CME364A: Convex Optimization

Erich Trieschman

2021 Winter quarter class notes

1 Optimization overview

Optimization problems have the form

$$\text{minimize } f_0(x)$$
$$\text{subject to } f_i(x) \le b_i$$

The objective function is f_0 and the constraints are f_i Functions can be optimized locally or globally. The true distinction within optimization problems isn't between linear and nonlinear programming, but rather convex and nonconvex programming. This field studies methods for using known convex optimization solutions in as broad a way as possible to reach solutions (or bounds of solutions) for convex and nonconvex optimization problems alike. We can optimize

1.1 Least squares

Least squares is its own type of optimization problem without constraints:

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

- There exists an anlytical solution to this problem: $x = (A^T A)^{-1} x A^T b$

- There are more flexible techniques, including

    - Weighted least-squares: $f_0(x) = \sum_{i=1}^{k} w_i(a_i^T x - b_i)^2$ for weights, w_i
    - Regularized least-squares: $f_0(x) = \sum_{i=1}^{k}(a_i^T x - b_i)^2 + \rho \sum_{i=1}^{n} x_i^2$

1.2 Linear programming

Linear programs have the form

$$\text{minimize } c^T x$$
$$\text{subject to } a_i^T x \le b_i$$

There is no analytical solution to linear programs, but variety of effective algorithms to do so.

1.3 Convex optimization

An optimization problem where the objective and constraint functions are convex

$$f_i(\alpha x + (1 - \alpha)y) \le \alpha f_i(x) + (1 - \alpha)f_i(y), \text{ for all } x, y \in \mathbb{R}^n,\ \alpha \in \mathbb{R},\ 0 \le \alpha \le 1$$

2 Convex sets

2.1 Definitions

2.1.1 Affine sets

Affine set: for $x_1, x_2 \in C \subset \mathbb{R}^n$, $\theta x_1 + (1 - \theta)x_2 \in C$
Affine combination: for points, $x_1, \ldots, x_k \in C$, $\theta_1 x_1 + \cdots + \theta_k x_k$, where $\sum \theta_i = 1$, is an affine combination. An affine set contains every affine combination of its points

2.1.2   Convex sets

Convex set: a set, C is convex if for $x_1, x_2 \in C \subset \mathbb{R}^n$, $\theta x_1 + (1-\theta)x_2 \in C$

Convex combination: for points, $x_1, \ldots, x_k \in C$, $\theta_1 x_1 + \cdots + \theta_k x_k$, where $\sum \theta_i = 1$, is a convex combination. C convex $\Leftrightarrow$ every convex combination $\in C$

Convex hull: The set of all convex combinations of points in C: $\mathrm{conv}C = \{\theta_1 x_1 + \cdots + \theta_k x_k \mid x_i \in C, \theta_i \geq 0, i = 1, \ldots, k, \sum \theta_i = 1\}$

2.1.3   Cones

Cone: a set C is a cone if $\forall x \in C$ and $\forall \theta \geq 0$, we have $\theta x \in C$

Convex cone: a set is a convex cone if it is i) convex and ii) a cone. i.e., $\forall x, y \in C$ and $\forall \theta_1, \theta_2 \geq 0$, we have $\theta_1 x + \theta_2 y \in C$

Connic combination: for points, $x_1, \ldots, x_k \in C$, $\theta_1, \ldots, \theta_k \geq 0$, $\theta_1 x_1, \cdots + \theta_k x_k$ is a connic combination. C convex $\Leftrightarrow$ every convex combination $\in C$

Connic hull: The set of all connic combinations of points in C: $\{\theta_1 x_1 + \cdots + \theta_k x_k \mid x_i \in C, \theta_i \geq 0, i = 1, \ldots, k\}$

2.2   Key set examples

Simple examples below; more complex examples covered in subsections

- The empty set , any single point (i.e., singleton) $\{x_0\}$, and the whole space $\mathbb{R}^n$ are affine (hence, convex) subsets of $\mathbb{R}^n$

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

- A line segment is convex, but not affine (unless it reduces to a point)

- A ray, which has the form $\{x_0 + \theta v \mid \theta \geq 0\}$,where $v \neq 0$,is convex, but not affine. It is a convex cone if its base $x_0 = 0$

- Any subspace is affine, and a convex cone (hence convex)

2.2.1   Hyperplanes and halfspaces

Hyperplane: $\{x \mid a^T x = b\}$

Analytically, this is an affine set. Geometrically, this is the set of points with a constant inner product. It can also be written in the form below, for x_0 a point in the solution space. This indicates that a hyperplane is composed of an offset, x_0, and a vector orthogonal to a.

$$\{x \mid a^T x = b\} \iff \{x \mid a^T(x - x_0) = 0\} \iff x_0 + a^\perp$$

Halfspace: A hyperplane divides $\mathbb{R}^n$ into two halfspaces.

The halfspace extending in the $-a$ direction is $\{x \mid a^T x \leq b\}$. A vector making an obtuse angle with a is in this halfspace.

The halfspace extending in the a direction is $\{x \mid a^T x \geq b\}$. A vector making an acute angle with a is in this halfspace.

Halfspaces are convex, but not affine

2.2.2   Euclidean balls and ellipsoids

Euclinean balls are convex sets. They have the form

$$B(x_c, r) = \{x \mid \|x - x_c\|_2 \leq r\} \iff \{x \mid (x - x_c)^T(x - x_c) \leq r^2\} \iff \{x + ru \mid \|u\|_2 \leq 1\}$$

Ellipsoids are convex sets of the form

$$\mathbb{E} = \{x \mid (x - x_c)^T P^{-1}(x - x_c) \leq r^2\} \iff \{x + Au \mid \|u\|_2 \leq 1\}$$

Where P is SPD and determines how far the ellipsoid extends in each direction. A can be thought of (WLOG) as $P^{1/2}$

2.2.3   Norm balls and cones

Norm balls and cones are convex

Norm ball: $\{x \mid \|x - x_c\|_p \leq r\} \in \mathbb{R}^n$

Norm cone: $\{(x, t) \mid \|x\|_p \leq t\} \in \mathbb{R}^{n+1}$

2.2.4   Polyhedra

Polyhedra are convex sets. A polyhedron is the intersection of a finite number of halfspaces and hyperplanes. Affine sets, rays, line segments, and halfspaces are all polyhedra.

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

Simplexes are a special type of polyhedra with k independent vectors defining the set:

$$\mathcal{C} = \{\theta_0 v_0 + \cdots + \theta_k v_k \mid \theta \succeq 0, 1^T \theta = 1\}$$

2.2.5 Positive semidefinite cone

The set of symmetric semidefinite matrices S_+^n is a convex cone. i.e., if $\theta_1, \theta_2 \geq 0$ and $A, B \in S_+^n$, then $\theta_1 A + \theta_2 B \in S_+^n$. This can easily be seen by the rules of positive semidefiniteness:

$$x^T(\theta_1 A + \theta_2 B)x = \theta_1 x^T A x + \theta_2 x^T B x \geq 0 \text{ when } A \succeq 0, B \succeq 0, \ \theta_1, \theta_2 \geq 0$$

2.3 Operations preserving convexity

- **Intersection:** The intersection of a finite (or infinite) number of convex sets is also convex
  If S_1, S_2 convex $\Longrightarrow S_1 \cap S_2$ convex
  If S_α convex for all $\alpha \in \mathcal{A} \Longrightarrow \bigcap_{\alpha \in \mathcal{A}} S_\alpha$ convex

- **Affine functions:** An affine transformation of a convex set is convex. This can be thought of as scaling a convex set or shifting a convex set
  If S convex and $f(x)$ affine, then $f(S) = \{f(x) \mid x \in S\}$ is convex.

- **Perspective functions:** The perspective function scales vectors so that the last component is 1, and then drops the last component. The result of a perspective function applied to a convex set is a convex set.

- **Linear-fractional functions:** Linear-fractional functions are the composition of the perspective function with an affine function. They preserve convexity

2.4 Generalized inequality

Generalized inequality is a partial ordering in $\mathbb{R}^n$ that has many of the same properties as standard ordering in $\mathbb{R}$ ($<, >, \leq, \geq$). Those properties include i) preserved under addition, ii) transitive, iii) preserved under nonnegative scaling, iv) reflexive, v) antisymmetric, vi) preserved under limits (see p44).
We use *proper cones* to define this ordering. A cone, K is proper if it satisifes

- K is convex

- K is closed

- K is solid (i.e., nonempty interior)

- K is pointed (i.e., contains no full line, only rays)

A generalized inequality can be defined with a proper cone, K, as

$$x \preceq_K y \Longleftrightarrow y - x \in K$$
$$x \prec_K y \Longleftrightarrow y - x \in int K$$

Two useful examples of generalized inequalities:

- Nonnegative orthant: componentwise inequality

- Positive semidefinite cone: Difference in PSD matrices is also PSD (i.e., nonnegative eigenvalues)

2.4.1 Minimum and minimal elements

In generalized inequalities, not all values are comparable. This motivates different definitions for minimum elements.

- $x \in S$ is *minimum* if for every $y \in S$, we have $x \preceq_K y$. In set notation, $x \in S$ is minimum of $S \Longleftrightarrow S \subseteq x + K$

- $x \in S$ is *minimal* if $y \preceq_K x$ for $y \in S$, then $y = x$. In set notation, $x \in S$ is minimal in $S \Longleftrightarrow (x - K) \cap S = \{x\}$

Visually, think of how far negative you can go in each direction. If you reach a single point in this effort, it's **minimum**. If you reach a line, those points on the line are **minimal**

2.5 Separating hyperplanes

The **separating hyperplane theroem** result states: Suppose C and D are nonempty disoint convex sets ($C \cap D = \emptyset$). Then, there exist $a \neq 0$ and b, such that $a^T x \leq b \ \forall x \in C$ and $a^T x \geq b \ \forall x \in D$. The hyperplane $\{x \mid a^T x = b\}$ is called the separating hypterplane for sets C and D. See proof on p. 46.

2.6 Supporting hyperplanes

For $C \subseteq \mathbb{R}^n$ and $x_0 \in bdC$, the **supporting hyperplane** is the hyperplane $\{x \mid a^T x = a^T x_0\}$ that satisfies $a^T x \le a^T x_0 \ \forall x \in C$.

This is equivalent to saying that the point x_0 and the set C are separated by the hyperplane $\{x \mid a^T x = a^T x_0\}$. The geometric interpretation is that the hyperplane $\{x \mid a^T x = a^T x_0\}$ is tangent to C at x_0, and the halfspace $\{x \mid a^T x \le a^T x_0\}$ contains C.

The **supporting hyperplane theorem**, states that for any nonempty convex set C, and any $x_0 \in bdC$, there exists a supporting hyperplane to C at x_0. The supporting hyperplane theorem is readily proved from the separating hyperplane theorem.

2.7 Dual cones

For cone K, the **dual cone**, K^* is defined as

$$K^* = \{y \mid x^T y \ge 0 \ \forall x \in K\}$$

- K^* is always a cone and always convex, even when K is not convex

- Geometrically, $y \in K^* \iff -y$ is normal of a hyperplane supporting K at the origin

- for proper cone, K, $x \preceq_K y \iff \lambda^T x \le \lambda^T y$ for all $\lambda \preceq_{K^*} 0$

3 Convex functions

3.1 Definitions

A function, $f : \mathbb{R}^n \to \mathbb{R}$ is **convex** if $dom f$ is a convex set and if

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

- Geometrically, this can be interpreted as a convex function being a function where a line segment connecting any two points on the function lies above the graph.

- Another interpretation: f is convex $\iff \forall x \in dom f$ the function $g(t) = f(x + tv)$ is convex (requiring $x + tv \in dom f$)

- The extended value extension of a convex function sets $f(x) = \infty$ for all $x \notin dom f$

We can verify convexity through one of three approaches

- Directly verify the inequality

- Restrict the function to an arbitrary line and verify convexity of the resulting function of one variable

- Verify that the Hessian is positive semidefinite (second-order conditions)

3.1.1 First-order conditions

For differentiable f with convex domain:

$$f \text{ convex } \iff f(y) \ge f(x) + \nabla f(x)^T (y - x) \ \forall x, y \in dom f$$

This is the first-order Taylor approximation of f near x. For convex sets, the first-order Taylor approximation is a global underestimator of the function

3.1.2 Second-order conditions

For twice-differentiable f with convex domain:

$$f \text{ convex } \iff \nabla^2 f(x) \succeq 0$$

The second-order condition states that the Hessian of f must be a positive semidefinite matrix for f to be concave. Geometrically, this means we require positive, upward curvature of f.
We can evaluate PSD matrix by considering its eigenvalues. Note: rank-1 matrices are not positive semidefinite, so if we can rewrite a Hessian as a rank-1 matrix, we can show that it is not convex.

3.1.3 Examples

- *Exponential:* e^{ax} is convex on $\mathbb{R}$

- *Powers:* x^a is convex on $\mathbb{R}_{++}$ when $a \geq 1, a \leq 0$, concave otherwise

- *Powers of absolute value:* $|x|^p$ is convex on $\mathbb{R}$ for $p \geq 1$

- *Logarithm:* $log(x)$ is concave on $\mathbb{R}_{++}$

- *Negative entropy:* $xlog(x)$ is convex on $\mathbb{R}_{++}$ or $\mathbb{R}$

- *Norms:* every norm $\|x\|_x$ is convex on $\mathbb{R}^n$

- *Max function:* $f(x) = max\{x_1, \ldots, x_n$ is convex on $\mathbb{R}^n$

- *Quadratic-over-linear:* x^2/y is convex on $\mathbb{R} \times \mathbb{R}_{++} = \{(x,y) \in \mathbb{R}^2 \mid y > 0\}$

- *Log-sum-exp:* $f(x) = log(\sum_i e^{x_i})$ is convex on $\mathbb{R}^n$

- *Geometric mean:* $f(x) = (\prod_{i=1}^n x_1)^{1/n}$ is concave on $\mathbb{R}^n_{++}$

- *Log-determinant:* $f(X) = log(det(X))$ is concave on $\mathbb{S}^n_{++}$

3.1.4 Sublevel sets

Sublevel sets of a convex function are convex. The α-**sublevel set** of a function f is defined as

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

Note: The converse is not true. A function can have all its sublevel sets convex, but not be convex
Note: If f concave, then the α-superlevel set given by $\{x \in dom f \mid f(x) \geq \alpha\}$ is convex

3.1.5 Epigraph

The graph of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined in $\mathbb{R}^{n+1}$ as $\{(x, f(x)) \mid x \in dom f\}$
The epigraph of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined in $\mathbb{R}^{n+1}$ as $\{(x, t) \mid f(x) \leq t\}$

3.1.6 Jensen's inequality

The basic convexity inequality is sometimes referred to as Jensen's inequality

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

It can easily be extended to convex combinations of more than two points

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

And even to infinite sums, integrals and expected values.

3.2 Operations preserving convexity

- *Nonnegative weighted sums:* $f = w_1 f_1 + \cdots + w_m f_m$ convex for f_i convex and $w_i \geq 0$

- *Composition with an affine mapping:* $g(x) = f(Ax + b)$ convex for f convex and $dom g = \{x \mid Ax + b \in dom g\}$

- *Point-wise maximum and supremum:* $f(x) = max\{f_1(x), f_2(x)\}$ convex for f_1, f_2 convex and $dom f = dom f_1 \cap dom f_2$ convex

- *Composition:* $f(x) = h(g(x))$ convex under certain conditions for scalar composition and vector composition

- *Minimization* $g(x) = \inf_{y \in C} f(x, y)$ convex for f convex, C a convex set, and $g(x) > -\infty$ for all x

- *Perspective of a function:* $g(x, t) = tf(x/t)$ convex for $f : \mathbb{R}^n \to R$ convex with $dom g = \{(x, t) \mid x/t \in dom f, t > 0\}$

3.2.1 Scalar convexity-preserving function composition

In the scalar case we consider $h : \mathbb{R} \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}$
The convexity of f is determined by $f'' \geq 0$, and the second derivative of the composition is

$$f''(x) = h''(g(x))g'(x)^2 + h'(g(x))g''(x)$$

How can we determine if $f(x) = h(g(x))$ is convex ($f''(x) \geq 0$)?

- f convex if: h convex ($h'' \geq 0$) and h nondecreasing ($h' \geq 0$) and g is convex ($g'' \geq 0$)

- f convex if: h convex ($h'' \geq 0$) and h nonincreasing ($h' \leq 0$) and g is concave ($g'' \leq 0$)

- f concave if: h concave ($h'' \geq 0$) and h nondecreasing ($h' \geq 0$) and g is concave ($g'' \geq 0$)

- f concave if: h concave ($h'' \geq 0$) and h nonincreasing ($h' \geq 0$) and g is convex ($g'' \geq 0$)

3.2.2 Vector convexity-preserving function composition

In the vector case we consider $h : \mathbb{R}^k \to \mathbb{R}$, $g_i : \mathbb{R}^n \to \mathbb{R}$
Similarly, the convexity of f is determined by $f'' \geq 0$, and the second derivative of the composition is

$$f''(x) = g'(x)^T \nabla^2 h(g(x))g'(x)^2 + \nabla h(g(x))^T g''(x)$$

How can we determine if $f(x) = h(g(x))$ is convex ($f''(x) \geq 0$)?

- f convex if: h convex ($\nabla^2 h \succeq 0$) and for reach argument, i,

    - h is nondecreasing in that argument ($\partial h / \partial x_i \geq 0$) and g_i is convex ($g'' \geq 0$)
    - h is nonincreasing in that argument ($\partial h / \partial x_i \leq 0$) and g_i is concave ($g'' \leq 0$)

- f concave if: h concave ($\nabla^2 h \preceq 0$) and for reach argument, i,

    - h is nondecreasing in that argument ($\partial h / \partial x_i \geq 0$) and g_i is concave ($g'' \leq 0$)
    - h is nonincreasing in that argument ($\partial h / \partial x_i \leq 0$) and g_i is convex ($g'' \geq 0$)

3.3 Conjugate function

For function f, the conjugate of the function, f^* is defined as

$$f^*(y) = \sup_{x \in dom f} (y^T x - f(x))$$

We see immediately that f^* is a convex function, since it is the pointwise supremum of a family of convex (indeed, affine) functions of y

3.4 Quasiconvex functions

A function f is quasiconvex if its domain and all its sublevel sets are convex. A function is quasiconcave if $-f$ is quaiconvex (i.e., every superlevel set is convex). A function that is both quasiconvex and quasiconcave is called quasilinear.

Jensen's inequality for quasiconvex functions characterizes quasiconvexity:

$$f \text{ quasiconvex} \iff f(\theta x + (1 - \theta)y) \leq \max\{f(x), f(y)\}$$

TODO: Additional properties of quasiconvex functions can be found in the textbook, including

- First order conditions: quasiconvexity can be characterized by first order conditions

- Second order conditions: quasiconvexity can be characterized by second order conditions

- Nonnegative weighted maximum

- Composition: g quasiconvex and h nondecreasing, then f = h(g(x)) is quasiconvex

- Minimization

3.5 Log-concave functions

A function $f : \mathbb{R}^n \to \mathbb{R}$ is log-concave if $f(x) > 0$ for all $x \in dom f$ and $\log f$ is concave. f is log-convex if $\log f$ is convex. Log concavity can be characterized by the following inequality

$$f(\theta x + (1 - \theta)y) \geq f(x)^\theta f(y)^{1-\theta} \text{ for } x, y \in dom f \text{ and } 0 \leq \theta \leq 1$$

TODO: properties of log-concave functions include

- Twice differentiable property
- log-convexity and log-concavity are closed under multiplication and positive scaling
- In some special cases, log-concavity is preserved by integration

3.6 Convexity with respect to generalized inequalities

TODO

3.7 Disciplinned Convex Programming

Algorithmic methods for writing functions to determine concavity and convexity

- Stanford DCP tutorial
- CVXPy DCP tutorial

4 Applied concepts

4.1 How to set up an applied problem

- Write down all variables
- Write down all constants
- Write out the size of all variables and constants (e.g., $x \in \mathbb{R}^n$)
- Write out optimization problem in a DCP compliant way
- Write out all constraints in a DCP compliant way

4.2 Constructing the dual of a perturbed problem

Take the original primal problem and its dual. Assume convexity and strong duality

$$
\begin{aligned}
\textbf{Primal:} \quad & min. \ f_0(x) \\
& s.t. \ f_i(x) \leq 0 \\
& \quad h_i(x) = 0 \\
\textbf{Dual:} \quad & min. \ g(\lambda, \nu) = \inf_x \{L(\lambda, \nu)\} = \inf_x \{f_0(x) + \sum_i \lambda_i f_i(x) + \sum_j \nu_j h_j(x)\} \\
& s.t. \ \lambda \preceq 0 \\
\textbf{Perturbed primal:} \quad & min. \ f_0(x) \\
& s.t. \ f_i(x) \leq u_i \\
& \quad h_i(x) = v_i \\
\textbf{Perturbed dual:} \quad & min. \ g_{pert}(\lambda, \nu) = \inf_x \{L_{pert}(\lambda, \nu)\} \\
& = \inf_x \{f_0(x) + \sum_i \lambda_i (f_i(x) - u_i) + \sum_j \nu_j (h_j(x) - v_i)\} \\
& = \inf_x \{f_0(x) + \sum_i \lambda_i f_i(x) + \sum_j \nu_j h_j(x)\} - \lambda^T u - \nu^T v \\
& = g(\lambda, \nu) - \lambda^T u - \nu^T v, \text{ written as a function of the original lagrangian dual problem} \\
& s.t. \ \lambda \preceq 0
\end{aligned}
$$

Now, by properties of strong and weak duality we have

$$f_0(\tilde{x}) \geq g_{pert}(\tilde{\lambda}, \tilde{\mu}), \text{ by weak duality for any feasible values}$$
$$p^*_{pert}(u, v) \geq g_{pert}(\lambda, \mu), \text{ by weak duality, since the prior result holds for all values}$$
$$p^*_{pert}(u, v) \geq g_{pert}(\lambda^*, \mu^*), \text{ the optimal values of the original problem}$$
$$= g(\lambda^*, \mu^*) - \lambda^{*T}u - \nu^{*T}v, \text{ rewriting equality from above}$$
$$= p^*(0, 0) - \lambda^{*T}u - \nu^{*T}v, \text{ from strong duality of } p^*(0, 0) = g(\lambda^*, \nu^*)$$

5 Assorted techniques

5.1 ...for constructing convex problems

- Monotonicity of functions
- Condition on maximum of functions $\implies$ condition on each individual function within the maximum
- homogenous functions
- change of variables: perspective, $g(x, t) = tf(x/t)$
- change of varibles: replace a log of a variable with a new variable (e.g., $y = \log x$). This may lead to creating a perspective function
- change of variables: replace a product of two variables with a single variable (e.g., $z = x/y$, $z = xy$)
- scale variables in the problem – can the variables be scaled without changing the objective or constraints?
- or scale variables after solving the problem, i.e., $x = \tilde{x}/(1^T \tilde{x})$
- keep things in norms and try to rewrite norms as subcomponents
- When working with functions of the form $(x_i)_-$ and $(x_i)_+$, breaking it into two vectors: $x_+ \succeq 0, x_- \succeq 0, x = x_+ - x_-$
- The geometric mean, $g(u, v) = \sqrt{uv}$ is concave
- Change of variables for normal distribution: $\theta = \Sigma^{-1}$, $\omega = \Sigma^{-1}\mu$
- Logarithms and maximums: $\max(\log f(x), \log g(x)) = \log \max(f(x), g(x))$

5.2 ...for coding in CVXPY

- Define variable constraints in the variable definition `cp.Variable(n, nonneg=True)` or `cp.Variable((m, n), PSD=True)`
- Axis-level norms: `cp.norm(u, 2, axis=0)`
- sum of squares: `cp.sum_squares()`

6 Additional resources