KC Border
Fall 2020

Lecture 5: Convex Analysis and Support Functions

5.1 Geometry of the Euclidean inner product

The Euclidean inner product of p and x is defined by

$$p \cdot x = \sum_{i=1}^{m} p_i x_i$$

Properties of the inner product include:

1. $p \cdot p \geqslant 0$ and $p \neq 0 \implies p \cdot p > 0$
2. $p \cdot x = x \cdot p$
3. $p \cdot (\alpha x + \beta y) = \alpha(p \cdot x) + \beta(p \cdot y)$
4. $\|p\| = (p \cdot p)^{1/2}$
5. $p \cdot x = \|p\| \, \|x\| \, \cos\theta$, where θ is the angle between p and x.

To see that

$$x \cdot y = \|x\| \, \|y\| \, \cos\theta,$$

where θ is the angle between x and y, orthogonally project y on the space spanned by x. That is, write $y = \alpha x + z$ where $z \cdot x = 0$. Thus

$$z \cdot x = (y - \alpha x) \cdot x = y \cdot x - \alpha x \cdot x = 0 \qquad \implies \qquad \alpha = x \cdot y / x \cdot x.$$

Referring to Figure 5.1 we see that

$$\cos\theta = \alpha\|x\|/\|y\| = x \cdot y/\|x\| \, \|y\|.$$

For a nonzero $p \in \boldsymbol{R}^{\mathrm{m}}$,

$$\{x \in \boldsymbol{R}^{\mathrm{m}} : p \cdot x = 0\}$$

is a linear subspace of dimension $m - 1$. It is the subspace of all vectors x making a right angle with p.

A set of the form

$$\{x \in \boldsymbol{R}^{\mathrm{m}} : p \cdot x = c\}, \qquad p \neq 0$$

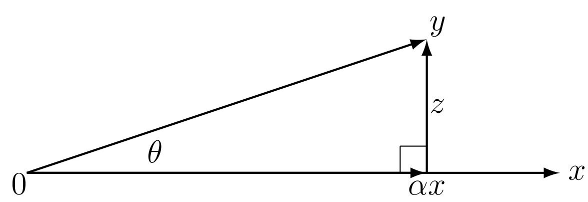

Figure 5.1. Dot product and angles

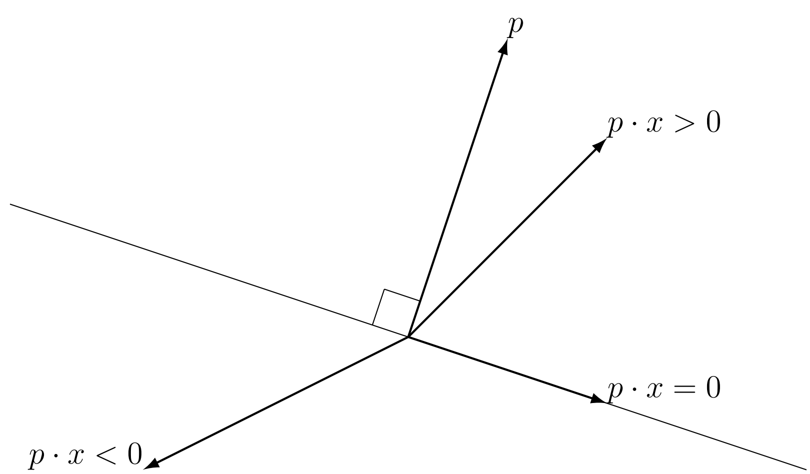

Figure 5.2. Sign of the dot product

is called a **hyperplane**. To visualize the hyperplane $H = \{x : p \cdot x = c\}$ start with the vector $\alpha p \in H$, where $\alpha = c/p \cdot p$. Draw a line perpendicular to p at the point αp. For any x on this line, consider the right triangle with vertices $0, (\alpha p), x$. The angle x makes with p has cosine equal to $\|\alpha p\| / \|x\|$, so $p \cdot x = \|p\| \, \|x\| \, \|\alpha p\| / \|x\| = \alpha p \cdot p = c$. That is, the line lies in the hyperplane H. See Figure 5.3.

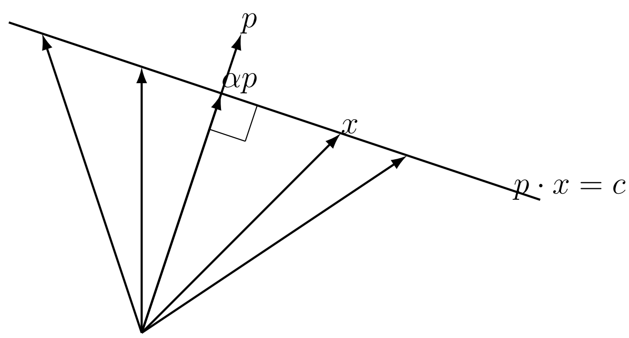

Figure 5.3. A hyperplane.

5.2 Production sets

We now consider a way to describe producers that can potentially produce many commodities. Multiproduct producers are by far more common than single-product producers.

If there are m commodities altogether (inputs, outputs, intermediate goods), a point y in $\boldsymbol{R}^{\mathrm{m}}$ can be used to represent a **production plan**, where y_i indicates the quantity of good i used or produced. The **sign convention** is that $y_i > 0$ indicates that good i is an output and $y_i < 0$ indicates that it is an input. The **technology set** $Y \subset \boldsymbol{R}^{\mathrm{m}}$ is the set of **feasible** plans.

$$
\begin{aligned}
x \geqq y &\iff x_i \geqslant y_i,\ i = 1, \dots, n \\
x > y &\iff x_i \geqslant y_i,\ i = 1, \dots, n \text{ and } x \neq y \\
x \gg y &\iff x_i > y_i,\ i = 1, \dots, n
\end{aligned}
$$

Orderings on $\boldsymbol{R}^{\mathrm{n}}$.

A plan $y \in Y$ is **(technologically) efficient** if there is no $y' \in Y$ such that $y' > y$. A **transformation function** is a function $T \colon \boldsymbol{R}^{\mathrm{m}} \to \boldsymbol{R}$ such that $T(y) \geqslant 0$ if and only if $y \in Y$ and $T(y) = 0$ if and only if y is efficient.

5.3 Maximizing a Linear Function

Given a production set $Y \subset \boldsymbol{R}^{\mathrm{m}}$ obeying our sign convention, and a nonzero vector of prices p, the profit maximization problem is to

$$\underset{y \in Y}{\text{maximize}} \; p \cdot y.$$

This is because outputs have a positive sign, so if $y_j > 0$, then $p_j y_j$ is output j's contribution to revenue, and if $y_j < 0$, then $p_j y_j$ is input j's contribution to costs. Note that for inputs, the price and wage are the same thing.

The **profit function** assigns the maximized profit to each p:

$$\pi(p) = \sup_{y \in Y} p \cdot y.$$

Geometrically it amount to finding the "highest" hyperplane orthogonal to p that touches Y. See Figure 5.4.

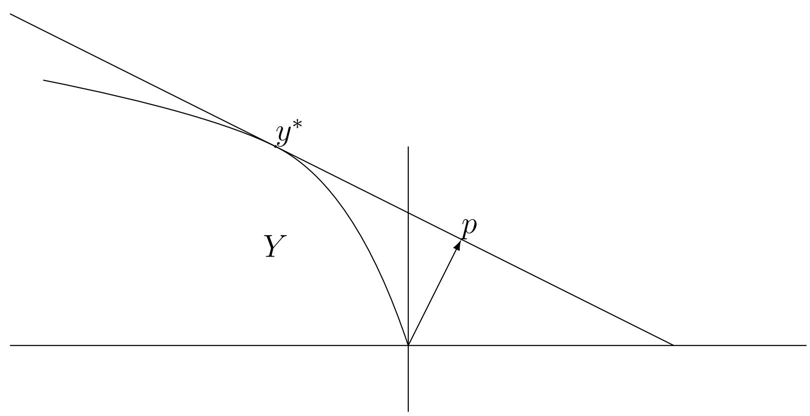

Figure 5.4. Maximizing profit

5.4 Profit and cost functions

Let A be a subset of $\boldsymbol{R}^{\mathrm{m}}$. Convex analysts may give one of two definitions for the **support function** of A as either an infimum or a supremum. Recall that the **supremum** of a set of real numbers is its least upper bound and the **infimum** is its greatest lower bound. By convention, if A has no upper bound, $\sup A = \infty$ and if A has no lower bound, then $\inf A = -\infty$. For the empty set, $\sup A = -\infty$ and $\inf A = \infty$; otherwise $\inf A \leqslant \sup A$. (This makes a kind of sense: Every real number λ is an upper bound for the empty set, since there is no member of the empty set that is greater than λ. Thus the least upper bound must be $-\infty$. Similarly, every real number is also a lower bound, so the infimum is ∞.) Thus support functions (as infima or suprema) may assume the values ∞ and $-\infty$.

By convention, $0 \cdot \infty = 0$; if $\lambda > 0$ is a real number, then $\lambda \cdot \infty = \infty$ and $\lambda \cdot (-\infty) = -\infty$; and if $\lambda < 0$ is a real number, then $\lambda \cdot \infty = -\infty$ and $\lambda \cdot (-\infty) = \infty$. These conventions are used to simplify statements involving positive homogeneity.

Rather than choose one definition, I shall give the two definitions different names derived by the sort of economic interpretation I want to give them.

Profit maximization

The **profit function** π_A of A is defined by

$$\pi_A(p) = \sup_{y \in A} p \cdot y.$$

Cost minimization

The **cost function** c_A of A is defined by

$$c_A(p) = \inf_{y \in A} p \cdot y.$$

5.5 Introduction to convex analysis

A subset of a vector space is **convex** if it includes the line segment joining any two of its points. That is, C is convex if for each pair x, y of points in C, the **line segment**

$$\{\lambda x + (1 - \lambda)y : \lambda \in [0, 1]\}$$

is included in C. Intuitively a convex set has no holes or dents. (This is why my car's license plate reads CONVEX. It aspires to be free of dents.)

The **convex hull** of a set $A \subset \boldsymbol{R}^n$, denoted $\operatorname{co} A$, is the smallest convex set that includes A. You can think of it as filling in any holes or dents. It consists of all points of the form

$$\sum_{i=1}^{m} \lambda_i x_i$$

where each $x_i \in A$, each $\lambda_i > 0$, and $\sum_{i=1}^{m} \lambda_i = 1$. The **closed convex hull**, $\overline{\operatorname{co}}\, A$, is the smallest closed convex set that includes A. It is the **closure** of $\operatorname{co} A$.

5.5.1 Carathéodory's Theorem *In the above sum, m need be no larger than $n + 1$.* (Remember, n is the dimension of the space.)

5.6 Concave and convex functions

An extended real-valued function f on a convex set C is **concave** if its **hypograph**

$$\{(x, \alpha) \in \boldsymbol{R}^m : f(x) \geqslant \alpha\}$$

is a convex set, or equivalently if

$$f\Big((1 - \lambda)x + \lambda y\Big) \geqslant (1 - \lambda)f(x) + \lambda f(y), \qquad (0 < \lambda < 1).$$

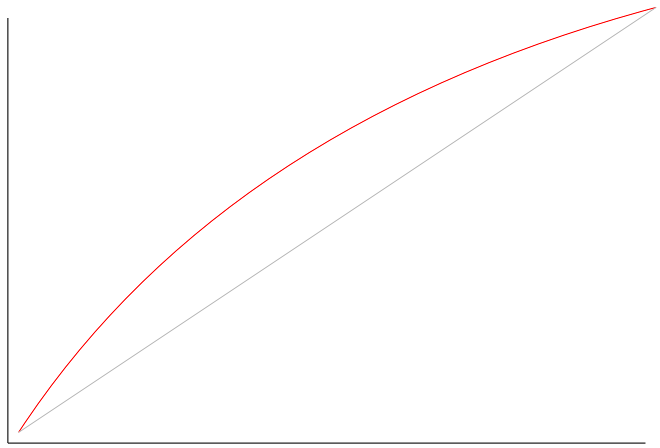

Figure 5.5. A concave function

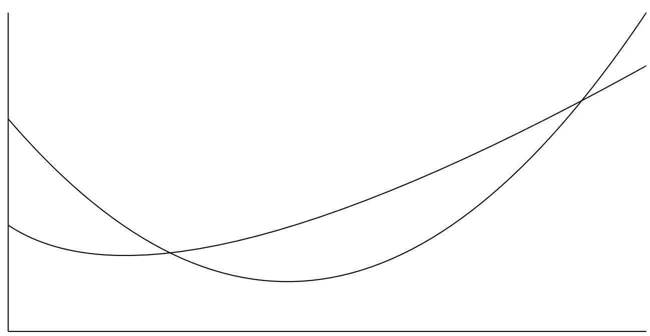

Figure 5.6. The supremum of convex functions is convex

An extended real-valued function f on a convex set C is **convex** if its **epigraph**

$$\{(x,\alpha) \in \boldsymbol{R}^{\mathrm{m}} : f(x) \leqslant \alpha\}$$

is a convex set. Equivalently if

$$f\Big((1-\lambda)x + \lambda y\Big) \leqslant (1-\lambda)f(x) + \lambda f(y), \qquad (0 < \lambda < 1).$$

A function f is convex if and only if $-f$ is concave.

A function $f \colon C \to \boldsymbol{R}$ on a convex subset C of a vector space is:

- **concave** if

$$f\Big(\lambda x + (1-\lambda)y\Big) \geqslant \lambda f(x) + (1-\lambda)f(y)$$

- **strictly concave** if

$$f\Big(\lambda x + (1-\lambda)y\Big) > \lambda f(x) + (1-\lambda)f(y)$$

for all x, y in C with $x \neq y$ and all $0 < \lambda < 1$. It is **convex** if $-f$ is concave, etc.

5.6.1 Proposition *The pointwise supremum of a family of convex functions is convex. The pointwise infimum of family of concave functions is concave.*

To see why this is true, note that the epigraph of the supremum of a family is the intersection of their epigraphs; and the intersection of convex sets is convex.

5.7 Convexity of the profit function

Proposition: π_A is convex, lower semicontinuous, and positively homogeneous of degree 1.

Proposition: c_A is concave, upper semicontinuous, and positively homogeneous of degree 1.

"Proof": Fix p^0, p^1, define $p^\lambda = \lambda p^1 + (1-\lambda)p^0$, and for $\lambda \in [0,1]$, let x^λ maximize $p^\lambda \cdot x$ over A. Then

$$\pi_A(p^0) = p^0 \cdot x^0 \geqslant p^0 \cdot x^\lambda$$
$$\pi_A(p^1) = p^1 \cdot x^1 \geqslant p^1 \cdot x^\lambda$$

So

$$(1-\lambda)\pi_A(p^0) \geqslant (1-\lambda)p^0 \cdot x^\lambda$$
$$\lambda\pi_A(p^1) \geqslant \lambda p^1 \cdot x^\lambda$$

Adding gives:

$$\begin{aligned}\lambda\pi_A(p^1) + (1-\lambda)\pi_A(p^0) &\geqslant \Big(\lambda p^1 + (1-\lambda)p^0\Big) \cdot x^\lambda \\ &= p^\lambda \cdot x^\lambda \\ &= \pi_A(p^\lambda) \\ &= \pi_A\Big(\lambda p^1 + (1-\lambda)p^0\Big)\end{aligned}$$

■

Why the quotes in "proof?" What if no maximizer exists?

Bona fide proof:

$$\pi_A(p) = \sup_{x \in A} f_x(p) \quad \text{where} f_x(p) = p \cdot x.$$

Each f_x is convex (in fact linear), so therefore the supremum is convex. (See my notes on maximization.) ■

Positive homogeneity of π_A is obvious given the conventions on multiplication of infinities. To see that it is convex, let g_x be the linear (hence convex) function defined by $g_x(p) = x \cdot p$. Then $\pi_A(p) = \sup_{x \in A} g_x(p)$. Since the pointwise supremum of a family of convex functions is convex, π_A is convex. Also each g_x is continuous, hence lower semicontinuous, and the supremum of a family of lower semicontinuous functions is lower semicontinuous. See my notes on maximization.

Proposition: The set

$$\{p \in \boldsymbol{R}^{\mathrm{m}} : \pi_A(p) < \infty\}$$

is a closed convex cone, called the **effective domain** of π_A, and denoted $\operatorname{dom} \pi_A$.

The effective domain will always include the point 0 provided A is nonempty. By convention $\pi_\varnothing(p) = -\infty$ for all p, and we say that π_A is **improper**. If $A = \boldsymbol{R}^{\mathrm{m}}$, then 0 is the only point in the effective domain of π_A.

It is easy to see that the effective domain $\operatorname{dom} \pi_A$ of π_A is a cone, that is, if $p \in \operatorname{dom} \pi_A$, then $\lambda p \in \operatorname{dom} \pi_A$ for every $\lambda \geqslant 0$. (Note that $\{0\}$ is a (degenerate) cone.)

It is also straightforward to show that $\operatorname{dom} \pi_A$ is convex. For if $\pi_A(p) < \infty$ and $\pi_A(q) < \infty$, for $0 \leqslant \lambda \leqslant 1$, by convexity of π_A, we have

$$\pi_A\Big(\lambda x + (1-\lambda)y\Big) \leqslant \lambda \pi_A(p) + (1-\lambda)\pi_A(q) < \infty.$$

The closedness of $\operatorname{dom} \pi_A$ is more difficult.

Positive homogeneity of c_A is obvious given the conventions on multiplication of infinities. To see that it is concave, let g_x be the linear (hence concave) function defined by $g_x(p) = x \cdot p$. Then $c_A(p) = \inf_{x \in A} g_x(p)$. Since the pointwise infimum of a family of concave functions is concave, c_A is concave. Also each g_x is continuous, hence upper semicontinuous, and the infimum of a family of upper semicontinuous functions is upper semicontinuous. See my notes on maximization.

Proposition: The set

$$\{p \in \boldsymbol{R}^{\mathrm{m}} : c_A(p) > -\infty\}$$

is a closed convex cone, called the **effective domain** of c_A, and denoted $\operatorname{dom} c_A$.

The effective domain will always include the point 0 provided A is nonempty. By convention $c_\varnothing(p) = \infty$ for all p, and we say that $c_\varnothing$ is **improper**. If $A = \boldsymbol{R}^{\mathrm{m}}$, then 0 is the only point in the effective domain of c_A.

It is easy to see that the effective domain $\operatorname{dom} c_A$ of c_A is a cone, that is, if $p \in \operatorname{dom} c_A$, then $\lambda p \in \operatorname{dom} c_A$ for every $\lambda \geqslant 0$. (Note that $\{0\}$ is a (degenerate) cone.)

It is also straightforward to show that $\operatorname{dom} c_A$ is convex. For if $c_A(p) > -\infty$ and $c_A(q) > -\infty$, for $0 \leqslant \lambda \leqslant 1$, by concavity of c_A, we have

$$c_A\Big(\lambda x + (1-\lambda)y\Big) \geqslant \lambda c_A(p) + (1-\lambda)c_A(q) > -\infty.$$

The closedness of $\operatorname{dom} c_A$ is more difficult.

5.8 When do maximizers and minimizers exist?

Let $K \subset \boldsymbol{R}^{\mathrm{n}}$ be closed and bounded. Let $f \colon K \to \boldsymbol{R}$ be continuous. Then f has a maximizer and minimizer in K.

More generally, let K be a compact subset of a metric space. If f is upper semicontinuous, then f has a maximizer in K, and if f is lower semicontinuous, then f has a minimizer in K.

(See my notes on maximization.)

5.9 Recoverability

Separating Hyperplane Theorem If A is a nonempty closed convex set, and x does not belong to A, then there is a nonzero p satisfying

$$p \cdot x > \pi_A(p).$$

For a proof see my notes. From this theorem we easily get the next proposition.

Proposition: The closed convex hull $\overline{\mathrm{co}}\, A$ of satisfies

$$\overline{\mathrm{co}}\, A = \{y \in \boldsymbol{R}^{\mathrm{m}} : (\forall p \in \boldsymbol{R}^{\mathrm{m}})\ [\, p \cdot y \leqslant \pi_A(p)\,]\}.$$

Proposition: If f is continuous on its effective domain, convex, and positively homogeneous of degree 1, define

$$A = \{y \in \boldsymbol{R}^{\mathrm{m}} : (\forall p \in \boldsymbol{R}^{\mathrm{m}})\ [\, p \cdot y \leqslant f(p)\,]\}.$$

Then A is closed and convex and

$$f = \pi_A.$$

Separating Hyperplane Theorem If A is a nonempty closed convex set, and x does not belong to A, then there is a nonzero p satisfying

$$p \cdot x < c_A(p).$$

For a proof see my notes. From this theorem we easily get the next proposition.

Proposition: The closed convex hull $\overline{\mathrm{co}}\, A$ of satisfies

$$\overline{\mathrm{co}}\, A = \{y \in \boldsymbol{R}^{\mathrm{m}} : (\forall p \in \boldsymbol{R}^{\mathrm{m}})\ [\, p \cdot y \geqslant c_A(p)\,]\}.$$

Proposition: If f is continuous on its effective domain, concave, and positively homogeneous of degree 1, define

$$A = \{y \in \boldsymbol{R}^{\mathrm{m}} : (\forall p \in \boldsymbol{R}^{\mathrm{m}})\ [\, p \cdot y \geqslant f(p)\,]\}.$$

Then A is closed and convex and

$$f = c_A.$$

5.10 Recovering an input requirement set

The **input requirement set** is the set of inputs that allow the producer to produce an output level of at least y. Knowing the cost function as a function of w, the vector of inout wages. Figure 5.7 shows the intersection of a few sets of the form $\{x \in \boldsymbol{R}^{\mathrm{m}} : w \cdot x \geqslant c(w;y)\}$ for the production function $f(x_1,x_2) = x_1^{1/2}x_2^{1/2}$. But if the input requirement set is not convex, you will recover its closed convex hull, see Figure 5.8, for the production function $f(x_1,x_2) = \max\{x_1^{5/6}x_2^{1/6},\ x_1^{1/6}x_2^{5/6}\}$.

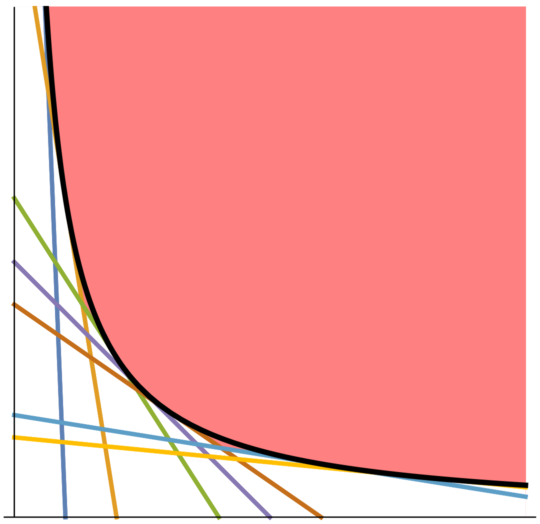

Figure 5.7.

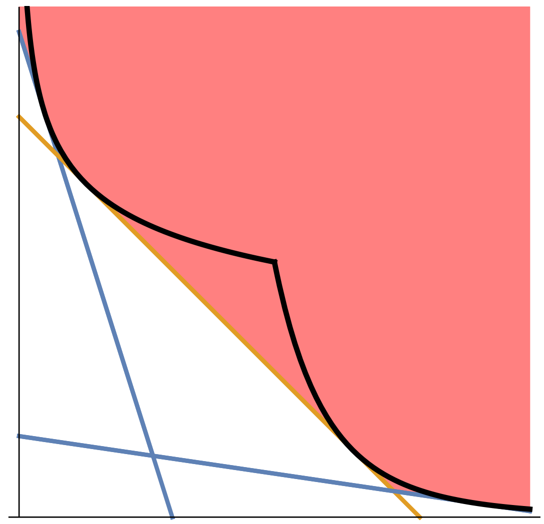

Figure 5.8.

5.11 Concavity and maxima

5.11.1 Proposition *If f is a concave function on a convex set and x^* is a local maximizer, then it is a global maximizer.*

Proof: Prove the contrapositive: Suppose x^* is not a global maximizer. Let

$$f(\hat{x}) > f(x^*).$$

Then for $0 < \lambda < 1$,

$$\lambda f(\hat{x}) + (1-\lambda)f(x^*) > f(x^*).$$

By concavity

$$f(\lambda\hat{x} + (1-\lambda)x^*) \geqslant \lambda f(\hat{x}) + (1-\lambda)f(x^*) > f(x^*),$$

but $\lambda\hat{x} + (1-\lambda)x^* \to x^*$ as $\lambda \to 0$, so x^* is not a local maximizer. ∎

5.12 Supergradients and first order conditions

5.12.1 Definition *Let $f\colon \boldsymbol{R}^{\mathrm{n}} \to \boldsymbol{R}$ be concave. A vector p is a* ***supergradient*** *of f at the point x if for every y it satisfies the* ***supergradient inequality****,*

$$f(x) + p \cdot (y - x) \geqslant f(y).$$

Similarly, if f is convex, then p is a ***subgradient*** *of f at x if*

$$f(x) + p \cdot (y - x) \leqslant f(y)$$

for every y.

For concave f, the set of all supergradients of f at x is called the ***superdifferential*** *of f at x, and is denoted $\partial f(x)$. If the superdifferential is nonempty at x, we say that f is* ***superdifferentiable*** *at x.*

For convex f the same symbol $\partial f(x)$ denotes the set of subgradients and is called the ***subdifferential****. If it is nonempty we say that f is* ***subdifferentiable****.*

5.12.2 Theorem (Gradients are supergradients) *Assume f is concave on a convex set $C \subset \boldsymbol{R}^{\mathrm{n}}$, and differentiable at the point x. Then for every y in C,*

$$f(x) + f'(x) \cdot (y - x) \geqslant f(y). \tag{1}$$

If instead f is convex, then the above inequality is reversed.

Proof: Let $y \in C$. Rewrite the definition of concavity as

$$f\Big(x + \lambda(y - x)\Big) \geqslant f(x) + \lambda\Big(f(y) - f(x)\Big).$$

Rearranging and dividing by $\lambda > 0$,

$$\frac{f\Big(x + \lambda(y - x)\Big) - f(x)}{\lambda} \geqslant f(y) - f(x).$$

Letting $\lambda \downarrow 0$, the left hand side converges to $f'(x) \cdot (y - x)$. ∎

For concave/convex functions the first order conditions for an extremum are sufficient.

5.12.3 Theorem (First order conditions for concave functions) *Suppose f is concave on a convex set $C \subset \boldsymbol{R}^{\mathrm{n}}$. A point x^* in C is a global maximum point of f if and only 0 belongs to the superdifferential $\partial f(x^*)$.*

Suppose f is convex on a convex set $C \subset \boldsymbol{R}^{\mathrm{n}}$. A point x^* in C is a global minimum point of f if and only 0 belongs to the superdifferential $\partial f(x^*)$.

Proof: Note that x^* is a global maximum point of f if and only if

$$f(x^*) + 0 \cdot (y - x^*) \geqslant f(y)$$

for all y in C, but this is just the supergradient inequality for 0. ∎

5.12.4 Corollary *If f is concave and $f'(x^*) = 0$, then x^* is a global maximizer. If f is convex and $f'(x^*) = 0$, then x^* is a global minimizer.*

Proof: The graph of a concave function lies below a horizontal line at x^*, see (1). ∎

A function need not be differentiable to have sub/supergradients.

5.12.5 Theorem (Subdifferentiability) *A convex function on a convex subset of $\boldsymbol{R}^{\mathrm{n}}$ is subdifferentiable at each point of the relative interior of its domain.*

A concave function on a convex subset of $\boldsymbol{R}^{\mathrm{n}}$ is superdifferentiable at each point of the relative interior of its domain.

5.12.6 Fact *If f is concave, then f differentiable at x if and only if its superdifferential $\partial f(x)$ is a singleton, in which case $\partial f(x) = f'(x)$.*

If f is convex, then f differentiable at x if and only if its subdifferential $\partial f(x)$ is a singleton, in which case $\partial f(x) = f'(x)$.

5.13 Jensen's Inequality

5.13.1 *Let f be a concave function, and let X be a random variable taking values in the domain of f, with $|\boldsymbol{E}\, X| < \infty$. Then*

$$f(\boldsymbol{E}\, X) \geqslant \boldsymbol{E}\, f(X).$$

"Proof:": Evaluate (1) at $\boldsymbol{E}\, X$:

$$f(\boldsymbol{E}\, X) + f'(\boldsymbol{E}\, X)(X - \boldsymbol{E}\, X) \geqslant f(X) \quad \text{for all} X$$

and take expectations:

$$f(\boldsymbol{E}\, X) + f'(\boldsymbol{E}\, X)\underbrace{\boldsymbol{E}(X - \boldsymbol{E}\, X)}_{=0} \geqslant \boldsymbol{E}\, f(X).$$

(The result is true even f is not differentiable at $\boldsymbol{E}\, X$.) ∎

5.14 Concavity and second derivatives

5.14.1 Fact *Let f be differentiable on an open interval in $\boldsymbol{R}$.*

Then f is concave if and only if f' is nonincreasing.

If f is concave and twice differentiable at x, then $f''(x) \leqslant 0$.

If f is everywhere twice differentiable with $f'' \leqslant 0$, then f is concave.

If f is everywhere twice differentiable with $f'' < 0$, then f is strictly concave.

5.14.2 Fact *If $f\colon C \subset \boldsymbol{R}^{\mathrm{n}} \to \boldsymbol{R}$ is twice differentiable, then the Hessian H_f is everywhere negative semidefinite if and only if f is concave.*

If H_f is everywhere negative definite, then f is strictly concave.

5.15 The subdifferential of the profit function

$$\partial\pi_A(p) = \{x \in \overline{\text{co}}\, A : p \cdot x = \pi_A(p)\}$$

Extremizers are subgradients

Proposition: If $\tilde{y}(p)$ maximizes p over A, that is, if $\tilde{y}(p)$ belongs to A and $p \cdot \tilde{y}(p) \geqslant p \cdot y$ for all $y \in A$, then $\tilde{y}(p)$ is a subgradient of π_A at p. That is,

$$\pi_A(p) + \tilde{y}(p) \cdot (q - p) \leqslant \pi_A(q) \qquad (*)$$

for all $q \in \boldsymbol{R}^{\mathrm{m}}$.

To see this, note that for any $q \in \boldsymbol{R}^{\mathrm{m}}$, by definition we have

$$q \cdot \tilde{y}(p) \leqslant \pi_A(q).$$

Now add $\pi_A(p) - p \cdot \tilde{y}(p) = 0$ to the left hand side to get the subgradient inequality.

Note that $\pi_A(p)$ may be finite for a closed convex set A, and yet there may be no maximizer. For instance, let

$$A = \{(x, y) \in \boldsymbol{R}^2 : x < 0,\ y < 0,\ xy \geqslant 1\}.$$

Then for $p = (1, 0)$, we have $\pi_A(p) = 0$ as $(1, 0) \cdot (-1/n, -n) = -1/n$, but $(1, 0) \cdot (x, y) = x < 0$ for each $(x, y) \in A$. Thus there is no maximizer in A.

It turns out that if there is no maximizer of p, then π_A has no subgradient at p. In fact, the following is true, but I won't present the proof, which relies on the Separating Hyperplane Theorem. (See my notes for a proof.)

Theorem: If A is closed and convex, then x is a subgradient of π_A at p if and only if $x \in A$ and x maximizes p over A.

Proposition: If $\hat{y}(p)$ minimizes p over A, that is, if $\hat{y}(p)$ belongs to A and $p \cdot \hat{y}(p) \leqslant p \cdot y$ for all $y \in A$, then $\hat{y}(p)$ is a supergradient of c_A at p. That is,

$$c_A(p) + \hat{y}(p) \cdot (q - p) \geqslant c_A(q) \qquad (*)$$

for all $q \in \boldsymbol{R}^{\mathrm{m}}$.

To see this, note that for any $q \in \boldsymbol{R}^{\mathrm{m}}$, by definition we have

$$q \cdot \hat{y}(p) \geqslant c_A(q).$$

Now add $c_A(p) - p \cdot \hat{y}(p) = 0$ to the left hand side to get the supergradient inequality.

Note that $x_A(p)$ may be finite for a closed convex set A, and yet there may be no minimizer. For instance, let

$$A = \{(x, y) \in \boldsymbol{R}^2 : x > 0,\ y > 0,\ xy \geqslant 1\}.$$

Then for $p = (1, 0)$, we have $\pi_A(p) = 0$ as $(1, 0) \cdot (1/n, n) = 1/n$, but $(1, 0) \cdot (x, y) = x > 0$ for each $(x, y) \in A$. Thus there is no minimizer in A.

It turns out that if there is no minimizer of p, then c_A has no supergradient at p. In fact, the following is true, but I won't present the proof, which relies on the Separating Hyperplane Theorem. (See my notes for a proof.)

Theorem: If A is closed and convex, then x is a supergradient of c_A at p if and only if $x \in A$ and x minimizes p over A.

5.16 Hotelling's Lemma

If x is the unique profit maximizer at prices p in the convex set A, then π_A is differentiable at p and $\nabla \pi_A(p) = x$.

5.17 Consequence of Hotelling's Lemma

$$\Big[D_{ij}\pi^*(p)\Big] = \Big[D_j y_i^*(p)\Big] = \Big[\tfrac{\partial y_i^*(p)}{\partial p_j}\Big]$$

is symmetric and positive semidefinite. Consequently,

$$D_i y_i^*(p) = \frac{\partial y_i^*(p)}{\partial p_i} \geqslant 0.$$

Comparative statics

Proposition: Consequently, if A is closed and convex, and $\tilde{y}(p)$ is the unique maximizer of p over A, then π_A is differentiable at p and

$$\tilde{y}(p) = \pi_A'(p). \qquad (**)$$

One way to see this is to consider q of the form $p \pm \lambda e^i$, where e^i is the i^{th} unit coordinate vector, and $\lambda > 0$.
The subgradient inequality for $q = p + \lambda e^i$ is

$$\tilde{y}(p) \cdot \lambda e^i \leqslant \pi_A(p + \lambda e^i) - \pi_A(p)$$

and for $q = p - \lambda e^i$ is

$$-\tilde{y}(p) \cdot \lambda e^i \leqslant \pi_A(p - \lambda e^i) - \pi_A(p).$$

Dividing these by λ and $-\lambda$ respectively yields

$$y_i^*(p) \leqslant \frac{\pi_A(p + \lambda e^i) - \pi_A(p)}{\lambda}$$
$$y_i^*(p) \geqslant \frac{\pi_A(p - \lambda e^i) - \pi_A(p)}{\lambda}.$$

so

$$\tfrac{\pi_A(p-\lambda e^i) - \pi_A(p)}{\lambda} \leqslant y_i^*(p) \leqslant \tfrac{\pi_A(p+\lambda e^i) - \pi_A(p)}{\lambda}.$$

Letting $\lambda \downarrow 0$ yields $\tilde{y}_i(p) = D_i \pi_A(p)$.

Proposition: Consequently, if A is closed and convex, and $\hat{y}(p)$ is the unique minimizer of p over A, then c_A is differentiable at p and

$$\hat{y}(p) = c_A'(p). \qquad (**)$$

One way to see this is to consider q of the form $p \pm \lambda e^i$, where e^i is the i^{th} unit coordinate vector, and $\lambda > 0$.
The supergradient inequality for $q = p + \lambda e^i$ is

$$\hat{y}(p) \cdot \lambda e^i \geqslant c_A(p + \lambda e^i) - c_A(p)$$

and for $q = p - \lambda e^i$ is

$$-\hat{y}(p) \cdot \lambda e^i \geqslant c_A(p - \lambda e^i) - c_A(p).$$

Dividing these by λ and $-\lambda$ respectively yields

$$y_i^*(p) \geqslant \frac{c_A(p + \lambda e^i) - c_A(p)}{\lambda}$$
$$y_i^*(p) \leqslant \frac{c_A(p - \lambda e^i) - c_A(p)}{\lambda}.$$

so

$$\tfrac{c_A(p+\lambda e^i) - c_A(p)}{\lambda} \leqslant y_i^*(p) \leqslant \tfrac{c_A(p-\lambda e^i) - c_A(p)}{\lambda}.$$

Letting $\lambda \downarrow 0$ yields $\hat{y}_i(p) = D_i c_A(p)$.

Proposition: Thus if π_A is twice differentiable at p, that is, if the maximizer $\tilde{y}(p)$ is differentiable with respect to p, then the i^{th} component satisfies

$$D_j y_i^*(p) = D_{ij}\pi_A(p). \tag{***}$$

Consequently, the matrix

$$\Big[D_j y_i^*(p)\Big]$$

is positive semidefinite.

In particular,

$$D_i\tilde{y}_i \geqslant 0.$$

Even without twice differentiability, from the subgradient inequality, we have

$$\begin{aligned}\pi_A(p) + \tilde{y}(p)\cdot(q-p) &\leqslant \pi_A(q)\\ \pi_A(q) + \tilde{y}(q)\cdot(p-q) &\leqslant \pi_A(p)\end{aligned}$$

so adding the two inequalities, we get

$$\Big(\tilde{y}(p) - \tilde{y}(q)\Big)\cdot(p-q) \geqslant 0.$$

Proposition: Thus if q differs from p only in its i^{th} component, say $q_i = p_i + \triangle p_i$, then we have

$$\triangle\tilde{y}_i\triangle p_i \geqslant 0.$$

Dividing by the positive quantity $(\triangle p_i)^2$ does not change this inequality, so

$$\frac{\triangle\tilde{y}_i}{\triangle p_i} \geqslant 0.$$

Proposition: Thus if c_A is twice differentiable at p, that is, if the minimizer $\hat{y}(p)$ is differentiable with respect to p, then the i^{th} component satisfies

$$D_j y_i^*(p) = D_{ij}c_A(p). \tag{***}$$

Consequently, the matrix

$$\Big[D_j y_i^*(p)\Big]$$

is negative semidefinite.

In particular,

$$D_i\hat{y}_i \leqslant 0.$$

Even without twice differentiability, from the supergradient inequality, we have

$$\begin{aligned}c_A(p) + \hat{y}(p)\cdot(q-p) &\geqslant c_A(q)\\ c_A(q) + \hat{y}(q)\cdot(p-q) &\geqslant c_A(p)\end{aligned}$$

so adding the two inequalities, we get

$$\Big(\hat{y}(p) - \hat{y}(q)\Big)\cdot(p-q) \leqslant 0.$$

Proposition: Thus if q differs from p only in its i^{th} component, say $q_i = p_i + \triangle p_i$, then we have

$$\triangle\hat{y}_i\triangle p_i \leqslant 0.$$

Dividing by the positive quantity $(\triangle p_i)^2$ does not change this inequality, so

$$\frac{\triangle\hat{y}_i}{\triangle p_i} \leqslant 0.$$

5.18 Appendix: A Separating Hyperplane Theorem

5.18.1 Strong Separating Hyperplane Theorem *Let K and C be disjoint nonempty convex subsets of a Euclidean space. Suppose K is compact and C is closed. Then there exists a nonzero p that strongly separates K and C.*

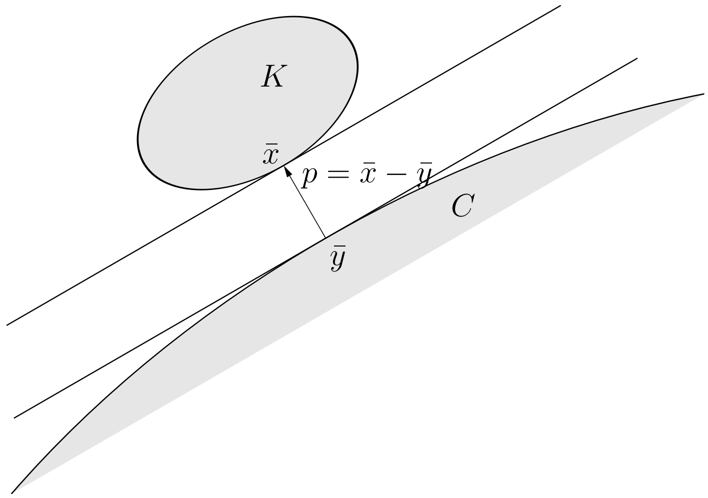

Figure 5.9. Minimum distance and separating hyperplanes.

Proof: Define $f \colon K \to \boldsymbol{R}$ by

$$f(x) = \inf\{d(x,y) : y \in C\},$$

that is $f(x)$ is the distance from x to C. The function f is continuous. To see this, observe that for any y, the distance $d(x',y) \leqslant d(x',x) + d(x,y)$, and $d(x,y) \leqslant d(x,x') + d(x',y)$. Thus $|d(x,y) - d(x',y)| \leqslant d(x,x')$, so $|f(x) - f(x')| \leqslant d(x,x')$. Thus f is actually Lipschitz continuous.

Since K is compact, f achieves a minimum on K at some point $\bar{x}$.

I next claim that there is some point $\bar{y}$ in C such that $d(\bar{x},\bar{y}) = f(\bar{x}) = \inf\{d(\bar{x},y) : y \in C\}$. That is, $\bar{y}$ achieves the infimum in the definition of f, so the infimum is actually a minimum. The proof of this is subtler than you might imagine (particularly in an arbitrary Hilbert space). See Theorem 5.18.3 below.

Put $p = \bar{x} - \bar{y}$. See Figure 5.9. Since K and C are disjoint, we must have $p \neq 0$. Then $0 < \|p\|^2 = p \cdot p = p \cdot (\bar{x} - \bar{y})$, so $p \cdot \bar{x} > p \cdot \bar{y}$. What remains to be shown is that $p \cdot \bar{y} \geqslant p \cdot y$ for all $y \in C$ and $p \cdot \bar{x} \leqslant p \cdot x$ for all $x \in K$:

So let y belong to C. Since $\bar{y}$ minimizes the distance (and hence the square of the distance) to $\bar{x}$ over C, for any point $z = \bar{y} + \lambda(y - \bar{y})$ (with $0 < \lambda \leqslant 1$) on the line segment between y and $\bar{y}$ we have

$$(\bar{x} - z) \cdot (\bar{x} - z) \geqslant (\bar{x} - \bar{y}) \cdot (\bar{x} - \bar{y}).$$

Rewrite this as

$$\begin{aligned}
0 &\geqslant (\bar{x} - \bar{y}) \cdot (\bar{x} - \bar{y}) - (\bar{x} - z) \cdot (\bar{x} - z) \\
&= (\bar{x} - \bar{y}) \cdot (\bar{x} - \bar{y}) - \Big(\bar{x} - \bar{y} - \lambda(y - \bar{y})\Big) \cdot \Big(\bar{x} - \bar{y} - \lambda(y - \bar{y})\Big) \\
&= (\bar{x} - \bar{y}) \cdot (\bar{x} - \bar{y}) - (\bar{x} - \bar{y}) \cdot (\bar{x} - \bar{y}) + 2\lambda(\bar{x} - \bar{y}) \cdot (y - \bar{y}) - \lambda^2 (y - \bar{y}) \cdot (y - \bar{y}) \\
&= 2\lambda(\bar{x} - \bar{y}) \cdot (y - \bar{y}) - \lambda^2 (y - \bar{y}) \cdot (y - \bar{y}) \\
&= 2\lambda p \cdot (y - \bar{y}) - \lambda^2 (y - \bar{y}) \cdot (y - \bar{y}).
\end{aligned}$$

Divide by $\lambda > 0$ to get

$$2p \cdot (y - \bar{y}) - \lambda(y - \bar{y}) \cdot (y - \bar{y}) \leqslant 0.$$

Letting $\lambda \downarrow 0$, we conclude $p \cdot \bar{y} \geqslant p \cdot y$.

A similar argument for $x \in K$ completes the proof. ∎

This proof is a hybrid of several others. The manipulation in the last series of inequalities appears in von Neumann and Morgenstern [6, Theorem 16.3, pp. 134–38], and is probably older. The role of the parallelogram identity in related problems is well known, see for instance, Hiriart-Urruty and Lemaréchal [3, pp. 41, 46] or Rudin [5, Theorem 12.3, p. 293]. A different proof for $\boldsymbol{R}^{\mathrm{n}}$ appears in Rockafellar [4, Corollary 11.4.2, p. 99].

Theorem 5.18.1 is true in general locally convex spaces, where p is interpreted as a continuous linear functional and $p \cdot x$ is replaced by $p(x)$. (But remember, compact sets can be rare in such spaces.) Roko and I give a proof of the general case in [1, Theorem 5.79, p. 207], or see Dunford and Schwartz [2, Theorem V.2.10, p. 417].

5.18.2 Corollary *Let C be a nonempty closed convex subset of a Euclidean space. Assume that the point x does not belong to C. Then there exists a nonzero p that strongly separates x and C. In other words, there exists a nonzero p satisfying $p \cdot x > \pi_C(p)$.*

5.18.3 Theorem (Nearest Point Theorem) *If C is a nonempty closed convex subset of a Euclidean space H, then for each $x \in H$ there exists a unique vector $\pi_{C}(x) \in C$ satisfying $\|x - \pi_{C}(x)\| \leq \|x - y\|$ for all $y \in C$.*

Proof: We can assume that $x = 0$. Put $d = \inf_{u \in C} \|u\|$ and then select a sequence $\{u_n\} \subset C$ such that $\|u_n\| \to d$. From the parallelogram law

$$\begin{aligned}\|u_n - u_m\|^2 &= 2\|u_n\|^2 + 2\|u_m\|^2 - 4\left\|\tfrac{u_n + u_m}{2}\right\|^2 \\ &\leq 2\|u_n\|^2 + 2\|u_m\|^2 - 4d^2 \xrightarrow[n,m \to \infty]{} 0,\end{aligned}$$

we see that $\{u_n\}$ is a Cauchy sequence. Thus there is some $u \in H$ with $u_n \to u$ in H. Since C is closed we have $u \in C$ and by continuity of the norm, $\|u\| = d$. This establishes the existence of a point in C nearest zero.

For the uniqueness of the nearest point, assume that some $v \in C$ satisfies $\|v\| = d$. Then using the parallelogram law once more, we get

$$0 \leq \|u - v\|^2 = 2\|u\|^2 + 2\|v\|^2 - 4\left\|\tfrac{u+v}{2}\right\|^2 \leq 2d^2 + 2d^2 - 4d^2 = 0,$$

so $\|u - v\| = 0$ or $u = v$. ∎

5.18.4 Lemma (Parallelogram Law) *If X is an inner product space, then for each $x, y \in X$ we have*

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2.$$

Proof: Note that $\|x+y\|^2 = \|x\|^2 + 2x \cdot y + \|y\|^2$ and $\|x-y\|^2 = \|x\|^2 - 2x \cdot y + \|y\|^2$. Adding these two identities yields $\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2$. ∎

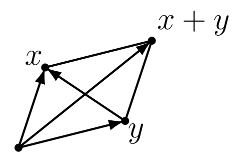

Figure 5.10.

The parallelogram law, which is a simple consequence of the Pythagorean Theorem, asserts that the sum of the squares of the lengths of the diagonals of a parallelogram is equal to the sum of the squares of the lengths of the sides. Consider the parallelogram with vertices $0, x, y, x+y$. Its diagonals are the segments $[0, x+y]$ and $[x, y]$, and their lengths are $\|x+y\|$ and $\|x-y\|$. It has two sides of length $\|x\|$ and two of length $\|y\|$. See Figure 5.10.

5.18.1 Weak Separating Hyperplane Theorem

Let A, B be disjoint nonempty convex subsets of $\boldsymbol{R}^{\mathrm{n}}$. Then there exists a nonzero $p \in \boldsymbol{R}^{\mathrm{n}}$ properly separating A and B.

References

[1] C. D. Aliprantis and K. C. Border. 2006. *Infinite dimensional analysis: A hitchhiker's guide*, 3d. ed. Berlin: Springer–Verlag.

[2] N. Dunford and J. T. Schwartz. 1957. *Linear operators: Part I.* New York: Interscience.

[3] J.-B. Hiriart-Urruty and C. Lemaréchal. 2001. *Fundamentals of convex analysis*. Grundlehren Text Editions. Berlin: Springer–Verlag.

[4] R. T. Rockafellar. 1970. *Convex analysis*. Number 28 in Princeton Mathematical Series. Princeton: Princeton University Press.

[5] W. Rudin. 1973. *Functional analysis*. New York: McGraw Hill.

[6] J. von Neumann and O. Morgenstern. 1944. *The theory of games and economic behavior*. Princeton: Princeton University Press. Reprint. New York: John Wiley & Sons, 1967.