Chapter 1

Singular Value Decomposition

1.1 Motivation

Assume for a moment that X is a matrix with n rows and p columns and *rank* $r \leq \min(n, p)$. In many circumstances, the problem we face is the following: we need to find a *low rank* approximation of matrix X, that is another matrix M with n rows and p columns, but with rank k, possibly much smaller than r and which is "as close as possible" to X. In order to make this problem a mathematical problem, we need to define the notion(s) of closeness we have in mind. We consider two matrix *norms*: the so-called *operator norm*, and the *Frobenius* or *Hilbert-Schmidt norm*. Thus our problem becomes finding the best rank-k approximation of a given matrix with respect to either the operator norm or the Hilbert-Schmidt norm.

In this lesson, we first define Singular Value Decompositions (SVD) (Section 1.3) and connect the notion with Spectral Decomposition of symmetric matrices (Section 1.4). Then, we show that every matrix enjoys a *Singular Value Decomposition* (SVD) (Section 1.5). In Section 1.6, we introduce the Hilbert-Schmidt norm and inner product. This allows us to endow the linear space of matrices with an inner product. Moreover, from the singular value decomposition of a matrix X, we can derive an orthonormal basis for the space of matrices endowed with the Hilbert-Schmidt inner product (Proposition 1.7). In that adapted orthonormal basis, matrix X has a *sparse* expansion with only r non-zero coefficients. We also point out that the so-called singular values are closely related to the two matrix norms. In Section 1.7, we revisit the proof the existence of SVD (Section 1.5) to establish a variational characterization of positive singular values. In Section 1.8, we connect the SVD of X with the spectral decompositions of the Gramian matrices $X^\top X$ and $XX^\top$.

In Section 1.9, we show that the singular value decomposition is the key ingredient in the determination of the best low rank approximations of a given matrix with respect to the two matrix norms we are working with. It is possible and easy to build a rank-k approximations of X from its singular value decomposition. We can check that. for each rank $k \leq r$, those rank-k *truncated* Singular Value Decompositions provide simultaneously optimal approximations of the matrix X with respect to both norms under rank constraints. This is the content of the Eckart-Young-Mirsky Theorem (Theorem 1.2 in Section 1.9).

The singular value decomposition is the cornerstone of many methods used in Exploratory Data Analysis. For example, *Principal Component Analysis*, *Correspondence Analysis*, *Multiple Correspondence Analysis*, *Canonical Correlation Analysis*, and so on. All those methods can be viewed as wrappers around singular value decomposition. But do not forget that the singular value decomposition is used well beyond Exploratory Data Analysis. It can be used as a *dimension reduction* technique, for example in image analysis. it can also be used in settings like Actuarial Science to unveil the structure and the evolution of *life tables*.

1.2 Convention

1.3 Definition of SVD

Definition 1.1 (Definition of Singular Value Decomposition)**.** Let $X \in \mathcal{M}_{n,p}$ be a real matrix with n rows and p columns.
A tuple of matrices (U, D, V) with $U \in \mathcal{M}_{n,n}, D \in \mathcal{M}_{n,p}, V \in \mathcal{M}_{p,p}$ is a SVD of X iff

$$X = U \times D \times V^\top \qquad \text{with} \begin{cases} U \times U^\top & = \mathrm{Id}_n \\ V \times V^\top & = \mathrm{Id}_p \\ D & \text{non-negative diagonal.} \end{cases}$$

- The columns of U are called *left singular vectors*.
- The columns of V are called *right singular vectors*.
- The diagonal coefficients of D are called the *singular values*.

Note that the definition makes no claim of uniqueness with respect to singular vectors. It is clear from the definition that multiplying the k^{th} column of U and V by -1 leads to another SVD.

Notwithstanding the fact that we have not established the existence of the SVD, we can already state the next proposition.

Proposition 1.1. *If $X \in \mathcal{M}_{n,p}$ has a SVD $X = U \times D \times V^\top$, then the ranks of X and D are equal. If X has rank r, it has exactly r positive singular values.*

We may also define the *thin* SVD.

Definition 1.2. If $X \in \mathcal{M}_{n,p}$ has SVD $X = U \times D \times V^\top$ and rank r, let $U^{[r]}$, $D^{[r]}$, and $V^{[r]}$ be constructed by picking the first r columns of U and V, and the first r rows and columns of D, then

$$X = U^{[r]} \times D^{[r]} \times {V^{[r]}}^\top$$

is a thin SVD of X.

Example 1.1. Matrix X has rank 3 (full column rank).

$$\underbrace{\begin{bmatrix} 3.33 & 0.77 & 2.33 \\ 0.98 & 1.94 & 1.71 \\ -0.74 & -0.60 & 1.06 \\ 1.65 & 0.17 & 2.90 \end{bmatrix}}_{X} = \underbrace{\begin{bmatrix} -0.71 & 0.28 & -0.48 \\ -0.42 & 0.27 & 0.86 \\ -0.02 & -0.75 & 0.16 \\ -0.56 & -0.54 & -0.04 \end{bmatrix}}_{U^{[3]}} \times \underbrace{\begin{bmatrix} 5.66 & 0.00 & 0.00 \\ 0.00 & 1.87 & 0.00 \\ 0.00 & 0.00 & 1.58 \end{bmatrix}}_{D^{[3]}} \times \underbrace{\begin{bmatrix} -0.65 & -0.25 & -0.71 \\ 0.48 & 0.59 & -0.65 \\ -0.59 & 0.76 & 0.27 \end{bmatrix}}_{{V^{[3]}}^\top}$$

1.4 SVD and spectral decomposition of symmetric matrices

Symmetric matrices are known to be diagonalizable in an orthonormal basis, see Bhatia (1997), Horn and Johnson (1990). In words, if $W \in \mathcal{M}_{n,n}$ is symmetric, there exists an orthogonal matrix $O \in \mathcal{M}_{n,n}$ ($O \times O^\top = O^\top \times O = \mathrm{Id}_n$) and a diagonal matrix $\Lambda \in \mathcal{M}_{n,n}$ such that

$$W = O \times \Lambda \times O^\top$$

or equivalently $O^\top \times W \times O = \Lambda$.

The diagonal coefficients of Λ are the *eigenvalues* of W, the columns of O are the *eigenvectors* of W. The non-decreasing rearrangement of the diagonal coefficients of Λ (the sorted eigenvalues) are unique. The factorization $O \times \Lambda \times O^\top$ is called the *spectral decomposition* of W.

⚠ A symmetric matrix is not necessarily semi-definite positive. So, eigenvalues are not necessarily non-negative. Indeed a symmetric matrix is definite (resp. semi-definite) positive iff all its eigenvalues are positive (resp. non-negative).

Exercise 1.1. Check the preceding assertion on $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

The singular value decomposition and the spectral decomposition are intimately related. This can be stated in several ways. The SVD of a matrix X is related to the spectral decomposition of a derived matrix.

Proposition 1.2. *Let $X \in \mathcal{M}_{n,p}$ be a real matrix. Let the square symmetric $A \in \mathcal{M}_{n+p,n+p}$ be defined as*

$$A = \begin{bmatrix} 0 & \vdots & X \\ \cdots & & \cdots \\ X^\top & \vdots & 0 \end{bmatrix}$$

Then $rank(A) = 2rank(X)$. Assume that s_k is a positive singular value of X. Then s_k and $-s_k$ are eigenvalues of X, and all non-zero eigenvalues of A are obtained that way. Moreover, let u_k (resp. v_k) be the k^{th} column of U (resp. V), vectors

$$\begin{bmatrix} u_k \\ \cdots \\ v_k \end{bmatrix} \quad \textit{and} \quad \begin{bmatrix} u_k \\ \cdots \\ -v_k \end{bmatrix}$$

are eigenvectors of A with associated eigenvalues s_k and $-s_k$.

Remark 1.1. Proposition 1.2 can be used to reduce SVD computation to spectral decomposition (and vice versa).

The proof of Proposition 1.2 is left as an exercise.

1.5 Existence and first properties of Singular Value Decomposition

The next theorem Theorem 1.1 tells us that any real-valued matrix has a Singular Value Decomposition.

Theorem 1.1 (Existence of Singular Value Decomposition)**.** *Let $X \in \mathcal{M}_{n,p}$ be a real matrix with n rows and p columns. X has a SVD complying with Definition 1.1.*

We repeat the warning concerning absence of uniqueness.

Remark 1.2. The SVD decomposition is not unique.

$M := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is orthogonal but it is not a rotation

$$\begin{aligned} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \end{aligned}$$

To establish Theorem 1.1, we will use the matrix norm.

Definition 1.3 (Operator norm)**.** For $M \in \mathcal{M}_{n,p}$

$$\|M\|_{\text{op}} := \sup_{u\in\mathbb{R}^n \|u\|\leq 1} \sup_{v\in\mathbb{R}^p \|v\|\leq 1} u^\top M v = \sup_{v\in\mathbb{R}^p \|v\|\leq 1} \|Mv\|$$

Checking that $\|\cdot\|_{\text{op}}$ is indeed a norm is left as an exercise.

Remark 1.3. The above defined norm is just *one* operator norm among many possible ones. Indeed it is the operator norm for linear functions from $\mathbb{R}^p$ to $\mathbb{R}^n$ both equipped with Euclidean norms. If we had equipped $\mathbb{R}^p$ and $\mathbb{R}^n$ with different norms (say ℓ_∞^p, ℓ_1^n), we would have defined another operator norm, namely:

$$\sup_{\|x\|_1\leq 1, \|y\|_\infty \leq 1} x^\top M y = \max_{i\leq n} \sum_{j=1}^{p} |M_{i,j}|$$

Operator norms satisfy a sub-multiplicative condition.

Proposition 1.3.

$$\|A \times B\|_{op} \leq \|A\|_{op} \times \|B\|_{op}$$

Proof. Proposition 1.3 follows if we rely on the variational characterization of the operator norm.

Let $A \in \mathcal{M}_{n,p}$ and $B \in \mathcal{M}_{p,\ell}$

$$\begin{aligned}
\|A \times B\|_{\text{op}} &= \sup_{x\in\mathbb{R}^\ell \|x\|\leq 1} \|A \times Bx\|_2 \\
&\leq \sup_{x\in\mathbb{R}^\ell \|x\|\leq 1} \sup_{y\in\mathbb{R}^p \|y\|\leq 1} \|A \times y\|_2 \|A \times Bx\|_2 \\
&\leq \|A\|_{\text{op}} \times \|B\|_{\text{op}}
\end{aligned}$$

□

Exercise 1.2.

- Check that the operator norm of an orthogonal matrix is 1.
- Check that if $UDV^\top$ is a SVD of X, then \$$\|X\|_{\text{op}} = \|D\|_{\text{op}}$.
- Check that if $UDV^\top$ is a SVD of X, $\|X\|_{\text{op}}$ is equal to the largest coefficient in D.

The proof of Theorem 1.1 proceeds by induction on $\max(n,p)$ the largest dimension of the matrix.

Proof. If $\max(n,p) = 1$, nothing to do. The matrix is (almost) its own SVD.

Assume now that the SVD exists for any matrix in $\mathcal{M}_{n,p}$ with $\max(n,p) \leq m\}$

Let $Z \in \mathcal{M}_{n,p}$ with $\max(n,p) = m+1$ and rank $r \leq \min(n,p)$.
If $\|Z\|_{\text{op}} = 0$, $\text{rank}(Z) = 0$, $Z = 0$ and there is nothing to prove.
If $\|Z\|_{\text{op}} > 0$, by compactness of unit balls $B_2^n(1)$ (resp. $B_2^p(1)$) in $\mathbb{R}^n$ (resp. $\mathbb{R}^p$), there exist unit vectors $\hat{u} \in \mathbb{R}^n, \hat{v} \in \mathbb{R}^p$ such that:

$$\|Z\|_{\text{op}} = \hat{u}^\top Z \hat{v} = \sup_{u \in B_2^n(1), v \in B_2^p(1)} u^\top Z v \,.$$

Let

- $s_1 = \|Z\|_{\text{op}}$
- A, B be matrices such that

$$\begin{bmatrix} \hat{u} & \vdots & A \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} \hat{v} & \vdots & B \end{bmatrix}$$

  are *orthognal* matrices with dimensions $n \times n$ and $p \times p$ (the existence of such matrices is warranted by the Gram-Schmidt orthogonalization procedure).

Decomposing matrices into blocks, we obtain:

$$\begin{bmatrix} \hat{u}^\top \\ \cdots \\ A^\top \end{bmatrix} \times Z \times \begin{bmatrix} \hat{v} & \vdots & B \end{bmatrix} = \begin{bmatrix} s_1 & \vdots & w^\top \\ \cdots & & \cdots \\ 0 & \vdots & A^\top Z B \end{bmatrix} .$$

Let Y denote the matrix on the right-hand side.
Observe that $A^\top Z \hat{v} = 0$ since:

- $Z\hat{v}$ is colinear with $\hat{u}$ and
- the columns of A are orthogonal to $\hat{u}$.

Now, we have to check that $w = 0$.
Because multiplying by orthogonal matrices does not change operator norm, we have

$$\|Y\|_{\text{op}} = \|Z\|_{\text{op}} = s_1 \,.$$

$$\|Y\|_{\text{op}} \geq \frac{\left\| Y \begin{bmatrix} s_1 \\ \cdots \\ w \end{bmatrix} \right\|}{\left\| \begin{bmatrix} s_1 \\ \cdots \\ w \end{bmatrix} \right\|} \geq \frac{s_1^2 + w^\top w}{\sqrt{s_1^2 + w^\top w}} = \sqrt{s_1^2 + w^\top w}$$

In order to have $s_1 \geq \|Y\|_{\text{op}}$, we need to have $w = 0$.
Matrix $A^\top Z B$ (with rank $r-1$) has $n-1$ rows and $p-1$ columns. Thus, $A^\top Z B$ satisfy the induction hypothesis ($\max(n-1, p-1) \leq m$).
Note also that $\|A^\top Z B\|_{\text{op}} \leq \|Y\|_{\text{op}}$.

This entails the existence of orthogonal matrices U' and V' such that ${U'}^\top A^\top Z B V'$ is equal to a diagonal non-negative matrix D' with $r-1$ non null-coefficients.

This entails:

$$\begin{bmatrix} 1 & \vdots & 0 \\ \cdots & & \cdots \\ 0 & \vdots & {U'}^\top \end{bmatrix} \times \begin{bmatrix} \hat{u}^\top \\ \cdots \\ A^\top \end{bmatrix} \times Z \times \begin{bmatrix} \hat{v} & \vdots & B \end{bmatrix} \times \begin{bmatrix} 1 & \vdots & 0 \\ \cdots & & \cdots \\ 0 & \vdots & V' \end{bmatrix} = \begin{bmatrix} s_1 & \vdots & 0 \\ \cdots & & \cdots \\ 0 & \vdots & D' \end{bmatrix}$$

Matrices

$$\begin{bmatrix} 1 & \vdots & 0 \\ \cdots & & \cdots \\ 0 & \vdots & {U'}^\top \end{bmatrix} \times \begin{bmatrix} \hat{u}^\top \\ \cdots \\ A^\top \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} \hat{v} & \vdots & B \end{bmatrix} \times \begin{bmatrix} 1 & \vdots & 0 \\ \cdots & & \cdots \\ 0 & \vdots & V' \end{bmatrix}$$

are orthogonal with dimensions $n \times n$ and $p \times p$.

$$Z = \underbrace{\begin{bmatrix} \hat{u} & \vdots & A \end{bmatrix} \times \begin{bmatrix} 1 & \vdots & 0 \\ \cdots & & \cdots \\ 0 & \vdots & U' \end{bmatrix}}_{\text{orthogonal} \in \mathcal{M}_{n,n}} \times \underbrace{\begin{bmatrix} s_1 & \vdots & 0 \\ \cdots & & \cdots \\ 0 & \vdots & D' \end{bmatrix}}_{\text{diagonal}} \times \underbrace{\begin{bmatrix} 1 & \vdots & 0 \\ \cdots & & \cdots \\ 0 & \vdots & {V'}^\top \end{bmatrix} \times \begin{bmatrix} \hat{v}^\top \\ \cdots \\ B^\top \end{bmatrix}}_{\text{orthogonal} \in \mathcal{M}_{p,p}}$$

□

From the proof of Theorem 1.1, we pocket the next proposition.

Proposition 1.4 (Singular values and operator norm)**.** *For any matrix $X \in \mathcal{M}_{n,p}$, the largest singular value is equal to the operator norm.*

Exercise 1.3. Prove directly (without resorting to the SVD Theorem) that if $X \in \mathcal{M}_{n,p}$, there exists a unit vector $u \in \mathbb{R}^p$ (endowed with Euclidean norm) such that

$$\|X\|_{\text{op}}^2 u = X^\top X u,$$

that is $\|X\|_{\text{op}}^2$ is an eigenvalue of $X^\top X$.

See Theorem 2.3.1 in Golub and Van Loan (1996).

1.6 More matrix norms : Hilbert-Schmidt norms

(Golub and Van Loan 1996, sec. 2.3)

Definition 1.4 (Hilbert-Schmidt-Frobenius norm)**.** Let $A \in \mathcal{M}_{n,p}$, then the Hilbert-Schmidt-Frobenius norm of A is defined by:

$$\|A\|_{\text{HS}}^2 = \sum_{i \leq n, j \leq p} A_{i,j}^2$$

Proposition 1.5. *The next statement*

$$\langle A, B \rangle_{\text{HS}} \overset{def}{=} \text{Trace}\left(A \times B^\top\right)$$

defines a inner product *over* $\mathcal{M}_{n,p}$*. This inner product is called the Hilbert-Schmidt-Frobenius inner product.*
The Hilbert-Schmidt-Frobenius norm is related to the Hilbert-Schmidt-Frobenius inner product by:

$$\|A\|_{\text{HS}} = \sup_{B : \|B\|_{\text{HS}} \leq 1} \langle A, B \rangle_{\text{HS}}$$

Checking Proposition 1.5 amounts to verify that $\text{Trace}\left(A \times B^\top\right)$ is indeed a inner product, that $\text{Trace}\left(A \times A^\top\right) = \|A\|_{\text{HS}}^2$ and that the supremum in $\sup_{B : \|B\|_{\text{HS}} \leq 1} \langle A, B \rangle_{\text{HS}}$ is achieved by picking $B = \frac{1}{\|A\|_{\text{HS}}} A$.

Proof. To check that $\text{Trace}\left(A \times B^\top\right)$ is an inner product, we first have to check linearity with respect to both arguments.

This follows immediately from the linearity of the trace operator and from the fact that matrix multiplication is distributive with respect to matrix addition and compatible with scalar multiplication.

Symmetry follows from the fact that transposition does not change the trace.

Finally, positive definiteness follows from the fact that

$$\text{Trace}\left(A \times A^\top\right) = \sum_{i,j} A_{i,j}^2 \geq 0$$

with equality iff $A = 0$.

The fact that $\text{Trace}\left(A \times A^\top\right) = \|A\|_{\text{HS}}^2$ for those from the more general identity:

$$\text{Trace}\left(A \times B^\top\right) = \sum_{i,j} A_{i,j} B_{i,j}\,.$$

Finally, by Cauchy-Schwarz inequality for inner products:

$$\langle A, B \rangle_{\text{HS}} \leq \|A\|_{\text{HS}} \times \|B\|_{HS}$$

with equality iff A and B are colinear. □

The Hilbert-Schmidt norm is also connected with the singular values.

Proposition 1.6 (Singular values and Hilbert-Schmidt norm)**.** *Let $X \in \mathcal{M}_{n,p}$ have singular values $s_1 \geq s_2 \geq \ldots \geq s_{n\wedge p}$ then*

$$\|X\|_{\text{HS}}^2 = \sum_{i=1}^{n\wedge p} s_i^2$$

In words, the Hilbert-Schmidt norm is the Euclidean norm of the vector built from the singular values.

Proof. The Hilbert-Schmidt norm is invariant by mutiplication by orthogonal matrices. Hence, if $U \times D \times V^\top$ is a SVD of X,

$$\|X\|_{\text{HS}} = \|D\|_{\text{HS}}$$

The proposition follows by observing that $\|D\|_{\text{HS}}^2$ is the sum of the squared singular values of X. □

Proposition 1.7 (A basis for $\mathcal{M}_{n,p}$)**.** *Let $u_1, \ldots, u_n$ and $v_1, \ldots, v_p$ form two orthonormal bases of $\mathbb{R}^n$ and $\mathbb{R}^p$ (endowed with Euclidean norm), then $\left(u_i \times v_j^\top\right)_{i\leq n, j\leq p}$ forms an orthonormal basis of $\mathcal{M}_{n,p}$ endowed with the Hilbert-Schmidt inner product.*

Proof.

$$\begin{aligned}\langle u_i v_j^\top, u_k v_\ell^\top \rangle_{\text{HS}} &= \text{Trace}\left(u_i v_j^\top v_\ell u_k^\top\right) \\ &= \langle v_j, v_\ell \rangle \,\text{Trace}\left(u_i u_k^\top\right) \\ &= \langle v_j, v_\ell \rangle \times \langle u_i, u_k \rangle\end{aligned}$$

□

Corollary 1.1. *The left and right singular vectors in the full SVD of an $n \times p$ matrix form an orthonormal basis of $\mathcal{M}_{n,p}$ endowed with the Hilbert-Schmidt inner product.*

Matrix X has a simple expansion in this basis.

$$X = \sum_{k\leq r} s_k u_k v_k^\top$$

1.7 Variational characterizations of singular values

Proposition 1.4 provides a variational characterization of the first singular value. Proposition 1.8 asserts that all positive singular values have a variational characterization. This is also pocketed from the proof of Theorem 1.1.

Proposition 1.8. *Let $X \in \mathcal{M}_{n,p}$ with rank r and thin SVD: $X = U \times D \times V^\top$.*
Let $u_1, \dots, u_r$ (resp. $v_1, \dots, v_r$) denote the column vectors of U (resp. V).
Let $s_1 \geq s_2 \geq \dots \geq s_r$ denote the positive singular values of X.
Then for $k \leq r$,

$$s_k = \sup_{\substack{u \perp span(u_1, \dots u_{k-1}) \\ v \perp span(v_1, \dots v_{k-1})}} \frac{u^\top X v}{\|u\| \|v\|}$$

1.8 Further connections between SVD and spectral decomposition

Another connection between SVD and spectral decomposition relies on the fact that the Gramian matrices $A = X^\top X$ and $M = XX^\top$ are symmetric positive semi-definite (SDP), that is, for every $u \in \mathbb{R}^p, v \in \mathbb{R}^n$,

$$u^\top A u = (Xu)^\top (Xu) = \langle Xu, Xu \rangle \geq 0 \qquad v^\top M v = (X^\top v)^\top (X^\top v) = \langle X^\top v, X^\top v \rangle \geq 0$$

The SVD of X and the spectral decomposition of $X^\top X$ and $XX^\top$ are tightly related.
Let $X = U \times D \times V^\top$, then

$$A = X^\top \times X = V \times D \times U^\top \times U \times D \times V^\top = V \times D^2 \times V^\top$$

so $V \times D^2 \times V^\top$ is the spectral decomposition of the SDP matrix $X^\top X$. In words, the eigenvectors of $X^\top X$ are the right singular vectors of X.

The non-zero eigenvalues of $X^\top X$ are the squared non-zero singular values of X.

The eigenvectors of $XX^\top$ are the left singular vectors of X. And the non-zero eigenvalues of $XX^\top$ are the squared non-zero singular values of X.

1.9 Eckart-Young-Mirsky Theorems

The SVD theorem is a *factorization* theorem: it allows to break up a matrix into simpler objects. There are many other factorizations.
But, there is something special about SVD.
SVD delivers a sequence of *best approximations with given rank*.
The low-rank approximations are simultaneously optimal with respect to a large collection of norms (including operator and Hilbert-Schmidt norms)
This is the content of the *Eckart-Young-Mirsky Theorem*

Definition 1.5 (Truncated SVD)**.** If $X \in \mathcal{M}_{n,p}$ has SVD $X = U \times D \times V^\top$ and rank r, let $k \leq r$ and $U^{[k]}$, $D^{[k]}$, and $V^{[k]}$ be constructed by picking the first k columns of U and V, and the first k rows and columns of D, then

$$U^{[k]} \times D^{[k]} \times {V^{[k]}}^\top$$

is a SVD of a matrix from $\mathcal{M}_{n,p}$ with rank k.
$U^{[k]} \times D^{[k]} \times {V^{[k]}}^\top$ is called a rank-k truncated SVD of X.

Note that the truncated SVD have sparse expansions in the orthonormal basis defined by the singular vectors.

$$U^{[k]} \times D^{[k]} \times {V^{[k]}}^\top = \sum_{j \leq k} s_j u_j v_j^\top$$

Theorem 1.2 (Optimality of Truncated SVD)**.** *The* truncated SVD *leads to the best rank-k approximation with respect to both the Hilbert-Schmidt norm and the operator norm.*

Proof. Best approximation in operator norm

Assume without loss of generality that $n \geq p$

Let $Z^{[k]}$ be the rank-k truncated SVD of Z. The SVD of $Z - Z^{[k]}$ is (up to a permutation of the columns of U, V, and the singular values):

$$U \times \text{diag} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ s_{k+1} \\ \vdots \\ s_p \end{bmatrix} \times V^\top$$

The operator norm of the difference $Z - Z^{[k]}$ is equal to s_{k+1}

Now, let W be of rank k. The null space of W, $\ker(W)$ has dimension $p - k$. Its intersection with the linear span of the first $k + 1$ columns of V has dimension at least 1.

$$\dim\left(\ker(W) \cap \text{span}(v_1, \ldots, v_{k+1})\right) \geq 1$$

Let y be a unit vector in this intersection.

$$y = \sum_{i=1}^{k+1} \langle y, v_i \rangle v_i$$

and $\sim_{i=1}^{k+1} \langle y, v_i \rangle^2 = 1$.

By definition o y, $Wy = 0$ and

$$Zy = \sum_{i=1}^{k+1} s_i \langle y, v_i \rangle u_i$$

$$\|Zy\|^2 = \sum_{i=1}^{k+1} s_i^2 \langle v_i, y \rangle^2 \geq s_{k+1}^2$$

which proves that $\|Z - W\|_{\text{op}} \geq \|(Z - W)y\| \geq s_{k+1}$ □

Proposition 1.9. *Let $A, A', A'' \in \mathcal{M}_{n,p}$ satisfying $A = A' + A''$, then for $i, j \geq 1$*

$$s_i(A') + s_j(A'') \geq s_{i+j-1}(A)$$

with the convention $s_k(X) = 0$ for $k \geq \text{rank}(X)$.

This is called Weyl's additive inequality for singular values.

Proof. For any $X \in \mathcal{M}_{n,p}$, let $X^{(k)}$ be the rank-k truncated SVD of X.

$$\begin{aligned} s_i(A') + s_j(A'') &= \left\|A' - A'^{[i-1]}\right\|_{\text{op}} + \left\|A'' - A''^{[j-1]}\right\|_{\text{op}} \\ &\geq \left\|A - \underbrace{A'^{[i-1]} + A''^{[j-1]}}_{\text{has rank} \leq i+j-2}\right\|_{\text{op}} \\ &\geq \left\|A - A^{[i+j-2]}\right\|_{\text{op}} \\ &= s_{i+j-1}(A)\,. \end{aligned}$$

where the last inequality follows from the optimality of the rank-truncated SVD for approximation with respect the operator norm. □

Remark. We used the optimality of the truncated SVD for approximating with respect to operator norm to prove Proposition 1.9. It is interesting to note that if we assume Proposition 1.9, the optimality of the truncated SVD for approximating with respect to operator norm follows easily.

Indeed, let $X \in \mathcal{M}_{n,p}$. Let $B \in \mathcal{M}_{n,p}$ be of rank $k \leq \text{rank}(X)$, then $s_{k+1}(B) = 0$. Then, by Proposition 1.9:

$$s_1(X - B) + s_{k+1}(B) \geq s_{k+1}(X)$$

which translates into

$$\|X - B\|_{\text{op}} \geq s_{k+1}(X) = \left\|X - X^{[k]}\right\|_{\text{op}}$$

We may now proceed to the proof tof the second part of Theorem 1.2

Proof. Best approximation in Hilbert-Schmidt norm
Let $X \in \mathcal{M}_{n,p}$. Let $B \in \mathcal{M}_{n,p}$ be of rank $k \leq \text{rank}(X)$, then $s_{k+1}(B) = 0$.

$$\begin{aligned}\|X - B\|_{\text{HS}}^2 &= \sum_{i=1}^{\min(n,p)} \left(s_i(X - B)\right)^2 \\ &= \sum_{i=1}^{\min(n,p)} \left(s_i(X - B) + s_{k+1}(B)\right)^2 \\ &\geq \sum_{i=1}^{\min(n,p)} \left(s_{i+k}(X)\right)^2\end{aligned}$$

where the inequality follows from Proposition 1.9.
Now,

$$\begin{aligned}\|X - B\|_{\text{HS}}^2 &\geq \sum_{i=k+1}^{\min(n,p)} \left(s_i(X)\right)^2 \\ &= \left\|X - X^{[k]}\right\|_{\text{HS}}^2\end{aligned}$$

□ □

1.10 SVD and QR factorizations

Definition 1.6. Let $Z \in \mathcal{M}_{n,p}$. The QR-factorization of Z is a couple of matrices (Q, R) such that

$$Z = Q \times R$$

and

- $Q \in \mathcal{M}_{n,p}$ has pairwise orthogonal columns with unit norm ($Q^T \times Q = \text{Id}_p$) and
- R is upper-triangular with non-negative diagonal coefficients.
- If Z has full column rank, then R is invertible.

If Z has thin SVD $U \times D \times V^\top$ and QR factorization $Q \times R$ then

$$\begin{aligned}Z^\top Z &= R^\top R \\ &= VD^2V^\top\end{aligned}$$

so, R is a Cholesky factor of the Gramian matrix $Z^\top Z$ while $VD^2V^\top$ is a spectral decomposition of $Z^\top Z$.

If Z has full column rank ($Z^T Z$ is invertible), then

$$\begin{aligned} Z(Z^\top Z)^{-1} Z^\top &= QQ^\top \\ &= UU^\top \end{aligned}$$

The column space of U and the column space of Q are identical, and $QQ^\top, UU^\top$ both define the orthogonal projection on the column space of Z.

1.11 SVD and pseudo-inverse

What if Z does not have full column rank? Defining the column space of Z still makes sense. And making sense of the Ordinary Least Squares problem for rank-deficient designs is also an important problem (in High Dimensional Statistics).

The notion of pseudo-inverse of a matrix addresses these issues.

Definition 1.7. Let $X \in \mathcal{M}_{n,p}$ have positive rank. A *pseudo-inverse* of X is a matrix $Y \in \mathcal{M}_{p,n}$ such that:

a. $X \times Y \times X = X$
b. $Y \times X \times Y = Y$
c. $X \times Y$ is symmetric
d. $Y \times X$ is symmetric

Notwithstanding existence, we first check:

Proposition 1.10. *The pseudo-inverse of a matrix is unique.*

This pseudo-inverse is called the Moore-Penrose pseudo-inverse.

Proof. Assume that Z and Y satisfy the four conditions in Definition A.1.

$$
\begin{aligned}
Z &= ZXZ && \text{b)}\\
&= Z(XZ)^\top && \text{d)}\\
&= ZZ^\top X^\top\\
&= ZZ^\top (XYX)^\top && \text{a)}\\
&= Z(XZ)^\top (XY)^\top\\
&= ZXZXY && \text{c) and d)}\\
&= ZXY && \text{a)}\\
&= (ZX)^\top YXY && \text{c) and b)}\\
&= X^\top Z^\top (YX)^\top Y && \text{c)}\\
&= (XYX)^\top Z^\top X^\top Y^\top Y && \text{a)}\\
&= (YX)^\top X^\top Z^\top X^\top Y^\top Y\\
&= (YX)^\top X^\top Y^\top Y && \text{a)}\\
&= (XYX)^\top Y^\top Y\\
&= X^\top Y^\top Y && \text{a)}\\
&= (YX)^\top Y\\
&= YXY && \text{c)}\\
&= Y && \text{b)}
\end{aligned}
$$

□

What about existence? When X has full column rank, the pseudo-inverse of X is $(X^\top X)^{-1}X^\top$ (check this). When X does not have full column rank, the pseudo-inverse can be obtained from the SVD.

Proposition 1.11. *Let X have positive rank and $X = U \times D \times V^\top$ be a thin SVD. Then*

$$V \times D^{-1} \times U^\top$$

is the *(Moore-Penrose) pseudo-inverse of X.*

Proof. Remember that in a thin SVD of a positive rank X, the square matrix D is invertible.

Properties 3. and 4. are thus readily checked: $X \times Y = UU^\top$ and $Y \times X = V \times V^\top$. Note that both matrices are orthogonal projectors on the SEVs spanned by the columns of U and V respectively.

To check 1., $X \times Y \times X = U \times D \times V^\top \times V \times V^\top = X \times Y \times X = U \times D \times \mathrm{Id}_r \times V^\top = U \times D \times V^\top = X$.

Property 2. follows in a similar way. □

In words, the pseudo-inverse is obtained from the SVD by exchanging the left and right singular vectors, and by taking inverses of the positive singular values.

1.12 References

Horn and Johnson (1990) and Bhatia (1997) are classics on Matrix Analysis. Their content goes far beyond Singular Value Decomposition. Golub and Van Loan (1996) is a classic on matrix computations. Hsing and Eubank (2015) presents functional extensions of PCA.