

We assume real matrices $\mathbf{A} \in \mathbb{R}^{m \times n}$ for this problem

1 SVD

$$\mathbf{A} = \mathbf{U} \mathbf{S} \mathbf{V}^T \quad (1)$$

where $\mathbf{U} \in \mathbb{R}^{m \times m}$, $\mathbf{V} \in \mathbb{R}^{n \times n}$, $\mathbf{S} \in \mathbb{R}^{m \times n}$, \mathbf{U}, \mathbf{V} are orthonormal and \mathbf{S} is a rectangular diagonal matrix, i.e.

$$\mathbf{S} = \begin{cases} \hat{\mathbf{S}}_{m \times n} & \text{if } m = n \\ \begin{bmatrix} \hat{\mathbf{S}}_{n \times n} \\ \mathbf{O}_{(m-n) \times n} \end{bmatrix} & \text{if } m > n \\ \begin{bmatrix} \hat{\mathbf{S}}_{m \times m} & \mathbf{O}_{m \times (n-m)} \end{bmatrix} & \text{if } m < n \end{cases} \quad (2)$$

where $\hat{\mathbf{S}}$ is a diagonal matrix with its diagonal entries called singular values of \mathbf{A} , and \mathbf{O} is a zero matrix of appropriate dimension.

2 Solution

Let's calculate $\mathbf{S}^T \mathbf{S}$ for all the above cases.

$$\mathbf{S}^T \mathbf{S} = \begin{cases} \hat{\mathbf{S}}_{n \times m} \hat{\mathbf{S}}_{m \times n} & = \hat{\mathbf{S}}_{n \times n} \hat{\mathbf{S}}_{n \times n} & = \hat{\mathbf{S}}_{n \times n}^2 & \text{if } m = n \\ \begin{bmatrix} \hat{\mathbf{S}}_{n \times n} & \mathbf{O}_{n \times (m-n)} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{S}}_{n \times n} \\ \mathbf{O}_{(m-n) \times n} \end{bmatrix} & = \hat{\mathbf{S}}_{n \times n} \hat{\mathbf{S}}_{n \times n} + \mathbf{O}_{n \times (m-n)} \mathbf{O}_{(m-n) \times n} & = \hat{\mathbf{S}}_{n \times n}^2 & \text{if } m > n \\ \begin{bmatrix} \hat{\mathbf{S}}_{m \times m} \\ \mathbf{O}_{(n-m) \times m} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{S}}_{m \times m} & \mathbf{O}_{m \times (n-m)} \end{bmatrix} & = \begin{bmatrix} \hat{\mathbf{S}}_{m \times m} \hat{\mathbf{S}}_{m \times m} & \hat{\mathbf{S}}_{m \times m} \mathbf{O}_{m \times (n-m)} \\ \mathbf{O}_{(n-m) \times m} \hat{\mathbf{S}}_{m \times m} & \mathbf{O}_{(n-m) \times m} \mathbf{O}_{m \times (n-m)} \end{bmatrix} & = \begin{bmatrix} \hat{\mathbf{S}}_{m \times m}^2 & \mathbf{O}_{m \times (n-m)} \\ \mathbf{O}_{(n-m) \times m} & \mathbf{O}_{(n-m) \times (n-m)} \end{bmatrix} & \text{if } m < n \end{cases} \quad (3)$$

Hence $\mathbf{S}^T \mathbf{S}$ is always a diagonal matrix with the non-zero diagonal entries equal to the square of non-zero singular values of \mathbf{A} . We will use this result below. Let $\mathbf{D} = \mathbf{S}^T \mathbf{S}$ be this diagonal matrix.

(a) Now, consider $\mathbf{A}^T \mathbf{A}$,

$$\begin{aligned}
\mathbf{A}^T \mathbf{A} &= (\mathbf{U} \mathbf{S} \mathbf{V}^T)^T \mathbf{U} \mathbf{S} \mathbf{V}^T \\
&= (\mathbf{V}^T)^T \mathbf{S}^T \mathbf{U}^T \mathbf{U} \mathbf{S} \mathbf{V}^T \\
&= \mathbf{V} \mathbf{S}^T \mathbf{S} \mathbf{V}^T & (\mathbf{U}^T \mathbf{U} = \mathbf{I}) \\
&= (\mathbf{V}^T)^{-1} \mathbf{D} \mathbf{V}^T & (\mathbf{V}^T \mathbf{V} = \mathbf{I})
\end{aligned} \tag{4}$$

With this process, we have obtained a *diagonalization* of $\mathbf{A}^T \mathbf{A}$ with transformation matrix \mathbf{V}^T . Hence, from linear algebra theory, diagonal entries of \mathbf{D} are the eigenvalues of $\mathbf{A}^T \mathbf{A}$. Now, we know that non-zero diagonal entries of \mathbf{D} are equal to the square of non-zero singular values of \mathbf{A} . Hence, the eigenvalues of $\mathbf{A}^T \mathbf{A}$ are equal to the square of non-zero singular values of \mathbf{A} which implies that the non-zero singular values of \mathbf{A} are the square roots of non-zero eigenvalues of $\mathbf{A}^T \mathbf{A}$.

Similarly, consider $\mathbf{A} \mathbf{A}^T$,

$$\begin{aligned}
\mathbf{A} \mathbf{A}^T &= \mathbf{U} \mathbf{S} \mathbf{V}^T (\mathbf{U} \mathbf{S} \mathbf{V}^T)^T \\
&= \mathbf{U} \mathbf{S} \mathbf{V}^T (\mathbf{V}^T)^T \mathbf{S}^T \mathbf{U}^T \\
&= \mathbf{U} \mathbf{S} \mathbf{S}^T \mathbf{U}^T & (\mathbf{V}^T \mathbf{V} = \mathbf{I}) \\
&= (\mathbf{U}^T)^{-1} \mathbf{D}^T \mathbf{U}^T & (\mathbf{U}^T \mathbf{U} = \mathbf{I}) \\
&= (\mathbf{U}^T)^{-1} \mathbf{D} \mathbf{U}^T & (\mathbf{D}^T = \mathbf{D} \text{ as } \mathbf{D} \text{ is diagonal})
\end{aligned} \tag{5}$$

With this process, we have obtained a *diagonalization* of $\mathbf{A} \mathbf{A}^T$ with transformation matrix \mathbf{U}^T . Hence, from linear algebra theory, diagonal entries of \mathbf{D} are the eigenvalues of $\mathbf{A} \mathbf{A}^T$. Now, we know that non-zero diagonal entries of \mathbf{D} are equal to the square of non-zero singular values of \mathbf{A} . Hence, the eigenvalues of $\mathbf{A} \mathbf{A}^T$ are equal to the square of non-zero singular values of \mathbf{A} which implies that the non-zero singular values of \mathbf{A} are the square roots of non-zero eigenvalues of $\mathbf{A} \mathbf{A}^T$.

(b) The Frobenius norm of a matrix $\|\mathbf{A}\|_F$ is given by

$$\begin{aligned}
\|\mathbf{A}\|_F^2 &= \sum_i \sum_j a_{ij}^2 \\
&= \text{trace}(\mathbf{A}^T \mathbf{A}) \\
&= \text{trace}(\mathbf{V} \mathbf{S}^T \mathbf{S} \mathbf{V}^T) & (\text{From 4}) \\
&= \text{trace}(\underbrace{\mathbf{V}^T}_A \underbrace{\mathbf{V} \mathbf{S}^T \mathbf{S}}_B) & (\text{Using } \text{trace}(\mathbf{A} \mathbf{B}) = \text{trace}(\mathbf{B} \mathbf{A})) \\
&= \text{trace}(\mathbf{S}^T \mathbf{S}) & (\mathbf{V}^T \mathbf{V} = \mathbf{I}) \\
&= \text{trace}(\mathbf{D})
\end{aligned} \tag{6}$$

Now $\text{trace}(\mathbf{D})$ is sum of non-zero diagonal entries of \mathbf{D} . As these entries are equal to the square of non-zero singular values of \mathbf{A} , the squared Frobenius norm of a matrix is equal to the sum of squares of its singular values.

(c) Let's try to understand this through an example

```

>> A = [1 2 3 4; 5 6 7 8; 9 10 11 12]
A =
     1     2     3     4
     5     6     7     8
     9    10    11    12

>> [U,S,V] = svd(A)
U =

    -0.2067    -0.8892     0.4082
    -0.5183    -0.2544    -0.8165
    -0.8298     0.3804     0.4082

S =

    25.4368         0         0         0
         0     1.7226         0         0
         0         0     0.0000         0

V =

    -0.4036     0.7329     0.5241     0.1592
    -0.4647     0.2898    -0.8174     0.1784
    -0.5259    -0.1532     0.0626    -0.8343
    -0.5870    -0.5962     0.2307     0.4968

>> [V_1, D_1] = eig(A'*A)
V_1 =

    -0.1732    -0.5196     0.7329     0.4036
    -0.1564     0.8219     0.2898     0.4647
     0.8323    -0.0849    -0.1532     0.5259
    -0.5028    -0.2173    -0.5962     0.5870

D_1 =

    -0.0000         0         0         0
         0     0.0000         0         0
         0         0     2.9674         0
         0         0         0    647.0326

>> [U_1, D_2] = eig(A*A')
U_1 =

    -0.4082     0.8892     0.2067
     0.8165     0.2544     0.5183

```

```

    -0.4082    -0.3804     0.8298
D_2 =
    -0.0000         0         0
         0     2.9674         0
         0         0    647.0326

```

Firstly, we observe that the diagonal entries of \mathbf{D}_1 and \mathbf{D}_2 are in ascending order but the diagonal entries of \mathbf{S} are in descending order. According to [MATLAB documentation](#) “`eig` does not always return the eigenvalues and eigenvectors in sorted order”.

Secondly, we observe that $\mathbf{U} = -\mathbf{U}_1$, and some columns of \mathbf{V} also have different signs. In general, if \mathbf{v} is an eigenvector of a matrix with some eigenvalue then $-\mathbf{v}$ is also an eigenvector of that matrix with same eigenvalue. As \mathbf{U}, \mathbf{V} are eigenvectors here, we can't be sure about the sign.

Hence, to rectify this, we need to fix the order and use consistent signs for the eigenvectors

To fix the first issue, we can simply sort the eigenvalues and reorder columns of \mathbf{U}, \mathbf{V} appropriately.

For the second issue, we know that $\mathbf{A}\mathbf{v}_i = \sigma_i\mathbf{u}_i$ where $\mathbf{v}_i, \mathbf{u}_i$ are respective columns of \mathbf{U}, \mathbf{V} after reordering. So, we can check this relation for each non-zero singular value and if there is a sign difference we flip the signs of any one vector.

The code is given below

```

A = [1 2 3 4; 5 6 7 8; 9 10 11 12];
[U, S, V] = svd_eig(A, 1e-6);
A - U*S*V'
function [C, D] = sort_eig(A, B)
    [b, i] = sort(diag(B), 'descend');
    D = B(i,i);
    C = A(:,i);
end

function [U, S, V] = svd_eig(A, epsilon)
    s = size(A);
    m = s(1); n = s(2);
    ATA = A'*A;
    AAT = A*A';
    [V_1, D_1] = eig(ATA);
    [V, D_1] = sort_eig(V_1, D_1);
    [U_1, D_2] = eig(AAT);
    [U, D_2] = sort_eig(U_1, D_2);
    if n > m
        S = D_1;
    else
        S = D_2;
    end

```

```

end
S = sqrt(S(1:m, 1:n));
sign = ones(1,m);
for i = 1:min(m,n)
    if S(i,i) > epsilon % else we assume the singular value is zero
        if norm(U(:,i) - A*V(:,i)/S(i,i)) > epsilon
            sign(i) = -1;
        end
    end
end
U = U.*sign;
end
% this gives the following output for A - U*S*V'
1.0e-07 *
-0.1002    0.1585   -0.0164   -0.0419
 0.2005   -0.3171    0.0328    0.0839
-0.1002    0.1585   -0.0164   -0.0419

```