CPSC 542F WINTER 2017

Convex Analysis and Optimization

Lecture Notes

SUMMER RESEARCH INTERNSHIP, UNIVERSITY OF WESTERN ONTARIO

GITHUB.COM/LAURETHTEX/CLUSTERING

This research was done under the supervision of Dr. Pauline Barmby with the financial support of the MITACS Globalink Research Internship Award within a total of 12 weeks, from June 16th to September 5th of 2014.

First release, August 2014

Contents

1. Convex Sets

1.1 Convexity

1.1.1 Cone

Definition 1.1.1 — Cone. A set $K \in \mathbb{R}^n$, when $x \in K$ implies $\alpha x \in K$.

A non convex cone can be hyper-plane.
For convex cone $x + y \in K, \forall x, y \in K$.
Cone don't need to be "pointed". e.g.
Direct sums of cones $C_1 + C_2 = \{x = x_1 + x_2 | x_1 \in C_1, x_2 \in C_2\}$.

■ **Example 1.1** $S_1^n\{X|X = X^n, \lambda(x) \geq 0\}$
A matrix with positive eigenvalues. ■

Operations preserving convexity

Intersection $C \cap_{i \in \mathbb{I}} C_i$
Linear map Let $A : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map. If $C \in \mathbb{R}^n$ is convex, so is $A(C) = \{Ax \forall x \in C\}$
Inverse image $A^{-1}(D) = \{x \in \mathbb{R} | Ax \in D\}$

Operations that induce convexity

Convex hull on $S = \cap\{C|S \in C, Cisconvex\}$

■ **Example 1.2** $Co\{x_1, x_2, \cdots, x_m\} = \{\sum_{i=1}^m \alpha_i x_i | \alpha \in \delta_m\}$ ■

For a convex set $x \in C \Rightarrow x = \sum \alpha_i x_i$.

Theorem 1.1.1 — Carathéodory's theorem. If a point $x \in \mathbb{R}^d$ lies in the convex hull of a set P, there is a subset P' of P consisting of $d + 1$ or fewer points such that x lies in the convex hull of P'. Equivalently, x lies in an r-simplex with vertices in P.

1.2 Convex Functions

Definition 1.2.1 — Convex function. Let $C \in \mathbb{R}^n$ be convex, $f : C \to \mathbb{R}$ is convex on f if $x, y \in C \times C$. $\forall \alpha \in (0,1)$, $f(\alpha x + (1-\alpha)y) \leq f(\alpha x) + f((1-\alpha)y)$

Definition 1.2.2 — Strictly Convex function. Let $C \in \mathbb{R}^n$ be convex, $f : C \to \mathbb{R}$ is strictly convex on f if $x, y \in C \times C$. $\forall \alpha \in (0,1)$, $f(\alpha x + (1-\alpha)y) \langle f(\alpha x) + f((1-\alpha)y)$

Definition 1.2.3 — Strongly convex. $f : C \to \mathbb{R}$ is strongly convex with modules $u \geq 0$ if $f - \frac{1}{2}u||\cdot||^2$ is convex.

Interpretation: There is a convex quadratic $\frac{1}{2}u||\cdot||^2$ that lower bounds f.

■ **Example 1.3** $\min_{x \in C} f(x) \leftrightarrow \min \bar{f}(x)$ Useful to turn this into an unconstrained problem.

$$\bar{f}(x) = \begin{cases} f(x) & if x \in C \\ \infty & elsewhere \end{cases}$$

■

Definition 1.2.4 A function $f : \mathbb{R}^n \to \mathbb{R} \cup \infty\ \bar{\mathbb{R}}$ is convex if $x, y \in \mathbb{R}^n \times \mathbb{R}^n$, $\forall x, y, \bar{f}(\alpha x + (1-\alpha)y) \leq f(\alpha x) + f((1-\alpha)y)$

Definition 1 is equivalent to definition 2 if $f(x) = \infty$.

■ **Example 1.4** $f(x) = \sup_{j \in J} f_j(x)$ ■

1.2.1 Epigraph

Definition 1.2.5 — Epigraph. For $f : \mathbb{R}^n \to \bar{R}$, its epigraph $epi(f) \in \mathbb{R}^{n+1} is the set epi(f) \{(x, \alpha) | f(x) \in \alpha\}$

Next: a function is convex i.f.f. its epigraph is convex.

Definition 1.2.6 A function $f : C \to \mathbb{R}, C \in \mathbb{R}^n$ is convex if $\forall x, y \in C$, $f(ax + (1-a)x) \leq af(x) + (1-a)f(x) \quad \forall a \in (0,1)$.
Strict convex: $x \neq y \Rightarrow f(ax + (1-a)x) \leq af(x) + (1-a)f(x)$

R f is convex $\Rightarrow -f$ is concave.

Level set: $S_\alpha f = \{x | f(x) \leq \alpha\}$.
$S_\alpha f$ is convex $\Leftrightarrow f$ is convex.

Definition 1.2.7 — Strongly convex. $f : C \to \mathbb{R}$ is strongly convex with modules μ if $\forall x, y \in C$, $\forall \alpha \in (0,1)$, $f(ax + (1-a)x) \leq af(x) + (1-a)f(x) - \frac{1}{2\mu}\alpha(1-\alpha)\|x - y\|^2$.

R

- f is 2nd-differentiable, f ix convex $\iff \nabla^2 f(x) \rangle 0$.
- f is strongly convex $\iff \nabla^2 f(x) \rangle \mu I \iff x \geq \mu$

Definition 1.2.8 — 2. $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is convexif $x, y \in \mathbb{R}, \alpha \in (0,1), f(ax+(1-a)x) \leq af(x) + (1-a)f(x)$.

The effective domain of f is $domf = \{x | f(x) \langle +\infty\}$

■ **Example 1.5 — Iudcator function.** $\delta_c(x) = \begin{cases} 0 & x \in C \\ +\infty & elsewhere \end{cases}$.

$dom\delta_c(x) = C$ ■

Definition 1.2.9 — Epigraph. The epigraph of f is $epif = \{(x, \alpha) | f(x) \leq \alpha\}$

The graph of $epif$ is $\{(x, f(x) | x \in domf\}$.

Definition 1.2.10 — III. A function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is convex if $epif$ is convex.

Lemma 1.2.1 $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is convex $\iff$ $\forall x, y \in \mathbb{R}^n, \alpha \in (0,1), f(ax+(1-a)x) \leq af(x) + (1-a)f(x)$.

Proof. $\Rightarrow$ take $x, y \in domf$, $(x, f(x)) \in epif$,$(y, f(y)) \in epif$.
Because $epif$ is convex, $\alpha(x, f(x)) + (1-\alpha)(y, f(y)) \in epif$. $\alpha x + (1-\alpha y), \alpha f(x) + (1-\alpha)f(y) \in epif$ ■

■ **Example 1.6 — Distance.** Distance to a convexset $d_c(x) = \inf_{z\in C}\{\|z-x\|\}$. Take any two sequence $\{y_k\} and \{\bar{y}_k\} \subset C$ s.t. $\|y_k - x\| \to d_c(x)$, $\|\bar{y}_k - \bar{x}\| \to d_c(\bar{x})$. $z_k = \alpha y_k + (1-\alpha)\bar{y}_k$.

$$
\begin{aligned}
d_c(\alpha x + (1-\alpha)\bar{x}) &\leq \|z_k - \alpha x - (1-\alpha)\bar{x}\| \\
&= \|\alpha(y_k - x) + (1-\alpha)(\bar{y}_k - \bar{x})\| \\
&\leq \alpha\|y_k - x\| + (1-\alpha)\|\bar{y}_k - \bar{x}\|
\end{aligned}
$$

Take $k \to \infty$, $d_c(\alpha x + (1-\alpha)\bar{x}) \leq \alpha d(x) + (1-\alpha)d(\bar{x})$ ■

■ **Example 1.7 — Eigenvalues.** Let $X \in S^n := \{n \times n$ symmetric matrix$\}$. $\lambda_1(x) \geq \lambda_2(X) \geq \ldots \geq \lambda_n(x)$.
$f_k(x) = \sum_1^n \lambda_i(x)$.
Equivalent characterization

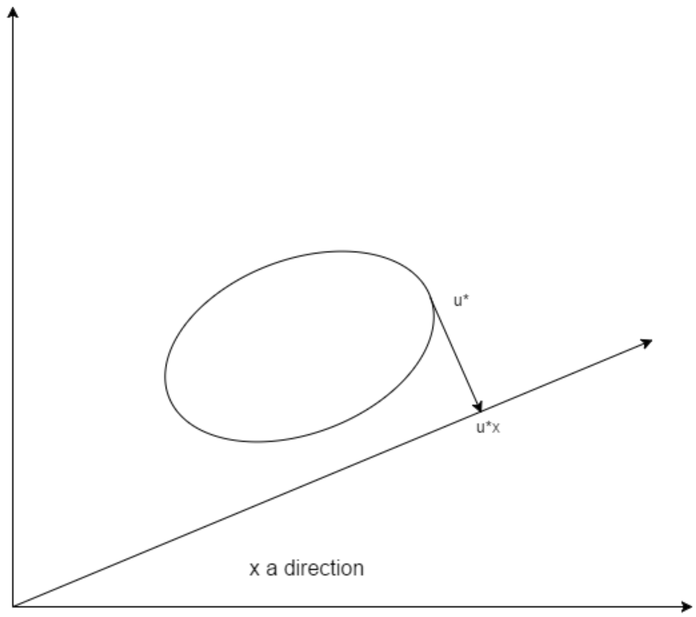
$$
\begin{aligned}
f_k(x) &= \max\{\sum_i v_i^T X v_i | v_i \perp v_j, i \neq j\} \\
&= \max\{tr(V^T X V | V^T V = I_k\}
\end{aligned}
$$

$$\max\{tr(VV^T X)\} \text{by circularity}$$

Note $\langle A, B\rangle = tr(A, B)$ is true for symmetric matrix.
$\langle A, A\rangle = |A|_F^2 = \sum_i A_{ii}^2$ ■

1.3 Support Function

Take a set $C \in \mathbb{R}^n$, not necessarily convex.The support function is $\sigma_C = \mathbb{R}^n \to \bar{\mathbb{R}}$. $\sigma_C(x) =$



$\sum\{\langle x,u\rangle | u \in C\}$.

Fact 1.3.1 The support function binds the supporting hyper-plane.

Supporting functions are

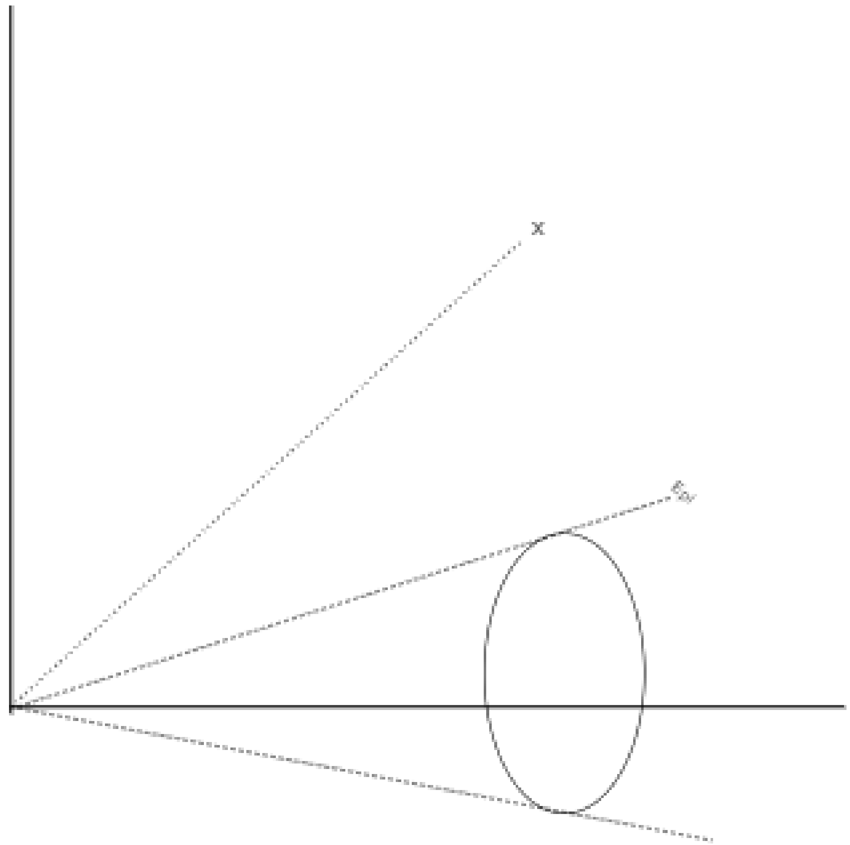
- Positively homogeneous
  $\sigma_C(\alpha x) = \alpha\sigma_C(x)\forall\alpha\rangle 0$
  $\sigma_C(\alpha x) = \sup_{u\in C}\langle \alpha x, u\rangle = \alpha \sup_{u\in C}\langle x,u\rangle = \alpha\sigma_C(x)$
- Sub-linear(a special case of convex, linear combination holds $\forall\alpha$.
  $\sigma_C(\alpha x + (1-\alpha)y) = \sup_{u\in C}\langle \alpha x + (1-\alpha)y, u\rangle \leq \alpha \sup_{u\in C}\langle x,u\rangle + (1-\alpha)\sup_{u\in C}\langle y,u\rangle$

■ **Example 1.8 — L2-norm.** $\|x\| = \sup_{u\in C}\{\langle x,u\rangle, u \in \mathbb{R}^n\}$.
$\|x\|_p = \sup\{\langle x,u\rangle, u \in B_q\}$ where $\frac{1}{p} + \frac{1}{q} = 1$. $B_q = \{\|x\|_q \leq 1\}$.
The norm is

- Positive homogeneous
- sub-linear
- If $0 \in C$, σ_C is non-negative.
- If C is central-symmetric, $\sigma_C(0) = 0$ and $\sigma_C(x) = \sigma_C(-x)$

■

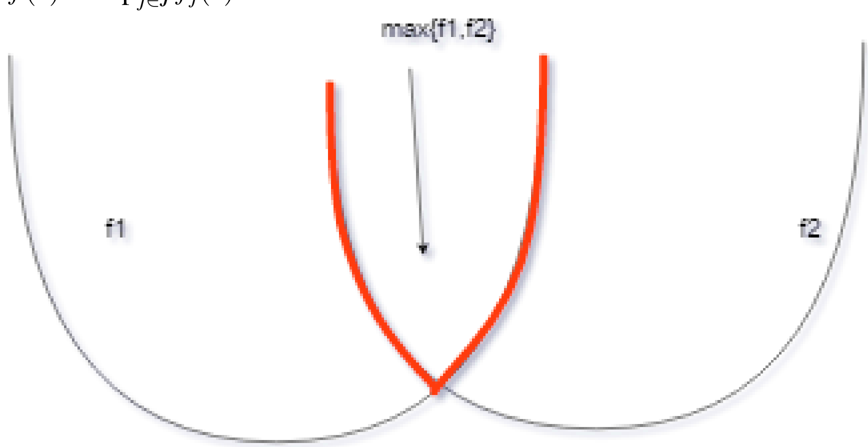
Fact 1.3.2 — Epigraph of a support function is a cone. Epigraph of a support function $epi\sigma_c = \{(x,t)|\sigma_c(x) \leq t\}$. Take any $\alpha > 0$, $\alpha(x,t) = (\alpha x, \alpha t)$.
$\sigma_c(\alpha x) = \alpha\sigma_c(x) \leq \alpha t$.
$\Rightarrow \alpha(x,t) \in epi\sigma_C$



$\Rightarrow epi\sigma_c$ is a cone.

1.4 Operations Preserve Convexity of Functions

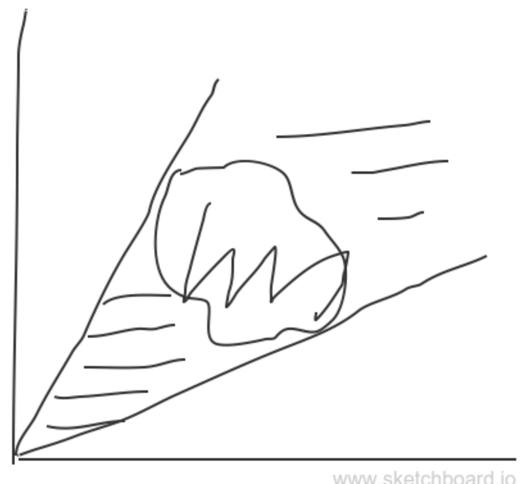
1. Positive affine transformation
   $f_1, f_2, \dots, f_k \in cvx\mathbb{R}^n$.
   $f = \alpha_1 f_1 + \alpha_2 f_2 + \ldots + \alpha_k f_k$
2. Supremum of functions. Let $\{f_i\}_{i\in I}$ be arbitrary family of functions. If $\exists x \sup_{j\in J} f_j(x) < \infty \Leftrightarrow$
   $f(x) = \sup_{j\in J} f_j(x)$



3. Composition with linear map.
   $f \in cvx\mathbb{R}^n, A : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map. $f \circ A(x) = f(Ax) \in cvx\mathbb{R}^n$

$$
\begin{aligned}
f \circ A(x) &= f(A(\alpha x + (1-\alpha)y)) \\
&= f(A\alpha x + (1-\alpha)Ay) \\
&\leq \alpha f(Ax) + (a-\alpha)f(Ay)
\end{aligned}
$$

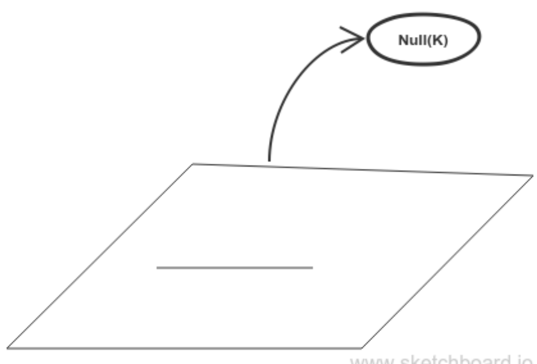
Observation 1.4.1 — Special case: Cone generated by $X \subset \mathbb{R}$. $cone(x) = \{\alpha x | x \in ch(x), \alpha > 0\} = \{y | x_i \in X, \alpha \in \mathbb{R}^n_+, y = \alpha^T x\}$



■ **Example 1.9 — Finitely generated cone.** $X = \{e_1, e_2, \ldots, e_n\}$ where $e_i \in \mathbb{R}^n$ is a basis vector. Then $cone(X) = \mathbb{R}$ ■

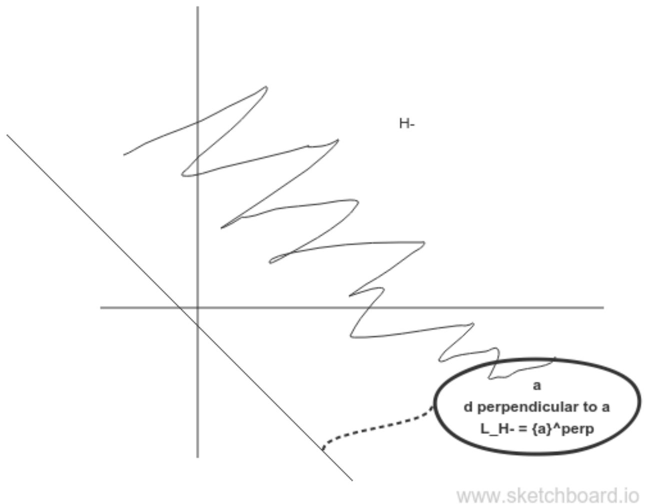
■ **Example 1.10 — Polyhedral cone.** For linear map A, B, $K = \{x \in \mathbb{R}^n | Ax \leq 0, Bx = 0\}$ ■

Generalize notion of linear subspace, for example $null(B)$ is special cone.



■ **Example 1.11 — Lineality space.** Linear space of a convex set C, $L_c = \{d | x + \alpha d \in C \forall x \in C, \forall \alpha\}$.

Lineality space of polyhedral $cone(K)$ is $L_k = null \begin{pmatrix} A \\ B \end{pmatrix}$. K is pointed iff $rank \begin{pmatrix} A \\ B \end{pmatrix} = n$.



■

Polar of a convex cone K, $K^c = \{y | \langle y, x\rangle \leq 0, \forall x \in K\}$
Polar cone and linear map

- $C \subset \mathbb{R}^n$ closed convex cone
- $A : \mathbb{R}^n \to \mathbb{R}^m$ linear map
- $K = A^{-1}(C) = \{x | Ax \in C\}$
- K^c $\{z | z = A^T y, y \in C^c\}$

■ Example 1.12

$$
\begin{aligned}
C &= \{0\} \\
K &= \{x | Ax = 0\} = null(A) \\
K^c &= \{z | z = A^T y, y \in \mathbb{R}^n = \{0\}\} = range(A^T)
\end{aligned}
$$

■

■ Example 1.13

$$
\begin{aligned}
C &= \mathbb{R}^n = \{y | y_i \leq 0\} \\
K &= \{x | Ax \leq 0\} \\
K^c &= \{z | z = A^T y, y \geq 0\}
\end{aligned}
$$

■

Take any $x \in K, z \in K^c$, $\langle x, z\rangle = \langle x, A^T y\rangle = \langle Ax, y\rangle \leq 0$

Definition 1.4.1 — Projection onto a closed convex set. $P_c(x) = \arg\min \frac{1}{2}|z - x|^2$

Claim 1.4.2 Projection onto C is unique.
Suppose $y_1, y_2 \in Pc(x)$, if $y_1 \neq y_2$, take $y_0 = \frac{1}{2}(y_1 + y_2)$.

Proof.

$$\begin{aligned}\|x-\frac{1}{2}|y_1+y_2|\|^2 &= \frac{1}{4}\|x-y_1+x-y_2\|^2\\ &= \frac{1}{2}(\|x-y_1\|+\|x-y_2\|-\frac{1}{2}\|y_1-y_2\|^2)\\ &< d-\frac{1}{4}\|y_1-y_2\|^2\end{aligned}$$

■

Theorem 1.4.3 — Projection Theorem. $y_x = P_c(x) \Leftrightarrow \langle x-y_x, y-y_x\rangle \leq 0 \quad \forall y \in C$,
Suppose C is an affine manifold. $\langle x-y_x, y-y_x\rangle = 0$

Proposition 1.4.4 1. $\{x \in \mathbb{R}^n | x = P_c(x)\} = c$
2. $P_c \circ P_c = P_c$. (Idempotent)
3. $\|P_c(x_1)-P_c(x_2)\|^2 \leq \langle P_c(x_1), P_c(x_2), x_1, x_2\rangle$
   (a) $0 \leq \langle P_c(x_1)-P_c(x_2), x_1-x_2\rangle$
   (b) Apply Cauchy to (3), $\|P_c(x_1)-P_c(x_2)\| \leq \|P_c(x_1)-P_(x_2)\| \cdot \|x_1-x_2\| \Rightarrow \|P_c(x_1)-P_c(x_2)\| \leq \|x_1-x_2\|$
   (c) If $0 \in C \quad \Rightarrow \quad \|P_c(x)\| \leq \|x\|$

1.5 Notes on January 23rd

Projection onto convex set $C \in R^n$ $P_c(x) = \arg\min_{z\in C} \frac{1}{2}\|z-x\|^2$

Theorem 1.5.1 — Projection Theorem. $y_x = P_c(x) \Leftrightarrow \langle x-y_x, y-y_x\rangle \leq 0 \quad \forall y \in C$,
Suppose C is an affine manifold. $\langle x-y_x, y-y_x\rangle = 0$

Projection onto convex closed set $K \in \mathbb{R}^n$. $y_x = P_K(x) \Leftrightarrow y_x \in K, x-y_x \in K^\perp, \langle x-y_x, y_x\rangle = 0$

■ **Example 1.14** Now any ?? $K = \mathbb{R}^n_+$, $y_K = \max\{0, x_1\}$.

$$x-y_x = \begin{cases} 0 & x \geq 0 \\ x & x < 0 \end{cases}$$

■

Proof of ?? $\Rightarrow$
By PT, $y_x \in P_c(x) \quad \Rightarrow \quad \langle x-y_x, y-y_x\rangle \quad \forall y \in C$. Because $y \in K$, $y = \alpha y_x \forall \alpha > 0$.

$$\begin{aligned}0 &\geq \langle \alpha y_x - y_x, x-y_x\rangle\\ &= \langle(\alpha-1)y_x, x-y_x\rangle\\ &= (\alpha-1)\langle y_x, x-y_x\rangle\end{aligned}$$

$\alpha - 1 \leq 0 \Rightarrow \langle y_x, x-y_x\rangle = 0$
$(*)$ implies $< y, x-y_n > \leq 0 \forall K$ ■

Consequences
$P_k(x) = 0 \iff x \in K^o$

$P_K(\alpha x) = \alpha P_K(x) \forall \alpha > 0$

$$P_K(\alpha x) = \arg\min_{z \in C} \frac{1}{2}\|z - \alpha x\| = \arg\min_{z \in C} \frac{1}{2}\|z/\alpha - \alpha x\| \alpha P_K(\alpha x) = \ldots ???$$

Proposition 1.5.2 The following are equal

1. $x = x_1 + x_2$ with $x_1 \in K, x_2 \in K^o, \langle x_1, x_2 \rangle = 0$
2. $x_1 = P_x(x), x_2 = P_{K^o(x)}$

1.6 Optimal Conditions for Unconstrained optimization

$\max_{x \in \mathbb{R}^n} f(x) \quad f : \mathbb{R}^n \to \mathbb{R}$ continuous on all $\mathbb{R}^n$.

■ **Example 1.15** $f(x) = e^{-x}$
The function is not coercive ■

Theorem 1.6.1 — Weierstrass Extreme Value . Every continuous function on compact set achieves its extreme value on that set.

R Caution to apply on $\mathbb{R}^n$, it's closed but not compact.

Definition 1.6.1 — coercive. The function $f : \mathbb{R}^n \to \mathbb{R}$ is coercive if for any sequence in $\{x^k\}\mathbb{R}^n$ diverges $\Rightarrow f(\{x^k\})$ also diverges.

Theorem 1.6.2 — Coercivity and compactness. Let $f : \mathbb{R}^n \to \mathbb{R}$ and all $\mathbb{R}$. Then f is coercive $\iff S_x \|x| f(x) \leq \alpha\|$ are compact for each $\alpha \in Z$

Proof. Assume S_x not bdd, $\exists \{x^k\} \in S_x$ such that $\|x^k\| \to \infty$. Since f is coercive, $\|f(x^k)\| \to \infty$ ■

Theorem 1.6.3 — Coercivity implies. let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous, If f is coercive, P has at least one global solution

Proof. Coercivity implies S_x is compact
Wierestrauss implies $\min\{f(x) | x \in S_x\}$ has at least one minimizer.
$\arg\min_{x \in \mathbb{R}^n} f(x) \supset \{f(x) | x \in S_x\}$ ■

■ **Example 1.16 — Simple function not coersive.** $f(x) = \frac{1}{2}\|Ax - b\|^2$
If A is not full rank, f is not continuous. Take $x^k \in NULL(A)$, $\|x^k\| \to \infty$ but $f(x^k) = \frac{1}{2}\|b\|^2$. There's still global maxima even the function is not coercive. ■

1.6.1 First Order Optimality Condition

Reduce to a one dimensional case. Take $\phi(\alpha) = f(x + \alpha d)$ for some $x, d \in \mathbb{R}^n$. Direction of derivative, $f'(x,d) = \lim_{\alpha \to 0} \frac{f(x+\alpha d) - f(x)}{\alpha}$. When f is differentiable, then $f(x,d) = \nabla f(x)^T d = \phi'(\alpha)$ If $f'(x,d) < 0$, by continuity of f, $\exists \bar{\alpha} > 0$ s.t. $\frac{f(x+\alpha d) - f(x)}{\alpha} < 0 \; \forall \alpha \in (0, \bar{\alpha})$.
$\Rightarrow f(x + \alpha d) < f(x) \forall \alpha \in (0, \bar{\alpha})$.

Lemma 1.6.4 — 1st order optimality. Let $f\mathbb{R}^n \to \mathbb{R}$ be a local solution to P. Then $f'(x,d) \geq 0, \forall d st f'(x,d) \exists$

Theorem 1.6.5 Let f differentiable at $\bar{x} \in$. If $\bar{x}$ is a local min. $\nabla f(\bar{x}) = 0$

Proof. $0 \leq f'(\bar{x},d) = \nabla f(\bar{x})^T d \forall d$.
Take $d = -\nabla f(\bar{x})$
$0 \leq \nabla f(\bar{x})^T d = -\nabla f(\bar{x})^T \nabla f(\bar{x})$. $\|\nabla f(\bar{x})\| \leq 0 \Rightarrow \|\nabla f(\bar{x})\| = 0$ ■