# A Tutorial on Convex Optimization

Haitham Hindi
Palo Alto Research Center (PARC), Palo Alto, California
email: hhindi@parc.com

*Abstract*— In recent years, convex optimization has become a computational tool of central importance in engineering, thanks to it's ability to solve very large, practical engineering problems reliably and efficiently. The goal of this tutorial is to give an overview of the basic concepts of convex sets, functions and convex optimization problems, so that the reader can more readily recognize and formulate engineering problems using modern convex optimization. This tutorial coincides with the publication of the new book on convex optimization, by Boyd and Vandenberghe [7], who have made available a large amount of free course material and links to freely available code. These can be downloaded and used immediately by the audience both for self-study and to solve real problems.

## I. INTRODUCTION

Convex optimization can be described as a fusion of three disciplines: optimization [22], [20], [1], [3], [4], convex analysis [19], [24], [27], [16], [13], and numerical computation [26], [12], [10], [17]. It has recently become a tool of central importance in engineering, enabling the solution of very large, practical engineering problems reliably and efficiently. In some sense, convex optimization is providing new indispensable computational tools today, which naturally extend our ability to solve problems such as least squares and linear programming to a much larger and richer class of problems.

Our ability to solve these new types of problems comes from recent breakthroughs in algorithms for solving convex optimization problems [18], [23], [29], [30], coupled with the dramatic improvements in computing power, both of which have happened only in the past decade or so. Today, new applications of convex optimization are constantly being reported from almost every area of engineering, including: control, signal processing, networks, circuit design, communication, information theory, computer science, operations research, economics, statistics, structural design. See [7], [2], [5], [6], [9], [11], [15], [8], [21], [14], [28] and the references therein.

The objectives of this tutorial are:

1) to show that there is a straight forward, systematic rules and facts, which when mastered, allow one to quickly deduce the convexity (and hence tractability) of many problems, often by inspection;
2) to review and introduce some canonical optimization problems, which can be used to model problems and for which reliable optimization code can be readily obtained;
3) emphasize the modeling and formulation aspect; we do not discuss the aspects of writing custom codes.

We assume that the reader has a working knowledge of linear algebra and vector calculus, and some (minimal) exposure to optimization.

Our presentation is quite informal. Rather than provide details for all the facts and claims presented, our goal is instead to give the reader a flavor for what is possible with convex optimization. Complete details can be found in [7], from which all the material presented here is taken. Thus we encourage the reader to skip sections that might not seem clear and continue reading; the topics are not all interdependent.

*Motivation*

A vast number of design problems in engineering can be posed as constrained optimization problems, of the form:

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p. \end{array} \quad (1)$$

where $x$ is a vector of decision variables, and the functions $f_0$, $f_i$ and $h_i$, respectively, are the cost, inequality constraints, and equality constraints. However, such problems can be very hard to solve in general, especially when the number of decision variables in $x$ is large. There are several reasons for this difficulty: First, the problem "terrain" may be riddled with local optima. Second, it might be very hard to find a feasible point (*i.e.*, an $x$ which satisfy all equalities and inequalities), in fact, the feasible set which needn't even be fully connected, could be empty. Third, stopping criteria used in general optimization algorithms are often arbitrary. Forth, optimization algorithms might have very poor convergence rates. Fifth, numerical problems could

cause the minimization algorithm to stop all together or wander.

It has been known for a long time [19], [3], [16], [13] that if the $f_i$ are all *convex*, and the $h_i$ are affine, then the first three problems disappear: any local optimum is, in fact, a global optimum; feasibility of convex optimization problems can be determined unambiguously, at least in principle; and very precise stopping criteria are available using *duality*. However, convergence rate and numerical sensitivity issues still remain a potential problem.

It was not until the late '80's and '90's that researchers in the former Soviet Union and United States discovered that if, in addition to convexity, the $f_i$ satisfied a property known as *self-concordance*, then issues of convergence and numerical sensitivity could be avoided using *interior point methods* [18], [23], [29], [30], [25]. The self-concordance property is satisfied by a very large set of important functions used in engineering. Hence, it is now possible to solve a large class of convex optimization problems in engineering with great efficiency.

## II. CONVEX SETS

In this section we list some important convex sets and operations. It is important to note that some of these sets have different representations. Picking the right representation can make the difference between a tractable problem and an intractable one.

We will be concerned only with optimization problems whose decision variables are vectors in $\mathbf{R}^n$ or matrices in $\mathbf{R}^{m \times n}$. Throughout the paper, we will make frequent use of informal sketches to help the reader develop an intuition for the geometry of convex optimization.

A function $f : \mathbf{R}^n \to \mathbf{R}^m$ is *affine* if it has the form linear plus constant $f(x) = Ax + b$. If $F$ is a matrix valued function, *i.e.*, $F : \mathbf{R}^n \to \mathbf{R}^{p \times q}$, then $F$ is affine if it has the form

$$F(x) = A_0 + x_1 A_1 + \cdots + x_n A_n$$

where $A_i \in \mathbf{R}^{p \times q}$. Affine functions are sometimes loosely refered to as linear.

Recall that $S \subseteq \mathbf{R}^n$ is a *subspace* if it contains the plane through any two of its points and the origin, *i.e.*,

$$x, y \in S, \quad \lambda, \mu \in \mathbf{R} \implies \lambda x + \mu y \in S.$$

Two common representations of a subspace are as the range of a matrix

$$
\begin{aligned}
\text{range}(A) &= \{Aw \mid w \in \mathbf{R}^q\} \\
&= \{w_1 a_1 + \cdots + w_q a_q \mid w_i \in \mathbf{R}\}
\end{aligned}
$$

where $A = [\begin{array}{ccc} a_1 & \cdots & a_q \end{array}]$; or as the nullspace of a matrix

$$
\begin{aligned}
\text{nullspace}(B) &= \{x \mid Bx = 0\} \\
&= \{x \mid b_1^T x = 0, \ldots, b_p^T x = 0\}
\end{aligned}
$$

where $B = [\begin{array}{ccc} b_1 & \cdots & b_p \end{array}]^T$. As an example, let $\mathbf{S}^n = \{X \in \mathbf{R}^{n \times n} \mid X = X^T\}$ denote the set of symmetric matrices. Then $\mathbf{S}^n$ is a subspace, since symmetric matrices are closed under addition. Another way to see this is to note that $\mathbf{S}^n$ can be written as $\{X \in \mathbf{R}^{n \times n} \mid X_{ij} = X_{ji}, \forall i, j\}$ which is the nullspace of linear function $X - X^T$.

A set $S \subseteq \mathbf{R}^n$ is *affine* if it contains line through any two points in it, *i.e.*,

$$x, y \in S, \quad \lambda, \mu \in \mathbf{R}, \quad \lambda + \mu = 1 \implies \lambda x + \mu y \in S.$$
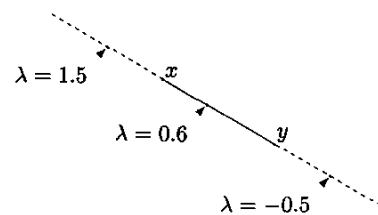


Geometrically, an affine set is simply a subspace which is not necessarily centered at the origin. Two common representations for of an affine set are: the range of affine function

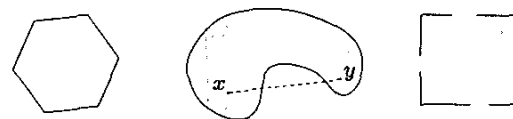$$S = \{Az + b \mid z \in \mathbf{R}^q\},$$

or as the solution of a set of linear equalities:

$$
\begin{aligned}
S &= \{x \mid b_1^T x = d_1, \ldots, b_p^T x = d_p\} \\
&= \{x \mid Bx = d\}.
\end{aligned}
$$

A set $S \subseteq \mathbf{R}^n$ is a *convex set* if it contains the line segment joining any of its points, *i.e.*,

$$x, y \in S, \quad \lambda, \mu \geq 0, \quad \lambda + \mu = 1 \implies \lambda x + \mu y \in S$$

convex          not convex



Geometrically, we can think of convex sets as always bulging outward, with no dents or kinks in them. Clearly subspaces and affine sets are convex, since their definitions subsume convexity.

A set $S \subseteq \mathbf{R}^n$ is a *convex cone* if it contains all rays passing through its points which emanate from the

origin, *as well as* all line segments joining any points on those rays, *i.e.*,

$$x, y \in S, \quad \lambda, \mu \geq 0, \quad \Longrightarrow \lambda x + \mu y \in S$$

Geometrically, $x, y \in S$ means that $S$ contains the entire 'pie slice' between $x$, and $y$.
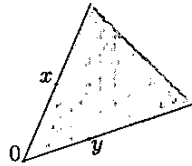


The nonnegative orthant, $\mathbf{R}_+^n$ is a convex cone. The set $\mathbf{S}_+^n = \{X \in \mathbf{S}^n \mid X \succeq 0\}$ of symmetric positive semidefinite (PSD) matrices is also a convex cone, since any positive combination of semidefinite matrices is semidefinite. Hence we call $\mathbf{S}_+^n$ the *positive semidefinite cone*.

A convex cone $K \subseteq \mathbf{R}^n$ is said to be *proper* if it is closed, has nonempty interior, and is *pointed*, *i.e.*, there is no line in $K$. A proper cone $K$ defines a *generalized inequality* $\preceq_K$ in $\mathbf{R}^n$:

$$x \preceq_K y \quad \Longleftrightarrow \quad y - x \in K$$

(strict version $x \prec_K y \quad \Longleftrightarrow \quad y - x \in \mathrm{int}\, K$).
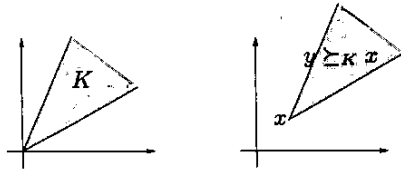


This formalizes our use of the $\succeq$ symbol:

- $K = \mathbf{R}_+^n$: $x \preceq_K y$ means $x_i \leq y_i$
  (componentwise vector inequality)
- $K = \mathbf{S}_+^n$: $X \preceq_K Y$ means $Y - X$ is PSD

These are so common we drop the $K$.

Given points $x_i \in \mathbf{R}^n$ and $\theta_i \in \mathbf{R}$, then $y = \theta_1 x_1 + \cdots + \theta_k x_k$ is said to be a

- *linear combination* for any real $\theta_i$
- *affine combination* if $\sum_i \theta_i = 1$
- *convex combination* if $\sum_i \theta_i = 1$, $\theta_i \geq 0$
- *conic combination* if $\theta_i \geq 0$

The linear (resp. affine, convex, conic) *hull* of a set $S$ is the set of all linear (resp. affine, convex, conic) combinations of points from $S$, and is denoted by $\mathrm{span}(S)$ (resp. $\mathbf{Aff}(S)$, $\mathbf{Co}(S)$, $\mathbf{Cone}(S)$). It can be shown that this is the smallest such set containing $S$.

As an example, consider the set $S = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$. Then $\mathrm{span}(S)$ is $\mathbf{R}^3$; $\mathbf{Aff}(S)$ is the hyperplane passing through the three

points; $\mathbf{Co}(S)$ is the *unit simplex* which is the triangle joining the vectors along with all the points inside it; $\mathbf{Cone}(S)$ is the nonnegative orthant $\mathbf{R}_+^3$.

Recall that a *hyperplane*, represented as $\{x \mid a^T x = b\}$ ($a \neq 0$), is in general an affine set, and is a subspace if $b = 0$. Another useful representation of a hyperplane is given by $\{x \mid a^T(x - x_0) = 0\}$, where $a$ is *normal vector*; $x_0$ lies on hyperplane. Hyperplanes are convex, since they contain all lines (and hence segments) joining any of their points.

A *halfspace*, described as $\{x \mid a^T x \leq b\}$ ($a \neq 0$) is generally convex and is a convex cone if $b = 0$. Another useful representation is $\{x \mid a^T(x - x_0) \leq 0\}$, where $a$ is *(outward) normal vector* and $x_0$ lies on boundary. Halfspaces are convex.
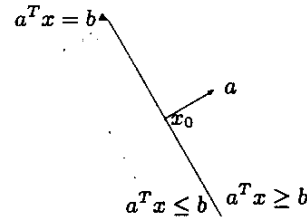


We now come to a very important fact about properties which are preserved under intersection: Let $\mathcal{A}$ be an arbitrary index set (possibly uncountably infinite) and $\{S_\alpha \mid \alpha \in \mathcal{A}\}$ a collection of sets , then we have the following:

$$S_\alpha \text{ is } \begin{pmatrix} \text{subspace} \\ \text{affine} \\ \text{convex} \\ \text{convex cone} \end{pmatrix} \Longrightarrow \bigcap_{\alpha \in \mathcal{A}} S_\alpha \text{ is } \begin{pmatrix} \text{subspace} \\ \text{affine} \\ \text{convex} \\ \text{convex cone} \end{pmatrix}$$

In fact, every closed convex set $S$ is the (usually infinite) intersection of halfspaces which contain it, *i.e.*,

$S = \bigcap \{\mathcal{H} \mid \mathcal{H} \text{ halfspace}, S \subseteq \mathcal{H}\}$. For example, another way to see that $\mathbf{S}_+^n$ is a convex cone is to recall that a matrix $X \in \mathbf{S}^n$ is positive semidefinite if $z^T X z \geq 0$, $\forall z \in \mathbf{R}^n$. Thus we can write

$$\mathbf{S}_+^n = \bigcap_{z \in \mathbf{R}^n} \left\{ X \in \mathbf{S}^n \;\middle|\; z^T X z = \sum_{i,j=1}^n z_i z_j X_{ij} \geq 0 \right\}.$$
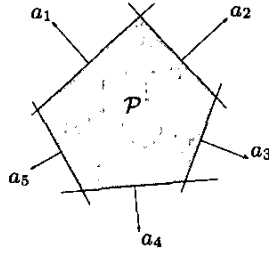
Now observe that the summation above is actually *linear* in the components of $X$, so $\mathbf{S}_+^n$ is the infinite intersection of halfspaces containing the origin (which are convex cones) in $\mathbf{S}^n$.

We continue with our listing of useful convex sets.

A *polyhedron* is intersection of a finite number of halfspaces

$$\mathcal{P} = \{x \mid a_i^T x \leq b_i, \; i = 1, \ldots, k\}$$
$$= \{x \mid Ax \preceq b\}$$

where $\preceq$ above means componentwise inequality.



A bounded polyhedron is called a *polytope*, which also has the alternative representation $\mathcal{P} = \mathbf{Co}\{v_1, \ldots, v_N\}$, where $\{v_1, \ldots, v_N\}$ are its vertices. For example, the nonnegative orthant $\mathbf{R}_+^n = \{x \in \mathbf{R}^n \mid x \succeq 0\}$ is a polyhedron, while the *probability simplex* $\{x \in \mathbf{R}^n \mid x \succeq 0, \sum_i x_i = 1\}$ is a polytope.

If $f$ is a norm, then the *norm ball* $B = \{x \mid f(x - x_c) \leq 1\}$ is convex, and the *norm cone* $C = \{(x, t) \mid f(x) \leq t\}$ is a convex cone. Perhaps the most familiar norms are the $\ell_p$ norms on $\mathbf{R}^n$:

$$\|x\|_p = \begin{cases} \left(\sum_i |x_i|^p\right)^{1/p} & ; p \geq 1, \\ \max_i |x_i| & ; p = \infty \end{cases}$$

The corresponding norm balls (in $\mathbf{R}^2$) look like this:
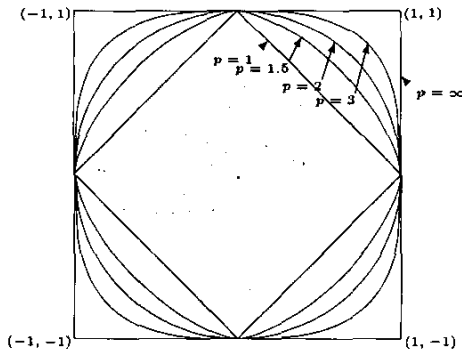


Another workhorse of convex optimization is the *ellipsoid*. In addition to their computational convenience, ellipsoids can be used as crude models or approximates for other convex sets. In fact, it can be shown that any bounded convex set in $\mathbf{R}^n$ can be approximated to within a factor of $n$ by an ellipsoid.

There are many representations for ellipsoids. For example, we can use

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

where the parameters are $A = A^T \succ 0$, the center $x_c \in \mathbf{R}^n$.
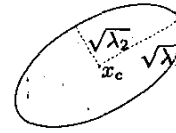


In this case, the semiaxis lengths are $\sqrt{\lambda_i}$, where $\lambda_i$ eigenvalues of $A$; the semiaxis directions are eigenvectors of $A$ and the volume is proportional to $(\det A)^{1/2}$. However, depending on the problem, other descriptions might be more appropriate, such as ($\|u\| = \sqrt{u^T u}$)

- image of unit ball under affine transformation: $\mathcal{E} = \{Bu + x_c \mid \|u\| \leq 1\}$; vol. $\propto \det B$
- preimage of unit ball under affine transformation: $\mathcal{E} = \{x \mid \|Ax - b\| \leq 1\}$; vol. $\propto \det A^{-1}$
- sublevel set of nondegenerate quadratic: $\mathcal{E} = \{x \mid f(x) \leq 0\}$ where $f(x) = x^T C x + 2d^T x + e$ with $C = C^T \succ 0$, $e - d^T C^{-1} d < 0$; vol. $\propto (\det A)^{-1/2}$

It is an instructive exercise to convert among representations.

The *second-order cone* is the norm cone associated with the Euclidean norm.

$$\begin{aligned} S &= \{(x, t) \mid \sqrt{x^T x} \leq t\} \\ &= \left\{(x, t) \,\middle|\, \begin{bmatrix} x \\ t \end{bmatrix}^T \begin{bmatrix} I & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ t \end{bmatrix} \leq 0, \, t \geq 0 \right\} \end{aligned}$$
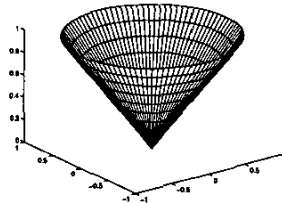


The image (or preimage) under an *affinte transformation* of a convex set are convex, *i.e.*, if $S$, $T$ convex, then so are the sets

$$\begin{aligned} f^{-1}(S) &= \{x \mid Ax + b \in S\} \\ f(T) &= \{Ax + b \mid x \in T\} \end{aligned}$$

An example is coordinate projection $\{x \mid (x, y) \in S$ for some $y\}$. As another example, a constraint of the form

$$\|Ax + b\|_2 \leq c^T x + d,$$

where $A \in \mathbf{R}^{k \times n}$, a *second-order cone constraint*, since it is the same as requiring the affine function $(Ax + b, c^T x + d)$ to lie in the second-order cone in $\mathbf{R}^{k+1}$. Similarly, if $A_0, A_1, \ldots, A_m \in \mathbf{S}^n$, solution set of the *linear matrix inequality* (LMI)

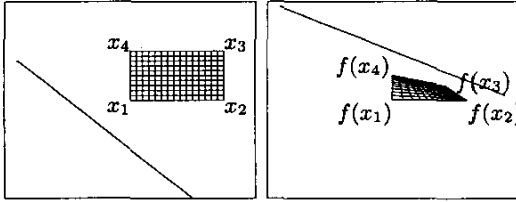$$F(x) = A_0 + x_1 A_1 + \cdots + x_m A_m \succeq 0$$

is convex (preimage of the semidefinite cone under an affine function).

A *linear-fractional* (or projective) function $f : \mathbf{R}^m \to \mathbf{R}^n$ has the form

$$f(x) = \frac{Ax + b}{c^T x + d}$$

and domain $\mathbf{dom}\, f = \mathcal{H} = \{x \mid c^T x + d > 0\}$. If $C$ is a convex set, then its *linear-fractional transformation* $f(C)$ is also convex. This is because linear fractional transformations preserve line segments: for $x, y \in \mathcal{H}$,

$$f([x, y]) = [f(x), f(y)]$$

Two further properties are helpful in visualizing the geometry of convex sets. The first is the *separating hyperplane theorem*, which states that if $S$, $T \subseteq \mathbf{R}^n$ are convex and disjoint $(S \cap T = \emptyset)$, then there exists a hyperplane $\{x \mid a^T x - b = 0\}$ which *separates* them.
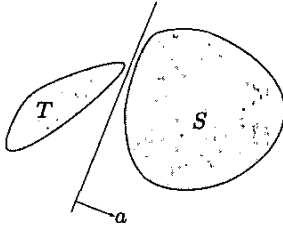
The second property is due to the *supporting hyperplane theorem* which states that there exists a supporting hyperplan at every point on the boundary of a convex set, where a supporting hyperplane $\{x \mid a^T x = a^T x_0\}$ *supports* $S$ at $x_0 \in \partial S$ if

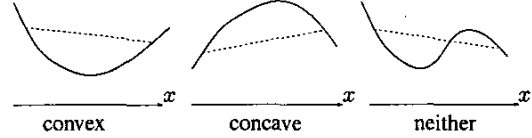$$x \in S \Rightarrow a^T x \leq a^T x_0$$
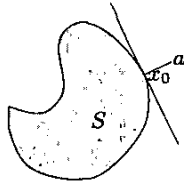
## III. CONVEX FUNCTIONS

In this section, we introduce the reader to some important convex functions and techniques for verifying convexity. The objective is to sharpen the reader's ability to recognize convexity.

### A. Convex functions

A function $f : \mathbf{R}^n \to \mathbf{R}$ is *convex* if its domain $\mathbf{dom}\, f$ is convex and for all $x, y \in \mathbf{dom}\, f$, $\theta \in [0, 1]$

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

$f$ is *concave* if $-f$ is convex.

Here are some simple examples on $\mathbf{R}$) are: $x^2$ is convex $(\mathbf{dom}\, f = \mathbf{R})$; $\log x$ is concave $(\mathbf{dom}\, f = \mathbf{R}_{++})$; and $f(x) = 1/x$ is convex $(\mathbf{dom}\, f = \mathbf{R}_{++})$.
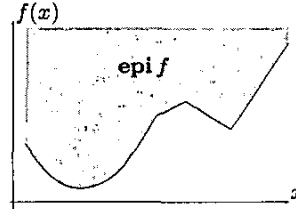
It is convenient to define the *extension* of a convex function $f$

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

Note that $f$ still satisfies the basic definition on for all $x, y \in \mathbf{R}^n$, $0 \leq \theta \leq 1$ (as an inequality in $\mathbf{R} \cup \{+\infty\}$). We will use same symbol for $f$ and its extension, *i.e.*, we will implicitly assume convex functions are extended.

The *epigraph* of a function $f$ is

$$\mathbf{epi}\, f = \{(x, t) \mid x \in \mathbf{dom}\, f, \; f(x) \leq t \}$$

The $(\alpha$-$)$*sublevel set* of a function $f$ is

$$S_\alpha \triangleq \{x \in \mathbf{dom}\, f \mid f(x) \leq \alpha\}$$

Form the basic definition of convexity, it follows that if $f$ is a convex function if and only if its epigraph $\mathbf{epi}\, f$ is a convex set. It also follows that if $f$ is convex, then its sublevel sets $S_\alpha$ are convex (the converse is false - see quasiconvex functions later).

The convexity of a differentiable function $f : \mathbf{R}^n \to \mathbf{R}$ can also be characterized by conditions on its gradient $\nabla f$ and Hessian $\nabla^2 f$. Recall that, in general, the gradient yields a first order Taylor approximation at $x_0$:

$$f(x) \simeq f(x_0) + \nabla f(x_0)^T (x - x_0)$$

We have the following *first-order condition*: $f$ is convex if and only if for all $x, x_0 \in \mathbf{dom}\, f$,

$$f(x) \geq f(x_0) + \nabla f(x_0)^T (x - x_0),$$

*i.e.*, the first order approximation of $f$ is a *global underestimator*.
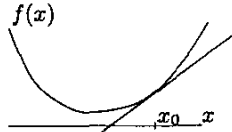


$f(x_0) + \nabla f(x_0)^T (x - x_0)$

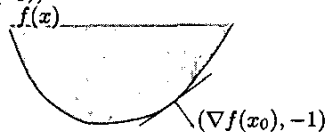To see why this is so, we can rewrite the condition above in terms of the epigraph of $f$ as: for all $(x, t) \in \text{epi} f$,

$$\begin{bmatrix} \nabla f(x_0) \\ -1 \end{bmatrix}^T \begin{bmatrix} x - x_0 \\ t - f(x_0) \end{bmatrix} \leq 0,$$

*i.e.*, $(\nabla f(x_0), -1)$ defines supporting hyperplane to epi $f$ at $(x_0, f(x_0))$



Recall that the Hessian of $f$, $\nabla^2 f$, yields a second order Taylor series expansion around $x_0$:

$$f(x) \simeq f(x_0) + \nabla f(x_0)^T (x - x_0) + \frac{1}{2}(x - x_0)^T \nabla^2 f(x_0)(x - x_0)$$

We have the following necessary and sufficient *second order condition*: a twice differentiable function $f$ is convex if and only if for all $x \in \text{dom} f$, $\nabla^2 f(x) \succeq 0$, *i.e.*, its Hessian is positive semidefinite on its domain.

The convexity of the following functions is easy to verify, using the first and second order characterizations of convexity:

- $x^\alpha$ is convex on $\mathbf{R}_{++}$ for $\alpha \geq 1$;
- $x \log x$ is convex on $\mathbf{R}_+$;
- *log-sum-exp function* $f(x) = \log \sum_i e^{x_i}$ (tricky!)
- *affine functions* $f(x) = a^T x + b$ where $a \in \mathbf{R}^n, b \in \mathbf{R}$ are convex and concave since $\nabla^2 f \equiv 0$.
- *quadratic functions* $f(x) = x^T P x + 2q^T x + r$ ($P = P^T$) whose Hessian $\nabla^2 f(x) = 2P$ are convex $\Longleftrightarrow$ $P \succeq 0$; concave $\Longleftrightarrow P \preceq 0$

Further *elementary properties* are extremely helpful in verifying convexity. We list several:

- $f : \mathbf{R}^n \to \mathbf{R}$ is convex iff it is convex on all lines: $\tilde{f}(t) \triangleq f(x_0 + th)$ is convex in $t \in \mathbf{R}$, for all $x_0, h \in \mathbf{R}^n$
- nonnegative sums of convex functions are convex: $\alpha_1, \alpha_2 \geq 0$ and $f_1, f_2$ convex $\Longrightarrow \alpha_1 f_1 + \alpha_2 f_2$ convex
- nonnegative infinite sums, integrals: $p(y) \geq 0$, $g(x, y)$ convex in $x$ $\Longrightarrow \int p(y) g(x, y) dy$ convex

- pointwise supremum (maximum): $f_\alpha$ convex $\Longrightarrow \sup_{\alpha \in A} f_\alpha$ convex (corresponds to intersection of epigraphs)
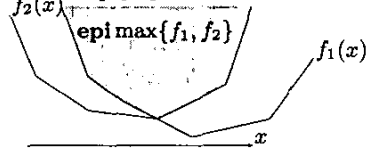


- affine transformation of domain: $f$ convex $\Longrightarrow f(Ax + b)$ convex

The following are important examples of how these further properties can be applied:

- *piecewise-linear functions*: $f(x) = \max_i \{a_i^T x + b_i\}$ is convex in $x$ (epi $f$ is polyhedron)
- *maximum distance* to any set, $\sup_{s \in S} \|x - s\|$, is convex in $x$ [$(x - s)$ is affine in $x$, $\| \cdot \|$ is a norm, so $f_s(x) = \|x - s\|$ is convex].
- *expected value*: $f(x, u)$ convex in $x$ $\Longrightarrow g(x) = \mathbf{E}_u f(x, u)$ convex
- $f(x) = x_{[1]} + x_{[2]} + x_{[3]}$ is convex on $\mathbf{R}^n$ ($x_{[i]}$ is the $i$th largest $x_j$). To see this, note that $f(x)$ is the sum of the largest triple of components of $x$. So $f$ can be written as $f(x) = \max_i c_i^T x$, where $c_i$ are the set of all vectors with components zero except for three 1's.
- $f(x) = \sum_{i=1}^m \log(b_i - a_i^T x)^{-1}$ is convex (dom $f = \{x \mid a_i^T x < b_i, i = 1, \ldots, m\}$)

Another convexity preserving operation is that of *minimizing over some variables*. Specifically, if $h(x, y)$ is convex in $x$ and $y$, then

$$f(x) = \inf_y h(x, y)$$

is convex in $x$. This is because the operation above corresponds to projection of the epigraph, $(x, y, t) \to (x, t)$.
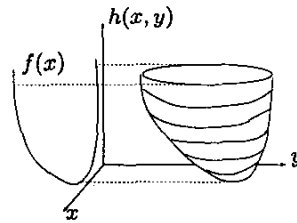


An important example of this is the *minimum distance function* to a set $S \subseteq \mathbf{R}^n$, defined as:

$$\text{dist}(x, S) = \inf_{y \in S} \|x - y\|$$
$$= \inf_y \|x - y\| + \delta(y|S)$$

where $\delta(y|S)$ is the *indicator function* of the set $S$, which is infinite everywhere except on $S$, where it is zero. Note that in contrast to the maximum distance function defined earlier, $\text{dist}(x, S)$ is *not* convex in

general. However, if $S$ is convex, then $\text{dist}(x, S)$ is convex since $\|x - y\| + \delta(y|S)$ is convex in $(x, y)$.

The technique of *composition* can also be used to deduce the convexity of differentiable functions, by means of the chain rule. For example, one can show results like: $f(x) = \log \sum_i \exp g_i(x)$ is convex if each $g_i$ is; see [7].

*Jensen's inequality*, a classical result which essentially a restatement of convexity, is used extensively in various forms. If $f : \mathbf{R}^n \to \mathbf{R}$ is convex

- two points: $\theta_1 + \theta_2 = 1$, $\theta_i \geq 0 \implies$
  $f(\theta_1 x_1 + \theta_2 x_2) \leq \theta_1 f(x_1) + \theta_2 f(x_2)$
- more than two points: $\sum_i \theta_i = 1$, $\theta_i \geq 0 \implies$
  $f(\sum_i \theta_i x_i) \leq \sum_i \theta_i f(x_i)$
- continuous version: $p(x) \geq 0$, $\int p(x)\, dx = 1 \implies$
  $f(\int x p(x)\, dx) \leq \int f(x) p(x)\, dx$
  or more generally, $f(\mathbf{E}\, x) \leq \mathbf{E}\, f(x)$.

The simple techniques above can also be used to verify the convexity of *functions of matrices*, which arise in many important problems.

- $\text{Tr}\, A^T X = \sum_{i,j} A_{ij} X_{ij}$ is linear in $X$ on $\mathbf{R}^{n \times n}$
- $\log \det X^{-1}$ is convex on $\{X \in \mathbf{S}^n \mid X \succ 0\}$
  *Proof:* Verify convexity along lines. Let $\lambda_i$ be the eigenvalues of $X_0^{-1/2} H X_0^{-1/2}$
  $$\begin{aligned} f(t) &\triangleq \log \det(X_0 + tH)^{-1} \\ &= \log \det X_0^{-1} + \log \det(I + t X_0^{-1/2} H X_0^{-1/2})^{-1} \\ &= \log \det X_0^{-1} - \sum_i \log(1 + t\lambda_i) \end{aligned}$$
  which is a convex function of $t$.
- $(\det X)^{1/n}$ is concave on $\{X \in \mathbf{S}^n \mid X \succ 0\}$
- $\lambda_{\max}(X)$ is convex on $\mathbf{S}^n$ since, for $X$ symmetric, $\lambda_{\max}(X) = \sup_{\|y\|_2 = 1} y^T X y$, which is a supremum of linear functions of $X$.
- $\|X\|_2 = \sigma_1(X) = (\lambda_{\max}(X^T X))^{1/2}$ is convex on $\mathbf{R}^{m \times n}$ since, by definition, $\|X\|_2 = \sup_{\|u\|_2 = 1} \|Xu\|_2$.

The *Schur complement* technique is an indispensible tool for modeling and transforming nonlinear constraints into convex LMI constraints. Suppose a symmetric matrix $X$ can be decomposed as

$$X = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$$

with $A$ invertible. The matrix $S \triangleq C - B^T A^{-1} B$ is called the Schur complement of $A$ in $X$, and has the following important attributes, which are not too hard to prove:

- $X \succ 0$ if and only if $A \succ 0$ and $S \triangleq C - B^T A^{-1} B \succ 0$
- if $A \succ 0$, then $X \succeq 0$ if and only if $S \succeq 0$

For example, the following second-order cone constraint on $x$

$$\|Ax + b\| \leq e^T x + d$$

is equivalent to LMI

$$\begin{bmatrix} (e^T x + d)I & Ax + b \\ (Ax + b)^T & e^T x + d \end{bmatrix} \succeq 0.$$

This shows that all sets that can be modeled by an SOCP constraint can be modeled as LMI's. Hence LMI's are more "expressive". As another nontrivial example, the quadratic-over-linear function

$$f(x, y) = x^2/y, \quad \text{dom}\, f = \mathbf{R} \times \mathbf{R}_{++}$$

is convex because its epigraph $\{(x, y, t) \mid y > 0, x^2/y \leq t\}$ is convex: by Schur complememts, it's equivalent to the solution set of the LMI constraints

$$\begin{bmatrix} y & x \\ x & t \end{bmatrix} \succeq 0, \quad y > 0.$$

The concept of $K$-*convexity* generalizes convexity to vector- or matrix-valued functions. As mentioned earlier, a pointed convex cone $K \subseteq \mathbf{R}^m$ induces generalized inequality $\preceq_K$. A function $f : \mathbf{R}^n \to \mathbf{R}^m$ is $K$-convex if for all $\theta \in [0, 1]$

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

In many problems, especially in statistics, *log-concave functions* are frequently encountered. A function $f :$ $\mathbf{R}^n \to \mathbf{R}_+$ is log-concave (log-convex) if $\log f$ is concave (convex). Hence one can transform problems in log convex functions into convex optimization problems. As an example, the normal density, $f(x) = \text{const} \cdot e^{-(1/2)(x - x_0)^T \Sigma^{-1}(x - x_0)}$ is log-concave.

## B. Quasiconvex functions

The next best thing to a convex function in optimization is a quasiconvex function, since it can be minimized in a few iterations of convex optimization problems.

A function $f : \mathbf{R}^n \to \mathbf{R}$ is *quasiconvex* if every sublevel set $S_\alpha = \{x \in \text{dom}\, f \mid f(x) \leq \alpha\}$ is convex. Note that if $f$ is convex, then it is automatically quasiconvex. Quasiconvex functions can have 'locally flat' regions.
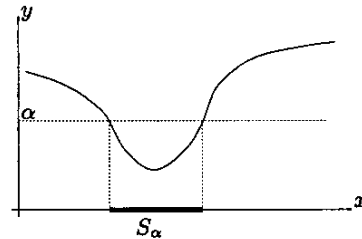
We say $f$ is *quasiconcave* if $-f$ is quasiconvex, *i.e.*, superlevel sets $\{x \mid f(x) \geq \alpha\}$ are convex. A function which is both quasiconvex and quasiconcave is called *quasilinear*.

We list some examples:

- $f(x) = \sqrt{|x|}$ is quasiconvex on $\mathbf{R}$
- $f(x) = \log x$ is quasilinear on $\mathbf{R}_+$
- linear fractional function,

$$f(x) = \frac{a^T x + b}{c^T x + d}$$

is quasilinear on the halfspace $c^T x + d > 0$
- $f(x) = \frac{\|x - a\|_2}{\|x - b\|_2}$ is quasiconvex on the halfspace $\{x \mid \|x - a\|_2 \leq \|x - b\|_2\}$

Quasiconvex functions have a lot of similar features to convex functions, but also some important differences. The following *quasiconvex properties* are helpful in verifying quasiconvexity:
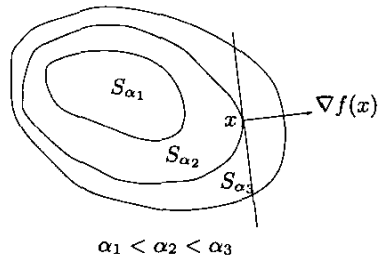
- $f$ is quasiconvex if and only if it is quasiconvex on lines, *i.e.*, $f(x_0 + th)$ quasiconvex in $t$ for all $x_0, h$
- modified Jensen's inequality: $f$ is quasiconvex iff for all $x, y \in \mathbf{dom} f$, $\theta \in [0, 1]$,

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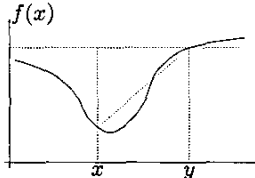


- for $f$ differentiable, $f$ quasiconvex $\iff$ for all $x, y \in \mathbf{dom} f$

$$f(y) \leq f(x) \Rightarrow (y - x)^T \nabla f(x) \leq 0$$



$\alpha_1 < \alpha_2 < \alpha_3$

- positive multiples
  $f$ quasiconvex, $\alpha \geq 0 \Longrightarrow \alpha f$ quasiconvex
- pointwise supremum $f_1, f_2$ quasiconvex $\Longrightarrow$ $\max\{f_1, f_2\}$ quasiconvex
  (extends to supremum over arbitrary set)
- affine transformation of domain
  $f$ quasiconvex $\Longrightarrow f(Ax + b)$ quasiconvex

- linear-fractional transformation of domain
  $f$ quasiconvex $\Longrightarrow f\left(\frac{Ax + b}{c^T x + d}\right)$ quasiconvex
  on $c^T x + d > 0$
- composition with monotone increasing function:
  $f$ quasiconvex, $g$ monotone increasing
  $\Longrightarrow g(f(x))$ quasiconvex
- sums of quasiconvex functions are **not** quasiconvex in general
- $f$ quasiconvex in $x$, $y \Longrightarrow g(x) = \inf_y f(x, y)$ quasiconvex in $x$ (projection of epigraph remains quasiconvex)

## IV. Convex optimization problems

In this section, we will discuss the formulation of optimization problems at a general level. In the next section we will focus on useful optimization problems with specific structure.

Consider the following optimization in *standard form*

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

where $f_i, h_i : \mathbf{R}^n \to \mathbf{R}$; $x$ is the *optimization variable*; $f_0$ is the *objective* or *cost function*; $f_i(x) \leq 0$ are the *inequality constraints*; $h_i(x) = 0$ are the *equality constraints*. Geometrically, this problem corresponds to the minimization of $f_0$, over a set described by as the intersection of 0-sublevel sets of $f_i, i = 1, \ldots, m$ with surfaces described by the 0-solution sets of $h_i$.

A point $x$ is *feasible* if it satisfies the constraints; the *feasible set* $C$ is the set of all feasible points; and the problem is *feasible* if there are feasible points. The problem is said to be *unconstrained* if $m = p = 0$ The *optimal value* is denoted by $f^* = \inf_{x \in C} f_0(x)$, and we adopt the convention that $f^* = +\infty$ if the problem is infeasible). A point $x \in C$ is an *optimal point* if $f(x) = f^*$ and the *optimal set* is $X_{\mathrm{opt}} = \{x \in C \mid f(x) = f^*\}$.

As an example consider the problem

$$\begin{array}{ll} \text{minimize} & x_1 + x_2 \\ \text{subject to} & -x_1 \leq 0 \\ & -x_2 \leq 0 \\ & 1 - \sqrt{x_1 x_2} \leq 0 \end{array}$$
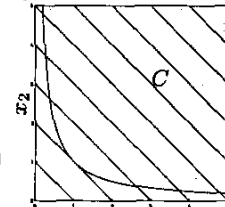


The objective function is $f_0(x) = [1 \ 1]^T x$; the feasible set $C$ is half-hyperboloid; the optimal value is $f^* = 2$; and the only optimal point is $x^* = (1, 1)$.

In the standard problem above, the *explicit* constraints are given by $f_i(x) \leq 0$, $h_i(x) = 0$. However, there are also the *implicit* constraints: $x \in \mathbf{dom} f_i$, $x \in \mathbf{dom} h_i$,

*i.e.*, $x$ must lie in the set

$$D = \mathbf{dom}\, f_0 \cap \cdots \cap \mathbf{dom}\, f_m \cap \mathbf{dom}\, h_1 \cap \cdots \cap \mathbf{dom}\, h_p$$

which is called the *domain of the problem*. For example,

$$\begin{array}{ll}\text{minimize} & -\log x_1 - \log x_2 \\ \text{subject to} & x_1 + x_2 - 1 \leq 0\end{array}$$

has the implicit constraint $x \in D = \{x \in \mathbf{R}^2 \mid x_1 > 0, x_2 > 0\}$.

A *feasibility problem* is a special case of the standard problem, where we are interested merely in finding any feasible point. Thus, problem is really to

- either find $x \in C$
- or determine that $C = \emptyset$ .

Equivalently, the feasibility problem requires that we either solve the inequality / equality system

$$\begin{array}{ll} f_i(x) \leq 0, & i = 1, \ldots, m \\ h_i(x) = 0, & i = 1, \ldots, p \end{array}$$

or determine that it is inconsistent.

An optimization problem in standard form is a *convex optimization problem* if $f_0, f_1, \ldots, f_m$ are all convex, and $h_i$ are all affine:

$$\begin{array}{ll}\text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & a_i^T x - b_i = 0, \quad i = 1, \ldots, p.\end{array}$$

This is often written as

$$\begin{array}{ll}\text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b\end{array}$$

where $A \in \mathbf{R}^{p \times n}$ and $b \in \mathbf{R}^p$. As mentioned in the introduction, convex optimization problems have three crucial properties that makes them fundamentally more tractable than generic nonconvex optimization problems:

1) *no local minima*: any local optimum is necessarily a global optimum;

2) *exact infeasibility detection*: using duality theory (which is not cover here), hence algorithms are easy to initialize;

3) *efficient numerical solution methods* that can handle very large problems.

Note that often seemingly 'slight' modifications of convex problem can be very hard. Examples include:

- convex maximization, concave minimization, *e.g.*

$$\begin{array}{ll}\text{maximize} & \|x\| \\ \text{subject to} & Ax \preceq b\end{array}$$

- nonlinear equality constraints, *e.g.*

$$\begin{array}{ll}\text{minimize} & c^T x \\ \text{subject to} & x^T P_i x + q_i^T x + r_i = 0, \quad i = 1, \ldots, K\end{array}$$

- minimizing over non-convex sets, *e.g.*, Boolean variables

$$\begin{array}{ll}\text{find} & x \\ \text{such that} & Ax \preceq b, \\ & x_i \in \{0, 1\}\end{array}$$

To understand global optimality in convex problems, recall that $x \in C$ is *locally optimal* if it satisfies

$$y \in C, \ \|y - x\| \leq R \implies f_0(y) \geq f_0(x)$$

for some $R > 0$. A point $x \in C$ is *globally optimal* means that

$$y \in C \implies f_0(y) \geq f_0(x).$$

For convex optimization problems, any local solution is also global. [*Proof sketch:* Suppose $x$ is locally optimal, but that there is a $y \in C$, with $f_0(y) < f_0(x)$. Then we may take small step from $x$ towards $y$, *i.e.*, $z = \lambda y + (1 - \lambda)x$ with $\lambda > 0$ small. Then $z$ is near $x$, with $f_0(z) < f_0(x)$ which contradicts local optimality.]

There is also a *first order condition* that characterizes optimality in convex optimization problems. Suppose $f_0$ is differentiable, then $x \in C$ is optimal iff

$$y \in C \implies \nabla f_0(x)^T (y - x) \geq 0$$

So $-\nabla f_0(x)$ defines supporting hyperplane for $C$ at $x$. This means that if we move from $x$ towards any other feasible $y$, $f_0$ does not decrease.

contour lines of $f_0$
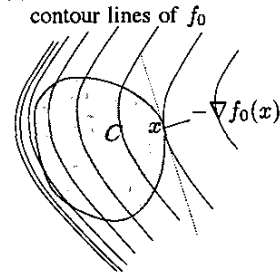


Many standard convex optimization algorithms assume a linear objective function. So it is important to realize that any convex optimization problems can be converted to one with a linear objective function. This is called putting the problem into *epigraph form*. It involves rewriting the problem as:

$$\begin{array}{ll}\text{minimize} & t \\ \text{subject to} & f_0(x) - t \leq 0, \\ & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p\end{array}$$

where the variables are now $(x, t)$ and the objective has essentially be moved into the inequality constraints. Observe that the new objective is *linear*: $t = e_{n+1}^T(x, t)$

$(e_{n+1}$ is the vector of all zeros except for a 1 in its $(n + 1)th$ component). Minimizing $t$ will result in minimizing $f_0$ with respect to $x$, so any optimal $x^\star$ of the new problem with be optimal for the epigraph form and vice versa.
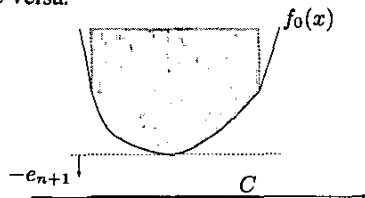


So the linear objective is 'universal' for convex optimization.

The above trick of introducing the extra variable $t$ is known as the method of *slack variables*. It is very helpful in transforming problems into canonical form. We will see another example in the next section.

A convex optimization problem in standard form *with generalized inequalities*:

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \preceq_{K_i} 0, \quad i = 1, \ldots, L \\ & Ax = b \end{array}$$

where $f_0 : \mathbf{R}^n \to \mathbf{R}$ are all convex; the $\preceq_{K_i}$ are generalized inequalities on $\mathbf{R}^{m_i}$; and $f_i : \mathbf{R}^n \to \mathbf{R}^{m_i}$ are $K_i$-convex.

A *quasiconvex optimization problem* is exactly the same as a convex optimization problem, except for one difference: the objective function $f_0$ is *quasiconvex*.

We will see an important example of quasiconvex optimization in the next section: linear-fractional programming.

## V. CANONICAL OPTIMIZATION PROBLEMS

In this section, we present several canonical optimization problem formulations, which have been found to be extremely useful in practice, and for which extremely efficient solution codes are available (often for free!). Thus if a real problem can be cast into one of these forms, then it can be considered as essentially solved.

### A. Conic programming

We will present three canonical problems in this section. They are called *conic* problems because the inequalities are specified in terms of affine functions and generalized inequalities. Geometrically, the inequalities are feasible if the range of the affine mappings intersects the cone of the inequality.

The problems are of increasing expressivity and modeling power. However, roughly speaking, each added level of modeling capability is at the cost of longer computation time. Thus one should use the least complex form for the problem at hand.

A general *linear program* (LP) has the form

$$\begin{array}{ll} \text{minimize} & c^T x + d \\ \text{subject to} & Gx \preceq h \\ & Ax = b, \end{array}$$

where $G \in \mathbf{R}^{m \times n}$ and $A \in \mathbf{R}^{p \times n}$.

A problem that subsumes both linear and quadratic programming is the *second-order cone program* (SOCP):

$$\begin{array}{ll} \text{minimize} & f^T x \\ \text{subject to} & \|A_i x + b_i\|_2 \le c_i^T x + d_i, \quad i = 1, \ldots, m \\ & Fx = g, \end{array}$$

where $x \in \mathbf{R}^n$ is the optimization variable, $A_i \in \mathbf{R}^{n_i \times n}$, and $F \in \mathbf{R}^{p \times n}$. When $c_i = 0$, $i = 1, \ldots, m$, the SOCP is equivalent to a quadratically constrained quadratic program (QCQP), (which is obtained by squaring each of the constraints). Similarly, if $A_i = 0$, $i = 1, \ldots, m$, then the SOCP reduces to a (general) LP. Second-order cone programs are, however, more general than QCQPs (and of course, LPs).

A problem which subsumes linear, quadratic and second-order cone programming is called a *semidefinite program* (SDP), and has the form

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x_1 F_1 + \cdots + x_n F_n + G \preceq 0 \\ & Ax = b, \end{array}$$

where $G, F_1, \ldots, F_n \in \mathbf{S}^k$, and $A \in \mathbf{R}^{p \times n}$. The inequality here is a linear matrix inequality. As shown earlier, since SOCP constraints can be written as LMI's, SDP's subsume SOCP's, and hence LP's as well. (If there are multiple LMI's, they can be stacked into one large block diagonal LMI, where the blocks are the individual LMIs.)

We will now consider some examples which are themselves very instructive demonstrations of modeling and the power of convex optimization.

First, we show how slack variables can be used to convert a given problem into canonical form. Consider the *constrained* $\ell_\infty$- *(Chebychev) approximation*

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|_\infty \\ \text{subject to} & Fx \preceq g. \end{array}$$

By introducing the extra variable $t$, we can rewrite the problem as:

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & Ax - b \preceq t\mathbf{1} \\ & Ax - b \succeq -t\mathbf{1} \\ & Fx \preceq g, \end{array}$$

which is an LP in variables $x \in \mathbf{R}^n$, $t \in \mathbf{R}$ (1 is the vector with all components 1).

Second, as (more extensive) example of how an SOCP might arise, consider linear programming with *random constraints*

minimize $c^T x$
subject to $\mathbf{Prob}(a_i^T x \leq b_i) \geq \eta$, $\quad i = 1, \ldots, m$

Here we suppose that the parameters $a_i$ are independent Gaussian random vectors, with mean $\bar{a}_i$ and covariance $\Sigma_i$. We require that each constraint $a_i^T x \leq b_i$ should hold with a probability (or confidence) exceeding $\eta$, where $\eta \geq 0.5$, *i.e.*,

$$\mathbf{Prob}(a_i^T x \leq b_i) \geq \eta.$$

We will show that this probability constraint can be expressed as a second-order cone constraint. Letting $u = a_i^T x$, with $\sigma^2$ denoting its variance, this constraint can be written as

$$\mathbf{Prob}\left( \frac{u - \bar{u}}{\sigma} \leq \frac{b_i - \bar{u}}{\sigma} \right) \geq \eta.$$

Since $(u - \bar{u})/\sigma$ is a zero mean unit variance Gaussian variable, the probability above is simply $\Phi((b_i - \bar{u})/\sigma)$, where

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} \, dt$$

is the cumulative distribution function of a zero mean unit variance Gaussian random variable. Thus the probability constraint can be expressed as

$$\frac{b_i - \bar{u}}{\sigma} \geq \Phi^{-1}(\eta),$$

or, equivalently,

$$\bar{u} + \Phi^{-1}(\eta)\sigma \leq b_i.$$

From $\bar{u} = \bar{a}_i^T x$ and $\sigma = (x^T \Sigma_i x)^{1/2}$ we obtain

$$\bar{a}_i^T x + \Phi^{-1}(\eta)\|\Sigma_i^{1/2} x\|_2 \leq b_i.$$

By our assumption that $\eta \geq 1/2$, we have $\Phi^{-1}(\eta) \geq 0$, so this constraint is a second-order cone constraint. In summary, the stochastic LP can be expressed as the SOCP

minimize $c^T x$
subject to $\bar{a}_i^T x + \Phi^{-1}(\eta)\|\Sigma_i^{1/2} x\|_2 \leq b_i$, $\quad i = 1, \ldots, m$.

We will see examples of using LMI constraints below.

## B. Extensions of conic programming

We now present two more canonical convex optimization formulations, which can be viewed as extensions of the above conic formulations.

A *generalized linear fractional program* (GLFP) has the form:

minimize $f_0(x)$
subject to $Gx \preceq h$
$\qquad\qquad Ax = b$

where the objective function is given by

$$f_0(x) = \max_{i=1,\ldots,r} \frac{c_i^T x + d_i}{e_i^T x + f_i},$$

with $\mathbf{dom}\, f_0 = \{x \mid e_i^T x + f_i > 0, \ i = 1, \ldots, r\}$. The objective function is the pointwise maximum of $r$ quasiconvex functions, and therefore quasiconvex, so this problem is quasiconvex.

A *determinant maximization program* (maxdet) has the form:

minimize $c^T x - \log \det G(x)$
subject to $G(x) \succeq 0$
$\qquad\qquad F(x) \succeq 0$
$\qquad\qquad Ax = b$

where $F$ and $G$ are linear (affine) matrix functions of $x$, so the inequality constraints are LMI's. By flipping signs, one can pose this as a maximization problem, which is the historic reason for the name. Since $G$ is affine and, as shown earlier, the function $-\log \det X$ is convex on the symmetric semidefinite cone $\mathbf{S}_+^n$, the maxdet program is a convex problem.

We now consider an example of each type. First, we consider an optimal transmitter power allocation problem where the variables are the transmitter powers $p_k$, $k = 1, \ldots, m$. Given $m$ transmitters and $mn$ receivers all at the same frequency. Transmitter $i$ wants to transmit to its $n$ receivers labeled $(i, j)$, $j = 1, \ldots, n$, as shown below (transmitter and receiver locations are fixed):
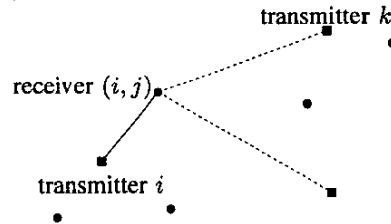


Let $A_{ijk} \in \mathbf{R}$ denote the path gain from transmitter $k$ to receiver $(i, j)$, and $N_{ij} \in \mathbf{R}$ be the (self) noise power of receiver $(i, j)$ So at receiver $(i, j)$ the signal power is $S_{ij} = A_{iji} p_i$ and the noise plus interference

power is $I_{ij} = \sum_{k \neq i} A_{ijk} p_k + N_{ij}$. Therefore the signal to interference/noise ratio (SINR) is

$$S_{ij}/I_{ij}.$$

The optimal transmitter power allocation problem is then to choose $p_i$ to maximize smallest SINR:

$$\begin{array}{ll} \text{maximize} & \min_{i,j} \dfrac{A_{iji} p_i}{\sum_{k \neq i} A_{ijk} p_k + N_{ij}} \\ \text{subject to} & 0 \leq p_i \leq p_{\mathsf{max}} \end{array}$$

This is a GLFP.

Next we consider two related ellipsoid approximation problems. In addition to the use of LMI constraints, this example illustrates techniques of computing with ellipsoids and also how the choice representation of the ellipsoids and polyhedra can make the difference between a tractable convex problem, and a hard intractable one.

First consider computing a *minimum volume ellipsoid around points* $v_1, \ldots, v_K \in \mathbf{R}^n$. This is equivalent to finding the minimum volume ellipsoid around the polytope defined by the convex hull of those points, represented in vertex form.
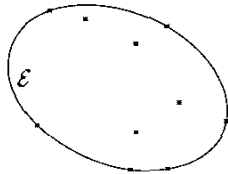


For this problem, we choose the representation for the ellipsoid as $\mathcal{E} = \{x \mid \|Ax - b\| \leq 1\}$, which is parametrized by its center $A^{-1}b$, and the shape matrix $A = A^T \succ 0$, and whose volume is proportional to $\det A^{-1}$. This problem immediately reduces to:

$$\begin{array}{ll} \text{minimize} & \log \det A^{-1} \\ \text{subject to} & A = A^T \succ 0 \\ & \|A v_i - b\| \leq 1, \quad i = 1, \ldots, K \end{array}$$

which is a convex optimization problem in $A$, $b$ ($n + n(n+1)/2$ variables), and can be written as a maxdet problem.

Now, consider finding the *maximum volume ellipsoid in a polytope* given in inequality form

$$\mathcal{P} = \{x \mid a_i^T x \leq b_i, \ i = 1, \ldots, L\}$$
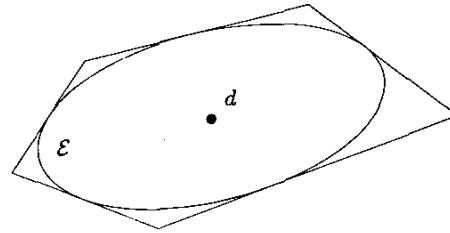


For this problem, we represent the ellipsoid as $\mathcal{E} = \{By + d \mid \|y\| \leq 1\}$, which is parametrized by its center $d$ and shape matrix $B = B^T \succ 0$, and whose volume is proportional to $\det B$. Note the containment condition means

$$\begin{aligned} \mathcal{E} \subseteq \mathcal{P} \iff & \ a_i^T(By + d) \leq b_i \text{ for all } \|y\| \leq 1 \\ \iff & \ \sup_{\|y\| \leq 1} (a_i^T By + a_i^T d) \leq b_i \\ \iff & \ \|B a_i\| + a_i^T d \leq b_i, \ i = 1, \ldots, L \end{aligned}$$

which is a convex constraint in $B$ and $d$. Hence finding the maximum volume $\mathcal{E} \subseteq \mathcal{P}$: is convex problem in variables $B$, $d$:

$$\begin{array}{ll} \text{maximize} & \log \det B \\ \text{subject to} & B = B^T \succ 0 \\ & \|B a_i\| + a_i^T d \leq b_i, \ i = 1, \ldots, L \end{array}$$

Note, however, that minor variations on these two problems, which are convex and hence easy, resutl in problems that are *very* difficult:

- compute the maximum volume ellipsoid inside polyhedron given in vertex form $\mathrm{Co}\{v_1, \ldots, v_K\}$
- compute the minimum volume ellipsoid containing polyhedron given in inequality form $Ax \preceq b$

In fact, just checking whether a given ellipsoid $\mathcal{E}$ covers a polytope described in inequality form is extremely difficult (equivalent to *maximizing* a convex quadratic s.t. linear inequalities).

### C. Geometric programming

In this section we describe a family of optimization problems that are *not* convex in their natural form. However, they are an excellent example of how problems can sometimes be transformed to convex optimization problems, by a change of variables and a transformation of the objective and constraint functions. In addition, they have been found tremendously useful for modeling real problems, *e.g.*circuit design and communication networks.

A function $f : \mathbf{R}^n \to \mathbf{R}$ with $\mathrm{dom}\, f = \mathbf{R}_{++}^n$, defined as

$$f(x) = c x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n},$$

where $c > 0$ and $a_i \in \mathbf{R}$, is called a *monomial function*, or simply, a *monomial*. The exponents $a_i$ of a monomial can be any real numbers, including fractional or negative, but the coefficient $c$ must be nonnegative. A sum of monomials, *i.e.*, a function of the form

$$f(x) = \sum_{k=1}^{K} c_k x_1^{a_{1k}} x_2^{a_{2k}} \cdots x_n^{a_{nk}},$$

where $c_k > 0$, is called a *posynomial function* (with $K$ terms), or simply, a *posynomial*. Posynomials are closed under addition, multiplication, and nonnegative scaling. Monomials are closed under multiplication and division. If a posynomial is multiplied by a monomial, the result is a posynomial; similarly, a posynomial can be divided by a monomial, with the result a posynomial.

An optimization problem of the form

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 1, \quad i = 1, \ldots, m \\ & h_i(x) = 1, \quad i = 1, \ldots, p \end{array}$$

where $f_0, \ldots, f_m$ are posynomials and $h_1, \ldots, h_p$ are monomials, is called a *geometric program* (GP). The domain of this problem is $\mathcal{D} = \mathbf{R}_{++}^n$; the constraint $x \succ 0$ is implicit.

Several extensions are readily handled. If $f$ is a posynomial and $h$ is a monomial, then the constraint $f(x) \leq h(x)$ can be handled by expressing it as $f(x)/h(x) \leq 1$ (since $f/h$ is posynomial). This includes as a special case a constraint of the form $f(x) \leq a$, where $f$ is posynomial and $a > 0$. In a similar way if $h_1$ and $h_2$ are both nonzero monomial functions, then we can handle the equality constraint $h_1(x) = h_2(x)$ by expressing it as $h_1(x)/h_2(x) = 1$ (since $h_1/h_2$ is monomial). We can maximize a nonzero monomial objective function, by minimizing its inverse (which is also a monomial).

We will now transform geometric programs to convex problems by a change of variables and a transformation of the objective and constraint functions.

We will use the variables defined as $y_i = \log x_i$, so $x_i = e^{y_i}$. If $f$ a monomial function of $x$, *i.e.*,

$$f(x) = c x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n},$$

then

$$\begin{aligned} f(x) &= f(e^{y_1}, \ldots, e^{y_n}) \\ &= c(e^{y_1})^{a_1} \cdots (e^{y_n})^{a_n} \\ &= e^{a^T y + b}, \end{aligned}$$

where $b = \log c$. The change of variables $y_i = \log x_i$ turns a monomial function into the exponential of an affine function.

Similarly, if $f$ a posynomial, *i.e.*,

$$f(x) = \sum_{k=1}^{K} c_k x_1^{a_{1k}} x_2^{a_{2k}} \cdots x_n^{a_{nk}},$$

then

$$f(x) = \sum_{k=1}^{K} e^{a_k^T y + b_k},$$

where $a_k = (a_{1k}, \ldots, a_{nk})$ and $b_k = \log c_k$. After the change of variables, a posynomial becomes a sum of exponentials of affine functions.

The geometric program can be expressed in terms of the new variable $y$ as

$$\begin{array}{ll} \text{minimize} & \sum_{k=1}^{K_0} e^{a_{0k}^T y + b_{0k}} \\ \text{subject to} & \sum_{k=1}^{K_i} e^{a_{ik}^T y + b_{ik}} \leq 1, \quad i = 1, \ldots, m \\ & e^{g_i^T y + h_i} = 1, \quad i = 1, \ldots, p, \end{array}$$

where $a_{ik} \in \mathbf{R}^n$, $i = 0, \ldots, m$, contain the exponents of the posynomial inequality constraints, and $g_i \in \mathbf{R}^n$, $i = 1, \ldots, p$, contain the exponents of the monomial equality constraints of the original geometric program.

Now we transform the objective and constraint functions, by taking the logarithm. This results in the problem

$$\begin{array}{ll} \text{minimize} & \tilde{f}_0(y) = \log \left( \sum_{k=1}^{K_0} e^{a_{0k}^T y + b_{0k}} \right) \\ \text{subject to} & \tilde{f}_i(y) = \log \left( \sum_{k=1}^{K_i} e^{a_{ik}^T y + b_{ik}} \right) \leq 0, \quad i = 1, \ldots, m \\ & \tilde{h}_i(y) = g_i^T y + h_i = 0, \quad i = 1, \ldots, p. \end{array}$$

Since the functions $\tilde{f}_i$ are convex, and $\tilde{h}_i$ are affine, this problem is a convex optimization problem. We refer to it as a *geometric program in convex form*. Note that the transformation between the posynomial form geometric program and the convex form geometric program does not involve any computation; the problem data for the two problems are the same.

## VI. NONSTANDARD AND NONCONVEX PROBLEMS

In practice, one might encounter convex problems which do not fall into any of the canonical forms above. In this case, it is necessary to develop custom code for the problem. Developing such codes requires gradient, for optimal performance, Hessian information. If only gradient information is available, the ellipsoid, subgradient or cutting plane methods can be used. These are reliable methods with exact stopping criteria. The same is true for interior point methods, however these also require Hessian information. The payoff for having Hessian information is much faster convergence; in practice, they solve most problems at the cost of a dozen

least squares problems of about the size of the decision variables. These methods are described in the references.

Nonconvex problems are more common, of course, than nonconvex problems. However, convex optimization can still often be helpful in solving these as well: it can often be used to compute lower; useful initial starting points; etc.

## VII. CONCLUSION

In this paper we have reviewed some essentials of convex sets, functions and optimization problems. We hope that the reader is convinced that convex optimization dramatically increases the range of problems in engineering that can be modeled and solved exactly.

## ACKNOWLEDGEMENT

I am deeply indebted to Stephen Boyd and Lieven Vandenberghe for generously allowing me to use material from [7] in preparing this tutorial.

## REFERENCES

[1] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear Programming. Theory and Algorithms.* John Wiley & Sons, second edition, 1993.

[2] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization. Analysis, Algorithms, and Engineering Applications.* Society for Industrial and Applied Mathematics, 2001.

[3] D. P. Bertsekas. *Nonlinear Programming.* Athena Scientific, 1999.

[4] D. Bertsimas and J. N. Tsitsiklis. *Introduction to Linear Optimization.* Athena Scientific, 1997.

[5] S. Boyd and C. Barratt. *Linear Controller Design: Limits of Performance.* Prentice-Hall, 1991.

[6] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory.* Society for Industrial and Applied Mathematics, 1994.

[7] S.P. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, 2003. In Press. Material available at www.stanford.edu/~boyd.

[8] D. Colleran, C. Portmann, A. Hassibi, C. Crusius, S. Mohan, T. Lee S. Boyd, and M. Hershenson. Optimization of phase-locked loop circuits via geometric programming. In *Custom Integrated Circuits Conference (CICC)*, San Jose, California, September 2003.

[9] M. A. Dahleh and I. J. Diaz-Bobillo. *Control of Uncertain Systems: A Linear Programming Approach.* Prentice-Hall, 1995.

[10] J. W. Demmel. *Applied Numerical Linear Algebra.* Society for Industrial and Applied Mathematics, 1997.

[11] G. E. Dullerud and F. Paganini. *A Course in Robust Control Theory: A Convex Approach.* Springer, 2000.

[12] G. Golub and C. F. Van Loan. *Matrix Computations.* Johns Hopkins University Press, second edition, 1989.

[13] M. Grötschel, L. Lovasz, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization.* Springer, 1988.

[14] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction.* Springer, 2001.

[15] M. del Mar Hershenson, S. P. Boyd, and T. H. Lee. Optimal design of a CMOS op-amp via geometric programming. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 20(1):1–21, 2001.

[16] J.-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of Convex Analysis.* Springer, 2001. Abridged version of *Convex Analysis and Minimization Algorithms* volumes 1 and 2.

[17] R. A. Horn and C. A. Johnson. *Matrix Analysis.* Cambridge University Press, 1985.

[18] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4(4):373–395, 1984.

[19] D. G. Luenberger. *Optimization by Vector Space Methods.* John Wiley & Sons, 1969.

[20] D. G. Luenberger. *Linear and Nonlinear Programming.* Addison-Wesley, second edition, 1984.

[21] Z.-Q. Luo. Applications of convex optimization in signal processing and digital communication. *Mathematical Programming Series B*, 97:177–207, 2003.

[22] O. Mangasarian. *Nonlinear Programming.* Society for Industrial and Applied Mathematics, 1994. First published in 1969 by McGraw-Hill.

[23] Y. Nesterov and A. Nemirovskii. *Interior-Point Polynomial Methods in Convex Programming.* Society for Industrial and Applied Mathematics, 1994.

[24] R. T. Rockafellar. *Convex Analysis.* Princeton University Press, 1970.

[25] C. Roos, T. Terlaky, and J.-Ph. Vial. *Theory and Algorithms for Linear Optimization. An Interior Point Approach.* John Wiley & Sons, 1997.

[26] G. Strang. *Linear Algebra and its Applications.* Academic Press, 1980.

[27] J. van Tiel. *Convex Analysis. An Introductory Text.* John Wiley & Sons, 1984.

[28] H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors. *Handbook of Semidefinite Programming.* Kluwer Academic Publishers, 2000.

[29] S. J. Wright. *Primal-Dual Interior-Point Methods.* Society for Industrial and Applied Mathematics, 1997.

[30] Y. Ye. *Interior Point Algorithms. Theory and Analysis.* John Wiley & Sons, 1997.