Mathematical Programming Glossary Supplement: Convex Cones, Sets, and Functions

Harvey J. Greenberg
University of Colorado at Denver
http://www.cudenver.edu/~hgreenbe/

June 10, 2003

This elaborates on convex analysis. Its importance in mathematical programming is due to properties, such as *every local minimum is a global minimum*. Any basic text on nonlinear programming defines and illustrates convex sets and functions. The classics of Grünbaum [2] and Rockafellar [3] are still outstanding books to have for in-depth understanding. This note is simply a digest of fundamental concepts and facts.

Definition 1 The (closed) *line segment* joining two points, u, v, is given by:

$$[u, v] = \{\lambda u + (1 - \lambda)v \colon\ 0 \le \lambda \le 1\}.$$

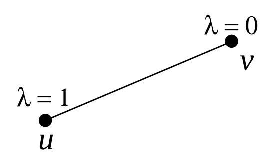

Figure 1: Line segment of u, v

The *open line segment* is $(u, v) = \{\lambda u + (1 - \lambda)v \colon\ 0 < \lambda < 1\}$ (simply the exclusion of the end points, u, v).

Definition 2 A set S is *convex* if $u, v \in S$ implies $[u, v] \subseteq S$.
$\mathbf{R}^n$, $[u, v]$, (u, v), and $\emptyset$ are convex sets.

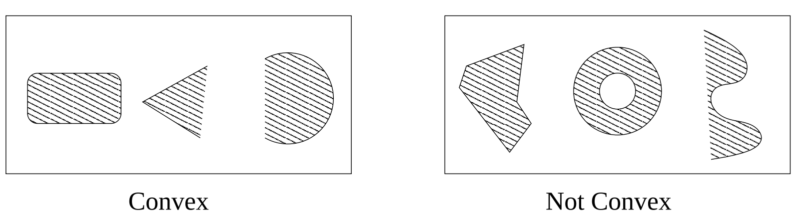

Figure 2: Convex sets have no holes or wrinkly boundaries

The intersection of (a possibly infinite number of) convex sets is a convex set. Once we establish conditions on each constraint function (g_i) such that $\{x \in X \colon\ g_i(x) \le 0\}$ is a convex set, we will have established the convexity of the intersection given by the standard vector inequality,

$\{x \in X\colon g(x) \le 0\}$. When the feasible region has no holes or wrinkles, we can expect good behavior from algorithms designed to move toward an optimum. The straight line from any feasible point to an optimum is completely within the feasible region. Thus, think of convexity as a “betweeness” property.

Definition 3 A *convex combination* of a set of points, $x^1, \dots, x^k \in X$ is a vector x given by

$$x = \sum_{i=1}^{k} \lambda_i x^i \text{ for some } \lambda \ge 0 \text{ such that } \sum_{i=1}^{k} \lambda_i = 1.$$

Each point in the line segment, $[u, v]$, is a convex combination of u and v. Each point in the unit square with corner points, (0,0), (0,1), (1,1), (1,0), is a convex combination of the corner points.

Definition 4 The *convex hull* of a set X, denoted $\mathrm{convh}(X)$, is the set of all convex combinations of the points in X.

If X is a convex set, $\mathrm{convh}(X) = X$; otherwise, the convex hull is a proper superset of X, which “fills in” holes and wrinkles. If X is a finite set, $\mathrm{convh}(X)$ is a *polytope.*

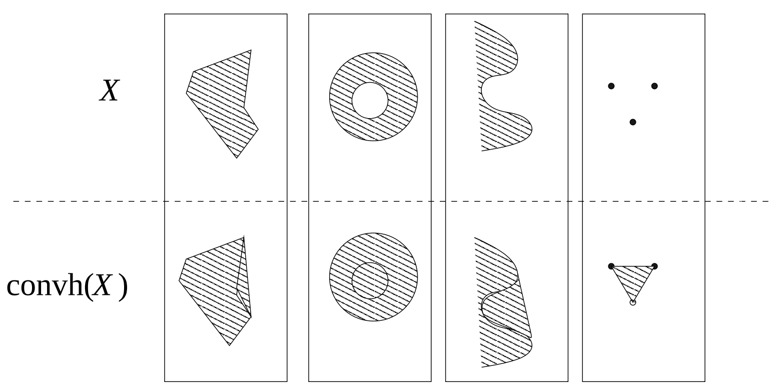

Figure 3: Convex hulls

Definition 5 Suppose $x \in S$. Then, x is an *extreme point* of S if there does not exist $u, v \in S$ such that $x \in (u, v)$.

The extreme points of a line segment are its two end points. The extreme points of a square are its four corner points. Every extreme point is a boundary point, but not conversely. Every extreme point of X is an extreme point of $\mathrm{convh}(X)$, but not conversely. If S is a closed, bounded, convex set with extreme points, $\mathrm{ext}(S)$, then $S = \mathrm{convh}(\mathrm{ext}(S))$ — i.e., S is the set of all convex combinations of its extreme points — i.e., every point in S can be written as a convex combination of its extreme points.

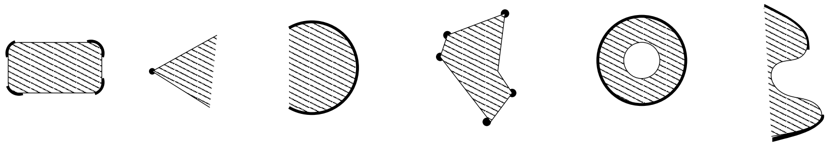

Figure 4: Extreme points are darkened boundary points

Definition 6 A *hyperplane* is the set $H = \{x\colon\ ax = b\}$ for some $a \neq 0$. Its *halfspaces* are

closed halfspaces: $H^{+} = \{x\colon\ ax \geq b\}$, $\quad H^{-} = \{x\colon\ ax \leq b\}$;

open halfspaces: $H^{++} = \{x\colon\ ax > b\}$, $H^{--} = \{x\colon\ ax < b\}$.

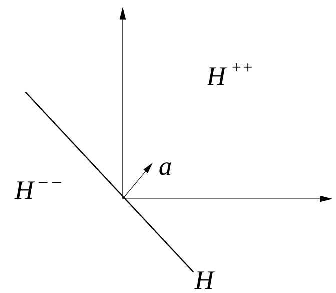

Figure 5: Halfspaces with $b = 0$

A hyperplane and each of its halfspaces are convex sets.

A *polyhedron* is the intersection of a finite number of halfspaces. This can be written as $\{x \in \mathbf{R}^n\colon\ Ax \leq b\}$, where A is an $m \times n$ matrix with no null rows, and b is an m-vector. The i-th inequality, $A_{i\bullet}x \leq b_i$ defines the halfspace associated with the hyperplane $\{x\colon\ A_{i\bullet}x = b_i\}$. Every polyhedron has a finite number of extreme points. A single hyperplane or halfspace is a polyhedron with no extreme point. An orthant is a polyhedron with one extreme point — the origin. If a polyhedron is bounded, it is a polytope. (Some authors reverse the definition of polyhedron and polygon, where it is the polyhedron that is a bounded polytope, but the definition here is more widely used.) A succinct introduction, with focus on computational problems, is given by Fukuda [1].

Definition 7 H is a *supporting hyperplane* at $x \in S$ if $x \in H \cap S$ and $S \subseteq H^{\pm}$.

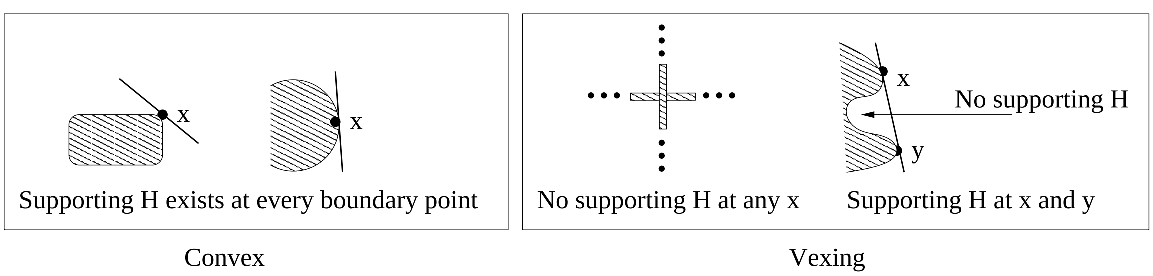

Figure 6: Supporting hyperplanes can exist only at boundary points

In fact, every closed, convex set equals the intersection of the halfspaces of its supporting hyperplanes. In the case of a polyhedron, the number of supporting hyperplanes can be limited to be finite. On the other hand, the disc requires an infinite number, one for each of the extreme points that comprise the circumference.

If we consider maximizing or minimizing a linear function, cx, over a set, S, we obtain a supporting hyperplane $H = \{x\colon cx = cx^*\}$, where x^* optimizes cx on S. By definition of maximum, we must have $S \subseteq H^-$, and by definition of minimum, we must have $S \subseteq H^+$, so in either case, H supports S at x^*. Convexity in this context is an inverse property: *if S is convex and $\bar{x}$ is any boundary point in S, there exists a linear function, cx, for which $c\bar{x}$ is the maximum value of cx on S (and $\bar{x}$ minimizes $-cx$).* We cannot make this claim if S is not convex.

An important foundation for linear programming (LP) is that if an optimum is attained in the polyhedron, $\{x\colon Ax = b, x \geq 0\}$, it must be attained at an extreme point. Instead of the inverse question, we are given c, and we assert that cx^* occurs at an extreme point (if there is an optimum). This follows immediately from the following theorem if the feasible region of the LP is bounded. It can fail when the feasible region is not bounded, such as when there are no extreme points. (The theory of LP is both geometric and algebraic, and the connection is made by assuming $\text{rank}(A) = m$, in which case pathologies vanish and the Supporting Hyperplane Theorem applies indirectly.)

Supporting Hyperplane Theorem. Suppose S is non-empty, closed, and bounded. Then, every supporting hyperplane of S contains an extreme point of S.

Definition 8 H is a (strictly) *separating hyperplane* between two sets, S and X in $\mathbf{R}^n$, if either

$$S \subset H^{--} \quad \text{and} \quad X \subset H^{++}$$
$$\text{or} \quad S \subset H^{++} \quad \text{and} \quad X \subset H^{--}$$

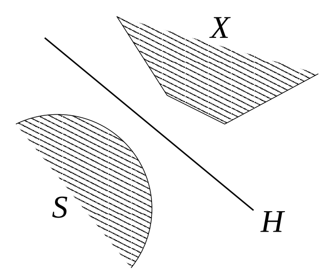

Figure 7: Separating hyperplane

Separating Hyperplane Theorem. Suppose S and X are closed, convex sets in $\mathbf{R}^n$. Then, $S \cap X = \emptyset$ iff there exists a separating hyperplane between them.

The usual way to prove this is by first considering $X = \{x\}$ (a single point not in S). We can then identify the point in S that is closest to x and draw the line segment between them. A separating hyperplane is the perpendicular bisector of that line segment. The convexity of S ensures that it cannot "wrap around" with points in the same halfspace as x. This case is used as a lemma to prove the more general Separating Hyperplane Theorem.

One application is when we are given two finite sets of points, and we want to know if there exists a separating hyperplane between them. This would then serve as a way of knowing when a new point belongs in the class represented by S versus the class represented by X. Such a separating hyperplane exists iff their convex hulls are disjoint — i.e., there exists H for which $S \subset H^{--}$ and $X \subset H^{++}$ iff $\text{convh}(S) \cap \text{convh}(X) = \emptyset$.

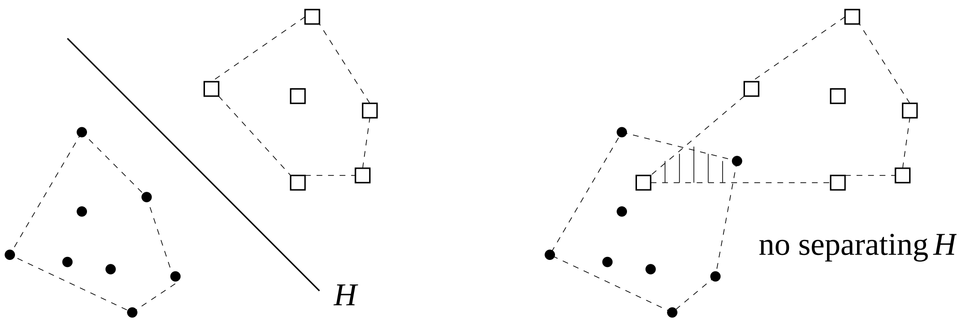

Figure 8: Separating data clusters ↔ disjoint convex hulls

Definition 9 A *ray* through a point $x \neq 0$ is the halfline emanating from the origin:

$$R(x) = \{\lambda x\colon\ \lambda \geq 0\}.$$

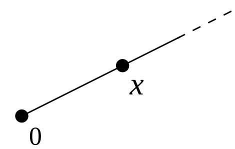

Figure 9: Ray through x

Definition 10 A (closed) *cone*, C, is a set such that $x \in C$ implies $R(x) \subseteq C$. A *convex cone* is a cone that is a convex set. A *polyhedral cone* is a cone that is a polyhedron.

$\mathbb{R}^n$, $\{0\}$ and $\emptyset$ are convex cones. The union of rays, such as two or more coordinate axes, is a cone, but it is not convex. A convex cone can also be defined as a cone C with the property that $u, v \in C$ implies $u + v \in C$. Convex cones are closed under intersection; cones (not necessarily convex) are also closed under union. If C is a polyhedral cone, there exists a matrix A such that $C = \{x\colon\ Ax \leq 0\}$. The quadratic surface, $\{x \in \mathbb{R}^2\colon\ ax_1^2 + bx_1x_2 + cx_2^2 = 0\}$, is a cone, but not polyhedral.

Definition 11 An *extreme ray* of a convex cone, C, is a ray in C that cannot be expressed as sum of two other rays in C.

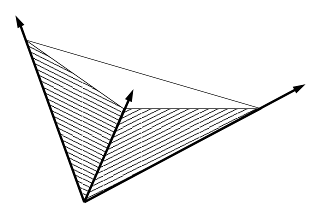

Figure 10: Extreme rays are arrows

Given the convex, polyhedral cone $C = \{x \in \mathbb{R}^n\colon\ Ax = 0, x \geq 0\}$ for which $C \neq \{0\}$, let $S = \{x \in C\colon\ \sum_{j=1}^n x_j = 1\}$. Then, for any $x \in C \setminus \{0\}$, $R(x)$ is an extreme ray of C iff $x / \sum_{j=1}^n x_j$ is an extreme point of S.

Definition 12 The *recession cone* of a polyhedron $P = \{x\colon\ Ax \le b\}$ is $\text{rec}(P) = \{h\colon\ Ah \le 0\}$.

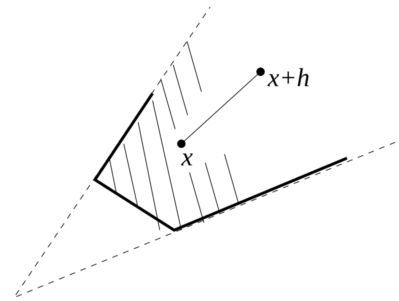

Figure 11: Recession cone

The recession cone is sometimes called the *characteristic cone*. When P is bounded, $\text{rec}(P) = \{0\}$. Otherwise, every vector in the recession cone defines a feasible halfline rooted at ***any*** point in P. In general, $\text{rec}(P) = \{h\colon\ x + h \in P \text{ for all } x \in P\}$, so if $P = \{x\colon\ Ax = b, x \ge 0\}$, $\text{rec}(P) = \{h\colon\ Ah = 0, h \ge 0\}$.

Minkowski-Weyl Polyhedron Decomposition Theorem. If P is a convex polyhedron,

$$P = \text{convh}(\text{ext}(P)) + \text{rec}(P).$$

This means

$$x \in P \leftrightarrow \exists \mu, \lambda \ge 0, \sum_{i=1}^{m} \lambda_i = 1 \text{ such that } x = \sum_{i=1}^{m} \lambda_i v_i + \sum_{i=1}^{M} \mu_i r_i,$$

where $\{v_i\}_1^m$ is the set of extreme points of P, and $\{R(r_i)\}_1^M$ is the set of extreme rays of P. Further, Caratheodory's Theorem says that we can find coefficients such that at most $n+1$ of them are positive.

Definition 13 A *simplex* in $\mathbb{R}^n$ is $\text{convh}\{v_0, v_1, \ldots, v_n\}$ such that $v_1 - v_0, \ldots, v_n - v_0$ are linearly independent. The *dimension* of the simplex is n.

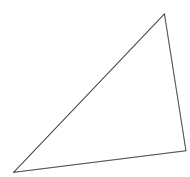

Figure 12: Simplex in $\mathbb{R}^2$

If $n = 0$, the simplex is just a point, and for $n = 1$, it is a line segment with distinct end points. A simplex is a triangle in 2-space and a tetrahedron in 3-space. A common simplex is the convex hull of the origin and unit vectors, $\{0, \mathrm{e}_1, \ldots, \mathrm{e}_n\}$, where e_i is the i-th column of the identity matrix. Any $n+1$ distinct extreme points of a polyhedron satisfy the linear independence property. Thus, another way to state Caratheodory's Theorem for a polytope (where there are no rays) is to say that every point belongs to a simplex (possibly of lower dimension) defined by (at most) $n+1$ extreme points.

Definition 14 The *dimension* of a convex set is the maximum of the dimensions of simplices it contains.

The intersection of a polyhedron, $P = \{x\colon\ Ax \geq b\}$, with one of its supporting hyperplanes is a *face* of P. The extreme points of P comprise the faces of dimension zero. Faces of dimension 1 are the *edges* of P. In general a face F, of P, is either $\emptyset$, all of P, or satisfies:

$$F = \{x \in P\colon\ A_{i\bullet}x = b_i \text{ for } i \in I\} \text{ and } \exists x \in F\colon\ A_{i\bullet}x < b_i \text{ for } i \notin I.$$

If P has dimension n, the faces of dimension $n-1$ are called *facets*. If each facet is a simplex, the polyhedron is *simplicial*.

Definition 15 A function, $f : X \mapsto Y$ is *convex* if X (its domain) is a convex set, and $u, v \in X$ implies

$$f(\lambda u + (1-\lambda)v) \leq \lambda f(u) + (1-\lambda)f(v) \text{ for } 0 \leq \lambda \leq 1.$$

It is *strictly convex* if strict inequality holds for $u \neq v$.

A function is (strictly) *concave* if $-f$ is (strictly) convex; equivalently, if

$$f(\lambda u + (1-\lambda)v) \geq \lambda f(u) + (1-\lambda)f(v) \text{ for } 0 \leq \lambda \leq 1.$$

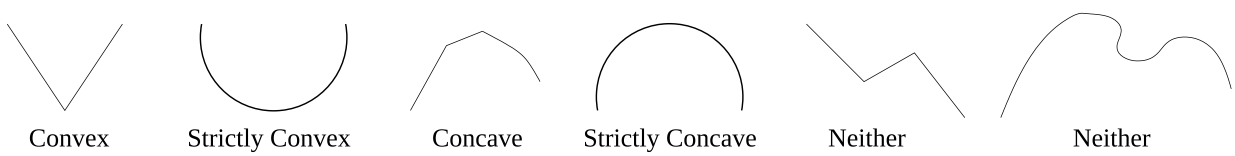

Figure 13: Functions on $\mathbb{R}$

Linear affine functions (of the form $ax - b$) are both convex and concave on any convex subset of $\mathbb{R}^n$. The square Euclidean norm, $||x - c||^2$, is convex. More generally, a quadratic function, x^tQx, is convex on $\mathbb{R}^n$ if it is never negative (i.e., Q is positive semi-definite); it is strictly convex if $x^tQx > 0$ for all $x \neq 0$ (i.e., Q is positive definite), such as $\sum_{j=1}^n x_j^2$. x^tQx is concave if it is never negative (i.e., Q is negative semi-definite); it is strictly concave if $x^tQx < 0$ for all $x \neq 0$ (i.e., Q is negative definite). It is neither convex nor concave on $\mathbb{R}^n$ if $u^tQu > 0$ for some u and $v^tQv < 0$ for some v (e.g., $x_1^2 - x_2^2$).

Sums of convex functions are convex. In particular, $cx + x^tQx$ is convex if Q is positive semi-definite. By extension, if f_1 and f_2 are each convex on $\mathbb{R}$, $f(x_1, x_2) = f_1(x_1) + f_2(x_2)$ is convex on $\mathbb{R}^2$.

When f is differentiable, convex functions obey the inequality

$$f(x + h) \geq f(x) + h^t \nabla f(x)$$

(opposite inequality for concave function). If second derivatives exist, the hessian is positive semi-definite (negative semi-definite for concave function). The inequality means that extrapolation along the gradient underestimates the value of a convex function, and overestimates the value of a concave function.

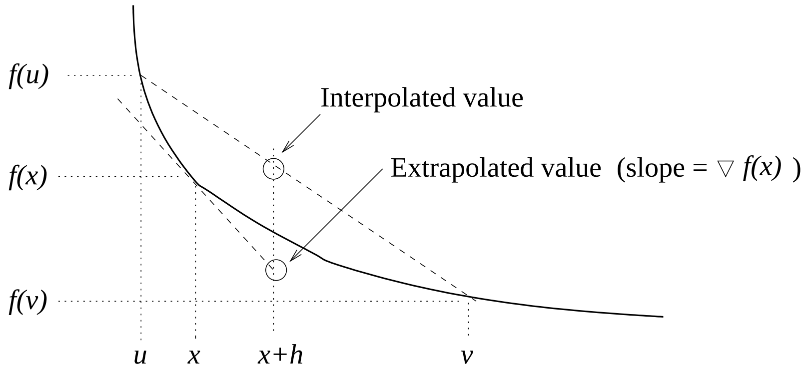

Figure 14: Interpolation overestimates and extrapolation underestimates a convex function

We can also characterize functions as sets, which provides a unified treatment of convex analysis.

Definition 16 Suppose $f : X \mapsto \mathbb{R}$. The *graph*, *epigraph*, and *hypograph* are defined respectively:

$$\begin{aligned} \text{grph}(f, X) &= \{(x, z)\colon x \in X, z = f(x)\} \\ \text{epi}(f, X) &= \{(x, z)\colon x \in X, z \geq f(x)\} \\ \text{hypo}(f, X) &= \{(x, z)\colon x \in X, z \leq f(x)\} \end{aligned}$$

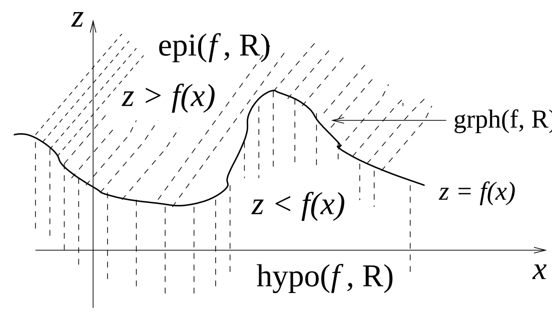

Figure 15: Sets of a function

Then,

$$\begin{array}{llll} f \text{ is convex on } X & \leftrightarrow & \text{epi}(f, X) & \text{is a convex set} \\ f \text{ is concave on } X & \leftrightarrow & \text{hypo}(f, X) & \text{is a convex set} \\ f \text{ is affine on } X & \leftrightarrow & \text{grph}(f, X) & \text{is a convex set} \end{array}$$

The standard form of a mathematical program is

$$\max f(x)\colon x \in X, g(x) \leq 0, h(x) = 0,$$

where $\emptyset \neq X \subseteq \mathbb{R}^n$, $f : X \mapsto \mathbb{R}$, $g : X \mapsto \mathbb{R}^m$, and $h : X \mapsto \mathbb{R}^M$. The *feasible region* is denoted $F = \{x \in X\colon g(x) \leq 0, h(x) = 0\}$. The *optimal region* is denoted $X^* = \{x^* \in F\colon f(x^*) \geq f(x) \text{ for all } x \in F\}$.

A *convex program* is a mathematical program for which X is a convex set, g is convex on X, h is affine on X, and f is concave (for maximization) on X. Key properties:

- The feasible region is a convex set because

$$\begin{array}{lcl} g(u), g(v) \le 0 & \to & g(\lambda u + (1-\lambda)v) \le \lambda g(u) + (1-\lambda) g(v) \le 0 \\ h(u) = h(v) = 0 & \to & h(\lambda u + (1-\lambda)v) = \lambda h(u) + (1-\lambda) h(v) = 0 \end{array}$$
$$\text{for } 0 \le \lambda \le 1.$$

- The optimal region is a convex set because

$$X^* = \bigcap_{x \in F} \{x^* \in F\colon\ f(x^*) \ge f(x)\}.$$

- Every local optimum is global because

$$x + h \in F \text{ and } f(x+h) > f(x) \to x + \varepsilon h \in F \text{ and}$$

$$f(x + \varepsilon h) \ge (1-\varepsilon) f(x) + \varepsilon f(x+h) > f(x) \text{ for all } \varepsilon \in (0,1)$$

  (so x could not be a local maximum in F unless it is a global maximum, which disallows the contradictory strict inequality).

- When $X = \mathbf{R}^n$ and f, g are differentiable such that $g(x) < 0$ for some x such that $h(x) = 0$, the Lagrange Multiplier Rule is both necessary and sufficient for x^* to be (globally) optimal:

$$\textit{There exists } \lambda \in \mathbf{R}^m \textit{ and } \mu \in \mathbf{R}^M \textit{ for which } \lambda \ge 0, \lambda g(x^*) = 0, \textit{ and}$$
$$\nabla f(x^*) - \lambda \nabla g(x^*) - \mu \nabla h(x^*) = 0.$$

There are some pathologies. For example, while every convex function is continuous on the interior of its domain, it can have a discontinuity at the boundary. An example is $X = [0, \infty)$ and $f(x) = x$ if $x > 0$, but $f(0) = K > 0$. For the most part, convex functions are well behaved from an optimization view. All linear functions are convex, as are a large class of quadratics, which includes the square Euclidean norm, $||x - c||^2$. All norms are convex. Further, the additivity property allows us to build up more classes of convex functions.

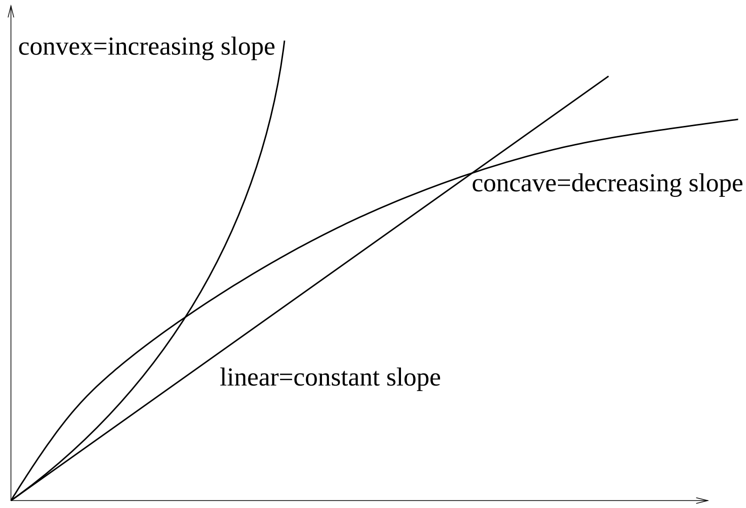

Figure 16: From an economics view, a linear function yields a constant return to scale, a convex function yields an increasing return to scale, and a concave function yields a decreasing return to scale.

References

[1] K. Fukuda. *Frequently Asked Questions in Polyhedral Computation.* Swiss Federal Institute of Technology, http://www.ifor.math.ethz.ch/~fukuda/fukuda.html, 2000.

[2] B. Grünbaum. *Convex Polytopes.* Wiley-Interscience, London, UK, 1967.

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