Matrix Games and Optimization

The theory of two-person games is largely the work of John von Neumann, and was developed somewhat later by von Neumann and Morgenstern [3] as a tool for economic analysis. Two-person zero-sum games provide a nice example of optimization and an opportunity to apply some of the linear algebra.

A two-person game is called a *constant-sum game* if the total payout is the same, each time the game is played. In such cases, we can subtract half the total payout from the payout to each player and record only the difference. Then the total payout appears to be zero, and such games are called *zero-sum games*. We can then suppose that whatever one player wins is paid by the other player. Except for the final section, we shall consider only two-person, zero-sum games.

1 Deterministic Solutions

In this two-person game, the first player, call him P1, selects a row of the I by J real matrix A, say i, and the second player selects a column of A, say j. The second player, call her P2, pays the first player A_{ij}. If some $A_{ij} < 0$, then this means that the first player pays the second. Since whatever the first player wins, the second loses, and vice versa, we need only one matrix to summarize the situation.

1.1 Optimal Pure Strategies

In our first example, the matrix is

$$A = \begin{bmatrix} 7 & 8 & 4 \\ 4 & 7 & 2 \end{bmatrix}. \tag{1.1}$$

The first player notes that by selecting row $i = 1$, he will get at least 4, regardless of which column the second player plays. The second player notes that, by playing column $j = 3$, she will pay the first player no more than 4, regardless of which row the first player plays. If the first player then begins to play $i = 1$ repeatedly, and the second player notices this consistency, she will still have no motivation to play any column except $j = 3$, because the other pay-outs are both worse than 4. Similarly, so long as the second player is playing $j = 3$ repeatedly, the first player has no motivation to play anything other than $i = 1$, since he will be paid less if he switches. Therefore, both players adopt a *pure strategy* of $i = 1$ and $j = 3$. This game is said to be *deterministic* and the entry $A_{1,3} = 4$ is a *saddle-point* because it is the maximum of its column and the minimum of its row. We then have

$$\max_i \min_j A_{ij} = 4 = \min_j \max_i A_{ij}.$$

Not all such two-person games have saddle-points, however.

1.2 Optimal Randomized Strategies

Consider now the two-person game with pay-off matrix

$$A = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix}. \tag{1.2}$$

The first player notes that by selecting row $i = 2$, he will get at least 2, regardless of which column the second player plays. The second player notes that, by playing column $j = 2$, she will pay the first player no more than 3, regardless of which row the first player plays. If both begin by playing in this conservative manner, the first player will play $i = 2$ and the second player will play $j = 2$.

If the first player plays $i = 2$ repeatedly, and the second player notices this consistency, she will be tempted to switch to playing column $j = 1$, thereby losing only 2, instead of 3. If she makes the switch and the first player notices, he will be motivated to switch his play to row $i = 1$, to get a pay-off of 4, instead of 2. The second player will then soon switch to playing $j = 2$ again, hoping that the first player sticks with $i = 1$. But the first player is not stupid, and quickly returns to playing $i = 2$. There is no saddle-point in this game.

For such games, it makes sense for both players to select their play at random, with the first player playing $i = 1$ with probability p and $i = 2$ with probability $1 - p$, and the second player playing column $j = 1$ with probability q and $j = 2$ with probability $1 - q$. These are called *randomized strategies*.

When the first player plays $i = 1$, he expects to get $4q + (1 - q) = 3q + 1$, and when he plays $i = 2$ he expects to get $2q + 3(1 - q) = 3 - q$. Since he plays $i = 1$ with probability p, he expects to get

$$p(3q + 1) + (1 - p)(3 - q) = 4pq - 2p - q + 3 = (4p - 1)q + 3 - 2p.$$

He notices that if he selects $p = \frac{1}{4}$, then he expects to get $\frac{5}{2}$, regardless of what the second player does. If he plays something other than $p = \frac{1}{4}$, his expected winnings will depend on what the second player does. If he selects a value of p less than $\frac{1}{4}$, and $q = 1$ is selected, then he wins $2p + 2$, but this is less than $\frac{5}{2}$. If he selects $p > \frac{1}{4}$ and $q = 0$ is selected, then he wins $3 - 2p$, which again is less than $\frac{5}{2}$. The maximum of these minimum pay-offs occurs when $p = \frac{1}{4}$ and the *max-min* win is $\frac{5}{2}$.

Similarly, the second player, noticing that

$$p(3q + 1) + (1 - p)(3 - q) = (4q - 2)p + 3 - q,$$

sees that she will pay out $\frac{5}{2}$ if she takes $q = \frac{1}{2}$. If she selects a value of q less than $\frac{1}{2}$, and $p = 0$ is selected, then she pays out $3 - q$, which is more than $\frac{5}{2}$. If, on the other hand, she selects a value of q that is greater than $\frac{1}{2}$, and $p = 1$ is selected, then she

will pay out $3q+1$, which again is greater than $\frac{5}{2}$. The only way she can be certain to pay out no more than $\frac{5}{2}$ is to select $q = \frac{1}{2}$. The minimum of these maximum pay-outs occurs when she chooses $q = \frac{1}{2}$, and the *min-max* pay-out is $\frac{5}{2}$.

This leads us to the question of whether or not there will always be probability vectors for the players that will lead to the equality of the max-min win and the min-max pay-out.

Exercise 1.1 *Suppose that there are two strains of flu virus and two types of vaccine. The first vaccine, call it V1, is* 0.85 *effective against the first strain (F1) and* 0.70 *against the second (F2), while the second vaccine (V2) is* 0.60 *effective against F1 and* 0.90 *effective against F2. The public health service is the first player, P1, and nature is the second player, P2. The service has to decide what percentage of the vaccines manufactured and made available to the public are of type V1 and what percentage are of type V2, while not knowing what percentage of the flu virus is F1 and what percentage is F2. Set this up as a matrix game and determine how the public health service should proceed.*

We make a notational change at this point. From now on the letters p and q will denote probability column vectors, and not individual probabilities, as in this section.

1.3 The Min-Max Theorem

Let A be an I by J pay-off matrix. Let

$$P = \{p = (p_1, ..., p_I) \,|\, p_i \geq 0, \sum_{i=1}^{I} p_i = 1\},$$

$$Q = \{q = (q_1, ..., q_J) \,|\, q_j \geq 0, \sum_{j=1}^{J} q_j = 1\},$$

and

$$R = A(Q) = \{Aq \,|\, q \in Q\}.$$

The first player selects a vector p in P and the second selects a vector q in Q. The expected pay-off to the first player is

$$E = \langle p, Aq \rangle = p^T Aq.$$

Let

$$m_0 = \max_{r \in R} \min_{p \in P} \langle p, r \rangle,$$

and

$$m^0 = \min_{p \in P} \max_{r \in R} \langle p, r \rangle.$$

Clearly, we have

$$\min_{p\in P}\langle p, r\rangle \leq \langle p, r\rangle \leq \max_{r\in R}\langle p, r\rangle,$$

for all $p \in P$ and $r \in R$. It follows that $m_0 \leq m^0$. The Min-Max Theorem, also known as the Fundamental Theorem of Game Theory, asserts that $m_0 = m^0$.

Theorem 1.1 The Fundamental Theorem of Game Theory *Let A be an arbitrary real I by J matrix. Then there are vectors $\hat{p}$ in P and $\hat{q}$ in Q such that*

$$p^T A\hat{q} \leq \hat{p}^T A\hat{q} \leq \hat{p}^T Aq, \tag{1.3}$$

for all p in P and q in Q.

The quantity $\omega = \hat{p}^T A\hat{q}$ is called the *value of the game*. Notice that if P1 knows that P2 plays according to the mixed-strategy vector $\hat{q}$, P1 could examine the entries $(A\hat{q})_i$, which are his expected pay-offs should he play strategy i, and select the one for which this expected pay-off is largest. It follows from the inequalities in (1.3) that

$$(A\hat{q})_i \leq \omega$$

for all i, and

$$(A\hat{q})_i = \omega$$

for all i for which $\hat{p}_i > 0$. However, if P2 notices what P1 is doing, she can abandon $\hat{q}$ to her advantage.

2 Non-Constant-Sum Games

In this section we consider non-constant-sum games. These are more complicated and the mathematical results more difficult to obtain than in the constant-sum games. Such non-constant-sum games can be used to model situations in which the players may both gain by cooperation, or, when speaking of economic actors, by collusion [1]. We begin with the most famous example of a non-constant-sum game, the Prisoners' Dilemma.

2.1 The Prisoners' Dilemma

Imagine that you and your partner are arrested for robbing a bank and both of you are guilty. The two of you are held in separate rooms and given the following options by the district attorney: (1) if you confess, but your partner does not, you go free, while he gets three years in jail; (2) if he confesses, but you do not, he goes free and you get the three years; (3) if both of you confess, you each get two years; (4)

if neither of you confesses, each of you gets one year in jail. Let us call you player number one, and your partner player number two. Let strategy one be to remain silent, and strategy two be to confess.

Your pay-off matrix is

$$A = \begin{bmatrix} -1 & -3 \\ 0 & -2 \end{bmatrix}, \tag{2.4}$$

so that, for example, if you remain silent, while your partner confesses, your pay-off is $A_{1,2} = -3$, where the negative sign is used because jail time is undesirable. From your perspective, the game has a deterministic solution; you should confess, assuring yourself of no more than two years in jail. Your partner views the situation the same way and also should confess. However, when the game is viewed, not from one individual's perspective, but from the perspective of the pair of you, we see that by sticking together you each get one year in jail, instead of each of you getting two years; if you cooperate, you both do better.

2.2 Two Pay-Off Matrices Needed

In the case of non-constant-sum games, one pay-off matrix is not enough to capture the full picture. Consider the following example of a non-constant-sum game. Let the matrix

$$A = \begin{bmatrix} 5 & 4 \\ 3 & 6 \end{bmatrix} \tag{2.5}$$

be the pay-off matrix for Player One (P_1), and

$$B = \begin{bmatrix} 5 & 6 \\ 7 & 2 \end{bmatrix} \tag{2.6}$$

be the pay-off matrix for Player Two (P_2); that is, $A_{1,2} = 4$ and $B_{2,1} = 7$ means that if P_1 plays the first strategy and P_2 plays the second strategy, then P_1 gains four and P_2 gains seven. Notice that the total pay-off for each play of the game is not constant, so we require two matrices, not one.

Player One, considering only the pay-off matrix A, discovers that the best strategy is a randomized strategy, with the first strategy played three quarters of the time. Then P_1 has expected gain of $\frac{9}{2}$. Similarly, Player Two, applying the same analysis to his pay-off matrix, B, discovers that he should also play a randomized strategy, playing the first strategy five sixths of the time; he then has an expected gain of $\frac{16}{3}$. However, if P_1 switches and plays the first strategy all the time, while P_2 continues with his randomized strategy, P_1 expects to gain $\frac{29}{6} > \frac{27}{6}$, while the expected gain of P_2 is unchanged. This is very different from what happens in the case of a constant-sum game; there, the sum of the expected gains is constant, and equals zero for a

zero-sum game, so P_1 would not be able to increase his expected gain, if P_2 plays his optimal randomized strategy.

2.3 An Example: Illegal Drugs in Sports

In a recent article in Scientific American [4], Michael Shermer uses the model of a non-constant-sum game to analyze the problem of doping, or illegal drug use, in sports, and to suggest a solution. He is a former competitive cyclist and his specific example comes from the Tour de France. He is the first player, and his opponent the second player. The choices are to cheat by taking illegal drugs or to stay within the rules. The assumption he makes is that a cyclist who sticks to the rules will become less competitive and will be dropped from his team.

Currently, the likelihood of getting caught is low, and the penalty for cheating is not too high, so, as he shows, the rational choice is for everyone to cheat, as well as for every cheater to lie. He proposes changing the pay-off matrices by increasing the likelihood of being caught, as well as the penalty for cheating, so as to make sticking to the rules the rational choice.

3 Learning the Game

In our earlier discussion we saw that the matrix game involving the pay-off matrix

$$A = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix} \tag{3.7}$$

is not deterministic. The best thing the players can do is to select their play at random, with the first player playing $i = 1$ with probability p and $i = 2$ with probability $1 - p$, and the second player playing column $j = 1$ with probability q and $j = 2$ with probability $1 - q$. If the first player, call him P1, selects $p = \frac{1}{4}$, then he expects to get $\frac{5}{2}$, regardless of what the second player, call her P2, does; otherwise his fortunes depend on what P2 does. His optimal mixed-strategy (column) vector is $[1/4, 3/4]^T$. Similarly, the second player notices that the only way she can be certain to pay out no more than $\frac{5}{2}$ is to select $q = \frac{1}{2}$. The minimum of these maximum pay-outs occurs when she chooses $q = \frac{1}{2}$, and the *min-max* pay-out is $\frac{5}{2}$.

Because the pay-off matrix is two-by-two, we are able to determine easily the optimal mixed-strategy vectors for each player. When the pay-off matrix is larger, finding the optimal mixed-strategy vectors is not a simple matter. As we have seen, one approach is to obtain these vectors by solving a related linear-programming problem. In this section we consider other approaches to finding the optimal mixed-strategy vectors.

3.1 An Iterative Approach

In [2] Gale presents an iterative approach to learning how best to play a matrix game. The assumptions are that the game is to be played repeatedly and that the two players adjust their play as they go along, based on the earlier plays of their opponent.

Suppose, for the moment, that P1 knows that P2 is playing the randomized strategy q, where, as earlier, we denote by p and q probability column vectors. The entry $(Aq)_i$ of the column vector Aq is the expected pay-off to P1 if he plays strategy i. It makes sense for P1 then to find the index i for which this expected pay-off is largest and to play that strategy every time. Of course, if P2 notices what P1 is doing, she will abandon q to her advantage.

After the game has been played n times, the players can examine the previous plays and make estimates of what the opponent is doing. Suppose that P1 has played strategy i n_i times, where $n_i \geq 0$ and $n_1 + n_2 + ... + n_I = n$. Denote by p^n the probability column vector whose ith entry is n_i/n. Similarly, calculate q^n. These two probability vectors summarize the tendencies of the two players over the first n plays. It seems reasonable that an attempt to learn the game would involve these probability vectors.

For example, P1 could see which entry of q^n is the largest, assume that P2 is most likely to play that strategy the next time, and play his best strategy against that play of P2. However, if there are several strategies for P2 to choose, it is still unlikely that P2 will choose this strategy the next time. Perhaps P1 could do better by considering his long-run fortunes and examining the vector Aq^n of expected pay-offs. In the exercise below, you are asked to investigate this matter.

Exercise 3.1 *Suppose that both players are attempting to learn how best to play the game by examining the vectors p^n and q^n after n plays. Devise an algorithm for the players to follow that will lead to optimal mixed strategies for both. Simulate repeated play of a particular matrix game to see how your algorithm performs. If the algorithm does its job, but does it slowly, that is, it takes many plays of the game for it to begin to work, investigate how it might be speeded up.*

References

[1] Dorfman, R., Samuelson, P., and Solow, R. (1958) *Linear Programming and Economic Analysis*. New York: McGraw- Hill.

[2] Gale, D. (1960) *The Theory of Linear Economic Models*. New York: McGraw-Hill.

[3] von Neumann, J., and Morgenstern, O. (1944) *Theory of Games and Economic Behavior*. New Jersey: Princeton University Press.

[4] Shermer, M. (2008) "The Doping Dilemma" *Scientific American*, April 2008, pp. 82–89.