last copy

VECTOR SPACES AND PROJECTIONS

BU-547-M David M. Allen January 1975

Abstract

This is largely a condensation of the material in Appendix I, The Analysis of Variance, by Henry Scheffé. Notation has been changed to be similar to that used by C. R. Rao and S. K. Mitra, Generalized Inverse of Matrices and Its Application. I have added a few comments on projections.

\- - - - - - -

Definition 1: A vector is an ordered n-tuple

$$\underset{\sim}{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{1}$$

of real numbers. The number x_i is called the i^{th} component (or element) of $\underset{\sim}{x}$. We shall write $\underset{\sim}{x} = \underset{\sim}{x}(n)$ if we need to emphasize that $\underset{\sim}{x}$ is a vector with n components.

Definition 2: We define the sum $\underset{\sim}{x} + \underset{\sim}{y}$ of two vectors

$$\underset{\sim}{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{and} \quad \underset{\sim}{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \text{to be the vector} \quad \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix} .$$

Associative and commutative properties:

If $\underline{x}$, $\underline{y}$, and $\underline{z}$ are vectors, then

$$(\underline{x}+\underline{y}) + \underline{z} = \underline{x} + (\underline{y}+\underline{z}) = \underline{x} + \underline{y} + \underline{z}$$

and

$$\underline{x} + \underline{y} = \underline{y} + \underline{x} .$$

<u>Definition 3</u>: The product $c\underline{x}$ of a vector

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{by a scalar } c \text{ is the vector} \quad \begin{bmatrix} cx_1 \\ cx_2 \\ \vdots \\ cx_n \end{bmatrix} .$$

This operation is called <u>scalar multiplication</u>.

<u>Notation</u>: Given a set of vectors $\{\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_s\}$ let X be the augmented matrix $(\underline{x}_1 | \underline{x}_2 | \cdots | \underline{x}_s)$. We regard "the set of vectors $\{\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_s\}$" and "the columns of X" synomously.

We say that the vector $\underline{z}$ is a <u>linear combination</u> of the columns of X with coefficient vector $\underline{c} = \underline{c}(s)$ if

$$\underline{z} = X\underline{c} .$$

<u>Definition 4</u>: The <u>scalar product</u> (or <u>inner product</u>) of two vectors

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{and} \quad \underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \text{is the scalar} \quad \underline{x}'\underline{y} = \underline{y}'\underline{x}$$

in consonance with matrix transposition and multiplication.

<u>Properties</u>: If c is a scalar and $\underline{x}$, $\underline{y}$, and $\underline{z}$ are vectors, then $\underline{x}'(\underline{y}+\underline{z}) = \underline{x}'\underline{y} + \underline{x}'\underline{z}$, $\underline{x}'(c\underline{y}) = (c\underline{x})'\underline{y} = c(\underline{x}'\underline{y})$, $\underline{x}'\underline{x} \geq 0$. We denote a vector with all components zero by $\underline{0}$. The vector $\underline{x} = \underline{0}$ if and only if $\underline{x}'\underline{x} = 0$.

<u>Defintion 5</u>: The norm (or length) of a vector $\underline{x}$ written $\|\underline{x}\|$, is defined to be

$$\|\underline{x}\| = (\underline{x}'\underline{x})^{\frac{1}{2}} \quad .$$

It follows that $\|c\underline{x}\| = |c| \; \|\underline{x}\|$.

Law of cosines: For all vectors $\underline{x}$ and $\underline{y}$

$$\|\underline{x}-\underline{y}\|^2 = \|\underline{x}\|^2 + \|\underline{y}\|^2 - 2\|\underline{x}\| \; \|\underline{y}\| \cos \alpha$$

where α is the angle between $\underline{x}$ and $\underline{y}$.

Algebraic reduction shows that

$$\|\underline{x}\| \; \|\underline{y}\| \cos \alpha = x'y \; .$$

<u>Definition 6</u>: Two vectors $\underline{x}$ and $\underline{y}$ are said to be orthogonal if and only if $\underline{x}'\underline{y} = 0$.

<u>Notation</u>: We write $\underline{x} \perp \underline{y}$ to mean that the vectors $\underline{x}$ and $\underline{y}$ are orthogonal. More generally, we write $X \perp Y$ to mean that every column of X is orthogonal to every column of Y.

<u>Definition 7</u>: The <u>vector space</u> $\mathcal{M}(X)$ <u>spanned by columns of</u> X is the set of all linear combinations of the columns of X. We write $Y \in \mathcal{M}(X)$ to mean every column of Y is an element of $\mathcal{M}(X)$.

<u>Definition 8</u>: The columns of X are said to be <u>linearly dependent</u> if there exists a nonnull vector $\underline{c} = \underline{c}(s)$ such that

$$X\underline{c} = \underline{0} .$$

If no such $\underline{c}$ exists, the columns of X are called <u>linearly independent</u>. The columns of X are linearly independent if and only if none of its columns is a linear combination of the others.

<u>Lemma 1</u>: If the columns of X are linearly independent and $\underline{x}$ is a nonzero vector which is not a linear combination of them, the columns of $[X|x]$ are linearly independent.

<u>Definition 9</u>: A basis for $\mathcal{M}(X)$ is a set of linearly independent vectors that span $\mathcal{M}(X)$.

<u>Lemma 2</u>: Every vector space has a basis.

<u>Definition 10</u>: The rank of a vector space $\mathcal{M}(X)$ is the number of vectors in any basis.

<u>Theorem 1</u>: Any two bases for a vector space contain the same number of vectors.

Lemma 3: If $\underset{\sim}{x} \in \mathcal{M}(X)$ and the columns of A span $\mathcal{M}(X)$, then the vector $\underset{\sim}{c}$ in

$$x = A\underset{\sim}{c}$$

is unique only if the columns of A are linearly independent.

Definition 11: If the columns of A are a basis of $\mathcal{M}(X)$ and $A'A = I$ then the columns of A are called an orthonormal basis of $\mathcal{M}(X)$.

Lemma 4: If the columns of X are pairwise orthogonal and nonzero they are linearly independent.

Lemma 5: If $\mathcal{M}(X)$ has rank r, any linearly independent set of r vectors in $\mathcal{M}(X)$ is a basis for $\mathcal{M}(X)$.

Lemma 6: We can always construct an orthonormal basis of $\mathcal{M}(X)$.

Lemma 7: If the columns of A_1 are an orthonormal basis of $\mathcal{M}(X)$, we can construct A_2 such that $[A_1|A_2]$ is an orthonormal basis of $\mathcal{M}(I_n)$.

Definition 12: The vector $\underset{\sim}{x}$ is said to be orthogonal to $\mathcal{M}(X)$ (we write $\underset{\sim}{x} \perp \mathcal{M}(X)$) if and only if $\underset{\sim}{x}$ is orthogonal to every vector in $\mathcal{M}(X)$.

Lemma 8: A vector $\underset{\sim}{x}$ is orthogonal to $\mathcal{M}(X)$ if and only if $\underset{\sim}{x} \perp X$.

Lemma 9: For any vector $\underset{\sim}{y}$ there exists vectors $\underset{\sim}{x}$ and $\underset{\sim}{z}$ such that $\underset{\sim}{y} = \underset{\sim}{x} + \underset{\sim}{z}$, $\underset{\sim}{x} \in \mathcal{M}(X)$, $\underset{\sim}{z} \perp \mathcal{M}(X)$. This decomposition is unique.

Proof: Let the columns of A be an orthonormal basis of $\mathcal{M}(X)$, $x = AA'y$, and $z = (I-AA')y$. $X = AC$ for some matrix C, thus $AA'X = AA'AC = AC = X$. Thus $x \in \mathcal{M}(X)$, $z \perp \mathcal{M}(X)$, and $y = x + z$. To prove uniqueness suppose $y = x^* + z^*$ where $x^* \in \mathcal{M}(X)$, and $z^* \perp \mathcal{M}(X)$. Then $(x-x^*) + (z-z^*) = 0$. But $(x-x^*) \in \mathcal{M}(X)$ on one hand, and $(x-x^*) = -(z-z^*) \perp \mathcal{M}(X)$ on the other. Thus $x - x^* = 0$, $x^* = x$, and $z = z^*$.

The proof of uniqueness of x implies that x is in fact independent of the particular orthonormal basis used in its definition.

Definition 15: Given a vector y, the vector $x = AA'y$ defined in Lemma 9 which is such that $(y-x) \perp \mathcal{M}(X)$, is called the projection of y on $\mathcal{M}(X)$.

Definition 16: The matrix AA' defined in Lemma 9 is called a projection operator onto $\mathcal{M}(X)$. We denote AA' by P_X. The concept of a projection is so important to our study of linear models that the symbol P will not be used in any other context. It can be shown that P is symetric, idempotent, and $PX = X$.

Theorem 2: Given a fixed vector y, a fixed $\mathcal{M}(X)$ and a variable vector $x \in \mathcal{M}(X)$, then $\|y - x\|$ has a minimum value. This minimum is attained if and only if x is the projection of y onto $\mathcal{M}(X)$.

Proof: Write $y - x = (y-P_X y) + (P_X y-x)$. Then

$$\|y - x\|^2 = \|y - P_X y\|^2 + 2y'(I-P_X)(P_X y-x) + \|P_X y - x\|^2$$

$$= \|y - P_X y\|^2 + \|P_X y - x\|^2 .$$

The first term is fixed and the second term is nonnegative, and is zero if and only if $x = P_X y$.

<u>Theorem 3</u>: If $\underline{y}$ is orthogonal to every $\underline{w}$ orthogonal to $\mathcal{M}(X)$, then $\underline{y} \in \mathcal{M}(X)$.

<u>Proof</u>: Let $\underline{y} = \underline{x} + \underline{z}$ where $\underline{x} \in \mathcal{M}(X)$ and $\underline{z} \perp \mathcal{M}(X)$, and take the scalar product with $\underline{z}$ to get $\underline{z}'\underline{y} = z'x + z'z$. But $\underline{z}'\underline{x} = 0$, and by the hypothesis of the theorem $\underline{z}'\underline{y} = 0$; thus $\underline{z}'\underline{z} = 0$, $\underline{z} = \underline{0}$, and hence $\underline{y} = x \in \mathcal{M}(X)$.

<u>Definition 17</u>: The set of all vectors orthogonal to $\mathcal{M}(X)$ is called the ortho-complement of $\mathcal{M}(X)$. The orthogonal complement of $\mathcal{M}(X)$ is denoted by $\mathcal{O}(X)$.

<u>Lemma 10</u>: If $\mathcal{M}(X)$ has rank r, $\mathcal{O}(X)$ has rank n - r.

<u>Proof</u>: Let the columns of A_1 by an orthonormal basis for $\mathcal{M}(X)$ and the columns of $[A_1 | A_2]$ be an orthonormal basis of $\mathcal{M}(I_n)$. For any $\underline{x}$ there exist $\underline{c}_1 = \underline{c}_1(r)$ and $\underline{c}_2 = \underline{c}_2(n-r)$ such that $\underline{x} = A_1\underline{c}_1 + A_2\underline{c}_2$. $\underline{x} \perp \mathcal{M}(X)$ if and only if $\underline{c}_1 = 0$, i.e., if and only if it is in the (n-r) dimensional space of vectors of the form $A_2\underline{c}_2$.