

Mathematics for Data Science

M1 IDD

Exam - January 8, 2025 / Examen - 8 janvier 2025



English version

This exam is available in both English and French. Students can write their answers in either language. If students suspect that a question contains an error, they shall indicate it explicitly on their paper, and proceed while taking this error into account.

Duration: 2 hours.

Contents: Four exercises. English version on pages 1-4, French version on pages 5-8.

Allowed documents: One A4 sheet of handwritten or typed notes (two-sided).

Foreword

- Dimensions of vectors and matrices will always be assumed to be greater than or equal to 1.
- The notation $\|\cdot\|$ will be used for the Euclidean vector norm $\|\mathbf{x}\| = \sqrt{\mathbf{x}^T \mathbf{x}}$.
- Given a vector $\mathbf{x} \in \mathbb{R}^n$ with $n \geq 1$, the i th coordinate of this vector will be denoted by x_i or $[\mathbf{x}]_i$.
- For any integer $n \geq 1$, the zero vector in $\mathbb{R}^n$ will be denoted by $\mathbf{0}_{\mathbb{R}^n}$.
- The canonical basis vectors in $\mathbb{R}^n$ will be denoted by $\mathbf{e}_1, \dots, \mathbf{e}_n$, so that $\mathbf{e}_j$ is the vector such that $[\mathbf{e}_j]_i = 0$ if $i \neq j$ and $[\mathbf{e}_j]_j = 1$.
- Given any vector $\mathbf{a} \in \mathbb{R}^n$, the gradient of $\mathbf{x} \mapsto \mathbf{a}^T \mathbf{x}$ at any $\mathbf{x} \in \mathbb{R}^n$ is equal to $\mathbf{a}$.

Exercise 1: Extreme points and convex sets

In this exercise, we investigate the notion of extreme points of convex sets. Given a convex set $\mathcal{C}$ in $\mathbb{R}^n$, a vector $\bar{x} \in \mathcal{C}$ is called an *extreme point* of $\mathcal{C}$ if

$$\forall (x, y) \in \mathcal{C}^2, \forall \alpha \in [0, 1], \quad [\bar{x} = \alpha x + (1 - \alpha)y] \implies [x = y = \bar{x}].$$

The set of extreme points of $\mathcal{C}$ will be denoted by $\text{ext}(\mathcal{C})$.

a) In this question, we aim at showing that $\bar{x} \in \text{ext}(\mathcal{C})$ if and only if $\mathcal{C} \setminus \{\bar{x}\}$ is a convex set.

- i) Consider $\bar{x} \in \text{ext}(\mathcal{C})$. For any $(x, y) \in (\mathcal{C} \setminus \{\bar{x}\})^2$ and any $\alpha \in [0, 1]$, justify that $\alpha x + (1 - \alpha)y \neq \bar{x}$.
- ii) Use the result from question a)i) to conclude that $\mathcal{C} \setminus \{\bar{x}\}$ is a convex set.
- iii) Conversely, suppose that $\mathcal{C} \setminus \{\bar{x}\}$ is a convex set for some $\bar{x} \in \mathcal{C}$. Show that if there exist $(x, y) \in (\mathcal{C} \setminus \{\bar{x}\})^2$ and $\alpha \in [0, 1]$ such that $\bar{x} = \alpha x + (1 - \alpha)y$, then we obtain a contradiction.
- iv) Conclude from the result of question a)iii) that $\bar{x} \in \text{ext}(\mathcal{C})$.

b) Justify that the set $\{x \in \mathbb{R}^n \mid x_i = 0 \quad \forall i = 2, \dots, n\}$ is a convex set with no extreme point.

c) In this question, we consider the set

$$\mathcal{B}_1 = \left\{ x \in \mathbb{R}^n \mid \sum_{i=1}^n |x_i| \leq 1 \right\}. \quad (1)$$

- i) Justify that the set $\mathcal{B}_1$ is convex.
 - ii) Show further that $\mathcal{B}_1 = \text{conv}(e_1, \dots, e_n, -e_1, \dots, -e_n)$, where $e_1, \dots, e_n$ are the canonical basis vectors in $\mathbb{R}^n$ and $\text{conv}(\cdot)$ denotes the convex hull of those vectors.
 - iii) Using the result from question b)ii), give all extreme points of $\mathcal{B}_1$.
- d) Finally, we consider

$$\mathcal{B}_\infty = \left\{ x \in \mathbb{R}^n \mid \max_{i=1, \dots, n} |x_i| \leq 1 \right\}.$$

- i) Assuming *for this question only* that $n = 1$, find the extreme points of $\mathcal{B}_\infty$.
- ii) Given two convex sets $\mathcal{C}_1$ and $\mathcal{C}_2$ of $\mathbb{R}^m$, show that

$$\text{ext}(\mathcal{C}_1 \times \mathcal{C}_2) = \text{ext}(\mathcal{C}_1) \times \text{ext}(\mathcal{C}_2).$$

$$\text{where we recall that } \mathcal{A} \times \mathcal{B} = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \mid x \in \mathcal{A}, y \in \mathcal{B} \right\}.$$

- iii) Use the result from question c)i) to establish that

$$\text{ext}(\mathcal{B}_\infty) = \{x \in \mathbb{R}^n \mid [x]_i \in \{-1, +1\} \quad \forall i = 1, \dots, n\}.$$

Exercise 2: Linear optimization and extreme points

In this exercise, we consider the following optimization problem

$$\begin{aligned} & \text{minimize}_{\mathbf{x} \in \mathbb{R}^n} && \mathbf{g}^T \mathbf{x} \\ & \text{subject to} && \mathbf{x} \in \mathcal{B}_1, \end{aligned} \quad (2)$$

where $\mathbf{g} \in \mathbb{R}^n$ and $\mathcal{B}_1$ is defined by (1).

- Justify that problem (2) is a convex optimization problem.
- Justify that problem (2) has a solution. Is that solution necessarily unique?
- Write down the first-order optimality conditions for problem (2).
- Suppose that $\mathbf{g} = \mathbf{e}_i$ (i th vector of the canonical basis). Show that $-\mathbf{e}_i$ is a solution of problem (2).

Exercise 3: Linear optimization and duality

In this exercise, we consider a linear optimization problem over the set $\mathcal{B}_\infty$, as defined in Exercise 1. A possible formulation of such a problem is

$$\begin{aligned} & \text{minimize}_{\mathbf{x} \in \mathbb{R}^n} && \mathbf{g}^T \mathbf{x} \\ & \text{subject to} && -1 \leq x_i \leq 1 \quad \forall i = 1, \dots, n. \end{aligned} \quad (3)$$

- Is problem (3) a convex optimization problem in standard form? If not, provide an equivalent reformulation in standard form.
- Write down the Lagrangian associated with the problem (or its reformulation if necessary per question a)).
- Define the dual problem of (3) (or its reformulation if necessary per question a)), and show that it is equivalent to a linear program.
- Justify that constraint qualification holds for problem (3) and its dual.
- Write down the KKT conditions for problem (3).
- Use those conditions to find a solution to the problem when $\mathbf{g} = \mathbf{e}_1$ (first vector of the canonical basis).
- We now consider a variant on problem (3) where the constraints are put in the objective under the form of a so-called barrier function. We define the function $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ by

$$\forall \mathbf{x} \in \mathbb{R}^n, \quad f(\mathbf{x}) = \begin{cases} \mathbf{g}^T \mathbf{x} - \sum_{i=1}^n \ln(x_i + 1) - \sum_{i=1}^n \ln(1 - x_i) & \text{if } -1 < x_i < 1 \quad \forall i = 1, \dots, n, \\ +\infty & \text{otherwise.} \end{cases}$$

- What is the (effective) domain of the function f ?
- Justify that the function f is convex.
- Suppose, as in question f), that $\mathbf{g} = \mathbf{e}_1$. Justify that the problem $\text{minimize}_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$ is not equivalent to problem (3).

Exercise 4: Concentration inequality

In this exercise, we suppose that the vector $\mathbf{g}$ used in Exercises 2 and 3 is randomly generated according to a standard Gaussian distribution. Our goal is to study properties of $\|\mathbf{g}\|$, which is an approximation of the optimal value of both problems (2) and (3), and to show that it is close to $\sqrt{n}$ with high probability.

To this end, we will rely on a form of Bernstein's inequality. Let $y_1, \dots, y_n$ be n independent, mean zero, subexponential random variables. Then, for any $u \geq 0$, it holds that

$$\mathbb{P} \left(\left| \frac{1}{n} \sum_{i=1}^n y_i \right| \geq u \right) \leq 2 \exp \left(-\frac{cn}{K^2} \min(u^2, u) \right), \quad (4)$$

where $K = \max \{1, \max_i \|y_i\|_{\Psi_1}\}$ and $c > 0$ is a universal constant.

- a) Why is the inequality (4) called a concentration inequality?
- b) Consider now $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$. By definition, the coordinates of $\mathbf{g}$, denoted by $g_1, \dots, g_n$, are independent, subgaussian random variables such that $\mathbb{E}[g_i^2] = 1$.
 - i) Justify that the variables $y_i = g_i^2 - 1$ are independent, mean zero and subexponential.
 - ii) By applying (4) show that

$$\mathbb{P} \left(\left| \frac{1}{n} \|\mathbf{g}\|^2 - 1 \right| \geq u \right) \leq 2 \exp \left(-\frac{cn}{K^4} \min(u^2, u) \right)$$

for any $u \geq 0$.

- iii) Using the result from question b)ii) together with the property

$$|z - 1| \geq \delta \Rightarrow |z^2 - 1| \geq \max(\delta, \delta^2)$$

valid for any $z \geq 0$ and $\delta \geq 0$, show that

$$\mathbb{P} \left(\left| \frac{1}{\sqrt{n}} \|\mathbf{g}\| - 1 \right| \geq \delta \right) \leq 2 \exp \left(-\frac{cn}{K^4} \delta^2 \right)$$

for any $\delta \geq 0$.

- iv) Conclude from the previous question that

$$\mathbb{P} (|\|\mathbf{g}\| - \sqrt{n}| \geq t) \leq 2 \exp \left(-\frac{ct^2}{K^4} \right)$$

for any $t \geq 0$.

- c) Let $\mathbf{s} \in \mathbb{R}^n$ be a random vector uniformly distributed on the sphere of radius $\sqrt{n}$ in $\mathbb{R}^n$. Given any $t \geq 0$, what is

$$\mathbb{P} (|\|\mathbf{s}\| - \sqrt{n}| \geq t)$$

equal to?

- d) Explain how the results in this exercise highlight that standard Gaussian vectors concentrate around the sphere of radius $\sqrt{n}$ in $\mathbb{R}^n$.

Version française

Ce sujet d'examen est fourni en versions anglaise et française. Les étudiants sont libres de choisir l'une ou l'autre de ces deux langues pour composer.

Si les étudiants pensent constater une erreur dans l'une des questions, ils sont invités à l'indiquer explicitement sur leur copie, et à poursuivre leur composition en tenant compte de cette erreur.

Durée : 2 heures.

Contenu : Quatre exercices en versions anglaise (pages 1 à 4) et française (pages 5 à 8).

Documents autorisés : Une feuille A4 recto-verso de notes manuscrites ou imprimées.

Avant-propos

- Les dimensions des vecteurs ou matrices seront toujours supposées supérieures ou égales à 1.
- La notation $\|\cdot\|$ désignera la norme euclidienne vectorielle $\|x\| = \sqrt{x^T x}$.
- Pour tout vecteur $x \in \mathbb{R}^n$ avec $n \geq 1$, la i -ème coordonnée de ce vecteur sera notée x_i .
- Pour tout entier $n \geq 1$, la notation $\mathbf{0}_{\mathbb{R}^n}$ désignera le vecteur nul de $\mathbb{R}^n$.
- Les vecteurs de la base canonique de $\mathbb{R}^n$ seront notés $e_1, \dots, e_n$, où e_j est le vecteur vérifiant $[e_j]_i = 0$ pour $i \neq j$ et $[e_j]_j = 1$.
- Pour tout vecteur $a \in \mathbb{R}^n$, le gradient de la fonction $x \mapsto a^T x$ en tout point $x \in \mathbb{R}^n$ est égal au vecteur a .

Exercice 1 : Points extrémaux et ensembles convexes

Dans cet exercice, on s'intéresse à la notion de points extrémaux. Étant donné un ensemble convexe $\mathcal{C}$ de $\mathbb{R}^n$, on dit qu'un vecteur $\bar{x} \in \mathcal{C}$ est un *point extrémal* de $\mathcal{C}$ si

$$\forall (x, y) \in \mathcal{C}^2, \forall \alpha \in [0, 1], \quad [\bar{x} = \alpha x + (1 - \alpha)y] \implies [x = y = \bar{x}].$$

L'ensemble des points extrémaux de $\mathcal{C}$ sera noté $\text{ext}(\mathcal{C})$.

a) Le but de cette question est de démontrer que $\bar{x} \in \text{ext}(\mathcal{C})$ si et seulement si $\mathcal{C} \setminus \{\bar{x}\}$ est un ensemble convexe.

- i) Soit $\bar{x} \in \text{ext}(\mathcal{C})$. Pour tous $(x, y) \in (\mathcal{C} \setminus \{\bar{x}\})^2$ et tout $\alpha \in [0, 1]$, justifier que $\alpha x + (1 - \alpha)y \neq \bar{x}$.
- ii) En utilisant le résultat de la question a)i), montrer alors que $\mathcal{C} \setminus \{\bar{x}\}$ est un ensemble convexe.
- iii) Réciproquement, supposons que $\mathcal{C} \setminus \{\bar{x}\}$ soit convexe pour $\bar{x} \in \mathcal{C}$. Montrer que s'il existe $(x, y) \in (\mathcal{C} \setminus \{\bar{x}\})^2$ et $\alpha \in [0, 1]$ tels que $\bar{x} = \alpha x + (1 - \alpha)y$, alors on aboutit à une contradiction.
- iv) Dédire du résultat de la question a)iii) que $\bar{x} \in \text{ext}(\mathcal{C})$.

b) Justifier que l'ensemble $\{x \in \mathbb{R}^n \mid x_i = 0 \ \forall i = 2, \dots, n\}$ est un ensemble convexe qui ne possède pas de point extrémal.

c) Dans cette question, on s'intéresse à l'ensemble

$$\mathcal{B}_1 = \left\{ x \in \mathbb{R}^n \mid \sum_{i=1}^n |x_i| \leq 1 \right\}. \quad (5)$$

- i) Justifier que l'ensemble $\mathcal{B}_1$ est un ensemble convexe.
- ii) Montrer que $\mathcal{B}_1 = \text{conv}(e_1, \dots, e_n, -e_1, \dots, -e_n)$, où $e_1, \dots, e_n$ désignent les vecteurs de la base canonique de $\mathbb{R}^n$ et $\text{conv}(\cdot)$ l'enveloppe convexe de ces vecteurs.
- iii) En utilisant le résultat de la question b)ii), donner tous les points extrémaux de $\mathcal{B}_1$.

d) On considère finalement l'ensemble

$$\mathcal{B}_\infty = \left\{ x \in \mathbb{R}^n \mid \max_{i=1, \dots, n} |x_i| \leq 1 \right\}.$$

- i) On suppose *uniquement pour cette question* que $n = 1$. Trouver alors l'ensemble des points extrémaux de $\mathcal{B}_\infty$.
- ii) Soient deux ensembles convexes $\mathcal{C}_1$ et $\mathcal{C}_2$ de $\mathbb{R}^m$. Montrer que

$$\text{ext}(\mathcal{C}_1 \times \mathcal{C}_2) = \text{ext}(\mathcal{C}_1) \times \text{ext}(\mathcal{C}_2),$$

$$\text{où } \mathcal{A} \times \mathcal{B} = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \mid x \in \mathcal{A}, y \in \mathcal{B} \right\}.$$

- iii) Utiliser le résultat de la question c)i) pour montrer que

$$\text{ext}(\mathcal{B}_\infty) = \{x \in \mathbb{R}^n \mid [x]_i \in \{-1, +1\} \ \forall i = 1, \dots, n\}.$$

Exercice 2 : Optimisation linéaire et points extrémaux

Dans cet exercice, on s'intéresse au problème d'optimisation

$$\begin{array}{ll} \text{minimiser}_{\mathbf{x} \in \mathbb{R}^n} & \mathbf{g}^T \mathbf{x} \\ \text{sous contraintes} & \mathbf{x} \in \mathcal{B}_1, \end{array} \quad (6)$$

où $\mathbf{g} \in \mathbb{R}^n$ et $\mathcal{B}_1$ est l'ensemble défini par (5).

- Justifier que le problème (6) est un problème d'optimisation convexe.
- Justifier que le problème (6) possède une solution. Cette solution est-elle toujours unique ?
- Écrire les conditions d'optimalité à l'ordre un pour le problème (6).
- Supposons que $\mathbf{g} = \mathbf{e}_i$ (i -ème vecteur de la base canonique). Montrer alors que $-\mathbf{e}_i$ est solution du problème (6).

Exercice 3 : Optimisation linéaire et dualité

Dans cet exercice, on considère un problème d'optimisation linéaire sur l'ensemble $\mathcal{B}_\infty$ introduit dans l'exercice 1, et plus précisément la formulation suivante :

$$\begin{array}{ll} \text{minimiser}_{\mathbf{x} \in \mathbb{R}^n} & \mathbf{g}^T \mathbf{x} \\ \text{sous contraintes} & -1 \leq x_i \leq 1 \quad \forall i = 1, \dots, n. \end{array} \quad (7)$$

- Le problème (7) est-il un problème d'optimisation convexe en forme standard ? Dans le cas contraire, donner une reformulation en forme standard.
- Écrire le lagrangien associé au problème (7) (ou à sa reformulation en forme standard si celui-ci n'est pas en forme standard).
- Définir le problème dual de (7), et montrer que ce problème est équivalent à un programme linéaire.
- Justifier qu'il y a qualification des contraintes pour le problème (7) et son dual.
- Écrire les conditions de KKT associées au problème (7).
- Utiliser les conditions de KKT pour trouver une solution au problème lorsque $\mathbf{g} = \mathbf{e}_1$ (premier vecteur de la base canonique).
- On considère une variante du problème (7) où les contraintes sont incorporées dans la fonction objectif via des termes dits de barrière. On définit ainsi la fonction $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ suivante :

$$\forall \mathbf{x} \in \mathbb{R}^n, \quad f(\mathbf{x}) = \begin{cases} \mathbf{g}^T \mathbf{x} - \sum_{i=1}^n \ln(x_i + 1) - \sum_{i=1}^n \ln(1 - x_i) & \text{si } -1 < x_i < 1 \quad \forall i = 1, \dots, n, \\ +\infty & \text{sinon.} \end{cases}$$

- Quel est le domaine (effectif) de la fonction f ?
- Justifier que la fonction f est convexe.
- On suppose comme dans la question f) que $\mathbf{g} = \mathbf{e}_1$. Justifier que le problème $\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$ n'est pas équivalent au problème (7).

Exercice 4 : Inégalité de concentration

Dans cet exercice, on suppose que le vecteur $\mathbf{g}$ utilisé dans les exercices 2 et 3 est généré aléatoirement selon une distribution gaussienne standard. On cherche à étudier les propriétés de $\|\mathbf{g}\|$, que l'on peut voir comme une estimation des valeurs optimales des problèmes (6) et (7). On va voir que cette valeur est proche de $\sqrt{n}$ avec forte probabilité.

Pour cela, on se basera sur une inégalité de Bernstein. Soient $y_1, \dots, y_n$ n variables aléatoires sous-exponentielles, indépendantes et d'espérance nulle. Alors, pour tout $u \geq 0$, on a

$$\mathbb{P} \left(\left| \frac{1}{n} \sum_{i=1}^n y_i \right| \geq u \right) \leq 2 \exp \left(-\frac{cn}{K^2} \min(u^2, u) \right), \quad (8)$$

où $K = \max \{1, \max_i \|y_i\|_{\Psi_1}\}$ et $c > 0$ est une constante "universelle".

- a) Pourquoi dit-on que l'inégalité (8) est une inégalité de concentration ?
- b) Soit un vecteur gaussien $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$. Par définition, les coordonnées de $\mathbf{g}$, notées $g_1, \dots, g_n$, sont des variables aléatoires sous-gaussiennes, indépendantes et telles que $\mathbb{E}[g_i^2] = 1$.
 - i) Justifier que les variables $y_i = g_i^2 - 1$ sont sous-exponentielles, indépendantes et d'espérance nulle.
 - ii) En appliquant (8), montrer que

$$\mathbb{P} \left(\left| \frac{1}{n} \|\mathbf{g}\|^2 - 1 \right| \geq u \right) \leq 2 \exp \left(-\frac{cn}{K^4} \min(u^2, u) \right)$$

pour tout $u \geq 0$.

- iii) En utilisant le résultat de la question b)ii) ainsi que la propriété

$$|z - 1| \geq \delta \Rightarrow |z^2 - 1| \geq \max(\delta, \delta^2)$$

valable pour tous $z \geq 0$ et $\delta \geq 0$, montrer que

$$\mathbb{P} \left(\left| \frac{1}{\sqrt{n}} \|\mathbf{g}\| - 1 \right| \geq \delta \right) \leq 2 \exp \left(-\frac{cn}{K^4} \delta^2 \right)$$

pour tout $\delta \geq 0$.

- iv) En déduire que

$$\mathbb{P} (|\|\mathbf{g}\| - \sqrt{n}| \geq t) \leq 2 \exp \left(-\frac{ct^2}{K^4} \right)$$

pour tout $t \geq 0$.

- c) Soit $\mathbf{s} \in \mathbb{R}^n$ un vecteur aléatoire uniformément distribué sur la sphère de rayon $\sqrt{n}$ dans $\mathbb{R}^n$. Pour tout $t \geq 0$, que vaut

$$\mathbb{P} (|\|\mathbf{s}\| - \sqrt{n}| \geq t) ?$$

- d) Expliquer comment les résultats de l'exercice illustrent que les vecteurs gaussiens standards se concentrent autour de la sphère de rayon $\sqrt{n}$ de $\mathbb{R}^n$.