Tutorials for Lecture Course on Modeling and System Identification (MSI)
Albert-Ludwigs-Universität Freiburg – Winter Term 2023-2024

Emergency Guide to Linear Algebra

Recall of important Matrix Properties and Operation

Prof. Dr. Moritz Diehl, Tobias Schöls, Katrin Baumgärtner, Alexander Petrov, Reworked by Jakob Harzer

1 Motivation (or why would you do this?)

Matrices are common in many fields of engineering, i.e. measurements are often stored as a matrix, for example series of voltage measurements. On top of that formulating the math that is used to process these data as matrix operations is usually more compact and convenient. Therefore you will have to deal with matrices a lot during this course. However, we understand that matrices might not be intuitive for everyone, especially if you have not dealt with them in a long time. This tutorial is meant to get you used to working with matrices (again).
Along with this tutorial, we also provide a jupyter notebook that gives examples on how to use `PYTHON` to perform the operations in each of the sections.

1.1 Warm-Up Exercises

The following exercises are meant to refresh your memory and get you used to matrices again. We recommend you calculate the tasks by hand first and then check the result using `PYTHON`.

$$A = \begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix} \qquad B = \begin{bmatrix} 5 & 7 \\ 8 & 6 \end{bmatrix} \qquad C = \begin{bmatrix} 0 & 1 & 5 & 6 \\ 1 & 0 & 3 & 1 \end{bmatrix} \qquad v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$(A+B)v = \tag{1}$$

$$Av + Bv = \tag{2}$$

$$(A+B)C = \tag{3}$$

$$AA^{-1} = \tag{4}$$

$$v^\top v = \tag{5}$$

$$vv^\top = \tag{6}$$

$$A(BC) = \tag{7}$$

$$(AB)C = \tag{8}$$

$$A^\top = \tag{9}$$

$$(Av)^\top = \tag{10}$$

$$v^\top A^\top = \tag{11}$$

$$v^\top A^\top A v = \tag{12}$$

$$\sum_{i=1}^{2} v_i = \tag{13}$$

$$\begin{bmatrix} 1 & 1 \end{bmatrix} v = \tag{14}$$

Convert the following system of equations into it's equivalent matrix form $Ax = b$ by defining the matrix A and the vector b.

$$\begin{aligned} 3x_1 + 2x_2 + \quad 6x_3 - 5 &= 0 \\ 4x_2 + 0.5x_3 \qquad &= 10 \end{aligned}$$

$$A = \qquad\qquad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad\qquad b = \tag{15}$$

2 Matrix and Vector Properties and Operations

2.1 Norm of a Vector

In linear algebra norms are functions that compute the length or a similar measure of a vector. There are several ways to define a norm. We will only use two:

- **Euclidean norm** Most common norm definition, straight-line distance between two points (here $\mathbf{x}$ and the origin).

$$\begin{aligned} \|x\|_2 &= \sqrt{x_1^2 + \cdots + x_n^2} \\ \|x\|_2^2 &= x_1^2 + \cdots + x_n^2 = x^\top x \end{aligned}$$

- **1-norm**

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

Calculate both the euclidian and the 1-norm of the vector $v = \begin{bmatrix} 2 & -4 & 4 \end{bmatrix}^\top$.

$$\|v\|_2 = \qquad\qquad \|v\|_1 = \tag{16}$$

2.2 Rank of a Matrix

The rank of a matrix is the number of linear independent rows. This is equivalent to saying the rank of a matrix is the number of independent columns. A matrix is said to have full rank if all rows or columns are linearly independent, that is the rank matches the dimension of that matrix. For linear equation systems this means that a unique solution exists.

Compute the rank of the remaining two matrices

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{does not have full rank, since it contains a zero row.}$$

$$\begin{bmatrix} 3 & 4 & 1 \\ 5 & 7 & 9 \\ 6 & 8 & 2 \end{bmatrix} \tag{17}$$

$$\begin{bmatrix} 8 & 1 & 6 \\ 3 & 5 & 7 \\ 4 & 9 & 2 \end{bmatrix} \tag{18}$$

2.3 Inverse

A square matrix $A \in \mathbb{R}^{n \times n}$ is called invertible if there exists a matrix $B \in \mathbb{R}^{n \times n}$ such that

$$AB = BA = \mathbb{I}_n$$

where $\mathbb{I}_n$ is a n-by-n identity matrix. If B exists, it is unique and called the inverse of A, denoted by A^{-1}. Note that non-square matrices do not have an inverse.
Let A be a square matrix. Then the following statements are equivalent:

- A is invertible.
- A has full rank.
- The determinant of A is not zero.
- A has only non-zero eigenvalues.

If A is not invertible, then A is called singular or degenerate.

Calculating the Inverse For a non-singular 2-by-2 matrix the inverse can be calculated in the closed form

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

but for larger matrices more complicated methods are needed.

Solution of Linear Systems With the above we find a solution to the linear system $Ax = b$ by performing the following calculation

$$Ax = b \Leftrightarrow A^{-1}Ax = A^{-1}b \Leftrightarrow \mathbb{I}x = A^{-1}b \Leftrightarrow x = A^{-1}b$$

Solve following linear system using the calculation above.

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \tag{19}$$

2.4 Eigenvalues and Eigenvectors

Vectors that do not change the direction when multiplied with A are called *eigenvectors* here denoted as $\mathbf{v}$. When A is multiplied with one of its eigenvectors the result is just a scalar multiple of that eigenvector. This can be formulated in a formula as

$$A\mathbf{v} = \lambda\mathbf{v}$$

where $\mathbf{v}$ is an eigenvector and λ is the corresponding *eigenvalue*.[1]
As an example, consider the following equation $Ax = b$ where A is defined as

$$A = \begin{bmatrix} 1.25 & -0.75 \\ -0.75 & 1.25 \end{bmatrix}$$

This is a linear map from $\mathbb{R}^2$ to $\mathbb{R}^2$. Its eigenvalues are $\lambda_1 = 0.5$ and $\lambda_2 = 2$ the corresponding eigenvectors are $v_1 = -\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \end{bmatrix}^\top$ and $v_2 = -\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \end{bmatrix}^\top$. In the 2-D case this can be visualized by the deformation of a unit circle (figure 2) and a unit square (figure 1).

From math class you may remember that the eigenvalues are the roots of the characteristic polynomial. For this class you do not need to compute them by hand and you can rely on `PYTHON` to find them for you in the exercises.
Optional: Come up with your own matrix and plot the deformation of the unit square/circle, similar to the figures above.

2.5 Outlook on Covariance Ellipsoids

Throughout this course you will be dealing with some random variables that follow a certain distribution. The spread of this distribution and the correlation between different variables is contained in a covariance matrix. To visualize the spread of distributions and the relation between different variables you will use deformed unit circles (similar to what you have seen above). These ellipses are called confidence ellipses. [Don't panic! You have seen this already.] Take a look at the figure 3 below. The thing we are talking about is the light blue circle around the blue dot that marks the (most likely) position.

[1] `http://math.mit.edu/~gs/linearalgebra/ila0601.pdf`

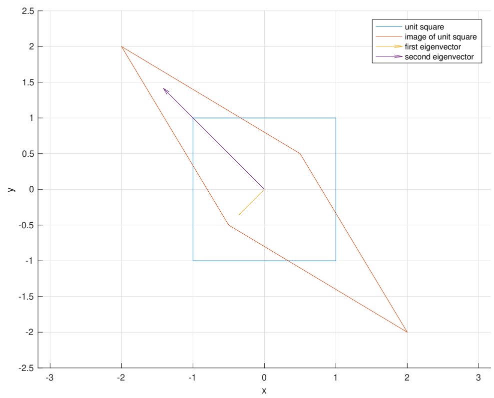

Figure 1: Deformation of a unit square through example transformation.

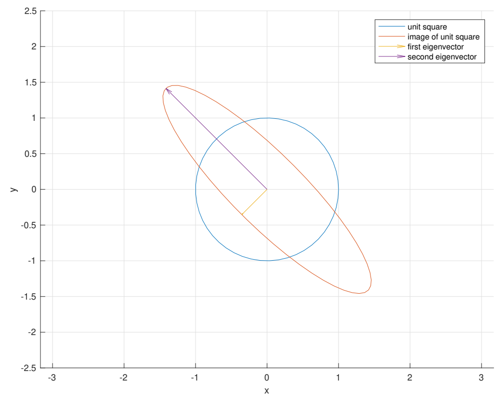

Figure 2: Deformation of a unit circle through example transformation.

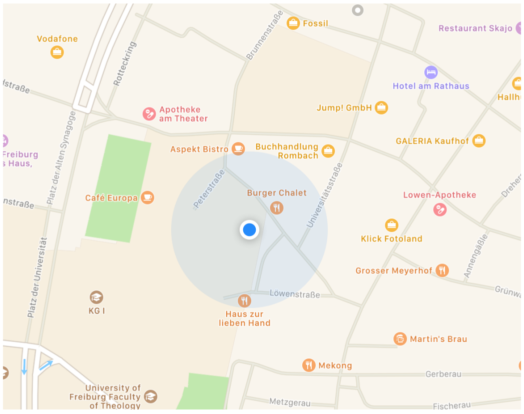

Figure 3: Example Confidence Ellipse: location estimate and area in which the actual position is most likely to lie in.

3 Special Matrices

3.1 Symmetric Matrices

A matrix A is called *symmetric* if it is equal to its transpose, i.e. $A = A^\top$. An example for this is the matrix

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

Please note that only square matrices can be symmetric and that the product of a matrix with its transpose is symmetric. Thus, for any $B \in \mathbb{R}^{m \times n}$ it holds

$$B^\top B = B^\top \left(B^\top\right)^\top = \left(B^\top B\right)^\top$$

where we used $(AB)^\top = B^\top A^\top$ and $\left(A^\top\right)^\top = A$. In addition, symmetric matrices have only real eigenvalues.

Fill in the gaps or compute the the following symmetric matrices

$$\begin{bmatrix} a & b & \dots \\ \dots & d & e \\ c & \dots & f \end{bmatrix} \tag{20}$$

$$\begin{bmatrix} v \\ w \end{bmatrix} \begin{bmatrix} v & w \end{bmatrix} = \tag{21}$$

$$\begin{bmatrix} v & w \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} = \tag{22}$$

3.2 Positive/Negative (Semi-)Definite Matrices

If a symmetric matrix has no negative eigenvalue (all are positive or zero) it is called *positive semi-definite* (PSD). The same holds for *positive definite* matrices only that the zero is not allowed as eigenvalue. Similarly a *negative definite* matrix has only

strictly negative eigenvalues and a *negative semi-definite* has no positive eigenvalue (all negative or zero).
An alternative definition of positive/negative (semi-)definiteness is the following: Let $M \in \mathbb{R}^{n\times n}$ be a symmetric matrix. If for all $x \in \mathbb{R}^n$, $x \neq 0$, it holds

$x^\top Mx < 0$, then M is called negative-definite.

$x^\top Mx \leq 0$, then M is called negative-semi-definite.

$x^\top Mx > 0$, then M is called positive-definite

$x^\top Mx \geq 0$, then M is called positive-semi-definite.

If none of the above is true, then M is called indefinite.

Square symmetric matrices of dimension n we sometimes use the symbol $\mathbb{S}_n$, i.e. $\mathbb{S}_n = \{A \in \mathbb{R}^{n\times n} | A = A^\top\}$. For any symmetric matrix $A \in \mathbb{S}_n$ we write $A \succcurlyeq 0$ if it is a positive semi-definite matrix, i.e. all its eigenvalues are larger or equal to zero, and $A \succ 0$ if it is positive definite, i.e. all its eigenvalues are positive. This notation is also used for *matrix inequalities* that allow us to compare two symmetric matrices $A, B \in \mathbb{S}_n$, where we define for example $A \succcurlyeq B$ by $A - B \succcurlyeq 0$.
A *positive-definite* matrix is always invertible. The inverse of a positive-definite matrix is also positive-definite.

For *positive semi-definite* matrices, the following properties hold:

- For any matrix $A \in \mathbb{R}^{m\times n}$, it holds that $A^\top A$ is positive semi-definite (PSD).
- For M PSD, it holds that for all $r > 0$ that rM is PSD.
- If M is PSD, then $A^\top MA$ is also PSD.

Determine if the matrices below are positive semi-definite and give a short reason.

$$\begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix} \tag{23}$$

$$\begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 3 & 2 \end{bmatrix} \tag{24}$$

$$\begin{bmatrix} 8 & 1 & 6 \\ 3 & 5 & 7 \end{bmatrix} \tag{25}$$

$$\begin{bmatrix} 8 & 3 \\ 1 & 5 \\ 6 & 7 \end{bmatrix} \begin{bmatrix} 8 & 1 & 6 \\ 3 & 5 & 7 \end{bmatrix} \tag{26}$$

$$\begin{bmatrix} 8 & 1 & 6 \\ 3 & 5 & 7 \end{bmatrix} \begin{bmatrix} 8 & 3 \\ 1 & 5 \\ 6 & 7 \end{bmatrix} \tag{27}$$

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \tag{28}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{29}$$

3.3 Orthogonal Matrices

Square matrices that if multiplied with their own transpose equal the identity matrix are called *orthogonal matrix*. In mathematical terms this is expressed as: If $AA^\top = A^\top A = \mathbb{I}$, then A is called an orthogonal matrix. This is equivalent to:

$$A^\top = A^{-1}$$

Examples for such matrices are:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} 0.866 & -0.5 \\ 0.5 & 0.866 \end{bmatrix} \qquad \begin{bmatrix} 0.7071 & -0.7071 \\ 0.7071 & 0.7071 \end{bmatrix} \qquad \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}$$

Orthogonal matrices have interesting properties:

- An orthogonal matrix is always invertible.
- The determinant of an orthogonal matrix is always ± 1.

Check if the following matrices are orthogonal.

$$\begin{bmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \tag{30}$$

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{31}$$

$$\frac{1}{3} \begin{bmatrix} 2 & -2 & 1 \\ 1 & 2 & 2 \\ 2 & 1 & -2 \end{bmatrix} \tag{32}$$

3.4 Upper/Lower Triangular Matrices

If all entries of a square matrix above the main diagonal are zero this matrix is called a *lower triangular* matrix. Similarly if all entries of a square matrix below the main diagonal are zero this matrix is called a *upper triangular* matrix. Examples for such matrices are:

$$\text{upper triangular matrix:} \quad \begin{bmatrix} 1 & 99 & 5 \\ 0 & 3 & 6 \\ 0 & 0 & 1 \end{bmatrix} \qquad \text{lower triangular matrix:} \quad \begin{bmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f \end{bmatrix} \tag{33}$$

Some properties are:

- The transpose of an upper triangular matrix is a lower triangular matrix and vice versa.
- The determinant of a triangular matrix equals the product of the diagonal entries.

Calculate the determinant of the matrices above. What do you notice?

Upper and lower triangular matrices play an important role when solving linear equation systems, as a linear system in this form is easy to solve. This is illustrated by the following task:

Solve the this system of equations for x_1, x_2, and x_3

$$\begin{bmatrix} 1 & 99 & 5 \\ 0 & 3 & 6 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 120 \\ 13 \\ 2 \end{bmatrix}$$

$$x_1 = \qquad\qquad x_2 = \qquad\qquad x_3 = \tag{34}$$

3.5 Diagonal Matrices

Matrices in which the off-diagonal entries are zero are called *diagonal matrix*, i.e. for a diagonal matrix any entry $d_{i,j}$ with $i \neq j$ is 0.

Fill out the gaps such that this is a diagonal matrix.

$$\begin{bmatrix} 1 & \dots & \dots \\ \dots & 2 & \dots \\ \dots & \dots & 3 \end{bmatrix} \tag{35}$$

This definition also applies for non-square matrices. To be a little bit more specific: Only entries $d_{i,j}$ with $i = j$ may be non-zero. For example

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \end{bmatrix}$$

are both diagonal matrices.
For diagonal matrices, the following properties hold:

- The sum of diagonal matrices is again diagonal
- The product $C = AB$ of two diagonal matrices A and B is again a diagonal matrix where the diagonal entry $c_{i,i}$ is given by the product of the corresponding diagonal entries in A and B, i.e. $c_{i,i} = a_{i,i} \cdot b_{i,i}$.
- The inverse of a diagonal square matrix is defined if all diagonal entries are non zero. The inverse is then given by a diagonal matrix with inverse of the diagonal entries.

Please do the following calculations (without PYTHON *).*

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 7 \end{bmatrix} + \begin{bmatrix} 4 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 5 \end{bmatrix} = \tag{36}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 7 \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 5 \end{bmatrix} = \tag{37}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 7 \end{bmatrix}^{-1} = \tag{38}$$

4 Matrix and Vector Valued Functions

In linear algebra functions are not defined for scalars but for vectors or matrices. This works similarly but might be a bit unintuitive at first.
Examples for such functions that take the vector $x = [x_1\, x_2]^T$ as input are

$$f(x) = \begin{bmatrix} 3x_2 \\ x_1 + x_2 \\ x_1 x_2 \end{bmatrix} \qquad g(x) = \begin{bmatrix} x_1 & 0 & 0 \\ 0 & -x_2 & 0 \\ 0 & 0 & x_1 x_2 \end{bmatrix}$$

Notice that f is vector-valued and g is matrix-valued.

Create a function h that inverts the order of the input vector $x = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}^T$.

$$h(x) = \tag{39}$$

4.1 Linear and Affine Functions

Any function that can be written as $\mathbf{f}(\mathbf{x}) = A\mathbf{x}$ is called a *linear function*. If a linear function is extended by a constant term it becomes an *affine function* and has the form $\mathbf{f}(\mathbf{x}) = A\mathbf{x} + b$.

Please reformulate the following functions in the form $\mathbf{f}(\mathbf{x}) = A\mathbf{x} + b$, where $\mathbf{f} = \begin{bmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_n(\mathbf{x}) \end{bmatrix}$

$$f(\mathbf{x}) = \begin{bmatrix} f(\mathbf{x}) \end{bmatrix} = 5x_1 + 7x_2 + 9 \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad A = \qquad b = \tag{40}$$

$$g(\mathbf{x}) = \begin{bmatrix} g_1(\mathbf{x}) \\ g_2(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} 5x_1 + 7x_2 + 9 \\ 24x_1 + 23x_3 - 42 \end{bmatrix} \qquad \mathbf{x} = \qquad A = \qquad b = \tag{41}$$

$$h(x) = \begin{bmatrix} h_1(\mathbf{x}) \\ h_2(\mathbf{x}) \\ h_3(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} 5x_1 + 7x_2 + 9 \\ x_2 + \frac{1}{2} \\ 25x_1 - 49x_2 + 81 \end{bmatrix} \qquad \mathbf{x} = \qquad A = \qquad b = \tag{42}$$

4.2 Quadratic Functions

Quadratic functions have a a slightly different structure than their scalar complements.

$$\mathbf{f}(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x} + B\mathbf{x} + c$$

Please reformulate the following functions in the form $\mathbf{f}(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x} + B\mathbf{x} + c$

$$f_1(\mathbf{x}) = 7x_1^2 + 4x_1x_2 + 2x_2^2 \qquad \mathbf{x} = \qquad A = \qquad B = \qquad c = \tag{43}$$

$$g_1(\mathbf{x}) = f_1(\mathbf{x}) + 5x_1 + 7x_2 + 9 \qquad \mathbf{x} = \qquad A = \qquad B = \qquad c = \tag{44}$$

4.3 Derivatives

Derivatives are very common and have many applications. For a function $\mathbf{f}: \mathbb{R}^n \to \mathbb{R}^m$ we define the derivative with respect to its parameter vector $\mathbf{x}$ as follows (instead $\mathbf{f}(\mathbf{x})$ we write $\mathbf{f}$ here for cleaner notation):

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial \mathbf{f}}{\partial x_1} & \cdots & \frac{\partial \mathbf{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{m \times n}$$

The matrix above is called the Jacobian matrix. For a scalar function $f: \mathbb{R}^n \to \mathbb{R}$, we define the gradient vector as

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix} \in \mathbb{R}^n$$

Thus, we have $\nabla f(x) = \frac{\partial f}{\partial x}(x)^T$. Based on the above definitions, we can derive a number of differentiation rules. The list below includes some important rules that will be handy for this course. Let A be a matrix of appropriate size.

$$\begin{aligned}
\mathbf{f} = \mathbf{x}: \quad & \frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \mathbb{I}_n \\
\mathbf{f} = A\mathbf{x}: \quad & \frac{\partial \mathbf{f}}{\partial \mathbf{x}} = A \\
\mathbf{f} = \mathbf{x}^\top A\mathbf{x}: \quad & \frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \mathbf{x}^\top A + (A\mathbf{x})^\top = \mathbf{x}^\top (A + A^\top) \\
\mathbf{f} = \mathbf{x}^\top A^\top A\mathbf{x}: \quad & \frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \mathbf{x}^\top A^\top A + (A^\top A\mathbf{x})^\top = 2\mathbf{x}^\top A^\top A
\end{aligned}$$

Above we considered the partial derivatives of f, i.e. when calculating $\frac{\partial f}{\partial x_i}$, we consider all other variables x_j, $j \neq i$, to be constants. However, this is not always reasonable. Regard for example a function $f(x,t)$ that depends on the position x and the time t. The position, however, changes with time, we actually have $x = x(t)$. Rather than calculating $\frac{\partial f}{\partial t}$, we would in this case be interested in the so-called *total derivative*. Here, it is given by

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x}\frac{\partial x}{\partial t}$$

where we made use of the chain rule. Note that for scalar functions, the partial and total derivative coincide.