EE 227A: Convex Optimization and Applications	**January 17**

Lecture 1: **Introduction**

Lecturer: Anh Pham

Reading assignment: Chapter 1 of BV[1].

Course outline and organization

- Course web page: `http://www.eecs.berkeley.edu/~elghaoui/Teaching/EE227A/index.html`

  We will use bspace (to post homeworks) and Piazza.

- Textbook (BV) freely available at

  `http://www.stanford.edu/~boyd/cvxbook/`

  Background material is also available at

  `https://inst.eecs.berkeley.edu/~ee127a/book/login/index.html`

  The login and password for this URL are both set to `ee127a-web`.

- Instructor: L. El Ghaoui (`elghaoui@eecs`). Office hours: Thu 2-3pm, 421 SDH.

- The course is a 4-unit course.

- Homeworks: there will be about 5 homeworks, which require the use of Matlab. Typically the HW is given on a Wednesday, and due in class two weeks later. If you turn your homework late, there will be a 2 point penalty per day, with a maximum delay until the next class following the due date (that is, if a HW is due on a Wednesday, you have to turn it on the following Friday at the latest).

- There will be about 5 quizzes, one project, and no final. Instructions on the project fill be given soon.

1.1 Optimization problems

1.1.1 The model

The mathematical programming model is of the form

$$p^* := \min_x \; f_0(x) \text{ subject to } f_i(x) \leq 0, \;\; i = 1, \ldots, m. \tag{1.1}$$

[1]In the course, the acronym refers to the book *Convex optimization,* by S. Boyd and L. Vandenberghe, Cambridge, 2004.

- $x \in \mathbf{R}^n$ is the *decision variable*
- $f_0 : \mathbf{R}^n \to \mathbf{R}$ is the *objective* function.
- $f_i : \mathbf{R}^n \to \mathbf{R}$, $i = 1, \ldots, m$ represent the *constraints*.
- p^* is the *optimal value*.

The term "programming" (or "program") does not refer to a computer code. It is used mainly for historical purposes. A more rigorous (but less popular) term is "optimization problem". The term "subject to" is often replaced by a colon.

The set

$$\mathcal{D} = \{x \in \mathbf{R}^n \ : \ f_i(x) \leq 0, \ \ i = 1, \ldots, m\}$$

is called the *feasible set*. Any $x \in \mathcal{D}$ is called feasible (with respect to the specific optimization problem at hand).

Sometimes, the model is described in terms of the feasible set, as follows:

$$\min_{x \in \mathcal{D}} \ f_0(x).$$

Also, sometimes a maximization problem is considered:

$$\max_{x \in \mathcal{D}} \ f_0(x).$$

Finally, it may be useful to distinguish between "structure" constraints (such as non-negativity constraints on variables) and constraints involving problem data, with the notation

$$\min_{x \in \mathcal{X}} \ f_0(x) \text{ subject to } f_i(x) \leq 0, \ \ i = 1, \ldots, m,$$

where $\mathcal{X} \subseteq \mathbf{R}^n$ describes the "structure" constraints.

1.1.2 Examples

- Least-squares problem:

  $$\min_x \ \|Ax - b\|_2^2$$

  where $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, and $\| \cdot \|_2$ denotes the Euclidean norm.

  This problem arises in many situations, for example in statistical estimation problems such as linear regression. The problem dates back many years, at least to Gauss (1777-1855), who solved it to predict the trajectory of the planetoid Ceres.

- Linear programming problem:

  $$\min \ c^T x \ : \ a_i^T x \leq b_i, \ \ i = 1, \ldots, m,$$

  where $c \in \mathbf{R}^n$, $a_i \in \mathbf{R}^n$, $b_i \in \mathbf{R}$ $(i = 1, \ldots, m)$. This corresponds to the case where the functions f_i $(i = 0, \ldots, m)$ in (1.1) are all affine (that is, linear plus a constant term).

  This problem was introduced by Dantzig in the 40's in the context of logistical problems arising in military operations. This model of computation is perhaps the most widely used optimization problem today.

- Quadratic programming problem:

$$\min \ \|x\|_2^2 + c^T x \ : \ a_i^T x \le b_i, \ \ i = 1, \ldots, m,$$

  which can be thought of as a generalization of both the least-squares and linear programming problems.

  QP's are popular in many areas, such as finance, where the linear term in the objective refers to the expected negative return on an investment, and the squared term corresponds to the risk (or variance of the return).

  This model was introduced by Markowitz (who was a student of Dantzig) in the 50's, to model investment problems. Markowitz won the Nobel prize in Economics in 1990, mainly for this work.

1.1.3 Solution

The *optimal set* of problem (1.1) is defined as the set of *feasible* points x^* such that $p^* = f_0(x^*)$:

$$\mathcal{X}_{\text{opt}} := \{x \in \mathbf{R}^n \ : \ f_i(x) \le 0, \ \ i = 1, \ldots, m, \ \ p^* = f_0(x)\}.$$

The ϵ-suboptimal set is defined as

$$\mathcal{X}_\epsilon := \{x \in \mathbf{R}^n \ : \ f_i(x) \le 0, \ \ i = 1, \ldots, m, \ \ f_0(x) \le p^* + \epsilon\}.$$

(With our notation, $\mathcal{X}_0 = \mathcal{X}_{\text{opt}}$.)

A point z is *locally optimal* if there is a value $R > 0$ such that z is optimal for problem

$$\min_x \ f_0(x) \text{ subject to } f_i(x) \le 0, \ \ i = 1, \ldots, m, \ \ \|z - x\|_2 \le R.$$

In other words, x minimizes f_0, but only for nearby points on the feasible set.

1.1.4 Convexity

A set $\mathcal{C}$ is convex if it contains the line segments between any two of its points:

$$\forall\, x, y \in \mathcal{C}, \ \ \forall\, \lambda \in [0,1], \ \ \lambda x + (1-\lambda) y \in \mathcal{C}.$$

A function $f : \mathbf{R}^n \to \mathbf{R}$ is *convex* if

$$\forall\, x, y \in \mathbf{R}^n, \ \ \forall\, \lambda \in [0,1], \ \ f(\lambda x + (1-\lambda) y) \le \lambda f(x) + (1-\lambda) f(y).$$

In other words, the graph of the function is always below the chord joining any two points. That is, a function is convex if and only if its *epigraph*

$$\mathbf{epi} f := \left\{(x,t) \in \mathbf{R}^{n+1} \ : \ t \ge f(x)\right\}$$

is convex. (Check this statement.)

The optimization problem (1.1) is convex if every function involved $f_0, f_1, \ldots, f_m$, is convex.

The examples presented in section (1.1.2) are all convex. Examples of non-convex problems include *combinatorial optimization* problems, where (some if not all) variables are constrained to be boolean, or integers. (Such problems arise for example when discrete choices are to be made, such as in crew assignment in the airline industry.)

1.1.5 Complexity

In this course, complexity of an optimization problem refers to the difficulty of solving the problem on a computer. At this stage we do not define this notion precisely.

The complexity of an optimization problem depends on its structure. Two seemingly similar problem may require a widely different computational effort to solve. Some problems are "NP-hard", which roughly means that they cannot be solved in reasonable time on a computer.

As an example, the quadratic programming problem seen above is "easy" to solve, however the apparently similar problem

$$\min\ c^T x - \|x\|_2^2\ :\ a_i^T x \leq b_i,\ \ i = 1, \ldots, m,$$

is NP-hard.

In the early days of optimization, it was thought that linearity was what distinguished a hard problem from an easy one. Today, it appears that convexity is the relevant notion. Roughly speaking, a convex problem is easy. In this course, we will refine this statement.

1.2 A brief history of convex optimization

Theory:

- 19-th century: optimization models are used mostly in physics, with the concept of energy as the objective function. No attempt (with the notable exception of Gauss' algorithm for least-squares) is made to actually solve these problems numerically.
- 1900-70: a great effort in made in mathematics to build the theory of optimization. The emphasis is on convex analysis, which allows to describe the optimality conditions of a convex problem.
- 1990's: theory of semi-definite programming, interactions with algebraic geometry.

Algorithms:

- 1947: simplex algorithm for linear programming (Dantzig).
- 1960s: early interior-point methods (Fiacco & McCormick, Dikin, ...).
- 1970s: ellipsoid method and other subgradient methods
- 1980s: polynomial-time interior-point methods for linear programming (Karmarkar 1984).
- late 1980s–now: polynomial-time interior-point methods for nonlinear convex optimization (Nesterov & Nemirovski 1994).
- Since 2000, attention was given to first-order methods, which are slower than interior-point (second-order) methods, but scale much better with problem size.

Applications (of convex optimization):

- before 1990: mostly in operations research; few in engineering (except least-squares); statistics is a big user of nonlinear optimization methods such as Newton-Raphson.
- since 1990: many applications: machine learning & statistics, control, signal processing, communications, circuit design, etc; new problem classes (semidefinite and second-order cone programming, robust optimization).

1.3 Course objectives

The course will emphasize models, not algorithms. It will cover a specific class of optimization models, based on the notion of convexity. We will briefly cover some algorithms for solving convex problems. We will also study robust optimization, which is a method to handle uncertainty in decision-making. We will also cover algorithms for large-scale convex optimization.

Exercises

1. *Linear objective is universal.* Show that we can always re-formulate any optimization problem as another optimization problem with linear objective. If the original problem is convex, is the new problem convex as well?

2. *Slalom problem.* Consider a two-dimensional skier that goes down a slope at constant velocity. At specific points along the slope, indexed by positions x_i, $i = 1, \ldots, n$, the skier has to pass through gates of widths $\sigma_i > 0$ and middle position y_i, $i = 1, \ldots, n$. We assume that the skier comes from uphill situated very far away from the start of the gate, with its initial direction set at a given angle.

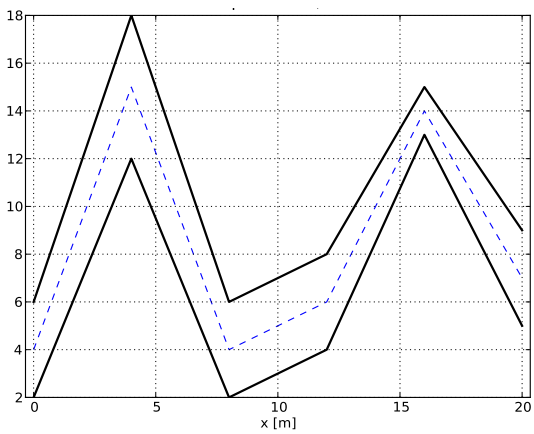
   Find the path that minimizes the time to reach the bottom of the slope. Your answer should come in the form of an optimization problemn. (Fig. 1.1 illustrates the problem, with "downhill" to the right.)

3. *Asset liability management.* A company faces the following *net cash flow requirements:*

Month	Jan	Feb	Mar	Apr	May	Jun
Net cash flow (in \$ k)	-150	-100	200	-200	50	300

   To meet these cash-flow requirements, there are several sources of funds: a) a line of credit (with maximum at \$100k, and interest rate 1% per month); b) it can issue a bond, in the form of a 90-day commercial paper bearing a total interest of 2% for the 3-month period; c) excess funds available can be invested at 0.3% per month.

   The variables in this decision problem (with decision horizon from January to June) are:



i	x_i	y_i	σ_i
0	0	4	2
1	4	5	3
2	8	4	2
3	12	6	2
4	16	5	1
5	20	7	2

Tab. 1: Sample course.

Figure 1.1: Slalom problem with $n = 5$ obstacles. Corresponding data on the right.

- Balance on the credit line x_i for month $i = 1, 2, 3, 4, 5$.
- Amount y_i of commercial paper issued ($i = 1, 2, 3$).
- Excess funds z_i for month $i = 1, 2, 3, 4, 5$.
- z_6, the company's wealth in June.

Finally, the *decision problem* is to maximize the terminal wealth (z_6) subject to cash-flow balance equations (as determined by the cash-flow requirements), and the bounds on variables. Show that the problem can be formulated as a linear program, which you will determine.