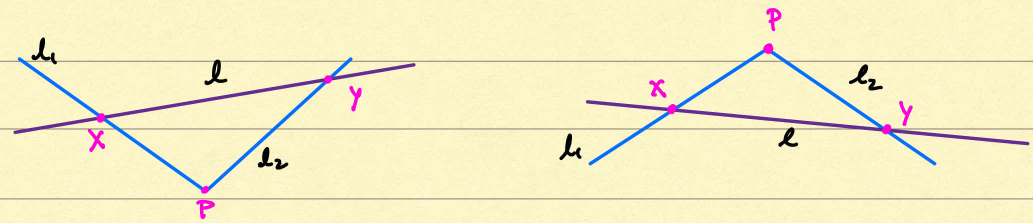
Sep 18, 2018 Lecture 11

Convex functions

Observation: Consider $\mathbb{R}^2$ regarded as the usual cartesian plane of analytic geometry. Suppose l_1 and l_2 are two lines in $\mathbb{R}^2$ meeting at a point T, X a point on l_1 to the left of P (i.e., the first coordinate of T is less than the first coordinate of P) and Y a point on l_2 to the right of P. Let l be the line joining X and Y. Then P lies below l if and only if the slope of l_1 is less than the slope of l_2.



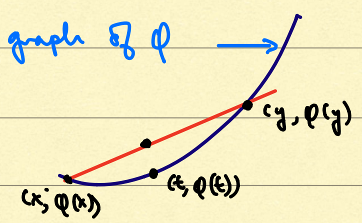
The proof is elementary and the details are left to you. Here is a sketch. Let R be the point on l whose x-coordinate is the same as that of P. We want to find conditions when the y-coordinate of R is larger than that of P. So let $X=(\alpha,\beta)$, $Y=(\gamma,\delta)$, $P=(p,q)$, $R=(p,s)$. Let $m_1 = \dfrac{\beta-q}{\alpha-p}$, $m_2 = \dfrac{\delta-q}{\gamma-p}$. The eqn of l is $y = \beta + \dfrac{\delta-\beta}{\gamma-\alpha}(x-\alpha)$. This means $s = \beta + \dfrac{\delta-\beta}{\gamma-\alpha}(p-\alpha)$. You have to show $q<s$ if and only if $m_1 < m_2$.

Definition (Convex function): Let (a,b) be an open interval in $\mathbb{R}$, $-\infty \le a < b \le \infty$. A function $\varphi : (a,b) \longrightarrow \mathbb{R}$ is said to be convex if

$$\varphi\left((1-\lambda)x + \lambda y\right) \le (1-\lambda)\varphi(x) + \lambda\varphi(y) \qquad (1)$$

whenever $x, y \in (a,b)$ and $0 \le \lambda \le 1$.

This is equivalent to saying that if $x, y \in (a,b)$ and $t \in (x,y)$, then $(t, \varphi(t))$ lies below or on the line connecting $(x, \varphi(x))$ and $(y, \varphi(y))$.



From our observation above, the condition for convexity of φ is equivalent to the condition

$$\frac{\varphi(t)-\varphi(s)}{t-s} \le \frac{\varphi(u)-\varphi(t)}{u-t} \qquad (2)$$

whenever $a < s < t < u < b$.

Theorem: If φ is convex on (a,b) then φ is continuous on (a,b).

Proof:

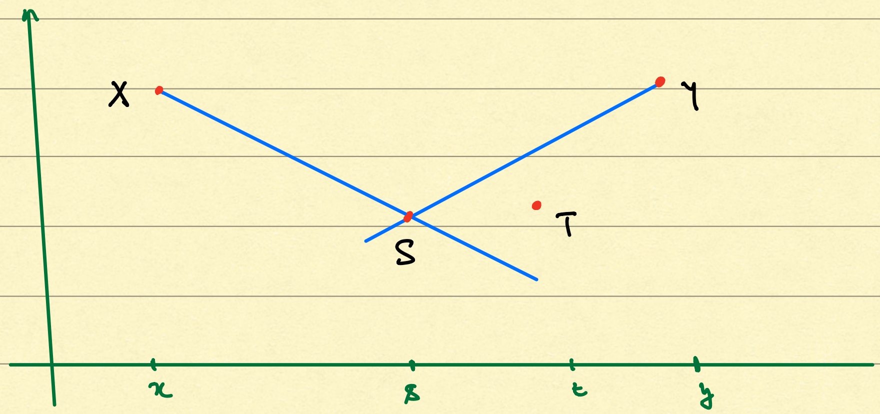
Let us prove φ is right continuous. The proof for left continuity is similar.

Suppose we $a < x < s < t < y < b$. We wish to examine $\lim_{t \to s} \varphi(t)$. Let

$$X = (x, \varphi(x)), \quad Y = (y, \varphi(y))$$
$$S = (s, \varphi(s)), \quad T = (t, \varphi(t)).$$

Then X is the left most point, Y the right most, and S is to the left of T.

Since S lies below the line joining X and T, therefore T lies above the line joining S and X. On the other hand T lies below the line joining S and Y.



Now let $t \longrightarrow s$. Since T is in the wedge in the picture, it follows that $T \longrightarrow S$ as $t \longrightarrow s$. Thus $\lim_{t \to s^+} \varphi(t) = \varphi(s)$. A similar argument would show that $\lim_{t \to s^-} \varphi(t) = \varphi(s)$. q.e.d.

Remark: Suppose $\lambda_1, \lambda_2, \dots, \lambda_n$ are non-negative numbers such that $\lambda_1 + \dots + \lambda_n = 1$, and $\varphi : (a,b) \longrightarrow \mathbb{R}$ a convex function. It is easy to see by induction that for $t_1, \dots, t_n \in (a,b)$,

$$\varphi\left(\sum_{i=1}^{n} \lambda_i t_i\right) \le \sum_{i=1}^{n} \lambda_i \varphi(t_i). \qquad \text{——— } (*)$$

The data (λ_i), (t_i) gives us a probability measure λ on $((a,b), \mathcal{L})$ where $\mathcal{L}$ is the Lebesgue σ-algebra on (a,b), namely

$$\lambda(E) = \sum_{j=1}^{k} \lambda_{i_j} \qquad \text{where } E \cap \{t_1, \dots, t_n\} = \{t_{i_1}, \dots, t_{i_k}\}$$

with the t_{i_j}'s distinct for $j = 1, \dots, k$. With this measure it is clear that $(*)$ is equivalent to saying

$$\varphi\left(\int_{(a,b)} t\, d\lambda\right) \le \int_{(a,b)} \varphi(t)\, d\lambda . \qquad (**)$$

More generally we have the following:

Theorem (Jensen's inequality): Let $(X, \mathcal{M}, \mu)$ be a measure space with $\mu(X)=1$, $f : X \longrightarrow (a,b)$ a measurable function and $\varphi\ (a,b) \longrightarrow \mathbb{R}$ a convex function. Then

$$\varphi\left(\int_X f\, d\mu\right) \le \int_X (\varphi \circ f)\, d\mu$$

Proof:

Let $t = \int_X f\, d\mu$. Then it is easy to see that $t \in (a,b)$. [Hint: Use the fact that if σ is a non-zero measure on $(Z, \mathcal{F})$, and $g > 0$ is a m'ble function on Z, then $\int_Z g\, d\sigma > 0$. (How will you prove this?) Use also the fact that if $\sigma(Z) = 1$, then for any constant c, $\int_Z c\, d\sigma = c$.]

Let $$\beta = \sup\left\{ \frac{\varphi(t) - \varphi(u)}{t-u} \;\middle|\; a < u < t \right\} \qquad (+)$$

Then $$\beta \le \frac{\varphi(s) - \varphi(t)}{s-t} \quad \forall\ s > t. \qquad (++) \quad (\text{by } (2))$$

It is then easy to see that

$$\varphi(s) \ge \varphi(t) + \beta(s-t) \qquad \forall\, s \in (a,b).$$

Indeed if $s \in [t,b)$ then the above follows from $(++)$ and if $s \in (a,t]$ then the above follows from $(+)$.

This means

$$\phi(f(x)) \geq \phi(t) + \beta(f(x) - t) \qquad \forall\, x \in X.$$

Integrating we get

$$\int_X \phi \circ f \, d\mu \geq \phi(t) + \beta\left(\int_X f \, d\mu - t\right)$$

$$= \phi\left(\int_X f \, d\mu\right) + \beta(t - t)$$

$$= \phi\left(\int_X f \, d\mu\right).$$

q.e.d.

Remark: In view of (2), a differentiable function ϕ on (a, b) is convex if and only if ϕ' is an increasing function. In particular $t \longmapsto e^t$ is convex. So is $t \longmapsto t^p$ if $p > 1$.

Conjugate exponents: If p and q are positive real numbers such that $p + q = pq$, or equivalently

$$\frac{1}{p} + \frac{1}{q} = 1$$

then p and q are called a pair of conjugate exponents.

It is clear from the relation above that $1 < p, q < \infty$.

Note that $p = q$ if and only if $p = 2$ (or $q = 2$).

We extend the definition of conjugate exponents to the case $p = 1$ and $p = \infty$ by setting $q = \infty$ in the first case and $q = 1$ in the second case. Note that with this extended definition, the relationship $\frac{1}{p} + \frac{1}{q} = 1$ continues to hold given our conventions regarding division by 0 and ∞.

Our interest is in L^p-spaces for $p \geq 1$. The spaces $L^p(\mu)$ and $L^q(\mu)$, p, q conjugate, share an interesting relation.

To see this we first need:

Theorem: Let p and q be a pair of conjugate exponents with $1<p<\infty$. Let $(X, \mathcal{M}, \mu)$ be a measure space, and f, g measurable functions on X taking values in $[0,\infty]$. Then

(a) (Hölder's inequality)

$$\int_X fg \, d\mu \leq \left\{ \int_X f^p \, d\mu \right\}^{1/p} \left\{ \int_X g^q \, d\mu \right\}^{1/q}$$

(b) (Minkowski's inequality)

$$\left\{ \int_X (f+g)^p \, d\mu \right\}^{1/p} \leq \left\{ \int_X f^p \, d\mu \right\}^{1/p} + \left\{ \int_X g^q \, d\mu \right\}^{1/q}$$

Remark: If $p=q=2$ then Hölder's inequality is the Cauchy-Schwarz inequality.

Proof:

(a) Let $A = \int_X f^p \, d\mu$ and $B = \int_X g^q \, d\mu$. If either A or B is ∞, then the Hölder inequality is clearly true. So assume both A and B are finite. If $A=0$, then clearly $f=0$ a.e., and hence $\int_X fg \, d\mu = 0$, and Hölder is trivially true. Similarly if $B=0$. So let us assume A and B are positive real numbers. Set

$$F = \frac{f}{A} \quad \text{and} \quad G = \frac{g}{B}.$$

Then

$$\int_X F^p \, d\mu = \int_X G^q \, d\mu = 1.$$

For $x \in \{F \neq 0\} \cap \{G \neq 0\}$ let r, s be real numbers such that $e^{r/p} = F(x)$ and $e^{s/q} = G(x)$. Since $t \mapsto e^t$ is convex we have $e^{r/p + s/q} \le \frac{1}{p} e^r + \frac{1}{q} e^s$. This means

$$F(x) G(x) \le \frac{1}{p} F(x)^p + \frac{1}{q} G(x)^q.$$

The above inequality is trivially true if $x \in \{F=0\} \cup \{G=0\}$, and hence is true for all $x \in X$. Integrate w.r.t μ to get

$$\int_X FG \, d\mu \le \frac{1}{p} \int_X F^p d\mu + \frac{1}{q} \int_X G^q d\mu = \frac{1}{p} + \frac{1}{q} = 1.$$

Hölder's inequality follows easily.

(b). By Hölder we have

$$\int_X f(f+g)^{p-1} d\mu \le \left\{\int_X f^p d\mu\right\}^{1/p} \left\{\int_X (f+g)^{(p-1)q} \, d\mu\right\}^{1/q}.$$

$$= \left\{\int_X f^p d\mu\right\}^{1/p} \left\{\int_X (f+g)^p \, d\mu\right\}^{1/q}$$

Reversing the role of f and g we get

$$\int_X g(f+g)^{p-1} d\mu \le \left\{\int_X g^p d\mu\right\}^{1/p} \left\{\int_X (f+g)^p d\mu\right\}^{1/q}.$$

Adding the two inequalities we get

$$(***) \qquad \int_X (f+g)^p d\mu \le \left(\left\{\int_X f^p d\mu\right\}^{1/p} + \left\{\int_X g^p d\mu\right\}^{1/p} \right) \left\{\int_X (f+g)^p d\mu\right\}^{1/q}$$

Now Minkowski is clearly true if either the left side of Minkowski is zero or its right side is ∞. So let us assume $\int_X (f+p)^p d\mu \neq 0$ and that both $\int_X f^p d\mu$ as well as $\int_X g^p d\mu$ are finite. Since $t \mapsto t^p$ is convex, we see that $\left(\frac{f+g}{2}\right)^p \le \frac{1}{2} f^p + \frac{1}{2} g^p$. This means $\int_X (f+p)^p d\mu < \infty$. Hence $(***)$ yields

$$\left\{ \int_X (f+g)^p \, d\mu \right\}^{1-\frac{1}{q}} \leq \left\{ \int_X f^p \, d\mu \right\}^{\frac{1}{p}} + \left\{ \int_X g^p \, d\mu \right\}^{\frac{1}{p}}.$$

i.e.

$$\left\{ \int_X (f+g)^p \, d\mu \right\}^{\frac{1}{p}} \leq \left\{ \int_X f^p \, d\mu \right\}^{\frac{1}{p}} + \left\{ \int_X g^p \, d\mu \right\}^{\frac{1}{p}}.$$

q.e.d.