The Singular Value Decomposition in Symmetric (Löwdin) Orthogonalization and Data Compression

The SVD is the most generally applicable of the orthogonal-diagonal-orthogonal type matrix decompositions

Every matrix, even nonsquare, has an SVD

The SVD contains a great deal of information and is very useful as a theoretical and practical tool

****

1 Preliminaries

Unless otherwise indicated, all vectors are column vectors

$$u \in \mathbb{R}^n \qquad \Longrightarrow \qquad u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \in \mathbb{R}^{n \times 1}$$

Definition 1.1 Let $u \in \mathbb{R}^n$, so that $u = (u_1, u_2, \ldots u_n)^T$. The (Euclidean) *norm* of u is defined as

$$\|u\|_2 = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2} = \left(\sum_{j=1}^{n} u_j^2 \right)^{1/2}$$

Definition 1.2 A vector $u \in \mathbb{R}^n$ is a *unit vector* or *normalized* if

$$\|u\|_2 \;=\; 1$$

==

Definition 1.3 Let $A = (a_{ij}) \in \mathbb{R}^{m \times n}$. The *transpose* A^T of A is the matrix $(a_{ji}) \in \mathbb{R}^{n \times m}$.

==

Example 1.4

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

Definition 1.5 (Matrix Multiplication) Let $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$. Then the product AB is defined element-wise as

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

and the matrix $$AB \quad \in \quad \mathbb{R}^{m \times p}$$

==

Definition 1.6 Let $u, v \in \mathbb{R}^n$. Then the *inner product* of u and v, written $\langle u, v \rangle$ is defined as

$$\langle u, v \rangle \quad = \quad \sum_{j=1}^{n} u_j v_j = u^T v$$

==

Note that this notation permits us to write matrix multiplication as entry-wise inner products of the rows and columns of the matrices

==

If we denote the i^{th} row of A by ${}_iA$ and the j^{th} column of B by B_j we have

$$(AB)_{ij} \quad = \quad \langle \, ({}_iA)^T \, , \, B_j \, \rangle \quad = \quad {}_iAB_j$$

Example 1.7

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

$$= \begin{pmatrix} -1\cdot(2)+1\cdot(0)+0\cdot(6) & -1\cdot(3)+1\cdot(-2)+0\cdot(-3) \\ \\ 3\cdot(2)+-2\cdot(0)+1\cdot(6) & 3\cdot(3)+-2\cdot(-2)+1\cdot(-3) \end{pmatrix}$$

$$= \begin{pmatrix} -2 & -5 \\ 12 & 10 \end{pmatrix}$$

==

Definition 1.8 Two vectors $u, v \in \mathbb{R}^n$ are *orthogonal* if

$$\langle u\,,\,v\rangle \quad = \quad u^T v \quad = \quad \begin{pmatrix} u_1 & u_2 & \cdots & u_n \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

$$= \quad u_1 v_1 + u_2 v_2 + \cdots + u_n v_n \quad = \quad 0$$

==

If u, v are orthogonal and both $\|u\|_2 = 1$ and $\|v\|_2 = 1$, then we say u and v are *orthonormal*

Recall that the n-dimensional *identity matrix* is

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

We'll write I for the identity matrix when the size is clear from the context.

==

Definition 1.9 A square matrix $Q \in \mathbb{R}^{n \times n}$ is *orthogonal* if $Q^T Q = I$.

==

This definition means that the columns of an orthogonal matrix A are mutually orthogonal unit vectors in $\mathbb{R}^n$

==

Alternatively, the columns of A are an orthonormal basis for $\mathbb{R}^n$

==

Now Definition 1.9 shows that Q^T is the left-inverse of Q

But since matrix multiplication is associative, Q^T is the right-inverse (and hence the inverse) of Q - indeed, let P be a right-inverse of Q (so that $QP = I$); then

$$(Q^T Q)P = Q^T(QP) \qquad \Longleftrightarrow \qquad IP = Q^T I \qquad \Longleftrightarrow \qquad P = Q^T$$

==

The SVD is applicable to even nonsquare matrices with complex entries, but for clarity we will restrict our initial treatment to real square matrices

****

2 Structure of the SVD

Definition 2.1 Let $A \in \mathbb{R}^{n\times n}$. Then the *(full) singular value decomposition* of A is

$$A = U\Sigma V^T = \left(\begin{array}{c|c|c|c} & & & \\ U_1 & U_2 & \cdots & U_m \\ & & & \end{array}\right) \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & & \ddots & 0 \\ 0 & \cdots & & \sigma_n \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{pmatrix} \left(\begin{array}{c} (V_1)^T \\ \hline (V_2)^T \\ \hline \vdots \\ \hline (V_n)^T \end{array}\right).$$

where U, V are orthogonal matrices and Σ is diagonal

The σ_i's are the *singular values* of A, by convention arranged in nonincreasing order

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0;$$

the columns of U are termed *left singular vectors* of A; the columns of V are called *right singular vectors* of A

Since U and V are orthogonal matrices, the columns of each form orthonormal (mutually orthogonal, all of length 1) bases for $\mathbb{R}^n$

==

We can use these bases to illuminate the fundamental property of the SVD:

==

For the equation $Ax = b$, the SVD makes every matrix diagonal by selecting the right bases for the range and domain

==

Let $b, x \in \mathbb{R}^n$ such that $Ax = b$, and expand b in the columns of U and x in the columns of V to get

$$b' = U^T b, \quad x' = V^T x.$$

Then we have

$$b = Ax \quad \Longleftrightarrow \quad \begin{aligned} U^T b &= U^T A x \\ &= U^T (U \Sigma V^T) x \\ &= (U^T U) \Sigma (V^T x) \\ &= I \Sigma x' \\ &= \Sigma x' \end{aligned}$$

or

$$\boldsymbol{b = Ax} \quad \Longleftrightarrow \quad \boldsymbol{b' = \Sigma x'}$$

==

Let $y \in \mathbb{R}^n$, then the action of left multiplication of y by A (computing $z = Ay$) is decomposed by the SVD into three steps

$$\begin{aligned} z &= Ay \\ &= (U \Sigma V^T)\, y = U \Sigma (V^T y) \\ &= U \Sigma\, c \qquad (c := V^T y) \\ &= U w \qquad (w := \Sigma\, c) \end{aligned}$$

$c = V^T y$ is the *analysis* step, in which the components of y, *in the basis of* $\mathbb{R}^n$ *given by the columns of* V, are computed

==

$w = \Sigma c$ is the *scaling* step in which the components $c_i,\ i \in \{1, 2, \ldots, n\}$ are dilated

==

$z = Uw$ is the *synthesis* step, in which z is assembled by scaling each of the $\mathbb{R}^n$-basis vectors u_i by w_i and summing

So how do we find the matrices U, Σ, and V in the SVD of some $A \in \mathbb{R}^{n \times n}$?

==

Since $\quad V^T V = I = U^T U, \quad A = U \Sigma V^T \quad$ yields

$$AV = U\Sigma \qquad \text{and} \tag{1}$$

$$U^T A = \Sigma V^T \qquad \text{or, taking transposes}$$

$$A^T U = V\Sigma \tag{2}$$

==

Or, for each $j \in \{1, 2, \ldots, n\}$,

$$A v_j = \sigma_j u_j \qquad \text{from Equation 1} \tag{3}$$

$$A^T u_j = \sigma_j v_j \qquad \text{from Equation 2} \tag{4}$$

==

Now we multiply Equation 3 by A^T to get

$$\begin{aligned} A^T A v_j &= A^T \sigma_j u_j \\ &= \sigma_j A^T u_j \\ &= \sigma_j^2 v_j \end{aligned}$$

==

So the v_j's are the eigenvectors of $A^T A$ with corresponding eigenvalues σ_j^2

==

Note that $\quad (A^T A)_{ij} \quad = \quad {}_iAA_j \quad$ or

$$A^T A = \begin{pmatrix} {}_1AA_1 & {}_1AA_2 & \cdots & {}_1AA_n \\ {}_2AA_1 & {}_2AA_2 & & \vdots \\ \vdots & & \ddots & \\ {}_nAA_1 & \cdots & & {}_nAA_n \end{pmatrix} \tag{5}$$

$A^T A$ is a matrix of inner products of columns of A - often called the *Gram matrix of* A

We'll see the Gram matrix again when considering applications

Let's do an example:

$$A = \begin{pmatrix} 1 & 0 & -1 \\ 1 & 1 & 0 \\ -1 & 0 & -1 \end{pmatrix} \qquad \Longrightarrow \qquad A^T = \begin{pmatrix} 1 & 1 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & -1 \end{pmatrix}$$

$$\Longrightarrow \qquad A^T A = \begin{pmatrix} 3 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

===================================

To find the eigenvectors v and the corresponding eigenvalues λ for $B := A^T A$, we solve

$$Bx = \lambda x \qquad \Longleftrightarrow \qquad (B - \lambda I)x = 0$$

for λ and x

The standard technique for finding such λ and v is to first note that we are looking for the λ that make the matrix

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

singular

==

This is most easily done by solving det$(B - \lambda I) = 0$:

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

$$= -\lambda^3 + 6\lambda^2 - 10\lambda + 4 = 0$$

$$\Longleftrightarrow$$

$$\sigma_1^2 = \lambda_1 = 2 + \sqrt{2}$$

$$\sigma_2^2 = \lambda_2 = 2$$

$$\sigma_3^2 = \lambda_3 = 2 - \sqrt{2}$$

Now (for a gentle first step) we'll find a vector v_2 so that $A^T A v_2 = 2v_2$

==

We do this by finding a basis for the nullspace of

$$A^T A - 2I = \begin{pmatrix} 3-2 & 1 & 0 \\ 1 & 1-2 & 0 \\ 0 & 0 & 2-2 \end{pmatrix} \quad = \quad \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

==

Certainly any vector of the form $\begin{pmatrix} 0 \\ 0 \\ t \end{pmatrix}, \quad t \in \mathbb{R}$, is mapped to zero by

$$A^T A - 2I$$

==

So we can set $v_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

To find v_1 we find a basis for the nullspace of

$$A^TA - (2+\sqrt{2})I = \begin{pmatrix} 1-\sqrt{2} & 1 & 0 \\ 1 & -1-\sqrt{2} & 0 \\ 0 & 0 & -\sqrt{2} \end{pmatrix}$$

which row-reduces ($R2 \quad \longleftarrow \quad (1+\sqrt{2})R1 + R2,$ then $R3 \quad \longleftrightarrow \quad R2$) to

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

==

So any vector of the form $\begin{pmatrix} s \\ \\ (-1+\sqrt{2})s \\ \\ 0 \end{pmatrix}$ is mapped to zero by

$$A^TA - (2+\sqrt{2})I$$

==

so $v_1' = \begin{pmatrix} 1 \\ \\ -1+\sqrt{2} \\ \\ 0 \end{pmatrix}$ spans the nullspace of $A^TA - \lambda_1 I,$ but $\|v_1'\| \neq 1$

So we set $v_1 = \frac{v_1'}{\|v_1'\|} = \frac{1}{\sqrt{4-2\sqrt{2}}} \begin{pmatrix} 1 \\ -1+\sqrt{2} \\ 0 \end{pmatrix}$

==

We *could* find v_3 in a similar manner, but in this particular case there's a quicker way...

==

$$v_3 = \begin{pmatrix} -(v_1)_2 \\ (v_1)_1 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{4-2\sqrt{2}}} \begin{pmatrix} 1-\sqrt{2} \\ 1 \\ 0 \end{pmatrix}$$

Certainly $v_3 \perp v_2$ and by construction $v_3 \perp v_1$ - recall the theorem from linear algebra symmetric matrices must have orthogonal eigenvectors

We've found $V = \left(\begin{array}{c|c|c} & & \\ v_1 & v_2 & v_3 \\ & & \end{array} \right) = \begin{pmatrix} \frac{1}{\sqrt{4-2\sqrt{2}}} & 0 & \frac{1-\sqrt{2}}{\sqrt{4-2\sqrt{2}}} \\ \frac{-1+\sqrt{2}}{\sqrt{4-2\sqrt{2}}} & 0 & \frac{1}{\sqrt{4-2\sqrt{2}}} \\ 0 & 1 & 0 \end{pmatrix}$

==

And of course

$$\Sigma = \begin{pmatrix} \sqrt{2+\sqrt{2}} & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & \sqrt{2-\sqrt{2}} \end{pmatrix}$$

==

Now, how do we find U?

If $\sigma_n > 0,\ \Sigma$ is invertible and

$$U = AV\Sigma^{-1}$$

==

So we have

$$U = \begin{pmatrix} 1 & 0 & -1 \\ 1 & 1 & 0 \\ -1 & 0 & -1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{4-2\sqrt{2}}} & 0 & \frac{1-\sqrt{2}}{\sqrt{4-2\sqrt{2}}} \\ \frac{-1+\sqrt{2}}{\sqrt{4-2\sqrt{2}}} & 0 & \frac{1}{\sqrt{4-2\sqrt{2}}} \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2+\sqrt{2}}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2-\sqrt{2}}} \end{pmatrix}$$

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

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

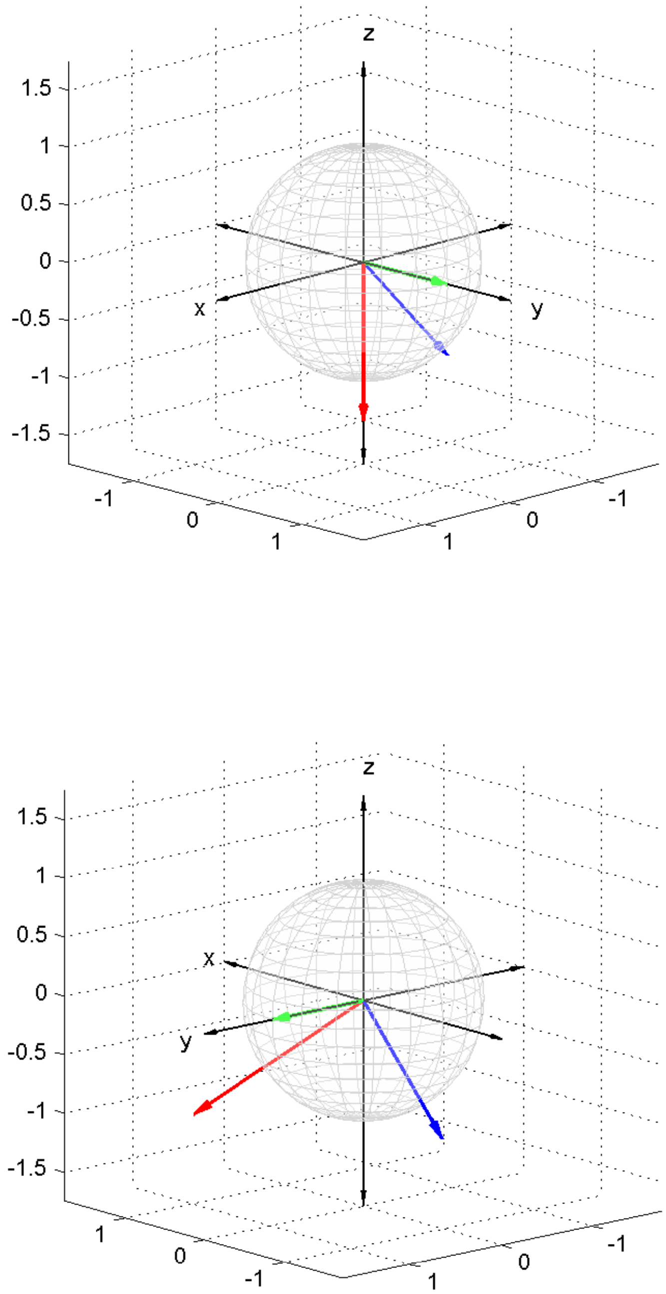
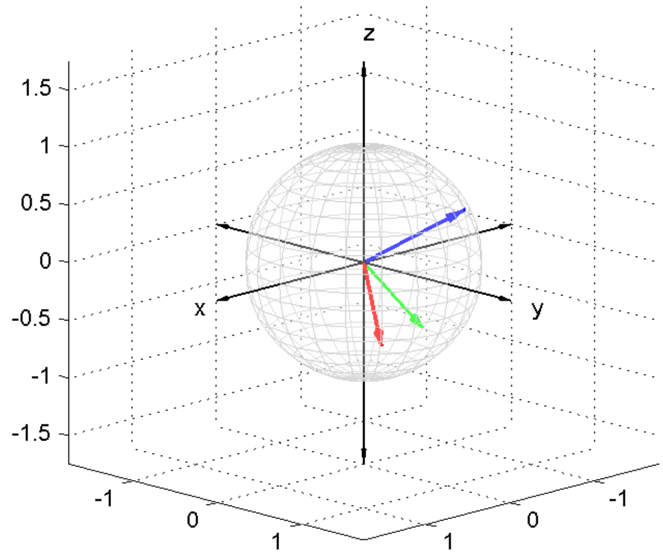


Figure 1: The columns of A in the unit sphere



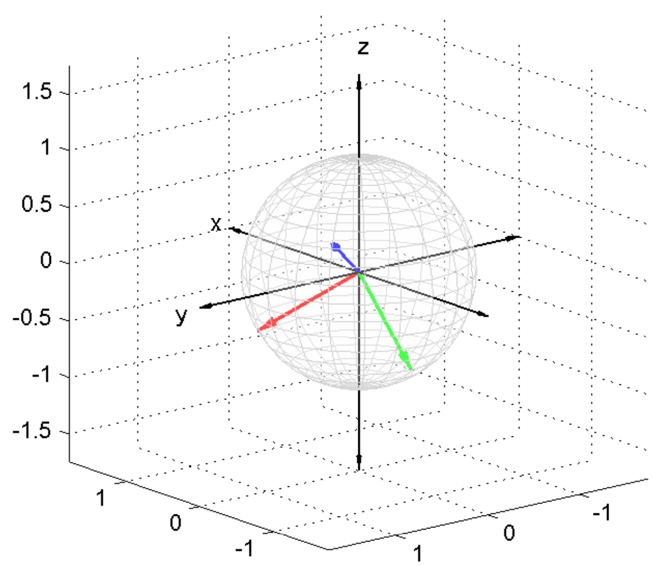


Figure 2: The columns of U in the unit sphere

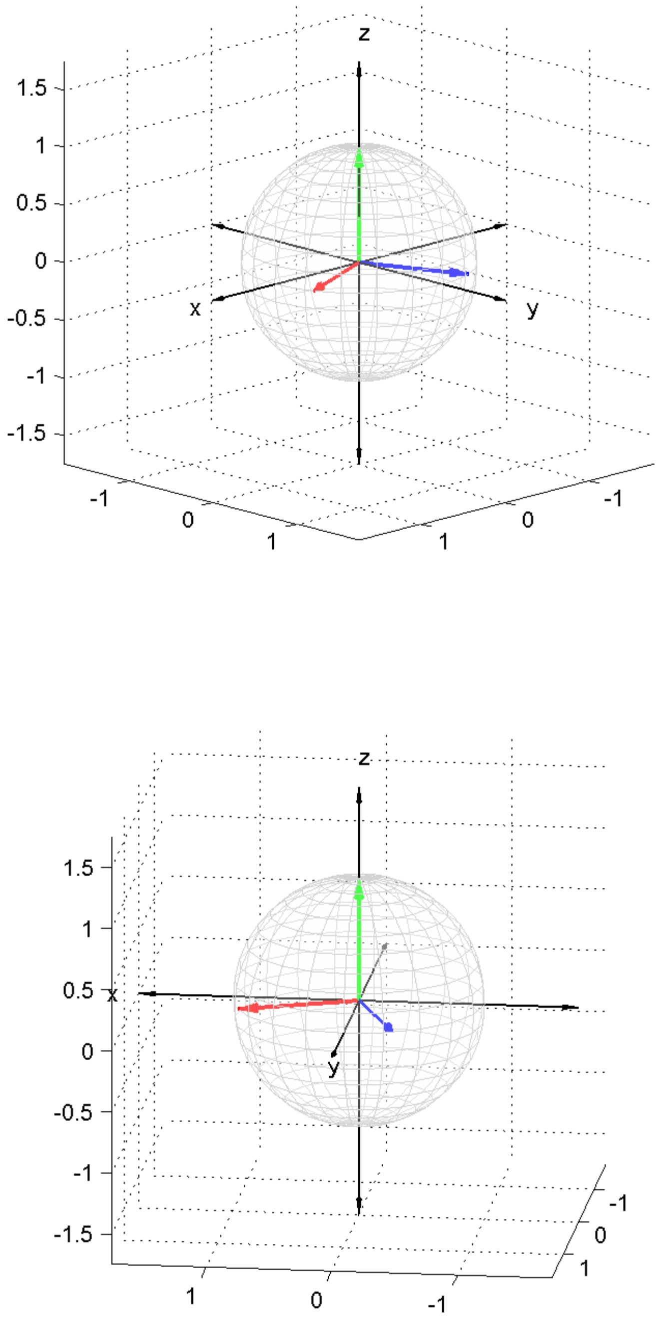


Figure 3: The columns of V in the unit sphere

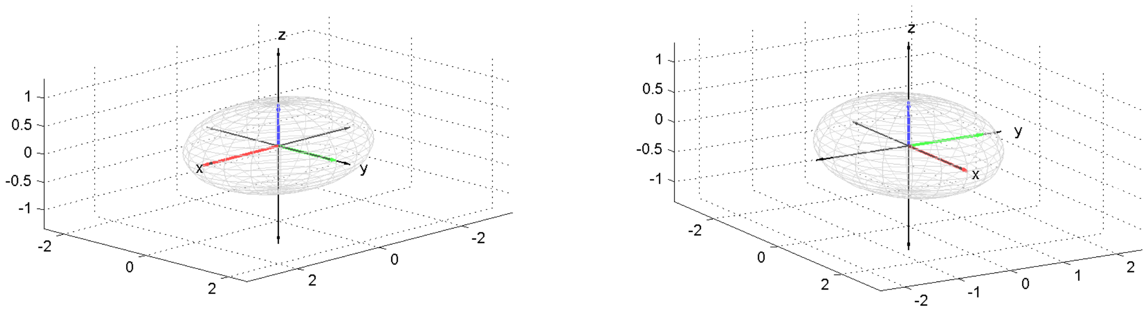


Figure 4: The columns of Σ in the ellipse formed by Σ acting on the unit sphere by left-multiplication

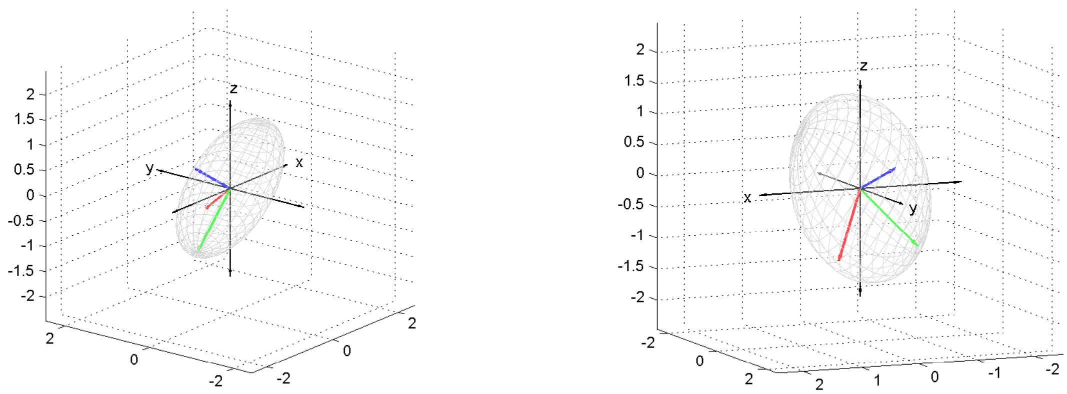


Figure 5: The columns of $AV = U\Sigma$ in the ellipse formed by A acting on the unit sphere by left-multiplication

==

Note that the columns of U and V are orthogonal (as are, of course, the columns of Σ)

Note that in practice, the SVD is computed more efficiently than by the direct method we used here; usually by (OK, get ready for the gratuitous mathspeak)

reducing A to bidiagonal form $U_1 B V_1^T$ by elementary reflectors or Givens rotations and

directly computing the SVD of B $(= U_2 \Sigma V_2^T)$

then the SVD of A is $\quad (U_1 U_2)\, \Sigma\, (V_2^T V_1^T)$

==

If $\sigma_n = 0,$ then A is singular and the entire process above must be modified slightly but carefully.

==

If r is the rank of A (the number of nonzero rows of the row-echelon form of A) then

$n - r$ singular values of A are zero (equivalently if there are $n - r$ zero rows in the row-echelon form of A), so

Σ^{-1} is not defined, and we define the *pseudo-inverse* Σ^{+} of Σ as

$$\Sigma^{+} = diag(\sigma_1^{-1}, \sigma_2^{-1}, \ldots, \sigma_r^{-1}, 0, \ldots, 0)$$

Thus we can define the first r columns of U via $AV\Sigma^+$ and to complete U we choose any $n - r$ orthonormal vectors which are also orthogonal to span$\{u_1, u_2, \ldots, u_r\}$, via, for example, Gram-Schmidt

==

Recall that the SVD is defined for even nonsquare matrices

==

In this case, the above process is modified to permit U and V to have different sizes

==

If $A \in \mathbb{R}^{m \times n}$, then

$$U \in \mathbb{R}^{m \times m}$$

$$\Sigma \in \mathbb{R}^{m \times n}$$

$$V \in \mathbb{R}^{n \times n}$$

In the case $m > n$:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{m1} & \cdots & & a_{mn} \end{pmatrix}$$

$$= \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} & \cdots & u_{1m} \\ u_{21} & u_{22} & \cdots & u_{2n} & \cdots & u_{2m} \\ \vdots & & & \ddots & & \vdots \\ u_{m1} & \cdots & & u_{mn} & \cdots & u_{mm} \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & & \ddots & 0 \\ 0 & \cdots & & \sigma_n \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{pmatrix} \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & & \vdots \\ \vdots & & \ddots & \\ v_{n1} & \cdots & & v_{nn} \end{pmatrix}$$

or, in another incarnation of the SVD (the *reduced* SVD)

$$A = \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & & \vdots \\ \vdots & & \ddots & \\ u_{m1} & \cdots & & u_{mn} \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & \sigma_n \end{pmatrix} \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & & \vdots \\ \vdots & & \ddots & \\ v_{n1} & \cdots & & v_{nn} \end{pmatrix}$$

where the matrix U is no longer square (so it can't be orthogonal) but still has orthonormal columns

If $m < n$:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & \cdots & & a_{2n} \\ \vdots & & & \ddots & & \vdots \\ a_{m1} & \cdots & & & & a_{mn} \end{pmatrix}$$

$$= \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & & \vdots \\ \vdots & & \ddots & \\ u_{n1} & \cdots & & u_{nn} \end{pmatrix}$$

$$\times \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots & & \vdots \\ 0 & \cdots & & \sigma_n & 0 & \cdots & 0 \end{pmatrix}$$

$$\times \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n} & \cdots & v_{1m} \\ v_{21} & v_{22} & \cdots & v_{2n} & \cdots & v_{2m} \\ \vdots & & \ddots & & & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nn} & \cdots & v_{nm} \\ \vdots & & & & \ddots & \vdots \\ v_{m1} & \cdots & & v_{mn} & \cdots & v_{mm} \end{pmatrix}$$

In which case the reduced SVD is

$$A = \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & & \vdots \\ \vdots & & \ddots & \\ u_{n1} & \cdots & & u_{nn} \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & \sigma_n \end{pmatrix} \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n} & \cdots & v_{1m} \\ v_{21} & v_{22} & \cdots & v_{2n} & \cdots & v_{2m} \\ \vdots & & \ddots & & & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nn} & \cdots & v_{nm} \end{pmatrix}$$

****

3 Properties of the SVD

Recall r is the rank of A; the number of nonzero singular values of A

==

$$\begin{aligned} \text{range } (A) &= \text{span } \{u_1, u_2, \ldots, u_r\} \\ \text{range } (A^T) &= \text{span } \{v_1, v_2, \ldots, v_r\} \\ \text{null } (A) &= \text{span } \{v_{r+1}, v_{r+2}, \ldots, v_n\} \\ \text{null } (A^T) &= \text{span } \{u_{r+1}, u_{r+2}, \ldots, u_m\} \end{aligned}$$

$$\text{For } A \in \mathbb{R}^{n\times n}, \quad |\det A| = \prod_{i=1}^{n} \sigma_i$$

==

The SVD of an $m \times n$ matrix A leads to an easy proof that the image of the unit sphere S^{n-1} under left-multiplication by A is a hyperellipse with semimajor axes of length $\sigma_1, \sigma_2, \ldots, \sigma_n$

==

The *condition number* of an $m \times n$ matrix A, with $m \geq n$, is

$$\kappa(A) = \frac{\sigma_1}{\sigma_n}$$

Used in numerics, $\kappa(A)$ is a measure of how close A is to being singular *with respect to floating-point computation*

==

The *2-norm* of A is $\|A\|_2 := \sup \left\{ \|Ax\|_2 \;\middle|\; \|x\|_2 = 1 \right\}$

The *Frobenius norm* of A is $\|A\|_F := \left(\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \right)^{1/2}$

We have

$$\|A\|_2 = \sigma_1 \quad \text{and} \quad \|A\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2}$$

since both matrix norms are invariant under orthogonal transformations (multiplication by orthogonal matrices)

===

Note that although the singular values of A are uniquely determined, the left and right singular vectors are only determined up to a sequence of sign choices for the columns of either U or V

===

So the SVD is not generally unique, there are $2^{(\max m,n)}$ possible SVD's for a given matrix A

===

If we fix signs for, say, column 1 of V, then the sign for column 1 of U is determined - recall $AV = U\Sigma$

4 Symmetric Orthogonalization

For nonsingular A, the matrix $L := UV^T$ is called the *symmetric orthogonalization* of the matrix A

==

L is unique since any sequence of sign choices for the columns of V determines a sequence of signs for the columns of U

==

$$
\begin{aligned}
L_{ij} &= U_{i1}(V^T)_{1j} + U_{i2}(V^T)_{2j} + U_{i3}(V^T)_{3j} + \cdots + U_{in}(V^T)_{ni} \\
&= U_{i1}V_{j1} + U_{i2}V_{j2} + U_{i3}V_{j3} + \cdots + U_{in}V_{jn}
\end{aligned}
$$

==

Like Gram-Schmidt orthogonalization, it takes as input a linearly independent set (the columns of A) and outputs an orthonormal set

(Classical) Gram-Schmidt is unstable due to repeated subtractions; Modifed Gram-Schmidt remedies this

==

But occasionally we want to disturb the original set of vectors as little as possible

==

Theorem 4.1 Over all orthogonal matrices Q, , $\|A - Q\|_F$ is minimized when $Q = L$.

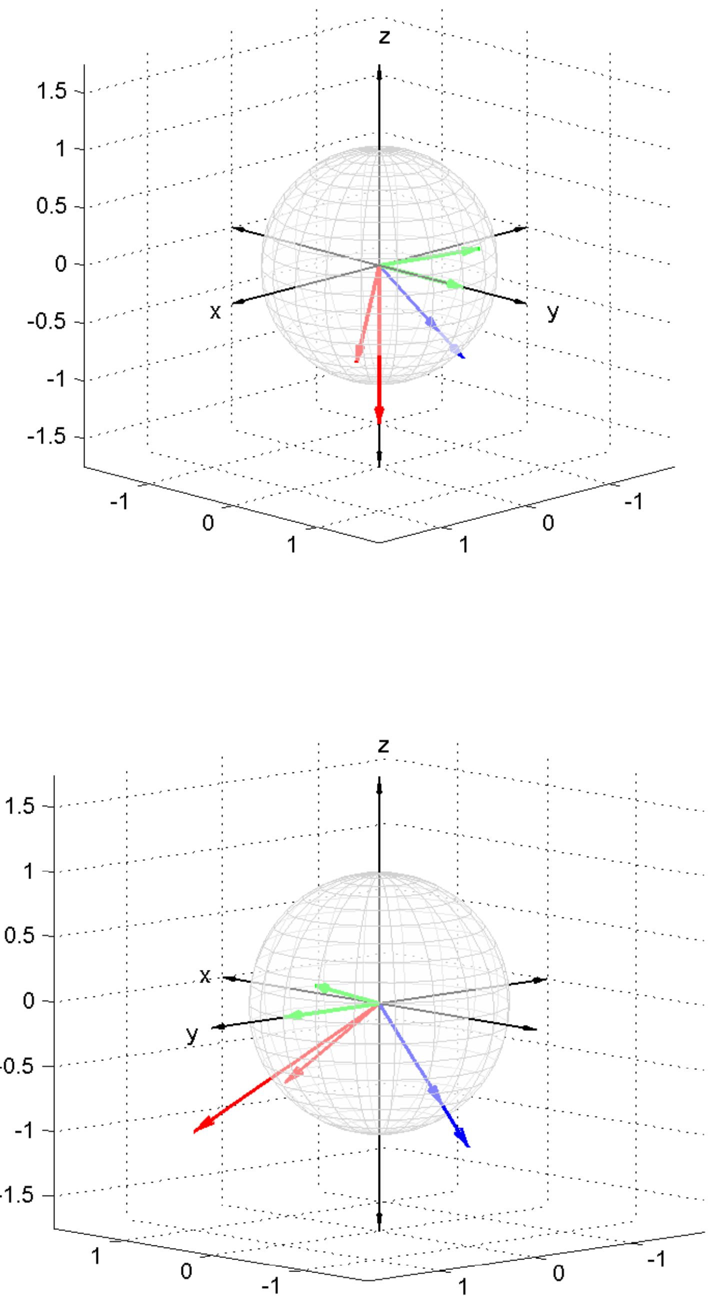


Figure 6: The columns of $L := UV^T$ and the columns of A

5 Applications of the SVD

Symmetric Orthogonalization was invented by a Swedish chemist, Per-Olov Löwdin, for the purpose of orthogonalizing hybrid electron orbitals

==

Also has application in 4G wireless communication standard, Orthogonal Frequency-Division Multiplexing (OFDM)

==

Nonorthogonal carrier waves with ideal properties, good time-frequency localization, orthogonalized in this manner have maximal TF-localization among all orthogonal carriers

==

Carrier waves are continuous (complex-valued) functions and not matrices, but there is an inner product defined for pairs of carrier waves via integration

==

With that inner product, the Gram matrix of the set of carrier waves can be computed

The symmetrically orthogonalized Gram matrix is then used to provide coefficients for linear combinations of the carrier waves

==

These linear combinations are orthogonal (hence suitable for OFDM) and optimally TF-localized

==

The SVD also has a natural application to finding the least squares solution to $Ax = b$ (i.e., a vector x with minimal $\|Ax - b\|_2$) where $Ax = b$ is inconsistent (e.g., $A \in \mathbb{R}^{m\times n}$, $m > n$, $r = n$)

==

But perhaps the most visually striking property of the SVD comes from an application in image compression

We can rewrite Σ as

$$\begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & \sigma_n \end{pmatrix} =$$

$$\underbrace{\begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & 0 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & 0 \end{pmatrix}}_{\Sigma_1} + \underbrace{\begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & 0 \end{pmatrix}}_{\Sigma_2} + \quad \cdots + \underbrace{\begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & \sigma_n \end{pmatrix}}_{\Sigma_n}$$

$$= \Sigma_1 + \Sigma_2 + \cdots + \Sigma_n$$

==

Now consider the SVD

$$A = \left(U_1 \middle| U_2 \middle| \cdots \middle| U_n \right) \left(\Sigma_1 + \Sigma_2 + \cdots + \Sigma_n \right) \begin{pmatrix} (V_1)^T \\ \hline (V_2)^T \\ \hline \vdots \\ \hline (V_n)^T \end{pmatrix}$$

and focus on, say, the first term

$$\left(\begin{array}{c|c|c|c} U_1 & U_2 & \cdots & U_n \end{array}\right) \left(\Sigma_1 \right) \left(\begin{array}{c} (V_1)^T \\ \hline (V_2)^T \\ \hline \vdots \\ \hline (V_n)^T \end{array}\right)$$

$$= \left(\begin{array}{c|c|c|c} \sigma_1 U_1 & 0 & \cdots & 0 \end{array}\right) \left(\begin{array}{c} (V_1)^T \\ \hline (V_2)^T \\ \hline \vdots \\ \hline (V_n)^T \end{array}\right)$$

$$= \left(\begin{array}{c|c|c|c} \sigma_1 U_1 & 0 & \cdots & 0 \end{array}\right) \left(\begin{array}{c} (V_1)^T \\ \hline 0 \\ \hline \vdots \\ \hline 0 \end{array}\right)$$

$$= \sigma_1 U_1 (V_1)^T$$

In general $$U\Sigma_k V^T = \sigma_k U_k (V_k)^T$$

So $$A = \sum_{j=1}^{n} \sigma_j U_j (V_j)^T$$

which is an expression of A as a sum of rank-one matrices

In this representation of A, we can consider partial sums

==

For any k with $1 \leq k \leq n$, define

$$A^{(k)} = \sum_{j=1}^{k} \sigma_j U_j (V_j)^T$$

==

This amounts to discarding the smallest $n - k$ singular values *and their corresponding singular vectors*, and storing only the V_j's and the $s_j U_j$'s

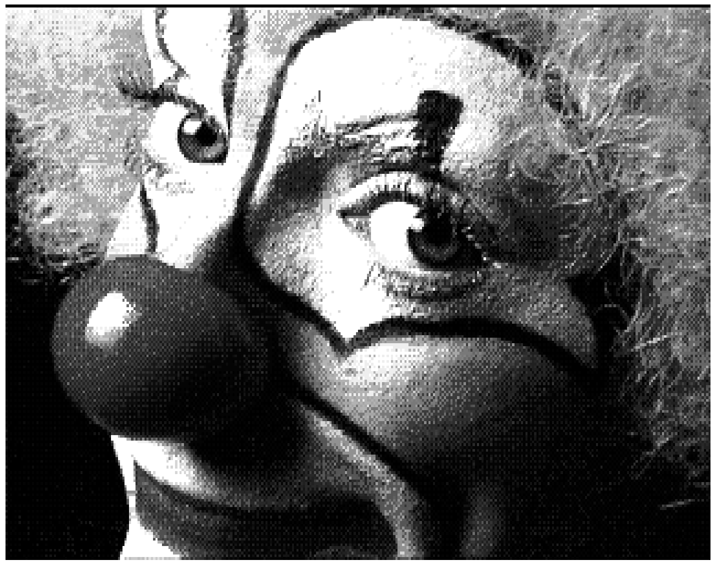
==

Theorem 5.1 Among all rank-k matrices P, $\|A - P\|_F$ is minimized for $P = A^{(k)}$

==

Theorem 5.1 says that the k^{th} partial sum of $A^{(n)}$ captures as much of the "energy" of A as possible

Example Consider the 320-by-200-pixel image below



This is stored as a 320×200 matrix of grayscale values, between 0 (black) and 1 (white), denoted by A_{clown}

==

We can take the SVD of A_{clown}

By Theorem 5.1, $A_{\text{clown}}^{(k)}$ is the best rank-k approximation to A_{clown}, measured by the Frobenius norm

==

Storage required for $A_{\text{clown}}^{(k)}$ is a total of $(320 + 200) \cdot k$ bytes for storing $\sigma_1 u_1$ through $\sigma_k u_k$ and v_1 through v_k

==

$320 \cdot 200 = 64,000$ bytes required to store A_{clown} explicitly

==

Now consider the rank-20 approximation to the original image, and the difference between the images

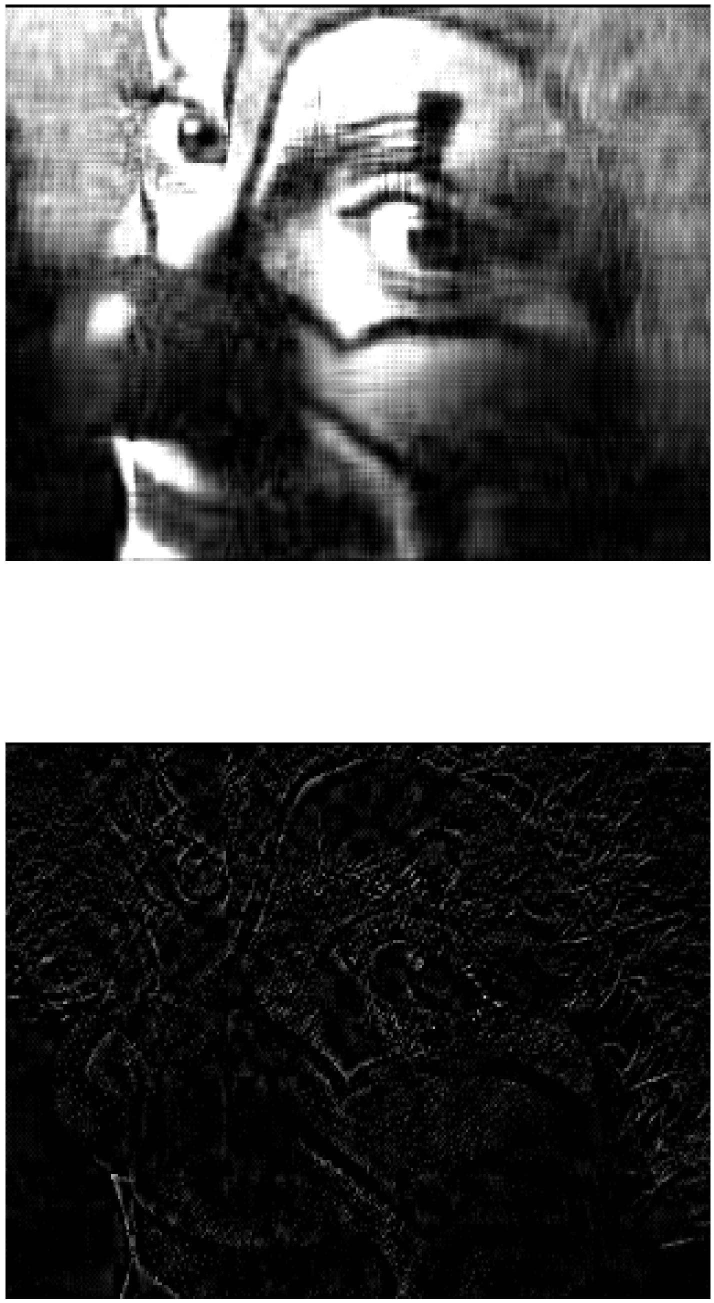


Figure 7: Rank-20 approximation $A^{(20)}_{\text{clown}}$ and $A - A^{(20)}_{\text{clown}}$

The original image took 64 kb, while the low-rank approximation required $(320 + 200) \cdot 20 \;=\; 10.4$ kb, a compression ratio of .1625

==

The SVD can also make you rich - but that's a topic for another time...

==

For further investigation, see

"Numerical Linear Algebra" by Trefethen

"Applied Numerical Linear Algebra" by Demmel

"Matrix Analysis" by Horn and Johnson

"Matrix Computations" by Golub and van Loan