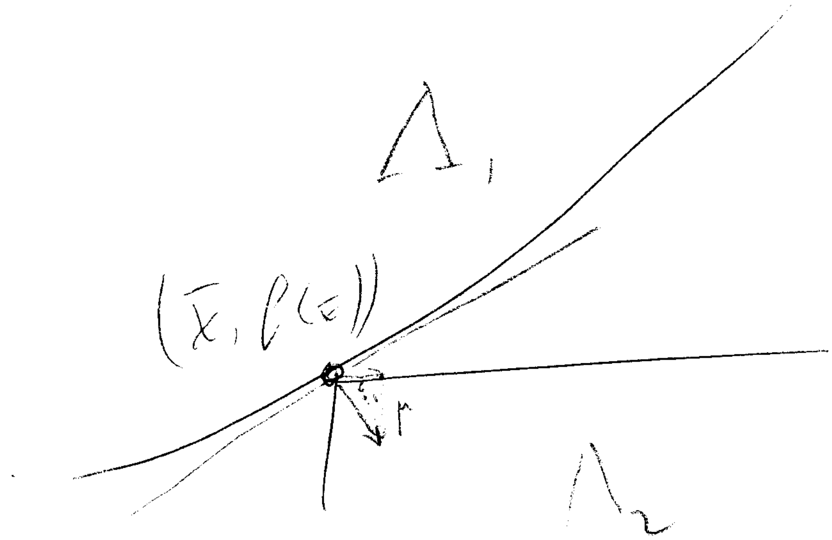


Proof without translation:

Want:

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

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

$$\Lambda_1 = \{ (x,y) : x \in \mathbb{R}^n, \; y > f(x) \}$$
$$\Lambda_2 = \{ (x,y) : x \in C, \; y \leq f(\bar{x}) \}$$

Convex, [illegible] Λ_1 [illegible]

$$\xi_0^T x + \mu y \leq \alpha \qquad (x,y) \in \Lambda_1$$
$$\xi_0^T x + \mu y \geq \alpha \qquad (x,y) \in \Lambda_2.$$

$(\bar{x}, y) \in \Lambda_2$, so $\alpha \leq \xi_0^T \bar{x} + \mu f(\bar{x})$.

$(\bar{x}, y)$ ~~not large~~ in Λ_1 ~~so $y\mu \leq 0$~~ for all $y > f(\bar{x})$

so $\alpha \geq \xi_0^T \bar{x} + \mu y$ So $\alpha = \xi_0^T \bar{x} + \mu f(\bar{x})$,

and $\mu \leq 0$.

If $\mu = 0$: $\xi_0^T x \leq \xi_0^T \bar{x}$ $\forall x$ in $\mathbb{R}^n$. $\therefore \xi_0 = 0$.

So $\mu < 0$.

Let $\xi = -\dfrac{\xi_0}{\mu}$ $\qquad \xi^T x - y \leq \xi^T \bar{x} - f(\bar{x})$ ie $\xi^T(x-\bar{x}) + f(\bar{x}) \leq f($

$\forall x \in \mathbb{R}^n$

Also $\qquad \xi^T x - y \geq \xi^T \bar{x} - f(\bar{x})$ ie $\xi^T(x-\bar{x}) \geq y - f(x)$

$\forall (x,y)$ in $C \Rightarrow \xi^T(x-\bar{x}) \geq ($

Corollary 1 If C open, $\bar{x}$ is an optimal solution

$\iff \exists$ zero subgradient of f at $\bar{x}$.

In particular, if $C = \mathbb{R}^n$, then $\bar{x}$ is a global minimum if and only if there exists a zero subgradient of f at $\bar{x}$.

Proof: All directions out of $\bar{x}$ are valid, so $\xi^T d \geq 0 \ \forall d$

$\Rightarrow \xi = 0$. //

Corollary 2 If in addition, f is differentiable, $\bar{x}$ is an optimal soln

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

Furthermore, if C is open then $\bar{x}$ is an optimal soln if and only if $\nabla f(\bar{x}) = 0$.

Lemma Let $f: C \to \mathbb{R}$ be convex on the convex set $C \subseteq \mathbb{R}^n$. Let $\bar{x} \in C$. The set of subgradients of f at $\bar{x}$ is a convex set.

Proof Let ξ^1, ξ^2 be subgradients to f at $\bar{x}$

Let $\xi = \lambda \xi^1 + (1-\lambda)\xi^2, \quad 0 \leq \lambda \leq 1$

Then $\forall x \in C$, $f(\bar{x}) + \xi^T(x - \bar{x}) = f(\bar{x}) + \lambda(\xi^1)^T(x-\bar{x}) + (1-\lambda)(\xi^2)^T(x-\bar{x})$

$= \lambda\left(f(\bar{x}) + (\xi^1)^T(x-\bar{x})\right) + (1-\lambda)\left(f(\bar{x}) + (\xi^2)^T(x-\bar{x})\right)$

$\leq \lambda f(x) + (1-\lambda) f(x)$ since ξ^1, ξ^2 are subgradients

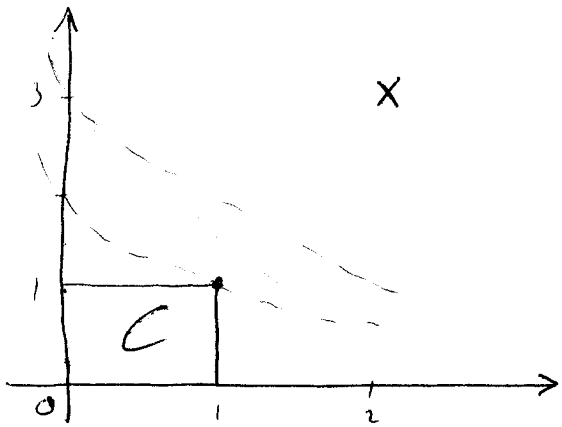
$= f(x)$. True $\forall x \in C$. $\square$

Eg: $\min \max\{x^2+1, 3-x\}$, $x \in \mathbb{R}$.
At $x=1$, set [illegible] is $[-1,2] \ni 0$.

Example.

$$\min \quad (x_1-2)^2 + (x_2-3)^2$$

$$\text{s.t.} \quad \begin{aligned} x_1 &\leq 1 \\ x_2 &\leq 1 \\ x_1 &\geq 0 \\ x_2 &\geq 0 \end{aligned}$$



$$\nabla f = \begin{pmatrix} 2(x_1-2) \\ 2(x_2-3) \end{pmatrix}$$

$$\nabla f\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}-2\\-4\end{pmatrix}$$

$$\nabla f\begin{pmatrix}0\\0\end{pmatrix} = \begin{pmatrix}-4\\-6\end{pmatrix} \qquad \nabla f\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}-2\\-6\end{pmatrix}$$

$\nabla f\begin{pmatrix}1\\1\end{pmatrix}$ makes positive inner product with any feasible direction out of $\begin{pmatrix}1\\1\end{pmatrix}$.

$\nabla f\begin{pmatrix}0\\0\end{pmatrix}$ makes negative inner product with some feasible directions.

Eg A nondifferentiable function, with a subgradient ξ satisfying $\xi^T(x-\bar{x}) \geq 0 \ \forall x \in X$ at the optimal $\bar{x}$:

$$\min \quad |x_1 - x_2| + \tfrac{1}{2}(x_1 + x_2 + \tfrac{1}{2})^2$$

$$\text{s.t.} \qquad x_i \geq 0, \ i = 1, 2.$$

Optimal soln is the origin.

As $x_1 - x_2 \to 0_+$, $\quad \nabla f(x) = \begin{bmatrix} 1 + x_1 + x_2 + \frac{1}{2} \\ -1 + x_1 + x_2 + \frac{1}{2} \end{bmatrix}$

As $x_1 - x_2 \to 0_-$, $\quad \nabla f(x) = \begin{bmatrix} -1 + x_1 + x_2 + \frac{1}{2} \\ 1 + x_1 + x_2 + \frac{1}{2} \end{bmatrix}$.

So, two subgradients of f at the origin are:

$$\xi^1 = \begin{bmatrix} 3/2 \\ -\frac{1}{2} \end{bmatrix}, \quad \xi^2 = \begin{bmatrix} -\frac{1}{2} \\ 3/2 \end{bmatrix}$$ (by evaluating the gradient at the limits)

Let $\xi = \frac{1}{2}\xi^1 + \frac{1}{2}\xi^2 = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$.

Then $\xi^T(x - \bar{x}) \geq 0 \ \forall x \in X$, since $\bar{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $x \geq 0$.

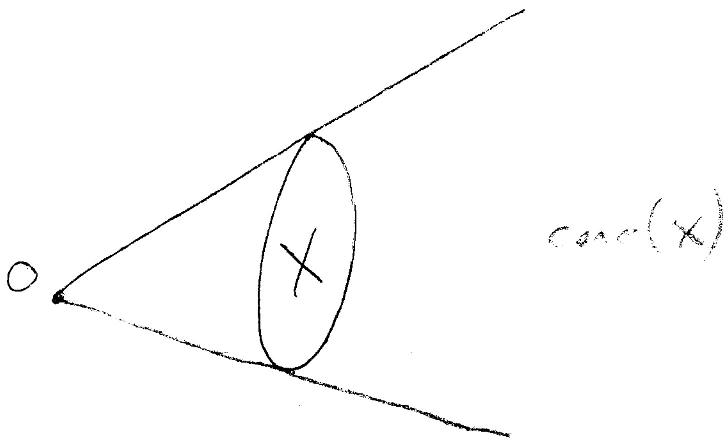
Convex Cones

Lemma Assume X is a convex set. Then the set

$$\text{cone}(X) := \{\gamma x : x \in X,\ \gamma \geq 0\}$$

is a convex cone.

Proof See text. $\boxed{/\!/}$



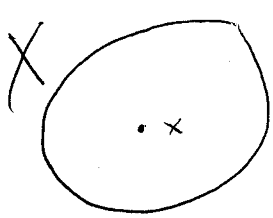
cone(X) is the cone generated by X.

Given a convex set X and a point $x \in X$:

Defn The cone of feasible directions of X at x (or the radial cone) is $K_X(x) := \text{cone}(X - x)$. (It is a convex cone.)

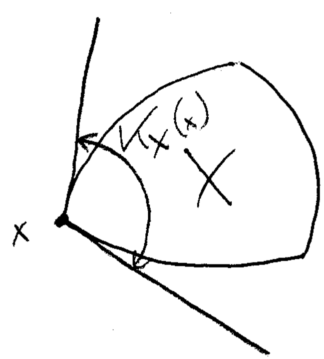
Eg:

(i)



Cone is everything

(ii)



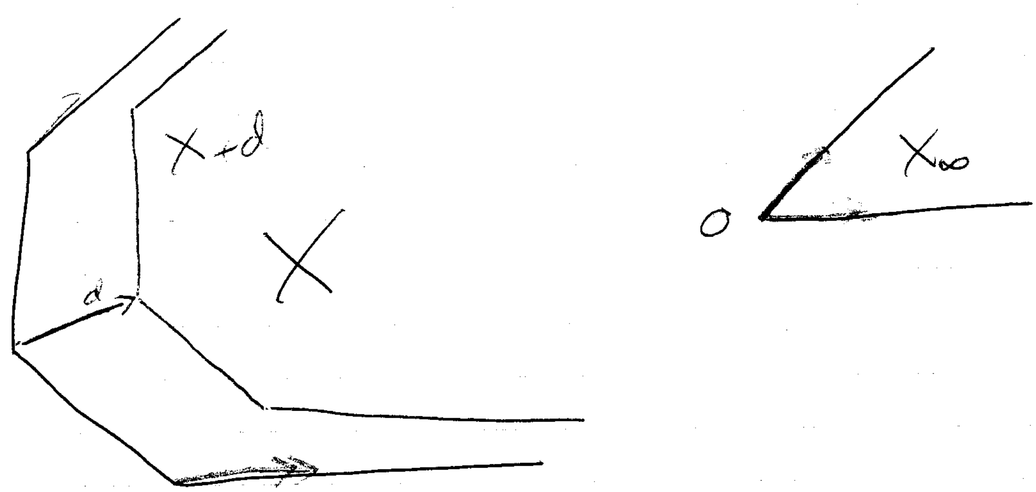
DEFN Let $X \subseteq \mathbb{R}^n$ be a convex set. The set

$$X_\infty := \{ d \in \mathbb{R}^n : X + d \subseteq X \}$$

is the RECESSION CONE of X. (It is a convex cone.)

These are the directions you can go in forever from any point in X.

Eg:



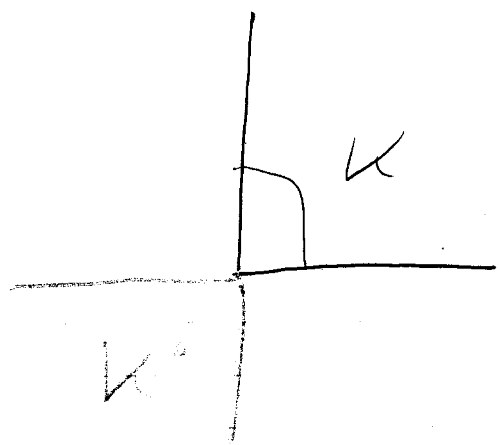
<u>DEFN</u> Let K be a cone in $\mathbb{R}^n$. The set

$$K^\circ := \{y \in \mathbb{R}^n : y^T x \le 0 \ \forall x \in K\}$$

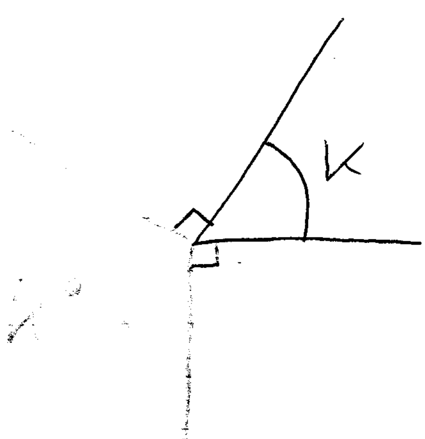
is the POLAR CONE of K.

<u>Eg</u>: (i) $K = \mathbb{R}^n_+$.

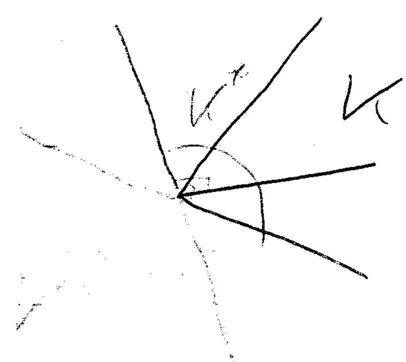
Then $K^\circ = \{y \in \mathbb{R}^n : y \le 0\}$.



(ii)



(iii)



<u>DEFN</u> The negative of the polar cone is $K^* := -K^\circ$ and is the DUAL CONE

Eg: Let C be a convex set, and let $\bar{x} \in C$.

Our optimality conditions for $\min f(x)$ s.t. $x \in C$, f convex

involved finding a subgradient satisfying $\xi^T(x-\bar{x}) \geq 0 \quad \forall x \in C$.

$$\text{Now, } K_C(\bar{x}) = \{\gamma d : d \in X - \bar{x}, \gamma \geq 0\}$$

cone of feasible directions.

$$= \{\gamma d : d = x - \bar{x}, \gamma \geq 0, x \in C\}$$

$$= \{\gamma(x-\bar{x}) : \gamma \geq 0, x \in C\}$$

So requiring $\xi^T(x-\bar{x}) \geq 0 \quad \forall x \in C$ is equivalent to

requiring $\xi^T d \geq 0 \quad \forall d \in K_C(\bar{x})$

So $-\xi$ is in the polar cone to the cone of feasible directions.

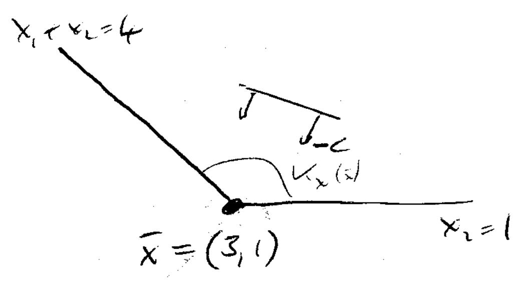
Or, ξ is in the dual cone to the cone of feasible directions.

<u>DEFN</u> Given a convex set X and a point $x \in X$, the NORMAL CONE to X at x is $N_X(x) := [\text{cone}(X-x)]^\circ$, the polar cone to the cone of feasible directions.

Eg: Look at an LP:

$$\begin{aligned} \min \quad & x_1 + 2x_2 \\ \text{s.t.} \quad & x_1 + x_2 \geq 4 \\ & \cancel{x_1}\ x_2 \geq 1 \end{aligned}$$

$X = \{x \in \mathbb{R}^2 : x_1 + x_2 \geq 4, x_2 \geq 1\}$



$$K_X(\bar{x}) = \{d : d_1 - d_2 \geq 0, d_2 \geq 0\}$$

Polar cone: $K_X^\circ(\bar{x}) = \{d \in \mathbb{R}^2 : d_1 \leq 0, d_1 + d_2 \geq 0\}$

$$= \{d \in \mathbb{R}^2 : d = \gamma_1 \begin{bmatrix} 0 \\ -1 \end{bmatrix} + \gamma_2 \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \gamma_1, \gamma_2 \geq 0\}$$

$\bar{x}$ is optimal if $-c = -\nabla f(\bar{x}) \in K_X^\circ(\bar{x})$, ie, $-c = \gamma_1 \begin{bmatrix} 0 \\ -1 \end{bmatrix} + \gamma_2 \begin{bmatrix} -1 \\ -1 \end{bmatrix}$,

or $\gamma_1 \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \gamma_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\gamma_1, \gamma_2 \geq 0$. Exactly LP optimality conditions,

<u>THEOREM</u> For every ~~cl~~ convex cone $K \subseteq \mathbb{R}^n$, the polar cone is closed and convex, and $K^\circ = (cl(K))^\circ$. Further, if K is a closed convex cone then $K^{\circ\circ} = K$.

<u>Proof</u>: See text. $\square$

<u>FARKAS LEMMA</u>

Let A be $m \times n$ matrix and let $K = \{x \in \mathbb{R}^n : Ax \leq 0\}$

Then $K^\circ = \{y \in \mathbb{R}^n : y = A^T\lambda, \lambda \in \mathbb{R}^m, \lambda \geq 0\}$.

Equivalently, exactly one of the following two systems has a solution:

(i) $Ax \leq 0$ and $c^T x > 0$
(ii) $A^T\lambda = c, \quad \lambda \geq 0$.

<u>Proof</u> See text. $\square$

Why are these two formulations equivalent?

$y \in K^\circ \Leftrightarrow$ ~~$y = A^T$~~

$\Leftrightarrow c$ such that ~~2~~ (i) holds.

Also, $y \in K^\circ \Leftrightarrow y = A^T\lambda, \lambda \geq 0$, and $y^T x \leq 0 \ \forall x \in K$
$\Leftrightarrow$ (ii) holds, (i) doesn't.

$\underbrace{y \notin K^\circ}_{\text{(ii) doesn't hold, with } y = c.} \Leftrightarrow \underbrace{\exists x \in K \text{ with } x^T y > 0}_{\text{(i) holds, with } c = y}$

Example (Ex 2.30, page 31)

Arbitrage:

Stocks with costs: $c_1 = 2 \quad c_2 = 3 \quad c_3 = 6$

Possible scenarios for final value of stocks:

Scenario 1: $a_{11} = 5, \quad a_{12} = 1, \quad a_{13} = 3$

Scenario 2: $a_{21} = 1, \quad a_{22} = 7, \quad a_{23} = 7$

Can we buy (or sell short) the stocks so that we are guaranteed a profit?

borrow stock & sell at c_j, buy [illegible] [illegible] borrowed from

Let x_j = amount purchased of stock j, $j = 1, 2, 3$.
x_j can be positive or negative (or zero).

We pay $\sum_j c_j x_j$

Return under scenario 1: $\sum_j a_{1j} x_j$

Return under scenario 2: $\sum_j a_{2j} x_j$

We'd like
$$\left.\begin{array}{l} \sum_j c_j x_j < 0 \\ \sum_j a_{ij} x_j \geq 0 \quad \text{for } i = 1, 2. \end{array}\right\} \quad \begin{array}{l} c^T x < 0 \\ Ax \geq 0. \end{array}$$

Here:
$$\begin{aligned} 2x_1 + 3x_2 + 6x_3 &< 0 \\ 5x_1 + x_2 + 3x_3 &\geq 0 \\ x_1 + 7x_2 + 7x_3 &\geq 0 \end{aligned}$$

One soln: $x_1 = 1 \quad x_2 = 1 \quad x_3 = -1.$

Example (eg 2.31, page 31)

$\mathbb{S}^n$: set of symmetric matrices of dimension $n \times n$

$\mathbb{S}^n_+$: set of all psd $n \times n$ matrices.

$\mathbb{S}^n_+$ is a convex cone: exercise.

To define polar cone, need to define inner product

$$\langle A, B \rangle := \text{trace}(AB) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} b_{ij}, \text{ since symmetric.}$$

(Frobenius inner product)

Polar cone:

$$[\mathbb{S}^n_+]^\circ = \{ B \in \mathbb{S}^n : tr(BA) \leq 0 \quad \forall A \in \mathbb{S}^n_+ \}.$$

Then $[\mathbb{S}^n_+]^\circ = -\mathbb{S}^n_+$

Proof: Let $B \in -\mathbb{S}^n_+$, $A \in \mathbb{S}^n_+$. Want to show $tr(BA) \leq 0$.

Let $d^1, \ldots, d^n$ be evecs of A, with evals $\lambda_1, \ldots, \lambda_n$.
norm 1.

Then $A = \sum_{i=1}^n \lambda_i d^i d^{iT}$

So $tr(BA) = tr(B \sum_{i=1}^n \lambda_i d^i d^{iT}) = \sum_{i=1}^n \lambda_i \, tr(B d^i d^{iT})$

$= \sum_{i=1}^n \lambda_i \, tr(d^{iT} B d^i)$

$= \sum_{i=1}^n \lambda_i \, d^{iT} B d^i$

≤ 0 since $-B$ psd.

Thus $[\mathbb{S}^n_+]^\circ \supseteq -\mathbb{S}^n_+$.

NLP 524

Conversely, let $B \in [\mathbb{S}^n_+]^\circ$.

Pick $d \in \mathbb{R}^n$, let $A = dd^T \in \mathbb{S}^n_+$.

Since $B \in [\mathbb{S}^n_+]^\circ$, we have

$$0 \geq tr(BA) = tr(Bdd^T) = tr(d^TBd) = d^TBd$$

Thus $-d^TBd \geq 0$ for any $d \in \mathbb{R}^n$.

So $-B \in \mathbb{S}^n_+$ $\square$