MA3231 Linear Programming
W. J. Martin
October 9, 2018

The Affine Scaling Method

Overview

Given a linear programming problem in equality form with full rank constraint matrix and a strictly positive feasible solution x^0, we transform the problem to a new one with feasible solution $\mathbf{1} = [1,\ 1,\ \cdots, 1]^\top$. The original problem and scaled problem

Original:		Scaled:	
max	$c^\top x$	**max**	$\hat{c}^\top x$
s. t.	$Ax = b$	**s. t.**	$\hat{A}x = b$
	$x \geq 0$		$x \geq 0$

are related as follows: X^0 is the diagonal matrix with the entries of x^0 down the diagonal; $\hat{A} = AX^0$; and $\hat{c} = X^0 c$.

In the scaled problem, we

- project the gradient $\hat{c}$ onto the subspace $\hat{A}x = 0$ to obtain a search direction d;
- find the largest scalar β such that $\mathbf{1} + \beta d$ is still non-negative (the other feasibility conditions are already guaranteed by choice of d);
- obtain new (scaled) solution $\hat{x} = \mathbf{1} + r\beta d$ where $0 < r < 1$ is a global constant;
- update our solution x^0 to the next solution $x^1 = X^0 \hat{x}$.

This process is repeated, as needed, until we are sufficiently close to an optimal solution (e.g., as one may check by solving for a dual vector using complementary slackness conditions).

One iteration of the affine scaling method

We assume we have an $m \times n$ matrix A with full row rank, an m-vector b and an n-vector c. (If A does not have full row rank, first row reduce, and if $Ax = b$ is feasible, then replace A by a smaller matrix.) We also have a fixed step size r, $0 < r < 1$.

Given a strictly positive feasible solution x^0 to the problem (so $Ax^0 = b$), we obtain a better strictly positive feasible solution x^1 as follows:

Step 1: (**Scale**) Let $\hat{A} = AX^0$ and $\hat{c} = X^0 c$ where X^0 is the diagonal matrix with entries from x^0 down the diagonal. (Now we have a scaled problem with $\mathbf{1}$ as a feasible solution.)

Step 2: (**Find projection**) Compute the projection matrix onto the nullspace of $\hat{A}$:

$$P = I - \hat{A}^\top (\hat{A}\hat{A}^\top)^{-1} \hat{A}$$

Step 3: (**Find search direction**) The search direction (in the scaled problem) is then

$$d = P\hat{c}.$$

Step 4: (**Compute ratios**) Let $\beta = \min \left\{ -\frac{1}{d_j} \ : \ d_j < 0 \right\}$. (If this set is empty, the problem is unbounded.)

Step 5: (**Scaled update**) The improved solution in the scaled problem is

$$\hat{x} = \mathbf{1} + r\beta d$$

where r is the prescribed step size $(0 < r < 1)$.

Step 6: (**Next iterate**) Now scale back to the original problem and take

$$x^1 = X^0 \hat{x}.$$

Note that the above algorithm brings us from x^0 to x^1. To obtain further improved solutions, we iterate this with x^0 replaced by the current solution. The general step is described with only a slight change of notation and is given at the end of these notes, just before the exercises.

Some explanations

In Step 1, we simply write down the data $(\hat{A}, \hat{c})$ for the scaled problem. Note that the right-hand side vector does not change. But that's okay because

$$\hat{A}\mathbf{1} = (AX^0)\mathbf{1} = A(X^0\mathbf{1}) = Ax^0 = b$$

so $\mathbf{1}$ is indeed a feasible solution to the scaled problem and it is "far away" from the boundaries $x_j = 0$. Moreover, X^0 is a diagonal matrix, so it is symmetric: $(X^0)^\top = X^0$. Thus

$$\hat{c}^\top\mathbf{1} = (X^0 c)^\top\mathbf{1} = (c^\top(X^0)^\top)\mathbf{1} = c^\top(X^0\mathbf{1}) = c^\top x^0.$$

More generally, the objective value of any solution $\hat{x}$ to the scaled problem is equal to the objective value of the corresponding solution $x = X^0\hat{x}$ in the original problem.

In Step 2, we do the most work. We need to project the gradient $\hat{c}$ onto the subspace $\{x : \hat{A}x = 0\}$ in order to have a feasible search direction. As we worked out in class, the projection matrix P does just the trick.

In Step 3, we take $d = P\hat{c}$ so that $\hat{A}d = 0$ and, for any scalar β, the vector $\hat{x} = \mathbf{1} + \beta d$ satisfies

$$\hat{A}\hat{x} = \hat{A}(\mathbf{1} + \beta d) = \hat{A}\mathbf{1} + \beta\hat{A}d = \hat{A}\mathbf{1} = b$$

as established above. So, no matter how far we travel in this direction, the equality constraints will still be satisfied.

From this, we know in Step 4 that any vector of the form $\hat{x} = \mathbf{1} + \beta d$ satisfies the equality constraints. So in order to stay feasible, we simply need $\mathbf{1} + \beta d \geq 0$. This requires $1 + \beta d_j \geq 0$ for all j. Since we are going to take β positive (to improve the objective value), the only coordinates j that concern us are those with $d_j < 0$, and each of these gives us an upper limit $\beta \leq -1/d_j$.

In Step 5, we finally use our global step size r. We know that $\mathbf{1} + \beta d \geq 0$ and is otherwise feasible. But we need a *strictly positive* feasible solution. So we only move a fraction of the maximum possible distance. For example, with $r = 0.9$, we only go "90% of the way to the wall". Thus our update is $\mathbf{1} + r(\beta d)$.

The transformed problem has now served its purpose of pushing the search direction away from the boundaries. We discard it and transform the solution $\hat{x}$ back to the original LP to get the next iterate x^1. As explained in Step 1, this is simply achieved through scaling by X^0.

Example 1: Consider the problem

$$\begin{array}{llllllllll}
\textbf{max} & 3 & x_1 & + & x_2 & & & & & \\
\textbf{s.t.} & 5 & x_1 & +2 & x_2 & + & x_3 & & = & 2 \\
 & & x_1 & +2 & x_2 & & & + \; x_4 & = & 2 \\
 & & x_1, & & x_2, & & x_3, & x_4 & \geq & 0
\end{array}$$

with initial feasible solution $x^0 = \left(\frac{1}{5}, \frac{2}{5}, \frac{1}{5},\ 1\right)$. So our current objective value is $\zeta = 1$.

We have

$$A = \begin{bmatrix} 5 & 2 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \quad c = \begin{bmatrix} 3 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad x^0 = \begin{bmatrix} 1/5 \\ 2/5 \\ 1/5 \\ 1 \end{bmatrix}.$$

Step 1: Scale the problem:

$$X^0 = \begin{bmatrix} 1/5 & 0 & 0 & 0 \\ 0 & 2/5 & 0 & 0 \\ 0 & 0 & 1/5 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad \hat{A} = AX^0 = \begin{bmatrix} 1 & 4/5 & 1/5 & 0 \\ 1/5 & 4/5 & 0 & 1 \end{bmatrix}, \quad \hat{c} = X^0 c = \begin{bmatrix} 3/5 \\ 2/5 \\ 0 \\ 0 \end{bmatrix}.$$

Step 2: Compute P:

$$P = I - \hat{A}^\top \left(\hat{A}\hat{A}^\top\right)^{-1} \hat{A}.$$

$$\hat{A}\hat{A}^\top = \frac{21}{25} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad \left(\hat{A}\hat{A}^\top\right)^{-1} = \frac{25}{63} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

We eliminate fractions where possible to write

$$\hat{A}^\top \left(\hat{A}\hat{A}^\top\right)^{-1} = \frac{5}{63} \begin{bmatrix} 5 & 1 \\ 4 & 4 \\ 1 & 0 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = \frac{5}{63} \begin{bmatrix} 9 & -3 \\ 4 & 4 \\ 2 & -1 \\ -5 & 10 \end{bmatrix}$$

so that

$$\hat{A}^\top \left(\hat{A}\hat{A}^\top\right)^{-1} \hat{A} = \frac{1}{63} \begin{bmatrix} 9 & -3 \\ 4 & 4 \\ 2 & -1 \\ -5 & 10 \end{bmatrix} \begin{bmatrix} 5 & 4 & 1 & 0 \\ 1 & 4 & 0 & 5 \end{bmatrix} = \frac{1}{63} \begin{bmatrix} 42 & 24 & 9 & -15 \\ 24 & 32 & 4 & 20 \\ 9 & 4 & 2 & -5 \\ -15 & 20 & -5 & 50 \end{bmatrix}.$$

Let's call this last matrix Q. Then we have

$$P - I - Q = \frac{1}{63} \begin{bmatrix} 21 & -24 & -9 & 15 \\ -24 & 31 & -4 & -20 \\ -9 & -4 & 61 & 5 \\ 15 & -20 & 5 & 13 \end{bmatrix}.$$

Step 3: Find search direction d:

$$d = P\hat{c} = \frac{1}{63}\begin{bmatrix} 21 & -24 & -9 & 15 \\ -24 & 31 & -4 & -20 \\ -9 & -4 & 61 & 5 \\ 15 & -20 & 5 & 13 \end{bmatrix}\begin{bmatrix} 3/5 \\ 2/5 \\ 0 \\ 0 \end{bmatrix} = \frac{1}{315}\begin{bmatrix} 15 \\ -10 \\ -35 \\ 5 \end{bmatrix}$$

which simplifies to $d = \frac{1}{63}(\ 3,\ -2,\ -7,\ 1)$.

Step 4: Ratios:

$$\beta = \min\left\{\frac{-1}{d_j}\ :\ d_j < 0\right\} = \min\left\{\frac{63}{7}, \frac{63}{2}\right\} = 9.$$

So $\beta = 9$.

Step 5: Scaled update (using step size $r = 1/2$):

$$\hat{x} = \mathbf{1} + r\beta d = \frac{1}{14}\begin{bmatrix} 14 \\ 14 \\ 14 \\ 14 \end{bmatrix} + \frac{1}{14}\begin{bmatrix} 3 \\ -2 \\ -7 \\ 1 \end{bmatrix} = \frac{1}{14}\begin{bmatrix} 17 \\ 12 \\ 7 \\ 15 \end{bmatrix}$$

using $r\beta \cdot \frac{1}{63} = \cdot\frac{1}{2} \cdot 9 \cdot \frac{1}{63} = \frac{1}{14}$.

Step 6: Next iterate:
Now we go back to the original problem using X^0:

$$x^1 = X^0\hat{x} = \begin{bmatrix} 1/5 & 0 & 0 & 0 \\ 0 & 2/5 & 0 & 0 \\ 0 & 0 & 1/5 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 17/14 \\ 12/14 \\ 7/14 \\ 15/14 \end{bmatrix} = \frac{1}{70}\begin{bmatrix} 17 \\ 24 \\ 7 \\ 75 \end{bmatrix}$$

with objective value $\zeta = 15/14$. As you can see, the numbers will get worse after this, so we use a computer if we want to go further.

Example 2: Consider the problem

$$\begin{array}{lrcl} \textbf{max} & x_1 - x_2 & & \\ & 2x_1 + x_2 & = & 4 \\ & x_1, \quad x_2 & \geq & 0 \end{array}$$

with initial feasible solution $x^0 = (1, 2)$ and step size $r = 1/2$.

We have

$$A = \begin{bmatrix} 2 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 4 \end{bmatrix}, \quad c = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad x^0 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Step 1: Scale the problem:

$$X^0 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \qquad \hat{A} = AX^0 = \begin{bmatrix} 2 & 2 \end{bmatrix}, \quad \hat{c} = X^0 c = \begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

Step 2: Compute $P = I - \hat{A}^\top \left(\hat{A}\hat{A}^\top\right)^{-1} \hat{A}$:

$$\hat{A}\hat{A}^\top = \begin{bmatrix} 2 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \end{bmatrix}, \qquad P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \end{bmatrix} \begin{bmatrix} 1/8 \end{bmatrix} \begin{bmatrix} 2 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{1}{8} \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix}$$

$$P = \begin{bmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{bmatrix}$$

Step 3: Find search direction d:

$$d = P\hat{c} = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 3 \\ -3 \end{bmatrix} = \begin{bmatrix} 3/2 \\ -3/2 \end{bmatrix}$$

Step 4: Ratios:

$$\beta = \min \left\{ \frac{-1}{d_j} \ : \ d_j < 0 \right\} = \frac{2}{3}.$$

Step 5: Scaled update:

$$\hat{x} = \mathbf{1} + r\beta d = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot \begin{bmatrix} 3 \\ -3 \end{bmatrix} = \begin{bmatrix} 3/2 \\ 1/2 \end{bmatrix}$$

Step 6: Next iterate:

$$x^1 = X^0 \hat{x} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 3/2 \\ 1/2 \end{bmatrix} = \begin{bmatrix} 3/2 \\ 1 \end{bmatrix}.$$

In this iteration, our objective value has improved from -1 to $+1/2$.

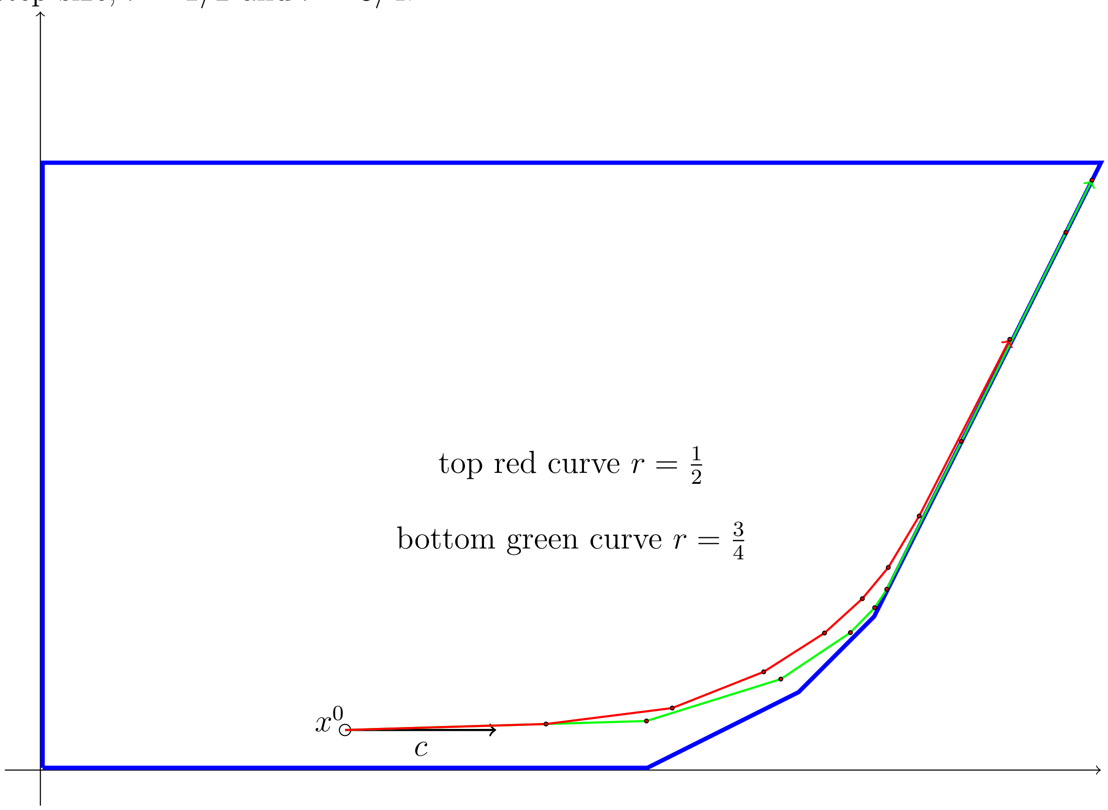
Pictorial Example

The advantage of interior point methods is that they tend to avoid the combinatorial complexity of basic feasible solutions. One can imagine a high-dimensional polyhedron having large clusters of nearby vertices all of which are suboptimal. The simplex method can only move from one vertex to a neighbor, possibly contributing to inefficiency.

To illustrate how an interior point approach can "smooth out" all these corners, let's look at a simple two-dimensional example with three nearby basic feasible solutions far away from optimality:

$$\begin{array}{rl} \textbf{maximize} & 2x_1 \\ \textbf{subject to} & 2x_1 - \ \ x_2 \ \le \ 20 \\ & x_1 - \ \ x_2 \ \le 9 \\ & x_1 - 2x_2 \ \le 8 \\ & x_2 \ \le 8 \\ & x_1, \quad x_2 \ \ge 0 \end{array}$$

We will skip the conversion to equality form. After introducing slack variables $x_3, \ldots, x_6$, the initial strictly feasible solution $(4, \frac{1}{2})$ gives us $x^0 = (4, \frac{1}{2}, \frac{15}{2}, \frac{10}{2}, \frac{11}{2}, \frac{25}{2})$. The objective direction c is horizontal, pointing directly to the right. But the optimal solution is at $(14, 8)$. Here we give the first eight iterates $x^1, \ldots, x^8$ of the affine scaling method with two choices of step size, $r = 1/2$ and $r = 3/4$.



Initially, we showed how to move from initial solution x^0 to a better solution x^1. A generic iteration looks like this:

Step 1: (**Scale**) Let $\hat{A} = AX^k$ and $\hat{c} = X^k c$ where X^k is the diagonal matrix with entries from the most recent solution x^k down the diagonal. (Now we have a scaled problem with $\mathbf{1}$ as a feasible solution.)

Step 2: (**Find projection**) Compute the projection matrix onto the nullspace of $\hat{A}$:

$$P = I - \hat{A}^\top (\hat{A}\hat{A}^\top)^{-1} \hat{A}$$

Step 3: (**Find search direction**) The search direction (in the scaled problem) is then

$$d = P\hat{c}.$$

Step 4: (**Compute ratios**) Let $\beta = \min\left\{ -\frac{1}{d_j} \ : \ d_j < 0 \right\}$. (If this set is empty, the problem is unbounded.)

Step 5: (**Scaled update**) The improved solution in the scaled problem is

$$\hat{x} = \mathbf{1} + r\beta d$$

where r is the prescribed step size ($0 < r < 1$).

Step 6: (**Next iterate**) Now scale back to the original problem to obtain the next iterate

$$x^{k+1} = X^k \hat{x}.$$

Exercises

1.) Consider the linear programming problem

$$\textbf{max} \ -x_1 - x_2 \qquad \textbf{subject to} \quad x_1 + 2x_2 \leq 2, \quad x_1, x_2 \geq 0.$$

(i) Starting with $x^0 = (1/2,\ 1/4)$, apply two iterations of the affine scaling method using $r = 1/2$. (First, convert to equality form.) For each iteration, give

- the constraint matrix $\hat{A}$ and objective vector $\hat{c}$ for the scaled problem;
- the projection matrix P and search direction d for this iteration;
- the ratio computation from Step 4;

- the next iterate, both in scaled form $\hat{x}$ and as a solution x^{k+1} to the original problem above. (So you will be finding x^1 and x^2.)

(ii) On a sheet of graph paper, make a careful (and large!) drawing of the feasible region in $\mathbb{R}^2$. For x^0 and each of the next two iterates, x^1, and x^2, plot both the gradient of the objective function (namely $c^\top = [-1 \ -1]$) at that point as well as the scaled step direction $x^{k+1} - x^k$.

2.) Repeat the steps of Exercise 1 (again with $r = 1/2$) for the linear programming problem

$$\textbf{max } x_1 + 2x_2 \qquad \textbf{subject to} \quad 0 \le x_1 \le 1, \quad 0 \le x_2 \le 1.$$

starting with $x^0 = (3/4,\ 1/4)$ (a column vector). (You may wish to use MAPLE to check your computations and to follow the trajectory further.)

3.) Perform two iterations of the affine scaling method for the problem

$$\begin{array}{rl} \textbf{maximize} & x_1 + 2x_2 + 2x_3 \\ \textbf{subject to} & x_1 - x_2 \qquad\qquad = 0 \\ & x_1 + x_2 + \sqrt{2}x_3 = 4 \\ & x_1, \quad x_2, \quad x_3 \ \ge 0 \end{array}$$

with step size $r = 1/2$ and initial solution $x^0 = (1,\ 1,\ \sqrt{2})$.

4.) Perform two iterations of the affine scaling method for the problem **max** x_1 subject to $x_1 + x_2 = 7$, $x_1, x_2 \ge 0$ with step size $r = \frac{3}{4}$ and initial solution $x^0 = (3, 4)$.

5.) Perform two iterations of the affine scaling method for the problem **max** $x_1 - x_2 + 3x_3$ subject to $x_1 + 2x_2 + x_3 = 5$ (all variables nonnegative) with step size $r = \frac{1}{2}$ and initial solution $x^0 = (1, 1, 2)$.

6.) Perform two iterations of the affine scaling method for the problem max $c^\top x$ **s.t.** $Ax = b$, $x \ge 0$ with step size $r = 1/2$ and initial solution $x^0 = (5, 3, 2)$ where

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \end{bmatrix}, \qquad b = \begin{bmatrix} 5 \\ 7 \end{bmatrix}, \qquad c = (2, 2, -1).$$

7.) Perform two iterations of the affine scaling method for the problem

$$\begin{array}{rl} \textbf{maximize} & x_2 - x_3 \\ \textbf{subject to} & x_1 + 4x_2 - 4x_3 = -2 \\ & 2x_1 + 2x_2 + x_3 = 8 \\ & x_1, \quad x_2, \quad x_3 \ \ge 0 \end{array}$$

with step size $r = 1/2$ and initial solution $x^0 = (2, 1, 2)$.

8.) Perform two iterations of the affine scaling method for the problem

$$\textbf{max} - x_1 + 2x_2 \qquad \textbf{s.t.} \quad 2x_1 + 5x_2 = 3 \qquad x_1, x_2 \geq 0$$

with step size $r = 0.75$ and initial solution $x^0 = (1, 0.2)$.

9.) Perform two iterations of the affine scaling method for the problem

$$\begin{array}{rrrrl} \textbf{maximize} & 2x_1 & & - \quad x_3 & \\ \textbf{subject to} & 2x_1 - & x_2 & & = 0 \\ & & x_2 + & \frac{1}{2}x_3 & = 2 \\ & x_1, & x_2, & x_3 & \geq 0 \end{array}$$

with step size $r = 1/2$ and initial solution $x^0 = (\frac{1}{2}, 1, 2)$.

10.) Perform two iterations of the affine scaling method for the problem **min** $-2x_1 - x_2$ **s.t.** $3x_1 + 5x_2 = 13$, $x_1, x_2 \geq 0$ with step size $r = 1/2$ and initial solution $x^0 = (1, 2)$.

11.) Perform two iterations of the affine scaling method for the problem

$$\begin{array}{rrrrl} \textbf{maximize} & x_1 + & 2x_2 + & 2x_3 & \\ \textbf{subject to} & x_1 - & x_2 & & = 0 \\ & x_1 + & x_2 + & \sqrt{2}x_3 & = 4 \\ & x_1, & x_2, & x_3 & \geq 0 \end{array}$$

with step size $r = 1/2$ and initial solution $x^0 = (1, 1, \sqrt{2})$.

12.) More generically, consider the n-dimensional optimization problem with one equality constraint $\sum_{j=1}^{n} x_j = 1$, objective function $\max c^\top x$, step size $r = 1/2$, and initial solution x^0 with positive entries summing to one. For convenience, also assume that $\sum_{\ell=1}^{n} c_\ell (x_\ell^0)^2 = \|x^0\|^2$ and that the variables are ordered in such a way that $c_1 \geq c_2 \geq \cdots \geq c_n$. Compare the next iterate x^1 produced by the affine scaling method against the corresponding unscaled update $x^0 + \frac{1}{2}\beta' c'$ where c' is the projection of the objective gradient c onto the affine subspace containing the feasible region and β' is computed so as to prevent any entry of $x^0 + \beta' c'$ from going negative.