# Matrix Algebra Marathon I. (marathon01-matrix.pdf)

Tsuyoshi Kato

April 5, 2020

1

## 1 Notation

We denote vectors by bold-faced lower-case letters and matrices by bold-faced upper-case letters. Entries of vectors and matrices are not bold-faced. The transposition of a matrix $\boldsymbol{A}$ is denoted by $\boldsymbol{A}^\top$, and the inverse of $\boldsymbol{A}$ is by $\boldsymbol{A}^{-1}$. The $n \times n$ identity matrix is denoted by $\boldsymbol{I}_n$. We use $\boldsymbol{E}_{ij}$ to denote a matrix in which $(i,j)$th entry is one and all the others are zero. The $n$-dimensional vector all of whose entries are one is denoted by $\mathbf{1}_n$. We use $\mathbb{R}$ and $\mathbb{N}$ to denote the set of real and natural numbers, $\mathbb{R}^n$ and $\mathbb{N}^n$ to denote the set of $n$-dimensional real and natural vectors, and $\mathbb{R}^{m\times n}$ to denote the set of $m \times n$ real matrices. The set of real nonnegative numbers is denoted by $\mathbb{R}_+$. For any $n \in \mathbb{N}$, we use $\mathbb{N}_n$ to denote the set of natural numbers less than or equal to $n$. We use $\mathbb{S}^n$ to denote the set of symmetric $n \times n$ matrices. $\mathbb{S}^n_+$ to denote the set of symmetric positive semi-definite $n \times n$ matrices, and $\mathbb{S}^n_{++}$ to denote the set of symmetric strictly positive definite $n \times n$ matrices. $\mathbb{O}^{m\times n}$ is used to denote the set of $m \times n$ orthonormal matrices, i.e. $\mathbb{O}^{m\times n} := \{\boldsymbol{A} \in \mathbb{R}^{m\times n} \,|\, \boldsymbol{A}^\top\boldsymbol{A} = \boldsymbol{I}_n\}$. The definition implies that $\mathbb{O}^{m\times n} = \emptyset$ if $m < n$. The $n$-dimensional probabilistic simplex is denoted by $\boldsymbol{\Delta}_n := \{\boldsymbol{x} \in \mathbb{R}^n_+ \,|\, \boldsymbol{x}^\top\mathbf{1}_n = 1\}$. The symbols $\le$ and $\ge$ are used to denote not only the standard inequalities between scalars, but also the componentwise inequalities between vectors.

## 2 Exercises

A sample code is given in some simulations, but readers may make their own codes. When readers try one of exercises, they may use the results of the previous exercises.

### 2.1 Inner-Product of Vectors

**Definition 2.1.1.** The inner-product of vectors is defined as

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \sum_{i=1}^n x_i y_i$$

where $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$.

**Exercise 2.1.2.** (m0201020-dot.tex) Show that $\forall \boldsymbol{x}, \forall \boldsymbol{y} \in \mathbb{R}^n$,

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle = \boldsymbol{x}^\top \boldsymbol{y} = \boldsymbol{y}^\top \boldsymbol{x}.$$

Hint: Substitute Definition 2.1.1 into $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ and $\langle \boldsymbol{y}, \boldsymbol{x} \rangle$.

**Exercise 2.1.3.** (m0201030-dot.tex) For $\forall a \in \mathbb{R}$ and $\forall \boldsymbol{x}, \forall \boldsymbol{y} \in \mathbb{R}^n$, show that

$$\langle a\boldsymbol{x}, \boldsymbol{y} \rangle = a \langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{x}, a\boldsymbol{y} \rangle .$$

Hint: Substitute Definition 2.1.1 into each side.

**Exercise 2.1.4.** (m0201040-dot.tex) For $\forall a, \forall b \in \mathbb{R}$, $\forall \boldsymbol{x}, \forall \boldsymbol{y} \in \mathbb{R}^n$, show that

$$\langle a\boldsymbol{x}, b\boldsymbol{y} \rangle = ab \langle \boldsymbol{x}, \boldsymbol{y} \rangle .$$

Hint: Substitute Definition 2.1.1 into each side.

**Simulation 2.1.5.** (m0201050-dot.tex) Generate a sample and verify Exercise 2.1.4. The following code can be used:

```
n = 3;
a = randn(1);   b = randn(1);
x = randn(n,1); y = randn(n,1);
lhs = dot(a*x,b*y);
rhs = a*b*dot(x,y);
tsassert( norm(lhs-rhs) < 1e-8 );
```

An error message is displayed if lhs is not equal to rhs.

**Exercise 2.1.6.** (m0201060-dot.tex) For $\forall \boldsymbol{x}, \forall \boldsymbol{y}, \forall \boldsymbol{z} \in \mathbb{R}^n$, show that

$$\langle \boldsymbol{x}, \boldsymbol{y} + \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{x}, \boldsymbol{z} \rangle .$$

Hint: Substitute Definition 2.1.1 into each side.

**Simulation 2.1.7.** (m0201070-dot.tex) Generate a sample and verify Exercise 2.1.6. The following code can be used:

```
n = 3;
x = randn(n,1);
y = randn(n,1);
z = randn(n,1);
```

```
lhs = dot(x,y+z);
rhs = dot(x,y)+dot(x,z);
tsassert( norm(lhs-rhs) < 1e-8 );
```

An error message is displayed if lhs is not equal to rhs.

## 2.2 $\ell_2$-Norm

**Definition 2.2.1.** The $\ell_2$-norm of vectors is defined as

$$\|\boldsymbol{x}\| := \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$$

where $\boldsymbol{x} \in \mathbb{R}^n$.

**Exercise 2.2.2.** (m0202020-l2norm.tex) Show that, $\forall \boldsymbol{x} \in \mathbb{R}^n$,

$$\|\boldsymbol{x}\|^2 = \sum_{i=1}^n x_i^2.$$

Hint: Substitute Definition 2.2.1 into lhs, and then use Definition 2.1.1.

**Simulation 2.2.3.** (m0202030-l2norm.tex) Generate a sample and verify Exercise 2.2.2. The following code can be used:

```
n = 3;
x = randn(n,1);
lhs = norm(x).^2;
rhs = sum(x.^2);
tsassert( norm(lhs-rhs) < 1e-8 );
```

An error message is displayed if lhs is not equal to rhs.

**Exercise 2.2.4.** (m0202040-l2norm.tex)[homogeneousness for $\ell_2$-norm] For $\forall a \in \mathbb{R}$, $\forall \boldsymbol{x} \in \mathbb{R}^n$, derive the equality:

$$\|a\boldsymbol{x}\| = \|a\| \, \|\boldsymbol{x}\|$$

where $\|a\|$ denotes the absolute value of $a$.
Hint: Substitute Definition 2.2.1 into both sides.

**Simulation 2.2.5.** (m0202050-l2norm.tex) Generate a sample and verify Exercise 2.2.4. The following code can be used:

```
n = 3;
a = randn(1,1);
x = randn(n,1);
lhs = norm(a*x);
rhs = abs(a)*norm(x);
tsassert( norm(lhs-rhs) < 1e-8 );
```

An error message is displayed if lhs is not equal to rhs.

**Exercise 2.2.6.** (m0202060-l2norm.tex)[definiteness for $\ell_2$-norm] It is known that $\|\boldsymbol{x}\| = 0$ holds if and only if $\boldsymbol{x} = \boldsymbol{0}$. Show this fact.
Hint: You can see, by using the definition of the $\ell_2$-norm,

$$\boldsymbol{x} = \boldsymbol{0} \Longrightarrow \|\boldsymbol{x}\| = 0$$

which proves $\Longleftarrow$.
To show $\Longrightarrow$, assume the existance of a non-zero vector $\boldsymbol{x}$ such that $\|\boldsymbol{x}\| = 0$ to lead to a contradiction. Using the fact that non-zero vectors $\boldsymbol{x}$ satisfy that $\exists k \in \mathbb{N}_n$ such that $x_k \neq 0$, check the norm of such a vector as

$$\|\boldsymbol{x}\| = \sqrt{\sum_{i=1}^n x_i^2} \geq \sqrt{x_k^2} = \|x_k\| > 0.$$

Readers will then notice that this contradicts the assumption of $\|\boldsymbol{x}\| = 0$.

**Simulation 2.2.7.** (m0202070-l2norm.tex) Generate samples and verify Exercise 2.2.6. The following code can be used:

```
ntrys = 10000;   n = 3;
for i_try=1:ntrys
  x = floor(rand(n,1)*3);
  if all(x == 0.0),
    tsassert( norm(x) == 0 );
  else
    tsassert( norm(x) > 0 );
  end
end
```

An error message is displayed if lhs is not equal to rhs.

**Exercise 2.2.8.** (m0202080-l2norm.tex) Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. Derive the equality:

$$\|\boldsymbol{x} + \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 + 2\langle \boldsymbol{x}, \boldsymbol{y} \rangle .$$

Hint: Substitute Definition 2.2.1 into both sides.

**Exercise 2.2.9.** (m0202090-l2norm.tex) Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. Derive the equality:

$$\|\boldsymbol{x} - \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 - 2\langle \boldsymbol{x}, \boldsymbol{y} \rangle .$$

Hint: Substitute Definition 2.2.1 into both sides.

**Exercise 2.2.10.** (m0202100-l2norm.tex)[$\star$,Cauchy-Schwartz inequality] Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. Show that

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle \leq \|\boldsymbol{x}\| \, \|\boldsymbol{y}\| .$$

**Exercise 2.2.11.** (m0202110-l2norm.tex)[$\star$,triangle inequality for $\ell_2$-norm] For $\forall \boldsymbol{x}, \forall \boldsymbol{y} \in \mathbb{R}^n$, show that

$$\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\| .$$

**Simulation 2.2.12.** (m0202120-l2norm.tex) Generate samples and verify Exercise 2.2.11. The following code can be used:

```
ntrys = 10000;   n = 3;
for i_try=1:ntrys
  x = randn(n,1);
  y = randn(n,1);
  tsassert( norm(x+y) <= norm(x)+norm(y) );
end
```

An error message is displayed if lhs is not equal to rhs.

**Simulation 2.2.13.** (m0202130-l2norm.tex) Generate samples and verify Exercise 2.2.10. The following code can be used:

```
ntrys = 10000;   n = 3;
for i_try=1:ntrys
  x = randn(n,1);
  y = randn(n,1);
  tsassert( dot(x,y) <= norm(x)*norm(y) );
end
```

An error message is displayed if lhs is not equal to rhs.

**Exercise 2.2.14.** (m0202140-l2norm.tex) Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ be unit vectors (i.e. $\|\boldsymbol{x}\| = \|\boldsymbol{y}\| = 1$). Show that

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle \leq 1.$$

Hint: Substitute the assumption $\|\boldsymbol{x}\| = \|\boldsymbol{y}\| = 1$ into the result of Exercise 2.2.10.

## 2.3 $\ell_1$-Norm of Vectors

**Definition 2.3.1.** The $\ell_1$-norm of vectors is defined as

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

where $\boldsymbol{x} \in \mathbb{R}^n$.

**Exercise 2.3.2.** (m0203020-l1norm.tex)[homogeneousness for $\ell_1$-norm] Let $a \in \mathbb{R}$ and $\boldsymbol{x} \in \mathbb{R}^n$. Show that

$$\|a\boldsymbol{x}\|_1 = |a| \, \|\boldsymbol{x}\|_1 .$$

Hint: Substitute the definition.

**Exercise 2.3.3.** (m0203030-l1norm.tex)[non-negativeness for $\ell_1$-norm] For $\forall \boldsymbol{x} \in \mathbb{R}^n$, show that

$$\|\boldsymbol{x}\|_1 \geq 0.$$

Hint: Look at the definition of $\ell_1$-norm.

**Exercise 2.3.4.** (m0203040-l1norm.tex)[definiteness for $\ell_1$-norm] It is known that $\|\boldsymbol{x}\|_1 = 0$ is hold if and only if $\boldsymbol{x} = \boldsymbol{0}$. Show this fact.

Hint: You can see, by using the definition of the $\ell_1$-norm,

$$\boldsymbol{x} = \boldsymbol{0} \Longrightarrow \|\boldsymbol{x}\|_1 = 0$$

which proves $\Longrightarrow$. To show $\Longrightarrow$, you may use proof by contradiction. Namely, assume the existence of non-zero vector $\boldsymbol{x}$ such that $\|\boldsymbol{x}\|_1 = 0$, and find a contradictoin.

**Exercise 2.3.5.** (m0203050-l1norm.tex)[$\star$,triangle inequality for $\ell_1$-norm] For $\forall \boldsymbol{x}, \forall \boldsymbol{y} \in \mathbb{R}^n$, show that

$$\|\boldsymbol{x} + \boldsymbol{y}\|_1 \leq \|\boldsymbol{x}\|_1 + \|\boldsymbol{y}\|_1 .$$

Hint: If you substitute the definition in the lhs, you get

$$\text{lhs} = \sum_{i=1}^{n} |x_i + y_i| .$$

Then, use the fact of $\forall x, \forall y \in \mathbb{R}$, $|x + y| \leq |x| + |y|$.

## 2.4 $\ell_\infty$-Norm of Vectors

**Definition 2.4.1.** The $\ell_\infty$-norm of vectors is defined as

$$\|\boldsymbol{x}\|_\infty := \max_{i \in \mathbb{N}_n} |x_i|$$

where $\boldsymbol{x} \in \mathbb{R}^n$.

**Exercise 2.4.2.** (m0204020-linfnorm.tex)[homogeneousness for $\ell_\infty$-norm] Let $a \in \mathbb{R}$ and $\boldsymbol{x} \in \mathbb{R}^n$. Show that

$$\|a\boldsymbol{x}\|_\infty = |a| \, \|\boldsymbol{x}\|_\infty .$$

Hint: Substitute the definition.

**Exercise 2.4.3.** (m0204030-linfnorm.tex)[non-negativeness for $\ell_\infty$-norm] For $\forall \boldsymbol{x} \in \mathbb{R}^n$, show that

$$\|\boldsymbol{x}\|_\infty \geq 0.$$

Hint: Look at the definition of $\ell_\infty$-norm.

**Exercise 2.4.4.** (m0204040-linfnorm.tex)[definiteness for $\ell_\infty$-norm] It is known that $\|\boldsymbol{x}\|_\infty = 0$ is hold if and only if $\boldsymbol{x} = \boldsymbol{0}$. Show this fact.

Hint: Apparently, you can see, by using the definition of the $\ell_\infty$-norm,

$$\boldsymbol{x} = \boldsymbol{0} \Longrightarrow \|\boldsymbol{x}\|_\infty = 0.$$

which proves $\Longleftarrow$.

To show $\Longrightarrow$, you may use proof by contradiction. Namely, assume the existence of non-zero vector $\boldsymbol{x}$ such that $\|\boldsymbol{x}\|_\infty = 0$, and find a contradiction.

**Exercise 2.4.5.** (m0204050-linfnorm.tex)[$\star$,triangle inequality for $\ell_\infty$-norm] For $\forall \boldsymbol{x}, \forall \boldsymbol{y} \in \mathbb{R}^n$, show that

$$\|\boldsymbol{x}+\boldsymbol{y}\|_\infty \leq \|\boldsymbol{x}\|_\infty + \|\boldsymbol{y}\|_\infty .$$

Hint: If you substitute the definition into the lhs, you get

$$\text{lhs} = \max_{i\in\mathbb{N}_n} |x_i + y_i| .$$

Using the fact of $\forall x, \forall y \in \mathbb{R}$, $|x+y| \leq |x| + |y|$, you will see that

$$\max_{i\in\mathbb{N}_n} |x_i + y_i| \leq \max_{i\in\mathbb{N}_n} (|x_i| + |y_i|) .$$

It holds that

$$\max_{i\in\mathbb{N}_n} (|x_i| + |y_i|) \leq \max_{i\in\mathbb{N}_n} |x_i| + \max_{j=1}^{n} |y_j| .$$

Consider why the above inequality holds.

## 2.5 Basic Matrix Identities

**Exercise 2.5.1.** (m0205010-basic.tex)[associativity] Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}, \boldsymbol{B} \in \mathbb{R}^{n\times p}, \boldsymbol{C} \in \mathbb{R}^{p\times q}$. Show that $(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$.

**Exercise 2.5.2.** (m0205020-basic.tex)[left distributivity] Let $\boldsymbol{A} \in \mathbb{R}^{m\times p}, \boldsymbol{B}, \boldsymbol{C} \in \mathbb{R}^{p\times n}$. Show that $\boldsymbol{A}(\boldsymbol{B}+\boldsymbol{C}) = \boldsymbol{AB} + \boldsymbol{AC}$.

**Exercise 2.5.3.** (m0205030-basic.tex) Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$. Denote by $\boldsymbol{I}_m$ the $m\times m$ identity matrix. Show that $\boldsymbol{I}_m \boldsymbol{A} = \boldsymbol{A}$.

**Definition 2.5.4.** The *transpose* of an $m \times n$ matrix $\boldsymbol{A}$, to be denoted by $\boldsymbol{A}^\top$, is the $n \times m$ marix whose $(j, i)$-th entry is $(i, j)$-th entry of $\boldsymbol{A}$. Namely,

$$\left(\begin{bmatrix} A_{1,1} & \dots & A_{1,n} \\ \vdots & \ddots & \vdots \\ A_{m,1} & \dots & A_{m,n} \end{bmatrix}\right)^\top = \begin{bmatrix} A_{1,1} & \dots & A_{1,m} \\ \vdots & \ddots & \vdots \\ A_{n,1} & \dots & A_{n,m} \end{bmatrix}$$

**Exercise 2.5.5.** (m0205050-basic.tex) Let $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m\times n}$. Show that $(\boldsymbol{A}+\boldsymbol{B})^\top = \boldsymbol{A}^\top + \boldsymbol{B}^\top$.

**Exercise 2.5.6.** (m0205060-basic.tex) Let $\boldsymbol{A} \in \mathbb{R}^{m\times k}$ and $\boldsymbol{B} \in \mathbb{R}^{k\times n}$. Show $(\boldsymbol{AB})^\top = \boldsymbol{B}^\top \boldsymbol{A}^\top$. If it is too hard, readers may first assume $\boldsymbol{A}$ is $2 \times 2$, and then extend the proof to a general setting (i.e. $\boldsymbol{A} \in \mathbb{R}^{m\times k}$ and $\boldsymbol{B} \in \mathbb{R}^{k\times n}$).

**Exercise 2.5.7.** (m0205070-basic.tex) Let $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m\times p}, \boldsymbol{C} \in \mathbb{R}^{p\times n}$. Show that $(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{C} = \boldsymbol{AC}+\boldsymbol{BC}$.

**Exercise 2.5.8.** (m0205080-basic.tex) Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$. Denote by $\boldsymbol{I}_n$ the $n \times n$ identity matrix. Show that $\boldsymbol{A}\boldsymbol{I}_n = \boldsymbol{A}$.

**Exercise 2.5.9.** (m0205090-basic.tex) Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{B} \in \mathbb{R}^{n\times p}$, and $\boldsymbol{C} \in \mathbb{R}^{n\times q}$. Show that

$$\boldsymbol{A}[\boldsymbol{B}, \boldsymbol{C}] = [\boldsymbol{AB}, \boldsymbol{AC}] . \tag{2.5.9.1}$$

**Exercise 2.5.10.** (m0205100-basic.tex) Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{B} \in \mathbb{R}^{p\times n}$, and $\boldsymbol{C} \in \mathbb{R}^{n\times q}$. Show that

$$\begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{bmatrix} \boldsymbol{C} = \begin{bmatrix} \boldsymbol{AC} \\ \boldsymbol{BC} \end{bmatrix} . \tag{2.5.10.1}$$

**Exercise 2.5.11.** (m0205110-basic.tex) Let $\boldsymbol{A} \in \mathbb{R}^{m\times p}$, $\boldsymbol{B} \in \mathbb{R}^{m\times q}$, $\boldsymbol{C} \in \mathbb{R}^{p\times n}$, and $\boldsymbol{D} \in \mathbb{R}^{q\times n}$. Show that

$$[\boldsymbol{A}, \boldsymbol{B}] \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{D} \end{bmatrix} = \boldsymbol{AC} + \boldsymbol{BD}. \tag{2.5.11.1}$$

**Exercise 2.5.12.** (m0205120-basic.tex) Let $\boldsymbol{A}^{(1,1)} \in \mathbb{R}^{m_1\times n_1}$, $\boldsymbol{A}^{(1,2)} \in \mathbb{R}^{m_1\times n_2}$,$\boldsymbol{A}^{(2,1)} \in \mathbb{R}^{m_2\times n_1}$,$\boldsymbol{A}^{(2,1)} \in \mathbb{R}^{m_2\times n_1}$,$\boldsymbol{B}^{(1)} \in \mathbb{R}^{n_1\times p}$, and$\boldsymbol{B}^{(2)} \in \mathbb{R}^{n_2\times p}$. Show that

$$\begin{bmatrix} \boldsymbol{A}^{(1,1)} & \boldsymbol{A}^{(1,2)} \\ \boldsymbol{A}^{(2,1)} & \boldsymbol{A}^{(2,2)} \end{bmatrix} \begin{bmatrix} \boldsymbol{B}^{(1)} \\ \boldsymbol{B}^{(2)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}^{(1,1)}\boldsymbol{B}^{(1)} + \boldsymbol{A}^{(1,2)}\boldsymbol{B}^{(2)} \\ \boldsymbol{A}^{(2,1)}\boldsymbol{B}^{(1)} + \boldsymbol{A}^{(2,2)}\boldsymbol{B}^{(2)} \end{bmatrix} . \tag{2.5.12.1}$$

**Exercise 2.5.13.** (m0205130-basic.tex) Let $\boldsymbol{A}^{(1)} \in \mathbb{R}^{m\times n_1}$, $\boldsymbol{A}^{(2)} \in \mathbb{R}^{m\times n_2}$,$\boldsymbol{B}^{(1,1)} \in \mathbb{R}^{n_1\times p_1}$,$\boldsymbol{B}^{(1,2)} \in \mathbb{R}^{n_1\times p_2}$,$\boldsymbol{B}^{(2,1)} \in \mathbb{R}^{n_2\times p_1}$, and$\boldsymbol{B}^{(2,2)} \in \mathbb{R}^{n_2\times p_2}$ . Show that

$$\begin{aligned} &\left[\boldsymbol{A}^{(1)}, \boldsymbol{A}^{(2)}\right] \begin{bmatrix} \boldsymbol{B}^{(1,1)} & \boldsymbol{B}^{(1,2)} \\ \boldsymbol{B}^{(2,1)} & \boldsymbol{B}^{(2,2)} \end{bmatrix} \\ &= \left[\boldsymbol{A}^{(1)}\boldsymbol{B}^{(1,1)} + \boldsymbol{A}^{(2)}\boldsymbol{B}^{(2,1)}, \boldsymbol{A}^{(1)}\boldsymbol{B}^{(1,2)} + \boldsymbol{A}^{(2)}\boldsymbol{B}^{(2,2)}\right] . \end{aligned} \tag{2.5.13.1}$$

**Exercise 2.5.14.** (m0205140-basic.tex) Show that $(\boldsymbol{ABC})^\top = \boldsymbol{C}^\top \boldsymbol{B}^\top \boldsymbol{A}^\top$ where $\boldsymbol{A} \in \mathbb{R}^{m\times n}, \boldsymbol{B} \in \mathbb{R}^{n\times p}, \boldsymbol{C} \in \mathbb{R}^{p\times q}$.

Hint: Use $(\boldsymbol{AB})^\top = \boldsymbol{B}^\top \boldsymbol{A}^\top$.

**Simulation 2.5.15.** (m0205150-basic.tex) Generate a sample and verify $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$ where $\boldsymbol{A}$ is a non-singular square matrix. The following code can be used:

```
n = 5;
A = rand(n,n); iA = inv(A);
tsassert( norm(A*iA-eye(n)) < 1e-8 );
tsassert( norm(iA*A-eye(n)) < 1e-8 );
```

An error message is displayed if lhs is not equal to rhs.

**Definition 2.5.16.** The *inverse* of a square $n \times n$ matrix $\boldsymbol{A}$, to be denoted by $\boldsymbol{A}^{-1}$, is the $n \times n$ marix such that

$$\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_n.$$

The matrix $\boldsymbol{A}$ is said to be *non-singular* if $\boldsymbol{A}^{-1}$ exists.

If the inverse of $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ exists, then it is unique and it holds that $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}_n$.

**Exercise 2.5.17.** (m0205170-basic.tex) Let $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$ be non-singular. Show $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$.
Hint: Derive the equality: $(\boldsymbol{AB})(\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}) = \boldsymbol{I}$.

**Simulation 2.5.18.** (m0205180-basic.tex) Generate two square matrices as a sample and verify $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$ in Exercise 2.5.17. The following code can be used:

```
n = 3;
A = rand(n,n); iA = inv(A);
B = rand(n,n); iB = inv(B);
lhs = inv(A*B);
rhs = iB'*iA';
tsassert( norm(lhs-rhs) < 1e-8 );
```

**Exercise 2.5.19.** (m0205190-basic.tex) Show that $(\boldsymbol{ABC})^{-1} = \boldsymbol{C}^{-1}\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$ where $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C} \in \mathbb{R}^{n\times n}$ are non-singular.
Hint: Use $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$.

**Exercise 2.5.20.** (m0205200-basic.tex) Show $(\boldsymbol{A}^\top)^{-1} = (\boldsymbol{A}^{-1})^\top$ where $\boldsymbol{A}$ is a non-singular square matrix.
Hint: Transpose the both sides of $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}$.

**Simulation 2.5.21.** (m0205210-basic.tex) Generate a non-singular square matrix $\boldsymbol{A}$ as a sample and verify $(\boldsymbol{A}^\top)^{-1} = (\boldsymbol{A}^{-1})^\top$. The following code can be used:

```
n = 3;
A = rand(n,n);
lhs = inv(A');
rhs = inv(A)';
tsassert( norm(lhs-rhs) < 1e-8 );
```

**Exercise 2.5.22.** (m0205220-basic.tex) Show the Woodbury formula

$$\begin{aligned}&(\boldsymbol{A} + \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1}\\ &= \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D} + \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{C}\boldsymbol{A}^{-1}.\end{aligned}$$

where $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, are $\boldsymbol{D}$ are matrices with the correct size.
Hint: Show that $(\boldsymbol{A} + \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})(\boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D} + \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{C}\boldsymbol{A}^{-1}) = \boldsymbol{I}$. If it is still hard to derive this equality, readers may see the solution which is given in Wikipedia currently(Oct 12, 2013).

**Simulation 2.5.23.** (m0205230-basic.tex) Generate a sample and verify the Woodbury formula given in Exercise 2.5.22.

**Exercise 2.5.24.** (m0205240-basic.tex) Let $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n] \in \mathbb{R}^{m\times n}$. Note that $j$-th column of $\boldsymbol{A}$ is $\boldsymbol{a}_j$. Let $k \in \mathbb{N}_n$ Show that

$$\boldsymbol{A}\boldsymbol{e}_k = \boldsymbol{a}_k$$

where $\boldsymbol{e}_k$ is a unit vector with $k$-th entry one and the other entries zero.

**Exercise 2.5.25.** (m0205250-basic.tex) Let $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m]^\top \in \mathbb{R}^{m\times n}$. Note that $i$-th row of $\boldsymbol{A}$ is $\boldsymbol{a}_i^\top$. Show that, $\forall k \in \mathbb{N}_m$,

$$\boldsymbol{e}_k^\top \boldsymbol{A} = \boldsymbol{a}_k^\top$$

where $\boldsymbol{e}_k$ is a unit vector with $k$-th entry one and the other entries zero.

**Exercise 2.5.26.** (m0205260-basic.tex) Let $\boldsymbol{x} \in \mathbb{R}^n$ and $k \in \mathbb{N}_n$. Show that

$$\langle \boldsymbol{e}_k, \boldsymbol{x} \rangle = x_k$$

where $\boldsymbol{e}_k$ is a unit vector with $k$-th entry one and the other entries zero.

**Exercise 2.5.27.** (m0205270-basic.tex) Let $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m]^\top \in \mathbb{R}^{n\times m}$. Note that $i$-th column of $\boldsymbol{A}$ is $\boldsymbol{a}_i$. Let $\boldsymbol{x} \in \mathbb{R}^n$ and $k \in \mathbb{N}_n$. Show that

$$\left\langle \boldsymbol{e}_k, \boldsymbol{A}^\top \boldsymbol{x} \right\rangle = \langle \boldsymbol{a}_k, \boldsymbol{x} \rangle$$

where $\boldsymbol{e}_k$ is a unit vector with $k$-th entry one and the other entries zero.

## 2.6 Statistics

**Exercise 2.6.1.** (m0206010-stat.tex) Given $\ell$ vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_\ell \in \mathbb{R}^n$, the mean vector is defined by

$$\boldsymbol{m} := \frac{1}{\ell}\sum_{i=1}^{\ell} \boldsymbol{x}_i. \tag{2.6.1.1}$$

Show that the $n \times \ell$ matrix $\boldsymbol{X} := [\boldsymbol{x}_1, \dots, \boldsymbol{x}_\ell]$ satisfies

$$\boldsymbol{m} = \frac{1}{\ell}\boldsymbol{X}\boldsymbol{1}_\ell. \tag{2.6.1.2}$$

**Exercise 2.6.2.** (m0206020-stat.tex) Given $\ell$ vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_\ell \in \mathbb{R}^n$, the covariance matrix $\boldsymbol{C} \in \mathbb{S}_+^n$ is defined by

$$\boldsymbol{C} := \frac{1}{\ell}\sum_{i=1}^{\ell}(\boldsymbol{x}_i - \boldsymbol{m})(\boldsymbol{x}_i - \boldsymbol{m})^\top. \tag{2.6.2.1}$$

where

$$\boldsymbol{m} := \frac{1}{\ell}\sum_{i=1}^{\ell}\boldsymbol{x}_i. \tag{2.6.2.2}$$

Show that the $n \times \ell$ matrix $\boldsymbol{X} := [\boldsymbol{x}_1, \dots, \boldsymbol{x}_\ell]$ satisfies

$$\boldsymbol{C} + \boldsymbol{m}\boldsymbol{m}^\top = \frac{1}{\ell}\boldsymbol{X}\boldsymbol{X}^\top. \tag{2.6.2.3}$$

**Exercise 2.6.3.** (m0206030-stat.tex) Let $\boldsymbol{v} := [v_1, \dots, v_\ell]^\top \in \boldsymbol{\Delta}_\ell$ where $\boldsymbol{\Delta}_\ell$ denotes the $\ell$-dimensional probabilistic simplex: $\boldsymbol{\Delta}_\ell := \{\boldsymbol{x} \in \mathbb{R}_+^\ell \,|\, \boldsymbol{x}^\top\boldsymbol{1}_\ell = 1\}$. Given $\ell$ vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_\ell \in \mathbb{R}^n$, the weighted mean vector is defined by

$$\boldsymbol{m} := \sum_{i=1}^{\ell} v_i\boldsymbol{x}_i. \tag{2.6.3.1}$$

Show that the $n \times \ell$ matrix $\boldsymbol{X} := [\boldsymbol{x}_1, \dots, \boldsymbol{x}_\ell]$ satisfies

$$\boldsymbol{m} = \boldsymbol{X}\boldsymbol{v}. \tag{2.6.3.2}$$

## 2.7 Idempotent Matrices

**Exercise 2.7.1.** (m0207010-idem.tex) Show that the $\ell \times \ell$ matrix

$$\boldsymbol{K} := \boldsymbol{I} - \frac{1}{\ell}\boldsymbol{1}_\ell\boldsymbol{1}_\ell^\top$$

satisfies $\boldsymbol{K}^2 = \boldsymbol{K}$.

**Exercise 2.7.2.** (m0207020-idem.tex) Let $\boldsymbol{v} \in \boldsymbol{\Delta}_\ell$ where $\boldsymbol{\Delta}_\ell$ denotes the $\ell$-dimensional probabilistic simplex: $\boldsymbol{\Delta}_\ell := \{\boldsymbol{x} \in \mathbb{R}_+^\ell \,|\, \boldsymbol{x}^\top\boldsymbol{1}_\ell = 1\}$. Show that the $\ell \times \ell$ matrix

$$\boldsymbol{K} := \boldsymbol{I} - \boldsymbol{1}_\ell\boldsymbol{v}^\top$$

satisfies $\boldsymbol{K}^2 = \boldsymbol{K}$.

**Exercise 2.7.3.** (m0207030-idem.tex) Define an $\ell \times \ell$ matrix

$$\boldsymbol{K} = \boldsymbol{I} - \frac{1}{\ell}\boldsymbol{1}_\ell\boldsymbol{1}_\ell^\top. \tag{2.7.3.1}$$

Denote, by $\boldsymbol{m} \in \mathbb{R}^n$ and $\boldsymbol{C} \in \mathbb{S}^n$, the mean vector and the covariance matrix of $\ell$ vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_\ell \in \mathbb{R}^n$, respectively. Show that

$$\boldsymbol{C} = \frac{1}{\ell}\boldsymbol{X}\boldsymbol{K}\boldsymbol{X}^\top \tag{2.7.3.2}$$

where $\boldsymbol{X} := [\boldsymbol{x}_1, \dots, \boldsymbol{x}_\ell]$.

## 2.8 Trace

**Definition 2.8.1.** Let $\boldsymbol{A}$ be an $n \times n$ square matrix. The trace of $\boldsymbol{A}$ is defined as

$$\mathrm{tr}(\boldsymbol{A}) = \sum_{i=1}^{n} A_{i,i}. \tag{2.8.1.1}$$

**Exercise 2.8.2.** (m0208020-trace.tex) For $\forall a \in \mathbb{R}$ and $\forall \boldsymbol{X} \in \mathbb{R}^{n\times n}$, show that

$$\mathrm{tr}(a\boldsymbol{X}) = a\mathrm{tr}(\boldsymbol{X}).$$

Hint: The rhs can be rearranged as

$$a\mathrm{tr}(\boldsymbol{X}) = a\sum_{i=1}^{n} X_{i,i}.$$

On the other hand, we can rearrange the lhs by considering the trace of

$$a\boldsymbol{X} = a\begin{bmatrix} X_{1,1} & \dots & X_{1,n} \\ \vdots & \ddots & \vdots \\ X_{n,1} & \dots & X_{n,n} \end{bmatrix} = \begin{bmatrix} aX_{1,1} & \dots & aX_{1,n} \\ \vdots & \ddots & \vdots \\ aX_{n,1} & \dots & aX_{n,n} \end{bmatrix}.$$

**Exercise 2.8.3.** (m0208030-trace.tex) For $\forall a, \forall b \in \mathbb{R}$ and $\forall \boldsymbol{X}, \forall \boldsymbol{Y} \in \mathbb{R}^{n\times n}$, show that

$$\mathrm{tr}(a\boldsymbol{X} + b\boldsymbol{Y}) = a\mathrm{tr}(\boldsymbol{X}) + b\mathrm{tr}(\boldsymbol{Y}).$$

Hint: Note that

$$\mathrm{tr}(a\boldsymbol{X} + b\boldsymbol{Y}) = \sum_{i=1}^{n}(aX_{i,i} + bY_{i,i}).$$

and

$$\mathrm{tr}(\boldsymbol{X}) = \sum_{i=1}^{n} X_{i,i}, \qquad \mathrm{tr}(\boldsymbol{Y}) = \sum_{i=1}^{n} Y_{i,i}.$$

**Exercise 2.8.4.** (m0208040-trace.tex) For $\forall \boldsymbol{a} \in \mathbb{R}^k$ and $\forall \boldsymbol{X}_1, \dots, \forall \boldsymbol{X}_k \in \mathbb{R}^{n\times n}$, show that

$$\mathrm{tr}(\sum_{i=1}^{k} a_i\boldsymbol{X}_i) = \sum_{i=1}^{k} a_i\mathrm{tr}(\boldsymbol{X}_i).$$

**Exercise 2.8.5.** (m0208050-trace.tex) Show that $\mathrm{tr}(\boldsymbol{AB}) = \mathrm{tr}(\boldsymbol{BA})$ for $\boldsymbol{A} \in \mathbb{R}^{m\times n}, \boldsymbol{B} \in \mathbb{R}^{n\times m}$. Then, show that, $\mathrm{tr}(\boldsymbol{ABC}) = \mathrm{tr}(\boldsymbol{CAB}) = \mathrm{tr}(\boldsymbol{BCA})$ for $\boldsymbol{A} \in \mathbb{R}^{m\times n}, \boldsymbol{B} \in \mathbb{R}^{n\times p}, \boldsymbol{C} \in \mathbb{R}^{p\times m}$.

**Simulation 2.8.6.** (m0208060-trace.tex) Generate samples and verify $\mathrm{tr}(\boldsymbol{AB}) = \mathrm{tr}(\boldsymbol{BA})$ and $\mathrm{tr}(\boldsymbol{ABC}) = \mathrm{tr}(\boldsymbol{CAB}) = \mathrm{tr}(\boldsymbol{BCA})$ in Exercise 2.8.5.

**Exercise 2.8.7.** (m0208070-trace.tex) Let $\boldsymbol{A}^{m\times n}$ such thet $\boldsymbol{A}^\top \boldsymbol{A}$ is non-singular (i.e. the inverse exists). Show that

$$\mathrm{tr}(\boldsymbol{A}(\boldsymbol{A}^\top\boldsymbol{A})^{-1}\boldsymbol{A}^\top) = n.$$

**Exercise 2.8.8.** (m0208080-trace.tex) For $\forall \boldsymbol{x}, \forall \boldsymbol{y} \in \mathbb{R}^n$, show that

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \mathrm{tr}(\boldsymbol{x}\boldsymbol{y}^\top) \tag{2.8.8.1}$$

**Exercise 2.8.9.** (m0208090-trace.tex) Let $\mathbb{R}^n_+$ denote the set of $n$-dimensional nonnegative vectors. For $\forall \boldsymbol{x} \in \mathbb{R}^n_+$, show the equalities:

$$\|\boldsymbol{x}\|_1 = \langle \boldsymbol{x}, \boldsymbol{1}_n \rangle = \mathrm{tr}(\boldsymbol{x}\boldsymbol{1}_n^\top). \tag{2.8.9.1}$$

## 2.9 Determinant

**Simulation 2.9.1.** (m0209010-det.tex) The determinant of a $2 \times 2$ matrix is given in

$$\det\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right) = a_{11}a_{22} - a_{12}a_{21}.$$

Use the function `det` to verify the correctness.

**Simulation 2.9.2.** (m0209010-det.tex) Generate samples and verify a equality $\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$ for symmetric matrices $\boldsymbol{A}$ and $\boldsymbol{B}$.

**Exercise 2.9.3.** (m0209010-det.tex) Show $\det(\boldsymbol{A}^{-1}) = \frac{1}{\det(\boldsymbol{A})}$ for any symmetric matrix $\boldsymbol{A}$. Readers may use (C.2), (C.12) in Bishop book [1] and the fact $\det \boldsymbol{I} = 1$.

**Simulation 2.9.4.** (m0209010-det.tex) Verify the equality $\det \boldsymbol{A} = \det \boldsymbol{A}^\top$ with a square matrix generated randomly.

**Simulation 2.9.5.** (m0209010-det.tex) Generate samples and verify

$$\det(\boldsymbol{I}_N + \boldsymbol{A}\boldsymbol{B}^\top) = \det(\boldsymbol{I}_M + \boldsymbol{A}^\top\boldsymbol{B})$$

and

$$\det(\boldsymbol{I}_n + \boldsymbol{a}\boldsymbol{b}^\top) = 1 + \langle \boldsymbol{a}, \boldsymbol{b} \rangle,$$

where $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{N\times M}$ and $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$.

**Simulation 2.9.6.** (m0209010-det.tex) Let $\boldsymbol{a} \in \mathbb{R}^n$ and $\boldsymbol{A} = \mathrm{diag}(\boldsymbol{a})$. Verify the equality:

$$\det(\boldsymbol{A}) = \prod_{i=1}^n a_i \tag{2.9.6.1}$$

## 2.10 Inner-Product of Matrices

**Definition 2.10.1.** The inner-product of matrices is defined as

$$\langle \boldsymbol{X}, \boldsymbol{Y} \rangle := \sum_{i=1}^m \sum_{j=1}^n X_{i,j} Y_{i,j}$$

where $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{m\times n}$.

**Exercise 2.10.2.** (m0210020-mdot.tex) For $\forall \boldsymbol{X}, \forall \boldsymbol{Y} \in \mathbb{R}^{m\times n}$, derive the equalities:

$$\langle \boldsymbol{X}, \boldsymbol{Y} \rangle = \langle \boldsymbol{Y}, \boldsymbol{X} \rangle = \mathrm{tr}(\boldsymbol{X}^\top \boldsymbol{Y}) = \mathrm{tr}(\boldsymbol{Y}^\top \boldsymbol{X}) \tag{2.10.2.1}$$

**Simulation 2.10.3.** (m0210030-mdot.tex) Generate a sample and verify Exercise 2.10.2. The following code can be used:

```
m = 3; n = 2;
X = rand(m,n); Y = rand(m,n);
side1 = sum(sum(X.*Y));
side2 = sum(sum(Y.*X));
side3 = trace(X'*Y));
side4 = trace(Y'*X));
tsassert( norm(side1-side2) < 1e-8 );
tsassert( norm(side1-side3) < 1e-8 );
tsassert( norm(side1-side4) < 1e-8 );
```

**Exercise 2.10.4.** (m0210040-mdot.tex) For $\forall a \in \mathbb{R}$ and $\forall \boldsymbol{X}, \forall \boldsymbol{Y} \in \mathbb{R}^{m\times n}$, show that

$$\langle a\boldsymbol{X}, \boldsymbol{Y} \rangle = a \langle \boldsymbol{X}, \boldsymbol{Y} \rangle = \langle \boldsymbol{X}, a\boldsymbol{Y} \rangle.$$

**Exercise 2.10.5.** (m0210050-mdot.tex) For $\forall a, \forall b \in \mathbb{R}$, $\forall \boldsymbol{X}, \forall \boldsymbol{Y} \in \mathbb{R}^{m\times n}$, show that

$$\langle a\boldsymbol{X}, b\boldsymbol{Y} \rangle = ab \langle \boldsymbol{X}, \boldsymbol{Y} \rangle.$$

**Exercise 2.10.6.** (m0210060-mdot.tex) For $\forall \boldsymbol{X}, \forall \boldsymbol{Y}, \forall \boldsymbol{Z} \in \mathbb{R}^{m\times n}$, show that

$$\langle \boldsymbol{X}, \boldsymbol{Y}+\boldsymbol{Z}\rangle = \langle \boldsymbol{X}, \boldsymbol{Y}\rangle + \langle \boldsymbol{X}, \boldsymbol{Z}\rangle . \tag{2.10.6.1}$$

**Exercise 2.10.7.** (m0210070-mdot.tex) For $\forall \boldsymbol{X} \in \mathbb{R}^{m\times n}$, show that

$$\langle \boldsymbol{X}, \boldsymbol{E}_{i,j}\rangle = \operatorname{tr}\left(\boldsymbol{X}^\top \boldsymbol{E}_{i,j}\right) = X_{i,j} \tag{2.10.7.1}$$

where $\boldsymbol{E}_{i,j}$ denotes an $m\times n$ matrix in which $(i,j)$th entry is one and all the others are zero.

## 2.11 Frobenius Norm

**Definition 2.11.1.** The Frobenius norm of a matrix $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ is defined as

$$\|a\boldsymbol{X}\|_{\mathrm{F}} := \sqrt{\langle \boldsymbol{X}, \boldsymbol{X}\rangle}.$$

In Matlab, the function `norm(X,'fro')` computes the Frobenius norm.

**Exercise 2.11.2.** (m0211020-fronorm.tex)[homogeneousness for Frobenius norm] Let $a \in \mathbb{R}$ and $\boldsymbol{X} \in \mathbb{R}^{m\times n}$. Show that

$$\|a\boldsymbol{X}\|_{\mathrm{F}} = |a|\,\|\boldsymbol{X}\|_{\mathrm{F}} .$$

**Exercise 2.11.3.** (m0211030-fronorm.tex)[non-negativeness for Frobenius norm] For $\forall \boldsymbol{X} \in \mathbb{R}^{m\times n}$, show that

$$\|\boldsymbol{X}\|_{\mathrm{F}} \geq 0.$$

**Exercise 2.11.4.** (m0211040-fronorm.tex)[definiteness for Frobenius norm] It is known that $\|\boldsymbol{X}\|_{\mathrm{F}} = 0$ holds only if $\boldsymbol{X} = \boldsymbol{O}$. Show this fact.

This property can be used for checking whether a matrix is zero or not.

**Exercise 2.11.5.** (m0211060-fronorm.tex) Let $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{m\times n}$. Derive the equality:

$$\|\boldsymbol{X}+\boldsymbol{Y}\|_{\mathrm{F}}^2 = \|\boldsymbol{X}\|_{\mathrm{F}}^2 + \|\boldsymbol{Y}\|_{\mathrm{F}}^2 + 2\langle \boldsymbol{X}, \boldsymbol{Y}\rangle . \tag{2.11.5.1}$$

**Exercise 2.11.6.** (m0211070-fronorm.tex) Let $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{m\times n}$. Derive the equality:

$$\|\boldsymbol{X}-\boldsymbol{Y}\|_{\mathrm{F}}^2 = \|\boldsymbol{X}\|_{\mathrm{F}}^2 + \|\boldsymbol{Y}\|_{\mathrm{F}}^2 - 2\langle \boldsymbol{X}, \boldsymbol{Y}\rangle . \tag{2.11.6.1}$$

**Exercise 2.11.7.** (m0211080-fronorm.tex)[$\star$,triangle inequality for Frobenius norm] For $\forall \boldsymbol{X}, \forall \boldsymbol{Y} \in \mathbb{R}^{m\times n}$, show that

$$\|\boldsymbol{X}+\boldsymbol{Y}\|_{\mathrm{F}} \leq \|\boldsymbol{X}\|_{\mathrm{F}} + \|\boldsymbol{Y}\|_{\mathrm{F}} .$$

## 2.12 Derivatives of Functions of Vectors

**Simulation 2.12.1.** (m0212010-vecsca.tex) The derivative of a function $f : \mathbb{R} \to \mathbb{R}$ is defined by

$$\nabla_x f(x) := \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}.$$

By using this definition, verify the following four equations with randomly generated samples:

$$\begin{aligned} \nabla_x (x^n) &= nx^{n-1}, \\ \nabla_x \exp(ax) &= a\exp(ax), \\ \nabla_x \sin(x) &= \cos(x), \\ \nabla_x \cos(x) &= -\sin(x). \end{aligned} \tag{2.12.1.1}$$

An example of the code verifying $\nabla_x(\sin(x)) = \cos(x)$ is given below:

```
x   = rand(1,10);

% A small value
h   = 1e-7;

% Analytical derivative.
df1 = cos(x);

% Numerical derivative.
df2 = (sin(x+h)-sin(x))./h;

tsassert( norm(df1-df2) < 1e-4 );
```

This code checks the differences between the analytical gradient and the numerical gradient at ten points chosen randomly. The differences may not be zero, but must be sufficiently small.

The derivative of a function $f : \mathbb{R}^n \to \mathbb{R}$ is denoted by

$$\nabla_{\boldsymbol{x}} f(\boldsymbol{x}) := \left[\nabla_{x_1} f(\boldsymbol{x}), \dots, \nabla_{x_n} f(\boldsymbol{x})\right]^\top .$$

**Exercise 2.12.2.** (m0212020-vecsca.tex) If we define a function $f : \mathbb{R}^2 \to \mathbb{R}$ as $f(\boldsymbol{x}) = 3x_1 - x_2$, show that the derivative is given by

$$\nabla_{\boldsymbol{x}} f(\boldsymbol{x}) = [3, -1]^\top .$$

Hint:

$$\text{lhs} = \nabla_{\boldsymbol{x}} f(\boldsymbol{x}) = \begin{bmatrix} \nabla_{x_1} f(\boldsymbol{x}) \\ \nabla_{x_2} f(\boldsymbol{x}) \end{bmatrix} = \begin{bmatrix} \nabla_{x_1}(3x_1 - x_2) \\ \nabla_{x_2}(3x_1 - x_2) \end{bmatrix} .$$

**Exercise 2.12.3.** (m0212030-vecsca.tex) Let a function $f : \mathbb{R}^2 \to \mathbb{R}$ be defined as

$$f(\boldsymbol{x}) := 3x_1^2 - x_2$$

Show that the derivative is given by

$$\nabla_{\boldsymbol{x}} f(\boldsymbol{x}) = [6x_1, -1]^\top . \tag{2.12.3.1}$$

Hint:

$$\text{lhs} = \nabla_{\boldsymbol{x}} f(\boldsymbol{x}) = \begin{bmatrix} \nabla_{x_1} f(\boldsymbol{x}) \\ \nabla_{x_2} f(\boldsymbol{x}) \end{bmatrix} = \begin{bmatrix} \nabla_{x_1}(3x_1^2 - x_2) \\ \nabla_{x_2}(3x_1^2 - x_2) \end{bmatrix}.$$

**Exercise 2.12.4.** (m0212040-vecsca.tex) Let a function $f : \mathbb{R}^n \to \mathbb{R}$ be defined as

$$f(\boldsymbol{x}) := \sum_{i=1}^{n} \sin(x_i).$$

Show that the derivative is given by

$$\nabla_{\boldsymbol{x}} f(\boldsymbol{x}) = [\cos(x_1), \dots, \cos(x_n)]^\top . \tag{2.12.4.1}$$

Hint:

$$\begin{aligned} \nabla_{\boldsymbol{x}} f(\boldsymbol{x}) &= \begin{bmatrix} \nabla_{x_1} f(\boldsymbol{x}) \\ \vdots \\ \nabla_{x_n} f(\boldsymbol{x}) \end{bmatrix} \\ &= \begin{bmatrix} \nabla_{x_1} \left(\sum_{i=1}^n \sin(x_i)\right) \\ \vdots \\ \nabla_{x_n} \left(\sum_{i=1}^n \sin(x_i)\right) \end{bmatrix} = \begin{bmatrix} \nabla_{x_1} \sin(x_1) \\ \vdots \\ \nabla_{x_n} \sin(x_n) \end{bmatrix}. \end{aligned}$$

**Simulation 2.12.5.** (m0212050-vecsca.tex) Generate a sample and verify (**??**). Note that $\forall i \in \mathbb{N}_n$,

$$\nabla_{x_i} f(\boldsymbol{x}) = \lim_{h \to 0} \frac{f(\boldsymbol{x} + h\boldsymbol{e}_i) - f(\boldsymbol{x})}{h}.$$

where $\boldsymbol{e}_i$ is a unit vector with $i$-th entry one and the other entries zero. The following code is an example to verify the derivative:

```
n = 3;
h = 1e-7;
x = randn(n,1);

% Analytic gradient
df1 = cos(x);

% Compute a numrical gradient
df2 = zeros(n,1);
for i=1:n
  e_i    = zeros(n,1); e_i(i) = 1;
  df2_i  = (sum(sin(x+h*e_i))-sum(sin(x)))/h;
  df2(i) = df2_i;
end

tsassert( norm(df1-df2)/n < 1e-4 );
```

**Exercise 2.12.6.** (m0212060-vecsca.tex) Show $\nabla_{\boldsymbol{x}} \langle \boldsymbol{a}, \boldsymbol{x} \rangle = \boldsymbol{a}$.

Hint:

$$\begin{aligned} \nabla_{\boldsymbol{x}} \langle \boldsymbol{a}, \boldsymbol{x} \rangle &= \begin{bmatrix} \nabla_{x_1} \langle \boldsymbol{a}, \boldsymbol{x} \rangle \\ \vdots \\ \nabla_{x_n} \langle \boldsymbol{a}, \boldsymbol{x} \rangle \end{bmatrix} \\ &= \begin{bmatrix} \nabla_{x_1} \left(\sum_{i=1}^n a_i x_i\right) \\ \vdots \\ \nabla_{x_n} \left(\sum_{i=1}^n a_i x_i\right) \end{bmatrix} = \begin{bmatrix} \nabla_{x_1} a_1 x_1 \\ \vdots \\ \nabla_{x_n} a_n x_n \end{bmatrix}. \end{aligned}$$

**Exercise 2.12.7.** (m0212070-vecsca.tex) Generate a sample and verify $\nabla_{\boldsymbol{x}} \langle \boldsymbol{a}, \boldsymbol{x} \rangle = \boldsymbol{a}$ in Exercise 2.12.6.

**Exercise 2.12.8.** (m0212080-vecsca.tex) Let $\boldsymbol{x} \in \mathbb{R}^n$. Show that

$$\nabla_{\boldsymbol{x}} \|\boldsymbol{x}\|^2 = 2\boldsymbol{x}.$$

Hint:

$$\begin{aligned} \nabla_{\boldsymbol{x}} \|\boldsymbol{x}\|^2 &= \begin{bmatrix} \nabla_{x_1} \|\boldsymbol{x}\|^2 \\ \vdots \\ \nabla_{x_n} \|\boldsymbol{x}\|^2 \end{bmatrix} \\ &= \begin{bmatrix} \nabla_{x_1} \left(\sum_{i=1}^n x_i^2\right) \\ \vdots \\ \nabla_{x_n} \left(\sum_{i=1}^n x_i^2\right) \end{bmatrix} = \begin{bmatrix} \nabla_{x_1} x_1^2 \\ \vdots \\ \nabla_{x_n} x_n^2 \end{bmatrix}. \end{aligned}$$

**Exercise 2.12.9.** (m0212090-vecsca.tex) Let $\boldsymbol{x}, \boldsymbol{b} \in \mathbb{R}^n$. Show that

$$\nabla_{\boldsymbol{x}} \|\boldsymbol{x} + \boldsymbol{b}\|^2 = 2(\boldsymbol{x} + \boldsymbol{b}).$$

Hint:

$$\begin{aligned} \nabla_{\boldsymbol{x}} \|\boldsymbol{x}\|^2 &= \begin{bmatrix} \nabla_{x_1} \|\boldsymbol{x} + \boldsymbol{b}\|^2 \\ \vdots \\ \nabla_{x_n} \|\boldsymbol{x} + \boldsymbol{b}\|^2 \end{bmatrix} \\ &= \begin{bmatrix} \nabla_{x_1} \left(\sum_{i=1}^n (x_i + b_i)^2\right) \\ \vdots \\ \nabla_{x_n} \left(\sum_{i=1}^n (x_i + b_i)^2\right) \end{bmatrix} = \begin{bmatrix} \nabla_{x_1} (x_1 + b_1)^2 \\ \vdots \\ \nabla_{x_n} (x_n + b_n)^2 \end{bmatrix}. \end{aligned}$$

**Exercise 2.12.10.** (m0212100-vecsca.tex) Let $a \in \mathbb{R}$ and $\boldsymbol{x}, \boldsymbol{b} \in \mathbb{R}^n$. Show that

$$\nabla_{\boldsymbol{x}} \|a\boldsymbol{x} + \boldsymbol{b}\|^2 = 2a(a\boldsymbol{x} + \boldsymbol{b}).$$

**Exercise 2.12.11.** (m0212110-vecsca.tex) Let $\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$. Show that

$$\nabla_{\boldsymbol{x}} \langle \boldsymbol{a}, \boldsymbol{x} + \boldsymbol{b} \rangle = \boldsymbol{a}.$$

**Exercise 2.12.12.** (m0212120-vecsca.tex) Let $\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$. Show that

$$\nabla_{\boldsymbol{x}} \langle \boldsymbol{x} + \boldsymbol{a}, \boldsymbol{x} + \boldsymbol{b} \rangle = 2\boldsymbol{x} + \boldsymbol{a} + \boldsymbol{b}.$$

**Exercise 2.12.13.** (m0212130-vecsca.tex) Let $\boldsymbol{x}, \boldsymbol{a} \in \mathbb{R}^n$ . Show that

$$\nabla_{\boldsymbol{x}} \langle \boldsymbol{a}, \boldsymbol{x} \rangle^2 = 2 \langle \boldsymbol{a}, \boldsymbol{x} \rangle \boldsymbol{a}.$$

**Exercise 2.12.14.** (m0212140-vecsca.tex) Let $\boldsymbol{x}, \boldsymbol{a}_1, \dots, \boldsymbol{a}_m \in \mathbb{R}^n$ . Show that

$$\nabla_{\boldsymbol{x}} \sum_{i=1}^{m} \langle \boldsymbol{a}_i, \boldsymbol{x} \rangle^2 = 2 \sum_{i=1}^{m} \langle \boldsymbol{a}_i, \boldsymbol{x} \rangle \boldsymbol{a}_i.$$

**Exercise 2.12.15.** (m0212150-vecsca.tex) Let $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. Show that

$$\nabla_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x}\|^2 = 2\boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x}.$$

Hint: Notice that the result of Exercise 2.12.14 can be used by letting

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1^\top \\ \vdots \\ \boldsymbol{a}_m^\top \end{bmatrix}.$$

**Exercise 2.12.16.** (m0212160-vecsca.tex) Let $\boldsymbol{x}, \boldsymbol{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Show that

$$\nabla_{\boldsymbol{x}} (\langle \boldsymbol{a}, \boldsymbol{x} \rangle + b)^2 = 2(\langle \boldsymbol{a}, \boldsymbol{x} \rangle + b)\boldsymbol{a}.$$

**Exercise 2.12.17.** (m0212170-vecsca.tex) Let $\boldsymbol{x}, \boldsymbol{a}_1, \dots, \boldsymbol{a}_m \in \mathbb{R}^n$ and $\boldsymbol{b} = [b_1, \dots, b_m]^\top \in \mathbb{R}^m$. Show that

$$\nabla_{\boldsymbol{x}} \left( \sum_{i=1}^{m} (\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle + b_i)^2 \right) = 2 \sum_{i=1}^{m} (\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle + b_i)\boldsymbol{a}_i.$$

**Exercise 2.12.18.** (m0212180-vecsca.tex) Let $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{b} \in \mathbb{R}^m$ and $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. Show that

$$\nabla_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}\|^2 = 2\boldsymbol{A}^\top (\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}).$$

Hint: Notice that the result of Exercise 2.12.14 can be used by letting

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1^\top \\ \vdots \\ \boldsymbol{a}_m^\top \end{bmatrix}.$$

**Exercise 2.12.19.** (m0212190-vecsca.tex) Let $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{A} \in \mathbb{S}^n$ where $\mathbb{S}^n$ denotes the set of symmetric $n \times n$ matrices. Show that

$$\nabla_{\boldsymbol{x}} (\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}) = 2\boldsymbol{A}\boldsymbol{x}.$$

**Exercise 2.12.20.** (m0212200-vecsca.tex) Let $\boldsymbol{x}, \boldsymbol{b} \in \mathbb{R}^n$ and $\boldsymbol{A} \in \mathbb{S}^n$ where $\mathbb{S}^n$ denotes the set of symmetric $n \times n$ matrices. Show that

$$\nabla_{\boldsymbol{x}} ((\boldsymbol{x} + \boldsymbol{b})^\top \boldsymbol{A}(\boldsymbol{x} + \boldsymbol{b})) = 2\boldsymbol{A}(\boldsymbol{x} + \boldsymbol{b}).$$

**Exercise 2.12.21.** (m0212210-vecsca.tex) Let $\boldsymbol{x}, \boldsymbol{a} \in \mathbb{R}^n$ and $\boldsymbol{B} \in \mathbb{R}^{n \times n}$. Show that

$$\nabla_{\boldsymbol{x}} (\operatorname{tr} \left( \boldsymbol{x}\boldsymbol{a}^\top + \boldsymbol{B} \right)) = \boldsymbol{a}.$$

**Exercise 2.12.22.** (m0212220-vecsca.tex) Let $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{A} \in \mathbb{S}^n$ where $\mathbb{S}^n$ denotes the set of $n \times n$ symmetric matrices. Show that

$$\nabla_{\boldsymbol{x}} \operatorname{tr}(\boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^\top) = 2\boldsymbol{A}\boldsymbol{x}.$$

**Exercise 2.12.23.** (m0212230-vecsca.tex) Let $\boldsymbol{x}, \boldsymbol{b} \in \mathbb{R}^n$ and $\boldsymbol{A} \in \mathbb{S}^n$ where $\mathbb{S}^n$ denotes the set of $n \times n$ symmetric matrices. Show that

$$\nabla_{\boldsymbol{x}} \operatorname{tr}(\boldsymbol{A}(\boldsymbol{x} + \boldsymbol{b})(\boldsymbol{x} + \boldsymbol{b})^\top) = 2\boldsymbol{A}(\boldsymbol{x} + \boldsymbol{b}).$$

**Exercise 2.12.24.** (m0212240-vecsca.tex) Let $k \in \mathbb{N}_n$. Define a function $f : \mathbb{R}^n \to \mathbb{R}$ as $f(\boldsymbol{x}) := \langle \boldsymbol{e}_k, \boldsymbol{x} \rangle$. Show that $\forall i \in \mathbb{N}_n$

$$\nabla_{x_i} (f(\boldsymbol{x})) = \delta_{i,k}$$

where $\delta_{i,k}$ is the Kronecker delta.

## 2.13 Derivatives of Vector-Valued Functions

**Exercise 2.13.1.** (m0213010-scavec.tex) Let $\boldsymbol{a} = [a_1, a_2]^\top, \boldsymbol{b} = [b_1, b_2]^\top \in \mathbb{R}^2$. Let a vector-valued function $\boldsymbol{f} : \mathbb{R} \to \mathbb{R}^2$ be defined as

$$\boldsymbol{f}(x) = [\sin(b_1 x), \sin(b_2 x)]^\top .$$

Show that

$$\nabla_x \langle \boldsymbol{a}, \boldsymbol{f}(x) \rangle = a_1 b_1 \cos(b_1 x) + a_2 b_2 \cos(b_2 x).$$

**Exercise 2.13.2.** (m0213020-scavec.tex) Let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$. Let a vector-valued function $\boldsymbol{f} : \mathbb{R} \to \mathbb{R}^n$ be defined as

$$\boldsymbol{f}(x) = [\sin(b_1 x), \sin(b_2 x), \ldots, \sin(b_n x)]^\top.$$

Show that

$$\nabla_x \langle \boldsymbol{a}, \boldsymbol{f}(x) \rangle = \sum_{i=1}^n a_i b_i \cos(b_i x).$$

**Exercise 2.13.3.** (m0213030-scavec.tex) The derivative of a vector-valued function $\boldsymbol{f} : \mathbb{R} \to \mathbb{R}^n$ is denoted by

$$\nabla_x \boldsymbol{f}(x) := [\nabla_x f_1(x), \ldots, \nabla_x f_n(x)]^\top.$$

Let $\boldsymbol{a} \in \mathbb{R}^n$. Show that

$$\nabla_x \langle \boldsymbol{a}, \boldsymbol{f}(x) \rangle = \langle \boldsymbol{a}, \nabla_x \boldsymbol{f}(x) \rangle.$$

**Exercise 2.13.4.** (m0213040-scavec.tex) Let $\boldsymbol{f} : \mathbb{R} \to \mathbb{R}^n$ and $\boldsymbol{g} : \mathbb{R} \to \mathbb{R}^n$. Show that

$$\nabla_x \langle \boldsymbol{f}(x), \boldsymbol{g}(x) \rangle = \langle \nabla_x \boldsymbol{f}(x), \boldsymbol{g}(x) \rangle + \langle \boldsymbol{f}(x), \nabla_x \boldsymbol{g}(x) \rangle. \tag{2.13.4.1}$$

**Simulation 2.13.5.** (m0213050-scavec.tex) Use the following functions as an example:

$$\boldsymbol{f}(x) = [\sin(a_1 x), \sin(a_2 x)]^\top,$$
$$\boldsymbol{g}(x) = [\sin(b_1 x), \sin(b_2 x)]^\top,$$

to numerically verify (2.13.4.1).

**Exercise 2.13.6.** (m0213060-scavec.tex) Let $\boldsymbol{f} : \mathbb{R} \to \mathbb{R}^n$ and $\boldsymbol{g} : \mathbb{R} \to \mathbb{R}^n$. Show that

$$\nabla_x (\boldsymbol{f}(x) + \boldsymbol{g}(x)) = \nabla_x \boldsymbol{f}(x) + \nabla_x \boldsymbol{g}(x).$$

**Exercise 2.13.7.** (m0213070-scavec.tex) Let $a, b \in \mathbb{R}$ and $\boldsymbol{f}, \boldsymbol{g} : \mathbb{R} \to \mathbb{R}^n$. Show that

$$\nabla_x (a\boldsymbol{f}(x) + b\boldsymbol{g}(x)) = a\nabla_x \boldsymbol{f}(x) + b\nabla_x \boldsymbol{g}(x).$$

**Exercise 2.13.8.** (m0213080-scavec.tex) Let $\boldsymbol{f} : \mathbb{R} \to \mathbb{R}^n$ and $\boldsymbol{A} \in \mathbb{R}^{m\times n}$. Define a vector-valued function $\boldsymbol{g} : \mathbb{R} \to \mathbb{R}^m$ as $\boldsymbol{g}(x) = \boldsymbol{A}\boldsymbol{f}(x)$. Show that

$$\nabla_x (\boldsymbol{g}(x)) = \boldsymbol{A}\nabla_x \boldsymbol{f}(x). \tag{2.13.8.1}$$

**Simulation 2.13.9.** (m0213090-scavec.tex) Verify (2.13.8.1) numerically. The following code uses a definition $\boldsymbol{f}(x) := [\sin(a_1 x), \ldots, \sin(a_n x)]^\top$ is used as an example to verify (2.13.8.1):

```
m = 4; n = 3; h = 1e-5;
A = randn(m,n); a = randn(n,1);
x = randn(1,1);

% Analytical gradient
df1 = A*cos(a*x);

% Numerical gradient
df2 = (A*sin(a*(x+h))-A*sin(a*x))/h;

tsassert( norm(df1-df2)/m < 1e-4 );
```

**Exercise 2.13.10.** (m0213100-scavec.tex) Let $\boldsymbol{f} : \mathbb{R} \to \mathbb{R}^n$, $\boldsymbol{b} \in \mathbb{R}^m$ and $\boldsymbol{A} \in \mathbb{R}^{m\times n}$. Define a vector-valued function $\boldsymbol{h} : \mathbb{R} \to \mathbb{R}^m$ as $\boldsymbol{h}(x) := \boldsymbol{A}\boldsymbol{f}(x) + \boldsymbol{b}$. Show that

$$\nabla_x (\boldsymbol{h}(x)) = \boldsymbol{A}\nabla_x \boldsymbol{f}(x).$$

**Exercise 2.13.11.** (m0213110-scavec.tex) Let $\boldsymbol{f} : \mathbb{R} \to \mathbb{R}^n$, $\boldsymbol{g} : \mathbb{R} \to \mathbb{R}^m$ and $\boldsymbol{A} \in \mathbb{R}^{m\times n}$. Define a vector-valued function $\boldsymbol{h} : \mathbb{R} \to \mathbb{R}^m$ as $\boldsymbol{h}(x) := \boldsymbol{A}\boldsymbol{f}(x) + \boldsymbol{g}(x)$. Show that

$$\nabla_x (\boldsymbol{h}(x)) = \boldsymbol{A}\nabla_x \boldsymbol{f}(x) + \nabla_x \boldsymbol{g}(x).$$

**Exercise 2.13.12.** (m0213120-scavec.tex) Let $\boldsymbol{x} \in \mathbb{R}^n$. Show that, $\forall i \in \mathbb{N}_n$,

$$\nabla_{x_i} (\boldsymbol{x}) = \boldsymbol{e}_i$$

where $\boldsymbol{e}_i$ is a unit vector with $i$-th entry one and the other entries zero.

## 2.14 Derivatives of Functions of Matrices

**Exercise 2.14.1.** (m0214010-matsca.tex) Let $\boldsymbol{X}, \boldsymbol{A} \in \mathbb{R}^{m\times n}$. Show that, $\forall i \in \mathbb{N}_m$, $\forall j \in \mathbb{N}_n$,

$$\nabla_{X_{i,j}} (\mathrm{tr}(\boldsymbol{A}^\top \boldsymbol{X})) = A_{i,j}.$$

We denote the derivative of a function $f : \mathbb{R}^{m\times n} \to \mathbb{R}$ by

$$\nabla_{\boldsymbol{X}} f(\boldsymbol{X}) := \begin{bmatrix} \nabla_{X_{1,1}}(f(\boldsymbol{X})) & \ldots & \nabla_{X_{1,n}}(f(\boldsymbol{X})) \\ \vdots & \ddots & \vdots \\ \nabla_{X_{m,1}}(f(\boldsymbol{X})) & \ldots & \nabla_{X_{m,n}}(f(\boldsymbol{X})) \end{bmatrix}.$$

**Exercise 2.14.2.** (m0214020-matsca.tex) Let $\boldsymbol{X}, \boldsymbol{A} \in \mathbb{R}^{m\times n}$. Show that

$$\nabla_{\boldsymbol{X}} (\mathrm{tr}(\boldsymbol{A}^\top \boldsymbol{X})) = \boldsymbol{A}.$$

**Exercise 2.14.3.** (m0214030-matsca.tex) Let $\boldsymbol{X}, \boldsymbol{A} \in \mathbb{R}^{m\times n}$. Show that

$$\nabla_{\boldsymbol{X}}(\mathrm{tr}(\boldsymbol{X}\boldsymbol{A}^\top)) = \boldsymbol{A}.$$

**Exercise 2.14.4.** (m0214040-matsca.tex) Let $\boldsymbol{X}, \boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m\times n}$. Show that

$$\nabla_{\boldsymbol{X}}(\mathrm{tr}(\boldsymbol{A}^\top(\boldsymbol{X}+\boldsymbol{B}))) = \boldsymbol{A}.$$

**Exercise 2.14.5.** (m0214050-matsca.tex) Let $\boldsymbol{X}, \boldsymbol{A} \in \mathbb{R}^{n\times n}$. Show that

$$\nabla_{\boldsymbol{X}}(\mathrm{tr}(\boldsymbol{X})) = \boldsymbol{I}_n.$$

**Exercise 2.14.6.** (m0214060-matsca.tex) Let $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{A} \in \mathbb{S}^n$. Show that

$$\nabla_{\boldsymbol{X}}(\mathrm{tr}(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^\top)) = 2\boldsymbol{X}\boldsymbol{A}.$$

**Exercise 2.14.7.** (m0214070-matsca.tex) Let $\boldsymbol{X}, \boldsymbol{A} \in \mathbb{R}^{m\times n}$. Show that

$$\nabla_{\boldsymbol{X}}(\langle \boldsymbol{A}, \boldsymbol{X}\rangle) = \boldsymbol{A}.$$

**Exercise 2.14.8.** (m0214080-matsca.tex) Let $\boldsymbol{X}, \boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m\times n}$. Show that

$$\nabla_{\boldsymbol{X}}(\langle \boldsymbol{A}, \boldsymbol{X}+\boldsymbol{B}\rangle) = \boldsymbol{A}.$$

**Exercise 2.14.9.** (m0214090-matsca.tex) Let $\boldsymbol{X} \in \mathbb{R}^{m\times n}$. Show that

$$\nabla_{\boldsymbol{X}}(\|\boldsymbol{X}\|_{\mathrm{F}}^2) = 2\boldsymbol{X}.$$

**Exercise 2.14.10.** (m0214100-matsca.tex) Let $\boldsymbol{X}, \boldsymbol{B} \in \mathbb{R}^{m\times n}$. Show that

$$\nabla_{\boldsymbol{X}}(\|\boldsymbol{X}+\boldsymbol{B}\|_{\mathrm{F}}^2) = 2(\boldsymbol{X}+\boldsymbol{B}).$$

**Exercise 2.14.11.** (m0214110-matsca.tex) Let $a \in \mathbb{R}$ and $\boldsymbol{X}, \boldsymbol{B} \in \mathbb{R}^{m\times n}$. Show that

$$\nabla_{\boldsymbol{X}}(\|a\boldsymbol{X}+\boldsymbol{B}\|_{\mathrm{F}}^2) = 2a(a\boldsymbol{X}+\boldsymbol{B}).$$

**Exercise 2.14.12.** (m0214120-matsca.tex) Let $\boldsymbol{X}, \boldsymbol{B} \in \mathbb{R}^{m\times n}$. Show that

$$\nabla_{\boldsymbol{X}}(\langle \boldsymbol{X}, \boldsymbol{X}+\boldsymbol{B}\rangle) = 2\boldsymbol{X}+\boldsymbol{B}.$$

**Exercise 2.14.13.** (m0214130-matsca.tex) Let $\boldsymbol{X}, \boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m\times n}$. Show that

$$\nabla_{\boldsymbol{X}}(\langle \boldsymbol{X}+\boldsymbol{A}, \boldsymbol{X}+\boldsymbol{B}\rangle) = 2\boldsymbol{X}+\boldsymbol{A}+\boldsymbol{B}.$$

## 2.15 Derivatives of Matrix-Valued Functions

Consdier a matrix-valued function $\boldsymbol{F} : \mathbb{R} \to \mathbb{R}^{m\times n}$ expressed as

$$\boldsymbol{F}(x) := \begin{bmatrix} F_{1,1}(x) & \dots & F_{1,n}(x) \\ \vdots & \ddots & \vdots \\ F_{m,1}(x) & \dots & F_{m,n}(x) \end{bmatrix}.$$

The derivative of the function is denoted by

$$\nabla_x \boldsymbol{F}(x) := \begin{bmatrix} \nabla_x F_{1,1}(x) & \dots & \nabla_x F_{1,n}(x) \\ \vdots & \ddots & \vdots \\ \nabla_x F_{m,1}(x) & \dots & \nabla_x F_{m,n}(x) \end{bmatrix}.$$

**Exercise 2.15.1.** (m0215010-scamat.tex) Let $\boldsymbol{a} = [a_1, a_2]^\top, \boldsymbol{b} = [b_1, b_2]^\top \in \mathbb{R}^2$. Define a matrix-valued function $\boldsymbol{F} : \mathbb{R} \to \mathbb{R}^{2\times 2}$ expressed as

$$\boldsymbol{F}(x) := \begin{bmatrix} \sin(a_1 b_1 x) & \sin(a_1 b_2 x) \\ \sin(a_2 b_1 x) & \sin(a_2 b_2 x) \end{bmatrix}.$$

Show that the derivative is expressed as

$$\nabla_x \boldsymbol{F}(x) := \begin{bmatrix} a_1 b_1 \cos(a_1 b_1 x) & a_1 b_2 \cos(a_1 b_2 x) \\ a_2 b_1 \cos(a_2 b_1 x) & a_2 b_2 \cos(a_2 b_2 x) \end{bmatrix}.$$

**Simulation 2.15.2.** (m0215020-scamat.tex) Verify Exercise 2.15.1 numerically. After completing Subsection 2.20, readers may understarnd the following code veryfying Exercise 2.15.1:

```
a = randn(2,1); b = randn(2,1);
x = randn(1,1); h = 1e-5;

% Analytical gradient
df1 = sin(a*b'*x);

% Numerical gradient
term1 = (a*b').*cos(a*b'*(x+h));
term2 = (a*b').*cos(a*b'*x);
df2   = (term1-term2)/h;

tsassert( norm(df1-df2,'fro') < 1e-4 );
```

**Exercise 2.15.3.** (m0215030-scamat.tex) Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{F} : \mathbb{R} \to \mathbb{R}^{m\times n}$. Show that

$$\nabla_x \langle \boldsymbol{A}, \boldsymbol{F}(x)\rangle = \langle \boldsymbol{A}, \nabla_x \boldsymbol{F}(x)\rangle$$

**Exercise 2.15.4.** (m0215040-scamat.tex) Let $\boldsymbol{A} \in \mathbb{R}^{k\times m}$ and $\boldsymbol{F} : \mathbb{R} \to \mathbb{R}^{k\times n}$. Show that

$$\nabla_x(\boldsymbol{A}^\top \boldsymbol{F}(x)) = \boldsymbol{A}^\top \nabla_x(\boldsymbol{F}(x)).$$

**Exercise 2.15.5.** (m0215050-scamat.tex) Let $\boldsymbol{F} : \mathbb{R} \to \mathbb{R}^{m\times k}$ and $\boldsymbol{G} : \mathbb{R} \to \mathbb{R}^{k\times n}$. Show that

$$\nabla_x(\boldsymbol{F}(x)\boldsymbol{G}(x)) = \nabla_x(\boldsymbol{F}(x))\boldsymbol{G}(x) + \boldsymbol{F}(x)\nabla_x(\boldsymbol{G}(x)).$$

## 2.16 Diagonal Matrices

**Definition 2.16.1.** A *diagonal matrix* is a square matrix whose all the off-diagonal entries are zero. Such a matrix with diagonal entries $\boldsymbol{d} := [d_1, \dots, d_n]^\top \in \mathbb{R}^n$ is denoted by

$$\operatorname{diag}(\boldsymbol{d}) = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{bmatrix}. \tag{2.16.1.1}$$

**Exercise 2.16.2.** (m0216020-diag.tex) Let $p \in \mathbb{N}$ and $\boldsymbol{x} \in \mathbb{R}^n$. Show that

$$(\operatorname{diag}(\boldsymbol{x}))^p = \operatorname{diag}([x_1^p, \dots, x_n^p]).$$

Hint: Use induction. First assume

$$(\operatorname{diag}(\boldsymbol{x}))^{p-1} = \operatorname{diag}([x_1^{p-1}, \dots, x_n^{p-1}]),$$

to get

$$\begin{aligned} \text{lhs} &= (\operatorname{diag}(\boldsymbol{x}))^p = (\operatorname{diag}(\boldsymbol{x}))^{p-1} \operatorname{diag}(\boldsymbol{x}) \\ &= \operatorname{diag}([x_1^{p-1}, \dots, x_n^{p-1}]) \operatorname{diag}([x_1, \dots, x_n]) \\ &= \begin{bmatrix} x_1^{p-1} & & \boldsymbol{O} \\ & \ddots & \\ \boldsymbol{O} & & x_n^{p-1} \end{bmatrix} \begin{bmatrix} x_1 & & \boldsymbol{O} \\ & \ddots & \\ \boldsymbol{O} & & x_n \end{bmatrix}. \end{aligned}$$

**Exercise 2.16.3.** (m0216030-diag.tex) Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. Show that

$$\operatorname{diag}(\boldsymbol{x}) \operatorname{diag}(\boldsymbol{y}) = \operatorname{diag}(\boldsymbol{y}) \operatorname{diag}(\boldsymbol{x}).$$

**Exercise 2.16.4.** (m0216040-diag.tex) Let $\boldsymbol{x} \in \mathbb{R}^n$. Show that

$$\operatorname{diag}(\boldsymbol{x})\mathbf{1}_n = \boldsymbol{x}$$

## 2.17 Orthonormal Matrices

**Definition 2.17.1.** An $m \times n$ matrix $\boldsymbol{P}$ is said to be *orthonormal* if the matrix satisfies $\boldsymbol{P}^\top \boldsymbol{P} = \boldsymbol{I}_n$. Symbol $\mathbb{O}^{m\times n}$ is used to denote the set of $m \times n$ orthonormal matrices.

**Exercise 2.17.2.** (m0217020-ortho.tex) Let $\boldsymbol{P} = [\boldsymbol{p}_1, \dots, \boldsymbol{p}_n] \in \mathbb{O}^{m\times n}$. Show that $\forall i, \forall j \in \mathbb{N}_n$,

$$\langle \boldsymbol{p}_i, \boldsymbol{p}_j\rangle = \delta_{i,j}$$

where $\delta_{i,j}$ is the Kronecker delta.

Hint: Express the $(i, j)$-th entry of both sides of $\boldsymbol{P}^\top \boldsymbol{P} = \boldsymbol{I}$.

**Exercise 2.17.3.** (m0217030-ortho.tex) Let $\boldsymbol{P} = [\boldsymbol{p}_1, \dots, \boldsymbol{p}_n] \in \mathbb{O}^{m\times n}$. Show that $\forall k \in \mathbb{N}_n$,

$$\boldsymbol{P}^\top \boldsymbol{p}_k = \boldsymbol{e}_k \tag{2.17.3.1}$$

where $\boldsymbol{e}_k$ is a unit vector with $k$-th entry one and the other entries zero.

Hint: Observe

$$\text{LHS} = \boldsymbol{P}^\top \boldsymbol{p}_k = \begin{bmatrix} \boldsymbol{p}_1^\top \\ \vdots \\ \boldsymbol{p}_n^\top \end{bmatrix} \boldsymbol{p}_k = \begin{bmatrix} \langle \boldsymbol{p}_1, \boldsymbol{p}_k\rangle \\ \vdots \\ \langle \boldsymbol{p}_n, \boldsymbol{p}_k\rangle \end{bmatrix}$$

and substitute the result of Exercise 2.17.2.

**Exercise 2.17.4.** (m0217040-ortho.tex) Let $\boldsymbol{P} \in \mathbb{O}^{n\times n}$. Show that

$$\boldsymbol{P}^\top = \boldsymbol{P}^{-1}. \tag{2.17.4.1}$$

Hint: Right-multiply, by $\boldsymbol{P}^{-1}$, the both sides of $\boldsymbol{P}^\top \boldsymbol{P} = \boldsymbol{I}$.

**Exercise 2.17.5.** (m0217050-ortho.tex) Let $\boldsymbol{P} \in \mathbb{O}^{n\times n}$. Show that

$$\boldsymbol{P}\boldsymbol{P}^\top = \boldsymbol{I}_n. \tag{2.17.5.1}$$

Hint: Substitute the result of Exercise 2.17.4 into LHS.

**Exercise 2.17.6.** (m0217060-ortho.tex) Let $\boldsymbol{P} \in \mathbb{O}^{n\times n}$. Show that

$$\det \boldsymbol{P} \in \{\pm 1\}.$$

Hint: From the definition $\boldsymbol{P}^\top \boldsymbol{P} = \boldsymbol{I}$, we have

$$\begin{aligned} 1 &= \det(\boldsymbol{I}) = \det(\boldsymbol{P}^\top \boldsymbol{P}) = \det(\boldsymbol{P}^\top) \det(\boldsymbol{P}) \\ &= (\det(\boldsymbol{P}))^2. \end{aligned}$$

Consider why each equality of this follows.

**Exercise 2.17.7.** (m0217070-ortho.tex) Let $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{P} \in \mathbb{O}^{n\times n}$. Show that

$$\|\boldsymbol{P}\boldsymbol{x}\| = \|\boldsymbol{x}\|.$$

Hint:

$$\|\boldsymbol{P}\boldsymbol{x}\|^2 = \langle \boldsymbol{P}\boldsymbol{x}, \boldsymbol{P}\boldsymbol{x}\rangle = \mathrm{tr}(\boldsymbol{x}^\top \boldsymbol{P}^\top \boldsymbol{P}\boldsymbol{x}) = \mathrm{tr}(\boldsymbol{x}^\top \boldsymbol{x}) = \|\boldsymbol{x}\|^2.$$

Consider why each equality of this follows.

**Exercise 2.17.8.** (m0217080-ortho.tex) Let $\boldsymbol{P} \in \mathbb{O}^{m\times k}$ and $\boldsymbol{Q} \in \mathbb{O}^{k\times n}$ where $m \geq k \geq n$. Show that $\boldsymbol{P}\boldsymbol{Q} \in \mathbb{O}^{m\times n}$.

## 2.18 Spectral Decomposition

**Theorem 2.18.1.** For any $\boldsymbol{A} \in \mathbb{S}^n$, there exist $\boldsymbol{U} \in \mathbb{O}^{n\times n}$ and $\boldsymbol{\lambda} \in \mathbb{R}^n$ such that

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$$

where $\boldsymbol{\Lambda} := \mathrm{diag}(\boldsymbol{\lambda})$. This is said to be the *spectral decomposition.* In this note, we assume $\lambda_1 \geq \cdots \geq \lambda_n$ without loss of generality. Each entry of $\boldsymbol{\lambda}$ is called an *eigenvalue*, and each column of $\boldsymbol{U}$ is called an *eigenvector*.

**Simulation 2.18.2.** (m0218020-eig.tex) Generate two matrices $\boldsymbol{U}$ and $\boldsymbol{\Lambda}$ by the following code:

```
n = 5;

% Generate a symmetric matrix.
A = randn(n); A = 0.5*(A+A');

% Compute the spectral decomposition of A.
[U,Lam] = eig(A);
[tmp1,r_srt] = sort(-diag(Lam));
U = U(:,r_srt); Lam = Lam(r_srt,r_srt);
```

Then, confirm that $\boldsymbol{U}$ is orthonormal, $\boldsymbol{\Lambda}$ is diagonal, and $\boldsymbol{A}$ is reconstructed by $\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$.

**Exercise 2.18.3.** (m0218030-eig.tex) Denote the spectral decomposition of $\boldsymbol{A} \in \mathbb{S}^n$ by $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$. Let $\lambda_i$ be the $i$-th diagonal entry of $\boldsymbol{\Lambda}$ and $\boldsymbol{u}_i$ be the $i$-th column in $\boldsymbol{U}$. Then, show the equality

$$\boldsymbol{A} = \sum_{i=1}^n \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^\top. \tag{2.18.3.1}$$

**Exercise 2.18.4.** (m0218040-eig.tex) Denote the spectral decomposition of $\boldsymbol{A} \in \mathbb{S}^n$ by $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$. Let $\lambda_i$ be the $i$-th diagonal entry of $\boldsymbol{\Lambda}$ and $\boldsymbol{u}_i$ be the $i$-th column in $\boldsymbol{U}$. Then, show that $\forall i \in \mathbb{N}_n$,

$$\boldsymbol{u}_i^\top \boldsymbol{A}\boldsymbol{u}_i = \lambda_i.$$

Hint: Substitute the spectral decomposition into lhs as

$$\mathrm{LHS} = \boldsymbol{u}_i^\top \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top \boldsymbol{u}_i = \boldsymbol{e}_i^\top \boldsymbol{\Lambda}\boldsymbol{e}_i \tag{2.18.4.1}$$

Consider why the last equality follows.

**Exercise 2.18.5.** (m0218050-eig.tex) Denote the spectral decomposition of $\boldsymbol{A} \in \mathbb{S}^n$ by $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$. Let $\lambda_i$ be the $i$-th diagonal entry of $\boldsymbol{\Lambda}$ and $\boldsymbol{u}_i$ be the $i$-th column in $\boldsymbol{U}$. Then, show that $\forall i, \forall j \in \mathbb{N}_n$,

$$\boldsymbol{u}_i^\top \boldsymbol{A}\boldsymbol{u}_j = \lambda_i \delta_{i,j} \tag{2.18.5.1}$$

where $\delta_{i,j}$ is the Kronecker delta.

Hint: Substitute the spectral decomposition into lhs as

$$\begin{aligned}\mathrm{lhs} &= \boldsymbol{u}_i^\top \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top \boldsymbol{u}_j = \boldsymbol{e}_i^\top \boldsymbol{\Lambda}\boldsymbol{e}_j = \mathrm{tr}(\boldsymbol{\Lambda}\boldsymbol{e}_j\boldsymbol{e}_i^\top) \\ &= \mathrm{tr}(\boldsymbol{\Lambda}\boldsymbol{E}_{i,j}^\top) = \langle \boldsymbol{\Lambda}, \boldsymbol{E}_{i,j}\rangle .\end{aligned} \tag{2.18.5.2}$$

Consider why each equality follows.

**Exercise 2.18.6.** (m0218060-eig.tex) Denote the spectral decomposition of $\boldsymbol{A} \in \mathbb{S}^n$ by $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$. Show that

$$\mathrm{tr}\boldsymbol{A} = \mathrm{tr}\boldsymbol{\Lambda}. \tag{2.18.6.1}$$

Hint: Consider why each equality of the following equation follows:

$$\mathrm{tr}(\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top) = \mathrm{tr}(\boldsymbol{\Lambda}\boldsymbol{U}^\top\boldsymbol{U}).$$

Then, recall that $\boldsymbol{U}$ is orthonormal, and apply the definition of orthonormal matrices to the above equation.

**Exercise 2.18.7.** (m0218070-eig.tex) Denote the spectral decomposition of $\boldsymbol{A} \in \mathbb{S}^n$ by $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$. Let $\lambda_i$ be the $i$-th diagonal entry of $\boldsymbol{\Lambda}$. Show that

$$\det\boldsymbol{A} = \prod_{i=1}^n \lambda_i.$$

Hint: Readers may use (2.9.6.1). It is hold that

$$\det(\boldsymbol{A}) = \det(\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top) = \det(\boldsymbol{U})\det(\boldsymbol{\Lambda})\det(\boldsymbol{U}^\top)$$

Consider why each equality of this follows.

**Definition 2.18.8.** Let $\boldsymbol{A} \in \mathbb{S}^n$. Denote the spectral decomposition of $\boldsymbol{A} \in \mathbb{S}^n$ by $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$. Let $\lambda_i$ be the $i$-th diagonal entry of $\boldsymbol{\Lambda}$ The matrix power of $\boldsymbol{A}$ by $p \in \mathbb{R}$ is defined as

$$\boldsymbol{A}^p := \boldsymbol{U}\boldsymbol{\Lambda}^p\boldsymbol{U}^\top$$

where

$$\boldsymbol{\Lambda}^p = \mathrm{diag}\left([\lambda_1^p, \ldots, \lambda_n^p]\right).$$

**Exercise 2.18.9.** (m0218090-eig.tex) Denote the spectral decomposition of $\boldsymbol{A} \in \mathbb{S}^n$ by $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$. Let $\lambda_i$ be the $i$-th diagonal entry of $\boldsymbol{\Lambda}$ For $\forall p \in \mathbb{R}$, show that

$$\mathrm{tr}(\boldsymbol{A}^p) = \sum_{i=1}^n \lambda_i^p.$$

**Exercise 2.18.10.** (m0218100-eig.tex) Denote the spectral decomposition of $\boldsymbol{A} \in \mathbb{S}^n$ by $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$. Let $\lambda_i$ be the $i$-th diagonal entry of $\boldsymbol{\Lambda}$ For $\forall p \in \mathbb{R}$, show that

$$\det(\boldsymbol{A}^p) = \prod_{i=1}^n \lambda_i^p.$$

## 2.19 Positive Definite Matrices

**Exercise 2.19.1.** Do nothing.

We use $\mathbb{S}_+^n$ to denote the set of symmetric positive semi-definite $n \times n$ matrices, and $\mathbb{S}_{++}^n$ to denote the set of symmetric strictly positive definite $n \times n$ matrices.

**Definition 2.19.2.** An $n \times n$ symmetric matrix is said to be *positive semi-definite* if $\forall \boldsymbol{x} \in \mathbb{R}^n$,

$$\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} \geq 0.$$

**Definition 2.19.3.** An $n \times n$ symmetric matrix is said to be *strictly positive definite* if $\forall \boldsymbol{x} \in \mathbb{R}^n$,

$$\boldsymbol{x} \neq \boldsymbol{0}_n \qquad \Longrightarrow \qquad \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} > 0.$$

**Exercise 2.19.4.** (m0219040-pd.tex) Show that, for any $\boldsymbol{A} \in \mathbb{S}_{++}^n$ and $\boldsymbol{x} \in \mathbb{R}^n$, show that

$$\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} > 0 \qquad \Longrightarrow \qquad \boldsymbol{x} \neq \boldsymbol{0}_n. \tag{2.19.4.1}$$

Hint: Set $\boldsymbol{x} = \boldsymbol{0}_n$, and then check if $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} > 0$.

**Exercise 2.19.5.** (m0219050-pd.tex) Show that, for $\forall \boldsymbol{A} \in \mathbb{S}_+^n$, every eigen-value is non-negative.
Hint: Use the result of Exercise 2.18.4.

**Exercise 2.19.6.** (m0219060-pd.tex) For $\forall \boldsymbol{A} \in \mathbb{S}_{++}^n$, every eigen-value is positive.
Hint: Use the result of Exercise 2.18.4.

**Exercise 2.19.7.** (m0219063-pd.tex) Let $\boldsymbol{A}$ be an $n \times n$ symmetric matrix Denote its spectral decomposition by $\boldsymbol{A} = \boldsymbol{U}\mathrm{diag}(\boldsymbol{\lambda})\boldsymbol{U}^\top$, where $\boldsymbol{U} \in \mathbb{O}^{n\times n}$ and $\boldsymbol{\lambda} \in \mathbb{R}^n$. Show that the matrix $\boldsymbol{A}$ is positive semi-definite if $\boldsymbol{\lambda} \geq \boldsymbol{0}$.
Hint: To derive the inequality $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} \geq 0$ for an arbitrary vector $\boldsymbol{x} \in \mathbb{R}^n$, use the result of Exercise 2.18.3 as

$$\begin{aligned}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} &= \boldsymbol{x}^\top \left(\sum_{i=1}^n \lambda_i \boldsymbol{u}_i\boldsymbol{u}_i^\top\right)\boldsymbol{x} \\ &= \sum_{i=1}^n \lambda_i \boldsymbol{x}^\top \boldsymbol{u}_i\boldsymbol{u}_i^\top \boldsymbol{x} = \sum_{i=1}^n \lambda_i \langle \boldsymbol{x}, \boldsymbol{u}_i\rangle^2 .\end{aligned}$$

**Exercise 2.19.8.** (m0219067-pd.tex) Let $\boldsymbol{A}$ be an $n \times n$ symmetric matrix Denote its spectral decomposition by $\boldsymbol{A} = \boldsymbol{U}\mathrm{diag}(\boldsymbol{\lambda})\boldsymbol{U}^\top$, where $\boldsymbol{U} \in \mathbb{O}^{n\times n}$ and $\boldsymbol{\lambda} \in \mathbb{R}^n$. Show that the matrix $\boldsymbol{A}$ is strictly positive definite if $\boldsymbol{\lambda} > \boldsymbol{0}$.
Hint: The proof can be done by deriving the inequality $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} > 0$ for an arbitrary non-zero $n$-dimensional vector $\boldsymbol{x}$. This can be proved in the following steps:

- Show $\boldsymbol{U}^\top \in \mathbb{O}^{n\times n}$,
- Denote by $\boldsymbol{u}_i$ $i$-th column in $\boldsymbol{U}$, and let $\boldsymbol{z} := \boldsymbol{U}^\top \boldsymbol{x}$. Then, express the $i$-th entry in $\boldsymbol{z}$ as $z_i = \langle \boldsymbol{u}_i, \boldsymbol{x}\rangle$.
- Show $\|\boldsymbol{z}\| > 0$,
- Show $\boldsymbol{z} \neq \boldsymbol{0}$, which means one of $n$ entries in $\boldsymbol{z}$ is non-zero.
- The rest of the proof is similar to the answer of Exercise 2.19.7 in which the result of Exercise 2.18.3 is used.

**Exercise 2.19.9.** (m0219070-pd.tex) Show that the determinant of every positive semidefinite matrix is non-negative.
Hint: Combine Exercise 2.19.5 with Exercise 2.18.7.

**Exercise 2.19.10.** (m0219080-pd.tex) The determinant of every strictly positive definite matrix is strictly positive.
Hint: Combine Exercise 2.19.6 with Exercise 2.18.7.

**Exercise 2.19.11.** (m0219090-pd.tex) For any symmetric matrix $\boldsymbol{A} \in \mathbb{S}^n$, show that $\boldsymbol{A}^2$ is positive semi-definite.

Note: This exercise is not made based on any literature, but TK thinks this is true.

**Exercise 2.19.12.** (m0219100-pd.tex) For any symmetric matrix $\boldsymbol{A} \in \mathbb{S}^n$ such that $\det\boldsymbol{A} \neq 0$, show that $\boldsymbol{A}^2$ is strictly positive definite.
Hint: Show the following facts and combine them:

- From the result of Exercise 2.18.7, the assumption $\det\boldsymbol{A} \neq 0$ implies that every eigenvalue is not zero.
- If we denote, by $\lambda_i$, the $i$-th eigenvalue of $\boldsymbol{A}$, the eigenvalues of $\boldsymbol{A}^2$ are given by $\lambda_1^2, \dots, \lambda_n^2$.

Note: This exercise is not made based on any literature, but TK thinks this is true.

## 2.20 Hadamard Product

**Definition 2.20.1.** Let $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{m\times n}$. The *Hadamard product* between $\boldsymbol{X}$ and $\boldsymbol{Y}$, denoted by $\boldsymbol{Z}$, is defined as

$$\boldsymbol{Z} = \boldsymbol{X} \odot \boldsymbol{Y} \quad \Longleftrightarrow \quad \forall i \in \mathbb{N}_m, \forall j \in \mathbb{N}_n, Z_{i,j} = X_{i,j}Y_{i,j}.$$

**Exercise 2.20.2.** (m0220020-hada.tex) Show that the Harmard product is *commutative*.
Hint: See `http://en.wikipedia.org/wiki/Commutative_property` to learn the definition of 'commutative' (Oct 12, 2013).

**Exercise 2.20.3.** (m0220030-hada.tex) Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. Show the equalities:

$$\boldsymbol{x} \odot \boldsymbol{y} = \text{diag}(\boldsymbol{x})\boldsymbol{y} = \text{diag}(\boldsymbol{y})\boldsymbol{x} = \text{diag}(\boldsymbol{x} \odot \boldsymbol{y})\mathbf{1}_n.$$

**Simulation 2.20.4.** (m0220040-hada.tex) Verify the equalities in Exercise 2.20.3 numerically. The operater `.*` corresponds to the operator of Hadamard product $\odot$.

**Exercise 2.20.5.** (m0220050-hada.tex) Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. Show that

$$\langle \mathbf{1}_n, \boldsymbol{x} \odot \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle .$$

**Exercise 2.20.6.** (m0220060-hada.tex) Let $\boldsymbol{x} \in \mathbb{R}^n$. Show that

$$\langle \mathbf{1}_n, \boldsymbol{x} \odot \boldsymbol{x} \rangle = \|\boldsymbol{x}\|^2.$$

**Exercise 2.20.7.** (m0220070-hada.tex) Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n_+$. Show that

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \|\boldsymbol{x} \odot \boldsymbol{y}\|_1.$$

**Exercise 2.20.8.** (m0220080-hada.tex) Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{x} \in \mathbb{R}^n$. Show that

$$\boldsymbol{A}\,\text{diag}(\boldsymbol{x}) = \boldsymbol{A} \odot (\mathbf{1}_m \boldsymbol{x}^\top).$$

**Exercise 2.20.9.** (m0220090-hada.tex) Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{x} \in \mathbb{R}^m$. Show that

$$\text{diag}(\boldsymbol{x})\boldsymbol{A} = (\boldsymbol{x}\mathbf{1}_n^\top) \odot \boldsymbol{A}.$$

**Exercise 2.20.10.** (m0220100-hada.tex) Let $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{m\times n}$. Show that =

$$\langle \boldsymbol{X}, \boldsymbol{Y} \rangle = \mathbf{1}_m^\top (\boldsymbol{X} \odot \boldsymbol{Y}) \mathbf{1}_n = \left\langle \mathbf{1}_m \mathbf{1}_n^\top, \boldsymbol{X} \odot \boldsymbol{Y} \right\rangle .$$

**Exercise 2.20.11.** (m0220110-hada.tex) Let $\boldsymbol{X} \in \mathbb{R}^{m\times n}$. Show that

$$\|\boldsymbol{X}\|_{\mathrm{F}}^2 = \mathbf{1}_m^\top (\boldsymbol{X} \odot \boldsymbol{X}) \mathbf{1}_n = \left\langle \mathbf{1}_m \mathbf{1}_n^\top, \boldsymbol{X} \odot \boldsymbol{X} \right\rangle .$$

## 2.21 Vec Operator

**Definition 2.21.1.** Let $\boldsymbol{A} = [\boldsymbol{a}_1, \dots, \boldsymbol{a}_n] \in \mathbb{R}^{m\times n}$. The vec operator is defined as

$$\text{vec}(\boldsymbol{A}) := \begin{bmatrix} \boldsymbol{a}_1 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix}.$$

In Matlab, $\text{vec}(\boldsymbol{A})$ can be obtained by `A(:)`.

**Exercise 2.21.2.** (m0222020-vec.tex) Let $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{m\times n}$. Show that

$$\langle \boldsymbol{X}, \boldsymbol{Y} \rangle = \langle \text{vec}\boldsymbol{X}, \text{vec}\boldsymbol{Y} \rangle .$$

**Simulation 2.21.3.** (m0222030-vec.tex) Generate a sample and verify Exercise 2.21.2. The following code can be used:

```
m = 3; n = 2;
X = randn(m,n); Y = randn(m,n);
lhs = sum(sum(X.*Y));
rhs = dot(X(:),Y(:));
tsassert( norm(lhs-rhs) < 1e-8 );
```

**Exercise 2.21.4.** (m0222040-vec.tex) Let $\boldsymbol{X} \in \mathbb{R}^{m\times n}$. Show that

$$\|\boldsymbol{X}\|_{\mathrm{F}} = \|\text{vec}\boldsymbol{X}\| .$$

**Simulation 2.21.5.** (m0222050-vec.tex) Generate a sample and verify Exercise 2.21.4. The following code can be used:

```
m = 3; n = 2;
X = randn(m,n);
lhs = norm(X,'fro');
rhs = norm(X(:));
tsassert( norm(lhs-rhs) < 1e-8 );
```

# References

[1] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer Science+Business Media, LLC, New York, USA, 2006.