

Mathematics for Machine Learning

Additional Exercises

Marc Peter Deisenroth, A. Aldo Faisal, Cheng Soon Ong

Last update: 2020-05-16

In this living document, we provide additional exercises (including solutions) for the mathematics chapters of our book *Mathematics for Machine Learning*, published by Cambridge University Press (2020). Possible solutions are shown in blue. They may not be unique or optimal.

If you find mistakes, please raise a github issue at

<https://github.com/mml-book/mml-book.github.io/issues>.

Chapter 2

1. Find all solutions of the inhomogeneous system of linear equations $\mathbf{Ax} = \mathbf{b}$, where

(a)

$$\mathbf{A} := \begin{bmatrix} 1 & 2 \\ 3 & 0 \\ -1 & 2 \end{bmatrix}, \quad \mathbf{b} := \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

To determine the general solution of the inhomogeneous system of linear equations, a good start is to compute the reduced row echelon form of the augmented system $[\mathbf{A}|\mathbf{b}]$:

$$\begin{aligned} & \left[\begin{array}{cc|c} 1 & 2 & 1 \\ 3 & 0 & 0 \\ -1 & 2 & 1 \end{array} \right] \begin{array}{l} -3R_1 \\ +R_1 \end{array} \\ \rightsquigarrow & \left[\begin{array}{cc|c} 1 & 2 & 1 \\ 0 & -6 & -3 \\ 0 & 4 & 2 \end{array} \right] \begin{array}{l} +\frac{1}{3}R_2 \\ \cdot (-\frac{1}{6}) \\ +\frac{2}{3}R_2 \end{array} \rightsquigarrow \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 \end{array} \right] \end{aligned}$$

From the last row of this augmented system, we see that $0x_1 + 0x_2 = 0$, which is always true. From the other rows, we obtain $x_1 = 0$ and $x_2 = \frac{1}{2}$, so that

$$\mathbf{x} = \begin{bmatrix} 0 \\ \frac{1}{2} \end{bmatrix}$$

is the unique solution of the system of linear equations $\mathbf{Ax} = \mathbf{b}$.

(b)

$$\mathbf{A} := \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 2 \end{bmatrix}, \quad \mathbf{b} := \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

The general solution consists of a particular solution of the inhomogeneous system and all solutions of the homogeneous system $\mathbf{Ax} = \mathbf{0}$. An efficient way to determine the general solution is via the reduced row echelon form (RREF) of the augmented system $[\mathbf{A}|\mathbf{b}]$:

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 0 & 2 & 2 & 1 \end{array} \right] \begin{array}{l} -R_2 \\ \cdot \frac{1}{2} \end{array} \rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & \frac{1}{2} \end{array} \right]$$

- From the RREF, we can read out a *particular solution* (not unique) by using the pivot columns as

$$\mathbf{x}_p = \begin{bmatrix} 0 \\ \frac{1}{2} \\ 0 \end{bmatrix} \in \mathbb{R}^3.$$

Here, we set x_1 to the right-hand side of the augmented RREF in the first row, and x_2 to the right-hand side of the augmented RREF in the second row. Since $\mathbf{x}_p \in \mathbb{R}^3$ (otherwise the matrix-vector multiplication $\mathbf{A}\mathbf{x} = \mathbf{b}$ would not be defined), the third coordinate $x_3 = 0$.

- Next, we determine all solutions of the homogeneous system of linear equations $\mathbf{A}\mathbf{x} = \mathbf{0}$. From the left-hand side of the augmented RREF, we can immediately read out the solutions as

$$\lambda \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \quad \lambda \in \mathbb{R},$$

where we used the Minus-1 trick.

- Putting everything together, we obtain the set of all solutions of the system $\mathbf{A}\mathbf{x} = \mathbf{b}$ as

$$\mathcal{S} = \left\{ \mathbf{x} \in \mathbb{R}^3 : \mathbf{x} = \begin{bmatrix} 0 \\ \frac{1}{2} \\ 0 \end{bmatrix} + \lambda \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \quad \lambda \in \mathbb{R} \right\}.$$

2. Compute the matrix products \mathbf{AB} , if possible, where

(a)

$$\mathbf{A} := \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & 2 \end{bmatrix}, \quad \mathbf{B} := \begin{bmatrix} 4 & -1 & 2 \\ 0 & 2 & 1 \end{bmatrix}$$

This matrix multiplication is not defined since $\mathbf{A} \in \mathbb{R}^{2 \times 3}$ and $\mathbf{B} \in \mathbb{R}^{2 \times 3}$. For the matrix product to be defined, the “neighboring” dimensions (columns of \mathbf{A} and rows of \mathbf{B}) would need to match. Here, they are 2 and 3.

(b)

$$\mathbf{A} := \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & 2 \end{bmatrix}, \quad \mathbf{B} := \begin{bmatrix} 4 & -1 \\ 2 & 0 \\ 2 & 1 \end{bmatrix}$$

$$\mathbf{AB} = \begin{bmatrix} 14 & 2 \\ 2 & 2 \end{bmatrix},$$

where (for example) $14 = 1 \cdot 4 + 2 \cdot 2 + 3 \cdot 2$.

3. Find the intersection $L_1 \cap L_2$, where L_1 and L_2 are affine spaces (subspaces that are offset from $\mathbf{0}$) defined as

$$L_1 := \underbrace{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}}_{=: \mathbf{p}_1} + \underbrace{\text{span}\left[\begin{bmatrix} -3 \\ -2 \\ 1 \end{bmatrix}\right]}_{=: U_1}, \quad L_2 := \underbrace{\begin{bmatrix} 10 \\ 6 \\ -2 \end{bmatrix}}_{=: \mathbf{p}_2} + \underbrace{\text{span}\left[\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 5 \\ 4 \\ 1 \end{bmatrix}\right]}_{=: U_2}.$$

$$\mathbf{x} \in L_1 \iff \mathbf{x} = \mathbf{p}_1 + \alpha \mathbf{b}_1$$

for some $\alpha \in \mathbb{R}$. We defined \mathbf{b}_1 as the basis vector of U_1 . Similarly,

$$\mathbf{x} \in L_2 \iff \mathbf{x} = \mathbf{p}_2 + \beta_1 \mathbf{c}_1 + \beta_2 \mathbf{c}_2$$

for some $\beta_1, \beta_2 \in \mathbb{R}$ and $U_2 = \text{span}[\mathbf{c}_1, \mathbf{c}_2]$. Therefore, for all $\mathbf{x} \in L_1 \cap L_2$ both conditions must hold and we arrive at

$$\mathbf{x} \in L_1 \cap L_2 \iff \exists \alpha, \beta_1, \beta_2 \in \mathbb{R} : \alpha \mathbf{b}_1 - \beta_1 \mathbf{c}_1 - \beta_2 \mathbf{c}_2 = \mathbf{p}_2 - \mathbf{p}_1$$

which leads to the inhomogeneous system of linear equations $\mathbf{A}\boldsymbol{\lambda} = \mathbf{b}$ where $\boldsymbol{\lambda} = [\alpha, \beta_1, \beta_2]^\top$ and

$$\mathbf{A} := \begin{bmatrix} -3 & -1 & -5 \\ -2 & -1 & -4 \\ 1 & -1 & -1 \end{bmatrix}, \quad \mathbf{b} := \mathbf{p}_2 - \mathbf{p}_1 = \begin{bmatrix} 9 \\ 6 \\ -3 \end{bmatrix}$$

We bring the augmented system $[\mathbf{A}|\mathbf{b}]$ into reduced row echelon form using Gaussian elimination:

$$\begin{bmatrix} 1 & 0 & 1 & -3 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and read out the particular solution $\alpha = -3 \Rightarrow \boldsymbol{\xi} = \mathbf{p}_1 - 3\mathbf{b}_1 = [10, 6, -2]^\top = \mathbf{p}_2$.

To find the general solution, we need to look at the intersection of the direction spaces $U_1 \cap U_2$. The corresponding RREF that we obtain is identical to the submatrix

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

of the reduced row echelon form of the augmented system. We obtain $\beta_1 = -2\beta_2$, such that

$$U_1 \cap U_2 = \text{span}\left[\begin{bmatrix} -3 \\ -2 \\ 1 \end{bmatrix}\right].$$

We then arrive at the final solution

$$L_1 \cap L_2 = \begin{bmatrix} 10 \\ 6 \\ -2 \end{bmatrix} + \text{span}\left[\begin{bmatrix} -3 \\ -2 \\ 1 \end{bmatrix}\right] = L_1,$$

i.e., $L_1 \subseteq L_2$.

Chapter 3

1. Consider \mathbb{R}^3 with $\langle \cdot, \cdot \rangle$ defined for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$ as

$$\langle \mathbf{x}, \mathbf{y} \rangle := \mathbf{x}^\top \mathbf{A} \mathbf{y}, \quad \mathbf{A} := \begin{bmatrix} 4 & 2 & 1 \\ 0 & 4 & -1 \\ 1 & -1 & 5 \end{bmatrix}.$$

Is $\langle \cdot, \cdot \rangle$ an inner product?

We will show that $\langle \cdot, \cdot \rangle$ is not symmetric, i.e., $\langle \mathbf{x}, \mathbf{y} \rangle \neq \langle \mathbf{y}, \mathbf{x} \rangle$.

We choose $\mathbf{x} := [1, 1, 0]^\top$ and $\mathbf{y} := [1, 2, 0]^\top$. Then $\langle \mathbf{x}, \mathbf{y} \rangle = 16$ and $\langle \mathbf{y}, \mathbf{x} \rangle = 14 \neq 16$.

In general, we can see directly that \mathbf{A} is not symmetric. Similarly, for a symmetric \mathbf{A} , we would need to check that it is positive definite (e.g., via the eigenvalues of \mathbf{A}).

Chapter 4

1. Compute the determinants of the following matrices:

(a)

$$\mathbf{A} := \begin{bmatrix} 1 & 0 & -3 & 0 & 9 \\ 3 & 7 & 10 & 3 & 17 \\ 4 & 0 & 11 & 0 & 1 \\ 6 & 0 & 8 & 0 & -3 \\ 5 & 1 & 6 & -1 & 8 \end{bmatrix}$$

$$\begin{vmatrix} 1 & 0 & -3 & 0 & 9 \\ 3 & 7 & 10 & 3 & 17 \\ 4 & 0 & 11 & 0 & 1 \\ 6 & 0 & 8 & 0 & -3 \\ 5 & 1 & 6 & -1 & 8 \end{vmatrix} = \begin{vmatrix} 1 & 0 & -3 & 0 & 9 \\ 18 & 10 & 28 & 0 & 41 \\ 4 & 0 & 11 & 0 & 1 \\ 6 & 0 & 8 & 0 & -3 \\ 5 & 1 & 6 & -1 & 8 \end{vmatrix}$$

where we added 3 times the last row to the second row. Now, we develop the determinant about the fourth column:

$$\begin{aligned} \det(\mathbf{A}) &= (-1)(-1)^{4+5} \begin{vmatrix} 1 & 0 & -3 & 9 \\ 18 & 10 & 28 & 41 \\ 4 & 0 & 11 & 1 \\ 6 & 0 & 8 & -3 \end{vmatrix} \stackrel{2^{\text{nd}} \text{ col}}{=} 10 \begin{vmatrix} 1 & -3 & 9 \\ 4 & 11 & 1 \\ 6 & 8 & -3 \end{vmatrix} \\ &= 10(-33 - 18 + 288 - 594 - 8 - 36) = -4010, \end{aligned}$$

where we can use the Sarrus rule.

(b)

$$\mathbf{B} := \begin{bmatrix} 2 & 0 & 4 & 5 \\ 1 & 1 & 1 & 1 \\ 9 & 2 & 0 & 0 \\ 0 & 0 & 2 & 3 \end{bmatrix}$$

$$\begin{aligned} \begin{vmatrix} 2 & 0 & 4 & 5 \\ 1 & 1 & 1 & 1 \\ 9 & 2 & 0 & 0 \\ 0 & 0 & 2 & 3 \end{vmatrix} &\stackrel{\text{col } 2}{=} \begin{vmatrix} 2 & 4 & 5 \\ 9 & 0 & 0 \\ 0 & 2 & 3 \end{vmatrix} - 2 \begin{vmatrix} 2 & 4 & 5 \\ 1 & 1 & 1 \\ 0 & 2 & 3 \end{vmatrix} = -9 \begin{vmatrix} 4 & 5 \\ 2 & 3 \end{vmatrix} - 2 \left(-2 \begin{vmatrix} 2 & 5 \\ 1 & 1 \end{vmatrix} + 3 \begin{vmatrix} 2 & 4 \\ 1 & 1 \end{vmatrix} \right) \\ &= -9(12 - 10) - 2(-2 \cdot (2 - 5) + 3(2 - 4)) = -18 - 2(6 - 6) = -18. \end{aligned}$$

We could have seen that the second 3×3 -matrix after the development about the 2nd column is rank deficient (the third row is the first row minus twice the second row), which results in a determinant of 0.

2. Consider an endomorphism $\Phi : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ with transformation matrix

$$\mathbf{A} = \begin{bmatrix} 4 & 0 & -2 \\ 1 & 3 & -2 \\ 1 & 2 & -1 \end{bmatrix}, \quad \lambda \in \mathbb{R}$$

- (a) Compute the characteristic polynomial of \mathbf{A} and determine all eigenvalues.
We have

$$\begin{aligned}
 p(\lambda) &= \det(\mathbf{A} - \lambda \mathbf{I}) = \begin{vmatrix} 4-\lambda & 0 & -2 \\ 1 & 3-\lambda & -2 \\ 1 & 2 & -1-\lambda \end{vmatrix} \stackrel{\text{1st row}}{=} (4-\lambda) \begin{vmatrix} 3-\lambda & -2 \\ 2 & -1-\lambda \end{vmatrix} - 2 \begin{vmatrix} 1 & 3-\lambda \\ 1 & 2 \end{vmatrix} \\
 &= (4-\lambda)((3-\lambda)(-1-\lambda) + 4) - 2(2 - (3-\lambda)) \\
 &= (4-\lambda)(3-\lambda)(-1-\lambda) + 4(4-\lambda) - 4 + 2(3-\lambda) \\
 &= -\lambda^3 + 6\lambda^2 - 11\lambda + 6.
 \end{aligned}$$

Now, we need to find the eigenvalues, i.e., the roots of $p(\lambda)$:

$$\begin{aligned}
 -\lambda^3 + 6\lambda^2 - 11\lambda + 6 &= 0 \\
 \iff \lambda^3 - 6\lambda^2 + 11\lambda - 6 &= 0 \\
 \iff (\lambda - 1)(\lambda - 2)(\lambda - 3) &= 0
 \end{aligned}$$

Therefore, the eigenvalues are 1, 2, 3.

- (b) Compute bases of all eigenspaces.

We use Gaussian eliminatin to determine $E_1 = \ker(\mathbf{A} - \mathbf{I})$

$$\begin{bmatrix} 3 & 0 & -2 \\ 1 & 2 & -2 \\ 1 & 2 & -2 \end{bmatrix} \begin{array}{l} -3R_2 \\ -R_2 \end{array} \rightsquigarrow \begin{bmatrix} 0 & -6 & 4 \\ 1 & 2 & -2 \\ 0 & 0 & 0 \end{bmatrix} \begin{array}{l} \cdot(-\frac{1}{6}) \\ +\frac{1}{3}R_2 \mid \text{swap with } R_1 \end{array} \rightsquigarrow \begin{bmatrix} 1 & 0 & -\frac{2}{3} \\ 0 & 1 & -\frac{3}{2} \\ 0 & 0 & 0 \end{bmatrix}$$

Therefore,

$$E_1 = \text{span}\left[\begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}\right].$$

We use again Gaussian elimination to determine $E_2 = \ker(\mathbf{A} - 2\mathbf{I})$

$$\begin{bmatrix} 2 & 0 & -2 \\ 1 & 1 & -2 \\ 1 & 2 & -3 \end{bmatrix} \begin{array}{l} -2R_2 \\ -R_2 \end{array} \rightsquigarrow \begin{bmatrix} 0 & -2 & 2 \\ 1 & 1 & -2 \\ 0 & 1 & -1 \end{bmatrix} \begin{array}{l} +2R_3 \\ -R_3 \mid \text{move to } R_1 \\ \text{move to } R_2 \end{array} \rightsquigarrow \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

and we obtain

$$E_2 = \text{span}\left[\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right]$$

Finally, $E_3 = \ker(\mathbf{A} - 3\mathbf{I})$, which we compute via Gaussian elimination:

$$\begin{bmatrix} 1 & 0 & -2 \\ 1 & 0 & -2 \\ 1 & 2 & -4 \end{bmatrix} \begin{array}{l} -R_1 \\ -R_1 \end{array} \rightsquigarrow \begin{bmatrix} 1 & 0 & -2 \\ 0 & 0 & 0 \\ 0 & 2 & -2 \end{bmatrix} \begin{array}{l} \text{swap with } R_3 \\ \cdot\frac{1}{2} \end{array} \rightsquigarrow \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

such that

$$E_3 = \text{span}\left[\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}\right]$$

- (c) Determine a transformation matrix \mathbf{B} such that $\mathbf{B}^{-1}\mathbf{A}\mathbf{B}$ is a diagonal matrix and provide this diagonal matrix.

The desired matrix \mathbf{B} consists of all eigenvectors (as the columns of the matrix), and is given by

$$\begin{bmatrix} 2 & 1 & 2 \\ 2 & 1 & 1 \\ 3 & 1 & 1 \end{bmatrix}.$$

The corresponding diagonal matrix is then

$$\mathbf{B}^{-1}\mathbf{A}\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

Note that this is the diagonal matrix with the eigenvalues on the diagonal. If you compute $\mathbf{B}^{-1}\mathbf{A}\mathbf{B}$ and should get the same answer.

3. Consider the matrix

$$\mathbf{A} = \begin{bmatrix} -3 & 4 & 4 \\ -5 & 9 & 5 \\ -7 & 4 & 8 \end{bmatrix}$$

The aim is to find a matrix $\mathbf{M} \in \mathbb{R}^{3 \times 3}$ such that $\mathbf{M}^2 = \mathbf{A}$ (a “square root” of \mathbf{A}).

- (a) Find an invertible matrix \mathbf{P} and a diagonal matrix \mathbf{D} such that $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$.

The characteristic polynomial of \mathbf{A} is $p(\lambda) = -\lambda^3 + 14\lambda^2 - 49\lambda + 36$. An obvious root of this polynomial is 1, and we can factorize $p(\lambda) = -(\lambda-1)(\lambda-4)(\lambda-9)$, which gives us the eigenvalues 1, 4, 9.

We use Gaussian elimination to compute eigenspace $E_1 = \ker(\mathbf{A} - 1\mathbf{I})$, and we get $E_1 = \text{span}\left[\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\right]$.

Similarly, we get $E_4 = \text{span}\left[\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}\right]$ and $E_9 = \text{span}\left[\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}\right]$. We then define the invertible matrix \mathbf{P} and the diagonal matrix \mathbf{D} as

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \\ 1 & -1 & 1 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 9 \end{bmatrix}$$

so that $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$.

- (b) Let \mathbf{M} be in $\mathbb{R}^{3 \times 3}$ and let us assume that $\mathbf{M}^2 = \mathbf{A}$. Let us consider $\mathbf{N} = \mathbf{P}^{-1}\mathbf{M}\mathbf{P}$. Show that $\mathbf{N}^2 = \mathbf{D}$. Then prove that \mathbf{N} commutes with \mathbf{D} , i.e., $\mathbf{N}\mathbf{D} = \mathbf{D}\mathbf{N}$.

Exploiting the associativity of matrix multiplication, we obtain

$$\mathbf{N}^2 = (\mathbf{P}^{-1}\mathbf{M}\mathbf{P})(\mathbf{P}^{-1}\mathbf{M}\mathbf{P}) = \mathbf{P}^{-1}\mathbf{M}(\mathbf{P}\mathbf{P}^{-1})\mathbf{M}\mathbf{P} = \mathbf{P}^{-1}\mathbf{M}^2\mathbf{P} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D}$$

and, therefore,

$$\mathbf{N}\mathbf{D} = \mathbf{N}(\mathbf{N}^2) = \mathbf{N}^3 = (\mathbf{N}^2)\mathbf{N} = \mathbf{D}\mathbf{N}.$$

- (c) Explain that \mathbf{N} is thus necessarily diagonal.

Hint: Note that all the diagonal values of \mathbf{D} are distinct.

Intuitively, as \mathbf{D} is diagonal, the product $\mathbf{N}\mathbf{D}$ multiplies the columns of \mathbf{N} while $\mathbf{D}\mathbf{N}$ multiplies the rows of \mathbf{N} . But as $\mathbf{N}\mathbf{D} = \mathbf{D}\mathbf{N}$, and \mathbf{D} has different values on the diagonal, then \mathbf{N} has to be diagonal. Let us prove this result formally.

Let us denote by $n_{i,j}$ the coefficient of matrix \mathbf{N} at row i and column j and let d_i denote the i^{th} coefficient on the diagonal of \mathbf{D} . Note that in our example, i and j will be ranged in $\{1, 2, 3\}$, but this result extends to matrices of arbitrary size. Let i and j be in $\{1, 2, 3\}$. The coefficient of \mathbf{ND} at row i and column j is equal to $n_{i,j}d_j$, while that of \mathbf{DN} is equal to $d_i n_{i,j}$. The matrix equality $\mathbf{ND} = \mathbf{DN}$ yields

$$\forall i, j \in \{1, 2, 3\}: n_{i,j}d_j = n_{i,j}d_i,$$

i.e.,

$$\forall i, j \in \{1, 2, 3\}: n_{i,j}(d_j - d_i) = 0. \quad (1)$$

In general, a product is null if and only if at least one of its factors is null. But as all the values on the diagonal of \mathbf{D} are different, (1) is equivalent to

$$\forall i, j \in \{1, 2, 3\}: (i \neq j) \implies (n_{i,j} = 0),$$

which ensures that \mathbf{N} is diagonal. Note that if two values on the diagonal of \mathbf{D} were equal, \mathbf{N} would not necessarily be diagonal and we would have infinitely many candidates for \mathbf{N} , and thus as many for \mathbf{M} .

- (d) What can you say about \mathbf{N} 's possible values? Compute a matrix \mathbf{M} , whose square is equal to \mathbf{A} . How many different such matrices are there?

We can write \mathbf{N} as $\mathbf{N} = \text{diag}(n_1, n_2, n_3)$ and $\mathbf{N}^2 = \mathbf{D}$ requires that $n_1^2 = 1$, $n_2^2 = 4$ and $n_3^2 = 9$. As all diagonal values are positive, we have exactly two distinct square roots for each one. Therefore, we have 8 possible values for \mathbf{N} that we gather in the following set:

$$\{\text{diag}(n_1, n_2, n_3) \mid n_1 \in \{-1, +1\}, n_2 \in \{-2, +2\}, n_3 \in \{-3, +3\}\}.$$

Now, let us set $\mathbf{N} = \text{diag}(1, 2, 3)$ and compute the product $\mathbf{M} = \mathbf{PNP}^{-1}$. First, Gaussian elimination gives us

$$\mathbf{P}^{-1} = \frac{1}{2} \begin{bmatrix} 3 & -1 & -1 \\ 1 & 0 & -1 \\ -1 & 1 & 1 \end{bmatrix},$$

and we find one square root of \mathbf{A} as

$$\mathbf{M} = \mathbf{PNP}^{-1} = \begin{bmatrix} 0 & 1 & 1 \\ -1 & 3 & 1 \\ -2 & 1 & 3 \end{bmatrix}.$$

We can check that \mathbf{M}^2 indeed equals \mathbf{A} . We can choose amongst the 8 different possible values of \mathbf{N} to find a new square root of \mathbf{A} . Hence, there are equally many different such matrices \mathbf{M} .

4. <https://github.com/mml-book/mml-book.github.io/issues/338>

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$. Show that:

- \mathbf{AA}^\top and $\mathbf{A}^\top \mathbf{A}$ have identical non-zero eigenvalues.
- If \mathbf{q} is an eigenvector of \mathbf{AA}^\top then $\mathbf{A}^\top \mathbf{q}$ is an eigenvector of $\mathbf{A}^\top \mathbf{A}$.
- If \mathbf{p} is an eigenvector of $\mathbf{A}^\top \mathbf{A}$ then \mathbf{Ap} is an eigenvector of \mathbf{AA}^\top .
- We start by showing that if $\lambda \neq 0$ is an eigenvalue of \mathbf{AA}^\top then it is also a non-zero eigenvalue of $\mathbf{A}^\top \mathbf{A}$.

Let $\lambda \neq 0$ be an eigenvalue of \mathbf{AA}^\top and \mathbf{q} be a corresponding eigenvector, i.e., $(\mathbf{AA}^\top)\mathbf{q} = \lambda\mathbf{q}$. Then

$$(\mathbf{A}^\top \mathbf{A})\mathbf{A}^\top \mathbf{q} = \mathbf{A}^\top (\mathbf{AA}^\top \mathbf{q}) = \mathbf{A}^\top (\lambda\mathbf{q}) = \lambda\mathbf{A}^\top \mathbf{q}.$$

We now need to show that $\mathbf{A}^\top \mathbf{q} \neq \mathbf{0}$ before we can conclude that λ is an eigenvalue of $\mathbf{A}^\top \mathbf{A}$.

Assume $\mathbf{A}^\top \mathbf{q} = \mathbf{0}$. Then it would follow that $\mathbf{AA}^\top \mathbf{q} = \mathbf{0}$, which contradicts $\mathbf{AA}^\top \mathbf{q} = \lambda\mathbf{q} \neq \mathbf{0}$ since \mathbf{q} is an eigenvector of \mathbf{AA}^\top with associated eigenvalue λ . Therefore, $\mathbf{q} \neq \mathbf{0}$, which implies that $\mathbf{A}^\top \mathbf{q} \neq \mathbf{0}$.

Therefore, λ is an eigenvalue of $\mathbf{A}^\top \mathbf{A}$ with $\mathbf{A}^\top \mathbf{q}$ as the corresponding eigenvector.

- Let us now consider the case where $\lambda \neq 0$ is an eigenvalue of $\mathbf{A}^\top \mathbf{A}$. We want to show that λ is also an eigenvalue of $\mathbf{A}\mathbf{A}^\top$.
Let $\lambda \neq 0$ be an eigenvalue of $\mathbf{A}^\top \mathbf{A}$ and \mathbf{p} be a corresponding eigenvector, i.e., $(\mathbf{A}^\top \mathbf{A})\mathbf{p} = \lambda\mathbf{p}$. Then

$$(\mathbf{A}\mathbf{A}^\top)\mathbf{A}\mathbf{p} = \mathbf{A}(\mathbf{A}^\top \mathbf{A})\mathbf{p} = \mathbf{A}(\lambda\mathbf{p}) = \lambda\mathbf{A}\mathbf{p}.$$

Similar to above, we now need to show that $\mathbf{A}\mathbf{p} \neq \mathbf{0}$ before we can draw our conclusions.

Assume $\mathbf{A}\mathbf{p} = \mathbf{0}$. Then $\mathbf{0} = \mathbf{A}\mathbf{p} = \mathbf{A}^\top \mathbf{A}\mathbf{p} = \lambda\mathbf{p}$ with $\lambda \neq 0$. This contradicts our assumption that \mathbf{p} is an eigenvector of $\mathbf{A}^\top \mathbf{A}$. Therefore $\mathbf{A}\mathbf{p} \neq \mathbf{0}$.

Therefore, $\lambda \neq 0$ is an eigenvalue of $\mathbf{A}\mathbf{A}^\top$, and a corresponding eigenvector is $\mathbf{A}\mathbf{p}$.

Chapter 5

1. Consider

$$\mathbf{f} := \mathbf{A}\mathbf{x},$$

where $\mathbf{A} \in \mathbb{R}^{3 \times 2}$ and $\mathbf{x} \in \mathbb{R}^2$. Compute the partial derivative

$$\frac{\partial \mathbf{f}}{\partial \mathbf{A}}.$$

- We start by determining the dimension of the partial derivative. Knowing the dimensions of \mathbf{A} and \mathbf{x} , it follows that $\mathbf{f} \in \mathbb{R}^3$. Therefore, $\frac{\partial \mathbf{f}}{\partial \mathbf{A}} \in \mathbb{R}^{3 \times (3 \times 2)}$.
- We look at every element of $\mathbf{f} := [f_1, f_2, f_3]^\top$ and determine the corresponding partial derivatives. By definition,

$$f_i = \sum_{j=1}^2 A_{ij}x_j$$

for $i = 1, 2, 3$. Therefore,

$$\begin{aligned} \frac{\partial f_i}{\partial A_{ij}} &= x_j \\ \frac{\partial f_i}{\partial A_{kj}} &= 0 \end{aligned}$$

for $k \neq i$. This then gives

$$\begin{array}{cccccc} \frac{\partial f_1}{\partial A_{11}}=x_1 & \frac{\partial f_1}{\partial A_{12}}=x_2 & \frac{\partial f_1}{\partial A_{21}}=0 & \frac{\partial f_1}{\partial A_{22}}=0 & \frac{\partial f_1}{\partial A_{31}}=0 & \frac{\partial f_1}{\partial A_{32}}=0 \\ \frac{\partial f_2}{\partial A_{11}}=0 & \frac{\partial f_2}{\partial A_{12}}=0 & \frac{\partial f_2}{\partial A_{21}}=x_1 & \frac{\partial f_2}{\partial A_{22}}=x_2 & \frac{\partial f_2}{\partial A_{31}}=0 & \frac{\partial f_2}{\partial A_{32}}=0 \\ \frac{\partial f_3}{\partial A_{11}}=0 & \frac{\partial f_3}{\partial A_{12}}=0 & \frac{\partial f_3}{\partial A_{21}}=0 & \frac{\partial f_3}{\partial A_{22}}=0 & \frac{\partial f_3}{\partial A_{31}}=x_1 & \frac{\partial f_3}{\partial A_{32}}=x_2 \end{array}$$

We now have all our 18 entries that we need to construct our $3 \times 3 \times 2$ partial derivative, which can be done in the following way (where we store the partial derivatives in $d\mathbf{f}$):

$$d\mathbf{f}[i, j, k] = \frac{\partial f_i}{\partial A_{jk}}.$$

From above, we see that

$$df[:, :, 1] = \frac{\partial f}{\partial \mathbf{A}_{:,1}} = \begin{bmatrix} x_1 & 0 & 0 \\ 0 & x_1 & 0 \\ 0 & 0 & x_1 \end{bmatrix} \in \mathbb{R}^{3 \times 3}, \quad df[:, :, 2] = \frac{\partial f}{\partial \mathbf{A}_{:,2}} = \begin{bmatrix} x_2 & 0 & 0 \\ 0 & x_2 & 0 \\ 0 & 0 & x_2 \end{bmatrix} \in \mathbb{R}^{3 \times 3},$$

which is what we expect if we compute the partial derivative of a vector $\mathbf{f} \in \mathbb{R}^3$ with respect to a column vector $\mathbf{A}_{:,i} \in \mathbb{R}^3$ of matrix \mathbf{A} .

- An alternative approach is to vectorize \mathbf{A} , compute the partial derivatives, and then re-assemble them afterwards. Here, we define a vector

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_6 \end{bmatrix} := \frac{\begin{bmatrix} A_{11} \\ A_{21} \\ A_{31} \\ A_{12} \\ A_{22} \\ A_{32} \end{bmatrix}}{\begin{bmatrix} A_{11} \\ A_{21} \\ A_{31} \\ A_{12} \\ A_{22} \\ A_{32} \end{bmatrix}} \in \mathbb{R}^6,$$

which consists of the stacked columns of \mathbf{A} . Using this vector, we obtain the elements of \mathbf{f} as

$$\begin{aligned} f_1 &= a_1 x_1 + a_4 x_2 \\ f_2 &= a_2 x_1 + a_5 x_2 \\ f_3 &= a_3 x_1 + a_6 x_2. \end{aligned}$$

The partial derivative of $\mathbf{f} \in \mathbb{R}^3$ with respect to $\mathbf{a} \in \mathbb{R}^6$ results in the 3×6 matrix

$$\frac{\partial \mathbf{f}}{\partial \mathbf{a}} = \left[\begin{array}{ccc|ccc} x_1 & 0 & 0 & x_2 & 0 & 0 \\ 0 & x_1 & 0 & 0 & x_2 & 0 \\ 0 & 0 & x_1 & 0 & 0 & x_2 \end{array} \right] \in \mathbb{R}^{3 \times 6}.$$

We can now get the desired partial derivative as

$$df[:, :, 1] = \begin{bmatrix} x_1 & 0 & 0 \\ 0 & x_1 & 0 \\ 0 & 0 & x_1 \end{bmatrix}, \quad df[:, :, 2] = \begin{bmatrix} x_2 & 0 & 0 \\ 0 & x_2 & 0 \\ 0 & 0 & x_2 \end{bmatrix}.$$

Chapter 6

Chapter 7