Linear Programming

Jitkomut Songsiri

Department of Electrical Engineering
Faculty of Engineering
Chulalongkorn University

CUEE

March 6, 2023

Outline

Standard form

Standard form

a general linear program has the form

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Gx \preceq h \\ & Ax = b \end{array}$$

where $G \in \mathbf{R}^{m \times n}$ and $A \in \mathbf{R}^{p \times n}$

- n optimization variables: $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$
- the objective function: $c^T x = \sum_{i=1}^n c_i x_i$
- the inequality constraint: $\sum_{j=1}^n g_{ij} x_j \leq h_i$ for $i = 1, 2, \ldots, m$
- the equality constraint: $\sum_{j=1}^n a_{ij} x_j = b_i$ for $i = 1, 2, \ldots, p$
- the objective function and constraint functions are *linear* in x

called **linear program** (LP) or **linear optimization problem**

Another standard form

LP can also be represented in another form

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \succeq 0 \end{array}$$

using the facts that

- any $x \in \mathbf{R}$ can be written $x = x^+ - x^-$
- $a^T x \leq b \Longleftrightarrow a^T x + s = b, \quad s \succeq 0$

note: we assume A is fat and has full row rank

exercise: transform into the two general forms

$$\begin{array}{ll} \text{minimize} & 2x_1 - x_2 + x_3 \\ \text{subject to} & -3x_1 + x_2 - 5x_3 \leq 3 \\ & 2x_2 + 7x_3 \geq 10 \\ & 3x_2 + 4x_3 = 2 \end{array}$$

Mixed integer programming

if $x \in \mathbf{Z}^n$ (integers) the problem on page 4 is called an **integer linear programming (ILP)**

if some components of x are integers and some are real numbers, the problem is called a **mixed integer linear programming**

examples of integer programing:

- x represents quantities, countable units (pieces)
- number of sale products
- number of persons assigned on a work schedule
- $x \in \{0, 1\}$: **binary integer programming**
- x is status of a functioning unit in factory, '1' is on, '0' is off

Geometrical interpretation

- hyperplane: solution set of a linear equation with coefficient vector $a \neq 0$

$$\{x \mid a^T x = b\}$$

- halfspace: solution set of a linear inequality with coefficient vector $a \neq 0$

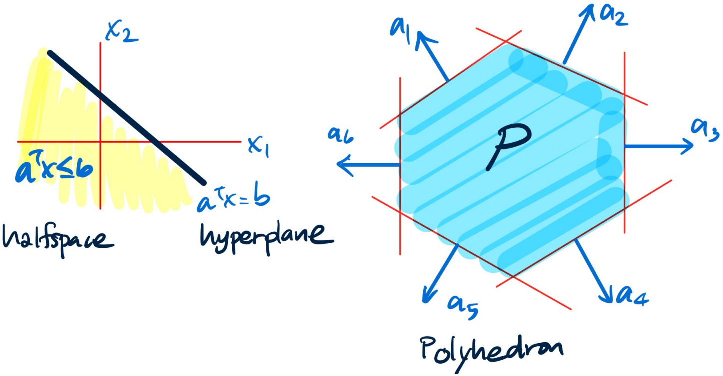
$$\{x \mid a^T x \leq b\}$$

  we say a is the **normal vector**

- polyhedron: solution set of a finite number of linear inequalities

$$\{x \mid a_1^T x \leq b_1, \;\; a_2^T x \leq b_2, \;\; \ldots, \;\; a_m^T x \leq b_m\} = \{x \mid Ax \leq b\}$$

  intersection of a finite number of halfspaces

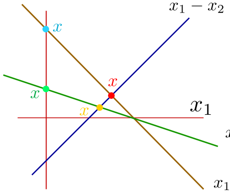


extreme point of $\mathcal{C}$

a vector $x \in \mathcal{C}$ is an extreme point (or a vertex) if we cannot find $y, z \in \mathcal{C}$ both different from x and a scalar $\alpha \in [0, 1]$ such that $x = \alpha y + (1 - \alpha)z$

Solving LPs graphically

LP 1 (left) and LP 2 (right, with non-negative constraints)



$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x_1 + x_2 \leq 6 \\ & x_1 - x_2 \leq 3 \\ & x_1 + 3x_2 \geq 6 \end{array}$$

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x_1 + x_2 \leq 6 \\ & x_1 - x_2 \leq 3 \\ & x_1 + 3x_2 \geq 6 \\ & x_1, x_2 \geq 0 \end{array}$$

- LP 1: feasible set is unbounded but the problem is bounded below for some c

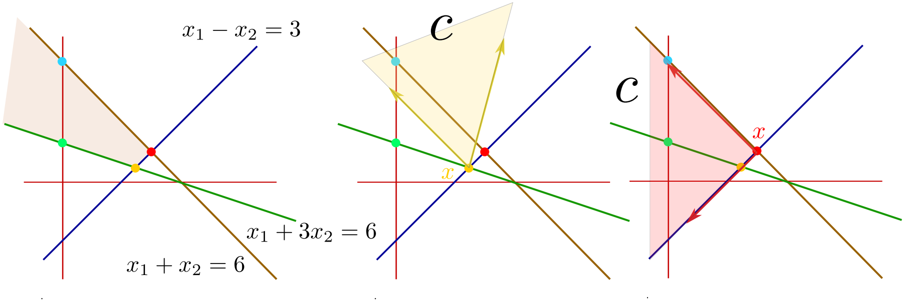
$$c = (0,1), x^\star = \quad c = (-1,0), x^\star = \quad c = (-1,1), x^\star = \quad c = (1,3), x^\star =$$

- LP 2: feasible set is a bounded polyhedron
  - $x^\star = x$ if
  - $x^\star = x$ if
  - $x^\star$ is not unique if

$x^\star = x$ if

$x^\star = x$ if

the directions of c that lead LP 1 to have $x^\star$ at vertices x or x



- for other directions of c than the two cases above, the problem is unbounded below
- for 2-dimensional problems, solutions can be sketched graphically
- LP properties depend on both the objective direction and the feasible set

Properties and simple LPs

Properties

refer to the standard form on page 5

- an LP may not have a solution (constraints are inconsistent or the feasible set is unbounded)
- we assume A is full row rank; if not, considering $Ax = b$
  - depending on A, the system could be inconsistent (hence, no extreme points), or
  - $Ax = b$ contains redundant equations, which can be removed
- if a standard LP has a finite optimal solution then

  a solution can always be chosen from among the vertices of the feasible set

  (called **basic feasible solutions**)
- the dual of an LP is also an LP
- solutions of some simple LPs can be analytically inspected

Simple linear programs

minimize $c^T x$ over each of these simple sets

we can derive an explicit solution of these LPs

- **box constraint:** $l \preceq x \preceq u$
- **probability simplex** (or budget allocation): $\mathbf{1}^T x = 1, x \succeq 0$
- **not all budget is used:** $\mathbf{1}^T x \leq 1, x \succeq 0$
- **halfspace:** $a^T x \leq b$

draw the constraint set and inspect the solution for a given c

Some problems may not look like an LP

example 1: functions that involve ℓ_1 and ℓ_∞ norms

$$\text{minimize} \quad \|Fx - g\|_1 \quad \text{subject to} \quad \|x\|_\infty \leq 1$$

(minimize a cost measured by 1-norm having a worst-case budget constraint)

by introducing u; imposing the constraint: $-u \preceq Fx - g \preceq u$; and noting that

$$\|Fx - g\|_1 = \sum_{i=1}^{m} |f_i^T x - g_i| \leq \mathbf{1}^T u$$

the problem is equivalent to the LP

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T u \\ \text{subject to} & -u \preceq Fx - g \preceq u, \\ & -\mathbf{1} \preceq x \preceq \mathbf{1} \end{array}$$

Example

finding a probability mass function (pmf) of a discrete random variable y

- y takes n possible values as a_i for $i = 1, 2, \ldots, n$ with $0 < a_1 < a_2 < \cdots < a_n$
- $p = (p_1, p_2, \ldots, p_n)$ is a pmf of y: $\mathbf{prob}(y = a_i) = p_i$ for $i = 1, 2, \ldots, n$

given scalar parameters, $a \in \mathbf{R}^n$, $\alpha > 0$ and b, find $p \in \mathbf{R}^n$ from the optimization

$$\begin{array}{ll} \text{maximize} & \mathbf{prob}(y \geq \alpha) \\ \text{subject to} & \mathbf{E}[y] = b \end{array}$$

(find the pmf of y that maximizes the probability and satisfies a given mean)

express the problem as LP with variable p

(recognize that the objective and constraint are linear in p)

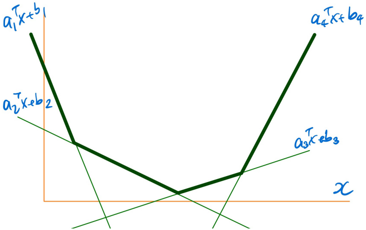
Applications

Linear programs in applications

- piecewise-linear minimization
- ℓ_1-norm and ℓ_∞-norm approximation
- sparse recovery
- separating two sets using hyperplane

Piecewise-linear minimization

a problem of minimizing a piecewise-linear function is in the form:



$$\text{minimize} \quad f(x) := \max_{i=1,2,\ldots,m} (a_i^T x + b_i)$$

$f : \mathbf{R}^n \to \mathbf{R}$ is called a **piecewise-linear** function

- f is obtained by taking a point-wise maximum of m affine functions (convex)
- it is equivalent to LP (with variables x and auxiliary scalar variable t)

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & a_i^T x + b_i \leq t, \quad i = 1, 2, \ldots, m \end{array}$$

Piecewise-linear minimization

$$\text{minimize} \quad c^T z \quad \text{subject to} \quad Gz \preceq h$$

where

$$z = \begin{bmatrix} x \\ t \end{bmatrix}, c = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, G = \begin{bmatrix} a_1^T & -1 \\ a_2^T & -1 \\ \vdots & \vdots \\ a_m^T & -1 \end{bmatrix}, h = \begin{bmatrix} -b_1 \\ -b_2 \\ \vdots \\ -b_m \end{bmatrix}$$

example: minimize $\sum_{i=1}^m \max\{0, a_i^T x + b_i\}$ (related to ReLU function)

can be cast as an LP

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T t \\ \text{subject to} & 0 \preceq t \\ & a_i^T x + b_i \leq t_i, \quad i = 1, 2, \dots, m \end{array}$$

with variable $t \in \mathbf{R}^m$

ℓ_1-norm and ℓ_∞-norm approximations

given $A \in \mathbf{R}^{m \times n}$ and $b \in \mathbf{R}^m$

- ℓ_1-norm approximation: minimize $\|Ax - b\|_1$

  equivalent LP:

  $$\begin{array}{ll} \text{minimize} & \mathbf{1}^T u \\ \text{subject to} & -u \preceq Ax - b \preceq u \end{array}$$

  with variable x and auxiliary variable u

- ℓ_∞-norm (or Chebyshev) approximation: minimize $\|Ax - b\|_\infty$

  equivalent LP:

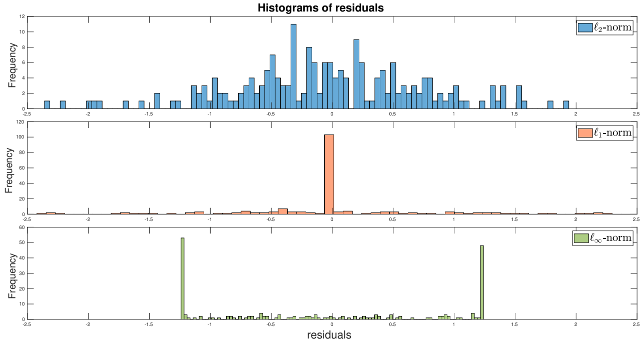
  $$\begin{array}{ll} \text{minimize} & t \\ \text{subjec to} & -t\mathbf{1} \preceq Ax - b \preceq t\mathbf{1} \end{array}$$

  with variable x and auxiliary variable t

ℓ_1- and ℓ_∞-norm approximation results

compare histograms of residuals $Ax - b$ for

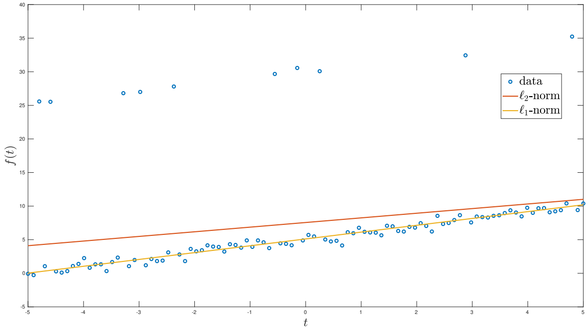
$$x_{\rm ls} = \operatorname{argmin} \|Ax - b\|_2, \quad x_1 = \operatorname{argmin} \|Ax - b\|_1, \quad x_\infty = \operatorname{argmin} \|Ax - b\|_\infty$$



example of $A \in \mathbf{R}^{200\times 100}$: residuals of 1-norm approximation is concentrated at zero

Estimation with outliers

fitting $f(t) = \alpha + \beta t$ to data containing 10% outliers

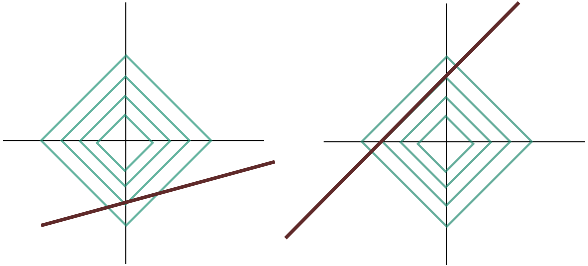


- ℓ_2-norm approximation tends to reduce large residuals occurred from outliers
- ℓ_1-norm has less penalty than ℓ_2 when residuals are large; it is more robust to outliers

Sparse recovery

given $A \in \mathbf{R}^{m \times n}$ (sensor matrix) with $m < n$ and $y \in \mathbf{R}^m$ (measurements)

$$\begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & Ax = y \end{array}$$



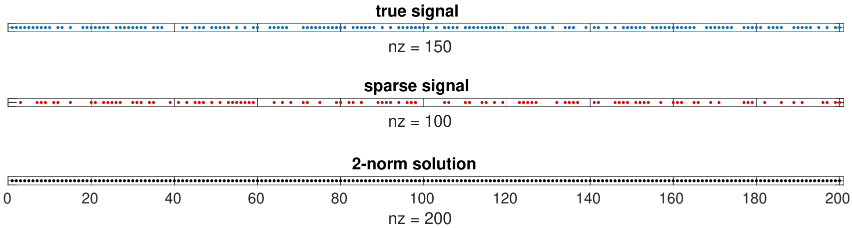
- estimate a sparse signal x that gives the model output matched with measurements
- the constraint makes sense when A is fat (many feasible points)
- equivalent LP (with variables $x, u \in \mathbf{R}^n$)

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T u \\ \text{subject to} & -u \preceq x \preceq u \\ & Ax = y \end{array}$$

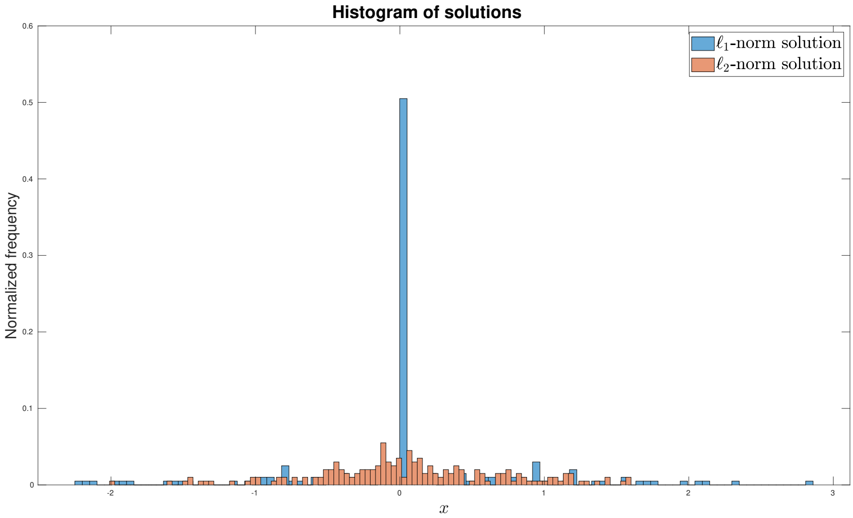
Example of sparse signal estimation

given $A \in \mathbf{R}^{100\times 200}, y \in \mathbf{R}^{100}$ with $y = Ax +$ noise

- the ground-truth signal x has 30 nonzero components
- ℓ_1-norm estimate is sparse while ℓ_2-norm estimate is generally dense
- estimated sparsity is close to the true zero locations



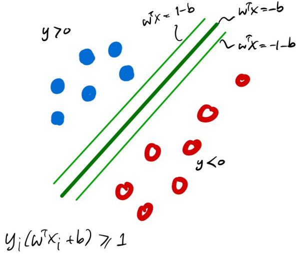
ℓ_1-norm estimate is generally sparser than ℓ_2-norm estimate



Seperating two sets using hyperplane

given: a set of points $\{x_1, \ldots, x_N\}$ with binary labels $y_i \in \{-1, 1\}$

problem: find a hyperplane that strictly separates the two data classes



$$
\begin{aligned}
w^T x_i + b > 0, &\quad \text{if } y_i = 1 \\
w^T x_i + b < 0, &\quad \text{if } y_i = -1
\end{aligned}
$$

the two sets of inequalities can be merged into a single set of N inequalities

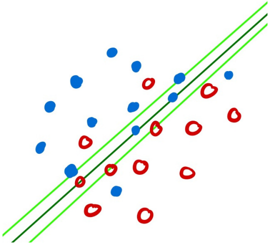
$$y_i(w^T x_i + b) \geq 1, \quad i = 1, 2, \ldots, N$$

since the inequality is homogenous in w and b

many feasible solutions can be found (if the two sets are separable)

Linear separation of non-separable sets

when two sets cannot be strictly separable



$$\underset{w,b}{\text{minimize}} \quad \sum_{i=1}^{N} \max\{0, 1 - y_i(w^T x_i + b)\}$$

equivalent LP: with variables $w \in \mathbf{R}^n, b \in \mathbf{R}, z \in \mathbf{R}^N$

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T z \\ \text{subject to} & 1 - y_i(x_i^T w + b) \leq z_i, \quad i = 1, 2, \ldots, N \\ & z_i \geq 0, \quad i = 1, 2, \ldots, N \end{array}$$

- no penalization when $y_i(w^T x_i + b) \geq 1$
- it is a heuristic method for minimizing # of misclassified points
- a piecewise-linear minimization problem with variables w, b
- related to a soft-margin SVM (but no cost on the hyperplane margin)

Algorithms

Modeling softwares

- accept linear programs in standard notation
- recognize problems that can be converted to LPs
- express the problem in the format required by LP solvers
- examples of modeling packages
  - CVX, YALMIP (on MATLAB)
  - CVXPY, CVXOPT (on Python)
  - AMPL

Numerical methods

- simplex (by Dantzig): move along the vertices of polyhedron when the objective is decreasing
- interior-point: move through the interior points of the feasible region
- many libraries/solvers (both commercial and open-source) on the market
  - `linprog` in MATLAB
  - Pulp or `scipy.optimize.linprog` in Python
  - Gurobi

Sentivity analysis

Perturbed problem

perturbed version of the standard LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Gx \preceq h + u \\ & Ax = b + v \end{array}$$

question: we aim to get information about the sensitivity of the solution with respect to changes in problem data

- how does $p^\star(u, v)$ of the perturbed problem change upon the values of u_i and v_i ?
- if $u_i > 0$, the inequality is loosen, but if $u_i < 0$, the inequality is tighten

Global and local sensitivity analysis

the analysis requires the duality result of LP

λ, ν are Lagrange mulipliers corresponding to inequality and equality, respective

- global analysis: we can derive a lower bound of the perturbed optimal value

$$p^\star(u,v) \geq p^\star(0,0) - \lambda^{\star T} u - \nu^{\star T} v$$

- local analysis: Lagrange multipliers give the rate of change in $p^\star(u,v)$ at $(0,0)$

$$\frac{\partial p^\star(0,0)}{\partial u_i} = -\lambda_i^\star, \quad \frac{\partial p^\star(0,0)}{\partial v_i} = -\nu_i^\star$$

References

1. L. Vandenberghe, *Linear programming*, Lecture notes of EE236A, UCLA
2. D. Bertsimas and J.N. Tsitsiklis, *Introduction to Linear Optimization*, Athena Scientific, 1997
3. I. Griva, S.G. Nash, and A. Sofer, Linear and Nonlinear Optimization, 2nd edition, SIAM, 2009
4. S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge, 2004