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 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, where

   (a)

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

   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 $[\boldsymbol{A}|\boldsymbol{b}]$:

   $$\left[\begin{array}{cc|c} 1 & 2 & 1 \\ 3 & 0 & 0 \\ -1 & 2 & 1 \end{array}\right] \begin{array}{l} \\ -3R_1 \\ +R_1 \end{array}$$

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

   From the last row of this augmented system, we see that $0\boldsymbol{x}_1 + 0\boldsymbol{x}_2 = 0$, which is always true. From the other rows, we obtain $x_1 = 0$ and $x_2 = \frac{1}{2}$, so that

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

   is the unique solution of the system of linear equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$.

   (b)

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

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

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

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

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

  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 $\boldsymbol{x}_p \in \mathbb{R}^3$ (otherwise the matrix-vector multiplication $\boldsymbol{Ax} = \boldsymbol{b}$ would not be defined), the third coordinate $x_3 = 0$.
- Next, we determine all solutions of the homogeneous system of linear equations $\boldsymbol{Ax} = \boldsymbol{0}$. 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 $\boldsymbol{Ax} = \boldsymbol{b}$ as

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

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

   (a)

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

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

   (b)

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

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

   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 $\boldsymbol{0}$) defined as

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

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

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

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

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

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

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

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

We bring the augmented system $[\boldsymbol{A}|\boldsymbol{b}]$ 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 $\alpha = -3 \Rightarrow \boldsymbol{\xi} = \boldsymbol{p}_1 - 3\boldsymbol{b}_1 = [10, 6, -2]^\top = \boldsymbol{p}_2$.
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

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

We then arrive at the final solution

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

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

Chapter 3

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

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

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

   We will show that $\langle \cdot, \cdot \rangle$ is not symmetric, i.e., $\langle \boldsymbol{x}, \boldsymbol{y} \rangle \neq \langle \boldsymbol{y}, \boldsymbol{x} \rangle$.
   We choose $\boldsymbol{x} := [1, 1, 0]^\top$ and $\boldsymbol{y} = [1, 2, 0]^\top$. Then $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 16$ and $\langle \boldsymbol{y}, \boldsymbol{x} \rangle = 14 \neq 16$.
   In general, we can see directly that $\boldsymbol{A}$ is not symmetric. Similarly, for a symmetric $\boldsymbol{A}$, we would need to check that it is positive definite (e.g., via the eigenvalues of $\boldsymbol{A}$).

Chapter 4

1. Compute the determinants of the following matrices:

   (a)

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

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

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

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

   where we can use the Sarrus rule.

   (b)

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

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

   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 $\Phi : \mathbb{R}^3 \to \mathbb{R}^3$ with transformation matrix

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

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

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

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

$$-\lambda^3 + 6\lambda^2 - 11\lambda + 6 = 0$$
$$\iff \lambda^3 - 6\lambda^2 + 11\lambda - 6 = 0$$
$$\iff (\lambda-1)(\lambda-2)(\lambda-3) = 0$$

Therefore, the eigenvalues are $1, 2, 3$.

(b) Compute bases of all eigenspaces.
We use Gaussian eliminatin to determine $E_1 = \ker(\boldsymbol{A} - \boldsymbol{I})$

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

Therefore,

$$E_1 = \operatorname{span}[\begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}].$$

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

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

and we obtain

$$E_2 = \operatorname{span}[\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}]$$

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

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

such that

$$E_3 = \operatorname{span}[\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}]$$

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

The desired matrix $\boldsymbol{B}$ 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

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

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

3. Consider the matrix

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

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

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

The characteristic polynomial of $\boldsymbol{A}$ 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 $E_1 = \ker(\boldsymbol{A} - 1\boldsymbol{I})$, and we get $E_1 = \operatorname{span}[\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}]$.

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

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

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

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

Exploiting the associativity of matrix multiplication, we obtain

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

and, therefore,

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

(c) Explain that $\boldsymbol{N}$ is thus necessarily diagonal.

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

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

Let us denote by $n_{i,j}$ the coefficient of matrix $\boldsymbol{N}$ at row i and column j and let d_i denote the i^{th} coefficient on the diagonal of $\boldsymbol{D}$. 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 $\boldsymbol{ND}$ at row i and column j is equal to $n_{i,j}d_j$, while that of $\boldsymbol{DN}$ is equal to $d_i n_{i,j}$. The matrix equality $\boldsymbol{ND} = \boldsymbol{DN}$ yields

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

i.e.,

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

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 $\boldsymbol{D}$ are different, (1) is equivalent to

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

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

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

We can write $\boldsymbol{N}$ as $\boldsymbol{N} = \operatorname{diag}(n_1, n_2, n_3)$ and $\boldsymbol{N}^2 = \boldsymbol{D}$ 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 $\boldsymbol{N}$ that we gather in the following set:

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

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

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

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

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

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

4. `https://github.com/mml-book/mml-book.github.io/issues/338`
Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. Show that:

- $\boldsymbol{AA}^\top$ and $\boldsymbol{A}^\top\boldsymbol{A}$ have identical non-zero eigenvalues.
- If $\boldsymbol{q}$ is an eigenvector of $\boldsymbol{AA}^\top$ then $\boldsymbol{A}^\top\boldsymbol{q}$ is an eigenvector of $\boldsymbol{A}^\top\boldsymbol{A}$.
- If $\boldsymbol{p}$ is an eigenvector of $\boldsymbol{A}^\top\boldsymbol{A}$ then $\boldsymbol{Ap}$ is an eigenvector of $\boldsymbol{AA}^\top$.

- We start by showing that if $\lambda \neq 0$ is an eigenvalue of $\boldsymbol{AA}^\top$ then it is also a non-zero eigenvalue of $\boldsymbol{A}^\top\boldsymbol{A}$.
Let $\lambda \neq 0$ be an eigenvalue of $\boldsymbol{AA}^\top$ and $\boldsymbol{q}$ be a corresponding eigenvector, i.e., $(\boldsymbol{AA}^\top)\boldsymbol{q} = \lambda\boldsymbol{q}$. Then

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

We now need to show that $\boldsymbol{A}^\top\boldsymbol{q} \neq 0$ before we can conclude that λ is an eigenvalue of $\boldsymbol{A}^\top\boldsymbol{A}$.
Assume $\boldsymbol{A}^\top\boldsymbol{q} = \boldsymbol{0}$. Then it would follow that $\boldsymbol{AA}^\top\boldsymbol{q} = \boldsymbol{0}$, which contradicts $\boldsymbol{AA}^\top\boldsymbol{q} = \lambda\boldsymbol{q} \neq \boldsymbol{0}$ since $\boldsymbol{q}$ is an eigenvector of $\boldsymbol{AA}^\top$ with associated eigenvalue λ. Therefore, $\boldsymbol{q} \neq \boldsymbol{0}$, which implies that $\boldsymbol{A}^\top\boldsymbol{q} \neq \boldsymbol{0}$.
Therefore, λ is an eigenvalue of $\boldsymbol{A}^\top\boldsymbol{A}$ with $\boldsymbol{A}^\top\boldsymbol{q}$ as the corresponding eigenvector.

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

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

  Similar to above, we now need to show that $\boldsymbol{A}\boldsymbol{p} \neq \boldsymbol{0}$ before we can draw our conclusions.
  Assume $\boldsymbol{A}\boldsymbol{p} = \boldsymbol{0}$. Then $\boldsymbol{0} = \boldsymbol{A}\boldsymbol{p} = \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{p} = \lambda \boldsymbol{p}$ with $\lambda \neq 0$. This contradicts our assumption that $\boldsymbol{p}$ is an eigenvector of $\boldsymbol{A}^\top \boldsymbol{A}$. Therefore $\boldsymbol{A}\boldsymbol{p} \neq \boldsymbol{0}$.
  Therefore, $\lambda \neq 0$ is an eigenvalue of $\boldsymbol{A}\boldsymbol{A}^\top$, and a corresponding eigenvector is $\boldsymbol{A}\boldsymbol{p}$.

Chapter 5

1. Consider

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

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

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

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

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

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

$$\frac{\partial f_i}{\partial A_{ij}} = x_j$$
$$\frac{\partial f_i}{\partial A_{kj}} = 0$$

  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 df):

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

From above, we see that

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

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

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

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_6 \end{bmatrix} := \left[\begin{array}{c} A_{11} \\ A_{21} \\ A_{31} \\ \hline A_{12} \\ A_{22} \\ A_{32} \end{array}\right] \in \mathbb{R}^6,$$

which consists of the stacked columns of $\boldsymbol{A}$. Using this vector, we obtain the elements of $\boldsymbol{f}$ 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 $\boldsymbol{f} \in \mathbb{R}^3$ with respect to $\boldsymbol{a} \in \mathbb{R}^6$ results in the 3×6 matrix

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

We can now get the desired partial derivative as

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

Chapter 6

Chapter 7