

# Numerical Approximations of Derivatives

## 1 Overview

**Question:** How do we calculate derivatives numerically?

Often, finding the derivatives of functions analytically can be hard. Instead, we can approximate derivatives numerically by using discrete differences or finite differences. In fact, as long as we can evaluate the function, we can always approximate the derivative.

## 2 Derivative (numerical) approximation

In general, there are two types of approximations for derivatives: forward (backward) differences and central differences. Both types of approximation are derived from the limit definition of derivatives.

### 2.1 The forward/backward difference approximation

Here, we drop the limit operation to approximate the analytical derivative as:

$$\begin{aligned}\frac{df}{dx} &= \lim_{\Delta x \rightarrow 0} \frac{\Delta f}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \\ &\approx \frac{\Delta f}{\Delta x} = \frac{f(x + \Delta x) - f(x)}{\Delta x},\end{aligned}\tag{1}$$

for a small  $\Delta x$ . This approximation is called *the forward difference approximation* of the first derivative. Its analogous counterpart, the *backward difference*, is given by:

$$\frac{df}{dx} \approx \frac{\Delta f}{\Delta x} = \frac{f(x) - f(x - \Delta x)}{\Delta x}.\tag{2}$$

## 2.2 The central difference approximation

We can also use a slightly modified definition of the analytical derivative of a scalar function of a single scalar variable, which is given by:

$$\frac{df}{dx} = \lim_{\Delta x \rightarrow 0} \frac{\Delta f}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x - \Delta x)}{2\Delta x}. \quad (3)$$

We can drop the limit in the calculation of Equation 3 to obtain the *central approximation* of the derivative which is given by:

$$\frac{df}{dx} \approx \frac{\Delta f}{\Delta x} \approx \frac{f(x + \Delta x) - f(x - \Delta x)}{2\Delta x}. \quad (4)$$

## 3 Representing function values on grids

Numerically, the value of functions are represented as a discrete grid of values, which can be stored in multi-dimensional arrays. The array's dimensionality depends on the type of function and on implementation decisions. For example, a scalar function of a scalar variable can be stored in a 1-D array while a scalar function of two variables can be stored in a  $M \times N$  matrix (i.e., a multidimensional array). Here,  $M$  is the number of rows in the matrix and  $N$  is the number of columns. Similarly, a scalar function of three variables can be stored in a  $M \times N \times K$  array. Figure 1 illustrates the values of the Rosenbrock function stored in a  $18 \times 18$  matrix.

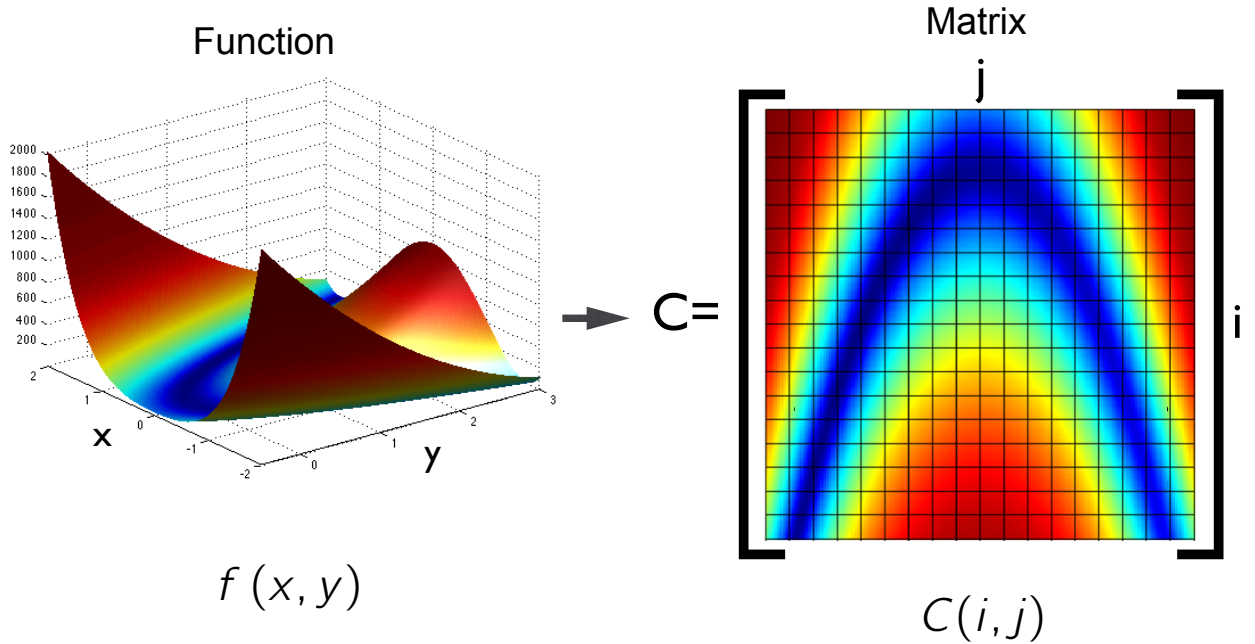


Figure 1: The Rosenbrock function stored in a  $18 \times 18$  matrix  $C$ .

## 4 Discrete derivatives on grids

Given the matrix representation of the function, we can approximate the function's derivatives (gradient, Jacobian) by means of discrete differences. The derivative approximation can be calculated at all points of the function's domain (if the derivative exists, of course). Figure 2 shows a sample region of the matrix and the function values.

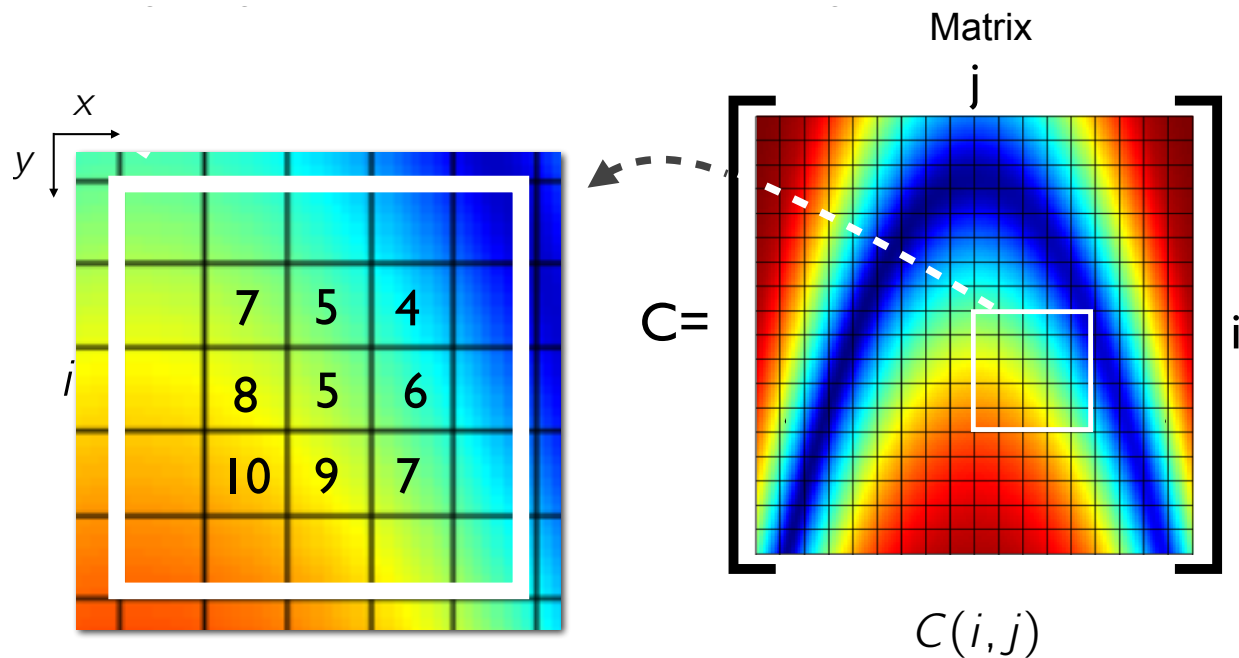


Figure 2: A sample region of function values from the matrix.

To calculate the gradient at a given position  $(x, y)$ , we treat the matrix as a grid of values and calculate the required derivatives using finite differences.

**Important:** When working with Cartesian coordinate systems and multi-dimensional arrays, be aware of the relationship between array indices  $(i, j)$  and Cartesian coordinates  $(x, y)$  as they are not the same!

When calculating the derivative's discrete approximations, we need to convert between the Cartesian coordinates and their corresponding locations in the matrix indexing system, which uses  $(i, j)$  for (row,column) indexing. We also need to be aware of the location of the origin of the Cartesian coordinate that we are using, e.g., the  $(x, y) = (0, 0)$  is at the top-left corner of the coordinate system with the y-axis pointing downwards, the  $(x, y) = (0, 0)$  is at the bottom-left corner of the coordinate system and the y-axis points upwards. Two examples of these grid representations are shown in Figure 3.

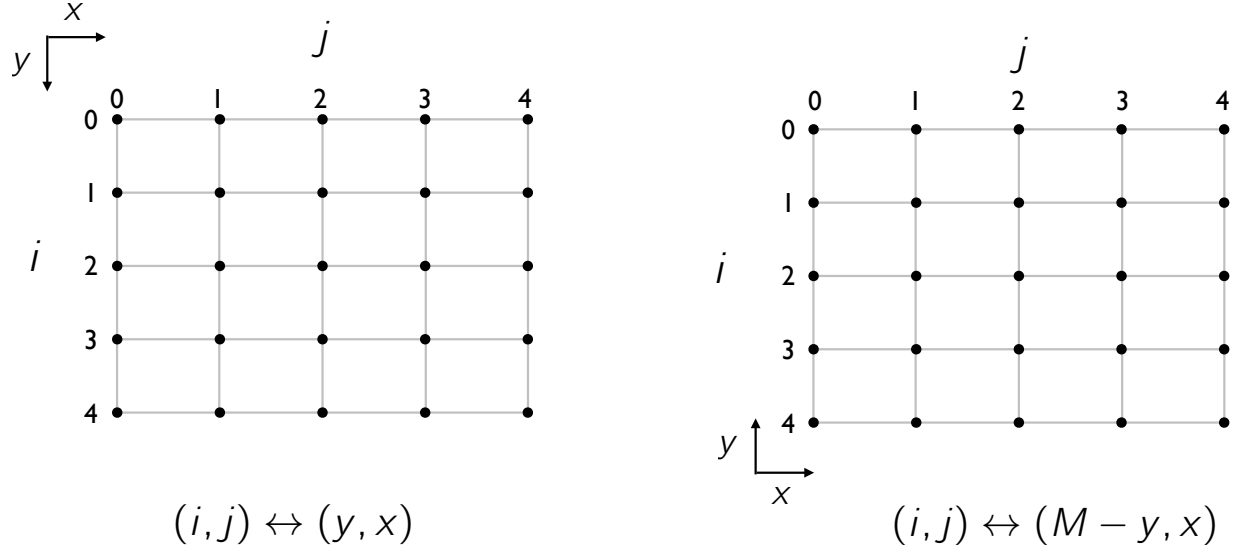
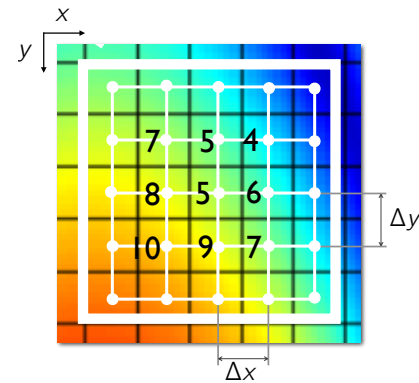
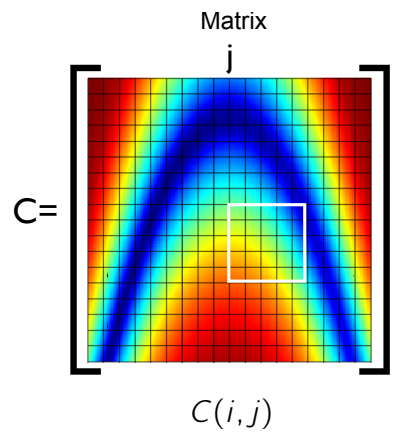
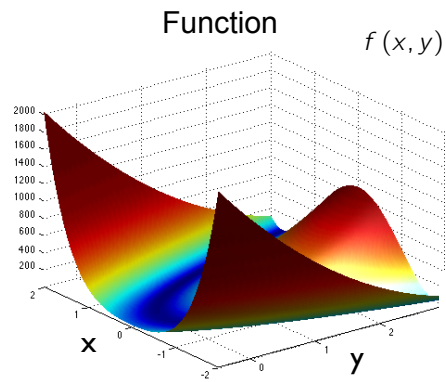


Figure 3: Matrix indexing vs. Cartesian coordinates.

Figure 4 shows a graphical overview of the gradient calculation using central derivatives.



Calculating the gradient of a function on a discrete grid

$$\nabla f(x, y) = \begin{bmatrix} \frac{\partial f(x, y)}{\partial x} \\ \frac{\partial f(x, y)}{\partial y} \end{bmatrix} \approx \begin{bmatrix} \frac{\Delta f}{\Delta x} \\ \frac{\Delta f}{\Delta y} \end{bmatrix}$$

$$\Delta x = \Delta y = 1$$

$$\frac{\partial f(x, y)}{\partial x} \approx \frac{\Delta f}{\Delta x} = \frac{f(x + \Delta x) - f(x - \Delta x)}{2\Delta x} = \frac{C_{i, j+1} - C_{i, j-1}}{2} = \frac{6 - 8}{2} = -1$$

$$\frac{\partial f(x, y)}{\partial y} \approx \frac{\Delta f}{\Delta y} = \frac{f(y + \Delta y) - f(y - \Delta y)}{2\Delta y} = \frac{C_{i+1, j} - C_{i-1, j}}{2} = \frac{9 - 5}{2} = 2$$

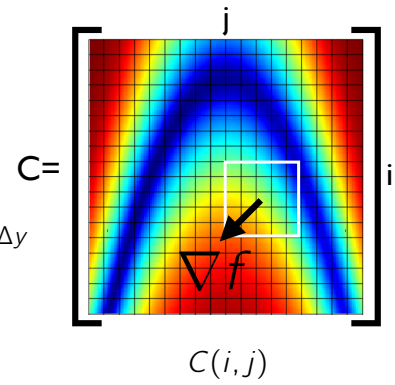
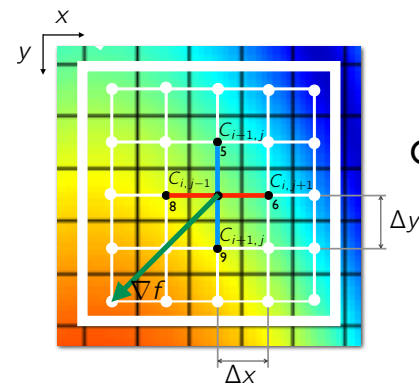
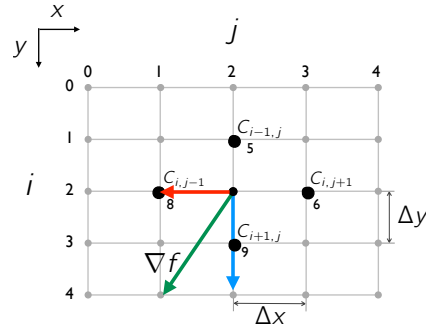


Figure 4: Gradient approximation using central differences.