1 Projections

We would like to take \( d = -c^k \), in order to decrease the objective function value as quickly as possible. Unfortunately, this may lead to a point which violates the constraint \( A^k z = b \). For example, in the problem \((LP1^k)\), taking \( \beta = 0.5 \) gives the point

\[
\begin{align*}
z &= e - 0.5c^k \\
&= (1, 1, 1, 1) - 0.5(-0.8, -0.1, 0.2, 1.9) \\
&= (1.4, 1.05, 0.9, 0.05)
\end{align*}
\]

which does not satisfy the constraints. We need to move in a direction \( d \) which satisfies \( A^k d = 0 \). (A direction \( d \) which satisfies \( A^k d = 0 \) is said to be in the nullspace of the matrix \( A^k \).) We then get

\[
A^k (e + \beta d) = A^k e + \beta A^k d = A^k e = b,
\]

because \( e \) is feasible in \((P^k)\). Therefore, we project the direction \( c^k \) onto the nullspace of \( A^k \). Algebraically, this means that we take

\[
d = -P_{A^k} c^k \tag{1}
\]

where

\[
P_{A^k} = (I - A^k A^k T (A^k A^k T)^{-1} A^k), \tag{2}
\]

and \( I \) denotes the identity matrix. (Aside: We assume that the rows of \( A \) are linearly independent. Under this assumption, the projection matrix is well-defined. Note that we need this assumption to hold in order to be able to obtain a basic feasible solution.) For the problem \((LP1)\), we have

\[
A^k A^k T = \begin{bmatrix} 0.8 & 0 & 0.2 & 0 \\ 0 & 0.1 & 0 & 1.9 \end{bmatrix} \begin{bmatrix} 0.8 & 0 \\ 0 & 0.1 \\ 0.2 & 0 \\ 0 & 1.9 \end{bmatrix} = \begin{bmatrix} 0.68 & 0 \\ 0 & 3.62 \end{bmatrix}
\]

so

\[
(A^k A^k T)^{-1} = \begin{bmatrix} 1.4706 & 0 \\ 0 & 0.2762 \end{bmatrix}
\]
and

\[ P_A^k = I - \begin{bmatrix} 0.8 & 0 \\ 0 & 0.1 \\ 0.2 & 0 \\ 0 & 1.9 \end{bmatrix} \begin{bmatrix} 1.4706 & 0 \\ 0 & 0.2762 \\ 0.8 & 0.2 \\ 0 & 1.9 \end{bmatrix} \begin{bmatrix} 0.8 & 0 \end{bmatrix} \]

\[ = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.941 & 0 & 0.235 & 0 \\ 0 & 0.003 & 0 & 0.052 \\ 0.235 & 0 & 0.059 & 0 \\ 0 & 0.052 & 0 & 0.997 \end{bmatrix} \]

\[ = \begin{bmatrix} 0.059 & 0 & -0.235 & 0 \\ 0 & 0.997 & 0 & -0.052 \\ -0.235 & 0 & 0.941 & 0 \\ 0 & -0.052 & 0 & 0.003 \end{bmatrix} \]

So, in the problem \((LP1^k)\), we use the direction

\[ d = -P_A^k c^k \]

\[ = - \begin{bmatrix} 0.059 & 0 & -0.235 & 0 \\ 0 & 0.997 & 0 & -0.052 \\ -0.235 & 0 & 0.941 & 0 \\ 0 & -0.052 & 0 & 0.003 \end{bmatrix} \begin{bmatrix} -0.8 \\ -0.1 \\ 0.2 \\ 1.9 \end{bmatrix} \]

\[ = \begin{bmatrix} 0.0942 \\ 0.1985 \\ -0.3762 \\ -0.0109 \end{bmatrix} \]

This gives a new point of the form

\[ z^{new} = e + \beta d \]

\[ = (1, 1, 1, 1) + \beta(.0941, .1994, -.3765, -.0160) \]

Taking \(\beta = 2\) gives

\[ z^{new} = (1.1882, 1.3988, 0.237, 0.968) \]

and then we get a new iterate for \((P)\):

\[ x^{k+1} = D^k z^{new} \]

\[ = (0.95, 0.14, 0.05, 1.86) \]

with value -0.82.
2 Choosing the steplength

We have $z^{new} = e + \beta d$, and we need to have $z^{new} > 0$. Thus, we need to select $\beta$ so that $1 + \beta d_i > 0$ for each component $i$. So, we need to pick

$$\beta \leq 1/(-d_i) \text{ if } d_i < 0,$$

so we take

$$\beta = 0.9/\max\{-d_i : d_i < 0\}. \quad (3)$$

This will result in moving 0.9 of the way to the boundary of the feasible region of the rescaled problem $(P^k)$. Notice that one component of $z^{new}$ will be equal to 0.1.

(This discussion on choosing the steplength should remind you of the minimum ratio rule in the simplex algorithm.)

3 One iteration of the algorithm

Recall that when we looked at the revised simplex method, we found out that it was not necessary to calculate the whole of the simplex tableau. In the same spirit of efficiency, it is not necessary to calculate the complete projection matrix $P_A$. In practice, we would perform one iteration as follows:

1. Given current iterate $x^k > 0$ feasible in $(P)$.
2. Calculate $D^k, A^k = AD^k, c^k = D^k c$.
3. Calculate $w^k = A^k c^k$.
4. Calculate $v^k = (A^k A^{kT})^{-1} w^k$. (Aside: In practice, we would not calculate the inverse explicitly, but we would solve the system of equations $(A^k A^{kT}) v^k = w^k$ in order to find $v^k$.)
5. Calculate $g^k = A^{kT} v^k$.
6. Calculate the direction $d^k$ in the problem $(P^k)$: $d^k = -c^k + g^k (= -P_A c^k)$.
7. Calculate step length $\beta$ using equation (3).
8. Update $z^{new} = e + \beta d^k$.
9. Update $x^{k+1} = D^k z^{new}$. 

4 Monotonicity of the algorithm

The algorithm does decrease the objective function at each iteration. This is because a projection matrix $M$ is idempotent, that is, it is symmetric and $MM = M$. We then get that the objective function value of $z_{new}$ is

$$c^T z_{new} = c^T (e - \beta P_A c^k) = c^T e - \beta c^T P_A c^k$$

$$= c^T e - \beta c^T P_A P_A c^k = c^T e - \beta c^T P^T A P_A c^k$$

and this must be smaller than $c^T e$ since it subtracts off the square of the length of the vector $P_A c^k$.

5 Getting a dual solution

It can be shown that

- $d^k = 0$ if and only if $x^k$ is a vertex.
- $c - A^T v^k \geq 0$ at the optimal vertex. The optimal vertex is the only one where we have $c - A^T v^k \geq 0$.

A possible dual iterate is

$$v^k = (A^k A^{kT})^{-1} A^k c^k.$$ (4)

Notice that the dual slacks are then $c - A^T v^k$, and that $d^k = -D^k (c - A^T v^k)$. Thus, the elements of $d^k$ give the product between the primal variable $d^k_i$ and the corresponding dual slack, so they give a measure of the complementary slackness. At optimality, we will have dual feasibility with this choice. It can be shown that this choice is actually dual optimal if $x^k$ is optimal for the primal problem.

6 Stopping the algorithm

We stop the simplex algorithm when all the costs in the tableau are nonnegative. Unfortunately, we don’t have such a simple stopping rule for the primal affine scaling algorithm. However, there are some possibilities:

- Is $c - A^T v^k \geq -\epsilon e$ for some $\epsilon$ such as $10^{-6}$. (So almost dual feasible.)
- Is the change in the objective function from one iteration to the next smaller than some tolerance? That is, do we have $c^T x^k - c^T x^{k+1} \leq \epsilon$, or alternatively, $(c^T x^k - c^T x^{k+1})/c^T x^k \leq \epsilon$. This criterion indicates that we are not making much progress, so we must be very close to the optimal.
- Is $d^k$ a small vector? This indicates that we are close to getting complementary slackness. (See the last section.)