MATH3027: Optimization 2022

Week 1: Mathematical Preliminaries

Prof. Richard Wilkinson
Please send any comments or mistakes to r.d.wilkinson@nottingham.ac.uk

This week we will briefly recall some of the most important mathematical preliminaries required for this module. We recap basic definitions from linear algebra, some topological concepts such as the definition of closed, open, bounded and compact sets, and the gradient and directional derivatives of functions of multiple variables.

Look out for the penguins in the notes - these are exercises for you to attempt. We will discuss their solution in class.

Algebra

We will need some basic algebraic concepts during the module. All of this should be familiar to you from the first year linear algebra module.

Vector Spaces

The space $\mathbb{R}^n$ is the set of n -dimensional column vectors $\mathbf{x}$ with real components endowed with the component-wise addition operator:

$$\mathbf{x} + \mathbf{y} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix},$$

and the scalar-vector product

$$\lambda \mathbf{x} = \lambda \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{pmatrix}.$$

The vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ denote the standard/canonical basis, i.e.,

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

$\mathbf{1}$ and $\mathbf{0}$ denote all ones and all zeros column vectors, respectively.

Important Subsets of $\mathbb{R}^n$.

- Nonnegative orthant:

$$\mathbb{R}^n_+ = \left\{ (x_1, x_2, \ldots, x_n)^T : x_1, x_2, \ldots, x_n \geq 0 \right\}$$

- Positive orthant:

$$\mathbb{R}^n_{++} = \left\{ (x_1, x_2, \ldots, x_n)^T : x_1, x_2, \ldots, x_n > 0 \right\}.$$

- If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, the closed line segment between x and y is given by

$$[\mathbf{x}, \mathbf{y}] = \{\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x}) : \alpha \in [0, 1]\}$$

  and the open line segment $(\mathbf{x}, \mathbf{y})$ is defined as

$$(\mathbf{x}, \mathbf{y}) = \{\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x}) : \alpha \in (0, 1)\}$$

  for $\mathbf{x} \neq \mathbf{y}$. Note that $(\mathbf{x}, \mathbf{x}) = \emptyset$.

- The unit-simplex:

$$\Delta_n = \left\{ \mathbf{x} \in \mathbb{R}^n : \mathbf{x} \geq \mathbf{0}, \mathbf{1}^T \mathbf{x} = 1 \right\},$$

  i.e., $\mathbf{x}$ such that $\sum_{i=1}^n x_i = 1$ with $x_i \geq 0$ for all i.

The Space $\mathbb{R}^{m \times n}$. The set of all real valued $m \times n$ matrices is a vector space, and is denoted by $\mathbb{R}^{m \times n}$. The $n \times n$ identity matrix is denoted by $\mathbf{I}_n$.

Inner Products and Norms

Vector spaces are not rich enough to interest us in practical settings. We need to equip them with a sense of geometry. We use norms to give us a sense of distance, and inner-products to give us a sense of distance and angle.

Definition (Inner Product)**.** *An inner product on* $\mathbb{R}^n$ *is a map* $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ *with the following properties:*

1. *Symmetry:* $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$ *for any* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.
2. *Additivity:* $\langle \mathbf{x}, \mathbf{y} + \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle$ *for any* $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$.
3. *Homogeneity:* $\langle \lambda\mathbf{x}, \mathbf{y} \rangle = \lambda \langle \mathbf{x}, \mathbf{y} \rangle$ *for any* $\lambda \in \mathbb{R}$ *and* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.
4. *Positive definiteness:* $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ *for any* $\mathbf{x} \in \mathbb{R}^n$ *and* $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ *if and only if* $x = 0$.

An ***inner product space*** *is a vector space equipped with an inner-product.*

Examples:

- The usual "dot product":

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \sum_{i=1}^{n} x_i y_i \text{ for any } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n .$$

- The "weighted dot product":

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{w}} = \sum_{i=1}^{n} w_i x_i y_i$$

  where $\mathbf{w} \in \mathbb{R}^n_{++}$.

Definition (Norms)**.** *A norm* $\|\cdot\|$ *on* $\mathbb{R}^n$ *is a function* $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ *satisfying:*

1. *Nonnegativity:* $\|\mathbf{x}\| \geq 0$ *for any* $\mathbf{x} \in \mathbb{R}^n$ *and* $\|\mathbf{x}\| = 0$ *if and only if* $\mathbf{x} = 0$.
2. *Positive homogeneity:* $\|\lambda\mathbf{x}\| = |\lambda| \|\mathbf{x}\|$ *for any* $\mathbf{x} \in \mathbb{R}^n$ *and* $\lambda \in \mathbb{R}$.
3. *Triangle inequality:* $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ *for any* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

A ***normed space*** *is a vector space equipped with a norm.*

Any inner-product space is also a normed space (the converse is false), as we can generate a norm from the inner product $\langle \cdot, \cdot \rangle$ by setting

$$\|\mathbf{x}\| \equiv \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}, \text{ for all } \mathbf{x} \in \mathbb{R}^n .$$

The norm associated with the dot-product is the so-called Euclidean norm or ℓ_2-norm:

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} \text{ for all } \mathbf{x} \in \mathbb{R}^n .$$

Other useful norms are the ℓ_p-norm and ℓ_∞-norm

$$\|\mathbf{x}\|_p \equiv \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \qquad \|\mathbf{x}\|_\infty \equiv \max_{i=1,2,\ldots,n} |x_i| .$$

It can be shown that

$$\|\mathbf{x}\|_\infty = \lim_{p \to \infty} \|\mathbf{x}\|_p .$$

Theorem (Cauchy-Schwartz Inequality). *For any* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|$$

Proof. For any $\lambda \in \mathbb{R}$:

$$\|\mathbf{x} + \lambda \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2\lambda \langle \mathbf{x}, \mathbf{y} \rangle + \lambda^2 \|\mathbf{y}\|^2 ,$$

leading to (why?)

$$\langle \mathbf{x}, \mathbf{y} \rangle^2 \leq \|\mathbf{x}\|^2 \|\mathbf{y}\|^2 ,$$

establishing the desired result. □

We can also define norms on the space of matrices $\mathbb{R}^{m \times n}$. The most commonly used norm is the Frobenius norm:

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2}, \qquad \mathbf{A} \in \mathbb{R}^{m \times n}$$

Eigenvalues and Eigenvectors

Definition (Eigenvalues and eigenvectors). *Let* $\mathbf{A} \in \mathbb{R}^{n \times n}$*. Then scalar* λ *is an eigenvalue of* $\mathbf{A}$ *if there exists* $\mathbf{v} \in \mathbb{R}^n$ *such that*

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

The vector $\mathbf{v}$ *is called an eigenvector of* $\mathbf{A}$ *corresponding to eigenvalue* λ*.*

In general, real-valued matrices can have complex eigenvalues, but when the matrix is symmetric the eigenvalues are necessarily real. We will denote the eigenvalues of a symmetric $n \times n$ matrix $\mathbf{A}$ as

$$\lambda_1(\mathbf{A}) \geq \lambda_2(\mathbf{A}) \geq \ldots \geq \lambda_n(\mathbf{A}) .$$

The maximum eigenvalue is also denoted by $\lambda_{\max}(\mathbf{A})$ $(= \lambda_1(\mathbf{A}))$ and the minimum eigenvalue is also denoted by $\lambda_{\min}(\mathbf{A})$ $(= \lambda_n(\mathbf{A}))$.

You will have seen the following result, which is one of the most important theorems of linear algebra, proved during the first year linear algebra module.

Theorem (Spectral Factorization Theorem). *Let* $\mathbf{A} \in \mathbb{R}^{n\times n}$ *be an* $n \times n$ *symmetric matrix. Then there exists an orthogonal matrix* $\mathbf{U} \in \mathbb{R}^{n\times n}$ $\left(\mathbf{U}^T\mathbf{U} = \mathbf{U}\mathbf{U}^T = \mathbf{I}_n\right)$ *and a diagonal matrix* $\mathbf{D} = \operatorname{diag}(d_1, d_2, \ldots, d_n)$ *for which*

$$\mathbf{U}^T\mathbf{A}\mathbf{U} = \mathbf{D} \qquad \textit{or equivalently} \qquad \mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^\top .$$

The columns of the matrix $\mathbf{U}$ constitute an orthogonal basis comprising eigenvectors of $\mathbf{A}$ and the diagonal elements of $\mathbf{D}$ are the corresponding eigenvalues.

Topology

Many of the results we prove in the module will rely upon some basic topological concepts, such as open, closed, compact and bounded sets.

The **open ball** with center $c \in \mathbb{R}^n$ and radius r :

$$B(c, r) = \{\mathbf{x} : \|\mathbf{x} - c\| < r\} .$$

The **closed ball** with center c and radius r :

$$B[c, r] = \{\mathbf{x} : \|\mathbf{x} - c\| \leq r\} .$$

Definition (Interior Point). *Given a set* $U \subseteq \mathbb{R}^n$, *a point* $\mathbf{c} \in U$ *is called an interior point of* U *if there exists* $r > 0$ *for which* $B(\mathbf{c}, r) \subseteq U$. *The set of all interior points of a given set* U *is called the interior of the set and is denoted by* $\operatorname{int}(U)$*:*

$$\operatorname{int}(U) = \{\mathbf{x} \in U : B(\mathbf{x}, r) \subseteq U \textit{ for some } r > 0\} .$$

An open set is a set that contains only interior points. Meaning that

$$U = \operatorname{int}(U) .$$

Examples of open sets are open balls (hence the name...) and the positive orthant $\mathbb{R}^n_{++}$. The union of any number of open sets is an open set and the intersection of a finite number of open sets is open.

Closed Sets. A set $U \subseteq \mathbb{R}^n$ is closed if it contains all the limits of convergent sequences of vectors in U, that is, if $\{\mathbf{x}_i\}_{i=1}^{\infty} \subseteq U$ satisfies $\mathbf{x}_i \longrightarrow \mathbf{x}^*$ as $i \longrightarrow \infty$, then $\mathbf{x}^* \in U$. A known result states that U is closed iff its complement U^c is open. Examples of closed sets are the closed ball $B[\mathbf{c}, r]$, closed lines segments, the nonnegative orthant $\mathbb{R}^n_+$ and the unit simplex Δ_n. What about $\mathbb{R}^n$? $\emptyset$?

Boundary Points. Given a set $U \subseteq \mathbb{R}^n$, a boundary point of U is a vector $\mathbf{x} \in \mathbb{R}^n$ satisfying the following: any neighbourhood of $\mathbf{x}$ contains at least one point in U and at least one point in its complement U^c. The set of all boundary points of a set U is denoted by $\mathrm{bd}(U)$.

Definition (Closure). *The closure of a set $U \subseteq \mathbb{R}^n$ is denoted by $\mathrm{cl}(U)$ and is defined to be the smallest closed set containing U :*

$$\mathrm{cl}(U) = \bigcap \{T : U \subseteq T, T \text{ is closed}\}.$$

Another equivalent definition of $\mathrm{cl}(U)$ is:

$$\mathrm{cl}(U) = U \cup \mathrm{bd}(U)$$

Boundedness and Compactness. A set $U \subseteq \mathbb{R}^n$ is called bounded if there exists $M > 0$ for which $U \subseteq B(0, M)$. A set $U \subseteq \mathbb{R}^n$ is called compact if it is closed and bounded. Examples of compact sets are: closed balls, unit simplex, and closed line segments.

Calculus

Definition (Directional Derivative). *Let f be a function defined on a set $S \subseteq \mathbb{R}^n$. Let $\mathbf{x} \in \mathrm{int}(S)$ and let $\mathbf{d} \in \mathbb{R}^n$. If the limit*

$$f'(\mathbf{x}; \mathbf{d}) := \lim_{t \to 0^+} \frac{f(\mathbf{x} + t\mathbf{d}) - f(\mathbf{x})}{t}$$

*exists, then it is called the **directional derivative** of f at x along the direction $\mathbf{d}$ and is denoted by $f'(\mathbf{x}; \mathbf{d})$.*

Definition (Partial Derivative). *For any $i = 1, 2, \ldots, n$, if the limit*

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = \lim_{t \to 0} \frac{f(\mathbf{x} + t\mathbf{e}_i) - f(\mathbf{x})}{t}$$

*exists, then its value is called the i-th **partial derivative** and is denoted by $\frac{\partial f}{\partial x_i}(\mathbf{x})$.*

Definition (Gradient). *If all the partial derivatives of a function f exist at a point $\mathbf{x} \in \mathbb{R}^n$, then the **gradient** of f at $\mathbf{x}$ is*

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}) \\ \frac{\partial f}{\partial x_2}(\mathbf{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{x}) \end{pmatrix}.$$

Definition (Continuous Differentiability). *A function f defined on an open set $U \subseteq \mathbb{R}^n$ is called continuously differentiable over U if all the partial derivatives exist and are continuous on U. In that case,*

$$\boxed{f'(\mathbf{x}; \mathbf{d}) = \nabla f(\mathbf{x})^T \mathbf{d}.}$$

Twice Differentiability and the Hessian

A function f defined on an open set $U \subseteq \mathbb{R}^n$ is called twice continuously differentiable over U if all the second order partial derivatives exist and are continuous over U. In that case, for any $i \neq j$ and any $\mathbf{x} \in U$:

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x})$$

Definition (The Hessian). *The Hessian of f at a point $\mathbf{x} \in U$ is the $n \times n$ matrix:*

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & & \vdots \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

For twice continuously differentiable functions, the Hessian is a symmetric matrix.

Theorem (Linear Approximation Theorem). *Let $f : U \to \mathbb{R}$ be defined on an open set $U \subseteq \mathbb{R}^n$. Suppose that f is twice continuously differentiable over U. Let $\mathbf{x} \in U$ and $r > 0$ satisfy $B(\mathbf{x}, r) \subseteq U$. Then for any $\mathbf{y} \in B(\mathbf{x}, r)$, there exists $\xi \in [\mathbf{x}, \mathbf{y}]$ such that:*

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) + \frac{1}{2} (\mathbf{y} - \mathbf{x})^T \nabla^2 f(\xi) (\mathbf{y} - \mathbf{x}).$$

This is an approximation in the sense that we can write

$$\boxed{f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) + o(||\mathbf{y} - \mathbf{x}||)}.$$

Here, $o(\cdot)$ is the 'little-o' notation: $o(\cdot) : \mathbb{R}_+ \longrightarrow \mathbb{R}$ is a one-dimensional function satisfying $\frac{o(t)}{t} \longrightarrow 0$ as $t \longrightarrow 0^+$. So here, the error in the linear approximation, denoted here as $o(\|\mathbf{x} - \mathbf{y}\|)$, goes to zero faster than $\|\mathbf{x} - \mathbf{y}\|$.

It will sometimes be useful to write this linear apprxomation in an alternative form. For any $\mathbf{h} \in B(\mathbf{0}, r)$, there exists $t \in [0, 1]$ such that

$$f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) = \nabla f(\mathbf{x})^\top \mathbf{h} + \frac{1}{2}\mathbf{h}^\top \nabla^2 f(\mathbf{x} + t\mathbf{h})\mathbf{h}.$$

Theorem (Quadratic Approximation Theorem). *Let $f : U \longrightarrow \mathbb{R}$ be defined on an open set $U \subseteq \mathbb{R}^n$. Suppose that f is twice continuously differentiable over U. Let $\mathbf{x} \in U$ and $r > 0$ satisfy $B(\mathbf{x}, r) \subseteq U$. Then for any $\mathbf{y} \in B(\mathbf{x}, r)$:*

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) + \frac{1}{2}(\mathbf{y} - \mathbf{x})^T \nabla^2 f(\mathbf{x})(\mathbf{y} - \mathbf{x}) + o\left(\|\mathbf{y} - \mathbf{x}\|^2\right)$$

where again, $o(\|\mathbf{x} - \mathbf{y}\|^2)$ is an error term that goes to zero faster than $\|\mathbf{x} - \mathbf{y}\|^2$.

Finally, a reminder that the fundamental theorem of calculus says that for a differentiable function $f(x)$, we have that

$$\int_a^b \frac{\mathrm{d}}{\mathrm{d}x} f(x) dx = f(b) - f(a).$$

Gradient of Linear and Quadratic Functions

There are two results that we will use repeatedly during the module:

$$\begin{aligned} f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x} \quad &\Rightarrow \quad \nabla f(\mathbf{x}) = \mathbf{a} \\ f(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x} \quad &\Rightarrow \quad \nabla f(\mathbf{x}) = (A + A^\top)\mathbf{x} \quad \Rightarrow \quad \nabla^2 f(\mathbf{x}) = A + A^\top \end{aligned}$$

We can prove these results in a labourious manner by writing out the linear and quadratic forms as sums. However, using the linear approximation theorem

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^\top \mathbf{h} + o(||\mathbf{h}||)$$

gives us a more way elegant to derive the results. The second term in the linear approximation theorem is $\nabla f(\mathbf{x})^\top \mathbf{h} = \langle \nabla f(\mathbf{x}), \mathbf{h} \rangle = f'(\mathbf{x}; \mathbf{h})$, which is the directional derivative of f at $\mathbf{x}$ in the direction $\mathbf{h}$. Thus, if for a given f we can write

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + \langle \mathbf{g}, \mathbf{h} \rangle + o(||\mathbf{h}||)$$

as $\mathbf{h} \longrightarrow 0$, then we can conclude that $\mathbf{g}$ is the gradient of f at $\mathbf{x}$.

Lets first consider the linear function

$$f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x}$$

If we consider $f(\mathbf{x}+\mathbf{h})$ we can see that we have

$$
\begin{aligned}
f(\mathbf{x}+\mathbf{h}) &= \mathbf{a}^\top(\mathbf{x}+\mathbf{h}) \\
&= \mathbf{a}^\top\mathbf{x}+\mathbf{a}^\top\mathbf{h} \\
&= f(\mathbf{x}) + \langle \mathbf{a}, \mathbf{h}\rangle + o(||\mathbf{h}||)
\end{aligned}
$$

where we have used the fact that 0 is $o(\mathbf{h})$. Thus the derivative is $\nabla f(\mathbf{x}) = \mathbf{a}$.

Now consider the quadratic form

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

for an $n \times n$ matrix $\mathbf{A}$. Then

$$
\begin{aligned}
f(\mathbf{x}+\mathbf{h}) &= (\mathbf{x}+\mathbf{h})^\top\mathbf{A}(\mathbf{x}+\mathbf{h}) \\
&= f(\mathbf{x}) + \mathbf{h}^\top\mathbf{A}\mathbf{x} + \mathbf{x}^\top\mathbf{A}\mathbf{h} + \mathbf{h}^\top\mathbf{A}\mathbf{h} \\
&= f(\mathbf{x}) + \langle \mathbf{A}\mathbf{x}+\mathbf{A}^\top\mathbf{x}, \mathbf{h}\rangle + o(||\mathbf{h}||)
\end{aligned}
$$

and so $\nabla_{\mathbf{w}} f(\mathbf{w}) = \mathbf{A}\mathbf{x}+\mathbf{A}^\top\mathbf{x}$.

Checklist

The idea of this checklist is to help you to self-evaluate your progress and understanding of the subject, and to give you some guidance on where to focus. If you can tick all the boxes it means you're doing alright, otherwise you need to study a bit more, grab a book, watch the videos, or seek help from classmates or me (the lecturer). Try to fill as many gaps as quickly as possible.

And remember to do the exercises!

Learning Outcome	Check
I understand what $\mathbb{R}^n$ is, vectors, and subspaces	
I understand the properties defining an inner product	
I can identify a norm and give examples	
I can prove the Cauchy-Schwartz Inequality	
I know the definition of eigenvalue and eigenvector	
What can be said about the eigenvalues of a symmetric matrix?	
I can compute the eigenvalues of a matrix	
I understand the spectral factorization theorem	
What's the definition of an open ball? of a closed ball?	
What is the link between directional derivatives and the gradient?	
I understand what is the Hessian and how to compute it	
I understand the Linear and Quadratic Approximation Theorems	
I can compute the gradient and directional derivatives of a function	
I can write down a linear and quadratic approximation to a function	

Exercises

1. Show that the dot product satisfies the definition of an inner-product.
2. Why is $\ell_{1/2}$ not a norm?
3. Fill in the missing detail near the 'why?' in the proof of the Cauchy-Schwartz algorithm.
4. Prove $||\mathbf{A}||_F$ is a norm on the vector space of matrices?
5. Prove that for real symmetric matrix $\mathbf{A}$ we have $\mathrm{Tr}(\mathbf{A}) = \sum_{i=1}^n \lambda_i(\mathbf{A})$ (the trace of $\mathbf{A}$) and $\det(\mathbf{A}) = \Pi_{i=1}^n \lambda_i(\mathbf{A})$.
6. If $\mathbf{A}$ is a 2×2 symmetric matrix with $\mathrm{Tr}(\mathbf{A}) = 3$ and $\det(\mathbf{A}) = 1$, what are the eigenvalues of $\mathbf{A}$?
7. Some topological questions:

$$\begin{aligned}
\mathrm{int}\left(\mathbb{R}^n_+\right) &=? \\
\mathrm{int}(B[\mathbf{c}, r]) &=? \\
\mathrm{int}([\mathbf{x}, \mathbf{y}]) &=? \\
(\mathbf{c} \in \mathbb{R}^n, r \in \mathbb{R}_{++}), \mathrm{bd}(B(\mathbf{c}, r)) &= \\
(\mathbf{c} \in \mathbb{R}^n, r \in \mathbb{R}_{++}), \mathrm{bd}(B[\mathbf{c}, r]) &= \\
\mathrm{bd}\left(\mathbb{R}^n_{++}\right) &= \\
\mathrm{bd}\left(\mathbb{R}^n_+\right) &= \\
\mathrm{bd}\left(\mathbb{R}^n\right) &= \\
\mathrm{bd}\left(\Delta_n\right) &= \\
\mathrm{cl}\left(R^n_{++}\right) &= \\
(\mathbf{x} \neq \mathbf{y}), \mathrm{cl}((\mathbf{x}, \mathbf{y})) &=
\end{aligned}$$

8. Consider the function $f(\mathbf{x}) = x^2y^3 - 4xz$. What is the gradient of f? What is the directional derivative of f in the direction $(-1, 2, 0)$ at the point $\mathbf{x} = (1, 1, 1)^\top$? Thus find a linear and a quadratic approximation to the function f at the point $(1, 1, 1)$.
9. Find direction $\mathbf{d} \in \mathbb{R}^n$ with $||\mathbf{d}|| = 1$ to minimize $f'(\mathbf{x}; \mathbf{d})$. Hint: use Cauchy-Schwartz! Why might it be useful to know which direction has the steepest descent?
10. Consider the following approximation to $f(\mathbf{x})$

$$\tilde{f}(\mathbf{x}) = f(\mathbf{y}) + \nabla f(\mathbf{y})^\top (\mathbf{x} - \mathbf{y}) + \frac{1}{2h}||\mathbf{x} - \mathbf{y}||^2.$$

where $h > 0$ is a constant. Find the value of $\mathbf{x}$ that minimizes $\tilde{f}(\mathbf{x})$. Given this result, can you think of an iterative approach for minimizing $f(\mathbf{x})$?

Repeat these steps using the quadratic approximation

$$\tilde{f}(\mathbf{x}) = f(\mathbf{y}) + \nabla f(\mathbf{y})^\top (\mathbf{x} - \mathbf{y}) + \frac{1}{2h}(\mathbf{x} - \mathbf{y})^\top \nabla^2 f(\mathbf{y})(\mathbf{x} - \mathbf{y})$$