Still Simpler Way of Introducing Interior-Point method for Linear Programming

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

January 7, 2015

Abstract

Linear Programming is now included in Algorithm undergraduate and postgraduate courses for Computer Science majors. It is possible to teach interior-point methods directly with just minimal knowledge of Algebra and Matrices.

1 Introduction

Terlaky [7] and Lesaja [3] have suggested simple ways to teach interior-point methods. In this paper a still simpler way is being suggested. 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. In this paper, it is suggested if appropriate material is selected from these books then it becomes very easy to teach interior-point methods as the first or only method for Linear programming in Computer Science Courses.

Canonical Linear Programming Problem is to

minimise cx^T subject to $Ax = b$ and $x \geq 0$.

Here A is an $n \times m$ matrix, b and c are n-dimensional and x is an m-dimensional vector.

Remark 1 maximise cx^T is equivalent to minimise $-cx^T$.

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 (surplus) variable $\gamma \geq 0$.

Thus, we assume that there are n constraints and m variables, with $m > n$ (more variables and fewer constraints)— basically slack or surplus are added or subtracted to convert inequalities into equalities.

*E-mail: ssax@iitk.ac.in

We first use pivoting to make first term of all but the first equation as zero. Basically, we multiply i^{th}equation by $-\frac{a_{11}}{a_{i1}}$ and subtract the first equation. In similar way we make first two terms of all but the first two equations as zero– multiply i^{th}equation (for $i \neq 2$) by $-\frac{a_{22}}{a_{21}}$ and subtract the second equation. And so on. In case, if in any equation all coefficients become zero, we drop those equations. As a result, in the end all remaining equations will be linearly independent. Or the resulting matrix will have full row rank.

Remark We may have to interchange two columns (interchange two variables), in case, for example, if a diagonal term of an equation becomes zero.

From convexity, it is sufficient to obtain a locally optimal solution, as local optimality will imply global optimality.

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

> maximise by^T subject to $A^T y + s = c$, with slack variables $s \geq 0$ and variables y are unconstrained.

Claim 1 $by^T \leq cx^T$. *The equality will hold if and only if,* $s_i x_i = 0$ *for all is.*

Remark Thus if value of both primal and dual are the same, then both are optimal.

Proof: $s = c - A^T y$, or $x^T s = x^T c - x^T(A^T y) = c^T x - (x^T A^T)y = c^T x - b^T y$. As $x, s \geq 0$, we have $x^T s \geq 0$ or $c^T x \geq b^T y$.

Equality will hold if $x^T s = 0$ or $\sum_i s_i x_i = 0$ but as $s_i, x_i \geq 0$, we want each term (product) $s_i x_i = 0$. ∎

Thus, if we are able to find a solution of following equations (last one is not linear, else, an inversion of matrix would have been sufficient), we will be getting optimal solutions of both the original and the dual problems.

$$Ax = b, A^T y + s = c, x_i s_i = 0$$

subject to $x \geq 0, s \geq 0$.

We will relax the last condition to get something like (duality gap):

$$x_i s_i \approx \mu$$

with parameter $\mu \geq 0$. Thus, we will be solving (the exact last equation will be derived in the next section):

$$Ax = b, A^T y + s = c, x_i s_i \approx \mu$$

subject to $x \geq 0, s \geq 0$.

Remark Thus, $by^T - cx^T \approx m\mu$. If μ is very small, then in the case of rationals, the solution will be exact.

2 Use of Newton Raphson Method

Let us assume that we have an initial solution (x, s, y) (we will discuss how to get an initial solution in Section 5) which satisfies

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

We will use the Newton-Raphson method[5] to get "better" solution. Let us choose the next values as $x+h, y+k, s+f$. Then we want:

1. $A(x+h) = b$ or $Ax + Ah = b$ but as $Ax = b$, we get $Ah = 0$.
2. $A^T(y+k)+(s+f) = c$, from $A^Ty+s = c$, we get $A^Tk+f = c-A^Ty-s = 0$ or $A^Tk+f = 0$
3. $(x_i+h_i)(s_i+f_i) \approx \mu$ or $x_is_i+h_is_i+f_ix_i+h_if_i \approx \mu$. Or approximately, $x_is_i+h_is_i+f_ix_i = \mu$ (neglecting the non-linear h_if_i term). Thus, the equation we will be solving is

$$h_is_i + f_ix_i = \mu - x_is_i$$

Thus, we have a system of linear equations for h_i, k_i, f_i. We next show that these can be solved by "inverting" a matrix.

But first observe that from the third equation,

Observation 1 $(x_i+h_i)(s_i+f_i) = \mu + h_if_i$

Theorem 1 *Following equations have a unique solution:*

1. $Ah = 0$
2. $A^Tk + f = 0$
3. $h_is_i + f_ix_i = \mu - x_is_i$

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. in X the diagonal entries will be $x_1, x_2, \ldots, x_m$). We will also use e to denote a column vector of all ones (usually of length m).

Then in the new notation, the last equation is:

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

Let us look at this equation in more detail.

$$\begin{aligned}
Sh + Xf &= \mu e - XSe \\
h + S^{-1}Xf &= S^{-1}\mu e - S^{-1}XSe \text{ pre-multiply by } S^{-1} \\
h + S^{-1}Xf &= \mu S^{-1}e - X\cancel{S^{-1}}\cancel{S}e \text{ diagonal matrices commute} \\
h + S^{-1}Xf &= \mu S^{-1}e - x \text{ as } Xe = x \\
Ah + AS^{-1}Xf &= \mu AS^{-1}e - Ax \text{ pre-multiply by } A \\
AS^{-1}Xf &= \mu AS^{-1}e - b \text{ but } Ax = b \text{ and } Ah = 0 \\
-AS^{-1}XA^Tk &= \mu AS^{-1}e - b \text{ using } f = -A^Tk \\
b - AS^{-1}e &= (AS^{-1}XA^T)k
\end{aligned}$$

As XS^{-1} is diagonal with positive items and as A has full rank, if $W = \sqrt{XS^{-1}}$ (each diagonal term is $\sqrt{x_i/s_i}$) then $AS^{-1}XA^T = (AW)(AW)^T = AW^2A^T$ is invertible (see appendix). The last equation can thus be used to get the value of matrix k after inverting the matrix $AS^{-1}XA^T$, or

$$k = (AS^{-1}XA^T)^{-1}\left(b - AS^{-1}e\right)$$

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, the above system has a unique solution. ∎

Claim 2 $\sum_i h_i f_i = 0$ *or equivalently* $h^T f = f^T h = 0$

Proof: As $A^T k + f = 0$, we get $h^T A^T k + h^T f = 0$ but $h^T A^T = (Ah)^T = 0$, hence $h^T f = 0$ follows. ∎

3 Invariants in each Iteration

We will maintain following invariants:

1. $Ax^T = b$, with $x > 0$ (strict inequality)
2. $A^T y + s = c$ with $s > 0$ (strict inequality)
3. If μ is the "approximate duality gap" then $\sigma \leq \frac{2}{3} < \sqrt{3} - 1$ where $\sigma^2 = \sum_i \left(\frac{x_i s_i}{\mu} - 1\right)^2$.

At end of this iteration we want duality gap $\mu' \leq (1 - \delta)\mu$. We will see that δ can be chosen as $\delta = \Theta\left(\frac{1}{\sqrt{m}}\right)$.

We first show that strict inequality invariants hold (in σ' we have $x + h, s + f$ and same μ):

Fact 1 *If* $\sigma' < 1$ *then* $x + h > 0$ *and* $s + f > 0$

Proof: We first show that the product $(x_i + h_i)(s_i + f_i)$ is term-wise positive. From Observation 1, $(x_i + h_i)(s_i + f_i) = \mu + h_i s_i$.

From $\sigma' < 1$ we get $\sigma'^2 < 1$. But (using Observation 1):

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

As the sum is at most one, it follows that each term of the summation must be less than one, or $\left|\frac{h_i f_i}{\mu}\right| < 1$ or $-\mu < h_i f_i < \mu$. In particular $\mu + h_i f_i > 0$.

Thus the product $(x_i + h)(s_i + f)$ is term-wise positive.

Assume for contradiction that both $x_i + h_i < 0$ and $s_i + f_i < 0$. But as $s_i > 0$ and $h_i > 0$, we have $s_i(x_i + h_i) + x_i(s_i + f_i) < 0$, or $\mu + x_i s_i < 0$. Which is impossible as μ, x_i, s_i are all non-negative, a contradiction. ∎

We have to still show that the "approximate duality gap" μ decreases as desired.

Let us define three new variables:

$$\begin{aligned} H_i &= h_i \sqrt{\frac{s_i}{x_i \mu}} \\ F_i &= f_i \sqrt{\frac{x_i}{s_i \mu}} \end{aligned}$$

Observe that $\sum_i H_i F_i = \sum \frac{h_i f_i}{\mu} = 0$ (see Claim 2).

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

From, the proof of Fact 1, we also observe that $\sigma'^2 = \sum_i \left(\frac{h_i f_i}{\mu}\right)^2$, or $\sigma'^2 = \sum_i (H_i F_i)^2$

And finally

$$\begin{aligned} \sigma'^2 &= \sum_i (H_i F_i)^2 \\ &\leq \sum_i \frac{(H_i^2 + F_i^2)^2}{4} \quad \text{using AM-GM inequality-see Appendix} \\ &\leq \frac{1}{4} \left(\sum_i (H_i^2 + F_i^2) \right)^2 \quad \text{more positive terms} \\ &\leq \frac{1}{4} \left(\sum_i (H_i + F_i)^2 \right)^2 \quad \text{from Claim 2} \\ &= \frac{1}{4} \left(\sum_i \frac{\mu}{x_i s_i} \left(\frac{x_i s_i}{\mu} - 1 \right)^2 \right)^2 \\ &\leq \left(\max \frac{\mu}{x_i s_i} \right)^2 \frac{1}{4} \left(\sum \left(\frac{x_i s_i}{\mu} - 1 \right)^2 \right)^2 \\ &\leq \frac{\sigma^4}{4} \left(\max \frac{\mu}{x_i s_i} \right)^2 \end{aligned}$$

As $\sigma^2 = \sum \left(\frac{x_i s_i}{\mu} - 1\right)^2$, each individual term is at most σ^2 or

$$\left| \frac{x_i s_i}{\mu} - 1 \right| \leq \sigma$$

Thus,

$$\begin{array}{rcl} -\sigma & \leq \frac{x_i s_i}{\mu} - 1 & \leq \sigma \\ 1-\sigma & \leq \frac{x_i s_i}{\mu} & \leq 1+\sigma \end{array}$$

In particular $\frac{\mu}{x_i s_i} \leq \frac{1}{1-\sigma}$ or

$$\max \frac{\mu}{x_i s_i} \leq \frac{1}{1-\sigma}$$

Thus, $\sigma'^2 \leq \left(\frac{1}{1-\sigma}\right)^2 \frac{\sigma^4}{4}$ or $\sigma' \leq \frac{1}{2}\frac{\sigma^2}{1-\sigma}$

We summarise our observations as:

Observation 2

$$\sigma' \leq \frac{1}{2}\frac{\sigma^2}{1-\sigma}$$

For $\sigma' < 1$, $\frac{\sigma^2}{1-\sigma} < 2$ or $\sigma^2 < 2 - 2\sigma$ or $\sigma^2 + 2\sigma - 2 < 0$ or $\sigma = \frac{-2 \pm \sqrt{2^2 - 4\times 1 \times (-2)}}{2} = \frac{-2+\sqrt{12}}{2} = -1 + \sqrt{3} = \sqrt{3} - 1$.

Remark Thus $\sigma \leq \frac{2}{3}$ is more than enough.

Let us finally try to get bounds on δ (and hence μ). Let us assume $\mu' = \mu(1-\delta)$ then if σ'' corresponds to $x+h, s+f$ and μ' we have

$$\begin{aligned} \sigma''^2 &= \sum_i \left(\frac{(x_i+h_i)(s_i+f_i)}{\mu(1-\delta)} - 1\right)^2 \\ &= \sum_i \left(\frac{(x_i+h_i)(s_i+f_i) - \mu(1-\delta)}{\mu(1-\delta)}\right)^2 \\ &= \sum_i \left(\frac{(x_i+h_i)(s_i+f_i) - \mu}{\mu(1-\delta)} + \frac{\delta}{1-\delta}\right)^2 \\ &= \sum_i \left(\frac{h_i f_i}{\mu(1-\delta)} + \frac{\delta}{1-\delta}\right)^2 \text{ From Observation 1} \\ &= \frac{1}{(1-\delta)^2} \sum_i \left(\frac{h_i f_i}{\mu} + \delta\right)^2 \\ &= \frac{1}{(1-\delta)^2} \left(\sum_i \left(\frac{h_i f_i}{\mu}\right)^2 + m\delta^2 + \frac{2\delta}{\mu}\sum h_i f_i\right) \\ &= \frac{1}{(1-\delta)^2} \left(\sum_i \left(\frac{h_i f_i}{\mu}\right)^2 + m\delta^2\right) \text{ From Claim 2} \\ &= \frac{1}{(1-\delta)^2}\left(\sigma'^2 + m\delta^2\right) \end{aligned}$$

Thus observe that

Observation 3

$$\sigma'' = \frac{1}{1-\delta}\sqrt{\sigma'^2 + m\delta^2}$$

We want to choose δ such that $\sigma'' \leq \frac{2}{3}$. As $\sigma'' = \frac{1}{1-\delta}\sqrt{\sigma'^2 + m\delta^2} \leq \frac{1}{1-\delta}\sqrt{m\delta^2} = \frac{\delta\sqrt{m}}{1-\delta}$

We want $\frac{\delta\sqrt{m}}{1-\delta} < \frac{2}{3}$ or $\frac{\delta}{1-\delta} < \frac{2}{3\sqrt{m}}$. We can thus choose $\delta = \frac{1}{4\sqrt{m}}$.

4 Summary

Let us assume that initial duality gap is μ_0 and final duality gap is μ_f, as after each iteration, $\mu' \leq (1-\delta)\mu$, thus after r iterations, $\mu_f \leq (1-\delta)^r \mu_0$, or

$$\log \frac{\mu_0}{\mu_f} = -r \log(1-\delta) \approx -r(-\delta)$$

or

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

As $1 - \sigma \leq \frac{x_i s_i}{\mu} \leq 1 + \sigma$, we have (in last inequality we use $\sigma < \frac{2}{3}$).

$$\mu(1-\sigma) \leq x_i s_i \leq \mu(1+\sigma) < \frac{5}{3}\mu$$

Thus, when μ becomes very small, even the products $x_i s_i$s will be very small. The above method will give a "polynomial time" algorithm even if $\mu_0 = 2^{L^2}$ and $\mu_f = \frac{1}{2^{L^2}}$, with $L = \sum_{i,j} \log_2(1 + A[i,j]) + \sum_i \log_2(1 + b[i])$ (the size of input in bits[6, p43]).

5 Initial Solution

This section is based on description of Mehlhorn[4].

Let us first assume that there is a number (say) W such that there is an optimal solution x^* for which each $x_i \leq W$; we will see later (see Section 6) how to find such a number in case all entries of A and b are integers. If e is a column vector (of length m) of all ones, then $e^T x^* < mW$.

Thus [1, p430] (see also [2, p128-129]) an optimal solution of the problem

minimise cx^T subject to $Ax = b$, $e^T x < mW$ and $x \geq 0$.

will also be a solution of the original problem (without $e^T x < mW$ constraint).

Remark In case some x_i turns out to be larger than W, then it is a sign that W was not large enough.

Let us replace (scale) variables x_i by

$$\frac{mWx'_i}{m+2}$$

then the problem becomes:

minimise $\frac{mW}{m+2}cx'^T$ subject to $Ax' = d$, $e^T x' < m+2$ and $x' \geq 0$ with $d = b(m+2)/(mW)$

We add a new variable x'_{m+1} and replace $e^T x' < m+2$ by $e^T x' + x'_{m+1} = m+2$ (with $x_{m+1} \geq 0$). Or dropping primes, the problem is equivalent to:

minimise cx^T subject to $Ax = d$, $e^T x + x_{m+1} = m+2$ and $x \geq 0$.

Consider a starting solution x_0 s.t. all components of x_0 are strictly positive (say all $x_i = 1$ or $x = e$, $x_{m+1} = 1$). Define a vector $\rho = d - Ae$. Let x_{m+2} be one more new variable. Then $Ax + x_{m+2}\rho = d$ with $x \geq 0, x_{m+2} \geq 0$ has a solution with $x = e$ and $x_{m+1} = x_{m+2} = 1$. For this choice, $e^T x + x_{m+1} + x_{m+2} = m+2$ is also true. We want a solution in which $x_{m+2} = 0$. Thus, we try to minimise $cx^T + Mx_{m+2}$ for a large M.

Remark It is sufficient to choose $M > mW \times \max|c_i|$.

We thus consider the artificial primal problem:

minimise $cx^T + Mx_{m+2}$ subject to $Ax + \rho x_{m+2} = d$, $e^T x + x_{m+1} + x_{m+2} = m+2$ and $x \geq 0, x_{m+1} \geq 0$ and $x_{m+2} \geq 0$.

Remark If in optimal solution $x_{m+2} > 0$, then either there is no feasible solution, or the value of M chosen was not large enough.

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

maximise $dy^T + (m+2)y_{n+1}$ subject to
$A^T y + ey_{n+1} + s = c$,
$\rho^T y + y_{n+1} + s_{m+2} = M$
$y_{n+1} + s_{m+1} = 0$ with slack variables $s \geq 0, s_{m+1} > 0, s_{m+2} > 0$ and variables y are unconstrained.

To get an initial solution, as $x_{m+1} = 1$, we try $s_{m+1} = \mu/x_{m+1} = \mu$. Then from the last equation $y_{n+1} = -s_{m+1} = -\mu$. The simplest choice will be to choose all other $y = 0$ then from first equation $s = c + e\mu$ which is again a positive number (if μ is larger than all $-c_i$s). To satisfy the second equation we must choose $s_{m+2} = M - y_{n+1} = M + \mu$. Observe that all slack variables are positive (provided μ is large enough).

For this choice, $x_i s_i = c_i + \mu$ or $\frac{x_i s_i}{\mu} - 1 = \frac{c_i}{\mu}$ and $x_{m+1}s_{m+1} = \mu$ and $\frac{x_{m+2}s_{m+2}}{\mu} - 1 = \frac{M}{\mu}$. Thus $\sigma^2 = \frac{1}{\mu^2}\left(\sum c_i^2 + M^2\right)$. We can make $\sigma^2 < \frac{1}{4}$ by choosing $\mu^2 = 4\left(\sum c_i^2 + M^2\right)$.

6 Integer Case

This section assumes some more knowledge of algebra– determinants and Cramer's rule and some knowledge of geometry.

If A is an $n \times n$ matrix then $\det |A| = \sum_{\pi} (-1)^{\pi} a_{1\pi(1)} a_{2\pi(2)} \ldots a_{n\pi(n)}$ will be sum of all possible (products) of permutations π (with appropriate sign). Clearly $\det |A| = \sum_{\pi} (-1)^{\pi} a_{1\pi(1)} a_{2\pi(2)} \ldots a_{n\pi(n)} \leq \sum_{\pi} \left| a_{1\pi(1)} a_{2\pi(2)} \ldots a_{n\pi(n)} \right|$. If[1] each $a_{ij} \leq U$, then $\det|A| \leq n! U^n$.

Cramer's rule says that solution of equation $Ax = b$ (for $n \times n$ non-singular matrix A) is $x_i = \det(A_i)/\det(A)$ where A_i is obtained by replacing i^{th}column of A by b.

As all constraints are linear, solution space will be a convex polytope and (by convexity) for optimal solution it is sufficient to look at corner points. At each corner point exactly n components of x will be non-zero; remaining $m-n$ x_i will be zero. Thus, at optimal solution $x_i = \det(A'_i)/\det(A')$, where A' is obtained by keeping only n columns of A. If we assume that each $|b_i| \leq U$, the maximum value of the determinant can be $n!U^n$. If all entries are integers, then determinant has to be at least one if it is non-zero.

Thus, each x_i is between $1/(n!U^n)$ and $n!U^n$. Or we can choose $W = n!U^n < (nU)^n$.

Acknowledgement

This work was inspired by an informal lecture given by Nisheeth Vishnoi at IIT Kanpur. I also wish to thank students of CS602 (2014-2015 batch) for their helpful comments and questions when I was teaching this material. Thanks also to Kurt Mehlhorn for pointing out that method for finding initial solution (of an earlier version) may not work and for his suggestion of using method of [1, p430] instead in Section 5.

References

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

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

[3] 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.

[4] Kurt Mehlhorn, Private Communication, 2014.

[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).

[1](see also e.g. [1, pp 373-374], [2, p75] or [6, pp 43-44])

Appendix: Result from Algebra

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

Consider the equation $(AA^T)y^T = 0$. Pre-multiplying by y we get $yAA^Ty^T = 0$ or $(yA)(yA)^T = 0$ or the dot product $< yA, yA >= 0$ which, for real vectors (matrices) means, that each term of yA is (individually) zero, or y is identically zero.

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 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., i^{th}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 $\alpha_i = \beta_i x_i$.

Arithmetic-Geometric Inequality follows from:

$$(x + y)^2 \geq (x + y)^2 - (x - y)^2 = 4xy$$

or taking square roots (of the first and the last) $x + y \geq 2\sqrt{xy}$