Optimization I	**Summer 2011** **Lecture 5**

The Simplex Method

1 An iteration of the simplex method

1.1 Preliminaries

The following facts will be useful:

- Let $A \in \mathbb{R}^{m \times n}, d \in \mathbb{R}^n$. Then $Ad = \sum_{j=1}^n A_j d_j$.
- Supppose $B(1), \ldots, B(m)$ are indices such that $d_j = 0$ if $j \neq B(1), \ldots, B(m)$. Let d_B denote the vector $\begin{bmatrix} d_{B(1)} \\ \vdots \\ d_{B(m)} \end{bmatrix}$, and let $\mathbf{B} = \begin{bmatrix} | & | & & | \\ A_{B(1)} & A_{B(2)} & \ldots & A_{B(m)} \\ | & | & & | \end{bmatrix}$.
  Then $Ad = \mathbf{B} d_B$.
- If $\mathbf{B}$ is an invertible matrix, then $\mathbf{B}^{-1} A_{B(i)} = e_i$, where e_i is the i-th unit vector.

We also recall from last time that x is a basic feasible solution to an LP in standard form, if and only if x satisfies $Ax = b, x \geq 0$, and there exist indices $B(1), \ldots, B(m)$ such that $\mathbf{B} = \begin{bmatrix} | & | & & | \\ A_{B(1)} & A_{B(2)} & \ldots & A_{B(m)} \\ | & | & & | \end{bmatrix}$ contains m linearly independent columns, and $x_j = 0$ if $j \neq B(1), \ldots, B(m)$.

1.2 Choosing which nonbasic variable to increase

Let d be a vector, which has a 1 in position j and 0 for every other nonbasic variable. We consider the vector $x + \theta d$ – this is the vector that increases x_j by θ, and keeps all other nonbasic variables at 0 (and we will determine what it does to the basic variables in a moment).

We want $x + \theta d$ to be a basic feasible solution, that is better than our current bfs x.

Let's first check what we need to do to make sure $x + \theta d$ is feasible:

- We need $A(x + \theta d) = b$. But our current solution is feasible, i.e., $Ax = b$. So we need $Ad = 0$. Recall that d has a 1 in position j and 0 for every other nonbasic variable, and we have not yet determined the other entries. So $Ad = \mathbf{B} d_B + A_j$. Hence, we need $\mathbf{B} d_B = -A_j$, or (since $\mathbf{B}$ is invertible!) $d_B = -\mathbf{B}^{-1} A_j$.
- We need $x + \theta d \geq 0$. For the nonbasic variables, we are sure this holds for any $\theta \geq 0$. For the basic variables, we need $x_B + \theta d_B \geq 0$. There are two things that can happen:
  - There exists some i such that $x_{B(i)} = 0$ and $d_{B(i)} < 0$. Then, we must let $\theta = 0$. Note that it must be the case that the current bfs is degenerate, since $x_{B(i)} = 0$!
  - There exists no i such that $x_{B(i)} = 0$ and $d_{B(i)} < 0$. Then, we can find some $\theta > 0$, such that $x + \theta d$ is a feasible solution.

How does moving in the direction d change the objective value?

$$c^T(x + \theta d) = c^T x + \theta c^T d = c^T x + \theta(c_j + c_B^T d_B) = c^T x + \theta(c_j - c_B^T \mathbf{B}^{-1} A_j).$$

We note a couple of things:

- You may have expected that increasing x_j by θ increases the objective by θx_j, but we see it only increases the objective by $\theta(c_j - c_B^T \mathbf{B}^{-1} A_j)$! This is caused by the fact that we have to simultaneously change the basic variables, to ensure that $Ax = b$.
- We call $c_j - c_B^T \mathbf{B}^{-1} A_j$ the **reduced cost** of the variable j, and denote it by $\bar{c}_j$.
- Note that if $j = B(i)$, then $\bar{c}_j = 0$, since $\mathbf{B}^{-1} A_{B(i)} = e_i$, so $c_j - c_B^T e_i = c_j - c_{B(i)} = 0$.

Theorem 1. *Let $A_{B(1)}, \ldots, A_{B(m)}$ be a basis, let x be the corresponding basic feasible solution and let $\bar{c}_j = c_j - c_B^T \mathbf{B}^{-1} A_j$ for $j = 1, \ldots, n$. If x is not optimal, then there exists some j such that $\bar{c}_j < 0$.*

Proof. Suppose y is feasible and $c^T y < c^T x$.

Let $d = y - x$. We know three things: $Ad = 0$, $c^T d < 0$, and $d_j \geq 0$ for $j \neq B(1), \ldots, B(m)$.

So

$$\mathbf{B} d_B + \sum_{j \neq B(1), \ldots, B(m)} A_j d_j = 0 \Leftrightarrow d_B = -\mathbf{B}^{-1} \sum_{j \neq B(1), \ldots, B(m)} A_j d_j$$

and

$$c^T d < 0 \Leftrightarrow \sum_{j \neq B(1), \ldots, B(m)} c_j d_j + c_B^T d_B < 0.$$

We can substitute $d_B = -\mathbf{B}^{-1} \sum_{j \neq B(1), \ldots, B(m)} A_j d_j$ into this inequality, to get

$$\begin{aligned} & \sum_{j \neq B(1), \ldots, B(m)} c_j d_j - c_B^T \mathbf{B}^{-1} \sum_{j \neq B(1), \ldots, B(m)} A_j d_j < 0 \\ \Leftrightarrow \quad & \sum_{j \neq B(1), \ldots, B(m)} d_j (c_j - c_B^T \mathbf{B}^{-1} A_j) < 0 \\ \Leftrightarrow \quad & \sum_{j \neq B(1), \ldots, B(m)} d_j \bar{c}_j < 0. \end{aligned}$$

So, since $d_j \geq 0$ for $j \neq B(1), \ldots, B(m)$ it cannot be the case that $\bar{c}_j \geq 0$ for all j. □

1.3 Moving to a new bfs from a (non-degenerate) bfs.

In the previous subsection, we found out how to determine if there is a nonbasic variable x_j, such that increasing x_j will improve the solution: we just check if $\bar{c}_j = c_j - c_B^T \mathbf{B}^{-1} A_j < 0$.

Suppose we find j such that $\bar{c}_j < 0$, and suppose the current solution x is non-degenerate. Let d be the vector we described above, with $d_j = 1, d_{j'} = 0$ for every $j' \neq j, B(1), \ldots, B(m)$ and $d_B = -\mathbf{B}^{-1} A_j$. Let $x + \theta d$. How large can θ be? We need

$$x_B + \theta d_B \geq 0$$

- if $d_B \geq 0$, then we let $\theta^* = \infty$. Since $\bar{c}_j < 0$, this means the objective decreases by $\bar{c}_j \theta^*$ – in other words, the optimal objective value is $-\infty$.
- Otherwise, let

  $$\theta^* = \min_{i = 1, \ldots, m : d_{B(i)} < 0} -x_{B(i)} / d_{B(i)}.$$

  Note that $\theta^* > 0$ because we assumed that x is non-degenerate.

In the first case, we have detected that the optimal value is $-\infty$. In the second case, we move to the new solution $x^{new} = x + \theta^* d$.

- Note that x_j was non-basic in the old solution, and that $x_j^{new} > 0$. We say that j **enters the basis**.

- Note that it must be the case that there is some ℓ such that $x_{B(\ell)} > 0$ and $x^{new}_{B(\ell)} = 0$: Indeed, there is some ℓ such that $\theta^* = -x_{B(\ell)}/d_{B(\ell)}$, and for this ℓ, $x^{new}_{B(\ell)} = 0$. We say that $x_{B(\ell)}$ **leaves the basis**.

- So, if our old basis matrix is $\mathbf{B} = \begin{bmatrix} | & | & & | & & | \\ A_{B(1)} & A_{B(2)} & \ldots & A_{B(\ell)} & \ldots & A_{B(m)} \\ | & | & & | & & | \end{bmatrix}$. Then our new basis matrix is $\mathbf{B}^{new} = \begin{bmatrix} | & | & & | & & | \\ A_{B(1)} & A_{B(2)} & \ldots & A_j & \ldots & A_{B(m)} \\ | & | & & | & & | \end{bmatrix}$.

Lemma 2. *$\mathbf{B}^{new}$ is a basis corresponding to the new solution $x^{new} = x + \theta^* d$.*

Proof. Note that $\mathbf{B}^{new}$ contains m columns, and that it contains all the columns that correspond to the non-zero variables in x^{new}. To check that the columns are linearly independent, we must check that A_j is not a linear combination of $A_{B(1)}, \ldots, A_{B(\ell-1)}, A_{B(\ell+1)}, \ldots, A_{B(m)}$.

Suppose by contradiction it is a linear combination. Then $A_j = \sum_{k \neq \ell} \gamma_k A_{B(k)}$ or $\mathbf{B}^{-1} A_j = \sum_{k \neq \ell} \gamma_k \mathbf{B}^{-1} A_{B(k)}$. Note that $\mathbf{B}^{-1} A_{B(k)}$ is the k-th unit vector e_k, so the right hand side is $\sum_{k \neq \ell} \gamma_k e_k$. Note that the right hand side must have a 0 in the ℓ-th entry, since e_ℓ is not part of the sum. On the other hand, the left hand side is $-d_B$ and we know that $d_{B(\ell)} > 0$, so its ℓ-th entry is non-zero. $\square$

1.4 An iteration of the simplex method

We can now put together an iteration of the simplex method:

1. We start with a basis $A_{B(1)}, \ldots, A_{B(m)}$ and an associated bfs x.
2. We compute $\bar{c}_j = c_j - c_B^T \mathbf{B}^{-1} A_j$ for every non-basic variable. If $\bar{c}_j \geq 0$ for all j, we conclude x is optimal, and the algorithm terminates.
3. Otherwise, choose j such that $\bar{c}_j < 0$. Variable x_j enters the basis.
4. Let $u = B^{-1} A_j$ (note that this is $-d_B$). If $u_i \leq 0$ for all i, then the optimal value is $-\infty$, and the algorithm terminates.
5. Otherwise, set $\theta^* = \min_{i: u_i > 0} x_{B(i)}/u_i$, and let ℓ be such that $\theta^* = x_{B(\ell)}/u_\ell$. Variable $x_{B(\ell)}$ leaves the basis.
6. The new basis is obtained by replacing $A_{B(\ell)}$ by A_j, and the new bfs by setting $x^{new}_j = \theta^*$ and $x^{new}_{B(i)} = x_{B(i)} + \theta^* u_i$ for $i = 1, \ldots, m$.

Theorem 3. *Suppose we have an initial basic feasible solution and there is no degeneracy. Then the simplex method terminates in finite time and it either finds either an optimal solution, or it correctly reports that the optimal value is $-\infty$.*

Proof. If the Simplex Method terminates in Step 2, then $\bar{c}_j \geq 0$ for all j. Hence, by Theorem 1, the current bfs is optimal. If the Simplex Method terminates in Step 4, then the vector d (which we recall has $d_B = -u$, $d_j = 1$ and $d_{j'} = 0$ for all $j' \neq j, B(1), \ldots, B(m)$) satisfies that $x + \theta d$ is feasible for all $\theta \geq 0$ and $c^T d = \bar{c}_j < 0$. Hence the optimum is $-\infty$.

Finally, by Lemma 2, in each iteration, our current solution x is a basic feasible solution. Since there is no degeneracy, the objective value $c^T x$ strictly decreases in each iteration. So we consider each basic feasible solution at most once, and since there is a finite number of bfs, the algorithm must have a finite number of iterations. $\square$

We are left with a couple of questions.

- How do we find an initial bfs? [Initialization]

- What do we do if our current bfs is degenerate? [Anticycling Rules]

  Note that we can still execute our iteration in Section 1.4; the only problem is that we may have $\theta^* = 0$. This means that we do not actually move to a different bfs, just to a different basis. We'll call this a *degenerate pivot*. Sometimes you have to do such a sequence of basis changes before you can move to a different solution. The danger, however, is cycling: the algorithm may keep changing bases but never leave the current solution.

  In practice, this fortunately does not happen. We'll also see a way to deal with this problem in theory.

  An example of cycling is given here: http://glossary.computing.society.informs.org/notes/cycling.pdf (We'll see how to interpret these tables next week).

- If we have multiple non-basic variables with $\bar{c}_j < 0$, which one should we choose? [Pivot Selection]

- How fast is this algorithm? [Implementation Issues, Number of Iterations]