Lecture 2

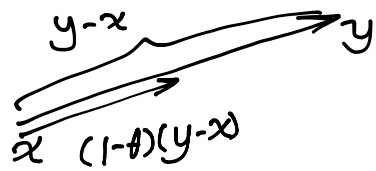
Today's convex functions

We are interested b/c for a convex function, every local minimum is a global minimum.

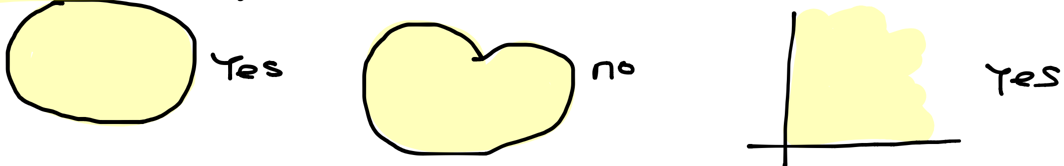
Def: A set $\Omega \subseteq E^n$ is convex if $\forall x, y \in \Omega$ and $\theta \in [0,1]$.
$\theta x + (1-\theta) y \in \Omega$

This means that if $x, y \in \Omega$, then the line segment joining $x$ to $y$ is contained in $\Omega$.

$$\theta x + (1-\theta) y = x + \underbrace{(1-\theta)}_{\text{between 0 and 1}}(y-x)$$




since as picture shows, the set of points $\{\theta x + (1-\theta) y : 0 \le \theta \le 1\}$ is exactly the line segment from $x$ to $y$.



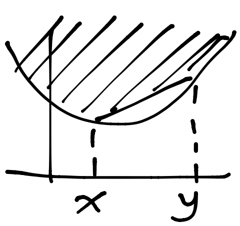

x-axis in x-y coordinate is convex.

A function $f$ on a convex set $\Omega \in E^n$ is convex if for any $x, y \in \Omega$ and $\theta \in [0,1]$. So $f(\theta x + (1-\theta) y) \le \theta f(x) + (1-\theta) f(y)$

strictly convex if $f(\theta x) + (1-\theta) y) < \theta f(x) + (1-\theta) f(y)$

- $f$ is concave if $-f$ is convex.
  strictly concave if $-f$ is strictly convex.

e.g.




$f$ on the line segment joining $x$ and $y$ $\le$ linear function on line segment that equals $f$ at end pts.

i.e. is below

Remark: $f$ is a convex function iff $\{(x, z) \in E^{n+1} : z \ge f(x)\}$ is a convex set.

To prove ① Assume the function is convex and show that set is convex. (using def of convex set & convex function).
② Assume set is convex, deduce that function is convex.

Properties of convex functions:
① If $f_1$ & $f_2$ are convex functions
(always understood to be defined on a convex set $\Omega$)
Then $f_1 + f_2$ is convex.

I must show that for any $x, y$ in $\Omega$, any $\theta \in [0,1]$
$(f_1+f_2)(\theta x + (1-\theta) y) \le \theta (f_1+f_2)(x) + (1-\theta)(f_1+f_2)(y)$

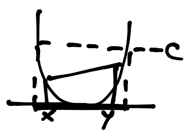
True b/c:

$(f_1+f_2)(\theta x+(1-\theta)y)=f_1(\theta x+(1-\theta)y)+f_2(\theta x+(1-\theta)y)$
$\qquad\qquad \le \theta f_1(x)+(1-\theta)f_1(y)+\theta f_2(x)+(1-\theta)f_2(y)$
$\qquad\qquad =\theta(f_1+f_2)(x)+(1-\theta)(f_1+f_2)(y)$

② If $f$ is convex and $a>0$, then $a\cdot f>0$ is convex.
Pf: similar but easier.

Note: by combining the above:
if $f_1,\dots,f_k$ are convex and $a_1,\dots,a_k\ge 0$ then $a_1f_1+\dots+a_kf_k$ is convex.

③ If $f$ is a convex function on convex set $\Omega$, then for any number $c$, $\{x\in\Omega: f(x)\le c\}$ is a convex set.

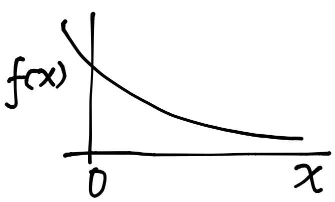




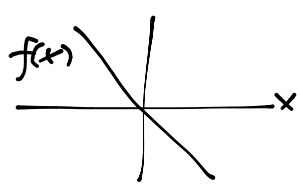
Proof: Sps $x,y\in\{z\in\Omega: f(z)<c\}$
i.e. $f(x)\le c$ and $f(y)\le c$

Then for any $\theta\in[0,1]$, by convexity $f(\theta x,(1-\theta)y))\le\theta f(x)+(1-\theta)f(y)\le\theta\cdot c+(1-\theta)\cdot c=c$

so $\theta x+(1-\theta)y\in\{z\in\Omega: f(z)\le c\}$
so this set is convex




Note: convex functions need not "curve up" in all directions.

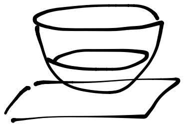



④ If $f$ is a $C^1$ function on a convex set $\Omega$, then $f$'s convex if
$\boxed{f(y)\ge f(x)+\nabla f(x)(y-x)\ ,\ \forall x,y\in\Omega}$

Note: If I write $y=x+v$ then above becomes
$\boxed{f(x+v)\ge \underbrace{f(x)+\nabla f(x)\cdot v}\ ,\ \forall x,v\in\Omega}$

first order Taylor approximation of $f$ near $x$.

equation for "tangent plane" to $f$ at $x$ (as function of $v$).



"function $\geq$ every tangent plane"

Proof:

① Assume $f$ is convex

Fix $x, y \in \Omega$, have to show $\star$

By convexity:

$$f(\theta x+(1-\theta)y) \leq \theta f(x)+(1-\theta)f(y), \forall \theta \in [0,1]$$

$$\| \quad f(x+(1-\theta)(y-x))$$

Subtract $f(x)$ from both sides:

$$f(x+(1-\theta)(y-x)) - f(x) \leq (1-\theta)(f(y)-f(x))$$

so $$\frac{f(x+(1-\theta)(y-x))-f(x)}{(1-\theta)} \leq f(y)-f(x)$$

$$\lim_{\theta\to 1} \frac{f(x+(1-\theta)(y-x))-f(x)}{(1-\theta)} \leq f(y)-f(x)$$

(directional derivative)

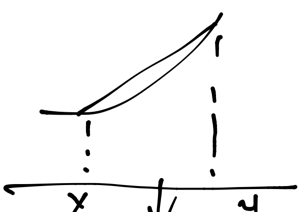
since not hard to check that

$$\lim_{\theta\to 1} \frac{f(x+(1-\theta)(y-x))-f(x)}{1-\theta} = \nabla f(x)(y-x)$$

(check it!)

this proves $\star$

$$\frac{d}{d\theta}\Big|_{\theta=1} f(x+(1-\theta)(y-x))$$




$$x_\theta = \theta x + (1-\theta)y$$

Now we assume $\star$ and try to show that $f$ is convex.

Fix $x$ and $y \in \Omega$ and $\theta \in [0,1]$.

Let $x_\theta = \theta x + (1-\theta)y$

By condition $\star$

$$f(x) \geq f(x_\theta) + \nabla f(x_\theta)(x - x_\theta) \qquad (1)$$

and

$$f(y) \geq f(x_\theta) + \nabla f(x_\theta)(y - x_\theta) \qquad (2)$$

Also, $x - x_\theta = (1-\theta)(x-y)$ and $y - x_\theta = \theta(y-x)$

so: $\theta(x-x_\theta)+(1-\theta)(y-x_\theta) = 0$

$\theta$(eqn 1) + $(1-\theta)$(eqn 2) $\Longrightarrow$ $\boxed{\theta f(x)+(1-\theta)f(y) \geq f(x_\theta)}$

so $f$ is convex. ■

⑤ If $f$ is a $C^2$ function on a convex set $\Omega$, then $f$ is convex iff

$$\boxed{v^T \nabla^2 f(x) v \geq 0, \ \forall x \in \Omega, \text{ and } v \in E^n} \qquad (\star\star)$$

i.e. $\nabla^2 f$ positive semi-definite.

Proof: ① convexity $\Rightarrow$ $(\star\star)$

Sufficient to show that

if $(\star\star)$ not true $\Rightarrow f$ not convex (contrapositive)

if $(\star\star)$ not true then $\exists\, x \in \Omega$ and $v \in E^n$ s.t. $v^T \nabla^2 f(x) v < 0$.

By Taylor's Theorem:

$$f(x+hv) = f(x) + h\nabla f(x) v + \tfrac{1}{2} h^2 v^T \nabla^2 f(x) v + O(h^2 |v|^2)$$

This implies that if $h$ is small enough, $\frac{1}{2} h^2 v^T \nabla^2 f(x) v + o(h^2 |v|^2) < 0$

so $f(x+hv) < f(x) + \nabla f(x)(hv)$

So $f$ is non-convex by property ④

Now assume $(\star\star)$. We'll show that property ④ holds.

Fix $x, y$. Goal: $f(y) \geq f(x) + \nabla f(x)(y-x)$

Let $g(s) = f(x + s(y-x))$

Then $f(y) = g(1)$

$f(x) = g(0)$

By MVT, $\dfrac{g(1)-g(0)}{1-0} = g'(\theta)$ for some $\theta \in (0,1)$

$$g(1) - g(0) = g'(\theta) = g'(0) + (g'(\theta) - g'(0))$$
$$= g'(0) + \theta g''(s) \text{ for some } s \in (0, \theta)$$

then one can check that

$g'(0) = \nabla f(x)(y-x)$ and $g''(\theta) = (y-x)^T \nabla^2 f(x + s(y-x))(y-x)$

has the form $v^T \nabla^2 f(z) v$, so $\cdots \geq 0$

Put together to get $f(y) - f(x) \geq \nabla f(x)(y-x)$

∎

---

Thm 1: If $f$ is a convex function on a convex set $\Omega$, then

① $\Gamma = \{x \in \Omega : f(x) = \inf_\Omega f\}$ is convex, and

② every local min is a global min

Proof: ① Let $c = \inf_\Omega f$

Then $\Gamma = \{x \in \Omega : f(x) \leq c\}$

and this is convex by property ③

Next: ② sufficied to show that if $x$ is not a global min, then $x$ is not a local min, then $\exists\, y$ s.t. $f(y) < f(x)$.

so $f(\theta x+(1-\theta)y)\le\theta f(x)+(1-\theta)f(y)<f(x), \forall\theta\in[0,1]$ which implies that $x$ is not a local min.

Thm 2: If $f$ is convex on $\Omega$ and $x^*$ is a point where $\boxed{\nabla f(x^*)(y-x^*)\ge 0}$ for all $y\in\Omega$, then $x^*$ is a global min.

Proof: If this holds, by property ④

$$f(y)\ge f(x^*)+\nabla f(x^*)(y-x^*)\ge f(x^*), \forall y$$

Corollary: if $\nabla f(x^*)=0 \Rightarrow x^*$ global min.

Review of linear algebra:

1). Transpose of matrix

$$(AB)^T=B^TA^T$$

2). linear dependent/in-

basis ...

3). determinants, eigenvalues, eigenvectors

$$d\begin{pmatrix} a & b \\ c & d\end{pmatrix}=ad-bc=\begin{vmatrix} a & b \\ c & d\end{vmatrix}$$

$$\det\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33}\end{pmatrix}=a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33}\end{vmatrix}-a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33}\end{vmatrix}+a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32}\end{vmatrix}$$

Property:

$A$ not invertible $\Leftrightarrow \det A=0$

$\Leftrightarrow \exists$ some nonzero vector $v$ s.t. $Av=0$

A number $\lambda$ and nonzero vector $v$ are eigenvalue & eigenvector if

$$Av=\lambda v$$

equivalently $(A-I\lambda)v=0$ where $I$ = Identity matrix = $\begin{pmatrix} 1 & & ? \\ & 1 & \\ ? & & 1\end{pmatrix}$

Combing above:

$\lambda$ is an eigenvalue $\Leftrightarrow \det(A-\lambda I)=0$

n-th-order polynomial in variable $\lambda$.

---

Strategy for finding e-values/vectors.

① compute polynomial $\det(A-\lambda I)$

② find roots $\lambda_1,\cdots,\lambda_k$ there are eigenvalues

③ For each $\lambda_i$, find all solutions of $Av_i=\lambda_i v_i$.

Def: A matrix $S$ is symmetric if $S^T=S$. as $S=\nabla^2 f(x)$

IMPORTANT FACT:

- if $S$ symmetric, then all eigenvalues are real
- there is an orthonormal basis of eigenvectors.

E.g. $S=\begin{pmatrix} 5 & -1 & -1 \\ -1 & 5 & -1 \\ -1 & -1 & 5 \end{pmatrix}$

Find e-values/vectors

eigenvalues: $\det(S-\lambda I)=(5-\lambda)^3-(5-\lambda)=-(\lambda-3)(\lambda-6)^2$

Find eigenvectors:

i. $\lambda=3$. must solve $(S-3I)v=0$

i.e. $\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}=\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

$v_1=v_2=v_3=1$

so $v=c\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ where $c$ is a constant

ii. $\lambda=6$

$S-6I=\begin{pmatrix} -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \end{pmatrix}=-\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$

$v=\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$ is an eigenvector if $v_1+v_2+v_3=0$

i.e. $v$ orthogonal to $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

So any vector $v$ s.t. $v_1+v_2+v_3=0$ is an eigenvector when $\lambda=6$.

But I have to think a bit to find orthogonal eigenvectors.

e.g. $v=\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$ & $\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$ are e-vectors

but not orthogonal.

But $\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$ & $\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$ are orthogonal

---

orthonormal basis of eigenvectors

$\frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, -\frac{1}{\sqrt{6}}\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$

Def: A symmetric matrix $S$ is positive semi-definite if $v^TSv\geq 0$ for all $v\in E^n$.

positive definite if $v^TSv>0$ for all nonzero $v\in E^n$.

Fact:

- $S$ positive semi-definite $\Leftrightarrow$ all eigenvalues $\geq 0$
- $S$ positive definite $\Leftrightarrow$ all eigenvalues $>0$

Why?: Fix $S$ symmetric & let $\lambda_1,\cdots,\lambda_n$ eigenvalues, and $w_1,\cdots,w_n \Rightarrow$ orthnormal basis of eigenvectors.

i.e. $w^T w_k = \delta_{jk} = \begin{cases} 1 & \text{if } j=k \\ 0 & \text{if not} \end{cases}$

- Any vector $v$ = a linear combination of $w_1, \cdots, w_n$.

  e.g. $v = a_1 w_1 + \cdots + a_n w_n$

  so $v^T v = (a_1 w_1^T + \cdots + a_n w_n^T)(a_1 w_1 + \cdots + a_n w_n) = a_1^2 + \cdots + a_n^2.$

  and $v^T S v = (a_1 w_1^T + \cdots + a_n w_n^T) S (a_1 w_1 + \cdots + a_n w_n)$

  $= (a_1 w_1^T + \cdots + a_n w_n^T)(a_1 S w_1 + \cdots + a_n S w_n)$

  $= (a_1 w_1^T + \cdots + a_n w_n^T)(a_1 \lambda_1 w_1 + \cdots + a_n \lambda_1 w_n)$

  $v^T S v = \lambda_1 a_1^2 + \cdots + \lambda_n a_n^2$

Recall last week

To find optimal polynomial approximation to a function in $g$, given $g(x_1), \cdots, g(x_m)$, we tried to minimize

$$f(a) = \sum_{k=1}^{n} (a_0 + a_1 x_k + \cdots + a_n x_k^n - g(x_k))^2$$

we rewrite as $f(a) = a^T Q a - 2b^T a - c$ for some matrix $Q$, vector $b$, number $c$

Q: is this function convex?