1 SVD

Let $A$ be an arbitrary $m \times n$ real matrix. We can write $A$ as

$$A = USV^T, \quad (1.1)$$

where

- $U$ is an orthonormal $m \times m$ matrix,
- $V$ is an orthonormal $n \times n$ matrix, and
- $S$ is a diagonal matrix with nonnegative entries on the diagonal, arranged in decreasing order.

This matrix factorization is called the SVD (singular value decomposition) of $A$. In these notes we will not construct the SVD, but will discuss its properties.

You may also want to look the first 3 sections in http://www.cs.princeton.edu/courses/archive/spring12/cos598c/svdchapter.pdf for a different approach (but it is not required reading).

2 A simple factorization

2.1 4 basic subspaces

We check that $\mathbb{R}^n$ can be decomposed orthogonally into $N(A)$ and $R(A^T)$, and $\mathbb{R}^m$ can be decomposed orthogonally into $N(A^T)$ and $R(A)$. First, check that

$$N(A) = R(A^T)^\perp. \quad (2.1)$$

To do this, assume $x \in N(A)$. Then if $z \in R(A^T)^\perp$, $z = A^Ty$ for some $y$ in $\mathbb{R}^m$, and so

$$z^T x = y^T A x = 0, \quad (2.2)$$

so $x \in R(A^T)^\perp$, and so $N(A) \subset R(A^T)^\perp$. On the other hand, if $x \in R(A^T)^\perp$, for every $y \in \mathbb{R}^m$,

$$x^T (A^T y) = y^T (Ax) = 0; \quad (2.3)$$

thus $Ax = 0$ (any vector whose inner product with everything is 0 is 0), and so $x \in N(A)$, and so $R(A^T)^\perp \subset N(A)$, so $R(A^T)^\perp = N(A)$.

A similar argument shows $R(A)^\perp = N(A^T)$.

2.2 A simple factorization

Let $Q_1$ be an orthonormal basis for $R(A^T)$, $Q_2$ be an orthonormal basis for $N(A)$, $P_1$ be an orthonormal basis for $R(A)$, $P_2$ be an orthonormal basis for $N(A^T)$. Set $r_1 = \text{dimension}(R(A^T))$ and $r_2 = \text{dimension}(R(A))$; we know $r_1 = r_2$. We have seen this before, but with our current tools it is easy:

$$y \in R(A^T) \text{ and } Ay = 0 \implies y = 0. \quad (2.4)$$

So the columns of $AQ_1$ are independent, else there is a nonzero vector $b$ such that $0 = AQ_1 b$; but also $Q_1 b \in R(A^T)$, so $Q_1 b = 0$, and since $Q_1$ independent, $b = 0$. Thus $r_1 \leq r_2$. The other inequality works the same way with everything transposed.

Let $Q = (Q_1 \quad Q_2)$, and $P = (P_1 \quad P_2)$. Then

$$P^T AQ = \begin{pmatrix} P_1^T \\ P_2^T \end{pmatrix} A \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} = \begin{pmatrix} P_1^T AQ_1 & P_1^T AQ_2 \\ P_2^T AQ_1 & P_2^T AQ_2 \end{pmatrix} \quad (2.5)$$

$$= \begin{pmatrix} r \underline{\begin{pmatrix} P_1^T AQ_1 \\ P_2^T AQ_1 \end{pmatrix}} \\ n-r \underline{\begin{pmatrix} 0 \\ 0 \end{pmatrix}} \end{pmatrix}, \quad (2.6)$$
where \( C = P_1^T AQ_1 \) is an invertible \( r \times r \) matrix. Then
\[
A = P \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} Q^T. \tag{2.7}
\]
This is not quite what we wanted, as \( C \) is not necessarily diagonal or nonnegative, but it gives a flavor of the construction of an SVD. It also gives a useful decomposition: removing the zero blocks gives that for any matrix \( A \) of rank \( r \), we can write \( A = WZ \), with \( W \) an \( m \times r \) matrix and \( Z \) an \( r \times n \) matrix; for example, take \( W = P_1 \) and \( Z = CQ_1^T \).

### 3 Frobenious Norm

Given an \( m \times n \) matrix \( A \), the Frobenious norm of \( A \) is defined by
\[
\|A\|_F^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij}^2. \tag{3.1}
\]
This is equivalent to “rasterizing” the matrix, and taking its norm considered as a vector in \( \mathbb{R}^{mn} \). It is also equivalent to taking the sum of the square norm of each row of \( A \), or the sum of the square norm of each column of \( A \). This means that if \( O \) is an \( m \times m \) orthonormal matrix,
\[
\|OA\|_F = \|A\|_F, \tag{3.2}
\]
and if \( O \) is an \( n \times n \) orthonormal matrix,
\[
\|AO\|_F = \|A\|_F. \tag{3.3}
\]

### 4 Best low rank approximation

In this section, we assume that it is possible to find an SVD of an arbitrary matrix \( A \); this will be proven later. Recall that we specified that the diagonal entries of \( S \) should be arranged in decreasing order. Then the best rank \( r \) approximation to \( A \) is given by
\[
\arg \min_{B \in \mathbb{R}^{m \times n}, \text{rank}(B) = r} \|A - B\|_F^2 = U_r S_r V_r^T, \tag{4.1}
\]
where \( U_r \) is the first \( r \) columns of \( U \), \( V_r \) is the first \( r \) columns of \( V \), and \( S_r \) is the upper left \( r \times r \) block of \( S \).

To see this:

- By the orthonormal invariance of the Frobenious norm, we may assume \( A \) takes on the properties of \( S \), so that \( A \) is diagonal, nonnegative, and the diagonal entries decrease in value. In other words, make the change of variable \( C = U^T BV \), and solve
  \[
  \arg \min_{C \in \mathbb{R}^{m \times n}, \text{rank}(C) = r} \|S - C\|_F^2. \tag{4.2}
  \]

- By the last line in 2.2, we can write \( C = WZ \), where \( W \) is an \( m \times r \) matrix with orthonormal columns and \( Z \) is an \( r \times n \) matrix. Our problem becomes
  \[
  \arg \min_{W \in \mathbb{R}^{m \times r}, \text{W o.n.}, \ Z \in \mathbb{R}^{r \times n}} \|S - WZ\|_F^2. \tag{4.3}
  \]

We will now finish the argument twice. Once for the case \( r = 1 \), and once for the general case. The two cases are exactly the same; I write the argument twice because it appears cleaner in the case \( r = 1 \), and it may be easier for you to not get lost in details. On the other hand, it may be confusing to you to see something broken into two cases when the cases are the same; in that case, skip right to the general case.
4.1 \( r = 1 \), for simplicity

Our problem becomes

\[
\arg \min_{w \in \mathbb{R}^m, \|w\| = 1} \|S - wz^T\|_F^2 = \arg \min_{w \in \mathbb{R}^m, \|w\| = 1} \sum_{j=1}^{\min(m,n)} \|S_j - z(j)w\|^2. \tag{4.4}
\]

Here the limit of summation is \( \min(m, n) \) because \( S_j \) is the zero vector after that, if \( m < n \). Each entry of \( z \) can be changed independently of any other entry in \( z \) (this is not true of \( w \), because of the constraint that \( w \) has norm 1); but for any fixed \( w \) with \( \|w\| = 1 \),

\[
z(j) = z(j, w) = \arg \min_{\alpha \in \mathbb{R}} \|S_j - \alpha w\|^2 = w^T S_j = S_{jj} w(j). \tag{4.5}
\]

This is just the least squares solution. Then

\[
\arg \min_{w \in \mathbb{R}^m, \|w\| = 1} \sum_{j=1}^{\min(m,n)} \|S_j - z(j)w\|^2 = \arg \min_{w \in \mathbb{R}^m, \|w\| = 1} \sum_{j=1}^{\min(m,n)} \|S_j - S_{jj} w(j)w\|^2, \tag{4.6}
\]

and following our standard manipulations:

\[
\arg \min_{w \in \mathbb{R}^m, \|w\| = 1} \sum_{j=1}^{\min(m,n)} S_{jj}^2 - S_{jj}^2 w(j)^2 = \arg \min_{w \in \mathbb{R}^m, \|w\| = 1} - \sum_{j=1}^{\min(m,n)} S_{jj}^2 w(j)^2 = \arg \max_{w \in \mathbb{R}^m, \|w\| = 1} \sum_{j=1}^{\min(m,n)} S_{jj}^2 w(j)^2 \tag{4.7}
\]

Recall the diagonal entries of \( S \) are written in decreasing (non-increasing) order, so \( S_{11} \geq S_{jj} \) for \( j > 1 \). So for any \( w \) with \( \|w\| = 1 \),

\[
\sum_{j=1}^{\min(m,n)} S_{jj}^2 w(j)^2 \leq \sum_{j=1}^{\min(m,n)} S_{11}^2 w(j)^2 = S_{11}^2; \tag{4.9}
\]

but this is attained for \( w \) with a 1 in the first position and 0 otherwise. Tracing back through equations (4.5) and (4.2), we see that \( C_{11} = S_{11} \), and \( C_{ij} = 0 \) for \( i, j \neq 1 \), as desired.

4.2 General \( r \)

This is the same argument as above. For any \( W \in \mathbb{R}^{m \times r} \) and \( Z \in \mathbb{R}^{r \times n} \)

\[
\|S - WZ\|_F^2 = \sum_{j=1}^{\min(m,n)} \|S_j - WZ_j\|^2. \tag{4.10}
\]

In the minimization problem, each column of \( Z \) can be changed independently of any other column of \( Z \) (this is not true of \( W \), because of the constraint that \( W \) has orthonormal columns). For any \( W \) with orthonormal columns,

\[
Z_j = \arg \min_{\alpha \in \mathbb{R}^r} \|S_j - W\alpha\|^2 = W^T S_j, \tag{4.11}
\]

so

\[
\arg \min_{W \in \mathbb{R}^{m \times r}, \text{ o.n.}} \sum_{j=1}^n \|S_j - WZ_j\|^2 = \arg \min_{W \in \mathbb{R}^{m \times r}, \text{ o.n.}} \sum_{j=1}^{\min(m,n)} \|S_j - WW^T S_j\|^2 \tag{4.12}
\]
\[
\begin{align*}
\min_{W \in \mathbb{R}^{m \times r}, \text{or.}} \sum_{j=1}^{\min(m,n)} (||S_j||^2 - ||WW^T S_j||^2) &= \max_{W \in \mathbb{R}^{m \times r}, \text{or.}} \sum_{j=1}^{\min(m,n)} (||WW^T S_j||^2 - ||W W^T S_j||^2) \\
\min_{W \in \mathbb{R}^{m \times r}, \text{or.}} \sum_{j=1}^{\min(m,n)} ||W W^T S_j||^2 &= \max_{W \in \mathbb{R}^{m \times r}, \text{or.}} \sum_{j=1}^{\min(m,n)} ||W^T S_j||^2 \\
\end{align*}
\]

Since \(W^T S_j\) is just \(S_{jj}\) times the \(j\)th row of \(W\),
\[
= \max_{W \in \mathbb{R}^{m \times r}, \text{or.}} \sum_{j=1}^{\min(m,n)} S_{jj}^2 \sum_{i=1}^{r} |W_{ji}|^2.
\]

Consider the problem
\[
\max_{\sum a_j = r, \sum_{0 \leq a_j \leq 1}} \sum_{j=1}^{\min(m,n)} S_{jj}^2 a_j.
\]

This problem has maximum value \(\sum_{j=1}^{r} S_{jj}^2\), obtained when \(a_j = 1\) for \(j \in \{1, ..., r\}\) (recall \(S_{jj}\) are nonincreasing in \(j\)). But since \(W\) has orthonormal columns, \(\sum_{i=1}^{r} |W_{ji}|^2 \leq 1\) for every \(j \in \{1, ..., m\}\), and \(\sum_{j=1}^{m} \sum_{i=1}^{r} |W_{ji}|^2 = r\). That is, \(\sum_{i=1}^{r} |W_{ji}|^2\) satisfies the constraints of the problem (4.16), so
\[
\max_{W \in \mathbb{R}^{m \times r}, \text{or.}} \sum_{j=1}^{\min(m,n)} S_{jj}^2 \sum_{i=1}^{r} |W_{ji}|^2 \leq \max_{\sum a_j = r, \sum_{0 \leq a_j \leq 1}} \sum_{j=1}^{\min(m,n)} S_{jj}^2 a_j = \sum_{j=1}^{r} S_{jj}^2.
\]

On the other hand, for any \(W\) with an orthonormal \(r \times r\) block in the first \(r\) entries, but zero otherwise, \(\sum_{j=1}^{m} S_{jj}^2 \sum_{i=1}^{r} |W_{ji}|^2 = \sum_{j=1}^{r} S_{jj}^2\), so any such \(W\) is a maximizer. For such a \(W\), \(WW^T\) has the identity in the first \(r \times r\) block, but zeros elsewhere, and so as before, following back through the equations shows \(C = S\) for the upper left \(r \times r\) block, and is 0 elsewhere, as desired.

## 5 Problems

1. Suppose \(W\) is an \(m \times n\) matrix with orthonormal columns. What is \(||W||_F||\)?
2. Check carefully that if \(O\) is an orthonormal \(m \times m\) matrix, and if \(A\) is an \(m \times n\) matrix, \(||OA||_F = ||A||_F\).
3. If \(A = USV^T\) is an SVD of \(A\), write a formula for \(||A||_F\) in terms of the diagonal elements of \(S\).
4. Suppose \(W\) is an \(m \times n\) matrix with orthonormal columns. Show that the norm of each row of \(W\) is less than or equal to 1.
5. Work through the manipulations in equations (4.13), (4.14), and (4.15)
6. Find the flaw in the following argument going straight from (4.2) to the conclusion: Since \(S\) is zero for non-diagonal entries, it is clear that you pay a penalty for any off diagonal entry in \(C\) in (4.2). Thus we may conclude that \(C\) is also diagonal. The only diagonal matrices of rank \(r\) are nonzero in precisely \(r\) locations on the diagonal; since \(S_{jj}\) is non-decreasing in \(j\), keeping the first \(r\) entries gives the least error.