Essential Mathematics for Economics

Alexis Akira Toda

University of California San Diego

Chapter 0

Road map

Typical problem in economics

- N goods indexed by $n = 1, \dots, N$
- When agent consumes $x_n \geq 0$ units of good n, derives utility
$$u(x_1, \dots, x_N)$$
- Unit price of good n is $p_n > 0$
- Agent has disposable income $w > 0$
- What is optimal choice of $x = (x_1, \dots, x_N)$?

Setting up problem mathematically

- If agent consumes x_n units of good n, expenditure is $p_n x_n$
- Hence total expenditure is $p_1 x_1 + \cdots + p_N x_N$
- Mathematically, problem is

$$\begin{aligned} &\text{maximize} && u(x_1, \dots, x_N) \\ &\text{subject to} && p_1 x_1 + \cdots + p_N x_N \leq w, \\ & && (\forall n) x_n \geq 0 \end{aligned}$$

- This problem is called *utility maximization problem* (UMP)
- One of most basic constrained optimization problems studied in economics

Many questions to ask

1. How do we define solution?
2. Does solution exist?
3. What are necessary or sufficient conditions that characterize solution?
4. Is solution unique?
5. How do we compute solution?
6. How does solution change if we change parameters p_n or w?

Chapter 1

Existence of Solutions

Constrained minimization

- We would like to solve

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && x \in C, \end{aligned}$$

  where
  - C is *constraint set*
  - f is *objective function* from C to $\mathbb{R} = (-\infty, \infty)$
- We say $\bar{x} \in C$ is *solution* if $f(\bar{x}) \leq f(x)$ for all $x \in C$
- $\bar{x}$ is also called *minimizer* or *minimum*, and we write

$$f(\bar{x}) = \min_{x \in C} f(x),$$

$$\bar{x} \in \arg\min_{x \in C} f(x)$$

Maximization

- We focus on minimization because maximization problems can be turned into minimization
- For example, consider

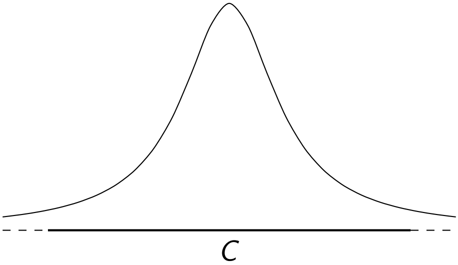
$$\begin{array}{ll} \text{maximize} & g(x) \\ \text{subject to} & x \in C \end{array}$$

- We can convert to

$$\begin{array}{ll} \text{minimize} & f(x) = -g(x) \\ \text{subject to} & x \in C \end{array}$$

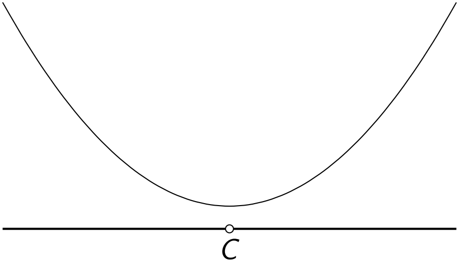
Not all minimization problems have solutions

- Constraint set unbounded



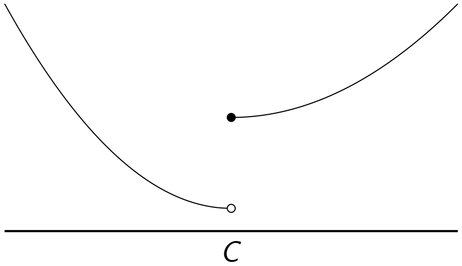
Not all minimization problems have solutions

- Constraint set has hole



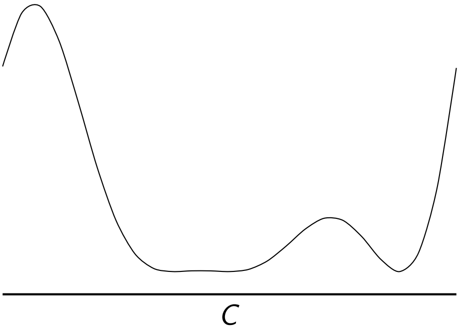
Not all minimization problems have solutions

- Graph of objective function has gap



Not all minimization problems have solutions

- Minimum exists, but not unique



Real number system

- $\mathbb{N} = \{1, 2, \dots\}$: set of *natural numbers*
- $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$: set of *integers*
- $\mathbb{Q} = \{m/n : m \in \mathbb{Z}, n \in \mathbb{N}\}$: set of *rational numbers*
- $\mathbb{R}$: set of *real numbers*
- We assume you know what $\mathbb{R}$ is
- Essentially, $\mathbb{R}$ is set on which we can do addition, subtraction, multiplication, and division, and has some continuity property ($\sqrt{2}$ is not in $\mathbb{Q}$ but is in $\mathbb{R}$)

Some terminology

- *Absolute value* of $x \in \mathbb{R}$ is denoted by
$$|x| = \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x < 0 \end{cases}$$
- $A \subset \mathbb{R}$ is *bounded above* if there exists $b \in \mathbb{R}$ such that $x \leq b$ for all $x \in A$
- b is called *upper bound* of A
- Bounded below/lower bound analogous
- If both bounded above and below, we just say *bounded*: there exists $b \geq 0$ such that $|x| \leq b$ for all $x \in A$

Extended real numbers

- Often convenient to consider set of *extended real numbers* that includes plus or minus infinity: $\pm\infty$
- Rules of algebra:

$$
\begin{aligned}
&x \pm \infty = \pm\infty \text{ if } x \in \mathbb{R}, \\
&\infty + \infty = \infty, \\
&x \times (\pm\infty) = \pm\infty \text{ if } x > 0, \\
&x \times (\pm\infty) = \mp\infty \text{ if } x < 0, \\
&x/(\pm\infty) = 0 \text{ if } x \in \mathbb{R}
\end{aligned}
$$

- Note: $\infty - \infty$ and ∞/∞ are undefined, though convenient to define $0 \times \infty = 0$

Least-upper-bound property

- If $x \leq a$ ($x \geq a$) for all $x \in A$ and $a \in A$, we call a *maximum (minimum)* of A
- Defining property of $\mathbb{R}$ is *least-upper-bound property*: if A is bounded above, there exists least upper bound
- More precisely, if $\emptyset \neq A \subset \mathbb{R}$ is bounded above and B is set of upper bounds of A, then $\alpha = \min B$ exists
- Least upper bound α is called *supremum* of A and is denoted by $\alpha = \sup A$
- Symmetric argument shows that if A is bounded below, then greatest lower bound exists, called *infimum* of A and denoted by $\inf A$

Convergence of sequences

- Real sequence $(x_1, x_2, \dots) = \{x_k\}_{k=1}^{\infty}$ is function from $\mathbb{N}$ to $\mathbb{R}$
- We say $\{x_k\}$ *converges* to x if

$$(\forall \epsilon > 0)(\exists K > 0)(\forall k \geq K) \quad |x_k - x| < \epsilon$$

- In words: give me any error tolerance $\epsilon > 0$; I can take K large enough such that error between x_k and x is less than ϵ if $k \geq K$
- Write $\lim_{k\to\infty} x_k = x$ or $x_k \to x$ $(k \to \infty)$ and call x *limit* of $\{x_k\}$
- We say $\{x_k\}_{k=1}^{\infty} \subset \mathbb{R}$ *converges to infinity* if

$$(\forall \epsilon > 0)(\exists K > 0)(\forall k \geq K) \quad x_k > \epsilon$$

Monotone sequences are convergent

- $\{x_k\}$ is *monotone increasing (decreasing)* if $x_1 \le x_2 \le \cdots$ $(x_1 \ge x_2 \ge \cdots)$

Proposition

If $\{x_k\}_{k=1}^{\infty} \subset [-\infty, \infty]$ *is monotone, it is convergent.*

Proof.

- Without loss of generality (wlog), assume $\{x_k\}$ increasing
- Let $x = \sup \{x_k : k \in \mathbb{N}\}$
- If $x < \infty$, take any $\epsilon > 0$; then by definition of supremum, $(\exists K) x_K \in (x - \epsilon, x]$
- By monotonicity, $x_k \in (x - \epsilon, x]$ for all $k \ge K$, so $|x_k - x| < \epsilon$ and $x_k \to x$
- If $x = \infty$, analogous argument $\square$

Limit superior and inferior

- Let $\{x_k\}_{k=1}^{\infty} \subset [-\infty, \infty]$ be any sequence
- Define

$$\alpha_k = \sup \{x_k, x_{k+1}, \dots\} = \sup_{l \geq k} x_l$$

- Since the set $\{x_l : l \geq k\}$ decreasing with k, clearly $\{\alpha_k\}_{k=1}^{\infty}$ is decreasing sequence in $[-\infty, \infty]$
- Hence by previous proposition, limit

$$\alpha := \lim_{k \to \infty} \alpha_k = \lim_{k \to \infty} \sup_{l \geq k} x_l$$

exists, called *limit superior* of $\{x_k\}$ and denoted by

$$\alpha = \limsup_{k \to \infty} x_k$$

- Limit inferior analogous

The space $\mathbb{R}^N$

- We are often interested in functions of several variables
- Let $\mathbb{R}^N$ denote set of N-tuples of real numbers $x = (x_1, \dots, x_N) = (x_n)$
- For $x, y \in \mathbb{R}^N$, define sum entrywise by $x + y = (x_n + y_n)$
- For $\alpha \in \mathbb{R}$ and $x \in \mathbb{R}^N$, define scalar multiplication entrywise by $\alpha x = (\alpha x_n)$
- In general, we call set X (real) *vector space* if sum $x + y$ and scalar product αx are defined and belong to X for all vectors $x, y \in X$ and scalar $\alpha \in \mathbb{R}$

Linear functions

- If X is vector space and $f : X \to \mathbb{R}$, we say f is *linear* if f preserves addition and scalar multiplication:

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$$

  for all $x, y \in X$ and $\alpha, \beta \in \mathbb{R}$
- An obvious example of linear function $f : \mathbb{R}^N \to \mathbb{R}$ is

$$f(x) = a_1 x_1 + \cdots + a_N x_N = \sum_{n=1}^{N} a_n x_n,$$

  where $a_1, \dots, a_N \in \mathbb{R}$
- We can prove converse too, because if f linear, write $x = x_1 e_1 + \cdots + x_N e_N$ for unit vectors $\{e_n\}$, and

$$f(x) = f(x_1 e_1 + \cdots + x_N e_N) = x_1 f(e_1) + \cdots + x_N f(e_N)$$

Inner product

- Expression of form $a_1x_1 + \cdots + a_Nx_N$ appears so often that it deserves special name and notation
- Let $x = (x_1, \dots, x_N)$ and $y = (y_1, \dots, y_N)$ be two vectors in $\mathbb{R}^N$
- Then
$$\langle x, y\rangle := x_1y_1 + \cdots + x_Ny_N = \sum_{n=1}^{N} x_ny_n$$
is called *inner product* of x and y
- Other common notations are (x, y), $x \cdot y$, and $\langle x \mid y\rangle$, etc.
- Fixing x, inner product $\langle x, y\rangle$ is linear in y, so we have
$$\langle x, \alpha_1y_1 + \alpha_2y_2\rangle = \alpha_1 \langle x, y_1\rangle + \alpha_2 \langle x, y_2\rangle$$

Euclidean norm

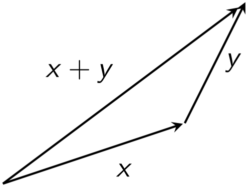
- To do analysis, it is convenient to have notion of size of vector or distance between two vectors
- Motivated by Pythagorean theorem in elementary geometry, *(Euclidean) norm* of $x \in \mathbb{R}^N$ is defined by

$$\|x\| := \sqrt{\langle x, x \rangle} = \sqrt{x_1^2 + \cdots + x_N^2}$$

- Euclidean norm is also called ℓ^2 norm for reason that will become clear later

Normed space

- More generally, for real vector space X, function $\|\cdot\| : X \to \mathbb{R}$ is called *norm* if:
  1. (Nonnegativity) $\|x\| \geq 0$ for all $x \in X$, with equality if and only if $x = 0$
  2. (Positive homogeneity) $\|\alpha x\| = |\alpha| \, \|x\|$ for all $\alpha \in \mathbb{R}$ and $x \in X$
  3. (Triangle inequality) $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in X$
- Vector space X equipped with norm $\|\cdot\|$ is called *normed space*



Examples of norms

- There are many norms

$$
\begin{aligned}
&(\ell^1 \text{ norm}) && \|x\|_1 := \sum_{n=1}^N |x_n|, \\
&(\ell^\infty \text{ or sup norm}) && \|x\|_\infty := \max_n |x_n|, \\
&(\ell^p \text{ norm for } p \geq 1) && \|x\|_p := \left(\sum_{n=1}^N |x_n|^p\right)^{1/p}
\end{aligned}
$$

- Proofs that $\|\cdot\|_1$ and $\|\cdot\|_\infty$ are norms straightforward
- Proof that $\|\cdot\|_p$ is norm uses Minkowski inequality, discussed much later

Equivalence of norms

Theorem

Let $\|\cdot\|_1$, $\|\cdot\|_2$ be two norms on $\mathbb{R}^N$. Then there exist constants $0 < c \le C$ such that

$$c\,\|x\|_1 \le \|x\|_2 \le C\,\|x\|_1$$

for all $x \in \mathbb{R}^N$.

- See textbook for proof (complicated)
- Hence in $\mathbb{R}^N$, it does not matter which norm to use to define convergence: $(\forall \epsilon > 0)(\exists K > 0)(\forall k \ge K)\,\|x_k - x\| < \epsilon$
- To see equivalence of Euclidean (ℓ^2) and sup (ℓ^∞) norms, note

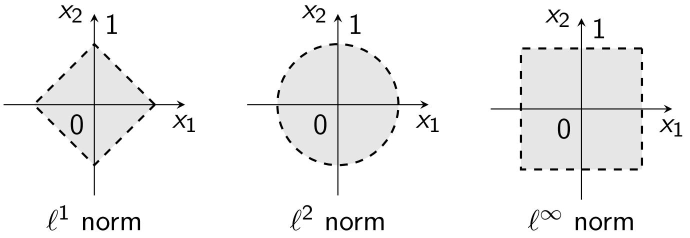
$$\|x\|_2 = \sqrt{\sum_{n=1}^N x_n^2} \ge |x_n| \implies \|x\|_2 \ge \|x\|_\infty\,,$$
$$\|x\|_2 = \sqrt{\sum_{n=1}^N x_n^2} \le \sqrt{N\,\|x\|_\infty^2} = \sqrt{N}\,\|x\|_\infty$$

Balls

- For $x \in \mathbb{R}^N$ and $\epsilon > 0$, set

$$B_\epsilon(x) := \left\{ y \in \mathbb{R}^N : \|y - x\| < \epsilon \right\}$$

  is called *ball* with center x and radius ϵ
- Shape of ball depends on norm used



Open sets

- Let X be normed space and $A \subset X$
- We say x is *interior point* of A if there exists $\epsilon > 0$ such that $B_\epsilon(x) \subset A$ (we can draw ball with center x and radius ϵ that is entirely contained in A)
- If every $x \in A$ is interior point of A, we say that A is *open set*
- We often use symbols U and V to denote open set because French word for "open" is *ouvert* but letter O is confusing due to resemblance to 0
- By definition, empty set $\emptyset$ and entire space X are open

Complement, closed sets

- For $A \subset X$, let $A^c := X \backslash A = \{x \in X : x \notin A\}$ denote its *complement*
- We say that A is *closed set* if A^c is open
- We often use symbol F to denote closed set because French word for "closed" is *fermé*
- By definition, both $\emptyset, X$ are closed

Examples

- Interval $(a, b) = \{x \in \mathbb{R} : a < x < b\}$ is open
- Interval $[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$ is closed
- Interval $(a, b] = \{x \in \mathbb{R} : a < x \leq b\}$ is neither open nor closed
- ϵ-ball is open

Proof.

- Let $y \in B_\epsilon(x)$; by definition, $\|y - x\| < \epsilon$; define $\delta := \epsilon - \|y - x\| > 0$
- If $z \in B_\delta(y)$, then by triangle inequality,

$$\|z - x\| \leq \|z - y\| + \|y - x\| < \delta + \|y - x\| = \epsilon,$$

  so $z \in B_\epsilon(x)$
- Therefore $B_\delta(y) \subset B_\epsilon(x)$, so $B_\epsilon(x)$ is open $\square$

Unions and intersections of open sets

Proposition

Any union of open sets is open: if I is any set and U_i is open for each $i \in I$, so is $\bigcup_{i \in I} U_i$. Any finite intersection of open sets is open: if U_j is open for each $j = 1, \dots, J$, so is $\bigcap_{j=1}^{J} U_j$.

Proof.

- Suppose U_i open for each $i \in I$ and let $U = \bigcup_{i \in I} U_i$
- If $x \in U$, then $x \in U_i$ for some i; since U_i is open, we can take some $\epsilon > 0$ such that $B_\epsilon(x) \subset U_i \subset U$, so U is open
- Suppose U_j is open for each $j = 1, \dots, J$ and let $U = \bigcap_{j=1}^{J} U_j$
- If $x \in U$, then in particular $x \in U_j$, so we can take $\epsilon_j > 0$ such that $B_{\epsilon_j}(x) \subset U_j$
- Let $\epsilon = \min_j \epsilon_j$; then $B_\epsilon(x) \subset B_{\epsilon_j}(x) \subset U_j$ for all j, so $B_\epsilon(x) \subset \bigcap_{j=1}^{J} U_j = U$ and U is open $\square$

Unions and intersections of closed sets

Corollary

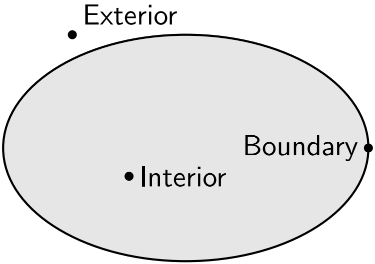
Any intersection of closed sets is closed: if I is any set and for each $i \in I$ the set F_i is closed, so is $\bigcap_{i \in I} F_i$. Any finite union of closed sets is closed: if for each $j = 1, \dots, J$ the set F_j is closed, so is $\bigcup_{j=1}^{J} F_j$.

Proof.

Let $U_i = F_i^c$ and apply $\left(\bigcap_{i \in I} F_i\right)^c = \bigcup_{i \in I} F_i^c$ etc. $\square$

Interior, closure, boundary

- Largest open set included in A is called *interior* of A and is denoted by $\operatorname{int} A$
- Smallest closed set including A is called *closure* of A and is denoted by $\operatorname{cl} A$
- The set $\operatorname{cl} A \setminus \operatorname{int} A$ is called *boundary* of A and is denoted by ∂A



Continuous functions

- Earlier discussion suggests that minimization problem may not have solution if graph of function has "gaps"
- Continuous functions have no gaps in their graphs, which avoids this problem
- It is often convenient to allow function f to take values in extended real numbers $[-\infty, \infty]$ instead of just $\mathbb{R}$
- Example: instead of saying $\log x$ is defined for $x > 0$, it is convenient to define $\log 0 = -\infty$

Continuous functions

- In $\bar{\mathbb{R}} = [-\infty, \infty]$, we declare *open intervals* to be
  - $(a, b) = \{x \in \bar{\mathbb{R}} : a < x < b\}$ for $-\infty \le a < b \le \infty$,
  - $(a, \infty] = \{x \in \bar{\mathbb{R}} : a < x \le \infty\}$ for $-\infty \le a < \infty$, and
  - $[-\infty, b) = \{x \in \bar{\mathbb{R}} : -\infty \le x < b\}$ for $-\infty < b \le \infty$
- We say $\{x_k\}_{k=1}^{\infty} \subset \bar{\mathbb{R}}$ converges to x if

$$(\forall \text{open interval } I \ni x)(\exists K > 0)(\forall k \ge K) \quad x_k \in I$$

- Generalization of previous definitions

Continuous functions

- Let X be normed space and $A \subset X$
- We say $f : A \to [-\infty, \infty]$ is *continuous* at $x_0 \in A$ if

$$(\forall \text{open interval } I \ni f(x_0))(\exists \delta > 0)(\forall x \in A \cap B_\delta(x_0)) \quad f(x) \in I$$

- In words, if $x \in A$ is sufficiently close to x_0 in sense that $\|x - x_0\| < \delta$, then function value $f(x)$ is close to $f(x_0)$ in sense that $f(x)$ is contained in neighborhood I of $f(x_0)$

Proposition

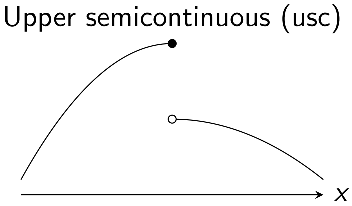
$f : A \to [-\infty, \infty]$ is continuous at $x_0 \in A$ if and only if for any sequence $\{x_k\}_{k=1}^\infty \subset A$ with $x_k \to x_0$, we have $f(x_k) \to f(x_0)$.

Semicontinuous functions

- Sometimes, asking for continuity is too much
- Let X be normed space, $A \subset X$, and $f : A \to [-\infty, \infty]$
- We say f is *upper semicontinous (usc)* at $x_0 \in A$ if

$$(\forall y > f(x_0))(\exists \delta > 0)(\forall x \in A \cap B_\delta(x_0)) \quad f(x) < y$$

- We say f is *lower semicontinuous (lsc)* if $-f$ is usc
- Intuitively, upper (lower) semicontinuous functions are those that function value can suddenly jump upward (downward)



Sequential characterization

Proposition

$f : A \to [-\infty, \infty]$ is upper (lower) semicontinuous at $x_0 \in A$ if and only if for any sequence $\{x_k\}_{k=1}^{\infty} \subset A$ with $x_k \to x_0$, we have $\limsup_{k\to\infty} f(x_k) \le f(x_0)$ ($\liminf_{k\to\infty} f(x_k) \ge f(x_0)$).

Proof.

Similar to continuous functions $\square$

Sequential compactness

- Previous observations suggest solution to minimization problem may not exist if constraint is unbounded or has hole or function is not continuous
- Does solution exist if constraint set closed and bounded and function continuous? Yes!
- We say $S \subset X$ is *sequentially compact* if every sequence in S has subsequence converging to point in S, that is, if $\{x_k\}_{k=1}^{\infty} \subset S$, we can take $x \in S$ and indices $k_1 < k_2 < \cdots$ such that $x_{k_l} \to x \in S$ as $l \to \infty$

Bolzano-Weierstrass theorem

Theorem (Bolzano-Weierstrass theorem)

A set $S \subset \mathbb{R}^N$ is sequentially compact if and only if it is closed and bounded.

Proof.

- If S unbounded, we can take $\{x_k\} \subset S$ such that $\|x_k\| \to \infty$
- Then for any $x \in S$ and subsequence, we have $\|x_{k_l} - x\| \geq \|x_{k_l}\| - \|x\| \to \infty$, so $\{x_{k_l}\}$ does not converge to x; hence S not sequentially compact
- Suppose S closed and bounded; if $N = 1$, can take convergent subsequence by finding $\{x_{k_l}\}$ such that $x_{k_l} \to \limsup x_k$
- For general N, use mathematical induction and pass to subsequence $\square$

Extreme value theorem

Theorem (Extreme value theorem)

Let $\emptyset \neq S \subset \mathbb{R}^N$ be sequentially compact and $f : S \to [-\infty, \infty]$ be lower (upper) semicontinuous. Then f attains a minimum (maximum) over S.

Proof.

- Let $m = \inf_{x \in S} f(x)$
- Take sequence $\{x_k\} \subset S$ such that $f(x_k) \to m$
- Since S is sequentially compact, there is subsequence such that $x_{k_l} \to x$ for some $x \in S$
- Since f is lsc, we obtain

$$m \leq f(x) \leq \liminf_{l \to \infty} f(x_{k_l}) = m,$$

so $f(x) = m$ □

Important points

- Closed sets include boundary; open sets do not
- All norms are equivalent in $\mathbb{R}^N$; use whatever convenient (usually $\ell^1, \ell^2, \ell^\infty$ norms)
- In $\mathbb{R}^N$, bounded sequence has convergent subsequence (Bolzano-Weierstrass); proof is by induction on dimension N
- Extreme value theorem: continuous functions achieve minima and maxima on closed and bounded sets (existence of solution guaranteed)

Chapter 2

One-variable Optimization

Introduction

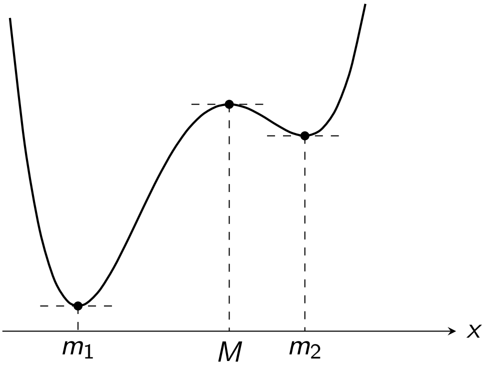
- We would like to solve

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && x \in C \end{aligned}$$

- In practice, we are not only interested in proving existence of solution but also in its characterization
- Some terminology:
  - x is *feasible* if $x \in C$
  - $\bar{x}$ is *(global) solution* if $f(\bar{x}) \le f(x)$ for all $x \in C$
  - $\bar{x} \in C$ is *local solution* if there exists neighborhood $U \subset C$ of x such that $f(\bar{x}) \le f(x)$ for all $x \in U$
  - If inequality strict whenever $x \neq \bar{x}$, then $\bar{x}$ is strict local solution

Local solutions need not be global

- m_1 is global minimum
- m_2 is local minimum but not global minimum
- M is local maximum but not global maximum



Differentiation

- Powerful tool for solving nonlinear optimization problems is *differentiation* (taking derivatives)
- Basically, linear approximation
- Suppose for some p, q, we have

$$f(x) \approx p(x-a)+q$$

- Requiring exact value at $x=a$, get $q=f(a)$
- Solve for p, and require good approximation as $x \to a$:

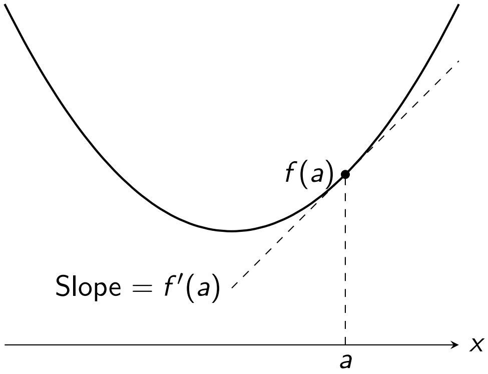
$$p=f'(a) := \lim_{x \to a} \frac{f(x)-f(a)}{x-a}$$

- $f'(a)$ is *derivative* of f at a

First-order approximation

- Hence first-order approximation is

$$f(x) \approx f(a) + f'(a)(x - a)$$



Some terminology

- $f : (a, b) \to \mathbb{R}$ is *differentiable* if $f'(x)$ exists for all $x \in (a, b)$
- If f differentiable and $f'(x)$ continuous in x, we say f is *continuously differentiable* or C^1
- High-order derivatives denoted by f'', f''', etc.
- If f is r times continuously differentiable (so $f, f', f'', \dots, f^{(r)}$ all exist and are continuous), we say f is C^r

Some remarks

- Differentiable functions are continuous
- Continuous functions need not be differentiable (e.g., $f(x) = |x|$)
- Differentiable functions need not have continuous derivatives, for example

$$f(x) = \begin{cases} x^2 \sin(1/x) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0 \end{cases}$$

- Check above example

Necessary condition

- Let $C \subset \mathbb{R}$, and consider

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && x \in C \end{aligned}$$

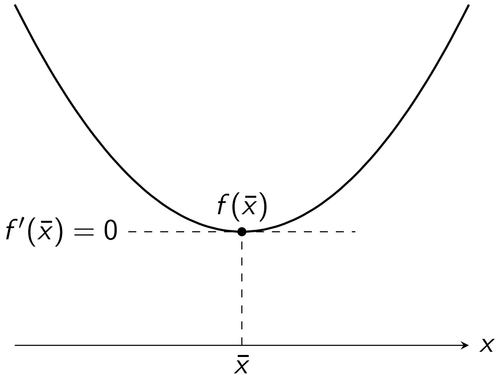
Proposition (Necessity of first-order condition)

If $\bar{x} \in \operatorname{int} C$ is local solution and f is differentiable at $\bar{x}$, then $f'(\bar{x}) = 0$.

Proof

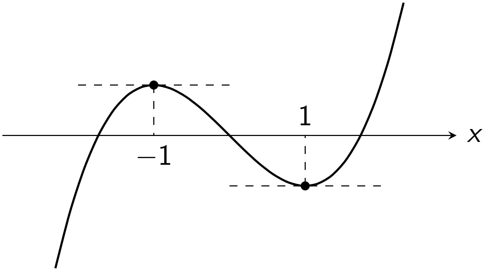
- Since $\bar{x}$ is interior point of C, we have $x + h \in C$ for small enough $|h|$
- Since $\bar{x}$ attains the minimum of f in a neighborhood of $\bar{x}$, we have
$$f(\bar{x} + h) \geq f(\bar{x})$$
for sufficiently small $|h|$
- Subtracting $f(\bar{x})$ from both sides and dividing by $h > 0$, we obtain
$$\frac{f(\bar{x} + h) - f(\bar{x})}{h} \geq 0$$
- Letting $h \to 0$ and using definition of derivative, we get $f'(\bar{x}) \geq 0$
- Reverse inequality similar $\square$

First-order condition is necessary



First-order condition is not sufficient

- Consider $f(x) = x^3/3 - x$
- Then $f'(x) = x^2 - 1 = (x - 1)(x + 1)$, so $f(x) = 0$ at $x = \pm 1$
- But neither point (global) minimum nor maximum



Mean value theorem

Proposition (Mean value theorem)

Let f be continuous on $[a, b]$ and differentiable on (a, b). Then there exists $c \in (a, b)$ such that

$$\frac{f(b) - f(a)}{b - a} = f'(c).$$

Proof.

- Let $\phi(x) := f(x) - f(a) - \frac{f(b)-f(a)}{b-a}(x - a)$
- Then $\phi(a) = \phi(b) = 0$, so achieve some minimum or maximum at $c \in (a, b)$
- Then $\phi'(c) = 0$ implies claim $\square$

Taylor's theorem

- In mean value theorem, changing notation to $b \to x$ and $c \to \xi$, there exists ξ between a and x such that

$$f'(\xi) = \frac{f(x) - f(a)}{x - a} \iff f(x) = f(a) + f'(\xi)(x - a)$$

- Taylor's theorem is generalization: for second order (most useful),

$$f(x) = f(a) + f'(a)(x - a) + \frac{1}{2} f''(\xi)(x - a)^2$$

- More generally, if f is C^n, we can take ξ such that

$$f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(a)}{k!}(x - a)^k + \frac{f^{(n)}(\xi)}{n!}(x - a)^n$$

Sufficient condition

- So far, we have seen that for interior optimum, $f'(\bar{x}) = 0$ is necessary
- Is there simple sufficient condition?
- Yes, if f is convex or concave
- We say f is *convex* if for all x_1, x_2 and $\alpha \in [0, 1]$, we have

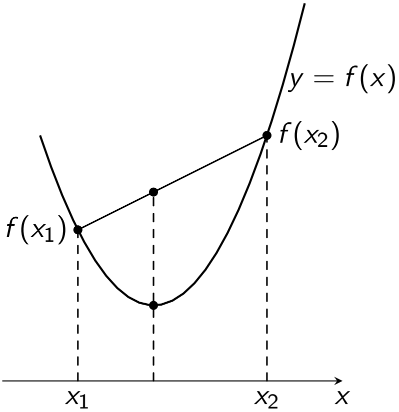
$$f((1-\alpha)x_1 + \alpha x_2) \leq (1-\alpha)f(x_1) + \alpha f(x_2)$$

- Graphically, function is convex if segment joining points $(x_1, f(x_1))$ and $(x_2, f(x_2))$ lies above graph of f
- f is *concave* if inequality flipped:

$$f((1-\alpha)x_1 + \alpha x_2) \geq (1-\alpha)f(x_1) + \alpha f(x_2)$$

Convex function

- f is *convex* if for all x_1, x_2 and $\alpha \in [0, 1]$, we have
$$f((1-\alpha)x_1 + \alpha x_2) \leq (1-\alpha)f(x_1) + \alpha f(x_2)$$
- Can prove: if f is C^2, then convex if and only if $f'' \geq 0$



Sufficiency of first-order condition for convex f

Proposition

Let f be C^2 and convex (concave). If $f'(\bar{x}) = 0$, then $\bar{x}$ is the minimum (maximum) of f.

Proof.

- Suppose f is convex, so $f''(x) \geq 0$
- Applying Taylor's theorem for $n = 2$, for any x there exists ξ such that

$$f(x) = f(\bar{x}) + f'(\bar{x})(x - \bar{x}) + \frac{1}{2} f''(\xi)(x - \bar{x})^2$$

- Since by assumption $f'(\bar{x}) = 0$ and $f''(\xi) \geq 0$, we obtain $f(x) \geq f(\bar{x})$, so $\bar{x}$ is minimum of f
- Same argument for maximum $\square$

Characterization of local solution

Proposition

Let $U \subset \mathbb{R}$ be open and $f : U \to \mathbb{R}$ be C^2. Then following statements are true.

1. *If $\bar{x} \in U$ is a local minimum, then $f'(\bar{x}) = 0$ and $f''(\bar{x}) \geq 0$.*
2. *If $f'(\bar{x}) = 0$ and $f''(\bar{x}) > 0$, then $\bar{x}$ is a strict local minimum.*

Proof.

Similar to convex case $\square$

Optimal savings problem

- We consider example with step-by-step analysis
- Agent lives for two dates indexed by $t = 1, 2$
- At $t = 1$, agent endowed with initial wealth $w > 0$
- Needs to decide how much to consume when gross interest rate is R
- Utility function is

$$U(c_1, c_2) = \frac{c_1^{1-\gamma}}{1-\gamma} + \beta \frac{c_2^{1-\gamma}}{1-\gamma},$$

where $0 < \gamma \neq 1$ is curvature parameter and $\beta > 0$ is discount factor

Optimal savings problem

- Letting $c_1 = c$, savings is $w - c$
- Hence consumption at $t = 2$ is $c_2 = R(w - c)$
- Objective function is

$$\begin{aligned} f(c) &:= \frac{c^{1-\gamma}}{1-\gamma} + \beta \frac{(R(w-c))^{1-\gamma}}{1-\gamma} \\ &= \frac{1}{1-\gamma}\left(c^{1-\gamma} + \beta R^{1-\gamma}(w-c)^{1-\gamma}\right) \end{aligned}$$

- Derivatives are

$$\begin{aligned} f'(c) &= c^{-\gamma} - \beta R^{1-\gamma}(w-c)^{-\gamma}, \\ f''(c) &= -\gamma(c^{-\gamma-1} + \beta R^{1-\gamma}(w-c)^{-\gamma-1}) \end{aligned}$$

Optimal savings problem

- Clearly $f''(c) < 0$, so f is concave
- First-order condition is

$$\begin{aligned} f'(c) = 0 &\iff c^{-\gamma} = \beta R^{1-\gamma}(w-c)^{-\gamma} \\ &\iff c = (\beta R^{1-\gamma})^{-1/\gamma}(w-c) \\ &\iff c = \frac{w}{1+(\beta R^{1-\gamma})^{1/\gamma}} \end{aligned}$$

- Since f concave, first-order condition is sufficient for optimality, so this c is optimal consumption

Important points

- Differentiation is basically linear approximation
- Taylor's theorem allows polynomial approximation of smooth functions ($n = 1, 2$ most useful)
- At interior optimum, $f'(\bar{x}) = 0$ (first-order condition)
- For convex/concave functions, first-order condition is sufficient for optimality

Chapter 3

Multi-variable Unconstrained Optimization

Introduction

- We would like to solve

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && x \in C \end{aligned}$$

- In previous slides, we learned how to do this when $C \subset \mathbb{R}$ and solution is interior
- We now consider case $C \subset \mathbb{R}^N$
- Generalization is conceptually straightforward, but we need to use vectors and matrices to make notation manageable

Linear maps and matrices

- Let $f : \mathbb{R}^N \to \mathbb{R}^M$ be *linear map*, meaning
  1. for each $x \in \mathbb{R}^N$, map f associates vector $f(x) \in \mathbb{R}^M$,
  2. f is linear (preserves addition and scalar multiplication): $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$ for all $x, y \in \mathbb{R}^N$ and $\alpha, \beta \in \mathbb{R}$
- Let $f_m(x)$ be m-th entry of f; then f_m linear, so we can write

$$f_m(x) = a_{m1}x_1 + \cdots + a_{mN}x_N$$

  for some $a_{m1}, \dots, a_{mN}$
- Hence linear map f has one-to-one correspondence with numbers (a_{mn}); we write

$$A = (a_{mn}) = \begin{bmatrix} a_{11} & \cdots & a_{1n} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} & \cdots & a_{mN} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{M1} & \cdots & a_{Mn} & \cdots & a_{MN} \end{bmatrix}$$

  and call it *matrix*

Linear maps and matrices

- If $f : \mathbb{R}^N \to \mathbb{R}^M$ linear, can write $f(x) = Ax$, where $A = (a_{mn})$ is $M \times N$ matrix and

$$(Ax)_m = a_{m1}x_1 + \cdots + a_{mN}x_N = \sum_{n=1}^{N} a_{mn}x_n$$

- $\mathcal{M}_{M,N}(\mathbb{R})$: set of $M \times N$ real matrices, can identify as $\mathbb{R}^{MN}$
- If $M = N$, then A called *square matrix*; then $f : \mathbb{R}^N \to \mathbb{R}^N$ is self map
- $f : \mathbb{R}^M \to \mathbb{R}^N$ defined by $f(x) = 0$ is clearly linear; corresponding matrix is *null matrix* and write $A = 0$
- Identity map $f : \mathbb{R}^N \to \mathbb{R}^N$ defined by $f(x) = x$ also linear; corresponding matrix is *identity matrix* and write $A = I$

Composition of linear maps

- Consider two linear maps $f : \mathbb{R}^N \to \mathbb{R}^M$ and $g : \mathbb{R}^M \to \mathbb{R}^L$
- Since f, g linear, we can find $A = (a_{mn}) \in \mathcal{M}_{M,N}$ and $B = (b_{lm}) \in \mathcal{M}_{L,M}$ such that $f(x) = Ax$ and $g(y) = By$
- We can also consider composition of these two maps, $h = g \circ f$ defined by $h(x) := g(f(x))$
- Easy to see $h : \mathbb{R}^N \to \mathbb{R}^L$ is linear, so can write $h(x) = Cx$ with $C = (c_{ln}) \in \mathcal{M}_{L,N}$
- Using definition $h(x) = g(f(x)) = B(Ax)$, easy to see

$$c_{ln} = \sum_{m=1}^{M} b_{lm} a_{mn}$$

Matrix multiplication

- If $f : \mathbb{R}^N \to \mathbb{R}^M$ and $g : \mathbb{R}^M \to \mathbb{R}^L$ linear, so is $h = g \circ f : \mathbb{R}^N \to \mathbb{R}^L$
- $h(x) = B(Ax)$, so we define *matrix multiplication* by $C = BA$, where

$$c_{ln} = \sum_{m=1}^{M} b_{lm} a_{mn}$$

- Can use all standard rules of algebra such as $B(A_1 + A_2) = BA_1 + BA_2$, $A(BC) = (AB)C$, etc.
- Proofs immediate by carrying out algebra or thinking about linear maps

Inner product

- Identify 1×1 matrix as scalar, so $\mathcal{M}_1(\mathbb{R}) = \mathbb{R}$
- Then for $x, y \in \mathbb{R}^N$, can write inner product as

$$\langle x, y \rangle = x_1 y_1 + \cdots + x_N y_N = \begin{bmatrix} x_1 & \cdots & x_N \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix},$$

product of *row* and *column* vectors

Transpose

- Let $x = (x_m) \in \mathbb{R}^M$, $y = (y_n) \in \mathbb{R}^N$, $A = (a_{mn}) \in \mathcal{M}_{M,N}$
- Then

$$\langle x, Ay \rangle = \sum_{m=1}^{M} x_m \left(\sum_{n=1}^{N} a_{mn} y_n \right) = \sum_{m=1}^{M} \sum_{n=1}^{N} x_m a_{mn} y_n$$

- Right-hand side is also $\langle A'x, y \rangle$, where $A' := (a_{nm}) \in \mathcal{M}_{N,M}$
- A' is called *transpose* of A
- Hence we can write inner product as $\langle x, y \rangle = x'y$
- If matrix product AB defined, then by definition

$$\left\langle (AB)'x, y \right\rangle = \langle x, ABy \rangle = \left\langle A'x, By \right\rangle = \left\langle B'A'x, y \right\rangle,$$

so $(AB)' = B'A'$

Differentiation

- For one-variable function, we defined derivative by

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

- Not useful for multi-variable function, because cannot divide by vector h
- But we can write

$$f(x+h) - f(x) \approx f'(x)h$$

- More precisely,

$$\lim_{|h| \to 0} \frac{|f(x+h) - f(x) - f'(x)h|}{|h|} = 0.$$

Differentiation

- Motivated by this, we say $f : \mathbb{R}^N \to \mathbb{R}^M$ *differentiable* at x if there exists matrix $A \in \mathcal{M}_{M,N}$ such that

$$\lim_{\|h\| \to 0} \frac{\|f(x+h) - f(x) - Ah\|}{\|h\|} = 0$$

- By letting $h = te_n$ and, can show $A = (a_{mn})$ satisfies

$$a_{mn} = \frac{\partial f_m}{\partial x_n}(x) := \lim_{t \to 0} \frac{f_m(x_1, \dots, x_n + t, \dots, x_N) - f_m(x)}{t},$$

- Hence A is matrix of *partial derivatives*; we write $A = Df(x) = (\partial f_m(x)/\partial x_n)$ and call *Jacobian*

Some terminology

- We already defined "differentiable"
- If partial derivatives $\partial f_m(x)/\partial x_n$ exist, we say "f is partially differentiable"
- If f is partially differentiable and partial derivatives are continuous, we say "f is C^1"
- Can prove

$$\begin{aligned} \text{differentiable} &\implies \text{partially differentiable}, \\ C^1 &\implies \text{differentiable} \end{aligned}$$

- C^r means f is r times continuously differentiable
- If f is C^r, order of taking partial derivatives doesn't matter

Chain rule

- For one-variable functions, chain rule is $(g(f(x)))' = g'(f(x))f'(x)$
- We generalize this for multi-variable functions

Proposition

Let $U \subset \mathbb{R}^N$ and $V \subset \mathbb{R}^M$ be open. Let $f : U \to V$ be differentiable at $a \in U$ and $g : V \to \mathbb{R}^L$ be differentiable at $b := f(a) \in V$. Then $g \circ f : U \to \mathbb{R}^L$ defined by $(g \circ f)(x) = g(f(x))$ is differentiable at a and

$$\underbrace{D(g \circ f)(a)}_{L \times N} = \underbrace{Dg(b)}_{L \times M} \underbrace{Df(a)}_{M \times N}.$$

- Intuition: differentiation is linear approximation, and composition of linear maps is matrix product

Gradient

- We would like to solve

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && x \in C \end{aligned}$$

- If f one-variable function, first-order condition was $f'(x) = 0$
- If f partially differentiable, Jacobian is

$$Df(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_N} \end{bmatrix}$$

- Its transpose

$$\nabla f(x) := Df(x)^\top = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_N} \end{bmatrix}$$

and is called *gradient*

Necessary condition

Proposition (Necessity of first-order condition)

If $\bar{x} \in \operatorname{int} C$ *is local solution and* f *is differentiable at* $\bar{x}$*, then* $\nabla f(\bar{x}) = 0$*.*

Proof.

- Take any $v \in \mathbb{R}^N$ and define $\phi : \mathbb{R} \to \mathbb{R}^N$ by $\phi(t) = \bar{x} + vt$
- Since $\bar{x}$ is an interior point of C, the function

$$g(t) := (f \circ \phi)(t) = f(\bar{x} + vt)$$

  is well defined for t close enough to 0
- Since $\bar{x}$ is local solution, clearly $t = 0$ is local minimum of g
- Hence by previous result and chain rule,

$$0 = g'(0) = Df(\bar{x})v = \langle \nabla f(\bar{x}), v \rangle$$

- Since $v \in \mathbb{R}^N$ arbitrary, we obtain $\nabla f(\bar{x}) = 0$ □

Important points

- Differentiation is basically linear approximation
- Chain rule: $D(g \circ f) = (Dg)(Df)$: obvious because differentiation is linear approximation and composition of linear maps is matrix multiplication
- At interior optimum, $\nabla f(\bar{x}) = 0$ (first-order condition)
- We will talk about sufficient conditions much later

Chapter 4

Introduction to Constrained Optimization

Introduction

- We would like to solve

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && x \in C \end{aligned}$$

- In previous slides, we learned how to do this when $C \subset \mathbb{R}^N$ and solution is interior
- Assumption of interior solution unsatisfactory, because constraints bind in most problems

Utility maximization problem

- Consider utility maximization problem

$$\begin{aligned} &\text{maximize} && u(x_1, \dots, x_N) \\ &\text{subject to} && p_1 x_1 + \cdots + p_N x_N \le w, \\ &&& (\forall n) x_n \ge 0 \end{aligned}$$

- Constraint set is

$$C = \left\{ x \in \mathbb{R}_+^N : \langle p, x \rangle \le w \right\},$$

- If agent likes goods, budget constraint $\langle p, x \rangle \le w$ will bind, so $\langle p, x \rangle = w$
- We will learn general approach when constraints can bind

One linear constraint

- To build intuition, we start from one linear constraint
- Problem is

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && \langle a, x \rangle \le c, \end{aligned}$$

  where f: differentiable, $a \neq 0$
- Constraint set is

$$C = \left\{ x \in \mathbb{R}^N : \langle a, x \rangle \le c \right\}.$$

- Suppose $\bar{x} \in C$ is local solution

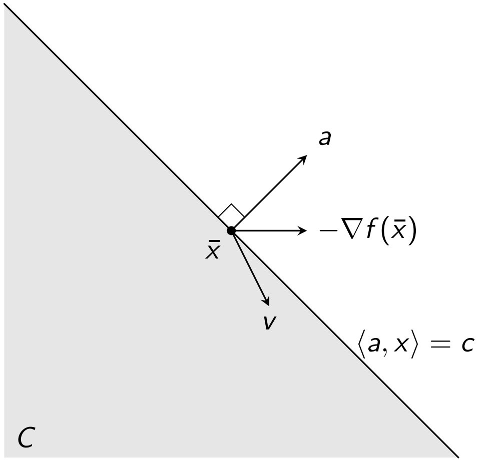
One linear constraint

- If $\langle a, \bar{x} \rangle < c$, then $\bar{x}$ is interior point of C
- Then we already know $\nabla f(\bar{x}) = 0$ is necessary
- Hence assume constraint binds, and $\langle a, \bar{x} \rangle = c$
- Consider moving towards direction v from solution $\bar{x}$
- Since $\bar{x}$ is on boundary, we have $\langle a, \bar{x} \rangle = c$
- Hence point $x = \bar{x} + tv$ is feasible for small $t > 0$ if and only if

$$\langle a, \bar{x} + tv \rangle \leq c = \langle a, \bar{x} \rangle \iff \langle a, v \rangle \leq 0$$

- Hence for feasibility, vectors a, v must form obtuse angle

One linear constraint



Necessary condition

- Since $\bar{x}$ is solution, we have $f(\bar{x}+tv) \geq f(\bar{x})$ for small $t > 0$
- Hence by chain rule,

$$0 \leq \lim_{t \downarrow 0} \frac{f(\bar{x}+tv)-f(\bar{x})}{t} = \langle \nabla f(\bar{x}), v \rangle \iff \langle -\nabla f(\bar{x}), v \rangle \leq 0$$

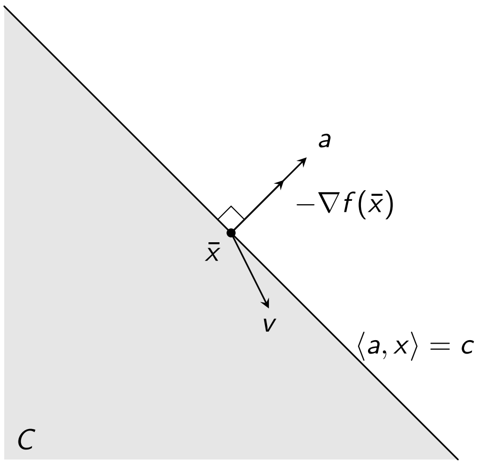
- We obtain following general principle for optimality:

If a and v form obtuse angle, then so do $-\nabla f(\bar{x})$ and v

- Only case $-\nabla f(\bar{x})$ and v form obtuse angle whenever a and v do so is when $-\nabla f(\bar{x})$ and a point to same direction
- Hence there exists $\lambda \geq 0$ such that

$$-\nabla f(\bar{x}) = \lambda a \iff \nabla f(\bar{x}) + \lambda a = 0$$

Necessary condition



Necessary condition with one constraint

Proposition

Consider the optimization problem

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && \langle a, x\rangle \le c, \end{aligned}$$

where $f : \mathbb{R}^N \to \mathbb{R}$ *is differentiable,* $0 \neq a \in \mathbb{R}^N$*, and* $c \in \mathbb{R}$*. If* $\bar{x}$ *is a local solution, then there exists* $\lambda \ge 0$ *such that*

$$\nabla f(\bar{x}) + \lambda a = 0.$$

Multiple linear constraints

- We next consider optimization problem

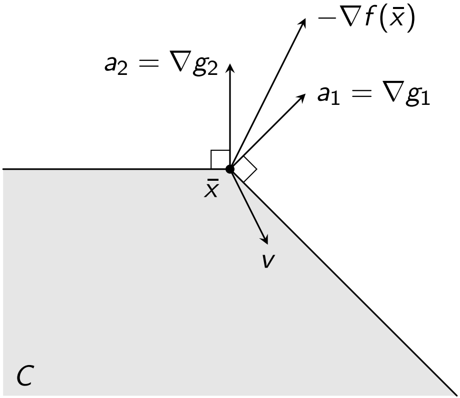
$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && \langle a_1, x\rangle \le c_1, \\ & && \langle a_2, x\rangle \le c_2, \end{aligned}$$

where f differentiable, $a_1, a_2 \neq 0$, and c_1, c_2 are constants
- Let $\bar{x}$ be local solution
- Constraint set is

$$C = \{x : g_1(x) \le 0, g_2(x) \le 0\},$$

where $g_i(x) = \langle a_i, x\rangle - c_i$ for $i = 1, 2$ are *affine*
- Assume both constraints are active (bind) at solution

Multiple linear constraints



Necessary condition with two constraints

- Principle

  If a_i and v form obtuse angle, then so do $-\nabla f(\bar{x})$ and v

  still valid
- By looking at picture, for $\bar{x}$ to be solution, it is necessary that $-\nabla f(\bar{x})$ lies between a_1 and a_2
- This is true if and only if there are numbers $\lambda_1, \lambda_2 \geq 0$ such that

$$\begin{aligned} & -\nabla f(\bar{x}) = \lambda_1 a_1 + \lambda_2 a_2 \\ \iff & \nabla f(\bar{x}) + \lambda_1 \nabla g_1(\bar{x}) + \lambda_2 \nabla g_2(\bar{x}) = 0 \end{aligned}$$

Karush-Kuhn-Tucker theorem

Theorem (KKT theorem with linear constraints)

Consider the optimization problem

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && g_i(x) \le 0 && (i = 1, \dots, I), \end{aligned}$$

where f is differentiable and $g_i(x) = \langle a_i, x \rangle - c_i$ is affine with $a_i \neq 0$. If $\bar{x}$ is a local solution, then there exist Lagrange multipliers $\lambda_1, \dots, \lambda_I$ such that

$$\begin{aligned} &\textit{(First-order condition)} && \nabla f(\bar{x}) + \sum_{i=1}^{I} \lambda_i \nabla g_i(\bar{x}) = 0, \\ &\textit{(Complementary slackness)} && (\forall i)\ \lambda_i \ge 0,\ g_i(\bar{x}) \le 0,\ \lambda_i g_i(\bar{x}) = 0. \end{aligned}$$

Remembering KKT theorem

1. Express problem as

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && g_i(x) \le 0 && (i = 1, \dots, I), \end{aligned}$$

2. Define Lagrangian

$$L(x, \lambda) := f(x) + \sum_{i=1}^{I} \lambda_i g_i(x)$$

3. Pretend taking unconstrained FOC, and derive

$$0 = \nabla L(x, \lambda) = \nabla f(x) + \sum_{i=1}^{I} \lambda_i \nabla g_i(x)$$

4. Complementary slackness is just $\lambda_i g_i(x) = 0$ for all i

William Karush (1917-1997)

- A version of KKT theorem appeared in 1939 master's thesis (U of Chicago) of William Karush, who became teaching prof at California State U
- Received no attention, because applied mathematics gained respect only after World War II
- Rediscovered by Princeton profs Harold Kuhn (1925-2014) and Albert Tucker (1905-1995) in 1950 conference paper, so often called "Kuhn-Tucker theorem"
- We should obviously call Karush-Kuhn-Tucker theorem

Constrained maximization

- What if problem is maximization

$$\begin{aligned} &\text{maximize} && f(x) \\ &\text{subject to} && g_i(x) \geq 0 && (i = 1, \dots, I)? \end{aligned}$$

- Append minus sign to convert to minimization:

$$\begin{aligned} &\text{minimize} && -f(x) \\ &\text{subject to} && -g_i(x) \leq 0 && (i = 1, \dots, I) \end{aligned}$$

- Then KKT conditions are

$$-\nabla f(\bar{x}) - \sum_{i=1}^{I} \lambda_i \nabla g_i(\bar{x}) = 0,$$
$$(\forall i)\ \lambda_i(-g_i(\bar{x})) = 0,$$

same as minimization after putting minus sign!

Tips for formulating problems

- For minimization problems, use format

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && g(x) \le 0 \end{aligned}$$

- For maximization problems, use format

$$\begin{aligned} &\text{maximize} && f(x) \\ &\text{subject to} && g(x) \ge 0 \end{aligned}$$

- In either case, Lagrangian is $L(x, \lambda) = f(x) + \lambda g(x)$ with $\lambda \ge 0$
- First-order condition is $\nabla_x L(x, \lambda) = 0$
- Always stick to this convention to avoid stupid mistakes

Utility maximization problem

- As application and illustration of KKT theorem, we provide step-by-step analysis of utility maximization problem
- Consider

$$\begin{aligned} &\text{maximize} && u(x) = \alpha \log x_1 + (1-\alpha) \log x_2 \\ &\text{subject to} && p_1 x_1 + p_2 x_2 \le w, \\ &&& x_1, x_2 \ge 0 \end{aligned}$$

- Here
  - $\alpha \in (0, 1)$ is preference parameter,
  - $p_1, p_2 > 0$ are prices of goods,
  - $w > 0$ is disposable income of agent

Existence of solution

- Define constraint set by

$$C := \left\{x \in \mathbb{R}^2_+ : p_1x_1 + p_2x_2 \le w\right\}$$

- Clearly C is nonempty, closed, and bounded
- $u : \mathbb{R}^2_+ \to [-\infty, \infty)$ is continuous
- Hence by extreme value theorem, solution $\bar{x}$ exists
- If $\bar{x}_1 = 0$ or $\bar{x}_2 = 0$, we have $u(\bar{x}) = -\infty$, which is clearly not optimum
- Hence $\bar{x} \gg 0$

Formulating problem

- Because it is maximization problem, we need to convert to format

$$\begin{aligned} &\text{maximize} && f(x) \\ &\text{subject to} && g(x) \geq 0 \end{aligned}$$

- Thus problem is

$$\begin{aligned} &\text{maximize} && \alpha \log x_1 + (1-\alpha) \log x_2 \\ &\text{subject to} && w - p_1 x_1 - p_2 x_2 \geq 0, \\ &&& x_1 \geq 0 \\ &&& x_2 \geq 0 \end{aligned}$$

Deriving KKT conditions

- Define Lagrangian by

$$\begin{aligned} L(x,\lambda,\mu) = \alpha \log x_1 + (1-\alpha)\log x_2 \\ + \lambda(w - p_1x_1 - p_2x_2) + \mu_1x_1 + \mu_2x_2 \end{aligned}$$

- First-order conditions are

$$\begin{aligned} 0 &= \frac{\partial L}{\partial x_1} = \frac{\alpha}{x_1} - \lambda p_1 + \mu_1, \\ 0 &= \frac{\partial L}{\partial x_2} = \frac{1-\alpha}{x_2} - \lambda p_2 + \mu_2 \end{aligned}$$

- Complementary slackness conditions are

$$\begin{aligned} &\lambda(w - p_1x_1 - p_2x_2) = 0, \\ &\mu_1x_1 = \mu_2x_2 = 0 \end{aligned}$$

Solving KKT conditions

- Since we argued $\bar{x} \gg 0$, complementary slackness implies $\mu_1 = \mu_2 = 0$
- Solving for first-order condition, get $x_1 = \frac{\alpha}{\lambda p_1}$, $x_2 = \frac{1-\alpha}{\lambda p_2}$
- Substituting into remaining complementary slackness condition, get

$$\frac{\alpha}{\lambda} + \frac{1-\alpha}{\lambda} = w \iff \lambda = \frac{1}{w}$$

- Therefore

$$(x_1, x_2) = \left(\frac{\alpha w}{p_1}, \frac{(1-\alpha)w}{p_2}\right)$$

- We know solution exists, and we arrived at unique candidate using only necessary condition, so this must be (unique) solution

Nonnegativity constraints

- In many economic applications such as UMP, some constraints are nonnegative: $x \geq 0$
- In previous example, we used $\log 0 = -\infty$ to rule out solutions of form $x_1 = 0$ or $x_2 = 0$
- We seek to provide more general sufficient condition for dropping nonnegativity constraints in

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && x \in C \end{aligned}$$

Dropping nonnegativity constraints

Proposition

Let $f:\mathbb{R}^N_+\to(-\infty,\infty]$ *be continuous and* $C\subset\mathbb{R}^N_+$*. Suppose that*

1. C *is a convex set, so* $x_1,x_2\in C$ *and* $t\in[0,1]$ *imply* $(1-t)x_1+tx_2\in C$*; furthermore, there exists* $x_0\gg 0$ *such that* $x_0\in C$*,*
2. f *is differentiable on* $\mathbb{R}^N_{++}$ *with partial derivatives that are uniformly bounded above, so there exists* $b\geq 0$ *such that* $\max_n\sup_{x\in C}\frac{\partial f}{\partial x_n}\leq b$*,*
3. f *satisfies the Inada condition with respect to* x_n*, so*

$$\lim_{y\to x}\frac{\partial f}{\partial x_n}(y)=-\infty$$

whenever $x=(x_1,\dots,x_N)$ *satisfies* $x_n=0$*.*

If $\bar{x}\in C$ *is a solution, then* $\bar{x}_n>0$*.*

Proof

- Since C is convex and $x_0 \in C$, we may define $g : [0,1] \to (-\infty, \infty]$ by $g(t) = f(x(t))$, where $x(t) := (1-t)\bar{x} + tx_0$
- By assumption, g is continuous on $[0,1]$ and differentiable on $(0,1]$
- Applying chain rule and using uniform boundedness, we get

$$\begin{aligned} g'(t) &= \sum_{n=1}^{N} \frac{\partial f}{\partial x_n}(x(t))(x_{0n} - \bar{x}_n) \\ &\le \frac{\partial f}{\partial x_n}(x(t))(x_{0n} - \bar{x}_n) + (N-1)b \left\| x_0 - \bar{x} \right\|, \end{aligned}$$

  where $\left\| \cdot \right\|$ is supremum (l^∞) norm

Proof

- If $\bar{x}_n = 0$, then $x_{0n} - \bar{x}_n > 0$, so letting $t \downarrow 0$ and using Inada condition, we obtain $\lim_{t \downarrow 0} g'(t) = -\infty$
- In particular, $g'(t) < 0$ for sufficiently small t
- By mean value theorem, we can take $s \in (0, t)$ such that

$$\begin{aligned} & g(t) - g(0) = g'(s)(t - 0) = g'(s)t < 0 \\ \implies & f(x(t)) = g(t) < g(0) = f(\bar{x}), \end{aligned}$$

which is contradiction $\square$

Important points

- Consider

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && g_i(x) \le 0 && (i = 1, \dots, I) \end{aligned}$$

- KKT theorem: if $\bar{x}$ local solution, then

$$\text{(First-order condition)} \qquad \nabla f(\bar{x}) + \sum_{i=1}^{I} \lambda_i \nabla g_i(\bar{x}) = 0,$$

$$\text{(Complementary slackness)} \qquad (\forall i)\ \lambda_i \ge 0,\ g_i(\bar{x}) \le 0,\ \lambda_i g_i(\bar{x}) = 0$$

- One of most important theorems in economics
- For maximization, remember to flip inequality for constraint: $g_i(x) \ge 0$

Chapter 5

Vector Space, Matrix, and Determinant

Vector space

Solving linear equations

Determinant

Vector space

- Roughly speaking, *vector space* is set on which addition and scalar multiplication are defined
- Thus if V is vector space, for each vector $v, w \in \mathsf{V}$, there corresponds sum
$$v + w \in \mathsf{V},$$
and for each $v \in \mathsf{V}$ and scalar α, there corresponds scalar multiplication
$$\alpha v \in \mathsf{V}$$
- By "scalar", for practical purposes we use either set of real numbers $\mathbb{R}$ or set of complex numbers $\mathbb{C}$
- See standard textbooks for precise axioms

Examples of vector spaces

- Typical example of vector space is N-dimensional Euclidean space $\mathbb{R}^N$
- Other examples are

$$\begin{aligned}
\mathrm{V}_1 &:= \{v : \mathbb{R} \to \mathbb{R} : v \text{ is a continuous function}\}, \\
\mathrm{V}_2 &:= \{v : \mathbb{R} \to \mathbb{R} : v \text{ is a bounded continuous function}\}, \\
\mathrm{V}_3 &:= \{v : \mathbb{R} \to \mathbb{R} : v \text{ is a polynomial}\}, \\
\mathrm{V}_4 &:= \{v : \mathbb{R} \to \mathbb{R} : v \text{ is a polynomial of degree} \le N-1\},
\end{aligned}$$

  etc., where addition and scalar multiplication of functions are defined pointwise
- If subset $\mathrm{W} \subset \mathrm{V}$ is itself vector space, we say W is *subspace* of V
- Obviously, $\mathrm{V}_2, \mathrm{V}_3$ are subspaces of V_1 and V_4 is subspace of V_3

Linear combination, span

- If $v_1, \dots, v_K \in \mathrm{V}$ and $\alpha_1, \dots, \alpha_K \in \mathbb{R}$, then

$$v := \alpha_1 v_1 + \cdots + \alpha_K v_K = \sum_{k=1}^{K} \alpha_k v_k \in \mathrm{V}$$

  is *linear combination* of $\{v_k\}$ with coefficients $\{\alpha_k\}$
- The set

$$\operatorname{span}[v_1, \dots, v_K] := \left\{ v = \sum_{k=1}^{K} \alpha_k v_k : (\forall k) \alpha_k \in \mathbb{R} \right\}$$

  is *span* of $\{v_k\}$
- If $\operatorname{span}[v_1, \dots, v_K] = \mathrm{V}$, we say $\{v_k\}$ *spans* V
- If V has finite set of vectors $\{v_k\}$ that spans V, we say V is *finite dimensional*

Linear independence, dimension

- Set of vectors $\{v_k\}$ is *linearly independent* if

$$\sum_{k=1}^{K} \alpha_k v_k = 0 \implies (\forall k)\alpha_k = 0$$

- Otherwise ($\sum_{k=1}^{K} \alpha_k v_k = 0$ for nontrivial $\{\alpha_k\}$), *linearly dependent*
- If $\{v_k\}_{k=1}^{K}$ linearly independent and spans V, we say $\{v_k\}$ is *basis* of V
- K is *dimension* of V and we write $\dim \mathrm{V} = K$
- Clearly $\dim \mathbb{R}^N = N$

Maps

- Let V, W be general sets
- $\phi : \mathrm{V} \to \mathrm{W}$ is *one-to-one* (or *injective*) if $v_1 \neq v_2 \implies \phi(v_1) \neq \phi(v_2)$
- ϕ is *onto* (or *surjective*) if for all $w \in \mathrm{W}$, there exists $v \in \mathrm{V}$ such that $\phi(v) = w$
- If ϕ is both one-to-one and onto, we say it is *bijective*
- If ϕ bijective, then for each $w \in \mathrm{W}$, there exists unique $v \in \mathrm{V}$ such that $\phi(v) = w$, which we denote as $v = \phi^{-1}(w)$
- The map $\phi^{-1} : \mathrm{W} \to \mathrm{V}$ is called *inverse* of ϕ

Isomorphism

- Roughly speaking, when bijective map $\phi : \mathrm{V} \to \mathrm{W}$ preserves properties that we are interested in, we call it *isomorphism*
- If V, W are vector spaces (which are characterized by linearity), bijection $\phi : \mathrm{V} \to \mathrm{W}$ is isomorphism if it is linear:

$$\phi(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 \phi(v_1) + \alpha_2 \phi(v_2)$$

- Two sets that are isomorphic can be regarded as identical, as long as we are concerned with properties that we are interested in
- Can show any N-dimensional (real) vector space is isomorphic to $\mathbb{R}^N$
- For example, space of polynomials with degree $\le N-1$ is isomorphic to $\mathbb{R}^N$ through

$$v(x) = \sum_{n=1}^{N} \alpha_n x^{n-1} \longleftrightarrow \alpha = (\alpha_1, \dots, \alpha_N) \in \mathbb{R}^N$$

Solving linear equations

- In practice, we often want to solve $Ax = b$
- If we define $\phi : \mathbb{R}^N \to \mathbb{R}^N$ by $\phi(x) = Ax$, then can write $\phi(x) = b$
- If ϕ bijective, we may solve $x = \phi^{-1}(b)$
- Clearly ϕ^{-1} linear, so has matrix representation denoted by A^{-1}, called *inverse* of A
- Thus $x = A^{-1}b$
- But argument vacuous unless we know how to compute

Solving linear equations

- If $N = 1$, can solve $ax = b$ as $x = b/a$ if $a \neq 0$
- If $N = 2$, $Ax = b$ is

  $$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= b_1, \\ a_{21}x_1 + a_{22}x_2 &= b_2, \end{aligned}$$

  and we can solve by eliminating one variable from two equations
- This process involves *elementary row operations*
  1. swapping two equations,
  2. multiplying equation by nonzero scalar, and
  3. adding scalar multiple of equation to another

Swapping equations

- Let $P = I$ (identity matrix), and define $P(i,j) = (p_{mn})$ by setting $p_{ii} = p_{jj} = 0$ and $p_{ij} = p_{ji} = 1$ in P
- For instance, if $N = 3$ and $(i,j) = (2,3)$, we have
$$P(i,j) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$
- Then swapping rows i and j of $Ax = b$ corresponds to
$$P(i,j)Ax = P(i,j)b$$
- Note that $P(i,j)^2 = I$, so multiplying $P(i,j)$ from left, we recover $Ax = b$, so these equations are equivalent

Multiplying equation

- Let $Q = I$, and define $Q(i; c) = (q_{mn})$ by setting $q_{ii} = c$ in Q
- For instance, if $N = 3$ and $i = 2$, we have

$$Q(i; c) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- Then multiplying row i of $Ax = b$ by $c \neq 0$ corresponds to

$$Q(i; c)Ax = Q(i; c)b$$

- Note that $Q(i; 1/c)Q(i; c) = I$, so multiplying $Q(i; 1/c)$ from left, we recover $Ax = b$

Adding scalar multiple of equation

- Let $R = I$, and define $R(i,j;c) = (r_{mn})$ by setting $r_{ij} = c$ in R
- For instance, if $N = 3$ and $(i,j) = (2,3)$, we have

$$R(i,j;c) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix}$$

- Then adding c times row j of $Ax = b$ to row i corresponds to

$$R(i,j;c)Ax = R(i,j;c)b$$

- Note that $R(i,j;-c)R(i,j;c) = I$, so multiplying $R(i,j;-c)$ from left, we recover $Ax = b$

Gaussian elimination

- Multiplying P, Q, R matrices from left leaves equation equivalent
- Find (i,j) such that $a_{ij} \neq 0$; if $j \neq 1$, consider equation $AP(1,j)P(1,j)x = b$
- By redefining $AP(1,j)$ as A and $P(1,j)x$ as x (swapping x_1 and x_j), we may assume $a_{i1} \neq 0$ for some i
- If $i \neq 1$, consider equation $P(i,1)Ax = P(i,1)b$; by redefining $P(i,1)A$ as A and $P(i,1)b$ as b (swapping rows 1 and i), we may assume $a_{11} \neq 0$
- Multiply $Q(1,1;1/a_{11})$ from left to $Ax = b$; then we may assume $a_{11} = 1$

Gaussian elimination

- For each $m = 2, \dots, N$, multiply $R(m, 1; -a_{m1})$ from left to $Ax = b$; then we may assume $a_{m1} = 0$ for all $m > 1$
- System of equations can now be written as

$$\begin{bmatrix} 1 & A_{12} \\ 0 & \tilde{A} \end{bmatrix} \begin{bmatrix} x_1 \\ \tilde{x} \end{bmatrix} = \begin{bmatrix} b_1 \\ \tilde{b} \end{bmatrix},$$

- Continuing this process, we may write $Ax = b$ equivalently as

$$(LAP)Px = Lb,$$

where L is product of finitely many $P(i,j)$, $Q(i;c)$, $R(i,j;c)$ matrices, P is product of finitely many $P(i,j)$ matrices, and

$$LAP = \begin{bmatrix} I_r & B \\ 0_{N-r,r} & 0_{N-r,N-r} \end{bmatrix}$$

for some $0 \le r \le N$ and $B \in \mathcal{M}_{r,N-r}$

Gaussian elimination

- Write $y = Px$, $c = Lb$, and partition $(LAP)Px = Lb$ as

$$\begin{bmatrix} I & B \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix},$$

  which is equivalent to $y_1 + By_2 = c_1$ and $c_2 = 0$
- Therefore, there exists solution if and only if $c_2 = 0$, in which case solution takes form $y_1 = c_1 - By_2$ for any $y_2 \in \mathbb{R}^{N-r}$
- There exists unique solution if and only if $r = N$, in which case $y = Px = Lb \iff x = PLb$ (because $P^2 = I$)
- Number r in Gaussian elimination algorithm is called *rank* of matrix A (which is uniquely determined by A)

Determinant

- Although Gaussian elimination is practical for computational purposes, it does not provide theoretical insights
- We define *determinant* of square matrices
- $A \in \mathcal{M}_N$ can be written as $A = [a_1, \dots, a_N]$
- Consider function $D : \mathcal{M}_N \to \mathbb{R}$ satisfying
    1. (Multi-linearity) For each n, $D(\dots, x_n, \dots)$ is linear in $x_n \in \mathbb{R}^N$: for all $x_n, y_n \in \mathbb{R}^N$ and $\alpha, \beta \in \mathbb{R}$, we have
    $$D(\dots, \alpha x_n + \beta y_n, \dots) = \alpha D(\dots, x_n, \dots) + \beta D(\dots, y_n, \dots)$$
    2. (Alternation) For each $m < n$, sign of D flips whenever we flip columns m, n:
    $$D(\dots, x_m, \dots, x_n, \dots) = -D(\dots, x_n, \dots, x_m, \dots)$$
    3. (Normalization) $D(I) = 1$

Determinant

- It turns out that these properties uniquely determine D
- For $N = 1$, we can write $A = (a)$ (scalar), so it must be

$$D(A) = D(a) = D(aI) = aD(I) = a$$

- For general N we need a few lemmas

Lemma
If A has two identical columns, then $D(A) = 0$

Proof.
By flipping two identical columns,

$$D(A) = D(\dots, a, \dots, a \dots) = -D(\dots, a, \dots, a \dots) = -D(A),$$

so $D(A) = 0$ $\square$

Determinant

Lemma

If columns of A are linearly dependent, then $D(A) = 0$

Proof.

- By assumption, there is nontrivial linear combination

$$\sum_{n=1}^{N} \alpha_n a_n = 0$$

- $\alpha_j \neq 0$ for some j, so we may write $a_j = -\frac{1}{\alpha_j} \sum_{n \neq j} \alpha_n a_n$ for some j
- Using multi-linearity and previous lemma, get $D(A) = 0$ $\square$

The case $N = 2$

- Suppose $N = 2$ and $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$
- Using properties of D, we get

$$\begin{aligned}
&D(A) \\
&= aD\begin{pmatrix} 1 & b \\ 0 & d \end{pmatrix} + cD\begin{pmatrix} 0 & b \\ 1 & d \end{pmatrix} \\
&= abD\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} + adD\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + bcD\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + cdD\begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \\
&= adD\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - bcD\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\
&= ad - bc,
\end{aligned}$$

so D uniquely determined

Laplace expansion formula

- General case proceeds by induction
- Let $A = (a_{mn}) \in \mathcal{M}_N$. For fixed i, define

$$D_N(A) = \sum_{m=1}^{N} (-1)^{m+i} a_{mi} D_{N-1}(A_{mi}),$$

  where A_{mi} is $(N-1) \times (N-1)$ submatrix of A obtained by removing row m and column i
- We can show $D_N(A)$ does not depend on i and is unique function satisfying three properties
- Unique value $D(A)$ is called *determinant* of A and is denoted by $\det A$ or $|A|$

Laplace expansion formula

- For $N = 2$ and $i = 1$, we may compute

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = a(d) - c(b) = ad - bc$$

- For $N = 3$ and $i = 1$, we may compute

$$\begin{aligned} \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} &= a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - d \begin{vmatrix} b & c \\ h & i \end{vmatrix} + g \begin{vmatrix} b & c \\ e & f \end{vmatrix} \\ &= a(ei - fh) - d(bi - ch) + g(bf - ce), \end{aligned}$$

etc.

Dropping normalization

Lemma

If $F : \mathcal{M}_N \to \mathbb{R}$ satisfies multi-linearity and alternation, then $F(A) = |A| \, F(I)$.

Proof.

- Repeatedly using multi-linearity and alternation as we did for 2×2 case, we may write $F(A) = g(A)F(I)$ for some function g independent of F
- If $F(I) = 1$, then by uniqueness it must be $F = \det$, so $g(A) = \det A = |A|$
- Hence $F(A) = |A| \, F(I)$ □

Determinant of product

Proposition

If $A, B \in \mathcal{M}_N$, then $|AB| = |A|\,|B| = |BA|$.

Proof.

- Fix $A \in \mathcal{M}_N$ and define $F : \mathcal{M}_N \to \mathbb{R}$ by $F(X) = |AX|$
- Using linearity of $X \mapsto AX$, we can see that F satisfies multi-linearity and alternation
- Hence by previous lemma, we obtain

$$|AX| = F(X) = |X|\,F(I) = |X|\,|A| = |A|\,|X|$$

- Setting $X = B$, we obtain $|AB| = |A|\,|B|$
- Interchanging role of A, B, get $|BA| = |B|\,|A| = |A|\,|B| = |AB|$ $\square$

Block matrices

- We may write matrices in blocks, for example

$$A=\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

- *Block upper triangular*:

$$A=\begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$$

- *Block diagonal*:

$$A=\begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}$$

Determinant of block upper triangular matrix

Proposition

If A is block upper triangular, then $|A| = |A_{11}| \, |A_{22}|$.

Proof.

- Let $A_{11} \in \mathcal{M}_r$; for general matrix $X \in \mathcal{M}_r$, define

$$F(X) = \begin{vmatrix} X & A_{12} \\ 0 & A_{22} \end{vmatrix}$$

- Then F satisfies multi-linearity and alternation, so $F(X) = |X| \, F(I)$
- Hence suffices to show $F(I) = |A_{22}|$

Determinant of block upper triangular matrix

Proof.

- Now
$$F(I)=\begin{vmatrix} I & A_{12} \\ 0 & A_{22} \end{vmatrix}=\begin{vmatrix} I & 0 \\ 0 & A_{22} \end{vmatrix}$$
by subtracting some multiples of first r columns from last $N-r$ columns
- If we view last expression as function of A_{22}, all properties of D satisfied, so $F(I)=|A_{22}|$ □

Determinant of upper triangular matrix

- We say square matrix $A=(a_{mn})$ is *upper triangular* if $a_{mn}=0$ whenever $m>n$, so A can be written as

$$A=\begin{bmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_{NN} \end{bmatrix}$$

- Obviously, upper triangular matrix is block upper triangular with N diagonal blocks of size 1×1
- Hence $|A|=a_{11}\cdots a_{NN}$: determinant is product of diagonal entries

Formula for inverse matrix

- Let $A = (a_{mn})$ be square matrix
- Let A_{mn} be submatrix of A obtained by removing row m and column n
- Then $c_{mn} := (-1)^{m+n} |A_{mn}|$ is called (m, n) *cofactor* of A
- The matrix $C = (c_{mn})$ is called *cofactor matrix*

Proposition

Let A be square matrix and C be cofactor matrix. Then A is invertible if and only if $|A| \neq 0$, in which case $A^{-1} = \frac{1}{|A|} C'$.

Proof

- By definition of cofactor, for each i Laplace expansion formula implies
$$|A| = \sum_{m=1}^{N} a_{mi} c_{mi}$$
- Let $A[i \leftarrow j]$ be matrix obtained by replacing column i with column j
- If $i \neq j$, since column j appears twice in $A[i \leftarrow j]$, we have
$$0 = |A[i \leftarrow j]| = \sum_{m=1}^{N} a_{mj} c_{mi}$$

Proof

- Define Kronecker's delta by $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$
- Combining cases $i = j$ (hence $A[i \leftarrow j] = A$) and $i \neq j$, we obtain
$$\delta_{ij} |A| = \sum_{m=1}^{N} c_{mi} a_{mj}$$
- Collecting terms into a matrix, we obtain $|A| I = C'A$
- Therefore if $|A| \neq 0$, then A is invertible and claim holds
- Conversely, if A is invertible, then $1 = |I| = \left|AA^{-1}\right| = |A| \left|A^{-1}\right|$, so it must be $|A| \neq 0$ $\square$

2×2 case

- Let A be 2×2 and
$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$
- Cofactor matrix is
$$C = \begin{bmatrix} d & -c \\ -b & a \end{bmatrix}$$
- Inverse matrix is
$$A^{-1} = \frac{1}{|A|} C' = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Equivalent conditions of invertibility

Theorem

Let A be a square matrix. Then the following conditions are equivalent.

1. *A is invertible.*
2. *The column vectors of A are linearly independent.*
3. *For any b, the equation $Ax = b$ has a unique solution.*
4. *A has full rank.*
5. $|A| \neq 0$.

Order of operations

- How many operations are required to solve $Ax = b$?
- With Gaussian elimination, for each i and $m \neq i$, we subtract constant multiple of row i from row m, which involves N numbers; repeating this for each m and iterating over i, order of operations is $N \times N \times N = N^3$
- If we use Gaussian elimination to compute A^{-1} first (so applying Gaussian elimination to $b = e_n$ for each n) and compute $x = A^{-1}b$, order of operations is $N^3 \times N = N^4$
- With Laplace expansion to compute $|A|$, letting $o(n)$ be order for computing determinant of $A \in \mathcal{M}_n$, then Laplace expansion implies $o(n) = no(n-1)$, so $o(n) = n!$; thus computing A^{-1} requires $N^2 \times (N-1)! \sim (N+1)!$ operations
- Hence Laplace expansion is impractical

Chapter 6

Spectral Theory

Introduction

- In economic analysis, we often want to know behavior of matrix power A^k as $k \to \infty$
- For instance, linearization of economic models often imply dynamics
$$x_t = Ax_{t-1} + u_t,$$
where x_t is vector of state variables, A is square matrix, and u_t is vector of shocks
- Iterating this, we obtain
$$x_t = u_t + Au_{t-1} + \cdots + A^{t-1}u_1 + A^t x_0$$
- Thus if $\lim_{t\to\infty} A^t = 0$, then $A^t x_0 \to 0$, so initial condition becomes irrelevant as time goes by

Analysis for diagonal matrix

- We say square matrix $A = (a_{mn})$ is *diagonal* if $a_{mn} = 0$ whenever $m \neq n$, so we can write

$$A = \operatorname{diag}[d_1, \dots, d_N] := \begin{bmatrix} d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_N \end{bmatrix}$$

- If A is diagonal, straightforward calculation shows $A^k = \operatorname{diag}[d_1^k, \dots, d_N^k]$ for all $k \in \mathbb{N}$
- Hence $A^k \to 0$ as $k \to \infty$ if and only if $|d_n| < 1$ for all n
- We generalize this argument for any square matrix

Eigenvalue and eigenvector

- Let A be square matrix (real or complex)
- If there is vector $v \neq 0$ and scalar α such that $Av = \alpha v$, we say α is *eigenvalue* of A and v is *eigenvector* corresponding to α
- If $Av = \alpha v$, by iteration we may compute $A^k v = \alpha^k v$, so we can easily understand behavior of $A^k v$ as $k \to \infty$

Characterization of eigenvalues

- By definition, α is eigenvalue if and only if there exists $v \neq 0$ such that
$$Av = \alpha v \iff (\alpha I - A)v = 0$$
- By previous results, such $v \neq 0$ exists if and only if $|\alpha I - A| = 0$
- For any complex number $z \in \mathbb{C}$, define function $\Phi_A : \mathbb{C} \to \mathbb{C}$ by $\Phi_A(z) = |zI - A|$
- Then by applying Laplace expansion of determinant and induction, Φ_A is polynomial of degree N with leading coefficient 1
- By fundamental theorem of algebra, $\Phi_A(z) = 0$ has exactly N roots if we count multiplicity, so any $A \in \mathcal{M}_N(\mathbb{C})$ has exactly N eigenvalues
- Polynomial Φ_A is called *characteristic polynomial* of A

2×2 case

- Let A be 2×2 and
$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$
- Then
$$\begin{aligned} \Phi_A(z) = |zI - A| &= \begin{vmatrix} z-a & -b \\ -c & z-d \end{vmatrix} \\ &= z^2 - (a+d)z + ad - bc \end{aligned}$$

Eigenvalues need not be real

- Even if A is real matrix, eigenvalues (hence eigenvectors) need not be real
- Example: let
$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix},$$
- Then characteristic polynomial and roots are
$$z^2 - 2(\cos\theta)z + 1 = 0 \iff z = \cos\theta \pm i\sin\theta,$$
which are complex whenever $\sin\theta \neq 0$
- Hence when we discuss eigenvalues and eigenvectors, we always consider complex vector space $\mathbb{C}^N$ unless otherwise specified

Eigenvalues of upper triangular matrix

Proposition

If $A = (a_{mn})$ is upper triangular (so $a_{mn} = 0$ whenever $m > n$), then the eigenvalues of A are the diagonal entries $a_{11}, \dots, a_{NN}$.

Proof.

- If A is upper triangular, so is $zI - A$
- n-th diagonal entry of $zI - A$ is $z - a_{nn}$
- Since determinant of upper triangular matrix is product of diagonal entries, we have

$$\Phi_A(z) = |zI - A| = (z - a_{11}) \cdots (z - a_{NN}) \qquad \square$$

Change of basis

- We usually take standard basis $\{e_1, \dots, e_N\}$ in $\mathbb{R}^N$ or $\mathbb{C}^N$, but that is not necessary
- Suppose we take different basis $\{p_1, \dots, p_N\}$
- By definition, $\{p_n\}$ is linearly independent, so $P = [p_1, \dots, p_N]$ is invertible
- Let x be any vector and $y = P^{-1}x$; then

  $$x = PP^{-1}x = Py = y_1 p_1 + \dots + y_N p_N,$$

  so entries of y can be interpreted as coordinates of x when expressed with basis P
- Then $x \mapsto Ax$ becomes

  $$y = P^{-1}x \mapsto P^{-1}Ax = (P^{-1}AP)(P^{-1}x) = (P^{-1}AP)y,$$

  so linear map $x \mapsto Ax$ has matrix representation $B = P^{-1}AP$ under basis P

Similarity

- When there exists invertible matrix P such that $B = P^{-1}AP$, we say A, B are *similar*
- When A, B are similar, they can be regarded as identical because they can be mapped to each other by change of basis
- For instance, characteristic polynomial of $B = P^{-1}AP$ is

$$\begin{aligned}\Phi_B(z) &= |zI - B| = \left|zI - P^{-1}AP\right| \\ &= \left|P^{-1}(zI - A)P\right| = |zI - A| = \Phi_A(z),\end{aligned}$$

so A, B have identical eigenvalues

Diagonalization

- For analysis, often useful to find matrix P such that $P^{-1}AP$ is simple
- For instance, let $B = P^{-1}AP$ and suppose computing B^k is easy (e.g., diagonal)
- Then $B^k = (P^{-1}AP)^k = P^{-1}A^kP$, so we may compute $A^k = PB^kP^{-1}$
- Simplest matrices of all is diagonal

Proposition

If the eigenvalues of the square matrix A are distinct, A is diagonalizable.

Proof

- Let $\{\alpha_n\}_{n=1}^N$ be eigenvalues and $Ap_n = \alpha_n p_n$; we show $\{p_n\}$ is linearly independent
- Suppose $\sum_{n=1}^N x_n p_n = 0$, and multiply $B_m := \prod_{n \neq m}(A - \alpha_n I)$
- Then

$$0 = B_m 0 = B_m \sum_{n=1}^N x_n p_n = x_m \prod_{n \neq m} (\alpha_m - \alpha_n) p_m,$$

  so $x_m = 0$ for all m
- Let $P = [p_1, \dots, p_N]$, which is invertible
- Stacking $Ap_n = \alpha_n p_n$ as column vectors, we obtain

$$AP = A[p_1, \dots, p_N] = [\alpha_1 p_n, \dots, \alpha_N p_N] = P \operatorname{diag}[\alpha_1, \dots, \alpha_N],$$

  so $P^{-1}AP = \operatorname{diag}[\alpha_1, \dots, \alpha_N]$ □

Inner product and norm

- When eigenvalues not distinct, matrix may not be diagonalizable; need additional structure
- For real vector space V, we say $\langle \cdot, \cdot \rangle : \mathrm{V} \times \mathrm{V} \to \mathbb{R}$ is *inner product* if
  1. (Nonnegativity) $\langle x, x \rangle \geq 0$ for all $x \in \mathrm{V}$, with equality if and only if $x = 0$,
  2. (Symmetry) $\langle x, y \rangle = \langle y, x \rangle$ for all $x, y \in \mathrm{V}$,
  3. (Linearity) $\langle x, y \rangle$ is linear in y
- Real vector space equipped with inner product $\langle \cdot, \cdot \rangle$ is called inner product space
- Obvious example is $\mathbb{R}^N$, but there are many more
- Can show Cauchy-Schwarz $\|x\| \|y\| \geq |\langle x, y \rangle|$ and triangle inequality $\|x + y\| \leq \|x\| + \|y\|$, so inner product space is automatically normed space

Example

- Let $a < b$ and $w : [a, b] \to (0, \infty)$ be positive continuous function
- Let V be space of continuous functions defined on $[a, b]$
- For $f, g \in \mathrm{V}$, define

$$\langle f, g \rangle = \int_a^b f(x)g(x)w(x)\,\mathrm{d}x$$

- Then V is inner product space

Complex inner product space

- When V is complex vector space, symmetry is replaced by
  2' (Conjugate symmetry) $\langle x, y\rangle = \overline{\langle y, x\rangle}$
- Here $\bar{\alpha}$ denotes complex conjugate of the scalar $\alpha \in \mathbb{C}$
- For example, if $V = \mathbb{C}^N$ and $x, y \in \mathbb{C}^N$, inner product is defined by
$$\langle x, y\rangle = x^* y = \bar{x}' y = \sum_{n=1}^{N} \bar{x}_n y_n$$
- Here $x^* = \bar{x}'$ is transpose of complex conjugate of x, or *conjugate transpose* for short

Gram-Schmidt orthonormalization

- Vectors $x, y \in \mathrm{V}$ satisfying $\langle x, y\rangle = 0$ are called *orthogonal*
- If any two vectors of $\{v_k\}_{k=1}^K$ are orthogonal and $\|v_k\| = 1$ for all k, we say that $\{v_k\}_{k=1}^K$ is *orthonormal*
- From any linearly independent $\{v_k\}_{k=1}^K$, we may construct orthonormal vectors $\{u_k\}_{k=1}^K$ as follows
  1. Define $u_1 = v_1/\|v_1\|$, so $\|u_1\| = 1$
  2. Proceed by induction; suppose $u_1, \dots, u_k$ have already been defined and $\operatorname{span}[u_1, \dots, u_k] = \operatorname{span}[v_1, \dots, v_k]$
  3. Define $v = v_{k+1} - \sum_{l=1}^k \langle v_{k+1}, u_l\rangle u_l$ and $u_{k+1} = v/\|v\|$
  4. Then clearly $\|u_{k+1}\| = 1$ and $\langle u_{k+1}, u_l\rangle = 0$ for all $l = 1, \dots, k$
  5. Continuing this process, we obtain desired orthonormal vectors $\{u_k\}$

Conjugate transpose, unitary matrix

- For complex A, let $A^* = \bar{A}'$ be *conjugate transpose*
- Using property of inner product,

$$\langle A^* x, y\rangle = (A^* x)^* y = x^*(A^*)^* y = x^* A y = \langle x, Ay\rangle$$

- Let $\{u_1, \dots, u_N\}$ be orthonormal basis and $U = [u_1, \dots, u_N]$
- Then $\langle u_m, u_n\rangle = \delta_{mn}$ (Kronecker delta), so $U^* U = I$
- Hence $U^* = U^{-1}$ and $U^* U = UU^* = I$; such matrix called *unitary matrix*
- If $P = U$ is real, then called *orthogonal matrix*

Schur triangularization theorem

Theorem (Schur triangularization theorem)

For any $A \in \mathcal{M}_N(\mathbb{C})$, there exists a unitary matrix U such that $U^{-1}AU = U^*AU$ is upper triangular.

Proof.

- Trivial if $N = 1$ by taking $U = (1)$
- General case is by induction; suppose true up to $N - 1$
- Let u_1 be eigenvector of A, so $Au_1 = \alpha_1 u_1$, with $\|u_1\| = 1$
- Use Gram-Schmidt to construct unitary $U_0 = [u_1, \dots, u_N]$
- Then
$$U_0^* A U_0 = \begin{bmatrix} \alpha_1 & b_1^* \\ 0 & A_1 \end{bmatrix},$$
and apply induction hypothesis to A_1 $\square$

Spectral theorem

- Schur triangularization theorem has many applications
- One of them is to diagonalize *self-adjoint matrices*, which satisfy $A^* = A$
- If $A = (a_{mn})$ real, it is self-adjoint if it is symmetric: $A' = A$ and $a_{mn} = a_{nm}$

Corollary (Spectral theorem)

A self-adjoint matrix is diagonalizable by a unitary matrix.

Proof.

- By Schur, can take unitary U such that U^*AU is upper triangular
- Then $(U^*AU)^* = U^*A^*U = U^*AU$ lower triangular, so diagonal □

Diagonalization of real symmetric matrices

Proposition

If $A \in \mathcal{M}_N(\mathbb{C})$ is self-adjoint, then

1. *for any $x \in \mathbb{C}^N$, the quadratic form $\langle x, Ax \rangle$ is real,*
2. *all eigenvalues of A are real.*

Proof.

- Note that $\overline{\langle x, Ax \rangle} = \langle Ax, x \rangle = \langle A^* x, x \rangle = \langle x, Ax \rangle$
- If $\alpha \in \mathbb{C}$ an eigenvalue of A, so $Av = \alpha v$ for some $v \neq 0$, then

$$\mathbb{R} \ni \langle v, Av \rangle = \langle v, \alpha v \rangle = \alpha \langle v, v \rangle = \alpha \|v\|^2$$

- Therefore $\alpha = \langle v, Av \rangle / \|v\|^2$ is also real $\square$

Corollary

A real symmetric matrix is diagonalizable by an orthogonal matrix.

Quadratic form

- For $A \in \mathcal{M}_N(\mathbb{R})$ and $x \in \mathbb{R}^N$, inner product $\langle x, Ax \rangle$ is called *quadratic form*
- Since $\langle x, Ax \rangle$ is scalar, we have $\langle x, Ax \rangle = \langle Ax, x \rangle = \langle x, A'x \rangle$
- Hence

$$\langle x, Ax \rangle = \frac{1}{2}(\langle x, Ax \rangle + \langle Ax, x \rangle) = \left\langle x, \left(\frac{A + A'}{2}\right) x \right\rangle,$$

  so without loss of generality we may assume A is symmetric
- We say A is *positive semidefinite* (psd) if $\langle x, Ax \rangle \geq 0$ for all x
- We say A is *positive definite* (pd) if $\langle x, Ax \rangle > 0$ for all $x \neq 0$
- Negative definite/semidefinite defined analogously

Characterization of positive (semi)definite matrices

Proposition

A real symmetric matrix is positive semidefinite (definite) if and only if all eigenvalues are nonnegative (positive).

Proof.

- We can take orthogonal matrix P such that $P'AP = \operatorname{diag}[\alpha_1, \dots, \alpha_N]$
- For any x, let $y = P'x$; since $PP' = I$, we have

$$\begin{aligned}\langle x, Ax \rangle &= x'Ax = x'PP'APP'x \\ &= y' \operatorname{diag}[\alpha_1, \dots, \alpha_N] y = \sum_{n=1}^{N} \alpha_n y_n^2\end{aligned}$$

- Last expression is nonnegative (positive) for all x (hence for all y) if and only if all α_n's are nonnegative (positive) $\square$

Characterization using principal minors

- For square A, determinant of matrix obtained by keeping first k rows and columns of A is called k-th *principal minor*
- For example, if $A = (a_{mn})$ is $N \times N$, first principal minor is a_{11}, second principal minor is $a_{11}a_{22} - a_{12}a_{21}$, and N-th principal minor is $|A|$, etc.

Proposition

A real symmetric matrix is positive definite if and only if its principal minors are all positive.

Proof.

By induction on N $\square$

Second-order optimality condition

- When we discussed multi-variable optimization, we only considered first-order condition
- This is because we need matrices for second-order condition
- Let $U \subset \mathbb{R}^N$ be open and $f : U \to \mathbb{R}$ be C^2
- Fix some $a \in U$, let $x \in U$ be sufficiently close to a, and define $g : [0,1] \to \mathbb{R}$ by $g(t) = f(a + t(x-a))$
- Then $g(0) = f(a)$ and $g(1) = f(x)$, so applying chain rule and Taylor, get

  $$f(x) = f(a) + \langle \nabla f(a), x - a \rangle + \frac{1}{2} \left\langle x - a, \nabla^2 f(\xi)(x-a) \right\rangle,$$

  where $\xi = (1-\theta)a + \theta x$ for some $0 < \theta < 1$
- $\nabla^2 f = \left(\frac{\partial^2 f}{\partial x_m \partial x_n} \right)$ is matrix of second partial derivatives of f, known as *Hessian*

Second-order optimality condition

Proposition

Let $U \subset \mathbb{R}^N$ be open and $f : U \to \mathbb{R}$ be C^2. Then:

1. *If $\bar{x} \in U$ is a local minimum, then $\nabla f(\bar{x}) = 0$ and $\nabla^2 f(\bar{x})$ is positive semidefinite.*
2. *If $\nabla f(\bar{x}) = 0$ and $\nabla^2 f(\bar{x})$ is positive definite, then $\bar{x}$ is a strict local minimum.*

Proof of necessity

- Let $\bar{x}$ be a local minimum; by FOC, we have $\nabla f(\bar{x}) = 0$
- Take any $v \in \mathbb{R}^N$; then for small enough $t > 0$, letting $a = \bar{x}$ and $x = a + tv$ in Taylor, we obtain

$$f(\bar{x}) \le f(x) = f(\bar{x}) + t \langle \nabla f(\bar{x}), v \rangle + \frac{1}{2} t^2 \left\langle v, \nabla^2 f(\bar{x} + \theta tv) v \right\rangle$$
$$\implies 0 \le \left\langle v, \nabla^2 f(\bar{x} + \theta tv) v \right\rangle$$

- Letting $t \to 0$ and noting that f is C^2, we obtain $\left\langle v, \nabla^2 f(\bar{x}) v \right\rangle \ge 0$, so $\nabla^2 f(\bar{x})$ is psd $\square$

Proof of sufficiency

- Suppose $\nabla f(\bar{x}) = 0$ and $\nabla^2 f(\bar{x})$ is pd
- Since determinant of matrix is continuous in its entries, signs of principal minors of $\nabla^2 f(x)$ remain same if x is sufficiently close to $\bar{x}$
- Hence $\nabla^2 f(x)$ is pd in neighborhood of $\bar{x}$
- Let $\|v\| = 1$ and $x = \bar{x} + tv$ for sufficiently small $t > 0$; by Taylor,

$$\begin{aligned} f(x) &= f(\bar{x}) + t \langle \nabla f(\bar{x}), v \rangle + \frac{1}{2} t^2 \left\langle v, \nabla^2 f(\bar{x} + \theta t v) v \right\rangle \\ &= f(\bar{x}) + \frac{1}{2} t^2 \left\langle v, \nabla^2 f(\bar{x} + \theta t v) v \right\rangle > f(\bar{x}), \end{aligned}$$

so $\bar{x}$ is local minimum $\square$

Matrix norm

- Since $\mathcal{M}_N(\mathbb{R})$ (set of $N \times N$ matrices) can be viewed as $\mathbb{R}^{N^2}$, we may use norms to measure sizes of matrices
- But distinctive property of matrices is that they can be multiplied
- We define *matrix norm* as follows
  1. (Nonnegativity) $\|A\| \geq 0$, with equality if and only if $A = 0$,
  2. (Positive homogeneity) $\|\alpha A\| = |\alpha| \, \|A\|$,
  3. (Triangle inequality) $\|A + B\| \leq \|A\| + \|B\|$,
  4. (Submultiplicativity) $\|AB\| \leq \|A\| \, \|B\|$
- When submultiplicativity is dropped, we call $\|\cdot\|$ *vector norm*
- For any norm $\|\cdot\|$ on $\mathbb{R}^N$, we can define *operator norm* on $\mathcal{M}_N(\mathbb{R})$ as follows

Operator norm

Proposition

For any norm $\|\cdot\|$ on $\mathbb{R}^N$, $\|A\| := \sup_{x \neq 0} \|Ax\| / \|x\|$ is matrix norm on $\mathcal{M}_N(\mathbb{R})$.

Proof.

- Nonnegativity and positive homogeneity easy
- Triangle inequality: Note that $\|Ax\| \le \|A\| \|x\|$ for all x, so
$$\|(A+B)x\| = \|Ax + Bx\| \le \|Ax\| + \|Bx\| \le (\|A\| + \|B\|) \|x\|$$
- Dividing both sides by $\|x\|$ and taking supremum, we obtain $\|A+B\| \le \|A\| + \|B\|$
- Submultiplicativity: For all x, we have
$$\|ABx\| = \|A(Bx)\| \le \|A\| \|Bx\| \le \|A\| \|B\| \|x\|$$
- Dividing both sides by $\|x\|$ and taking supremum, we obtain $\|AB\| \le \|A\| \|B\|$

Example: ℓ^∞ norm

- Let $\|\cdot\|$ denote ℓ^∞ norm and $A = (a_{mn})$
- Then
$$\|Ax\| = \max_m \left| \sum_{n=1}^N a_{mn} x_n \right|$$
- Taking maximum over all x with $\|x\| = \max_n |x_n| = 1$, we get
$$\|A\| = \max_m \sum_{n=1}^N |a_{mn}|$$

Spectral radius

- Let $A \in \mathcal{M}_N(\mathbb{C})$
- Set of eigenvalues $\{\alpha_n\}_{n=1}^N$ is called *spectrum* of A
- Largest absolute value of all eigenvalues,

$$\rho(A) := \max_n |\alpha_n|,$$

  is called *spectral radius*
- As we shall see below, spectral radius is important measure of matrix

Convergence of matrix power

Proposition

Let $A \in \mathcal{M}_N(\mathbb{C})$*. Then* $\lim_{k\to\infty} A^k = 0$ *if and only if* $\rho(A) < 1$*.*

Proof of necessity.

- By Schur, we may assume that A is upper triangular; then diagonal entries of A are eigenvalues
- If $A^k \to 0$, then $\alpha^k \to 0$ for all eigenvalues, so $|\alpha| < 1$ for all α and $\rho(A) < 1$ $\square$

Proof of sufficiency

- Conversely, suppose A is upper triangular and $r := \rho(A) < 1$
- Write $A = D + T$, where D is diagonal and T is upper triangular with zero diagonal entries
- Then $|A| = |D| + |T| \le rI + |T|$ entrywise
- Since T upper triangular with zero diagonal entries, we can easily check $|T|^N = 0$
- Hence by binomial theorem, for $k \ge N$ we have

$$\begin{aligned} 0 &\le \left|A^k\right| \le |A|^k \le (rI + |T|)^k \\ &= \sum_{l=0}^{k} \binom{k}{l} r^{k-l} |T|^l = \sum_{l=0}^{N-1} \binom{k}{l} r^{k-l} |T|^l, \end{aligned}$$

which tends to 0 as $k \to \infty$ because $0 \le r < 1$ and $\binom{k}{l}$ is polynomial of k with degree at most $N - 1$ $\square$

Gelfand spectral radius formula

Theorem (Gelfand spectral radius formula)

Let $\|\cdot\|$ be any matrix norm on $\mathcal{M}_N(\mathbb{C})$. Then $\rho(A) \le \left\|A^k\right\|^{1/k}$ and $\rho(A) = \lim_{k\to\infty} \left\|A^k\right\|^{1/k}$.

Proof of first statement.

- If $Av = \alpha v$, then $A^k v = \alpha^k v$ for all k
- For $V = (v, \dots, v)$, we have $A^k V = \alpha^k V$, so

$$|\alpha|^k \|V\| = \left\|A^k V\right\| \le \left\|A^k\right\| \|V\| \implies |\alpha|^k \le \left\|A^k\right\|$$

- Since α is any eigenvalue, $\rho(A) \le \left\|A^k\right\|^{1/k}$ $\square$

Proof of second statement

- Take any $\epsilon > 0$ and let $\tilde{A} = \frac{1}{\rho(A)+\epsilon} A$
- Then $\rho(\tilde{A}) = \frac{\rho(A)}{\rho(A)+\epsilon} < 1$, so $\lim_{k\to\infty} \tilde{A}^k = 0$
- Therefore $\left\| \tilde{A}^k \right\| < 1$ for large enough k, and hence $\left\| A^k \right\| \le (\rho(A) + \epsilon)^k$
- Taking k-th root, letting $k \to \infty$, and $\epsilon \downarrow 0$, we obtain $\limsup_{k\to\infty} \left\| A^k \right\|^{1/k} \le \rho(A)$
- Since $\rho(A) \le \left\| A^k \right\|^{1/k}$, it follows that $\rho(A) = \lim_{k\to\infty} \left\| A^k \right\|^{1/k}$ $\square$

Matrix series

- By Gelfand, "size" of matrix power A^k is approximately $\rho(A)^k$
- Suppose power series $f(z) = \sum_{k=0}^{\infty} a_k z^k$ converges for $|z| < r$
- Then matrix series

$$f(A) = \sum_{k=0}^{\infty} a_k A^k$$

  well defined if $\rho(A) < r$
- Example:

$$\exp(A) := \sum_{k=0}^{\infty} \frac{1}{k!} A^k$$

Important points

- Eigenvalue and eigenvector: $Av = \alpha v$
- Any matrix can be upper triangularized by unitary matrix (Schur)
- Real symmetric matrix can be diagonalized by orthogonal matrix
- Real symmetric matrix is positive definite if and only if all eigenvalues positive, related to second-order optimality condition
- Gelfand spectral radius formula $\lim \left\|A^k\right\|^{1/k} = \rho(A)$, so matrix power A^k behaves like $\rho(A)^k$

Chapter 7

Metric Space and Contraction

Metric space

Completeness and Banach space

Contraction mapping theorem

Blackwell's sufficient condition

Perov contraction

Implicit function theorem

Metric space

- Recall that normed space is vector space V equipped with norm $\|\cdot\|$
- For any two elements v_1, v_2 of V, we may define distance by

$$d(v_1, v_2) := \|v_1 - v_2\|$$

- Using properties of norm, we can easily show that d satisfies:
  1. (Nonnegativity) $d(v_1, v_2) \geq 0$, with equality if and only if $v_1 = v_2$,
  2. (Symmetry) $d(v_1, v_2) = d(v_2, v_1)$,
  3. (Triangle inequality) $d(v_1, v_3) \leq d(v_1, v_2) + d(v_2, v_3)$
- In general, if V is equipped with $d : \mathrm{V} \times \mathrm{V} \to \mathbb{R}$ satisfying above properties, we say (V, d) is *metric space*

Space of bounded functions

- Let $X \subset \mathbb{R}^N$ be nonempty and V be space of bounded functions on X:
$$\mathrm{V} = \{v : X \to \mathbb{R} : v \text{ is bounded}\}$$
- For $v \in \mathrm{V}$, define
$$\|v\| = \sup_{x \in X} |v(x)|$$
- Then $(\mathrm{V}, \|\cdot\|)$ is normed space
- $\|\cdot\|$ is called *supremum norm* or *sup norm*

Proof

- Since $v \in \mathrm{V}$ is bounded, clearly $0 \le \|v\| < \infty$; if $\|v\| = 0$, then $|v(x)| = 0$ for all $x \in X$, so $v = 0$
- If $\alpha \in \mathbb{R}$ and $v \in \mathrm{V}$, then

$$\|\alpha v\| = \sup_{x \in X} |\alpha v(x)| = |\alpha| \sup_{x \in X} |v(x)| = |\alpha| \, \|v\|$$

- Noting that $|v(x)| \le \|v\|$ for all $x \in X$, for $v_1, v_2 \in \mathrm{V}$, we have

$$|v_1(x) + v_2(x)| \le |v_1(x)| + |v_2(x)| \le \|v_1\| + \|v_2\|$$

- Taking supremum of left-hand side over $x \in X$, we obtain

$$\|v_1 + v_2\| \le \|v_1\| + \|v_2\| \; \square$$

Examples

- Let V be space of bounded functions on X
- For any subset $\mathrm{V}_1 \subset \mathrm{V}$ and $v_1, v_2 \in \mathrm{V}_1$, define *sup distance*

$$d(v_1, v_2) = \|v_1 - v_2\|$$

- Then (V_1, d) is metric space
- Examples:
  - Set of bounded increasing functions
  - Set of bounded convex functions

Topology on metric spaces

- If (V, d) is metric space, define (open) *ball* with center $v \in \mathrm{V}$ and radius $\epsilon > 0$ by

$$B_\epsilon(v) := \{w \in \mathrm{V} : d(v, w) < \epsilon\}$$

- Then we may define convergence of sequences in V and topology (open sets) of V exactly as $\mathbb{R}^N$
- For instance, $U \subset \mathrm{V}$ is *open* if and only if for any $v \in U$, we can take $\epsilon > 0$ such that $B_\epsilon(v) \subset U$
- Similarly, a sequence $\{v_k\}_{k=1}^\infty \subset \mathrm{V}$ *converges* to $v \in \mathrm{V}$ if and only if $d(v_k, v) \to 0$ as $k \to \infty$
- All previous results for $\mathbb{R}^N$ generalize to metric spaces, with identical proofs

Complete metric space and Banach space

- Let (V, d) be metric space
- We say that sequence $\{v_k\}_{k=1}^\infty \subset \mathrm{V}$ is *Cauchy* if

$$(\forall \epsilon > 0)(\exists K > 0)(\forall k, l \geq K) \quad d(v_k, v_l) < \epsilon$$

- Can show Cauchy sequences in $\mathbb{R}^N$ are convergent, called *completeness* of $\mathbb{R}^N$
- When metric space (V, d) is complete (Cauchy sequences are convergent), we call it *complete metric space*
- Normed space $(V, \cdot)$ can be viewed as metric space with sup distance $d(v_1, v_2) = \|v_1 - v_2\|$; complete normed spaces are called *Banach spaces*

Space of bounded functions is Banach

Proposition

The normed space $(\mathrm{V}, \|\cdot\|)$ *of bounded functions is complete and hence Banach.*

Proof.

- Let $\{v_k\}_{k=1}^{\infty} \subset \mathrm{V}$ be Cauchy; since $|v(x)| \le \|v\|$ for all x,

$$(\forall \epsilon > 0)(\exists K > 0)(\forall k, l \ge K)(\forall x \in X) \quad |v_k(x) - v_l(x)| < \epsilon$$

- Therefore for fixed $x \in X$, $\{v_k(x)\}$ is Cauchy in $\mathbb{R}$ and convergent; let $v(x)$ be its limit
- Letting $l \to \infty$, we obtain

$$(\forall \epsilon > 0)(\exists K > 0)(\forall k \ge K)(\forall x \in X) \quad |v_k(x) - v(x)| \le \epsilon$$

- Taking supremum over $x \in X$, we obtain

$$(\forall \epsilon > 0)(\exists K > 0)(\forall k \ge K) \quad \|v_k - v\| \le \epsilon$$

$\square$

Closed subsets of complete metric space

- Let (V, d) be complete metric space
- Let $V_1 \subset V$ be closed
- Then clearly (V_1, d) is complete metric space
- Examples:
  - Set of bounded increasing functions
  - Set of bounded convex functions
- Note: above examples are complete metric spaces but not Banach (because increasing or convex functions do not form vector space)

Space of bounded continuous functions is Banach

Corollary

The space of bounded continuous functions is Banach. Any closed subset of it is a complete metric space.

Proof.

- Let $X \subset \mathbb{R}^N$ and bX be Banach space of bounded functions on X with sup norm $\|\cdot\|$
- Let bcX be space of bounded continuous functions on X, which is normed space; let $\{v_k\}_{k=1}^{\infty}$ be Cauchy in bcX
- Then it is Cauchy in bX and converges to some v; thus suffices to show v is continuous
- For any $\epsilon > 0$, since $v_k \to v$ in bX, we can take K such that $\|v - v_k\| < \epsilon/3$ for $k > K$

Proof

- Fix such k and take any $x \in X$; by continuity, we can take neighborhood U of x such that $|v_k(y) - v_k(x)| < \epsilon/3$ for $y \in U$
- Then

$$\begin{aligned}
&|v(y) - v(x)| \\
&= |v(y) - v_k(y) + v_k(y) - v_k(x) + v_k(x) - v(x)| \\
&\le |v(y) - v_k(y)| + |v_k(y) - v_k(x)| + |v_k(x) - v(x)| \\
&\le \|v - v_k\| + \frac{\epsilon}{3} + \|v - v_k\| < \epsilon,
\end{aligned}$$

so v is continuous $\square$

Contraction

- In general, if V is set and T is function from V to itself ($T : \mathrm{V} \to \mathrm{V}$), we say that T is *self map* or *operator*
- If T is self map on V and $v \in \mathrm{V}$ satisfies $T(v) = v$, we say v is *fixed point* of T
- For metric space (V, d), we say $T : \mathrm{V} \to \mathrm{V}$ is *contraction* with modulus β if $\beta \in [0, 1)$ and

  $$d(T(v_1), T(v_2)) \le \beta d(v_1, v_2)$$

  for all $v_1, v_2 \in \mathrm{V}$
- Intuitively, when we apply T, distance between two points shrinks by factor $\beta < 1$

Contraction mapping theorem

Theorem (Contraction mapping theorem)

Let (V, d) *be a complete metric space and* $T : \mathrm{V} \to \mathrm{V}$ *be a contraction with modulus* $\beta \in [0, 1)$*. Then*

1. T *has a unique fixed point* $v^* \in \mathrm{V}$*,*
2. *for any* $v_0 \in \mathrm{V}$*, we have* $v^* = \lim_{k\to\infty} T^k(v_0)$*, and*
3. *the approximation error* $d(T^k(v_0), v^*)$ *has order of magnitude* β^k*.*

- Contraction mapping theorem is also called Banach fixed point theorem

Proof

- By definition, contraction is (uniformly) continuous
- Take any $v_0 \in V$ and define $v_k = T(v_{k-1})$ for $k \geq 1$
- Since T is contraction, we have

$$\begin{aligned} d(v_k, v_{k-1}) &= d(T(v_{k-1}), T(v_{k-2})) \leq \beta d(v_{k-1}, v_{k-2}) \\ &\leq \cdots \leq \beta^{k-1} d(v_1, v_0) \end{aligned}$$

- If $k > l \geq K$, by triangle inequality we have

$$\begin{aligned} d(v_k, v_l) &\leq d(v_k, v_{k-1}) + \cdots + d(v_{l+1}, v_l) \\ &\leq (\beta^{k-1} + \cdots + \beta^l) d(v_1, v_0) \\ &= \frac{\beta^l - \beta^k}{1-\beta} d(v_1, v_0) \leq \frac{\beta^l}{1-\beta} d(v_1, v_0) \leq \frac{\beta^K}{1-\beta} d(v_1, v_0) \end{aligned}$$

- Since $0 \leq \beta < 1$, we have $\beta^K \to 0$ as $K \to \infty$, so $\{v_k\}$ is Cauchy and $v^* = \lim_{k \to \infty} v_k$ exists
- Since $d(T(v_k), v_k) = d(v_{k+1}, v_k) \leq \beta^k d(v_1, v_0)$, letting $k \to \infty$ and using the continuity of T, we get $d(T(v^*), v^*) = 0$, so $T(v^*) = v^*$

Proof

- To show uniqueness, suppose v_1, v_2 are fixed points of T, so $T(v_1) = v_1$ and $T(v_2) = v_2$
- Since T is contraction, we have

$$\begin{aligned}0 &\le d(v_1, v_2) = d(T(v_1), T(v_2)) \le \beta d(v_1, v_2) \\ &\implies (\beta - 1) d(v_1, v_2) \ge 0\end{aligned}$$

- Since $\beta < 1$, it must be $d(v_1, v_2) = 0$ and hence $v_1 = v_2$
- Finally, take any v_0 and let $v_k = T^k(v_0)$; then

$$\begin{aligned}d(v_k, v^*) &= d(T(v_{k-1}), T(v^*)) \le \beta d(v_{k-1}, v^*) \\ &\le \cdots \le \beta^k d(v_0, v^*)\end{aligned}$$

- Letting $k \to \infty$ we have $v_k \to v^*$, and error has order of magnitude β^k $\square$

Blackwell's sufficient condition

Proposition (Blackwell's sufficient condition)

Let X be a set and V be a space of functions on X with the following properties:

(a) *(Upward shift) For $v \in \mathrm{V}$ and $c \in \mathbb{R}_+$, we have $v + c \in \mathrm{V}$.*

(b) *(Bounded difference) For all $v_1, v_2 \in \mathrm{V}$, we have*

$$d(v_1, v_2) := \sup_{x \in X} |v_1(x) - v_2(x)| < \infty.$$

Suppose that (V, d) is a complete metric space and $T : \mathrm{V} \to \mathrm{V}$ satisfies

1. *(Monotonicity) $v_1 \le v_2$ implies $Tv_1 \le Tv_2$,*
2. *(Discounting) there exists $\beta \in [0, 1)$ such that, for all $v \in \mathrm{V}$ and $c \in \mathbb{R}_+$, we have $T(v + c) \le Tv + \beta c$.*

Then T is a contraction with modulus β.

Proof

- Take any $v_1, v_2 \in \mathrm{V}$ and let $c = d(v_1, v_2) \geq 0$
- For any $x \in X$, we have

$$v_1(x) = v_1(x) - v_2(x) + v_2(x) \leq v_2(x) + c,$$

  so $v_1 \leq v_2 + c \in \mathrm{V}$ by upward shift property
- Using monotonicity and discounting, we obtain

$$Tv_1 \leq T(v_2 + c) \leq Tv_2 + \beta c \implies Tv_1 - Tv_2 \leq \beta c$$

- Interchanging role of v_1, v_2, we obtain $Tv_2 - Tv_1 \leq \beta c$
- Thus $|(Tv_1)(x) - (Tv_2)(x)| \leq \beta d(v_1, v_2)$ for any $x \in X$
- Taking supremum over x, we obtain $d(Tv_1, Tv_2) \leq \beta d(v_1, v_2)$, so T is contraction with modulus β $\square$

Vector-valued metric space

- Let V be set, $I \in \mathbb{N}$, and $d : \mathrm{V} \times \mathrm{V} \to \mathbb{R}^I$
- We say d is *vector-valued metric* if:
    1. (Nonnegativity) $d(v_1, v_2) \geq 0$, with equality if and only if $v_1 = v_2$,
    2. (Symmetry) $d(v_1, v_2) = d(v_2, v_1)$,
    3. (Triangle inequality) $d(v_1, v_3) \leq d(v_1, v_2) + d(v_2, v_3)$
- In conditions 1 and 3, note that for $a = (a_1, \dots, a_I) \in \mathbb{R}^I$ and $b = (b_1, \dots, b_I) \in \mathbb{R}^I$, we write $a \leq b$ if and only if $a_i \leq b_i$ for all i
- Set V endowed with vector-valued metric d is called *vector-valued metric space*
- Obviously, $I = 1$ corresponds to metric space

Complete vector-valued metric space

- Let $\|\cdot\|$ denote supremum norm on $\mathbb{R}^I$, so $\|a\| = \max_i |a_i|$ for $a = (a_1, \dots, a_I) \in \mathbb{R}^I$
- Note that supremum norm is monotone: if $a, b \in \mathbb{R}^I$ and $0 \le a \le b$, then $\|a\| = \max_i a_i \le \max_i b_i = \|b\|$
- If (V, d) is vector-valued metric space and we define $\|d\| : \mathrm{V} \times \mathrm{V} \to \mathbb{R}$ by $\|d\|(v_1, v_2) = \|d(v_1, v_2)\|$, then $(\mathrm{V}, \|d\|)$ is metric space in usual sense
- To see why, nonnegativity and symmetry are obvious, and

$$\begin{aligned}
&\|d\|(v_1, v_3) \\
&= \|d(v_1, v_3)\| \le \|d(v_1, v_2) + d(v_2, v_3)\| \\
&\le \|d(v_1, v_2)\| + \|d(v_2, v_3)\| = \|d\|(v_1, v_2) + \|d\|(v_2, v_3)
\end{aligned}$$

- We say (V, d) is complete if $(\mathrm{V}, \|d\|)$ is complete

Perov contraction

- Let $\|\cdot\|$ be supremum norm as well as operator norm for $I \times I$ matrices
- Recall that for square matrix A, spectral radius $\rho(A)$ is largest absolute value of all eigenvalues
- Let (V, d) be vector-valued metric space; we say $T : \mathrm{V} \to \mathrm{V}$ is *Perov contraction* with coefficient matrix $B \geq 0$ if $\rho(B) < 1$ and
$$d(Tv_1, Tv_2) \leq Bd(v_1, v_2)$$
for all $v_1, v_2 \in \mathrm{V}$
- Here $B \geq 0$ means $B = (b_{ij})$ is nonnegative: $b_{ij} \geq 0$ for all i, j

Perov contraction theorem

Theorem (Perov contraction theorem)

Let (V, d) *be a complete vector-valued metric space and* $T : \mathrm{V} \to \mathrm{V}$ *be a Perov contraction with coefficient matrix* $B \geq 0$*. Then*

1. T *has a unique fixed point* $v^* \in \mathrm{V}$*,*
2. *for any* $v_0 \in \mathrm{V}$*, we have* $v^* = \lim_{k\to\infty} T^k v_0$*, and*
3. *for any* $\beta \in (\rho(B), 1)$*, the approximation error* $d(T^k v_0, v^*)$ *has order of magnitude* β^k*.*

Proof.

- Almost identical to proof of contraction mapping theorem
- Monotonicity of sup norm $\|\cdot\|$ and Gelfand spectral radius formula $\rho(B) = \lim_{k\to\infty} \left\|B^k\right\|^{1/k}$ play key role □

Sufficient condition for Perov contraction

Proposition

Let X be a set and V be a space of functions $v : X \to \mathbb{R}^I$ with the following properties:

(a) *(Upward shift) For $v \in \mathrm{V}$ and $c \in \mathbb{R}^I_+$, we have $v + c \in \mathrm{V}$.*

(b) *(Bounded difference) For all $u, v \in \mathrm{V}$ and i, we have*

$$d_i(u, v) := \sup_{x \in X} |u_i(x) - v_i(x)| < \infty.$$

Let $d = (d_1, \dots, d_I)$. Suppose that (V, d) is a complete vector-valued metric space and $T : \mathrm{V} \to \mathrm{V}$ satisfies

1. *(Monotonicity) $u \leq v$ implies $Tu \leq Tv$,*
2. *(Discounting) there exists a nonnegative matrix $B \in \mathcal{M}_I(\mathbb{R})$ with $\rho(B) < 1$ such that, for all $v \in \mathrm{V}$ and $c \in \mathbb{R}^I_+$, we have $T(v + c) \leq Tv + Bc$.*

Then T is a Perov contraction with coefficient matrix B.

Comparative statics

- When solving economic problems, we often encounter equations like $f(x, y) = 0$, where y is endogenous variable and x is exogenous variable
- Oftentimes y does not have explicit expression, but we might be interested in how y changes with x
- Such exercise is called *comparative statics*
- Implicit function theorem allows us to compute derivative $\mathrm{d}y/\mathrm{d}x$

Implicit function theorem

Theorem (Implicit function theorem)

Let $f : \mathbb{R}^M \times \mathbb{R}^N \to \mathbb{R}^N$ be C^1. If $f(x_0, y_0) = 0$ and $D_y f(x_0, y_0)$ is invertible, then there exist neighborhoods U of x_0 and V of y_0 and a function $g : U \to V$ such that

1. *for all $x \in U$, $f(x, y) = 0 \iff y = g(x)$,*
2. *g is C^1, and*
3. *$D_x g(x) = -[D_y f(x, y)]^{-1} D_x f(x, y)$, where $y = g(x)$.*

Proof.

- Proof is application of inverse function theorem
- Proof of inverse function theorem is hard and uses contraction mapping theorem □

Remembering implicit function theorem

- No need to remember precise statement of implicit function theorem, but important to know how to apply
- Simple way to remember assumption and statement of implicit function theorem: start from equation $f(x,y)=0$
- Set $y=g(x)$, differentiate $f(x,g(x))=0$ applying chain rule, and derive

$$D_x f + D_y f D_x g = 0 \iff D_x g = -[D_y f]^{-1} D_x f$$

- For this equation to be meaningful, we need $D_y f$ to be invertible, which is exactly assumption

Chapter 8

Nonnegative Matrices

Model of employment-unemployment

- Suppose worker can be either employed or unemployed
  - If employed, worker becomes unemployed with probability $p \in (0,1)$ next period
  - If unemployed, worker becomes employed with probability $q \in (0,1)$ next period
- Let $x_t = (e_t, u_t)$ be (row) probability vector of being employed and unemployed at time t, where $u_t = 1 - e_t$; then

$$\begin{aligned} e_{t+1} &= (1-p)e_t + qu_t, \\ u_{t+1} &= pe_t + (1-q)u_t \end{aligned}$$

- Collecting these equations into vector, we obtain $x_{t+1} = x_t P$, where

$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$$

Model of employment-unemployment

- Since $x_t = x_0 P^t$, suffices to know behavior of P^t as $t \to \infty$
- Characteristic polynomial of P is

$$\begin{aligned}\Phi_P(x) = |xI - P| &= \begin{vmatrix} x-1+p & -p \\ -q & x-1+q \end{vmatrix} \\ &= x^2 + (p+q-2)x + 1 - p - q \\ &= (x-1)(x+p+q-1)\end{aligned}$$

- Since eigenvalues are 1 and $1 - p - q \in (-1, 1)$, can diagonalize to compute P^t
- We omit details, but easy to show

$$P^t \to \frac{1}{p+q} \begin{bmatrix} q & p \\ q & p \end{bmatrix}$$

and hence $x_t \to \frac{1}{p+q}(q, p)$, so worker eventually unemployed with probability $\frac{p}{p+q}$

Markov process

- When random variable is indexed by time, we call *stochastic process*
- For stochastic process $\{X_t\}_{t=0}^{\infty}$, when distribution of X_t conditional on past information $X_{t-1}, X_{t-2}, \dots$ depends only on most recent past (X_{t-1}), we say $\{X_t\}$ is *Markov process*
- For example, vector autoregression (VAR)

$$X_t = AX_{t-1} + u_t$$

(where A is a matrix and the shock u_t is IID over time) is Markov process

Markov chain

- When Markov process $\{X_t\}$ takes on finitely many values, it is called *finite-state Markov chain*
- Let $\{X_t\}$ be (finite-state) Markov chain and $n = 1, \dots, N$ index values $\{x_n\}_{n=1}^N$ process can take
- We write $X_t = x_n$ when state at t is n
- Since there are finitely many states, distribution of X_t conditional on X_{t-1} is multinomial
- Hence Markov chain is completely characterized by *transition probability (stochastic) matrix* $P = (p_{nn'})$, where $p_{nn'}$ is probability of transitioning from state n to n'
- Clearly, we have $p_{nn'} \geq 0$ and $\sum_{n'=1}^N p_{nn'} = 1$

Unconditional distribution of Markov chain

- Let $\{X_t\}$ be Markov chain with transition probability matrix P
- If X_0 distributed according to distribution $\mu = (\mu_1, \dots, \mu_N)$, what is distribution of X_t?
- Using Markov property,

$$\Pr(X_1 = n') = \sum_{n=1}^{N} \Pr(X_0 = n) p_{nn'} = \sum_{n=1}^{N} \mu_n p_{nn'}$$

- Collecting into vector, distribution of X_1 is μP
- By induction, distribution of X_t is μP^t
- What is long run behavior as $t \to \infty$?

Invariant distribution of Markov chain

Theorem

Let $P = (p_{nn'})$ be a stochastic matrix such that $p_{nn'} > 0$ for all n, n'. Then there exists a unique invariant distribution π such that $\pi = \pi P$, and $\lim_{t\to\infty} \mu P^t = \pi$ for all initial distribution μ.

Proof.

- Let $\Delta = \left\{x \in \mathbb{R}^N_+ : \sum_{n=1}^N x_n = 1\right\}$ be set of all multinomial distributions
- Since $\Delta \subset \mathbb{R}^N$ is closed and $\mathbb{R}^N$ is complete metric space with ℓ^1 norm, Δ is complete metric space
- View $x \in \mathbb{R}^N$ as row vector and define $T : \Delta \to \mathbb{R}^N$ by $T(x) = xP$
- Let us show $T\Delta \subset \Delta$

Proof

- Note that if $x \in \Delta$, since $p_{nn'} \geq 0$ for all n, n', we have $xP \geq 0$
- Since $\sum_{n'=1}^{N} p_{nn'} = 1$, we have

$$\sum_{n'=1}^{N} (xP)_{n'} = \sum_{n'=1}^{N} \sum_{n=1}^{N} x_n p_{nn'} = \sum_{n=1}^{N} x_n \sum_{n'=1}^{N} p_{nn'} = \sum_{n=1}^{N} x_n = 1,$$

  so $T(x) = xP \in \Delta$
- Next we show T is contraction
- Since $P \gg 0$, we can take $\epsilon > 0$ such that $p_{nn'} - \epsilon > 0$ for all n, n'
- Let $q_{nn'} = \frac{p_{nn'} - \epsilon}{1 - N\epsilon} > 0$ and $Q = (q_{nn'})$
- Since $\sum_{n'} p_{nn'} = 1$, we obtain $\sum_{n'} q_{nn'} = 1$, so Q is also stochastic matrix; letting J be matrix with all entries equal to 1, we have $P = (1 - N\epsilon)Q + \epsilon J$

Proof

- For $\mu, \nu \in \Delta$, we have
$$\mu P - \nu P = (1 - N\epsilon)(\mu Q - \nu Q) + \epsilon(\mu J - \nu J)$$
- Since all entries of J are 1 and vectors μ, ν sum to 1, we have $\mu J = \nu J = 1 = (1, \dots, 1)$
- Therefore letting $0 < \beta = 1 - N\epsilon < 1$, we get
$$\begin{aligned}
\|T(\mu) - T(\nu)\| &= \|\mu P - \nu P\| = \beta \|\mu Q - \nu Q\| \\
&= \beta \sum_{n'=1}^{N} |(\mu Q)_{n'} - (\nu Q)_{n'}| = \beta \sum_{n'=1}^{N} \left| \sum_{n=1}^{N} (\mu_n - \nu_n) q_{nn'} \right| \\
&\le \beta \sum_{n'=1}^{N} \sum_{n=1}^{N} |\mu_n - \nu_n| \, q_{nn'} = \beta \sum_{n=1}^{N} |\mu_n - \nu_n| \sum_{n'=1}^{N} q_{nn'} \\
&= \beta \sum_{n=1}^{N} |\mu_n - \nu_n| = \beta \|\mu - \nu\|
\end{aligned}$$
and T is contraction

Nonnegative matrices

- Recall convention for vector inequalities: for real matrices $A = (a_{mn})$ and $B = (b_{mn})$ of the same size, we write $A \leq B$ ($A \ll B$) if $a_{mn} \leq b_{mn}$ ($a_{mn} < b_{mn}$) for all m, n
- Reverse inequalities $\geq, \gg$ are defined analogously
- If $A \geq 0$ ($A \gg 0$), we say A is *nonnegative* (*positive*)
- For example, stochastic matrices are nonnegative
- Nonnegative matrices often appear in economics, for instance input-output analysis

Spectral radius of nonnegative matrices

Proposition

For $A, B \in \mathcal{M}_N(\mathbb{C})$, if $0 \le |A| \le B$, then $\rho(A) \le \rho(|A|) \le \rho(B)$.

Proof.

- Let $\|\cdot\|$ denote supremum norm on $\mathbb{C}^N$ as well as operator norm induced by it
- Then by triangle inequality for complex numbers, we have $\|A^k\| \le \left\||A|^k\right\| \le \|B^k\|$
- Taking $1/k$-th power and letting $k \to \infty$, by Gelfand spectral radius formula, we obtain $\rho(A) \le \rho(|A|) \le \rho(B)$ $\square$

Perron's theorem

Theorem (Perron's theorem)

If $A \in \mathcal{M}_N(\mathbb{R})$ is positive, the following statements are true.

1. $\rho(A) > 0$, *which is an eigenvalue of A (called the Perron root.*
2. *There exist $x, y \gg 0$ (called the right and left Perron vectors) such that $Ax = \rho(A)x$ and $y'A = \rho(A)y'$.*
3. *The vectors x, y are unique up to scalar multiplication (in $\mathbb{C}^N$).*
4. *If x, y are chosen such that $y'x = 1$, then $\lim_{k\to\infty}[\frac{1}{\rho(A)}A]^k = xy'$.*

- Generalization for when $A = P$ is stochastic matrix

Proof

- Let $\alpha = \rho(A)$, λ be eigenvalue of A with $|\lambda| = \alpha$, and $u = (u_1, \dots, u_N)' \neq 0$ be corresponding eigenvector
- Let $v = (|u_1|, \dots, |u_N|)' > 0$ be vector of absolute values
- Since $Au = \lambda u$, taking absolute value of each entry and noting that A is positive, we obtain

$$\alpha |u_m| = \left| \sum_{n=1}^{N} a_{mn} u_n \right| \leq \sum_{n=1}^{N} a_{mn} |u_n| \iff \alpha v \leq Av$$

- To show $Av = \alpha v$, suppose to contrary $w := Av > \alpha v$
- Then $w - \alpha v > 0$, so multiplying A from left and noting that $A \gg 0$, we obtain

$$A(w - \alpha v) \gg 0 \iff Aw \gg \alpha Av = \alpha w$$

Proof

- Since A is finite-dimensional, we can take $\epsilon > 0$ such that $Aw \geq (1+\epsilon)\alpha w$
- Multiplying both sides from left by A^{k-1}, we obtain

$$A^k w \geq (1+\epsilon)\alpha A^{k-1} w \geq \cdots \geq [(1+\epsilon)\alpha]^k w$$

- Let $\|\cdot\|$ be sup norm as well as operator norm induced by it; then

$$\left\|A^k\right\| \|w\| \geq \left\|A^k w\right\| \geq [(1+\epsilon)\alpha]^k \|w\| \implies \left\|A^k\right\|^{1/k} \geq (1+\epsilon)\alpha$$

- Letting $k \to \infty$, by Gelfand, we obtain $\alpha \geq (1+\epsilon)\alpha$, which is contradiction
- Therefore $Av = \alpha v$, so A has positive eigenvector

Proof

- Let $x = v \gg 0$ be right Perron vector of A
- Then $\sum_{n=1}^N a_{mn} x_n = \alpha x_m$
- Define $D = \operatorname{diag}[x_1, \dots, x_N]$ and $P = \frac{1}{\alpha} D^{-1} A D \gg 0$
- Comparing (m, n) entry, we obtain $p_{mn} = \frac{a_{mn} x_n}{\alpha x_m}$, so

$$\sum_{n=1}^N p_{mn} = \sum_{n=1}^N \frac{a_{mn} x_n}{\alpha x_m} = 1$$

- Thus P is positive stochastic matrix, and rest of proof follows from previous case $\square$

Irreducible nonnegative matrices

- Perron's theorem generalizes to irreducible nonnegative matrices
- Irreducibility is best understood with stochastic matrices
- Let $\{X_t\}_{t=0}^{\infty}$ be finite-state Markov chain with state space $\{x_1, \dots, x_N\}$ and transition probability matrix $P = (p_{mn})$
- If we write $P^k = (p_{mn}^{(k)})$ for $k = 1, 2, \dots$, we obtain

$$\Pr(X_{t+k} = x_n \mid X_t = x_m) = p_{mn}^{(k)}$$

- We say Markov chain is *irreducible* if for each (m, n) pair, we have $p_{mn}^{(k)} > 0$ for some k
- In other words, irreducibility means that starting from any state m, we may transition to any other state n some time in future with positive probability

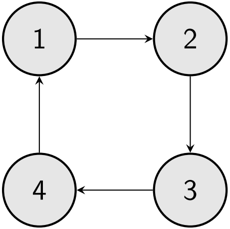
Directed graph and adjacency matrix

- More generally, irreducibility is related to directed graphs or networks
- Let $\{1, \dots, N\}$ be finite set, and for each (m, n) pair, suppose we can determine whether some property holds or not; example:
  - "person m likes person n",
  - "chapter m is required to understand chapter n",
  - "in Markov chain, it is possible to transition from state m to n in one step"
- For each (m, n) pair, define $a_{mn} = 1$ (0) if property holds (does not hold)
- Mathematically, directed graph is defined by *adjacency matrix* $A = (a_{mn})$ such that $a_{mn} \in \{0, 1\}$ for all m, n

Example: four seasons

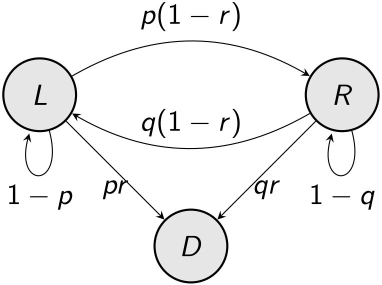
- Let $\{1, 2, 3, 4\}$ denote four seasons (spring, summer, fall, winter)
- Let $a_{mn} = 1$ if season n immediately follows season m, and set $a_{mn} = 0$ otherwise
- Thus adjacency matrix and graph are

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$



Example: animal crossing river

- Animal randomly crosses a river
- Conditional on being on left (right) side of river, it attempts to cross with probability p (q)
- Each time animal crosses river, it drowns with probability r
- Let $\{L, R, D\}$ denote states left, right, and drown
- Transition probability matrix P and graph are



$$P = \begin{bmatrix} 1-p & p(1-r) & pr \\ q(1-r) & 1-q & qr \\ 0 & 0 & 1 \end{bmatrix}$$

Equivalent characterizations of irreducibility

- For $A \in \mathcal{M}_N(\mathbb{C})$, we say $A = (a_{mn})$ is *irreducible* if for all $m \neq n$, there exist $k \in \mathbb{N}$ and indices $m = i_0, i_1, \dots, i_k = n$ such that $a_{i_l i_{l+1}} \neq 0$ for all $l = 0, \dots, k$

Proposition

For $A \in \mathcal{M}_N(\mathbb{C})$, the following conditions are equivalent.

1. *The complex matrix A is irreducible.*
2. *The nonnegative matrix $|A|$ is irreducible.*
3. *For all $m \neq n$, there exist $k \in \{1, \dots, N-1\}$ and indices $m = i_0 \neq i_1 \neq \dots \neq i_k = n$ such that $a_{i_l i_{l+1}} \neq 0$ for all $l = 0, \dots, k$.*
4. $\sum_{k=0}^{N-1} |A|^k \gg 0$.
5. $(I + |A|)^{N-1} \gg 0$.

Perron-Frobenius theorem

Theorem (Perron-Frobenius theorem)

If $A \in \mathcal{M}_N(\mathbb{R})$ is nonnegative and irreducible, the following statements are true.

1. *$\rho(A)$ is an eigenvalue of A (called the Perron root).*
2. *There exist $x, y \gg 0$ (called the right and left Perron vectors) such that $Ax = \rho(A)x$ and $y'A = \rho(A)y'$.*
3. *The vectors x, y are unique up to scalar multiplication (in $\mathbb{C}^N$).*

- Many interesting applications in economics

Chapter 9

Convex Sets

Convex sets

Convex hull

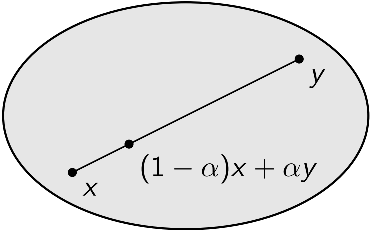
Hyperplanes and half spaces

Separation of convex sets

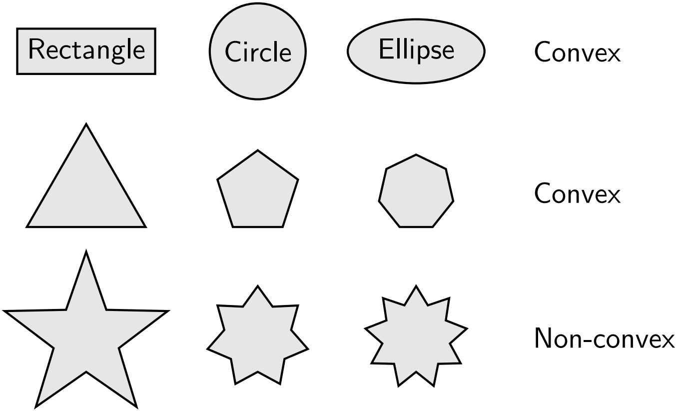
Cone and dual cone

Convex sets

- We say $C \subset \mathbb{R}^N$ is *convex* if line segment joining any two points in C is entirely contained in C
- More formally, C is convex if for any $x, y \in C$ and $\alpha \in [0, 1]$, we have $(1-\alpha)x + \alpha y \in C$



Examples

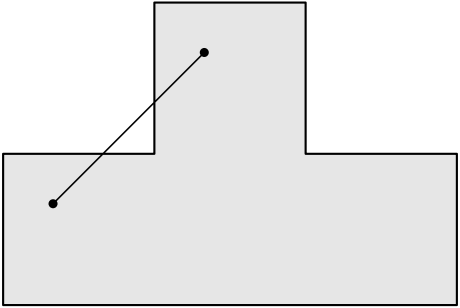


My favorite joke

- Chinese character for "convex"

My favorite joke

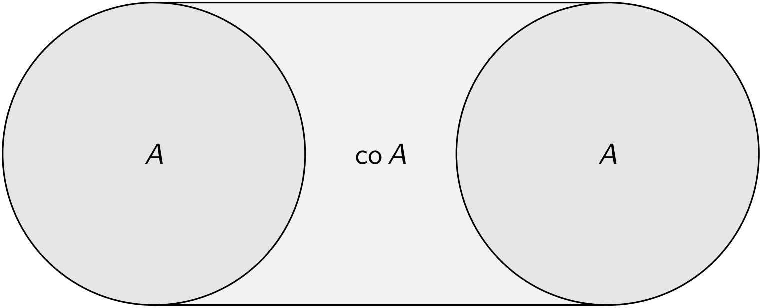
- Chinese character for "convex"
- is not convex



Convex hull

- Let $A \subset \mathbb{R}^N$ be any set
- Smallest convex set that includes A is called *convex hull* of A and is denoted by $\operatorname{co} A$
- To see $\operatorname{co} A$ is well defined, let $\{C_i\}_{i \in I}$ be collection of all convex sets containing A and $C = \bigcap_{i \in I} C_i$
- For any $x, y \in C$ and $\alpha \in [0, 1]$, since $x, y \in C_i$ and C_i is convex, we have $(1 - \alpha)x + \alpha y \in C_i$
- Hence $(1 - \alpha)x + \alpha y \in C$, so C is convex
- But clearly $A \subset C$, and C was intersection of all such convex sets, so C is smallest convex set containing A

Example



Convex combination

- Let $x_k \in \mathbb{R}^N$ for $k = 1, \dots, K$
- Take any numbers α_k for $k = 1, \dots, K$ such that $\alpha_k \geq 0$ and $\sum_{k=1}^K \alpha_k = 1$
- Point of form $x = \sum_{k=1}^K \alpha_k x_k$ is called *convex combination* of $\{x_k\}_{k=1}^K$ with weights (or coefficients) $\{\alpha_k\}_{k=1}^K$

Lemma

Let $A \subset \mathbb{R}^N$ be any set. Then $\operatorname{co} A$ *consists of all convex combinations of points of A.*

- Actually, in above lemma, we may set $K = N + 1$ when forming convex combination (*Carathéodory theorem*)

Hyplerplanes and half spaces

- In $\mathbb{R}^2$, equation of line is $a_1x_1 + a_2x_2 = c$
- In $\mathbb{R}^3$, equation of plane is $a_1x_1 + a_2x_2 + a_3x_3 = c$
- In $\mathbb{R}^N$, *hyperplane* is

$$\left\{x \in \mathbb{R}^N : \langle a, x\rangle = c\right\}$$

- *Half spaces*:

$$H^+ = \left\{x \in \mathbb{R}^N : \langle a, x\rangle \geq c\right\},$$
$$H^- = \left\{x \in \mathbb{R}^N : \langle a, x\rangle \leq c\right\}$$

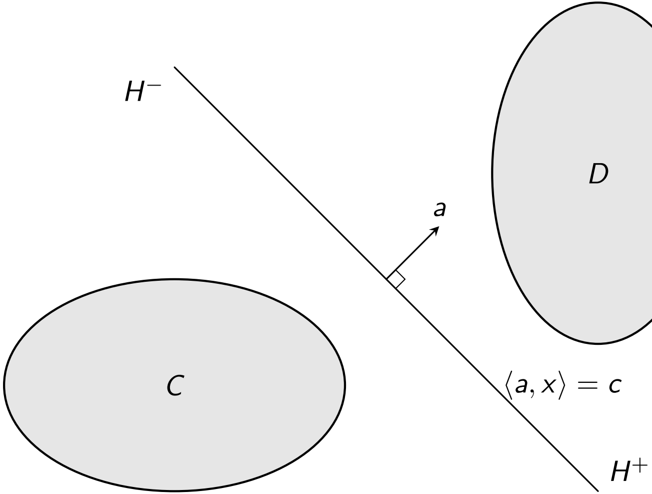
Separation of sets

- Let C, D be two (not necessarily convex) sets
- We say that hyperplane H: $\langle a, x\rangle = c$ *separates* C, D if $C \subset H^-$ and $D \subset H^+$:

$$\begin{aligned} x \in C &\implies \langle a, x\rangle \le c, \\ x \in D &\implies \langle a, x\rangle \ge c. \end{aligned}$$

- Then we call H a *separating hyperplane*

Separation of sets



Separating hyperplane theorem

- Clearly C, D can be separated if and only if

$$\sup_{x \in C} \langle a, x \rangle \leq \inf_{x \in D} \langle a, x \rangle$$

- We say C, D can be *strictly separated* if

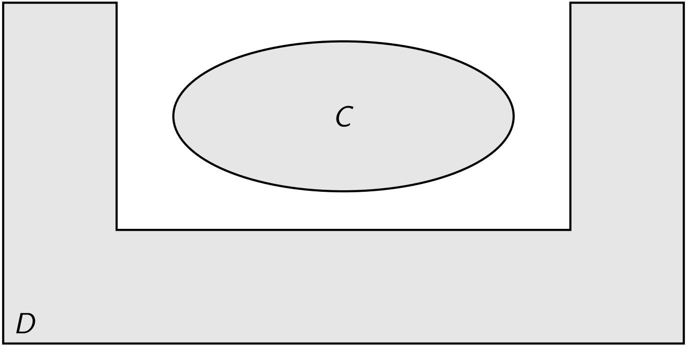
$$\sup_{x \in C} \langle a, x \rangle < \inf_{x \in D} \langle a, x \rangle$$

- One of most important theorems applied in economics is

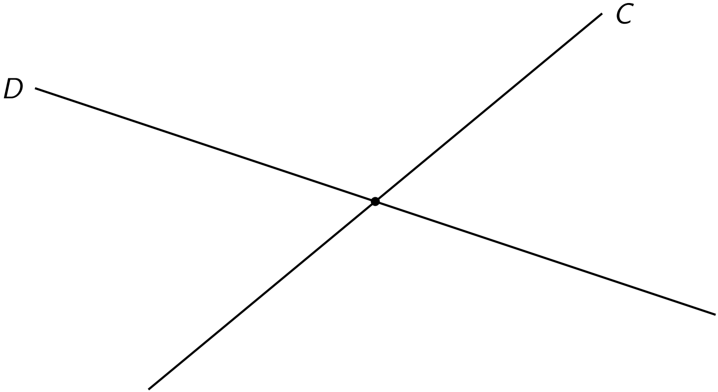
Theorem (Separating hyperplane theorem)

Let $C, D \subset \mathbb{R}^N$ be nonempty, convex, and $C \cap D = \emptyset$. Then there exists a hyperplane that separates C, D. If in addition C, D are closed and one of them is bounded, then they can be strictly separated.

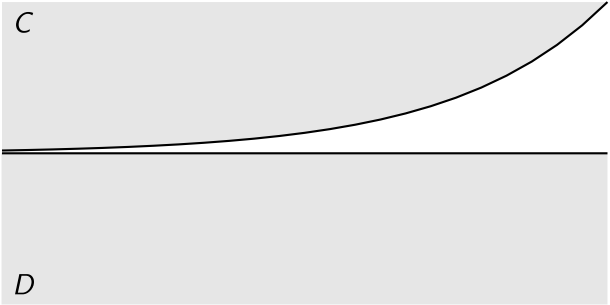
Necessity of convexity for separation



Necessity of empty intersection for separation



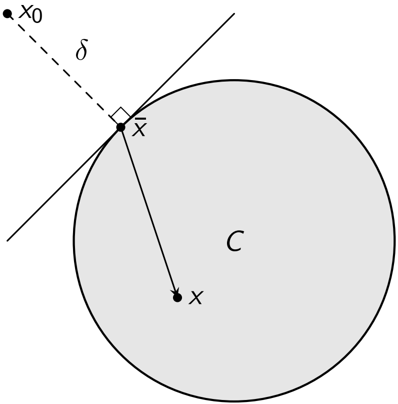
Necessity of boundedness for strict separation



Proof of separating hyperplane theorem

Lemma

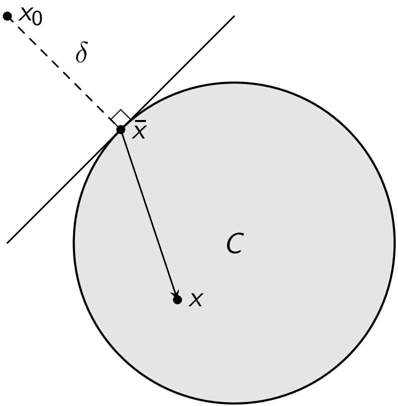
Let $C \subset \mathbb{R}^N$ be nonempty, closed, and convex. Then for any $x_0 \in \mathbb{R}^N$, the minimum distance problem $\min_{x \in C} \|x - x_0\|$ has a unique solution $x = \bar{x}$. Furthermore, for any $x \in C$ we have $\langle x_0 - \bar{x}, x - \bar{x} \rangle \leq 0$.



Proof of separating hyperplane theorem

Proposition

Let $C \subset \mathbb{R}^N$ be nonempty and convex and $x_0 \notin \operatorname{int} C$. Then there exist $0 \neq a \in \mathbb{R}^N$ and $c \in \mathbb{R}$ such $\langle a, x\rangle \leq c \leq \langle a, x_0\rangle$ for any $x \in C$. If $x_0 \notin \operatorname{cl} C$, then the above inequalities can be made strict.



Proof of separating hyperplane theorem

- Define set

$$E = C - D := \{z = x - y : x \in C, y \in D\}$$

- Since C, D are nonempty and convex, so is E
- Since $C \cap D = \emptyset$, we have $0 \notin E$
- By above Proposition, there exists $a \neq 0$ such that $\langle a, z \rangle \leq 0 = \langle a, 0 \rangle$ for all $z \in E$
- By definition of E, we have

$$\langle a, x - y \rangle \leq 0 \iff \langle a, x \rangle \leq \langle a, y \rangle$$

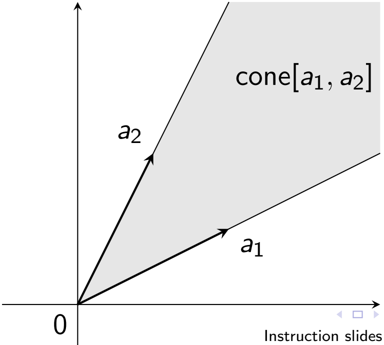
  for all $x \in C$ and $y \in D$
- Taking supremum over $x \in C$ and infimum over $y \in D$, we obtain claim $\square$

Cone

- We say $C \subset \mathbb{R}^N$ is *cone* if $x \in C$ implies $\lambda x \in C$ for all $\lambda > 0$
- Graphically, ray originating from 0 and passing through x is contained in C
- Example: *polyhedral cone*

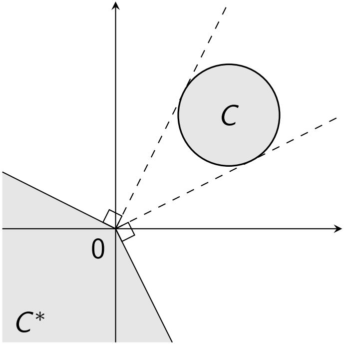
$$C = \text{cone}[a_1, \dots, a_K] := \left\{ x = \sum_{k=1}^{K} \alpha_k a_k : (\forall k) \alpha_k \geq 0 \right\},$$

where $a_1, \dots, a_K \in \mathbb{R}^N$



Dual cone

- Let $C \subset \mathbb{R}^N$ be any nonempty set
- The set
$$C^* = \left\{ y \in \mathbb{R}^N : (\forall x \in C) \, \langle x, y \rangle \leq 0 \right\}$$
is called *dual cone* of C
- Dual cone C^* consists of all vectors that make obtuse angle with any vector in C



Properties of dual cone

Proposition

Let $\emptyset \neq C \subset D$. Then

1. *the dual cone C^* is a nonempty closed convex cone,*
2. *$C^* \supset D^*$, and*
3. *$C^* = (\operatorname{co} C)^*$.*

Proof.

- Clearly $0 \in C^*$, so $C^* \neq \emptyset$
- If $y \in C^*$, then by definition $\langle x, y \rangle \leq 0$ for all $x \in C$
- Then for any $\lambda > 0$ and $x \in C$, we have $\langle x, \lambda y \rangle = \lambda \langle x, y \rangle \leq 0$, so $\lambda y \in C^*$ and C^* is cone
- C^* is intersection of half spaces $H_x^- := \left\{ y \in \mathbb{R}^N : \langle x, y \rangle \leq 0 \right\}$, so it is closed and convex $\square$

Proof

- If $C \subset D$ and $y \in D^*$, then $\langle x, y \rangle \leq 0$ for all $x \in D$, so in particular for all $x \in C$; hence $y \in C^*$
- Setting $D = \operatorname{co} C$, clearly $C^* \supset (\operatorname{co} C)^*$
- To prove reverse inclusion, take any $x \in \operatorname{co} C$; then there exists convex combination $x = \sum_{k=1}^K \alpha_k x_k$ such that $x_k \in C$ for all k
- If $y \in C^*$, it follows that

$$\langle x, y \rangle = \left\langle \sum \alpha_k x_k, y \right\rangle = \sum \alpha_k \langle x_k, y \rangle \leq 0,$$

so $y \in (\operatorname{co} C)^*$ and $C^* \subset (\operatorname{co} C)^*$ $\square$

Dual dual cone

- Let $C^{**} := (C^*)^*$ be dual cone of dual cone
- C and C^{**} closely related

Proposition

Let $C \subset \mathbb{R}^N$ be a nonempty cone. Then $C^{**} = \operatorname{cl} \operatorname{co} C$.

Proof of $\operatorname{cl} \operatorname{co} C \subset C^{**}$.

- If $x \in C$, then $\langle x, y \rangle \leq 0$ for all $y \in C^*$; hence $x \in C^{**}$
- But $C^{**} = (C^*)^*$ is closed convex cone, so $\operatorname{cl} \operatorname{co} C \subset C^{**}$ $\square$

Proof of $\operatorname{cl}\operatorname{co} C \supset C^{**}$

- If $x \notin \operatorname{cl}\operatorname{co} C$, by separating hyperplane theorem we can take $a \neq 0$ and $c \neq 0$ such that
$$\sup_{z \in \operatorname{cl}\operatorname{co} C} \langle a, z \rangle < c < \langle a, x \rangle$$
- In particular,
$$\sup_{z \in C} \langle a, z \rangle < c < \langle a, x \rangle$$
- Since C is cone, for any $\lambda > 0$ we have $\lambda z \in C$ and
$$\lambda \langle a, z \rangle = \langle a, \lambda z \rangle < c < \langle a, x \rangle$$
- Letting $\lambda \to \infty$, it must be $\langle a, z \rangle \leq 0$ for all $z \in C$, and hence $a \in C^*$
- Letting $\lambda \to 0$, get $\langle a, x \rangle > c \geq 0$, so $x \notin C^{**}$ $\square$

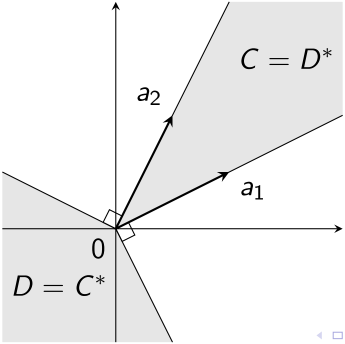
Farkas' lemma

Proposition (Farkas' lemma)

Let $\{a_k\}_{k=1}^K \subset \mathbb{R}^N$ *be vectors and define the sets* $C, D \subset \mathbb{R}^N$ *by*

$$C = \operatorname{cone}[a_1, \dots, a_K],$$
$$D = \left\{ y \in \mathbb{R}^N : (\forall k) \langle a_k, y \rangle \le 0 \right\}.$$

Then $D = C^*$ *and* $C = D^*$*.*



Proof

- For any $x \in C$, by definition of polyhedral cone, we can take $\{\alpha_k\}_{k=1}^K \subset \mathbb{R}_+$ such that $x = \sum_k \alpha_k a_k$
- Then for any $y \in D$, we have

$$\langle x, y \rangle = \sum_k \alpha_k \langle a_k, y \rangle \leq 0,$$

  so $y \in C^*$, which shows $D \subset C^*$
- Conversely, let $y \in C^*$; since $a_k \in C$, we get $\langle a_k, y \rangle \leq 0$ for all k, so $y \in D$, which shows $C^* \subset D$
- Therefore $D = C^*$
- Since C is closed convex cone, by previous proposition, we get

$$C = \operatorname{cl} \operatorname{co} C = C^{**} = (C^*)^* = D^* \square$$

Chapter 10

Convex Functions

Convex function

- Previously we introduced convex functions of single variable and showed that first-order necessary condition for optimality is actually sufficient
- We discuss properties of convex and quasi-convex functions in general setting
- For $f : \mathbb{R}^N \to (-\infty, \infty]$, its *epigraph* is

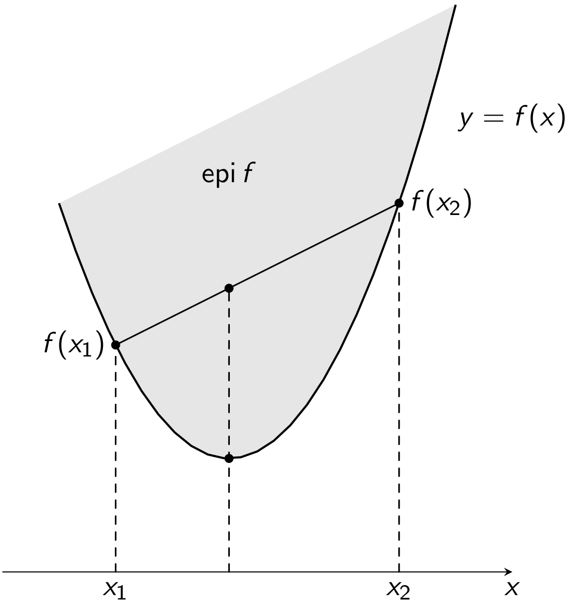
$$\operatorname{epi} f := \left\{(x, y) \in \mathbb{R}^N \times \mathbb{R} : f(x) \le y\right\}$$

- We say f is *convex function* if $\operatorname{epi} f$ is convex set
- Easy to show that f is convex if and only if for any $x_1, x_2 \in \mathbb{R}^N$ and $\alpha \in [0, 1]$, we have *convex inequality*

$$f((1-\alpha)x_1 + \alpha x_2) \le (1-\alpha)f(x_1) + \alpha f(x_2)$$

- If inequality strict whenever $x_1 \ne x_2$ and $\alpha \in (0, 1)$, we say f is *strictly convex*

Convex function



Quasi-convex function

- Set of form
$$L_f(y) := \left\{x \in \mathbb{R}^N : f(x) \le y\right\}$$
is called *lower contour set* of f at level y
- We say that f is *quasi-convex* if lower contour sets are convex for all values of y
- Easy to show that f is quasi-convex if and only if for any $x_1, x_2 \in \mathbb{R}^N$ and $\alpha \in [0,1]$, we have
$$f((1-\alpha)x_1 + \alpha x_2) \le \max\{f(x_1), f(x_2)\}$$
- If inequality strict whenever $x_1 \ne x_2$ and $\alpha \in (0,1)$, we say f is *strictly quasi-convex*

Uniqueness of solution with strict quasi-convexity

Proposition

If $C \subset \mathbb{R}^N$ is nonempty and convex and $f : C \to \mathbb{R}$ is strictly quasi-convex, then the solution to $\min_{x \in C} f(x)$ is unique.

Proof.

- Suppose to contrary that there are two solutions $x_1 \neq x_2$
- Take any $\alpha \in (0,1)$ and let $x = (1-\alpha)x_1 + \alpha x_2$
- Since C is convex, we have $x \in C$
- Since f is strictly quasi-convex, we obtain

$$\begin{aligned} f(x) &= f((1-\alpha)x_1 + \alpha x_2) \\ &< \max\{f(x_1), f(x_2)\} = f(x_1) = \min_{x \in C} f(x), \end{aligned}$$

which is contradiction □

Concave and quasi-concave functions

- f is *concave* if $-f$ is convex, so

$$f((1-\alpha)x_1 + \alpha x_2) \geq (1-\alpha)f(x_1) + \alpha f(x_2)$$

- f is *quasi-concave* if $-f$ is quasi-convex, so

$$f((1-\alpha)x_1 + \alpha x_2) \geq \min\{f(x_1), f(x_2)\}$$

- Strict (quasi-)concavity analogous
- If f strictly quasi-concave, maximum is unique

Convex functions are quasi-convex

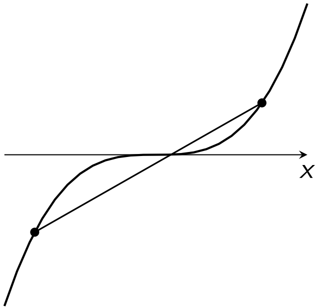
- Let f be convex
- If $x_1, x_2 \in L_f(y)$ and $\alpha \in [0,1]$, we have

$$\begin{aligned} f((1-\alpha)x_1+\alpha x_2) &\le (1-\alpha)f(x_1)+\alpha f(x_2) \\ &\le (1-\alpha)y+\alpha y = y, \end{aligned}$$

  so $(1-\alpha)x_1+\alpha x_2 \in L_f(y)$
- Hence $L_f(y)$ is convex set, so f quasi-convex

Quasi-convex functions are not necessarily convex

- Consider $f(x) = x^3$
- Clearly f is quasi-convex
- But f is not convex



Convexity-preserving operations

- There are many operations that preserve convexity
- Useful for constructing convex functions

Proposition

For each $i = 1, \dots, I$, let $f_i : \mathbb{R}^N \to (-\infty, \infty]$ be convex. Then for any $\beta_i \geq 0$, the function $f := \sum_{i=1}^{I} \beta_i f_i$ is convex.

Proof.

- Take any x_1, x_2 and $\alpha \in [0, 1]$
- Then

$$\begin{aligned} f((1-\alpha)x_1 + \alpha x_2) &= \sum_{i=1}^{I} \beta_i f_i((1-\alpha)x_1 + \alpha x_2) \\ &\leq \sum_{i=1}^{I} \beta_i ((1-\alpha) f_i(x_1) + \alpha f_i(x_2)) \\ &= (1-\alpha) f(x_1) + \alpha f(x_2) \end{aligned}$$

□

Convexity-preserving operations

Proposition

Let I be a nonempty set, and for each $i \in I$, suppose that $f_i : \mathbb{R}^N \to (-\infty, \infty]$ is (quasi-)convex. Then $f := \sup_{i \in I} f_i$ is (quasi-)convex.

Proof.

- Suppose that each f_i is convex
- Since $f_i \leq f$, it follows that

$$f_i((1-\alpha)x_1 + \alpha x_2) \leq (1-\alpha)f(x_1) + \alpha f(x_2)$$

- Taking the supremum over $i \in I$ in the left-hand side, we obtain

$$f((1-\alpha)x_1 + \alpha x_2) \leq (1-\alpha)f(x_1) + \alpha f(x_2)$$

- Proof for quasi-convexity is similar $\square$

Example: support function

- Let $\emptyset \neq A \subset \mathbb{R}^N$
- For each $a \in A$, linear function $f_a(x) := \langle a, x \rangle$ is clearly convex
- Hence by Proposition, function $h_A := \sup_{a \in A} f_a$ defined by $h_A(x) = \sup_{a \in A} \langle a, x \rangle$ is convex
- h_A is called *support function* of set A

Convexity-preserving operations

Proposition

If $f : \mathbb{R}^N \to \mathbb{R}^M$ is convex map and $\phi : \mathbb{R}^M \to \mathbb{R}$ is monotone (quasi-)convex function, then $g := \phi \circ f$ is (quasi-)convex.

Proof.

- Suppose ϕ is convex and take any $x_1, x_2 \in \mathbb{R}^N$ and $\alpha \in [0,1]$
- Since f is convex map, applying ϕ to $f((1-\alpha)x_1 + \alpha x_2) \leq (1-\alpha)f(x_1) + \alpha f(x_2)$, we obtain

$$\begin{aligned}
&g((1-\alpha)x_1 + \alpha x_2) \\
&= \phi(f((1-\alpha)x_1 + \alpha x_2)) \\
&\leq \phi((1-\alpha)f(x_1) + \alpha f(x_2)) && (\because \phi \text{ monotone}) \\
&\leq (1-\alpha)\phi(f(x_1)) + \alpha\phi(f(x_2)) && (\because \phi \text{ convex}) \\
&= (1-\alpha)g(x_1) + \alpha g(x_2)
\end{aligned}$$

- Proof when ϕ is quasi-convex is similar

□

Convexity-preserving operations

Proposition

Let X, Y be vector spaces, $f : X \times Y \to (-\infty, \infty]$ be (quasi-)convex, and define $g : Y \to [-\infty, \infty]$ by $g(y) = \inf_{x \in X} f(x, y)$. Then g is (quasi-)convex.

Proof.

- Suppose f is convex and take $y_1, y_2 \in Y$ and $\alpha \in [0, 1]$
- For each $j = 1, 2$, take any $u_j > g(y_j)$; by the definition of g, we can take x_j such that $g(y_j) \leq f(x_j, y_j) \leq u_j$
- Define $x = (1 - \alpha)x_1 + \alpha x_2$ and similarly for y; using definition of g and convexity of f, we obtain

$$g(y) \leq f(x, y) \leq (1-\alpha)f(x_1, y_1)+\alpha f(x_2, y_2) \leq (1-\alpha)u_1+\alpha u_2$$

- Letting $u_j \downarrow g(y_j)$, we obtain

$$g(y) \leq (1 - \alpha)g(y_1) + \alpha g(y_2)$$

□

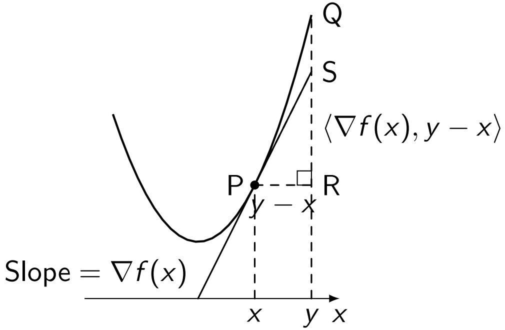
First-order characterization of convexity

Proposition

Let $U \subset \mathbb{R}^N$ be an open convex set and $f : U \to \mathbb{R}$ be differentiable. Then f is (strictly) convex if and only if

$$f(y) - f(x) \geq (>) \langle \nabla f(x), y - x \rangle$$

for all $x \neq y$.



Sufficiency of first-order condition

Proposition (Sufficiency of first-order condition for convex minimization)

Let $U \subset \mathbb{R}^N$ be open and convex and $f : U \to \mathbb{R}$ be convex and differentiable. If $\nabla f(\bar{x}) = 0$, then $f(\bar{x}) = \min_{x \in U} f(x)$.

Proof.

- Take any $x \in U$
- Since f is convex and $\nabla f(\bar{x}) = 0$, by previous proposition, we have
$$f(x) - f(\bar{x}) \geq \langle \nabla f(\bar{x}), x - \bar{x} \rangle = 0$$
- Therefore $f(\bar{x}) \leq f(x)$ and $f(\bar{x}) = \min_{x \in U} f(x)$ □

Second-order characterization of convexity

Proposition (Second-order characterization of convexity)

Let $U \subset \mathbb{R}^N$ be an open convex set and $f : U \to \mathbb{R}$ be C^2. Then f is convex if and only if the Hessian

$$\nabla^2 f(x) = \left[\frac{\partial^2 f(x)}{\partial x_m \partial x_n}\right]$$

is positive semidefinite for all x. Furthermore, if $\nabla^2 f$ is positive definite for all x, then f is strictly convex.

Proof.

Use Taylor and first-order characterization $\square$

First-order characterization of quasi-convexity

Proposition

Let $U \subset \mathbb{R}^N$ be an open convex set and $f : U \to \mathbb{R}$ be differentiable. Then f is quasi-convex if and only if

$$f(y) \le f(x) \implies \langle \nabla f(x), y - x \rangle \le 0$$

for all $x \neq y$.

Second-order characterization of quasi-convexity

Proposition

Let $U \subset \mathbb{R}^N$ be an open convex set and $f : U \to \mathbb{R}$ be C^2. Then the following statements are true.

1. *If f is quasi-convex, then for all x and $v \neq 0$, we have*

$$\langle \nabla f(x), v \rangle = 0 \implies \left\langle v, \nabla^2 f(x) v \right\rangle \geq 0.$$

2. *If for all x and $v \neq 0$ we have*

$$\langle \nabla f(x), v \rangle = 0 \implies \left\langle v, \nabla^2 f(x) v \right\rangle > 0,$$

then f is strictly quasi-convex.

Continuity of convex functions

Theorem

Let $U \subset \mathbb{R}^N$ be an open convex set and $f : U \to \mathbb{R}$ be convex. Then f is continuous.

- In finite-dimensional spaces, convex functions are continuous except at boundary points
- To see why convex function need not be continuous at boundary points, consider

$$f(x) = \begin{cases} \infty & \text{if } x < 0 \text{ or } x > 1 \\ 0 & \text{if } 0 \le x < 1 \\ 1 & \text{if } x = 1 \end{cases}$$

- Proof of this theorem is hard (see textbook)

Homogeneous quasi-convex functions

- We say $f : \mathbb{R}^N \to [-\infty, \infty]$ is *homogeneous (of degree 1)* if $f(\lambda x) = \lambda f(x)$ for all x and $\lambda > 0$
- Following theorem shows homogeneous quasi-convex functions are automatically convex, which is nice (proof is hard)

Theorem

Let $C \subset \mathbb{R}^N$ be a nonempty convex cone. Let $f : C \to (-\infty, \infty]$ be

1. *quasi-convex,*
2. *homogeneous, and*
3. *either $f(x) > 0$ for all $x \in C \setminus \{0\}$ or $f(x) < 0$ for all $x \in C \setminus \{0\}$.*

Then f is convex.

Example

- Let $1 \leq p < \infty$ and define $f : \mathbb{R}^N \to \mathbb{R}$ by

$$f(x) = \|x\|_p := \left(\sum_{n=1}^N |x_n|^p \right)^{1/p}$$

- Let $\phi(y) = \frac{1}{p} y^p$ for $y \geq 0$
- Then $\phi'(y) = y^{p-1} \geq 0$ and $\phi''(y) = (p-1)y^{p-2} \geq 0$, so ϕ is increasing and convex
- Hence

$$g(x) := \phi(f(x)) = \frac{1}{p} \sum_{n=1}^N |x_n|^p$$

  is convex, and $f = \phi^{-1} \circ g$ is quasi-convex
- Since f is homogeneous, it is convex
- Can use this to show ℓ^p norm is indeed norm

Important points

- Convex functions: epigraph is convex
- Quasi-convex functions: lower contour sets are convex
- Convex functions are quasi-convex, but not vice versa
- Strict quasi-convexity implies uniqueness of solution to minimization problem
- There are many convexity-preserving operations
- Monotonic transformation of quasi-convex functions are quasi-convex, so quasi-concave functions are suitable for modeling utility
- Homogeneous quasi-convex functions are convex

Chapter 11

Nonlinear Programming

Introduction

- We would like to solve

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in C \end{array}$$

- When objective function f or constraint set C don't have particular structure, we say *nonlinear programming problem*
- Recall
  - $\bar{x} \in C$ is *(global) solution* if $f(\bar{x}) \leq f(x)$ for all $x \in C$
  - $\bar{x}$ is *local solution* if there exists open neighborhood U of $\bar{x}$ such that $f(\bar{x}) \leq f(x)$ for all $x \in C \cap U$
  - $\bar{x}$ is *strict local solution* if above inequality strict whenever $x \neq \bar{x}$

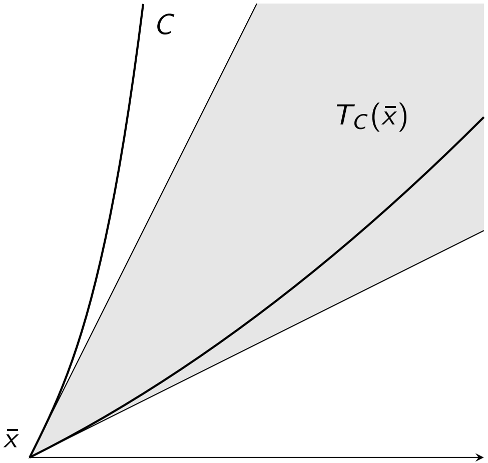
Tangent cone

- To derive first-order necessary condition, we define tangent cone
- Let $\emptyset \neq C \subset \mathbb{R}^N$ be constraint set and $\bar{x} \in C$
- *Tangent cone* of C at $\bar{x}$ is

$$T_C(\bar{x}) := \Big\{ y \in \mathbb{R}^N : (\exists) \{\alpha_k\} \geq 0, \{x_k\} \subset C, \\ \lim_{k \to \infty} x_k = \bar{x}, y = \lim_{k \to \infty} \alpha_k (x_k - \bar{x}) \Big\}$$

- Intuitively, tangent cone of C at $\bar{x}$ consists of all directions y that can be approximated by that from $\bar{x}$ to another point in C

Tangent cone



Tangent cone

Lemma

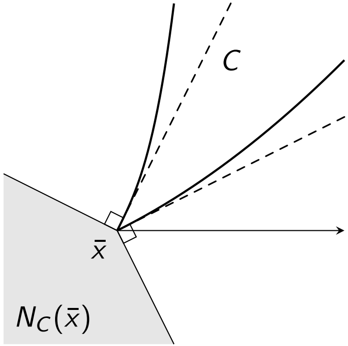
$T_C(\bar{x})$ is a nonempty closed cone.

Proof.

- $0 \in T_C(\bar{x})$, so nonempty
- If $\alpha_k(x_k - \bar{x}) \to y$, then $\beta\alpha_k(x_k - \bar{x}) \to \beta y$ for any $\beta > 0$, so $T_C(\bar{x})$ is cone
- Can show closedness by usual sequential argument $\square$

Normal cone

- Dual cone of tangent cone,
$$N_C(\bar{x}) = (T_C(\bar{x}))^* = \left\{z \in \mathbb{R}^N : (\forall y \in T_C(\bar{x}))\, \langle y, z\rangle \le 0\right\},$$
is called *normal cone* of C at $\bar{x}$



First-order necessary condition

Theorem (First-order necessary condition)

If f is differentiable and $\bar{x}$ is a local solution, then $-\nabla f(\bar{x}) \in N_C(\bar{x})$.

Proof.

- Let $y \in T_C(\bar{x})$ and take sequence such that $\alpha_k \geq 0$, $x_k \to \bar{x}$, and $\alpha_k(x_k - \bar{x}) \to y$
- Since $\bar{x}$ is local solution and f is differentiable, we have

$$0 \leq f(x_k) - f(\bar{x}) = \langle \nabla f(\bar{x}), x_k - \bar{x} \rangle + o(\|x_k - \bar{x}\|)$$

- Multiplying both sides by $\alpha_k \geq 0$, we get

$$\begin{aligned} 0 &\leq \langle \nabla f(\bar{x}), \alpha_k(x_k - \bar{x}) \rangle + \|\alpha_k(x_k - \bar{x})\| \cdot \frac{o(\|x_k - \bar{x}\|)}{\|x_k - \bar{x}\|} \\ &\to \langle \nabla f(\bar{x}), y \rangle + \|y\| \cdot 0 = \langle \nabla f(\bar{x}), y \rangle \end{aligned}$$

□

Inequality and equality constraints

- Consider minimization problem

$$\begin{aligned} &\text{minimize} && f(x) && \\ &\text{subject to} && g_i(x) \le 0 && (i = 1, \dots, I), \\ & && h_j(x) = 0 && (j = 1, \dots, J), \end{aligned}$$

  where f, g_i, h_j's are differentiable
- Constraint set is

$$C = \left\{ x \in \mathbb{R}^N : (\forall i) g_i(x) \le 0, (\forall j) h_j(x) = 0 \right\}$$

- We wish to study shape of C around $\bar{x} \in C$

Linearizing cone

- Let $\bar{x} \in C$
- *Active set* is set of indices of binding constraints, $I(\bar{x}) = \{i : g_i(\bar{x}) = 0\}$
- Since
$$g_i(x) \approx g_i(\bar{x}) + \langle \nabla g_i(\bar{x}), x - \bar{x} \rangle ,$$
for $i \in I(\bar{x})$, condition $g_i(x) \leq 0$ is approximately same as $\langle \nabla g_i(\bar{x}), y \rangle \leq 0$ for $y = x - \bar{x}$
- Similarly, $h_j(x) = 0$ approximately same as $\langle \nabla h_j(\bar{x}), y \rangle = 0$ for $y = x - \bar{x}$
- Motivated by this, define *linearizing cone* by
$$L_C(\bar{x}) = \Big\{ y \in \mathbb{R}^N : (\forall i \in I(\bar{x}))\ \langle \nabla g_i(\bar{x}), y \rangle \leq 0, \\ (\forall j)\ \langle \nabla h_j(\bar{x}), y \rangle = 0 \Big\}$$

Linearizing cone

Proposition

If $\bar{x} \in C$, then $\operatorname{co} T_C(\bar{x}) \subset L_C(\bar{x})$.

Proof.

- Let $y \in T_C(\bar{x})$ and take sequence such that $\alpha_k \geq 0$, $x_k \to \bar{x}$, and $\alpha_k(x_k - \bar{x}) \to y$
- By same argument as showing necessary condition, we obtain $\langle \nabla g_i(\bar{x}), y \rangle \leq 0$ for $i \in I(\bar{x})$ and $\langle \nabla h_j(\bar{x}), y \rangle = 0$, so $y \in L_C(\bar{x})$
- Therefore $T_C(\bar{x}) \subset L_C(\bar{x})$
- Since $L_C(\bar{x})$ is convex cone, we get $\operatorname{co} T_C(\bar{x}) \subset L_C(\bar{x})$ $\square$

Karush-Kuhn-Tucker theorem

Theorem (Karush-Kuhn-Tucker theorem for nonlinear programming)

Suppose that f, g_i, h_j's are differentiable and $\bar{x}$ is a local solution. If $L_C(\bar{x}) \subset \operatorname{co} T_C(\bar{x})$, then there exist $\lambda \in \mathbb{R}^I_+$ and $\mu \in \mathbb{R}^J$ such that

$$\nabla f(\bar{x}) + \sum_{i=1}^{I} \lambda_i \nabla g_i(\bar{x}) + \sum_{j=1}^{J} \mu_j \nabla h_j(\bar{x}) = 0,$$

$$(\forall i)\ \lambda_i \geq 0,\ g_i(\bar{x}) \leq 0,\ \lambda_i g_i(\bar{x}) = 0.$$

Proof

- We know co $T_C(\bar{x}) \subset L_C(\bar{x})$; by assumption, $L_C(\bar{x}) \subset \text{co } T_C(\bar{x})$; hence $L_C(\bar{x}) = \text{co } T_C(\bar{x})$
- Hence normal cone is

$$N_C(\bar{x}) = (T_C(\bar{x}))^* = (\text{co } T_C(\bar{x}))^* = (L_C(\bar{x}))^*$$

- By Farkas' lemma, $N_C(\bar{x}) = (L_C(\bar{x}))^*$ equals polyhedral cone generated by $\{\nabla g_i(\bar{x})\}_{i \in I(\bar{x})}$ and $\{\pm \nabla h_j(\bar{x})\}$
- Since $-\nabla f(\bar{x}) \in N_C(\bar{x})$, FOC holds
- Complementary slackness follows from feasibility $\square$

Constraint qualification

- Condition of sort "$L_C(\bar{x}) \subset \operatorname{co} T_C(\bar{x})$" is called *constraint qualification*
- It is necessary condition for deriving KKT conditions
- In general, we cannot omit those; example:

$$\begin{aligned} &\text{minimize} && x \\ &\text{subject to} && -x^3 \le 0 \end{aligned}$$

- Solution is clearly $\bar{x} = 0$, but FOC violated because

$$\nabla_x L(\bar{x}, \lambda) = 1 - 3\lambda\bar{x}^2 = 1 \neq 0$$

Constraint qualification

Guignard (GCQ) $L_C(\bar{x}) \subset \operatorname{co} T_C(\bar{x})$.

Abadie (ACQ) $L_C(\bar{x}) \subset T_C(\bar{x})$.

Mangasarian-Fromovitz (MFCQ) $\{\nabla h_j(\bar{x})\}_{j=1}^J$ are linearly independent, and there exists $y \in \mathbb{R}^N$ such that $\langle \nabla g_i(\bar{x}), y\rangle < 0$ for all $i \in I(\bar{x})$ and $\langle \nabla h_j(\bar{x}), y\rangle = 0$ for all j.

Slater (SCQ) g_i's are convex, $h_j(x) = \langle a_j, x\rangle - c_j$ with $\{a_j\}_{j=1}^J$ linearly independent, and there exists $x_0 \in \mathbb{R}^N$ such that $g_i(x_0) < 0$ for all i and $h_j(x_0) = 0$ for all j.

Linear independence (LICQ) The set of vectors

$$\{\nabla g_i(\bar{x})\}_{i \in I(\bar{x})} \cup \{\nabla h_j(\bar{x})\}_{j=1}^J$$

is linearly independent.

Constraint qualification

Theorem

The following implication holds for constraint qualifications:

$$\text{LICQ or SCQ} \implies \text{MFCQ} \implies \text{ACQ} \implies \text{GCQ}.$$

- No need to remember detail of each, but remember this:
  1. If all constraints linear (so g_i, h_j are affine), then GCQ automatically holds, and hence no need to check (see Problem)
  2. Slater is for convex optimization, and it requires existence of point satisfying strict inequalities (usually not hard to check)
  3. Most textbooks only list LICQ or Slater
- Most economic problems have linear constraints (e.g., budget or nonnegativity constraints), so OK

Sufficiency of FOC for convex programming

- KKT theorem is only necessary condition for optimality
- We may derive sufficient conditions under additional structure (e.g., convexity)

Theorem (KKT theorem for convex programming)

Consider the constrained minimization problem, where f, g_i's are differentiable and convex and h_j's are affine. If $\bar{x}, \lambda, \mu$ satisfy the KKT conditions, then $\bar{x}$ is a solution.

Proof

- Since f, g_i's are convex, h_j's are affine, and $\lambda \geq 0$, Lagrangian
$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{I} \lambda_i g_i(x) + \sum_{j=1}^{J} \mu_j h_j(x)$$
is convex in x
- Since FOC holds, we have $\nabla_x L(\bar{x}, \lambda, \mu) = 0$, so $\bar{x}$ achieves minimum of L
- Therefore, for any feasible x, it follows that
$$\begin{aligned} f(\bar{x}) &= f(\bar{x}) + \sum_{i=1}^{I} \lambda_i g_i(\bar{x}) + \sum_{j=1}^{J} \mu_j h_j(\bar{x}) \\ &= L(\bar{x}, \lambda, \mu) \leq L(x, \lambda, \mu) \\ &= f(x) + \sum_{i=1}^{I} \lambda_i g_i(x) + \sum_{j=1}^{J} \mu_j h_j(x) \leq f(x) \end{aligned}$$
- Therefore $\bar{x}$ is solution

Recipe for solving convex minimization problems

1. Verify the functions f, g_i's are differentiable and convex and h_j's are affine
2. Define Lagrangian

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{I} \lambda_i g_i(x) + \sum_{j=1}^{J} \mu_j h_j(x)$$

   Derive first-order condition and complementary slackness condition
3. Solve these conditions; if there is a solution $\bar{x}$, it is solution to minimization problem
4. Note that for necessity, you need to check Slater; but for sufficiency, you don't need to

Sufficiency of FOC for quasi-convex programming

Theorem (KKT theorem for quasi-convex programming)

Consider the minimization problem, where f, g_i's are differentiable and quasi-convex and h_j's are affine. If the Slater condition holds, $\bar{x}, \lambda, \mu$ satisfy the KKT conditions, and $\nabla f(\bar{x}) \neq 0$, then $\bar{x}$ is a solution.

- Proof is harder and uses first-order characterization of quasi-convex functions
- Important because objective function is often quasi-convex in economic problems
- Unlike convex case, need to check Slater condition and $\nabla f(\bar{x}) \neq 0$ (which are usually easy)

Utility maximization problem

- As example, consider utility maximization problem

$$\begin{aligned} &\text{maximize} && u(x) \\ &\text{subject to} && \langle p, x\rangle \le w, \\ &&& x \ge 0 \end{aligned}$$

- Assume $\nabla u \gg 0$, and to prevent zero consumption, assume *Inada condition* $\partial u/\partial x_n \to \infty$ as $x_n \to 0$ for each n
- Lagrangian is $L(x, \lambda) = u(x) + \lambda(w - \langle p, x\rangle)$
- If $p \gg 0$ and $w > 0$, Slater condition trivial
- Hence if u is quasi-concave, then FOC $\nabla u(x) = \lambda p$ sufficient for optimality

Parametric optimization problem

- Utility maximization problem contains price vector p and income w as parameters
- Consider parametric optimization problem

$$\begin{aligned} &\underset{x}{\text{minimize}} && f(x,\theta) \\ &\text{subject to} && g_i(x,\theta) \le 0 && (i=1,\dots,I), \end{aligned}$$

  where $\theta \in \Theta$ (some subset of Euclidean space) is parameter
- Under some regularity conditions discussed in draft, we can show solution $x(\theta)$ is differentiable in parameter θ (*parametric differentiability*)
- Remembering each condition is not worth your time, but essentially
  - f is locally strictly quasi-convex in x, and
  - $\{\nabla g_i\}_{i \in I(\bar{x})}$ is linearly independent

Envelope theorem

- Using parametric differentiability and chain rule, can show *Envelope theorem*
- Essentially, to find rate of change of optimal value, just differentiate Lagrangian with respect to parameter

Theorem (Envelope theorem)

Consider parametric optimization problem as above. Let $\phi(\theta) = f(x(\theta), \theta)$ *be the local minimum value function and*

$$L(x, \lambda, \theta) = f(x, \theta) + \sum_{i=1}^{I} \lambda_i g_i(x, \theta)$$

the Lagrangian. Then ϕ *is differentiable and*

$$\nabla \phi(\theta) = \nabla_\theta L(x(\theta), \lambda(\theta), \theta).$$

Example: utility maximization problem

- Consider utility maximization problem

$$\begin{aligned} &\text{maximize} && u(x) \\ &\text{subject to} && \langle p, x \rangle \le w \end{aligned}$$

- Maximum value $v(p, w)$ is called *indirect utility function*
- Lagrangian is $L(x, \lambda) = u(x) + \lambda(w - \langle p, x \rangle)$
- By envelope theorem, get

$$\begin{aligned} \nabla_p v(p, w) &= \nabla_p L = -\lambda x, \\ \nabla_w v(p, w) &= \nabla_w L = \lambda \end{aligned}$$

- Therefore demand satisfies

$$x(p, w) = -\frac{\nabla_p v(p, w)}{\nabla_w v(p, w)},$$

which is called *Roy's identity*

Parametric continuity

- Sufficient conditions for parametric differentiability are rather strong
- In many applications, we may not need differentiability but only continuity
- For instance, in utility maximization problem

$$\begin{aligned} &\text{maximize} && u(x) \\ &\text{subject to} && \langle p, x\rangle \le w, \end{aligned}$$

  we may be interested only in continuity of solution $x(p, w)$

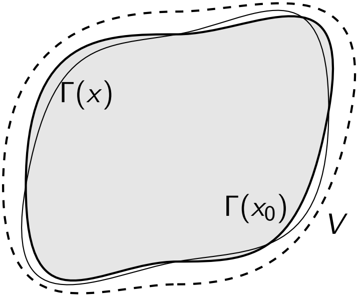
Correspondence

- In utility maximization problem, solution need not be unique unless utility function quasi-concave
- Let X, Y be nonempty sets; if for each $x \in X$ there corresponds subset $\Gamma(x) \subset Y$, we say Γ is *correspondence* from X to Y and write $\Gamma : X \twoheadrightarrow Y$
- Clearly, function f can be viewed as correspondence Γ by considering singleton $\Gamma(x) = \{f(x)\}$
- For any property P (e.g., nonempty, compact, or convex, etc.), we say Γ is P-valued if $\Gamma(x)$ satisfies property P for all $x \in X$

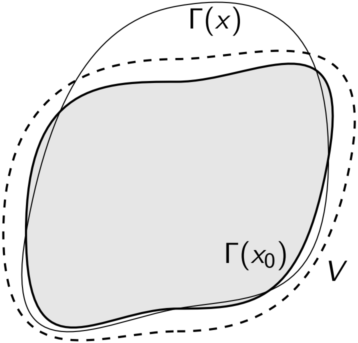
Continuity of correspondence

- Two natural notions of continuity, upper and lower hemicontinuity
- Let X, Y be sets and $\Gamma : X \twoheadrightarrow Y$
- We say Γ is *upper hemicontinuous (uhc)* at x_0 if for any open $V \supset \Gamma(x_0)$, there exists open $U \ni x_0$ such that $x \in U$ implies $\Gamma(x) \subset V$
- We say Γ is *lower hemicontinuous (lhc)* at x_0 if for any open V with $\Gamma(x_0) \cap V \neq \emptyset$, there exists open $U \ni x_0$ such that $x \in U$ implies $\Gamma(x) \cap V \neq \emptyset$
- If both uhc and lhc, we say *continuous*
- Intuitively, uhc correspondences can "expand" but not "shrink", whereas lhc correspondences can "shrink" but not "expand"

Upper hemicontinuity

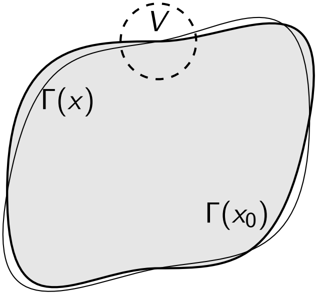


(a) UHC.

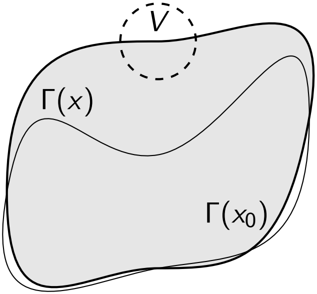


(b) Not UHC.

Lower hemicontinuity



(a) LHC.



(b) Not LHC.

Sequential characterization of UHC

Proposition (Sequential characterization of upper hemicontinuity)

Let $\Gamma : X \twoheadrightarrow Y$ be nonempty. Then the following conditions are equivalent.

1. *Γ is upper hemicontinuous at x and $\Gamma(x)$ is compact.*
2. *For any sequence $\{(x_k, y_k)\} \subset X \times Y$ with $x_k \to x$ and $y_k \in \Gamma(x_k)$, there exists a convergent subsequence $\{y_{k_l}\}$ such that $y_{k_l} \to y \in \Gamma(x)$.*

- *Intuitively, sequence "cannot escape Γ"*
- *UHC: can expand but not shrink*

Sequential characterization of LHC

Proposition (Sequential characterization of lower hemicontinuity)

Let $\Gamma : X \twoheadrightarrow Y$ be nonempty. Then the following conditions are equivalent.

1. *Γ is lower hemicontinuous at x.*
2. *For any sequence $\{x_k\}$ with $x_k \to x$ and any $y \in \Gamma(x)$, there exists a subsequence $\{x_{k_l}\} \subset X$ and a sequence $\{y_l\} \subset Y$ such that $y_l \in \Gamma(x_{k_l})$ for all l and $y_l \to y$.*

- Intuitively, "whatever point in destination, can choose sequence to get there"
- LHC: can shrink but not expand

Maximum theorem

- Following maximum theorem guarantees parametric continuity of maximum value and solution

Theorem (Maximum theorem)

Let X, Y be nonempty metric spaces, $f : X \times Y \to \mathbb{R}$ be a continuous function, and $\Gamma : X \twoheadrightarrow Y$ be a nonempty, compact, continuous correspondence. Let

$$\begin{aligned} f^*(x) &= \max_{y \in \Gamma(x)} f(x,y), \\ \Gamma^*(x) &= \operatorname*{arg\,max}_{y \in \Gamma(x)} f(x,y) \neq \emptyset, \end{aligned}$$

which exist by the extreme value theorem. Then $f^* : X \to \mathbb{R}$ is continuous and $\Gamma^* : X \twoheadrightarrow Y$ is upper hemicontinuous.

Proof.

Immediate from following two lemmas □

USC and UHC lemma

Lemma

Let $f : X \times Y \to \mathbb{R}$ be upper semicontinuous and $\Gamma : X \twoheadrightarrow Y$ be nonempty, compact, and upper hemicontinuous. Then $f^*(x) = \max_{y \in \Gamma(x)} f(x, y)$ is upper semicontinuous.

Proof.

- Take any sequence $\{x_k\}$ with $x_k \to x$; take subsequence $\{x_{k_l}\}$ such that $f^*(x_{k_l}) \to \limsup_{k\to\infty} f^*(x_k)$
- For each l, take $y_{k_l} \in \Gamma(x_{k_l})$ such that $f(x_{k_l}, y_{k_l}) = f^*(x_{k_l})$; since Γ is uhc and compact, by taking subsequence if necessary, we may assume $y_{k_l} \to y \in \Gamma(x)$
- Since f is usc, we have

$$\begin{aligned} f^*(x) \geq f(x, y) &\geq \limsup_{l\to\infty} f(x_{k_l}, y_{k_l}) \\ &= \lim_{l\to\infty} f^*(x_{k_l}) = \limsup_{k\to\infty} f^*(x_k) \end{aligned}$$

□

LSC and LHC lemma

Lemma

Let $f: X \times Y \to \mathbb{R}$ be lower semicontinuous and $\Gamma: X \twoheadrightarrow Y$ be nonempty and lower hemicontinuous. Then $f^*(x) = \sup_{y \in \Gamma(x)} f(x, y)$ is lower semicontinuous.

Proof.

- Take any sequence $\{x_k\}$ with $x_k \to x$ and any $u < f^*(x)$; by definition of f^*, we can take $y \in \Gamma(x)$ such that $f(x, y) > u$; by taking subsequence if necessary, assume $f^*(x_k) \to \liminf_{k \to \infty} f^*(x_k)$
- Since Γ is lhc, we may take subsequence $\{x_{k_l}\}$ and a sequence $\{y_l\}$ such that $y_l \in \Gamma(x_{k_l})$ for all l and $y_l \to y$; then $f^*(x_{k_l}) \geq f(x_{k_l}, y_l)$
- Since f is lower semicontinuous, we have

$$\liminf_{k \to \infty} f^*(x_k) = \liminf_{l \to \infty} f^*(x_{k_l}) \geq \liminf_{l \to \infty} f(x_{k_l}, y_l) \geq f(x, y) > u$$

$\square$

Important points

- For necessity of KKT conditions, we need constraint qualifications
  - If all constraints linear, then (luckily) no need to check
  - If all constraints convex, then Slater is usually most convenient
- Sufficiency of KKT conditions:
  - for convex programming, get for free
  - for quasi-convex programming, get under Slater and $\nabla f(x) \neq 0$
- Parametric differentiability and envelope theorem
- Continuity concepts for correspondences: uhc and lhc
- Maximum theorem: for parametric maximization, if objective function and feasible correspondence continuous, then
  - value function continuous
  - solution set upper hemicontinuous

Chapter 12

Introduction to Dynamic Programming

Introduction

- So far, we have only considered maximization or minimization of given function subject to some constraints
- Such problem is sometimes called *static* optimization problem because there is only one decision to make
- In some cases, writing down or evaluating objective function itself may be complicated
- Furthermore, in many problems, decision maker makes multiple decisions over time instead of single decision
- We will discuss several examples

Knapsack problem

- You break into jewelry shop to steal jewelry
- Your knapsack has size (capacity) S, which is integer
- Types of jewelry: $i = 1, \dots, I$
- Type i jewelry has size s_i and worth w_i
- Your goal is to maximize total value $\sum_{i=1}^{I} w_i n_i$ of stolen jewelry, where n_i: number of type i jewelry to pack

Formulating problem

- Knapsack problem is simple constrained optimization problem:

$$\begin{aligned} &\text{maximize} && \sum_{i=1}^{I} w_i n_i \\ &\text{subject to} && \sum_{i=1}^{I} s_i n_i \le S, \\ &&& (\forall i) n_i \in \mathbb{Z}_+ \end{aligned}$$

- However, cannot be solved by KKT theorem because n_i is contrained to be integer and cannot apply calculus

Dynamic programming formulation

- We solve knapsack problem by dynamic programming
- Let $V(S)$ be maximum total value of jewelry that can be packed in size S knapsack (*value function*)
- Clearly $V(S) = 0$ if $S < \min_i s_i$ since you cannot pack anything in this case
- If you put anything at all in your knapsack (so $S \geq \min_i s_i$), clearly you start packing with some type of jewelry
- If you put object i first (with $s_i \leq S$), then you get value w_i and remaining size $S - s_i$
- By definition of value function, if continue packing optimally, you get total value $V(S - s_i)$ from the remaining space

Bellman equation

- Therefore maximum value that you can get (if you first pack object i) is
$$w_i + V(S - s_i)$$
- To pack optimally, need to choose i that maximizes this
- Hence
$$V(S) = \max_{i:s_i \le S} [w_i + V(S - s_i)],$$
which is called *Bellman equation*
- In principle, can iterate Bellman equation backwards starting from $V(S) = 0$ for $S < \min_i s_i$ to find maximum value
- This process is called *backward induction* or *value function iteration*

Example

- Let $I = 3$ (three types), $(s_1, s_2, s_3) = (1, 2, 5)$, and $(w_1, w_2, w_3) = (1, 3, 8)$
- Then

$$\begin{aligned}
V(0) &= 0, \\
V(1) &= w_1 + V(0) = 1, \\
V(2) &= \max_i [w_i + V(2 - s_i)] = \max \{1 + V(1), 3 + V(0)\} \\
&= \max \{2, 3\} = 3, \\
V(3) &= \max_i [w_i + V(3 - s_i)] = \max \{1 + V(2), 3 + V(1)\} \\
&= \max \{4, 4\} = 4, \\
V(4) &= \max \{1 + V(3), 3 + V(2)\} = \max \{5, 6\} = 6, \\
V(5) &= \max \{1 + V(4), 3 + V(3), 8 + V(0)\} = \max \{7, 7, 8\} = 8
\end{aligned}$$

- No closed-form solution, but writing computer program to solve numerically is straightforward

Shortest path problem

- There are finitely many locations indexed by $i = 1, \dots, I$
- Traveling directly from i to $j \neq i$ costs $c_{ij} \geq 0$
  - (If there is no direct route from i to j, simply define $c_{ij} = \infty$)
- You want to find cheapest way to travel from any point i to any other point j

Dynamic programming formulation

- Let $V_N(i,j)$ be minimum cost to travel from i to j in at most N steps
- For convenience, allow possibility $i=j$ (staying at same location) and set $c_{ii}=0$
- Let k be first connection (including possibly $k=i$); traveling from i to k costs c_{ik}, and now need to travel from k to j in at most $N-1$ steps
- If continue optimally, cost from k to j is (by definition of value function) $V_{N-1}(k,j)$
- Therefore Bellman equation is

$$V_N(i,j)=\min_k \{c_{ik}+V_{N-1}(k,j)\}$$

Optimal savings problem

- Time is indexed by $t = 0, 1, \dots, T$
- Initial wealth $w_0 > 0$
- At each point in time, you can either consume some of your wealth or save it at gross interest rate $R > 0$
- You cannot go in debt; what is optimal consumption-saving plan?

Dynamic programming formulation

- Let w_t be wealth at beginning of time t
- If consume c_t, budget constraint is

$$w_{t+1} = R(w_t - c_t)$$

- For concreteness, assume that the utility function is

$$U_T(c_0, \dots, c_T) = \sum_{t=0}^{T} \beta^t \log c_t,$$

  where subscript T in U_T means that planning horizon is T
- Clearly we have

$$U_T(c_0, \dots, c_T) = \log c_0 + \beta U_{T-1}(c_1, \dots, c_T)$$

Dynamic programming formulation

- Let $V_T(w)$ be maximum utility when you start with initial wealth w and planning horizon is T
- If $T = 0$, you consume everything, so $V_0(w) = \log w$
- If $T > 0$ and you consume c this period, by budget constraint you have wealth $w' = R(w - c)$ next period and planning horizon will be $T - 1$
- Therefore Bellman equation is

$$V_T(w) = \max_{0 \le c \le w} [\log c + \beta V_{T-1}(R(w - c))]$$

Value function iteration

- In principle, we can compute $V_T(w)$ by iterating backwards from $T = 0$ using $V_0(w) = \log w$
- Let us compute $V_1(w)$, for example
- Using Bellman for $T = 1$ and $V_0(w) = \log w$, we have

$$
\begin{aligned}
V_1(w) &= \max_{0 \le c \le w} [\log c + \beta V_0(R(w - c))] \\
&= \max_{0 \le c \le w} [\log c + \beta \log(R(w - c))]
\end{aligned}
$$

- Right-hand side inside brackets is concave in c, so we can maximize by setting derivative equal to zero: FOC is

$$
\frac{1}{c} + \beta \frac{-1}{w - c} = 0 \iff w - c = \beta c \iff c = \frac{w}{1 + \beta}
$$

Value function iteration

- Therefore value function for $T = 1$ is

$$\begin{aligned} V_1(w) &= \log \frac{w}{1+\beta} + \beta \log \left(R \frac{\beta w}{1+\beta} \right) \\ &= (1+\beta) \log w + \text{constant}, \end{aligned}$$

  where "constant" is some constant that depends only on given parameters β and R
- For general T, we may guess that functional form of V_T is

$$V_T(w) = (1 + \beta + \dots + \beta^T) \log w + \text{constant},$$

  and apply mathematical induction to confirm it
- See draft for details

Abstract formulation

Definition

A *dynamic program* is a tuple $\mathcal{D} = \{\mathsf{X}, \mathsf{A}, \Gamma, \mathsf{V}, H\}$, where

- X is a nonempty set called the *state space*,
- A is a nonempty set called the *action space*,
- $\Gamma : \mathsf{X} \twoheadrightarrow \mathsf{A}$ is a nonempty correspondence called the *feasible correspondence*, with its graph denoted by

$$\mathsf{G} := \{(x, a) \in \mathsf{X} \times \mathsf{A} : a \in \Gamma(x)\},$$

- V is a nonempty space of functions $v : \mathsf{X} \to [-\infty, \infty]$ called the *value space*,
- $H : \mathsf{G} \times \mathsf{V} \to [-\infty, \infty]$ is a function called the *aggregator*, which is increasing in the last argument:

$$v_1 \leq v_2 \implies H(x, a, v_1) \leq H(x, a, v_2)$$

Idea of abstract dynamic program

- Given state $x \in \mathsf{X}$, decision maker can take some actions $a \in \mathsf{A}$
- Let $\Gamma(x) \subset \mathsf{A}$ denote all possible actions
- Let $v(x')$ be continuation value that decision maker expects when next state is $x' \in \mathsf{X}$; write $v \in \mathsf{V}$
- Now given current state x, action $a \in \Gamma(x)$, and continuation value v, decision maker should be able to evaluate reward (utility); write it $H(x, a, v) \in [-\infty, \infty]$

Bellman operator

- Let $\mathcal{D} = \{\mathsf{X}, \mathsf{A}, \Gamma, \mathsf{V}, H\}$ be dynamic program
- Without loss of generality, we consider maximization problems
- Hence given $v \in \mathsf{V}$, define function $Tv : \mathsf{X} \to [-\infty, \infty]$ by

$$(Tv)(x) := \sup_{a \in \Gamma(x)} H(x, a, v)$$

- Operator T defined on value space V is called the *Bellman operator*

Bellman equation

- Let $\mathcal{D} = \{\mathsf{X}, \mathsf{A}, \Gamma, \mathsf{V}, H\}$ be dynamic program with Bellman operator T
- We say that $v \in \mathsf{V}$ is *value function* of $\mathcal{D}$ if v is fixed point of T, that is, $v = Tv$
- Equation $v = Tv$, or equivalently

  $$v(x) = \sup_{a \in \Gamma(x)} H(x, a, v),$$

  is called *Bellman equation*
- Condition $v = Tv$ is also called *principle of optimality*: optimal policy has property that whatever initial state and actions are, remaining actions must constitute optimal policy with regard to state resulting from first action

Example: knapsack problem

- State space is $\mathsf{X} = \{0, 1, \dots\} = \mathbb{Z}_+$, with state denoted by $S \in \mathsf{X}$
- Action space is $\mathsf{A} = \{0, 1, \dots, I\}$, with action denoted by $i \in \mathsf{A}$ (where "0" corresponds to packing nothing)
- Feasible correspondence is $\Gamma(S) = \{i = 1, \dots, I : s_i \le S\}$ if this is nonempty and $\Gamma(S) = \{0\}$ otherwise
- Value space V is set of all functions $v : \mathsf{X} \to \mathbb{R}$ with $v(S) = 0$ for $S < \min_i s_i$
- Aggregator is

$$H(S, i, v) = \begin{cases} w_i + v(S - s_i) & \text{if } i \ge 1, \\ v(S) & \text{if } i = 0 \end{cases}$$

Example: shortest path problem

- State space is $\mathsf{X} = \mathbb{N} \times \{1, \dots, I\}^2$, with state denoted by $(n, i, j) \in \mathsf{X}$ (where n is number of trips allowed and i, j denote origin and destination)
- Action space is $\mathsf{A} = \{1, \dots, I\}$, with action denoted by transit point $k \in \mathsf{A}$
- Feasible correspondence is $\Gamma(n, i, j) = \mathsf{A}$, entire space
- Value space V is set of all functions $v : \mathsf{X} \to [0, \infty]$
- Aggregator is

$$H(n, i, j, k, v) = \begin{cases} c_{ik} + v(n-1, k, j) & \text{if } n > 1, \\ c_{ij} & \text{if } n = 1 \text{ and } k = j, \\ \infty & \text{if } n = 1 \text{ and } k \neq j \end{cases}$$

Example: optimal savings problem

- State space is $\mathsf{X} = \mathbb{Z}_+ \times \mathbb{R}_+$, with state denoted by $(T, w) \in \mathsf{X}$ (where T is horizon and $w \geq 0$ is wealth)
- Action space is $\mathsf{A} = \mathbb{R}_+$, with action denoted by consumption $c \in \mathsf{A}$
- Feasible correspondence is $\Gamma(T, w) = [0, w]$
- Value space V is set of all functions $v : \mathsf{X} \to [-\infty, \infty)$
- Aggregator is

$$H(T, w, c, v) = \begin{cases} \log c + \beta v(T-1, R(w-c)) & \text{if } T \geq 1, \\ \log c & \text{if } T = 0 \end{cases}$$

Finite-horizon dynamic programs

- In general, analysis of dynamic programs is case-by-case basis
- For finite-horizon DPs, we have existence and uniqueness

Proposition

Let $\mathcal{D} = \{\mathsf{X}, \mathsf{A}, \Gamma, \mathsf{V}, H\}$ be a dynamic program with Bellman operator $T : \mathsf{V} \to \mathsf{V}$. Suppose that

1. *there exists a sequence of subsets $\emptyset = \mathsf{X}_0 \subset \mathsf{X}_1 \subset \cdots \subset \mathsf{X}_n \subset \cdots \subset \mathsf{X}$ with $\bigcup_{n=1}^{\infty} \mathsf{X}_n = \mathsf{X}$,*
2. *for any n, $x \in \mathsf{X}_n$, $a \in \Gamma(x)$, and $v_1, v_2 \in \mathsf{V}$ with $v_1 = v_2$ on X_{n-1}, we have $H(x, a, v_1) = H(x, a, v_2)$.*

Then $\mathcal{D}$ has a unique value function.

Proof

- Take any $v_0 \in \mathrm{V}$ and define $v_n = T^n v_0$; by condition 2, for $x \in \mathrm{X}_1$, value of $H(x, a, v)$ does not depend on v_0
- Therefore for $x \in \mathrm{X}_1$, value of

$$v_1(x) = (Tv_0)(x) = \sup_{a \in \Gamma(x)} H(x, a, v_0)$$

  also does not depend on v_0
- In particular, setting $v_0 = v_1$, we obtain $v_1 = Tv_1$ on X_1
- We prove $v_n = Tv_n$ on X_n by induction; suppose claim is true up to some n, and let $u_n = Tv_n$
- By induction hypothesis, we have $v_n = u_n$ on X_n, so by condition 2, for $x \in \mathrm{X}_{n+1}$, we have $H(x, a, v_n) = H(x, a, u_n)$, and therefore

$$\begin{aligned} v_{n+1}(x) = (Tv_n)(x) &= \sup_{a \in \Gamma(x)} H(x, a, v_n) = \sup_{a \in \Gamma(x)} H(x, a, u_n) \\ &= (Tu_n)(x) = (T^2 v_n)(x) = (Tv_{n+1})(x) \end{aligned}$$

Proof

- Next, define $v \in \mathrm{V}$ by $v(x) = v_n(x)$ if $x \in \mathrm{X}_n$
- To see that v is well defined, suppose $x \in \mathrm{X}_m \cap \mathrm{X}_n$ for some $m < n$
- Then by condition 1 $\mathrm{X}_m \subset \mathrm{X}_n$, and by what we have just proved $v_n = T^{n-m} v_m = v_m$ on X_m, so value of v is unambiguous
- Furthermore, because $\{\mathrm{X}_n\}$ cover entire space X, we have $x \in \mathrm{X}_n$ for some n, so v is defined on entire X
- Thus $v \in \mathrm{V}$ is well defined

Proof

- To show that v is fixed point of T, take any $x \in \mathrm{X}$
- Then by condition 1, we have $x \in \mathrm{X}_n$ for some n, so $v(x) = v_n(x) = (Tv_n)(x) = (Tv)(x)$
- Since x is arbitrary, $v = Tv$
- If u, v are fixed points of T, then on X_1, we have $H(x, a, u) = H(x, a, v)$, so $u = Tu = Tv = v$
- Using condition 2 and applying induction, we have $u = v$ on X_n for all n, and hence $u = v$ on X $\square$

Important points

- Bellman equation is

$$v(x) = \sup_{a \in \Gamma(x)} H(x, a, v),$$

  where
  - $x \in \mathsf{X}$: state
  - $a \in \Gamma(x) \subset \mathsf{A}$: action ($\Gamma$: feasible correspondence),
  - $v \in \mathsf{V}$: value function,
  - H: aggregator
- Principle of optimality is, first action needs to be optimal fixing remaining plan
- To formulate dynamic programming problems, need a lot of practice for identifying state space X, action space A, value space V, and aggregator H
- Unique value function for finite-horizon dynamic programs

Chapter 13

Contraction Methods

Introduction

- Many interesting dynamic programs are *infinite-horizon*
- Example is optimal savings problem:

$$
\begin{aligned}
\text{maximize} \quad & \mathrm{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t) \\
\text{subject to} \quad & (\forall t) w_{t+1} = R(z_t, z_{t+1})(w_t - c_t) + y(z_{t+1}) \\
& (\forall t) 0 \le c_t \le w_t, \\
& w_0 > 0,\ z_0 \text{ given}
\end{aligned}
$$

- Here
  - $\{z_t\}$ is Markov chain with transition probability matrix P
  - $R(z, z') \ge 0$ is gross return on wealth conditional on $z \to z'$
  - $y(z) \ge 0$ is non-financial income in state z
- How to study such problems?

Markov dynamic program

- Let $\mathcal{D} = \{\mathsf{X}, \mathsf{A}, \Gamma, \mathsf{V}, H\}$ be dynamic program (state, action, feasible correspondence, value, aggregator)
- We say $\mathcal{D}$ is additive *Markov dynamic program* (MDP) if
  - state space can be written as $\mathsf{X} \times \mathsf{Z}$, where $\mathsf{Z} = \{1, \dots, Z\}$ is finite set associated with stochastic matrix $P = (P(z, z'))_{z, z' \in \mathsf{Z}}$,
  - aggregator takes additive (expected utility) form

$$H(x, z, a, v) = r(x, z, a) + \beta \sum_{z'=1}^{Z} P(z, z') v(g(x, z, z', a), z'),$$

where $r : \mathsf{X} \times \mathsf{Z} \times \mathsf{A} \to [-\infty, \infty)$ is *reward function*, $g : \mathsf{X} \times \mathsf{Z}^2 \times \mathsf{A} \to \mathsf{X}$ is *law of motion* or *transition function*, and $\beta \in [0, 1)$ is *discount factor*

- Note that summation is conditional expectation $\mathrm{E}[v(x', z') \mid z]$ with $x' = g(x, z, z', a)$

Bellman operator of MDP

- By definition, Bellman operator T is

$$\begin{aligned}(Tv)(x,z) &:= \sup_{a\in\Gamma(x,z)} H(x,z,a,v)\\ &= \sup_{a\in\Gamma(x,z)} \left\{r(x,z,a)+\beta\,\mathrm{E}_z[v(x',z')]\right\}\end{aligned}$$

- Here $\mathrm{E}_z = \mathrm{E}[\cdot \mid z]$ denotes conditional expectation and it is understood that $x' = g(x,z,z',a)$
- We write additive Markov dynamic program as

$$\mathcal{D} = \{\mathsf{X}, \mathsf{Z}, P, \mathsf{A}, \Gamma, \mathsf{V}, r, g, \beta\}$$

Example: optimal savings problem

- For optimal savings problem, we may identify each object of additive MDP as:
- State space is $\mathsf{X} = [0, \infty)$, where state is wealth $w \in \mathsf{X}$
- Action space is $\mathsf{A} = [0, \infty)$, where action is consumption $c \in \mathsf{A}$
- Feasible correspondence is $\Gamma(w, z) = [0, w]$
- Reward is utility $r(w, z, c) = u(c)$
- Transition function is

$$g(w, z, z', c) = R(z, z')(w - c) + y(z')$$

Existence and uniqueness of value function for bounded MDP

- Let $b\mathsf{X}$ or $b(\mathsf{X})$ be space of all bounded functions defined on X, which is Banach endowed with sup norm $\|\cdot\|$

Theorem

Let $\mathcal{D} = \{\mathsf{X}, \mathsf{Z}, P, \mathsf{A}, \Gamma, \mathsf{V}, r, g, \beta\}$ be an additive Markov dynamic program, where $\mathsf{V} = b(\mathsf{X} \times \mathsf{Z})$. Suppose that $r \in b(\mathsf{X} \times \mathsf{Z} \times \mathsf{A})$, so r is bounded. Then the Bellman operator T is a contraction with modulus $\beta \in [0,1)$. Consequently, the following statements are true.

1. *$\mathcal{D}$ has a unique value function v, which is the unique fixed point of T.*
2. *For any $v_0 \in \mathsf{V}$, we have $v = \lim_{k\to\infty} T^k v_0$.*
3. *The approximation error $\left\|T^k v_0 - v\right\|$ has order of magnitude β^k.*

Proof

- By contraction mapping theorem, suffices to show T is contraction
- We verify Blackwell's sufficient conditions
- (Upward shift) If $v \in \mathrm{V} = b(\mathrm{X} \times \mathrm{Z})$, then v is bounded, so for any $c \geq 0$, we have $v + c \in \mathrm{V}$
- (Bounded difference) If $v_1, v_2 \in \mathrm{V}$, then triangle inequality implies $\|v_1 - v_2\| \leq \|v_1\| + \|v_2\| < \infty$
- (Self map) If $v \in \mathrm{V}$, then

$$\begin{aligned} |(Tv)(x,z)| &= \left| \sup_{a \in \Gamma(x,z)} \left\{ r(x,z,a) + \beta \, \mathrm{E}_z[v(x',z')] \right\} \right| \\ &\leq \|r\| + \beta \|v\| < \infty, \end{aligned}$$

so $T : \mathrm{V} \to \mathrm{V}$

Proof

- (Monotonicity) If $v_1, v_2 \in \mathrm{V}$ and $v_1 \leq v_2$ pointwise, then

$$\begin{aligned}(Tv_1)(x,z) &= \sup_{a\in\Gamma(x,z)} \left\{r(x,z,a) + \beta \, \mathrm{E}_z[v_1(x',z')]\right\} \\ &\leq \sup_{a\in\Gamma(x,z)} \left\{r(x,z,a) + \beta \, \mathrm{E}_z[v_2(x',z')]\right\} = (Tv_2)(x,z),\end{aligned}$$

  so $Tv_1 \leq Tv_2$

- (Discounting) If $v \in \mathrm{V}$ and $c \geq 0$, then

$$\begin{aligned}(T(v+c))(x,z) &= \sup_{a\in\Gamma(x,z)} \left\{r(x,z,a) + \beta \, \mathrm{E}_z[v(x',z') + c]\right\} \\ &= \sup_{a\in\Gamma(x,z)} \left\{r(x,z,a) + \beta \, \mathrm{E}_z[v(x',z')]\right\} + \beta c \\ &= (Tv)(x,z) + \beta c,\end{aligned}$$

  so $T(v+c) = Tv + \beta c$ (in particular, $\leq$) $\square$

Sequential and recursive formulations

- Let $\mathcal{D}$ be additive MDP
- Bellman equation is

$$v(x,z) = \sup_{a \in \Gamma(x,z)} \left\{ r(x,z,a) + \beta \operatorname{E}_z[v(x',z')] \right\}$$

- When we write Bellman equation, we formulate problem *recursively*
- Alternatively, can formulate MDP *sequentially* as

$$\begin{aligned} &\text{maximize} && \operatorname{E}_{z_0} \sum_{t=0}^{\infty} \beta^t r(x_t, z_t, a_t) \\ &\text{subject to} && (\forall t) x_{t+1} = g(x_t, z_t, z_{t+1}, a_t) \\ & && (\forall t) a_t \in \Gamma(x_t, z_t), \\ & && (x_0, z_0) \in \mathrm{X} \times \mathrm{Z} \text{ given} \end{aligned}$$

Sequential and recursive formulations

- What is relation between value function of Bellman and solution to sequential problem?
- We say stochastic process of state-action pair $\{(x_t, a_t)\}_{t=0}^{\infty}$ is *feasible* if $a_t \in \Gamma(x_t, z_t)$ for all t given initial state x_0 and Markov chain $\{z_t\}_{t=0}^{\infty}$
- Function $\sigma : \mathrm{X} \times \mathrm{Z} \to \mathrm{A}$ satisfying $\sigma(x, z) \in \Gamma(x, z)$ is called (feasible) *policy function*
- Given $v \in \mathrm{V}$, if

$$\sigma(x, z) \in \operatorname*{arg\,max}_{a \in \Gamma(x,z)} \left\{ r(x, z, a) + \beta \, \mathrm{E}_z[v(x', z')] \right\},$$

we say σ is *v-greedy*

Equivalence of sequential and recursive formulations

Theorem

Let everything be as in Theorem and $v \in \mathsf{V}$ be the unique fixed point of the Bellman operator T. Then the following statements are true.

1. *The supremum value $\bar{v}(x_0, z_0)$ of the sequential dynamic program is well-defined and finite.*
2. *We have $v(x, z) = \bar{v}(x, z)$ for all $(x, z) \in \mathsf{X} \times \mathsf{Z}$.*
3. *If a v-greedy policy σ exists and we define the state-action process $\{(x_t, a_t)\}_{t=0}^{\infty}$ by $a_t = \sigma(x_t, z_t)$ for all t, then $\{(x_t, a_t)\}_{t=0}^{\infty}$ solves the sequential dynamic program.*

- Sequential and recursive formulations equivalent
- Hence we will focus on recursive formulation because more tractable

Proof

- Since r bounded, value of objective function in sequential problem is bounded as

$$\left| \mathrm{E}_{z_0} \sum_{t=0}^{\infty} \beta^t r(x_t, z_t, a_t) \right| \le \sum_{t=0}^{\infty} \beta^t \|r\| = \frac{\|r\|}{1-\beta} < \infty$$

- Therefore objective function is well defined and supremum value exists and finite, denoted by $\bar{v}$
- To prove $v = \bar{v}$, we show $v \le \bar{v}$ and $v \ge \bar{v}$
- Take any (x_0, z_0) and feasible $\{(x_t, a_t)\}$
- Then by Bellman,

$$v(x_t, z_t) \ge r(x_t, z_t, a_t) + \beta \, \mathrm{E}_{z_t}[v(x_{t+1}, z_{t+1})]$$

- Iterating over $t = 0, \dots, T$, get

$$v(x_0, z_0) \ge \mathrm{E}_{z_0} \sum_{t=0}^{T-1} \beta^t r(x_t, z_t, a_t) + \mathrm{E}_{z_0} \beta^T v(x_T, z_T)$$

Proof

- Noting $\|v\| < \infty$ and $\beta \in [0, 1)$, we have

$$\left| \mathrm{E}_{z_0} \beta^T v(x_T, z_T) \right| \le \beta^T \|v\| \to 0$$

- Hence letting $T \to \infty$, get

$$v(x_0, z_0) \ge \mathrm{E}_{z_0} \sum_{t=0}^{\infty} \beta^t r(x_t, z_t, a_t)$$

- Taking supremum over all feasible $\{(x_t, a_t)\}$, get $v(x_0, x_0) \ge \bar{v}(x_0, z_0)$
- To show reverse inequality, take any $\epsilon > 0$
- Then Bellman implies

$$v(x_t, z_t) \le r(x_t, z_t, a_t) + \beta \, \mathrm{E}_{z_t}[v(x_{t+1}, z_{t+1})] + (1 - \beta)\epsilon$$

for some $a_t \in \Gamma(x_t, z_t)$

Proof

- Iterating over $t = 0, \dots, T$, get

$$v(x_0, z_0) \le \mathrm{E}_{z_0} \sum_{t=0}^{T-1} \beta^t r(x_t, z_t, a_t) + \mathrm{E}_{z_0} \beta^T v(x_T, z_T) + (1 - \beta^T)\epsilon$$

- Letting $T \to \infty$, we obtain

$$v(x_0, z_0) \le \mathrm{E}_{z_0} \sum_{t=0}^{\infty} \beta^t r(x_t, z_t, a_t) + \epsilon \le \bar{v}(x_0, z_0) + \epsilon$$

- Letting $\epsilon \downarrow 0$, we obtain $v(x_0, z_0) \le \bar{v}$ $\square$

Properties of value function

- In many applications, we are not only interested in proving existence (and uniqueness) of value function but also establishing properties such as
  - continuity,
  - monotonicity,
  - convexity/concavity
- Following simple lemma very useful

Very simple lemma

Lemma

Let (V, d) *be a complete metric space and* $T : \mathsf{V} \to \mathsf{V}$ *a contraction with a unique fixed point* $v \in \mathsf{V}$. *If* $\emptyset \neq \mathsf{V}_1 \subset \mathsf{V}$ *is closed and* $T\mathsf{V}_1 \subset \mathsf{V}_1$, *then* $v \in \mathsf{V}_1$.

Proof.

- Since V_1 is closed, (V_1, d) is complete metric space
- Since $T : \mathsf{V}_1 \to \mathsf{V}_1$ is contraction, it has unique fixed point $v_1 \in \mathsf{V}_1$
- v_1 is also fixed point of $T : \mathsf{V} \to \mathsf{V}$
- Since v is unique, we must have $v = v_1 \in \mathsf{V}_1$ $\square$

Application: continuity of value function

Proposition

Let X, A *be topological spaces,* r, g *continuous, and* Γ *nonempty, compact, and continuous. Then the value function* v *is continuous and the policy correspondence* σ *is nonempty and uhc.*

Proof.

- Let $\mathrm{V}_1 \subset \mathrm{V}$ be space of bounded continuous functions equipped with sup norm $\|\cdot\|$
- Then V_1 is closed subset of V and hence Banach
- Under maintained assumptions, for $v \in \mathrm{V}_1$, maximum theorem implies $Tv \in \mathrm{V}_1$, so $T\mathrm{V}_1 \subset \mathrm{V}_1$
- By simple lemma, $v \in \mathrm{V}_1$ and hence v is continuous
- Since Γ is nonempty and compact, by extreme value theorem, policy correspondence σ is nonempty, and it is uhc by maximum theorem $\square$

Partial order

- For set X, we say binary relation $\leq$ is *partial order* if
  1. (Reflexivity) $x \leq x$ for all $x \in X$,
  2. (Antisymmetry) if $x \leq y$ and $y \leq x$, then $x = y$,
  3. (Transitivity) if $x \leq y$ and $y \leq z$, then $x \leq z$
- A set with partial order is called a *partially ordered set*
- Examples:
  - Euclidean space $X = \mathbb{R}^N$ is partially ordered Banach space by declaring $x \leq y$ whenever $x_n \leq y_n$ for all n
  - Function space is partially ordered by declaring $v_1 \leq v_2$ whenever $v_1(x) \leq v_2(x)$ for all x
  - "Set of sets" declare $A \leq B$ if $A \subset B$

Application: monotonicity of value function

Proposition

Let $\mathcal{D}$ be a bounded additive Markov dynamic program. Suppose that X *is partially ordered and* Γ, r, g *are monotone in the sense that, for all* $x_1 \le x_2$, $z, z' \in \mathsf{Z}$, *and* $a \in \Gamma(x_1, z)$, *we have*

$$\begin{aligned} \Gamma(x_1, z) &\subset \Gamma(x_2, z), \\ r(x_1, z, a) &\le r(x_2, z, a), \\ g(x_1, z, z', a) &\le g(x_2, z, z', a). \end{aligned}$$

Then the value function is monotone:
$x_1 \le x_2 \implies v(x_1, z) \le v(x_2, z)$.

Proof

- Let $\mathrm{V}_1 \subset \mathrm{V}$ be set of bounded monotone functions, which is closed
- If $v \in \mathrm{V}_1$, then for any $x_1 \leq x_2$, we have

$$\begin{aligned}
(Tv)(x_1, z) &= \sup_{a \in \Gamma(x_1, z)} \left\{ r(x_1, z, a) + \beta \, \mathrm{E}_z[v(g(x_1, z, z', a), z')] \right\} \\
&\leq \sup_{a \in \Gamma(x_1, z)} \left\{ r(x_2, z, a) + \beta \, \mathrm{E}_z[v(g(x_2, z, z', a), z')] \right\} \\
&\leq \sup_{a \in \Gamma(x_2, z)} \left\{ r(x_2, z, a) + \beta \, \mathrm{E}_z[v(g(x_2, z, z', a), z')] \right\} \\
&= (Tv)(x_2, z),
\end{aligned}$$

where first inequality uses monotonicity of r, g, v and second inequality uses the monotonicity of Γ
- Therefore Tv is monotone and $T\mathrm{V}_1 \subset \mathrm{V}_1$, so claim follows from simple lemma $\square$

Application: concavity of value function

Proposition

Let everything be as in Proposition and suppose that the state space X *and the action space* A *are vector spaces. If r, g are concave in (x, a), then the value function is monotone and concave in x.*

Proof.

- Let V_2 be space of bounded monotone concave function, which is closed
- Recall that if f convex map and ϕ monotone convex function, then $\phi \circ f$ convex
- Hence if f concave map and ϕ monotone concave function, then $\phi \circ f$ concave (by carefully looking at proof)

Proof

- Hence
$$r(x,z,a)+\beta\,\mathrm{E}_z[v(g(x,z,z',a),z')]$$
concave in (x,a)
- By discussion of convexity-preserving operations,
$$(Tv)(x,z)=\sup_{a\in\Gamma(x,z)}\left\{r(x,z,a)+\beta\,\mathrm{E}_z[v(g(x,z,z',a),z')]\right\}$$
is concave in x
- Therefore Tv is monotone and concave and $T\mathrm{V}_2\subset \mathrm{V}_2$, so claim follows from simple lemma $\square$

Unbounded rewards

- Although solving additive Markov dynamic programs based on contraction principle is elegant, reward function needs to be bounded
- However, some reward functions commonly used in applications are unbounded
- For instance, consider optimal savings problem with utility

$$u(c) = \begin{cases} \frac{c^{1-\gamma}}{1-\gamma} & \text{if } 0 < \gamma \neq 1, \\ \log c & \text{if } \gamma = 1, \end{cases}$$

  where parameter $\gamma > 0$ governs risk aversion
- This u is unbounded above if $0 < \gamma < 1$, unbounded below if $\gamma > 1$, and unbounded both from above and below if $\gamma = 1$

Restricting spaces

- Sometimes we may get around unboundedness by restricting spaces
- In optimal savings, suppose u is strictly increasing, bounded above, and income is always positive, so $\underline{y} := \min_{z \in Z} y(z) > 0$; then

$$\underline{u} := u(\underline{y}) > -\infty \quad \text{and} \quad \bar{u} := u(\infty) < \infty$$

- Due to budget constraint, agent is guaranteed to have wealth $w_t \geq \underline{y} > 0$, so we may restrict state space to $X = [\underline{y}, \infty)$
- For any feasible state-action process $\{(w_t, c_t)\}$, value agent gets is restricted to range

$$\frac{\underline{u}}{1-\beta} \leq \mathrm{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t) \leq \frac{\bar{u}}{1-\beta}$$

- Therefore, without loss of generality we may restrict value space to v with $\frac{\underline{u}}{1-\beta} \leq v(x,z) \leq \frac{\bar{u}}{1-\beta}$, and can apply previous results

Stochastic growth model

- Another example is stochastic growth model:

$$\begin{aligned} &\text{maximize} && \mathrm{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t) \\ &\text{subject to} && (\forall t) w_{t+1} = g(w_t, z_t, z_{t+1}, c_t), \\ &&& (\forall t) 0 \le c_t \le w_t, \\ &&& w_0 > 0,\ z_0 \text{ given} \end{aligned}$$

- Common example is

$$g(w, z, z', c) = A(z, z') k^\alpha + (1-\delta) k,$$

where $k := w - c$ is capital, $A(z, z') > 0$ is productivity, $\alpha \in (0,1)$ governs decreasing returns to scale, and $\delta \in (0,1)$ is capital depreciation rate
- Easy to show $\{w_t\}$ bounded, so can allow utility functions that are unbounded above

State-dependent discounting

- Discount factor $\beta \in [0, 1)$ in Markov dynamic program governs patience of decision maker
- When β is large (small), decision maker puts relatively more (less) weight on future rewards and thus can be considered more (less) patient
- For some applications, we may want to consider situations where patience changes over time
- For instance, if decision maker is head of dynasty, even if parent is patient and lives frugally, child may be impatient and spend extravagantly
- We thus consider more general setting where discount factor $\beta(z, z')$ could be state dependent: just change aggregator to

$$\begin{aligned} &H(x, z, a, v) \\ &\qquad = r(x, z, a) + \sum_{z'=1}^{Z} P(z, z')\beta(z, z')v(g(x, z, z', a), z') \end{aligned}$$

Dynamic programming with state-dependent discounting

Theorem

Let $\mathcal{D}$ be a bounded additive Markov dynamic program with state-dependent discounting. If the matrix $B := (P(z,z')\beta(z,z'))$ has spectral radius $\rho(B) < 1$, the following statements are true.

1. *The Bellman operator T is a Perov contraction with coefficient matrix B.*
2. *$\mathcal{D}$ has a unique value function v, which is the unique fixed point of T.*
3. *For any $v_0 \in \mathrm{V}$, we have $v = \lim_{k\to\infty} T^k v_0$.*
4. *For any $\gamma \in (\rho(B), 1)$, the approximation error $\left\| T^k v_0 - v \right\|$ has order of magnitude γ^k.*
5. *If the policy correspondence σ is nonempty, the state-action process generated by σ achieves the maximum of the sequential problem.*

Weighted supremum norm

- As we have seen, common problems have unbounded utility
- Sometimes we may get around by restricting state space or value space, but such approaches ad hoc and lack generality
- Slightly more general approach is to use *weighted supremum norm*
- Let $\psi(x,z)>0$ and set $\tilde{v}=v/\psi$ in Bellman:

$$\begin{aligned}&\psi(x,z)\tilde{v}(x,z)\\&\quad=\sup_{a\in\Gamma(x,z)}\left\{r(x,z,a)+\mathrm{E}_z[\beta(z,z')\psi(x',z')\tilde{v}(x',z')]\right\}\end{aligned}$$

Modified Bellman equation

- Dividing by $\psi(x,z)>0$, get

$$\begin{aligned}&(\tilde{T}\tilde{v})(x,z)\\&:=\sup_{a\in\Gamma(x,z)}\left\{\tilde{r}(x,z,a)+\mathrm{E}_z\left[\beta(z,z')\frac{\psi(x',z')}{\psi(x,z)}\tilde{v}(x',z')\right]\right\},\end{aligned}$$

  where $\tilde{r}:=r/\psi$
- To make $\tilde{T}$ (Perov) contraction, all we need is to control ratio $\psi(x',z')/\psi(x,z)$
- Hence define

$$\tilde{\beta}(z,z'):=\beta(z,z')\sup_{x\in\mathsf{X}}\sup_{a\in\Gamma(x,z)}\frac{\psi(g(x,z,z',a),z')}{\psi(x,z)}$$

  and let $B:=(P(z,z')\tilde{\beta}(z,z'))$

Unique fixed point with weighted supremum norm

Theorem

Let $\mathcal{D}$ be additive Markov dynamic program associated with function $\psi : \mathsf{X} \times \mathsf{Z} \to (0, \infty)$, where V is space of all function v satisfying

$$\sup_{x \in \mathsf{X}} \frac{|v(x,z)|}{\psi(x,z)} < \infty.$$

For $v_1, v_2 \in \mathsf{V}$, define the vector-valued metric $d : \mathsf{V} \to \mathbb{R}^Z_+$ by

$$d_z(v_1, v_2) = \sup_{x \in \mathsf{X}} \frac{|v_1(x,z) - v_2(x,z)|}{\psi(x,z)}.$$

Let B be as above. If $\rho(B) < 1$, then following statements are true.

1. *The (modified) Bellman operator T ($\tilde{T}$) is a Perov contraction on V ($b(\mathsf{X} \times \mathsf{Z})$) with coefficient matrix B.*
2. *$\mathcal{D}$ has a unique value function $v = \psi \tilde{v}$, where $\tilde{v}$ is the unique fixed point of $\tilde{T}$ in $b(\mathsf{X} \times \mathsf{Z})$.*

Example: optimal savings with unbounded utility

- Consider optimal savings problem, where u could be unbounded from both above and below
- Consider weight function $\psi(w,z)=w+b$, where $b>0$; then

$$\begin{aligned}\frac{\psi(g(w,z,z',c),z')}{\psi(w,z)}&=\frac{R(z,z')(w-c)+y(z')+b}{w+b}\\&\le\frac{R(z,z')w+y(z')+b}{w+b}\le\max\{1,R(z,z')\}+\frac{y(z')}{b}\end{aligned}$$

- Letting $b\to\infty$, RHS arbitrarily close to $\max\{1,R(z,z')\}$
- Hence sufficient condition for existence of a solution is $u(w)/(w+b)$ is bounded above (concavity of u suffices) and that $\tilde{\beta}(z,z'):=\beta\max\{1,R(z,z')\}$ satisfies assumption of theorem

Optimal savings with CRRA utility

- Consider optimal savings problem with $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$ with $0 < \gamma < 1$
- If we consider weight function $\psi(w, z) = (w + b)^{1-\gamma}$ for $b > 0$, by similar argument we may set

$$\tilde{\beta}(z, z') := \beta \max \left\{1, R(z, z')^{1-\gamma}\right\}$$

- Satisfying assumptions of Theorem becomes even easier (because $R^{1-\gamma} < R$ whenever $R > 1$)

Numerical dynamic programming

- Almost all dynamic programming problems do not admit closed-form solutions and must be solved numerically
- Consider Markov dynamic program described above, and for simplicity assume $x, a \in \mathbb{R}$
- Because computer can accept only finitely many objects, first step to solve problem is to discretize state space X
- Take some N, and let $\mathsf{X}_N = \{x_1, \dots, x_N\}$ be finite grid, where $x_1 < \cdots < x_N$

Parameterizing value function

- We parameterize value function by finitely many numbers $\{v(x_n, z)\}_{n=1}^{N}{}_{z=1}^{Z} \in \mathbb{R}^{NZ}$
- Then value space is $\mathrm{V}_N := \mathbb{R}^{NZ}$, which is Banach
- Suppose we use some interpolation/extrapolation method to evaluate v on entire state space X, for instance linear interpolation on interval $[x_1, x_N]$ and extrapolation by constants outside
- With slight abuse of notation, we use same symbol V_N to denote space of functions defined on entire X by interpolation/extrapolation

Bellman operator

- Bellman operator T is

$$(Tv)(x,z) := \max_{a\in\Gamma(x,z)} \left\{ r(x,z,a) + \beta \operatorname{E}_z[v(x',z')] \right\}$$
$$= \max_{a\in\Gamma(x,z)} \left\{ r(x,z,a) + \beta \sum_{z'=1}^{Z} P(z,z')v(g(x,z,z,'a),z') \right\}$$

- If $v \in \mathrm{V}_N$ and we use particular interpolation/extrapolation method to evaluate $v(g(x,z,z,'a),z')$, then computing RHS for each (w_n,z) pair, we obtain new numbers $\{(Tv)(x_n,z)\}_{n=1}^{N}{}_{z=1}^{Z}$
- Thus we may view T as self map from V_N to V_N
- By Blackwell, T is contraction with modulus β
- Hence T has unique fixed point in V_N, which could be thought of as approximation to true value function $v \in \mathrm{V}$

Example: stochastic growth model

- For stochastic growth model, Bellman operator is

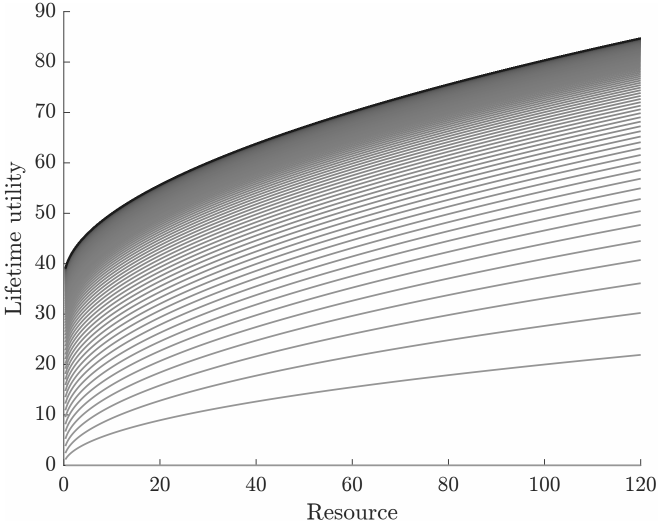
$$(Tv)(w,z) = \max_{0\le k\le w}\left\{u(w-k)+\beta\sum_{z'=1}^{Z} P(z,z')v(A(z,z')k^\alpha+(1-\delta)k,z')\right\}$$

- Two-state Markov chain with $Z=\{1,2\}$ with transition probability $P(z,z')=0.8$ if $z=z'$ and $P(z,z')=0.2$ if $z\neq z'$
- Productivity is

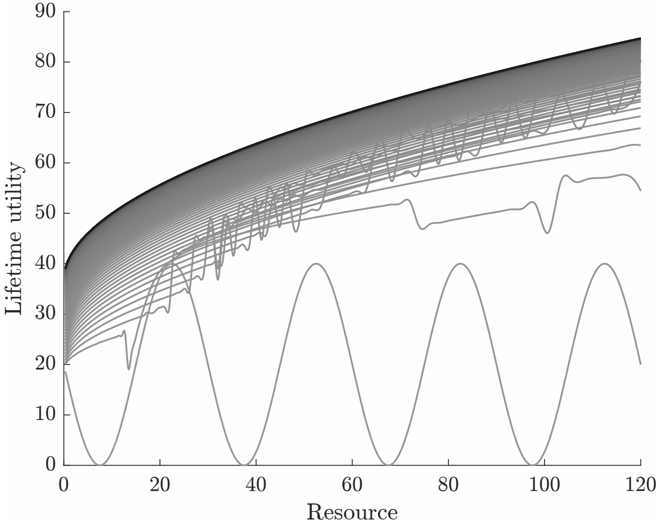
$$A(z,z')=\begin{cases}1.1 & \text{if } z'=1,\\ 0.9 & \text{if } z=1,\end{cases}$$

  so state 1 is high-productivity state
- Set $\alpha=0.36$ and $\delta=0.08$, $\beta=0.95$, and $\gamma=0.5$
- Use 100-point exponential grid on $[0,120]$ to numerically solve stochastic growth model by value function iteration

Value function iteration



Value function iteration



Optimistic policy iteration

- Value function iteration (VFI) is slow because it maximizes at each iteration
- One way to get around is to perform optimization step only occasionally
- For instance, take $m \in \mathbb{N}$ and suppose we update k-th value function v_k using the Bellman operator

$$v_{k+1} := (Tv_k)(x,z) = \sup_{a \in \Gamma(x,z)} \left\{ r(x,z,a) + \beta \operatorname{E}_z[v_k(x',z')] \right\}$$

  only when $k = ml$ for $l = 0, 1, \dots$
- Otherwise, skip optimization step as

$$v_{k+1} := r(x,z,a) + \beta \operatorname{E}_z[v_k(x',z')],$$

  where we use optimal action a from last optimization step
- *Optimistic policy iteration* (OPI)

Convergence of optimistic policy iteration

Theorem

Let everything be as before. If $v_0 \in \mathrm{V}$ satisfies $v_0 \leq Tv_0$, then the sequence $\{v_k\}_{k=0}^{\infty}$ obtained by optimistic policy iteration converges to the value function v.

- In general, cannot show optimistic policy operator is contraction, but convergence guaranteed if $Tv_0 \geq v_0$
- If r bounded, by adding positive constant if necessary, without loss of generality we may assume that $r \geq 0$
- If we start from $v_0 \equiv 0$, then clearly

$$(Tv_0)(x,z) = (T0)(x,z) = \max_{a \in \Gamma(x,z)} r(x,z,a) \geq 0 = v_0(x,z),$$

so $Tv_0 \geq v_0$ holds

Optimistic policy iteration with $m = 10$

- Takes more iterations (205 instead of 196) but much faster (3 sec instead of 18 sec)