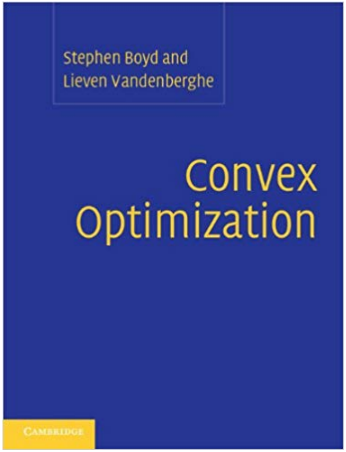
# CS495
# Optimiztaion

Waleed A. Yousef, Ph.D.,

Human Computer Interaction Lab.,
Computer Science Department,
Faculty of Computers and Information,
Helwan University,
Egypt.

April 19, 2019

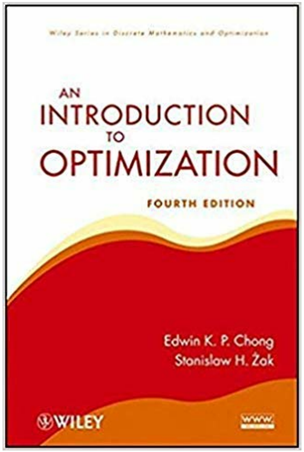
Lectures follow:
Boyd and Vandenberghe (2004)




Boyd, S., & Vandenberghe, L. (2004). Convex Optimization. Cambridge: Cambridge University Press.

Book and Stanford course: http://web.stanford.edu/~boyd/cvxbook/

Some examples from:
Chong and Zak (2013)




Chong, E. K., & Zak, S. (2001). An introduction to optimization: Wiley-Interscience.

Other complementary rigorous books for optimization/convex analysis are:
Bagirov (2014)
Bertsekas (2003)
Borwein (2006)
Dattorro (2009)
Hiriart (2001)
Rockafellar (1997)

Luenberger (1968) is a very interesting treatment for the optimization problem from the point of view of functional analysis and converging sequences.

# Course Objectives

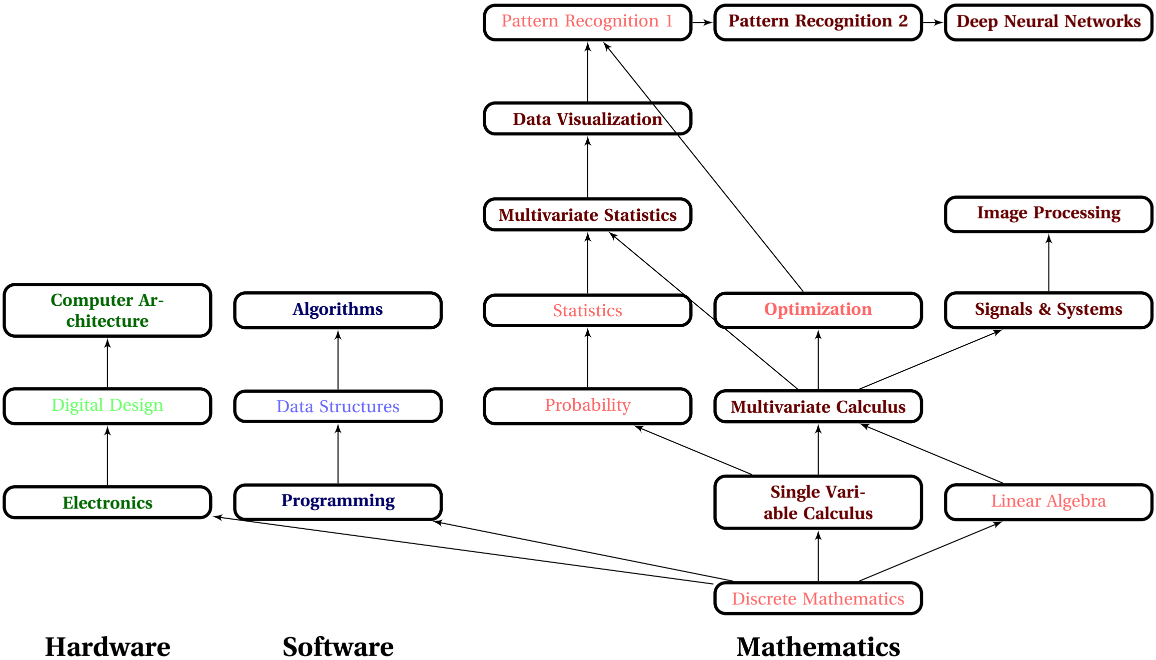
- Developing rigorous mathematical treatment for mathematical optimization.
- Building intuition, in particular to practical problems.
- Developing computer practice to using optimization SW.

## Prerequisites

1. Discrete Mathematics
2. Calculus (single variable)
3. Calculus (multi variable)
4. Linear Algebra

---

5. Some Real Analysis and some Topology: Wade (2000); Kreyszig (1978) are rigorous and wonderfully lucid; Rudin (1976) is the reference of references but very terse.


Pattern Recognition 1
Pattern Recognition 2
Deep Neural Networks
Data Visualization
Multivariate Statistics
Image Processing
Computer Architecture
Algorithms
Statistics
Optimization
Signals & Systems
Digital Design
Data Structures
Probability
Multivariate Calculus
Electronics
Programming
Single Variable Calculus
Linear Algebra
Discrete Mathematics
Hardware
Software
Mathematics

# Chapter 1
# Introduction
# Snapshot on Optimization

# Contents

# Chapter 1

# Introduction

## 1.1 Mathematical Optimization

**Definition 1** *A mathematical optimization problem or just optimization problem, has the form (Boyd and Vandenberghe, 2004):*

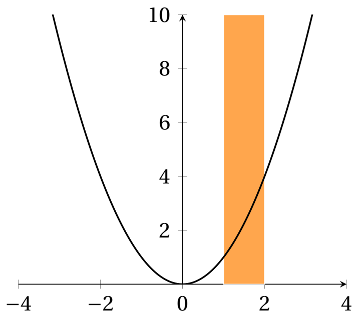
$$
\begin{aligned}
\underset{x}{\text{minimize}} \quad & f_0(x) & \\
\text{subject to:} \quad & f_i(x) \le 0, & i = 1, \dots, m \\
& h_i(x) = 0, & i = 1, \dots, p,
\end{aligned}
$$

$$
\begin{aligned}
x &= (x_1, \dots, x_n) \in \mathbf{R}^n, && \textit{(optimization variable)} \\
f_0 &: \mathbf{R}^n \mapsto \mathbf{R}, && \textit{(objective (cost/utility) function)} \\
f_i &: \mathbf{R}^n \mapsto \mathbf{R}, && \textit{(inequality constraints (functions))} \\
h_i &: \mathbf{R}^n \mapsto \mathbf{R}, && \textit{(equality constraints (functions))} \\
\mathcal{D} &: \bigcap_{i=1}^{m} \mathbf{dom}\, f_i \cap \bigcap_{i=1}^{p} \mathbf{dom}\, h_i && \textit{(feasible set)} \\
&= \left\{x \mid x \in \mathbf{R}^n \wedge f_i(x) \le 0 \wedge h_i(x) = 0\right\} && \\
x^* &: \left\{x \mid x \in \mathcal{D} \wedge f_0(x) \le f_0(z)\ \forall z \in \mathcal{D}\right\} && \textit{(solution)}
\end{aligned}
$$

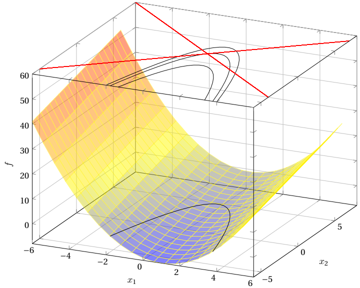
- minimize $f_0 \equiv$ maximize $-f_0$.
- $f_i \le 0 \equiv -f_i \ge 0$.
- 0s can be replaced of course by constants $b_i$, $c_i$
- unconstrained problem when $m = p = 0$.

**Example 2** *:*

$$
\begin{aligned}
\underset{x}{\text{minimize}} \quad & x^2 \\
\text{subject to:} \quad & x \le 2 \wedge x \ge 1.
\end{aligned}
$$




$x^* = 1$.
If the constraints are relaxed, then $x^* = 0$.




$$
\begin{array}{lll}
\underset{x}{\text{minimize}} & f_0(x) & \\
\text{subject to:} & f_i(x) \le 0, & i = 1, \ldots, m \\
 & h_i(x) = 0, & i = 1, \ldots, p,
\end{array}
$$

$x = (x_1, \ldots, x_n) \in \mathbf{R}^n$, (*optimization variable*)

$f_0 : \mathbf{R}^n \mapsto \mathbf{R}$, (*objective (cost/utility) function*)

$f_i : \mathbf{R}^n \mapsto \mathbf{R}$, (*inequality constraints (functions)*)

$h_i : \mathbf{R}^n \mapsto \mathbf{R}$, (*equality constraints (functions)*)

$$\mathcal{D} : \bigcap_{i=1}^{m} \mathbf{dom}\, f_i \cap \bigcap_{i=1}^{p} \mathbf{dom}\, h_i \qquad (\textit{feasible set})$$

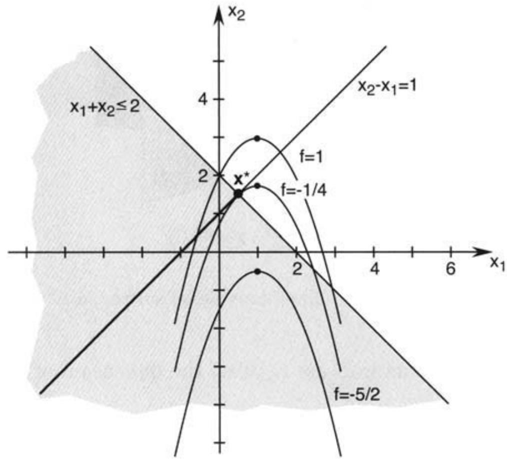
$$= \{x \mid x \in \mathbf{R}^n \wedge f_i(x) \le 0 \wedge h_i(x) = 0\}$$

$$x^* : \{x \mid x \in \mathcal{D} \wedge f_0(x) \le f_0(z) \; \forall z \in \mathcal{D}\} \qquad (\textit{solution})$$

**Example 3** (*Chong and Zak, 2013, Ex. 20.1, P. 454*):

$$
\begin{array}{ll}
\underset{x}{\text{minimize}} & (x_1 - 1)^2 + x_2 - 2 \\
\text{subject to:} & x_2 - x_1 = 1 \\
 & x_1 + x_2 \le 2.
\end{array}
$$

*No global minimizer:* $\partial z / \partial x_2 = 1 \ne 0$. *However,* $z|_{(x_2 - x_1 = 1)} = (x_1 - 1)^2 + (x_1 - 1)$, *which attains a minima at* $x_1 = 1/2$.




$x* = (1/2, 3/2)'$. (Let's see animation)

### 1.1.1 Motivation and Applications

- *optimization problem* is an abstraction of how to make "best" possible choice of $x \in \mathbf{R}^n$.
- *constrains* represent trim requirements or specifications that limit the possible choices.
- *objective function* represents the *cost* to minimize or the *utility* to maximize for each $x$.

**Examples:**

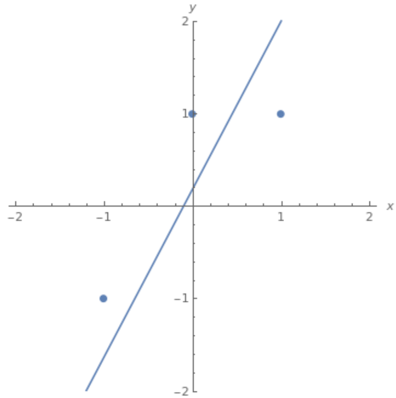
| | *Any problem* | *Portfolio Optimization* | *Device Sizing* | *Data Science* |
|---|---|---|---|---|
| $x \in \mathbf{R}^n$ | choice made | investment in capitals | dimensions | parameters |
| $f_i,\ h_i$ | firm requirements /conditions | overall budget | engineering constraints | regularizer |
| $f_0$ | cost (or utility) | overall risk | power consumption | error |

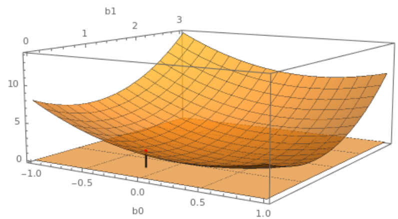
- Amazing variety of practical problems. In particular, data science: two sub-fields: construction and assessment.
- The construction of: Least Mean Square (LMS), Logistic Regression (LR), Support Vector Machines (SVM), Neural Networks(NN), Deep Neural Networks (DNN), etc.
- Many techniques are for solving the optimization problem:
  - Closed form solutions: convex optimization problems
  - Numerical solutions: Newton's methods, Gradient methods, Gradient descent, etc.
  - "Intelligent" methods: particle swarm optimization, genetic algorithms, etc.

**Example 4 (Machine Learning: construction)** *:* *Let's suppose that the best regression function is $Y = \beta_0 + \beta_1 X$, then for the training dataset $(x_i, y_i)$ we need to minimize the MSE.*

- Half of ML field is construction: NN, SVM, etc.
- In DNN it is an optimization problem of millions of parameters.
- Let's see animation.
- Where are Probability, Statistics, and Linear Algebra here? Let's re-visit the chart.
- Is the optimization problem solvable:
  - closed form? (LSM)
  - numerically and guaranteed? (convex and linear)
  - numerically but not guaranteed? (non-convex):
    * numerical algorithms, e.g., GD,
    * local optimization,
    * heuristics, swarm, and genetics,
    * brute-force with exhaustive search

$$\underset{\beta_o, \beta_1}{\text{minimize}} \sum_i (\beta_o + \beta_1 x_i - y_i)^2$$



### 1.1.2 Solving Optimization Problems

- A *solution method* for a class of optimization problems is an algorithm that computes a solution.
- Even when the *objective function* and constraints are smooth, e.g., polynomials, the solution is very difficult.
- There are three classes where solutions exist, theory is very well developed, and amazingly found in many practical problems:

Linear $\subset$ Quadratic $\subset$ Convex $\subset$ Non-linear (not linear and not known to be convex!)

- For the first three classes, the problem can be solved very reliably in hundreds or thousands of variables!

## 1.2 Least-Squares and Linear Programming

### 1.2.1 Least-Squares Problems

A *least-squares* problem is an optimization problem with no constraints (i.e., $m = p = 0$), and an objective in the form:

$$\underset{x}{\text{minimize}}\, f_0(x) = \sum_{i=1}^{k} (a_i' x - b_i)^2 = \| A_{k\times n} x_{n\times 1} - b_{k\times 1} \|^2.$$

The solution is given in **closed form** by:

$$x = (A'A)^{-1} A' b$$

- Good algorithms in many SC SW exist; it is a very mature technology.
- Solution time is $O(n^2 k)$.
- Easily solvable even for hundreds or thousands of variables.
- More on that in the Linear Algebra course.
- Many other problems reduce to typical LS problem:
  - Weighted LS (to emphasize some observations)

$$\underset{x}{\text{minimize}}\, f_0(x) = \sum_{i=1}^{k} w_i (a_i' x - b_i)^2.$$

  - Regularization (to penalize for over-fitting)

$$\underset{x}{\text{minimize}}\, f_0(x) = \sum_{i=1}^{k} (a_i' x - b_i)^2 + \rho \sum_{j=1}^{n} x_j^2.$$

### 1.2.2 Linear Programming

A *linear programming* problem is an optimization problem with objective and all constraint functions are linear:

$$
\begin{array}{lll}
\underset{x}{\text{minimize}} & f_0(x) = C'x & \\
\text{subject to:} & a_i'x \leq b_i, & i = 1, \ldots, m \\
 & h_i'x = g_i, & i = 1, \ldots, p,
\end{array}
$$

- **No** closed form solution as opposed to LS.
- Very robust, reliable, and effective set of methods for numerical solution; e.g., Dantzig's simplex, and interior point.
- Complexity is $\simeq O(n^2 m)$.
- Similar to LS, we can solve a problem of thousands of variables.
- Example is *Chebyshev minimization* problem:

$$\underset{x}{\text{minimize}} \, f_0(x) = \max_{i=1,\ldots,k} |a_i'x - b_i|,$$

- The objective is different from the LS: minimize the maximum error. **Ex:**
- After some tricks, requiring familiarity with optimization, it is equivalent to a LP:

$$
\begin{array}{lll}
\underset{x}{\text{minimize}} & t & \\
\text{subject to:} & a_i'x - t \leq b_i, & i = 1, \ldots, k \\
 & -a_i'x - t \leq -b_i, & i = 1, \ldots, k
\end{array}
$$

## 1.3 Convex Optimization

A *convex optimization* problem is an optimization problem with objective and all constraint function are convex:

$$\begin{array}{llllr}
\underset{x}{\text{minimize}} & f_0(x) & & & \\
\text{subject to:} & f_i(x) \le 0, & & i = 1, \ldots, m & \\
& h_i(x) = 0, & & i = 1, \ldots, p, & \\
& f_i(\alpha x + \beta y) \le \alpha f_i(x) + \beta f_i(y), & \alpha + \beta = 1, & 0 \le \alpha,\ 0 \le \beta, & 0 \le i \le m \\
& h_i(x) = a_i' x + b_i & & 0 \le i \le p &
\end{array}$$

- The LP and LS are special cases; however, only LS has closed-form solution.
- Very robust, reliable, and effective set of methods, including *interior point methods.*
- Complexity is almost: $O\big(\max(n^3, n^2 m, F)\big)$, where $F$ is the cost of evaluating 1st and 2nd derivatives of $f_i$ and $h_i$.
- Similar to LS and LP, we can solve a problem of thousands of variables.
- However, it is not as very mature technology as the LP and LS yet.
- There are many practical problems that can be re-formulated as convex problem **BUT** requires mathematical skills; but once done the problem is solved. **Hint:** realizing that the problem is convex requires more mathematical maturity than those required for LP and LS.

## 1.4 Nonlinear Optimization

A *non-linear optimization* problem is an optimization problem with objective and constraint functions are non-linear **BUT** not known to be convex (**so far**). Even simple-looking problems in 10 variables can be extremely challenging. Several approaches for solutions:

**Local Optimization** : starting at initial point in space, using differentiablity, then navigate

- does not guarantee global optimal.
- affected heavily by initial point.
- depends heavily on numerical algorithm and their parameters.
- More art than technology.
- In contrast to convex optimization, where a lot of art and mathematical skills are required to formulate the problem as convex; then numerical solution is straightforward.

**Global Optimization** : the true global solution is found; the compromise is complexity.

- The complexity goes exponential with dimensions.
- Sometimes it is worth it when: the cost is huge, not in real time, and dimensionality is low.

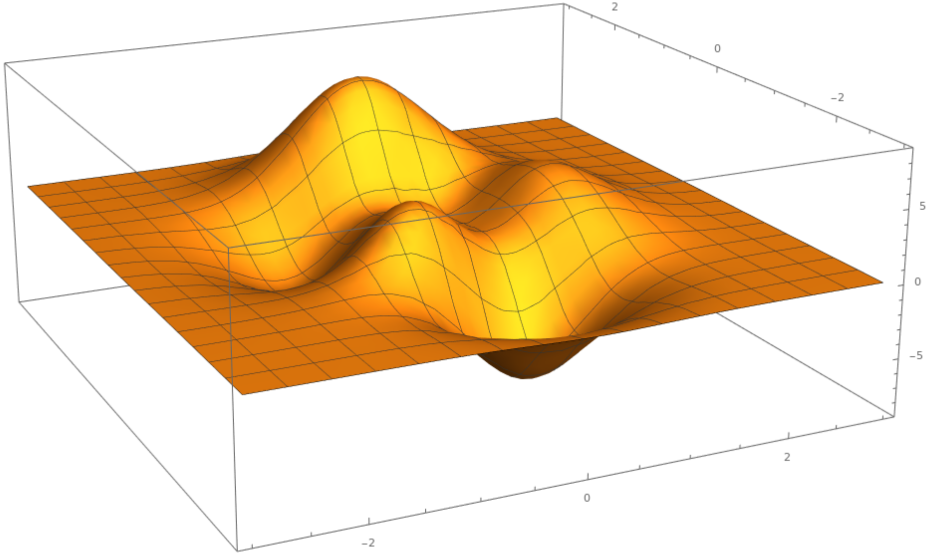
**Role of Convex Optimization** :

- Approximate the non-linear function to a convex one, finding the exact solution, then using it as a starting point for the original problem. (Also does not guarantee optimality)
- Setting bounds on the global solution.

**Evolutionary Computations** : Genetic Algorithm (GA), Simulated Annealing (SA), Particle Swarm Optimization (PSO), etc.

**Example 5 (Nonlinear Objective Function)** *: (Chong and Zak, 2013, Ex. 14.3, P.290)*

$$f(x,y) = 3(1-x)^2 e^{-x^2-(y+1)^2} - 10e^{-x^2-y^2}\left(-x^3 + \frac{x}{5} - y^5\right) - \frac{1}{3}e^{-(x+1)^2-y^2}$$



# Part I

# Theory

# Chapter 2

# Convex sets

# 2.1 Affine and convex sets

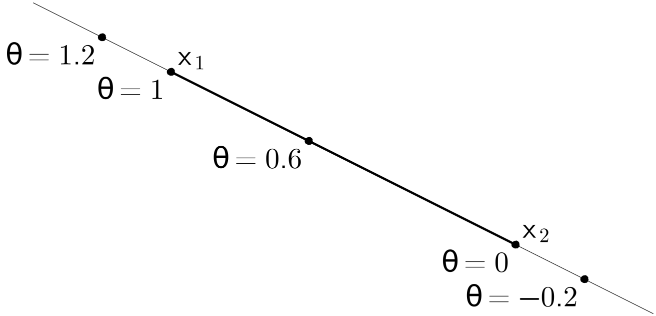
## 2.1.1 Lines and line segments

**Definition 6 (line and line segment)** *Suppose* $x_1 \neq x_2 \in \mathbf{R}^n$*. Points of the form*

$$\begin{aligned} y &= \theta x_1 + (1-\theta)x_2 \\ &= x_2 + \theta(x_1 - x_2), \end{aligned}$$

*where* $\theta \in \mathbf{R}$*, form the line passing through* $x_1$ *and* $x_2$*.*

- As usual, this is a definition for high dimensions taken from a proof for $n \leq 3$.
- We have done it many times: angle, norm, cardinality of sets, etc.
- if $0 \leq \theta \leq 1$ this forms a line segment.



### 2.1.2 Affine sets

**Definition 7 (Affine sets)** *A set $C \subset \mathbf{R}^n$ is affine if the line through any two distinct points in $C$ lies in $C$. I.e., $\forall x_1, x_2 \in C$ and $\theta \in \mathbf{R}$, we have $\theta x_1 + (1-\theta)x_2 \in C$. In other words, $C$ contains any linear combination (summing to one) of any two points in $C$.*

**Examples:** what about line, line segment, circle, disk, strip, first quadrant?

**Corollary 8** *Suppose $C$ is an affine set, and $x_1, \ldots, x_k \in C$, then $C$ contains every general affine combination of the form $\theta_1 x_1 + \ldots + \theta_k x_k$, where $\theta_1 + \ldots + \theta_k = 1$.*

**Wrong Proof.** Suppose $y_1, y_2 \in C$, then

$$x = \sum_{i=1}^{k} \theta_i x_i = \sum_{i=1}^{k} \theta_i(\alpha_i y_1 + (1-\alpha_i)y_2);$$

and the summation of the coefficients will be

$$\sum_{i=1}^{k} \theta_i \alpha_i + \sum_{i=1}^{k} \theta_i(1-\alpha_i) = \sum_{i=1}^{k} \theta_i(\alpha_i + 1 - \alpha_i) = \sum_{i=1}^{k} \theta_i = 1.$$

Where is the bug? ■

**Correct Proof.** **base:** $k = 3$.

$$\begin{aligned} x &= \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 \\ &= (1-\theta_3)\left(\frac{\theta_1}{1-\theta_3}x_1 + \frac{\theta_2}{1-\theta_3}x_2\right) + \theta_3 x_3. \\ &= (1-\theta_3)(\cdot \in C) + \theta_3(\cdot \in C). \end{aligned}$$

**induction:** suppose it is true for some $k \geq 3$; i.e., $\sum_{i=1}^{k} \theta_i x_i \in C$ when $\sum_{i=1}^{k} \theta_i = 1$ . Then

$$\begin{aligned} x &= \sum_{i=1}^{k+1} \theta_i x_i \\ &= \sum_{i=1}^{k} \theta_i x_i + \theta_{k+1} x_{k+1} \\ &= (1-\theta_{k+1}) \sum_{i=1}^{k} \theta_i / (1-\theta_{k+1}) x_i + \theta_{k+1} x_{k+1} \\ &= (1-\theta_{k+1})(\cdot \in C) + \theta_{k+1}(\cdot \in C), \\ &\qquad \text{(from the induction hypothesis)} \end{aligned}$$

which completes the proof. ■

**Definition 9 (Subspace from Linear Algebra)** *a set $V \subset \mathbf{R}^n$ of vector (here points) is a subspace if it is closed under sums and scalar multiplication. I.e., $\forall v_1, v_2 \in V$ and $\forall \alpha, \beta \in \mathbf{R}$ we have $\alpha v_1 + \beta v_2 \in V$.*

**Remember:**

- $\alpha + \beta$ not necessarily equals 1
- $\alpha = 0, \beta = 0 \rightarrow \mathbf{0} \in V$.
- Any subspace $V$ is the solution set of $A_{m\times n} x_{n\times 1} = 0$, which is $\mathcal{N}(A)$ (the null space of $A$). Geometry? I.e., $V = \{x \mid Ax = 0\}$
- $\mathbf{rank}(A) = n - \dim(V)$.

**Corollary 10** .

1. *If $C$ is affine, then $V = C - x_0 = \{x - x_0 \mid x, x_0 \in C\}$ is a subspace.*
2. *If $V$ is a subspace, then $C = V + x_0 = \{x + x_0 \mid x \in V\}$ is affine $\forall x_0$.*
3. *An affine set $C$ can be represented as the solution set of a nonhomogeneous linear system $Ax = b$, where $V = C - x_0$ is $\mathcal{N}(A)$.*
4. *The solution set of any nonhomogeneous system is an affine set. (Ex. 2.1)*

**Proof.**

1. Suppose $x_1, x_2, x_0 \in C$, an affine set. Both $x_1 - x_0$ and $x_2 - x_0$, by construction, $\in V$; then

$$\begin{aligned} x &= \alpha(x_1 - x_0) + \beta(x_2 - x_0) + x_0 \\ &= \alpha x_1 + \beta x_2 + (1 - \alpha - \beta)x_0 \in C \end{aligned}$$

Then $x - x_0 = \alpha(x_1 - x_0) + \beta(x_2 - x_0) \in V$; hence $V$ is a subspace.

2. Suppose $x_1, x_2 \in V$, a subspace. Both $x_1 + x_0$ and $x_2 + x_0$, by construction, $\in C$; then

$$\begin{aligned} x &= \theta(x_1 + x_0) + (1 - \theta)(x_2 + x_0) \\ &= \theta x_1 + (1 - \theta)x_2 + x_0 = (\cdot \in V) + x_0 \in C \end{aligned}$$

3. If $C$ is affine and $x_0 \in C$, then

$$\begin{aligned} C - x_0 &= \{x \mid Ax = 0\} \qquad \text{(since it is a subspace)} \\ C &= \{x + x_0 \mid A(x + x_0) = Ax_0\} \\ C &= \{c \mid Ac = b\}. \end{aligned}$$

4. $C = \{x \mid Ax = b\}$; if $x_0 \in C$ then $Ax_0 = b$ and

$$C - x_0 = \{x - x_0 \mid A(x - x_0) = b - Ax_0 = 0\}.$$

Hence, $C - x_0$ is a subspace and $C$ is affine.

■

**Proof of the book.** Suppose $x_1, x_2 \in C$, where $C = \{x \mid Ax = b\}$. Then

$$A(\theta x_1 + (1-\theta)x_2) = \theta A x_1 + (1-\theta) A x_2 = \theta b + (1-\theta) b = b,$$

which means $\theta x_1 + (1-\theta)x_2 \in C$ as well. ■

**Remark 1** *:*

- *The dimension of affine is defined to be the dimension of the associate subspace.*
- *affine is a subspace plus offset.*
- *every subspace is affine but not the vice versa; i.e., subspace is a special case of affine.*

**Definition 11 (affine hull)** *The "smallest" set of all affine combinations of some set $C$ (not necessarily affine) is called the affine hull* (**aff** $C$)*:*

$$\textbf{aff } C = \{\sum_{i=1}^{k} \theta_i x_i \mid x_i \in C,\ \sum_i \theta_i = 1\}.$$

**Corollary 12** **aff** $C$ *is affine.*

**Proof.** For $x_1 = \sum_i \alpha_i x_i$, $\sum_i \alpha_i = 1$, and $x_2 = \sum_i \beta_i x_i$, $\sum_i \beta_i = 1$, we have

$$\theta x_1 + (1-\theta) x_2 = \theta \sum_i \alpha_i x_i + (1-\theta) \sum_i \beta_i x_i = \sum_i (\theta \alpha_i + (1-\theta)\beta_i) x_i$$

$$\sum_i (\theta\alpha_i + (1-\theta)\beta_i) = \theta \sum_i \alpha_i + (1-\theta) \sum_i \beta_i = \theta + (1-\theta) = 1.$$

Hence, **aff** $C$ is affine as well. ■

**Example 13** *Construct the affine hull of the set* $C = \{(-1,0),(1,0),(0,1)\}$

$$\begin{aligned} \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 =& (1-\theta_3)\Big(\frac{\theta_1}{1-\theta_3} x_1 + \frac{\theta_2}{1-\theta_3} x_2\Big) + \theta_3 x_3 \\ =& (1-\alpha_3)\big((1-\alpha_2)x_1 + \alpha_2 x_2\big) + \alpha_3 x_3 \qquad = (1-\alpha_2)(1-\alpha_3)x_1 + \alpha_2(1-\alpha_3)x_2 + \alpha_3 x_3, \end{aligned}$$

$$\begin{array}{lll} \theta_3 = \alpha_3 & \theta_2 = \alpha_2(1-\alpha_3) & \theta_1 = 1-\theta_2-\theta_3 = (1-\alpha_2)(1-\alpha_3) \\ \alpha_3 = \theta_3 & \alpha_2 = \theta_2/(1-\theta_3) & \alpha_1 = 1-\alpha_2 = \theta_1/(1-\theta_3). \end{array}$$

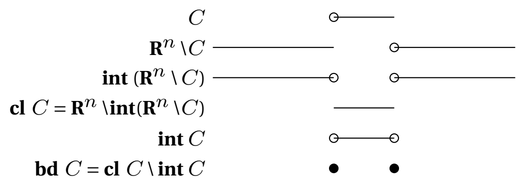
**HW:** Derive expressions for $\alpha_i$ and $\theta_i$ for $n$-point combination.

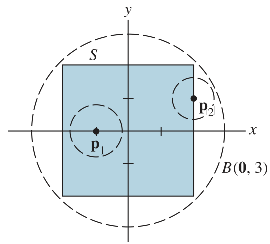
### 2.1.3 Affine dimension and relative interior

**Definition 14 (some basic topology in $\mathbf{R}^n$)** :

1. *The ball of radious $r$ and center $x$ in the norm $\|\cdot\|$.*
$$B(x,r) = \{y \mid \|y-x\| \le r\}.$$
2. *An element $x \in C \subseteq \mathbf{R}^n$ is called an interior point of $C$ if $\exists\, \varepsilon > 0$ for which*
$$B(x,\epsilon) = \{y \mid \|y-x\|_2 \le \varepsilon\} \subseteq C.$$
*I.e., $\exists$ a ball centered at $x$ that lies entirely in $C$.*
3. *The set of all points interior to $C$ is called the interior of $C$ and is denoted* **int** $C$.
4. *A set $C$ is open if* **int** $C = C$*. I.e., every point in $C$ is an interior point.*
5. *A set $C$ is closed if its complement is open*
$$\mathbf{R}^n \setminus C = \{x \in \mathbf{R}^n \mid x \notin C\}$$
6. *The closure of a set $C$ is defined as*
$$\mathbf{cl}\ C = \mathbf{R}^n \setminus \mathbf{int}(\mathbf{R}^n \setminus C).$$
7. *The boundary $C$ is defined as*
$$\mathbf{bd}\ C = \mathbf{cl}\ C \setminus \mathbf{int}\ C.$$

**Corollary 15** *A boundary point (a point $x \in$* **bd**$C$*) satisfies: $\forall \epsilon > 0, \exists\, y \in C$ and $z \notin C$ s.t. $y, z \in B(x,\epsilon)$.*




**Definition 16 (alter. equiv. def.)** :

- **int** $C$ *and* **bd** $C$ *are defined as 2,3, corollary.* *(It is obvious that:* **int** $C \cap$ **bd** $C = \phi$.*)*
- *$C$ is open if* **int** $C = C \quad \equiv \quad C \cap$ **bd** $C = \phi$.
- *$C$ is closed if* **bd** $C \subseteq C$.
- **cl** $C =$ **bd** $C \cup$ **int** $C$.

**Definition 17** *We define the affine dimension of a set $C$ as the dimension of its affine hull.*

**Example 18** *The unit circle in $\mathbf{R}^2$, i.e., $\{x \mid x_1^2 + x_2^2 = 1\}$ has affine hull of whole $\mathbf{R}^2$. So its affine dimension is 2. However, it has a dimensionality of 1 since it is parametric in just one parameter (manifold).*

**Definition 19** *We define the relative interior of the set $C$, denoted* **relint** $C$*, as its interior relative to* **aff** $C$

$$\textbf{relint } C = \{x \in C \mid B(x, r) \cap \textbf{aff } C \subseteq C \textit{ for some } r > 0\},$$

*and its relative boundary, denoted* **relbd** $C$ *is defined as*

$$\textbf{relbd } C = \textbf{cl } C \setminus \textbf{relint } C.$$

**Example 20** *Consider a square in the $(x_1, x_2)$-plane in $\mathbf{R}^3$, defined as:*

$$C = \{x \in \mathbf{R}^3 \mid -1 \le x_1 \le 1,\ -1 \le x_2 \le 1,\ x_3 = 0\}.$$

*Then:*

$$\begin{aligned}
\textbf{int } C &= \Phi \\
\textbf{cl } C &= \mathbf{R}^n \setminus \textbf{int}\left(\mathbf{R}^n \setminus C\right) = C \\
\textbf{bd } C &= \textbf{cl } C \setminus \textbf{int } C = C \\
\textbf{aff } C &= \{x \in \mathbf{R}^3 \mid x_3 = 0\} \\
\textbf{relint } C &= \{x \in \mathbf{R}^3 \mid -1 < x_1 < 1,\ -1 < x_2 < 1,\ x_3 = 0\} \\
\textbf{relbd } C &= \{x \in \mathbf{R}^3 \mid \max\{|x_1|, |x_2|\} = 1,\ x_3 = 0\}
\end{aligned}$$

### 2.1.4 Convex sets

**Definition 21 (convex set)** *A set $C$ is convex if the line segment between any two points in $C$ lies in $C$; i.e., if for any $x_1$, $x_2 \in C$ and any $\theta$ with $0 \le \theta \le 1$, we have*

$$\theta x_1 + (1-\theta)x_2 \in C.$$

**Corollary 22** *Suppose $C$ is convex set, and $x_1, \ldots, x_k \in C$, then $C$ contains every general convex combination (also called mixture); i.e.,*

$$\sum_i \theta_i x_i \in C,\ \sum_i \theta_i = 1,\ \theta_i \ge 0.$$

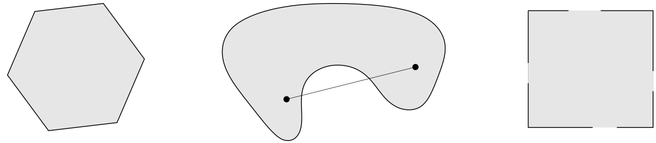
**Proof.** identical to proof of corollary 8. ■

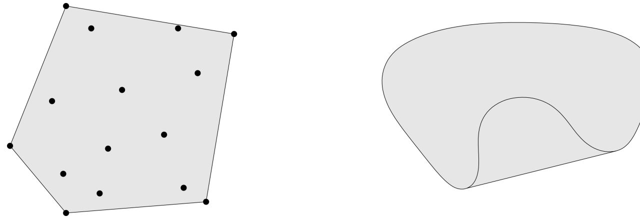
**Definition 23 (convex hull)** *The "smallest" set of all convex combinations of some set $C$ (not necessarily convex) is called the convex hull (**conv** $C$)*

$$\textbf{conv}\, C = \left\{ \sum_{i=1}^{k} \theta_i x_i \mid x_i \in C,\ \sum_i \theta_i = 1,\ \theta_i \ge 0 \right\}.$$

**Corollary 24** **conv** *$C$ is convex.*

**Proof.** identical to proof of corollary 12. ■





**Example 25** *Revisit example 13.*

**Example 26 (Applications)** *: Suppose $X \in C$ is a r.v., $C$ is convex. Then $\mathrm{E}\, X \in C$ if it exists:*

$$\mathrm{E}\, X = \sum_{i=1}^{n} p_i x_i$$

$$\mathrm{E}\, X = \sum_{i=1}^{\infty} p_i x_i$$

$$\mathrm{E}\, X = \int_C f_X(x) x\, dx \qquad \text{(Riemann sum)}$$

### 2.1.5 Cones

**Definition 27** *A set $C$ is called a cone (or nonnegative homogeneous) if $\forall\ x \in C,\ \theta \geq 0$ we have $\theta x \in C$; and it is a convex cone if it is convex in addition to being a cone.*

**Definition 28** *A point of the form $\sum_{i=1}^{k} \theta_i x_i,\ \theta_i \geq 0$ is called a conic combination.*

**Corollary 29** *A set $C$ is a convex cone if and only if it contains all conic combinations of its elements; i.e.,*

$$\sum_i \theta_i x_i \in C\ \forall x_i \in C \text{ and } \theta_i \geq 0.$$

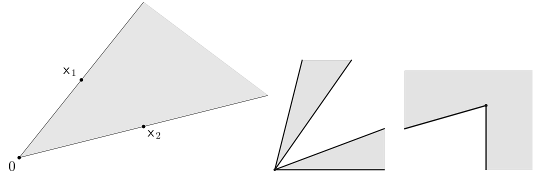
**Proof.**
**Sufficiency:** is obvious. Choosing $\sum_i \theta_i = 1$ implies $C$ is convex; and setting $\theta_i = 0\ \forall i > 1$ implies $C$ is cone.
**Necessity:** Since $C$ is convex cone, then $\forall x_i \in C, \theta_i \geq 0$ we have:

$$\theta_i x_i \in C \qquad \text{(cone)}$$

$$\sum_i (1/n)(\theta_i x_i) \in C \qquad \text{(convex)}$$

$$n \sum_i (1/n)(\theta_i x_i) = \sum_i \theta_i x_i \in C \qquad \text{(cone)}$$

■

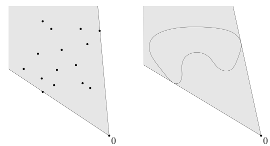



**Definition 30** *A conic hull of a set $C$ is the minimum set of all conic combination:*

$$\textbf{cone } C = \{\sum_i \theta_i x_i \mid x_i \in C,\ \theta_i \geq 0,\ i = 1, \ldots, n\}.$$

**Corollary 31** **cone** *$C$ is convex cone.*

**Proof.** If $y \in \textbf{cone } C, \alpha \geq 0$, then $\alpha y = \alpha \sum_i \theta_i x_i = \sum_i (\alpha\theta_i) x_i \in \textbf{cone } C$. And if $y_1, y_2 \in \textbf{cone } C$ then $\alpha y_1 + (1-\alpha) y_2 = \alpha \sum_i \theta_i x_i + (1-\alpha) \sum_i \mu_i x_i = \sum_i (\alpha\theta_i + (1-\alpha)\mu_i) x_i \in \textbf{cone } C$ ■



## 2.2 Some important examples

### Fast Revision

- Each of the sets: $\phi$, $x_0$ (a singleton), $\mathbf{R}^n$ are affine and convex.
- Any line is affine. If it passes through zero, it is a subspace and a convex cone.
- Any subspace is convex cone but not vise versa.
- A line segment is convex, but not affine (unless it reduces to a singleton).
- A *ray*, $\{x_0 + \theta v \mid \theta \geq 0, v \neq 0\}$ is convex but not affine. It is convex cone if $x_0 = 0$.

### 2.2.1 Hyperplanes and halfspaces

**Definition 32** *A hyperplane is a set of the form*

$$\begin{aligned} \mathcal{S} &= \{x \mid a'x = b\}, && a, b \in \mathbf{R}^n,\ a \neq 0 \\ &\equiv \{x \mid a'(x - x_0) = 0\}, && a'x_0 = b. \end{aligned}$$

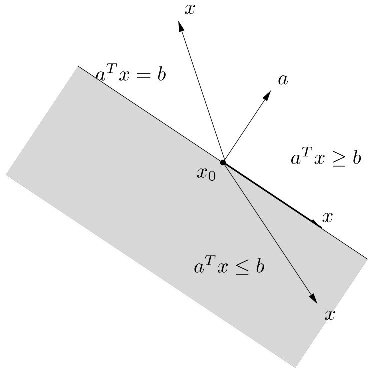
- Vectors with inner product with $a$ is $b$: $\frac{a'}{\|a\|}x = \frac{b}{\|a\|}$. I.e., from $\mathbf{0}$, walk a distance $\frac{b}{\|a\|}$ (either + or -) in the direction of $a$, then draw perpendicular line.

**Definition 33** *A closed halfspace is the region generated by the hyperplane and defined as:*

$$\begin{aligned} \mathcal{H} &= \{x \mid a'x \leq b\}, && a, b \in \mathbf{R}^n,\ a \neq 0 \\ &\equiv \{x \mid a'(x - x_0) \leq 0\}, && a'x_0 = b. \end{aligned}$$

- region of all vectors with projection $< b/\|a\|$.
- Vectors with obtuse angle with $a$: ($\cos\theta = \frac{a'x}{\|a\|\|x\|}$).
- Line passing with $p_0$ and $\perp$ on $\mathcal{S}$:

$$\begin{aligned} x &= p_0 + \theta\overline{a} && \text{(parametric eq.)} \\ a'x_0 &= a'p_0 + \theta_0\|a\| \\ \theta_0 &= (b - a'p_0)/\|a\| && (x_0 \text{ pt. of intersection.}) \\ x_0 - p_0 &= \frac{(b - a'p_0)}{\|a\|}\overline{a}. \end{aligned}$$




**Corollary 34** *$\mathcal{S}$ is affine, $\mathcal{H}$ is convex and not affine,* **int** *$\mathcal{H} = \mathcal{H} \setminus \mathcal{S}$, and* **bd** *$\mathcal{H} = \mathcal{S}$.*

**Proof.** $\mathcal{S}$ is affine done. $\mathcal{H}$ is convex: take $0 \leq \theta \leq 1$ $\theta a'x_1 + (1-\theta)a'x_2 \leq \theta b + (1-\theta)b = b$. (why not affine?!)

$$\begin{aligned} y &= x + ru, \quad 0 \leq \|u\| \leq 1 && (y \in B(x, r)) \\ a'y &= a'x + ra'u = b - (b - a'x) + r\|a\|\|u\|\cos(a, u) \end{aligned}$$

If $b = a'x$, i.e., $x \in \mathcal{S}$, $a'u > 0$ or $< 0$ (depending on the angle) and hence $a'y > b$ or $< b$. Then $\mathcal{S} \subseteq$ **bd** $\mathcal{H}$.

If $a'x < b$, i.e., $x \in \mathcal{H} \setminus \mathcal{S}$, $\exists\, r < \frac{b - a'x}{\|a\|}$, s.t. $a'y < b$. Hence: **int** $\mathcal{H} = \mathcal{H} \setminus \mathcal{S}$ and **bd** $\mathcal{H} = \mathcal{S}$. ■

### 2.2.2 Euclidean balls and ellipsoids

**Definition 35** *A Euclidean ball in* $\mathbf{R}^n$ *is the set:*

$$\begin{aligned} B(x_c, r) &= \left\{x = x_c + ru \mid \|u\|_2 \le 1\right\} \\ &= \left\{x \mid \|x - x_c\|_2 / r \le 1\right\} \\ &= \left\{x \mid (x - x_c)'(x - x_c)/r^2 \le 1\right\}. \end{aligned}$$

**Definition 36** *Ellipsoid in* $\mathbf{R}^n$ *is the set:*

$$\begin{aligned} \mathcal{E} &= \left\{x = x_c + Au \mid \|u\|_2 \le 1,\ A \succ 0\right\} \\ &= \left\{x \mid \|A^{-1}(x - x_c)\|_2 \le 1, A \succ 0\right\} \\ &= \left\{x \mid (x - x_c)'(A^{-1})'A^{-1}(x - x_c) \le 1\right\} \end{aligned}$$

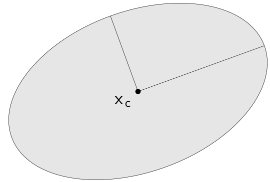
**Spectral decomposition for** $A = A'$.

$$\begin{aligned} Au &= (\lambda_1 v_1 v_1' + \lambda_2 v_2 v_2' + \cdots + \lambda_n v_n v_n')u \\ &= \lambda_1 v_1 (v_1' u) + \lambda_2 v_2 (v_2' u) + \cdots + \lambda_n v_n (v_n' u), \end{aligned}$$

which reduces to a Ball when $\lambda_i = r$.

**Remark 2** *A does not have to be symmetric; put:*

$$\begin{aligned} P^{-1} &= (A^{-1})'A^{-1} = V\Sigma^{-1/2}\Sigma^{-1/2}V' && \textit{symmetric} \\ P^{1/2}u_2 &= Au_1 && \textit{is bijection} \\ \|u_2\|^2 &= u_1'A'P^{-1/2}P^{-1/2}Au_1 = \|u_1\|^2 \end{aligned}$$




**Remark 3 (Contours of $\mathcal{N}(\mu, \Sigma)$)** *:*

$$f_X(x) = \frac{1}{((2\pi)^p |\Sigma|)^{1/2}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}$$

**Corollary 37** *An ellipsoid, hence a ball, is convex*

**Proof.** For $x_1, x_2 \in \mathcal{E}, 0 \le \theta \le 1$,

$$\begin{aligned} x_1 &= x_c + Au_1, \|u_1\| \le 1 \\ x_2 &= x_c + Au_2, \|u_2\| \le 1 \\ x &= \theta(x_c + Au_1) + (1 - \theta)(x_c + Au_2) \\ &= x_c + A(\theta u_1 + (1 - \theta)u_2) \\ \|u\| &= \|(\theta u_1 + (1 - \theta)u_2)\| \\ &\le \theta\|u_1\| + (1 - \theta)\|u_2\| \\ &\le \theta + (1 - \theta) = 1. \end{aligned}$$

■

### 2.2.3 Norm balls and norm cones

**Definition 38 (Norm)** *Let $x, y \in \mathbf{R}^n$, $t \in \mathbf{R}$; a function $f : \mathbf{R}^n \mapsto \mathbf{R}_+$ with* **dom** *$f = \mathbf{R}^n$ is called a norm if*

1. $f(x) = 0 \to x = 0$ *(definite)*
2. $f(tx) = |t| f(x)$ *(homogeneous)*
3. $f(x+y) \le f(x) + f(y)$ *(triangle inequality)*

*$f(0) = 0$ is implied from (2)* *(positive definite)*

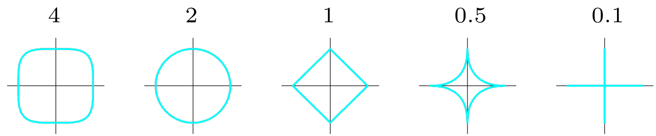
**Definition 39 ($L^p$-norm ($\|\cdot\|_p$))** *is defined as*

$$\|x\|_p = \Big(\sum_{i=1}^{n} |x_i|^p\Big)^{1/p} = \big(|x_1|^p + \cdots + |x_n|^p\big)^{1/p}.$$

**Proof of $\|\cdot\|_p$ is a norm.** :

1. $\Big(\sum_{i=1}^{n} |x_i|^p\Big)^{1/p} = 0 \to \sum_{i=1}^{n} |x_i|^p = 0 \to x_i = 0.$
2. $\|tx\|_p = \Big(\sum_{i=1}^{n} |tx_i|^p\Big)^{1/p} = |t|\Big(\sum_{i=1}^{n} |x_i|^p\Big)^{1/p} = |t|\|x\|_p$
3. $\|x+y\|_p \le \|x\|_p + \|y\|_p$ :
   (Minkowski's inequality proof (Kreyszig, 1978))

**counter example for** $p < 1$: $\|(0,1)\|_{1/2} + \|(1,0)\|_{1/2} = 1 + 1 = 2$, whereas $\|(1,1)\|_{1/2} = (1+1)^{1/(1/2)} = 4.$ ■




$$\{x \in \mathbf{R}^2 \mid \|x\|_p = c,\ p = 4, 2, 1, 0.5, 0.1\}$$

- $L_1$-norm, Manhatan dist., Taxicab, abs. value

$$\|x\|_1 = \Big(\sum_{i=1}^{n} |x_i|\Big).$$

- $L_2$-norm, Euclidean distance (most meaningful)

$$\|x\|_2 = \Big(\sum_{i=1}^{n} |x_i|^2\Big)^{1/2}.$$

- $L_\infty$-norm

$$\|x\|_\infty = \lim_{p\to\infty}\Big(\sum_{i=1}^{n} |x_i|^p\Big)^{1/p} = \max_i |x_i|$$

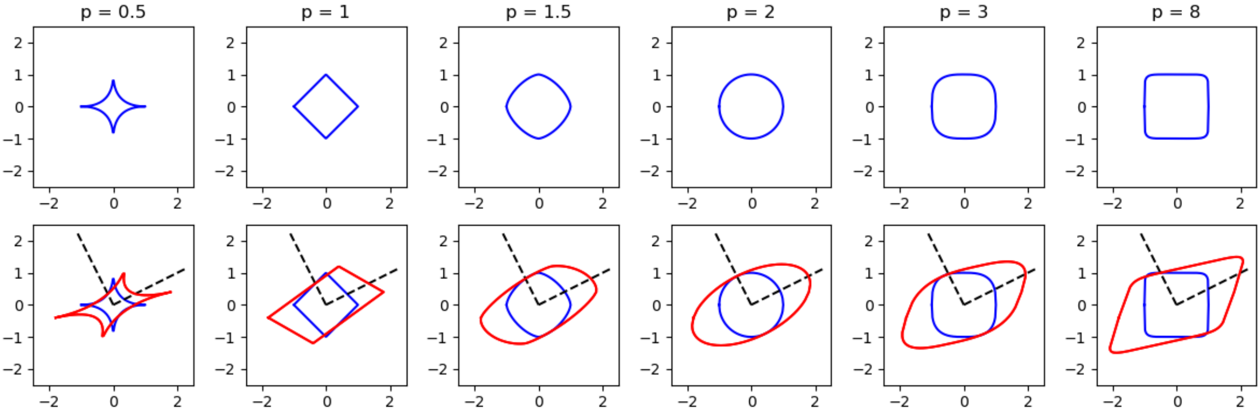
**Corollary 40 (properties of $\|\cdot\|_p$)** :

1. *$|x_i| < \|x\|_p \ \forall p < \infty$.*
2. *$L_P$-norm is monotonic in $p$.*

**Proof.** is HW. ■

**Definition 41 (A general norm ellipsoid in $\mathbb{R}^n$)** *is the set generated by a norm ball, for any norm $\|\cdot\|$, of radius $r$, centered at $x_c$, and transformed by any symmetric matrix $A \succ 0$:*

$$\mathcal{E} = \left\{x = x_c + Au \mid \|u\|_p \le 1,\ A \succ 0\right\} \equiv \left\{x \mid \|A^{-1}(x - x_c)\|_p \le 1,\ A \succ 0\right\}. \tag{2.1}$$




- $A = \lambda_1 v_1 v_1' + \lambda_2 v_2 v_2',\ v_1 = (2,1)'/\sqrt{5},\ v_2 = (-1,2)'/\sqrt{5},\ \lambda_1 = 2,\ \lambda_2 = 1.$
- The unit ball intersect with the ellipsoid at $v_2$; why? The ellipsoids, of course, no longer have unit $L_p$-norm.

**Corollary 42** *The general ellipsoid* (2.1) *is convex.*

**Definition 43 (Metric)** *Let $x, y, z \in S$, a function $\delta : S \times S \mapsto \mathbf{R}_+$ is called a metric on $S$ if:*

1. $\delta(x,y) = 0 \leftrightarrow x = y$ *(positive definite)*
2. $\delta(x,y) = \delta(y,x)$ *(symmetric)*
3. $\delta(x,y) \le \delta(x,z) + \delta(z,y)$ *(triangle inequality)*

**Lemma 44** *: $\delta(x,y) = f(x-y)$ is a metric ($\delta(x,0) = f(x)$):*

$$\begin{aligned}
\delta(x,y) &= f(x-y) = 0 \leftrightarrow x-y=0 \leftrightarrow x=y\\
\delta(x,y) &= f(x-y) = f(-1(y-x)) = f(y-x) = \delta(y,x)\\
\delta(x,y) &= f(x-y) = f((x-z)+(z-y))\\
&\le f(x-z) + f(z-y) = \delta(x,z) + \delta(y,z)
\end{aligned}$$

**Definition 45 (Loss)** *Let $x \in S_1$ (called set of nature), and $y \in S_2$ (called set of actions); then a function $L : S_1 \times S_2 \mapsto \mathbf{R}_+$ is called a loss incurred from assigning the action $y$ based on the truth of nature $x$. For details on loss and utility theory see Berger (1993).*

**Remark 4** *:*

- *Not any metric defines a norm; e.g., $\delta(x,y) = I_{x \neq y}$: First, prove it is a metric (HW). Then: $\delta(x,0) = f(x) = 1 \neq 10 = f(10x) = \delta(10x,0) \quad \forall x \neq 0$.*
- *Why? metric suites any set even categorical.*
- *Loss do not have to follow metric properties at all; e.g., $L(P,N) \neq L(N,P)$ in medical classification prediction.*

**Definition 46** *The norm cone associated with any norm $\|x\|$, $x \in \mathbf{R}^n$ is the set*

$$C = \left\{(x,t)' \in \mathbf{R}^{n+1} \mid \|x\| \le t,\ t > 0\right\}$$

**Example 47** *The second-order cone is the norm cone for the Euclidean norm; i.e.,*

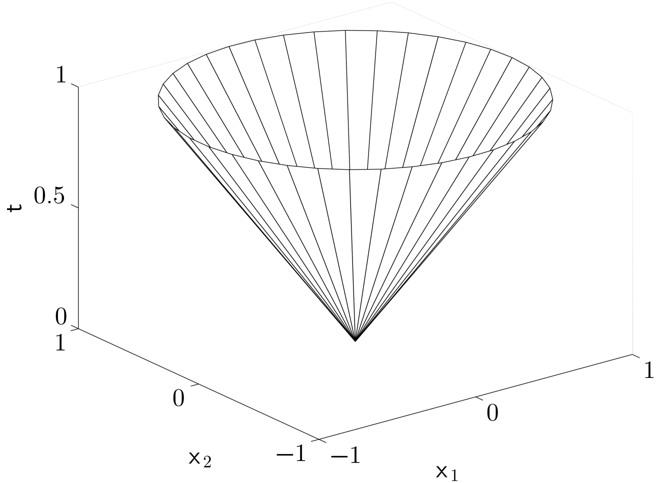
$$\begin{aligned} C &= \left\{(x,t)' \in \mathbf{R}^{n+1} \mid \|x\|_2 \le t\right\} \\ &= \left\{(x,t)' \mid x'x \le t, t > 0\right\} \end{aligned}$$

**Corollary 48** *The norm cone is convex.*

**Proof.** : given $p_i = (x_i, t_i)$, $\|x_i\| \le t_i$, $i = 1, 2$, then

$$\begin{aligned} p &= \theta(x_1,t_1) + (1-\theta)(x_2,t_2) = (x,t) \\ &= (\theta x_1 + (1-\theta)x_2,\ \theta t_1 + (1-\theta)t_2) \\ \|x\| &= \|\theta x_1 + (1-\theta)x_2\| \\ &\le \theta\|x_1\| + (1-\theta)\|x_2\| \\ &\le \theta t_1 + (1-\theta)t_2 = t. \end{aligned}$$

■




To imagine it: pay attention to that the radius is the same as the height $t$. Therefore, the cross section is convex and in the $z$-direction is convex as well. E.g.,

$$C = \left\{(x,t^2) \in \mathbf{R}^{n+1} \mid \|x\| \le t,\ t > 0\right\}$$

is convex; however:

$$C = \left\{(x,\sqrt{t}) \in \mathbf{R}^{n+1} \mid \|x\| \le t,\ t > 0\right\}$$

is not.

### 2.2.4 Polyhedra

(Remember the early definition 1.2.2).

**Definition 49** *A polyhedron is defined as the solution set of a finite number of linear qualities and inequalities:*

$$\mathcal{P} = \{x \mid a_j' x \le b_j,\ j = 1, \cdots, m, \quad c_j' x = d_j,\ j = 1, \cdots, p\}.$$

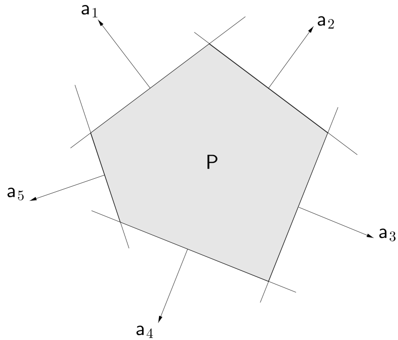
*For short notation, we write*

$$\mathcal{P} = \{x \mid Ax \preceq b,\ Cx = d\}, \quad A = \begin{pmatrix} a_1' \\ \vdots \\ a_m' \end{pmatrix}, \quad C = \begin{pmatrix} c_1' \\ \vdots \\ c_m' \end{pmatrix}.$$

*The polyhedron is called polytope if it is bounded.*

**Example 50** *The nonnegative orthant*

$$\mathbf{R}_+^n = \{x \in \mathbf{R}^n \mid x_i \ge 0,\ i = 1, \cdots, n\} = \{x \in \mathbf{R}^n \mid x \succeq 0\}$$



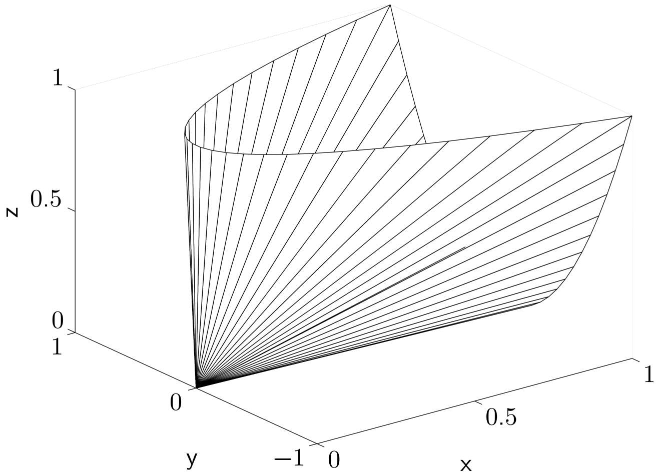

**Corollary 51** *The polyhedron is convex.*

**Proof.** $x_1, x_2 \in \mathcal{P}$, $x = \theta x_1 + (1-\theta)x_2$, $0 \le \theta \le 1$. Then:

$$\begin{aligned} a_j' x &= \theta a_j' x_1 + (1-\theta) a_j' x_2 \\ &\le \theta b_j + (1-\theta) b_j = b_j \\ c_j' x &= \theta c_j' x_1 + (1-\theta) c_j' x_2 \\ &= \theta d_j + (1-\theta) d_j = d_j. \end{aligned}$$

Hence, all conditions are satisfied; the proof is complete. ■

### 2.2.5 The positive semidefinite cone



## 2.3 Operations that preserve convexity

## 2.4 Generalized inequalities

## 2.5 Separating and supporting hyperplanes

## 2.6 Dual cones and generalized inequalities

# Part II

# Applications

# Part III

# Algorithms

# Bibliography


Bagirov, A. (2014), *Introduction to nonsmooth optimization : theory, practice and software*, Cham: Springer.

Berger, J. O. (1993), *Statistical decision theory and Bayesian analysis*, New York: Springer-Verlag, 2nd ed.

Bertsekas, D. (2003), *Convex analysis and optimization*, Belmont, Mass: Athena Scientific.

Borwein, J. (2006), *Convex analysis and nonlinear optimization : theory and examples*, New York: Springer.

Boyd, S. and Vandenberghe, L. (2004), *Convex Optimization*, Cambridge: Cambridge University Press.

Chong, E. K. and Zak, Stanislaw, H. (2013), *An Introduction to Optimization*, Wiley-Interscience, 4th ed.

Dattorro, J. (2009), *Convex Optimization & Euclidean Distance Geometry*, California: $\mathcal{M}\epsilon\beta OO$ Publishing.

Hiriart (2001), *Fundamentals of convex analysis*, Berlin New York: Springer.

Kreyszig, E. (1978), *Introductory functional analysis with applications*, New York: Wiley.

Luenberger, D. (1968), *Optimization by vector space methods*, New York: Wiley.

Rockafellar, R. (1997), *Convex analysis*, Princeton, N.J: Princeton University Press.

Rudin, W. (1976), *Principles of mathematical analysis*, New York: McGraw-Hill, 3rd ed.

Wade, W. R. (2000), *An introduction to analysis*, Upper Saddle River, NJ: Prentice Hall, 2nd ed.