Applications of Farkas alternative and Duality theorem

This handout gives three different applications of the Farkas alternative/Duality theorem. It reveals a well-known property of so-called Markov matrices, very popular in applied science; establishes equivalence between LPs and two-person games, and shows an application in finance, central in portfolio theory.

Elementary probability: little background. We will need most elementary facts from probability theory. A discrete probability space is the set $\{1, \ldots, n\}$ and a vector $p \in \mathbb{R}^n_+$, such that $\sum_{j=1}^n p_j = 1$. The elements of the set are usually called elementary events, or outcomes, and p_j the probability of the jth outcome. If all p_j are rational, let M be their least common multiple, let $p_j = M_j/M$. Then a "physical" model for a discrete probability space is having a bag containing M marbles in n different colours, with M_j marbles in each colour, and then pulling a marble out at random. If one wants to repeat the experiment, the marble that has been pulled out must be returned to the bag first.

If f is a function, whose value f_j depends only on the colour of the marble that gets pulled out, it is called a *random variable.* One can think of f simply as an n-vector $(f_1, \ldots, f_n)$. The *expectation*, or expected value, of f

$$E[f] = \sum_{j=1}^{n} f_j p_j = f \cdot p.$$

If the same experiment is repeated independently a great number N times, with the random outcomes $j_1, j_2, \ldots, j_N$, then the limit

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} f(j_k) = E[f],$$

and it converges pretty quickly with high probability. This is the meaning of the expectation: averaging over the results of many independent trials is the same as averaging once over the probability space.

Also, if we have two probability spaces $\{1, \ldots, m\}$ and $\{1, \ldots, n\}$ with probability vectors q and p, respectively, we can consider pairs of events (i, j) and assign to them probabilities $q_i p_j$, thereby calling the events i and j *independent.* A function $A(i, j)$ is represented by a matrix a_{ij}, and its expectation is

$$E[A] = \sum_{i=1}^{m} \sum_{j=1}^{n} q_i a_{ij} p_j = q \cdot Ap = p \cdot A^T q.$$

Markov matrices (optional reading). Markov matrices are square matrices, whose columns are probability vectors. Namely, $A \in \mathbb{R}^{n\times n}$ is Markov if its elements are non-negative and the sum of the elements in each column is one:

$$a_{ij} \geq 0,\ \forall\, i, j \quad \text{and} \quad \sum_{i=1}^{n} a_{ij} = 1,\ \forall\, j. \tag{1}$$

Markov matrices provide the simplest model for discrete random processes. For instance, consider a system (machine, creature), which can find itself in one of n distinct states. Let us further call them *pure states.* Suppose you never know exactly in which pure state the system is (as if happens all over the place ... you never know *for certain* in what kind of mood/state your partner is ... and all over quantum physics), but instead can only say that the system is in the pure state j with some probability p_j, $j = 1, \ldots, n$. So each $p_j \geq 0$ and $\sum_j p_j = 1$. The vector p is called *mixed state* and will coincide with a pure state when one of its components is 1 and the rest are 0. (Bold symbols are not used for vectors here – dimensions should be clear from the context.)

Now, some external factor, which is further referred to as a *kick* makes the system change its state. The effect of the kick also bears uncertainty and is expressed in terms of the matrix A as follows. If the system is in a pure state j, after the kick it will find itself in some other state i with probability a_{ij}. (Thus a_{ij} is a conditional probability.) But as you never know the original state of the system, after it has been kicked all you can say is that now it is described by a mixed state with the probability vector $p' = Ap$.

If you take *pretty much* any Markov matrix and *any* initial state p_0 and start kicking it again and again, it will turn out that very quickly the result will converge to a *steady state* p_* which will not depend on p_0, but will depend only on A. Mathematically, it is equivalent to the following statement: All eigenvalues of Markov matrices are at most one in the absolute value, *and* one of the eigenvalues always *equals* one. Moreover, the corresponding eigenvector is such that all its components are non-negative. "Pretty much" above is a non-degeneracy assumption: the unique steady state will exist if 1 is a simple (i.e. not multiple) eigenvalue which is true for almost all Markov matrices – let us not address this any further as we do nt need it for what we are about to show.

We first prove that if λ is an eigenvalue of A, then $|\lambda| \leq 1$. If λ is an eigenvalue and x is the corresponding eigenvector, then $Ax = \lambda x$. Without loss of generality, as eigenvectors are defined up to a multiplier – if x is a an eigenvector, so is Cx, for any C – we can assume $\sum_j |x_j| = 1$. Then

$$\begin{aligned} |\lambda| &= |\lambda| \sum_j |x_j| = \sum_j |\lambda x_j| = \sum_j |(Ax)_j| \\ &\leq \sum_{j,k} a_{jk}|x_k| = \sum_k |x_k| = 1, \end{aligned}$$

by (1). Recall, all $a_{jk} \geq 0$.

Now, let us show that there, in fact, exists $x \in \mathbb{R}^n_+ : x = Ax$ (indeed, not just an eigenvector, but a nonnegative one). We can use Farkas as follows: formulate the desired statement as the positive side of the alternative and show that the other side of the alternative is absurd. So, one side of Farkas is "$(A-I)x = 0$ (I is the $n \times n$ identity matrix) has a nonzero solution $x \geq 0$ and $x \cdot e = 1$, where $e = (1, \ldots, 1)$. Indeed, as an eigenvector is determined up to a multiplier, we can assume that x is a probability vector, so the sum of its components equals 1, which is expressed in terms of the dot product with e.

Denote

$$\tilde{A} = \begin{pmatrix} A - I \\ e^T \end{pmatrix}, \quad \tilde{b} = (0, 1) \in \mathbb{R}^{n+1}.$$

The Farkas alternative to $\tilde{A}x = \tilde{b}$ has a solution $x \geq 0$ is: $\exists \tilde{y} = (y, z) \in \mathbb{R}^{n+1}$ such that $\tilde{y} \cdot \tilde{b} < 0$, i.e. $z < 0$, and

$$\tilde{y}^T \tilde{A} \geq 0, \text{ i.e. } y^T(A - I) \geq -ze^T > 0.$$

Here comes the contradiction: $y^T(A-I) > 0$ implies that for every j, $y_j < y \cdot a^j$, where a^j is the jth column of A. Let some y_m be the maximum of all y_j, i.e. $y_m \geq y_j$, $\forall j$. We should in particular have $y_m < y \cdot a^m$, but this cannot happen, as $y \cdot a^m \leq y_m e \cdot a^m = y_m \sum_j a_{jm} = y_m$.

Zero-sum two-person games. A zero-sum two-person game is given by an $m \times n$ matrix A. There are two players P and Q. A single game goes as follows. Player P chooses a column $j \in \{1, \ldots, n\}$ and player Q independently chooses a row $i \in \{1, \ldots, m\}$. After the choices have been revealed, P gets a_{ij} pounds form Q. So, A is called the *payoff matrix*. The game is called *zero-sum* because the payoff for Q is given by $-A$ (or, in fact, $-A^T$ if Q now will become a column-player and P a row-player.)

E.g.

$$A = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \tag{2}$$

is a payoff matrix for a game where both P and Q secretly write "black" or "white" on a piece of paper and then show each other what they've written. If it is the same word, Q pays P a pound. Otherwise, P pays Q a pound.

The question that P asks is as follows. The game is going to be played many times. What is P's best strategy, so that no matter what Q does, P's expected payoff (over many games to be played) is *at least* V_1? And how to maximize V_1? The question what Q asks is what is Q's best strategy, so that no matter what P does, P's expected payoff is *no more* than V_2, and how to minimize V_2. Looks like linear programming.

Particular moves j for P and i for Q are called *pure strategies*. Sticking to the same pure strategy, i.e. the same move all the time is usually not great. E.g. in the example (2) if P always writes "white", Q has a strategy of always writing "black", so after N games Q wins N pounds. On the other hand, if Q always writes "black", then P can always write "black" as well, and win N pounds after N games. Hence, in terms of the above question of what the best strategy is, which was asked *regardless of what the opponent does*, pure strategies fail.

But what if every times he plays, P secretly tosses a fair coin and puts down "white" if it falls heads and "black" if it falls tails? There is nothing that Q can do about it. Whatever Q does, the expected payoff for P will be at least zero, and in fact, equal to zero. This is the solution of the game for P, and the same for Q.

To formalise it, a *mixed* strategy for P is a probability vector $p \in \mathbb{R}^n_+$. Each time the game is played, P will play a move j with probability p_j. A *mixed* strategy for Q is a probability vector $q \in \mathbb{R}^m_+$. Each time the game is played, Q will play a move i with probability q_i. The expected payoff of a single game for P is $q \cdot Ap$. The task for P to have an expected payoff of at least V_1, no matter what Q does is then

$$Ap \geq e^m V_1, \quad \max V_1, \tag{3}$$

where $e^m = (1, \dots, 1) \in \mathbb{R}^m$. Similarly, Q's task to have an expected payoff of at least V_2, no matter what P does is

$$A^T q \leq e^n V_2, \quad \min V_2, \tag{4}$$

where $e^n = (1, \dots, 1) \in \mathbb{R}^n$. It turns out that the two questions above are just the dual/primal pair of LPs. Here comes the main theorem:

Von Neumann equilibrium theorem. *There exist best strategies p^*, q^* for P and Q, such that $\max V_1 = \min V_2 = V$, and V is called the value of the game. I.e., every zero-sum two-person game has a solution in mixed strategies.*

Proof. Follows from strong duality for LPs. First off, let us add the same large number a to each entry of A, so they all become positive. If we can prove the theorem now and get the value V and optimal strategies p^*, q^*, the original problem will have the same best strategies and value $V - a$. So without loss of generality assume that all entries of A are positive. Then, as p, q are probability vectors (non-negative, components summing to 1), we can assume that both V_1 and V_2 are *positive*. Then we can divide the first set of inequalities (3) by V_1 and the second set of inequalities (4) by V_2. Denoting $x = p/V_1$, $y = q/V_2$, the two problems become:

$$Ax \geq e^m, \quad \min e^n \cdot x, \qquad A^T y \leq e^m, \quad \max e^m \cdot y. \tag{5}$$

Indeed, $e^n \cdot x = \sum_{j=1}^n p_j/V_1 = 1/V_1$, $e^m \cdot y = \sum_{i=1}^m q_i/V_2 = 1/V_2$. To maximize V_1 (minimize V_2) is to minimize $e^n \cdot x$ (maximize $e^m \cdot y$). But above we clearly have a diet/manufacturing dual/primal pair, and both problems are certainly feasible. Therefore, by Strong duality, there exist optimisers x^*, y^* delivering equal values, co $V_1 = V_2 = V$. Now $p^* = x^* V$, $q^* = y^* V$. Done.

As a practical recipe, given a game A, (i) add a large positive number a to each entry to get a positive matrix A'. Then (ii) solve one of the LPs in (5) to get, say x^*, then find shadow prices y^*. Observe here that the problem for y^* is the manufacturing problem, and therefore if you solve it via the simplex method, the shadow prices x^* will appear in the final tableau in the slack variables' columns in the objective row, so there is no need even for the dual simplex method. (iii) Now let $V' = \frac{1}{e^n \cdot x^*} = \frac{1}{e^m \cdot y^*}$. Then $p^* = x^* V'$, $q^* = y^* V'$. Finally, (iv) $V = V' - a$.

To apply it to the example (2), add 2 to each component of the payoff matrix, get for Q the LP max $y_1 + y_2$, s.t. $y_1 + 3y_2 \leq 1,\ y_2 + 3y_1 \leq 1$; solve graphically, get $y_1 = y_2 = 1/4$. Thus the value $V = \frac{1}{1/4+1/4} - 2 = 0$, and it is achieved when $q_1 = q_2 = 1/2$, same for P from symmetry.

Of special theoretical interest are *symmetric* games, where the payoff matrices for both players would look the same, i.e. $A = -A^T$ (so $m = n$). Such games, no wonder, have zero values and the best strategy p^* for P is best for Q, and, of course, from symmetry, the other way around.

To show that the value of the game is zero, look back at the conditions (3,4). By von Neumann's theorem $V_1 = V_2$, the game's value. So $Ve^n \leq Ap$ and $Ve^n \geq A^T q$. Dot-multiply the first identity by p and the second by q. Now use $p \cdot e^n = q \cdot e^n = 1$, as p, q are probability vectors, as well as the fact that $p \cdot Ap = A^T p \cdot p = -p \cdot Ap$, as $A^T = -A$, so $p \cdot Ap = q \cdot Aq = 0$. Thus $V \leq 0$ and $V \geq 0$, i.e. $V = 0$.

Therefore, for symmetric games von Neumann's theorem is even simpler: *a strategy p is optimal iff $Ap \geq 0$.*

To round things up, we have used linear programmes to solve games. Let us now show that every solvable primal/dual LP pair is equivalent to a zero-sum game. Once again the MP/DP form is better here, because it requires both x and y to be non-negative. Let us add to feasibility $Ax \leq b$, $A^T y \geq c$ the strong duality statement $c^T x - b^T y \geq 0$ (of course, only equality can occur there, by weak duality). In the matrix form this all is

$$\begin{bmatrix} 0 & -A & b \\ A^T & 0 & -c \\ -b^T & c^T & 0 \end{bmatrix} \begin{bmatrix} y \\ x \\ 1 \end{bmatrix} \geq 0$$

The big matrix $\tilde{A}$ now has the property $\tilde{A}^T = -\tilde{A}$. If we have a feasible solution $(y, x, 1)$ to this system of inequalities, then its positive multiple will also be a solution, so $(y, x, 1)$ can be scaled into a probability vector. We've shown that solvable LPs are reducible to symmetric games.

Finance: main theorem of asset pricing. Suppose, one has n financial assets $S_1, \ldots, S_n$ (bonds, stocks, etc., as well as debts which can be sold as well...), and today $p_j \in \mathbb{R}$ is a price of the jth asset S_j. (Assume one can sell and buy freely, without any extra charges.) Suppose, tomorrow the world may find itself in one of m possible states. In which state the world will be – we don't know, but each may come with positive probability. What we do know, however, is that $a_{ij} \in \mathbb{R}$ would be the value of the jth asset in the state of the world i. The $m \times n$ matrix A with entries the a_{ij} will be called the *payoff* matrix. Such a discrete model is called in finance the *Arrow-Debreu* model. See e.g. http://dybfin.wustl.edu/research/papers/arbetc7.pdf for economist's description.

Suppose, we acquire θ_j units of asset j. The n-vector θ is further referred to as a *portfolio.* In fact, $\theta \in \mathbb{R}^n$, not in $\mathbb{R}^n_+$ i.e one is allowed to assume "short" positions as well. (Practically, this means borrowing θ_j units of asset j today under an obligation of returning them in kind tomorrow, and immediately selling them today. Tomorrow one will then buy the asset at a new price and return it to the lender. This is called selling short. If the security goes down in price, one makes a profit.) Today's cost of the portfolio is $p \cdot \theta$. Tomorrow its possible values are $A\theta$, the ith component being the cost if the world finds itself in the state i. A portfolio is called *arbitrage*, or "free lunch" if either today its cost $p \cdot \theta$ is *negative*, while in all states of tomorrow it is *non-negative,* or if today's cost is zero, while in *all* states of tomorrow it is *non-negative* and in *some* state it is *positive.*

Mathematically, the existence of arbitrage is clearly expressed as follows. With

$$B = \begin{pmatrix} -p^T \\ A \end{pmatrix}, \quad \exists\, \theta \in \mathbb{R}: \quad B\theta \geq 0, \;\; \text{and for some } i = 1, \ldots, m+1, \;\; (B\theta)_i > 0. \tag{6}$$

Main theorem of asset pricing (optional). *There is no arbitrage if and only if there exists $q \in \mathbb{R}^m_{++}$ (i.e. $q_i > 0$, $\forall\, i = 1, \ldots, m$) such that*

$$p = A^T q. \tag{7}$$

The vector q is called *state price vector*, or linear pricing rule. Suppose, one of our assets, say S_1 a risk-free asset, such as one pound sterling under zero inflation (if inflation or interest rates exist they can easily be taken into account as well). Then all entries the first column of A, as well as p_1 equal 1. The state price vector q, if there is no free lunch, becomes then a probability vector: $\sum_{i=1}^m q_i = 1$. So the rough meaning of the theorem is that there is no free lunch if and only if today's asset pricing has taken into account all possible scenarios for tomorrow with *positive* probabilities. (Alternatively, if there is something about tomorrow that you know and nobody else knows, so it has not been reflected in today's asset pricing but tomorrow it will affect them – this knowledge entitles you to a free lunch.)

Proof: consists in showing that the Farkas alternative to (6) is "there exists $y \in \mathbb{R}^{m+1}_{++}$ such that $B^T y = 0$. If this is shown, since y will be determined up to a positive multiplier, one can assume that $y_1 = 1$, i.e. let $y = (1, q)$. Then the condition $B^T y = 0$ becomes precisely (7). So (6) and (7) are the opposite sides of Farkas, cannot happen together, and one *must* happen.

So, all what's left is to find the Farkas alternative to (6). Observe that the condition of positivity of some component $(B\theta)_i$ and non-negativity of the rest of the components in (6), as θ there is defined up to a positive multiplier, can be interpreted as $e \cdot B\theta = 1$, where $e = (1, \ldots, 1) \in \mathbb{R}^{m+1}$. To take care of technicality before we apply Farkas: as $\theta \geq 0$ is *not* required, we have to let $\theta = u - v$, with $u, v \in \mathbb{R}^n_+$. Besides, the condition $B\theta \geq 0$ is equivalent to $B\theta - Is = 0$, where $s \in \mathbb{R}^{m+1}_+$ and I is the identity. (You may recall the general principle: dealing with sign-unconstrained variables on one sides results in equations, rather than inequalities on the other (dual) side; dealing with inequalities rather than equations on one side results in sign constraints for the dual variables).

So the front side of the Farkas alternative, corresponding to (6) is $\tilde{A}x = b$ has a solution $x \geq 0$, with $x = (u, v, s) \in \mathbb{R}^{2n+m+1}_+$, $b = (1, 0) \in \mathbb{R}^{1+m}$, and

$$\tilde{A} = \begin{pmatrix} e^T B & -e^T B & 0 \\ B & -B & -I \end{pmatrix}.$$

The opposite side of Farkas is: there exists $\tilde{y} = (z, y) \in \mathbb{R}^{1+m}$, such that $z < 0$ and

$$\begin{cases} z e^T B + y^T B \geq 0, \\ -z e^T B - y^T B \geq 0, \\ y \leq 0. \end{cases}$$

This means $-B^T(y + ze) = 0$, with $y \leq 0$, $z < 0$. Changing y to $y' = -(y + ze) > 0$ now proves the claim: $y' > 0$ and $By' = 0$. Replace the notation y' with y and we are done.