

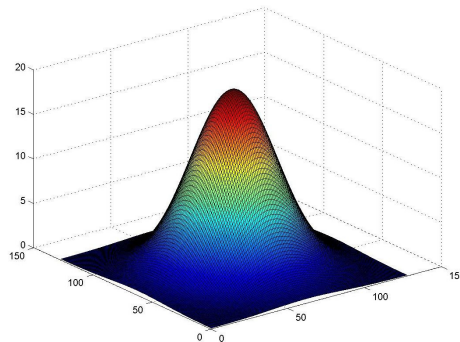
Add Taylor and Basic optimization idea at the end. Make Optimization (Lin+NonLin) evtl. a section

9.2 Multivariate Calculus Basics

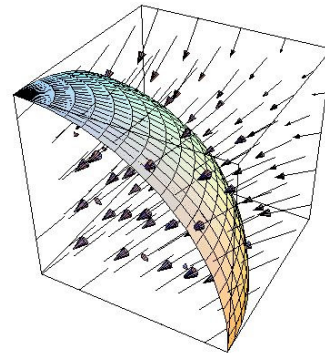
In this section we briefly introduce the basics of the multivariate calculus, i.e. function which take more than one variables as input. More specifically, we will deal with the different forms of the derivatives of multivariate functions. This will provide the basis for most of the methods presented. We will need it for parameter estimation, algorithm analysis, error propagation, reconstruction and optimization methods. Furthermore, multivariate calculus is extremely important in many other applications.

The section will be divided in two parts. The first one will consider analysis of *multivariate real-valued functions*, i.e. functions which take more parameters and give one single scalar value as the result. These functions are also known as scalar-fields. Here we will present concepts which can be applied to multivariate real-valued functions: the *partial derivative*, the *gradient* and the *Hessian* of the function. The second part is about *multivariate vector-valued functions*, also known as vector-fields. These functions return vectors as the result. Here we will present the concept of the *Jacobian*.

The topics are presented here very briefly. For more details you can - and we encourage you to - refer to [4] and [1] (both in german) for more details.



(a) Image of a scalar-field



(b) Image of a vector-field,

www.math.armstrong.edu/mmacalc/animations/

9.2.1 Multivariate Real-Valued Functions

In the first part we will consider real-valued multivariate functions. These are, as described above, functions that take several arguments as input (thus multivariate), and return a single scalar as the result. More formally, we will describe such a function f as $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Multivariate real-valued functions are also denoted as *scalar fields*.

In the following we will assume a multivariate real-valued function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, or more specifically $f(x_1, x_2, \dots, x_n) = y$. Moreover we assume that the function f is sufficiently *nice* so that it fulfills all the necessary properties required for anything we want to do with it.

We will now introduce the three most important¹ derivative concepts for scalar fields. These are the *partial derivative*, the *gradient* and the *Hessian*.

¹at least in our context

9.2.1.1 Partial Derivative

Definition 9.1 (Partial Derivative)

- The result of applying the partial derivative with respect to x_i (the i -th input variable of f) is the derivative of the function f where we assume that all other variables x_j with $i \neq j$ are constant. The derivatives with respect to these variables thus vanish.
- After we obtain the formula for the partial derivative w.r.t. x_i has to be evaluated for a certain value $a \in \mathbb{R}^n$:

$$\frac{\delta f}{\delta x_i}(a) =: \delta_{x_i} f(a)$$

Example 9.2 Given the function $f(x_1, x_2, x_3) = x_1^4 + x_1 x_2 x_3 + 2x_2 x_3^2$ we can compute the following partial derivatives:

$$\delta_{x_1} f(x) = 4x_1^3 + x_2 x_3 \quad (9.1)$$

$$\delta_{x_2} f(x) = x_1 x_3 + 2x_3^2 \quad (9.2)$$

$$\delta_{x_3} f(x) = x_1 x_2 + 4x_2 x_3 \quad (9.3)$$

Now we could evaluate the three partial derivatives for any $x = (x_1, x_2, x_3)^T$.

The most important properties of the partial derivative are summarized as follows:

- Partial derivatives give the derivatives along the main coordinate system axes.
- Higher partial derivatives are also possible:

$$\delta_{x_1 x_2} f(a) = \delta_{x_2} \delta_{x_1} f(a) = \frac{\delta}{\delta x_2} \left(\frac{\delta f}{\delta x_1}(x) \right)$$

- The operation of the partial derivative can be interpreted as an operator F taking the function $f \in L(\mathbb{R}^n \rightarrow \mathbb{R})$ as an argument and producing the function $\delta_i f(x) \in L(\mathbb{R}^n \rightarrow \mathbb{R})$. Here $L(\mathbb{R}^n \rightarrow \mathbb{R})$ denotes the space of multivariate real-valued functions. This way of thinking about the partial derivative is not necessary, it is sort of interesting and gives further insight.

9.2.1.2 Gradient

Definition 9.3 (Gradient) Taking all partial derivatives $\delta_{x_i} f(x)$ of a function and putting them together in a vector gives a thing called a gradient:

$$\text{grad } f(x) = \begin{bmatrix} \delta_{x_1} f(x) \\ \delta_{x_2} f(x) \\ \vdots \\ \delta_{x_n} f(x) \end{bmatrix}$$

Like the partial derivative, the gradient is a function that can be evaluated for all $x \in \mathbb{R}^n$.

Example 9.4 For the function $f(x_1, x_2, x_3) = x_1^4 + x_1x_2x_3 + 2x_2x_3^2$ we already computed the partial derivatives in the example above. Now we use them to compose the gradient.

$$\text{grad } f(x) = \begin{bmatrix} \delta_{x_1} f(x) \\ \delta_{x_2} f(x) \\ \delta_{x_3} f(x) \end{bmatrix} = \begin{bmatrix} 4x_1^3 + x_2x_3 \\ x_1x_3 + 2x_3^2 \\ x_1x_2 + 4x_2x_3 \end{bmatrix}$$

We can now evaluate the gradient at any point a where the function f is defined. Let's use $a = (1, 2, 3)^T$ for the above example. This gives

$$\text{grad } f(a) = \begin{bmatrix} \delta_{x_1} f(a) \\ \delta_{x_2} f(a) \\ \delta_{x_3} f(a) \end{bmatrix} = \begin{bmatrix} 4 \cdot 1^3 + 2 \cdot 3 \\ 1 \cdot 3 + 2 \cdot 3^2 \\ 1 \cdot 2 + 4 \cdot 2 \cdot 3 \end{bmatrix} = \begin{bmatrix} 10 \\ 21 \\ 26 \end{bmatrix}$$

Here come some properties of the gradient:

- The gradient $\text{grad } f(x)$ shows for every point x in the direction of the steepest slope of the function f at the point x .
- The magnitude of the gradient $\text{grad } f(x)$ gives the "steepness" of the function f at the point x .
- The gradient can be interpreted as the first derivative of a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$.
- The gradient can be interpreted as an operator F taking the function $f \in L(\mathbb{R}^n \rightarrow \mathbb{R})$ and producing the function $\text{grad } f(x) \in L(\mathbb{R}^n \rightarrow \mathbb{R})$.

9.2.1.3 Directional Derivative

The *directional derivative* is the generalization of the concept of the partial derivative. We have already seen that the partial derivative gives the derivative along the main coordinate system axes x_1, x_2, \dots, x_n . Now we want to have a derivative along an arbitrary vector v .

Definition 9.5 (Directional Derivative) Given a vector v with $\|v\|_2 = 1$. Then the derivative along v is

$$\delta_v f(x) = \text{grad } f(x) \cdot v = \sum_{i=1}^n f_{x_i}(x) v_i$$

The directional derivative can be interpreted as the projection of the gradient $\text{grad } f(x)$ onto the vector v

9.2.1.4 Application: Linear Approximation of Real-Valued Functions

Given an arbitrary real-valued function $f : \mathbb{R}^n \rightarrow \mathbb{R}$. In case that the function evaluation is complicated we can approximate the function in the neighborhood of x_0 using the Taylor expansion. For that we can use either a simple linear approximation or higher-order approximations, e.g. the quadratic approximation.

Given the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and the point x_0 we can approximate f linearly in the neighborhood of x_0 by using the gradient:

$$f(x) \approx f(x_0) + \text{grad } f(x_0) \cdot (x - x_0)$$

If the whole stuff here is unknown, so is probably Taylor expansion as well... introduce maybe

9.2.1.5 Hessian

Know that we know that the gradient is first derivative of a scalar-field, how about higher order derivatives? We will see that this concept is encapsulated in the so called *Hessian* (matrix) of the function

Definition 9.6 (Hessian) For a real-valued function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ we can define a $n \times n$ matrix containing second partial derivatives. It shall be called the *Hessian of the function* f and denoted

$$Hess_f(x) = \begin{bmatrix} \delta_{x_1x_1}f(x) & \delta_{x_1x_2}f(x) & \dots & \delta_{x_1x_n}f(x) \\ \delta_{x_2x_1}f(x) & \delta_{x_2x_2}f(x) & \dots & \delta_{x_2x_n}f(x) \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{x_nx_1}f(x) & \delta_{x_nx_2}f(x) & \dots & \delta_{x_nx_n}f(x) \end{bmatrix}$$

Properties of the Hessian:

- The Hessian is a symmetric matrix.
- The Hessian can be interpreted as the second derivative for real-valued functions.

9.2.1.6 Application: Approximation by Taylor Expansion

As we have shown before, we can use the gradient for the linear approximation of a non-linear function f :

$$f(x) \approx f(x_0) + grad f(x_0) \cdot (x - x_0)$$

With the Hessian we can achieve quadratic approximation:

$$f(x) \approx f(x_0) + grad f(x_0) \cdot (x - x_0) + \frac{1}{2}(x - x_0)^T Hess_f(x_0)(x - x_0)$$

9.2.2 Multivariate Vector-Valued Functions

Now we will discuss multivariate vector-valued functions. These, in contrast to real-valued functions, return a vector as the result. More formally, f is $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$. A synonym for vector-valued functions is *vector fields*.

In this part we will introduce the so called *Jacobian* of a function. We will see that for vector-valued functions the Jacobian corresponds to the concept of the first derivative.

In the following we will assume a multivariate vector-valued function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, or more specifically $f(x) = y$ with $x = (x_1, x_2, \dots, x_n)^T$ and $y = (y_1, y_2, \dots, y_m)^T$. Again we assume that f is sufficiently nice.

9.2.2.1 Jacobian

In order to intuitively explain the Jacobian of a vector-valued function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $f(x) = y$, we will first note that we can think about the function f as m real-vector functions $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$, s.t. $f_1(x) = y_1, \dots, f_m(x) = y_m$. Now, in order to compute the first derivative of f , we compute the first derivatives of the real-valued functions f_i , i.e. their gradients. Stacked on top of each other as rows in the matrix they give the first derivative of the function f .

Definition 9.7 (Jacobian) *The Jacobian $J_f(x)$ of the function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is the generalization of the gradient for vector-valued functions.*

$$J_f(x) = \begin{bmatrix} \text{grad } f_1(x) \\ \text{grad } f_2(x) \\ \vdots \\ \text{grad } f_m(x) \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & \dots & \frac{\partial f_1}{\partial x_n}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) & \dots & \frac{\partial f_2}{\partial x_n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \frac{\partial f_m}{\partial x_2}(x) & \dots & \frac{\partial f_m}{\partial x_n}(x) \end{bmatrix}$$

The Jacobian $J_f(x)$ for $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a $m \times n$ matrix.

Please notice that the Jacobian, same as the gradient and the Hessian for real-valued functions, is itself a function that must be evaluated at a certain point.

Properties of the Jacobian

- The Jacobian $J_f(x)$ of the function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ can be interpreted as the first derivative of the vector-valued function f . Thus $J_f(x)$ can be used for the linear approximation of f .
- The operation of building the Jacobian can be interpreted as an operator J taking the function $f \in L(\mathbb{R}^n \rightarrow \mathbb{R}^m)$ and producing a linear, vector-valued function $J_f \in L(\mathbb{R}^n \rightarrow \mathbb{R}^m)$.

9.2.2.2 Application: Linear Approximation

- The Jacobian $J(x)$ of the function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ can be interpreted as the linear approximation of that function.
- This can be used for local approximation around some value x_0 .

- Hence in the neighbourhood of x_0 we can set:

$$f(x) \approx \tilde{f}(x) = f(x_0) + J(x_0) \cdot (x - x_0)$$