NLP: Introduction and Basic Terminologies

from Chapter 2 of *Convex Optimization* by S. Boyd and L. Vandenberghe

Optimization Lab.

IE department
Seoul National University

7th September 2009

Optimization

$$\begin{array}{lll} \min & f_0(x) & \text{"Objective"} \\ \text{sub. to} & f_i(x) \leq b_i \quad i = 1, \ldots, m, & \text{"Constraints"} \end{array} \qquad (1)$$

where $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ for $i = 0, 1, \ldots, m$.

Definition

A solution x is called feasible if it satisfies all constraints. A feasible solution x^* is called optimal if its objective value is minimum: $f_0(x^*) \leq f_0(x)$ for all feasible x.

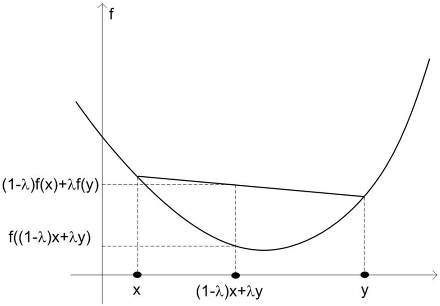
- Tractability of (1), namely possibility of an *efficient* solution method for (1), depends on characteristics of f_i's.
- In general, easy to devise a problem whose only feasibility problem is believed to have no efficient method.

Convex optimization

Definition

Optimization (1) is convex if f_i's are all convex: domains are convex and $\forall\ x, y\ \forall\ 0 \le \lambda \le 1$, *we have*

$$f_i\left((1-\lambda)x + \lambda y\right) \le (1-\lambda) f_i(x) + \lambda f_i(y), \qquad (2)$$



Convex optimization. Why care?

- Easy!

  “In fact the great watershed in optimization isn’t between linearity and nonlinearity, but convexity and nonconvexity.” - Rockafellar

  We can find global optima in polynomial time of input sizes of problem size and numerical accuracy (modulo some technical conditions).

- Prevalent!

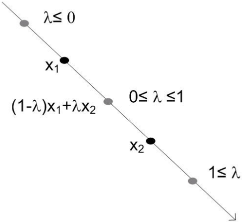
  We are still discovering new applications that can be formulated as a convex optimization problem.

The goal

- To develop the skills and background needed to recognize, formulate, and solve nonlinear programs.
  1. Overview of key properties of nonlinear programs, especially of convex programs.
  2. To understand principles of nonlinear algorithms based on the key properties.

Lines and affine sets

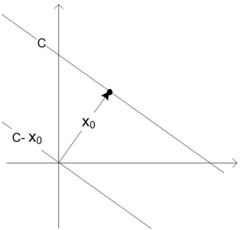
- A line through x_1 and x_2 is the set of points that $y = (1-\lambda)x_1 + \lambda x_2 = x_1 + \lambda(x_2 - x_1)$ where $\lambda \in \mathbb{R}$. Thus, y is the sum of the base point x_2 and the direction $x_1 - x_2$ scaled by λ.



- A set C is called *affine* if it contains the line through any two distinct points of itself: for any x_1 , $x_2 \in C$, with $x_1 \neq x_2$, and $\lambda \in \mathbb{R}$, we have $\lambda x_1 + (1-\lambda)x_2 \in C$.

Lines and affine sets(*cont'd*)

- If C is affine, then for *any* $x_0 \in C$, $V = C - x_0 = \{x - x_0 | x \in C\}$ is a subspace as closed for scalar multiplication and addition: $\forall\ x, y \in C$ and $\forall \lambda \in \mathbb{R}$, $\lambda(x - x_0) = \lambda x + (1 - \lambda)x_0 - x_0 \in C - x_0$, and $x - x_0 + y - x_0$ $= 2(\frac{1}{2}x + \frac{1}{2}y - x_0) \in C - x_0$. Thus, an affine set C is a translation of a subspace V, $C = V + x_0$. The *dimension* of C, dim C is defined as the dimension of V.



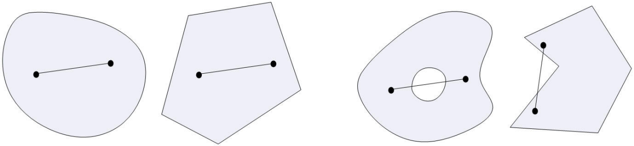
Lines and affine sets(*cont'd*)

- An affine combination of the points $x_1, \ldots, x_k$ is a point of the form: $\lambda_1 x_1 + \cdots + \lambda_k x_k$ with $\sum_{i=1}^{k} \lambda_i = 1$.
- The affine hull of a set C, denoted by aff C, is defined to be the set of all affine combinations of points in C. Thus, aff C is the smallest affine set that contains C. (Why?)
- The *affine dimension* of a set C is defined as dim(aff C). e.g. $C = \{x \in \mathbb{R}^2 | x_1^2 + x_2^2 = 1\} \Rightarrow \dim(\text{aff } C) = 2$, while the "dimension" of C is < 2 in usual senses.
- Relative interior of set C, relint C:

$$x \in \text{relint } C \;\Leftrightarrow\; \exists \delta > 0 \; s.t. \; B(x, \delta) \cap \text{aff } C \subseteq C.$$

Line segments and convex sets

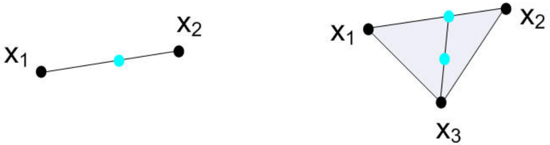
- The line segment between x_1 and x_2 is the set of points $y = (1-\lambda)x_1 + \lambda x_2 = x_1 + \lambda(x_2 - x_1)$ where $0 \le \lambda \le 1$.
- A set is called *convex* if it contains the line segment between any two points in the set: for any $x_1, x_2 \in C$, and for any $0 \le \lambda \le 1$, we have $(1-\lambda)x_1 + \lambda x_2 \in C$.



Line segments and convex sets

- A *convex combination* of $x_1, \dots, x_k$ is a point of the form: $\lambda_1 x_1 + \cdots + \lambda_k x_k$ with $\sum_{i=1}^k \lambda_i = 1, \forall \lambda_i \geq 0$.

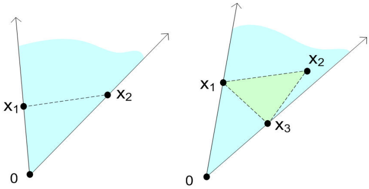
$$\lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3 = (\lambda_1 + \lambda_2)\frac{\lambda_1 x_1 + \lambda_2 x_2}{\lambda_1 + \lambda_2} + \lambda_3 x_3.$$



- The *convex hull* of a set C, denoted by conv$C$, is defined to be the set of convex combinations of points from C. Thus, conv$C$ is the smallest convex set that contains C.
- The *dimension* of a convex set is defined to be its affine dimension.

Cones

- A *cone* is a set closed under positive scalar multiplication: $\forall x \in C$ and $\forall \lambda \geq 0$, $\lambda x \in C$.
- A *convex cone* is a set which is convex as well as a cone.
- A *conic combination* of points $x_1, \ldots, x_k$ is a point of the form: $\lambda_1 x_1 + \cdots + \lambda_k x_k$ with $\lambda_i \geq 0\ \forall i$.
- The *conic hull* of a set C, $\mathrm{cone}C$ is defined to be the set of conic combinations of points from C.



Thus, $\mathrm{cone}C$ is the smallest convex cone containing C.

- We say a set S is a is *finitely generated* cone if $S = \mathrm{cone}C$ with $|C| < \infty$.

Examples

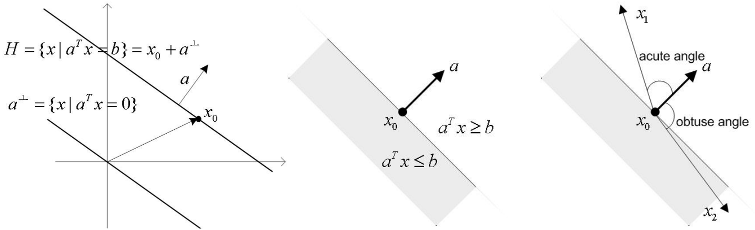
- The empty set $\emptyset$, any single point $\{x_0\}$, and $\mathbb{R}^n$ are affine (hence convex).
- Any subspace is affine, and a convex cone.
- Any line is affine. If it passes through zero, then it is a subspace (hence a convex cone).
- A ray, $\{x_0 + \lambda v | \lambda \geq 0\}$, where $v \neq 0$, is convex, but not affine.

Hyperplanes and halfspaces

Definition

A hyperplane is a set $H = \{x : a^T x = b\}$, where $a \in \mathbb{R}^n$, $a \neq 0$, and $b \in \mathbb{R}$.

For any $x_0 \in H$, $H = \{x | a^T(x - x_0) = 0\} = x_0 + a^{\perp}$, where $a^{\perp} = \{x | a^T x = 0\}$.
A (closed) halfspace is a set of the form $\{x : a^T x \leq b\}$, where $a \neq 0$.



Polyhedra

Definition

A polyhedron P is the intersection of a finite number of halfspaces:

$$P = \{x | a_i^T x \leq b_i,\ i = 1, ..., m\} = \{x | Ax \leq b\}.$$

Definition

A simplex C is the convex hull of a set of affinely indep vectors:

$$C = conv\{v_0, ..., v_k\} = \{\lambda_0 v_0 + ... + \lambda_k v_k | \lambda \geq 0, \mathbf{1}^T \lambda = 1\}$$

where $v_0, ..., v_k \in \mathbb{R}^n$ *are affinely independent.*

Remark

For any $x_1, x_2, \ldots, x_k \in \mathbb{R}^n$, $x_2 - x_1, x_3 - x_1, \ldots, x_n - x_1$ *are linearly independent* $\Leftrightarrow$ $x_1 - x_2, x_3 - x_2, \ldots, x_n - x_2$ *are linearly independent* $\Leftrightarrow \cdots \Leftrightarrow$ $x_1 - x_n, x_2 - x_n, \ldots, x_{n-1} - x_n$ *are linearly independent. In this case, we say* x_1, $x_2, \ldots, x_n$ *are affinely independent.*

Euclidean balls and ellipsoids

Definition

A (Euclidean) ball in $\mathbb{R}^n$ is

$$B(x_c, r) = \{x : \|x - x_c\|_2 \leq r\} = \{x_c + ru : \|u\|_2 \leq 1\}$$

where $r > 0$, and $\|\cdot\|_2$ denotes the Euclidean norm. The vector x_c is the center of the ball and the scalar r is its radius.

Definition

An ellipsoid is

$$E = \{x : (x - x_c)^T P^{-1} (x - x_c) \leq 1\}$$

where $P = P^T \succ 0$. (Notice it is a ball with radius r when $P = r^2 I$.)

The length of axes are $\sqrt{\lambda_i}$ where λ_i are eigenvalues of P. The triangle property of the norm implies the convexity of an ellipsoid.

Euclidean balls and ellipsoids

Theorem

An ellipsoid can be represented as

$$\{Au + b : \|u\|_2 \leq 1\}$$

where A is nonsingular.

Proof. For $u \in B(0,1)$, let $x = Au + b$ or $u = A^{-1}(x-b)$. Then,

$$u^T I u \leq 1 \Leftrightarrow (x-b)^T(A^{-1})^T A^{-1}(x-b) = (x-b)^T(A^{-1})^T A^{-1}(x-b) \leq 1$$

By denoting $(A^{-1})^T A^{-1}$, symmetric and positive-definite, by P^{-1}, we get

$$(x-b)^T P^{-1}(x-b) \leq 1. \quad \square$$

The positive semidefinite cones

- We say a symmetric matrix $M \in \mathbb{S}^n$ is *positive semidefinite*, or PSD if

$$x^T M x \geq 0, \ \forall x \in \mathbb{R}^n.$$

- The set of PSD matrices $\mathbb{S}^n_+$ is a convex cone: for nonnegative $\alpha, \beta \in \mathbb{R}$ and $M, N \in \mathbb{S}^n_+$,

$$x^T(\alpha M + \beta N)x = \alpha x^T M x + \beta x^T N x \geq 0.$$

- Positive semidefinite cones in $\mathbb{S}^2$

$$X = \begin{bmatrix} x & y \\ y & z \end{bmatrix} \Longleftrightarrow x \geq 0, z \geq 0, xz \geq y^2.$$

Convexity is preserved under intersection

Namely, S_α is convex for $\alpha \in A \Rightarrow \cap_{\alpha \in A} S_\alpha$ is convex.

Example

- $\mathbb{S}^n_+ = \cap_{z \neq 0} \{X \in \mathbb{S}^n : z^T X z \geq 0\}$ *where* $\{X \in \mathbb{S}^n | z^T X z \geq 0\}$ *is a linear function of* X*. Hence* $\mathbb{S}^n_+$ *is convex.*
- *Let* $p(t) = \sum_{k=1}^m x_k \cos kt$*. Then,*

$$S = \{x \in \mathbb{R}^m : |p(t)| \leq 1 \text{ for } |t| \leq \pi/3\}$$

can be expressed as $S = \bigcap_{|t| \leq \pi/3} S_t$ *where*

$$S_t = \{x | -1 \leq (\cos t, ..., \cos mt)^T x \leq 1\}.$$

Affine transformations

Definition

An affine transformation $f : \mathbb{R}^n \to \mathbb{R}^m$ *is a translated linear transformation, i.e., has the form* $f(x) = Ax + b$*, where* $A \in \mathbb{R}^{m \times n}$ *and* $b \in \mathbb{R}^m$*.*

The image and inverse image of a convex set under an affine transformation f is convex:

- $S \subseteq \mathbb{R}^n$ *convex* $\Rightarrow f(S) = \{f(x) | x \in S\}$ is convex.
- $C \subseteq \mathbb{R}^m$ *convex* $\Rightarrow f^{-1}(C) = \{x \in \mathbb{R}^n | f(x) \in C\}$ is convex.

Affine transformations(*cont'd*)

Example

1. *Scaling and translation preserve convexity.*
2. *So does a projection:* $[I_m \vdots O_n] \begin{bmatrix} x \\ y \end{bmatrix} = x$, *where* $x \in \mathbb{R}^m, y \in \mathbb{R}^n$
3. S_1, S_2 *convex* $\Rightarrow$ *So are their sum* $S_1 + S_2$, *product* $S_1 \times S_2 := \{(x_1, x_2) | x_1 \in S_1, x_2 \in S_2\}$, *and*
4. *partial sum,* $S := \{(x, y_1 + y_2) | (x, y_1) \in S_1, (x, y_2) \in S_2\}$.

Affine transformations(*cont'd*)

Example

$$\begin{aligned} Polyhedron &= \{x \mid Ax \leq b, Cx = d\} \\ &= f^{-1}(\mathbb{R}_+^m \times \{0\}), \end{aligned}$$

where $f(x) = (b - Ax, d - Cx)$.

Example

$$\begin{aligned} Ellipsoid &= \{x \mid (x - x_c)^T P^{-1}(x - x_c) \leq 1\}, P \in \mathbb{S}_{++}^n \\ &= \text{image of } \{u \mid \|u\|_2 \leq 1\} \\ &\quad \text{under affine transform } f(u) = P^{1/2}u + x_c \\ &= \text{inverse image of unit ball} \\ &\quad \text{under affine transform } g(x) = P^{-1/2}(x - x_c). \end{aligned}$$

Linear-fractional and perspective functions

Definition

A perspective function $P : \mathbb{R}^{n+1} \to \mathbb{R}^n$, *with dom* $P = \mathbb{R}^n \times \mathbb{R}_{++}$ *is defined as*

$$P(z, t) = z/t.$$

- If $C \subseteq \mathrm{dom} P$ is convex, $P(C)$ is convex.
- If $C \subseteq \mathbb{R}^n$ is convex. $P^{-1}(C)$ is convex.

Linear-fractional and perspective functions(*cont'd*)

Definition

Let $g : \mathbb{R}^n \to \mathbb{R}^{m+1}$ *be an affine transformation* $g(x) = \begin{bmatrix} A \\ c^T \end{bmatrix} x + \begin{bmatrix} b \\ d \end{bmatrix}$.
The function $f : \mathbb{R}^n \to \mathbb{R}^m$ *given by* $f = P \circ g$, *i.e.,*

$$f(x) = (Ax + b)/(c^T x + d), \quad domf = \{x : c^T x + d > 0\},$$

is called a linear-fractional (or projective) function.

- Both image and inverse image of a convex set under linear-fractional are convex.

$$x \xrightarrow{\text{affine transform}} \begin{bmatrix} A \\ c^T \end{bmatrix} x + \begin{bmatrix} b \\ d \end{bmatrix} \xrightarrow{\text{perspective}} \frac{Ax+b}{c^T x+d}.$$

Minimum and minimal elements

In general, $\leq$ is not a linear ordering: there can be x and y such that $x \nleq y$ and $y \ngeq x$.

Definition

A point $x \in S$ is the minimum element of S if $x \leq y, \forall y \in S$. Equivalently, $x \in S$ is the minimum element iff $S \subseteq x + \mathbb{R}^n_+$. A point $x \in S$ is a minimal element of S if $y \in S, y \leq x \Rightarrow y = x$. Equivalently, $x \in S$ is a minimal element iff $(x - \mathbb{R}^n_+) \cap S = \{x\}$.

Separating hyperplane theorem

Theorem

Suppose C and D are disjoint convex sets. Then, $\exists a \neq 0,\ b$ s.t.

$$a^T x \leq b, \forall x \in C, a^T x \geq b, \forall x \in D.$$

Then $\{x|a^T x = b\}$ is called a separating hyperplane for C and D.
In some cases, a *strict* separation can be established: Let C be a closed convex set and $x_0 \notin C$. Then there exists a hyperplane $\{x|a^T x = b\}$ that strictly separates x_0 from C, namely, $a^T x \leq b$ for every $x \in C$ and $a^T x_0 > b$.

Supporting hyperplanes

Definition

Suppose $x_0 \in C \subseteq \mathbb{R}^n$. If $a \neq 0$ satisfies $a^T x \leq a^T x_0, \forall x \in C$, then the hyperplane $\{x | a^T x = a^T x_0\}$ is called a supporting hyperplane to C at $x = x_0$.

Separating hyperplane theorem immediately implies the following.

Theorem

For any nonempty convex set C and any $x_0 \in \text{bd } C$, there exists a supporting hyperplane to C at x_0.

Homework

2.1, 2.2, 2.9, 2.10, 2.12, 2.19, 2.28