MATH60005/70005: Optimisation Coursework (AY 2526)

Instructions: read this first![1]

- This coursework has a total of 20 marks and accounts for 10% of the module.
- **Submission deadline:** Monday, November 24, 13:00 UK time, via Blackboard dropbox.
- **Submit a single PDF file typeset in LaTeX, Word, or similar.**
- The coursework requires basic coding capabilities that are available in all modern languages. **You are welcome to use any language of your preference**. You don't need to include your code in the submission, we want to assess your answers/plots/mathematical reasoning.
- **Marking criteria:** Full marks will be awarded for work that 1) is mathematically correct, 2) shows an understanding of material presented in lectures, 3) gives details of all calculations, implementation, and reasoning, and 4) is presented in a logical and clear manner.
- Do not discuss your answers publicly via our forum. If you have any queries regarding your interpretation of the questions, please contact the lecturer at `dkaliseb@imperial.ac.uk`. Consulting with the lecturer excludes debugging code or checking whether an answer is correct.
- Beware of plagiarism regulations; the College takes this very seriously. This is a **group-based assessment** with groups of 1 to 3 students. **Make a single group submission displaying the CID of every group member on the front page. Do not include your name.**

Part I: Gradient-Enhanced Polynomial Regression

Consider the task of approximating a function $V : \mathbb{R}^d \mapsto \mathbb{R}$ from a set of m random samples $\mathcal{D} = \{\mathbf{x}_i, V(\mathbf{x}_i)\}_{i=1}^m$ via linear least squares approximation. We will fit polynomial models of the form

$$V_\theta(\mathbf{x}) = \sum_{j=1}^{n} \theta_j \phi_j(\mathbf{x}) = \theta^\top \Phi(\mathbf{x}) \,,$$

where $\theta = (\theta_1, \ldots, \theta_n)^\top$ are the expansion coefficients to be learned, and $\Phi(\mathbf{x}) = (\phi_1(\mathbf{x}), \ldots, \phi_n(\mathbf{x}))^\top$ is a known set of polynomial basis functions in $\mathbb{R}^d$ to be specified later.

I.a) **[3 marks]** Formulate the approximation problem above as a linear least squares of the form

$$\min_{\theta \in \mathbb{R}^n} \|\mathbf{A}\theta - \mathbf{b}\|_2^2 \,,$$

providing general but precise expressions for $\mathbf{A}$ and $\mathbf{b}$.

[1] The ability to read and follow instructions is a highly valued skill in the job market and elsewhere.

Using your previous formulation, fit a model for the 1-d function data given in `trainingIa.dat`[2]. Here, the set $\Phi(x) = (1, x, x^2, \ldots, x^{n-1})$ corresponds to the 1-d monomial basis up to degree $n-1$. Present a plot displaying MSE $:= \frac{1}{m_{val}} \sum_{k=1}^{m_{val}} |V_\theta(x_k) - V(x_k)|^2$ in the validation data versus degree n, with $n = 1, \ldots, 20$. Use all the available training data, and measure your validation errors using `validationIa.dat`. What is the smallest n for which you achieve an error of 10^{-3} on the validation test? Fixing the value of n, plot the MSE in validation versus number of training points. Show the function that is learned at that value of n.

I.b) **[4 marks]** Now, assume that your training dataset has been enriched with derivative information, $\mathcal{D} = \{\mathbf{x}_i, V(\mathbf{x}_i), \nabla V(\mathbf{x}_i)\}_{i=1}^m$, available in the file `trainingIb.dat`. Using the same model $V_\theta(\mathbf{x})$ and monomial basis as in part I.a), reformulate the linear least squares problem to incorporate this data. Repeat the plots and analysis, and compare both regressions using the validation dataset `validationIa.dat`. What do you observe?

Part II: A First Approach to Dynamic Optimisation

Consider a discrete dynamical process of the form

$$\begin{aligned} x_0 &= \bar{x} \in \mathbb{R}\,, \\ x_i &= a x_{i-1} + b u_i\,, \quad i = 1, \ldots, N\,, \end{aligned} \tag{S}$$

where $a, b \in \mathbb{R}$. The variables x_i and u_i denote the internal state of the system and a control variable at discrete time i, respectively. The sequence $\mathbf{x}_{\bar{x}}^{\mathbf{u}} := \{x_i\}_{i=0}^N \in \mathbb{R}^{N+1}$ is the trajectory of the system departing from the initial condition $\bar{x}$ associated with the control sequence $\mathbf{u} := \{u_i\}_{i=1}^N \in \mathbb{R}^N$.

Given an initial condition $\bar{x}$, and parameters a, b, and N, our goal is to find an optimal sequence of controls $\mathbf{u}$ which drives the trajectory of the system $\mathbf{x}_{\bar{x}}^{\mathbf{u}}$ to zero while balancing the amount of control energy spent on this task. We express our goal as the **dynamic optimisation problem**

$$\begin{aligned} &\min_{\mathbf{u} \in \mathbb{R}^N} \ \|\mathbf{x}_{\bar{x}}^{\mathbf{u}}\|_2^2 + \gamma \|\mathbf{u}\|_2^2\,, \qquad \gamma > 0\,, \\ &\text{subject to} \\ &\qquad x_0 = \bar{x}\,, \\ &\qquad x_i = a x_{i-1} + b u_i\,, \quad i = 1, \ldots, N. \end{aligned} \tag{DO}$$

II.a) **[3 marks]** Reformulate (DO) as a regularised linear least squares problem for $\mathbf{u}$. Discuss the existence and uniqueness of an optimal solution $\mathbf{u}^*$ to this problem. Show that any $\mathbf{u}^*$ solving the associated regularised problem satisfies $\|\mathbf{u}^*\| \leq \|\mathbf{u}\|$ for any $\mathbf{u}$ solving the unregularised linear least squares problem (that is, with $\gamma = 0$).

For $N = 50$, $a = 1$, $b = -0.01$, and $\bar{x} = 1$, generate two plots: one for the optimal control signals $\mathbf{u}^*$, and another for the associated optimal trajectories $\mathbf{x}_{\bar{x}}^{\mathbf{u}^*}$, for $\gamma = 10^{-3}, 10^{-2}, 0.1, 1$. What is the effect of increasing γ on both the control and the trajectories?

[2] Data format is $(x_i, V(x_i))$

II.b) **[2 marks]** Now, in addition to the system (S), consider a second system (S') given by

$$\begin{aligned} y_0 &= \bar{y} \in \mathbb{R}\,, \\ y_i &= c y_{i-1} + d v_i\,, \quad i = 1, \ldots, N\,, \end{aligned} \tag{S'}$$

and consider the problem

$$\min_{(\mathbf{u},\mathbf{v}) \in \mathbb{R}^{2N}} \; \|\mathbf{x}^{\mathbf{u}}_{\bar{x}}\|_2^2 + \|\mathbf{y}^{\mathbf{v}}_{\bar{y}}\|_2^2 + \gamma \|\mathbf{u} - \mathbf{v}\|_2^2\,, \qquad \gamma > 0\,, \tag{MO}$$

that is, we want to drive both $\mathbf{x}$ and $\mathbf{y}$ to zero with the respective control signals $\mathbf{u}$ and $\mathbf{v}$ close to each other. Formulate (MO) as a regularised linear least squares problem. For $N = 50$, $a = 1$, $b = 0.05$, $\bar{x} = 1$, $c = 0.2$, $d = -0.5$, and $\bar{y} = 1$, solve the regularised linear least squares problem for $\gamma = 1$ and show two plots for state trajectories and controls.

II.c) **[4 marks]** Using the regularised least squares formulation from the previous part, if every parameter of the problem is fixed except for γ, we consider the following functionals evaluated at the corresponding optimal solution:

$$\mathcal{J}_1(\gamma) := \|\mathbf{x}^*(\gamma)\|_2^2 + \|\mathbf{y}^*(\gamma)\|_2^2\,, \qquad \mathcal{J}_2(\gamma) := \|\mathbf{u}^*(\gamma) - \mathbf{v}^*(\gamma)\|_2^2\,.$$

Using the same parameters as in part II.b), solve the regularised least squares problem for values of $10^{-5} \leq \gamma \leq 10^5$ (take the exponent increasing in steps of 0.1). Show a plot of $\mathcal{J}_1(\gamma)$ and $\mathcal{J}_2(\gamma)$ versus γ. Plot the curve $(\mathcal{J}_1(\gamma), \mathcal{J}_2(\gamma))$. This curve is known as the *Pareto front* and is a relevant object in multi-objective optimisation. Draw a conclusion from these plots regarding the impact of γ on the total cost.

II.d) **[4 marks]** We wish to promote sparsity in the control penalty by considering an ℓ_1 norm penalty in the cost:

$$\min_{(\mathbf{u},\mathbf{v}) \in \mathbb{R}^{2N}} \; \|\mathbf{x}^{\mathbf{u}}_{\bar{x}}\|_2^2 + \|\mathbf{y}^{\mathbf{v}}_{\bar{y}}\|_2^2 + \gamma_2 \|\mathbf{u} - \mathbf{v}\|_2^2 + \gamma_1 \|\mathbf{u} - \mathbf{v}\|_1\,, \qquad \gamma_1, \gamma_2 \geq 0\,.$$

However, we have not yet discussed how to deal with the non-differentiability of the ℓ_1 norm at the origin (we will do so by the end of term!). Instead, we propose the following approximation to the ℓ_1 norm ($\epsilon > 0$):

$$\|\mathbf{u}\|_1 \approx \sum_{i=1}^{N} \mathcal{L}_\epsilon(u_i)\,, \qquad \mathcal{L}_\epsilon(u_i) := \begin{cases} \frac{1}{2} u_i^2 & \text{if } |u_i| \leq \epsilon \\ \epsilon\left(|u_i| - \frac{1}{2}\epsilon\right) & \text{otherwise.} \end{cases}$$

Explain in your own words (you may use plots) the meaning of $\mathcal{L}_\epsilon(u_i)$ as a regulariser. Is it a differentiable function?

Implement a gradient descent method with constant stepsize (describe all your settings) to find the optimal solution to

$$\min_{(\mathbf{u},\mathbf{v}) \in \mathbb{R}^{2N}} \; \|\mathbf{x}^{\mathbf{u}}_{\bar{x}}\|_2^2 + \|\mathbf{y}^{\mathbf{v}}_{\bar{y}}\|_2^2 + \gamma_2 \|\mathbf{u} - \mathbf{v}\|_2^2 + \gamma_1 \sum_{i=1}^{N} \mathcal{L}_\epsilon(u_i - v_i)\,, \qquad \gamma_1, \gamma_2, \epsilon > 0\,.$$

Using the same parameters as in part IIb), compare trajectories and controls plots for

i) $\epsilon = 1$, $\gamma_2 = 1$, $\gamma_1 = 0$, (same as part II.b)

ii) $\epsilon = 1$, $\gamma_2 = 0$, $\gamma_1 = 1$.

iii) $\epsilon = .1$, $\gamma_2 = 0$, $\gamma_1 = 1$.

What do you observe?