Two Purposes for Matrix Factorization: A Historical Appraisal*

Lawrence Hubert[†]
Jacqueline Meulman[‡]
Willem Heiser[§]

Abstract. Matrix factorization in numerical linear algebra (NLA) typically serves the purpose of restating some given problem in such a way that it can be solved more readily; for example, one major application is in the solution of a linear system of equations. In contrast, within applied statistics/psychometrics (AS/P), a much more common use for matrix factorization is in presenting, possibly spatially, the structure that may be inherent in a given data matrix obtained on a collection of objects observed over a set of variables. The actual components of a factorization are now of prime importance and not just as a mechanism for solving another problem. We review some connections between NLA and AS/P and their respective concerns with matrix factorization and the subsequent rank reduction of a matrix. We note in particular that several results available for many decades in AS/P were more recently (re)discovered in the NLA literature. Two other distinctions between NLA and AS/P are also discussed briefly: how a generalized singular value decomposition might be defined, and the differing uses for the (newer) methods of optimization based on cyclic or iterative projections.

Key words. rank reduction, matrix factorization, matrix decomposition, singular value decomposition, cyclic projection

AMS subject classifications. 62H25, 65F15

PII. S0036144598340483

1. Introduction. Matrix factorization in the context of numerical linear algebra (NLA) generally serves the purpose of rephrasing through a series of easier subproblems a task that may be relatively difficult to solve in its original form. For example, given the typical linear system $Ax = b$ for $A \in R^{n \times n}$, x and $b \in R^n$, a factorization of A as LU for L a unit lower triangular matrix (thus, with ones along its main diagonal) and an upper triangular U is a mechanism for characterizing what occurs in Gaussian elimination. We replace $Ax = b$ by two (easier to solve) triangular systems: find y so $Ly = b$ and then find x so $Ux = y$. The point being made here is that the factorization of A as LU has no real importance in and of itself other than as a computationally convenient means for obtaining a solution to the original linear system.

In contrast to its usage as a mechanism for obtaining another end, within the field of applied statistics/psychometrics (AS/P), matrix factorization also plays a very major role but usually not just for the purpose of solving systems of equations (although many exemplars for that specific application exist as well). Typically, a matrix $A \in R^{n \times p}$ represents a data matrix containing numerical observations on n objects (subjects) over p attributes (variables), or possibly $B \in R^{p \times p}$ and the entries are some measure of proximity between attributes, such as the correlation between columns of A.[1] The major purpose of a matrix factorization in this context is to obtain some form of lower-rank (and therefore simplified) approximation to A (or possibly to B) for understanding the structure of the data matrix, particularly the relationship within the objects and within the attributes, and how the objects relate to the attributes. If we can further interpret the matrix factorization geometrically and actually present the results spatially through coordinates obtained from the components of the factorization, we will be able to better communicate to others what structure may be present in the original data matrix. In any case, matrix factorizations are again directed toward the issue of simplicity, but now the actual components making up a factorization are of prime concern and not solely as a mechanism for solving another problem.

An Example.   To give a brief introductory example of how matrix factorization might be used for understanding the structure inherent in a data matrix, a (small) 8×8 correlation matrix, **B**, is presented below among eight physical variables measured for 305 girls. The rows and columns of **B** correspond in numerical order to the variables: height (1), arm span (2), length of forearm (3), length of lower leg (4), weight (5), bitrochanteric diameter (6), chest girth (7), and chest width (8). This matrix has been used repeatedly in the literature to illustrate a range of factor-analytic methods and is based on raw data originally collected in the 1930s. We obtained the correlation matrix from the classic factor analysis text by Harman [21, pp. 81–84].

$$
\mathbf{B} = \begin{bmatrix}
1.00 & 0.85 & 0.81 & 0.86 & 0.47 & 0.40 & 0.30 & 0.38 \\
0.85 & 1.00 & 0.88 & 0.83 & 0.38 & 0.33 & 0.28 & 0.41 \\
0.81 & 0.88 & 1.00 & 0.80 & 0.38 & 0.32 & 0.24 & 0.34 \\
0.86 & 0.83 & 0.80 & 1.00 & 0.44 & 0.33 & 0.33 & 0.36 \\
0.47 & 0.38 & 0.38 & 0.44 & 1.00 & 0.76 & 0.73 & 0.63 \\
0.40 & 0.33 & 0.32 & 0.33 & 0.76 & 1.00 & 0.58 & 0.58 \\
0.30 & 0.28 & 0.24 & 0.33 & 0.73 & 0.58 & 1.00 & 0.54 \\
0.38 & 0.41 & 0.34 & 0.36 & 0.63 & 0.58 & 0.54 & 1.00
\end{bmatrix}.
$$

The best (in a least-squares sense) rank-2 approximation to **B** is given below and is based on retaining the largest two eigenvalues (4.67 and 1.77) and their corresponding eigenvectors from the complete eigenvector/eigenvalue decomposition of **B**:

$$
\begin{bmatrix}
.40 & .28 \\
.39 & .33 \\
.38 & .34 \\
.39 & .30 \\
.35 & -.39 \\
.31 & -.40 \\
.29 & -.44 \\
.31 & -.31
\end{bmatrix}
\begin{bmatrix}
4.67 & 0.0 \\
0.0 & 1.77
\end{bmatrix}
\begin{bmatrix}
.40 & .39 & .38 & .39 & .35 & .31 & .29 & .31 \\
.28 & .33 & .34 & .30 & -.39 & -.40 & -.44 & -.31
\end{bmatrix}
$$

---

[1] For now, we use matrix notation common in NLA, but later, when quoting original sources, we will adopt the notation of these authors as a way of acknowledging their contributions.

$$
= \begin{bmatrix}
.89 & .89 & .88 & .88 & .46 & .38 & .32 & .43 \\
.88 & .89 & .89 & .88 & .41 & .33 & .26 & .38 \\
.87 & .89 & .87 & .86 & .38 & .30 & .24 & .35 \\
.87 & .88 & .86 & .86 & .43 & .36 & .29 & .40 \\
.46 & .41 & .38 & .43 & .85 & .79 & .77 & .73 \\
.38 & .33 & .30 & .36 & .79 & .73 & .72 & .66 \\
.32 & .27 & .25 & .29 & .78 & .73 & .74 & .66 \\
.42 & .38 & .35 & .40 & .73 & .68 & .66 & .62
\end{bmatrix}.
$$

Although one may attempt to interpret the given correlations directly, in many cases it is a lot easier to do so on the basis of the lower-rank (now 2) approximation to **B** just given. The latter could again be subjected to a variety of other factorization strategies that reproduce **B** just as well, and which in addition would hopefully suggest nice substantive interpretations. For example, the factorization

$$
\begin{bmatrix}
.93 & .05 \\
.95 & -.02 \\
.93 & -.04 \\
.93 & .02 \\
.45 & .81 \\
.37 & .78 \\
.29 & .79 \\
.41 & .68
\end{bmatrix}
\begin{bmatrix}
.93 & .95 & .93 & .93 & .45 & .37 & .29 & .41 \\
.05 & -.02 & -.04 & .02 & .81 & .78 & .79 & .68
\end{bmatrix}
$$

provides a clear split of the variables into two groups: the first four of height, arm span, length of forearm, and length of lower leg, which constitute a "lankiness" collection, and the last four of weight, bitrochanteric diameter, chest girth, and chest width, which constitute a "stockiness" collection. This last factorization was obtained from the rank reduction results to be presented in section 2.2, and in particular from (2.1). A further discussion of this specific factorization just given is delayed until that section and then provided in a footnote to (2.1). It is also possible to present the eight variables in a two-dimensional space based on the coordinates for the eight variables obtained from the matrix used in this particular factorization. The subdivision of the variables into two distinct groups of lankiness and stockiness would be very clear graphically from this representation.

Over the last half century or so, the separate literatures in the fields of NLA and AS/P have evolved in rather separate fashions, although there have been a few significant historical ties because of a common interest in matrix factorization. The main body of the present paper will try to make this latter point more explicitly by discussing a 1995 contribution to *SIAM Review* by Chu, Funderlic, and Golub [5] (hereafter designated as CFG) entitled "A rank-one reduction formula and its applications to matrix factorizations," which presented a versatile rank-1 reduction formula attributed to Wedderburn and discussed how it might be used to unify and organize in a coherent framework an array of different matrix factorization schemes prominent in the NLA literature. In all cases, CFG's integration is directed toward the ultimate purpose of solving systems of equations. On the other side, we will note in some detail the connections between CFG and work available for some decades in the AS/P literature, specifically work by the late Louis Guttman [17, 18, 20]. In a manner very parallel to unifying matrix factorization methods for solving systems of equations presented by CFG, Guttman's basic results on rank reduction have served the field of AS/P for some time as the primary means for organizing matrix factorization methods when the purpose is dimension reduction and data interpretation.

There are several such organizing presentations available, but probably none is more aggressively comprehensive than the textbook by Horst [27]. Horst reviews and integrates some 50 years of data representation through matrix factorization and does so almost exclusively through the mechanism of Guttman's rank reduction results. Some of this integration will be reviewed briefly in the next section.

The fields of NLA and AS/P, as noted above, have had some very significant historical ties. The prime example is probably the development of what has become *the* major tool for both NLA (the singular value decomposition (SVD) of a matrix) and AS/P (in which it is more commonly called the Eckart–Young decomposition of a matrix after Eckart and Young [12], which as Stewart [37, p. 563] comments is probably an incorrect attribution given the precedent of Schmidt's earlier work [36]. In the last section of this paper, we point to two of these historical issues. We also present a brief review of a generalization of the SVD that has been developed for the purpose of data representation, as well as for uses in a data representation context for another technique of importance in NLA that involves applications of iterative or cyclic projection methods. Pointing out these connections may be of some interdisciplinary interest for those likely to rely on the SVD and associated methods only indirectly along the route to constructing a solution to another problem that is of more primary concern.[2]

2. The CFG and Guttman Rank Reduction Theorems.

2.1. CFG. The CFG review paper is built around three main theorems that we reference below as CFG1, CFG2, and CFG3, with CFG1 being the main result from which the various matrix factorizations reviewed are unified.

CFG1. If $A \in R^{m \times n}$, and $x \in R^n$, $y \in R^m$ are vectors such that $\omega = y^T A x \neq 0$, then the matrix $B := A - \omega^{-1} A x y^T A$ has rank exactly one less than the rank of A.

CFG2. Let $u \in R^m$ and $v \in R^n$. Then the rank of the matrix $B = A - \sigma^{-1} u v^T$ is less than that of A if and only if there are vectors $x \in R^n$ and $y \in R^m$ such that $u = Ax, v = A^T y$, and $\sigma = y^T A x$, in which case rank$(B) = $ rank$(A) -1$.

CFG3. Suppose $U \in R^{m \times k}$, $R \in R^{k \times k}$, and $V \in R^{n \times k}$. Then rank$(A - U R^{-1} V^T)$ $= $ rank$(A) - $ rank$(U R^{-1} V^T)$ if and only if there exist $X \in R^{n \times k}$ and $Y \in R^{m \times k}$ such that $U = AX$, $V = A^T Y$, and $R = Y^T A X$.

A timeline for the appearance of these rank reduction results in the NLA and AS/P literatures is given in Figure 2.1, which provides a short summary of the discussion in this section.

Historically, CFG attribute CFG1 to Wedderburn and his classic 1934 text [46, p. 69]. The converse, present in CFG2, is noted to be part of an exercise in Householder [26, p. 33, Exercise 34] with due reference to Wedderburn. However, CFG attribute the entire characterization in CFG2 to a posthumous paper by Egerváry [13] (who apparently was unaware of Wedderburn's earlier result), but also comment that Householder, given that drafts of his book were available in 1960, may have independently discovered the converse in CFG2. The comprehensive result in CFG3,

---

[2]We will not delve further into biographical issues in this paper and will resist the temptation to explore the connections between AS/P and one of the founders of modern NLA, A. S. Householder (to whose memory the CFG contribution is dedicated). In his very early career, Householder was very involved with psychometrics and the relevant University of Chicago community (Householder was on the editorial board of and a prolific contributor to *Psychometrika* in the 1930s and early 1940s). With Gale Young he popularized to the mathematical community the lower rank approximations of a matrix based on the SVD in Householder and Young [27], and as a result of his further joint work with Young in 1938 [48], truly began the whole field of what is now known as classical multidimensional scaling in AS/P.
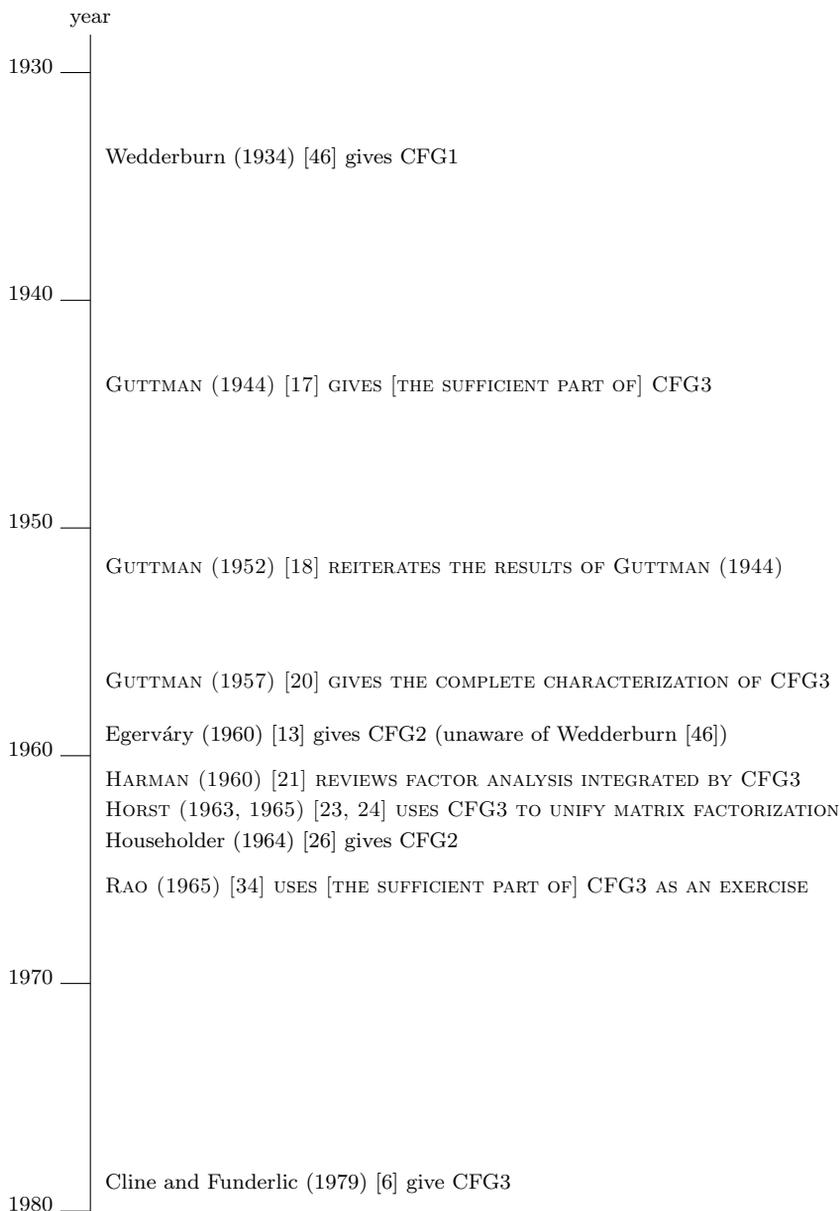
year

1930

Wedderburn (1934) [46] gives CFG1

1940

GUTTMAN (1944) [17] GIVES [THE SUFFICIENT PART OF] CFG3

1950

GUTTMAN (1952) [18] REITERATES THE RESULTS OF GUTTMAN (1944)

GUTTMAN (1957) [20] GIVES THE COMPLETE CHARACTERIZATION OF CFG3

Egerváry (1960) [13] gives CFG2 (unaware of Wedderburn [46])

1960

HARMAN (1960) [21] REVIEWS FACTOR ANALYSIS INTEGRATED BY CFG3
HORST (1963, 1965) [23, 24] USES CFG3 TO UNIFY MATRIX FACTORIZATION
Householder (1964) [26] gives CFG2

RAO (1965) [34] USES [THE SUFFICIENT PART OF] CFG3 AS AN EXERCISE

1970

Cline and Funderlic (1979) [6] give CFG3

1980

Fig. 2.1 *A timeline for the appearance of the rank reduction results in the NLA and AS/P litera-*
tures. The results are referred to in both the text and the figure by the acronyms CFG1,
CFG2, and CFG3. Those results that appeared in the AS/P literature are indicated by
small capital letters.

which obviously encompasses both CFG1 and CFG2, is attributed to Cline and Fun-
derlic [6]. As we will see below in greater detail, the CFG3 result was available in
complete form in the AS/P literature by 1957 in Guttman [20] with the sufficient
part available as early as 1944 [17], and from the discussion of its use by Guttman,
it was known and applied for some years prior to this. More pointedly, the sufficient

condition in CFG3 has been a long-standing unifying principle in AS/P. For example, the 1963 applied text *Matrix Algebra for Social Scientists* by Horst [23] includes a complete chapter devoted to its applications (Chapter 22, pp. 477–487, "The general matrix reduction theorems"); the seminal text in linear statistical inference by Rao [34] includes the sufficient condition of CFG3 merely as an exercise (that we repeat verbatim below in the original notation [34, p. 55, Exercise 4]), with an attribution (to be discussed more fully later) to Lagrange's theorem.

Lagrange's Theorem. Let $\mathbf{S}$ be any square matrix of order n and rank $r > 0$, and $\mathbf{X}$, $\mathbf{Y}$ be column vectors such that $\mathbf{X'SY} \neq 0$. Then the residual matrix

$$\mathbf{S}_1 = \mathbf{S} - \frac{\mathbf{SYX'S}}{\mathbf{X'SY}}$$

is exactly of rank $r - 1$.

More generally, if $\mathbf{S}$ is $n \times m$ of rank $r > 0$, $\mathbf{A}$ and $\mathbf{B}$ are of order $s \times n$ and $s \times m$, respectively, where $s \leq r$, and $\mathbf{ASB'}$ is nonsingular, the residual matrix

$$\mathbf{S}_1 = \mathbf{S} - \mathbf{SB'}(\mathbf{ASB'})^{-1}\mathbf{AS}$$

is exactly of rank $(r - s)$ and $\mathbf{S_1}$ is Gramian if $\mathbf{S}$ is Gramian.[3]

The primary emphasis in CFG is on the repeated use of CFG1 to obtain a variety of matrix factorizations and to unify a substantial literature that has gone before. They summarize this usage as follows.

For $A \in R^{m \times n}$, rank$(A) = \gamma$, and vectors $x_1, \ldots, x_\gamma \in R^n$ and $y_1, \ldots, y_\gamma \in R^m$,

$$(2.1) \qquad A = \Phi\Omega^{-1}\Psi^T,$$

where $\Omega := \text{diagonal}\{\omega_1, \ldots, \omega_\gamma\}$, $\Phi := [\phi_1, \ldots, \phi_\gamma] \in R^{m \times \gamma}$, and $\Psi := [\psi_1, \ldots, \psi_\gamma] \in R^{n \times \gamma}$, with

$$\phi_k := A_k x_k, \ \psi_k := A_k^T y_k,$$

and letting $A_1 := A$, $A_{k+1} := A_k - \omega_k^{-1} A_k x_k y_k^T A_k$, where (it is assumed that $x_1, \ldots, x_\gamma$ and $y_1, \ldots, y_\gamma$ are chosen so) $\omega_k := y_k^T A_k x_k \neq 0$.[4]

Alternatively, if $u_1, \ldots, u_\gamma \in R^n$ and $v_1, \ldots, v_\gamma \in R^m$ are defined by $u_1 := x_1$ and $v_1 := y_1$, and for $k > 1$,

$$u_k := x_k - \sum_{i=1}^{k-1} \left(\frac{v_i^T A x_k}{v_i^T A u_i} \right) u_i,$$

---

[3]The statement that "$\mathbf{S_1}$ is Gramian if $\mathbf{S}$ is Gramian" needs the further qualification of "and $\mathbf{A} = \mathbf{B}$" to be correct. In this summary statement, and probably to economize on words in the exercise, Rao put together the sufficient condition of CFG3 with a more specific result from Guttman [17] that dealt separately with Gramian matrices. We label this latter Guttman result as G3 in section 2.2. This observation that the constraint of equality is needed for the matrices $\mathbf{A}$ and $\mathbf{B}$ to make the phrasing given by Rao completely correct was pointed out to us by Yoshio Takane [39].

[4]In the earlier numerical example of section 1, (2.1) was used to generate a factorization for the rank-2 approximation of the given correlation matrix. In (2.1), the matrix A is interpreted as the 8×8 rank-2 approximation to $\mathbf{B}$ (so $m = n = 8$ and $\gamma = 2$), and the vectors $x_1 = y_1 = [11110000]^T$ and $x_2 = y_2 = [00001111]^T$. The factorization of A that is given corresponds to $A = (\Phi\Omega^{-\frac{1}{2}})(\Psi\Omega^{-\frac{1}{2}})^T$, where in this case

$$\Phi\Omega^{-\frac{1}{2}} = \Psi\Omega^{-\frac{1}{2}} = \begin{bmatrix} .93 & .95 & .93 & .93 & .45 & .37 & .29 & .41 \\ .05 & -.02 & -.04 & .02 & .81 & .78 & .79 & .68 \end{bmatrix}^T.$$

$$v_k := y_k - \sum_{i=1}^{k-1} \left(\frac{y_k^T A u_i}{v_i^T A u_i} \right) v_i,$$

then for $U := [u_1, \ldots, u_\gamma]$ and $V := [v_1, \ldots, v_\gamma]$,

$$(2.2) \qquad\qquad\qquad V^T A U = \Omega,$$

and the factorization in (2.1) can be rewritten as

$$(2.3) \qquad\qquad\qquad A = (AU)\Omega^{-1}(V^T A).$$

Depending on the initial matrix A and the choice of the vector sets $x_1, \ldots, x_\gamma$ and $y_1, \ldots, y_\gamma$, a variety of factorizations of A ensue. (In fact, even the ubiquitous Gram–Schmidt orthogonalization process is a very simple special case: let A be the identity matrix and suppose the two collections, $x_1, \ldots, x_\gamma$ and $y_1, \ldots, y_\gamma$, are identical and contain the vectors for which an orthogonal basis is desired. The columns of $U = V$ define the set of basis vectors with the orthogonality condition given by (2.2).) To allow some comparisons later when the factorization of A for purposes of data representation are reviewed, we list three of the more common (and simplest) ones below in schematic summary form (assuming the condition $\omega_k \neq 0$ obtains in the recursive process), but will not discuss, for example, CFG's further extension to the QR decomposition, the Lanczos process, or to ABS methods; the latter are more relevant to the solution of systems of equations, but usage in the context of matrix factorization with the aim of data representation would be more limited.

(a) For $A \in R^{m \times n}$ of rank γ, if X and Y are upper trapezoidal matrices in $R^{n \times \gamma}$ and $R^{m \times \gamma}$, respectively, then (2.3) provides a trapezoidal LDU decomposition where AU and $V^T A$ are lower and upper trapezoidal matrices, respectively.[5]

(b) For $A \in R^{n \times n}$ of rank n, if X and Y are both the identity matrix, then (2.3) provides an LDM^T factorization of A, where $A = V^{-T} \Omega U^{-1}$ for V^{-T} and U^{-T} unit lower triangular matrices (with ones on the main diagonal).

If $A \in R^{n \times n}$ is symmetric, then $U = V$ and $A = U^{-T} \Omega U^{-1}$.

If $A \in R^{n \times n}$ is symmetric and positive definite, then $A = (\Omega^{\frac{1}{2}} U^{-1})^T (\Omega^{\frac{1}{2}} U^{-1})$ gives the Cholesky factorization of A.

(c) If $A \in R^{m \times n}$ is of rank γ, the singular value factorization of A can be given as $A = P \Omega Q^T$, where $P \in R^{m \times \gamma}$ and $P^T P = I$; $Q \in R^{n \times \gamma}$ and $Q^T Q = I$; $\Omega \in R^{\gamma \times \gamma}$ and is diagonal with positive (and nonincreasing) entries along the main diagonal. If $X = Q$ and $Y = P$, then the SVD of A is retrieved by (2.2) since $U = Q$ and $V = P$, so $V^T A U = \Omega$.

2.2. Guttman. The three papers by Guttman mentioned earlier [17, 18, 20] are rather remarkable in that they served to completely characterize the issue of rank reduction for the purpose of data representation within AS/P by 1957, with the sufficiency clearly delineated in 1944 and completely integrated within the literature by the 1950s (e.g., see Harman's 1960 comprehensive review [21]). We state the major theorems below in their original forms (as well as using most of the original notation) as G1, G2, and G3. We proceed thereafter to make several historical comments and place these rank reduction results in the framework of interpreting the structure of a data matrix.

---

[5]An upper (lower) trapezoidal matrix is one in which the (i, j) entry is zero for $j < i$ $(i < j)$, which naturally generalizes the notion of an upper or lower triangular matrix when A is square.

G1 [17, p. 4]: Let $\mathbf{S}$ be any matrix of order $n \times N$ and of rank $r > 0$. Let $\mathbf{X}$ and $\mathbf{Y}$ be of orders $s \times n$ and $s \times N$, respectively (where $s \leq r$), and such that $\mathbf{XSY}'$ is nonsingular. Then the residual matrix

$$\mathbf{S_1} = \mathbf{S} - \mathbf{SY}'(\mathbf{XSY}')^{-1}\mathbf{XS}$$

is exactly of rank $r - s$.

Guttman comments that G1 can be considered as generalizing a result due to Lagrange, and he references Wedderburn [46, p. 68] with an apparent allusion to the introductory phrase "The Lagrange method of reducing quadratic and bilinear forms to a normal form is ...." Wedderburn uses this phrase before discussing the rank-1 reduction result (which was given earlier as CFG1). Guttman then presents what he explicitly labels as Lagrange's theorem: Let $\mathbf{S}$ be any matrix of order $n \times n$ and of rank $r > 0$. Let $\mathbf{x}$ and $\mathbf{y}$ be row vectors of n elements each and such that $\mathbf{xSy}' \neq 0$. Then the residual matrix

$$\mathbf{S_1} = \mathbf{S} - \frac{\mathbf{Sy}'\mathbf{xS}}{\mathbf{xSy}'}$$

is exactly of rank $r - 1$.

Although Guttman gave the label of Lagrange's theorem only to this latter special case of G1 (and by implication indicated that the more general form of G1 and its proof were his own), as we noted earlier in the exercise from Rao [34, p. 55] given verbatim, Rao attached the label of Lagrange's theorem to Guttman's G1 result as well. Rao provided no reference to Guttman in this context, although he does so for another exercise [34, Exercise 14, p. 57] dealing with the minimum rank possible for a (reduced) correlation matrix, but then lists a reference to a completely different paper (Guttman [19]).

Guttman only gave a proof for the sufficiency of the rank reduction result of G1 in his 1944 paper, but he came back to prove its necessity in a short article in 1957 [20]. Thus, the complete characterization given earlier in CFG3 was available in the AS/P literature by 1957—and Guttman should appropriately be given credit for it.[6]

The next two results from Guttman [17, pp. 11, 12], G2 and G3, also concern the issue of rank reduction but not in terms of the original $n \times N$ matrix $\mathbf{S}$. Many (if not most) uses of matrix factorizations for purposes of data representation rely on matrices that must be Gramian (e.g., correlation or covariance matrices), and it is very relevant to know how such factorizations relate back to the original data matrix from which the Gramian matrices were constructed. G2 concerns a factorization of the $n \times n$ positive semidefinite (Gramian) matrix $\mathbf{SS}'$ and how this could then be related to a factorization of $\mathbf{S}$; G3 is a restatement of G1, but for a matrix that is assumed to be Gramian. We restate G2 from Guttman [17, p. 11] in a slight variant form that deals just with the cross-product matrix $\mathbf{SS}'$ and we include an explicit

---

[6]Although Guttman's results appear in the AS/P literature, that does not preclude them from being rediscovered and republished in this same literature, where eventually no credit at all is given. As a case in point, Overall [33] presented some of Guttman's results independently and noted in passing, "It has been called to the attention of this writer that Guttman (1944, 1952) has presented what is essentially the same general factor model in somewhat different form" (p. 652). However, secondary sources are not always so careful in attribution, and when providing a separate chapter and accompanying Fortran program in their classic text, Cooley and Lohnes (1971) [7, Chapter 5] reference Overall [33] but fail to mention any of Guttman's original contributions. Unfortunately, when cited for applications in the substantive literature, Cooley and Lohnes [7] is now made the primary (and only) reference for methodology directly attributable to Guttman.

representation for one of the defining matrices ($\mathbf{P}$) that appears only in Guttman's proof.

G2: If $\mathbf{S}$ is of order $n \times N$ and of rank r, $\mathbf{F}$ is of order $n \times r$ (and of rank r), and $\mathbf{SS}' = \mathbf{FF}'$, then there is a unique matrix $\mathbf{P}$ of order $r \times N$ such that

$$\mathbf{S} = \mathbf{FP}.$$

The matrix $\mathbf{P} = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{S}$ satisfies $\mathbf{PP}' = \mathbf{I}$ (i.e., $\mathbf{P}$ has orthonormal rows).

G3: Let $\mathbf{G}$ be a Gramian matrix of order $n \times n$ and of rank $r > 0$. Let $\mathbf{X}$ be of order $s \times n$ and such that $\mathbf{XGX}'$ is nonsingular. Then the residual matrix

$$\mathbf{G_1} = \mathbf{G} - \mathbf{GX}'(\mathbf{XGX}')^{-1}\mathbf{XG}$$

is of rank $r - s$ and is Gramian.

The original impetus for the Guttman results in G1, G2, and G3 was to provide a unifying framework for the methods of factor analysis that were then popular, particularly by incorporating the centroid method of Thurstone (e.g., [40]) and the principal components characterization from Hotelling [25]. Both of these methods were usually implemented with a correlation matrix in which one begins with a data matrix (to use Guttman's notation) $\mathbf{S}$ of order $n \times N$, where N typically refers to subjects and n to variables. The entries within each row of $\mathbf{S}$ are assumed to be standardized to a mean of zero and standard deviation of one, so the $n \times n$ (positive semidefinite) correlation matrix $\mathbf{R}$ is equal to $(\frac{1}{N})\mathbf{SS}'$. (At times, $\mathbf{R}$ is further discounted by replacing the ones along the main diagonal by what are called communality estimates, but we will ignore this possibility here.)

Given $\mathbf{R}$, the result in G3 could then be applied, and typically in a one-at-a-time fashion, using a row vector $\mathbf{x}$ of order $1 \times n$. What would result would be a direct analogue of the iterated use of the rank-1 Wedderburn reductions of (2.1). Here, $\mathbf{R} = \mathbf{\Lambda\Omega\Lambda}'$, where r is the rank of $\mathbf{R}$, $\mathbf{\Lambda}$ is $n \times r$ of rank r, and $\mathbf{\Omega}$ is an $r \times r$ diagonal matrix with positive entries along the diagonal. (The entries in $\mathbf{\Omega}$ are positive since the original matrix and the residual matrices constructed from it are all Gramian.) Expressing $\mathbf{R}$ as $(\mathbf{\Lambda\Omega}^{\frac{1}{2}})(\mathbf{\Lambda\Omega}^{\frac{1}{2}})'$, further effort could then be directed toward the simplification of the "factors" by, for example, locating an orthogonal $r \times r$ (rotation) matrix $\mathbf{T}$ such that $\mathbf{R} = (\mathbf{\Lambda\Omega}^{\frac{1}{2}}\mathbf{T})(\mathbf{\Lambda\Omega}^{\frac{1}{2}}\mathbf{T})'$, which would hopefully provide a more substantively understandable result than the original factors may have.

As noted above, one of Guttman's primary motivations for his 1944 paper was to provide a formal justification for the then-popular centroid method of factor analysis, i.e., that the rank of the correlation matrix was actually reduced by one at each stage in the sequential process. In our current notation, the Thurstone method can be characterized through the one-at-a-time use of G3 beginning with a $1 \times n$ vector $\mathbf{x}$ having constant entries of $+1$. Entries in the subsequent vectors chosen to effect the rank reduction would be ± 1, where the choice of signs was determined from the pattern of entries in the residual matrix. Comments in Guttman [17] and elsewhere (e.g., Horst [23] and Harman [21]) with regard to this centroid strategy generally considered it a poor approximation to what could be generated from Hotelling's method that would choose successive unit length vectors to produce a rank reduction by identifying (through an iterative strategy) the eigenvector associated with the largest eigenvalue for each of the residual matrices successively obtained. At the time, however, the centroid method was computationally much less demanding than Hotelling's iterative (or power) method for obtaining each of the principal components (again, one-at-a-time

and reducing rank at each iteration); for this reason alone, the centroid method was a very common factorization strategy until electronic computing capabilities became more widely available.

The 1952 Guttman paper [18] was motivated by a lack of attention to his 1944 contribution since a number of individuals (e.g., Thurstone [41, 42] and Holzinger [22]) were publishing methods of factor analysis based on what were called multiple group strategies that could have been unified by Guttman's earlier results. Multiple group methods choose the successive vectors in the Wedderburn reductions to reflect groups of variables that should be clustered together, with the resulting binary (1/0) vectors merely reflecting group membership, or they do so simultaneously by using what are in effect Guttman's G3 results.[7] In their simultaneous extraction, the matrix $\mathbf{GX}$ represents the covariances of the variables with the factors, and the matrix $\mathbf{XGX}'$ is the variance-covariance matrix between the factors. Guttman [18] reiterates the theorems from his 1944 article and points out how the multiple group methods were merely a special case of his earlier results. In the course of discussing how these multiple group methods could be incorporated, Guttman shows his enthusiasm, as do CFG, for unification, but now in the context of data representation: "It was shown how this theorem [referring to G3] includes all previous correlation factoring methods, and in addition provides new techniques that enable as many common factors to be extracted in one operation as one wishes. Different methods differ solely by their choice of the weight matrix $\mathbf{X}$" [18, p. 210].

And, once the necessity of G1 was shown in 1957, he added, "All possible factoring methods, whether directly on the score matrix or on the correlation matrix, can differ only in the choice of weight matrices $\mathbf{X}$ and $\mathbf{Y}$" [20, p. 81].

Guttman goes on to argue how important the choice is for the vectors in $\mathbf{X}$ whether used one at a time or simultaneously for the purposes of data representation. Specifically, the vectors should be chosen to avoid the necessity of "cleaning up" the factorization by rotational strategies of the type alluded to earlier.

3. Two (Other) Connections between the Processes of NLA and Methods for Data Representation.

3.1. Data Representation Uses of an Alternative Strategy for Solving Linear Systems of Equations through Iterative Projection.
Matrix factorization may be the most well known strategy for solving linear systems of equations, but another method, typically attributed to Kaczmarz [31] (e.g., see Bodewig [3, pp. 163–164] or, more recently, Deutsch [9, pp. 107–108]) and based on an iterative projection strategy also has some very close connections with several more recent approaches in the AS/P literature to the representation of a data matrix. The latter rely on a close relative to the Kaczmarz strategy and what is now commonly referred to in AS/P as Dykstra's method for solving linear inequality constrained weighted least-squares tasks (e.g., see Dykstra [10]).

Kaczmarz's method can be characterized as follows. Given $\mathbf{A} = \{a_{ij}\}$ of order $m \times n$, $\mathbf{x}' = \{x_1, \ldots, x_n\}$, $\mathbf{b}' = \{b_1, \ldots, b_m\}$, and assuming the linear system $\mathbf{Ax} = \mathbf{b}$ is consistent, define the set $C_i = \{\mathbf{x} \mid a_{ij}x_j = b_i\}$ for $1 \leq i \leq m$. The projection of any $n \times 1$ vector $\mathbf{y}$ onto C_i is simply $\mathbf{y} - (\mathbf{a}_i'\mathbf{y} - b_i)\mathbf{a}_i(\mathbf{a}_i'\mathbf{a}_i)^{-1}$, where $\mathbf{a}_i' = \{a_{i1}, \ldots, a_{in}\}$. Beginning with a vector $\mathbf{x}_0$ and successively projecting $\mathbf{x}_0$ onto C_1, and that result

---

[7]As noted in footnote 4, the factorization for the rank-2 approximation of the correlation matrix given as an example in section 1 is based on binary vectors reflecting the group membership of "lankiness" and "stockiness" for the eight physical variables.

onto C_2, and so on, and cyclically and repeatedly reconsidering projections onto the sets $C_1, \ldots, C_m$, leads at convergence to a vector $\mathbf{x}_0^*$ that is closest (in vector 2-norm) to $\mathbf{x}_0$ while it satisfies $\mathbf{A}\mathbf{x}_0^* = \mathbf{b}$.

Dykstra's method can be characterized as follows. Given $\mathbf{A} = \{a_{ij}\}$ of order $m \times n$, $\mathbf{x}_0' = \{x_{01}, \ldots, x_{0n}\}$, $\mathbf{b}' = \{b_1, \ldots, b_m\}$, and $\mathbf{w}' = \{w_1, \ldots, w_n\}$, where $w_j > 0$ for all j, find $\mathbf{x}_0^*$ such that $\mathbf{a}_i'\mathbf{x}_0^* \leq b_i$ for $1 \leq i \leq m$ and $\sum_{i=1}^{n} w_i(x_{0i} - x_{0i}^*)^2$ is minimized. Again, (re)define the (closed convex) sets $C_i = \{\mathbf{x} \mid a_{ij}x_j \leq b_i\}$, and when a vector $\mathbf{y} \notin C_i$, its projection onto C_i (in the metric defined by the weight vector $\mathbf{w}$) is $\mathbf{y} - (\mathbf{a}_i'\mathbf{y} - b_i)\mathbf{a}_i\mathbf{W}^{-1}(\mathbf{a}_i'\mathbf{W}^{-1}\mathbf{a}_i)^{-1}$, where $\mathbf{W}^{-1} = \mathrm{diag}\{w_1^{-1}, \ldots, w_n^{-1}\}$. We again initialize the process with the vector $\mathbf{x}_0$ and each set $C_1, \ldots, C_m$ is considered in turn. If the vector being carried forward to this point when C_i is (re)considered does not satisfy the constraint defining C_i, a projection onto C_i occurs. The sets $C_1, \ldots, C_m$ are cyclically and repeatedly considered, but with one difference from the operation of Kaczmarz's method—each time a constraint set C_i is revisited, any changes from the previous time C_i was reached are first "added back." This last process ensures convergence to an optimal solution $\mathbf{x}_0^*$ (see Dykstra [10]).

The Dykstra method currently serves as the major computational tool for a variety of newer data representation devices in AS/P. For example, and first considering an arbitrary rectangular data matrix, Dykstra and Robertson [11] use it to fit a least-squares approximation constrained by entries within rows and within columns that are monotonic with respect to given row and column orders.[8] For an arbitrary symmetric proximity matrix $\mathbf{A}$ (of order $p \times p$ and with diagonal entries typically set to zero), a number of applications of Dykstra's method have been discussed for approximating $\mathbf{A}$ in a least-squares sense by $\mathbf{A}_1 + \cdots + \mathbf{A}_K$, where K is typically small (e.g., 2 or 3) and each $\mathbf{A}_k$ is patterned in a particularly informative way that can be characterized by a set of linear inequality constraints that its entries should satisfy. We note three exemplar classes of patterns that $\mathbf{A}_k$ might have, all with a substantial history in the AS/P literature. In each instance, Dykstra's method can be used to fit the additive structures satisfying the inequality constraints once they are identified, possibly through an initial combinatorial optimization task seeking an optimal reordering of a given (residual) data matrix, or in some instances in a heuristic form to identify the constraints to impose in the first place. We merely give the patterns sought in $\mathbf{A}_k$ and refer the reader to sources that develop the representations in more detail.

(a) Order constraints (Hubert and Arabie [28]): The entries in $\mathbf{A}_k = \{a_{ij(k)}\}$ should satisfy the anti-Robinson constraints: there exists a permutation on the first p integers $\rho(\cdot)$ such that $a_{\rho(i)\rho(j)(k)} \leq a_{\rho(i)\rho(j')(k)}$ for $1 \leq i < j < j' \leq p$, and $a_{\rho(i)\rho(j)(k)} \leq a_{\rho(i')\rho(j)(k)}$ for $1 \leq i < i' < j' \leq p$.

(b) Ultrametric and additive trees (Hubert and Arabie [29]): The entries in $\mathbf{A}_k$ should be represented by an ultrametric: for all i, j, and h, $a_{ij(k)} \leq \max\{a_{ih(k)}, a_{jh(k)}\}$; or by an additive tree: for all i, j, h, and l, $a_{ij(k)} + a_{hl(k)} \leq \max\{a_{ih(k)} + a_{jl(k)}, a_{il(k)} + a_{jh(k)}\}$.

(c) Linear and circular unidimensional scales (Hubert, Arabie, and Meulman [30]): The entries in $\mathbf{A}_k$ should be represented by a linear unidimensional scale: $a_{ij(k)} = |x_j - x_i|$ for some set of coordinates $x_1, \ldots, x_n$; or a circular unidimensional scale: $a_{ij(k)} = \min\{|x_j - x_i|, x_0 - |x_j - x_i|\}$ for some set of coordinates $x_1, \ldots, x_n$ and x_0 representing the circumference of the circular structure.

---

[8]Or to give a reference for an even broader usage, Van der Lans [44, section 3.3] applies Dykstra's method to approximate a rectangular matrix under a more general set of inequality constraints constructed from a model assumed to be responsible for generating the matrix.

3.2. Alternative Notions for How the SVD Should Be Generalized. As noted earlier, in both NLA and AS/P, the SVD of a matrix has become a major computational tool. For NLA it often serves as a means for rephrasing some given problem into a simpler or more transparent form, and for AS/P it has become the mechanism for effecting a very wide variety of geometric representations for a rectangular data matrix. Given the ubiquity of the SVD in both NLA and AS/P, it should not be surprising that separate generalizations of the SVD have been suggested and that both have been labeled with the generic title of the generalized singular value decomposition (GSVD). Here we rephrase from Golub and Van Loan [15, p. 466] the most standard generalization in NLA, which was introduced in Van Loan [45]:

GSVD. If we have $A \in R^{m \times n}$ with $m \geq n$ and $B \in R^{p \times n}$, then there exist orthogonal $U \in R^{m \times m}$ and $V \in R^{p \times p}$ and an invertible $X \in R^{n \times n}$ such that

$$U^T A X = C = \mathrm{diag}(c_1, \ldots, c_n), \qquad c_i \geq 0,$$

and

$$V^T B X = S = \mathrm{diag}(s_1, \ldots, s_q), \qquad s_i \geq 0,$$

where $q = \min(p, n)$.

The emphasis here is obviously on the simultaneous diagonalization of A and B, with such decompositions relevant when a particular task includes two such matrices that need to be simplified. The prime example provided by Golub and Van Loan [15, pp. 580–582] is in the solution to a least-squares minimization task with a quadratic inequality constraint. Extensions of such simultaneous diagonalizations to three (or more) matrices have also been developed, most notably by De Moor and Golub [8], and some attempts have been made in the AS/P literature to use these generalizations directly for a variety of data representation tasks (e.g., see Takane [38]).

In AS/P, however, it is almost universal that the GSVD refers to the decomposition of a matrix that involves the use of different metrics in the normalizations of the components of the decomposition (also mentioned in Van Loan [45] as a second possible generalization of the SVD), as well as to a subsequent application to find lower-rank approximations (or generalized Eckart–Young decompositions) that minimize a weighted least-squares loss criterion. Or even more generally, and following the comprehensive review given in Rao [35], the GSVD in this form is a way of approaching the task of approximating a given matrix by one contained within a specified subclass that minimizes a suitable (unitarily invariant) norm of their difference.

To give the GSVD in the form most familiar in AS/P, we repeat a result from Gower and Hand [16, p. 237], with a slight variation on the notation they use.

GSVD in the metrics **P** and **Q**: Given any rectangular $p \times q$ matrix **X** and symmetric positive definite matrices **P** and **Q** (with orders $p \times p$ and $q \times q$), then

$$(3.1) \qquad\qquad \mathbf{X} = \mathbf{S}\boldsymbol{\Sigma}\mathbf{T}',$$

where **S** is $p \times p$ and satisfies the normalization $\mathbf{S}'\mathbf{P}\mathbf{S} = \mathbf{I}$, **T** is $q \times q$ and satisfies the normalization $\mathbf{T}'\mathbf{Q}\mathbf{T} = \mathbf{I}$, and $\boldsymbol{\Sigma}$ is $p \times q$ and diagonal (in the extended sense) with nonnegative and nonincreasing values from the left to the right.

Gower and Hand go on to provide a generalized Eckart–Young factorization and lower rank approximation that we restate as follows: If for the factorization in (3.1) $\mathbf{S}_{(r)}$ denotes and is formed from the first r columns of **S**, $\boldsymbol{\Sigma}_{(r)}$ denotes and is formed from the $r \times r$ upper left diagonal matrix selected from $\boldsymbol{\Sigma}$, and $\mathbf{T}'_{(r)}$ denotes and is formed from the first r rows of $\mathbf{T}'$, then

$$(3.2) \qquad\qquad \hat{\mathbf{X}} = \mathbf{S}_{(r)}\boldsymbol{\Sigma}_{(r)}\mathbf{T}'_{(r)}$$

is a rank-r matrix minimizing the weighted least-squares criterion of

$$\text{Trace}[(\mathbf{P}^{\frac{1}{2}}(\mathbf{X} - \hat{\mathbf{X}})\mathbf{Q}^{\frac{1}{2}})(\mathbf{P}^{\frac{1}{2}}(\mathbf{X} - \hat{\mathbf{X}})\mathbf{Q}^{\frac{1}{2}})'].$$

In practice, $\hat{\mathbf{X}}$ would be obtained from a standard singular value factorization of $\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{Q}^{\frac{1}{2}}$, with the appropriate deletion of rows and columns in the components of the factorization to obtain a rank-r (unweighted) least-squares approximation; i.e., we first construct $\mathbf{P}^{\frac{1}{2}}\hat{\mathbf{X}}\mathbf{Q}^{\frac{1}{2}}$, and from the latter, $\hat{\mathbf{X}}$ can be retrieved.

The first application of the GSVD in the form just presented appears to be in a paper by Young [47], but it now serves as the computational engine for a number of popular data representation methods, e.g., various forms of what might generically be called biplots. To be more explicit, suppose we interpret the rectangular $p \times q$ matrix $\mathbf{X} = \{x_{ij}\}$ in (3.1) as a data matrix on p objects (subjects) and q variables (attributes), so the entry x_{ij} refers to the ith object's score on the jth variable, $1 \leq i \leq p$ and $1 \leq j \leq q$. The lower-rank approximation to $\mathbf{X}$ given by $\hat{\mathbf{X}} = \{\hat{x}_{ij}\}$ in (3.2) provides a mechanism for displaying this approximation graphically in terms of a joint plot (or a biplot) of the objects and variables in a common r-dimensional Euclidean space. Specifically, if $\hat{\mathbf{X}} = \mathbf{GH}'$, where $\mathbf{G} = \mathbf{S}_{(r)}\mathbf{\Sigma}_{(r)}^{\alpha}$ and $\mathbf{H} = \mathbf{T}_{(r)}\mathbf{\Sigma}_{(r)}^{(1-\alpha)}$, where α is some chosen constant, $0 \leq \alpha \leq 1$, the rows of $\mathbf{G}$ provide r-dimensional coordinates for each of the p objects (which are typically represented as points in the joint plot); the rows of $\mathbf{H}$ provide r-dimensional coordinates for each of the q variables (which are typically represented as vectors in the joint plot). Irrespective of the chosen value for α, the approximating entry $\hat{x}_{ij}$ is the inner product of the ith row of $\mathbf{G}$ and the jth row of $\mathbf{H}$; consequently, considering just the column vector representing the jth variable in $\hat{\mathbf{X}}$, graphically these inner products are proportional to the lengths of the orthogonal projections of the objects onto the vector.[9]

The GSVD has probably found its most widespread use in the data analysis technique known as correspondence analysis, sometimes limited (unnecessarily) to the analysis of a two-way contingency table $\mathbf{F} = \{f_{ij}\}$, expressing the relationships between the categories of two categorical variables U and V, where the entry f_{ij} denotes the count of individual units of observation (objects) in category i of variable U and in category j of variable V. Classical "Analyse des Correspondances" à la Benzécri [1, 2], however, stands for a much more general technique to analyze any type of positive measure of correspondence; thus, correspondence analysis is best described as the analysis of a correspondence table, characterized as any table containing positive entries. Whatever interpretation is given to the technique, the marginals of the two-way table play an important role in that they define the metrics. In terms of lower-rank approximation, the product $\mathbf{GH}'$ of row scores $\mathbf{G}$ and column scores $\mathbf{H}$ approximates $\mathbf{M_r}^{-1}(\mathbf{F} - \frac{1}{N}\mathbf{M_r}\mathbf{u_r}\mathbf{u_c}'\mathbf{M_c})\mathbf{M_c}^{-1}$, where $\mathbf{M_r}$ and $\mathbf{M_c}$ denote diagonal matrices with the row and column marginals, respectively, along the main diagonal; $\mathbf{u_r}$ and $\mathbf{u_c}$ are vectors of ones of size R (number of rows) and C (number of columns);

---

[9]The joint representation of rows and columns as points and vectors in a common space originates with Tucker [43] and has found applications in psychometrics in the analysis of preference data (Carroll [4]) before the display became well known in applied statistics as the biplot (through Gabriel [14]). The prefix "bi" in the term "biplot" refers to two sets of different entities, objects and variables, and not to two dimensions, as is sometimes incorrectly assumed. This same connotation for the term "bi" is used by Kruskal [32] in his characterization of principal components analysis as a bilinear model.

and N is the total frequency.[10]

REFERENCES

[1] J.-P. BENZÉCRI, *L'Analyse des données*, Dunod, Paris, 1973.

[2] J.-P. BENZÉCRI, *Correspondence Analysis Handbook*, Marcel Dekker, New York, 1992.

[3] E. BODEWIG, *Matrix Calculus*, North–Holland, Amsterdam, 1956.

[4] J. D. CARROLL, *Individual differences and multidimensional scaling*, in Multidimensional Scaling: Theory and Applications in the Behavioral Sciences, R. N. Shepard, A. K. Romney, and S. B. Nerlove, eds., Seminar Press, New York, 1972, pp. 105–155.

[5] M. T. CHU, R. E. FUNDERLIC, AND G. H. GOLUB, *A rank-one reduction formula and its applications to matrix factorizations*, SIAM Rev., 37 (1995), pp. 512–530.

[6] R. E. CLINE AND R. E. FUNDERLIC, *The rank of a difference of matrices and associated generalized inverses*, Linear Algebra Appl., 24 (1979), pp. 185–215.

[7] W. W. COOLEY AND P. R. LOHNES, *Multivariate Data Analysis*, John Wiley, New York, 1971.

[8] B. L. R. DE MOOR AND G. H. GOLUB, *The restricted singular value decomposition: Properties and applications*, SIAM J. Matrix Anal. Appl., 12 (1991), pp. 401–425.

[9] F. DEUTSCH, *The method of alternating orthogonal projections*, in Approximation Theory, Spline Functions and Applications, S. P. Singh, ed., Kluwer Academic Publishers, Dordrecht, the Netherlands, 1992, pp. 105–121.

[10] R. L. DYKSTRA, *An algorithm for restricted least squares regression*, J. Amer. Statist. Assoc., 78 (1983), pp. 837–842.

[11] R. L. DYKSTRA AND R. ROBERTSON, *An algorithm for isotonic regression for two or more independent variables*, Ann. Statist., 10 (1982), pp. 708–716.

[12] C. ECKART AND G. YOUNG, *The approximation of one matrix by another of lower rank*, Psychometrika, 1 (1936), pp. 211–218.

[13] E. EGERVÁRY, *On rank-diminishing operators and their applications to the solution of linear equations*, Z. Angew. Math. Phys., 11 (1960), pp. 376–386.

[14] K. R. GABRIEL, *The biplot graphic display of matrices with application to principal components analysis*, Biometrika, 58 (1971), pp. 453–467.

[15] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, 3rd ed., The Johns Hopkins University Press, Baltimore, 1996.

[16] J. C. GOWER AND D. J. HAND, *Biplots*, Chapman and Hall, London, 1996.

[17] L. GUTTMAN, *General theory and methods for matric factoring*, Psychometrika, 9, (1944), pp. 1–16.

[18] L. GUTTMAN, *Multiple group methods for common-factor analysis: Their basis, computation, and interpretation*, Psychometrika, 17 (1952), pp. 209–222.

[19] L. GUTTMAN, *Some necessary conditions for common factor analysis*, Psychometrika, 19 (1954), pp. 149–161.

[20] L. GUTTMAN, *A necessary and sufficient formula for matrix factoring*, Psychometrika, 22 (1957), pp. 79–81.

[21] H. H. HARMAN, *Modern Factor Analysis*, The University of Chicago Press, Chicago, 1960.

[22] K. J. HOLZINGER, *A simple method of factor-analysis*, Psychometrika, 9 (1944), pp. 257–262.

[23] P. HORST, *Matrix Algebra for Social Scientists*, Holt, Rinehart and Winston, New York, 1963.

[24] P. HORST, *Factor Analysis of Data Matrices*, Holt, Rinehart and Winston, New York, 1965.

[25] H. HOTELLING, *Analysis of a complex of statistical variables into principal components*, J. Educ. Psych., 24 (1933), pp. 417–441 and 498–520.

[26] A. S. HOUSEHOLDER, *The Theory of Matrices in Numerical Analysis*, Blaisdell, New York, 1964.

[27] A. S. HOUSEHOLDER AND G. YOUNG, *Matrix approximations and latent roots*, Amer. Math. Monthly, 45 (1938), pp. 165–171.

---

[10]The Gower and Hand [16] reference, as well as the comprehensive survey of Rao [35], details many uses for the GSVD. We might mention another particular application, because of the general form of the rank reduction results discussed in the main body of the paper, which is discussed and used in some depth by Gower and Hand [16, pp. 245–246]: For given matrices $\mathbf{Y}$, $\mathbf{A}$, and $\mathbf{B}$ of orders $p \times q$, $p \times s$, and $q \times t$, respectively (where for convenience $s \leq p$ and $t \leq q$), the rank-r matrix $\mathbf{\Gamma}$ (which we can identify with the search for $\hat{\mathbf{X}}$ in (3.2)) of order $s \times t$ that minimizes $\text{Trace}[(\mathbf{Y} - \mathbf{A}\mathbf{\Gamma}\mathbf{B}')(\mathbf{Y} - \mathbf{A}\mathbf{\Gamma}\mathbf{B}')']$ is obtained from the GSVD of $(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{Y}\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}$ (which we can identify with $\mathbf{X}$ in (3.1)) in the metrics $\mathbf{P} = \mathbf{A}'\mathbf{A}$ and $\mathbf{Q} = \mathbf{B}'\mathbf{B}$ (and if the notation of (3.2) is used), with the normalizations of $\mathbf{S}'_{(\mathbf{r})}(\mathbf{A}'\mathbf{A})\mathbf{S}_{(\mathbf{r})} = \mathbf{I}$ and $\mathbf{T}'_{(\mathbf{r})}(\mathbf{B}'\mathbf{B})\mathbf{T}_{(\mathbf{r})} = \mathbf{I}$.

[28] L. J. Hubert and P. Arabie, *The analysis of proximity matrices through sums of matrices having (anti-)Robinson forms*, Brit. J. Math. Statist. Psych., 47 (1994), pp. 1–40.

[29] L. J. Hubert and P. Arabie, *Iterative projection strategies for the least-squares fitting of tree structures to proximity data*, Brit. J. Math. Statist. Psych., 48 (1995), pp. 281–317.

[30] L. J. Hubert, P. Arabie, and J. Meulman, *Linear and circular unidimensional scaling for symmetric proximity matrices*, Brit. J. Math. Statist. Psych., 50 (1997), pp. 253–284.

[31] S. Kaczmarz, *Angenäherte Auflösung von Systemen linearer Gleichungen*, Bull. Internat. Acad. Pol. Sci. Lett., A35 (1937), pp. 355–357.

[32] J. B. Kruskal, *Factor analysis and principal components analysis: Bilinear methods*, in International Encyclopedia of Statistics, W. H. Kruskal and J. M. Tanur, eds., The Free Press, New York, 1978, pp. 307–330.

[33] J. E. Overall, *Orthogonal factors and uncorrelated factor scores*, Psych. Rep., 10 (1962), 651–662.

[34] C. R. Rao, *Linear Statistical Inference and Its Applications*, John Wiley, New York, 1965.

[35] C. R. Rao, *Matrix approximations and reduction of dimensionality in multivariate statistical analysis*, in Multivariate Analysis, V, P. R. Krishnaiah, ed., North–Holland, Amsterdam, 1980, pp. 3–22.

[36] E. Schmidt, *Zur Theorie der linearen und nichtlinearen Integralgleichungen. I. Teil. Entwicklung willkürlichen Funktionen nach System vorgeschriebener*, Math. Ann., 63 (1907), pp. 433–476.

[37] G. W. Stewart, *On the early history of the singular value decomposition*, SIAM Rev., 35 (1993), pp. 551–566.

[38] Y. Takane, *The use of PSVD and QSVD in psychometrics*, in Bull. Internat. Statist. Inst., Proceedings of the 51st Session, Tome LVII, Book 2, 1997, pp. 255–258.

[39] Y. Takane, *personal communication*, June 17, 1998.

[40] L. L. Thurstone, *The Vectors of the Mind*, University of Chicago Press, Chicago, 1935.

[41] L. L. Thurstone, *A multiple group method of factoring the correlation matrix*, Psychometrika, 10 (1945), pp. 73–78.

[42] L. L. Thurstone, *Note about the multiple group method*, Psychometrika, 14 (1949), pp. 43–45.

[43] L. R. Tucker, *Intra-individual and inter-individual multidimensionality*, in Psychological Scaling: Theory and Applications, H. Guilliksen and S. Messick, eds., John Wiley, New York, 1960, pp. 155–167.

[44] I. A. Van der Lans, *Nonlinear Multivariate Analysis for Multiattribute Preference Data*, DSWO Press, Leiden, the Netherlands, 1992.

[45] C. F. Van Loan, *Generalizing the singular value decomposition*, SIAM J. Numer. Anal., 13 (1976), pp. 76–83.

[46] J. H. M. Wedderburn, *Lectures on Matrices, Colloquium Publications*, Vol. 17, American Mathematical Society, New York, 1934.

[47] G. Young, *Maximum likelihood estimation and factor analysis*, Psychometrika, 6 (1941), pp. 49–53.

[48] G. Young and A. S. Householder, *Discussion of a set of points in terms of their mutual distances*, Psychometrika, 3 (1938), pp. 11–22.