Introduction to convex optimization I

Sergio García[1]

The University of Edinburgh, UK

June 2018

[1] Based on lecture notes by Dr Paresh Date, Brunel University

Outline

- Introduction to convex problems

Outline

- Introduction to convex problems
- Special classes of convex problems
    1. Linear programming
    2. Convex quadratic programming

The convex optimization problem I

- The problems of interest are of the form

$$\begin{aligned} &\text{minimize} \quad f_0(\mathbf{x}), \\ &\text{subject to } f_i(\mathbf{x}) \le 0,\ i = 1, 2, \ldots, m, \\ &\text{and } h_i(\mathbf{x}) = 0,\ i = 1, 2, \ldots, p, \end{aligned} \tag{1}$$

  where the functions $f_i : \mathrm{dom}\,(f_i) \supseteq \mathbb{R}^n \mapsto \mathbb{R},\ i = 0, 1, 2, \ldots, m$ are convex; $h_i(\mathbf{x}) = \mathbf{a}_i^\top \mathbf{x} - \mathbf{b}_i,\ i = 1, 2, \ldots, p$ are affine.

- Maximization of a concave function subject to convex constraints is also a convex optimization problem.

The convex optimization problem II

- The set
$$\mathcal{D} = \bigcap \text{dom}(f_i) \bigcap \text{dom}(h_i)$$
is the *domain* of the optimization problem (1). $\mathcal{D}$ is obviously convex.
- A point $\mathbf{x} \in \mathcal{D}$ is said to be a *feasible* point for (1) if it satisfies $f_i(\mathbf{x}) \leq 0$, $i = 1, 2, \ldots, m$, $h_i(\mathbf{x}) = 0$, $i = 1, 2, \ldots, p$. The set of all feasible points $\mathcal{F}$ is called the feasible set or the constraint set.

The convex optimization problem II

- The set
$$\mathcal{D} = \bigcap \mathrm{dom}(f_i) \bigcap \mathrm{dom}(h_i)$$
is the *domain* of the optimization problem (1). $\mathcal{D}$ is obviously convex.
- A point $\mathbf{x} \in \mathcal{D}$ is said to be a *feasible* point for (1) if it satisfies $f_i(\mathbf{x}) \leq 0$, $i = 1, 2, \ldots, m$, $h_i(\mathbf{x}) = 0$, $i = 1, 2, \ldots, p$. The set of all feasible points $\mathcal{F}$ is called the feasible set or the constraint set.
- The *optimal value* $p^\star$ is defined as
$$p^\star = \inf \{f_0(\mathbf{x}), \mathbf{x} \in \mathcal{F}\},$$
where $p^\star$ is $-\infty$ if the problem is unbounded from below.
- A point $x^\star$ is said to be an optimal point if it is feasible and $f_0(\mathbf{x}^\star) = p^\star$.

The convex optimization problem III

- We say that (1) is solvable and the optimum is attained if $\mathbf{x}^\star$ exists; the problem is unsolvable if $\mathcal{F}$ is empty or if $p^\star = -\infty$.
- A point $\mathbf{x}$ is $\epsilon-$suboptimal if it is feasible and $f_0(\mathbf{x}) \leq p^\star + \epsilon$, where $\epsilon > 0$.

The convex optimization problem III

- We say that (1) is solvable and the optimum is attained if $\mathbf{x}^\star$ exists; the problem is unsolvable if $\mathcal{F}$ is empty or if $p^\star = -\infty$.
- A point $\mathbf{x}$ is $\epsilon-$suboptimal if it is feasible and $f_0(\mathbf{x}) \leq p^\star + \epsilon$, where $\epsilon > 0$.
- A feasible point $\mathbf{x}_l^\star$ is said to be *locally optimal* if there exists $r > 0$ such that

$$f_0(\mathbf{x}_l^\star) = \inf \left\{ f_0(\mathbf{x}), \mathbf{x} \in \mathcal{F}, \|\mathbf{x} - \mathbf{x}_l^\star\|_2 \leq r \right\},$$

  and is said to be *globally optimal* if it is optimal over all $\mathbf{x} \in \mathcal{F}$.

The convex optimization problem III

- We say that (1) is solvable and the optimum is attained if $\mathbf{x}^\star$ exists; the problem is unsolvable if $\mathcal{F}$ is empty or if $p^\star = -\infty$.
- A point $\mathbf{x}$ is $\epsilon-$suboptimal if it is feasible and $f_0(\mathbf{x}) \leq p^\star + \epsilon$, where $\epsilon > 0$.
- A feasible point $\mathbf{x}_l^\star$ is said to be *locally optimal* if there exists $r > 0$ such that

$$f_0(\mathbf{x}_l^\star) = \inf\left\{f_0(\mathbf{x}), \mathbf{x} \in \mathcal{F}, \|\mathbf{x} - \mathbf{x}_l^\star\|_2 \leq r\right\},$$

  and is said to be *globally optimal* if it is optimal over all $\mathbf{x} \in \mathcal{F}$.
- **For convex optimization problems, any local optimum is also a global optimum, and the set of points which achieves this optimum is convex.**
- This means: if we are *searching* for an optimum, we can stop once we find a local one. There is no better optimum *out there* in the domain.

A simple equivalent formulation

- Note that problem (1) is also equivalent to

$$
\begin{aligned}
&\text{minimize} \quad t, \\
&\text{subject to } f_0(\mathbf{x}) \leq t,\ f_i(\mathbf{x}) \leq 0,\ i = 1, 2, \ldots, m, \\
&\text{and } h_i(\mathbf{x}) = 0,\ i = 1, 2, \ldots, p,
\end{aligned} \tag{2}
$$

- No further constraint on new decision variable t means that we can simply set $t^\star = f_0(\mathbf{x}^\star)$. This is also called *epigraph* formulation.

A simple equivalent formulation

- Note that problem (1) is also equivalent to

$$\begin{aligned} &\text{minimize} \quad t, \\ &\text{subject to } f_0(\mathbf{x}) \le t,\ f_i(\mathbf{x}) \le 0,\ i = 1, 2, \ldots, m, \\ &\text{and } h_i(\mathbf{x}) = 0,\ i = 1, 2, \ldots, p, \end{aligned} \tag{2}$$

- No further constraint on new decision variable t means that we can simply set $t^\star = f_0(\mathbf{x}^\star)$. This is also called *epigraph* formulation.
- This added variable t comes handy in many cases when $f_0(\mathbf{x})$ itself is less convenient to deal with, as we shall see.

Applications of convex optimization

- Within OR, convex optimization problems occur in supply chain planning, capacity location, financial portfolio optimization, asset and liability management, $\cdots$
- Elsewhere, they also occur in data analysis (curve fitting), signal processing, control system design, structural optimization, antenna array design, $\cdots$

Applications of convex optimization

- Within OR, convex optimization problems occur in supply chain planning, capacity location, financial portfolio optimization, asset and liability management, $\cdots$
- Elsewhere, they also occur in data analysis (curve fitting), signal processing, control system design, structural optimization, antenna array design, $\cdots$
- Special types of (extremely useful) convex optimization problems: linear programming (LP), quadratic programming (QP) and semi-definite programming (SDP).
- Very significant body of theoretical research as well as software implementation exists for each of these.

The linear programming problem

- In LP, both the objective function and the constraint functions are linear:

$$\begin{aligned} &\text{minimize } \mathbf{c}^\top \mathbf{x} \\ &\text{subject to } A\mathbf{x} = \mathbf{b}, \\ &\mathbf{x} \geq 0, \end{aligned} \tag{3}$$

- The vectors $\mathbf{c}, \mathbf{b}$ and the matrix A are the problem parameters specifying the objective function and the constraint functions.

The linear programming problem

- In LP, both the objective function and the constraint functions are linear:

$$\begin{aligned} &\text{minimize } \mathbf{c}^\top \mathbf{x} \\ &\text{subject to } A\mathbf{x} = \mathbf{b}, \\ &\mathbf{x} \geq 0, \end{aligned} \tag{3}$$

- The vectors $\mathbf{c}, \mathbf{b}$ and the matrix A are the problem parameters specifying the objective function and the constraint functions.
- Applied convex programming starts with LP; simplex method of Dantzig $\sim$ 1947-48 made mathematical optimization tractable.
- Still a work-horse within financial optimization. You will learn about solving large scale LPs in this course.

Example of LP: the diet problem

- Suppose that there are m basic nutrients;
- A healthy diet needs b_j units of j^{th} nutrient per day.
- There are n different food items available, with one unit of item i containing a_{ji} units of nutrient j.
- Price of food item i is c_i per unit.

Example of LP: the diet problem

- Suppose that there are m basic nutrients;
- A healthy diet needs b_j units of j^{th} nutrient per day.
- There are n different food items available, with one unit of item i containing a_{ji} units of nutrient j.
- Price of food item i is c_i per unit.
- How do we minimize the cost of food per day, while keeping the diet healthy?

The diet problem (continued)

- This leads, precisely, to

$$\begin{aligned} &\text{minimize } \mathbf{c}^\top \mathbf{x} \\ &\text{subject to } A\mathbf{x} = \mathbf{b}, \\ &\mathbf{x} \geq 0, \end{aligned} \tag{4}$$

  where x_i is the number of units of food item i to be purchased.

- There might be other linear constraints on $\mathbf{x}$, *e.g.* on the number of units of any one food item purchased.

The diet problem (continued)

- This leads, precisely, to

$$
\begin{aligned}
&\text{minimize } \mathbf{c}^\top \mathbf{x}\\
&\text{subject to } A\mathbf{x} = \mathbf{b},\\
&\mathbf{x} \geq 0,
\end{aligned}
\tag{4}
$$

  where x_i is the number of units of food item i to be purchased.

- There might be other linear constraints on $\mathbf{x}$, *e.g.* on the number of units of any one food item purchased.
- Note: Increasing the number of food items from, say, 20 to 200 makes very little difference in computational complexity, but $\cdots$
- Saying 'use any 10 out of 20 food items' makes obtaining an exact solution 'far more difficult'/ practically impossible.

The Quadratic programming problem

- In QP, the objective function is convex quadratic and the constraint functions are linear, i.e. the problem is of the form

$$\begin{aligned} &\text{minimize } \frac{1}{2}\mathbf{x}^\top P\mathbf{x} + \mathbf{q}^\top \mathbf{x} + r \\ &\text{subject to } G\mathbf{x} \leq \mathbf{h}, \\ &\qquad A\mathbf{x} = \mathbf{b}. \end{aligned} \tag{5}$$

- The matrices P, G, A, vectors $\mathbf{q}$, $\mathbf{h}$ and the scalar r are the problem parameters.

The Quadratic programming problem

- In QP, the objective function is convex quadratic and the constraint functions are linear, i.e. the problem is of the form

$$\begin{aligned} &\text{minimize } \frac{1}{2}\mathbf{x}^\top P\mathbf{x} + \mathbf{q}^\top \mathbf{x} + r \\ &\text{subject to } G\mathbf{x} \preceq \mathbf{h}, \\ &\qquad A\mathbf{x} = \mathbf{b}. \end{aligned} \tag{5}$$

- The matrices P, G, A, vectors $\mathbf{q}$, $\mathbf{h}$ and the scalar r are the problem parameters.
- The vector inequality (5) indicates that $G\mathbf{x} - \mathbf{h}$ has all non-negative elements.

The Quadratic programming problem

- In QP, the objective function is convex quadratic and the constraint functions are linear, i.e. the problem is of the form

$$\begin{aligned} \text{minimize } & \frac{1}{2}\mathbf{x}^\top P\mathbf{x} + \mathbf{q}^\top \mathbf{x} + r \\ \text{subject to } & G\mathbf{x} \le \mathbf{h}, \\ & A\mathbf{x} = \mathbf{b}. \end{aligned} \tag{5}$$

- The matrices P, G, A, vectors $\mathbf{q}$, $\mathbf{h}$ and the scalar r are the problem parameters.
- The vector inequality (5) indicates that $G\mathbf{x} - \mathbf{h}$ has all non-negative elements.
- The matrix P is required to be positive semi-definite for this problem to be convex ($\mathbf{x}^\top P\mathbf{x} \ge 0 \; \forall \, \mathbf{x}$).

Examples of QP: least squares data-fitting

- In *data fitting* problems,

$$\mathbf{b} = A\mathbf{x} + \mathbf{v}$$

where b is a vector of measurements, the perturbation $\mathbf{v}$ is assumed to be small and we are trying to find a vector $\mathbf{x}$ which minimizes the Euclidian norm of this perturbation. This leads to QP

$$\text{minimize } \|A\mathbf{x} - \mathbf{b}\|_2^2 \tag{6}$$

Examples of QP: least squares data-fitting

- In *data fitting* problems,

$$\mathbf{b} = A\mathbf{x} + \mathbf{v}$$

where b is a vector of measurements, the perturbation $\mathbf{v}$ is assumed to be small and we are trying to find a vector $\mathbf{x}$ which minimizes the Euclidian norm of this perturbation. This leads to QP

$$\text{minimize } \|A\mathbf{x} - \mathbf{b}\|_2^2 \tag{6}$$

- This has a closed-form solution if there are no constraints on $\mathbf{x}$; needs to be solved numerically if there are constraints, *e.g.* $\mathbf{x} \geq 0$.
- In interpolation problems, the matrix A has entries of the form $(A)_{ij} = \theta_i^{j-1}$ for given $\theta_i, i = 1, 2, \dots m$ and the problem is to find the coefficient vector $\mathbf{x}$ of a polynomial $p(\theta)$ of a prescribed degree n, which best matches the set of points (θ_i, b_i).

Data fitting in l_1 norm as LP

- Recall least squares data fitting; in general, minimizing any vector norm of $A\mathbf{x} - \mathbf{b}$ is a convex problem.

Data fitting in l_1 norm as LP

- Recall least squares data fitting; in general, minimizing any vector norm of $A\mathbf{x} - \mathbf{b}$ is a convex problem.
- In particular, since $\|\mathbf{z}\|_1 = \sum_i |z_i|$, we can re-formulate minimizing $\|A\mathbf{x} - \mathbf{b}\|_1$ over $\mathbf{x}$ as a linear program:

Data fitting in l_1 norm as LP

- Recall least squares data fitting; in general, minimizing any vector norm of $A\mathbf{x} - \mathbf{b}$ is a convex problem.
- In particular, since $\|\mathbf{z}\|_1 = \sum_i |z_i|$, we can re-formulate minimizing $\|A\mathbf{x} - \mathbf{b}\|_1$ over $\mathbf{x}$ as a linear program:
-

$$\begin{aligned} \text{minimize} \quad & \sum_i t_i \quad \text{subject to} \\ (A\mathbf{x} - \mathbf{b})_i \; & \leq \; t_i \\ (A\mathbf{x} - \mathbf{b})_i \; & \geq \; -t_i, \end{aligned}$$

with $t_1, \cdots t_n$ as auxiliary decision variables.

Data fitting in l_∞ norm as LP

- Infinity norm for a vector is defined by $\|\mathbf{z}\|_\infty = \max_i |z_i|$.
- We can re-formulate minimizing $\|A\mathbf{x} - \mathbf{b}\|_\infty$ as a linear program:

$$\begin{aligned} &\text{minimize} \quad t \quad \text{subject to} \\ &(A\mathbf{x} - \mathbf{b})_i \leq t \\ &(A\mathbf{x} - \mathbf{b})_i \geq -t, \end{aligned}$$

with t as a single auxiliary decision variable.

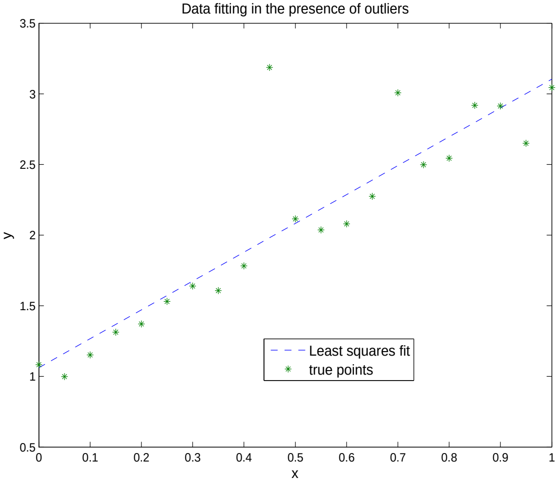
Data fitting: what should you use?

- Least squares is usually the quickest.
- If you want a solution robust to outliers: use l_1−norm.
- If you want to get the 'best worst case' fit: use l_∞−norm.

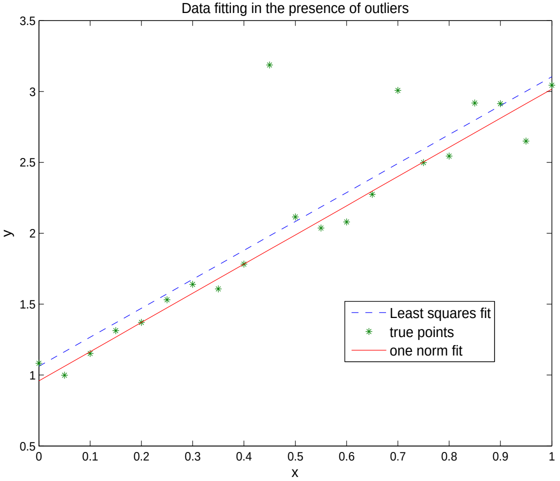
Data fitting: what should you use?

- Least squares is usually the quickest.
- If you want a solution robust to outliers: use l_1-norm.
- If you want to get the 'best worst case' fit: use $l_\infty-$norm.
- For the same set of points ($y = 2x + 1+$ random noise) with two outliers, we can compare the fits obtained by minimising different norms.

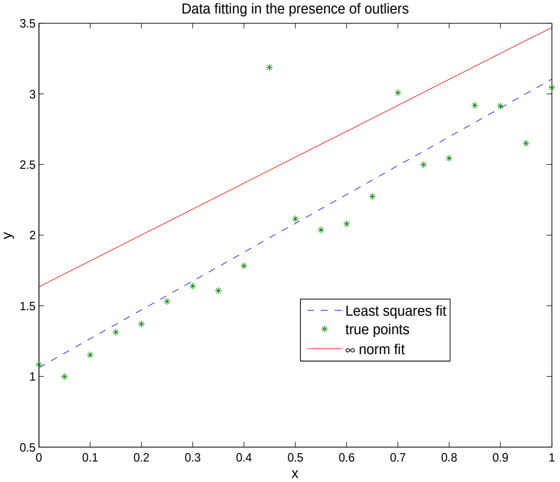
Data fitting with outliers: least squares fit



Least squares vs 1−norm fit



Least squares vs ∞-norm fit



Recognizing convex problems

- See if you can re-formulate the problem as LP/QP or SDP (next lecture);
- See if you can re-formulate it as a quasiconvex problem (next lecture);
- Can you arrive at your objective function and constraints via composition of simpler convex functions?
- Check convexity of functions via gradient/Hessian/ testing it on a line.

Recognizing convex problems - example

- Given a decision vector $\mathbf{x}$ specifying variables such as retail price and advertising spend, let the probability of consumer buying your product be defined by

$$f(x) = \frac{\exp(\mathbf{a}^\top \mathbf{x} + \mathbf{b})}{1 + \exp(\mathbf{a}^\top \mathbf{x} + \mathbf{b})}.$$

  How would you maximize $f(\mathbf{x})$ over $\mathbf{x}$? Assume that there are suitable constraints over $\mathbf{x}$, and $a^\top \mathbf{x} + \mathbf{b} \geq 0$.

Recognizing convex problems - example

- Given a decision vector $\mathbf{x}$ specifying variables such as retail price and advertising spend, let the probability of consumer buying your product be defined by

$$f(x) = \frac{\exp(\mathbf{a}^\top \mathbf{x} + \mathbf{b})}{1 + \exp(\mathbf{a}^\top \mathbf{x} + \mathbf{b})}.$$

  How would you maximize $f(\mathbf{x})$ over $\mathbf{x}$? Assume that there are suitable constraints over $\mathbf{x}$, and $a^\top \mathbf{x} + \mathbf{b} \geq 0$.

- $h(x) = {e^x}/{1+e^x}$ is concave and non-decreasing and $g(x) = a^\top \mathbf{x} + \mathbf{b}$. Hence $f(x) = h(g(x))$ is concave. Further, $\nabla(f) = 0 \Leftrightarrow \nabla(g) = 0$.

Recognizing convex problems - example

- Given a decision vector $\mathbf{x}$ specifying variables such as retail price and advertising spend, let the probability of consumer buying your product be defined by

$$f(x) = \frac{\exp(\mathbf{a}^\top \mathbf{x} + \mathbf{b})}{1 + \exp(\mathbf{a}^\top \mathbf{x} + \mathbf{b})}.$$

  How would you maximize $f(\mathbf{x})$ over $\mathbf{x}$? Assume that there are suitable constraints over $\mathbf{x}$, and $a^\top \mathbf{x} + \mathbf{b} \geq 0$.

- $h(x) = {}^{e^x}/_{1+e^x}$ is concave and non-decreasing and $g(x) = a^\top \mathbf{x} + \mathbf{b}$. Hence $f(x) = h(g(x))$ is concave. Further, $\nabla(f) = 0 \Leftrightarrow \nabla(g) = 0$.
- This is a simple linear programming problem if the constraints on $\mathbf{x}$ are affine.

Next steps

- Having looked at a few different types of convex optimization problems,

Next steps

- Having looked at a few different types of convex optimization problems,
- we will next look at one more special- and important- class of problems (semidefinite programs).
- Then we will look at some theoretical analysis of optimization and (finally!) how to actually solve these problems.
- This will also include a de-tour on modelling and solving quasiconvex optimization problems.