Ordinary Differential Equations - Final exam - Spring 2018 - NYU

This exam consists of three exercises and a problem:

- Exercise 1 is a list of six computational questions.
- Exercise 2 studies the time of existence for the maximal solution of a non-linear ODE.
- Exercise 3 is devoted to the qualitative study of some non-linear, autonomous ODE.
- The Problem is devoted to the "gradient descent" algorithm. There are three (almost) independent parts
  - In 4.1, we study the gradient descent in continuous time.
  - In 4.2, we study a numerical scheme for gradient descent.
  - In 4.3, we study the case of a gradient descent with some perturbative noise term.

Of course, **THIS IS WAY TOO MUCH** for 110 minutes. I recommend that you start by solving Exercise 1. Then, you are invited to first have a quick overview of the exam, choose the topics that you feel the most comfortable with and solve the related questions. An indicative weight of each part is given, the total weight being

$$30 + 20 + 30 + 40 + 40 + 20 = 180.$$

(It is plausible that a total of 100 points would roughly correspond to a grade of 100% but the scheme may be adjusted after grading.)

Allow me to remind you that:

- Every answer must be precisely justified, unless stated otherwise. Intellectual honesty is a good guideline for "how much justification should I provide?"
- Every answer must use **words**, and will preferably take the form of one or several **full sentences**. It is good practice to underline or to $\boxed{\text{box}}$ the key steps of an argument and the final result of a computation.
- Please use a **real pen**, not a pencil. Use scratch paper for your trial-and-error process and for uncertain computations. It is, of course, OK to strike out a paragraph.
- It is always OK to skip a question and to admit the result of a previous question. Indicate it clearly. In general, always refer precisely to the result(s) you are using, may it be the answer to a previous question or a result from class.

Good luck!

1 Computational questions (30)

For all the following ODE's with given initial condition, find the expression of the solution as a function of the time variable t. You do not have to justify existence, uniqueness, or to worry about the time of existence of the solutions, but you need to explain your computations.

1. $$x' = \frac{t}{3+x}, \quad x(0) = 1.$$

2. $$x'' - x' + x = 0, \quad x(0) = 1, \quad x'(0) = 0.$$

3. $$X' = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} X, \quad X(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

4. $$x' = t + tx^2, \quad x(0) = 0.$$

5. $$x' - x = e^t, \quad x(0) = 1.$$

6. $$tx' = x + te^{x/t}, \quad x(1) = 1.$$

2 Time of existence (20)

We consider the ODE

$$x' = t^2 x + (1 + \cos^2(t))x^2.$$

We denote by γ the maximal solution of this ODE with initial condition $\gamma(0) = 1$, defined on some interval (α, β).

1. Show that $\gamma(t)$ is always positive for t in (α, β).
2. Show that γ is increasing on (α, β).
3. Justify that $\alpha = -\infty$.
4. Justify that β is finite. *You may use the ODE $x' = x^2$ for comparison.*

3 Qualitative study (30)

We consider the ODE

$$x'' - \sin(x) = 0. \tag{1}$$

1. Re-write (1) as a first-order ODE with unknown function $X = (x, x')$.
2. Find a (non-trivial) conserved quantity.
3. Sketch the allure of the orbits in $\mathbb{R}^2$ with the system of coordinates (x, x').
   (a) Near the point $(\pi/2, 0)$.
   (b) Near the point $(0, 2)$.

   (Briefly justify your drawing.)
4. Explain why we can find a change of variables that would transform the two sketches drawn in the previous question onto one another.
5. Explain why, near the point $(0, 0)$, the flow of this ODE looks like

$$e^{tA}, \quad A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

6. Sketch the allure of the orbits near $(0, 0)$.

4 Around the gradient descent

Let $d \geq 1$ be the dimension. In this problem, E is a function from $\mathbb{R}^d$ to $\mathbb{R}$ of class C^1 such that:

- E is **Lipschitz**, with a Lipschitz constant denoted by L. By definition, it means

$$\forall x, y \in \mathbb{R}^d, \quad |\mathsf{E}(x) - \mathsf{E}(y)| \leq L\|x - y\|.$$

  You may use the following consequence: $\boxed{\forall x \in \mathbb{R}^d, \quad \|\nabla \mathsf{E}(x)\| \leq L.}$

- The gradient ∇E is **Lipschitz**, with a Lipschitz constant denoted by M. By definition, it means

$$\boxed{\forall x, y \in \mathbb{R}^d, \quad \|\nabla \mathsf{E}(x) - \nabla \mathsf{E}(y)\| \leq M\|x - y\|.}$$

- E is **α-convex** for some α. By definition, it means (we denote by $\langle a, b\rangle$ the scalar product of two vectors).

$$\boxed{\forall x, y \in \mathbb{R}^d, \quad \langle \nabla \mathsf{E}(x) - \nabla \mathsf{E}(y), x - y\rangle \geq \alpha\|x - y\|^2.}$$

Preliminary question

1. Show that, because E is α-convex, then E has at most one critical point.

In the following, we will denote by $X_{\min}$ the unique critical point, we assume that it exists and is the unique global minimizer of E.

4.1 Gradient descent in continuous time (40)

In this section, we fix X_0 in $\mathbb{R}^d$ and we study the ODE

$$\boxed{X'(t) = -\nabla \mathsf{E}(X(t)), \quad X(0) = X_0} \tag{2}$$

where X is an unknown function with values in $\mathbb{R}^d$. This is known as a "gradient descent".

4.1.1 Convergence to the minimizer

1. Explain why the maximal solution to the ODE (2) exists, is unique, and is defined for all times t in $(-\infty, +\infty)$.

2. Are there constant solutions to (2)? If yes, how many?

3. Show that either the solution is constant, or $\mathsf{E}(X(t))$ is (strictly) decreasing in t.

4. Is the equilibrium solution $X(t) \equiv X_{\min}$ stable?

5. Prove that $\lim_{t+\infty} X(t) = X_{\min}$ (for an initial condition close enough to $X_{\min}$, or, more difficult, for any choice of initial condition).

4.1.2 Speed of convergence

In this paragraph, we want to quantify the speed at which $X(t)$ tends to $X_{\min}$. We suppose that the initial condition (at time 0) X_0 is not equal to $X_{\min}$. For $t \geq 0$, we introduce the quantity

$$\mathcal{D}(t) := \|X(t) - X_{\min}\|^2.$$

1. Compute $\mathcal{D}'(t)$. *You may use one of the auxiliary results.*

2. Show that for $t \geq 0$ we have
$$D'(t) \leq -\alpha D(t)$$

3. Prove that
$$D(t) \leq \|X_0 - X_{\min}\|^2 e^{-\alpha t}$$
and conclude about the speed of convergence of $X(t)$ to $X_{\min}$.

4.2 Numerical study (40)

In this paragraph, we are interested in a numerical approach to gradient descent. It can be described as a sequence $\{X_n\}_{n\geq 0}$ defined as follows:

- We start at some point X_0.
- At each step $n \geq 0$, we chose a step-size $s_n \geq 0$ and we compute X_{n+1} in terms of X_n by

$$\boxed{X_{n+1} := X_n - s_n \nabla \mathsf{E}(X_n).} \tag{3}$$

4.2.1 A model case

In this paragraph only, we take $\mathsf{E}(X) = \|X\|^2$. For the two following choices of step-sizes, show that the numerical scheme defined above does **not** converge to the minimizer of E (here $X_{\min} = 0$ of course), unless we start at this point.

1. For a constant step-size $s_n = 1$.
2. For a step-size $s_n = n^{-100}$.

It is thus important to chose the step-size carefully.
In applied maths classes, the usual heuristics for step-sizes is to chose s_n such that

$$\begin{cases} \sum_n s_n & \text{diverges} \\ \sum_n s_n^2 & \text{converges.} \end{cases}$$

We will try to justify this heuristics.

4.2.2 Convergence to the minimizer

- We let $\{X_n\}_n$ be the sequence of points defined as above.
- We let $t \mapsto X(t)$ be the solution to the "gradient descent" ODE (2) with initial condition $X(0) = X_0$.
- We let $t_0 = 0$ and we let
$$\widetilde{X}_0 := X(t_0) = X(0) = X_0.$$
- For any $n \geq 0$, we define
$$t_{n+1} = t_n + s_n, \quad \widetilde{X}_{n+1} = X(t_{n+1}).$$

In other words: t_n is the time after n steps, $\widetilde{X}_n$ is the value of the "real solution" at time t_n while X_n is the value of the numerical solution after n steps.

1. Explain why, if we assume that the series $\sum_n s_n$ diverges, then $\widetilde{X}_n$ tends to the minimizer $X_{\min}$ as $n \to +\infty$.

   In the following, we will always assume that $\sum_n s_n$ diverges.

2. Show that $\widetilde{X}_n$ satisfies
$$\widetilde{X}_{n+1} = \widetilde{X}_n - \int_{t_n}^{t_{n+1}} \nabla \mathsf{E}(X(s))ds.$$
The next questions are devoted to the analysis of this numerical scheme, and are thus of "real analysis" spirit.

3. Show that we have
$$\widetilde{X}_{n+1} = \widetilde{X}_n - s_n \nabla \mathsf{E}(\widetilde{X}_n) + \varepsilon_n,$$
with an error term ε_n bounded by
$$\|\varepsilon_n\| \leq \frac{MLs_n^2}{2},$$
where L, M are the Lipschitz constants defined in the introduction.

4. Using α-convexity, show that
$$\|X_n - s_n \nabla \mathsf{E}(X_n) - \widetilde{X}_n + s_n \nabla \mathsf{E}(\widetilde{X}_n)\|^2 \leq \|X_n - \widetilde{X}_n\|^2 \left(1 - 2\alpha s_n + s_n^2 M^2\right),$$
where α, M are the constants defined in the introduction.

5. For any $n \geq 0$, we let V_n be the difference $V_n := X_n - \tilde{X}_n$. Prove that
$$\|V_{n+1}\| \leq \|V_n\| \sqrt{1 - 2\alpha s_n + s_n^2 M^2} + \frac{MLs_n^2}{2}.$$

6. Using the discrete version of Grönwall's lemma recalled in the "Auxiliary results" section, show that if $\sum_n s_n^2$ converges, then X_n tends to $X_{\min}$ as $n \to +\infty$.

4.3 A noisy version (20)

In this section, we fix the dimension $d = 1$ and we consider the ODE

$$\boxed{x_\varepsilon'(t) = -\mathsf{E}'(x_\varepsilon(t)) + \varepsilon A(t), \quad x_\varepsilon(0) = x_0} \tag{4}$$

where $t \mapsto x_\varepsilon(t)$ is an unknown function with real values, E satisfies the same assumptions as before (but in addition, we assume E to be of class C^2), ε is some fixed real parameter and A is a continuous function such that

$$T \mapsto \left| \int_0^T A(t)dt \right| \text{ is bounded.}$$

Let $\bar{x}$ be the solution to (4) when $\varepsilon = 0$. We look for an expression of x_ε as

$$x_\varepsilon = \bar{x} + \varepsilon \widetilde{x} + O(\varepsilon^2).$$

1. Write down (without rigorous justification) the ODE satisfied by $\tilde{x}$.

2. Write down an expression for $\tilde{x}$.

3. Show that $\widetilde{x}(t)$ is bounded as $t \to +\infty$.

Auxiliary results

- $(\arctan)'(x) = (1+x^2)^{-1}$
- If $t \mapsto F(t)$ is a C^1 function with values in $\mathbb{R}^d$, then the derivative of $\|F(t)\|^2$ is given by

$$\frac{d}{dt}\|F(t)\|^2 = 2\langle F(t), F'(t)\rangle$$

- For any real number x, we have

$$\lim_{n+\infty}\left(1+\frac{x}{n}\right)^n = e^x$$

- **Discrete Grönwall's lemma.** Let $\{u_n\}_n, \{a_n\}_n, \{b_n\}_n$ be sequences of non-negative numbers such that:

$$u_{n+1} \leq a_n u_n + b_n,$$

  then we have, for $n \geq 1$

$$u_n \leq e^{\sum_{k=0}^{n-1}\ln(a_k)}\left(u_0 + \sum_{k=0}^{n-1} b_k\right)$$