1 28th of September 2018 — F. Poloni

1.1 Singular value decomposition (SVD)

We are left with the task of reaching a (sort of) "eigenvalue decomposition" when the target matix is not symmetric.
There are two way to generalize the eigenvalue decomposition to a nonsymmetric matrix A (with something that always exists):

Definition 1.1 (Schur decomposition). *Let $A \in M(n, \mathbb{R})$, $\exists\ U \in M(n, \mathbb{R})$ orthogonal matrix and $T \in M(n, \mathbb{R})$ triangular matrix such that $A = UTU^T$ and this is called* ***Schur decomposition****.*

What is really important for us is the **Singular value decomposition**, every square matrix A can be written with **SVD** form.

Every square matrix A can be written as SVD.

Definition 1.2 (Singular value decomposition). *Let $A \in M(n, \mathbb{R})$, $\exists\ U, V \in M(n, \mathbb{R})$ orthogonal matrices (V not necessary equal to U) and $\Sigma \in Diag(n, \mathbb{R})$ such that $A = U\Sigma V^T$ and this is called* ***Singular Value Decomposition****.*

$$A = \begin{pmatrix} u_1 & u_2 & \cdots & u_n \end{pmatrix} \cdot \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{pmatrix} \cdot \begin{pmatrix} {v_1}^T \\ {v_2}^T \\ \vdots \\ {v_n}^T \end{pmatrix} = \sum_{i=1}^{n} u_i \sigma_i {v_i}^T =$$

$$= u_1 \sigma_1 v_1^T + u_2 \sigma_2 v_2^T + \cdots + u_n \sigma_n v_n^T$$

Where σ_i are called ***singular values*** *and they are sorted such that:*

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

General fact on singular values:

- *Singular values $\neq$ eigenvalues*
- *They are always positive and usually more spread apart than the eigenvalues.*

$$\sigma_1 \geq |\lambda_1| \ \ and \ \ |\lambda_m| \geq \sigma_m$$

λ_i is larger than the largest eigenvalue of a matrix A and λ_m is smaller than the smallest eigenvalue of a matrix A.

The SVD can be defined also for a rectangular matrix A:

Definition 1.3 (Rectangular matrices and SVD)**.** *Let $A \in M(m,n,\mathbb{R})$, there exist $U \in M(m,\mathbb{R})$ orthogonal, $V \in M(n,\mathbb{R})$ orthogonal and $\Sigma(m,n,\mathbb{R})$ diagonal in the sense that $\sum_{ij} = 0$ with $i \neq j$ (padded with zeros). Matrix A has a* ***SVD factorization****, where Σ has the following shape:*

- *case $m < n$ (e.g $m = 3, n = 5$)*

$$\begin{pmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 \end{pmatrix}$$

- *case $m > n$ (e.g $m = 5, n = 3$)*

$$\begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Definition 1.4 (Thin SVD)**.** *Let $A \in M(m,n,\mathbb{R})$, has a* ***thin SVD factorization****: we may restrict to compute only the first min(m, n) vectors that appear in this sum: thin SVD.*

$$A = \sum_{i=1}^{\min(m,n)} u_i \sigma_i {v_i}^T = u_1\sigma_1 v_1^T + u_2\sigma_2 v_2^T + \cdots + u_{min(m,n)}\sigma_{min(m,n)} v_{min(m,n)}^T$$

Something on Matlab ...

In Matlab the SVD decomposition is obtained through the command `svd(A)`, which return value is made of the three matrices U, Σ, V.
As an example, `[U, S, V] = svd(A)`. Notice that, if `svd(A)` is assigned to one variable, then such variable is an array of singular values.
The thin SVD can be compute with: `[U, S, V] = svd(A, 0)`

Computational costs
We are not going into details of algorithms for computing SVD, but we would like to add a consideration about the computational complexity of such an algorithm.

- `[U, S, V] = svd(A, 0)` (thin) costs $O(mn^2)$ ops for $A \in \mathbb{R}^{m\times n}$ or $A \in \mathbb{R}^{n\times m}$ with $m \geq n$
- `[U, S, V] = svd(A)` (non-thin) more expensive, because it has to store the large $m \times m$ factor. (But there are some tricks to store orthogonal matrices compactly, more about it later).

1.1.1 Properties of SVD

The SVD reveals rank, image, and kernel of a matrix.

Definition 1.5 (Rank)**.** *Let $A \in M(n, \mathbb{R})$ we call the **rank** of A the number of non-zero singular values.*

*Equivalently, the **rank** is the size of the column space.*

Property 1.1. *A matrix $A \in M(n, \mathbb{R})$ has rank r iff all its eigenvalues starting from the $r+1$-th are 0, formally iff $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_n = 0$.*

Thanks to Property 1.1, we can somehow talk about an "even thinner" SVD, where all the 0s in the bottom right part of the matrix Σ, cancel out the latter columns of U and the latter rows of V (aka columns of V^T). A pictorial representation of the shape of Σ can be found below.

$$\Sigma \quad = \quad \left(\begin{array}{cc} \begin{array}{cccc|} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \\ \hline \end{array} & \\ & \begin{array}{ccc} 0 & & \\ & \ddots & \\ & & 0 \end{array} \\ \hline \multicolumn{2}{c}{\mathbf{0}} \end{array}\right)$$

This factorization represents A as $\sum\limits_{i=1}^{r} U_i \sigma_i V_i$.

An attentive reader may notice that $Ax = \sum\limits_{i=1}^{r} U_i \sigma_i V_i x$, where the last three terms are dimensionally a scalar. It goes without saying that the image of A is the span of $U_1, U_2, \ldots U_r$, hence $rk(A) = r$.

Moreover, $\ker(A) = span(V_{r+1}, V_{r+2}, \ldots V_n)$, since V is orthgonal (proof: plugging in $x = V_j$, where $j > r$).

Definition 1.6 (Matrix norm)**.** *Let $A \in M(m, n, \mathbb{R})$. We define the **matrix norm** of A as*

$$\|A\| := \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \max_{z=1} \|Az\|$$

*Where the norm may be any of the ones defined in **??** second equality is introduced in order to work in a compact set, the one of normalized vectors z.*

Notice that $\|Ax\| \leq \|A\| \cdot \|x\|$.

Property 1.2. *Let A and $B \in M(n, m, \mathbb{R})$ and let $x \in \mathbb{R}^n$, the following holds, for any norm defined in **??**:*

- $\|A\| \geq 0$ *(and the equality holds iff* $A = 0$*);*

- $\|\alpha A\| = |\alpha| \|A\|, \forall \alpha \in \mathbb{R}$*;*

- $\|A + B\| \leq \|A\| + \|B\|$*;*

- $\|AB\| \leq \|A\| \|B\|$*;*

- $\|Ax\| \leq \|A\| \|x\|$.

Fact 1.3. *Let* $A \in (n, m, \mathbb{R})$ *and let* $U \in M(m, n, \mathbb{R})$ *orthogonal, in the case of* 2*-norm* $\|A\|_2 = \|AU\|_2 = \|UA\|_2$.

Proof. $\|UA\|_2 = \max\limits_{x \in R^n,\ x \neq 0} \frac{\|UAx\|_2}{\|x\|_2} \overset{(1)}{=} \max\limits_{x \in R^n,\ x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \|A\|_2$, where $\overset{(1)}{=}$ follows from a property of vector norms.

$\|AU\|_2 = \max\limits_{x \in R^n,\ x \neq 0} \frac{\|AUx\|_2}{\|x\|_2} \overset{(2)}{=} \max\limits_{y \in R^n,\ y \neq 0} \frac{\|Ay\|_2}{\|y\|_2} = \|A\|_2$, where $\overset{(2)}{=}$ follows from the substitution $y = Ux$. $\square$

Definition 1.7 (Frobenius norm)**.** *Let* $A \in M(n, m, \mathbb{R})$*, we term* ***Frobenius norm*** *of* A $\|A\|_F = \sqrt{\sum\limits_{i=1}^{n} \sum\limits_{j=1}^{m} {(A)_{ij}}^2}$.

Notice that all the properties enumerated in Property 1.2 hold for the Frobenius norm as well.

Fact 1.4. *Let* $A \in M(n, m, \mathbb{R})$ *and let* $A = U \Sigma V^T$ *be its singular value decomposition. The following hold:*

1. $\|A\|_2 = \|\Sigma\|_2 = \sigma_1$

2. $\|A\|_F = \|Sigma\|_F = \sum\limits_{i=1}^{\min n,m} {\sigma_i}^2$

Proof. 1. The first equality follows from Proposition 1.3, while the second is proved as follows:

$$\|\Sigma\|_2 = \max_{x\in\mathbb{R}^n,\ x\neq 0} \frac{\|\Sigma x\|_2}{\|x\|_2} = \max_{x\in\mathbb{R}^n,\ x\neq 0} \frac{\left\| \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \\ & & 0 & \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \right\|_2}{\left\| \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \right\|_2} = \max_{x\in\mathbb{R}^n,\ x\neq 0} \frac{\left\| \begin{pmatrix} \sigma_1 x_1 \\ \sigma_2 x_2 \\ \vdots \\ \sigma_n x_n \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right\|_2}{\left\| \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \right\|_2}$$

$$= \frac{\sqrt{(\sigma_1 x_1)^2 + (\sigma_2 x_2)^2 + \cdots + (\sigma_n x_n)^2}}{\sqrt{{x_1}^2 + {x_2}^2 + \cdots + {x_n}^2}} \leq \frac{\sqrt{(\sigma_1 x_1)^2 + (\sigma_1 x_2)^2 + \cdots + (\sigma_1 x_n)^2}}{\sqrt{{x_1}^2 + {x_2}^2 + \cdots + {x_n}^2}}$$

$$= \sqrt{{\sigma_1}^2} \cdot \frac{\sqrt{{x_1}^2 + {x_2}^2 + \cdots + {x_n}^2}}{\sqrt{{x_1}^2 + {x_2}^2 + \cdots + {x_n}^2}} = \sigma_1 \tag{1.1}$$

The equality is achieved if we pick $x = e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$.

2. The proof of this assertion is similar to the other and it is left to the reader.

□

Theorem 1.5 (Eckart-Younger)**.** *Let $A \in M(n, m, \mathbb{R})$ and let $A = U\Sigma V^T$ be its singular value decomposition.*

The solution of $\min_{rk(X)\leq k} \|A\| - X$ *is given by the* truncated SVD*:*

$$X = \begin{pmatrix} U^1 & U^2 & \cdots & U^k \end{pmatrix} \cdot \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_k \end{pmatrix} \cdot \begin{pmatrix} V^1 \\ V^2 \\ \vdots \\ V^k \end{pmatrix}$$

Where the norm is both $\|\cdot\|_2$ and $\|\cdot\|_F$.

Fact 1.6. *Let $A \in M(n, \mathbb{R})$ and let A be invertible. The following holds: $\|A^{-1}\| = \frac{1}{\sigma_n}$*

Proof. Since A is invertible, none of the σ_i is 0, hence the smaller (namely σ_n) is not 0.

$$A^{-1} = \left(U\Sigma V^T\right)^{-1} \overset{(1)}{=} {V^T}^{-1}\Sigma^{-1}U^{-1} = V \begin{pmatrix} \frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_n} \end{pmatrix} U^T$$

Where $\overset{(1)}{=}$ follows from the orthogonlaity of V and U.

Notice that this is *almost* an SVD, because the values on the diagonal are not sorted in a decreasing order.

Plugging in the norm, we have:

$$\left\|A^{-1}\right\| = \left\|V \begin{pmatrix} \frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_n} \end{pmatrix} U^T\right\| = \left\|\begin{pmatrix} \frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_n} \end{pmatrix}\right\| = \frac{1}{\sigma_n}$$

$\square$