Using the Singular Value Decomposition *

Emmett J. Ientilucci
Chester F. Carlson Center for Imaging Science
Rochester Institute of Technology
emmett@cis.rit.edu

May 29, 2003

Abstract

This report introduces the concept of factorizing a matrix based on the singular value decomposition. It discusses methods that operate on square-symmetric matrices such as spectral decomposition. The singular value decomposition technique is explained and related to solving linear systems of equations. Examples are presented based on over and under determined systems.

1 Introduction

The Singular Value Decomposition (SVD) is a widely used technique to decompose a matrix into several component matrices, exposing many of the useful and interesting properties of the original matrix. The decomposition of a matrix is often called a *factorization*. Ideally, the matrix is decomposed into a set of factors (often orthogonal or independent) that are optimal based on some criterion. For example, a criterion might be the reconstruction of the decomposed matrix. The decomposition of a matrix is also useful when the matrix is not of *full rank*. That is, the rows or columns of the matrix are linearly *dependent*. Theoretically, one can use Gaussian elimination to reduce the matrix to row echelon form and then count the number of nonzero rows to determine the rank. However, this approach is not practical when working in finite precision arithmetic. A similar case presents itself when using LU decomposition where L is in lower triangular form with 1's on the diagonal and U is in upper triangular form. Ideally, a *rank-deficient* matrix

*This file has version number v1.0, last revised on 5/29/03 may be decomposed into a smaller number of factors than the original matrix and still preserve all of the information in the matrix. The SVD, in general, represents an expansion of the original data in a coordinate system where the covariance matrix is diagonal.

Using the SVD, one can determine the dimension of the matrix range or more-often called the *rank*. The rank of a matrix is equal to the number of linear *independent* rows or columns. This is often referred to as a *minimum spanning set* or simply a *basis*. The SVD can also quantify the sensitivity of a linear system to numerical error or obtain a matrix inverse. Additionally, it provides solutions to least-squares problems and handles situations when matrices are either singular or numerically very close to singular.

2 Spectral Decomposition and Square-Symmetric Matrices

We now turn to the simple case of factoring matrices that are both *square* and *symmetric*. An example of a square-symmetric matrix would be the $k \times k$ co-variance matrix, $\mathbf{\Sigma}$. If matrix $\mathbf{A}$ has the property $\mathbf{A} = \mathbf{A}^T$, then it is said to be symmetric. If $\mathbf{A}$ is a square-symmetric matrix, then a useful decomposition is based on its *eigenvalues* and *eigenvectors*. That is,

$$\mathbf{AX} = \mathbf{X\Lambda} \tag{1}$$

where $\mathbf{X}$ is a matrix of eigenvectors and $\mathbf{\Lambda}$ is the *diagonal* matrix of eigenvalues. Diagonal in the sense that all the entries off the main diagonal are zero. The eigenvectors have the convenient mathematical property of orthogonality (*i.e.*, $\mathbf{e}^T\mathbf{e} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix) and span the entire space of $\mathbf{A}$. That is, form a basis or minimum spanning set.

The set of eigenvalues is called the *spectrum* of $\mathbf{A}$. If two or more eigenvalues of $\mathbf{A}$ are identical, the spectrum of the matrix is called *degenerate*. The "spectrum" nomenclature is an exact analogy with the idea of the spectrum of light as depicted in a rainbow. The brightness of each color of the spectrum tell us "how much" light of that wavelenght existed in the undispersed white light. For this reason, the procedure is often referred to as a *spectral decomposition*.

Additionally, the spectral decomposition can be re-formulated in terms of eigenvalue-eigenvector pairs. That is, first let $\mathbf{A}$ be a $k \times k$ symmetric matrix. Then $\mathbf{A}$ can be expressed in terms of its k eigenvalue-eigenvector pairs $(\lambda_i, \mathbf{e}_i)$ as the expansion

$$\mathbf{A} = \sum_{i=1}^{k} \lambda_i \mathbf{e}_i \mathbf{e}_i^T \tag{2}$$

This is a useful result in that it helps facilitate the computation of $\mathbf{A}^{-1}$, $\mathbf{A}^{1/2}$, and $\mathbf{A}^{-1/2}$. First, let the normalized eigenvectors be the columns of another matrix $\mathbf{P} = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k]$. Then

$$\mathbf{A} = \mathbf{P\Lambda P}^T = \sum_{i=1}^{k} \lambda_i \mathbf{e}_i \mathbf{e}_i^T \tag{3}$$

where $\mathbf{PP}^T = \mathbf{P}^T\mathbf{P} = \mathbf{I}$ and $\mathbf{\Lambda}$ is the diagonal matrix with eigenvalues λ_i on the diagonal (we will soon see that that this is the form of the SVD). Thus to compute the inverses, we simply take the reciprocal of the eigenvalues in the diagonal matrix $\mathbf{\Lambda}$. That is

$$\mathbf{A}^{-1} = \mathbf{P\Lambda}^{-1}\mathbf{P}^T = \sum_{i=1}^{k} \frac{1}{\lambda_i} \mathbf{e}_i \mathbf{e}_i^T \tag{4}$$

Similarly, the computation of the square root and inverse square root matrices are performed as follows:

$$\mathbf{A}^{1/2} = \mathbf{P\Lambda}^{1/2}\mathbf{P}^T = \sum_{i=1}^{k} \sqrt{\lambda_i} \mathbf{e}_i \mathbf{e}_i^T \tag{5}$$

$$\mathbf{A}^{-1/2} = \mathbf{P\Lambda}^{-1/2}\mathbf{P}^T = \sum_{i=1}^{k} \frac{1}{\sqrt{\lambda_i}} \mathbf{e}_i \mathbf{e}_i^T \tag{6}$$

3 The Singular Value Decomposition

The ideas that lead to the spectral decomposition can be extended to provide a decomposition for a rectangular, rather than a square, matrix. We can decompose a matrix that is not square nor symmetric by first considering a matrix $\mathbf{A}$ that is of dimension $m \times n$ where $m \geq n$. This assumption is made for convenience only; all the results will also hold if $m < n$. As it turns out, the vectors in the the expansion of $\mathbf{A}$ are the eigenvectors of the square matrices $\mathbf{AA}^T$ and $\mathbf{A}^T\mathbf{A}$. The former is a outer product and results in a matrix that is spanned by the row space of $\mathbf{A}$. The latter is a inner product and results in a matrix that is spanned by the column space (*i.e.*, the range) of $\mathbf{A}$.

The *singular values* are the nonzero square roots of the eigenvalues from $\mathbf{AA}^T$ and $\mathbf{A}^T\mathbf{A}$. The eigenvectors of $\mathbf{AA}^T$ are called the "left" singular vectors ($\mathbf{U}$) while the eigenvectors of $\mathbf{A}^T\mathbf{A}$ are the "right" singular vectors ($\mathbf{V}$). By retaining the nonzero eigenvalues $k = \min(m, n)$, a *singular value decomposition* (SVD) can be constructed. That is

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T \tag{7}$$

where $\mathbf{U}$ is an $m \times m$ orthogonal matrix ($\mathbf{U}^T\mathbf{U} = \mathbf{I}$), $\mathbf{V}$ is an $n \times n$ orthogonal matrix ($\mathbf{V}^T\mathbf{V} = \mathbf{I}$), and $\mathbf{\Lambda}$ is an $m \times n$ matrix whose off-diagonal entries are all 0's and whose diagonal elements satisfy

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0 \tag{8}$$

It can be shown that the rank of $\mathbf{A}$ equals the number of nonzero singular values and that the magnitudes of the nonzero singular values provide a measure of how close $\mathbf{A}$ is to a matrix of lower rank. That is, if $\mathbf{A}$ is nearly rank deficient (singular), then the singular values will be small. In general, the SVD represents an expansion of the original data $\mathbf{A}$ in a coordinate system where the covariance matrix $\mathbf{\Sigma}_A$ is diagonal.

Remember, this is called the singular value decomposition because the factorization finds values or eigenvalues or *characteristic roots* (all the same) that make the the following *characteristic equation* true or singular. That is

$$|\mathbf{A} - \lambda\mathbf{I}| = 0 \tag{9}$$

Using the determinant this way helps solve the linear system of equations thus generating an nth degree polynomial in the variable λ. This polynomial, that yields n-roots, is called the *characteristic polynomial*.

Equation (9) actually comes from the more *generalized eigenvalue equation* which has the form

$$\mathbf{Ax} = \lambda\mathbf{x} \tag{10}$$

which, when written in matrix form, is expressed as Eqn. (1) introduced earlier. This implies

$$\mathbf{Ax} - \lambda\mathbf{x} = 0 \tag{11}$$

or

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = 0 \tag{12}$$

The theory of simultaneous equations tells us that for this equation to be true it is necessary to have either $\mathbf{x} = 0$ or $|\mathbf{A} - \lambda\mathbf{I}| = 0$. Thus the motivation to solve Eqn. (9).

4 Examples

4.1 Inverses of Square-Symmetric Matrices

The covariance matrix $\boldsymbol{\Sigma}$ is an example of a square-symmetric matrix. Consider the following

$$\boldsymbol{\Sigma} = \begin{bmatrix} 2.2 & 0.4 \\ 0.4 & 2.8 \end{bmatrix}$$

The matrix is not singular since the determinant $|\boldsymbol{\Sigma}| = 6$ therefore $\boldsymbol{\Sigma}^{-1}$ exists. The eigenvalues and eigenvectors are obtained directly from $\boldsymbol{\Sigma}$ since it is already square. Furthermore, the left and right singular vectors $(\mathbf{U}, \mathbf{V})$ will be the same due to symmetry. We solve for the eigenvalues via Eqn. (9) to obtain $\lambda_1 = 3$ and $\lambda_2 = 2$ which are also the singular values in this case. We then compute the corresponding eigenvectors via Eqn. (10) to obtain $\mathbf{e}_1^T = [1/\sqrt{5}, 2/\sqrt{5}]$ and $\mathbf{e}_2^T = [2/\sqrt{5}, -1/\sqrt{5}]$. Finally we factor $\boldsymbol{\Sigma}$ into a singular value decomposition.

$$\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \end{bmatrix}^T = \begin{bmatrix} 2.2 & 0.4 \\ 0.4 & 2.8 \end{bmatrix}$$

It is now trivial to compute $\boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{\Sigma}^{-1/2}$.

$$\boldsymbol{\Sigma}^{-1} = \mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{V}^T = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \end{bmatrix}^T = \begin{bmatrix} 0.47 & -0.07 \\ -0.07 & 0.37 \end{bmatrix}$$

$$\boldsymbol{\Sigma}^{-1/2} = \mathbf{U}\boldsymbol{\Lambda}^{-1/2}\mathbf{V}^T = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} & 0 \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \end{bmatrix}^T = \begin{bmatrix} 0.68 & -0.05 \\ -0.05 & 0.60 \end{bmatrix}$$

4.2 Solving A System of Linear Equations

A set of linear algebraic equations can be written as

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{13}$$

where $\mathbf{A}$ is a matrix of coefficients $(m \times n)$, and $\mathbf{b}$ $(m \times 1)$ is some form of a system output vector. The vector $\mathbf{x}$ is what we usually solve for. If $m = n$ then there are as many equations as unknowns, and there is a good chance of solving for $\mathbf{x}$. That is

$$\mathbf{A}^{-1}\mathbf{A}\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} \tag{14}$$

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} \tag{15}$$

Here, we simply compute the inverse of **A**. This can prove to be a challenging task, however, for there are many situations where the inverse of **A** does not exist. In these cases we will approximate the inverse via the SVD which can turn a singular problem into a non-singular one.

Vector **x** in Eqn. 13 can also be solved for by using the transpose of **A**. That is

$$\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b} \tag{16}$$

$$\mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} \tag{17}$$

This is the form of the solution in a least-squares sense from standard multivariate regression theory where the inverse of **A** is express as

$$\mathbf{A}^{\dagger} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \tag{18}$$

where $\mathbf{A}^{\dagger}$ is called the More-Penrose pseudoinverse. We will see that the use of the SVD can aid in the computation of the generalized pseudoinverse.

4.2.1 Equal Number of Equations and Unknowns

This is the case when matrix **A** is square. We have already presented the case when **A** is both square and symmetric. But what if it is only square, or more importantly, square and singular or degenerate (*i.e.*, one of the rows or columns of the original matrix is a linear combination of another one) Here again we use SVD. Take for example the following matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$$

This matrix is square but not symmetric. Furthermore it is singular since the determinant $|\mathbf{A}| = 0$. This would imply $\mathbf{A}^{-1}$ does not exist. Using the SVD, however, we can approximate an inverse. The SVD approach tells us to compute eigenvalues and eigenvectors from the inner and outer product matrices:

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 5 & 5 \\ 5 & 5 \end{bmatrix} \quad \text{and} \quad \mathbf{A}\mathbf{A}^T = \begin{bmatrix} 2 & 4 \\ 4 & 8 \end{bmatrix}$$

The inner and outer product matrices are both symmetric. The eigenvalues from these matrices are $\lambda_1 = 0$ and $\lambda_2 = 10$. Consequently, the singular values of $\mathbf{A}$ are $\sigma_1 = 0$ and $\sigma_2 = \sqrt{10}$. Therefore the rank of $\mathbf{A}$ is 1. The decomposition is then expressed as

$$\mathbf{A} = \mathbf{U\Lambda V}^T = \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{-1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & \sqrt{10} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}^T = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$$

4.2.2 Underdetermined - Fewer Equations than Unknowns

$$\mathbf{B} = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}$$

$$\mathbf{B}^T\mathbf{B} = \begin{bmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{BB}^T = \begin{bmatrix} 11 & 1 \\ 1 & 11 \end{bmatrix}$$

The eigenvalues from $\mathbf{B}^T\mathbf{B}$ are $\lambda_1 = 12$, $\lambda_2 = 10$ and $\lambda_3 = 0$. The eigenvalues from $\mathbf{BB}^T$ are $\lambda_1 = 12$ and $\lambda_2 = 10$. Consequently, the non-zero singular values of $\mathbf{B}$ are $\sigma_1 = \sqrt{12}$ and $\sigma_2 = \sqrt{10}$. Therefore the rank of $\mathbf{B}$ is 2. The decomposition is then expressed as

$$\mathbf{B} = \mathbf{U\Lambda V}^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{30}} \\ \frac{2}{\sqrt{6}} & \frac{-1}{\sqrt{5}} & \frac{2}{\sqrt{30}} \\ \frac{1}{\sqrt{6}} & 0 & \frac{-5}{\sqrt{30}} \end{bmatrix}^T = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}$$

4.2.3 Overdetermined - More Equations than Unknowns

$$\mathbf{C} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix}$$

$$\mathbf{C}^T\mathbf{C} = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{CC}^T = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The eigenvalues from $\mathbf{C}^T\mathbf{C}$ are $\lambda_1 = 4$ and $\lambda_2 = 0$. The eigenvalues from $\mathbf{CC}^T$ are $\lambda_1 = 4$, $\lambda_2 = 0$ and $\lambda_3 = 0$. Consequently, the non-zero singular values of $\mathbf{C}$ are $\sigma_1 = \sqrt{4}$ and $\sigma_2 = 0$. Therefore the rank of $\mathbf{C}$ is 1. The decomposition is then expressed as

$$\mathbf{C} = \mathbf{U\Lambda V}^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{4} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}^T = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix}$$

4.2.4 Overdetermined: Least-Squares Solution

When we have a set of linear equations with more equations than unknowns, and we wish to solve for the vector $\mathbf{x}$, as in Eqn. 17, we usually do so in a least-squares sense. However, use of the SVD provides a numerically robust solution to the least-squares problem presented in Eqn. 17 which now becomes

$$\mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} = \mathbf{U\Lambda}^{-1}\mathbf{V}^T\mathbf{b} \tag{19}$$