Convex Optimization notes

Abiyaz Chowdhury
StonyBrook ID: 111580554

February 15, 2018

1 Introduction

Definition 1.1 A mathematical optimization problem has the form:
Minimize $f_0(x)$ subject to $f_i(x) \leq b_i \ \forall i \in \{1, 2, ...m\}$ where $x \in \mathbb{R}^n$ is the optimization variable and $f_0 : \mathbb{R}^n \to \mathbb{R}$ is the objective function. The functions $f_i : \mathbb{R}^n \to \mathbb{R}$ are the constraint functions. A vector $x^\star \in \mathbb{R}^n$ is optimal (or a solution) if $f_0(x^\star) \leq f_0(x) \ \forall x \in \mathbb{R}^n$ and also $f_i(x^\star) \leq b_i \ \forall i \in \{1, 2, ...m\}$.

Definition 1.2 An optimization problem in which the objective and the constraint functions $f_0, ...f_m$ are all linear (i.e. $f_i(\alpha x + \beta y) = \alpha f_i(x) + \beta f_i(y) \ \forall x, y \in \mathbb{R}^n, \ \forall \alpha, \beta \in \mathbb{R}$) is called a linear program.

Definition 1.3 A convex optimization problem is one in which the objective and constraint functions are convex, which means they satisfy the inequality $f_i(\alpha x + \beta y) \leq \alpha f_i(x) + \beta f_i(y) \ \forall x, y \in \mathbb{R}^n, \ \forall \alpha, \beta \in \mathbb{R}$ with $\alpha + \beta = 1, \alpha \geq 0, \beta \geq 0$.

Theorem 1.1 Every linear program is also a convex optimization problem.

Definition 1.4 A least-squares problem is an optimization problem with no constraints and an objective function which is of the form:
Minimize $f_0(x) = ||Ax - b||_2^2$ where $A \in \mathbb{R}^{k \times n}, k \geq n$ and $x \in \mathbb{R}^n$ is the optimization variable.

Theorem 1.2 The solution of the least squares problem is $x = (A^T A)^{-1} A^T b$

Theorem 1.3 Every least squares problem is also a convex optimization problem.

Theorem 1.4 Every linear program is of the form:
Minimize $c^T x$ subject to $a_i^T(x) \leq b_i, \ \forall i \in \{1, ...m\}$ where $c, a_1...a_m \in \mathbb{R}^n$ and $b_1...b_m \in \mathbb{R}$.

Definition 1.5 The Chebyshev approximation problem is to:
Minimize $\max_{i=1,...,k} |a_i^T x - b_i|$ where $x \in \mathbb{R}^n$ is the optimization variable, and $a_1, ..., a_k \in \mathbb{R}^n, b_1...b_k \in \mathbb{R}$.

Theorem 1.5 The Chebyshev approximation problem can be transformed to an equivalent linear program of the form:
Minimize t subject to $a_i^T x - t \leq b_i, \ \forall i \in \{1, ..., k\}$ and $-a_i^T x - t \leq -b_i, \ \forall i \in \{1, ..., k\}$ for $x \in \mathbb{R}^n$ and $t \in \mathbb{R}$.

2 Convex sets

Definition 2.1 A set $C \subseteq \mathbb{R}^n$ is affine if the line through any two distinct points in C lies in C, i.e. $\forall x_1, x_2 \in C$ and $\theta \in \mathbb{R}$, we have $\theta x_1 + (1 - \theta) x_2 \in C$.

Theorem 2.1 A set C is affine if and only if C contains the linear combination of two of the points in C whose coefficients sum to 1.

Definition 2.2 An affine combination of a set of points $x_1...x_k \in \mathbb{R}^n$ is of the form $\theta_1 x_1 + ... + \theta_k x_k$ where $\theta_1 + ... + \theta_k = 1$.

Theorem 2.2 A set C is affine if and only if C contains all affine combinations of any set of its points in C.

Theorem 2.3 If a set C is affine, and $x_0 \in C$, then the set $V = \{x - x_0 \mid x \in C\}$ is a subspace.

Theorem 2.4 If a set C is affine, and $x_0 \in C$, then the set $C = \{v + x_0 \mid v \in V\}$ for some subspace V.

Theorem 2.5 An affine set C is also a vector space if and only if it contains the zero vector.

Definition 2.3 The dimension of an affine set C is the dimension of a subspace V where $C = V + x_0$ where $x_0 \in C$.

Theorem 2.6 The dimension of an affine set C is unique.

Theorem 2.7 If $Ax = b$ is a system of linear equations where $A \in \mathbb{R}^{m \times n}, x \in \mathbb{R}^n, b \in \mathbb{R}^m$, then the solution set to this system is affine (provided a solution exists).

Definition 2.4 The affine hull of a set C is the set of all affine combinations of points in C.
aff $C = \{\theta_1 x_1 + ... + \theta_k x_k \mid x_1 ... x_k \in C, \theta_1 + ... + \theta_k = 1\}$

Theorem 2.8 The intersection of two or more affine sets is affine.

Theorem 2.9 Given a set C, the intersection of all affine sets containing C is the affine hull of C.

Theorem 2.10 (?) A set is affine if and only if it equals its affine hull.

Theorem 2.11 The affine hull of any set is affine.

Theorem 2.12 Let $x_1, x_2...$ be a sequence in C and let $\theta_1, \theta_2...$ be a sequence where each $\theta_i \in \mathbb{R}$, and $\sum_{i=1}^{\infty} \theta_i = 1$. If C is affine, then $\sum_{i=1}^{\infty} \theta_i x_i \in C$ if the series converges.

Theorem 2.13 Let $p : \mathbb{R}^n \to \mathbb{R}$ satisfy $\int_C p(x)\ dx = 1$, where $C \subseteq \mathbb{R}^n$ is affine. Then $\int_C p(x)x\ dx \in C$ if the integral exists.

Definition 2.5 The affine dimension of a set C is the dimension of its affine hull.

Example 2.1 The affine dimension of the unit circle in $\mathbb{R}^2$, i.e. $\{(x_1, x_2) \mid x_1^2 + x_2^2 = 1\}$ is 2.

Theorem 2.14 The affine dimension of a set $C \subseteq \mathbb{R}^n$ is n if and only if **aff** $C = \mathbb{R}^n$

Definition 2.6 The relative interior of the set C, denoted **relint** C is its interior relative to **aff** C:
relint $C = \{x \in C \mid B(x, r) \cap$ **aff** $C \subseteq C$ for some $r > 0\}$ where $B(x, r) = \{y \mid ||y - x|| \leq r\}$ where $|| \cdot ||$ is any norm.

Definition 2.7 The relative boundary of a set C is **cl** $C \setminus$ **relint** C where **cl** C is the closure of C.

Example 2.2 Consider the unit square in the xy-plane in $\mathbb{R}^3$, i.e. $C = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1 \in [-1, 1], x_2 \in [-1, 1], x_3 = 0\}$. Its affine hull is the xy-plane. The interior of C is empty, but its relative interior is **relint** $C = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1 \in (-1, 1), x_2 \in (-1, 1), x_3 = 0\}$. Its boundary is itself, and its relative boundary is $\{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid \max\{|x_1|, |x_2|\} = 1, x_3 = 0\}$.

Definition 2.8 A set $C \subseteq \mathbb{R}^n$ is convex if the line segment through any two distinct points in C lies in C, i.e. $\forall x_1, x_2 \in C$ and $\theta \in [0, 1]$, we have $\theta x_1 + (1 - \theta)x_2 \in C$.

Theorem 2.15 Every affine set is also convex.

Theorem 2.16 A set C is convex if and only if C contains the convex combination of any two points in C whose coefficients sum to 1.

Definition 2.9 A convex combination of a set of points $x_1...x_k \in \mathbb{R}^n$ is of the form $\theta_1 x_1 + ... + \theta_k x_k$ where $\theta_1 + ... + \theta_k = 1$ and each of the $\theta_1...\theta_k$ are all in $[0, 1]$.

Theorem 2.17 A set C is convex if and only if C contains all convex combinations of any set of its points in C.

Definition 2.10 The affine hull of a set C is the set of all affine combinations of points in C.
aff $C = \{\theta_1 x_1 + ... + \theta_k x_k \mid x_1...x_k \in C, \theta_1 + ... + \theta_k = 1\}$

Theorem 2.18 The intersection of two or more convex sets is convex.

Theorem 2.19 Given a set C, the intersection of all convex sets containing C is the convex hull of C.

Theorem 2.20 (?) A set is convex if and only if it equals its convex hull.

Theorem 2.21 The convex hull of any set is convex.

Theorem 2.22 Let $x_1, x_2...$ be a sequence in C and let $\theta_1, \theta_2...$ be a sequence where each $\theta_i \in [0, 1]$, and $\sum_{i=1}^{\infty} \theta_i = 1$. If C is convex, then $\sum_{i=1}^{\infty} \theta_i x_i \in C$ if the series converges.

Theorem 2.23 Let $p : \mathbb{R}^n \to \mathbb{R}$ satisfy $p(x) \geq 0$ for all $x \in C$ and $\int_C p(x)\ dx = 1$, where $C \subseteq \mathbb{R}^n$ is convex. Then $\int_C p(x)x\ dx \in C$ if the integral exists.

Theorem 2.24 If $C \subseteq \mathbb{R}^n$ is convex and x is a random vector with $p(x) = 1$, then $\mathbf{E}[x] \in C$.

Definition 2.11 A set C is a cone, or nonnegative homogenous, if for all $x \in C$ and $\theta \geq 0$ we have $\theta x \in C$. C is a convex cone, if it is convex and a cone.

Theorem 2.25 C is a convex cone if and only if for all $x_1, x_2 \in C$, and $\theta_1, \theta_2 \geq 0$, we have $\theta_1 x_1 + \theta_2 x_2 \in C$.

Definition 2.12 A conic combination of a set of points $x_1, ..., x_k \in \mathbb{R}^n$ is of the form $\theta_1 x_1 + ... + \theta_k x_k$ where each of the $\theta_i \geq 0$.

Theorem 2.26 A set C is a convex cone if and only if C contains all conic combinations of its elements.

Theorem 2.27 Let $x_1, x_2...$ be a sequence in C and let $\theta_1, \theta_2...$ be a sequence where each $\theta_i \geq 0$. If C is a convex cone, then $\sum_{i=1}^{\infty} \theta_i x_i \in C$ if the series converges.

Theorem 2.28 Let $p : \mathbb{R}^n \to \mathbb{R}$ satisfy $p(x) \geq 0$, where $C \subseteq \mathbb{R}^n$ is a convex cone. Then $\int_C p(x)x\ dx \in C$ if the integral exists.

Definition 2.13 The conic hull of a set C is the set of all conic combinations of points in C.
$\{\theta_1 x_1 + ... + \theta_k x_k \mid x_1...x_k \in C, \theta_i \geq 0\}$

Theorem 2.29 (?) The intersection of two or more convex cones is a convex cone.

Theorem 2.30 Given a set C, the intersection of all conic cones containing C is the conic hull of C.

Theorem 2.31 (?) A set is a conic cone if and only if it equals its conic hull.

Theorem 2.32 The conic hull of any set is a convex set.

Example 2.3 The empty set $\varnothing$, any singleton in $\mathbb{R}^n$, and $\mathbb{R}^n$ itself are all affine (hence, convex) subsets of $\mathbb{R}^n$.

Example 2.4 Any line is affine. If it passes through zero, it is a subspace, and hence also a convex cone.

Example 2.5 The line segment connecting two distinct points is convex, but not affine.

Example 2.6 The ray $\{(x_0 + \theta v \mid \theta \geq 0\}$ where $v \neq 0$ is convex, but not affine. It is a convex cone if its base $x_0 = 0$.

Example 2.7 Any subspace is affine, and also a convex cone.

Definition 2.14 A hyperplane is a set of the form $\{x \mid a^T x = b\}$ where $a \in \mathbb{R}^n, a \neq 0, b \in \mathbb{R}$.

Theorem 2.33 The above hyperplane definition can be equivalently stated as $\{x \mid a^T(x - x_0) = 0\} = x_0 + a^\perp$ where $a^\perp$ is the orthogonal complement of a.

Definition 2.15 A halfspace of $\mathbb{R}^n$ is of the form $\{x \mid a^T x \leq b\}$ where $a \neq$.

Theorem 2.34 The above halfspace definition can be equivalently stated as $\{x \mid a^T(x-x_0) leq 0\}$ where x_0 is a point on the hyperplane $a^T x = b$.

Theorem 2.35 Halfspaces are convex, but not affine.

Theorem 2.36 A hyperplane divides $\mathbb{R}^n$ into two disjoint halfspaces. Vectors fall into one or the other of the two halfspaces depending on whether the angle they make with a normal vector to the halfspace is acute or obtuse. Vectors normal to a normal lie in the hyperplane.

Theorem 2.37 The boundary of a halfspace is the associated hyperplane.

Definition 2.16 A Euclidean ball in $\mathbb{R}^n$ has the form $B(x_c, r) = \{x \mid ||x - x_c||_2 \leq r\}$ where $r > 0$ and $|| \cdot ||$ denotes the Euclidean norm.

Theorem 2.38 An equivalent definition of a Euclidean ball is $B(x_c, r) = \{x_c + ru \mid ||u||_2 \leq 1\}$.

Theorem 2.39 Euclidean balls are convex. In fact, balls are convex with respect to any norm.

Definition 2.17 An ellipsoid is a set of the form $\mathcal{E} = \{x \mid (x - x_c)^T P^{-1}(x - x_c) \leq 1\}$ where $P = P^T \succ 0$ is symmetric and positive definite and x_c is the center of the ellipse.

Theorem 2.40 The lengths of the semi-axes of $\mathcal{E}$ are given by $\sqrt{\lambda_i}$ where λ_i are the eigenvalues of P.

Theorem 2.41 A ball is an ellipsoid with $P = r^2 I$.

Theorem 2.42 An equivalent definition of an ellipsoid is $\mathcal{E} = \{x_c + Au \mid ||u||_2 \leq 1\}$ where A is square and nonsingular. (By taking $A = P^{1/2}$, we get the ellipsoid defined earlier.

Theorem 2.43 When the matrix in the above definition is symmetric positive definite but singular, the set in the definition above is called a degenerate ellipsoid and its affine dimension is equal to the rank of A.

Theorem 2.44 Ellipsoids (including degenerate ones) are convex.

Definition 2.18 The norm cone associated with the norm $|| \cdot ||$ is the set $C = \{(x, t) \mid ||x|| \leq t\} \subseteq \mathbb{R}^{n+1}$.

Theorem 2.45 Norm cones are convex cones.

Example 2.8 The second-order cone, defined by $C = \{(x, t) \in \mathbb{R}^{n+1} \mid ||x||_2 \leq t\}$ is the norm cone for the Euclidean norm. It is also known as the quadratic cone, the Lorentz cone, or the ice-cream cone.

Definition 2.19 A polyhedron is the solution set of a finite number of linear equalities and inequalities. $\mathcal{P} = \{x \mid a_j^T x \leq b_j, j = 1, ..., m, c_j^T x = d_j, j = 1, ..., p\}$. Another notation is $\mathcal{P} = \{x \mid Ax \succ b, Cx \preceq d\}$. where $\preceq$ denotes componentwise vector inequality.

Theorem 2.46 A polyhedron is the intersection of a finite number of halfspaces and hyperplanes.

Theorem 2.47 Affine sets (subspaces, hyperplanes, lines), rays, line segments, and halfspaces are all polyhedra.

Theorem 2.48 Polyhedra are convex sets.

Example 2.9 The nonnegative orthant is the set of points with nonnegative components, $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n \mid x_i \geq 0, i = 1, ..., n\}$. The nonnegative orthant is both a polyhedron and a cone.

Definition 2.20 Suppose the $k+1$ points $v_0, ..., v_k \in \mathbb{R}^n$ are affinely independent, which means $v_1 - v_0, ..., v_k - v_0$ are linearly independent. The simplex determined by them is $C = \mathbf{conv}\{v_0, ..., v_k\} = \{\theta_0 v_0 + ... + \theta_k v_k \mid \theta \succeq 0, \mathbf{1}^T = 1\}$

Theorem 2.49 Simplices are also polyhedra and a simplex from $k+1$ points has affine dimension k.

Example 2.10 A 1D simplex is a line segment, a 2D simplex is a triangle, and a 3D simplex is a tetrahedron.

Theorem 2.50 The convex hull of a finite set is a polyhedron, and bounded, but not necessarily represented easily as polyhedra. Likewise, polyhedra are not necessarily easily represented as convex hulls of a set of points.

Example 2.11 The unit ball in the ℓ_∞-norm in $\mathbb{R}^n$ is $C = \{x \mid |x_i| \leq 1, i = 1, ..., n\}$. This set can be described in polyhedron form with $2n$ inequalities, or in convex hull form with 2^n points.

Theorem 2.51 The set of symmetric $n \times n$ matrices, $S^n = \{X \in \mathbb{R}^{n \times n} \mid X = X^T\}$ is a subspace with dimension $n(n+1)/2$.

Definition 2.21 $S^n_+ = \{X \in S^n \mid X \succeq 0\}$ denotes the set of symmetric positive semidefinite matrices. $S^n_{++} = \{X \in S^n \mid X \succ 0\}$ denotes the set of symmetric positive definite matrices.

Theorem 2.52 The set S^n_+ is a convex cone.

Theorem 2.53 Affine transformations preserve convexity.

Theorem 2.54 Inverse images of affine transformations preserve convexity.

Theorem 2.55 Projections preserve convexity.

Theorem 2.56 The sum of two convex sets is convex (the sum is denoted by $S_1 + S_2 = \{x_1 + x_2 \mid x_1 \in S_1, x_2 \in S_2\}$).

Theorem 2.57 The partial sum of two convex sets is convex (with respect to one of the sets) is convex (the partial sum of two vectors adds only some of the components.)

Example 2.12 The polyhedron $\{x \mid Ax \preceq b, Cx = d\}$ can be expressed as the inverse image of the Cartesian product of the nonnegative orthant and the origin under the affine function $f(x) = (b - Ax, d - Cx)$: $\{x \mid f(x) \in \mathbb{R}^m_+ \times \{0\}\}$.

3 Convex functions

Definition 3.1 A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if $\mathbf{dom}(f)$ is a convex set and if for all $x, y \in \mathbf{dom}(f)$ with $\theta \in [0, 1]$, we have $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$.

Definition 3.2 A function $f : \mathbb{R}^n \to \mathbb{R}$ is strictly convex if $\mathbf{dom}(f)$ is a convex set and if for all $x, y \in \mathbf{dom}(f)$ with $x \neq y$ and $\theta \in (0, 1)$, we have $f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta) f(y)$.

Definition 3.3 f is concave if $-f$ is convex and f is strictly concave if $-f$ is strictly convex.

Theorem 3.1 No function can be both strictly concave strictly convex, but both convexity and concavity is possible. In fact, any function is affine if and only if it is both convex and concave.

Theorem 3.2 A function is convex if and only if it is convex when restricted to any line that intersects its domain.