5 Singular Value Decomposition and Matrix Rank Reduction

This note presents the basics of singular value decomposition for matrices, using quadratically constrained quadratic optimization as a tool, and model reduction as the main motivation.

5.1 Singular Value Decomposition

Singular value decomposition can be expressed in many useful forms. The following theorem does this in terms of singular vectors and singular numbers. Recall that $X'$ denotes the Hermitian conjugation of a complex matrix $X$ (transposition followed by complex conjugation). We will also use the Kronecker Delta notation

$$\delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

**Theorem 5.1** Every $m$-by-$n$ complex matrix $M$ can be represented in the form

$$M = \sigma_1 u_1 v_1' + \cdots + \sigma_r u_r v_r' = \sum_{i=1}^{r} \sigma_i u_i v_i', \quad (5.1)$$

where vectors $v_i \in \mathbb{C}^n$, $u_i \in \mathbb{C}^m$, and real numbers $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ are such that $v_i' v_j = u_i' u_j = \delta_{ij}$. Such decomposition takes place if and only if vectors $\{v_i\}$ form a complete orthonormalized sequence of eigenvectors of $M'M$ corresponding to the ordered non-zero eigenvalues of $M'M$, i.e. $r = \text{rank}(M)$, $M'M v_i = \sigma_i^2 v_i$, $v_i' v_j = \delta_{ij}$, and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, in which case $u_i = \sigma_i^{-1} M v_i$. In particular, when $M$ is a real matrix, the vectors $v_i, u_i$ in (5.1) can be chosen to be real as well.

Representation (5.1) is usually referred to as singular value decomposition (SVD) of $M$. Vectors $v_i$ are called right singular vectors of $M$, vectors $u_i$ are called left singular vectors of $M$ while $\sigma_i$ are called singular numbers (or singular values) of $M$. According to Theorem 5.1, singular numbers are uniquely defined by $M$, while singular vectors are not (for example, multiplying both $v_i$ and $u_i$ by the same scalar $c$ such that $|c| = 1$ preserves the identity in (5.1)). In general, $\sigma_k(M)$ denotes the $k$-th singular number of $M$ when $k \leq \text{rank}(M)$, and $\sigma_k(M) = 0$ when $k > \text{rank}(M)$.

A compact matrix form of SVD (5.1) is given by $M = U \Sigma V'$ where

$$U = [u_1 \ u_2 \ \ldots \ u_r], \quad V = [v_1 \ v_2 \ \ldots \ v_r]$$
are the matrices formed by the singular vectors as columns, and
\[
\Sigma = \begin{bmatrix}
\sigma_1 & 0 \\
0 & \sigma_2 \\
& \ddots \\
& & \sigma_r
\end{bmatrix}
\]
is the diagonal matrix formed by the singular numbers. Note that orthonormality of \((u_i)\) and \(v_i\) means that matrices \(U, V\) are orthogonal, in the sense that \(U'U\) and \(V'V\) are identity matrices. Numerical SVD can be performed by using MATLAB’s `svd.m` function, where

\[
[U, S, V] = \text{svd}(M)
\]
produces a full version of the decomposition \(M = USV'\), in which \(U, V\) are unitary (i.e. square orthogonal) matrices, and \(S\) is a diagonal rectangular matrix, possibly with some zero diagonal elements.

A proof of Theorem 5.1 for the case when \(M\) is a real matrix will be presented later, after appropriate machinery is developed in the next section.

5.2 Quadratically Constrained Quadratic Optimization

Let \(b, g : \mathcal{V} \times \mathcal{V} \mapsto \mathbb{R}\) be two symmetric bilinear forms on a real vector space \(\mathcal{V}\). Consider the task of maximizing quadratic form \(g(v, v)\) subject to quadratic constraint \(b(v, v) = 1\).

Using the classical directional differentiation technique yields the following necessary condition of optimality in such optimization.

**Lemma 5.1** If \(a \in \mathcal{V}\) is an argument of maximum of \(g(v, v)\) subject to \(b(v, v) = 1\) then
\[
g(a, w) = g(a, a) b(a, w) \quad \forall \ w \in \mathcal{V}. \quad (5.2)
\]

**Proof.** For every \(w \in \mathcal{V}\) the function \(f_w : \mathbb{R} \mapsto \mathbb{R}\) defined by \(f_w(t) = b(a + tw, a + tw)\) is quadratic (hence continuous), and such that \(f_w(0) = b(a, a) = 1\). Therefore there exists \(r_w > 0\) such that \(f_w(t) > 0\) whenever \(|t| < r_w\). For \(|t| < r_w\) and \(v_t = f_w(t)^{-1}/(a + tw)\) we have

\[
b(v_t, v_t) = \frac{b(a + tw, a + tw)}{f_w(t)} = 1,
\]

\[
g(v_t, v_t) = \frac{g(a + tw, a + tw)}{b(a + tw, a + tw)} = \frac{g(a, a) + 2tg(a, w) + t^2g(w, w)}{1 + 2tb(a, w) + t^2b(w, w)}.
\]

Since \(a\) is assumed to maximize \(g(v, v)\) subject to \(b(v, v) = 1\), the rational expression for \(g(v_t, v_t)\) has a local maximum at \(t = 0\), hence its derivative at zero is zero, i.e.

\[
2g(a, w) - 2b(a, w)g(a, a) = 0.
\]
The result of Lemma 5.1 can be generalized to the case when a sequence of quadratically constrained quadratic optimizations is performed.

**Lemma 5.2** Assume that vectors $v_1, \ldots, v_r \in \mathcal{V}$ are such that $v = v_k$ maximizes $g(v, v)$ subject to $b(v, v) = 1$ and $b(v_i, v) = 0$ for all $i < k$. Then $b(v_i, v_k) = \delta_{ik}$, $g(v_i, v_k) = g(v_k, v_k)\delta_{ik}$, and

$$g(v_k, w) = g(v_k, v_k)b(v_k, w) \quad \forall \ w \in \mathcal{V}$$

(5.3)

for all $i, k \in \{1, \ldots, r\}$.

**Proof.** By induction: the case $r = 1$ is covered by Lemma 5.1. Once proven for $r = d - 1 \geq 1$, to make the induction step to $r = d$, consider the optimization task defining $v_d$. Applying Lemma 5.1 to $\mathcal{V}$ replaced by $\mathcal{V}_d = \{v \in \mathcal{V} : b(v_i, v) = 0 \text{ for } i < d\}$ yields

$$g(v_d, w) = g(v_d, v_d)b(v_d, w) \quad \forall \ w : b(v_i, w) = 0 \ (i < d).$$

Combining this with $b(v_i, v_d) = \delta_{id}$ yields (5.3) for $k = d$.

### 5.3 Proof of Theorem 5.1

Given a real $m$-by-$n$ matrix $M$, define vectors $v_1, \ldots, v_r$ by the iterative process described in the formulation of Lemma 5.2 with $\mathcal{V} = \mathbb{R}^n$, $b(v, w) = v'w$ (i.e. $b(v, v) = |v|^2$) and $g(v, w) = v'M'Mw$ (i.e. $g(v, v) = |Mv|^2$), stopping either at $r = n$ or when $g(v_{r+1}, v_{r+1}) = 0$, whichever comes first. Note that the existence of such maximizers $v_k$ follows from the fact that the constraint $b(v, v) = |v|^2 = 1$ defines a compact subset in $\mathbb{R}^n$, and that the function $v \mapsto g(v, v) = |Mv|^2$ is continuous. According to Lemma 5.2, $v_i$ are orthonormal vectors satisfying condition (5.3), which means that $v'_kM'Mw = |Mv_k|^2v'_kw$ for all $w \in \mathbb{R}^n$, i.e. $M'Mv_k = \sigma_k^2v_k$, where $\sigma_k = |Mv_k| > 0$ are monotonically non-increasing. Hence $u_i = \sigma_i^{-1}Mv_i$ are orthonormal vectors as well.

For $v = v_k$,

$$\left(\sum \sigma_iu_i'v_i\right) v = \sigma_ku_kv_kv_k = \sigma_ku_k = Mv.$$ 

In addition, for $v$ such that $v'_iv = 0$,

$$\left(\sum \sigma_iu_i'v_i\right) v = 0 = Mv.$$ 

Hence the identity (5.1) holds.

The proof establishes an important interpretation of the right singular vectors $v_k$: $v_1$ is the vector for which the length is maximally amplified by the action of $M$; $v_k$ is the vector which, among those orthogonal to $v_i$ for all $i < k$, has its length maximally amplified by the action of $M$.
5.4 Matrix Rank Reduction

Given a matrix $M$ and a positive integer $k$ consider the matrix rank reduction problem as the task of finding a matrix $\hat{M}$ which has rank less than $k$ and minimizes the L2 gain $\|M - \hat{M}\|$ of the error $M - \hat{M}$ of approximating $M$ by $\hat{M}$. When $k > \text{rank}(M)$, the task is trivial, as the optimal approximation is $\hat{M} = M$. The problem becomes important when $k$ is significantly less than $\text{rank}(M)$. Despite the fact that the set of all matrices of rank less than a given number $k$ is not convex, which makes L2 gain optimal matrix rank reduction a non-convex optimization task, the problem has a simple solution in terms of the SVD of $M$.

**Theorem 5.2** Let $M$ be a matrix with SVD (5.1). Let $k$ be a positive integer not larger than $r$. Then the minimal value of $\|M - \hat{M}\|$ over matrices $\hat{M}$ of rank less than $k$ equals $\sigma_k$, and is achieved at

$$\hat{M}_k = \sum_{i<k} \sigma_i u_i v_i' .$$

**Proof.** To show that $\|M - \hat{M}_k\| = \sigma_k$, i.e. $\|(M - \hat{M}_k)v\|^2 \leq \sigma_k^2 |v|^2$, represent $v$ in the form

$$v = x_1 v_1 + \cdots + x_r v_r + e,$$

where $x_i \in \mathbb{R}$ and $v'_i e = 0$ for all $i$ (and hence $Me = 0$). Then

$$\sigma_k^2 |v|^2 = \sigma_k^2 (|e|^2 + \sum_{i=1}^r |x_i|^2) \geq \sum_{i=k}^r \sigma_i^2 |x_i|^2 = \sigma_k^2 (|x_k| u_k + \cdots + \sigma_r x_r u_r|^2 = \| (M - \hat{M}_k)v \|^2 .$$

To show that $\|M - \hat{M}\| \geq \sigma_k$ whenever $\text{rank}(M) < k$, let $\mathcal{V}_k$ be the subspace of all linear combinations

$$v = \sum_{i=1}^k x_k v_k$$

of vectors $v_1, \ldots, v_k$. On one hand,

$$|Mv|^2 = |\sigma_1 x_1 u_1 + \cdots + \sigma_k x_k u_k|^2 = \sigma_1^2 |x_1|^2 + \cdots + \sigma_k^2 |x_k|^2 \geq \sigma_k^2 (|x_1|^2 + \cdots + |x_k|^2) = \sigma_k^2 |v|^2 ,$$

i.e. $|Mv| \geq \sigma_k |v|$ for all $v \in \mathcal{V}_k$. On the other hand, the dimension of $\mathcal{V}_k$ equals $k$, and hence at least one non-zero vector $v_0 \in \mathcal{V}_k$ must belong to the null-space of $M$, as its dimension equals $n - \text{rank}(M) > n - k$. For this vector we have

$$|(M - \hat{M})v| = |Mv| \geq \sigma_k |v| ,$$

which proves that $\|M - \hat{M}\| \geq \sigma_k$. 

\[\Box\]