Version: June 19, 2007

Notes for Applied Multivariate Analysis with MATLAB

These notes were written for use within Quantitative Psychology courses at the University of Illinois, Champaign. The expectation is that for Psychology 406/7 (Statistical Methods I and II), the material up through Section 0.1.12 be available to a student. For Multivariate Analysis (Psychology 594) and Covariance Structure and Factor Models (Psychology 588), the remainder of the notes are relevant, with particular emphasis on Singular Value Decomposition (SVD) and Eigenvector/Eigenvalue Decomposition (Spectral Decomposition).

Contents

List of Figures

0.1 Necessary Matrix Algebra Tools

The strategies of multivariate analysis tend to be confusing unless specified compactly in matrix terms. Therefore, we will spend some significant amount of time on these topics because, in fact, most of multivariate analysis falls out directly once we have these tools under control. Remember the old Saturday Night Live skit with Hans and Franz, "listen to me now, and believe me later". I have a goal in mind of where I would like you all to be — at the point of understanding and being able to work with what is called the Singular Value Decomposition (SVD) of a matrix, and to understand the matrix topics that lead up to the SVD. Very much like teaching how to use some word-processing program, where we need to learn all the various commands and what they do, an introduction to the matrix tools can seem a little disjointed. But just like word-processing comes together more meaningfully when required to do your own manuscripts from beginning to end, once we proceed into the techniques of multivariate analysis per se, the wisdom of this preliminary matrix excursion will be apparent.

0.1.1 Preliminaries

A *matrix* is merely an array of numbers; for example,

$$\begin{pmatrix} 4 & -1 & 3 & 1 \\ 4 & 6 & 0 & 2 \\ 7 & 2 & 1 & 4 \end{pmatrix}$$

is a matrix. In general, we denote a matrix by an uppercase (capital) boldface letter such as $\mathbf{A}$ (or using a proofreader representation

on the blackboard, a capital letter with a wavy line underneath to indicate boldface):

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1V} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2V} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{U1} & a_{U2} & a_{U3} & \cdots & a_{UV} \end{pmatrix}$$

This matrix has U rows and V columns and is said to have *order* $U \times V$. An arbitrary element a_{uv} refers to the element in the u^{th} row and v^{th} column, with the row index always preceding the column index (and therefore, we might use the notation of $\mathbf{A} = \{a_{uv}\}_{U \times V}$ to indicate the matrix $\mathbf{A}$ as well as its order).

A 1×1 matrix such as $(4)_{1\times 1}$ is just an ordinary number, called a *scalar*. So without loss of any generality, numbers are just matrices. A *vector* is a matrix with a single row or column; we denote a column vector by a lowercase boldface letter, e.g., $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, and so on. The vector

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_U \end{pmatrix}_{U\times 1}$$

is of order $U \times 1$; the column indices are typically omitted since there is only one. A row vector is written as

$$\mathbf{x}' = (x_1, \ldots, x_U)_{1\times U}$$

with the prime indicating the *transpose* of $\mathbf{x}$, i.e., the interchange of row(s) and column(s). This transpose operation can be applied to any matrix; for example,

$$\mathbf{A} = \begin{pmatrix} 1 & -1 \\ 3 & 7 \\ 4 & 1 \end{pmatrix}_{3\times 2}$$

$$\mathbf{A}' = \begin{pmatrix} 1 & 3 & 4 \\ -1 & 7 & 1 \end{pmatrix}_{2\times 3}$$

If a matrix is *square*, defined by having the same number of rows as columns, say U, and if the matrix and its transpose are equal, the matrix is said to be *symmetric*. Thus, in $\mathbf{A} = \{a_{uv}\}_{U\times U}$, $a_{uv} = a_{vu}$ for all u and v. As an example,

$$\mathbf{A} = \mathbf{A}' = \begin{pmatrix} 1 & 4 & 3 \\ 4 & 7 & -1 \\ 3 & -1 & 3 \end{pmatrix}$$

For a square matrix $\mathbf{A}_{U\times U}$, the elements a_{uu}, $1 \leq u \leq U$, lie along the *main* or *principal* diagonal. The sum of main diagonal entries of a square matrix is called the *trace*; thus,

$$\text{trace}(\mathbf{A}_{U\times U}) \equiv \text{tr}(\mathbf{A}) = a_{11} + \cdots + a_{UU}$$

A number of special matrices appear periodically in the notes to follow. A $U \times V$ matrix of all zeros is called a *null* matrix, and might be denoted by

$$\emptyset = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}$$

Similarly, we might at times need an $U \times V$ matrix of all ones, say $\mathbf{E}$:

$$\mathbf{E} = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$$

A *diagonal* matrix is square with zeros in all the off main-diagonal positions:

$$\mathbf{D}_{U\times U} = \begin{pmatrix} a_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_U \end{pmatrix}_{U\times U}$$

Here, we again indicate the main diagonal entries with just one index as $a_1, a_2, \ldots, a_U$. If all of the main diagonal entries in a diagonal matrix are 1s, we have the *identity* matrix denoted by $\mathbf{I}$:

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

To introduce some useful operations on matrices, suppose we have two matrices $\mathbf{A}$ and $\mathbf{B}$ of the same $U \times V$ order:

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1V} \\ \vdots & \ddots & \vdots \\ a_{U1} & \cdots & a_{UV} \end{pmatrix}_{U\times V}$$

$$\mathbf{B} = \begin{pmatrix} b_{11} & \cdots & b_{1V} \\ \vdots & \ddots & \vdots \\ b_{U1} & \cdots & b_{UV} \end{pmatrix}_{U\times V}$$

As a definition for equality of two matrices of the same order (and for which it only makes sense to talk about equality), we have:

$\mathbf{A} = \mathbf{B}$ if and only if $a_{uv} = b_{uv}$ for all u and v.

Remember, the "if and only if" statement (sometimes abbreviated as "iff") implies two conditions:

if $\mathbf{A} = \mathbf{B}$, then $a_{uv} = b_{uv}$ for all u and v;

if $a_{uv} = b_{uv}$ for all u and v, then $\mathbf{A} = \mathbf{B}$.

Any definition by its very nature implies an "if and only if" statement.

To add two matrices together, they first have to be of the same order (referred to as conformal for addition); we then do the addition component by component:

$$\mathbf{A} + \mathbf{B} = \begin{pmatrix} a_{11} + b_{11} & \cdots & a_{1V} + b_{1V} \\ \vdots & \ddots & \vdots \\ a_{U1} + b_{U1} & \cdots & a_{UV} + b_{UV} \end{pmatrix}_{U \times V}$$

To preform scalar multiplication of a matrix $\mathbf{A}$ by, say, a constant c, we again do the multiplication component by component:

$$c\mathbf{A} = \begin{pmatrix} ca_{11} & \cdots & ca_{1V} \\ \vdots & \ddots & \vdots \\ ca_{U1} & \cdots & ca_{UV} \end{pmatrix} = c \begin{pmatrix} a_{11} & \cdots & a_{1V} \\ \vdots & \ddots & \vdots \\ a_{U1} & \cdots & a_{UV} \end{pmatrix}$$

Thus, if one wished to define the difference of two matrices, we could proceed rather obviously as follows:

$$\mathbf{A} - \mathbf{B} \equiv \mathbf{A} + (-1)\mathbf{B} = \{a_{uv} - b_{uv}\}$$

One of the more important matrix operations is multiplication where two matrices are said to be conformal for multiplication if the

number of rows in one matches the number of columns in the second. For example, suppose $\mathbf{A}$ is $U \times V$ and $\mathbf{B}$ is $V \times W$; then, because the number of columns in $\mathbf{A}$ matches the number of rows in $\mathbf{B}$, we can define $\mathbf{AB}$ as $\mathbf{C}_{U \times W}$, where $\{c_{uw}\} = \{\Sigma_{k=1}^{V} a_{uk} b_{kw}\}$. This process might be referred to as row (of $\mathbf{A}$) by column (of $\mathbf{B}$) multiplication; the following simple example should make this clear:

$$\mathbf{A}_{3\times 2} = \begin{pmatrix} 1 & 4 \\ 3 & 1 \\ -1 & 0 \end{pmatrix}, \ \mathbf{B}_{2\times 4} = \begin{pmatrix} -1 & 2 & 0 & 1 \\ 1 & 0 & 1 & 4 \end{pmatrix};$$

$$\mathbf{AB} = \mathbf{C}_{3\times 4} =$$

$$\begin{pmatrix} 1(-1)+4(1) & 1(2)+4(0) & 1(0)+4(1) & 1(1)+4(4) \\ 3(-1)+1(1) & 3(2)+1(0) & 3(0)+1(1) & 3(1)+1(4) \\ -1(-1)+0(1) & -1(2)+0(0) & -1(0)+0(1) & -1(1)+0(4) \end{pmatrix} =$$

$$\begin{pmatrix} 3 & 2 & 4 & 17 \\ -2 & 6 & 1 & 7 \\ 1 & -2 & 0 & -1 \end{pmatrix}$$

Some properties of matrix addition and multiplication follow, where the matrices are assumed conformal for the operations given:

(A) matrix addition is commutative:

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$$

(B) matrix addition is associative:

$$\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$$

(C) matrix multiplication is right and left distributive over matrix addition:

$$\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$$

$$(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC}+\mathbf{BC}$$

(D) matrix multiplication is associative:

$$\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$$

In general, $\mathbf{AB} \neq \mathbf{BA}$ even if both products are defined. Thus, multiplication is not commutative as the following simple example shows:

$$\mathbf{A}_{2\times 2} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix};\ \mathbf{B}_{2\times 2} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix};\ \mathbf{AB} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix};\ \mathbf{BA} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

In the product $\mathbf{AB}$, we say that $\mathbf{B}$ is *premultiplied* by $\mathbf{A}$ and $\mathbf{A}$ is *postmultiplied* by $\mathbf{B}$. Thus, if we pre- or postmultiply a matrix by the identity, the same matrix is retrieved:

$$\mathbf{I}_{U\times U}\mathbf{A}_{U\times V} = \mathbf{A}_{U\times V};\ \mathbf{A}_{U\times V}\mathbf{I}_{V\times V} = \mathbf{A}_{U\times V}$$

If we premultiply $\mathbf{A}$ by a diagonal matrix $\mathbf{D}$, then each row of $\mathbf{A}$ is multiplied by a particular diagonal entry in $\mathbf{D}$:

$$\mathbf{D}_{U\times U}\mathbf{A}_{U\times V} = \begin{pmatrix} d_1 a_{11} & \cdots & d_1 a_{1V} \\ \vdots & \ddots & \vdots \\ d_U a_{r1} & \cdots & d_U a_{UV} \end{pmatrix}$$

If $\mathbf{A}$ is post-multiplied by a diagonal matrix $\mathbf{D}$, then each column of $\mathbf{A}$ is multiplied by a particular diagonal entry in $\mathbf{D}$:

$$\mathbf{A}_{U\times V}\mathbf{D}_{V\times V} = \begin{pmatrix} d_1 a_{11} & \cdots & d_V a_{1V} \\ \vdots & \ddots & \vdots \\ d_1 a_{U1} & \cdots & d_V a_{UV} \end{pmatrix}$$

Finally, we end this section with a few useful results on the transpose operation and matrix multiplication and addition:

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}';\ (\mathbf{ABC})' = \mathbf{C}'\mathbf{B}'\mathbf{A}';\ \ldots$$

$$(\mathbf{A}')' = \mathbf{A};\ (\mathbf{A}+\mathbf{B})' = \mathbf{A}'+\mathbf{B}'$$

0.1.2 The Data Matrix

A very common type of matrix encountered in multivariate analysis is what is referred to as a data matrix containing, say, observations for N subjects on P variables. We will typically denote this matrix by $\mathbf{X}_{N\times P} = \{x_{ij}\}$, with a generic element of x_{ij} referring to the observation for subject or row i on variable or column j ($1 \le i \le N$ and $1 \le j \le P$):

$$\mathbf{X}_{N\times P} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1P} \\ x_{21} & x_{22} & \cdots & x_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NP} \end{pmatrix}$$

All right-thinking people always list subjects as rows and variables as columns, conforming also to the now-common convention for computer spreadsheets.

Any matrix in general, including a data matrix, can be viewed either as a collection of its row vectors or of its column vectors,

and these interpretations can be generally useful. For a data matrix $\mathbf{X}_{N\times P}$, let $\mathbf{x}_i' = (x_{i1}, \ldots, x_{iP})_{1\times P}$ denote the row vector for subject i, $1 \leq i \leq N$, and let $\mathbf{v}_j$ denote the $N \times 1$ column vector for variable j:

$$\mathbf{v}_j = \begin{pmatrix} x_{1j} \\ \vdots \\ x_{Nj} \end{pmatrix}_{N\times 1}$$

Thus, each subject could be viewed as providing a vector of coordinates $(1 \times P)$ in P-dimensional "variable space", where the P axes correspond to the P variables; or each variable could be viewed as providing a vector of coordinates $(N \times 1)$ in "subject space", where the N axes correspond to the N subjects:

$$\mathbf{X}_{N\times P} = \begin{pmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_N' \end{pmatrix} = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_P \end{pmatrix}$$

0.1.3 Inner Products

The *inner product* (also called the dot or scalar product) of two vectors , $\mathbf{x}_{U\times 1}$ and $\mathbf{y}_{U\times 1}$, is defined as

$$\mathbf{x}'\mathbf{y} = (x_1, \ldots, x_U) \begin{pmatrix} y_1 \\ \vdots \\ y_U \end{pmatrix} = \sum_{u=1}^{U} x_u y_u$$

Thus, the inner product of a vector with itself is merely the sum of squares of the entries in the vector: $\mathbf{x}'\mathbf{x} = \sum_{u=1}^{U} x_u^2$. Also, because

an inner product is a scalar and must equal it own transpose (i.e., $\mathbf{x}'\mathbf{y} = (\mathbf{x}'\mathbf{y})' = \mathbf{y}'\mathbf{x}$), we have the end result that

$$\mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x}$$

If there is an inner product, there should also be an *outer product* defined as the $U \times U$ matrices given by $\mathbf{x}\mathbf{y}'$ or as $\mathbf{y}\mathbf{x}'$. As indicated by the display equations below, $\mathbf{x}\mathbf{y}'$ is the transpose of $\mathbf{y}\mathbf{x}'$:

$$\mathbf{x}\mathbf{y}' = \begin{pmatrix} x_1 \\ \vdots \\ x_U \end{pmatrix} (y_1, \ldots, y_N) = \begin{pmatrix} x_1 y_1 & \cdots & x_1 y_U \\ \vdots & \ldots & \vdots \\ x_U y_1 & \cdots & x_U y_U \end{pmatrix}$$

$$\mathbf{y}\mathbf{x}' = \begin{pmatrix} y_1 \\ \vdots \\ y_U \end{pmatrix} (x_1, \ldots, x_U) = \begin{pmatrix} y_1 x_1 & \cdots & y_1 x_U \\ \vdots & \ldots & \vdots \\ y_U x_1 & \cdots & y_U x_U \end{pmatrix}$$

A vector can be viewed as a geometrical vector in U dimensional space. Thus, the two 2×1 vectors

$$\mathbf{x} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}; \ \mathbf{y} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$$

can be represented in the two-dimensional Figure 1 below, with the entries in the vectors defining the coordinates of the endpoints of the arrows.

The *Euclidean distance* between two vectors, $\mathbf{x}$ and $\mathbf{y}$, is given as:

$$\sqrt{\sum_{u=1}^{U} (x_u - y_u)^2} = \sqrt{(\mathbf{x} - \mathbf{y})'(\mathbf{x} - \mathbf{y})}$$

and the *length* of any vector is the Euclidean distance between the vector and the origin. Thus, in Figure 1, the distance between $\mathbf{x}$ and $\mathbf{y}$ is $\sqrt{10}$ with respective lengths of 5 and $\sqrt{17}$.

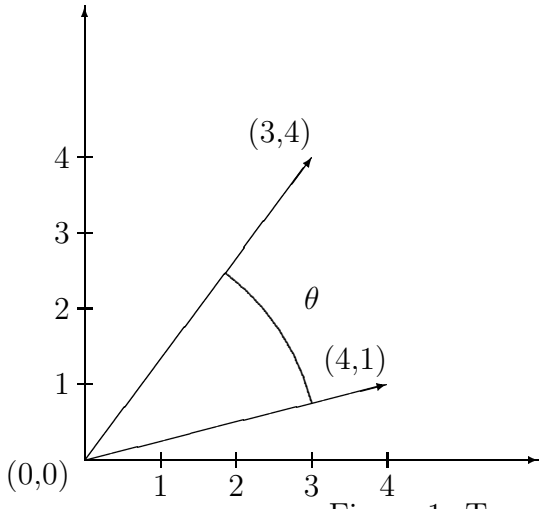

Figure 1: Two vectors plotted in two-dimensional space

The cosine of the angle between the two vectors $\mathbf{x}$ and $\mathbf{y}$ if defined by:

$$\cos(\theta) = \frac{\mathbf{x}'\mathbf{y}}{(\mathbf{x}'\mathbf{x})^{1/2}(\mathbf{y}'\mathbf{y})^{1/2}}$$

Thus, in the figure we have

$$\cos(\theta) = \frac{\begin{pmatrix} 3 & 4 \end{pmatrix}\begin{pmatrix} 4 \\ 1 \end{pmatrix}}{5\sqrt{17}} = \frac{16}{5\sqrt{17}} = .776$$

The cosine value of .776 corresponds to an angle of 39.1 degrees or .68 radians; these later values can be found with the inverse (or arc) cosine function (on, say, a hand calculator, or using MATLAB as we suggest in the next section).

When the means of the entries in $\mathbf{x}$ and $\mathbf{y}$ are zero (i.e., deviations from means have been taken), then $\cos(\theta)$ is the correlation between

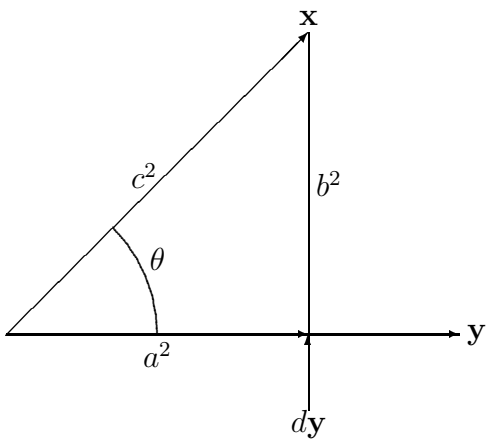

Figure 2: Illustration of projecting one vector onto another

the entries in the two vectors. Vectors at right angles have $\cos(\theta) = 0$, or alternatively, the correlation is zero.

Figure 2 shows two generic vectors, $\mathbf{x}$ and $\mathbf{y}$, where without loss of any real generality, $\mathbf{y}$ is drawn horizontally in the plane and $\mathbf{x}$ is projected at a right angle onto the vector $\mathbf{y}$ resulting in a point defined as a multiple d of the vector $\mathbf{y}$. The formula for d that we demonstrate below is based on the Pythagorean theorem that $c^2 = b^2 + a^2$:

$$c^2 = b^2 + a^2 \Rightarrow \mathbf{x}'\mathbf{x} = (\mathbf{x} - d\mathbf{y})'(\mathbf{x} - d\mathbf{y}) + d^2\mathbf{y}'\mathbf{y} \Rightarrow$$

$$\mathbf{x}'\mathbf{x} = \mathbf{x}'\mathbf{x} - d\mathbf{x}'\mathbf{y} - d\mathbf{y}'\mathbf{x} + d^2\mathbf{y}'\mathbf{y} + d^2\mathbf{y}'\mathbf{y} \Rightarrow$$

$$0 = -2d\mathbf{x}'\mathbf{y} + 2d^2\mathbf{y}'\mathbf{y} \Rightarrow$$

$$d = \frac{\mathbf{x}'\mathbf{y}}{\mathbf{y}'\mathbf{y}}$$

The diagram in Figure 2 is somewhat constricted in the sense that the angle between the vectors shown is less than 90 degrees; this allows the constant d to be positive. Other angles might lead to negative d when defining the projection of $\mathbf{x}$ onto $\mathbf{y}$, and would merely indicate the need to consider the vector $\mathbf{y}$ oriented in the opposite (negative) direction. Similarly, the vector $\mathbf{y}$ is drawn with a larger length than $\mathbf{x}$ which gives a value for d that is less than 1.0; otherwise, d would be greater than 1.0 indicating a need to stretch $\mathbf{y}$ to represent the point of projection onto it.

There are other formulas possible based on this geometric information: the length of the projection is merely d times the length of $\mathbf{y}$; and $\cos(\theta)$ can be given as the length of $d\mathbf{y}$ divided by the length of $\mathbf{x}$, which is $d\sqrt{\mathbf{y}'\mathbf{y}}/\sqrt{\mathbf{x}'\mathbf{x}} = \mathbf{x}'\mathbf{y}/(\sqrt{\mathbf{x}'\mathbf{x}}\sqrt{\mathbf{y}'\mathbf{y}})$.

0.1.4 Determinants

To each square matrix, $\mathbf{A}_{U\times U}$, there is an associated scalar called the *determinant* of $\mathbf{A}$ that is denoted by $|\mathbf{A}|$ or $\det(\mathbf{A})$. Determinants up to a 3×3 can be given by formula:

$$\det(\begin{pmatrix} a \end{pmatrix}_{1\times 1}) = a;\ \det(\begin{pmatrix} a & b \\ c & d \end{pmatrix}_{2\times 2}) = ad - bc;$$

$$\det(\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}_{3\times 3}) = aei + dhc + gfb - (ceg + fha + idb)$$

Beyond a 3×3 we can use a recursive process illustrated below. This requires the introduction of a few additional matrix terms that we now give: for a square matrix $\mathbf{A}_{U \times U}$, define $\mathbf{A}_{uv}$ to be the $(n-1) \times (n-1)$ submatrix of $\mathbf{A}$ constructed by deleting the u^{th} row and v^{th} column of $\mathbf{A}$. We call $\det(\mathbf{A}_{uv})$ the *minor* of the entry a_{uv}; the signed minor of $(-1)^{u+v} \det(\mathbf{A}_{uv})$ is called the *cofactor* of a_{uv}. The recursive algorithm would chose some row or column (rather arbitrarily), and find the cofactors for the entries in it; the cofactors would then be weighted by the relevant entries and summed.

As an example, consider the 4×4 matrix

$$\begin{pmatrix} 1 & -1 & 3 & 1 \\ -1 & 1 & 0 & -1 \\ 3 & 2 & 1 & 2 \\ 1 & 2 & 4 & 3 \end{pmatrix}$$

and choose the second row. The expression below involves the weighted cofactors for 3×3 submatrices that can be obtained by formulas. Beyond a 4×4 there will be nesting of the processes:

$$(-1)((-1)^{2+1}) \det(\begin{pmatrix} -1 & 3 & 1 \\ 2 & 1 & 2 \\ 2 & 4 & 3 \end{pmatrix}) + (1)((-1)^{2+2}) \det(\begin{pmatrix} 1 & 3 & 1 \\ 3 & 1 & 2 \\ 1 & 4 & 3 \end{pmatrix}) +$$

$$(0)((-1)^{2+3}) \det(\begin{pmatrix} 1 & -1 & 1 \\ 3 & 2 & 2 \\ 1 & 2 & 3 \end{pmatrix}) + (-1)((-1)^{2+4}) \det(\begin{pmatrix} 1 & -1 & 3 \\ 3 & 2 & 1 \\ 1 & 2 & 4 \end{pmatrix}) =$$

$$5 + (-15) + 0 + (-29) = -39$$

Another strategy to find the determinant of a matrix is to reduce it a form in which we might note the determinant more or less by simple

inspection. The reductions could be carried out by operations that have a known effect on the determinant; the form which we might seek is a matrix that is either *upper-triangular* (all entries below the main diagonal are all zero), *lower-triangular* (all entries above the main diagonal are all zero), or diagonal. In these latter cases, the determinant is merely the product of the diagonal elements. Once found, we can note how the determinant might have been changed by the reduction process and carry out the reverse changes to find the desired determinant.

The properties of determinants that we could rely on in the above iterative process are as follows:

(A) if *one* row of $\mathbf{A}$ is multiplied by a constant c, the new determinant is $c\det(\mathbf{A})$; the same is true for multiplying a column by c;

(B) if two rows or two columns of a matrix are interchanged, the sign of the determinant is changed;

(C) if two rows or two columns of a matrix are equal, the determinant is zero;

(D) the determinant is unchanged by adding a multiple of some row to another row; the same is true for columns;

(E) a zero row or column implies a zero determinant;

(F) $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$

0.1.5 Linear Independence/Dependence of Vectors

Suppose I have a collection of K vectors each of size $U \times 1$, $\mathbf{x}_1, \ldots, \mathbf{x}_K$. If no vector in the set can be written as a linear combination of the remaining ones, the set of vectors is said to be *linearly independent*; otherwise, the vectors are *linearly dependent*. As an example,

consider the three vectors:

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 4 \\ 0 \end{pmatrix};\ \mathbf{x}_2 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix};\ \mathbf{x}_3 = \begin{pmatrix} 3 \\ 7 \\ 1 \end{pmatrix}$$

Because $2\mathbf{x}_1 + \mathbf{x}_2 = \mathbf{x}_3$, we have a linear dependence among the three vectors; however, $\mathbf{x}_1$ and $\mathbf{x}_2$, or, $\mathbf{x}_2$ and $\mathbf{x}_3$, are linearly independent.

If the U vectors (each of size $U \times 1$), $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_U$, are linearly independent, then the collection defines a *basis*, i.e., any vector can be written as a linear combination of $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_U$. For example, using the *standard basis*, $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_U$, where $\mathbf{e}_u$ is a vector of all zeros except for a single one in the u^{th} position, any vector $\mathbf{x}' = (x_1, \ldots, x_U)$ can be written as:

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_U \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + x_U \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} =$$

$$x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_U\mathbf{e}_U$$

Bases that consist of *orthogonal* vectors (where all inner products are zero) are important later in what is known as principal components analysis. The standard basis involves orthogonal vectors, and any other basis may always be modified by what is called the Gram-Schmidt orthogonalization process to produce a new basis that does contain all orthogonal vectors.

0.1.6 Matrix Inverses

Suppose $\mathbf{A}$ and $\mathbf{B}$ are both square and of size $U \times U$. If $\mathbf{AB} = \mathbf{I}$, then $\mathbf{B}$ is said to be an inverse of $\mathbf{A}$ and is denoted by $\mathbf{A}^{-1} (\equiv \mathbf{B})$.

Also, if $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$, then $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ holds automatically. If $\mathbf{A}^{-1}$ exists, the matrix $\mathbf{A}$ is said to be *nonsingular*; if $\mathbf{A}^{-1}$ does not exist, $\mathbf{A}$ is *singular*.

An example:

$$\begin{pmatrix} 1 & 3 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} -1/5 & 3/5 \\ 2/5 & -1/5 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} -1/5 & 3/5 \\ 2/5 & -1/5 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Given a matrix $\mathbf{A}$, the inverse $\mathbf{A}^{-1}$ can be found using the following four steps:

(A) form a matrix of the same size as $\mathbf{A}$ containing the minors for all entries of $\mathbf{A}$;

(B) multiply the matrix of minors by $(-1)^{u+v}$ to produce the matrix of cofactors;

(C) divide all entries in the cofactors matrix by $\det(\mathbf{A})$;

(D) the transpose of the matrix found in (C) gives $\mathbf{A}^{-1}$.

As a mnemonic device to remember these four steps, we have the phrase "My Cat Does Tricks" for Minor, Cofactor, Determinant Division, Transpose" (I tried to work "my cat turns tricks" into the appropriate phrase but failed with the second to the last "t"). Obviously, an inverse exists for a matrix $\mathbf{A}$ if $\det(\mathbf{A}) \neq 0$, allowing the division in step (C) to take place.

An example: for

$$\mathbf{A} = \begin{pmatrix} 1 & 3 & 2 \\ 0 & 1 & 1 \\ 0 & 2 & 1 \end{pmatrix}; \ \det(\mathbf{A}) = -1$$

Step (A), the matrix of minors:

$$\begin{pmatrix} -1 & 0 & 0 \\ -1 & 1 & 2 \\ 1 & 1 & 1 \end{pmatrix}$$

Step (B), the matrix of cofactors:

$$\begin{pmatrix} -1 & 0 & 0 \\ 1 & 1 & -2 \\ 1 & -1 & 1 \end{pmatrix}$$

Step (C), determinant division:

$$\begin{pmatrix} 1 & 0 & 0 \\ -1 & -1 & 2 \\ -1 & 1 & -1 \end{pmatrix}$$

Step (D), matrix transpose:

$$\mathbf{A}^{-1} = \begin{pmatrix} 1 & -1 & -1 \\ 0 & -1 & 1 \\ 0 & 2 & -1 \end{pmatrix}$$

We can easily verify that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$:

$$\begin{pmatrix} 1 & 3 & 2 \\ 0 & 1 & 1 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 & -1 \\ 0 & -1 & 1 \\ 0 & 2 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

As a very simple instance of the mnemonic in the case of a 2×2 matrix with arbitrary entries:

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

the inverse exists if $\det(\mathbf{A}) = ad - bc \neq 0$:

$$\mathbf{A}^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Several properties of inverses are given below that will prove useful in our continuing presentation:

(A) if $\mathbf{A}$ is symmetric, then so is $\mathbf{A}^{-1}$;

(B) $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$; or, the inverse of a transpose is the transpose of the inverse;

(C) $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$; $(\mathbf{ABC})^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}$; or, the inverse of a product is the product of inverses in the opposite order;

(D) $(c\mathbf{A})^{-1} = (\frac{1}{c})\mathbf{A}^{-1}$; or, the inverse of a scalar times a matrix is the scalar inverse times the matrix inverse;

(E) the inverse of a diagonal matrix, is also diagonal with the entries being the inverses of the entries from the original matrix (assuming none are zero):

$$\begin{pmatrix} a_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_U \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{a_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{a_U} \end{pmatrix}$$

0.1.7 Matrices as Transformations

Any $U \times V$ matrix $\mathbf{A}$ can be seen as transforming a $V \times 1$ vector $\mathbf{x}_{V\times 1}$ to another $U \times 1$ vector $\mathbf{y}_{U\times 1}$:

$$\mathbf{y}_{U\times 1} = \mathbf{A}_{U\times V}\mathbf{x}_{V\times 1}$$

or,

$$\begin{pmatrix} y_1 \\ \vdots \\ y_U \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1V} \\ \vdots & \ddots & \vdots \\ a_{U1} & \cdots & a_{UV} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_V \end{pmatrix}$$

where $y_u = a_{u1}x_1 + a_{u2}x_2 + \cdots + a_{uV}x_V$. Alternatively, $\mathbf{y}$ can be written as a linear combination of the columns of $\mathbf{A}$ with weights given by $x_1, \ldots, x_V$:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_U \end{pmatrix} = x_1 \begin{pmatrix} a_{11} \\ \vdots \\ a_{U1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ \vdots \\ a_{U2} \end{pmatrix} + \cdots + x_V \begin{pmatrix} a_{1V} \\ \vdots \\ a_{UV} \end{pmatrix}$$

To indicate one common usage for matrix transformation in a data context, suppose we consider our data matrix $\mathbf{X} = \{x_{ij}\}_{N\times P}$, where x_{ij} represents an observation for subject i on variable j. We would like to use matrix transformations to produce a standardized matrix $\mathbf{Z} = \{(x_{ij} - \bar{x_j})/s_j\}_{N\times P}$, where $\bar{x_j}$ is the mean of the entries in the j^{th} column and s_j is the corresponding standard deviation; thus, the columns of $\mathbf{Z}$ all have mean zero and standard deviation one. A matrix expression for this transformation could be written as follows:

$$\mathbf{Z}_{N\times P} = (\mathbf{I}_{N\times N} - (\frac{1}{N})\mathbf{E}_{N\times N})\mathbf{X}_{N\times P}\mathbf{D}_{P\times P}$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{E}$ contains all ones, and $\mathbf{D}$ is a diagonal matrix containing $\frac{1}{s_1}, \frac{1}{s_2}, \ldots, \frac{1}{s_P}$, along the main diagonal positions. Thus, $(\mathbf{I}_{N\times N} - (\frac{1}{N})\mathbf{E}_{N\times N})\mathbf{X}_{N\times P}$ produces a matrix with columns deviated from the column means; a postmultiplication by $\mathbf{D}$ carries out the within column division by the standard deviations. Finally, if we define the expression $(\frac{1}{N})(\mathbf{Z}'\mathbf{Z})_{P\times P} \equiv \mathbf{R}_{P\times P}$, we have the familiar correlation coefficient matrix among the P variables.

0.1.8 Matrix and Vector Orthogonality

Two vectors, $\mathbf{x}$ and $\mathbf{y}$, are said to be *orthogonal* if $\mathbf{x}'\mathbf{y} = 0$, and would lie at right angles when graphed. If, in addition, $\mathbf{x}$ and $\mathbf{y}$ are both of unit length (i.e., $\sqrt{\mathbf{x}'\mathbf{x}} = \sqrt{\mathbf{y}'\mathbf{y}} = 1$), then they are said to be *orthonormal*. A square matrix $\mathbf{T}_{U\times U}$ is said to be *orthogonal* if its rows form a set of mutually orthonormal vectors. An example (called a Helmert matrix of order 3) follows:

$$\mathbf{T} = \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ -1/\sqrt{6} & -1/\sqrt{6} & 2/\sqrt{6} \end{pmatrix}$$

There are several nice properties of orthogonal matrices that we will see again in our various discussions to follow:

(A) $\mathbf{T}\mathbf{T}' = \mathbf{T}'\mathbf{T} = \mathbf{I}$;

(B) the columns of $\mathbf{T}$ are orthonormal;

(C) $\det(\mathbf{T}) = \pm 1$;

(D) if $\mathbf{T}$ and $\mathbf{R}$ are orthogonal, then so is $\mathbf{T}\mathbf{R}$;

(E) vectors lengths do not change under an orthogonal transformation: to see this, let $\mathbf{y} = \mathbf{T}\mathbf{x}$; then

$$\mathbf{y}'\mathbf{y} = (\mathbf{T}\mathbf{x})'(\mathbf{T}\mathbf{x}) = \mathbf{x}'\mathbf{T}'\mathbf{T}\mathbf{x} = \mathbf{x}'\mathbf{I}\mathbf{x} = \mathbf{x}'\mathbf{x}$$

0.1.9 Matrix Rank

An arbitrary matrix, $\mathbf{A}$, of order $U \times V$ can be written either in terms of its U rows, say, $\mathbf{r}'_1, \mathbf{r}'_2, \ldots, \mathbf{r}'_U$ or its V columns, $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_V$, where

$$\mathbf{r}'_u = \begin{pmatrix} a_{u1} & \cdots & a_{uV} \end{pmatrix};\ \mathbf{c}_v = \begin{pmatrix} a_{1v} \\ \vdots \\ a_{Uv} \end{pmatrix}$$

and

$$\mathbf{A}_{U\times V} = \begin{pmatrix} \mathbf{r}'_1 \\ \mathbf{r}'_2 \\ \vdots \\ \mathbf{r}'_U \end{pmatrix} = \begin{pmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_V \end{pmatrix}$$

The maximum number of linearly independent rows of $\mathbf{A}$ and the maximum number of linearly independent columns is the same; this common number is defined to be the *rank* of $\mathbf{A}$. A matrix is said to be of *full rank* is the rank is equal to the minimum of U and V.

Matrix rank has a number of useful properties:

(A) $\mathbf{A}$ and $\mathbf{A}'$ have the same rank;

(B) $\mathbf{A}'\mathbf{A}$, $\mathbf{A}\mathbf{A}'$, and $\mathbf{A}$ have the same rank;

(C) the rank of a matrix is unchanged by a pre- or postmultiplication by a nonsingular matrix;

(D) the rank of a matrix is unchanged by what are called elementary row and column operations: (a) interchange of two rows or two columns; (2) multiplication or a row or a column by a scalar; (3) addition of a row (or column) to another row (or column). This is true because any elementary operation can be represented by a premultiplication (if the operation is to be on rows) or a postmultiplication (if the operation is to be on columns) of a nonsingular matrix.

To give a simple example, suppose we wish to perform some elementary row and column operations on the matrix

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 2 \\ 3 & 2 & 4 \end{pmatrix}$$

To interchange the first two rows of this latter matrix, interchange the first two rows of an identity matrix and premultiply; for the first two columns to be interchanged, carry out the operation on the identity and post-multiply:

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 2 \\ 3 & 2 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 2 \\ 1 & 1 & 1 \\ 3 & 2 & 4 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 2 \\ 3 & 2 & 4 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 2 & 3 & 4 \end{pmatrix}$$

To multiply a row of our example matrix (e.g., the second row by 5), multiply the desired row of an identity matrix and premultiply; for multiplying a specific column (e.g., the second column by 5), carry out the operation of the identity and post-multiply:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 2 \\ 3 & 2 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 5 & 0 & 10 \\ 3 & 2 & 4 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 2 \\ 3 & 2 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 5 & 1 \\ 1 & 0 & 2 \\ 3 & 10 & 4 \end{pmatrix}$$

To add one row to a second (e.g., the first row to the second), carry out the operation on the identity and premultiply; to add one column to a second (e.g., the first column to the second), carry out the operation of the identity and post-multiply:

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 2 \\ 3 & 2 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & 3 \\ 3 & 2 & 4 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 2 \\ 3 & 2 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 2 \\ 3 & 5 & 4 \end{pmatrix}$$

In general, by performing elementary row and column operations, any $U \times V$ matrix can be reduced to a *canonical form*:

$$\begin{pmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix}$$

The rank of a matrix can then be found by counting the number of ones in the above matrix.

Given an $U \times V$ matrix, $\mathbf{A}$, there exist s nonsingular elementary row operation matrices, $\mathbf{R}_1, \ldots, \mathbf{R}_s$, and t nonsingular elementary column operation matrices, $\mathbf{C}_1, \ldots, \mathbf{C}_t$ such that $\mathbf{R}_s \cdots \mathbf{R}_1 \mathbf{A} \mathbf{C}_1 \cdots \mathbf{C}_t$ is in canonical form. Moreover, if $\mathbf{A}$ is square ($U \times U$) and of full rank (i.e., $\det(\mathbf{A}) \neq 0$), then there are s nonsingular elementary row

operation matrices, $\mathbf{R}_1, \ldots, \mathbf{R}_s$, and t nonsingular elementary column operation matrices, $\mathbf{C}_1, \ldots, \mathbf{C}_t$, such that $\mathbf{R}_s \cdots \mathbf{R}_1 \mathbf{A} = \mathbf{I}$ or $\mathbf{A}\mathbf{C}_1 \cdots \mathbf{C}_t = \mathbf{I}$. Thus, $\mathbf{A}^{-1}$ can be found either as $\mathbf{R}_s \cdots \mathbf{R}_1$ or as $\mathbf{C}_1 \cdots \mathbf{C}_t$. In fact, a common way in which an inverse is calculated "by hand" starts with both $\mathbf{A}$ and $\mathbf{I}$ on the same sheet of paper; when reducing $\mathbf{A}$ step-by-step, the same operations are then applied to $\mathbf{I}$, building up the inverse until the canonical form is reached in the reduction of $\mathbf{A}$.

0.1.10 Using Matrices to Solve Equations

Suppose we have a set of U equations in V unknowns:

$$\begin{array}{ccccccc} a_{11}x_1 & + & \cdots & + & a_{1V}x_1 & = & c_1 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots \\ a_{U1}x_1 & + & \cdots & + & a_{UV}x_V & = & c_U \end{array}$$

If we let

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1V} \\ \vdots & \ddots & \vdots \\ a_{U1} & \cdots & a_{UV} \end{pmatrix}; \ \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_V \end{pmatrix}; \ \mathbf{c} = \begin{pmatrix} c_1 \\ \vdots \\ c_U \end{pmatrix}$$

then the equations can be written as follows: $\mathbf{A}_{U\times V}\mathbf{x}_{V\times 1} = \mathbf{c}_{U\times 1}$. In the simplest instance, $\mathbf{A}$ is square and nonsingular, implying that a solution may be given simply as $\mathbf{x} = \mathbf{A}^{-1}\mathbf{c}$. If there are fewer (say, $S \leq \min(U, V)$ linearly independent) equations than unknowns (so, S is the rank of $\mathbf{A}$), then we can solve for S unknowns in terms of the constants $c_1, \ldots, c_U$ and the remaining $V - S$ unknowns. We will see how this works in our discussion of obtaining eigenvectors that correspond to certain eigenvalues in a section to follow. Generally,

the set of equations is said to be *consistent* if a solution exists, i.e., a linear combination of the column vectors of $\mathbf{A}$ can be used to define $\mathbf{c}$:

$$x_1 \begin{pmatrix} a_{11} \\ \vdots \\ a_{U1} \end{pmatrix} + \cdots + x_V \begin{pmatrix} a_{1V} \\ \vdots \\ a_{UV} \end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_U \end{pmatrix}$$

or the augmented matrix $(\mathbf{A}\ \mathbf{c})$ has the same rank as $\mathbf{A}$; otherwise no solution exists and the system of equations is said to be *inconsistent*.

0.1.11 Quadratic Forms

Suppose $\mathbf{A}_{U\times U}$ is symmetric and let $\mathbf{x}' = (x_1, \ldots, x_U)$. A *quadratic form* is defined by

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{u=1}^{U}\sum_{v=1}^{U} a_{uv}x_u x_v =$$

$$a_{11}x_1^2 + a_{22}x_2^2 + \cdots + a_{UU}x_U^2 + 2a_{12}x_1x_2 + \cdots + 2a_{1U}x_1x_U + \cdots + 2a_{(U-1)U}x_{U-1}x_U$$

For example, $\Sigma_{u=1}^{U}(x_u - \bar{x})^2$, where $\bar{x}$ is the mean of the entries in $\mathbf{x}$, is a quadratic form since it can be written as

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_U \end{pmatrix}' \begin{pmatrix} (U-1)/U & -1/U & \cdots & -1/U \\ -1/U & (U-1)/U & \cdots & -1/U \\ \vdots & \vdots & \ddots & \vdots \\ -1/U & -1/U & \cdots & (U-1)/U \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_U \end{pmatrix}$$

Because of the ubiquity of sum-of-squares in statistics, it should be no surprise that quadratic forms play a central role in multivariate analysis.

A symmetric matrix $\mathbf{A}$ (and associated quadratic form) are called *positive definite*(p.d.) if $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$ (the zero vector);

if $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x}$, then $\mathbf{A}$ is *positive semi-definite*(p.s.d). We could have negative definite, negative semi-definite, and indefinite forms as well. Note that a correlation or covariance matrix is at least positive semi-definite, and satisfies the stronger condition of being positive definite if the vectors of the variables on which the correlation or covariance matrix is based, are linearly independent.

0.1.12 Multiple Regression

One of the most common topics in any beginning statistics class is *multiple regression* that we now formulate (in matrix terms) as the relation between a dependent random variable Y and a collection of K independent variables, $X_1, X_2, \ldots, X_K$. Suppose we have N subjects on which we observe Y, and arrange these values into an $N \times 1$ vector:

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{pmatrix}$$

The observations on the K independent variables are also placed in vectors:

$$\mathbf{X}_1 = \begin{pmatrix} X_{11} \\ X_{21} \\ \vdots \\ X_{N1} \end{pmatrix};\ \mathbf{X}_2 = \begin{pmatrix} X_{12} \\ X_{22} \\ \vdots \\ X_{N2} \end{pmatrix};\ \ldots;\ \mathbf{X}_K = \begin{pmatrix} X_{1K} \\ X_{2K} \\ \vdots \\ X_{NK} \end{pmatrix}$$

It would be simple if the vector $\mathbf{Y}$ were linearly dependent on $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_K$ since then

$$\mathbf{Y} = b_1\mathbf{X}_1 + b_2\mathbf{X}_2 + \cdots + b_K\mathbf{X}_K$$

for some values $b_1, \ldots, b_K$. We could always write for *any* values of $b_1, \ldots, b_K$:

$$\mathbf{Y} = b_1\mathbf{X}_1 + b_2\mathbf{X}_2 + \cdots + b_K\mathbf{X}_K + \mathbf{e}$$

where

$$\mathbf{e} = \begin{pmatrix} e_1 \\ \vdots \\ e_N \end{pmatrix}$$

is an error vector. To formulate our task as an optimization problem (least-squares), we wish to find a good set of weights, $b_1, \ldots, b_K$, so the length of $\mathbf{e}$ is minimized, i.e., $\mathbf{e}'\mathbf{e}$ is made as small as possible.

As notation, let

$$\mathbf{Y}_{N\times 1} = \mathbf{X}_{N\times K}\mathbf{b}_{K\times 1} + \mathbf{e}_{N\times 1}$$

where

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 & \ldots & \mathbf{X}_K \end{pmatrix};\ \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_K \end{pmatrix}$$

To minimize $\mathbf{e}'\mathbf{e} = (\mathbf{Y} - \mathbf{X}\mathbf{b})'(\mathbf{Y} - \mathbf{X}\mathbf{b})$, we use the vector $\mathbf{b}$ that satisfies what are called the normal equations:

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y}$$

If $\mathbf{X}'\mathbf{X}$ is nonsingular (i.e., $\det(\mathbf{X}'\mathbf{X}) \neq 0$; or equivalently, $\mathbf{X}_1, \ldots, \mathbf{X}_K$ are linearly independent), then

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

The vector that is "closest" to $\mathbf{Y}$ in our least-squares sense, is $\mathbf{Xb}$; this is a linear combination of the columns of $\mathbf{X}$ (or in other jargon, $\mathbf{Xb}$ defines the *projection* of $\mathbf{Y}$ into the space defined by (all linear combinations of) the columns of $\mathbf{X}$.

In statistical uses of multiple regression, the estimated variance-covariance matrix of the regression coefficients, $b_1, \ldots, b_K$, is given as $(\frac{1}{N-K})\mathbf{e}'\mathbf{e}(\mathbf{X}'\mathbf{X})^{-1}$, where $(\frac{1}{N-K})\mathbf{e}'\mathbf{e}$ is an (unbiased) estimate of the error variance for the distribution from which the errors are assumed drawn. Also, in multiple regression instances that usually involve an additive constant, the latter is obtained from a weight attached to an independent variable defined to be identically one.

In multivariate multiple regression where there are, say, T dependent variables (each represented by an $N \times 1$ vector), the dependent vectors are merely concatenated together into an $N \times T$ matrix, $\mathbf{Y}_{N\times T}$; the solution to the normal equations now produces a matrix $\mathbf{B}_{K\times T} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ of regression coefficients. In effect, this general expression just uses each of the dependent variables separately and adjoins all the results.

0.2 Eigenvectors and Eigenvalues

Suppose we are given a square matrix, $\mathbf{A}_{U\times U}$, and consider the polynomial $\det(\mathbf{A} - \lambda\mathbf{I})$ in the unknown value λ, referred to as Laplace's expansion:

$$\det(\mathbf{A} - \lambda\mathbf{I}) = (-\lambda)^U + S_1(-\lambda)^{U-1} + \cdots + S_{U-1}(-\lambda)^{-1} + S_U(-\lambda)^0$$

where S_u is the sum of all $u \times u$ principal minor determinants. A *principal* minor determinant is obtained from a submatrix formed from $\mathbf{A}$ that has u diagonal elements left in it. Thus, S_1 is the trace of $\mathbf{A}$ and S_U is the determinant.

There are U roots, $\lambda_1, \ldots, \lambda_U$, of the equation $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$, given that the left-hand-side is a U^{th} degree polynomial. The roots are called the *eigenvalues* of $\mathbf{A}$. There are a number of properties of eigenvalues that prove generally useful:

(A) $\det \mathbf{A} = \Pi_{u=1}^{U} \lambda_u$; $\text{trace}(\mathbf{A}) = \Sigma_{u=1}^{U} \lambda_u$;

(B) if $\mathbf{A}$ is symmetric with real elements, then all λ_u are real;

(C) if $\mathbf{A}$ is positive definite, then all λ_u are positive (strictly greater than zero); if $\mathbf{A}$ is positive semi-definite, then all λ_u are nonnegative (greater than or equal to zero);

(D) if $\mathbf{A}$ is symmetric and positive semi-definite with rank R, then there are R positive roots and $U - R$ zero roots;

(E) the nonzero roots of $\mathbf{AB}$ are equal to those of $\mathbf{BA}$; thus, the trace of $\mathbf{AB}$ is equal to the trace of $\mathbf{BA}$;

(F) eigenvalues of a diagonal matrix are the diagonal elements themselves;

(G) for any $U \times V$ matrix $\mathbf{B}$, the ranks of $\mathbf{B}$, $\mathbf{B}'\mathbf{B}$, and $\mathbf{B}\mathbf{B}'$ are all the same. Thus, because $\mathbf{B}'\mathbf{B}$ (and $\mathbf{B}\mathbf{B}'$) are symmetric and positive semi-definite (i.e., $\mathbf{x}'(\mathbf{B}'\mathbf{B})\mathbf{x} \geq 0$ because $(\mathbf{B}\mathbf{x})'(\mathbf{B}\mathbf{x})$ is a sum-of-squares which is always nonnegative), we can use (D) to find the rank of $\mathbf{B}$ by counting the positive roots of $\mathbf{B}'\mathbf{B}$.

We carry through a small example below:

$$\mathbf{A} = \begin{pmatrix} 7 & 0 & 1 \\ 0 & 7 & 2 \\ 1 & 2 & 3 \end{pmatrix}$$

$$S_1 = \text{trace}(\mathbf{A}) = 17$$

$$S_2 = \det(\begin{pmatrix} 7 & 0 \\ 0 & 7 \end{pmatrix}) + \det(\begin{pmatrix} 7 & 1 \\ 1 & 3 \end{pmatrix}) + \det(\begin{pmatrix} 7 & 2 \\ 2 & 3 \end{pmatrix}) = 49 + 20 + 17 = 86$$

$$S_3 = \det(\mathbf{A}) = 147 + 0 + 0 - 7 - 28 - 0 = 112$$

Thus,

$$\det(\mathbf{A} - \lambda\mathbf{I}) = (-\lambda)^3 + 17(-\lambda)^2 + 86(-\lambda)^1 + 112 =$$

$$-\lambda^3 + 17\lambda^2 - 86\lambda + 112 = -(\lambda - 2)(\lambda - 8)(\lambda - 7) = 0$$

which gives roots of 2, 8, and 7.

If λ_u is an eigenvalue of $\mathbf{A}$, then the equations $[\mathbf{A} - \lambda_u\mathbf{I}]\mathbf{x}_u = 0$ have a nontrivial solution (i.e., the determinant of $\mathbf{A} - \lambda_u\mathbf{I}$ vanishes, and so the inverse of $\mathbf{A} - \lambda_u\mathbf{I}$ does not exist). The solution is called an *eigenvector* (associated with the corresponding eigenvalue), and can be characterized by the following condition:

$$\mathbf{A}\mathbf{x}_u = \lambda_u\mathbf{x}_u$$

An eigenvector is determined up to a scale factor only, so typically we normalize to unit length (which then gives a $\pm$ option for the two possible unit length solutions).

We continue our simple example and to find the corresponding eigenvalues: when $\lambda = 2$, we have the equations (for $[\mathbf{A} - \lambda\mathbf{I}]\mathbf{x} = \mathbf{0}$)

$$\begin{pmatrix} 5 & 0 & 1 \\ 0 & 5 & 2 \\ 1 & 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

with an arbitrary solution of

$$\begin{pmatrix} -\frac{1}{5}a \\ -\frac{2}{5}a \\ a \end{pmatrix}$$

Choosing a to be $+\frac{5}{\sqrt{30}}$ to obtain one of the two possible normalized solutions, we have as our final eigenvector for $\lambda = 2$:

$$\begin{pmatrix} -\frac{1}{\sqrt{30}} \\ -\frac{2}{\sqrt{30}} \\ \frac{5}{\sqrt{30}} \end{pmatrix}$$

For $\lambda = 7$ we will use the normalized eigenvector of

$$\begin{pmatrix} -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \\ 0 \end{pmatrix}$$

and for $\lambda = 8$,

$$\begin{pmatrix} \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{pmatrix}$$

One of the interesting properties of eigenvalues/eigenvectors for a symmetric matrix $\mathbf{A}$ is that if λ_u and λ_v are distinct eigenvalues,

then the corresponding eigenvectors, $\mathbf{x}_u$ and $\mathbf{x}_v$, are orthogonal (i.e., $\mathbf{x}_u'\mathbf{x}_v = 0$). We can show this in the following way: the defining conditions of

$$\mathbf{A}\mathbf{x}_u = \lambda_u \mathbf{x}_u$$

$$\mathbf{A}\mathbf{x}_v = \lambda_v \mathbf{x}_v$$

lead to

$$\mathbf{x}_v'\mathbf{A}\mathbf{x}_u = \mathbf{x}_v'\lambda_u \mathbf{x}_u$$

$$\mathbf{x}_u'\mathbf{A}\mathbf{x}_v = \mathbf{x}_u'\lambda_v \mathbf{x}_v$$

Because $\mathbf{A}$ is symmetric and the left-hand-sides of these two expressions are equal (they are one-by-one matrices and equal to their own transposes), the right-hand-sides must also be equal. Thus,

$$\mathbf{x}_v'\lambda_u \mathbf{x}_u = \mathbf{x}_u'\lambda_v \mathbf{x}_v \Rightarrow$$

$$\mathbf{x}_v'\mathbf{x}_u\lambda_u = \mathbf{x}_u'\mathbf{x}_v\lambda_v$$

Due to the equality of $\mathbf{x}_v'\mathbf{x}_u$ and $\mathbf{x}_u'\mathbf{x}_v$, and by assumption, $\lambda_u \neq \lambda_v$, the inner product $\mathbf{x}_v'\mathbf{x}_u$ must be zero for the last displayed equality to hold.

In summary of the above discussion, for every real symmetric matrix $\mathbf{A}_{U\times U}$, there exists an orthogonal matrix $\mathbf{P}$ (i.e., $\mathbf{P}'\mathbf{P} = \mathbf{P}\mathbf{P}' = \mathbf{I}$) such that $\mathbf{P}'\mathbf{A}\mathbf{P} = \mathbf{D}$, where $\mathbf{D}$ is a diagonal matrix containing the eigenvalues of $\mathbf{A}$, and

$$\mathbf{P} = \begin{pmatrix} \mathbf{p}_1 & \dots & \mathbf{p}_U \end{pmatrix}$$

where $\mathbf{p}_u$ is a normalized eigenvector associated with λ_u for $1 \leq u \leq U$. If the eigenvalues are not distinct, it is still possible to choose the eigenvectors to be orthogonal. Finally, because $\mathbf{P}$ is an orthogonal matrix (and $\mathbf{P'AP} = \mathbf{D} \Rightarrow \mathbf{PP'APP'} = \mathbf{PDP'}$), we can finally represent $\mathbf{A}$ as

$$\mathbf{A} = \mathbf{PDP'}$$

In terms of the small numerical example being used, we have for $\mathbf{P'AP} = \mathbf{D}$:

$$\begin{pmatrix} -\frac{1}{\sqrt{30}} & -\frac{2}{\sqrt{30}} & \frac{5}{\sqrt{30}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix} \begin{pmatrix} 7 & 0 & 1 \\ 0 & 7 & 2 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} -\frac{1}{\sqrt{30}} & -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{30}} & \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{6}} \\ \frac{5}{\sqrt{30}} & 0 & \frac{1}{\sqrt{6}} \end{pmatrix} =$$

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 8 \end{pmatrix}$$

and for $\mathbf{PDP'} = \mathbf{A}$:

$$\begin{pmatrix} -\frac{1}{\sqrt{30}} & -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{30}} & \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{6}} \\ \frac{5}{\sqrt{30}} & 0 & \frac{1}{\sqrt{6}} \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 8 \end{pmatrix} \begin{pmatrix} -\frac{1}{\sqrt{30}} & -\frac{2}{\sqrt{30}} & \frac{5}{\sqrt{30}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix} =$$

$$\begin{pmatrix} 7 & 0 & 1 \\ 0 & 7 & 2 \\ 1 & 2 & 3 \end{pmatrix}$$

The representation of $\mathbf{A}$ as $\mathbf{PDP'}$ leads to several rather nice computational "tricks". First, if $\mathbf{A}$ is p.s.d., we can define

$$\mathbf{D}^{1/2} \equiv \begin{pmatrix} \sqrt{\lambda_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sqrt{\lambda_U} \end{pmatrix}$$

and represent $\mathbf{A}$ as

$$\mathbf{A} = \mathbf{P}\mathbf{D}^{1/2}\mathbf{D}^{1/2}\mathbf{P}' = \mathbf{P}\mathbf{D}^{1/2}(\mathbf{P}\mathbf{D}^{1/2})' = \mathbf{L}\mathbf{L}', \text{say.}$$

In other words, we have "factored" $\mathbf{A}$ into $\mathbf{L}\mathbf{L}'$, for

$$\mathbf{L} = \mathbf{P}\mathbf{D}^{1/2} = \begin{pmatrix} \sqrt{\lambda_1}\mathbf{p}_1 & \sqrt{\lambda_2}\mathbf{p}_2 & \dots & \sqrt{\lambda_U}\mathbf{p}_U \end{pmatrix}$$

Secondly, if $\mathbf{A}$ is p.d., we can define

$$\mathbf{D}^{-1} \equiv \begin{pmatrix} \frac{1}{\lambda_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\lambda_U} \end{pmatrix}$$

and represent $\mathbf{A}^{-1}$ as

$$\mathbf{A}^{-1} = \mathbf{P}\mathbf{D}^{-1}\mathbf{P}'$$

To verify,

$$\mathbf{A}\mathbf{A}^{-1} = (\mathbf{P}\mathbf{D}\mathbf{P}')(\mathbf{P}\mathbf{D}^{-1}\mathbf{P}') = \mathbf{I}$$

Thirdly, to define a "square root" matrix, let $\mathbf{A}^{1/2} \equiv \mathbf{P}\mathbf{D}^{1/2}\mathbf{P}'$. To verify, $\mathbf{A}^{1/2}\mathbf{A}^{1/2} = \mathbf{P}\mathbf{D}\mathbf{P}' = \mathbf{A}$.

There is a generally interesting way to represent the multiplication of two matrices considered as collections of column and row vectors, respectively, where the final answer is a sum of outer products of vectors. This view will prove particularly useful in our discussion of

principal component analysis. Suppose we have two matrices $\mathbf{B}_{U\times V}$, represented as a collection of its V columns:

$$\mathbf{B} = \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_V \end{pmatrix}$$

and $\mathbf{C}_{V\times W}$, represented as a collection of its V rows:

$$\mathbf{C} = \begin{pmatrix} \mathbf{c}_1' \\ \mathbf{c}_2' \\ \vdots \\ \mathbf{c}_V' \end{pmatrix}$$

The product $\mathbf{BC} = \mathbf{D}$ can be written as

$$\mathbf{BC} = \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_V \end{pmatrix} \begin{pmatrix} \mathbf{c}_1' \\ \mathbf{c}_2' \\ \vdots \\ \mathbf{c}_V' \end{pmatrix} =$$

$$\mathbf{b}_1\mathbf{c}_1' + \mathbf{b}_2\mathbf{c}_2' + \cdots + \mathbf{b}_V\mathbf{c}_V' = \mathbf{D}$$

As an example, consider the *spectral decomposition* of $\mathbf{A}$ considered above as $\mathbf{PDP}'$, and where from now on, without loss of any generality, the diagonal entries in $\mathbf{D}$ are ordered as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_U$. We can represent $\mathbf{A}$ as

$$\mathbf{A}_{U\times U} = \begin{pmatrix} \sqrt{\lambda_1}\mathbf{p}_1 & \dots & \sqrt{\lambda_U}\mathbf{p}_U \end{pmatrix} \begin{pmatrix} \sqrt{\lambda_1}\mathbf{p}_1' \\ \vdots \\ \sqrt{\lambda_U}\mathbf{p}_U' \end{pmatrix} =$$

$$\lambda_1\mathbf{p}_1\mathbf{p}_1' + \cdots + \lambda_U\mathbf{p}_U\mathbf{p}_U'$$

If $\mathbf{A}$ is p.s.d. and of rank R, then the above sum obviously stops at R components. In general, the matrix $\mathbf{B}_{U\times U}$ that is a rank K ($\leq R$)

least-squares approximation to $\mathbf{A}$ can be given by

$$\mathbf{B} = \lambda_1 \mathbf{p}_1 \mathbf{p}_1' + \cdots + \lambda_k \mathbf{p}_K \mathbf{p}_K'$$

and the value of the loss function:

$$\sum_{v=1}^{U} \sum_{u=1}^{U} (a_{uv} - b_{uv})^2 = (\lambda_{K+1}^2 + \cdots + \lambda_U^2)^{\frac{1}{2}}$$

0.3 The Singular Value Decomposition of a Matrix

The *singular value decomposition* (SVD) or the *basic structure* of a matrix refers to the representation of *any* rectangular $U \times V$ matrix, say, $\mathbf{A}$, as a triple product:

$$\mathbf{A}_{U \times V} = \mathbf{P}_{U \times R} \mathbf{\Delta}_{R \times R} \mathbf{Q}'_{R \times V}$$

where the R columns of $\mathbf{P}$ are orthonormal; the R rows of $\mathbf{Q}'$ are orthonormal; $\mathbf{\Delta}$ is diagonal with ordered positive entries, $\delta_1 \geq \delta_2 \geq \cdots \geq \delta_R > 0$; and R is the rank of $\mathbf{A}$. Or, alternatively, we can "fill up" this decomposition as

$$\mathbf{A}_{U \times V} = \mathbf{P}^*_{U \times U} \mathbf{\Delta}^*_{U \times V} \mathbf{Q}^{*'}_{V \times V}$$

where the columns of $\mathbf{P}^*$ and rows of $\mathbf{Q}^{*'}$ are still orthonormal, and the diagonal matrix $\mathbf{\Delta}$ forms the upper-left-corner of $\mathbf{\Delta}^*$:

$$\mathbf{\Delta}^* = \begin{pmatrix} \mathbf{\Delta} & \emptyset \\ \emptyset & \emptyset \end{pmatrix}$$

here, $\emptyset$ represents an appropriately dimensioned matrix of all zeros. In analogy to the least-squares result of the last section, if a rank K ($\leq R$) matrix approximation to $\mathbf{A}$ is desired, say $\mathbf{B}_{U \times V}$, the first K ordered entries in $\mathbf{\Delta}$ are taken:

$$\mathbf{B} = \delta_1 \mathbf{p}_1 \mathbf{q}_1' + \cdots + \delta_K \mathbf{p}_K \mathbf{q}_K'$$

and the value of the loss function:

$$\sum_{v=1}^{V} \sum_{u=1}^{U} (a_{uv} - b_{uv})^2 = \delta_{K+1}^2 + \cdots + \delta_R^2$$

This latter result of approximating one matrix (least-squares) by another of lower rank, is referred to as the Ekart-Young theorem in the psychometric literature.

Once one has the SVD of a matrix, a lot of representation needs can be expressed in terms of it. For example, suppose $\mathbf{A} = \mathbf{P}\boldsymbol{\Delta}\mathbf{Q}'$; the spectral decomposition of $\mathbf{A}\mathbf{A}'$ can then be given as

$$(\mathbf{P}\boldsymbol{\Delta}\mathbf{Q}')(\mathbf{P}\boldsymbol{\Delta}\mathbf{Q}')' = \mathbf{P}\boldsymbol{\Delta}\mathbf{Q}'\mathbf{Q}\boldsymbol{\Delta}\mathbf{P}' = \mathbf{P}\boldsymbol{\Delta}\boldsymbol{\Delta}\mathbf{P}' = \mathbf{P}\boldsymbol{\Delta}^2\mathbf{P}'$$

Similarly, the spectral decomposition of $\mathbf{A}'\mathbf{A}$ is expressible as $\mathbf{Q}\boldsymbol{\Delta}^2\mathbf{Q}'$.

0.4 Common Multivariate Methods in Matrix Terms

In this section we give a very brief overview of some common methods of multivariate analysis in terms of the matrix ideas we have introduced thus far in this chapter. Later chapters (if they ever get writtten) will come back to these topics and develop them in more detail.

0.4.1 Principal Components

Suppose we have a data matrix $\mathbf{X}_{N \times P} = \{x_{ij}\}$, with x_{ij} referring as usual to the observation for subject i on variable or column j:

$$\mathbf{X}_{N\times P} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1P} \\ x_{21} & x_{22} & \cdots & x_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NP} \end{pmatrix}$$

The columns can be viewed as containing N observations on each of P random variables that we denote generically by $X_1, X_2, \ldots, X_P$. We let $\mathbf{A}$ denote the $P \times P$ sample covariance matrix obtained among the variables from $\mathbf{X}$, and let $\lambda_1 \geq \cdots \geq \lambda_P \geq 0$ be its P eigenvalues and $\mathbf{p}_1, \ldots, \mathbf{p}_P$ the corresponding normalized eigenvectors. Then, the linear combination

$$\mathbf{p}_k' \begin{pmatrix} X_1 \\ \vdots \\ X_P \end{pmatrix}$$

is called the k^{th} (sample) *principal component.*

There are (at least) two interesting properties of principal components to bring up at this time:

A) The k^{th} principal component has maximum variance among all linear combinations defined by unit length vectors orthogonal to $\mathbf{p}_1, \ldots, \mathbf{p}_{k-1}$; also, it is uncorrelated with the components up to $k-1$;

B) $\mathbf{A} \approx \lambda_1 \mathbf{p}_1 \mathbf{p}_1' + \cdots + \lambda_K \mathbf{p}_K \mathbf{p}_K'$ gives a least-squares rank K approximation to $\mathbf{A}$ (a special case of the Ekart-Young theorem for an arbitrary symmetric matrix).

0.4.2 Discriminant Analysis

Suppose we have a one-way analysis-of-variance (ANOVA) layout with J groups (n_j subjects in group j, $1 \leq j \leq J$), and P measure-

ments on each subject. If x_{ijk} denotes person i, in group j, and the observation of variable k ($1 \leq i \leq n_j$; $1 \leq j \leq J$; $1 \leq k \leq P$), then define the Between-Sum-of-Squares matrix

$$\mathbf{B}_{P\times P} = \{\sum_{j=1}^{J} n_j(\bar{x}_{\cdot jk} - \bar{x}_{\cdot\cdot k})(\bar{x}_{\cdot jk'} - \bar{x}_{\cdot\cdot k'})\}_{P\times P}$$

and the Within-Sum-of-Squares matrix

$$\mathbf{W}_{P\times P} = \{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(x_{ijk} - \bar{x}_{\cdot jk})(x_{ijk'} - \bar{x}_{\cdot jk'})\}_{P\times P}$$

For the matrix product $\mathbf{W}^{-1}\mathbf{B}$, let $\lambda_1, \ldots, \lambda_T \geq 0$ be the eigenvectors ($T = \min(P, J-1)$, and $\mathbf{p}_1, \ldots, \mathbf{p}_T$ the corresponding normalized eigenvectors. Then, the linear combination

$$\mathbf{p}_k' \begin{pmatrix} X_1 \\ \vdots \\ X_P \end{pmatrix}$$

is called the k^{th} *discriminant function*. It has the valuable property of maximizing the univariate F-ratio subject to being uncorrelated with the earlier linear combinations. A variety of applications of discriminant functions exists in classification that we will come back to later. Also, standard multivariate ANOVA significance testing is based on various functions of the eigenvalues $\lambda_1, \ldots, \lambda_T$ and their derived sampling distributions.

0.4.3 Canonical Correlation

Suppose the collection of P random variables that we have observed over the N subjects is actually in the form of two "batteries", $X_1, \ldots, X_Q$

and $X_{Q+1}, \ldots, X_P$, and the observed covariance matrix $\mathbf{A}_{P\times P}$ is partitioned into four parts:

$$\mathbf{A}_{P\times P} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}'_{12} & \mathbf{A}_{22} \end{pmatrix}$$

where $\mathbf{A}_{11}$ is $Q \times Q$ and represents the observed covariances among the variables in the first battery; $\mathbf{A}_{22}$ is $(P-Q) \times (P-Q)$ and represents the observed covariances among the variables in the second battery; $\mathbf{A}_{12}$ is $Q \times (P-Q)$ and represents the observed covariances between the variables in the first and second batteries. Consider the following two equations in unknown vectors $\mathbf{a}$ and $\mathbf{b}$, and unknown scalar λ:

$$\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}'_{12}\mathbf{a} = \lambda\mathbf{a}$$

$$\mathbf{A}_{22}^{-1}\mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{b} = \lambda\mathbf{b}$$

There are T solutions to these expressions (for $T = \min(Q, (P-Q))$), given by normalized unit-length vectors, $\mathbf{a}_1, \ldots, \mathbf{a}_T$ and $\mathbf{b}_1, \ldots, \mathbf{b}_T$; and a set of common $\lambda_1 \geq \cdots \geq \lambda_T \geq 0$.

The linear combinations of the first and second batteries defined by $\mathbf{a}_k$ and $\mathbf{b}_k$ are the k^{th} *canonical variates* and have squared correlation of λ_k; they are uncorrelated with all other canonical variates (defined either in the first or second batteries). Thus, $\mathbf{a}_1$ and $\mathbf{b}_1$ are the first canonical variates with squared correlation of λ_1; among all linear combinations defined by unit-length vectors for the variables in the two batteries, this squared correlation is the highest it can be. (We note that the coefficient matrices $\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}'_{12}$ and $\mathbf{A}_{22}^{-1}\mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$ are not symmetric; thus, special symmetrizing and

equivalent equation systems are typically used to obtain the solutions to the original set of expressions.)

0.4.4 Algebraic Restrictions on Correlations

A matrix $\mathbf{A}_{P\times P}$ that represents a covariance matrix among a collection of random variables, $X_1, \ldots, X_P$ is p.s.d.; and conversely, any p.s.d. matrix represents the covariance matrix for some collection of random variables. We partition $\mathbf{A}$ to isolate its last row and column as

$$\mathbf{A} = \begin{pmatrix} \mathbf{B}_{(P-1)\times(P-1)} & \mathbf{g}_{(P-1)\times 1} \\ \mathbf{g}' & a_{PP} \end{pmatrix}$$

$\mathbf{B}$ is the $(P-1)\times(P-1)$ covariance matrix among the variables $X_1, \ldots, X_{P-1}$; $\mathbf{g}$ is $(P-1)\times 1$ and contains the cross-covariance between the the first $P-1$ variables and the P^{th}; a_{PP} is the variance for the P^{th} variable.

Based on the observation that determinants of p.s.d. matrices are nonnegative, and a result on expressing determinants for partitioned matrices (that we do not give here), it must be true that

$$\mathbf{g}'\mathbf{B}^{-1}\mathbf{g} \leq a_{PP}$$

or if we think correlations rather than merely covariances (so the main diagonal of $\mathbf{A}$ consists of all ones):

$$\mathbf{g}'\mathbf{B}^{-1}\mathbf{g} \leq 1$$

Given the correlation matrix $\mathbf{B}$, the possible values the correlations in $\mathbf{g}$ could have are in or on the ellipsoid defined in $P-1$ dimensions by $\mathbf{g}'\mathbf{B}^{-1}\mathbf{g} \leq 1$. The important point is that we do not have a "box"

in $P-1$ dimensions containing the correlations with sides extending the whole range of ± 1; instead, some restrictions are placed on the observable correlations that gets defined by the size of the correlation in $\mathbf{B}$. For example, when $P = 3$, a correlation between variables X_1 and X_2 of $r_{12} = 0$ gives the "degenerate" ellipse of a circle for constraining the correlation values between X_1 and X_2 and the third variable X_3 (in a two-dimensional r_{13} versus r_{23} coordinate system); for $r_{12} = 1$, the ellipse flattens to a line in this same two-dimensional space.

Another algebraic restriction that can be seen immediately is based on the formula for the partial correlation between two variables, "holding the third constant":

$$\frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

Bounding the above by ± 1 (because it is a correlation) and "solving" for r_{12}, gives the algebraic upper and lower bounds of

$$r_{12} \leq r_{13}r_{23} + \sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}$$

$$r_{13}r_{23} - \sqrt{(1 - r_{13}^2)(1 - r_{23}^2)} \leq r_{12}$$

0.4.5 The Biplot

Let $\mathbf{A} = \{a_{ij}\}$ be an $n \times m$ matrix of rank r. We wish to find a second matrix $\mathbf{B} = \{b_{ij}\}$ of the same size, $n \times m$, but of rank t, where $t \leq r$, such that the least squares criterion, $\Sigma_{i,j}(a_{ij} - b_{ij})^2$, is as small as possible overall all matrices of rank t.

The solution is to first find the singular value decomposition of $\mathbf{A}$ as $\mathbf{UDV}'$, where $\mathbf{U}$ is $n \times r$ and has orthonormal columns, $\mathbf{V}$ is $m \times r$ and has orthonormal columns, and $\mathbf{D}$ is $r \times r$, diagonal, with positive values $d_1 \geq d_2 \geq \cdots \geq d_r > 0$ along the main diagonal. Then, $\mathbf{B}$ is defined as $\mathbf{U}^*\mathbf{D}^*\mathbf{V}^{*\prime}$, where we take the first t columns of $\mathbf{U}$ and $\mathbf{V}$ to obtain $\mathbf{U}^*$ and $\mathbf{V}^*$, respectively, and the first t values, $d_1 \geq \cdots \geq d_t$, to form a diagonal matrix $\mathbf{D}^*$.

The approximation of $\mathbf{A}$ by a rank t matrix $\mathbf{B}$, has been one mechanism for representing the row and column objects defining $\mathbf{A}$ in a low-dimensional space of dimension t through what can be generically labeled as a biplot (the prefix "bi" refers to the representation of both the row and column objects together in the same space). Explicitly, the approximation of $\mathbf{A}$ and $\mathbf{B}$ can be written as

$$\mathbf{B} = \mathbf{U}^*\mathbf{D}^*\mathbf{V}^{*\prime} = \mathbf{U}^*\mathbf{D}^{*\alpha}\mathbf{D}^{(\mathbf{1}-\alpha)}\mathbf{V}^{*\prime} = \mathbf{PQ}' \ ,$$

where α is some chosen number between 0 and 1, $\mathbf{P} = \mathbf{U}^*\mathbf{D}^{*\alpha}$ and is $n \times t$, $\mathbf{Q} = (\mathbf{D}^{(\mathbf{1}-\alpha)}\mathbf{V}^{*\prime})'$ and is $m \times t$.

The entries in $\mathbf{P}$ and $\mathbf{Q}$ define coordinates for the row and column objects in a t-dimensional space that, irrespective of the value of α chosen, have the following characteristic:

If a vector is drawn from the origin through the i^{th} row point and the m column points are projected onto this vector, the collection of such projections is proportional to the i^{th} row of the approximating matrix $\mathbf{B}$. The same is true for projections of row points onto vectors from the origin through each of the column points.

0.4.6 The Procrustes Problem

Procrustes (the subduer), son of Poseidon, kept an inn benefiting from what he claimed to be a wonderful all-fitting bed. He lopped off excessive limbage from tall guests and either flattened short guests by hammering or stretched them by racking. The victim fitted the bed perfectly but, regrettably, died. To exclude the embarrassment of an initially exact-fitting guest, variants of the legend allow Procrustes two, different-sized beds. Ultimately, in a crackdown on robbers and monsters, the young Theseus fitted Procrustes to his own bed. (Gower and Dijksterhuis, 2004)

Suppose we have two matrices, $\mathbf{X}_1$ and $\mathbf{X}_2$, each considered (for convenience) to be of the same size, $n \times p$. If you wish, $\mathbf{X}_1$ and $\mathbf{X}_2$ can be interpreted as two separate p-dimensional coordinate sets for the same set of n objects. Our task is to match these two configurations optimally, with the criterion being least-squares: find a transformation matrix, $\mathbf{T}_{p\times p}$, such that $\parallel \mathbf{X}_1\mathbf{T} - \mathbf{X}_2 \parallel$ is minimized, where $\parallel \cdot \parallel$ denotes the sum-of-squares of the incorporated matrix, i.e., if $\mathbf{A} = \{a_{uv}\}$, then $\parallel \mathbf{A} \parallel = \text{trace}(\mathbf{A}'\mathbf{A}) = \Sigma_{u,v}\, a_{uv}^2$. For convenience, assume both $\mathbf{X}_1$ and $\mathbf{X}_2$ have been normalized so $\parallel \mathbf{X}_1 \parallel = \parallel \mathbf{X}_2 \parallel = 1$, and the columns of $\mathbf{X}_1$ and $\mathbf{X}_2$ have sums of zero.

Two results are central:

(a) When $\mathbf{T}$ is unrestricted, we have the multivariate multiple regression solution

$$\mathbf{T}^* = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1\mathbf{X}_2 \ ;$$

(b) When $\mathbf{T}$ is orthogonal, we have the Schönemann solution done for his thesis in the Quantitative Division at Illinois in 1965 (published in *Psychometrika* in 1966):

for the SVD of $\mathbf{X}_2'\mathbf{X}_1 = \mathbf{USV}'$, we let $\mathbf{T}^* = \mathbf{VU}'$.

0.4.7 Matrix Rank Reduction

Lagrange's Theorem (as inappropriately named by C. R. Rao, because it should really be attributed to Guttman) can be stated as follows:

Let $\mathbf{G}$ be a nonnegative-definite (i.e., a symmetric positive semi-definite) matrix of order $n \times n$ and of rank $r > 0$. Let $\mathbf{B}$ be of order $n \times s$ and such that $\mathbf{B}'\mathbf{GB}$ is non-singular. Then the residual matrix

$$\mathbf{G_1} = \mathbf{G} - \mathbf{GB}(\mathbf{B}'\mathbf{GB})^{-1}\mathbf{B}'\mathbf{G} \tag{1}$$

is of rank $r - s$ and is nonnegative definite.

Intuitively, this theorem allows you to "take out" "factors" from a covariance (or correlation) matrix.

0.4.8 Torgerson Metric Multidimensional Scaling

Let $\mathbf{A}$ be a symmetric matrix of order $n \times n$. Suppose we want to find a matrix $\mathbf{B}$ of rank 1 (of order $n \times n$) in such a way that the sum of the squared discrepancies between the elements of $\mathbf{A}$ and the corresponding elements of $\mathbf{B}$ (i.e., $\Sigma_{j=1}^{n}\Sigma_{i=1}^{n}(a_{ij} - b_{ij})^2$) is at a minimum. It can be shown that the solution is $\mathbf{B} = \lambda\mathbf{kk}'$ (so all columns in $\mathbf{B}$ are multiples of $\mathbf{k}$), where λ is the largest eigenvalue of $\mathbf{A}$ and $\mathbf{k}$ is the corresponding normalized eigenvector. This theorem can be generalized. Suppose we take the first r largest eigenvalues and the corresponding normalized eigenvectors. The eigenvectors are collected in an $n \times r$ matrix $\mathbf{K} = \{\mathbf{k}_1, \ldots, \mathbf{k}_r\}$ and the eigenvalues in a diagonal matrix $\mathbf{\Lambda}$. Then $\mathbf{K\Lambda K}'$ is an $n \times n$ matrix of rank r and

is a least-squares solution for the approximation of $\mathbf{A}$ by a matrix of rank r. It is assumed, here, that the eigenvalues are all positive. If $\mathbf{A}$ is of rank r by itself and we take the r eigenvectors for which the eigenvalues are different from zero collected in a matrix $\mathbf{K}$ of order $n \times r$, then $\mathbf{A} = \mathbf{K\Lambda K'}$. Note that $\mathbf{A}$ could also be represented by $\mathbf{A} = \mathbf{LL'}$, where $\mathbf{L} = \mathbf{K\Lambda^{1/2}}$ (we factor the matrix), or as a sum of r $n \times n$ matrices — $\mathbf{A} = \lambda_1 \mathbf{k_1 k_1'} + \cdots + \lambda_r \mathbf{k_r k_r'}$.

Metric Multidimensional Scaling – Torgerson's Model (Gower's Principal Coordinate Analysis)

Suppose I have a set of n points that can be perfectly represented spatially in r dimensional space. The i^{th} point has coordinates $(x_{i1}, x_{i2}, \ldots, x_{ir})$. If $d_{ij} = \sqrt{\Sigma_{k=1}^{r}(x_{ik} - x_{jk})^2}$ represents the Euclidean distance between points i and j, then

$$d_{ij}^* = \sum_{k=1}^{r} x_{ik}x_{jk}, \text{where}$$

$$d_{ij}^* = -\frac{1}{2}(d_{ij}^2 - A_i - B_j + C); \tag{2}$$

$$A_i = (1/n) \sum_{j=1}^{n} d_{ij}^2;$$

$$B_j = (1/n) \sum_{i=1}^{n} d_{ij}^2;$$

$$C = (1/n^2) \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}^2.$$

Note that $\{d_{ij}^*\}_{n \times n} = \mathbf{XX'}$, where $\mathbf{X}$ is of order $n \times r$ and the entry in the i^{th} row and k^{th} column is x_{ik}.

So, the Question: If I give you $\mathbf{D} = \{d_{ij}\}_{n\times n}$, find me *a* set of coordinates to do it. The Solution: Find $\mathbf{D}^* = \{d^*_{ij}\}$, and take its Spectral Decomposition. This is *exact* here.

To use this result to obtain a spatial representation for a set of n objects given *any* "distance-like" measure, p_{ij}, between objects i and j, we proceed as follows:

(a) Assume (i.e., pretend) the Euclidean model holds for p_{ij}.

(b) Define p^*_{ij} from p_{ij} using (1).

(c) Obtain a spatial representation for p^*_{ij} using a suitable value for r, the number of dimensions (at most, r can be no larger than the number of positive eigenvalues for $\{p^*_{ij}\}_{n\times n}$):

$$\{p^*_{ij}\} \approx \mathbf{X}\mathbf{X}'$$

(d) Plot the n points in r dimensional space.

0.4.9 A Guttman Multidimensional Scaling Result

I. If $\mathbf{B}$ is a symmetric matrix of order n, having all its elements non-negative, the following quadratic form defined by the matrix $\mathbf{A}$ must be positive semi-definite:

$$\sum_{i,j} b_{ij}(x_i - x_j)^2 = \sum_{i,j} x_i a_{ij} x_j,$$

where

$$a_{ij} = \begin{cases} \Sigma^n_{k=1;k\neq i}\, b_{ik} & (i = j) \\ -b_{ij} & (i \neq j) \end{cases}$$

If all elements of $\mathbf{B}$ are positive, then $\mathbf{A}$ is of rank $n-1$, and has one smallest eigenvalue equal to zero with an associated eigenvector

having all constant elements. Because all (other) eigenvectors must be orthogonal to the constant eigenvector, the entries in these other eigenvectors must sum to zero.

This Guttman result can be used for a method of multidimensional scaling (mds), and is one that seems to get reinvented periodically in the literature. Generally, this method has been used to provide rational starting points in iteratively-defined nonmetric mds.

0.4.10 A Few General MATLAB Routines to Know About

For Eigenvector/Eigenvalue Decompositions:

$[\mathbf{V}, \mathbf{D}] = \text{eig}(\mathbf{A})$, where $\mathbf{A} = \mathbf{VDV}'$, for $\mathbf{A}$ square; $\mathbf{V}$ is orthogonal and contains eigenvectors (as columns); $\mathbf{D}$ is diagonal and contains the eigenvalues (ordered from *smallest to largest*).

For Singular Value Decompositions:

$[\mathbf{U}, \mathbf{S}, \mathbf{V}] = \text{svd}(\mathbf{B})$, where $\mathbf{B} = \mathbf{USV}'$; the columns of $\mathbf{U}$ and the rows of $\mathbf{V}'$ are orthonormal; $\mathbf{S}$ is diagonal and contains the nonnegative singular values (ordered from *largest to smallest*).

The help comments for the Procrustes routine in the Statistics Toolbox are given verbatim below. Note the very general transformation provided in the form of a MATLAB Structure that involves optimal rotation, translation, and scaling.

```
>> help procrustes
 PROCRUSTES Procrustes Analysis
    D = PROCRUSTES(X, Y) determines a linear transformation (translation,
    reflection, orthogonal rotation, and scaling) of the points in the
    matrix Y to best conform them to the points in the matrix X.  The
    "goodness-of-fit" criterion is the sum of squared errors.  PROCRUSTES
    returns the minimized value of this dissimilarity measure in D.  D is
    standardized by a measure of the scale of X, given by
```

```
   sum(sum((X - repmat(mean(X,1), size(X,1), 1)).^2, 1))

i.e., the sum of squared elements of a centered version of X.  However,
if X comprises repetitions of the same point, the sum of squared errors
is not standardized.

X and Y are assumed to have the same number of points (rows), and
PROCRUSTES matches the i'th point in Y to the i'th point in X.  Points
in Y can have smaller dimension (number of columns) than those in X.
In this case, PROCRUSTES adds columns of zeros to Y as necessary.

[D, Z] = PROCRUSTES(X, Y) also returns the transformed Y values.

[D, Z, TRANSFORM] = PROCRUSTES(X, Y) also returns the transformation
that maps Y to Z.  TRANSFORM is a structure with fields:
   c:  the translation component
   T:  the orthogonal rotation and reflection component
   b:  the scale component
That is, Z = TRANSFORM.b * Y * TRANSFORM.T + TRANSFORM.c.

Examples:

   % Create some random points in two dimensions
   n = 10;
   X = normrnd(0, 1, [n 2]);

   % Those same points, rotated, scaled, translated, plus some noise
   S = [0.5 -sqrt(3)/2; sqrt(3)/2 0.5]; % rotate 60 degrees
   Y = normrnd(0.5*X*S + 2, 0.05, n, 2);

   % Conform Y to X, plot original X and Y, and transformed Y
   [d, Z, tr] = procrustes(X,Y);
   plot(X(:,1),X(:,2),'rx', Y(:,1),Y(:,2),'b.', Z(:,1),Z(:,2),'bx');

   % Compute a procrustes solution that does not include scaling:
   trUnscaled.T = tr.T;
   trUnscaled.b = 1;
   trUnscaled.c = mean(X) - mean(Y) * trUnscaled.T;
   ZUnscaled = Y * trUnscaled.T + repmat(trUnscaled.c,n,1);
   dUnscaled = sum((ZUnscaled(:)-X(:)).^2) ...
                       / sum(sum((X - repmat(mean(X,1),n,1)).^2, 1));
```