Appendix A: Algorithms

Pivoting

The pivot operation was presented in Chapter 2 in terms of substitution in the tableau equations. The same result is obtained more mechanically as follows:

Let r be the **pivot row** (the row labelled by the variable leaving the basis) and c be the **pivot column** (the column labelled by the variable entering the basis). Let a_{ij} be the entry in row i and column j of the original tableau, and a'_{ij} be the corresponding entry of the new tableau.

(1) $a'_{rc} = 1/a_{rc}$.

(2) For each j except c, $a'_{rj} = a_{rj}a'_{rc}$.

(3) For each i except r, $a'_{ic} = -a_{ic}a'_{rc}$.

(4) For each i except r and each j except c, $a'_{ij} = a_{ij} + a_{rj}a'_{ic}$.

(5) Interchange the labels of row r and column c.

LINEAR will refuse to perform a pivot if a_{rc} is too close to zero.

The Simplex Method

The discussion below assumes that we are dealing with a maximization problem. For a minimization problem, all references to signs in the objective row would be reversed.

LINEAR's version of the Simplex Method involves three phases:

Phase 0 removes artificial variables from the basis.

Phase 1 finds a basic feasible solution.

Phase 2 finds an optimal solution.

It is easiest to understand these in reverse order.

Phase 2. Consider the first tableau of our example from Chapter 2:

	x_1	x_2	x_3	
p	-20	-30	-15	0
m_1	1.5	1	1	100
m_2	1	2	1	100

There are no artificial variables, and the basic solution for this tableau, obtained by setting all the nonbasic variables to 0, is feasible. Thus we can start with Phase 2. Now instead of the basic solution, consider a solution where x_2 is assigned a positive value, while the other nonbasic variables stay at 0. According to the first equation of the tableau,

$$p - 20x_1 - 30x_2 - 15x_3 = 0$$

so p would increase by 30 for every unit increase of x_2. This is good news, since we want to make p as large as possible. However, there are limits to the amount of increase we can obtain in this way. The reason is that the values of m_1 and m_2 are also affected by increasing x_2. If x_2 was made too large, m_1 or m_2 could become negative and the solution would no longer be feasible. According to the tableau equations, m_1 becomes 0 when $x_2 = 100/1 = 100$, while m_2 becomes 0 when $x_2 = 100/2 = 50$. Since neither can be allowed to become negative, the best feasible solution that can be obtained in this way (increasing x_2 while keeping the other nonbasic variables at 0) is $x_1 = 0, x_2 = 50, x_3 = 0, m_1 = 50, m_2 = 0, p = 1500$.

This new solution is also a basic solution, not for the original tableau but for another one. Since x_2 is no longer 0, it must be a basic variable in the new tableau. On the other hand, m_2 has become 0, so it could be nonbasic. In fact, the new tableau is the one that we obtained earlier by pivoting, x_2 entering the basis and m_2 leaving. The discussion above provides motivation for performing that particular pivot.

It is instructive to see what would happen if some of the entries in the x_2 column of the tableau were changed in sign. If the entry in the p row and x_2 column were positive instead of negative, increasing x_2 would decrease p instead of increasing it. Therefore we would not want x_2 to enter the basis at this point, but instead would choose a different nonbasic variable with a negative entry in the p row. What if none of the nonbasic variables had a negative entry in the p row? Then there would be no way at all to increase p by changing the nonbasic variables while keeping them ≥ 0, and the current basic solution would be optimal.

Next, suppose instead that the entry in the m_2 row and x_2 column were negative instead of positive. Then increasing x_2 would increase m_2, so that m_2 would not put a limit on this increase. Only m_1 would provide such a limit, and so the next tableau would have x_2 basic and m_1 nonbasic.

Finally, suppose that the entries in the x_2 column in both m_1 and m_2 rows were negative instead of positive. Then there would be no limit to the amount x_2 (and consequently p) could increase. The problem would be **unbounded**.

From this example, we can see the strategy for Phase 2 of the simplex method:

(1) Choose a nonbasic, non-artificial variable x_E with a negative entry in the objective row.

— If there is no such x_E, conclude that the current basic solution is optimal.

(2) Divide each positive element in the x_E column into the corresponding element in the constants column (obtaining the value of x_E which would drive the corresponding basic variable to 0). Let x_L be the basic variable for the row where the minimum quotient is obtained.

— If there are no positive elements in the x_E column, conclude that the problem is unbounded because there is no limit to the amount x_E can be increased.

(3) Pivot with x_L leaving the basis and x_E entering, and return to step (1).

This strategy is (almost) guaranteed to stop eventually, either with an optimal solution or with the conclusion that the problem is unbounded: the number of possible choices for basic variables is finite, and if the value of p keeps increasing each time we pivot, the process must stop eventually. If it stops in step (1) we have an optimal solution, while if it stops in step (2) we can conclude that the problem is unbounded.

The reason I say "almost" is that if some basic variable happens to be 0, and the entry in this variable's row and the x_E column is positive, then the allowed amount of increase of x_E according to step (2) is 0. The pivot will make a new basis and a new tableau, but the values of the variables in the basic solution will still be exactly the same. This situation is called **degeneracy**.

Most of the time degeneracy is harmless: a few pivots might leave the solution unchanged, but eventually either a non-degenerate pivot will be found, the current solution will be seen to be optimal, or the problem will be seen to be unbounded. However, there is the theoretical possibility of the method **cycling**, repeating an endless loop through the same set of tableaus. This is quite rare in practice, but does occasionally happen in problems with a large amount of degeneracy. There are several methods of preventing cycling. LINEAR uses the **perturbation** method, which has advantages of simplicity and speed. The command '`Change Add Perturbation`' adds a small random amount (between 0 and 1/65536) to each RHS entry. This almost certainly removes the degeneracy, so that it will be easy to find the optimal solution to the new problem. Then the perturbation can be removed, resulting (usually) in the optimal solution to the original problem. There is a slight possibility that the solution obtained will be slightly infeasible, so that we would have to continue the solution process.

The next question to consider is which entering variable to select if there are several candidates. A number of methods of doing this have been used. LINEAR's approach is to choose the one that will produce the largest change in the objective function.

Phase 1. Consider the following tableau.

	x_1	x_2	x_3	
p	-20	-30	-15	0
m_1	-1.5	1	-1	-50
m_2	1	-2	-1	-100
m_3	1	1	1	80

Here there are no artificial variables, but the basic solution is not feasible. Basic variables will be called "negative" or "nonnegative" if their values in the current basic solution are < 0 or ≥ 0, respectively. We would like to increase the negative basic variables m_1 and m_2, if possible making them positive. The **target row** method of Phase 1 starts by choosing one of the negative basic variables (e.g. m_2) as a "temporary objective": its row is the "target row". A nonbasic, non-artificial variable with a negative entry in the target row (e.g. x_3) is chosen to enter the basis.

The choice of variable to leave the basis is more complicated than in Phase 2, because negative and nonnegative basic variables must be considered separately. Consider what happens as x_3 is increased. A negative basic variable hitting 0 constitutes progress towards a feasible solution. This would happen for m_1 at $x_3 = 50$ and for m_2 at $x_3 = 100$. There would be no advantage in increasing x_3 past 100, where there are no more negative basic variables approaching 0. Thus, considering the negative basic variables only, we would wish to increase x_3 to 100. Unfortunately, however, that would make the nonnegative basic variable m_3 negative. Since the goal of Phase 1 is to reduce the number of negative basic variables, x_3 must not be increased past the value 80 that makes the nonnegative basic variable m_3 hit 0. Thus x_3 should be increased to 80. A pivot is performed with x_3 entering and m_3 (the variable that hits 0 when $x_3 = 80$) leaving the basis.

A positive entry in the entering column indicates that the corresponding basic variable decreases as the entering variable increases, while a negative entry indicates that it increases. We can ignore negative basic variables with a positive or zero entry in the entering column, and nonnegative basic variables with a negative or zero entry. Let a be the largest value of the entering variable for which a negative basic variable hits 0, and let b be the smallest value for which a nonnegative basic variable hits 0 (or $+\infty$ if there are none). Then we will want to increase the entering variable to the minimum of a and b. The basic variable that produced this minimum value will leave the basis.

After the pivot, the same target row can be kept if its RHS entry is still negative. If the RHS entry is no longer negative, another target row must be selected. When there are no possible targets, Phase 1 is done — the basic solution is feasible, and Phase 2 may begin.

It may happen that there is no possible choice for the entering variable, i.e. every nonbasic non-artificial variable has a nonnegative entry in the target row. This indicates that the problem is **infeasible**. To see this, consider the equation corresponding to the target row. For example, if the -2 and -1 in the m_2 row of our example were changed in sign, this equation would be

$$m_2 + x_1 + 2x_2 + x_3 = -100$$

This equation would have to be true in any feasible solution. However, the left side would have to be ≥ 0 (since the variables are ≥ 0 in a feasible solution), while the right side is < 0. Thus there can be no feasible solutions.

Here, then, is the strategy for Phase 1:

(1) Choose a target row (not the objective) with a negative RHS entry.

— If there are none, go on to Phase 2.

(2) Choose a nonbasic, non-artificial variable x_E with a negative entry in the target row.

— If there is no such x_E, conclude that the problem is infeasible.

(3) For each row with negative entries in both the RHS and the x_E column, divide the RHS entry by the x_E entry. Let a be the maximum of these quotients. (There will be at least one, arising from the target row)

(4) For each row with nonnegative entry in the RHS and positive entry in the x_E column, divide the RHS entry by the x_E entry. Let b be the minimum of these quotients, or $+\infty$ if there are none.

(5) Take the minimum of a and b. Let x_L be the basic variable for the row where this value occurred.

(6) Pivot with x_L leaving the basis and x_E entering.

(7) If the RHS entry for the target row is still negative, go to step (2). Otherwise, go to step (1).

This strategy is (almost) guaranteed to stop eventually, since at each pivot either

(a) the number of negative basic variables decreases, or

(b) the number of negative basic variables stays the same, and the value of the target variable increases. Again the "almost" refers to the remote possibility of cycling.

LINEAR actually uses a variation of the target row method: the target row is actually obtained by adding together up to 10 possible target rows. The rationale behind this is to attempt to improve many negative basic variables at a time, rather than perhaps just one.

Phase 0. Here there are artificial variables in the basis. We want to remove them from the basis. We can assume, for convenience, that the RHS entries in all the artificial rows are ≤ 0 — this can be arranged by multiplying those rows with positive RHS entries by -1. That is equivalent to changing the sign of the corresponding basic variable, and can be reversed later. Now the situation can be compared to that of Phase 1 — we want to increase the negative basic artificial variables to 0, while not making any nonnegative basic variables negative. We select a basic artificial variable as "target", and choose an entering variable with negative coefficient in that row. The main difference with Phase 1 is that we do not want artificial variables to become positive. Thus in the same example considered for Phase 1, if m_1 and m_2 were artificial we would only increase x_3 to 50, and m_1 would leave the basis.

Another difference with Phase 1 is that if there is no entering variable, the problem need not be infeasible if the RHS entry in the target row is zero. Such a row would correspond to an equation such as

$$m_2 + x_1 + 2x_2 + x_3 = 0$$

This simply means that m_2, x_1, x_2 and x_3 must all be 0 in any feasible solution. To get the artificial variable m_2 out of the basis, we could pivot with any of x_1, x_2 and x_3 entering the basis. Another case that could arise would have zero entries in the target row for the RHS and all non-artificial nonbasic variables. In this case the artificial variable is automatically 0 in any feasible solution (and the corresponding equality is a logical consequence of the other equality constraints). It is then harmless to leave the artificial variable in the basis.

Here, then, is a strategy for Phase 0:

(0) Change signs in any row corresponding to an artificial variable where the RHS entry is positive.

(1) Choose a target row corresponding to a basic artificial variable. If there are none, go on to Phase 2.

(2) Choose a nonbasic, non-artificial variable x_E with a negative entry in the target row.

— If there are none, then

— if the RHS entry in the target row is not 0, conclude that the problem is infeasible.

— if the RHS entry is 0 but there are nonzero entries in non-artificial columns, pivot on one of them and go to step 1.

— otherwise ignore this artificial variable and go to step 1.

(3) For each artificial row with negative entry in the x_E column and each row with nonnegative entry in the RHS and positive entry in the x_E column, divide the RHS entry by the x_E entry. Take the minimum of these quotients. Let x_L be the basic variable for the row where this value occurred.

(4) Pivot with x_L leaving the basis and x_E entering.

(5) If the pivot was in the target row, go to step (1). Otherwise go to step (2).