# 4 SingularValue Decomposition (SVD)

The singular value decomposition of a matrix $A$ is the factorization of $A$ into the product of three matrices $A = UDV^T$ where the columns of $U$ are orthonormal and similarly the columns of $V$ are orthonormal. The matrix $D$ is diagonal and has positive (real) diagonal entries. The SVD is useful in many tasks. Here we mention two examples. First the rank of a matrix $A$ can be read off from its SVD. It is just the size of $D$. The second example is that for a square and invertible matrix $A$, the inverse of $A$ is $VD^{-1}U^T$ .

To motivate the SVD, we treat the rows of an $n \times d$ matrix $A$ as $n$ points in a $d$-dimensional space and consider the problem of finding the best $k$ dimensional subspace with respect to the set of points. Here best means minimize the sum of the squares of the perpendicular distances of the points to the subspace. We begin with a special case of the problem where the subspace is 1-dimensional, a line through the origin. We will see later that we can find the best-fitting $k$-dimensional subspace by $k$ applications of the best fitting line algorithm. Finding the best fitting line through the origin with respect to a set of points $\left\{ \left( x_i, y_i \right) | 1 \le i \le n \right\}$ in the plane means minimizing the sum of squared distances of the points to the line. Here distance is measured perpendicular to the line. The problem is called the *best least squares fit*.

In the best least squares fit, one is minimizing the distance to a subspace. A similar problem is to find the function that best fits some data. Here one variable y is a function of the variables $x_1, x_2, \cdots, x_n$ and one wishes to minimize the vertical distance to the subspace of the $x_i$ rather than the perpendicular distance to the subspace.

Let $y = mx$ be a line through the origin. Project the point $\left( x_i, y_i \right)$ onto the line y=mx. Then

$$x_i^2 + y_i^2 = \left(\text{length of projection}\right)^2 + \left(\text{distance of point to the line}\right)^2 .$$

See Figure 4.1. Thus

$$\left(\text{distance of point to the line}\right)^2 = x_i^2 + y_i^2 - \left(\text{length of projection}\right)^2 .$$

To minimize the sum of the squares of the distances to the line, one could minimize $\sum_{i=1}^{n} \left( x_i^2 + y_i^2 \right)$ minus the sum of the squares of the lengths of the projections of the points onto the line. However, $\sum_{i=1}^{n} \left( x_i^2 + y_i^2 \right)$ is a constant (independent of the line), so minimizing the sum of the squares of the distances is equivalent to maximizing the sum of the squares of the lengths of the projections onto the line. Similarly for best-fit subspaces, we could maximize the sum of the squared lengths of the projections onto the subspace instead of minimizing the sum of squared distances to the subspace.

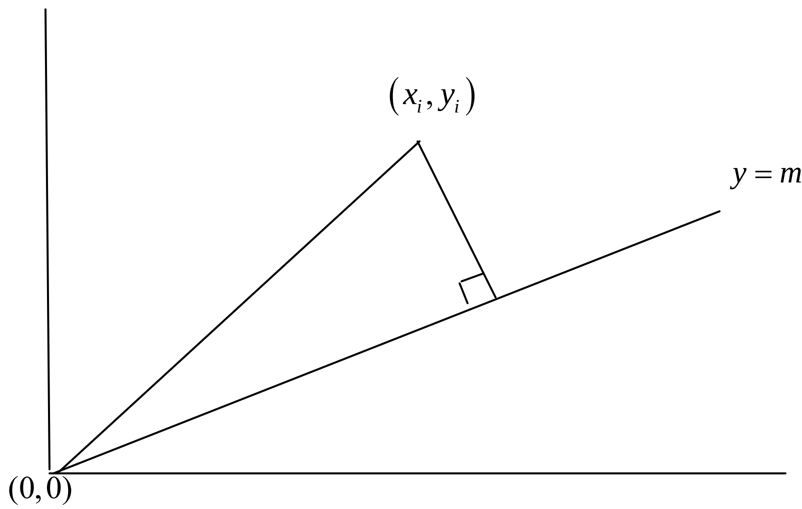

**Figure 4.1**: The projection of the point $(x_i, y_i)$ onto the line y=mx.

## 4.1 Singular vectors

We now define the *singular vectors* of an $n \times d$ matrix *A*. Consider the rows of *A* as *n* points in a *d* dimensional space. Consider the best fit line through the origin. Let **v** be a unit vector along this line. The length of the projection of the $i^{th}$ row of *A*, $\mathbf{a_i}$, onto $\mathbf{v}$ is $|\mathbf{a_i} \cdot \mathbf{v}|$ and from this we see that the sum of length squared of the projections is $|A\mathbf{v}|^2$. The best fit line is the one maximizing $|A\mathbf{v}|^2$ and hence minimizing the sum of the squared distances of the points to the line.

With this in mind, define the *first singular vector*, $\mathbf{v_1}$ of A, which is a column vector, as the best fit line through the origin for the *n* points in *d* space which are the rows of *A*. Thus

$$\mathbf{v_1} = \arg\max_{|\mathbf{v}|=1} |A\mathbf{v}| .$$

The value $\sigma_1(A) = |A\mathbf{v_1}|$ is called the *first singular value* of *A*. Note that $\sigma_1^2$ is the sum of the squares of the projections of the points to the line determined by $v_1$.

The greedy approach to find the best fit 2-dimensional subspace for a matrix *A*, takes $\mathbf{v_1}$ as the first basis vector for the 2-dimenional subspace and finds the best 2-dimensional subspace containing $\mathbf{v_1}$. The fact that we are using sum of squared distances helps us. For every 2-dimensional subspace containing $\mathbf{v_1}$, the sum of squared lengths of the projections onto the subspace is just the sum of squared projections onto $\mathbf{v_1}$ plus the sum of squared projections along a vector perpendicular to $\mathbf{v_1}$ in the subspace. Thus, instead of looking for the best 2-dimensional subspace containing $\mathbf{v_1}$, we look for a unit vector, call it $\mathbf{v_2}$, perpendicular to $\mathbf{v_1}$ which maximizes $|A\mathbf{v}|^2$ among all such unit vectors. The same argument shows that in pursuing a greedy strategy for finding the best three or higher dimensional subspaces, we should define $\mathbf{v_3}, \mathbf{v_4}, \ldots$ in a similar manner. This is what the following definitions capture. There is no apriori guarantee that the greedy approach gives the best fit. But, in fact, the greedy algorithm does work and yields the best-fit subspaces of every dimension.

The *second singular vector*, $\mathbf{v_2}$, is defined by the best fit line perpendicular to $\mathbf{v_1}$:

$$v_2 = \arg\max_{v \perp v_1, |v|=1} |Av|$$

.

The value $\sigma_2(A) = |A\mathbf{v_2}|$ is called the *second singular value* of *A***.** The *third singular vector* $\mathbf{v_3}$ is defined similarly by

$$\mathbf{v_3} = \arg\max_{\mathbf{v} \perp \mathbf{v_1}, \mathbf{v_2}, |\mathbf{v}|=1} |A\mathbf{v}|$$

and so on. The process stops when we have found

$$\mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_r}$$

as singular vectors and

$$\arg\max_{\mathbf{v} \perp \mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_r}, |\mathbf{v}|=1} |A\mathbf{v}| = 0 .$$

Now, we give a simple proof that the greedy algorithm indeed finds the best subspaces of every dimension.

**Theorem 4.1:** Let *A* be an $n \times d$ matrix and suppose $\mathbf{v_1}, \mathbf{v_2}, \ldots \mathbf{v_r}$ are the singular vectors defined above. For $1 \le k \le r$, let $V_k$ be the subspace spanned by $\mathbf{v_1}, \mathbf{v_2}, \ldots \mathbf{v_k}$. Then for each $k$, $V_k$ is the best-fit $k$ dimensional subspace for *A*.

**Proof**: The statement is obviously true for $k = 1$. For $k = 2$, let $W$ be a best-fit 2-dimensional subspace for *A*. For any basis $\mathbf{w_1}, \mathbf{w_2}$ of $W$, $|A\mathbf{w_1}|^2 + |A\mathbf{w_2}|^2$ is the sum of squared lengths of the projections of the rows of *A* onto $W$. Now, choose a basis $\mathbf{w_1}, \mathbf{w_2}$ of $W$ so that $\mathbf{w_2}$ is perpendicular to $\mathbf{v_1}$. If $\mathbf{v_1}$ is perpendicular to $W$, any unit vector in $W$ will do as $\mathbf{w_2}$. If not, we choose $\mathbf{w_2}$ to be the unit vector in $W$ perpendicular to the projection of $\mathbf{v_1}$ onto $W$.
Since $\mathbf{v_1}$ was chosen to maximize $|Av_1|^2$, $|A\mathbf{w_1}|^2 \le |A\mathbf{v_1}|^2$ and since $\mathbf{v}_2$ was chosen to maximize $|Av_2|^2$ over all $\mathbf{v}$ perpendicular to $\mathbf{v_1}$, $|A\mathbf{w}_2|^2 \le |Av_2|^2$. Thus

$$|A\mathbf{w_1}|^2 + |A\mathbf{w_2}|^2 \le |A\mathbf{v_1}|^2 + |A\mathbf{v_2}|^2 .$$

Hence, $V_2$ is at least as good as $W$ and so is a best-fit 2-dimensional subspace.

For general $k$, proceed by induction. By the induction hypothesis $V_{k-1}$ is a best-fit *k-1* dimensional subspace. Suppose $W$ is a best-fit $k$ dimensional subspace. Choose a basis $\mathbf{w_1}, \mathbf{w_2}, \ldots \mathbf{w_k}$ of $W$ so that $\mathbf{w_k}$ is perpendicular to $\mathbf{v_1}, \mathbf{v_2}, \ldots \mathbf{v_{k-1}}$. Then

$$|A\mathbf{w_1}|^2 + |A\mathbf{w_2}|^2 + \ldots |A\mathbf{w_k}|^2 \le |A\mathbf{v_1}|^2 + |A\mathbf{v_2}|^2 + \ldots |A\mathbf{v_{k-1}}|^2 + |A\mathbf{w_k}|^2$$

by the fact that $V_{k-1}$ is an optimal *k-1* dimensional subspace. Since $\mathbf{w_k}$ is perpendicular to $\mathbf{v_1}, \mathbf{v_2}, \ldots \mathbf{v_{k-1}}$. By the definition of $\mathbf{v_k}$, $\left|Aw_k\right|^2 \le \left|Av_k\right|^2$ and thus

$$|A\mathbf{v_1}|^2 + |A\mathbf{v_2}|^2 + \ldots + |A\mathbf{v_{k-1}}|^2 + |A\mathbf{w_k}|^2 \le |A\mathbf{v_1}|^2 + |A\mathbf{v_2}|^2 + \ldots |A\mathbf{v_{k-1}}|^2 + |A\mathbf{v_k}|^2,$$

proving that $V_k$ is at least as good as $W$ and hence is optimal.

■

Note that the $n$-vector $A\mathbf{v_i}$ is really a list of lengths (with signs) of the projections of the rows of $A$ onto $\mathbf{v_i}$. Thus, think of $|A\mathbf{v_i}| = \sigma_i(A)$ as the ``component'' of the matrix $A$ along $\mathbf{v_i}$. For this interpretation to make sense, it should be true that adding up the squares of the components along each of the $\mathbf{v_i}$, gives the square of the ``whole content of the matrix $A$''. This is the matrix analogy of decomposing a vector into its components along orthogonal directions. The length squared of the whole vector then would be the sum of squares of its components. This is indeed the case.

Consider one row, say $\mathbf{a_j}$, of $A$. Since $\mathbf{v_1}, \mathbf{v_2}, \ldots \mathbf{v_r}$ span the space of all rows of A, $\mathbf{a_j} \cdot \mathbf{v} = 0$ for all $\mathbf{v}$ perpendicular to $\mathbf{v_1}, \mathbf{v_2}, \ldots \mathbf{v_r}$. Thus, for each row $\mathbf{a_j}$, $\sum_{i=1}^{r} (\mathbf{v_i} \cdot \mathbf{a_j})^2 = |\mathbf{a_j}|^2$. Summing over all rows $j$,

$$\sum_{j=1}^{n} |\mathbf{a_j}|^2 = \sum_{i=1}^{r} \sum_{j=1}^{n} (\mathbf{v_i} \cdot \mathbf{a_j})^2 = \sum_{i=1}^{r} |A\mathbf{v_i}|^2 = \sum_{i=1}^{r} \sigma_i^2(A).$$

But $\sum_{j=1}^{n} |\mathbf{a_j}|^2 = \sum_{j=1}^{n} \sum_{k=1}^{d} a_{jk}^2$, the sum of squares of all the entries of $A$. There is an important norm associated with this, namely, the Frobenius norm of $A$, denoted $\|A\|_F$ which is defined as

$$\|A\|_F = \sqrt{\sum_{j,k} a_{jk}^2}\ .$$

Thus, we have shown that the sum of squares of the singular values of $A$ is indeed the square of the ``whole content of $A$'', i.e., the sum of squares of all the entries.

**Lemma 4.1**: For any matrix $A$, the sum of squares of the singular values equals the Frobenius norm. That is, $\sum \sigma_i^2(A) = \|A\|_F^2$.

**Proof**: By the preceding discussion.

■

For each i, $A\mathbf{v_i}$ is a list of lengths of projections of the rows of $A$ onto $\mathbf{v_i}$. $A$ can be described fully by how it transforms the vectors $\mathbf{v_i}$. Every vector $\mathbf{v}$ can be written as a linear combination of $\mathbf{v_1}, \mathbf{v_2}, \ldots \mathbf{v_r}$ and a vector perpendicular to all the $\mathbf{v_i}$. Thus, $A\mathbf{v}$ is the same linear combination of $A\mathbf{v_1}, A\mathbf{v_2}, \ldots A\mathbf{v_r}$. So the $A\mathbf{v_1}, A\mathbf{v_2}, \ldots A\mathbf{v_r}$ form a fundamental set of vectors associated with *A*. We normalize them to length one by

$$\mathbf{u_i} = \frac{1}{\sigma_i(A)} A\mathbf{v_i}.$$

The vectors $\mathbf{u_1}, \mathbf{u_2}, \ldots \mathbf{u_r}$ are called the *left singular vectors* of *A*. The $\mathbf{v_i}$ are called the *right singular vectors*. The SVD theorem below will fully explain the reason for these terms.

For any matrix *A*, the sequence of singular values is unique and if there are no ties, then the sequence of singular vectors is unique also. When ties exist, they are broken arbitrarily. Independent of how the ties are broken, the set of singular values is always unique. However, when some set of singular values are equal the corresponding singular vectors span some subspace. Any set of orthonormal vectors spanning this subspace can be used as the singular vectors.

## 4.2 Singular Value Decomposition (SVD)

Let *A* be an $n \times d$ matrix with singular vectors $\mathbf{v_1}.\mathbf{v_2}, \cdots, \mathbf{v_r}$ and corresponding singular values $\sigma_1, \sigma_2, \cdots, \sigma_r$. Then $\mathbf{u_i} = \frac{1}{\sigma_i} A\mathbf{v_i}$ for $i = 1, 2, \cdots, r$ are the *left singular vectors* and *A* can be decomposed into a sum of rank one matrices as

$$A = \sum_{i=1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T.$$

The decomposition is called the *singular value decomposition*, *SVD*, of *A*. In matrix notation $A = UDV^T$ where the columns of *U* and *V* consist of the left and right singular vectors, respectively, and $D$ is a diagonal matrix whose entries are the singular values of *A*.

Before proceeding, we prove a simple lemma that in order to show two matrices *A* and *B* are identical it suffices to show that $A\mathbf{v} = B\mathbf{v}$ for all $\mathbf{v}$. The lemma says that in the abstract, a matrix A can be viewed as a transformation which maps vector $\mathbf{v}$ onto $A\mathbf{v}$.

**Lemma 4.2**: Matrices *A* and *B* are identical if and only if for all vectors **v**, $A\mathbf{v} = B\mathbf{v}$.

**Proof**: Clearly if *A*=*B* then $A\mathbf{v} = B\mathbf{v}$ for all **v**. For the converse, suppose that $A\mathbf{v} = B\mathbf{v}$ for all $\mathbf{v}$. Let $\mathbf{e_i}$ be the vector that is all zeros except for the i^th^ component which has value 1. Now $A\mathbf{e_i}$ is the i^th^ column of *A* and thus *A*=*B* if for each $i$, $A\mathbf{e_i} = B\mathbf{e_i}$.

■

Clearly, the right singular vectors are orthogonal by definition. We now show that the left singular vectors are also orthogonal and that $A = \sum_{i=1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$ .

**Theorem 4.2**: Let *A* be a rank *r* matrix.

(1) The left singular vectors of *A*, $\mathbf{u_1}, \mathbf{u_2}, \cdots, \mathbf{u_r}$ , are orthogonal, and

(2) $A = \sum_{i=1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$

**Proof**: The proof is by induction on *r*. For $r = 1$, there is only one $\mathbf{u_i}$ so Condition (1) is trivially true. Condition (2) simplifies to $A = \sigma_1 \mathbf{u_1} \mathbf{v_1}^T$ . To prove Condition (2), it suffices to prove that for every vector $\mathbf{v}$, $A\mathbf{v} = \sigma_1 \mathbf{u_1} \mathbf{v_1}^T \mathbf{v}$ . Write $\mathbf{v}$ as $\mathbf{v} = a\mathbf{v_1} + \mathbf{w}$ , where $a$ is a scalar and $\mathbf{w}$ is perpendicular to $\mathbf{v_1}$ . Since $r = 1$

$$\arg\max_{\mathbf{v}\perp\mathbf{v_1}\, |\mathbf{v}=1|} |A\mathbf{v}| = 0 .$$

Since ***w*** is perpendicular to $\mathbf{v_1}$ and $\arg\max_{\mathbf{v}\perp\mathbf{v_1}\, |\mathbf{v}=1|} |A\mathbf{v}| = 0$**,** it follows that $A\mathbf{w} = 0$ . Thus,

$$A\mathbf{v} = A\left(a\mathbf{v_1} + \mathbf{w}\right) = aA\mathbf{v_1} = a\sigma_1 \mathbf{u_1} .$$

Similarly,

$$\sigma_1 \mathbf{u_1} \mathbf{v_1}^T \mathbf{v} = \sigma_1 \mathbf{u_1} \mathbf{v}_1^T \left(a\mathbf{v_1} + w\right) = \sigma_1 \mathbf{u_1} a \mathbf{v_1}^T \mathbf{v_1} = \sigma_1 a \mathbf{u_1} .$$

Thus, by Lemma 4.1, $A = \sigma_1 \mathbf{u_1} \mathbf{v_1}^T$ . This finishes the proof in the case r=1.

Next, we prove the inductive part. Consider the matrix

$$B = A - \sigma_1 \mathbf{u_1} \mathbf{v_1}^T .$$

Apply the implied algorithm in the definition of singular value decomposition to *B*. We claim that a run of this algorithm is identical to a run of the algorithm on A for its second and later singular vectors/values. To see this, first observe that $B\mathbf{v_1} = 0$ . So, the first right singular vector, call it $\mathbf{z}_,$ of B will be perpendicular to $\mathbf{v_1}$ since if it had a component $\mathbf{z_1}$ along $\mathbf{v_1}$ , then, $|B\frac{\mathbf{z}-\mathbf{z_1}}{|\mathbf{z}-\mathbf{z_1}|}| = \frac{|B\mathbf{z}|}{|\mathbf{z}-\mathbf{z_1}|} > |B\mathbf{z}|$ , contradicting the $\arg\max$ definition. But for any $\mathbf{v}$ perpendicular to $\mathbf{v_1}$ , $B\mathbf{v} = A\mathbf{v}$ . Thus, the top singular vector of *B* is indeed a second singular vector of *A*. Now repeat this argument to show that a run of the algorithm on *B* is the same as a run on *A* for its second and later singular vectors; this is left as an exercise.

Thus, there is a run of the algorithm that finds that *B* has right singular vectors $\mathbf{v_2}, \mathbf{v_3}, \cdots, \mathbf{v_r}$ and corresponding left singular vectors $\mathbf{u_2}, \mathbf{u_3}, \cdots, \mathbf{u_r}$ . By the induction hypothesis, $B = \sum_{i=2}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$ and $\mathbf{u_2}, \mathbf{u_3} \cdots, \mathbf{u_r}$ are orthogonal. It follows that $A = \sum_{i=1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$ .

It still remains to prove that $\mathbf{u_1}$ is orthogonal to the other $\mathbf{u_i}$. Suppose not and for some $i \geq 2$, $\mathbf{u_1}^T\mathbf{u_i} \neq 0$. Without loss of generality assume that $\mathbf{u_1}^T\mathbf{u_i} > 0$. The proof is symmetric for the case where $\mathbf{u_1}^T\mathbf{u_i} < 0$. Now, for infinitesimally small $\varepsilon > 0$, the vector

$$A\left(\frac{\mathbf{v}_1 + \varepsilon \mathbf{v_i}}{\left|\mathbf{v_1} + \varepsilon \mathbf{v_i}\right|}\right) = \frac{\sigma_1 \mathbf{u_1} + \varepsilon \sigma_i \mathbf{u_i}}{\sqrt{1+\varepsilon^2}}$$

has length at least as large as its component along $u_1$ which is

$$\left(\sigma_1 + \varepsilon \sigma_i \mathbf{u_1}^T \mathbf{u_i}\right)\left(1 - \tfrac{\varepsilon^2}{2} + O\left(\varepsilon^4\right)\right) = \sigma_1 + \varepsilon \sigma_i \mathbf{u_1}^T \mathbf{u_i} - O\left(\varepsilon^2\right) > \sigma_1$$

which contradicts the definition of $\sigma_1$. Thus $\mathbf{u_1}, \mathbf{u_2}, \cdots, \mathbf{u_r}$ are orthogonal.

■

## 4.3 Best rank k approximations

There are two important matrix norms, the Frobenius norm denoted $\|A\|_F$ (seen already) and the 2-norm denoted $\|A\|_2$. The 2-norm of the matrix $A$ is given by

$$\max_{|\mathbf{v}|=1} \left|A\mathbf{v}\right|$$

and thus equals the largest singular value of the matrix.

Let $A$ be an $n \times d$ matrix and think of the rows of $A$ as points in *d*-dimensional space. The Frobenius norm of $A$ is the sum of the squared distance of the points to the origin. The two norm is the sum of squared distances to the origin along the direction that maximizes this quantity.

Let

$$A = \sum_{i=1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$$

be the SVD of $A$. For $k \in \{1, 2, , \cdots, r\}$, let

$$A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$$

be the sum truncated after *k* terms. It is clear that $A_k$ has rank *k*. We will see that it is the best rank *k* approximation to $A$, when the error is measured in either of the 2- norm or the Frobenius norm.

**Lemma** 4.3: The rows of $A_k$ are precisely the projections of the (corresponding) rows of $A$ onto the subspace $V_k$ defined in Theorem 4.1.

**Proof** : For an arbitrary row vector $a$, its projection onto $V_k$ is given by $\sum_{i=1}^{k}\left(a\cdot\mathbf{v_i}\right)\mathbf{v_i}^T$ since the $\mathbf{v_i}$ are orthonormal. Thus, the matrix whose rows are the projections of the rows of $A$ onto $V_k$ is given by $A\left(\sum_{i=1}^{k}\mathbf{v_i}\mathbf{v_i}^T\right)$. This last expression simplifies to

$$A\left(\sum_{i=1}^{k}\mathbf{v_i}\mathbf{v_i}^T\right)=\sum_{j=1}^{r}\sigma_j(A)\mathbf{u}_j\mathbf{v}_j^T\left(\sum_{i=1}^{k}\mathbf{v_i}\mathbf{v_i}^T\right)=\sum_{i=1}^{k}\sum_{j=1}^{r}\sigma_j(A)\mathbf{u}_j\mathbf{v}_j^T\mathbf{v}_i\mathbf{v}_i^T=\sum_{i=1}^{k}\sigma_i(A)\mathbf{u_i}\mathbf{v_i}^T=A_k$$

using orthogonality.

■

The matrix $A_k$ is the best rank k approximation to $A$ in both the Frobenius and the 2-norm. First we show that the matrix $A_k$ is the best rank k approximation to $A$ in the Frobenius norm.

**Theorem 4.3**: For any matrix $B$ of rank at most k, we have

$$\| A-A_k \|_F \leq \| A-B \|_F \ .$$

**Proof**: We use the fact that the subspaces $V_k$ defined in Theorem 4.1 are the best-fit subspaces. Suppose *B* minimizes $\| A-B \|_F^2$ among all rank *k* or less matrices. Let *V* be the space spanned by the rows of *B*. The dimension of *V* is at most *k*. We may assume that each row of *B* is the projection of the corresponding row of *A* onto *V*, since this is the vector in *V* closest to the row of *A*. Thus, $\| A-B \|_F^2$ is at least the sum of squared distances of the rows of *A* to $V_k$ by the optimality of $V_k$. Now, Theorem 4.3 follows from Lemma 4.3. **CLARIFY**

■

Next we tackle 2-norm.

**Lemma 4.4**: $\left\|A-A_k\right\|_2^2=\sigma_{k+1}^2$.

**Proof**: $A-A_k=\sum_{i=k+1}^{r}\sigma_i u_i v_i^T$ . Let $\mathbf{v}$ be the top singular vector of $A-A_k$ . Express $\mathbf{v}$ as a linear combination of $\mathbf{v_1},\mathbf{v_2},\ldots\mathbf{v_r}$: $\mathbf{v}=\sum_{i=1}^{r}\alpha_i\mathbf{v_i}$ . Then,

$$\left|(A-A_k)\mathbf{v}\right|=\left|\sum_{i=k+1}^{r}\sigma_i u_i v_i^T\sum_{j=1}^{r}\alpha_j v_j\right|=\left|\sum_{i=k+1}^{r}\alpha_i\sigma_i u_i v_i^T\mathbf{v_i}\right|=\left|\sum_{i=k+1}^{r}\alpha_i\sigma_i\mathbf{u_i}\right|=\sqrt{\sum_{i=k+1}^{r}\alpha_i^2\sigma_i^2}\ .$$

The v maximizing this last quantity is subject to the constraint that $|\mathbf{v}|^2=\sum_{i=1}^{r}\alpha_i^2=1$ occurs when $\alpha_{k+1}=1$ and the rest of the $a_i$ are 0. This proves the Lemma. **CLARIFY**

■

The next theorem states that $A_k$ is the best rank *k* approximation to *A* in 2-norm.

**Theorem 4.4**:

$$\left\|A-A_k\right\|_2 = \min_{\text{rank}(B)\le k} \left\|A-B\right\|_2$$

**Proof**: By Lemma 4.4, $\left\|A-A_k\right\|_2^2 = \sigma_{k+1}^2$. Now suppose there is some matrix *B* of rank at most *k* such that *B* is a better 2-norm approximation to A than $A_k$, that is $\left\|A-B\right\|_2 < \sigma_{k+1}$. The null space $N(B)$ of *B* (the set of vectors *v* such that $Bv = 0$) has dimension at least *d-k*. By a dimension argument, it follows that there exists a $z \ne 0$ in

$$N(B) \bigcap span\{v_1, v_2, \cdots, v_{k+1}\}.$$

Scale *z* so that $|z| = 1$. We now show that for this vector z which lies in the space of the first k+1 singular vectors, $(A-B)z \ge \sigma_{k+1}(A)$. First

$$\left\|A-B\right\|_2^2 \ge \left|(A-B)z\right|^2.$$

Since $Bz = 0$,

$$\left\|A-B\right\|_2^2 \ge \left|Az\right|^2.$$

Since *z* is in the $span\{v_1, v_2, \cdots, v_{k+1}\}$

$$\left\|A-B\right\|_2^2 \ge \sum_{i=1}^{k+1} \sigma_i^2 \left(v_i^T z\right)^2 \ge \sigma_{k+1}^2$$

contradicting that $\| A-B \|_2 < \sigma_{k+1}$. This proves the Theorem.

■

## 4.4 Algorithm for Computing the Singular Value Decomposition

Computing the singular value decomposition is an important branch of numerical analysis in which there has been many sophisticated developments over a long period of time. Here we present an ``in-principle'' method to establish that the approximate SVD of a matrix *A* may be computed in polynomial time. The reader is referred to numerical analysis texts for more details. The method we present, called the Power Method, is conceptually simple. The word power refers to taking high powers of the matrix $B = AA^T$. We see by direct multiplication that if the SVD of $A$ is $\sum_i \sigma_i \mathbf{u}_i \mathbf{v}_i^T$, then

$$AA^T = \left(\sum_i \sigma_i \mathbf{u_i} \mathbf{v_i}^T\right)\left(\sum_j \sigma_j \mathbf{v_j} \mathbf{u_j}^T\right) = \sum_{i,j} \sigma_i \sigma_j \mathbf{u_i} \mathbf{v_i}^T \mathbf{v_j} \mathbf{u_j}^T = \sum_{i,j} \sigma_i \sigma_j \mathbf{u_i} (\mathbf{v}_i^T \cdot \mathbf{v_j}) \mathbf{u_j}^T = \sum_i \sigma_i^2 \mathbf{u_i} \mathbf{u_i}^T,$$

since $\mathbf{v_i}^T \mathbf{v_j}$ is just the dot product of the two vectors and it is zero unless $i = j$. [Caution :

$\mathbf{v_i v_j}^T$ is a matrix and is not zero even for $i \neq j$.] Using the same kind of calculation,

$$B^k = \sum_i \sigma_i^{2k} \mathbf{u_i u_i}^T .$$

As $k$ increases, $\frac{\sigma_i^{2k}}{\sigma_1^{2k}}$, goes to zero and $B^k$ is approximately equal to

$$\sigma_1^{2k} \mathbf{u_1 u_1}^T$$

provided $\sigma_i(A) < \sigma_1(A)$.

This suggests a way of finding $\sigma_1$ and $\mathbf{u_1}$, by successively powering $B$. But, computing $B^k$ costs $k$ matrix multiplications when done in a straight-forward manner or $O(\log k)$ when done by successive squaring. Instead we compute

$$B^k \mathbf{u}$$

where $\mathbf{u}$ is a random unit length vector. Each increase in $k$ requires a matrix-vector product which takes time proportional to the number of non-zero entries in $B$. Further saving may be achieved by writing

$$B^k \mathbf{u} = AA^T \left( B^{k-1} \mathbf{u} \right).$$

Now the cost is proportional to the number of nonzero entries in $A$. From $B^k \mathbf{u}$, we can recover $\mathbf{u_1}$ since, $B^k \mathbf{u} \approx \sigma_1^{2k} \mathbf{u_1} (\mathbf{u}_1^T \cdot \mathbf{u})$; so it is a scaler multiple of $\mathbf{u_1}$.

If there is a significant gap between the first and second singular values of a matrix, then the above argument should apply and the power method will quickly converge to the first left singular vector. Suppose there is no significant gap. In the extreme case, there may be ties for the top singular value. Then the above argument does not work. The theorem below says that even with ties, the power method converges to some vector in the span of those singular vectors corresponding to the ``nearly highest'' singular values.

**Theorem 4.5**: Suppose $A$ is an $n \times d$ matrix and $\mathbf{u}$ is a random unit length vector. Let $V$ be the space spanned by the left singular vectors of $A$ corresponding to singular values greater than $(1-\varepsilon)\sigma_1$. Let $k$ be any positive integer. With probability at least 9/10, the unit vector

$$\frac{\left(AA^T\right)^k \mathbf{u}}{\left|\left(AA^T\right)^k \mathbf{u}\right|}$$

has a component in $V$ of length at least $1 - 200ne^{-4\epsilon k}$

.

**Proof**: Let

$$A = \sum_{i=1}^{m} \sigma_i \mathbf{u_i v_i}^T$$

be the SVD of $A$. If the rank of $A$ is less than $n$, then complete $\{\mathbf{u_1}, \mathbf{u_2}, \ldots \mathbf{u_m}\}$ into a basis $\{\mathbf{u_1}, \mathbf{u_2}, \ldots \mathbf{u_n}\}$ of $n$-space with vectors other than the singular vectors of $A$. Write $\mathbf{u}$ in the basis of the $\mathbf{u_i}'s$ as

$$\mathbf{u} = \sum_{i=1}^{n} a_i \mathbf{u_i} .$$

Since $(AA^T)^k = \sum_i \sigma_i^{2k} \mathbf{u_i} \mathbf{u_i}^T$, we have $(AA^T)^k \mathbf{u} = \sum_i \sigma_i^{2k} a_i \mathbf{u_i}$ . For a random vector $\mathbf{u}$ picked independently of $A$ the $\mathbf{u_i}$ are fixed vectors and picking u at random is equivalent to picking random $a_i$. Thus, we know from **????** that $|a_1| \geq \frac{1}{10\sqrt{n}}$ with probability at least $\frac{9}{10}$.

Suppose

$$\sigma_r \geq (1-\varepsilon)\sigma_1 > \sigma_{r+1},$$

so $r$ is the last singular value which is at least $(1-\varepsilon)\sigma_1$. Then V is the span of $\{\mathbf{u_1}, \mathbf{u_2}, \ldots \mathbf{u_r}\}$. We have

$$|(AA^T)^k \mathbf{u}|^2 = |\sum_{i=1}^{m} \sigma_i^{2k} a_i \mathbf{u_i}|^2 = \sum_{i=1}^{m} \sigma_i^{4k} a_i^2 .$$

We split the last sum into two parts, one corresponding to the high singular values (up to the $r$ th) and the other corresponding to the low ones. We have (using $1-\epsilon \leq e^{-\epsilon}$ and $\sum_i a_i^2 = |\mathbf{u}| = 1$)

(Component of $(AA^T)^k \mathbf{u}$ perpendicular to V)$^2$ =

$$\sum_{i=r+1}^{m} \sigma_i^{4k} a_i^2 \leq (1-\varepsilon)^{4k} \sigma_1^{4k} \sum_{i=r+1}^{m} a_i^2 \leq e^{-4\varepsilon k} \sigma_1^{4k}$$

(Component of $(AA^T)^k \mathbf{u}$ in V)$^2$ =

$$\left|\sum_{i=1}^{r} \sigma_i^{2k} a_i u_i\right|^2 \geq a_1^2 \sigma_1^{4k} \geq \frac{1}{100n} \sigma_1^{4k} .$$

From these two, the theorem follows.

■

# Exercises

**Exercise 4.1**: Let $A$ be a square $n \times n$ matrix whose rows are orthonormal. Prove that the columns of $A$ are orthonormal.

**Exercise 4.2**: (Best fit functions versus best least squares fit) In many experiments one collects the value of a parameter at various instances of time. Let $y_i$ be the value of the parameter $y$ at time $x_i$. Suppose we wish to construct the best linear approximation to the data in the sense that we wish to minimize the mean square error. Here error is measured vertically rather than perpendicular to the line. Develop formulas for *m* and *b* to minimize the mean square error of the points $\left\{\left(x_i, y_i\right) | 1 \le i \le n\right\}$ to the line $y = mx + b$.

■

**Exercise 4.3**: Given five observed parameters, height, weight, age, income, and blood pressure of *n* people how would one find the best least squares fit subspace of the form

3(height)+4(weight) – 2(age) -1.72(income) + 2.89(BP) = 0?

If there is a good best fit 4-dimensional subspace, then one can think of the points as lying close to a 4-dimensional sheet rather than points lying in 5-dimensions. Why is it better to use the perpendicular distance rather than vertical distance? What is vertical distance anyway?

■

**Exercise 4.4**: What is the best fit line for each of the following set of points?

(a) $\left\{(0,1),(1,0)\right\}$

(b) $\left\{(0,1),(2,0)\right\}$

(c) The rows of : $\begin{pmatrix} 17 & 4 \\ -2 & 26 \\ 11 & 7 \end{pmatrix}$

**Exercise 4.5**: Suppose *A* is a $m \times d$ matrix with block diagonal structure where the blocks $B_1, B_2, \cdots, B_k$ are $\frac{m}{k} \times \frac{d}{k}$ and all entries of each $B_i$ are $a_i$ with $a_1 > a_2 > \cdots > a_k$. Show that *A* has exactly *k* nonzero singular vectors $\mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_k}$ where $v_i$ has the value $(\frac{k}{d})^{1/2}$ in coordinates $\left(i-1\right)\frac{d}{k}+1, \left(i-1\right)\frac{d}{k}+2, \cdots, i\frac{d}{k}$ and 0 elsewhere. In other words, the singular vectors exactly identify the blocks of the diagonal.

**Hint:** By symmetry, the top singular vector's component must be constant in each block**.**

■

**Exercise 4.6**: Prove that the left singular vectors of A are the right singular vectors of $A^T$.

■

**Exercise 4.7**: Interpret the right and left singular vectors for the document term matrix.

**Exercise 4.8**: Verify that the sum of rank one matrices $\sum_{i=1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$ can be written as $UDV^T$, where the $\mathbf{u_i}$ are the columns of $U$ and $\mathbf{v_i}$ are the columns of $V$. To do this, first verify that for any two matrices P and Q, we have

$$PQ = \sum_i \mathbf{p_i q_i}$$

where, $\mathbf{p_i}$ is the $i$ th column of P and $\mathbf{q_i}$ is the $i$ th row of Q.

■

**Exercise 4.9**: Let $A$ be a matrix. Suppose we have an algorithm for finding

$$\mathbf{v_1} = \arg\max_{|\mathbf{v}|=1} |A\mathbf{v}|.$$

Describe an algorithm to find the SVD of $A$.

■

**Exercise 4.10**:

(a) Show that the rank of A is $r$.

(b) Show that $\left|\mathbf{u_1}^T A\right| = \max_{|\mathbf{u}|=1} \left|\mathbf{u^T} A\right| = \sigma_1$.

Hint: Use SVD.

■

**Exercise 4.11:** If $\sigma_1, \sigma_2, \ldots \sigma_r$ are the singular values of $A$ and $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_r$ are the corresponding right singular vectors, show that

(a) $A^T A = \sum_{i=1}^{r} \sigma_i^2 v_i v_i^T$

(b) $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_r$ are eigen-vectors of $A^T A$.

(c) Assuming that the set of eigen-vectors of a matrix is unique, conclude that the set of singular values of any matrix is unique.

See the appendix for the definition of eigen vectors.

■

**Exercise 4.12**: Computational Problem : Compute the SVD of ????. Or write a program to implement the power method and SVD (run the power
method for a fixed number of steps…)…………..More Details

**Exercise 4.13**: Suppose A is a square invertible matrix and the SVD of A is $A = \sum_i \sigma_i u_i v_i^T$.

Prove that the inverse of A is $\sum_i \frac{1}{\sigma_i} v_i u_i^T$.

**Exercise 4.14**: Suppose A is square, but not necessarily invertible and has SVD $A = \sum_{i=1}^{r} \sigma_i u_i v_i^T$.

Let $B = \sum_i \frac{1}{\sigma_i} v_i u_i^T$. Show that $Bv = v$ for all v in the span of the right singular vectors of A. For this reason B is sometimes called the pseudo inverse of A and can play the role of $A^{-1}$ in many applications.

**Exercise 4.15**:

(a) For any matrix A, show that $\sigma_k \leq \frac{\|A\|_F}{\sqrt{k}}$.

(b) Prove that there exists a matrix B of rank at most k such that $\|A-B\|_2 \leq \frac{\|A\|_F}{\sqrt{k}}$.

(c) Can the 2-norm on the left hand side be replaced by Frobenius norm?

**Exercise 4.16**: Suppose a $n \times d$ matrix $A$ is given and you are allowed to preprocess A. Then you will be given a number of d-vectors $x_1, x_2, ...., x_m$ and for each of these you must find the vector $Ax_i$ approximately, in the sense that you must find a vector $u_i$ satisfying $|u_i - Ax_i| \leq \varepsilon \|A\|_F |x_i|$ (here ε >0 is a given error bound.) Describe an algorithm which accomplishes this in time $O\left(\frac{d+n}{\varepsilon^2}\right)$ per $x_i$ (not counting the pre-processing time).

**Exercise 4.17**: Constrained Least Squares Problem using SVD – Golub and van Loan, Chapter 12 :

Use SVD to solve : Given A,b and M, find a vector x with |x|<M minimizing |Ax-b|.

(More explanation or algorithm as in GL chapter 12 should be given…..)