Optimisation Coursework

02017784,

November 10, 2023

Part I

i)

We show the function, $f(x,y) = x^2 - 2xy^2 + \frac{1}{2}y^4$ is not coercive. Indeed, taking the trajectory $(x,y) = (\alpha^2, \alpha)$ with $\alpha \in \mathbb{R}$. As $\alpha \to \infty$, $\|(\alpha^2,\alpha)\| \to \infty$, but

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

Thus $\lim\limits_{\alpha\to\infty} f(\alpha^2,\alpha) = -\infty$, so the function is not coercive.

ii)

We first find the gradient and hessian of f,

$$\begin{aligned} \nabla f(x,y) &= \begin{pmatrix} 2x - 2y^2 \\ -4xy + 2y^3 \end{pmatrix} \\ \nabla^2 f(x,y) &= \begin{pmatrix} 2 & -4y \\ -4y & -4x + 6y^2 \end{pmatrix} \end{aligned}$$

So the stationary points of f, are the solutions to

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

From the first equation we have $x = y^2$ and from the second we have either $y = 0$ or $2x = y^2$. So either $x = 0^2 = 0$ or $x = 2x = 0$, thus the only stationary point is at the origin $x^* = (0,0)$. The hessian at the origin is

$$\nabla^2 f(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

this is diagonal and with eigenvalues $\lambda = 2, 0$ which are both non-negative, so the hessian is positive semi-definite. Thus the stationary point is either a local minimum or a saddle point. We now show it is a saddle point by comparing two trajectories. Consider the trajectories (α^2,α) and $(0,\beta)$ as $\alpha,\beta \to 0^+$ from above.

From `i)`, we know $f(\alpha^2,\alpha) = -\frac{1}{2}\alpha^4 < 0$ for positive α, and $\lim\limits_{\alpha\to 0^+} f(\alpha^2,\alpha) = (0,0) = x^*$. Also, $f(0,\beta) = \frac{1}{2}\beta^4 > 0$ for positive β, and $\lim\limits_{\beta\to 0^+} f(0,\beta) = (0,0) = x^*$. Hence our function approaches $x^* = (0,0)$ from both above and below, so is neither a local minima or maxima and $(0,0)$ is a saddle point.

Part II

i

To denoise the signal, we consider the regularised least squares formulation,

$$\min_{\mathbf{x}\in\mathbb{R}^{1000}} \|\mathbf{x}-\mathbf{s}\|_2^2 + \lambda\|\mathbf{L}\mathbf{x}\|_2^2$$

With $\mathbf{A} = \mathbf{I}_{1000} \in \mathbb{R}^{1000\times 1000}$ the identity matrix, $\mathbf{b} = \mathbf{s} \in \mathbb{R}^{1000}$ the noisy signal and $\mathbf{L} \in \mathbb{R}^{999\times 1000}$ given by,

$$\mathbf{L} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 0 \\ 0 & 0 & 0 & \cdots & 1 & -1 \end{pmatrix}$$

From the lectures, the solution is given by,

$$\mathbf{x}_{\mathbf{RLS}} = (\mathbf{I} - \lambda\mathbf{L}^T\mathbf{L})^{-1}\mathbf{s}$$

We can also equivocate this to an ordinary least squares problem of the form,

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

With $\hat{\mathbf{A}} = \begin{pmatrix} \mathbf{I}_{1000} \\ \mathbf{L} \end{pmatrix} \in \mathbb{R}^{1999\times 1000}$ and $\hat{\mathbf{b}} = \begin{pmatrix} \mathbf{s} \\ \mathbf{0} \end{pmatrix} \in \mathbb{R}^{1999}$

We can plot the **RLS** solution for different values of λ shown in `figure 1`. We can see that for small values of λ, the denoised signal closer fits our data but is more noisy, compared to larger values of λ, where the denoised signals are smoother but a worse fit to the noisy signal.

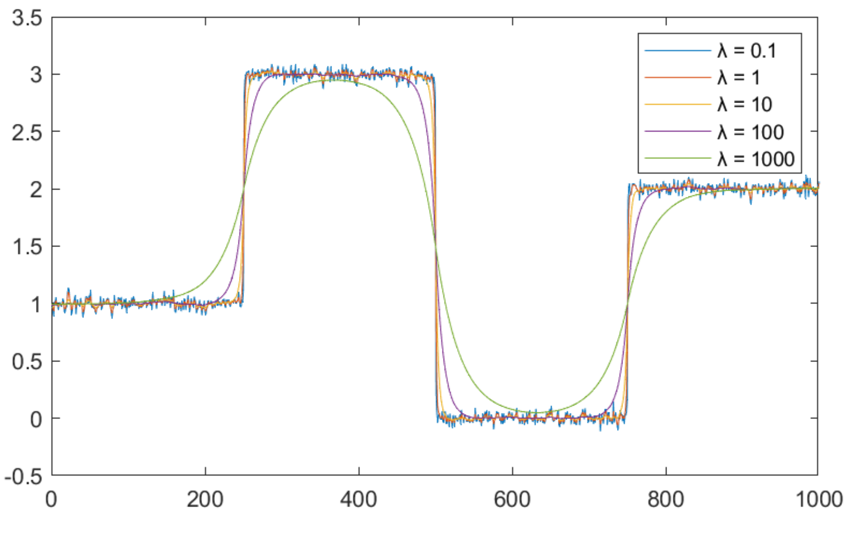

Figure 1: plot of the regularised least squares solutions for different values of λ

iia)

We consider the weighted least squares problem given by,

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

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

since $w_i > 0$. Writing $\mathbf{W} = \mathrm{Diag}(\mathbf{w})$ and $\mathbf{W}^{\frac{1}{2}} = \mathrm{Diag}\left(\sqrt{w_1}, \ldots, \sqrt{w_{1999}}\right)$, we can express the problem by,

$$\min_{\mathbf{x}\in\mathbb{R}^{1000}} \|\mathbf{W}^{\frac{1}{2}}\hat{\mathbf{A}}\mathbf{x} - \mathbf{W}^{\frac{1}{2}}\hat{\mathbf{b}}\|_2^2$$

The least squares solution for this is,

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

iib)

We employ this algorithm on `MATLAB`, with the plot for $\mathbf{x}^*$ shown in `figure 2` for different values of λ. The solutions are in general much sharper than those for the regularized least squares solution from `part 1`.

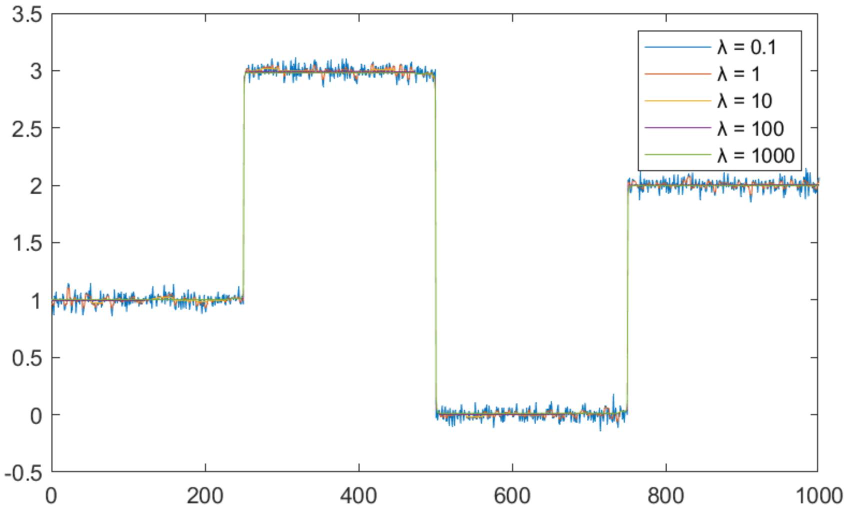

Figure 2: caption here

Part III

i)

We first seek a general form for $\nabla^2 g(\mathbf{y})$ for $\mathbf{y} = (y_1, \ldots, y_m) \in \mathbb{R}^m$.

$$\begin{aligned}\frac{\partial}{\partial y_i} g(\mathbf{y}) &= \frac{\partial}{\partial y_i} \sum_{j=1}^{m} \sqrt{y_j^2 + \eta^2} \\ &= \frac{y_i}{\sqrt{y_i^2 + \eta^2}}\end{aligned}$$

which is dependent on y_i only, so

$$\frac{\partial^2}{\partial y_i \partial y_j} g(\mathbf{y}) = \begin{cases} 0 & i \neq j \\ \frac{\eta^2}{\sqrt{y_i^2+\eta^2}^3} & i = j \end{cases}$$

Thus we can write $\nabla^2 g(\mathbf{y}) = \mathrm{Diag}\left(\frac{\eta^2}{\sqrt{y_1^2+\eta^2}^3}, \ldots, \frac{\eta^2}{\sqrt{y_1^2+\eta^2}^3}\right)$, a diagonal matrix. Since the matrix is symmetric and all diagonal elements are non-negative, for any $\mathbf{y} \in \mathbb{R}^m$ we can write $\nabla^2 g(\mathbf{y}) = \mathbf{S}^T\mathbf{S}$, with $\mathbf{S} = \mathrm{Diag}\left(\frac{\eta}{(y_1^2+\eta^2)^{3/4}}, \ldots, \frac{\eta}{(y_1^2+\eta^2)^{3/4}}\right)$. Now we let $\mathbf{z}, \mathbf{x} \in \mathbb{R}^n$, then

$$\begin{aligned}\mathbf{z}^T \nabla^2 f(\mathbf{x}) \mathbf{z} &= \mathbf{z}^T \mathbf{A}^T \nabla^2 g(\mathbf{A}\mathbf{x} - \mathbf{b}) \mathbf{A}\mathbf{z} \\ &= \mathbf{z}^T \mathbf{A}^T \mathbf{S}^T \mathbf{S} \mathbf{A} \mathbf{z} \\ &= (\mathbf{S}\mathbf{A}\mathbf{z})^T \mathbf{S}\mathbf{A}\mathbf{z} \\ &= \|\mathbf{S}\mathbf{A}\mathbf{z}\|^2 \\ &\geq 0\end{aligned}$$

Thus $\nabla^2 f(\mathbf{x}) \succeq 0$ for all $\mathbf{x} \in \mathbb{R}^n$.

iiB)

To show $f \in C_L^{1,1}$, we find a bound $\|\nabla^2 f(\mathbf{x})\| \leq L$. Using the result provided,

$$\begin{aligned}\|\nabla^2 f(\mathbf{x})\| &= \|\mathbf{A}^T \nabla^2 g(\mathbf{A}\mathbf{x} - \mathbf{b}) \mathbf{A}\| \\ &\leq \|\mathbf{A}^T\| \|\nabla^2 g(\mathbf{y})\| \|\mathbf{A}\|\end{aligned}$$

Where we have let $\mathbf{y} = \mathbf{A}\mathbf{x} - \mathbf{b} \in \mathbb{R}^m$. Now for any matrix $\mathbf{M}$, $\|\mathbf{M}\| = \sqrt{\lambda_{\max}(\mathbf{M}^T\mathbf{M})}$ with $\lambda_{\max}(\mathbf{M}^T\mathbf{M})$, the largest eigenvalue of $\mathbf{M}^T\mathbf{M}$. If $\mathbf{M}$ is symmetric, then $\lambda_{\max}(\mathbf{M}^T\mathbf{M}) = \lambda_{\max}(\mathbf{M}^2) = \lambda_{\max}(\mathbf{M})^2$ (result used in problem sheets). So applying this for $\mathbf{M} = \nabla^2 g(\mathbf{y})$,

$$\begin{aligned}\|\nabla^2 g(\mathbf{y})\| &= \lambda_{\max}(\nabla^2 g(\mathbf{y}))^2 \\ &= \max_{y_i} \left(\frac{\eta^2}{\sqrt{y_i^2 + \eta^2}^3}\right)^2 \\ &\leq \frac{\eta^2}{\sqrt{(\eta^2)^3}} \\ &\leq \frac{1}{\eta}\end{aligned}$$

Also, $\|\mathbf{A}\| = \sqrt{\lambda_{\max}(\mathbf{A}^T\mathbf{A})} = \sqrt{\lambda_{\max}(\mathbf{A}\mathbf{A}^T)} = \|\mathbf{A}^T\|$ since $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T$ share the same non-zero eigenvalues.

So altogether,

$$\|\nabla^2 f(\mathbf{x})\| \leq \frac{\|\mathbf{A}\|^2}{\eta}$$

So we have $f \in C_L^{1,1}$ with $L = \frac{\|\mathbf{A}\|^2}{\eta}$