Introduction to optimization

Geir Dahl
CMA, Dept. of Mathematics and Dept. of Informatics
University of Oslo

The plan

- 1. The basic concepts
- 2. Some useful tools
- 3. LP (linear programming = linear optimization)

Literature:

- Vanderbei: *Linear programming*, 2001 (2008).
- Bertsekas: *Nonlinear programming*, 1995.
- Boyd and Vandenberghe: *Convex optimization*, 2004.

What we do in optimization!

- we study and solve optimization problems!

The typical problem:

- given a function $f : \mathbb{R}^n \to \mathbb{R}$
- and a subset $S \subseteq \mathbb{R}^n$
- find a point (vector) $x^* \in S$ which minimizes (or maximizes) f over this set S.
- S is often the solution set of a system of linear, or nonlinear, equations and inequalities: this complicates things!

The work:

- find such x^*
- construct a suitable algorithm
- analyze algorithm
- analyze problem: prove theorems on properties

Areas – depending on properties of f and S:

- linear optimization (LP=linear programming)
- nonlinear optimization
- discrete optimization (combinatorial opt.)
- stochastic optimization
- optimal control
- multicriteria optimization

Optimization: branch of applied mathematics, so

theory – algorithms – applications

The basic concepts

- feasible point: a point $x \in S$, and S is called the feasible set
- global minimum (point): a point $x^* \in S$, satisfying

$$f(x^*) = \min\{f(x) : x \in S\}$$

- local minimum (point): a point $x^* \in S$, satisfying

$$f(x^*) = \min\{f(x) : x \in N \cap S\}$$

  for some (suitable small) neighborhood N of x^*.
- local/global maximum (point): similar.
- f: objective function, cost function

Optimal: minimum or maximum

Some useful tools

Tool 1: Existence:

- a minimum (or maximum) may not exist.
- how can we prove the existence?

Theorem

(*Extreme value theorem*) A continuous function on a compact (closed and bounded) subset of $\mathbb{R}^n$ attains its (global) maximum and minimum.

- very important result, but it does not tell us how to find an optimal solution.

Tool 2: local approximation – optimality criteria

- First order Taylor approximation:

$$f(x+h) = f(x) + \nabla f(x)^T h + \|h\| O(h)$$

where $O(h) \to 0$ as $h \to 0$.

- Second order Taylor approximation:

$$f(x+h) = f(x) + \nabla f(x)^T h + (1/2) h^T H_f(x) h + \|h\|^2 O(h)$$

where $O(h) \to 0$ as $h \to 0$.

Linear optimization (LP)

- linear optimization is to maximize (or minimize) a linear function in several variables subject to constraints that are linear equations and linear inequalities.
- many applications

Example: production planning

$$
\begin{array}{lrrl}
\text{maximize} & 3x_1 + & 5x_2 & \\
\text{subject to} & & & \\
& x_1 & & \le 4 \\
& & 2x_2 & \le 12 \\
& 3x_1 + & 2x_2 & \le 18 \\
& \multicolumn{3}{c}{x_1 \ge 0,\ x_2 \ge 0.}
\end{array}
$$

Application: linear approximation

Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. Recall: ℓ_1-norm; $\|y\|_1 = \sum_{i=1}^n |y_i|$.
The linear approximation problem

$$\min\{\|Ax - b\|_1 : x \in \mathbb{R}^n\}$$

may be solved as the following LP problem

$$\begin{array}{lll} \text{min} & \sum_{i=1}^m z_i & \\ \text{subject to} & & \\ & a_i^T x - b_i \leq z_i & (i \leq m) \\ & -(a_i^T x - b_i) \leq z_i & (i \leq m) \end{array}$$

LP problems in matrix form:

$$\begin{array}{lrcl} \max & c^T x & & \\ \text{subject to} & & & \\ & Ax & \leq & b \\ & x & \geq & O \end{array}$$

The inequality $Ax \leq b$ is a vector inequality and means that $\leq$ holds componentwise (for every component).

Analysis/algorithm: based on linear algebra.

LP is closely tied to theory/methods for solving systems of linear inequalities. Such systems have the form

$$Ax \leq b.$$

The simplex algorithm

- the simplex method is a general method for solving LP problems.
- developed by George B. Dantzig around 1947 in connection with the investigation of transportation problems for the U.S. Air Force.
- discussions on duality with John von Neumann
- the work was published in 1951.

Example

$$
\begin{array}{llllllll}
\max & 5x_1 & + & 4x_2 & + & 3x_3 & & \\
\text{subject to} & & & & & & & \\
& 2x_1 & + & 3x_2 & + & x_3 & \leq 5 \\
& 4x_1 & + & x_2 & + & 2x_3 & \leq 11 \\
& 3x_1 & + & 4x_2 & + & 2x_3 & \leq 8 \\
\end{array}
$$

$$x_1, x_2, x_3 \geq 0.$$

First, we convert to equations by introducing slack variables for every $\leq$-inequality, so e.g. the first ineq. is replaced by

$$w_1 = 5 - 2x_1 - 3x_2 - x_3\,, \quad w_1 \geq 0.$$

Problem rewritten as a "dictionary":

$$\begin{array}{lrcrcrcrcr}
\max & \eta & = & & & 5x_1 & + & 4x_2 & + & 3x_3 \\
\text{subj. to} & & & & & & & & & \\
 & w_1 & = 5 & - & & 2x_1 & - & 3x_2 & - & x_3 \\
 & w_2 & = 11 & - & & 4x_1 & - & x_2 & - & 2x_3 \\
 & w_3 & = 8 & - & & 3x_1 & - & 4x_2 & - & 2x_3 \\
 & & & & & \multicolumn{5}{l}{x_1, x_2, x_3, w_1, w_2, w_3 \geq 0.}
\end{array}$$

- left-hand side: dependent variables = basic variables.
- right-hand side: independent variables = nonbasic variables.

Initial solution: Let $x_1 = x_2 = x_3 = 0$, so $w_1 = 5, w_2 = 11, w_3 = 8$.

We always let the nonbasic variables be equal to zero. The basic variables are then *uniquely* determined. ("Basis property" in matrix version).

Not optimal! For instance, we can increase x_1 while keeping $x_2 = x_3 = 0$. Then

- η (the value of the objective function) will increase
- new values for the basic variables, determined by x_1
- the more we increase x_1, the more η increases!
- but, careful! The w_j's approach 0!

Maximum increase of x_1: avoid the basic variables to become negative. From $w_1 = 5 - 2x_1$, $w_2 = 11 - 4x_1$ and $w_3 = 8 - 3x_1$ we get $x_1 \leq 5/2$, $x_1 \leq 11/4$, $x_1 \leq 8/3$ so we can increase x_1 to the smallest value, namely 5/2.

This gives the new solution $x_1 = 5/2$, $x_2 = x_3 = 0$ and therefore $w_1 = 0, w_2 = 1, w_3 = 1/2$. And now $\eta = 25/2$. Thus: an improved solution!!

How to proceed? The dictonary is well suited for testing optimality, so we must transform to a new dictionary.

- We want x_1 and w_1 to "switch sides". So: x_1 should go *into* the basis, while w_1 goes *out* of the basis.This can be done by using the w_1-equation in order to eliminate x_1 from all other equations.
- Equivalent: we may use elementary row operations on the system in order to eliminate x_1: (i) solve for x_1: $x_1 = 5/2 - (1/2)w_1 - (3/2)x_2 - (1/2)x_3$, and (ii) add a suitable multiple of this equation to the other equations so that x_1 disappears and is replaced by twerms with w_1.

Remember: elementary row operations do not change the solution set of the linear system of equations.

Result:

$$
\begin{array}{lllllllll}
\eta & = 12.5 & - & 2.5w_1 & - & 3.5x_2 & + & 0.5x_3 \\
\hline
x_1 & = 2.5 & - & 0.5w_1 & - & 1.5x_2 & - & 0.5x_3 \\
w_2 & = 1 & + & 2w_1 & + & 5x_2 & & \\
w_3 & = 0.5 & + & 1.5w_1 & + & 0.5x_2 & - & 0.5x_3
\end{array}
$$

We have performed a pivot: the use of elementary row operations (or elimination) to switch two variables (one into and one out of the basis).

Repeat the process: not optimal solution as we can increase η by increasing x_3 from zero! May increase to $x_3 = 1$ and then $w_3 = 0$ (while the other basic variables are nonnegative). So, pivot: x_3 goes into the basis, and w_3 leaves the basis.

This gives the new dictionary:

$$\begin{array}{rcrcrcrcr}
\eta & = 13 & - & w_1 & - & 3x_2 & - & w_3 \\
\hline
x_1 & = 2 & - & 2w_1 & - & 2x_2 & + & w_3 \\
w_2 & = 1 & + & 2w_1 & + & 5x_2 & & \\
x_3 & = 1 & + & 3w_1 & + & x_2 & - & 2w_3
\end{array}$$

Here we see that all coefficients of the nonbasic variables are nonpositive in the η-equation. Then every increase of one or more nonbasic variables will result in a solution where $\eta \leq 13$.

Conclusion: we have found an optimal solution! It is $w_1 = x_2 = w_3 = 0$ and $x_1 = 2, w_2 = 1, x_3 = 1$. The corresponding value of η is 13, and this is called the optimal value.

The simplex method – comments

- geometry: from vertex to adjacent vertex
- phase 1 problem: first feasible solution
- the dictionary approach good for understanding
- in practice: the revised simplex method used
- relies on numerical linear algebra techniques
- main challenges: (degeneracy), pivot rule, update basis efficiently
- commercial systems like CPLEX routinely solves large-scale problems in a few seconds

Matrix version: basis B: $A = [\, B \;\; N \,]$, $Ax = b$ becomes $Bx_B + Nx_N = b$ so $x_B = B^{-1}b - B^{-1}Nx_N$.

The fundamental theorem of LP

Theorem

For every LP problem the following is true:

- If there is no optimal solution, then the problem is either nonfeasible or unbounded.
- If the problem is feasible, there exist a basic feasible solution.
- If the problem has an optimal solution, then there exist an optimal basic solution.

Duality theory

- associated to every LP problem there is another, related, LP problem called the the dual problem
- so primal (P) and dual problem (D).
- the dual may be used to, easily, find bounds on the optimal value in (P)
- may find optimal solution of (P) by solving (D)!

The dual problem

Consider the LP problem (P), the primal problem, given by

$$\max\{c^T x : Ax \leq b,\ x \geq O\}.$$

We define the dual problem (D) like this:

$$\min\{b^T y : A^T y \geq c,\ y \geq O\}.$$

- max and min
- y associated with the constraints in (P)
- constraint ineq. reversed
- c and b switch roles

Lemma

(*Weak duality*) If $x = (x_1, \ldots, x_n)$ is feasible in (P) and $y = (y_1, \ldots, y_m)$ is feasible in (D) we have

$$c^T x \leq b^T y.$$

Proof: From the constraints in (P) and (D) we have

$$c^T x \leq (A^T y)^T x = y^T A x \leq y^T b = b^T y.$$

□

The duality theorem

Theorem

If (P) has an optimal solution x^*, then (D) has an optimal solution and

$$\max\{c^T x : Ax \leq b,\ x \geq O\} = \min\{b^T y : A^T y \geq c,\ y \geq O\}$$

Comments:

- (P) and (D) have the same optimal value when (P) has an optimal solution.
- If (P) (resp. (D)) is unbounded, then (D) (resp. (P)) has no feasible solution.

Interior point methods

- these are alternatives to simplex methods
- may be faster for certain problem instances
- roots in nonlinear optimization

Main idea (in primal-dual int. methods):

- based on duality: solve both (P) and (D) at once
- a special treatment of the optimality property called complementary slack: $x_j z_j = 0$ etc., relaxed compl. slack: $x_j z_j = \mu$ etc.
- solution parameterized by $\mu > 0$, e.g. $x(\mu)$
- convergence: $x(\mu) \to x^*$ as $\mu \to 0$.
- Newton's method etc. , efficient, polynomial method
- more on this in Nonlinear optimization lectures