CHAPTER 1

INTRODUCTION TO OPTIMIZATION

General reading on your own

Homework

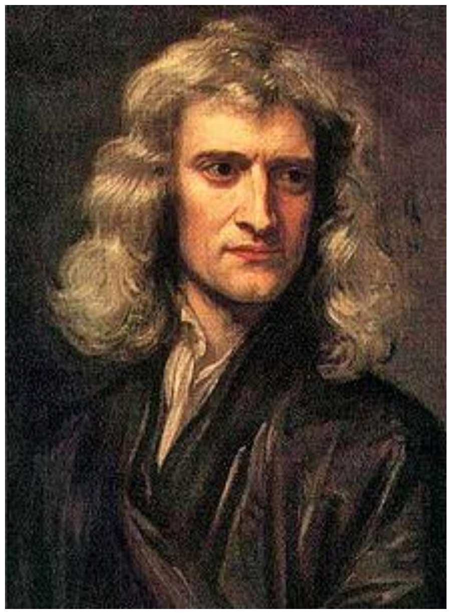
1.1, 1.8, 1.19

CHAPTER 2

CLASSICAL OPTIMIZATION TECHNIQUES

This chapter is a revision of what you already learned in your math undergraduate curriculum. We are going through it to ensure that you have a systematic understanding of the mathematical basis of the optimization theory.

Most of this work was started by Isaac Newton and was further developed through the 18th Century.



Isaac Newton (1642-1726), Wikipedia

Single-Variable Optimization with no Constraints

A function *f(x)* has a local minimum at x* if $f(x^*) < f(x^* + h)$ where *h* is a small negative or positive disturbance around *x**.

Similarly, a function *f(x)* has a local maximum at *x** if $f(x^*) > f(x^* + h)$ where *h* is a small negative or positive disturbance around x*.

Necessary Condition: Given a piecewise smooth function *f(x)* that is defined in the interval, $a \leq x \leq b$. If the function has an extremum *x** within this period, then,

$$\frac{df(x^*)}{dx} = f'(x^*) = 0$$

Proof: This can be shown using the function quotient

Note: The theory is not valid to for some cases such as $f(x)=x^3$

Sufficient Condition: Given a piecewise smooth function *f(x)* that is defined in the interval $a \leq x \leq b$.

If $f'(x^*) = f''(x^*) = \cdots = f^{n-1}(x^*) = 0$ and $f^n(x^*) \neq 0$, then $f(x^*)$ is,

1. *Minimum* if $f^n(x^*) > 0$ and n is even
2. *Maximum* if $f^n(x^*) < 0$ and n is even
3. Neither minimum or maximum if *n* is odd (*inflection/saddle point*)

Proof: This can be easily shown using Taylor's theorem:

Brook Taylor (1685-1731), Wikipedia

$$f(x + \Delta x) = f(x) + \frac{df(x)}{dx}\Delta x + \frac{1}{2!}\frac{d^2 f(x)}{dx^2}(\Delta x)^2 + \cdots + \frac{1}{(n-1)!}\frac{d^{n-1} f(x)}{dx^{n-1}}(\Delta x)^{n-1}$$

Note: The theory cannot make distinction between local and global extrema.

Example 2.1: Analyze the extrema of $f(x) = 5x^6 - 36x^5 + \frac{165}{2}x^4 - 60x^3 + 36$ in the interval in the interval $-1 \leq x \leq 4$

$$f'(x) = 30x^5 - 180x^4 + 330x^3 - 180x^2 = 30x^2(x^3 - 6x^2 + 11x - 6)$$
$$= 30x^2(x-1)(x-2)(x-3)$$

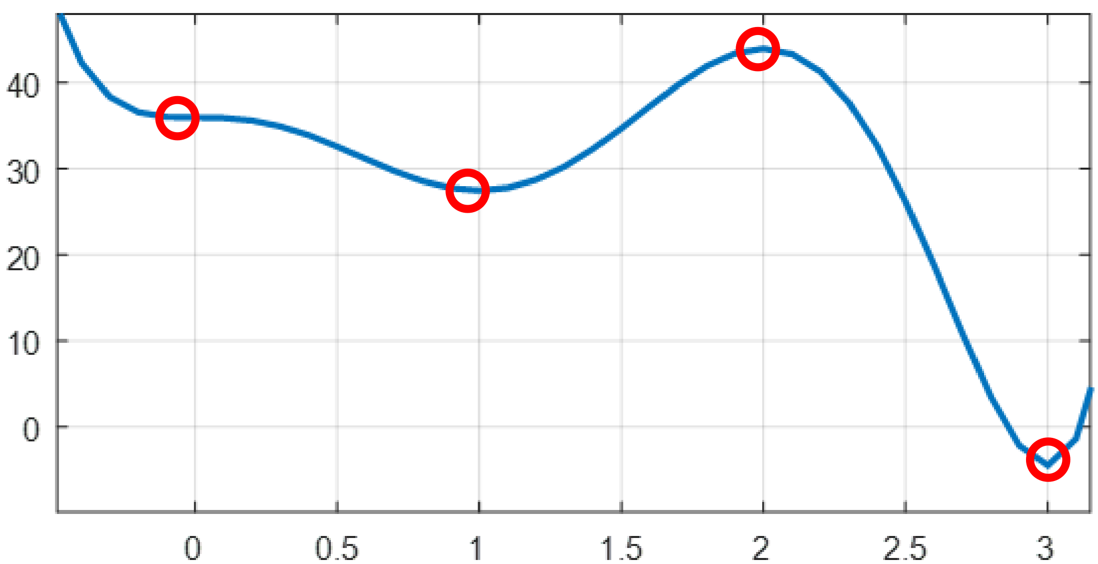
$$f''(x) = 150x^4 - 720x^3 + 990x^2 - 360x$$

$$f'''(x) = 600x^3 - 2160x^2 + 1980x - 360$$

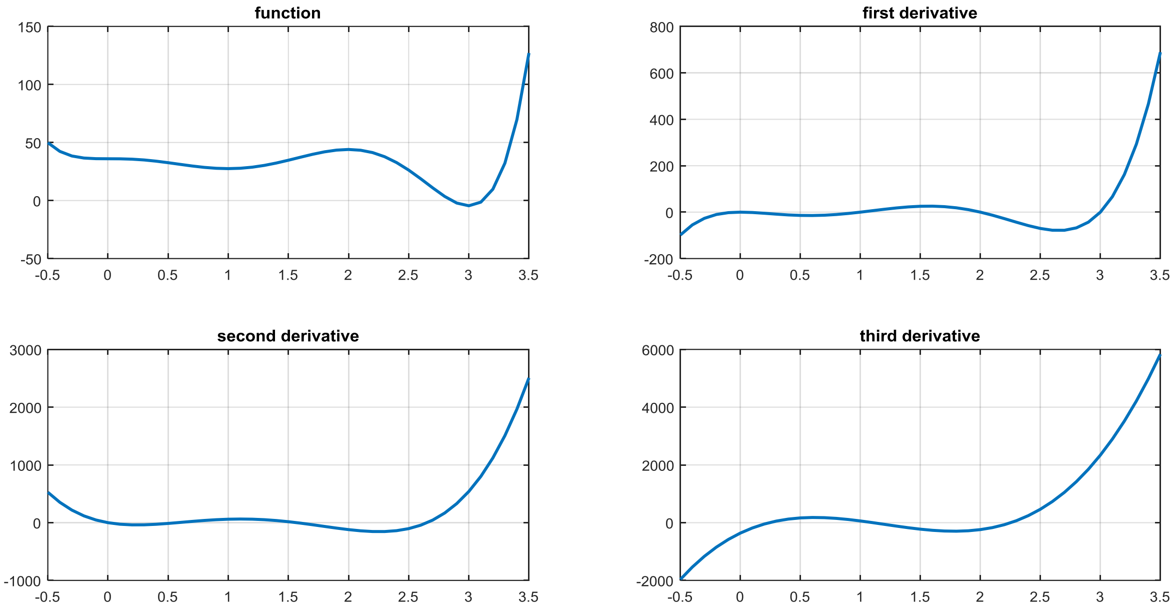
$f'(x)$ shows that there are possibilities for extremum at 0, 1, 2, and 3. The following can help determine their nature.

x	$f(x)$	$f'(x)$	$f''(x)$	$f'''(x)$	Observation
0	36	0	0	-360	First nonzero derivative is odd, *inflection*
1	27.5	0	60		First nonzero derivative is even, the value is positive: *local minimum*
2	44	0	-120		First nonzero derivative is even, the value is negative: *local maximum*
3	5.5	0	540		First nonzero derivative is even, the value is positive: *local minimum*

Zoomed view



Function and derivative comparison



Homework: 2.2, 2.4, 2.5, 2.11

Multivariable Optimization with no Constraints

Necessary Condition: Given a piecewise smooth function *f(x)* that is defined in the interval $a \leq x \leq b$. If the function has an extremum *x** within this period, then,

$$\frac{\partial f(x^*)}{\partial x_1} = \frac{\partial f(x^*)}{\partial x_1} = \cdots = \frac{\partial f(x^*)}{\partial x_n} = 0$$

In we think in terms of x as one-dimensional array of variables,

$$\frac{\partial f(x^*)}{\partial x} = \nabla f(x^*) = 0$$

Proof: This can be shown using Taylor's theorem,

$$f(x^* + h) = f(x^*) + \sum_{i=1}^{n} h_i \frac{\partial f(x^*)}{\partial x_i} + R_1(x^*, h)$$

Sufficient Condition: Given a piecewise smooth function *f(x)* that is defined in the interval $a \leq x \leq b$, a sufficient condition for a point x* to be an extremum is that the matrix of second partial derivatives (Hessian matrix, *J*) of f(x) when evaluated at x* is,

1. *Minimum* if $J(x^*)$ is positive definite
2. *Maximum* if $J(x^*)$ is negative definite
3. *Inflection/saddle point* if $J(x^*)$ is indefinite*

Proof: This can be shown using Taylor's theorem,

$$f(x^* + h) = f(x^*) + \sum_{i=1}^{n} h_i \frac{\partial f(x^*)}{\partial x_i} + \frac{1}{2!} \sum_{i=1}^{n} \sum_{j=1}^{n} h_i h_j \frac{\partial^2 f(x^*)}{\partial x_i \partial x_j}$$

Note: The theory cannot make distinction between local and global extrema.

It may be more useful to express the above equation in a matrix form.

$$f(x^* + h) = f(x^*) + [h]^T \left\{ \frac{\partial f(x^*)}{\partial x_i} \right\} + \frac{1}{2!} [h]^T [J(x^*)] \{h\}$$

where *J* is the *Hessian matrix* of the function.

$$J(x^*) = \begin{bmatrix} \frac{\partial^2 f(x^*)}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f(x^*)}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x^*)}{\partial n \partial x_1} & \cdots & \frac{\partial^2 f(x^*)}{\partial x_n \partial x_n} \end{bmatrix}$$

Note: The Hessian matrix is always *square* and *symmetric* (*why?*)

Otto Hesse (1811-1874), Wikipedia

Definition: A matrix is positive definite if all eigenvalues are positive. This means that

$$|A - \lambda\, I| = 0$$

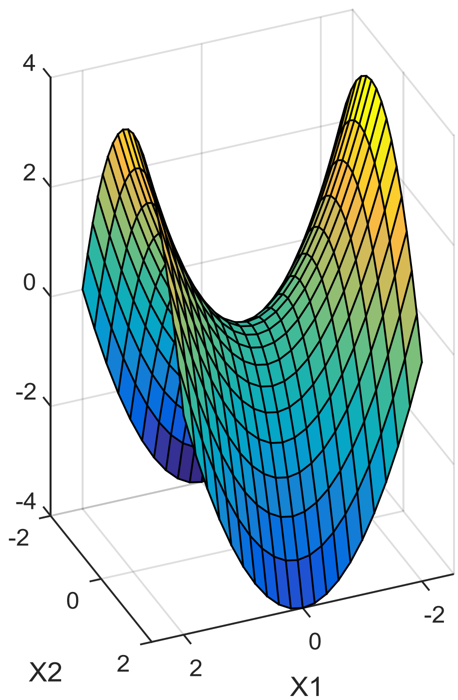
A matrix is negative definite if all eigenvalues are negative.

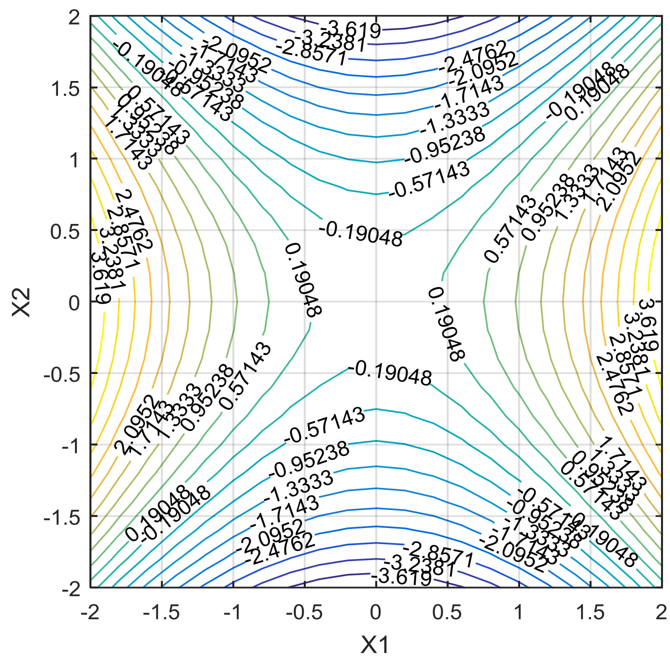
Note: Eigenvalues can be calculated using *eig* function in MATLAB.

---

*A saddle point is maximum in one variable and minimum in another. An example

$$f(x) = {x_1}^2 - {x_2}^2$$





Example 2.2: Analyze the extrema of $f(x_1, x_2) = 2 - {x_1}^2 - x_1 x_2 - {x_2}^2$ in the interval in the interval $-2 \leq x_1 \leq 2$ and $-2 \leq x_2 \leq 2$

This is a *quadratic* function.

$$\frac{\partial f}{\partial x} = \nabla f = \begin{Bmatrix} -2x_1 - x_2 \\ -x_1 - 2x_2 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}$$

The first equation is zero at $x_1 = -x_2/2$

The second equation is zero at $x_1 = -2x_2$

Therefore, extremum is *only* possible at *(0, 0)*

$$\frac{\partial^2 f}{\partial x^2} = J = \begin{bmatrix} -2 & -1 \\ -1 & -2 \end{bmatrix}$$

This matrix is independent of x_1 and x_2 (why?)

Test the eigenvalues of *J:*

$$|J - \lambda\, I| = 0$$

$$\left|\begin{bmatrix} -2 & -1 \\ -1 & -2 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right| = 0$$

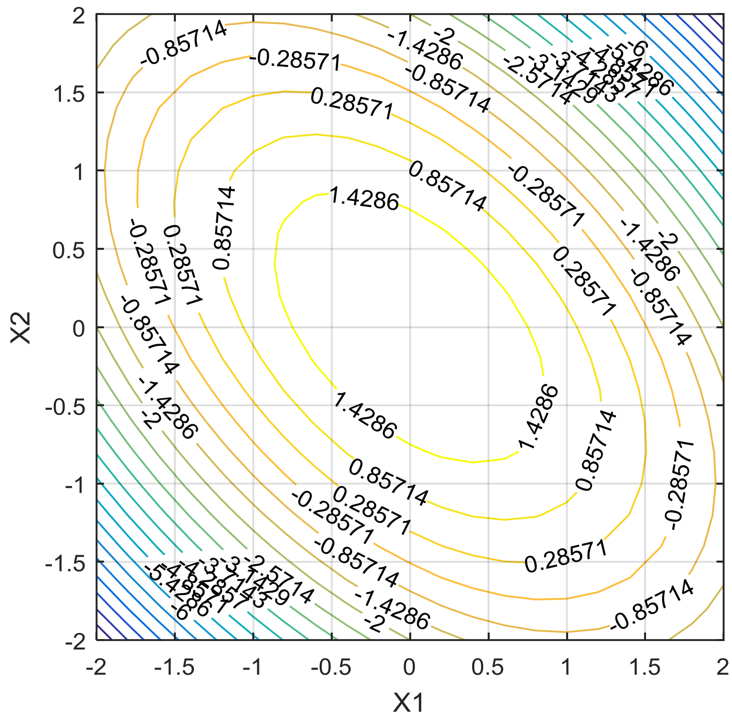
$$\left|\begin{bmatrix} -2 - \lambda & -1 \\ -1 & -2 - \lambda \end{bmatrix}\right| = 0$$

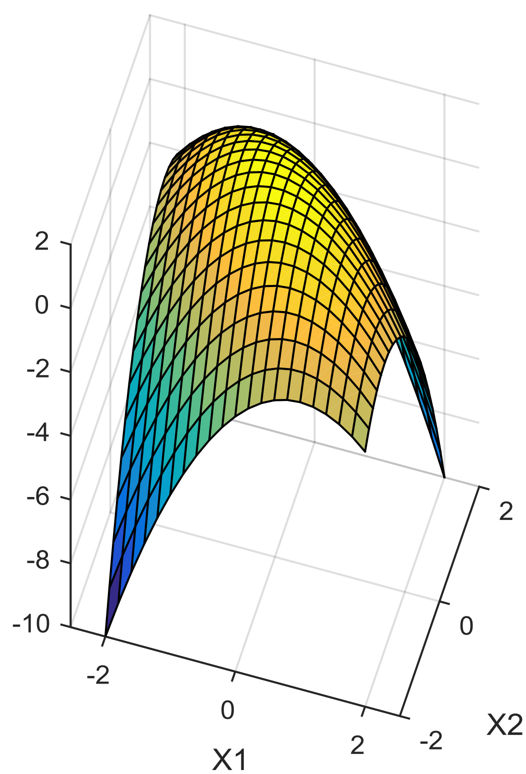
$$(-2 - \lambda)^2 - 1 = 0$$

$$\lambda^2 + 4\lambda + 3 = 0$$

The eigenvalues are *(-3, -1)*, which shows that the matrix is negative definite.

This means that *(0, 0)* is maximum.


X1
X2
-0.85714
-0.28571
0.28571
0.85714
1.4286
-1.4286
-2
-2.5714


X1
X2

Example 2.3: Analyze the extrema of $f(x_1, x_2) = (x_1 - 1)^2(x_2 + 1) - x_2$ in the interval in the interval $-1 \leq x_1 \leq 3$ and $-2 \leq x_2 \leq 2$

$$\frac{\partial f}{\partial x} = \nabla f = \begin{Bmatrix} 2(x_1 - 1)(x_2 + 1) \\ (x_1 - 1)^2 - 1 \end{Bmatrix} = \begin{Bmatrix} 2(x_1 - 1)(x_2 + 1) \\ {x_1}^2 - 2x_1 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}$$

The first equation is zero at $x_1 = 1 \; or \; x_2 = -1$

The second equation is zero at $x_1 = 0 \; or \; 2$

Therefore, $x_1 = 1$ cannot be accepted as using it wll result in a nonzero value of the second equation of ∇f.

Therefore, the extremum are possible at (0, -1) and (2, -1).

To assess the nature of these two points, we derive the Hessian matrix.

$$\frac{\partial^2 f}{\partial x^2} = J = \begin{bmatrix} 2(x_2 + 1) & 2(x_1 - 1) \\ 2(x_1 - 1) & 0 \end{bmatrix}$$

Substituting in (0, -1)

$$J = \begin{bmatrix} 0 & -2 \\ -2 & 0 \end{bmatrix}$$

The eigenvalues are (-2, 2), which shows that the matrix is *indefinite*.

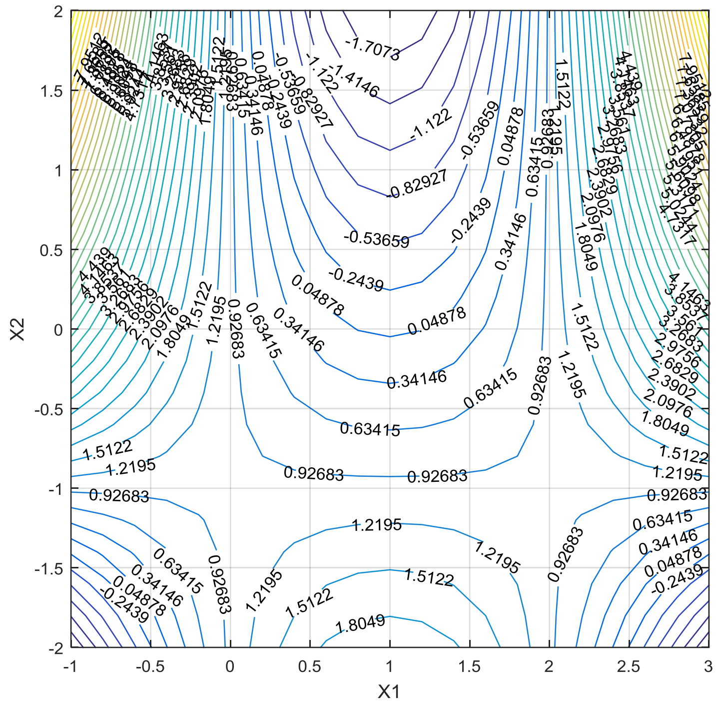
This means that *(0, -1) is a saddle point.*

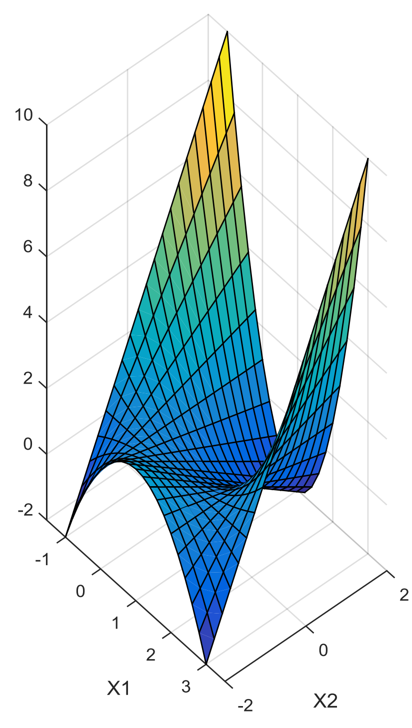
Substituting in (2, -1)

$$J = \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix}$$

The eigenvalues are (-2, 2), which shows that the matrix is *indefinite*.

This means that *(2, -1) is a saddle point*.





Homework: 2.12, 2.18, 2.21, 2.26, 2.32