Convex Optimization

Varun Varanasi

October 12, 2023

Contents

1 Introduction

1.1 Mathematical Optimization

A generally mathematical optimization problem has the form

$$\begin{aligned} &\text{minimize} \quad f_0(x) \\ &\text{s.t.} \quad f_i(x) \leq b_i \qquad i = 1, ...m \end{aligned}$$

with the following features:

- Optimization Variable: $x \in \mathbb{R}^n$
- Objective Function: $f_0 : \mathbb{R}^n \to \mathbb{R}$
- Constraint Functions: $f_i : \mathbb{R}^n \to \mathbb{R}$
- Constraint Limits: b_i

An optimal solution x^* is one such that it obtains the smallest realization of the objective function for all vectors that satisfy the constraints. Problems are classified according to the structure of the objective function and constraints. Convex functions obey the following property:

$$f_i(\alpha x + \beta y) \leq \alpha f_i(x) + \beta f_i(y) \qquad \forall x, y \in \mathbb{R}^n \quad \alpha, \beta \in \mathbb{R}, \quad \alpha + \beta = 1$$

Since convexity is a more general statement then linearity, any linear program is a convex optimization problem.

1.2 Least-squares and Linear Programming

1.2.1 Least-squares problems

Least squares problems have no constraints and are optimizing a sum of squares.

$$\text{minimize} \quad f_0(x) = ||Ax - b||_2^2 = \sum_{i=1}^{k} (a_i^T x - b_i)^2$$

where $A \in \mathbb{R}^{k \times n}, x \in \mathbb{R}^n$.

The solution to the least squares problem is a solution to a series of linear equations:

$$\begin{aligned} \left(AA^T\right) x = A^T b \\ x = (AA^T)^{-1} A^T b \end{aligned}$$

Computational solutions for these classes of problems are quite fast and accurate. Recognizing a problems is least squares involves showing that the objective function is quadratic and the quadratic form is positive semi-definite.

1.2.2 Linear Programming

Linear programming refers to the set of problems with a linear objective function and linear constraints.

$$\begin{aligned} &\text{minimize} \quad c^T(x) \\ &\text{s.t.} \quad a_i^T x \leq b_i \qquad i = 1, ...m \end{aligned}$$

where $c, a_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$.

Linear programming problems do not have simple analytical solutions, but existing algorithms are quite reliable.

1.3 Convex Optimization

Convex optimization problems take the form

$$\begin{aligned} &\text{minimize} \quad f_0(x) \\ &\text{s.t.} \quad f_i(x) \le b_i \qquad i = 1, ...m \end{aligned}$$

where f_0, f_i are all convex.

$$f_i(\alpha x + \beta y) \le \alpha f_i(x) + \beta f_i(y) \qquad \forall x, y \in \mathbb{R}^n \quad \alpha, \beta \in \mathbb{R}, \quad \alpha + \beta = 1$$

Like linear programming, convex problems do not have a general solution. Current solutions involve inner-point methods, but despite their reliability, these solutions are not as fleshed out as their counterparts in linear programming and least-squares. The challenge of convex optimization is recognizing which problems are convex afterwhich the machinary of technology can solve your problem.

1.4 Non-linear Optimization

Non-linear optimization is a general term for problems that do not fall into the earlier classifications. These problems do not have simple solutions. Instead, you compromise on your solution to suite your needs. For example, local optimization finds an optimal point among the feasible points around it. This method is fast and scalable, but comes at the cost finding the actual optimal point and is highly reliant on initial parameters. Global optimization on the other hand sacrifices efficiency to produce an optimal solution.

Convex optimization can assist in solving non-linear problems by using convex optimization to solve for the initial point in the local optimization problem. This method involves solving a simpler convex problem and using that solution as an initializing point for a local optimization solution on the original problem. Convex optimization can also provide bounds for global optimization problems by relaxing constraints or solving dual problems.

2 Convex Sets

2.1 Affine and Convex Sets

2.1.1 Affine Sets

A set C is said to be affine if any line through two distinct points in C lies in C.

$$\theta x_1 + (1 - \theta) x_2 \in C \qquad \forall x_1, x_2 \in C, \quad \theta \in [0, 1]$$

Note that the line refers to the infinite line between points x_1 and x_2, not just the line segment. For multiple points, we can refer to the affince combination $\theta_1 x_1 + ... + \theta_k x_k$ where $\sum \theta_i = 1$. An affine set containts every affine combination of its points.

For an affine set C, consider $x_0 \in C$. The set V defined as

$$V = C - x_0 = \{x - x_0 | x \in C\}$$

is a subspace (closed under addition and scalar multiplication). We can therefore define an affine set as an offset of a subspace:

$$C = V + x_0 = \{v + x_0 | v \in V\}$$

Since V is independent of our choice of x_0, we can think about the dimension of C as the dimension of V where $V = C - x_0$ for aribtrary $x_0 \in C$.

The solution to a system of linear equations is an affine set. Conversely, any affine set can be represented as the solution to a system of linear equations.

For a set $C \subseteq \mathbb{R}^n$, the affine hull of C is affine combination of all points in C.

$$\textbf{Aff } C = \Big\{ \sum_i \theta_i x_i | x_i \in C, \sum \theta_i = 1 \Big\}$$

The affine hull is the smallest affine set that contains C. Affine dimension is given by the dimension of its affine hull.

2.1.2 Convex Sets

A set is convex if the line segment between any two points in C lies in C.

$$\forall x, y \in C, \quad 0 \leq \theta \leq 1 \qquad \theta x + (1-\theta) y \in C$$

Conceptually, you can think of a convex set as the set of points such that each point can see eachother with a path that lies in the set. Like an affine combination, a convex combination of the points $x_1 ... x_k$ is the point $\theta_1 x_1 + ... + \theta_k x_k$ such that $\theta_1 + ... + \theta_k$ where $\theta_i \geq 0$. A set is convex if and only if it contains every convex combination of points in the set. Intuitively, this is akin to considering a weighted average of points in the set where the weights are given by θ_i. This combination can be generalized to infinite points without issue.

A convex hull of a set C is the set of all convex combinations of points in C.

$$\textbf{conv } C = \Big\{ \theta_1 x_1 + ... + \theta_k x_k | x_i \in C, \theta_i \geq 0, \sum \theta_i = 1 \Big\}$$

The convex hull is the smallest convex set that contains C.

2.1.3 Cones

A set C is called a cone or non-negative homogeneous if for every $x \in C$ and $\theta \geq 0$, $\theta x \in C$. Sets can be both convex and a cone in which case we refer to them as convex cones.

$$\forall x_1, x_2 \in C, \quad \theta_1, \theta_2 \geq 0 \qquad \theta_1 x_1 + \theta_2 x_2 \in C$$

As with other sets, we can define a conic combination (nonnegative linear combination) as the combination $\theta_1 x_1 + ... + \theta_k x_k$ where $\theta_1, ..., \theta_k \geq 0$. If a point exist in a convex cone C, then every convex combination of that point also exists in C. Conversely, a set is a a convex cone if and only if it contains all convex combinations of its elements.

A conic hull, the conic combination of all points in C, is the smallest convex cone that contains a set C.

$$\Big\{ \theta_1 x_1 + ... + \theta_k x_k | x_i \in C, \theta_1 \geq 0 \Big\}$$

2.2 Important Examples

- The empty set $\emptyset$
- Singleton $\{x_0\}$
- The whole space $\mathbb{R}^n$
- Any line (affine hence convex) If it passes through 0 it is a subspace and thus a convex cone as well
- Line segement (convex but not affine)
- A ray (convex but not affine) If it's base is 0 then it is a convex cone
- Any subspace is affine and a convex cone

2.2.1 Hyperplanes and Halfspaces

A hyperplane is defined as the set

$$\{x|a^T x = b\}$$

where $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Conceptually, it is equivalent to the solution set of linear equations for each component of x (affine set!). Geometrically, it is the space with a constant inner product to a vector a or a hyperplane with normal vector a and offset b.

Hyperplanes divide the space into two half-spaces. These spaces are convex but not affine.

$$\{x|a^T x \leq b\} \quad \text{or}\{x|a^T(x - x_0) \leq 0, a^T x_0 = b\} \qquad \text{closed halfspace}$$

2.2.2 Euclidean Balls and Ellipsoids

A euclidean ball in $\mathbb{R}^n$ has the form:

$$B(x_c, r) = \{x|||x - x_c||_2 \leq r\} = \{x|(x - x_c)^T(x - x_c) \leq r\}$$

The euclidean ball represents the space of all points with a radius r of the center point x_c. You can use the homogeniety and triangle inquality of norms to show that the euclidean ball is convex.

Ellipsoids are a similar family of sets that take the form:

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

where P is a positive definite matrix and x_c is the center of ellipsoid. The lengths of the semi-axis of $\mathcal{E}$ are given by the eigenvalues of P $\sqrt{\lambda_i}$. An ellipsoid is a generalization of the ball where $P = r^2 I$.

2.2.3 Norm Balls and Norm Cones

For a given norm on $\mathbb{R}^n$, the norm ball defined as $\{x|||x - x_c|| \leq r\}$ is convex. The norm cone, the set $C = \{(x,t)|||x|| \leq t\} \subseteq \mathbb{R}^{n+1}$ is a convex cone.

2.2.4 Polyhedra

Polyhedron are the solution set for a finite set of linear equalities and inequalities.

$$\mathcal{P} = \{x|a_j^T x \leq b_j, c_i^T x = d_i\}$$

Alternatively, a polyhedron can be understood as the intersection of a finite number of halfspaces and hyperplanes. Polyhedra are convex.

For a set of affinely independent (linearly independent with additional condition that the weights add to 1) points, we define a simplex to be the set

$$C = \textbf{conv}\{v_0, ..., v_k\} = \{\theta_0 v_0 + ... + \theta_k v_k | \theta \succeq 0, \mathbf{1}^T \theta = 1\}$$

The affine dimension of this simplex is k and is sometimes referred to as k-dimensional simplex. One dimensional simplex is a line, 2-d simplex is a triangle, 3-d simplex is a tetrahedron. Simplex can be understood as a polyhedron by redefining it as a series of linear equalities and inequalities.

2.2.5 The positive semidefinite cone

Let S^n denote the space of symmetric matricies. Similarly, S^n_+ denotes the set of symmetric positive semidefinite matricies and S^n_{++} is the set of symmetric positive definite matricies. The set S^n_+ is a cone since

$$x^T(\theta_1 A + \theta_2 B)x = \theta_1 x^T A x + \theta_2 x^T B x \geq 0$$

where $A, B \in S^n_+$ and $\theta_1, \theta_2 \geq 0$

2.3 Operations that Preserve Convexity

- Intersections: $S_1 \cap S_2$
- Affine Functions
  - $f(S) = \{f(x) : x \in S\}$ where S is convex and f is affine
  - $f^{-1}(S) = \{x : f(x) \in S\}$ where S is convex and f is affine
- Scaling: $\alpha S = \{\alpha x : x \in S\}$
- Translation: $S + a = \{x + a : x \in S\}$
- Projection: $T = \{x_1 \in \mathbb{R}^m | (x_1, x_2) \in S, x_2\ in \mathbb{R}^n\}$
- Summation: $S_1 + S_2 = \{x + y | x \in S_1, y \in S_2\}$
- Cartesian Product: $S_1 \times S_2 = \{(x_1, x_2) : x_1 \in S_1, x_2 \in S_2\}$
- Partial Sum: $S = \{(x, y_1 + y_2) : (x, y_1) \in S_1, (x, y_2) \in S_2\}$

2.4 Seperating and Supporting Hyperplanes

Theorem 2.1 (Seperating Hyperplane Theorem) *Suppose C and D are nonempty disjoint convex sets. Then there exists $a \neq 0$ and b such that $a^T x \leq b$ for all $x \in C$ and $a^T x \geq b$ for all $x \in D$.*

This hyperplane $\{a^T x = b\}$ is said to be the seperating hyperplane for sets C and D.

Proof in Boyd 2.5

Converse of the seperating hyperplane theorem is not necessarily true.

Definition 1 (Supporting Hyperplane) *Suppose $C \subseteq \mathbb{R}^n$ and x_0 is a point on the boundary of C.*

$$x_0 \in \textbf{bd}\ C = \textbf{cl}\ C \backslash \textbf{int}\ C$$

If $a \neq 0$ satisfies $a^T x \leq a^T x_0$ for all $x \in C$, then the hyperplane $\{x | a^T x = a^T x_0\}$ is called the supporting hyperplane to C at point x_0.

Geometrically, you can interpret the supporting hyperplane as tangent to the set C at point x_0.

Theorem 2.2 (Supporting Hyperplane Theorem) *For any nonempty convex set C and any $x_0 \in \textbf{bd}\ C$, there exists a supporting hyperplane to C at x_0*

If a set is closed, has a nonempty interior, and has a supporting hyperplane at everypoint in its boundary, then it is convex.

3 Convex Functions

3.1 Basic Properties and Examples

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if **dom** f is a convex set and if for all $x, y \in$ **dom** f and $0 \leq \theta \leq 1$,

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$

The geometric interpretation of this definition is that the chord between $(x, f(x))$ and $(y, f(y))$ lies above the curve f. A function is said to be strictly convex is the inequality is strict for all $x \neq y$. A concave function is one where $-f$ is convex. Strictness is given by the same inequality conditions.

By definition affine functions will always meet the equality. Therefore, affine functions are both convex and concave and any function that is both convex and concave is affine.

A function is convex if and only if it is convex when restricted to any line that intersects its domain. This is useful tool in determining the convexity of a function by simply restricting it to a line.

3.1.1 Extended-value Extensions

It is helpful to extend a convex function to all of $\mathbb{R}^n$ by setting its value to ∞ outside of its domain.

$$\tilde{f}(x) = \begin{cases} f(x) & x \in \textbf{dom } f \\ \infty & x \notin \textbf{dom } f \end{cases}$$

Defining the extension allows us to smoothly operate on all x instead of the domain of f.

3.1.2 First Order Conditions

Suppose f is differentiable (∇f exists at each point on the domain). Then f is convex if and only if domain f is convex and for all x,y

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

Proof in Boyd 3.1.3

The right hand side is simply the first order taylor approximation of $f(y)$ near x. The first order taylor approximation is a global underestimator of the function. If the first order taylor approximation is always a global underestimator of the function, then the function is convex. This inequality also states that if $\nabla f(x) = 0$, then x is a global minimizer of the function.

Equality is dropped when considering strict convexity. Similarly, the inequality is flipped for concavity.

3.1.3 Second Order Conditions

Now assume that f is twice differentiable. f is convex if and only if **dom** f is convex and the hessian is positive semidefinite.

$$\nabla^2 f(x) \succeq 0 \qquad \forall x \in \textbf{dom } f$$

This condition corresponds to the graph of the function having an upward curvature at x.

Concavity is represented by the same conditions, but the Hessian must obey $\nabla^2 f(x) \preceq 0$. Strict convexity is given by $\nabla^2 f(x) \succ 0$, but the converse is not true.

3.1.4 Examples

Proofs in Boyd 3.1.5

- Exponential: e^{ax} is convex on R
- Powers: x^a is convex on $\mathbb{R}_{++}$ when $a \geq 1$ or $a \leq 0$
- Powers of absolute value: $|x|^p$ for $p \geq 1$
- Logarithm: $\log x$
- Negative Entropy: $x \log x$
- Norms
- Max Function
- Quadratic-over-linear function: $\frac{x^2}{y}$
- Log-sum-exp: $\log(e^{x_1} + ... + e^{x_n})$
- Geometric Mean
- Log-determinant:

3.1.5 Sublevel Sets

The α-sublevel set of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$C_\alpha = \{x \in \textbf{dom } f | f(x) \leq \alpha\}$$

Sublevel sets of a convex function are convex sets. However, the converse is not true. Despite all of a function's sublevel sets being convex the function does not necessarily have to be convex. Sublebel property is a useful way of establishing a set as convex. Specifically, by identifying a set as a sublevel set of a convex function, we can determine that the set is convex.

3.1.6 Epigraph

The epigraph of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$\textbf{epi } f = \{(x,t) | x \in \textbf{dom } f, f(x) \leq t\}$$

A function is convex if and only if its epigraph is convex. A function is concave if and only if its hypograph ($\{(x,t)|t \leq f(x)\}$) is convex.

3.1.7 Jensen's Inequality and Extensions

$$f(\theta_1 x_1 + ... + \theta_k x_k) \leq \theta_1 f(x_1) + ... + \theta_k f(x_k)$$

3.2 Operations that Preserve Convexity

Proofs in Boyd 3.2

- Nonnegative Weighted Sum: $\tilde{f} = w_1 f_1 + ... + w_m f_m$ for m nonnegative weights and convex functions
- Composition with Affine Mapping: $g(x) = f(Ax + b)$, if f is convex (or concave) so is g
- Pointwise Maximum: $f(x) = \max\{f_1(x), f_2(x)\}$
- Pointwise Supremum: $g(x) = \sup_{y \in \mathcal{A}} f(x,y)$ where $f(x,y)$ is convex in x for each $y \in \mathcal{A}$
- Composition: Consider $f(x) = h(g(x))$
  - f is convex if h is convex and nondecreasing and g is convex
  - f is convex if h is convex and nondecreasing and g is concave
  - f is concave if h is concave and nondecreasing and g is concave
  - f is concave if h is concave and nondecreasing and g is convex
  - Note that the above relations generalize for vector composition
- Minimization: $g(x) = \inf_{y \in C} f(x,y)$ for f convex and C convex
- Perspective: $g(x,t) = tf(\frac{x}{t})$ if f is convex

4 Convex Optimization Problems

4.1 Optimization Problems

Recall the general form of an optimization problem:

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0 \qquad i = 1, ...m \\ & h_i(x) = 0 \qquad i = 1, ...p \end{aligned}$$

with the following features:

- Optimization Variable: $x \in \mathbb{R}^n$
- Objective Function: $f_0 : \mathbb{R}^n \to \mathbb{R}$
- Inequality Constraints: $f_i : \mathbb{R}^n \to \mathbb{R}$
- Equality Constraints: $h_i : \mathbb{R}^n \to \mathbb{R}$

Domain

The domain of an optimization problem is the set of points for which the objective function and constraits are defined.

$$\mathcal{D} = \bigcap_{i=0}^{m} \textbf{dom } f_i \cap \bigcap_{i=1}^{p} \textbf{dom } h_i$$

Feasible Set

A feasible point is one that obeys the constraints and a problem is said to be feasible if the domain contains at least one point. The set of fesabile points is called the feasible set or the constraint set. The optimal value for a problem p^* is defined as:

$$p^* = \inf\{f_0(x)|f_i(x) \leq 0, h_i(x) = 0\}$$

If the problem is infeasible, we have $p^* = \infty$ (inf of an empty set is ∞). If a problem has feasible points x_k such that $f(x_k) \to -\infty$ as $k \to \infty$, the we say the problem is unbounded below and $p^* = -\infty$.

Optimal Point

An optimal point is one that satisfies the optimal value for a problem $f(x^*) = p^*$.

$$X_{opt} = \{x|f(x) = p^*, f_i(x) \leq 0, h_i(x) = 0\}$$

A problem is said to be solvable if X_{opt} is non-empty.

A feasible solution that is obeys $f_0(x) \leq p^* + \epsilon$ is said to be ϵ-sub optimal and all feasible points satisfying this condition constitute the ϵ-sub optimal set.

The point x is said to be locally optimal if there is an $R > 0$ such that

$$f_0(x) = \inf\{f_0(z)|f_i(z) \leq 0, h_i(z) = 0, ||z - x||_2 \leq R\}$$

This is equivalent to solving the original optimization problem with an additional norm constraint.

Constraints are said to be active at x if the inequality is saturated (ie. $f_i(x) = 0$). A redundant constraint is one that can be removed without impacting the problem.

Feasibility Problems

If we set the objective function to 0 the outcome is either 0 if the feasible set is non-empty or ∞ if the feasible set is empty.

$$\begin{aligned} \text{find} \quad & x \\ \text{s.t.} \quad & f_i(x) \leq 0 \qquad i = 1, ...m \\ & h_i(x) = 0 \qquad i = 1, ...p \end{aligned}$$

4.2 Convex Optimization

A convex optimization problem takes the form

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0 \qquad i = 1, ...m \\ & a^T x = b_i \qquad i = 1, ...p \end{aligned}$$

where $f_0, ..., f_m$ are all convex functions and the equality constraint is affine. It is clear that the feasible set of a convex optimization problem is convex since it is the intersection of convex sets. Therefore, convex optimization is simply minimization over a convex set.

4.2.1 Local and Global Optima

In convex optimization, any locally optimal point is also globally optimal.

4.2.2 Optimality Criterion

For a differentiable function f_0, a point $x \in X$ the feasile set is optimal if and only if

$$\nabla f_0(x)^T (y - x) \geq 0 \qquad \forall y \in X$$

Conceptually, you can think of this as the diference between y and x in the first order taylor approximation of y.

If the problem is unconstrained, the necessary and sufficient condition reduces to $\nabla f_0(x) = 0$.

If the problem only has equivalence constraints, the feasible set is affine. If the function is non-negative, you can represent its optimality condition as

$$\nabla f_0(x) + A^T v = 0$$

4.3 Linear Optimization Problems

A linear program is an optimization problem with affine constraints and an affine objective.

$$\begin{aligned} \text{minimize} \quad & c^T x + d \\ \text{s.t.} \quad & G(x) \preceq h \\ & Ax = b \end{aligned}$$

These are a subclass of convex optimization problems. The standard form of linear programming only contains component wise nonnegativity constraints

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{s.t.} \quad & x \succeq 0 \\ & Ax = b \end{aligned}$$

If there are no equality constraints, we refer to the problem as an inequality form linear program with constraint $Ax \preceq b$.

4.3.1 Linear Fractional Programming

This class of problems refers to problems involving the minimization of the ratio of affine functions over a polyhedron.

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{s.t.} \quad & G(x) \preceq h \\ & Ax = b \end{aligned}$$

$$f_0(x) = \frac{c^T x + d}{e^T x + f}$$

Note that the objective function is quasiconvex. If the feasible set is non-empty, the problem can be rewritten as a linear program:

$$\begin{aligned} \text{minimize} \quad & c^T y + dz \\ \text{s.t.} \quad & Gy - hz \preceq 0 \\ & Ay - bz = 0 \\ & e^T y + fz = 1 \\ & z \geq 0 \end{aligned}$$

4.4 Quadratic Optimization Problems

5 Sections Skipped

- Relative interior of affine sets (Boyd 2.1.3)
- Simplex as polyhedron argument (Boyd 2.2.4)
- Convex hull description of polyhedra (Boyd 2.2.4)
- Linear-fractional and perspective functions (Boyd 2.3.3)
- Generalized Inequalities (Boyd 2.4)
- Dual Cones and generalized inequalities (Boyd 2.6)
- Representation as pointwise supremum of affine functions (Boyd 3.2.3)
- Equivalent Problems and Equivalent Convex Problems (Boyd 4.1.3 + Boyd 4.2.4)
- Quasiconvex Optimization (Boyd 4.2.5)