MATH H110- – Fall 2013
Singular Value Decomposition (SVD)

Rishi Sharma

December 13, 2013

Let $A \in \mathbb{R}^{n \times m}$ be any nonzero map $A : \mathbb{R}^m \to \mathbb{R}^n$.

Claim: $A^\intercal A$ has the same kernel as A
Proof:
Let $x \in ker(A) \implies A^\intercal Ax = A^\intercal 0 = 0 \implies x \in ker(A^\intercal A)$
Let $x \in ker(A^\intercal A) \implies x^\intercal A^\intercal Ax = 0 = ||Ax||^2 \geq 0 \implies x \in ker(A) \quad \square$

Note: We get the same result for $A^\intercal$ that $ker(AA^\intercal) = ker(A^\intercal)$

Corollary: By Rank-Nullity,

$$\begin{aligned} \underset{m\times n}{A^\intercal}\ \underset{n\times m}{A} \in \mathbb{R}^{m\times m} \quad &\implies \quad rank(A^\intercal A) = m - nul(A) \\ \underset{n\times m}{A}\ \underset{m\times n}{A^\intercal} \in \mathbb{R}^{n\times n} \quad &\implies \quad rank(AA^\intercal) = n - null(A^\intercal) \end{aligned}$$

A major portion of SVD is that $A^\intercal A$ and $AA^\intercal$ will have the same rank and range spaces with similar properties.

Claim: $A^\intercal A$ and $AA^\intercal$ are symmetric (self-adjoint), **positive semi-definite** matrices $\implies$

1. Both matrices are diagonalizable
2. All eigenvalues are real
3. All eigenvalues are non-negative

Note that a square matrix S is positive semi-definite if $x^\intercal Sx \geq 0$.

Proof: First two properties are easy: $(A^\intercal A)^\intercal = A^\intercal A^{\intercal^\intercal} = A^\intercal A$ and $(AA^\intercal)^\intercal = A^{\intercal^\intercal} A^\intercal = AA^\intercal$. Thus properties 1 and 2 follow from the Spectral Theorem of self-adjoint operators. Also, $x^\intercal A^\intercal Ax = ||Ax||^2 \geq 0$ and $x^\intercal AA^\intercal x = ||Ax||^2 \geq 0$, so $A^\intercal A$ and $AA^\intercal$ are positive semi-definite matrices. We will show that all eigenvalues of a positive semi-definite matrix are positive. Let y be an eigenvector of a positive semi-definite matrix S with eigenvalue λ

$$y^\intercal Sy = \lambda y^\intercal y = \lambda ||y|| \geq 0 \implies \lambda \geq 0 \quad \square$$

Now we will proceed to show similarities between $A^\intercal A$ and $AA^\intercal$

Claim: $A^\intercal A$ and $AA^\intercal$ have the same eigenvalues λ
Proof:
Let $x \in \mathbb{R}^m$ be an eigenvector of $A^\intercal A$ with eigenvalue λ. Then $AA^\intercal Ax = A\lambda x = \lambda Ax \implies Ax$ is an eigenvector of $AA^\intercal$ with eigenvalue λ.
Let $x \in \mathbb{R}^n$ be an eigenvector of $AA^\intercal$ with an eigenvalue λ. Then $A^\intercal AA^\intercal x = A^\intercal \lambda x = \lambda A^\intercal x \implies A^\intercal x$ is an eigenvector of $A^\intercal A$ with eigenvalue λ. $\square$

The square roots of these shared eigenvalues $\{\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_r}\}$ are referred to as the singular values of A. Note that we disclude 0 from this set of eigenvalues $\{\lambda_1, \ldots, \lambda_r\}$, because as we noted above the null spaces of $A^\intercal A$ and $AA^\intercal$ are different if $n \neq m$. It is more common to write these values as $\{\sigma_1^2, \ldots, \sigma_r^2\}$, which we know we can do because these values are positive. Now we will show that the eigenspaces that correspond to each eigenvalue have the same dimension. This is an interesting result because the eigenspaces $\{W_{\lambda_1}, \ldots, W_{\lambda_r}\} \in \mathbb{R}^m$ and $\{W'_{\lambda_1}, \ldots, W'_{\lambda_r}\} \in \mathbb{R}^n$

Claim: $dim(W_{\lambda_i}) = dim(W'_{\lambda_i}) = n_i \; \forall \; i \in \{1, \ldots, r\}$.
Proof: Let $\{u_1, \ldots, u_{n_i}\}$ be a basis for W_{λ_i}. We will show that $\{Au_1, \ldots, Au_{n_i}\} \in W'_{\lambda_i}$ (we saw that in the previous proof) is a basis for W'_{λ_i}

$$\sum_{j=1}^{n_i} c_j Au_j = A\sum_{j=1}^{n_i} c_j u_j = 0 \implies \sum_{j=1}^{n_i} c_j u_j \in ker(A) \cap W_{\lambda_i} \equiv ker(A^\intercal A) \cap W_{\lambda_i} = \{0\} \implies c_j = 0 \; \forall \; j$$

shows that $\{Au_1, \ldots, Au_{n_i}\}$ are linearly independent. Now let some $w' \in W'_{\lambda_i}$ (remember that means that $AA^\intercal w' = \lambda_i w'$), then

$$A^\intercal w' \in W_{\lambda_i} \implies A^\intercal w' = \sum_{j=1}^{n_i} v_j u_j \implies AA^\intercal w' = \sum_{j=1}^{n_i} v_j Au_j = \lambda_i w' \implies w' = \frac{1}{\lambda_i}\sum_{j=1}^{n_i} v_j Au_j$$

which shows that $\{Au_1, \ldots, Au_{n_i}\}$ is a spanning set, and thus a basis of W'_{λ_i} $\square$

So now that we have a dimension match between eigenspaces corresponding to nonzero eigenvalues, and we also know that $A^\intercal A$ and $AA^\intercal$ are diagonalizable, so lets notice

$$\begin{aligned} \mathbb{R}^m &= W_{\lambda_0} \oplus W_{\lambda_1} \oplus \cdots \oplus W_{\lambda_r} \\ \mathbb{R}^n &= W'_{\lambda_0} \oplus W'_{\lambda_1} \oplus \cdots \oplus W'_{\lambda_r} \end{aligned}$$

where each of the spaces correspond to one another except for the kernels, represented by W_{λ_0} and W'_{λ_0}.

Let's now take note of how the original matrix A acts on vectors in these eigenspace W_{λ_i} (eigenspace of $A^\intercal A$ with eigenvalue λ_i), and we will be ready for the SVD Theorem. Suppose $u, v \in W_{\lambda_i}$ such that $u \perp v \implies u^\intercal v = 0$. Then A will preserve orthogonality by $(Au)^\intercal Av = u^\intercal A^\intercal Av = \lambda_i u^\intercal v = 0$. Consider another $u \in W_{\lambda_i}$ such that $||U||^2 = u^\intercal u = 1$. As we see, A will not preserve norm, and $(Au)^\intercal Au = u^\intercal A^\intercal Au = \lambda_i u^\intercal u = \lambda_i = \sigma_i^2$, which are our singular values.

<u>SVD Theorem</u>:
We construct an orthonormal basis for W_{λ_i} of vectors $v_1^i, \ldots, v_{n_i}^i$.
We construct an orthonormal basis for W'_{λ_i} of vectors $u_1^i, \ldots, u_{n_i}^i$ by setting $u_j^i = \frac{Av_j^i}{\sigma_i}$ where $\sigma_i = \sqrt{\lambda_i}$
Then we have

$$A\left[v_1^i \cdots v_{n_i}^i\right] = \left[u_1^i \cdots u_{n_i}^i\right]\sigma_i$$

and if we chain together the orthonormal bases of all the eigenspaces (a direct sum decomposition $\mathbb{R}^k$), we get an orthornormal basis of the $\mathbb{R}^k$ where $k = \{m, n\}$. We must be careful with the subspaces W_{λ_0} and W'_{λ_0}, which are the kernels of $A^\intercal A$ and $AA^\intercal$ respectively. We get the result

$$A\left[v_1^1 \cdots v_{n_i}^1 v_1^2 \cdots v_{n_2}^2 \cdots v_1^r \cdots v_{n_r}^r v_1^0 \cdots v_{n_0}^0\right] = \left[u_1^1 \cdots u_{n_i}^1 u_1^2 \cdots u_{n_2}^2 \cdots u_1^r \cdots u_{n_r}^r u_1^0 \cdots u_{n'_0}^0\right] \begin{bmatrix} \sigma_1 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

MATRIX Σ DOES NOT LOOK LIKE THAT I'M JUST TOO LAZY TO FORMAT IT

We can rewrite this in the form of the SVD, where $U = \left[u_1^1 \cdots u_{n_i}^1 u_1^2 \cdots u_{n_2}^2 \cdots u_1^r \cdots u_{n_r}^r u_1^0 \cdots u_{n'_0}^0\right]$ and $V = \left[v_1^i \cdots v_{n_i}^i v_1^2 \cdots v_{n_2}^2 \cdots v_1^r \cdots v_{n_r}^r v_1^0 \cdots v_{n_0}^0\right]$, and Σ is the diagonal matrix of singular values.

$$A = U\Sigma V^{-1} = U\Sigma V^\intercal$$

where U and $V^\intercal$ are both orthonormal matrices that are bases for eigenspaces of $AA^\intercal$ and $A^\intercal A$ respectively.

We can also look at the Dyadic expansion of the SVD, which says

$$A = UDV^\intercal = \sum_{i=1}^{r} \sigma_i u_i v_i^\intercal$$

which can be used to see that the matrix generated by any individual singular value is a rank-1 approximation of A by the outer product.

Polar Decomposition
The SVD leads immediately to the polar decomposition of a matrix into a rotation and a scaling such that

$$A = \Theta R$$

where Θ is unitary and R is a positive semi-definite Hermitian matrix. We define this from the SVD

$$A = U\Sigma V^\intercal$$

by setting

$$\begin{aligned} \Theta &= UV^\intercal \\ R &= V\Sigma V^\intercal \end{aligned}$$

It can easily be verified that this satisfies the conditions on Θ and R. Intuitively, A is decomposes into R, which stretches the space along a set of orthonormal axes, and a rotation Θ.