Game Theory, Alive

Anna R. Karlin and Yuval Peres

Two chapters from draft of upcoming book May 14, 2015

Please send comments and corrections to

`karlin@cs.washington.edu`

1

Two-person zero-sum games

We begin with the theory of **two-person zero-sum games**, developed in a seminal paper by John von Neumann and Oskar Morgenstern. In these games, one player's loss is the other player's gain. The central theorem for two-person zero-sum games is that even if each player's strategy is known to the other, there is an amount that one player can guarantee as her expected gain, and the other, as his maximum expected loss. This amount is known as the **value** of the game.

1.1 Examples

Consider the following game:

Example 1.1.1 (***Pick-a-Hand*****, a betting game).** There are two players, Chooser (player I), and Hider (player II). Hider has two gold coins in his back pocket. At the beginning of a turn, he† puts his hands behind his back and either takes out one coin and holds it in his left hand (strategy $L1$), or takes out both and holds them in his right hand (strategy $R2$). Chooser picks a hand and wins any coins the hider has hidden there. She may get nothing (if the hand is empty), or she might win one coin, or two. How much should Chooser be willing to pay in order to play this game?

The following matrix summarizes the payoffs to Chooser in each of the cases.

		Hider	
		$L1$	$R2$
Chooser	L	1	0
	R	0	2

---

† In all two-person games, we adopt the convention that player I is female and player II is male.

How should Hider and Chooser play? Imagine that they are conservative and want to optimize for the worst case scenario. Hider can guarantee himself a loss of at most 1 by selecting action L1, whereas if he selects R2, he has the potential to lose 2. Chooser cannot guarantee herself any positive gain since, if she selects L, in the worst case, Hider selects R2, whereas if she selects R, in the worst case, Hider selects L1.

Now consider expanding the possibilities available to the players by incorporating randomness. Suppose that Hider selects $L1$ with probability y_1 and $R2$ with probability $y_2 = 1 - y_1$. Hider's expected loss is y_1 if Chooser plays L, and $2(1 - y_1)$ if Chooser plays R. Thus Hider's worst-case expected loss is $\max(y_1, 2(1 - y_1))$. To minimize this, Hider will choose $y_1 = 2/3$. Thus, **no matter how Chooser plays, Hider can guarantee himself an expected loss of at most 2/3.** See Figure 1.1.

Similarly, suppose that Chooser selects L with probability x_1 and R with probability $x_2 = 1 - x_1$. Then Chooser's worst-case expected gain is $\min(x_1, 2(1 - x_1))$. To maximize this, she will choose $x_1 = 2/3$. Thus, **no matter how Hider plays, Chooser, can guarantee herself an expected gain of at least 2/3.**

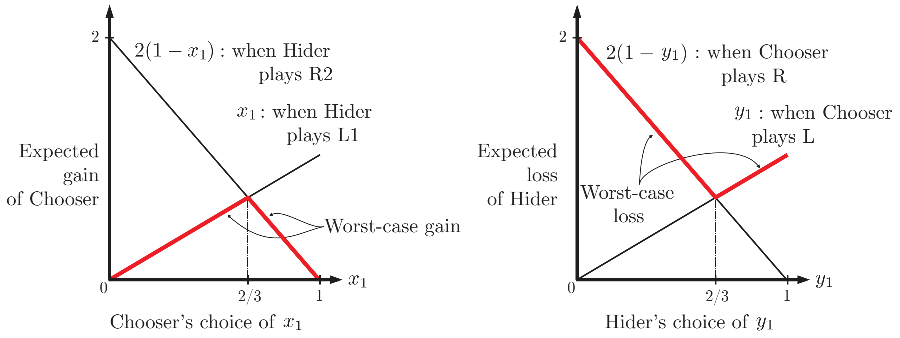

Fig. 1.1. The left side of the figure shows the worst-case expected gain of Chooser as a function of x_1, the probability with which she plays L. The right side of the figure shows the worst-case expected loss of Hider as a function of y_1, the probability with which he plays L1. (In this example, the two graphs "look" the same because the payoff matrix is symmetric. See Example 1.1.2 for a game where the two graphs are different.)

Notice that without some extra incentive, it is not in Hider's interest to play *Pick-a-hand* because he can only lose by playing. To be enticed into joining the game, Hider will need to be paid at least 2/3. Conversely, Chooser should be willing to pay any sum below 2/3 to play the game. Thus, we say that the **value** of this game is 2/3; we can think if it as being equivalent to the following 1 by 1 game.

	Hider
Chooser	2/3

Exercise 1.1.2 (**Another Betting Game**)**.** Consider the betting game with the following payoff matrix:

		player II	
		L	R
player I	T	0	2
	B	5	1

Draw graphs for this game analogous to those shown in Figure 1.1, and determine the value of the game.

1.2 Definitions

A two-person zero-sum game can be represented by an $m \times n$ **payoff matrix** $A = (a_{ij})$, whose rows are indexed by the m possible actions of player I, and whose columns are indexed by the n possible actions of player II. Player I selects an action i and player II selects an action j, each unaware of the other's selection. Their selections are then revealed and player II pays player I the amount a_{ij}.

If player I selects action i, in the worst case her gain will be $\min_j a_{ij}$, and thus the largest gain she can guarantee is $\max_i \min_j a_{ij}$. Similarly, if II selects action j, in the worst case his loss will be $\max_i a_{ij}$, and thus the smallest loss he can guarantee is $\min_j \max_i a_{ij}$. It follows that

$$\max_i \min_j a_{ij} \leq \min_j \max_i a_{ij} \tag{1.1}$$

since player I can guarantee gaining the left hand side and player II can guarantee losing no more than the right hand side. (For a formal proof, see Lemma 1.5.3.) As in Example 1.1.1, without randomness, the inequality is usually strict.

A strategy in which each action is selected with some probability is a **mixed strategy**. A mixed strategy for player I is determined by a vector

$(x_1, \ldots, x_m)^T$ where x_i represents the probability of playing action i. The set of mixed strategies for player I is denoted by

$$\Delta_m = \left\{ \mathbf{x} \in \mathbb{R}^m : x_i \geq 0, \sum_{i=1}^{m} x_i = 1 \right\}.$$

Similarly, the set of mixed strategies for player II is denoted by

$$\Delta_n = \left\{ \mathbf{y} \in \mathbb{R}^n : y_j \geq 0, \sum_{j=1}^{n} y_j = 1 \right\}.$$

A mixed strategy in which a particular action is played with probability 1 is called a **pure strategy**. Observe that in this vector notation, pure strategies are represented by the standard basis vectors, though we often identify the pure strategy $\mathbf{e}_i$ with the corresponding action i.

If player I employs strategy $\mathbf{x}$ and player II employs strategy $\mathbf{y}$, the expected gain of player I (which is the same as the expected loss of player II) is

$$\mathbf{x}^T A\mathbf{y} = \sum_i \sum_j x_i a_{ij} y_j.$$

Thus, if player I employs strategy $\mathbf{x}$, she can guarantee herself an expected gain of

$$\min_{\mathbf{y} \in \Delta_n} \mathbf{x}^T A\mathbf{y} = \min_j (\mathbf{x}^T A)_j \tag{1.2}$$

since for any $\mathbf{z} \in \mathbb{R}^n$, we have $\min_{\mathbf{y} \in \Delta_n} \mathbf{z}^T \mathbf{y} = \min_j z_j$.

A conservative player will choose $\mathbf{x}$ to maximize (1.2), that is, to maximize her worst case expected gain. This is a safety strategy.

Definition 1.2.1. A mixed strategy $\mathbf{x}^* \in \Delta_m$ is **a safety strategy for player I** if the maximum over $\mathbf{x} \in \Delta_m$ of the function

$$\mathbf{x} \mapsto \min_{\mathbf{y} \in \Delta_n} \mathbf{x}^T A\mathbf{y}$$

is attained at $\mathbf{x}^*$. The value of this function at $\mathbf{x}^*$ is the **safety value for player I** . Similarly, a mixed strategy $\mathbf{y}^* \in \Delta_n$ is **a safety strategy for player II** if the minimum over $\mathbf{y} \in \Delta_n$ of the function

$$\mathbf{y} \mapsto \max_{\mathbf{x} \in \Delta_m} \mathbf{x}^T A\mathbf{y}$$

is attained at $\mathbf{y}^*$. The value of this function at $\mathbf{y}^*$ is the **safety value for player II** .

Remark. For the existence of safety strategies see Lemma 1.5.3.

Safety strategies might appear conservative, but the following celebrated theorem shows that the two players' safety values coincide.

Theorem 1.2.2. ***von Neumann's minimax Theorem***
For any finite two-person zero-sum game, there is a number V, called the value *of the game, satisfying*

$$\max_{\mathbf{x}\in\Delta_m} \min_{\mathbf{y}\in\Delta_n} \mathbf{x}^T A\mathbf{y} = V = \min_{\mathbf{y}\in\Delta_n} \max_{\mathbf{x}\in\Delta_m} \mathbf{x}^T A\mathbf{y}. \tag{1.3}$$

We will prove the minimax theorem in §1.5.

Remarks:

(i) It is easy to check that the left hand side of equation (1.3) is upper bounded by the right hand side, i.e.

$$\max_{\mathbf{x}\in\Delta_m} \min_{\mathbf{y}\in\Delta_n} \mathbf{x}^T A\mathbf{y} \leq \min_{\mathbf{y}\in\Delta_n} \max_{\mathbf{x}\in\Delta_m} \mathbf{x}^T A\mathbf{y}. \tag{1.4}$$

(See the argument for equation 1.1 and Lemma 1.5.3). The magic of zero-sum games is that, in mixed strategies, this inequality becomes an equality.

(ii) If $\mathbf{x}^*$ is a safety strategy for player I and $\mathbf{y}^*$ is a safety strategy for player II, then it follows from Theorem 1.2.2 that:

$$\min_{\mathbf{y}\in\Delta_n} (\mathbf{x}^*)^T A\mathbf{y} = V = \max_{\mathbf{x}\in\Delta_m} \mathbf{x}^T A\mathbf{y}^*. \tag{1.5}$$

In words, this means: that the mixed strategy $\mathbf{x}^*$ yields player I an expected gain of at least V, no matter how II plays, and the mixed strategy $\mathbf{y}^*$ yields player II an expected loss of at most V, no matter how I plays. Therefore, from now on, we will refer to the safety strategies in zero-sum games as **optimal strategies**.

1.3 Simplifying and solving zero-sum games

In this section, we will discuss techniques that help us understand zero-sum games and solve them (that is, find their value and determine optimal strategies for the two players).

1.3.1 Pure optimal strategies: saddle points

Given a zero-sum game, the first thing to check is whether or not there is a pair of optimal strategies that is pure.

For example, in the following game, by playing action 1, player I guarantees himself a payoff at least 2 (since that is the smallest entry in the row). Similarly, by playing action 1, player II guarantees himself a loss of at most 2. Thus, the value of the game is 2.

		player II	
		action 1	action 2
player I	action 1	2	3
	action 2	1	0

Definition 1.3.1. A **saddle point**† of a payoff matrix A is a pair (i^*, j^*) such that

$$\max_i a_{ij^*} = a_{i^*j^*} = \min_j a_{i^*j} \tag{1.6}$$

If (i^*, j^*) is a saddle point, then $a_{i^*j^*}$ is the value of the game. A saddle point is also called a **pure Nash equilibrium**: given the action pair (i^*, j^*), neither player has an incentive to deviate. See §1.4 for a more detailed discussion of Nash equilibria.

1.3.2 Equalizing payoffs

Most zero-sum games do not have pure optimal strategies. At the other extreme, some games have a pair $(\mathbf{x}^*, \mathbf{y}^*)$ of optimal strategies that are fully mixed, that is, where each action is assigned positive probability. In this case, it must be that against $\mathbf{y}^*$, player I obtains the same payoff from each action. If not, say $(A\mathbf{y}^*)_1 > (A\mathbf{y}^*)_2$, then player I could increase her gain by moving probability from action 2 to action 1: this contradicts the optimality of $\mathbf{x}^*$. Applying this observation to both players enables us to solve for optimal strategies by equalizing payoffs. Consider, for example, the following payoff matrix, where each row and column is labelled with the probability that the corresponding action is played in the optimal strategy.

---

† The term saddle point comes from the continuous setting where a function $f(x, y)$ of two variables has a point (x^*, y^*) at which locally $\max_x f(x, y^*) = f(x^*, y^*) = \min_y f(x^*, y)$. Thus, the surface resembles a saddle that curves up the the y direction and curves down in the x direction.

		player II	
		y_1	$1-y_1$
player I	x_1	3	0
	$1-x_1$	1	4

Equalizing the gains for player I's actions, we obtain

$$3y_1 = y_1 + 4(1 - y_1)$$

or $y_1 = 2/3$. Thus, if player II plays $(2/3, 1/3)$, his loss will not depend on player I's actions; it will be 2 no matter what I does.

Similarly, equalizing the losses for player II's actions, we obtain

$$3x_1 + (1 - x_1) = 4(1 - x_1)$$

or $x_1 = 1/2$. So if player I plays $(1/2, 1/2)$, his gain will not depend on player II's action; again, it will be 2 no matter what II does. We conclude that the value of the game is 2.

See Corollary 1.4.2 for a general version of the equalization principle.

Exercise 1.3.2. Show that any 2-by-2 game has a pair of optimal strategies that are both pure or both fully mixed. Show that this can fail for 3-by-3 games.

1.3.3 The technique of domination

Domination is a technique for reducing the size of a game's payoff matrix, enabling it to be more easily analyzed. Consider the following example.

Example 1.3.3 (***Plus One***)**.** Each player chooses a number from $\{1, 2, \ldots, n\}$ and writes it down; then the players compare the two numbers. If the numbers differ by one, the player with the higher number wins \$1 from the other player. If the players' choices differ by two or more, the player with the higher number pays \$2 to the other player. In the event of a tie, no money changes hands.

The payoff matrix for the game is:

		player II									
		1	2	3	4	5	6	$\cdots$			n
player I	1	0	-1	2	2	2	2	$\cdots$			2
	2	1	0	-1	2	2	2	$\cdots$			2
	3	-2	1	0	-1	2	2	$\cdots$			2
	4	-2	-2	1	0	-1	2	$\cdots$			2
	5	-2	-2	-2	1	0	-1	2			2
	6	-2	-2	-2	-2	1	0	2			2
	$n-1$	-2	-2	$\cdots$						0	-1
	n	-2	-2	$\cdots$					-2	1	0

In this payoff matrix, every entry in row 4 is at most the corresponding entry in row 1. Thus player I has no incentive to play 4 since it is **dominated** by row 1. In fact, rows 4 through n are all dominated by row 1, and hence player I can ignore those strategies.

By symmetry, we see that player II need never play any of strategies 4 through n. Thus, in *Plus One* we can search for optimal strategies in the reduced payoff matrix:

		player II		
		1	2	3
player I	1	0	-1	2
	2	1	0	-1
	3	-2	1	0

To analyze the reduced game, let $\mathbf{x}^T = (x_1, x_2, x_3)$ be player I's mixed strategy. For $\mathbf{x}$ to be optimal, each component of

$$\mathbf{x}^T A = (x_2 - 2x_3, \; -x_1 + x_3, \; 2x_1 - x_2) \tag{1.7}$$

must be at least the value of the game. In this game, there is complete symmetry between the players. This implies that the payoff matrix is **anti-symmetric**: the game matrix is square and $a_{ij} = -a_{ji}$ for every i and j.

Claim 1.3.4. *If the payoff matrix of a zero-sum game is anti-symmetric, then the game has value 0.*

Proof. This is intuitively clear by symmetry. Formally, suppose that the value of the game is V. Then there is a vector $\mathbf{x} \in \Delta_n$ such that for all

$\mathbf{y} \in \Delta_n$, $\mathbf{x}^T A\mathbf{y} \geq V$. In particular

$$\mathbf{x}^T A\mathbf{x} \geq V. \tag{1.8}$$

Taking the transpose of both sides yields $\mathbf{x}^T A^T \mathbf{x} = -\mathbf{x}^T A\mathbf{x} \geq V$. Adding this latter inequality to (1.8) yields $V \leq 0$. Similarly, there is a $\mathbf{y} \in \Delta_n$ such that for all $\tilde{\mathbf{x}} \in \Delta_n$ we have $\tilde{\mathbf{x}}^T A\mathbf{y} \leq V$. Taking $\tilde{\mathbf{x}} = \mathbf{y}$ yields in the same way that $0 \leq V$. □

We conclude that for any optimal strategy $\mathbf{x}$ in *Plus One*

$$\begin{aligned} x_2 - 2x_3 &\geq 0 \\ -x_1 + x_3 &\geq 0 \\ 2x_1 - x_2 &\geq 0, \end{aligned}$$

Thus $x_2 \geq 2x_3$, $x_3 \geq x_1$, and $2x_1 \geq x_2$. If one of these inequalities was strict, then adding the first, twice the second and the third, we could deduce $0 > 0$, so in fact each of them must be an equality. Solving the resulting system, with the constraint $x_1 + x_2 + x_3 = 1$, we find that the optimal strategy for each player is $(1/4, 1/2, 1/4)$.

Summary of domination

We say a row ℓ of a two-person zero-sum game dominates row i if $a_{\ell j} \geq a_{ij}$ for all j. When row i is dominated, then there is no loss to player I if she never plays it. More generally, we say that subset I of rows dominates row i if some convex combination of the rows in I dominates row i, i.e., there is a probability vector $(\beta_\ell)_{\ell \in I}$ such that for every j

$$\sum_{\ell \in I} \beta_\ell a_{\ell j} \geq a_{ij}. \tag{1.9}$$

Similar definitions hold for columns.

Exercise 1.3.5. Prove that if equation (1.9) holds, then player I can safely ignore row i.

1.3.4 The use of symmetry

Another way to simplify the analysis of a game is via the technique of **symmetry**. We illustrate a symmetry argument in the following example:

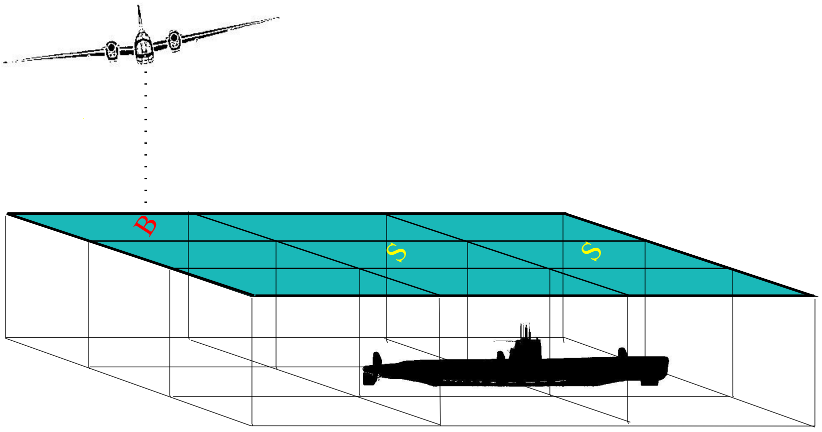

Fig. 1.2. The bomber chooses one of the nine squares to bomb. She cannot see which squares represent the location of the submarine.

Example 1.3.6 (***Submarine Salvo***)**.** A submarine is located on two adjacent squares of a three-by-three grid. A bomber (player I), who cannot see the submerged craft, hovers overhead and drops a bomb on one of the nine squares. She wins \$1 if she hits the submarine and \$0 if she misses it. (See Figure 1.2.) There are nine pure strategies for the bomber and twelve for the submarine, so the payoff matrix for the game is quite large. To determine some, but not all, optimal strategies, we can use symmetry arguments to simplify the analysis.

There are three types of moves that the bomber can make: She can drop a bomb in the center, in the middle of one of the sides, or in a corner. Similarly, there are three types of positions that the submarine can assume: taking up the center square, taking up a corner square and the adjacent square clockwise, or taking up a corner square and the adjacent square counter-clockwise. It is intuitive (and true) that both players have optimal strategies that assign equal probability to actions of the same type (e.g. *corner-clockwise*). To see this, observe that in *Submarine Salvo* a 90 degree rotation describes a permutation π of the possible submarine positions and a permutation σ of the possible bomber actions. Clearly π^4 (rotating by 90 degrees four times) is the identity and so is σ^4. For any bomber strategy $\mathbf{x}$, let $\pi\mathbf{x}$ be the rotated row strategy. (Formally $(\pi\mathbf{x})_i = x_{\pi(i)}$). Clearly, the probability that the bomber will hit the submarine if they play $\pi\mathbf{x}$ and $\sigma\mathbf{y}$ is the same as it is when they play $\mathbf{x}$ and $\mathbf{y}$, and therefore

$$\min_{\mathbf{y}} \mathbf{x}^T A\mathbf{y} = \min_{\mathbf{y}} (\pi\mathbf{x})^T A(\sigma\mathbf{y}) = \min_{\mathbf{y}} (\pi\mathbf{x})^T A\mathbf{y}.$$

Thus, if V is the value of the game and $\mathbf{x}$ is optimal, then $\pi^k\mathbf{x}$ is also optimal for all k.

Fix any submarine strategy $\mathbf{y}$. Then $\pi^k\mathbf{x}$ gains at least V against $\mathbf{y}$, hence so does

$$\mathbf{x}^* = \frac{1}{4}(\mathbf{x} + \pi\mathbf{x} + \pi^2\mathbf{x} + \pi^3\mathbf{x}).$$

Therefore $\mathbf{x}^*$ is an optimal rotation-invariant strategy.

Using these equivalences, we may write down a more manageable payoff matrix:

		submarine		
		center	corner-clockwise	corner-counterclockwise
bomber	corner	0	1/4	1/4
	midside	1/4	1/4	1/4
	middle	1	0	0

Note that the values for the new payoff matrix are different from those in the standard payoff matrix. They incorporate the fact that when, say, the bomber is playing *corner* and the submarine is playing *corner-clockwise*, there is only a one-in-four chance that there will be a hit. In fact, the pure strategy of corner for the bomber in this reduced game corresponds to the mixed strategy of bombing each corner with probability 1/4 in the original game. Similar reasoning applies to each of the pure strategies in the reduced game.

Since the right two columns yield the same payoff to the submarine, it's natural for the submarine to give them the same weight. This yields the mixed strategy of choosing uniformly one of the eight positions containing a corner. We can use domination to simplify the matrix even further. This is because for the bomber, the strategy *midside* dominates that of *corner* (because the submarine, when touching a corner, must also be touching a midside). This observation reduces the matrix to:

		submarine	
		center	corner
bomber	midside	1/4	1/4
	middle	1	0

Now note that for the submarine, *corner* dominates *center*, and thus we obtain the reduced matrix:

		submarine
		corner
bomber	midside	1/4
	middle	0

The bomber picks the better alternative — technically, another application of domination — and picks *midside* over *middle*. The value of the game is 1/4; the bomber's optimal strategy is to hit one of the four mid-sides with probability 1/4 each, and the optimal submarine strategy is to hide with probability 1/8 each in one of the eight possible pairs of adjacent squares that exclude the center. The symmetry argument is generalized in Exercise 1.21.

Remark. It is perhaps surprising that in *Submarine Salvo* there also exist optimal strategies that do not assign equal probability to all actions of the same type. (See exercise 1.15.)

1.4 Nash equilibria, equalizing payoffs and optimal strategies

.

A notion of great importance in game theory is **Nash equilibrium**. In §1.3.1, we introduced pure Nash equilibria. In this section, we introduce mixed Nash equilibria.

Definition 1.4.1. A pair of strategies $(\mathbf{x}^*, \mathbf{y}^*)$ is a **Nash equilibrium** in a zero-sum game with payoff matrix A if

$$\min_{\mathbf{y} \in \Delta_n} (\mathbf{x}^*)^T A \mathbf{y} = (\mathbf{x}^*)^T A \mathbf{y}^* = \max_{\mathbf{x} \in \Delta_m} \mathbf{x}^T A \mathbf{y}^*. \tag{1.10}$$

Thus, $\mathbf{x}^*$ is a **best response** to $\mathbf{y}^*$ and vice versa.

Remark. If $\mathbf{x}^* = \mathbf{e}_{i^*}$ and $\mathbf{y}^* = \mathbf{e}_{j^*}$, then by Equation (1.2), this definition coincides with Definition 1.3.1.

Proposition 1.4.2. *Let $\mathbf{x} \in \Delta_m$ and $\mathbf{y} \in \Delta_n$ be a pair of mixed strategies. The following are equivalent:*

(i) *The vectors $\mathbf{x}$ and $\mathbf{y}$ are in Nash equilibrium.*
(ii) *There are V_1, V_2 such that:*

$$\sum_i x_i a_{ij} \begin{cases} = V_1 & \text{for every } j \text{ such that } y_j > 0 \\ \geq V_1 & \text{for every } j \text{ such that } y_j = 0. \end{cases} \tag{1.11}$$

and

$$\sum_j a_{ij} y_j \begin{cases} = V_2 & \textit{for every } i \textit{ such that } x_i > 0 \\ \leq V_2 & \textit{for every } i \textit{ such that } x_i = 0 \end{cases} \tag{1.12}$$

(iii) *The vectors* **x** *and* **y** *are optimal.*

Remark. If (1.11) and (1.12) hold, then

$$V_1 = \sum_j y_j \sum_i x_i a_{ij} = \sum_i x_i \sum_j a_{ij} y_j = V_2.$$

Proof. (i) is equivalent to (ii): Clearly, **y** is a best response to **x** if and only if **y** assigns positive probability only to actions that yield II the minimum loss given **x**; this is precisely (1.11). The argument for (1.12) is identical. Thus (i) and (ii) are equivalent.

(ii) implies (iii): Player I guarantees herself a gain of at least V_1 and player II guarantees himself a loss of at most V_2. Since $V_1 = V_2$, these are optimal.

(iii) implies (i): Let $V = \mathbf{x}^T A \mathbf{y}$ be the value of the game. Since playing **x** guarantees I a gain of a least V, player II has no incentive to deviate from **y**. Similarly for player II.

□

1.4.1 A first glimpse of incomplete information

Example 1.4.3 (A random game). Consider the zero-sum two-player game in which the game to be played is randomized by a fair coin toss. If the toss comes up heads, the payoff matrix is given by A^H, and if tails, it is given by A^T.

$$A^H = \text{player I} \begin{array}{c|cc} & \multicolumn{2}{c}{\text{player II}} \\ & L & R \\ \hline U & 4 & 1 \\ D & 3 & 0 \end{array} \qquad A^T = \text{player I} \begin{array}{c|cc} & \multicolumn{2}{c}{\text{player II}} \\ & L & R \\ \hline U & 1 & 3 \\ D & 2 & 5 \end{array}$$

If the players don't know the outcome of the coin flip before playing, they are merely playing the game given by the average matrix

$$\frac{1}{2}A^H + \frac{1}{2}A^T = \begin{pmatrix} 2.5 & 2 \\ 2.5 & 2.5 \end{pmatrix}$$

which has a value of 2.5. If both players know the outcome of the coin flip, then (since A^H has a value of 1 and A^T has a value of 2) the value is 1.5 — player II is able to use the additional information to reduce his losses.

But now suppose that only I is told the result of the coin toss, but she must reveal her move first. If I adopts the simple strategy of picking the best row in whichever game is being played, and II realizes this and counters, then I has a payoff of only 1.5, less than the payoff if she ignores the extra information! See §**??** for a detailed analysis of this and related games.

This example demonstrates that sometimes the best strategy is to ignore extra information, and play as if it were unknown. For example, during World War II, Polish and British cryptanalysts had broken the secret code the Germans were using (the Enigma machine), and could therefore decode the Germans' communications. This created a challenging dilemma for the Allies: acting on the decoded information could reveal to the Germans that their code had been broken, which could lead them to switch to more secure encryption.

1.5 Proof of Von Neumann's minimax theorem*

We now prove the von Neumann Minimax Theorem. The proof will rely on a basic theorem from convex geometry.

Definition 1.5.1. A set $K \subseteq \mathbb{R}^d$ is **convex** if, for any two points $\mathbf{a}, \mathbf{b} \in K$, also lies in K. In other words, for every pair of points $\mathbf{a}, \mathbf{b} \in K$,

$$\{p\,\mathbf{a} + (1-p)\mathbf{b} : p \in [0,1]\} \in K,$$

Theorem 1.5.2 (**The Separating Hyperplane Theorem**)**.** *Suppose that $K \subseteq \mathbb{R}^d$ is closed and convex. If $\mathbf{0} \notin K$, then there exists $\mathbf{z} \in \mathbb{R}^d$ and $c \in \mathbb{R}$ such that*

$$0 < c < \mathbf{z}^T\mathbf{v}$$

for all $\mathbf{v} \in K$.

Here $\mathbf{0}$ denotes the vector of all 0's, and $\mathbf{z}^T\mathbf{v}$ is the usual dot product $\sum_i z_i v_i$. The theorem says that there is a **hyperplane** (a line in two dimensions, a plane in three dimensions, or, more generally, an affine $\mathbb{R}^{d-1}$-subspace in $\mathbb{R}^d$) that separates $\mathbf{0}$ from K. In particular, on any continuous path from $\mathbf{0}$ to K, there is some point that lies on this hyperplane. The separating hyperplane is given by $\{\mathbf{x} \in \mathbb{R}^d : \mathbf{z}^T\mathbf{x} = c\}$. The point $\mathbf{0}$ lies in the half-space $\{\mathbf{x} \in \mathbb{R}^d : \mathbf{z}^T\mathbf{x} < c\}$, while the convex body K lies in the complementary half-space $\{\mathbf{x} \in \mathbb{R}^d : \mathbf{z}^T\mathbf{x} > c\}$.

Recall first that the **(Euclidean) norm of v** is the (Euclidean) distance between $\mathbf{0}$ and $\mathbf{v}$, and is denoted by $\|\mathbf{v}\|$. Thus $\|\mathbf{v}\| = \sqrt{\mathbf{v}^T\mathbf{v}}$. A subset of a metric space is **closed** if it contains all its limit points, and **bounded** if it is

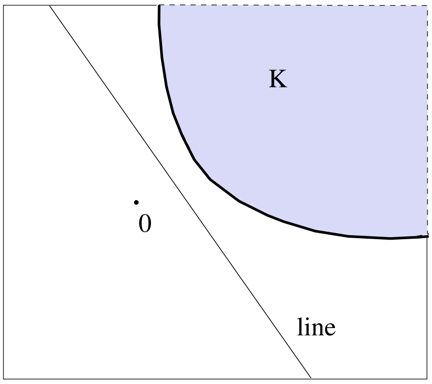

Fig. 1.3. Hyperplane separating the closed convex body K from $\mathbf{0}$.

contained inside a ball of some finite radius R. In what follows, the metric is the Euclidean metric.

Proof of Theorem 1.5.2. Let $B_r = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| \le r\}$ be the ball of radius r centered at $\mathbf{0}$. If we choose r so that B_r intersects K, the function $\mathbf{w} \mapsto \|\mathbf{w}\|$, considered as a map from $K \cap B_r$ to $[0, \infty)$, is continuous, with a domain that is nonempty, closed and bounded (see Figure 1.4). Thus the map attains its infimum at some point $\mathbf{z}$ in K. For this $\mathbf{z} \in K$ we have

$$\|\mathbf{z}\| = \inf_{\mathbf{w} \in K} \|\mathbf{w}\|.$$

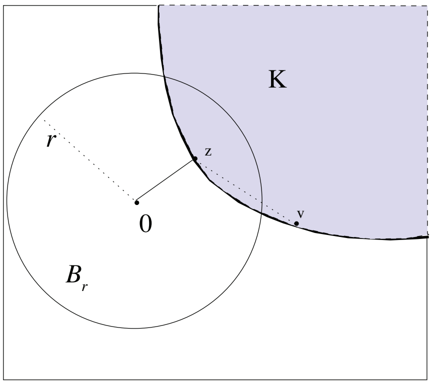

Fig. 1.4. Intersecting K with a ball to get a nonempty closed bounded domain.

Let $\mathbf{v} \in K$. Because K is convex, for any $\varepsilon \in (0,1)$, we have that $\varepsilon\mathbf{v} + (1-\varepsilon)\mathbf{z} = \mathbf{z} - \varepsilon(\mathbf{z} - \mathbf{v}) \in K$. Since $\mathbf{z}$ has the minimum norm of any point in K,

$$\|\mathbf{z}\|^2 \leq \|\mathbf{z} - \varepsilon(\mathbf{z} - \mathbf{v})\|^2.$$

Multiplying this out, we get

$$\|\mathbf{z}\|^2 \leq \|\mathbf{z}\|^2 - 2\varepsilon \mathbf{z}^T(\mathbf{z} - \mathbf{v}) + \varepsilon^2 \|\mathbf{z} - \mathbf{v}\|^2.$$

Cancelling $\|\mathbf{z}\|^2$ and rearranging terms we get

$$2\varepsilon \mathbf{z}^T(\mathbf{z} - \mathbf{v}) \leq \varepsilon^2 \|\mathbf{z} - \mathbf{v}\|^2 \quad \text{or} \quad \mathbf{z}^T(\mathbf{z} - \mathbf{v}) \leq \frac{\varepsilon}{2}\|\mathbf{z} - \mathbf{v}\|^2.$$

Letting ε approach 0, we find

$$\mathbf{z}^T(\mathbf{z} - \mathbf{v}) \leq 0 \quad \text{which means that} \quad \|\mathbf{z}\|^2 \leq \mathbf{z}^T\mathbf{v}.$$

Since $z \in K$ and $\mathbf{0} \notin K$, the norm $\|\mathbf{z}\| > 0$. Choosing $c = \frac{1}{2}\|\mathbf{z}\|^2$, we get $0 < c < \mathbf{z}^T\mathbf{v}$ for all $\mathbf{v} \in K$. □

We will also need the following simple lemma:

Lemma 1.5.3. *Let X and Y be closed and bounded sets in $\mathbb{R}^d$. Let $f : X \times Y \to \mathbb{R}$ be continuous. Then*

$$\max_{\mathbf{x} \in X} \min_{\mathbf{y} \in Y} f(\mathbf{x}, \mathbf{y}) \leq \min_{\mathbf{y} \in Y} \max_{\mathbf{x} \in X} f(\mathbf{x}, \mathbf{y}). \tag{1.13}$$

Proof. We first prove the lemma for the case where X and Y are finite sets (with no assumptions on f). Let $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) \in X \times Y$. Clearly

$$\min_{\mathbf{y} \in Y} f(\tilde{\mathbf{x}}, \mathbf{y}) \leq f(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) \leq \max_{\mathbf{x} \in X} f(\mathbf{x}, \tilde{\mathbf{y}}).$$

Because the inequality holds for any $\tilde{\mathbf{x}} \in X$,

$$\max_{\tilde{\mathbf{x}} \in X} \min_{\mathbf{y} \in Y} f(\tilde{\mathbf{x}}, \mathbf{y}) \leq \max_{\mathbf{x} \in X} f(\mathbf{x}, \tilde{\mathbf{y}}).$$

Minimizing over $\tilde{\mathbf{y}} \in Y$, we obtain (1.13).

To prove the lemma in the general case, we just need to verify the existence of the relevant maxima and minima. Since continuous functions achieve their minimum on compact sets, $g(\mathbf{x}) = \min_{\mathbf{y} \in Y} f(\mathbf{x}, \mathbf{y})$ is well-defined. The continuity of f and compactness of $X \times Y$ imply that f is uniformly continuous on $X \times Y$. In particular,

$$\forall \epsilon \ \exists \delta \ : \ |\mathbf{x}_1 - \mathbf{x}_2| < \delta \Longrightarrow |f(\mathbf{x}_1, \mathbf{y}) - f(\mathbf{x}_2, \mathbf{y})| \leq \epsilon$$

and hence $|g(\mathbf{x}_1) - g(\mathbf{x}_2)| \leq \epsilon$. Thus, $g : X \to \mathbb{R}$ is a continuous function and $\max_{\mathbf{x} \in X} g(\mathbf{x})$ exists. □

We can now prove:

Theorem 1.5.4 (Von Neumann's Minimax Theorem). *Let A be an $m \times n$ payoff matrix, and let $\Delta_m = \{\mathbf{x} \in \mathbb{R}^m : \mathbf{x} \geq \mathbf{0}, \sum_i x_i = 1\}$ and $\Delta_n = \{\mathbf{y} \in \mathbb{R}^n : \mathbf{y} \geq \mathbf{0}, \sum_j y_j = 1\}$. Then*

$$\max_{\mathbf{x}\in\Delta_m} \min_{\mathbf{y}\in\Delta_n} \mathbf{x}^T A\mathbf{y} = \min_{\mathbf{y}\in\Delta_n} \max_{\mathbf{x}\in\Delta_m} \mathbf{x}^T A\mathbf{y}.$$

As we discussed earlier quantity is called the ***zero-sum game!value*** *of the two-person zero-sum game with payoff matrix A.*

Proof. The inequality

$$\max_{\mathbf{x}\in\Delta_m} \min_{\mathbf{y}\in\Delta_n} \mathbf{x}^T A\mathbf{y} \leq \min_{\mathbf{y}\in\Delta_n} \max_{\mathbf{x}\in\Delta_m} \mathbf{x}^T A\mathbf{y}$$

follows immediately from the Lemma 1.5.3 because $f(\mathbf{x},\mathbf{y}) = \mathbf{x}^T A\mathbf{y}$ is a continuous function in both variables and $\Delta_m \subset \mathbb{R}^m$, $\Delta_n \subset \mathbb{R}^n$ are closed and bounded.

We will prove the other inequality by contradiction. Suppose that

$$\max_{\mathbf{x}\in\Delta_m} \min_{\mathbf{y}\in\Delta_n} \mathbf{x}^T A\mathbf{y} < \lambda < \min_{\mathbf{y}\in\Delta_n} \max_{\mathbf{x}\in\Delta_m} \mathbf{x}^T A\mathbf{y}.$$

Define a new game with payoff matrix $\hat{A}$ given by $\hat{a}_{i,j} = a_{ij} - \lambda$. For this new game, since each payoff in the matrix is reduced by λ, the expected payoffs for every pair of mixed strategies are also reduced by λ and hence:

$$\max_{\mathbf{x}\in\Delta_m} \min_{\mathbf{y}\in\Delta_n} \mathbf{x}^T \hat{A}\mathbf{y} < 0 < \min_{\mathbf{y}\in\Delta_n} \max_{\mathbf{x}\in\Delta_m} \mathbf{x}^T \hat{A}\mathbf{y}. \tag{1.14}$$

Each mixed strategy $\mathbf{y} \in \Delta_n$ for player II yields a gain vector $\hat{A}\mathbf{y} \in \mathbb{R}^m$. Let K denote the set of all vectors which dominate the gain vectors $\hat{A}\mathbf{y}$, that is,

$$K = \left\{\hat{A}\mathbf{y} + \mathbf{v} : \mathbf{y} \in \Delta_n,\ \mathbf{v} \in \mathbb{R}^m, \mathbf{v} \geq \mathbf{0}\right\}.$$

The set K is convex and closed: this follows from the fact that Δ_n, the set of probability vectors corresponding to mixed strategies $\mathbf{y}$ for player II, is closed, bounded and convex, and the set $\{\mathbf{v} \in \mathbb{R}^m, \mathbf{v} \geq \mathbf{0}\}$ is closed and convex. (See Exercise 1.19.) Also, K cannot contain the $\mathbf{0}$ vector, because if $\mathbf{0}$ was in K, there would be some mixed strategy $\mathbf{y} \in \Delta_n$ such that $\hat{A}\mathbf{y} \leq \mathbf{0}$. But this would imply that $\max_{\mathbf{x}\in\Delta_m} \mathbf{x}^T \hat{A}\mathbf{y} \leq 0$, contradicting the right-hand side of (1.14).

Thus K satisfies the conditions of the separating hyperplane theorem

(Theorem 1.5.2), which gives us $\mathbf{z} \in \mathbb{R}^m$ and $c > 0$ such that $\mathbf{z}^T\mathbf{w} > c > 0$ for all $\mathbf{w} \in K$. That is,

$$\mathbf{z}^T(\hat{A}\mathbf{y} + \mathbf{v}) > c > 0 \text{ for all } \mathbf{y} \in \Delta_n \text{ and } \mathbf{v} \geq \mathbf{0}. \tag{1.15}$$

We claim also that $\mathbf{z} \geq \mathbf{0}$. If not, say $z_j < 0$ for some j, then for $\mathbf{v} \in \mathbb{R}^m$ with v_j sufficiently large and $v_i = 0$ for all $i \neq j$, we would have $\mathbf{z}^T(\hat{A}\mathbf{y}+\mathbf{v}) = \mathbf{z}^T\hat{A}\mathbf{y} + z_jv_j < 0$ for some $\mathbf{y} \in \Delta_n$ which would contradict (1.15).

It also follows from (1.15) that not all of the z_i's can be zero. Thus $s = \sum_{i=1}^m z_i$ is strictly positive, so that $\tilde{\mathbf{x}} = \frac{1}{s}(z_1, \ldots, z_m)^T = \mathbf{z}/s \in \Delta_m$, with $\tilde{\mathbf{x}}^T\hat{A}\mathbf{y} > c/s > 0$ for all $\mathbf{y} \in \Delta_n$.

In other words, $\tilde{\mathbf{x}}$ is a mixed strategy for player I that gives a positive expected gain against any mixed strategy of player II. This contradicts the left hand inequality of (1.14). □

1.6 Zero-sum games with infinite action spaces*

Theorem 1.6.1. *Consider a zero-sum game in which the players' action spaces are* $[0,1]$ *and the gain is* $A(x,y)$ *when player I chooses action* x *and player II chooses action* y. *Suppose that* $A(x,y)$ *is continuous on* $[0,1]^2$. *Let* $\Delta = \Delta_{[0,1]}$ *be the space of probability distributions on* $[0,1]$. *Then*

$$\max_{F\in\Delta}\min_{G\in\Delta}\int\int A(x,y)dF(x)dG(y) = \min_{G\in\Delta}\max_{F\in\Delta}\int\int A(x,y)dF(x)dG(y) \tag{1.16}$$

Proof. If there is a matrix (a_{ij}) for which

$$A(x,y) = a_{\lceil nx\rceil,\lceil ny\rceil} \tag{1.17}$$

then (1.16) reduces to the finite case. If A is continuous, then there are functions A_0 and A_1 of the form (1.17) so that $A_0 \leq A \leq A_1$ and $|A_1 - A_0| \leq \epsilon$. This implies (1.16) with infs and sups in place of min and max. The existence of the maxima and minima follows from compactness of $\Delta_{[0,1]}$ as in the proof of Lemma 1.5.3. □

Remark. The previous theorem applies in any setting where the action spaces are compact metric spaces and the payoff function is continuous.

Exercise 1.6.2. Two players each choose a number in $[0,1]$. If they choose the same number, the payoff is 0. Otherwise, the player that chose the lower number pays \$1 to the player who chose the higher number, unless the higher number is 1, in which case the payment is reversed. Show that this game has no mixed Nash equilibrium. Show that the safety values for players I and II are -1 and 1 respectively.

Remark. The game from the previous exercise shows that the continuity assumption on the payoff function $A(x, y)$ cannot be removed. See also Exercise 1.22.

Notes

The theory of two-person zero-sum games was first laid out in a 1928 paper by John von Neumann [vN28], where he proved the minimax theorem (Theorem 1.5.4). The foundations were further developed in the book *The Theory of Games and Economic Behavior*, by Neumann and Morgenstern [vNM53], first published in 1944. The original proof of the minimax theorem used a fixed point theorem. A proof based on the separating hyperplane theorem (Theorem 1.5.2) was given by Weyl [Wey50], and an inductive proof was given by Owen [Owe67]. Subsequently, many other minimax theorems were proved, such as Theorem 1.6.1, due to Glicksberg [Gli52], and Sion's minimax theorem [Sio58]. An influential example of a zero-sum game on the unit square with discontinuous payoff functions and without a value is in [**?**]. An important class of continuous games that are not discussed in the text are games of timing. See e.g., [Gar00].

More detailed accounts of this material in this chapter can be found in Ferguson [Fer08], Karlin [Kar59] and Owen [Owe95], among others.

In §1.3, we present techniques for simplifying and solving zero-sum games by hand. However, for large games, there are efficient algorithms for finding optimal strategies and the value of the game based on linear programming. See e.g., [MG07] for an introduction to linear programming.

Exercise 1.2 is from [Kar59]. Exercise 1.17 comes from [HS89]. Exercise 1.18 is an example of a class of recursive games studied in [Eve57].

Exercises

1.1 Show that all saddle points in a zero-sum game (assuming there is at least one) result in the same payoff to player I.

1.2 Show that if a zero-sum game has a saddle point in every two by two submatrix, then it has a saddle point.

1.3 Find the value of the following zero-sum game and determine some optimal strategies for each of the players.

$$\begin{pmatrix} 8 & 3 & 4 & 1 \\ 4 & 7 & 1 & 6 \\ 0 & 3 & 8 & 5 \end{pmatrix}$$

1.4 Find the value of the zero-sum game given by the following payoff matrix, and determine some optimal strategies for each of the players.

$$\begin{pmatrix} 0 & 9 & 1 & 1 \\ 5 & 0 & 6 & 7 \\ 2 & 4 & 3 & 3 \end{pmatrix}$$

1.5 Find the value of the zero-sum game given by the following payoff matrix and determine *all* optimal strategies for both players.

$$\begin{pmatrix} 3 & 0 \\ 0 & 3 \\ 2 & 2 \end{pmatrix}$$

1.6 Given a 5 by 5 zero-sum game, such as the following, how would you quickly determine by hand if it has a saddle point?

$$\begin{pmatrix} 20 & 1 & 4 & 3 & 1 \\ 2 & 3 & 8 & 4 & 4 \\ 10 & 8 & 7 & 6 & 9 \\ 5 & 6 & 1 & 2 & 2 \\ 3 & 7 & 9 & 1 & 5 \end{pmatrix}$$

1.7 Give an example of a two-player zero-sum game where there are no pure Nash equilibria. Can you give an example where all the entries of the payoff matrix are different?

1.8 Define a zero-sum game in which one player's optimal strategy is pure and the other player's optimal strategy is mixed.

1.9 Player II is moving an important item in one of three cars, labeled 1, 2, and 3. Player I will drop a bomb on one of the cars of his choosing. He has no chance of destroying the item if he bombs the wrong car. If he chooses the right car, then his probability of destroying the item depends on that car. The probabilities for cars 1, 2, and 3 are equal to 3/4, 1/4, and 1/2.

Write the 3×3 payoff matrix for the game, and find some optimal winning strategies for each of the players.

1.10 Using the result of Corollary 1.4.2 give an exponential time algorithm to solve an n by m two-person zero-sum game. Hint: Consider each possibility for which subset S of player I strategies have $x_i > 0$ and which subset of player II strategies T have $y_j > 0$.

1.11 Consider the following two-person zero-sum game. Both players simultaneously call out one of the numbers $\{2, 3\}$. Player 1 wins if the sum of the numbers called is odd and player 2 wins if their sum is even. The loser pays the winner the product of the two numbers called (in dollars). Find the payoff matrix, the value of the game, and an optimal strategy for each player.

1.12 Consider the four mile stretch of road shown in Figure **??**. There are three locations at which restaurants can be opened: Left, Central, and Right. Company I opens a restaurant at one of these locations and company II opens two restaurants (both restaurants can be at the same location). A customer is located at a uniformly random location along the four mile stretch. He walks to the closest location at which there is a restaurant, and then into one of the restaurants there, chosen uniformly at random. The payoff to company I is the probability that the customer visits a company I restaurant.

Determine the value of the game, and find some optimal mixed strategies for the companies.

1.13 Bob has a concession at Yankee Stadium. He can sell 500 umbrellas at \$10 each if it rains. (The umbrellas cost him \$5 each.) If it shines, he can sell only 100 umbrellas at \$10 each and 1000 sunglasses at \$5 each. (The sunglasses cost him \$2 each.) He has \$2500 to invest in one day, but everything that isn't sold is trampled by the fans and is a total loss.

This is a game against nature. Nature has two strategies: rain and shine. Bob also has two strategies: buy for rain or buy for shine.

Find the optimal strategy for Bob assuming that the probability for rain is 50%.

1.14 The number picking game: Two players I and II pick a positive integer each. If the two numbers are the same, no money changes hands. If the players' choices differ by 1 the player with the lower number pays \$1 to the opponent. If the difference is at least 2 the player with the higher number pays \$2 to the opponent. Find the value of this zero-sum game and determine optimal strategies for both players. (Hint: use domination.)

1.15 Show that in *Submarine Salvo* the submarine has an optimal strategy where all choices containing a corner and a clockwise adjacent site are excluded.

1.16 A zebra has four possible locations to cross the Zambezi river, call them a, b, c, and d, arranged from north to south. A crocodile can wait (undetected) at one of these locations. If the zebra and the crocodile choose the same location, the payoff to the crocodile (that is, the chance it will catch the zebra) is 1. The payoff to the crocodile is 1/2 if they choose adjacent locations, and 0 in the remaining cases, when the locations chosen are distinct and non-adjacent.

(a) Write the payoff matrix for this game.
(b) Can you reduce this game to a 2×2 game?
(c) Find the value of the game (to the crocodile) and optimal strategies for both.

1.17 **Generalized Matching Pennies** Consider a directed graph $G = (V, E)$ with nonnegative weights w_{ij} on each edge (i, j). Let $W_i = \sum_j w_{ij}$. Each player chooses a vertex, say i for player I and j for player II. Player I receives a payoff of w_{ij} if $i \neq j$, and loses $W_i - w_{ii}$ if $i = j$. Thus, the payoff matrix A has entries $a_{ij} = w_{ij} - 1_{\{i=j\}} W_i$. If $n = 2$ and the w_{ij}'s are all 1, this game is called Matching Pennies.

- Show that the game has value 0.
- Deduce that for some $x \in \Delta_n$, $\mathbf{x}^T A = 0$.

1.18 **A recursive zero-sum game.** An inspector can inspect a facility on just one occasion, on one of the days $1, \ldots, n$. The worker at the facility can cheat or be honest on any given day. The payoff to the inspector is 1 if he inspects while the worker is cheating. The payoff is -1 if the worker cheats and is not caught. The payoff is also -1 if the inspector inspects but the worker did not cheat, and there is at least one day left. This leads to the following matrices Γ_n for the game with n days: the matrix Γ_1 is shown on the left, and the matrix Γ_n is shown on the right.

		worker	
		cheat	honest
inspector	inspect	1	0
	wait	-1	0

		worker	
		cheat	honest
inspector	inspect	1	-1
	wait	-1	Γ_{n-1}

Find the optimal strategies and the value of Γ_n.

1.19 Prove that if set $G \subseteq \mathbb{R}^d$ is compact and $H \subseteq \mathbb{R}^d$ is closed, then $G + H$ is closed. (This fact is used in the proof of the minimax theorem to show that that the set K is closed.)

1.20 Find two sets $F_1, F_2 \subset \mathbb{R}^2$ that are closed such that $F_1 - F_2$ is not closed.

1.21 * Consider a zero-sum game A and suppose that π and σ are permutations of I's strategies $\{1, \ldots, m\}$ and player II's strategies $\{1, \ldots, n\}$ respectively such that

$$a_{\pi(i)\sigma(j)} = a_{ij} \tag{E1.1}$$

for all i and j. Show that there exist optimal strategies $\mathbf{x}^*$ and $\mathbf{y}^*$ such that $x_i^* = x_{\pi(i)}^*$ for all i and $y_j^* = y_{\sigma(j)}^*$ for all j.

1.22 Two players each choose a positive integer. The player that chose the lower number pays \$1 to the player who chose the higher number (with no payment in case of a tie). Show that this game has no Nash equilibrium. Show that the safety values for players I and II are -1 and 1 respectively.

1.23 Two players each choose a number in $[0, 1]$. Suppose that $A(x, y) = |x - y|$.

- Show that the value of the game is $1/2$.
- More generally, suppose that $A(x, y)$ is a convex function in each of x and y, and continuous. Show that player I has an optimal strategy supported on 2 points and player II has an optimal pure strategy.

1.24 Consider a zero-sum game in which the strategy spaces are $[-1, 1]$, and the gain of player I when she plays x and player II plays y is

$$A(x, y) = \log \frac{1}{|x - y|}.$$

Show that I picking $X = \cos \Theta$, where Θ is Uniform on $[-1, 1]$, and II using the same strategy is a pair of optimal strategies.

2

Adaptive decision making

Suppose that two players are playing multiple rounds of the same game. How would they adapt their strategies to the outcomes of previous rounds? This fits into the broader framework of adaptive decision making which we develop next and later apply to games. In particular, we'll see an alternative proof of the Minimax Theorem (see Theorem 2.4.2). We start with a very simple setting.

2.1 Binary prediction with expert advice and a perfect expert

Example 2.1.1. [Predicting the Stock Market] Consider a trader trying to predict whether the stock market will go up or down each day. Each morning, for T days, he solicits the opinions of n experts, who each make up/down predictions. Based on their predictions, the trader makes a choice between up and down, and buys or sells accordingly.

In this section, we assume that at least one of the experts is perfect, that is, predicts correctly every day, but the trader doesn't know which one it is. What should the trader do to minimize the number of mistakes he makes in T days?

First approach – Follow the Majority of Leaders: On any day, call the experts who have never made a mistake *leaders*. By following the majority opinion among the leaders, the trader is guaranteed never to make more than $\log_2 n$ mistakes: Each mistake the trader makes eliminates at least half of leaders and, obviously, never eliminates the perfect expert.

Second approach – Follow a Random Leader (FRL): Perhaps surprisingly, following a random leader yields a slightly better guarantee: For any n, the number of mistakes made by the trader is at most $\mathrm{H}_n - 1$ in expectation.† We verify this by induction on the number of leaders. The case of a single leader is clear. Consider the first day on which some number of experts, say $k > 0$, make a mistake. Then by the induction hypothesis, the expected number of mistakes the trader ever makes is at most

$$\frac{k}{n} + \mathrm{H}_{n-k} - 1 \leq \mathrm{H}_n - 1.$$

This analysis is tight for $T \geq n$. Suppose that for $1 \leq i < n$, on day i, only expert i makes a mistake. Then the probability that the trader makes a mistake that day is $1/(n - i + 1)$. Thus, the expected number of mistakes he makes is $\mathrm{H}_n - 1$.

Remark. We can think of this setting as an extensive-form zero-sum game between an adversary and a trader. The adversary chooses the daily advice of the experts and the actual outcome on each day t, and the trader chooses a prediction each day based on the experts' advice. In this game, the adversary seeks to maximize his gain, the number of mistakes the trader makes.

Next we derive a lower bound on the expected number of mistakes made by any trader algorithm, by presenting a strategy for the adversary.

Proposition 2.1.2. *In the setting of Example 2.1.1 with at least one perfect expert, there is an adversary strategy that causes any trader algorithm to incur at least* $\lfloor \log_4 n \rfloor$ *mistakes in expectation.*

Proof. Let $2^k \leq n < 2^{k+1}$, Define E_0 to be the first 2^k experts, and let E_t be the experts in E_{t-1} that predicted correctly on day t.

Now suppose that on day t, for $1 \leq t \leq k$, half of the experts in E_{t-1} predict up and half predict down, and the rest of the experts predict down. Suppose also that the truth is equally likely to be up or down. Then no matter how the trader chooses up or down, with probability $1/2$, he makes a mistake. Thus, in the first k days, the algorithm makes $k/2$ mistakes in expectation. In other words, the expected number of mistakes is at least $\lfloor \log_2 n \rfloor / 2 \geq \lfloor \log_4 n \rfloor$.

□

To prove a matching upper bound, we will take a middle road between following the majority of the leaders (ignoring the minority) and FRL (which weights the minority too highly).

---

† Recall that $\mathrm{H}_n = \sum_{i=1}^{n} \frac{1}{i} \in (\ln n, \ln n + 1)$.

Third approach – Boosted Majority of Leaders: Given any function $p : [1/2, 1] \to [1/2, 1]$, consider the trader algorithm A_p: When the experts are split on their advice in proportion $(x, 1-x)$ with $x \geq 1/2$, follow the majority with probability $p(x)$.

If $p(x) = 1$ for all $x > 1/2$ we get the deterministic majority vote, while if $p(x) = x$ we get FRL.

For which $a > 1$ can we prove an upper bound of $\log_a n$ on the expected number of mistakes? To do so, by induction, we need to verify two inequalities for all $x \in [1/2, 1]$:

$$\log_a(nx) + 1 - p(x) \leq \log_a n \tag{2.1}$$

$$\log_a(n(1-x)) + p(x) \leq \log_a n \tag{2.2}$$

The LHS of (2.1) is an upper bound on the expected mistakes of A_p assuming the majority is right (using the induction hypothesis) and the LHS of (2.2) is an upper bound on the expected mistakes of A_p assuming the minority is right.

Adding these inequalities and setting $x = 1/2$ gives $2\log_a(1/2) + 1 \leq 0$, that is, $a \leq 4$. We already know this, since $\lfloor \log_4 n \rfloor$is a lower bound for the worst case performance. Setting $a = 4$, the two required inequalities become

$$p(x) \geq 1 + \log_4 x$$

$$p(x) \leq -\log_4(1-x)$$

We can easily satisfy both of these inequalities, e.g. by taking $p(x) = 1 + \log_4 x$, since $x(1-x) \leq 1/4$.

2.2 Nobody's perfect

Unfortunately, the assumption that there is a perfect expert is unrealistic. In the setting of Example 2.1.1, let L_i^t be the *cumulative loss* (i.e., total number of mistakes) incurred by expert i on the first t days. Denote

$$L_*^t = \min_i L_i^t \quad \text{and} \quad S_j = \{t \mid L_*^t = j\}.$$

Suppose that, for each t, on day $t+1$ the trader follows the majority opinion of the *leaders*, i.e., those experts with $L_i^t = L_*^t$. Then during S_j, by the discussion of the case where there is a perfect expert, his loss is at most $\log_2 n + 1$. Thus, for any number T of days, the trader's loss is bounded by $(\log_2 n + 1)(L_*^T + 1)$. Similarly, the expected loss of FRL is at most $\mathrm{H}_n(L_*^T + 1)$ and the expected loss of Boosted Majority is at most $(\log_4 n + 1)(L_*^T + 1)$.

Remark. There is an adversary strategy that ensures any trader algorithm that only uses the advice of the leading experts will incur an expected loss that is at least $\lfloor \log_4(n) \rfloor L_*^T$ in T steps. See Exercise 2.1.

2.2.1 Weighted Majority

There are strategies that guarantee the trader an asymptotic loss that is at most twice that of the best expert. One such strategy is based on weighted majority, where the weight assigned to an expert is decreased by a factor $1-\epsilon$ each time he makes a mistake.

Weighted Majority Algorithm

Fix $\epsilon \in [0,1]$. On each day t, associate a weight w_i^t with each expert i. Initially, set $w_i^0 = 1$ for all i.

Each day t, follow the ***weighted majority*** opinion: Let U_t be the set of experts predicting up on day t, and D_t the set predicting down. Predict "up" on day t if $W_U(t-1) = \sum_{i \in U_t} w_i^{t-1} \geq W_D(t-1) = \sum_{i \in D_t} w_i^{t-1}$ and "down" otherwise.

At the end of day t, for each i such that expert i predicted incorrectly on day t, set

$$w_i^t = (1-\epsilon) w_i^{t-1} \tag{2.3}$$

Thus, $w_i^t = (1-\epsilon)^{L_i^t}$.

For the analysis of this algorithm, we will use the following facts:

Lemma 2.2.1. *Let* $\epsilon \in [0, 1/2]$. *Then* $\epsilon \leq -\ln(1-\epsilon) \leq \epsilon + \epsilon^2$.

Proof. Taylor expansion gives

$$-\ln(1-\epsilon) = \sum_{k \geq 1} \frac{\epsilon^k}{k} \geq \epsilon.$$

On the other hand,

$$\sum_{k \geq 1} \frac{\epsilon^k}{k} \leq \epsilon + \frac{\epsilon^2}{2} \sum_{k=0}^{\infty} \epsilon^k \leq \epsilon + \epsilon^2$$

since $\epsilon \leq 1/2$.

□

Theorem 2.2.2. *Suppose there are* n *experts. Let* $L(T)$ *be the number of mistakes made by the Weighted Majority Algorithm in* T *steps with* $\epsilon \leq \frac{1}{2}$, *and let L_i^T be the number of mistakes made by expert i in T steps. Then for any sequence of up/down outcomes and for* every *expert i, we have*

$$L(T) \leq 2(1+\epsilon)L_i^T + \frac{2\ln n}{\epsilon}. \tag{2.4}$$

Proof. Let $W(t) = \sum_i w_i^t$ be the total weight on all the experts after the t^{th} day. If the algorithm incurs a loss on the t^{th} day, say by predicting up instead of correctly predicting down, then $W_U(t-1) \geq \frac{1}{2}W(t-1)$. But in that case

$$W(t) \leq W_D(t-1) + (1-\epsilon)W_U(t-1) \leq \left(1 - \frac{\epsilon}{2}\right) W(t-1).$$

Thus, after a total loss of $L = L(T)$,

$$W(T) \leq \left(1 - \frac{\epsilon}{2}\right)^L W(0) = \left(1 - \frac{\epsilon}{2}\right)^L n.$$

Now consider expert i who incurs a loss of $L_i = L_i^T$. His weight at the end is

$$w_i^T = (1-\epsilon)^{L_i},$$

which is at most $W(T)$. Thus

$$(1-\epsilon)^{L_i} \leq \left(1 - \frac{\epsilon}{2}\right)^L n.$$

Taking logs and negating, we have

$$-L_i \ln(1-\epsilon) \geq -L \ln\left(1 - \frac{\epsilon}{2}\right) - \ln n. \tag{2.5}$$

Applying Lemma 2.2.1, we obtain that for $\epsilon \in (0, 1/2]$,

$$L_i(\epsilon + \epsilon^2) \geq L\frac{\epsilon}{2} - \ln n$$

or

$$L(T) \leq 2(1+\epsilon)L_i^T + \frac{2\ln n}{\epsilon}.$$

□

Remarks:

1. It follows from (2.5) that for all $\epsilon \in (0, 1]$

$$L(T) \leq \frac{|\ln(1-\epsilon)|L_i^T + \ln n}{|\ln(1 - \frac{\epsilon}{2})|}. \tag{2.6}$$

If we know in advance that there is an expert with $L_i^T = 0$, then letting $\epsilon \uparrow 1$ recovers the result of the first approach to Example 2.1.1.

2. There are cases where the Weighted Majority Algorithm incurs at least twice the loss of the best expert. In fact, this holds for every deterministic algorithm. See Exercise 2.2.

2.3 Multiple choices and varying costs

"I hear the voices, and I read the front page, and I know the speculation. But I'm the decider, and I decide what is best."
George W. Bush

In the previous section, the decision maker used the advice of n experts to choose between two options, and the cost of any mistake was the same. We saw that a simple deterministic algorithm could guarantee that the number of mistakes was not much more than twice that of any expert. One drawback of the Weighted Majority algorithm is that it treats a majority of 51% with the same reverence as a majority of 99%. With careful randomization, we can avoid this pitfall, and show that the decision maker can do almost as well as the best expert.

In this section, the decider faces multiple options, e.g., which stock to buy, rather than just up or down, now with varying losses. We will refer to the options of the decider as *actions*: This covers the task of prediction with expert advice, as the i^{th} action could be "follow the advice of expert i".

Example 2.3.1 (**Route-picking**). Each day you choose one of a set of n routes from your house to work. Your goal is to minimize the time it takes to get to work. However, you do not know ahead of time how much traffic, and hence how long each route will take. Once you choose your route, you incur a loss equal to the latency on the route you selected. This continues for T days. Let L_i^T be the total latency you would have incurred over the T days if you had taken the same route every day, say i, for some $1 \leq i \leq n$. Can we find a strategy for choosing a route each day such that the total latency incurred is not too much more than $\min_i L_i^T$?

The following setup captures the stock-market and route-picking examples above and many others.

Definition 2.3.2 (**Sequential adaptive decision making**). On day t, a decider $\mathcal{D}$ chooses a probability distribution $\mathbf{p}^t = (p_1^t, \ldots, p_n^t)$ over a set of n actions, e.g., stocks to own or routes to drive. (The choice of $\mathbf{p}^t$ can depend on the history, i.e., the prior losses of each action and prior actions taken by $\mathcal{D}$.) The losses $\boldsymbol{\ell}^t = (\ell_1^t, \ldots, \ell_n^t) \in [0, 1]^n$ of each action on day t are then revealed.

Given the history, $\mathcal{D}$'s expected loss on day t is $\mathbf{p}^t \cdot \boldsymbol{\ell}^t = \sum_{i=1}^n p_i^t \ell_i^t$. The total expected loss $\mathcal{D}$ incurs in T days is

$$L_{\mathcal{D}}^T = \sum_{t=1}^T \mathbf{p}^t \cdot \boldsymbol{\ell}^t.$$

(See the chapter notes for a more precise interpretation of $L_{\mathcal{D}}^T$ in the case where losses depend on prior actions taken by $\mathcal{D}$.)

Remark. In stock-picking examples, $\mathcal{D}$ could have a fraction p_i^t of his portfolio in stock i instead of randomizing.

Definition 2.3.3. The *regret of a decider $\mathcal{D}$ in T steps against loss sequence* $L = \{\boldsymbol{\ell}^t\}_{t=1}^T$ is defined as the difference between the total expected loss of the decider and the total loss of the best single action, that is

$$\mathcal{R}_T(\mathcal{D}, L) := L_{\mathcal{D}}^T - \min_i L_i^T,$$

where $L_i^T = \sum_{t=1}^T \ell_i^t$.

We define the **regret of a decider $\mathcal{D}$** as

$$\mathcal{R}_T(\mathcal{D}) := \max_L \mathcal{R}_T(\mathcal{D}, L). \tag{2.7}$$

Perhaps surprisingly, there exist algorithms with regret that is **sublinear in** T, i.e., the average regret per day tends to 0. We will see one below.

Discussion

Let $L = \{\ell_1^t\}_{t=1}^T$ be a sequence of loss vectors. A natural goal for a decision-making algorithm $\mathcal{D}$ is to minimize its worst case loss, i.e., $\max_L L_{\mathcal{D}}^T$. But this is a dubious measure of the quality of the algorithm, since on a worst-case sequence, there may be nothing any decider can do. This motivates evaluating $\mathcal{D}$ by its performance gap

$$\max_L (L_{\mathcal{D}}^T - \mathcal{B}(L)),$$

where $\mathcal{B}(L)$ is a benchmark loss for L. The most obvious choice for the benchmark is $\mathcal{B}^* = \sum_{t=1}^T \min \ell_i^t$, but this is too ambitious: E.g., if $n = 2$, $\{\ell_1^t\}_{t=1}^T$ are independent unbiased bits, and $\ell_2^t = 1 - \ell_1^t$, then $\mathbb{E}\left[L_{\mathcal{D}}^T - \mathcal{B}^*\right] = T/2$ since $\mathcal{B}^* = 0$. Instead, in the definition of regret, we employ the benchmark $\mathcal{B}(L) = \min_i L_i^T$. At first sight, this benchmark looks weak, since why should choosing the same action every day be a reasonable option? We give two answers: (1) Often there really is a better action and the goal of the decision algorithm is to learn its identity without losing too much in the process. (2) Alternative decision algorithms (e.g., use action 1 on odd days, action 2 on even days, except if one action has more than double the cumulative loss of the other), can be considered experts and incorporated into the model as additional actions. We show below that the regret of any decision algorithm is at most $\sqrt{T \log n/2}$ when there are n actions to choose from. Note that this grows linearly in T if n is exponential in T. To see that this is unavoidable, recall that in the binary prediction setting, if we include 2^T experts making all possible predictions, one of them will make no mistakes, and we already know that for this case, any decision algorithm will incur worst-case regret at least $T/2$.

2.3.1 The Multiplicative Weights Algorithm

We now present an algorithm for adaptive decision making, with regret that is $o(T)$ as $T \to \infty$. The algorithm is a randomized variant of the Weighted Majority Algorithm; it uses the weights in that algorithm as probabilities. The algorithm and its analysis in Theorem 2.3.6 deal with the case where the decider incurs losses only.

Multiplicative Weights algorithm (MW)

Fix $\epsilon < 1/2$ and n possible actions.
On each day t, associate a weight w_i^t with the i^{th} action.
Initially, $w_i^0 = 1$ for all i.
On day t, use the mixed strategy $\mathbf{p}^t$, where

$$p_i^t = \frac{w_i^{t-1}}{\sum_k w_k^{t-1}}.$$

For each action i, with $1 \leq i \leq n$, observe the loss $\ell_i^t \in [0,1]$ and update the weight w_i^t as follows:

$$w_i^t = w_i^{t-1} \exp(-\epsilon \ell_i^t). \tag{2.8}$$

In the next proof we will use the following lemma.

Lemma 2.3.4 (Hoeffding Lemma). *Suppose that X is a random variable with distribution F such that $a \leq X \leq a+1$, for some $a \leq 0$, and $\mathbb{E}[X] = 0$. Then for any λ,*

$$\mathbb{E}\left[e^{\lambda X}\right] \leq e^{\lambda^2/8}.$$

For a proof, see Appendix A1.2.1. For the reader's convenience, we prove here the following slightly weaker version.

Lemma 2.3.5. *Let X be a random variable with $\mathbb{E}[X] = 0$ and $|X| \leq 1$. Then*

$$\mathbb{E}\left[e^{\lambda X}\right] \leq e^{\lambda^2/2}.$$

Proof. By convexity of the function $f(x) = e^{\lambda x}$, we have

$$e^{\lambda x} \leq \frac{(1+x)e^{\lambda} + (1-x)e^{-\lambda}}{2} = \ell(x)$$

for $x \in [-1, 1]$. Thus, since $|X| \leq 1$ and $\mathbb{E}[X] = 0$, we have

$$\begin{aligned}\mathbb{E}\left[e^{\lambda X}\right] \leq \mathbb{E}[\ell(X)] = \frac{e^{\lambda} + e^{-\lambda}}{2} &= \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k)!} \\ &\leq \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{2^k k!} = e^{\lambda^2/2}.\end{aligned}$$

□

Theorem 2.3.6. *Consider the Multiplicative Weights algorithm with n actions. Define*

$$L_{MW}^T := \sum_{t=1}^{T} \mathbf{p}^t \cdot \boldsymbol{\ell}^t,$$

where $\boldsymbol{\ell}^t \in [0,1]^n$. Then, for every loss sequence $\{\boldsymbol{\ell}^t\}_{t=1}^T$, and every action i, we have

$$L_{MW}^T \leq L_i^T + \frac{T\epsilon}{8} + \frac{\log n}{\epsilon},$$

where $L_i^T = \sum_{t=1}^T \ell_i^t$. In particular, taking $\epsilon = \sqrt{\frac{8 \log n}{T}}$, we obtain that for all i,

$$L_{MW}^T \leq L_i^T + \sqrt{\frac{1}{2} T \log n},$$

i.e., the regret of MW in T steps is at most $\sqrt{\frac{1}{2} T \log n}$.

Proof. Let $W^t = \sum_{1 \leq i \leq n} w_i^t = \sum_{1 \leq i \leq n} w_i^{t-1} \exp(-\epsilon \ell_i^t)$. Then

$$\frac{W^t}{W^{t-1}} = \sum_i \frac{w_i^{t-1}}{W^{t-1}} \exp(-\epsilon \ell_i^t) = \sum_i p_i^t \exp(-\epsilon \ell_i^t) = \mathbb{E}\left[e^{-\epsilon X_t}\right], \tag{2.9}$$

where X_t is the loss the algorithm incurs at time t, i.e.,

$$\mathbb{P}\left[X_t = \ell_i^t\right] = p_i^t.$$

Let

$$\overline{\ell}^t := \mathbb{E}\left[X_t\right] = \mathbf{p}^t \cdot \boldsymbol{\ell}^t.$$

By Hoeffding's Lemma (Lemma 2.3.4), we have

$$\mathbb{E}\left[e^{-\epsilon X_t}\right] = e^{-\epsilon\overline{\ell}^t}\mathbb{E}\left[e^{-\epsilon(X_t-\overline{\ell}^t)}\right] \le e^{-\epsilon\overline{\ell}^t}e^{\epsilon^2/8}.$$

so plugging back into (2.9), we obtain

$$W^t \le e^{-\epsilon\overline{\ell}^t}e^{\epsilon^2/8}\,W^{t-1} \quad \text{and thus} \quad W^T \le e^{-\epsilon L^T}e^{T\epsilon^2/8}n.$$

On the other hand,

$$W^T \ge w_i^T = e^{-\epsilon L_i^T},$$

so combining these two inequalities, we obtain

$$e^{-\epsilon L_i^T} \le e^{-\epsilon L^T}e^{T\epsilon^2/8}n.$$

Taking logs, we obtain

$$L_{MW}^T \le L_i^T + \frac{T\epsilon}{8} + \frac{\log n}{\epsilon}.$$

□

The bound of Theorem 2.3.6 is optimal as T and n go to infinity.

Proposition 2.3.7. *Consider a loss sequence L in which all losses are independent and equally likely to be 0 or 1. Then for any algorithm $\mathcal{D}$,*

$$\mathcal{R}_T(\mathcal{D}, L) = \frac{1}{2}\sqrt{T\gamma_n}\cdot(1+o(1)) \qquad \text{as } T \to \infty \tag{2.10}$$

where

$$\gamma_n = \mathbb{E}\left[\max_{1\le i\le n} Y_i\right] \qquad \text{and} \qquad Y_i \sim N(0,1).$$

Moreover,

$$\gamma_n = \sqrt{2\log n}\,(1+o(1)) \qquad \text{as} \qquad n \to \infty. \tag{2.11}$$

Proof. Action i's loss, L_i^T, is binomial with parameters T and $1/2$, and thus by the Central Limit Theorem $\frac{L_i^T - T/2}{\sqrt{T/4}}$ converges in law to a normal (0,1) random variable Y_i. Let $L_*^T = \min_i L_i^T$. Then as $T \to \infty$,

$$\frac{\mathbb{E}\left[L_*^T - T/2\right]}{\sqrt{T/4}} \to \mathbb{E}\left[\min_{1 \le i \le n} Y_i\right] = -\gamma_n,$$

which proves (2.10). See Exercise 2.6 for the proof of (2.11). □

2.4 Using adaptive decision making to play zero-sum games

Consider a two-person zero-sum game with payoff matrix $A = \{a_{ij}\}$. Suppose T rounds of this game are played. We can apply the Multiplicative Weights Update (MW) algorithm to the decision-making process of player II. In round t, he chooses a mixed strategy $\mathbf{p}^t$, i.e., column j is assigned probability p_j^t. Knowing $\mathbf{p}^t$ and the history of play, player I chooses a row i_t. The loss of action j in round t is $\ell_j^t = a_{i_t j}$, and the total loss of action j in T rounds is $L_j^T = \sum_{t=1}^T a_{i_t j}$.

The following proposition bounds the total loss $L_{MW}^T = \sum_{t=1}^T (A\mathbf{p}^t)_{i_t}$ of player II.

Proposition 2.4.1. *Suppose the $m \times n$ payoff matrix $A = \{a_{ij}\}$ has entries in $[0,1]$ and player II is playing according to the MW algorithm. Let $\mathbf{x}_{emp}^T \in \Delta_m$ be a row vector representing the empirical distribution of actions taken by player I in T steps, i.e. the i^{th} coordinate of $\mathbf{x}_{emp}^T$ is $\frac{|\{t \mid i_t = i\}|}{T}$. Then the total loss satisfies*

$$L_{MW}^T \le T \min_{\mathbf{y}} \mathbf{x}_{emp}^T A\mathbf{y} + \sqrt{\frac{T \log n}{2}}.$$

Proof. It follows from Corollary **??** that player II's loss over the T rounds satisfies

$$L_{MW}^T \le L_j^T + \sqrt{\frac{T \log n}{2}}.$$

The proposition then follows from the fact that

$$\min_j L_j^T = T \min_j (\mathbf{x}_{emp}^T A)_j = T \min_{\mathbf{y}} \mathbf{x}_{emp}^T A\mathbf{y}.$$

□

Remark. Suppose player I uses the mixed strategy ξ (a row vector) in all T

rounds. If player II knows this, he can guarantee an expected loss of

$$\min_{\mathbf{y}} \xi A\mathbf{y},$$

which could be lower than v, the value of the game. In this case, $E(\mathbf{x}_{emp}^T) = \xi$, so even with no knowledge of ξ, the proposition bounds player II's expected loss by

$$T \min_{\mathbf{y}} \xi A\mathbf{y} + \sqrt{\frac{T \log n}{2}}.$$

Next, as promised, we rederive the Minimax Theorem as a corollary of Proposition 2.4.1.

Theorem 2.4.2 (Minimax Theorem)**.** *Let $A = \{a_{ij}\}$ be the payoff matrix of a zero-sum game. Let*

$$v_I = \max_{\mathbf{x}} \min_{\mathbf{y}} \mathbf{x}^T A\mathbf{y} = \max_{\mathbf{x}} \min_{j} (\mathbf{x}^T A)_j$$

and

$$v_{II} = \min_{\mathbf{y}} \max_{\mathbf{x}} \mathbf{x}^T A\mathbf{y} = \min_{\mathbf{y}} \max_{i} (A\mathbf{y})_i$$

be the safety values of the players. Then $v_I = v_{II}$.

Proof. By adding a constant to all entries of the matrix and scaling, we may assume that all entries of A are in $[0, 1]$.

From Lemma 1.5.3, we have $v_{\text{I}} \leq v_{\text{II}}$.

Suppose that, in round t, player II plays the mixed strategy $\mathbf{p}^t$ given by the MW algorithm, and player I plays a best response, i.e.

$$i_t = \text{argmax}_i (A\mathbf{p}^t)_i.$$

Then

$$\overline{\ell}^t = \max_i (A\mathbf{p}^t)_i \geq \min_{\mathbf{y}} \max_i (A\mathbf{y})_i = v_{\text{II}},$$

whence

$$L_{MW}^T \geq T v_{\text{II}}. \tag{2.12}$$

(Note that the proof of (2.12) did not rely on any property of the MW algorithm.) On the other hand, from Proposition 2.4.1, we have

$$L_{MW}^T \leq T \min_{\mathbf{y}} \mathbf{x}_{emp}^T A\mathbf{y} + \sqrt{\frac{T \log n}{2}},$$

and since $\min_{\mathbf{y}} \mathbf{x}_{emp}^T A\mathbf{y} \le v_{\mathrm{I}}$, we obtain

$$Tv_{\mathrm{II}} \le Tv_{\mathrm{I}} + \sqrt{\frac{T\log n}{2}},$$

and hence $v_{\mathrm{II}} \le v_I$.

□

2.5 Adaptive decision-making as a zero-sum game

Our goal in this section is to characterize the minimax regret in the setting of Definition 2.3.2, i.e.,

$$\min_{\mathcal{D}_{[0,1]}} \max_{\{\boldsymbol{\ell}^t\}} \mathcal{R}_T(\mathcal{D}_{[0,1]}, \{\boldsymbol{\ell}^t\}) \tag{2.13}$$

as the value of a finite zero-sum game between a decider and an adversary. In (2.13), the sequence of loss vectors $\{\boldsymbol{\ell}^t\}_{t=1}^T$ is in $[0,1]^{nT}$, and $\mathcal{D}_{[0,1]}$ is a sequence of functions $\{\mathbf{p}^t\}_{t=1}^T$, where

$$\mathbf{p}^t : [0,1]^{n(t-1)} \to \Delta_n$$

maps the losses from previous rounds to the decider's current mixed strategy over actions.

2.5.1 Minimax regret is attained in {0,1} losses

Given $\{\ell_i^t : 1 \le i \le n, 1 \le t \le T\}$, denote by $\{\hat{\ell}_i^t\}$ the sequence of independent $\{0,1\}$-valued random variables with

$$\mathbb{E}\left[\hat{\ell}_i^t\right] = \ell_i^t.$$

Theorem 2.5.1 ("Replacing losses by coin tosses"). *For any decision strategy $\mathcal{D}$ that is defined only for $\{0,1\}$ losses, a corresponding decision strategy $\mathcal{D}_{[0,1]}$ is defined as follows: For each t, given $\{\boldsymbol{\ell}^j\}_{j=1}^{t-1}$, applying $\mathcal{D}$ to $\{\hat{\boldsymbol{\ell}}^j\}_{j=1}^{t-1}$ yields $\hat{\mathbf{p}}^t$. Use $\mathbf{p}^t = \mathbb{E}\left[\hat{\mathbf{p}}^t\right]$ at time t in $\mathcal{D}_{[0,1]}$. Then*

$$\mathbb{E}\left[\mathcal{R}_T(\mathcal{D}, \{\hat{\boldsymbol{\ell}}^t\})\right] \ge \mathcal{R}_T(\mathcal{D}_{[0,1]}, \{\boldsymbol{\ell}^t\}). \tag{2.14}$$

Proof. We have

$$\mathbb{E}_t\left[\hat{\mathbf{p}}^t \cdot \hat{\boldsymbol{\ell}}^t\right] = \hat{\mathbf{p}}^t \cdot \mathbb{E}_t\left[\hat{\boldsymbol{\ell}}^t\right] = \hat{\mathbf{p}}^t \cdot \boldsymbol{\ell}^t,$$

since $\hat{\mathbf{p}}^t$ is determined by history prior to t. Thus

$$\mathbb{E}\left[\hat{\mathbf{p}}^t \cdot \hat{\boldsymbol{\ell}}^t\right] = \mathbf{p}^t \cdot \boldsymbol{\ell}^t.$$

Also,

$$\mathbb{E}\left[\min_i \hat{L}_i^T\right] \leq \min_i \mathbb{E}\left[\hat{L}_i^T\right] = \min_i L_i^T.$$

Thus,

$$\sum_t \mathbb{E}\left[\hat{\mathbf{p}}^t \cdot \hat{\boldsymbol{\ell}}^t\right] - \mathbb{E}\left[\min_i \hat{L}_i^T\right] \geq \sum_t \mathbf{p}^t \cdot \boldsymbol{\ell}^t - \min_i L_i^T,$$

yielding (2.14). □

Remark. From an algorithmic perspective, there is no need to compute $\mathbf{p}^t = \mathbb{E}\left[\hat{\mathbf{p}}^t\right]$ in order to implement $\mathcal{D}_{[0,1]}$'s decision at time t. Rather, $\mathcal{D}_{[0,1]}$ can simply use $\hat{\mathbf{p}}^t$ at step t.

2.5.2 Optimal adversary strategy

The adaptive decision-making problem can be seen as a two-player zero-sum game as follows. The pure strategies† of the adversary (player I) are the loss vectors $\{\boldsymbol{\ell}^t\} \in \{0,1\}^{nT}$. The pure strategies for the decider $\mathcal{D}$ (player II) are $\{a_t\}_{t=1}^T$, where $a_t : \{0,1\}^{n(t-1)} \to [n]$.

By the minimax theorem

$$\min_{\mathcal{D}} \max_{\{\boldsymbol{\ell}^t\}} \mathcal{R}_T(\mathcal{D}, \{\boldsymbol{\ell}^t\}) = \max_{\mathcal{L}} \min_{\mathcal{D}} \mathcal{R}_T(\mathcal{D}, \mathcal{L})$$

By Theorem **??**, we may restrict attention to behavioral strategies for the decider, player II, i.e. $\mathcal{D} = \{\mathbf{p}^t\}_{t=1}^T$, as in the previous section.

Remark. Formally, a behavioral randomized decision strategy $\mathcal{D}$ could depend on previous actions of the decider, i.e., $\mathbf{p}^t$ could be a function of $\{a_s\}_{s=1}^{t-1}$ as well as the losses $\{\boldsymbol{\ell}^s\}_{s=1}^{t-1}$. Clearly,

$$\mathbb{E}\left[\mathcal{R}_T(\mathcal{D}, L)\right] = \mathcal{R}_T(\overline{\mathcal{D}}, L),$$

where in $\overline{\mathcal{D}}$, each $\mathbf{p}^t$ is replaced by its average over past actions $\overline{\mathbf{p}^t} = \mathbb{E}\left[\mathbf{p}^t\right]$. This justifies the restriction to decision strategies that are independent of previous actions.

Optimal adversary strategy

An adversary strategy $\mathcal{L}$ is a probability distribution over loss sequences $\{\boldsymbol{\ell}^t\}_{t=1}^T$. We say that adversary strategy $\mathcal{L}$ is **balanced** if, for every time

---

† These are *oblivious* strategies, which do not depend on previous decider actions. See the notes for a discussion of *adaptive*, i.e., non-oblivious, adversary strategies.

t and every history of losses through time $t-1$, the expected loss of each expert is the same, i.e., for all pairs of actions i and j, $\mathbb{E}_t\left[\ell_i^t\right]=\mathbb{E}_t\left[\ell_j^t\right]$.

Proposition 2.5.2. *Let* $\mathcal{R}_T(\mathcal{L}) := \min_{\mathcal{D}} \mathcal{R}_T(\mathcal{D},\mathcal{L})$. *Then*

$$\max_{\mathcal{L}}\min_{\mathcal{D}}\mathcal{R}_T(\mathcal{D},\mathcal{L})=\max_{\mathcal{L}}\mathcal{R}_T(\mathcal{L})$$

is attained in ***balanced strategies****.*

Proof. Clearly $\min_{\mathcal{D}} \mathcal{R}_T(\mathcal{D},\mathcal{L})$ is achieved by choosing, at each time step t, the action which has the smallest expected loss in each step, given the history of losses. We claim that for every $\mathcal{L}$ that is not balanced at time t for some history, there is an alternative strategy $\tilde{\mathcal{L}}$ that is balanced at t and has

$$\mathcal{R}_T(\tilde{\mathcal{L}})\geq\mathcal{R}_T(\mathcal{L}).$$

Construct such a $\tilde{\mathcal{L}}$ as follows: Pick $\{\boldsymbol{\ell}^t\}$ according to $\mathcal{L}$ conditioned on the history. Let $\tilde{\boldsymbol{\ell}}^s=\boldsymbol{\ell}^s$ for all $s\neq t$. At time t, define

$$\tilde{\ell}_i^t=\ell_i^t\theta_i$$

where

$$\theta_i\in\{0,1\}\qquad\text{and}\qquad\mathbb{E}_t\left[\theta_i\right]=\frac{\min_j\mathbb{E}_t\left[\ell_j^t\right]}{\mathbb{E}_t\left[\ell_i^t\right]}.$$

This change ensures that at time t all experts have expected loss equal to $\min_j \mathbb{E}_t\left[\ell_j^t\right]$ The best response strategy of $\mathcal{D}$ is still a best response at time t. But the benchmark loss $\min_j \mathbb{E}\left[L_j^T\right]$ is weakly reduced. ☐

Against a balanced adversary, to calculate regret, $\mathcal{D}$ is irrelevant. Taking a uniform $\mathcal{D}$, we have

$$\mathcal{R}_T(\mathcal{L})=\mathcal{R}_T(\mathcal{D},\mathcal{L})=\frac{1}{n}\sum_{i=1}^{n}L_i^T-\min_i L_i^T. \tag{2.15}$$

2.5.3 The case of two actions

For $n=2$, (2.15) reduces to

$$\mathcal{R}_T(\mathcal{L})=\frac{1}{2}(L_1^T+L_2^T)-\min(L_1^T,L_2^T)=\frac{1}{2}|L_1^T-L_2^T|.$$

Clearly, $X^t = \ell_1^t - \ell_2^t$ defines a random walk $S_T = \sum_{t=1}^T X^t$. We have $X^t \in \{-1, 0, 1\}$ with $\mathbb{E}_t\left[X^t\right] = 0$. To maximize $\mathbb{E}|S_T| = 2\mathcal{R}_T(\mathcal{L})$, we let $X^t \in \{-1, 1\}$. Specifically, $\boldsymbol{\ell}^t$ is i.i.d., equally likely to be $(1, 0)$ or $(0, 1)$. Thus, by the Central Limit Theorem, we have

$$\mathcal{R}_T = \mathcal{R}_T(\mathcal{L}) = \frac{1}{2}\mathbb{E}|S_T| = \frac{1}{2}\sqrt{T}\,\mathbb{E}|Z|(1 + o(1))$$

with $Z \sim N(0, 1)$, so $\mathbb{E}|Z| = \sqrt{\frac{2}{\pi}}$.

Optimal decider

To find the optimal $\mathcal{D}$, we consider an initial integer gap $h \geq 0$ between the actions, and define

$$r_T(h) = \min_{\mathcal{D}} \max_{\mathcal{L}} \mathbb{E}\left[L_{\mathcal{D}}^T - \min\{L_1^T + h, L_2^T\}\right]$$

where $L_{\mathcal{D}}^T = \sum_{t=1}^T \boldsymbol{\ell}^t \cdot \mathbf{p}^t$. By the minimax theorem,

$$r_T(h) = \max_{\mathcal{L}} \max_{\mathcal{D}} \mathbb{E}\left[L_{\mathcal{D}}^T - \min\{L_1^T + h, L_2^T\}\right].$$

As in the discussion above, the optimal adversary is balanced, so we have

$$r_T(h) = \frac{1}{2}(L_1^T + L_2^T) - \min(L_1^T + h, L_2^T) = \frac{1}{2}\Big(|L_1^T + h - L_2^T| - h\Big).$$

Again, the adversary's optimal strategy is to select $\boldsymbol{\ell}^t$ i.i.d., equally likely to be $(1, 0)$ or $(0, 1)$, so

$$r_T(h) = \frac{1}{2}\mathbb{E}\left[|h + S_T| - h\right]. \tag{2.16}$$

We now calculate the optimal strategy for $\mathcal{D}$. To emphasize the dependence on T and h, write $q_T(h) = p_1^1$, the chance that the optimal $\mathcal{D}$ chooses action 1 in the first step.

Observe that for $h > 0$,

$$r_T(h) = \max\Big(r_{T-1}(h + 1) + q_T(h),\, r_{T-1}(h - 1) - 1 + 1 - q_T(h)\Big)$$

At optimality, the decider will choose $q_T(h)$ to equalize these costs, if he can, in which case:

$$q_T(h) = \frac{r_{T-1}(h - 1) - r_{T-1}(h + 1)}{2}.$$

Thus by (2.16)

$$q_T(h) = \frac{1}{4}\mathbb{E}\Big[|h - 1 + S_{T-1}| - |h + 1 - S_{T-1}| + 2\Big].$$

Since

$$\mathbb{E}\Big[|h-1+S_{T-1}|-|h+1-S_{T-1}|\Big]=\begin{cases}-2 & \text{if } S_{T-1}+h>0\\ 0 & \text{if } S_{T-1}+h=0\\ 2 & \text{if } S_{T-1}+h<0\end{cases}$$

we conclude that

$$q_T(h)=\mathbb{P}\,[S_{T-1}+h<0]+\frac{1}{2}\mathbb{P}\,[S_{T-1}+h=0]. \tag{2.17}$$

In other words, $q_T(h)$ is the probability that the action currently lagging by h will be the leader in T steps.

Theorem 2.5.3. *For $n=2$, with losses in $\{0,1\}$, the minimax optimal regret is*

$$\mathcal{R}_T=\sqrt{\frac{T}{2\pi}}\,(1+o(1)).$$

The optimal adversary strategy is to take $\boldsymbol{\ell}^t$ i.i.d., equally likely to be $(1,0)$ or $(0,1)$.

The optimal decision strategy $\{\mathbf{p}^t\}_{t=1}^T$ is determined as follows: First, $\mathbf{p}^1=(1/2,1/2)$. For $t\in[1,T-1]$, let $L_{i_t}^t=\min(L_1^t,L_2^t)$ and $h_t=|L_1^t-L_2^t|$. At time $t+1$, take the leading action i_t with probability $p_{i_t}^{t+1}=1-q_{T-t}(h_t)$ and the lagging action $3-i_t$ with probability $p_{3-i_t}^{t+1}=q_{T-t}(h_t)$.

Let Φ denote the standard normal distribution function. By the central limit theorem

$$q_T(h)=\Phi(-h/\sqrt{T})(1+o(1)) \qquad \text{as } T\to\infty$$

so the optimal algorithm is easy to implement.

2.5.4 Adaptive versus oblivious adversaries

In the preceding, we assumed the adversary is oblivious, i.e., selects the loss vectors $\boldsymbol{\ell}^t=\boldsymbol{\ell}^t(\mathcal{D})$ independently of the actions of the decider. (He can still use the mixed strategy of $\mathcal{D}$ but not the actual random choices.)

A more powerful adversary is *adaptive*, i.e., can select loss vectors $\boldsymbol{\ell}^t=\boldsymbol{\ell}^t(\mathcal{D},a_1,a_2,\ldots,a_{t-1})$ that depend on previous actions. With the (standard) definition of regret that we used, for every $\mathcal{D}$, adaptive adversaries cause the same worst-case regret as oblivious ones; both simply equal the maximum over individual loss sequences $\max_L\mathcal{R}_T(\mathcal{D},L)$. For this reason, it is often noted that low regret algorithms like Multiplicative Weights work against adaptive adversaries as well as against oblivious ones.

Against adaptive adversaries, the notion of regret we use here is

$$\mathcal{R}_T(\mathcal{D}, L) = \mathbb{E}\left[L_{\mathcal{D}}^T - \min_i \sum_{t=1}^{T} \ell_i^t(a_1, \dots a_{t-1})\right]. \tag{2.18}$$

An alternative known as **policy regret** is

$$\mathcal{R}_T^*(\mathcal{D}, L) = \mathbb{E}\left[L_{\mathcal{D}}^T - \min_i \sum_{t=1}^{T} \ell_i^t(i, \dots i)\right]. \tag{2.19}$$

The notion of regret in (2.18) is useful in the setting of learning from expert advice where it measures the performance of the decider relative to the performance of the best expert. Next we give two examples where policy regret is more appropriate.

(i) Let $\mathcal{L} = \{\boldsymbol{\ell}^t\}$ be any oblivious loss sequence. Imposing a switching cost can be modeled as an adaptive adversary $\tilde{\mathcal{L}}$ defined by

$$\tilde{\ell}_i^t = \ell_i^t + \mathbb{1}_{\{a_{t-1} \neq a_{t-2}\}} \quad \forall i, t.$$

The usual regret will ignore the switching cost, i.e.,

$$\mathcal{R}_T(\mathcal{D}, \mathcal{L}) = \mathcal{R}_T(\mathcal{D}, \tilde{\mathcal{L}}) \quad \forall \mathcal{D},$$

but policy regret will take it into account. E.g. if $\ell_i^t = \mathbb{1}_{\{i = t \mod 2\}}$, then $\mathcal{R}_T(MW, \tilde{\mathcal{L}}) = O(1)$, but $\mathcal{R}_T^*(MW, \tilde{\mathcal{L}}) = T/2 + O(1)$.

(ii) Consider a decider playing repeated Prisoner's Dilemma (as discussed in Example **??** and §**??**) for T rounds.

		player II	
		cooperate	defect
player I	cooperate	$(-1, -1)$	$(-10, 0)$
	defect	$(0, -10)$	$(-8, -8)$

Suppose that the loss sequence L is defined by a opponent playing

Tit-for-Tat.† In this case, defining $a_0 = C$, the losses at time t are:

$$\ell_i^t(a_{t-1}) = \begin{cases} 1 & (a_{t-1}, i) = (C, C) \\ 0 & (a_{t-1}, i) = (C, D) \\ 10 & (a_{t-1}, i) = (D, C) \\ 8 & (a_{t-1}, i) = (D, D). \end{cases}$$

Since it is a dominant strategy to defect in prisoner's dilemma,

$$L^T_{\text{defect}} > L^T_{\text{cooperate}}.$$

(This holds for any opponent, not just Tit-for-Tat.) Thus, for any decider $\mathcal{D}$,

$$\mathcal{R}_T(\mathcal{D}, L) = \sum_t \mathbb{1}_{\{a_t=C\}}(\mathbb{1}_{\{a_{t-1}=C\}} + 2\mathbb{1}_{\{a_{t-1}=D\}}).$$

Minimizing regret will lead the decider towards defecting every round and incurring a loss of $8(T-1)$. However, minimizing policy regret will lead the decider to cooperating for $T-1$ rounds yielding a loss of $T-1$.

Notes

The origins of the material in this chapter go back to the work of Hannan [Han57], who was motivated by the question of how to play a repeated game. These issues received renewed attention starting in the 1990s due to their applicability in machine learning settings. For in-depth expositions of this topic, see the book by Cesa-Bianchi and Lugosi [CBL06] and the surveys by Blum and Mansour (Chapter 4 of [Nis07]) and Arora, Hazan and Kale [AHK12].

The Weighted Majority algorithm and its analysis in §2.2.1 are due to Littlestone and Warmuth [LW94]. The Multiplicative Weights algorithm discussed in §2.3.1 and variants thereof are from Littlestone and Warmuth [LW94], Cesa-Bianchi et al [CBFH$^+$97] and Freund and Schapire [FS97]. A suite of decision-making algorithms closely related to Multiplicative Weights achieves similar or the same regret bounds and go under different names including Exponential Weights, Hedge, etc. The use of the Multiplicative Weights algorithm to play zero-sum games discussed in §2.4 is due to Grigoriadis and Khachiyan [] and Freund and Schapire [FS97]. Theorem 2.5.3 is due to [Cov65]. Our exposition follows [**?**], where optimal strategies for three experts are also determined. The notion of policy regret discussed in §2.5.4 is due to Arora, Dekel and Tiwari [ADT12].

Some important extensions of the material in this chapter include the "bandit" setting and "swap regret". The bandit setting addresses the situation where the decider learns his loss each round, but does not learn the losses of actions he did not choose. Surprisingly, it is possible to achieve regret $O(\sqrt{Tn \log n})$ in the bandit setting [**?**].

---

† Recall Definition **??**: Tit-for-Tat is the strategy in which the player cooperates in round 1 and in every round thereafter plays the strategy his opponent played in the previous round.

In swap regret, the benchmark loss is strengthened to allow for replacing each action i selected by the player by a corresponding action $f(i)$ as opposed to a replacing all actions $i \in [n]$ by some action j. Hart and Mas-Collel showed how to achieve sublinear swap regret using Blackwell's Approachability Theorem [**?**]. When players playing a game use sublinear swap regret algorithms, play can be shown to converge to a correlated equilibrium. A final topic related to the material in this chapter is that of online convex optimization. See, e.g., the survey by Shalev-Shwartz [SS11].

One of the first strategies proposed (by Brown [**?**]) for repeated play of a 2-player zero-sum game is known as *fictitious play* or *Follow the Leader*: At time t, each player plays a best response to the empirical distribution of play by their opponent in the first $t - 1$ rounds. Julia Robinson showed that if both players use fictitious play, their empirical distributions converge to optimal strategies [**?**]. However, as discussed in §2.1, this strategy does not achieve sublinear regret in the setting of binary prediction with expert advice. A variant, known as *Follow the Perturbed Leader*, in which each action/expert is initially given a random loss and then henceforth the leader is followed, achieves essentially the same bounds as the Multiplicative Weights algorithm, and has the advantage of also achieving low policy regret? Follow the Perturbed Leader was analyzed by Hannan [Han57] and Kalai and Vempala [**?**].

sublinear regret - Hannan consistent

Exercise 2.3 is due to Avrim Blum.

Exercises

2.1 Consider the setting of §2.2, and suppose that u_t (respectively d_t) is the number of leaders voting up (respectively down) at time t. Consider any trader algorithm $\mathcal{A}$ that decides up or down at time t with probability p_t, where $p_t = p_t(u_t, d_t)$. Then there is an adversary strategy that ensures that any such trader algorithm $\mathcal{A}$ incurs an expected loss of at least $\lfloor \log_4(n) \rfloor L_*^T$.

Hint: Adapt the adversary strategy in Proposition 2.1.2, ensuring that no expert incurs more than one mistake during S_0. Repeat.

2.2 Show that there are cases where any deterministic algorithm in the experts setting makes at least twice as many mistakes as the best expert, i.e., for some T, $L(T) \geq 2L_*^T$.

2.3 Consider the following variation on Weighted Majority:

> On each day t, associate a weight w_i^t with each expert i.
> Initially, when $t = 1$, set $w_i^1 = 1$ for all i.
> Each day t, follow the ***weighted majority*** opinion: Let U_t be the set of experts predicting up on day t, and D_t the set predicting down. Predict "up" on day t if $W_U(t) = \sum_{i \in U_t} w_i^t \geq W_D(t) = \sum_{i \in D_t} w_i^t$ and "down" otherwise.

On day $t+1$, for each i such that (a) expert i predicted incorrectly on day t, and (b) $w_i^t \geq \frac{1}{4n}\sum_j w_j^t$, set

$$w_i^{t+1} = \frac{1}{2} w_i^t \tag{E2.1}$$

Show that for every contiguous subsequence of days, say $\tau, \tau+1, \ldots, \tau+r$, the number of mistakes made by the algorithm during those days is $O(m + \log n)$, where m is the fewest number of mistakes made by any expert on days $\tau, \tau+1, \ldots, \tau+r$.

2.4 Consider the sequential adaptive decision-making setting of §2.3 with unknown time horizon T. Adapt the MW algorithm by changing the parameter ϵ over time, to a new value at $t = 2^j$ for $j = 0, 1, 2, \ldots$. (This is a "doubling trick".) Show that the sequence of ϵ values can be chosen so that for every action i,

$$L^T \leq L_i^T + \frac{\sqrt{2}}{\sqrt{2}-1}\sqrt{\frac{1}{2} T \log n}.$$

2.5 Generalize the results of §2.5.4 for $n = 2$ to the case where the time horizon T is geometric with parameter δ, i.e., the process stops with probability δ in every round:

- Determine the minimax optimal adversary and the minimax regret.
- Determine the minimax optimal decision algorithm.

2.6 (a) For Y a normal random variable, $N(0,1)$, show that

$$e^{-\frac{y^2}{2}(1+o(1))} \leq \mathbb{P}\left[Y > y\right] \leq e^{-\frac{y^2}{2}} \quad \text{as} \quad y \to \infty.$$

(b) Suppose that $Y_1, \ldots, Y_n$ are i.i.d. $N(0,1)$ random variables. Show that

$$\mathbb{E}\left[\max_{1 \leq i \leq n} Y_i\right] = \sqrt{2 \log n}\,(1 + o(1)) \quad \text{as} \quad n \to \infty. \tag{E2.2}$$

Solution:

(a)

$$\int_y^\infty e^{-x^2/2}\,dx \leq \frac{1}{y}\int_y^\infty x e^{-x^2/2}\,dx = \frac{1}{y} e^{-y^2/2}$$

$$\text{and} \quad \int_y^{y+1} e^{-x^2/2}\,dx \geq e^{-\frac{(y+1)^2}{2}}.$$

(b) Let $M_n = \mathbb{E}\left[\max_{1\leq i\leq n} Y_i\right]$. Then by a union bound

$$\mathbb{P}\left[M_n \geq \sqrt{2\log n} + \frac{x}{\sqrt{2\log n}}\right] \leq ne^{-(\log n + x)} = e^{-x}.$$

On the other hand,

$$\mathbb{P}\left[Y_i > \sqrt{2\alpha\log n}\right] = n^{-\alpha+o(1)}$$

$$\text{so} \quad \mathbb{P}\left[M_n \geq \sqrt{2\alpha\log n}\right] = \left(1 - n^{-\alpha+o(1)}\right)^n \to 0 \qquad \text{for } \alpha < 1.$$

2.7 Consider an adaptive adversary with bounded memory, that is $\ell_i^t = \ell_i^t(a_{t-m}, \ldots, a_{t-1})$ for constant m. Consider a decider that divides time into blocks of length b, and uses a fixed action, determined by the Multiplicative Weights Algorithm in each block. Show that the policy regret of this decider is $O(\sqrt{Tb\log n} + Tm/b)$. Then optimize over b.

Appendix 1

Some useful mathematical lemmas

Lemma A1.0.4. *For any* $n \times n$ *stochastic matrix* Q *(a matrix is stochastic if all of its rows are probability vectors), there is a row vector* $\boldsymbol{\pi} \in \Delta_n$ *such that*

$$\boldsymbol{\pi} = \boldsymbol{\pi} Q.$$

Proof. This is a special case of Brouwer's Fixed Point Theorem **??**, but there is a simple direct proof: Let $\mathbf{v} \in \Delta_n$ arbitrary and define

$$\mathbf{v}_T = \frac{1}{T}\mathbf{v}(I + Q + Q^2 + \ldots + Q^{T-1}).$$

Then

$$\mathbf{v}_T Q - \mathbf{v}_T = \frac{1}{T}\mathbf{v}(Q^T - I) \longrightarrow 0$$

as $T \to \infty$, so any limit point $\boldsymbol{\pi}$ of $\mathbf{v}_T$ must satisfy $\boldsymbol{\pi} = \boldsymbol{\pi} Q$.

□

Remark. In fact, $\mathbf{v}_T$ converges for any $\mathbf{v} \in \Delta_n$. See Exercise A1.0.5.

Exercise A1.0.5. Prove that $\mathbf{v}_T$ from Lemma A1.0.4 converges for any $\mathbf{v} \in \Delta_n$.

Solution:

Let P be any $n \times n$ stochastic matrix (possibly reducible) and denote $Q_T = \frac{1}{T}\sum_{t=0}^{T-1} P^t$. Given a probability vector $v \in \Delta_n$ and $T > 0$, we define $v_T = vQ_T$. Then $\|v_T(I-P)\|_1 = \|v(I-P^T)\|_1/T \le 2/T$, so any subsequential limit point z of v_T satisfies $z = zP$. To see that v_T actually converge, an additional argument is needed. With $I - P$ acting on row vectors in $\mathbb{R}^n$ by multiplication from the right, we claim that the kernel and the image of $I-P$ intersect only in 0. Indeed, if $z = w(I-P)$ satisfies $z = zP$, then $z = zQ_T = w(I - P^T)$ must satisfy $\|z\|_1 \le 2\|w\|_1/T$ for every T, so necessarily $z = 0$. Since the dimensions of $\mathrm{Im}(I-P)$ and $\mathrm{Ker}(I-P)$ add up to n, it follows that any vector $v \in \mathbb{R}^n$ has a unique representation $v = u+z$ (*) with $u \in \mathrm{Im}(I-P)$ and $z \in \mathrm{Ker}(I-P)$. Therefore $v_T = vQ_T = uQ_T + z$, so writing $u = x(I-P)$ we conclude that $\|v_T - z\|_1 \le 2\|x\|_1/T$. If $v \in \Delta_n$ then also $z \in \Delta_n$ due to z being the limit of v_T; The non-negativity of the entries of z is not obvious from the representation (*) alone.

A1.1 The Second Moment Method

Lemma A1.1.1. *Let X be a nonnegative random variable. Then*

$$\mathbb{P}(X > 0) \ge \frac{(\mathbb{E}[X])^2}{\mathbb{E}[X^2]}.$$

Proof. The lemma follows from the following version of the Cauchy Schwartz inequality:

$$(\mathbb{E}[XY])^2 \le \mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right]. \tag{1.1}$$

Applying (1.1) to X and $Y = \mathbb{1}_{X>0}$, we obtain

$$(\mathbb{E}[X])^2 \le \mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right] = \mathbb{E}\left[X^2\right]\mathbb{P}(X > 0).$$

Finally, we prove (1.1). Without loss of generality $\mathbb{E}\left[X^2\right]$ and $\mathbb{E}\left[Y^2\right]$ are both positive. Letting $U = X/\sqrt{\mathbb{E}[X^2]}$ and $V = Y/\sqrt{\mathbb{E}[Y^2]}$, and using the fact that $2|UV| \le U^2 + V^2$, we obtain

$$2\mathbb{E}[|UV|] \le \mathbb{E}\left[U^2\right] + \mathbb{E}\left[V^2\right] = 2.$$

Therefore,

$$[\mathbb{E}[UV]]^2 \le 1$$

which is equivalent to (1.1). □

A1.2 The Hoeffding-Azuma Inequality

Lemma A1.2.1 (Hoeffding Lemma)**.** *Suppose that X is a random variable with distribution F such that $a \le X \le a+1$, for some $a \le 0$, and* $\mathbb{E}[X] = 0$. *Then for any λ,*

$$\mathbb{E}\left[e^{\lambda X}\right] \le e^{\lambda^2/8}.$$

Proof. Let

$$\Psi(\lambda) = \log \mathbb{E}\left[e^{\lambda X}\right].$$

Observe that

$$\Psi'(\lambda) = \frac{\mathbb{E}\left[Xe^{\lambda X}\right]}{\mathbb{E}\left[e^{\lambda X}\right]} = \int x dF_\lambda,$$

where

$$F_\lambda(u) = \frac{\int_{-\infty}^{u} e^{\lambda x} dF}{\int_{-\infty}^{\infty} e^{\lambda x} dF}.$$

Also,

$$\begin{aligned}\Psi''(\lambda) &= \frac{\mathbb{E}\left[e^{\lambda X}\right]\mathbb{E}\left[X^2 e^{\lambda X}\right] - \left(\mathbb{E}\left[Xe^{\lambda X}\right]\right)^2}{\left(\mathbb{E}\left[e^{\lambda X}\right]\right)^2} \\ &= \int x^2 dF_\lambda - \left(\int x dF_\lambda\right)^2 = \mathrm{Var}(X_\lambda),\end{aligned}$$

where X_λ has law F_λ.

For any random variable Y with $a \le Y \le a+1$, we have

$$\mathrm{Var}(Y) \le \mathbb{E}\left[\left(Y - a - \frac{1}{2}\right)^2\right] \le \frac{1}{4}.$$

In particular,

$$|\Psi''(\lambda)| \le 1/4$$

for all λ. Since $\Psi(0) = \Psi'(0) = 0$, it follows that $|\Psi'(\lambda)| \le \frac{|\lambda|}{4}$, and thus

$$\Psi(\lambda) \le \left| \int_0^\lambda \frac{\theta}{4} d\theta \right| = \frac{\lambda^2}{8}$$

for all λ. □

Theorem A1.2.2 (Hoeffding-Azuma Inequality)**.** *Let* $S_t = \sum_{i=1}^{t} X_i$ *be a martingale, i.e.* $\mathbb{E}\left[S_{t+1}|H_t\right] = S_t$ *where* $H_t = (X_1, X_2, \ldots, X_t)$ *represents the history. If all* $|X_t| \le 1$*, then*

$$\mathbb{P}\left[S_t \ge R\right] \le e^{-R^2/2t}.$$

Proof. Since $-\frac{1}{2} \le \frac{X_{t+1}}{2} \le \frac{1}{2}$, the previous lemma gives

$$\mathbb{E}\left[e^{\lambda X_{t+1}}|H_t\right] \le e^{\frac{(2\lambda)^2}{8}} = e^{\lambda^2/2}$$

so

$$\mathbb{E}\left[e^{\lambda S_{t+1}}|H_t\right] = e^{\lambda S_t}\mathbb{E}\left[e^{\lambda X_{t+1}}|H_t\right] \le e^{\lambda^2/2} e^{\lambda S_t}.$$

Taking expectations

$$\mathbb{E}\left[e^{\lambda S_{t+1}}\right] \leq e^{\lambda^2/2}\mathbb{E}\left[e^{\lambda S_t}\right],$$

so by induction on t

$$\mathbb{E}\left[e^{\lambda S_t}\right] \leq e^{t\lambda^2/2}.$$

Finally, by Markov's Inequality,,

$$\mathbb{P}\left[S_t \geq R\right] = \mathbb{P}\left[e^{\lambda S_t} \geq e^{\lambda R}\right] \leq e^{-\lambda R}e^{t\lambda^2/2}.$$

Optimizing we choose $\lambda = R/t$, so

$$\mathbb{P}\left[S_t \geq R\right] \leq e^{-R^2/2t}.$$

□

Bibliography

[AABR09] Jacob Abernethy, Alekh Agarwal, Peter L. Bartlett, and Alexander Rakhlin. A stochastic view of optimal regret through minimax duality. In *COLT 2009 - The 22nd Conference on Learning Theory, Montreal, Quebec, Canada, June 18-21, 2009*, 2009.

[ACBFS02] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002.

[ADT12] Raman Arora, Ofer Dekel, and Ambuj Tewari. Online bandit learning against an adaptive adversary: from regret to policy regret. In *Proceedings of the 29th International Conference on Machine Learning, ICML 2012, Edinburgh, Scotland, UK, June 26 - July 1, 2012*, 2012. 42

[AG03] Steve Alpern and Shmuel Gal. *The theory of search games and rendezvous.* International Series in Operations Research & Management Science, 55. Kluwer Academic Publishers, Boston, MA, 2003.

[AH81] Robert Axelrod and William D. Hamilton. The evolution of cooperation. *Science*, 211(4489):1390–1396, 1981.

[AHK12] Sanjeev Arora, Elad Hazan, and Satyen Kale. The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing*, 8(1):121–164, 2012. 42

[Ake70] George A Akerlof. The market for "lemons": Quality uncertainty and the market mechanism. *The quarterly journal of economics*, pages 488–500, 1970.

[Ansa] Hex information. `http://www.cs.unimaas.nl/icga/games/hex/`.

[Ansb] V. V. Anshelevich. The game of Hex: An automatic theorem proving approach to game programming. `http://home.earthlink.net/~vanshel/VAnshelevich-01.pdf`.

[Arr51] Kenneth J. Arrow. *Social Choice and Individual Values.* Cowles Commission Monograph No. 12. John Wiley & Sons Inc., New York, N. Y., 1951.

[Arr02] A. Arratia. On the descriptive complexity of a simplified game of Hex. *Log. J. IGPL*, 10:105–122, 2002.

[ARS$^+$03] Micah Adler, Harald Räcke, Naveen Sivadasan, Christian Sohler, and Berthold Vöcking. Randomized pursuit-evasion in graphs. *Combin. Probab. Comput.*, 12(3):225–244, 2003. Combinatorics, probability and computing (Oberwolfach, 2001).

[Aum87] Robert J. Aumann. Correlated equilibrium as an expression of Bayesian rationality. *Econometrica*, 55(1):1–18, jan 1987. `http://www.jstor.org/pss/`

`1911154`.

[Axe85] Robert Axelrod. *The Evolution of Cooperation.* Basic Books, 387 Park Avenue So., New York, NY 10016, 1985.

[Axe97] Robert Axelrod. The evolution of strategies in the iterated prisoner's dilemma. In *The dynamics of norms*, Cambridge Stud. Probab. Induc. Decis. Theory, pages 1–16. Cambridge Univ. Press, Cambridge, 1997.

[Ban51] Thøger Bang. A solution of the "plank problem.". *Proc. Amer. Math. Soc.*, 2:990–993, 1951.

[BBC69] A. Beck, M. Bleicher, and J. Crow. *Excursions into Mathematics.* Worth, 1969.

[BCG82a] E. R. Berlekamp, J. H. Conway, and R. K. Guy. *Winning Ways for Your Mathematical Plays*, volume 1. Academic Press, 1982.

[BCG82b] E. R. Berlekamp, J. H. Conway, and R. K. Guy. *Winning Ways for Your Mathematical Plays*, volume 2. Academic Press, 1982.

[BK99] Jeremy Bulow and Paul Klemperer. The generalized war of attrition. *American Economic Review*, pages 175–189, 1999.

[Bou02] Charles L. Bouton. Nim, a game with a complete mathematical theory. *Ann. of Math. (2)*, 3(1-4):35–39, 1901/02.

[BPP$^+$14] Yakov Babichenko, Yuval Peres, Ron Peretz, Perla Sousi, and Peter Winkler. Hunter, Cauchy rabbit, and optimal Kakeya sets. *Trans. Amer. Math. Soc.*, 366(10):5567–5586, 2014.

[Bro00] C. Browne. *Hex Strategy: Making the Right Connections.* A. K. Peters, 2000.

[BSW05] Itai Benjamini, Oded Schramm, and David B. Wilson. Balanced Boolean functions that can be evaluated so that every input bit is unlikely to be read. In *Proc. 37th Symposium on the Theory of Computing*, 2005.

[Car92] John L. Cardy. Critical percolation in finite geometries. *J. Phys. A*, 25(4):L201–L206, 1992.

[CBFH$^+$97] Nicolò Cesa-Bianchi, Yoav Freund, David Haussler, David P. Helmbold, Robert E. Schapire, and Manfred K. Warmuth. How to use expert advice. *J. ACM*, 44(3):427–485, May 1997. 42

[CBL06] Nicolò Cesa-Bianchi and Gabor Lugosi. *Prediction, Learning, and Games.* Cambridge University Press, New York, NY, USA, 2006. 42

[CD06] Xi Chen and Xiaotie Deng. Settling the complexity of two-player nash equilibrium. In *IEEE Foundations of Computer Science*, volume 6, page 47th, 2006.

[CDT06] Xi Chen, Xiaotie Deng, and Shang-Hua Teng. Computing nash equilibria: Approximation and smoothed complexity. In *Foundations of Computer Science, 2006. FOCS'06. 47th Annual IEEE Symposium on*, pages 603–612. IEEE, 2006.

[CH13] Shuchi Chawla and Jason D Hartline. Auctions with unique equilibria. In *ACM Conference on Electronic Commerce*, pages 181–196, 2013.

[Cla71] Edward H Clarke. Multipart pricing of public goods. *Public choice*, 11(1):17–33, 1971.

[Cla00] Eward H Clarke. *Revelation and the Provision of Public Goods.* iUniverse, 2000.

[CM88] Jacques Cremer and Richard P McLean. Full extraction of the surplus in bayesian and dominant strategy auctions. *Econometrica: Journal of the Econometric Society*, pages 1247–1257, 1988.

[Con89] Antonio Coniglio. Fractal structure of Ising and Potts clusters: exact results. *Phys. Rev. Lett.*, 62(26):3054–3057, 1989.

[Con01] J. H. Conway. *On Numbers and Games.* A K Peters Ltd., Natick, MA, second edition, 2001.

[Cov65] Thomas M. Cover. Behavior of sequential predictors of binary sequences. In *Proceedings of the 4th Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, pages 263–272, 1965. 42

[Das13] Constantinos Daskalakis. On the complexity of approximating a nash equilibrium. *ACM Transactions on Algorithms (TALG)*, 9(3):23, 2013.

[dCMdC90] Marie Jean Antoine Nicolas de Caritat (Marquis de Condorcet). Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix (eng: Essay on the application of analysis to the probability of majority decisions), 1785. In *The French Revolution Research Collection.* Pergamon Press, Headington Hill Hall, Oxford OX3 0BW UK, 1990. `http://gallica.bnf.fr/scripts/ConsultationTout.exe?E=0&O=N041718`.

[DDT12] Constantinos Daskalakis, Alan Deckelbaum, and Christos Tzamos. The complexity of optimal mechanism design. *arXiv preprint arXiv:1211.1703*, 2012.

[DGP09] Constantinos Daskalakis, Paul W Goldberg, and Christos H Papadimitriou. The complexity of computing a nash equilibrium. *SIAM Journal on Computing*, 39(1):195–259, 2009.

[DS84] Peter G. Doyle and J. Laurie Snell. *Random walks and electric networks*, volume 22 of *Carus Mathematical Monographs.* Mathematical Association of America, Washington, DC, 1984.

[Dub57] L. E. Dubins. *A discrete evasion game.* Princeton University Press, Princeton, N.J., 1957.

[EOS05] Benjamin Edelman, Michael Ostrovsky, and Michael Schwarz. Internet advertising and the generalized second price auction: Selling billions of dollars worth of keywords. Technical report, National Bureau of Economic Research, 2005.

[EOS07] Benjamin Edelman, Michael Ostrovsky, and Michael Schwarz. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *American Economic Review*, 97(1):242–259, 2007.

[Eve57] H. Everett. Recursive games. In *Contributions to the theory of games, vol. 3*, Annals of Mathematics Studies, no. 39, pages 47–78. Princeton University Press, Princeton, N. J., 1957. 19

[Fer08] Thomas Ferguson. *Game Theory.* 2008. `http://www.math.ucla.edu/~tom/Game_Theory/Contents.html`. 19

[FPT04] Alex Fabrikant, Christos Papadimitriou, and Kunal Talwar. The complexity of pure Nash equilibria. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, pages 604–612. ACM, New York, 2004.

[FS97] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.*, 55(1):119–139, August 1997. 42

[FV93] Dean P. Foster and Rakesh V. Vohra. A randomization rule for selecting forecasts. *Operations Research*, 41(4):704–709, Jul 1993. `http://www.jstor.org/pss/171965`.

[FV97] Dean P. Foster and Rakesh V. Vohra. Calibrated learning and correlated equilibrium. *Games Econom. Behav.*, 21(1-2):40–55, 1997.

[FV99] Dean P. Foster and Rakesh V. Vohra. Calibration, expected utility and local optimality. Discussion Papers 1254, Northwestern University, Center for

Mathematical Studies in Economics and Management Science, March 1999. `http://ideas.repec.org/p/nwu/cmsems/1254.html`.

[Gal79] D. Gale. The game of Hex and the Brouwer fixed-point theorem. *Amer. Math. Monthly*, 86:818–827, 1979.

[Gar59] M. Gardner. The game of Hex. In *Hexaflexagons and Other Mathematical Diversions: The First Scientific American Book of Puzzles and Games*, pages 73–83. Simon and Schuster, 1959.

[Gar00] Andrey Garnaev. *Search games and other applications of game theory*, volume 485 of *Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag, Berlin, 2000. 19

[GBOW88] S. Goldwasser, M. Ben-Or, and A. Wigderson. Completeness theorems for non-cryptographic fault-tolerant distributed computing. In *Proc. of the 20th STOC*, pages 1–10, 1988.

[GDGJFJ10] Julio González-Díaz, Ignacio García-Jurado, and M. Gloria Fiestras-Janeiro. *An introductory course on mathematical game theory*, volume 115 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI; Real Sociedad Matemática Española, Madrid, 2010.

[Gea04] J. Geanakoplos. *Three Brief Proofs of Arrow's Impossibility Theorem.* Cowles Commission Monograph No. 1123R4. Cowles Foundation for Research in Economics, Yale University, Box 208281, New Haven, Connecticut 06520-8281., 1996 (updated 2004). `http://cowles.econ.yale.edu/`.

[Gin00] H. Gintis. *Game Theory Evolving; A problem-centered introduction to modeling strategic interaction.* Princeton University Press, Princeton, New Jersey, 2000.

[Gli52] I. L. Glicksberg. A further generalization of the Kakutani fixed theorem, with application to Nash equilibrium points. *Proc. Amer. Math. Soc.*, 3:170–174, 1952. 19

[GO80] E. Goles and J. Olivos. Periodic behaviour of generalized threshold functions. *Discrete Math.*, 30(2):187–189, 1980.

[Gra99] Peter Grassberger. Pair connectedness and shortest-path scaling in critical percolation. *J. Phys. A*, 32(35):6233–6238, 1999.

[Gro79] Theodore Groves. Efficient collective choice when compensation is possible. *The Review of Economic Studies*, 46(2):227–241, 1979.

[GS62] D. Gale and L. S. Shapley. College Admissions and the Stability of Marriage. *Amer. Math. Monthly*, 69(1):9–15, 1962.

[Hal35] Philip Hall. On representatives of subsets. *J. London Math. Soc*, 10(1):26–30, 1935.

[Han57] James Hannan. Approximation to bayes risk in repeated play. *Contributions to the Theory of Games*, 3:97–139, 1957. 42, 43

[Har68] Garrett Hardin. The tragedy of the commons. *science*, 162(3859):1243–1248, 1968.

[Har12] J. Hartline. *Approximation in economic design.* 2012.

[HMC00] Sergiu Hart and Andreu Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5):1127–1150, sep 2000. `http://www.jstor.org/pss/2999445`.

[HN12] Sergiu Hart and Noam Nisan. Approximate revenue maximization with multiple items. *arXiv preprint arXiv:1204.1846*, 2012.

[HS89] Sergiu Hart and David Schmeidler. Existence of correlated equilibria. *Mathematics of Operations Research*, 14(1):18–25, 1989. 19

[Isa55] Rufus Isaacs. The problem of aiming and evasion. *Naval Res. Logist. Quart.*, 2:47–67, 1955.

[Isa65] Rufus Isaacs. *Differential games. A mathematical theory with applications to warfare and pursuit, control and optimization.* John Wiley & Sons, Inc., New York-London-Sydney, 1965.

[Kar57] Samuel Karlin. An infinite move game with a lag. In *Contributions to the theory of games, vol. 3*, Annals of Mathematics Studies, no. 39, pages 257–272. Princeton University Press, Princeton, N. J., 1957.

[Kar59] S. Karlin. *Mathematical methods and theory in games, programming and economics*, volume 2. Addison-Wesley, 1959. 19

[Kle99] Paul Klemperer. Auction theory: A guide to the literature. *Journal of economic surveys*, 13(3):227–286, 1999.

[KN02] H. W. Kuhn and S. Nasar, editors. *The Essential John Nash.* Princeton University Press, 2002.

[Kon] Denés Konig. Gráfok és mátrixok. matematikai és fizikai lapok, 38: 116–119, 1931.

[Kri09] Vijay Krishna. *Auction theory.* Academic press, 2009.

[Las31] E. Lasker. *Brettspiele der Völker, Rätsel und mathematische Spiele.* Berlin, 1931.

[Leh64] Alfred Lehman. A solution of the Shannon switching game. *J. Soc. Indust. Appl. Math.*, 12:687–725, 1964.

[LLP$^+$99] Andrew J. Lazarus, Daniel E. Loeb, James G. Propp, Walter R. Stromquist, and Daniel H. Ullman. Combinatorial games under auction play. *Games Econom. Behav.*, 27(2):229–264, 1999.

[LLPU96] Andrew J. Lazarus, Daniel E. Loeb, James G. Propp, and Daniel Ullman. Richman games. In Richard J. Nowakowski, editor, *Games of No Chance*, volume 29 of *MSRI Publications*, pages 439–449. Cambridge Univ. Press, Cambridge, 1996.

[LP09] László Lovász and Michael D. Plummer. *Matching theory.* AMS Chelsea Publishing, Providence, RI, 2009. Corrected reprint of the 1986 original [MR0859549].

[LPSA94] Robert Langlands, Philippe Pouliot, and Yvan Saint-Aubin. Conformal invariance in two-dimensional percolation. *Bull. Amer. Math. Soc. (N.S.)*, 30(1):1–61, 1994.

[LPW09] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times.* American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson.

[LR57] R. Duncan Luce and Howard Raiffa. *Games and decisions: introduction and critical survey.* John Wiley & Sons, Inc., New York, N. Y., 1957. A study of the Behavioral Models Project, Bureau of Applied Social Research, Columbia University;.

[LW94] Nick Littlestone and Manfred K. Warmuth. The weighted majority algorithm. *Information and Computation*, 108(2):212–261, February 1994. 42

[Man96] Richard Mansfield. Strategies for the Shannon switching game. *Amer. Math. Monthly*, 103(3):250–252, 1996.

[Mar98] Donald A. Martin. The determinacy of Blackwell games. *J. Symbolic Logic*, 63(4):1565–1581, 1998.

[MG07] Jiri Matousek and Bernd Gärtner. *Understanding and using linear programming.* Springer Science & Business Media, 2007. 19

[Mil04] Paul Robert Milgrom. *Putting auction theory to work.* Cambridge University Press, 2004.

[MM05] Flávio Marques Menezes and Paulo Klinger Monteiro. *An introduction to auction theory*. Oxford University Press, 2005.

[Moo10] E. H. Moore. A generalization of the game called nim. *Ann. of Math.* (Ser. 2), 11:93–94, 1909–1910.

[MS96] Dov Monderer and Lloyd S. Shapley. Potential games. *Games Econom. Behav.*, 14(1):124–143, 1996.

[Mye81] Roger B Myerson. Optimal auction design. *Mathematics of operations research*, 6(1):58–73, 1981.

[Nas50] John F. Nash, Jr. Equilibrium points in n-person games. *Proc. Nat. Acad. Sci. U. S. A.*, 36:48–49, 1950.

[Nis07] Noam Nisan. *Algorithmic game theory*. Cambridge University Press, 2007. 42

[NS03] Abraham Neyman and Sylvain Sorin, editors. *Stochastic games and applications*, volume 570 of *NATO Science Series C: Mathematical and Physical Sciences*, Dordrecht, 2003. Kluwer Academic Publishers.

[O'N94] Barry O'Neill. Game theory models of peace and war. In *Handbook of game theory with economic applications, Vol. II*, volume 11 of *Handbooks in Econom.*, pages 995–1053. North-Holland, Amsterdam, 1994.

[OS04] Ryan O'Donnell and Rocco Servedio. On decision trees, influences, and learning monotone decision trees. Technical Report CUCS-023-04, Columbia University, Dept. of Computer Science, 2004. `http://www1.cs.columbia.edu/~library/2004.html`.

[OSSS05] Ryan O'Donnell, Mike Saks, Oded Schramm, and Rocco Servedio. Every decision tree has an influential variable. In *Proc. of the 46th Annual Symposium on Foundations of Computer Science (FOCS)*, 2005. `http://arxiv.org/abs/cs/0508071`.

[Owe67] Guillermo Owen. An elementary proof of the minimax theorem. *Management Sci.*, 13:765, 1967. 19

[Owe95] Guillermo Owen. *Game theory*. Academic Press, Inc., San Diego, CA, third edition, 1995. 19

[Pou11] William Poundstone. *Prisoner's dilemma*. Anchor, 2011.

[PS12a] Panagiota Panagopoulou and Paul Spirakis. Playing a game to bound the chromatic number. *Amer. Math. Monthly*, 114(5):373–387, 2012. `http://arxiv.org/math/0508580`.

[PS12b] Panagiota N. Panagopoulou and Paul G. Spirakis. Playing a game to bound the chromatic number. *Amer. Math. Monthly*, 119(9):771–778, 2012.

[PSSW07] Yuval Peres, Oded Schramm, Scott Sheffield, and David B. Wilson. Random-turn Hex and other selection games. *Amer. Math. Monthly*, 114(5):373–387, 2007. `http://arxiv.org/math/0508580`.

[PSSW09] Yuval Peres, Oded Schramm, Scott Sheffield, and David B. Wilson. Tug-of-war and the infinity Laplacian. *J. Amer. Math. Soc.*, 22(1):167–210, 2009. `http://arxiv.org/abs/math.AP/0605002`.

[Rei81] S. Reisch. Hex ist PSPACE-vollständig. *Acta Inform.*, 15:167–191, 1981.

[Ros73] R.W. Rosenthal. A class of games possessing pure-strategy Nash equilibria. *International Journal of Game Theory*, 2:65–67, 1973.

[Rub98] Ariel Rubinstein. *Modeling bounded rationality*, volume 1. MIT press, 1998.

[Rud64] Walter Rudin. *Principles of Mathematical Analysis*. MH, NY, third edition, 1964.

[Saa90] D. G. Saari. The Borda dictionary. *Social Choice and Welfare*, 7(4):279–317, 1990.

[Saa06] Donald G. Saari. Which is better: the Condorcet or Borda winner? *Social Choice and Welfare*, 26(1):107–129, 2006.

[Sha53] C. E. Shannon. Computers and automata. *Proc. Inst. Radio Eng.*, 41:1234–1241, 1953.

[Sha79] Adi Shamir. How to share a secret. *Comm. ACM*, 22(11):612–613, 1979.

[Sio58] Maurice Sion. On general minimax theorems. *Pacific J. Math.*, 8:171–176, 1958. 19

[Sky04] Brian Skyrms. *The stag hunt and the evolution of social structure.* Cambridge University Press, 2004.

[SL96] B. Sinervo and C. M. Lively. The rock-paper-scissors game and the evolution of alternative male strategies. *Nature*, 380:240–243, March 1996. `http://adsabs.harvard.edu/cgi-bin/nph-bib_query?bibcode=1996Natur.380..240S&db_key=GEN`.

[Smi01] S. Smirnov. Critical percolation in the plane. I. Conformal invariance and Cardy's formula. II. Continuum scaling limit, 2001. `http://www.math.kth.se/~stas/papers/percol.ps`.

[Spr36] R. Sprague. Über mathematische kampfspiele. *Tôhoku Math. J.*, 41:438–444, 1935–36.

[Spr37] R. Sprague. Über zwei abarten von Nim. *Tôhoku Math. J.*, 43:351–359, 1937.

[SS74] Lloyd Shapley and Herbert Scarf. On cores and indivisibility. *Journal of mathematical economics*, 1(1):23–37, 1974.

[SS05] Oded Schramm and Jeffrey E. Steif. Quantitative noise sensitivity and exceptional times for percolation, 2005. `http://arxiv.org/abs/math/0504586`.

[SS11] Shai Shalev-Shwartz. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4(2):107–194, 2011. 43

[Ste00] I. Stewart. Hex marks the spot. *Sci. Amer.*, 283:100–103, Sep. 2000.

[Stu02] Bernd Sturmfels. *Solving systems of polynomial equations*, volume 97 of *CBMS Regional Conference Series in Mathematics.* Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, 2002.

[SW57a] Maurice Sion and Philip Wolfe. On a game without a value. In *Contributions to the theory of games, vol. 3*, Annals of Mathematics Studies, no. 39, pages 299–306. Princeton University Press, Princeton, N. J., 1957.

[SW57b] Maurice Sion and Philip Wolfe. On a game without a value. In *Contributions to the theory of games, vol. 3*, Annals of Mathematics Studies, no. 39, pages 299–306. Princeton University Press, Princeton, N. J., 1957.

[SW01] Stanislav Smirnov and Wendelin Werner. Critical exponents for two-dimensional percolation. *Math. Res. Lett.*, 8(5-6):729–744, 2001. `http://arxiv.org/abs/math/0109120`.

[SW05] Oded Schramm and David B. Wilson. SLE coordinate changes. *New York J. Math.*, 11:659–669, 2005.

[TvS02] Theodore L Turocy and Bernhard von Stengel. Game theory. in encyclopedia of information systems. Technical report, CDAM Research Report LSE-CDAM-2001-09, 2002.

[Var07] Hal R Varian. Position auctions. *International Journal of Industrial Organization*, 25(6):1163–1178, 2007.

[Vic61] William Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance*, 16(1):8–37, 1961.

[vLW01] J. H. van Lint and R. M. Wilson. *A course in combinatorics.* Cambridge

University Press, Cambridge, second edition, 2001.

[vN28] J. v. Neumann. Zur Theorie der Gesellschaftsspiele. *Math. Ann.*, 100(1):295–320, 1928. 19

[vN53] John von Neumann. A certain zero-sum two-person game equivalent to the optimal assignment problem. In *Contributions to the theory of games, vol. 2*, Annals of Mathematics Studies, no. 28, pages 5–12. Princeton University Press, Princeton, N. J., 1953.

[vNM53] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behaviour*. Princeton University Press, Princeton, NJ., 3rd edition, 1953. 19

[vR00] J. van Rijswijck. Computer Hex: Are bees better than fruitflies? Master's thesis, University of Alberta, 2000.

[vR02] J. van Rijswijck. Search and evaluation in Hex, 2002. `http://www.cse.iitb.ac.in/~nikhilh/NASH/y-hex.pdf`.

[Wey50] Hermann Weyl. Elementary proof of a minimax theorem due to von Neumann. In *Contributions to the Theory of Games*, Annals of Mathematics Studies, no. 24, pages 19–25. Princeton University Press, Princeton, N.J., 1950. 19

[Wil86] J. D. Williams. *The Compleat Strategyst: Being a primer on the theory of games of strategy*. Dover Publications Inc., New York, second edition, 1986.

[Win69] Robert L. Winkler. Scoring rules and the evaluation of probability assessors. *Journal of the American Statistical Association*, 64:1073–1078, 1969. `http://www.jstor.org/pss/2283486`.

[WM68] Robert L. Winkler and Allan H. Murphy. "good" probability assessors. *J. Applied Meteorology*, 7(5):751–758, 1968.

[Wyt07] W. A. Wythoff. A modification of the game of Nim. *Nieuw Arch. Wisk.*, 7:199–202, 1907.

[Yan] Jing Yang. Hex solutions. `http://www.ee.umanitoba.ca/~jingyang/`.

[YLP03] J. Yang, S. Liao, and M. Pawlak. New winning and losing positions for Hex. In J. Schaeffer, M. Muller, and Y. Bjornsson, editors, *Computers and Games: Third International Conference, CG 2002, Edmonton, Canada, July 25-27, 2002, Revised Papers*, pages 230–248. Springer-Verlag, 2003.

[Zif99] Robert M. Ziff. Exact critical exponent for the shortest-path scaling function in percolation. *J. Phys. A*, 32(43):L457–L459, 1999.

Index