MATH60005 Optimisation Coursework

CID: 02024600, 02045099, 02025959

Part 1

i)

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

f is coercive if it satisfies

$$\lim_{\|\mathbf{x}\| \longrightarrow \infty} f(\mathbf{x}) = \infty$$

for any $\mathbf{x} = (x,y)^\top$

We will consider the trajectory $\mathbf{x} = (\alpha^2, \alpha)^\top$ as $\alpha \longrightarrow \infty$.

On this trajectory,

$$\begin{aligned} f(\mathbf{x}) &= (\alpha^2)^2 - 2(\alpha^2)(\alpha)^2 + \frac{1}{2}(\alpha)^4 \\ &= \alpha^4 - 2\alpha^4 + \frac{1}{2}\alpha^4 \\ &= -\frac{1}{2}\alpha^4 \end{aligned}$$

Hence, as $\alpha \longrightarrow \infty$, $\|\mathbf{x}\| \longrightarrow \infty$ but $f(\mathbf{x}) \longrightarrow -\infty$. Thus, f is not coercive as it does not go to ∞ for all large $\mathbf{x}$.

ii)

To find the stationery points of f we first consider

$$\begin{aligned} \nabla f(\mathbf{x}) &= \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \\ &= \begin{bmatrix} 2x - 2y^2 \\ 2y^3 - 4xy \end{bmatrix} \end{aligned}$$

Hence, we look for solutions of the system of equations given by (1)

$$\begin{aligned} 2x - 2y^2 &= 0 \\ 2y^3 - 4xy &= 0 \end{aligned} \tag{1}$$

The first equation has solution $x = y^2$ and the second equation has two solutions, $4x = y^2$ and $y = 0$. So f has one stationery point at $(0,0)$.

$$\begin{aligned} \nabla^2 f(\mathbf{x}) &= \begin{bmatrix} 2 & -4y \\ -4y & 6y^2 - 4x \end{bmatrix} \\ \implies \nabla^2 f(0,0) &= \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \succeq 0 \end{aligned}$$

$\nabla^2 f(0,0)$ is a positive semi-definite matrix so we can conclude that $(0,0)$ is either a local minimum or a saddle point. We will consider two trajectories to determine the nature of the stationery point.

Here $(0,0) \in \mathbb{R}^2$ is a local minimum of f over $\mathbb{R}^2$ if there exists $r > 0$ for which $f(\mathbf{x}) \geq f(0,0)$ for any $\mathbf{x} \in \mathbb{R}^2 \cap B((0,0), r)$.

For our purposes, we take $0 < \varepsilon < r$:

Trajectory 1: $\mathbf{x} = (\varepsilon^2, \varepsilon)^\top$

$$f(\varepsilon^2, \varepsilon) = -\frac{1}{2}\varepsilon^4 < 0$$

Trajectory 2: $\mathbf{x} = (0, \varepsilon)^\top$

$$f(0, \varepsilon) = \frac{1}{2}\varepsilon^4 > 0$$

Hence, we conclude $(0,0)$ is a saddle point.

Part 2

i)

From the lecture notes, our denoising problem is of the form

$$\min_{\mathbf{x} \in \mathbb{R}^{1000}} \|\mathbf{x} - \mathbf{s}\|^2 + \lambda \|\mathbf{L}\mathbf{x}\|^2$$

So, $\mathbf{A} = \mathbf{I} \in \mathbb{R}^{1000 \times 1000}$, $\mathbf{b} = \mathbf{s} \in \mathbb{R}^{1000}$, and

$$\mathbf{L} = \begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -1 \end{bmatrix} \in \mathbb{R}^{999 \times 1000} \tag{2}$$

This can be cast as an equivalent ordinary least squares problem

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

where

$$\hat{\mathbf{A}} = \begin{bmatrix} I \\ \lambda^{\frac{1}{2}}\mathbf{L} \end{bmatrix} \in \mathbb{R}^{1999 \times 1000} \quad \mathbf{I} \in \mathbb{R}^{1000 \times 1000} \quad \mathbf{L} \in \mathbb{R}^{999 \times 1000}$$

$$\hat{\mathbf{b}} = \begin{bmatrix} \mathbf{s} \\ \mathbf{0} \end{bmatrix} \quad \mathbf{s} \in \mathbb{R}^{1000} \quad \mathbf{0} \in \mathbb{R}^{999}$$

The solution to the denoising problem can be seen in Figure 1. For small λ the solution still exhibits a lot of noise but maintains the clear step-like nature of the noisy signal. In contrast, for large λ the solution becomes a clear smooth function, losing the step-like nature of the nosy signal but maintaining the peak and troughs. This smoothing of the curve comes from the assumption that the denoised function is a smooth curve in order for us to pick L as we did in (2). At first glance, the optimal value is $\lambda = 10$ as it balances the step-like nature with the denoising, showing only slight rounding at the edges of 'steps'.

iia)

$$\begin{aligned} \sum_{i=1}^{1999} w_i |\hat{\mathbf{a}}_i^\top \mathbf{x} - \hat{b}_i|^2 &= \|\mathbf{W}^{\frac{1}{2}}(\hat{\mathbf{A}}\mathbf{x} - \hat{\mathbf{b}})\|^2 \\ &= \|\mathbf{W}^{\frac{1}{2}}\hat{\mathbf{A}}\mathbf{x} - \mathbf{W}^{\frac{1}{2}}\hat{\mathbf{b}}\|^2 \end{aligned}$$

From the lecture notes, this linear least squares problem has solution

$$\begin{aligned} \mathbf{x}^* &= (\hat{\mathbf{A}}^\top \mathbf{W} \hat{\mathbf{A}})^{-1} \hat{\mathbf{A}}^\top \mathbf{W}^{\frac{1}{2}} (\mathbf{W}^{\frac{1}{2}} \hat{\mathbf{b}}) \\ &= (\hat{\mathbf{A}}^\top \mathbf{W} \hat{\mathbf{A}})^{-1} \hat{\mathbf{A}}^\top \mathbf{W} \hat{\mathbf{b}} \end{aligned}$$

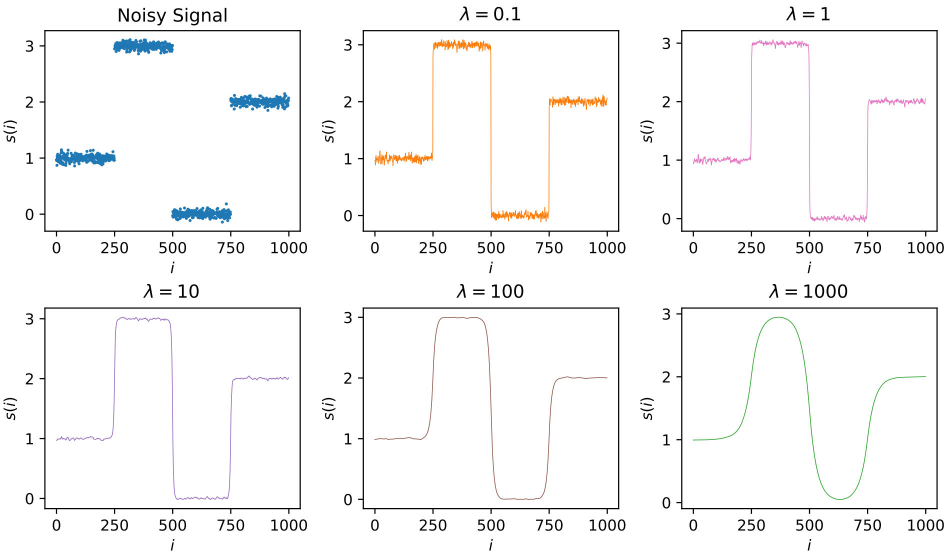

Figure 1: Solution to Denoising Problem for $\lambda = 0.1, 1, 10, 100, 1000$

iib)

Using the algorithm given in the question we approximate the solution of

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

as seen in Figure 2.

The main difference between this solution and the solution to the denoising problem is the decrease in smoothing the graph at the ends of 'steps'. This solution provides graphs with clearly defined steps even for high values of λ. In our method for (iib) we have not assumed the smoothness of the function as we did in (iia), hence we see graphs that much more closely show a step function similar to the input s.

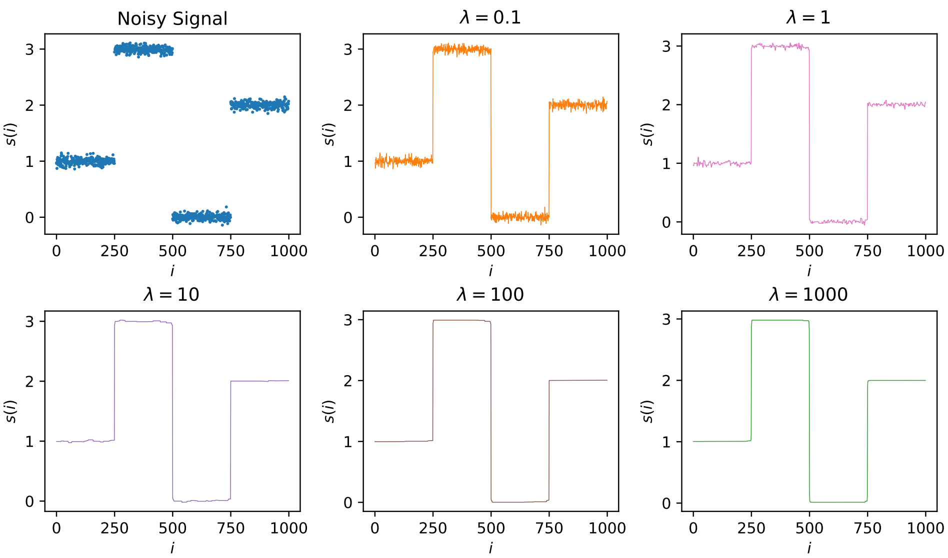

Figure 2: Solution to Linear Least Squares Problem for $\lambda = 0.1, 1, 10, 100, 1000$

Part 3

i)

$$g(\mathbf{y}) := \sum_{i=1}^{m} \sqrt{y_i^2 + \eta^2}$$

$$\nabla g(\mathbf{y}) = \left(\frac{y_1}{\sqrt{y_1^2 + \eta^2}}, \dots, \frac{y_m}{\sqrt{y_m^2 + \eta^2}} \right)$$

$$\begin{aligned}
(\nabla^2 g(\mathbf{y}))_{i,i} &= \frac{\partial}{\partial y_i} \left(\frac{y_i}{\sqrt{y_i^2 + \eta^2}} \right) \\
&= \frac{1}{\sqrt{y_i^2 + \eta^2}} - \frac{y_i^2}{(y_i^2 + \eta^2)^{\frac{3}{2}}} \\
&= \frac{y_i^2 + \eta^2}{(y_i^2 + \eta^2)^{\frac{3}{2}}} - \frac{y_i^2}{(y_i^2 + \eta^2)^{\frac{3}{2}}} \\
&= \frac{\eta^2}{(y_i^2 + \eta^2)^{\frac{3}{2}}}
\end{aligned} \tag{3}$$

Note also that $(\nabla^2 g(\mathbf{y}))_{i,j} = 0$ for $i \neq j$

$$\implies \nabla^2 g(\mathbf{y}) = \begin{bmatrix} \frac{\eta^2}{(y_1^2+\eta^2)^{\frac{3}{2}}} & & \\ & \ddots & \\ & & \frac{\eta^2}{(y_m^2+\eta^2)^{\frac{3}{2}}} \end{bmatrix}$$

Here, we can see the eigenvalues of this matrix are the diagonal elements i.e. $\frac{\eta^2}{(y_i^2+\eta^2)^{\frac{3}{2}}}$, which are all non-negative. From lecture notes, we can deduce that $\nabla^2 g(\mathbf{y})$ is positive semi-definite or $\forall \mathbf{x} \in \mathbb{R}^n$:

$$\mathbf{x}^\top \nabla^2 g(\mathbf{y}) \mathbf{x} \geq 0$$

Now consider:

$$\begin{aligned}
\mathbf{x}^\top \nabla^2 f(\mathbf{x}) \mathbf{x} &= \mathbf{x}^\top \mathbf{A}^\top \nabla^2 g(\mathbf{A}\mathbf{x} - \mathbf{b}) \mathbf{A}\mathbf{x} \\
&= (\mathbf{A}\mathbf{x})^\top \nabla^2 g(\mathbf{A}\mathbf{x} - \mathbf{b}) \mathbf{A}\mathbf{x}
\end{aligned} \tag{4}$$

Relabelling $\mathbf{A}\mathbf{x} = \mathbf{w}$

$$(\mathbf{A}\mathbf{x})^\top \nabla^2 g(\mathbf{A}\mathbf{x} - \mathbf{b}) \mathbf{A}\mathbf{x} = \mathbf{w}^\top \nabla^2 g(\mathbf{A}\mathbf{x} - \mathbf{b}) \mathbf{w} \geq 0 \tag{5}$$

Hence, $\nabla^2 f(\mathbf{x}) \succcurlyeq 0$ as required.

iiB)

Firstly, consider $\left\| \nabla^2 g(\mathbf{A}\mathbf{x} - \mathbf{b}) \right\|$:

$$\begin{aligned}
\left\| \nabla^2 g(\mathbf{A}\mathbf{x} - \mathbf{b}) \right\| &= \left(\lambda_{\max} \left[(\nabla^2 g(\mathbf{A}\mathbf{x} - \mathbf{b}))^\top \nabla^2 g(\mathbf{A}\mathbf{x} - \mathbf{b}) \right] \right)^{\frac{1}{2}} \\
&= \left(\max_i \left[\frac{\eta^4}{(y_i^2 + \eta^2)^3} \right] \right)^{\frac{1}{2}} \\
&\leq \left(\frac{\eta^4}{(\eta^2)^3} \right)^{\frac{1}{2}} \\
&= \frac{1}{\eta}
\end{aligned} \tag{6}$$

Now consider $\left\|\nabla^2 f(\mathbf{x})\right\|$:

$$\begin{aligned}
\left\|\nabla^2 f(\mathbf{x})\right\| &= \left\|\mathbf{A}^\top \nabla^2 g(\mathbf{A}\mathbf{x}-\mathbf{b})\mathbf{A}\right\| \\
&\leq \left\|\mathbf{A}^\top\right\| \left\|\nabla^2 g(\mathbf{A}\mathbf{x}-\mathbf{b})\right\| \|\mathbf{A}\| \\
&= \|\mathbf{A}\|^2 \left\|\nabla^2 g(\mathbf{A}\mathbf{x}-\mathbf{b})\right\| \\
&\leq \frac{\|\mathbf{A}\|^2}{\eta}
\end{aligned} \tag{7}$$

Finally, from lectures we know that:

$$\left\|\nabla^2 f(\mathbf{x})\right\| \leq L = \frac{\|\mathbf{A}\|^2}{\eta} \iff f \in C_L^{1,1}$$
