Optimal Transport Math Seminar: Section 4.2

Vytis Krupovnickas

April 11, 2024

Abstract

Introduction and outline to lecture: First Talk over Definition 4.4, Proposition 4.5, and half of proposition 4.6; Second Talk over other half of proposition 4.6, and proposition 4.7. The goal is to provide some definitions of convex functions and use some ideas from convex analysis to characterize the subdifferential of transport maps.

1 Preliminary Results from Convex Analysis

Outline of section:

The goal of this section is to provide various results from Convex Analysis. We cover the definition of a convex conjugate, and the propositions following from it (Propositions 4.5, 4.6, 4.7) which are used to characterize the subgradient.

Definition 4.4. The convex conjugate of a proper function $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is defined by

$$\phi^*(y) = sup_{x \in \mathbb{R}^n}(x \cdot y - \phi(x))$$

.

Then we can characterize the subdifferential, as given in proposition 4.5. The subdifferential is a set associated with a convex function that generalizes the concept of the derivative.

2 Propositions

Proposition 4.5. Let ϕ be a proper, lower semi-continuous, convex function on $\mathbb{R}^n$. Then for all $x, y \in \mathbb{R}^n$

$$x \cdot y = \phi(x) + \phi^*(y) \iff y \in \partial\phi(x).$$

Proof:

Since $\phi^*(y) \geq x \cdot y - \phi(x)$ for all x, y we have:

$$x \cdot y = \phi(x) + \phi^*(y) \iff x \cdot y \geq \phi(x) + \phi^*(y)$$

Which is simply obtained from rearranging $\phi^*(y) \geq x \cdot y - -\phi(x)$ for $x \cdot y$. Then,

$$\iff x \cdot y \geq \phi(x) + y \cdot z - \phi(z) \; \forall z \in \mathbb{R}^d$$

Where we are using a modified version of Definition 4.4. to define the conjugate. Finally,

$$\iff \phi(z) \geq \phi(x) + y \cdot (z - x) \; \forall z \in \mathbb{R}^d$$

Where we have subtracted $x \cdot y$ from both sides and added $\phi(z)$ to both sides. And thus:

$$\iff y \in \partial\phi(x)$$

This concludes the proof to the proposition.

Proposition 4.6. If $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex then (1) ϕ is almost everywhere differentiable and (2) whenever ϕ is differentiable $\partial\phi(x) = \nabla\phi(x)$.

Proof: Let $x \in int(Dom(\phi))$ and δ^* be such that $\bar{B}(x, \delta^*) \subset int(Dom(\phi))$. We show that ϕ is Lipschitz continuous on $B(x, \delta^*/4)$. Then, by Rademacher's theorem, ϕ is differentiable almost everywhere on $B(x, \delta^*/4)$, and therefore differentiable almost everywhere on $int(Dom(\phi))$. This will complete the proof of (1).

What is said above is that we want to consider a point x in the interior of the domain of ϕ. We center a ball at x with a radius δ, so that the closure of the ball $(\bar{B}(x, \delta))$ is contained within the interior of the domain of ϕ (again, denoted $int(Dom(\phi))$). Our goal is to show that ϕ is Lipschitz continuous on a smaller ball $B(x, \delta^*/4)$. What is means is that we need to show that ϕ does not vary too rapidly within the bigger ball.

More precisely, Lipschitz continuity on $B(x, \delta/4)$ is proven by first showing that ϕ is uniformly bounded on the larger ball $\bar{B}(x, \delta/2)$. This is important because a uniformly bounded convex function on a compact set is Lipschitz continuous.

We show ϕ is Lipschitz on $B(x, \delta^*/4)$ by first showing that ϕ is uniformly bounded on $\bar{B}(x, \delta^*/2)$. By the Minkowshi-Carathéodory Theorem (this theorem expresses any point y in the ball $B(x, \delta)$ so that we can write y as the sum $\sum_{i=0}^n \lambda_i x_i$) there exists $x_{ii=0}^n \subset \partial B(x, \delta^*)$ such that for all $y \in B(x, \delta^*)$ there exists $\lambda_{ii=0}^n \subset [0, 1]$ with $\sum_{i=0}^n \lambda_i = 1$ and $y = \sum_{i=0}^n \lambda_i x_i$. So,

$$\phi(y) = \phi(\sum_{i=0}^n \lambda_i x_i) \leq \sum_{i=0}^n \lambda_i \phi(x_i) \leq max_{i=0,\ldots,n} |\phi(x_i|.$$

Here we are using the convexity of ϕ to show that $\phi(y)$ is less than or equal to a convex combination of the values of ϕ at the boundary points, which is bounded further by the maximum value of ϕ at these points. Thus we establish the uniform boundness of ϕ on $\bar{B}(x, \delta^*/2)$.

Now for $y \in B(x, \delta^*)$ and $y' = x - (y - x) = 2x - y$ we have $y' \in B(x, \delta^*)$ and $x = \frac{1}{2}y' + \frac{1}{2}y$. Therefore $\phi(x) \leq \frac{1}{2}\phi(y') + \frac{1}{2}\phi(y)$. In particular,

$$\phi(y) \geq 2\phi(x) - \phi(y') \geq 2\phi(x) - max_{i=0,\ldots,n} |\phi(x_i)|$$

Here, we are making a symmetry argument. For any point y in the ball $B(x, \delta^*)$, a symmetric point y' is constructed such that x is the midpoint between y and y'. We again use the convexity of ϕ to show that $\phi(y)$ is bounded below by the maximum value of ϕ at its boundary points.

We have shown that

$$2\phi(x) - max_{i=0,\ldots,n} |\phi(x_i)| \leq \phi(y) \leq max_{i,\ldots,n} |\phi(x_i)| \; \forall y \in B(x, \delta^*).$$

Rademacher's theorem states that a Lipschitz continuous function on an open subset of $\mathbb{R}^n$ is differentiable almost everywhere on that set. Since ϕ is shown to be Lipschitz continuous on $B(x, \delta^*/4)$, it follows that ϕ is differentiable almost everywhere on this ball, and by extension, almost everywhere on $int(Dom(\phi))$. This completes the proof of part (1) of the statement.

Part (2) of the statement is a consequence of the definition of the subdifferential and the differentiability of ϕ. When ϕ is differentiable at a point x, its subdifferential $\partial\phi(x)$ consists of a single element, which is the gradient $\nabla\phi(x)$. This follows from the properties of convex functions and their subdifferentials.

To show ϕ is Lipschitz on $B(x, \delta^*/4)$ let $x_1, x_2 \in B(x, \delta^*/4)(x_1 \neq x_2)$ and take x_3 to be the point of intersection of the line through x_1 and x_2 with $\partial B(x, \delta^*/2)$; there are two possibilities for x_3, we choose the option where x_2 lies between x_1 and x_3. Let $\lambda = |\frac{x_2 - x_3}{x_1 - x_3}| \in (0, 1)$. Note that $\lambda \in (0, 1)$ because x_2 is between x_1 and x_3. Now,

$$\lambda x_1 + (1 - \lambda)x_3 = \lambda x_2 + \lambda(x_1 - x_2) + (1 - \lambda)x_2 + (1 - \lambda)(x_3 - x_2)$$

Here we are expressing x_2 as a convex combination of x_1 and x_3: $\lambda x_1 + (1 - \lambda)x_3$. This is equivalent to saying that x_2 lies on the line segment between x_1 and x_3. The previous and next few steps are rewriting the convex combination to obtain x_2:

$$\begin{aligned}
&= x_2 + \frac{|x_3 - x_2|(x_1 - x_2)}{|x_3 - x_1|} + \frac{(|x_3 - x_1| - |x_3 - x_2|)(x_3 - x_2)}{|x_3 - x_1|} \\
&= x_2 + \frac{1}{|x_3 - x_1|}(|x_3 - x_2|(x_1 - x_2) + |x_2 - x_1|(x_3 - x_2)) \\
&= x_2
\end{aligned}$$

The last step uses the fact that $\frac{x_2 - x_1}{|x_2 - x_1|} = \frac{x_3 - x_2}{|x_3 - x_2|}$, which means that the vectors $x_2 - x_1$ and $x_3 - x_2$ are parallel and in the same direction. So by convexity of ϕ,

$$\begin{aligned}
\phi(x_2) - \phi(x_1) &\leq (1 - \lambda)(\phi(x_3) - \phi(x_1)) \\
&= \frac{|x_1 - x_3| - (x_2 - x_3)}{|x_1 - x_3|}(\phi(x_3) - \phi(x_1))
\end{aligned}$$

We can bound the difference $\phi(x_3) - \phi(x_1)$ by the maximum norm of ϕ on the ball $B(x, \bar{\delta}/2)$, denoted by $M = ||\phi||_{L^\infty B(x,\bar{\delta}/2)}$. Additionally, we use the fact that $|x_1 - x_3| \geq \delta^*/4$ to get:

$$\leq \frac{8M|x_1 - x_2|}{\delta^*}$$

Switching x_1 and x_2 implies that $|\phi(x_2) - \phi(x_1)| \leq \frac{8M|x_1 - x_2|}{\delta^*}$ hence ϕ is Lipschitz continuous, with constant $L = \frac{8M}{\delta^*}$ in $B(x, \delta^*/4)$.

For (2) we assume that ϕ is differentiable at x. Then the linear approximation of ϕ at x is given by $\phi(x) + \nabla\phi(x) \cdot (z - x)$. Then,

$$\begin{aligned}
\phi(x) + \nabla\phi(x) \cdot (z - x) &= \phi(x) + lim_{h \to 0^+} \frac{\phi(x + (z - x)h) - \phi(x)}{h} \\
&= \phi(x) + lim_{h \to 0^+} \frac{\phi((1 - h)x + hz) - \phi(x)}{h}
\end{aligned}$$

Then, using the convexity of ϕ, we have:

$$\leq \phi(x) + lim_{h\to 0^+}\frac{(1-h)\phi(x) + h\phi(z) - \phi(x)}{h}$$
$$= \phi(z).$$

This shows that $\nabla\phi(x) \in \partial\phi(x)$, where $\partial\phi(x)$ is the subdifferential of ϕ at x. Now if $y \in \partial\phi(x)$ then

$$\phi(x) + y \cdot (z - x) \leq \phi(z)$$

for all $z \in \mathbb{R}^n$. Let $z = x + hw$, where $h > 0$ and $w \in \mathbb{R}^n$. Then, we have:

$$y \cdot w \leq \frac{\phi(x + hw) - \phi(x)}{h}$$

for all $h > 0$ and $w \in \mathbb{R}^n$. Letting $h \to 0^+$ we have $y \cdot w \leq \nabla\phi(x) \cdot w$ for all $w \in \mathbb{R}^n$. Substituting $w \mapsto -w$ we have $y \cdot w = \nabla\phi(x) \cdot w$ for all $w \in \mathbb{R}^n$. Hence $y = \nabla\phi(x)$. This shows that if ϕ is differentiable at x, then $\nabla\phi(x)$ is the unique element of the subdifferential $\partial\phi(x)$, which means that $\partial\phi(x) = \nabla\phi(x)$.

Proposition 4.7. Let $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be proper. Then the following are equivalent:

1. ϕ is convex and lower semi-continuous;
2. $\phi = \psi^*$ for some proper function ψ;
3. $\psi^{**} = \phi$.

Proof: First we want to show that ϕ is convex, using the definition of convexity provided in the Definitions section. First define $x_1, x_2 \in \mathbb{R}^n, t \in [0, 1]$, and then define ϕ using the definition of convexity:

$$\phi(tx_1 + (1-t)x_2) = \psi^*(tx_1 + (1-t)x_2)$$

Which is using the identity $\phi = \psi^*$. Then, from definition 4.4.:

$$= sup_{y\in\mathbb{R}^n}((tx_1 + (1-t)x_2) \cdot y - \psi(y))$$

Then, we know that the supremum is less than or equal to the sum of the supremums of the function:

$$\leq sup_{y\in\mathbb{R}^n}(tx_1 \cdot y - t\psi(y)) + sup_{y\in\mathbb{R}^n}((1-t)x_2 \cdot y - (1-t)\psi(y))$$

Finally we can use the linearity property of the Legendre-Fenchel transform to obtain:

$$= t\psi^*(x_1) + (1-t)\psi^*(x_2)$$

And from the proposition:

$$= t\psi(x_1) + (1-t)\psi(x_2)$$

Now we want to show that ϕ is lower semi-continuous. Let $x_m \to x$, then:

$$lim_{m\to\infty}inf\phi(x_m) = lim_{m\to\infty}inf sup_{y\in\mathbb{R}^n}(x_m \cdot y - \phi(y)) \geq lim_{m\to\infty}(x_m \cdot y - \psi(y)) = x \cdot y - \psi(y).$$

This follows from the definition of lower semi-continuity. Taking the supremum over $y \in \mathbb{R}^n$ implies $liminf_{m\to\infty}\phi(x_n) \geq \phi(x)$. This show that the second condition of

Proposition 4.7 impres the first condition. Next we want to show the first condition implies the third condition. Now that we've shown ϕ is a lower semi-continuous and convex function, we want to show that $\phi(x) = \phi^{**}(x)$. $\phi^* \geq x \cdot y - \phi(x)$ for all $y \in \mathbb{R}^n$ from Proposition 4.5., we obtain:

$$\phi(x) \geq sup_{y \in \mathbb{R}^n}(x \cdot y - \phi^*(y)) = \phi^{**}(x)$$

And thus we show that $\phi(x)$ is greater than or equal to $\phi^{**}(x)$, so now we must show that $\phi \leq \phi^{**}(x)$, and from this can conclude $\phi(x) = \phi^{**}(x)$. Let $x \in int(Dom(\phi))$, then since ϕ can be bounded below by an affine function passing through $\phi(x)$ as ϕ is convex then $\partial\phi(x) \neq \varnothing$. Then, let $y_0 \in \partial\phi(x)$. By Proposition 4.5, $x \cdot y_0 = \phi(x) + \phi^*(y_0)$. Now we can describe:

$$\phi(x) = x \cdot y_0 - \phi^*(y_0) \leq sup_{y \in \mathbb{R}^n}(x \cdot y - \phi^*(y)) = \phi^{**}(x)$$

Where we have simply rearranged the equation to obtain and relation for $\phi(x)$ and we have again used Definition 4.4. Now the proposition is proved for any ϕ with $int(Dom(\phi)) = \mathbb{R}^d$. Now we must cover our bases and prove this for all other ϕ for a small part ϵ. Define $\psi_\epsilon = \frac{|x|^2}{\epsilon}$, so:

$$\phi_\epsilon(x) = inf_{y \in \mathbb{R}^d}(\phi(x-y) + \psi_\epsilon(y)) = inf_{y \in \mathbb{R}^d}(\phi(y) + \psi_\epsilon(x-y)).$$

Now we want to show $\phi_\epsilon = \phi_\epsilon^{**}$ on $\mathbb{R}^d$. To do this, we only need to show that $\phi_{epsilon}$ is convex, lower semi-continuous, and $int(Dom(\phi_\epsilon)) = \mathbb{R}^d$. First, let's handle convexity, as before:

$$\phi_\epsilon(tx_1 + (1-t)x_2) = inf_{y \in \mathbb{R}}(\phi(tx_1 + (1-t)x_2 - y) + \phi_\epsilon(y))$$

Where instead of using the supremum, we use the infimum, from the definition provided before. Then we defined two points $y_1.y_2$, where we define $y_1 = x_1 + y$ and $y_2 = x_2 - y$, which we use to compare to a convex combination of $\phi(x_1)$ and $\phi(x_2)$. Thus:

$$= inf_{y_1,y_2 \in \mathbb{R}^d}(\phi(t(x_1 - y_1) + (1-t)(x_2 - y_2)) + \psi_\epsilon(ty_1 + (1-t)y_2))$$
$$\leq inf_{y_1,y_2 \in \mathbb{R}^d}(t(\phi(x_1 - y_1) + \psi_\epsilon(y_1)) + (1-t)(\phi(x_2 - y_2) + \phi_\epsilon(y_2)))$$
$$= t\phi_\epsilon(x_1) + (1-t)\phi_\epsilon(x_2)$$

Where we follow the same logic as before in showing ϕ is convex. Now we wish to show ϕ_ϵ is lower semi-continuous. This is simple; the pointwise limit (defined before) of an arbitrary collection of lower semi-continuous functions is itself lower semi-continuous, so ϕ_ϵ is lower semi-continuous. Finally, if we let $x \in \mathbb{R}^d$ (and recall ϕ is proper) then there exists $y_0 \in \mathbb{R}^d$ such that $\phi(y_0) < \infty$, so $\phi_\epsilon(x) \leq \phi(y_0) + \psi_\epsilon(x - y_0)$. ϕ is finite everywhere, so it follows that $\phi_\epsilon(x)$ is finite. Thus, $x \in Dom(\phi_\epsilon)$ and $int(Dom(\phi_\epsilon)) = \mathbb{R}^d$. Now we show that $liminf_{\epsilon \to 0}\phi_\epsilon(x) \geq \phi(x)$. This is the last step in showing that $\phi(x) = \phi^{**}(x)$, as we can show the constraints of the inequalities do not leave room for a discontinuity. Fix $x \in \mathbb{R}^n$ and note $\phi_\epsilon(x) \leq A + \frac{B}{\epsilon}$. Since ϕ is convex, it is bounded from below by an affine function (again, from the definition of convexity), so we can say $\phi(z) \geq a \cdot z + b \forall z \in \mathbb{R}^n$, which logically follows. Let y_ϵ be a minimizing sequence, such that $\phi_\epsilon(x) \geq \phi(x - y_\epsilon) + \psi_\epsilon(y_\epsilon) - \epsilon$. Then:

$$A + \frac{B}{\epsilon} \geq \phi_\epsilon(x)$$

(From the definition before)

$$\geq a \cdot (x - y_\epsilon) + \frac{|y_\epsilon|^2}{\epsilon} - \epsilon$$

Here, we are using the definition $\phi(z) \geq a \cdot z + b$ and $\phi_\epsilon(x) \geq \phi(x - y_\epsilon) + \psi_\epsilon(y_\epsilon) - \epsilon$. Here, $z = (x - y_\epsilon)$ and $b = \frac{|y_\epsilon|^2}{\epsilon} - \epsilon$. We express the function $\psi_\epsilon(y_\epsilon)$ as $\frac{|y_\epsilon|^2}{\epsilon}$. $\frac{|y_\epsilon|^2}{\epsilon}$ represents the magnitude of the minimizing sequence relative to ϵ, which is an important quantity for understanding the behavior of the expression as ϵ approaches 0. Then:

$$\geq -\frac{(1+\epsilon)|a|^2}{\epsilon} - \frac{|x|^2}{2} + \frac{|y_\epsilon|^2}{2\epsilon} - \epsilon$$

This implies that $|y_\epsilon| = O(1)$. Now let $\epsilon_n \to 0$ be a subsequence with $liminf_{\epsilon \to 0} \phi_\epsilon(x) = lim_{n \to \infty} \phi_{\epsilon_n}(x)$. We have determined y_{ϵ_n} is bounded, so there exists another subsequence and some $y \in \mathbb{R}^n$ such that $y_{\epsilon_n} \to y$. Furthermore,

$$lim_{n \to \infty} \phi_{\epsilon_n}(x) = lim_{n \to \infty} (\phi(x - y_{\epsilon_n}) + \phi_{\epsilon_n}(y_{\epsilon_n})) \geq \begin{cases} \phi(x) & \text{if } y = 0 \\ +\infty & \text{else} \end{cases}$$

The right hand side is greater than $\phi(x)$ in both cases, so $lim_{n \to \infty} \phi_{\epsilon_n}(x) \geq \phi(x)$. Since $\phi(x) \geq \phi_\epsilon(x)$ then:

$$\phi^{**}(x) = sup_{y \in \mathbb{R}^n} inf_{z \in \mathbb{R}^n} (y \cdot (x - z) + \phi(z) \geq sup_{y \in \mathbb{R}^n} inf_{z \in \mathbb{R}^n} (y \cdot (x - z) + \phi_\epsilon(z)) = \phi^{**}(x)$$

Thus,

$$\phi^{**}(x) \geq lim_{\epsilon \to 0} inf \phi^{**}(x) = lim_{\epsilon \to 0} inf \phi_\epsilon(x) \geq \phi(x)$$

Thus, we conclude the proof.

3 Definitions

3.1 Convex Functions

A convex function is a function such that a line segment between any two points on the graph lie above the graph. More rigorously, we define a function as convex if any of the following two conditions hold: 1. For all $0 \leq t \leq 1$ and all $x_1, x_2 \in X$:

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2)$$

2. For all $0 \leq t \leq 1$ and all $x_1, x_2 \in X$ such that $x_1 \neq x_2$:

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2)$$

These are essentially equivalent statements. This definition is important for the proof of Proposition 4.7.

3.2 Subgradient

The subgradient is a generalization of the derivative of convex functions at points which are not necessarily differentiable (for example, the inflection point of an absolute value function). We can describe a line which intersects non-differentiable points on this graph without necessarily touching the graph, and from this obtain a slope which we call the subderivative. More rigorously, we define the subderivative of a convex function $f : I \to \mathbb{R}$ at a point x_0 in the open interval I as a real number c such that

$$f(x) - f(x_0) \geq c(x - x_0)$$

for all $x \in I$. Additionally, the set of subderivatives at x_0 for a convex function is a nonempty closed interval $[a, b]$, where a and b are one-sided limits defined as:

$$a = \lim_{x \to x_0^-} \frac{f(x) - f(x_0)}{x - x_0}$$

and:

$$b = \lim_{x \to x_0^+} \frac{f(x) - f(x_0)}{x - x_0}$$

We denote the subdifferential as $\partial f(x_0)$.

3.3 Legendre-Fenchel Transform/Legendre Transform

In $\mathbb{R}$, we define an interval $I \cup \mathbb{R}$ and a convex function f as $f : I \to \mathbb{R}$. The Legendre transform of f is then the function $f^* : I^* \to \mathbb{R}$, where f^* is defined as:

$$f^*(x^*) = sup_{x \in I}(x^* x - f(x))$$

and I^* as:

$$I^* = \{x^* \in \mathbb{R} : f^* x^* < \infty\}.$$

In $\mathbb{R}^n$, we describe the Legendre transform for a function $f : X \to \mathbb{R}$ on a convex set $X \subset \mathbb{R}^n$ as $f^* : X^* \to \mathbb{R}$ with domain:

$$X^* = \{x^* \in \mathbb{R}^n : sup_{x \in X}(\langle x^*, x \rangle - f(x)) < \infty\}$$

where

$$f^*(x^*) = sup_{x \in X}(\langle x^*, x \rangle) - f(x))$$

Here $\langle x^*, x \rangle$ is the dot product of x^* and x, which is defined:

$$\langle u, v \rangle = \int_a^b u(x) v(x) dx$$

over some interval $[a, b]$. This is essentially the same as Definition 4.4.

3.4 Upper and Lower Semi-Continuity

f is upper semi-continuous at x_0 if and only if

$$limsup_{x \to x_0} \leq f(x_0).$$

f is lower semi-continuous at x_0 if and only if

$$limsup_{x \to x_0} \geq f(x_0).$$

3.5 Lipschitz Continuous

Lipschitz continuity describes the rate at which a function's value changes as its input varies. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to be Lipschitz continuous if there exists a constant K for any two points x_1 and x_2 such that:

$$|f(x_1) - f(x_2)| \leq K|x_1 - x_2|.$$

A function is said to be Lipschitz if there exists a double cone whose origin can be moved along the graph such that the graph remains outside the cones, in another sense.

3.6 Rademacher's Theorem

If $U \subset \mathbb{R}^n$ is open and $f : U \to \mathbb{R}$ is Lipschitz continuous then f us differentiable almost everywhere on U. Or, the points in U at which f is not differentiable form a set of Lebesque measure zero.

The Lebesque measure is a way of assigning a measure to subsets of higher dimensional Euclidean spaces.

3.7 Minkowski-Carathéodory Theorem

If a point x lies in the convex hull $Conv(P)$ of a set $P \subset \mathbb{R}^d$, then x can be written as the convex combination of at most $d+1$ points in P. x can be written as the complex combination of at most $d+1$ external points in P, as non-external points can be removed from P without changing the membership of x in the convex hull.

A convex hull is the smallest convex set which contains a shape. It is also the set of all convex combinations of points in the subset of a Euclidean space. A convex combination is a linear combination of points where all coefficients are non-negative and sum to 1. This is similar to a summed weighted average.

3.8 Pointwise Limit/Pointwise Convergence

Pointwise convergence describes a way for a sequence of functions to converge towards a particular function, and is weaker than continuous convergence. We call the function f the pointwise limit of the sequence (f_n) for a set X and topological space Y if the sequence (f_n) has the same domain as X and codomain as Y, written as $f : X \to Y$, which we write as:

$$lim_{n\to\infty} f_n = f pointwise$$

such that:

$$\forall x \in X. lim_{n\to\infty} f_n(x) = f(x).$$

3.9 Minimizing Sequence

A minimizing sequence $y_n \in M$ for a given element x of a metric space $X = (X, \rho)$ is a sequence for which:

$$\rho(x, y_n) =\to \rho(x, M) = inf \rho(x, y) : y \in M.$$

3.10 Proper Function, Infimum, Supremum, Uniformly Bounded, Affine Function, Subsequence

A function is called proper if its preimage is a compact set; essentially, if the inverse of the function is closed and has no "holes", then it is a proper function.

For two sets S and P, where S is a subset of P, the infimum is the greatest element in P which is less than or equal to every element of S. The supremum is the least element in P which is greater than or equal to every element of S.

A uniformly bounded family of functions is a family of bounded functions that can be bounded by the same constant, which is greater than or equal to the absolute value of any value of the functions in the family. In a metric space, let Y be a metric space with metric d, then the set:

$$\mathcal{F} = f_i : X \to Y, i \in I$$

is called uniformly bounded if there exists an element a from Y and a real number M such that:

$$d(f_i(x), a) \leq M \,\forall i \in I \,\forall x \in X.$$

Affine functions are essentially a summation of linear functions, defined as:

$$f(x_1, ..., x_n) = A_1 x_1 + ... + A_n x_n + b.$$

A subsequence is a sequence within a sequence that can be derived by deleting some or no elements of another sequence without changing the order of that sequence.

3.11 Notation

int(Dom) means the interior of the domain of.

The double bar, $||x||$, represents the norm of a function, defined as:

$$||x|| = \sqrt{x_1^2 + x_2^2 + ... + x_n^2}.$$

$\bar{x}$ represents the mean of a set in probability and statistics.

References

Terence Tao Analysis I, II