Q: Does f strictly convex correspond to H p.d.?

A: Not quite:

~~eg: $f(x) = x^4$, so f strictly convex~~
~~Then $D^2 f(x) = 12x^2$, so $D^2 f(0) = 0$, not p.d.~~

eg $f(x_1, x_2) = x_1^4 + x_2^2$. So f is strictly convex

Then $D^2 f(x) = \begin{pmatrix} 12 x_1^2 & 0 \\ 0 & 2 \end{pmatrix}$

Not p.d. when $x_1 = 0$ – $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ is evec with 0 eval.

We do have

Theorem ~~H p.d. $\Rightarrow$ f strictly convex~~
$C \subseteq \mathbb{R}^n$ convex, nonempty. $f : C \to \mathbb{R}$, smooth.

If ~~the~~ H p.d. for all $x \in C$, f is strictly convex.

Proof Very similar to previous theorem //.

<u>Theorem</u> $C \subseteq \mathbb{R}^n$, nonempty, convex. $f: C \to \mathbb{R}$.

Consider the problem
$$\min f(x) \quad \text{s.t. } x \in C.$$

~~#~~ Let $\bar{x} \in C$ be a local optimum.

i) if f is convex, then $\bar{x}$ is a ~~go~~ global optimal solution.

ii) if f is strictly convex, then $\bar{x}$ is the unique global optimal solution.

<u>Proof</u> i) Assume f convex:

$\bar{x}$ is a local optimal solution $\therefore \exists\, \varepsilon > 0$ s.t. ~~A~~

$$x \in N_\varepsilon(\bar{x}) \Rightarrow f(x) \geq f(\bar{x}).$$

Assume $\bar{x}$ is not global optimum, so $\exists\, \hat{x}$ s.t. $f(\hat{x}) \overset{2}{<} f(\bar{x})$

Since f convex,
$$f(\lambda \bar{x} + (1-\lambda)\hat{x}) \overset{1}{\leq} \lambda f(\bar{x}) + (1-\lambda) f(\hat{x}) \quad \forall \lambda \in (0,1)$$
$$\overset{2}{<} \lambda f(\bar{x}) + (1-\lambda) f(\bar{x})$$
$$= f(\bar{x}) \qquad \forall \lambda \in (0,1).$$

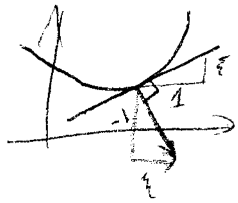
By taking λ close enough to one we get a contradiction.

ii) First $\overset{1}{\text{ineq}}$ becomes strict, second $\overset{2}{}$ becomes $\leq$, so get

$$f(\hat{x}) \leq f(\bar{x}) \text{ and } f(\lambda \bar{x} + (1-\lambda)\hat{x}) < f(\bar{x}).$$

Subgradients after differentiable convex fcns

p34 Definition of subgradient. Draw picture



p35 Definition of supporting hyperplane

p36 f convex, $x \in \mathrm{int}(C)$ $\Rightarrow \exists\, \xi_{\bar{x}}$ s.t. $f(x) \geq f(\bar{x}) + \xi_{\bar{x}}^T(x-\bar{x}) \quad \forall x \in C$.

Idea of proof: Separate $(\bar{x}, f(\bar{x}))$ from epi(f).

p38 $\exists$ subgradient $\forall x \in \mathrm{int}(C) \Rightarrow f$ convex on int C.

p40 f differentiable at $\bar{x}$ $\Rightarrow$ only subgrad at $\bar{x}$ is $\nabla f(\bar{x})$.

p48 $\min_{x \in C} f(x)$ $\quad \bar{x}$ optimal $\Leftrightarrow \exists\, \xi$ s.t. $f(x) \geq f(\bar{x}) + \xi^T(x-\bar{x})$
$\xi^T(x-\bar{x}) \geq 0 \quad \forall x \in C$.

Proof: Involved. See text. (Separate two convex sets: epigraph, and better points.

p50 Corollary 1. C open, $\bar{x}$ optimal $\Leftrightarrow \exists$ zero subgradient of f at $\bar{x}$.

34

Subgradients.

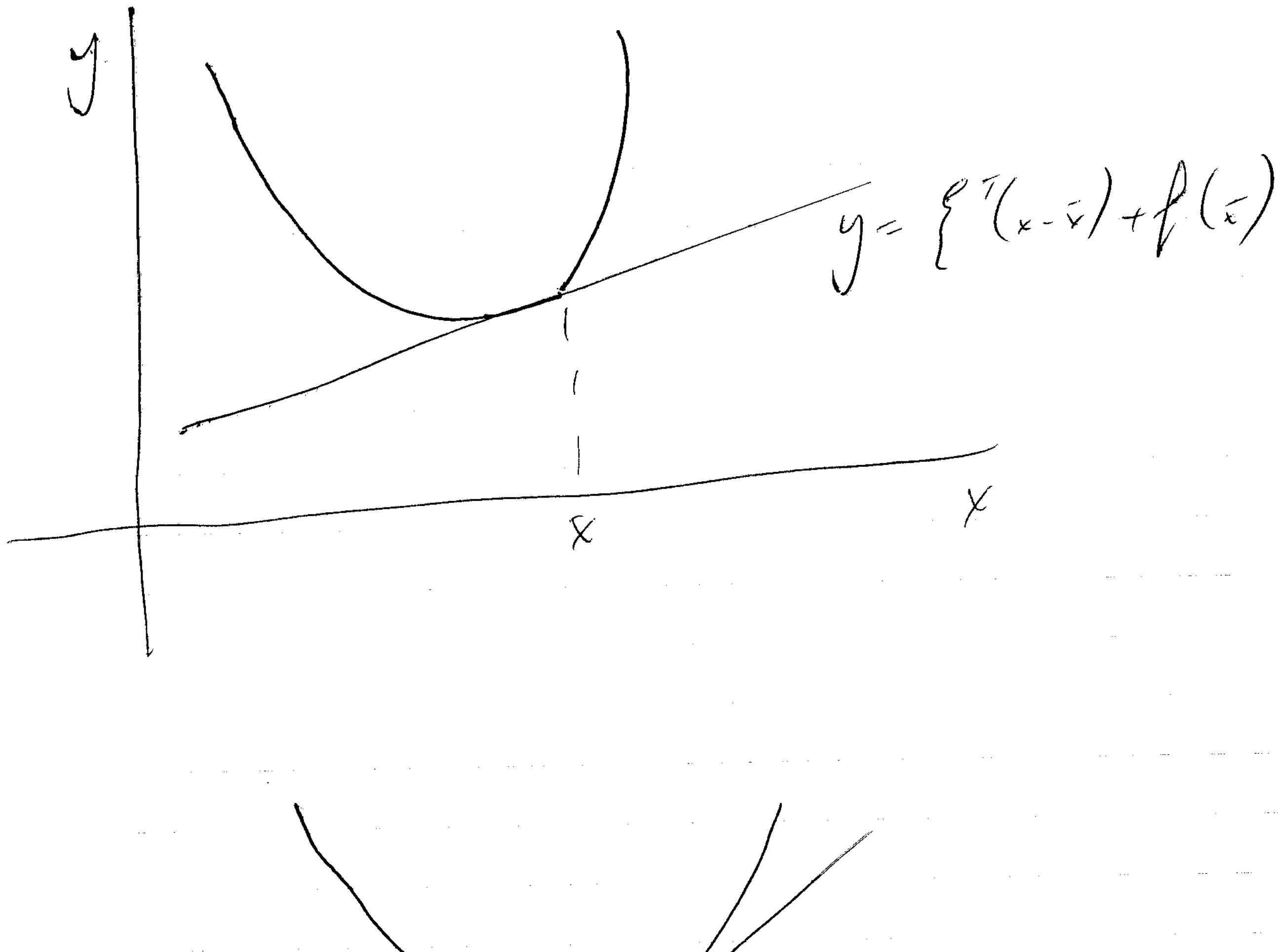
Defn Let C be a nonempty convex set in $\mathbb{R}^n$

Let $f: C \to \mathbb{R}^1$ be convex.

ξ is called a <u>subgradient</u> of f at $\bar{x} \in C$ if

$$f(x) \geq \xi^T(x - \bar{x}) + f(\bar{x}) \quad \forall x \in C.$$

Picture:



Defn: Let S be a nonempty set in E_n, and let $\bar{x} \in \partial S$. A hyperplane $H = \{x : p^T(x-\bar{x}) = 0\}$ is called a <u>supporting hyperplane</u> of S at $\bar{x}$ if ~~either~~ $S \subset H^+$, i.e., ~~said that~~ $p^T(x-\bar{x}) \geq 0$ for each $x \in S$. ~~or~~ ~~else $S \subset H^-$~~

If $S \not\subset H$, H is called a <u>proper</u> supporting hyperplane.

<u>Theorem</u> Let C be a nonempty convex set in $\mathbb{R}^n$, and let $f: C \to \mathbb{R}$ be convex. Let $\bar{x} \in \operatorname{int} C$. Then $\exists$ a vector $\xi \in \mathbb{R}^n$ such that the hyperplane

$$H = \{(x,y) : y = f(\bar{x}) + \xi^T(x-\bar{x})\}$$

supports epi f at $[\bar{x}, f(\bar{x})]$ ~~ie~~. In particular,

$$f(x) \geq f(\bar{x}) + \xi^T(x-\bar{x}) \quad \forall x \in C.$$

<u>Proof</u> epi f is convex. $[\bar{x}, f(\bar{x})] \in \partial(\text{epi } f)$.

Therefore we can <u>separate</u> $[\bar{x}, f(\bar{x})]$ from C, ~~ie~~

that is, $\exists (\xi_0, \mu) \neq 0$ such that

$$(*) \qquad \xi_0^T(x-\bar{x}) + \mu(y - f(\bar{x})) \leq 0 \quad \forall (x,y) \in \text{epi } f.$$

We must have $\mu \leq 0$ - consider large y.

Suppose $\mu = 0$:

Then $\xi_0^T(x-\bar{x}) \leq 0 \quad \forall x \in S$.

Now $\bar{x} \in \operatorname{int} S$. Therefore $\exists \lambda > 0$ s.t. $\bar{x} + \lambda\xi_0 \in S$

~~But~~ So, $\lambda \xi_0^T \xi_0 \leq 0 \Rightarrow \xi_0 = 0$ $\#$ to fact that $(\xi_0, \mu) \neq 0$.

So $\mu < 0$.

So divide through by μ in $(*)$ and rearrange:

(Let $\xi := \frac{1}{\mu}\xi_0$)

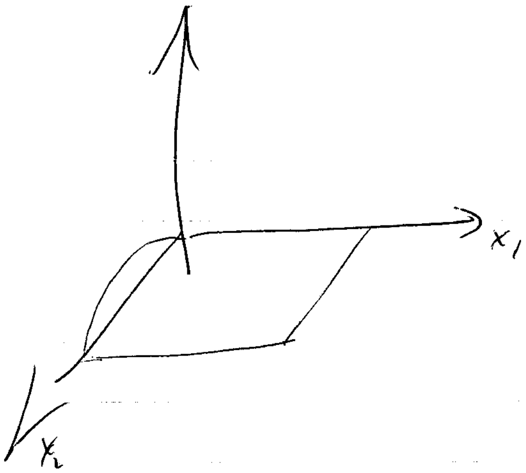
$$y \geq f(\bar{x}) + \xi^T(x - \bar{x}) \quad \forall (x,y) \in \text{epi } f.$$

//.

~~Thus $f(x) \geq f$~~

Converse not true in general:

$$C = \{(x_1, x_2) : 0 \leq x_1, x_2 \leq 1\}$$



$$f(x) = \begin{cases} 0 & x_1 > 0 \\ x_2(1-x_2) & x_1 = 0 \end{cases}$$

Theorem Let C be a nonempty convex set in $\mathbb{R}^n$, and let $f: C \to \mathbb{R}$. Suppose for each $\bar{x} \in \operatorname{int} C$, $\exists$ subgradient vector ξ such that

$$f(x) \geq f(\bar{x}) + \xi^T (x - \bar{x}) \quad \text{for each } x \in C.$$

Then, f is convex on $\operatorname{int} C$.

Pf ~~Let C be a nonempty convex set in $\mathbb{R}^n$, and let $f: C \to \mathbb{R}$. Suppo~~ (same proof as '[illegible] $\Leftarrow$ [illegible]

Let $x_1, x_2 \in \operatorname{int} S$. int S is convex,

so $\lambda x_1 + (1-\lambda) x_2 \in \operatorname{int} S$.

By assumption, $\exists$ subgradient ξ of f at $\lambda x_1 + (1-\lambda) x_2$.

So we have

$$f(x_1) \geq f(\lambda x_1 + (1-\lambda) x_2) + (1-\lambda) \xi^T (x_1 - x_2)$$

$$f(x_2) \geq f(\lambda x_1 + (1-\lambda) x_2) + \lambda \xi^T (x_2 - x_1).$$

Multiplying ineqs by λ and $1-\lambda$ resp, we obtain:

$$\lambda f(x_1) + (1-\lambda) f(x_2) \geq f(\lambda x_1 + (1-\lambda) x_2) \quad //.$$

Lemma $C \subseteq \mathbb{R}^n$ convex, nonempty

$f : C \to \mathbb{R}$ convex, f diff at $\bar{x} \in \text{int } S$.

Then the collection of subgradients of f at $\bar{x}$ is the singleton set $\{\nabla f(\bar{x})\}$.

Pf f has some subgradient at $\bar{x}$. (skip this)

Let ξ be a subgradient of f at $\bar{x}$.

For λ suff small, for any d,

$$f(\bar{x}+\lambda d) \geq f(\bar{x}) + \lambda \xi^T d$$

$$f(\bar{x}+\lambda d) = f(\bar{x}) + \lambda \nabla f(\bar{x})^T d + \lambda \|d\| \alpha(\bar{x}; \lambda d)$$

Thus,
$$0 \geq (\xi - \nabla f(\bar{x}))^T d - \|d\| \alpha(\bar{x}, \lambda d) \quad \text{for } \lambda > 0, \ \lambda \text{ suff small}$$

So $0 \geq (\xi - \nabla f(\bar{x}))^T d$ (limit as $\lambda \to 0$).

True $\forall d$. So choose $d = \xi - \nabla f(\bar{x})$.

Thus $\xi = \nabla f(\bar{x})$. //

Proof of Cor L, without using subgradients

$\bar{x}$ optimal

$$\Leftrightarrow \nabla f(\bar{x})^T (x - \bar{x}) \geq 0 \quad \forall x \in C.$$

Know $\quad f(x) \geq f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x})$.

$$f(x) = f(\bar{x}) + \underbrace{\nabla f(\bar{x})^T (x - \bar{x})}_{\text{this } < 0} + \|x - \bar{x}\| \alpha(x; x - \bar{x})$$

$\Rightarrow f(x) < f(\bar{x})$ for x close enough.

$\Rightarrow$ x not optimal.

So $\quad x$ optimal $\Rightarrow \nabla f(\bar{x})^T (x - \bar{x}) \geq 0 \quad \forall x \in C$.

Conversely:

Assume $\quad \nabla f(\bar{x})^T (x - \bar{x}) \geq 0 \quad \forall x \in C$.

Then $\quad f(x) \geq f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x})$

$\quad \geq f(\bar{x})$.

Thus, $\bar{x}$ optimal.

Theorem $f : \mathbb{R}^n \to \mathbb{R}$, convex, $C \subseteq \mathbb{R}^n$ convex

Consider the convex program $\min f(x)$ s.t. $x \in C$

$\bar{x} \in C$ is an optimal solution $\Longleftrightarrow$

$\exists$ subgradient ξ s.t.

$$f(x) \geq f(\bar{x}) + \xi^T(x - \bar{x}) \quad \forall x \in \mathbb{R}^n$$

$$\xi^T(x - \bar{x}) \geq 0 \quad \forall x \in C.$$

Proof $\Longleftarrow$: Assume we have a subgradient.

So $f(x) \geq f(\bar{x}) + \xi^T(x - \bar{x}) \quad \forall x \in \mathbb{R}^n$

$\geq f(\bar{x}) \quad \forall x \in C$ //

$\Longrightarrow$: ~~Assume f convex~~

Let $\bar{x}$ be an optimal solution.

Construct the following two sets:

$$\Lambda_1 = \{(x - \bar{x}, y) : x \in \mathbb{R}^n, y > f(x) - f(\bar{x})\}$$

$$\Lambda_2 = \{(x - \bar{x}, y) : x \in C, y \leq 0\}$$

Λ_1 and Λ_2 are both convex, and $\Lambda_1 \cap \Lambda_2 = \phi$. (since $\bar{x}$ optimal)

We can separate Λ_1 and Λ_2, so $\exists (\xi_0, \mu) \neq 0$

and scalar α such that

(1) $\xi_0^T(x-\bar{x}) + \mu y \leq \alpha \qquad x \in \mathbb{R}^n,\ y > f(x) - f(\bar{x})$

(2) $\xi_0^T(x-\bar{x}) + \mu y \geq \alpha \qquad x \in C,\ y \leq 0.$

$x = \bar{x},\ y = 0 \Rightarrow \alpha \leq 0$ from (2).

$x = \bar{x},\ y = \varepsilon > 0$ in (1) $\Rightarrow \alpha \geq \mu\varepsilon$

True $\forall \varepsilon > 0$, so $\mu \leq 0$, ~~and $\alpha \geq 0$~~

Also, get $\alpha \geq 0$ so $\alpha = 0$.

$\mu = 0$: Consider (1): Get $\xi_0^T(x-\bar{x}) \leq 0 \quad \forall x \in \mathbb{R}^n$

Setting $x = \bar{x} + \xi_0 \Rightarrow \xi_0^T\xi_0 \leq 0 \Rightarrow \xi_0 = 0$ $\#$

(to $(\xi_0, \mu) \neq 0$)

So $\mu < 0$.

Let $\xi := -\xi_0/\mu$, divide (1) and (2) by μ:

(1') $y \geq \xi^T(x-\bar{x}) \qquad x \in \mathbb{R}^n,\ y > f(x) - f(\bar{x})$

(2') $y \leq \xi^T(x-\bar{x}) \qquad x \in C,\ y \leq 0$

Let $y = 0$ in (2'): $\xi^T(x-\bar{x}) \geq 0 \quad \forall x \in C.$

From (1'): $f(x) \geq f(\bar{x}) + \xi^T(x-\bar{x}) \quad \forall x \in \mathbb{R}^n$ //.