Recitation 2. Solution

Focus: QR decomposition, SVD, least squares.

1. Say that a square matrix A is factored as a product $A = BC^{-1}$. Perform the same column operation on both B and C, for example add the first column to the second: $B \mapsto B'$, $C \mapsto C'$. Show that this operation does not change the result, that is $A = BC^{-1} = B'(C')^{-1}$.

Solution: A column operation on B corresponds to multiplying B on the right by an appropriate invertible matrix, say M. Assuming this rule for a moment, we observe that then $B' = BM$ and $C' = CM$, because we perform the same column operation on both matrices. Therefore, plugging these formulas in, we get $B'(C')^{-1} = (BM)(CM)^{-1}$, and since taking inverses switches the order of factors, we can continue the string of equalities as $B'(C')^{-1} = BMM^{-1}C^{-1} = BC^{-1} = A$. So we achieved the desired result.

In order to understand why multiplying by M performs column operations on B, first write B as a block matrix $B = \begin{pmatrix} b_1 & \cdots & b_n \end{pmatrix}$, where b_i are columns of B. Then we can multiply B by M as block matrices (M is considered to have only trivial 1×1 blocks), and each column of BM would be expressed as a linear combination of columns of B with coefficients from some column of M. For example, if $M = (m_{ij})$, then the first column of BM is:

$$\begin{pmatrix} b_1 & \cdots & b_n \end{pmatrix} \begin{pmatrix} m_{11} \\ \vdots \\ m_{n1} \end{pmatrix} = m_{11}b_1 + m_{21}b_2 + \cdots + m_{n1}b_n.$$

This should have concluded explanation, but it might still look confusing, so let's consider several examples. First, take a 3×3 matrix B written in its block form $B = \begin{pmatrix} b_1 & b_2 & b_3 \end{pmatrix}$ and try to find such a matrix M that multiplication by M on the right would add twice the first column to the third. I claim that the following matrix works (blank spaces are zeroes):

$$M = \begin{pmatrix} 1 & 0 & 2 \\ & 1 & 0 \\ & & 1 \end{pmatrix}.$$

Let's check it (don't forget that the b_i are all column 3-vectors):

$$BM = \begin{pmatrix} b_1 & b_2 & b_3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 2 \\ & 1 & 0 \\ & & 1 \end{pmatrix} = \begin{pmatrix} b_1 & b_2 & 2b_1 + b_3 \end{pmatrix}.$$

For the second example, I want to scale up the second column of B by a factor of $\sqrt{5}$. Then I need to take the following matrix:

$$M = \begin{pmatrix} 1 & & \\ & \sqrt{5} & \\ & & 1 \end{pmatrix}.$$

You can check that BM looks like B with its second column scaled by direct block multiplication as above.

Finally, let's say that I want to switch the second and the third column (although this operation will not be used in the remainder of the worksheet). Then I need to take an appropriate *permutation matrix*:

$$M = \begin{pmatrix} 1 & & \\ & 0 & 1 \\ & 1 & 0 \end{pmatrix}.$$

2. *Finding a QR decomposition.* Write the following matrix A as a product $A = QR$ for some orthogonal Q and upper-triangular R:

$$A = \begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix}.$$

Solution: In this problem, I want to apply the previous one by writing A as a product AI^{-1}, where I is the identity matrix, and then perform column operations on both A and I simultaneously until the first matrix becomes orthogonal and we get $A = QR$. In order to orthogonalize the first factor, I will use a modified version of Gram-Schmidt process.

Recall that Gram-Schmidt algorithm goes as follows:

- Normalize the first vector.
- Subtract a suitable scalar multiple of the first vector from all the rest vectors so that they become orthogonal to the first vector.
- Normalize the second vector.
- Subtract a suitable scalar multiple of the first vector from the third and later vectors so that they become orthogonal to the second vector. Note that they stay orthogonal to the first vector.
- Etc.

For the purpose of hand calculation, I find it bothersome to normalize vectors before orthogonalizing, because I wouldn't like to carry a lot of square roots around. So I will use the following algorithm for the given 2×2 matrix:

- Subtract a suitable scalar multiple of the first vector from the second so that the result becomes orthogonal to the first vector.
- Normalize the first vector.
- Normalize the second vector.
- Now we have Q explicitly. Compute R.

Note that since we are subtracting multiples of the first column to the second, the second factor becomes upper-triangular as required. More generally, since Gram-Schmidt algorithm subtracts multiples of one column from those to the right, the second factor stays upper-triangular. See the first example in the first problem.

First step. Find a scalar λ such that $\begin{pmatrix} 4 \\ 3 \end{pmatrix} - \lambda \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ is orthogonal to $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$. I claim that $\lambda = \frac{\begin{pmatrix} 1 \\ 2 \end{pmatrix} \cdot \begin{pmatrix} 4 \\ 3 \end{pmatrix}}{\begin{pmatrix} 1 \\ 2 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix}} = \frac{4+6}{1+4} = 2$ works. Note that dots denote dot product here. Then subtracting the twice the first column from the second in both factor gives us to the following equality (do not forget that the second matrix is inverted!):

$$A = \begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}^{-1}.$$

Now columns of the first factor are orthogonal.

Second step. The magnitude of the first column is $\sqrt{1+2^2} = \sqrt{5}$, so we divide the first column of both matrices by $\sqrt{5}$:

$$A = \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{\sqrt{5}} & 2 \\ \frac{2}{\sqrt{5}} & -1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & -2 \\ 0 & 1 \end{pmatrix}^{-1}.$$

Third step. The magnitude of the second column is also $\sqrt{5}$, so we divide the second column of both matrices by $\sqrt{5}$:

$$A = \begin{pmatrix} \frac{1}{\sqrt{5}} & 2 \\ \frac{2}{\sqrt{5}} & -1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & -2 \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{5}} \\ 0 & \frac{1}{\sqrt{5}} \end{pmatrix}^{-1}.$$

Fourth step. Compute the inverse. For that, we first compute the quatity $ad - bc$, because we will need to divide by it. So $ad - bc = \frac{1}{\sqrt{5}} \cdot \frac{1}{\sqrt{5}} - 0 = \frac{1}{5}$. Dividing by it means multiplying by 5. So apply our formula for inverses of 2×2 matrices:

$$A = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{5}} \\ 0 & \frac{1}{\sqrt{5}} \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \end{pmatrix} 5 \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ 0 & \frac{1}{\sqrt{5}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} \sqrt{5} & 2\sqrt{5} \\ 0 & \sqrt{5} \end{pmatrix}.$$

3. *Finding an SVD.* Consider matrix A:

$$A = \begin{pmatrix} 1 & 3 & -1 \\ -2 & -6 & 2 \\ 3 & 9 & -3 \end{pmatrix}.$$

a) Describe its column space.
b) Express A as an outer product of two vectors. Is this decomposition unique?
c) Find a compact and a full form SVD for this matrix.

Solution:

a) $\mathrm{Col}(A) = \mathrm{Span}\begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}$, because all other columns are scalar multiples of the first one.

b) $A = \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix} \begin{pmatrix} 1 & 3 & -1 \end{pmatrix}$. The decomposition is not unique, because for example, we can multiply the column by 5 and divide the row by 5, and the result will not change.

c) A compact form SVD follows almost immediately from the column-row decomposition – what remains is to normalize the row and teh column. Note that (1) denotes a 1×1 matrix as opposed to a scalar 1.

$$A = \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix} (1) \begin{pmatrix} 1 & 3 & -1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{14}} \\ \frac{-2}{\sqrt{14}} \\ \frac{3}{\sqrt{14}} \end{pmatrix} \left(\sqrt{154}\right) \begin{pmatrix} \frac{1}{\sqrt{11}} & \frac{3}{\sqrt{11}} & \frac{-1}{\sqrt{11}} \end{pmatrix}.$$

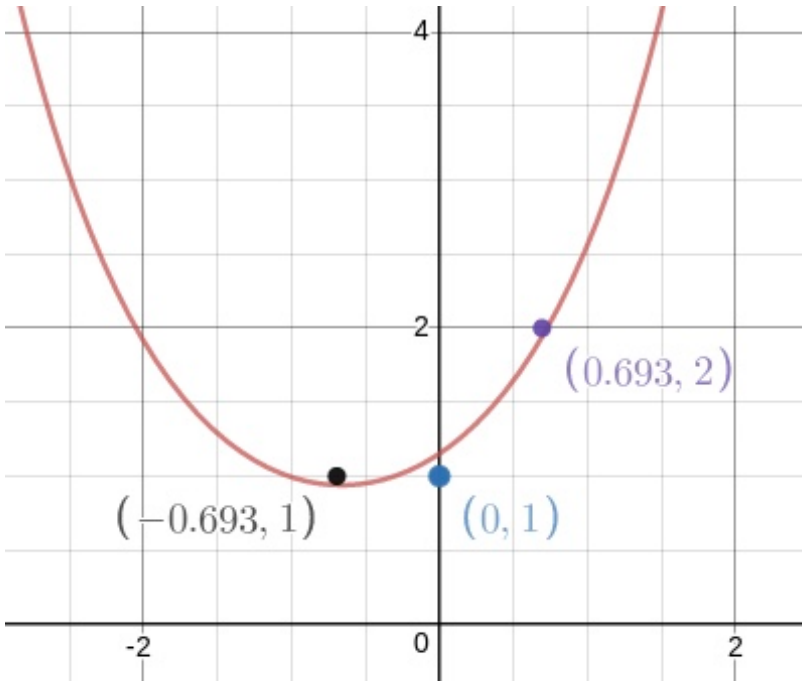
A full form SVD can be found by picking two complementary orthogonal vector to the column, and then to the row. Here is an example for the column:

$$A = \begin{pmatrix} \frac{1}{\sqrt{14}} & \frac{1}{\sqrt{3}} & \frac{5}{\sqrt{42}} \\ \frac{-2}{\sqrt{14}} & \frac{-1}{\sqrt{3}} & \frac{4}{\sqrt{42}} \\ \frac{3}{\sqrt{14}} & \frac{-1}{\sqrt{3}} & \frac{1}{\sqrt{42}} \end{pmatrix} \begin{pmatrix} \sqrt{154} & & \\ & 0 & \\ & & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{11}} & \frac{3}{\sqrt{11}} & \frac{-1}{\sqrt{11}} \\ * & * & * \\ * & * & * \end{pmatrix}.$$

4. *Least squares approximation.* Consider the set of functions of the form $f(x) = ae^{x} + be^{-x}$, where a and b vary over real numbers. In this space of functions, use the least squares algorithm to approximate the unknown function U that takes the following values:

$$U\left(\ln \frac{1}{2}\right) = 1,\ U(0) = 1,\ U(\ln 2) = 2.$$

You will get the following picture:



a) Write the sum $S(a,b)$ of squared errors.
b) Write the condition of finding local minimum using partial derivatives with respect to a and b.
c) Write the condition above as a matrix equation and solve this equation.

Solution:

a) $S(a,b) = \sum_{i=1}^{n} \left(f(x_i) - y_i\right)^2$. In our case, $n = 3$ and the datapoints are given by the values of U, so we plug this in: $S(a,b) = \left(\frac{1}{2}a + 2b - 1\right)^2 + (a+b-1)^2\left(2a + \frac{1}{2}b - 2\right)^2$.

b) First compute partial derivative with respect to a:

$$\frac{\partial S}{\partial a}(a,b) = 2\sum_{i=1}^{n}\left(ae^{x_i} + be^{-x_i} - y_i\right)e^{x_i} = 2\sum_{i=1}^{n}\left(ae^{2x_i} + b - y_i e^{x_i}\right).$$

Combine similar summands:

$$\frac{\partial S}{\partial a}(a,b) = 2\left(\sum_{i=1}^{n} e^{2x_i}\right)a + 2\left(\sum_{i=1}^{n} 1\right)b - 2\left(\sum_{i=1}^{n} y_i e^{x_i}\right).$$

Now we can plug in our datapoints:

$$\frac{\partial S}{\partial a}(a,b) = \frac{21}{4}a + 3b - \frac{11}{2}.$$

Similarly for b:

$$\frac{\partial S}{\partial b}(a,b) = 2\left(\sum_{i=1}^{n} 1\right)a + 2\left(\sum_{i=1}^{n} e^{-2x_i}\right)b - 2\left(\sum_{i=1}^{n} y_i e^{-x_i}\right).$$

And now with the given data:

$$\frac{\partial S}{\partial b}(a,b) = 3a + \frac{21}{4}b - 4.$$

The condition for finding an extremum is for all the partial derivatives to vanish, so:

$$\begin{cases} \frac{21}{4}a + \ \ 3b = \frac{11}{2}, \\ \ \ 3a + \frac{21}{4}b = 4. \end{cases}$$

c) $\begin{pmatrix} \frac{21}{4} & 3 \\ 3 & \frac{21}{4} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \frac{11}{2} \\ 4 \end{pmatrix}$. Solution can be obtained by multiplying both sides by the inverse of $\begin{pmatrix} \frac{21}{4} & 3 \\ 3 & \frac{21}{4} \end{pmatrix}$ on the left. The resulting coefficients are $a = \frac{30}{33}$ and $b = \frac{8}{33}$. You can find the plot of the corresponding function above.