Optimization

Exercises

1. Prove Rolle's Theorem.

Since f is a continuous function and $[a, b]$ is a compact set, it attains a maximum and a minimum. There therefore exists $\bar{x}^\star$, $\underline{x}^\star$ such that $f(\bar{x}^\star) \geq f(x)$ and $f(\underline{x}^\star) \leq f(x)$ for all $x \in [a, b]$. If $\bar{x}^\star, \underline{x}^\star \in (a, b)$, then these optima are interior and $f'(\bar{x}^\star) = f'(\underline{x}^\star) = 0$ and we are done. If $\bar{x}^\star, \underline{x}^\star \notin (a, b)$, then $\bar{x}^\star, \underline{x}^\star \in \{a, b\}$. Since $f(a) = f(b)$, $f(\bar{x}^\star) = f(\underline{x}^\star) = f(x)$ for all $x \in (a, b)$. The function is therefore constant and $f'(x) = 0$ for all $x \in [a, b]$. ■

2. Characterize the stationary point(s) of[1]

$$f(x_1, x_2) = x_1^2 + x_2^2$$

Are these points maxima, minima, or saddle points?

We look for points for which

$$\frac{\partial f(x_1, x_2)}{\partial x_1} = \frac{\partial f(x_1, x_2)}{\partial x_2} = 0$$

We have

$$\frac{\partial f(x_1, x_2)}{\partial x_1} = 2x_1 = 0 \qquad \frac{\partial f(x_1, x_2)}{\partial x_2} = 2x_2 = 0$$

which can only be satisfied for $x_1 = x_2 = 0$.

To check the concavity/convexity of the function, we construct the Hessian

$$\frac{\partial^2 f(x_1, x_2)}{\partial x_1^2} = 2 \qquad \frac{\partial^2 f(x_1, x_2)}{\partial x_2^2} = 2 \qquad \frac{\partial^2 f(x_1, x_2)}{\partial x_1 \partial x_2} = 0$$

$$H_f = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

For arbitrary $\boldsymbol{x} \begin{pmatrix} x_1 & x_2 \end{pmatrix}$, we can check the Hessian's definiteness by taking

$$\boldsymbol{x}^T H_f \boldsymbol{x} = 2x_1^2 + 2x_2^2 > 0$$

which implies $f(x_1, x_2)$ is convex. The critical point is a minimum.

[1] Carter 5.10

3. Characterize local optima and solve[2]

$$\max_{x_1,x_2} f(x_1,x_2) = 3x_1x_2 - x_1^3 - x_2^3$$

We first find the stationary points, taking

$$\frac{\partial f(x_1,x_2)}{\partial x_1} = 3x_2 - 3x_1^2 = 0$$

and

$$\frac{\partial f(x_1,x_2)}{\partial x_2} = 3x_1 - 3x_2^2 = 0$$

which obtain for $x_1 = x_2 = 0$ and $x_1 = x_2 = 1$. We now need to check if the function is concave at either of these points. This will identify whether or not the point is a local maximum. To do so, we build the Hessian.

$$\frac{\partial^2 f(x_1,x_2)}{\partial x_1^2} = -6x_1 \qquad \frac{\partial^2 f(x_1,x_2)}{\partial x_2^2} = -6x_2 \qquad \frac{\partial^2 f(x_1,x_2)}{\partial x_1 \partial x_2} = 3$$

$$H_f(x_1,x_2) = \begin{pmatrix} -6x_1 & 3 \\ 3 & -6x_2 \end{pmatrix}$$

To check the definiteness of $f(x_1,x_2)$ at $(0,0)$, take $\boldsymbol{x} \begin{pmatrix} x_1 & x_2 \end{pmatrix}$ and

$$\boldsymbol{x}^T H_f(0,0)\boldsymbol{x}$$

or

$$\begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 0 & 3 \\ 3 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 6x_1x_2$$

so the matrix is *indefinite*, depending on the values of $\boldsymbol{x}$. $(0,0)$ is a saddle point.

Now check definiteness at $(1,1)$.

$$\begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} -6 & 3 \\ 3 & -6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = -6x_1^2 + 6x_1x_2 - 6x_2^2$$

We can simplify this expression as follows

$$\begin{aligned} -6x_1^2 + 6x_1x_2 - 6x_2^2 &= -6(x_1^2 - x_1x_2 + x_2^2) \\ &= -6(x_1^2 + x_2^2 - x_1x_2) \end{aligned}$$

The Hessian will be negative semidefinite as long as

$$x_1^2 + x_2^2 - x_1x_2 \geq 0$$

[2]Carter Example 5.8

or equivalently

$$\frac{x_1 x_2}{x_1^2 + x_2^2} \leq 1 \tag{1}$$

Since $x_1 x_2 \leq \max\left\{x_1^2, x_2^2\right\}$,

$$x_1 x_2 \leq x_1^2 + x_2^2$$

and Inequality 1 holds. So $(1,1)$ is a local maximum. Note, however, that as $x_1, x_2 \to -\infty f(x_1, x_2) \to \infty$. So there is no global maximum of the function $f(x_1, x_2)$ and no solution to the above problem.

4. Prove that the least squares objective function is convex, implying that the first order conditions are sufficient to characterize the $\boldsymbol{\beta}$ that solves the least squares estimator.

We're given an $m \times n$ matrix X with

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix}$$

and a $n \times 1$ vector $\boldsymbol{y}$.

Our problem is as follows:

$$\min_{\boldsymbol{\beta}} f_{\text{SSE}}(\boldsymbol{\beta}) = \sum_i \epsilon_i^2 = \sum_{i=1}^n \left(y_i - \sum_j x_{ij}\beta_j \right)^2$$

where $\boldsymbol{\beta}$ is an $m \times 1$ vector. Taking first partials gives

$$\begin{aligned} \frac{\partial f_{\text{SSE}}(\boldsymbol{\beta})}{\partial \beta_k} &= 2\sum_{i=1}^n \left(y_i - \sum_j x_{ij}\beta_j \right)(-x_{ik}) \\ &= 2\sum_{i=1}^n \sum_j x_{ij}\beta_j x_{ik} - 2\sum_{i=1}^n y_i x_{ik} \end{aligned}$$

Taking second order partials with respect to some β_h,

$$\frac{\partial^2 f_{\text{SSE}}(\boldsymbol{\beta})}{\partial \beta_k \partial \beta_h} = 2\sum_{i=1}^n x_{ih} x_{ik}$$

We can combine store these partials in the Hessian

$$H_f = 2\begin{pmatrix} \sum_{i=1}^n x_{i1}x_{i1} & \cdots & \sum_{i=1}^n x_{i1}x_{im} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_{im}x_{i1} & \cdots & \sum_{i=1}^n x_{im}x_{im} \end{pmatrix}$$

The columns of X are $n \times 1$ vectors $\boldsymbol{x}_k$, so this matrix can be equivalently expressed as

$$H_f = 2 \begin{pmatrix} \boldsymbol{x}_1^T \boldsymbol{x}_1 & \cdots & \sum_{i=1}^n \boldsymbol{x}_1^T \boldsymbol{x}_m \\ \vdots & \ddots & \vdots \\ \boldsymbol{x}_m^T \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_m^T \boldsymbol{x}_m \end{pmatrix}$$

or alternatively as

$$H_f = 2X^T X$$

We seek to show this matrix is positive definite, or

$$\boldsymbol{x}^T H_f \boldsymbol{x} > 0$$

for all $\boldsymbol{x} \neq 0$. We have

$$\begin{aligned} \boldsymbol{x}^T 2X^T X\boldsymbol{x} &> 0 \\ 2\boldsymbol{x}^T X^T X\boldsymbol{x} &> 0 \\ 2(X\boldsymbol{x})^T X\boldsymbol{x} &> 0 \\ (X\boldsymbol{x})^T X\boldsymbol{x} &> 0 \end{aligned}$$

which holds by the positive definiteness of the inner product. ■

5. Let $g : \mathbb{R} \to \mathbb{R}$ be a monotonic transformation of $f : X \to \mathbb{R}$. Show that $g \circ f$ has the same local maxima as f.

If $\boldsymbol{x}^\star$ is a local maximum of f, then $f(\boldsymbol{x}^\star) \geq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in S(\boldsymbol{x}^\star)$ where S is a neighborhood around $\boldsymbol{x}^\star$. Since g is a monotonic function

$$f(\boldsymbol{x}^\star) \geq f(\boldsymbol{x}) \implies g(f(\boldsymbol{x}^\star)) \geq g(f(\boldsymbol{x}))$$

which implies $\boldsymbol{x}^\star$ is a local maximum of $g \circ f$ as desired. ■

6. Consider the problem

$$\begin{aligned} \max_{x_1, x_2} \quad & x_1 x_2 \\ \text{subject to} \quad & x_1 + x_2 = 1 \end{aligned} \tag{2}$$

   Think about the geometry of the problem. What is the constraint set? Then solve it using the method of Legrange.[3]

Note that the objective can be rewritten

$$\max_{x_1} x_1(1 - x_1)$$

[3] Carter Example 5.14

which makes it easy to verify the function is concave along the constraint set

$$\frac{\partial^2 f}{\partial x_1^2} = -2 < 0$$

This also makes it easy to solve the problem using techniques from univariate calculus. To be thorough, we'll write the Legrangian

$$\mathcal{L} = x_1 x_2 - \lambda(1 - x_1 - x_2)$$

$$\frac{\partial \mathcal{L}}{\partial x_1} = x_2 + \lambda$$

$$\frac{\partial \mathcal{L}}{\partial x_2} = x_1 + \lambda$$

Yielding $x_1 = x_2 = -\lambda$. This yields a single possibility for x_1 and x_2, we must have $x_1 = x_2 = 1/2$.