14 Convex Optimization

Many optimization objectives generated by LTI system design and analysis are characterized by convexity of the underlying constraints. This lecture is devoted to recognizing and working with convex constraints.

14.1 Basic Definitions

Definition 14.1 *A subset Ω of $V = \mathbb{R}^n$ is called convex if*

$$cv_1 + (1-c)v_2 \in \Omega \quad \forall v_1, v_2 \in \Omega, c \in [0,1] .$$

In other words, a set is convex whenever the line segment connecting any two points of Ω lies completely within Ω. In many applications, the elements of Ω are, formally speaking, not vectors but other mathematical objects, such as matrices, polynomials, etc. What matters, however, is that Ω is a subset of a set V such that a one-to-one correspondence between $\mathbb{R}^n$ and V is established for some n. For example, the set S^n of all symmetric $n \times n$ matrices is a vector space, because of the natural one-to-one correspondence between S^n and $R^{n(n+1)/2}$. Using this definition directly, in some situations it would be rather difficult to check whether a given set is convex. The following simple statement is of a great help.

Lemma 14.2 *Let K be a set of affine functions on $V = \mathbb{R}^n$, i.e. elements $f \in K$ are functions $f : V \to \mathbb{R}$ such that*

$$f(cv_1 + (1-c)v_2) = cf(v_1) + (1-c)f(v_2) \quad \forall c \in \mathbb{R},\ v_1, v_2 \in V.$$

Then the subset Ω of V defined by

$$\Omega = \{v \in V : f(v) \geq 0 \quad \forall f \in K\}$$

is convex. In other word, any set defined by linear inequalities is convex.

Proof: Let $v_1, v_2 \in \Omega$ and $c \in [0,1]$. Since $f(v_1) \geq 0$ and $f(v_2) \geq 0$ for all $f \in K$, and $c \geq 0$ and $1-c \geq 0$, we conclude that

$$f(cv_1 + (1-c)v_2) = cf(v_1) + (1-c)f(v_2) \geq 0 \quad \forall f \in K .$$

Hence $cv_1 + (1-c)v_2 \in \Omega$.

■

Here is an example of how Lemma 14.2 can be used.

Example. Let us prove that the subset $\Omega = \mathbf{S}^n_+$ of the set $V = \mathbf{S}^n$ of symmetric n-by-n matrices, consisting of all positive semidefinite matrices, is convex. Note that doing this via the “nonnegative eigenvalues” definition of positive semidefiniteness would be difficult. Luckily, there is another definition: a matrix $M \in \mathbf{S}_+^n$ is positive semidefinite if and only if $x^T Mx \geq 0 \quad \forall x \in \mathbb{C}^n$. Note that any $x \in \mathbb{C}^n$ defines an affine (actually, a linear) function $f = f_x : \mathbf{S}^n \to \mathbb{R}$ according to $f_x(M) = x^T Mx$. Hence, $\mathbf{S}_+^n$ is a subset of $\mathbf{S}^n$ defined by some (infinite) set of linear inequalities. According to Lemma 14.2, $\mathbf{S}_+^n$ is a convex set.

14.1.1 Convex Functions

Definition 14.3 *$f : \Omega \to \mathbb{R}$ is said to be convex if the following two conditions hold:*

(i) $\Omega \subset \mathbb{R}^n$ is convex;

(ii) the inequality

$$f(cv_1 + (1-c)v_2) \leq cf(v_1) + (1-c)f(v_2)$$

holds for all $v_1, v_2 \in \Omega, c \in [0, 1]$.

We say f is concave if $-f$ is convex.

Note that condition (ii) has the meaning that any segment connecting two points on the graph of f lies above the graph of f. The definition of a convex function does not help much with proving that a given function is convex. The following statements are of great help in establishing convexity of functions.

Lemma 14.4 *Let $\Omega \subset \mathbb{R}^n$ be a (open) convex subset of $\mathbb{R}^n$. Let $f : \Omega \to \mathbb{R}$ be differentiable. Then f is convex iff*

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

for all $x, y \in \Omega$.

This lemma shows that for a convex function, the first order Taylor approximation is in fact a global underestimator of the function. Conversely, if the first order Taylor approximation of a function is always a global underestimator of the function then the function is convex.

The inequality above shows that from local information about a convex function (i.e. its derivative at a pt), we can derive global information (i.e. a global underestimator). This is perhaps the most important property of convex functions, and explains some of the remarkable properties of convex optimization.

Now, let us call a function $f : \Omega \to \mathbb{R}$ defined on a subset Ω of $\mathbb{R}^n$ twice differentiable at a point v_0 if there exists a symmetric matrix $W \in \mathbf{S}_{\mathbb{R}}^n$ and a row vector p such that

$$\frac{f(v) - f(v_0) - p(v - v_0) - 1/2(v - v_0)^T W(v - v_0)}{||v - v_0||^2} \to 0 \; as \; v \to v_0 \; v \in \Omega$$

in which case $p = f'(v_0)$ is called the first derivative of f at v_0 and $W = f''(v_0)$ is called the second derivative of f at v_0.

Lemma 14.5 *Let $\Omega \subset \mathbb{R}^n$ be a convex subset of $\mathbb{R}^n$. Let $f : \Omega \to \mathbb{R}$ be a function which is twice differentiable and has a positive semidefinite second derivative $f''(v_0) \geq 0$ at any point $v_0 \in \Omega$. Then f is convex.*

For example, let Ω be the positive quadrant in $\mathbb{R}^2$, i.e. the set of vectors $[x; y] \in \mathbb{R}^2$ with positive components $x > 0, y > 0$. Obviously Ω is convex. Let the function $f : \Omega \to \mathbb{R}$ be defined by $f(x, y) = 1/xy$. According to the previous Lemma, f is convex, because the second derivative

$$W(x,y) = \begin{bmatrix} d^2f/dx^2 & d^2f/dydx \\ d^2f/dxdy & d^2f/dy^2 \end{bmatrix} = \begin{bmatrix} 2/x^3y & 1/x^2y^2 \\ 1/x^2y^2 & 2/xy^3 \end{bmatrix}$$

is positive definite on Ω.

Lemma 14.6 *Let $\Omega \subset V$ be a convex set of a $V = \mathbb{R}^n$. Let P be a set of affine functions on V such that*

$$f(v) = \sup_{p \in P} p(v) < \infty \quad \forall v \in \Omega \ .$$

Then $f : \Omega \to \mathbb{R}$ is a convex function.

In addition to Lemma 14.5 and Lemma 14.6, which help establishing convexity “from scratch”, the following statements can be used to derive convexity of one function from convexity of other functions.

Lemma 14.7 *Let V be a vector space, $\Omega \subset V$.*

(a) If $f : \Omega \to \mathbb{R}$ and $g : \Omega \to \mathbb{R}$ are convex functions then $h : \Omega \to \mathbb{R}$ defined by $h(v) = f(v) + g(v)$ is convex as well.

(b) If $f : \Omega \to \mathbb{R}$ is a convex function and $c > 0$ is a positive real number then $h : \Omega \to \mathbb{R}$ defined by $h(v) = cf(v)$ is convex.

(c) If $f : \Omega \to \mathbb{R}$ is a convex function, U is a vector space, and $L : U \to V$ is an affine function, i.e.

$$L(cu_1 + (1 - c)u_2) = cL(u_1) + (1 - c)L(u_2) \quad \forall c \in \mathbb{R},\ u_1, u_2 \in U \ ,$$

then the set $L^{-1}(\Omega) = \{u \in U : L(u) \in \Omega\}$ is convex, and the function $f \circ L : L^{-1}(\Omega) \to \mathbb{R}$ defined by $(f \circ L)(u) = f(L(u))$ is convex.

Example. Simple example of convex ftns

- e^{ax} is convex in $\mathbb{R}$.
- $|x|^p,\ p \geq 1$ is convex on $\mathbb{R}$.
- $\log x$ is concave on $\{x \mid x > 0\}$.
- $x \log x$ (negative entropy) is convex on $\{x \mid x > 0\}$.

- $\log \mathbf{erfc}(x) = \log\left(\frac{2}{\sqrt{\pi}} \int_{-\infty}^{x} e^{-u^2/2} du\right)$ is concave on $\mathbb{R}$.
- Every norm on $\mathbb{R}^n$ is convex.
- $f(x) = \max_i x_i$ is convex on $\mathbb{R}^n$.
- quadratic function: $f(x) = (1 = 2)xTPx + qTx + r$ (with $P \in S^n$)

  $$\nabla f(x) = Px + q, \quad \nabla^2 f(x) = P$$

  convex if $P \succeq 0$
- least-squares objective: $f(x) = ||Ax - b||_2^2$

  $$\nabla f(x) = 2A^T(Ax - b), \quad \nabla^2 f(x) = 2A^T A$$

  convex for any A.
- maximum eigenvalue of symmetric matrix: for $X \in S^n$,

  $$\lambda_{\max}(X) = \sup_{||y||_2=1} y^T X y$$

  is convex.
- $\exp g(x)$ is convex if g is convex.
- $1/g(x)$ is convex if g is concave and positive.

14.1.2 Quasi-Convex Functions

Definition 14.8 *Let $\Omega \subset V$ be a subset of a vector space. A function $f : \Omega \to \mathbb{R}$ is called quasi-convex if its level sets*

$$\Omega = \{v. : f(v) < \gamma\}$$

are convex for all γ.

It is easy to prove that any convex function is quasi-convex. However, there are many important quasi-convex functions which are not convex. For example, let $\Omega = \{(x, y) : x > 0, y > 0\}$ be the positive quadrant in $\mathbb{R}^2$. The function $f : \Omega \to \mathbb{R}$ defined by $f(x, y) = -xy$ is not convex but quasi-convex. A rather general definition leading to quasi-convex functions is given as follows.

Lemma 14.9 *Let $\Omega \subset V$ be a subset of a vector space. Let $P = \{(p, q)\}$ be a set of pairs of affine functions $p, q : \Omega \to \mathbb{R}$ such that*

(a) inequality $p(v) \geq 0$ holds for all $v \in \Omega$, $(p, q) \in P$;

(b) for any $v \in \Omega$ there exists $(p, q) \in P$ such that $p(v) > 0$.

Then the function $f : \Omega \to \mathbb{R}$ defined by

$$f(v) = \inf\{\lambda : \lambda p(v) \geq q(v)\ \forall (p,q) \in P\} \tag{1}$$

is quasi-convex.

For example, the largest generalized eigenvalue function $f(v) = \lambda_{\max}(\alpha, \beta)$ defined on the set $\Omega = \{v\}$ of pairs $v = (\alpha, \beta)$ of matrices $\alpha, \beta \in \mathbf{S}^n$ such that α is positive semidefinite and $\alpha \neq 0$, is quasi-convex. To prove this, recall that

$$\lambda_{\max}(\alpha, \beta) = \inf\{\lambda : \lambda x^T \alpha x \geq x^T \beta x\ \forall x \in \mathbb{C}^n\} .$$

This is a representation of $\lambda_{\max}$ in the form (1) with $(p,q) = (p_x, q_x)$ defined by an $x \in \mathbb{C}^n$ according to

$$p_x(v) = x^T \alpha x, q_x(v) = x^T \beta x \quad \text{ where } v = (\alpha, \beta) .$$

Since for any $\alpha > 0$ there exists $x \in \mathbb{C}$ such that $x^T \alpha x > 0$, Lemma 14.9 implies that $\lambda_{\max}$ is quasi-concave on Ω.

14.2 Standard Convex Optimization Setups

There exist a variety of significantly different tasks commonly referred to as convex optimization problems.

14.2.1 Minimization of a Convex Function

The standard general form of a convex optimization problem is minimization of a convex function $f : \Omega \to \mathbb{R}$. The remarkable feature of such optimization is that for any point $v \in \Omega$ which is not a minimum of f and for any number $\gamma \in (\inf(f), f(v))$ there exists a vector u such that $v + tu \in \Omega$ and $f(v + tu) \leq f(v) + t(\gamma - f(v))$ for all $t \in [0, 1]$. (In other words, f is decreasing quickly in the direction u.) In particular, any local minimum of a convex function is its global minimum.

Proposition 14.10 *Suppose that $f : \Omega \to \mathbb{R}$ is convex. If f has a local minimum at $x_0 \in \mathbf{S}$ then $f(x_0)$ is also the global minimum of f. If f is strictly convex, then x_0 is moreover unique.*

Proof: Let f be convex and suppose that f has a local minimum at $x_0 \in \mathbf{S}$. Then for all $x \in \mathbf{S}$ and $\alpha \in (0, 1)$ sufficiently small,

$$f(x0) \leq f((1-\alpha)x_0 + \alpha x) = f(x_0 + \alpha(x - x_0)) \leq (1-\alpha)f(x_0) + \alpha f(x). \tag{2}$$

This implies that

$$0 \leq \alpha(f(x) - f(x_0)) \tag{3}$$

or $f(x_0) \leq f(x)$. So $f(x_0)$ is a global minimum. If f is strictly convex, then the second inequality in (2) is strict so that (3) becomes strict for all $x \in \mathbf{S}$. Hence, x_0 is unique.

■

It is very important to emphasize that proposition 14.10 does not make any statement about *existence* of optimal solutions $x_0 \in \mathbf{S}$ which minimize f. It merely says that all local minima of f are also global minima. It therefore suffices to compute local minima of a convex function f to actually determine its global minimum. Proposition 14.10 does not hold for quasi-convex functions.

While it is reasonable to expect that convex optimization problems are easier to solve, and reducing a given design setup to a convex optimization is frequently a major step, it must be understood clearly that convex optimization problems are useful only when the task of calculating $f(v)$ for a given v (which includes checking that $v \in \Omega$.) is not too complicated.

For example, let X be any finite set and let $g : X \to \mathbb{R}$ be any real-valued function on X. Minimizing g on X can be very tricky when the size of X is large (because there is very little to offer apart from the random search). However, after introducing the vector space V of all functions $v : X \to \mathbb{R}$, the convex set Ω can be defined as the set of all probability distributions on X, i.e. as the set of all $v \in V$ such that

$$v(x) \geq 0 \;\forall x, \quad \sum_{x \in X} v(x) = 1,$$

and $f : \Omega \to \mathbb{R}$ can be defined by

$$f(v) = \sum_{x \in X} g(x)v(x) \; .$$

Then f is convex and, formally speaking, minimization of g on X is "equivalent" to minimization of f on Ω, in the sense that the argument of minimum of f is a function $v \in \Omega$. which is non-zero only at those $x \in X$ for which $g(x) = \min(g)$. However, unless some nice simplification takes place, $f(v)$ is "difficult" to evaluate for any particular v (the "brute force" way of doing this involves calculation of $g(x)$ for all $x \in X$), this "reduction" to the convex optimization does not make much sense.

14.2.2 Linear Programs

As it follows from Lemma 14.2, a convex set Ω can be defined by a family of linear inequalities. Similarly, according to Lemma 14.6, a convex function can be defined as supremum of a family of affine functions. The problem of finding the minimum of f on Ω when Ω is a subset of $\mathbb{R}^n$ defined by a finite family of linear inequalities, i.e.

$$\Omega = \{v \in \mathbb{R}^n : a_i^T v \leq b_i, i = 1, \cdots, m\} \;, \tag{4}$$

and $f : \Omega \to \mathbb{R}$ is defined as supremum of a finite family of affine functions,

$$f(v) = \max_{i=1,\cdots,k} c_i^T v + d_i, \tag{5}$$

where a_i, c_i are given vectors in $\mathbb{R}^n$, and b_i, d_i are given real numbers, is referred to as a linear program. In fact, all linear programs defined by (4), (5) can be reduced to the case when f is a linear function, by appending an extra component v_{n+1} to v, so that the new decision variable becomes

$$\bar{v} = \begin{bmatrix} v \\ v_{n+1} \end{bmatrix} \in \mathbb{R}^{n+1} ,$$

introducing the additional linear inequalities

$$\bar{c}_i^T \bar{v} = c_i^T v - v_{n+1} \leq -d_i ,$$

and defining the new objective function $\bar{f}$ by

$$\bar{f}(\bar{v}) = v_{n+1} .$$

Most linear programming optimization engines would work with the setup (4), (5), where $f(v) = Cv$ is a linear function. The common equivalent notation in this case is

$$\min\ Cv \quad \text{subject to } Av \leq B ,$$

where a_i^T are the rows of A, b_i are the elements of the column vector B, and the inequality $Av \leq B$ is understood component-wise.

14.2.3 Semidefinite Programs

A semidefinite program is typically defined by an affine function $\alpha : \mathbb{R}^n \to \mathbf{S}_{\mathbb{R}}^N$ and a vector $c \in \mathbb{R}^n$, and is formulated as

$$\min\ Cv \quad \text{subject to } \alpha(v) \geq 0 .$$

Note that in the case when

$$\alpha(v) = \begin{bmatrix} b_1 - a_1^T v & & 0 \\ & \ddots & \\ 0 & & b_N - a_N^T v \end{bmatrix}$$

is a diagonal matrix valued function, the special semidefinite program becomes a general linear program. Therefore, linear programming is a special case of semidefinite programming.

Since a single matrix inequality $\alpha \geq 0$ represents an infinite number of inequalities $x^T \alpha x \geq 0$, semidefinite programs can be used to represent constraints much more efficiently than linear programs. On the other hand, software for solving general semidefinite programs appears to be not as well developed as in the case of linear programming.