54

Vector-Valued Functions of Several Real Variables

> Auch die Chemiker müssen sich allmählich an den Gedanken gewöhnen, dass ihnen die theoretische Chemie ohne die Beherrschung der Elemente der höheren Analysis ein Buch mit sieben Siegeln blieben wird. Ein Differential- oder Integralzeichen muss aufhören, für den Chemiker eine unverständliche Hieroglyphe zu sein, ... wenn er sich nicht der Gefahr aussetzen will, für die Entwicklung der theoretischen Chemie jedes Verständnis zu verlieren. (H. Jahn, Grundriss der Elektrochemie, 1895)

54.1 Introduction

We now turn to the extension of the basic concepts of real-valued functions of one real variable, such as Lipschitz continuity and differentiability, to vector-valued functions of several variables. We have carefully prepared the material so that this extension will be as natural and smooth as possible. We shall see that the proofs of the basic theorems like the Chain rule, the Mean Value theorem, Taylor's theorem, the Contraction Mapping theorem and the Inverse Function theorem, extend almost word by word to the more complicated situation of vector valued functions of several real variables.

We consider functions $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ that are vector valued in the sense that the value $f(x) = (f_1(x), \ldots, f_m(x))$ is a vector in $\mathbb{R}^m$ with components $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ for $i = 1, \ldots, m$, where with $f_i(x) = f_i(x_1, \ldots, x_n)$ and $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. As usual, we view $x = (x_1, \ldots, x_n)$ as a n-column vector and $f(x) = (f_1(x), \ldots, f_m(x))$ as a m-column vector.

As particular examples of vector-valued functions, we first consider *curves*, which are functions $g : \mathbb{R} \to \mathbb{R}^n$, and *surfaces*, which are functions $g : \mathbb{R}^2 \to \mathbb{R}^n$. We then discuss composite functions $f \circ g : \mathbb{R} \to \mathbb{R}^m$, where $g : \mathbb{R} \to \mathbb{R}^n$ is a curve and $f : \mathbb{R}^n \to \mathbb{R}^m$, with $f \circ g$ again being a curve. We recall that $f \circ g(t) = f(g(t))$.

The inputs to the functions reside in the n dimensional vector space $\mathbb{R}^n$ and it is worthwhile to consider the properties of $\mathbb{R}^n$. Of particular importance is the notion of Cauchy sequence and convergence for sequences $\{x^{(j)}\}_{j=1}^{\infty}$ of vectors $x^{(j)} = (x_1^{(j)}, \ldots, x_n^{(j)}) \in \mathbb{R}^n$ with coordinates $x_k^{(j)}$, $k = 1, \ldots, n$. We say that the sequence $\{x^{(j)}\}_{j=1}^{\infty}$ is a *Cauchy sequence* if for all $\epsilon > 0$, there is a natural number N so that

$$\|x^{(i)} - x^{(j)}\| \leq \epsilon \quad \text{for } i, j > N.$$

Here $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^n$, that is, $\|x\| = (\sum_{i=1}^n x_i^2)^{1/2}$. Sometimes, it is convenient to work with the norms $\|x\|_1 = \sum_{i=1}^n |x_i|$ or $\|x\|_\infty = \max_{i=1,\ldots,n} |x_i|$. We say that the sequence $\{x^{(j)}\}_{j=1}^{\infty}$ of vectors in $\mathbb{R}^n$ *converges* to $x \in \mathbb{R}^n$ if for all $\epsilon > 0$, there is a natural number N so that

$$\|x - x^{(i)}\| \leq \epsilon \quad \text{for } i > N.$$

It is easy to show that a convergent sequence is a Cauchy sequence and conversely that a Cauchy sequence converges. We obtain these results applying the corresponding results for sequences in $\mathbb{R}$ to each of the coordinates of the vectors in $\mathbb{R}^n$.

Example 54.1. The sequence $\{x^{(i)}\}_{i=1}^{\infty}$ in $\mathbb{R}^2$, $x^{(i)} = (1 - i^{-2}, \exp(-i))$, converges to $(1, 0)$.

54.2 Curves in $\mathbb{R}^n$

A function $g : I \to \mathbb{R}^n$, where $I = [a, b]$ is an interval of real numbers, is a *curve* in $\mathbb{R}^n$, see Fig. 54.1. If we use t as the independent variable ranging over I, then we say that the curve $g(t)$ is *parametrized* by the variable t. We also refer to the set of points $\Gamma = \{g(t) \in \mathbb{R}^n : t \in I\}$ as the curve Γ parameterized by the function $g : I \to \mathbb{R}^n$.

Example 54.2. The simplest example of a curve is a straight line. The function $g : \mathbb{R} \to \mathbb{R}^2$ given by

$$g(t) = \bar{x} + tz,$$

where $z \in \mathbb{R}^2$ and $\bar{x} \in \mathbb{R}^2$, is a straight line in $\mathbb{R}^2$ through the point $\bar{x}$ with direction z, see Fig. 54.2.

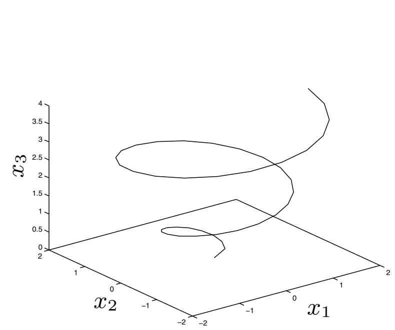

Fig. 54.1. The curve $g : [0, 4] \to \mathbb{R}^3$ with $g(t) = \big(t^{1/2}\cos(\pi t),\, t^{1/2}\sin(\pi t), t\big)$

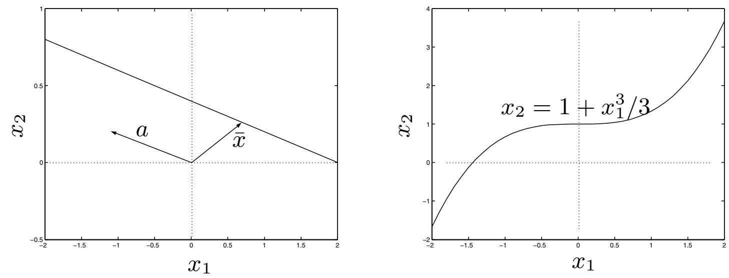

Fig. 54.2. On the *left*: the curve $g(t) = \bar{x} + ta$. On the *right*: a curve $g(t) = (t, f(t))$

Example 54.3. Let $f : [a, b] \to \mathbb{R}$ be given, and define $g : [a, b] \to \mathbb{R}^2$ by $g(t) = (g_1(t), g_2(t)) = (t, f(t))$. This curve is simply the graph of the function $f : [a, b] \to \mathbb{R}$, see Fig. 54.2.

54.3 Different Parameterizations of a Curve

It is possible to use different parametrizations for the set of points forming a curve. If $h : [c, d] \to [a, b]$ is a one-to-one mapping, then the composite function $f = g \circ h : [c, d] \to \mathbb{R}^2$ is a *reparameterization* of the curve $\{g(t) : t \in [a, b]\}$ given by $g : [a, b] \to \mathbb{R}^2$.

Example 54.4. The function $f : [0, \infty) \to \mathbb{R}^3$ given by

$$f(\tau) = (\tau\cos(\pi\tau^2), \tau\sin(\pi\tau^2), \tau^2),$$

is a reparameterization of the curve $g : [0, \infty) \to \mathbb{R}^3$ given by

$$g(t) = (\sqrt{t}\cos(\pi t), \sqrt{t}\sin(\pi t), t),$$

obtained setting $t = h(\tau) = \tau^2$. We have $f = g \circ h$.

54.4 Surfaces in $\mathbb{R}^n$, $n \geq 3$

A function $g : Q \to \mathbb{R}^n$, where $n \geq 3$ and Q is a subdomain of $\mathbb{R}^2$, may be viewed to be a *surface* S in $\mathbb{R}^n$, see Fig. 54.3. We write $g = g(y)$ with $y = (y_1, y_2) \in Q$ and say that S is parameterized by $y \in Q$. We may also identify the surface S with the set of points $S = \{g(y) \in \mathbb{R}^n : y \in Q\}$, and reparameterize S by $f = g \circ h : \tilde{Q} \to \mathbb{R}^n$ if $h : \tilde{Q} \to Q$ is a one-to-one mapping of a domain $\tilde{Q}$ in $\mathbb{R}^2$ onto Q.

Example 54.5. The simplest example of a surface $g : \mathbb{R}^2 \to \mathbb{R}^3$ is a plane in $\mathbb{R}^3$ given by

$$g(y) = g(y_1, y_2) = \bar{x} + y_1 b_1 + y_2 b_2, \quad y \in \mathbb{R}^2,$$

where $\bar{x}$, b_1, $b_2 \in \mathbb{R}^3$.

Example 54.6. Let $f : [0,1] \times [0,1] \to \mathbb{R}$ be given, and define $g : [0,1] \times [0,1] \to \mathbb{R}^3$ by $g(y_1, y_2) = (y_1, y_2, f(y_1, y_2))$. This is a surface, which is the graph of $f : [0,1] \times [0,1] \to \mathbb{R}$. We also refer to this surface briefly as the surface given by the function $x_3 = f(x_1, x_2)$ with $(x_1, x_2) \in [0,1] \times [0,1]$.

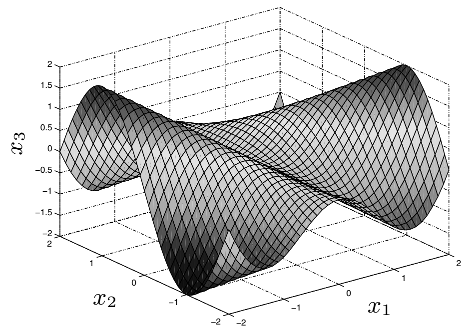

Fig. 54.3. The surface $s(y_1, y_2) = \big(y_1, y_2, y_1 \sin\big((y_1 + y_2)\pi/2\big)\big)$ with $-1 \leq y_1, y_2 \leq 1$, or briefly the surface $x_3 = x_1 \sin\big((x_1 + x_2)\pi/2\big)$ with $-1 \leq x_1, x_2 \leq 1$

54.5 Lipschitz Continuity

We say that $f : \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz continuous on $\mathbb{R}^n$ if there is a constant L such that

$$\|f(x) - f(y)\| \leq L\|x - y\| \quad \text{for all } x, y \in \mathbb{R}^n. \qquad (54.1)$$

This definition extends easily to functions $f : A \to \mathbb{R}^m$ with the domain $D(f) = A$ being a subset of $\mathbb{R}^n$. For example, A may be the unit n-cube $[0,1]^n = \{x \in \mathbb{R}^n : 0 \leq x_i \leq 1, i = 1, \ldots, n\}$ or the unit n-disc $\{x \in \mathbb{R}^n : \|x\| \leq 1\}$.

To check if a function $f : A \to \mathbb{R}^m$ is Lipschitz continuous on some subset A of $\mathbb{R}^n$, it suffices to check that the component functions $f_i : A \to \mathbb{R}$ are Lipschitz continuous. This is because

$$|f_i(x) - f_i(y)| \leq L_i\|x - y\| \quad \text{for } i = 1, \ldots, m,$$

implies

$$\|f(x) - f(y)\|^2 = \sum_{i=1}^{m} |f_i(x) - f_i(y)|^2 \leq \sum_{i=1}^{m} L_i^2\|x - y\|^2,$$

which shows that $\|f(x) - f(y)\| \leq L\|x - y\|$ with $L = (\sum_i L_i^2)^{\frac{1}{2}}$.

Example 54.7. The function $f : [0,1] \times [0,1] \to \mathbb{R}^2$ defined by $f(x_1, x_2) = (x_1 + x_2, x_1x_2)$, is Lipschitz continuous with Lipschitz constant $L = 2$. To show this, we note that $f_1(x_1, x_2) = x_1 + x_2$ is Lipschitz continuous on $[0,1] \times [0,1]$ with Lipschitz constant $L_1 = \sqrt{2}$ because $|f_1(x_1, x_2) - f_1(y_1, y_2)| \leq |x_1 - y_1| + |x_2 - y_2| \leq \sqrt{2}\|x - y\|$ by Cauchy's inequality. Similarly, $f_2(x_1, x_2) = x_1x_2$ is Lipschitz continuous on $[0,1] \times [0,1]$ with Lipschitz constant $L_2 = \sqrt{2}$ since $|x_1x_2 - y_1y_2| = |x_1x_2 - y_1x_2 + y_1x_2 - y_1y_2| \leq |x_1 - y_1| + |x_2 - y_2| \leq \sqrt{2}\|x - y\|$.

Example 54.8. The function $f : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$f(x_1, \ldots, x_n) = (x_n, x_{n-1}, \ldots, x_1),$$

is Lipschitz continuous with Lipschitz constant $L = 1$.

Example 54.9. A linear transformation $f : \mathbb{R}^n \to \mathbb{R}^m$ given by an $m \times n$ matrix $A = (a_{ij})$, with $f(x) = Ax$ and x a n-column vector, is Lipschitz continuous with Lipschitz constant $L = \|A\|$. We made this observation in Chapter *Analytic geometry in* $\mathbb{R}^n$. We repeat the argument:

$$\begin{aligned} L = \max_{x \neq y} \frac{\|f(x) - f(y)\|}{\|x - y\|} &= \max_{x \neq y} \frac{\|Ax - Ay\|}{\|x - y\|} \\ &= \max_{x \neq y} \frac{\|A(x - y)\|}{\|x - y\|} = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \|A\|. \end{aligned}$$

Concerning the definition of the matrix norm $\|A\|$, we note that the function $F(x) = \|Ax\|/\|x\|$ is homogeneous of degree zero, that is, $F(\lambda x) = F(x)$ for all non-zero real numbers λ, and thus $\|A\|$ is the maximum value of $F(x)$ on the closed and bounded set $\{x \in \mathbb{R}^n : \|x\| = 1\}$, which is a finite real number.

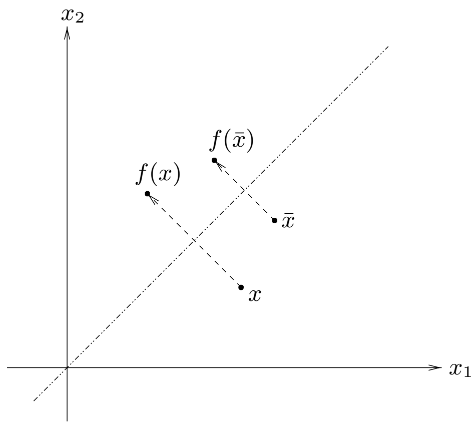

Fig. 54.4. Illustration of the mapping $f(x_1, x_2) = (x_2, x_1)$, which is clearly Lipschitz continuous with $L = 1$

We recall that if A is a diagonal $n \times n$ matrix with diagonal elements λ_i, then $\|A\| = \max_i |\lambda_i|$.

54.6 Differentiability: Jacobian, Gradient and Tangent

We say that $f : \mathbb{R}^n \to \mathbb{R}^m$ is *differentiable at* $\bar{x} \in \mathbb{R}^n$ if there is a $m \times n$ matrix $M(\bar{x}) = (m_{ij}(\bar{x}))$, called the *Jacobian* of the function $f(x)$ at $\bar{x}$, and a constant $K_f(\bar{x})$ such that for all x close to $\bar{x}$,

$$f(x) = f(\bar{x}) + M(\bar{x})(x - \bar{x}) + E_f(x, \bar{x}), \tag{54.2}$$

where $E_f(x, \bar{x}) = (E_f(x, \bar{x})_i)$ is an m-vector satisfying $\|E_f(x, \bar{x})\| \leq K_f(\bar{x})\|x - \bar{x}\|^2$. We also denote the Jacobian by $Df(\bar{x})$ or $f'(\bar{x})$ so that $M(\bar{x}) = Df(\bar{x}) = f'(\bar{x})$. Since $f(x)$ is a m-column vector, or $m \times 1$ matrix, and x is a n-column vector, or $n \times 1$ matrix, $M(\bar{x})(x - \bar{x})$ is the product of the $m \times n$ matrix $M(\bar{x})$ and the $n \times 1$ matrix $x - \bar{x}$ yielding a $m \times 1$ matrix or a m-column vector.

We say that $f : A \to \mathbb{R}^m$, where A is a subset of $\mathbb{R}^n$, is *differentiable on* A if $f(x)$ is differentiable at $\bar{x}$ for all $\bar{x} \in A$. We say that $f : A \to \mathbb{R}^m$ is *uniformly differentiable on* A if the constant $K_f(\bar{x}) = K_f$ can be chosen independently of $\bar{x} \in A$.

We now show how to determine a specific element $m_{ij}(\bar{x})$ of the Jacobian using the relation (54.2). We consider the coordinate function $f_i(x_1, \ldots, x_n)$ and setting $x = \bar{x} + se_j$, where e_j is the j^{th} standard basis vector and s is a small real number, we focus on the variation of $f_i(x_1, \ldots, x_n)$ as the

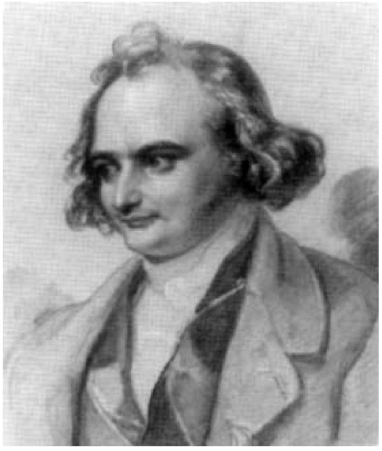

Fig. 54.5. Carl Jacobi (1804–51): "It is often more convenient to possess the ashes of great men than to possess the men themselves during their lifetime" (on the return of Descarte's remains to France)

variable x_j varies in a neighborhood of $\bar{x}_j$. The relation (54.2) states that for small non-zero real numbers s,

$$f_i(\bar{x} + se_j) = f_i(\bar{x}) + m_{ij}(\bar{x})s + E_f(\bar{x} + se_j, \bar{x})_i, \qquad (54.3)$$

where $\|x - \bar{x}\|^2 = \|se_j\|^2 = s^2$ implies

$$|E_f(\bar{x} + se_j, \bar{x})_i| \leq K_f(\bar{x})s^2.$$

Note that by assumption $\|E_f(x, \bar{x})\| \leq K_f(\bar{x})\|x - \bar{x}\|^2$, and so each coordinate function $E_f(\bar{x} + se_j, \bar{x})_i$ satisfies $|E_f(x, \bar{x})_i| \leq K_f(\bar{x})\|x - \bar{x}\|^2$.

Now, dividing by s in (54.3) and letting s tend to zero, we find that

$$m_{ij}(\bar{x}) = \lim_{s \to 0} \frac{f_i(\bar{x} + se_j) - f_i(\bar{x})}{s}, \qquad (54.4)$$

which we can also write as

$$m_{ij}(\bar{x}) = \qquad (54.5)$$
$$\lim_{x_j \to \bar{x}_j} \frac{f_i(\bar{x}_1, \ldots, \bar{x}_{j-1}, x_j, \bar{x}_{j+1}, \ldots, \bar{x}_n) - f_i(\bar{x}_1, \ldots, \bar{x}_{j-1}, \bar{x}_j, \bar{x}_{j+1}, \ldots, \bar{x}_n)}{x_j - \bar{x}_j}.$$

We refer to $m_{ij}(\bar{x})$ as the *partial derivative* of f_i with respect to x_j at $\bar{x}$, and we use the alternative notation $m_{ij}(\bar{x}) = \frac{\partial f_i}{\partial x_j}(\bar{x})$. To compute $\frac{\partial f_i}{\partial x_j}(\bar{x})$ we freeze all coordinates at $\bar{x}$ but the coordinate x_j and then let x_j vary in a neighborhood of $\bar{x}_j$. The formula

$$\frac{\partial f_i}{\partial x_j}(\bar{x}) = \tag{54.6}$$
$$\lim_{x_j \to \bar{x}_j} \frac{f_i(\bar{x}_1, \dots, \bar{x}_{j-1}, x_j, \bar{x}_{j+1}, \dots, \bar{x}_n) - f_i(\bar{x}_1, \dots, \bar{x}_{j-1}, \bar{x}_j, \bar{x}_{j+1}, \dots, \bar{x}_n)}{x_j - \bar{x}_j},$$

states that we compute the partial derivative with respect to the variable x_j by keeping all the other variables $x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n$ constant. Thus, computing partial derivatives should be a pleasure using our previous expertise of computing derivatives of functions of one real variable!

We may express the computation alternatively as follows:

$$\frac{\partial f_i}{\partial x_j}(\bar{x}) = m_{ij}(\bar{x}) = g'_{ij}(0) = \frac{dg_{ij}}{ds}(0), \tag{54.7}$$

where $g_{ij}(s) = f_i(\bar{x} + se_j)$.

Example 54.10. Let $f : \mathbb{R}^3 \to \mathbb{R}$ be given by $f(x_1, x_2, x_3) = x_1 e^{x_2} \sin(x_3)$. We compute

$$\frac{\partial f}{\partial x_1}(\bar{x}) = e^{\bar{x}_2} \sin(\bar{x}_3), \quad \frac{\partial f}{\partial x_2}(\bar{x}) = \bar{x}_1 e^{\bar{x}_2} \sin(\bar{x}_3),$$
$$\frac{\partial f}{\partial x_3}(\bar{x}) = \bar{x}_1 e^{\bar{x}_2} \cos(\bar{x}_3),$$

and thus

$$f'(\bar{x}) = (e^{\bar{x}_2} \sin(\bar{x}_3), \bar{x}_1 e^{\bar{x}_2} \sin(\bar{x}_3), \bar{x}_1 e^{\bar{x}_2} \cos(\bar{x}_3))$$

Example 54.11. If $f : \mathbb{R}^3 \to \mathbb{R}^2$ is given by $f(x) = \begin{pmatrix} \exp(x_1^2 + x_2^2) \\ \sin(x_2 + 2x_3) \end{pmatrix}$, then

$$f'(x) = \begin{pmatrix} 2x_1 \exp(x_1^2 + x_2^2) & 2x_2 \exp(x_1^2 + x_2^2) & 0 \\ 0 & \cos(x_2 + 2x_3) & 2\cos(x_2 + 2x_3) \end{pmatrix}.$$

We have now shown how to compute the elements of a Jacobian using the usual rules for differentiation with respect to one real variable. This opens a whole new world of applications to explore. The setting is thus a differentiable function $f : \mathbb{R}^n \to \mathbb{R}^m$ satisfying for suitable $x, \bar{x} \in \mathbb{R}^n$:

$$f(x) = f(\bar{x}) + f'(\bar{x})(x - \bar{x}) + E_f(x, \bar{x}), \tag{54.8}$$

with $\|E_f(x, \bar{x})\| \le K_f(\bar{x}) \|x - \bar{x}\|^2$, where $f'(\bar{x}) = Df(\bar{x})$ is the Jacobian $m \times n$ matrix with elements $\frac{\partial f_i}{\partial x_j}$:

$$f'(\bar{x}) = Df(\bar{x}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\bar{x}) & \frac{\partial f_1}{\partial x_2}(\bar{x}) & \dots & \frac{\partial f_1}{\partial x_n}(\bar{x}) \\ \frac{\partial f_2}{\partial x_1}(\bar{x}) & \frac{\partial f_2}{\partial x_2}(\bar{x}) & \dots & \frac{\partial f_2}{\partial x_n}(\bar{x}) \\ \dots & \dots & \dots & \\ \frac{\partial f_m}{\partial x_1}(\bar{x}) & \frac{\partial f_m}{\partial x_2}(\bar{x}) & \dots & \frac{\partial f_m}{\partial x_n}(\bar{x}) \end{pmatrix}.$$

Sometimes we use the following notation for the Jacobian $f'(x)$ of a function $y = f(x)$ with $f : \mathbb{R}^n \to \mathbb{R}^m$:

$$f'(x) = \frac{dy_1, \dots, dy_m}{dx_1, \dots, dx_n}(x) \tag{54.9}$$

The function $x \to \hat{f}(x) = f(\bar{x}) + f'(\bar{x})(x - \bar{x})$ is called the *linearization* of the function $x \to f(x)$ at $x = \bar{x}$. We have

$$\hat{f}(x) = f'(\bar{x})x + f(\bar{x}) - f'(\bar{x})\bar{x} = Ax + b,$$

with $A = f'(\bar{x})$ a $m\times n$ matrix and $b = f(\bar{x}) - f'(\bar{x})\bar{x}$ a m-column vector. We say that $\hat{f}(x)$ is an *affine transformation*, which is a transformation of the form $x \to Ax + b$, where x is a n-column vector, A is a $m \times n$ matrix and b is a m-column vector. The Jacobian $\hat{f}'(x)$ of the linearization $\hat{f}(x) = Ax + b$ is a constant matrix equal to the matrix A, because the partial derivatives of Ax with respect to x are simply the elements of the matrix A.

If $f : \mathbb{R}^n \to \mathbb{R}$, that is $m = 1$, then we also denote the Jacobian f' by ∇f, that is,

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

In words, $\nabla f(\bar{x})$ is the n-row vector or $1 \times n$ matrix of partial derivatives of $f(x)$ with respect to $x_1, x_2, \dots, x_n$ at $\bar{x}$. We refer to $\nabla f(\bar{x})$ as the *gradient* of $f(x)$ at $\bar{x}$. If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\bar{x}$, we thus have

$$f(x) = f(\bar{x}) + \nabla f(\bar{x})(x - \bar{x}) + E_f(x, \bar{x}), \tag{54.10}$$

with $|E_f(x, \bar{x})| \le K_f(\bar{x})\|x - \bar{x}\|^2$, and $\hat{f}(x) = f(\bar{x}) + \nabla f(\bar{x})(x - \bar{x})$ is the linearization of $f(x)$ at $x = \bar{x}$. We may alternatively express the product $\nabla f(\bar{x})(x - \bar{x})$ of the n-row vector ($1 \times n$ matrix) $\nabla f(\bar{x})$ with the n-column vector ($n \times 1$ matrix) $(x - \bar{x})$ as the scalar product $\nabla f(\bar{x}) \cdot (x - \bar{x})$ of the n-vector $\nabla f(\bar{x})$ with the n-vector $(x - \bar{x})$. We thus often write (54.10) in the form

$$f(x) = f(\bar{x}) + \nabla f(\bar{x}) \cdot (x - \bar{x}) + E_f(x, \bar{x}). \tag{54.11}$$

Example 54.12. If $f : \mathbb{R}^3 \to \mathbb{R}$ is given by $f(x) = x_1^2 + 2x_2^3 + 3x_3^4$, then

$$\nabla f(x) = (2x_1, 6x_2^2, 12x_3^3).$$

Example 54.13. The equation $x_3 = f(x)$ with $f : \mathbb{R}^2 \to \mathbb{R}$ and $x = (x_1, x_2)$ represents a surface in $\mathbb{R}^3$ (the graph of the function f). The linearization

$$\begin{aligned} x_3 &= f(\bar{x}) + \nabla f(\bar{x}) \cdot (x - \bar{x}) \\ &= f(\bar{x}) + \frac{\partial f}{\partial x_1}(\bar{x})(x_1 - \bar{x}_1) + \frac{\partial f}{\partial x_2}(\bar{x})(x_2 - \bar{x}_2) \end{aligned}$$

with $\bar{x} = (\bar{x}_1, \bar{x}_2)$, represents the *tangent plane* at $x = \bar{x}$, see Fig. 54.6.

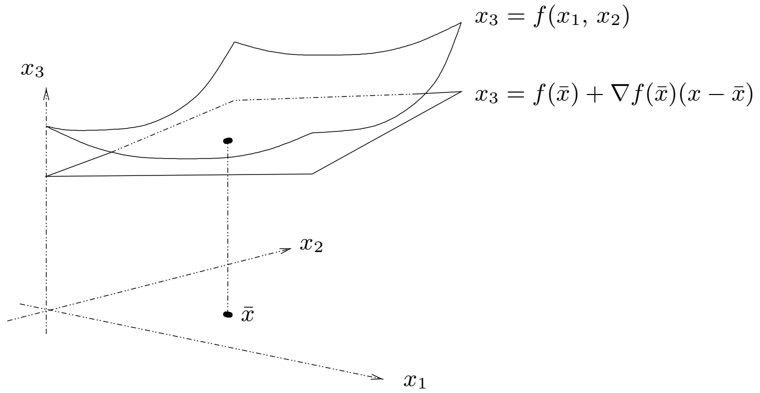

Fig. 54.6. The surface $x_3 = f(x_1, x_2)$ and its tangent plane

Example 54.14. Consider now a curve $f : \mathbb{R} \to \mathbb{R}^m$, that is, $f(t) = (f_1(t), \ldots, f_m(t))$ with $t \in \mathbb{R}$ and we have a situation with $n = 1$. The linearization $t \to \hat{f}(t) = f(\bar{t}) + f'(\bar{t})(t - \bar{t})$ at $\bar{t}$ represents a straight line in $\mathbb{R}^m$ through the point $f(\bar{t})$ and the Jacobian $f'(\bar{t}) = (f_1'(\bar{t}), \ldots, f_m'(\bar{t}))$ gives the direction of the *tangent* to the curve $f : \mathbb{R} \to \mathbb{R}^m$ at $f(\bar{t})$, see Fig. 54.7.

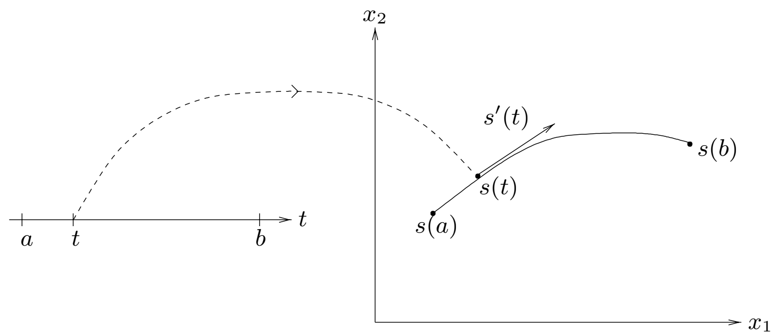

Fig. 54.7. The tangent $s'(t)$ to a curve given by $s(t)$

54.7 The Chain Rule

Let $g : \mathbb{R}^n \to \mathbb{R}^m$ and $f : \mathbb{R}^m \to \mathbb{R}^p$ and consider the composite function $f \circ g : \mathbb{R}^n \to \mathbb{R}^p$ defined by $f \circ g(x) = f(g(x))$. Under suitable assumptions of differentiability and Lipschitz continuity, we shall prove a *Chain rule* generalizing the Chain rule of Chapter *Differentiation rules* in the case

$n = m = p = 1$. Using linearizations of f and g, we have

$$f(g(x)) = f(g(\bar{x})) + f'(g(\bar{x}))(g(x) - g(\bar{x})) + E_f(g(x), g(\bar{x}))$$
$$= f(g(\bar{x})) + f'(g(\bar{x}))g'(\bar{x})(x - \bar{x}) + f'(g(\bar{x}))E_g(x, \bar{x}) + E_f(g(x), g(\bar{x})),$$

where we may naturally assume that

$$\|E_f(g(x), g(\bar{x}))\| \leq K_f \|g(x) - g(\bar{x})\|^2 \leq K_f L_g^2 \|x - \bar{x}\|^2,$$

and $\|f'(g(\bar{x}))E_g(x, \bar{x})\| \leq \|f'(g(\bar{x}))\| K_g \|x - \bar{x}\|^2$, with suitable constants of differentiability K_f and K_g and Lipschitz constant L_g. We have now proved:

Theorem 54.1 (The Chain rule) *If $g : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\bar{x} \in \mathbb{R}^n$, and $f : \mathbb{R}^m \to \mathbb{R}^p$ is differentiable at $g(\bar{x}) \in \mathbb{R}^m$ and further $g : \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz continuous, then the composite function $f \circ g : \mathbb{R}^n \to \mathbb{R}^p$ is differentiable at $\bar{x} \in \mathbb{R}^n$ with Jacobian*

$$(f \circ g)'(\bar{x}) = f'(g(\bar{x}))g'(\bar{x}).$$

The Chain rule has a wealth of applications and we now turn to harvest a couple of the most basic examples.

54.8 The Mean Value Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable on $\mathbb{R}^n$ with a Lipschitz continuous gradient, and for given $x, \bar{x} \in \mathbb{R}^n$ consider the function $h : \mathbb{R} \to \mathbb{R}$ defined by

$$h(t) = f(\bar{x} + t(x - \bar{x})) = f \circ g(t),$$

with $g(t) = \bar{x} + t(x - \bar{x})$ representing the straight line through $\bar{x}$ and x. We have

$$f(x) - f(\bar{x}) = h(1) - h(0) = h'(\bar{t}),$$

for some $\bar{t} \in [0, 1]$, where we applied the usual Mean Value theorem to the function $h(t)$. By the Chain rule we have

$$h'(t) = \nabla f(g(t)) \cdot g'(t) = \nabla f(g(t)) \cdot (x - \bar{x}),$$

and we have now proved:

Theorem 54.2 (Mean Value theorem) *Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable on $\mathbb{R}^n$ with a Lipschitz continuous gradient ∇f. Then for given x and $\bar{x}$ in $\mathbb{R}^n$, there is $y = x + \bar{t}(x - \bar{x})$ with $\bar{t} \in [0, 1]$, such that*

$$f(x) - f(\bar{x}) = \nabla f(y) \cdot (x - \bar{x}).$$

With the help of the Mean Value theorem we express the difference $f(x) - f(\bar{x})$ as the scalar product of the gradient $\nabla f(y)$ with the difference $x - \bar{x}$, where y is a point somewhere on the straight line between x and $\bar{x}$.

We may extend the Mean Value theorem to a function $f : \mathbb{R}^n \to \mathbb{R}^m$ to take the form

$$f(x) - f(\bar{x}) = f'(y)(x - \bar{x}),$$

where y is a point on the straight line between x and $\bar{x}$, which may be different for different rows of $f'(y)$. We may then estimate:

$$\|f(x) - f(\bar{x})\| = \|f'(y) \cdot (x - \bar{x})\| \leq \|f'(y)\|\|x - \bar{x}\|,$$

and we may thus estimate the Lipschitz constant of f by $\max_y \|f'(y)\|$ with $\|f'(y)\|$ the (Euclidean) matrix norm of $f'(y)$.

Example 54.15. Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by $f(x) = \sin(\sum_{j=1}^n x_j)$. We have

$$\frac{\partial f}{\partial x_i}(\bar{x}) = \cos\left(\sum_{j=1}^n \bar{x}_j\right) \quad \text{for } i = 1, \ldots, n,$$

and thus $|\frac{\partial f}{\partial x_i}(\bar{x})| \leq 1$ for $i = 1, \ldots, n$, and therefore

$$\|\nabla f(\bar{x})\| \leq \sqrt{n}.$$

We conclude that $f(x) = \sin(\sum_{j=1}^n x_j)$ is Lipschitz continuous with Lipschitz constant $\sqrt{n}$.

54.9 Direction of Steepest Descent and the Gradient

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a given function and suppose we want to study the variation of $f(x)$ in a neighborhood of a given point $\bar{x} \in \mathbb{R}^n$. More precisely, let x vary on the line through $\bar{x}$ in a given direction $z \in \mathbb{R}^n$, that is assume that $x = \bar{x} + tz$ where t is a real variable varying in a neighborhood of 0. Assuming f to be differentiable, the linearization formula (54.8) implies

$$f(x) = f(\bar{x}) + t\nabla f(\bar{x}) \cdot z + E_f(x, \bar{x}), \tag{54.12}$$

where $|E_f(x, \bar{x})| \leq t^2 K_f \|z\|^2$ and $\nabla f(\bar{x}) \cdot z$ is the scalar product of the gradient $\nabla f(\bar{x}) \in \mathbb{R}^n$ and the vector $z \in \mathbb{R}^n$. If $\nabla f(\bar{x}) \cdot z \neq 0$, then the linear term $t\nabla f(\bar{x}) \cdot z$ will dominate the quadratic term $E_f(x, \bar{x})$ for small t. So the linearization

$$\hat{f}(x) = f(\bar{x}) + t\nabla f(\bar{x}) \cdot z$$

will be a good approximation of $f(x)$ for $x = \bar{x} + tz$ close to $\bar{x}$. Thus if $\nabla f(\bar{x}) \cdot z \neq 0$, then we get good information on the variation of $f(x)$ along the line $x = \bar{x} + tz$ by studying the linear function $t \to f(\bar{x}) + t\nabla f(\bar{x}) \cdot z$ with slope $\nabla f(\bar{x}) \cdot z$. In particular, if $\nabla f(\bar{x}) \cdot z > 0$ and $x = \bar{x} + tz$ then $\hat{f}(x)$ increases as we increase t and decreases as we decrease t. Similarly, if $\nabla f(\bar{x}) \cdot z < 0$ and $x = \bar{x} + tz$ then $\hat{f}(x)$ decreases as we increase t and increases as we decrease t.

Alternatively, we may consider the composite function $F_z : \mathbb{R} \to \mathbb{R}$ defined by $F_z(t) = f(g_z(t))$ with $g_z : \mathbb{R} \to \mathbb{R}^n$ given by $g_z(t) = \bar{x} + tz$. Obviously, $F_z(t)$ describes the variation of $f(x)$ on the straight line through $\bar{x}$ with direction z, with $F_z(0) = f(\bar{x})$. Of course, the derivative $F_z'(0)$ gives important information on this variation close to $\bar{x}$. By the Chain rule we have

$$F_z'(0) = \nabla f(\bar{x}) z = \nabla f(\bar{x}) \cdot z,$$

and we retrieve $\nabla f(\bar{x}) \cdot z$ as a quantity of interest. In particular, the sign of $\nabla f(\bar{x}) \cdot z$ determines if $F_z(t)$ is increasing or decreasing at $t = 0$.

We may now ask how to choose the direction z to get maximal increase or decrease. We assume $\nabla f(\bar{x}) \neq 0$ to avoid the trivial case with $F_z'(0) = 0$ for all z. It is then natural to normalize z so $\|z\| = 1$ and we study the quantity $F_z'(0) = \nabla f(\bar{x}) \cdot z$ as we vary $z \in \mathbb{R}^n$ with $\|z\| = 1$. We conclude that the scalar product $\nabla f(\bar{x}) \cdot z$ is maximized if we choose z in the direction of the gradient $\nabla f(\bar{x})$,

$$z = \frac{\nabla f(\bar{x})}{\|\nabla f(\bar{x})\|},$$

which is called the direction of *steepest ascent*. For this gives

$$\max_{\|z\|=1} F_z'(0) = \nabla f(\bar{x}) \cdot \frac{\nabla f(\bar{x})}{\|\nabla f(\bar{x})\|} = \|\nabla f(\bar{x})\|.$$

Similarly, the scalar product is minimized if we choose z in the opposite direction of the gradient $\nabla f(\bar{x})$,

$$z = -\frac{\nabla f(\bar{x})}{\|\nabla f(\bar{x})\|},$$

which is called the direction of *steepest descent*, see Fig. 54.8. For then

$$\min_{\|z\|=1} F_z'(0) = -\nabla f(\bar{x}) \cdot \frac{\nabla f(\bar{x})}{\|\nabla f(\bar{x})\|} = -\|\nabla f(\bar{x})\|.$$

If $\nabla f(\bar{x}) = 0$, then $\bar{x}$ is said to be a *stationary point*. If $\bar{x}$ is a stationary point, then evidently $\nabla f(\bar{x}) \cdot z = 0$ for any direction z and

$$f(x) = f(\bar{x}) + E_f(x, \bar{x}).$$

The difference $f(x) - f(\bar{x})$ is then quadratically small in the distance $\|x - \bar{x}\|$, that is $|f(x) - f(\bar{x})| \leq K_f \|x - \bar{x}\|^2$, and $f(x)$ is very close to the constant value $f(\bar{x})$ for x close to $\bar{x}$.

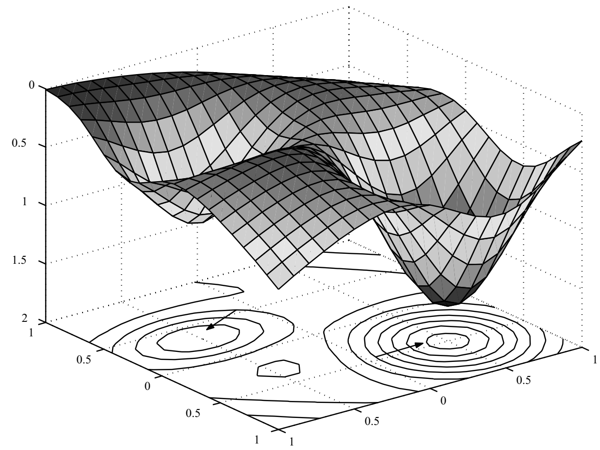

Fig. 54.8. Directions of steepest descent on a "hiking map"

54.10 A Minimum Point Is a Stationary Point

Suppose $\bar{x} \in \mathbb{R}^n$ is a *minimum point* for the function $f : \mathbb{R}^n \to \mathbb{R}$, that is

$$f(x) \geq f(\bar{x}) \quad \text{for } x \in \mathbb{R}^n. \tag{54.13}$$

We shall show that if $f(x)$ is differentiable at a minimum point $\bar{x}$, then

$$\nabla f(\bar{x}) = 0. \tag{54.14}$$

For if $\nabla f(\bar{x}) \neq 0$, then we could move in the direction of steepest descent from $\bar{x}$ to a point x close to $\bar{x}$ with $f(x) < f(\bar{x})$, contradicting (54.13). Consequently, in order to find minimum points of a function $f : \mathbb{R}^n \to \mathbb{R}$, we are led to try to solve the equation $g(x) = 0$, where $g = \nabla f : \mathbb{R}^n \to \mathbb{R}^n$. Here, we interpret $\nabla f(x)$ as a n-column vector.

A whole world of applications in mechanics, physics and other areas may be formulated as solving equations of the form $\nabla f(x) = 0$, that is as finding stationary points. We shall meet many applications below.

54.11 The Method of Steepest Descent

Let $f : \mathbb{R}^n \to \mathbb{R}$ be given and consider the problem of finding a minimum point $\bar{x}$. To do so it is natural to try a *method of Steepest Descent*: Given an approximation $\bar{y}$ of $\bar{x}$ with $\nabla f(\bar{y}) \neq 0$, we move from $\bar{y}$ to a new point y in the direction of steepest descent:

$$y = \bar{y} - \alpha \frac{\nabla f(\bar{y})}{\|\nabla f(\bar{y})\|},$$

where $\alpha > 0$ is a step length to be chosen. We know that $f(y)$ decreases as α increases from 0 and the question is just to find a reasonable value of α. This can be done by increasing α in small steps until $f(y)$ doesn't decrease anymore. The procedure is then repeated with $\bar{y}$ replaced by y. Evidently, the method of Steepest Descent is closely connected to Fixed Point Iteration for solving the equation $\nabla f(x) = 0$ in the form

$$x = x - \alpha \nabla f(x)$$

where we let $\alpha > 0$ include the normalizing factor $1/\|\nabla f(\bar{y})\|$.

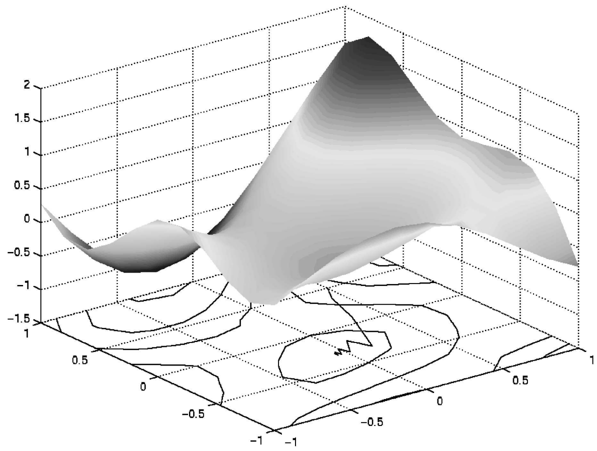

Fig. 54.9. The method of Steepest Descent for $f(x_1, x_2) = x_1 \sin(x_1 + x_2) + x_2 \cos(2x_1 - 3x_2)$ starting at $(.5, .5)$ with $\alpha = .3$

54.12 Directional Derivatives

Consider a function $f : \mathbb{R}^n \to \mathbb{R}$, let $g_z(t) = \bar{x} + tz$ with $z \in \mathbb{R}^n$ a given vector normalized to $\|z\| = 1$, and consider the composite function $F_z(t) = f(\bar{x} + tz)$. The Chain rule implies

$$F_z'(0) = \nabla f(\bar{x}) \cdot z,$$

and

$$\nabla f(\bar{x}) \cdot z$$

is called the *derivative of* $f(x)$ *in the direction* z at $\bar{x}$, see Fig. 54.10.

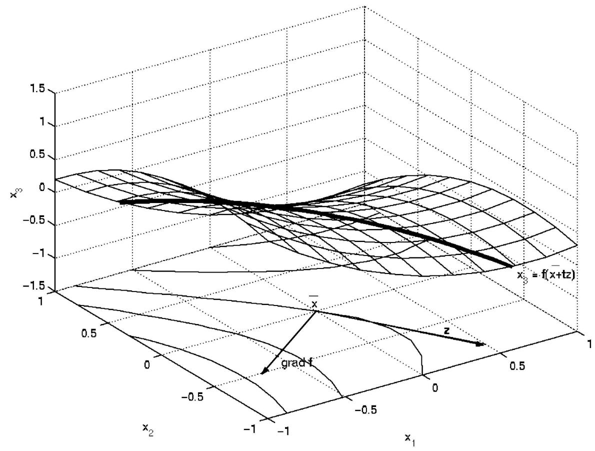

Fig. 54.10. Illustration of directional derivative

54.13 Higher Order Partial Derivatives

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable on $\mathbb{R}^n$. Each partial derivative $\frac{\partial f}{\partial x_i}(\bar{x})$ is a function of $\bar{x} \in \mathbb{R}^n$ may be itself be differentiable. We denote its partial derivatives by

$$\frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i}(\bar{x}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\bar{x}), \quad i, j = 1, \ldots, n, \, \bar{x} \in \mathbb{R}^n,$$

which are called the *partial derivatives of* f *of second order at* $\bar{x}$. It turns out that under appropriate continuity assumptions, the order of differentiation does not matter. In other words, we shall prove that

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(\bar{x}) = \frac{\partial^2 f}{\partial x_i \partial x_j}(\bar{x}).$$

We carry out the proof in the case $n = 2$ with $i = 1$ and $j = 2$. We rewrite the expression

$$A = f(x_1, x_2) - f(\bar{x}_1, x_2) - f(x_1, \bar{x}_2) + f(\bar{x}_1, \bar{x}_2), \tag{54.15}$$

as

$$A = f(x_1, x_2) - f(x_1, \bar{x}_2) - f(\bar{x}_1, x_2) + f(\bar{x}_1, \bar{x}_2), \tag{54.16}$$

by shifting the order of the two mid terms. First, we set $F(x_1, x_2) = f(x_1, x_2) - f(\bar{x}_1, x_2)$ and use (54.15) to write

$$A = F(x_1, x_2) - F(x_1, \bar{x}_2).$$

The Mean Value theorem implies

$$A = \frac{\partial F}{\partial x_2}(x_1, y_2)(x_2 - \bar{x}_2) = \left(\frac{\partial f}{\partial x_2}(x_1, y_2) - \frac{\partial f}{\partial x_2}(\bar{x}_1, y_2)\right)(x_2 - \bar{x}_2)$$

for some $y_2 \in [\bar{x}_2, x_2]$. We use the Mean value theorem once again to get

$$A = \frac{\partial^2 f}{\partial x_1 \partial x_2}(y_1, y_2)(x_1 - \bar{x}_1)(x_2 - \bar{x}_2),$$

with $y_1 \in [\bar{x}_1, x_1]$. We next rewrite A using (54.16) in the form

$$A = G(x_1, x_2) - G(\bar{x}_1, x_2),$$

where $G(x_1, x_2) = f(x_1, x_2) - f(x_1, \bar{x}_2)$. Using the Mean Value theorem twice as above, we obtain

$$A = \frac{\partial^2 f}{\partial x_2 \partial x_1}(z_1, z_2)(x_1 - \bar{x}_1)(x_2 - \bar{x}_2),$$

where $z_i \in [\bar{x}_i, x_i]$, $i = 1, 2$. Assuming the second partial derivatives are Lipschitz continuous at $\bar{x}$ and letting x_i tend to $\bar{x}_i$ for $i = 1, 2$ gives

$$\frac{\partial^2 f}{\partial x_1 \partial x_2}(\bar{x}) = \frac{\partial^2 f}{\partial x_2 \partial x_1}(\bar{x}).$$

We have proved the following fundamental result:

Theorem 54.3 *If the partial derivatives of second order of a function $f : \mathbb{R}^n \to \mathbb{R}$ are all Lipschitz continuous, then the order of application of the derivatives of second order is irrelevant.*

The result directly generalizes to higher order partial derivatives: if the derivatives are Lipschitz continuous, then the order of application doesn't matter. What a relief!

54.14 Taylor's Theorem

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ has Lipschitz continuous partial derivatives of order 2. For given $x, \bar{x} \in \mathbb{R}^n$, consider the function $h : \mathbb{R} \to \mathbb{R}$ defined by

$$h(t) = f(\bar{x} + t(x - \bar{x})) = f \circ g(t),$$

where $g(t) = \bar{x} + t(x - \bar{x})$ is the straight line through $\bar{x}$ and x. Clearly $h(1) = f(x)$ and $h(0) = f(\bar{x})$, so the Taylor's theorem applied to $h(t)$ gives

$$h(1) = h(0) + h'(0) + \frac{1}{2}h''(\bar{t}),$$

for some $\bar{t} \in [0,1]$. We compute using the Chain rule:

$$h'(t) = \nabla f(g(t)) \cdot (x - \bar{x}) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(g(t))(x_i - \bar{x}_i),$$

and similarly by a further differentiation with respect to t:

$$h''(t) = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(g(t))(x_i - \bar{x}_i)(x_j - \bar{x}_j).$$

We thus obtain

$$f(x) = f(\bar{x}) + \nabla f(\bar{x}) \cdot (x - \bar{x}) + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(y)(x_i - \bar{x}_i)(x_j - \bar{x}_j), \quad (54.17)$$

for some $y = \bar{x} + \bar{t}(x - \bar{x})$ with $t \in [0,1]$. The $n \times n$ matrix $H(\bar{x}) = (h_{ij}(\bar{x}))$ with elements $h_{ij}(\bar{x}) = \frac{\partial^2 f}{\partial x_i \partial x_j}(\bar{x})$ is called the *Hessian* of $f(x)$ at $x = \bar{x}$. The Hessian is the matrix of all second partial derivatives of $f : \mathbb{R}^n \to \mathbb{R}$. With matrix vector notation with x a n-column vector, we can write

$$\sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(y)(x_i - \bar{x}_i)(x_j - \bar{x}_j) = (x - \bar{x})^\top H(y)(x - \bar{x}).$$

We summarize:

Theorem 54.4 (Taylor's theorem) *Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable with Lipschitz continuous Hessian $H = (h_{ij})$ with elements $h_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$. Then, for given x and $\bar{x} \in \mathbb{R}^n$, there is $y = x + \bar{t}(x - \bar{x})$ with $\bar{t} \in [0,1]$, such that*

$$\begin{aligned} f(x) &= f(\bar{x}) + \nabla f(\bar{x}) \cdot (x - \bar{x}) + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(y)(x_i - \bar{x}_i)(x_j - \bar{x}_j) \\ &= f(\bar{x}) + \nabla f(\bar{x}) \cdot (x - \bar{x}) + \frac{1}{2}(x - \bar{x})^\top H(y)(x - \bar{x}). \end{aligned}$$

54.15 The Contraction Mapping Theorem

We shall now prove the following generalization of the Contraction Mapping theorem.

Theorem 54.5 *If $g : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous with Lipschitz constant $L < 1$, then the equation $x = g(x)$ has a unique solution $\bar{x} = \lim_{i\to\infty} x^{(i)}$, where $\{x^{(i)}\}_{i=1}^{\infty}$ is a sequence in $\mathbb{R}^n$ generated by Fixed Point Iteration: $x^{(i)} = g(x^{(i-1)})$, $i = 1, 2, \ldots$, starting with any initial value $x^{(0)}$.*

The proof is word by word the same as in the case $g : \mathbb{R} \to \mathbb{R}$ considered in Chapter *Fixed Points and Contraction Mappings*. We repeat the proof for the convenience of the reader. Subtracting the equation $x^{(k)} = g(x^{(k-1)})$ from $x^{(k+1)} = g(x^{(k)})$, we get

$$x^{(k+1)} - x^{(k)} = g(x^{(k)}) - g(x^{(k-1)}),$$

and using the Lipschitz continuity of g, we thus have

$$\|x^{(k+1)} - x^{(k)}\| \leq L\|x^{(k)} - x^{(k-1)}\|.$$

Repeating this estimate, we find that

$$\|x^{(k+1)} - x^{(k)}\| \leq L^k\|x^{(1)} - x^{(0)}\|,$$

and thus for $j > i$

$$\begin{aligned}\|x^{(i)} - x^{(j)}\| &\leq \sum_{k=i}^{j-1} \|x^{(k)} - x^{(k+1)}\| \\ &\leq \|x^{(1)} - x^{(0)}\| \sum_{k=i}^{j-1} L^k = \|x^{(1)} - x^{(0)}\| L^i \, \frac{1 - L^{j-i}}{1 - L}.\end{aligned}$$

Since $L < 1$, $\{x^{(i)}\}_{i=1}^{\infty}$ is a Cauchy sequence in $\mathbb{R}^n$, and therefore converges to a limit $\bar{x} = \lim_{i\to\infty} x^{(i)}$. Passing to the limit in the equation $x^{(i)} = g(x^{(i-1)})$ shows that $\bar{x} = g(\bar{x})$ and thus $\bar{x}$ is a fixed point of $g : \mathbb{R}^n \to \mathbb{R}^n$. Uniqueness follows from the fact that if $\bar{y} = g(\bar{y})$, then $\|\bar{x} - \bar{y}\| = \|g(\bar{x}) - g(\bar{y})\| \leq L\|\bar{x} - \bar{y}\|$ which is impossible unless $\bar{y} = \bar{x}$, because $L < 1$.

Example 54.16. Consider the function $g : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $g(x) = (g_1(x), g_2(x))$ with

$$g_1(x) = \frac{1}{4 + |x_1| + |x_2|}, \quad g_2(x) = \frac{1}{4 + |\sin(x_1)| + |\cos(x_2)|}.$$

We have

$$|\frac{\partial g_i}{\partial x_j}| \leq \frac{1}{16},$$

and thus by simple estimates

$$\|g(x) - g(y)\| \leq \frac{1}{4}\|x - y\|,$$

which shows that $g : \mathbb{R}^2 \to \mathbb{R}^2$ is Lipschitz continuous with Lipschitz constant $L_g \leq \frac{1}{4}$. The equation $x = g(x)$ thus has a unique solution.

54.16 Solving $f(x) = 0$ with $f : \mathbb{R}^n \to \mathbb{R}^n$

The Contraction Mapping theorem can be applied as follows. Suppose $f : \mathbb{R}^n \to \mathbb{R}^n$ is given and we want to solve the equation $f(x) = 0$. Introduce

$$g(x) = x - Af(x),$$

where A is some non-singular $n \times n$ matrix with constant coefficients to be chosen. The equation $x = g(x)$ is then equivalent to the equation $f(x) = 0$. If $g : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous with Lipschitz constant $L < 1$, then $g(x)$ has a unique fixed point $\bar{x}$ and thus $f(\bar{x}) = 0$. We have

$$g'(x) = I - Af'(x),$$

and thus we are led to choose the matrix A so that

$$\|I - Af'(x)\| \leq 1$$

for x close to the root $\bar{x}$. The ideal choice seems to be:

$$A = f'(\bar{x})^{-1},$$

assuming that $f'(\bar{x})$ is non-singular, since then $g'(\bar{x}) = 0$. In applications, we may seek to choose A close to $f'(\bar{x})^{-1}$ with the hope that the corresponding $g'(x) = I - Af'(x)$ will have $\|g'x)$ TS^{c} $\|$ small for x close to the root $\bar{x}$, leading to a quick convergence. In Newton's method we choose $A = f'(x)^{-1}$, see below.

Example 54.17. Consider the initial value problem $\dot{u}(t) = f(u(t))$ for $t > 0$, $u(0) = u_0$, where $f : \mathbb{R}^n \to \mathbb{R}^n$ is a given Lipschitz continuous function with Lipschitz constant L_f, and as usual $\dot{u} = \frac{du}{dt}$. Consider the backward Euler method

$$U(t_i) = U(t_{i-1}) + k_i f(U(t_i)), \tag{54.18}$$

where $0 = t_0 < t_1 < t_2 \ldots$ is a sequence of increasing discrete time levels with time steps $k_i = t_i - t_{i-1}$. To determine $U(t_i) \in \mathbb{R}^n$ satisfying (54.18) having already determined $U(t_{i-1})$, we have to solve the nonlinear system of equations

$$V = U(t_{i-1}) + k_i f(V) \tag{54.19}$$

in the unknown $V \in \mathbb{R}^n$. This equation is of the form $V = g(V)$ with $g(V) = U(t_{i-1}) + k_i f(V)$ and $g : \mathbb{R}^n \to \mathbb{R}^n$.

Therefore, we use the Fixed Point Iteration

$$V^{(m)} = U(t_{i-1}) + k_i f(V^{(m-1)}), \quad m = 1, 2, \ldots,$$

choosing say $V^{(0)} = U(t_{i-1})$ to try to solve for the new value. If L_f denotes the Lipschitz constant of $f : \mathbb{R}^n \to \mathbb{R}^n$, then

$$\|g(V) - g(W)\| = \|k_i(f(V) - f(W))\| \leq k_i L_f \|V - W\|, \quad V, W \in \mathbb{R}^n,$$

TS^{c} Is there an opening parenthesis missing here?

Editor's or typesetter's annotations (will be removed before the final TEX run) and thus $g : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous with Lipschitz constant $L_g = k_i L_f$. Now $L_g < 1$ if the time step k_i satisfies $k_i < 1/L_f$ and thus the Fixed Point Iteration to determine $U(t_i)$ in (54.18) converges if $k_i < 1/L_f$. This gives a method for numerical solution of a very large class of initial value problems of the form $\dot{u}(t) = f(u(t))$ for $t > 0$, $u(0) = u_0$. The only restriction is to choose sufficiently small time steps, which however can be a severe restriction if the Lipschitz constant L_f is very large in the sense of requiring massive computational work (very small time steps). Thus, caution for large Lipschitz constants L_f!!

54.17 The Inverse Function Theorem

Suppose $f : \mathbb{R}^n \to \mathbb{R}^n$ is a given function and let $\bar{y} = f(\bar{x})$, where $\bar{x} \in \mathbb{R}^n$ is given. We shall prove that if $f'(\bar{x})$ is non-singular, then for $y \in \mathbb{R}^n$ close to $\bar{y}$, the equation

$$f(x) = y \tag{54.20}$$

has a unique solution x. Thus, we can define x as a function of y for y close to $\bar{y}$, which is called the inverse function $x = f^{-1}(y)$ of $y = f(x)$. To show that (54.20) has a unique solution x for any given y close to $\bar{y}$, we consider the Fixed Point iteration for $x = g(x)$ with $g(x) = x - (f'(\bar{x}))^{-1}(f(x) - y)$, which has the fixed point x satisfying $f(x) = y$ as desired. The iteration is

$$x^{(j)} = x^{(j-1)} - (f'(\bar{x}))^{-1}(f(x^{(j-1)}) - y), \quad j = 1, 2, \dots,$$

with $x^{(0)} = \bar{x}$. To analyze the convergence, we subtract

$$x^{(j-1)} = x^{(j-2)} - (f'(\bar{x}))^{-1}(f(x^{(j-2)}) - y)$$

and write $e^j = x^{(j)} - x^{(j-1)}$ to get

$$e^j = e^{j-1} - (f'(\bar{x}))^{-1}(f(x^{(j-1)}) - f(x^{(j-2)}) \quad \text{for } j = 1, 2, \dots .$$

The Mean Value theorem implies

$$f_i(x^{(j-1)}) - f_i(x^{(j-2)}) = f'(z)e^{j-1},$$

where z lies on the straight line between $x^{(j-1)}$ and $x^{(j-2)}$. Note there might be possibly different z for different rows of $f'(z)$. We conclude that

$$e^j = \left(I - (f'(\bar{x}))^{-1} f'(z)\right) e^{j-1}.$$

Assuming now that

$$\|I - (f'(\bar{x}))^{-1} f'(z)\| \leq \theta, \tag{54.21}$$

where $\theta < 1$ is a positive constant, we have

$$\|e^j\| \leq \theta \|e^{j-1}\|.$$

As in the proof of the Contraction Mapping theorem, this shows that the sequence $\{x^{(j)}\}_{j=1}^{\infty}$ is a Cauchy sequence and thus converges to a vector $x \in \mathbb{R}^n$ satisfying $f(x) = y$.

The condition for convergence is obviously (54.21). This condition is satisfied if the coefficients of the Jacobian $f'(x)$ are Lipschitz continuous close to $\bar{x}$ and $f'(\bar{x})$ is non-singular so that $(f'(\bar{x}))^{-1}$ exists, and we restrict y to be sufficiently close to $\bar{y}$.

We summarize in the following (very famous):

Theorem 54.6 (Inverse Function theorem) *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ and assume the coefficients of $f'(x)$ are Lipschitz continuous close to $\bar{x}$ and $f'(\bar{x})$ is non-singular. Then for y sufficiently close to $\bar{y} = f(\bar{x})$, the equation $f(x) = y$ has a unique solution x. This defines x as a function $x = f^{-1}(y)$ of y.*

Carl Jacobi (1804–51), German mathematician, was the first to study the role of the determinant of the Jacobian in the inverse function theorem, and also gave important contributions to many areas of mathematics including the budding theory of first order partial differential equations.

54.18 The Implicit Function Theorem

There is an important generalization of the Inverse Function theorem. Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ be a given function with value $f(x, y) \in \mathbb{R}^n$ for $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. Assume that $f(\bar{x}, \bar{y}) = 0$ and consider the equation in $x \in \mathbb{R}^n$,

$$f(x, y) = 0,$$

for $y \in \mathbb{R}^m$ close to $\bar{y}$. In the case of the Inverse Function theorem we considered a special case of this situation with $f : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ defined by $f(x, y) = g(x) - y$ with $g : \mathbb{R}^n \to \mathbb{R}^n$.

We define the Jacobian $f'_x(x, y)$ of $f(x, y)$ with respect to x at (x, y) to be the $n \times n$ matrix with elements

$$\frac{\partial f_i}{\partial x_j}(x, y).$$

Assuming now that $f'_x(\bar{x}, \bar{y})$ is non-singular, we consider the Fixed Point iteration:

$$x^{(j)} = x^{(j-1)} - (f'_x(\bar{x}, \bar{y}))^{-1} f(x^{(j-1)}, y),$$

connected to solving the equation $f(x, y) = 0$. Arguing as above, we can show this iteration generates a sequence $\{x^{(j)}\}_{j=1}\infty$ that converges to $x \in \mathbb{R}^n$ satisfying $f(x, y) = 0$ assuming $f'_x(x, y)$ is Lipschitz continuous for x close to $\bar{x}$ and y close to $\bar{y}$. This defines x as a function $g(y)$ of y for y close to $\bar{y}$. We have now proved the (also very famous):

Theorem 54.7 (Implicit Function theorem) *Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ with $f(x, y) \in \mathbb{R}^n$ and $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, and assume that $f(\bar{x}, \bar{y}) = 0$. Assume that the Jacobian $f'_x(x, y)$ with respect to x is Lipschitz continuous for x close to $\bar{x}$ and y close to $\bar{y}$, and that $f'_x(\bar{x}, \bar{y})$ is non-singular. Then for y close to $\bar{y}$, the equation $f(x, y) = 0$ has a unique solution $x = g(y)$. This defines x as a function $g(y)$ of y.*

54.19 Newton's Method

We next turn to *Newton's method* for solving an equation $f(x) = 0$ with $f : \mathbb{R}^n \to \mathbb{R}^n$, which reads:

$$x^{(i+1)} = x^{(i)} - f'(x^{(i)})^{-1} f(x^{(i)}), \quad \text{for } i = 0, 1, 2, \ldots, \tag{54.22}$$

where $x^{(0)}$ is an initial approximation. Newton's method corresponds to Fixed Point iteration for $x = g(x)$ with $g(x) = x - f'(x)^{-1} f(x)$. We shall prove that Newton's method converges quadratically close to a root $\bar{x}$ when $f'(\bar{x})$ is non-singular. The argument is the same is as in the case $n = 1$ considered above. Setting $e^i = \bar{x} - x^{(i)}$, and using $\bar{x} = \bar{x} - f'(x^{(i)})^{-1} f(\bar{x})$ if $f(\bar{x}) = 0$, we have

$$\begin{aligned}
\bar{x} - x^{(i+1)} &= \bar{x} - x^{(i)} - f'(x^{(i)})^{-1}(f(\bar{x}) - f(x^{(i)})) \\
&= \bar{x} - x^{(i)} - f'(x^{(i)})^{-1}(f'(x^{(i)}) + E_f(x^{(i)}, \bar{x})) = f'(x^{(i)})^{-1} E_f(x^{(i)}, \bar{x}).
\end{aligned}$$

We conclude that

$$\|\bar{x} - x^{(i+1)}\| \leq C \|\bar{x} - x^{(i)}\|^2$$

provided

$$\|f'(x^{(i)})^{-1}\| \leq C,$$

where C is some positive constant. We have proved the following fundamental result:

Theorem 54.8 (Newton's method) *If $\bar{x}$ is a root of $f : \mathbb{R}^n \to \mathbb{R}^n$ such that $f(x)$ is uniformly differentiable with a Lipschitz continuous derivative close to $\bar{x}$ and $f'(\bar{x})$ is non-singular, then Newton's method for solving $f(x) = 0$ converges quadratically if started sufficiently close to $\bar{x}$.*

In concrete implementations of Newton's method we may rewrite (54.22) as

$$\begin{aligned}
f'(x^{(i)}) z &= -f(x^{(i)}), \\
x^{(i+1)} &= x^{(i)} + z,
\end{aligned}$$

where $f'(x^{(i)}) z = -f(x^{(i)})$ is a system of equations in z that is solved by Gaussian elimination or by some iterative method.

Example 54.18. We return to the equation (54.19), that is,

$$h(V) = V - k_i f(V) - U(t_{i-1}) = 0.$$

To apply Newton's method to solve the equation $h(V) = 0$, we compute

$$h'(v) = I - k_i f'(v),$$

and conclude that $h'(v)$ will be non-singular at v, if $k_i < \|f'(v)\|^{-1}$. We conclude that Newton's method converges if k_i is sufficiently small and we start close to the root. Again the restriction on the time step is connected to the Lipschitz constant L_f of f, since L_f reflects the size of $\|f'(v)\|$.

54.20 Differentiation Under the Integral Sign

Finally, we show that if the limits of integration of an integral are independent of a variable x_1, then the operation of taking the partial derivative with respect x_1 can be moved past the integral sign. Let then $f : \mathbb{R}^2 \to \mathbb{R}$ be a function of two real variables x_1 and x_2 and consider the integral

$$\int_0^1 f(x_1, x_2)\, dx_2 = g(x_1),$$

which is a function $g(x_1)$ of x_1. We shall now prove that

$$\frac{dg}{dx_1}(\bar{x}_1) = \int_0^1 \frac{\partial f}{\partial x_1}(\bar{x}_1, x_2)\, dx_2, \qquad (54.23)$$

which is referred to as "differentiation under the integral sign". The proof starts by writing

$$f(x_1, x_2) = f(\bar{x}_1, x_2) + \frac{\partial f}{\partial x_1}(\bar{x}_1, x_2)(x_1 - \bar{x}_1) + E_f(x_1, \bar{x}_1, x_2),$$

where we assume that

$$|E_f(x_1, \bar{x}_1, x_2)| \le K_f(\bar{x}_1 - x_1)^2.$$

Taylor's theorem implies this is true provided the second partial derivatives of f are bounded. Integration with respect to x_2 yields

$$\int_0^1 f(x_1, x_2)\, dx_2 = \int_0^1 f(\bar{x}_1, x_2)\, dx_2 + (x_1 - \bar{x}_1)\int_0^1 \frac{\partial f}{\partial x_1}(\bar{x}_1, x_2)\, dx_2 + \int_0^1 E_f(x_1, \bar{x}_1, x_2)\, dx_2.$$

Since

$$\left|\int_0^1 E_f(x_1,\bar{x}_1,x_2)\,dx_2\right| \le K_f(\bar{x}_1 - x_1)^2$$

(54.23) follows after dividing by $(x_1 - \bar{x}_1)$ and taking the limit as x_1 tends to $\bar{x}_1$. We summarize:

Theorem 54.9 (Differentiation under the integral sign) *If the second partial derivatives of $f(x_1, x_2)$ are bounded, then for $x_1 \in \mathbb{R}$,*

$$\frac{d}{dx_1}\int_0^1 f(x_1,x_2)\,dx_2 = \int_0^1 \frac{\partial f}{\partial x_1}(x_1,x_2)\,dx_2 \tag{54.24}$$

Example 54.19.

$$\frac{d}{dx}\int_0^1 (1+xy^2)^{-1}\,dy = \int_0^1 \frac{\partial}{\partial x}(1+xy^2)^{-1}\,dy = -\int_0^1 \frac{y^2}{(1+xy^2)^2}\,dy.$$

Chapter 54 Problems

54.1. Sketch the following surfaces in $\mathbb{R}^3$: (a) $\Gamma = \{x : x_3 = x_1^2 + x_2^2\}$, (b) $\Gamma = \{x : x_3 = x_1^2 - x_2^2\}$, (c) $\Gamma = \{x : x_3 = x_1 + x_2^2\}$, (d) $\Gamma = \{x : x_3 = x_1^4 + x_2^6\}$. Determine the tangent planes to the surfaces at different points.

54.2. Determine whether the following functions are Lipschitz continuous or not on $\{x : |x| < 1\}$ and determine Lipschitz constants:

- (a) $f : \mathbb{R}^3 \to \mathbb{R}^3$ where $f(x) = x|x|^2$,
- (b) $f : \mathbb{R}^3 \to \mathbb{R}$ where $f(x) = \sin|x|^2$,
- (c) $f : \mathbb{R}^2 \to \mathbb{R}^3$ where $f(x) = (x_1, x_2, \sin|x|^2)$,
- (d) $f : \mathbb{R}^3 \to \mathbb{R}$ where $f(x) = 1/|x|$,
- (e) $f : \mathbb{R}^3 \to \mathbb{R}^3$ where $f(x) = x\sin(|x|)$, (optional)
- (f) $f : \mathbb{R}^3 \to \mathbb{R}$ where $f(x) = \sin(|x|)/|x|$. (optional)

54.3. For the functions in the previous exercise, determine which are contractions in $\{x : |x| < 1\}$ and find their fixed points (optional).

54.4. Linearize the following functions on $\mathbb{R}^3$ at $x = (1, 2, 3)$:

- (a) $f(x) = |x|^2$,
- (b) $f(x) = \sin(|x|^2)$,
- (c) $f(x) = (|x|^2, \sin(x_2))$,
- (d) $f(x) = (|x|^2, \sin(x_2), x_1x_2\cos(x_3))$.

54.5. Compute the determinant of the Jacobian of the following functions: (a) $f(x) = (x_1^3 - 3x_1x_2^2, 3x_1x_2^2 - x_2^3)$, (b) $f(x) = (x_1e^{x_2}\cos(x_3), x_1e^{x_2}\sin(x_3), x_1e^{x_2})$.

54.6. Compute the second order Taylor polynomials at $(0,0,0)$ of the following functions $f : \mathbb{R}^3 \rightarrow \mathbb{R}$: (a) $f(x) = \sqrt{1 + x_1 + x_2 + x_3}$, (b) $f(x) = (x_1 - 1)x_2x_3$, (c) $f(x) = \sin(\cos(x_1x_2x_3))$, (d) $\exp(-x_1^2 - x_2^2 - x_3^2)$, (e) try to estimate the errors in the approximations in (a)-(d).

54.7. Linearize $f \circ s$, where $f(x) = x_1x_2x_3$ at $t = 1$ with (a) $s(t) = (t, t^2, t^3)$, (b) $s(t) = (\cos(t), \sin(t), t)$, (c) $s(t) = (t, 1, t^{-1})$.

54.8. Evaluate $\int_0^\infty y^n e^{-xy}\,dy$ for $x > 0$ by repeated differentiation with respect to x of $\int_0^\infty e^{-xy}\,dy$.

54.9. Try to minimize the function $u(x) = x_1^2 + x_2^2 + 2x_3^2$ by starting at $x = (1,1,1)$ using the method of steepest descent. Seek the largest step length for which the iteration converges.

54.10. Compute the roots of the equation $(x_1^2 - x_2^2 - 3x_1 + x_2 + 4, 2x_1x_2 - 3x_2 - x_1 + 3) = (0,0)$ using Newton's method.

54.11. Generalize Taylor's theorem for a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ to third order.

54.12. Is the function $f(x_1, x_2) = \dfrac{x_1^2 - x_2^2}{x_1^2 + x_2^2}$ Lipschitz continuous close to $(0,0)$?

> Jacobi and Euler were kindred spirits in the way they created their mathematics. Both were prolific writers and even more prolific calculators; both drew a great deal of insight from immense algorithmical work; both laboured in many fields of mathematics (Euler, in this respect, greatly surpassed Jacobi); and both at any moment could draw from the vast armoury of mathematical methods just those weapons which would promise the best results in the attack of a given problem. (Sciba)