Short Course
Robust Optimization and Machine Learning

Lecture 2: Convex Optimization

Laurent El Ghaoui

EECS and IEOR Departments
UC Berkeley

Spring seminar TRANSP-OR, Zinal, Jan. 16-19, 2012

Outline

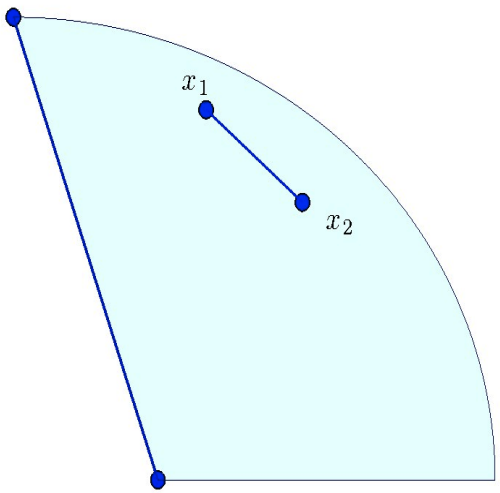
Outline

Convex sets

Definition

A set $\mathcal{C}$ in $\mathbf{R}^n$ is convex if, or any two pair of points, such as x_1 and x_2, the line segment joining the two points is entirely in the set. Mathematically:

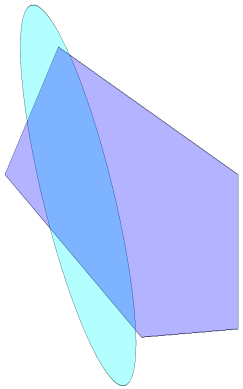
$$\forall\, x_1, x_2 \in \mathcal{C}, \;\; \forall\, \lambda \in [0,1] \;:\; \lambda x_1 + (1 - \lambda x_2) \in \mathcal{C}.$$



Points x_1,x_2 are in the set, so the line between them is.

Convex sets

Intersection rule



The intersection of convex sets is convex: if $\mathcal{C}_\alpha$ is a family of convex sets indexed by $\alpha \in \mathcal{A}$, then $\cap_{\alpha \in \mathcal{A}} \mathcal{C}_\alpha$ is convex. (We allow for infinite sets $\mathcal{A}$.)

Examples:

- the second-order cone $\{(y, t) :: \|y\|_2 \leq t\}$ is convex, as it is the intersection of half-spaces of the form $u^T y \leq t$ with u spanning the unit Euclidean sphere.
- the cone of positive, semi-definite (PSD) matrices is convex, since a $n \times n$ symmetric matrix X is PSD iff $z^T X z \geq 0$ for every n-vector z.

Convex sets

Affine transformation rule

Affine transformations of sets are convex. If $\mathcal{C} \subseteq \mathbf{R}^m$ is convex, then for any $\hat{x} \in \mathbf{R}^n$ and $R \in \mathbf{R}^{n \times m}$, the set $\{\hat{x} + Ru \ : \ u \in \mathcal{C}\}$ is convex.

Example: For a given $m \times n$ matrix A, m-vector b, n-vector c, and scalar d, the set

$$\left\{ x \in \mathbf{R}^n \ : \ \|Ax + b\|_2 \leq c^T x + d \right\}$$

is convex.

Convex sets

Example: chance constraint

If $a \in \mathbf{R}^n$ is a Gaussian random variable with mean $\hat{a}$ and covariance matrix Σ, then the set of points $x \in \mathbf{R}^n$ such that

$$\mathbf{Prob}\{a \ : \ a^T x \leq b\} \geq 0.99$$

is convex, and can be expressed via the second-order cone condition

$$\hat{a}^T x + \kappa \underbrace{\sqrt{x^T \Sigma x}}_{=\|\Sigma^{1/2} x\|_2} \leq b,$$

where $\kappa = \Phi^{-1}(0.99) \approx 2.33$, with Φ the CDF of the standard Gaussian.

Convex functions

Extended value functions

Extended value functions and domain:

- The domain of a function f, denoted by $\mathbf{dom} f$, is the set of points where it is finite.
  Example: $f : x \to \log x$, $\mathbf{dom}\, f = \mathbf{R}_{++}$.
- We can extend functions outside the domain using the values $\pm\infty$.
  Example: $f : x \to \log x$, $x > 0$, $+\infty$ otherwise.

Convex functions

Definition

A (extended value) function $f : \mathbf{R}^n \to \mathbf{R}$ is convex if

$$\forall\, x_1, x_2, \quad \forall \lambda \in [0,1] \;:\; f(\lambda x_1 + (1 - \lambda x_2)) \leq \lambda f(x_1) + (1-\lambda) f(x_2).$$

(This requires the domain to be convex.)

Examples:

- $f : x \to 1/x$ if $x > 0$, $+\infty$ otherwise, is convex.
- $f : x \to 1/x$ if $x \neq 0$, $+\infty$ otherwise, is not convex.

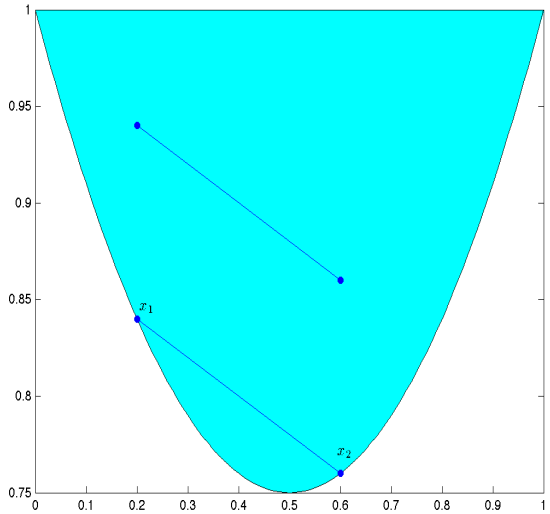
Convex functions

Epigraph

Epigraph of a function:

$$\mathbf{epi}\, f := \left\{ (x,t) \in \mathbf{R}^n \times \mathbf{R} \ : \ f(x) \leq t \right\}.$$

$$f \text{ convex} \iff \mathbf{epi}\, f \text{ is convex.}$$

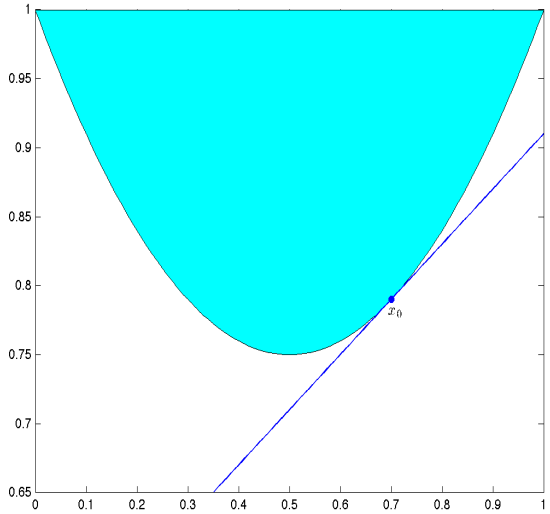


Convex functions

Alternate characterizations

f differentiable, convex $\iff$

$$\forall\, x, x_0 \;:\; f(x) \geq f(x_0) + \nabla f(x_0)^T (x - x_0).$$



Affine approximation to f at any point is global lower bound on f.

Convex functions

Alternate characterizations

f twice differentiable, convex if and only if its Hessian

$$\nabla^2 f(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right)_{1 \leq ij \leq n}$$

is positive semidefinite (PSD) for every x.

In general, this condition is hard to check. For *quadratic* functions though, it is easy, as seen next.

Quadratic functions

Representation via symmetric matrices

A quadratic function $q : \mathbf{R}^n \to \mathbf{R}$ can be represented as

$$q(x) = \frac{1}{2} x^T Q x + b^T x + c,$$

for appropriate symmetric matrix Q, vector $b \in \mathbf{R}^n$, and scalar c.

Positive semidefinite matrices

Flashback from linear algebra:

- A square matrix A has n (possibly non-distinct) eigenvalues, which are (in general complex) numbers that solve $\det(\lambda I - A) = 0$.
- Symmetric matrices have real eigenvalues only.
- A symmetric matrix Q is said to be *positive semi-definite* (PSD) if

$$\forall\, x \;:\; x^T Q x \geq 0.$$

  We write $Q \succeq 0$.

Eigenvalue decomposition for symmetric matrices

Theorem (EVD of symmetric matrices)

We can decompose any symmetric $p \times p$ matrix Q as

$$Q = U \Lambda U^T = \sum_{i=1}^{p} \lambda_i u_i u_i^T,$$

where $\Lambda = \mathbf{diag}(\lambda_1, \ldots, \lambda_p)$, with $\lambda_1 \geq \ldots \geq \lambda_n$ the eigenvalues, and $U = [u_1, \ldots, u_p]$ is a $p \times p$ orthogonal matrix ($U^T U = I_p$) that contains the eigenvectors of Q.

Corollary

$$Q \succeq 0 \iff \textit{every eigenvalue is non-negative.}$$

Convex functions

Convex quadratic functions

A quadratic function $q : \mathbf{R}^n \to \mathbf{R}$ can be represented as

$$q(x) = \frac{1}{2}x^T Qx + b^T x + c,$$

for appropriate symmetric matrix Q, vector $b \in \mathbf{R}^n$, and scalar c.
Since $\nabla^2 q(x) = Q$:

Theorem (Convex quadratic functions)

q convex $\iff Q \succeq 0$.

Convex functions

Some properties

- Pointwise maxima of convex functions are convex: if f_α is a family of convex functions sets indexed by $\alpha \in \mathcal{A}$, then f with values $\max_{\alpha \in \mathcal{A}} f_\alpha(x)$ is convex.
- Conversely, any convex function can be represented as maxima of affine ones. (Hence, "convex means max-linear").
- If f is convex and component-wise monotone increasing and $g_1, \ldots, g_k$ are convex, then $h = f \circ g$ with values $h(x) = f(g_1(x), \ldots, g_k(x))$ is convex.
  Proof: the epigraph of h can be represented via convex constraints:

$$h(x) \leq t \iff \exists\, u \in \mathbf{R}^k \;:\; f(u) \leq t, \;\; g_i(x) \leq u_i, \;\; i = 1, \ldots, k.$$

Convex functions

Examples

The following functions are convex:

- Norms, such as l_1, l_2 and l_∞. Hence, for example, $x \to \|Ax + b\|_2$, with A a matrix and b a vector, is convex.
- The function $x \to \lambda_{\max}(F(x))$, with $F(x) = F_0 + x_1 F_1 + \ldots + x_m F_m$ and affine combination of given symmetric matrices, and $\lambda_{\max}$ is the largest eigenvalue, is convex.

Convex problems

Definition

The problem in standard form

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

is convex if the functions $f_0, \ldots, f_m$ are all convex. Here $A \in \mathbf{R}^{p \times n}$, $b \in \mathbf{R}^p$ are given.

Note that only affine equality constraints are allowed.

Definition

Maximization problems

The problem in standard form

$$\max_x \; f_0(x) \text{ subject to } \quad \begin{array}{l} f_i(x) \leq 0, \quad i = 1, \ldots, m, \\ Ax = b, \end{array}$$

is convex if

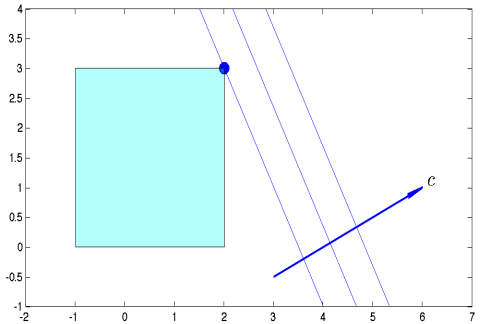
- The function f_0 is concave.
- The functions $f_1, \ldots, f_m$ are all convex.

Problem classes

Linear programming

Linear programming (LP) involves the minimization of a linear function over a polytope:

$$\min_x \; c_0^T x \;:\; c_i^T x + d_i \geq 0, \quad i = 1, \ldots, m.$$



The problem

$$\begin{array}{ll} \min\limits_x & 3x_1 + 1.5x_2 \\ \text{s.t.} & -1 \leq x_1 \leq 2, \\ & 0 \leq x_2 \leq 3 \end{array}$$

is an LP, since the objective and constraint functions are all affine.

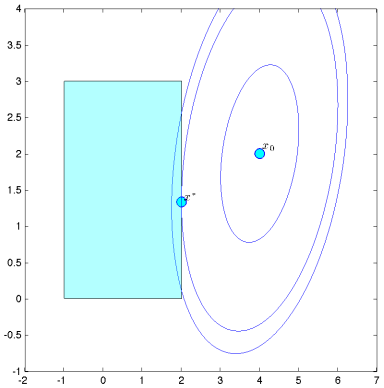
Problem classes

Quadratic programming

Quadratic programming (QP) involves the minimization of a quadratic convex function over a polytope.

$$\min_x \; c^T x + x^T Q x \;:\; c_i^T x + d_i \geq 0, \quad i = 1, \ldots, m.$$

Here, Q must be PSD.



The problem

$$\begin{array}{ll} \min_x & x_1^2 - x_1 x_2 + 2x_2^2 \\ & -3x_1 - 1.5x_2 \\ \text{s.t.} & -1 \leq x_1 \leq 2, \\ & 0 \leq x_2 \leq 3 \end{array}$$

is a QP, since objective is quadratic convex, and the the constraint functions are all affine.

Example of a quadratic program

The slalom problem

The problem seen in Overview lecture can be formulated as a QP:

$$\min_{v} \; v_1 u_0 + \sum_{i=1}^{n} \frac{(y_i + v_{i+1} - v_i)^2}{\sigma_i^2} \; : \; \|v\|_\infty \leq c.$$

where σ_i (resp. y_i) is the horizontal (resp. vertical) distance between the middle of the gates.

Problem classes

Second-order cone programming

Second-order cone programming (SOCP) generalizes LP and QP via the inclusion of Euclidean norms in the constraint functions.

$$\min_x \; c_0^T x \;:\; \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \quad i = 1, \ldots, m.$$

Application: Chance-constrained linear programming

$$\min_x \; c^T x \;:\; \mathbf{Prob}\{a_i^T x \leq b\} \geq 0.99, \quad i = 1, \ldots, m,$$

where each a_i is a Gaussian random variable with mean $\hat{a}_i$ and covariance matrix Σ_i, $i = 1, \ldots, m$.

Problem classes

Semi-definite programming programming

Semi-definite programming (SDP) involves the minimization of a linear function over the constraint that a symmetric matrix affine in the decision variables be positive-semidefinite:

$$\min_x \; c_0^T x \; : \; F_0 + \sum_{i=1}^n x_i F_i \succeq 0.$$

(Here, $F_0, \ldots, F_n$ are given symmetric matrices.)

Application: relaxation of Boolean problems, see later.

Outline

Lagrangian

Consider the (non-necessarily convex) "primal" problem in standard form

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

We can write the problem as an unconstrained one:

$$p^* := \min_x \; \max_{y \geq 0} \; \mathcal{L}(x, y),$$

where $\mathcal{L}$ is the Lagrange function :

$$\mathcal{L}(x, y) := f_0(x) + \sum_{i=1}^{m} y_i f_i(x).$$

Min-max inequality

Theorem (Min-max inequality)

For any function $\mathcal{L}$ of two variables:

$$\min_x \max_y \mathcal{L}(x, y) \geq \max_y \min_x \mathcal{L}(x, y).$$

Proof.

In the LHS, maximizing Player y has full information on x (y^* is a function of x). The RHS is better (smaller) for the minimizing Player x, as then y has no information on x. ■

Weak duality

Weak duality is the process of applying the min-max inequality to the problem in unconstrained form:

$$p^* = \min_x \max_{y \geq 0} \mathcal{L}(x, y) \geq d^* := \max_{y \geq 0} \min_x \mathcal{L}(x, y).$$

The dual problem is

$$d^* = \max_{y \geq 0} G(y), \text{ where } G(y) := \min_x \mathcal{L}(x, y)$$

is called the dual function .

Since G is concave, the dual problem is convex (but not necessarily easy to solve!)

Example

LP duality

Linear programming:

$$p^* := \min \; c^T x \; : \; Ax \leq b.$$

Lagrangian:

$$\mathcal{L}(x, y) = c^T x + y^T (Ax - b).$$

Dual is also an LP

$$p^* \geq d^* = \max_{y \geq 0} \; \min_{x} \; \mathcal{L}(x, y) = \max_{y \geq 0, \; A^T y + c = 0} \; -b^T y.$$

Turns out to be equivalent ($p^* = d^*$), as seen later.

Example of weak duality

A Boolean problem

Let $W = W^T$ be a given $n \times n$ matrix.

$$p^* := \max_x \; x^T W x \; : \; x_i^2 = 1, \quad i = 1, \ldots, n$$

- Arises in *e.g.*, segmentation problems (with W the Laplacian matrix of a graph).
- Hard combinatorial problem.
- Duality yields a provably good, non-trivial bound.

Example of weak duality

Dual of Boolean problem

We can express the problem in an *unconstrained* way

$$p^* = \min_x \max_y \mathcal{L}(x, y)$$

where $\mathcal{L}$ is the Lagrangian:

$$\mathcal{L}(x, y) := x^T W x + \sum_{i=1}^{n} y_i (1 - x_i^2)$$

Dual function:

$$G(y) := \max_x \mathcal{L}(x, y) = \begin{cases} \sum_{i=1}^{n} y_i & \text{if } diag(y) - W \text{ is PSD,} \\ +\infty & \text{otherwise.} \end{cases}$$

Dual problem is an SDP:

$$p^* \leq d^* := \max_y G(y) = \max_y \sum_{i=1}^{n} y_i \ : \ \mathbf{diag}(y) \succeq W.$$

Example of weak duality

CVX syntax CVX is a matlab prototyping tool for convex optimization [1].

Here is a matlab snippet that solves for the dual bound, assuming W, n are in the workspace:

```
cvx_begin
    variable y(n,1);
    minimize( sum(y) );
    subject to
        diag(y)-W == semidefinite(n) ;
cvx_end
dstar = sum(y);
```

Strong duality

Definition & main result

Recall weak duality:

$$p^* = \min_x \max_{y \geq 0} \mathcal{L}(x, y) \geq d^* := \max_{y \geq 0} \min_x \mathcal{L}(x, y).$$

We say that strong duality holds when $p^* = d^*$.

- It holds when the primal problem is convex and strictly feasible.
- It holds for LPs that are feasible.
- It holds in some very specific non-convex problems. For example, if $m = 1$ and both f_0, f_1 are quadratic.

Example of strong duality

Bidual of Boolean problem

Return to Lagrange relaxation ("dual") of Boolean problem:

$$d^* = \max_y \sum_{i=1}^n y_i \ : \ \mathbf{diag}(y) \succeq W.$$

Express it as max-min problem:

$$d^* = \max_y \min_{X \succeq 0} \sum_{i=1}^n y_i + \mathbf{Tr}\, X(\mathbf{diag}(y) - W).$$

Here we use the fact that for any symmetric matrix Q,

$$\max_{X \succeq 0} \mathbf{Tr}\, QX = \begin{cases} 0 & \text{if } Q \preceq 0, \\ +\infty & \text{otherwise.} \end{cases}$$

Exchanging min and max leads to the so-called *bidual* :

$$p^{**} := \max_X \mathbf{Tr}\, WX \ : \ X \succeq 0, \ \ X_{ii} = 1, \ \ i = 1, \ldots, n.$$

Example of strong duality

Bidual of Boolean problem

Bidual:

$$\max_X \ \mathbf{Tr}\, WX \ : \ X \succeq 0, \quad X_{ii} = 1, \quad i = 1, \ldots, n.$$

- Also an SDP, with objective value equal to d^* (*strong duality holds*).
- Can obtain this directly by expressing original problem in terms of matrix variable $X := xx^T$, and dropping the rank constraint on X.
- Suggests a way to generate primal feasible points from bidual variable X, using distribution with mean 0 and covariance matrix given by X.
- Allows to prove that $(2/\pi)d^* \leq p^* \leq d^*$ (independent of problem size!).

Example of strong duality

CVX syntax

Here is a matlab snippet that solves the bidual, assuming W, n are in the workspace:

```
cvx_begin
    variable X(n,n) symmetric;
    maximize( trace(X*W) );
    subject to
        X == semidefinite(n) ;
        diag(X) == 1;
cvx_end
dstar = trace(X*W);
```

Alternatively, we can use the `dual` command in the original dual problem:

```
cvx_begin
    variable y(n,1);
    dual variable X
    minimize( sum(y) );
    subject to
        diag(y)-W == semidefinite(n) : X;
cvx_end
```

Optimality conditions

If the primal problem is convex, and both primal and dual are strictly feasible, then

- Strong duality holds;
- both problems are attained;
- the "Karush-Kuhn-Tucker" (KKT) conditions

$$x^* \in \arg\min_x \mathcal{L}(x, y^*), \quad y_i^* f_i(x^*) = 0, \quad i = 1, \ldots, m.$$

characterize optimality (primal-dual feasible pairs are optimal iff they satisfy them).

Applications of duality

- Sensitivity analysis of convex problems.
- Stopping criteria in convex optimization algorithms.
- Decomposition methods for large-scale convex optimization.
- Bounds and approximations for non-convex problems.
- For convex problems, leads to often surprising connections.

Outline

References

S. Boyd and M. Grant.
The CVX optimization package, 2010.

S. Boyd and L. Vandenberghe.
Convex Optimization.
Cambridge University Press, 2004.