J. TOMAL

Shad2024 Lecture Notes:

THE NEWTON-RAPHSON METHOD

July 16, 2024

Department of Mathematics and Statistics, Thompson Rivers University

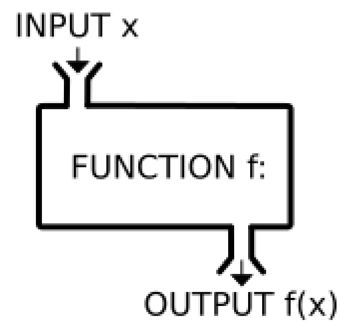
1
Essential Concepts

1.1 Functions

Definition of a function

A functioon f from a set X to a set Y is an assignment of one element of Y to each element of X, where

- The set X is called the ***domain*** of the function.
- The set Y is called the ***codomain*** of the function.
- Notation: $f : X \to Y$ or $X \xrightarrow{f} Y$ or $x \mapsto f(x)$



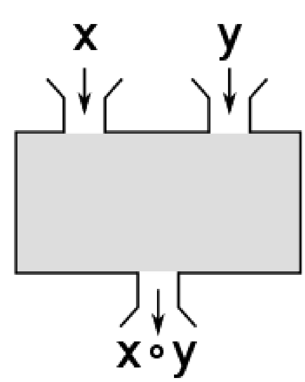
Example 1.1: (Probability density functions and likelihood functions)

Identify the independent variables of the following functions:

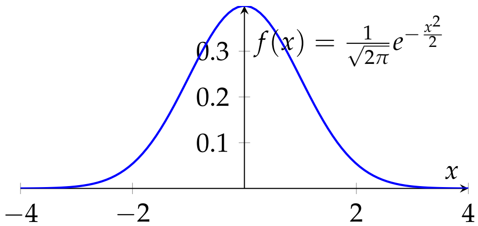
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$P(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i - \alpha - \beta x_i - \gamma x_i^2)^2}$$

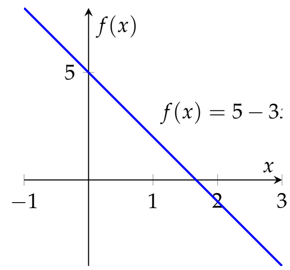
$$L(\alpha, \beta, \gamma, \sigma^2) = \prod_{i=1}^{n} P(y_i) \quad = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i - \gamma x_i^2)^2}$$



A multivariate function $f : U \to f(x_1, \ldots, x_n)$, where $x_1 \in X_1, \ldots, x_n \in X_n$, the domain U has the form $U \subseteq X_1 \times \ldots \times X_n$, $X_1 \times \ldots \times X_n$ is the Cartesian product of $X_1, \ldots, X_n$

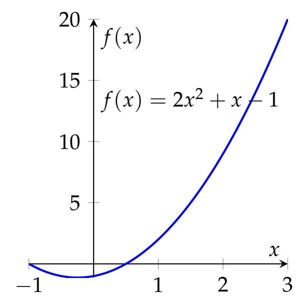


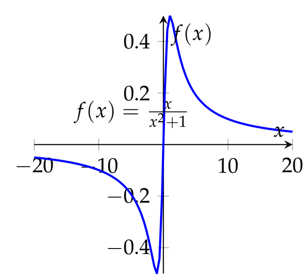
1.2 *Common types of functions*



Common types of functions

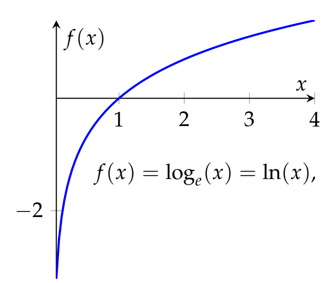
- ***Linear Functions:*** Represented by $f(x) = ax + b$, where a and b are constants. The graph is a straight line.
- ***Quadratic Functions:*** Represented by $f(x) = ax^2 + bx + c$, where a, b, and c are constants. The graph is a parabola.
- ***Polynomial Functions:*** Represented by $f(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$, where $a_n, a_{n-1}, \ldots, a_1, a_0$ are constants and n is a non-negative integer. The graph varies depending on the degree n.
- ***Rational Functions:*** Represented by $f(x) = \frac{P(x)}{Q(x)}$, where $P(x)$ and $Q(x)$ are polynomial functions and $Q(x) \neq 0$. The graph can have vertical and horizontal asymptotes.
- ***Exponential Functions:*** Represented by $f(x) = a \cdot b^x$, where a and b are constants and $b > 0$. The graph increases (if $b > 1$) or decreases (if $0 < b < 1$)
- ***Logarithmic Functions:*** Represented by $f(x) = a \cdot \log_b(x) + c, \quad x > 0$, where a, b, and c are constants and $b > 0$. The graph is the inversse of an exponential function. When $a = 1, b = e = 2.718281828459, c = 0$, f is called the ***natural logrithm*** of x, denoted as $f(x) = \ln(x)$, the graph is strictly increasing on its domain.
- ...





1.3 *Limits*

A limit describes the value that a function (or sequence) approaches as the input (or index) approaches some value. Limits are fundamental to calculus and mathematical analysis and are used to define concepts like continuity, derivatives, and integrals.



* $e = \lim_{n\to\infty}(1 + \frac{1}{n})^n$ or $e = \sum_{n=0}^{\infty} \frac{1}{n!}$

Definition

The limit of $f(x)$ as x approaches c is L, denoted as $\lim_{x \to c} f(x) = L$, if for every number $\epsilon > 0$, there exists a number $\delta > 0$ such that $0 < |x - c| < \delta$ implies $|f(x) - L| < \epsilon$.

Intuitive Understanding: As x gets closer and closer to c from both sides, $f(x)$ gets closer and closer to L. The value of the limit is what $f(x)$ approaches, not necessarily what $f(x)$ equals when $x = c$.

1.4 *Derivatives*

A derivative represents the rate of change of a function with respect to a variable. It is the slope of the function's graph at any given point and indicates how the function's value changes as the input changes.

Definition

The derivative of a function $f(x)$ at a point x is given by:

$$f'(x) = lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

This limit, if it exists, defines the instantaneous rate of change of the function at x.

Derivatives can be denoted in several ways, depending on the context:

- ***Leibniz Notation:*** $\frac{dy}{dx}$ or $\frac{df}{dx}$ if $y = f(x)$. This notation emphasizes the derivative as a ratio of differentials.
- ***Lagrange Notation:*** $f'(x)$ This notation is compact and often used when the function is explicitly named.
- ***Newton Notation:*** $\dot{y}$ for the first derivative and $\ddot{y}$ for the second derivative, primarily used in physics.

Basic Differentiation Rules:

- Power Rule:
$$\frac{d}{dx}[x^n] = nx^{n-1}$$
Example: If $f(x) = x^3$, then
$$f'(x) = 3x^2$$
- Product Rule:
$$\frac{d}{dx}[u(x)v(x)] = u'(x)v(x) + u(x)v'(x)$$
Example: If $f(x) = x^2 \sin(x)$, then
$$f'(x) = 2x \sin(x) + x^2 \cos(x)$$
- Quotient Rule:
$$\frac{d}{dx}\left[\frac{u(x)}{v(x)}\right] = \frac{u'(x)v(x) - u(x)v'(x)}{[v(x)]^2}$$
Example: If $f(x) = \frac{x}{\sin(x)}$, then
$$f'(x) = \frac{\sin(x) - x\cos(x)}{\sin^2(x)}$$
- Chain Rule:
$$\frac{d}{dx}[g(h(x))] = g'(h(x))h'(x)$$
Example: If $f(x) = \sin(x^2)$, then
$$f'(x) = \cos(x^2) \cdot 2x$$

Common Derivatives:

$$[c]' = 0$$
$$[e^x]' = e^x$$
$$[a^x]' = a^x \ln(a)$$
$$[\ln(x)]' = \frac{1}{x}$$
$$[\log_a(x)]' = \frac{1}{x\ln(a)}$$
$$[\sin(x)]' = \cos(x)$$
$$[\cos(x)]' = -\sin(x)$$
$$[\tan(x)]' = \sec^2(x)$$
$$[\sin^{-1}(x)]' = \frac{1}{\sqrt{1-x^2}}$$
$$[\cos^{-1}(x)]' = -\frac{1}{\sqrt{1-x^2}}$$
$$[\tan^{-1}(x)]' = \frac{1}{1+x^2}$$

2

The Newton-Raphson Method

2.1 Introduction: Root-Finding Methods

Objective: To Find the location of ***roots*** of a function.

* **A Root** of a function is a value x at which $f(x) = 0$, or the x values at which the graph of f intersects the x-axis (x-intercepts).

Example 2.1: Find the roots of the following functions.

a) $f(x) = 5 - 3x$

b) $f(x) = 2x^2 + x - 1$

Now what about roots of $f(x) = x^3 - 5x + 1$?

* The discriminant of the general cubic equation $ax^3 + bx^2 + cx + d = 0, a \neq 0$ exists and can be written as

$$\frac{4(b^2 - 3ac)^3 - (2b^3 - 9abc + 27a^2d)^2}{27a^2}.$$

But the solution for x can be complecated. Hence, we introduce the ***Newton-Raphson Method***.

Newton-Raphson Method

Newton-Raphson Method is a numerical analysis technique for finding a root of a function.
It involves finding successive tangent lines to the graph of f, following a certain algorithm until we get close enough to the root.

2.2 Derive the iteration formula

A function f is given and assume r is a root of f that we wish to approximate, we also assume f is differentiable in an interval containing r.

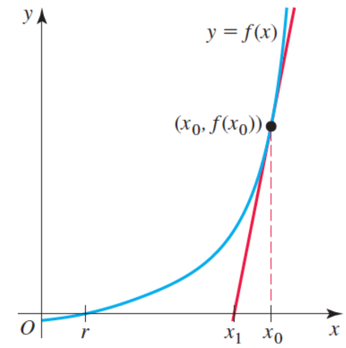
Suppose an initial approximation of r is also given, call that x_0.

$\rightsquigarrow$ **Algorithm**:

1) Draw a tangent line to the curve of f at the point $(x_0, f(x_0))$:

(What is the equation of this tangent line at $(x_0, f(x_0))$?)

2) Continue the tangent line such that it intersects the x-axis at a point, call that point x_1. x_1 is our new approximation:

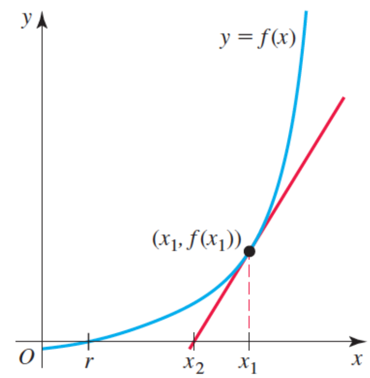


(a)

$$x_1 =$$

3) Now repeat steps 1 and 2 for x_1:

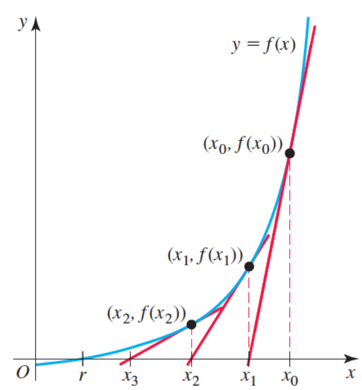
Draw a tangent line to the curve of f at the point $(x_1, f(x_1))$, continue the tangent line such that it intersects the x-axis at x_2 and repeat the steps.



(b)

$$x_2 =$$

Repeating the algorithm for each point, we will obtain a sequence of points, $\{x_0, x_1, x_2, x_3, ...\}$, that ideally get closer and closer to the root r.



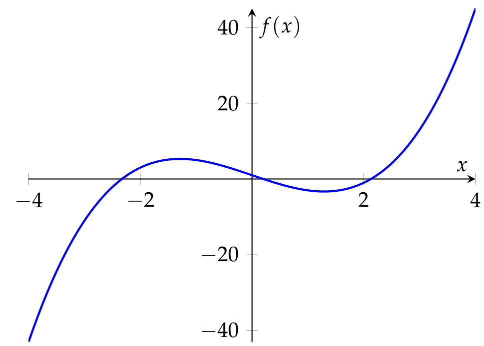
$$x_3 =$$

This pattern continues: If a previous approximation is known, say x_n, then the new approximation is calculated by the following formula:

$$x_{n+1} = \tag{2.1}$$

Example 2.2: Find the root of cubic equation

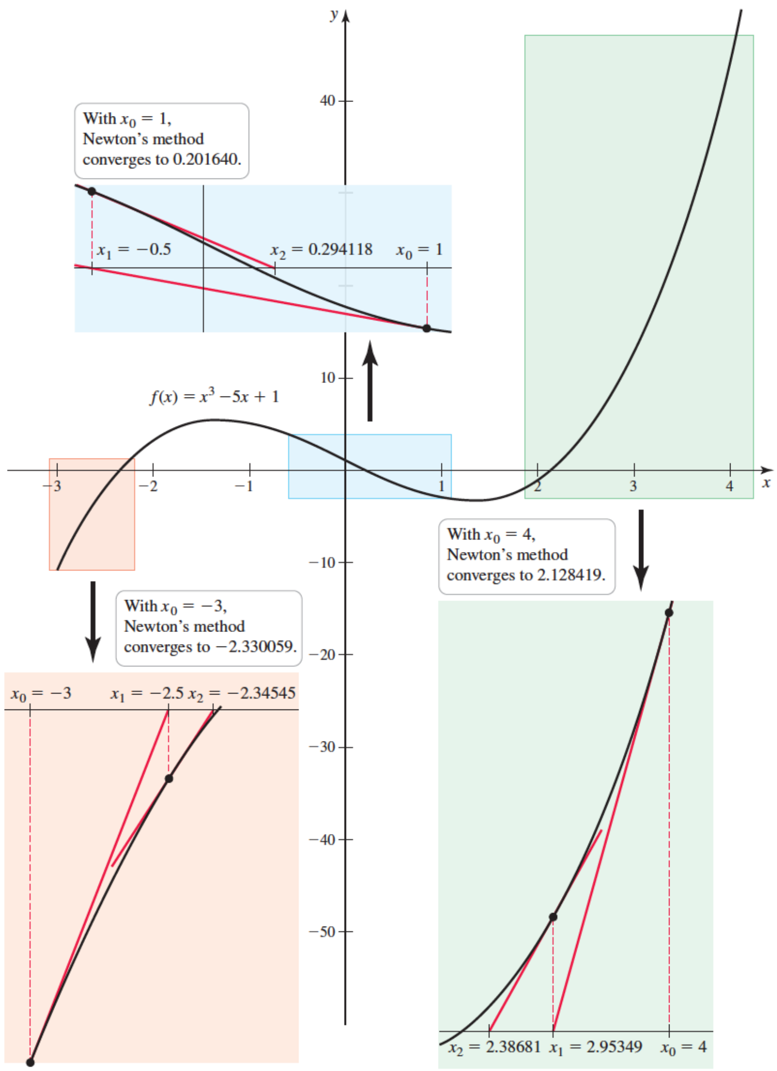
Find the root of $f(x) = x^3 - 5x + 1 = 0$ with inital approximations $x_0 = -3, x_0 = 1$ and $x_0 = 4$ with two steps.



$$f(x) = x^3 - 5x + 1$$

$$f'(x) =$$

$$x_{n+1} =$$



Remark 1. Newton's method is an example of a repetitive loop calculation called an *iteration*. It is mainly done by calculators and computers and it is included in many scientific computing software.

Remark 2. When to stop?

They are different way to decide when to terminate the iterations. Either the number of iterations are given, or the number of agreeing digits between two successive approximations are given, for instance continue the iterations until two successive approximations agree to 4 digits. The effectiveness of the algorithm is to get close enough to the root (small error) as quickly as possible (not many iterations).

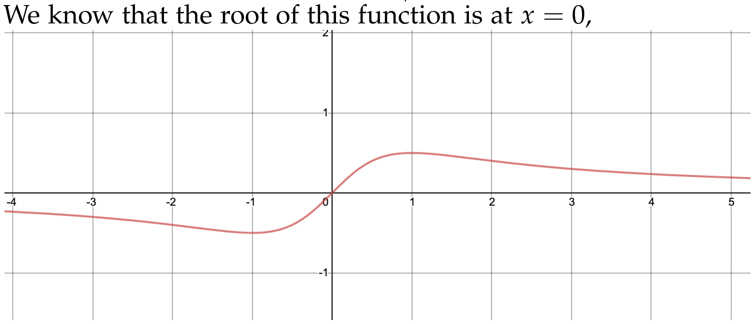
Example 2.3:

Use Newton's method to estimate the value for $\sqrt{10}$. Stop calculating approximations when two successive approximations agree to four digits to the right of the decimal point.

Remark 3. Newton's method is not always working. The location of the initial approximation is important.

Example 2.4:

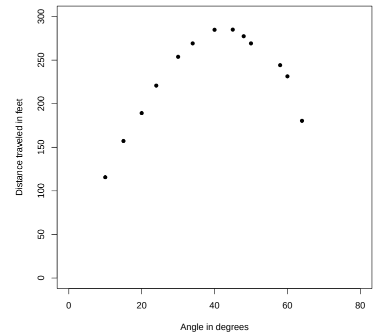
Consider the function $f(x) = \dfrac{x}{x^2+1}$:
We know that the root of this function is at $x = 0$,



Now let's apply Newton's method with two initial approximations $x_0 = 2$ and $x_0 = \frac{1}{\sqrt{3}}$:

n	x
1	5.33333333333333333333
2	11.05533063427800269906
3	22.29306186322140567443
4	44.67601861966209455121
5	89.39682642589238930955
6	178.81602781036031290 58
7	357.64324065024497904 32
8	715.29207350917434544 6
9	1,430.58694308443561571 7
10	2,861.17528419713292443 1

2.3 Quadratic Model



Example 2.5:

Problem Statement:
We aim to model the distance a baseball travels as a function of the hitting angle using a quadratic function.
Model Specification:
Let

- Y_i denote the random variable of the distance traveled by the baseball (in feet) for the i-th observation. The actual observed data of Y_i is denoted by y_i.
- X_i denote the random variable of the hitting angle (in degrees) for the i-th observation. The actual observed data of X_i is denoted by x_i

The relationship between the distance and the hitting angle is modeled as: $Y_i = \alpha + \beta X_i + \gamma X_i^2 + \epsilon_i$
where:

- α, β, γ are the parameters to be estimated.
- ϵ_i represents the error term, assumed to be normally distributed with mean 0 and variance σ^2, i.e. $\epsilon_i \sim N(0, \sigma^2)$.

Objective:
Given a dataset $\{(y_i, x_i)\}_{i=1}^n$, our goal is to estimate the parameters α, β, γ, and σ^2.

Because of $Y_i \sim N(\alpha + \beta X_i + \gamma X_i^2, \sigma^2)$,
The probability density function of $Y_i = y_i$ is:

$$P(Y_i = y_i | x_i, \alpha, \beta, \gamma, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y_i - \alpha - \beta x_i - \gamma x_i^2)^2}$$

The likelihood function of $(\alpha, \beta, \gamma, \sigma^2)$ is:

$$\begin{aligned} L(\alpha, \beta, \gamma, \sigma^2 | \{(y_i, x_i)\}_{i=1}^n) &= \prod_{i=1}^n P(Y_i = y_i | x_i, \alpha, \beta, \gamma, \sigma^2) \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \alpha - \beta x_i - \gamma x_i^2)^2} \\ &= (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \alpha - \beta x_i - \gamma x_i^2)^2} \end{aligned}$$

Taking the log on both sides of the equation yields the following log likelihood function:

$$l(\alpha,\beta,\gamma,\sigma^2) = \ln L(\alpha,\beta,\gamma,\sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i - \gamma x_i^2) \tag{2.2}$$

We want to find the value of $(\alpha,\beta,\gamma,\sigma^2)$ such that equation 2.2 gets its maximum.

Consider the gradient of $l(\alpha,\beta,\gamma,\sigma^2)$:

$$\nabla l(\mathbf{\Theta}) = \begin{pmatrix} \frac{\partial l}{\partial \alpha} \\ \frac{\partial l}{\partial \beta} \\ \frac{\partial l}{\partial \gamma} \\ \frac{\partial l}{\partial \sigma^2} \end{pmatrix}, \qquad \text{where } \mathbf{\Theta} = \begin{pmatrix} \alpha \\ \beta \\ \gamma \\ \sigma^2 \end{pmatrix}$$

$$\begin{aligned}
\frac{\partial l}{\partial \alpha} &= -\frac{1}{2\sigma^2}\sum_{i=1}^{n} 2(y_i - \alpha - \beta x_i - \gamma x_i^2)(-1) \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i - \gamma x_i^2) \\
\frac{\partial l}{\partial \beta} &= -\frac{1}{2\sigma^2}\sum_{i=1}^{n} 2(y_i - \alpha - \beta x_i - \gamma x_i^2)(-x_i) \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i)(y_i - \alpha - \beta x_i - \gamma x_i^2) \\
&= \frac{1}{\sigma^2}\left(\sum_{i=1}^{n} x_i y_i - \alpha\sum_{i=1}^{n} x_i - \beta\sum_{i=1}^{n} x_i^2 - \gamma\sum_{i=1}^{n} x_i^3\right) \\
\frac{\partial l}{\partial \gamma} &= -\frac{1}{2\sigma^2}\sum_{i=1}^{n} 2(y_i - \alpha - \beta - \gamma x_i^2)(-x_i^2) \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i^2)(y_i - \alpha - \beta - \gamma x_i^2) \\
&= \frac{1}{\sigma^2}\left(\sum_{i=1}^{n} x_i^2 y_i - \alpha\sum_{i=1}^{n} x_i^2 - \beta\sum_{i=1}^{n} x_i^3 - \gamma\sum_{i=1}^{n} x_i^4\right) \\
\frac{\partial l}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} - \frac{(-1)(\sigma^2)^{-2}}{2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i - \gamma x_i^2)^2 \\
&= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i - \gamma x_i^2)^2
\end{aligned}$$

The Hessian matrix of the likelihood function is:

$$\nabla^2 l(\mathbf{\Theta}) = \begin{bmatrix}
\frac{\partial^2 l}{\partial \alpha^2} & \frac{\partial^2 l}{\partial \alpha \partial \beta} & \frac{\partial^2 l}{\partial \alpha \partial \gamma} & \frac{\partial^2 l}{\partial \alpha \partial \sigma^2} \\
\frac{\partial^2 l}{\partial \beta \partial \alpha} & \frac{\partial^2 l}{\partial \beta^2} & \frac{\partial^2 l}{\partial \beta \partial \gamma} & \frac{\partial^2 l}{\partial \beta \partial \sigma^2} \\
\frac{\partial^2 l}{\partial \gamma \partial \alpha} & \frac{\partial^2 l}{\partial \gamma \partial \beta} & \frac{\partial^2 l}{\partial \gamma^2} & \frac{\partial^2 l}{\partial \gamma \partial \sigma^2} \\
\frac{\partial^2 l}{\partial \sigma^2 \partial \alpha} & \frac{\partial^2 l}{\partial \sigma^2 \partial \beta} & \frac{\partial^2 l}{\partial \sigma^2 \partial \gamma} & \frac{\partial^2 l}{(\partial \sigma^2)^2}
\end{bmatrix}$$

Calculate the second-order partial derivatives yields:

$$
\nabla^2 l(\mathbf{\Theta})
$$

$$
= \begin{bmatrix}
-\frac{n}{\sigma^2} & -\frac{\sum x_i}{\sigma^2} & -\frac{\sum x_i^2}{\sigma^2} & -\frac{\left(\sum y_i - n\alpha - \beta\sum x_i - \gamma\sum x_i^2\right)}{\sigma^4} \\
-\frac{\sum x_i}{\sigma^2} & -\frac{\sum x_i^2}{\sigma^2} & -\frac{\sum x_i^3}{\sigma^2} & -\frac{\left(\sum x_i y_i - \alpha\sum x_i - \beta\sigma x_i^2 - \gamma\sigma x_i^3\right)}{\sigma^4} \\
-\frac{\sum x_i^2}{\sigma^2} & -\frac{\sum x_i^3}{\sigma^2} & -\frac{\sum x_i^4}{\sigma^2} & -\frac{\left(\sum x_i^2 y_i - \alpha\sum x_i^2 - \beta\sum x_i^3 - \gamma\sum x_i^4\right)}{\sigma^4} \\
-\frac{\left(\sum y_i - n\alpha - \beta\sum x_i - \gamma\sum x_i^2\right)}{\sigma^4} & -\frac{\left(\sum x_i y_i - \alpha\sum x_i - \beta\sigma x_i^2 - \gamma\sigma x_i^3\right)}{\sigma^4} & -\frac{\left(\sum x_i^2 y_i - \alpha\sum x_i^2 - \beta\sum x_i^3 - \gamma\sum x_i^4\right)}{\sigma^4} & \frac{\partial^2 l}{(\partial\sigma^2)^2}
\end{bmatrix}
$$

The last element

$$
\begin{aligned}
\frac{\partial^2 l}{(\partial\sigma^2)^2} &= -\frac{n}{2}(-1)(\sigma^2)^{-2} + \frac{1}{2}(-2)(\sigma^3)^{-3}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i - \gamma x_i^2)^2 \\
&= \frac{n}{2\sigma^4} - \frac{\sum_{i=1}^{n}(y_i - \alpha - \beta x_i - \gamma x_i^2)^2}{\sigma^6}
\end{aligned}
$$

Hence, using the Newton-Raphson method, the iteration formula for $\mathbf{\Theta}$ is:

$$
\mathbf{\Theta}^{\text{new}} = \mathbf{\Theta}^{\text{old}} - \left(\nabla^2 l(\mathbf{\Theta}^{\text{old}})\right)^{-1}\left(\nabla l(\mathbf{\Theta}^{\text{old}})\right) \tag{2.3}
$$