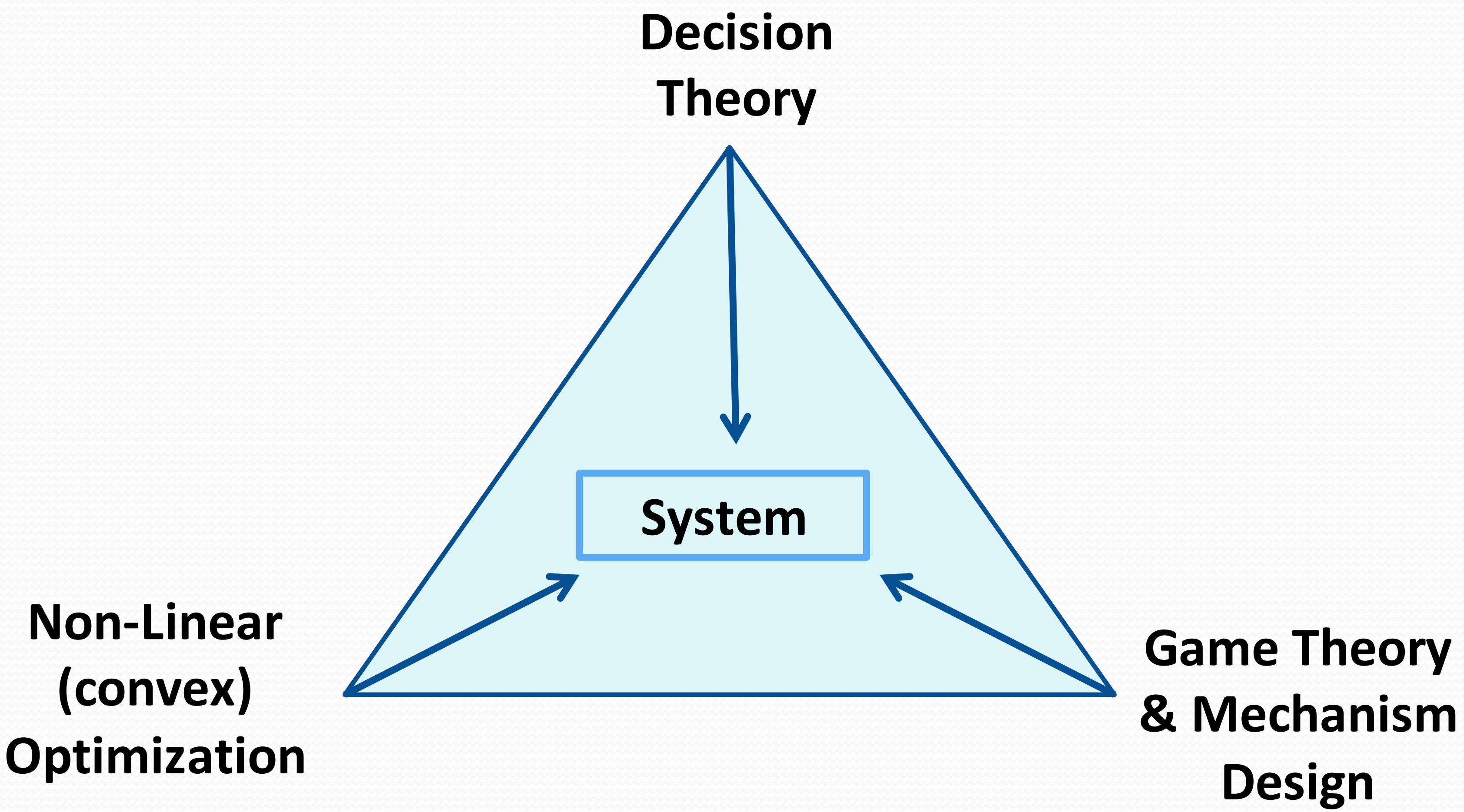
Master-MIRI Topics on Optimization and Machine Learning (TOML)

José M. Barceló Ordinas
Departament d'Arquitectura de Computadors
(UPC)

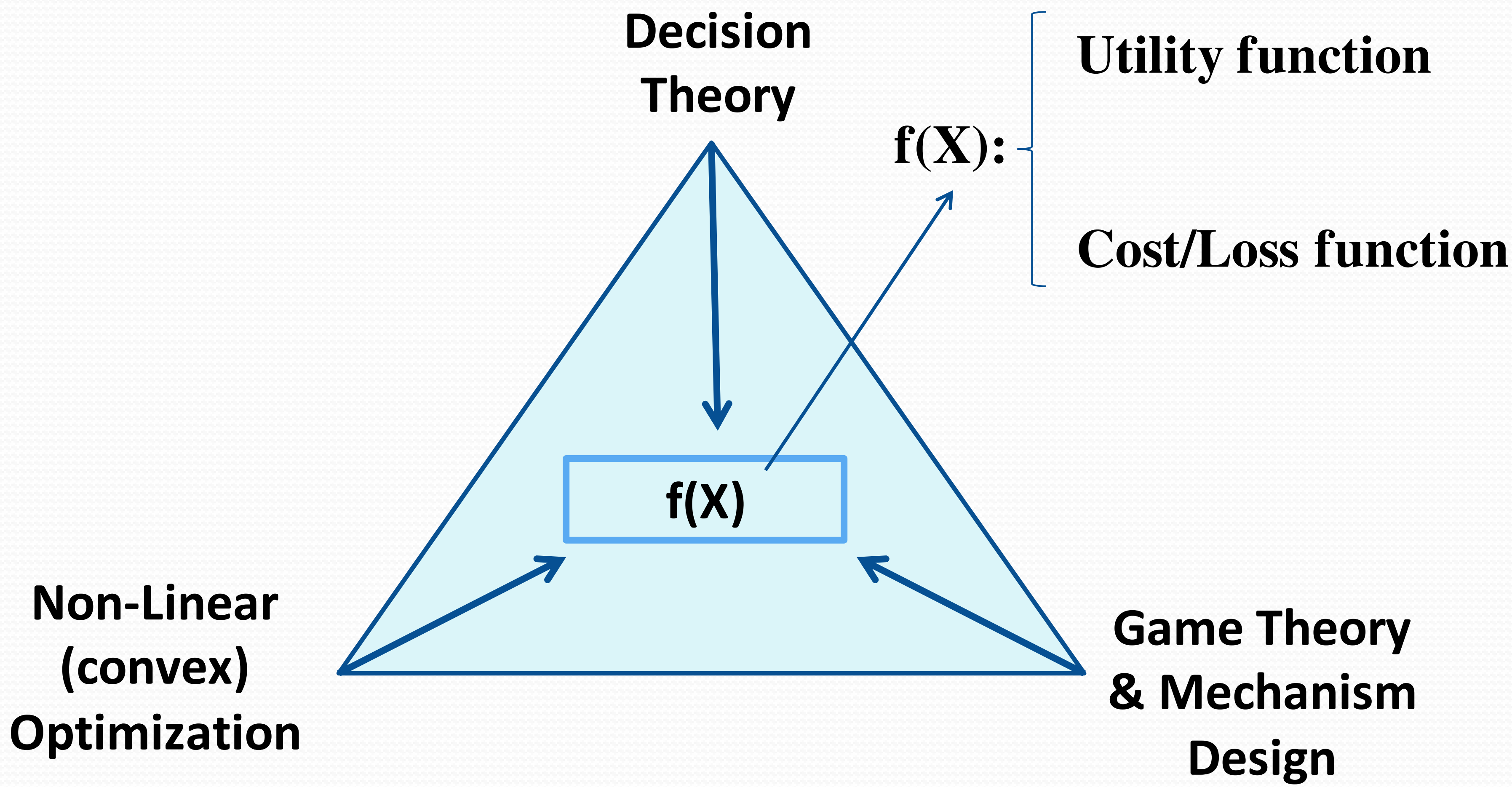
- **Topic 1: Convex Optimization Basics**
  - **General Objective:**
    - Many networking problems can be solve using optimization tools. This topic 1 gives basic knowledge and skills on the basic background to understand, formulate and recognize convex optimization problems.
    - **Textbook:** “Convex Optimization”, Stephen Boyd and Lieven Vendenberghe, Ed. Cambridge University Press

- **Topic 1: Convex Optimization Basics**
  - Specific Objectives:
    - Understand what a **convex set** is and the operations that preserve convexity
    - Define and operate with **convex functions**
    - Identify and formulate basic **convex optimization problems** (linear, quadratic and geometric optimization problems)
    - Understand **duality** and **Karush-Khun-Tucker (KKT) optimality conditions**
    - Some methods to solve COP’s: **Gradient descent methods** and **Interior Point methods**

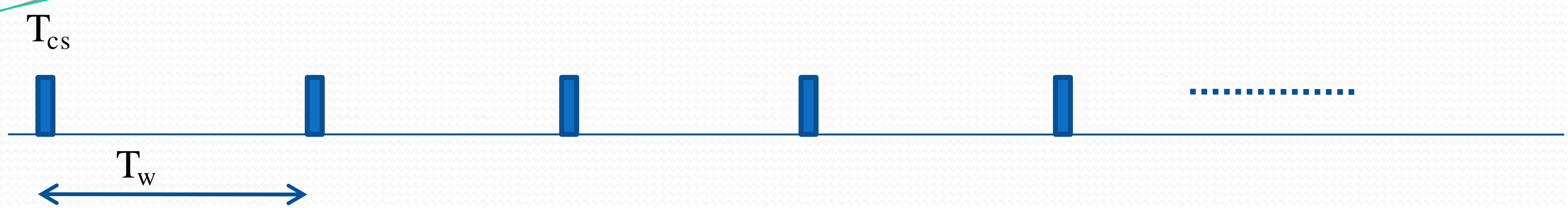
Optimization, Game Theory and Decision Theory



Optimization, Game Theory and Decision Theory



Example – Duty-cycle MAC protocols



In general MAC, contention-based, the duty cycle is defined as the quotient between the carrier sense (CS) period over the CS + Sleep period, i.e., $DC=T_{cs}/T_w$,

$DC \sim 1$ if $T_{cs} \sim T_w$, the extreme case is that the radio is always ON and DC=1
$DC \sim 0$ if $T_{cs} << T_w$, where we can make T_w as large as we want and then the radio is OFF, of course, not transmitting any packet

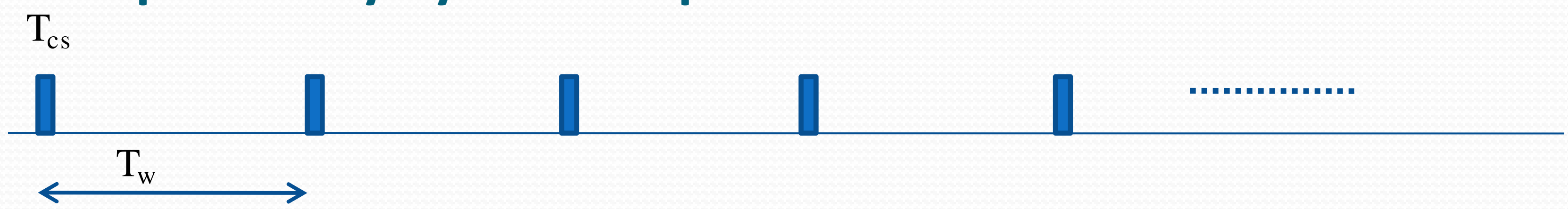
In general, without specifying how intrinsically works the MAC protocol, e.g. WiseMAC, D-MAC, B-MAC, …. We can obtain formulas for the delay and energy consumption. Let's say that the packets are generated randomly, CBR with period $T>>T_w$. Nodes have their T_w not synchronized, e.g., like WiseMAC.

Average Delay, assuming h hops: $\mathbf{L= h\ (T_w/2) + T_{DATA}}$, where T_{DATA} is the transmision time for a packet, e.g. $T_{DATA} = T_{cw} + T_{guard} + T_{msg}$. → $\mathbf{L= a\ T_w + b}$

Average Energy Consumption: $\mathbf{E = E_{cs} + E_{rx} + E_{tx} + E_{others} = E_{idle}\ T_{cs}/T_w + d}$, where d is the energy consumed by Rx/Tx packets and E_{others} accounts for other factors

→ $\mathbf{E = c\ /T_w + d}$

Example – Duty-cycle MAC protocols



Summarizing: $DC=T_{cs}/T_w$,

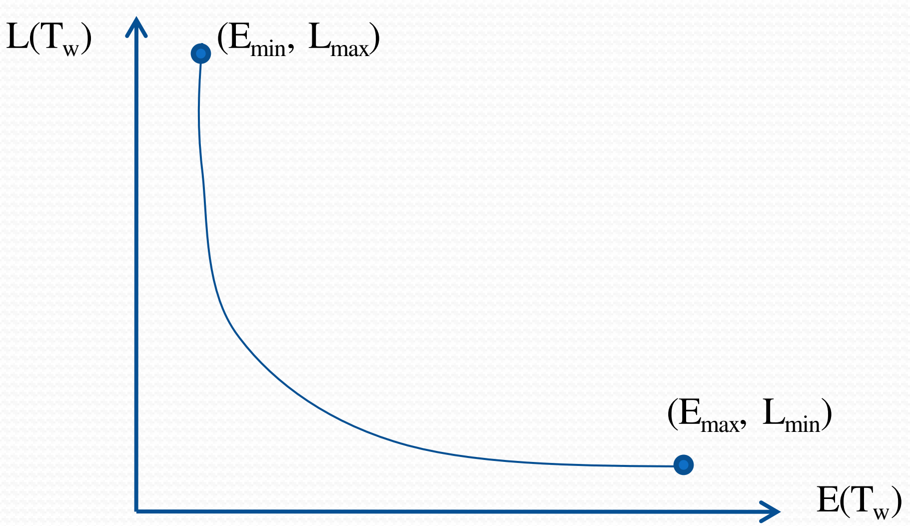
DC ~ 1 if $T_{cs} \sim T_w$, the extreme case is that the radio is always ON and DC=1

DC ~ 0 if $T_{cs} << T_w$, where we can make T_w as large as we want and then the radio is OFF, of course, not transmitting any packet

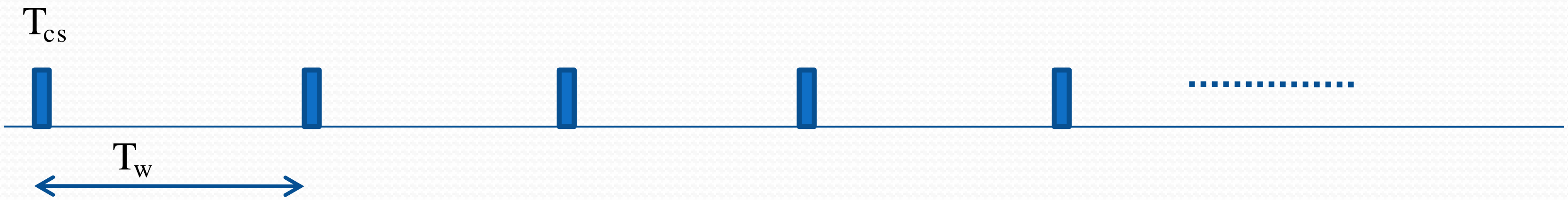
Average Delay, $\mathbf{L= a\ T_w + b}$

Average Energy Consumption: $\mathbf{E = c\ /T_w + d}$

- If $T_{cs} \sim T_w$ (DC ~ 1) → $L(T_w)=L_{min}$ and $E(T_w)=E_{max}$.
- If $T_{cs} << T_w$ (DC ~ 0) → $L(T_w)=L_{max}$ and $E(T_w)=E_{min}$, where L_{max} can be $\propto$ if $T_w= \propto$



Example – Duty-cycle MAC protocols

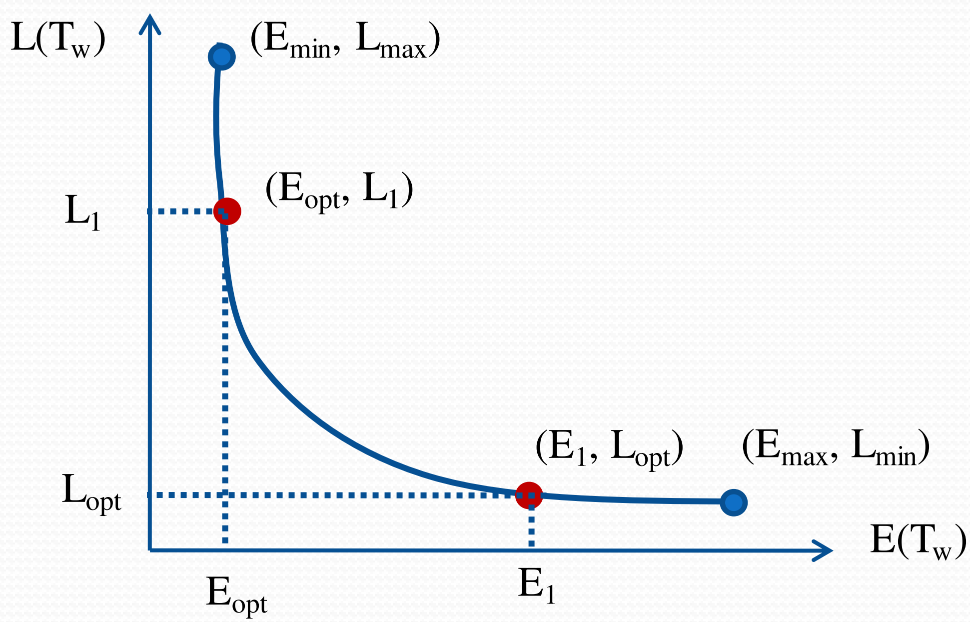


Let us **optimize (Non-Linear)** the system using the following models:

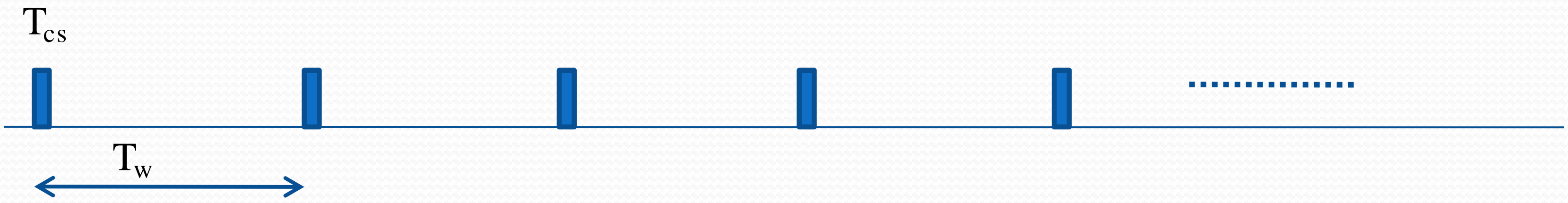
(1) Minimize $E(T_w) = c / T_w + d$
s.t. $L(T_w) \leq L_{th}$
var T_w

(2) Minimize $L(T_w) = a\, T_w + b$
s.t. $E(T_w) \leq E_{th}$
var T_w

Optimization Problem (1) results in T_{w1}, corresponding to the pair $(E(T_w), L(T_w)) = (E_{opt}, L_1)$
Optimization Problem (2) results in T_{w2}, corresponding to the pair $(E(T_w), L(T_w)) = (E_1, L_{opt})$

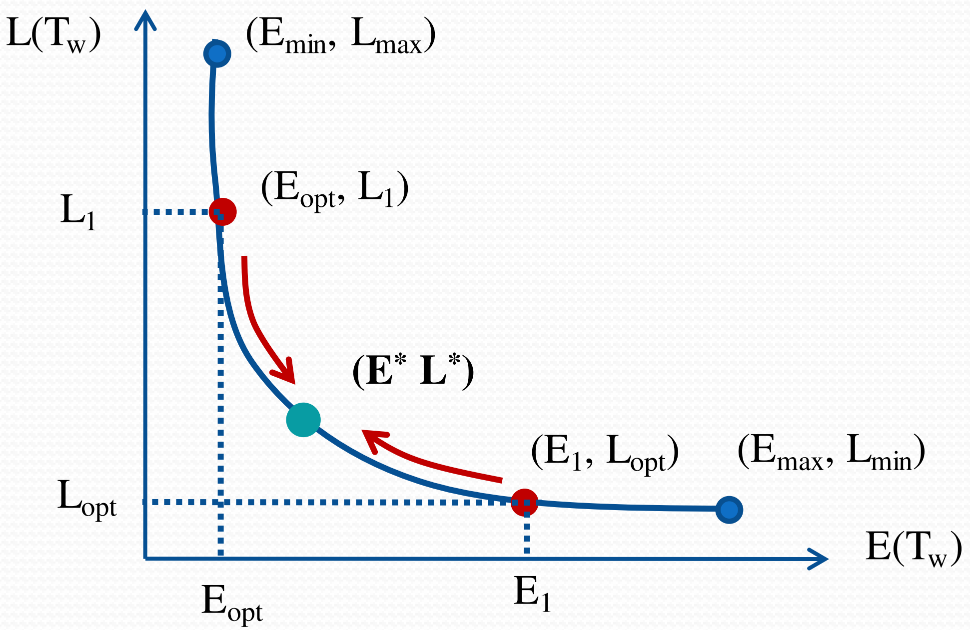


Example – Duty-cycle MAC protocols



It is clear that we are interested in finding a "fair" operation point in which we trade off energy consumption and latency according to the specific characteristics of the MAC considered.

Game Theory gives us a method to find this operation point.

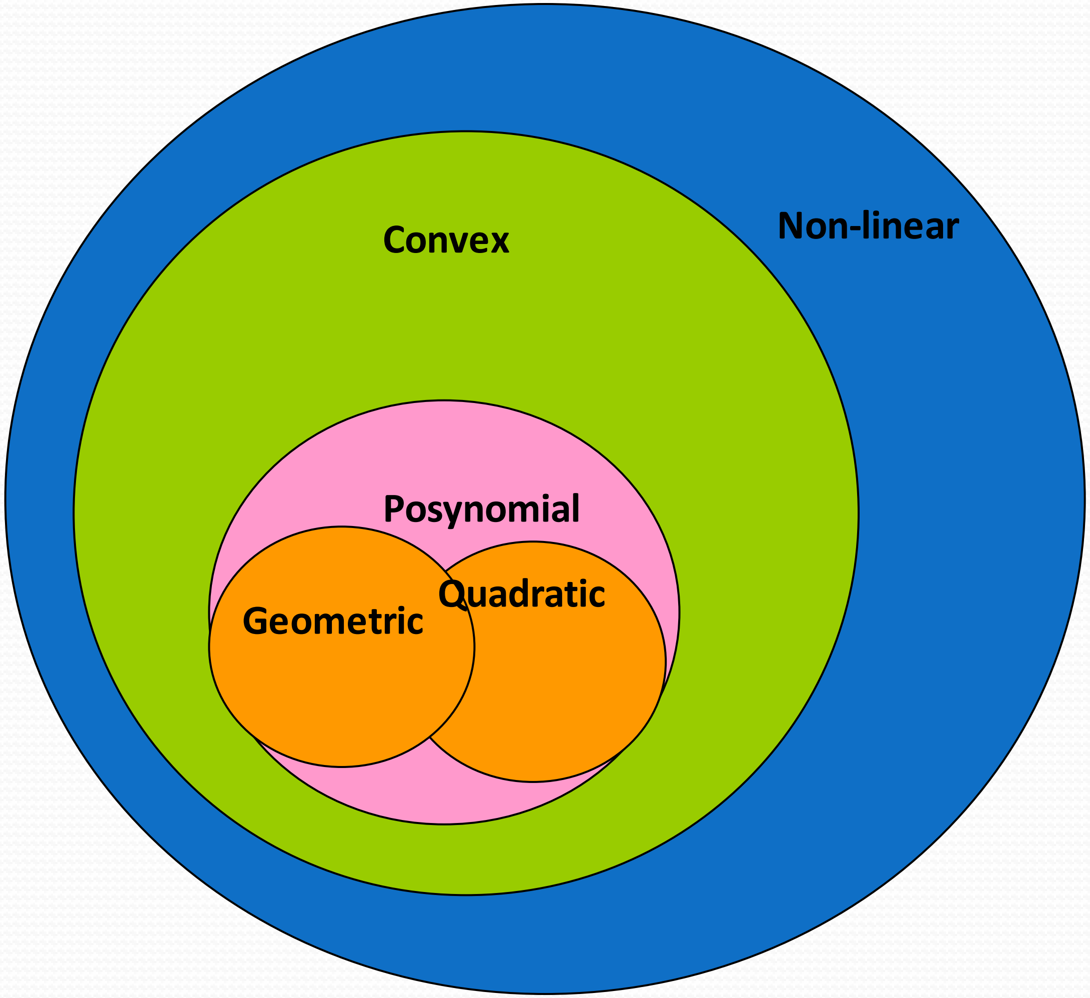


Example – Other issues

- **Calibration of sensors:** sensors does not come calibrated and if they come they have been calibrated on chambers (ideal conditions) and not in the field (uncontrolled environment),
  - Use of regression methods such as multiple linear regression methods (MLR) or support vector regression (SVR) → optimization problem
- **Graph Signal Processing:** the signal captured by the network is smooth. In order to analyse the signal we can use the Laplacian of the graph formed by the sensor network. Finding the optimal Laplacian according to the smothness of the signal is an optimization problem,
- **Power control** in celular/ad hoc networks is an optimization problem
- **Resource allocation** is an optimization problem
- **Optimal routing** is an optimization poblem

Convex Optimization Problems. Convex Sets

In mathematics, computer science or electrical engineering, **optimization,** is the selection of a best element with regard some criteria (constraint) from some set of available alternatives.



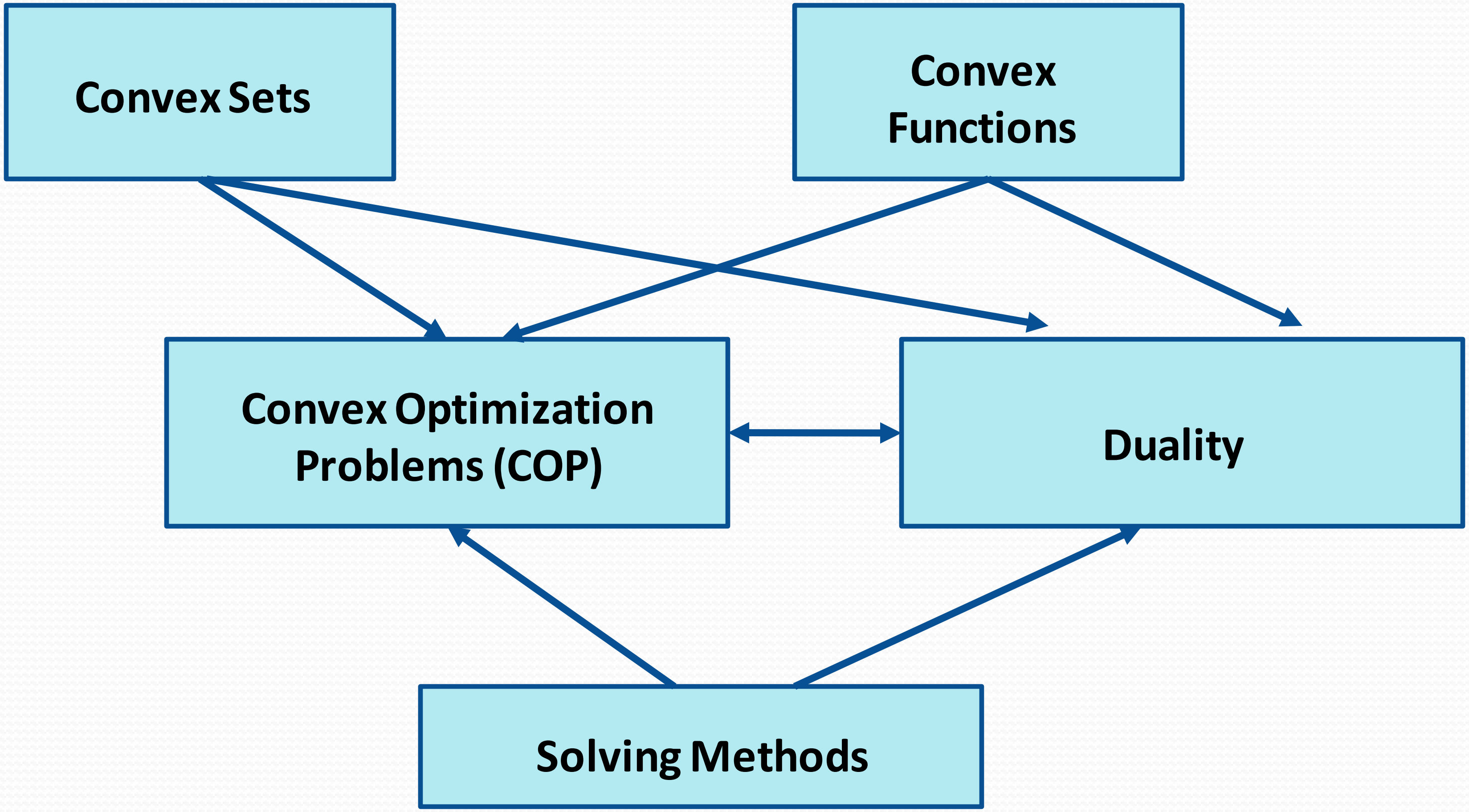
Convex Optimization Problems:

a local minimum is a global minimum

Non-Linear Optimization non-convex

There are no effective methods to solve the problem → bounds, approximations, heuristics, …

Convex Optimization Problems. Convex Sets



Vector Space:

A vector space is a set of objects called vectors that are closed under finite vector addition and scalar multiplication. Then, for V to be a vector space ($x,y,z \in V$ and $s,t \in R$):

i) **Commutativity**: $x+y=y+x$
ii) **Associativity**: $(x+y)+z = x+(y+z)$
iii) **Additive identity**: $0+x=x+0=x$
iv) **Existence of inverse**: $x+(-x)=0$
v) **Associativity of scalar multiplication**: $s(tx)=(st)x$
vi) **Distributivity of scalar sums**: $(s+t)x=sx+tx$
vii) **Distributivity of vector sums**: $s(x+y) = sx+sy$
viii) **Scalar multiplication identity**: $1x=x1=x$

Examples:

- Coordinate space: $x \in R^n$, $x=(x_1,\ldots, x_n)$ and $t \in R$ is a vector space
- Set of matrices R^{mn}, is a vector space over R, (addition of matrices and multiplication of scalar over matrices)
- P_n, set of polynomials over R (coefficients in R) and of order less or equal of n
- Set of continuous functions $f: R^n \rightarrow R$, where $(f+g)(x) = f(x)+g(x)$ and $(tf)(x)=tf(x)$

Affine Space

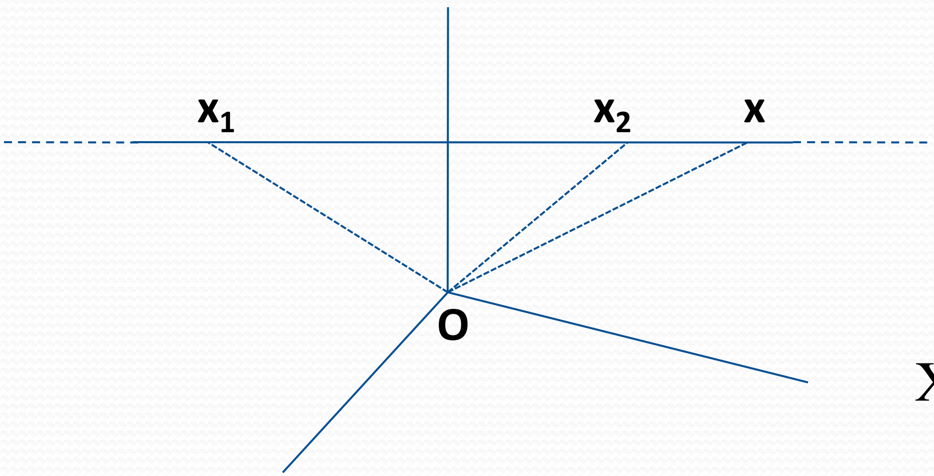
A geometric structure composed of points, a vector space and operations (e.g. translations) over these points which form a vector space.

Operations are over two points to form a vector, or over a point and a vector to translate the point. Specifically, an affine space A is a map:

$$f: A \times V \longrightarrow A$$
$$(a,v) \qquad\qquad a+v$$

with V a vector space that underlines the affine space A. The mapping f has the following properties: Identity, Associativity and Uniqueness.

- For example:



$$\vec{OX} = \vec{OX_1} + \vec{X_1X} = \vec{OX_1} + \lambda \vec{X_1X_2}$$

$$\downarrow$$

$$X = X_1 + \lambda(X_2 - X_1) = (1-\lambda)X_1 + \lambda X_2$$

- **Affine sets:**

A **set is affine** if it contains the line through any two distinct points in the set. For any $x_1,x_2 \in C \subset R^N$ and $\theta \in R$, the linear combination of these two points, also lies in A:

$$\theta_1 x_1+\theta_2 x_2 = \theta x_1+(1-\theta)x_2 \in A \quad \text{with} \quad \theta_1+\theta_2=1 \text{ and } \theta,\theta_1,\theta_2 \in R$$

Let us extend the case of having k points $x_1,x_2,\ldots,x_k \in A$ and let $\theta_1, \theta_2, \ldots, \theta_k \in R$. Then:

$$\theta_1 x_1+ \theta_2 x_2+ \ldots +\theta_k x_k \in A \quad \text{if} \quad \theta_1+\theta_2+ \ldots+\theta_k=1$$

is an **affine combination** of the points $x_1, x_2, \ldots, x_k$.

If A is an affine set and $x_0 \in A$, then $V=A-x_0=\{x-x_0 | x \in A\}$ is a subspace (close under sums and scalar multiplication) and:

$$A=V+x_0=\{v+x_0 | v \in V\} \quad \text{(subspace with an offset)}$$

and its dimension is the dimension of the subspace V.

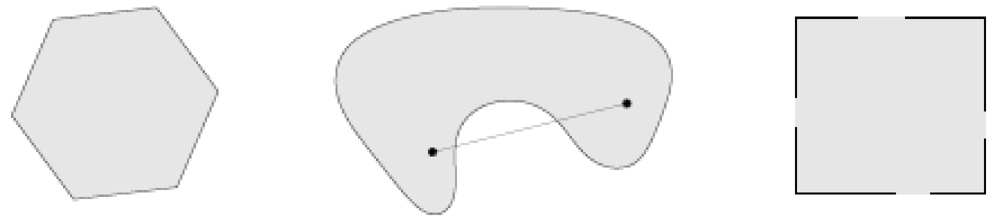
- **Exercise:**
  - Proof that the set of solutions of the linear system $Ax=b$ (with $A \in R^{mn}$) is an affine set,
- **Convex Sets**

A set is a convex set if **the line segment** between ANY two points of C **lies in C**. That means that if $x_1, x_2 \in C$ and $0 \leq \theta \leq 1$, then:

$$\theta_1 x_1 + \theta_2 x_2 = \theta x_1 + (1-\theta) x_2 \in C \quad \text{with} \quad \theta_1 + \theta_2 = 1 \text{ and } 0 \leq \theta, \theta_1, \theta_2 \leq 1$$

A point $x = \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_k x_k \in C$ with $\theta_1 + \theta_2 + \ldots + \theta_k = 1$, $\theta_i \geq 0$ is a **convex combination of the points** $x_1, x_2, \ldots, x_k$.

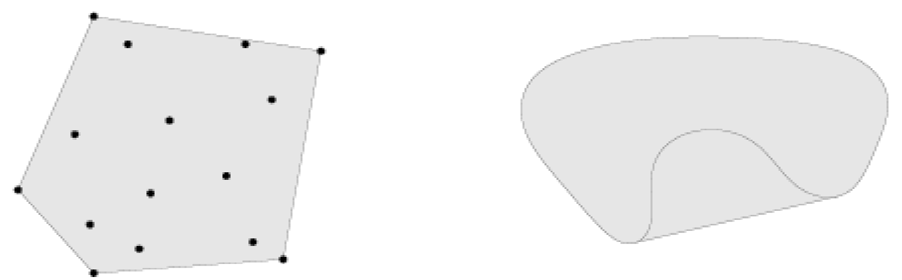
examples (one convex, two nonconvex sets)



- **Convex Hull:**
  - Let $C \subset R^n$ be an arbitrary set,

The **convex hull** conv(C) is defined as:

$$Conv(C)=\{x \mid x=\Sigma_{i=1,..k} \lambda_i x_i,\ x_i \in C,\ \lambda_i \geq 0,\ \Sigma_{i=1,..k} \lambda_i = 1,\ k \geq 1\}$$



Other definitions of **convex hull:**

i. Let M be a nonempty subset in R^n. Then among all convex sets containing M (these sets exist, e.g., R^n itself) there exists the smallest one, namely, the intersection of all convex sets containing M.

ii. Conv (M) = {the set of all convex combinations of vectors from M}

- **Cones:**

A **set C is a cone** if for every $x \in C$ and $\theta \geq 0$, we have $\theta x \in C$

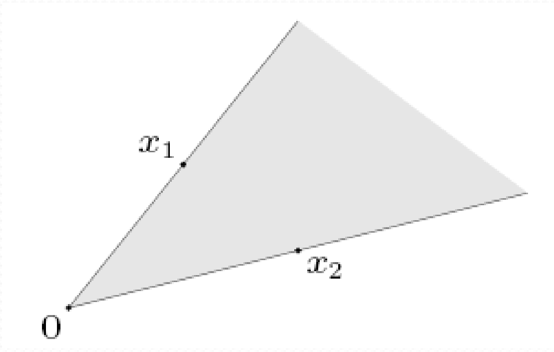
A **convex cone** is a convex set and a cone, that is: for any $x_1, x_2 \in C$ and $\theta_1 \geq 0$, $\theta_2 \geq 0$, we have $\theta_1 x_1 + \theta_2 x_2 \in C$

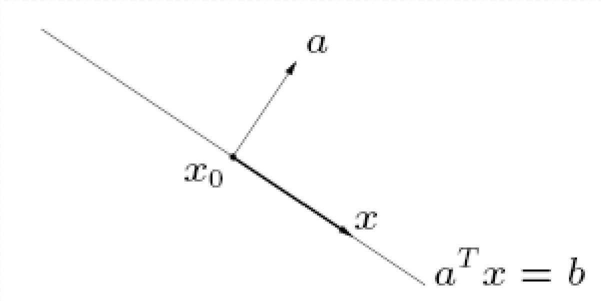
- **Hyperplanes and halfspaces are convex:**

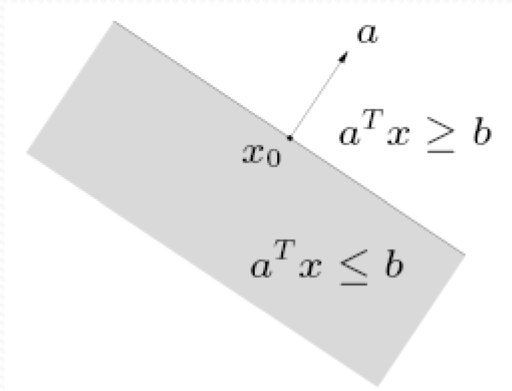
Hyperplane: subspace of dimension n-1 of a space of dimension n.
An **hyperplane** is a set of the form:

$H=\{x \mid a^T x = b\} = \{x \mid a^T(x-x_0)=0\} = x_0 + a^\perp$ with $a^\perp = \{v \mid a^T v = 0\}$ (orthogonal complement of a)

A **halfspace** is expressed as $H=\{x \mid a^T x \leq b\}$ or $H=\{x \mid a^T x \geq b\}$







Norms and inner product:

Inner product on R^N: Let be $x,y \in R^N$, $x=(x_1,\ldots, x_N)$ and $y=(y_1,\ldots, y_N)$, then $<x,y> = x^T y = \Sigma_{i=1,\ldots,N} x_i y_i$

Angle between vectors: angle=$\angle(x,y) = \cos^{-1} (<x,y> / (||x||_2 \; ||y||_2))$. Two vectors x, y are **orthogonal** if $\angle(x,y)$ =90º, then $x^T y = 0$

A **norm** is a function f: $R^N \rightarrow R$, such that for each $x \in R^N$

i) f is non-negative: $f(x) \geq 0$

ii) f is definite: $f(x) = 0$ only if $x=0$

iii) f is homogeneous: $f(tx) = |t| f(x)$, with $t \in R$

iv) f satisfies triangle inequality: $f(x+y) \leq f(x) + f(y)$

<u>A norm is a measure of the length of a vector x</u>.

Euclidean Norm or l-2 norm of: $x \in R^N$, $x=(x_1,\ldots, x_N)$, then $||x||_2 = (<x,x>)^{1/2} = (x^T x)^{1/2} = (\Sigma_{i=1,\ldots,N} x_i^2)^{1/2}$

Cauchy-Schwartz inequality: $|x^T y| \leq ||x||_2 \; ||y||_2$ for any $x,y \in R^N$,

Other norms:

l_1-norm or sum-absolute norm: $||x||_1 = \Sigma_{i=1,\ldots,N} |x_i|$

l_∞-norm or Chebyshev norm: $||x||_\infty = \max(|x_1|,\ldots, |x_N|)$

l_p-norm: $||x||_p = (\Sigma_{i=1,\ldots,N} x_i^p)^{1/p}$

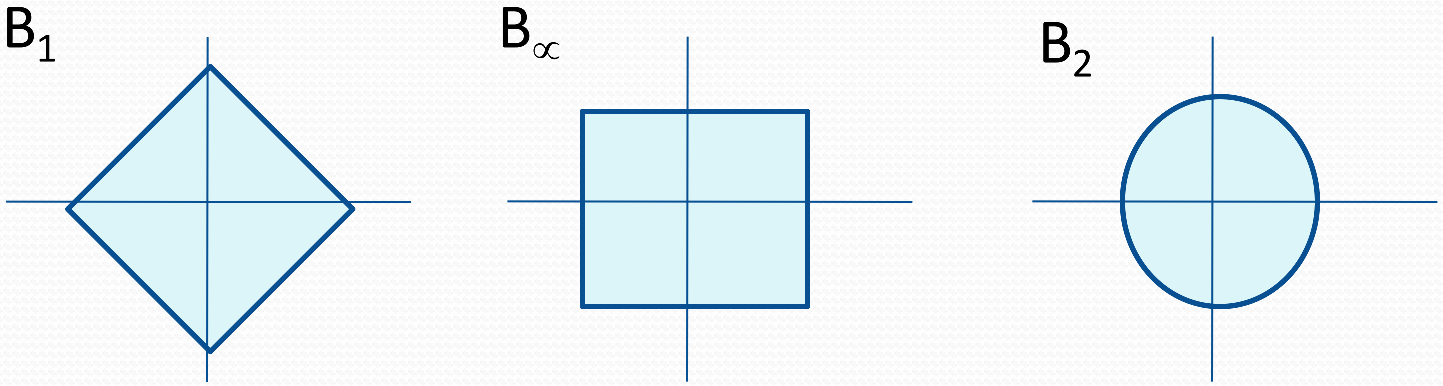
- **Other examples of convex sets**

Any **Norm Ball $\mathbf{B_p}$** is convex. Let be $\mathbf{x}, \mathbf{x_c} \in \mathbf{R^N}$ vectors, where $\mathbf{x_c}$ is the center (e.g. vector $x_c=0$). Then a Norm Ball of radius **r** is defined as

$$B_p=\{x \mid \|x-\mathbf{x_c}\|_p \le r\}$$

where

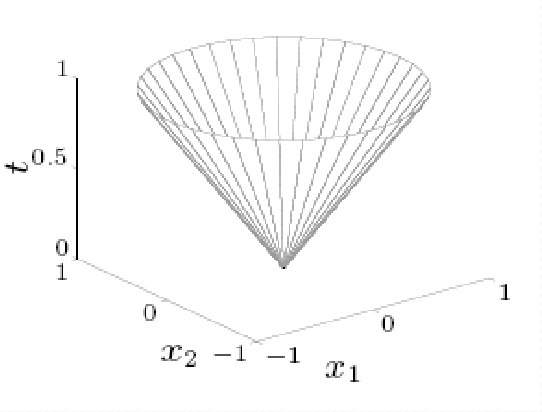
$$\|x\|_1 = \sum_{i=1}^{n} |x_i| \qquad \|x\|_\infty = \max_{1\le i\le n} |x_i| \qquad \|x\|_p = (\sum_{i=1}^{n} |x_i|^p)^{1/p}$$



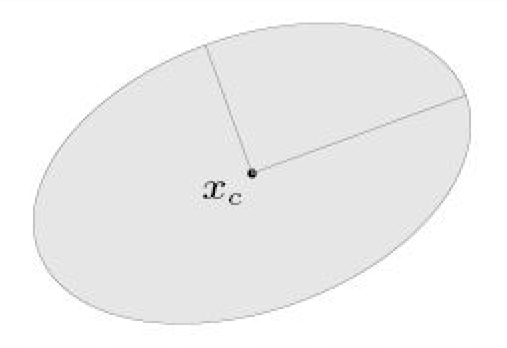
- **Proof than an Euclidean Ball $B_p=\{x \mid ||x-x_c||_2 \leq r\}$ is convex.**
- **A norm cone associated to norm $||\cdot||_p$:**

  Set $C=\{(x,t) \mid ||x||_p \leq t\}$

  - For example: $\{(x_1,x_2,t) \mid (x_1^2+x_2^2)^{1/2} \leq t\}$



- **Ellipsoid:** $E=\{ x \mid (x-x_c)^T P^{-1} (x-x_c) \leq 1 \}$ with $P=P^T>0$ (P is symmetric and positive definite) and vector x_c the centre of the ellipsoid.



- **Other examples of convex sets**
  - Let's have $\Sigma\theta_i=1$ for i=1,2, ... (infinite terms) and let $x_1,x_2,\ldots \in C$ be a convex set, then

$$\sum_{i=1}^{\infty} \theta_i x_i \in C$$

  - Moreover, let x be a random vector with $x \in C$ with probability one. Then E(x) is convex, since:

$$\int_C p(x)dx = 1 \qquad E(x) = \int_C p(x)x dx \in C$$

  - E.g. Bernoulli distribution: $p(x_1)=\theta_1$ and $p(x_2)=\theta_2=1-\theta_1$. Then,

$$E(x) = \theta_1 x_1 + \theta_2 x_2 = \theta_1 x_1 + (1-\theta_1)x_2 \in C$$

Operations that preserve convexity

These operations allow us to construct convex sets from other convex sets

Intersection: $S_1, S_2, \ldots$, convex, then $\cap S_i$ is convex

Affine functions:

$f: R^m \rightarrow R^n$ is affine if $f(x)=Ax+b$
(sum of a linear function and a constant).

Then if $S \subset R^m$ is convex and f is affine, $f(S)=\{f(x) \mid x \in S\}$ is convex
The inverse image of S under f, $f^{-1}(S)=\{x \mid f(x) \in S\}$ also is convex

Examples: **scaling** $aS=\{ax \mid x \in S\}$ and **translation** $S+a=\{x+a \mid x \in S\}$

Examples: **sum of two sets** $S_1+S_2=\{x_1+x_2 \mid x_1 \in S_1, x_2 \in S_2\}$

Examples: **Cartesian product** $S_1 x S_2=\{ (x_1,x_2) \mid x_1 \in S_1, x_2 \in S_2\}$

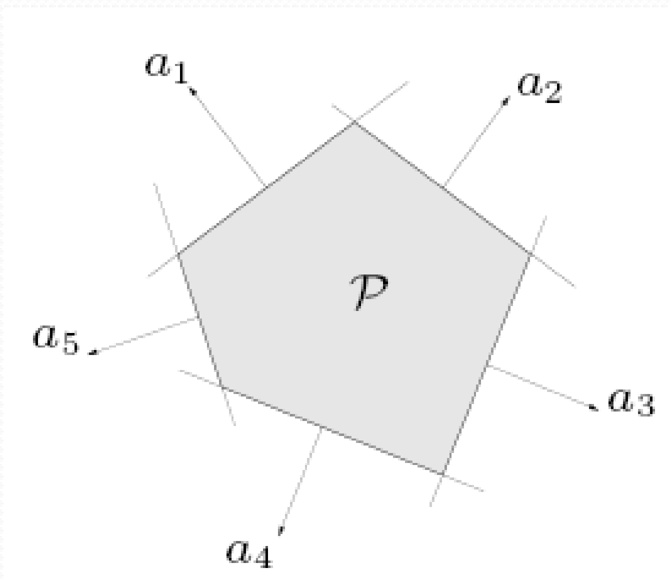
Examples: **Projection** $T=\{x_1 \in R^m \mid (x_1,x_2) \in S \text{ for some } x_2 \in R^n\}$

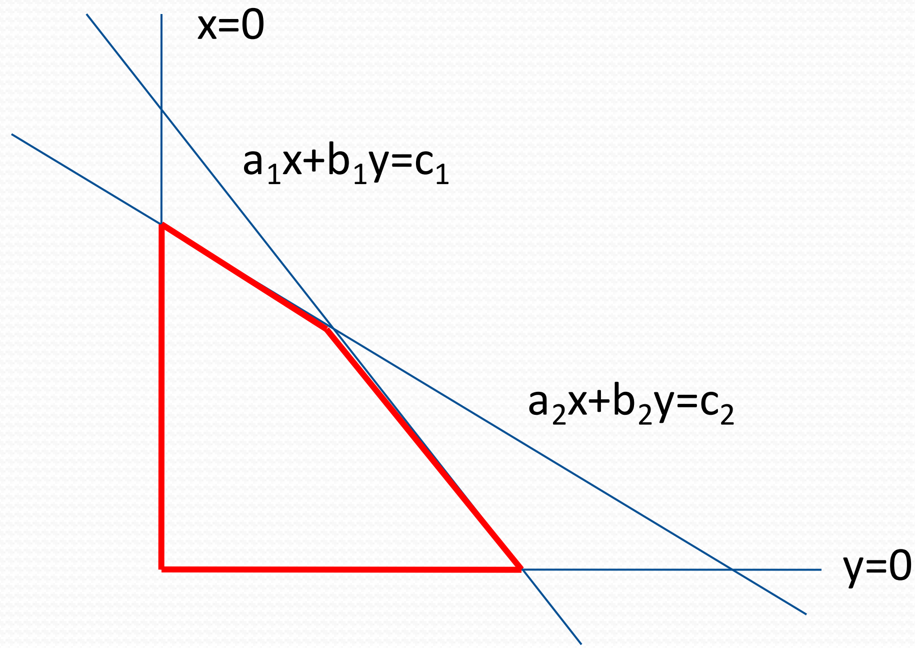
Other examples of convex sets

- A **polyhedra** is the intersection of a finite set of halfspaces and hyperplanes
- Affine sets such as rays, line segments an halfspaces are polyhedra

$$P=\{x \mid Ax \leq b, Cx=d\}$$

- Let's define a polytope as a bounded polyhedra (convex hull of the intersection of hyperplanes)





- **Separating hyperplane theorem**

If are S_1, S_2 are convex and $S_1 \cap S_2 = \emptyset$, then there exists $a \neq 0$ and b such as $a^T x \leq b$ for all $x \in S_1$ and $a^T x \geq b$ for all $x \in S_2$.

Then, the hyperplane $\{x \mid a^T x = b\}$ separates sets S_1 and S_2.

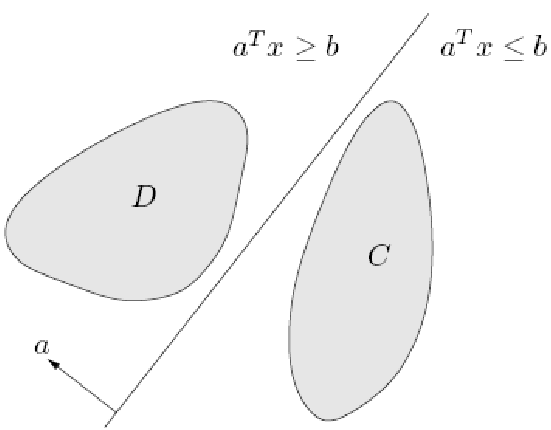


Figure 2.19 The hyperplane $\{x \mid a^T x = b\}$ separates the disjoint convex sets C and D. The affine function $a^T x - b$ is nonpositive on C and nonnegative on D.

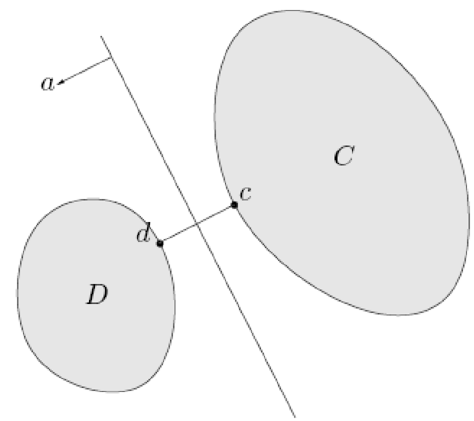
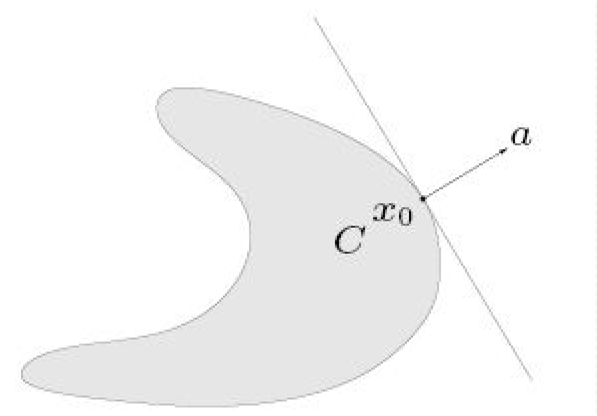


Figure 2.20 Construction of a separating hyperplane between two convex sets. The points $c \in C$ and $d \in D$ are the pair of points in the two sets that are closest to each other. The separating hyperplane is orthogonal to, and bisects, the line segment between c and d.

- **Supporting hyperplanes**

Let C is a set and x_0 a point of the boundary of C.
If $a \neq 0$ satisfies $a^T x \leq a^T x_0$ for all $x \in C$ then the hyperplane

$\{x \mid a^T x = a^T x_0\}$ is called the **supporting hyperplane to C at x_0**



The **supporting hyperplane theorem** states that for any convex set S and any x_0 belonging to the boundary of S, there exists a supporting hyperplane.