

A NOVEL INTERPRETATION OF LEAST SQUARES SOLUTION

JACK-KANG CHAN

Norden Systems
75 Maxess Road
Melville, New York 11747

(Received March 5, 1991 and in revised form August 21, 1991)

Abstract. We show that the well-known least squares (LS) solution of an overdetermined system of linear equations is a convex combination of all the non-trivial solutions weighed by the squares of the corresponding denominator determinants of the Cramer's rule. This **Least Squares Decomposition (LSD)** gives an alternate statistical interpretation of least squares, as well as another geometric meaning. Furthermore, when the singular values of the matrix of the overdetermined system are not small, the LSD may be able to provide flexible solutions. As an illustration, we apply the LSD to interpret the LS-solution in the problem of source localization.

Key Words and Phrases: least squares decomposition, least squares solution, overdetermined system, source localization.

1980 Mathematics Subject Classification: 15A09, 15A15, 65F20, 56F40

1. **INTRODUCTION.** Given an overdetermined system of linear equations

$$Ax = b \tag{1}$$

where $A = [a_{ij}]$ is an $m \times n$ matrix with $m \geq n$ and $\text{rank}(A) = n$, and $b = [b_j]$ is an m -column vector, and $x = [x_j]$ is the unknown n -column vector. For simplicity, we consider only real numbers. There are at most $\binom{m}{n}$ non-trivial (NT) solutions given by the familiar Cramer's rule:

$$x_j[i_1 \dots i_n] = \frac{\det[A: i_1 \dots i_n; b: j]}{\det[A: i_1 \dots i_n]} \quad j = 1, \dots, n \tag{2}$$
$$1 \leq i_1 < \dots < i_n \leq m$$

where $\det[A: i_1 \dots i_n]$, assumed non-zero, is the $n \times n$ minor formed from A by taking the rows i_1, \dots, i_n , and $\det[A: i_1 \dots i_n; b: j]$ is the previous determinant with its j th column replaced by the corresponding b_{i_1}, \dots, b_{i_n} from the vector b .

The least squares (LS) solution is [3]:

$$x_{LS} = (A'A)^{-1}A'b \tag{3}$$

where $(A'A)^{-1}A'$ is the *generalized inverse* of A . Using our notations, the LS-solution can be rewritten as

$$x_{j[LS]} = \frac{\det[A'A; A'b: j]}{\det[A'A]} \quad j = 1, \dots, n \quad (3')$$

Furthermore, it is well-known that the generalized inverse and the LS-solution can be expressed in terms of the *singular-value decomposition* (SVD) [3] which suggests a way of obtaining more accurate solutions by taking the sum over those large singular values only.

2. LEAST SQUARES DECOMPOSITION.

THEOREM.

$$\det[A'A; A'b: j] = \sum_{1 \leq i_1 < \dots < i_n \leq m} \det[A: i_1 \dots i_n]' \cdot \det[A: i_1 \dots i_n; b: j] \quad (4)$$

$$\det[A'A] = \sum_{1 \leq i_1 < \dots < i_n \leq m} |\det[A: i_1 \dots i_n]|^2 \geq 0 \quad (5)$$

For $j = 1, \dots, n$,

$$x_{j[LS]} = \sum_{1 \leq i_1 < \dots < i_n \leq m} \frac{|\det[A: i_1 \dots i_n]|^2 \cdot x_{j[i_1 \dots i_n]}}{\sum_{1 \leq k_1 < \dots < k_n \leq m} |\det[A: k_1 \dots k_n]|^2} \quad (6)$$

PROOF. We have assumed that $A'A$ is non-singular (since $\text{rank}(A) = n$). Note that Eq. (5) is a special case of Eq. (4), and Eq. (6) is the consequences of Eqs. (2), (3), (4) and (5). The case when $m = n$ is obvious, because in this case, there is only one term in the summations. The case when $n = 2$ can be verified easily by direct evaluation. The proof for the general case is notationally lengthy. In order to illustrate the spirit, we will prove Eq. (4) for $m = 4$ and $n = 3$ in the following. Note that if A is complex, we simply replace all the transposes by the conjugate transposes.

Suppose $A = [a_{ij}]_{4 \times 3}$ and $b = [b_j]_{4 \times 1}$. Then

$$A'A = \begin{bmatrix} \sum_1^4 a_{i1}^2 & \sum_1^4 a_{i1}a_{i2} & \sum_1^4 a_{i1}a_{i3} \\ \sum_1^4 a_{i2}a_{i1} & \sum_1^4 a_{i2}^2 & \sum_1^4 a_{i2}a_{i3} \\ \sum_1^4 a_{i3}a_{i1} & \sum_1^4 a_{i3}a_{i2} & \sum_1^4 a_{i3}^2 \end{bmatrix}$$

and

$$A'b = \begin{bmatrix} \sum_1^4 a_{i1}b_i \\ \sum_1^4 a_{i2}b_i \\ \sum_1^4 a_{i3}b_i \end{bmatrix}$$

Let $D_{123} = \det[A: 123]' \cdot \det[A: 123; b: 1]$. Therefore,

$$\begin{aligned} D_{123} &= \begin{vmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{vmatrix} \begin{vmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{vmatrix} \\ &= \begin{vmatrix} \sum_{1,2,3} a_{i1}b_i & \sum_{1,2,3} a_{i1}a_{i2} & \sum_{1,2,3} a_{i1}a_{i3} \\ \sum_{1,2,3} a_{i2}b_i & \sum_{1,2,3} a_{i2}^2 & \sum_{1,2,3} a_{i2}a_{i3} \\ \sum_{1,2,3} a_{i3}b_i & \sum_{1,2,3} a_{i3}a_{i2} & \sum_{1,2,3} a_{i3}^2 \end{vmatrix} \end{aligned}$$

$$\begin{aligned}
 & - \begin{vmatrix} \sum_1^4 a_{i1}b_i - a_{41}b_4 & \sum_{1,2,3} a_{i1}a_{i2} & \sum_{1,2,3} a_{i1}a_{i3} \\ \sum_1^4 a_{i2}b_i - a_{42}b_4 & \sum_{1,2,3} a_{i2}^2 & \sum_{1,2,3} a_{i2}a_{i3} \\ \sum_1^4 a_{i3}b_i - a_{43}b_4 & \sum_{1,2,3} a_{i3}a_{i2} & \sum_{1,2,3} a_{i3}^2 \end{vmatrix} \\
 & - \begin{vmatrix} \sum_1^4 a_{i1}b_i & \sum_{1,2,3} a_{i1}a_{i2} & \sum_{1,2,3} a_{i1}a_{i3} \\ \sum_1^4 a_{i2}b_i & \sum_{1,2,3} a_{i2}^2 & \sum_{1,2,3} a_{i2}a_{i3} \\ \sum_1^4 a_{i3}b_i & \sum_{1,2,3} a_{i3}a_{i2} & \sum_{1,2,3} a_{i3}^2 \end{vmatrix} - \begin{vmatrix} a_{41}b_4 & \sum_{1,2,3} a_{i1}a_{i2} & \sum_{1,2,3} a_{i1}a_{i3} \\ a_{41}b_4 & \sum_{1,2,3} a_{i2}^2 & \sum_{1,2,3} a_{i2}a_{i3} \\ a_{41}b_4 & \sum_{1,2,3} a_{i3}a_{i2} & \sum_{1,2,3} a_{i3}^2 \end{vmatrix} \\
 & - \begin{vmatrix} \sum_1^4 a_{i1}b_i & \sum_1^4 a_{i1}a_{i2} & \sum_1^4 a_{i1}a_{i3} \\ \sum_1^4 a_{i2}b_i & \sum_1^4 a_{i2}^2 & \sum_1^4 a_{i2}a_{i3} \\ \sum_1^4 a_{i3}b_i & \sum_1^4 a_{i3}a_{i2} & \sum_1^4 a_{i3}^2 \end{vmatrix} - \begin{vmatrix} \sum_1^4 a_{i1}b_i & \sum_1^4 a_{i1}a_{i2} & a_{41}a_{43} \\ \sum_1^4 a_{i2}b_i & \sum_1^4 a_{i2}^2 & a_{42}a_{43} \\ \sum_1^4 a_{i3}b_i & \sum_1^4 a_{i3}a_{i2} & a_{43}^2 \end{vmatrix} \\
 & - \begin{vmatrix} \sum_1^4 a_{i1}b_i & a_{41}a_{42} & \sum_{1,2,3} a_{i1}a_{i3} \\ \sum_1^4 a_{i2}b_i & a_{42}^2 & \sum_{1,2,3} a_{i2}a_{i3} \\ \sum_1^4 a_{i3}b_i & a_{43}a_{42} & \sum_{1,2,3} a_{i3}^2 \end{vmatrix} - \begin{vmatrix} a_{41}b_4 & \sum_{1,2,3} a_{i1}a_{i2} & \sum_{1,2,3} a_{i1}a_{i3} \\ a_{42}b_4 & \sum_{1,2,3} a_{i2}^2 & \sum_{1,2,3} a_{i2}a_{i3} \\ a_{43}b_4 & \sum_{1,2,3} a_{i3}a_{i2} & \sum_{1,2,3} a_{i3}^2 \end{vmatrix} \\
 & = \det[A'A; A'b: 1] - \begin{vmatrix} \sum_1^4 a_{i1}b_i & \sum_1^4 a_{i1}a_{i2} & a_{41}a_{43} \\ \sum_1^4 a_{i2}b_i & \sum_1^4 a_{i2}^2 & a_{42}a_{43} \\ \sum_1^4 a_{i3}b_i & \sum_1^4 a_{i3}a_{i2} & a_{43}^2 \end{vmatrix} \\
 & - a_{42} \begin{vmatrix} \sum_1^4 a_{i1}b_i & a_{41} & \sum_{1,2,3} a_{i1}a_{i3} \\ \sum_1^4 a_{i2}b_i & a_{42} & \sum_{1,2,3} a_{i2}a_{i3} \\ \sum_1^4 a_{i3}b_i & a_{43} & \sum_{1,2,3} a_{i3}^2 \end{vmatrix} - b_4 \begin{vmatrix} a_{41} & \sum_{1,2,3} a_{i1}a_{i2} & \sum_{1,2,3} a_{i1}a_{i3} \\ a_{42} & \sum_{1,2,3} a_{i2}^2 & \sum_{1,2,3} a_{i2}a_{i3} \\ a_{43} & \sum_{1,2,3} a_{i3}a_{i2} & \sum_{1,2,3} a_{i3}^2 \end{vmatrix}
 \end{aligned}$$

For the third determinant, multiply Column 2 by a_{43} and add it to Column 3. For the fourth determinant, multiply Column 1 by a_{42} and add it to Column 2, then multiply Column 1 by a_{43} and add it to Column 3. We then have,

$$D_{123} = \det[A'A; A'b: 1] - \begin{vmatrix} \sum_1^4 a_{i1}b_i & \sum_1^4 a_{i1}a_{i2} & a_{41}a_{43} \\ \sum_1^4 a_{i2}b_i & \sum_1^4 a_{i2}^2 & a_{42}a_{43} \\ \sum_1^4 a_{i3}b_i & \sum_1^4 a_{i3}a_{i2} & a_{43}^2 \end{vmatrix}$$

$$- \begin{vmatrix} \sum_1^4 a_{1j}b_i & a_{41}a_{42} & \sum_1^4 a_{1j}a_{13} \\ \sum_1^4 a_{2j}b_i & a_{42}^2 & \sum_1^4 a_{12}a_{13} \\ \sum_1^4 a_{3j}b_i & a_{43}a_{42} & \sum_1^4 a_{13}^2 \end{vmatrix} - \begin{vmatrix} a_{41}b_4 & \sum_1^4 a_{11}a_{12} & \sum_1^4 a_{11}a_{13} \\ a_{42}b_4 & \sum_1^4 a_{12}^2 & \sum_1^4 a_{12}a_{13} \\ a_{43}b_4 & \sum_1^4 a_{13}a_{12} & \sum_1^4 a_{13}^2 \end{vmatrix}$$

Similarly, we obtain D_{124} , D_{134} , and D_{234} . Adding up all these expressions, we note that the sum of the second determinants in the above expression gives $-\det[A'A; A'b; 1]$, and similarly for the third and the fourth determinants. Therefore,

$$\begin{aligned} D_{123} + D_{124} + D_{134} + D_{234} &= 4 \cdot \det[A'A; A'b; 1] - \det[A'A; A'b; 1] \\ &\quad - \det[A'A; A'b; 1] - \det[A'A; A'b; 1] \\ &= \det[A'A; A'b; 1] \end{aligned}$$

which is the required LSD (Eq. (4)).

If all the singular values of A are not small, we cannot reduce the summation in the SVD. However, the Least Square Decomposition (LSD) (Eq. (6)) suggests that we may still get a better answer by summing those NT-solutions whose Cramer denominator determinants are large in magnitude. We will verify the LSD formulas and the above idea via an example in the next Section. In fact, the LSD has the same form as a well-known result in statistics: If $\hat{\Theta}_1, \dots, \hat{\Theta}_r$ are unbiased estimators of Θ with variances $\sigma_1^2, \dots, \sigma_r^2$ respectively, then the linear unbiased minimum variance estimator of Θ is well known to be [2]:

$$\hat{\Theta} = \frac{\sum_{i=1}^r \frac{\sigma_i^2}{\sum_{j=1}^r \sigma_j^2} \hat{\Theta}_i}{\sum_{i=1}^r \frac{\sigma_i^2}{\sum_{j=1}^r \sigma_j^2}} \quad (7)$$

Eq. (7) is also the result of minimizing

$$\sum_{i=1}^r \sum_{j=1}^{N_j} (\Theta_{ij} - \Theta)^2 / \sigma_i^2 \quad (8a)$$

where

$$\hat{\Theta}_i = \left(\sum_{j=1}^{N_j} \Theta_{ij} \right) / N_j \quad (8b)$$

by the method of least squares [2]. Thus, if we interpret the squares of the Cramer denominator determinants as the σ_i^{-2} 's, the NT-solutions are the estimates of the "true" solutions, then Eq. (6) and Eq. (7) are identical. Therefore, the LSD has a second meaning of "least squares" (Eq. (8))!

Eq. (6) gives a simple geometric interpretation, namely, the LS-solution is the "center of mass" among the NT-solutions. The LS-solution is therefore lying near those NT-solutions whose denominator determinants are large in magnitude (or, small variances).

3. AN EXAMPLE. Consider

$$\begin{bmatrix} -1 & 2 & -3 \\ 4 & 0 & 5 \\ 3 & -2 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 7 \\ 8 \\ 9 \\ 10 \end{bmatrix}$$

As shown in Table 1, there are 4 NT-solutions (see Columns 1 to 4 in Table 1(a)). The LS-solution is computed from $(A'A)x = (A'b)$,

$$\begin{bmatrix} 26 & -8 & 23 \\ -8 & 9 & -7 \\ 23 & -7 & 35 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 52 \\ 6 \\ 9 \end{bmatrix}$$

as shown in Column 5. We have, from Table 1(a),

$$(44)^2(208/44) + (1)^2(57/1) + (-5)^2(-55/-5) + (23)^2(161/23) = 13187$$

$$(44)^2(114/44) + (1)^2(-34/1) + (-5)^2(-60/-5) + (23)^2(138/23) = 8456$$

$$(44)^2(-96/44) + (1)^2(-44/1) + (-5)^2(-10/-5) + (23)^2(-92/23) = -6334$$

and

$$(44)^2 + (1)^2 + (-5)^2 + (23)^2 = 2491.$$

Hence all the LSD formulas have been verified. Furthermore, the ratio of the magnitude squared (or, just the magnitude) of the determinants in descending order provides information about the significant number of terms in the LSD sum. In our case, we have

$$(44)^2 : (23)^2 : (-5)^2 : (1)^2 = 1 : 0.27324 : 0.01291 : 0.00052$$

Thus, we may define a "condition number" as the ratio of the largest determinant squared to the smallest. The singular values can be computed [1]: $s_1 = 7.501111$, $s_2 = 2.926371$, $s_3 = 2.273693$. Note that $s_1^2 s_2^2 s_3^2 = 2491$. None of these singular values are small, because the rank of the matrix is 3. In this case, the SVD gives no further improvement, but the LSD is still flexible as illustrated in Table 1(b), where the LSD is summing up the NT-solutions with large denominator determinants in descending order of magnitude. Thus, we may first find the SVD of the system to see how many singular values are small to do the necessary rank reduction, then we apply the LSD for an ultimate improvement.

(a) NT-Solutions and LS-Solutions

	Cramer's rule: NT-Solutions				LS (SVD)-Solution det=2491
	Rows 1,2,3 det=44	Rows 1,2,4 det=1	Rows 1,3,4 det=-5	Rows 2,3,4 det=23	
x_1	208/44 = 4.72727	57/1 = 57	-55/-5 = 11	161/23 = 7	13187/2491 = 5.29386
x_2	114/44 = 2.59091	-34/1 = -34	-60/-5 = 12	138/23 = 6	8456/2491 = 3.39462
x_3	-96/44 = -2.18182	-44/1 = -44	-10/-5 = 2	-92/23 = -4	-6334/2491 = -2.54275

(b) LSD Solutions

	LSD Solutions			
	1 det : 44	2 dets : 44, 23	3 dets : 44, 23, -5	4 dets : 44, 23, -5, 1
x_1	208/44 = 4.72727	12855/2465 = 5.21501	13130/2490 = 5.27309	13187/2491 = 5.29386
x_2	114/44 = 2.59091	8190/2465 = 3.32251	8490/2490 = 3.40964	8456/2491 = 3.39462
x_3	-96/44 = -2.18182	-6340/2465 = -2.57201	-6290/2490 = -2.52610	-6334/2491 = -2.54275

Table 1. Example of Least Squares Decompositions.

4. AN APPLICATION TO SOURCE LOCALIZATION. In navigation and sonar we adopt sensors (at least 3) to receive signals from a source. Let us first consider a 2-dimensional problem with a constant sound speed for simplicity. Suppose we can estimate the time delays between every two sensors. The locus of the source falls on a hyperbola with these two sensors as its foci. Thus every two sensors determine a hyperbola. Suppose there are n sensors ($n \geq 3$), then the intersection of all the hyperbolas will give the source location. This is the well known technique of **hyperbolic fixing**. However, due to noisy time delay measurements, these hyperbolas do not intersect at a unique point. Usually, the source is far away from the sensors and the hyperbolas may be approximated by their asymptotes. The problem is now reduced to a system of pairs of 2×2 linear equations. A least-squares solution gives the source location. With the LSD theorem, we interpret the LS-solution as the weighted sum of all possible source locations according to their denominator determinants. Each denominator is proportional to the tangent of the angle between the two hyperbolas. Thus, if the hyperbolas intersect at almost a right angle, the source location is more accurate than those intersections at small angles. This angle interpretation is simple and intuitive, and it is justified by the LSD theorem. Furthermore, we can just select those solutions with large denominators only. Note that the SVD method has no improvement because the rank is always 2. The optimal location is the "center of mass" of the possible locations.

Moreover, instead of using hyperbolas, Schmidt [4] has shown that the source location is the focus of a conic passing through the 3 sensors, hence the source is on the focal line. With more than 3 sensors, we have more than one focal lines. The intersections of these focal lines give the source location(s). Thus, we actually solving linear equations, not just an approximation using asymptotes as in the hyperbolic fixing technique. We can use the LSD to interpret the angles between the focal lines as a measure of the accuracy of the solutions. Formulations of the localization in 3-dimensions using Schmidt's method and other equivalent methods to get LS-solutions have been done [5], [6]. We can interpret all these LS-solutions using the LSD similar to the 2-dimensional case.

ACKNOWLEDGEMENT

The author would like to thank Professor George Bachman of Polytechnic University for his proof reading of the manuscript and valuable suggestions, and to the Referee for his/her valuable comments.

REFERENCES

- [1] J.J. Dongarra, C.B. Moler, J.R. Bunch, and G.W. Stewart, *LINPACK User's Guide*, Chap. 11, SIAM, Philadelphia, PA, 1979.
- [2] A. Hald, *Statistical Theory with Engineering Applications*, pp. 243-245, Wiley, New York, NY, 1952.
- [3] C.L. Lawson and R.J. Hanson, *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, NJ, 1974.
- [4] R.O. Schmidt, *A new approach to geometry of range difference location*, IEEE Trans. Aerosp. Electron., Vol. AES-8, Nov. 1972, 821-835.
- [5] H.C. Schau, A.Z. Robinson, *Passive source localization employing intersecting spherical surfaces from time-of-arrival differences*, IEEE Trans. Acoust., Speech, Signal Process., Vol. ASSP-35, Aug. 1987, pp. 1223-1225.
- [6] J.O. Smith, J.S. Abel, *Closed-form least-squares source location estimation from range-difference measurements*, IEEE Trans. Acoust., Speech, Signal Process., Vol. ASSP-35, Dec. 1987, pp. 1661-1669.