Summer term 2024

Ramp-up Mathematics — Analysis
Extra Homework Sheet

Here are some more exercises you can do to prepare for the exam.

1. Recall that for a square matrix $M \in \mathbb{R}^{n\times n}$ it holds that $\mathrm{Tr}(M) := \sum_{k=i}^{n} M_{kk}$ and that the inner product for matrices $A, B \in \mathbb{R}^{m\times n}$ is defined by $\langle A, B\rangle := \mathrm{Tr}(AB^T)$.
   Prove or disprove that for this inner product it holds that $\langle A, B\rangle = \langle A^T, B^T\rangle$
2. Define $F : \mathbb{R}^{m\times n} \to \mathbb{R}$ by $F(R) = \mathrm{Tr}(RR^T)$.
   (a) Calculate the derivative $DF(R)[H]$! (You can equip $\mathbb{R}^{m\times n}$ with any of your favorite norms. Recall that all norms on finite-dimensional vector spaces are equivalent to each other.)
   (b) What is the gradient $\nabla F(R) \in \mathbb{R}^{m\times n}$?
3. Let $C = \{x \in \mathbb{R}^n \mid x_i > 0 \forall i\}$ and define $d : C \times C \to [0, \infty[$ by
   $$d(x, y) := \sum_{i=1}^{n} \frac{(x_i - y_i)^2}{x_i y_i}.$$
   Which of the axioms of a metric are fulfilled and which are not?
4. Let
   $$A = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \vdots \\ 1 & \cdots & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{n\times n}.$$
   Calculate $\|A\|_{1\to 1}$ (i.e. the operator norm of A when $\mathbb{R}^n$ is equipped with the 1-norm).
5. Let $A \in \mathbb{R}^{n\times n}$ and define $F : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times m}$ by $F(R) = RAR^T$. Calculate the derivative $DF(R)[H]$!
6. Give a definition of a norm $\|\cdot\|_{\text{mynorm}}$ on $\mathbb{R}^2$ which fulfills $\|\binom{1}{0}\|_{\text{mynorm}} = 10$ and $\|\binom{0}{1}\|_{\text{mynorm}} = 1$.
7. Let $F : \mathbb{R}^n \to \mathbb{R}$ be defined by $F(x) = \sqrt{\langle a, x\rangle}$ for $a \in \mathbb{R}^n$. Calculate $\nabla F(x)$ at some x with $\langle a, x\rangle > 0$.
8. We define a map from $\mathbb{R}^2$ to $\mathbb{R}$ by $\|(x, y)\|_* := |x - y|$. Argue why this is *not* a norm on $\mathbb{R}^2$.
9. Let $a_i \in \mathbb{R}^n$, $i = 1, \ldots, m$ and $F : \mathbb{R}^n \to \mathbb{R}$ be defined by $F(x) = \sum_{i=1}^{m} (\langle a_i, x\rangle)^3$. Calculate $\nabla F(x)$.

10. Let $n \in \mathbb{N}$ and consider the map

$$\begin{aligned} \mathbb{R}^{n\times n} &\to [0,\infty) \\ A &\mapsto F(A) := \left(\operatorname{Tr}(A^T A)\right)^{1/2}. \end{aligned}$$

(a) Show that $F(A)$ is positive definite.

Hint: You may freely use that there are an invertible matrix $P \in \mathbb{R}^{n\times n}$ and $\lambda_1, \lambda_2, ..., \lambda_n \in [0,\infty)$ such that $A^T A = P^{-1} D P$ with the diagonal matrix

$$D = \operatorname{diag}(\lambda_1, ..., \lambda_n) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \lambda_n \end{pmatrix} \in \mathbb{R}^{n\times n}.$$

(b) Show that $F(A)$ is homogeneous.

(c) Show the triangle inequality $F(A+B) \le F(A) + F(B)$.

Recall also the Cauchy–Schwarz inequality

$$\left|\sum_{j=1}^{n} a_j b_j\right| \le \left(\sum_{j=1}^{n} a_j^2\right)^{1/2} \left(\sum_{j=1}^{n} b_j^2\right)^{1/2}$$

for all $a_1, ..., a_n, b_1, ..., b_n \in \mathbb{R}$.

(d) Let $x, y \in \mathbb{R}^n$ and recall the notation $\|x\|_2 := (\sum_{j=1}^{n} |x_j|^2)^{1/2}$. Show that $F(x\, y^T) = \|x\|_2\, \|y\|_2$ of the matrix $x\, y^T \in \mathbb{R}^{n\times n}$.

11. Let $n = 2$ and consider the map

$$\begin{aligned} \mathbb{R}^{2\times 2} &\to \mathbb{R} \\ A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} &\mapsto G(A) := a_{11}a_{22} - a_{21}a_{12}. \end{aligned}$$

Note that $G(A)$ is just the determinant of A.

(a) Show that for $h \in \mathbb{R}$ the derivative $DG(A)[h\mathbf{1}_2]$ along $h\mathbf{1}_2 = \begin{pmatrix} h & 0 \\ 0 & h \end{pmatrix}$ is

$$DG(A)[h\mathbf{1}_2] = h\operatorname{Tr}(A) = h(a_{11} + a_{22}).$$

(b) Let $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$. Compute $DG(A)[h\mathbf{1}_2]$.

Possible ways so solve the problems:

1. The claim is true: We could (for example) use that the trace is invariant under cyclic permutations and get

$$\langle A^T, B^T \rangle = \mathrm{Tr}(A^T B) = \mathrm{Tr}(BA^T) = \langle B, A \rangle = \langle A, B \rangle.$$

   One could also use the sum formula:

$$\langle A^T, B^T \rangle = \sum_{k=1}^{m} \sum_{j=1}^{n} (A^T)_{jk} (B^T)_{jk} = \sum_{k=1}^{m} \sum_{j=1}^{n} A_{kj} B_{kj} = \langle A, B \rangle.$$

2. (a) We calculate

$$\begin{aligned} F(R+H) &= \mathrm{Tr}((R+H)(R+H)^T) \\ &= \mathrm{Tr}(RR^T) + \mathrm{Tr}(RH^T) + \mathrm{Tr}(HR^T) + \mathrm{Tr}(HH^T) \end{aligned}$$

   and thus

$$F(R+H) - F(R) = \mathrm{Tr}(RH^T) + \mathrm{Tr}(HR^T) + \mathrm{Tr}(HH^T)$$

   Since $\mathrm{Tr}(HH^T)/\|H\| \to 0$ for $H \to 0$ we get

$$DF(R)[H] = \mathrm{Tr}(RH^T) + \mathrm{Tr}(HR^T).$$

   (Could further be simplified to $DF(R)[H] = 2\,\mathrm{Tr}(RH^T)$ or $= 2\,\mathrm{Tr}(HR^T)$ or $= 2\,\mathrm{Tr}(H^T R)$ or $= 2\,\mathrm{Tr}(R^T H)$)

   (b) By 1. and $\mathrm{Tr}(A) = \mathrm{Tr}(A^T)$we have $DF(R)[H] = \mathrm{Tr}(RH^T) + \mathrm{Tr}(HR^T) = 2\,\mathrm{Tr}(R^T H)$. By the definition of the inner product for matrices we get $DF(R)[H] = 2\langle H, R \rangle$ and thus $\nabla F(R) = 2R$.

3. Symmetry is fullfilled:

$$d(x,y) = \sum_{i=1}^{n} \tfrac{(x_i - y_i)^2}{x_i y_i} = \sum_{i=1}^{n} \tfrac{(y_i - x_i)^2}{y_i x_i} = d(y,x).$$

   Positivity and definiteness is fulfilled as well:

$$\begin{aligned} d(x,y) \geq 0 \text{ since } x_i \geq 0, \qquad d(x,y) = 0 &\iff \text{for all } i\text{: } \tfrac{(x_i - y_i)^2}{x_i y_i} = 0 \\ &\iff \text{for all } i\text{: } \quad (x_i - y_i)^2 = 0 \\ &\iff \text{for all } i\text{:} x_i = y_i. \end{aligned}$$

   The triangle inequality is not fulfilled: We consider $n = 1$

$$\begin{aligned} d(1,2) &= (1-2)^2/2 = \tfrac{1}{2}, \\ d(2,3) &= (2-3)^2/3 = \tfrac{1}{3} \end{aligned}$$

   but

$$d(1,3) = (1-3)^2/3 = \tfrac{4}{3} > \tfrac{1}{2} + \tfrac{1}{3} = d(1,2) + d(2,3).$$

4. The norm $\|A\|_{1\to 1}$ is the column sum norm, more precisely, the maximal 1-norm of the columns. The k-th column has 1-norm k, and hence the maximum is achieved at $k = n$ and we have $\|A\|_{1\to 1} = n$.

5. We calculate

$$\begin{aligned} F(R+H) = (R+H)^T A(R+H) &= R^T A(R+H) + H^T A(R+H) \\ & R^T AR + R^T AH + H^T AR + H^T AH \\ &= F(R) + R^T AH + H^T AR + H^T AH. \end{aligned}$$

Since $\|H^T AH\|/\|H\| \leq \|H^T\|\|A\|\|H\|/\|H\| = \|H^T\|\|A\| \to 0$ for $H \to 0$ and $H \mapsto R^T AH + H^T AR$ is linear in H we get

$$DF(R)[H] = R^T AH + H^T AR.$$

6. We could, for example, define

$$\|\begin{pmatrix} x \\ y \end{pmatrix}\|_{\text{mynorm}} := 10|x| + |y|.$$

(Another possibility would be $\|\binom{x}{y}\|_{\text{mynorm}} := \sqrt{100|x|^2 + |y|^2}$, and there infinitely many more...)

7. We calculate

$$F(x+h) = \sqrt{\langle a, x\rangle + \langle a, h\rangle}.$$

Now we use differentiability of the function $g(t) = \sqrt{t}$: it holds that

$$\sqrt{t+\varepsilon} = g(t+\varepsilon) = g(t) + g'(t)\varepsilon + \varphi(\varepsilon) = \sqrt{t} + \tfrac{1}{2\sqrt{t}}\varepsilon + \varphi(\varepsilon)$$

with $\varphi(\varepsilon)/\varepsilon \to 0$ for $\varepsilon \to 0$ (this is exactly the same as doing a Taylor expansion up to first order). Using this we get

$$F(x+h) = \sqrt{\langle a, x\rangle + \langle a, h\rangle} = \sqrt{\langle a, x\rangle} + \tfrac{1}{2\sqrt{\langle a,x\rangle}}\langle a, h\rangle + \varphi(\langle a, h\rangle)$$

with $\varphi(\langle a, h\rangle)/\langle a, h\rangle \to 0$ for $h \to 0$. We set $\Phi(h) = \varphi(\langle a, h\rangle)$ and get from $\langle a, h\rangle \leq \|a\|\|h\|$ that

$$\frac{\Phi(h)}{\|h\|} \leq \frac{\varphi(\langle a, h\rangle)\|a\|}{\langle a, h\rangle} \to 0$$

for $h \to 0$.

Thus, we have

$$DF(x)[h] = \tfrac{1}{2\sqrt{\langle a,x\rangle}}\langle a, h\rangle = \langle \tfrac{1}{2\sqrt{\langle a,x\rangle}} a, h\rangle$$

and hence

$$\nabla F(x) = \tfrac{1}{2\sqrt{\langle a,x\rangle}} a.$$

8. This map is not positive definite since $\|(1, 1)\|_* = |1 - 1| = 0$.

9. One calculates

$$\begin{aligned} F(x+h) &= \sum_{i=1}^{m} (\langle a_i, x\rangle + \langle a_i, h\rangle)^3 \\ &= \sum_{i=1}^{m} \langle a_i, x\rangle^3 + 3\langle a_i, x\rangle^2\langle a_i, h\rangle + 3\langle a_i, x\rangle\langle a_i, h\rangle^2 + \langle a_i, h\rangle^3 \\ &= F(x) + \sum_{i=1}^{m} 3\langle a_i, x\rangle^2\langle a_i, h\rangle + \Phi(h). \end{aligned}$$

By Cauchy–Schwarz,

$$
\begin{aligned}
\frac{|\Phi(h)|}{\|h\|} &= \frac{\sum_{i=1}^{m} 3\langle a_i, x\rangle \langle a_i, h\rangle^2 + \langle a_i, h\rangle^3}{\|h\|} \\
&\leq \frac{\sum_{i=1}^{m} 3\langle a_i, x\rangle \|a_i\|^2 \|h\|^2 + \|a_i\|^3 \|h\|^3}{\|h\|} \to 0
\end{aligned}
$$

for $h \to 0$. Hence,

$$
DF(x)[h] = \sum_{i=1}^{m} 3\langle a_i, x\rangle^2 \langle a_i, h\rangle = \langle \sum_{i=1}^{m} 3\langle a_i, x\rangle^2 a_i, h\rangle
$$

and therefore

$$
\nabla F(x) = \sum_{i=1}^{m} 3\langle a_i, x\rangle^2 a_i.
$$

10. This exercise is a little longer.

   (a) Positive definiteness: If $A = 0$, then $F(A) = 0$. If $F(A) = 0$, then $0 = \operatorname{Tr}(A^T A) = \operatorname{Tr}(D) = \sum_j \lambda_j$ by cyclicity of the trace. Since $\lambda_j \geq 0$ for all $j \in \{1, 2, ..., n\}$, $\lambda_1 = \lambda_2 = ... = \lambda_n = 0$, i.e., $A = 0$.

   (b) Homogeneity: For $\lambda \in \mathbb{R}$, we have $\|\lambda A\|_F = \sqrt{\operatorname{Tr}\left((\lambda A)^T \lambda A\right)} = \sqrt{\lambda^2 \operatorname{Tr}(A^T A)} = |\lambda| \|A\|_F$.

   (c) Triangle inequality: Since $\operatorname{Tr}(A^T B) = \operatorname{Tr}(B^T A)$ by cyclicity and $\operatorname{Tr}(C) = \operatorname{Tr}(C^T)$,

$$
\operatorname{Tr}(A+B)^T(A+B) = \operatorname{Tr}(A^T A) + \operatorname{Tr}(B^T B) + 2\operatorname{Tr}(A^T B).
$$

   Let $A = (a_{j,\ell})_{j,\ell=1}^n$, $B = (b_{j,\ell})_{j,\ell=1}^n$, $\tilde{a}_{j,\ell} = a_{\ell,j}$, and $\tilde{b}_{j,\ell} = b_{\ell,j}$. By Cauchy–Schwarz,

$$
\begin{aligned}
\operatorname{Tr}(A^T B) = \sum_{j,\ell} \tilde{a}_{j,\ell} b_{\ell,j} &\leq \left(\sum_{j,\ell} (\tilde{a}_{j,\ell})^2\right)^{1/2} \left(\sum_{j,\ell} b_{j,\ell}^2\right)^{1/2} \\
&= \left(\sum_{j,\ell} \tilde{a}_{j,\ell} a_{\ell,j}\right)^{1/2} \left(\sum_{j,\ell} \tilde{b}_{\ell,j} b_{j,\ell}\right)^{1/2} = F(A)F(B).
\end{aligned}
$$

   Thus,

$$
F(A+B)^2 \leq F(A)^2 + F(B)^2 + 2F(A)F(B) = (F(A) + F(B))^2,
$$

   as desired.

   (d) Let $A = x\, y^T$. Then, by associativity of matrix multiplication and cyclicity of the trace,

$$
\operatorname{Tr} A^T A = \operatorname{Tr}(y\, x^T x\, y^T) = \|x\|_2^2 \cdot \operatorname{Tr}(y\, y^T) = \|x\|_2^2 \|y\|_2^2.
$$

   Thus, $F(x\, y^T) = \|x\|_2 \|y\|_2$.

11. This exercise is a little longer.

(a) We compute

$$G(A + h\mathbf{1}_2) - G(A) = h(a_{11} + a_{22}) + h^2.$$

Since

$$\lim_{h \to 0} \frac{h^2}{\|h\mathbf{1}_2\|} = 0$$

for any matrix norm (as all norms on finite dimensional vector spaces are equivalent), we obtain $DG(A)[h\mathbf{1}_2] = h\,\mathrm{Tr}(A) = h(a_{11} + a_{22})$.

(b) We have $DG(A)[h\mathbf{1}_2] = h\,\mathrm{Tr}(A) = 5h$.