Optimization Models

EECS 127 / EECS 227AT

Laurent El Ghaoui

EECS department
UC Berkeley

Fall 2018

LECTURE 17

Convex Optimization Problems

The great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity."

R. Tyrrell Rockafellar

Outline

1. Overview
2. Nomenclature
   - Standard forms
   - Feasibility and boundedness
   - Optimal set
   - Local and global optima
   - 2D Examples
3. Examples
4. Problem transformations
   - Monotone objective transformation
   - Adding slack variables
   - Replacing nonlinear equalities with inequalities
   - Hidden convexity

Context

- So far we've seen linear algebra and conic models.
- These are computationally efficient, reliable.
- They cover quite a lot of practical problems — however they are very specific ...

Are there more models that offer a similar good trade-off between computational efficiency and modelling power?

- Conic models are part of a broader class called "convex programming" (CP) models.
- Convex optimization offers (roughly) the same degree of reliability as linear algebra, but covers much more.
- The formalism allows for deep theoretical and practical insight, which is not true for general nonlinear programming models.

Trouble with nonlinear models

General optimization model:

$$\begin{aligned} p^* = \min_x \quad & f_0(x) \\ \text{subject to:} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m, \end{aligned}$$

with f_i's arbitrary nonlinear functions.

- Algorithms may deliver very suboptimal solutions for unconstrained problems.
- Can fail for constrained problems—find no feasible point even though one exists.
- Still, a useful paradigm, but beyond the scope of the class.

- Here we focus on models that are reliable, fool-proof;
- such models can help solve more complicated ones (*e.g.*, via "successive convex approximation" methods)

Convex problem

Standard form

$$p^* = \min_{x \in \mathbb{R}^n} f_0(x) \text{ subject to: } \quad \begin{aligned} f_i(x) &\leq 0, \quad i = 1, \ldots, m, \\ Ax &= b, \end{aligned}$$

where

- $f_0, \ldots, f_m$ are convex functions;
- The equality constraints are affine, and represented via the matrix $A \in \mathbb{R}^{q \times n}$ and vector $b \in \mathbb{R}^q$.

There is an implicit constraint on x, that it belongs to domain of the problem, which is the set

$$\mathcal{D} \doteq \bigcap_{i=0}^{m} \operatorname{dom} f_i.$$

Since f_i's are convex, the domain is convex.

Implicit constraints

Example

$$\min_x \; f_0(x) \; : \; x^\top C x \leq 1,$$

with $C = C^\top \succeq 0$, and

$$f_0(x) \doteq \begin{cases} -\sum_{i=1}^n \log x_i & \text{if } x > 0 \\ +\infty & \text{otherwise.} \end{cases}$$

- the problem is convex, since f_0 is, and C is PSD;
- the domain of problem is $\mathcal{D} = \mathbb{R}^n_{++}$.
- the constraint $x > 0$ is implicit.

Conic optimization problems are convex

The class of convex problems includes the conic optimization problems we have seen so far: LP, QP, QCQP, and SOCP.

Since the SOCP class includes all the others, it suffices to show that SOCPs are convex problems. Recall the SOCP model:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & c^\top x \\ \text{s.t.:} \quad & \|A_i x + b_i\|_2 \leq c_i^\top x + d_i, \quad i = 1, \ldots, m, \end{aligned}$$

with A_i, b_i, c_i, d_i matrices of appropriate size. Convexity of the problem stems from the fact that for every i,

$$x \to \|A_i x + b_i\|_2 - (c_i^\top x + d_i)$$

is a convex function, as seen in lecture 16.

Other standard forms of convex problems

- A convex optimization problem can equivalently be defined as a problem where one minimizes a convex objective function, subject to the restriction $x \in \mathcal{X}$, i.e. that the decision variable must belong to a convex set $\mathcal{X}$:

$$p^* = \min_{x \in \mathcal{X}} f_0(x).$$

  Of course, there is still the implicit constraint that x belongs to the domain of f_0.

- The problem is said to be *unconstrained*, when $\mathcal{X} = \mathbb{R}^n$.
- Solving the optimization problem means finding the optimal minimal value p^* of the objective, and possibly also a *minimizer*, or optimal solution, that is a vector $x^* \in \mathcal{X}$ such that $f_0(x^*) = p^*$.

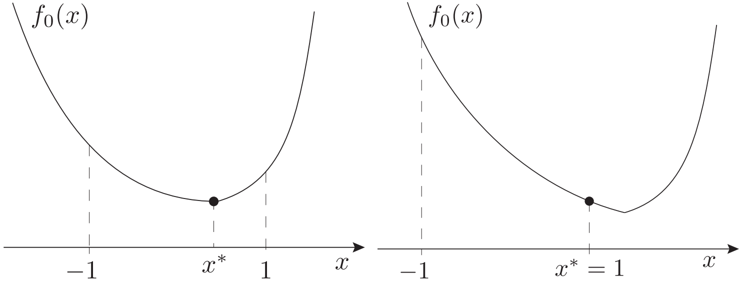
Feasibility problems

- If $\mathcal{X}$ is the empty set, we say that the problem is *infeasible*: no solution that satisfies the constraints exists. In such a case it is customary to set $p^* = +\infty$. When $\mathcal{X}$ is nonempty, we say that the problem is *feasible*.
- Often, we do not know in advance if the feasible set $\mathcal{X}$ is empty or not; the task of determining if this is the case or not is referred to as a *feasibility* problem.

If the problem is feasible and $p^* = -\infty$ we say that the problem is *unbounded below*. Notice that it can also happen that the problem is feasible but still no optimal solution exists, in which case we say that the optimal value p^* is not *attained* at any finite point.

Active vs inactive constraints

- If $x^* \in \mathcal{X}_{\rm opt}$ is such that $f_i(x^*) < 0$, we say that the i-th inequality constraint is *inactive* (or *slack*) at the optimal solution x^*.
- Conversely, if $f_i(x^*) = 0$, we say that the i-th inequality constraint is *active* at x^* .



Optimal set

- The *optimal set* (or, set of solutions) is defined as the set of feasible points for which the objective function attains the optimal value:

$$\mathcal{X}_{\mathrm{opt}} = \{x \in \mathcal{X} : \, f_0(x) = p^*\} \, .$$

  We shall also write, using the "argmin" notation,

$$\mathcal{X}_{\mathrm{opt}} = \arg\min_{x \in \mathcal{X}} \, f_0(x).$$

- For convex problems, the optimal set is a convex set (see proof below).

Local and Global Optima

Theorem 1

Consider the optimization problem: $\min_{x \in \mathcal{X}} f_0(x)$*. If* f_0 *is a convex function and* $\mathcal{X}$ *is a convex set, then any locally optimal solution is also globally optimal. Moreover, the set* $\mathcal{X}_{\rm opt}$ *of optimal points is convex.*

Proof. Let $x^* \in \mathcal{X}$ be a local minimizer of f_0, let $p^* = f_0(x^*)$, and consider any point $y \in \mathcal{X}$. We need to prove that $f_0(y) \geq f_0(x^*) = p^*$. By convexity of f_0 and $\mathcal{X}$ we have that, for $\theta \in [0, 1]$, $x_\theta = \theta y + (1 - \theta)x^* \in \mathcal{X}$, and

$$f_0(x_\theta) \leq \theta f_0(y) + (1 - \theta) f_0(x^*).$$

Subtracting $f_0(x^*)$ from both sides of this equation, we obtain

$$f_0(x_\theta) - f_0(x^*) \leq \theta(f_0(y) - f_0(x^*)).$$

Since x^* is a local minimizer, the left-hand side in this inequality is nonnegative for all small enough values of $\theta > 0$. We thus conclude that the right hand side is also nonnegative, i.e., $f_0(y) \geq f_0(x^*)$, as claimed. Also, the optimal set is convex, since it can be expressed as the p^*-sublevel set of a convex function:

$$\mathcal{X}_{\rm opt} = \{x \in \mathcal{X} : f_0(x) \leq p^*\}.$$

2D Examples

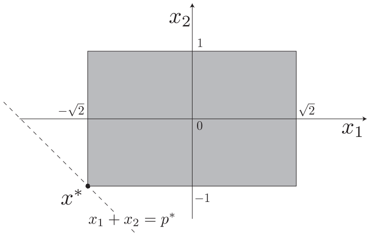
Consider the problem

$$\begin{aligned} p^* = \min_{x \in \mathbb{R}^2} \quad & x_1 + x_2 \\ \text{subject to:} \quad & x_1^2 \leq 2 \\ & x_2^2 \leq 1. \end{aligned}$$

The feasible set $\mathcal{X}$ for this problem is nonempty and it is given by the rectangle $[-\sqrt{2}, \sqrt{2}] \times [-1, 1]$. The optimal objective value is $p^* = -\sqrt{2} - 1$, which is attained at the unique optimal point

$$x^* = [-\sqrt{2} \;\; -1]^\top,$$

.



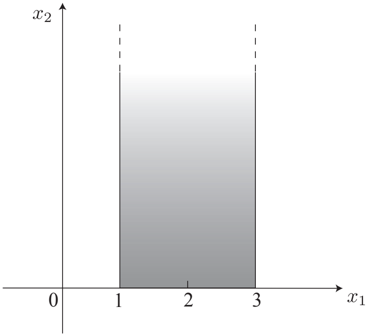
2D Examples

Consider the problem

$$\begin{aligned} p^* = \min_{x \in \mathbb{R}^2} \quad & x_2 \\ \text{subject to:} \quad & (x_1 - 2)^2 \leq 1 \\ & x_2 \geq 0. \end{aligned}$$

The feasible set $\mathcal{X}$ for this problem is nonempty. The optimal objective value is $p^* = 0$, which is attained at infinitely many optimal points: the optimal set is

$$\mathcal{X}_{\text{opt}} = \{(x_1, 0) \ : \ x_1 \in [1, 3]\}.$$



2D Examples

- The problem

$$\begin{aligned} p^* = \min_{x\in\mathbb{R}^2} \quad & \mathrm{e}^{x_1} \\ \text{subject to:} \quad & x_2 \geq (x_1-1)^2+1 \\ & x_2 - x_1 + \tfrac{1}{2} \leq 0 \end{aligned}$$

  is unfeasible, thus, by convention, $p^* = +\infty$.

- The problem

$$\begin{aligned} p^* = \min_{x\in\mathbb{R}} \quad & \mathrm{e}^{-x} \\ \text{subject to:} \quad & x \geq 0 \end{aligned}$$

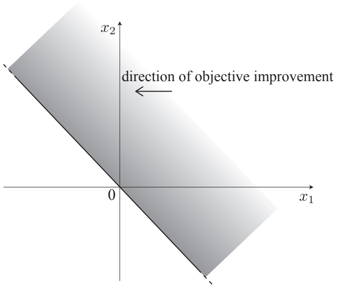
  is feasible, and the optimal objective value is $p^* = 0$. However, the optimal set is empty, since p^* is not attained at any finite point (it is only attained in the limit, as $x \to \infty$).

2D Examples

Consider the problem

$$\begin{aligned} p^* = \min_{x \in \mathbb{R}^2} \quad & x_1 \\ \text{subject to:} \quad & x_1 + x_2 \geq 0. \end{aligned}$$

The feasible set is a halfspace, and the problem is unbounded below ($p^* = -\infty$). No optimal solution exists, since the optimal value is attained asymptotically for $x_1 \to -\infty$, $x_1 \geq -x_2$.



2D Examples

- Consider the problem

$$p^* = \min_{x \in \mathbb{R}} \quad (x+1)^2.$$

This is an unconstrained problem, for which $p^* = 0$ is attained at the (unique) optimal point $x^* = -1$.

- Consider next a constrained version of the problem, where

$$\begin{aligned} p^* = \min_{x \in \mathbb{R}} & \quad (x+1)^2 \\ \text{subject to:} & \quad x > 0. \end{aligned}$$

This problem is feasible, and has optimal value $p^* = 1$. However, this optimal value is not attained by any feasible point: it is attained in the limit by a point x that tends to 0, but 0 does not belong to the feasible set $(0, +\infty)$.

Problem transformations

- An optimization problem can be transformed, or reformulated, into an *equivalent* one by means of several useful "tricks," such as:
  - monotone transformation of the objective (e.g., scaling, logarithm, squaring) and constraint functions;
  - change of variables;
  - addition of slack variables;
  - epigraphic reformulation;
  - replacement of equality constraints with inequality ones;
  - elimination of inactive constraints;
  - discovering hidden convexity;
  - etc.
- By the term "equivalent" referred to two optimization problems, we here mean informally that the optimal objective value and optimal solutions (if they exist) of one problem can be easily obtained from the optimal objective value and optimal solutions of the other problem, and vice-versa.

Monotone objective transformation

- Consider an optimization problem in standard form (1). Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a continuous and strictly increasing function over $\mathcal{X}$, and consider the transformed problem

$$\begin{aligned} g^* = \min_{x \in \mathbb{R}^n} \quad & \varphi(f_0(x)) \\ \text{subject to:} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m, \\ & Ax = b. \end{aligned} \tag{1}$$

- The original and transformed problems have the same set of optimal solutions.

Example: the least-squares problem, with objective $f_0(x) = \|Ax - y\|_2$, where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ are given. We apply the result with the function $z \geq 0 \to \phi(z) = z^2$, which is increasing on $\mathbb{R}_+$.

Monotone objective transformation

Proof. Suppose x^* is optimal for problem (1), i.e., $f_0(x^*) = p^*$. Then, x^* is feasible for problem (1), thus it holds that $\varphi(f_0(x^*)) = \varphi(p^*) \geq g^*$.
Assume next that $\tilde{x}^*$ is optimal for problem (1), i.e., $\varphi(f_0(\tilde{x}^*)) = g^*$. Then, $\tilde{x}^*$ is feasible for problem (1), thus it holds that $f_0(\tilde{x}^*) \geq p^*$.
Now, since φ is continuous and strictly increasing over $\mathcal{X}$, it has a well-defined inverse φ^{-1}, thus we may write $\varphi(f_0(\tilde{x}^*)) = g^* \quad \Leftrightarrow \quad \varphi^{-1}(g^*) = f_0(\tilde{x}^*)$, which yields

$$\varphi^{-1}(g^*) \geq p^*.$$

Since φ is strictly increasing and $\varphi(\varphi^{-1}(g^*)) = g^*$, the latter relation also implies that $g^* \geq \varphi(p^*)$, which implies that it must be $\varphi(p^*) = g^*$. This means that for any optimal solution x^* of problem (1) it holds that

$$\varphi(f_0(x^*)) = g^*,$$

which implies that x^* is also optimal for problem (1). Vice-versa, for any optimal solution $\tilde{x}^*$ of problem (1) it holds that

$$f_0(\tilde{x}^*) = \varphi^{-1}(g^*) = p^*,$$

which implies that $\tilde{x}^*$ is also optimal for problem (1).

Monotone objective transformation

Example: logistic regression

A random variable $y \in \{-1, 1\}$ has a distribution modelled as

$$p = \mathbf{Prob}(y = 1) = \frac{\exp(w^\top x + b)}{\exp(w^\top x + b) + 1} = 1 - \mathbf{Prob}(y = -1).$$

where $w \in \mathbb{R}^n$, $b \in \mathbb{R}$ are parameters, and $x \in \mathbb{R}^n$ contains explanatory variables (features). The estimation problem is to estimate w, b from given observations (xi, yi), $i = 1, \ldots, m$.

In the "maximum likelihood" approach, we maximize the likelihood function

$$L(w, b) \doteq \prod_{i=1}^{m} \left(\frac{\exp y_i(w^\top x_i + b)}{\exp y_i(w^\top x_i + b) + 1} \right)$$

Function L is not concave in (w, b), but $\log L$ is, since the log-sum-exp function is convex:

$$\log L(w, b) = -\sum_{i=1}^{m} \log\left(1 + \exp(-y_i(w^\top x_i + b))\right).$$

Addition of slack variables

- Equivalent problem formulations are also obtained by introducing new "slack" variables into the problem. We here describe a typical case that arises when a constraint or the objective involves the sum of terms, as in the following problem

$$\begin{array}{rl} p^* = \min\limits_{x} & f_0(x) + \sum\limits_{i=1}^{r} \varphi_i(x) \\ \text{s.t.} & x \in \mathcal{X}. \end{array} \tag{2}$$

- Introducing slack variables t_i, $i = 1, \ldots, p$, we reformulate this problem as

$$\begin{array}{rl} g^* = \min\limits_{x,t} & \sum\limits_{i=1}^{r} t_i \\ \text{s.t.} & x \in \mathcal{X} \\ & \varphi_i(x) \le t_i \quad i = 1, \ldots, r, \end{array} \tag{3}$$

where this new problem has the original variable x, plus the vector of slack variables $t = (t_1, \ldots, t_r)$.

Addition of slack variables

What is problem equivalence?

Problem (2) and (3) are equivalent in the following sense:

1. If x is feasible for (2), then x, $t_i = \varphi_i(x)$, $i = 1, \ldots, r$, is feasible for (3);
2. If x, t is feasible for (3), then x is feasible for (2);
3. If x^* is optimal for (2), then x^*, $t_i^* = \varphi_i(x^*)$, $i = 1, \ldots, r$, is optimal for (3);
4. If x^*, t^* is optimal for (3), then x^* is optimal for (2);
5. $g^* = p^*$.

Example: the LASSO problem

$$p^* = \min_x \|Ax - y\|_2^2 + \|x\|_1$$

is equivalent to the QP

$$p^* = \min_{x,t} \|Ax - y\|_2^2 + \sum_{i=1}^n t_i \ : \ t_i \geq x_i \geq -t_i, \quad i = 1, \ldots, n.$$

At optimum, we have $t_i^* = |x_i^*|$, $i = 1, \ldots, n$.

Generality of the linear objective

- A common use of the slack variable "trick" described above consists in transforming a convex optimization problem of the form (1), with generic convex objective f_0, into an *equivalent* convex problem having *linear* objective.
- Introducing a new slack variable $t \in \mathbb{R}$, problem (1) is reformulated as

$$\begin{aligned} t^* = \min_{x \in \mathbb{R}^n, t \in \mathbb{R}} \quad & t \\ \text{subject to:} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, q \\ & f_0(x) \leq t. \end{aligned} \tag{4}$$

- Problem (4) has a linear objective in the augmented variables (x, t), and it is usually referred to as the *epigraphic* reformulation of the original problem (1).
- Any convex optimization problem can thus be reformulated in the form of an equivalent convex problem with linear objective.

Substituting equality constraints with inequality constraints

- In certain cases, we can substitute an equality constraint of the form $b(x) = u$ with an inequality constraint $b(x) \leq u$.
- This can be useful, in some cases, for gaining convexity. Indeed, if $b(x)$ is a convex function, then the set described by the *equality* constraint $\{x : b(x) = u\}$ is non-convex in general (unless b is affine); contrary, the set described by the *inequality* constraint $\{x : b(x) \leq u\}$ is the sublevel set of a convex function, hence it is convex.
- We give a sufficient condition under which such a substitution can be safely performed.

Substituting equality constraints with inequality constraints

- Consider a (non necessarily convex) problem of the form

$$p^* = \min_{x \in \mathcal{X}} \quad f_0(x) \qquad (5)$$
$$\text{s.t.} \quad b(x) = u,$$

where u is a given scalar, together with the related problem in which the equality constraint is substituted by an inequality one:

$$g^* = \min_{x \in \mathcal{X}} \quad f_0(x) \qquad (6)$$
$$\text{s.t.} \quad b(x) \leq u.$$

- Clearly, since the feasible set of the first problem is included in the feasible set of the second problem, it always holds that $g^* \leq p^*$.
- It actually holds that $g^* = p^*$, under the following conditions:
  - f_0 is nonincreasing over $\mathcal{X}$ (i.e., $f_0(x) \leq f_0(y) \Leftrightarrow x \geq y$ elementwise)
  - b is nondecreasing over $\mathcal{X}$, and
  - the optimal value p^* is attained at some optimal point x^*, and the optimal value g^* is attained at some optimal point $\tilde{x}^*$.

Substituting equality constraints with inequality constraints

Analogously, if the problem is in maximization form

$$\max_{x \in \mathcal{X}} \; f_0(x) \quad \text{s.t.} \quad b(x) = u,$$

then a sufficient condition for replacing the equality constraint with an inequality one is that both f_0 and b are nondecreasing over $\mathcal{X}$.

Remark 1

Observe that while we have $p^* = g^*$ (under the stated hypotheses) and every optimal solution of (5) is also optimal for (6), the converse is not necessarily true, that is problem (6) may have optimal solutions which are not feasible for the original problem (5). However, this converse implication holds if the objective is strictly monotone.

Example: Budget constraint in portfolio optimization

$$
\begin{aligned}
\max_x \quad & \hat{r}^\top x \\
\text{s.t.} \quad & x^\top \Sigma x \leq \sigma^2 \\
& \mathbf{1}^\top x + \phi(x) = 1,
\end{aligned}
$$

- $\phi(x)$: a function measuring the cost of transaction, which is a nondecreasing function of x.
- The last constraint expresses the fact that, assuming a unit initial capital in cash, the sum of the invested amounts, $\mathbf{1}^\top x$, must be equal to the initial capital, minus the expense for transaction costs.
- If $\hat{r} > 0$, the objective function in the above problem is increasing in x, and the left-hand side of the equality constraint is nondecreasing in x.
- We can write the problem equivalently by substituting the equality constraint with the inequality one $\mathbf{1}^\top x + \phi(x) \leq 1$.
- If ϕ is a convex function, the modified problem is convex, whereas the original formulation is not convex in general (unless ϕ is affine).

Hidden convexity

- Sometimes a problem as given is not convex, but we can transform it into an equivalent problem that is.
- An approach is to "relax" the problem into a convex one, and then prove that the relaxation is "exact".

None of these approaches is full-proof and a guaranteed path to finding a convex expression of a given problem.

Hidden convexity

An example: assignment problems

Revisit the swimming team assignment problem of lecture 10. Variable is $X \in \{0,1\}^{4\times 5}$; assignment problem is:

$$\min_{X \in \{0,1\}^{4\times 5}} \operatorname{trace} M^\top X \;:\; \begin{array}{l} X\mathbf{1} \le \mathbf{1} \text{ (at most one swimmer per stroke)} \\ X^\top \mathbf{1} = \mathbf{1} \text{ (one stroke per swimmer)} \end{array}$$

The above problem is not convex (since the feasible set is not). However we can relax it into an LP by changing the constraints $X \in \{0,1\}^{4\times 5}$ into $X \in [0,1]^{4\times 5}$.

It turns out that the relaxation is exact, in the sense that the optimal value of the non-convex and convex problems coincide.