Visualization of Matrix Singular Value Decomposition

CLIFF LONG
Bowling Green State University
Bowling Green, OH 43403

A real matrix is frequently used as a finite representation of a real function of two variables, especially as a tool for studying continuous functions in numerical analysis and computer graphics. It is also advantageous to use continuous functions to provide visualization for matrix techniques such as **singular value decomposition** (SVD). We will illustrate how this factorization technique can be thought of as providing least square best fit approximations to functions of two variables. The basic theory of SVD (sometimes called basic structure of a matrix) will be presented, one simple example given for clarification (similar to those found in [**7**]), and then a matrix representation of a sculptured head of Abe Lincoln will be used to illustrate the geometry involved. For ease in understanding, we'll restrict our attention to real matrices and refer the reader to [**2**], [**4**], [**9**], and [**11**] for the proofs.

Singular value decomposition of a matrix is a technique which represents any given matrix as a sum of rank 1 matrices, i.e., it yields a finite series expansion for a matrix. For example, the matrix

$$A=\begin{bmatrix} 3.01 & 0.01 & -2.99 \\ 2.99 & -0.01 & -3.01 \\ 2.00 & -4.00 & 2.00 \end{bmatrix} \tag{1}$$

can be written as the sum of three rank 1 matrices in a rather obvious decomposition:

$$A=\begin{bmatrix} 3 & 0 & -3 \\ 3 & 0 & -3 \\ 0 & 0 & 0 \end{bmatrix}+\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 2 & -4 & 2 \end{bmatrix}+\begin{bmatrix} .01 & .01 & .01 \\ -.01 & -.01 & -.01 \\ 0 & 0 & 0 \end{bmatrix}.$$

A less obvious decomposition (which results from the theorem stated below) is:

$$A=6\begin{bmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix}+2\sqrt{6}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix}+\frac{\sqrt{6}}{100}\begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \\ 0 & 0 & 0 \end{bmatrix}. \tag{2}$$

In matrix singular value decomposition, the matrix need not be square nor real, and the rank 1 matrices are chosen, normalized and ordered for usefulness in solving problems. The theory of singular value decomposition is not new (according to [**10**, p. 78] it was established for real and square matrices in the 1870s by Beltrami and Jordan and later developments are referenced in [**8**]). However, its current importance and extensive use is due to the existence of an efficient and numerically stable algorithm developed by Golub in the 1960s ([**5**], [**6**]). The technique is regularly used in solving least square problems and computing pseudoinverses of matrices. It is certain to be used even more extensively now that good computer programs are readily available (e.g., Moler [**4**] and software packages such as EISPACK, LINPACK and IMSL). An application to digital image processing by Andrews and Patterson [**1**] inspired my interest in SVD, and comments such as [it is] "The most reliable method for computing the coefficients for general least square problems..." [**4**, p. 195] and "... it is not nearly as famous as it should be" [**11**, p. 142] have kept me going.

The key theorem for SVD of matrices is the following.

THEOREM. *Any real matrix A can be factored as $A = PSQ^T$, where P and Q are orthogonal and S is diagonal with diagonal elements $\sigma_i \geqslant 0$ (called the singular values of A)* [**9**, *p. 18*].

COROLLARY. *Any real $m \times n$ matrix A can be expressed as a finite sum of rank 1 matrices in normalized form, that is, $A = \sigma_1 R_1 + \sigma_2 R_2 + \cdots + \sigma_k R_k$, where $k = \min(m, n)$ and*

(1) $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0 = \sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n$, *rank* $A = r \leqslant k$.
(2) $R_i = \bar{p}_i \bar{q}_i^T$ *where $\bar{p}_i$ is the ith column of P and a unit eigenvector of AA^T, and $\bar{q}_i$ is the ith column of Q and a unit eigenvector of A^TA.*
(3) *each R_i has the sum of the squares of its elements equal to* 1 (*this follows from* 2).

The proofs of these results depend on the fact that A^TA and AA^T are real, square and symmetric, and each has nonnegative eigenvalues and a complete set of orthogonal eigenvectors. (In fact A^TA and AA^T have precisely the same nonzero eigenvalues and the square roots of these are singular values σ_i, $1 \leqslant i \leqslant k$ of both A and A^T.)

To illustrate how the decomposition stated in the Corollary proceeds from the Theorem, consider our previous example. The matrix A is first factored as in the Theorem:

$$A = PSQ^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 6 & 0 & 0 \\ 0 & 2\sqrt{6} & 0 \\ 0 & 0 & \frac{\sqrt{6}}{100} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}. \tag{3}$$

Next, this factorization can be written as the sum $\Sigma\, \sigma_i \bar{p}_i q_i^T$, where the σ_i are the diagonal entries of S, $\bar{p}_i$ the ith column of P and q_i the ith row of Q.

$$A = 6 \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix} + 2\sqrt{6} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix}$$
$$+ \frac{\sqrt{6}}{100} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}. \tag{4}$$

Multiplication of the $\bar{p}_i \bar{q}_i^T$ yields the desired decomposition $A = \sigma_1 R_1 + \sigma_2 R_2 + \sigma_3 R_3$, which we have noted in (2). As an illustration of some of the matrix ideas used in obtaining the Corollary from the Theorem, we now establish part 2 for a square matrix A and for $\bar{q}_1$, the first column of Q, i.e., we'll show that $A^TA\bar{q}_1 = \sigma_1^2 \bar{q}_1$. Since P is orthogonal and S diagonal, we have $A^TA = (PSQ^T)^T(PSQ^T) = (QS^TP^T)(PSQ^T) = QS^TP^{-1}PSQ^{-1} = QS^TSQ^{-1} = QS^2Q^{-1}$. But then, if $\bar{e}_1$ is the column vector $[1 \quad 0 \quad \cdots \quad 0]^T$, we have

$$A^TA\bar{q}_1 = QS^2Q^{-1}\bar{q}_1 = QS^2\bar{e}_1 = Q\sigma_1^2\bar{e}_1 = \sigma_1^2\bar{q}_1.$$

It is equally easy to show that the matrix $R_i = \bar{p}_i \bar{q}_i^T$ has the sum of the squares of its elements equal to 1 (i.e., it has Frobenius norm $\|R\|_F$ equal to 1). This property allows the SVD of a matrix A of rank r to be used to find an $m \times n$ matrix B of rank $l < r$ that minimizes $\|B - A\|_F$. This

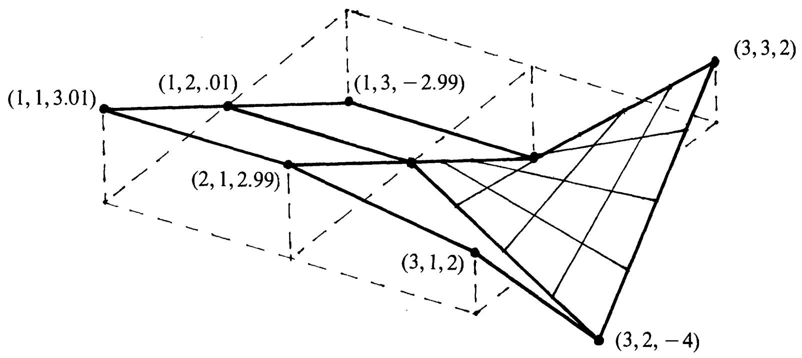

FIGURE 1

problem was posed and solved by Eckart and Young [**3**] who show that the first l terms of the SVD of A sum up to just such a matrix B [**9**, p. 26]. Thus the sum of the first l terms of the singular value decomposition is a best rank l approximation to the matrix A in the sense of the Frobenius norm (i.e., in the least squares sense). In fact,

$$\|A - \sigma_1 R_1 - \sigma_2 R_2 - \cdots - \sigma_l R_l\|_F = \sigma_{l+1}^2 + \cdots + \sigma_r^2.$$

While the Corollary suggests a way to determine the factorization $A = PSQ^T$, this method is not numerically stable for nearly singular matrices and should be replaced by an algorithmic approach such as that of Golub.

Now for some geometry! To a real $m \times n$ matrix A, we can associate a surface containing the points (x_i, y_j, a_{ij}) where (x_i, y_j) are lattice points on a rectangular grid. For example the matrix A given by (1) can be associated to the surface shown in FIGURE 1. This surface can be thought of as a piecewise hyperbolic function $z = ax + bxy + cy + d$ on the 3×3 grid points (x_i, y_j, a_{ij}).

Alternatively, given a continuous real function f defined over a rectangular grid, we may associate a real matrix A with entries the function values $f(x_i, y_j) = a_{ij}$ and treat the matrix A as a finite approximation to the surface $z = f(x, y)$. The singular value decomposition of this matrix A then gives a further approximation to the surface. Conversely, the related surface can be used to "visualize" the singular value decomposition. (Similar visualization techniques have been used for one-variable Taylor series and Fourier series expansions and should be utilized more often in the two-variable setting now that 3D computer graphics programs are more readily available.)

For illustrative purposes, we obtained a finite approximation to a bust of Abe Lincoln (using a crude homemade scanning device which allowed for a 49×36 matrix). The original sculpture and finite approximation (called ABE) are shown in FIGURE 2. The related matrix A was then factored to produce a finite expansion of ABE using rank 1 matrices from a SVD. The surfaces shown in FIGURE 3 represent surface approximations to ABE by keeping only a specified number of terms from the finite series expansion. The surface marked A_1 represents the approximation $A \doteq \sigma_1 R_1$, the surface A_2 represents the approximation of A by the two-term decomposition $A \doteq \sigma_1 R_1 + \sigma_2 R_2$, and so on. It is somewhat surprising that the rank 5 approximation to ABE (of a possible 36) is so recognizable. This means that the tail-end terms of the series $A = \sum \sigma_i R_i$ are not all that important, and suggests that the matrix might be somewhat ill-conditioned (actually $\sigma_1/\sigma_6 \doteq 33$ and the ratio σ_1/σ_{36}, called the condition number, is approximately 2000). Note that two different approximation techniques are used on the original sculpture. The first is the grid size which determines the matrix size $m \times n$. The second is related to the relative sizes of the singular values σ_i of the matrix. Thus our first step reduced ABE to 49×36 real numbers and our second step for surface A_5 reduced him to just $5 \times (49 + 36) + 5 = 430$ real numbers.

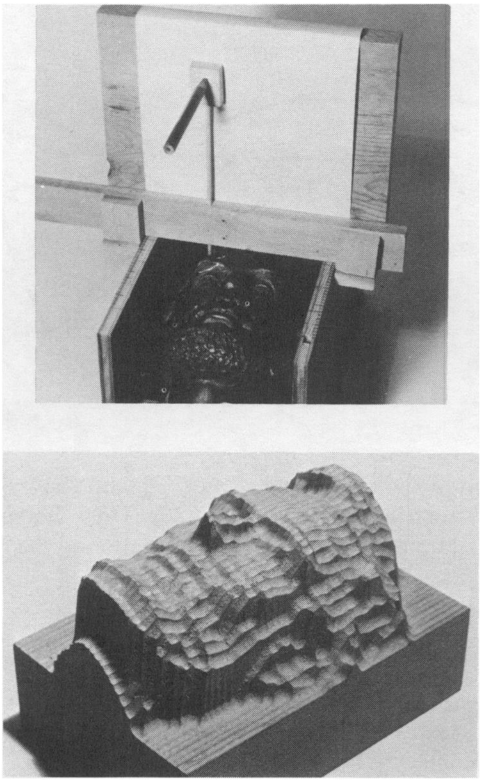

FIGURE 2

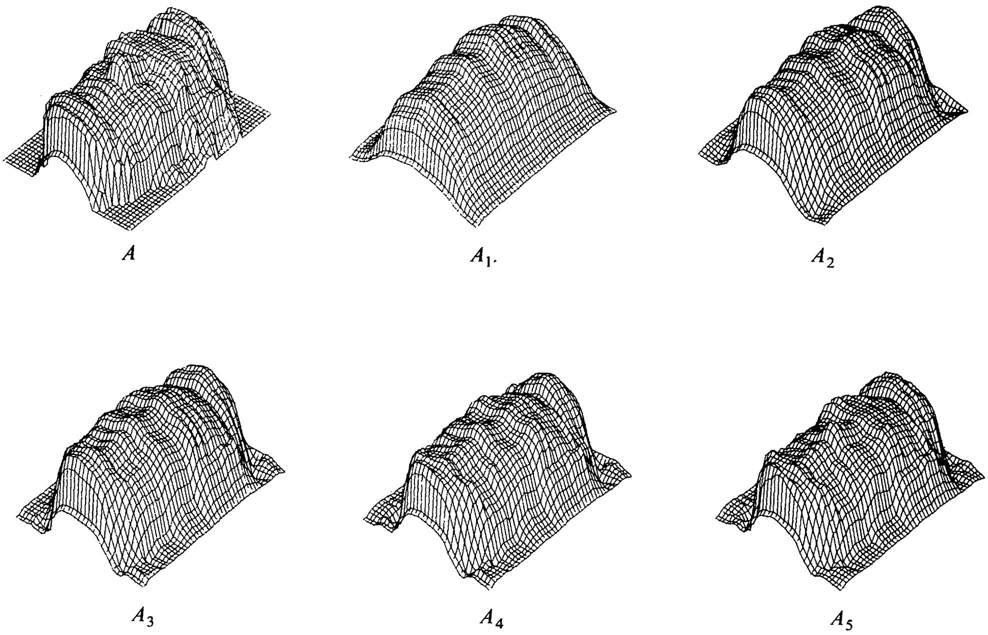

FIGURE 3. Singular value decomposition of "*ABE.*"

These procedures for reducing a continuous image to a finite set of real numbers are of particular importance in image processing techniques [**1**]. Specifically, dropping the tail-end terms of the singular value decomposition can be associated with eliminating the "snowy" feature of a TV picture (i.e., noise elimination from a picture transmission). The essential information of the picture is carried by the earlier terms of the decomposition and associated with the larger singular values, while the more random noise (unless of significant size) is associated with the smaller singular values and discarded.

The omission of small singular values is also significant for handling problems involving inverses of ill-conditioned matrices. (Many least square approximation problems fall into this category.) If a matrix A is decomposed as in the Theorem, $A = PSQ^T$, and if A^{-1} exists, then since P is orthogonal and S is diagonal, it follows that $A^{-1} = (PSQ^T)^{-1} = QS^{-1}P^T$ where S^{-1} is a diagonal matrix with ith diagonal entry σ_i^{-1}. This is a factorization of A^{-1} as in the Theorem, so the Corollary applies. Thus if we know the singular value decomposition of a nonsingular matrix, then we also have a decomposition of A^{-1}. For example, the matrix A in (1) is shown in factored form in (3), and its SVD derived in (4). From the above discussion, we have

$$
\begin{aligned}
A^{-1} &= \begin{bmatrix} 3.01 & .01 & -2.99 \\ 2.99 & -.01 & -3.01 \\ 2 & -4 & 2 \end{bmatrix}^{-1} \\
&= \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} \frac{1}{6} & 0 & 0 \\ 0 & \frac{\sqrt{6}}{12} & 0 \\ 0 & 0 & \frac{100}{6}\sqrt{6} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix} \\
&= \frac{1}{6} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix} + \frac{\sqrt{6}}{12} \begin{bmatrix} \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \\
&\quad + \frac{100}{6}\sqrt{6} \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix} \\
&= \frac{1}{6} R_1^T + \frac{\sqrt{6}}{12} R_2^T + \frac{100\sqrt{6}}{6} R_3^T \\
&= \sigma_1^{-1} R_1^T + \sigma_2^{-1} R_2^T + \sigma_3^{-1} R_3^T.
\end{aligned}
$$

This shows that the least significant rank 1 matrix in the SVD of A (i.e., R_3 which has smallest

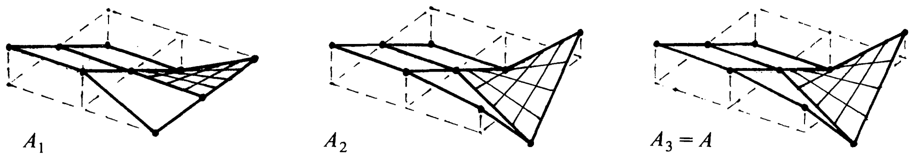

FIGURE 4(a). **Rank r approximations, A_r, to A.**

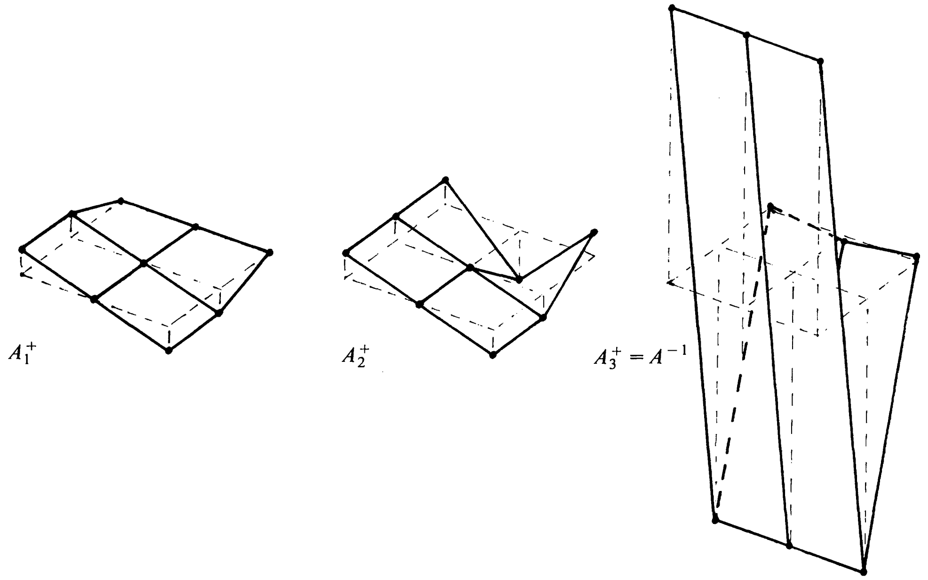

FIGURE 4(b). **Corresponding pseudoinverses, A_r^+ (not to scale).**

coefficient σ_3) becomes the most significant rank 1 matrix in the SVD of A^{-1}. Thus if the condition number of a matrix is large ($\sigma_n \ll \sigma_1$), the inverse is dominated by an insignificant part of the original matrix. This suggests that small changes in A (such as round-off errors or other noise) can seriously affect A^{-1} and such matrices are then called ill-conditioned. Rather than allow this noise to dominate the inverse, it seems more appropriate to ignore it and replace the corresponding diagonal terms of S^{-1} by 0. When this is done, the matrix A^{-1} is essentially replaced by an effective pseudoinverse. The decision of which values σ_i^{-1} to replace by 0 depends not only on the ratio σ_1/σ_i but also on the order of computer machine precision and the application involved.

For our 3×3 matrix A, the surfaces of FIGURE 4 show how A^{-1} is dominated by the smallest term of the singular value decomposition, and suggest that while A_2^+ might be a good replacement for A^{-1} in certain applications, this decision should not be taken lightly. The shape of the surface is being emphasized in FIGURE 4 with the scales chosen for viewing convenience. The maximum surface height for A_3^+ is actually about 200 times that of A_2^+.

When a matrix A is either square singular or nonsquare, then A^{-1} fails to exist and a pseudoinverse of A is given by $A^+ = QS^+P^T$ where S^+ is diagonal with $d_i = \sigma_i^{-1}$ if $\sigma_i \neq 0$ and $d_i = 0$ if $\sigma_i = 0$. Thus if A has rank r and singular value decomposition $A = \sum_{i=1}^n \sigma_i R_i$ then

$$A^+ = \sigma_1^{-1} R_1^T + \sigma_2^{-1} R_2^T + \cdots + \sigma_{r-1}^{-1} R_{r-1}^T + \sigma_r^{-1} R_r^T.$$

We show in FIGURE 5 the pseudoinverse A^+ of ABE, and A_1^+ through A_5^+, the pseudoinverses

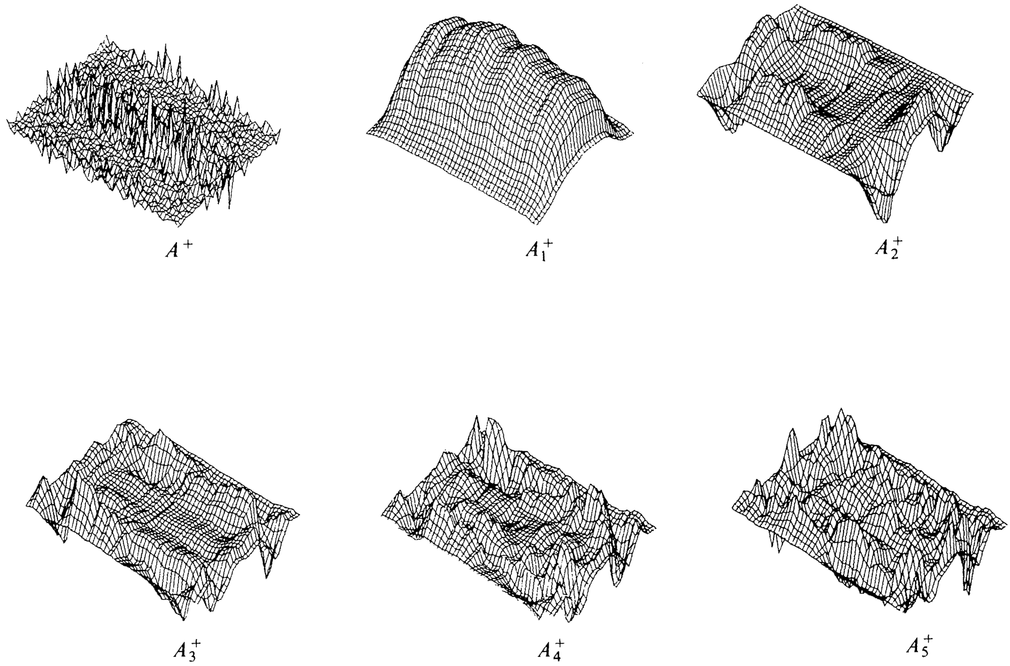

FIGURE 5. **Pseudoinverses of "ABE" and A_r.**

of our rank 1 through rank 5 approximations of ABE. The heights are again adjusted for viewing convenience (in fact A^+ has terms of much larger magnitude than the others). It is the "shape" of the matrices which is emphasized. It should be noted that while A_5 is a good approximation to A, A_5^+ is not a good approximation to A^+. Listing the singular values for a given problem frequently aids the user in deciding on an effective rank e for a matrix, and then A_e^+ is used in place of A^+. These substitutions provide the reliability for the SVD method in solving least square problems, since small changes in the original matrices are not allowed to dominate the pseudoinverse.

References

[1] H. C. Andrews and C. L. Patterson, Outer product expansions and their uses in digital image processing, Amer. Math. Monthly, 82 (1975) 1–12.

[2] P. J. Davis, Circulant Matrices, John Wiley and Sons, New York, 1979.

[3] C. Eckart and G. Young, The approximation of one matrix by another of lower rank, Psychometrika, 1 (1936) 211–218.

[4] G. E. Forsythe, M. A. Malcolm, and C. B. Moler, Computer Methods for Mathematical Computations, Prentice-Hall, Englewood Cliffs, NJ, 1977.

[5] G. H. Golub and W. Kahan, Calculating the singular values and pseudoinverse of a matrix, SIAM J. Numer. Anal., 2 (1965) 205–224.

[6] G. H. Golub and C. Reinsch, Singular value decomposition and least squares solutions, Numer. Math., 14 (1970) 403–420.

[7] P. E. Green, Mathematical Tools for Applied Multivariate Analysis, Academic Press, New York, 1976.

[8] V. C. Kema and A. J. Laub, The singular value decomposition: its computation and some applications, IEEE Transactions on Automatic Control, 25 (1980) 164–176.

[9] C. L. Lawson and R. J. Hanson, Solving Least Squares Problems, Prentice-Hall, Englewood Cliffs, NJ, 1974.

[10] C. C. MacDuffee, The Theory of Matrices, Springer, Berlin, 1933.

[11] G. Strang, Linear Algebra and Its Applications, 2nd ed., Academic Press, New York, 1980.