Lecture 4: Convex optimization problems

- Linear programming
- Convex sets and functions
- Semidefinite programming
- Duality
- Algorithms

Linear programming (LP)

optimization problem with linear cost function and affine constraints

Linear program in a standard form:

$$\text{minimize} \quad c^\top x \quad \text{subject to} \quad Gx \le h \quad \text{and} \quad Ax = b \qquad \text{(LP)}$$

c, G, h, A, b are given (problem data)
x is an unknown vector of optimization variables

Contrary to least-squares and least-norm, (LP) has no analytic solution

however, it can be solved very efficiently by iterative methods.

Note: recurrent theme — use of quickly convergent iterative methods.

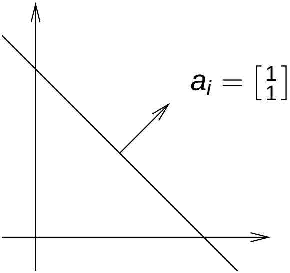
Even for LS and LN problems, iterative methods may have advantage.

Geometric interpretation of LP

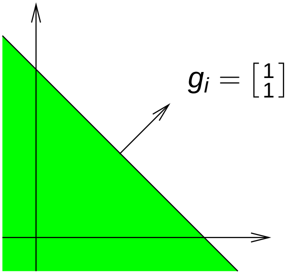
Let $a_i^\top$ be the ith row of A, and $g_i^\top$ be the ith row of G

$a_i^\top x = b_i$ is a hyperplane, perpendicular to a_i (assuming $a_i \neq 0$)

$g_i^\top x \geq h_i$ is a half space (assuming $h_i \neq 0$)



$\{x \mid \begin{bmatrix}1 & 1\end{bmatrix} x = 1\}$



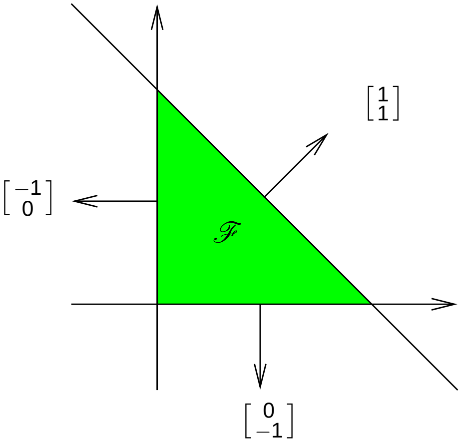
$\{x \mid \begin{bmatrix}1 & 1\end{bmatrix} x \leq 1\}$

Feasible set of (LP)

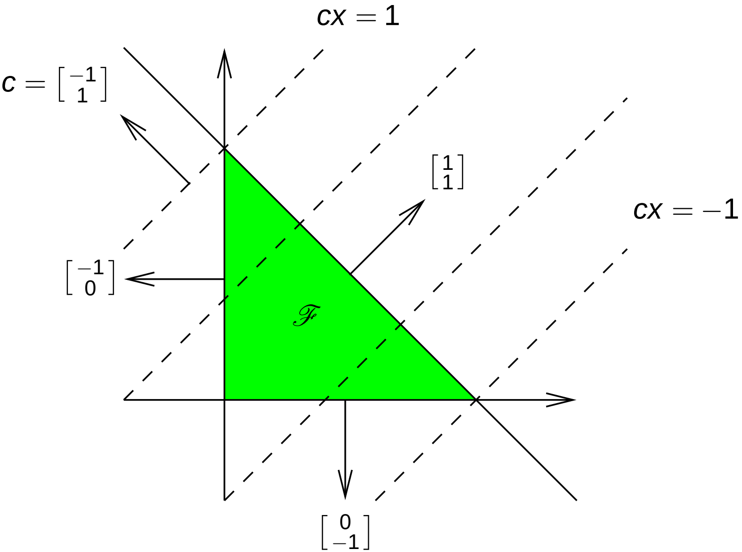
$$\mathscr{F} = \{\, x \mid Gx \geq h,\ Ax = b \,\}$$

intersection of a finite number of half spaces and hyperplanes

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 0 & -1 \\ -1 & 0 \end{bmatrix}}_{G} x \leq \underbrace{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}}_{h}$$



level curves of the objective functions $cx = \text{const}$ are hyperplanes ($\perp c$)



Example: ℓ_∞ approximation problems

The ℓ_∞ approximation problem

$$\text{minimize} \quad \|Ax - b\|_\infty \qquad \text{where} \quad \|e\|_\infty := \max\{\, |e_1|, \dots, |e_m| \,\}$$

is equivalent to the linear program

$$\text{minimize} \quad \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} t \\ x \end{bmatrix} \quad \text{subject to} \quad \begin{bmatrix} -\mathbf{1} & A \\ -\mathbf{1} & -A \end{bmatrix} \begin{bmatrix} t \\ x \end{bmatrix} \leq \begin{bmatrix} b \\ -b \end{bmatrix}$$

where $\mathbf{1} = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}^\top$.

Example: ℓ_1 approximation problems

The ℓ_1 approximation problem

$$\text{minimize} \quad \|Ax - b\|_1 \qquad \text{where} \quad \|e\|_1 := |e_1| + \cdots + |e_m|$$

is equivalent to the linear program

$$\text{minimize} \quad \begin{bmatrix} \mathbf{1}^\top & 0 \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} \quad \text{subject to} \quad \begin{bmatrix} -I & A \\ -I & -A \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} \leq \begin{bmatrix} b \\ -b \end{bmatrix}$$

Standard trick in formulating LPs: introducing “slack” variables.

Linear programming algorithms

- Simplex method (Dantzig, 1947)

  Exploits the fact that one of the vertexes of $\mathscr{F}$ is a solution.

  Searches over the vertexes using a heuristic rule.

  Very efficient in practice although there is no theoretical proof for its efficiency.

- Interior point methods (Karmarkar, 1984)

  Searches inside $\mathscr{F}$, using the Newton method.

  Efficient in practice with theoretical proof for efficiency.

Convex sets

$\mathscr{S} \subseteq \mathbb{R}^n$ is convex if

$$a, b \in \mathscr{S} \quad \Longrightarrow \quad \alpha a + \beta b \in \mathscr{S}, \text{ for all } \alpha, \beta \in \mathbb{R}, \ \alpha + \beta = 1$$

$\{\, x \mid x = \alpha a + \beta b, \ \alpha + \beta = 1 \,\}$ is the line segment between a and b

$\mathscr{S}$ convex if it contains line segments between any two points in $\mathscr{S}$

Examples:

- subspaces
- half spaces
- balls and ellipses
- polyhedra

Ellipsoids

2-norm unit ball in $\mathbb{R}^n$:

$$\mathscr{U} = \{\, x \in \mathbb{R}^n \mid \|x\|_2 \leq 1 \,\}$$

Ellipsoid

$$\mathscr{E} := \{\, Ax + c \mid \|x\|_2 \leq 1 \,\}$$

an image of an affine function $f(x) = Ax + c$ to $\mathscr{U}$

A and c are parameters: A determines the shape and c is the center

Another representation

$$\mathscr{E} := \{\, x \in \mathbb{R}^n \mid (x - c)^\top V(x - c) \leq 1 \,\}$$

where V is a positive definite matrix ($V = (A^\top A)^{-1}$).

Operations that preserve convexity

Checking whether a set is convex can be done using

1. the definition
2. operations that preserved convexity, applied on basic convex sets

Operations that preserve convexity:

- intersection
- projection
- affine mapping

Convex functions

$f : \mathbb{R}^n \to \mathbb{R}$ is convex if

$$f(\alpha a + \beta b) \leq \alpha f(a) + \beta f(b), \qquad \text{for all } a, b \text{ and } \alpha + \beta = 1$$

Link with convex sets — epigraph $\{\, \big(x, f(x)\big) \mid \text{ for all } x \,\}$

f is convex if and only if its epigraph is convex

Examples:

- linear and affine functions
- quadratic functions
- exponential

calculus of convex functions (operations that preserve convexity)

Convex optimization problems

$$\text{minimize} \quad f(x) \quad \text{subject to} \quad g(x) \leq 0 \quad \text{and} \quad h(x) = 0$$

where f and g_i are convex and h is affine

Important property: local minima are global

Examples:

- Least-squares and least-norm
- Linear programming
- Second order cone programming
- Semidefinite programming

How to recognize that a problem is convex?

Semidefinite programming

$$\text{minimize} \quad c^\top x \quad \text{subject to} \quad G(x) \le 0 \quad \text{and} \quad Ax = b \qquad \text{(SDP)}$$

where

$$G(x) = G_0 + G_1 x_1 + \cdots + G_n x_n$$

$G(x) \le 0$ is called a linear matrix inequality (LMI)

LP is a special case of (SDP) with diagonal $G(x)$.

Interior point methods for LP can be generalized to solve SDP.

Example: eigenvalue minimization

$$\text{minimize} \quad \lambda_{\max}(A(x)), \qquad \text{where} \quad A(x) = A_0 + A_1 x_1 + \cdots + A_n x_n$$

is equivalent to

$$\text{minimize} \quad t \quad \text{subject to} \quad A(x) \preceq tI$$

because $\lambda_{\max}(A(x)) < t$ is equivalent to $A(x) \preceq tI$

Example: matrix norm minimization

$$\text{minimize} \quad \|A(x)\|_2, \qquad \text{where} \quad A(x) = A_0 + A_1 x_1 + \cdots + A_n x_n$$

is equivalent to

$$\text{minimize} \quad t \quad \text{subject to} \quad \begin{bmatrix} tI & A(x) \\ A^\top(x) & tI \end{bmatrix} \preceq 0$$

because

$$\|A(x)\|_2 \leq t \quad \iff \quad A^\top(x)A(x) \preceq t^2 I \quad \iff \quad \begin{bmatrix} tI & A(x) \\ A^\top(x) & tI \end{bmatrix} \preceq 0$$

Schur complement

Convert a quadratic matrix equation into an LMI.

$$X = \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix} \geq 0 \qquad \Longleftrightarrow \qquad \begin{cases} A \geq 0 \\ C - B^\top A^{-1} B \geq 0 \end{cases}$$

$S := C - B^\top A^{-1} B$ is the Schur complement of A in X.

Lagrange duality

Consider an optimization problem

$$\text{minimize} \quad f(x) \quad \text{subject to} \quad g(x) \leq 0 \quad \text{and} \quad h(x) = 0 \qquad (1)$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^p$, and $f : \mathbb{R}^n \to \mathbb{R}^m$.

The Lagrangian L for (1) is the function defined by

$$L(x, \lambda, \nu) = f(x) + \lambda^\top g(x) + \nu^\top h(x)$$

The variables λ and ν are called Lagrange multipliers associated with the constraints.

Note: L is a weighted sum of the cost function and the functions, defining the constants.

Lagrange dual function

$$d(\lambda,\nu) := \text{minimize}_x \quad L(x,\lambda,\nu)$$

Independent of f, g, h, the function $-d$ is convex (d is concave).

Lower bound property of d: if $\lambda \geq 0$ and x is a feasible point, then

$$f(x) \geq L(x,\lambda,\nu) \geq d(\lambda,\nu)$$

Therefore, $f(x^*) \geq d(\lambda,\nu)$, where x^* is an optimal point for (1).

Example: linear programming

$$\text{minimize} \quad c^\top x \quad \text{subject to} \quad x \leq 0 \quad \text{and} \quad Ax = b$$

The Lagrangian is

$$\begin{aligned} L(x,\lambda,\nu) &= c^\top x + \lambda^\top x + \nu^\top (Ax - b) \\ &= -b^\top \nu + (c + A^\top \nu - \lambda)^\top x \end{aligned}$$

The Lagrange dual function is

$$d(\lambda,\nu) = \text{minimize}_x \, L(x,\lambda,\nu) = \begin{cases} -b^\top \nu, & c + A^\top \nu - \lambda = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

Lower bound

$$c^\top x^* \geq -b^\top \nu, \qquad \text{if } c + A^\top \nu - \lambda = 0 \quad \text{and} \quad \lambda \geq 0$$

$$(\iff \; c + A^\top \nu \geq 0)$$

Weak and strong duality

Lagrange dual problem

$$\text{maximize} \quad d(\lambda, \nu) \quad \text{subject to} \quad \lambda \geq 0$$

finds best lower bound $d(\lambda^*, \nu^*)$ on the original (primal) problem

- Weak duality: $d(\lambda^*, \nu^*) \leq f(x^*)$
- Strong duality: $d(\lambda^*, \nu^*) = f(x^*)$

Under mild conditions,

strong duality holds for convex optimization problems.

Example: linear programming

$$\text{minimize} \quad c^\top x \quad \text{subject to} \quad x \leq 0 \quad \text{and} \quad Ax = b$$

Lower bound

$$-b^\top \nu, \quad \text{subject to} \quad c + A^\top \nu \geq 0$$

Dual problem

$$\text{maximize} \quad -b^\top \nu \quad \text{subject to} \quad A^\top \nu \geq -c$$

again a linear program.

Karush-Kuhn-Tucker optimality conditions

Necessary optimality conditions:

1. primal feasibility: $g(x) \leq 0$, $h(x) = 0$
2. dual feasibility: $\lambda \geq 0$
3. complementary slackness: $\lambda_i g_i(x) = 0$, for $i = 1, \ldots, m$
4. gradient of the Lagrangian w.r.t. x is zero $\nabla_x L(x, \lambda, \nu) = 0$

For a convex problem they are necessary and sufficient.

Sensitivity analysis

Unperturbed problem

$$\text{minimize} \quad f(x) \quad \text{subject to} \quad g(x) \le 0 \quad \text{and} \quad h(x) = 0 \tag{2}$$

Perturbed problem

$$\text{minimize} \quad f(x) \quad \text{subject to} \quad g(x) \le u_i \quad \text{and} \quad h(x) = v_i \tag{3}$$

The perturbations u and v are parameters.

The dual problem of (3) is

$$p^*(u,v) := \text{maximize} \quad d(\lambda,\nu) - u^\top\lambda - v^\top\nu \quad \text{subject to} \quad \lambda \ge 0$$

where d is the dual function of (2).

We are interested in $p^*(u,v)$ as a function of u and v.

Let λ^* and ν^* be optimal points for the unperturbed probelm.

We have $d(\lambda^*, \nu^*) p^*(0,0)$, so that

$$p^*(u,v) \geq p^*(0,0) - u^\top \lambda^* - v^\top \nu^*$$

where λ^* and ν^* are dual optimal.

> Assuming that strong duality holds, λ^* and ν^* show the sensitivity of the optimal value of the unperturbed problem to perturbations.

Algorithms

- Unconstrained minimization

  steepest descent, Newton method, line search, trust region

- Minimization with equality constraints

- Minimization with inequality constraints

  barrier functions, primal-dual methods

Unconstrained minimization

$$\text{minimize} \quad f(x), \qquad (f \text{ twice differentiable})$$

Minimization methods produce

- a sequence $x^{(k)}$, $k = 0, 1, \ldots$
- starting from a given initial point $x^{(0)}$
- convergent to a minimum point

First order optimality condition

$$\nabla f(x) = 0$$

In general, the condition is only necessary.

For a convex problem, it is necessary and sufficient.

General form of a minimization method

Given initial point $x^{(0)}$

For $k = 1, 2, \ldots$ (till convergence)

- Find search direction Δx.
- Choose step size $t > 0$.
- Update $x := x + t\Delta x$.

Search direction: steepest descent, Newton, quasi-Newton, ...

Step size: exact line search

$$t = \arg\min_{t>0} f(x + t\Delta x)$$

or heuristic rules (backtracking, ...).

Normalized steepest descent step

$$\Delta x = \arg\min_{\|v\|=1} \nabla f^\top(x) v$$

unit norm step with most negative directional derivative

- 2-norm: gradient descent

$$\Delta x = -\nabla f^\top(x)$$

- 1-norm: coordinate descent

$$\Delta x = -\frac{\partial}{\partial x_i} f(x) e_i$$

where $\frac{\partial}{\partial x_i} f(x) = \|\nabla f(x)\|_\infty$

Newton step

$$\Delta x = -\left(\nabla^2 f(x)\right)^{-1} \nabla f(x)$$

minimizes the second order approximation of f

$$\widehat{f}(x+v) \approx f(x) + \nabla^\top(x)v + \frac{1}{2}v^\top \nabla^2 f(x) v$$

The Newton step is affine invariant:

change of coordinates $y = Tx$ results in $\Delta y = T\Delta x$.

The steepest descent step is not affine invariant.

Convergence analysis: (under suitable conditions)

- the steepest descent method is linearly convergent
- Newton's method is quadratically convergent

References

Introductory texts:

- Boyd and Vandenberghe, Convex optimization (available online)
- J. Nocedal & Wright, Numerical optimization

Advanced texts:

- Boyd *et al.*., Linear matrix inequalities in system and control theory (available online)