

Dimension Reduction in Multivariate Linear Regression

William W.S. Chen

*Department of Statistics, the George Washington University
Washington D.C. 20013
williamwschen@gmail.com*

ABSTRACT

We defined some elementary terminology. It includes the vector space, linear combination, set of independent vectors, dependent vectors, basis of vector space, and direct sum of subspaces. This theory can help us lower the dimension of a given vector spaces. We apply to multivariate linear multiple regression analysis. It not only simplifies the computation and eases the interpretation, but also reduce the rate of errors. Cook (2010) developed an envelope model for the same reason. The main objective in that model is decomposing the covariance matrix into the sum of two matrices, each of whose column spaces either contains, or is orthogonal to, the subspace containing the mean. In other words, break the covariance matrix into the direct sum of the subspaces.

Mathematical Subject Classification: 15-00

Keywords and Phrases: basis of the vector space, direct sum of subspaces, decompose the covariance matrix, envelope model, finite dimensional vector space, linear combinations of vector, linearly dependent vectors, linearly independent vectors, multivariate linear regression, span of the vector space.

1 Introduction

Author worked for Internal Revenue Service for a long period of time. That organization has over one-hundred thousand variables available for study. It is not uncommon to study more than a hundred variables simultaneously at only one time. For example, in a case study there are 130 variables needed to find a relationship between the variables and predict its future trend. We need to run a multivariate regression analysis. Our experiences are calling a conference meeting with the subject matter group members. Using their suggestions, we can break down the variables into smaller groups. Let say 130 variables are cut into six groups of 20 variables each, and 10 variables in the remainder one. In this way, it is not only easier to run the program, but it will also help to interpret the output, and reduce the chance to make errors. Cook (2010) developed the envelope model that decomposed the covariance matrix into the direct sum of two matrices, each of whose column spaces either contains or is orthogonal to the subspace containing the mean. The objective of his model is consistent with ours.

2 Linear Transformation

Let V be n -dimensional vector space over a field F and let v_1, \dots, v_n be a basis of V over F . If $T \in A(V)$, then T is determined on any vector as soon as we know its action on a basis of V . Since T maps V into V , v_1T, v_2T, \dots, v_nT must all be in V . As elements of V , each of these is a realization in a unique way as a linear combination of v_1, \dots, v_n over F . Thus,

$$v_1T = \alpha_{11}v_1 + \alpha_{12}v_2 + \dots + \alpha_{1n}v_n$$

$$v_2T = \alpha_{21}v_1 + \alpha_{22}v_2 + \dots + \alpha_{2n}v_n$$

$$v_nT = \alpha_{n1}v_1 + \alpha_{n2}v_2 + \dots + \alpha_{nn}v_n$$

Where each $\alpha_{ij} \in F$. This system of equation can be written more compactly as

$$v_iT = \sum_{j=1}^n \alpha_{ij}v_j \quad i=1,2,\dots,n$$

The ordered set of n^2 numbers α_{ij} in F completely describes T . They will serve as the measure of representing T .

Definition 2.1

Let V be an n dimensional vector space over F and let v_1, \dots, v_n

be a basis for v over F . If $T \in A(V)$ then the matrix of T in the basis v_1, \dots, v_n written as $m(T)$, is

$$m(T) = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdot & \cdot & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdot & \cdot & \alpha_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \alpha_{n1} & \alpha_{n2} & \cdot & \cdot & \alpha_{nn} \end{pmatrix}$$

$$\text{where } v_iT = \sum_{j=1}^n \alpha_{ij}v_j$$

Example 2.1 Let the vector space V of 2×2 matrix over R and the following usual basis E of V .

$$E = \left\{ E_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, E_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, E_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, E_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}$$

Let $M = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ and T_1 be the linear transformation on V defined by $T_1(E) = ME$. Find the matrix representation of T_1 relative to the above usual basis of V .

$$T_1(E_1) = ME_1 = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 3 & 0 \end{pmatrix} = 1E_1 + 0E_2 + 3E_3 + 0E_4$$

$$T_1(E_2) = ME_2 = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 3 \end{pmatrix} = 0E_1 + 1E_2 + 0E_3 + 3E_4$$

$$T_1(E_3) = ME_3 = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 4 & 0 \end{pmatrix} = 2E_1 + 0E_2 + 4E_3 + 0E_4$$

$$T_1(E_4) = ME_4 = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 2 \\ 0 & 4 \end{pmatrix} = 0E_1 + 2E_2 + 0E_3 + 4E_4$$

Since $\dim V=4$, any matrix representation of a linear transformation on V must be a 4-square matrix.

Hence $|T_1|_E = \begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \\ 3 & 0 & 4 & 0 \\ 0 & 3 & 0 & 4 \end{pmatrix}$

Example 2.2 Let $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $T_2: V \rightarrow V$ be the linear transformation on V defined by $T_2(E) = EM$. Find the matrix representation of T_2 in the above usual basis of V .

$$T_2(E_1) = E_1M = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} = aE_1 + bE_2 + 0E_3 + 0E_4$$

$$T_2(E_2) = E_2M = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} c & d \\ 0 & 0 \end{pmatrix} = cE_1 + dE_2 + 0E_3 + 0E_4$$

$$T_2(E_3) = E_3M = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ a & b \end{pmatrix} = 0E_1 + 0E_2 + aE_3 + bE_4$$

$$T_2(E_4) = E_4M = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ c & d \end{pmatrix} = 0E_1 + 0E_2 + cE_3 + dE_4$$

$$\text{Hence } |T_2|_E = \begin{pmatrix} a & b & 0 & 0 \\ c & d & 0 & 0 \\ 0 & 0 & a & b \\ 0 & 0 & c & d \end{pmatrix} = \begin{pmatrix} M & 0 \\ 0 & M \end{pmatrix}$$

3 Define Terms

In this section we will define a sequence of useful terminology.

It includes vector space, subspace of a vector space, linear combination of the vector, span of the vectors, set of linear independent vectors, basis of the vector space, dimension of vector space. All these terms are required for us to understand dimension reduction. Define the vector space as follow.

Definition 3.1: A nonempty set V is said to be a vector space over a field F if V is an abelian group under an operation which we denote by $+$, and if for every $\alpha \in F, v \in V$ there is defined an element written as $\alpha v \in V$ subject to:

- (1) $\alpha(v + w) = \alpha v + \alpha w$, (2) $(\alpha + \beta)v = \alpha v + \beta v$, (3) $\alpha(\beta v) = (\alpha\beta)v$,
- (4) $1v = v$ for all $\alpha, \beta \in F, v, w \in V$.

Where 1 represents the unit element of F under multiplication.

definition 3.2: If V is a vector space over F if $W \subset V$ then W is a subspace of V if under the operation of V, W , itself, forms

a vector space over F . Equivalently, W is a subspace of V whenever $w_1, w_2 \in W, \alpha\beta \in F$ implies that $\alpha w_1 + \beta w_2 \in W$.

Definition 3.3: A linear combination of the vector space

is a sum of scalar multiples of these vectors, that is ,

$c_1v_1 + c_2v_2 + \dots + c_nv_n$, for some scalar coefficients, $c_1, c_2, \dots, c_n \in F$. If S is a set of vectors in V , a linear combination of vectors in S is a vector of the form $c_1v_1 + c_2v_2 + \dots + c_nv_n$ with $i \in \mathbb{N}, v_i \in S, c_i \in F$

Definition 3.4: The span of the vectors $c_1, c_2, \dots, c_n \in V$ is the set of all linear combinations of these vectors, denoted by $span(c_1, c_2, \dots, c_n)$. If S is a (finite or infinite) set of vectors in V , then the span of S , denoted by $span(S)$, is the set of all linear combinations of vectors in S . If $V = span(S)$, then S spans the vector space V .

Definition 3.5: A (finite or infinite) set of vectors S in V is linearly independent if the only linear combination of distinct vectors in S that produces the zero vector is a trivial linear combination. That is, if v_i are distinct vectors in S and $c_1v_1 + c_2v_2 + \dots + c_nv_n = 0$, then $c_1 = c_2 = \dots = c_n = 0$. Vectors that are not

linearly independent are linear dependent. That is, there exist distinct vectors $v_1, v_2, \dots, v_n \in S$ and c_1, c_2, \dots, c_n not all 0 such that

$$c_1 v_1 + c_2 v_2 + \dots + c_n v_n = 0.$$

Definition 3.6: Let V be a vector space over a field F . A set of vectors B in a vector space V is a basis for V if (1) B is a linearly independent set and (2) $\text{Span}(B)=V$.

Definition 3.7: The number of vectors in a basis for a vector space V is the dimension of V denoted by $\text{dim}(V)$.

Definition 3.8: The sum of two subspaces is direct if and only if the intersection of these subspaces is zero.

Since our objective is reduction the dimension of multivariate regression vector space using base theory. The following theorems can be found in Mal[^]Cev(1963) linear algebra. We only claim the theorems and not repeat its proof.

Theorem 3.1 The dimension of the sum of two linear subspaces is the sum of their dimensions minus the dimension of their intersection.

Proof: see Mal[^]Cev(1963), page 50. Alternatively, see Cook,R.D.,

Li,B. and Chiaromonte,F.(2010), page 936, Corollary 3.1.

Theorem 3.2 The dimension of a direct sum of subspaces is the sum of their dimension.

Proof: By theorem 3.1 the dimension of the sum is the sum of the dimensions minus the dimension of the intersection. From the definition 3.8, the intersection of the subspaces is zero

and has dimension zero. Therefore, the dimension of the direct sum of two subspaces is the sum of their dimensions.

If the number of summands is greater than 2, the proof carried out by induction. From theorem 3.2 we have the following.

Theorem 3.3 If a subspace A is the direct sum of subspaces

A_1, \dots, A_n , then taking a basis a_{i1}, \dots, a_{im_i} of each subspace A_i

for $i=1, \dots, n$ and combining these bases into one system,

$a_{11}, \dots, a_{1m_1}, \dots, a_{s1}, \dots, a_{sm_1}$ we obtain a Basis of the subspace A .

The above definitions and three theorems are the foundation of our applications. More of the related facts will be discussed in the section 4. We now give some examples to demonstrate its the real meaning and its possible applications.

Example 3.1 If we are given that

$$A_1 = \begin{pmatrix} 2 & 3 \\ 1 & 5 \end{pmatrix}, A_2 = (1 \ 2), A_3 = (3) \text{ then}$$

$$A_1 \oplus A_2 \oplus A_3 = \begin{pmatrix} 2 & 3 & 0 & 0 & 0 \\ 1 & 5 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 3 \end{pmatrix}$$

Example 3.2 Let $\mathbf{B} = \{V_1, \dots, V_n\}$ be a basis for V then

$$V = \text{span}(V_1) \oplus \dots \oplus \text{span}(V_n)$$

Example 3.3 In the direct sum $R(x; 2) \times R^{2 \times 2}$

$$(2x^2 + 7, \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}) + 3(x^2 + 4x - 2, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}) = (5x^2 + 12x + 1, \begin{pmatrix} 1 & 5 \\ 0 & 4 \end{pmatrix})$$

4 Some Facts

Although in section 3 we have defined the terms of basis and dimension of a vector space, it may not be good enough in actual application. We realize that these two terms are critical and useful to us. We list more facts to help us to more deeply understand these two terms. Fact 1: Every vector space has a basis.

Fact 2: The standard basis for F^n is a basis for F^n , and so $\dim F^n = n$. Fact 3: A basis B in a vector space V is the largest set of linearly independent vectors in V that contain B , and it is the smallest set of vectors in V that contains B and spans V . Fact 4: The empty set is a basis for the trivial vector space $\langle 0 \rangle$, and $\dim \langle 0 \rangle = 0$. Fact 5: If S is a linearly independent set in a vector space V , then S can be expanded, if necessary, to a basis for V . Fact 6: If the set $S = \langle v_1, \dots, v_p \rangle$ spans a vector space V , then some subset of S forms a basis for V . For example, if one of the vectors, say v_i , is a linear combination of the remaining vectors, then the set formed from S by removing v_i , will be "closer" to a basis for V . This process can be continued until the remaining vectors form a basis for V .

Fact 7: If a vector space V has a basis containing n vectors, then every basis of V must contain n vectors. Similarly if V has infinite basis, then every basis of V must be infinite. So the dimension of V is unique. Fact 8: Let $\dim(V) = n$ and let S be a set containing n vectors. The following are equivalent: (8.1) S is a basis for V . (8.2) S span V . (8.3) S is linear independent. Fact 9: If $\dim(V) = n$, then any subset of V containing more than n vectors is linearly dependent. Fact 10: If $B = \langle b_1, \dots, b_p \rangle$ is a basis for a vector space V , then each $x \in V$ can be expressed as a unique linear combination of the vectors in B . That is, for each $x \in V$ there is a unique set of scalars c_1, c_2, \dots, c_p such that $x = c_1 b_1 + c_2 b_2 + \dots + c_p b_p$. The concept of direct sum is the critical important to us. In section 3 we have clearly defined the direct sum and three important theorems. We found that the following fact can help us in real applications. Fact 1: $W = W_1 \oplus W_2$ if and only if $W = W_1 + W_2$ and $W_1 \cap W_2 = \langle 0 \rangle$ Fact 2: If W is a subspace of V ,

then there exists a subspace U of V such that $V=W \oplus U$. Note that U is not usually unique. Fact 3: Let $W=W_1+\dots+W_n$. The following statements are equivalent: (3.1) Let $W=W_1 \oplus \dots \oplus W_n$. That is, for all $i=1\dots n$,

$$\text{we have } W_i \cap \sum_{j \neq i} W_j = \langle 0 \rangle. \quad (3.2) \quad W_i \cap \sum_{j=1}^{i-1} W_j = \langle 0 \rangle \text{ for all } i=2,\dots,n$$

(3.3) For each $w \in W$, w can be expressed in exactly one way as a sum of vectors in $W_1 \dots W_n$. That is, there exists a unique $w_i \in W_i$, such that $w=w_1+\dots+w_n$. (3.4) The subspace

W_i , for $i=1\dots n$ are independent. (3.5) If B_i is an ordered basis for W_i , then $B = \bigcup_{i=1}^n B_i$ is an

ordered basis for W . Fact 4: If B is a basis for V and B is partitioned into disjoint subset B_i , for $i=1\dots n$. then $V = \text{span}(B_1) \oplus \dots \oplus \text{span}(B_n)$. Fact 5: If S is a linearly independent subset of V and S is

partitioned into disjoint subsets S_i , for $i=1\dots n$, then the subspaces $\text{span}(S_1), \dots, \text{span}(S_n)$ are independent. Fact 6: If V is finite dimensional and $V=W_1+\dots+W_n$, then $\dim(V)=\dim(W_1)+\dots+\dim(W_n)$ if and only if $V=W_1 \oplus \dots \oplus W_n$.

5 Concluding Remarks

The Multivariate general linear multiple regression model is given by $Y=XB+E$ where Y is $n \times p$, X is $n \times q$, B is $q \times p$, and E is $n \times p$ error matrix. If we move E matrix to the other side of equation and combine with Y matrix, then we can view X matrix as a linear transformation on the parameter matrix B to the response matrix Y . This paper is seeking some square submatrix W_i such that $W_i \cap \sum_{j \neq i} W_j = \langle 0 \rangle$

for $i=1,2,\dots,p$ and $W=W_1 \oplus \dots \oplus W_p$. We are sure that such square matrix exist as can be seen from example 2.2. However, the method to identify such submatrix from the given data set may not be trivial. Compare to the Cook (2010) envelope model, we found that our basic objective to reduce the dimension of vector space is identical. Envelope model attempt to decompose the covariance matrix into the direct sum of two submatrix, each of whose column space either contain or is orthogonal to the subspace containing the mean. The only way to do this is to create a split based on the eigenvector of the covariance. Of course, this will lead to a large computer related computation.

REFERENCES

- [1] Anderson, T.W. (1958). An Introduction to Multivariate Statistical Analysis. Second Edition, Wiley, New York.
- [2] Cook, R.D. (2007). Fisher lecture: Dimension Reduction in Regression (with discussion). Statist. Sci. 22 p1-26.

- [3] Cook, R.D., Li, B. and Chiaromonte, F. (2010). Envelope Models for Parsimonious and Efficient Multivariate Linear Regression. *Statistica Sinica* 20, p927-1010.
- [4] Cook, R.D., Li, B. and Chiaromonte, F. (2007). Dimension Reduction without matrix inversion. *Biometrika*, 94, p569-584.
- [5] Halmos, P.R. (1974,1987). *Finite-Dimensional Vector Space*. Published by Springer-Verlag, New York Inc.
- [6] *Handbook of Linear Algebra*. (2007). Edited by Leslie Hogben, Associate editors, Richard Brualdi, Anne Greenbaum, Roy Mathias. Chapman & Hall/CRC, Taylor & Francis Group, Boca Raton, London, New York.
- [7] Mal'cev, A.I.(1963). *Foundations of Linear Algebra*. Translated from the Russian by Thomas Craig Brown, Edited by Roberts, J.B. Published by W.H. Freeman and Company, San Francisco and London.