MATH60005/70005: Optimisation (Autumn 23-24)

Instructions: read this first!

- This coursework has a total of 20 marks and accounts for 10% of the module.
- Students who want to take the final exam (90%) must submit this coursework.
- **Submission deadline:** Monday November 20th, 13:00 UK time, via Blackboard drop box.
- Submit a single file, ideally a pdf typed in LaTeX or similar.
- **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 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
- Beware of plagiarism regulations. This is a **group-based assessment** with groups from 1 to 3 students. **Make a single submission for the coursework indicating in the front page the CID of every group member. Do not include your name.**

Questions

Part I: Unconstrained Optimisation (6 marks)

Let $f(x,y) = x^2 - 2xy^2 + \frac{1}{2}y^4$.

i) [**3 marks**] Is the function f coercive? Explain your answer.

ii) [**3 marks**] Find the stationary points of f and classify them according to whether they are saddle points, strict/nonstrict local/global minimum/maximum points.

Part II: Linear Least Squares - Denoising (8 marks)

You are given the noisy signal $\mathbf{s} \in \mathbb{R}^{1000}$ shown in Figure 1 (corresponding to the file `signal.mat`), and the goal is to use linear least squares to denoise it.

i) [**3 marks**] Write a regularised least squares formulation for the denoising problem of the form

$$\min_{\mathbf{x}\in\mathbb{R}^{1000}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda\|\mathbf{L}\mathbf{x}\|_2^2\,, \qquad \lambda > 0\,,$$

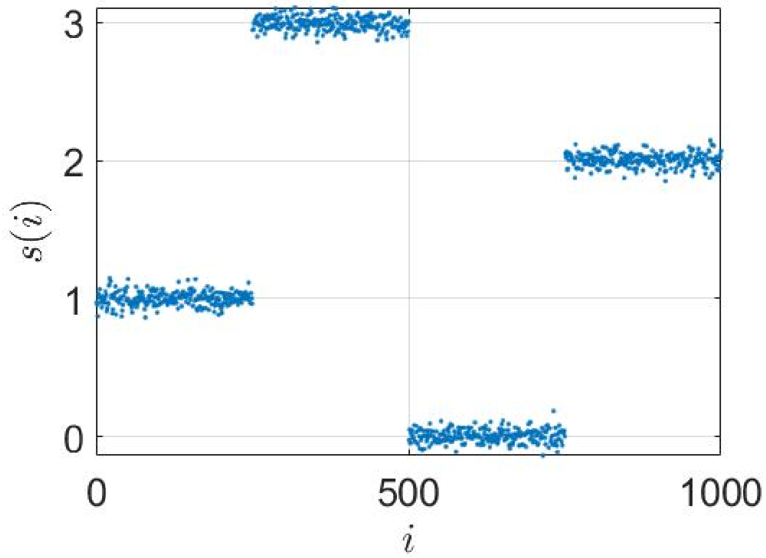

Figure 1: A noisy signal composed of piecewise constant components.

where $\|\mathbf{L}\mathbf{x}\|^2 = \sum_{i=1}^{999}(x_i - x_{i+1})^2$. Give precise definitions for $\mathbf{A}, \mathbf{b}$, and $\mathbf{L}$. Cast an equivalent ordinary least squares problem of the form

$$\min_{\mathbf{x}\in\mathbb{R}^{1000}} \|\hat{\mathbf{A}}\mathbf{x} - \hat{\mathbf{b}}\|_2^2\,,$$

giving precise definitions for $\hat{\mathbf{A}}$ and $\hat{\mathbf{b}}$. Plot the solution of the regularised least squares problems for values of $\lambda = 0.1, 1, 10, 100, 1000$. Describe what you observe in terms of denoising level.

ii) We will now develop an alternative solution to the denoising problem. We consider instead the problem

$$\min_{\mathbf{x}\in\mathbb{R}^{1000}} \|\hat{\mathbf{A}}\mathbf{x} - \hat{\mathbf{b}}\|_1\,,$$

with the same $\hat{\mathbf{A}}$ and $\hat{\mathbf{b}}$ as before, and where $\|\mathbf{x}\|_1 = \sum |x_i|$. Note that this problem is nonsmooth and we cannot directly apply least squares or gradient-based techniques.

iia) [**3 marks**] We will construct an algorithm to approximate the solution of the problem above. For this, we consider the weighted least squares problem

$$\min_{\mathbf{x}\in\mathbb{R}^{1000}} \sum_{i=1}^{1999} w_i |\hat{\mathbf{a}}_i{}^\top \mathbf{x} - \hat{b}_i|^2\,,$$

where $\hat{\mathbf{a}}_i{}^\top$ denotes the $i-$th row of the matrix $\hat{\mathbf{A}}$ and $\mathbf{w} \in \mathbb{R}^{1999}$ is a given weight vector with positive entries. Proceeding similarly as in the linear least squares problem, derive an expression for the solution of this problem in terms of $\hat{\mathbf{A}}, \hat{\mathbf{b}}$ and $\mathbf{W}$, which is a diagonal matrix with $\mathbf{w}$ in the diagonal. We call this solution $\mathbf{x}^* = \mathbf{x}^*(\hat{\mathbf{A}}, \hat{\mathbf{b}}, \mathbf{W})$.

iib) **[2 marks]** We now construct the following algorithm:

Step 0: Set a vector $\mathbf{w}^0 = \mathbf{1} \in \mathbb{R}^{1999}$, a vector full of 1's. Set $k = 0$, and a tolerance δ.

Step 1: Solve $\mathbf{x}^k = \mathbf{x}^*(\hat{\mathbf{A}}, \hat{\mathbf{b}}, \mathbf{W^k})$.

Step 2: Update the weights according to

$$w_i^{k+1} = \frac{1}{\max\{\delta, |\hat{\mathbf{a}}_i{}^\top \mathbf{x}^k - \hat{b}_i|\}}.$$

Step 3: Set $k = k + 1$ and go back to Step 1.

Plot the final solution $\mathbf{x}^*$ of this algorithm for $\delta = 10^{-5}$, 100 iterations, and values of $\lambda = 0.1, 1, 10, 100, 1000$. This iterative procedure approximates the solution of the problem

$$\min_{\mathbf{x}\in\mathbb{R}^{1000}} \|\hat{\mathbf{A}}\mathbf{x} - \hat{\mathbf{b}}\|_1 .$$

Discuss the differences you observe in your results with respect to the regularized least squares from part i).

Part III: Gradient Descent (6 marks) Consider the problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \left\{ f(\mathbf{x}) \equiv \sum_{i=1}^{m} \sqrt{\left(\mathbf{a}_i^\top \mathbf{x} - b_i\right)^2 + \eta^2} \right\},$$

where $\mathbf{a}_i \in \mathbb{R}^n, b_i \in \mathbb{R}, i = 1, 2, \ldots, m$, and $\eta > 0$. If We define $\mathbf{A} \in \mathbb{R}^{m\times n}$ and the vector $\mathbf{b} \in \mathbb{R}^m$ as

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_1^\top \\ \mathbf{a}_2^\top \\ \vdots \\ \mathbf{a}_m^\top \end{pmatrix}, \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix},$$

and the function $g : \mathbb{R}^m \to \mathbb{R}$ as $g(\mathbf{y}) = \sum_{i=1}^m \sqrt{y_i^2 + \eta^2}$, we can express

$$\nabla f(\mathbf{x}) = \mathbf{A}^\top \nabla g(\mathbf{A}\mathbf{x} - \mathbf{b}), \quad \text{and} \quad \nabla^2 f(\mathbf{x}) = \mathbf{A}^\top \nabla^2 g(\mathbf{A}\mathbf{x} - \mathbf{b})\mathbf{A}.$$

i) **[2 marks]** Show that $\nabla^2 f(\mathbf{x}) \succeq 0$ for all $\mathbf{x} \in \mathbb{R}^n$.

Choose only one of the following two questions:

iiA) **[4 marks]** Show that if $\mathbf{A}$ has full column rank, then there exists a unique global minimiser.
Hint: for existence of a minimiser, you can use that for any $\mathbf{z} \in \mathbb{R}^m$, it holds that $\|\mathbf{z}\|_1 \geq \|\mathbf{z}\|_2$. For uniqueness, use the following result: if a function f is such that $\nabla^2 f(\mathbf{x}) \succ 0$ for all $\mathbf{x} \in \mathbb{R}^n$, then a stationary point of f is necessarily a strict global minimum.

iiB) **[4 marks]** Show that $f \in C_L^{1,1}$ with $L = \frac{\|A\|^2}{\eta}$.
Hint: you can use that for two matrices $\mathbf{A}$ and $\mathbf{B}$ of compatible dimensions, it holds that $\|\mathbf{AB}\| \leq \|\mathbf{A}\|\|\mathbf{B}\|$.