Technical Report No. 91-5
Technical Memorandum KL001
October 1990

**Least Squares Estimation Principle
and its
Geometrical Interpretation**

by
Khosrow Lashkari
Acoustics Division

I. LEAST SQUARES ESTIMATION

Least squares estimation is a widely used technique in analyzing and processing engineering and scientific data.

The least squares estimation problem can be stated as follows: given the observed/measured data sequence $x_i, i = 1, 2 \ldots n$ the objective is to estimate a related sequence $y_j, j = 1, \ldots m$ such that the estimated sequence $\hat{y}_j, j = 1, \ldots m$ is as close as possible to the original sequence in the Euclidean norm.

Let

$$\hat{y}_j = f_j(\mathbf{x}) \qquad i = 1, \ldots m \tag{1.1}$$

be the estimated sequence where $\mathbf{x} = [x_1 \ldots x_n]'$ is the given data vector and $f_j(\mathbf{x})$ is in general a non linear multi variable function. The error vector in this estimation is given by

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} \tag{1.2}$$

with

$$\mathbf{y} = [y_1 \ldots y_m]' \quad \text{and} \quad \hat{\mathbf{y}} = [\hat{y}_1 \ldots \hat{y}_m]'$$

The objective of least squares estimation is to determine the functions f_j such that the least squares error criterion

$$E = |\mathbf{e}|^2 = \sum_{j=1}^{m} (y_j - \hat{y}_j)^2 \tag{1.3}$$

is minimum. In linear least squares functions f_i are linear and $\hat{y}_j$ is a linear combination of the measurements $x_i, \quad i = 1, \ldots n$.

Least squares technique was originally developed independently by the German mathematician Gauss and French mathematician Legendre.

The concept of least squares analysis is probably best explained by Gauss himself in Theoria Motus Corporum Celestium [1] (Theory of the Motion of Heavenly Bodies):

> If the astronomical observations and other quantities on which the computation of orbits is based were absolutely correct, the elements also, whether deduced from three or four observations would be strictly accurate (so far indeed as the motion is supposed to take place exactly according to the laws of Kepler) and, therefore, if other observations were used, they might be confirmed but not corrected. But since all our measurements and observations are nothing more than approximations to the truth, the same must be true of all calculations resting upon them, and the highest aim of all computations concerning concrete phenomena must be to approximate, as nearly as practicable to the truth. But this can be accomplished in no other way than by a suitable combination of more observations than the number absolutely requisite for the determination of the unknown quantities.

Gauss goes on to say that

> ... the most probable value of the unknown quantities will be that in which the sum of the squares of the differences between the actually observed and the computed values multiplied by numbers that measure the degree of precision is a minimum.

NOTATIONS

For the sake of clarity, the following conventions are adopted in this document:

1. Scalars are represented by lowercase letters such as a, x, y and so on. Indexed scalars are denoted as x_1, x_2, and so on.

2. Vectors are denoted by boldface lowercase letters such as $\mathbf{a}, \mathbf{b}, \mathbf{x}$ and so on. Indexed vectors are denoted as $\mathbf{x}_1, \mathbf{x}_2$ and so on.

3. Matrices are denoted by boldface uppercase letters such as $\mathbf{A}, \mathbf{B}$ and so on. Indexed matrices are denoted as $\mathbf{M}_1, \mathbf{M}_2$, and so on.

ORGANIZATION OF THE DOCUMENT

Section 2.0 introduces the linear least squares problem. Section 2.1 uses basic principles to derive the least squares solution for an overdetermined system of linear equations. Section 2.2 discusses orthogonal projections and provides a new interpretation for the least squares solutions in terms of these projections. Section 2.3 introduces singular value decomposition and gives an alternate interpretation for the least squares technique in terms of subspaces spanned by orthonormal basis functions.

Finally, section 2.4 introduces the general case of weighted least squares and derives the solution for this general case.

2.0 INTRODUCTION

In this section we will develop the theory of linear least squares estimation. Here the functions f_j are assumed to be linear and of the form

$$\hat{y}_j = f_j(x) = \sum_{k=1}^{n} f_{jk} x_k = \mathbf{f}'_j \mathbf{x} \qquad j = 1, \ldots m \tag{2.1}$$

Where

$$\mathbf{f}'_j = [f_{jl} \ldots f_{jn}] \tag{2.2a}$$

The estimated sequence $\hat{y}$ can be expressed as

$$\hat{\mathbf{y}} = \mathbf{F}\mathbf{x} \tag{2.3}$$

where $\mathbf{F}$ is an $m \times n$ matrix of the form

$$\mathbf{F} = \begin{bmatrix} \mathbf{f}'_1 \\ \mathbf{f}'_2 \\ \vdots \\ \mathbf{f}'_m \end{bmatrix} \tag{2.4}$$

and $\mathbf{x}$ and $\hat{\mathbf{y}}$ are defined as before.

2.1 Least Squares Solution of an Overdetermined System of Linear Equations

Let

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{2.1.1}$$

be an overdetermined set of linear equations where $\mathbf{A}, \mathbf{x}$ and $\mathbf{b}$ have the following dimensions.

$$
\begin{aligned}
\mathbf{A} &= m \times n = \text{coefficient matrix} \quad m > n \\
\mathbf{x} &= n \times 1 = \text{vector of the unknowns} \\
\mathbf{b} &= m \times 1 = \text{RHS vector of known quantities} \\
m &= \text{number of rows} \\
n &= \text{number of columns}
\end{aligned} \tag{2.1.2}
$$

The solution to this problem can be found by explicitly writing out the error and minimizing it by setting the partial derivatives relative to components of $\mathbf{x}$ equal to zero. The error $\mathbf{e}$ is defined as the difference between the RHS and the LHS. Let $\mathbf{x}*$ be the least squares solution vector, then the error vector is given by:

$$\mathbf{e} = \mathbf{b} - \mathbf{A}\mathbf{x}* \quad (m \times 1) \tag{2.1.3}$$

The least squares cost function is given by:

$$E = \mathbf{e}'\mathbf{e} = |\mathbf{e}|^2 \tag{2.1.4}$$

where:

$$
\begin{aligned}
\mathbf{e}' &= \text{transpose of } \mathbf{e}; \text{ and} \\
|\mathbf{e}| &= \text{norm of } \mathbf{e}
\end{aligned}
$$

Substituting for **e**, we get:

$$E = \mathbf{b}'\mathbf{b} - 2\mathbf{b}'\mathbf{A}\mathbf{x} + \mathbf{x}'\mathbf{A}'\mathbf{A}\mathbf{x} \tag{2.1.5}$$

Differentiating with respect to **x**, we get the gradient vector

$$\nabla E = -2\mathbf{A}'\mathbf{b} + 2\mathbf{A}'\mathbf{A}\mathbf{x} \tag{2.1.6}$$

Setting the derivatives to zero we get:

$$\mathbf{A}'\mathbf{A}\mathbf{x} = \mathbf{A}'\mathbf{b} \tag{2.1.7}$$

solving for **x** we get:

$$
\begin{array}{cccc}
\mathbf{x}* \;=\; & (\mathbf{A}'\mathbf{A})^{-1} & \mathbf{A}' & \mathbf{b} \\
\uparrow & \uparrow & \uparrow & \uparrow \\
| & | & | & | \\
| & | & | & | \\
n \times 1 & n \times n & n \times m & m \times 1
\end{array}
\tag{2.1.8}
$$

Note that $\mathbf{A}'\mathbf{A}$ is an $n \times n$ square matrix and the above solution exists only if this matrix is nonsingular. In case **A** is square (say $n \times n$) and invertible, then

$$(\mathbf{A}'\mathbf{A})^{-1} = \mathbf{A}^{-1}(\mathbf{A}')^{-1} \tag{2.1.9}$$

and we get:

$$\mathbf{x}* = \mathbf{A}^{-1}\mathbf{b} \tag{2.1.10}$$

which is the standard solution.

Matrix $(\mathbf{A}'\mathbf{A})$ is nonsingular if and only if **A** is full rank. If $r = \text{rank}(\mathbf{A}) < n$, then equation 2.1.7 is an underdetermined set of linear equations and the least squares solution $\mathbf{x}*$ is not unique. This is because $\mathbf{q} = \mathbf{A}'\mathbf{b}$ is a vector in the row space of **A** and $\mathbf{A}'\mathbf{A}\mathbf{x}$ is also a vector in the row space of **A** and since **A** is not full rank equation $\mathbf{A}'\mathbf{y} = \mathbf{q}$ does not have a unique solution for **y** which implies that equation $\mathbf{A}\mathbf{x} = \mathbf{y}$ does not have a unique solution for **x**.

Note also that for nonsquare **A**, the relationship $\mathbf{A}\mathbf{x} = \mathbf{b}$ is not an exact equality, rather it is an approximate relationship, thus it can be more appropriately written as:

$$\mathbf{A}\mathbf{x} \approx \mathbf{b} \tag{2.1.11}$$

To summarize:

1) if **A** is not full rank, then least squares solution is not unique

2) if **A** is full rank, that is rank $(\mathbf{A}) = n$, then least squares solution $\mathbf{x}*$ is unique and is given by (2.1.8).

2.2 Orthogonal (Minimum Distance) Projections and Least Squares Solutions

The least squares solution to an overdetermined system of linear equations was derived using the first principles. This solution can also be derived using the theory of orthogonal projections. In the following we will look at the geometric interpretations of the least squares solution and present an alternative method of derivation using this theory.

To introduce the geometric approach we look at the $m \times n$ matrix $\mathbf{A}$ as an operator that maps the n-dimensional vector $\mathbf{x}$ in $\Re^n$ to an m-dimensional vector $\mathbf{Ax}$ in a subspace of $\Re^m$. The particular subspace where $\mathbf{x}$ is mapped to is the subspace spanned by the columns of $\mathbf{A}$. If the rank of $\mathbf{A}$ is $r(r \leq n)$, then the subspace is r-dimensional. To see what $\mathbf{Ax}$ means, let $x_1, x_2, \ldots x_n$ be the components of vector $\mathbf{x}$. Also let $\mathbf{a}_1, \mathbf{a}_2, \ldots \mathbf{a_n}$ denote the n columns of matrix $\mathbf{A}$, then $\mathbf{Ax}$ can be written as:

$$\mathbf{Ax} = \sum_{i=1}^{n} x_i \mathbf{a}_i \tag{2.2.1}$$

Thus to form $\mathbf{Ax}$ one takes the oblique set of vectors $\{\mathbf{a}_1, \mathbf{a}_2, \ldots \mathbf{a}_n\}$ and forms $\mathbf{Ax}$ by the parallelogram rule, i.e. by weighting the vectors $\mathbf{a}_i$ by x_i and adding up these vectors. Considering that:

$$\mathbf{x} = \sum_{i=1}^{n} x_i \mathbf{c_i}$$

where $\mathbf{c}'_i s$ are the unit vectors of the standard Cartesian coordinates, we can see that one merely replaces $\mathbf{c}'_i s$ with $\mathbf{a}'_i s$ to get $\mathbf{Ax}$.

Geometric Interpretation

The error vector in (2.1.3) after substituting the least squares solution from (2.1.7) is given by:

$$\mathbf{e*} = \mathbf{b} - \mathbf{Ax*} = [\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}']\mathbf{b} \tag{2.2.2}$$

Also from (2.1.6) the equation for solving $\mathbf{x*}$ can be rewritten as:

$$\nabla E = \mathbf{A}'(\mathbf{b} - \mathbf{Ax*}) = \mathbf{A}'\mathbf{e*} = 0 \tag{2.2.3}$$

Equation (2.2.3) implies that the error vector $\mathbf{e*}$ corresponding to the (optimum) least squares solution is orthogonal to the column space of the coefficient matrix $\mathbf{A}$. This is referred to as the principle of orthogonality and equation (2.2.3) is also referred to as the normal equations. Using this principle we can show that the solution given by (2.2.3) is the least square solution. To see this consider any other solution $\mathbf{x}$ to (2.1.1) or more precisely to (2.1.11). Then we can show that:

$$|\mathbf{e}| = |\mathbf{b} - \mathbf{Ax}| \geq |\mathbf{b} - \mathbf{Ax}*| = |\mathbf{e}*| \tag{2.2.4}$$

Let

$$\mathbf{e} = \mathbf{b} - \mathbf{Ax} = \mathbf{b} - \mathbf{Ax}^* + \mathbf{A}(\mathbf{x}^* - \mathbf{x}) = \mathbf{e}^* + \mathbf{A}(\mathbf{x}^* - \mathbf{x})$$
$$|\mathbf{e}|^2 = |\mathbf{b} - \mathbf{Ax}|^2 = |\mathbf{e}^*|^2 + |\delta|^2 + 2\delta'\mathbf{e}^* \tag{2.2.5}$$

where

$$\delta = \mathbf{A}(\mathbf{x}^* - \mathbf{x})$$

Now since $\mathbf{Ax}$ and $\mathbf{Ax*}$ both lie in the column space of $\mathbf{A}$ so does δ and since $\mathbf{e*}$ is orthogonal to this space therefore it must be orthogonal to δ too, that is:

$$\delta'\mathbf{e}^* = (\mathbf{x}^* - \mathbf{x})'\mathbf{A}'\mathbf{e}^* = 0 \tag{2.2.6}$$

Using (2.2.6) in (2.2.5) we get:

$$|\mathbf{e}|^2 = |\mathbf{e}^*|^2 + |\delta|^2 \geq |\mathbf{e}^*|^2 \tag{2.2.7}$$

From which (2.2.3) follows easily.

Now we proceed to show that the least squares solution $\mathbf{x}*$ is such that $\hat{\mathbf{b}} = \mathbf{Ax}^*$ is the orthogonal projection of $\mathbf{b}$ onto the column space of $\mathbf{A}$. It is obvious that $\hat{\mathbf{b}}$ lies in the column space of $\mathbf{A}$. We have also seen that the error vector $\mathbf{e}* = \mathbf{b} - \hat{\mathbf{b}}$ is orthogonal to the column space of $\mathbf{A}$ and therefore to $\hat{\mathbf{b}}$. Hence using (2.2.6) we get the Pythogoras theorem in the hyperspace

$$|\mathbf{b}|^2 = |\hat{\mathbf{b}} + \mathbf{e}^*|^2 = |\hat{\mathbf{b}}|^2 + |\mathbf{e}^*|^2$$

Thus $\hat{\mathbf{b}}$ is orthogonal to $\mathbf{e}* = \mathbf{b} - \hat{\mathbf{b}}$ and $\mathbf{b} = \hat{\mathbf{b}} + \mathbf{e}*$, which means that $\mathbf{b}$ has been decomposed into two orthogonal components, one $(\hat{\mathbf{b}})$ in the column space of $\mathbf{A}$ and the other $(\mathbf{b} - \hat{\mathbf{b}})$ orthogonal to this space. From this we conclude that $\hat{\mathbf{b}}$ is the orthogonal projection of $\mathbf{b}$ onto the column space of $\mathbf{A}$. Substituting the least squares solution from (2.1.8) we get:

$$\hat{\mathbf{b}} = \mathbf{Ax}* = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{b} \equiv \mathbf{Pb} \qquad (2.2.8)$$

where the matrix

$$\mathbf{P} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' \qquad (2.2.9)$$

is called the orthogonal projection matrix because it projects $\mathbf{b}$ onto the column space of $\mathbf{A}$. Equation $\mathbf{Ax} \approx \mathbf{b}$ is approximate but equation $\mathbf{Ax} = \hat{\mathbf{b}}$ is exact. Figure 1 illustrates these concepts.

Figure 1:   Orthogonal decomposition of $\mathbf{b}$ into $\overset{\wedge}{\mathbf{b}}$ and $\mathbf{e}^*$

2.3 Singular Value Decomposition and Least Squares Solutions

From the above discussion we know that least squares solutions are related to orthogonal projections. This discussion however does not provide a method of solving for $\mathbf{x}$. In this section we will use singular value decomposition to first shed more light into the nature of the solution and second to give a method of finding the solution vector $\mathbf{x}$. The singular value decomposition of an $m \times n$ matrix $\mathbf{A}$ is given as:

$$\mathbf{A} = \mathbf{U} \; \Sigma \; \mathbf{V}'$$

$m \times n$

$m \times m$

$m \times n$

$n \times n \qquad (2.3.1)$

6

The column space of the $\mathbf{U}$ matrix forms an orthonormal basis for $\Re^m$ and those of $\mathbf{V}$ an orthonormal basis for $\Re^n$, that is

$$\mathbf{U}'\mathbf{U} = \mathbf{I}(m \times m) \qquad (2.3.2)$$

$$\mathbf{V}'\mathbf{V} = \mathbf{I}(n \times n) \qquad (2.3.3)$$

where $\mathbf{I}$ is the identity matrix and Σ is an $m \times n$ diagonal matrix of the form:

$$\sum = \left[\begin{array}{cccc|c}
\sigma_1 & & 0 & 0 & 0 \\
0 & \sigma_2 & & 0 & 0 \\
\cdots & \cdots & \cdots & & \\
0 & 0 & & \sigma_r & \\
\hline
0 & 0 & \cdots & \cdots & 0 \\
0 & 0 & \cdots & \cdots & 0
\end{array}\right] \qquad (2.3.4)$$

where r is the number of nonzero singular values and is equal to the rank of $\mathbf{A}$.

$$r = \mathrm{rank}(\mathbf{A}) \qquad (2.3.5)$$

Matrix Σ can also be written as:

$$\Sigma = \left[\begin{array}{c|c}
\mathbf{D} & O \\
\hline
O & O
\end{array}\right] \qquad (2.3.6)$$

Where $\mathbf{D}$ is a diagonal matrix of singular values σ_1 to σ_r.

$$\mathbf{D} = \left[\begin{array}{ccc}
\sigma_1 & & O \\
& \ddots & \\
O & & \sigma_r
\end{array}\right] \qquad (2.3.7)$$

Columns of $\mathbf{U}$ and $\mathbf{V}$ form orthonormal basis functions for $\Re^n$ and $\Re^m$ that are different from the standard Cartesian systems in these spaces (i.e. the system whose basis vectors are $(1,0,0..0), (0,1,0,0,.0), \ldots$ and $(0,0,\ldots,1)$). The Cartesian system defined by the columns of $\mathbf{U}$ for example can be obtained from the standard system by rotating this system through $(m-1)$ consecutive rotations with one each aligning one axis from the standard Cartesian system from columns of $\mathbf{V}$. The combined effect of these $(m-1)$ axial rotations is one complex rotation. The two Cartesian systems however have the same origin because under the transformation $\mathbf{y} = \mathbf{U}\mathbf{x}$ the origin (point $(0,0,..,0)$) maps into itself. We now use the SVD representation of $\mathbf{A}$ to find a solution for the overdetermined set of linear equations $\mathbf{A}\mathbf{x} \approx \mathbf{b}$. To do this we first dwell a little bit on the meaning of the transformation $\mathbf{x} \to \mathbf{A}\mathbf{x}$. We mentioned in section 2.2 equation (2.2.1) that to get $\mathbf{A}\mathbf{x}$ from $\mathbf{x}$ one has to replace the unit vectors of the Cartesian system with the columns of $\mathbf{A}$, this maps vector $\mathbf{x}$ in $\Re^n$ to vector $\mathbf{A}\mathbf{x}$ in an r-dimensional subspace of $\Re^m$ where $r = \mathrm{rank}(\mathbf{A})$. Using singular value decomposition we will show that the action of operator $\mathbf{A}$ corresponds to two rotations (one in $\Re^n$ and the other from $\Re^n$ to the r- dimensional subspace of $\Re^m$) and a dilation.

Let $\mathbf{z} = \mathbf{A}\mathbf{x}$, then

$$\mathbf{z} = \mathbf{A}\mathbf{x} = \mathbf{U}\sum\mathbf{V}'\mathbf{x} = \mathbf{U}\sum\mathbf{y} = \mathbf{W}\mathbf{y} \qquad (2.3.8)$$

where:

$$\mathbf{y} = \mathbf{V}'\mathbf{x} \qquad (2.3.9a)$$

$$\mathbf{W} = \mathbf{U}\Sigma \qquad (2.3.9b)$$

Let us first absorb the meaning of $\mathbf{V}'\mathbf{x}$. If we denote columns of $\mathbf{V}$ by $\mathbf{v}_1, \mathbf{v}_2$ to $\mathbf{v}_n$, then the $i-th$ component of $\mathbf{y}$, y_i can be written as

$$y_i = \mathbf{v}_i'\mathbf{x} \qquad (2.3.10)$$

that is y_i is the inner product of two vectors $\mathbf{v}_i$ and $\mathbf{x}$. Reflecting on this a little bit, it becomes clear that $\mathbf{y}$ is exactly the same vector $\mathbf{x}$ only in the coordinate system defined by columns of $\mathbf{V}$, y_i of equation (2.3.10) is the $i-th$ component of $\mathbf{x}$ in the new Cartesian coordinate system. From this perspective it is clear that:

$$|\mathbf{y}| = |\mathbf{V}'\mathbf{x}| = |\mathbf{x}| \qquad (2.3.11)$$

This can be shown using (2.3.10) as follows:

$$|\mathbf{y}|^2 = |\mathbf{V}'\mathbf{x}|^2 = \sum_i (\mathbf{v}_i'\mathbf{x})^2 = \sum_i \left[\sum_j v_{ij} x_j \right]^2 = \sum_i \sum_j \sum_k v_{ij} v_{ik} x_j x_k$$

$$|\mathbf{y}|^2 = \sum_j \sum_k x_j x_k \cdot \sum_i v_{ij} v_{ik} \qquad (2.3.12)$$

From orthonormality of $\mathbf{v}_i$'s we have:

$$\sum_i v_{ij} v_{ik} = \delta_{jk} = \begin{cases} 1 & j = k \\ 0 & \text{else} \end{cases} \qquad (2.3.13)$$

Thus (2.3.12) can be written as:

$$|\mathbf{y}|^2 = \sum_j \sum_k x_j x_k \delta_{jk} = \sum_j^2 = |\mathbf{x}|^2 \qquad (2.3.14)$$

Since all our references are with respect to the standard Cartesian system (SCS), therefore we must look for an interpretation of $\mathbf{y}$ in this system. One way to do this is to rotate SCS through an angle θ such that it coincides with the coordinate system defined by the $\mathbf{v}_i$'s. Since SCS must be kept unchanged what we have to do is to rotate $\mathbf{x}$ by $-\theta$ to get $\mathbf{y}$, by doing so the relationship of $\mathbf{y}$ to SCS will be exactly the same as $\mathbf{x}$ with respect to $\mathbf{V}$. Remember that θ is the angle between the Cartesian coordinate system of $\mathbf{V}$ and the SCS that is the angle that SCS must be rotated to coincide with $\mathbf{V}$. Note that vector $\mathbf{V}\mathbf{x}$ is $\mathbf{x}$ rotated $+\theta$ degrees, while $\mathbf{V}'\mathbf{x}$ is obtained by rotating $\mathbf{x}$ by $-\theta$ degrees. This is a direct consequence of orthonormality of $\mathbf{V}$ whereby $\mathbf{V}'$ equals to the left inverse of $\mathbf{V}$.

$$\mathbf{V}'\mathbf{V} = I$$
$$\mathbf{V}^{-1}\mathbf{V} = I$$

Continuing from equation (2.3.8), the next step to get from $\mathbf{y}$ to $\mathbf{A}\mathbf{x}$ is to map $\mathbf{y}$ by the operator $\mathbf{U}\Sigma$ to the r- dimensional subspace of $\Re^m$. This r-dimensional subspace is defined by the first r orthonormal columns of $\mathbf{U}$. If $\{\mathbf{u}_1 \ldots \mathbf{u}_r\}$ are the first r columns of $\mathbf{U}$, then the basis of this subspace is $\{\sigma_1 \mathbf{u}_1, \ldots \sigma_r \mathbf{u}_r\}$. If $r = n$, then operator $\mathbf{W}$ corresponds to a rotation ϕ of the SCS to that defined by the columns of $\mathbf{W}$ (try to visualize the case where m=3 and r=n=2). Transformation $\mathbf{y} \rightarrow \mathbf{W}\mathbf{y}$ in general requires changing from subspace $\Re^n$ to an r-dimensional hyper space of $\Re^m$ (change of planes for r=n=2 and m=3). Thus the mapping $\mathbf{x} \rightarrow \mathbf{A}\mathbf{x}$ corresponds to three consecutive operations: i) rotation of vector $\mathbf{x}$ in $\Re^n$ by amount $-\theta$ determined by $\mathbf{V}$ to get $\mathbf{y} = \mathbf{V}'\mathbf{x}$, ii) rotation of $\mathbf{y}$ from $\Re^n$ by ϕ to the r- dimensional subspace of $\Re^m$ defined by the first r columns of $\mathbf{U}$ to get $\mathbf{w}$ iii) dilation of vector $\mathbf{w}$ by matrix Σ. If we let $\mathbf{z} = \Sigma\mathbf{w}$, then the $i-th$ component of $\mathbf{z}$, z_i is related to the $i-th$ component of $\mathbf{w}$, w_i by the dilation relationship $z_i = \sigma_i w_i$. Figure 2 illustrates these operations geometrically.
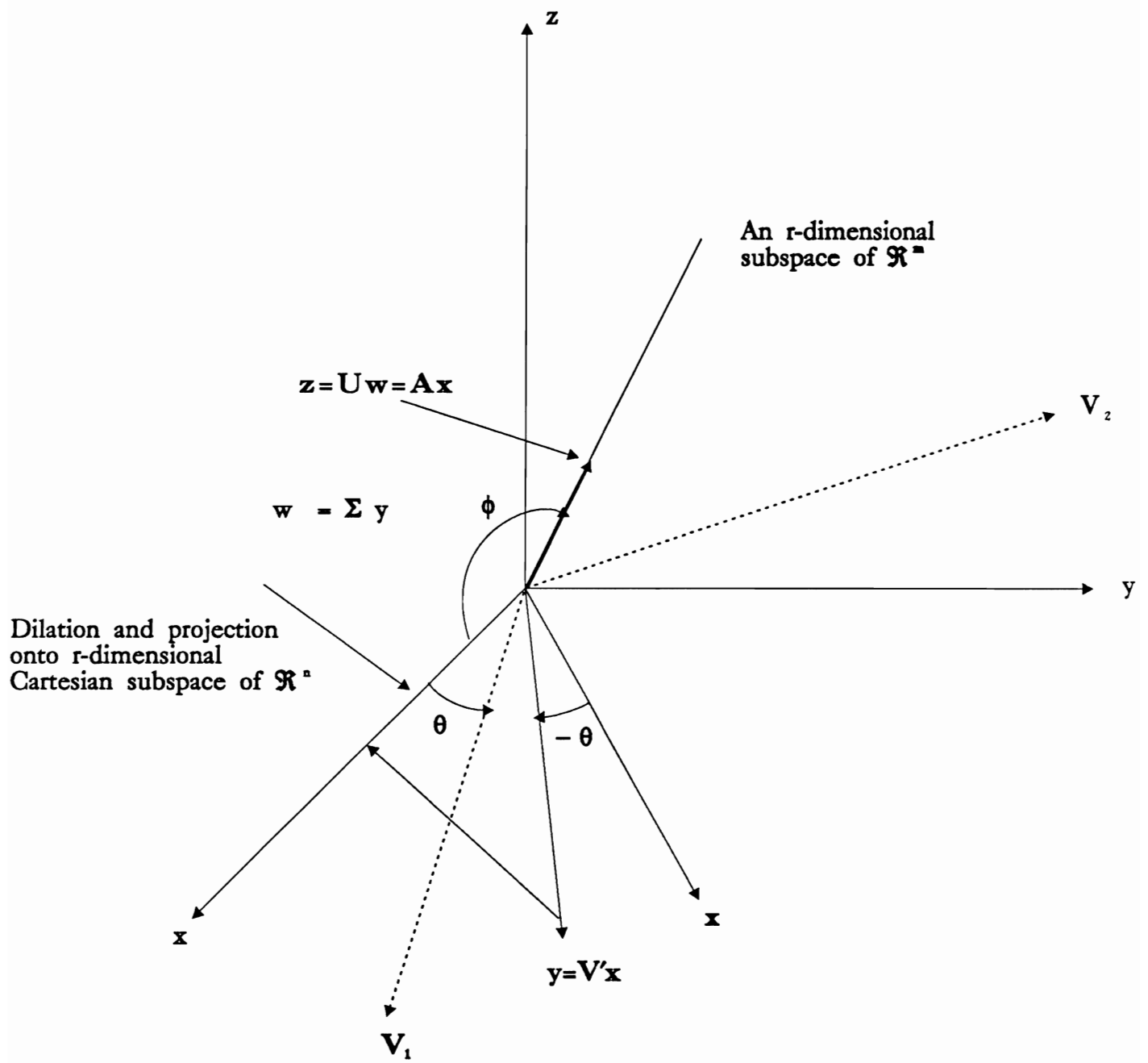
Figure 2. Geometric interpretation of singular value decomposition for r=1, n=2, and m=3.

Using singular value decomposition of $\mathbf{A}$ from equation (2.3.1), the generalized inverse of $\mathbf{A}$, $\mathbf{A}^{-1}$ can be expressed as:

$$\mathbf{A}^{-1} = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}'$$

(2.3.15)

where $\mathbf{\Sigma}^{-1}$ is an $n \times m$ matrix and is defined as:

$$\mathbf{\Sigma}^{-1} = \left[\begin{array}{c|c} \mathbf{D}^{-1} & O \\ \hline O & O \end{array} \right]$$

(2.3.16)

It can be shown by direct substitution that the solution

$$\mathbf{x}^* = \mathbf{A}^{-1}\mathbf{b} = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}'\mathbf{b}$$

(2.3.17)

will always satisfy equation (2.1.7), that is it is a least squares solution.

Figure 3 gives a geometrical interpretation for equation (2.3.17). Figures 4a and 4b provide schematic interpretations for the forward and inverse operations.

2.4 Weighted Least Squares (WLS)

In ordinary least squares, components of the error vector are weighted equally in forming the total least squares error. In weighted least squares, error components are weighted according to their significance. The cost function of WLS is:

$$E = \sum_i \sum_j w_{ij} e_i e_j = \mathbf{e}'\mathbf{W}\mathbf{e}$$

(2.4.1)

where $\mathbf{e} = \mathbf{b} - \mathbf{A}\mathbf{x} = $ the error vector

$\mathbf{W} = \{w_{ij}\}$ is the weighting matrix ≥ 0

(2.4.2)

We will show that the solution to the weighted least squares problem is obtained by solving the following system of equations:

$$\mathbf{S}\mathbf{A}\mathbf{x} = \mathbf{S}\mathbf{b}$$

(2.4.3)

where $\mathbf{S}$ is the square root of $\mathbf{W}$:

$$\mathbf{S}'\mathbf{S} = \mathbf{W}$$

(2.4.4)

This follows from (2.4.1) if we substitute for $\mathbf{e} = \mathbf{b} - \mathbf{A}\mathbf{x}$

$$E = \mathbf{e}'\mathbf{W}\mathbf{e} = (\mathbf{b} - \mathbf{A}\mathbf{x})'\mathbf{W}(\mathbf{b} - \mathbf{A}\mathbf{x}) = \mathbf{x}'(\mathbf{A}'\mathbf{W}\mathbf{A})\mathbf{x} - 2(\mathbf{b}'\mathbf{W}\mathbf{A})\mathbf{x} + \mathbf{b}'\mathbf{W}\mathbf{b}$$
$$\nabla E = 2(\mathbf{A}'\mathbf{W}\mathbf{A})\mathbf{x} - 2(\mathbf{A}'\mathbf{W}'\mathbf{b}) = 0$$
$$\mathbf{x} = (\mathbf{A}'\mathbf{W}\mathbf{A})^{-1}(\mathbf{A}'\mathbf{W}'\mathbf{b})$$

(2.4.5)

Substituting for $\mathbf{W}$, we get:

$$\mathbf{x} = (\mathbf{A}'\mathbf{S}'\mathbf{S}\mathbf{A})^{-1}(\mathbf{A}'\mathbf{S}'\mathbf{S}\mathbf{b}) = [(\mathbf{S}\mathbf{A})'(\mathbf{S}\mathbf{A})]^{-1}(\mathbf{S}\mathbf{A})'\mathbf{S}\mathbf{b}$$

(2.4.6)

which shows that (equation 8, section 2.1) $\mathbf{x}$ is the least squares solution of the following system of equations:

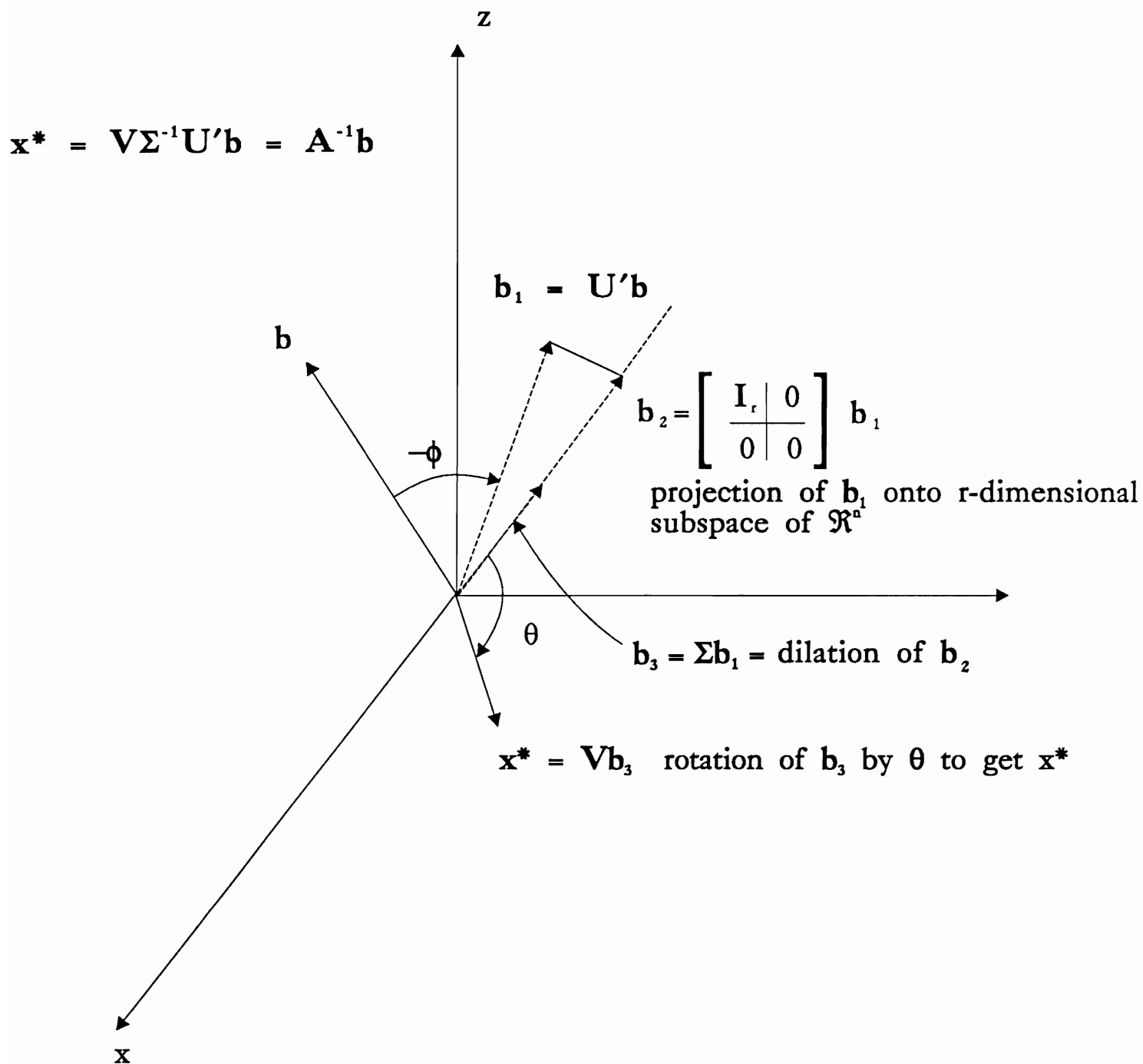$$\mathbf{S}\mathbf{A}\mathbf{x} \approx \mathbf{S}\mathbf{b}$$

10

$$x^* = V\Sigma^{-1}U'b = A^{-1}b$$

$b_1 = U'b$

$-\phi$

$$b_2 = \left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array} \right] b_1$$

projection of b_1 onto r-dimensional subspace of $\Re^n$

θ

$b_3 = \Sigma b_1 =$ dilation of b_2

$x^* = Vb_3$ rotation of b_3 by θ to get x^*

Figure 3.   Geometric interpretation of least squares solution in terms of singular value decomposition.

$$b = Ax = U \Sigma V'x = U \Sigma y = UW$$

Rotation in $\Re^n$ by $-\theta$	Dilation and projection into $\Re^r$	Rotation by $+\varphi$ from $\Re^r$ onto r-dimensional subspace of $\Re^m$

x $\Re^n$ → [Rotation in $\Re^n$ by $-\theta$] $y = V'x$ → [Dilation and projection into $\Re^r$] $W = \Sigma y$ → [Rotation by $+\varphi$ from $\Re^r$ onto r-dimensional subspace of $\Re^m$] → b
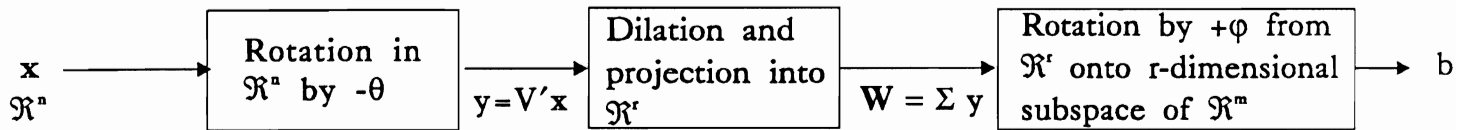
Figure 4a.   Schematic illustration of the linear operation Ax=b.

$$x = A^{-1}b = V \Sigma^{-1} U'b$$

b → [Rotation in $\Re^m$ by $-\varphi$] $U'b$ → [Dilation by Σ^{-1} and projection onto $\Re^r$] → [Rotation by $+\theta$ from $\Re^r$ onto an r-dimensional subspace of $\Re^n$] → x
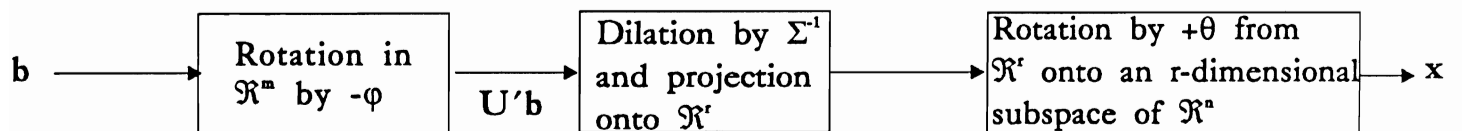
Figure 4b.   Schematic illustration of the inverse operation
x =A^{-1}b where A^{-1} is the generalized inverse.

12

REFERENCES

1. K.G. Gauss, "Theory of Motion of Heavenly Bodies," Dover, New York, 1963.