A Still Simpler Way of Introducing the Interior-Point Method for Linear Programming

Kurt Mehlhorn*
Max-Planck-Institut für Informatik,
Saarbrücken,
Germany

Sanjeev Saxena†
Computer Science and Engineering,
Indian Institute of Technology,
Kanpur, INDIA-208 016

December 8, 2021

Abstract

Linear programming is now included in algorithm undergraduate and postgraduate courses for computer science majors. We show that it is possible to teach interior-point methods directly to students with just minimal knowledge of linear algebra.

1 Introduction

Terlaky [7] and Lesaja [4] have suggested simple ways to teach interior-point methods. In this paper, we suggest a still simpler way. Most material required to teach interior-point methods is available in popular text books [5, 8]. However, these books assume knowledge of calculus, which is not really required. If appropriate material is selected from these books, then it becomes feasible to teach interior-point methods as the first or only method for linear programming.

The canonical *linear programming problem* is to

$$\text{minimize } c^T x \text{ subject to } Ax = b \text{ and } x \geq 0. \tag{1}$$

Here, A is an $n \times m$ matrix, b and c are n-dimensional, and x is an m-dimensional vector. A *feasible solution* is any vector x with $Ax = b$ and $x \geq 0$. The problem is *feasible* if there is a feasible solution, and *infeasible* otherwise. The problem is *unbounded* if for every real z, there is a feasible x with $c^T x \leq z$, and *bounded* otherwise. Infeasible problems are bounded.

Remark 1. Maximize $c^T x$ is equivalent to minimize $-c^T x$.

Remark 2. Constraints of type $\alpha_1 x_1 + \ldots + \alpha_n x_n \leq \beta$ can be replaced by $\alpha_1 x_1 + \ldots + \alpha_n x_n + \gamma = \beta$ with a new (slack) variable $\gamma \geq 0$. Similarly, constraints of type $\alpha_1 x_1 + \ldots + \alpha_n x_n \geq \beta$ can be replaced by $\alpha_1 x_1 + \ldots + \alpha_n x_n - \gamma = \beta$ with a (surplus) variable $\gamma \geq 0$.

We first prepare the problem by deleting superfluous equations and making the rows of A linearly independent. Assume first that A contains a row i in which all entries are equal to zero. If b_i is also

*E-mail: mehlhorn@mpi-inf.mpg.de
†E-mail: ssax@iitk.ac.in

zero, we simply delete the row. If b_i is nonzero, the system of equations has no solution, and we declare the problem infeasible and stop. Now, every row of A contains a nonzero entry, in particular, the first row. We may assume that a_{11} is nonzero. Otherwise, we interchange two columns. We multiply the ith equation by $-\frac{a_{11}}{a_{i1}}$ and subtract the first equation. In this way, the first entry of all equations but the first becomes zero. If any row of A becomes equal to the all zero vector, we either delete the equation or declare the problem infeasible. We now proceed in the same way with the second equation. We first make sure that a_{22} is nonzero by interchanging columns if necessary. Then we multiply the ith equation (for $i > 2$) by $-\frac{a_{22}}{a_{21}}$ and subtract the second equation. And so on. In the end, all remaining equations will be linearly independent. Equivalently, the resulting matrix will have full row-rank.

We now have n constraints in m variables with $m \geq n$. If $m = n$, the system $Ax = b$ has a unique solution (recalling that A has full row-rank and is hence invertible). We check whether this solution is nonnegative. If so, we have solved the problem. Otherwise, we declare the problem infeasible. So, we may from now on assume $m > n$ (more variables than constraints).

We consider another problem, the *dual problem*, which is

$$\text{maximize } b^T y, \text{ subject to } A^T y + s = c, \text{ with slack variables } s \geq 0 \text{ and unconstrained variables } y. \tag{2}$$

Remark 3. The vector y has m components and the vector s has n components. We will call the original problem the *primal problem*.

Claim 1 (Weak Duality). *If x is any solution of $Ax = b$ with $x \geq 0$ and (y, s) is a solution of $A^T y + s = c$ with $s \geq 0$, then*
1. $x^T s = c^T x - b^T y$, and
2. $b^T y \leq c^T x$, with equality if and only if $s_i x_i = 0$ for all is.

Proof. We multiply $s = c - A^T y$ with x^T from the left and obtain

$$x^T s = x^T c - x^T (A^T y) = c^T x - (x^T A^T) y = c^T x - (Ax)^T y = c^T x - b^T y.$$

As $x, s \geq 0$, we have $x^T s \geq 0$, and hence, $c^T x \geq b^T y$.

Equality will hold if $x^T s = 0$, or equivalently, $\sum_i s_i x_i = 0$. Since $s_i, x_i \geq 0$, $\sum_i s_i x_i = 0$ if and only if $s_i x_i = 0$ for all i. ■

Remark 4. If x is a feasible solution of the primal and (y, s) is a feasible solution of the dual, the difference $c^T x - b^T y$ is called the *objective value gap* of the solution pair.

Remark 5. Thus, if the value of the primal and the dual problem are the same, then both are optimal. Actually, from the Strong Duality Theorem, if both primal and dual solutions are optimal, then the equality will hold. We will prove the Strong Duality Theorem in Section 5 (Corollary 2).

Remark 6. *We will proceed under the assumption that the primal as well as the dual problem are both bounded and feasible.* We come back to this point in Section 4. If the primal is unbounded, the dual is infeasible. If the dual is unbounded, the primal is infeasible. If the primal is feasible and bounded, the dual is feasible and bounded. The primal is unbounded if it is feasible and the homogeneous problem "minimize $c^T x$ subject to $Ax = 0$ and $x \geq 0$" has a negative objective value. Equivalently, if the problem "minimize 0 subject to $c^T x = -1$, $Ax = 0$, and $x \geq 0$" is feasible".

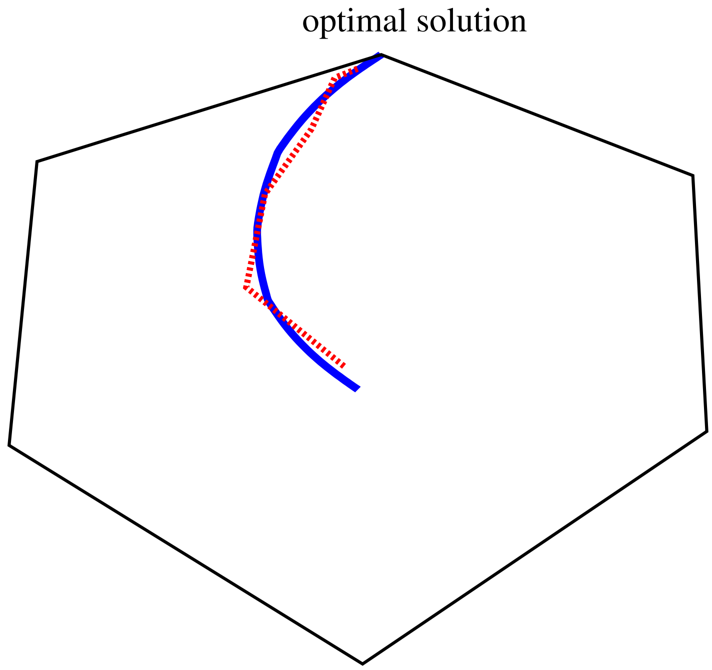

Figure 1: The interior of the polygon comprises all points (x, y, s) satisfying $Ax = b$ and $A^T y + s = c$, $x > 0$, and $s > 0$. The blue (bold) line consists of all points in this polygon with $x_i s_i = \mu$ for all i and some $\mu > 0$. The optimal solution is a vertex of the polygon corresponding to $x_i s_i = 0$ for all i. The red (dashed) line illustrates the steps of the algorithm. It follows the blue (bold) line in discrete steps.

Claim 1 implies, that if we are able to find a solution to the following system of equations and inequalities

$$Ax = b,\ A^T y + s = c,\ x_i s_i = 0 \text{ for all } i,\ x \geq 0,\ s \geq 0,$$

we will get optimal solutions of both the original and the dual problem. Notice that the constraints $x_i s_i = 0$ are nonlinear and hence it is not clear whether we have made a step towards the solution of our problem. The idea is now to relax the conditions $x_i s_i = 0$ to the conditions $x_i s_i \approx \mu$ (with the exact form of this equation derived in the next section), where $\mu \geq 0$ is a parameter. We obtain

$$(P_\mu) \quad Ax = b,\ A^T y + s = c,\ x_i s_i \approx \mu \text{ for all } i,\ x > 0,\ s > 0.$$

We will show:

1. (initial solution) For a suitable, μ, it is easy to find a solution to the problem P_μ. This will be the subject of Section 4.
2. (iterative improvement) Given a solution (x, y, μ) to P_μ, one can find a solution (x', y', s') to $P_{\mu'}$, where μ' is substantially smaller than μ. This will be the subject of Section 2. Applying this step repeatedly, we can make μ arbitrarily small.
3. (final rounding) Given a solution (x, y, μ) to P_μ for sufficiently small μ, one can extract the exact solutions for the primal and the dual problem. This will be the subject of Section 5.

Remark 7. For the iterative improvement, it is important that $x > 0$ and $s > 0$. For this reason, we replace the constraints $x \geq 0$ and $s \geq 0$ by $x > 0$ and $s > 0$ when defining problem P_μ (see Figure 1).

Remark 8. Note that $x_i s_i \approx \mu$ for all i implies $b^T y - c^T x \approx m\mu$ by Claim 1. Thus, repeated application of iterative improvement will make the gap between the primal and dual objective values arbitrarily small.

2 Iterative Improvement: Use of the Newton-Raphson Method

This section and the next follow Roos et al [5] (see also Vishnoi [9]).

Let us assume that we have a solution (x, y, s) to

$$Ax = b \text{ and } A^T y + s = c \text{ and } x > 0 \text{ and } s > 0.$$

We will use the Newton-Raphson Method [5] to get a "better" solution. Let us choose the next values as $x' = x + h$, $y' = y + k$, and $s' = s + f$. You should think of the steps h, k, and f as small values. Then we want, ignoring the positivity constraints for x' and s' for the moment:

1. $Ax' = A(x + h) = b$, or equivalently, $Ax + Ah = b$. Since $Ax = b$, this is tantamount to $Ah = 0$.
2. $A^T y' + s' = A^T(y + k) + (s + f) = c$. Since $A^T y + s = c$, we get $A^T k + f = c - A^T y - s = 0$.
3. $x'_i s'_i = (x_i + h_i)(s_i + f_i) \approx \mu'$, or equivalently, $x_i s_i + h_i s_i + f_i x_i + h_i f_i \approx \mu'$. We drop the quadratic term $h_i f_i$ (if the steps h_i and f_i are small, the quadratic term $h_i f_i$ will be very small) and turn the approximate equality into an equality, i.e., we require $x_i s_i + h_i s_i + f_i x_i = \mu'$ for all i.

Thus, we have a system of linear equations for h_i, k_i, f_i, namely,

$$\text{system (S)} \qquad \begin{aligned} Ah &= 0 \\ A^T k + f &= 0 \\ h_i s_i + f_i x_i &= \mu' - x_i s_i \quad \text{for all } i \end{aligned}$$

We show in Theorem 1 that system (S) can be solved by "inverting" a matrix.

Remark 9. Note that there are m variables h_i, n variables k_j, and m variables f_i for a total of $2m + n$ unknowns. Also note that $Ah = 0$ constitutes n equations, $A^T k + f = 0$ constitutes m equations, and $h_i s_i + f_i x_i = \mu' - x_i s_i$ for all i comprises m equations. So we have $2m + n$ equations and the same number of unknowns. Also note that the x_i and s_i are *not* variables in this system, but fixed values.

Before we show that the system has a unique solution, we make some simple observations. From the third group of equations, we conclude

Claim 2. *$(x_i + h_i)(s_i + f_i) = \mu' + h_i f_i$, and $(x + h)^T(s + f) = m\mu' + h^T f$.*

Proof. From the third group of equations, we obtain

$$(x_i + h_i)(s_i + f_i) = x_i s_i + h_i s_i + f_i x_i + h_i f_i = \mu' + h_i f_i.$$

Summation over i yields

$$(x + h)^T(s + f) = \sum_i (x_i + h_i)(s_i + f_i) = \sum_i (\mu' + h_i f_i) = m\mu' + h^T f.$$

■

Claim 3. *$h^T f = f^T h = \sum_i h_i f_i = 0$, i.e., the vectors h and f are orthogonal to each other.*

Proof. Multiplying $A^T k + f = 0$ by h^T from the left, we obtain $h^T A^T k + h^T f = 0$. Since $h^T A^T = (Ah)^T = 0$, the equality $h^T f = 0$ follows. ∎

Claim 4. $c^T(x+h) - b^T(y+k) = (x+h)^T(s+f) = m\mu'$.

Proof. From Claims 2 and 3, $(x+h)^T(s+f) = m\mu' + h^T f = m\mu'$. Also, applying Claim 1 to the primal solution $x' = x+h$ and to the dual solution $(y', s') = (y+k, s+f)$ yields $c^T(x+h) - b^T(y+k) = (x+h)^T(s+f)$. ∎

Remark 10. Thus, $m\mu'$ is the objective value gap of the updated solution.

Theorem 1. *The system (S) has a unique solution.*

Proof. We will follow Vanderbei [8] and use capital letters (e.g. X) in this proof (only) to denote a diagonal matrix with entries of the corresponding row vector (e.g. X has the diagonal entries $x_1, x_2, \ldots, x_m$). We will also use e to denote a column vector of all ones (usually of length m).

Then, in the new notation, the last group of equations becomes

$$Sh + Xf = \mu' e - XSe.$$

Let us look at this equation in more detail.

$$\begin{aligned}
Sh + Xf &= \mu' e - XSe & & \\
h + S^{-1}Xf &= S^{-1}\mu' e - S^{-1}XSe & & \text{pre-multiply by } S^{-1} \\
h + S^{-1}Xf &= \mu' S^{-1}e - XS^{-1}Se & & \text{diagonal matrices commute} \\
h + S^{-1}Xf &= \mu' S^{-1}e - x & & \text{as } Xe = x \\
Ah + AS^{-1}Xf &= \mu' AS^{-1}e - Ax & & \text{pre-multiply by } A \\
AS^{-1}Xf &= \mu' AS^{-1}e - b & & \text{since } Ax = b \text{ and } Ah = 0 \\
-AS^{-1}XA^T k &= \mu' AS^{-1}e - b & & \text{using } f = -A^T k \\
b - \mu' AS^{-1}e &= (AS^{-1}XA^T)k & &
\end{aligned}$$

As XS^{-1} is diagonal with positive items, the matrix $W = \sqrt{XS^{-1}}$ is well-defined. Note that the diagonal terms are $\sqrt{x_i/s_i}$; since $x > 0$ and $s > 0$, we have $x_i/s_i > 0$ for all i. Thus, $AS^{-1}XA^T = AW^2A^T = (AW)(AW)^T$. Since A has full rank, $(AW)(AW)^T$, and hence $AS^{-1}XA^T$, is invertible (see Appendix). Thus,

$$k = (AS^{-1}XA^T)^{-1}\left(b - \mu' AS^{-1}e\right).$$

Then, we can find f from $f = -A^T k$. And to get h, we use the equation: $h + S^{-1}Xf = \mu' S^{-1}e - x$, i.e.,

$$h = -XS^{-1}f + \mu' S^{-1}e - x.$$

Thus, system (S) has a unique solution. ∎

Remark 11. What have we achieved at this point? Given feasible solutions (x, y, s) to the primal and dual problem, we can compute a solution $(x', y', s') = (x+h, y+k, s+f)$ to $Ax' = b$ and $A^T y' + s' = c$ that also satisfies $h^T f = 0$ and $x'^T s = m\mu'$ for any prescribed parameter μ'. Why do we not simply choose $\mu' = 0$ and be done? It is because we have ignored that we want $x' > 0$ and $s' > 0$. We will attend to these constraints in the next section.

3 Invariants in each Iteration

Recall that we want to construct solutions (x, y, s) to P_μ for smaller and smaller values of μ. The solution to P_μ will satisfy the following invariants. The first two invariants state that x is a positive solution of the primal and (y, s) is a solution to the dual with positive s. The third invariant formalized the condition $x_i s_i \approx \mu$ for all i.

1. (primal feasibility) $Ax^T = b$ with $x > 0$ (strict inequality).
2. (dual feasibility) $A^T y + s = c$ with $s > 0$ (strict inequality).
3. $\sigma^2 := \sum_i \left(\frac{x_i s_i}{\mu} - 1\right)^2 \le \frac{1}{4}$.

Remark 12. Even though the variance of $x_i s_i$ is $\frac{1}{m}\sum_i (x_i s_i - \mu)^2$, we still use the notation σ^2.

We need to show

$$x' > 0 \text{ and } s' > 0 \text{ and } \sigma'^2 := \sum_i \left(\frac{x'_i s'_i}{\mu'} - 1\right)^2 \le \frac{1}{4}.$$

We will do so for $\mu' = (1 - \delta)\mu$ and $\delta = \Theta\left(\frac{1}{\sqrt{m}}\right)$. Claim 2 gives us an alternative expression for σ'^2, namely,

$$\sigma'^2 = \sum_i \left(\frac{(x_i + h_i)(s_i + f_i)}{\mu'} - 1\right)^2 = \sum_i \left(\frac{h_i f_i}{\mu'}\right)^2 \tag{3}$$

We first show that the positivity invariants hold if σ' is less than one.

Fact 1. *If $\sigma' < 1$, then $x' > 0$, and $s' > 0$.*

Proof. We first observe that each product $x'_i s'_i = (x_i + h_i)(s_i + f_i) = \mu' + h_i f_i$ is positive. From $\sigma' < 1$, we get $\sigma'^2 < 1$. Since $\sigma'^2 = \sum_i (h_i f_i/\mu')^2$, each term of the summation must be less than one, and hence, $-\mu' < h_i f_i < \mu'$. In particular, $\mu' + h_i f_i > 0$ for every i. Thus, each product $(x_i + h)(s_i + f)$ is positive.

Assume for the sake of a contradiction that both $x_i + h_i < 0$ and $s_i + f_i < 0$. But as $s_i > 0$ and $x_i > 0$, this implies $s_i(x_i + h_i) + x_i(s_i + f_i) < 0$, or equivalently, $\mu' + x_i s_i < 0$, which is impossible because μ', x_i, s_i are all non-negative. This is a contradiction. ■

We next show $\sigma' \le 1/2$. We first establish

Claim 5. $\frac{\mu}{x_i s_i} \le \frac{1}{1-\sigma}$ *for all i and* $\sum_i \left|1 - \frac{x_i s_i}{\mu}\right| \le \sqrt{m} \cdot \sigma$.

Proof. As $\sigma^2 = \sum_i (1 - x_i s_i/\mu)^2$, each individual term in the sum is at most σ^2. Thus, $|1 - x_i s_i/\mu| \le \sigma$, and hence, $x_i s_i/\mu \ge 1 - \sigma$, and further, $\mu/x_i s_i \le 1/(1 - \sigma)$.

For the second claim, we have to work harder. Consider any m reals z_1 to z_m. Then $(\sum_i |z_i|)^2 \le m \sum_i z_i^2$; this is the frequently used inequality between the one-norm and the two-norm of a vector. Indeed,

$$m \sum_i z_i^2 - \left(\sum_i z_i\right)^2 = m \sum_i z_i^2 - \sum_i z_i^2 - 2\sum_{i<j} z_i z_j = (m-1)\sum_i z_i^2 - 2\sum_{i<j} z_i z_j = \sum_{i<j}(z_i - z_j)^2 \ge 0.$$

We apply the inequality with $z_i = 1 - x_i s_i/\mu$ and obtain the second claim. ■

Let us define two new quantities

$$H_i = h_i \sqrt{\frac{s_i}{x_i \mu'}} \quad \text{and} \quad F_i = f_i \sqrt{\frac{x_i}{s_i \mu'}}.$$

Observe that $\sum_i H_i F_i = \sum \frac{h_i f_i}{\mu'} = 0$ (from Claim 3) and $\sum_i (H_i F_i)^2 = \sum_i \left(\frac{h_i f_i}{\mu'}\right)^2 = \sigma'^2$. Also,

$$\begin{aligned} H_i + F_i &= \sqrt{\frac{1}{x_i s_i \mu'}}\,(h_i s_i + f_i x_i) = \sqrt{\frac{1}{x_i s_i \mu'}}\,(\mu' - \mu + \mu - x_i s_i) \\ &= \sqrt{\frac{\mu}{x_i s_i}\frac{\mu}{\mu'}}\left(\frac{\mu'}{\mu} - 1 + 1 - \frac{x_i s_i}{\mu}\right) = \sqrt{\frac{\mu}{x_i s_i (1-\delta)}}\left(-\delta + 1 - \frac{x_i s_i}{\mu}\right). \end{aligned} \tag{4}$$

Finally,

$$\begin{aligned} \sigma'^2 &= \sum_i (H_i F_i)^2 = \frac{1}{4}\left(\sum_i (H_i^2 + F_i^2)^2 - \sum_i (H_i^2 - F_i^2)^2\right) \\ &\le \frac{1}{4}\sum_i (H_i^2 + F_i^2)^2 && \text{since } \sum_i (H_i^2 - F_i^2)^2 \ge 0 \\ &\le \frac{1}{4}\left(\sum_i (H_i^2 + F_i^2)\right)^2 && \text{more positive terms} \\ &= \frac{1}{4}\left(\sum_i (H_i + F_i)^2\right)^2 && \text{since } H^T F = 0 \\ &= \frac{1}{4}\left(\sum_i \frac{\mu}{x_i s_i (1-\delta)}\left(-\delta + 1 - \frac{x_i s_i}{\mu}\right)^2\right)^2 && \text{by (4)} \\ &\le \frac{1}{4(1-\delta)^2(1-\sigma)^2}\left(\sum_i \left(-\delta + 1 - \frac{x_i s_i}{\mu}\right)^2\right)^2 && \text{since } \mu/(x_i s_i) \le 1/(1-\sigma) \\ &\le \frac{1}{4(1-\delta)^2(1-\sigma)^2}\left(m\delta^2 - 2\delta \sum_i \left(1 - \frac{x_i s_i}{\mu}\right) + \sum_i \left(1 - \frac{x_i s_i}{\mu}\right)^2\right)^2 && \text{remove inner square} \\ &\le \frac{1}{4(1-\delta)^2(1-\sigma)^2}\left(m\delta^2 + 2\delta \sum_i \left|1 - \frac{x_i s_i}{\mu}\right| + \sum_i \left(1 - \frac{x_i s_i}{\mu}\right)^2\right)^2 \\ &\le \frac{1}{4(1-\delta)^2(1-\sigma)^2}\left(m\delta^2 + 2\delta\sqrt{m}\cdot\sigma + \sigma^2\right)^2 && \text{by Claim 5} \\ &= \frac{1}{4(1-\delta)^2(1-\sigma)^2}\left(\left(\sqrt{m}\delta + \sigma\right)^2\right)^2, && \text{forming inner square} \end{aligned}$$

and hence,

$$\sigma' \le \frac{\left(\sqrt{m}\delta + \sigma\right)^2}{2(1-\sigma)(1-\delta)} \le \frac{\left(\sqrt{m}\delta + 1/2\right)^2}{2(1-1/2)(1-\delta)} \overset{!}{\le} \frac{1}{2}, \tag{5}$$

where the second inequality holds since the bound for σ' is increasing in σ, and $\sigma \le 1/2$. We need to choose δ such that the last inequality holds. This is why we put an exclamation mark on top of the

$\leq$-sign. Setting $\delta = c/\sqrt{m}$ for some to be determined constant c yields the requirement

$$\frac{(c+1/2)^2}{(1-\delta)} \overset{!}{\leq} \frac{1}{2}, \quad \text{or equivalently,} \quad (2c+1)^2 \overset{!}{\leq} 2\left(1 - \frac{c}{\sqrt{m}}\right).$$

This holds true for $c = 1/8$ and all $m \geq 1$. Thus, $\delta = 1/(8\sqrt{m})$.

Remark 13. Why do we require $\sigma \leq 1/2$ in the invariant? Let us formulate the bound as $\sigma \leq \sigma_0$ for some to be determined σ_0. Then, the inequality (5) becomes

$$\frac{\left(\sqrt{m}\delta + \sigma_0\right)^2}{2(1-\sigma_0)(1-\delta)} \overset{!}{\leq} \sigma_0.$$

We want this to hold for $\delta = \frac{c}{\sqrt{m}}$ and some $c > 0$. In order for the inequality to hold for $c = 0$, we need $\sigma_0 \leq 2(1-\sigma_0)$, or equivalently, $\sigma_0 \leq 2/3$. Since we want it to hold for some positive c, we need to choose a smaller σ_0; $1/2$ is a nice number smaller than $2/3$.

4 Initial Solution

This section follows Bertsimas and Tsitsiklis [1, p430]; see also Karloff [3, p128-129]. We have to deal with three problems: first, how to find an initial solution; second, how to make sure that we are dealing with a bounded problem; third, how to guarantee the third condition of the invariant for the initial solution. There are standard solutions for the first two problems.

Let us assume that we know a number W such that if (1) is bounded, there is an optimal solution x^* with $x_i^* < W$ for all i. Let e be the column vector of length m of all ones. We may then add the constraint $e^T x < mW$ to our problem without changing the optimal objective value. If (1) is unbounded, the additional constraint makes it bounded.

The standard solution for the second problem is the *big M method*. In the big M method, we introduce a new variable $z \geq 0$, change $Ax = b$ into $Ax + bz = b$ and the objective into "minimize $c^T x + Mz$", where M is a big number. We also have the constraint $e^T x^* < mW$. Note that $x = 0$ and $z = 1$ is a feasible solution to the modified problem. We solve the modified problem. If $z^* = 0$ in an optimal solution, we have also found the optimal solution to the original problem. If $z^* > 0$ in an optimal solution and M was chosen big enough, the original problem is infeasible.

We will see in Section 7 how to find the numbers W and M. We will now give the details and also show how to fulfill the third condition of the invariant for the initial solution, namely, $\sigma^2 = \sum_i (x_i s_i/\mu - 1)^2 \leq 1/4$.

We add two new nonnegative variables x_{m+1} and x_{m+2} and the constraint "$e^T x + x_{m+1} + x_{m+2} = (m+2)W$". Here, x_{m+2} is used for the big M method, and x_{m+1} is the slack variable for the constraint $e^T x + x_{m+2} \leq (m+2)W$. The new constraint can be satisfied by setting all variables to W. We are aiming for a particularly simple initial solution, namely $x_i = 1$ for $1 \leq i \leq m+2$ and, therefore, scale the variable x_i by $x_i = Wx'_i$.

Then, $e^T x + x_{m+1} + x_{m+2} = (m+2)W$ becomes $e^T Wx' + Wx'_{m+1} + Wx'_{m+2} = (m+2)W$, or equivalently, $e^T x' + x'_{m+1} + x'_{m+2} = m+2$.

$Ax = b$ becomes $WAx' = b$, or equivalently, $Ax' = \frac{1}{W} \cdot b$.

Finally, $c^T x$ becomes $c^T Wx' = Wc^T x'$. As W is a constant, the problem is equivalent to minimizing $c^T x'$. After replacing primed variables with unprimed variables, the problem is

$$\text{minimize } c^T x, \text{ subject to } Ax = d,\ e^T x + x_{m+1} + x_{m+2} = m+2 \text{ and } x \geq 0 \text{ with } d = \frac{1}{W} \cdot b.$$

We now come to the big M part. Let $\rho = d - Ae$. Then, $Ax + \rho x_{m+2} = d$ holds for $x_i = 1$, $1 \le i \le m+2$, and $x_{m+1} = x_{m+2} = 1$. We want a solution in which $x_{m+2} = 0$. Thus, we minimize $cx^T + Mx_{m+2}$ for a large M. We thus consider the artificial primal problem

$$\text{minimize } cx^T + Mx_{m+2}, \text{ subject to} \quad \begin{array}{ccccccc} Ax & & & + & \rho x_{m+2} & = & d \\ e^T x & + & x_{m+1} & + & x_{m+2} & = & m+2 \\ x \ge 0 & & x_{m+1} \ge 0 & & x_{m+2} \ge 0 & . & \end{array} \tag{6}$$

Remark 14. $W = 2(nU)^n$ suffices if all entries of A and b are integral and $U \ge \max_{ij} \left|a_{ij}\right|$ and $U \ge \max_i |b_i|$ as we will see in Section 7. Assume we also know a number $L > 0$ such that in every optimal solution x^* to (6), either $x^*_{m+2} = 0$ or $x^*_{m+2} > L$. Then $M = 4mU/L$ suffices, if also $U \ge \max_i |c_i|$. Indeed, if our original problem is feasible, then there is a feasible solution to (6) with $x_{m+2} = 0$. The objective value of this solution is less than or equal to $(m+2)U \le 2mU$ since $e^T x + x_{m+1} + x_{m+2} = m+2$ and $m \ge 2$. On the other hand, if $x^*_{m+2} > 0$ in an optimal solution to (6), then $x^*_{m+2} > L$, and hence the optimal objective value is larger than $ML - 2mU = 2mU$. Thus, our original problem is feasible if and only if x^*_{m+2} in every optimal solution to (6). We will see in Section 7 how to determine L.

Remark 15. Assume $x^*_{m+2} = 0$ in an optimal solution to (6). Then our original problem is feasible by the preceding remark. For x^*_{m+1} we distinguish two cases. If $x^*_{m+1} > 0$, then our original problem is bounded. If $x^*_{m+1} = 0$, the problem may be bounded or unbounded. Remark 6 explains how to distinguish these cases.

The dual problem (with new dual variables y_{n+1}, s_{m+1} and s_{m+2}) is

$$\text{maximize } d^T y + (m+2)y_{n+1}, \text{ subject to} \quad \begin{aligned} A^T y + e y_{n+1} + s &= c, \\ \rho^T y + y_{n+1} + s_{m+2} &= M \\ y_{n+1} + s_{m+1} &= 0 \end{aligned} \tag{7}$$

with slack variables $s \ge 0$, $s_{m+1} \ge 0$, $s_{m+2} \ge 0$ and unconstrained variables y.

Which initial solution should we choose? Recall that we also need to satisfy the third part of the invariant for some choice of μ, i.e., $\sum_{1 \le i \le m+2} (x_i s_i/\mu - 1)^2 \le 1/4$. Also, recall that we set x_i to 1 for all i. As $x_{m+1} = 1$, we choose $s_{m+1} = \mu/x_{m+1} = \mu$. Then, from the last equation, $y_{n+1} = -s_{m+1} = -\mu$. The simplest choice for the other ys is $y = 0$. Then, from the first equation, $s = c + e\mu$, and from the second equation $s_{m+2} = M - y_{n+1} = M + \mu$. Observe that all slack variables are positive (provided μ is large enough). For this choice,

$$\begin{aligned} \frac{x_i s_i}{\mu} - 1 &= \frac{c_i}{\mu} && \text{for } i \le m \\ \frac{x_{m+1} s_{m+1}}{\mu} - 1 &= 0 \\ \frac{x_{m+2} s_{m+2}}{\mu} - 1 &= \frac{M}{\mu}. \end{aligned}$$

Thus, $\sigma^2 = \left(M^2 + \sum c_i^2\right)/\mu^2$. We can make $\sigma^2 \le 1/4$ by choosing $\mu^2 = 4\left(M^2 + \sum c_i^2\right)$.

Summary: Let us summarize what we have achieved.

- For the artificial primal problem and its dual, we have constructed solutions $(x^{(0)}, y^{(0)}, s^{(0)})$ that satisfy the invariants for $\mu^{(0)} = 2\left(M^2 + \sum c_i^2\right)^{1/2}$.
- From the initial solution, we can construct a sequence of solutions $(x^{(i)}, y^{(i)}, s^{(i)})$ and corresponding $\mu^{(i)}$ such that
  - $x^{(i)}$ is a solution to the artificial primal,
  - $(y^{(i)}, s^{(i)})$ is a solution to its dual,
  - $\mu^{(i)} = (1-\delta) \cdot \mu^{(i-1)} = (1-\delta)^i \cdot \mu^{(0)}$, and $\sum_j \left(x_j^{(i)} s_j^{(i)}/\mu^{(i)} - 1\right)^2 \leq 1/4$.

  For $i \geq 1$, the difference between the primal and the dual objective value is exactly $(m+2)\mu^{(i)}$ (Claim 4). The gap decreases by a factor $1 - \delta = 1 - 1/(8\sqrt{m+2})$ in each iteration, and hence, can be made arbitrarily small.

In the next section, we will exploit this fact and show how to extract the optimal solution. Before doing so, we show the existence of an optimal solution.

Remark 16. Existence of an Optimal Solution: This paragraph requires some knowledge of calculus, namely continuity and accumulation point. Our sequence $(x^{(i)}, y^{(i)}, s^{(i)})$ has an accumulation point (this is clear for the sequence of x^i since the x-variables all lie between 0 and $m+2$ and we ask the reader to accept it for the others). Then there is a converging subsequence. Let (x^*, y^*, s^*) be its limit point. Then x^* and (y^*, s^*) are feasible solutions of the artificial primal and its dual respectively, and $x_i s_i = 0$ for all i by continuity.

5 Finding the Optimal Solution

This section is similar to [10, Theorem 5.3] and to the approach in [5, Section 3.3]. Let us assume that we know a positive number L such that any nonzero coordinate of an optimal solution to either primal or dual is at least L. We will see later (Section 7) how to find such a number in case all entries of A and b are integers.

Consider our sequence of iterates. We show: (1) if some x_i becomes sufficiently small, then $x_i^* = 0$ in all optimal solutions, and if some s_i becomes sufficiently small, then $s_i^* = 0$ in all optimal solutions. (2) If μ is sufficiently small, then either x_i or s_i will be sufficiently small.

Lemma 1. *Let (x, y, s) and μ satisfy the invariants. Let x^* be any optimal solution of the primal and (y^*, s^*) be any optimal solution of the dual. Assume that the smallest nonzero value of x_i^* and s_i^* is at least L.*

1. *If $x_i < \frac{L}{4m}$, then $x_i^* = 0$ in every optimal solution.*
2. *If $s_i < \frac{L}{4m}$, then $s_i^* = 0$ in every optimal solution.*

Proof. By the third part of our invariant, we have $\sigma^2 = \sum_i (\frac{x_i s_i}{\mu} - 1)^2 \leq \frac{1}{4}$. Thus, $(\frac{x_i s_i}{\mu} - 1)^2 \leq \frac{1}{4}$, and hence, $\mu/2 \leq x_i s_i \leq 3\mu/2 \leq 2\mu$ for all i. Further, $x^T s = \sum_i x_i s_i \leq 2m\mu$. By the first two parts of the invariant, x is a feasible solution of the primal and (y, s) a feasible solution to the dual.

Since x^* is an optimal solution, $c^T x \geq c^T x^*$. We apply Claim 1 first to the solution pair x and (y, s) and then to the pair x^* and (y, s) to obtain

$$x^T s = c^T x - b^T y \geq c^T x^* - b^T y = (x^*)^T s.$$

Assume $x_i < L/(4m)$. Since $x_i s_i \geq \mu/2$, we have $s_i \geq \mu/(2x_i) > 2m\mu/L \geq 1/L \cdot x^T s$. If $x_i^* > 0$, then $x_i^* \geq L$, and hence,

$$(x^*)^T s \geq x_i^* s_i > L \cdot 1/L \cdot x^T s = x^T s \geq (x^*)^T s,$$

a contradiction. Thus, $x_i < L/(4m)$ implies $x_i^* = 0$ in every optimal solution.

Since (y^*, s^*) is an optimal solution, $b^T y^* \geq b^T y$. We apply Claim 1 first to the solution pair x and (y, s) and then to the pair x and (y^*, s^*) to obtain

$$x^T s = c^T x - b^T y \geq c^T x - b^T y^* = x^T s^*.$$

Assume $s_i < L/(4m)$. Since $x_i s_i \geq \mu/2$, we have $x_i \geq \mu/(2s_i) > 2m\mu/L \geq 1/L \cdot x^T s$. If $s_i^* > 0$, then $s_i^* \geq L$, and hence,

$$x^T s^* \geq x_i s_i^* > 1/L \cdot x^T s \cdot L = x^T s \geq x^T s^*,$$

a contradiction. Thus, $s_i < L/(4m)$ implies $s_i^* = 0$ in every optimal solution. ∎

We now define two set of indices

$$B = \{i \,|\, s_i = 0 \text{ in all optimal solutions}, 1 \leq i \leq m\}, \text{ and}$$
$$N = \{i \,|\, x_i = 0 \text{ in all optimal solutions}, 1 \leq i \leq m\}.$$

Clearly, $B \cup N \subseteq \{1, 2, \ldots, m\}$.

Theorem 2 (Strong Duality). *For each i, either $x_i^* = 0$ in every optimal solution or $s_i^* = 0$ in every optimal solution. Thus, $c^T x^* - b^T y^* = (x^*)^T s^* = 0$, and $B \cup N = \{1, 2, \ldots, m\}$.*

Proof. As $x_i s_i < 2\mu$, if $\mu \leq \frac{L^2}{32m^2}$, then $x_i s_i < 2\mu \leq 2\frac{L^2}{32m^2} = \frac{L^2}{16m}$. Then, either $x_i < \sqrt{\frac{L^2}{16m}} = \frac{L}{4m}$ or $s_i < \sqrt{\frac{L^2}{16m}} = \frac{L}{4m}$, and hence, either $i \in B$ or $i \in N$ by the Lemma above . ∎

Remark 17. By the Strict Complementarity Theorem (see e.g. [6, pp 77-78] or [10, pp 20-21]), there are optimal solutions x^* and (y^*, s^*) in which $x_i^* > 0$ or $s_i^* > 0$; thus, both these conditions can not hold simultaneously. Thus, $B \cap N = \emptyset$. Further, from Theorem 2, the above partition is unique (see also [2]).

Remark 18. In the integer case (Section 7), if $x_i^* > 0$ or $s_i^* > 0$, then $x_i^* \geq \frac{1}{W}$ and $s_i^* \geq \frac{1}{W}$. Or, the lower bound $L = \frac{1}{W}$.

Let (x^*, y^*, s^*) be any optimal solution. As soon as $\mu < \frac{L^2}{32m^2}$, we can determine the optimal partition (B, N), i.e., $x_i^* = 0$ for $i \in N$, $s_i^* = 0$ for $i \in B$ and $B \cup N = \{1, \ldots, m\}$. We split the variables x into x_B and x_N, the variables s into s_B and s_N, the vector c into c_B and c_N, and our matrix A into A_B and A_N. Then our system (ignoring the nonnegativity constraints) becomes

$$A_B x_B + A_N x_N = b \quad \text{and} \quad A_B^T y + s_B = c_B \quad \text{and} \quad A_N^T y + s_N = c_N.$$

Since we know that $x_N^* = 0$ and $s_B^* = 0$ in every optimal solution, the system simplifies to

$$A_B x_B = b \quad \text{and} \quad A_B^T y = c_B \quad \text{and} \quad A_N^T y + s_N = c_N. \tag{8}$$

This is a system of $n + |B| + |N| = n + m$ equations in $|B| + m + |N| = n + m$ unknowns that is satisfied by every optimal solution.

Let us concentrate on the equation $A_B x_B = b$. If this equation has a unique solution, call it x_B^*, then (x_B^*, x_N^*) with $x_N^* = 0$ must be the optimal solution, as there is an optimal solution, every optimal solution satisfies $A_B x_B + A_N x_N = b$ and $x_N = 0$ in every optimal solution. In particular, $x_B^* \geq 0$. Note that if $A_B x_B = b$ has a unique solution, we can find it by Gaussian elimination.

What can we do if $A_B x_B = b$ has an entire solution set? We describe a simple method, which, however, is not the most efficient. There are more efficient methods, see, for example, [5, Section 3.3.5] or [10, Section 5.2.2], which do not increase the asymptotic running time. If $|B| < n$, the problem

$$\max c_B^T x_B, \quad \text{subject to } A_B x_B = b, x \geq 0$$

has fewer variables than the original primal, and we simply use the interior point method recursively on the smaller problem.

Fortunately, we can force the situation $|B| < n$ by using a technique called *perturbation*. Note that $|B| = n$, implies $N = \emptyset$. Thus $s_i^* = 0$ for all i in every optimal solution and hence the system $A^T y = c$ must have a solution. Thus we are guaranteed $N \neq \emptyset$ if $A^T y = c$ does not have a solution. Assume, it does. Note that A^T has n columns, c is an m-vector, and $m > n$. Instead of working with the objective direction c, we solve the problem for the direction $c' = c + c''$, where $c'' = (\varepsilon, \varepsilon^2, \ldots, \varepsilon^m)$, and ε is positive, but very close to zero. Geometrically, we perturb the optimal direction slightly so as to guarantee that the optimal solution is in a vertex of the feasible region and hence unique, see Figure 2. Moreover, if ε is small enough, the optimal solution for cost vector c' is also an optimal solution for cost vector c. Using the techniques from Section 7, one can compute an explicit value for ε. We will refrain from doing so. The perturbation also guarantees that $A^T y = c'$ does not have a solution for any positive sufficiently small ε.[1] Thus $N \neq \emptyset$ and hence $|B| < n$.

We are thus guaranteed that we eliminate at least one primal variable. We now use recursion to solve the smaller problem. As the number of variables decreases after every call, there can be at most $O(m)$ such calls or the running time will go up by a multiplicative factor of m.

6 Complexity

Let us assume that the initial value of μ is μ_0 and that we want to decrease μ to μ_f. Since every iteration decreases μ by the factor $(1 - \delta)$, we have $\mu = (1 - \delta)^r \mu_0$ after r iterations. The smallest r such that $(1 - \delta)^r \leq \mu_f$ is given by

$$\ln \frac{\mu_0}{\mu_f} \approx -r \ln(1 - \delta) \approx -r(-\delta),$$

or equivalently,

$$r = O\left(\frac{1}{\delta} \log \frac{\mu_0}{\mu_f}\right) = O\left(\sqrt{m} \log \frac{\mu_0}{\mu_f}\right).$$

If W is an upper bound on the coordinates of the optimal solution to our primal problem and L is a lower bound on a nonzero x_{m+2}^* in an optimal solution to (6), then from Section 4,

$$\mu_0^2 = 4\left(M^2 + \sum c_i^2\right) \leq 4\left(\frac{16m^2U^2}{L^2} + mU^2\right) \leq 68\frac{m^2U^2}{L^2}.$$

[1] Assume $A^T y = c''$ has a solution. Since A has more columns than rows, there is a nonzero x such that $Ax = 0$. Multiplying $A^T y = c''$ by x^T from the left yields $x^T c'' = x^T A^T y = (Ax)^T y = 0^T y = 0$. Next note that $x^T c'' = \sum_{1 \leq i \leq m} x_i \varepsilon^i$, i.e., ε is a zero of the m-th degree polynomial with coefficients x_m to x_1 and constant coefficient zero. Since a polynomial of degree m has at most m real zeros, we have $x^T c'' \neq 0$ for all sufficiently small positive ε, a contradiction.

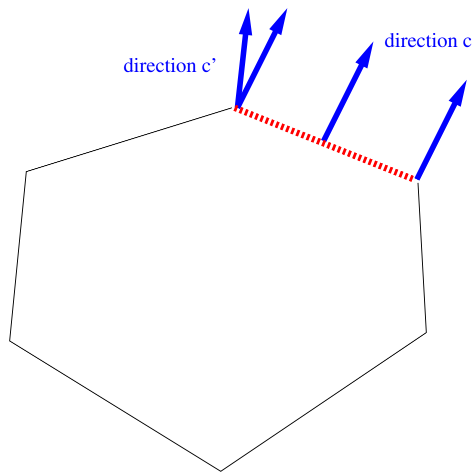

Figure 2: The cost vector c is orthogonal to the red (dashed) facet of the feasible region and hence all points on the red (dashed) facet are optimal. The cost vector c' is a small perturbation of c. With respect to cost vector c', the optimal solution is unique and also an optimal solution for cost vector c.

From Section 5, $\mu_f \geq \frac{L^2}{32m^2}$. Thus, the number of iterations will be

$$r = O\left(\sqrt{m}\log\frac{\mu_0}{\mu_f}\right) = O\left(\sqrt{m}\log\frac{mU/L}{L^2/m^2}\right) = O\left(\sqrt{m}(\log m + \log U + \log\frac{1}{L})\right).$$

For the integer case, as $\log L = O(n(\log n + \log U))$, the number of iterations will be

$$O\left(\sqrt{m}(\log m + n\,(\log n + \log U)\right).$$

7 The Bounds

In the previous sections, we used upper bounds on the components of an optimal solution and lower bounds on the nonzero components of an optimal solution. In this section, we derive these bounds. It assumes more knowledge of linear algebra, namely, determinants and Cramer's rule, and some knowledge of geometry. Unless stated otherwise, we assume that all entries of A and b are integers bounded by U in absolute value.

The determinant of a $n \times n$ matrix A is a sum of products, namely,

$$\det A = \sum_{\pi}(-1)^{\pi} a_{1\pi(1)}a_{2\pi(2)}\ldots a_{n\pi(n)}.$$

The summation is over all permutations π (with the appropriate sign) of n elements and the product corresponding to a permutation π selects the $\pi(i)$-th element in row i for each i. Each product is at most U^n. As there are $n!$ summands, we have $|\det A| \leq n!U^n < 2(nU)^n$; the 2 is only needed for $n = 1$, see [1, pp 373-374], [3, p75] or [6, pp 43-44].

Cramer's rule states that the solution of the equation $Ax = b$ (for a $n \times n$ non-singular matrix A) is $x_i = (\det A_i)/\det A$, where A_i is obtained by replacing the ith column of A with b.

Assume that the primal is bounded. As all constraints are linear, the solution space will be a convex polytope, and (by convexity) there will be an optimal solution that is a vertex. For each vertex, there is a submatrix A' of A obtained by keeping only n columns of A such that the corresponding coordinates of the vertex are $x_i = (\det A_i')/\det A'$. The remaining $m - n$ coordinates are zero. If we assume that each $|b_i| \le U$, the maximum value of $\left|\det A_i'\right|$ is no more than $n!U^n$. Also, $|\det A'| \ge 1$ since a nonzero integer is at least one in absolute value. Thus, $x_i^* < W = 2(nU)^n$ for the coordinates of vertex solutions of the original primal.

In the rest of this section, we mainly discuss bounds for the artificial problem. Let us next ask how small a nonzero coordinate of a vertex solution of the artificial primal problem (6) can be? The constraint system is

$$\begin{array}{ccccccc} Ax & & & + & (\frac{1}{W}b - Ae)x_{m+2} & = & \frac{1}{W}b \\ e^T x & + & x_{m+1} & + & x_{m+2} & = & (m+2). \end{array}$$

Any vertex solution is determined by some $(n+1)\times(n+1)$ nonsingular submatrix B of the left-hand side. In the column corresponding to x_{m+2}, the entries are bounded by $(m+1)U$, and all other entries are bounded by U. Since any product in the determinant formula for B can contain only one value of the column for x_{m+2}, we have $|\det B| \le (n+1)!(m+1)U^{n+1}$. Consider next $\det B_i$ where B_i is obtained from B by replacing one of the columns with the right-hand side. We need to lower bound $|\det B_i|$. The matrix B_i may contain two columns with fractional values. If we multiply these columns with W, we obtain an integer matrix. Thus, $|\det B_i| \ge 1/W^2$ if nonzero. Thus, any nonzero coordinate of a vertex solution of (6) is greater than L, where

$$L = \frac{1}{W^2}\cdot\frac{1}{2m\left((n+1)U\right)^{n+1}} \ge \frac{1}{8m\left((n+1)U\right)^{3(n+1)}}.$$

The constraint system of the dual (7) is

$$\begin{array}{ccccccccccc} A^T y & + & ey_{n+1} & + & s & & & & & = & c, \\ (\frac{1}{W}b - Ae)^T y & + & y_{n+1} & & & & & + & s_{m+2} & = & M \\ & & y_{n+1} & & & + & s_{m+1} & & & = & 0. \end{array}$$

The constraint matrix has $m+2$ rows and $n+1+m+2$ columns. The last $m+2$ columns contain an identity matrix, all entries in the column for y_{n+1} are one, and in the first n columns most entries are bounded by U. In the row with right-hand side M, the entries are bounded by $(m+1)U$. Any vertex solution is determined by some $(m+2)\times(m+2)$ nonsingular submatrix B of the left-hand side. At most $n+1$ columns of B belong to the first $n+1$ columns of the left-hand side. The other columns of B contain the identity matrix. Thus, $\det B$ is equal to the determinant of a square submatrix of the first $n+1$ columns of the left-hand side. We conclude that $|\det B| \le (n+1)!(m+1)U^{n+1}$. Consider next $\det B_i$ where B_i is obtained from B by replacing one of the columns with the right-hand side. The matrix B_i may contain one column with fractional values. If we multiply this column with W, we obtain an integer matrix. Thus, $|\det B_i| \ge 1/W$, if nonzero. Thus, any nonzero coordinate of a vertex solution of (7) is also greater than L.

Claim 6. *If all entries of A and b are integral and bounded by U in absolute value, then the coordinates of each vertex solution of the primal problem (1) are less than W. Any nonzero coordinate of a vertex solution of the artificial primal and its dual is at least L.*

Remark 19. If the entries of A and b are rational numbers, we write the entries in each column (or row) with a common denominator. Pulling them out brings us back to the integral case. For example,

$$\begin{vmatrix} 2/3 & 4/5 \\ 1/3 & 6/5 \end{vmatrix} = \frac{1}{15}\begin{vmatrix} 2 & 4 \\ 1 & 6 \end{vmatrix}.$$

Thus, if the determinant is nonzero, it is at least $1/15$.

Remark 20. If the entries are reals, we approximate each a_{ij} by a rational number r_{ij} with $1 - 1/n \leq a_{ij}/r_{ij} \leq 1 + 1/n$. Then, any product of n a_{ij}s is upper bounded by $(1 + 1/n)^n$ times the product of the corresponding r_{ij}s and lower bounded by $(1 - 1/n)^n$ times the product. Since $(1 + 1/n)^n \leq e \approx 2.71$ (e here being Euler's number) and $(1 - 1/n)^n \geq 1/e$, we can use the bounds for the rational case to get bounds for the real case.

Acknowledgments

The first author thanks Andreas Karrenbauer and Ruben Becker for intensive discussions and Andreas Karrenbauer for teaching from an earlier draft of this note. The work of the second author was inspired by an informal lecture given by Nisheeth Vishnoi at IIT Kanpur. The second author also thanks the students of CS602 (2014-15 and 2015-2016 batches) for their helpful comments and questions. Thanks also to Romesh Saigal for very prompt replies to queries.

References

[1] D. Bertsimas and J.N. Tsitsiklis, Introduction to linear optimization, Athena Scientific, 1997.

[2] Harvey J. Greenberg, The use of the optimal partition in a linear programming solution for postoptimal analysis, Operations Research Letters, 15 (1994), 179-185.

[3] H. Karloff, Linear Programming, Birkhauser, 1991.

[4] Goran Lesaja, Introducing Interior-Point Methods for Introductory Operations Research Courses and/or Linear Programming Courses, The Open Operational Research Journal, 2009, 3, 1-12.

[5] C. Roos, T. Terlaky and J-P. Vial, Interior Point Methods for Linear Optimization, 2nd Ed, 2006, Springer.

[6] Romesh Saigal, Linear Programming, A Modern Integrated Analysis, Kluwer, 1995.

[7] Tamas Terlaky, An easy way to teach interior-point methods, European J of Operational Research, 130 (2001), 1-19

[8] R.J. Vanderbei, Linear Programming: Foundations and Extension, 1st Ed: Kluwer Academic Publishers, 1997 (3rd Ed: Springer).

[9] Nisheeth K. Vishnoi, Convex Optimization Notes, `http://theory.epfl.ch/vishnoi/Nisheeth-VishnoiFall2014-ConvexOptimization.pdf`

[10] Yinyu Ye, Interior Point Algorithms, Wiley, 1997

Appendix: Result from Algebra

Assume that A is $n \times m$ matrix and the rank of A is n, with $n < m$. Then, all n rows of A are linearly independent. Or, $\alpha_1 A_1 + \alpha_2 A_2 + \ldots + \alpha_n A_n = 0$ (0 here being a row vector of size m) has only one solution $\alpha_i = 0$. Thus, if x is any $n \times 1$ matrix (a column vector of size n), then $x^T A = 0$ implies $x = 0$. Note that $(x^T A)^T = A^T x$. Thus, $A^T x = 0$ implies $x = 0$.

As A is $n \times m$ matrix, A^T will be $m \times n$ matrix. The product AA^T will be an $n \times n$ square matrix.

Consider the equation $(AA^T)x = 0$. Pre-multiplying by x^T we get $x^T AA^T x = 0$ or $(A^T x)^T (A^T x) = 0$. Now, $(A^T x)^T (A^T x)$ is the squared length of the vector $A^T x$. If a vector has length zero, all its coordinates must be zero. Thus, $A^T x = 0$, and hence, $x = 0$ by the preceding paragraph.

Thus, the matrix AA^T has rank n and is invertible.

Also observe that if X is a diagonal matrix (with all diagonal entries non-zero) and if A has full row-rank, then AX will also have full row-rank. Basically, if the entries of X are $x_1, x_2, \ldots, x_n$ then the matrix AX will have rows as $x_1 A_1, x_2 A_2, \ldots, x_n A_n$ (i.e., ith row of A gets scaled by x_i). If rows of AX are not independent, then there are βs (not all zero) such that $\beta_1 x_1 A_1 + \beta_2 x_2 A_2 + \ldots + \beta_n x_n A_n = 0$, or there are αs (not all zero) such that $\alpha_1 A_1 + \alpha_2 A_2 + \ldots + \alpha_n A_n = 0$ with $\alpha_i = \beta_i x_i$.