DUALITY TOPICS IN ECONOMICS

MAXWELL B. STINCHCOMBE

Contents

Tentative schedule (first part)

Topic	Source	Week/date
Questions		Week 1, Jan. 16, 2014
Convergence, Cones, etc.	[Corbae et al., 2009]	Week 2, Jan. 23, 2014
TVS's	[Robertson and Robertson, 1980]	Week 3, Jan. 30, 2014
TVS's and Duality	[Robertson and Robertson, 1980]	Week 4, Feb. 6, 2014
Dynamic programming	[Smith and McCardle, 2002]	Week 5, Feb. 13, 2014
Monotone comparative statics	[Athey, 2002]	Week 6, Feb. 20, 2014
Monotone comparative statics	[Milgrom, 1994]	Week 6, Feb. 20, 2014

The idea is that we will cover the math background in a lecture format with some assigned problems to be divvied up between the students, and then we will divvy up the papers between students for presentation.

Overview

The next section begins with simple versions of the four questions this class is aimed at. The presumption is that you can do basic proofs using continuous functions. In a variety of places, we will mention measurable functions. These generalize continuous functions: every continuous function is measurable, and they consist of the class of functions that we can integrate. We will not seriously use them until we cover them, which will be in the second half of the course. However, it is worth talking about them early, if only to give you the motivation to go through learning what they are and how we use them.

1. Four Questions

We will study properties of value functions for dynamic programming, supermodularity in expected utility maximization problems, specification testing in econometrics, and informational continuity results for infinite games with differential information. Getting to the point that you know your way around these topics will mean that you are ready for a huge variety of (yet) more advanced topics.

1.A. Dynamic Programming.

Finite state Markovian dynamic programming: state space $S = \{1, 2, \ldots, \ell\}$, random state $X_t \in S$ at time points $t \in \{0, 1, \ldots\}$; for each $i \in S$, a finite action set $A(i)$ with possible state-action pairs $(i, a) \in \times_{i \in S} A(i)$; a Markovian transition, $P : \times_{i \in S} A(i) \to [0, 1]$, $P_{i,j}(a) := P(X_{t+1} = j | X_t = i, a_t = a)$; a stage utility function $(i, a) \mapsto u(i, a)$. The value to starting in $i \in S$ is the maximum/supremum expected discounted value, $v^*(i) := \max E\left(\sum_{t=0}^{\infty} \beta^t u(X_t, a_t) | X_0 = i\right)$, maximum taken over all possible choices of policies, functions from histories to actions $a_t \in A(X_t)$.

For an infinite state space, e.g. $S = [0, 1]$, need to specify exactly what we mean by $P_{i,j}(a)$, which will require the beginnings of measure theory.

Iterative procedure:

$$\boldsymbol{v}_0(i) := \max_{a \in A(i)} u(i, a), \tag{1}$$

$$\boldsymbol{v}_1(i) := \max_{a \in A(i)} \left[u(i, a) + \beta \sum_j \boldsymbol{v}_0(j) P_{i,j}(a) \right], \tag{2}$$

$$\vdots \qquad\qquad \vdots$$

$$\boldsymbol{v}_{k+1}(i) := \max_{a \in A(i)} \left[u(i, a) + \beta \sum_j \boldsymbol{v}_k(j) P_{i,j}(a) \right]. \tag{3}$$

The properties of $\boldsymbol{v}_0(\cdot)$ depend on properties of $u(\cdot, \cdot)$ and $\times_{i \in S} A(i)$. The properties of $\boldsymbol{v}_1(\cdot)$ depend on properties of $u(\cdot, \cdot)$, $\times_{i \in S} A(i)$, $\boldsymbol{v}_0(\cdot)$, and the $P_{i,j}(a)$. After this, the properties of $\boldsymbol{v}_{k+1}(\cdot)$ depend on properties of $u(\cdot, \cdot)$, $\times_{i \in S} A(i)$, $\boldsymbol{v}_k(\cdot)$, and the $P_{i,j}(a)$.

Question: What are the properties of the $\boldsymbol{v}_k$ and $\boldsymbol{v}^* = \lim_{k \to \infty} \boldsymbol{v}_k$? And what exactly do we mean by $\lim_{k \to \infty} \boldsymbol{v}_k$?

1.B. **Expected Utility Maximization.** In the middle of the previous problem, we have problems that look like

$$V(\beta) = \max_{a \in A} U(a, \beta) = \int_S u(i, a)\, d\beta(i) \text{ or} \tag{4}$$

$$a^*(\beta) = \operatorname*{argmax}_{a \in A} U(a, \beta) = \int_S u(i, a)\, d\beta(i). \tag{5}$$

Here $\beta \in \Delta(S)$, the set of probabilities on S, things are easy if S is finite, a bit more complicated if S is an infinite subset of $\mathbb{R}$.

Question: What are the joint properties of $u(\cdot, \cdot)$, orders on A, and orders on $\Delta(S)$ such that $[\beta' \succsim \beta] \Rightarrow [a^*(\beta') \succsim a^*(\beta)]$?

1.C. **Specification Testing.** A parametric model for the expectation of a random variable $Y \in \mathbb{R}^1$ given $X \in \mathbb{R}^k$ is a class of functions $f : X \times \Theta \to \mathbb{R}$ where the set of parameters, Θ, is a subset of $\mathbb{R}^p$. The model is correctly specified if for some $\theta_\circ \in \Theta$, $E\,(Y|X = x) = f(x, \theta_\circ)$ for a probability 1 set of realizations x.

We will cover, in some detail, the observation that the almost everywhere unique measurable function, $x \mapsto g(x)$, that solves the problem

$$\min_{x \mapsto g(x)} E\,(Y - g(X))^2 \tag{6}$$

has the property that $g(X) = E\,(Y|X)$. Now, this means that $E\,(Y - g(X)|X)$ is equal to 0 almost everywhere. With this in mind, let $\theta^* \in \Theta$ denote the best fitting parameter, e.g. the solution to

$$\min_{\theta \in \Theta} E\,(Y - f(X, \theta))^2. \tag{7}$$

Define $\varepsilon = Y - f(X, \theta^*)$.

The model is **correctly specified** if $e(x) := E\,(\varepsilon|X = x) = 0$ almost everywhere. Being a function of X, the random variable $e(X)$ is equal to 0 if and only if the inner product, $E\, e(X) \cdot h(X)$, is equal to 0 for any function $h(X)$ for which the expectation just given is finite. To check such a thing with data, $(Y_n, X_n) \in (\mathbb{R}^1 \times \mathbb{R}^k)^n$, we construct a sequence of estimators $\widehat{\theta}_n(Y_n, X_n) \to \theta^*$ and check whether we get the sample version of $E\, e(X)h(X)$ equal to 0 plus/minus noise for a "rich enough" class of functions, h.

Question: Can we find classes of functions, $\mathcal{H}$, that simultaneously have the property that $\sup_{h \in \mathcal{H}} E\, e(X) \cdot h(X) = 0$ implies that $e(\cdot)$ is 0 almost everywhere, and for which we can succesfully estimate $\sup_{h \in \mathcal{H}} E\, e(X) \cdot h(X) = 0$ using $\widehat{e}_n = Y_n - f(X_n, \widehat{\theta}_n)$?

1.D. **Games With Differential Information.** A random $\omega \in \Omega$ is drawn according to a distribution P, each i in a finite set of agents, I, observe signals, $s_i(\omega) \in S_i$, and as a function of what they observe, choose a mixed strategy, $\sigma_i \in \Delta(A_i)$. Payoffs are $u_i(\omega, a)$ where $a = (a_i)_{i \in I}$. The utility functions are measurable in ω, continuous in a, and integrable. Following [Milgrom and Weber, 1985], information is *continuous* if the joint distribution of the vector $s = (s_i)_{i \in I}$, has a density with respect to the product of its marginal distributions, which requires, *inter alia*, that there not be any continuously distributed common components in the signals. If this condition is violated, there is a three player game without a Nash equilibrium [Simon, 2003].

1.D.1. *Cournot competition with private information.* Firms $i \in I$ have increasing, convex cost functions $c_i(\cdot)$. As a function of the privately known, stochastic cost structures, $c_i(\cdot)$, and signals about demand conditions, ω_i, they pick their quantities $q_i \in [0, \overline{q}_i]$. Expected profits conditional on signal ω_i are $E\left(\pi_i(c,q)|\omega_i\right) = E\left([q_i p_i((q_j)_{j \in I}) - c_i(q_i)]|\omega_i\right)$ where $p_i(\cdot)$ is i's, possibly random, demand function. One might expect the cost functions and the signals about demand conditions to contain, in general, many continuously distributed common components, including input prices, technological knowledge, and other market information. The question of Nash equilibrium existence is open.

1.D.2. *Signaling games with finite signal spaces.* Two players observe their private information, ω_1 and ω_2. Then the sender chooses a_1 in a *finite* space A_1. The receiver observes a_1 and picks a point in $K(a_1)$, a compact metric space. Thus, the receiver's action space is $A_2 = \times_{a_1 \in A_1} K(a_1)$.

If the u_i are measurable in ω, continuous on each $K(a_1)$, and integrable, then the signaling game fits into the class considered here. When ω_2 is a degenerate random variable and A_1 is not finite, [Manelli, 1996] shows that, to accomodate limits of approximations and to guarantee equilibrium existence, the appropriate strategy space for the sender includes an expansion to allow for cheap talk. When ω_1 and ω_2 have a continuously distributed, commonly observable information, the question of Nash equilibrium existence is open.

1.D.3. *Purifications.* Randomization is crucial to the existence of saddle points in 0-sum games, and, more generally, to Nash's (1950) equilibrium existence theorem for finite games. Despite its crucial role in game theory, randomization is, to many, not an attractive behavioral assumption. [Bellman and Blackwell, 1949] and [Dvoretzky et al., 1951] were early studies of the extent to which randomization might not be needed in games. Harsanyi [Harsanyi, 1973] shows that, generically at least, mixed strategy equilibria are observationally equivalent to pure strategy equilibria of infinitesimal perturbations of the game. A stronger version of this result is in [Govindan et al., 2003].

Fix a game $\mathrm{G}(v) = (A_i, v_i)_{i \in I}$ where each A_i is finite, $v = (v_i)_{i \in I}$, and the utilities $v_i \in \mathbb{R}^A$, $A = \times_i A_i$. For all v outside a closed set having Lebesgue measure 0, every equilibrium of $\mathrm{G}(v) = (A_i, v_i)_{i \in I}$ is regular.

Let $(\omega_i)_{i \in I}$ be an independent collection of random vectors in $\mathbb{R}^A$ assigning, for all $i \in I$ and for any fixed strategy σ_{-i} of the players, mass 0 to the event that $(\omega_i(a_i, \cdot) - \omega_i(a_i', \cdot))$ lies in the hyperplane orthogonal to σ_{-i} (e.g. if the distribution of the ω_i has a density with respect to Lebesgue measure). A **perturbation** of $\mathrm{G}(v)$ is an incomplete information game in which each $i \in I$ observes the vector ω_i, picks an $a_i \in A_i$, and payoffs are $v_i(a) + \omega_i(a)$. A perturbation is a **δ-perturbation** if the distribution of the ω_i is within δ of point mass on 0 in the weak* topology.

For regular equilibria $\overline{\sigma}$, [Govindan et al., 2003] shows that for all $\epsilon > 0$, there is a $\delta > 0$ such that any δ-perturbation of $\mathrm{G}(v)$ has an essentially strict, hence pure strategy, equilibrium inducing a distribution on A that is within ϵ of $\overline{\sigma}$. The interpretation is that mixed strategies played in equilibrium are, observationally, impossible to distinguish from strict pure strategy equilibria in nearby games. These nearby games have independent idiosyncratic shocks to utilities, and the pure strategy equilibria of these nearby games "purify" the (regular) mixed equilibria of $\mathrm{G}(v)$.

The independence of the $\omega_i(a)$ rules out the existence of any continuously distributed common information, and is crucial to the existence of Harsanyi's purifications. [Radner and Rosenthal, 1982] give a generic game $\text{G}(v)$ and expand it using $\omega_i(a)$ that are uniformly distributed on a bounded triangle in $\mathbb{R}^A$. Posterior distributions are, with probability 1, atomless, which is part of what is needed for purification.[1] Because of the way that independence of the signals fails in Radner and Rosenthal's example, the game has a unique equilibrium in which the players randomize after seeing a probability 1 set of signals.

The finiteness of A is also crucial to exact purification. [Khan et al., 1999] present a game in which $A_i = [0,1]$ and exact purification is not possible, even when types are smoothly and independently distributed. They further show that exact purification is possible when the A_i are countable.

1.D.4. *Wars of attrition.* Two players have types t_i smoothly distributed in $(0,1)$. A bounded, increasing value function $v : (0,1) \to [0, \overline{v}]$ gives the value of an object, in terms of the cost of time spent fighting, to a player of type t_i. A pure strategy for i is a mapping b_i from $(0,1)$ to $A_i = [0, 2\cdot\overline{v}]$, with $b_i(t_i)$ being the time at which the player stops fighting for the object. If player j plays a strategy giving an atomless cdf F_j on A_j, the payoff to i of fighting until a is $u_i(t_i,(a,b_j)) = \int_0^a (v(t_i)-s)\,dF_j(s) - a(1-F_j(a))$.

In games with independent types, [Milgrom and Weber, 1985] show that the first order conditions $\partial u_i/\partial a = 0$ contain a great deal of information about pure strategy equilibria. The independence of types rules out any continuously distributed common information, and is the leading special case of Milgrom and Weber's informational diffuseness requirement. They show that if the joint distribution of the types is diffuse in the sense of having a density with respect to the product of its marginals, then the game has jointly continuous expected utility functions and compact strategy sets, leading to Nash equilibrium existence.

Questions: Strategies are measurable functions from S_i to $\Delta(A_i)$. Can we, and if so, how, do we measure the distance between these measurable functions so that expected payoffs are jointly continuous and the set of strategies compact? And what kinds of continuity do we have in the information structures, here modeled as the functions $s_i : \Omega \to S_i$?

1.E. **Review Problems.** Later, the problems will contain extensions of results given. For now, they are aimed at (me) being sure that we are all on the same page in terms of background.

On dynamic programming, §1.A.

DynPg 1. Suppose that $S = \{1, 2, \ldots, \ell\}$ and that $\boldsymbol{v}_k, \boldsymbol{v}^* : S \to \mathbb{R}$. If for each $k \in \mathbb{N}$, $\boldsymbol{v}_k$ is non-decreasing, that is, for each $s < s' \in S$, $\boldsymbol{v}_k(s) \leq \boldsymbol{v}_k(s')$. If for each $s \in S$, $\boldsymbol{v}_k(s) \to \boldsymbol{v}^*(s)$, then $\boldsymbol{v}^*$ is non-decreasing.

DynPg 2. Suppose that $S = \{1, 2, \ldots, \ell\}$ and that $\boldsymbol{v}_k, \boldsymbol{v}^* : S \to \mathbb{R}$. If for each $k \in \mathbb{N}$, $\boldsymbol{v}_k$ is decreasing increments, that is, for each $s < s' < s'' \in S$, $\boldsymbol{v}_k(s'') - \boldsymbol{v}_k(s') \leq \boldsymbol{v}_k(s') - \boldsymbol{v}_k(s)$. If for each $s \in S$, $\boldsymbol{v}_k(s) \to \boldsymbol{v}^*(s)$, then $\boldsymbol{v}^*$ is has decreasing increments.

[1]In related work, [Aumann et al., 1983] showed that each player having, with probability 1, a nonatomic posterior is sufficient for *approximate* purification.

DynPg 3. Suppose that $S = [0,1] \subset \mathbb{R}$ and that $\boldsymbol{v}_k, \boldsymbol{v}^* : S \to \mathbb{R}$. Suppose also that each $\boldsymbol{v}_k$ is continuous and that $\boldsymbol{v}_k(t) \to \boldsymbol{v}^*(t)$ for each $t \in [0,1]$. Show that $\boldsymbol{v}^*$ need not be continuous (at any point $t \in [0,1]$).

DynPg 4. Suppose that $S = [0,1] \subset \mathbb{R}$ and that $\boldsymbol{v}_k, \boldsymbol{v}^* : S \to \mathbb{R}$. Suppose also that each $\boldsymbol{v}_k$ is continuous and concave, and that $\boldsymbol{v}_k(t) \to \boldsymbol{v}^*(t)$ for each $t \in [0,1]$. Show that $\boldsymbol{v}^*$ is concave but may not be continuous.

DynPg 5. Suppose that $S = [0,1] \subset \mathbb{R}$ and that $\boldsymbol{v}_k, \boldsymbol{v}^* : S \to \mathbb{R}$. Suppose also that each $\boldsymbol{v}_k$ is Lipschitz continuous, that the Lipschitz constants for the $\boldsymbol{v}_k$ are uniformly bounded, and that $\boldsymbol{v}_k(t) \to \boldsymbol{v}^*(t)$ for each $t \in [0,1]$. Show that $\boldsymbol{v}^*$ is continuous.

On expected utility maximization, §1.B

Eu 1. Suppose that $i \in S \subset \mathbb{R}$, $a \in A \subset \mathbb{R}$, that $c : A \to \mathbb{R}$ is a monotonically increasing function, and that $u : S \times A \to \mathbb{R}$ has increasing differences, that is, for all $a' > a \in A$ and $i' > i \in S$, $u(a', i') - u(a, i') \geq u(a', i) - u(a, i)$.

a. If β' first order stochastically dominates β, what can we say about the comparison of the solutions to the problems

$$\max_{a \in A} U(a, \beta) = \left[\int_S u(i,a)\, d\beta'(i)\right] - c(a) \text{ and } \tag{8}$$

$$\max_{a \in A} U(a, \beta) = \left[\int_S u(i,a)\, d\beta(i)\right] - c(a)? \tag{9}$$

b. How would the answer change if $c(\cdot)$ is not monotonic?

On specification testing, §1.C

Spec 1. Suppose that X is a random variable taking values in $[0,1]$ and that it has a strictly positive density, $f_X(\cdot)$, i.e. that $P(X \in A) = \int_A f_X(t)\, dt$.

a. If $h(x) = x$, give some and characterize all of the functions $x \mapsto e(x)$ such that $E\, e(X)h(X) = 0$.

b. If $h(x) = x^2$, give some and characterize all of the functions $x \mapsto e(x)$ such that $E\, e(X)h(X) = 0$.

c. If $\mathcal{H} = \{x, x^2\}$, give some and characterize all of the functions $x \mapsto e(x)$ such that $E\, e(X)h(X) = 0$ for all $h \in \text{span}\,(\mathcal{H})$.

On differential information games, §1.D

DfG 1. [Ignorance is bliss] If $\theta = \theta_0$, the payoffs to players 1 and 2 to are given by the left matrix (below), if $\theta = \theta_1$, they are given by the right matrix, and $P(\theta = \theta_0) = P(\theta = \theta_1) = \frac{1}{2}$.

θ_0	L	R
A	$(8,0)$	$(4,2)$
B	$(6,4)$	$(2,0)$

θ_1	L	R
A	$(6,4)$	$(2,0)$
B	$(10,0)$	$(6,2)$

a. Suppose that neither player observes the value of θ, e.g., suppose that they observe constant signals. Give the stable equilibrium outcome.

b. Suppose that player 1 observes the value of θ before making any choice, but that player 2 observes only a constant signal. Give the stable equilibrium outcome.
c. Suppose now that player 1 can choose whether or not to observe the value of θ before making any choice, but that player 2 observes only a constant signal. Give the stable equilibrium outcome.

2. Convergence

The use of bi-linear functionals in the following will be crucial in defining convergence, to answer the questions raised in dynamic programming. It will later be crucial in defining cones and their associated partial orders, to answer the questions raised in expected utility maximization.

If $\mathfrak{X}$ is a real vector space (think $\mathbb{R}^\ell$ or $C([0,1])$ as we haven't formally defined this yet), and $\|\cdot\|$ is a norm on $\mathfrak{X}$, then $\mathfrak{X}^\dagger$ denotes the set of continuous linear functions $\boldsymbol{x} \mapsto \boldsymbol{x}^\dagger(\boldsymbol{x})$. We will, throughout, use the notation $\langle \boldsymbol{x}, \boldsymbol{x}^\dagger \rangle$ for $\boldsymbol{x}^\dagger(\boldsymbol{x})$. The mapping $(\boldsymbol{x}, \boldsymbol{x}^\dagger) \mapsto \langle \boldsymbol{x}, \boldsymbol{x}^\dagger \rangle$ is *bi-linear*, that is, for all $\boldsymbol{x}, \boldsymbol{x} \in \mathfrak{X}$, all $\boldsymbol{x}^\dagger, \boldsymbol{y}^\dagger \in \mathfrak{X}^\dagger$, and all $\alpha, \beta \in \mathbb{R}$, we have

$$\langle \alpha\boldsymbol{x} + \beta\boldsymbol{y}, \boldsymbol{x}^\dagger \rangle = \alpha\langle \boldsymbol{x}, \boldsymbol{x}^\dagger \rangle + \beta\langle \boldsymbol{y}, \boldsymbol{x}^\dagger \rangle, \text{ and} \tag{10}$$

$$\langle \boldsymbol{x}, \alpha\boldsymbol{x}^\dagger + \beta\boldsymbol{y}^\dagger \rangle = \alpha\langle \boldsymbol{x}, \boldsymbol{x}^\dagger \rangle + \beta\langle \boldsymbol{x}, \boldsymbol{y}^\dagger \rangle. \tag{11}$$

The first holds because $\boldsymbol{x}^\dagger$ is linear, the second holds because that is how we define scalar multiplication of and addition of functions.

2.A. Norm Convergence in $\mathbb{R}^\ell$.

Here $\boldsymbol{v}_k \in \mathfrak{X} = C(S) = M(S) = \mathbb{R}^S$ (continuous and measurable functions on S). We give $\mathfrak{X} = C(S)$ the norm $\|\boldsymbol{x}\| = \max_{i \in S} |\boldsymbol{x}(i)|$.

The norm distance between $\boldsymbol{x}$ and $\boldsymbol{y}$ is $d(\boldsymbol{x}, \boldsymbol{y}) := \|\boldsymbol{x} - \boldsymbol{y}\|$. The triangle inequality for norms is $\|\boldsymbol{x}\| + \|\boldsymbol{y}\| \geq \|\boldsymbol{x} + \boldsymbol{y}\|$, letting $\boldsymbol{x} = (\boldsymbol{r} - \boldsymbol{s})$ and $\boldsymbol{y} = (\boldsymbol{s} - \boldsymbol{t})$, we have $d(\boldsymbol{r}, \boldsymbol{t}) \leq d(\boldsymbol{r}, \boldsymbol{s}) + d(\boldsymbol{s}, \boldsymbol{t})$.

Definition 2.A.1. ***Strong convergence*** *or* ***norm convergence*** *is* $\boldsymbol{v}_k \to \boldsymbol{v}^*$ *iff* $\|\boldsymbol{v}_k - \boldsymbol{v}^*\| \to 0$.

> In the finite dimensional case, strong convergence will be, essentially, all that we will study. This is because there is a unique topology on $\mathbb{R}^\ell$ such that $(\boldsymbol{x}, \boldsymbol{y}) \mapsto (\boldsymbol{x} + \boldsymbol{y})$ and $(r, \boldsymbol{x}) \mapsto r\boldsymbol{x}$ are jointly continuous. We will develop the alternative approaches here because everything is simpler to visualize, but the equivalence will not generalize to infinite dimensional cases.

Part of the "essentially" in the above comes from things called **pseudo-metrics**.

Definition 2.A.2. *A* ***pseudo-metric*** *on a set* M *is a function* $\rho : M \times M \to \mathbb{R}_+$ *such that for all* $x, y, z \in M$, $\rho(x, y) = \rho(y, x)$ *and* $\rho(x, y) + \rho(y, z) \geq \rho(x, z)$. *If a pseudo-metric also satisfies* $\rho(x, y) = 0$ *iff* $x = y$, *then it is a* ***metric***

Example 2.A.1. *A simple pseudo-metric is* $\rho(\boldsymbol{x}, \boldsymbol{y}) = |\boldsymbol{x}(1) - \boldsymbol{y}(1)|$, *which measures only the distance between the first coordinates. Note that* $\boldsymbol{x} = (7, 9, -23)'$ *and* $\boldsymbol{y} = (7, -506, 10^{10})'$ *are very different, but that* $\rho(\boldsymbol{x}, \boldsymbol{y}) = 0$.

2.B. **The Dual Space.** Notation: $\mathfrak{X} = \mathbb{R}^\ell$; $\mathfrak{X}^\dagger$ the **continuous dual** of $\mathfrak{X}$, that is, the set of all norm-continuous linear functions $\boldsymbol{x}^\dagger : \mathfrak{X} \to \mathbb{R}$; the bi-linear $(\boldsymbol{x}, \boldsymbol{y}) \mapsto \langle \boldsymbol{x}, \boldsymbol{y} \rangle$ given by $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = \sum_{i \in S} \boldsymbol{x}(i)\boldsymbol{y}(i)$ (here the usual inner product). From a long time ago, we know the following, which says that $\mathfrak{X}$ and $\mathfrak{X}^\dagger$ are the same space. To say this in a fancy way, $\mathfrak{X}$ is **reflexive**, that is, equal to its own dual space. We call results of this form "representation theorems" because they tell us how to represent the continuous dual space.

Theorem 2.1 (Representation)**.** *If $\mathfrak{X} = \mathbb{R}^S$ and S is finite, then for any $\boldsymbol{y} \in \mathfrak{X}$, the function $\boldsymbol{x} \mapsto \langle \boldsymbol{x}, \boldsymbol{y} \rangle$ is continuous and linear, and if $\boldsymbol{x}^\dagger \in \mathfrak{X}^\dagger$, then there exists a unique $\boldsymbol{y} \in \mathbb{R}^S$ such that for all $\boldsymbol{x} \in \mathfrak{X}$, $\boldsymbol{x}^\dagger(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{y} \rangle$.*

Proof. Fill in. □

In this finite dimensional case, linearity guarantees continuity. When this is not true, we will need to distinguish between the **algebraic dual** of $\mathfrak{X}$, the set of all linear functions, continuous or not, and the continuous dual of $\mathfrak{X}$.

The norm $\mathfrak{X}^\dagger$ is $\|\boldsymbol{x}^\dagger\| := \max\{|\langle \boldsymbol{x}, \boldsymbol{x}^\dagger \rangle| : \|\boldsymbol{x}\| \leq 1\}$. You should be able to put things together to prove the following.

Lemma 2.2. *If $\boldsymbol{x}^\dagger \in \mathfrak{X}^\dagger$ is represented by $\boldsymbol{x}^\dagger(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{y} \rangle$, then $\|\boldsymbol{x}^\dagger\| = \|\boldsymbol{y}\|$.*

Strong (or norm) convergence in the dual space $\mathfrak{X}^\dagger$ is convergence of the norm difference to 0.

Notation: $\mathfrak{X}^{\dagger\dagger}$ is the dual space of $\mathfrak{X}^\dagger$. Since $\mathfrak{X}$ is reflexive, $\mathfrak{X} = \mathfrak{X}^{\dagger\dagger}$. This is another thing that will not be generally true when we get away from functions on a finite domain.

2.C. **Semi-Norms for Convergence.** For $\mathcal{S}$ a class of subsets of $\mathfrak{X}^\dagger$, we will talk about $\mathcal{S}$-convergence: $\boldsymbol{v}_k \to_{\mathcal{S}} \boldsymbol{v}^*$ if for all $F \in \mathcal{S}$,

$$\sup_{\boldsymbol{x}^\dagger \in F} |\langle \boldsymbol{v}_k, \boldsymbol{x}^\dagger \rangle - \langle \boldsymbol{v}^*, \boldsymbol{x}^\dagger \rangle| \to 0. \tag{12}$$

The functions $\boldsymbol{x} \mapsto \sup_{\boldsymbol{x}^\dagger \in F} |\langle \boldsymbol{x}, \boldsymbol{x}^\dagger \rangle|$ in (12) are **semi-norms**.

We have been using the norm $\|\boldsymbol{x}\| = \max_{i \in \{1,\ldots,\ell\}} |x_i|$. There will be many other norms depending on context. Here is the general definition.

Definition 2.C.1. *A function $p : \mathfrak{X} \to \mathbb{R}_+$ is a **semi-norm** if $(\forall \lambda \in \mathbb{R})(\forall \boldsymbol{x} \in \mathfrak{X})[p(\lambda \boldsymbol{x}) = |\lambda| \boldsymbol{x}$ and $(\forall \boldsymbol{x}, \boldsymbol{y} \in \mathfrak{X})[p(\boldsymbol{x} + \boldsymbol{y}) \leq p(\boldsymbol{x}) + p(\boldsymbol{y})]$. If a semi-norm also satisfies $p(\boldsymbol{x}) = 0$ iff $\boldsymbol{x} = 0$, then it is a **norm**.*

Note that taking $\lambda = -1$ yields $p(-\boldsymbol{x}) = p(\boldsymbol{x})$. This will reappear later as $A = \{\boldsymbol{x} : p(\boldsymbol{x}) < r\}$ being a *balanced* set, i.e. $\boldsymbol{x} \in A$ iff $-\boldsymbol{x} \in A$. You should check that $\|\boldsymbol{x}\| := \max_{i \in \{1,\ldots,\ell\}} |x_i|$ really is a norm. This is also a consequence of the following

Lemma 2.3. *For non-empty, norm bounded $F \subset \mathfrak{X}^\dagger$, $p_F(\boldsymbol{x}) := \sup_{\boldsymbol{x}^\dagger \in F} |\langle \boldsymbol{x}, \boldsymbol{x}^\dagger \rangle|$ is a semi-norm, and it is a norm if F spans $\mathfrak{X}^\dagger$.*

Proof. Fill in. □

From Micro I and/or II, we know that the supremum of a collection of linear functionals is a convex function. This means that each of the $p_F(\cdot)$ is convex. This is true of all semi-norms

Lemma 2.4. *If $p : \mathfrak{X} \to \mathbb{R}_+$ is a semi-norm, then for all $\boldsymbol{x}, \boldsymbol{y} \in \mathfrak{X}$ and all $\alpha \in (0, 1)$, $p(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \leq \alpha p(\boldsymbol{x}) + (1-\alpha)p(\boldsymbol{y})$.*

Proof. From the second property of semi-norms, $p(\alpha\boldsymbol{x}+(1-\alpha)\boldsymbol{y}) \leq p(\alpha\boldsymbol{x})+p((1-\alpha)\boldsymbol{y})$. From the first property, $p(\alpha\boldsymbol{x}) = |\alpha|p(\boldsymbol{x}) = \alpha p(\boldsymbol{x})$ and $p((1-\alpha)\boldsymbol{y})|(1-\alpha)|p(\boldsymbol{y}) = (1-\alpha)p(\boldsymbol{y})$. □

Of particular later use to us is the observation that for $r > 0$, the set $\{\boldsymbol{x} : p(\boldsymbol{x}) < r\}$ is balanced, convex, and contains the origin, 0.

2.D. **Uniform Convergence on Subsets of Dual Spaces.** We are about to define convergence of $\boldsymbol{v}_n$ to $\boldsymbol{v}^*$ in terms of convergence of a collection of semi-norms, $p_F(\boldsymbol{v}_n - \boldsymbol{v}^*)$, to 0. Looking ahead, we will define topologies induced by collections of semi-norms in the same way. Because semi-norms are convex, the sets $\{\boldsymbol{x} : p(\boldsymbol{x}) < \lambda\}$ are convex open subsets of the origin. The general term that will describe this situation is **local convexity** of the topology.

Returning to our convergence theme, if $\mathcal{S}$ is a collection of subsets of $\mathfrak{X}^\dagger$, then $\boldsymbol{v}_k \to_{\mathcal{S}} \boldsymbol{v}^*$ iff $p_F(\boldsymbol{v}_k - \boldsymbol{v}^*) \to 0$ for all $F \in \mathcal{S}$.

Keep in mind the following simple switch of perspective: Since $\mathfrak{X} = \mathfrak{X}^{\dagger\dagger}$, we are also using subsets of $\mathfrak{X}$ to define convergence in $\mathfrak{X}^\dagger$.

Example 2.D.1. *Let $U = \{\boldsymbol{x}^\dagger \in \mathfrak{X}^\dagger : \|\boldsymbol{x}^\dagger\| \leq 1\}$ be the norm-closed unit ball in $\mathfrak{X}^\dagger$ and let $\mathcal{S} = \{U\}$. From the results above, $\mathcal{S}$-convergence is norm convergence.*

Lemma 2.5. *If $\mathcal{S} = \{F \subset \mathfrak{X}^\dagger : F$ is norm bounded$\}$, then $\mathcal{S}$-convergence is equivalent to norm convergence.*

Proof. Fill in. □

Example 2.D.2. *If $\mathcal{S} = \{\{\boldsymbol{x}^\dagger\} : \boldsymbol{x}^\dagger \in \mathfrak{X}^\dagger\}$ is the class of singleton sets, then $\boldsymbol{v}_k \to_{\mathcal{S}} \boldsymbol{v}^*$ iff $(\forall \boldsymbol{x}^\dagger \in \mathfrak{X}^\dagger)[\langle \boldsymbol{v}_k, \boldsymbol{x}^\dagger\rangle \to \langle \boldsymbol{v}^*, \boldsymbol{x}^\dagger\rangle]$.*

The reason to look at the following in some detail is because we use a special class of $\boldsymbol{x}^\dagger$, one with a rich span.

Lemma 2.6. $(\forall \boldsymbol{x}^\dagger \in \mathfrak{X}^\dagger)[\langle \boldsymbol{v}_k, \boldsymbol{x}^\dagger\rangle \to \langle \boldsymbol{v}^*, \boldsymbol{x}^\dagger\rangle]$ *iff* $(\forall i \in S)[\boldsymbol{v}_k(i) \to \boldsymbol{v}^*(i)]$.

Proof. Taking $\boldsymbol{x}^\dagger = \boldsymbol{e}_i$ (the unit vector in the i'th direction) and noting that $\boldsymbol{v}(i) = \langle \boldsymbol{v}, \boldsymbol{e}_i\rangle$, we have $(\forall \boldsymbol{x}^\dagger \in \mathfrak{X}^\dagger)[\langle \boldsymbol{v}_k, \boldsymbol{x}^\dagger\rangle \to \langle \boldsymbol{v}^*, \boldsymbol{x}^\dagger\rangle] \Rightarrow (\forall i \in S)[\boldsymbol{v}_k(i) \to \boldsymbol{v}^*(i)]$.

To go in the opposite direction, note that any $\boldsymbol{x}^\dagger$ is of the form $\sum_{i\in S} \alpha_i \boldsymbol{e}_i$, $\alpha_i \in \mathbb{R}$, so that $\langle \boldsymbol{v}, \boldsymbol{x}^\dagger\rangle = \sum_{i\in S} \alpha_i \langle \boldsymbol{v}, \boldsymbol{e}_i\rangle$ so that $(\forall i \in S)[\boldsymbol{v}_k(i) \to \boldsymbol{v}^*(i)] \Rightarrow (\forall \boldsymbol{x}^\dagger \in \mathfrak{X}^\dagger)[\langle \boldsymbol{v}_k, \boldsymbol{x}^\dagger\rangle \to \langle \boldsymbol{v}^*, \boldsymbol{x}^\dagger\rangle]$. □

Lemma 2.7. *If $\mathcal{S}$ is the class of finite subsets of $\mathfrak{X}^\dagger$, then $\boldsymbol{v}_k \to_{\mathcal{S}} \boldsymbol{v}^*$ iff $(\forall i \in S)[\boldsymbol{v}_k(i) \to \boldsymbol{v}^*(i)]$.*

Crazy big classes of sets $\mathcal{S}$ can make convergence into a silly notion.

Lemma 2.8. *If $\mathcal{S}$ is the class of all subsets of $\mathfrak{X}^\dagger$ than $\boldsymbol{v}_k \to_{\mathcal{S}} \boldsymbol{v}^*$ iff $(\exists K)(\forall k \geq K)[\boldsymbol{v}_k = \boldsymbol{v}^*]$.*

Proof. Fill in. □

It is for this reason that we always restrict attention to classes $\mathcal{S}$ that contain only norm bounded sets. Here is an example of $\mathcal{S}$ with norm unboundedness only in one direction. Things are still bad.

Example 2.D.3. *If $\mathfrak{X} = \mathbb{R}^3$ and $\mathcal{S} = \{(1,0,0)', (0,1,0)', \{(0,0,r) : r \in \mathbb{R}\}\}$, then $\boldsymbol{v}_k \to_{\mathcal{S}} \boldsymbol{v}^*$ iff $\boldsymbol{v}_k(1) \to \boldsymbol{v}^*(1)$, $\boldsymbol{v}_k(2) \to \boldsymbol{v}^*(2)$, and $(\exists K)(\forall k \geq K)[\boldsymbol{v}_k(3) = \boldsymbol{v}^*(3)]$.*

Here is the result that all of these convergences except the last two are the same. We say that $\mathcal{S}$ spans if the span of the $\boldsymbol{x}^\dagger \in E$ for some $E \in \mathcal{S}$ is norm dense in $\mathfrak{X}^\dagger$.

Theorem 2.9. *$\|\boldsymbol{v}_k - \boldsymbol{v}^*\| \to 0$ iff for all spanning $\mathcal{S}$ consisting of norm bounded sets, $\boldsymbol{v}_k \to_{\mathcal{S}} \boldsymbol{v}^*$.*

Side comment: if $\mathcal{S}$ is **not** spanning, we end up with pseudo-metrics. For example, taking $\mathcal{S}$ to contain only the set $\{(1,0,0)\}$, we get the pseudo-metric of Example 2.A.1.

Proof. ($\Leftarrow$) is really easy because $\{U\}$ is a norm bounded and spanning set. For the other direction, suppose that $\|\boldsymbol{v}_k - \boldsymbol{v}^*\| \to 0$, i.e. $\sup\{|\langle \boldsymbol{x}, \boldsymbol{v}_k - \boldsymbol{v}^*\rangle| : \boldsymbol{x} \in U\} \to 0$. If F is a norm bounded set, then for some $\kappa > 0$, $F \subset \kappa \cdot U$.

$$\sup\{|\langle \boldsymbol{x}, \boldsymbol{v}_k - \boldsymbol{v}^*\rangle| : \boldsymbol{x} \in F\} \leq \kappa \sup\{|\langle \boldsymbol{x}, \boldsymbol{v}_k - \boldsymbol{v}^*\rangle| : \boldsymbol{x} \in U\} \to 0. \quad \square \tag{13}$$

2.E. **Finite Versus Infinite Dimensionality.** We will come back and prove many of the assertions given in this subsection. For now, we just want a taste of how things are different.

Let $\mathfrak{X} = C([0,1])$ with the norm $\|\boldsymbol{x}\| = \max_{t \in [0,1]} |\boldsymbol{x}(t)|$.

Theorem 2.10 (Representation)**.** *If $\mu, \nu \in \Delta([0,1])$ with $\mu(E) = 1$ and $\nu(E) = 0$ for some measurable $E \subset [0,1]$ and $\alpha, \beta \geq 0$, then the functional*

$$\boldsymbol{x} \mapsto \alpha \int_{[0,1]} \boldsymbol{x}(t)\, d\mu(t) - \beta \int_{[0,1]} \boldsymbol{x}(t)\, d\nu(t) \tag{14}$$

belongs to $\mathfrak{X}^\dagger$. For any $\boldsymbol{x}^\dagger \in \mathfrak{X}^\dagger$, there exists a pair $\mu, \nu \in \Delta([0,1])$ and a pair $\alpha, \beta \geq 0$ such that for all $\boldsymbol{x} \in \mathfrak{X}$,

$$\boldsymbol{x}^\dagger(\boldsymbol{x}) = \alpha \int_{[0,1]} \boldsymbol{x}(t)\, d\mu(t) - \beta \int_{[0,1]} \boldsymbol{x}(t)\, d\nu(t). \tag{15}$$

Further, the signed measure $\eta_{\boldsymbol{x}^\dagger}(A) := \alpha\mu(A) - \beta\nu(A)$ is uniquely determined by $\boldsymbol{x}^\dagger$, i.e. if $\boldsymbol{x}^\dagger \neq \boldsymbol{y}^\dagger$ then $\eta_{\boldsymbol{x}^\dagger}(A) \neq \eta_{\boldsymbol{y}^\dagger}(A)$ for some measurable A.

To put this another way, the dual space for $C([0,1])$ is the set of countably additive finite signed measures on $[0,1]$, denoted $ca([0,1])$. We again use the bracket notation, $\langle \boldsymbol{x}, \boldsymbol{x}^\dagger\rangle = \int \boldsymbol{x}(t)\, d\eta(t)$ where η is the signed measure associated with $\boldsymbol{x}^\dagger$. Notation: η (rather than $\boldsymbol{x}^\dagger$) is the appropriate notation for elements of $\mathfrak{X}^\dagger$. The bracket function, $(\boldsymbol{x}, \eta) \mapsto \langle \boldsymbol{x}, \eta\rangle$ is, once again, bi-linear. We will very often be interested in $\Delta([0,1])$, which is a spanning, strict subset of $C([0,1])^\dagger$.

Example 2.E.1 (Pointwise convergence)**.** *Let $\mathcal{S}$ be the class of singleton subsets of $\mathfrak{X}^\dagger$ consisting of point masses (on points in $[0,1]$), $\mathcal{S} = \{\{\delta_r\} : r \in [0,1]\}$. Then $\boldsymbol{x}_k \to_{\mathcal{S}} \boldsymbol{x}$ iff for all $r \in [0,1]$, $\boldsymbol{x}_k(r) \to \boldsymbol{x}(r)$. This is called, not surprisingly, **pointwise convergence** in $C([0,1])$.*

The following is a special case of a result true for all normed spaces — if we use the unit ball in $\mathfrak{X}^\dagger$ to define a norm on $\mathfrak{X}$, we arrive back at the original norm.

Lemma 2.11. *For $\boldsymbol{x} \in C([0,1])$, $\sup\{|\langle \boldsymbol{x}, \eta\rangle| : \|\eta\| \leq 1\} = \max_{t \in [0,1]} |\boldsymbol{x}(t)|$.*

Consider the norm distance on $\mathfrak{X}^\dagger$: $\|\eta\| = \sup\{|\langle \boldsymbol{x}, \eta\rangle : \|\boldsymbol{x}\| \leq 1\}$. One of the key approximation results in measure theory is that for probabilities $\mu, \nu \in \mathfrak{X}^\dagger$, the norm or strong distance, $\|\mu - \nu\|$ is equal to $2 \cdot \max_{E \subset [0,1]} |\mu(E) - \nu(E)|$ where the maximum is over measurable sets.

Reflexivity fails.[2]

Example 2.E.2. $C([0,1]) \subsetneq ca([0,1])^\dagger = C([0,1])^{\dagger\dagger}$. *To see why, first note that* $\eta \mapsto \int f\, d\eta$ *is continuous if* f *is continuous, but that the functionals* $\eta \mapsto \int 1_E\, d\eta$ *are linear and norm continuous.*

Now we go to that simple switch of perspective: we use subsets of $\mathfrak{X} \subsetneq \mathfrak{X}^{\dagger\dagger}$ to define convergence in $\mathfrak{X}^\dagger$ and in $\Delta([0,1]) \subset \mathfrak{X}^\dagger$.

Example 2.E.3. *Let* $\mathcal{S} = \{\{f\} : f \in C([0,1])\}$ *so that* $\mu_k \to_{\mathcal{S}} \mu$ *iff for all continuous* $f : [0,1] \to \mathbb{R}$, $\int f\, d\mu_k \to \int f\, d\mu$. *If* μ_k *is the uniform distribution on* $\{1/k, 2/k, \ldots, (k-1)/k, 1\} \subset [0,1]$, *then* $\mu_k \to_{\mathcal{S}} \lambda$ *but* $\|\mu_k - \lambda\| \equiv 2$.

The following is a version of a more general result called the Banach-Steinhaus theorem: pointwise convergence of linear functions to a linear function on a compact set implies uniform convergence. Without linearity, you should know many examples of sequences of function that are pointwise convergent to a continuous function that are not uniformly convergent.

Theorem 2.12. *A sequence* μ_k *in* $\Delta([0,1])$ weak* *converges to* μ *iff for all compact, spanning* $\mathcal{S}$, $\mu_k \to_{\mathcal{S}} \mu$.

Compact subsets of $C([0,1])$ are smaller than you might believe. The following should help dis-abuse you of the notion that "closed and bounded means compact." That only works in the finite dimensional case.

Lemma 2.13. *If* $E \subset C([0,1])$ *is compact then it is bounded, has empty interior, and is approximately flat, that is, for all* $\epsilon > 0$, *there exists a finite dimensional* $V \subset C([0,1])$ *such that* $E \subset V^\epsilon$.

You should know how to prove that if $V \subset \mathbb{R}^\ell$ is a dense vector subspace, then $V = \mathbb{R}^\ell$. The situation is very different in $C([0,1])$.

Theorem 2.14 (Weierstrass)**.** *The set of polynomials is a dense vector subspace of* $C([0,1])$.

More generally, for the span of any compact $K \subset C([0,1])$ is closed iff the span is finite dimensional.

Detour 2.1. *What is interesting and tremendously useful for econometrics is the existence of compact sets having a span that is dense in an infinite dimensional space. For example, consider the span of the closure of the set* $\{k_n t^n : n = 0, 1, \ldots\}$ *where* $k_n \downarrow 0$. *The approximately flat result just given shows that any compact set with dense span must poke arbitrarily small amounts into some dimensions. It is out of such considerations that one finds rates of convergence results for different kinds of approximation schemes.*

---

[2] If you would like to know more about the structure of the double dual of $C(K)$, K compact, Karlin has a series of papers on its structure. There are also characterizations using nonstandard analysis.

2.F. **Problems.**

On norm convergence and $\mathcal{S}$-convergence in $\mathbb{R}^\ell$

NC 1. Complete the proof of Theorem 2.10.

NC 2. Give a complete proof that if $\mathcal{S}$ is a spanning set of norm bounded subsets of $\mathbb{R}^\ell$, then $\boldsymbol{v}_k \to_{\mathcal{S}} \boldsymbol{v}^*$ if and only if $\|\boldsymbol{v}_k - \boldsymbol{v}^*\| \to 0$.

NC 3. For $\{\boldsymbol{x}_n^\dagger : n \in \mathbb{N}\}$ a countable, norm dense subset of $\mathbb{R}^\ell$, define $\rho(\boldsymbol{x}, \boldsymbol{y}) = \sum_{n\in\mathbb{N}} \frac{1}{2^n} \frac{|\langle \boldsymbol{x}-\boldsymbol{y}, \boldsymbol{x}_n^\dagger\rangle|}{1+|\langle \boldsymbol{x}-\boldsymbol{y}, \boldsymbol{x}_n^\dagger\rangle|}$. Show that $\rho(\cdot,\cdot)$ is a metric and that $\rho(\boldsymbol{v}_k, \boldsymbol{v}^*) \to 0$ iff $\|\boldsymbol{v}_k - \boldsymbol{v}^*\| \to 0$.

NC 4. The utility distance between bundles $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^\ell_+$ is given by $|u(\boldsymbol{x}) - u(\boldsymbol{x}')|$. For what functions $u(\cdot)$ is this expressible as $\mathcal{S}$-convergence? For what classes of sets $\mathcal{S}$?

On convergence in $C([0,1])$ and $ca([0,1])$

Cts 1. Suppose that $\mathcal{S} = \{\{f\} : f \in C([0,1])$ so that $\mu_k \to_{\mathcal{S}} \mu$ in $\Delta([0,1])$ iff $\int f\, d\mu_k \to \int f\, d\mu$ for all $f \in C([0,1])$.
  a. For $r \in [0,1]$, δ_r denotes point mass on r, that is, $\delta_r(A) = 1_A(r)$. Show that $\delta_{r_n} \to_{\mathcal{S}} \delta_r$ iff $|r_n - r| \to 0$ in $[0,1]$.
  b. Show that $\mu_k \to_{\mathcal{S}} \mu$ in $\Delta([0,1])$ iff all of the moments of μ_k converge to the moments of μ.

Cts 2. We now turn to $ca([0,1])$ keeping the same class $\mathcal{S}$ from the previous problem.
  a. Let $\eta_n = \delta_{0.5+1/n} - \delta_{0.5-1/n}$. Show that $\eta_n \to_{\mathcal{S}} 0$ but that $\|\eta_n\| \equiv 2$.
  b. Let $\mathcal{P} = \{f_n : n = 0, 1, 2, \ldots\}$ where $f_n(t) = t^n$ so that the span of $\mathcal{P}$ is the set of polynomials. Define $\rho(\eta, \eta') = \sum_{n\in\mathbb{N}} \frac{1}{2^n} \frac{|\langle \eta-\eta', f_n\rangle|}{1+|\langle \boldsymbol{x}-\boldsymbol{y}, \boldsymbol{x}_n^\dagger\rangle|}$. Show that $\rho(\mu_k, \mu) \to_{\mathcal{S}} 0$ in $\Delta([0,1])$ iff $\rho(\mu_k, \mu) \to 0$, but that the norm diameter of the ρ unit ball in $ca([0,1])$ is infinite.

3. Convex Cones, Partial Orders, and Pre-Orders

We will return now to $\mathfrak{X} = \mathbb{R}^S$, $S = \{1, 2, \ldots, \ell\}$, though the abstract definitions will be what we work with in the more general cases. We will think of $\Delta(S)$ as subset of $\mathfrak{X}^\dagger$.

3.A. **Cones.** The following allows more than things shaped like dunce caps in $\mathbb{R}^3$.

Definition 3.A.1. *$C \subset \mathfrak{X}$ is a **convex cone** if $(\forall \boldsymbol{x}, \boldsymbol{y} \in C)(\forall \alpha, \beta > 0)[\alpha\boldsymbol{x} + \beta\boldsymbol{y} \in C]$. A convex cone C is **pointed** if $0 \in C$, otherwise C is **blunt**. A convex cone C is **flat** if for some non-zero $\boldsymbol{x} \in \mathfrak{X}$, $\boldsymbol{x} \in C$ and $-\boldsymbol{x} \in C$, otherwise C is **salient**. A convex cone is **proper** if it is not contained in any strict vector subspace of $\mathfrak{X}$.*

In terms of set addition and scalar multiplication, this is $\alpha C + \beta C = C$ for $\alpha, \beta > 0$. Note that the intersection of any collection of convex cones is another convex cone and that $\mathfrak{X}$ is itself a convex cone. Putting these together in the usual fashion, for any $E \subset \mathfrak{X}$, there is a smallest convex cone containing E.

Lemma 3.1. *Every blunt convex cone is salient, and a cone is salient iff $C \cap -C \subset \{0\}$.*

Another way to say that $C \cap -C \subset \{0\}$ is "C contains no non-trivial linear subspace."

Example 3.A.1. $\{0\}$ *is a closed, pointed, proper, salient cone in* $\mathbb{R}^\ell$*; a non-trivial strict vector subspace of* $\mathbb{R}^\ell$ *is closed, pointed, and flat cone which is not proper; the closed half-spaces,* $\{\boldsymbol{x} : \langle \boldsymbol{x}, \boldsymbol{y}\rangle \geq 0\}$ *for some* $\boldsymbol{y} \neq 0$*, are flat, pointed, proper cones;* $\mathbb{R}^S_+$ *is a closed, pointed, proper salient cone;* $\mathbb{R}^S_{++}$ *is an open, blunt, salient cone;* $C = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 > 0 \vee [x_1 = 0 \wedge x_2 \geq 0]\}$ *is a proper, salient pointed cone that is neither closed nor open.*

3.B. **Partial Orders and Pre-orders.** Recall that a **partial order** on $\mathfrak{X}$ is one satisfying

(1) $\boldsymbol{x} \succsim \boldsymbol{x}$,
(2) $[\,[\boldsymbol{x} \succsim \boldsymbol{y}] \wedge [\boldsymbol{y} \succsim \boldsymbol{z}]\,] \Rightarrow [\boldsymbol{x} \succsim \boldsymbol{y}]$, and
(3) $[\,[\boldsymbol{x} \succsim \boldsymbol{y}] \wedge [\boldsymbol{y} \succsim \boldsymbol{y}]\,] \Rightarrow [\boldsymbol{x} = \boldsymbol{y}]$,

and that a **pre-order** satisfies only the first two properties.

Definition 3.B.1. *Associated with a convex cone,* C*, is the pre-order* $\succsim_C$ *defined by* $\boldsymbol{x} \succsim_C \boldsymbol{y}$ *iff* $\boldsymbol{x} - \boldsymbol{y} \in C$*.*

Lemma 3.2. *If* C *is a pointed, salient, convex cone, then* $\succsim_C$ *is a partial order. If* C *is flat, then* $\succsim_C$ *is a pre-order.*

More examples:

(1) If $C = \mathbb{R}^S_+$, then $\boldsymbol{x} \succsim_C \boldsymbol{y}$ iff $\boldsymbol{x} \geq \boldsymbol{y}$.
(2) If C is also flat, then the order can be defined, but it fails $[\,[\boldsymbol{x} \succsim \boldsymbol{y}] \wedge [\boldsymbol{y} \succsim \boldsymbol{y}]\,] \Rightarrow [\boldsymbol{x} = \boldsymbol{y}]$.
(3) If $C = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 > 0 \vee [x_1 = 0 \wedge x_2 \geq 0]\}$, then $\boldsymbol{x} \succsim_C \boldsymbol{y}$ iff $\boldsymbol{x}$ is weakly larger than $\boldsymbol{y}$ in the lexicographic order.
(4) If $M : \mathbb{R}^\ell \to \mathbb{R}^\ell$ is an invertible linear mapping, then $C = M(\mathbb{R}^\ell_+)$ is a closed, pointed, proper salient cone giving the obvious order.

Example 3.B.1. *The set of non-decreasing functions,* $C = \{\boldsymbol{x} \in \mathfrak{X} : \boldsymbol{x}(1) \leq \boldsymbol{x}(2) \leq \cdots \leq \boldsymbol{x}(\ell)\}$*, is a closed, pointed, convex cone that is flat (it contains* $(r, r, \cdots, r)$ *for any* $r \in \mathbb{R}$*).*

3.C. **Dual Cones.** There is what looks like another way to use cones to give pre-orders and partial orders. It uses the idea of a dual cone. First an example, then the definitions and general observations.

Example 3.C.1. *Let* $C \subset \mathfrak{X}$ *be the closed convex cone of non-decreasing functions and consider the following pre-order on* $\Delta(S) \subset \mathfrak{X}^\dagger$*:* $\mu \succsim \nu$ *iff for all* $\boldsymbol{x} \in C$*,* $\langle \boldsymbol{x}, \mu - \nu\rangle \geq 0$*. Since* $\langle \boldsymbol{x}, \mu - \nu\rangle \geq 0$ *iff* $\langle \boldsymbol{x}, \mu\rangle \geq \langle \boldsymbol{x}, \nu\rangle$*, this is the usual* ***first order stochastic dominance (FOSD)*** *order,* $\mu \succsim_{FOSD} \nu$*.*

Define $C^\dagger = \{\boldsymbol{x}^\dagger : (\forall \boldsymbol{x} \in C)[\langle \boldsymbol{x}, \boldsymbol{x}^\dagger\rangle \geq 0]\}$*. This is a closed convex cone, and we are writing* $\mu \succsim_{FOSD} \nu$ *iff* $(\mu - \nu) \in C^\dagger$*.*

These are the two patterns: one can start with a cone $C \subset \mathfrak{X}$ and define a pre-order by $(\boldsymbol{x} - \boldsymbol{y}) \in C$; or one can start with a cone $C \in \mathfrak{X}^\dagger$, define $C^\dagger$ as in the example, and define a pre-order by $(\boldsymbol{x}^\dagger - \boldsymbol{y}^\dagger) \in C^\dagger$.

Here is the formal name for what we did, a **dual cone**.

Definition 3.C.1. *If* $C \subset \mathfrak{X}$ *is a cone, then its* ***dual cone*** *is* $C^\dagger := \{\boldsymbol{x}^\dagger : (\forall \boldsymbol{x} \in C)[\langle \boldsymbol{x}, \boldsymbol{x}^\dagger\rangle \geq 0]\}$*.*

Detour 3.1. *We didn't really need to start with C being a cone — for any $E \subset \mathfrak{X}$, $E^\dagger$ is equal to $C^\dagger$ where C is the smallest cone containing E. From this kind of thinking we have the following two results: $(C_1 \cup C_2)^\dagger = C_1^\dagger \cap C_2^\dagger$; and the smallest cone containing both $C_1^\dagger$ and $C_2^\dagger$ is $(C_1 \cap C_2)^\dagger$.*

Dual cones are the only kind of closed convex cones.

Lemma 3.3. *If C is a closed convex cone, then $C = (C^\dagger)^\dagger$.*

Proof. Exercise in keeping track of things. □

From the definition, the bigger is C, the smaller the dual cone — there are more constraints, $\langle \boldsymbol{x}, \boldsymbol{x}^\dagger \rangle \geq 0$, to be satisfied by elements of the dual cone.

Example 3.C.2. *If $C = \mathfrak{X}$, then $C^\dagger = \{0\}$ (we saw this at work in the example of $\mathcal{S}$-convergence with $\mathcal{S}$ being too large). If $C = \{0\}$, then $C^\dagger = \mathfrak{X}^\dagger$.*

If $C = \{(x_1, x_2)' \in \mathbb{R}^2 : x_1 + x_2 \geq 0\}$ then $C^\dagger = \{(r, r)' \in \mathbb{R}^2 : r \geq 0\}$. If $C = \{r \cdot \boldsymbol{x} : r \geq 0\}$, $\boldsymbol{x} \neq 0$, then $C^\dagger = \{\boldsymbol{x}^\dagger \in \mathfrak{X}^\dagger : \langle \boldsymbol{x}, \boldsymbol{x}^\dagger \rangle \geq 0\}$ is a closed half space.

If $C = \mathbb{R}^\ell_+$, then $C^\dagger = \mathbb{R}^\ell_+$.

It is worth being more explicit about the dual cone of the class of non-decreasing functions.

Example 3.C.3. *If $C = \{(x_1, x_2, x_3)' \in \mathbb{R}^3 : x_1 \leq x_2 \leq x_3\}$, then $C^\dagger$ is the smallest cone containing $(-1, 1, 0)'$ and $(0, -1, 1)'$, i.e. $C^\dagger = \{\alpha(-1, 1, 0)' + \beta(0, -1, +1) : \alpha, \beta \geq 0\}$.*

Detour 3.2. *In $M_b([0, 1])$, the set of measurable, bounded functions on $[0, 1]$, the class, $C = NDcr$, of non-decreasing functions is cone that is closed under pointwise convergence. For $\mu, \nu \in \Delta([0, 1])$, $\mu \succsim_{FOSD} \nu$ iff for all $f \in C$, $\int f\, d\mu \geq \int f\, d\nu$. Again, this is $(\mu - \nu) \in C^\dagger$, but $C^\dagger$ is somewhat harder to describe.*

3.D. **Reminders About Supermodularity.** This material is at its easiest when actions and parameters are linearly ordered, which for us means that we can treat them as subsets of $\mathbb{R}$. The lattice formulations will come later.

Definition 3.D.1. *For linearly ordered A and Θ, a function $f : A \times \Theta \to \mathbb{R}$ is **supermodular** if for all $a' \succ a$ and all $\theta' \succ \theta$,*

$$f(a', \theta') - f(a, \theta') \geq f(a', \theta) - f(a, \theta), \tag{16}$$

equivalently

$$f(a', \theta') - f(a', \theta) \geq f(a, \theta') - f(a, \theta). \tag{17}$$

*It is **strictly supermodular** if the inequalities are strict.*

At θ, the benefit of increasing from a to a' is $f(a', \theta) - f(a, \theta)$, at θ', it is $f(a', \theta') - f(a, \theta')$. This assumption asks that benefit of increasing a be increasing in θ. A good verbal shorthand for this is that f **has increasing differences in** a **and** θ. Three sufficient conditions in the differentiable case are: $\forall a$, $f_a(a, \cdot)$ is nondecreasing; $\forall \theta$, $f_\theta(\cdot, \theta)$ is nondecreasing; and $\forall a, \theta$, $f_{a,\theta}(a, \theta) \geq 0$.

Theorem 3.4. *If $f : A \times \Theta \to \mathbb{R}$ is supermodular and $a^*(\theta)$ is the largest (or the smallest) solution to $\max_{a \in A} f(a, \theta)$, then $[\theta' \succ \theta] \Rightarrow [a^*(\theta') \succsim a^*(\theta)]$. If f is strictly supermodular, then for any $a' \in a^*(\theta')$ and any $a \in a^*(\theta)$, $a' \geq a$.*

Proof. Suppose that $\theta' \succ \theta$ but that $a' = a^*(\theta') \prec a = a^*(\theta)$. Because $a^*(\theta)$ and $a^*(\theta')$ are maximizers, $f(a', \theta') \geq f(a, \theta')$ and $f(a, \theta) \geq f(a', \theta)$. Since a' is the largest of the maximizers at θ' and $a \succ a'$, we know a bit more, that $f(a', \theta') > f(a, \theta')$. Adding the inequalities, we get $f(a', \theta') + f(a, \theta) > f(a, \theta') + f(a', \theta)$, or

$$f(a, \theta) - f(a', \theta) > f(a, \theta') - f(a', \theta'). \tag{18}$$

But $\theta' \succ \theta$ and $a \succ a'$ and supermodularity imply that this inequality must go the other way. The argument using strict supermodularity is similar. □

Sometimes the set of available choices also shifts with θ.

Theorem 3.5. *Suppose that A and Θ are non-empty subsets of $\mathbb{R}$, that $\Gamma(\theta) = [g(\theta), h(\theta)] \cap A$ where g and h are weakly increasing functions with $g \leq h \leq \infty$, that $f : A \times \Theta \to \mathbb{R}$ is supermodular. Then the smallest and the largest solutions to the problem $P(\theta) = \max_{a \in \Gamma(\theta)} f(a, \theta)$ are weakly increasing functions. Further, if f is strictly supermodular, then every selection from $\Psi(\theta) := \text{argmax}_{a \in \Gamma(\theta)} f(a, \theta)$ is weakly increasing.*

Proof. Fill it in. □

3.E. **The Monotone Likelihood Ratio Property (MLRP).** We are interested in $U(a, \theta) := \int u(i, a)\, dp_\theta(i)$ being supermodular an a and θ. A first step down this road is the MLRP.

Definition 3.E.1. *A family of strictly positive densities $\{f(s; \theta) : \theta \in \Theta\}$, $\Theta \subset \mathbb{R}$, on $\mathbb{R}$ or on $S = \{1, 2, \ldots, \ell\}$ has the **monotone likelihood ratio property (MLRP)** if there exists a $s \mapsto T(s) \in \mathbb{R}$ such that for any $\theta' > \theta$, $f(s; \theta')$ and $f(s; \theta)$ are the densities of different distributions, and $\frac{f(s;\theta')}{f(s;\theta)}$ is a non-decreasing function of $T(s)$.*

Comment: if $s \mapsto T(s)$ is monotonic, as it often/almost always is, we can simplify the assumption to $\frac{f(s;\theta')}{f(s;\theta)}$ is a nondecreasing function of s.

There is a little bit of an embarassment in trying to extend the previous definition to all densities/probabilities — we need to be careful about dividing by 0 and about what $\frac{0}{0}$ is going to mean if the densities go to 0 at different places. We will return to this.

Taking logarithms, the non-decreasing property of $s \mapsto \frac{f(s;\theta')}{f(s;\theta)}$ is supermodularity. Specifically, let $h(s; \theta) = \log(f(s; \theta))$, for $s' > s$,

$$\tfrac{f(s';\theta')}{f(s';\theta)} \geq \tfrac{f(s;\theta')}{f(s;\theta)} \text{ iff } [h(s'; \theta') - h(s'; \theta)] \geq [h(s; \theta') - h(s; \theta)]\,. \tag{19}$$

To put it another way, what we need for the MLRP is that for all $\theta' > \theta$, $h(\cdot; \theta') - h(\cdot; \theta)$ belongs to the cone of non-decreasing functions. That is, $h(\cdot; \theta') \succsim_{NDcr} h(\cdot; \theta)$.

Recall that we are interested in problems of the form

$$\max_{a \in A}\, U(a, \theta) := \int u(a, s) f(s; \theta)\, ds. \tag{20}$$

to have $U(\cdot, \cdot)$ supermodular, we need, for $a' > a$ and $\theta' > \theta$, to have

$$\int [u(a', s) - u(a, s)]\ f(s; \theta')\, ds \geq \int [u(a', s) - u(a, s)]\ f(s, \theta)\, ds. \tag{21}$$

Lemma 3.6. *If $u : A \times S \to \mathbb{R}$ is supermodular and $\{f(s, \theta) : \theta \in \Theta\}$ has the MLRP, then the inequality in* (21) *holds.*

Proof. Fill in. □

3.F. **Problems.**

Cones and stochastic dominance

Cone 1. If C is a cone of functions, then $\mu \succsim_{C^\dagger} \nu$ iff μ second order stochastically dominates ν. Give the cone and its dual, proving your results.

Expected utility maximization

Eu 1. Consider the following classes of portfolio choice problems,

$$\max_{x \in [0,w]} \int u(w - x + xs) f(s; \theta)\, ds \tag{22}$$

where $\{f(s; \theta) : \theta \in \Theta\}$, $\Theta \subset \mathbb{R}$, is a class of distributions on $[0, \infty)$ with the MLRP.

a. If $u(r) = \log(r)$, does the supermodularity analysis tell us whether or not $x^*(\theta)$ is an increasing or decreasing function?
b. If $u(r) = r^\gamma$, $0 < \gamma < 1$, does the supermodularity analysis tell us whether or not $x^*(\theta)$ is an increasing or decreasing function?
c. If $u(r) = r^\gamma$, $\gamma \geq 1$, does the supermodularity analysis tell us whether or not $x^*(\theta)$ is an increasing or decreasing function?
d. In the previous three problems, characterize, if possible, the set of θ for which x^* increases with w.

4. Topological Vector Spaces

At topological vector space (tvs) is a vector space with a topology making addition and multiplication by constants continuous. We haven't formally defined a vector space, or a topology, or continuity yet, so this is but a guideline to where we are headed.

At different times we will care about the following vector spaces and their duals, sometimes even their double duals: $C(K)$, the continuous functions on a compact metric space K, with a leading special case being $K = \{1, 2, \ldots, \ell\}$ in which case $C(K) = \mathbb{R}^\ell$; $C(\mathbb{R}^k)$, the set of continuous functions on $\mathbb{R}^k$; $C_b(M)$, the continuous bounded functions on a metric space M; $Lip(M)$, the set of Lipschitz continuous functions on a metric space M; $\mathbb{R}^\mathbb{N}$, the set of all real-valued sequences; $\ell_p \subset \mathbb{R}^\mathbb{N}$, $p \in [1, \infty)$, the set of infinite sequences $x = (x_1, x_2, \ldots)$ such that $\sum_t |x_t|^p < \infty$; $\ell_\infty \subset \cap_{p \in [1,\infty)} \ell_p$, the set of infinite sequences with $\sup_t |x_t| < \infty$; $C^k(G)$, the set of k-times continuously differentiable functions on (a usually open and well-behaved) $G \subset \mathbb{R}^\ell$; the set of multi-nomials on $\mathbb{R}^k$; $C^\infty(U)$), the space of infinitely differentiable functions on on open $U \subset \mathbb{R}^k$, and $D(U) \subset C^\infty(U)$, the set of compactly supported elements of $C^\infty(U)$; $ca(M)$, the countably additive finite signed measures on a metric space M, with a specially leading role being played by the subset $\Delta(M)$, the set of countably additive probabilities on M; $L^p(\Omega, \mathcal{F}, P)$, $p \in [1, \infty)$, the set of measurable,

$\mathbb{R}$-valued functions, X, such that $\int |X(\omega)|^p \, dP(\omega) < \infty$; $L((\Omega, \mathcal{F}, P); ca(A))$, the set of measurable functions from Ω to $ca(A)$, A a compact metric space, with a specially leading role played by the subset such that $\int |X(\omega)|^p \, dP(\omega) < \infty$; $L((\Omega, \mathcal{F}, P); \Delta(A))$.

As we go through the general definitions, seeing what is at work in these different vector spaces is a useful exercise.

We will give each of these spaces a topology, sometimes many topologies, for which the mappings $(\boldsymbol{x}, \boldsymbol{y}) \mapsto (\boldsymbol{x} + \boldsymbol{y})$ from $\mathfrak{X} \times \mathfrak{X}$ to $\mathfrak{X}$ and $(r, \boldsymbol{x}) \mapsto r \cdot \boldsymbol{x}$ from $\mathbb{R} \times \mathfrak{X}$ to $\mathfrak{X}$ are continuous. At that point, we will have made the vector space into a **topological vector space**. Time to define terms.

4.A. **Vector Spaces.** Eight axioms for a vector space.

Adding sets and multiplying them by scalars.

Spanning, linear independence, Hamel basis for a vector space.

Convex sets, balanced sets, absolutely convex sets, convex hull, absolutely convex hull, absorbent sets.

4.B. **Topological Spaces.** Three axioms for a topological space.

A base of neighborhoods (or neighborhood base).

Separate (i.e. Hausdorff) topological spaces.

Metric spaces (most common example for us).

Continuous functions, homeomorphisms.

4.C. **Vector Spaces with Compatible Topologies.** For a tvs, the neighborhood base at 0 tells us everything there is to know.

Locally convexity of a tvs (we will have this in all of our applications).

Examples: go through the spaces above giving topologies.

4.D. **Seminorms.** Repeating what we had above,

Definition 4.D.1. *A function $p : \mathfrak{X} \to \mathbb{R}_+$ is a **semi-norm** if $(\forall \lambda \in \mathbb{R})(\forall \boldsymbol{x} \in \mathfrak{X})[p(\lambda \boldsymbol{x}) = |\lambda| \boldsymbol{x}$ and $(\forall \boldsymbol{x}, \boldsymbol{y} \in \mathfrak{X})[p(\boldsymbol{x} + \boldsymbol{y}) \leq p(\boldsymbol{x}) + p(\boldsymbol{y})]$. If a semi-norm also satisfies $p(\boldsymbol{x}) = 0$ iff $\boldsymbol{x} = 0$, then it is a **norm**.*

Minkowski guages.

Topologies from semi-norms. Normed spaces. Metrizability.

5. Duality and the Hahn-Banach Theorem

5.A. **Linear Mappings.** We often use the word "functional" when we mean "function" or "mapping." This is a holdover from the early explorations of properties of spaces of functions, at that time, the notion that the domain was a set of functions seemed sort of weird, so one distinguished the domain by using "functional." And anyway, "functional analysis" sounds pretty sophisticated.

Definition linearity.

Continuity at the origin is necessary and sufficient for continuity.

Examples: discontinuous linear functionals; $ca([0,1])$ as $C([0,1])^\dagger$, weak* neighborhoods of 0 contain elements of arbitrarily large norm.

To work with a given a vector space $\mathfrak{X}$, it is often useful to know the form of $\mathfrak{X}^\dagger$. Such results are called representation theorems, that is, they are results that tell us how to represent the elements of $\mathfrak{X}^\dagger$.

5.B. **The Hahn-Banach Theorem.**

5.C. **Duality, Weak Topologies, and Weak* Topologies.**

5.D. **Polar Sets.**

5.E. **Cones and Dual Cones.**

5.F. **Finite Dimensional Subspaces.** They all "look" the same, and they are always **complemented**, something that will help give perspective on the econometrics that we will do.

6. Dualities in Measure Spaces

For $p, q \in [1, \infty]$ are a **dual pair** if $\frac{1}{p} + \frac{1}{q} = 1$. For $p \in [1, \infty)$ and p, q a dual pair, $L^q = L^q(\Omega, \mathcal{F}, P)$ is the dual of L^p with the bilinear funnctional being $\langle f, g \rangle = \int f(\omega)g(\omega)\, dP(\omega)$. However, the dual of L^∞ is larger than L^1.

In any case, for p, q a dual pair we have the weak topologies: $f_n \to f$ weakly in L^p if for all $g \in L^q$, $\langle f_n, g \rangle \to \langle f, g \rangle$ and $g_n \to g$ weakly in L^q if for all $f \in L^q$, $\langle f, g_n \rangle \to \langle f, g \rangle$.

A well-known example of weak convergence that is not norm convergence is $f_n(\omega) = \cos(n\omega)$ for $\omega \in [0, 1]$ with P being the uniform distribution. $\|f_n - f\|_p$ is approximately 1 for large n, but $f_n \to 0$ weakly.

References

[Athey, 2002] Athey, S. (2002). Monotone comparative statics under uncertainty. *The Quarterly Journal of Economics*, 117(1):187–223.

[Aumann et al., 1983] Aumann, R. J., Katznelson, Y., Radner, R., Rosenthal, R. W., and Weiss, B. (1983). Approximate purification of mixed strategies. *Math. Oper. Res.*, 8(3):327–341.

[Bellman and Blackwell, 1949] Bellman, R. and Blackwell, D. (1949). Some two-person games involving bluffing. *Proc. Nat. Acad. Sci. U. S. A.*, 35:600–605.

[Corbae et al., 2009] Corbae, D., Stinchcombe, M. B., and Zeman, J. (2009). *An introduction to mathematical analysis for economic theory and econometrics.* Princeton University Press, Princeton, NJ.

[Dvoretzky et al., 1951] Dvoretzky, A., Wald, A., and Wolfowitz, J. (1951). Elimination of randomization in certain statistical decision procedures and zero-sum two-person games. *Ann. Math. Statistics*, 22:1–21.

[Govindan et al., 2003] Govindan, S., Reny, P. J., and Robson, A. J. (2003). A short proof of Harsanyi's purification theorem. *Games Econom. Behav.*, 45(2):369–374. Special issue in honor of Robert W. Rosenthal.

[Harsanyi, 1973] Harsanyi, J. C. (1973). Games with randomly disturbed payoffs: a new rationale for mixed-strategy equilibrium points. *Internat. J. Game Theory*, 2(1):1–23.

[Khan et al., 1999] Khan, M. A., Rath, K. P., and Sun, Y. (1999). On a private information game without pure strategy equilibria. *J. Math. Econom.*, 31(3):341–359.

[Manelli, 1996] Manelli, A. M. (1996). Cheap talk and sequential equilibria in signaling games. *Econometrica*, 64(4):917–942.

[Milgrom, 1994] Milgrom, P. (1994). Comparing optima: Do simplifying assumptions affect conclusions? *Journal of Political Economy*, 102(3):607–15.

[Milgrom and Weber, 1985] Milgrom, P. R. and Weber, R. J. (1985). Distributional strategies for games with incomplete information. *Math. Oper. Res.*, 10(4):619–632.

[Radner and Rosenthal, 1982] Radner, R. and Rosenthal, R. W. (1982). Private information and pure-strategy equilibria. *Math. Oper. Res.*, 7(3):401–409.

[Robertson and Robertson, 1980] Robertson, A. P. and Robertson, W. (1980). *Topological vector spaces*, volume 53 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, second edition.

[Simon, 2003] Simon, R. S. (2003). Games of incomplete information, ergodic theory, and the measurability of equilibria. *Israel J. Math.*, 138:73–92.

[Smith and McCardle, 2002] Smith, J. E. and McCardle, K. F. (2002). Structural properties of stochastic dynamic programs. *Operations Research*, 50(5):796–809.