Mathematical background

- Functions
- Derivatives
- Optimization using numerical libraries
- Numerical approximations of derivatives

Functions

Definition: A function is a mathematical relationship in which the values of a single dependent variable are determined by the values of one or more independent variables. Function means the dependent variable is determined by the independent variable(s).[a].

[a] http://www.columbia.edu/itc/sipa/math/variables.html

Types of function (dimensionality of variables)

1. **A scalar function of a single scalar variable**, i.e., $f(x) \in \mathbb{R}$ and $x \in \mathbb{R}$.

2. **A scalar function of multiple scalar variables (or a scalar function of a vector variable)**. In this case, $f(\mathbf{x}) \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$.

3. **A vector function of a single scalar variable**, i.e., $\mathbf{f}(x) \in \mathbb{R}^n$ and $x \in \mathbb{R}$.

4. **A vector function of a vector variable**, i.e., $\mathbf{f}(\mathbf{x}) \in \mathbb{R}^n$ and $\mathbf{x} \in \mathbb{R}^n$.

2.1 A scalar function of a single scalar variable

In scalar functions of a single scalar variable, both the function value, $f(x)$, and its independent variable x are real numbers, i.e., $f(x) \in \mathbb{R}$ and $x \in \mathbb{R}$. Figure 1 shows an example.

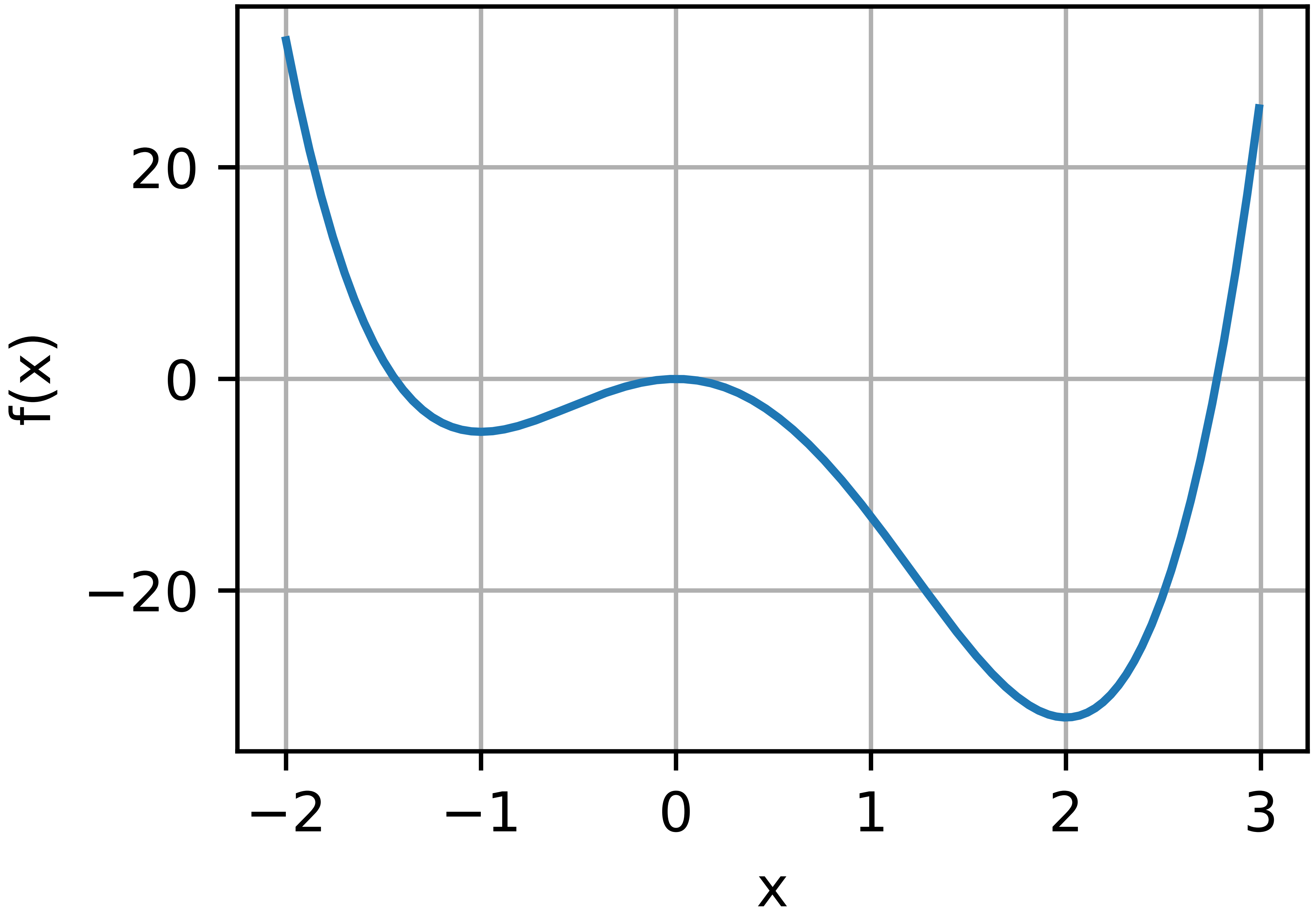


Figure 1: A scalar function of a single scalar variable, $f(x) = 3x^4 - 4x^3 - 12x^2$.

2.2 A scalar function of multiple scalar variables (or a scalar function of a vector variable)

In this case, the function value $f(\mathbf{x})$ is a (single) real number, which is dependent on multiple independent variables $\mathbf{x} = (x_1, \ldots, x_n)^\mathsf{T}$, i.e., $f(\mathbf{x}) \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$. A bold lowercase letter denotes a vector variable.

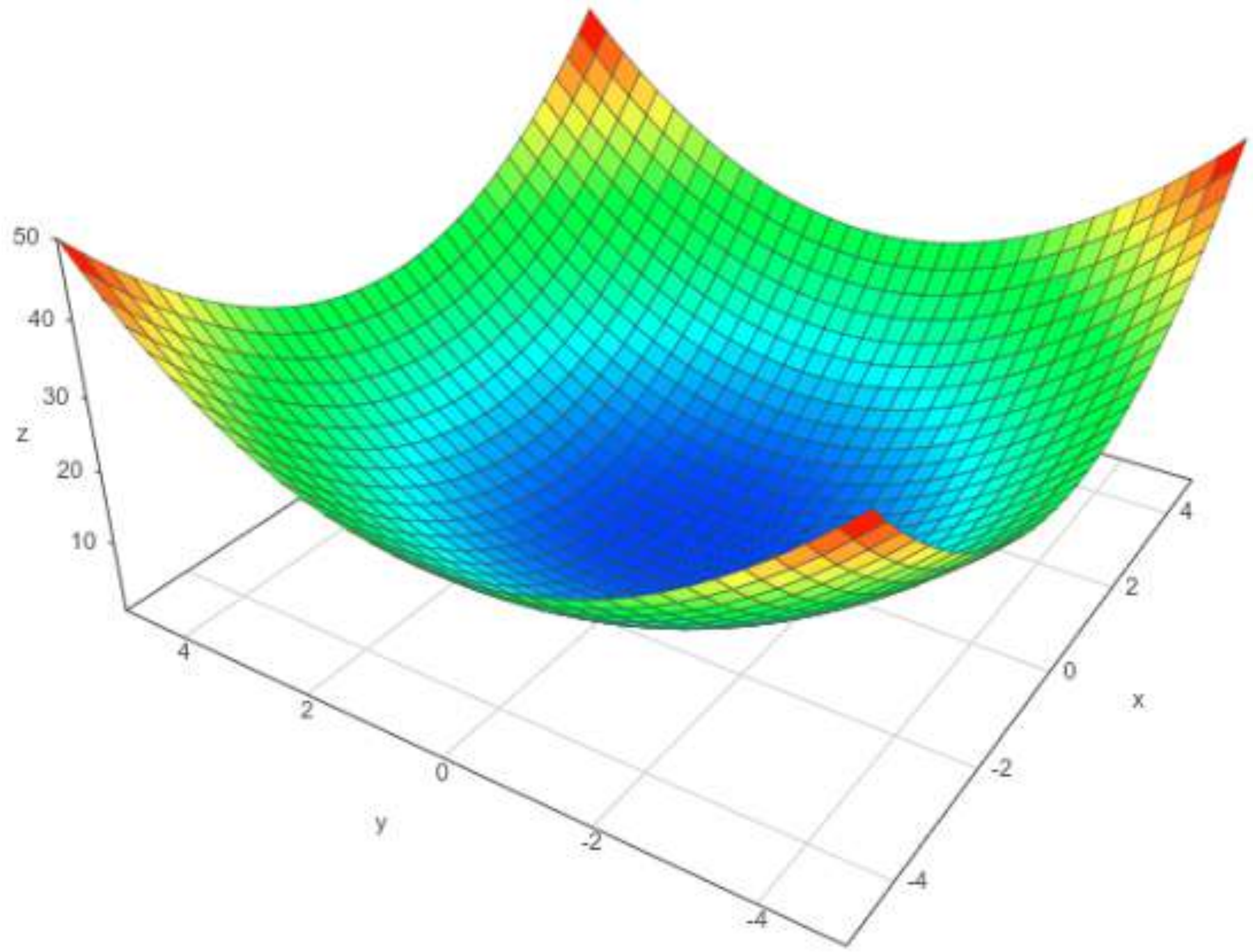


Figure 2: A scalar function of multiple variables (or vector variable), $f(x, y) = x^2 + y^2$.

2.3 A vector function of a single scalar variable

In these types of functions, it is the function value $\mathbf{f}(x)$ that is a vector $\mathbf{f}(x) = (x_1, \ldots, x_n)^\mathsf{T}$ (or multivariate). But, this value $\mathbf{f}(x)$ depends on a single scalar variable, i.e., $\mathbf{f}(x) \in \mathbb{R}^n$ and $x \in \mathbb{R}$. For example, the motion of a particle as a function of time.

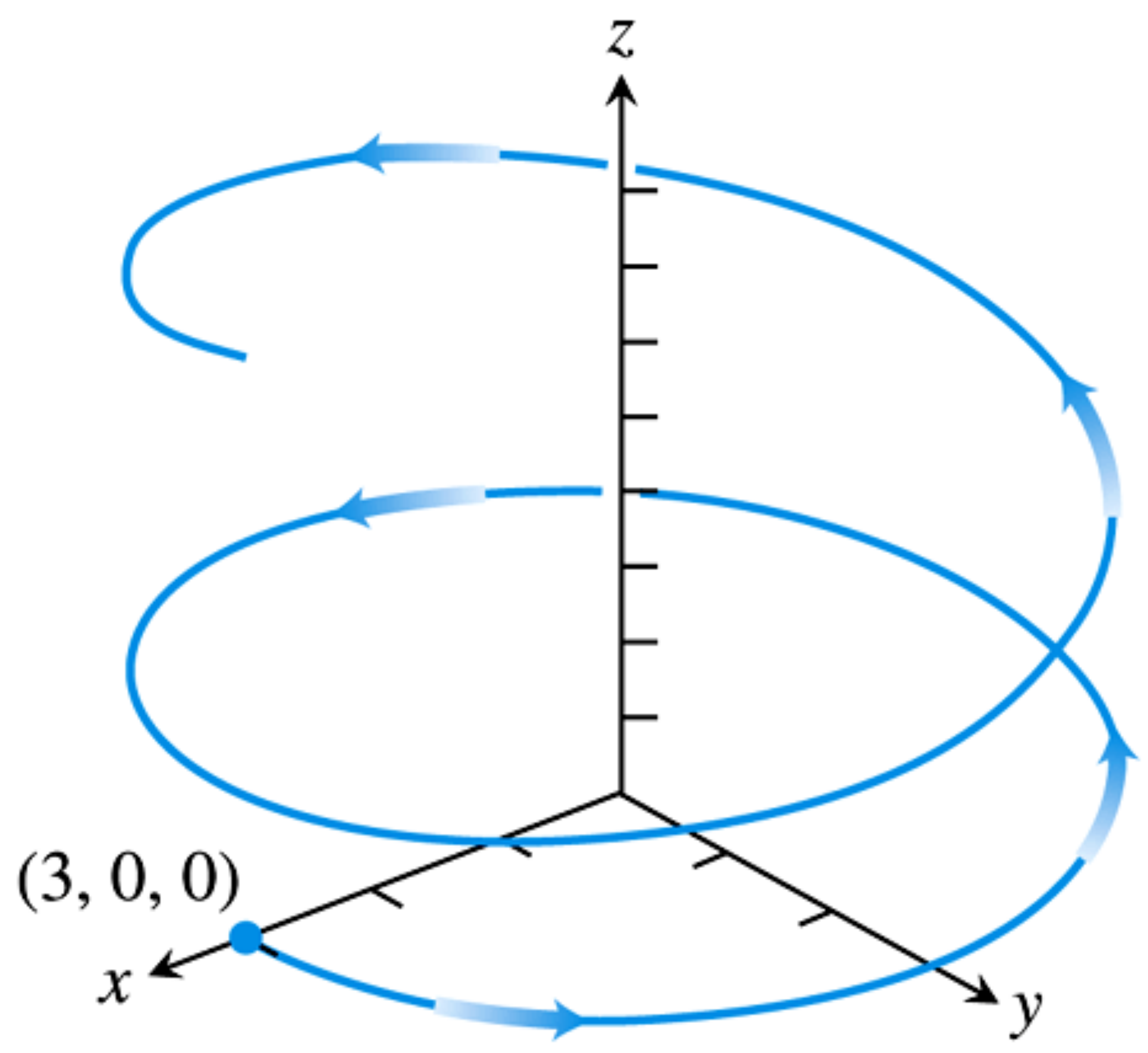


Figure 3: A vector function of a scalar variable. The path of a hang glider with position vector $\mathbf{r}(t) = (3\cos t)\,\mathbf{i} + (3\sin t)\,\mathbf{j} + t^2\,\mathbf{k}$.

2.4 A vector function of a vector variable

For some functions, both the dependent and independent variables are vectors (i.e., multiple dependent variables are dependent of multiple independent variables). In this case, $\mathbf{f}$ is a vector-valued function of a vector of variables, $\mathbf{x}$. Here, $\mathbf{f}(\mathbf{x}) = (f_1, f_2, \ldots, f_M)^\mathsf{T}$ and $\mathbf{x} = (x_1, x_2, \ldots, x_N)^\mathsf{T}$, with $\mathbf{f}(\mathbf{x}) \in \mathbb{R}^n$ and $\mathbf{x} \in \mathbb{R}^n$. Figure 4 shows an example of the motion of an articulated robot arm. The arm's pose and the 3-D location of its tip are given in terms of a set of joint angles. The figure also shows that a vector function of a vector variable describes a vector field.

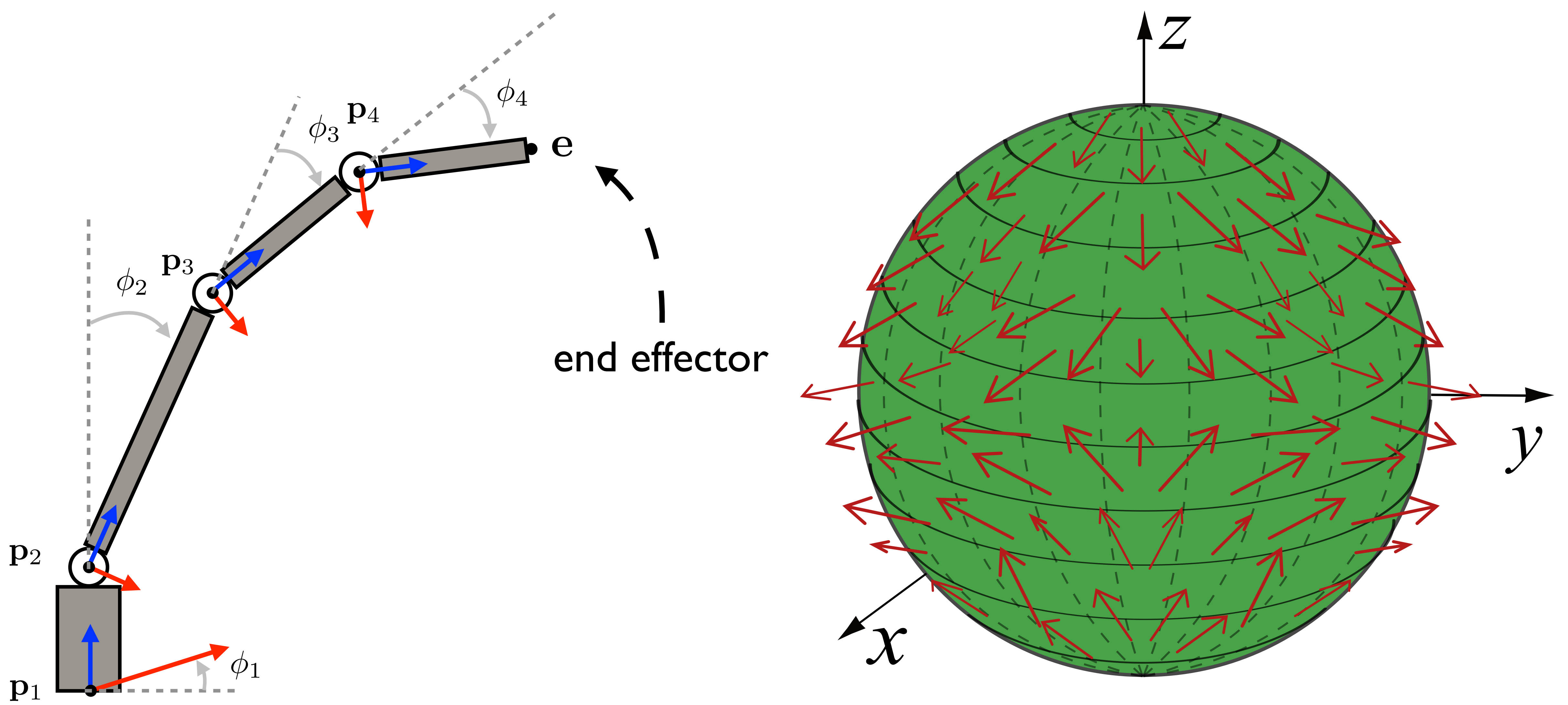


Figure 4: Vector functions of a vector variable. A robot arm and a vector field on a sphere (Figure from: `https://en.wikipedia.org/wiki/Vector_field`).

Summary: Types of function (dimensionality of variables)

1. **A scalar function of a single scalar variable**, i.e., $f(x) \in \mathbb{R}$ and $x \in \mathbb{R}$.

2. **A scalar function of multiple scalar variables (or a scalar function of a vector variable)**. In this case, $f(\mathbf{x}) \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$.

3. **A vector function of a single scalar variable**, i.e., $\mathbf{f}(x) \in \mathbb{R}^n$ and $x \in \mathbb{R}$.

4. **A vector function of a vector variable**, i.e., $\mathbf{f}(\mathbf{x}) \in \mathbb{R}^n$ and $\mathbf{x} \in \mathbb{R}^n$.

Derivatives

2 Derivative of a scalar function of a single scalar variable

Let f be a scalar function of a single variable x. The derivative of the function w.r.t. x is:

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x} = \lim_{\Delta x \to 0} \frac{f\left(x + \Delta x\right) - f\left(x\right)}{\Delta x}. \tag{1}$$

It describes the slope (i.e., rate of change) of the function at a point x.

Derivative of a scalar function of a single scalar variable

$$f(x) = x \sin x^2 + 1$$

$$\frac{df}{dx}(x) = 2x^2 \cos x^2 + \sin x^2$$

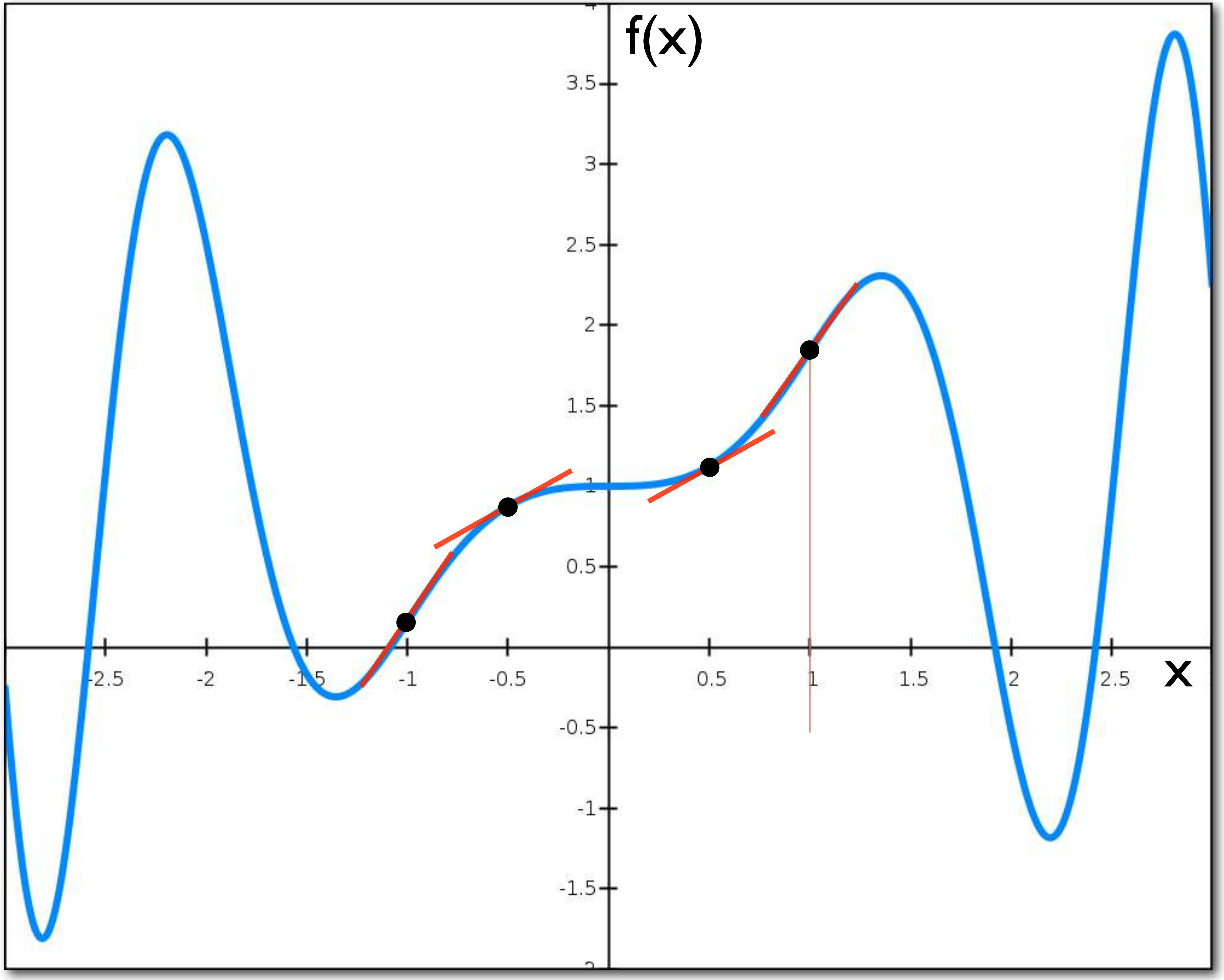


Figure 1: The derivative of the function $f(x) = x \sin x^2 + 1$ at $x = -1$, -0.5, 1, 0.5, and 1.

3 Derivative of a scalar function of multiple scalar variables (i.e., vector variable)

Let f be a scalar function of a vector variable. The vector variable is $\mathbf{x} = (x_1, x_2, \ldots, x_N)^\mathsf{T}$. This type of function is also called a scalar function of multiple variables or multi-variate function. The value of the function at a point $\mathbf{x}$ is given by $f(\mathbf{x})$ or $f(x_1, x_2, \ldots, x_N)$, and its derivative w.r.t. x is:

$$\frac{\mathrm{d}f}{\mathrm{d}\mathbf{x}} = \frac{\partial f}{\partial (x_1, x_2, \ldots, x_N)} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_N} \end{bmatrix}^\mathsf{T} = \boldsymbol{\nabla} f, \tag{4}$$

which is called the *Gradient* of f at x, and denoted by $\boldsymbol{\nabla} f$. Note that the components of $\boldsymbol{\nabla} f$ are simply the derivatives of function f w.r.t. each independent variable x_i (in isolation) calculated as described by Equation 1.

The gradient is a vector field because it yields a vector at each location $\mathbf{x} = (x_1, x_2, \ldots, x_N)^{\mathsf{T}}$ of the domain of the function.

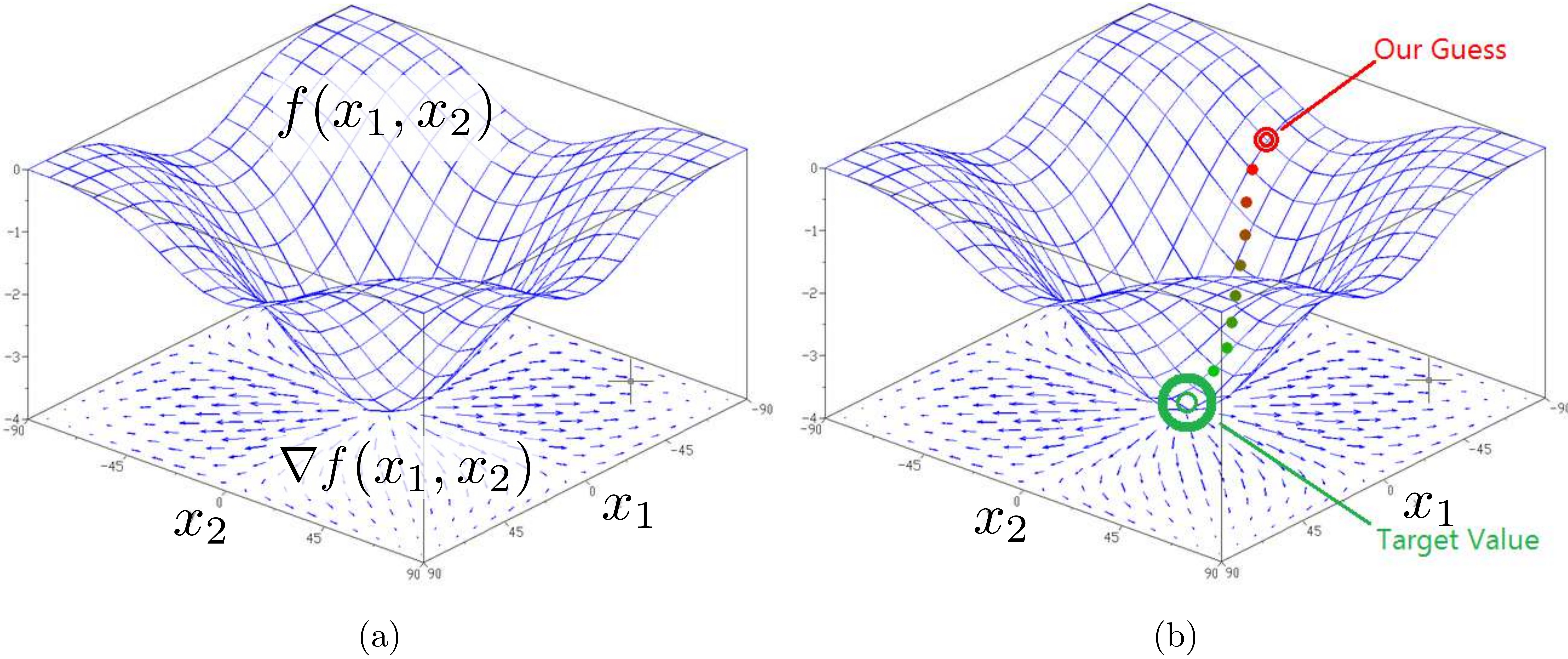


Figure 2: (a) Gradient vector field of two-variate function. (b) Gradient-descent method. In the gradient-descent method, we want to take the direction inverse to the gradient as the gradient (slope) points "uphill". Plots adapted from `https://goo.gl/zejgBw`.

4 Derivative of a vector function of a single scalar variable

Let $\mathbf{r}$ be a vector function representing the 3-D motion of a particle as a function of time t:

$$\mathbf{r} = [r_x \ r_y \ r_z]^\mathsf{T} . \tag{5}$$

Its derivative is given by:

$$\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \left[\frac{\mathrm{d}r_x}{\mathrm{d}t} \ \frac{\mathrm{d}r_y}{\mathrm{d}t} \ \frac{\mathrm{d}r_z}{\mathrm{d}t}\right]^\mathsf{T} , \tag{6}$$

which is also a vector field representing the velocity of the particle at time t. This vector field runs along the curve that describes the motion. Similar to the gradient, each component of the vector field in Equation 6 are calculated by Equation 1.

5 Derivative of a vector function of a vector variable

Some applications require us to calculate derivatives of vector quantifies with respect to other vector quantities. For example, if $\mathbf{f}$ is a vector-valued function of a vector of variables, $\mathbf{x}$. Here, $\mathbf{f}(\mathbf{x}) = (f_1, f_2, \ldots, f_M)^\mathsf{T}$ and $\mathbf{x} = (x_1, x_2, \ldots, x_N)^\mathsf{T}$.

The derivative of $\mathbf{f}$ w.r.t. $\mathbf{x}$ is:

$$\frac{\mathrm{d}\mathbf{f}}{\mathrm{d}\mathbf{x}} = J(\mathbf{f}, \mathbf{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_N} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_M}{\partial x_1} & \frac{\partial f_M}{\partial x_2} & \cdots & \frac{\partial f_M}{\partial x_N}, \end{bmatrix} \tag{7}$$

and is called the *Jacobian.*

The Jacobian gives us all we need to know about the variations of function with respect to all variables. It is hard to visualize the Jacobian. However, we can think of it as a

Optimization using numerical libraries

1 Overview

Question: Why do we calculate derivatives analytically if computers can calculate them numerically?

One answer to the above question is that numerical-optimization methods produce faster and more precise results when we provide them with the analytical derivative (or gradient, or Jacobian) of the objective function. This point is discussed further in these notes.

Its global minimum is at $(x, y) = (a, a^2)$, where $f(x, y) = 0$. The parameters are usually set to $a = 1$ and $b = 100$, i.e.:

$$f(x, y) = (1 - x)^2 + 100(y - x^2)^2. \tag{3}$$

The Rosenbrock function

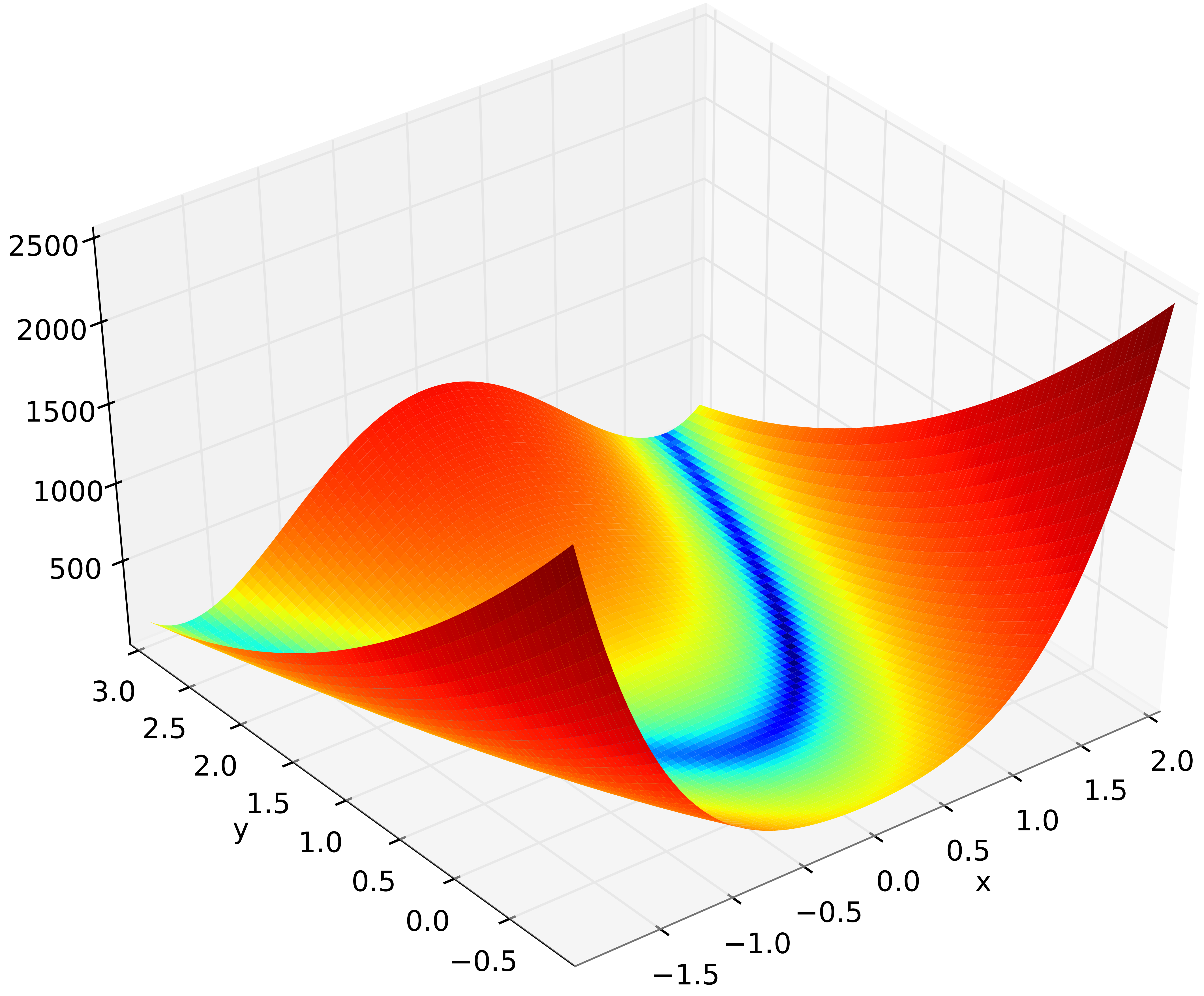


Figure 1: The Rosenbrock function. The global minimum is inside a long, narrow, parabolic shaped flat valley (Figure from `https://en.wikipedia.org/wiki/Rosenbrock_function`)

Optimizing without passing the analytical derivative

$$f(x, y) = (1 - x)^2 + 100(y - x^2)^2$$

```
def f(x):    # The Rosenbrock function
...     return (1.0 - x[0])**2 + 100.0*(x[1] - x[0]**2)**2
>>> optimize.minimize(f, [2, -1], method="CG")
     fun: 1.6...e-11
     jac: array([-6.15...e-06,   2.53...e-07])
 message: ...'Optimization terminated successfully.'
    nfev: 108
     nit: 13
    njev: 27
  status: 0
 success: True
       x: array([0.99999...,   0.99998...])
```

Optimizing passing the analytical derivative

$$f(x,y) = (1-x)^2 + 100(y-x^2)^2 \longrightarrow \nabla f(x,y) = \begin{bmatrix} -2(1-x) - 400\,y\,(y-x^2) \\ 200(y-x^2) \end{bmatrix}.$$

```
def jacobian(x):
                    Gx = -2.0*(1 - x[0]) - 400.0*x[0]*(x[1] - x[0]**2)
          Gy = 200.0*(x[1] - x[0]**2)
...       return np.array((Gx,Gy))
>>> optimize.minimize(f, [2, 1], method="CG", jac=jacobian)
     fun: 2.957...e-14
     jac: array([ 7.1825...e-07,  -2.9903...e-07])
 message: 'Optimization terminated successfully.'
    nfev: 16
     nit: 8
    njev: 16
  status: 0
 success: True
       x: array([1.0000...,  1.0000...])
```

When comparing the two optimization results, we can see that the second option is both more efficient (i.e., less function evaluations and less iterations) and more accurate (i.e., minimum was found at $[1,1]$ instead of at $[0.99999, 0.99999]$) than the first option.

Numerical Approximations of Derivatives

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.

2.1 The forward/backward difference approximation

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

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x} = \lim_{\Delta x \to 0} \frac{f\left(x + \Delta x\right) - f\left(x\right)}{\Delta x}$$

$$\approx \frac{\Delta f}{\Delta x} = \frac{f\left(x + \Delta x\right) - f\left(x\right)}{\Delta x}, \tag{1}$$

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

$$\frac{\mathrm{d}f}{\mathrm{d}x} \approx \frac{\Delta f}{\Delta x} = \frac{f\left(x\right) - f\left(x - \Delta x\right)}{\Delta x}. \tag{2}$$

2.1 The forward/backward difference approximation

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

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x} = \lim_{\Delta x \to 0} \frac{f\left(x + \Delta x\right) - f\left(x\right)}{\Delta x}$$

$$\approx \frac{\Delta f}{\Delta x} = \frac{f\left(x + \Delta x\right) - f\left(x\right)}{\Delta x}, \tag{1}$$

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

$$\frac{\mathrm{d}f}{\mathrm{d}x} \approx \frac{\Delta f}{\Delta x} = \frac{f\left(x\right) - f\left(x - \Delta x\right)}{\Delta x}. \tag{2}$$

Representing function values on grids

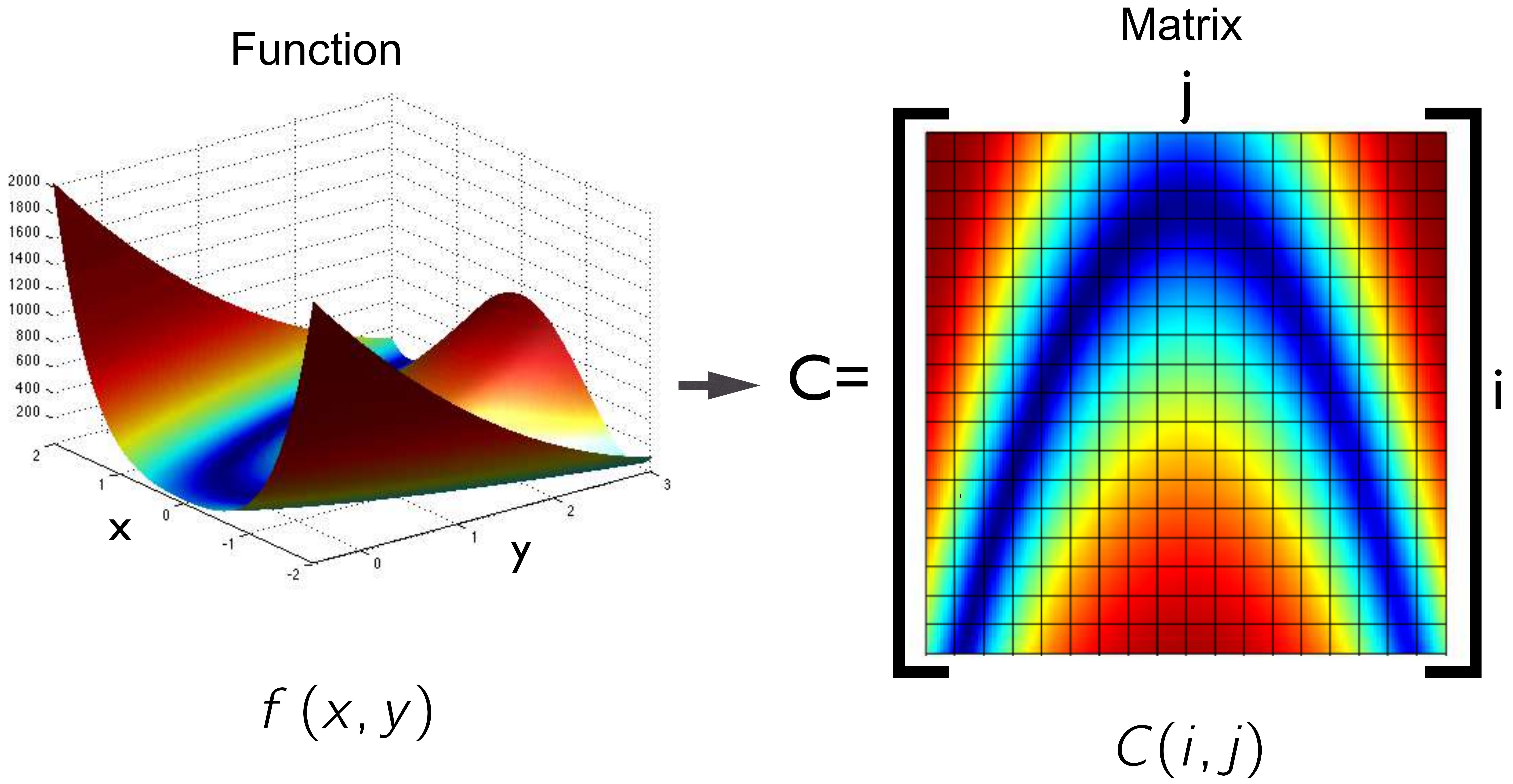


Figure 1: The Rosenbrock function stored in a 18×18 matrix C.

Discrete derivatives on grids

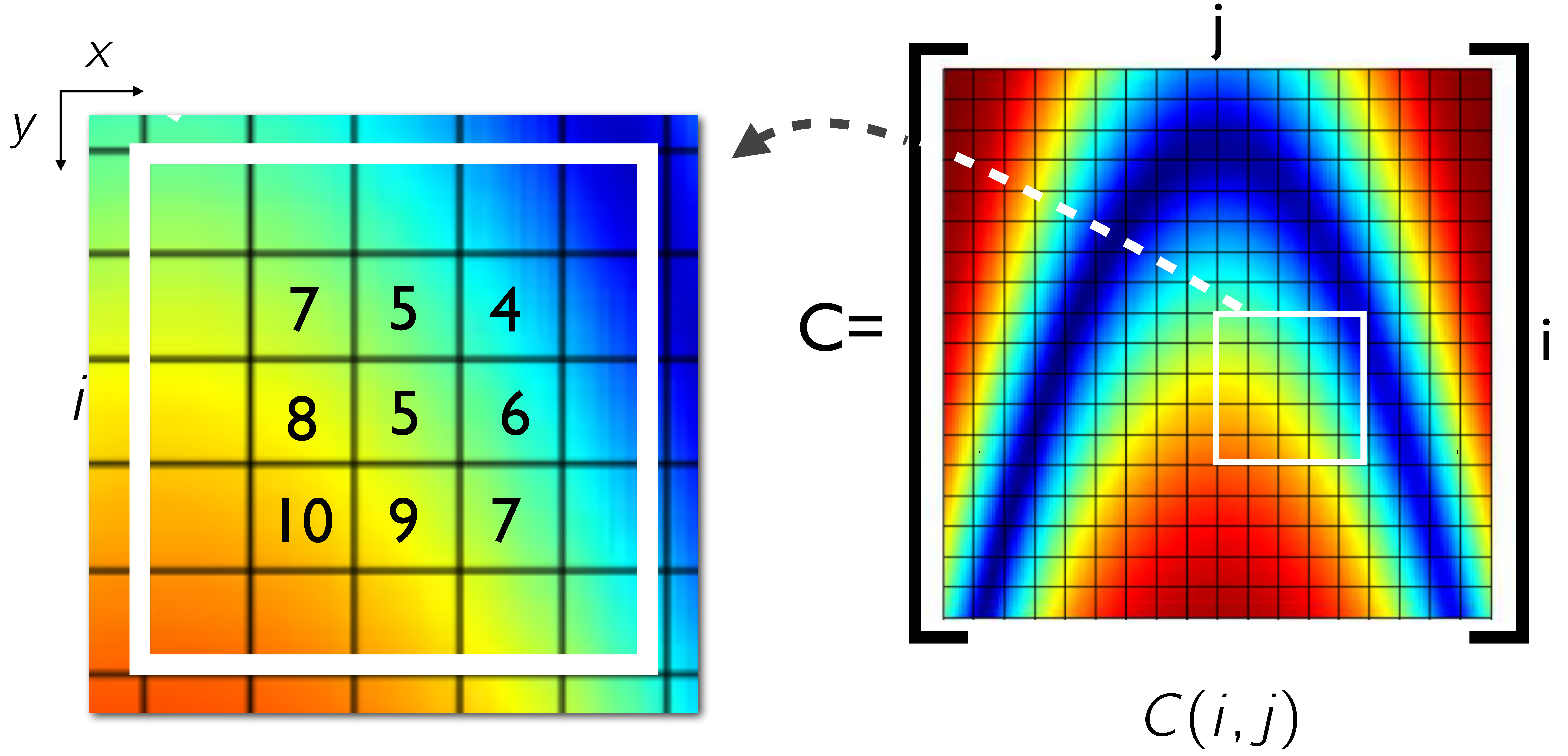


Figure 2: A sample region of function values from the matrix.

Discrete derivatives on grids

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!

Discrete derivatives on grids

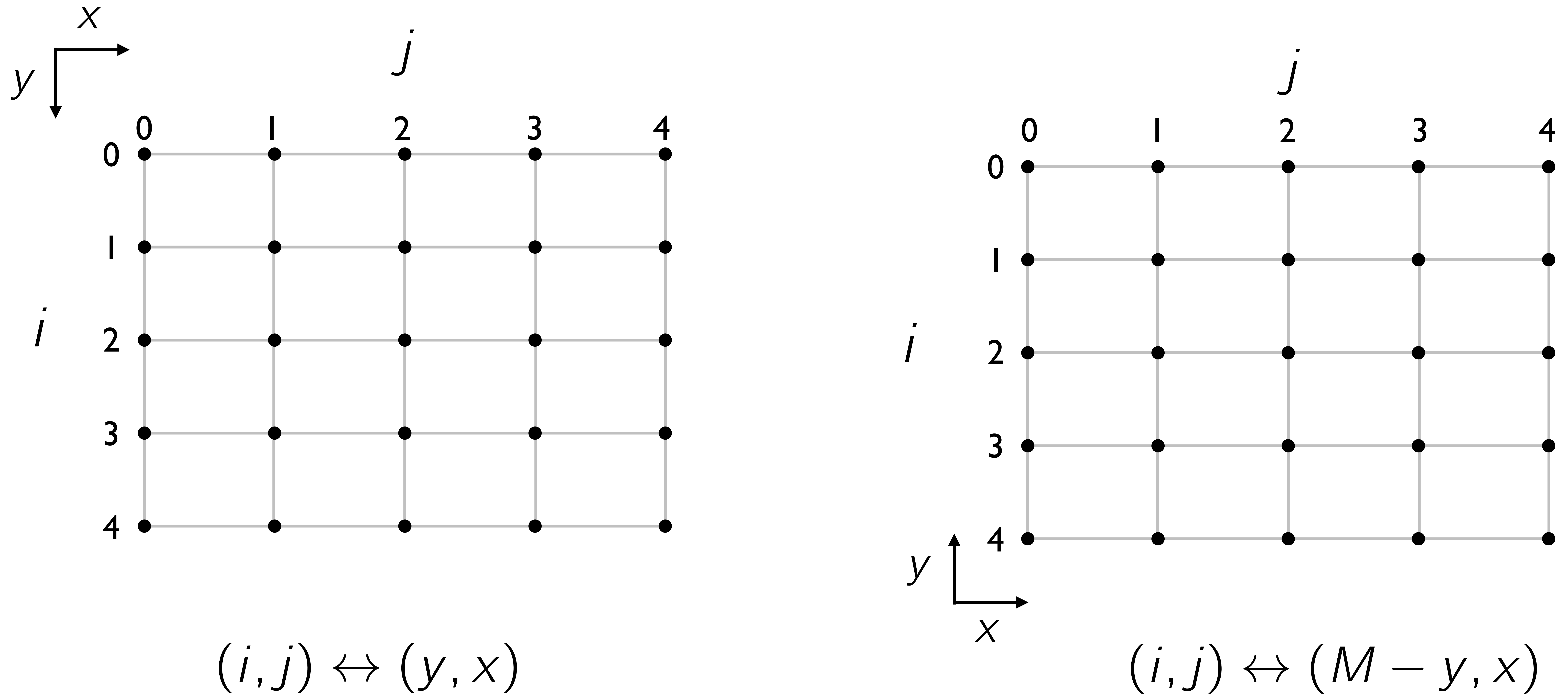
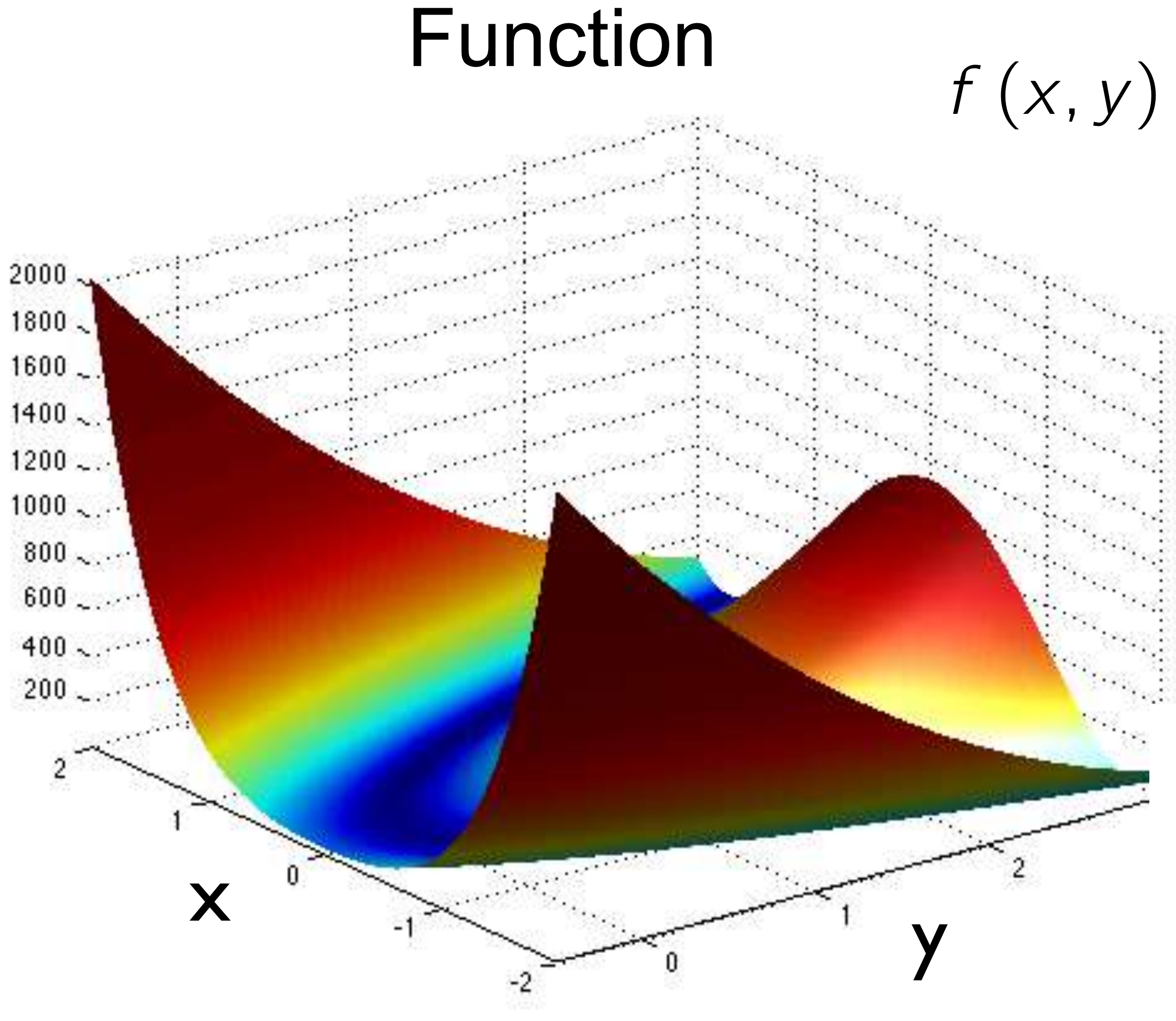
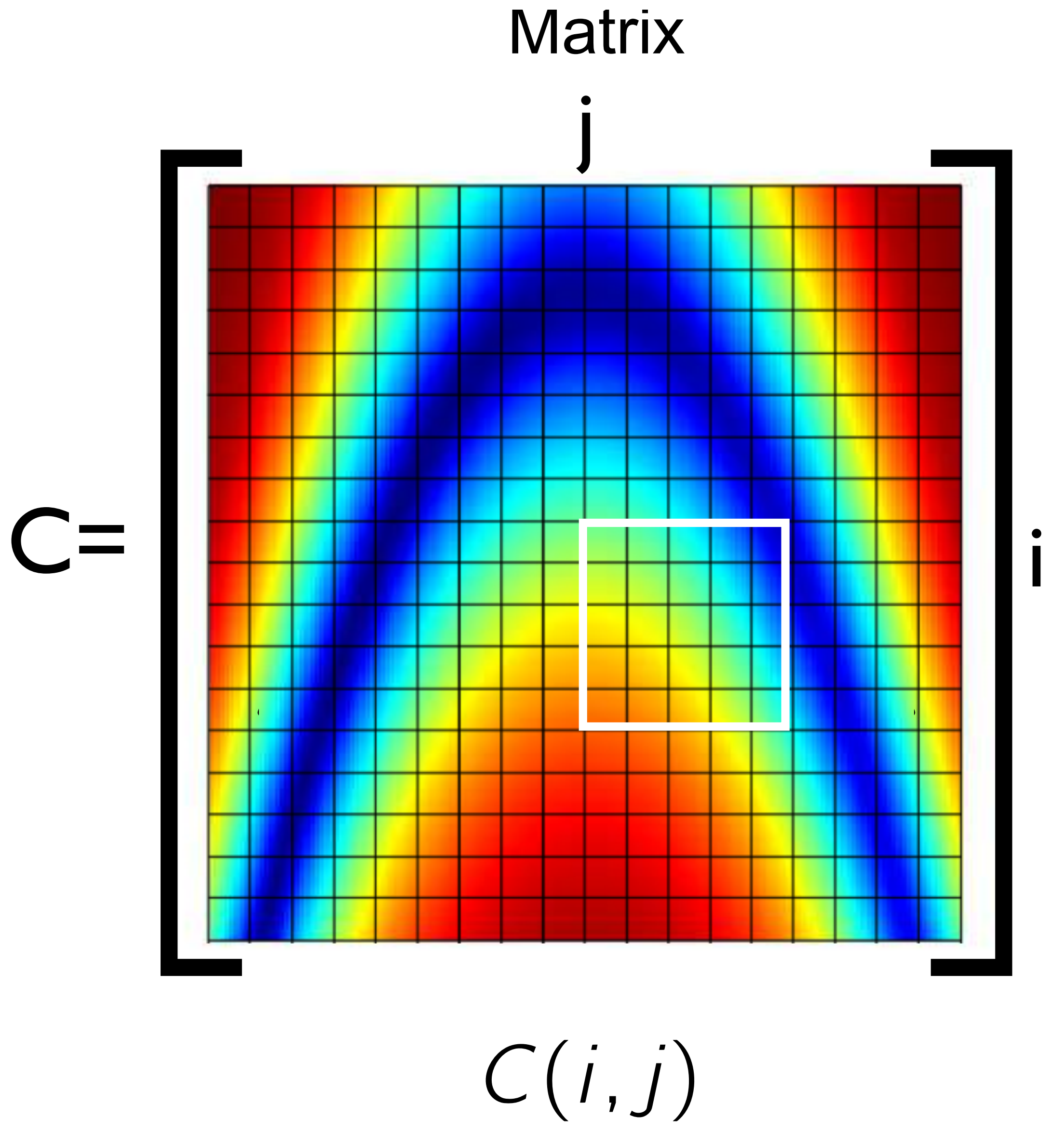
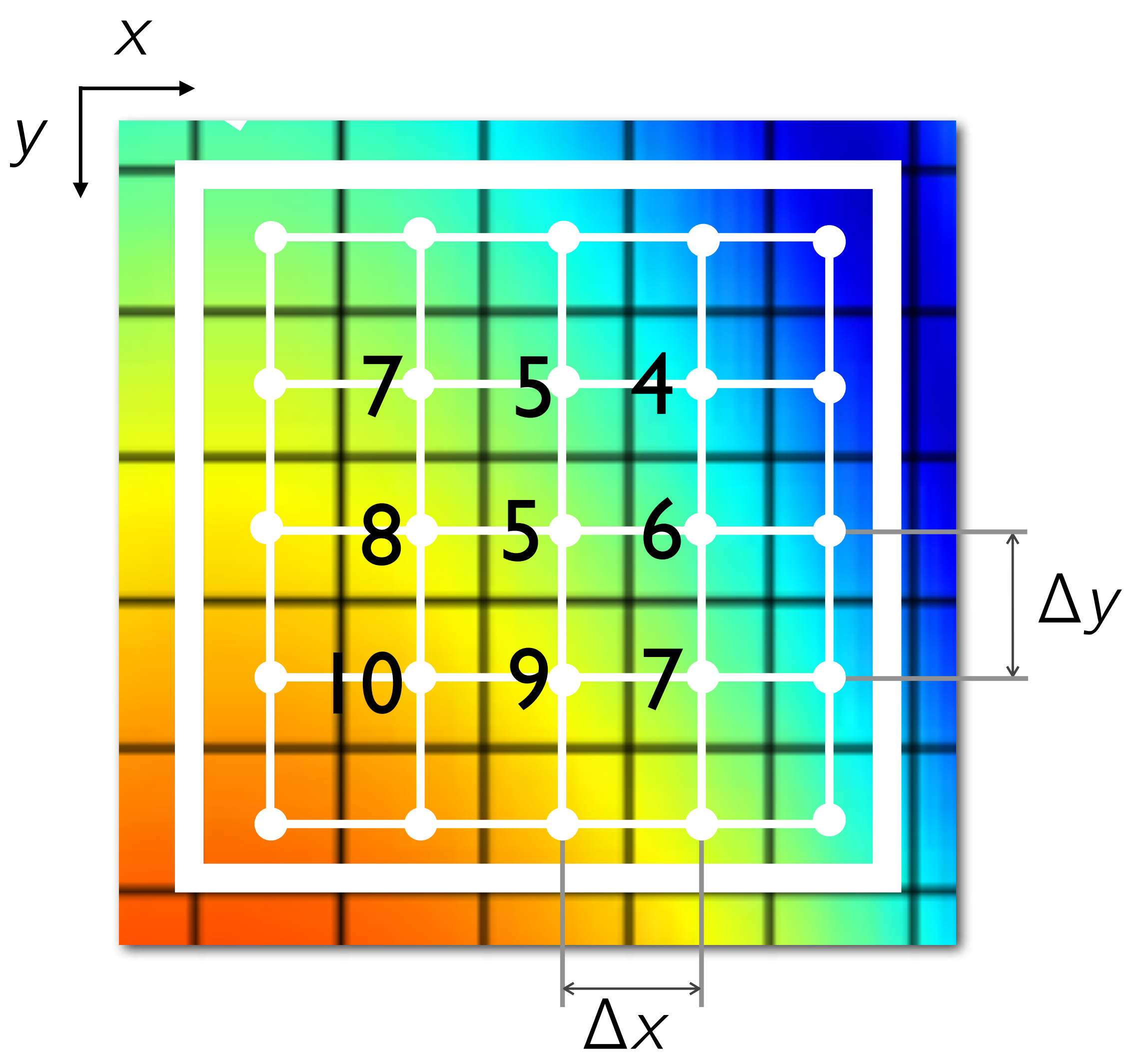


Figure 3: Matrix indexing vs. Cartesian coordinates.

Gradient approximation using central differences.

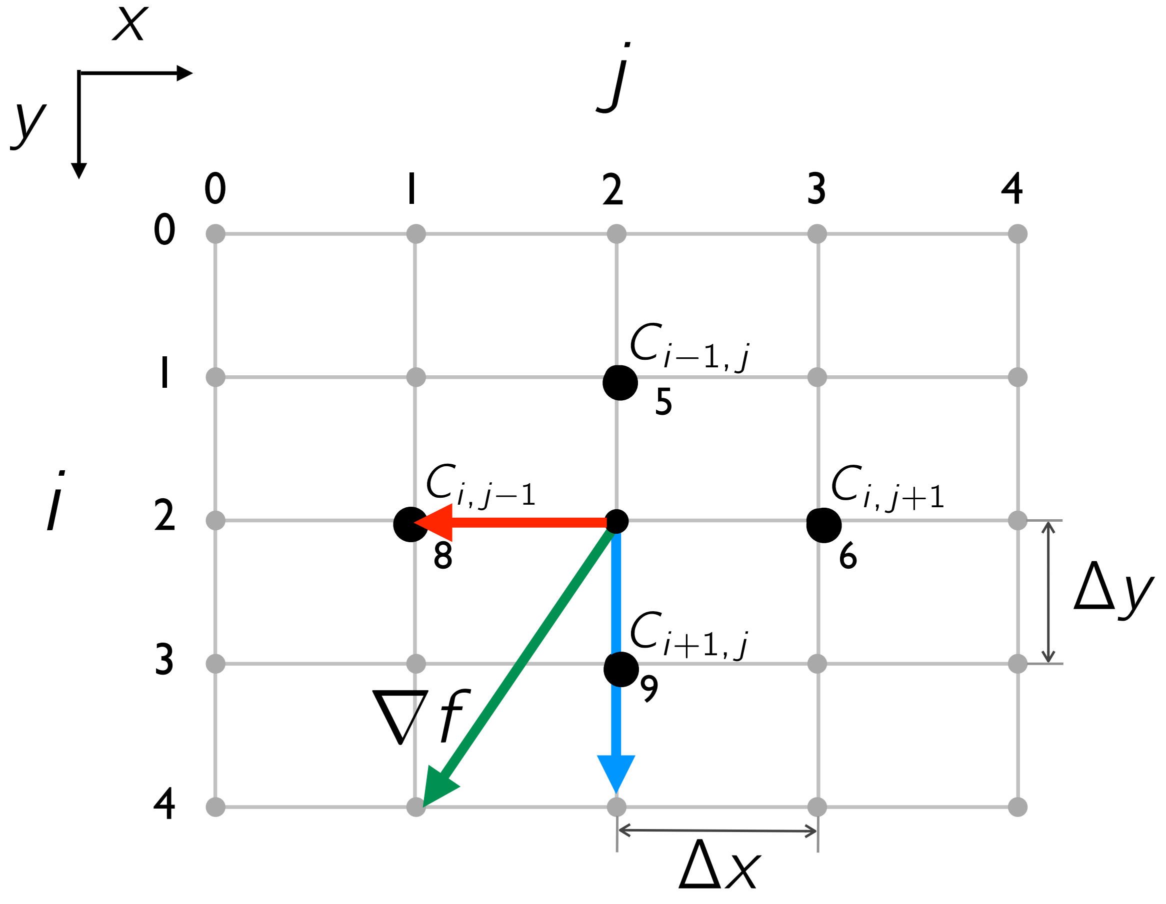


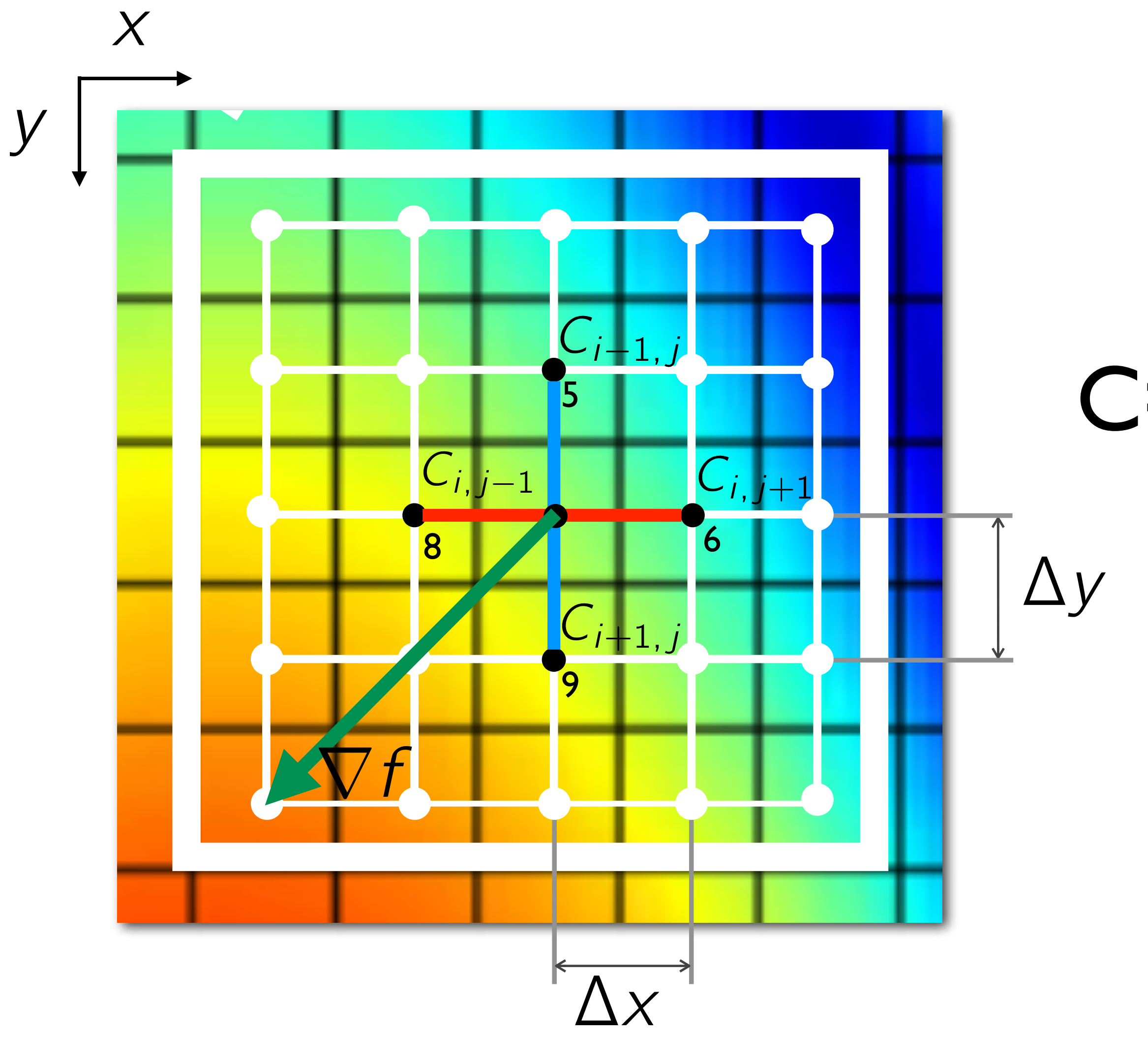


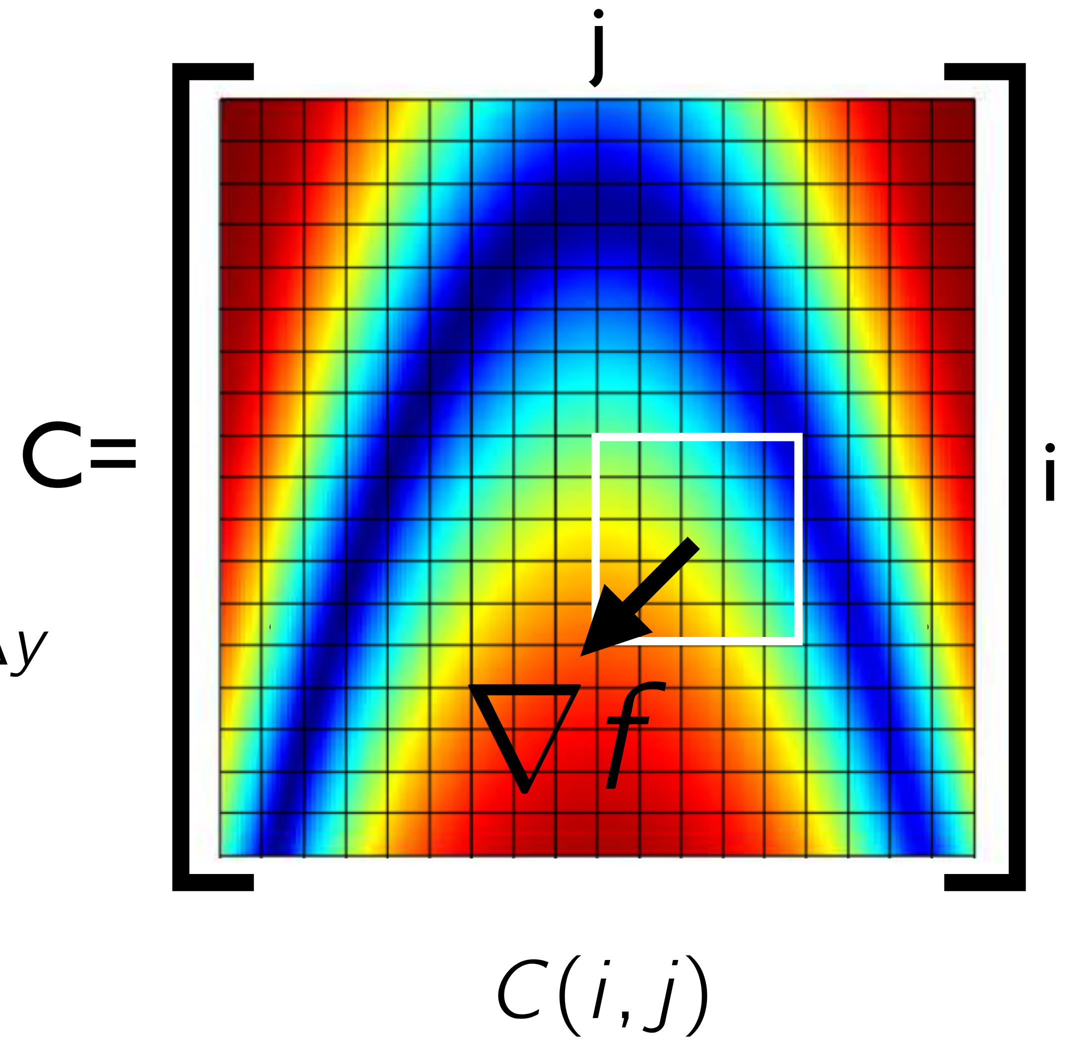


Calculating the gradient of a function on a discrete grid

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







$\Delta x = \Delta y = 1$

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

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

Overview

Four types of functions:

1. **A scalar function of a single scalar variable**, i.e., $f(x) \in \mathbb{R}$ and $x \in \mathbb{R}$.
2. **A scalar function of multiple scalar variables (or a scalar function of a vector variable)**. In this case, $f(\mathbf{x}) \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$.
3. **A vector function of a single scalar variable**, i.e., $\mathbf{f}(x) \in \mathbb{R}^n$ and $x \in \mathbb{R}$.
4. **A vector function of a vector variable**, i.e., $\mathbf{f}(\mathbf{x}) \in \mathbb{R}^n$ and $\mathbf{x} \in \mathbb{R}^n$.

One good reason for us to calculate analytical derivatives:

Summary: Numerical-optimization methods produce faster and more precise results when we provide them with the analytical derivative (or gradient, or Jacobian) of the objective function.

Calculating the gradient of a function on a discrete grid

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

$\Delta x = \Delta y = 1$