Chapter 1

Convex sets and functions

1.1 Introduction

Here is a motivating problem (taken from Meerschaert). Suppose a farm has 625 acres available for planting. The available crops are corn, wheat, and oats. The available resources include 1,000 acre-feet of water for the growing season, and 300 hours of labor per week. The crop requirements and profitability are detailed in the following table.

$Reqs/acre$	$corn$	$wheat$	$oats$
$water(acre-ft)$	3.0	1.0	1.5
$Labor(hrs/week)$	0.8	0.2	0.3
$Yield(dollars)$	400	200	250

The question is how to best allocate these resources to the crops to maximize profit. If x_1 denotes the number of acres of corn, x_2 denotes the number of acres of wheat, and x_3 denotes the number of acres of oats, our job is to maximize

$$400x_1 + 200x_2 + 250x_3,$$

subject to the constraints of positivity,

$$x_i \geq 0,$$

acreage,

$$x_1 + x_2 +_3 \leq 625,$$

water,

$$3x_1 + x_2 + 1.5x_3 \leq 1,000,$$

and labor availability,

$$.8x_1 + .2x_2 + .3x_3 \leq 300.$$

This problem is an example of a linear program. There are successful computational methods based largely on a geometric analysis of the problems. We will start by considering some of these geometric structures, talk a bit about more general problems, and then return to the special methods for linear programming. The ideas are nontrivial, but it is still surprising that modern techniques and broad applications, especially in economic problems, were not really developed until after WW2.

In the problem above, the set Ω of values x_1, x_2, x_3 satisfying the constraints is a convex set. The properties of such sets are developed first.

1.2 Basics of convex sets

A set $\Omega \subset \mathbb{R}^N$ is *convex* whenever two points U, V are in the set, then so is the line segment joining them. That is, $U, V \in \Omega$ implies $wU + (1-w)V \in \Omega$ for all $w \in [0,1]$. Here is a set of examples.

(1) A linear subspace

(2) An affine subspace.

$$\Omega = \{U_0 + V\}$$

If $X = U_0 + V_1$, $Y = U_0 + V_2$, then

$$tX + (1-t)Y = U_0 + tV_1 + (1-t)V_2, \quad 0 \leq t \leq 1.$$

(3) A ball of radius R centered at any point. By the triangle inequality

$$\|tX + (1-t)Y\| \leq t\|X\| + (1-t)\|Y\| \leq R.$$

(4) Any translate of a convex set, $U_0 + V$, where $V \in \Omega$.

(5) Half spaces. If $X^* \in \mathbb{R}^N$ then

$$F^{\pm} = \{Y \in \mathbb{R}^N, X^* \bullet Y \leq \alpha\}$$

is convex since

$$X^* \bullet (tY_1 + (1-t)Y_2) = tX^* \bullet Y_1 + (1-t)X^* \bullet Y_2 \leq t\alpha + (1-t)\alpha = \alpha.$$

In general the union of convex sets is not convex (easy), but convex sets are closed under intersections (also easy). Apply this to the linear programming example above. A set which is the intersection of a finite number of half spaces is a *convex polytope*.

If $w_j \geq 0$ and $\sum_{j=1}^J w_j = 1$, say that $V = \sum_{j=1}^J w_j U_j \in \Omega$ is a *convex combination* of the points U_j.

Lemma 1.2.1. *Suppose that $\Omega \subset R^N$ is convex and $U_1, \ldots, U_J \in \Omega$. If $w_j \geq 0$ and $\sum_{j=1}^J w_j = 1$, then $\sum_{j=1}^J w_j U_j \in \Omega$.*

Proof. Without loss of generality, assume that each $w_j > 0$. The proof is by induction on J, the case $J = 1$ being trivial. For $J > 1$ write

$$\sum_{j=1}^J w_j U_j = w_1 U_1 + \sum_{j=2}^J w_j U_j = w_1 U_1 + [\sum_{j=2}^J w_j]\Big[\Big(\sum_{j=2}^J \frac{w_j}{\sum_{j=2}^J w_j}\Big) U_j\Big].$$

Since $\sum_{j=2}^J (w_j/[\sum_{j=2}^J w_j]) = 1$, the term $\sum_{j=2}^J (w_j/[\sum_{j=2}^J w_j]) U_j$ is a convex combination of $J-1$ elements of Ω, hence is in Ω by the induction hypothesis. By the definition of convexity

$$\sum_{j=1}^J w_j U_j = w_1 U_1 + \sum_{j=2}^J w_j U_j \in \Omega.$$

□

Suppose $D \subset \mathbb{R}^N$ is any set. Then there is a smallest convex set $co(D)$ containing D, called the convex hull of D. It can be defined as the intersection of all convex sets containing D.

Proposition 1.2.2. *The set $co(D)$ is the set of all convex combinations of elements of D.*

Proof. Let Ω be the set of all convex combinations of elements of D. Obviously $\Omega \subset co(D)$. Also note that Ω is convex, so $co(D) \subset \Omega$. □

1.3 Convex functions

Earlier we saw that smooth functions $f : \mathbb{R}^N \to \mathbb{R}$ with everywhere positive semidefinite Hessians had nice properties, including the fact that critical points were global minimizers. Even in one variable the condition that a function have positive second derivative has some weaknesses. Consider the family of functions

$$f_\epsilon(x) = \int_0^x \tan^{-1}(t/\epsilon)\ dt, \quad -\infty < x < \infty, \quad \epsilon > 0.$$

This is a smooth function, with $f_\epsilon'(x) = \tan^{-1}(x/\epsilon)$ and

$$f_\epsilon''(x) = \frac{1/\epsilon}{1+(x/\epsilon)^2} = \frac{\epsilon}{x^2+\epsilon^2} > 0.$$

As $\epsilon \to 0^+$ the functions converge pointwise to $\frac{\pi|x|}{2}$, which has the same global minimizer as $f_\epsilon(x)$, but which is not differentiable. We would like some condition that captures the behavior of the 'positive second derivative test', but does not require differentiability. The techniques that are developed often prove to be effective ways to analyze problems.

Suppose $\Omega \subset \mathbb{R}^N$ is convex and $f : \Omega \to \mathbb{R}$. Say that f is *convex* if

$$f(tX + (1-t)Y) \le tf(X) + (1-t)f(Y), \quad X, Y \in \Omega, \quad 0 \le t \le 1.$$

Say that f is *strictly convex* if

$$f(tX + (1-t)Y) < tf(X) + (1-t)f(Y), \quad X \ne Y \in \Omega, \quad 0 < t < 1.$$

Notice that this definition simply says that the values of f on the line segment joining X and Y lie below the line segment joining $f(X)$ and $f(Y)$.

1.3.1 Convexity in one variable

Let's start with the case of one variable, when $\Omega = (\alpha, \beta)$.

Theorem 1.3.1. *If $f(x)$ is a convex function defined on (α, β), then f is continuous there.*

Proof. Suppose $r, s, t, u, v \in (\alpha, \beta)$ with $r < v < s < t < u$. Ignoring v for now, write

$$t = \frac{t-s}{u-s}u + \frac{u-t}{u-s}s.$$

Convexity of f gives

$$f(t) \le \frac{t-s}{u-s}f(u) + \frac{u-t}{u-s}f(s)$$

$$= \frac{t-s}{u-s}f(u) + \frac{u-t}{u-s}f(s) + \frac{t-s}{u-s}f(s) - \frac{t-s}{u-s}f(s),$$

or

$$f(t) \le f(s) + (t-s)\frac{f(u)-f(s)}{u-s}. \tag{1.3.1}$$

There is a version of this inequality for $f(s)$,

$$f(s) \le f(r) + (s-r)\frac{f(t)-f(r)}{t-r},$$

which may be rewritten as

$$f(t) \ge f(r) + (t-r)\frac{f(s)-f(r)}{s-r}$$

$$= f(r) + (t-s+s-r)\frac{f(s)-f(r)}{s-r},$$

or

$$f(t) \ge f(s) + (t-s)\frac{f(s)-f(r)}{s-r}. \tag{1.3.2}$$

The right hand sides of the inequalities (1.3.1) and (1.3.2) are functions of t with limits as $t \to s^+$. In both cases the limits are $f(s)$. Since $f(t)$ is sandwiched between these functions, the Squeeze Theorem gives

$$\lim_{t \to s^+} f(t) = f(s).$$

The case

$$\lim_{v \to s^-} f(r) = f(s)$$

is similar.

□

Theorem 1.3.2. *If $f(x)$ is a differentiable function defined on (α, β), then f is convex if and only if f lies above its tangent lines, that is*

$$f(x) + f'(x)(y-x) \le f(y), \quad x, y \in (\alpha, \beta).$$

Proof. Suppose first that f is convex. Then the defining inequality $f(ty + (1-t)x) \le tf(y) + (1-t)f(x)$ may be written as

$$f(t(y-x)+x) - f(x) \le t(f(y) - f(x)).$$

For $x \neq y$ and $t \neq 0$ this is the same as

$$\frac{f(t(y-x)+x)-f(x)}{t(y-x)}(y-x) \leq f(y)-f(x).$$

Take the limit as $h = t(x-y) \to 0$ to get

$$f'(x)(y-x) \leq f(y)-f(x),$$

as desired.

Conversely, suppose

$$f(x)+f'(x)(y-x) \leq f(y), \quad x, y \in (\alpha, \beta),$$

and let

$$w = tx + (1-t)y, \quad 0 < t < 1.$$

Then

$$f(w)+f'(w)(x-w) \leq f(x), \tag{1.3.3}$$
$$f(w)+f'(w)(y-w) \leq f(y).$$

Since

$$y - w = -\frac{t}{1-t}(x-w),$$

we find that

$$f(w)+f'(w)(x-w)\left(\frac{-t}{1-t}\right) \leq f(y). \tag{1.3.4}$$

Finally, multiply (1.3.3) by t and (1.3.4) by $(1-t)$ and add to get

$$f(w) \leq tf(x)+(1-t)f(y),$$

or

$$f(tx+(1-t)y) \leq tf(x)+(1-t)f(y).$$

□

When a function g has enough derivatives, there is a simple second derivative test that can be used to recognize convex functions.

Theorem 1.3.3. *Suppose $f(x)$ has a nonnegative (positive) second derivative on (α, β). Then f is convex (resp. strictly convex).*

Proof. For some z between x and y,Taylor's Theorem gives

$$f(y) = f(x) + f'(x)(y-x) + \frac{1}{2}(y-x)f''(z)(y-x).$$

By hypothesis,

$$f(y) \geq f(x) + f'(x)(y-x),$$

so f is convex by the previous result. □

This last theorem has a converse of sorts.

Theorem 1.3.4. *Suppose that f has two derivatives on (α, β), and $f''(x) < 0$ for all $x \in (\alpha, \beta)$. Then f is not convex on (α, β).*

Proof. Picking distinct points $x_1 < x_2$ in the interval (α, β), consider the function

$$g(x) = f(x) - f(x_1) - f'(x_1)(x - x_1).$$

The function $f'(x)$ is strictly decreasing on $[\alpha, \beta]$. This implies $g'(x) < 0$ for $x > \alpha$. Since $g(x_1) = 0$, it follows that $g(x_2) < 0$. This means

$$f(x_2) < f(x_1) + f'(x_1)(x_2 - x_1),$$

so g cannot be convex. □

Finally, here is the answer to a calculus student's prayers.

Theorem 1.3.5. *Suppose that $f : (\alpha, \beta) \to \mathbb{R}$ is convex, and $f'(x_1) = 0$ for some $x_1 \in (\alpha, \beta)$. Then x_1 is a global minimizer for f. If f is strictly convex, then f has at most one global minimizer.*

Proof. To see that x_1 is a global minimizer, simply use the fact that the tangent line to a convex function lies below the graph,

$$f(x_1) + f'(x_1)(x - x_1) \leq f(x)$$

to conclude that

$$f(x_1) \leq f(x), \quad \text{for all} \quad x \in (\alpha, \beta).$$

Suppose that f is strictly convex, with a global minimizer at x_1. If x_2 is distinct from x_1, and $f(x_1) = f(x_2)$, the defining inequality for strict convexity gives

$$f(tx_2 + (1-t)x_1) < tf(x_2) + (1-t)f(x_1) = f(x_1), \quad 0 < t < 1,$$

contradicting the assumption that x_1 is a global minimizer.

□

1.3.2 Convexity in several variables

The definition of convexity for a function $f : \mathbb{R}^N \to \mathbb{R}$ describes how f behaves on line segments in $\mathbb{R}^N$. Here is a way to strengthen the link to convex functions of one variable.

Lemma 1.3.6. *Suppose $\Omega \subset \mathbb{R}^N$ is convex, and $X, Y \in \Omega$. If $f : \Omega \to \mathbb{R}$ is (strictly) convex, then*

$$g(t) = f((1-t)X + tY), \quad 0 \le t \le 1,$$

is (strictly) convex on $[0,1]$.

Proof. Suppose $x_1, x_2 \in [0,1]$ and $0 \le t \le 1$. For $i = 1, 2$, let

$$Z_i = (1 - x_i)X + x_i Y.$$

Then

$$g(tx_1 + (1-t)x_2) = f([1 - tx_1 - (1-t)x_2]X + [tx_1 + (1-t)x_2]Y)$$

$$= f((1-t)([1-x_2]X + x_2 Y) + t([1-x_1]X + x_1 Y))$$

$$= f((1-t)Z_2 + tZ_1) \le (1-t)f(Z_2) + tf(Z_1) = (1-t)g(x_2) + tg(x_1),$$

as desired. □

Theorem 1.3.7. *Suppose $\Omega \subset \mathbb{R}^N$ is convex and $f : \Omega \to \mathbb{R}$ is convex. If $X_k \in \Omega$ and*

$$w_k \ge 0, \quad \sum_{k=1}^{K} w_k = 1,$$

then

$$f(\sum_{k=1}^{K} w_k X_k) \le \sum_{k=1}^{K} w_k f(X_k).$$

If f is strictly convex and all $w_k > 0$, then equality holds if and only if all X_k are the same.

Proof. Without loss of generality assume $0 < w_1 < 1$. By induction on K,

$$f(\sum_{k=1}^{K} w_k X_k) = f(w_1 X_1 + [\sum_{k=2}^{K} w_k] \sum_{k=2}^{K} \frac{w_k}{[\sum_{k=2}^{K} w_k]} X_k)$$

$$\leq w_1 f(X_1) + [\sum_{k=2}^{K} w_k] f(\sum_{k=2}^{K} \frac{w_k}{[\sum_{k=2}^{K} w_k]} X_k) \leq \sum_{k=1}^{K} w_k f(X_k).$$

□

Theorem 1.3.8. *Suppose $\Omega \subset \mathbb{R}^N$ is convex and $f : \Omega \to \mathbb{R}$ is convex. Any local minimizer of f is a global minimizer. If f is strictly convex this global minimizer is unique.*

Proof. Suppose Y is a local minimizer and $Z \neq Y$ with $f(Z) < f(Y)$. The fact that Y is a local minimizer means that for $t_1 > 0$ and sufficiently small, $f((1-t_1)Y + t_1 Z) \geq f(Y)$, but since $f(Z) < f(Y)$ the value $(1-t_1)f(Y) + t_1 f(Z)$ is strictly smaller than $f(Y)$, so

$$f((1-t_1)Y + tZ) > (1-t_1)f(Y) + t_1 f(Z),$$

contradicting the convexity of f.

The strictly convex result is similar. □

Theorem 1.3.9. *Suppose $\Omega \subset \mathbb{R}^N$ is an open convex set and $f : \Omega \to \mathbb{R}$ has first partial derivatives in Ω. Then*

(a) $f : \Omega \to \mathbb{R}$ is convex if and only if

$$f(X) + \nabla f(X) \bullet (Y - X) \leq f(Y), \quad X, Y \in \Omega,$$

(b) $f : \Omega \to \mathbb{R}$ is strictly convex if and only if

$$f(X) + \nabla f(X) \bullet (Y - X) < f(Y), \quad X \neq Y.$$

Proof. The proof is essentially the same as in one variable.

Suppose f is convex. Then for $0 \leq t \leq 1$

$$f((1-t)X + tY) \leq (1-t)f(X) + tf(Y),$$

or for $0 < t < 1$

$$\frac{f((1-t)X + tY) - f(X)}{t} = \frac{f(X + t(Y - X)) - f(X)}{t} \leq f(Y) - f(X).$$

Taking the limit as $t \to 0$ gives

$$\nabla f(X) \bullet (Y - X) \leq f(Y) - f(X).$$

Conversely, suppose

$$f(X) + \nabla f(X) \bullet (Y - X) \leq f(Y), \quad X, Y \in \Omega,$$

and let

$$W = tX + (1-t)Y, \quad 0 < t < 1.$$

Then

$$f(W) + \nabla f(W) \bullet (X - W) \leq f(X), \tag{1.3.5}$$

$$f(W) + \nabla f(W) \bullet (Y - W) \leq f(Y).$$

Since

$$Y - W = -\frac{t}{1-t}(X - W),$$

we find that

$$f(W) + \nabla f(W) \bullet (X - W)\left(\frac{-t}{1-t}\right) \leq f(Y). \tag{1.3.6}$$

Finally, multiply (1.3.5) by t and (1.3.6) by $(1-t)$ and add to get

$$f(W) \leq tf(X) + (1-t)f(Y),$$

or

$$f(tX + (1-t)Y) \leq tf(X) + (1-t)f(Y).$$

$\square$

Corollary 1.3.10. *Suppose $\Omega \subset \mathbb{R}^N$ is an open convex set and $f : \Omega \to \mathbb{R}$ is convex and has first partial derivatives in Ω. Then any critical point of f is a global minimizer.*

Proof. If X is a critical point then $\nabla f(X) = 0$. Thus for any $Y \in \Omega$,

$$f(X) + \nabla f(X) \bullet (Y - X) = f(X) \leq f(Y).$$

$\square$

Theorem 1.3.11. *Suppose $f(X)$ has continuous second partials on Ω. If $Hf(X)$ is positive semidefinite (resp. positive definite) for all $X \in \Omega$, then f is convex (resp. strictly convex).*

Proof. Taylor's Theorem gives

$$f(Y) = f(X) + \nabla f(X) \bullet (Y - X) + \frac{1}{2}(Y - X) \bullet Hf(Z)(Y - X).$$

By hypothesis,

$$f(Y) \geq f(X) + \nabla f(X) \bullet (Y - X),$$

so f is convex by the previous result. $\square$

Here are ways to see that a function is convex. Notice that (d) can use differentiable functions to produce nondifferentiable functions.

Theorem 1.3.12. *Suppose $\Omega \subset \mathbb{R}^N$ is convex and $f_k : \Omega \to \mathbb{R}$ are convex functions.*

(a) $f_1(X) + \cdots + f_K(X)$ is convex, and if at least one function is strictly convex, so is the sum.

(b) If $\alpha > 0$, then $\alpha f_k(X)$ is convex.

(c) If f is (strictly) convex and g is a (strictly) increasing convex function on the range of f, then $g(f(X))$ is (strictly) convex.

(d) The function $g(X) = \max f_1(X), \ldots, f_K(X)$ is convex.

Proof. (a) By induction it is sufficient to check

$$f_1(tX+(1-t)Y)+f_2(tX+(1-t)Y) \leq tf_1(X)+(1-t)f_1(Y)+tf_2(X)+(1-t)f_2(Y)$$

$$= t[f_1 + f_2](X) + (1 - t)[f_1 + f_2](Y).$$

(b) easy

(c) The hypotheses give

$$f(tX + (1 - t)Y) \leq tf(X) + (1 - t)f(Y),$$

and

$$g(f(tX + (1 - t)Y)) \leq g(tf(X) + (1 - t)f(Y)) \leq tg(f(X)) + (1 - t)g(f(Y)).$$

(d) Fixing t, X, Y, there is some k such that

$$g(tX + (1 - t)Y) = f_k(tX + (1 - t)Y) \leq tf_k(X) + (1 - t)f_k(Y)$$

$$\leq t \max_k f_k(X) + (1 - t) \max_k f_k(Y) = tg(X) + (1 - t)g(Y).$$

$\square$

Examples:

$$1) \quad f(x_1, x_2, x_3) = \exp(x_1^2 + x_2^2 + x_3^2),$$

2) For fixed vectors $A_k \in \mathbb{R}^N$, and $c_k > 0$,

$$f(X) = \sum_k c_k \exp(A_k \bullet X).$$