CPT Lecture Notes 3: Differentiable Functions

Levent Ülkü

August 29, 2019

Reminders: For any $n \in \mathbb{N}$, any $x, y \in \mathbb{R}^n$ and any $\varepsilon > 0$

$$\begin{aligned} d(x,y) &= \left[\sum_{i=1}^{n}(x_i - y_i)^2\right]^{1/2} \text{ and} \\ B(x,\varepsilon) &= \{z \in \mathbb{R}^n : d(x,z) < \varepsilon\} \end{aligned}$$

Let $S \subseteq \mathbb{R}^n$. S is open if for every $x \in S$ there is some $B(x,\varepsilon) \subset S$. S is closed if its complement is open. S is bounded if there exist numbers a_i, b_i, $i = 1, ..., n$, such that $S \subset \times_{i=1}^{n}[a_i, b_i]$. S is compact if and only if it is closed and bounded. S is convex if $tx + (1-t)x' \in S$ whenever $t \in [0,1]$ and $x, x' \in S$. Let $f : S \to \mathbb{R}^m$ and $x \in S$. Then f is continuous at x if for every $\varepsilon > 0$ there is some $\delta > 0$ such that $f(x') \in B(f(x), \varepsilon)$ whenever $x' \in B(x,\delta) \cap S$. We say that f is continuous if it is continuous at each $x \in S$.

Let $S \subseteq \mathbb{R}$, $x \in S$ and $f : S \to \mathbb{R}$. Then f is differentiable at x if

$$\lim_{u \to 0} \frac{f(x+u) - f(x)}{u}$$

exists. In this case the derivative of f at x is equal to this limit and it is denoted $f'(x)$.

In other words, f is differentiable at x if there is some $y \in \mathbb{R}$ such that for every $\varepsilon > 0$ there exists some $\delta > 0$ such that

$$\{x + u \in B_\delta(x) \cap S \textit{ and } u \neq 0\} \Rightarrow \left| \frac{f(x+u) - f(x)}{u} - y \right| < \varepsilon$$

and in this case $y = f'(x)$.

Let $S \subseteq \mathbb{R}^n$, $x \in S$ and $f : S \to \mathbb{R}$. The jth partial derivative of f at x is

$$D_j f(x) := \lim_{u \to 0} \frac{f(x + ue_j) - f(x)}{u}$$

if this limit exists.

If $D_j f(x)$ exists for all $j = 1, ..., n$, then the gradient of f at x is the vector

$$\nabla f(x) := \begin{bmatrix} D_1 f(x) \\ \vdots \\ D_n f(x) \end{bmatrix} \in \mathbb{R}^n.$$

If $S \subseteq \mathbb{R}^n$ and $f : S \to \mathbb{R}^m$ then for each $x \in S$ we will write

$$f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix} \in \mathbb{R}^m.$$

Let $S \subseteq \mathbb{R}^n$, $x \in S$ and $f : S \to \mathbb{R}^m$. Suppose that $D_i f_j(x)$ exists for all i and j, then the Jacobian of f at x is the $m \times n$ matrix

$$J_f(x) := \begin{bmatrix} D_1 f_1(x) & \cdots & D_n f_1(x) \\ \vdots & & \vdots \\ D_1 f_m(x) & \cdots & D_n f_m(x) \end{bmatrix}_{m \times n} = \begin{bmatrix} \nabla f_1(x)^T \\ \vdots \\ \nabla f_m(x)^T \end{bmatrix}$$

Now we will define differentiability of a function $f : S \to \mathbb{R}^m$ where $S \subseteq \mathbb{R}^n$. We must make sure that our definition overlaps with the earlier definition of differentiability when $m = n = 1$. For any vector x, $\|x\|$ is the *norm* of x, i.e., the distance of x to 0.

Let $S \subseteq \mathbb{R}^n$. Then $f : S \to \mathbb{R}^m$ is differentiable at $x \in S$ if the following two conditions hold:

1. $D_j f_i(x)$ exists for all i and j,
2. we have
$$\lim_{u \to 0} \frac{\|f(x+u) - f(x) - J_f(x)u\|}{\|u\|} = 0.$$

The second condition in the definition holds if and only if for every $\varepsilon > 0$ there is some $\delta > 0$ such that whenever $x + u \in B_\delta(x) \cap S$ and $u \neq 0_{n\times 1}$ we have

$$\frac{\|f(x+u) - f(x) - J_f(x)u\|}{\|u\|} < \varepsilon.$$

When $m = 1 = n$, the first condition implies the second, but in general this is not true.

Example: Let $S = \mathbb{R}^2$, $f : S \to \mathbb{R}$ and

$$f(x_1, x_2) = \begin{cases} x_1^3/(x_1^2 + x_2^2) & \text{if } (x_1, x_2) \neq (0,0), \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to compute $D_1 f(0,0) = 1$ and $D_2 f(0,0) = 0$. So the first condition in the definition of differentiability holds at $(0,0)$. Let us look at the second condition. At $x = (0,0)$ for any $u = (u_1, u_2) \neq (0,0)$ we have

$$\frac{\|f(x+u) - f(x) - J_f(x)u\|}{\|u\|} = \frac{\left|\frac{u_1^3}{u_1^2 + u_2^2} - u_1\right|}{(u_1^2 + u_2^2)^{1/2}}$$

and letting $u = (1/k, 1/k)$ we see that this expression does not converge to zero. So the second condition fails and f is not differentiable at $(0,0)$ even though its partials are well defined.

In the previous example $D_1 f(x)$ is not continuous at $x = (0, 0)$.

Theorem: Suppose $S \subseteq \mathbb{R}^n$, $x \in S$ and $f : S \to \mathbb{R}^m$. Suppose that for each i and j, $D_j f_i(x)$ exists and that $x \mapsto D_j f_i(x)$ is continuous on S. Then f is differentiable at x.

Theorem: (Implicit Function Theorem) Let $F : \mathbb{R}^2 \to \mathbb{R}$. Suppose that $D_1 F(x, y)$ and $D_2 F(x, y)$ exist and are continuous in an open U containing $(a, b) \in \mathbb{R}^2$. Suppose that $F(a, b) = 0$ and that $D_2 F(a, b) \neq 0$. Then there exists $\varepsilon > 0$ and $g : B_\varepsilon(a) \to \mathbb{R}$ such that $g(a) = b$, $F(x, g(x)) = 0$ for all $x \in B_\varepsilon(a)$ and

$$g'(a) = \frac{-D_1 F(a, b)}{D_2 F(a, b)}.$$

Example: Let $F(x, y) = x^2 + y^2 - 2$. Then $(a, b) = (1, 1)$ is a vector for which the result works, but $(2^{1/2}, 0)$ is not.

In higher dimensions we would like to come up, at least locally, with the map

$$(x_1, ..., x_n) \mapsto \begin{bmatrix} g_1(x_1, ..., x_n) \\ \vdots \\ g_m((x_1, ..., x_n) \end{bmatrix} \in \mathbb{R}^m$$

when we have a system of equations of the form

$$\begin{aligned} F_1(x_1, ..., x_n, y_1, ..., y_m) &= 0 \\ &\vdots \\ F_m(x_1, ..., x_n, y_1, ..., y_m) &= 0 \end{aligned}$$

and the Implicit Function addresses this problem.

Theorem: (Implicit Function Theorem) Suppose $F : \mathbb{R}^{n+m} \rightarrow \mathbb{R}^m$ is differentiable in an open set U containing $(a, b) \in \mathbb{R}^{n+m}$ and suppose that the entries of $J_F(x, y)$ are continuous at each $(x, y) \in U$. Let $A_{m\times n}$ and $B_{m\times m}$ be defined by $J_F(a, b)$ as follows:

$$
\begin{aligned}
A &= \begin{bmatrix} D_1F_1(a,b) & \cdots & D_nF_1(a,b) \\ \vdots & & \vdots \\ D_1F_m(a,b) & \cdots & D_nF_m(a,b) \end{bmatrix} \\
B &= \begin{bmatrix} D_{n+1}F_1(a,b) & \cdots & D_{n+m}F_1(a,b) \\ \vdots & & \vdots \\ D_{n+1}F_m(a,b) & \cdots & D_{n+m}F_m(a,b) \end{bmatrix}
\end{aligned}
$$

and suppose that $F(a, b) = 0$ and B is nonsingular. Then there exists $\varepsilon > 0$ and $g : B_\varepsilon(a) \rightarrow \mathbb{R}^m$ such that $g(a) = b$, $F(x, g(x)) = 0$ for all $x \in B_\varepsilon(a)$ and $J_g(a) = -B^{-1}_{m\times m}A_{m\times n}$.

Theorem: (Inverse Function Theorem) Suppose that $f : \mathbb{R}^m \to \mathbb{R}^m$ is differentiable in an open set U containing $b \in \mathbb{R}^m$. Suppose that the partials of f are continuos on U. Suppose that $f(b) = a$ and $J_f(b)$ is nonsingular. Then there exists some $\varepsilon > 0$ and some $g : B_\varepsilon(a) \to \mathbb{R}^m$ such that $g(a) = b$, $f(g(x)) = x$ for all $x \in B_\varepsilon(a)$ and g is differentiable at a with $J_g(a) = J_f(b)^{-1}$.

Proof: Define $F : \mathbb{R}^{2n} \to \mathbb{R}^n$ as $F(x, y) = x - f(y)$ and note that F is differentiable with continuos partials at each $(x, y) \in \mathbb{R}^n \times U$ and that

$$J_F(x, y) = [I \vdots - J_f(y)].$$

Now use the Implicit Function Theorem with $A = I$. ∎

Theorem: (Weierstrass Theorem) Let $S \subseteq \mathbb{R}^n$ be compact and let $f : S \rightarrow \mathbb{R}$ be continuous. Then there exists x' and x'' in S such that for all $x \in S$, $f(x') \leq f(x) \leq f(x'')$.

Theorem: Suppose $f : \mathbb{R}^n \to \mathbb{R}$ attains its maximum or minimum on an open set U at some $x \in U$. If $D_i f(x)$ exists, then $D_i f(x) = 0$.

Proof: (Sketch for $n = 1$) Suppose that f attains a maximum at x and that f is differentiable at x. Pick $\{u_k\}$ be a sequence such that $u_k > 0$ and $x \pm u_k \in U$ for each k and $u_k \to 0$. Note that such a sequence exists since U is open and $x \in U$. Now for each k we have

$$\frac{f(x+u_k) - f(x)}{u_k} \leq 0 \leq \frac{f(x-u_k) - f(x)}{-u_k}$$

and therefore letting $k \to \infty$ we get $f'(x) = 0$. Similarly if x is a minimizer. $\blacksquare$

Theorem: (Rolle's Theorem) Suppose $f : [a, b] \to \mathbb{R}$ is continuous at each $x \in [a, b]$ and that f is differentiable at each $x \in (a, b)$. Furthermore suppose that $f(a) = 0 = f(b)$. Then there exists some $c \in (a, b)$ such that $f'(c) = 0$.

Proof: Applying Weierstrass Theorem, we know that f attains its maximum and minimum on $[a, b]$. Let x' be the minimizer and x'' be the maximizer on $[a, b]$. There are three cases to consider. Case 1: $f(x') = f(x'') = 0$. Then let $c = \frac{a+b}{2}$. Case 2: $f(x'') \neq 0$. Then $x'' \in (a, b)$ which is open and by the previous result $f'(x'') = 0$. Case 3: $f(x') \neq 0$. Then $x' \in (a, b)$ which is open and by the previous result $f'(x') = 0$. ■

Theorem: (Mean Value Theorem) Suppose that $f : [a, b] \to \mathbb{R}$ is continuous at each $x \in [a, b]$ and that f is differentiable at each $x \in (a, b)$. Then there exists $c \in (a, b)$ such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

Proof: Let $g : [a, b] \to \mathbb{R}$ be given by

$$g(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a).$$

Then the conditions of Rolle's theorem are satisfied and $g'(c) = 0$ for some $c \in (a, b)$. Then $f'(c) = \frac{f(b)-f(a)}{b-a}$ and the proof is complete. ■

Theorem: (Mean Value Theorem) Let U be an open set in $\mathbb{R}^n$ containing a and b. Suppose that $(1-t)a + tb \in U$ whenever $t \in [0,1]$. Suppose that $f : U \to \mathbb{R}$ is differentiable at each $x \in U$. Then there exists $t' \in (0,1)$ such that $\nabla f(c') \cdot (b-a) = f(b) - f(a)$ where $c' = (1-t')a + t'b$.

Proof: Define $g : [0,1] \to \mathbb{R}$ by $g(t) = f((1-t)a + tb)$. Then g satisfies the conditions of the one dimensional Mean Value Theorem above. Furthermore $g(0) = f(a)$ and $g(1) = f(b)$ and there exists t' such that $g'(t') = f(b) - f(a)$. Since

$$\begin{aligned} g'(t') &= \sum_{i=1}^{n} D_i f((1-t')a + tb)(b_i - a_i) \\ &= \nabla f((1-t')a + tb) \cdot (b-a) \end{aligned}$$

the proof is complete. ■

Let U be open in $\mathbb{R}^n$ and let $f : U \to \mathbb{R}$. The Hessian of f at $x \in U$ is the $n \times n$ matrix

$$H_f(x) := \begin{bmatrix} D_1 D_1 f(x) & \cdots & D_n D_1 f(x) \\ \vdots & & \vdots \\ D_1 D_n f(x) & \cdots & D_n D_n f(x) \end{bmatrix}$$

Theorem: Suppose U is open in $\mathbb{R}^n$, $f : U \to \mathbb{R}$ and $x \mapsto D_i D_j f(x)$ is continuous at each $x \in U$ for every i and j. Then $H_f(x)$ is symmetric for each $x \in U$.

Example: Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$f(x_1, x_2) = \begin{cases} \frac{x_1^3 x_2 - x_1 x_2^3}{x_1^2 + x_2^2} & \text{if } (x_1, x_2) \neq (0,0) \\ 0 & \text{if } (x_1, x_2) = (0,0) \end{cases}$$

Then it is straightforward to check that $D_2 D_1 f(0,0) = -1 \neq 1 = D_1 D_2 f(0,0)$. The problem here is that $x \mapsto D_i D_j f(x)$ is not continuous at $(0,0)$.

Theorem: Suppose $U \subseteq \mathbb{R}^n$ is an open set containing a and b and $f : U \to \mathbb{R}$. Furthermore suppose that U contains $(1-t)a + tb$ for each $t \in [0,1]$. Suppose that both f and ∇f are differentiable on U. Then there is some $t' \in (0,1)$ such that

$$f(b) = f(a) + \nabla f(a) \cdot (b-a) + \frac{1}{2}(b-a)^T H_f(c')(b-a)$$

where $c' = (1-t')a + t'b$.

Proof: Let $g : [0,1] \to \mathbb{R}$ be defined by

$$\begin{aligned} g(t) &= f(b) - f((1-t)a + tb) - [\nabla f((1-t)a+tb)\cdot(b-a)](1-t) \\ &\quad -[f(b) - f(a) - \nabla f(a)\cdot(b-a)](1-t)^2 \end{aligned}$$

so that the conditions of Rolle's Theorem are satisfied. (Note that $g(0) = 0 = g(1)$.) Then there is some t' such that $g'(t') = 0$. Let $c' = (1-t')a + t'b$. Since

$$\begin{aligned} g'(t') &= -\nabla f(c')\cdot(b-a) \\ &\quad -\{(-1)\nabla f(c')\cdot(b-a) + (1-t')(b-a)^T H_f(c')(b-a)\} \\ &\quad -\{f(b) - f(a) - \nabla f(a)\cdot(b-a)\}(-1)2(1-t') \end{aligned}$$

we get the desired result by rearranging $g'(t') = 0$. ■

Let $S \subseteq \mathbb{R}^n$ be convex.

A function $f : S \longrightarrow \mathbb{R}$ is convex if for all $t \in [0, 1]$ and for all $x, x' \in S$, we have $f((1 - t)x + tx') \leq (1 - t)f(x) + tf(x')$.

A function $f : S \longrightarrow \mathbb{R}$ is strictly convex if for all $t \in (0, 1)$ and for all $x, x' \in S$ such that $x \neq x'$ we have $f((1 - t)x + tx') < (1 - t)f(x) + tf(x')$.

Note that a strictly convex function is convex.

Theorem: Suppose that $U \subseteq \mathbb{R}^n$ is open and convex and that $f : U \to \mathbb{R}$ is differentiable. Then
1. f is convex if and only if $f(x') \geq f(x) + \nabla f(x) \cdot (x' - x)$ for each $x, x' \in U$.
2. f is strictly convex if and only if $f(x') > f(x) + \nabla f(x) \cdot (x' - x)$ for each $x, x' \in U$ such that $x \neq x'$.

Proof: (1,$\Rightarrow$) Let f be convex. Pick $x, x' \in U$. Define $g : \mathbb{R} \to \mathbb{R}$ by $g(t) = f((1-t)x + tx')$. Note that g is differentiable, with $g'(t) = \nabla f((1-t)x + tx') \cdot (x' - x)$.
For all $t \in (0, 1]$ we have $f((1-t)x + tx') \leq (1-t)f(x) + tf(x')$.
Rearranging, we get

$$\frac{f((1-t)x + tx') - f(x)}{t} \leq f(x') - f(x).$$

The LHS of this inequality is precisely $(g(t) - g(0))/t$. Taking limit as $t \to 0$ we get $g'(0) \leq f(x') - f(x)$. It follows that $\nabla f(x) \cdot (x' - x) + f(x) \leq f(x')$ just as we desired.

$(1, \Leftarrow)$ Pick $x, x' \in U$ and $t \in [0,1]$. Let $x'' = (1-t)x + tx'$. Since U is convex, $x'' \in U$ and we have

$$\begin{array}{rcl} f(x) & \geq & f(x'') + \nabla f(x'')(x - x'') \\ f(x') & \geq & f(x'') + \nabla f(x'')(x' - x'') \end{array}$$

so that

$$(1-t)f(x) + tf(x') \geq f(x'').$$

Hence f is convex.

(2, $\Rightarrow$) Let f be strictly convex. Pick distinct $x, x' \in U$. Since strict convexity implies convexity, we know by (1) that $f(x') \geq f(x) + \nabla f(x) \cdot (x' - x)$. So we need to show that $f(x') \neq f(x) + \nabla f(x) \cdot (x' - x)$. Pick $t \in (0, 1)$ and suppose, towards a contradiction, that $f(x') = f(x) + \nabla f(x) \cdot (x' - x)$. Then

$$
\begin{aligned}
(1-t)f(x) + tf(x') &= f(x) + t(f(x') - f(x)) \\
&= f(x) + t[\nabla f(x) \cdot (x' - x)] \\
&= f(x) + \nabla f(x) \cdot (\underbrace{(1-t)x + tx'}_{=:x'' \in U} - x) \\
&\leq f(x) + f(x'') - f(x) \\
&= f(x'') \\
&= f((1-t)x + tx')
\end{aligned}
$$

where the weak inequality is by convexity of f. This contradicts strict convexity and the proof is complete.

(2, $\Leftarrow$) Pick $x \neq x'$ and $t \in (0,1)$. Let $x'' = (1-t)x + tx'$ and note that $x'' \in U \backslash \{x, x'\}$. Proceed as in (1, $\Leftarrow$) above. $\blacksquare$