The University of Texas at Austin
Department of Electrical and Computer Engineering

ECE 381V: Large-Scale Optimization II — Spring 2022

Lecture 3

Caramanis & Mokhtari Wednesday, January 26, 2022

Goal: In this lecture, we talk about different class of differentiable functions and their properties.

1 Classes of Differentiable Functions

Suppose $\mathcal{Q} \subseteq \mathbb{R}^n$. We introduce the notation $\mathcal{C}_L^{k,p}(\mathcal{Q}, \|.\|)$. We state that $f \in \mathcal{C}_L^{k,p}(\mathcal{Q}, \|.\|)$ if

1. f is k times continuously differentiable on $\mathcal{Q}$.
2. Its p-th derivative is Lipschitz continuous on $\mathcal{Q}$ with parameter L and with respect to the norm $\|.\|$
$$\|\nabla^p f(\mathbf{x}) - \nabla^p f(\mathbf{y})\|_* \ \leq \ L\|\mathbf{x}-\mathbf{y}\|, \qquad \text{for all } \mathbf{x}, \mathbf{y} \in \mathcal{Q}$$

Remark 1. *Note that smoothness or Lipschitz continuity of gradients or any derivative of the function can be defined/measured with respect to an arbitrary norm $\|.\|$. In this case, the difference between derivatives should be measured in the dual norm. Note that the dual norm $\|.\|_*$ of an arbitrary norm $\|.\|$ is defined as*

$$\|\mathbf{g}\|_* = \max_{\mathbf{x}\in\mathbb{R}^n} \{\mathbf{g}^\top \mathbf{x} \ \mid \ \|\mathbf{x}\| \leq 1\}$$

When we don't state the norm, we are using the standard Euclidean norm.

Note: In this definition, we don't assume convexity and the functions could be possibly nonconvex.

1.1 Functions with Lipschitz Continuous Gradients $\mathcal{C}_L^{1,1}(\mathbb{R}^n)$

An important class of differentiable functions are those that have Lipschitz continuous gradient for all points in $\mathbb{R}^n$ which we denote this class by $\mathcal{C}_L^{1,1}(\mathbb{R}^n)$. Indeed for this class we have

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \ \leq \ L\|\mathbf{x}-\mathbf{y}\|, \qquad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$

Lemma 1. *If $f \in \mathcal{C}_L^{1,1}(\mathbb{R}^n)$, then* for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *we have*

$$-\frac{L}{2}\|\mathbf{y}-\mathbf{x}\|^2 \leq f(\mathbf{y}) - f(\mathbf{x}) - \nabla f(\mathbf{x})^\top(\mathbf{y}-\mathbf{x}) \leq \frac{L}{2}\|\mathbf{y}-\mathbf{x}\|^2 \tag{1}$$

Proof. We know that

$$\begin{aligned}
f(\mathbf{y}) &= f(\mathbf{x}) + \int_0^1 \nabla f(s\mathbf{x} + (1-s)\mathbf{y})^\top(\mathbf{y}-\mathbf{x})ds \\
&= f(\mathbf{x}) + \nabla f(\mathbf{x})^\top(\mathbf{y}-\mathbf{x}) + \int_0^1 \big(\nabla f(s\mathbf{x} + (1-s)\mathbf{y}) - \nabla f(\mathbf{x})\big)^\top(\mathbf{y}-\mathbf{x})ds
\end{aligned}$$

Hence,

$$\begin{aligned}
|f(\mathbf{y}) - f(\mathbf{x}) - \nabla f(\mathbf{x})^\top (\mathbf{y}-\mathbf{x})| &\le \left| \int_0^1 \left(\nabla f(s\mathbf{x} + (1-s)\mathbf{y}) - \nabla f(\mathbf{x})\right)^\top (\mathbf{y}-\mathbf{x}) ds \right| \\
&\le \int_0^1 \left\| \left(\nabla f(s\mathbf{x} + (1-s)\mathbf{y}) - \nabla f(\mathbf{x})\right)^\top (\mathbf{y}-\mathbf{x}) \right\| ds \\
&\le \int_0^1 \left\| \nabla f(s\mathbf{x} + (1-s)\mathbf{y}) - \nabla f(\mathbf{x})) \right\| \|\mathbf{y}-\mathbf{x}\| ds \\
&\le \|\mathbf{y}-\mathbf{x}\| \int_0^1 L(1-s)\|\mathbf{y}-\mathbf{x}\| ds \\
&= \frac{L}{2}\|\mathbf{y}-\mathbf{x}\|^2,
\end{aligned}$$

and the proof is complete. □

Geometric interpretation of the above lemma. The above inequalities show that the function f at any point $\mathbf{x}$ can be bounded below and above by a quadratic function. To be more precise, consider a given point $\mathbf{x}_0$ and for which define the following quadratic functions

$$\phi_1(\mathbf{x};\mathbf{x}_0) := f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^\top (\mathbf{x}-\mathbf{x}_0) - \frac{L}{2}\|\mathbf{x}-\mathbf{x}_0\|^2$$

and

$$\phi_2(\mathbf{x};\mathbf{x}_0) := f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^\top (\mathbf{x}-\mathbf{x}_0) + \frac{L}{2}\|\mathbf{x}-\mathbf{x}_0\|^2$$

Based on the above lemma, for any $\mathbf{x}$ we have

$$\phi_1(\mathbf{x};\mathbf{x}_0) \le f(\mathbf{x}) \le \phi_2(\mathbf{x};\mathbf{x}_0)$$

Hence, even if a function is nonconvex, but has Lipschitz gradients it can be bounded below and above by quadratic functions. Note that the lower bound is concave, while the upper bound is convex.

Before stating the following theorem, we should note that $\mathcal{C}_L^{2,1}(\mathbb{R}^n) \subset \mathcal{C}_L^{1,1}(\mathbb{R}^n)$

Theorem 1. *A function f belongs to the class $\mathcal{C}_L^{2,1}(\mathbb{R}^n)$ if and only if $\|\nabla^2 f(\mathbf{x})\| \le L$,* for all $\mathbf{x} \in \mathbb{R}^n$.

Note: $\|\nabla^2 f(\mathbf{x})\| \le L \;\equiv\; -L\mathbf{I} \preceq \nabla^2 f(\mathbf{x}) \preceq L\mathbf{I}$.

Proof. By the mean value theorem we have

$$\nabla f(\mathbf{y}) = \nabla f(\mathbf{x}) + \int_0^1 \nabla^2 f(\tau\mathbf{y} + (1-\tau)\mathbf{x}) \; (\mathbf{y}-\mathbf{x}) \; d\tau$$

Hence, we have

$$\begin{aligned}
\|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\| &= \left\| \int_0^1 \nabla^2 f(\tau\mathbf{y} + (1-\tau)\mathbf{x}) \; (\mathbf{y}-\mathbf{x}) \; d\tau \right\| \\
&\le \left\| \int_0^1 \nabla^2 f(\tau\mathbf{y} + (1-\tau)\mathbf{x}) d\tau \right\| \|\mathbf{y}-\mathbf{x}\| \\
&\le \left(\int_0^1 \left\| \nabla^2 f(\tau\mathbf{y} + (1-\tau)\mathbf{x}) \right\| d\tau \right) \|\mathbf{y}-\mathbf{x}\| \\
&\le L\|\mathbf{y}-\mathbf{x}\|
\end{aligned}$$

For the other side, if f belongs to the class $\mathcal{C}_L^{2,1}(\mathbb{R}^n)$ then for any $\mathbf{s} \in \mathbb{R}^n$ and $\alpha > 0$ we have

$$\left\| \left(\int_0^\alpha \nabla^2 f(\mathbf{x} + \tau \mathbf{s}) \; d\tau \right) \mathbf{s} \right\| = \|\nabla f(\mathbf{x} + \alpha \mathbf{s}) - \nabla f(\mathbf{x})\| \leq L\|\mathbf{s}\|$$

Now by regrouping the terms and sending α to zero we have

$$\|\nabla^2 f(\mathbf{x})\| = \lim_{\alpha \to 0} \frac{\left\| \left(\int_0^\alpha \nabla^2 f(\mathbf{x} + \tau \mathbf{s}) \; d\tau \right) \mathbf{s} \right\|}{\|\mathbf{s}\|} \leq L.$$

□

Lemma 2. *Suppose $f \in \mathcal{C}_L^{2,1}(\mathbb{R}^n)$. Then, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\alpha \in [0, 1]$ we have*

$$f(\alpha \mathbf{x} + (1 - \alpha)\mathbf{y}) \geq \alpha f(\mathbf{x}) + (1 - \alpha) f(\mathbf{y}) - \frac{\alpha(1 - \alpha)L}{2}\|\mathbf{x} - \mathbf{y}\|^2$$

and

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

Proof. Exercise. □

Lemma 3. *Suppose $f \in \mathcal{C}_L^{2,1}(\mathbb{R}^n)$. Then, for any $\mathbf{x} \in \mathbb{R}^n$ we have*

$$f(\mathbf{x}) - f^* \geq \frac{1}{2L}\|\nabla f(\mathbf{x})\|_2^2,$$

where f^* is the minimum value of f on $\mathbb{R}^n$.

Proof. We have

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

We can minimize both sides of the following inequality with respect to $\mathbf{y}$ (the right hand side is a convex quadratic problem with respect to $\mathbf{y}$)

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

which implies that the optimal solution for the right hand side $\tilde{\mathbf{y}}$ is $\tilde{\mathbf{y}} = \mathbf{x} - \frac{1}{L}\nabla f(\mathbf{x})$. The optimal value of the left hand side is f^* and therefore we have

$$f^* \leq f(\mathbf{x}) - \frac{1}{2L}\|\nabla f(\mathbf{x})\|_2^2,$$

and the claim follows. □

The above result provides an upper bound for the optimal value of the objective function.

1.2 Functions with Lipschitz Continuous Hessian $\mathcal{C}_M^{2,2}(\mathbb{R}^n)$

Another important class of differentiable functions are those that are twice differentiable and have Lipschitz continuous Hessian for all points in $\mathbb{R}^n$ which we denote this class by $\mathcal{C}_M^{2,2}(\mathbb{R}^n)$. For this class we have

$$\|\nabla^2 f(\mathbf{y}) - \nabla^2 f(\mathbf{x})\| \;\leq\; M\|\mathbf{y}-\mathbf{x}\|, \qquad \text{for all } \mathbf{x},\mathbf{y} \in \mathbb{R}^n$$

Lemma 4. *If $f \in \mathcal{C}_M^{2,2}(\mathbb{R}^n)$, then* for all $\mathbf{x},\mathbf{y} \in \mathbb{R}^n$ *we have*

$$\left| f(\mathbf{y}) - f(\mathbf{x}) - \nabla f(\mathbf{x})^\top(\mathbf{y}-\mathbf{x}) - \frac{1}{2}(\mathbf{y}-\mathbf{x})^\top \nabla^2 f(\mathbf{x})(\mathbf{y}-\mathbf{x}) \right| \leq \frac{M}{6}\|\mathbf{y}-\mathbf{x}\|^3 \tag{2}$$

and

$$\|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x}) - \nabla^2 f(\mathbf{x})^\top(\mathbf{y}-\mathbf{x})\| \leq \frac{M}{2}\|\mathbf{y}-\mathbf{x}\|^2 \tag{3}$$

Proof. Similar to the proof of Lemma 1. For more details check Lemma 1.2.4 in the textbook. □

The first result shows that if a function belongs to $\mathcal{C}_M^{2,2}(\mathbb{R}^n)$ then it can be upper and lower bounded by a cubic function. The second result shows that its gradient at any point $\nabla f(\mathbf{y})$ can be approximated by the function $\nabla f(\mathbf{x}) + \nabla^2 f(\mathbf{x})^\top(\mathbf{y}-\mathbf{x})$ and the error would be $\frac{M}{2}\|\mathbf{y}-\mathbf{x}\|^2$.

Lemma 5. *If $f \in \mathcal{C}_M^{2,2}(\mathbb{R}^n)$, then*

$$\nabla^2 f(\mathbf{x}) - M\|\mathbf{y}-\mathbf{x}\|\mathbf{I} \;\preceq\; \nabla^2 f(\mathbf{y}) \;\preceq\; \nabla^2 f(\mathbf{x}) + M\|\mathbf{y}-\mathbf{x}\|\mathbf{I}$$

Proof. Note that based on the main definition we have

$$-M\|\mathbf{y}-\mathbf{x}\|\mathbf{I} \preceq \nabla^2 f(\mathbf{y}) - \nabla^2 f(\mathbf{x}) \preceq M\|\mathbf{y}-\mathbf{x}\|\mathbf{I}$$

and by regrouping the terms the claim follows. □

The above bound provides nice upper and lower bounds for the eigenvalues of the Hessian at a point $\mathbf{y}$ in terms of the eigenvalues of the Hessian at another point $\mathbf{x}$.

2 Classes of Convex Differentiable Functions

Now we proceed to a more specific class of differentiable functions that are also convex. We use the notation of $\mathcal{F}_L^{k,p}(\mathcal{Q}, \|.\|)$ for this class of functions. More precisely, suppose $\mathcal{Q} \subseteq \mathbb{R}^n$ and $\|.\|$ is an arbitrary norm. We state that $f \in \mathcal{F}_L^{k,p}(\mathcal{Q}, \|.\|)$ if

1. f is convex on $\mathcal{Q}$.
2. f is k times continuously differentiable on $\mathcal{Q}$.
3. Its p-th derivative is Lipschitz continuous on $\mathcal{Q}$ with parameter L and with respect to $\|.\|$, i.e.,

$$\|\nabla^p f(\mathbf{x}) - \nabla^p f(\mathbf{y})\|_* \;\leq\; L\|\mathbf{x}-\mathbf{y}\|, \qquad \text{for all } \mathbf{x},\mathbf{y} \in \mathcal{Q}$$

When we don't identify the norm, it means we are focusing on the Euclidean norm.

2.1 Convex Functions with Lipschitz Continuous Gradients $\mathcal{F}_L^{1,1}(\mathbb{R}^n, \|.\|)$

For this class of functions we have, f is convex and smooth, i.e.,

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_* \;\leq\; L\|\mathbf{x} - \mathbf{y}\|, \qquad \text{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

Theorem 2. *The statement that $f \in \mathcal{F}_L^{1,1}(\mathbb{R}^n)$ (it is convex and smooth with constant L) is equivalent to any of the following statements.*

1. $0 \leq f(\mathbf{y}) - f(\mathbf{x}) - \nabla f(\mathbf{x})^\top(\mathbf{y} - \mathbf{x}) \leq \frac{L}{2}\|\mathbf{y} - \mathbf{x}\|^2$
2. $f(\mathbf{y}) - f(\mathbf{x}) - \nabla f(\mathbf{x})^\top(\mathbf{y} - \mathbf{x}) \geq \frac{1}{2L}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_*^2$
3. $\frac{1}{2L}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_*^2 \leq (\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})^\top(\mathbf{x} - \mathbf{y})$
4. $0 \leq (\nabla f(\mathbf{x}) - \nabla f(\mathbf{y}))^\top(\mathbf{x} - \mathbf{y}) \leq L\|\mathbf{x} - \mathbf{y}\|^2$
5. $f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1-\alpha)f(\mathbf{y}) - \frac{\alpha(1-\alpha)L}{2}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|^2 \qquad \alpha \in [0,1]$
6. $0 \leq \alpha f(\mathbf{x}) + (1-\alpha)f(\mathbf{y}) - f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \leq \frac{\alpha(1-\alpha)L}{2}\|\mathbf{x} - \mathbf{y}\|^2 \qquad \alpha \in [0,1]$

Proof. See Theorem 2.1.5 in the text book. □