Homework #1, Econometrics III, Spring 2007
Maxwell B. Stinchcombe

All random objects are defined on a probability space $(\Omega, \mathcal{F}, P)$, and $L^2(\Omega, \mathcal{F}, P)$ is the set of $\mathbb{R}$-valued, measurable functions with $\|X\|_2 = \|X\| := \left(\int [X(\omega)]^2 \, dP(\omega)\right)^{\frac{1}{2}} < \infty$, that is, $L^2(\Omega, \mathcal{F}, P)$ is the set of random variables with finite variance. A partition $\mathcal{E} \subset \mathcal{F}$ of Ω is **non-null** if for all $E \in \mathcal{E}$, $P(E) > 0$.

Definition 1. *For any non-null partition $\mathcal{E}$ of Ω and $X \in L^1$, define the* ***conditional expectation of*** X ***given*** $\mathcal{E}$ *as the function*

$$X_{\mathcal{E}}(\omega) = \sum_{E_k \in \mathcal{E}} \left(\frac{1}{P(E_k)} \int_{E_k} X \, dP \right) \cdot 1_{E_k}(\omega). \tag{1}$$

The **Borel σ-field** in $\mathbb{R}^k$ is denoted $\mathcal{B}^k$ or $\mathcal{B}$, and defined as the smallest σ-field containing the open subsets of $\mathbb{R}^k$.

Let $\mathcal{G} \subset \mathcal{F}$ be a σ-field, $L^2(\mathcal{G}) := \{X \in L^2(\Omega, \mathcal{F}, P) : (\forall B \in \mathcal{B}) X^{-1}(B) \in \mathcal{G}\}$. It can be shown that $L^2(\mathcal{G})$ is a closed, linear subspace of $L^2(\Omega, \mathcal{F}, P)$, and that orthogonal projection onto $L^2(\mathcal{G})$ is a continuous linear mapping from $L^2(\Omega, \mathcal{F}, P)$ to its subspace, $L^2(\mathcal{G})$.

Definition 2. *For $Y \in L^2(\Omega, \mathcal{F}, P)$, the* ***conditional expectation of*** Y ***given*** $\mathcal{G}$, *written as $E(Y \big| \mathcal{G})$, is defined by $E(Y \big| \mathcal{G}) = \text{proj}_{L^2(\mathcal{G})}(Y)$.*

The geometry of orthogonal projection is that $E(Y \big| \mathcal{G})$ is the unique solution to the problem

$$\min_{f \in L^2(\mathcal{G})} \|Y - f\|_2. \tag{2}$$

When $\mathcal{G} = \sigma(X)$, $\sigma(X) := X^{-1}(\mathcal{B})$, this can be re-written as

$$\min_{\{f : f(X) \in L^2(\Omega, \mathcal{F}, P)\}} \|Y - f(X)\|_2. \tag{3}$$

The following is immediate from orthogonal projection, but extremely useful nonetheless. It is often used as the definition of the conditional expectation.

Theorem 3. *$E(Y \big| \mathcal{G})$ is the unique (up to sets of measure 0) $\mathcal{G}$-measurable random variable with the property that for all $A \in \mathcal{G}$, $\int_A Y \, dP = \int_A E(Y \big| \mathcal{G}) \, dP$.*

The **conditional expectation of** Y **given** X is written $E(Y \big| X)$ and defined by $E(Y \big| \sigma(X))$.

Theorem 4 (Strong Law of Large Numbers)**.** *If X_n is an iid sequence of random variables with finite expectation, $E\, X_n = \mu$, then $P(\{\omega : S_n(\omega) \to \mu\}) = 1$ where $S_n(\omega) := \frac{1}{n} \sum_{i \le n} X_i(\omega)$. If X_n is an iid sequence without finite expectation, $E\,|X_n| = \infty$, then $P(\{\omega : S_n(\omega) \text{ converges}\}) = 0$.*

1. Show the following using the projection definition of $E(Y|X)$.
   a. Let $\mathcal{G} = \{\emptyset, \Omega\}$. For any Y, $E(Y|\mathcal{G}) = EY$.
   b. Let $\mathcal{G} = \{\emptyset, E, E^c, \Omega\}$ and $Y = 1_B$. $E(Y|\mathcal{G}) = P(B|E) \cdot 1_E + P(B|E^c) \cdot 1_{E^c}$.
   c. Let $\mathcal{G} = \{\emptyset, E, E^c, \Omega\}$ and $Y \in L^2(\Omega, \mathcal{F}, P)$. $E(Y|\mathcal{G}) = E(Y|E) \cdot 1_E + E(Y|E^c) \cdot 1_{E^c}$ where $E(Y|E) := \frac{1}{P(E)} \int_E Y\,dP$ and $E(Y|E^c) := \frac{1}{P(E^c)} \int_{E^c} Y\,dP$.
   d. Show that when $\mathcal{G}$ is the smallest σ-field containing a finite non-null partition $\mathcal{E}$, the two definitions of conditional expecation are equivalent.
2. Let $(\Omega, \mathcal{F}, P) = ((0,1]^2, \mathcal{B}^2, \lambda^2)$, $Y = 1_{(0,\frac{1}{2}] \times (0,\frac{1}{2}]}$, and $X(\omega_1, \omega_2) = \omega_1 + \omega_2$. Find $E(Y|\sigma(X))$ and verify that your answer satisfies the condition given in Theorem 3.
3. Let $\Omega = \{(G,G), (G,B), (B,G), (B,B)\}$ with $\mathcal{F} = \mathcal{P}(\Omega)$, and $P(\{\omega\}) = \frac{1}{4}$. Let $\mathcal{G}$ be the smallest σ-field containing the set $\{(G,G), (G,B), (B,G)\}$. Let A be the event $\{(G,G)\}$. Find $P(A|\mathcal{G})$. [This is a version of the story about meeting an old acquaintance in the street, they have with them a girl child, they tell you that the girl is one of their two children, then some probabilist walking by and overhearing the conversation, asks you "What's the probability that the other child is a girl?" Most people incorrectly answer "$\frac{1}{2}$"]
4. Let $(\Omega, \mathcal{F}, P) = ((0,1], \mathcal{B}, \lambda)$. For each of the following rvs X and Y, find $E(Y|X)$.
   a. $Y(\omega) = \omega$, $X = 1_{(0,\frac{1}{3}]}(\omega)$.
   b. $Y(\omega) = \omega$, $X = 1_{(0,\frac{1}{3}]}(\omega) + 1_{(0.9,1]}(\omega)$.
   c. $Y(\omega) = \omega$, $X = 1_{(0,\frac{1}{3}]}(\omega) + 2 \cdot 1_{(0.9,1]}(\omega)$.
   d. $Y(\omega) = \omega^2$, $X = 1_{(0,\frac{1}{3}]}(\omega)$.
   e. $Y(\omega) = 1_{(\frac{1}{2},1]}(\omega)$, $X = \sum_{k \leq 2^n} \frac{k}{2^n} 1_{(\frac{k}{2^n}, \frac{k+1}{2^n}]}(\omega)$.
   f. $Y(\omega) = 1_{(\frac{1}{3},1]}(\omega)$, $X = \sum_{k \leq 2^n} \frac{k}{2^n} 1_{(\frac{k}{2^n}, \frac{k+1}{2^n}]}(\omega)$.
   g. $Y(\omega) = \omega$, $X = \sum_{k \leq 2^n} \frac{k}{2^n} 1_{(\frac{k}{2^n}, \frac{k+1}{2^n}]}(\omega)$.
5. Read/review the material in §1.4 below. You need not hand in any of the exercises in that subsection, but you should know how to do them. [The other material, on the relation of positive (semi-)definiteness of matrices to differentiable concavity, is sometimes useful, but we'll not use it in this class.]
6. Let $X : \Omega \to \mathbb{R}^k$ and $Y : \Omega \to \mathbb{R}$ be a random vector and a random variable where Y and each X_i, $X = (X_i)_{i=1}^k$, belong to $L^2(\Omega, \mathcal{F}, P)$. Let β^* be the solution to the problem

$$\min_{\beta} E(Y - X'\beta)^2.$$

Suppose that we have data, $X_1, \ldots, X_n$, and that as $T \to \infty$, the empirical moments of the data approach the true moments. Specifically, this means that for all $1 \leq i, j \leq k$,

$$P(\{\omega : \frac{1}{T} \sum_{t=1}^{T} X_{i,t}(\omega) X_{j,t}(\omega) \to E\,X_i X_j\}) = 1,$$

and for all i,

$$P(\{\omega : \frac{1}{T}\sum_{t=1}^{T} X_{i,t}(\omega)Y_t(\omega) \to E\, X_i Y\}) = 1.$$

Let $\widehat{\beta}_T(\omega)$ be the solution to the problem

$$\min_\beta \frac{1}{T}\sum_{t=1}^{T}(Y_t(\omega) - X_t(\omega)'\beta)^2.$$

a. Give conditions on the distribution of X under which β^* exists and is unique, and prove existence and uniqueness under the conditions that you give.

b. Under the conditions you just gave, show that $P(\{\omega : \widehat{\beta}_T(\omega) \to \beta^*\}) = 1$.

7. Let X, Y and $X_1, \ldots, X_n$ be as in the previous problem, and let Z be an $T \times \ell$, $\ell \geq k$, matrix of random variables with $P(\{\omega : Z'X(\omega) \to_T M\}) = 1$ where M is a full rank matrix. For each ω and T, the instrumental variable estimator, $\widehat{\beta}_{IV}$, of β solves $\min_\beta (Y - X\beta)'ZPZ'(Y - X\beta)$ where P is a positive definite $\ell \times \ell$ matrix. Give $\widehat{\beta}_{IV}$.

8. Let $\omega \mapsto X(\omega) \geq 0$ be a random variable with $E\, X < \infty$.

a. Show that for all ω and all $t \geq 0$, $X(\omega) \geq t \cdot 1_{\{X \geq t\}}(\omega)$. [Yes, this is very easy.]

Lemma 5 (Chebyschev)**.** *For integrable* $X \geq 0$, $P(X \geq t) \leq \frac{E\, X}{t}$.

b. Prove Chebyschev's inequality using the previous step.

c. Show that for some random variables and some t, Chebyschev's inequality is tight, that is, it is satisfied as an equality. [Check any random variable with $P(X = 0) = P(X = 2) = \frac{1}{2}$.]

d. Being tight is all fine and good, but for many random variables, one can get a much(!) better inequality. Show that if $Z \sim N(0, 1)$, then $P(|Z| \geq t) \leq \sqrt{2/\pi}(1/t)e^{-\frac{t^2}{2}}$. [As with Chebyschev's inequality, this follows by monotonicity of the integral.]

e. If X is a random variable, then $|X|$ and X^2 and $|X|^p$, $p \in [1, \infty)$ are all non-negative random variables, as are $|X - r|$ and $(X - r)^2$ and $|X - r|^p$, $p \in [1, \infty)$. [The same is true for $p \in (0, 1)$, but these are less useful except as a source of oddities.] When these rvs are integrable, Chebyshev's inequality applies. One most often takes $r = E\, X$. These yield

i. $P(|X - \mu| \geq t) = [1 - P(-t < X - \mu < t)] \leq \frac{E\,|X-\mu|}{t}$.

ii. $P((X - \mu)^2 \geq t^2) \leq \frac{\sigma_X^2}{t^2}$ where $\sigma_X^2 := \text{Var}\,(X)$. By taking square roots, this yields

iii. $P(|X - \mu| \geq t) \leq \frac{\sigma_X^2}{t^2}$.

iv. More generally, $P(|X|^p \geq t^p) \leq \frac{E\,|X|^p}{t^p}$ or $P(|X| \geq t) \leq \frac{E\,|X|^p}{t^p}$ (which sometimes goes by the name of Markov's inequality).

Suppose that X_n is an iid sequence in L^2, show that for all $\epsilon > 0$, $\lim_n P(|S_n - \mu| > \epsilon) = 0$. This is known an the **Weak Law of Large Numbers**.

9. The SLLN does <u>not</u> guarantee that for <u>all</u> ω, $S_n(\omega) \to \mu$.

Let $(\Omega, \mathcal{F}, P) = ((0, 1], \mathcal{B}, \lambda)$. For $k = 0, \ldots, 2^n - 1$, let $Y_{k,n}(\omega) = 1_{(k/2^n,(k+1)/2^n]}(\omega)$, and define $X_n(\omega) = \sum_{k \text{ even}} Y_{k,n}(\omega) - \sum_{k \text{ odd}} Y_{k,n}(\omega)$.

a. Show that $\{X_n : n \in \mathbb{N}\}$ is an iid sequence with $\lambda(\{X_n = -1\}) = \lambda(\{X_n = +1\}) = \frac{1}{2}$ so that $E\, X_n = 0$. Show that if $\omega = k/2^n$, then for all $m \geq n$, $X_m(\omega) = +1$, so that $S_n(\omega) \to \infty$. Thus, there is a dense set of exceptions to $S_n(\omega)$ converging.

b. There is more, it can be shown that there are uncountably many ω such that $S_n(\omega) \not\to 0$. Take a typical $\omega = 0.\omega_1\omega_2\omega_3\ldots$ where each ω_n is the n'th element of the binary expansion of ω. From the SLLN, $\lim_n \frac{1}{n}\#\{i \leq n : \omega_i = 1\} = \frac{1}{2}$ for a set of ω having probability 1. For any such ω, show that $\omega' = 0.0\omega_1 0\omega_2 0\omega_3\ldots$ has the property that $\lim_n \frac{1}{n}\#\{i \leq n : \omega'_i = 1\} = \frac{1}{4}$.

c. Start the previous process of adding 0's into every other spot after ω_N, and consider the union over N of these sets. Show that this gives a dense uncountable set of ω's for which $S_n(\omega) \not\to 0$.

10. (This problem is only for those of you that have had a course that teaches measure theory. The rest of you should read the problem, and be sure you understand what is being claimed.) Suppose that $(\Omega, \mathcal{F}, P) = ((0, 1], \mathcal{B}, \lambda)$. This problem gives a construction of a sequence, X_k, of independent $U[0, 1]$ random variables defined on $(0, 1]$.

For $n \in \mathbb{N}$ and $0 \leq k \leq 2^n - 1$, let $I(k, n) = (\frac{k}{2^n}, \frac{k+1}{2^n}]$ be the k'th dyadic interval of order n. Define $X_1(\omega) = 1_{(\frac{1}{2},1]}(\omega) = 1_{I(1,1)}$, $X_2(\omega) = 1_{I(1,2)}(\omega) + 1_{I(3,2)}(\omega)$, and more generally, $X_n(\omega) = \sum_{k \text{ odd}} 1_{I(k,n)}(\omega)$.

a. Show that X_n is an iid sequence.

b. Show that for all $r \in (0, 1]$, $\lim_n P(\{\omega : \sum_{t\leq n} \frac{X_t(\omega)}{2^t} \to s \text{ for some } s < r) = r$.

c. If $I = \{i_1, i_2, i_3, \ldots\} \subset \mathbb{N}$ is an infinite set of distinct integers, then

$$\lim_n P(\{\omega : \sum_{t\leq n} \frac{X_{i_t}(\omega)}{2^t} \to s \text{ for some } s < r\}) = r.$$

d. Partition $\mathbb{N}$ into a countable collection $I_k = \{i_{k,1}, i_{k,2}, \ldots$ of disjoint infinite subsets. For each k, define $X_k(\omega) = \lim_t \sum_{t\leq n} \frac{X_{i_t}(\omega)}{2^t}$. Show that X_k is a countable collection of iid $U[0, 1]$ random variables defined on the probability space $((0, 1], \mathcal{B}, \lambda)$.

There is a (much) more general version of this result: If $(\Omega, \mathcal{F}, P)$ is a non-atomic probability space, (M, d) is a complete, separable metric space, and μ is a probability distribution on $\mathcal{M}$, the smallest σ-field containing the open subsets of M, then there exists a measurable function $f : \Omega \to M$ such that $P(f^{-1}(B)) = \mu(B)$ for every $B \in \mathcal{M}$. Letting $M = [0, 1]^{\mathbb{N}}$ with the metric $d(x, y) = \sum_n \frac{1}{2^n} \min\{1, |x_n - y_n|\}$ allows you to go from this result to the previous, after some moderately involved work.

1. Differentiability and Concavity

This section develops the negative semi-definiteness of the matrix of second derivatives as being equivalent to the concavity of a twice continuously differentiable function. It also develops the determinant test for negative semi-definiteness. As we need them, we'll mention, but not prove, some basic facts about matrix multiplication and determinants.[1]

1.1. The two results.

Before giving the results, we need some terminology.

Definition 6. *A function $f : C \to \mathbb{R}$ is* ***strictly concave*** *if $\forall x, x' \in C$, $x \neq x'$, and all $\lambda \in (0,1)$, $f(x\lambda x') > f(x)\lambda f(x')$.*

An $n \times n$ matrix $\mathbf{A} = (a_{ij})_{i,j=1,\ldots,n}$ is symmetric matrix if $a_{ij} = a_{ji}$ for all i and j. A symmetric $\mathbf{A}$ is **negative semi-definite** if for all vectors $z \in \mathbb{R}^n$, $z^T\mathbf{A}z \leq 0$, it is **negative definite** if for all $z \neq 0$, $z^T\mathbf{A}z < 0$.

Theorem 7. *A twice continuously differentiable $f : \mathbb{R}^n \to \mathbb{R}$ defined on an open, convex set C is concave (respectively strictly concave) iff for all $x^\circ \in C$ $D^2_x f(x^\circ)$ is negative semi-definite (respectively negative definite).*

The **principal sub-matrices** of a symmetric $n \times n$ matrix $\mathbf{A} = (a_{ij})_{i,j=1,\ldots,n}$ are the $m \times m$ matrices $(a_{ij})_{i,j=1,\ldots,m}$, $m \leq n$. Thus, the 3 principal sub-matrices of the 3×3 matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 4 & \sqrt{3} \\ 0 & \sqrt{3} & 6 \end{bmatrix}$$

are

$$\begin{bmatrix} 3 \end{bmatrix}, \quad \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix}, \text{ and } \begin{bmatrix} 3 & 0 & 0 \\ 0 & 4 & \sqrt{3} \\ 0 & \sqrt{3} & 6 \end{bmatrix}.$$

Theorem 8. *A matrix $\mathbf{A}$ is negative semi-definite (respectively negative definite) iff the sign of m'th principal sub-matrix is either 0 or -1^m (respectively, the sign of the m'th principal sub-matrix is -1^m). It is positive semi-definite (respectively positive definite) if you replace "-1^m" with "$+1^m$" throughout.*

In the following two problems, use Theorem 7 and 8, even though we have not yet proved them.

Exercise 9. *The function $f : \mathbb{R}^2_+ \to \mathbb{R}$ defined by $f(x,y) = x^\alpha y^\beta$, $\alpha, \beta > 0$, is strictly concave on $\mathbb{R}^2_{++}$ if $\alpha + \beta < 1$, and is concave on $\mathbb{R}^2_{++}$ if $\alpha + \beta = 1$.*

Exercise 10. *The function $f : \mathbb{R}^2_+ \to \mathbb{R}$ defined by $f(x,y) = (x^p + y^p)^{1/p}$ is convex on $\mathbb{R}^2_{++}$ if $p \geq 1$ and is concave if $p \leq 1$.*

[1] Multiplying by -1 makes this a section on convexity, something that we'll mention only here.

1.2. **The one dimensional case, $f : \mathbb{R}^1 \to \mathbb{R}$.**

Exercise 11. *Suppose that $f : (a, b) \to \mathbb{R}$ is twice continuously differentiable. [Read the third part of this before starting the first two.]*

(1) Show that if $f''(x) \leq 0$ for all $x \in (a, b)$, then f is concave. [Hint: We know that f' is non-increasing. Pick x, y with $a < x < y < b$ and pick $\alpha \in (0, 1)$, define $z = \alpha x + (1 - \alpha)y$. Note that $(z - x) = (1 - \alpha)(y - x)$ and $(y - z) = \alpha(y - x)$. Show

$$f(z) - f(x) = \int_x^z f'(t)\, dt \geq f'(z)(z - x) = f'(z)(1 - \alpha)(y - x),$$
$$f(y) - f(z) = \int_z^y f'(t)\, dt \leq f'(z)(y - z) = f'(z)\alpha(y - x).$$

Therefore,

$$f(z) \geq f(x) + f'(z)(1 - \alpha)(y - x), \;\; f(z) \geq f(y) - f'(z)\alpha(y - x).$$

Multiply the lhs by α, the rhs by $(1 - \alpha)$, and ... .]

(2) Show that if f is concave, then $f''(x) \leq 0$ for all $x \in (a, b)$. [If not, then $f''(x^\circ) > 0$ for some $x^\circ \in (a, b)$ which implies that f'' is strictly positive on some interval $(a', b') \subset (a, b)$. Reverse the above argument.]

(3) Repeat the previous two problems for strict concavity, changing whatever needs to be changed.

1.3. **The multi-dimensional case, $f : \mathbb{R}^n \to \mathbb{R}$.**

Exercise 12. *Suppose that $f : C \to \mathbb{R}$ is twice continuously differentiable, C an open convex subset of $\mathbb{R}^n$.*

For each $y, z \in \mathbb{R}^n$, define $g_{y,z}(\lambda) = f(y+\lambda z)$ for those λ in the interval $\{\lambda : y+\lambda z \in C\}$.

(1) Show that f is (strictly) concave iff each $g_{y,z}$ is (strictly) concave.

(2) Show that $g''(\lambda) = z^T D^2_x f(x^\circ)z$ where $x^\circ = y + \lambda z$.

(3) Conclude that f is (strictly) concave iff for all $x^\circ \in C$, $D^2 f(x^\circ)$ is negative semi-definite (negative definite).

1.4. **A fair amount of matrix algebra background.** The previous has demonstrated that we sometimes want to know conditions on $n \times n$ symmetric matrices $\mathbf{A}$ such that $z^T\mathbf{A}z \leq 0$ for all z, or $z^T\mathbf{A}z < 0$ for all $z \neq 0$. We are trying to prove that a $\mathbf{A}$ is negative semi-definite (respectively negative definite) iff the sign of m'th principal sub-matrix is either 0 or -1^m (respectively, the sign of the m'th principal sub-matrix is -1^m). This will take a longish detour through eigenvalues and eigenvectors. The detour is useful for the study of linear regression too, so this section is also background for next semester's econometrics course.

Throughout, all matrices have only real number entries.

$|\mathbf{A}|$ denotes the determinant of the square $\mathbf{A}$. Recall that $\mathbf{A}$ is invertible, as a linear mapping, iff $|\mathbf{A}| \neq 0$. (If these statements do not make sense to you, you missed linear algebra and/or need to do some review.)

Exercise 13. *Remember, or look up, how to find determinants for* 2×2 *and* 3×3 *matrices.*

A vector $x \neq 0$ is an **eigenvector** and the number $\lambda \neq 0$ is an **eigenvalue**[2] for $\mathbf{A}$ if $\mathbf{A}x = \lambda x$. Note that $\mathbf{A}x = \lambda x$ iff $\mathbf{A}(rx) = \lambda(rx)$ for all $r \neq 0$. Therefore, we can, and do, normalize eigenvectors by $\|x\| = 1$, which corresponds to setting $r = 1/\|x\|$. There is still some ambiguity, since we could just as well set $r = -1/\|x\|$.

In general, one might need to consider λ's and x's that are imaginary numbers, that is $\lambda = a + bi$ with $i = \sqrt{-1}$. This means that x will need to be imaginary too. To see why, read on.

Lemma 14. $\mathbf{A} = \lambda x$*,* $x \neq 0$*, iff* $(\mathbf{A} - \lambda\mathbf{I})x = 0$ *iff* $|\mathbf{A} - \lambda\mathbf{I}| = 0$.

Proof. You should know why this is true. If not, you need some more review. □

Define $g(\lambda) = |\mathbf{A} - \lambda\mathbf{I}|$ so that g is an n'th degree polynomial in λ. The fundamental theorem of algebra tells us that any n'th degree polynomial has n roots, counting multiplicities, in the complex plane. To be a bit more concrete, this means that there are complex numbers λ_i, $i = 1, \ldots, n$ such that

$$g(y) = (\lambda_1 - y)(\lambda_2 - y)\cdots(\lambda_n - y).$$

The "counting multiplicities" phrase means that the λ_i need not be distinct.

Exercise 15. *Using the quadratic formula, show that if* $\mathbf{A}$ *is a symmetric* 2×2 *matrix, then both of the eigenvalues of* $\mathbf{A}$ *are real numbers. Give a* 2×2 *non-symmetric matrix with real entries having two imaginary eigenvalues.* [*This can be done with a matrix having only* 0*'s and* 1*'s as entries.*]

The conclusion about real eigenvalues in the previous problem is true for general $n\times n$ matrices, and we turn to this result.

From your trigonometry class (or from someplace else), $(a+bi)(c+di) = (ac-bd) + (ad+bd)i$ defines multiplication of complex numbers, and $(a+bi)^* := a - bi$ defines the complex conjugate of the number $(a+bi)$. Note that $rs = sr$ for all complex r, s. Further, $r = r^*$ iff r is a real number. By direct calculation, $(rs)^* = r^*s^*$ for any pair of complex numbers r, s. Complex vectors are vectors with complex numbers as their entries. Their dot product is defined in the usual way, $x\cdot y := \sum_i x_i y_i$. Notationally, $x\cdot y$ may be written $x^T y$. The next proof uses

Exercise 16. *If* r *is a complex number, then* $rr^* = 0$ *iff* $r = 0$*. If* x *is a complex vector, then* $x^T x^* = 0$ *iff* $x = 0$.

[2] "Eigen" is a german word meaning "own." Sometimes eigenvalues are called characteristic roots. The idea that we are building to is that the eigenvalues and eigenvectors tell us everything there is to know about the matrix A.

Lemma 17. *Every eigenvalue of a symmetric* $\mathbf{A}$ *is real, and distinct eigenvectors are real, and orthogonal to each other.*

Proof. The eigenvalue part: Suppose that λ is an eigenvalue and x an associated eigenvector so that

$$\mathbf{A}x = \lambda x. \tag{4}$$

Taking the complex conjugate of both sides,

$$\mathbf{A}x^* = \lambda^* x^* \tag{5}$$

because $\mathbf{A}$ has only real entries.

$$[\mathbf{A}x = \lambda x] \Rightarrow [(x^*)^T \mathbf{A}x = (x^*)^T \lambda x = \lambda x^T x^*],$$

$$[\mathbf{A}x^* = \lambda^* x^*] \Rightarrow [x^T \mathbf{A}x^* = x^T \lambda^* x^* = \lambda^* x^T x^*].$$

Subtracting,

$$(x^*)^T \mathbf{A}x - x^T \mathbf{A}x^* = (\lambda - \lambda^*) x^T x^*.$$

Since the matrix $\mathbf{A}$ is symmetric,

$$(x^*)^T \mathbf{A}x - x^T \mathbf{A}x^* = 0.$$

Since $x \neq 0$, $x^T x^* \neq 0$. Therefore,

$$[(\lambda - \lambda^*) x^T x^* = 0] \Rightarrow [(\lambda - \lambda^*) = 0],$$

which can only happen if λ is a real number.

The eigenvector part: From the previous part, all eigenvalues are real. Since $\mathbf{A}$ is real, this implies that all eigenvectors are also real.

Let $\lambda_i \neq \lambda_j$ be distinct eigenvalues and x_i, x_j their associated eigenvectors so that

$$\mathbf{A}x_i = \lambda_i x_i, \quad \mathbf{A}x_j = \lambda_j x_j.$$

Pre-multiplying by the appropriate vectors,

$$x_j^T \mathbf{A}x_i = \lambda_i x_j^T x_i, \quad x_i^T \mathbf{A}x_j = \lambda_j x_i^T x_j.$$

We know that $x_i^T x_j = x_j^T x_i$ (by properties of dot products). Because $\mathbf{A}$ is symmetric,

$$x_j^T \mathbf{A}x_i = x_i^T \mathbf{A}x_j.$$

Combining,

$$(\lambda_i - \lambda_j) x_j^T x_i = 0.$$

Since $(\lambda_i - \lambda_j) \neq 0$, we conclude that $x_i \cdot x_j = 0$, the orthogonality we were looking for. □

The following uses basic linear algebra definitions.

Exercise 18. *If the* $n \times n$ $\mathbf{A}$ *has* n *distinct eigenvalues, then its eigenvectors form an orthonormal basis for* $\mathbb{R}^n$.

A careful proof shows that if $\mathbf{A}$ has an eigenvalue λ_i with multiplicity $k \geq 2$, then we can pick k orthogonal eigenvectors spanning the k-dimensional set of all x such that

$\mathbf{A}x = \lambda_i x$. There will be infinitely many different ways of selecting such an orthogonal set. You either accept this on faith or go review a good matrix algebra textbook.

Exercise 19. *Find eigenvalues and eigenvectors for*

$$\begin{bmatrix} 4 & \sqrt{3} \\ \sqrt{3} & 6 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 3 & 0 & 0 \\ 0 & 4 & \sqrt{3} \\ 0 & \sqrt{3} & 6 \end{bmatrix}.$$

Let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of $\mathbf{A}$ (repeating any multiplicities), and let $u_1, \dots, u_n$ be a corresponding set of orthonormal eigenvectors. Let $\mathbf{Q} = (u_1, \dots, u_n)$ be the matrix with the eigenvectors as columns. Note that $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$ so that $\mathbf{Q}^{-1} = \mathbf{Q}^T$. Let Λ be the $n \times n$ matrix with $\Lambda_{ii} = \lambda_i$ and with 0's in the off-diagonal.

Exercise 20. *Show that* $\mathbf{Q}^T\mathbf{A}\mathbf{Q} = \Lambda$*, equivalently,* $\mathbf{A} = \mathbf{Q}\Lambda\mathbf{Q}^T$.

Expressing a symmetric matrix $\mathbf{A}$ in this form is called **diagonalizing the matrix**. We have shown that any symmetric matrix can be diagonalized so as to have its eigenvalues along the diagonal, and the matrix that achieves this is the matrix of eigenvectors.

Theorem 21. $\mathbf{A}$ *is negative (semi-)definite iff all of its eigenvalues are less than (or equal to)* 0.

Proof. $z^T\mathbf{A}z = z^T\mathbf{Q}^T\Lambda\mathbf{Q}z = v^T\Lambda v$, and the matrix $\mathbf{Q}$ is invertible. □

1.5. **The Alternating Signs Determinant Test for Concavity.** Now we have enough matrix algebra background to prove what we set out prove, $\mathbf{A}$ is negative semi-definite (respectively negative definite) iff the sign of m'th principal sub-matrix is either 0 or -1^m (respectively, the sign of the m'th principal sub-matrix is -1^m).

We defined $g(y) = |\mathbf{A} - y\mathbf{I}|$ so that g is an n'th degree polynomial in λ, and used the fundamental theorem of algebra (and some calculation) to tell us that

$$g(y) = (\lambda_1 - y)(\lambda_2 - y)\cdots(\lambda_n - y)$$

where the λ_i are the eigenvalues of $\mathbf{A}$. Note that $g(0) = |\mathbf{A} - 0\mathbf{I}| = |\mathbf{A}| = \lambda_1 \cdot \lambda_2 \cdots \lambda_n$, that is,

Lemma 22. *The determinant of a matrix is the product of its eigenvalues.*

We didn't use symmetry for this result.

Recall that the **principal sub-matrices** of a symmetric $n \times n$ matrix $\mathbf{A} = (a_{ij})_{i,j=1,\dots,n}$ are the $m \times m$ matrices $(a_{ij})_{i,j=1,\dots,m}$, $m \leq n$. The following is pretty obvious, but it's useful anyway.

Exercise 23. $\mathbf{A}$ *is negative definite iff for all* $m \leq n$ *and all non-zero* x *having only the first* m *components not equal to* 0, $x^T\mathbf{A}x < 0$.

Looking at $m = 1$, we must check if

$$(x_1, 0, 0, \ldots, 0) \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{pmatrix} x_1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = a_{11}x_1^2 < 0.$$

This is true iff the first principal sub-matrix of $\mathbf{A}$ has the same sign as $-1^m = -1^1 = -1$.

Looking at $m = 2$, we must check if

$$(x_1, x_2, 0, \ldots, 0) \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} < 0.$$

This is true iff the matrix

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

is negative definite, which is true iff all of its eigenvalues are negative. There are two eigenvalues, the product of two negative numbers is positive, so the $m = 2$ case is handled by having the sign of the determinant of the 2×2 principal submatrix being -1^2.

Looking at $m = 3$, we must check if

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

is negative definite, which is true iff all of its eigenvalues are negative. There are three eigenvalues, the product of three negative numbers is negative, so the $m = 3$ case is handled by having the sign of the determinant of the 3×3 principal submatrix being -1^3.

Continue in this fashion, and you have a proof of Theorem 8. Your job is to fill in the details for the negative semi-definite, the positive definite, and the positive semi-definite cases as well.

Exercise 24. *Prove Theorem* 8.