DTIC FILE COPY

(4)

AD-A201 342

CENTER FOR CYBERNETIC STUDIES

The University of Texas
Austin, Texas 78712

88 11 14 069

(4)

CCS Research Report 605

A GEOMETRIC PROPERTY OF THE LEAST SQUARES SOLUTION OF LINEAR EQUATIONS

by

Aharon Ben-Tal
Marc Teboulle

June 1988

The research for this paper was partly supported by National Science Foundation Grants SES 8520806 and SES 8722504 and U.S. Army Contract DAKF-15-87-C-0110, with the Center for Cybernetic Studies at The University of Texas at Austin. Reproduction in whole or in part is permitted for any purpose of the U.S. Government.

CENTER FOR CYBERNETIC STUDIES

A. Charnes, Director

College of Business Administration, 5.202
The University of Texas at Austin
Austin, Texas 78712-1177
(512) 471-1821

Abstract

We derive an explicit determinantal formula for the least squares (LS) solution of an overdetermined system of linear equations. From this formula it follows that the LS solution lies in the convex hull S of points, each of which is a solution of a square subsystem of the whole system. The results are extended to weighted LS solution; it is shown that the convex hull S in which the solution must lie is independent of the weighting matrix.

Key words. Linear equations, Least squares problems, Convexity.

1. Introduction

Consider the overdetermined system of linear equations

$$Ax = b \tag{1.1}$$

where A is $m \times n$ real matrix ($m > n$). The Least squares solution of (1.1) is obtained from the solution of the normal equation

$$A^T A x = A^T b \tag{1.2}$$

The $m \times n$ system (1.1) contains $\binom{m}{n}$ squared $n \times n$ subsystems (obtained by all possible choices of n rows of A out of m) which can be indexed by $j \in \{1, \ldots, \binom{m}{n}\} := P_{m,n}$. Let y_{I_j} be the solution of the j-th subsystem corresponding to rows of A indexed by $I_j = \{i_1, \ldots, i_n\} \subset P_{m,n}$, i.e.

$$A_{I_j} y_{I_j} = b_{I_j} \tag{1.3}$$

The solution to (1.3) exists and is unique if $\Delta_{I_j} = det A_{I_j} \neq 0$.

In this paper we obtain a new representation of the least squares solution x of (1.2) as follows:

$$x = \sum_{j \in P^+_{m,n}} \lambda_{I_j} y_{I_j} \tag{1.4}$$

where

$$P^+_{m,n} = \{j \in P_{m,n} : \Delta_{I_j} \neq 0\} \quad \text{and} \quad \lambda_{I_j} = \frac{\Delta^2_{I_j}}{\sum_{j \in P^+_{m,n}} \Delta^2_{I_j}}$$

The λ_{I_j}'s are weights satisfying: $\lambda_{I_j} > 0$ *and* $\sum_{j \in P^+_{m,n}} \lambda_{I_j} = 1$, hence from the representation (1.4)

$$x \in \text{ convex hull of } \{y_{I_j} : \ j \in P^+_{m,n}\} \tag{1.5}$$

We further show that the result (1.5) remains true for x being the solution of a *weighted least squares problem*

$$A^T M A x = A^T M b \tag{1.6}$$

where M is a positive definite diagonal matrix. For this case each y_{I_j} is still the solution of (1.3), *independent* of M. This result is useful for numerical methods whose $(k+1)$ st iterate x_{k+1} is the solution of a system

$$A^T M_k \, A \, x_{k+1} = A^T M_k \, b \tag{1.7}$$

The iterative procedure (1.7) is in the heart of Karmakar's Linear Programming method [4] and many related methods, see e.g. [3] and the references therein. From our result it follows that, starting from any x_0, the next iteration points $x_1, x_2, \ldots,$ all remain in the fixed compact polyhedral set $S = \mathrm{conv}\{y_{I_j} : \ j \in P^+_{m,n}\}$. The fact is important for proving convergence of the sequence $\{x_k\}$.

2. Some notations and preliminary results

In this section we introduce the notations used in this paper and we prove some results on the computation of determinants which will be necessary in order to derive our main results.

We denote by $\mathbb{R}^{m\times n}$ the space of $m \times n$ real matrices with $m > n$ and we assume that $A \in \mathbb{R}^{m\times n}$ with rank $A = n$. The transpose of A is denoted A^T. The matrix obtained from A by replacing column k with $x \in \mathbb{R}^m$ is denoted by:

$$A(k \to x).$$

The minor of A of order p corresponding to rows indexed by i_j and columns indexed by k_j is denoted by:

$$A\begin{pmatrix} i_1 & i_2 \ldots i_p \\ k_1 & k_2 \ldots k_p \end{pmatrix}$$

provided $1 \leq i_1 < i_2 < \ldots < i_p \leq m$ and $1 \leq k_1 < k_2 < \ldots < k_p \leq n$. If A is a square matrix $(m = n)$then the determinant of A is:

$$detA = A\begin{pmatrix} 1 & 2 \ldots n \\ 1 & 2 \ldots n \end{pmatrix}$$

An important formula for computing the determinant of the product of two rectangular matrices is the so called *Binet-Cauchy* formula see e.g. [2]. Let $D \in \mathbb{R}^{n\times m}$ and $E \in \mathbb{R}^{m\times n}$ be given matrices with $m \geq n$, and let $C := DE$ i.e. $C \in \mathbb{R}^{n\times n}$. Then,

$$detC = \sum_{1\leq i_1<i_2<\ldots<i_n\leq m} D\begin{pmatrix} 1 & 2 \ldots n \\ i_1 & i_2 \ldots i_n \end{pmatrix} E\begin{pmatrix} i_1 & i_2 \ldots i_n \\ 1 & 2 \ldots n \end{pmatrix} \tag{2.1}$$

A direct application of the Binet-Cauchy formula leads to the following result.

Lemma 2.1 *Let $A \in \mathbb{R}^{m\times n}$. Then,*

$$det(AA^T) = \sum_{1\le i_1<i_2<\ldots<i_n\le m} \left[A \begin{pmatrix} i_1 & i_2\ldots i_n \\ 1 & 2\ldots n \end{pmatrix} \right]^2$$

Proof. Let $D = A^T$ and $E = A$ in (2.1). Then,

$$detC = det(AA^T) = \sum_{1\le i_1<\ldots<i_n\le m} A^T \begin{pmatrix} 1 & 2\ldots n \\ i_1 & i_2\ldots i_n \end{pmatrix} A \begin{pmatrix} i_1 & i_2\ldots i_n \\ 1 & 2\ldots n \end{pmatrix}$$

Since

$$A^T \begin{pmatrix} 1 & 2\ldots n \\ i_1 & i_2\ldots i_n \end{pmatrix} = A \begin{pmatrix} i_1 & i_2\ldots i_n \\ 1 & 2\ldots n \end{pmatrix}$$

the result follows. □

Remark 2.1 An alternative proof of Lemma 2.1 is given by Linnik, [5, p. 25].

The next result is a simple multiplication rule between two specific matrices.

Lemma 2.2 *For all $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m\times n}$*

$$A^T A(k \to A^T b) = A^T \cdot A(k \to b)^1$$

Proof. We prove the result for $k = 1$. Suppose that column 1 is replaced in $A^T A$ by $A^T b$, then the first column of $A^T A$ is

$$\sum_{r=1}^{m} a_{rp} b_r, \quad p = 1, \ldots, n$$

Now replacing column 1 by b in A implies that the only column changed in the product $A^T \cdot A(1 \to b)$ is the first column, which is simply $\sum_{r=1}^{m} a_{rs} b_r, \quad s = 1, \ldots, n$. Hence,

$$A^T A(1 \to A^T b) = A^T \cdot A(1 \to b)$$

Repeating the same argument for each column $k = 1, \ldots m$, the result follows. □

In the sequel we will need the following notations. Let $N_{m,n}$ denote all subset of n integers out of $\{1, 2, \ldots, m\}$, $(m > n)$, i.e.

$$N_{m,n} = \{(i_1, \ldots, i_n : \ 1 \le i_1 < \ldots < i_n \le m\}$$

[1] The · notation is used here to indicate that in the RHS we replace column k with the vector b in the matrix A only

We order the members of $N_{m,n}$ by increasing lexicographic order. Let I_j be the j-th such index set, where $j \in P_{m,n} := \{1, 2, \ldots, \begin{pmatrix} m \\ n \end{pmatrix}\}$ Unless otherwise specified, all the summations in the rest of the paper are taken over $j \in P_{m,n}$. Using our new notations, Lemma 2.1 can be rewritten as:

$$\Delta := det A^T A = \sum \Delta_{I_j}^2$$

where Δ_{I_j} is the determinant of the $n \times n$ submatrix A_{I_j} (with rows indices in I_j). The corresponding subvector of $b \in \mathbb{R}^m$ is denoted $b_{I_j} \in \mathbb{R}^n$

Lemma 2.3 *Let $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. Then,*

$$det A^T A \, (k \to A^T b) = \sum det\, A_{I_j}\, det A_{I_j}\, (k \to b_{I_j})$$

Proof. From Lemma 2.2 we have:

$$det A^T A(k \to A^T b) = det A^T \cdot A(k \to b)$$

By Binet-Cauchy formula:

$$det A^T \cdot A(k \to b) = \sum det A_{I_j}^T \cdot det A_{I_j}(k \to b_{I_j})$$

and since $det A_{I_j}^T = det A_{I_j}$, the Lemma is proved. □

3. Results

Consider the system of linear equations

$$Ax = b \quad A \in \mathbb{R}^{m\times n},\ b \in \mathbb{R}^m,\ \ m > n \tag{3.1}$$

We assume that A is full column rank, i.e. $\operatorname{rank} A = n$. For each index set $I_j \in N_{m,n}$ $(j \in P_{m,n})$ there is an associated square subsystem of n linear equations out of (3.1):

$$A_{I_j} x = b_{I_j}$$

If $\Delta_{I_j} = det A_{I_j} \neq 0$ this system has a unique solution which we denote by y_{I_j}, i.e.

$$A_{I_j} y_{I_j} = b_{I_j}, \quad j \in P_{m,n}^+ = \{j \in P_{m,n} : \ \Delta_{I_j} \neq 0\} \tag{3.2}$$

Note that $P_{m,n}^+ \neq \emptyset$ since $\operatorname{rank} A = n$.

The *least Euclidean norm* solution of (3.1) (*least squares* solution for short) is the unique solution of the so called *normal equations*: (see e.g. [1])

$$A^T Ax = A^T b \tag{3.3}$$

The uniqueness follows from $\Delta := det A^T A \neq 0$ which is implied by rank$A = n$. The next result gives an explicit representation of the least squares solution as a *convex combination of the* y_{I_j}'s.

Theorem 3.1 *The least squares solution of the system (3.1) is given by*

$$x = \sum_{j \in P^+_{m,n}} \lambda_{I_j} y_{I_j}$$

where $\lambda_{I_j} = \frac{\Delta^2_{I_j}}{\Delta}$. *Moreover,* $\lambda_{I_j} > 0$, *and* $\sum_{j \in P^+_{m,n}} \lambda_{I_j} = 1$. *Hence,*

$$x \in conv \{y_{I_j} : \quad j \in P^+_{m,n}\}$$

Proof. Since $\Delta = det A^T A \neq 0$, the solution of (3.3) is unique and is given by the Cramer rule (componentwise)

$$x_k = \frac{det A^T A(k \rightarrow A^T b)}{det A^T A} \quad k = 1, \ldots, n. \tag{3.4}$$

By Lemma 2.3, (3.4) can be written as:

$$\begin{aligned} x_k &= \frac{\sum det A_{I_j} \cdot det A_{I_j}(k \rightarrow b_{I_j}}{det A^T A} \\ &= \frac{\sum_{j \in P^+_{m,n}} \Delta_{I_j} \cdot \Delta^k_{I_j}(b)}{\Delta} \end{aligned} \tag{3 5}$$

where

$$\Delta^k_{I_j}(b) := det A_{I_j}(k \rightarrow b_{I_j})$$

Defining $\lambda_{I_j} := \Delta^2_{I_j}/\Delta$, (3.5) can be written as:

$$x_k = \sum_{j \in P^+_{m,n}} \lambda_{I_j} \frac{\Delta^k_{I_j}(b)}{\Delta_{I_j}}, \quad k = 1, \ldots, n \tag{3.6}$$

Consider now the solution of the linear equation (3.2):

$$A_{I_j} y_{I_j} = b_{I_j} \quad \text{for} \quad j \in P^+_{m,n}$$

Using Cramer rule , the solution is given componentwise by

$$y_{I_j}^k = \frac{det A_{I_j}(k \to b_{I_j})}{det A_{I_j}} = \frac{\Delta_{I_j}^k(b)}{\Delta_{I_j}}, \quad k = 1, \ldots, n \tag{3.7}$$

Combining (3.6) and (3.7) it follows that

$$x_k = \sum_{j \in P_{m.n}^+} \lambda_{I_j} y_{I_j}^k, \quad k = 1, \ldots, n$$

Furthermore, by Lemma 2.1, $\Delta = \sum_{j \in P_{m.n}} \Delta_{I_j}^2 = \sum_{j \in P_{m.n}^+} \Delta_{I_j}^2$, hence

$$\lambda_{I_j} = \frac{\Delta_{I_j}^2}{\Delta} > 0, \text{ and } \sum_{j \in P_{m.n}^+} \lambda_{I_j} = 1. \ \square$$

Theorem 3.1 can be generalized to the solution of a *weighted least-squares* problem. Let $M \in \mathbb{R}^{m \times m}$ be the diagonal matrix

$$M = \text{diag}(\mu_1, \ldots, \mu_m)$$

with diagonal elements $\mu_i > 0, \ \ i = 1, \ldots, m$. The solution of the weighted least-squares problem with weight matrix M is obtained by solving the normal equation:

$$A^T M A = A^T M b \tag{3.8}$$

Since M is positive definite (3.8) can be written as:

$$A^T M^{1/2} M^{1/2} A = A^T M^{1/2} M^{1/2} b \tag{3.9}$$

where

$$M^{1/2} = \text{diag}(\sqrt{\mu}_1, \ldots, \sqrt{\mu}_m)$$

Define $\tilde{A} = M^{1/2}A, \quad \tilde{b} = M^{1/2}b$, then since $(M^{1/2})^T = M^{1/2}$, (3.9) is equivalent to

$$\tilde{A^T} A x = \tilde{A}^T \tilde{b} \tag{3.10}$$

i.e. x is the least squares solution of the linear equation

$$\tilde{A} x = \tilde{b}$$

Applying Theorem 3.1 we thus obtain

$$x = \sum_{j \in P_{m.n}^+} \tilde{\lambda}_{I_j} \tilde{y}_{I_j}$$

where y_{I_j} solves

$$\tilde{A}_{I_j}\tilde{y}_{I_j} = \tilde{b}_{I_j}$$

and

$$\tilde{\lambda}_{I_j} = \frac{\tilde{\Delta}^2_{I_j}}{\tilde{\Delta}} := \frac{det^2 \tilde{A}_{I_j}}{\sum det^2 \tilde{A}_{I_j}} \tag{3.11}$$

We show below that the determinants $\tilde{\Delta}_{I_j}$ and $\tilde{\Delta}$ can be expressed explicitly in terms of Δ_{I_j}, Δ and μ_j, and that the coefficients $\tilde{\lambda}_{I_j}$ can be expressed in terms of λ_{I_j} and μ_j. Therefore the least squares solution of the weighted problem can be obtained directly from the least norm solution of the original linear equation $Ax = b$ by a simple scaling of the convex combination coefficients λ_{I_j}.

Theorem 3.2 *The weighted least squares solution of (3.8) is given by*

$$x = \sum_{j \in P^+_{m,n}} \tilde{\lambda}_{I_j} y_{I_j}$$

where y_{I_j} solves (3.2), and where

$$\tilde{\lambda}_{I_j} = \frac{\pi^2_{I_j}\lambda_{I_j}}{\sum_{j \in P^+_{m,n}} \pi^2_{I_j}\lambda_{I_j}}, \quad j \in P^+_{m,n}$$

with

$$\pi_{I_j} := \prod_{j \in P^+_{m,n}} \mu^{1/2}_{I_j}$$

In particular, $x \in conv\,\{y_{I_j} : \; j \in P^+_{m,n}\}$

Proof. By definition

$$\tilde{A} = M^{1/2}A$$

and thus

$$\tilde{A}_{I_j} = M^{1/2}_{I_j} A_{I_j} \tag{3.12}$$

where $M^{1/2}_{I_j}$ is the $n \times n$ diagonal matrix with elements $\mu_k > 0$, $k \in I_j$, $j \in P_{m,n}$. Using (3.12) we compute:

$$\begin{aligned} \tilde{\Delta}_{I_j} &= det M^{1/2} \cdot det A_{I_j} \\ &= (\prod_{j \in P_{m,n}} \mu^{1/2}_{I_j})\Delta_{I_j} \\ &= \pi_{I_j}\Delta_{I_j}. \end{aligned}$$

Substituting the above in (3.11) and using λ_{I_j} as given in Theorem 3.1 we obtain

$$\begin{aligned}\tilde{\lambda}_{I_j} &= \frac{\pi_{I_j}^2 \Delta_{I_j}^2}{\sum \pi_{I_j}^2 \Delta_{I_j}^2} \\ &= \frac{\pi_{I_j}^2 \lambda_{I_j}}{\sum \pi_{I_j}^2 \lambda_{I_j}} \qquad (3.13)\end{aligned}$$

Now using (3.7) we have for $k = 1, \ldots, n$

$$\tilde{y}_{I_j}^k = \frac{\tilde{\Delta}_{I_j}^k(b)}{\tilde{\Delta}_{I_j}} = \frac{\tilde{\Delta}_{I_j}^k(b)}{\pi_{I_j} \Delta_{I_j}} \text{ by (3.13)} \qquad (3.14)$$

where

$$\begin{aligned}\tilde{\Delta}_{I_j}^k(b) &= det \tilde{A}_{I_j}(k \rightarrow \tilde{b}_{I_j} \\ &= det M_{I_j}^{1/2} A_{I_j}(k \rightarrow M_{I_j}^{1/2} b_{I_j}) \\ &= det M_{I_j}^{1/2} \cdot det A_{I_j}(k \rightarrow b_{I_j}) \qquad (3.15) \\ &= \pi_{I_j} \cdot \Delta_{I_j}^k(b) \qquad (3.16)\end{aligned}$$

The equality (3.15) following from the fact that for any diagonal matrix $D \in \mathbb{R}^{n \times n}$ and any matrix $F \in \mathbb{R}^{n \times n}$ and vector $c \in \mathbb{R}^n$ the following identity holds:

$$DF(k \rightarrow Dc) = D \cdot F(k \rightarrow c).$$

Substituting (3.16) in (3.14) it follows that $\tilde{y}_{I_j}^k = y_{I_j}^k$, for all $k = 1, \ldots, n$ and the proof is complete. □

Remark 3.1 The important fact to observe is that the y_{I_j}'s depend only on the data A, b and *not* on the matrix M, hence the weighted least squares solution is inside the compact polyhedral convex set conv $\{y_{I_j} : j \in P_{m,n}^+\}$ which is independent of the weighting matrix M.

References

[1] A. Ben-Israel and T. N. E. Greville, *Generalized Inverses: Theory and Applications*. Wiley Interscience, Krieger 1980.

[2] F. R. Gantmacher, *The theory of Matrices*, Chelsea Publishing Company; New-York, N.Y. 1959.

[3] P. E. Gill, W. Murray , M. A. Saunders, J.A. Tomlin, M. H. Wright, " On projected Newton barrier methods for linear programming and an equivalence to Karmaka'’s projective method". *Mathematical Programming*,**36**,(1986) 183-209.

[4] N. Karmakar, "A new polynomial-time algorithm for linear programming", *Combinatorica*, 4 (1984b) 373-395.

[5] Yu. V. Linnik, *Method of Least Squares and Principles of the Theory of Observations* . Pergamon Press, Oxford, 1961.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER CCS 606	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Geometric Property of the Lease Squares Solution of Linear Equations		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER CCS 605
7. AUTHOR(s)		8. CONTRACT OR GRANT NUMBER(s) NSF SES 8520806, DAK-15-87-C-0110
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Cybernetic Studies, UT Austin Austin, Texas 78712		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS USAREC Ft. Sheridan, IL 60037		12. REPORT DATE June 1988
		13. NUMBER OF PAGES 9
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Key Words: Linear equations, Least squares problems, Convexity.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We derive an explicit determinantal formula for the least squares (LS) solution of an overdetermined system of linear equations. From this formula it follows that the LS solution lies in the convex hull S of points, each of which is a solution of a square subsystem of the whole system. The results are extended to weighted LS solution; it is shown that the convex hull S in which the solution must lie is independent of the weighting matrix.