Mathematics Fundamentals

Lecture 3: Multivariable and Integral Calculus

Manuele Leonelli

School of Science and Technology, IE University

Multivariable Functions

Multivariable functions are functions that take more than one input. Instead of depending on a single variable, they depend on multiple variables.

Definition

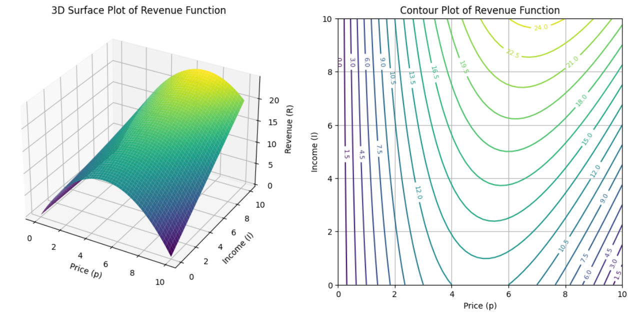
A multivariable function $f(x_1, \ldots, x_n)$ takes two inputs and assigns them a single output:

$$f : \mathbb{R}^n \to \mathbb{R}, \quad (x_1, \ldots, x_n) \mapsto f(x_1, \ldots, x_n)$$

- **Example:** The function $f(x, y) = x^2 + y^2$ describes a surface in three dimensions.
- Such functions are common in economics, physics, and machine learning, where outcomes depend on several factors.

Realistic Examples of Bivariate Functions

- **Revenue Function:** $R(p, I) = p \times D(p, I)$
  - p is the price per unit.
  - I is the consumer income.
  - $D(p, I) = a - bp + cI$ is the demand function, where a, b, c are constants.



Partial Derivatives

For a multivariable function $f(x_1, \ldots, x_n)$, we can compute the rate of change with respect to each variable while holding the other variables constant. These are called **partial derivatives**.

Definition

The partial derivative of $f(x_1, \ldots, x_n)$ with respect to x_i is:

$$\frac{\partial f}{\partial x_i} = \lim_{h \to 0} \frac{f(x_i + h, \boldsymbol{x}_{-i}) - f(x_i, \boldsymbol{x}_{-i})}{h}$$

The Gradient

The **gradient** of a multivariable function generalizes the concept of a derivative to multiple dimensions. It points in the direction of the steepest increase/decrease of the function.

Definition

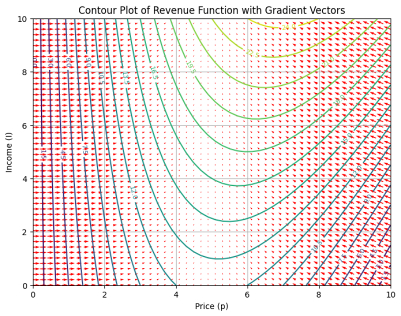
The gradient of a function $f(x_1, \dots, x_n)$ is the vector of its partial derivatives:

$$\nabla f(x_1, \dots, x_n) = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right)$$

- The gradient points in the direction of the steepest ascent.
- In optimization, we use the gradient to minimize or maximize functions.

Visualizing the Gradient

The gradient of a function $R(p, I)$ can be visualized on a contour plot, where each contour line represents a level of constant revenue.



- The contour lines represent levels of constant revenue.
- The red arrows show the direction and magnitude of the gradient, pointing towards higher revenue.

Critical Points and the Hessian Matrix

- For functions of many variables the critical points can be found solving the system of equations

$$\nabla f(x_1, \ldots, x_n) = (0, \ldots, 0)$$

- Critical points are used to determine local maxima and minima, but not all critical points are maxima or minima;
- The second derivative is now a symmetric matrix, called **Hessian**,

$$H(f(x_1, \ldots, x_n)) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)_{i,j},$$

whose eigenvalues tell us if a critical point is a maximum or minimum.

Why We Care About Optimization in High Dimensions

Key Applications

- **Training Neural Networks:**
  - Neural networks, especially deep learning models, involve optimizing millions or even billions of parameters.
  - Optimization techniques like gradient descent are essential for minimizing the loss function and improving model accuracy.
- **Large Language Models (LLMs):**
  - Models like GPT are trained using gradient descent to optimize parameters across vast datasets, enabling them to generate human-like text.
  - Fine-tuning LLMs on specific tasks also relies on optimization techniques to adapt the model to new data.

Introduction to Integration

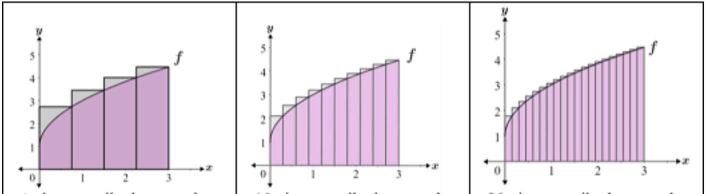
Integration is a fundamental concept in calculus used to solve practical problems like:

- **Calculating Areas:** Finding the area under curves, which is crucial in various applications.
- **Calculating Probabilities:** In probability theory, integrals are used to calculate the total probability under a probability density function (PDF).
- **Total Accumulation:** Determining the total accumulation of a quantity, such as distance traveled over time or total revenue generated.

The Idea Behind Integration

To understand integration, let's start with the concept of summing small pieces:

- Imagine dividing the area under a curve into small rectangles.
- The height of each rectangle is determined by the function value $f(x)$ at that point.
- The area of each rectangle is approximately $f(x) \times \Delta x$, where Δx is the width of the rectangle.



4 circumscribed rectangles

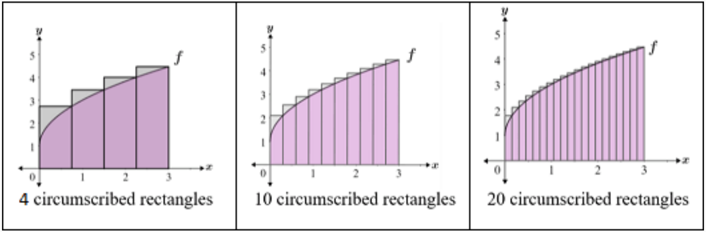
10 circumscribed rectangles

20 circumscribed rectangles

Riemann Sums and the Area Under the Curve

By summing the areas of all these rectangles, we approximate the total area under the curve:

- This sum is called a **Riemann sum**.
- As the rectangles become narrower (increasing in number), the approximation improves.
- In the limit, as the number of rectangles approaches infinity, the sum becomes the exact area under the curve.



The Fundamental Theorem of Calculus

The Fundamental Theorem of Calculus links differentiation and integration:

Fundamental Theorem of Calculus

If $f(x)$ is the derivative of $F(x)$, then:

$$\int_a^b f(x)\,dx = F(b) - F(a)$$

The operation of computing the area under the curve by summing the areas of infinitesimally small rectangles is what we define as **integration**.