That is, the function is not changing too soon/fast

Rate of change $\approx \dfrac{|f(x) - f(y)|}{\|x - y\|} \leq L$

That is, a linearly scaled upper bound on the absolute value of the difference of function values at those points

Lipschitz Continuity

- A function f is Lipschitz continuous with Lipschitz constant L if

$$|f(x) - f(y)| \leq L||x - y||$$

- Basically Lipschitz continuity limits how fast a function changes.

Lipschitz Continuity

(need not be necessarily differentiable)

- A function f is Lipschitz continuous with Lipschitz constant L if

$$|f(x) - f(y)| \leq L||x - y||$$

$y = x + hv$

- Basically Lipschitz continuity limits how fast a function changes.

IF f is differentiable

$$\left| \lim_{h \to 0} \frac{f(x+hv) - f(x)}{h} \right| = |D_v f(x)| \leq L$$

Lipschitz Continuity

- A function f is Lipschitz continuous with Lipschitz constant L if

$$|f(x) - f(y)| \leq L||x - y||$$

- Basically Lipschitz continuity limits how fast a function changes.
- Lipschitz continuity implies that any $x \in dom(f)$ the (sub)gradient $h \in \partial f(x)$ satisfies $||h|| \leq L$.

Lipschitz Continuity

- A function f is Lipschitz continuous with Lipschitz constant L if

$$|f(x) - f(y)| \leq L||x - y||$$

- Basically Lipschitz continuity limits how fast a function changes.
- Lipschitz continuity implies that any $x \in dom(f)$ the (sub)gradient $h \in \partial f(x)$ satisfies $||h|| \leq L$.

IF f is differentiable $\|\nabla^T f(x) v\| = \left| \lim_{h \to 0} \frac{f(x+hv) - f(x)}{h} \right| = |D_v f(x)| \leq L$

Lipschitz Continuity

- A function f is Lipschitz continuous with Lipschitz constant L if

$$|f(x) - f(y)| \leq L||x - y||$$

- Basically Lipschitz continuity limits how fast a function changes.
- Lipschitz continuity implies that any $x \in dom(f)$ the (sub)gradient $h \in \partial f(x)$ satisfies $||h|| \leq L$.
- Calculus of Lipschitz continuity (some properties):

Lipschitz Continuity

- A function f is Lipschitz continuous with Lipschitz constant L if

$$|f(x) - f(y)| \leq L||x - y||$$

- Basically Lipschitz continuity limits how fast a function changes.
- Lipschitz continuity implies that any $x \in dom(f)$ the (sub)gradient $h \in \partial f(x)$ satisfies $||h|| \leq L$.
- Calculus of Lipschitz continuity (some properties):

If f_1 is L_1 Lipschitz & f_2 is L_2 Lipschitz

Then $f_1 + f_2$ is $L_1 + L_2$ Lipschitz

Δ^{le} Ineq

$||a-b|| \leq ||a-c|| + ||c-b||$

Lipschitz Continuity

- A function f is Lipschitz continuous with Lipschitz constant L if

$$|f(x) - f(y)| \leq L||x - y||$$

- Basically Lipschitz continuity limits how fast a function changes.
- Lipschitz continuity implies that any $x \in dom(f)$ the (sub)gradient $h \in \partial f(x)$ satisfies $||h|| \leq L$.
- Calculus of Lipschitz continuity (some properties):
  - If f_1 is L_1-Lipschitz continuous and f_2 is L_2-Lipschitz continuous, then $f_1 + f_2$ is $L_1 + L_2$ Lipschitz continuous

Lipschitz Continuity

- A function f is Lipschitz continuous with Lipschitz constant L if

$$|f(x) - f(y)| \leq L||x - y||$$

- Basically Lipschitz continuity limits how fast a function changes.
- Lipschitz continuity implies that any $x \in dom(f)$ the (sub)gradient $h \in \partial f(x)$ satisfies $||h|| \leq L$.
- Calculus of Lipschitz continuity (some properties):
  - If f_1 is L_1-Lipschitz continuous and f_2 is L_2-Lipschitz continuous, then $f_1 + f_2$ is $L_1 + L_2$ Lipschitz continuous

Product of bounded L_1 & L_2 Lipschitz cts fns (Homework)

|f(x)g(x)-f(y)g(y)| $\leq$ Bnd|f(x)-f(y)|(g(x)-g(y)|

Lipschitz Continuity

- A function f is Lipschitz continuous with Lipschitz constant L if

$$|f(x) - f(y)| \leq L||x - y||$$

- Basically Lipschitz continuity limits how fast a function changes.
- Lipschitz continuity implies that any $x \in dom(f)$ the (sub)gradient $h \in \partial f(x)$ satisfies $||h|| \leq L$.
- Calculus of Lipschitz continuity (some properties):
  - If f_1 is L_1-Lipschitz continuous and f_2 is L_2-Lipschitz continuous, then $f_1 + f_2$ is $L_1 + L_2$ Lipschitz continuous
  - Product of two Lipschitz continuous and bounded functions is also Lipschitz continuous.

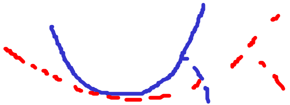
(Lipschitz) Smooth Functions

- A function f is called smooth if its gradient is Lipschitz continuous.

(Lipschitz) Smooth Functions

- A function f is called smooth if its gradient is Lipschitz continuous.

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\|$$



Roughly - the function has an upper bounded curvature $\left[v^T \nabla^2 f v \le L\right]$

Recall: Strong convexity is about lower bounded function curvature $\left[\lambda \le v^T \nabla^2 f v\right]$
(that is, the function cannot be too flat)

(Lipschitz) Smooth Functions

- A function f is called smooth if its gradient is Lipschitz continuous.
- Gradient of f being Lipschitz continuous implies

$$||\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})|| \leq L||\mathbf{x} - \mathbf{y}||$$

(Lipschitz) Smooth Functions

- A function f is called smooth if its gradient is Lipschitz continuous.
- Gradient of f being Lipschitz continuous implies

$$||\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})|| \leq L||\mathbf{x} - \mathbf{y}||$$

- This also implies that the Hessian satisfies

$$\nabla^2 f(\mathbf{x}) \preceq LI$$

Q: Example of function that is both strongly convex and Lipschitz smooth is x^2

- A function f is called smooth if its gradient is Lipschitz continuous.
- Gradient of f being Lipschitz continuous implies

$$||\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})|| \leq L||\mathbf{x} - \mathbf{y}||$$

- This also implies that the Hessian satisfies

$$\nabla^2 f(\mathbf{x}) \preceq LI$$

$$\nabla^2 f(x) \succeq mI$$

Strong convexity

$m \leq L$?

Quadratic upper bound ???????????????????

Strong convexity Quadratic lower bound $f(y) \geq f(x) + \nabla^T f(x)(y-x) + \frac{m}{2}||y-x||^2$

(Lipschitz) Smooth Functions

- A function f is called smooth if its gradient is Lipschitz continuous.
- Gradient of f being Lipschitz continuous implies

$$||\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})|| \leq L||\mathbf{x} - \mathbf{y}||$$

- This also implies that the Hessian satisfies

$$\nabla^2 f(\mathbf{x}) \preceq LI$$

- Lemma: If a convex function f is smooth (*i.e.*, has Lipschitz continuous gradients) then:

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla^\top f(\mathbf{x})(\mathbf{y} - \mathbf{x}) + \frac{L}{2}||\mathbf{y} - \mathbf{x}||^2$$

Q: Example of function that is both strongly convex and Lipschitz smooth is x^2

- A function f is called smooth if its gradient is Lipschitz continuous.
- Gradient of f being Lipschitz continuous implies

$$||\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})|| \leq L||\mathbf{x} - \mathbf{y}||$$

- This also implies that the Hessian satisfies

$$\nabla^2 f(\mathbf{x}) \preceq LI$$

Strong convexity $\nabla^2 f(x) \succeq mI$

- Lemma: If a convex function f is smooth (*i.e.*, has Lipschitz continuous gradients) then:

Quadratic upper bound

$$\boxed{f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla^\top f(\mathbf{x})(\mathbf{y} - \mathbf{x}) + \frac{L}{2}||\mathbf{y} - \mathbf{x}||^2}$$

$m \leq L$?

Strong convexity
Quadratic lower bound

$$f(y) \geq f(x) + \nabla^T f(x)(y-x) + \frac{m}{2}||y-x||^2$$

Examples of Smooth Functions

- Consider $f(x) = x^2$. Note that $\nabla f(x) = 2x$.

Examples of Smooth Functions

- Consider $f(x) = x^2$. Note that $\nabla f(x) = 2x$.

$$\nabla^2 f(x) = 2$$

$$\underbrace{1}_{m} \leq 2 \leq \underbrace{3}_{L}$$

For strong convexity

For Lipschitz smoothness

Examples of Smooth Functions

- Consider $f(x) = x^2$. Note that $\nabla f(x) = 2x$.
- The condition of Lipschitz smoothness asks for the existence of a L such that

$$||\nabla f(x) - \nabla f(y)|| \leq L||x - y||$$

Examples of Smooth Functions

- Consider $f(x) = x^2$. Note that $\nabla f(x) = 2x$.
- The condition of Lipschitz smoothness asks for the existence of a L such that

$$||\nabla f(x) - \nabla f(y)|| \leq L||x - y||$$

- In this case, this implies $2|y - x| \leq L|y - x|$ which means that $f(x) = x^2$ is Lipschitz smooth for any $L \geq 2$.

Examples of Smooth Functions

- Consider $f(x) = x^2$. Note that $\nabla f(x) = 2x$.
- The condition of Lipschitz smoothness asks for the existence of a L such that

$$||\nabla f(x) - \nabla f(y)|| \leq L||x - y||$$

- In this case, this implies $2|y - x| \leq L|y - x|$ which means that $f(x) = x^2$ is Lipschitz smooth for any $L \geq 2$.
- Is $f(x) = x^4$ Lipschitz smooth? What about $f(x) = x^3/3$? Let us study the latter. $\nabla f(x) = x^2$. For, Lipschitz continuity of $\nabla f(x)$, we ask the question: '*Does there exist an L such that $|x^2 - y^2| \leq L|x - y|$*'?

Examples of Smooth Functions

- Consider $f(x) = x^2$. Note that $\nabla f(x) = 2x$.
- The condition of Lipschitz smoothness asks for the existence of a L such that

$$||\nabla f(x) - \nabla f(y)|| \leq L||x - y||$$

- In this case, this implies $2|y - x| \leq L|y - x|$ which means that $f(x) = x^2$ is Lipschitz smooth for any $L \geq 2$.
- Is $f(x) = x^4$ Lipschitz smooth? What about $f(x) = x^3/3$? Let us study the latter. $\nabla f(x) = x^2$. For, Lipschitz continuity of $\nabla f(x)$, we ask the question: '*Does there exist an L such that* $\underbrace{|x^2 - y^2| \leq L|x - y|}$'?

Ans: Will require constraining the domain

$$|x-y||x+y| = |x^2-y^2| \overset{?}{\leq} L|x-y|$$

Ans: For any $|x| \leq \alpha$

$$|x+y| \leq 2\alpha = L$$

Examples of Smooth Functions

- Consider $f(x) = x^2$. Note that $\nabla f(x) = 2x$.
- The condition of Lipschitz smoothness asks for the existence of a L such that

$$||\nabla f(x) - \nabla f(y)|| \leq L||x - y||$$

- In this case, this implies $2|y - x| \leq L|y - x|$ which means that $f(x) = x^2$ is Lipschitz smooth for any $L \geq 2$.
- Is $f(x) = x^4$ Lipschitz smooth? What about $f(x) = x^3/3$? Let us study the latter. $\nabla f(x) = x^2$. For, Lipschitz continuity of $\nabla f(x)$, we ask the question: '*Does there exist an L such that* $|x^2 - y^2| \leq L|x - y|$'? This requires $|x + y| \leq L$. This means that globally, this is not Lipschitz continuous though it can be locally!
- Message: Only functions of asymptotically at most quadratic growth can be smooth globally.

Examples of Smooth Functions

- Consider $f(x) = x^2$. Note that $\nabla f(x) = 2x$.
- The condition of Lipschitz smoothness asks for the existence of a L such that

$$||\nabla f(x) - \nabla f(y)|| \leq L||x - y||$$

- In this case, this implies $2|y - x| \leq L|y - x|$ which means that $f(x) = x^2$ is Lipschitz smooth for any $L \geq 2$.
- Is $f(x) = x^4$ Lipschitz smooth? What about $f(x) = x^3/3$? Let us study the latter. $\nabla f(x) = x^2$. For, Lipschitz continuity of $\nabla f(x)$, we ask the question: '*Does there exist an L such that* $|x^2 - y^2| \leq L|x - y|$'? This requires $|x + y| \leq L$. This means that globally, this is not Lipschitz continuous though it can be locally!
- Message: Only functions of asymptotically at most quadratic growth can be smooth globally.

Q: What about this function? Study as homework

Better understanding of Lipschitz continuity and smoothness

- **Consider:** $f'(x) = |x|$

Better understanding of Lipschitz continuity and smoothness

- **Consider:** $f'(x) = |x|$

$$|f'(x) - f'(y)| \overset{?}{\leq} L\,|x-y|$$

Better understanding of Lipschitz continuity and smoothness

- **Consider:** $f'(x) = |x|$
- Since $|f'(x) - f'(y)| = ||x| - |y|| \leq |x - y|$, f' is Lipschitz continuous with $\underline{L = 1}$

(that is f is smooth and f' is L-continuous)

HOWEVER, DOES f '' (x) EVEN EXIST EVERYWHERE? THOUGH IT IS DIFFERENTIABLE ALMOST EVEYWHERE!

NOW HOW ABOUT THE LIPSCHITZ CONTINUITY OF f(x) ITSELF?

Better understanding of Lipschitz continuity and smoothness

- **Consider:** $f'(x) = |x|$
- Since $|f'(x) - f'(y)| = ||x| - |y|| \leq |x - y|$,
  f' is Lipschitz continuous with $L = 1$
- However, $f'(x)$ is not differentiable everywhere (not at 0). Such functions are called differentiable almost everywhere.

Better understanding of Lipschitz continuity and smoothness

- **Consider:** $f'(x) = |x|$
- Since $|f'(x) - f'(y)| = ||x| - |y|| \leq |x - y|$,
  f' is Lipschitz continuous with $L = 1$
- However, $f'(x)$ is not differentiable everywhere (not at 0). Such functions are called differentiable almost everywhere.

Eg: $\|x\|_1$ is also differentiable almost everywhere

Better understanding of Lipschitz continuity and smoothness

- **Consider:** $f'(x) = |x|$
- Since $|f'(x) - f'(y)| = ||x| - |y|| \leq |x - y|$,
  f' is Lipschitz continuous with $L = 1$
- However, $f'(x)$ is not differentiable everywhere (not at 0). Such functions are called differentiable almost everywhere.
- Is every sub-quadratic function Lipschitz smooth? Consider $f(x) = |x|^{3/2}$ on a closed set X s.t. $0 \in X$. Is this function smooth? Note that the gradient of $f'(x)$ is ∞ at $x = 0$.

Better understanding of Lipschitz continuity and smoothness

- **Consider:** $f'(x) = |x|$
- Since $|f'(x) - f'(y)| = ||x| - |y|| \leq |x - y|$,
  f' is Lipschitz continuous with $L = 1$
- However, $f'(x)$ is not differentiable everywhere (not at 0). Such functions are called differentiable almost everywhere.
- Is every sub-quadratic function Lipschitz smooth? Consider $f(x) = |x|^{3/2}$ on a closed set X s.t. $0 \in X$. Is this function smooth? Note that the gradient of $f'(x)$ is ∞ at $x = 0$.

  Not globally smooth
  Not smooth if 0 is in X
  But smooth in a bounded and closed interval
  not containing 0

$f'(x) = \frac{3}{2} x^{1/2}$ $(x > 0)$ $\quad f''(x) = \frac{3}{4} x^{-1/2}$

Lipschitz Continuity on closed and bounded functions

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

- If f is continuously differentiable almost everywhere, it is also Lipschitz continuous

[1]Theorem 1.41 of `https://moodle.iitb.ac.in/mod/resource/view.php?id=33825`

Lipschitz Continuity on closed and bounded functions

- If f is continuously differentiable almost everywhere, it is also Lipschitz continuous
- For functions over a bounded subset of $\Re^n$: f is continuous $\supseteq$ f is differentiable (almost everywhere) $=$ f is Lipschitz continuous[1] $\supseteq$ f is continuously differentiable $= \nabla f$ is continuous $\supseteq$ ∇f is differentiable (almost everywhere) $=$ f is smooth

[1] Theorem 1.41 of `https://moodle.iitb.ac.in/mod/resource/view.php?id=33825`

Lipschitz Continuity on closed and bounded functions

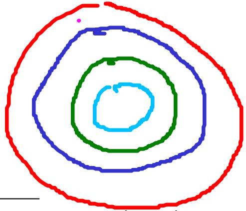
Eg: $x, e^x, |x|, f'(x) = |x|$

Eg: $x, |x|, e^x, f'(x) = |x|$

- If f is continuously differentiable almost everywhere, it is also Lipschitz continuous
- For functions over a bounded subset of $\Re^n$: f is continuous $\supseteq$ f is differentiable (almost everywhere) $=$ f is Lipschitz continuous[1] $\supseteq$ f is continuously differentiable $=$ ∇f is continuous $\supseteq$ ∇f is differentiable (almost everywhere) $=$ f is smooth

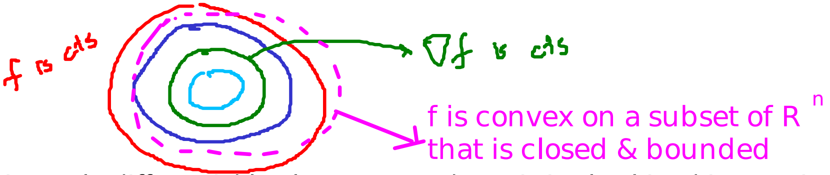
Eg: x

Eg: $e^x, x, f'(x) = |x|$



[1] Theorem 1.41 of https://moodle.iitb.ac.in/mod/resource/view.php?id=33825

Lipschitz Continuity on closed and bounded functions



- If f is continuously differentiable almost everywhere, it is also Lipschitz continuous
- For functions over a bounded subset of $\Re^n$: f is continuous $\supseteq$ f is differentiable (almost everywhere) $=$ f is Lipschitz continuous[1] $\supseteq$ f is continuously differentiable $=$ ∇f is continuous $\supseteq$ ∇f is differentiable (almost everywhere) $=$ f is smooth

[1] Theorem 1.41 of `https://moodle.iitb.ac.in/mod/resource/view.php?id=33825`

Lipschitz Continuity on closed and bounded functions

- If f is continuously differentiable almost everywhere, it is also Lipschitz continuous
- For functions over a bounded subset of $\Re^n$: f is continuous $\supseteq$ f is differentiable (almost everywhere) $=$ f is Lipschitz continuous[1] $\supseteq$ f is continuously differentiable $=$ ∇f is continuous $\supseteq$ ∇f is differentiable (almost everywhere) $=$ f is smooth
- Recall that a function is Lipschitz continuous if the norm of the (sub)gradient is bounded!

[1] Theorem 1.41 of `https://moodle.iitb.ac.in/mod/resource/view.php?id=33825`

Lipschitz Continuity on closed and bounded functions

- If f is continuously differentiable almost everywhere, it is also Lipschitz continuous
- For functions over a bounded subset of $\Re^n$: f is continuous $\supseteq$ f is differentiable (almost everywhere) $=$ f is Lipschitz continuous[1] $\supseteq$ f is continuously differentiable $=$ ∇f is continuous $\supseteq$ ∇f is differentiable (almost everywhere) $=$ f is smooth
- Recall that a function is Lipschitz continuous if the norm of the (sub)gradient is bounded!
- Also it holds that over a closed and bounded subset of $\Re^n$ that f is Lipschitz continuous $\supseteq$ f is convex

[1] Theorem 1.41 of https://moodle.iitb.ac.in/mod/resource/view.php?id=33825

More on Strong Convexity, Lipschitz Continuity and Smoothness

- Recall that a function f is strongly convex if there exists a $\mu > 0$ such that $f(x) - \mu/2||x||^2$ is convex.

More on Strong Convexity, Lipschitz Continuity and Smoothness

- Recall that a function f is strongly convex if there exists a $\mu > 0$ such that $f(x) - \mu/2||x||^2$ is convex.

Recall how we could construct a strongly convex function by adding a regularizer to increase its curvature

Q: Can we construct Lipschitz smooth function from a given function which may not be Lipschitz smooth?
Ans?: Subtract a regularizer to decrease curvature?

$$-g(x) + \frac{L}{2}||x||_2^2 \quad ?$$

Does this not require some assumption/property of g?

More on Strong Convexity, Lipschitz Continuity and Smoothness

- Recall that a function f is strongly convex if there exists a $\mu > 0$ such that $f(x) - \mu/2||x||^2$ is convex.
- Is there is a similar result for Lipschitz smooth functions?

More on Strong Convexity, Lipschitz Continuity and Smoothness

- Recall that a function f is strongly convex if there exists a $\mu > 0$ such that $f(x) - \mu/2||x||^2$ is convex.
- Is there is a similar result for Lipschitz smooth functions?

Ans: if g is convex we can do such a transformation
See how ubiquoutous convexity is!

More on Strong Convexity, Lipschitz Continuity and Smoothness

- Recall that a function f is strongly convex if there exists a $\mu > 0$ such that $f(x) - \mu/2||x||^2$ is convex.
- Is there is a similar result for Lipschitz smooth functions?
- A function f is Lipschitz smooth if there exists a $L > 0$ such that $L/2||x||^2 - f(x)$ is convex!

More on Strong Convexity, Lipschitz Continuity and Smoothness

- Recall that a function f is strongly convex if there exists a $\mu > 0$ such that $f(x) - \mu/2||x||^2$ is convex.
- Is there is a similar result for Lipschitz smooth functions?
- A function f is Lipschitz smooth if there exists a $L > 0$ such that $L/2||x||^2 - f(x)$ is convex!
- In fact there is an interesting duality between the two (more on that later).

More on Strong Convexity, Lipschitz Continuity and Smoothness

- Recall that a function f is strongly convex if there exists a $\mu > 0$ such that $f(x) - \mu/2||x||^2$ is convex.
- Is there is a similar result for Lipschitz smooth functions?
- A function f is Lipschitz smooth if there exists a $L > 0$ such that $L/2||x||^2 - f(x)$ is convex!
- In fact there is an interesting duality between the two (more on that later).
- If a Function f is strongly convex and g is convex, the function $f + g$ is strongly convex.

More on Strong Convexity, Lipschitz Continuity and Smoothness

- Recall that a function f is strongly convex if there exists a $\mu > 0$ such that $f(x) - \mu/2||x||^2$ is convex.
- Is there is a similar result for Lipschitz smooth functions?
- A function f is Lipschitz smooth if there exists a $L > 0$ such that $L/2||x||^2 - f(x)$ is convex!
- In fact there is an interesting duality between the two (more on that later).
- If a Function f is strongly convex and g is convex, the function $f + g$ is strongly convex.
- If a function f_1 is Lipschitz smooth and f_2 is Lipschitz smooth, then $f_1 + f_2$ is also Lipschitz smooth

Properties of ML Loss Functions

- Logistic Loss: $L(\theta) = \sum_{i=1}^{n} \log(1 + \exp(-y_i\theta^T x_i))$: Lipschitz Smooth (Proof sketch?)

Properties of ML Loss Functions

Proof sketch: Using composition

① $f_1(t) = \log(1+e^t)$ $\quad f_1'(t) = e^t/(1+e^t)$ $\quad f_1''(t) = \nabla^2 f(t) = \dfrac{e^t(1+e^t) - e^t \cdot e^t}{(1+e^t)^2}$

$$= \frac{e^t}{(1+e^t)^2} \overset{?}{\leq} L \quad \left(L = 1/2 \text{ or } 1/4\right)$$

- Logistic Loss: $L(\theta) = \sum_{i=1}^{n} \log(1 + \exp(-y_i\theta^T x_i))$: Lipschitz Smooth (Proof sketch?)

Can be shown to be also Lipschitz Continuous

Using Quotient Rule

$$\frac{d\left(\frac{a(t)}{b(t)}\right)}{dt} = \frac{a'(t)\,b(t) - b'(t)\,a(t)}{(b(t))^2}$$

② $f_2(\theta) = -y_i\theta^T x_i$ THe Hessian Is indeed upper bounded by an L = 0

$\sum_i f_1(f_2(\theta))$ would therefore be Lipschitz Smooth

In fact it is also convex

Properties of ML Loss Functions

- Logistic Loss: $L(\theta) = \sum_{i=1}^{n} \log(1 + \exp(-y_i\theta^T x_i))$: Lipschitz Smooth (Proof sketch?)
- Hinge Loss: $L(\theta) = \sum_{i=1}^{n} \max\{0, 1 - y_i\theta^T x_i\}$: Lipschitz Continuous

Properties of ML Loss Functions

Consider simpler function i) $\max(0, 1-t) = f(t)$

Case 1: $t_1 < 1$ & $t_2 \geq 1$ OR $t_1 \geq 1$ & $t_2 < 1$ → Can be done similarly

Case 2: $t_1 < 1$ & $t_2 < 1$

Case 3: $t_1 \geq 1$ & $t_2 \geq 1$

- Logistic Loss: $L(\theta) = \sum_{i=1}^{n} \log(1 + \exp(-y_i\theta^T x_i))$: Lipschitz Smooth (Proof sketch?)
- Hinge Loss: $L(\theta) = \sum_{i=1}^{n} \max\{0, 1 - \underbrace{y_i\theta^T x_i}_{t}\}$: Lipschitz Continuous

Wont expect smoothness (amounts to differentiability of its gradient almost everywhere) Case for Hinge Loss and L1 regularization

Case 1: $|f(t_1) - f(t_2)| = |1 - t_1 - 0| = |1 - t_1| \leq |t_2 - t_1|$ $(L = 1)$

Case 2: $|f(t_1) - f(t_2)| = |(1-t_1) - (1-t_2)| = |t_2 - t_1| \leq |t_2 - t_1|$ $(L = 1)$

Case 3: $|f(t_1) - f(t_2)| = 0 \leq |t_2 - t_1|$ $(L \leq 1)$

Subsequently, we consider a composition of the function f(.) above with a linear function exactly as in the previous case

Properties of ML Loss Functions

- Logistic Loss: $L(\theta) = \sum_{i=1}^{n} \log(1 + \exp(-y_i\theta^T x_i))$: Lipschitz Smooth (Proof sketch?)
- Hinge Loss: $L(\theta) = \sum_{i=1}^{n} \max\{0, 1 - y_i\theta^T x_i\}$: Lipschitz Continuous
- Multi-class Logistic Regression: $L(\theta_1, \cdots, \theta_k) = \sum_{i=1}^{n} -\theta_{y_i}^T x_i + \log(\sum_{c=1}^{k} \exp(\theta_c^T x_i))\}$: Lipschitz Smooth

Properties of ML Loss Functions

- Logistic Loss: $L(\theta) = \sum_{i=1}^{n} \log(1 + \exp(-y_i\theta^T x_i))$: Lipschitz Smooth (Proof sketch?)
- Hinge Loss: $L(\theta) = \sum_{i=1}^{n} \max\{0, 1 - y_i\theta^T x_i\}$: Lipschitz Continuous
- Multi-class Logistic Regression: $L(\theta_1, \cdots, \theta_k) = \sum_{i=1}^{n} -\theta_{y_i}^T x_i + \log(\sum_{c=1}^{k} \exp(\theta_c^T x_i))\}$: Lipschitz Smooth

  Proof Methodology: Generalization of the Logistic Loss case

Properties of ML Loss Functions

- Logistic Loss: $L(\theta) = \sum_{i=1}^{n} \log(1 + \exp(-y_i \theta^T x_i))$: Lipschitz Smooth (Proof sketch?)
- Hinge Loss: $L(\theta) = \sum_{i=1}^{n} \max\{0, 1 - y_i \theta^T x_i\}$: Lipschitz Continuous
- Multi-class Logistic Regression: $L(\theta_1, \cdots, \theta_k) = \sum_{i=1}^{n} -\theta_{y_i}^T x_i + \log(\sum_{c=1}^{k} \exp(\theta_c^T x_i))\}$: Lipschitz Smooth
- Least Squares: $L(\theta) = \sum_{i=1}^{n} (\theta^T x_i - y_i)^2$: Lipschitz Smooth and Strongly Convex!

Properties of ML Loss Functions

The first three are also convex

- Logistic Loss: $L(\theta) = \sum_{i=1}^{n} \log(1 + \exp(-y_i \theta^T x_i))$: Lipschitz Smooth (Proof sketch?)
- Hinge Loss: $L(\theta) = \sum_{i=1}^{n} \max\{0, 1 - y_i \theta^T x_i\}$: Lipschitz Continuous
- Multi-class Logistic Regression: $L(\theta_1, \cdots, \theta_k) = \sum_{i=1}^{n} -\theta_{y_i}^T x_i + \log(\sum_{c=1}^{k} \exp(\theta_c^T x_i))\}$: Lipschitz Smooth
- Least Squares: $L(\theta) = \sum_{i=1}^{n} \underbrace{(\theta^T x_i - y_i)^2}_{f(t) = t^2}$: Lipschitz Smooth and Strongly Convex!

Recap this is both Lipschitz Smooth and Strongly Convex

In a bounded domain, we also expect Least Squares to be Lipschitz continuous

Properties of ML Loss Functions

- Logistic Loss: $L(\theta) = \sum_{i=1}^{n} \log(1 + \exp(-y_i \theta^T x_i))$: Lipschitz Smooth (Proof sketch?)
- Hinge Loss: $L(\theta) = \sum_{i=1}^{n} \max\{0, 1 - y_i \theta^T x_i\}$: Lipschitz Continuous
- Multi-class Logistic Regression: $L(\theta_1, \cdots, \theta_k) = \sum_{i=1}^{n} -\theta_{y_i}^T x_i + \log(\sum_{c=1}^{k} \exp(\theta_c^T x_i))\}$: Lipschitz Smooth
- Least Squares: $L(\theta) = \sum_{i=1}^{n} (\theta^T x_i - y_i)^2$: Lipschitz Smooth and Strongly Convex!
- L2 Regularization: Lipschitz Smooth and Strongly Convex!

Properties of ML Loss Functions

- Logistic Loss: $L(\theta) = \sum_{i=1}^{n} \log(1 + \exp(-y_i \theta^T x_i))$: Lipschitz Smooth (Proof sketch?)
- Hinge Loss: $L(\theta) = \sum_{i=1}^{n} \max\{0, 1 - y_i \theta^T x_i\}$: Lipschitz Continuous
- Multi-class Logistic Regression: $L(\theta_1, \cdots, \theta_k) = \sum_{i=1}^{n} -\theta_{y_i}^T x_i + \log(\sum_{c=1}^{k} \exp(\theta_c^T x_i))\}$: Lipschitz Smooth
- Least Squares: $L(\theta) = \sum_{i=1}^{n} (\theta^T x_i - y_i)^2$: Lipschitz Smooth and Strongly Convex!
- L2 Regularization: Lipschitz Smooth and Strongly Convex!

same reasoning

$$\|x\|_2^2 = f(\|x\|_2)$$

where $\underbrace{f(t) = t^2}$

Properties of ML Loss Functions

- Logistic Loss: $L(\theta) = \sum_{i=1}^{n} \log(1 + \exp(-y_i\theta^T x_i))$: Lipschitz Smooth (Proof sketch?)
- Hinge Loss: $L(\theta) = \sum_{i=1}^{n} \max\{0, 1 - y_i\theta^T x_i\}$: Lipschitz Continuous
- Multi-class Logistic Regression: $L(\theta_1, \cdots, \theta_k) = \sum_{i=1}^{n} -\theta_{y_i}^T x_i + \log(\sum_{c=1}^{k} \exp(\theta_c^T x_i))\}$: Lipschitz Smooth
- Least Squares: $L(\theta) = \sum_{i=1}^{n} (\theta^T x_i - y_i)^2$: Lipschitz Smooth and Strongly Convex!
- L2 Regularization: Lipschitz Smooth and Strongly Convex!
- L1 Regularization: Lipschitz Continuous

Properties of ML Loss Functions

- Logistic Loss: $L(\theta) = \sum_{i=1}^{n} \log(1 + \exp(-y_i\theta^T x_i))$: Lipschitz Smooth (Proof sketch?)
- Hinge Loss: $L(\theta) = \sum_{i=1}^{n} \max\{0, 1 - y_i\theta^T x_i\}$: Lipschitz Continuous
- Multi-class Logistic Regression: $L(\theta_1, \cdots, \theta_k) = \sum_{i=1}^{n} -\theta_{y_i}^T x_i + \log(\sum_{c=1}^{k} \exp(\theta_c^T x_i))\}$: Lipschitz Smooth
- Least Squares: $L(\theta) = \sum_{i=1}^{n} (\theta^T x_i - y_i)^2$: Lipschitz Smooth and Strongly Convex!
- L2 Regularization: Lipschitz Smooth and Strongly Convex!
- L1 Regularization: Lipschitz Continuous

Recap: We showed that |x| is Lipschitz continuous

Properties of ML Loss Functions with Regularizers

- L1 Regularized Logistic Loss: Lipschitz Continuous and Convex

Properties of ML Loss Functions with Regularizers

Recap: Sum of Lipschitz continuous functions is Lipschitz continuous
Recap: Sum of convex functions is convex

- L1 Regularized Logistic Loss: Lipschitz Continuous and Convex

Properties of ML Loss Functions with Regularizers

- L1 Regularized Logistic Loss: Lipschitz Continuous and Convex
- L2 Regularized Logistic Loss: Lipschitz Smooth and Strongly Convex!

Properties of ML Loss Functions with Regularizers

- L1 Regularized Logistic Loss: Lipschitz Continuous and Convex
- L2 Regularized Logistic Loss: Lipschitz Smooth and Strongly Convex!
- L1 Regularized Hinge Loss: Lipschitz Continuous and Convex

Properties of ML Loss Functions with Regularizers

- L1 Regularized Logistic Loss: Lipschitz Continuous and Convex
- L2 Regularized Logistic Loss: Lipschitz Smooth and Strongly Convex!
- L1 Regularized Hinge Loss: Lipschitz Continuous and Convex
- L2 Regularized Hinge Loss: Lipschitz Continuous and Strongly Convex (On a Bounded set)

Properties of ML Loss Functions with Regularizers

- L1 Regularized Logistic Loss: Lipschitz Continuous and Convex
- L2 Regularized Logistic Loss: Lipschitz Smooth and Strongly Convex!
- L1 Regularized Hinge Loss: Lipschitz Continuous and Convex
- L2 Regularized Hinge Loss: Lipschitz Continuous and Strongly Convex (On a Bounded set)
- L2 Regularized Least Squares (Lasso): Lipschitz Smooth and Strongly Convex!

Properties of ML Loss Functions with Regularizers

- L1 Regularized Logistic Loss: Lipschitz Continuous and Convex
- L2 Regularized Logistic Loss: Lipschitz Smooth and Strongly Convex!
- L1 Regularized Hinge Loss: Lipschitz Continuous and Convex
- L2 Regularized Hinge Loss: Lipschitz Continuous and Strongly Convex (On a Bounded set)
- L2 Regularized Least Squares (Lasso): Lipschitz Smooth and Strongly Convex!
- L1 Regularized Least Squares: Lipschitz Continuous and Strongly Convex (On a bounded set)