The Singular-Value Theorem for $m \times n$ Matrices, Canonical Correlations, and Moore-Penrose Inverses

Stanley Sawyer — Washington University — January 8, 2007

1. Introduction We first prove the singular value decomposition theorem for matrices and then give two applications in statistics.

Theorem 1. *(Singular-Value Decomposition)* Let A be an arbitrary $m \times n$ matrix. Then there exist positive constants $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_r > 0$ for some integer r such that

$$A = R_1 D R_2' \tag{1.1}$$

where R_1 and R_2 are orthogonal matrices (R_1 is $m \times m$ and R_2 is $n \times n$) and D is the $m \times n$ matrix

$$D = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 & 0 & \ldots & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & & & \\ 0 & 0 & \ldots & \lambda_r & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & & \end{pmatrix} \tag{1.2}$$

The matrices R_1, D, R_2 are unique except for rotations of eigenspaces.

Proof. By the spectral theorem for symmetric matrices, there exist orthonormal bases $x_1, \ldots, x_m$ in R^m and $y_1, \ldots, y_n$ in R^n such that

$$\begin{aligned} (AA')x_i &= \mu_i x_i \quad (1 \le i \le m) \quad \text{and} \\ (A'A)y_j &= \nu_j y_j \quad (1 \le j \le n) \end{aligned} \tag{1.3}$$

In (1.3), $\mu_1 \ge \mu_2 \ge \ldots \mu_{r_1} > 0$ and $\mu_i = 0$ for $r_1 < i \le n$ for some $r_1 \le n$ with eigenvalues repeated according to multiplicity. Similarly, $\nu_1 \ge \nu_2 \ge \ldots \nu_{r_2} > 0$ and $\nu_j = 0$ for $r_2 < j \le m$ for some $r_2 \le m$. Assume $x_i' x_i = 1$ and $y_j' y_j = 1$ for definiteness.

Let $z_j = Ay_j$ for some $j \le r_2$. Then $(AA')z_j = A(A'A)y_j = A\nu_j y_j = \nu_j A y_j = \nu_j z_j$, so that $(AA')z_j = \nu_j z_j$. Also,

$$z_j' z_j = (Ay_j)'(Ay_j) = y_j'(A'Ay_j) = \nu_j y_i' y_i = \nu_j > 0$$

so that $z_j \ne 0$. Thus $z_j = Ay_j$ is an eigenvector of AA' with eigenvalue ν_j. This means that $\nu_j = \mu_k$ for some $k \le r_1$ and that $z_j = Ay_j$ is in the eigenspace of AA' for μ_k.

Similarly, $w_i = A'x_i$ satisfies $(A'A)w_i = A'(AA')x_i = A'\mu_i x_i = \mu_i A'x_i = \mu_i w_i$ and $w_i'w_i = x_i'(AA'x_i) = \mu_i x_i'x_i = \mu_i > 0$. Thus $w_i = A'x_i$ is an eigenvector of $A'A$ with eigenvalue μ_i, so that $\mu_i = \nu_k$ for some $k \le r_2$.

This implies that the sets $\{\mu_i\} = \{\nu_j\}$ for positive eigenvalues, but we cannot yet conclude that $\mu_i = \nu_i$ without accounting for multiple linearly-independent eigenvectors for the same eigenvalue. Assume that y_a, y_b are orthogonal eigenvectors of $A'A$ for the same eigenvalue and set $z_a = Ay_a$ and $z_b = Ay_b$. Then $z_a'z_b = y_a'(A'A)y_b = \nu_j y_a'y_b = 0$. This implies that the mapping $y \to Ay$ maps a basis for the eigenspace of ν_j for $A'A$ onto a basis for a subspace of the eigenspace for AA' for the same eigenvalue. The mapping $x \to A'x$ behaves similarly for eigenspaces of AA'. This means that the positive eigenvalues of $A'A$ and AA' also correspond taking into account multiplicity. Thus $r_1 = r_2 = r$ where r is the common value and $\mu_i = \nu_i$ for $1 \le i \le r$. The other eigenvalues of $A'A$ and AA' are zero.

Since $y \to Ay$ maps a basis for any positive eigenspace of $A'A$ onto a basis for eigenspace for the same eigenvalue for AA', it follows that we can assume that $Ay_i = c_i x_i$ for x_i, y_i in (1.3) for constants $c_i \ne 0$. Then $(Ay_i)'Ay_i = y_i'(A'Ay_i) = y_i'\nu_i y_i = \nu_i > 0$ since $y_i'y_i = 1$ and $(c_i x_i)'c_i x_i = c_i^2 x_i'x_i = c_i^2$ since $x_i'x_i = 1$. Thus $c_i^2 = \nu_i$ for $1 \le i \le r$. Since x_i in (1.3) could have been replaced by $-x_i$, we can assume $c_i > 0$, so that $c_i = \sqrt{\nu_i}$. Exactly the same argument holds for the mapping $x \to A'x$. Thus we have shown that

$$Ay_i = \lambda_i x_i \quad \text{and} \quad A'x_i = \lambda_i y_i \quad \text{for} \quad \lambda_i = \sqrt{\mu_i}, \quad 1 \le i \le r \tag{1.4}$$

We can use the Gramm-Schmidt process to extend y_j to a basis for R^m such that $Ay_i = 0$ for $r < j \le m$ and x_i to a basis for R^n such that $A'x_i = 0$ for $r < i \le n$.

If we write $y = \sum_{i=1}^n c_i y_i$ for an arbitrary $y \in R^n$, then

$$Ay = A\Big(\sum_{i=1}^n c_i y_i\Big) = \sum_{i=1}^n c_i A y_i = \sum_{i=1}^r c_i \lambda_i x_i$$

by (1.4). Similarly, if $Q = \sum_{j=1}^r \lambda_j x_j y_j'$,

$$Qy = \sum_{j=1}^r \sum_{i=1}^n c_i \lambda_j x_j y_j' y_i = \sum_{j=1}^r c_j \lambda_j x_j$$

since $y_j'y_i = 0$ if $j \ne i$ and $y_i'y_i = 1$. Thus $Ay = Qy$ for all $y \in R^n$. This implies $A = Q$ and

$$A \;=\; \sum_{i=1}^r \lambda_i x_i y_i' \tag{1.5}$$

Finally, define matrices

$$\begin{aligned} R_1 &= (\, x_1 \quad x_2 \quad \dots \quad x_m \,), \quad x_i \in R^m, \quad \text{and} \\ R_2 &= (\, y_1 \quad y_2 \quad \dots \quad y_n \,), \quad y_j \in R^n \end{aligned} \tag{1.6}$$

for column vectors x_i, y_j. Then R_1 is an $m \times m$ orthogonal matrix and R_2 is an $n \times n$ orthogonal matrix. Set $Q = R_1 D (R_2)'$ where D is the $m \times n$ matrix in (1.2). I claim that $Qy_j = Ay_j$ for $1 \le j \le n$. This follows from

$$Qy_j \;=\; R_1 D (R_2)' y_j \;=\; R_1 D \begin{pmatrix} y_1' \\ \dots \\ y_n' \end{pmatrix} y_j \;=\; R_1 D \begin{pmatrix} y_1' y_j \\ \dots \\ y_n' y_j \end{pmatrix} \;=\; R_1 D e_j$$

where e_j is the j^{th} unit vector in R^n. Then $De_j = \lambda_j e_j$ by (1.2) and

$$Qy_j \;=\; \lambda_j R_1 e_j \;=\; \lambda_j \, (\, x_1 \quad \dots \quad x_m \,) \begin{pmatrix} 0 \\ \dots \\ 1 \\ \dots \\ 0 \end{pmatrix} \;=\; \lambda_j x_j$$

However $\lambda_j x_j = Ay_j$ by (1.4), so that $Ay_j = Qy_j$ for $1 \le j \le n$. Thus $A = Q$ since $\{y_j\}$ is a basis and $A = Q = R_1 D (R_2)'$. This completes the proof of existence in Theorem 1.

The uniqueness of R_1, D, and R_2 follows from the identities

$$AA' = R_1 (DD') R_1' \qquad \text{and} \qquad A'A = R_2 (D'D)(R_2)'$$

in R^m and R^n, respectively. Thus the uniqueness of R_1 and R_2 follows from the uniqueness of the orthogonal matrix in the spectral theorem for symmetric matrices.

An equivalent form of Theorem 1 is

Corollary 1.1. Let A be an arbitrary $m \times n$ matrix. Then, there exists an integer r, orthonormal vectors $x_1, \dots, x_r$ in R^m, orthonormal vectors $y_1, \dots, y_r$ in R^m, and positive constants $\lambda_i > 0$ such that

$$A \;=\; \sum_{i=1}^{r} \lambda_i x_i y_i' \tag{1.7}$$

If the λ_i are distinct, the vectors x_i, y_j and constants $\lambda_i > 0$ in (1.7) are unique.

Exercise 1.1: Prove that Corollary 1.1 implies Theorem 1. (Note that (1.7) is the same as equation (1.5) in the proof of the theorem, so that Theorem 1 imples Corollary 1.1.)

We can use Theorem 1 to obtain an analog of the polar decomposition of complex numbers for matrices:

Corollary 1.2. Any $n \times n$ matrix A can be written

$$A = UM \tag{1.8}$$

where U is an $n \times n$ orthogonal matrix and M is positive semidefinite. The matrices U and M in (1.8) are unique except for rotations of the eigenspaces of M.

Proof. By (1.1), $A = R_1 D(R_2)' = (R_1(R_2)')(R_2 D(R_2)') = UM$ for $U = R_1(R_2)'$ and $M = R_2 D(R_2)'$.

2. Canonical Correlations Let $Y \in R^m$ and $X \in R^n$ be two vector-valued random variables. Assume for definiteness that the $m \times m$ matrix $A = \mathrm{Var}(Y)$ and the $n \times n$ matrix $B = \mathrm{Var}(X)$ are both invertible. Let C be the $m \times n$ matrix with entries

$$C_{ij} = \mathrm{Cov}(Y_i, X_j), \qquad 1 \le i \le m,\ 1 \le j \le n$$

The *first canonical correlation* of the vectors Y and X is

$$\lambda_1 \;=\; \max_{u,v} \mathrm{Corr}(u'Y, v'X) \;=\; \max_{u,v} \frac{\mathrm{Cov}(u'Y, v'X)}{\sqrt{\mathrm{Var}(u'Y)\,\mathrm{Var}(v'X)}} \tag{2.1}$$

Note $\lambda_1 \ge 0$ since the signs of u, v are arbitrary. Then

Theorem 2. The canonical correlation λ_1 in (2.1) is the largest diagonal entry λ_i in the matrix D in the singular value decomposition of the $m \times n$ matrix

$$G = A^{-1/2} C B^{-1/2} = R_1 D(R_2)' \tag{2.2}$$

in Theorem 1. Equivalently, the principal eigenvalue of the $m \times m$ matrix GG' and the $n \times n$ matrix $G'G$ is λ_1^2.

Remark. The matrix $E = \mathrm{Corr}(Y, X)$ with entries $E_{ij} = \mathrm{Corr}(Y_i, X_j)$ can be written $E = \big(\mathrm{Diag}(A)\big)^{-1/2} C \big(\mathrm{Diag}(B)\big)^{-1/2}$, which is not the same as the matrix G in (2.2). This is not a contradiction since, among other reasons, $\mathrm{Corr}(u'Y, v'X)$ is not a linear function of u and v.

Proof of Theorem 2. This will depend on

Lemma 2.1. If C is an arbitrary $m \times n$ matrix, then

$$\max_{u,v} \frac{u'Cv}{\sqrt{(u'u)(v'v)}} \;=\; \lambda_1 \tag{2.3}$$

where λ_1^2 is the largest eigenvalue of the matrices $C'C$ and CC'.

Proof of Theorem 2 given Lemma 2.1. If $A = \mathrm{Var}(Y)$, $B = \mathrm{Var}(X)$, and $C = \mathrm{Cov}(Y,X)$ as before, then

$$\begin{aligned}\max_{u,v} \frac{\mathrm{Cov}(u'Y, v'X)}{\sqrt{\mathrm{Var}(u'Y)\,\mathrm{Var}(v'X)}} \;&=\; \max_{u,v} \frac{u'Cv}{\sqrt{(u'Au)(v'Bv)}} \\ &=\; \max_{u,v} \frac{(A^{1/2}u)'(A^{-1/2}CB^{-1/2})(B^{1/2}v)}{\sqrt{(A^{1/2}u)'(A^{1/2}u)}\sqrt{(B^{1/2}v)'(B^{1/2}v)}} \\ &=\; \max_{u_1,v_1} \frac{u_1'Gv_1}{\sqrt{(u_1'u_1)(v_1'v_1)}} \;=\; \lambda_1\end{aligned}$$

for G in (2.2), where λ_1^2 is the principal eigenvalue of GG' and $G'G$ by Lemma 2.1.

This reduces the proof of Theorem 2 to the proof of Lemma 2.1, for which we give two proofs:

First Proof of Lemma 2.1. It is sufficient to restrict u, v in (2.3) so that $u'u = v'v = 1$. Then, by the method of Lagrange multipliers, the maximum is attained at a stationary point of

$$\phi(u,v) = u'Cv - \lambda(u'u) - \mu(v'v)$$

for constants λ and μ. At such a stationary point,

$$\frac{\partial\phi}{\partial u_i} = (Cv)_i - 2\lambda u_i = 0 \qquad \text{and} \qquad \frac{\partial\phi}{\partial v_j} = (u'C)_j - 2\mu v_j = 0$$

Thus $Cv = 2\lambda u$ and $C'u = 2\mu v$. Since $u'Cv = (C'u)'v$,

$$\begin{aligned}u'Cv &= u'(2\lambda u) = 2\lambda(u'u) = 2\lambda \\ &= (C'u)'v = (2\mu v)'v = 2\mu(v'v) = 2\mu\end{aligned} \tag{2.4}$$

Hence the maximum in (2.3) is $u'Cv = \lambda_1 = 2\lambda = 2\mu$ and $\lambda = \mu$. Similarly

$$(C'C)v = C'(Cv) = 2\lambda C'u = 4\lambda\mu v = \lambda_1^2 v$$
$$(CC')u = C(C'u) = 2\mu Cv = 4\lambda\mu u = \lambda_1^2 u$$

Thus λ_1^2 is the common largest eigenvalue of $C'C$ and CC'.

Second Proof of Lemma 2.1. By Theorem 1, we can write

$$C = R_1' D R_2$$

where R_1 and R_2 are orthogonal and D is as in (1.2). Then

$$\max_{u,v} \frac{u'Cv}{\sqrt{(u'u)(v'v)}} = \max_{u,v} \frac{(R_1u)'D(R_2v)}{\sqrt{(u'u)(v'v)}} = \max_{u_1,v_1} \frac{u_1'Dv_1}{\sqrt{(u_1'u_1)(v_1'v_1)}}$$

where $u_1 = R_1u$ and $v_1 = R_2v$. However, it follows from (1.2) that

$$\max\{\, u_1'Dv_1 : u_1'u_1 = v_1'v_1 = 1 \,\} = \lambda_1$$

where λ_1 is the largest value in the matrix D. It then follows from the proof of Theorem 1 that λ_1^2 is the largest eigenvalue of both $C'C$ and CC'.

Let $u_1 \in R^m$ and $v_1 \in R^n$ be vectors at which the maximum defining the first canonical correlation in (2.1) is attained. The *second canonical correlation* is

$$\lambda_2 = \max_{u,v}\Big\{ \operatorname{Corr}(u'Y, v'X) \;\Big|\; u'u_1 = v'v_1 = 0 \Big\} \tag{2.5}$$

Exercise 2.1: Show that if λ_2 is the second canonical correlation defined in (2.5), then λ_2^2 is the *second-largest* eigenvalue of the matrices $G'G$ and GG' for G in (2.2).

Exercise 2.2: Show that Theorem 2 remains valid if one or both of the matrices A and B are not invertible. (*Hint*: Show that you can use appropriate generalized inverses of A and B.)

3. Moore-Penrose Inverses By definition, if A is an $m\times n$ matrix, then G is a *Moore-Penrose generalized inverse* of A if G is $n\times m$ and A and G satisfy the four conditions:

$$\begin{array}{rl} \text{(i)} & AGA = A \\ \text{(ii)} & GAG = G \\ \text{(iii)} & (AG)' = AG \\ \text{(iv)} & (GA)' = GA \end{array} \tag{3.1}$$

Condition (i) in (3.1) implies that $AGAG = AG$ and $GAGA = GA$, so that AG and GA are both projections (in R^m and R^n, respectively). Conditions (iii-iv) then say that AG and GA are both *orthogonal* projections.

It follows from (i) that if y is in the range of A, then $x = Gy$ is a solution of $Ax = y$. A necessary and sufficient condition for y being in the range of A is that $(AG)y = y$.

Our next result states that Moore-Penrose inverses always exist:

Lemma 3.1. (Moore-Penrose Inverses) Let $A = R_1D(R_2)'$ be the $m \times n$ matrix in (1.1) and let G be the $n \times m$ matrix

$$G = R_2E'(R_1)' \tag{3.2}$$

In (3.2), E is the same as D in (1.2) except that each positive constant $\lambda_i > 0$ is replaced by $1/\lambda_i > 0$ in E. Then G is a Moore-Penrose generalized inverse of A.

Proof. Note $AGA = R_1D(R_2)'R_2E'(R_1)'R_1D(R_2)' = R_1DE'D(R_2)' = R_1D(R_2)' = A$ and similarly $GAG = G$. Also $AG = R_1D(R_2)'R_2E'(R_1)' = R_1DE'(R_1)'$, which is an orthogonal projection of rank r in R^m. Similarly, $GA = R_2E'(R_1)'R_1D(R_2)' = R_2E'D(R_2)'$ is an orthogonal projection in R^n.

Theorem 3. *(Uniqueness of Moore-Penrose Inverses)* Let A be an arbitrary $m \times n$ matrix. Then, the matrix G in (3.2) is the unique Moore-Penrose inverse of A.

Proof. Assume that G satisfies the conditions (i-iv) of the definition. Let $y_1, \ldots, y_n$ and $x_1, \ldots, x_m$ be the orthonormal bases in the proof of Theorem 1 for the singular-value decomposition applied to A. In particular $Ay_i = \lambda_i x_i$ and $A'x_i = \lambda_i y_i$ by (1.4). Since $y_1, \ldots, y_n$ is a basis for R^n, we can write

$$Gx_i \;=\; \sum_{k=1}^{n} c_{ik} y_k, \qquad 1 \le i \le m \tag{3.3}$$

The proof will be to show that, in fact,

$$Gx_i \;=\; \begin{cases} (1/\lambda_i)y_i & \text{if } 1 \le i \le r \\ 0 & r+1 \le i \le m \end{cases}$$

This implies that $G = R_2E'(R_1)'$ exactly as in the proof of $A = R_1D(R_2)'$ in Theorem 1. We will use the four conditions (i-iv) in the definition of a Moore-Penrose inverse in turn.

From $AGA = A$ and $Ay_i = \lambda_i x_i$, it follows that $AGAy_i = \lambda_i AGx_i = Ay_i = \lambda_i x_i$ and

$$(AG)x_i = x_i, \qquad 1 \le i \le r \tag{3.4}$$

Thus by (3.3) $x_i = A(Gx_i) = \sum_{k=1}^{n} c_{ik} Ay_k = \sum_{k=1}^{r} c_{ik}\lambda_k x_k = x_i$. Since $\{x_k\}$ is a basis, this implies $c_{ik} = 0$ for $1 \le i \le r$ for $k \ne i$ and $c_{ii} = 1/\lambda_i$. Thus by (3.3)

$$Gx_i = \frac{1}{\lambda_i} y_i + \sum_{k=r+1}^{n} c_{ik} y_k, \qquad 1 \le i \le r \tag{3.5}$$

From $AG = (AG)'$ and (3.4), it follows that AG must preserve the vector space spanned by $x_{r+1}, \dots, x_m$. Thus by (3.3)

$$\begin{aligned} (AG)x_i &= \sum_{k=r+1}^{m} d_{ik} x_k, \quad r+1 \le i \le m, \\ &= A(Gx_i) = A\Big(\sum_{k=1}^{n} c_{ik} y_k\Big) = \sum_{k=1}^{n} c_{ik} Ay_k = \sum_{k=1}^{r} \lambda_k c_{ik} x_k \end{aligned}$$

Thus $(AG)x_i = 0$ for $r+1 \le i \le m$ and also $c_{ik} = 0$ in (3.3) for $1 \le k \le r < i \le m$. In particular

$$Gx_i = \sum_{k=r+1}^{n} c_{ik} y_k, \qquad r+1 \le i \le m \tag{3.6}$$

From $(AG)x_i = x_i$ ($1 \le i \le r$) and $(AG)x_i = 0$ ($r < i \le m$) it follows that $AG = R_1(DE')(R_1)'$, but we will not need this.

From $GAG = G$ and (3.6), it follows that

$$Gx_i = GAGx_i = \sum_{k=r+1}^{n} c_{ik} GAy_k = 0, \qquad r+1 \le i \le m \tag{3.7}$$

From $GA = (GA)'$ and $GAy_j = G(Ay_j) = 0$ for $r+1 \le j \le n$, it follows that

$$\lambda_i Gx_i = GAy_i = \sum_{k=1}^{r} d_{ik} y_k, \qquad 1 \le i \le r \tag{3.8}$$

Thus $Gx_i = (1/\lambda_i)y_i$ for $1 \le i \le r$ by (3.5), and

$$Gx_i = \begin{cases} (1/\lambda_i)x_i & \text{for } 1 \le i \le r, \text{ by (3.5) and (3.8),} \\ 0 & r+1 \le i \le m \text{ by (3.7)} \end{cases}$$

This implies that $G = R_2E'(R_1)'$ by arguing as in the proof of Theorem 1.