Convexity and Duality

GIAN Short Course on Optimization: Applications, Algorithms, and Computation

Sven Leyffer

Argonne National Laboratory

September 12-24, 2016

Outline

Overview: Convexity and Duality

Convexity and duality are important optimization concepts.

- Convexity can replace the 2^{nd} order conditions
- Convexity guarantees global optimality
  ... snag: rarely holds in practice
- Duality is a transformation for (convex) optimization problems
- Duality can provide lower bounds & plays a role in MIPs

Start by discussing convexity

Outline

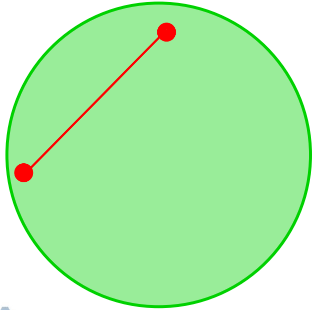
Convexity

Definition (Convex Set)

Set $\mathcal{S} \subset \mathbb{R}^n$ is convex, iff

$$x^{(0)}, x^{(1)} \in \mathcal{S} \;\Rightarrow\; (1-\theta)x^{(0)} + \theta x^{(1)} \in \mathcal{S} \quad \forall \theta \in [0,1]$$

A set is convex, if for any two points in the set, the line between the points also lies in the set.



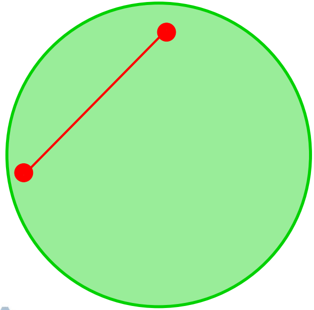
Convexity

Definition (Convex Set)

Set $\mathcal{S} \subset \mathbb{R}^n$ is convex, iff

$$x^{(0)}, x^{(1)} \in \mathcal{S} \;\Rightarrow\; (1-\theta)x^{(0)} + \theta x^{(1)} \in \mathcal{S} \quad \forall \theta \in [0,1]$$

A set is convex, if for any two points in the set, the line between the points also lies in the set.



Nonconvex

Convex

Examples of Convex Sets

Definition (Convex Set)

Set $\mathcal{S} \subset \mathbb{R}^n$ is convex, iff

$$x^{(0)}, x^{(1)} \in \mathcal{S} \;\Rightarrow\; (1-\theta)x^{(0)} + \theta x^{(1)} \in \mathcal{S} \quad \forall \theta \in [0,1]$$

More Examples of Convex Sets

- $\emptyset$, $\mathbb{R}^n$, single point, line, hyperplane $(a^T x = b)$
- Half-spaces are convex: $a^T x \geq b$
  $\Rightarrow$ any linear feasible set is convex
- Given $\hat{x} \in \mathbb{R}^n$, the ball, $\|x - \hat{x}\| \leq r$ is convex
- $\mathcal{S}_i \subset \mathbb{R}^n$, $i = 1, \ldots, m$ convex $\Rightarrow \mathcal{S} = \bigcap_{i=1}^{m} \mathcal{S}_i$ also convex

Convex Combination and Convex Hull

Definition (Convex Combination and Convex Hull)

- For $x^{(1)}, \dots, x^{(m)} \in \mathbb{R}^n$, the point

$$x_\lambda := \sum_{i=1}^{m} \lambda_i x^{(i)}, \quad \text{with } \sum_{i=1}^{m} \lambda_i = 1,\ \lambda_i \geq 0$$

  is convex combination of $x^{(1)}, \dots, x^{(m)}$
- Convex hull, $\text{conv}(\mathcal{S})$, of set $\mathcal{S} \subset \mathbb{R}^n$ is set of all convex combinations of all points in $\mathcal{S}$:

$$\left\{ x = \sum_{i=1}^{n} \lambda_i x^{(i)} \quad \text{with } \sum_{i=1}^{m} \lambda_i = 1,\ \lambda_i \geq 0,\ x^{(i)} \in \mathcal{S} \right\}$$

In $\mathbb{R}^2$, put nails into points, then snap rubber band around nails to get the convex hull.

Feasible Set of an LP or QP

Theorem (Convexity of Linear Feasible Sets)

Feasible set of an LP or QP is convex:

$$\mathcal{F}_{LP} := \{x \mid Ax = b,\ x \geq 0\}$$

Proof. Let $x^{(0)}, x^{(1)} \in \mathcal{F}_{LP}$ and $\theta \in [0,1]$.

$\Rightarrow (1-\theta)x^{(0)} + \theta x^{(1)} \geq 0$, because $x^{(0)}, x^{(1)} \geq 0$, $\theta, 1-\theta \geq 0$

Now consider linear constraints:

$$A\big((1-\theta)x^{(0)} + \theta x^{(1)}\big) = (1-\theta)Ax^{(0)} + \theta Ax^{(1)} = (1-\theta)b + \theta b = b$$

because $x^{(0)}, x^{(1)} \in \mathcal{F}_{LP}$ satisfy $Ax = b$.

Thus, we conclude that $(1-\theta)x^{(0)} + \theta x^{(1)} \in \mathcal{F}_{LP}$ $\square$

Extreme Points

Definition (Extreme Points)

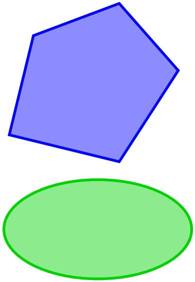
$x \in \mathcal{S}$ is extreme point, iff for any $x^{(0)}, x^{(1)} \in \mathcal{S}$

$$x = (1-\theta)x^{(0)} + \theta x^{(1)}, \quad \text{with } 0 < \theta < 1$$

implies $x = x^{(0)} = x^{(1)}$.

Extreme Points

- Vertices of feasible set
- Cannot be written as strict convex combination
- Play important role in algorithms:
  1. Iterates of Simplex method for LP
  2. Extreme points can be critical points in global optimization

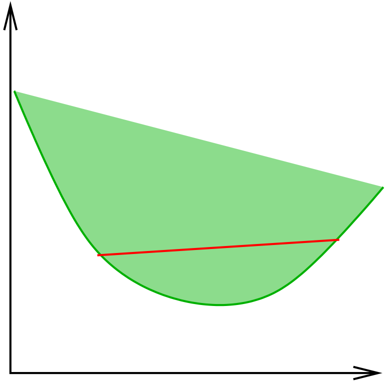


Convex Functions

Definition (Convex Function)

$f(x)$ is a convex function, iff its epigraph is convex set:

$$\Leftrightarrow \; f\left((1-\theta)x^{(0)} + \theta x^{(1)}\right) \le (1-\theta) f\left(x^{(0)}\right) + \theta f\left(x^{(1)}\right), \quad \forall \theta \in [0,1]$$

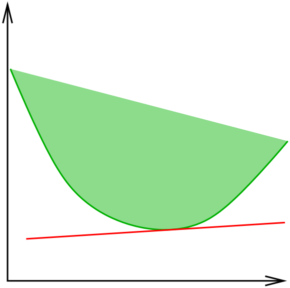


Alternative Definition of Convex Functions

Definition (Convex Differentiable Function)

$f(x)$ differentiable is convex, iff any tangent is supporting hyperplane:

$$f\left(x^{(1)}\right) \geq f\left(x^{(0)}\right) + \left(x^{(1)} - x^{(0)}\right)^T \nabla f\left(x^{(0)}\right)$$



Examples of convex functions:

- Linear functions: $a^T x + b$
- Quadratics with Hessian, $G \succeq 0$
- Norms, $\|x\|$
- Convex combinations of convex functions

Alternative Definition of Convex Functions

Definition (Convex Smooth Function)

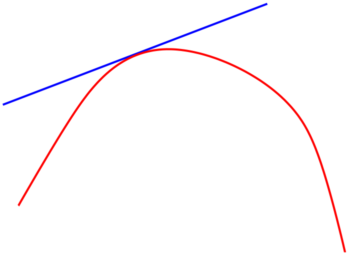
$f(x)$ twice continuously differentiable is convex, iff $\nabla^2 f(x) \succeq 0$ positive semi-definite.

More Examples of convex functions:

- Quadratics with Hessian, $G \succeq 0$

Definition (Concave Function)

$f(x)$ is concave, iff $-f(x)$ is convex.



Implications

- Convave function lies below any linear tangent
- Hessian of concave function is negative semi-definite

Convex Programming Problem

Convex Programming Problem note reversed sign

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & c(x) \leq 0, \end{array}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $c : \mathbb{R}^n \to \mathbb{R}^m$ convex functions.

Theorem (Global Solution of Convex Programs)

- *Local solution x^* of convex program is global solution.*
- *The set of global solutions is convex.*

Theorem (KKT Conditions are Necessary and Sufficient)

KKT conditions are necessary and sufficient for a global minimum of a convex program.

Proofs. Exercises.

Outline

Duality for Convex Programs

Duality is transformation for convex programs

$$(P) \quad \begin{cases} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & c(x) \le 0, \end{cases}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $c : \mathbb{R}^n \to \mathbb{R}^m$ smooth convex functions.

Theorem (Wolfe Dual)

If x^* solves (P), if $f(x)$ and $c(x)$ are smooth convex functions, and if MFCQ holds, then (x^*, y^*) solves the dual problem

$$(D) \quad \begin{cases} \underset{x,y}{\textit{maximize}} & \mathcal{L}(x,y) \\ \textit{subject to} & \nabla_x \mathcal{L}(x,y) = 0 \\ & y \ge 0, \end{cases}$$

where Lagrangian $\mathcal{L}(x,y) = f(x) + y^T c(x)$. Moreover, $f^* = \mathcal{L}^*$.

Proof of Wolfe Dual

Let x^* solve primal (satisfying MFCQ)

$$(P) \qquad \begin{cases} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & c(x) \leq 0, \end{cases}$$

$\Rightarrow \exists\ y^* \geq 0$ such that (x^*, y^*) is KKT point:

$$\nabla_x \mathcal{L}(x^*, y^*) = 0, \quad c_i(x^*) y_i^* = 0$$

$\Rightarrow (x^*, y^*)$ feasible in dual

$$(D) \qquad \begin{cases} \underset{x,y}{\text{maximize}} & \mathcal{L}(x, y) \\ \text{subject to} & \nabla_x \mathcal{L}(x, y) = 0 \\ & y \geq 0, \end{cases}$$

and $f^* = \mathcal{L}^*$ ... now show there is no better solution

Proof of Wolfe Dual

Let (x', y') be any other feasible point of dual

$$(D) \quad \begin{cases} \underset{x,y}{\text{maximize}} & \mathcal{L}(x,y) \\ \text{subject to} & \nabla_x \mathcal{L}(x,y) = 0 \\ & y \geq 0, \end{cases}$$

Now show $\mathcal{L}(x^*, y^*) \geq \mathcal{L}(x', y')$ to show optimality:

$$\mathcal{L}(x^*, y^*) = f^* \geq f^* + \sum_{i=1}^{m} y'_i c_i(x^*) = \mathcal{L}(x^*, y')$$

because $c_i(x^*) \leq 0$, and $y'_i \geq 0$ implies $\sum y'_i c_i(x^*) \leq 0$.
Since $\mathcal{L}(x,y)$ is convex, use supporting hyperplane result:

$$\mathcal{L}(x^*, y') \geq \mathcal{L}(x', y') + (x^* - x')^T \nabla_x \mathcal{L}(x', y') = \mathcal{L}(x', y')$$

Hence, $\mathcal{L}(x^*, y^*) \geq \mathcal{L}(x', y')$ and (x^*, y^*) is optimal.

Dual of Linear Program in Standard Form

Dual of LP in standard form

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

Lagrangian: $\mathcal{L}(x, y, z) = c^T x - y^T (Ax - b) - z^T x$

Wolfe-dual

$$\begin{aligned} \underset{x,y,z}{\text{maximize}} \quad & \mathcal{L}(x, y, z) \\ \text{subject to} \quad & \nabla_x \mathcal{L}(x, y, z) = 0 \\ & z \geq 0 \end{aligned}$$

Dual of Linear Program in Standard Form

Lagrangian: $\mathcal{L}(x,y,z) = c^T x - y^T(Ax-b) - z^T x$

Wolfe-dual

$$\begin{array}{ll} \underset{x,y,z}{\text{maximize}} & \mathcal{L}(x,y,z) \\ \text{subject to} & \nabla_x \mathcal{L}(x,y,z) = 0 \\ & z \geq 0 \end{array}$$

Use first-order condition to eliminate (x,z):

$$\nabla_x \mathcal{L}(x,y,z) = 0 \quad \Leftrightarrow \quad c - A^T y - z = 0 \quad \Leftrightarrow \quad c - A^T y = z \geq 0$$

Simplify objective (eliminate $z = c - A^T y$)

$$\mathcal{L}(x,y,z) = c^T x - y^T(Ax-b) - \left(c - A^T y\right)^T x = y^T b$$

Dual of Linear Program in Standard Form

Primal LP in standard form

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

Dual LP:

$$\begin{aligned} \underset{y}{\text{maximize}} \quad & b^T y \\ \text{subject to} \quad & A^T y \leq c \end{aligned}$$

- Any feasible point of dual gives lower bound on primal
- Given primal (dual) solution easy to get dual (primal) solution

Dual of Another Linear Program

Dual of LP in not in standard form

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & A^T x \geq b \\ & x \geq 0 \end{aligned}$$

Introduce multipliers $y \geq 0$ of $A^T x \geq 0$ and $z \geq 0$ for $x \geq 0$

Can show that dual LP is

$$\begin{aligned} \underset{y}{\text{maximize}} \quad & b^T y \\ \text{subject to} \quad & Ay \leq c \\ & y \geq 0 \end{aligned}$$

... see exercise this afternoon.

This primal/dual pair is often called the symmetric dual

Getting Dual Information from AMPL

Consider LP

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & a_i^T x = b_i \quad i \in \mathcal{E} \\ & a_i^T x \ge b_i \quad i \in \mathcal{I}, \end{aligned}$$

Question

How do we get duals (Lagrange multipliers) from AMPL

AMPL provides reduced costs (variable duals) and constraint multipliers in different formats:

```
display _varname, _var.rc;
display _conname, _con;
```

Note difference between variables and constraints!

Dual of Quadratic Program in Standard Form

Dual of QP

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \frac{1}{2}x^T Gx + g^T x \\ \text{subject to} & A^T x \geq b \end{array}$$

Lagrangian: $\mathcal{L}(x, y, z) = \frac{1}{2}x^T Gx + g^T x - y^T \left(A^T x - b\right)$

Wolfe-dual

$$\begin{array}{ll} \underset{x,y}{\text{maximize}} & \mathcal{L}(x, y) \\ \text{subject to} & \nabla_x \mathcal{L}(x, y) = 0 \\ & y \geq 0 \end{array}$$

... not as nice as LP dual

Dual of Quadratic Program in Standard Form

Lagrangian: $\mathcal{L}(x, y) = \frac{1}{2}x^T Gx + g^T x - y^T \left(A^T x - b\right)$

Wolfe-dual

$$\begin{array}{ll} \underset{x,y}{\text{maximize}} & \mathcal{L}(x, y) \\ \text{subject to} & \nabla_x \mathcal{L}(x, y) = 0 \\ & y \geq 0 \end{array}$$

As before, look at first-order condition

$$\nabla_x \mathcal{L}(x, y) = 0 \quad \Leftrightarrow \quad Gx + g - Ay = 0$$

$$\Leftrightarrow \quad Gx = Ay - g \quad \Leftrightarrow \quad x = G^{-1}(Ay - g)$$

... can eliminate x, provided G^{-1} nonsingular (i.e. positive def.)

Dual of Quadratic Program in Standard Form

Primal QP

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \frac{1}{2}x^T G x + g^T x \\ \text{subject to} & A^T x \geq b \end{array}$$

Eliminating $x = G^{-1}(Ay - g)$ gives ...

Dual QP

$$\begin{array}{ll} \underset{y}{\text{maximize}} & -\frac{1}{2}y^T(A^T G^{-1} A)y + y^T(b + A^T G^{-1} g) - \frac{1}{2}g^T G^{-1} g \\ \text{subject to} & y \geq 0 \end{array}$$

... bound constrained QP, but involves inverse G^{-1}

Discussion of Duality

$$\begin{cases} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & c(x) \le 0, \end{cases} \quad \xrightarrow{\text{dual}} \quad \begin{cases} \underset{x,y}{\text{maximize}} & \mathcal{L}(x,y) \\ \text{subject to} & \nabla_x \mathcal{L}(x,y) = 0 \\ & y \ge 0, \end{cases}$$

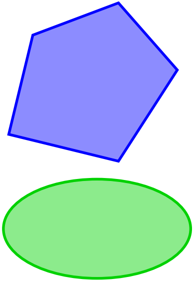
Remark (Discussion of Duality)

- *Duals* $y_i^* > 0$ *implies* $c_i(x^*) = 0$ *... indicates active set!*
- *Transformation interesting computationally, if*
  - $m \gg n$ *many more constraints than variables*
  - *Can easily eliminate primal variables, e.g. QP with* $G = I$
    ... e.g. bundle methods for nonsmooth optimization
- *Dual gives lower bound on primal optimum*
  ... use in MIP to cut-off branches ... any feasible point

Summary and Teaching Points

Convexity

- Linear Programs are convex
- Quadratic Programs with positive semi-definite Hessian are convex
- Rarely holds in practice
  … but really useful if it does
- Convex problems can be solved to global optimality
- Favorable complexity results & algorithms



Duality

- Transformation for convex optimization
- Creates problem that can provide lower bounds